



Universidade de Vigo

Trabajo Fin de Máster

---

# Análisis y Caracterización de las Estadísticas de Accidentes Laborales en España

---

María Navarro Jamarido

Máster en Técnicas Estadísticas

Curso 2022-2023



## Propuesta de Trabajo Fin de Máster

<p><b>Título en galego:</b> Análise e Caracterización das Estatísticas de Accidentes Laborais en España</p>
<p><b>Título en español:</b> Análisis y Caracterización de las Estadísticas de Accidentes Laborales en España</p>
<p><b>English title:</b> Analysis and Characterization of the Labor Accidents' Statistics in Spain</p>
<p><b>Modalidad:</b> Modalidad B</p>
<p><b>Autor/a:</b> María Navarro Jamardo, Universidade de Santiago de Compostela</p>
<p><b>Director/a:</b> Marta Sestelo Pérez, Universidade de Vigo</p>
<p><b>Tutor/a:</b> Daniel García Coego, Gradiant</p>
<p><b>Breve resumen del trabajo:</b></p> <p>En los últimos tiempos, la situación de los accidentes laborales en España se está agravando. Cada año el número de este tipo de accidentes aumenta y no se observan señales de que la tendencia vaya a cambiar.</p> <p>Este trabajo consiste en la elaboración de un informe sobre la accidentalidad laboral en España, examinando y exponiendo los resultados obtenidos a partir de la información proporcionada por el Gobierno a través de la web del Ministerio de Trabajo. El análisis de los datos tiene sobre todo carácter descriptivo; aunque también se incluyen análisis y predicciones a través del estudio de series temporales.</p>
<p><b>Recomendaciones:</b></p>
<p><b>Otras observaciones:</b></p>



Doña Marta Sestelo Pérez, Profesora ayudante de doctor de la Universidad de Vigo y don Daniel García Coego, Responsable de Industria en el Área de Sistemas Inteligentes de Gradiant informan que el Trabajo Fin de Máster titulado

**Análisis y Caracterización de las Estadísticas de Accidentes Laborales en España**

fue realizado bajo su dirección por doña María Navarro Jamardo para el Máster en Técnicas Estadísticas. Estimando que el trabajo está terminado, dan su conformidad para su presentación y defensa ante un tribunal.

En Santiago de Compostela, a 27 de enero de 2023.

La directora:  
Doña Marta Sestelo Pérez

El tutor:  
Don Daniel García Coego

La autora:  
Doña María Navarro Jamardo

---

**Declaración responsable.** Para dar cumplimiento a la Ley 3/2022, de 24 de febrero, de convivencia universitaria, referente al plagio en el Trabajo Fin de Máster (Artículo 11, [Disposición 2978 del BOE núm. 48 de 2022](#)), **el/la autor/a declara** que el Trabajo Fin de Máster presentado es un documento original en el que se han tenido en cuenta las siguientes consideraciones relativas al uso de material de apoyo desarrollado por otros/as autores/as:

- Todas las fuentes usadas para la elaboración de este trabajo han sido citadas convenientemente (libros, artículos, apuntes de profesorado, páginas web, programas, . . .)
- Cualquier contenido copiado o traducido textualmente se ha puesto entre comillas, citando su procedencia.
- Se ha hecho constar explícitamente cuando un capítulo, sección, demostración, . . . sea una adaptación casi literal de alguna fuente existente.

Y, acepta que, si se demostrara lo contrario, se le apliquen las medidas disciplinarias que correspondan.



# Índice general

<b>Resumen</b>	<b>IX</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Gradiant . . . . .	1
1.2. Motivación y Objetivos . . . . .	2
1.3. Descripción del proyecto . . . . .	4
<b>2. Construcción de los <i>datasets</i></b>	<b>7</b>
2.1. Descripción de los datos del Ministerio de Trabajo . . . . .	7
2.1.1. Descripción de los datos del INE . . . . .	12
2.2. Construcción de los nuevos <i>datasets</i> . . . . .	13
2.2.1. Esquema de los nuevos <i>datasets</i> . . . . .	13
2.2.2. Proceso de construcción y problemática durante el proceso . . . . .	18
<b>3. Análisis descriptivo de los datos sobre accidentes laborales en España</b>	<b>23</b>
3.1. Estado del arte . . . . .	23
3.2. Análisis de los accidentes laborales en España . . . . .	25
3.2.1. Características del accidentado . . . . .	28
3.2.2. Características de la empresa . . . . .	41
3.2.3. Características del accidente . . . . .	49
3.2.4. Características de la temporalidad . . . . .	73
<b>4. Análisis de las series temporales de accidentes laborales en España</b>	<b>83</b>
4.1. Modelos y técnicas empleadas . . . . .	84
4.1.1. Modelos de series temporales . . . . .	84
4.1.2. <i>Clustering</i> de series temporales . . . . .	88
4.2. Modelado y predicción del número de accidentes . . . . .	91
4.3. Agrupación de las series por provincias y sectores económicos . . . . .	99
4.3.1. Agrupación de las provincias . . . . .	104
4.3.2. Agrupación de los sectores económicos . . . . .	107
4.3.3. Agrupación conjunta . . . . .	111
<b>5. Conclusión</b>	<b>121</b>
<b>Bibliografía</b>	<b>127</b>



# Resumen

## Resumen en español

En los últimos tiempos, la situación de los accidentes laborales en España se está agravando. Cada año el número de este tipo de accidentes aumenta y no se observan señales de que la tendencia vaya a cambiar. Conviene, por tanto, que los especialistas en prevención de riesgos laborales dediquen sus esfuerzos a resolver esta situación y dispongan de herramientas que faciliten su labor. Con este trabajo se pretende informar de cuál es el estado real de los accidentes laborales en el país.

Este trabajo consiste en la construcción de unos *datasets* unificados y la elaboración de un informe sobre la accidentalidad laboral en España, examinando y exponiendo los resultados obtenidos a partir de la información proporcionada por el Gobierno a través de la web del Ministerio de Trabajo. El análisis de los datos tiene sobre todo carácter descriptivo, aunque también se incluyen análisis y predicciones a través del estudio de series temporales.

El informe, es decir, este propio documento, será público y se complementará con el acceso a gráficas interactivas y archivos de datos reorganizados a partir de los del Ministerio.

## English abstract

Lately, in the last decade, work accidents in Spain have grown remarkably. Every year the number of these types of accidents increases and there are no signs of a change of trend. In sight of this results, it's needed that specialists in occupational risk prevention dedicate their efforts to resolve this situation. To do so, it's essential that they can count with the right tools and information. In order to contribute, this work aims to inform about the real state of occupational accidents in the country.

This work consists of the construction of some unified datasets and the elaboration of a report on the number of work accidents in Spain, examining and exposing the results obtained from the information provided by the spanish Government through the website of the Ministry of Labour. The data analysis is mainly descriptive, although analysis and predictions through the study of time series are also included.

This report will be public and will be complemented by access to interactive graphics and data files reorganized from those of the Ministry.



# Capítulo 1

## Introducción

El presente capítulo se centra en la presentación del contexto en el que se realiza el trabajo, la motivación que lo ha impulsado y los objetivos que se pretenden lograr con él, introduciendo también los procedimientos llevados a cabo.

Como punto de partida se presenta el centro tecnológico Gradiant. Junto a su presentación, se incluye la de la línea de Industria, dentro de la cual se ha desarrollado este trabajo.

Posteriormente, se explica de forma breve la problemática causada por los accidentes laborales, tema central del trabajo, y la razón por la que es de interés estudiarlos. Se describe también el procedimiento llevado a cabo en el trabajo para analizar las estadísticas de este tipo de accidentes, que se desarrollan a lo largo del documento.

### 1.1. Gradiant

Gradiant, centro de Tecnologías de la Información y las Comunicaciones (TIC), tiene como objetivo mejorar la competitividad de las empresas mediante la transferencia de conocimiento y tecnologías en los ámbitos de la conectividad, inteligencia y seguridad. Con más de 140 profesionales y 14 patentes solicitadas, Gradiant ha desarrollado más de 340 proyectos diferentes de I+D+i, convirtiéndose en uno de los principales motores de la innovación en Galicia.

El Centro fue creado en 2008 y se conforma a partir de un patronato que agrupa a representantes del sector público y privado. Está formado por las Universidades de A Coruña, Santiago de Compostela y Vigo; las empresas Abanca, Altia, Arteixo Telecom, Egatel, Indra, Plexus, R, Telefónica, Televés, y la Asociación empresarial INEO.

El compromiso del centro con la calidad es una constante desde sus inicios. El Centro cuenta con los siguientes certificados: Sistema de Gestión de Calidad UNE-EN ISO 9001:2015, Sistema de Gestión de Proyectos de I+D+i UNE 166002:2014, Sistemas de Gestión de la Seguridad de la Información UNE-EN ISO/IEC 27001:2013. Además, forma parte del registro estatal de Centros de Innovación Tecnológica (Sello CIT).

Tras 14 años de actividad, Gradiant se sitúa como socio tecnológico de la industria orientado a sus necesidades en el ámbito de las TIC, aportando su experiencia nacional e internacional en tecnologías para la seguridad y la privacidad; el procesado de señales multimedia; internet de las cosas; la biometría y analítica de datos y los sistemas de comunicaciones avanzadas.

En cuanto a la línea de Industria del centro, donde se engloba este trabajo, su misión se encuentra totalmente alineada con la de Gradiant, pero aplicada al sector manufacturero. Por tanto, se puede definir como “contribuir al dinamismo innovador, el crecimiento y la mejora competitiva del tejido industrial gallego a través del desarrollo tecnológico y la innovación en el uso de las TIC”.

Su filosofía es ayudar a las empresas a generar negocio a través del aumento de su competitividad mediante la transformación digital, generando mejoras en sus procesos y facilitando la creación de nuevos modelos de negocio. Integran la transferencia de conocimiento y diseñan soluciones especializadas para la industria, sumando el valor añadido de los profesionales del Centro. Además, siendo completamente independientes, por lo que aportan una visión neutral acerca de las mejores soluciones para cubrir los objetivos de las empresas.

La idea propuesta inicialmente desde Gradient consistía en simplificar el acceso a los datos de accidentalidad laboral aportados por el Ministerio de Trabajo con la intención de facilitar su análisis. Especialmente, se buscaba proporcionar un método más adecuado para comparaciones temporales, ya que los datos originales se encuentran separados en ficheros anuales, de complejo análisis de forma conjunta. Por tanto, el objetivo inicial era construir un *dataset* unificado de dicha información, con el objetivo de usarlo para realizar un estudio pormenorizado de los datos y abrir tanto este *dataset* como los resultados de los análisis realizados al público general. Este objetivo se encuentra enmarcado dentro del plan estratégico de la línea de Industria de Gradient, que contempla actuaciones de apoyo a la innovación en el sector en el ámbito de procesos, negocio y personas, conteniéndose este trabajo en este último punto.

La idea principal de proyecto se ha mantenido, creando varios *datasets* unificados para los años 2014-2021, presentando también un informe de los accidentes laborales en el territorio español en el cual se muestran estadísticas que reflejan la situación de los últimos años. Tras estudiar los datos disponibles, se decidió incluir también otros tipos de análisis en el trabajo. De esta forma se presenta un informe acerca de los accidentes más completo de lo usual, incluyendo técnicas estadísticas más complejas que el análisis descriptivo. De todas formas, no se ha perdido la noción del objetivo inicial, que era presentar un informe público y útil acerca de la situación en España de los accidentes laborales.

## 1.2. Motivación y Objetivos

El trabajo conlleva de forma intrínseca riesgos para la seguridad y salud de los trabajadores. Estos riesgos son un inconveniente de mucha importancia que están presentes, en mayor o menor medida, en todos los puestos de trabajo. La exposición a estos potenciales peligros puede traer consigo accidentes que conllevan pérdidas (de material, económicas...) para las empresas y sobre todo lesiones, enfermedades e incluso la muerte de trabajadores.

Eliminar los riesgos en el ambiente laboral es una tarea imposible, pero sí que es posible reducirlos en general. Para ello, es imprescindible realizar un estudio adecuado de los mismos, para conocerlos y poder diseñar actuaciones preventivas dirigidas a limitar los potenciales peligros y preparar el espacio de trabajo para, en el caso de que ocurra el accidente, poder mitigar sus daños lo máximo posible.

El Gobierno de España, desde el Ministerio de Trabajo, proporciona información sobre los accidentes laborales que han tenido lugar en el país desde 1988 y han provocado la baja del trabajador. A medida que han ido avanzando los años, la información acerca de los accidentes proporcionada por el Ministerio ha sido cada vez más completa. Además, desde el año 2014 hasta la actualidad se presenta la información de forma más compacta y ordenada, en un formato más accesible.

Estos datos, que son registros oficiales, muestran una tendencia de aumento en el número de accidentes laborales en España en la última década. Reflejan también la disminución de accidentes debida al confinamiento en 2020, pero tras ellas se puede observar una nueva subida en 2021 y números en 2022 que se proyectan superiores a los del año anterior. Esta situación contrasta alarmanamente con la general de la UE, donde este tipo de incidentes han ido disminuyendo cada año desde 2010. Muchas organizaciones, como FUNCAS <sup>1</sup> o los propios gobiernos españoles y de sus regiones (mediante los in-

---

<sup>1</sup>Fundación Cajas de Ahorros. Última consulta el 29 de noviembre de 2022 [online](#).

formas oficiales, como ESENER (2020)<sup>2</sup>, y medios de comunicación (como por ejemplo RTVE <sup>3</sup>, entre muchos otros) se han hecho eco de esta situación, poniendo el foco sobre todo en los accidentes mortales.

Desde Gradient, como medida para ofrecer ayuda a profesionales del mundo de la prevención de riesgos laborales (PRL), se ha propuesto realizar un estudio de los datos anteriormente mencionados proporcionados por el Ministerio de Trabajo. La intención es analizar dichos datos y facilitar información a partir de ellos que proporcione un mejor conocimiento de la situación.

Este trabajo presenta diferentes objetivos, todos con el mismo fin de obtener una información más clara y útil sobre la accidentalidad laboral en España. Se enumeran a continuación:

1. Creación de un *dataset* unificado, que permite acceder a toda la información desde un mismo archivo.

Dado que los archivos proporcionados por el Ministerio cuentan con una gran cantidad de datos, que además no están ordenados siguiendo siempre las mismas categorías; en lugar de un solo *dataset* se construyen 3 archivos. Estos se han organizado por gravedad del accidente, territorio en el que tuvo lugar el accidente (CC.AA. y provincias) o sector económico al que pertenece la empresa del accidentado.

2. Creación de un documento informativo sobre los factores más relevantes en los accidentes laborales, basándose en la información proporcionada por los *datasets* construidos. Este informe será público y estará a disposición de todos los interesados en informarse sobre el tema.

Con esta idea se pretende proporcionar información útil, no solo en forma de tablas de datos, sino a través de estadísticas y análisis que puedan ser interesantes de cara a la prevención de riesgos laborales.

3. Construcción de gráficos (algunos incluidos en el informe y otros, interactivos, a los que se proporciona acceso) para complementar la información dada y poder así visualizar los datos de los que se dispone.

De forma ajena a la motivación específica de ayudar a reducir los accidentes en el ámbito laboral, desde el punto de vista académico este trabajo presenta también características interesantes. Por un lado, el tratamiento de la gran cantidad de datos proporcionados por el Ministerio permite enfrentarse con un problema de gran tamaño, una situación que dista bastante de los que se habitúan a ver en clase. A pesar de ello, muchas de las técnicas vistas a lo largo del máster son aplicables a estos datos, por lo que se presenta una oportunidad de aplicar técnicas aprendidas en diferentes asignaturas a un mismo proyecto.

Además, la construcción de los *datasets* y el análisis de los datos se ha hecho mediante el lenguaje de programación *Python* (Van Rossum and Drake 1995); una herramienta de mucha utilidad y comúnmente utilizada en la actualidad en este ámbito. Por tanto, realizando este trabajo se ha presentado la oportunidad de aprender un lenguaje de programación nuevo, diferente de los aprendidos a lo largo del Grado de Matemáticas o el propio Máster en Técnicas Estadísticas.

Para finalizar con la presentación del trabajo, cabe destacar que los datos de los que se dispone no contienen la información de cada accidente, sino que se presentan agrupados. Los grupos siguen diferentes categorías: grupos de edad de los trabajadores, sexo de los trabajadores, comunidad autónoma donde se encuentra la empresa, número de trabajadores de la empresa, etc. Esta característica limita las posibilidades del análisis, pues para empezar se está limitando la información, ya que no se dispone de intersecciones entre los datos; por ejemplo: es conocido el total de accidentes de mujeres y el total de accidentes por golpes en la cabeza, pero no se conoce (a través de estos datos) cuál es el

---

<sup>2</sup>Encuesta Europea de Empresas sobre Riesgos Nuevos y Emergentes. Última vez consultado el 29 de noviembre de 2022 [online](#).

<sup>3</sup>Radio Televisión Española. Última consulta el 29 de noviembre de 2022 [online](#).

total de mujeres accidentadas por golpes en la cabeza. Debido a esta situación, el trabajo se centrará principalmente en procesar y explicar los datos de una forma clara, analizando y representando los datos de forma que se pueda sacar el mayor provecho de ellos.

En el siguiente apartado se expone el planteamiento seguido durante el trabajo.

### 1.3. Descripción del proyecto

Tal y como se ha avanzado, la idea de este trabajo es resumir y analizar la información oficial proporcionada por el Ministerio de Trabajo acerca de los accidentes laborales en España, con la intención de presentar unos resultados útiles, fáciles de interpretar y que resulten de interés para los especialistas en PRL, las empresas que pretendan mejorar su gestión de riesgos laborales o para cualquier persona que pretenda informarse en el asunto.

A continuación se nombran y describen las etapas llevadas a cabo para el análisis de los datos:

#### ■ Tratamiento y acondicionamiento de los datos:

Tras la descarga de los archivos del Ministerio mediante el uso de un código de *Python* (el cual se puede modificar para descargar los documentos de años posteriores o anteriores), se ha procedido a la construcción de los tres *datasets* con los que se ha trabajado posteriormente. Estos se organizan según el nivel de gravedad del accidente, según la comunidad autónoma y provincia donde ha ocurrido el accidente y según el sector, sección o división al que pertenece la empresa del trabajador accidentado. En este proceso se han eliminado las hojas de los archivos Excel que no utilizamos, se han depurado los *NAs* presentes en dichos archivos y se han reestructurado los datos de forma que han quedado ordenados por años y, dentro de ellos, por diferentes categorías. Los tres nuevos archivos de datos se exportan finalmente a formato *.parquet*<sup>4</sup>, un formato *orientado-a-columna* diseñado para almacenar tablas de datos con un mayor número de columnas que filas (como es el caso tratado).

Durante este procedimiento se han encontrado diversas dificultades. Estos inconvenientes se explican en el siguiente capítulo, donde se expone detalladamente como ha sido el proceso de creación de los tres *datasets*.

Adicionalmente, también se han extraído archivos de la web del Instituto Nacional de Estadística (INE). Dichos documentos contienen información sobre la población activa y ocupada según sexo, grupo de edad y comunidad autónoma o sector a lo largo de los años desde 2014 hasta 2021. Gracias a esto, se han podido normalizar gran parte de los datos, en lugar de observar solo los valores absolutos. Este procedimiento permite observar los datos desde una perspectiva diferente, facilitando estudiar el porcentaje de accidentes por población ocupada. De esta forma se pueden establecer comparaciones respecto al número de accidentes entre diferentes categorías.

#### ■ Análisis y representación de los datos:

Una vez construidos los tres archivos de datos, se procede con el análisis de estos. Para ello, la intención ha sido dar respuesta a las preguntas más frecuentes a la hora de investigar un accidente laboral (teniendo en cuenta los datos disponibles).

Tal y como ya ha sido indicado, no se dispone de los datos por accidente, con lo cual el análisis se ha centrado mayoritariamente en calcular medidas de localización y dispersión de los datos, en representarlos de forma que se facilite su visualización y en clasificarlos según diferentes razonamientos, buscando relacionarlos entre ellos.

El análisis de los accidentes se ha estructurado de forma que responda a las siguientes cuestiones relacionadas con el accidente:

---

<sup>4</sup>Información [online](#) donde se pueden revisar las principales características del formato *.parquet*.

- Perfil de la persona accidentada.
- Perfil de la empresa a la que pertenece la persona accidentada.
- Características del accidente (tipo de lesión y materiales que la provocaron).
- Temporalidad del accidente.

Además, por medio del tratamiento de series de tiempo, se han llevado a cabo otros procesos. En primer lugar, se ha construido la serie de tiempo mensual desde enero de 2014 a diciembre de 2021 en la que se representa el número de accidentes laborales totales. A partir de ella se han obtenido predicciones para los meses futuros y se ha contrastado el resultado con los valores avanzados de 2022. En segundo lugar, se han construido las series de tiempo mensuales para cada una de las provincias españolas y para cada una de las secciones económicas consideradas en los archivos de datos. Posteriormente, se ha procedido a la agrupación de estas mediante métodos de *clustering*. El objetivo es organizar y agrupar las provincias y las secciones en función de sus similitudes respecto al número de accidentes laborales, facilitando así su gestión. De esta forma, se podrán aplicar medidas de prevención similares a aquellas provincias que presenten datos similares, por ejemplo. Además, esta agrupación también permite relacionar provincias con sectores económicos en función del número de accidentes laborales. Para este último caso, en lugar de considerar las secciones económicas, se han considerado las divisiones.

■ **Extracción de conclusiones:**

Finalmente, tras realizar todos los análisis mencionados en el apartado anterior, se describen las conclusiones extraídas de una forma organizada.

En este capítulo se recogen de forma conjunta y resumida los resultados vistos a lo largo del informe. La idea principal es que los análisis realizados para las diferentes categorías se puedan complementar entre ellos, para aportar una información más completa de la situación. Lógicamente no se repiten de nuevo todos los resultados obtenidos, pero sí se destacan los más relevantes y aquellos que se entienden mejor acompañados de otros resultados relacionados.

Finalmente, no a modo de conclusión, pero sí como finalidad del propio trabajo, el informe se hará público para que disponga de él toda persona interesada en la materia.



## Capítulo 2

# Construcción de los *datasets*

Este capítulo se centra en explicar de forma detallada como se han construido los *datasets* a partir de los cuales se ha realizado el análisis. En primer lugar, se describe el contenido de los datos de partida, proporcionados por el Ministerio en formato Excel. Posteriormente, se comenta el proceso seguido con el lenguaje de programación *Python* para transformar dichos archivos de Excel en los archivos de *parquet* finales, haciendo énfasis en los problemas encontrados a lo largo del proceso.

### 2.1. Descripción de los datos del Ministerio de Trabajo

Tal y como se ha ido avanzando, los datos utilizados para el análisis de accidentes laborales en España han sido obtenidos mayoritariamente a partir de la web<sup>1</sup> del Ministerio de Trabajo. Los datos consisten en 2 tipos de archivos de Excel, cada uno de ellos para los años entre 2014 y 2021. Dichos archivos de Excel se pueden encontrar en la web, seleccionando el año deseado, en los apartados A y D.

El estudio se ha limitado a los años indicados por una cuestión organizativa. Para los años anteriores a 2014 el formato en el que fueron almacenados los datos era diferente y, además, algunas de las características en las que se agrupan los accidentes no fueron consideradas hasta después del 2014. Por este motivo, considerar los datos de años previos al indicado, provocaría que los nuevos *datasets* creados tuviesen una gran cantidad de valores faltantes. En el caso de solo considerar las categorías presentes en todos los años desde 1988, se estaría perdiendo mucha información. Por tanto, la decisión que se tomó fue considerar los 8 años más recientes. Además, estos son los más cercanos a la situación actual y futura, que es la que interesa analizar y conocer mayormente.

A continuación se describe el contenido de los archivos de Excel obtenidos de la web del Ministerio:

- Archivos tipo A (ATR\_año\_A.xls): consisten en documentos de Excel con 29 hojas en cada uno de ellos (contando portada e índice). En cada una de las hojas con información del documento se puede encontrar una tabla en la cual se indican el número de accidentes leves, graves, mortales y totales que han tenido lugar, tanto durante la jornada como *in itinere*. Cada una de las tablas del documento clasifica estos accidentes según diferentes temáticas.

A modo de ejemplo se muestra en la Figura 2.1 una de las hojas de los documentos.

A continuación, se muestra una lista con todas las clasificaciones que se pueden encontrar en los archivos tipo A para todos los años seleccionados. En todo caso dan lugar a tablas que indican el número de accidentes según su gravedad (en jornada o *in itinere*) clasificados según la temática correspondiente:

---

<sup>1</sup>Estadísticas de accidentes de trabajo, Ministerio de Trabajo, Gobierno de España. Última vez consultado el 19 de enero de 2023 [online](#).

- 1.1 Por comunidad autónoma y provincia
- 1.2 Por sector, sección y división de actividad económica
- 2.1 Por sexo y edad del trabajador accidentado
- 2.2 Por situación profesional del trabajador accidentado
- 2.3 Por ocupación del trabajador accidentado
- 2.4 Por tipo de contrato y por antigüedad del puesto de trabajo
- 2.5 Por país de nacionalidad del trabajador accidentado
- 3.1 Por lugar del accidente (Figura 2.1)
- 3.2 Por tipo de empresa
- 3.3 Por organización preventiva de la empresa y evaluación de riesgos sobre el puesto
- 3.4 Por tamaño del centro
- 4.1 Por tipo de lugar donde se encontraba el trabajador accidentado
- 4.2 Por tipo de trabajo que realizaba el trabajador accidentado
- 4.3 Por actividad física específica que desempeñaba el trabajador accidentado
- 4.4 Por desviación que produjo el accidente
- 4.5 Por agente material de la desviación que produjo el accidente
- 4.6 Por forma o contacto que produjo la lesión
- 4.7 Por agente material de la forma o contacto que produjo la lesión
- 5.1 Por hora del día de ocurrencia del accidente
- 5.2 Por hora de la jornada laboral de ocurrencia del accidente
- 5.3 Por día de la semana de ocurrencia del accidente
- 5.4 Por mes de ocurrencia del accidente
- 6.1 Por descripción de la lesión
- 6.2 Por parte del cuerpo lesionada
- R.1 Recaídas por comunidad autónoma y provincia
- R.2 Recaídas por sector, sección y división de actividad económica

	A	B	C	D	E	F	G	H	I	J
1	<b>ACCIDENTES DE TRABAJO</b>									
2										
3	<b>ATR-A.3.1. ACCIDENTES DE TRABAJO CON BAJA, EN JORNADA E IN ITINERE, SEGÚN GRAVEDAD,</b>									
4	<b>POR LUGAR DEL ACCIDENTE</b>									
5										
6	<b>Año 2014</b>									
7		<b>EN JORNADA</b>				<b>IN ITINERE</b>				
8		<b>Total</b>	<b>Leves</b>	<b>Graves</b>	<b>Mortales</b>	<b>Total</b>	<b>Leves</b>	<b>Graves</b>	<b>Mortales</b>	
9	<b>TOTAL</b>	424.625	420.829	3.329	467	66.474	65.477	884	113	
10										
11	En el centro o lugar de trabajo habitual	365.938	363.438	2.276	224	-	-	-	-	
12	En otro centro o lugar de trabajo	18.741	18.271	396	74	-	-	-	-	
13	En desplazamiento en su jornada laboral	39.946	39.120	657	169	-	-	-	-	
14	Aj ir o volver del trabajo ( <i>in itinere</i> )	-	-	-	-	66.474	65.477	884	113	
15										
16	Se incluyen los accidentes que han causado baja en el periodo de referencia, <i>repcionados</i> por la Autoridad Laboral Provincial (véanse Fuentes y Notas Explicativas).									
17										

Figura 2.1: Hoja del documento *ATR\_2014\_A.xls* que muestra los accidentes de trabajo según gravedad clasificados por el lugar donde ocurrió el accidente (4 categorías).

Estos archivos, para todos los años considerados, contienen una última hoja (**S.1**) en la que simplemente se muestra una tabla que almacena el número de accidentes laborales que no causaron baja en los últimos años (desde 2006). Esta información no ha sido considerada para el análisis, pues toda la información dada en los archivos de Excel, salvo esas hojas, es respecto a los accidentes que produjeron la baja laboral del empleado o empleada.

Todas las hojas consideradas de los archivos de Excel tipo A tienen en común sus columnas:

En Jornada				<i>In Itinere</i>			
Total	Leve	Grave	Mortal	Total	Leve	Grave	Mortal

Las filas para cada hoja están compuestas por las categorías de cada temática. Por ejemplo, para la hoja **3.1**, se puede observar en la Figura 2.1 que las filas se corresponden con las categorías por lugar de accidente: en el centro o lugar de trabajo habitual, en otro centro o lugar de trabajo, en desplazamiento en su jornada laboral y al ir o volver del trabajo (*in itinere*).

Teniendo en cuenta que esta estructura se repite para los archivos de Excel de todos los años considerados, para todos ellos podemos extraer información sobre los accidentes laborales para las mismas categorías; en función también de la gravedad del accidente y si es en jornada o *in itinere*.

- Archivos tipo D (ATR\_año\_D.xls): consisten en documentos de Excel con 55 hojas en cada uno de ellos (contando portada e índice). Las 53 hojas con datos se dividen en 14 grupos y dentro de cada uno de esos grupos las hojas comparten las mismas columnas, pero varía la clasificación por filas.

Los grupos en los que se dividen los accidentes en el documento son los siguientes:

1. Según sexo y edad
2. Según tipo de contrato
3. Según grandes grupos de ocupación
4. Según antigüedad en el puesto de trabajo
5. Según país de nacionalidad
6. Según tamaño del centro donde ocurre el accidente
7. Según organización preventiva de la empresa y evaluación de riesgos sobre el puesto de trabajo
8. Según lugar del accidente
9. Según actividad física específica que desempeñaba el trabajador accidentado
10. Según desviación que desencadenó el accidente
11. Según forma o contacto que produjo la lesión
12. Según descripción de la lesión
13. Según parte del cuerpo lesionada
14. Según mes de ocurrencia del accidente

Dentro de cada uno de los grupos, como se ha dicho ya, la clasificación por columnas coincide y es la que indica el nombre de cada grupo. Por ejemplo, para las hojas del grupo 6 (según tamaño

del centro donde ocurre el accidente), las columnas son para todas las hojas del grupo: total, hasta 1 mes, de 1 a 9 trabajadores, de 10 a 25 trabajadores, de 26 a 49 trabajadores, de 50 a 99 trabajadores, de 100 a 249 trabajadores, de 250 a 499 trabajadores, de 500 a 1000 trabajadores, más de 1000 trabajadores y no consta.

Además de la clasificación por columnas, cada una de las hojas está también organizada por otra clasificación por filas. Para cada uno de los grupos, existen diferentes clasificaciones por filas. A continuación se enumeran para cada uno de los 14 grupos:

1. **1.1** por CCAA y provincias; **1.2** por sectores, secciones y divisiones; **1.3** por ocupación del trabajador; **1.4** por antigüedad en el puesto de trabajo; **1.5** por forma o contacto que produjo la lesión; **1.6** por nacionalidad del trabajador.
2. **2.1** por CCAA y provincias; **2.2** por sectores, secciones y divisiones; **2.3** por ocupación del trabajador; **2.4** por sexo y edad; **2.5** por antigüedad en el puesto de trabajo; **2.6** por forma o contacto que produjo la lesión; **2.7** por nacionalidad del trabajador.
3. **3.1** por CCAA y provincias; **3.2** por sectores, secciones y divisiones; **3.3** por antigüedad en el puesto de trabajo; **3.4** por forma o contacto que produjo la lesión.
4. **4.1** por CCAA y provincias; **4.2** por sectores, secciones y divisiones; **4.3** por forma o contacto que produjo la lesión.
5. **5.1** por CCAA y provincias; **5.2** por sectores, secciones y divisiones.
6. **6.1** por CCAA y provincias; **6.2** por sectores, secciones y divisiones.
7. **7.1** por CCAA y provincias; **7.2** por sectores, secciones y divisiones; **7.3** por tamaño del centro.
8. **8.1** por CCAA y provincias; **8.2** por sectores, secciones y divisiones; **8.3** por ocupación del trabajador.
9. **9.1** por CCAA y provincias; **9.2** por sectores, secciones y divisiones; **9.3** por forma o contacto que produjo la lesión.
10. **10.1** por CCAA y provincias; **10.2** por sectores, secciones y divisiones; **10.3** por forma o contacto que produjo la lesión.
11. **11.1** por CCAA y provincias; **11.2** por sectores, secciones y divisiones.
12. **12.1** por CCAA y provincias; **12.2** por sectores, secciones y divisiones; **12.3** por ocupación del trabajador; **12.4** por sexo y edad; **12.5** por forma o contacto que produjo la lesión.
13. **13.1** por CCAA y provincias; **13.2** por sectores, secciones y divisiones; **13.3** por ocupación del trabajador; **13.4** por sexo y edad; **13.5** por forma o contacto que produjo la lesión.
14. **14.1** por CCAA y provincias; **14.2** por sectores, secciones y divisiones; **14.3** por ocupación del trabajador; **14.4** por sexo y edad; **14.5** por forma o contacto que produjo la lesión.

Continuando con el ejemplo antes dado, en la primera de las hojas del grupo 6 (hoja **6.1**), las columnas son las anteriormente mencionadas y las filas consisten en las comunidades autónomas y provincias del territorio español.

Se muestran, a continuación, en las Figuras 2.2 y 2.3 varios ejemplos de las hojas de este tipo de documentos. La segunda de estas figuras se corresponde con el ejemplo del grupo 6 que ya ha sido introducido.

Como se puede observar en la enumeración anterior, no todos los grupos en los que se dividen las hojas de los archivos de Excel tipo D contienen el mismo número de elementos. Véase, por ejemplo, que el grupo 1 presenta 6 hojas (cada una con una clasificación diferente) y el grupo 2 tiene 7. De todas formas, aunque algunos grupos presenten más categorías que otro, sí que hay dos que están presentes para los 14 conjuntos de hojas en las que se divide el documento: clasificación por CC.AA. y provincias y clasificación por sectores, secciones y divisiones.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	ACCIDENTES DE TRABAJO														15	ATR	
2	ATR-D.1.4. ACCIDENTES DE TRABAJO CON BAJA EN JORNADA, SEGÚN SEXO Y EDAD, POR ANTIGÜEDAD EN EL PUESTO DE TRABAJO																
3	Año 2014																
4		TOTAL	De 16 y 17 años	De 18 y 19 años	De 20 a 24 años	De 25 y 29 años	De 30 a 34 años	De 35 y 39 años	De 40 a 44 años	De 45 y 49 años	De 50 a 54 años	De 55 y 59 años	De 60 a 64 años	Más de 65 años			
8	AMBOS SEXOS	424.625	398	2.371	21.755	41.356	58.807	69.935	65.263	59.238	51.554	36.390	16.740	818			
10	Hasta 1 mes	37.079	116	734	4.147	5.371	5.907	6.190	5.176	4.135	3.069	1.592	611	31			
11	De 1 a 2 meses	38.300	100	616	4.290	5.609	6.420	6.322	5.349	4.272	3.047	1.637	600	38			
12	De 3 a 4 meses	25.023	50	310	2.643	3.631	4.219	4.194	3.603	2.745	2.085	1.111	403	29			
13	De 5 a 7 meses	24.354	48	254	2.343	3.548	4.054	4.166	3.510	2.762	2.048	1.134	455	32			
14	De 8 a 12 meses	24.460	41	191	2.207	3.277	4.104	4.246	3.537	2.896	2.271	1.204	453	33			
15	De 1 a 3 años	57.434	32	230	3.903	7.583	9.378	10.381	8.445	6.964	5.522	3.409	1.492	95			
16	Más de 3 años y hasta 10 años	124.733	8	31	2.202	11.788	20.133	23.304	20.185	17.636	14.743	9.992	4.432	279			
17	Más de 10 años	93.242	3	5	20	549	4.592	11.132	15.458	17.828	18.769	16.311	8.294	281			
19	VARONES	291.408	320	1.713	15.484	29.354	42.886	51.324	46.607	39.494	31.745	22.206	9.856	419			
21	Hasta 1 mes	26.759	84	526	2.979	3.872	4.415	4.665	3.805	2.908	2.052	1.038	398	17			
22	De 1 a 2 meses	26.878	77	410	2.911	3.866	4.660	4.741	3.836	2.935	2.020	1.040	362	20			
23	De 3 a 4 meses	17.486	38	215	1.847	2.515	3.064	3.092	2.584	1.881	1.321	686	230	13			
24	De 5 a 7 meses	17.090	46	182	1.663	2.513	2.954	3.081	2.531	1.862	1.259	709	271	19			
25	De 8 a 12 meses	17.241	38	151	1.563	2.350	3.031	3.156	2.603	1.917	1.437	730	249	16			
26	De 1 a 3 años	39.291	30	203	2.874	5.378	6.845	7.441	5.840	4.540	3.242	2.018	835	45			
27	Más de 3 años y hasta 10 años	79.938	4	24	1.633	8.403	14.253	16.403	13.399	10.345	7.748	5.262	2.331	133			
28	Más de 10 años	66.725	3	2	14	457	3.664	8.745	12.009	13.106	12.666	10.723	5.180	158			

Figura 2.2: Hoja D.1.4 del documento *ATR\_2014\_D.xls* que muestra los accidentes de trabajo según edad y sexo clasificado por antigüedad en el puesto de trabajo.

	A	B	C	D	E	F	G	H	I	J	K	L	M		
1	ACCIDENTES DE TRABAJO													14	ATR
2	ATR-D.6.1. ACCIDENTES DE TRABAJO CON BAJA EN JORNADA, SEGÚN TAMAÑO DEL CENTRO DONDE OCURRE EL ACCIDENTE, POR COMUNIDAD AUTÓNOMA Y PROVINCIA														
3	Año 2014														
4		TOTAL	De 1 a 9 trabajad.	De 10 a 25 trabajad.	De 26 a 49 trabajad.	De 50 a 99 trabajad.	De 100 a 249 trabajad.	De 250 a 499 trabajad.	De 500 a 1000 trabajad.	Más de 1000 trabajad.	No consta				
8	TOTAL	424.625	99.935	75.082	52.236	49.555	55.091	30.337	18.950	33.136	10.303				
10	ANDALUCÍA	70.775	19.175	12.896	8.111	7.994	8.847	5.597	2.762	3.526	1.867				
11	Almería	6.395	1.642	977	770	707	952	545	382	155	265				
12	Cádiz	9.594	2.261	1.587	1.005	1.093	1.362	1.114	492	418	252				
13	Córdoba	7.204	2.373	1.508	925	796	845	363	173	123	198				
14	Granada	5.731	1.679	1.119	731	668	469	346	288	202	229				
15	Huelva	5.966	1.327	981	676	737	889	763	260	216	117				
16	Jaén	4.920	1.882	1.095	522	397	321	243	193	125	142				
17	Málaga	13.489	3.456	2.425	1.450	1.533	1.719	1.237	436	914	319				
18	Sevilla	17.476	4.555	3.204	2.132	2.063	2.270	986	548	1.373	345				
19	ARAGÓN	11.267	2.704	2.040	1.594	1.359	1.275	691	466	852	286				
21	Huesca	2.134	699	480	316	308	124	92	56	4	55				
22	Teruel	1.126	292	192	152	164	135	90	33	-	68				
23	Zaragoza	8.007	1.713	1.368	1.126	887	1.016	509	377	848	163				
24	ASTURIAS (PRINCIPADO DE)	10.388	2.105	1.948	1.224	1.293	1.187	1.174	595	347	515				
26	BALEARES (ILLES)	14.766	3.760	2.577	1.965	1.972	1.825	1.118	682	688	179				
28	CANARIAS	21.024	4.783	3.525	2.542	2.345	3.328	1.920	831	1.402	348				
29	Palmas (Las)	11.437	2.609	2.002	1.418	1.313	1.905	1.004	388	640	158				
30	S. C. Tenerife	9.587	2.174	1.523	1.124	1.032	1.423	916	443	762	190				
32	CANTABRIA	4.643	1.147	919	547	534	628	290	118	232	228				
34	CASTILLA-LA MANCHA	18.222	5.855	3.228	2.124	2.001	2.560	1.030	385	245	794				
35	Albacete	3.029	1.043	540	313	295	414	226	38	25	135				
36	Ciudad Real	4.200	1.607	777	472	415	430	223	36	65	175				
37	Cuenca	1.839	631	323	223	205	180	45	93	15	124				

Figura 2.3: Hoja D.6.1 del documento *ATR\_2014\_D.xls* que muestra los accidentes de trabajo según el número de trabajadores en la empresa clasificado por CC.AA. y provincias.

Dado que se pretende construir un *dataset* organizado que contenga toda la información posible en una misma tabla, se decidió considerar para cada grupo la clasificación por columnas correspondiente y solamente las clasificaciones por CC.AA. y provincias y clasificación por sectores, secciones y divisiones para las filas. De esta forma se tiene la misma información para cada grupo y es una forma de homogeneizar la organización del *dataset* que se construye. Además, la información que se está omitiendo no está del todo perdida, pues a partir de los archivos de Excel tipo A se han extraído tablas con esas clasificaciones (por ocupación del trabajador, antigüedad en el puesto de trabajo, forma de contacto que produjo la lesión, etc.) pero en función de la gravedad del accidente en lugar de las CC.AA. y provincias y de sectores, secciones y divisiones.

### 2.1.1. Descripción de los datos del INE

El análisis se realizará mayormente a partir de los datos descritos en el sub-apartado anterior. Para complementar esa información y tratar de obtener unas mejores conclusiones, ya se ha comentado que también se utilizarán los datos de población activa y ocupada de España entre los años 2014 y 2021.

La población activa de un país está compuesta por todos los habitantes en edad laboral que, o bien trabajan en un empleo remunerado (población ocupada), o bien se hallan en plena búsqueda de empleo (población en paro). Por tanto, la población activa se divide en empleados y desempleados, siendo los primeros los que componen la población ocupada.

Con estos datos se ha podido normalizar los accidentes totales en función de la población activa/ocupada de cada sexo o grupo de edad y poder así establecer comparaciones. Resulta interesante conocer la proporción de accidentes por población ocupada/activa; ya que, por ejemplo, resulta más interesante comparar la proporción de accidentes de mujeres con la proporción de hombres en función de la población ocupada/activa; pues comparar números totales mostraría una cantidad mucho mayor de accidentes de varones, lo cual viene dado en gran parte porque hay una mayor cantidad de varones ocupados/activos que de mujeres. De igual forma, también conviene expresar los resultados para las diferentes franjas de edad mediante proporciones.

Los datos de población ocupada y activa han sido extraídos de la página web del INE <sup>2</sup>. A continuación se mostrará y comentará la estructura del archivo del que se han obtenido los datos.

	A	B	C	D	E	F	G	H	I
1	Media de los cuatro trimestres del año								
2	Ocupados								
3									
4	Ocupados por sexo y grupo de edad. Valores absolutos y porcentajes respecto del total de cada sexo								
5	Unidades: Miles Personas								
6									
7	Ambos sexos								
8	Valor absoluto								
9		2021	2020	2019	2018	2017	2016	2015	2014
10	Total	19.773,6	19.202,4	19.779,3	19.327,7	18.824,8	18.341,5	17.866,0	
11	De 16 a 19 años	126,3	106,4	150,5	134,0	117,9	95,8	83,8	
12	De 20 a 24 años	869,0	777,5	888,2	855,3	803,2	724,5	718,9	
13	De 25 a 29 años	1.660,8	1.594,5	1.701,7	1.673,0	1.626,1	1.592,5	1.577,6	
14	De 30 a 34 años	2.017,0	1.979,5	2.119,1	2.113,9	2.131,9	2.183,8	2.262,0	
15	De 35 a 39 años	2.413,3	2.444,7	2.664,2	2.720,4	2.801,2	2.853,1	2.862,5	
16	De 40 a 44 años	3.030,5	3.020,0	3.130,7	3.127,2	3.039,9	2.938,6	2.824,3	
17	De 45 a 49 años	3.045,8	2.931,6	2.948,8	2.847,6	2.730,9	2.662,3	2.537,8	
18	De 50 a 54 años	2.714,6	2.638,4	2.649,7	2.550,1	2.455,8	2.340,7	2.255,9	
19	De 55 a 59 años	2.248,3	2.193,7	2.140,8	2.039,4	1.926,7	1.838,4	1.705,3	
20	De 60 a 64 años	1.367,4	1.271,3	1.174,3	1.075,4	1.015,0	952,9	889,4	
21	De 65 a 69 años	224,2	192,7	156,9	144,3	134,7	123,8	115,6	
22	70 y más años	56,3	52,1	54,5	47,0	41,5	35,0	33,0	
23									
24									
25	Notas:								
26									
27	Fuente:								
28	Instituto Nacional de Estadística								

Figura 2.4: Primeras columnas del archivo de Excel que almacena el número de población ocupada según grupos de edad y sexo.

En la Figura 2.4 se presentan las primeras columnas del archivo de Excel en el que se almacena la cantidad de población ocupada a lo largo de los años (2014-2021) clasificada según grupos de edad y sexo. Este primer bloque de columnas se corresponde con la clasificación por franjas de edad considerando ambos sexos. En las columnas posteriores se muestra la información análoga diferenciando entre varones y mujeres. En los tres casos (varones, mujeres y ambos) se presenta también la población ocupada total, que serán los datos utilizados para normalizar los resultados de accidentes laborales según el sexo del accidentado.

Los archivos presentados hasta ahora proporcionan información para calcular la proporción de accidentes de varias clases en función de la población activa/ocupada de cada una de ellas. Sin embargo, para categorías más específicas (como el tipo de contrato o número de trabajadores en la empresa, por ejemplo), las clases en las que el INE presenta la clasificación no coincide con la del Ministerio de Trabajo. Para estas situaciones, se ha consultado la página web del INE en cada caso, con la intención

<sup>2</sup>Resultados de la encuesta de población activa, INE. Última vez consultado el 23 de octubre de 2022 [online](#).

de obtener una idea aproximada que dé un contexto de cómo se reparte la población ocupada para esas categorías.

Posteriormente, cuando se expongan los resultados del análisis, se indica para qué clasificaciones es necesario consultar dicha web<sup>3</sup>.

## 2.2. Construcción de los nuevos *datasets*

Conocida la información de la que se dispone, el siguiente paso es seleccionar aquella con la que se desea trabajar.

Como se ha avanzado, todos los archivos descargados contienen información de los años entre 2014 y 2021, pues ya se ha argumentado que los archivos de estos años tienen la misma estructura, son los más recientes (cercaos al año en curso) y los que más información poseen. Por tanto, estos son los años sobre los que se ha realizado el análisis.

Sin embargo, a pesar de que los archivos de Excel del Ministerio considerados se organizan de la misma manera, también presentan detalles que no son comunes en todos los archivos de datos. Por esta razón, con la intención de crear los *datasets* unificados, se deben realizar modificaciones en ellos para poder unir las tablas de datos iniciales. A continuación, se explica detalladamente la reestructuración de los datos llevada a cabo.

Todo este procedimiento se ha realizado utilizando la biblioteca *numpy* (Harris *et al.* 2020) y la librería *pandas* (McKinney *et al.* 2010) de *Python*. La primera es una herramienta diseñada para crear y manipular vectores y matrices de grandes dimensiones, mientras que la segunda se especializa en el tratamiento de datos enfocado a su análisis. Ambas herramientas han resultado fundamentales tanto en la construcción de los *datasets* que se presenta a continuación, como en el análisis de estos.

### 2.2.1. Esquema de los nuevos *datasets*

La idea inicial del proyecto consistía, por una parte, en construir un solo *dataset* que unificase los datos disponibles sobre accidentalidad laboral en España. Con la información proporcionada por los datos del Ministerio de Trabajo y considerando la estructura y temática de estos, se consideró más apropiado construir, en lugar de uno, tres nuevos *dataframes* (*datasets* dispuesto por filas y columnas). El primero de ellos consiste en un *dataframe* en el que se han almacenado datos clasificados según la gravedad del accidente; el segundo *dataframe* clasifica los accidentes según las provincias y CC.AA. donde tuvieron lugar y, por último, el tercer *dataframe* clasifica los accidentes según el sector, sección o división de actividad económica al que pertenece la empresa del trabajador accidentado. Cada una de estas clasificaciones dará lugar a las filas de cada uno de los *dataframes*. Por otro lado, las columnas se obtendrán considerando otras clasificaciones ya mencionadas en los apartados anteriores.

A continuación se describe la estructura de cada uno de ellos:

#### ■ *Dataset* según la GRAVEDAD de los accidentes:

Como se comentaba en la sección anterior, los archivos de Excel tipo A del Ministerio de Trabajo contienen 26 hojas, en las cuales se muestran diversas clasificaciones de los accidentes laborales teniendo en cuenta la división por gravedad.

Todas las hojas (1.1, 1.2,...,3.1,...,6.2, R.1 y R.2) de estos archivos para todos los años desde 2014 a 2021 presentan las mismas columnas. Por otro lado, para cada año, cada una de las hojas tiene una clasificación diferente por filas, que ya han sido comentadas anteriormente cuando se introdujeron los archivos tipo A. Debido a esta estructuración, para cada año, las diferentes filas se pueden concatenar. De esta forma, para cada uno de los años considerados, se obtiene una

---

<sup>3</sup>INE. Última visualización el 1 de diciembre de 2022 [online](#)

tabla cuyas columnas consisten en la clasificación de accidentes por gravedad y sus filas son la concatenación de todas las filas de cada una de las hojas del documento del año correspondiente.

Por tanto, para cada año se ha logrado crear una tabla de datos cuyas columnas son:

En Jornada				<i>In Itinere</i>			
Total	Leve	Grave	Mortal	Total	Leve	Grave	Mortal

y cuyas filas consisten en la concatenación de todas las que se encuentran de las hojas de los archivos tipo A del año correspondiente.

Para mostrarlo de forma más clara, se ejemplifica en el Cuadro 2.2.1 a través del archivo de Excel tipo A correspondiente al año 2014. La situación es análoga para el resto de años. En las dos primeras hojas se clasifican los accidentes (además de por gravedad mediante las columnas) por comunidad autónoma y provincia, la primera, y por sector, sección y división de actividad económica, la segunda; con lo cual se unen todas las filas dando lugar a una única tabla. En el Cuadro 2.2.1 se puede ver, de forma resumida, la concatenación de estas dos primeras hojas.

	En Jornada				<i>In Itinere</i>			
	Total	Leve	Grave	Mortal	Total	Leve	Grave	Mortal
TOTAL								
Andalucía								
Almería								
Melilla								
...								
Agricultura, ...								
...								
Actividades ...								

Cuadro 2.1: Ejemplo de concatenación de dos hojas de los archivos de Excel *ATR\_2014\_A.xls*.

Como todas las hojas del archivo *ATR\_2014\_A.xls* presentan las mismas columnas, podemos concatenar las filas de todas ellas, al igual que hemos hecho con las dos anteriores.

Repetiendo el proceso con los archivos tipo A de cada año, se obtienen 8 *dataframes* con las mismas columnas; por tanto, es posible concatenar sus filas, generando así un nuevo *dataframe* con 8 columnas (En Jornada: total, leve, grave y mortal e *In Itinere*: total, leve grave y mortal) y 10732 filas que almacena los datos de todas las hojas de todos los años desde 2014 hasta 2021. Desde *Python*, este proceso se ha llevado a cabo concatenando todas las filas simultáneamente, una vez depuradas las distintas tablas de las hojas de Excel.

Nótese que el número de filas no es múltiplo de 8. Esto se debe a que, como ya se ha dicho, no para todos los años se tiene exactamente la misma clasificación (en algunos años se consideran más filas que en otros para algunas de las hojas). Esta cuestión se comentará a continuación, cuando se expliquen los inconvenientes encontrados a lo largo de la construcción de los *dataframes*.

El nuevo *dataframe* obtenido, ya compuesto por los datos de todos los años (de 2014 a 2021), se traspone, obteniendo así una tabla con 8 filas y 10732 columnas. Con este acto se consigue que el *dataframe* construido final tenga mayor número de columnas que de filas; de esta forma tiene la estructura perfecta para almacenarse en formato *parquet*. A dicho *dataset* se le ha denominado ***dfGRAVEDAD***. Como se deduce por como se ha construido, contiene una tabla en la que se clasifican los accidentes laborales intersecando la clasificación por gravedad (filas) con 26 otras clasificaciones (columnas), sacadas desde la hoja 1.1 a la R.2. de cada año.

■ ***Dataset* según la GEOGRAFÍA de los accidentes:**

El segundo de los *datasets* construidos ha sido obtenido a partir de los archivos Excel tipo D del Ministerio. Se ha visto anteriormente que estos archivos tienen una estructura diferente a los de tipo A y que no se podían estructurar de forma homogénea y organizada todas sus hojas. Por ese motivo se decidió aprovechar solamente, de cada uno de los 14 grupos del documento, las hojas que clasifican accidentes según CC.AA. y provincias y según sectores, secciones y divisiones económicas. Para crear el *dataset* según la geografía del accidente, se considerarán simplemente las 14 hojas (una de cada uno de los 14 grupos) que contienen la clasificación por geografía.

Conocidas ya las hojas de donde se obtendrá la información, el procedimiento para construir este segundo *dataset* es muy similar al del anterior.

