



Universidade de Vigo

Trabajo Fin de Máster

Análisis exploratorio de datos y generación de modelos predictivos del mercado de trabajo

Erik Martínez Sixto

Máster en Técnicas Estadísticas

Curso 2022-2023

Propuesta de Trabajo Fin de Máster

<p>Título en galego: Análise exploratoria de datos e xeneración de modelos predictivos do mercado de traballo</p>
<p>Título en español: Análisis exploratorio de datos y generación de modelos predictivos del mercado de trabajo</p>
<p>English title: Exploratory data analysis and generation of predictive models of the labour markets</p>
<p>Modalidad: Modalidad A</p>
<p>Autor/a: Erik Martínez Sixto, Universidade da Coruña</p>
<p>Director/a: Guillermo López Taboada, Universidade da Coruña</p>
<p>Breve resumen del trabajo:</p> <p>Este trabajo tiene como objetivo principal analizar y comprender el mercado laboral de Galicia al mismo tiempo que se desarrollan y se prueban una serie de herramientas para su análisis en esta u otras comunidades autónomas de España. La comprensión del mercado laboral es esencial para promover el crecimiento económico y la calidad de vida de los ciudadanos.</p> <p>Para ello se recopilaron datos de fuentes oficiales, creando con ellas bases de datos replicables y fácil acceso. Basándose en estos datos se analizaron diferentes aspectos del mercado laboral, como la proporción de población por su situación con respecto al mercado laboral (activos, parados e inactivos), los tipos de contratos, la distribución por sectores, etc. Los resultados revelan patrones interesantes, como una disminución en la proporción de contratos temporales, una fuerte estacionalidad en algunos sectores o la vuelta a niveles prepandémicos en la mayoría de indicadores. También se desarrolló un cuadro de mando en forma de aplicación shiny en R, el cual permite crear una amplia variedad de gráficas y resúmenes. Por último, se ajustaron modelos predictivos, identificando el mejor modelo para sectores especialmente estacionales, siendo los modelos SARIMAX y las redes neuronales artificiales los que mejor comportamiento han obtenido.</p> <p>En conclusión, este trabajo proporciona un análisis general del mercado laboral gallego y ofrece herramientas útiles para su estudio. Los resultados aun con sus limitaciones brindan una base sólida para futuras investigaciones, en esta u en otras comunidades autónomas de España.</p>
<p>Recomendaciones: Conocimientos de bases de datos SQL y NoSQL, y técnicas Machine Learning e IA, en particular con Spark, tanto en R como con Python.</p>
<p>Otras observaciones:</p>

Don Guillermo López Taboada, Catedrático de la Universidade da Coruña, informa que el Trabajo Fin de Máster titulado

Análisis exploratorio de datos y generación de modelos predictivos del mercado de trabajo

fue realizado bajo su dirección por don Erik Martínez Sixto para el Máster en Técnicas Estadísticas. Estimando que el trabajo está terminado, da su conformidad para su presentación y defensa ante un tribunal.

En Coruña, a 13 de Julio de 2023.



El director:
Don Guillermo López Taboada

El autor:
Don Erik Martínez Sixto

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

Resumen	IX
1. Introducción	1
1.1. Motivación	1
1.2. Objetivos	1
1.3. Estructura del trabajo	2
2. Construcción de los datasets	3
2.1. Encuesta de población activa	3
2.1.1. Datos de origen	3
2.1.2. Descarga de los datos	4
2.1.3. Limpieza y depuración	5
2.2. Afiliados a la seguridad social	5
2.2.1. Descarga de los datos	5
2.2.2. Limpieza y depuración	6
2.3. Datos complementarios	6
2.3.1. Datos del Producto Interior Bruto	6
2.3.2. Datos del Índice de Precios de Consumo	7
2.3.3. Datos de sobre eventos importantes	7
3. Cuadro de mando	9
3.1. Desarrollo del cuadro de mando	9
3.1.1. Cuadro de mando de datos de la EPA	10
3.1.2. Datos afiliados a la seguridad social	11
3.2. Descarga	11
4. Análisis descriptivo de los datos	13
4.1. Análisis de los datos de la EPA	13
4.1.1. Contexto español	13
4.1.2. Población y tipos de contrato. Galicia	19
4.1.3. Sector de ocupación. Galicia	23
4.1.4. Administración pública. Galicia	28
4.1.5. Autónomos. Galicia	29
4.2. Análisis de los datos de afiliados a la Seguridad Social	30
4.2.1. Contexto Español	30
4.2.2. Sexo y grupos de edad. Galicia	32
4.2.3. Sectores de ocupación. Galicia	33

5. Modelos predictivos	35
5.1. Modelos y técnicas empleadas	35
5.1.1. Fuerza de la estacionalidad	35
5.1.2. SARIMA	36
5.1.3. SARIMAX	37
5.1.4. Redes neuronales	37
5.2. Selección de sectores	37
5.3. Resultados	38
5.3.1. Modelos de afiliación para el conjunto de sectores	39
5.3.2. Modelos de afiliación para el conjunto de sectores. Covid	40
5.3.3. Modelos de afiliación para el sector servicios	41
5.3.4. Modelos de afiliación para el sector industria	42
5.3.5. Modelos de afiliación para el sector primario	43
6. Conclusiones	47
A. Lista de variables eliminadas de la EPA	51
B. Tabla resumen 1	53
C. Tabla resumen 2	55
D. Guía del cuadro de mando	57
D.1. Inicio	57
D.2. Análisis EPA	58
D.2.1. Temporalidad	58
D.2.2. Sector de ocupación	59
D.2.3. Horas	60
D.2.4. Trabajo público	61
D.2.5. Autónomos	63
D.3. Afiliados a la seguridad social	65
D.3.1. Por sexo y edad	65
D.3.2. Por sector de actividad	66
D.3.3. Serie temporal	67

Resumen

Este trabajo tiene como objetivo principal analizar y comprender el mercado laboral de Galicia al mismo tiempo que se desarrollan y se prueban una serie de herramientas para su análisis en esta u otras comunidades autónomas de España. La comprensión del mercado laboral es esencial para promover el crecimiento económico y la calidad de vida de los ciudadanos.

Para ello se recopilaron datos de fuentes oficiales, creando con ellas bases de datos replicables y fácil acceso. Con estos datos se analizaron diferentes aspectos del mercado laboral, como la proporción de población por su situación con respecto al mercado laboral (activos, parados e inactivos), los tipos de contratos, la distribución por sectores, etc. Los resultados revelan patrones interesantes, como una disminución en la proporción de contratos temporales, una fuerte estacionalidad en algunos sectores o la vuelta a niveles prepandémicos en la mayoría de estimadores. También se desarrolló un cuadro de mando en forma de aplicación shiny en R, el cual permite crear una amplia variedad de gráficas y resúmenes. Por último, se ajustaron modelos predictivos, identificando el mejor modelo para sectores especialmente estacionales, siendo los modelos SARIMAX y las redes neuronales artificiales los que mejor comportamiento han obtenido.

En conclusión, este trabajo proporciona un análisis general del mercado laboral gallego y ofrece herramientas útiles para su estudio. Los resultados aún con sus limitaciones brindan una base sólida para futuras investigaciones, en esta u otras comunidades autónomas de España.

English abstract

The main objective of this work is to analyze and understand the labor market of Galicia while developing and testing a series of tools to analyze this or other autonomous communities of Spain. A better understanding of the labor market is essential to promote economic growth and improve citizens' quality of life.

To achieve this, data has been collected from reliable sources, creating with them replicable and user-friendly databases. These data have been used to analyze different aspects of the labor market, such as population in relation to their situation in the labor market and contract types, labor sector distribution, and more. The results reveal interesting patterns, such as a decrease in the proportion of temporary contracts, strong seasonality in certain sectors, or a return to pre-pandemic levels in most indicators. Additionally, a user-friendly shiny dashboard application has been developed in R, allowing to create a wide variety of graphs and summaries. Furthermore, predictive models were adjusted, identifying the best model for particularly seasonal sectors, with SARIMAX models and artificial neural networks showing the best results.

In conclusion, this work provides a general analysis of the labor market in Galicia and offers useful tools for its study. Despite its limitations, the results provide a solid foundation for future research in this or other autonomous communities of Spain.

Capítulo 1

Introducción

1.1. Motivación

El mercado de laboral es una parte esencial de la estructura económica de cualquier país ya que el trabajo es la principal forma de creación de riqueza, bienes y servicios. En las últimas dos décadas hemos vivido transformaciones drásticas en las características y distribución en el mismo, impulsadas principalmente por avances tecnológicos como la automatización y la digitalización, que han traído consigo cambios significativos en cómo las personas trabajan y las necesidades del mercado laboral.

En este contexto de transformación, comprender cómo se están moldeando y adaptando los mercados laborales es fundamental. Por ello en este trabajo vamos a analizar cuáles son las fuentes principales de datos, qué información se puede sacar de ellas y algunas de las herramientas más interesantes para ello.

En este trabajo nos enmarcaremos principalmente en la comunidad autónoma de Galicia, una región con características y dinámicas propias, como la alta dispersión de la población, un progresivo envejecimiento, etc. Trataremos de entender cómo estos cambios están impactando en las ocupaciones predominantes, la empleabilidad y las condiciones laborales, la cual es información esencial de cara a poder tomar decisiones y desarrollar políticas específicas que impulsen el crecimiento económico, la equidad y la calidad de vida de los gallegos. Aunque no solo se trata de dar una imagen aproximada de las principales características del mercado laboral sino también de buscar datos de calidad sobre este ámbito y en desarrollar técnicas y tecnologías que faciliten el trabajo de cara a futuras investigaciones.

1.2. Objetivos

Para llevar a cabo este trabajo se han seguido una serie de objetivos clave creados a partir de las directrices de la propuesta del TFM. Aunque en la propuesta se planteaban tres grandes objetivos se decidió que de cara a la elaboración del trabajo se dividirían en cuatro, cada uno desarrollado en profundidad en su propio capítulo. Los objetivos son los siguientes:

1. **Construir un conjunto de bases de datos sobre el mercado laboral:** El primer objetivo es recopilar y crear un conjunto de bases de datos relevantes que proporcionen una visión amplia del mercado laboral en Galicia y al mismo tiempo permitan comparaciones con el resto de España. Estas bases de datos serán una herramienta fundamental para esta y futuras investigaciones.
2. **Realizar un análisis de los datos para caracterizar el mercado de trabajo:** A través del análisis exhaustivo de los datos recopilados, buscamos identificar patrones, tendencias y

características clave del mercado laboral en Galicia y ponerlos en contexto con respecto al resto de España.

- 3. Desarrollar una aplicación o cuadro de mando para explotar los datos recopilados:** Para facilitar la accesibilidad y utilidad de los datos recopilados, nos proponemos crear una aplicación o cuadro de mando que permita a los usuarios explorar y visualizar de manera interactiva la información relevante sobre el mercado laboral no solo en Galicia si no también en el resto de España. Esta herramienta proporcionará una interfaz intuitiva y amigable para analizar y comprender los datos de manera efectiva incluso para personas sin conocimientos amplios en estadística o programación.
- 4. Ajustar y comparar modelos predictivos sobre el comportamiento futuro del mercado laboral:** Nuestro último objetivo es desarrollar modelos predictivos que nos ayuden a comprender y anticipar el comportamiento futuro del mercado laboral en Galicia. Estos modelos se basarán en datos históricos y utilizarán técnicas avanzadas de análisis y pronóstico

Teniendo en cuenta que el principal objetivo de cualquier trabajo de fin de máster es mejorar las capacidades y conocimientos del alumno se ha buscado a mayores de estos objetivos principales mejorar y ampliar los conocimientos en distintos ámbitos de interés, como por ejemplo la obtención y explotación de datos reales, desarrollo de aplicaciones, ajuste de modelos de series temporales y un largo etc.

1.3. Estructura del trabajo

El trabajo esta compuesto de seis capítulos principales más anexos. Este capítulo 1 se centra en introducir las motivaciones del trabajo y sus objetivos. El capítulo 2 en la selección, descarga y proceso de creación de los datasets con información del mercado laboral. El capítulo 3 expondrá como se ha desarrollado, cuáles son sus principales funcionalidades y como descargar el cuadro de mando creado con R para dar a usuarios sin conocimientos de programación una opción sencilla para consultar algunas de las métricas y análisis más interesantes. El capítulo 4 expondrá el análisis de datos realizado sobre los datasets creados. El capítulo 5 está centrado en explicar el proceso de creación y selección de los mejores modelos predictivos de afiliados a la seguridad social. Por ultimo, el capítulo 6 se centra en las conclusiones del trabajo y las limitaciones y problemas que se han encontrado durante su desarrollo.

Capítulo 2

Construcción de los datasets

Este capítulo presenta las fuentes de datos usadas para el análisis de datos, la construcción del cuadro de mando y los modelos predictivos. El objetivo de los datasets es mejorar la usabilidad y acceso público de los datos. En primer lugar se describirán las fuentes y características principales de los datos de origen para después comentar cual ha sido su método de obtención y las transformaciones realizadas a través del lenguaje de programación R para obtener datasets en formato csv con codificación UTF-8. La colección de librerías *tidyverse* (Wickham et al., 2019) ha sido la principal herramienta para esta labor.

También cabe mencionar que se ha usado kaggle como repositorio de los distintos datasets creados. Kaggle (<https://www.kaggle.com/>) es la plataforma de ciencia de datos más extendida con más de 12 millones de usuarios y 50 mil datasets públicos. La cual ofrece de forma gratuita entornos de trabajo en Jupyter notebook dando acceso a una GPU de 4 cores y 30GB de RAM, pudiéndose usar en cada notebook 73GB de datos, lo cual seria equivalente a un servidor de tamaño pequeño o un portátil de gama alta. A mayores permite colgar de forma publica y sin coste grandes conjuntos de datos a los que puede accederse de forma gratuita solo con hacerse una cuenta, pudiendo descargarse o usarse directamente en una notebook.

2.1. Encuesta de población activa

La encuesta de población activa elaborada por el **INE(Instituto Nacional de Estadística)** es una de las principales fuentes de información sobre empleo y paro en España, por esta razón se ha seleccionado como una de las dos principales fuentes de información de esta investigación.

2.1.1. Datos de origen

Según el informe estandarizado (Instituto Nacional de Estadística, 2023) la **Encuesta de Población Activa**, también conocida por sus siglas **EPA**, es una encuesta elaborada de forma continua con periodicidad trimestral, que entrevista a la población que reside en viviendas familiares del territorio español. La EPA ofrece datos del mercado laboral de España desde 1964, aunque en esta investigación solo se usaron los datos que van de 2005 a finales de 2022. Esto debido a que durante una parte importante de desarrollo de este trabajo aun no habían liberado datos relativos a 2023, aunque a la fecha de entrega de este trabajo ya están liberados los datos del primer trimestre de 2023.

La encuesta ha sufrido múltiples modificaciones a lo largo del tiempo, tanto en su metodología como en otros aspectos, todo ello para poder adaptarse al cambiante mercado laboral español. Los cambios más significativos, en relación con esta investigación serían los de 2005 y los 2021. Siendo la modificación

2005 la que ha marcado las principales series y estimadores que en su mayoría se continuaron en la de 2021 (Instituto Nacional de Estadística, 2017, 2021b).

“La finalidad principal de la Encuesta de Población Activa (EPA) es conocer la actividad económica en lo relativo a su componente humano. Está orientada a dar datos de las principales categorías poblacionales en relación con el mercado de trabajo (ocupados, parados, activos, inactivos) y a obtener clasificaciones de estas categorías según diversas características. También posibilita confeccionar series temporales homogéneas de resultados. Por último, al ser las definiciones y criterios utilizados coherentes con los establecidos por los organismos internacionales que se ocupan de temas laborales, permite la comparación con datos de otros países” (Instituto Nacional de Estadística, 2021b, p. 6).

Para lograr este objetivo realizan un muestreo bietápico con estratificación de las unidades de primera etapa, a excepción de Ceuta y Melilla donde solo se hace el muestreo de una etapa, este ultimo dato no es relevante de cara a la parte de análisis de datos de esta investigación ya que el ámbito geográfico en el que nos centraremos en los siguientes capítulos es solo Galicia. Las unidades de primera etapa corresponderían a la secciones censales y las de segunda a las viviendas familiares principales y a los alojamientos fijos. Dentro de las viviendas se recoge información de todas las personas cuya residencia habitual sea esa vivienda. Para la estratificación de las unidades de muestreo se tiene en cuenta tanto criterios geográficos como socioeconómicos para así obtener una muestra lo más representativa de las características de la población (Instituto Nacional de Estadística, 2016, 2021a).

Con respecto al tamaño de la muestral hay que diferenciar aquí entre los datos que van de 2005 a 2021 y a partir de 2021, ya que aunque el número total de encuestados es similar no lo es el número de secciones censales, el cual ha ido aumentando en los últimos años para cumplir la normativa Europea. También es relevante mencionar que la muestra de Galicia es el doble de lo que le correspondería. Esto debido a un acuerdo firmado por el IGE (Instituto Galego de Estadística) y el INE (Instituto Nacional de Estadística, 2016, 2021a).

Por ultimo, hay que mencionar que aparte de las diferencias en la muestra también existen otros cambios entre los datos de antes y después de 2021, estos cambios están descritos en un documento de resumen que puede consultarse en Instituto Nacional de Estadística (2021c). El punto más relevante es que ciertos cambios en el cuestionario hicieron que algunas variables fuesen incorporadas y otras descontinuadas de un periodo a otro. Esto se comentará más adelante conjuntamente con la solución que se ha aplicado de cara a poder usar toda la serie temporal al completo.

2.1.2. Descarga de los datos

Los microdatos de la EPA están disponibles en Instituto Nacional de Estadística (2023a), desde esta web se pueden ir seleccionando los distintos resultados de las encuestas por trimestre como un fichero .zip el cual tiene los resultados en distintos formatos. Para simplificar el proceso y evitar posibles errores se programó una función en R para descargar los datos a desde la web, descomprimiese el archivo, cargase el CSV con los datos y los uniese en un único archivo CSV para el periodo que va de 2005 a 2020 (**EPA_05_20.csv**) y otro para el periodo que va del 2021 a 2022(**EPA_21_22.csv**).

A cada uno de estos datasets se les incluyo una variable para representar el año (*year*) y otra para representar el trimestre (*trim*), llamándose estos nuevos archivos **EPA_05_20_v2.csv** y **EPA_21_22_v2.csv**. Cabe mencionar que esta permitido el uso y distribución de los datos de la EPA siempre que no se desnaturalice su significado original y se referencie la fuente de origen, como se ha hecho en Kaggle a la hora de subir cada uno de los datasets. Pueden consultarse en los siguientes enlaces:

- <https://www.kaggle.com/datasets/erikmartinezsixto/epa-05-20>
- <https://www.kaggle.com/datasets/erikmartinezsixto/epa-21-22>

2.1.3. Limpieza y depuración

En esta paso se unificaron las versiones con año y trimestre de los datasets anteriores, creando **EPA_05_22.csv**, la cual contendría los datos desde el primer trimestre de 2005 hasta el cuarto de 2022. Posteriormente eliminamos aquellas variables que no tuvieran continuidad entre periodos y algunas de poco interés, creando así **EPA_05_22_MOD.csv**. Puede consultarse la lista de variables eliminadas en el [Apéndice A](#) o en la documentación del dataset en el enlace a Kaggle del final de este apartado.

Por último se crearon dos nuevas variables a partir de **EPA_05_22_MOD.csv**, que son *extraprov* que es una variable binaria que es 1 cuando el encuestado trabaja fuera de la provincia donde reside y cero cuando no, y se crea la variable *comu* para representar a la comunidad autónoma donde reside el entrevistado y que así sea más sencillo filtrar los datos. Este dataset se llama **EPA_05_22_MOD1.csv**. Se puede acceder a los 3 datasets de esta sección en el siguiente enlace:

- <https://www.kaggle.com/datasets/erikmartinezsixto/epa-05-22-mod>

A mayores existe un cuarto dataset basado en **EPA_05_22_MOD1.csv** llamado **EPA_app** el cual es una versión reducida que se usa en la aplicación de cuadro de mando creada para este trabajo. El cual solo está disponible a través del enlace de descarga de la aplicación de cuadro de mando del capítulo 3.

2.2. Afiliados a la seguridad social

La Tesorería General de la Seguridad Social proporciona a través de su herramienta online PX-Web (<https://w6.seg-social.es/PXWeb/pxweb/es/>) acceso a una serie de datos sobre afiliados medios mensuales a la seguridad social en situación de alta. Se trata de una información muy relevante de cara a analizar el mercado laboral.

En el informe estandarizado de [Treasorería General de la Seguridad Social \(2023a\)](#) se advierte que el número de afiliados no representa necesariamente el número de trabajadores, ya que solo se contabilizan aquellas relaciones laborales que exigen estar afiliado por ley. Por ejemplo, aquellos funcionarios que tengan cubiertas las contingencias sociales por MUFACE, ISFAS o MUGEJU no estarían incluidos entre los afiliados, cabe remarcar que paulatinamente estos casos acabarán desapareciendo ya que a partir de 2011 todos los funcionarios de nuevo ingreso están incluidos en el régimen general de la Seguridad Social, y todos los mutualistas acabarán jubilándose. También tiene un fuerte efecto sobre los datos que se contabilicen todas y cada una de las relaciones que exige la ley, por tanto una misma persona con dos trabajos tendría dos relaciones laborales y se contabilizaría dos veces. Sin embargo el uso del número de afiliados para medir el comportamiento y el buen estado del mercado laboral está ampliamente aceptado y se puede ver en multitud de investigaciones como por ejemplo [Maruri et al. \(2008\)](#) o en [Braña \(2007\)](#).

2.2.1. Descarga de los datos

La aplicación PX-Web:

“difunde información en formato de series temporales de filiación de trabajadores en media mensual y a fin de mes para las variables de clasificación: sexo, tramos de edad, nacionalidad, régimen, provincia, CCAA, tipo de contrato y actividad económica, mediante la consulta de tablas estadísticas multidimensionales (cubos), permitiendo al ciudadano personalizar las consultas, a través de la selección de las variables y los valores de estas, en las que esté interesado, obteniendo una mayor operatividad.” ([Treasorería General de la Seguridad Social, 2023b](#), p. 1)

La aplicación por tanto permite descargar los datos según distintas segregaciones, las elegidas para este trabajo han sido afiliados por provincia, sexo y grupo de edad y afiliados por provincia, régimen y Sección CNAE. La primera de ellas para hacer análisis exploratorio de los datos y la segunda para hacer análisis y modelos predictivos centrándonos en los sectores de ocupación con mayor estacionalidad. El primer dataset es **AfiliadosMedios_12_23_socio.csv** para el cual tenemos datos que van desde enero de 2012 hasta marzo de 2023 y el segundo es **AfiliadosMedios_09_23_sector.csv** que va desde enero de 2009 hasta marzo de 2023. Ambos están subidos a kaggle.

2.2.2. Limpieza y depuración

En este caso en los datos originales tenían algunas variables en formato carácter y otras en formato numérico por tanto primero se han factorizado las variables en formato carácter en ambos csv, y a mayores se les han añadido una variable año, mes y comunidad autónoma a la que pertenecen. En el caso de **AfiliadosMedios_09_23_sector.csv** también se ha creado una nueva variable llamada *act_pib* la cual recoge la categoría CNAE en base al sistema de agrupación que tienen los datos del PIB que se comentaran en la siguiente sección.

De esta forma se crearían los csv **AfiliadosMedios_12_23_Socio_Mod.csv** y **AfiliadosMedios_12_23_sector_Mod(2).csv**. Cabe mencionar que esta permitido el uso y distribución de los datos del número de afiliados a la Seguridad Social siempre que no se desnaturalice su significado original y se referencie la fuente de origen, como se ha hecho en Kaggle a la hora de subir cada uno de los datasets. Los cuatro archivos de datos de esta sección se pueden consultar y descargar en el siguiente enlace:

- https://www.kaggle.com/datasets/erikmartinezsixto/afiliados-seguridadsocial-spain?select=Afiliados+medios_09_23_sector.csv

2.3. Datos complementarios

A mayores de estos dos conjuntos de datos principales se han usado una serie de datos complementarios principalmente durante la generación de los modelos predictivos.

2.3.1. Datos del Producto Interior Bruto

El **PIB (Producto Interior Bruto)** es una de las variables predictoras usada en los modelos. Por ello se usa solo el PIB de Galicia. Como indica el [Instituto Galego de Estatística \(2022\) \(IGE\)](#) el PIB es el resultado de las actividades productivas. Se estima de forma anual y trimestral a partir de distintas fuentes como el INE, informes estadísticos de los distintos ministerios, etc. El IGE estima el PIB por tres vías:

- **Oferta:** la cual es la suma del valor agregado bruto de los diferentes sectores de actividades más los impuestos netos sobre las subvenciones de los productos.
- **Demanda:** es la suma de la demanda final de bienes y servicios más el saldo exterior de los mismo.
- **Rentas:** serían la suma de las remuneraciones de los asalariados, de los impuestos sobre la producción, las rentas mixtas, las importaciones netas de subvenciones y el excedente de explotación.

