



Universidade de Vigo

Trabajo Fin de Máster

Modelado y predicción del Customer Lifetime Value en Banca

Sergio Cambeiro Agulleiro

Máster en Técnicas Estadísticas

Curso 2019-2020

Propuesta de Trabajo Fin de Máster

Título en galego: Modelado e predición do Customer Lifetime Value na Banca
Título en español: Modelado y predicción del Customer Lifetime Value en Banca
English title: Modeling and prediction of Customer Lifetime Value in Banking
Modalidad: Modalidad B
Autor: Sergio Cambeiro Agulleiro, Universidade da Coruña
Director: Ricardo Cao Abad, Universidade da Coruña
Tutor: Jorge García Romarís, ABANCA - Data, Analytics & RTech
Breve resumen del trabajo: La adquisición y retención de clientes rentables es un desafío cada vez mayor para los bancos. Una de las métricas que más puede ayudar a la hora de conocer a los clientes en cuanto a rentabilidad es el customer lifetime value (CLV). Los objetivos de este trabajo son identificar las variables relevantes para el cálculo del CLV, desarrollar un modelo para el cálculo del CLV a partir de las variables relevantes y desarrollar un modelo predictivo para ayudar a definir estrategias de marketing.
Recomendaciones:
Otras observaciones:

Don Ricardo Cao Abad, Catedrático de la Universidade da Coruña, y don Jorge García Romarís, Director de ABANCA - Data, Analytics & RTech, informan que el Trabajo Fin de Máster titulado

Modelado y predicción del Customer Lifetime Value en Banca

fue realizado bajo su dirección por don Sergio Cambeiro Agulleiro para el Máster en Técnicas Estadísticas. Estimando que el trabajo está terminado, dan su conformidad para su presentación y defensa ante un tribunal.

En A Coruña, a 5 de septiembre de 2020.

El director:

El tutor:

Don Ricardo Cao Abad

Don Jorge García Romarís

El autor:

Don Sergio Cambeiro Agulleiro

Índice general

Resumen	IX
1. Customer lifetime value	1
1.1. Introducción	1
1.2. Definición	2
2. Fundamentación teórica	5
2.1. Relación entre variables	5
2.1.1. Análisis de componentes principales (ACP)	5
2.1.2. Coeficiente de correlación de Pearson	5
2.2. Análisis de grupos	6
2.2.1. Expectation maximization (EM) clustering	7
2.2.2. Agrupamiento jerárquico	9
2.3. Predicción	9
2.3.1. Regresión lineal simple	9
2.3.2. Modelo lineal generalizado (GLM)	10
2.3.3. Modelo aditivo generalizado (GAM)	10
2.4. Clasificación	11
2.4.1. Árboles de clasificación y regresión (CART)	11
2.4.2. Random forest	12
3. Análisis exploratorio de los datos	13
3.1. Data mart de clientes	13
3.2. Margen ordinario	21
3.3. Transformación de las variables	23
3.4. Variables finales	25
4. Resultados	35
4.1. Análisis de componentes principales	35
4.2. Coeficientes de correlación con MO_CLIENTE_CT	37
4.3. Cálculo de los segmentos	40
4.3.1. Segmentación en 8 clusters (<i>mclust</i>)	41
4.3.2. Segmentación en 4 clusters (<i>hclust</i>)	44
4.3.3. Segmentación en 8 clusters (<i>hclust</i>)	46
4.4. Predicción	49
4.4.1. Regresión lineal simple	49
4.4.2. Modelo lineal generalizado	51
4.4.3. Modelos aditivos generalizados	51
4.4.4. Otras pruebas	52
4.4.5. Segmentación según la precisión de la estimación	53
4.5. Árboles de clasificación	56