Para cada año, en el archivo correspondiente de Excel (por ejemplo, para 2014: ATR.2014.D.xls) hay 14 hojas en las que la clasificación de los accidentes que presentan tienen como filas las CC.AA. y provincias (junto con las Ciudades Autónomas) del estado español. Sin embargo, en cada una de las hojas, las columnas se corresponden con la temática de cada uno de los 14 grupos ya expuestos:

1. Según sexo y edad
2. Según tipo de contrato
3. Según grandes grupos de ocupación
4. Según antigüedad en el puesto de trabajo
5. Según país de nacionalidad
6. Según tamaño del centro donde ocurre el accidente
7. Según organización preventiva de la empresa y evaluación de riesgos sobre el puesto de trabajo
8. Según lugar del accidente
9. Según actividad física específica que desempeñaba el trabajador accidentado
10. Según desviación que desencadenó el accidente
11. Según forma o contacto que produjo la lesión
12. Según descripción de la lesión
13. Según parte del cuerpo lesionada
14. Según mes de ocurrencia del accidente

Teniendo esto en cuenta, para cada año se pueden concatenar las columnas de las 14 hojas, dando así lugar a una tabla cuyas filas son las CC.AA. y provincias, y las columnas están constituidas por las diferentes categorías en las que se clasifican los accidentes para las 14 temáticas.

Realizando esto para cada año, se obtienen 8 *datasets* que tienen en común todas sus filas. Esto permite que se puedan concatenar las columnas de los 8 *datasets* y formar así un *dataframe* final de 63 filas y 1777 columnas, conteniendo los datos desde el año 2014 hasta 2021. Al *dataframe* así generado se le ha denominado como ***dfGEOGRAFICO***.

En este caso, el *dataframe* construido ya posee más columnas que filas, por lo que no es necesario trasponerlo para convertirlo finalmente a formato *parquet*.

■ **Dataset según el SECTOR, SECCIÓN o DIVISIÓN de la empresa del accidentado:**

Tal y como comentamos para el caso anterior, los archivos tipo D pueden agrupar sus hojas en 14 categorías diferentes para clasificar los accidentes. Para construir el último *dataset* con información sobre accidentes laborales, se consideraron de esos grupos las hojas que clasificaban los accidentes según el sector, sección o división de actividad económica al que pertenecía el trabajador accidentado. La construcción de este nuevo *dataframe*, que se nombra como ***dfSECTORES***, se ha realizado de forma totalmente análoga al *dataframe* anterior (*dfGEOGRAFICO*); pues la estructura de las hojas consideradas era la misma. Por tanto, la organización y concatenación de las hojas del Excel se pudo llevar a cabo de la misma manera ya expuesta.

Al igual que los otros dos archivos construidos, este *dataset* también se ha guardado en formato *parquet*, con una estructura que contiene más columnas (1777) que filas (108). Nótese que tiene el mismo número de columnas que el documento *dfGEOGRAFICO*; pues como ya se ha dicho, las hojas de los archivos tipo D de Excel que almacenan los datos por CC.AA. y provincias contienen la misma información y organizada de igual forma que las hojas que clasifican por sectores, secciones y divisiones. El número de filas difiere porque las provincias y CC.AA. del Estado español son 63 (contando Ceuta y Melilla y sin repetir las comunidades que solo tienen una provincia) y, por otro lado, el número de sectores, secciones y divisiones considerado por el Ministerio de Trabajo es 108.

Para facilitar la visualización de estos tres nuevos archivos de datos, se incluyen a continuación fragmentos de las tablas construidas. Debido a su gran dimensión, tan solo se mostrarán algunas de las filas y columnas de las tablas de datos.

En las Figuras 2.5 y 2.6 se muestra el archivo *dfGRAVEDAD*, en primer lugar solamente sus primeras y últimas columnas y en segundo, una muestra de columnas aleatorias.

	2014	...										2021						
	ATR-A.1.1											ATR-A.2.1.2	ATR-A.2.4.1					
	TOTAL	ANDALUCÍA	Almería	Cádiz	Córdoba	Granada	Huelva	Jaén	Málaga	Sevilla	...	De 65 y más años MUJERES	TOTAL	Hasta 1 mes	Entre 1 y 2 meses	Entre 3 y 4 meses	Entre 5 y 7 meses	Entre 8 y 12 meses
total JORNADA	424625	70775	6395	9594	7204	5731	5966	4920	13489	17476	...	1122	523800	54531	56734	36248	35002	35157
leve JORNADA	420829	69914	6342	9498	7097	5628	5902	4835	13324	17288	...	1100	519293	54048	56345	35979	34745	34875
grave JORNADA	3329	786	38	84	99	98	58	78	155	176	...	22	3896	432	340	238	228	248
mortal JORNADA	467	75	15	12	8	5	6	7	10	12	...	0	611	51	49	31	29	34
total ITINERE	66474	11175	880	1223	738	1084	583	490	2601	3576	...	366	77323	7687	8378	5071	4823	4898
leve ITINERE	65477	10901	866	1199	710	1054	571	468	2550	3483	...	362	76292	7574	8297	5003	4757	4835
grave ITINERE	884	261	12	23	25	29	11	22	50	89	...	4	901	93	71	57	59	52
mortal ITINERE	113	13	2	1	3	1	1	0	1	4	...	0	130	20	10	11	7	11

Figura 2.5: Muestra del archivo *dfGRAVEDAD* en el que se incluyen todas sus filas y sus primeras y últimas columnas.

La tabla de muestra de la Figura 2.6 se ha obtenido con el siguiente código en *Python*:

```
lista=[]
for i in range(len(dfGRAVEDAD.columns)):
    lista.append(i)
muestra=sample(lista,10)

dfGRAVEDAD[dfGRAVEDAD.columns[[muestra]]]
```

Se ha seleccionado una muestra de 10 columnas para que fuese posible incluir la tabla completa en este documento.

	2017	2018	2014	2018	2021	2019	2014	2017	2014	
	ATR-A.2.5	ATR-A.4.7	ATR-A.6.1	ATR-A.2.3	ATR-A.2.3	ATR-A.4.5	ATR-A.4.6	ATR-R.2	ATR-A.4.7	ATR-A.4.2
	Resto de África	Herramientas Manuales - para soldar, pegar	Ahogamiento y sumersiones no mortales	52 Dependientes en tiendas y almacenes	24 Profesionales de las ciencias físicas, químicas, matemáticas/ny de las ingenierías	- móviles (andamios, escalas, barquilla)	Otro contacto contra un objeto inmóvil (trabajador en movimiento) - Sin especificar	20 Industria química	Elementos de edificios, de construcciones- puertas, paredes, ventanas, etc.	Demolición de todo tipo de construcciones
total JORNADA	289	581	15	23270	1232	7784	8421	140	10750	375
leve JORNADA	286	580	14	23191	1207	7552	8385	139	10688	362
grave JORNADA	3	1	1	65	22	218	36	1	55	10
mortal JORNADA	0	0	0	14	3	14	0	0	7	3
total ITINERE	24	0	1	6533	640	29	505	22	396	0
leve ITINERE	24	0	1	6480	624	29	504	22	389	0
grave ITINERE	0	0	0	50	14	0	1	0	7	0
mortal ITINERE	0	0	0	3	2	0	0	0	0	0

Figura 2.6: Muestra del archivo *dfGRAVEDAD* en el que se incluyen todas sus filas y 10 columnas aleatorias.

En las Figuras 2.8 y 2.7 se pueden ver ejemplos de segmentos de las tablas correspondientes a los archivos *dfGEOGRAFICO* y *dfSECTORES*, respectivamente.

	2014										... 2021										
	ATR-D.1.1										... ATR-D.1.1.2										
	TOTAL	16-17 años	18-19 años	20-24 años	25-29 años	30-34 años	35-39 años	40-44 años	45-49 años	50-54 años	...	20-24 años	25-29 años	30-34 años	35-39 años	40-44 años	45-49 años	50-54 años	55-59 años	60-64 años	+65 años
FILAS																					
TOTAL	424625	398	2371	21755	41356	58807	69935	65263	59238	51554	...	9380	14152	14880	17184	21389	21834	21291	17951	10544	1122
ANDALUCÍA	70775	104	519	4355	7549	10178	11664	10731	9632	8259	...	1369	2122	2257	2709	3267	3306	3109	2455	1352	162
Almería	6395	11	66	462	769	950	1108	947	773	628	...	199	260	267	267	365	377	352	220	137	22
Cádiz	9594	7	50	457	914	1341	1601	1480	1408	1184	...	135	238	293	321	406	390	382	281	138	19
Córdoba	7204	8	65	485	800	973	1050	998	1042	921	...	130	198	198	262	284	309	260	276	154	12
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Guipúzcoa	7294	6	28	266	561	914	1177	1208	1122	966	...	115	143	220	253	325	418	368	351	208	17
Vizcaya	11513	1	25	359	1012	1477	1896	1825	1664	1609	...	181	336	356	421	630	662	681	647	341	30
RIOJA (LA)	2865	1	16	157	300	373	420	444	386	377	...	74	118	132	165	213	250	247	190	127	15
Ceuta	610	0	1	27	45	65	86	114	93	92	...	15	24	20	27	30	34	31	30	10	1
Melilla	492	0	0	27	38	61	51	73	97	66	...	12	28	21	20	29	31	31	32	23	5

Figura 2.7: Muestra del archivo *dfGEOGRAFICO* en el que se incluyen sus primeras y últimas filas y columnas.

Además de estos tres archivos de datos principales, los Excel obtenidos de la web del INE con la información sobre población activa y ocupada también han sido guardados en formato *parquet*. En este

	2014										... 2021										
	ATR-D.1.2										... ATR-D.1.2.2										
	TOTAL	16-17 años	18-19 años	20-24 años	25-29 años	30-34 años	35-39 años	40-44 años	45-49 años	50-54 años	...	20-24 años	25-29 años	30-34 años	35-39 años	40-44 años	45-49 años	50-54 años	55-59 años	60-64 años	+65 años
FILAS																					
A Agricultura, ganadería, silvicultura y pesca	30250	73	412	2508	3416	4274	4662	4327	3910	3342	...	498	557	626	735	907	815	782	509	299	40
01 Agricultura, ganadería, caza y servicios relacionados con las mismas	24922	67	373	2218	2881	3493	3813	3514	3081	2650	...	483	533	603	701	850	772	733	471	285	39
02 Silvicultura y explotación forestal	2691	2	21	173	343	472	455	416	354	261	...	10	16	11	23	36	34	34	23	9	1
03 Pesca y acuicultura	2637	4	18	117	212	309	394	397	475	431	...	5	8	12	11	21	9	15	15	5	0
B Industrias extractivas	2717	0	5	41	252	420	579	677	405	203	...	1	2	5	3	4	1	4	2	1	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
96 Otros servicios personales	3091	1	21	220	408	483	480	409	414	313	...	83	176	189	232	278	224	208	159	75	11
T Actividades de hogares como empleadores de personal doméstico	2339	0	3	35	101	170	228	289	371	437	...	11	25	39	66	114	116	213	222	158	48
97 Actividades de los hogares como empleadores de personal doméstico	2339	0	3	35	101	170	228	289	371	437	...	11	25	39	66	114	116	213	222	158	48
U Actividades de organizaciones y organismos extraterritoriales	48	0	1	4	4	9	3	4	6	7	...	0	1	1	0	1	1	1	0	0	0
99 Actividades de organizaciones y organismos extraterritoriales	48	0	1	4	4	9	3	4	6	7	...	0	1	1	0	1	1	1	0	0	0

Figura 2.8: Muestra del archivo *dfSECTORES* en el que se incluyen sus primeras y últimas filas y columnas.

caso no se ha tenido que aplicar ninguna transformación a las tablas de datos proporcionadas por el INE, pues se ha mantenido para los documentos en formato *parquet* la estructura de las tablas del Excel.

### 2.2.2. Proceso de construcción y problemática durante el proceso

En la sección anterior se ha indicado la estructura que tendrán los *datasets* que almacenan la información sobre los accidentes laborales a analizar. En el proceso de creación de dichos archivos, han sido encontradas diversas cuestiones que han dificultado el trabajo. En esta nueva sección se detalla más el proceso de construcción de los *dataframes*, haciendo hincapié en los inconvenientes que han aparecido en el proceso y en su resolución.

El primer inconveniente ha tenido lugar antes del procesamiento de los datos. Tal y como ya se ha comentado, los archivos proporcionados por el Ministerio de Trabajo no presentan un formato totalmente unificado. Por los motivos ya expuestos, la decisión tomada fue prescindir de los datos más antiguos y restringir el estudio en los datos posteriores a 2014, año a partir del cual se cambió la estructura de los archivos por parte del Ministerio.

Entre los años 2014 y 2021 también existen algunas pequeñas variaciones, pero estas no están relacionadas con la estructura de los archivos. Ejemplo de estas modificaciones son la adición de categorías como “no consta” o “sin información” en algunas de las clasificaciones o la inclusión de países en las tablas de datos clasificadas por nacionalidad del trabajador. Este tipo de cambios provocan que haya que tener cuidado a la hora de construir los *datasets*, pues no para todos los años las tablas de clasificación de los accidentes contienen las mismas columnas. Esto implica que hay variaciones en

las cabeceras asociadas a las columnas en cada caso, pues la longitud de los encabezamientos no será idéntica y habrá que incluir o eliminar títulos según el año al que correspondan los datos. Salvando ese detalle, esta situación no supone ningún otro inconveniente a la hora de construir el *dataset* con la estructura determinada.

Tras seleccionar los años para el análisis, se procede con la construcción de los nuevos *datasets*.

En primer lugar, cabe destacar que los archivos de Excel fueron almacenados en diccionarios de Python. Se crearon dos diccionarios: uno con los archivos tipo A y otro con los archivos tipo D. Cada uno de ellos con 8 elementos, que se corresponden con los años

2014, 2015, 2016, 2017, 2018, 2019, 2020 y 2021.

Dentro de cada uno de los años, se almacenaba otro diccionario, cuyos elementos se corresponden con las hojas del Excel del año en cuestión (tanto para los del tipo A, como los del tipo D). De esta forma, para acceder a una tabla en concreto, solo es necesario indicar en el diccionario de que año es la tabla requerida y dentro del año indicar la hoja específica. Una vez construidos los diccionarios sobre los que se modifican las diferentes tablas, se eliminan aquellas que no se utilizan; como la portada y el índice (tanto para documentos tipo A, como D) y las tablas de los archivos tipo D que no han sido utilizadas (aquellas que no clasifiquen los accidentes por geografía o sectores). Con esta disposición de los datos, se procede a su modificación para poder construir los *datasets* finales.

Los primeros inconvenientes vienen dados porque, tal y como están construidas las tablas de los archivos de Excel, existen muchos espacios en blanco que al leerlos en *Python* se interpretan como *NA*s. Debido a esto, ha habido que examinar en qué filas y columnas se encontraban esos valores indeterminados para poder eliminarlos o sustituirlos. Primero, se ha utilizado un comando genérico para eliminar todas las filas y columnas que tuviesen algún *NA*, dado que el principal problema son las filas o columnas completamente vacías que aparecen en los documentos a modo de separación. Sin embargo, en algunos casos ocurre que hay datos faltantes para algunas categorías y eso se transforma en *NA*s al importar el archivo a *Python*. Esto provoca que al usar el comando anterior se eliminen columnas donde hay datos válidos. Por esta razón hubo que revisar las hojas de todos los documentos de Excel (desde 2014 hasta 2021) y modificar aquellos valores ausentes para que no se eliminasen esas columnas y filas al eliminar del documento las líneas con *NA*s.

Este inconveniente solo se ha presentado para los archivos tipo A, que dan lugar al *dataframe* *dfGRAVEDAD*. Para los archivos tipo D también se ha llevado a cabo la revisión de los documentos para evitar tener el mismo problema, pero en este caso no se han encontrado datos faltantes que se tradujesen como *NA*s en *Python*. Concretamente, las hojas que ha sido necesario modificar han sido las 4.1, 4.2, 4.3 y 5.2 de los documentos tipo A de cada uno de los 8 años considerados. La solución aplicada ha sido sustituir los valores faltantes por el valor 0, ya que en todos los casos la ausencia de valores se corresponde con la falta de casos.

En el proceso de revisión anterior se han detectado algunos caracteres indeseados. Concretamente, para los años 2014 y 2015, en las tablas de los archivos tipo A en las que se clasifican los accidentes por desviación que produjo el accidente (4.4) aparecen varios símbolos de almohadilla (#) que ha habido que eliminar.

Tras eliminar todos los caracteres indeseados, las modificaciones ahora se centran en reescribir los índices y columnas de las tablas. En este aspecto, se han resuelto diferentes casuísticas:

- En algunas de las tablas que presentan sus índices en diferentes categorías y, dentro de esas categorías, otras subcategorías, los títulos se presentan como se muestra en la Figura 2.9. El ejemplo mostrado en dicha figura se corresponde con la hoja 2.3 del *dataset* tipo A del año 2016, pero la situación es la misma para esa misma hoja en el resto de años. La cuestión es que

cada categoría viene indicada por una letra del abecedario y las subcategorías, por un número. Al importar las tablas a *Python*, se toma una sola columna como índice; por tanto se tuvieron que concatenar la primera y segunda columna de las tablas con este aspecto para que el índice contuviese el índice alfa-numérico y el nombre de la categoría. Las tablas que presentan esta organización de sus índices son las 2.3 de los documentos tipo A (para todos los años) y cualquiera de las tablas (en documentos tipo A o tipo D) en las que se presenten los accidentes clasificados por sectores, secciones o divisiones. En estos últimos documentos, las secciones son las organizadas por letras y las divisiones por números.

ACCIDENTES DE TRABAJO		ATR-A.2.3. ACCIDENTES DE TRABAJO CON BAJA, EN JORNADA E IN ITINERE, SEGÚN GRAVEDAD, POR OCUPACIÓN DEL TRABAJADOR ACCIDENTADO							
Año 2016		EN JORNADA				IN ITINERE			
TOTAL		Total	Leves	Graves	Mortales	Total	Leves	Graves	Mortales
TOTAL		489.065	484.912	3.657	496	77.170	76.045	992	133
<b>A Directores y gerentes</b>		<b>2.008</b>	<b>1.972</b>	<b>31</b>	<b>5</b>	<b>642</b>	<b>622</b>	<b>17</b>	<b>3</b>
11 Miembros del poder ejecutivo y de los cuerpos legislativos; directivos de la Administración Pública y organizaciones de interés social; directores ejecutivos		129	128	1	-	58	56	2	-
12 Directores de departamentos administrativos y comerciales		519	509	10	-	322	314	7	1
13 Directores de producción y operaciones		337	331	4	2	80	76	3	1
14 Directores y gerentes de empresas de alojamiento, restauración y comercio		390	386	3	1	83	80	3	-
15 Directores y gerentes de otras empresas de servicios		633	618	13	2	99	96	2	1
<b>B Técnicos y profesionales científicos e intelectuales de la salud y la enseñanza</b>		<b>12.759</b>	<b>12.609</b>	<b>144</b>	<b>6</b>	<b>5.837</b>	<b>5.733</b>	<b>101</b>	<b>3</b>
21 Profesionales de la salud		7.738	7.632	103	3	3.691	3.616	73	2
22 Profesionales de la enseñanza infantil, primaria, secundaria y postsecundaria		3.231	3.206	23	2	1.572	1.553	18	1
23 Otros profesionales de la enseñanza		1.790	1.771	18	1	574	564	10	-
<b>C Otros técnicos y profesionales científicos e intelectuales</b>		<b>3.576</b>	<b>3.521</b>	<b>43</b>	<b>12</b>	<b>2.605</b>	<b>2.571</b>	<b>28</b>	<b>6</b>
24 Profesionales de las ciencias físicas, químicas, matemáticas y de las ingenierías		1.054	1.031	17	6	749	737	10	2
25 Profesionales del derecho		127	124	2	1	154	152	1	1
26 Especialistas en organización de la Administración Pública y de las empresas y en la comercialización		807	796	9	2	610	603	6	1
27 Profesionales de las tecnologías de la información		296	289	6	1	591	582	7	2

Figura 2.9: Hoja 2.3 del documento tipo A de 2016.

- Aparte de la estructura anterior, existen otras hojas en las que las categorías engloban subcategorías. Es el caso de las hojas 2.1, 2.4, 4.5 y 4.7 de los documentos tipo A. En este caso, las tablas presentan subcategorías que se repiten para diferentes categorías.

Todos estos índices repetidos deben ser modificados, pues resulta esencial que cada índice y nombre de columna sea único para luego poder exportarlo a formato *parquet*. Para resolver esta cuestión, para cada una de las subcategorías cuyo nombre esté repetido, se le ha concatenado al título el nombre de la categoría (superior) a la que corresponde. Así, por ejemplo, en las tablas de sexo y edad las subcategorías ya no solo tienen como nombre “de 16 a 17 años” (por ejemplo), sino que pasaron a ser “de 16 a 17 años VARONES” y “de 16 a 17 años MUJERES”.

- Al eliminar las filas y columnas con *NAs*, se eliminaron también los títulos de las columnas de las tablas. En el caso de las tablas de los archivos A, ya se ha visto que las columnas son las mismas en todas las tablas:

En Jornada				In Itinere			
Total	Leve	Grave	Mortal	Total	Leve	Grave	Mortal

Por lo que con un solo bucle se pudieron modificar todas las tablas a la vez para nombrar todas las columnas. Concretamente, los nombres de columna introducidos fueron:

Total JORNADA	Leve JORNADA	Grave JORNADA	Mortal JORNADA	Total ITINERE	Leve ITINERE	Grave ITINERE	Mortal ITINERE
---------------	--------------	---------------	----------------	---------------	--------------	---------------	----------------

En el caso de los archivos tipo D, de los cuales solo se toman las hojas 1 y 2 de cada grupo (aquellas que clasifican por CC.AA./provincia y por sectores/secciones/divisiones), lo que tienen en común es el nombre de sus filas. Por tanto, para establecer el nombre de las columnas, se debe hacer hoja por hoja poniendo los nombres correspondientes según la clasificación pertinente. Se suma además otro detalle, y es que a lo largo de los años van variando las categorías consideradas, por ejemplo: en las hojas que clasifican los accidentes según la nacionalidad del accidentado, para los años 2014 y 2015 se consideraba la Unión Europea como nacionalidad, pero a partir de 2016 se empezaron a considerar los países que la componen de forma desglosada. Esto provoca que los títulos de las columnas varíen de un año a otro. Como este ejemplo, ocurren situaciones similares en otras clasificaciones debido a que en algunos años se incorporan categorías como “No consta”.

Debido a esto, los nombres de las columnas para cada una de las tablas se han puesto conjuntamente (usando un bucle) para los años en los que coincidían las categorías y, por separado, en los casos en los que las categorías diferían de año a año.

Todos estos procesos, aunque posteriormente se han llevado a cabo con comandos de las librerías *pandas* y *numpy* de *Python*, han requerido la revisión detallada de todas las hojas de los archivos de Excel para poder localizar en qué años, hojas o filas y columnas se encontraba la anomalía/problema que ha requerido solución.

Tras la depuración de los datos, están en disposición de unirlos para construir los 3 *datasets* finales tal y como ya se ha explicado en la sección anterior: para construir el *dataframe* *dfGRAVEDAD* se consideran las tablas de los archivos A traspuestas (con la intención de tener un *dataset* con más columnas que filas) y se concatenan sus columnas. Para construir los *dataframes* *dfGEOGRAFICO* y *dfSECTORES* se toman, en el caso del primero, las tablas de los archivos tipo D cuyas filas son las CC.AA. y provincias y, en el caso del segundo, las tablas de los archivos tipo D cuyas filas son los sectores, secciones y divisiones económicas. En ambos casos, se concatenan las columnas de las tablas pertinentes y se obtiene, en el caso de *dfGEOGRAFICO*, un *dataframe* cuyas filas clasifican los accidentes según CC.AA./provincias y, en el caso de *dfSECTORES*, un *dataframe* cuyas filas clasifican los accidentes según sectores/secciones/divisiones.

Los *dataframes* construidos presentan una limitación, que también poseen los datos originales extraídos de la web del Ministerio de Trabajo: solo se dispone de los datos agrupados, no se tiene acceso a los datos por accidente. Esto hace que se pierda mucha información, especialmente a la hora de intersecar características. Se tiene acceso a los accidentes por comunidades autónomas, por ejemplo, a través del documento *dfGEOGRAFICO*. En él se almacenan 14 tablas por año en las que se interseca la clasificación por CC.AA./provincias con otra clasificación, por ejemplo tipo de contrato en la segunda tabla. Teniendo esto en cuenta, de la tabla mencionada podríamos extraer el número de accidentes ocurridos en una comunidad o provincia según el tipo de contrato del accidentado (temporal, indefinido, a tiempo parcial...); pero si se quisiese intersecar esa información con otra clasificación (como puede ser el sexo del empleado), no se podría, pues no se dispone de la intersección de las tres categorías.

La agrupación de los datos también impide realizar algunas tareas que podrían resultar interesantes. La más notable sería realizar predicciones sobre la gravedad de nuevos accidentes: disponiendo de la información por individuos (edad del trabajador, sexo, tipo de contrato, nacionalidad, sector de actividad económica, etc.) obtener predicciones de la gravedad del posible accidente que pudiese sufrir. Desgraciadamente, con los datos tal y como están organizados, no se pueden llevar a cabo este tipo de procesos.

Sí permite hacer otro tipo de predicciones, por ejemplo: dado que se dispone de la cantidad de accidentes anuales y mensuales a lo largo de los 8 años considerados, se puede predecir la cantidad de accidentes laborales esperados en el futuro. Aunque útil, la información proporcionada por este tipo de predicciones es más limitada que en el caso de disponer de los datos originales por individuos.

Otra de las tareas que se harían posibles disponiendo de los datos por individuo (con un identificador para cada uno de estos) sería estudiar en profundidad las recaídas. Si en cada accidente se incorporase una variable que indicase con un código el individuo que sufre el accidente, se podría ver cuántas personas y con qué características sufren dichos accidentes. Esto aportaría una profundidad aún mayor al análisis. No solo se estudiarían las características de los accidentes, sino que también permitiría indagar en los perfiles que tienden a la reincidencia.

En conclusión, disponer de los datos por accidente laboral hubiese proporcionado mucha más información y flexibilidad a la hora de tratarlos y realizar el análisis de estos. A pesar de ello, se ha trabajado con los datos disponibles y se ha tratado de obtener un análisis completo que se muestra en la siguiente sección.

## Capítulo 3

# Análisis descriptivo de los datos sobre accidentes laborales en España

Dentro de los objetivos de este trabajo, el principal era llevar a cabo un análisis de los datos que extrajese información útil e interpretable acerca de la accidentalidad laboral en España. Este tercer capítulo se focaliza en ello y en él se muestran los resultados obtenidos con las diferentes técnicas utilizadas.

El lenguaje de programación usado para este procedimiento ha sido de nuevo *Python*; siendo las librerías *numpy*, *pandas*, *matplotlib* (Hunter 2007) y *seaborn* (Waskom et al 2017) las utilizadas principalmente.

### 3.1. Estado del arte

Cada año, es común que organizaciones como Comisiones Obreras, el Instituto Nacional de Seguridad y Salud en el Trabajo o el propio Ministerio de Trabajo publiquen estudios sobre la accidentalidad laboral en España.

Dado que la información presentada en este tipo de estudios puede resultar una cuestión delicada en muchas ocasiones, la información se tiende a presentar modificada, con la intención de anonimizar a los individuos del estudio y sus empresas. Dichas modificaciones no alteran los datos reales, pero sí ocultan parte de la información para que no se identifique a los trabajadores que sufren el accidente. Una solución sencilla a este inconveniente podría ser eliminar los nombres de las personas de los registros e identificarlas de forma numérica. Sin embargo, con el resto de información de cada accidente puede ser suficiente como para descifrar la persona que se corresponde con él, por lo que no es una forma eficiente de anonimizar a los trabajadores. Por este motivo, la situación más recurrente es presentar los datos sobre accidentalidad laboral agrupados en categorías, al igual que hace el Ministerio de Trabajo.

Tal y como se comentó en el capítulo anterior, la presentación de los datos de forma agrupada limita la información que se puede extraer para el análisis. Para evitar esta pérdida de información, se analizó la posibilidad de recuperar los datos originales de forma individual. Existen múltiples estudios que pretenden recuperar el conjunto de datos original (por individuos) a través de los datos agregados. Algunos de los que se han revisado para este trabajo en concreto son Zheng Li et al. (2020) y Fabrice Yaméogo et al. (2021). La idea general de los métodos de reconstrucción es generar un conjunto de datos sintético a partir de los datos agrupados. Los métodos más comunes lo llevan a cabo mediante algoritmos de optimización matemática y métodos de aprendizaje estadístico, de forma que se verifique que los valores agregados del conjunto de datos generado coincidan con los originales. En el primero de

los artículos mencionados, se presenta un nuevo método en el que se utilizan cópulas gaussianas para generar los nuevos conjuntos de datos, alegando que los métodos de optimización más comunes presentan limitaciones que tienen una difícil solución. Los autores afirman que dicho método, además, garantiza mejores resultados. Sin entrar en detalles sobre el algoritmo mediante el cual se pretende reproducir los datos, a continuación se comentará el esquema seguido para validar el modelo de recuperación de datos:

1. Se genera (o se recopila a partir de datos reales) un conjunto de datos que contenga los datos almacenados de forma individual.
2. A partir de los datos seleccionados se obtienen los datos agregados.
3. Se toman estos últimos y se les aplica a ellos el procedimiento para recuperar los datos originales.
4. Tras aplicar el algoritmo, se obtiene el conjunto de datos sintético en el que se presentan los datos por individuos.
5. Una vez obtenido, se compara con el archivo original y se contrasta el número de individuos que coinciden en ambos conjuntos de datos.

Es importante destacar que los mejores resultados que se muestran en Zheng Li et al. (2020) siguiendo este proceso consiguen reproducir solamente un 50 % de los datos originales; y tal como ya se comentó, los autores resaltan que es uno de los mejores resultados obtenidos en este tipo de procedimientos.

A pesar de que la idea de aplicar este método a los datos sobre accidentes laborales parece en principio interesante, tras revisar los resultados de los diferentes métodos, no parece apropiado. Como mucho solo se garantiza la reproducción de la mitad de los datos reales, lo cual es insuficiente en el caso a tratar. Pretendiendo realizar un análisis real de los accidentes laborales en España, no tendría sentido trabajar con unos datos sintéticos que no se correspondiesen totalmente con la realidad.

Respecto al análisis estadístico, sí resultaría más interesante trabajar con los datos de forma individual, pues como se ha reiterado múltiples veces a lo largo de este trabajo, permitiría profundizar más en el análisis. Sin embargo, dado que el objetivo principal del proyecto es proporcionar un análisis útil y real de la accidentalidad laboral en España, se deben utilizar los datos reales facilitados por el Ministerio que se proporcionan de forma agrupada. De hecho, así se han construido los archivos de datos expuestos en el capítulo anterior.

En general, todos los informes que se hacen con respecto a esa temática, disponen de la misma información que en este proyecto, es decir, disponen de la información agrupada.

El análisis que se aplica sobre ellos se basa principalmente en el estudio del número de accidentes para cada año. Para el año considerado, se estudia la evolución dentro de los diferentes grupos de actividad económica, por sexos, por grupo de edad y distinguiendo también por provincias de España. En el caso de todas estas categorías, es común calcular las proporciones de cada una de las clases en función del total. También es común que se analice la proporción de accidentes según otras características, como según el tipo de contrato, según la ocupación del trabajador o según la nacionalidad. En los informes de organizaciones como el CCOO (2022) <sup>1</sup> y el INSST (2021) <sup>2</sup> se pueden ver los documentos en los que se incluyen los análisis mencionados.

Cabe destacar, que en estos informes también se diferencian los accidentes según su gravedad (leve, grave o mortal) y si han tenido lugar durante la jornada o *in itinere*, desplazándose hacia el lugar de

---

<sup>1</sup>Análisis de las estadísticas de accidentes de trabajo y enfermedades profesionales en España en 2021, CCOO. Última vez consultada el 29 de noviembre de 2022 [online](#).

<sup>2</sup>Informe anual de accidentes de trabajo en España en 2020, INSST. Última vez consultado el 29 de noviembre de 2022 [online](#).

trabajo o de vuelta a casa. Esta coincidencia se debe, seguramente, que para dichos informes se han extraído los datos de la web del Ministerio de Trabajo, que es donde se pueden obtener los datos oficiales.

En el informe que se incluye a continuación, se reproducen estos análisis y se consideran otros a mayores, aprovechando las diferentes categorías de las que se dispone en los documentos creados; como las partes del cuerpo lesionadas, los materiales con los que se produjo la lesión, las horas y días en las que se producen los accidentes, etc. Son categorías mucho más específicas, pero que aportan una información que no se suele evaluar y puede proporcionar resultados de interés. De esta forma no solo se analizan las características de los empleados que sufren los accidentes, sino que también se analizan las características de los propios accidentes.

Además, en el análisis presentado se han considerado los datos a lo largo de 8 años simultáneamente, dando una perspectiva más general del asunto. Lo común en los informes sobre accidentalidad laboral es que se realicen anualmente, limitando así que se observen las variaciones a lo largo de los años.

La diferencia de este trabajo con los mencionados anteriormente no solo reside en que se haya utilizado una mayor cantidad y diversidad de información, sino que se han utilizado también nuevas técnicas para analizar los datos. Estas han permitido obtener otros enfoques de los resultados. En concreto, dos de los nuevos enfoques añadidos han sido el análisis y clasificación de series de tiempo en función del número de accidentes laborales y la implementación de metodología *clúster* para la agrupación de provincias y actividades económicas según el número de accidentes.

En conclusión, el informe realizado pretende aumentar y mejorar la información aportada en los informes de accidentalidad laboral, manteniendo también los análisis que se llevan a cabo comúnmente.

## 3.2. Análisis de los accidentes laborales en España

Actualmente, gracias a los grandes avances en tecnología, la imposición de normativas de seguridad y, recientemente, el aumento del teletrabajo (tras la pandemia de COVID-19), la cantidad de accidentes laborales en la Unión Europea está descendiendo notablemente. En contraposición, destaca que en España, lejos de descender, está aumentando todavía más. Hasta 2019, la progresión de los accidentes laborales que causan baja en España iba en aumento. Fue frenada en 2020, reduciéndose considerablemente el número de accidentados; pero al año siguiente, el número de accidentes siguió creciendo. Este aumento de los accidentes, ya de por sí grave, llama más la atención comparándolo con la situación general de la UE, que presenta un descenso de accidentes laborales notable en los últimos años.

En la Figura 3.1 se observa claramente la tendencia creciente a lo largo de los últimos años en España, salvo la caída en 2020. Nótese que para 2021, tras la bajada del año anterior, el número de accidentes laborales vuelve a estar por encima de la media de los últimos 8 años.

En el caso de considerar solo los accidentes mortales, el número de accidentes es mucho menor (siendo en general tan solo un 0.1 % de los accidentes totales, al rededor de 500 de un total de 500.000), pero la tendencia creciente continúa. En este caso ni siquiera se observa una caída en 2020. La Figura 3.2 muestra la cantidad de accidentes mortales a lo largo de los 8 años considerados en el estudio.

En la Figura 3.3 se muestra que la tendencia de crecimiento también se extiende a los accidentes *in itinere*. En este caso, vuelve a observarse una caída del número de accidentes en 2020 y una rápida subida de nuevo en 2021.

En vista de estas tres gráficas, resulta claro el crecimiento continuo de los accidentes laborales en España.

Por otro lado, la gráfica extraída de la web de Eurostat<sup>3</sup>, que se puede visualizar en la Figura 3.4,

---

<sup>3</sup>Accidents at work statistics, Eurostat. Última vez consultado el 29 de noviembre de 2022 [online](#).

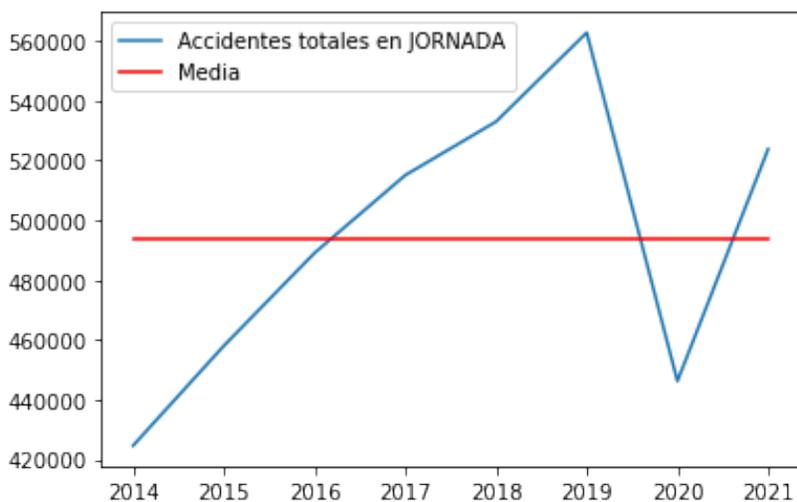


Figura 3.1: Número de accidentes laborales durante la jornada que causaron baja a lo largo de los años. Se representa en rojo la media de todos los años.

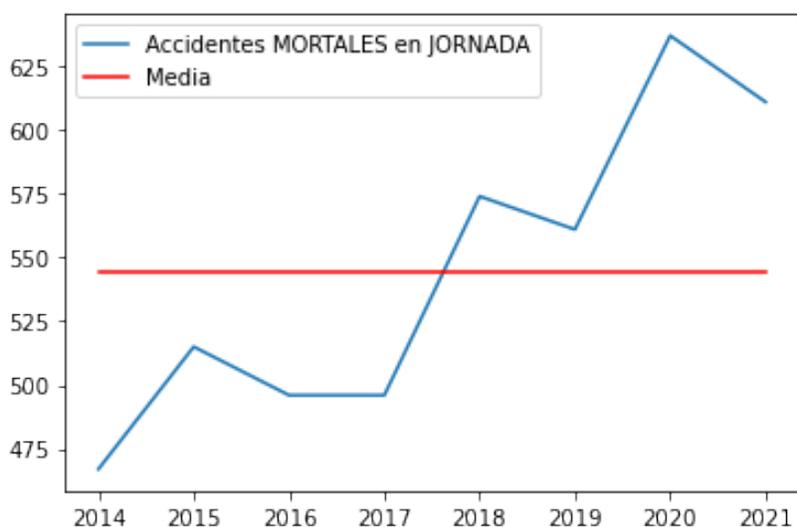


Figura 3.2: Número de accidentes laborales mortales durante la jornada. Se representa en rojo la media de todos los años.

muestra que entre los años 2010 y 2020 el número de accidentes descendió en la Unión Europea.

Gran parte de las actividades económicas han sufrido un notable descenso en el número de accidentes. Las que han aumentado su número, lo han hecho en una menor proporción, por lo que en el cómputo global se traduce en un descenso de la incidencia de casi 500 puntos (por cada 100.000 personas empleadas).

La comparación entre los accidentes en España y la UE no es del todo precisa, pues no se están comparando las mismas fechas. De todos modos, la idea general queda reflejada y es que en España está empeorando la situación en cuanto a accidentalidad laboral, mientras que en la Unión Europea se está avanzando en general hacia unos mejores resultados.

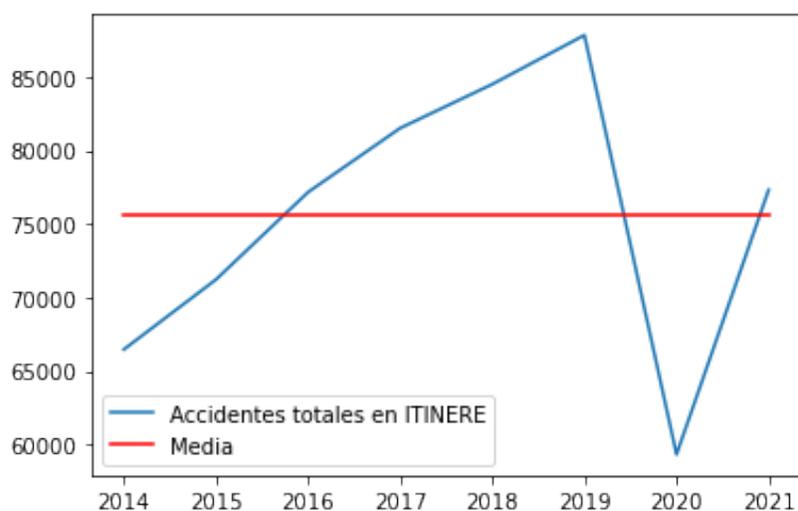


Figura 3.3: Número de accidentes laborales *in itinere* que causaron baja a lo largo de los años. Se representa en rojo la media de todos los años.

**Overall change in the incidence of non-fatal accidents at work by NACE division, EU, 2010-2020**  
(incidence rates per 100 000 persons employed)

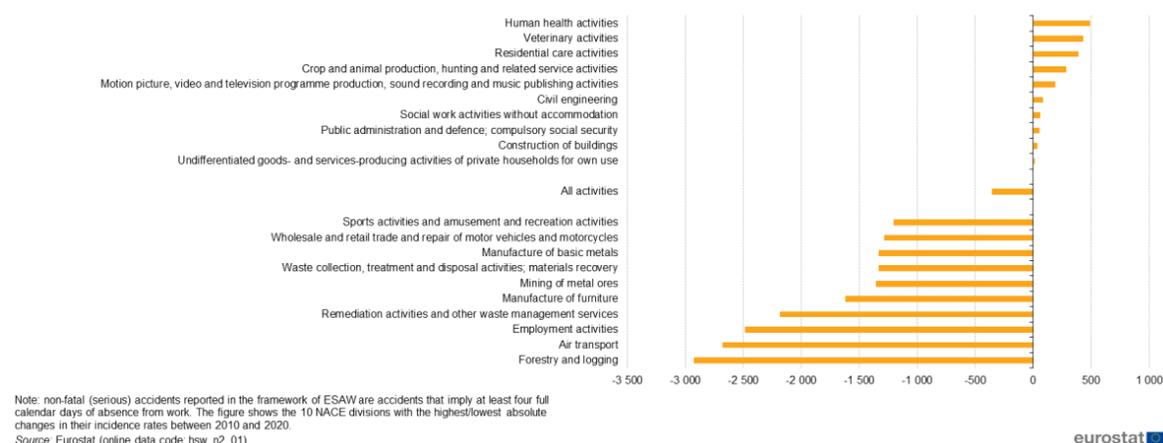


Figura 3.4: Variación del número de accidentes en las diferentes actividades económicas entre 2010 y 2020 en la UE.

En la web de Eurostat<sup>4</sup> se pueden observar referencias a esta situación. Concretamente, en el artículo adjuntado se presenta la situación en cuanto a número de accidentes laborales en la Unión Europea en 2020, comentando también el periodo transcurrido desde el año 2010. En el informe se exponen datos e información sobre la situación comentada: la bajada de accidentes, tanto fatales como no-fatales, en el conjunto de los países que forman la UE.

De cara a tener un mayor conocimiento del problema presentado, profundizar en las diferentes características que están presentes en los accidentes laborales aportará información sobre los perfiles de trabajadores con una mayor predisposición a sufrir accidentes, las actividades más peligrosas o los

<sup>4</sup>Accidents at work statistics, Eurostat. Última vez consultado el 31 de diciembre de 2022 [online](#).

periodos temporales donde más accidentes se sufren.

En las siguientes secciones se muestra la distribución de los accidentes laborales en España según sus diferentes características:

### 3.2.1. Características del accidentado

En los datos disponibles se puede encontrar la clasificación de los accidentes según diferentes categorías: sexo y grupo de edad del accidentado/a, la situación profesional, la ocupación y la nacionalidad de la persona accidentada. Conocer cómo se distribuyen los accidentes dentro de estas categorías proporcionará una información valiosa de cara a conocer los perfiles más propensos a sufrir accidentes laborales.

#### ■ Grupo de edad del accidentado/a

En cuanto a los grupos de edad, en la Figura 3.5 se puede ver el número de accidentes laborales para los mismos en los años 2014 y 2021,.

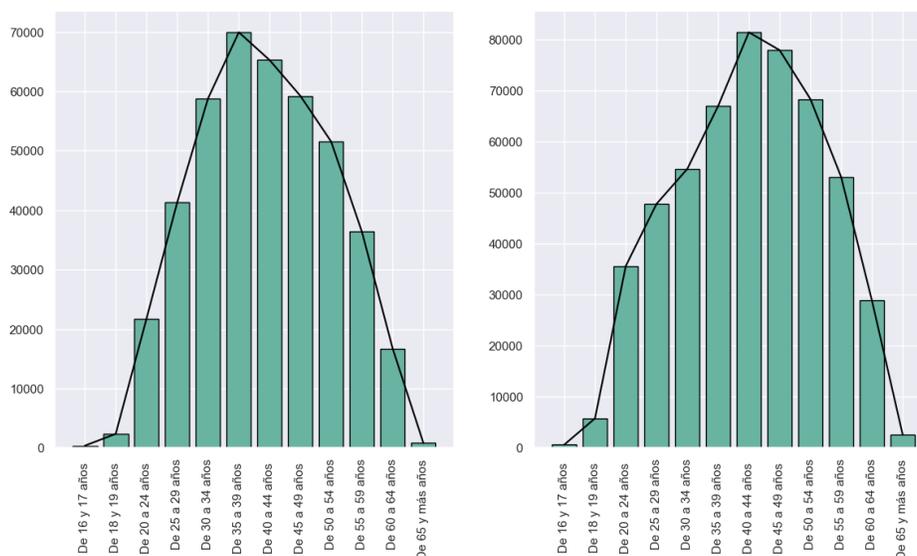


Figura 3.5: Número de accidentes durante la jornada en 2014 y 2021, en ese orden, según grupos de edad.

Simplemente visualizando los índices de cada una de las dos gráficas, ya se puede apreciar el aumento en número de accidentes desde 2014 a 2021. Sin embargo, la distribución en función de los grupos de edad apenas varía. Solamente destaca que el grupo con mayor número de accidentes pasa de ser de 35 a 39 años a ser el siguiente, de 40 a 44 años. En general, se observa cierta tendencia de aumento de edad en los trabajadores accidentados, seguramente provocada por el envejecimiento de la población.

En el siguiente enlace de [Datapane](#) se puede observar esta misma información para los años intermedios. En dicha web se puede acceder a las gráficas interactivas complementarias a este documento que no se han incluido en él. Las que no son interactivas se incluyen mediante enlaces a un repositorio de [Github](#). En ellas se muestran los diferentes resultados aquí mostrados pero a lo largo de todos los años entre 2014 y 2021. A medida que se vayan mencionando en el informe nuevas categorías, se adjuntan los enlaces que proporcionan acceso a las gráficas complementarias correspondientes.

En el diagrama de cajas y bigotes de la Figura 3.6 se pueden observar los valores entre los que se encuentran los diferentes grupos de edad a lo largo de los 8 años considerados, mostrando que la distribución general se mantiene más o menos similar en ese periodo de tiempo.

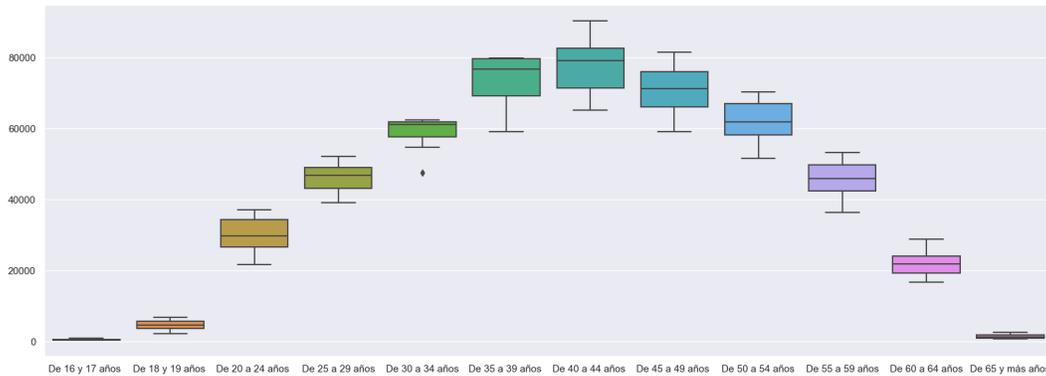


Figura 3.6: *Boxplot* que muestra la mediana y cuartiles del número de accidentes en jornada según grupos de edad para cada año.

En el caso de los accidentes *in itinere*, se puede apreciar una mayor variación en el número de accidentes según grupo de edad. Véase la Figura 3.7.

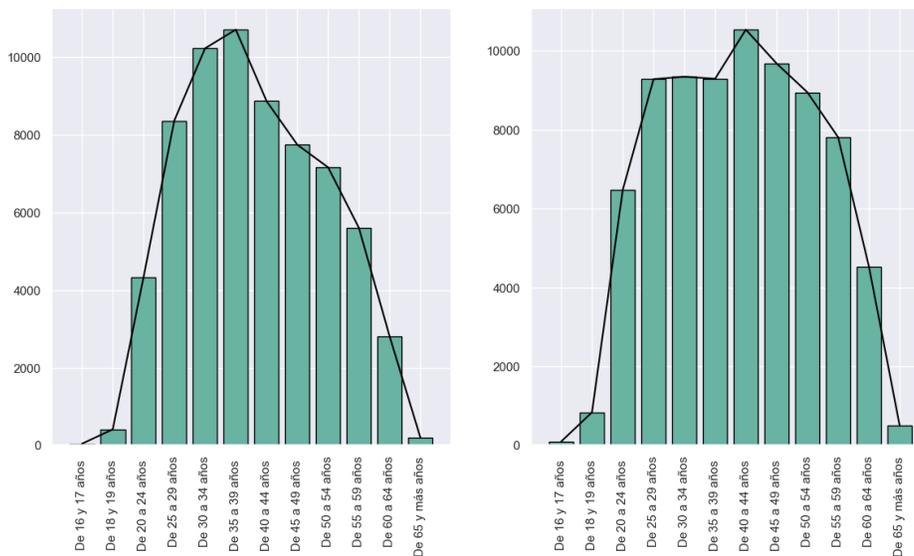


Figura 3.7: Número de accidentes *in itinere* en 2014 y 2021, en ese orden, según grupos de edad.

En este caso, no es tan notable que el grupo con mayor número de accidentes pase de ser de 35 a 39 a ser de 40 a 44. Lo más llamativo es que el número de accidentes ha aumentado en prácticamente todos los grupos de edad; destacando especialmente la subida en los mayores de 40.

De nuevo, al igual que con las gráficas para accidentes en jornada, se pueden visualizar los resultados para el [resto de años](#).

En la Figura 3.8 se puede observar el *boxplot* referente a los accidentes *in itinere*, que no muestra una distribución tan clara como la que se veía para los accidentes durante la jornada.

Un resultado notable que se observa en todas las gráficas mostradas hasta ahora es que los más jóvenes y los más mayores (los extremos) son los que presentan un menor volumen de accidentes laborales. Además, sin más que analizar los índices de dichas gráficas, se puede apreciar claramente el problema comentando inicialmente acerca de la subida en el número de accidentes laborales en España.

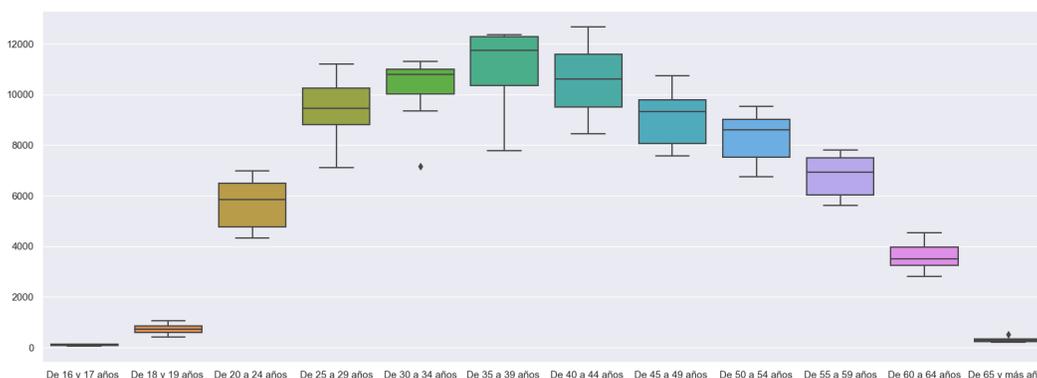


Figura 3.8: *Boxplot* que muestra la mediana y cuartiles del número de accidentes *in itinere* según grupos de edad para cada año.

De todas formas, es importante contextualizar la situación, pues como se puede observar en la Figura 3.9, esta subida viene ligada a un aumento del número de trabajadores ocupados en el país.

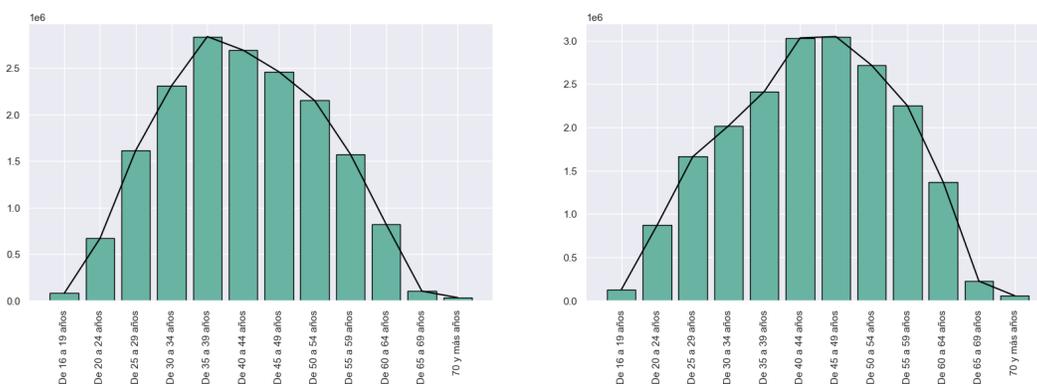


Figura 3.9: Población ocupada (en millones) en España en los años 2014 y 2021 respectivamente.