Las tres estimaciones obtienen resultados prácticamente idénticas en el conjunto de años seleccionados para la predicción, de 2009 a 2022. Por tanto es irrelevante cual de las tres escoger. En el caso de este trabajo se ha seleccionado la estimación basada en la oferta creando así el archivo **PIB_Galicia.csv**, el cual solo contiene dos variables una que representa el periodo y otra con el dato del PIB en miles de euros. Para facilitar su uso se hicieron un serie de modificaciones; se crearon un variable año y otra trimestre; como los datos de afiliados son mensuales para armonizar ambos datasets se asumió que dentro de cada trimestre el PIB se mantenía constante y sin variación, para ello se incluyeron nuevos casos para poder representar todos los meses de año; y por ultimo una variable mes. Este archivo se nombró como **pib_galicia_mod.csv**.

Para las predicciones del número de afiliados en ciertos sectores de actividad se considera que el PIB por sí solo sería demasiado general ya que no todos los sectores producen y tienen beneficios al mismo nivel. Por ello se descargó también los datos del *VEB* (Valor Engadido Bruto/Valor añadido bruto) el cual es uno de los componentes de la estimación de la oferta y está segregado por sector de actividad, el archivo **PIB(GPD)_sector_galicia.csv** contiene estos datos. Como en el caso anterior también se creó un archivo modificado añadiendo, año, trimestre y sector de actividad siguiendo la clasificación que se usa en los datos de afiliados. Como en el caso anterior los datos son trimestrales pero se añaden nuevos casos suponiendo que el *VEB* dentro del trimestre se mantiene constante y sin variación.

Se puede acceder a los datos originales desde [Instituto Galego de Estatística \(2023\)](#) o desde el siguiente enlace a Kaggle el cual contiene tanto los datos originales como los modificados:

- <https://www.kaggle.com/datasets/erikmartinezsixto/pibgpd-spain>

2.3.2. Datos del Índice de Precios de Consumo

El **IPC**(Índice de Precios de Consumo) es uno de los índices económicos más importantes, cuyo objetivo es “*medir la evolución de los precios de los bienes y servicios de consumo adquiridos por los hogares residentes en España*” ([Instituto Nacional de Estadística, 2023](#)). Aunque la forma de calcular el IPC ha ido cambiando a lo largo del tiempo, principalmente debido a que su calculo se basa en la medición de los precios medios de una cesta de la compra habitual de las familia españolas, y esta cesta habitual no es un concepto constante si no cambiante. Sin embargo para paliar esto y que el IPC represente lo más fiel posible la realidad económica el INE toma medidas como seleccionar siempre los artículos más consumidos, comprobar los establecimientos más visitados y comprobar la relevancia relativa de cada producto en la cesta.

Aunque solo se van a usar los datos del índice el archivo **IPC_spain.csv** contiene información también de la variación mensual y anual. Los datos del IPC son de carácter mensual y se pueden descargar segregados por provincia o por comunidad autónoma, en este caso se han seleccionado los datos por comunidad autónoma.

A partir de los datos originales se ha creado un archivo **IPC_spain_mod(1).csv** el cual contiene nuevas variables que son principalmente recodificaciones de variables ya existentes pero que estaban en formato carácter, aunque también se han añadido variables de año, mes y fecha. Ambos archivos pueden ser consultados desde el siguiente enlace:

- <https://www.kaggle.com/datasets/erikmartinezsixto/ipccpi-spain>

2.3.3. Datos de sobre eventos importantes

A mayores para los modelos predictivos se crearon dos variables binarias para marcar momentos temporales concretos.

- *santa*: representa si el mes en cuestión tiene Semana Santa o no. En los caso en que la semana se encuentre en un periodo entre dos meses se indicara en el mes en el que comienza la semana la semana santa. Datos obtenidos de [Luis Coronilla \(2023\)](#).
- *covid*: representa el momento temporal de influencia de las políticas de estado de alarma a causa de la pandemia de COVID-19, la más importante de estas medidas sería el confinamiento nacional que impedía salir del domicilio, excepto por causa de fuerza mayor, a la mayoría de la población. El principal problema para definir esta variable es que las medidas tomadas variaron bastante a lo largo del tiempo, aunque el confinamiento terminó en Galicia el 15 de junio todavía hubo algunas medidas que se mantuvieron en el tiempo. Por tanto la decisión que se ha tomado es considerar los meses que van desde marzo de 2020 hasta diciembre de 2020 como meses COVID.

Capítulo 3

Cuadro de mando para el análisis descriptivo de datos

Unos de los objetivos principales de esta investigación es crear un cuadro de mando de inteligencia de negocio, una herramienta pensada para la visualización y análisis de datos relevantes en el estudio del mercado laboral. Esta herramienta se ha desarrollado en el lenguaje R con el uso de las librerías *shinydashboard* (Chang et al., 2022), *tidyverse* (Wickham et al., 2019), *plotly* (Sievert, 2020) y *seasonal* (Sax, 2018) entre muchas otras.

3.1. Desarrollo del cuadro de mando

Para su desarrollo se usaran más de 15 librerías distintas con funciones tan dispares como generar tablas visualmente más atractivas para el usuario o poder usar variables de carácter temporal de forma más cómoda. Se hará referencia a las mas relevantes en este apartado

El cuadro de mando esta construido sobre la librería *shinydashboard* (Chang et al., 2022) la cual permite crear de forma sencilla un cuadro de mando a través de *Shiny* (Wickham, 2021). Shiny a su vez es un framework cuya propósito es crear aplicaciones web usando solo R.

Para el filtrado de datos y la creación de las tablas se ha usado *tidyverse* (Wickham et al., 2019) el cual es una colección de librerías especialmente diseñada para la ciencia de datos. Dentro de ella se encuentra *ggplot2* librería especializada en la creación de gráficas, la cual se uso junto a *plotly* (Sievert, 2020) para crear una gráficas interactivas y que el usuario pueda descargarlas como imágenes png.

También es relevante mencionar *seasonal* (Sax, 2018) la cual es una librería que da una interfaz en R para usar la tecnología X-13ARIMA-SEATS, la cual es una herramienta pensado para el análisis de series temporales con ajuste estacional. Dando acceso a una amplia variedad de modelos de análisis y predicción de series temporales. En este caso su uso principal sera realizar un análisis de descomposición estacional seleccionando automáticamente los parámetros sin intervención del usuario.

El cuadro de mando dentro de este trabajo tiene dos objetivos principales, por un lado servir como herramienta de trabajo en el capítulo 4 de análisis de datos, y por otro tener una herramienta simple para analizar datos del mercado laboral. En el análisis de datos solo se ha aprovechado una pequeña parte de la información que se puede obtener a través del cuadro de mando, el cual permite analizar información de la EPA y de afiliados a la seguridad social de forma más extensa y en otros ámbitos geográficos de España.

Durante el desarrollo tuvieron que quedarse varias ideas fuera, e incluso algunas de las gráficas y análisis del capítulo 4 no han podido ser incorporadas, aunque sí su mayoría. La principal razón para ello es el peso de los datos de la EPA. Por ejemplo el archivo EPA_05_22_MOD1.csv, que es el que se uso en el inicio del desarrollo de cuadro de mando, pesa sobre 3,5 GB y esta compuesto por más de 100 variables y sobre 10 millones de casos. Para poder reducir los tiempos de carga y ejecución se tomo la decisión de usar 16 de las variables y filtrar para usarse solo la población de Ocupados. Los datos de la EPA tienen mucha más información explotable de la que se ha podido usar durante el análisis de datos o la creación del cuadro de mando.

En las siguiente secciones se presenta de forma pormenorizada cuales son la utilidades del cuadro de mando. Adicionalmente, el apéndice D contiene una guía de uso del cuadro de mando, con capturas de pantalla sobre las cuales se explica como funciona la interfaz de usuario y todas las posibilidades del cuadro de mando.

3.1.1. Cuadro de mando de datos de la EPA

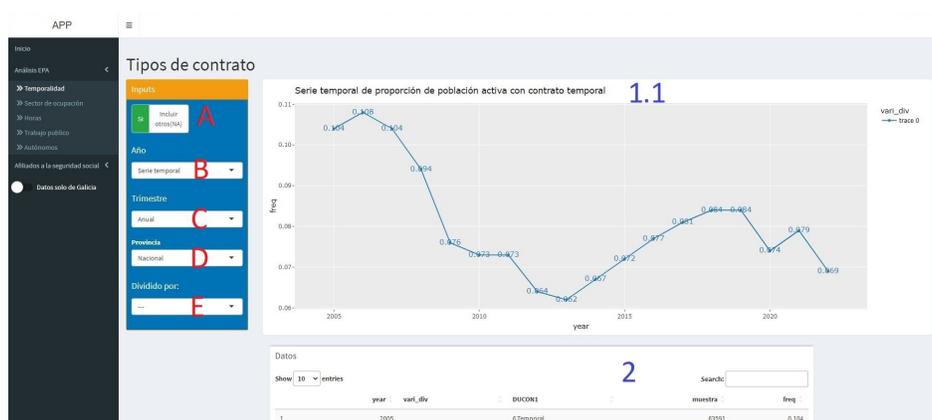


Figura 3.1: Pestañas de los datos de la EPA
Fuente: Captura de pantalla del cuadro de mando

Vamos a profundizar un poco en los análisis y gráficos programados en el cuadro de mando que usan como fuente de datos la EPA. Este apartado esta compuesto por 5 pestañas:

- **Tipos de contrato:** genera gráficas y tablas sobre la proporción de tipos de contrato de la población Ocupada, pudiendo filtrar por año, trimestre o provincia y pudiendo segregar por grupos de edad, sexo, nacionalidad, nivel de estudios o sector de actividad.
- **Sector de ocupación:** genera gráficas y tablas sobre la proporción de Ocupados en base al sector laboral, pudiendo usar los dos sistemas de clasificación que tiene incorporado el INE, filtrar por año, trimestre o provincia y pudiendo segregar por grupos de edad, sexo, nacionalidad o nivel de estudios.
- **Horas:** mapa con la media o mediana de horas por provincia tanto de contrato como las que el trabajador declara trabajar, se puede filtrar por año y trimestre.
- **Trabajo público:** pestaña resumen donde se agrupan diferentes gráficas sobre los trabajadores del sector público, como el porcentaje de trabajadores por tipo de administración y nivel de estudios o el porcentaje que representa sobre el resto de Ocupados, entre otras.

- **Autónomos:** pestaña resumen donde se agrupan diferentes gráficas sobre los trabajadores autónomos, como el porcentaje de trabajadores por tipo de sector de actividad y nivel de estudios o el porcentaje que representa sobre el resto de Ocupados, entre otras.

3.1.2. Datos afiliados a la seguridad social

- **Por sexo y edad:** serie temporal del número de afiliados pudiendo filtrar por sexo y grupos de edad.
- **Por sector de actividad:** serie temporal del número de afiliados pudiendo filtrar por sector de actividad y régimen de la seguridad social.
- **Serie temporal:** Gráficas de la serie temporal de afiliados por sector, junto con su descomposición con selección automática de parámetros (tendencia, error y estacionalidad)

3.2. Descarga

La aplicación esta disponible para su descarga a través de google drive como archivo .zip. Pasos para su descarga:

1. Bajar el archivo .zip *APP_ErikMartinez* en el siguiente enlace: <https://drive.google.com/drive/folders/1A5WieKlwUtdG01veMG1YSbVf4SuFrdCY?usp=sharing>
2. Descomprimir el archivo.
3. Abrir el archivo *app.R*.
4. Asegurarse de que el directorio de trabajo sea el mismo que en donde esta el archivo *app.R* y la carpeta *R*.
5. Ejecutar el archivo *app.R*.

La aplicación requiere una cantidad importante de recursos, especialmente al cargarse por primera vez, ya que de no estar ya instaladas, se requieren un buen numero de librerías. Una vez cargado la primera gráfica, y por tanto disponer de una copia de los datos de la aplicación en memoria pasa a ser más fluido su uso.

Con el fin de facilitar la evaluación de la herramienta se ha desarrollado una versión que accede a un subconjunto limitado, el 10% de los datos originales, de modo que se puede verificar el funcionamiento del cuadro de mando. Además, no es necesario, con esta versión, instalar la aplicación, ya que está disponible en un entorno de ejecución en modo Software-as-a-Service en la nube. Por tanto esta versión solo sirve para ver de forma rápida las utilidades generales de la herramienta.

https://skgxyl-erik-martinez.shinyapps.io/CuadroDeMandos_TFM_ErikMartinez/

Capítulo 4

Análisis descriptivo de los datos

Uno de los objetivos principales de este trabajo es realizar un análisis exploratorio caracterizando el mercado de trabajo gallego a través de variables de interés. En este capítulo profundizaremos en los análisis y resultados obtenidos. El lenguaje de programación usado para este capítulo vuelve a ser R con el uso de la colección de librerías de R *tidyverse* (Wickham et al., 2019), con mención especial a la *ggplot2* (Wickham, 2016) que ha sido la librería con la que se han creado la mayoría de gráficas de este capítulo.

4.1. Análisis de los datos de la EPA

Como ya se comentó en el capítulo 2 la EPA es una de las más importantes fuentes de información sobre el mercado de trabajo español y por consiguiente de sus comunidades autónomas.

4.1.1. Contexto español

Este apartado se centrará en hacer un breve análisis del mercado laboral español, aunque el ámbito geográfico de este trabajo es la comunidad autónoma de Galicia, tener el contexto del resto de país permitirá hacer comparaciones y entender en mayor profundidad las diferencias y similitudes entre Galicia y el conjunto de España.

Tipos de población y contratos

La EPA clasifica a todos sus encuestados en 7 categorías:

- Ocupados desempleados por insuficiencia de horas
- Resto de Ocupados
- Parados que buscan primer empleo
- Parados que han trabajado antes
- Inactivos 1 (desanimados)
- Inactivos 2 (junto con los desanimados forman los activos potenciales)
- Inactivos (resto de inactivos)

A partir de esta clasificación vamos a agrupar las primeras dos categorías como Ocupados, que serían aquellas personas mayores de 16 años con un empleo por cuenta ajena o trabajadores por cuenta propia; las dos siguientes categorías serían las relativas a los Parados, que serían aquellas personas mayores de 16 años sin trabajo, pero que se encuentran disponibles para trabajar y están en búsqueda activa de empleo o a la espera de incorporarse a uno; y por último las tres categorías restantes corresponderían a los Inactivos, que son aquellos mayores de 16 años que no pueden ser clasificados ni como ocupados ni como parados. (Instituto Nacional de Estadística, 2023)

Como se puede ver en la gráfica 4.1 la distribución es de alrededor un 45% de Ocupados, un 45% de Inactivos y sobre un 10% de parados. Evidentemente estos números van variando de un año a otro siendo en 2022 una 47,4% Ocupados, un 45,9% Inactivos y un 6,7% son Parados. La proporción de inactivos se ha mantenido bastante constante por los últimos 17 años y ha sido la proporción de parados la que más ha ido variando a lo largo del tiempo.

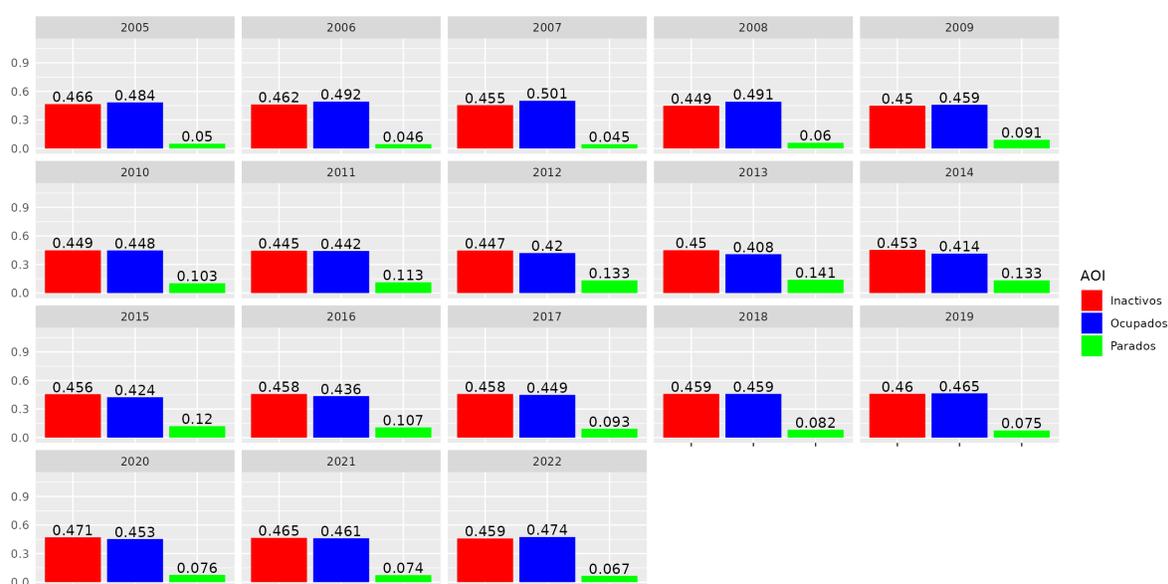


Figura 4.1: Proporción de población según AOI (si son Inactivos, Ocupados o Parados). España
Fuente: Elaboración propia a partir de datos de la EPA (INE)

Otro punto relevante es el tipo de contratos que se firman ya que tener un contrato temporal o uno indefinido como dice De la Rica (2010) tiene fuertes repercusiones en la vida de las personas, desde salarios más bajos, poco ahorro, dificultades en el acceso a la compra de vivienda, e incluso menores tasas de maternidad entre aquellas mujeres que tienen este tipo de contratos.

En la gráfica 4.2 se puede observar la proporción de Ocupados con contratos indefinidos o temporal. Hay que remarcar que dentro de los Ocupados no todos tienen contrato laboral, por ejemplo los que trabajan por cuenta propia o en una empresa familiar, pero en la gráfica 4.2 solo se está teniendo en cuenta a los trabajadores con contrato de asalariado. Observando la gráfica vemos una paulatina bajada de contratos temporales desde 2005 en adelante obteniendo el mínimo de la serie en 2013, aunque esto podría interpretarse como algo positivo si revisamos la gráfica 4.1 vemos que los años con menor porcentaje de temporalidad son también los años con mayor paro. Por tanto lo que estamos observando es probablemente que los principales afectados por despidos tras la crisis económica del 2008 eran aquellos con contrato temporal. La tendencia decreciente también aparece a partir de 2019 y en este caso sí que coincide con una proporción de parados estancada o decreciente probablemente

asociada a la situación excepcional que fue la crisis del COVID-19 y en 2022 la reforma laboral que trataba de limitar el uso tan extendido de los contratos temporales.

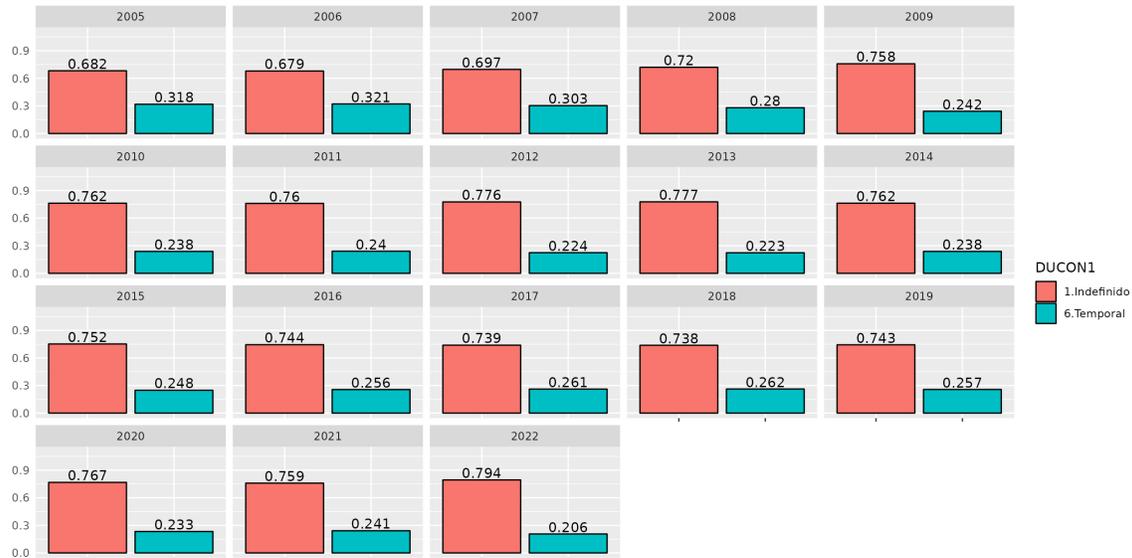


Figura 4.2: Proporción de tipos de contrato DUCON1 (indefinido o temporal). España

Fuente: Elaboración propia a partir de datos de la EPA (INE)

La EPA también tiene información sobre los contratos permanentes frente a los discontinuos. Como se puede observar en la gráfica 4.3 los contratos discontinuos representan una proporción minúscula del conjunto de contratos que son de carácter permanente, aunque parece haber habido un ligerísimo aumento en 2022.

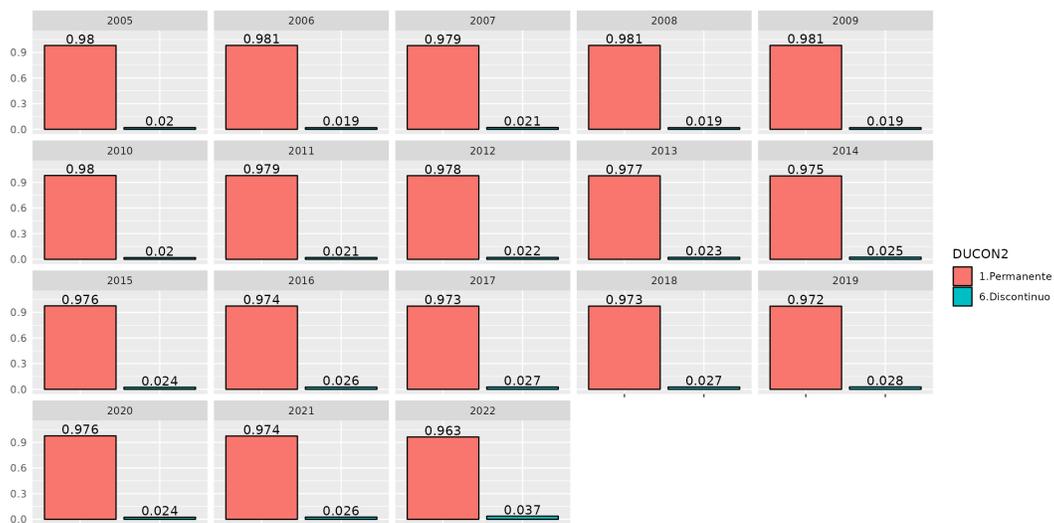


Figura 4.3: Proporción de tipos de contrato DUCON2 (permanente y discontinuo). España

Fuente: Elaboración propia a partir de datos de la EPA (INE)

Sector de ocupación

Otro punto relevante a la hora de analizar el mercado laboral es la distribución de trabajadores por sectores laborales. La EPA tiene dos variables para dividir a los Ocupados, por un lado estaría OCUP1 basado en la Clasificación Nacional de Ocupaciones 2011 (CNO-2011) y por el otro ACT1 basado en la Clasificación Nacional de Actividades Económicas 2009 (CNAE-2009). Pueden consultarse los categorías de las variables en la tabla 4.1. En este capítulo y en los siguientes nos vamos a centrar en los sectores a partir de la variable ACT1, la razón es que el sistema CNAE es el usado para segmentar el número de afiliados por sector de actividad, aun así la comparación directa entre ambas bases de datos es compleja ya que, aunque ambas usen el sistema CNAE la forma de agrupar los sectores es distinto. En el cuadro de mando creado en este trabajo es posible usar los dos tipos de clasificación en las tablas y gráficas.

Comenzaremos identificando cuales son los sectores con mayor proporción de Ocupados, la lista completa con los tres sectores con mayor proporción de ocupados por año puede consultarse en el Apéndice B. De forma consistente los sectores que ocupan los dos primeros puestos y representando alrededor del 50% de los Ocupados serian el *Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería* y la *Administración Pública, educación y actividades sanitarias*. Seguidas desde el 2005 hasta el 2008 por la *Construcción* y desde 2009 en adelante por el sector de la *Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros* siendo alrededor de una 12% de los ocupados.

En el caso del sector del *Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería* estaríamos hablando del sector servicios, y en el caso de *Administración Pública, educación y actividades sanitarias* sería lo que solemos denominar sector público o administración aunque no representan de forma perfecta estos dos sectores ya que por ejemplo las actividades sanitarias privadas como puede ser una clínica dental también entraría en esta categoría. Aun con sus limitaciones se usarán estas formas abreviadas a lo largo del trabajo.