4.5.1. Árboles de clasificación y regresión	56
4.5.2. Random forest	56
4.5.3. Resultados finales tras la clasificación	58
5. Conclusiones	59
A. Apéndice resultados	61
A.1. Tabla resultados del ACP	61
A.2. Tabla resultados del análisis de correlaciones entre cada variable y MO_CLIENTE_CT	71
A.3. Gráficos de dispersión entre las distintas variables y MO_CLIENTE_CT	88
A.4. Gráficos de densidades de las variables utilizadas para los cluster	117
A.5. Gráficos de dispersión de las variables utilizadas para los cluster	120
A.6. GAM. Funciones estimadas y funciones de densidad	122
A.7. Gráficas de la densidades de las variables diferenciando segmentos	128
A.8. Métricas del data mart de clientes	132
A.9. Gráficos de las variables finales	166
A.10. Aplicaciones que influyen en el margen ordinario de la entidad	170
A.11. Dendrogramas obtenidos en la sección 4.3 con una muestra de 100 clientes	173
Bibliografía	175

Resumen

Resumen en español

El conocimiento del cliente y cómo enfocar los recursos de la manera más eficiente, es uno de los desafíos más importantes de una entidad financiera. Por ello, en este trabajo se formulará un modelo para la predicción del valor de vida del cliente (en inglés customer lifetime value o CLV) utilizando únicamente la información que se podría solicitar a un cliente antes de formalizar una relación contractual con la entidad financiera, así como estudiar la posibilidad de segmentar los clientes en función del CLV con la intención de conocer cómo se podría incidir en los clientes para provocar un cambio de segmento y con ello mejorar la rentabilidad de la entidad.

Para ello, se determinarán las variables que tienen más relación con el CLV, se crearán los segmentos en función de esas variables y se realizará un proceso para poder obtener una predicción para cada posible cliente. Para la realización de este trabajo se empleó el software estadístico R así como la plataforma de analítica avanzada de la entidad, Teradata Vantage.

Este trabajo se ha realizado después de utilizar técnicas de seudonimización de datos personales, cumpliendo con la Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal (LOPD). También se ha utilizado un factor de modificación para no mostrar saldos reales por motivos de confidencialidad.

English abstract

Knowing the client and how to focus resources in the most efficient way, is one of the most important challenges for a financial institution. For this reason, in this work a model will be formulated for predicting the customer lifetime value (CLV) using only the information that could be requested from a customer before formalizing a contractual relationship with the financial entity. It was also studied the possibility of clustering clients according to the CLV, to know how to influence the clients to cause a change of segment and thereby improve the entity's profitability.

For this, the variables that are most related to the CLV will be determined, the segments will be created based on those variables and a process will be carried out to obtain a prediction for each possible client. To carry out this work, the R statistical software was used, as well as the entity's advanced analytics platform, Teradata Vantage.

This work has been carried out after using techniques of pseudonymisation of personal data, complying with Organic Law 15/1999 of December 13 on Protection of Personal Data. A modification factor has also been used to not show actual balances for confidentiality reasons.

Capítulo 1

Customer lifetime value

CLV son las siglas de customer lifetime value o el valor de vida de un cliente en español. El CLV es una métrica que indica el beneficio que genera un cliente a lo largo de su relación con una empresa.

1.1. Introducción

La adquisición y retención de clientes rentables es un desafío cada vez mayor para los bancos. Los datos demográficos del cliente, su comportamiento de compra o sus necesidades están evolucionando rápidamente. Los bancos ahora necesitan una vista de 360 grados de cada cliente con el fin de enfocar sus recursos de manera eficiente.

El CLV es un concepto simple en sí mismo, pero difícil de implementar en un contexto comercial complejo. Para obtener la máxima precisión en el cálculo, se debe tener en cuenta una gran cantidad de datos, por ejemplo:

- Nociones de adquisición y desgaste del cliente.
- Tenencia de diversos productos y servicios bancarios, su volumen y probabilidad de uso.
- Otras características de los clientes como geográfica, datos demográficos y de mercado.

Pero en ocasiones, se necesita obtener una predicción sin tener disponibles todos estos datos, como por ejemplo, a la hora de formalizar por primera vez una relación contractual con un cliente o, a mayor escala, para estudiar la viabilidad de una fusión con otra entidad. En estos escenarios, la predicción del CLV de los clientes sería de mucha importancia para estudiar la viabilidad de la operación pero la disponibilidad de información es menor.