Tan solo observando los índices de las gráficas, se intuye el aumento de población ocupada desde el año 2014 al 2020. Concretamente, según la documentación proporcionada por el INE, la población ocupada pasó de 17.3 millones de personas en 2014 a 19.7 en 2021. La [evolución](#) a través de los años y las gráficas concretas de cada año muestran este crecimiento general (con la salvedad de la bajada en 2020). El aumento se aprecia en prácticamente todas las franjas de edad, salvo en el grupo de 35 a 39 años. En este grupo era donde también se apreciaba una ligera bajada en el número de accidentes.

A pesar de que la mayor cantidad de accidentes se encuentra en los grupos de edad centrales, también se acaba de ver que es la franja de edad donde se encuentra la mayoría de la población ocupada. Con lo cual, a la hora de determinar cuáles son las edades para las que hay un mayor número de personas accidentadas, tiene más sentido estudiarlo en término de proporciones que de valores absolutos.

La Figura 3.10, referente al año 2021 muestra de forma aproximada las proporciones de accidentes en función de la cantidad de población ocupada para las siguientes franjas de edad (redondeadas a las milésimas):

- De 16 a 19 años: 0.051
- De 20 a 24 años: 0.041
- De 25 a 29 años: 0.029

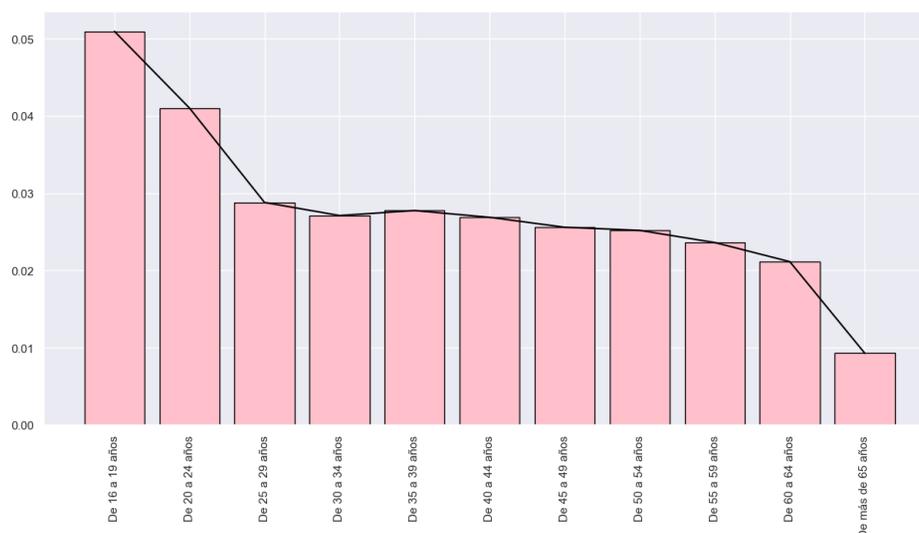


Figura 3.10: Proporción de accidentes laborales durante la jornada en función de la población ocupada según los grupos de edad en 2021.

- De 30 a 34 años: 0.027      ● De 35 a 39 años: 0.028      ● De 40 a 44 años: 0.027
- De 45 a 49 años: 0.026      ● De 50 a 54 años: 0.025      ● De 55 a 59 años: 0.024
- De 60 a 64 años: 0.021      ● Mayores de 65 años: 0.009

Los grupos de edad centrales, entre 25 y 55 años, no presentan una diferencia notable en proporción de accidentes entre ellos; rondan porcentajes entre el 2.9 y el 2.5%. Sin embargo, los grupos de edad más extremos, sí presentan unos resultados diferenciados: las personas mayores presentan una menor proporción de accidentes, sobre todo a partir de los 65 años; mientras que los más jóvenes superan claramente los porcentajes de los mayores de 25. Destaca sobre todo que las personas ocupadas entre 16 y 19 años son los que más accidentes laborales sufren (salvo en el año 2014), siendo uno de los grupos que presentan una menor cantidad de población ocupada. Por otro lado, los mayores de 65 también presentaban un bajo número de accidentes y aunque la cantidad de población ocupada en esa franja de edad es baja también, la proporción de accidentes es baja.

Con lo cual, a pesar de que en números absolutos las personas cuya edad se encuentra entre los 25 y 55 años sean las que sufren un mayor número de accidentes; en proporción, la situación de los jóvenes es más preocupante, pues prácticamente los duplican en el porcentaje de accidentes.

#### ■ Sexo del accidentado/a

A continuación se muestra el mismo análisis pero en función del sexo del empleado/a en lugar de su grupo de edad. En este caso, dado que se representa un número menor de categorías, se han desglosado los accidentes según su gravedad, aunque también se incluye el total.

Como se puede ver en la Figura 3.11, la mayoría de accidentes son leves, por lo que los graves y los mortales no tienen una especial incidencia. Aun así, es una información de la que se dispone y que resulta de interés en el caso de focalizarse en los accidentes de mayor gravedad, destacados en la

introducción de este trabajo por estar aumentando en los últimos años.

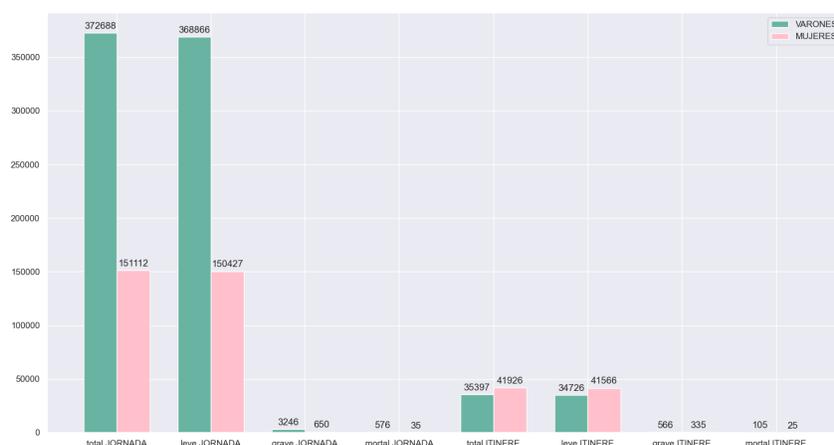


Figura 3.11: Número de accidentes según el sexo del trabajador/a y la gravedad del accidente en 2021.

En el último año, durante la jornada laboral, el número de accidentes laborales de varones supera por más del doble al número de accidentes de mujeres. Además, para los accidentes con mayor gravedad, la diferencia entre sexos es todavía mayor. Destacan, sobre todo, la diferencia entre los 576 fallecidos frente a las 35 mujeres fallecidas durante la jornada laboral.

*In itinere* también hay un número mayor de accidentes de varones que de mujeres, para aquellos que son más graves. Sin embargo, destaca que en cuanto a los accidentes leves, las mujeres sufren un número ligeramente superior de accidentes que los hombres: concretamente en 2021, las mujeres sufrieron 41586 accidentes de gravedad leve *in itinere*, mientras que los hombres, 34728.

A pesar de que las mujeres sufran más accidentes en los desplazamientos, el volumen de accidentes es mucho mayor durante la jornada, por lo que la mayor parte de accidentes los sufren varones.

Análogamente, esta gráfica se puede visualizar para los [anteriores años](#) (hasta 2014). Las conclusiones son similares a las ya comentadas, variando el número total de accidentes.

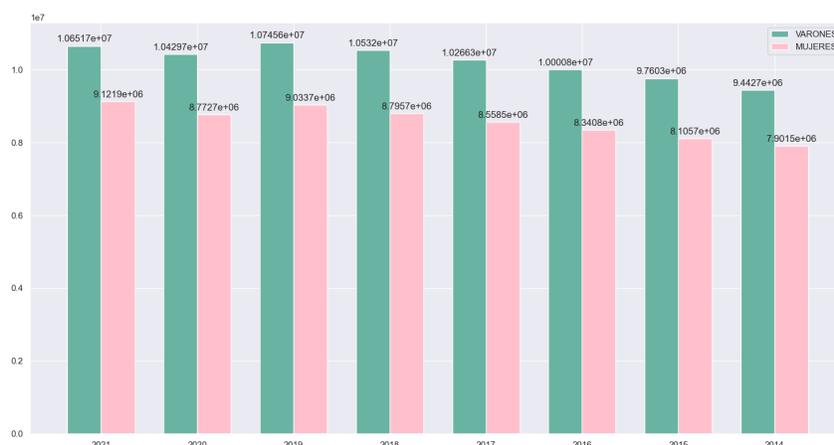


Figura 3.12: Población ocupada a lo largo de los años clasificada según el sexo.

Al igual que en el caso de los grupos de edad, y en general, es importante contextualizar los

resultados anteriores. En la Figura 3.12 se puede apreciar la población ocupada en España según el sexo de la persona. Se muestran, además de 2021, el resto de años considerados desde 2014.

De nuevo se puede ver el aumento de población ocupada en los últimos años. La nueva información es que el número de hombres ocupados es superior al de mujeres, tanto en 2021, como en los años anteriores. En estos años, de media, la diferencia entre varones ocupados y mujeres ocupadas es casi de un millón y medio de personas. De todas formas, aun habiendo un mayor número de varones ocupados, no compensa la gran diferencia en el número de accidentes entre ambos sexos.

Véase, por ejemplo, que en 2021 el número de mujeres ocupadas por cada hombre fue de 0.84 aproximadamente, mientras que el número de accidentes de mujeres por cada hombre accidentado fue de 0.55. Las proporciones para el resto de años son similares, verificándose que de media estos valores son 0.84 y 0.5, respectivamente. A pesar de la notable diferencia, con el paso de los años se ha ido reduciendo, aumentando el número de mujeres accidentadas por cada hombre accidentado.

En la Figura 3.13 se puede ver esto claramente a través de las proporciones de accidentados según ocupados en función del sexo.



Figura 3.13: Proporción de accidentes en función de ocupados según el sexo a lo largo de los años..

En media, el porcentaje de hombres accidentados es aproximadamente del 3.4 %, mientras que el de mujeres es del 1.7 %, la mitad. En la Figura 3.14 se puede apreciar como incluso los años que presentan peores resultados para el sexo femenino no son tan malos como los mejores resultados de los varones.

Con lo cual, no solo es mayor el número de accidentes laborales en el caso de los hombres, sino que también es mayor en proporción al número de personas ocupadas. Esta proporción es además nota-



Figura 3.14: *Boxplot* en el que se representan las proporciones entre 2014 y 2021 de accidentes frente a individuos ocupados según el sexo.

ble, pues la de las mujeres es aproximadamente la mitad. Estos resultados señalan que la problemática en la accidentalidad laboral pasa por una situación claramente más grave en el caso del sexo masculino.

En cuanto a los resultados según la población activa, la situación es similar a lo ya expuesto. La diferencia reside en que el número de personas activas es mayor que el de ocupadas, por lo que las proporciones de accidentes son menores en general. De todas formas, los resultados obtenidos fruto de la comparación entre proporciones son los mismos.

#### ■ Situación profesional del accidentado/a

Otro de los rasgos relevantes a la hora de estudiar el perfil de los trabajadores es su situación profesional, conocer si bien es un asalariado (del sector privado o público) o un autónomo (con o sin asalariados).

De nuevo, es importante contextualizar la situación. Para ello se han consultado las webs del INE<sup>5</sup> y del Ministerio de Trabajo<sup>6</sup> para obtener un número aproximado de trabajadores por cuenta ajena y por cuenta propia. De la primera de las referencias se ha obtenido que el número de asalariados en 2021 rondó los 16.6 millones de personas, de los cuales solo 3.44 millones trabajaron en el sector público. Además, también se puede ver que el número de trabajadores por cuenta propia fue de al rededor de 3.1-3.2 millones. En la segunda referencia, la información extraída es que a principio del año 2022 (31 de enero) el número de autónomos era aproximadamente de 3.3 millones.

En la Figura 3.15 se presentan las proporciones de accidentados en función de su situación profesional para los años 2014 y 2021. Se puede observar como los que sufren un mayor número de accidentes son los trabajadores por cuenta ajena en el sector privado; aunque también es cierto que constituyen gran parte de la población ocupada del país.

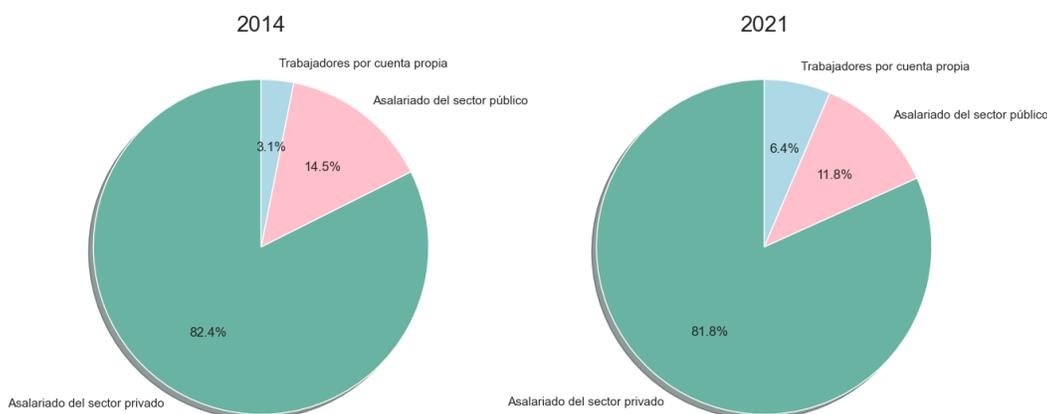


Figura 3.15: Proporciones del total de trabajadores según su situación profesional en 2014 y 2021, por ese orden.

Por otro lado, en España existe un número similar de trabajadores por cuenta propia y asalariados del sector público. Sin embargo, hay un número mayor de accidentes en el caso de los empleados del sector público. En la Figura 3.15 se puede observar esta diferencia, más notable en 2014 que en 2021. Los resultados para los años intermedios presentan una situación similar, pero se puede apreciar que

<sup>5</sup>INE. Última visualización el 24 de noviembre de 2022 [online](#). Seleccionar las categorías a consultar.

<sup>6</sup>Ministerio de Trabajo. Última vez consultado el 24 de noviembre de 2022 [online](#).

la diferencia entre ambas categorías se va reduciendo paulatinamente.

Resulta interesante considerar de nuevo los resultados en términos de proporciones. Considerando la cantidad de población ocupada según su ocupación obtenida en las referencias mencionadas y el Cuadro 3.1, obtenida a partir de los datos del Ministerio, se obtienen las siguientes proporciones para cada clase:

- Asalariados del sector privado: 0.0325
- Asalariados del sector público: 0.0183
- Trabajadores por cuenta propia: 0.0107.

Estos resultados corresponden al año más reciente del que se tiene datos, 2021.

	2014	2015	2016	2017	2018	2019	2020	2021
Asalariado (priv.)	349925	382715	415294	441923	463542	478828	367122	428219
Asalariado (públ.)	61423	62971	61560	61057	57932	50593	49315	62069
Autónomo	13277	12337	12211	12102	11503	33335	29758	33512

Cuadro 3.1: Número de accidentados según su situación profesional.

Lógicamente, las proporciones varían de año a año, pero se mantienen en unos rangos similares.

Se deduce entonces, que las personas autónomas sufren, en general, un número menor de accidentes. A pesar de ello, dentro de esta categoría, hay un grupo que destaca notablemente en el número de accidentes laborales sobre el otro.

Un trabajador por cuenta propia puede trabajar él solo para sí mismo o tener empleados contratados. El número de accidentes laborales varía mucho según el caso. En la Figura 3.16 se aprecia la diferencia existente entre estas dos categorías a lo largo de los años.



Figura 3.16: Número de accidentes laborales de trabajadores por su cuenta según tengan o no asalariados.

Destaca, en primera instancia, el aumento de accidentes en los últimos tres años, que llegan incluso a triplicar los casos de años anteriores. Además, resulta claro que los autónomos sin asalariados en sus

empresas o proyectos profesionales, sufren una cantidad mayor de accidentes. En la mayoría de los años considerados, el número de accidentes supera el triple de los accidentes de autónomos con asalariados.

De nuevo a partir del INE<sup>7</sup>, se puede extraer el número de personas que son autónomos y, específicamente, si tienen o no asalariados. Durante los 8 años considerados, el número de trabajadores por cuenta propia sin asalariados ha rondado los dos millones de personas, superándolas apenas por decenas de miles. Los autónomos con asalariados son muchos menos, ni siquiera llegan al millón, rondando los novecientos mil (en ocasiones por encima y en los últimos años más por debajo).

Con lo cual, dado que en número de accidentes triplican y en número de autónomos solo duplican (aproximadamente), no solo el número absoluto de accidentes es mayor en el caso de los autónomos sin asalariados, sino que en proporción la diferencia también destacable. Atendiendo solamente a los resultados de 2021, se tiene que el número de accidentes para autónomos sin asalariados es 24489 y con asalariados, 9023. Por otro lado, el número de personas que trabajan por cuenta propia sin empleados es 2082100, mientras que los que disponen de empleados son 944200. Por tanto, la proporción de accidentados de los primeros es 0.012 y de los segundos, 0.0096. Se tiene que tanto en proporción, como números totales, los autónomos que no disponen de empleados son más vulnerables a los accidentes laborales.

#### ■ Tipo de contrato del accidentado/a

Además de estudiar la influencia de la situación profesional del trabajador, el tipo de contrato del que este dispone también muestra diferencias en los accidentes laborales sufridos. En la Figura 3.17 se puede observar el número de accidentes durante la jornada según el tipo de contrato; diferenciado si el contrato es indefinido o temporal y, dentro de esta clasificación, si es a tiempo parcial, completo o fijo discontinuo. Los datos mostrados se corresponden con los años 2014 y 2021.

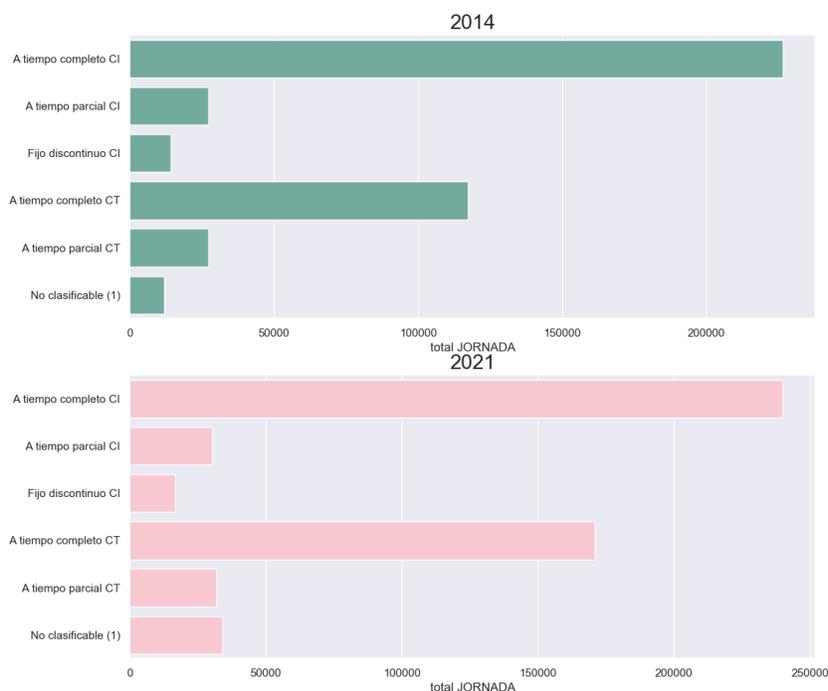


Figura 3.17: Número de accidentes en jornada en 2014 y 2021 según el tipo de contrato del trabajador.

<sup>7</sup>INE. Última visualización el 24 de noviembre de 2022 [online](#).

Los resultados para los [años intermedios](#), al igual que en la mayoría de las clasificaciones estudiadas hasta ahora, presentan resultados muy similares entre ellos (con la salvedad de que en general aumenta el número de accidentes en casi todas las categorías).

En la Figura 3.18 se puede ver la misma información, pero en este caso sobre los accidentes producidos *in itinere*.

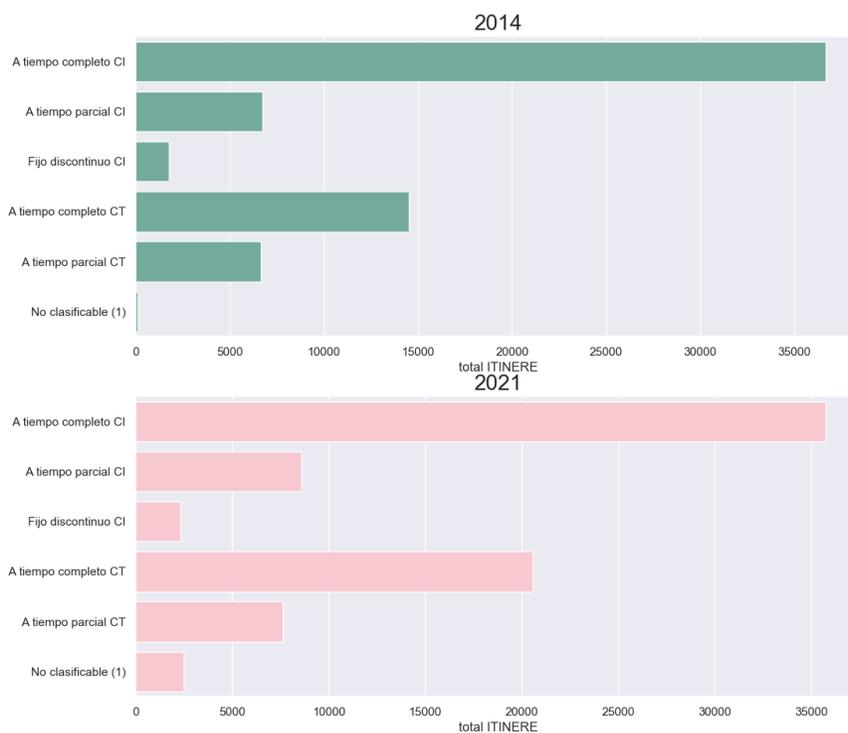


Figura 3.18: Número de accidentes *in itinere* en 2014 y 2021 según el tipo de contrato del trabajador.

De nuevo, los años intermedios tanto [en jornada](#) como [in itinere](#) presentan resultados similares. La distribución de los accidentes en las diferentes clases es muy parecida a la vista para los accidentes durante la jornada, con la salvedad de que estos últimos ocurren en un volumen mayor de veces; por ejemplo, en 2021 el número de accidentados con contrato indefinido a tiempo completo durante la jornada fue 240116 e *in itinere* fue 35767, mientras que en 2014 fueron 226536 y 36682, respectivamente.

Tanto durante la jornada como *in itinere*, la situación de los accidentes laborales es parecida. La gran mayoría de los accidentes le ocurren a personas con contratos a tiempo completo. En mayor medida se accidentan aquellas personas cuyo contrato es indefinido, comparándolo con aquellas con contrato temporal.

De nuevo, desde el INE<sup>8</sup> se puede obtener el número de ocupados en España según las categorías deseadas. Durante los 8 años considerados, el número de trabajadores con contrato indefinido ha ido aumentando, pasando de casi 11 millones en 2014 a 12.5 millones en 2021. Los trabajadores con contratos parciales también han aumentado, siendo 3.5 millones en 2014 y casi 4.2 en 2021. A pesar de estas variaciones, en general, la cantidad de trabajadores indefinidos supera por mucho el número de trabajadores con contrato temporal, de la misma manera que lo supera en número de accidentes laborales.

La información disponible en la web del INE indica el número total de trabajadores con contrato indefinido y con contrato temporal, pero los desglosa en unas categorías diferentes de las presentadas

<sup>8</sup>INE. Última vez consultado el 30 de noviembre de 2022 [online](#).

en los documentos del Ministerio de Trabajo, por lo que en este caso no se calcularán las proporciones en función de la población ocupada. Tampoco se dispone de información sobre la población activa española para realizar este tipo de cálculos.

#### ■ **Ocupación del accidentado/a**

El siguiente aspecto respecto al cual se analizan los accidentes de trabajo es la ocupación de los trabajadores. La ocupaciones que se consideran son las siguientes:

- **A.** Directores y gerentes
- **B.** Técnicos y profesionales científicos e intelectuales de la salud y la enseñanza
- **C.** Otros técnicos y profesionales científicos e intelectuales
- **D.** Técnicos; profesionales de apoyo
- **E.** Empleados de oficina que no atienden al público
- **F.** Empleados de oficina que atienden al público
- **G.** Trabajadores de los servicios de restauración y comercio
- **H.** Trabajadores de los servicios de salud y el cuidado de las personas
- **I.** Trabajadores de los servicios de protección y seguridad
- **J.** Trabajadores cualificados en el sector agrícola, ganadero, forestal y pesquero
- **K.** Trabajadores cualificados de la construcción, excepto operadores de máquinas
- **L.** Trabajadores cualificados de las industrias manufactureras, excepto operadores de instalaciones y máquinas
- **M.** Operadores de instalaciones y maquinaria fija, y montadores
- **N.** Conductores y operadores de maquinaria móvil
- **O.** Trabajadores no cualificados en servicios (excepto transportes)
- **P.** Peones de la agricultura, pesca, construcción, industrias manufactureras y transportes.

En la Figura 3.19 se pueden visualizar los resultados para 2021. Se observa la proporción del total de accidentes durante la jornada que sufre cada una de estas categorías en ese año. Los resultados para los 7 años anteriores muestran una distribución semejante, con pequeñas variaciones.

Las ocupaciones que destacan por tener en general una mayor accidentalidad laboral son las indicadas con las letras **G**, **L**, **O** y **P**, seguidas de **K** y **N**. La cuatro primeras se corresponden con servicios de restauración y comercio (51915 accidentes en 2021), industrias manufactureras (56903), trabajos no cualificados (50368) y trabajos como peones en agricultura, pesca, construcción, industria y transportes (68986). Las ocupaciones **K** y **N** se corresponden con trabajadores cualificados para la construcción (32886) y conductores y operadores de máquinas móviles (30674).

En general, durante la jornada, los trabajadores que mayor número de accidentes sufren son aquellos que trabajan como peones, tanto en el sector primario como secundario, trabajadores en industria no cualificados y también personal de servicios de restauración.

En la Figura 3.20 se puede observar un gráfico análogo al anterior pero que incluye la información en referencia a los accidentes que han tenido lugar en desplazamiento a o desde el lugar de trabajo.

Destacan de nuevo los trabajadores de servicios de restauración, con 11923 accidentes en 2021. Las siguientes ocupaciones más afectadas rondan los 5000-6000 accidentados y son aquellas indicadas con



Figura 3.19: *Treemap* en el que se muestra la proporción de accidentes durante la jornada en 2021 según la ocupación del accidentado.

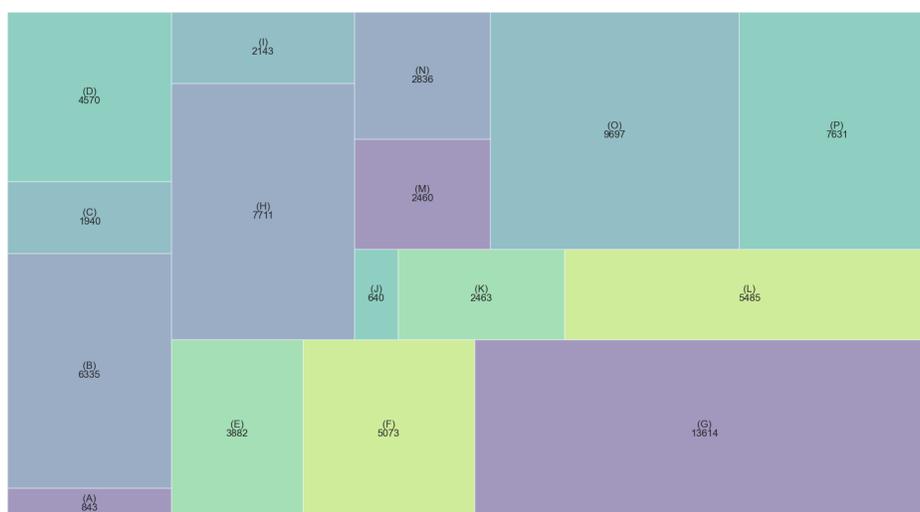


Figura 3.20: *Treemap* en el que se muestra la proporción de accidentes *in itinere* en 2021 según la ocupación del accidentado.

las letras **B**, **D**, **E**, **F**, **H**, **O** y **P**. Concretamente los empleados de oficina de cara al público (**F**) y los trabajadores no cualificados en servicios (**O**) son los más destacados de este grupo.

Nótese que los accidentes *in itinere*, a priori, no deberían estar especialmente relacionados con el trabajo desempeñado en el puesto de trabajo.

Se pueden consultar los *treemap* para el [resto de años](#) considerados, aunque no se aprecian variaciones demasiado notables entre ellos.

Tras la revisión del número de accidentes tanto en jornada como *in itinere*, parece que los grupos de los trabajadores no cualificados en servicios; y los peones en agricultura, pesca, construcción, industrias manufactureras y transportes son los más perjudicados, especialmente por la cantidad de accidentes que se sufren durante la jornada laboral.

Para esta clasificación resulta complicado encontrar datos sobre la cantidad de población ocupada correspondiente a cada ocupación, por lo que no es posible obtener los resultados en proporciones. De todas formas, la información dada mediante los valores totales es válida para extraer conclusiones útiles.

#### ■ Nacionalidad del accidentado/a

Para finalizar con el análisis de los accidentes según el perfil del accidentado, se estudia la distribución de los accidentes según la nacionalidad de los trabajadores.

La información más completa se puede encontrar en Datapane. Las gráficas que se presentan desglosan la información por países y continentes para los 8 años considerados: 2021, 2020, 2019, 2018, 2017, 2016, 2015 y 2014. En este documento solo se incluirán los resultados en función de los continentes y destacando también España, pues un gráfico (no interactivo) con todos los países sería difícil de visualizar e interpretar.

En la Figura 3.21 se muestran simplemente los resultados para el año 2021, pues en términos de continentes las diferencias que existen entre los distintos años considerados no son muy notables. En dicha figura se muestran tres gráficas: una primera en la que se comparan los accidentes según continentes, otra en la que se excluyen los trabajadores españoles del cómputo y una última en la que se comparan los accidentes según continentes distinguiendo por separado a los españoles.

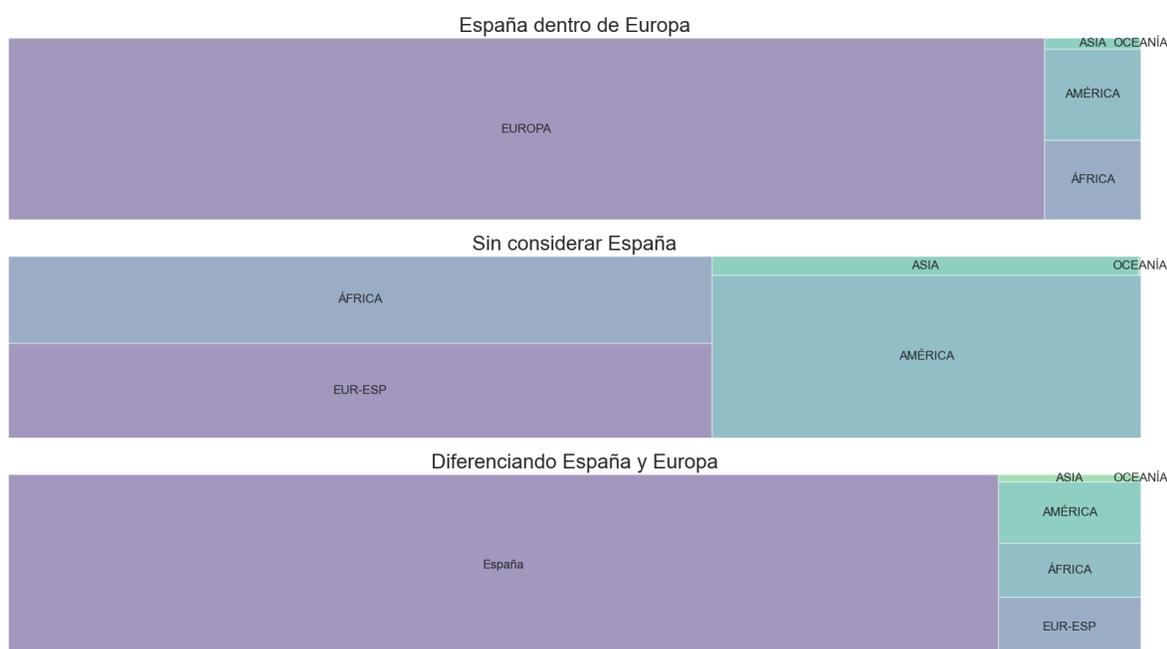


Figura 3.21: *Treemaps* que contienen información sobre el número de accidentados durante la jornada en el ámbito laboral según su nacionalidad.

En la primera de las tres gráficas se puede observar que Europa es el continente que más accidentados en el territorio español posee. Sin embargo, observando detenidamente las otras dos gráficas, se puede apreciar que la mayor parte de los europeos accidentados son españoles. De hecho, observando la última de ellas, se puede ver que España ocupa más del 80% del gráfico, lo que quiere decir que más del 80% de los accidentados en España durante la jornada laboral son españoles.

De nuevo, resulta interesante contextualizar, pues según el INE<sup>9</sup>, de los más de 19 millones de per-

<sup>9</sup>INE. Última vez consultado el 30 de noviembre de 2022 [online](#).

sonas que componen en 2021 la población ocupada del país, 16.6 millones son españoles, un porcentaje muy alto.

Dejando a un lado los españoles, los accidentes restantes se reparten mayoritariamente entre África, América y Europa en cantidades bastante similares. De Asia hay un número bastante menor y de Oceanía apenas hay accidentados.

Los números concretos de 2021 son los que se pueden visualizar en la siguiente tabla:

	España	Europa (sin ESP)	África	América	Asia	Oceanía
Durante jornada	457820	21373	19559	22349	2581	14
<i>In itinere</i>	71756	2341	1798	3472	291	1

En este caso, vuelve a ser interesante calcular las proporciones en función de la población ocupada de cada nacionalidad; sin embargo, no es posible, pues no se ha encontrado información oficial de la población ocupada con las mismas categorías por continentes. Por ejemplo, en el INE, la clasificación de extranjeros no distingue entre africanos, asiáticos y oceánicos. Dadas estas limitaciones, se muestran solamente la proporción de accidentes para españoles y para extranjeros en general. Se calcularán las del año 2021 tanto para accidentes durante la jornada, como para los que ocurren durante desplazamientos al comenzar y finalizar la jornada.

- Españoles: 0.0274 durante la jornada y 0.0043 *in itinere*.
- Extranjeros: 0.0276 durante la jornada y 0.00504 *in itinere*.

Tal y como se puede observar en función de estos resultados, la diferencia de accidentes en proporción entre españoles y extranjeros es muy pequeña. La proporción es ligeramente mayor en el caso de los extranjeros, pero en términos totales se traduce en una menor cantidad, debido a que son menos en general. De todas formas, no parece que la nacionalidad por sí sola provoque una gran diferencia en cuanto a la accidentalidad laboral en España, según los registros oficiales.

### 3.2.2. Características de la empresa

Se ha visto en la sección anterior que la proporción de accidentes varía notablemente según las características de los trabajadores. Entra dentro de lo razonable esperar que las características de las empresas también influyan en la proporción de accidentes laborales.

En general, este nuevo apartado se centra sobre todo en el estudio de los accidentes durante la jornada; pues se entiende que la organización de la empresa tendrá una influencia mucho menor en los accidentes *in itinere*.

A continuación se presenta cómo se distribuyen los accidentes en función de las características de las empresas en las que trabajan los accidentados.

#### ■ Lugar del accidente

En primer lugar se muestran las estadísticas en función del lugar donde tuvo lugar el accidente. Las tres categorías consideradas son: “centro habitual de trabajo”, “otro centro” y “desplazamiento al lugar de trabajo”. Esta última categoría no engloba los desplazamientos fuera del horario laboral para acudir al puesto de trabajo o marcharse de él sino que considera los desplazamientos por motivos laborales.

En la Figura 3.22 se puede observar la cantidad de accidentes laborales que tuvieron lugar para cada uno de los tres casos para cada uno de los 8 años considerados. Como se ha comentado desde el

principio del trabajo, de nuevo aquí se puede apreciar la subida en el número de accidentes general y la bajada del año 2020. Esto se puede observar en general en todas las gráficas que presentan datos a lo largo de los años.



Figura 3.22: Número de accidentes según el lugar donde ocurrieron.

En 2021, 57672 personas se accidentaron durante un desplazamiento; 29197, en un centro que no es el suyo y 436931 sufrieron el accidente en su propio centro de trabajo. Aunque superiores a los resultados de 2020, no alcanzan los valores máximos de 2019.

En este caso, no se dispone del número de población ocupada que se corresponde con cada categoría. Es más, un mismo trabajador puede, durante la jornada, pasar tiempo en su centro de trabajo y fuera de él. Por lo tanto, no se pueden obtener las proporciones en función de la población ocupada. Solo se podrán extraer conclusiones a partir de los números totales.

Resulta claro que la gran mayoría de accidentes laborales tienen lugar en el propio centro de trabajo. *A priori* parece razonable, pues es donde se espera que generalmente un trabajador pase más tiempo durante su jornada. La diferencia entre las otras dos categorías es mucho menor; pero aun así en todos los años se produce un mayor número de accidentes en desplazamiento, en comparación con los que ocurren en otros centros.

Para completar esta categoría, faltaría analizar los accidentes que tienen lugar al ir y volver del trabajo; pero estos son concretamente los accidentes clasificados como *in itinere*, que ya han sido analizados al inicio. De todas formas, se incluye la Figura 3.23 para comentarla con respecto a la anterior.

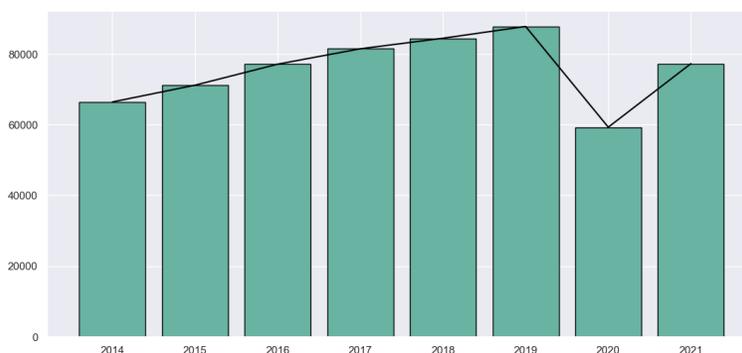


Figura 3.23: Número de accidentes *in itinere* a lo largo de los años.

De nuevo, se observa el aumento de accidentes, destacando 2019 y la bajada de 2020. En cuanto al

número de accidentes, es mucho menor que el número durante la jornada laboral en el centro de trabajo, pero supera en todo caso a los accidentes en desplazamiento durante la jornada (y por consiguiente a los que ocurren en un centro diferente del propio).

La mayor parte de los accidentes ocurren en el centro de trabajo, seguido con una distancia considerable por los accidentes acudiendo o yéndose de él.

#### ■ Tipo de empresa contratante del accidentado/a

Otra de las características que puede resultar interesante analizar es el tipo de empresa que contrata a los trabajadores que sufren los accidentes. El Ministerio considera tres opciones: que el empleado sea contratado por la propia empresa, por una subcontrata o por una Empresa de Trabajo Temporal (ETT). En la Figura 3.24 se puede observar el número de accidentes según el tipo de empresa contratante para cada uno de los años considerados.



Figura 3.24: Número de accidentes durante la jornada laboral según el tipo de empresa contratante.

De nuevo se observa una tendencia creciente, salvo por la bajada del 2020. Y al igual que en la clasificación anterior, hay una categoría que destaca por encima de todas. En este caso, la mayoría de los accidentes le ocurren a empleados de la propia empresa. De todas formas, esta situación seguramente se deba a que la gran mayoría de empleados españoles son contratados por la propia empresa en la que trabajan.

En la web del Ministerio de Trabajo<sup>10</sup> se puede ver la cantidad de contratos de trabajadores entre las ETT y las empresas usuarias de estas. En los últimos años (excluyendo el 2020) el número de este tipo de contratos ha superado los 4 millones; lo cual no quiere decir que 4 de los 17 millones de trabajadores en España se contraten mediante ETT, pues una parte de los trabajadores seguramente haya tenido más de uno de estos contratos a lo largo del año. De todas formas, estos datos sirven para contextualizar mínimamente la situación. Los contratados por ETT son minoría en comparación con las otras dos categorías, de las cuales “contratados por la propia empresa” es razonable que se corresponda con una gran parte de la situación de los empleados.

Por tanto, la gran cantidad de accidentes que presenta la categoría “contratados por la propia empresa” seguramente esté influido por el alto volumen de trabajadores que se encuentran en esa situación.

En la Figura 3.25 se puede ver información análoga a la ya presentada pero en referencia a los accidentes *in itinere*.

<sup>10</sup>Página del Ministerio de Trabajo que da acceso a los Excel con la información de cada año. Última vez consultado el 2 de enero de 2023 [online](#).

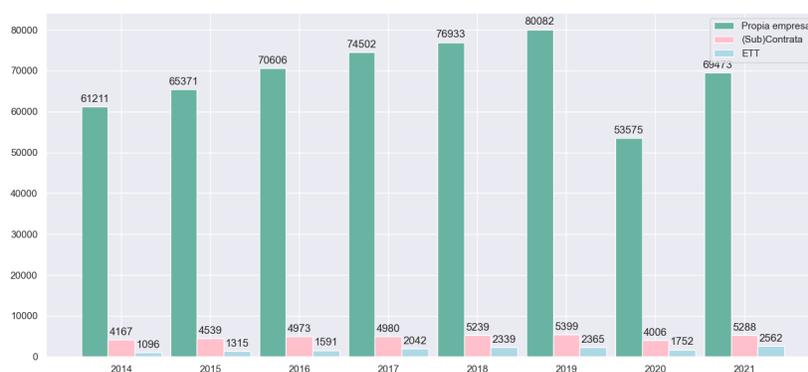


Figura 3.25: Número de accidentes *in itinere* según el tipo de empresa contratante.

El análisis es totalmente similar, con la salvedad de que el volumen de accidentes es menor que durante la jornada.

#### ■ Nivel de prevención laboral de la empresa

Un aspecto que también puede ser relevante a la hora de evaluar una empresa por sus accidentes laborales, es considerar el nivel de prevención de riesgos laborales que presentan. A continuación, en la Figura 3.26, se puede visualizar el número de accidentes laborales que han tenido lugar en las empresas según el tipo de servicio de prevención que poseen.

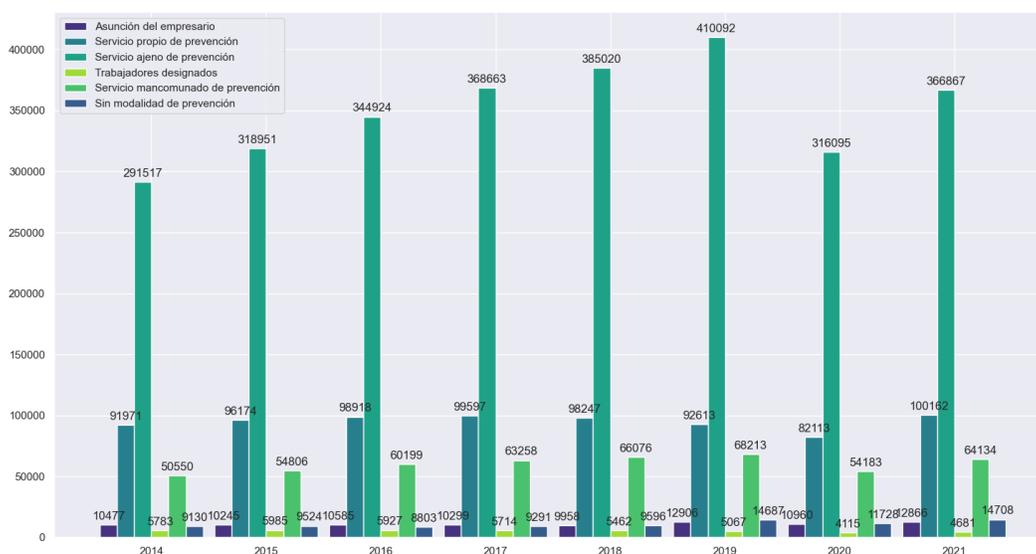


Figura 3.26: Número de accidentes durante la jornada según el tipo de servicio de prevención de riesgos laborales que posee la empresa.

Un detalle de esta gráfica es que la suma de las columnas de cada año da un valor superior al total de accidentes del propio año. Esto se debe a que hay accidentes que se contabilizan más de una vez, pues la empresa que contrata al trabajador posee más de un método de prevención.

A pesar de ello, la mayor parte de los accidentes ocurren en empresas cuyo servicio de prevención de riesgos laborales depende de un servicio ajeno. Esto ocurre así en gran medida porque la mayor

parte de las empresas españolas delegan en una empresa ajena su gestión de la seguridad laboral. El INSST<sup>11</sup> informa de que entre los años 2016 y 2019, alrededor del 90 % de las empresas españolas seguían este sistema, lo cual es útil para intuir que la mayoría de trabajadores en España están bajo el amparo de empresas ajenas en este aspecto. Resulta entonces clara la razón por la que una cantidad tan grande de los accidentes (366867 de un total de 523800 durante la jornada en 2021) ocurren en empresas con este tipo de prevención. En el informe del INSST también se puede ver que el resto de posibilidades están presentes en mucha menor medida en las empresas, siendo “la designación de un propio trabajador” el servicio de prevención más destacado de las restantes posibilidades. En la Figura 3.26 se puede ver que es uno de los métodos que menos accidentes laborales presenta. Por otro lado, destaca que, a pesar de contar con un porcentaje bajo de empresas que lo aplican, el servicio de prevención propio es el segundo método que más accidentes presenta a lo largo de los años.

Los servicios de mancomunado son los terceros que peores resultados totales muestran. El otro método considerado es la asunción de los riesgos por parte del propio empresario, que en números totales no destaca gravemente frente al resto de métodos.

En la Figura 3.27 se puede ver la misma información, pero en este caso restringiéndola solamente a los accidentes mortales. Los resultados son muy similares al caso anterior; siendo la mayor parte de los accidentados empleados de una empresa con servicio de prevención ajeno.

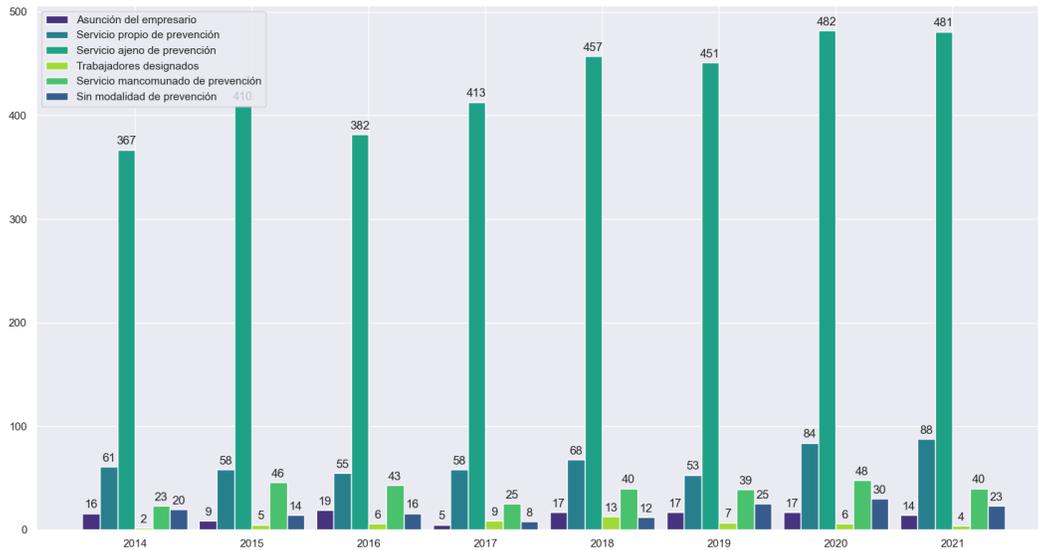


Figura 3.27: Número de accidentes laborales mortales durante la jornada según el tipo de servicio de prevención de riesgos laborales que posee la empresa.

Cabe destacar que, en este caso, no se observa un descenso de casos en 2020. Al inicio del análisis ya se comentó que, a pesar de que en general en 2020 hubo un descenso de los accidentes laborales, esto no ocurría para aquellos que acabaron en fallecimiento del trabajador.

También resulta interesante destacar la existencia de la Ley 31/1995, de 8 de noviembre, de prevención de Riesgos Laborales. Esta ley se aplica a todos los autónomos y empresas con al menos 1 asalariado y su cumplimiento obliga a la instauración de medidas de prevención de riesgos laborales. En relación con esto, llama la atención que en las gráficas anteriores haya accidentados cuyas empresas no tienen modalidad de prevención. Estos trabajadores se deben corresponder con autónomos que no tienen trabajadores a su cargo. Aunque el número de accidentados totales no sea especialmente alto en esta categoría, hay que tener en cuenta que el número de trabajadores considerados es menor que cualquiera de las otras categorías, pues solo deberían considerarse para ella los autónomos sin traba-

<sup>11</sup>INSST. Última vez consultado el 3 de diciembre de 2022 [online](#).

jadores a su cargo.

Además de contar con un servicio de prevención, es interesante tener en cuenta las medidas que se toman en dicho servicio. Una de las más interesantes y de las cuales recoge información el Ministerio es si las empresas realizan o no evaluaciones del riesgo existente en los puestos de trabajo de sus empleados.



Figura 3.28: Número de accidentes durante la jornada según la empresa realiza o no evaluaciones de riesgo.

En la Figura 3.28 se observa que el número de accidentes totales durante la jornada laboral fue mucho mayor en el caso de las empresas con evaluación de riesgos. En todos los años considerados, estas empresas duplican el número de accidentes en comparación con aquellas que no realizan la evaluación. Este resultado es sorprendente *a priori*, pero de nuevo hay que considerar el número de empresas que llevan a cabo cada uno de los métodos.

En el informe del INNST<sup>12</sup> se indica que en 2018-19 más del 70 % de las empresas realizaban un seguimiento de la eficacia de las medidas establecidas en la planificación de reducción de riesgos laborales. En dicho informe no se dispone de datos más recientes (2021, por ejemplo), pero sirven para dar una referencia de la situación. Es decir, prácticamente tres cuartos de las empresas españolas realizan evaluaciones sobre los métodos de prevención de riesgos aplicados a sus puestos de trabajo, lo que muy probablemente se traduce (no conocemos de qué tamaño son las empresas que sí llevan a cabo la evaluación y las que no) en que la mayoría de trabajadores trabajan en una empresa que realiza evaluaciones de los riesgos. Además, aunque no sepamos exactamente qué empresas aplican estas técnicas o no, la Ley 31/1995 de Prevención de Riesgos Laborales exige más requisitos a cumplir a aquellas empresas de mayor tamaño (con mayor número de trabajadores), de forma que se puede esperar que los trabajadores de las grandes empresas cuenten con evaluaciones de riesgos sobre su puesto de trabajo.

Por tanto, el alto número de accidentes de las empresas con evaluación de riesgos, que a priori parecía contradictorio, se entiende al tener en cuenta que una gran parte de las empresas españolas presentan este tipo de servicios.

#### ■ Tamaño de la empresa

Para finalizar esta subsección, se analiza el impacto que puede tener el tamaño de las empresas en el número de accidentes laborales que tienen lugar en ellas.

<sup>12</sup>INNST. Página 44. Última vez consultado el 7 de diciembre de 2022 [online](#).

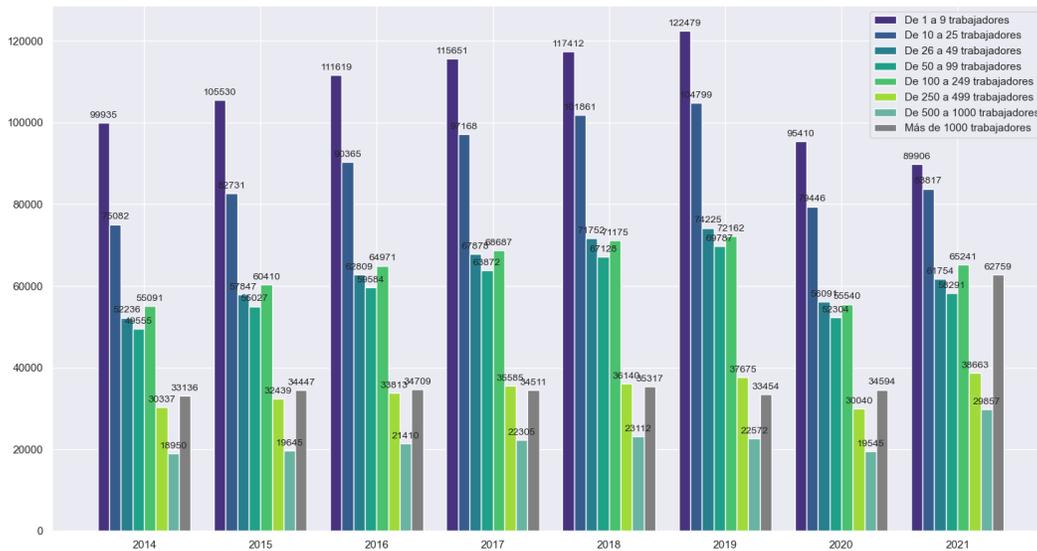


Figura 3.29: Número de accidentes durante la jornada según el número de empleados que presenta la empresa a la que pertenece el trabajador.