OCUP1	ACT1
Ocupaciones militares (códigos CNO-2011). Fuerzas armadas (códigos CNO-1994)	Agricultura, ganadería, silvicultura y pesca (códigos CNAE-09: 01, 02 y 03), (códigos CNAE-93: 01, 02 y 05)
Directores y gerentes (códigos CNO-2011). Dirección de las empresas y de las Administraciones Públicas (códigos CNO-1994)	Industria de la alimentación, textil, cuero, madera y papel (códigos CNAE-09: del 10 al 18), (códigos CNAE-93 del 15 al 22)
Técnicos y Profesionales científicos e intelectuales (códigos CNO-2011)	Industrias extractivas, refino de petróleo, industria química, farmacéutica, industria del caucho y materias plásticas, suministro energía eléctrica, gas, vapor y aire acondicionado, suministro de agua, gestión de residuos. Metalurgia (códigos CNAE-09: del 05 al 09, del 19 al 25, 35 y del 36 al 39), (códigos CNAE-93: del 10 al 14, del 23 al 28, 40 y 41)
Técnicos y Profesionales de apoyo (códigos CNO-2011)	Construcción de maquinaria, equipo eléctrico y material de transporte. Instalación y reparación industrial (códigos CNAE-09 del 26 al 33), (códigos CNAE-93 del 29 al 37)
Empleados contables, administrativos y otros empleados de oficina (códigos CNO-2011). Empleados de tipo administrativo (códigos CNO-1994)	Construcción (códigos CNAE-09: del 41 al 43), (código CNAE-93: 45)
Trabajadores de servicios de restauración, personales, protección y vendedores de comercio (códigos CNO-2011)	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería (códigos CNAE-09: del 45 al 47, 55 y 56), (códigos CNAE-93: 50, 51, 52 y 55)
Trabajadores cualificados en el sector agrícola, ganadero, forestal y pesquero (códigos CNO-2011). Trabajadores cualificados en la agricultura y en la pesca (códigos CNO-1994)	Transporte y almacenamiento. Información y comunicaciones (códigos CNAE-09 del 49 al 53 y del 58 al 63), (códigos CNAE-93 del 60 al 64)
Artesanos y trabajadores cualificados de las industrias manufactureras y la construcción (excepto operadores de instalaciones y maquinaria (códigos CNO-2011). Artesanos y trabajadores cualificados de las industrias manufactureras, la construcción, y la minería, excepto operadores de instalaciones y maquinaria (códigos CNO-1994)	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros (códigos CNAE-09: del 64 al 66, 68, del 69 al 75 y del 77 al 82), (códigos CNAE-93 del 65 al 67 y del 70 al 74)
Operadores de instalaciones y maquinaria, y montadores (códigos CNO-2011)	Administración Pública, educación y actividades sanitarias (códigos CNAE-09: 84, 85 y del 86 al 88), (códigos CNAE-93: 75, 80 y 85)
Ocupaciones elementales (códigos CNO-2011). Trabajadores no cualificados (códigos CNO-1994)	Otros servicios (códigos CNAE-09: del 90 al 93, del 94 al 96, 97 y 99), (códigos CNAE-93: del 90 al 93, 95 y 99)

Cuadro 4.1: Tabla con los valores de las variables OCUP1 y ACT1 de la EPA. **Fuente:** EPA (INE)

Sector Público

Vamos a analizar de forma pormenorizada el sector público español viendo cual es el porcentaje de trabajadores públicos sobre el total de trabajadores, la proporción de ellos con contrato temporal, las horas medias de trabajo y cuál es el porcentaje de trabajadores públicos por tipo de administración. Para ello seleccionaremos solo a los asalariados del sector público los cuales a su vez son parte del sector de *Administración Pública, educación y actividades sanitarias*, pero no representan al total de trabajadores del sector. Revisaremos solo los tres últimos años aunque en el cuadro de mando desarrollado para este trabajo es posible observar todos los años desde 2005 segmentando también por provincia, trimestre y sexo.

En primera instancia vamos a revisar la tabla 4.2 la cual contiene un resumen con los datos más interesantes sobre el sector público en los tres últimos años. El primer dato relevante es el porcentaje de trabajadores que pertenecen al sector público, que como se puede observar ha sido de al rededor de un 20 % del total de Ocupados. Con respecto al porcentaje de trabajadores con contratos temporales vemos que aunque varia bastante de un año a otro parece que se esta aproximando al 30 % de los asalariados del sector público. El numero de horas de trabajo semanal es de alrededor de 37 horas.

Año	Porcentaje de trabajadores del sector público	Porcentaje de trabajadores con contrato temporal	Horas semanales de trabajo (Declaradas por el encuestado)
2020	20,3 %	26,9 %	36,36 h
2021	20,5 %	30 %	37.51 h
2022	19,9 %	30.8 %	38.02 h

Cuadro 4.2: Tabla resumen sobre trabajadores del sector público.

Fuente: Elaboración propia a partir de datos la EPA (INE)

El la gráfica 4.4 se observa el porcentaje de trabajadores público por tipo de administración. Siendo las comunidades autónomas las que más trabajadores concentran, seguidas por las administraciones locales y finalmente por las administraciones centrales. Es evidente que como la mayoría de competencias del estado están transferidas a las comunidades autónomas son estas las que tienen el mayor número de empleados a su cargo.

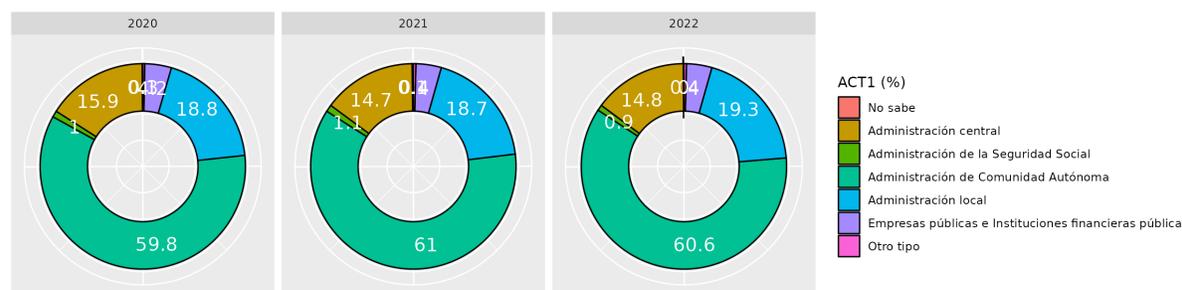


Figura 4.4: Porcentaje de trabajadores del sector público por tipo de administración. España

Fuente: Elaboración propia a partir de datos de la EPA (INE)

Autónomos

Otro punto relevante en el mercado laboral son los Autónomos o como los denomina la EPA, *Trabajador independiente o empresario sin asalariados*. Al igual que en la sección anterior revisaremos solo los tres últimos años aunque en el cuadro de mando es posible observar todos los años desde 2005.

Los autónomos representan un porcentaje bastante menor que el de los asalariados del sector público siendo solo sobre el 11 % de los Ocupados en 2020 y 2021 y parece que en 2022 han bajado. Con respecto a las horas tienen una media de horas de trabajo muy superior a por ejemplo los asalariados del sector público, y están rondando las casi 50 horas.

Año	Porcentaje de trabajadores autónomos	Horas semanales de trabajo (Declaradas por el encuestado)
2020	11,49 %	43,82 h
2021	11,18 %	49,7 h
2022	10,7 %	50,48 h

Cuadro 4.3: Tabla resumen sobre trabajadores Autónomos.
Fuente: Elaboración propia a partir de datos la EPA (INE)

Pasamos a revisar ahora la distribución de los autónomos por sector de actividad en la gráfica 4.5. Más de una cuarta parte de los autónomos trabajan en lo que comúnmente denominamos sector servicios, seguido por la *Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros* y por la *Agricultura, ganadería, silvicultura y pesca*.

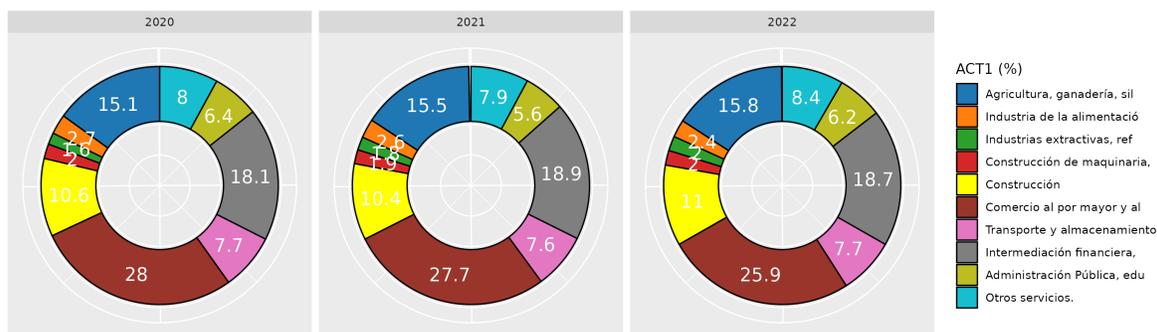


Figura 4.5: Porcentaje de autónomos por sector. España
Fuente: Elaboración propia a partir de datos de la EPA (INE)

4.1.2. Población y tipos de contrato. Galicia

Tras haber realizado un análisis pormenorizado del conjunto del país pasamos a centrarnos en el ámbito geográfico de esta investigación, Galicia. Lo primero es comprobar como está distribuida la población, lo cual puede observarse en la gráfica 4.6. La proporción de inactivos es mayor que la del conjunto del país pero las proporción de parados es menor, lo cual refleja una de las características demográficas más relevantes de la comunidad la cual como indica [Alberto José Pazo y María Pilar Moragón \(2018\)](#) es el excesivo envejecimiento de la población, la cantidad de personas de más de 65 años no deja de subir sin embargo la de personas de menos de 30 no hace más que bajar, esto sobretodo en las provincias de Lugo y Ourense, aumentando de esta forma el número de inactivos.

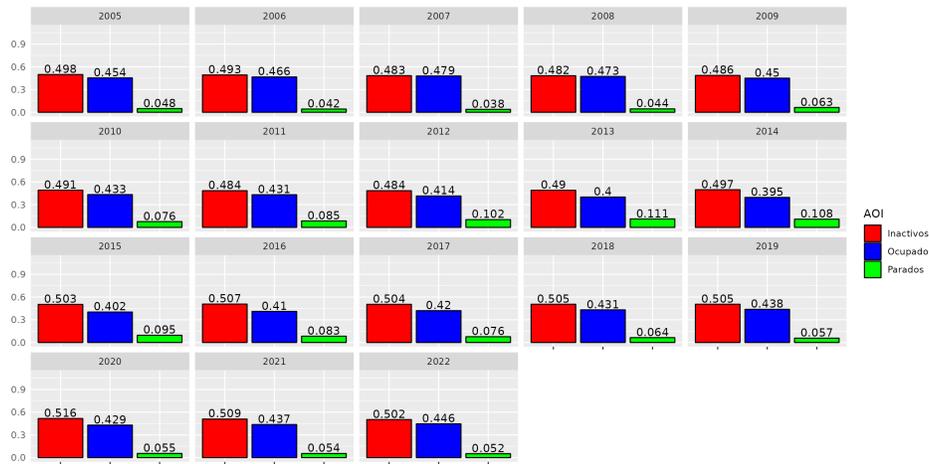


Figura 4.6: Proporción de población según AOI (si son inactivos, Ocupados o Parados). Galicia
Fuente: Elaboración propia a partir de datos de la EPA (INE)

Revisemos ahora la población teniendo en cuenta el sexo. Vemos que en número de mujeres inactivas es superior al de hombres, posiblemente vinculado a su mayor longevidad. Con respecto al número de parados vemos que mantiene proporciones similares la mayoría de los años.

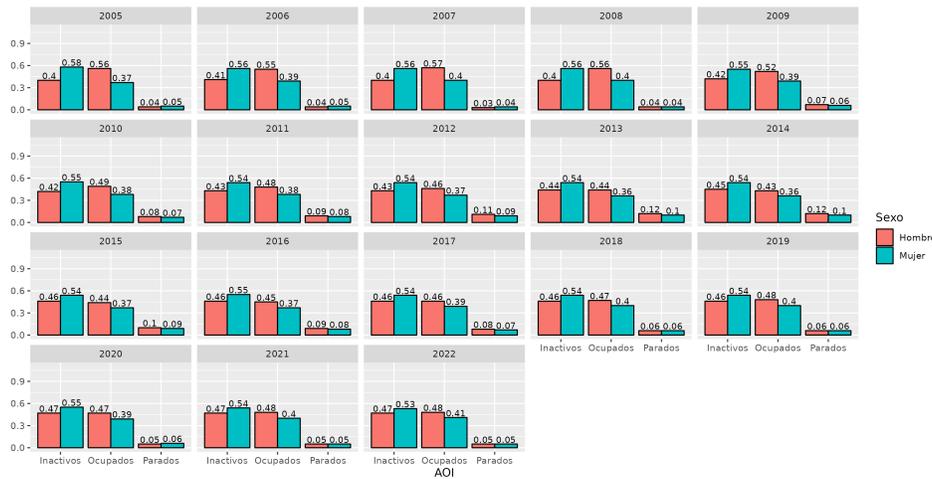


Figura 4.7: Proporción de población según AOI (si son inactivos, Ocupados o Parados) por sexo. Galicia
Fuente: Elaboración propia a partir de datos de la EPA (INE)

Para analizar la proporción de población por tipos de contrato, se va a estudiar la proporción de asalariados con contrato temporal a través de gráficas de líneas. Con el cuadro de mando desarrollado para esta investigación es posible acceder a los datos y gráficas de esta sección de forma anual, especificando trimestre o provincia y segregando por diferentes variables sociodemográficas.

En la gráfica 4.8 podemos ver la proporción de temporalidad la cual ha ido bajando paulatinamente desde 2006 hasta 2013 teniendo un repunte hasta 2019 y una bajada hasta alcanzar el mínimo de la serie en 2022, si comparamos con la gráfica 4.6 vemos claramente que el aumento de la proporción de contratos temporales ocurre al mismo tiempo que la proporción de parados disminuye, esto es debido principalmente a que “La temporalidad se traduce en mayor volatilidad de la creación de empleo. El saldo de los periodos de expansión y contracción es negativo, es decir la volatilidad tiende a provocar más paro” (Torres, 2018)

Por tanto estos datos ya nos están dejando entrever como la temporalidad es uno de los puntos claves del mercado laboral gallego al igual que lo es del mercado laboral español, y su relevancia a la hora de tratar de obtener un mercado laboral más fuerte y estable en el tiempo.

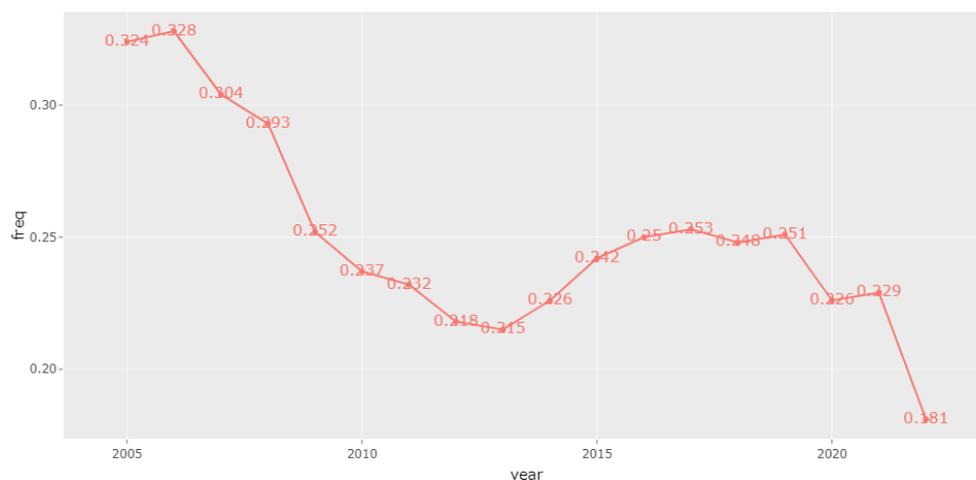


Figura 4.8: Proporción de asalariados con contrato temporal. Galicia

Fuente: Elaboración propia a partir de datos de la EPA (INE)

Al segregar por sexo podemos ver en la gráfica 4.9 como las tasas de temporalidad son mayores en las mujeres que en los hombres. Aunque ambos grupos siguen patrones similares y parece que el aumento o bajada en la temporalidad les ha afectado de forma equivalente. Reforzando más la idea de el influjo de la temporalidad en el paro.

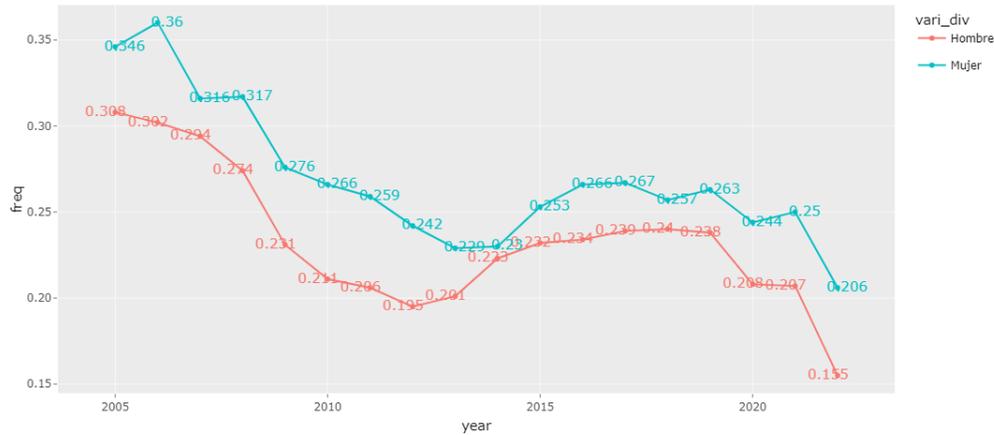


Figura 4.9: Proporción de asalariados con contrato temporal por sexo. Galicia
Fuente: Elaboración propia a partir de datos de la EPA (INE)

Segregando por grupos de edad (gráfica 4.10) podemos ver como la temporalidad esta asociada principalmente a los grupos más jóvenes, reduciéndose al aumentar la edad y alcanzando la proporción más baja con las personas de 60 a 64 años. El grupo de 65 o más tiene datos dispares y poco representativos ya que la muestra de de asalariados en este grupo es mucho menor que en el caso de los otros grupos de edad.

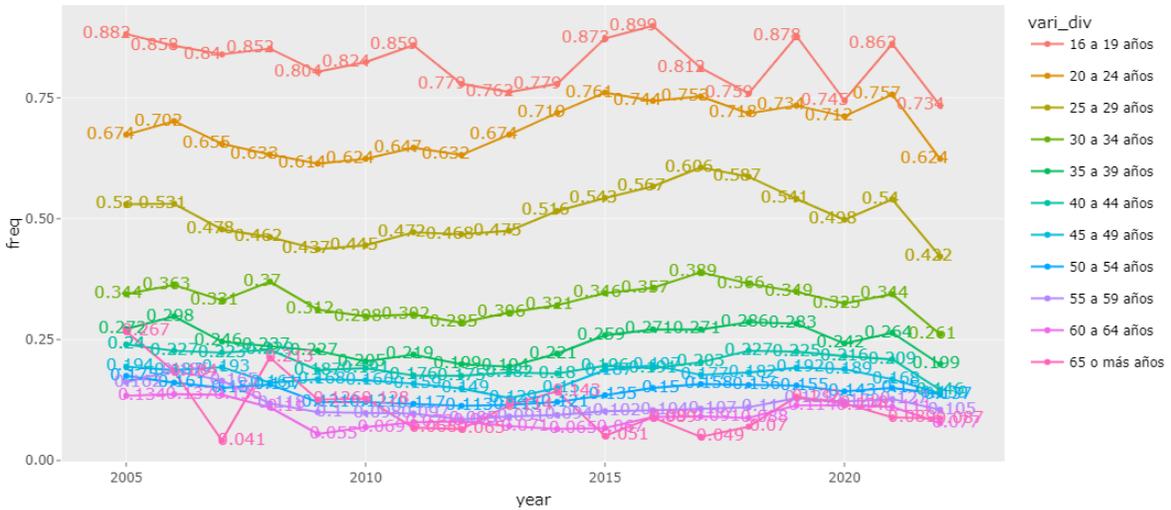


Figura 4.10: Proporción de asalariados con contrato temporal por grupo de edad. Galicia
Fuente: Elaboración propia a partir de datos de la EPA (INE)

Terminamos este apartado revisando la temporalidad por nivel de estudios (se puede consultar el significado de las categorías en la tabla 4.4). En la gráfica 4.11 (en la cual no se visualizan los datos de los asalariados analfabetos por tener muestras excesivamente pequeña) se puede observar como un mayor nivel de estudios está vinculando con menor temporalidad aunque nos encontramos con lo que a parece una anomalía que es que las personas con solo estudios secundarios de orientación general acabados son las que menos sufren de la temporalidad. Puede que esto esté relacionado con los grupos de edad ya que existe una tendencia a que las generaciones mas jóvenes estudien hasta aun mayor nivel al mismo tiempo que son las que mayor temporalidad sufren.

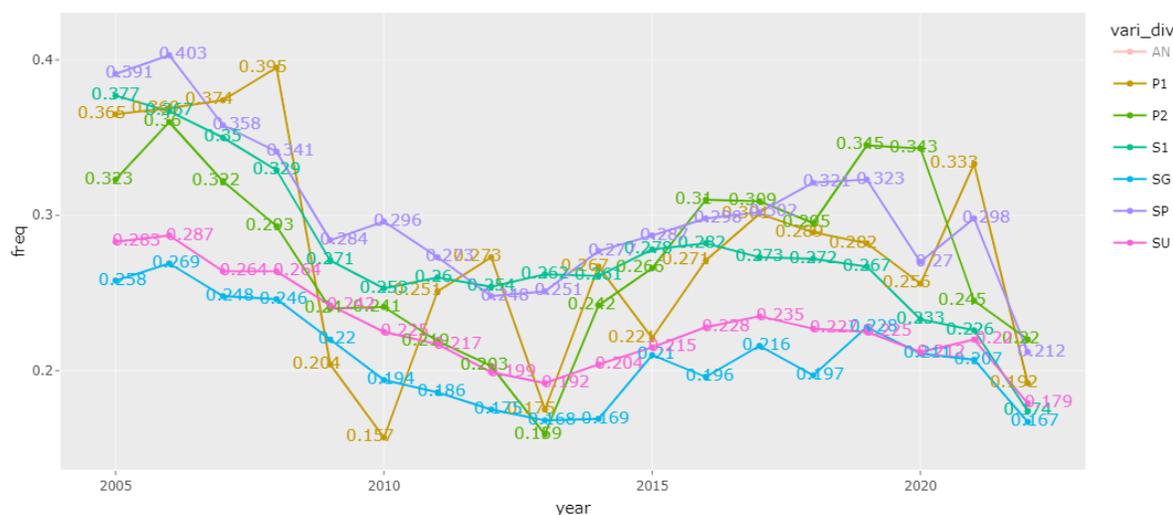


Figura 4.11: Proporción de asalariados con contrato temporal por nivel de estudios. Galicia

Fuente: Elaboración propia a partir de datos de la EPA (INE)

De nuevo es relevante no solo mirar los proporciones en bruto sino el comportamiento a lo largo de la serie, ya que mayor volatilidad nos está indicando mayor desempleo en estos grupos. Es evidente como las personas con solo educación primaria (P1 y P2) son las que más sufren esta problemática.

AN	Analfabetos (código 01 en CNED-2014), (código 80 en CNED-2000)
P1	Educación primaria incompleta (código 02 en CNED-2014), (código 11 en CNED-2000)
P2	Educación primaria (código 10 en CNED-2014), (código 12 en CNED 2000)
S1	Primera etapa de educación secundaria (códigos 21-24 en CNED-2014), (códigos 21-23, 31, 36* en CNED-2000)
SG	Segunda etapa de educación secundaria. Orientación general (código 32 en CNED-2014), (código 32 en CNED-2000)
SP	Segunda etapa de educación secundaria. Orientación profesional (incluye educación postsecundaria no superior) (códigos 33-35, 38**, 41 en CNED-2014), (códigos 33, 34, 41 en CNED-2000)
SU	Educación superior (códigos 51, 52, 61-63, 71-75, 81 en CNED-2014), (códigos 50-56, 59, 61 en CNED-2000)

Cuadro 4.4: Tabla con los valores de variable de nivel de estudios. **Fuente:** EPA (INE)

4.1.3. Sector de ocupación. Galicia

Vamos a centrarnos ahora en estudiar la proporción de Ocupados por sectores, al igual que en el apartado relativo a sectores para el conjunto de España vamos analizar de forma pormenorizada la serie y nos centraremos en los tres últimos años, en este caso con un análisis más profundo en el cual comprobaremos las diferencias por sexo, grupo de edad, nacionalidad y nivel de estudios. Cabe remarcar que con el cuadro de mando desarrollado para este trabajo, es posible revisar todos los gráficos y resultados desarrollados en este apartado pero para diferentes años, trimestres o provincias tanto de Galicia como del resto de España.