Todo esto pone de manifiesto la importancia del cálculo del CLV sin las limitaciones e inconvenientes de la mayoría de los modelos encontrados en la literatura, utilizando en su lugar un modelo de fácil interpretación con indicadores objetivos y fácilmente medibles.

En este trabajo se estudiarán las posibilidades de clasificación de clientes en segmentos según su CLV y se trabajará en la predicción del CLV de un cliente en potencia, en función de unas variables susceptibles de ser obtenidas antes de que este cliente en potencia sea realmente un cliente de la entidad bancaria.

Para explicar todo esto, primeramente será descrito el problema en el capítulo 1. Se comentarán algunos de los métodos de cálculo del CLV disponibles en la literatura y se explicará en detalle el

procedimiento que se propondrá en este trabajo. Además, también se profundizará en las motivaciones de predecir esta métrica. En el capítulo 2 se detallarán los fundamentos teóricos en los que se basa todo el procedimiento llevado a cabo en las distintas fases de este trabajo. En el capítulo 3, se describirá y analizará el conjunto de datos de la entidad financiera para poder llevar a cabo los objetivos de este trabajo. A continuación, en el capítulo 4 se detallarán los resultados obtenidos y en el capítulo 5 se incluirán las conclusiones alcanzadas con los resultados obtenidos.

Por último se añade un apéndice con varias entradas, al que se hace referencia desde distintos capítulos, para ayudar a la comprensión de las ideas expuestas.

1.2. Definición

Como ya se introdujo en el anterior capítulo, este trabajo gira en torno al concepto de **customer lifetime value** o **valor de vida del cliente**. El CLV es la medida que se utiliza para identificar **el beneficio potencial de un cliente para una empresa**. Es un valor muy interesante porque puede ser de mucha ayuda en la toma de decisiones en diversos departamentos de una empresa. Tiene utilidad en los siguientes departamentos:

- **Marketing.** Una vez que se tienen claros algunos aspectos como el público objetivo y las aspiraciones de una campaña de marketing, el CLV ayudará a determinar cuánto se puede gastar en esa campaña porque se tendrán estimaciones de lo que podrían aportar los nuevos clientes.
- **Atención al cliente.** La satisfacción de los clientes es un aspecto clave para una empresa y para ello la atención al cliente es fundamental. El problema muchas veces está en conocer cuánto debe invertir la empresa en la atención al cliente y es ahí donde el CLV puede ser de mucha ayuda.
- **Ventas.** Cada cliente tiene sus necesidades y generalmente una empresa suele tener varias opciones para satisfacerlas. A la hora de ofrecer una de las distintas opciones al cliente, el CLV ayudaría a elegir la que satisface al cliente que más beneficia también a la empresa.

Además, el CLV puede necesitarse en situaciones muy diferentes de una empresa como se puede comprobar en los artículos que se mencionan a continuación.

Ekinci et al. (2014) comentan los resultados obtenidos en su estudio que tiene como objetivo eliminar las limitaciones e inconvenientes de la mayoría de los modelos encontrados en la literatura a través de un modelo simple e industrial con **indicadores fácilmente medibles y objetivos**. Además, **se predice el valor potencial de los clientes en lugar de medir el valor actual**, que generalmente se ha utilizado en la mayoría de los estudios anteriores.

Glady et al. (2009) exponen que la definición y el modelado de la fidelización de los clientes han sido temas centrales durante muchos años en las relaciones entre una empresa y sus clientes. Por ello, con su predicción del CLV tienen como finalidad **detectar clientes que se están volviendo menos fieles y predecir así los clientes que tienen más probabilidades de abandonar la entidad**, también llamados *churners*. En definitiva, se utiliza el valor de vida del cliente para detectar cuándo el beneficio marginal de un cliente está disminuyendo.