Se muestran los resultados para los 8 años considerados en la Figura 3.29.

En general, los trabajadores que más accidentes laborales sufren son aquellos que trabajan en pequeñas empresas. Los trabajadores de empresas entre 1 y 9 trabajadores han superado los cien mil accidentes en casi todos los años hasta 2020, donde el número de accidentados descendió. Los trabajadores de empresas de entre 10 y 25 empleados son los segundos más accidentados. En general, se observa que cuanto menor es la empresa a la que pertenece el trabajador, mayor es el número de accidentes. La excepción notable a este criterio es que las empresas de más de mil trabajadores superan los resultados de las empresas de 250 a 499 y de 500 a 1000.

Por tanto, en números totales, los trabajadores de empresas pequeñas sufren más accidentes que los de empresas grandes; a excepción de aquellas entre 250 y 1000 empleados que son superadas por las de más de mil empleados. De todas formas, esto puede simplemente darse porque hay un mayor número de trabajadores en pequeñas empresas que en grandes empresas, por lo que se debe tener este aspecto en cuenta a la hora de compararlas.

En la página web del INE<sup>13</sup> se dispone de información acerca del número de empresas que hay según su tamaño, organizado por categorías. Por ejemplo, para 2021 se han extraído los siguientes datos:

- Empresas totales: 3.366.570
- Empresas sin asalariados: 1.879.126
- Empresas de 1 a 2 asalariados: 920.321
- Empresas de 3 a 5 asalariados: 304.095
- Empresas de 6 a 9 asalariados: 120.828
- Empresas de 10 a 19 asalariados: 75.522
- Empresas de 20 a 49 asalariados: 42.012

- Empresas de 50 a 99 asalariados: 11.959
- Empresas de 100 a 199 asalariados: 6.655
- Empresas de 200 a 249 asalariados: 1.380
- Empresas de 250 a 999 asalariados: 3.634
- Empresas de 1000 a 4999 asalariados: 867
- Empresas de 5000 o más asalariados: 171

Con lo cual resulta claro que el número de empresas disminuye a medida que aumenta el tamaño de estas. Pero dado que no indican el tamaño exacto, sino un intervalo, no es posible conocer el número preciso de trabajadores que se corresponden con cada tamaño de empresa. Aun así, se pueden hacer estimaciones mediante intervalos. Para ello, se multiplica el número de empresas por el número de empleados correspondientes; como el número de empleados se da en forma de intervalo, se obtiene un intervalo que contiene el número real de trabajadores para ese tamaño de empresas. A continuación se muestra el resultado para todos los tamaños de empresa en 2021:

- Trabajadores de empresas de 1 a 2 asalariados: 920321-1840642.
- Trabajadores de empresas de 3 a 5 asalariados: 912285-1520475.
- Trabajadores de empresas de 6 a 9 asalariados: 724968-1087452.
- Trabajadores de empresas de 10 a 19 asalariados: 755220-1434918.
- Trabajadores de empresas de 20 a 49 asalariados: 840240-2058588.
- Trabajadores de empresas de 50 a 99 asalariados: 597950-1183941.
- Trabajadores de empresas de 100 a 199 asalariados: 665500-1324345.
- Trabajadores de empresas de 200 a 249 asalariados: 276000-412620.
- Trabajadores de empresas de 250 a 999 asalariados: 908500-3630366.
- Trabajadores de empresas de 1000 a 4999 asalariados: 867000-4334133.
- Trabajadores de empresas de 5000 o más asalariados: más de 855000.

Entonces habrá, por ejemplo, entre 2557574 y 4448569 trabajadores en empresas de 1 a 9 asalariados y más de 1722000 trabajadores en empresas de más de 1000 asalariados.

Las cotas inferior y superior de las empresas de 1 a 9 empleados son las más altas; coincidente con que son el tipo de empresas con mayor número de accidentes. Las empresas de 10 a 19 y de 20 a 49 asalariados presentan unas de las cotas más altas también y se corresponden con los segundos grupos que más accidentes presentan.

Las empresas de 250 a 999 y de 1000 a 4999 asalariados también poseen de las cotas más altas en cuanto a número de trabajadores. Para comparar las primeras con los resultados de la Figura 3.29, habría que juntar los accidentes de las empresas de 250 a 499 empleados y los de las empresas de 500 a 999. De todas formas, la suma de ambas categorías no superaría el número de accidentes de las empresas más pequeñas (1-9 asalariados y 10-25 asalariados).

En general, los grupos que presentan (mediante cotas) un menor número de empleados, se corresponden también con un menor número de accidentes.

Dado que las categorías dadas por el INE y por el Ministerio no coinciden, los cálculos no han sido exactos, pero parece claro que, en general, el número de accidentes por número de empleados en la empresa está bastante relacionado con el número total de trabajadores que hay en empresas de dicho tamaño. Es decir, para aquellas empresas de las cuales hay más trabajadores, más accidentes ocurren. Concretamente, las empresas de menor tamaño son las que mayor número de accidentes presentan.

### 3.2.3. Características del accidente

Hasta el momento se han revisado qué características de los empleados y sus empresas derivan en un mayor o menor número de accidentes. Es decir, cómo se distribuyen los accidentes según los perfiles de los trabajadores y las empresas.

En este nuevo apartado, la idea es estudiar los diferentes tipos de accidentes y cuáles son aquellos más recurrentes. Se tienen en cuenta las lesiones que se pueden originar, las partes del cuerpo dañadas, el lugar donde tuvo lugar el accidente, la desviación que produce el accidente (suceso que interviene en la normal ejecución del trabajo) y los materiales con los que se produce la lesión.

#### ■ Lesión sufrida por el accidentado/a

Como primera división, se distinguen los accidentes según las lesiones que se provocan en ellos. Concretamente, el Ministerio de Trabajo considera la división en las siguientes categorías:

- 1 Heridas y lesiones superficiales
- 2 Fracturas de huesos
- 3 Dislocaciones, esguinces y torceduras
- 4 Amputaciones traumáticas (pérdida de partes del cuerpo)
- 5 Conmociones y lesiones internas
- 6 Quemaduras, escaldaduras y congelación
- 7 Envenenamientos e infecciones
- 8 Ahogamientos y asfixias
- 9 Efectos del ruido, la vibración y la presión
- 10 Efectos de las temperaturas extremas, la luz y la radiación
- 11 Trauma psíquico, choque traumático
- 12 Lesiones múltiples
- 13 Infartos, derrames cerebrales y otras patologías debidas a causas naturales
- 14 Otras lesiones especificadas no incluidas en anteriores apartado

En la Figura 3.30 se presenta un *treemap* en el que se muestra de forma proporcional la cantidad de accidentes que hay en función de la lesión a la que dieron lugar. Los datos de esta gráfica se corresponden con el año 2021 y con los accidentes que ocurrieron durante la jornada laboral. Los siete años anteriores muestran resultados similares.

Resulta claro que los dos tipos de lesiones más recurrentes son las heridas y lesiones superficiales y las dislocaciones, esguinces y torceduras. En ambos casos se corresponden con lesiones leves, lo cual coincide con el hecho de que la mayor parte de los accidentes laborales son de una gravedad leve.

En un segundo escalón se englobarían los accidentes que causan fracturas de huesos, conmociones y lesiones internas y envenenamientos e infecciones; situaciones de mayor gravedad, pero que se presentan en una medida bastante menor.

En las Figuras 3.31, 3.32 y 3.33 se puede visualizar la proporción del tipo de lesiones según la gravedad de los accidentes. La primera de ellas muestra los accidentes leves y sus resultados son muy similares a los totales, pues la mayoría de estos son leves. En la segunda gráfica se muestran los resultados para los accidentes graves y en la última, los resultados de los mortales.



Figura 3.30: *Treemap* en el que se muestra la proporción de accidentes durante la jornada laboral según la lesión que ocasionaron.

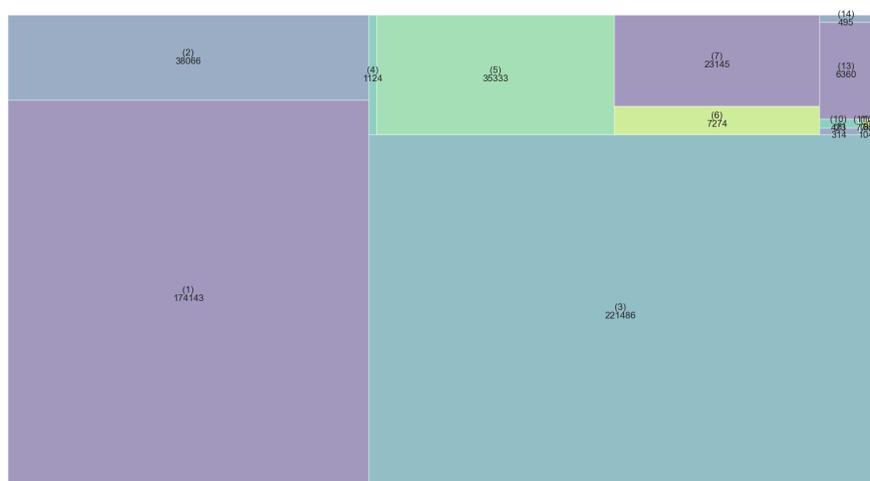


Figura 3.31: *Treemap* de proporción de lesiones de accidentes leves durante jornada.

Ya ha sido mencionado que los accidentes leves coinciden prácticamente con los totales. Por otro lado, los accidentes considerados como graves, presentan resultados notablemente diferentes. En este caso, las lesiones más destacables son aquellas que desembocaron en fracturas de huesos. Las menos destacadas fueron ahogamientos y asfixias; efectos del ruido, vibración y presión; efectos de temperaturas extremas, luz y radiación; y traumas psíquicos y lesiones múltiples. Estas también son las menos destacadas en cuanto a accidentes leves y en general. Finalmente, en cuanto a accidentes mortales, las lesiones que han provocado un mayor número de fallecidos han sido los infartos, derrames cerebrales y otras patologías debidas a causas naturales y otras lesiones no especificadas en el resto de apartados.

Los diferentes tipos de lesiones claramente dependen de la gravedad del accidente y, dado que la mayoría de accidentes laborales durante la jornada en España son leves, la mayor parte de las lesiones que tienen lugar se corresponden con torceduras, dislocaciones y esguinces y con heridas superficiales.

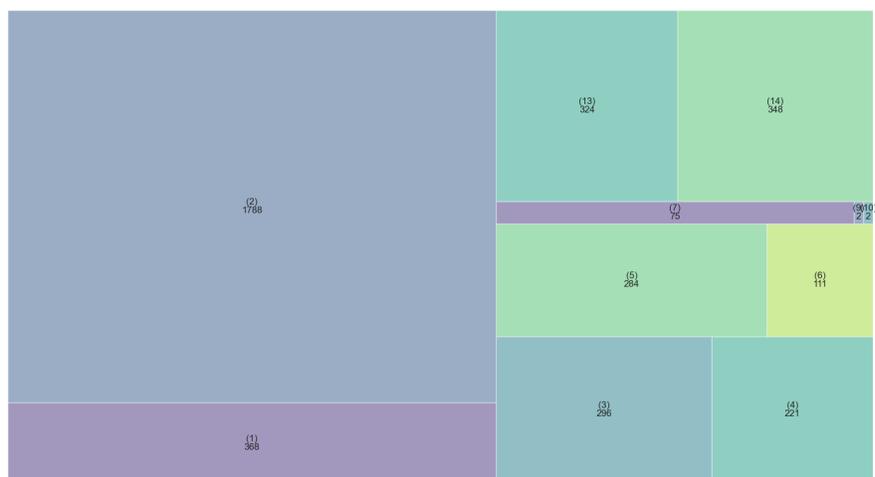


Figura 3.32: *Treemap* de proporción de lesiones de accidentes graves durante jornada.

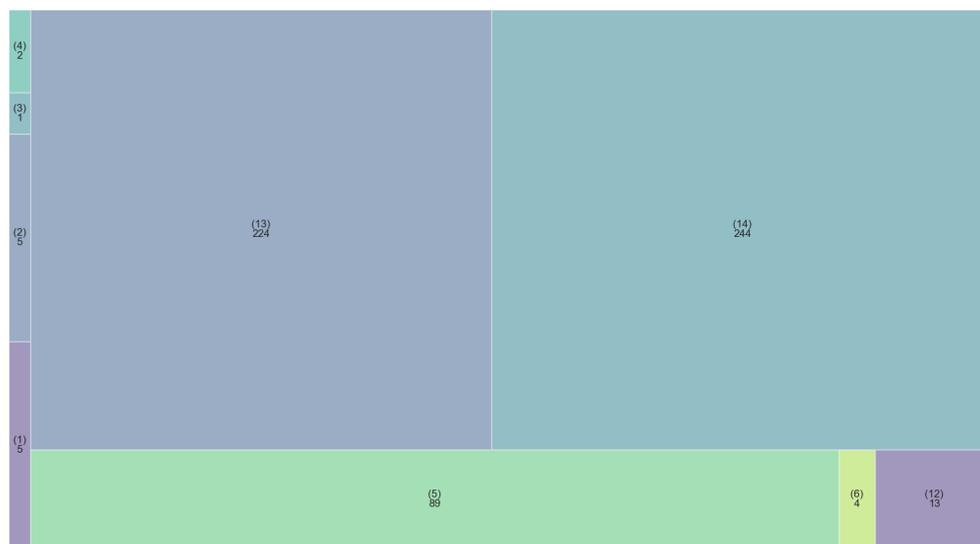


Figura 3.33: *Treemap* de proporción de lesiones de accidentes mortales durante jornada.

En la Figura 3.34 se pueden observar los resultados referentes a los accidentes que tienen lugar *in itinere*.

En este caso, las heridas superficiales y las torceduras también son las lesiones más recurrentes, seguidas de las fracturas de huesos, las conmociones y lesiones internas y los infartos y patologías por causas naturales. Unos resultados bastante similares a los vistos para los accidentes en jornada. La diferencia reside sobre todo en una mayor presencia de accidentes causados por infartos y derrames y una menor de quemaduras y congelación y envenenamientos e infecciones.

Aunque se hayan presentado los datos únicamente para 2021, los resultados que presentan los [años previos](#) son muy similares a los de 2021, por lo que los comentarios expuestos se pueden considerar a la situación general.

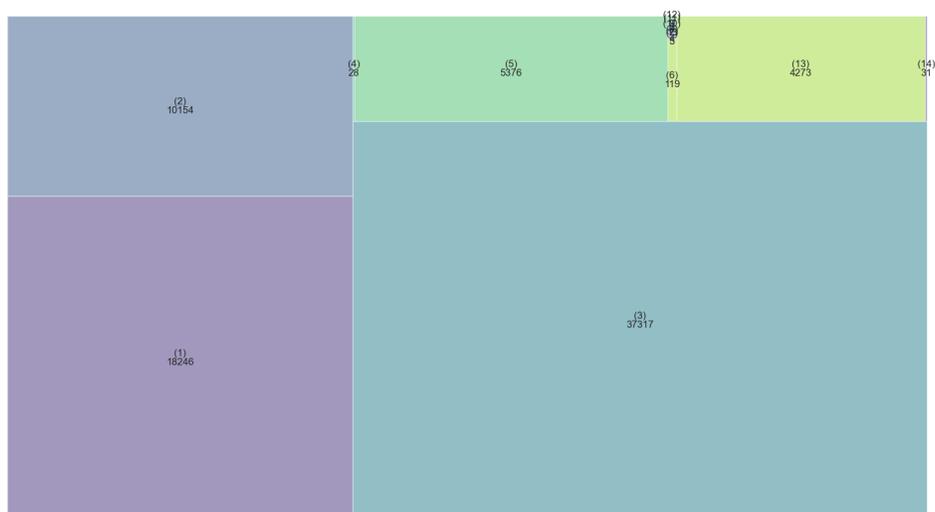


Figura 3.34: *Treemap* en el que se muestra la proporción de accidentes laborales *in itinere* según la lesión que ocasionaron.

#### ■ Parte del cuerpo dañada

La siguiente clasificación considerada es la división de los accidentes según la parte del cuerpo dañada. Se analizará el número de accidentes de los totales que afectaron a los siguientes miembros:

- 1 Cabeza, cerebro, nervios craneanos y vasos cerebrales
- 2 Zona facial
- 3 Ojo(s)
- 4 Oreja(s)
- 5 Dientes
- 6 Cabeza, múltiples partes afectadas
- 7 Cabeza - Sin especificar
- 8 Cuello, incluida la columna y las vértebras cervicales
- 9 Espalda, incluida la columna y las vértebras dorso-lumbares
- 10 Caja torácica, costillas, omóplatos y articulaciones acromioclaviculares
- 11 Región torácica, incluidos sus órganos
- 12 Región pélvica y abdominal, incluidos sus órganos
- 13 Tronco, múltiples partes afectadas
- 14 Tronco y órganos - Sin especificar
- 15 Hombro y articulaciones del húmero
- 16 Brazo, incluida la articulación del cúbito

- 17 Mano
- 18 Dedo(s)
- 19 Muñeca
- 20 Extremidades superiores, múltiples partes afectadas
- 21 Extremidades superiores - Sin especificar
- 22 Cadera y articulación de la cadera
- 23 Pierna, incluida la rodilla
- 24 Tobillo
- 25 Pie
- 26 Dedo(s) del pie
- 27 Extremidades inferiores, múltiples partes afectadas
- 28 Extremidades inferiores - Sin especificar
- 29 Todo el cuerpo y múltiples partes
- 30 Otras partes del cuerpo no incluidas en los anteriores apartados
- 31 Sin información

Nótese que al igual que en la clasificación anterior se incluye una categoría que incluye los accidentes en los que se dañaron múltiples partes del cuerpo (en el caso anterior: múltiples lesiones). Esto permite que la suma de todos los miembros del cuerpo dañados sea igual que el número de accidentes producidos.

De nuevo se presentan los resultados para el año más actual, 2021, pero se pueden observar los resultados de los [años anteriores](#).

En la Figura 3.35 se pueden visualizar los resultados en función de las partes del cuerpo dañadas en los accidentes que tuvieron lugar durante el periodo de jornada laboral.

Todas las partes del cuerpo consideradas han sido dañadas en algún accidente a lo largo del año 2021. No todas lo han sido en la misma medida, siendo las más destacadas la espalda (incluyendo columna y vértebras dorso-lumbares), las manos, los dedos, las piernas (incluyendo las rodillas), los tobillos y los pies. También constituyen una proporción interesante de los accidentes totales aquellos que involucran todo el cuerpo o múltiples partes del cuerpo dañadas. Los miembros que, en general, se ven menos afectados son la cabeza, cerebro, nervios craneanos y vasos cerebrales; las orejas; los dientes; la región torácica (incluyendo órganos); la región pélvica y abdominal (incluyendo órganos); el tronco; las extremidades superiores, en el caso de múltiples partes afectadas y otras partes del cuerpo no consideradas en las 29 categorías anteriores.

Por tanto, las extremidades son las partes del cuerpo generalmente más dañadas en los accidentes laborales. Tratando de conectarlo con la clasificación anterior, estos resultados parecen razonables, pues se había observado que las torceduras y esguinces eran unas de las lesiones más comunes en este tipo de accidentes.

En el caso de considerar solamente los accidentes de leve gravedad, los resultados son muy similares a los totales.

En el caso de los accidentes graves sí se observan algunas diferencias, como que la proporción de accidentes en múltiples partes y la proporción de accidentes en la cabeza, cerebro, nervios craneanos y vasos cerebrales son mayores que en el caso general. Los resultados son razonables, pues es de esperar

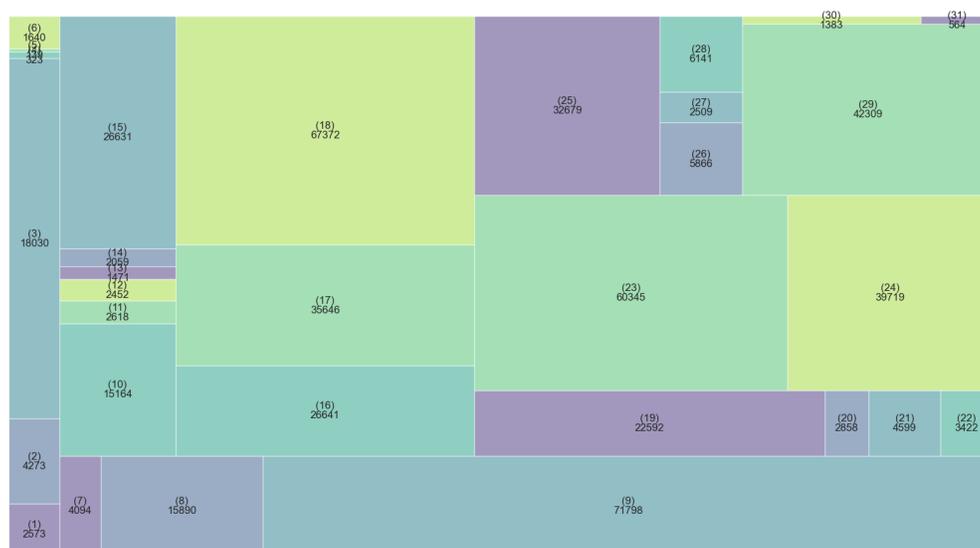


Figura 3.35: *Treemap* en el que se muestra la proporción de accidentes laborales durante la jornada según la parte del cuerpo lesionada en el accidente.

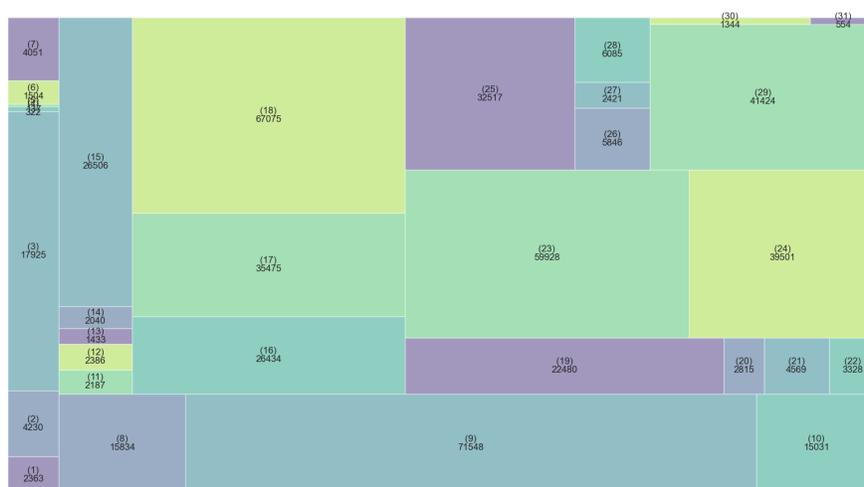


Figura 3.36: Proporción de accidentes laborales leves durante la jornada según la parte del cuerpo lesionada en el accidente.

que se consideren como accidentes más graves aquellos que afectan a varias partes del cuerpo o aquellos que afectan a las diferentes partes de la cabeza.

En el caso de los accidentes mortales, los resultados varían mucho más. En la Figura 3.38 se puede observar que muchas de las partes del cuerpo no se ven afectadas en este tipo de casos y otras lo hacen en una muy pequeña proporción. Sin embargo, dos categorías abarcan en torno al 75 % de los accidentes mortales: las lesiones en la región torácica (incluyendo órganos) y las lesiones en todo el cuerpo o múltiples partes. También destacan notablemente las lesiones en la cabeza.

A continuación, en las Figuras 3.39 y 3.40, se muestran los resultados en lo que se refiere a los accidentes *in itinere*. De nuevo, en este documento solo se incluyen las gráficas de 2021, pero también se dispone de los resultados de años anteriores, que presentan resultados casi idénticos.

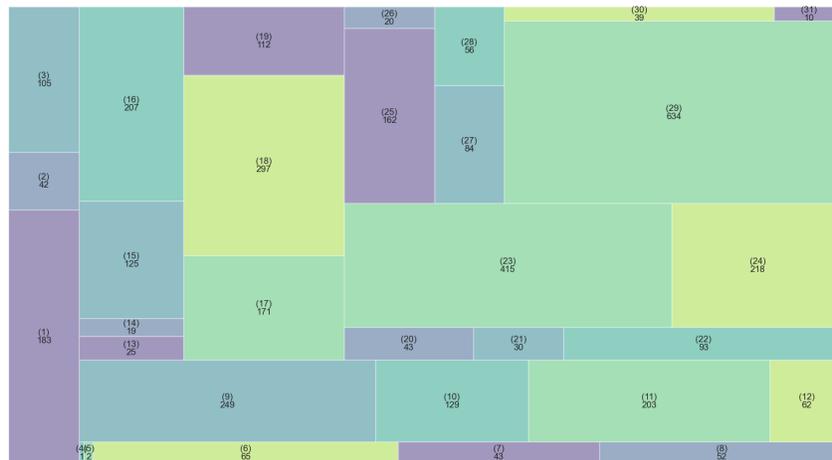


Figura 3.37: Proporción de accidentes laborales graves durante la jornada según la parte del cuerpo lesionada en el accidente.

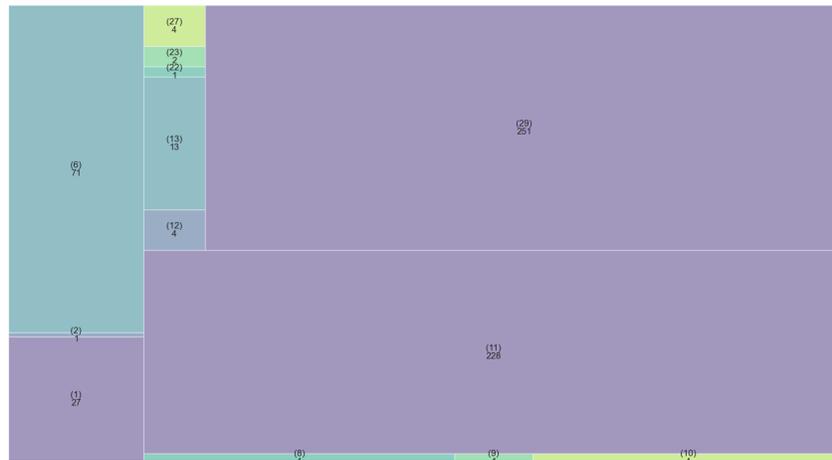


Figura 3.38: Proporción de accidentes laborales mortales durante la jornada según la parte del cuerpo lesionada en el accidente.

Los resultados no distan demasiado de los de los accidentes durante la jornada, pero sí que existen algunas diferencias. La proporción de accidentes en el tronco y órganos y en las extremidades superiores (brazos, manos y dedos) aumenta en comparación con los accidentes durante la jornada laboral. También se observa un aumento en la proporción de lesiones en la cadera y las extremidades inferiores (sin especificar). Por otro lado, disminuye la proporción de lesiones en los ojos, la caja torácica, costillas, omóplatos y articulaciones acromioclaviculares y en los pies.

En cuanto a accidentes mortales, la gran mayoría derivan de lesiones en todo el cuerpo o en múltiples partes. Algunos de ellos también son causados por lesiones en la cabeza (cerebro, nervios craneanos y vasos cerebrales o múltiples partes afectadas) y lesiones en la región torácica.

Al igual que en el caso de los accidentes durante la jornada laboral, los accidentes mortales son causados mayormente por lesiones en la cabeza o en la región del tórax. Relacionándolo con los resultados de las lesiones más comunes, se observa una conexión clara, pues gran parte de los accidentes mortales fueron causados por conmociones y lesiones internas, lesiones múltiples e infartos, derrames cerebrales y otras patologías debidas a causas naturales.

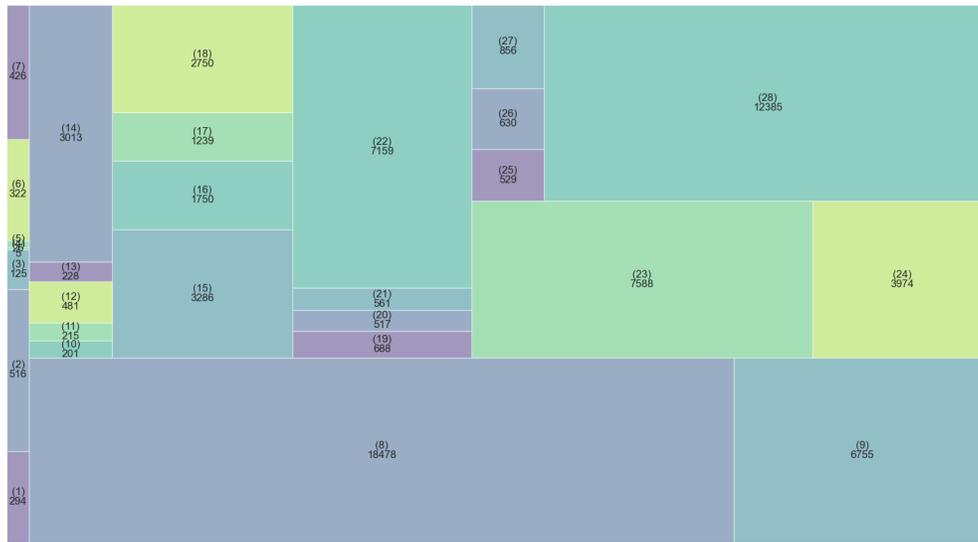


Figura 3.39: Proporción de accidentes laborales *in itinere* según la parte del cuerpo lesionada en el accidente.

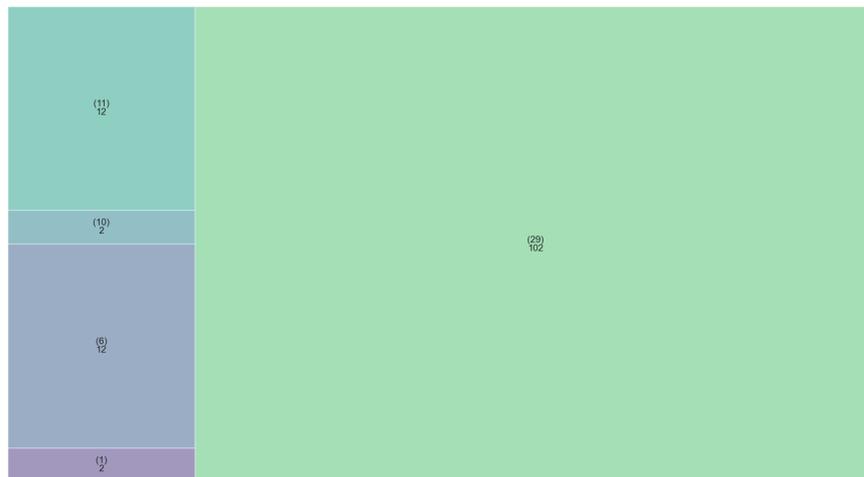


Figura 3.40: Proporción de accidentes laborales *in itinere* mortales según la parte del cuerpo lesionada en el accidente.

#### ■ Tipo de lugar donde ocurrió el accidente

Otra de las categorías que es interesante explorar, con la intención de tener un mayor conocimiento de los accidentes, es el lugar donde ocurrieron estos. Las categorías consideradas por el Ministerio de Trabajo son:

- 1 Zonas industriales
- 2 Obras, construcción, cantera, mina a cielo abierto
- 3 Lugares agrícolas, ganaderos, forestales, de piscicultura

- 4 Lugares del sector servicios, oficinas, zonas de ocio, etc
- 5 Centros sanitarios
- 6 Lugares públicos, medios de transporte
- 7 Domicilios
- 8 Lugares de actividades deportivas
- 9 En el aire, elevados - con excepción de las obras
- 10 Subterráneos - con excepción de las obras
- 11 En el agua, a bordo de todo tipo de navíos, excepto obras
- 12 En medio hiperbárico, bajo el agua – excepto obras
- 13 Otros tipos de lugar no incluidos en los anteriores apartados
- 14 Sin información

Tal y como se puede ver en la Figura 3.41, en torno a un 40 % de los accidentes que ocurren durante la jornada laboral tienen lugar en zonas industriales. Los lugares con un mayor volumen de accidentes laborales que siguen a este son, por orden, los lugares del sector servicios, los lugares públicos y medios de transporte, las obras, los centros sanitarios y los lugares agrícolas. Con menor frecuencia tienen lugar accidentes en domicilios y todavía es menor el número de accidentes laborales en lugares fuera de las obras en las que haya que estar elevado, bajo suelo, en el agua o en un medio hiperbárico, o bajo el agua.

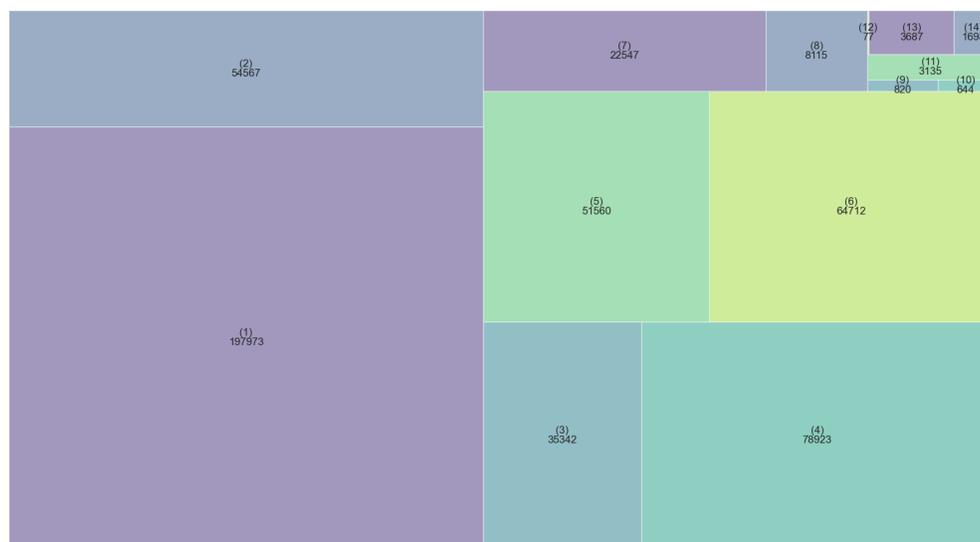


Figura 3.41: *Treemap* en el que se muestra la proporción de accidentes laborales durante la jornada según el lugar donde ocurrió en el accidente.

Como en el resto de categorías, los resultados entre los años 2014 y 2020 también están [disponibles](#). Los resultados a medida que avanzan los años no difieren excesivamente, pero sí es cierto que se percibe un descenso en la proporción de accidentes de lugares del sector servicios, oficinas, zonas de ocio, etc. De todas formas, en general los resultados son bastante similares a lo largo de los 8 años.

A continuación, en las tres gráficas siguientes, se presentan los resultados diferenciando la gravedad de los accidentes. A pesar de que se podría esperar una gran diferencia entre los accidentes graves y los mortales, como ocurre en algunas de las categorías ya vistas, en este caso no hay diferencias demasiado notables. Aun así, sí se aprecia alguna disparidad, como que a medida que aumenta la gravedad de los accidentes, se aprecia una mayor proporción de accidentes en lugares y transportes públicos y una menor proporción de accidentes en zonas industriales y obras, en lugares del sector servicios y en centros sanitarios.

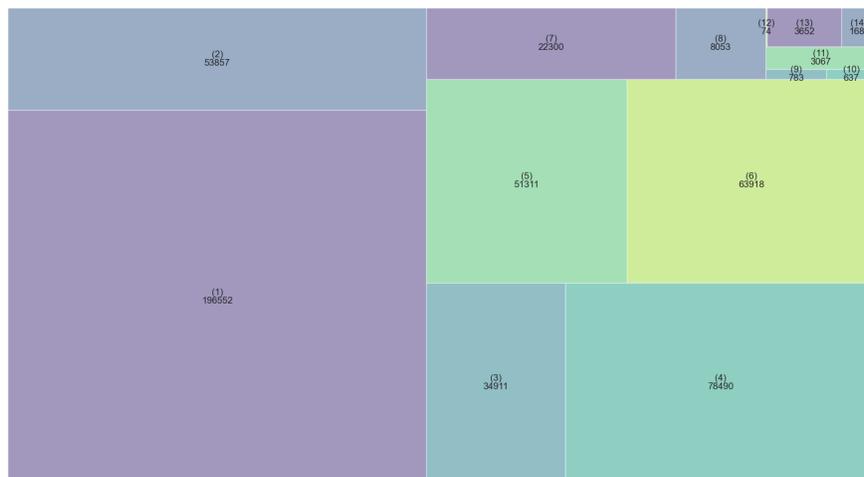


Figura 3.42: Proporción de accidentes laborales leves durante la jornada según el lugar de trabajo.

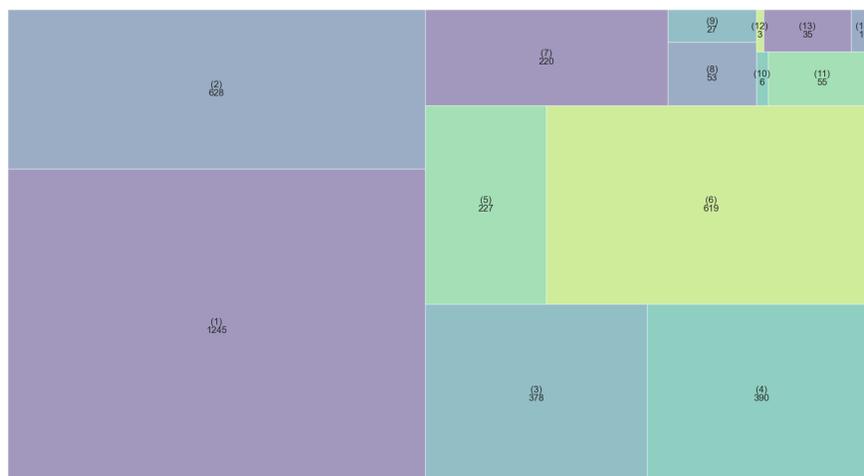


Figura 3.43: Proporción de accidentes laborales leves durante la jornada según el lugar de trabajo.

Los resultados que se muestran en cuanto a esta temática son en todo momento considerando solamente los accidentes durante la jornada laboral. Se han omitido los accidentes *in itinere*, pues la gran mayoría de los accidentes en este caso se engloban dentro de la categoría “Otros tipos de lugar no incluidos en los anteriores apartados”. Además, para las categorías consideradas, los resultados son muy parecidos para los 8 años considerados y no presentan información de especial interés para el caso *in itinere*.

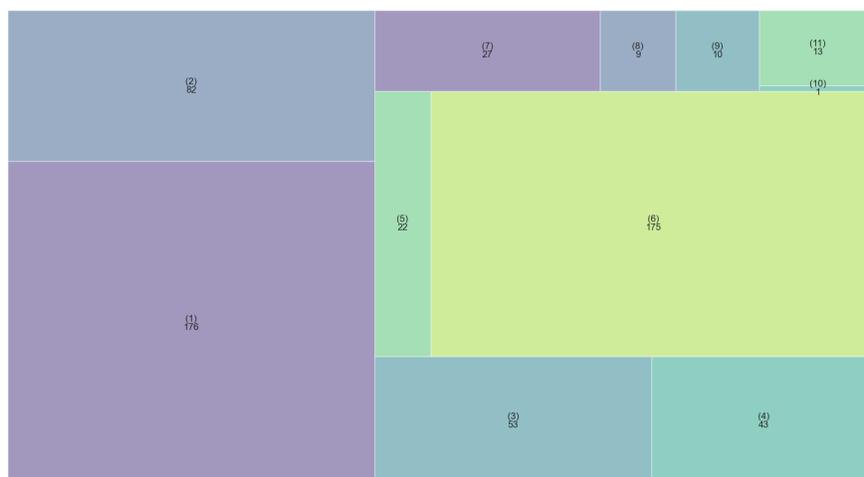


Figura 3.44: Proporción de accidentes laborales mortales durante la jornada según el lugar de trabajo.

#### ■ Trabajo que realizaba el accidentado/a

Otro aspecto interesante a tener en cuenta de cara a analizar los accidentes en el ámbito laboral es considerar el trabajo que estaba realizando el trabajador. Las diferentes categorías consideradas por el Ministerio son:

- 1 Tareas de producción, transformación, almacenamiento
- 2 Movimiento de tierras, construcción, demolición
- 3 Labores agrícolas, forestales, ganaderas, piscícolas
- 4 Servicios a empresas o a personas y trabajos intelectuales
- 5 Tareas de instalación, mantenimiento, limpieza, gestión de residuos, vigilancia
- 6 Circulación, actividades deportivas y artísticas
- 7 Otros tipos de trabajo no incluidos en los anteriores apartados
- 8 Sin información

En este caso tampoco se considerarán los accidentes *in itinere*, pues se entiende que no se está realizando ninguna acción más que desplazarse hacia el trabajo o desde él. Por otro lado, solo se incluyen en el documento las gráficas referentes a 2021, pero la información para el [resto de años](#) también se proporciona.

Como se puede ver en la Figura 3.45, las tareas de producción, transformación y almacenamiento son las que presentan una mayor proporción de accidentes respecto del total, seguidas de las tareas de instalación, mantenimiento, limpieza, gestión de residuos y vigilancia; y los servicios a empresas o a personas y trabajos intelectuales.

Las actividades que muestran menores proporciones son el movimiento de tierras, construcción, demolición y las labores agrícolas, forestales, ganaderas y piscícolas.

A continuación se muestran los resultados diferenciando los accidentes según su gravedad. Los resultados de los accidentes leves son prácticamente coincidentes con los resultados totales. En cuanto a aquellos con una mayor gravedad, se observa una disminución de la proporción de accidentes en tareas de producción, transformación y almacenamiento y en servicios a empresas o a personas y

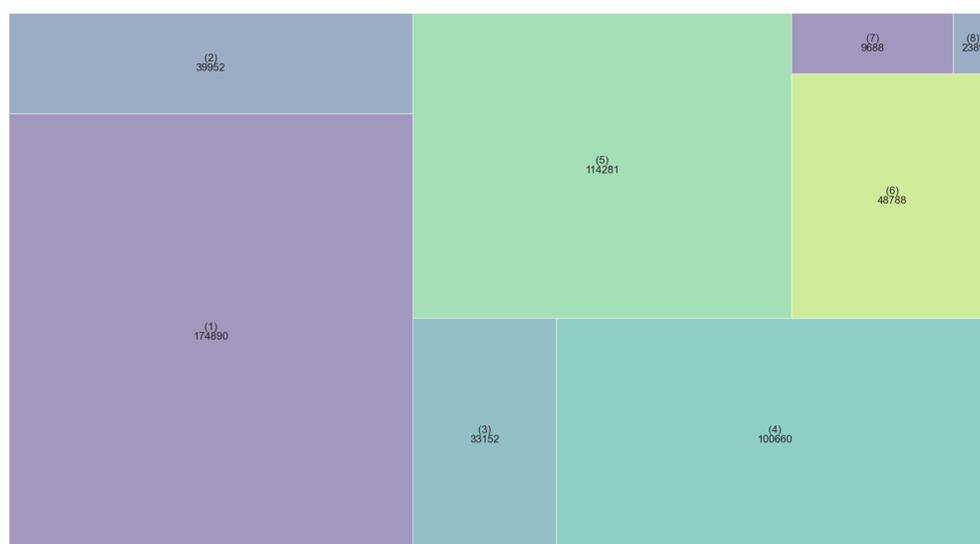


Figura 3.45: *Treemap* en el que se muestra la proporción de accidentes laborales durante la jornada según la actividad que estaba realizando el trabajador.

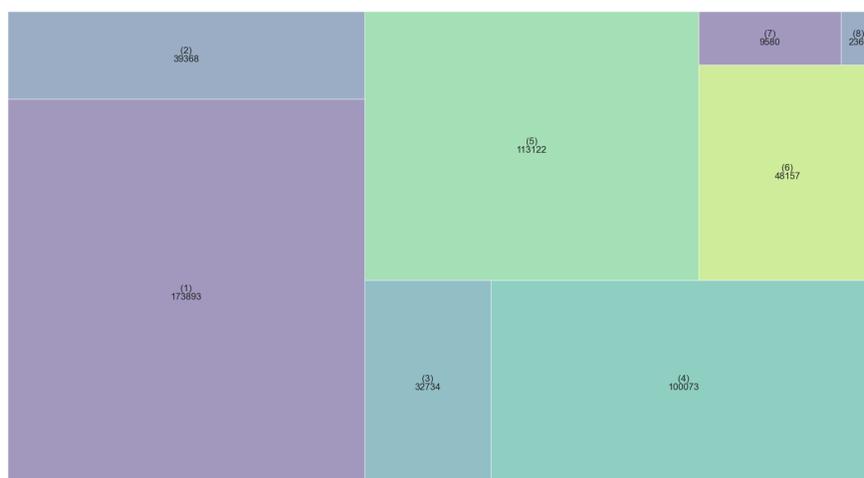


Figura 3.46: Proporción de accidentes laborales leves durante la jornada según la actividad que estaba realizando el trabajador.

trabajos intelectuales; sin embargo, la proporción de accidentes por movimiento de tierras, construcción y demolición y por labores agrícolas, forestales, ganaderas y piscícolas aumentan.

En cuanto a los accidentes mortales, las proporciones de las diferentes actividades se igualan más. Aquellas con una mayor proporción, la disminuyen y las que tenían una menor, la aumentan. De todas formas, hay dos que destacan por encima de las demás: tareas de instalación, mantenimiento, limpieza, gestión de residuos y vigilancia; y también circulación, actividades deportivas y artísticas.

Como resumen general, para todas las actividades consideradas ocurren accidentes de diferentes niveles de gravedad, con mayor o menor proporción. De todas formas, para cada nivel de gravedad existen algunas actividades que destacan por proporción de accidentes sobre las demás.

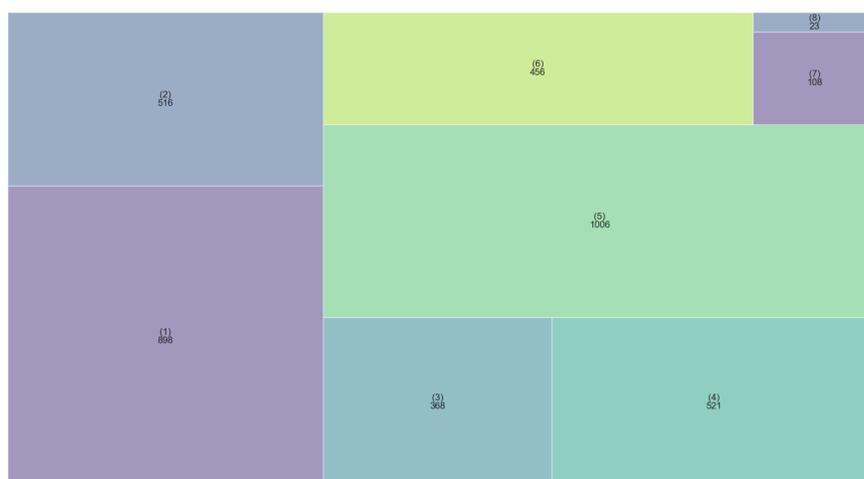


Figura 3.47: Proporción de accidentes laborales graves durante la jornada según la actividad que estaba realizando el trabajador.

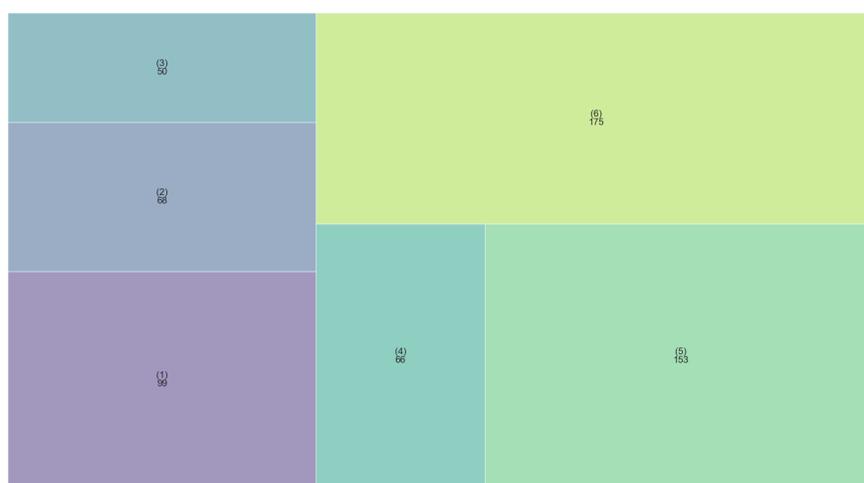


Figura 3.48: Proporción de accidentes laborales mortales durante la jornada según la actividad que estaba realizando el trabajador.

#### ■ Actividad física que realizaba el accidentado/a

La siguiente categoría que se analiza es el trabajo o actividad física concreta que realizaba el trabajador cuando sufrió el accidente que causó su baja. Las diferentes actividades físicas recogidas en los archivos del Ministerio de Trabajo son las siguientes:

- 1 Operaciones con máquinas
- 2 Trabajos con herramientas manuales
- 3 Conducir / estar a bordo de un medio de transporte o equipo de carga
- 4 Manipulación de objetos
- 5 Transporte manual

- 6 Movimiento
- 7 Estar presente
- 8 Otra actividad física específica no incluida en los anteriores apartados.

Dentro de cada una de estas categorías existen subcategorías, pero estas son bastantes y muy específicas, por lo que se ha limitado el estudio a las categorías superiores. Por mostrar un ejemplo, para la categoría “movimiento” se consideran las subcategorías: 30 Andar, correr, subir, bajar, etc.; 31 Entrar, salir; 32 Saltar, abalanzarse, etc.; 33 Arrastrarse, trepar, etc.; 34 Levantarse, sentarse, etc.; 35 Nadar, sumergirse; 36 Hacer movimientos en un mismo sitio y 37 Movimiento - Sin especificar. Aunque puede llegar a ser interesante estudiar en profundidad estas categorías, sería un contenido demasiado específico teniendo en cuenta la cantidad de clasificaciones consideradas en este trabajo.

En la Figura 3.49 se representan las proporciones en cuanto al número de accidentes laborales durante la jornada según la clase de actividad física que estaba realizando el trabajador o trabajadora. Los datos de esta gráfica se corresponden con el año 2021; la información para el [resto de años](#) está también disponible.

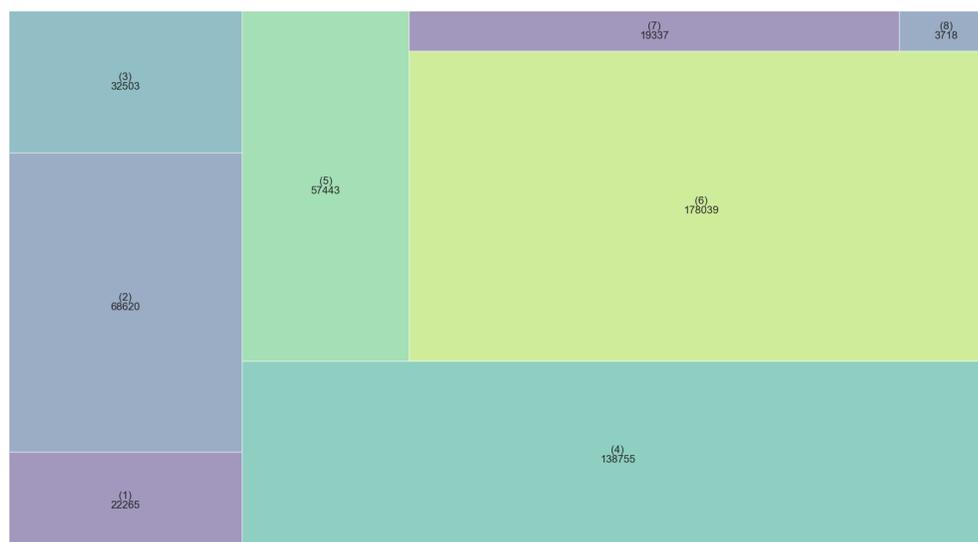


Figura 3.49: *Treemap* en el que se representa la cantidad de accidentes laborales durante la jornada según la actividad física que estaba realizando el trabajador.

Estar en movimiento o manipulando objetos son las dos actividades que presentan un mayor número de accidentes durante la jornada laboral. Los trabajos con herramientas manuales son también una actividad que provoca un alto número de accidentes, en comparación con el resto de categorías consideradas. Estar presente y las operaciones con máquinas son las actividades que presentan un menor volumen de accidentados; en el caso de la primera quizás se deba a que a priori conlleva un menor riesgo, pero en cuanto a la manipulación de máquinas seguramente se deba a que la cantidad de población ocupada que se dedica a trabajar con máquinas es menor que la que puede estar presente (que se puede entender que es toda).

Los resultados generales para años anteriores son muy similares a los de la Figura 3.49.

En las Figuras 3.50, 3.51 y 3.52 se pueden visualizar los resultados diferenciando la gravedad de los accidentes que tuvieron lugar durante la jornada.

Como en todos los casos vistos, la diferencia entre los accidentes leves y los totales es prácticamente nula, pues la gran mayoría de los accidentes se engloban bajo la categoría de leves.



Figura 3.50: Proporción de accidentes laborales leves durante la jornada según la actividad física que desempeñaba el trabajador.