Según se puede ver en el apéndice C los dos sectores que se han mantenido en primer y segundo puesto de forma intercalada han sido *Comercio al por mayor y al por menor y sus instalaciones y reparaciones.Reparación de automóviles, hostelería y Administración Pública, educación y actividades sanitarias*. Manteniendo proporciones muy similares y siendo entre los dos aproximadamente el 47% de los Ocupados en Galicia. El tercer puesto ha sido ocupado primero por la *Construcción* y a partir de 2009 por la *Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros*.

Segregando por sexo (gráfica 4.12) vemos diferencias destacables entre hombre y mujeres, los hombres tienen una distribución más uniforme entre los distintos sectores que las mujeres. El sector servicios (*Comercio al por mayor y al por menor y sus instalaciones y reparaciones.Reparación de automóviles, hostelería*) y administración (*Administración Pública, educación y actividades sanitarias*) siguen siendo los principales pero no representan más del 34% del total de Ocupados hombres sin embargo en el caso de las mujeres representa más de 65% del total de Ocupadas mujeres. En sectores como el de *Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros* los porcentajes son similares y la diferencia más destacables estaría en las industrias y la construcción que en el caso de los hombres representaría alrededor de un 33% del total y en de las mujeres sobre un 10%.

En la gráfica 4.13 podemos observar la segregación por edad. Al contrario que con el sexo aquí es más complicado ver diferencias claras entre grupos, aunque sí que parece que entre la gente joven el sector mayoritario es el sector servicios y con los grupos de más edad esto se va reduciendo y el que va aumentando es el sector de administración. También parece que en ciertos sectores que suelen ser considerados de carácter más bien físico como son *Agricultura, ganadería, silvicultura y pesca* y la construcción representan un porcentaje mayor en los grupos de edad más jóvenes al contrario que entre los grupos más mayores.

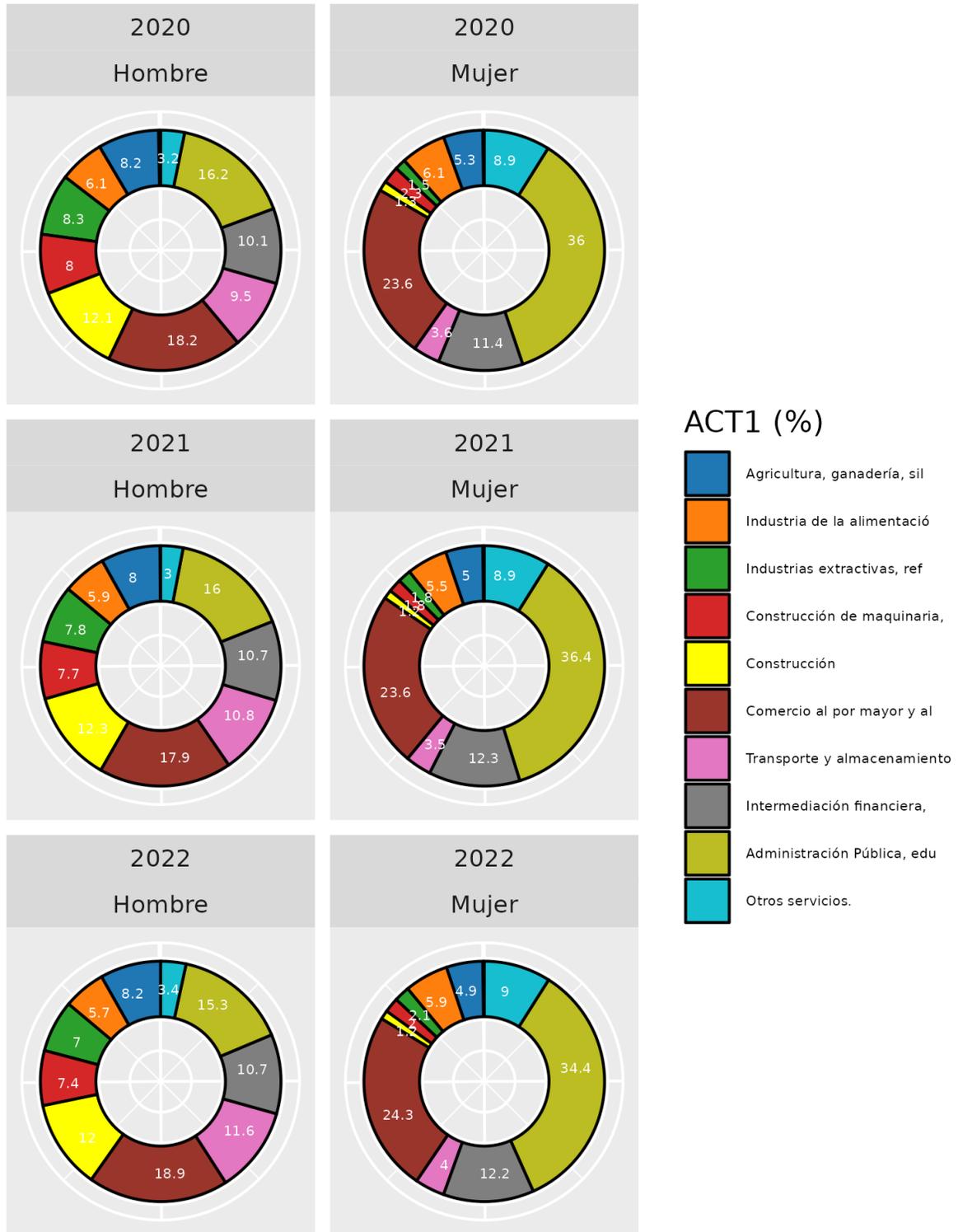


Figura 4.12: Porcentaje de Ocupados por sector de actividad y sexo. Galicia
 Fuente: Elaboración propia a partir de datos de la EPA (INE)

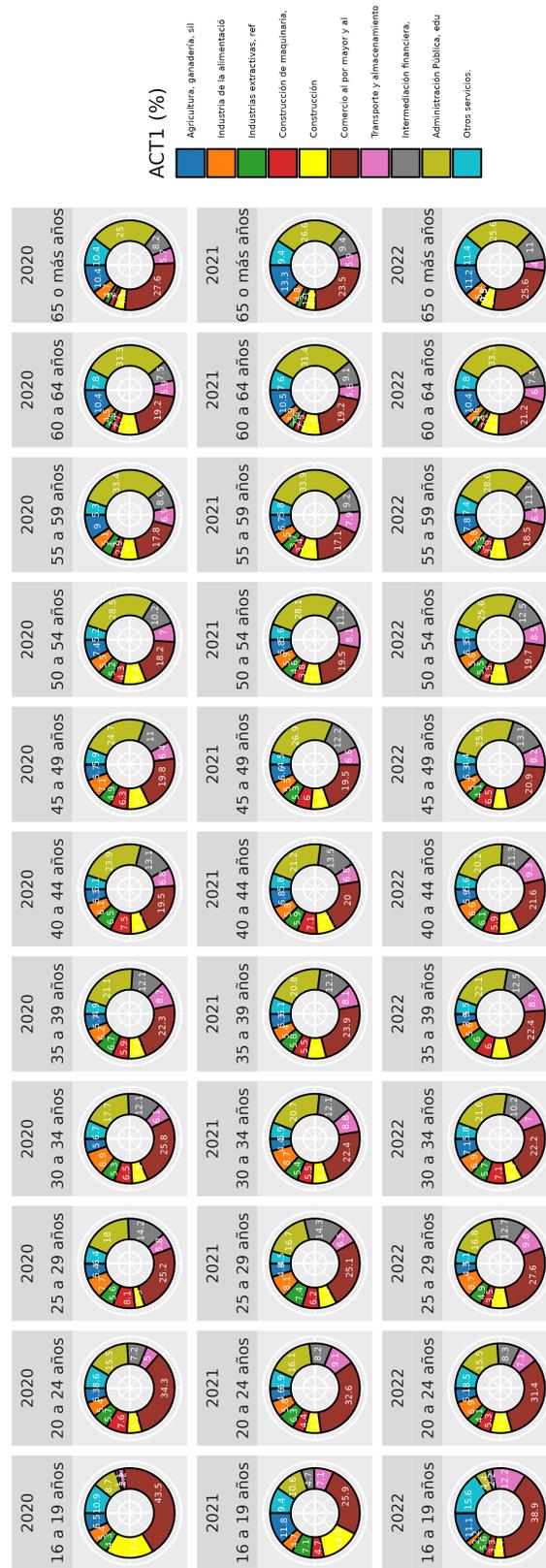


Figura 4.13: Porcentaje de Ocupados por sector de actividad y grupo edad. Galicia
Fuente: Elaboración propia a partir de datos de la EPA (INE)

Pasemos a comentar la gráfica 4.14 en la que se segrega en este caso por la situación de nacionalidad, dividiendo entre aquellos con solo nacionalidad española, los que tienen doble nacionalidad o los que solo tienen nacionalidad extranjera, estos últimos son de alrededor a un 4.41 % de la población Gallega y los dos primeros grupos un 95.59 % según los datos del padrón continuo de 2022 (Instituto Nacional de Estadística, 2023b). Es este caso si que vemos diferencias significativas en tres sectores principalmente, por un lado el sector servicios representa un porcentaje mayor en el caso de aquellos con doble nacionalidad o extranjeros, algo similar ocurre en el caso del sector primario (*Agricultura, ganadería, silvicultura y pesca*) el cual es casi el doble en el caso de las personas con doble nacionalidad y extranjeros. Por último, la otra diferencia significativa esta en el caso de la administración, mientras que para los que solo tienen nacionalidad española representa alrededor del 26 % en el caso de la población extranjera representa solo el 10 % o menos.

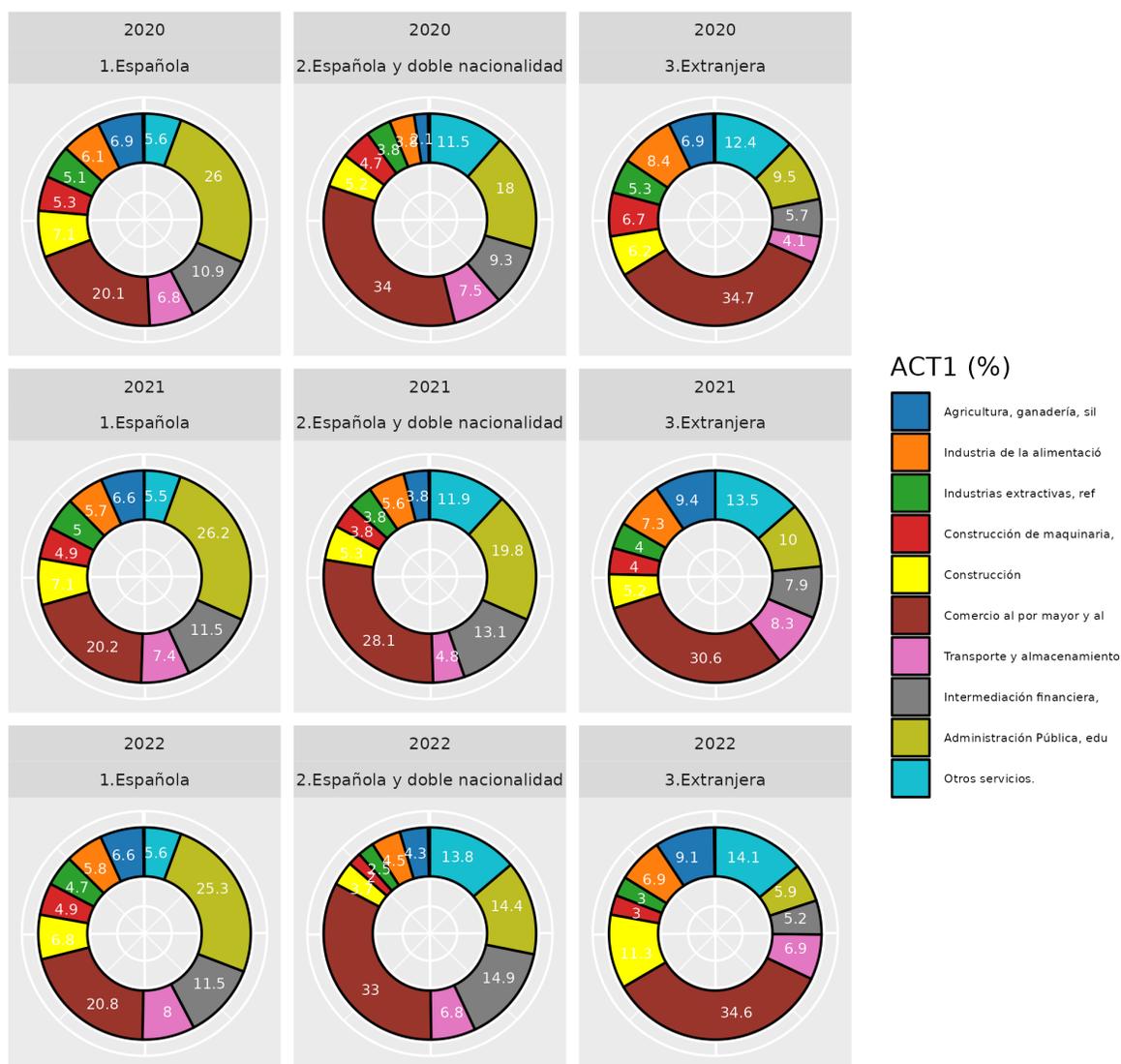


Figura 4.14: Porcentaje de Ocupados por sector de actividad y situación de nacionalidad. Galicia
Fuente: Elaboración propia a partir de datos de la EPA (INE)

En la gráfica 4.15 podemos ver la segregación por nivel de estudios, al igual que en el caso del conjunto del país no se comentará el grupo de analfabetos ya que la muestra de este grupo recogida en la EPA es excesivamente pequeña. La primera diferencia llamativa entre grupos sería la relativa al sector primario ya que a menor nivel de estudios mayor proporción de Ocupados en este sector, pasa lo mismo con la construcción y la *Industria de la alimentación, textil, cuero, madera y papel*. Algo similar pasa pero a la inversa con las administraciones, a mayor nivel de estudios más proporción de trabajadores en el sector, esto es debido a que la mayoría de profesiones vinculadas con este sector requieren de estudios superiores.

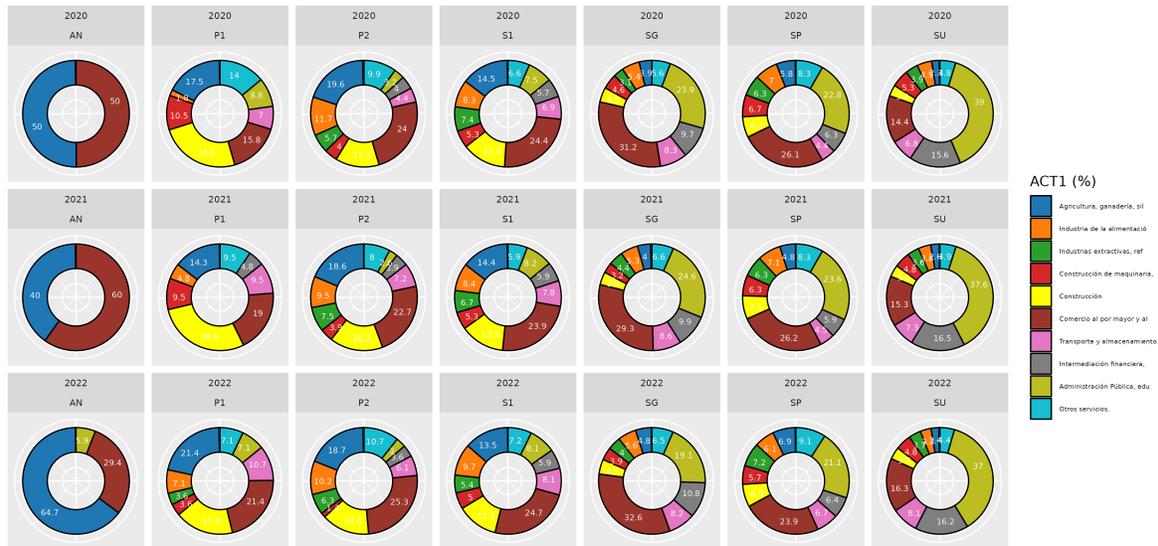


Figura 4.15: Porcentaje de Ocupados por sector de actividad y nivel de estudios. Galicia
 Fuente: Elaboración propia a partir de datos de la EPA (INE)

4.1.4. Administración pública. Galicia

Como hemos visto los trabajadores vinculados con la *Administración Pública, educación y actividades sanitarias* representan casi una cuarta parte de los Ocupados en Galicia pero como ya se menciona anteriormente este sector no está compuesto solo por trabajadores del sector público aquí también estarían trabajadores de empresas privadas con contratos con la administración; escuelas concertadas o privadas y farmacias o clínicas privadas. Por ello se ha incluido esta sección donde solo se analizaran datos de los asalariados del sector público.

Como podemos ver en la tabla 4.5 sobre el total de Ocupados en Galicia los trabajadores del sector público representan algo más del 20%. Al segregar por sexo vemos como las mujeres representan una mayor proporción sobre el total de mujeres que los hombres sobre el total de hombres. Con respecto a los contratos temporales son algo menos del 30% del total de trabajadores públicos estando por debajo de la media española. Y estos contratos son menos comunes entre hombres que entre mujeres. Con respecto a las horas trabajadas de media parecen estar alrededor de las 40 horas semanales independientemente del sexo, aunque los hombres parecen trabajar más horas de media que las mujeres. Los resultados en términos generales son muy similares a los del conjunto de España.

Año	Sexo	Porcentaje de trabajadores del sector público	Porcentaje de trabajadores con contrato temporal	Horas semanales de trabajo (Declaradas por el encuestado)
2020	Ambos	20,2%	25,1%	37,2 h
	Hombres	15,6%	18,9%	37,69 h
	Mujeres	25,1%	29,2%	36,88 h
2021	Ambos	20,3%	26,9%	39,19 h
	Hombres	15,4%	20,9%	39,73 h
	Mujeres	25,3%	30,7%	38,86 h
2022	Ambos	19,2%	27,6%	41 h
	Hombres	14,8%	20%	42,35 h
	Mujeres	23,6%	32,5%	40,14 h

Cuadro 4.5: Tabla resumen sobre trabajadores del sector público. Galicia

Fuente: Elaboración propia a partir de datos la EPA (INE)

En la gráfica 4.16 se puede ver el porcentaje de trabajadores públicos por tipo de administración. Los resultados son similares a los obtenidos para el conjunto de España, siendo predominante la administración de las comunidades autónomas, seguida por las administraciones locales y en tercer lugar la administración central. Resultado coherente teniendo en cuenta que la mayoría de las competencias en sanidad, educación, juventud, etc. están transferidas a las comunidades autónomas.

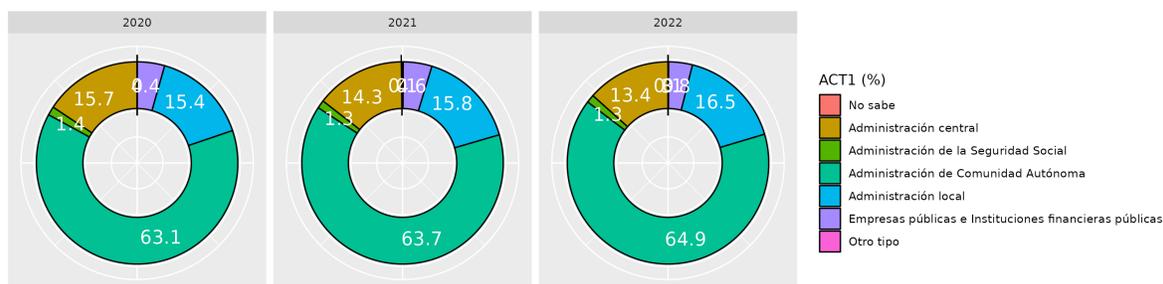


Figura 4.16: Porcentaje de trabajadores de sector público por tipo de administración. Galicia

Fuente: Elaboración propia a partir de datos de la EPA (INE)

4.1.5. Autónomos. Galicia

Los autónomos representan un porcentaje pequeño del total de Ocupados pero superior en Galicia que la media española. Con respecto a los datos segregados por sexo nos encontramos con que entre los hombres hay una proporción mayor que entre las mujeres, este de casi un punto porcentual. El otro dato relevante es el número de horas semanales de trabajo, el cual es bastante superior a la media gallega que se aproxima a las 40 horas semanales que a las 50 horas que los autónomos dedican de media.

Año	Sexo	Porcentaje de trabajadores autónomos	Horas semanales de trabajo (Declaradas por el encuestado)
2020	Ambos	13,13 %	46,66 h
	Hombres	14,76 %	47,38 h
	Mujeres	11,41 %	45,72 h
2021	Ambos	12,7 %	50,77 h
	Hombres	14,31 %	51,51 h
	Mujeres	11,04 %	49,78 h
2022	Ambos	12,3 %	52,87 h
	Hombres	14 %	53,96 h
	Mujeres	10,57 %	51,41 h

Cuadro 4.6: Tabla resumen sobre trabajadores autónomos. Galicia

Fuente: Elaboración propia a partir de datos la EPA (INE)

Con respecto al sector servicios nos encontramos con resultados similares a los del conjunto de España, representa sobre una cuarta parte de los trabajadores autónomos aunque no es en todos los años el sector principal. Mientras que sectores como el de la construcción siguen representando porcentajes similares, la principal diferencia se encuentra en el sector primario, el cual en el caso del resto de España representaba un porcentaje de sobre 15 % mientras que en Galicia representa sobre un 25 % de los autónomos.

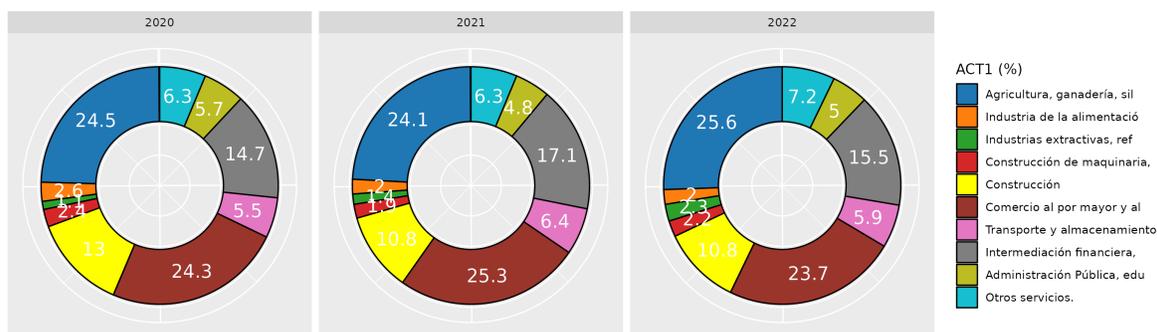


Figura 4.17: Porcentaje de trabajadores autónomos por sector de ocupación. Galicia

Fuente: Elaboración propia a partir de datos de la EPA (INE)

4.2. Análisis de los datos de afiliados a la Seguridad Social

Pasamos ahora a analizar los datos de afiliados a la seguridad social que además serán los datos que ajustaremos en el capítulo 5.

4.2.1. Contexto Español

Al igual que con los datos de la EPA vamos a hacer un breve análisis de la situación en el conjunto del territorio para poder hacer comparaciones con Galicia. En España había a fecha de Marzo de 2023 en torno a 20.4 millones de afiliados a la seguridad social, en torno a 10.8 son hombres y sobre 9.5 son mujeres, siendo esta la máxima de toda la serie estudiada.

En la gráfica 4.18 podemos ver la serie temporal desde 2012 hasta el primer trimestre de 2023 segregado por sexo. A primera vista es evidente que número de afiliados hombres es mayor que el de afiliados mujer a lo largo de toda la serie.

Además aunque la tendencia parece tener patrones similares de crecimiento, siendo algo más suave en el caso de las mujeres, es en el patrón estacional donde se observan las diferencias más destacables. En el caso de los hombres nos encontramos con que en la mayoría de los años hay un aumento constante de número de afiliados desde Enero hasta Agosto, y a partir de ahí una caída y un estancamiento hasta diciembre. Pero en caso de las mujeres parece que tras la caída en Agosto hay una fuerte recuperación en septiembre que se mantiene estable hasta hasta finales de año.