Haenlein et al. (2007) hacen referencia a que en años previos a la crisis financiera del 2008, la banca ya comenzó a enfrentar un conjunto de cambios que tuvieron un impacto general negativo en el margen y la rentabilidad de la industria. En respuesta a estos cambios, muchos bancos se han centrado en aumentar la escala de sus operaciones, lo que ha llevado a un **aumento de fusiones y adquisiciones de carteras de clientes**. Generalmente motivados por el supuesto de que el banco adquirente

pueda gestionar esa base de clientes de una forma más rentable de lo que el banco vendedor pudo. Este comportamiento fue otro de los impulsores de las mejoras sobre la valoración de los clientes y, por lo tanto, del interés en el cálculo del CLV.

Como vemos, las finalidades pueden ser diversas y en el caso de este trabajo, la más parecida de las anteriores sería la de Ekinçi et al. (2014). El objetivo en este trabajo, también es conseguir un modelo de predicción del valor potencial de los clientes a partir de indicadores que sean relativamente fácilmente obtenibles o, al menos, que se puedan tener antes de que el cliente formalice su relación con la entidad. Es decir, se buscan indicadores que se puedan obtener a partir de facturas del cliente por ejemplo o que se puedan calcular en función de los objetivos de inversión/financiación del cliente en la entidad.

Glady et al. (2009) definen el CLV como una función de todas las transacciones que un cliente va hacer, para los q productos que la entidad está vendiendo. Por lo tanto, el valor de vida del cliente i , hasta el período h desde el período t , es la suma de los flujos de efectivo netos $CF_{i,j,t+k}$, producidos por las transacciones del producto j , aplicando la tasa de descuento r (una tasa fija que representa el precio del dinero) definido como:

$$CLV_{i,t} = \sum_{k=1}^h \sum_{j=1}^q \frac{1}{(1+r)^k} CF_{i,j,t+k} \quad (1.1)$$

En nuestro caso, en lugar de disponer de los flujos de efectivo netos, tenemos directamente el margen ordinario del cliente. Y para comprender este margen es necesario indicar los tres principales márgenes de una entidad de crédito:

- En el entorno de las entidades de crédito, **el margen de intermediación** hace referencia al primer escalón de su cuenta de resultados, el cual representa la diferencia entre el producto obtenido por la colocación de los recursos (intereses cobrados o productos financieros) y el coste de captación de los mismos (intereses pagados o costes financieros). También se le denomina **margen financiero**.
- **El margen ordinario** se obtiene sumando al margen de intermediación, los ingresos derivados de otros productos bancarios. En realidad, el margen ordinario hace referencia al negocio generado por la entidad sin considerar costes no financieros. Los componentes de estos ingresos ordinarios son básicamente tres. En primer lugar, los ingresos netos por comisiones derivadas de las prestación de servicios (cobro de cheques, cajas de seguridad, domiciliaciones, fondos de inversión,...); en segundo lugar, los resultados netos de las diferencias de cambio y, por último, los ingresos netos vinculados a avales y créditos documentarios.
- **El margen de explotación** de una entidad de crédito es el resultado de restar al margen ordinario los gastos de explotación. El margen de explotación también se denomina resultado neto, puesto que de los resultados brutos, que son el margen de explotación, se deducen los gastos de explotación.

Por lo tanto en este trabajo, en lugar de los flujos de efectivo netos, se utiliza el margen ordinario que además de los ingresos financieros, deduce los costes financieros e incluye las comisiones de otros

productos bancarios. De esta forma, el valor de vida del cliente i , hasta el período h desde el período t , es el margen ordinario $MO_{i,t+k}$ aplicando la tasa de descuento r (una tasa fija definida por el equipo de gestión financiera de la entidad) definido como:

$$CLV_{i,t} = \sum_{k=1}^h \frac{1}{(1+r)^k} MO_{i,t+k} \quad (1.2)$$

Otro punto importante, es la unidad de tiempo a tener en cuenta. El conjunto de datos del que se dispone tiene frecuencia mensual así que se podría aplicar la fórmula utilizando el período mensual. El problema es que las entidades bancarias tienen numerosos ingresos periódicos en periodos superiores al mes (trimestrales, semestrales, etc) y esto podría provocar fluctuaciones por el simple hecho del mes que se esté considerando. Por ello, se decidió anualizar las variables para trabajar así con datos anuales y estos sí que ya deberían ser más homogéneos para cada período.