Al igual que en la gráfica anterior, se puede ver que la mayoría de los accidentes leves ocurren mientras el trabajador está manipulando objetos o simplemente mientras está presente.

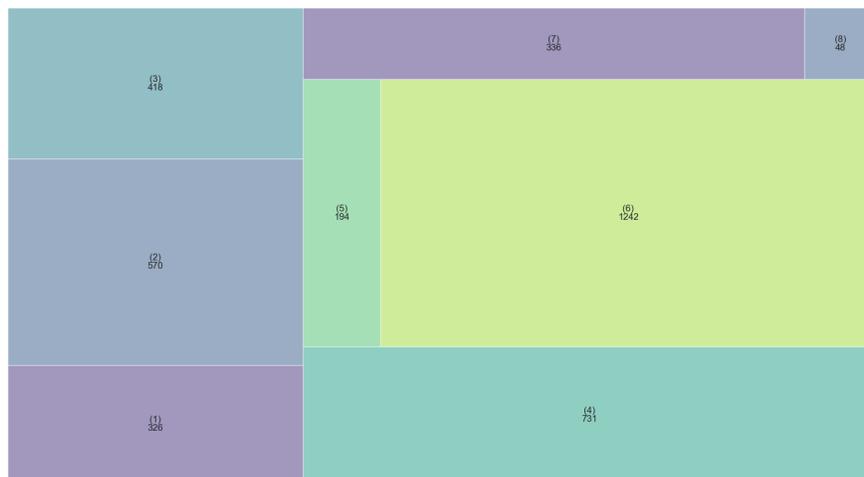


Figura 3.51: Proporción de accidentes laborales graves durante la jornada según la actividad física que desempeñaba el trabajador.

Al restringirnos a los accidentes graves, la proporción de accidentes de trabajadores que se encuentran manipulando objetos desciende. Por otro lado, se observa un aumento en la proporción de accidentes en aquellos en los que realizan operaciones con máquinas, trabajos con herramientas manuales y conducen o están a bordo de un medio de transporte o equipo de carga.

En cuanto a los accidentes mortales, el número de accidentes es menor, lógicamente, y además las proporciones varían. Los casos en los que el trabajador se encontraba en movimiento son los que presentan un mayor número de fallecimientos. Tras ellos se encuentran los accidentes en los que el trabajador estaba conduciendo o a bordo de un medio de transporte o equipo de carga.

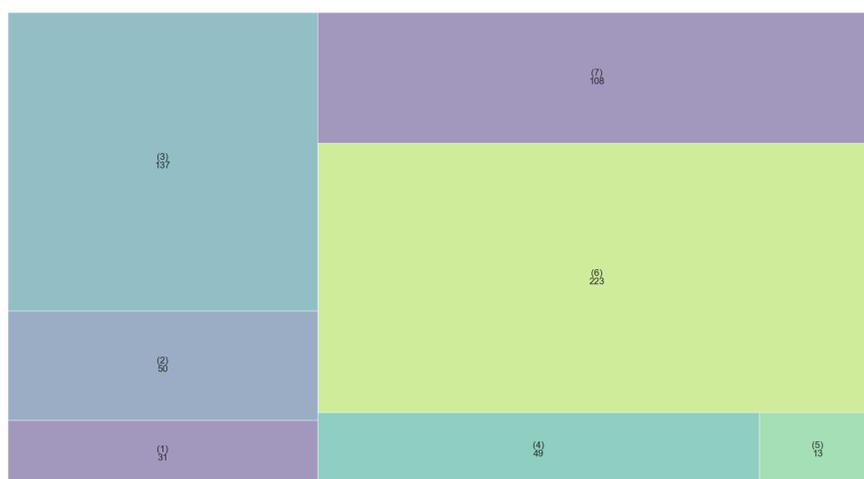


Figura 3.52: Proporción de accidentes laborales mortales durante la jornada según la actividad física que desempeñaba el trabajador.

En comparación con los accidentes de menos gravedad (leves o graves), destaca que la proporción de accidentes por estar simplemente presente es mayor. Esto seguramente se debe a que una gran proporción de accidentes mortales eran provocados por infartos, derrames, etc. y esto es frecuente que ocurra sin estar realizando ninguna actividad (simplemente estando presente).

Para finalizar con esta clasificación, se revisarán también los accidentes ocurridos *in itinere*. En la Figura 3.53 se muestra la proporción de accidentes según la actividad realizada por el trabajador de entre aquellos accidentes que tuvieron lugar fuera de la jornada laboral.

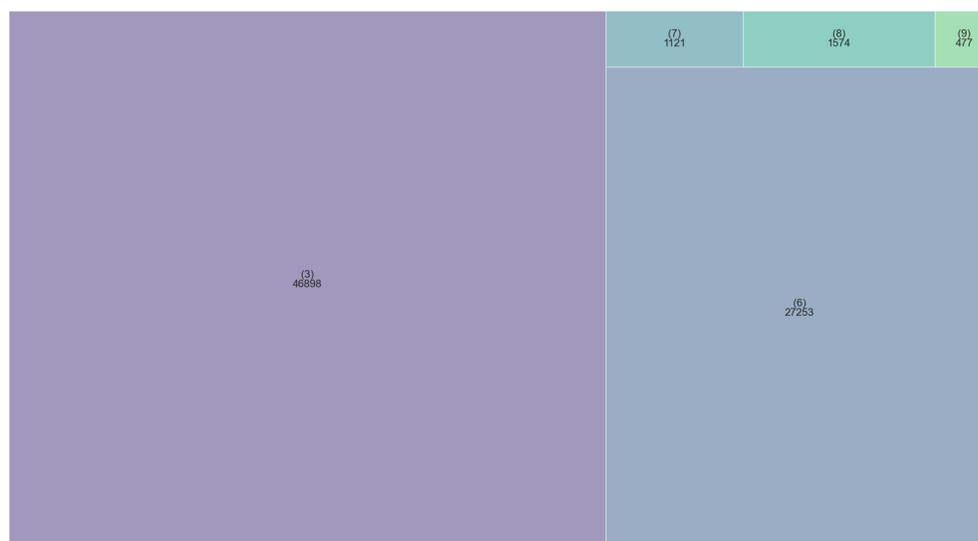


Figura 3.53: *Treemap* en el que se muestra la proporción de accidentes laborales *in itinere* según la actividad física que estaba realizando el trabajador.

En este caso, la mayoría de los accidentes ocurren cuando el trabajador se encuentra conduciendo o moviéndose; lo cual entra totalmente dentro de lo esperable, pues si está *in itinere* se entiende que

está desplazándose, ya sea en un medio de transporte o caminando.

Los accidentes que no se engloban en ninguna de estas dos categorías corresponden a trabajadores que se encontraban presentes, realizando una actividad física no incluida en la clasificación o no se recogió información de ellos (cuadro indicado con el número 9).

De nuevo, esta gráfica representa los resultados del año 2021. De todas formas, la situación en años anteriores es muy similar a la aquí presentada.

#### ■ Desviaciones que producen el accidente

En la siguiente categoría a analizar, se consideran las diferentes desviaciones que sufrieron los trabajadores y provocaron el accidente:

- 1 Desviación por problema eléctrico, explosión, fuego
- 2 Desviación por desbordamiento, vuelco, escape, derrame, emanación
- 3 Rotura, estallido, deslizamiento, caída, derrumbamiento de agente material
- 4 Pérdida de control total o parcial de equipos de trabajo o materiales
- 5 Caída de personas – Resbalón o tropezón con caída
- 6 Movimiento del cuerpo sin esfuerzo físico añadido
- 7 Movimiento del cuerpo como consecuencia de o con esfuerzo físico
- 8 Sorpresa, miedo, violencia, agresión, amenaza, presencia
- 9 Infartos, derrames cerebrales y otras causas estrictamente naturales

En la Figura 3.54 se puede ver la información relativa al año más reciente y los datos de los años anteriores hasta 2014 están disponibles *online*.



Figura 3.54: *Treemap* en el que se muestra la proporción de accidentes laborales durante la jornada según la desviación que provocó el accidente.

Además, se incluyen en las Figuras 3.55, 3.56 y 3.57 los resultados desglosados según la gravedad de los accidentes, también para el año 2021.

Las 4 categorías más recurrentes fueron:

- Pérdida de control total o parcial de equipos de trabajo o materiales
- Caída de personas – Resbalón o tropezón con caída
- Movimiento del cuerpo sin esfuerzo físico añadido
- Movimiento del cuerpo como consecuencia de o con esfuerzo físico,

siendo esta última la más destacada. Teniendo en cuenta estos tres últimos motivos, parece que muchos de los accidentes laborales fueron causados por desviaciones relacionadas con el movimiento de las personas.

En cuanto a los accidentes leves, la información es muy similar a la general.



Figura 3.55: Proporción de accidentes laborales leves durante la jornada según la desviación que provocó el accidente.



Figura 3.56: Proporción de accidentes laborales graves durante la jornada según la desviación que provocó el accidente.

En cuanto aumenta la gravedad de los accidentes (tanto graves como mortales), las desviaciones como infartos, derrames cerebrales y otras causas estrictamente naturales se presentan en una mayor

proporción. Sobre todo en el caso de los accidentes mortales se observa una mayor proporción. En el caso de los accidentes graves, la desviación que mayor número de accidentes provoca es la caída de personas (resbalón o tropezón con caída).

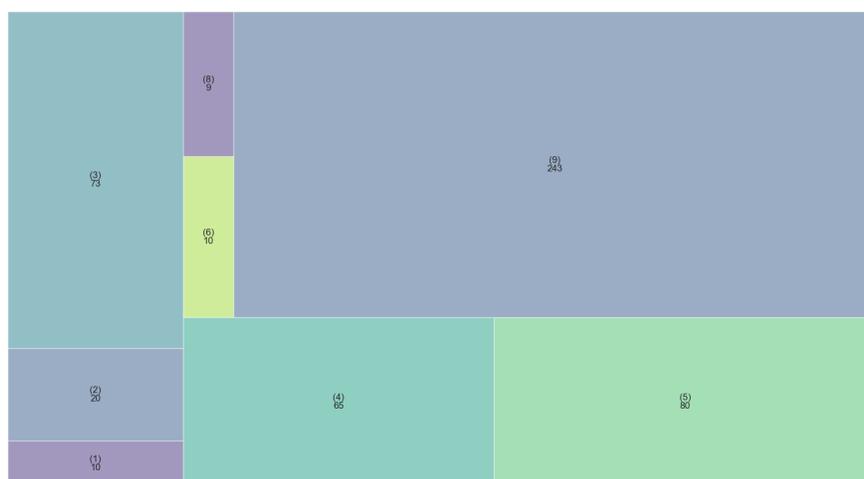


Figura 3.57: Proporción de accidentes laborales mortales durante la jornada según la desviación que provocó el accidente.

Se muestran también los resultados para el año 2021 de los accidentes *in itinere* (véase Figura 3.58).



Figura 3.58: Proporción de accidentes laborales *in itinere* según la desviación que provocó el accidente.

En estas circunstancias, la mayoría de los accidentes se debieron a movimiento del cuerpo sin esfuerzo físico añadido. Tras esta desviación, las tres restantes que fueron más recurrentes durante la jornada también lo son en el caso de los accidentes *in itinere*.

### ■ Agente material de la desviación que produce el accidente

La siguiente clasificación considerada divide los accidentes laborales según los agentes materiales de las desviaciones que produjeron el accidente. Por tanto, esta clasificación está totalmente relacionada con la anterior, a pesar de que en las tablas de datos proporcionadas por el Ministerio no se presente ninguna que contenga la intersección de ambas clasificaciones.

En las siguientes líneas se indican los diferentes grupos en los que se clasifican los accidentes según el agente material de la desviación:

- 1 Edificios, construcciones, superficies a nivel
- 2 Edificios, construcciones, superficies en altura
- 3 Edificios, construcciones, superficies en profundidad
- 4 Dispositivos distribución de materia, canalizaciones
- 5 Motores, dispositivos de transmisión, almacenamiento de energía
- 6 Herramientas manuales - sin motor
- 7 Herramientas sostenidas o guiadas con las manos-mecánicas
- 8 Herramientas manuales sin especificar en cuanto a motorización
- 9 Máquinas y equipos - portátiles o móviles
- 10 Máquinas y equipos - fijos
- 11 Dispositivos de traslado, transporte y almacenamiento
- 12 Vehículos terrestres
- 13 Otros vehículos de transporte
- 14 Materiales, objetos, productos, elementos de máquina, polvo, etc.
- 15 Sustancias químicas, explosivas, radioactivas, biológicas
- 16 Dispositivos y equipos de seguridad
- 17 Equipos oficina y personales, de deporte, armas, aparatos domésticos
- 18 Organismos vivos y seres humanos
- 19 Residuos en grandes cantidades
- 20 Fenómenos físicos y elementos naturales
- 21 Otros agentes materiales no incluidos en los anteriores apartados
- 22 No existe agente material

En la Figura 3.59 se presenta la proporción de cada una de estas clases en cuanto a accidentes laborales que han tenido lugar durante la jornada laboral en 2021. En función de esta representación, se puede apreciar que todas las categorías están presentes (en mayor o menor manera), pero las más destacadas son edificios, construcciones, superficies a nivel y materiales, objetos, productos, elementos de máquina, polvo, etc.

En la Figura 3.60 se presenta esta misma información, pero restringida a los accidentes que acabaron con el fallecimiento del trabajador. En este caso, las desviaciones producidas por vehículos terrestres y

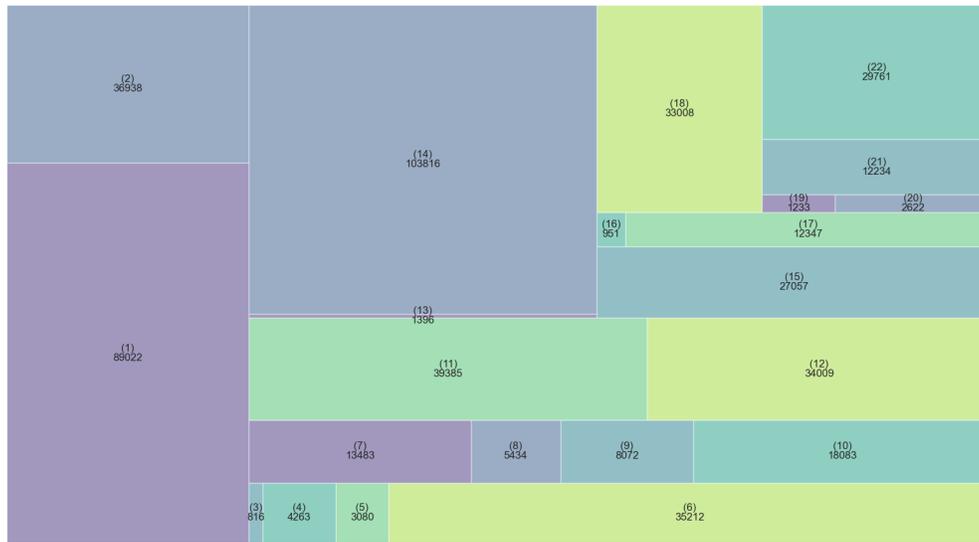


Figura 3.59: *Treemap* en el que se muestra la proporción de accidentes laborales durante la jornada según el agente material de la desviación que provocó el accidente.

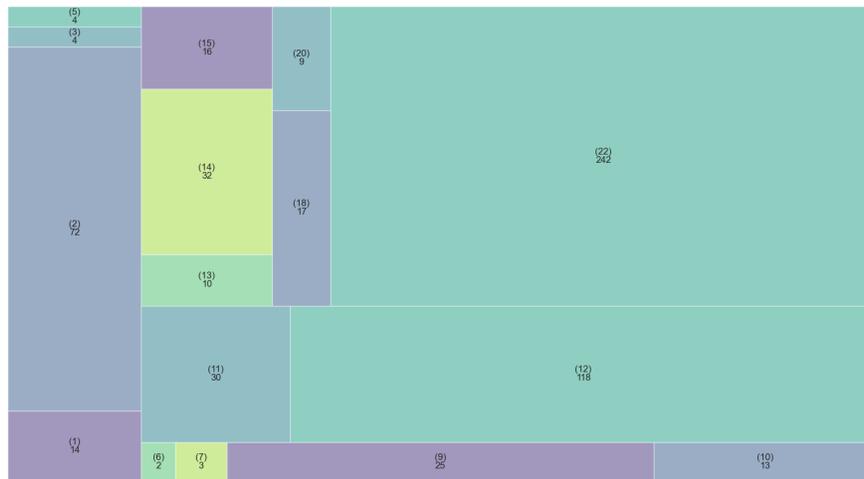


Figura 3.60: Proporción de accidentes laborales mortales durante la jornada según el agente material de la desviación que provocó el accidente.

edificios, construcciones, superficies en altura constituyen una parte importante de los agentes materiales causantes de los accidentes mortales. De todas formas, la clase que presenta una mayor proporción es: no existe agente material. Es un resultado esperable, dado que previamente se ha visto que los accidentes mortales más recurrentes se correspondían con infartos, derrames y otras causas naturales.

Tanto la Figura 3.59 como la 3.60 presentan información del año 2021, que no dista mucho del resultado de años anteriores.

A continuación, se muestran dos gráficas análogas en las que se presenta la misma información pero en referencia a los accidentes *in itinere* en el año 2021. Los resultados correspondientes a los años entre 2014 y 2021 también están disponibles.

El material que más accidentes desencadena son los edificios, construcciones y superficies a nivel

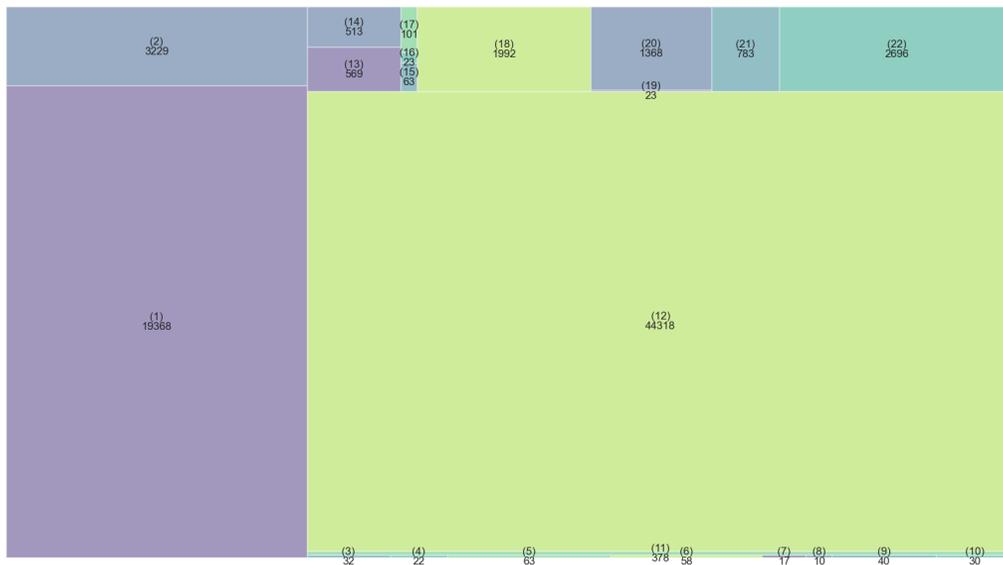


Figura 3.61: *Treemap* en el que se muestra la proporción de accidentes laborales *in itinere* según el agente material de la desviación que provocó el accidente.

y, en especial, los vehículos terrestres. Este último resultado era de esperar al estar considerando los accidentes que tienen lugar a la hora de desplazarse hacia o desde el lugar de trabajo.

Al restringirse a los accidentes mortales, la proporción de ellos debidos a vehículos terrestres aumenta mucho, siendo prácticamente el total. La segunda desviación que genera más accidentes consiste en aquella que ocurre por ningún agente material. Además de estas dos, en 2021 también hubo un accidente que se debe a cada uno de los siguientes agentes materiales: edificios, construcciones, superficies a nivel; edificios, construcciones, superficies en altura y máquinas y equipos (portátiles o móviles).

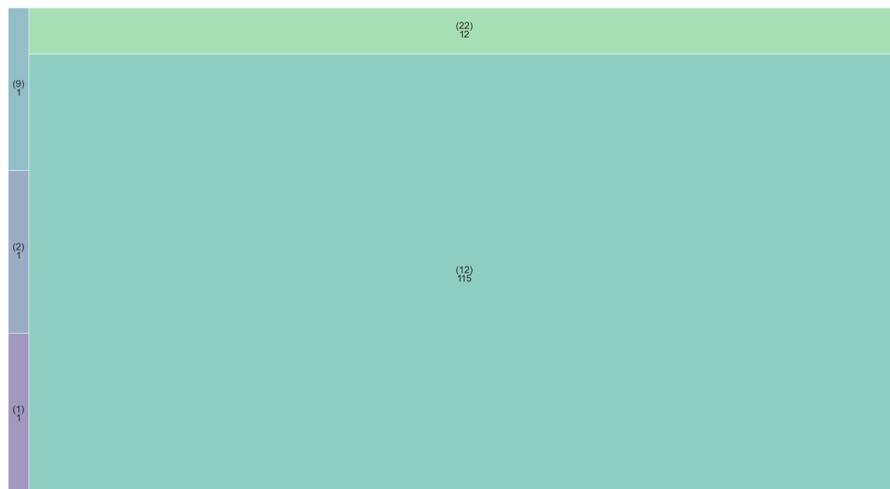


Figura 3.62: Proporción de accidentes laborales mortales *in itinere* según el agente material de la desviación que provocó el accidente.

### ■ Contacto o forma en la que se produjo la lesión

Para finalizar el análisis de accidentes laborales según las características de los propios accidentes, se revisa la distribución de los accidentes según el contacto o la forma en la que se produjo la lesión. De nuevo se presentan los datos de 2021, pero la relativa a [años anteriores](#) se puede consultar.

Las categorías consideradas en esta clasificación son:

- 1 Contacto con corriente eléctrica, fuego, temperatura, sustancias peligrosas
- 2 Ahogamiento, quedar sepultado, quedar envuelto
- 3 Choque o golpe contra objeto inmóvil (trabajador en movimiento)
- 4 Choque o golpe contra un objeto en movimiento, colisión
- 5 Contacto con agente material cortante, punzante, duro
- 6 Quedar atrapado, ser aplastado, sufrir una amputación
- 7 Sobreesfuerzo físico, trauma psíquico, radiaciones, ruido, luz presión
- 8 Mordeduras, patadas, etc. (de animales o personas)
- 9 Infartos, derrames cerebrales y otras causas estrictamente naturales
- 10 Accidentes de tráfico (1)
- 11 Otra forma o contacto no incluido en los anteriores apartados

En cuanto a los accidentes durante la jornada, se observa en la Figura 3.63 que la categoría que ha dado lugar a un mayor número de accidentes ha sido sobreesfuerzo físico, trauma psíquico, radiaciones, ruido, luz, presión. Tras ella, otras situaciones como choque o golpe contra objeto inmóvil (trabajador en movimiento), choque o golpe contra un objeto en movimiento y contacto con agente material cortante, punzante o duro son bastante recurrentes. Los infartos, derrames cerebrales y otras causas estrictamente naturales son muy poco recurrentes en el caso general.

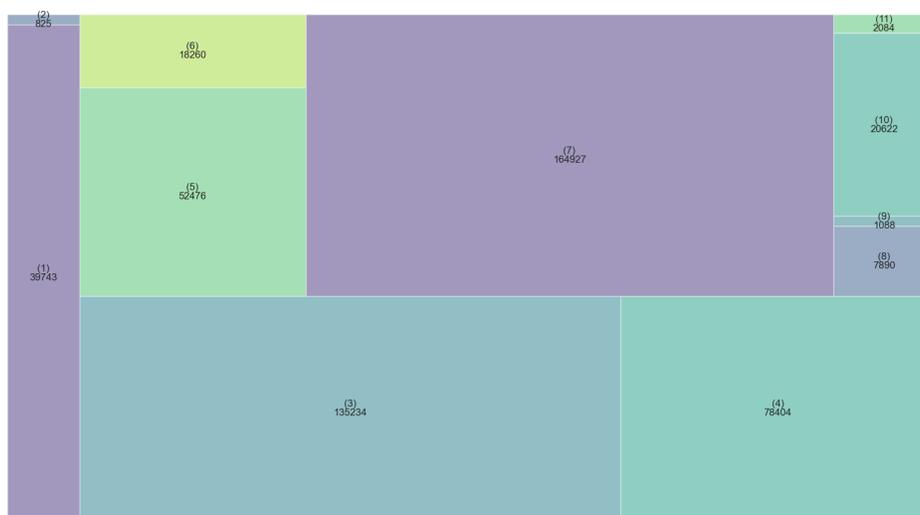


Figura 3.63: *Treemap* en el que se muestra la proporción de accidentes laborales durante la jornada según la forma o contacto que produjo la lesión que provocó el accidente.

Sin embargo, como se puede ver en la Figura 3.64, en el caso de los accidentes mortales es la que está más presente. Tras ella, choque o golpe contra objeto inmóvil, quedar atrapado, ser aplastado, sufrir una amputación y accidentes de tráfico dan también lugar a una gran parte de los accidentes mortales.

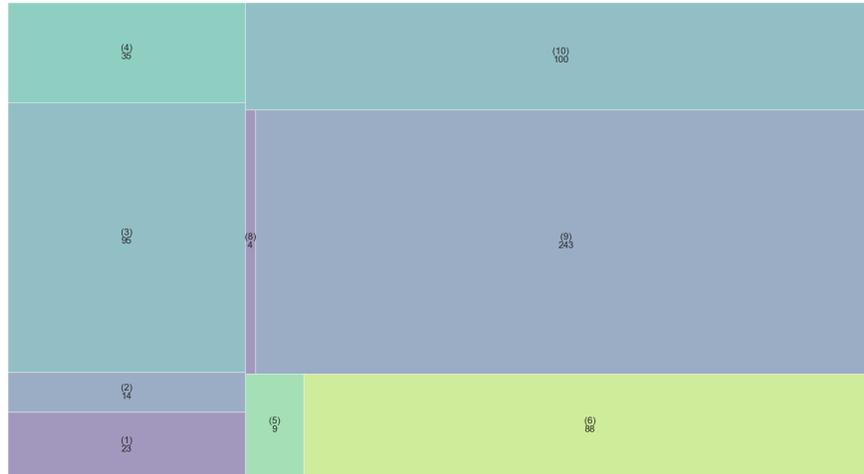


Figura 3.64: Proporción de accidentes laborales mortales durante la jornada según la forma o contacto que produjo la lesión que provocó el accidente.

Finalmente, en el caso de los accidentes *in itinere*, se puede observar en la Figura 3.65 que la mayoría son causados, como cabía esperar, en accidentes de tráfico. En un segundo plano, los accidentes por choques con objetos inmóviles también destacan.

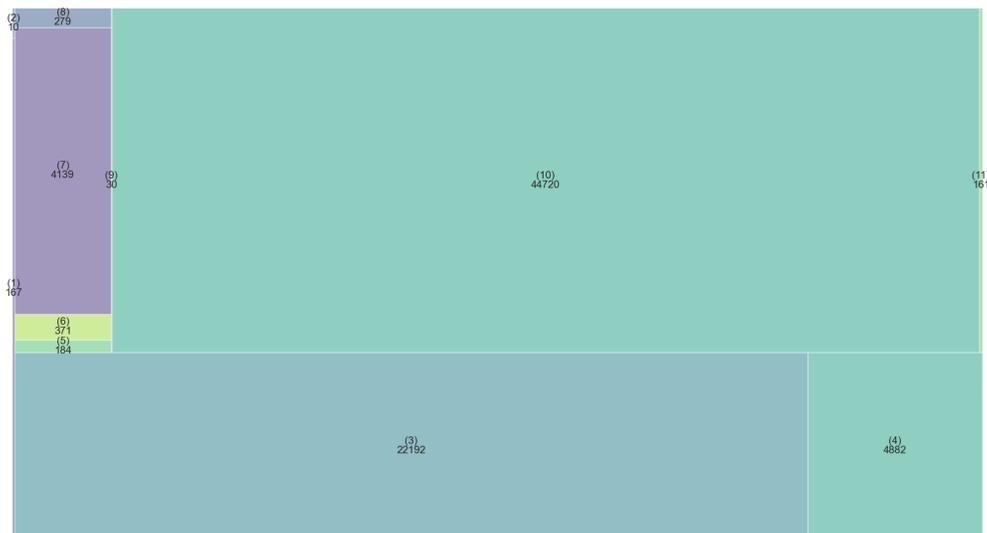


Figura 3.65: Treemap en el que se muestra la proporción de accidentes laborales *in itinere* según la forma o contacto que produjo la lesión que provocó el accidente.

Restringiendo solo a los accidentes mortales, prácticamente todos se deben a accidentes de tráfico; salvo un par de casos que se deben a choque o golpe contra objeto inmóvil y quedar atrapado, ser aplastado, sufrir una amputación. Además, varios también se deben a infartos, derrames cerebrales y

otras causas estrictamente naturales, aunque este caso es más común en los accidentes que no son *in itinere*.

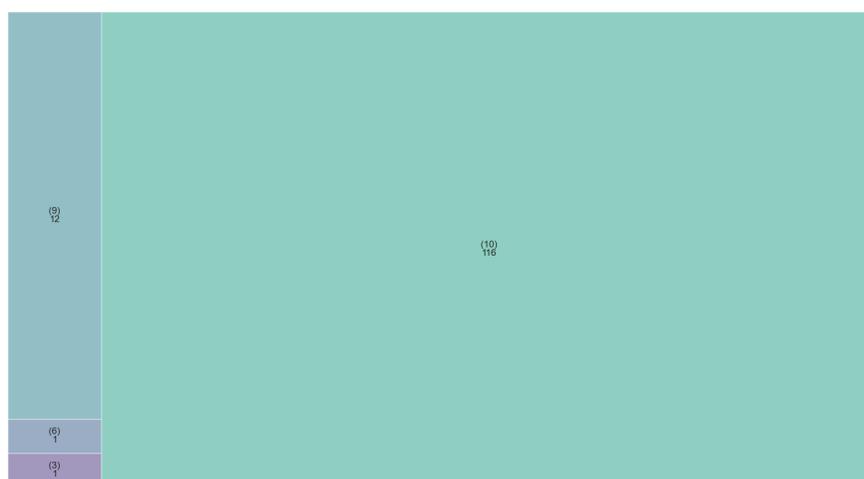


Figura 3.66: Proporción de accidentes laborales mortales *in itinere* según la forma o contacto que produjo la lesión que provocó el accidente.

### 3.2.4. Características de la temporalidad

Para finalizar el análisis de las diferentes características relacionadas con los accidentes laborales, en este apartado se muestra cómo se distribuyen estos en función del tiempo: las horas, días y meses en los que ocurren con más frecuencia este tipo de accidentes. Se presenta esta información a lo largo de los 8 años considerados en el estudio, pero prestando más atención al año 2021, más reciente.

#### ■ Hora del día en la que tuvo lugar el accidente

En primer lugar, se revisa el número de accidentes laborales según las horas del día.

En la Figura 3.67 se muestra el número de accidentes que tuvieron lugar a lo largo del 2021 clasificados por la hora en la que ocurrieron. La media anual de accidentes por franjas horarias es ligeramente superior a los 30000. Por encima de este valor se encuentran aquellas que transcurren a lo largo de la mañana: de 8 a 2 del mediodía, destacando sobre todo las horas entre 10 y 13. La suma de todas las horas de la madrugada (de 0 a 7) también se encuentra por encima de la media. Por la tarde y en la franja de 7 a 8 de la mañana, el número de accidentes de trabajadores que se encuentran en su jornada laboral es menor. La gráfica anterior muestra los resultados de 2021, para el [resto de años](#) también está disponible. Aunque en este caso, también se muestra esta información a través del *heatmap* de la Figura 3.68.

En el *heatmap* se puede apreciar que la distribución de los accidentes a lo largo de las horas no varía a penas de un año a otro. Sin embargo, sí que se aprecia que los colores cambian ligeramente a colores más claros desde 2014 a 2019, esto está relacionado con el aumento general de accidentes laborales que tuvo lugar en esos años. En 2020 el mapa vuelve a oscurecerse por la disminución de accidentes laborales ese año, pero la distribución a lo largo de las horas es prácticamente coincidente con la ya explicada para el caso de 2021.

En cuanto a los accidentes *in itinere*, los resultados varían bastante, como es de esperar, pues las horas de jornada no coinciden con las de *itinere* en las que los empleados se desplazan al o desde el puesto de trabajo.

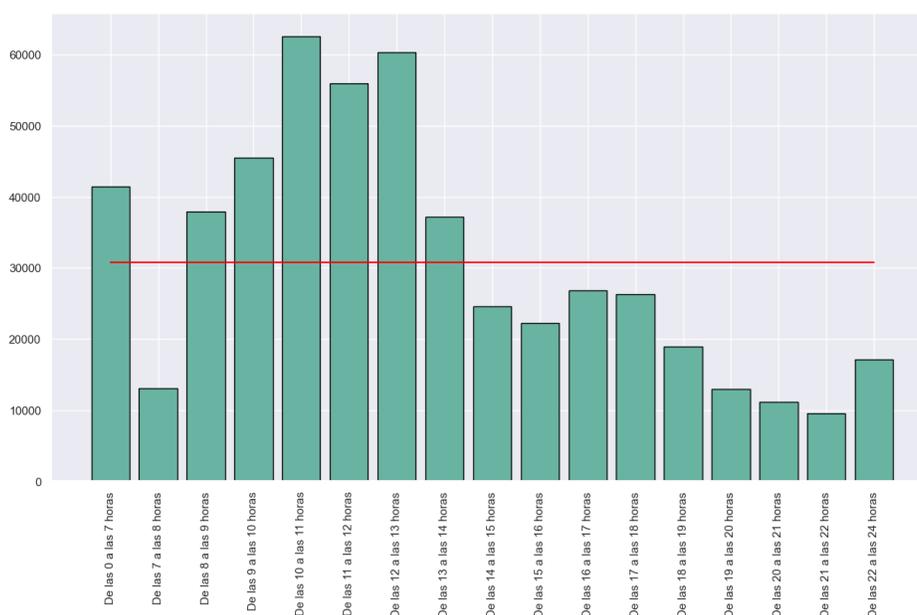


Figura 3.67: Número de accidentes laborales durante la jornada según la hora en la que ocurrieron.

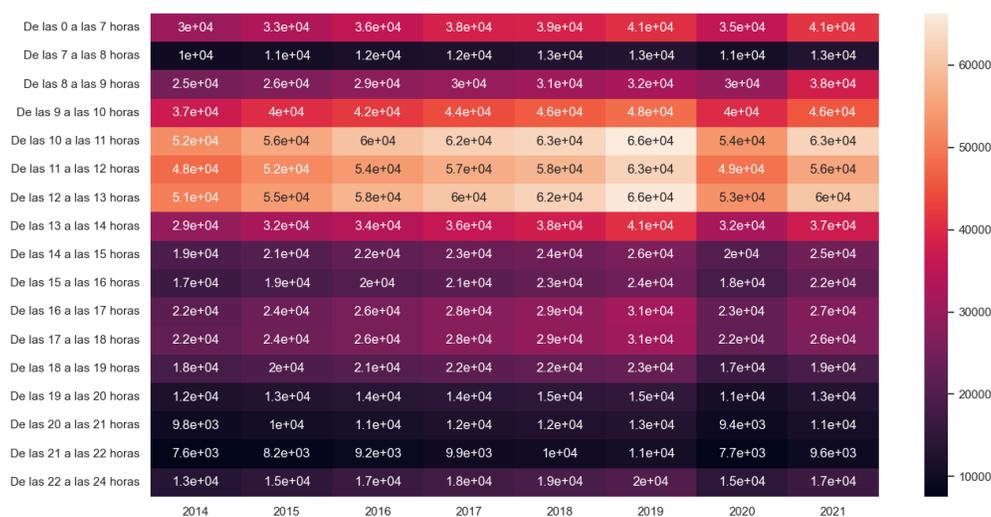


Figura 3.68: *Heatmap* del número de accidentes laborales durante la jornada según la hora del día en la que ocurrieron a lo largo de los años.

En la Figura 3.69 se pueden apreciar los resultados para el año 2021. Las horas en las que más accidentes ocurren este caso son la madrugada y las primeras horas de la mañana (de 7 a 8 y de 8 a 9). Además de 2 a 3 y de 3 a 4 de la tarde, horas de la comida en España, tienen lugar un número considerable de accidentes también. El resto de franjas horarias se encuentran en general bastante por debajo de la media.

En la Figura 3.70 se muestra el *heatmap* con los datos para todos los años desde 2014 hasta 2021. Al igual que en el caso de los accidentes durante la jornada, el patrón es prácticamente idéntico para todos los años, con la salvedad de que varía el número total de accidentes. En 2020 se observa la disminución

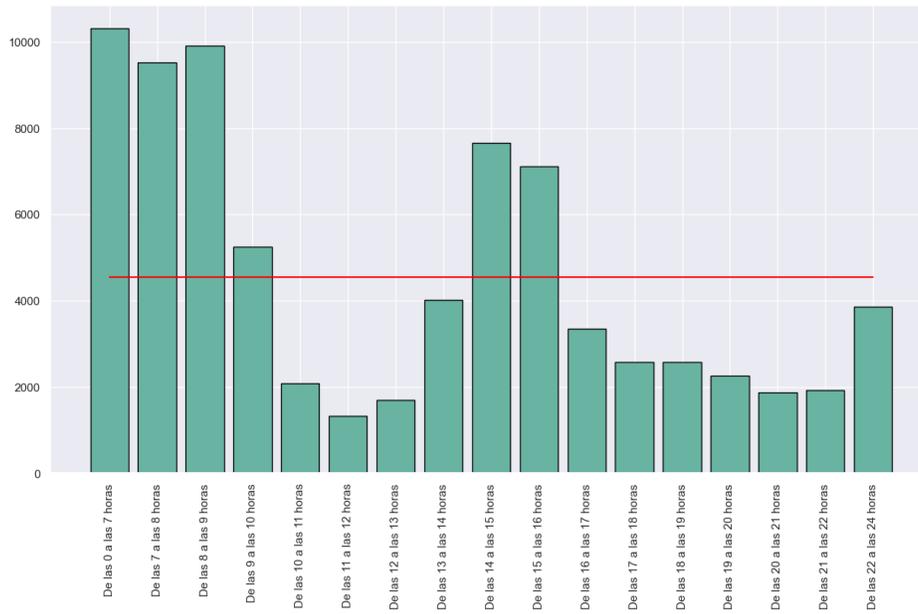


Figura 3.69: Número de accidentes laborales *in itinere* según la hora en la que ocurrieron.

de accidentes y de 2014 a 2019, el aumento de estos. Además, también se aprecia claramente que la franja de 8 a 9 es la que mayor número de accidentes presenta. Esto, sumado a que durante la jornada también presentaba un número ligeramente elevado de accidentes, lo sitúa como una de las franjas horarias donde más accidentes ocurren. De todas formas, el volumen de accidentes durante la jornada es en general mucho mayor que *in itinere*, por lo que las franjas de 10-11, 11-12 y 12-13 que despuntaban durante la jornada siguen siendo las que mayor número de accidentes presentan en total.

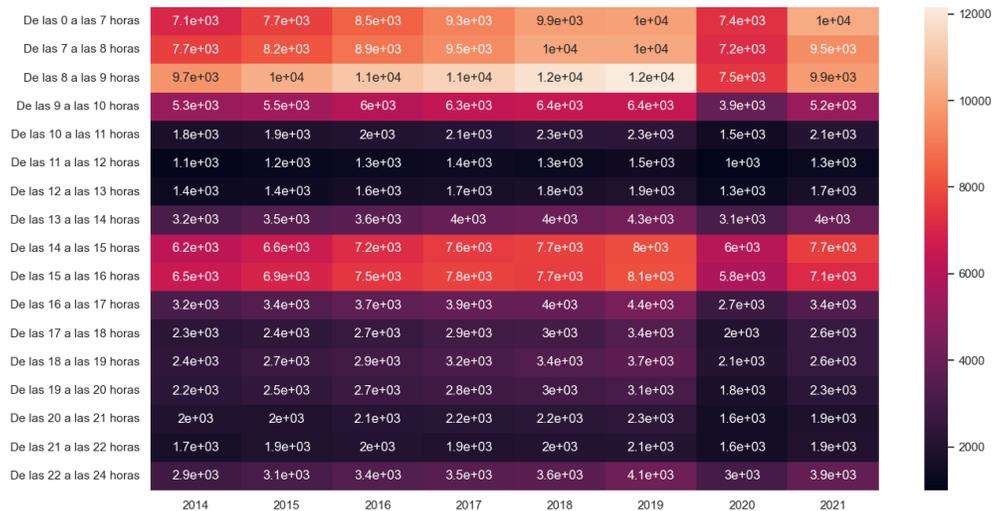


Figura 3.70: Heatmap del número de accidentes laborales *in itinere* según la hora del día en la que ocurrieron a lo largo de los años.

En cuanto a la diferenciación por gravedad del accidente, en este caso no se aprecia ninguna distinción notable. En la Figura 3.71 se muestran los resultados de 2021 para los tres tipos de gravedad

considerados: leves, graves y mortales. En la gráfica referente a accidentes mortales se aprecia quizás una distribución ligeramente más distinta, pero aun así la idea general es muy similar en los tres casos.

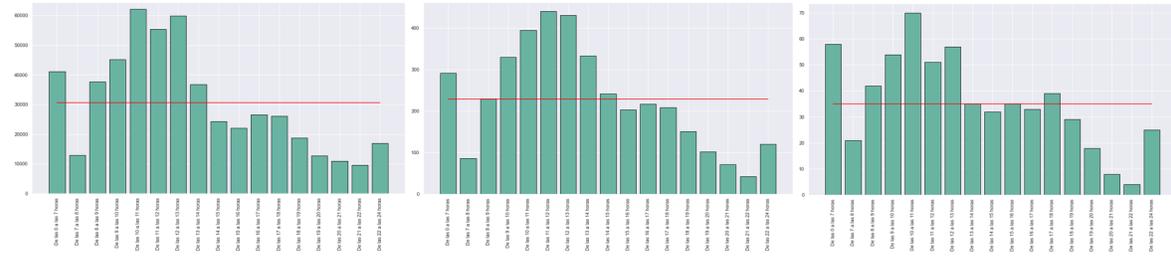


Figura 3.71: Número de accidentes laborales durante la jornada según la hora en la que ocurrieron y distinguiendo por gravedad del accidente, por orden: leves, graves y mortales.

■ Hora de la jornada en la que tuvo lugar el accidente

En la clasificación anterior se han considerado las horas en diferentes franjas que han abarcado las 24 horas diarias. A continuación se muestra el número de accidentes laborales, pero según las horas de la jornada laboral (desde la primera hasta completar la jornada de 8 horas, o las horas establecidas en el contrato del trabajador). También se consideran en una misma categoría todas las horas pasadas la octava.

En la Figura 3.72 se presentan los resultados para el año 2021, indicando el número de accidentes durante la jornada laboral según la hora de la propia jornada en la que ocurrieron.

Los resultados de los 7 años anteriores considerados en el estudio se pueden consultar. La distribución es prácticamente idéntica a la aquí mostrada, aunque cambia el número total de accidentes.

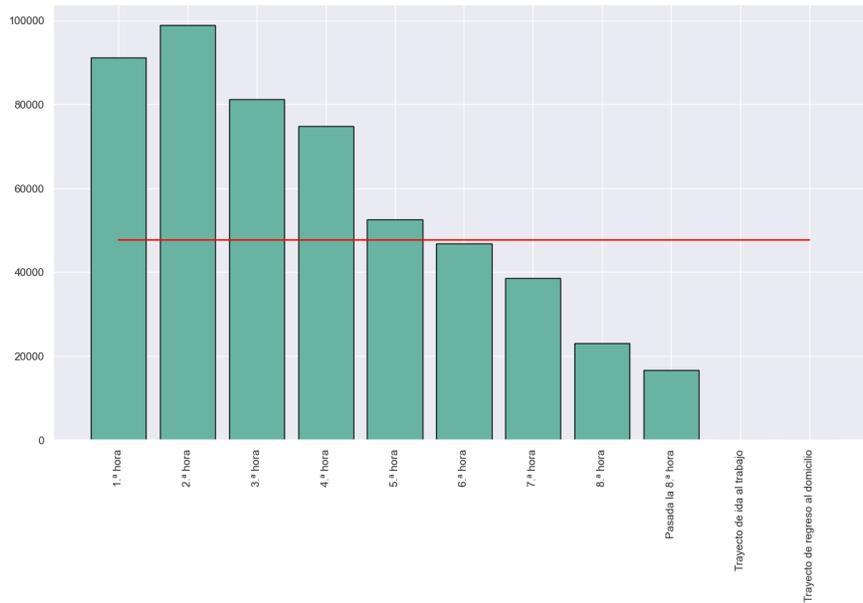


Figura 3.72: Número de accidentes laborales durante la jornada según la hora en la que ocurrieron.

Como regla general, a medida que avanzan las horas, el número de accidentes laborales disminuye.

Sin embargo, hay una excepción y es que la segunda hora de la jornada presenta un número ligeramente superior de accidentes que la primera.

Esta situación escalonada de los resultados seguramente esté relacionada con que todos los empleados tienen una primera hora de jornada laboral, pero dado que no todos los trabajadores tienen la misma duración de jornada, a medida que avanzan las horas hay menos trabajadores en la franja horaria (y, por tanto, menos personas que puedan sufrir un accidente en el ámbito laboral).

Nótese además que en esta clasificación también se consideran las categorías: trayecto de ida al trabajo y trayecto de regreso al domicilio. Dado que en la Figura 3.72 se consideran los accidentes durante la jornada, para estas categorías no se encuentra ningún accidente.

Para contrastar la afirmación anteriormente dada, en la Figura 3.73 se presenta un mapa de calor en el que se puede apreciar que se mantiene la misma distribución respecto a las horas de la jornada a lo largo de los años (con la salvedad de que el número total de accidentes varía).

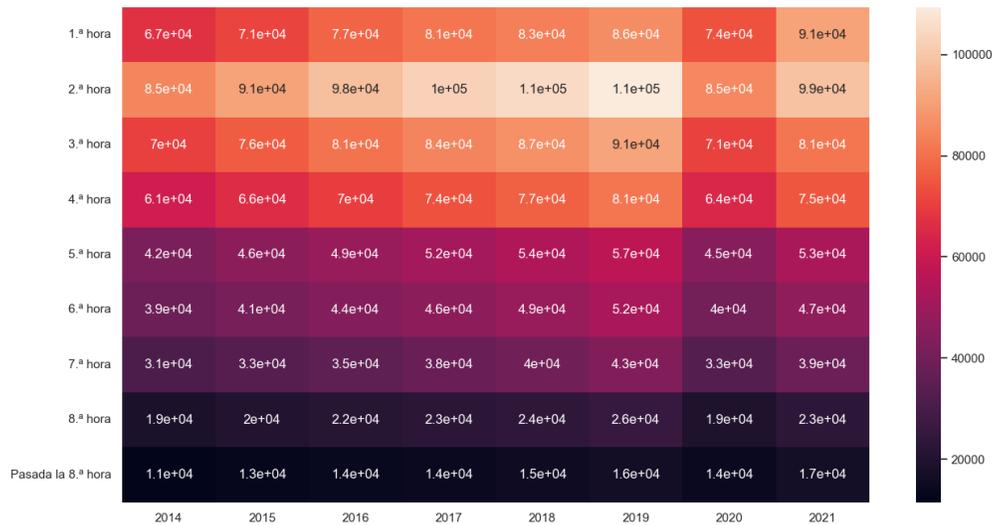


Figura 3.73: *Heatmap* del número de accidentes laborales durante la jornada según la hora de la jornada en la que ocurrieron a lo largo de los años.

En cuanto a la diferenciación de accidentes según la gravedad de estos, apenas se aprecia diferencia. Quizás la gráfica de los accidentes mortales se diferencia en mayor medida, pero esta situación viene influenciada por el menor tamaño de la población de accidentes mortales en comparación con las otras dos categorías. De nuevo, los datos mostrados corresponden al año 2021, pero son representativos de la situación general.

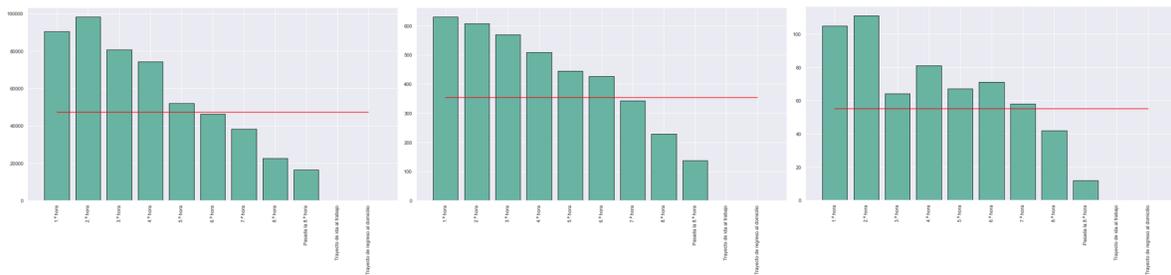


Figura 3.74: Número de accidentes laborales durante la jornada según la hora de la jornada en la que ocurrieron y distinguiendo por gravedad del accidente, por orden: leves, graves y mortales.

Respecto a la situación de los accidentes *in itinere*, se puede visualizar en la Figura 3.75 que el número de accidentes es mayor en el trayecto de ida al trabajo. Concretamente, en 2021, el número de accidentes en la ida duplica al número de accidentes en la vuelta a casa.

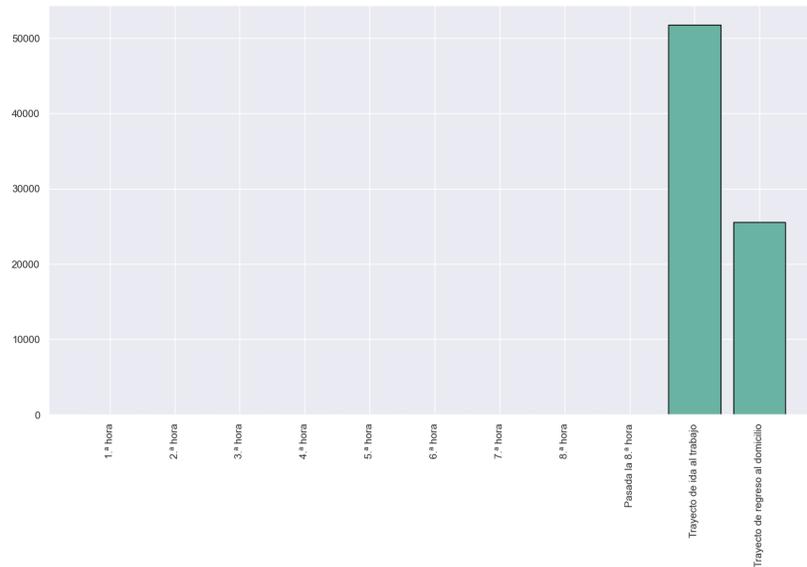


Figura 3.75: Número de accidentes laborales *in itinere* según la hora de la jornada en la que ocurrieron.

#### ■ Día de la semana en la que tuvo lugar el accidente

Continuando con el análisis de los accidentes según la temporalidad, la siguiente clasificación considerada distingue los accidentes laborales según el día de la semana en el que ocurrieron.

De nuevo se muestra la gráfica correspondiente a 2021, pero las correspondientes a los [años entre 2014 y 2020](#) también se encuentran disponibles.

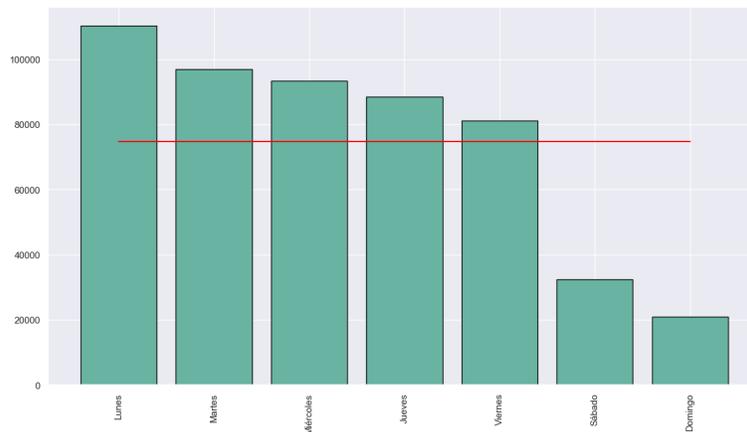


Figura 3.76: Número de accidentes laborales durante la jornada según el día de la semana en el que ocurrieron.

A medida que avanza la semana, van disminuyendo los accidentes laborales. La diferencia entre los 5 primeros días de la semana es notable, pero no excesivamente; sin embargo, el número de accidentes

que tienen lugar en sábado es muy inferior y todavía más en el caso del domingo. Este último resultado era de esperar, pues muchos trabajadores disfrutaban de sus días libres durante el fin de semana (con la excepción de la hostelería y ocupaciones similares).

En la Figura 3.77 se puede observar cómo esa misma distribución se mantiene a lo largo de los años, desde 2014.



Figura 3.77: *Heatmap* del número de accidentes laborales durante la jornada según el día de la semana en la que ocurrieron a lo largo de los años.

Se puede ver, además, cómo el número de accidentes aumenta hasta 2019, desciende en 2020 y vuelve a subir en 2021. En el caso de los accidentes *in itinere*, el análisis es muy similar. En la Figura 3.78 se muestran los resultados respectivos al año 2021.