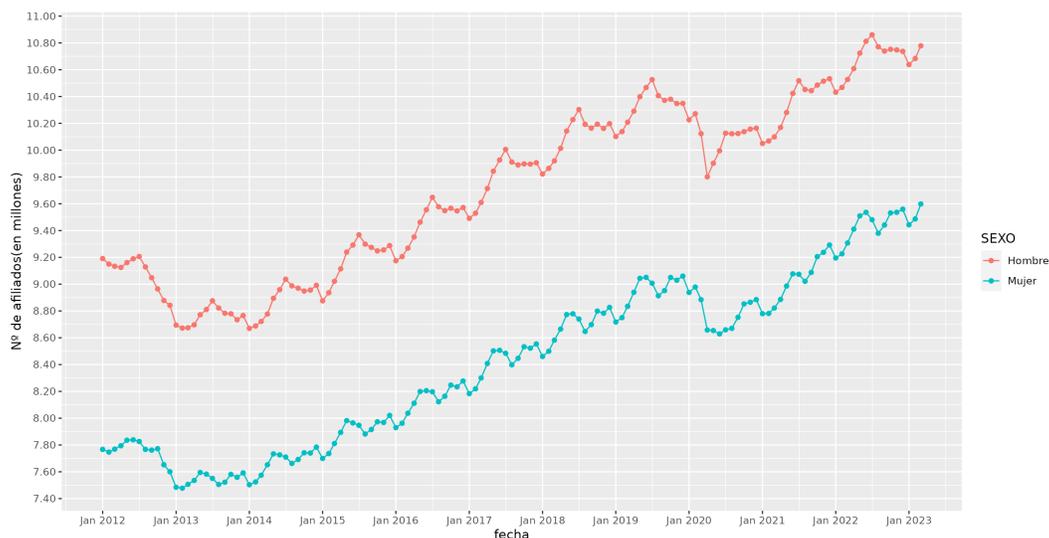


Figura 4.18: Número de afiliados medios por sexo. España

Fuente: Elaboración propia a partir de datos de Afiliados a la seguridad social (Seguridad social)

Pasamos ahora a revisar las series segmentadas por grupos quinquenales de edad en la gráfica 4.19. Lo más destacable en mi opinión es la fuerte estacionalidad de los grupos de edad más jóvenes en comparación con los más adultos. Se pueden observar también los efectos de la pandemia del COVID-19, habiendo una brusca bajada del número de afiliados en Marzo de 2020, con una recuperación desigual dependiendo del grupo de edad, los mayores de 45 parecen ser los que más rápidamente se han recuperado del evento. También se observa un aumento de los afiliados de 65 o más años posiblemente vinculado al paulatino aumento de la edad de jubilación que será de 67 años en 2027.

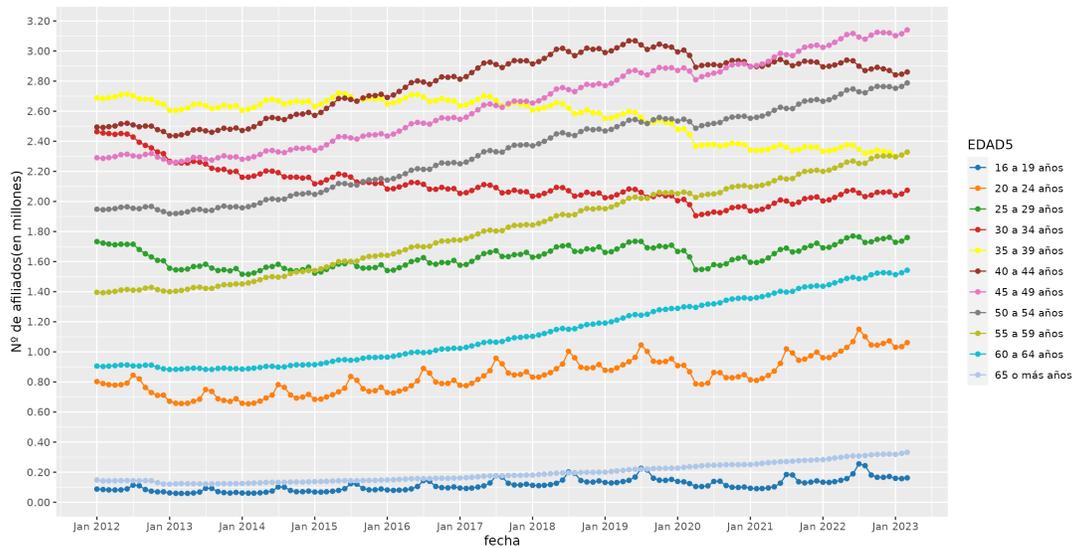


Figura 4.19: Número de afiliados medios por grupos de edad. España

Fuente: Elaboración propia a partir de datos de Afiliados a la seguridad social (Seguridad social)

En gráfico 4.20 se puede observar la serie de número de afiliados por sector de ocupación CNAE-09, aunque tanto los datos de la EPA como los datos de afiliados a la seguridad social usan el sistema CNAE-09 las agrupaciones por sector de los distintos trabajos es diferente lo que hace que no sean directamente equiparables. El sector con mayor número de afiliados es con una diferencia destacable el del *Comercio, transporte y hostelería*; seguido por la *Administración pública, educación y sanidad*, que también es el sector de mayor crecimiento en la última década; y en tercera posición estaría la *Industria* hasta 2014 y *Actividades profesionales, científicas y técnicas* posteriormente.

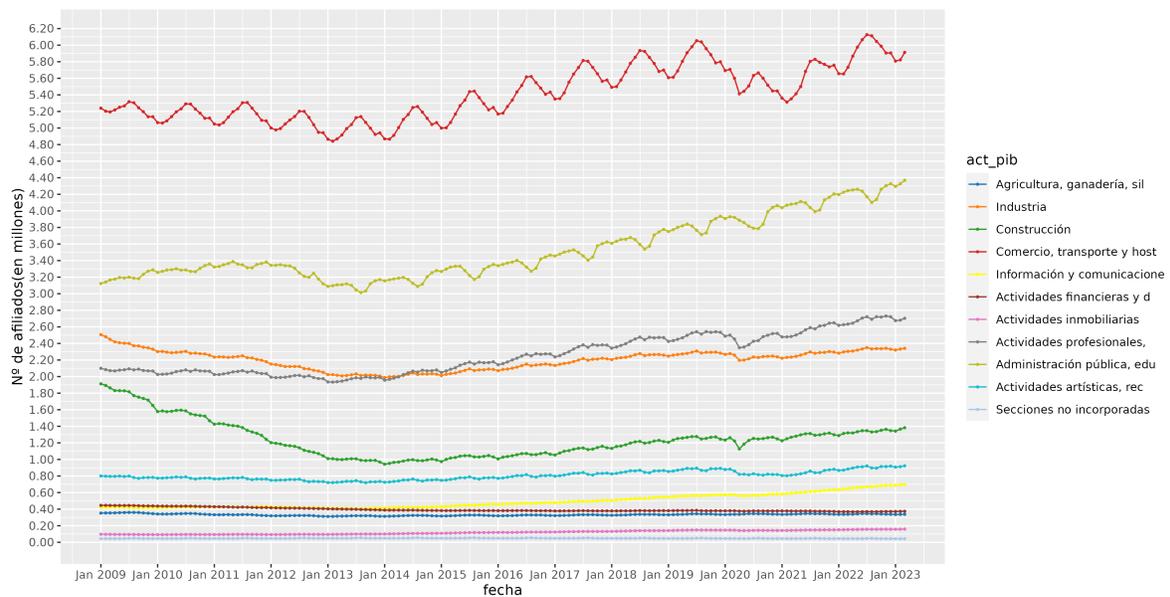


Figura 4.20: Número de afiliados medios por sector de ocupación. España

Fuente: Elaboración propia a partir de datos de Afiliados a la seguridad social (Seguridad social)

4.2.2. Sexo y grupos de edad. Galicia

En la gráfica 4.21 podemos observar las serie segregada por sexo, de nuevo nos encontramos con diferencias en el número de afiliados y en el comportamiento de la serie. En términos generales ambos tienen un comportamiento creciente hasta Agosto, a partir de este mes en el caso de los hombres el decrecimiento es constante hasta inicios del año siguiente pero en el caso de las mujeres tras Agosto viene un estancamiento hasta final de año. En ambos casos el efecto de la crisis del COVID-19 es evidente y el comportamiento en general parece bastante similar al visto para el conjunto del país.

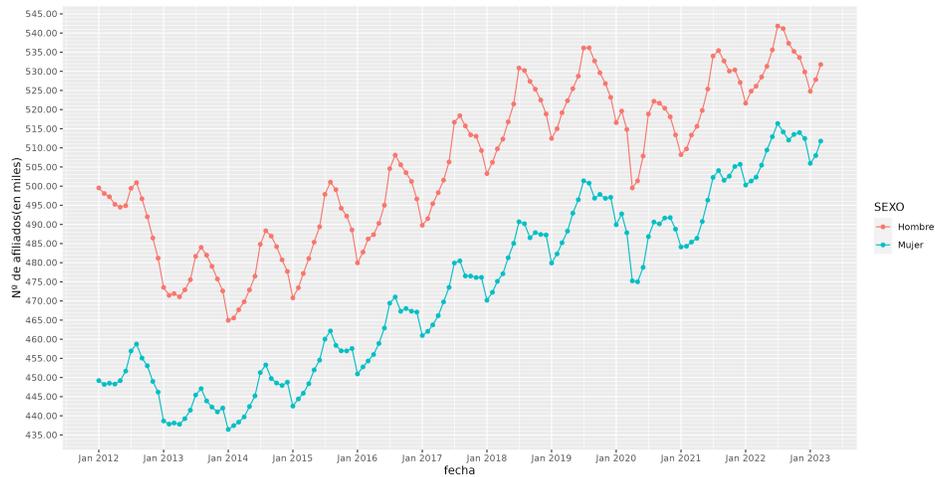


Figura 4.21: Número de afiliados medios por sexo. Galicia

Fuente: Elaboración propia a partir de datos de Afiliados a la seguridad social (Seguridad social)

En la gráfica 4.22 podemos observar el número de afiliados por grupo quinquenal de edad y al igual que lo que pasa con el resto de España en Galicia los grupos de edad más jóvenes sufren de mayor estacionalidad con respecto de los grupos más mayores. En prácticamente todos los grupos es evidente el efecto del COVID-19, y de nuevo los grupos de más edad parecen ser los más afectados.

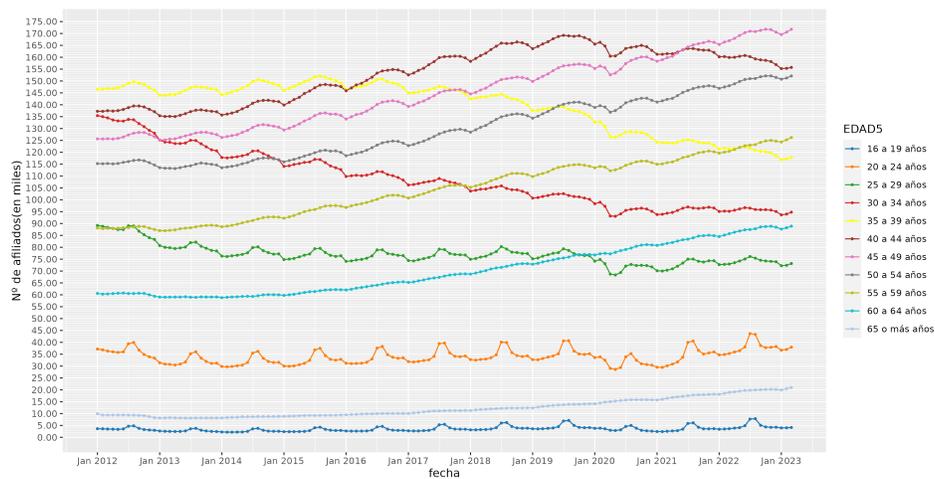


Figura 4.22: Número de afiliados medios por grupos de edad. Galicia

Fuente: Elaboración propia a partir de datos de Afiliados a la seguridad social (Seguridad social)

4.2.3. Sectores de ocupación. Galicia

Para terminar vamos a revisar los afiliados por sectores de ocupación en la gráfica 4.23. El sector del *Comercio, transporte y hostelería* es el que más número de afiliados tiene con bastante diferencia, también es el que parece tener una mayor estacionalidad y el efecto del COVID-19 fue bastante fuerte sobre el , aun no consiguiendo retomar los números prepandemia. El segundo sector con mayor número de afiliados sería la *Administración pública, educación y sanidad*, el cual ha tenido un crecimiento pronunciado durante la última década, la crisis del COVID-19 no parece que le afectara en demasía. El tercer lugar de forma clara y consistente esta la *Industria* al contrario que en el conjunto de España donde la industria ocupa la tercera posición hasta 2012 y a partir de ahí la cuarta, y es el sector de *Actividades profesionales, científicas y técnicas* el que mantiene una tercera posición estable y bastante constante desde 2014.

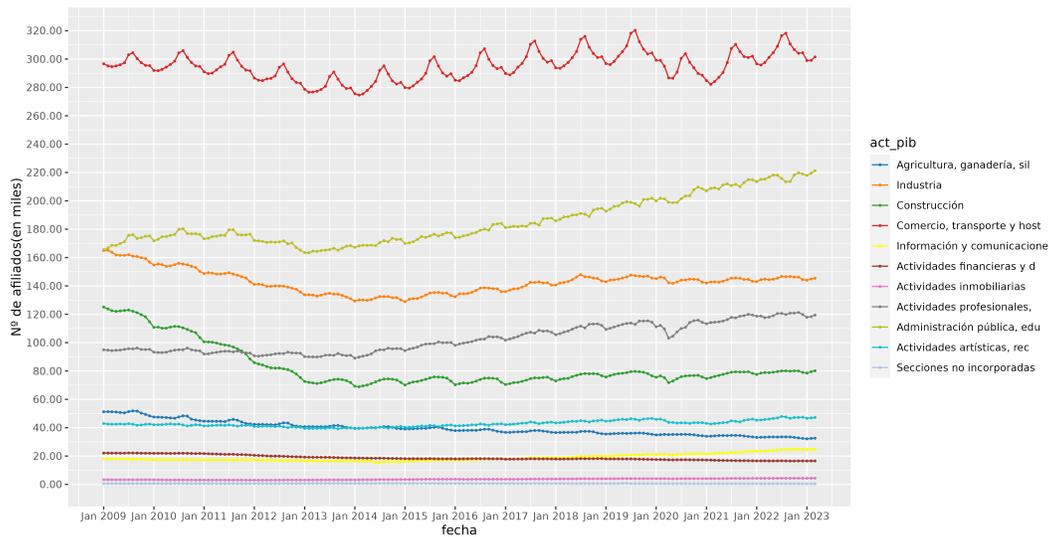


Figura 4.23: Número de afiliados medios por sector de ocupación. Galicia

Fuente: Elaboración propia a partir de datos de Afiliados a la seguridad social (Seguridad social)

Capítulo 5

Modelos predictivos

Este capítulo estará centrado en los modelos predictivos de la evolución del número afiliados a la Seguridad Social, ya que este es uno de los principales indicadores usados a la hora de analizar el estado del mercado laboral. Una de sus principales características que hemos observado en el análisis datos en el capítulo 4 es que la serie temporal esta sujeta a una fuerte estacionalidad, siendo los meses de verano de cada año los que mayor numero de afiliados tienen. Por ello en esta sección no solo trataremos de buscar un modelo que ajuste bien los datos de afiliados de Galicia, sino que trataremos también de ver cuales son los sectores de actividad en Galicia con mayor estacionalidad y crear modelos específicos para ellos.

Esto se hace con la intención no solo de tener un modelo predictor si no también para entender en mayor medida como es el comportamiento de estos sectores y así poder tomar medidas correctoras para mejorar la empleabilidad.

Todo el código en R relativo a este apartado es accesible a través de Kaggle, dentro de la notebook desarrollada hay enlaces a la búsqueda individual de los mejores modelos de cada tipo:

<https://www.kaggle.com/code/erikmartinezsixto/resultados-modelos-predictivos-tfm-erik-martinez>

5.1. Modelos y técnicas empleadas

En esta sección se explicarán los modelos y técnicas usadas tanto para discernir cuales son los sectores que mayor estacionalidad tienen como para crear modelos que ajusten los datos.

5.1.1. Fuerza de la estacionalidad

Una de las caracterizaras de la serie temporal que pueden ser analizadas es su estacionalidad (seasonal), para ello es necesario hacer una descomposición, la cual como indica su nombre descompone la serie temporal en sus componentes de tendencia-ciclo, estacionalidad y error. (Rob J. Hyndman y Yeasmin Khandakar, 2021)

Para ello usaremos la “Seasonal and Trend decomposition using Loess” conocida también por las siglas STL, la cual es una técnica que fue desarrollado por Cleveland et al. (1990) con la intención de crear un método de descomposición que fuese simple, versátil y robusto. Con la intención de obtener,

$$y_t = T_t + S_t + R_t \quad (5.1)$$

siendo y_t es la serie temporal original, el T_t es la tendencia suavizada, el S_t el componente estacional y R_t el error. La técnica se base en realizar un proceso iterativo de suavizado Loess para estimar la tendencia de la serie temporal, para después eliminar esta tendencia de la serie original y revelar el componente estacional y los residuos. Para obtener la estacionalidad se realiza el cálculo de la media móvil de esta serie sin tendencia. Y por ultimo, conociendo ahora tendencia y la estacionalidad se despeja el componente de error de la siguiente forma, $R_t = T_t + S_t + y_t$. Este proceso se repite hasta que la descomposición cumple el criterio de convergencia predefinido.

Una vez descompuesta la serie temporal nos interesa conocer la fuerza del componente estacional, para ello [Rob J. Hyndman y Yeasmin Khandakar \(2021\)](#) propone medir esta fuerza como,

$$F_s = \max\left(0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(T_t + T_t)}\right) \quad (5.2)$$

obteniendo así que las series temporales con una estacionalidad fuerte tendrían valores cercanos a uno y los que no tienen componente estacional cercanos a cero. De esta forma sería sencillo seleccionar aquellos sectores con una estacionalidad más fuerte.

5.1.2. SARIMA

Según [Rob J. Hyndman y Yeasmin Khandakar \(2021\)](#) los modelos ARIMA nacen como una combinación de modelos autoregresivos y modelos de media móvil. Los modelos autoregresivos se pronostica la variable y_t haciendo uso de una combinación de lineal de valores pasados de y_t . Estos modelos pueden ser descritos como,

$$AR(p) : y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (5.3)$$

siendo c es una constante; ϕ_1, ϕ_2, ϕ_p los coeficientes de autoregresion; $y_{t-1}, y_{t-2}, y_{t-p}$ son los valores anteriores de la serie temporal en los periodos $t-p$; por ultimo ε_t es el ruido blanco o error. Por tanto tendríamos un modelo $AR(p)$ donde p serie el paramento que definiría el orden del modelo, indicando la cantidad de valores pasados que se tendrán en cuenta a la hora de predecir el valor y_t .

Los modelos de media móvil en vez de usar los valores pasados para pronosticar usan los errores de pronósticos en modelos lineales hechos previamente. Estos modelos pueden ser descritos,

$$MA(q) : y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (5.4)$$

siendo c es una constante; $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-q}$ son los errores pasados; $\theta_1, \theta_2, \theta_p$ los coeficientes de media móvil. Tendríamos así un modelo $MA(q)$ donde q es el parametro del orden del modelo, indicando la cantidad de errores pasados ha considerar.

Al combinar ambos modelos con un proceso de diferenciación, solo en caso de que la serie original sea non-stationary (no estacionaria), obtendríamos un modelo AutoRegressive Integrated Moving Average más conocido por sus siglas ARIMA. El modelo puede ser descrito como,

$$ARIMA(p, d, q) : y'_t = c + \phi_1 y'_{t-1} + \dots + \phi'_p y'_{t-p} \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (5.5)$$

siendo y'_t la serie tras un proceso de diferenciación. Con esto tendríamos un modelo $ARIMA(p, d, q)$. Estos modelos son capaces de modelizar también series temporales con datos con fuerte estacionalidad, para eso se incluye términos adicionales para la parte estacional. Pudiendo describirse el modelo como,

$$ARIMA(p, d, q)(P, D, Q)_m \quad (5.6)$$

siendo los parámetros en minúsculos los relativos a la parte no estacional del modelo y en mayúscula los paramentos de la parte estacional. La selección de estos parámetros de forma manual es un proceso

complejo pero la función *ARIMA()* de la librería *fable* desarrollada por O'Hara-Wild et al. (2023) tiene programada una variante del algoritmo de Hyndman-Khandakar (Rob J. Hyndman y Yeasmin Khandakar, 2008) que permite seleccionar los parámetros para obtener el modelo con menor AIC (Akaike's Information Criterion).

5.1.3. SARIMAX

Los modelos SARIMA como hemos visto nos permiten modelar la serie temporal en base a los datos pasados de la propia serie, pero en muchas ocasiones querríamos añadir al modelo variables externas que permitan obtener predicciones más afinadas. Según Rob J. Hyndman y Yeasmin Khandakar (2021) para esto existen los modelos ARIMA con variables exógenas, también conocidos como ARIMAX. En caso de tener un componente estacional SARIMAX. Todas las variables deben ser estacionarias, es caso contrario aplicar un proceso de diferenciación.

En este caso los errores y coeficientes a estimar son más complejos que en el caso del SARIMA, para llevar acabo este proceso la librería *fable* en su función *ARIMA()* tiene incorporado un algoritmo para el calculo del mejor modelo, seleccionando automáticamente tanto los parámetros de la parte no estacional del modelo (p, d, q) como los parámetros de la parte estacional (P, D, Q) . A mayores para todos los modelos es calculado el AICc (Corrected Akaike's Information Criterion) el cual permite escoger entre todos los modelos con distintas variables predictoras cual es mejor.

5.1.4. Redes neuronales

Según Rob J. Hyndman y Yeasmin Khandakar (2021) los modelos de redes neuronales son métodos predictivos basados en modelos matemáticos simples que tratan de imitar el funcionamiento de las neuronas. Es una técnica muy interesante a la hora de ajustar modelos con relaciones no lineales entre las variable respuesta y las variables predictoras. La redes neuronales más simples son las que no tienen capas ocultas y que serían equivalentes a una regresión lineal pero de forma menos eficiente, por tanto las redes neuronales más interesantes serían aquellas con una o más capas ocultas.

En este trabajo usaremos una red neuronal feed-forward con una sola capa oculta, que es el modelo que esta programado en la función *NNETAR()* de la librería *fable*. Estos modelos generan variables ocultas que son combinaciones lineales de las variables predictoras, que son modificadas por una función no lineal en este caso la función sigmoideal (función logística), finalmente el modelo final es una combinación lineal de estas variables ocultas. Uno de los puntos a tener en cuenta es que los pesos iniciales dados a cada variable predictora son escogidos de manera aleatoria, haciendo el modelo aleatorio. Para solucionar esto lo que se hace es entrenar al modelo múltiples veces usando distintos pesos iniciales y usando finalmente la media de estos modelos.

5.2. Selección de sectores

A la hora de ajustar los modelos es interesante no solo ajustar modelos del conjunto de todos los trabajadores sino también de sectores de interés. Hacerlo de cada uno de los sectores sería complejo y requeriría una gran potencia computacional por ello se ha tomado la decisión de hacerlo tanto del conjunto de todos los afiliado y de los tres sectores con mayor estacionalidad. Ya que hemos vista en análisis previos que estacionalidad y temporalidad están relacionadas, y que a mayor temporalidad peores condiciones de trabajo.

Para la selección se usara la fuerza de le estacionalidad explicada con mayor profundidad en el apartado anterior, la cual esta incorporada en la librería *feast* desarrollada por O'Hara-Wild et al. (2022) a través de la función *feat_stl*. Los resultados se pueden ver en la tabla 5.1.

El sector del *Comercio, transporte y hostelería* es el que tiene un patrón estacional con mayor fuerza, algo esperable teniendo en cuenta el análisis de datos realizado en el capítulo 4. No solo estamos hablando del sector con mayor fuerza de estacionalidad si no también el sector con mayor número de afiliados de Galicia. En segunda posición aparece el conjunto de todos los afiliados de Galicia, demostrando que el mercado laboral gallego es estacional en su conjunto aunque haya sectores con más o menos estacionalidad. Posteriormente estarían la *Industria* y la *Agricultura, ganadería, silvicultura y pesca*. Los dos primeros sectores están entre los sectores con mayor número de afiliados de Galicia, solo quedando *Administración pública, educación y sanidad* fuera de esta lista, ya que como se observa este sector es el cuarto con menor fuerza de estacionalidad.