Capítulo 2

Fundamentación teórica

En este capítulo se hace una revisión de los fundamentos teóricos utilizados en este trabajo para llevar a cabo el estudio del conjunto de datos.

2.1. Relación entre variables

Se utilizan principalmente dos técnicas para estudiar cómo se relacionan las variables entre ellas. En primer lugar se realiza un análisis de componentes principales porque, dada la cantidad de variables disponible, se intenta realizar una reducción de la dimensión. Posteriormente, al ver que son necesarias demasiadas componentes principales para explicar un alto porcentaje de variabilidad de los datos, se calcula el coeficiente de correlación de Pearson entre el margen ordinario, nuestra variable objetivo, y cada una de las variables del conjunto de datos.

2.1.1. Análisis de componentes principales (ACP)

El análisis de componentes principales (ACP), como explicaban Mardia et al. (1979), es una técnica utilizada para describir un conjunto de datos en términos de nuevas variables (*componentes*) no correlacionadas. Las componentes se ordenan por la proporción de varianza que describen, por lo que la técnica es útil para reducir la dimensión de un conjunto de datos.

Con el ACP se estudian las relaciones que existen entre las variables de un conjunto de datos que están correlacionadas y se transforma el conjunto original de variables en otro conjunto de nuevas variables incorreladas entre sí, llamada conjunto de componentes principales. Es decir, si las variables originales están incorreladas no es necesario un análisis de componentes principales ya que las variables originales ya son las componentes principales.

Las nuevas variables son combinaciones lineales de las variables originales y se ordenan según la cantidad de variabilidad que explican. Por lo tanto, si tenemos una serie de variables (X_1, X_2, \dots, X_p) sobre un grupo de objetos o individuos, se trata de calcular un nuevo conjunto de variables (o componentes) incorreladas entre sí (Y_1, Y_2, \dots, Y_q) , cuyas varianzas vayan decreciendo progresivamente. De manera que con un número menor de componentes ($q < p$) se explica aproximadamente lo mismo que con las p variables originales.

2.1.2. Coeficiente de correlación de Pearson

El coeficiente de correlación de Pearson, pensado para variables cuantitativas continuas, es una medida del grado de covariación entre distintas variables relacionadas linealmente, como se explica en

Edwards AL (1976). Esto significa que puede haber variables fuertemente relacionadas, pero no de forma lineal, en cuyo caso no procede aplicar la correlación de Pearson. A diferencia de la covarianza, la correlación de Pearson es independiente de la escala de medida de las variables y su valor varía entre -1 y 1.

Suponiendo que se están estudiando dos variables aleatorias X e Y sobre una población, el coeficiente de correlación de Pearson se simboliza con la letra $\rho_{X,Y}$, siendo la expresión que nos permite calcularlo:

$$\rho_{X,Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}, \quad (2.1)$$

donde:

- σ_{XY} es la covarianza de X e Y
- σ_X es la desviación típica de la variable X
- σ_Y es la desviación típica de la variable Y

Además su interpretación es muy sencilla y se resume en los siguientes puntos, teniendo en cuenta que su valor varía en el intervalo $[-1, 1]$:

- Si $\rho = 1$ existe una correlación positiva perfecta. Se denomina relación directa e indica una dependencia total.
- Si $0 < \rho < 1$, existe una correlación positiva.
- Si $\rho = 0$, no existe relación lineal, aunque podrían existir relaciones no lineales.
- Si $-1 < \rho < 0$, existe una correlación negativa.
- Si $\rho = -1$, existe una correlación negativa perfecta. Se denomina relación inversa, cuando una de las variables aumenta, la otra disminuye en proporción constante.