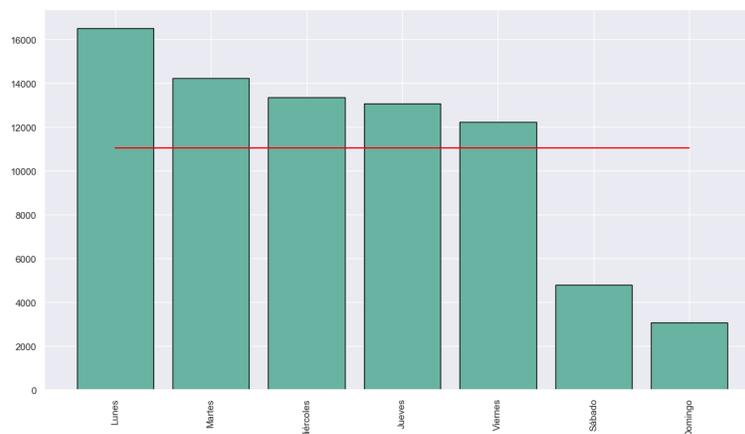


Figura 3.78: Número de accidentes laborales *in itinere* según el día de la semana en el que ocurrieron

### ■ Mes en el que tuvo lugar el accidente

Como último aspecto que revisar para finalizar el análisis de los accidentes laborales en cuanto a temporalidad, se muestra la clasificación de los accidentes según el mes en el que han ocurrido.

En primer lugar, se muestra de forma detallada los números correspondientes al año 2021 (Figura 3.79). También se incluye un mapa de calor en el que se puede observar la distribución de los accidentes a lo largo de los años considerados en el estudio. Además, se proporciona *online* el gráfico de barras de cada uno de los 8 años, desde 2014 a 2021.

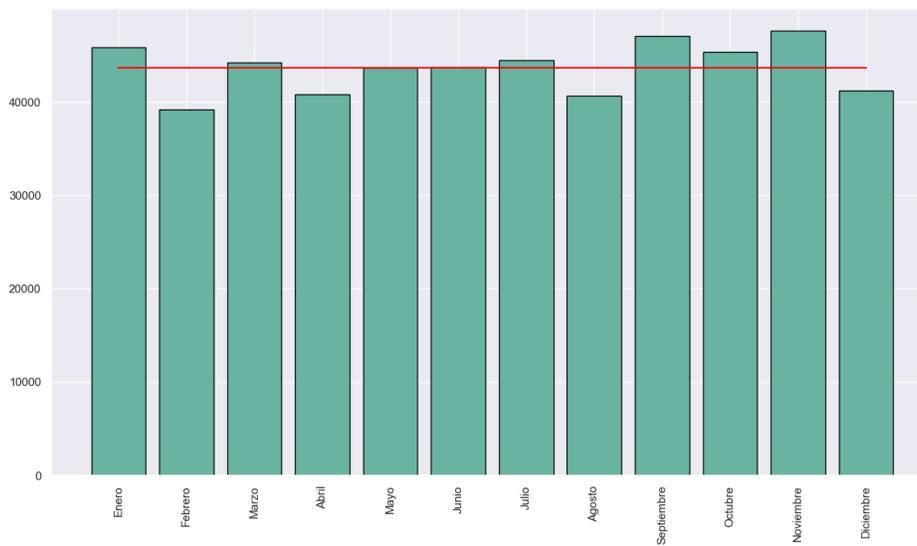


Figura 3.79: Número de accidentes laborales durante la jornada según el mes en el que tuvieron lugar.

En general, el número de accidentes laborales que tienen lugar cada mes ronda los 40000. Concretamente, la media de accidentes mensuales en 2021 fue 43650. Lógicamente, algunos meses se encuentran por encima y otros por debajo de la media. Estos últimos son en general los meses que coinciden con los periodos típicos de vacaciones: abril (Semana Santa), agosto (vacaciones de verano) y diciembre (Navidad). Febrero es también uno de los meses que presenta un menor número de accidentes, pero en este caso seguramente esté relacionado con que tiene un menor número de días (en el caso de 2021: 28).

Para comparar el resultado mostrado con el resto de años, véase la Figura 3.80. Se perciben cambios de colores entre los diferentes años pero es debido a que varía el número total de accidentes. De todas formas, se mantiene más o menos constante cuáles son los meses de colores más claros (más accidentes) y los de colores más oscuros (menos accidentes).

Cabe destacar que los meses desde marzo a junio de 2020 (y en general el 2020) presentan resultados distintos de los que se esperaban en función de los años anteriores. Esta situación se debe claramente a que ese periodo de tiempo coincide con los primeros meses de la pandemia por el COVID-19, de los cuales se pasaron varios meses en cuarentena en España. Es por este motivo que se observa una bajada tan drástica en el número de accidentes laborales durante esas fechas.

Se analizan ahora los accidentes laborales *in itinere*. En la Figura 3.81 se presenta un gráfico de barras en el que se muestra el número de accidentes por meses. Las gráficas de los años restantes también están disponibles.

Este año los resultados presentan un valor atípico en el mes de enero. Pues en dicho mes se superaron los diez mil accidentes *in itinere*. El dato destaca frente al resto de meses de ese mismo año, donde la segunda mayor cifra se alcanza en noviembre (7489). En la Figura 3.82 se puede observar los resultados de todos los años del estudio.

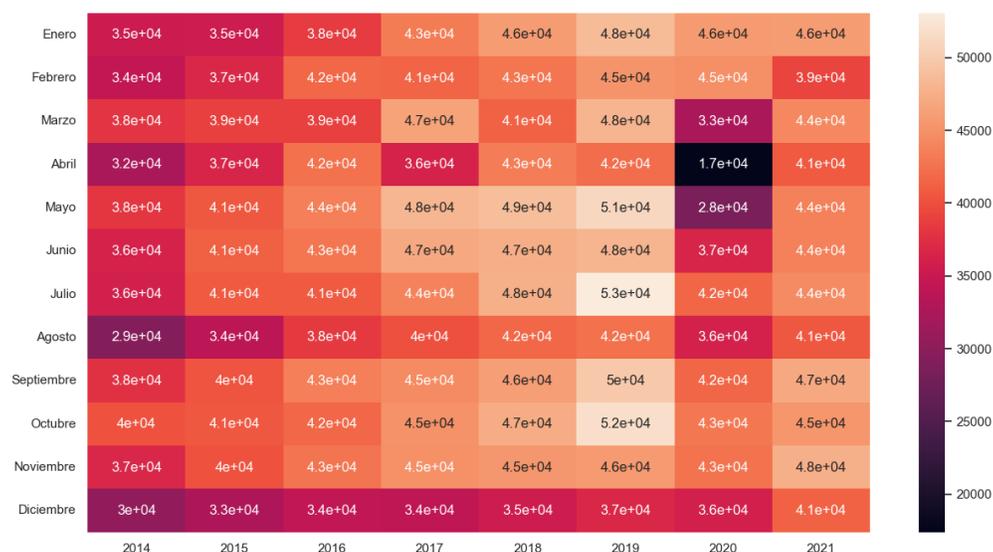


Figura 3.80: *Heatmap* del número de accidentes laborales durante la jornada según el mes en el que ocurrieron a lo largo de los años.

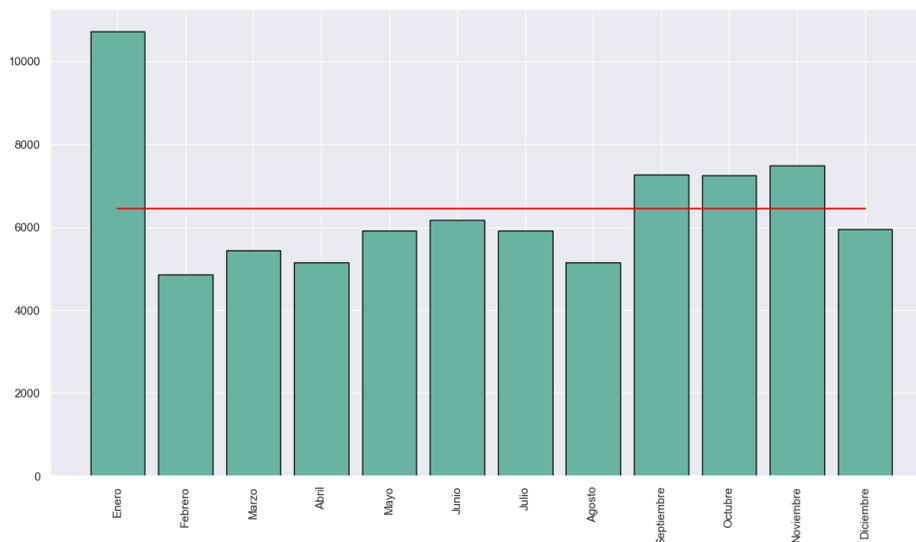


Figura 3.81: Número de accidentes laborales *in itinere* según el mes en el que tuvieron lugar.

Tan solo observando los colores de la gráfica, queda claro que el mes de enero de 2021 tuvo lugar un número atípico y muy alto de accidentes laborales. Es el único mes que alcanza las 5 cifras. En cuanto al resto de años, en 2020 se vuelve a apreciar la bajada de casos debido al confinamiento provocado por la crisis del COVID-19.

En cuanto a la distribución general, se muestra bastante homogénea a lo largo de los años. Aun así, se aprecia, al igual que para los accidentes durante la jornada, la bajada en febrero y los meses de vacaciones. Además, también se puede destacar el aumento de casos a medida que avanzan los años, sobre todo desde 2014 a 2019.

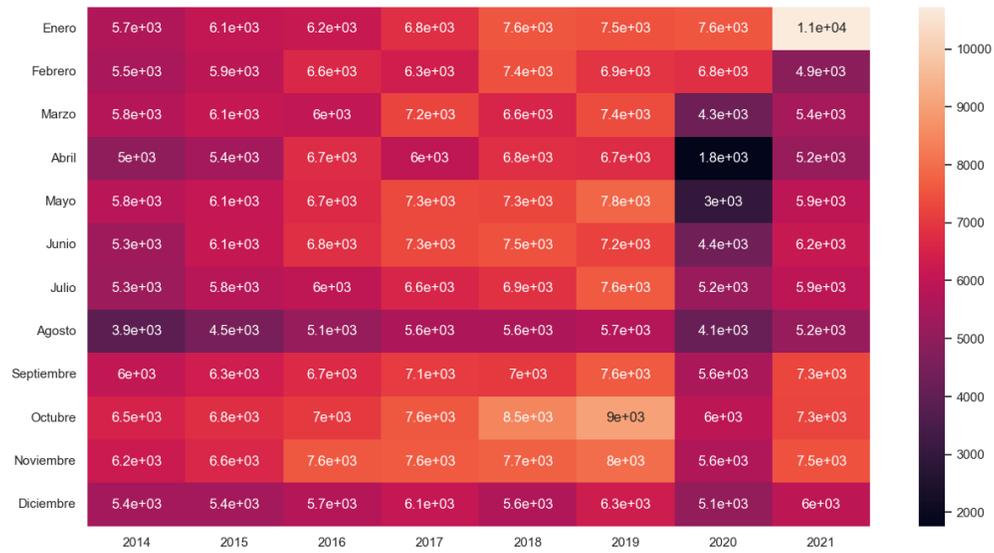


Figura 3.82: *Heatmap* del número de accidentes laborales *in itinere* según el mes en el que ocurrieron a lo largo de los años.

La conclusión general de todos los apartados considerados se incluye en el último capítulo de este documento, destinado ya a recopilar las conclusiones de los diferentes análisis realizados.

## Capítulo 4

# Análisis de las series temporales de accidentes laborales en España

En el capítulo anterior se ha presentado un análisis exhaustivo en el que se ha tratado de resumir y mostrar de forma útil y clara la información proporcionada en los archivos del Ministerio de Trabajo sobre los accidentes laborales que causan baja en España. Las estadísticas mostradas se centran principalmente en la distinción de accidentes según la gravedad de estos y si tuvieron lugar durante la jornada laboral o fuera de esta, mientras se acudía a ella o se abandonaba (*in itinere*).

En este nuevo capítulo se trata de complementar los resultados anteriores utilizando la información almacenada en los *datasets* *dfGEOGRAFICO* y *dfSECTORES*. Con la intención de mostrar los datos desde otro enfoque, se han puesto en práctica métodos estadísticos más complejos que proporcionan otra información distinta a la ya obtenida, ayudando a entender mejor la situación actual de los accidentes.

Son dos los análisis que se presentan en este nuevo capítulo. En primer lugar, aprovechando que se dispone de la cantidad de accidentes mensuales sufridos a lo largo de los 8 años del estudio, se ha construido la serie temporal correspondiente. El objetivo es modelizarla y obtener predicciones a partir de ella.

Por otro lado, se dispone de esta misma información pero correspondiente a cada una de las provincias y CC.AA. También se tiene para los sectores, secciones y divisiones económicas. Por tanto, se pueden obtener las series de tiempo correspondientes a cada una de las provincias o sectores económicos y modelizar cada una de ellas. Esto daría lugar al estudio de más de cien de series de tiempo, lo cual sería inabarcable en este trabajo. En su lugar, se ha llevado a cabo un proceso de *clustering*, en el que se pretende mostrar las similitudes entre las diferentes provincias y las diferentes secciones económicas en función de la distribución temporal de su número de accidentes. En definitiva, se pretende agrupar estas series buscando tendencias similares: picos en el número de accidentes, bajadas, resultados periódicos, etc.

Ambas técnicas tienen en común el uso de series temporales, que es una herramienta frecuentemente utilizada. Resulta útil tanto para predecir el futuro, como para entender el pasado, por lo que proporcionan una visión completa de la situación considerada.

En la primera sección de este capítulo se presentan los métodos que han sido utilizados. En primera instancia, se introduce de forma breve la teoría sobre la modelización de series de tiempo. Posteriormente, se incluye la teoría en cuanto a los métodos de *clustering*, concretando el método usado para este caso aplicado a la agrupación de series temporales.

Tras la introducción de los conceptos teóricos, se muestran los resultados correspondientes a ambas técnicas aplicadas a los datos del Ministerio de Trabajo.

## 4.1. Modelos y técnicas empleadas

En este apartado se introducen de forma teórica los modelos y métodos usados en esta segunda parte del análisis de los datos sobre accidentalidad laboral.

### 4.1.1. Modelos de series temporales

Las referencias principales de esta sección han sido Brockwell and Davis (2016) y, sobre todo, Cowpertwait and Metcalfe (2009). Para una información más completa acerca del modelado de series temporales, se recomienda consultar alguna de las anteriores referencias. Esta sección tiene como objetivo introducir el modelo con el cual se ha ajustado la serie de tiempo que se pretende estudiar: ARIMA estacional con intervención causada por el COVID-19. Para ello, se introducen previamente conceptos específicos de las series temporales y otros modelos más sencillos.

Una **serie de tiempo** (en tiempo discreto) es un conjunto de observaciones  $\{x_t\}_{t=1}^T$  tomadas a lo largo del tiempo en intervalos ya fijados hasta el instante final  $T$ . La parte esencial en un análisis de series de tiempo consiste en la selección de un modelo de probabilidad adecuado para los datos.

La base de estos modelos reside en que las observaciones  $\{x_t\}_{t=1}^T$  son la realización de un conjunto de variables aleatorias  $\{X_t\}_{t=1}^T$  (nos referiremos tanto al conjunto de variables como a las observaciones como serie de tiempo). Para seleccionar el modelo que mejor ajusta la serie existen varias categorías a las que se le debe prestar atención:

1. **Tendencia:** conviene determinar si la serie es estacionaria o, por lo contrario, presenta subidas y/o bajadas notables.
2. **Componente estacional:** resulta importante conocer si existe periodicidad en los datos, de forma que cada cierto periodo de tiempo haya una relación entre las componentes de la serie.
3. **Eventos o cambios en el comportamiento de la serie:** conviene conocer si la serie presenta algún cambio brusco de tendencia. De forma complementaria, también resulta interesante saber a qué se puede deber dicho cambio.
4. **Datos atípicos:** como en cualquier modelo que pretenda ajustar datos, conviene analizar si existen datos atípicos en la serie de tiempo considerada.

Todas estas características juegan un papel importante en la selección del modelo. Además, otro aspecto que interesa valorar es si la serie es homocedástica, es decir, si la variabilidad de la serie se mantiene estable a lo largo de todo el proceso. En el caso de que esto no ocurra, es común aplicarle alguna modificación a la serie. Por lo general se le suele aplicar la función logaritmo neperiano a los datos, para tratar de eliminar la heterocedasticidad de estos.

Presentados los aspectos a valorar, se introducen a continuación los modelos ARIMA con los que pretendemos modelizar la serie mensual de los accidentes laborales en España. Para ello, habrá que explicar previamente otros modelos esenciales para acabar exponiendo el que se desea ajustar.

Se incluyen a continuación una serie de definiciones básicas para introducir los procesos que definen las series de tiempo.

Se tienen las siguientes funciones para una serie de tiempo  $\{X_t\}_{t=1}^T$ :

- **Función de medias:**  $\mu_X(t) = \mathbb{E}(X_t)$ .

- **Función de autocovarianzas:**

$$\gamma_X(t, t+h) = \gamma_X(0, h) = \gamma_X(h) = \text{Cov}(X_t, X_{t+h}) = \mathbb{E}[(X_t - \mu_X(t))(X_{t+h} - \mu_X(t+h))].$$

Se dice que una serie  $\{X_t\}_{t=1}^T$  es **estacionaria** si la función  $\mu_X(t)$  es independiente de  $t$  y la función  $\gamma_X(t, t+h)$  es independiente de  $t$  y  $h$ . Por ejemplo, si  $\{X_t\}_{t=1}^T$  es una secuencia de variable incorreladas, con media nula y varianza  $\sigma^2$  para todo  $t \in \{1, \dots, T\}$ , estamos ante una serie estacionaria.

Concretamente, se define el **ruido blanco** como la secuencia de variables aleatorias incorreladas  $\{a_t\}_t$  que verifican que

$$\mu_X(t) = 0 \quad \forall t, \quad \sigma_t = \text{Var}(X_t) = \gamma(0) = \sigma^2 \quad \forall t \quad \text{y} \quad \gamma(h) = 0 \quad \text{si} \quad h \neq 0.$$

Considerando  $\{a_t\}_t$  como un proceso de ruido blanco y  $\{X_t\}_t$  una serie de tiempo, se definen:

- Los procesos **causales**:  $X_t = \psi_0 a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots$  cumpliendo  $\sum_{i=0}^{\infty} |\psi_i| < \infty$ .
- Los procesos **invertibles**:  $X_t = a_t + \pi_1 X_{t-1} + \pi_2 X_{t-2} + \dots$  cumpliendo  $\sum_{i=0}^{\infty} |\pi_i| < \infty$ .

Finalmente, antes de proceder con la definición de los distintos modelos, se introduce el **operador de retardo**  $B$  y el **operador de diferenciación**  $\Delta$ , que dada una secuencia de variables aleatorias  $\{X_t\}_t$  verifican

$$BX_t = X_{t-1} \quad \text{y} \quad \Delta X_t = X_t - X_{t-1} = (1 - B)X_t.$$

Estos operadores juegan un papel trascendental, pues permitirán modelar series no estacionarias y con estacionalidad<sup>1</sup>.

Tras la introducción de todos estos conceptos básicos, se procede entonces con la presentación de los modelos que se consideran posteriormente. Para ello habrá primero que introducir unos modelos más sencillos, que servirán de base para los más complejos:

Dada una secuencia de variables aleatorias de ruido blanco  $\{a_t\}_t$  independientes de la secuencia de variables  $\{X_t\}_t$  se definen los siguientes modelos:

- **Modelo autorregresivo**: Un modelo autorregresivo de orden  $p$ ,  $AR(p)$ , consiste en un proceso estacionario (e invertible) que verifica la fórmula

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + a_t, \quad \text{con} \quad \phi_i \in \mathbb{R} \quad \forall i \in \{1, \dots, p\} \quad \text{y} \quad \phi_p \neq 0.$$

Se puede escribir también como

$$\Phi_p(B)X_t = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)X_t = a_t.$$

- **Modelo de medias móviles**: Un modelo de medias móviles de orden  $q$ ,  $MA(q)$ , consiste en un proceso estacionario (y causal) que admite la representación

$$X_t = \theta_0 a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q}, \quad \text{con} \quad \theta_i \in \mathbb{R} \quad \forall i \in \{0, \dots, q\} \quad \text{y} \quad \theta_q \neq 0.$$

Se puede escribir también como

$$X_t = \Theta_q(B)a_t = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q)a_t.$$

Nótese que todos los parámetros que participan en estos modelos son números reales. La estimación de estos se obtiene mediante métodos de máxima verosimilitud, asumiendo en general gaussianidad, o con métodos de mínimos cuadrados condicionados.

El primero de estos dos modelos permite modelizar la serie de forma autorregresiva, haciendo que el valor de la serie en tiempo  $t$  dependa de instantes anteriores de la serie. Por otro lado, en el modelo

<sup>1</sup>Una serie presenta componente estacional si existe relación entre sus componentes cada cierto periodo de tiempo (estaciones).

de medias móviles, el valor de la serie en tiempo  $t$  solo depende del ruido blanco en el instante  $t$  y momentos anteriores. Combinando estos dos procesos surge un nuevo modelo de series temporales:

Un conjunto  $\{X_t\}_t$  sigue un modelo **ARMA**( $p, q$ ) si es estacionario y para cada instante  $t$  verifica

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p} = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q},$$

con  $\phi_p \neq 0$  y  $\theta_q \neq 0$ . La ecuación anterior también se puede escribir como

$$\Phi(B)X_t = \Theta(B)a_t,$$

siendo  $\Phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$  y  $\Theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q$ .

Resulta claro que un proceso **ARMA**( $p, 0$ ) es equivalente a un **AR**( $p$ ) y se corresponde con  $\Theta(z) = 1$ . Asimismo, un **ARMA**( $0, q$ ) es equivalente a un **MA**( $q$ ) y se corresponde con  $\Phi(z) = 1$ .

Otro aspecto que verifican los procesos **ARMA**( $p, q$ ) es que pueden ser (o no) causales y/o invertibles. Se cumple que un proceso  $\{X_t\}_t$  estacionario que sigue un modelo **ARMA**( $p, q$ ) es

$$\text{Causal} \iff 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p \neq 0 \forall z \text{ tal que } |z| \leq 1,$$

$$\text{Invertible} \iff 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q \neq 0 \forall z \text{ tal que } |z| \leq 1.$$

A continuación se introduce un nuevo modelo. Para introducirlo es interesante mencionar que, como ya se comentó anteriormente, las series pueden presentar una componente estacional. Esta situación se corresponde con cuando las componentes de la serie dependen de componentes anteriores pero de forma periódica. Considerando un periodo de duración  $s$ , se define el proceso **estacional ARMA**( $P, Q$ ) $_s$  como un proceso **ARMA**( $sP, sQ$ ), que sigue la fórmula:

$$X_t - \phi_1 X_{t-s} - \phi_2 X_{t-2s} - \cdots - \phi_p X_{t-ps} = a_t + \theta_1 a_{t-s} + \theta_2 a_{t-2s} + \cdots + \theta_q a_{t-qs},$$

también escrita como

$$\Phi(B)(1 - B^s)X_t = \Theta(B^s)a_t.$$

Hasta ahora, todos los modelos considerados corresponden a la modelización de series estacionarias, pero es importante destacar que no todas las series lo son. En efecto, la gran mayoría de series temporales no lo son y presentan distintas tendencias. Por tanto, los modelos **ARMA** no son suficientes para modelar todas las posibles series de tiempo. Cuando se presenta esta situación, se deben aplicar transformaciones a los datos dirigidas a obtener una nueva serie que sí sea estacionaria, verificando las condiciones para ser ajustada correctamente por un modelo **ARMA**. El método más común consiste en la diferenciación de la serie, dando lugar a lo que se conoce como modelos **ARIMA**.

Sea  $d \in \mathbb{Z}^+$ , se dice que  $\{X_t\}_t$  sigue un proceso **ARIMA**( $p, d, q$ ) si la serie  $Y_t := (1 - B)^d X_t$  sigue un proceso **ARMA**( $p, q$ ) causal. Esta definición se corresponde con que  $\{X_t\}_t$  verifique la fórmula

$$\Phi(B)(1 - B)^d X_t = \Theta(B)a_t,$$

donde  $\Phi(z)$  y  $\Theta(z)$  son polinomios de orden  $p$  y  $q$  respectivamente (como los expuestos para el modelo **ARMA**( $p, q$ )), cumpliendo que  $\Phi(z) \neq 0$  para  $|z| \leq 1$ .

Tras haber introducido los modelos anteriores, más sencillos, se presenta a continuación el modelo con el que se ajustará la serie de tiempo de los accidentes laborales en España, pues es el que más se ajusta a las características de la serie:

Dados  $d, D \in \mathbb{Z}^+$ , se dice que  $\{X_t\}_t$  sigue un proceso **estacional ARIMA**  $(p, d, q) \times (P, D, Q)_s$  de **periodo**  $s$  si la serie diferenciada  $Y_t = (1 - B)^d(1 - B^s)^D X_t$  es un proceso ARMA causal definido por

$$\Phi(B)\tilde{\Phi}(B^s)Y_t = \Theta(B)\tilde{\Theta}(B^s)a_t,$$

donde  $\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ ,  $\tilde{\Phi}(z) = 1 - \tilde{\phi}_1 z - \dots - \tilde{\phi}_p z^p$ ,  $\Theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$  y  $\tilde{\Theta}(z) = 1 + \tilde{\theta}_1 z + \dots + \tilde{\theta}_q z^q$ .

Se tiene por tanto, que

$$\Phi(B)\tilde{\Phi}(B^s)(1 - B)^d(1 - B^s)^D X_t = \Theta(B)\tilde{\Theta}(B^s)a_t.$$

Es decir,  $\{X_t\}_t$  sigue un modelo ARIMA estacional si después de aplicarle diferenciaciones regulares para eliminar la tendencia y diferenciaciones estacionales para eliminar la estacionalidad, la serie resultante sigue un proceso ARMA. Por tanto, este modelo captura tanto la dependencia regular entre las componentes de la serie, como la estacional. Resaltar que este proceso generaliza todos los anteriores.

Tanto este último modelo, como todos los expuestos en el trabajo, consideran la media de la serie como nula. En el caso de que la serie presente una media  $\mu$  distinta de cero, se debe ajustar la secuencia de variables aleatorias  $\{X_t - \mu\}_t$ .

Para finalizar, con la introducción de los conceptos teóricos necesarios para modelizar la serie mensual de accidentalidad laboral, se presenta a continuación una situación muy recurrente en el análisis de intervención de series y que, concretamente, se da en la serie que se pretende analizar. Es muy común que durante un periodo de tiempo en el que se observan valores de una serie, en algún punto ocurra algo que afecte al nivel de la serie. Un ejemplo de esto podría ser una serie de tiempo en la que se recojan las ventas semanales de un producto: en principio las ventas no son muy altas, pero a raíz de publicitarlo, las ventas suben y eso se refleja en la serie de tiempo a partir del momento en el que comenzó a publicitarse, con una subida del nivel de la serie.

Existen tres diferentes tipos de intervenciones:

- La serie sufre una intervención en el instante  $t = t_0$  de la serie y se modifica el nivel de la serie permanentemente.
- La serie sufre una intervención puntual en el instante  $t = t_0$  que solo afecta al nivel de la serie en ese punto.
- La serie sufre una intervención en el instante  $t = t_0$  que afecta al nivel de la serie en ese punto y en los posteriores, pero el efecto se va diluyendo hasta alcanzar de nuevo el nivel inicial.

La situación que más se asemeja al caso a tratar es la última. Es la que mejor representa el efecto causado por la pandemia del COVID-19 sobre la serie de accidentes laborales en España. Dado que la intención de esta sección es presentar de forma teórica el modelo con el que se ha ajustado dicha serie, se define a continuación solamente el caso oportuno.

Se considera  $\{X_t\}_t$  como el proceso que genera la serie de tiempo sin considerar que ha tenido lugar una intervención y se denota por  $\{Y_t\}_t$  el proceso que genera la serie intervenida (que es la que se observa). El modelo de la serie intervenida se escribe como

$$Y_t = \omega \sum_{j=0}^{\infty} \delta^j I_t(t_0 + j) + X_t, \text{ con } \omega \in \mathbb{R} \text{ y } \delta \in (0, 1),$$

donde  $\{X_t\}_t$  sigue un modelo de los ya introducidos y siendo

$$I_t(t_0) = \begin{cases} 1 & \text{si } t = t_0 \\ 0 & \text{si } t \neq t_0. \end{cases}$$

De esta forma, se obtiene que para todos los instantes anteriores a la intervención,  $t < t_0$ , la serie intervenida  $\{Y_t\}_t$  coincide con  $\{X_t\}_t$ :  $Y_t = X_t$ .

Para el caso  $t = t_0$ , se tiene  $Y_t = \omega + X_t$ . En este caso la intervención modifica el nivel de la serie en una cantidad  $\omega \in \mathbb{R}$  que debe ser estimada junto al resto de parámetros del modelo.

Dado el instante  $t = t_0 + j$ , por ejemplo, la serie intervenida tendría el siguiente valor  $Y_t = \omega\delta^j + X_t$ .

Es relevante destacar que el parámetro  $\delta$  debe encontrarse en el intervalo  $(0, 1)$  y se estimará junto al resto del modelo. Ocurre que, para cada instante que supera el tiempo  $t_0$ , el coeficiente  $\omega$  que modifica el nivel de la serie se va multiplicando por potencias cada vez más altas de  $\delta$  y dado que  $0 < \delta < 1$ , esto provoca que cada vez el efecto sobre el nivel de la serie sea menor. Finalmente, cuando los valores de la serie se encuentren lo suficientemente distanciados del efecto inicial, el nivel de la serie se restaurará a los valores que se esperaba que presentase si no hubiese sufrido ninguna intervención.

Como apreciación final, es importante mencionar que en el capítulo 3 de Cowpertwait and Metcalfe (2016) se puede encontrar los detalles de cómo identificar los modelos apropiados para modelar una serie. No se incluye en este trabajo este tipo de información, pues el interés en presentar los distintos modelos para series temporales reside simplemente en dar contexto a los métodos empleados.

En la sección 4.2 se ponen en práctica los modelos presentados para ajustar la serie objetivo del análisis.

#### 4.1.2. *Clustering* de series temporales

Tal y como ha sido mencionado al inicio del capítulo, además de modelizar la serie de los accidentes laborales mensuales en España, se ha trabajado también con las series mensuales de cada una de las provincias (y ciudades autónomas) y secciones económicas recogidas en los datos del Ministerio de Trabajo. La intención es buscar relaciones entre las series por medio del agrupamiento de estas.

Concretamente, el procedimiento que se ha realizado es la clusterización de las diferentes series. Este es un procedimiento recurrente en el análisis exploratorio de datos, pues no se poseen *a priori* categorías en las que englobar cada elemento, pero sí resulta interesante estudiar las similitudes entre ellas y examinar como se podrían agrupar en función de las mismas.

Para poder llevar a cabo procedimientos de *clustering*, es necesario establecer una medida de similitud entre los elementos que se pretenden agrupar. En el caso presente, la similitud entre las series se mide a través de la medida *Dynamic Time Warping* (DTW), que se podría traducir por deformación dinámica del tiempo. Por otro lado, el método de agrupamiento considerado ha sido *DTW barycenter averaging* (DBA).

En este sub apartado se presentará el funcionamiento de este algoritmo y las técnicas empleadas para agrupar mediante un método de clusterización las distintas series de tiempo en diferentes grupos.

Para ello, las referencias principales han sido Petitjean et al (2011) y Aghabozorgi et al (2015).

Como primera cuestión a abordar, conviene mencionar brevemente los métodos generales de agrupamiento en análisis no supervisado, para introducir a partir de ellos los específicos para series temporales. En primer lugar, se procede con la introducción del **método de particionamiento** conocido como **K-medias**. El algoritmo seguido para aplicar este método es el siguiente:

1. Para la inicialización del método se seleccionan  $k$  puntos diferentes a modo de centroides; siendo  $k$  el número prefijado de grupos que se quieren crear. A cada centroide se le asignan los individuos más cercanos a él, obteniéndose así los  $k$  *clústers* iniciales.
2. Se recalculan los centroides y se reasignan los grupos: para cada uno de los  $k$  grupos se obtiene un nuevo centroide, generalmente calculado como la media de los elementos del grupo. Una vez hecho esto, a cada elemento se le asigna el grupo de aquel centroide más próximo a él.
3. Se repite el paso 2 hasta que no se produzca ningún cambio a la hora de realizar las reasignaciones.

Para llevarlo a cabo, es interesante conocer qué valor de  $k$  hay que tomar. Generalmente, en el análisis exploratorio esto es desconocido, no se conoce previamente en cuantos grupos diferentes se dividen los datos. En el caso a tratar así ocurre, no hay información previa que permita conocer en cuantos grupos se podrán dividir las series que se pretenden clasificar. Es, por tanto, interesante aplicar otros métodos de forma complementaria para poder obtener un valor de  $k$  razonable.

En estos casos, es bastante útil aplicar **métodos jerárquicos de clusterización** a los datos. Existen dos grandes tipos en los que se engloban estos métodos: los aglomerativos, que a partir de los individuos van agrupándolos en clústers, y los divisivos, que parten de un grupo total y se va dividiendo en subgrupos. Los primeros son los más comunes y el algoritmo que los define es el siguiente:

1. Se considera cada observación como un *clúster* y la matriz de distancias total, que contiene las distancias entre todos los pares de observaciones (*clústers*) según la medida considerada.
2. Se selecciona el par de *clústers* más similar (que presenten la menor distancia entre ellos). Es relevante guardar la información de qué *clústers* se combinaron y a qué distancia.
3. Se unen los dos *clústers* anteriores en uno solo y se actualiza la matriz de distancias, eliminando los dos grupos anteriores e incluyendo este nuevo.
4. Se repiten los pasos 2 y 3 tantas veces como sea necesario (el número de elementos menos 1) hasta haber incluido todos los elementos en un mismo grupo.

Para realizar este procedimiento se suelen utilizar 3 métodos distintos para medir la distancia entre *clústers*, definiendo dicha distancia entre grupos como: la mínima de las distancias entre los elementos de uno y otro grupo; la máxima de esas distancias o, la más recurrente, la media de todas las distancias entre los elementos de los dos grupos considerados.

Con estos métodos aglomerativos se puede visualizar a qué distancias se unen los grupos, cómo estos van aumentando el número de elementos que poseen y cuáles se van agrupando antes. A partir de esto, se puede decidir un valor de  $k$  para emplear en el método de  $k$ -medias anteriormente expuesto.

Para mayor información acerca de los métodos de *clustering*, puede verse Johnson y Wichern (2007).

En cuanto a la aplicación de estos métodos a la clasificación de series temporales, hay varias consideraciones que tener en cuenta para poder llevar a la práctica estos procedimientos. Se presentan a continuación.

Un aspecto esencial en los métodos de *clustering* es definir la medida con la que se obtienen las distancias entre los elementos que se pretenden clasificar. En el caso de las series temporales hay varios aspectos que condicionan esta elección. Si las series presentan una longitud mayor de la que se puede almacenar o dimensiones muy altas (en el caso de las series multivariantes), el procedimiento de agrupamiento se puede ver comprometido. Sin embargo, el inconveniente más relevante es que la longitud de las series que se consideran no coincida. Además, también pueden resultar problemáticos los datos atípicos que las series puedan presentar.

Por todas estas razones, considerar la distancia euclídea como medida de similitud entre las series no es la elección más adecuada. La distancia euclídea muestra muy poca flexibilidad ante los datos atípicos y, en el caso de que la series no presenten la misma longitud, ni siquiera se puede aplicar.

En Aghabozorgi et al (2015) se presentan diferentes medidas que se pueden considerar para obtener las distancias entre series de tiempo. Entre ellas se encuentra una de las más comunes y la que se ha utilizado en este trabajo: **Dynamic Time Warping**. Esta medida fue introducida en Sakoe y Chiba (1978).

Se denotan mediante  $A = \{a_1, \dots, a_I\}$  y  $B = \{b_1, \dots, b_J\}$  dos series de tiempo de longitud  $I$  y  $J$  respectivamente. Se denota por  $F$  el conjunto (al que llamaremos camino) que recoge los pares formados entre los elementos de la serie  $A$  y  $B$ . Concretamente  $F = \{c(1), c(2), \dots, c(k), \dots, c(K)\}$ , donde  $c(k) = (i(k), j(k))$  representa la pareja de índices  $k$ -ésima del camino formada por el elemento  $i(k)$ -ésimo de la serie  $A$  y el elemento  $j(k)$ -ésimo de  $B$ . Dado que la longitud de las series no tiene por qué ser la misma, una de ellas tendrá más elementos y algunos de ellos se emparejarán con elementos

repetidos de la de menor tamaño. Por tanto, el número de parejas consideradas en el camino  $F$  será  $K = \max\{I, J\}$ . Existen múltiples caminos posibles, uno por cada combinación de parejas que se puede realizar con los elementos de  $A$  y  $B$  siempre que se cumplan las siguientes restricciones:

- Condición de monotonía:  $i(k-1) \leq i(k)$  y  $j(k-1) \leq j(k)$ .
- Condición de continuidad:  $i(k) - i(k-1) \leq 1$  y  $j(k) - j(k-1) \leq 1$ .

Teniendo en cuenta estas dos propiedades, se tiene que cumplir que

$$c(k-1) = \begin{cases} (i(k), j(k) - 1), \\ (i(k) - 1, j(k) - 1), \\ (i(k) - 1, j(k)). \end{cases}$$

- Condición de frontera:  $i(1) = j(1) = 1$ ,  $i(K) = I$  y  $j(K) = J$ .
- Condición de ajuste de ventana:  $|i(k) - j(k)| \leq r$ , donde  $r$  es un entero positivo que se selecciona de forma que la flexibilidad permitida en las tendencias temporales no sea excesiva. Es decir, permite regular la diferencia entre los periodos temporales de los elementos de las series que pretendemos agrupar.

En este trabajo se hará referencia a este valor como el **radio de Sakoe-Shiba**. En la aplicación del método al caso práctico a tratar no se ha impuesto restricciones sobre el valor de este radio.

Considerando estas características, se define entonces la distancia  $DTW$  entre dos series  $A$  y  $B$  como las definidas anteriormente como

$$D(A, B) = \min_F \left\{ \frac{\sum_{k=1}^K d(c(k)) \cdot w(k)}{\sum_{k=1}^K w(k)} \right\},$$

donde  $d(c(k)) = d(i(k), j(k))$  denota la distancia (euclídea, generalmente) que existe entre el elemento  $i(k)$ -ésimo de la serie  $A$  y el elemento  $j(k)$ -ésimo de la serie  $B$ . Los coeficientes  $w(k)$  son pesos asociados a cada par  $c(k)$  y que se introducen para contribuir a la flexibilidad del método.

Definida la métrica con la que se obtienen las distancias entre las series que se pretende agrupar, ya se puede proceder con el método de *clustering* apropiado.

El método de las  $k$ -medias presenta algunos inconvenientes a la hora de utilizar la distancia  $DTW$ , por culpa de la obtención de los centroides de los diferentes *clusters* en cada etapa. Por ese motivo, nace un nuevo método llamado **DTW baricenter averaging (DBA)**. La idea general de este procedimiento consiste en un proceso iterativo que a partir de una “serie media” inicial (centroide inicial) se va refinando con la intención de minimizar la distancia  $DTW$  con el conjunto de series que se pretende agrupar. Es un proceso similar al de  $k$ -medias, con la diferencia de cómo se calculan los centroides:

1. Al igual que en el método de  $k$ -medias se particiona el conjunto de series  $\mathbb{S} = \{S_1, \dots, S_n\}$  en  $k$  grupos. En la primera iteración se parte de  $k$  centroides iniciales y se agrupan junto a ellos las series con una menor distancia  $DTW$ .
2. Se recalculan los centroides a partir de las series contenidas en los grupos:
  - Para cada una de las series se obtienen los pares  $\{c(k)\}_k$  entre el centroide y las series, al igual que se hace para calcular la distancia  $DTW$  entre series.

- Cada coordenada de la serie centroide tiene asociada un valor de cada una de las series del grupo. Supongamos que hay  $m$  series en ese grupo, pues la coordenada  $l$ -ésima del nuevo centroide se obtiene como

$$\text{baricentro}[x_1, \dots, x_m] = \frac{x_1 + \dots + x_m}{m},$$

siendo  $x_i$  la coordenada de la serie  $S_i$  que se emparejó con la  $l$ -ésima del centroide.

3. Se reasignan los grupos, asignando a cada serie el centroide más cercano según la métrica  $DTW$ .
4. Se repiten los pasos 2 y 3 hasta la convergencia del método.

La elección de los centroides iniciales es relevante en el resultado final del método, por lo que existen diferentes formas de seleccionarlos. En este caso los centroides son series de tiempo que hay que determinar y uno de los aspectos a considerar es su longitud. Esta es recomendable (y razonable) que sea similar (o coincidente, si es posible) con la longitud de las series que se pretenden agrupar. En el caso de seleccionar los centroides de forma aleatoria, es un aspecto a considerar. La otra posibilidad más común es seleccionar como centroides algunas de las series que se pretende agrupar; en este caso ya no habría que preocuparse por la longitud de las series.

En las siguientes dos secciones se ponen en práctica los métodos presentados; tanto para la modelización de la serie a nivel nacional, como para la clasificación de las series por provincias y secciones económicas.

## 4.2. Modelado y predicción del número de accidentes

En el documento de *parquet* denominado *dfGEOGRAFICO* se encuentran para cada año, desde 2014 hasta 2021, el recuento de accidentes laborales en cada mes para cada CC.AA. y provincia. El recuento de accidentes mensuales para cada sector, sección y división económica se han almacena en el *dataset dfSECTORES*. En ambos casos se proporciona también el número de accidentes totales en España, sin considerar categorías.

Considerando solamente los accidentes mensuales totales, se construye la serie de los accidentes laborales mensuales desde enero de 2014 hasta diciembre de 2021 en España. El objetivo de esta sección es realizar un análisis de esta serie y obtener predicciones de los futuros meses. Todo esto se lleva a cabo a través de la modelización de la serie como un proceso ARIMA estacional. El procedimiento será llevado a cabo mediante el *software R* (R Core Team 2021). Concretamente se han usado las librerías *fpp2* (Hyndman 2020) , *tseries* (Trapletti 2022) y *TSA* (Chan and Ripley 2022) .

Una vez determinada la serie que se pretende analizar, se procede a su análisis. El primer paso consiste en representar la serie de forma gráfica. Este paso tan básico es fundamental para observar el comportamiento de la serie histórica y poder determinar el modelo que es apropiado ajustar.

En la Figura 4.1 se representa gráficamente la serie de tiempo. En el eje de abscisas se representa el avance de los años y en el eje de ordenadas se indica el número de accidentes correspondientes a los meses seleccionados para el estudio.

El número de accidentes correspondientes al primer año, 2014, se sitúan entre los 30000 y los 40000. Es notable como a medida que transcurre el tiempo estos valores van aumentando, llegando a rondar los 50000 accidentes en los meses del año 2019. Esta situación demuestra la existencia de una tendencia ascendente en la serie, lo cual implica la necesidad de diferenciación regular en su ajuste.

Continuando con las características de la serie, parece razonable suponer que presenta una estacionalidad de periodo  $s = 12$ . No era necesario observar la serie para asumir esta suposición; pues tiene sentido que en ella los resultados sean similares y tengan relación de año a año. Es decir, es razonable que los resultados de un mes concreto en un año tengan algún tipo de relación con los de ese mismo mes

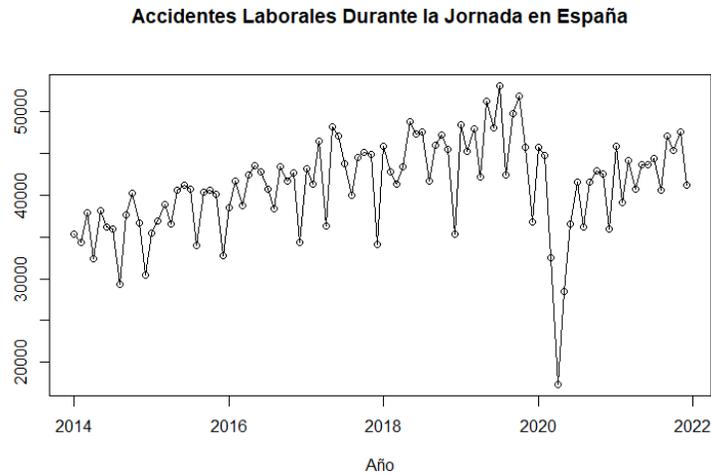


Figura 4.1: Representación de la serie de tiempo mensual del número de accidentes laborales en España.

en otros años. De todas formas, estas suposiciones es sensato contrastarlas mediante la representación de la serie. En este caso parece claro que cada 12 meses se repite un patrón similar. Siendo todavía más precisos, se puede apreciar que cada 4 meses se aprecia una bajada pronunciada del número de accidentes. Esto coincide en los meses de abril (mes 4), agosto (mes 8) y diciembre (mes 12), que se corresponden con los meses en donde la mayoría de trabajadores disfrutan de sus vacaciones, lo cual hace que el número de personas que se encuentran en su puesto de trabajo disminuya. Teniendo esto en cuenta, se podría considerar un modelo con una estacionalidad de periodo  $s = 4$ , aunque se muestra posteriormente en este trabajo que no proporciona un mejor ajuste que con  $s = 12$ .

Finalmente, cabe destacar la gran caída que tiene lugar a principios del año 2020, concretamente en los meses de marzo y abril. Esta caída coincide o, más bien, es provocada por la pandemia del COVID-19, que trajo consigo el confinamiento durante varios meses del año y la suspensión de múltiples contratos de trabajo mediante los ERTES. A pesar de que la caída tiene lugar principalmente en el mes de abril de 2020, el efecto de esta perdura en el tiempo afectando al nivel de la serie. Esto resulta claro, teniendo en cuenta que la serie presentaba un nivel ascendente hasta finales de 2019, alcanzando valores superiores a 50000 accidentes, y a partir del evento en abril de 2020, el nivel de la serie observada hasta ahora cae por debajo de los 45000 accidentes. Esta situación se puede (y debe) incluir en el ajuste de la serie mediante la inclusión de una intervención en el modelo. Concretamente, se ha considerado una intervención que tiene lugar en abril de 2020 y cuyo efecto se va diluyendo, aunque afectando a los meses posteriores en la serie.

Además de todas estas consideraciones básicas para ajustar el modelo, también se aplica la función logaritmo a la serie con la que se trabaja; ya que a pesar de no presentar problemas severos de heterocedasticidad, es muy común trabajar con la serie transformada y no afecta negativamente a los resultados.

Se ajusta por tanto un modelo  $ARIMA(p, d, q) \times (P, D, Q)_s$  con intervención. De esta forma se garantiza que se considere en el modelo tanto la tendencia ascendente como la estacionalidad, además de la caída en 2020. Nótese que no se ha especificado qué valor de  $s$  se considera, ya que se analizará tanto la posibilidad de que  $s = 12$  como de que  $s = 4$ . De todas formas, solo se documenta el proceso de construcción de la serie en el primer caso, pues el segundo es análogo. Simplemente, se muestran posteriormente los resultados de ambas, concretando cuál de ellas es más apropiada.

Además de decidir el valor de  $s$ , para determinar el modelo completamente es necesario seleccionar el valor de los parámetros  $p, d, q, P, D$  y  $Q$ . En el caso de los procesos ARMA, esto se puede realizar de una forma sencilla observando las distintas características de la serie, pero en este caso es necesario recurrir a la función *auto.arima* de *R*. Esta función solo proporciona los valores de  $p, q, P$  y  $Q$ , los de

$d$  y  $D$  deben ser indicados como argumentos de la función. Para conocer los valores más apropiados se procede como sigue:

- Con el propósito de obtener el valor de  $d$ , se representa la serie original (en este caso modificada mediante el logaritmo) diferenciada. En el caso de que diferenciando la serie una vez desaparezca la tendencia, entonces  $d = 1$ ; si no, se continúa diferenciando hasta que esto ocurra.

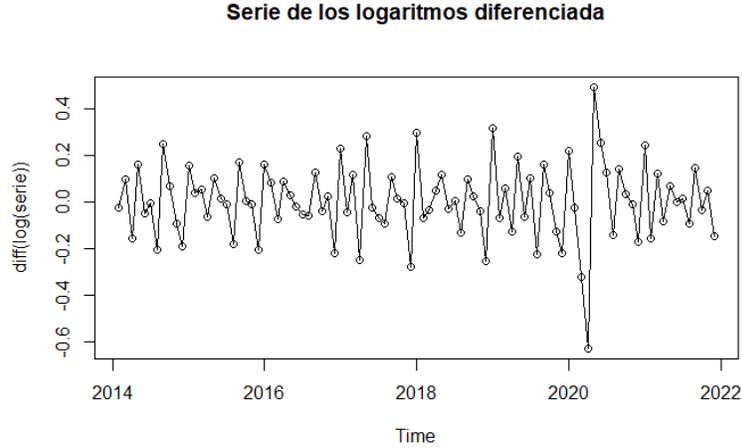


Figura 4.2: Serie mensual de los accidentes laborales en España diferenciada regularmente.

Con una sola diferenciación se obtiene una serie estacionaria, con un nivel estable; por tanto, se tomará  $d = 1$ .

- En cuanto el valor de  $D$ , como se puede ver en Coperwait and Metcalfe (2009), en general se toma  $D = 1$  y eso es lo que se ha aplicado aquí también. Se comenta además que los valores de  $P$  y  $Q$  suelen ser menores que 3.

Teniendo esto en cuenta, se obtiene con la función *auto.arima* el modelo  $ARIMA(1, 1, 1) \times (0, 1, 1)_{12}$ .

Una vez conocido el modelo que se toma en el ajuste, es momento de considerar la intervención que tiene lugar en la serie. Para ello se ha utilizado la función *arimax* de la librería *TSA* y se ha incluido en el ajuste la intervención. Además, se ha llevado a cabo un estudio de los datos atípicos y se ha incluido marzo de 2020 como tal. Observando la gráfica de la serie en la Figura 4.1 ya se podía haber intuido que había que incluir este valor de alguna forma en la modelización de la serie. Se podía haber considerado que la intervención tenía lugar en ese punto, que fue realmente en el mes en el que comenzó la cuarentena. Sin embargo, dado que la gran caída de los datos se presentaba en el mes siguiente, se consideró que era más apropiado considerar el evento en el mes de abril (siendo este el septuagésimo sexto mes de la serie).

Se tiene entonces que la serie  $\{X_t\}_t$  de accidentes laborales mensuales se ajusta mediante un modelo de la forma

$$\Phi(B)\tilde{\Phi}(B^{12})(1-B)(1-B^{12})\log(Y_t) = \Theta(B)\tilde{\Theta}(B^{12})a_t,$$

siendo

$$\log(Y_t) = \omega \sum_{j=0}^{\infty} \delta^j I_t(76+j) + \log(X_t).$$

Nótese que se considera el logaritmo neperiano en la fórmula, pues, como se ha indicado anteriormente, se aplica esta transformación para reducir la posible heterocedasticidad y trabajar con menores escalas.

Las funciones del modelo son en este caso

- $\Phi(z) = 1 - \phi_1 z$

- $\tilde{\Phi}(z) = 1$
- $\Theta(z) = 1 + \theta_1 z$
- $\tilde{\Theta}(z) = 1 + \tilde{\theta}_1 z$

y los parámetros estimados:

$$\hat{\phi}_1 = -0.2646, \hat{\theta}_1 = -0.6703, \hat{\tilde{\theta}}_1 = -0.9370, \hat{\omega} = -0.9064, \hat{\delta} = 0.4668,$$

que en este caso son todos significativos. Además, el modelo presenta otro parámetro, con valor  $IO = -0.3899397$ , que modeliza el dato atípico considerado en marzo de 2020 (indicando que existe una bajada).

Finalmente, tras el ajuste del modelo, se realiza una diagnosis de este para comprobar que es válido. En este caso, los resultados obtenidos son bastante aceptables: la serie diferenciada no presenta apenas dependencia y las innovaciones  $\{a_t\}_t$  son gaussianas.

Tras ajustar el modelo y teniendo en cuenta los parámetros estimados, se presenta a continuación la interpretación de la intervención en el modelo. Considerando en primer lugar la escala logarítmica, se tiene lo siguiente:

$$\log(Y_t) = -0.9064 \sum_{j=0}^{\infty} 0.4668^j I_t(76 + j) + \log(X_t).$$

Por tanto, en el instante  $t = 76$  (que se corresponde con abril de 2020) el número de accidentes se reduce en 0.9064 unidades logarítmicas. De forma análoga, cuando hayan pasado  $j$  meses del evento, la reducción en el número de accidentes respecto a la log-serie no intervenida es  $0.9064 \times (0.4668)^j$  unidades logarítmicas. Se puede ver que en los primeros meses los descensos (no adicionales) que tuvieron lugar fueron:

- mes de abril 2020: 0.9064 unidades logarítmicas.
- mes de mayo 2020: 0.4231 unidades logarítmicas.
- mes de junio 2020: 0.1975 unidades logarítmicas.
- mes de julio 2020: 0.0922 unidades logarítmicas.