Por tanto y para concluir esta sección los sectores que serán ajustados con los distintos modelos serán *Comercio, transporte y hostelería; Industria y Agricultura, ganadería, silvicultura y pesca*.

act_pib	trend_strength	seasonal_strength_year
Comercio, transporte y hostelería	0.9913100	0.9865087
TODAS LAS SECCIONES	0.9976363	0.9776253
Industria	0.9980817	0.9459560
Agricultura, ganadería, silvicultura y pesca	0.9976000	0.9422803
Construcción	0.9987448	0.9256823
Actividades profesionales, científicas y técnicas y otras	0.9976178	0.9196211
Actividades artísticas, recreativas y otros servicios	0.9948351	0.8924095
Secciones no incorporadas	0.9706117	0.8759778
Administración pública, educación y sanidad	0.9857099	0.7143966
Actividades financieras y de seguros	0.9987895	0.5614139
Actividades inmobiliarias	0.9963309	0.5190800
Información y comunicaciones	0.9821633	0.2901792

Cuadro 5.1: Tabla con los datos de fuerza de la tendencia y de la estacionalidad por sectores. Galicia
Fuente: Elaboración propia a partir de datos de Afiliados a la seguridad social(Seguridad social)

5.3. Resultados

Inicialmente se comenzó trabajando con 4 conjuntos de datos uno para la serie temporal de todos los afiliados en Galicia (conjunto) y tres para los distintos sectores seleccionados en la sección anterior (servicios, industria y primario). Estos datos se dividieron en dos conjuntos uno de entrenamiento desde 2009 hasta Diciembre de 2019 y otro de test desde Enero de 2020 hasta Diciembre de 2022. Pero tras observar los resultados se decidió añadir en algunos casos nuevos datasets pero partiendo los datos en Diciembre de 2020, incluyendo por tanto en el entrenamiento de los modelos el 2020, que fue el año con las principales restricciones COVID.

Con cada uno de estos dataset se han probado los distintos modelos que han sido comentados en las secciones anteriores:

- Para los modelos SARIMA se ha usado la función $ARIMA()$ de la librería *fable* desarrollada por O'Hara-Wild et al. (2023), la cual ya ha seleccionado parámetros óptimos.
- Para los modelos SARIMAX se ha usado también la función $ARIMA()$, la selección de los parámetros ha sido automática y se han ajustado todos los modelos posibles con las distintas combinaciones de las variables *ipc*, *pib* y *santa* en el dataset conjunto; *ipc*, *pib*, *santa* y *covid* en el dataset covid; *ipc*, *pib*, *santa* y *oferta* en los datasets servicios, industria y primario; *ipc*, *pib*, *santa*, *oferta* y *covid*.
- Con los modelos de redes neuronales se ha usado la función $NNETAR()$ de la librería *fable*. En este caso usando modelos que incluyen todas las variables que se mencionan a continuación; *ipc*, *pib* y *santa* en el dataset conjunto; *ipc*, *pib*, *santa* y *covid* en el dataset covid; *ipc*, *pib*, *santa* y *oferta* en los datasets servicios, industria y primario; *ipc*, *pib*, *santa*, *oferta* y *covid*. Se realizaron 1000 simulaciones por cada uno de estos modelos.

5.3.1. Modelos de afiliación para el conjunto de sectores

El modelo que mejor ajusta con una diferencia destacable en este dataset es el de redes neuronales. No solo tiene el menor error sino también tiene el mejor ajuste gráfico, consiguiendo de alguna forma ajustar la bajada del número de afiliados vinculada a la crisis del COVID-19. Posteriormente a este evento sobrestima el número de afiliados.

Modelo	Formula	RMSE	MAE	MPE	MAPE
neuro	afi_med~pib+ipc+santa	16836.792715608604	12798.425686359653	-1.0472074237358595	1.3150631939835584
sarima	afi_med ~pdq(1,1,1) + PDQ(0,1,1)	29422.27725468198	24698.893714085523	-2.5061967712991993	2.5557837535070846
sarimax	afi_med ~ipc + pdq(1,1,1) + PDQ(0,1,2)	33255.171256903006	28195.78256434254	-2.867448387674006	2.907179907804108

Cuadro 5.2: Errores de los mejores modelos de cada tipo. Dataset conjunto

Fuente: Elaboración propia

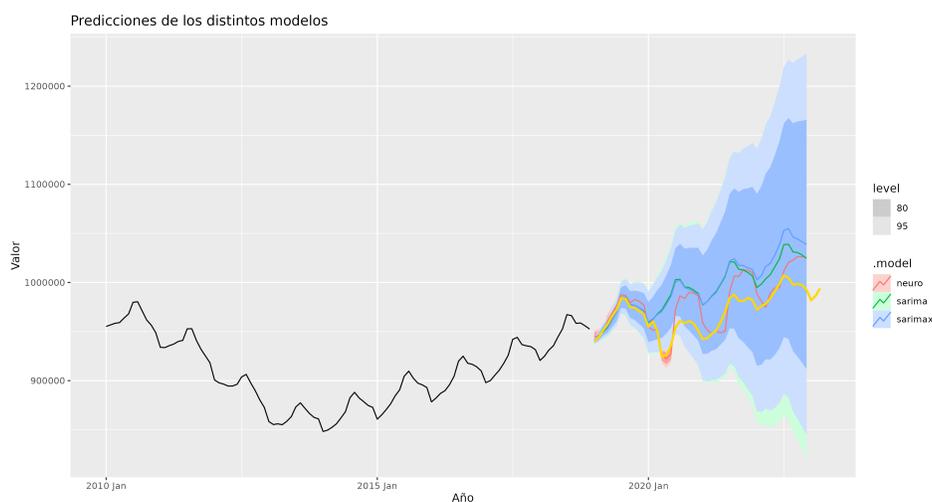


Figura 5.1: Predicciones del número de afiliados. Dataset conjunto. Galicia

Fuente: Elaboración propia

5.3.2. Modelos de afiliación para el conjunto de sectores. Covid

En este dataset los modelos son entrenados con el año covid por tanto se espera de ellos mejores predicciones que las obtenidas en el apartado anterior, sin embargo cabe resaltar que las medidas de error entre ambos datasets no pueden ser directamente comparadas en el caso del RMSE y el MAE, y el MPE y MAPE pueden ser medidas engañosas en algunos casos por penalizar de forma distinta los errores negativos y los positivos. (Rob J. Hyndman y Yeasmin Khandakar, 2021)

En este caso el modelo que mejor ajusta es el SARIMAX con las variables exogenas de PIB y COVID. En la gráfica 5.2 la concordancia entre la predicción y los datos reales es bastante evidente aunque no ajusta perfectamente los datos se encuentran dentro del intervalo de confianza. Y el modelo consigue representar muy bien el efecto de la estacionalidad.

Modelos	Formula	RMSE	MAE	MPE	MAPE
sarimax	$\text{afi_med} \sim \text{pib} + \text{covid} + \text{pdq}(1,1,0) + \text{PDQ}(0,1,1)$	6953.583067511953	5637.71122538747	0.17317043073253766	0.5776516042757226
neuro	$\text{afi_med} \sim \text{pib} + \text{ipc} + \text{santa} + \text{covid}$	16110.838709360687	13500.572523849589	1.0466132719023564	1.3689774972578783
sarima	$\text{afi_med} \sim \text{pdq}(1,1,1) + \text{PDQ}(0,1,1)$	43921.19073462394	40272.26402916414	4.0834894492282	4.0834894492282

Cuadro 5.3: Errores de los mejores modelos de cada tipo. Dataset covid

Fuente: Elaboración propia

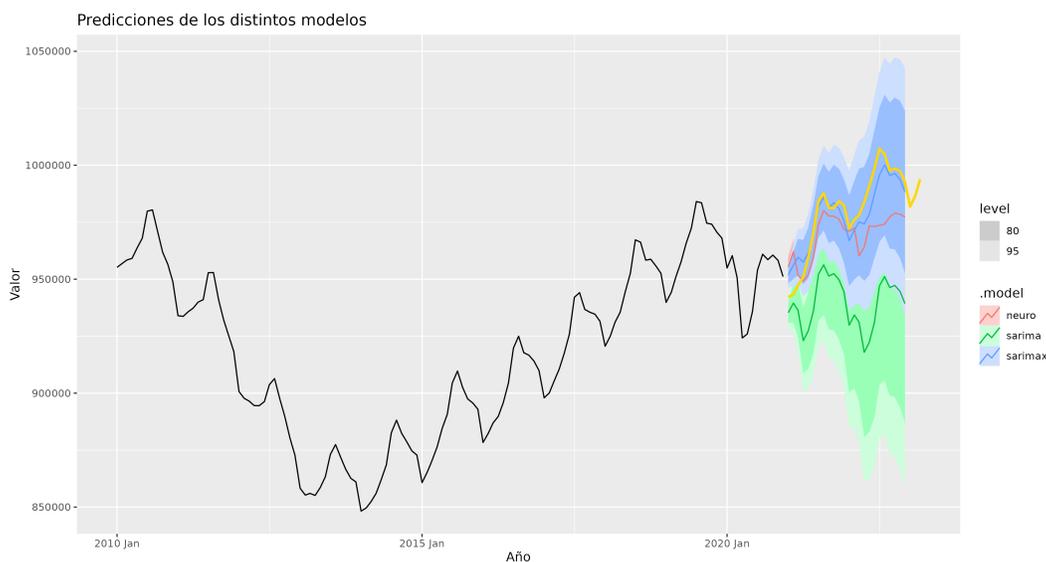


Figura 5.2: Predicciones del número de afiliados. Dataset covid. Galicia

Fuente: Elaboración propia

5.3.3. Modelos de afiliación para el sector servicios

De nuevo nos encontramos que el modelo que mejor ajusta entre los modelos que no están entrenados con datos del COVID sería la red neuronal. Consiguiendo un ajuste muy acertado. Todos los modelos predicen de forma adecuada el patrón anual que en el caso de este sector ha sido bastante continuado en el tiempo, aunque la tendencia ha ido cambiando es evidente que es un sector con fuerte estacionalidad, lo cual ayuda a su predicción por mantener estos patrones de forma estable.

Con respecto a los datos covid nos encontramos con que el en este caso es también la red neuronal la que consigue un menor error y mejor ajuste.

Modelo	Formulas	RMSE	MAE	MPE	MAPE
neuro	afi_med ~ pib + ipc + santa + oferta	3391.750010909394	2734.016473828668	-0.055029147548003164	0.9115656712830571
sarimax	afi_med ~ pib + santa + oferta + pdq(1,1,0) + PDQ(1,1,0)	10215.954950488816	8123.8692602481615	-2.511950291619311	2.7453278803601866
sarima	afi_med ~ pdq(1,1,1) + PDQ(1,1,0)	10431.941336457561	8238.79048454374	-2.577276017257476	2.7846380334987657

Cuadro 5.4: Errores de los mejores modelos de cada tipo. Dataset servicios

Fuente: Elaboración propia

Modelo	Formula	RMSE	MAE	MPE	MAPE
neuro	afi_med ~ pib + ipc + santa + oferta	3246.4028936916548	2728.7150815191176	0.48193809669675347	0.9026945332852346
sarimax	afi_med ~ covid + oferta + pib + santa + pdq(0,1,1) + PDQ(1,1,0)	13057.43806638949	11249.153158603302	3.36950258997617	3.69224775148942

Cuadro 5.5: Errores de los mejores modelos de cada tipo. Dataset servicios covid

Fuente: Elaboración propia

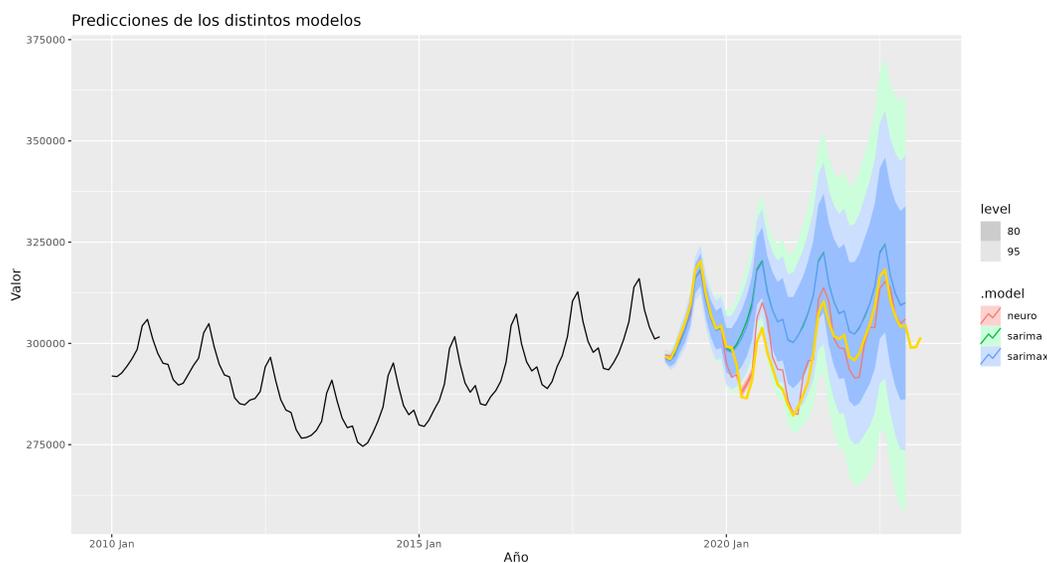


Figura 5.3: Predicciones del número de afiliados. Dataset servicios. Galicia

Fuente: Elaboración propia

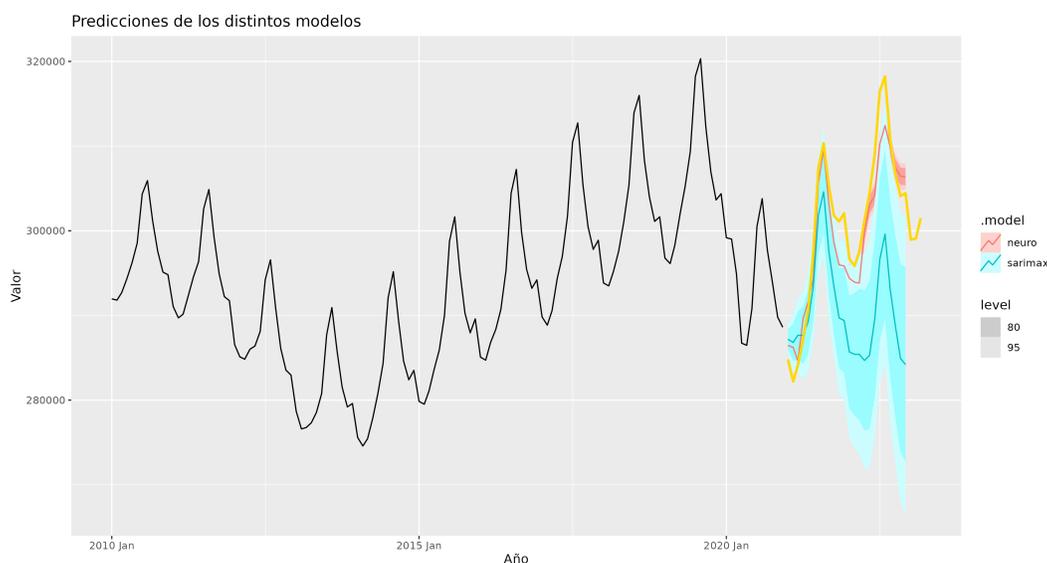


Figura 5.4: Predicciones del número de afiliados. Dataset servicios covid. Galicia
Fuente: Elaboración propia

5.3.4. Modelos de afiliación para el sector industria

En este caso el mejor modelo para los datos sin COVID es el SARIMA, aunque los tres modelos se comportan bastante mejor de media que en el resto de datasets. Ninguno consigue ajustar el año covid pero tras 2020 tanto el SARIMA como el SARIMAX consiguen un buen ajuste gráfico.

De nuevo si se entrenan los datos teniendo en cuenta el covid la predicción esta mucho mejor ajustada, siendo en este caso el SARIMAX con las variables covid y pib de nuevo el mejor modelo.

Modelo	Formula	RMSE	MAE	MPE	MAPE
sarima	$afi_med \sim pdq(1,1,2) + PDQ(1,1,1)$	1684.3254569114963	1315.9470501470198	0.15522569547440296	0.9099364195245411
sarimax	$afi_med \sim pib + pdq(1,1,2) + PDQ(1,1,1)$	1943.634922908178	1436.304452167485	-0.6417497420000302	0.9970705653921301
neuro	$afi_med \sim pib + ipc + santa + oferta$	3121.8532460737893	2441.935648718933	1.357971796922356	1.6851716869235676

Cuadro 5.6: Errores de los mejores modelos de cada tipo. Dataset industria
Fuente: Elaboración propia

Modelo	Formulas	RMSE	MAE	MPE	MAPE
sarimax	$afi_med \sim covid + pib + pdq(1,1,1) + PDQ(0,1,1)$	759.4323150963319	632.1023214970155	-0.3951167212385479	0.438913136736849
neuro	$afi_med \sim pib + ipc + santa + oferta$	2230.7305578409932	1885.6759937023417	1.0828446517610966	1.2973951076735482

Cuadro 5.7: Errores de los mejores modelos de cada tipo. Dataset industria covid
Fuente: Elaboración propia

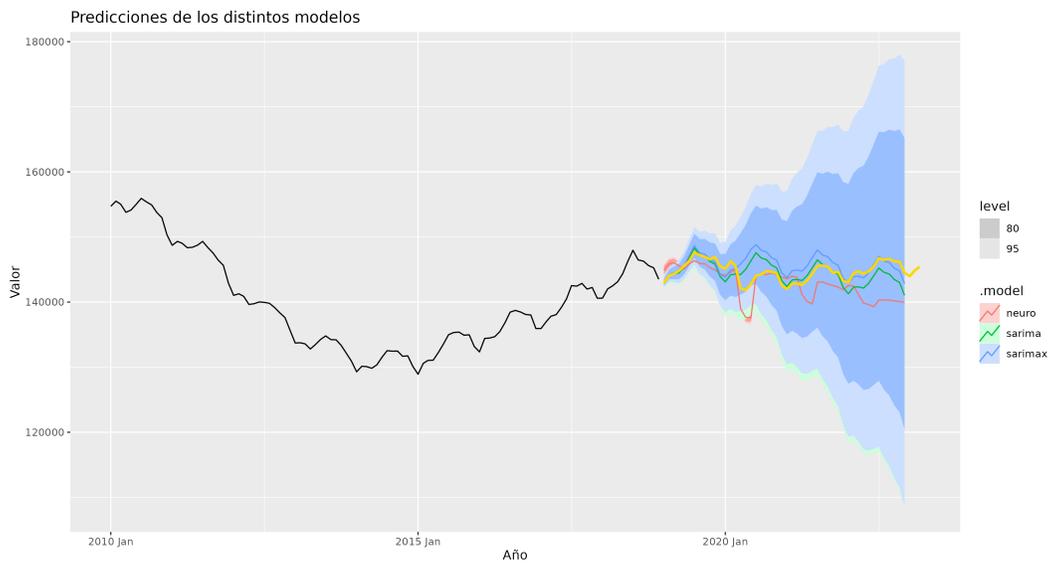


Figura 5.5: Predicciones del número de afiliados. Dataset industria. Galicia

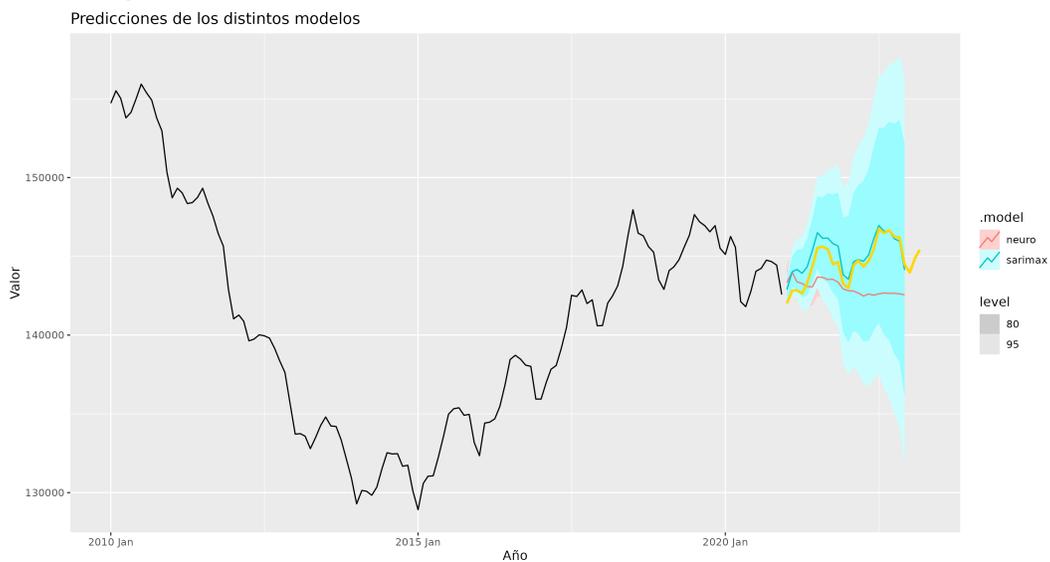


Figura 5.6: Predicciones del número de afiliados. Dataset industria covid. Galicia

Fuente: Elaboración propia

5.3.5. Modelos de afiliación para el sector primario

En este caso el mejor modelo es el SARIMA aunque el SARIMAX también consigue buenos resultados, ambos modelos predicen muy bien la tendencia aunque tenían dificultades para ajustar el comportamiento estacional.

En los datos covid el SARIMAX consigue de nuevo el mejor resultado ajustando muy bien tanto la tendencia como la estacionalidad.

Modelo	Formula	RMSE	MAE	MPE	MAPE
sarima	$\text{afi_med} \sim \text{pdq}(0,1,0) + \text{PDQ}(1,1,0)$	326.609948874297	250.3272508557728	-0.3896441032203489	0.725430790088648
sarimax	$\text{afi_med} \sim \text{oferta} + \text{pdq}(1,1,2) + \text{PDQ}(0,1,2)$	404.8996937298158	312.87236263723935	-0.6383821034107131	0.9133029778794395
neuro	$\text{afi_med} \sim \text{pib} + \text{ipc} + \text{santa} + \text{oferta}$	2279.800105504486	1901.3673390164247	-5.59121251630766	5.59121251630766

Cuadro 5.8: Errores de los mejores modelos de cada tipo. Dataset primario

Fuente: Elaboración propia

Modelo	Formula	RMSE	MAE	MPE	MAPE
sarimax	$\text{afi_med} \sim \text{pib} + \text{pdq}(0,1,0) + \text{PDQ}(1,1,0)$	133.28438802832736	112.85145581416509	-0.18939916983854976	0.3365703371086534
neuro	$\text{afi_med} \sim \text{pib} + \text{ipc} + \text{santa} + \text{oferta}$	430.74796255027167	350.6292773633307	0.46680532032534944	1.0375877492003762

Cuadro 5.9: Errores de los mejores modelos de cada tipo. Dataset primario covid

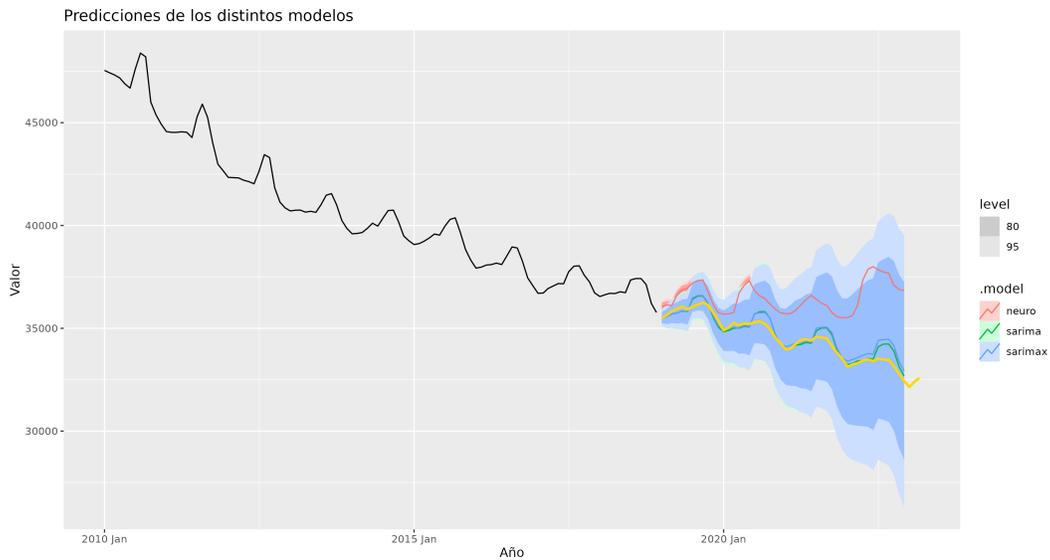
Fuente: Elaboración propia

Figura 5.7: Predicciones del número de afiliados. Dataset primario. Galicia

Fuente: Elaboración propia

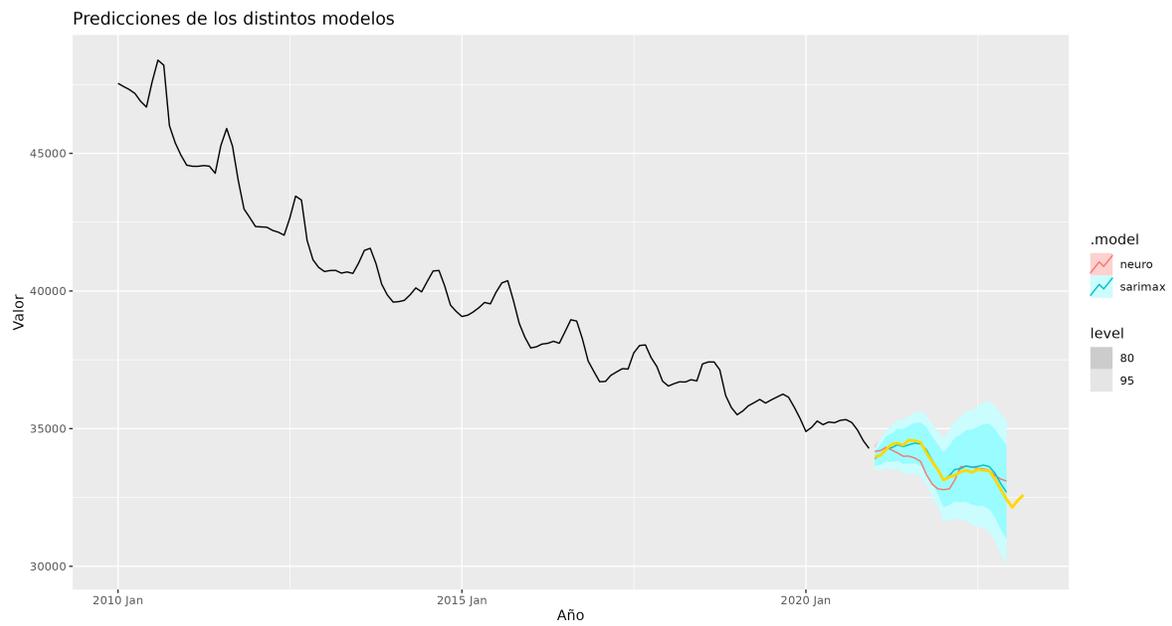


Figura 5.8: Predicciones del número de afiliados. Dataset primario covid. Galicia
Fuente: Elaboración propia

Capítulo 6

Conclusiones

En este trabajo se establecieron cuatro objetivos principales con el propósito de dar una imagen general del mercado laboral gallego al mismo tiempo que se creaban una serie de herramientas útiles para futuras investigaciones en el ámbito del mercado laboral español y sus comunidades autónomas. El primer objetivo se centra en crear datasets con información de calidad sobre el mercado laboral; el segundo en hacer un análisis de los mismos, obteniendo información general pero útil del mercado laboral gallego; el tercero se centra en el desarrollo de un cuadro de mando para la explotación de dichos datos; y por último el ajuste de modelos predictivos sobre el comportamiento del mercado laboral gallego comparando cuáles son las técnicas que mejor funcionan.