2.2. Análisis de grupos

El análisis de grupos o análisis cluster es un conjunto de técnicas multivariantes utilizadas para **clasificar un conjunto de individuos en grupos homogéneos**. Pertenecen, al igual que otras metodologías y que el análisis discriminante, al conjunto de técnicas que tiene por objetivo la clasificación de los individuos. La diferencia fundamental entre el análisis cluster y el discriminante reside en que en el análisis cluster los grupos son desconocidos a priori y son precisamente lo que queremos determinar; mientras que en el análisis discriminante, los grupos son conocidos y lo que pretendemos es saber en qué medida las variables disponibles nos discriminan esos grupos y nos pueden ayudar a clasificar los individuos en los grupos dados.

Se trata, fundamentalmente, de resolver el siguiente problema: Dado un conjunto de N individuos caracterizados por la información de n variables X_j , ($j = 1, 2, \dots, n$), nos planteamos el reto de ser capaces de clasificarlos de manera que los individuos pertenecientes a un grupo (cluster) (y siempre con

respecto a la información disponible) sean tan similares entre sí como sea posible, siendo los distintos grupos entre ellos tan disimilares como sea posible.

Con el análisis cluster se pretende encontrar un conjunto de grupos a los que ir asignando los distintos individuos por algún criterio de homogeneidad. Por lo tanto, se hace imprescindible definir una medida de similitud o bien de divergencia para ir clasificando los individuos en unos u otros grupos.

En las distintas pruebas realizadas se utilizaron principalmente dos medidas, que se definirán a continuación. Teniendo en cuenta que el conjunto de datos está formado por las observaciones de n variables y N individuos, podemos representar a los individuos en el espacio de las variables de manera que representamos al individuo i -ésimo por el vector:

$$W_i = \begin{pmatrix} x_{1i} \\ x_{2i} \\ \cdot \\ \cdot \\ x_{ni} \end{pmatrix}; \text{ y al individuo } j\text{-ésimo: } W_j = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \cdot \\ \cdot \\ x_{nj} \end{pmatrix}$$

La primera de las distancias utilizada en este trabajo fue la euclidiana, con las variables tipificadas para evitar la sensibilidad de esta distancia a las unidades de medida de las variables. Su expresión es la siguiente:

$$d(i, j) = (W_i - W_j)'(W_i - W_j)$$

Uno de los inconvenientes que puede tener es la sensibilidad a las unidades de medida de las variables y esto se solucionaba con la tipificación previa de las variables, pero otro inconveniente puede ser que las variables estén correlacionadas ya que en estos casos la distancia euclidiana obtendrá información redundante que distorsionaría los resultados. Por ello se realizaron también pruebas con la distancia de Mahalanobis que soluciona ambos inconvenientes. La distancia de Mahalanobis entre los individuos i y j la definimos por la expresión:

$$d(i, j) = (W_i - W_j)' V^{-1} (W_i - W_j)$$

donde la matriz V^{-1} asociada a la forma cuadrática es la inversa de la matriz de varianzas V . En cuanto a los métodos de cálculo de los clusters se emplearon los siguientes: model based clustering y el clustering jerarquizado.

2.2.1. Expectation maximization (EM) clustering

Uno de los métodos utilizados en este trabajo es el clustering basado en modelos (**Model Based Clustering**, Dempster AP et al. (1977)) que considera que las observaciones proceden de una distribución que es a su vez una combinación de dos o más componentes (clusters), cada uno con una distribución propia. En principio, cada cluster puede estar descrito por cualquier función de densidad, pero normalmente se asume que siguen una distribución multivariante normal. Para estimar los parámetros que definen la función de distribución de cada cluster (media y matriz de covarianza si