A pesar de que el modelo se haya ajustado considerando la transformación logaritmo neperiano, a la hora de interpretar los resultados la escala logarítmica puede no ser totalmente clara. A continuación se muestran los resultados a escala real. En este caso:

$$Y_t = \exp \left\{ -0.9064 \sum_{j=0}^{\infty} 0.4668^j I_t(76 + j) + \log(X_t) \right\} = \exp \left\{ -0.9064 \sum_{j=0}^{\infty} 0.4668^j I_t(76 + j) \right\} X_t.$$

Con lo cual, el coeficiente  $\exp \left\{ -0.9064 \sum_{j=0}^{\infty} 0.4668^j I_t(76 + j) \right\}$  designa la proporción de accidentes que ocurren en el mes  $j$ -ésimo tras la intervención. Por otro lado,

$$100 \times \left( 1 - \exp \left\{ -0.9064 \sum_{j=0}^{\infty} 0.4668^j I_t(76 + j) \right\} \right)$$

representa el tanto por ciento de accidentes que no tuvieron lugar en la serie intervenida de los que sí se esperaban en la original. Se tiene entonces que los descensos que tuvieron lugar en el mes del evento y los inmediatamente posteriores son:

- mes de abril 2020: 59.6034 %.
- mes de mayo 2020: 34.50091 %.
- mes de junio 2020: 17.92411 %.
- mes de julio 2020: 8.808482 %.

Por ejemplo, si en abril de 2020 en circunstancias normales hubiesen tenido lugar 40000 accidentes, considerando la intervención, el modelo ajustado indicaría que solamente ocurrieron 16158.64 accidentes.

Nótese que, a medida que van avanzando los meses, el efecto de la intervención en el modelo se va desvaneciendo hasta que, finalmente, el efecto que tenga sobre la serie sea casi nulo. En la Figura 4.1 se puede visualizar como esa situación es realmente lo que ocurre en la serie de valores observados.

Antes de proceder con las predicciones, se muestran en la Figura 4.3 las comparaciones entre los valores reales de la serie y los valores ajustados.

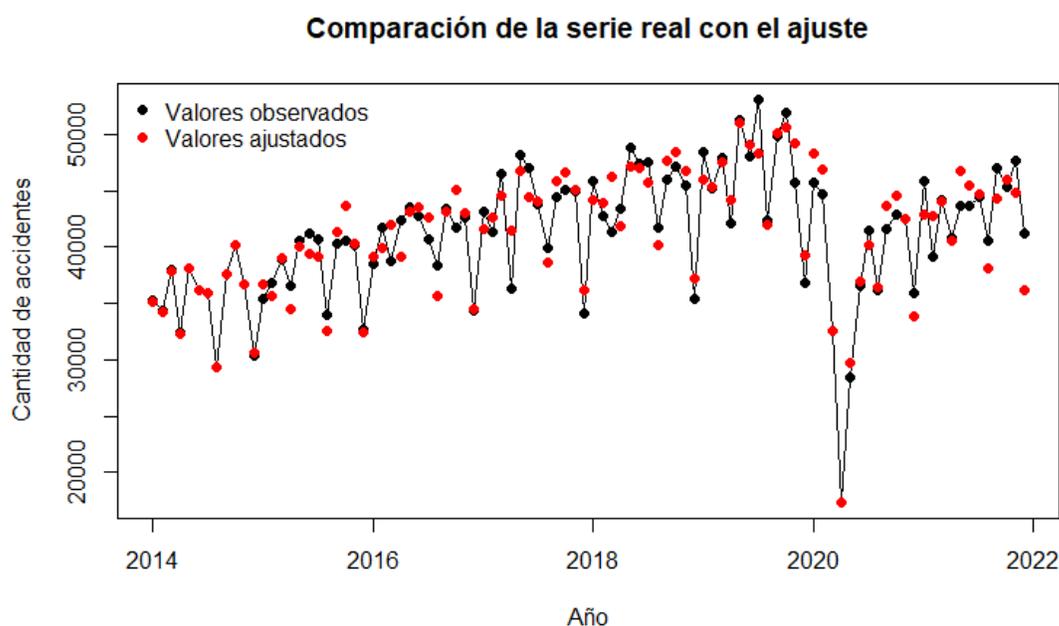


Figura 4.3: Representación de la serie mensual de los accidentes laborales en España junto al ajuste obtenido.

Nótese que los doce primeros valores de la serie coinciden, pues al considerar un modelo con un periodo de estacionalidad igual a  $s = 12$ , los 12 primeros valores de la serie presentan residuos artificiales (que se toman igual a 0).

En cuanto al resto de los meses, lógicamente, hay meses ajustados de forma más precisa que otros, pero en general parece un buen ajuste, incluso después de la intervención. Aun así, cabe destacar que varias de las caídas más pronunciadas después de la intervención las infra-ajusta, proporcionando un valor más bajo del observado.

Finalizado el análisis del ajuste, se muestran a continuación las predicciones de los 12 meses del año 2022, el siguiente a los años para los que se ha realizado el análisis. En la Figura 4.4 se puede ver la serie de los valores históricos junto con las nuevas predicciones, las cuales continúan la tendencia creciente de la serie. Los valores predichos para cada uno de los meses son

- Enero: 47769.00
- Febrero: 46294.61
- Marzo: 45106.56
- Abril: 44828.05
- Mayo: 49719.22
- Junio: 49214.41
- Julio: 49962.20
- Agosto: 43306.07
- Septiembre: 49771.88
- Octubre: 50021.42
- Noviembre: 49230.08
- Diciembre: 40582.63

Los meses en los que se predice un menor número de accidentes son abril, agosto y diciembre, los meses que a lo largo de la serie observada ya presentaban valores bajos en comparación con el resto de meses. El resto de valores parece seguir también un patrón similar al del resto de la serie, como se puede ver en la Figura 4.4.

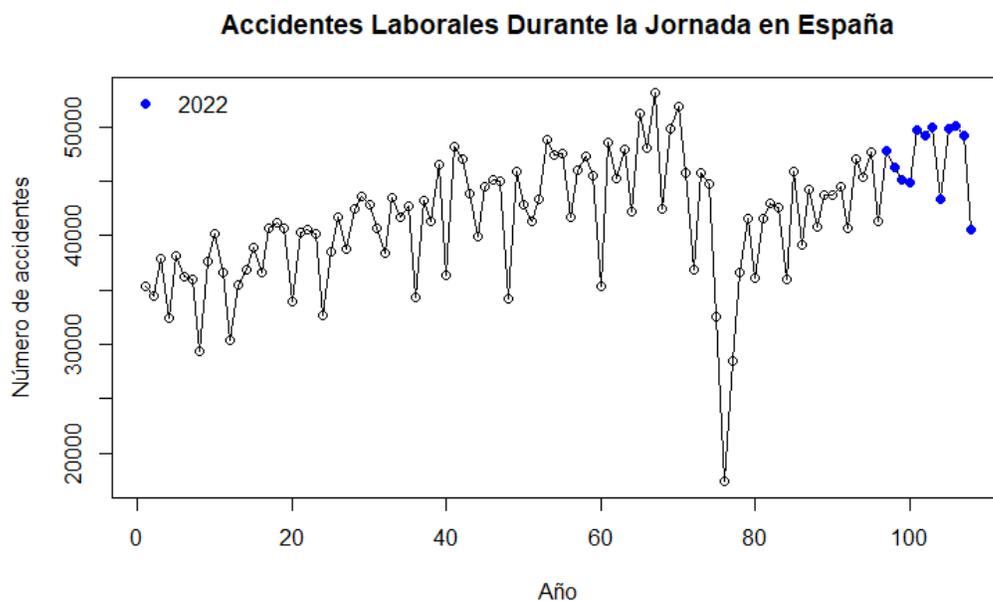


Figura 4.4: Representación de la serie mensual de los accidentes laborales en España junto con las predicciones de los 12 próximos meses.

Al principio del apartado se consideró la posibilidad de ajustar un modelo considerando  $s = 4$  como periodo de estacionalidad. La gráfica anterior no parece contradecir esta suposición de forma contundente, por lo que podría ser interesante barajar esta nueva opción. Se ha ajustado, por tanto, un modelo de estas características, repitiendo un proceso análogo al ya mostrado.

El nuevo ajuste considerado sigue un modelo  $ARIMA(0, 1, 1) \times (2, 1, 0)_4$  con los parámetros estimados:

$$\hat{\theta}_1 = -0.9368, \hat{\phi}_1 = -0.7889, \hat{\phi}_2 = -0.5713, \hat{\omega} = -0.8417, \hat{\delta} = 0.5868,$$

y como valor asociado al dato atípico en marzo de 2020,  $-0.4231$ . Todos los coeficientes son significativos y la diagnosis del modelo es medianamente razonable.

En las siguientes gráficas se representa tanto el ajuste del modelo sobre la serie real como las predicciones del modelo. En ambos casos se muestra en comparación con los resultados obtenidos con el modelo anterior.

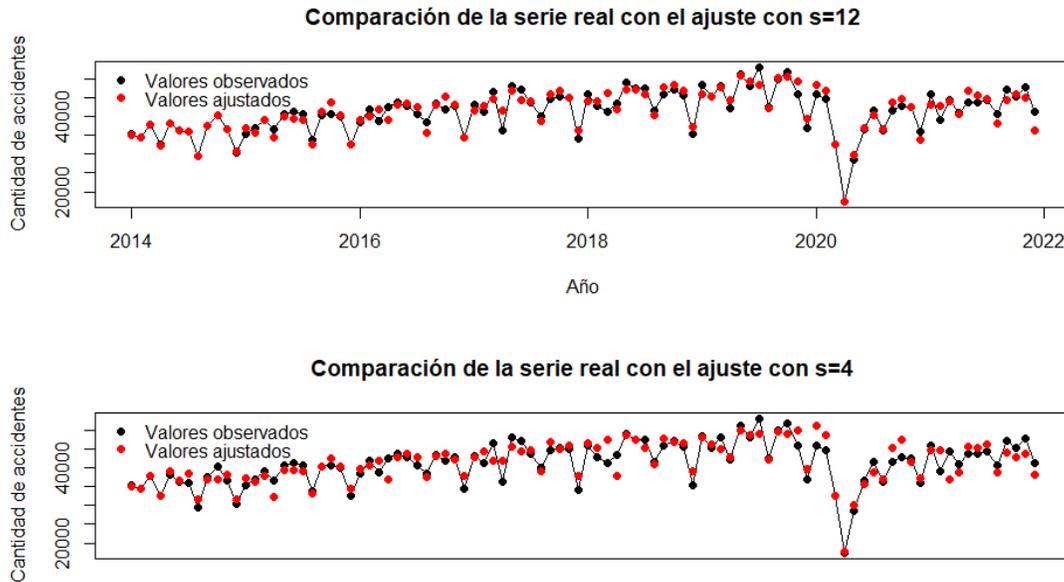


Figura 4.5: Representación de la serie mensual de los accidentes laborales en España junto a los ajustes obtenidos.

En la Figura 4.5 no se aprecia una especial diferencia entre los ajustes, por lo tanto, no se podría discriminar en función de esta gráfica cuál de los ajustes sería más adecuado.

En cuanto a la Figura 4.6, sí que parece que el primero de los ajustes se acerca más al patrón real de la serie; pues la segunda gráfica presenta un patrón mucho más repetitivo, que se aleja un poco del observado para los años anteriores.

De todas formas, para determinar realmente cuál de los ajustes es más correcto se analiza de forma numérica, estudiando qué ajuste presenta un menor error cuadrático medio y un mejor valor según el criterio de información de Akaike. Dado que el número de parámetros de los modelos no suele coincidir en general, es más correcto evaluarlos mediante el criterio corregido de Akaike, que valora este aspecto:

- Valor del AICC del modelo con  $s = 12$ : -214.2778
- Valor del AICC del modelo con  $s = 4$ : -203.2594

El valor del AICC se calcula como

$$AICC(\text{ajuste}) = -2\log\text{verosimilitud}(\text{ajuste}) + (\log(n) + 1) \times k,$$

siendo  $n$  el número de elementos de la serie y  $k$  el número de coeficientes del modelo. En Johnson y Wichern (2007) se profundiza en este aspecto.

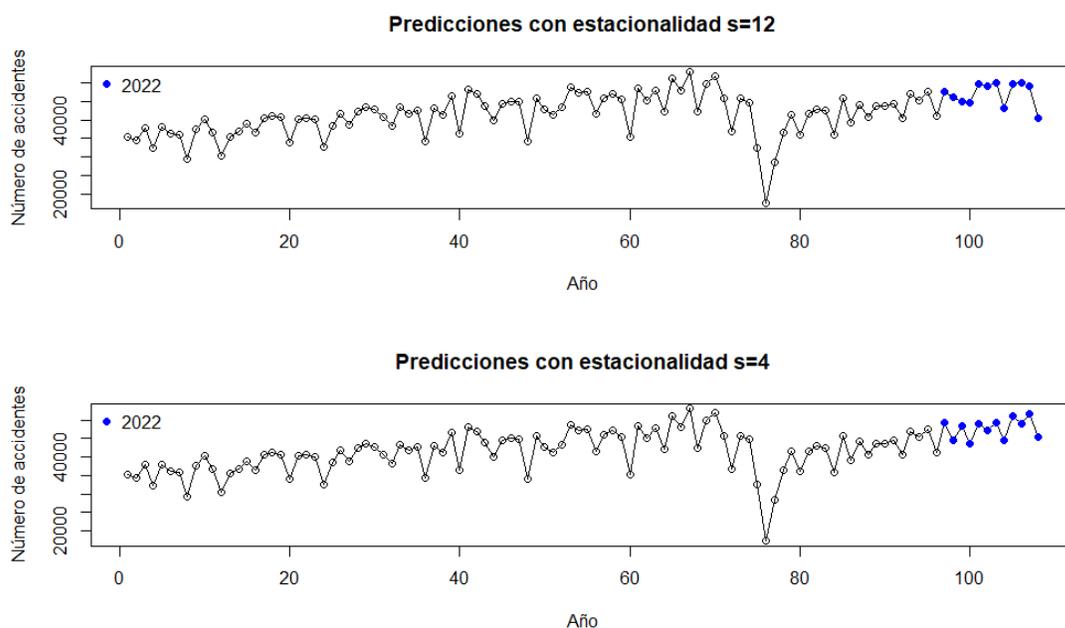


Figura 4.6: Predicciones para los meses de 2022 según los dos modelos ajustados.

Dado que el AICC del modelo con  $s = 12$  es menor que el del modelo con  $s = 4$  (por una diferencia mayor a 2 unidades), resulta claro que el primero de los modelos es el que mejor ajusta los datos analizados y el que se debe considerar para modelizar la serie.

En cuanto al error cuadrático medio, los resultados avalan la conclusión anterior:

- RMSE del modelo con  $s = 12$ : 1881.347
- RMSE del modelo con  $s = 4$ : 2549.917.

El primero de los modelos comete menos error en el ajuste. De todas formas, en la Figura 4.7 se muestra la comparación de las predicciones obtenidas mediante los dos modelos con los valores de los avances proporcionados por el Ministerio de Trabajo para los 10 primeros meses de 2022.

Se representan para ambos modelos las predicciones obtenidas a través de los ajustes. Sobre ellos se han indicado en color rojo los valores avanzados que se han publicado en la web del Ministerio<sup>2</sup>. Cabe destacar que estos datos no son del todo precisos, pues todavía no son los datos finales oficiales. De todas formas sí son unos valores que se aproximarán a los reales.

Llama la atención el valor tan bajo que presenta el valor avanzado de enero de 2022. Puede deberse a que es un dato atípico o a que, al tratarse de un avance y no los datos oficiales, pueda haber algún error de medición. Esta situación hará que el error en la predicción sea mayor; pero ocurre para cualquier ajuste, por lo que la comparación de los errores de predicción seguirá siendo válida.

En cuanto al resto de valores de las predicciones, parece que se ajustan mejor en ambas series. Y aunque el mes de marzo de 2022 parece que se predice mejor con el segundo modelo, para el resto de meses es el primer ajuste el que parece más acertado. De todas formas, la precisión de las predicciones se puede cuantificar mediante el RMSE:

- RMSE de las predicciones del modelo con  $s = 12$ : 4848.643
- RMSE de las predicciones del modelo con  $s = 4$ : 4974.751.

<sup>2</sup>Ministerio de Trabajo. última vez visitado el 3 de enero de 2023 [online](#).

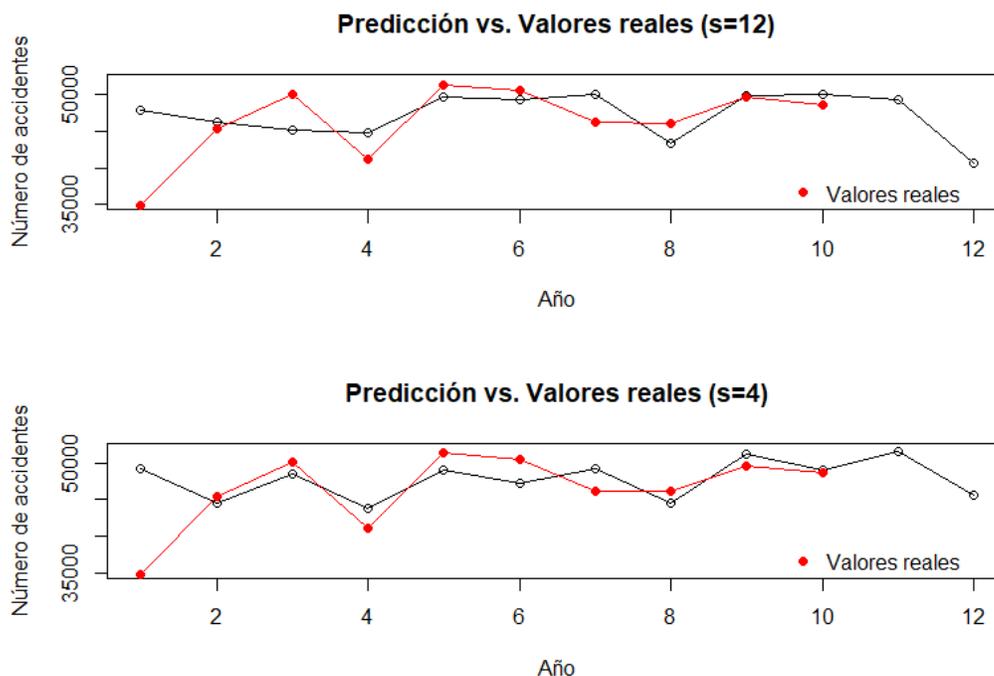


Figura 4.7: Predicciones para los meses de 2022 según los dos modelos ajustados.

Nuevamente, los resultados son mejores considerando el primero de los ajustes. En este caso, a pesar de que el mes de marzo presenta unos peores resultados, es en general mejor para predecir los nuevos valores de la serie, al igual que lo era para ajustarlos.

A pesar de que los resultados son más favorables al primero de los ajustes, en este caso la diferencia es menos notable que en las comparaciones anteriores. Aun así, considerando también esos resultados previos, se puede concluir que el modelo estacional  $ARIMA(1, 1, 1) \times (0, 1, 1)_{12}$  es el mejor de los considerados para modelar los datos y que en general presenta un ajuste bastante razonable.

### 4.3. Agrupación de las series por provincias y sectores económicos

Continuando con el tratamiento de series temporales, en esta última sección se pone en práctica un método de clusterización sobre las series de tiempo de las diferentes provincias españolas y secciones de actividad económica, con la intención de agruparlas según sus similitudes en cuanto a la distribución de los accidentes laborales.

Antes de llevar a cabo el agrupamiento de las series, se realizan también otros procedimientos con el propósito de entender los datos. En primer lugar, se presentan mapas de calor que muestran la cantidad de accidentes en las distintas provincias y secciones económicas a lo largo de los meses. No se ha realizado la agrupación a partir de estos gráficos, pero sí proporcionan una idea sobre qué se puede esperar de los datos. Posteriormente, se lleva a cabo el agrupamiento de las series de tiempo mediante un método jerárquico. Los resultados que se obtengan no serán los definitivos, pero son de utilidad para el procedimiento final. Este consiste en aplicar el método de *clustering* DBA a las series temporales consideradas. Los resultados del método jerárquico son útiles para este otro, pues facilita

la selección del número de *clústers* que se deben considerar para agrupar las series.

En las siguientes gráficas se muestran los valores de las series temporales de cada una de las provincias y ciudades autónomas españolas. Se muestran a través de un *heatmap*, por lo que se mostrará de colores claros aquellos meses en los que las provincias han presentado un alto número de accidentes y en colores oscuros, los meses con pocos accidentes.

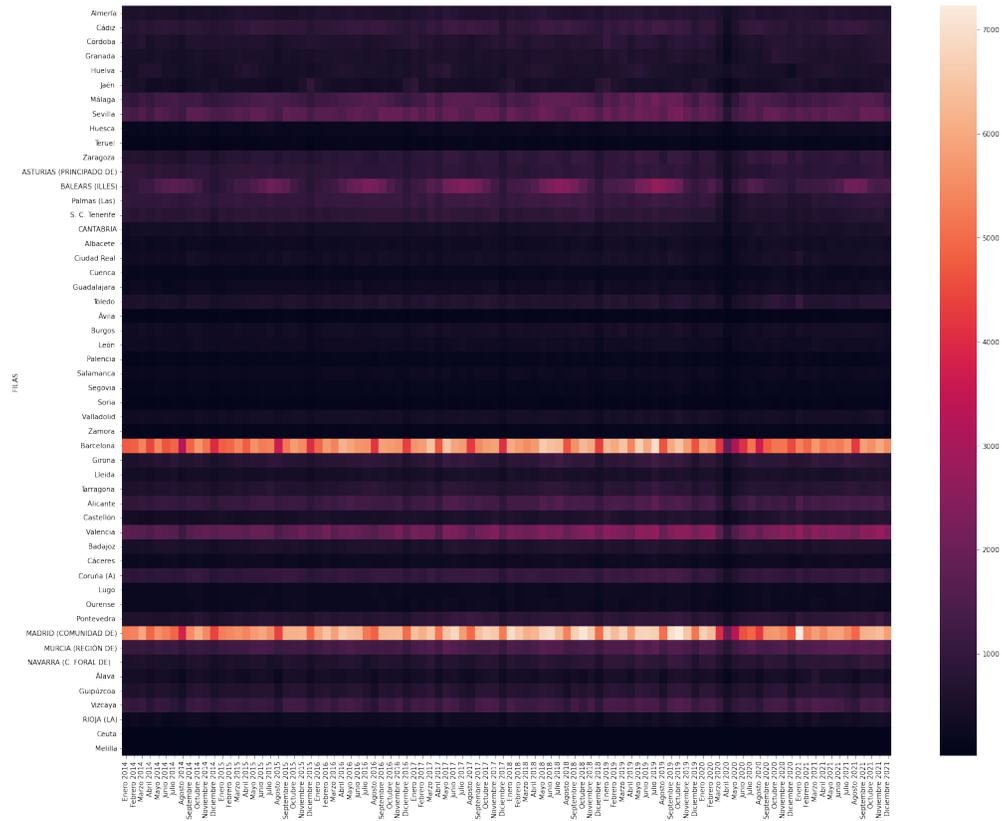


Figura 4.8: *Heatmap* que muestra el número de accidentes que tuvieron lugar en cada provincia entre 2014 y 2021.

Tal y como se puede visualizar en la Figura 4.8, hay dos provincias que presentan un muy alto número de accidentes, representadas en colores claros, y las demás provincias tienen colores oscuros asociados. De aquí se puede deducir que las provincias de Barcelona y Madrid son las que presentan un mayor número de accidentes en general, destacando muy notablemente sobre las demás. Algunas provincias como Sevilla, las Islas Baleares, Alicante, Vizcaya o, sobre todo, Valencia presentan un volumen mayor de accidentes a lo largo de la serie que las demás provincias, pero aun así están bastante distantes de Madrid y Barcelona.

Estos resultados no son en realidad sorprendentes, pues el número de población ocupada y de población en general en estas provincias está muy por encima de las demás; lo cual da una razón de peso para explicar la gran diferencia existente en número de accidentes entre las provincias. En cuanto a población en los últimos 8 años, según el INE<sup>3</sup>, Barcelona y Madrid presentan en torno a 6 y 7 millones de habitantes, respectivamente. Sin embargo, la siguiente provincia más poblada es Valencia, con en torno a 2.5 millones de habitantes, menos de la mitad de las dos provincias anteriores. En cuanto a la población ocupada, los resultados son similares. Madrid supera los 3 millones de ocupados en 2021

<sup>3</sup>Población por provincias según el INE. Última vez consultado el 4 de enero de 2023 [online](#).



En esta nueva gráfica ya se pueden apreciar muchos más matices que en la anterior. Por ejemplo, de forma general se puede observar como abril de 2020 fue para todas las provincias el mes con menor número de accidentes de todos los considerados. Un resultado que coincide con todos los obtenidos hasta ahora. Además, también destaca de forma bastante clara que, en muchas de las provincias, los meses de 2019 son los que presentaron un mayor número de accidentes, de acuerdo con los resultados vistos en el capítulo 3.

En cuanto a resultados más concretos, a simple vista es complejo diferenciar tendencias similares, pero a continuación se comentan algunas de las más destacables. Las provincias Islas Baleares, Cádiz y Castellón presentan un comportamiento claramente cíclico, mostrando un pico de accidentes en los meses de verano. Esto resulta razonable, pues son provincias que en verano reciben muchos turistas y la población ocupada aumenta con motivo, sobre todo, de reforzar el sector de la hostelería y restauración. Otro ejemplo de tendencias similares la muestran Asturias y León, que presentan un alto número de accidentes en general, sin mostrar bajadas destacadas en el nivel de accidentes, salvo la provocada por el COVID-19. Una situación similar se puede ver en las dos provincias canarias; pero en este caso sí que tiene lugar una bajada en el número de accidentes laborales a lo largo de 2020 y principios de 2021. Esta situación seguramente esté relacionada con la bajada del turismo tras la pandemia, lo cual trajo consigo menos visitas a las islas y menor actividad en el sector turismo, principal en las Islas Canarias. Con la clusterización de las series se pretenden obtener conclusiones similares a las anteriores.

Para complementar la información dada, en *Datapane* están disponibles dos mapas interactivos (uno desde 2014 a 2017 y otro de 2018 a 2021) en los que se puede visualizar el número de accidentes para cada provincia en cada mes a lo largo de los 8 años considerados.

De forma totalmente análoga a las series temporales de las provincias, se presenta a continuación el mapa de calor asociado a las secciones económicas consideradas en los archivos del Ministerio de Trabajo.

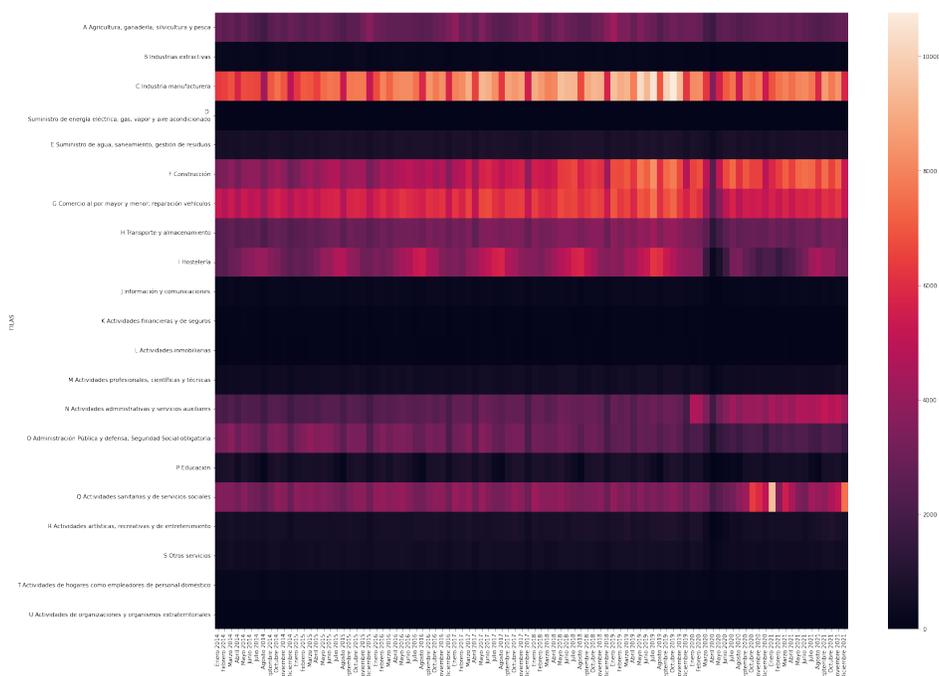


Figura 4.10: *Heatmap* que muestra el número de accidentes que tuvieron lugar en cada sección económica entre 2014 y 2021.

La Figura 4.10 presenta el mismo fenómeno que se visualizaba en la Figura 4.8: la diferencia tan

destacable entre el número de accidentes de las diferentes series hace que no se puedan apreciar las tendencias de muchas de ellas. En este caso, la construcción, el comercio al por mayor y menor y la reparación de vehículos, junto con sobre todo las industrias manufactureras representan una gran parte de la cantidad de accidentes laborales. Para resolver el inconveniente, se aplica la misma solución que anteriormente. En la Figura 4.11 se puede visualizar de nuevo el mapa de calor para los accidentes mensuales de cada una de las secciones económicas, pero esta vez estandarizando las series.

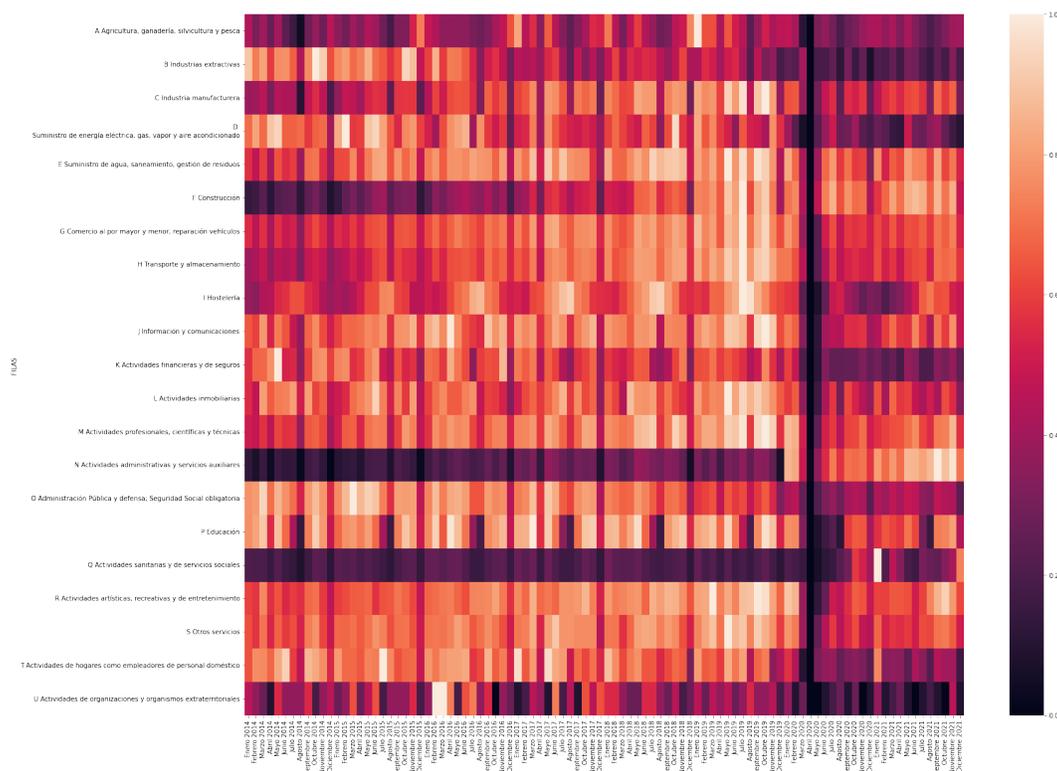


Figura 4.11: *Heatmap* que muestra el número de accidentes que tuvieron lugar en cada sección económica entre 2014 y 2021.

En la nueva gráfica ya sí se pueden observar de forma más clara las tendencias de cada una de las series. Algunas que se puede observar a simple vista son, por ejemplo, que la industria manufacturera y la construcción han ido aumentando su número de accidentes conforme avanzan los años o que, por ejemplo, las industrias extractivas y el suministro de energías han ido disminuyendo su número de accidentes. También resultan bastante llamativas las series correspondientes a la hostelería y la educación; la primera muestra picos de accidentes en los meses de verano, mientras que para la serie de educación son justamente esos meses donde se produce la bajada de accidentes (especialmente en agosto).

Los resultados vistos hasta ahora no solo han sido útiles para introducir el procedimiento principal de este capítulo, sino que ha sido una muestra muy clara de la necesidad de la estandarización de las series para poder agruparlas de forma razonable. De no hacerlo, los grupos se formarían en función del número de accidentes de las provincias/sectores económicos; que sería una clasificación válida, pero no es lo que se pretende obtener en este caso. La idea del procedimiento que se incluye a continuación es encontrar relaciones entre categorías en cuanto a la temporalidad de los accidentes laborales.

En primera instancia se procederá con la clasificación de las provincias, posteriormente con la de los

sectores económicos y, finalmente, combinando ambas categorías para estudiar las relaciones existentes entre provincias y sectores en cuanto a la temporalidad de los accidentes laborales.

### 4.3.1. Agrupación de las provincias

En esta sección se lleva a cabo la agrupación de las series mensuales de cada provincia. El agrupamiento de las series se realiza mediante el método DBA introducido anteriormente. Para poder poner en práctica este método es necesario tener en mente el número de grupos en los que se quieren dividir los elementos. Para resolver esta cuestión, se realiza primero una agrupación de las series mediante un método de *clustering* jerárquico. De esta forma, se puede observar para qué distancias se unen los grupos y decidir en función de eso cuantos grupos es razonable considerar. Nótese que, al aplicar posteriormente el algoritmo DBA, la distancia que se considera viene dada por la métrica DWT.

En primer lugar, se construye la matriz de distancias entre las diferentes series que se utiliza en el método jerárquico, considerando la distancia DWT sin imponer restricciones al radio de Sakoe-Shiba. A partir de esta matriz, se ha procedido con el método de agrupamiento. El árbol que muestra la clasificación de las series mensuales de las provincias se puede observar en la Figura 4.12.

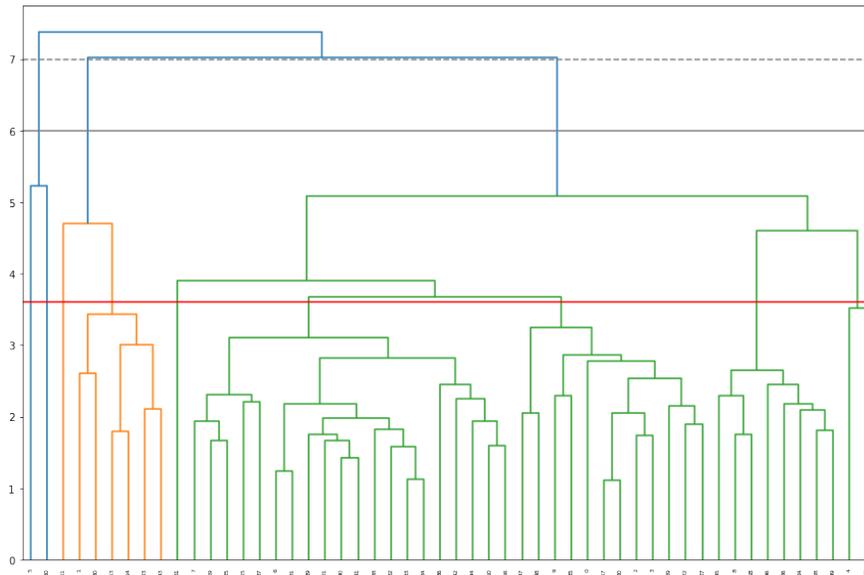


Figura 4.12: Agrupamiento mediante el método jerárquico de las series mensuales de las provincias.

En vista del resultado presentado en la gráfica anterior, parece razonable considerar el número de grupos que se generan trazando una recta para una distancia igual a 3.6. De esta forma se consideran las series temporales divididas en 9 grupos diferentes. Considerar menos grupos daría lugar a que provincias con series poco similares se agrupasen juntas. Por otro lado, considerando más grupos, estos contendrían un menor número de series (en algunos casos solamente una) por lo que las conclusiones que se podrían extraer serían más pobres.

Tras obtener este resultado, se puede proceder ya por fin con el método DBA para agrupar las series. Dado que con el método anterior se ha considerado apropiado seleccionar 9 *clústers*, se muestran a continuación los 9 grupos obtenidos mediante la aplicación del método basado en centroides:

- **GRUPO 1:** Almería, Zaragoza, Pontevedra, Murcia (Región de).
- **GRUPO 2:** Segovia, Valladolid, Navarra (C. Foral de).

- **GRUPO 3:** Sevilla, Asturias (Principado de), Palmas (Las), S. C. Tenerife, Guadalajara, León, Salamanca, Barcelona, Madrid (Comunidad de), Ceuta, Melilla.
- **GRUPO 4:** Huelva, Balears (Illes).
- **GRUPO 5:** Granada, Huesca, Ciudad Real, Cuenca, Toledo, Palencia, Álava, Rioja (La).
- **GRUPO 6:** Cádiz, Córdoba, Málaga, Ávila, Badajoz, Coruña (A), Lugo, Ourense
- **GRUPO 7:** Burgos, Soria, Guipúzcoa, Vizcaya.
- **GRUPO 8:** Teruel, Cantabria, Albacete, Zamora, Girona, Lleida, Tarragona, Alicante, Castellón, Valencia, Cáceres.
- **GRUPO 9:** Jaén.

*A priori* algunos de los grupos pueden parecer combinaciones extrañas, pero a continuación se analizará cada uno de ellos.

En el primero, dada la proximidad geográfica entre Almería y Murcia, podía esperarse que guardasen cierta relación. En cuanto a Pontevedra, podría relacionarse con estas dos por motivos relacionados con que las tres poseen costa. Sin embargo, con Zaragoza no se observa de forma sencilla una conexión. De todas formas, todas presentan similitudes en la serie temporal de accidentes laborales. En este caso, todas las provincias del grupo muestran un aumento progresivo en el número de accidentes, sobre todo hasta abril de 2020.

El segundo de los grupos es similar al primero en el aspecto de que la tendencia que siguen sus provincias es creciente. En este grupo el crecimiento parece más suave. En cuanto a la relación entre las componentes del grupo, Segovia y Valladolid son provincias de interior limítrofes; Navarra es también una provincia de interior, pero se encuentra algo distanciada de las anteriores.

Este tercer grupo es un de los que más provincias agrupa, considerando entre ellas a las que más accidentes registran (Madrid y Barcelona) y a las que menos (las ciudades autónomas). Estas provincias tienen en común que presentan una tendencia bastante monótona en cuanto a número de accidentes: en la Figura 4.9 se puede ver como sus filas son prácticamente todas de colores claros (meses con alto número de accidentes dentro de los números de cada provincia) salvo por los meses más próximos a abril de 2020. En el caso de las dos provincias canarias, esas tonalidades oscuras se alargan más allá de 2020; probablemente serían las primeras provincias en escindirse a un nuevo grupo si se considerasen más *clústers*.

El cuarto grupo considera la provincia andaluza Huelva (costera) y las Islas Baleares. Ambas presentan resultados cíclicos: Huelva presenta mayor número de accidentes en los meses de primavera, mientras que las Islas Baleares presentan un mayor número de accidentes en la temporada alta del turismo, en verano. Este grupo es un claro ejemplo de como funciona la distancia DWT y el algoritmo DBA, pues las dos series presentan un comportamiento similar pero para épocas del año diferentes. Esto es posible captarlo gracias a la flexibilidad que permite el algoritmo.

El siguiente de los grupos a analizar es el grupo 5. Este grupo está compuesto mayoritariamente por provincias de interior (solo Granada tiene frontera con el mar), por lo que puede existir una conexión entre las provincias por este motivo. En cuanto a la tendencia que presentan, es bastante plana. Sí que muestran un ligero crecimiento a medida que avanzan los años, pero predominan sobre todo en el *heatmap* de la Figura 4.9 los colores medios de la paleta para las filas de estas provincias. En la mayor parte de las provincias que pertenecen a este grupo, los valores más altos de su serie se encuentran tras la pandemia del COVID-19.

El sexto grupo considera 3 provincias gallegas (todas menos Pontevedra), 3 andaluzas (las costeras, Cádiz y Málaga y la interior, Córdoba), Ávila y Badajoz. Las provincias gallegas, Badajoz y Cádiz presentan series extremadamente similares: con tendencia creciente hasta abril de 2020 y con valores superiores al 0.4 en los casos más bajos. Las provincias restantes también muestran una tendencia creciente, pero con valores ligeramente más bajos de número de accidentes.

El grupo 7 contiene a Burgos, Soria, Guipúzcoa y Vizcaya. Todas bastante próximas, aunque las dos primeras interiores y las segundas, costeras. En cuanto a las similitudes entre sus series destaca que se asemeja bastante a las conclusiones generales del país: presentan una tendencia creciente, con picos más bajos en los meses de diciembre-enero, abril y agosto.

El octavo grupo considera las tres provincias valencianas, las tres catalanas, aparte de Barcelona, Teruel, Albacete, Cantabria, Zamora y Cáceres. Las ocho primeras están cerca geográficamente y las tres restantes están más próximas entre sí (sin estarlo demasiado). La similitud más destacable entre todas las series vuelve a ser de nuevo la presencia de una tendencia creciente parecida en todas las provincias.

Finalmente, solo queda considerar el más sencillo de los grupos, que considera únicamente la provincia de Jaén. En su serie se observan subidas en el número de accidentes en los meses de invierno, especialmente diciembre y enero. Sigue una tendencia cíclica similar a la del grupo 4. De hecho, haciendo la comprobación y construyendo ocho grupos en lugar de nueve, este último se une al grupo 4, formando con Huelva y las Islas Baleares el grupo de las provincias con series cíclicas.

En general, todos los grupos presentan una característica común, que es el crecimiento del número de accidentes a medida que van pasando los años. Esto es lógico, pues como se menciona al principio del informe, el número de accidentes en el país ha estado creciendo en los últimos años. Es por tanto lógico, que en las provincias españolas se observe una tendencia de crecimiento general. De todas formas, este crecimiento no ocurre de igual forma en todas las provincias y es ahí donde se encuentran las principales diferencias entre los grupos.

En las Figuras 4.13 y 4.14 se presentan las series temporales de los 9 grupos junto con el centroide de cada uno.

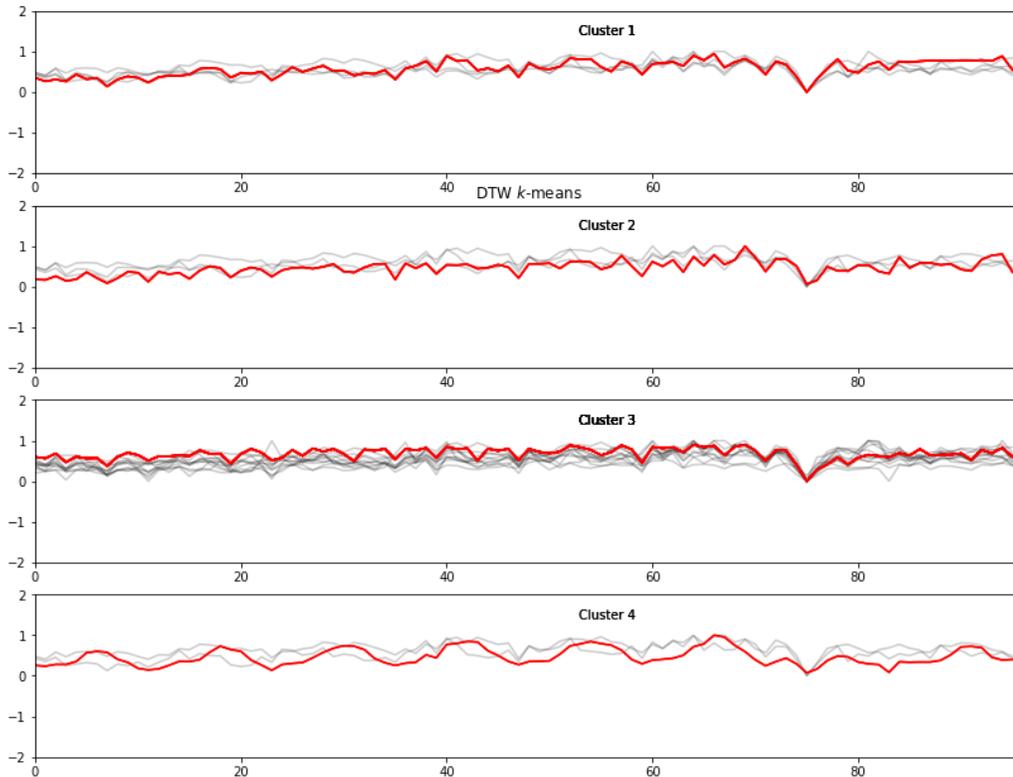


Figura 4.13: Representación de las series de cada grupo junto a su centroide.

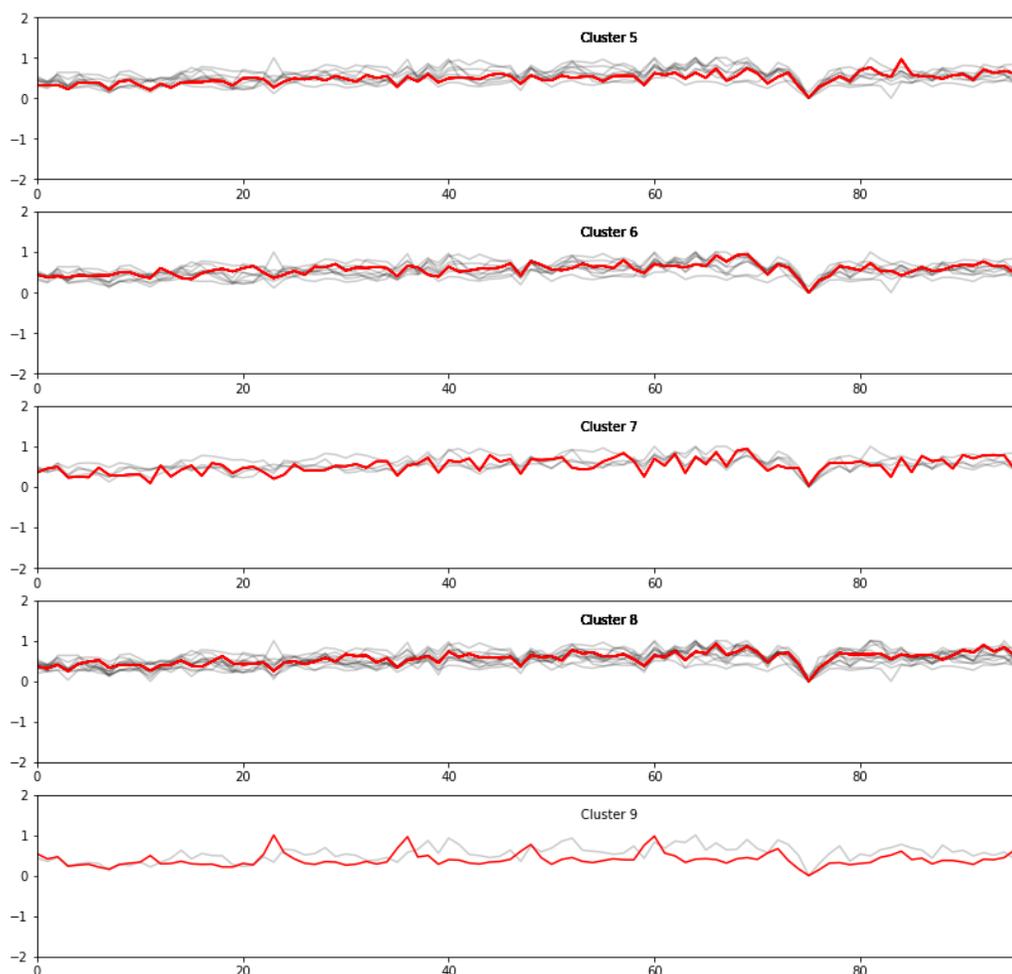


Figura 4.14: Representación de las series de cada grupo junto a su centroide.

En la mayoría de las representaciones gráficas es complicado distinguir un patrón claro. Sin embargo, sí que se aprecian las tendencias mencionadas en dos de los grupos: tanto el *clúster* 4, como el 9 se había mencionado que mostraban resultados cíclicos. En las representaciones anteriores se puede ver como los centroides de estos dos *clústers* y las series que forman parte de ellos presentan una serie periódica, con picos en los periodos de mayor número de accidentes y bajadas con el descenso de los accidentes.

Aparte de las características generales de cada grupo, en todos ellos se puede apreciar el gran descenso de los accidentes en abril de 2020.

### 4.3.2. Agrupación de los sectores económicos

A continuación se presentan los resultados referentes a las secciones económicas. El procedimiento es totalmente análogo al mostrado en la clasificación de las provincias. Por tanto, en primer lugar se construye la matriz de distancias, usando la métrica DWT de nuevo y con ella se aplica el método de agrupamiento jerárquico. A través de él se pretende obtener un número de grupos razonable para realizar la clasificación con el método DBA.

Antes de proceder con el método, se enumeran todos los sectores económicos que se consideran:

- A-Agricultura, ganadería, silvicultura y pesca
- B-Industrias extractivas
- C-Industria manufacturera
- D-Suministro de energía eléctrica, gas, vapor y aire acondicionado
- E-Suministro de agua, saneamiento, gestión de residuos
- F-Construcción
- G-Comercio al por mayor y menor; reparación vehículos
- H-Transporte y almacenamiento
- I-Hostelería
- J-Información y comunicaciones
- K-Actividades financieras y de seguros
- L-Actividades inmobiliarias
- M-Actividades profesionales, científicas y técnicas
- N-Actividades administrativas y servicios auxiliares
- O-Administración Pública y defensa; Seguridad Social obligatoria
- P-Educación
- Q-Actividades sanitarias y de servicios sociales
- R-Actividades artísticas, recreativas y de entretenimiento
- S-Otros servicios
- T-Actividades de hogares como empleadores de personal doméstico
- U-Actividades de organizaciones y organismos extraterritoriales.

Tras esta aclaración, se procede a mostrar los resultados del *clúster* jerárquico. En la Figura 4.15 se puede observar cómo, considerando como punto de corte una distancia igual a 4, se generan 6 grupos. Considerar una distancia ligeramente menor provocaría que se creasen varios grupos con un solo elemento y considerar distancias mayores provocaría una reducción del número de grupos. Ninguno de los dos casos parece mejorar la situación que se obtiene considerando 6 grupos. Por tanto, la agrupación mediante DBA se ha realizado considerando un número de *clústers* igual a 6.

A continuación se enumeran los grupos y los sectores que cada uno de ellos contiene:

- **GRUPO 1:** J-Información y comunicaciones, M-Actividades profesionales, científicas y técnicas, R-Actividades artísticas, recreativas y de entretenimiento, S-Otros servicios.
- **GRUPO 2:** B-Industrias extractivas, D-Suministro de energía eléctrica, gas, vapor y aire acondicionado, K-Actividades financieras y de seguros, O-Administración Pública y defensa; Seguridad Social obligatoria, T-Actividades de hogares como empleadores de personal doméstico.
- **GRUPO 3:** N-Actividades administrativas y servicios auxiliares, Q-Actividades sanitarias y de servicios sociales.

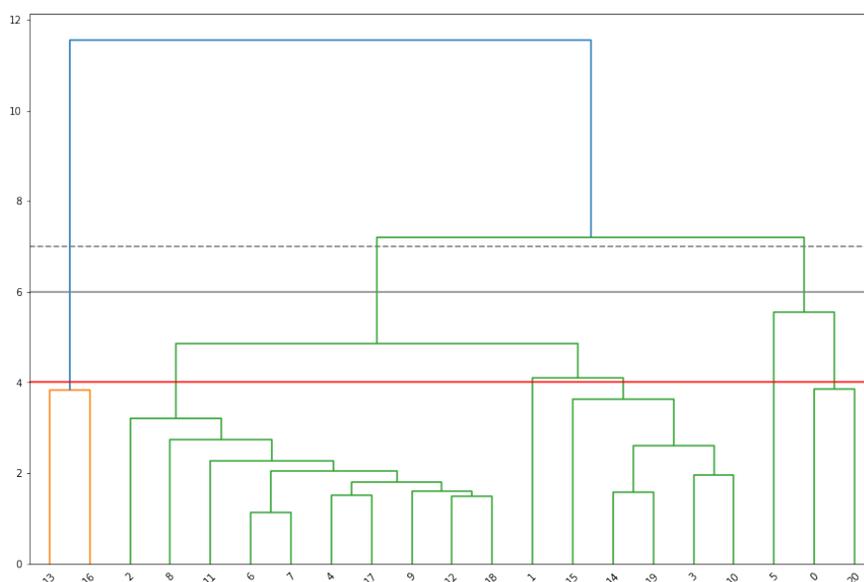


Figura 4.15: Agrupamiento mediante el método jerárquico de las series mensuales de los sectores.

- **GRUPO 4:** A-Agricultura, ganadería, silvicultura y pesca, U-Actividades de organizaciones y organismos extraterritoriales.
- **GRUPO 5:** P-Educación
- **GRUPO 6:** C-Industria manufacturera, E-Suministro de agua, saneamiento, gestión de residuos, F-Construcción, G-Comercio al por mayor y menor; reparación vehículos, H-Transporte y almacenamiento, I-Hostelería, L-Actividades inmobiliarias.