Comenzamos con el primero objetivo, como se ha descrito en el capítulo 2, los datasets creados son principalmente de dos fuentes, por un lado la **Encuesta de Población Activa** del INE y por el otro los datos del **numero de afiliados a la seguridad social** de la Tesorería General de la Seguridad Social, siendo estos datos de acceso libre y de altísima calidad. La EPA es una encuesta solvente con una muestra amplia y representativa (alrededor de 200.000 personas al trimestre) de la sociedad española y con una larga continuidad temporal en la mayoría de sus estimadores. El número de afiliados también es sumamente interesante como representación del número de trabajadores tanto segregados por características sociodemográficas como por sectores de actividad, dando información constante de como va fluctuando el mercado de trabajo.

Estos datos se han puesto en abierto a través de Kaggle, la plataforma de ciencia de datos más extendida con más de 12 millones de usuarios y 50 mil datasets públicos, siendo la primera vez que se publica un dataset de la EPA y afiliados de Seguridad Social en esta plataforma, pudiendo usarse gracias a las notebooks de la propia plataforma o directamente descargarse en un formato más simple y accesible de los que tenían en disposición el INE o la Tesorería General de la Seguridad Social. Dando acceso tanto los datos originales simplemente combinados en un único archivo o usar alguna de las versiones modificadas que dan mayor versatilidad y comodidad en su uso.

Sin embargo también ha habido ciertos desafíos y limitaciones en su creación, la más relevante sería la imposibilidad de unificar los datos de ambas fuentes en base a los sectores de ocupación, lo cual era el plan inicial. El número de afiliados da una visión general del comportamiento del mercado laboral mientras la EPA permite conocer información más concreta sobre las características y condiciones de los trabajadores que lo componen. Por desgracia la agrupación sectorial de la EPA, aun estando basada en la Clasificación de Actividades Económicas 2009 (**CNAE-2009**), es distinta a la de los datos de Afiliados a la Seguridad Social, y solamente es posible acceder a los datos del sector de actividad de la EPA sin agrupar asumiendo un coste de extracción y preparación de los datos.

El segundo objetivo era hacer un análisis de las bases de datos lo cual permitiera tener una imagen general del mercado laboral gallego, haciendo también comparaciones con el resto del territorio español. Los resultados del análisis de forma extensa se desarrollan en el capítulo 4.

Los análisis llevados a cabo, con el apoyo de la herramienta de cuadros de mando desarrollada, han permitido seleccionar aquellos casos de uso de más interés en el análisis de la realidad del mercado laboral gallego. No obstante, con la ayuda de la herramienta y los datasets es posible llevar a cabo un número muy superior de análisis dada la envergadura de campos disponibles.

Vamos a comentar los principales resultados con respecto a la EPA, cuyo análisis se ha centrado en los siguientes cuatro aspectos:

- **Categorización de población y de contrato.** Vemos que Galicia tiene una proporción de inactivos mayor que el resto del país pero menor proporción de parados, efectos probablemente vinculados al envejecimiento de la población. También mencionar que hay más mujeres inactivas que hombres. Otro de los puntos relevantes analizados en este apartado es la temporalidad, la proporción de ocupados con contrato temporal lleva en tendencia decreciente los últimos años, y al haber una fuerte relación entre el aumento de proporción de parados y la disminución de la proporción de personas con contratos temporales hemos llegado a la conclusión de que las personas con contrato temporal son más susceptibles de perder el empleo, posiblemente por su menor protección legal y coste de despido. También vemos algunas tendencias a lo que temporalidad se refiere, por ejemplo que las mujeres la sufren más de media que los hombres, que los más jóvenes son los que más tienen este tipo de contratos, y que con la edad va reduciéndose paulatinamente. El nivel de estudios también es un factor relevante y vemos como a mayor nivel de estudios suele haber en términos generales menor proporción de temporalidad.
- **Distribución de trabajadores por sector.** Los sectores de *Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería* conjuntamente con *Administración Pública, educación y actividades sanitarias* representarían en la mayoría del periodo analizado sobre 47 % del total de Ocupados. En los últimos 9 años el sector mayoritario ha sido la *Administración Pública, educación y actividades sanitarias*, y entre las mujeres Ocupadas este es con diferencia el principal sector de actividad. Hay diferencias reseñables por cuestión de nacionalidad, como que los extranjeros y aquellos con doble nacionalidad están menos vinculados con sectores como el de la administración y más con el sectores como el de Servicios. El nivel de estudios también es un factor relevante en la distribución por sectores, entre las personas con estudios superiores el sector de la *Administración Pública, educación y actividades sanitarias* representa cerca de un 40 % sin embargo en caso de las personas con solo la primera etapa de la educación primaria los sectores mayoritarios serían los Servicios y la Construcción.
- **El sector público.** Un 20 % del total de Ocupados son asalariado del sector público, si segregamos por sexo, sobre un 15 % del total de Ocupados hombre y sobre un 25 % del total de Ocupadas mujeres. En Galicia los contratos temporales entre asalariados del sector público es menor que el del resto de España pero aun así esta rondando el 30 %. La distribución por tipos de administración es de alrededor de un 64 % para las administraciones de las Comunidad Autónoma, un 16 % para la administración local y un 14 % de la administración central (el porcentaje restante se reparte entre la seguridad social, empresas publicas y otros)
- **Los autónomos.** Los autónomos en Galicia representa alrededor de un 13 % del total de Ocupados un porcentaje mayor que la media española. También trabajan de media más horas semanales que el resto de Ocupados, llegando casi a las 50. Con respecto a la distribución por sectores vemos diferencias significativas con respecto al resto de Ocupados, lo más destacable sería que sobre un 25 % de los autónomos trabajan en algo relacionado con el sector primario y que solo sobre 5 % trabajan en algo vinculado con la *Administración Pública, educación y actividades sanitarias*.

En cuanto al análisis de los datos de afiliación a la Seguridad Social las conclusiones más relevantes son:

- **Análisis por sexo y grupos de edad.** Una constante en el análisis de las series temporales hechas en este trabajo es el fuerte impacto de la crisis sanitaria del COVID-19 a comienzos de 2020, modificando el comportamiento de la tendencia pero no de la estacionalidad que tras 2020 sigue marcando los mismos patrones de crecimiento de número de afiliados hasta Agosto de cada año y su posterior decrecimiento hasta inicios del año siguiente. Sin embargo encontramos ligeras diferencias entre hombres y mujeres en este aspecto, por ejemplo que las mujeres tras una ligera bajada tras agosto se mantiene el número de afiliadas bastante estable hasta finales de año, el caso de los hombres esta bajada es más pronunciada. En el análisis segregado por grupo de edad también vemos patrones interesantes como que las series de los grupos de edad más mayores sufren de una estacionalidad mucho más suave en comparación con los grupos más jóvenes que la tienen más acentuada.
- **Análisis por sectores de ocupación,** el sector con mayor número de afiliados en este es el de *Comercio, transporte y hostelería* seguido por *Administración pública, educación y sanidad* y *Industria*. Vemos diferencias claras en el comportamiento por sector, por ejemplo el sector servicios sufre de una estacionalidad más fuerte que el resto de los sectores y el efecto del COVID-19 fue más destacable que por ejemplo en el caso de la Administración.

Por tanto el análisis de datos ha revelado importantes patrones y características del mercado laboral gallego. Hay una mayor proporción de inactivos, lo cual puede estar relacionado con el envejecimiento de la población en la región. Además, se evidencia una tendencia decreciente en la proporción de contratos temporales, lo cual indica una disminución en la precariedad laboral y en la vulnerabilidad de los trabajadores ante la pérdida de empleo, ya los contratos temporales están vinculados con menor protección legal y peores condiciones laborales. A su vez, es destacable que esta disminución en los contratos temporales no ha resultado en un aumento proporcional de desempleados como en años anteriores. Sin embargo se sigue observando que el mercado laboral gallego está regido por una estacionalidad, estando además sus principales sectores entre aquellos que más sufren de esta problemática, por tanto todavía existe un amplio margen de mejora en el mismo, tratando de disminuir la estacionalidad y temporalidad en pos de trabajos estables y de mayor calidad.

El tercer objetivo era la creación de una herramienta de cuadro de mando que permitiese a usuarios no experimentados acceder de forma fácil y cómoda a análisis y gráficas de los datos. Su desarrollo está explicado en el capítulo 3.

La aplicación está construida en el lenguaje de programación R a través de la librería *shinydashboard*, usando librerías como *tidyverse*, *plotly* y *seasonal* para el análisis de los datos y la creación de las gráficas. La aplicación ha sido probada en un portátil con una Intel(R) Core(TM) i7-7500U de 2 cores con 12 GB de RAM.

A lo largo de su desarrollo hubo que afrontar ciertos restos siendo el más destacable los problemas con la velocidad de carga y el rendimiento. La aplicación usa los datos en bruto y calcula todas y cada uno de los indicadores necesarios para el posterior análisis. Esto junto a la gran cantidad de datos de la EPA ocasionaban un rendimiento deficiente. Para solucionarlo se limitó los tipos de análisis a solo la población de Ocupados y se eliminaron las variables de menor interés, reduciendo así el número de datos y aumentando fluidez de la aplicación.

Aun con estas limitaciones la aplicación permite de forma muy sencilla acceder de forma ágil y cómoda a un montón de gráficos y tablas con datos que nos permiten caracterizar como está distribuido el trabajo, tipo de contrato, horas, entre otras. Aunque esta haya sido una versión inicial es posible

actualizarla con los nuevos datos de la EPA y de afiliados a la seguridad social que vayan saliendo en años siguientes o incluso incluir nuevos gráficos y análisis de forma simple.

El cuarto y último objetivo era ajustar modelos predictivos sobre el futuro del mercado laboral, para este apartado, que está desarrollado en mayor profundidad en el capítulo 5, se han probado varios tipos de modelos predictivos para estimar la evolución del número de afiliados a la Seguridad Social de cara a analizar que modelo tiene el mejor comportamiento y cuáles son los más útiles de cara a investigaciones posteriores.

Dependiendo del dataset usado obtenemos resultados dispares, un punto importante a mencionar es que los modelos que no han sido entrenados con datasets en los que está incluido el año 2020, año donde ocurrió la crisis del COVID-19, obtienen de media mejores resultados con errores más bajos y con ajustes gráficos claramente superiores. Durante el desarrollo de este apartado se han encontrado problemáticas concretas difíciles de solucionar como la excesiva capacidad de computación requerida por la mayoría de los modelos. Por ello se ha optado por usar solo tres tipos de modelos pero tratar de sacarles el mayor provecho a cada uno.

En los modelos entrenados con datos COVID (año 2020 en el conjunto de entrenamiento) nos encontramos que el modelo con mejor rendimiento es una SARIMAX con las variables *covid* y *pib*, esto nos está indicando que de no haber otro suceso catastrófico y difícilmente predecible las estimaciones del PIB conjuntamente con los datos históricos de afiliados a la Seguridad Social sería suficiente para obtener buenas predicciones en este respecto.

Cada sector tiene un modelo que mejor le funciona, para el sector servicios sería la red neuronal, para el sector industrial un modelo SARIMA en los datos sin COVID y uno SARIMAX con el *pib* y el *covid* con los datos con COVID, siendo este último modelo el que ajusta mejor. Por último para el sector Primario en los datos sin COVID el mejor modelo sería un SARIMA y en los datos con COVID sería un SARIMAX con el PIB.

Viendo estos resultados sería interesante de cara a futuras investigaciones profundizar en la relación de estas variables de PIB y afiliados a la seguridad social y también cuáles son los principales métodos de predecir del PIB. Con respecto a los sectores estudiados también sería relevante incluir en un futuro otras variables vinculadas a su actividad dentro de los modelos, por ejemplo número de turista o días con lluvia al mes en el caso del sector servicios, o el calendarios de cultivos en el sector primario.

En resumen y ya para concluir, el trabajo realizado proporciona un valioso análisis del mercado laboral gallego, revelando importantes tendencias y desafíos. Los resultados resaltan la necesidad de continuar investigando y profundizando en aspectos específicos, pero a la vez da un marco de referencia para continuar profundizando en ellos. Los dataset creados, con el cuadro de mando y los modelos ajustados servirían como punto de partida para futuras investigaciones y para informar en la toma de decisiones políticas y empresariales en la región. En última instancia, comprender mejor el mercado laboral gallego es fundamental para promover el crecimiento económico, la equidad y el bienestar de sus habitantes y por ello este trabajo puede dar un punto de partida interesante.

Apéndice A

Lista de variables eliminadas de la EPA

Variable	Descripción
NCONY	Número de orden del cónyuge o pareja en el hogar
NPADRE	Número de orden del padre en el hogar
NMADRE	Número de orden de la madre en el hogar
SIDI1	Situación inactividad autopercibida en la que estaba la semana pasada
SIDI2	Situación inactividad autopercibida en la que estaba la semana pasada
SIDI3	Situación inactividad autopercibida en la que estaba la semana pasada
SIDAC1	Situación de actividad autopercibida en la que estaba la semana pasada
SIDAC2	Situación de actividad autopercibida en la que estaba la semana pasada
TRAREM	Si ha realizado un trabajo remunerado durante la semana pasada
AYUDFA	Ayuda familiar. Realización de trabajo no remunerado empresa familiar
AUSENT	A pesar de no haber trabajado ¿tenía un empleo o negocio?
RZNOTB	Razones por las que no trabajó, teniendo empleo
VINCUL	Vinculación con el empleo de personas con empleo ausentes en la semana de referencia
NUEVEM	Ha encontrado empleo
DAUSVAC	Días de ausencia por vacaciones, días festivos, horario flexible o similar

Variable	Descripción
DAUSENF	Días de ausencia por enfermedad
DAUSOTR	Días de ausencia por otras razones
TRAANT	Si trabajó en algún momento el año pasado
EMPANT	Si ha realizado antes algún trabajo
DTANT	Número de meses transcurridos desde que dejó su último empleo
OCUPA	Ocupación u oficio que desempeñaba en su último empleo, si hace menos de un año que lo dejó
ACTA	Actividad del establecimiento donde trabajaba, si hace menos de un año que lo dejó
SITUA	Situación profesional que tenía en su anterior trabajo, si hace menos de un año que lo dejó (*)
MUN1	Lugar de residencia hace un año
PRORE1	Si es España: provincia
REPAIRE1	Región del país extranjero de residencia anterior
TCONTM	Duración en meses del contrato o relación laboral temporal (si ha trabajado un mes o más)
TCONTD	Duración en días del contrato o relación laboral temporal (si ha trabajado menos de un mes)
DREN	Tiempo en meses desde la renovación del contrato
DCOM	Tiempo en meses en la empresa

Cuadro A.1: Tabla con los nombres y descripción de las variables eliminadas en EPA_05_22_MOD.csv

Fuente: La EPA(INE)

Apéndice B

Tabla resumen: los tres sectores de actividad con mayor proporción de contratados por año. España

year	ACT1	freq
2005	Comercio al por mayor y al por menor y sus instalaciones y reparaciones.Reparación de automóviles, hostelería	0.22
	Administración Pública, educación y actividades sanitarias	0.206
	Construcción	0.117
2006	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.22
	Administración Pública, educación y actividades sanitarias	0.204
	Construcción	0.121
2007	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.221
	Administración Pública, educación y actividades sanitarias	0.204
	Construcción	0.124
2008	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.226
	Administración Pública, educación y actividades sanitarias	0.211
	Construcción	0.113
2009	Administración Pública, educación y actividades sanitarias	0.232
	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.226
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.115
2010	Administración Pública, educación y actividades sanitarias	0.24
	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.226
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.115
2011	Administración Pública, educación y actividades sanitarias	0.247
	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.231
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.115
2012	Administración Pública, educación y actividades sanitarias	0.248
	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.234
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.117
2013	Administración Pública, educación y actividades sanitarias	0.246
	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.237
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.12
2014	Administración Pública, educación y actividades sanitarias	0.248
	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.238

year	ACT1	freq
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.121
2015	Administración Pública, educación y actividades sanitarias	0.247
	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.239
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.12
2016	Administración Pública, educación y actividades sanitarias	0.245
	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.239
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.122
2017	Administración Pública, educación y actividades sanitarias	0.244
	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.236
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.122
2018	Administración Pública, educación y actividades sanitarias	0.247
	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.234
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.121
2019	Administración Pública, educación y actividades sanitarias	0.249
	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.231
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.123
2020	Administración Pública, educación y actividades sanitarias	0.26
	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.218
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.126
2021	Administración Pública, educación y actividades sanitarias	0.263
	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.213
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.128
2022	Administración Pública, educación y actividades sanitarias	0.26
	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.219
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.125

Cuadro B.1: Tabla resumen con los sectores de actividad con mayor proporción de trabajadores en España

Fuente: Elaboración propia a partir de datos de la EPA(INE)

Apéndice C

Tabla resumen: los tres sectores de actividad con mayor proporción de contratados por año. Galicia

year	ACT1	freq
2005	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.213
	Administración Pública, educación y actividades sanitarias	0.197
	Agricultura, ganadería, silvicultura y pesca	0.107
	Construcción	0.107
2006	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.21
	Administración Pública, educación y actividades sanitarias	0.206
	Construcción	0.112
2007	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.217
	Administración Pública, educación y actividades sanitarias	0.2
	Construcción	0.116
2008	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.22
	Administración Pública, educación y actividades sanitarias	0.199
	Construcción	0.106
2009	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.223
	Administración Pública, educación y actividades sanitarias	0.208
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.098
2010	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.222
	Administración Pública, educación y actividades sanitarias	0.217
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.098
2011	Administración Pública, educación y actividades sanitarias	0.228
	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.223
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.1
2012	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.229
	Administración Pública, educación y actividades sanitarias	0.229
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.099
2013	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.235
	Administración Pública, educación y actividades sanitarias	0.228
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.1
2014	Administración Pública, educación y actividades sanitarias	0.239

year	ACT1	freq
	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.233
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.103
2015	Administración Pública, educación y actividades sanitarias	0.245
	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.227
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.101
2016	Administración Pública, educación y actividades sanitarias	0.24
	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.225
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.101
2017	Administración Pública, educación y actividades sanitarias	0.236
	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.232
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.104
2018	Administración Pública, educación y actividades sanitarias	0.243
	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.233
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.112
2019	Administración Pública, educación y actividades sanitarias	0.241
	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.223
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.112
2020	Administración Pública, educación y actividades sanitarias	0.25
	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.211
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.111
2021	Administración Pública, educación y actividades sanitarias	0.252
	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.21
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.118
2022	Administración Pública, educación y actividades sanitarias	0.243
	Comercio al por mayor y al por menor y sus instalaciones y reparaciones. Reparación de automóviles, hostelería	0.218
	Intermediación financiera, seguros, actividades inmobiliarias, servicios profesionales, científicos, administrativos y otros	0.117

Cuadro C.1: Tabla resumen con los sectores de actividad con mayor proporción de trabajadores en Galicia

Fuente: Elaboración propia a partir de datos de la EPA(INE)

Apéndice D

Guía del cuadro de mando

En el siguiente apéndice explicaremos con imágenes como funciona la interfaz de usuario del cuadro de mando; los inputs o entradas del usuario estarán señalizados en la imagen con una letra en color rojo, las salidas que da el programa tanto el formato gráfica o en formato tabla se señalaran con un número en azul.

D.1. Inicio

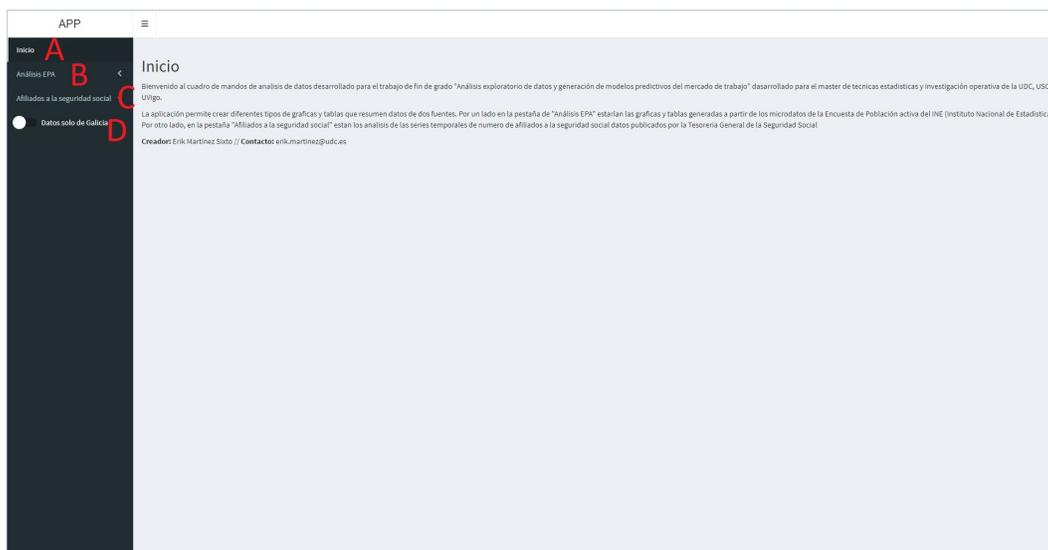


Figura D.1: Página de Inicio

Fuente: Captura de pantalla del cuadro de mando

- **A:** Botón de inicio.
- **B:** Menú desplegable para gráficas y análisis de los datos de la Encuesta de Población Activa del INE.
- **C:** Menú desplegable para gráficos y análisis de los datos de Afiliados de la seguridad social de la tesorería general de la seguridad social.
- **D:** Botón "Datos solo de Galicia". Al pulsarlo todos los análisis y gráficas (excepto el de Horas) se centran solo en Galicia y en sus provincias.