se asume que son de tipo normal) se recurre al algoritmo de Expectation-Maximization (EM), McLachlan y Peel, (2000) . Este resuelve distintos modelos en los que el volumen, forma y orientación de las distribuciones pueden considerarse iguales para todos los clusters o distintas para cada uno. Por ejemplo, un posible modelo es: volumen constante, forma variable, orientación variable. El paquete `mclust` emplea el método de máxima verosimilitud para ajustar todos estos modelos con distinto número, k , de clusters y selecciona el mejor en base al Bayesian Information Criterion (BIC). Se pueden elegir 14 modelos con diferentes características geométricas. Se asume una distribución gaussiana, por lo tanto, los clusters estarán centrados en el vector de medias μ_k y con diferentes características geométricas, como volumen, forma y orientación determinadas por la matriz de covarianzas Σ_k . En la tabla 2.1 se indican las características de los 14 posibles modelos teniendo en cuenta que $\Sigma_k = \lambda_k D_k A_k D_k^T$, siendo λ_k un escalar que controla el volumen del elipsoide, A_k una matriz diagonal y D_k una matriz ortogonal, Banfield y Raftery (1993), Celeux y Govaert (1995). En la figura 2.1 se muestran gráficamente las características geométricas.

Modelo	Σ_k	Distribución	Volumen	Forma	Orientación
EII	λI	Esférica	Igual	Igual	NA
VII	$\lambda_k I$	Esférica	Variable	Igual	NA
EEI	λA	Diagonal	Igual	Igual	Ejes de coordenadas
VEI	$\lambda_k A$	Diagonal	Variable	Igual	Ejes de coordenadas
EVI	λA_k	Diagonal	Igual	Variable	Ejes de coordenadas
VVI	$\lambda_k A_k$	Diagonal	Variable	Variable	Ejes de coordenadas
EEE	$\lambda D A D^T$	Elipsoidal	Igual	Igual	Igual
EVE	$\lambda D A_k D^T$	Elipsoidal	Igual	Variable	Igual
VEE	$\lambda_k D A D^T$	Elipsoidal	Variable	Igual	Igual
VVE	$\lambda_k D A_k D^T$	Elipsoidal	Variable	Variable	Igual
EEV	$\lambda D_k A D_k^T$	Elipsoidal	Igual	Igual	Variable
VEV	$\lambda_k D_k A D_k^T$	Elipsoidal	Variable	Igual	Variable
EVV	$\lambda D_k A_k D_k^T$	Elipsoidal	Igual	Variable	Variable
VVV	$\lambda_k D_k A_k D_k^T$	Elipsoidal	Variable	Variable	Variable

Cuadro 2.1: Tabla de características de los 14 posibles modelos de `mclust`.

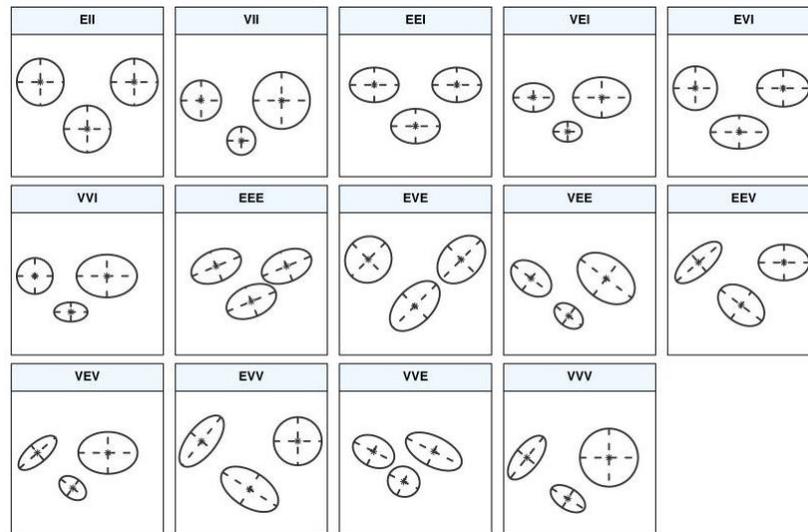


Figura 2.1: Gráfico con las características geométricas de los posibles modelos de mclust.