Comparando estos resultados con el mapa de calor de la Figura 4.11 se pueden realizar los siguientes comentarios:

El primer grupo contiene actividades cuyas series son bastante monótonas durante los 6-7 primeros años. Las 4 muestran un ligero aumento desde 2014 hasta 2020, coincidiendo con la situación general del país; pero tras la gran caída en el número de accidentes en abril de 2020, el número de accidentes es menor en cada uno de los sectores del grupo. Cabe destacar que, en este grupo, todos los sectores presentan en general un bajo número de accidentes (como se puede ver en la Figura 4.10). Para observar las tendencias se deben normalizar las series, por eso cuando se mencionan subidas o bajadas se hace siempre de forma relativa a los valores que presenta cada serie.

El grupo 2 también contiene sectores económicos que presentan un bajo número de accidentes en total. En cuanto a las tendencias que presentan en el número de accidentes, se puede apreciar en todas ellas un descenso claro de accidentes a lo largo de los 8 años. En el caso de los sectores B y D ya se había comentado anteriormente, pero se puede comprobar que el resto de series del grupo también lo verifican.

El tercero de los grupos solo contiene dos series (de los sectores N y Q). En el caso de las actividades sanitarias y servicios sociales (Q), la serie normalizada presenta valores muy bajos a lo largo de los 8 años, salvo por el mes de enero de 2021. En el caso de las actividades administrativas y servicios auxiliares ocurre algo similar, pero en esta serie son varios los meses de 2021 donde se pueden ver valores más altos. Por tanto, resulta claro que las tendencias de estas dos series son muy similares: presentando un pico muy alto de accidentes en comparación con el resto de valores de su serie.

El grupo 4 contiene también solo dos secciones: la A y la U. La segunda de ellas muestra una serie con bastantes subidas y bajadas, intercalando en el mapa de calor colores claros y oscuros. De todas

formas, en 2015-2016 parece que tiene una etapa en la que el número de accidentes parece mayor en general. Por otro lado, la serie del sector A no muestra tanta irregularidad, pero también presenta una etapa con un mayor número de accidentes que el resto de la serie: en 2018-2019.

El grupo 5 solamente contiene una serie, que es la correspondiente a la sección de educación. Como se menciona anteriormente, esta serie es periódica, mostrando un número de accidentes más o menos constante durante el curso y mostrando descensos en los meses de vacaciones: diciembre, abril, junio y julio y, sobre todo, agosto.

El último de los grupos contiene varias de las secciones con mayor número de accidentes. Sin embargo, otras presentan números bastante más discretos. De todas formas, todas las series tienen en común que tienen una tendencia creciente, aunque sí es cierto que en algunas series como la C o la F (que se corresponden con las secciones que más accidentes contabilizan) ese crecimiento es más notable.

Esta agrupación muestra una característica destacable frente a la de las provincias. Mientras que en la primera realizada se observaba que en general todas las provincias mostraban tendencias crecientes en sus series, en el caso de las secciones se puede comprobar que algunas han aumentado sus casos de accidentes laborales, pero otras los han disminuido. Sin embargo, el aumento general es algo que se sigue cumpliendo, lo cual prueba que el aumento de casos en las diferentes secciones ha superado a la disminución de otras.

En las Figuras 4.16 y 4.17 se pueden ver las representaciones de cada grupo junto a su centroide.

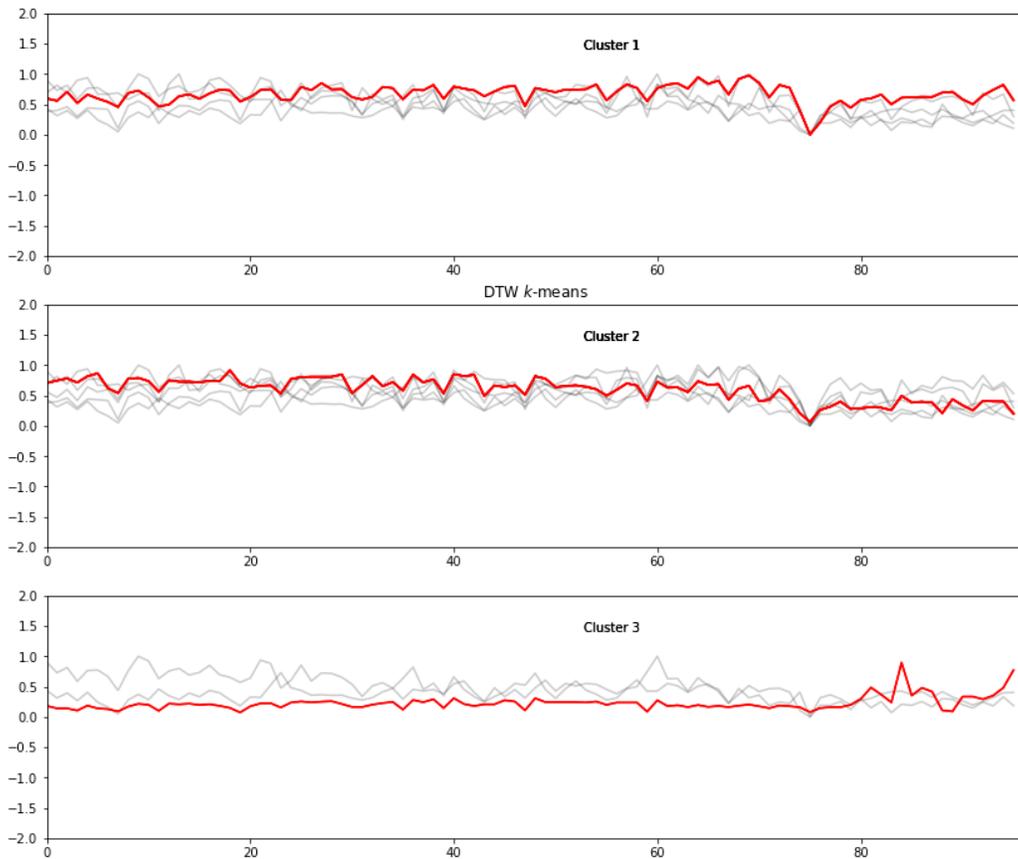


Figura 4.16: Representación de las series de cada grupo junto a su centroide.

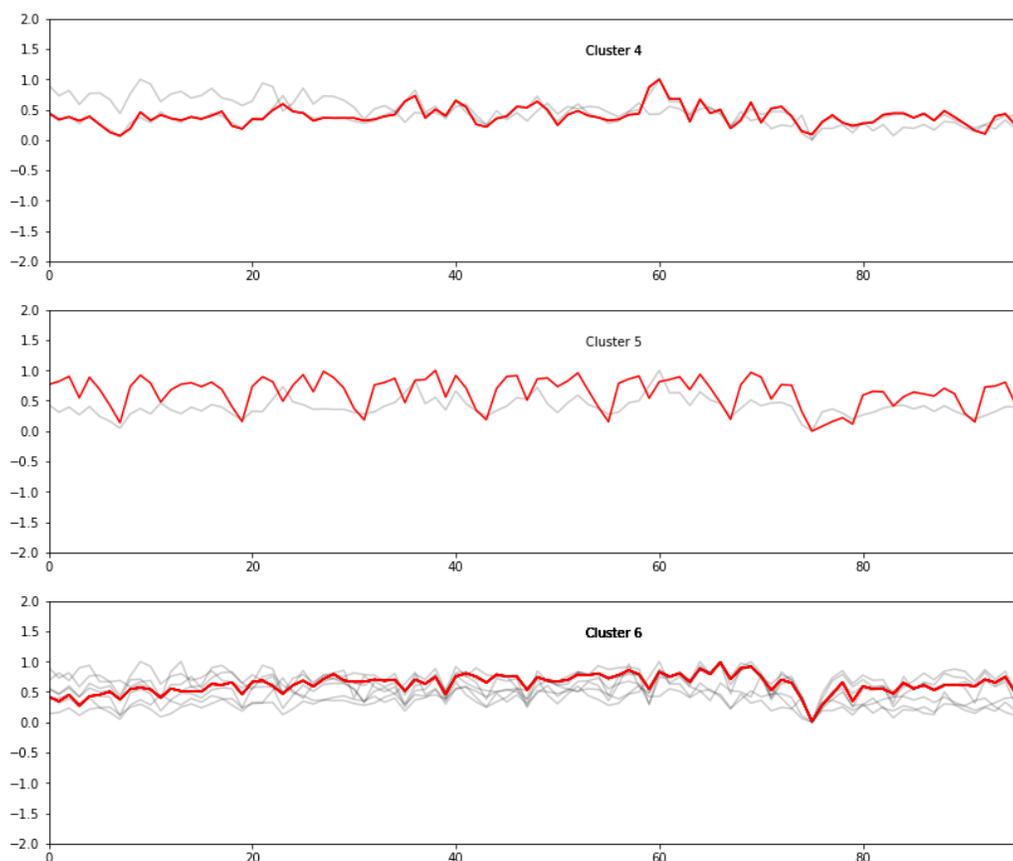


Figura 4.17: Representación de las series de cada grupo junto a su centroide.

Al contrario que en la clasificación de las provincias, en este caso no se refleja en todos los *clústers* el efecto de abril de 2020. Solamente en los dos primeros y en el último se puede apreciar de forma clara esa bajada en el centroide y las series que pertenecen al grupo.

Además, en algunas de las representaciones se puede apreciar de forma clara las tendencias comentadas anteriormente. Por ejemplo, el centroide del grupo 3 muestra claramente la tendencia estacionaria general de la serie interrumpida por un pico en el número de accidentes. En esa misma línea, el centroide del grupo 5 representa de forma notoria la periodicidad correspondiente al sector educativo.

### 4.3.3. Agrupación conjunta

Esta última sección está también dedicada a métodos de agrupación *clustering*. En este caso, la intención es establecer relaciones entre las provincias y las diferentes secciones económicas. Con el resultado de este agrupamiento se espera obtener información útil para que, a la hora de organizar medidas de PRL sea conocido qué regiones y actividades económicas tienen necesidades similares.

En un primer intento de agrupar las provincias y las secciones se detectó un inconveniente: cada una de las secciones económicas engloba divisiones económicas menores que presentan series con resultados diferentes. Por ejemplo, la sección “A-Agricultura, ganadería, silvicultura y pesca” considera en una misma categoría la ganadería y la pesca, cuando estas actividades presentan periodos de actividad y características diferentes. Por tanto, teniendo esta situación en consideración, se tomó la decisión de

considerar en la agrupación de esta sección las provincias y las divisiones económicas, en lugar de las secciones.

A continuación se incluye el mapa de calor (Figura 4.18) de las series ya estandarizadas de cada una de las divisiones económicas que se consideran.



Figura 4.18: *Heatmap* que muestra el número de accidentes que tuvieron lugar en cada división económica entre 2014 y 2021.

Debido a la gran cantidad de divisiones que se consideran (89), resulta muy costoso poder analizar detalladamente la gráfica anterior. De todas formas, se pueden extraer algunas conclusiones generales: en abril de 2020 prácticamente todas las divisiones presentan su menor número de accidentes de toda la serie y en los meses de 2019, tienen lugar los resultados más altos en general (con excepciones de algunas series como las de actividades sanitarias o actividades relacionadas con el empleo).

El mapa de calor de las series sin estandarizar en este caso proporciona conclusiones todavía más pobres que en los casos anteriores. La mayoría de resultados se ven eclipsados por el número de accidentes ocurridos en enero de 2021 para la división de actividades sanitarias. Esto provoca que la mayor parte del mapa sea de color oscuro, sin permitir observar tendencias destacables. Se incluye en la Figura 4.19 para que se puedan observar las divisiones con un mayor número de accidentes, pero realmente será el gráfico anterior el que se use como apoyo para las conclusiones (al igual que en los dos apartados anteriores).

Actividades como la construcción de edificios, la ingeniería civil, la construcción especializada, el comercio de vehículos de motor, los servicios de comidas y bebidas y las relacionadas con la administración pública y defensa son algunas de las más destacadas en cuanto a número de accidentes laborales. Estos resultados coinciden con lo obtenido en el análisis de las series por sectores; y es que las divisiones que presentan mayor número de accidentes se engloban dentro de los sectores con más accidentes.

Tras esta breve introducción de las divisiones económicas (la lista completa se puede encontrar en *Datapané*), se procede con la agrupación de estas junto a las provincias y ciudades autónomas.



Figura 4.19: *Heatmap* que muestra el número de accidentes que tuvieron lugar en cada división económica entre 2014 y 2021.

El procedimiento seguido es análogo al mostrado en los dos casos anteriores, realizando primero una agrupación jerárquica para deducir a partir de ella un número razonable de grupos para aplicar al algoritmo DBA.

En la Figura 4.20 se incluye el dendrograma en el cual se puede visualizar para qué distancias se han ido uniendo los grupos. En este caso, la distancia entre series ha sido más grande en general, por lo que también se ha considerado una distancia más grande para decidir el número de *clústers* para el algoritmo DBA. Para una distancia de 7.5, se ha obtenido un total de 9 grupos, que para el caso a tratar parece razonable.

A continuación, se enumera cada uno de los grupos obtenidos:

#### ■ GRUPO 1:

- Guipúzcoa
- Vizcaya
- 13-Industria textil
- 16-Industria de la madera y del corcho, excepto muebles; cestería y espartería
- 17-Industria del papel
- 20-Industria química
- 21-Fabricación de productos farmacéuticos
- 22-Fabricación de productos de caucho y plásticos
- 23-Fabricación de otros productos minerales no metálicos
- 24-Metalurgia; fabricación de productos de hierro

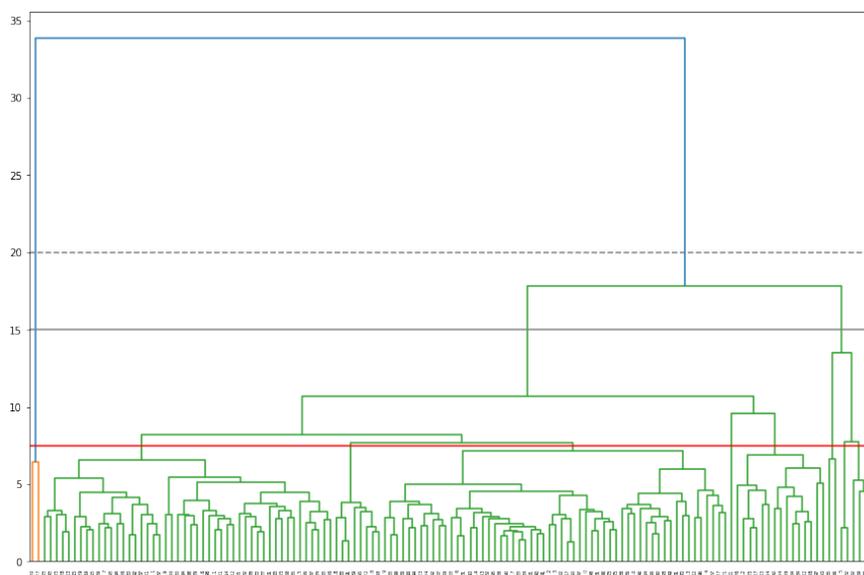


Figura 4.20: Agrupamiento mediante el método jerárquico de las series de provincias y divisiones.

- 25-Fabricación de productos metálicos, excepto maquinaria y equipo acero y ferroaleaciones
- 26-Fabricación de productos informáticos, electrónicos y ópticos
- 27-Fabricación de material y equipo eléctrico
- 28-Fabricación de maquinaria y equipo n.c.o.p.
- 31-Fabricación de muebles
- 32-Otras industrias manufactureras.

Este grupo contiene muchas de las actividades económicas que se engloban dentro del sector industria. Además, contiene las provincias costeras vascas Guipúzcoa y Vizcaya. Este resultado es muy razonable, pues junto al sector servicios, la industria es el sector económico más potente del País Vasco<sup>5</sup>, tanto históricamente como en la actualidad. En general, la tendencia de las series de este grupo es creciente.

■ **GRUPO 2:**

- Almería
- Granada
- Málaga
- Teruel
- Zaragoza
- Cantabria
- Albacete
- Guadalajara
- Ávila
- Salamanca

<sup>5</sup>Instituto Cultural Vasco. Última vez visitado el 22 de enero de 2023 [online](#).

- Soria
- Zamora
- Girona
- Lleida
- Tarragona
- Alicante
- Castellón
- Valencia
- Cáceres
- Coruña (A)
- Lugo
- Ourense
- Pontevedra
- Murcia (Región de)
- Melilla
- 02-Silvicultura y explotación forestal
- 33-Reparación e instalación de maquinaria y equipo
- 42-Ingeniería civil
- 43-Actividades de construcción especializada
- 49-Transporte terrestre y por tubería
- 53-Actividades postales y de correos
- 59-Actividades cinematográficas, de vídeo y de programas de televisión, grabación de sonido y edición musical
- 71-Servicios técnicos de arquitectura e ingeniería; ensayos y análisis técnicos
- 74-Otras actividades profesionales, científicas y técnicas
- 75-Actividades veterinarias
- 81-Servicios a edificios y actividades de jardinería
- 82-Actividades administrativas de oficina y otras actividades auxiliares a las empresas
- 88-Actividades de servicios sociales sin alojamiento.

Este segundo grupo no tiene una interpretación tan clara como el anterior. Presenta un alto número de provincias diferentes, aunque en varios casos se han incluido todas las provincias de algunas CC.AA. y en otros, varias de una misma comunidad. Además, las divisiones económicas incluidas son bastante dispares; pero tras observar su serie en el mapa de calor 4.18 se ha apreciado que todas presentan una tendencia creciente en su serie de número de accidentes. Esta situación coincide con la general de todas las provincias españolas: tendencias crecientes de las series, cosas que no ocurre con todas las divisiones y secciones económicas.

■ **GRUPO 3:**

- Ceuta
- 08-Otras industrias extractivas
- 14-Confeción de prendas de vestir
- 15-Industria del cuero y del calzado

- 18-Artes gráficas y reproducción de soportes grabados
- 29-Fabricación de vehículos de motor, remolques y semirremolques
- 30-Fabricación de otro material de transporte
- 36-Captación, depuración y distribución de agua
- 60-Actividades de programación y emisión de radio y televisión
- 61-Telecomunicaciones
- 62-Programación, consultoría y otras actividades relacionadas con la informática
- 69-Actividades jurídicas y de contabilidad
- 72-Investigación y desarrollo
- 73-Publicidad y estudios de mercado
- 84-Administración Pública y defensa; Seguridad Social obligatoria
- 85-Educación
- 97-Actividades de los hogares como empleadores de personal doméstico.

El grupo 3 es similar al anterior, en el sentido de que contiene muchas actividades económicas que *a priori* no guardan especial relación. De todas formas, al igual que en el grupo 2, también se puede intuir una tendencia similar en sus series: se aprecian descensos en los 3 meses principales de vacaciones (abril, agosto y diciembre). En algunos casos, como en fabricación de otro material de transporte, confección de prendas de vestir, industria del cuero o, sobre todo, en educación es más notable que en otras; pero en general parece que es la tendencia más destacable del grupo. La ciudad autónoma Ceuta también presenta esta característica.

■ **GRUPO 4:**

- 78-Actividades relacionadas con el empleo
- 86-Actividades sanitarias

Este grupo es uno de los más sencillos y claros de interpretar. Ambas muestran un muy bajo número de accidentes (relativo a cada serie) a lo largo de la mayor parte del periodo considerado. A partir del año 2020 crece el número de accidentes, pero más notablemente en el caso de las actividades sanitarias.

■ **GRUPO 5:**

- Baleares (Illes)
- 07-Extracción de minerales metálicos
- 50-Transporte marítimo y por vías navegables interiores
- 51-Transporte aéreo
- 55-Servicios de alojamiento
- 77-Actividades de alquiler
- 79-Actividades de agencias de viajes, operadores turísticos, servicios de reservas y actividades relacionadas con los mismos
- 90-Actividades de creación, artísticas y espectáculos.

El quinto de los grupos está formado por la provincia de las Islas Baleares y diversas actividades relacionadas con el turismo y actividades sociales, lo cual guarda bastante relación. Por otro lado, también pertenecen a este grupo las divisiones de extracción de minerales metálicos y las relacionadas con transportes.

■ **GRUPO 6:**

- 03-Pesca y acuicultura
- 05-Extracción de antracita, hulla y lignito
- 09-Actividades de apoyo a las industrias extractivas
- 35-Suministro de energía eléctrica, gas, vapor y aire acondicionado
- 58-Edición
- 64-Servicios financieros, excepto seguros y fondos de pensiones
- 65-Seguros, reaseguros y fondos de pensiones, excepto Seguridad Social obligatoria
- 99-Actividades de organizaciones y organismos extraterritoriales.

Este grupo presenta actividades económicas de diferentes sectores, por lo que en principio no se encuentra una relación clara entre todos ellos. Sin embargo, en cuanto a la tendencia de sus series, se observa claramente que todas muestran una tendencia decreciente en el número de accidentes laborales a lo largo de los años considerados.

■ **GRUPO 7:**

- 06-Extracción de crudo de petróleo y gas natural
- 12-Industria del tabaco
- 19-Coquerías y refino de petróleo

El grupo 7 contiene solo tres divisiones económicas, con bastante relación entre ellas. Además, sus series muestran una periodicidad muy similar, como era de esperar, pues la actividad de estas 3 divisiones está bastante relacionada.

■ **GRUPO 8:**

- Cádiz
- Sevilla
- Asturias (Principado de)
- Palmas (Las)
- S. C. Tenerife
- León
- Barcelona
- Badajoz
- Madrid (Comunidad de)
- 38-Recogida, tratamiento y eliminación de residuos; valorización
- 45-Venta y reparación de vehículos de motor y motocicletas
- 46-Comercio al por mayor e intermediarios del comercio, excepto de vehículos de motor y motocicletas
- 47-Comercio al por menor, excepto de vehículos de motor y motocicletas
- 52-Almacenamiento y actividades anexas al transporte
- 56-Servicios de comidas y bebidas
- 68-Actividades inmobiliarias
- 87-Asistencia en establecimientos residenciales

- 91-Actividades de bibliotecas, archivos, museos y otras actividades culturales
- 92-Actividades de juegos de azar y apuestas
- 93-Actividades deportivas, recreativas y de entretenimiento
- 94-Actividades asociativas
- 95-Reparación de ordenadores, efectos personales y artículos de uso doméstico
- 96-Otros servicios personales.

En este grupo se engloban prácticamente las mismas provincias que el grupo 3 de la clasificación por provincias. Con respecto a las divisiones económicas, en este grupo se consideran principalmente actividades el sector servicios. Todas estas series presentan valores altos a lo largo de toda su serie, sobre todo hasta abril de 2020, donde hay una gran bajada. A partir de esta, además, disminuye el número de accidentes en comparación con los años anteriores, pero se muestra una tendencia al crecimiento.

■ **GRUPO 9:**

- Córdoba
- Huelva
- Jaén
- Huesca
- Ciudad Real
- Cuenca
- Toledo
- Burgos
- Palencia
- Segovia
- Valladolid
- Navarra (C. Foral de)
- Álava
- Rioja (La)
- 01-Agricultura, ganadería, caza y servicios relacionados con las mismas
- 10-Industria de la alimentación
- 11-Fabricación de bebidas
- 37-Recogida y tratamiento de aguas residuales
- 39-Actividades de descontaminación y otros servicios de gestión de residuos
- 41-Construcción de edificios
- 63-Servicios de información
- 66-Actividades auxiliares a los servicios financieros y a los seguros
- 70-Actividades de las sedes centrales; actividades de consultoría de gestión empresarial
- 80-Actividades de seguridad e investigación.

Finalmente, en el grupo 9 se almacenan las provincias restantes, que salvo Huelva, son todas interiores. Además, tienen en común que son provincias que no destacan por tener un alto número de habitantes. Sin embargo, las divisiones económicas incluidas no parecen tener una relación común. De todas formas, todas parecen presentar cierta periodicidad en sus series, con una tendencia creciente general. En el caso de las provincias, hay algunas como Huelva y Jaén que sí que presentan una periodicidad más destacada. En el resto de provincias esa característica no se observa tan claramente, pero sí es cierto que son relativamente notables las bajadas en los periodos vacacionales.

En las Figuras 4.21 y 4.22 se puede observar la representación de las series de cada uno de los *clústers* de forma conjunta, representándose también el centroide de cada grupo.

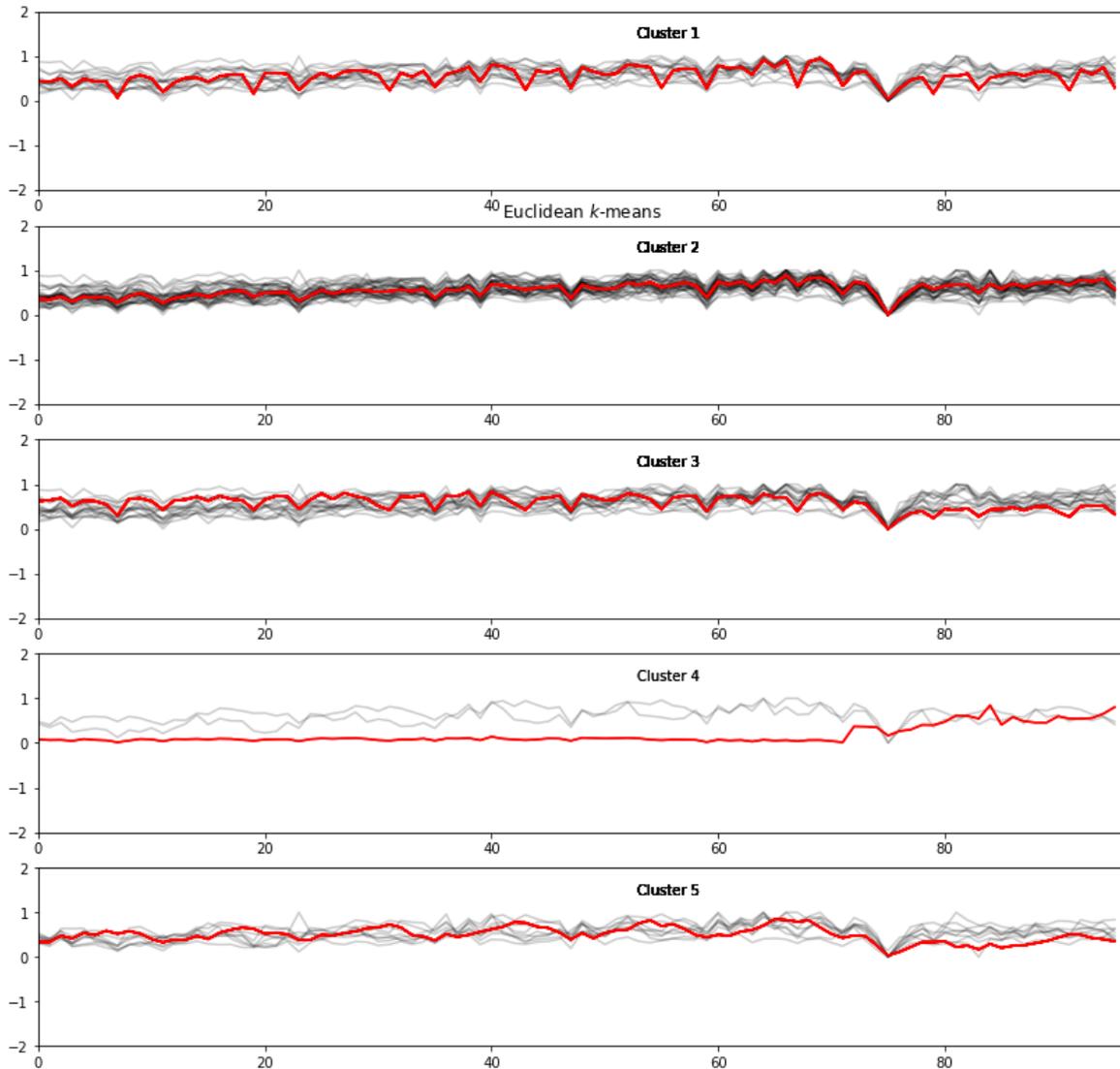


Figura 4.21: Representación de las series de cada grupo junto a su centroide.

En algunas de estas gráficas presenta cierta dificultad encontrar los patrones comentados anteriormente. Sin embargo, hay otras en las que sus características son más notables. Ejemplo de esto

es el grupo 4, en cuya gráfica se puede apreciar el aumento de accidentes hacia el final del periodo considerado que se comentaba. En la gráfica correspondiente al grupo 6 también se puede apreciar la tendencia descendiente. En las demás, las tendencias son más complicadas de diferenciar. De todas formas, sí hay un detalle bastante notable en todas, y es la gran bajada en abril de 2020.

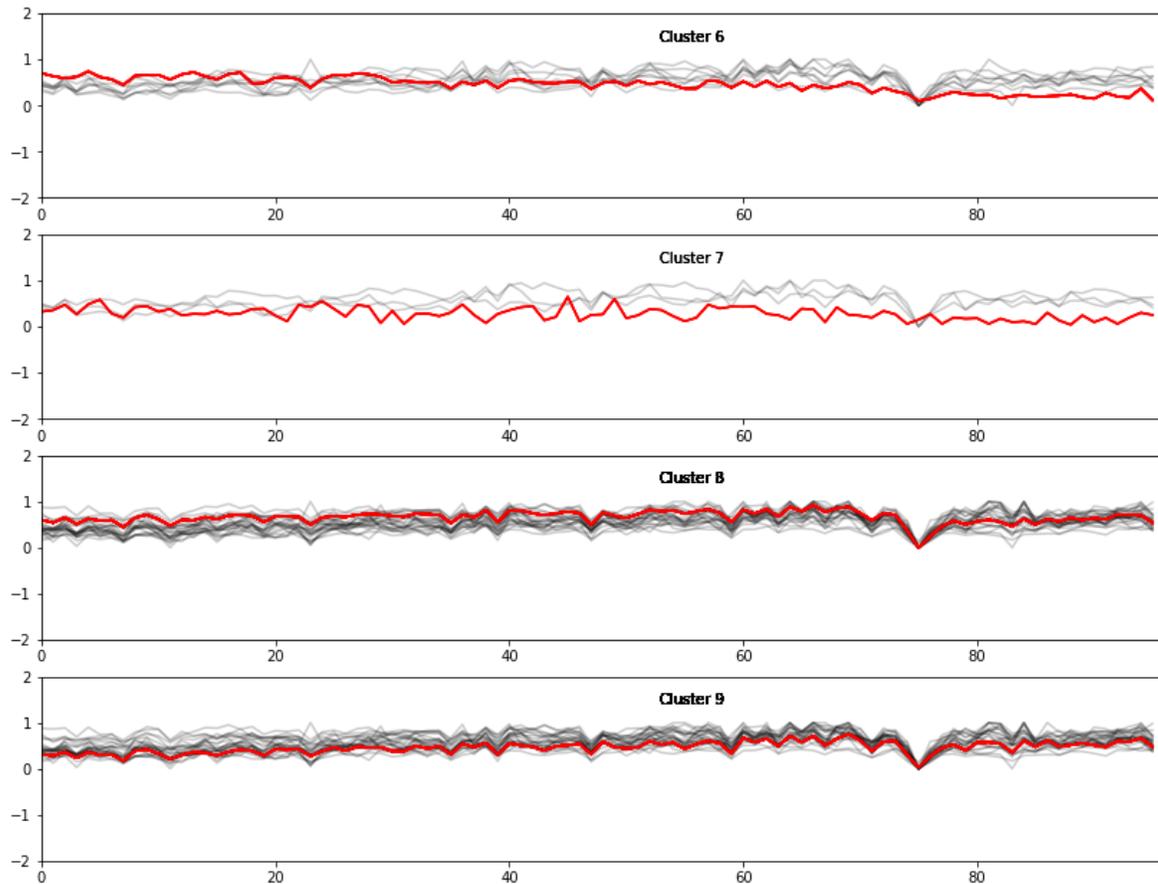


Figura 4.22: Representación de las series de cada grupo junto a su centroide.

Al igual que los resultados del resto de secciones, los obtenidos en esta última se comentan conjuntamente en el capítulo final.

## Capítulo 5

# Conclusión

En este último capítulo se incluyen las conclusiones acerca de los resultados mostrados a lo largo del documento. La idea general es repasar los resultados más relevantes y tratar de encontrar relaciones entre ellos. De esta forma, los resultados extraídos cobran un mayor sentido al complementarse con los datos obtenidos en otras categorías.

Comenzando por los primeros datos presentados, cabe destacar el resultado más claro: el notable aumento del número de accidentes laborales en España. Como punto inicial y para mostrar una visión global, se ha presentado la situación general y se ha observado que tanto los accidentes durante la jornada como *in itinere* están en aumento. Este resultado se ha ido repitiendo a lo largo de todas las categorías estudiadas. En todos los gráficos presentados que muestran la evolución de los datos durante los 8 años considerados, se ha podido ver como ha ido teniendo lugar ese aumento. Junto con este aspecto, hay otro que se ha repetido también: la gran caída en la cantidad de accidentes en 2020. Para cualquiera de las categorías consideradas, el efecto provocado por la pandemia del COVID-19 ha sido notable. La única categoría que muestra un resultado diferente ante este fenómeno se pudo observar en el análisis de las series temporales según el sector o división económica. Ahí se ha podido ver cómo para el sector sanitario esto produjo un pico de bajas laborales, aunque este efecto no fue tan inmediato como el que se observa en el resto de casos. Por tanto, resulta claro que este evento ha sido el único que ha ralentizado el continuo crecimiento del número de accidentes laborales en España.

Centrando el tema ya en clasificaciones más específicas, en primer lugar se han analizado los accidentes en función de las características del trabajador o trabajadora accidentada. El primer resultado que se extrajo mostraba que los jóvenes eran los más propensos a sufrir accidentes laborales, quizás relacionado con su inexperiencia, la falta de formación o incluso porque en muchas ocasiones están contratados en situaciones precarias. Por otro lado, las personas mayores de 60 mostraban la menor proporción de accidentados, pues normalmente suelen estar en posiciones más privilegiadas tras años de trabajo. Estos son los que presentan una menor proporción de ocupación junto a los jóvenes, por lo que ambos representan los dos extremos en esta clasificación.

Con respecto al sexo, también se fue detectada una categoría con una proporción de accidentes que destacaba sobre la otra. A pesar de que es cierto que el número de hombres ocupados es superior al de mujeres, la diferencia en el número de accidentados es mucho mayor; por lo que la conclusión es que los hombres son más propensos a sufrir accidentes laborales. Este resultado seguramente esté influido por los oficios que personas de uno y otro sexo tienen más frecuentemente. En el caso de actividades relacionadas con la construcción, por ejemplo, es extraño encontrarse con mujeres ejerciendo y, como se ha visto, esta es una de las actividades que más accidentes presentan. Se puede deducir entonces que algunos de los oficios regularmente asociados al género masculino son potencialmente más peligrosos y tienen parte de influencia en que el número de accidentes laborales sea más elevado en el caso de los varones.

Las siguientes clasificaciones sobre las que se han presentado datos han sido el número de accidentes según la ocupación del trabajador/a y según su tipo de contrato. En ambos casos se ha podido ver que las categorías con un mayor número de ocupados han sido las que han presentado un mayor número de accidentes. El resultado obtenido más destacable ha sido que el número de accidentes de los autónomos sin asalariados dobla al de aquellos que sí los tienen. Seguramente esta situación tiene lugar debido a que aquellos autónomos con empleados pueden delegar el trabajo en ellos y de esta forma evitan muchos de los riesgos relacionados con el trabajo.

Considerando la distribución de los accidentes por ocupación del accidentado, se ha podido observar que hay algunos grupos de ocupación que presentan un mayor riesgo. Los resultados expuestos guardan bastante relación con los que se pueden observar en el mapa de calor de las series de tiempo por sectores económicos. Aquellos sectores que presentaban un mayor número de accidentes: construcción, industria manufacturera, comercio al por mayor y transporte, hostelería y pesca y agricultura, se corresponden con las profesiones que presentan el volumen de accidentes más alto a lo largo de estos últimos 8 años.

Cuando se restringen los resultados a los accidentes *in itinere*, las diferencias en el número de accidentes entre las diferentes ocupaciones es menos notable. Estos accidentes ocurren fuera de la jornada laboral, por lo que la actividad que se realiza durante esta no influye de forma tan directa en el accidente. Además, el número de accidentes *in itinere* es mucho menor que los que ocurren durante la jornada, por lo que son estos últimos en los que urge poner más énfasis en su solución.

Finalizando el apartado de características del accidentado/a, se presentan el número de accidentes por continente de procedencia. En *Datapane* se pueden encontrar también los resultados por países. De todas formas, el resultado más claro es que la mayoría de accidentes registrados le ocurren a españoles. De nuevo, este resultado viene ligado a que la gran mayoría de ocupados del país son españoles. El resto del continente europeo, África y América muestran números similares entre ellos, mientras que el número de asiáticos accidentados es muy inferior y el de oceánicos prácticamente despreciable.

Por tanto, se deduce que las características del trabajador que más influyen en la accidentalidad laboral son la edad y el sexo de este. Siendo los jóvenes y los hombres los que mayores proporciones de accidentes presentan.

Los resultados en cuanto a las características de la empresa no han sido muy reveladores: en general, las categorías con un mayor número de personas ocupadas son las que presentan mayor número de accidentes. No destaca especialmente ninguna característica que rompa esa relación.

En primer lugar, se han estudiado los accidentes que causan baja según el lugar donde ocurrió dicho accidente. El centro habitual de trabajo ha sido la localización en la que la gran mayoría de los accidentes han tenido lugar, lo que coincide con el lugar de trabajo donde mayormente se suelen encontrar los trabajadores durante la jornada laboral.

Tras el estudio de esta clasificación, se consideró el estudio de los accidentes según la empresa contratante del trabajador/a. Al igual que en el caso anterior, la categoría mayoritaria (la propia empresa) fue la que presentó un mayor número de accidentes.

Respecto a las categorías según la prevención laboral de la empresa, los resultados se vuelven a repetir: las categorías que engloban un mayor número de empleados son las que presentan también un mayor número de bajas. En este caso particular, cabe destacar que un mismo individuo puede pertenecer a varias categorías. En la prevención de riesgos laborales existen 4 especialidades diferentes: Seguridad en el Trabajo, Higiene Industrial, Ergonomía y Psicología Aplicada y Medicina del Trabajo. Es muy común que las empresas recurran a diferentes servicios para las diferentes categorías. Concretamente, la medicina del trabajo suele estar externalizada, por lo que la gran mayoría de las empresas cuentan con servicios de prevención externa, que es justamente la categoría que mayor número de accidentes presenta.

De igual forma, las empresas que cuentan con evaluación de los riesgos presentan el doble de

accidentes que las que no, pues las primeras representan una proporción mucho mayor del total de las empresas.

Como última característica relacionada con las empresas se estudió la distribución de los accidentes según el número de empleados de estas. Los resultados siguieron la tónica general de la sección: los grupos mayoritarios presentan mayor número de accidentes.

Se deduce, por tanto, que las características que han sido evaluadas de las empresas no son un factor notable en la distribución de los accidentes. Sí es cierto que se observan categorías que a nivel total destacan mucho sobre el resto, pero en términos relativos no se observan estas diferencias.

La siguiente categoría considerada presenta el número de accidentes en función de las características de los propios accidentes. En cada caso se han mostrado los resultados totales y también diferenciando según la gravedad de los propios accidentes. Debido a que la gran mayoría son leves, los resultados generales y estos son muy similares. En el caso de los accidentes graves, se aprecian variaciones en los resultados y, al restringirse al caso de los accidentes mortales, los resultados son completamente diferentes. Muestra de esta situación son, por ejemplo, los resultados en función de las partes del cuerpo lesionadas y los tipos de lesiones: en general, las lesiones en las extremidades y las bajas por esguinces o torceduras son las situaciones que tienen lugar más frecuentemente en el caso de los accidentes laborales; sin embargo, si solo se consideran los accidentes que provocaron el fallecimiento del trabajador/a, las lesiones más comunes afectan a múltiples partes del cuerpo, a la región torácica y a la cabeza, siendo los infartos, derrames cerebrales y otras patologías las causas más comunes.

Esta diferencia de resultados se puede apreciar en la mayoría de clasificaciones relacionadas con las características del accidente. Además, en varias también existe una gran diferencia entre los datos presentados para los accidentes en jornada y los accidentes *in itinere*. Esta disparidad se debe sobre todo a que los accidentes que ocurren desplazándose hacia o desde el lugar de trabajo tienen lugar en medios de transporte. Por este motivo, algunas de las categorías consideradas, como “actividad física que realizaba el trabajador”, “la desviación que provocó el accidente” o “el agente material” están limitadas en cuanto a los resultados que se puede obtener, pues los posibles riesgos que pueden sufrir no son similares, en general, a los que sufren en su puesto de trabajo. Esta diferencia se ve reflejada en los resultados, donde se puede apreciar que los accidentes que suceden *in itinere* están en general relacionados con accidentes de tráfico.

Teniendo en cuenta esta situación, se puede extraer una conclusión que parece clara, y es que es necesario establecer medidas de prevención de riesgos laborales a distintos niveles. Las medidas preventivas que se deben aplicar para evitar las lesiones leves no pueden ser las mismas que se aplican para vigilar y evitar accidentes de carácter más grave. De igual forma, no se pueden implantar los mismos métodos de prevención durante la jornada y fuera de ella. Por ejemplo, de cara a ser previsores y evitar patologías o episodios provocados por infartos o derrames, es útil que las empresas ofrezcan a sus empleados revisiones médicas de forma periódica. Con la intención de reducir los accidentes más leves, es importante dar una buena formación inicial para evitar las malas prácticas en el puesto de trabajo. Esta medida, concretamente, también resulta interesante considerarla de cara a la reducción de accidentes de los más jóvenes, que cuentan con menos experiencia laboral. Con respecto a los accidentes *in itinere*, la influencia de la empresa no es tan directa, pero aun así existen medidas que se pueden tomar para mejorar la situación, como por ejemplo facilitar un medio de transporte para llegar al lugar de trabajo.

Es de interés en este caso volver a mencionar la existencia de las diferentes especialidades de PRL, pues anteriormente se ha mencionado que en una misma empresa se disponía de distintos métodos de prevención y estos resultados reflejan la necesidad de ello. De cara a ser más precisos con la prevención y actuación sobre los diferentes riesgos y accidentes, conviene que las medidas se tomen de forma específica para solucionar cada problema y no de forma general.

El último bloque de características analizadas ha mostrado los resultados del número de accidentes laborales según diferentes categorías relacionadas con la temporalidad de estos.

Los resultados obtenidos no han mostrado ningún aspecto especialmente destacable o inesperado.

Durante la jornada se ha podido ver que la mayor parte de los accidentes ocurren a lo largo de la mañana, mientras que *in itinere* las horas con mayor número de accidentes son entre las 7 y las 9 de la mañana y las 2 y las 4 de la tarde. En el primer caso, el resultado coincide con las horas de trabajo más comunes, por lo que era de esperar que a esas horas la cantidad de accidentes fuese superior a la media general. En cuanto al caso *in itinere*, las horas con mayor número de accidentes también coinciden con las más comunes para acudir o abandonar el lugar de trabajo.

En cuanto a la clasificación por horas de la jornada laboral, de nuevo las horas con mayor número de accidentes coinciden con las que mayor número de trabajadores engloban. En este caso hay que considerar que no todos los puestos de trabajo cuentan con contratos de 8 horas al día, muchos tienen una duración menor; por lo que el número de trabajadores que pueden ser accidentados en la séptima, octava hora o más allá es menor que el número de empleados que potencialmente pueden sufrir un accidente a primera o segunda hora. Por tanto, que el número de accidentes sea mayor en las primeras horas de la jornada parece razonable si se tiene en consideración la situación expuesta.

La situación en cuanto a los días de la semana es similar. Durante el fin de semana el número de accidentes baja muy notablemente con respecto a los accidentes durante la semana (que van descendiendo a medida que esta avanza) y, en general, es durante el fin de semana cuando es menor el número de personas que se encuentran trabajando.

Análoga a la presentada es la situación en cuanto a los meses del año: en general solo se observan ligeras bajadas en el número de accidentes en los meses de febrero, abril, agosto y diciembre, que son los meses donde los trabajadores suelen coger sus vacaciones (o en el caso de febrero, simplemente tiene menos días que los demás meses). Es por eso que en estos 4 meses la cantidad de empleados en sus puestos de trabajo suele ser menor y entra dentro de lo razonable que tengan lugar menos accidentes.

Solo cabe destacar dos datos atípicos: la ya comentada caída en abril de 2020 y, solo con respecto a los accidentes *in itinere*, un gran aumento en enero de 2021.

A primera vista, los resultados y gráficas incluidas en el informe muestran una clara influencia de la temporalidad en el número de accidentes, pues resulta evidente que hay horas, días y meses donde el número de accidentes es bastante superior a la media. Sin embargo, analizando los resultados se ha podido observar que esas situaciones donde el número de accidentes laborales es mayor coinciden con momentos donde es esperable encontrar una mayor cantidad de trabajadores en sus puestos de trabajo. Por tanto, estos resultados tan destacables de forma general, se mimetizan con los demás al considerarlos de forma relativa.

Tras el análisis descriptivo de los datos, los resultados mostrados mediante el análisis de las series de tiempo han sido bastante representativos de lo ya comentado desde el inicio. En la serie a nivel estatal se pudo observar de nuevo la tendencia creciente de los últimos años, solo interrumpida por la caída en 2020. Además, se ha realizado un ajuste de la serie mediante un modelo ARIMA estacional mostrando la periodicidad anual (periodo  $s = 12$ ), lo cual ha permitido obtener predicciones a partir de ella para el año 2022. Dado que se dispone de los datos avanzados de ese mismo año, se ha podido estimar el error cometido por la predicción y los resultados obtenidos no distaron mucho de los avances proporcionados por el Ministerio de Trabajo. Dichos resultados muestran que el número de accidentes laborales continúa en aumento incluso tras la intervención de abril de 2020, alcanzando ya valores similares a los obtenidos en 2019. Reflejan, por tanto, una situación preocupante, a la que conviene poner freno, pues no se observan indicios de que el número de accidentes vaya a descender.

En las series a nivel provincial se puede ver también esa tendencia creciente. Aun presentando otras tendencias, como puede ser periódicas en algunos casos, en las series de todas las provincias se puede ver como a medida que avanzan los años, el número de accidentes va en aumento. Con la salvedad de nuevo del descenso provocado por la pandemia del COVID-19.

Sin embargo, las series correspondientes a las secciones y divisiones económicas no todas muestran esta tendencia. De todas formas, dado que el cómputo general demuestra crecimiento en el número de accidentes, aquellas secciones o divisiones en las que disminuyen los accidentes se sobre-compensan con aquellas que aumentan el número de accidentes.

A partir de la normalización de las series mencionadas, se ha procedido al agrupamiento de estas. Esto facilita la detección de patrones y tendencias similares entre las diferentes series, que ya han sido expuestas en el capítulo anterior. El descubrimiento de estos patrones puede resultar útil a la hora de planificar las estrategias a seguir por los expertos en PRL en las diferentes provincias y sectores económicos. Conociendo las épocas más problemáticas, se pueden reforzar las medidas preventivas aplicadas en los puestos de trabajo pertinentes o tratar de repartir las medidas y recursos ya existentes de forma óptima para que las épocas más conflictivas reduzcan el número de accidentes. Asimismo, el agrupamiento de las series es interesante porque relaciona provincias y divisiones económicas que presentan tendencias similares en cuanto al número de accidentes. Esto permite detectar planes de prevención que se pueden aplicar de forma simultánea a diferentes provincias o diferentes secciones o divisiones.

Con la intención de dar cierre a esta sección y al trabajo, se comentan a continuación las conclusiones más destacables presentadas en este último capítulo.

En primer lugar, resulta obvio que el resultado más claro y repetido ha sido el aumento de los accidentes a lo largo de estos últimos años. Resaltando además con las predicciones obtenidas a través del modelo ARIMA estacional que la tendencia creciente se espera que continúe. Tras el efecto de la pandemia en el año 2020, el volumen de accidentes laborales está al nivel del año 2019 o incluso superior. De todas formas, es importante tener en cuenta que este aumento de accidentes ha estado acompañado también por un aumento de la población ocupada en España.

Un resultado que también se ha repetido bastante es que aquellas categorías con mayor cantidad de accidentes se corresponden en general con un mayor número de población ocupada que se engloba dentro de esa categoría. Es por esta razón que se ha tratado de mostrar los resultados en función de la proporción de accidentes en lugar de números totales. Aunque de todas formas, por el simple hecho de que ciertas categorías engloben gran parte de la población ocupada, ya merecen una atención especial por parte de los especialistas de PRL.

Las categorías que, tras analizar tanto los números totales como las proporciones, han mostrado los resultados más alarmantes han sido los trabajadores jóvenes, los varones y los sectores de la construcción, la industria manufacturera y el comercio al por mayor y menor y reparación de vehículos. Estas secciones económicas están, en general, más asociadas a hombres, lo cual encaja con los resultados comentados. Teniendo esto en cuenta, quizás se debería priorizar entonces la reducción de accidentes en estas categorías, como idea para tratar de empezar a frenar ese crecimiento del número de accidentes y comenzar su descenso al igual que en la UE de forma global.

Como último comentario, cabe destacar que los resultados de un año a otro, para los 8 años considerados, a penas varían en cuanto a la distribución de los accidentes. En general, las proporciones se mantienen medianamente constantes; sin embargo, sí que varía el número total de accidentes anuales, tal y como se ha comentado a lo largo de todo el informe. Por tanto, los resultados presentados para 2021 son extrapolables al resto de años del estudio y a la situación general actual.

De cara al futuro, conviene continuar elaborando esta clase de informes. Conocer como evolucionan los resultados en cuanto al número de accidentes laborales es crucial de cara a la toma de decisiones por parte de los especialistas del PRL. Los resultados mostrados sirven como referencia de cara a modificar las medidas aplicadas en la prevención de riesgos en el ámbito laboral.

En vista de lo observado en este informe, quizás no sea necesario realizar todos los informes considerando periodos tan amplios de tiempo, pues los resultados son similares en general para todos los años. De todas formas, sí es conveniente realizar periódicamente informes de estas características, para vigilar cambios relevantes en la distribución de los accidentes a través de los años. El aspecto que sí es conveniente mostrar en todo caso es la evolución del número total de accidentes, para detectar si la tendencia creciente actual continúa aumentando o comienza a reducirse, como es deseable.



# Bibliografía

- [1] Aghabozorgi S. and Shirkorshidi A. S. and Ying Wah T, (2015) Time-Series Clustering - A Decade Review. Elsevier Science Ltd., 53(C), 16–38.
- [2] Brockwell, P. J. and Davis, R. A. (2016) Introduction to Time Series and Forecasting. Springer Cham.
- [3] Chan K-S. and Ripley B. (2022). TSA: Time Series Analysis. R package version 1.3.1.
- [4] Cowpertwait, P.S.P. and Metcalfe, A.V. (2009) Introductory Time Series with R. Springer.
- [5] Comisiones Obreras (2022) Análisis de las estadísticas de accidentes de trabajo y enfermedades profesionales en España en 2021. Confederación Sindical de CCOO.
- [6] Fabrice Yaméogo B., Gastineau P., Hankach P., and Vandanjon P.-O. (2021) Comparing Methods for Generating a Two-Layered Synthetic Population. Transportation Research Record, 2675(1), 136–147.
- [7] Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D. *et al.* (2020). Array programming with NumPy. Nature, 585, 357–362.
- [8] Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science and Engineering, 9(3), 90–95.
- [9] Hyndman R: (2020). fpp2: Data for “Forecasting: Principles and Practice” (2nd Edition). R package version 2.4.
- [10] Instituto Nacional de Seguridad y Salud en el Trabajo (INSST), O.A., M.P. (2021) Informe anual de accidentes de trabajo en España 2020. Servicio de Ediciones y Publicaciones del INSST.
- [11] Instituto Nacional de Seguridad y Salud en el Trabajo (INSST), O.A., M.P. (2020) ESENER (Encuesta europea de empresas sobre riesgos nuevos y emergentes) 2019. Datos de España. Servicio de Ediciones y Publicaciones del INSST.
- [12] Johnson R. A. and Wichern D. W. (2007) Introductory Time Series with R. Pearson.
- [13] McKinney, W. *et al.* (2010). Data structures for statistical computing in Python. In Proceedings of the 9th Python in Science Conference, 445, 51–56.
- [14] Petitjean F., Ketterlin A. and Gançarski P.(2011) A global averaging method for dynamic time warping, with applications to clustering. Pattern Recognition, 44(3), 678-693.
- [15] R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- [16] Sakoe H. and Chiba S.(1978) A global averaging method for dynamic time warping, with applications to clustering. IEEE Transactions on Acoustics, Speech, and Signal Processing, 44(3), 678-693.

- [17] Trapletti A. and Hornik K. (2022). tseries: Time Series Analysis and Computational Finance. R package version 0.10-52.
- [18] Van Rossum, G. and Drake Jr, F. L. (1995). Python reference manual. Centrum voor Wiskunde en Informatica Amsterdam.
- [19] Waskom M., Botvinnik O., O’Kane, D., Hobson, P. et al. (2017). mwaskom/seaborn: v0.8.1 (September 2017). Zenodo.
- [20] Zheng Li, Yue Zhao and Jialin Fu (2020) SynC: A Copula based Framework for Generating Synthetic Data from Aggregated Sources. International Conference on Data Mining Workshops (ICDMW), 571-578.