D.2. Análisis EPA

D.2.1. Temporalidad

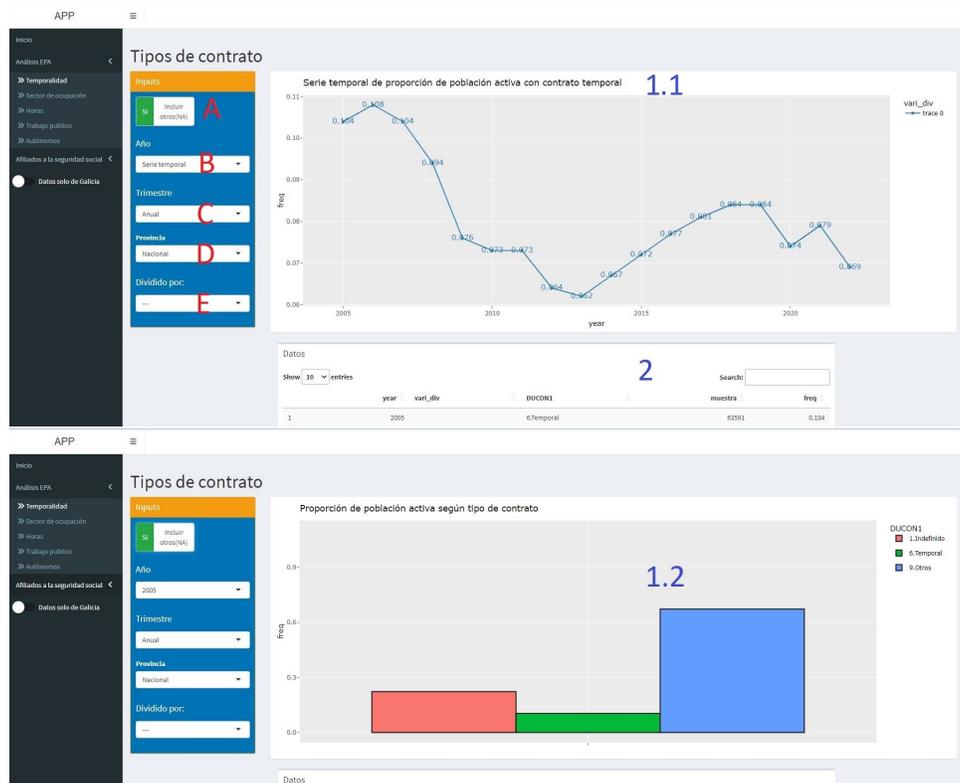


Figura D.2: Página de Temporalidad
Fuente: Captura de pantalla del cuadro de mando

- **A:** Botón *Incluir otros(NA)* sirve para decidir si las proporciones se calculan sobre el total de población de Ocupados dividiendo en tres grupos (Sí) o sobre el total de asalariados y dividiendo en dos grupos (No).
- **B:** Seleccionar *Año*. Filtra los datos por año concreto o crea una serie temporal con solo la proporción de temporalidad.
- **C:** Seleccionar *Trimestre*. Filtrar los datos por trimestre de la encuesta.
- **D:** Seleccionar *Provincia*. Filtrar los datos por provincia.
- **E:** Seleccionar *Dividir por*. Todas las gráficas y tablas pueden segmentadas en grupos en base a las variables: *Grupos de edad, Sexo, Nacionalidad, Nivel de estudios y Sección CNAE*.
- **1.1:** Gráfico de líneas con la proporción de temporalidad (Solo si en *Año* esta seleccionado *Serie temporal*)
- **1.2:** Gráfico de barras con la proporción de población de Ocupados según contrato (Solo si en *Año* esta seleccionado algún año)
- **2:** Tabla resumen de la información que se visualiza en las gráficas, añadiendo la columna muestra para saber en numero de individuos que están en cada categoría.

D.2.2. Sector de ocupación

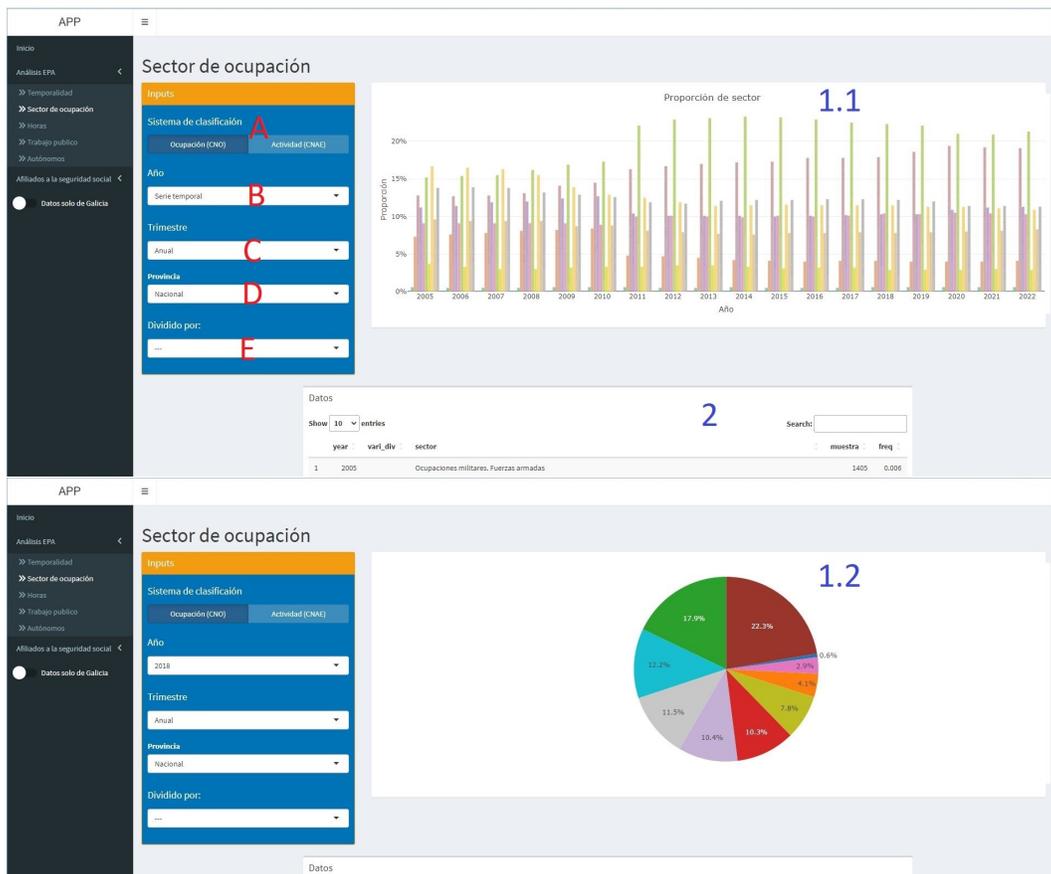


Figura D.3: Página de Sector de ocupación

Fuente: Captura de pantalla del cuadro de mando

- **A:** Botón de selección de *Sistema de clasificación*. Sirve para seleccionar el sistema de clasificación por sectores de ocupación o por sectores de actividad.
- **B:** Seleccionar *Año*. Filtra los datos por año concreto o crea una gráfico can todos los años.
- **C:** Seleccionar *Trimestre*. Filtrar los datos por trimestre de la encuesta.
- **D:** Seleccionar *Provincia*. Filtrar los datos por provincia.
- **E:** Seleccionar *Dividir por*. Todas las gráficas y tablas pueden segmentadas en grupos en base a las variables: *Grupos de edad, Sexo, Nacionalidad y Nivel de estudios*.
- **1.1:** Gráfico de barras con la proporción de población de Ocupados por sector de actividad (Solo si en *Año* esta seleccionado *Serie temporal*)
- **1.2:** Gráfico de sectores con la proporción de población de Ocupados por sector de actividad (Solo si en *Año* esta seleccionado algún año)
- **2:** Tabla resumen de la información que se visualiza en las gráficas, añadiendo la columna muestra para saber en numero de individuos que están en cada categoría.

D.2.3. Horas



Figura D.4: Página de Horas
Fuente: Captura de pantalla del cuadro de mando

- **A:** Seleccionar *Año*. Filtrar los datos por año concreto.
- **B:** Seleccionar *Trimestre*. Filtrar los datos por trimestre de la encuesta.
- **C:** Botón *Horas de trabajo semanal*. Sirve para seleccionar si se quiera la media de horas o la mediana.
- **D:** Botón *Tipos horas de trabajo semanal*. Sirve para seleccionar si se quieren usar las horas por contrato o las horas que el encuestado afirma trabajar (*Reales*).
- **1:** Mapa interactivo que permite ver las horas y al pasar el cursor sobre las provincias.

D.2.4. Trabajo público



Figura D.5: Página de Trabajo público
Fuente: Captura de pantalla del cuadro de mando

- **A:** Seleccionar *Año*. Filtra los datos por año concreto o crea un gráfico con todos los años.
- **B:** Seleccionar *Trimestre*. Filtrar los datos por trimestre de la encuesta.
- **C:** Seleccionar *Provincia*. Filtrar los datos por provincia.
- **D:** Seleccionar *Sexo*. Filtrar los datos por sexo.
- **1.1:** Gráfico de barras con el porcentaje de trabajadores públicos por tipo de administración. Todos los años (Solo si en *Año* esta seleccionado *Serie temporal*)
- **1.2:** Gráfico de sectores con el porcentaje de trabajadores públicos por tipo de administración (Solo si en *Año* esta seleccionado algún año)
- **2.1:** Gráfico de barras con el porcentaje de trabajadores públicos por nivel de estudios. Todos los años (Solo si en *Año* esta seleccionado *Serie temporal*)
- **2.2:** Gráfico de sectores con el porcentaje de trabajadores públicos por nivel de estudios (Solo si en *Año* esta seleccionado algún año)

- **3:** Porcentaje de de trabajadores públicos sobre el total de población de Ocupados (Solo si en *Año* esta seleccionado algún año)
- **4:** Porcentaje de trabajadores públicos con un contrato temporal (Solo si en *Año* esta seleccionado algún año)
- **5:** Media de horas de trabajo de los trabajadores públicos (Solo si en *Año* esta seleccionado algún año)

D.2.5. Autónomos



Figura D.6: Página de Autónomos
Fuente: Captura de pantalla del cuadro de mando

- **A:** Seleccionar *Año*. Filtra los datos por año concreto o crea una gráfico con todos los años.
- **B:** Seleccionar *Trimestre*. Filtrar los datos por trimestre de la encuesta.
- **C:** Seleccionar *Provincia*. Filtrar los datos por provincia.
- **D:** Seleccionar *Sexo*. Filtrar los datos por sexo.
- **1.1:** Gráfico de barras con el porcentaje de autónomos por sector de activad. Todos los años (Solo si en *Año* esta seleccionado *Serie temporal*)
- **1.2:** Gráfico de sectores con el porcentaje de autónomos por sector de activad (Solo si en *Año* esta seleccionado algún año)
- **2.1:** Gráfico de barras con el porcentaje de autónomos por nivel de estudios. Todos los años (Solo si en *Año* esta seleccionado *Serie temporal*)
- **2.2:** Gráfico de sectores con el porcentaje de autónomos por nivel de estudios (Solo si en *Año* esta seleccionado algún año)

- **3:** Porcentaje de autónomos sobre el total de población de Ocupados (Solo si en *Año* esta seleccionado algún año)
- **4:** Media de horas de trabajo de los autónomos (Solo si en *Año* esta seleccionado algún año)

D.3. Afiliados a la seguridad social

D.3.1. Por sexo y edad



Figura D.7: Página de Afiliados a la seguridad social por sexo y edad

Fuente: Captura de pantalla del cuadro de mando

- A: Selector múltiple de *Provincia*. Filtrar los datos por provincia.
- B: Seleccionar *Sexo*. Filtrar los datos por sexo.
- C: Seleccionar *Edad*. Filtra los datos por grupo de edad.
- B: Seleccionar *Rango de fechas*. Filtrar los datos por rangos de fecha.
- 1: Gráfico de líneas del número de afiliados a la Seguridad Social cada línea representa una provincia.
- 2: Tabla resumen de la información que se visualiza en las gráficas.

D.3.2. Por sector de actividad



Figura D.8: Página de Afiliados a la seguridad social por sector de actividad
Fuente: Captura de pantalla del cuadro de mando

- A: Selector múltiple de *Provincia*. Filtrar los datos por provincia.
- B: Seleccionar *Sector de actividad*. Filtrar los datos por sector de actividad.
- C: Seleccionar *Régimen*. Filtra los datos por el régimen de la Seguridad Social.
- B: Seleccionar *Rango de fechas*. Filtrar los datos por rangos de fecha.
- 1: Gráfico de líneas del número de afiliados a la Seguridad Social cada línea representa una provincia.
- 2: Tabla resumen de la información que se visualiza en las gráficas.

D.3.3. Serie temporal



Figura D.9: Página de Serie temporal
Fuente: Captura de pantalla del cuadro de mando

- **A:** Selector de *Provincia*. Filtrar los datos por provincia.
- **B:** Seleccionar *Tipos de clasificación*. Ambos sistemas están basados en el sistema CNAE pero el primero esta menos agregado y el segundo con más nivel de agregación.

- **C**: Seleccionar *Sector de actividad*. Filtra los datos por sector de actividad.
- **B**: Seleccionar *Régimen*. Filtrar los datos por régimen de la seguridad social.
- **1**: Gráfico con la serie temporal de afiliados a la Seguridad Social.
- **2**: Gráfico con la tendencia de la serie temporal de afiliados a la seguridad social obtenido con X-13ARIMA-SEATS.
- **3**: Gráfico con el error de la serie temporal de afiliados a la seguridad social obtenido con X-13ARIMA-SEATS.
- **4**: Gráfico de estacionalidad de afiliados a la seguridad social.
- **5**: Gráfico con la estacionalidad de la serie temporal de afiliados a la seguridad social obtenido con X-13ARIMA-SEATS.

Índice de figuras

3.1. Pestañas de los datos de la EPA	10
4.1. Proporción de población según AOI (si son Inactivos, Ocupados o Parados). España . . .	14
4.2. Proporción de tipos de contrato DUCON1 (indefinido o temporal). España	15
4.3. Proporción de tipos de contrato DUCON2 (permanente y discontinuo). España	15
4.4. Porcentaje de trabajadores del sector público por tipo de administración. España	17
4.5. Porcentaje de autónomos por sector. España	18
4.6. Proporción de población según AOI (si son inactivos, Ocupados o Parados). Galicia . . .	19
4.7. Proporción de población según AOI (si son inactivos, Ocupados o Parados) por sexo. Galicia	19
4.8. Proporción de asalariados con contrato temporal. Galicia	20
4.9. Proporción de asalariados con contrato temporal por sexo. Galicia	21
4.10. Proporción de asalariados con contrato temporal por grupo de edad. Galicia	21
4.11. Proporción de asalariados con contrato temporal por nivel de estudios. Galicia	22
4.12. Porcentaje de Ocupados por sector de actividad y sexo. Galicia	24
4.13. Porcentaje de Ocupados por sector de actividad y grupo edad. Galicia	25
4.14. Porcentaje de Ocupados por sector de actividad y situación de nacionalidad. Galicia . . .	26
4.15. Porcentaje de Ocupados por sector de actividad y nivel de estudios. Galicia	27
4.16. Porcentaje de trabajadores de sector público por tipo de administración. Galicia	28
4.17. Porcentaje de trabajadores autónomos por sector de ocupación. Galicia	29
4.18. Número de afiliados medios por sexo. España	30
4.19. Número de afiliados medios por grupos de edad. España	31
4.20. Número de afiliados medios por sector de ocupación. España	31
4.21. Número de afiliados medios por sexo. Galicia	32
4.22. Número de afiliados medios por grupos de edad. Galicia	32
4.23. Número de afiliados medios por sector de ocupación. Galicia	33
5.1. Predicciones del número de afiliados. Dataset conjunto. Galicia	39
5.2. Predicciones del número de afiliados. Dataset covid. Galicia	40
5.3. Predicciones del número de afiliados. Dataset servicios. Galicia	41
5.4. Predicciones del número de afiliados. Dataset servicios covid. Galicia	42
5.5. Predicciones del número de afiliados. Dataset industria. Galicia	43
5.6. Predicciones del número de afiliados. Dataset industria covid. Galicia	43
5.7. Predicciones del número de afiliados. Dataset primario. Galicia	44
5.8. Predicciones del número de afiliados. Dataset primario covid. Galicia	45
D.1. Página de Inicio	57
D.2. Página de Temporalidad	58
D.3. Página de Sector de ocupación	59
D.4. Página de Horas	60

D.5. Página de Trabajo público	61
D.6. Página de Autónomos	63
D.7. Página de Afiliados a la seguridad social por sexo y edad	65
D.8. Página de Afiliados a la seguridad social por sector de actividad	66
D.9. Página de Serie temporal	67

Índice de cuadros

4.1.	Tabla con los valores de las variables OCUP1 y ACT1 de la EPA. Fuente: EPA (INE)	16
4.2.	Tabla resumen sobre trabajadores del sector público. Fuente: Elaboración propia a partir de datos la EPA (INE)	17
4.3.	Tabla resumen sobre trabajadores Autónomos. Fuente: Elaboración propia a partir de datos la EPA (INE)	18
4.4.	Tabla con los valores de variable de nivel de estudios. Fuente: EPA (INE)	22
4.5.	Tabla resumen sobre trabajadores del sector público. Galicia Fuente: Elaboración propia a partir de datos la EPA (INE)	28
4.6.	Tabla resumen sobre trabajadores autónomos. Galicia Fuente: Elaboración propia a partir de datos la EPA (INE)	29
5.1.	Tabla con los datos de fuerza de la tendencia y de la estacionalidad por sectores. Galicia Fuente: Elaboración propia a partir de datos de Afiliados a la seguridad social(Seguridad social)	38
5.2.	Errores de los mejores modelos de cada tipo. Dataset conjunto Fuente: Elaboración propia	39
5.3.	Errores de los mejores modelos de cada tipo. Dataset covid Fuente: Elaboración propia	40
5.4.	Errores de los mejores modelos de cada tipo. Dataset servicios Fuente: Elaboración propia	41
5.5.	Errores de los mejores modelos de cada tipo. Dataset servicios covid Fuente: Elaboración propia	41
5.6.	Errores de los mejores modelos de cada tipo. Dataset industria Fuente: Elaboración propia	42
5.7.	Errores de los mejores modelos de cada tipo. Dataset industria covid Fuente: Elaboración propia	42
5.8.	Errores de los mejores modelos de cada tipo. Dataset primario Fuente: Elaboración propia	44
5.9.	Errores de los mejores modelos de cada tipo. Dataset primario covid Fuente: Elaboración propia	44
A.1.	Tabla con los nombres y descripción de las variables eliminadas en EPA_05_22_MOD.csv Fuente: La EPA(INE)	52
B.1.	Tabla resumen con los sectores de actividad con mayor proporción de trabajadores en España Fuente: Elaboración propia a partir de datos de la EPA(INE)	54
C.1.	Tabla resumen con los sectores de actividad con mayor proporción de trabajadores en Galicia Fuente: Elaboración propia a partir de datos de la EPA(INE)	56

Bibliografía

- Alberto José Pazo y María Pilar Moragón (2018). El despoblamiento en Galicia: la visualización de la “catástrofe”. *Ager. Revista de estudios sobre despoblación y desarrollo rural*, 1(24):123–154.
- Braña, Francisco y Antón, J. (2007). Determinantes de la inserción laboral de los jóvenes en España. revisión y nueva evidencia empírica. *Anuales de economía aplicada*, 10(1):44–67.
- Chang, W., Ribeiro, B. B., Studio, A., and Chang, M. W. (2022). Package shinydashboard.
- Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I. (1990). Stl: A seasonal-trend decomposition. *J. Off. Stat.*, 6(1):3–73.
- De la Rica, S. (2010). La penalización salarial de la temporalidad: ¿Qué efectos tiene en las decisiones familiares? *Dolado, JJ y F. Felgueroso (coordinadores): Propuesta para la reactivación laboral en España. Fedea. Madrid.—(2015):”Políticas activas de empleo: una panorámica”, Fedea Policy Papers–2015/01. Fedea. Madrid.*
- Instituto Galego de Estatística (2022). Contas económicas anuais. Revisión estadística 2019. https://www.ige.gal/estatico/pdfs/s3/metodoloxias/met_contas_economicas_gl.pdf. Accedido el 01 de Mayo de 2023.
- Instituto Galego de Estatística (2023). Cadro macroeconómico de Galicia. Revisión estadística 2019. Serie 1995-2020. <https://www.ige.gal/igebdt/selector.jsp?COD=9606&c=0307007001&idioma=gl>. Accedido el 27 de Abril de 2023.
- Instituto Nacional de Estadística (2016). Encuesta de población Activa. Diseño de la Encuesta y Evaluación de la calidad de los datos. Informe Técnico. https://www.ine.es/inebaseDYN/epa30308/docs/epa05_disenc.pdf. Accedido el 31 de Mayo de 2023.
- Instituto Nacional de Estadística (2017). Encuesta de población Activa, Metodología 2005. Descripción general de la encuesta. <https://www.ine.es/inebaseDYN/epa30308/docs/resumetepa.pdf>. Accedido el 31 de Mayo de 2023.
- Instituto Nacional de Estadística (2021a). Encuesta de población Activa. Diseño de la Encuesta y Evaluación de la calidad de los datos. Informe Técnico. <https://www.ine.es/inebaseDYN/epa30308/docs/resumetepa21.pdf>. Accedido el 31 de Mayo de 2023.
- Instituto Nacional de Estadística (2021b). Encuesta de población Activa, Metodología 2021 Descripción general de la encuesta. <https://www.ine.es/inebaseDYN/epa30308/docs/resumetepa21.pdf>. Accedido el 31 de Mayo de 2023.
- Instituto Nacional de Estadística (2021c). Resumen de los cambios en la Encuesta de Población Activa (EPA) en 2021. https://www.ine.es/inebaseDYN/epa30308/docs/epa_cambios2021.pdf. Accedido el 31 de Mayo de 2023.

- Instituto Nacional de Estadística (2023a). Encuesta de Población Activa. Descarga de microdatos. https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176918&menu=resultados&idp=1254735976595#!tabs-1254736030639. Accedido el 08 de Febrero de 2023.
- Instituto Nacional de Estadística (2023b). Estadística del Padrón continuo. Resultados. https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177012&menu=resultados&idp=1254734710990. Accedido el 12 de Julio de 2023.
- Instituto Nacional de Estadística (2023). Informe Metodológico Estandarizado EPA. <https://www.ine.es/dynt3/metadatos/es/RespuestaDatos.html?oe=30308>. Accedido el 31 de Mayo de 2023.
- Instituto Nacional de Estadística (2023). Informe Metodológico Estandarizado IPC. <https://www.ine.es/dynt3/metadatos/es/RespuestaDatos.html?oe=30138>. Accedido el 04 de Junio de 2023.
- Luis Coronilla (2023). Cuando se celebra la Semana Santa? <https://www.ige.gal/igebdt/selector.jsp?COD=9606&c=0307007001&idioma=gl>. Accedido el 20 de Mayo de 2023.
- Maruri, I. A., Claramunt, C. O., and Guillén, A. S. (2008). Inmigración, mercado de trabajo y seguridad social: evidencia empírica española e implicaciones de política económica. *Revista del Ministerio de Trabajo e Inmigración*, 81:17–39.
- O’Hara-Wild, M., Hyndman, R., and Wang, E. (2022). *Feasts: Feature Extraction and Statistics for Time Series*. <http://feasts.tidyverts.org/>, <https://github.com/tidyverts/feasts/>.
- O’Hara-Wild, M., Hyndman, R., and Wang, E. (2023). Fable. <https://fable.tidyverts.org/index.html>. Accedido el 26 de Junio de 2023.
- Rob J. Hyndman y Yeasmin Khandakar (2008). Automatic Time Series Forecasting. The forecast Package for R. *Journal of Statistical Software*, 27(3):1–22.
- Rob J. Hyndman y Yeasmin Khandakar (2021). Forecasting: principles and practice, 3rd edition. [OTexts.com/fpp3](https://otexts.com/fpp3). Accedido el 24 de Junio de 2023.
- Sax, Christoph y Eddelbuettel, D. (2018). Seasonal adjustment by X-13ARIMA-SEATS in R. *Journal of Statistical Software*, 87(11):1–17.
- Sievert, C. (2020). *Interactive Web-Based Data Visualization with R, plotly, and shiny*. Chapman and Hall/CRC.
- Tesorería General de la Seguridad Social (2023a). Estadística de afiliación a la Seguridad Social. Informe metodológico estandarizado. https://www.seg-social.es/wps/wcm/connect/wss/b7e3b61f-4001-404f-b7e0-b2aaf91a785d/2023-02+IME.pdf?MOD=AJPERES&CONVERT_T0=linktext&CACHEID=ROOTWORKSPACE.Z18_2G50H38209D640QTQ570VB2000-b7e3b61f-4001-404f-b7e0-b2aaf91a785d-oomPec7. Accedido el 01 de Mayo de 2023.
- Tesorería General de la Seguridad Social (2023b). Notas explicativas referentes a los PX-web de Afiliados en alta laboral. https://www.seg-social.es/wps/wcm/connect/wss/3827e6b3-7a01-4258-94e7-c0a4d553e434/20230524+Notas+explicativas+PX+WEB+actualizada.pdf?MOD=AJPERES&CONVERT_T0=linktext&CACHEID=ROOTWORKSPACE.Z18_2G50H38209D640QTQ570VB2000-3827e6b3-7a01-4258-94e7-c0a4d553e434-oxi6yQk. Accedido el 01 de Mayo de 2023.
- Torres, R. (2018). El mercado laboral español: situación y desafíos estructurales. *Cuadernos de información económica*, 1(267):1–12.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wickham, H. (2021). *Mastering shiny*. O'Reilly Media, Inc.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., çois, R. F., Golemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686.