2.2.2. Agrupamiento jerárquico

El otro de los métodos utilizados es el agrupamiento jerárquico. Este es un método de análisis de grupos puntuales, el cual busca construir una jerarquía de grupos. Puede tener dos tipos de estrategias de agrupamiento, aglomerativas o divisivas. En este trabajo se utilizaron las aglomerativas, con la cual cada observación comienza en su propio grupo y a partir de ahí se van fusionando los grupos según un criterio de enlace formando grupos cada vez más grandes hasta unir todas las observaciones. Para representar cómo se van uniendo las observaciones se suele utilizar un dendrograma. Respecto al criterio de enlace también hay múltiples posibilidades. El más utilizado en este trabajo para el agrupamiento jerárquico fue el de agrupamiento por encadenamiento completo, también llamado agrupamiento del vecino más lejano porque con este método la distancia entre dos grupos se define como la mayor distancia posible entre los elementos de cada grupo, es decir, la distancia entre dos grupos es la distancia entre los elementos más distantes de cada grupo.

2.3. Predicción

Para la predicción se utilizó el análisis de regresión, para estimar las relaciones entre las variables del conjunto de datos. Con ello, se estudia la relación entre la variable dependiente **margen ordinario** y las variables independientes o predictoras.

2.3.1. Regresión lineal simple

La regresión lineal simple es el modelo matemático más sencillo para aproximar la relación de dependencia entre una variable dependiente Y , las variables independientes X_i y un término aleatorio ε , expuesto en Seber (1977). Este modelo puede ser expresado como:

$$Y_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

donde:

- Y_t : variable dependiente o respuesta.
- X_1, X_2, \dots, X_p : variables explicativas o independientes.
- $\beta_0, \beta_1, \beta_2, \dots, \beta_p$: parámetros, miden la influencia que las variables explicativas tienen sobre la variable dependiente.

y donde β_0 es el intercepto o término *constante*, las β_i ($i > 0$) son los parámetros respectivos a cada variable independiente, y p es el número de parámetros independientes a tener en cuenta en la regresión.

2.3.2. Modelo lineal generalizado (GLM)

El modelo lineal generalizado (GLM, por sus siglas en inglés) es una generalización flexible de la regresión lineal ordinaria que permite variables respuesta que tienen modelos de distribución de errores distintos de una distribución normal, como se comenta en McCullagh y Nelder (1989). El GLM generaliza la regresión lineal al permitir que el modelo lineal esté relacionado con la variable respuesta a través de una función de enlace y que la magnitud de la varianza de cada medición sea una función de su valor predicho.

Por lo tanto, mientras que en un modelo de regresión lineal múltiple tenemos que:

$$E[Y] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p,$$

es decir que existe una relación lineal entre las X y $E[Y]$, en los modelos lineales generalizados esa relación en lugar de ser entre la media de la variable respuesta y los predictores es entre una función de la media de la variable respuesta y los predictores:

$$g(E[Y]) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p,$$

en donde la función g dependerá de la distribución de probabilidad de la variable respuesta.

En el caso de este trabajo, para los resultados que se muestran en el capítulo 4, la distribución de probabilidad de nuestra variable respuesta es la normal $N(\mu, \sigma^2)$, la función link es la identidad $X^T \beta = E[Y]$ (aunque se hicieron pruebas con otros vínculos como el logaritmo).

2.3.3. Modelo aditivo generalizado (GAM)

Los modelos descritos en la anterior sección tienen la limitación de que utilizan un predictor lineal. Este predictor lineal presenta limitaciones ya que muchas variables no se relacionan de forma lineal entre ellas. Para solucionar esto, en Hastie y Tibshirani (1990), introducen los GAM, una extensión de los GLM que permite incorporar relaciones no lineales. En los modelos GAM, la relación de cada predictor X_i con la media de la variable respuesta $g(E[Y])$ no es directa, sino que se hace a través de una función $f_i(X_i)$ que tiene la siguiente forma:

$$g(E[Y]) = \beta_0 + f_1(X_1) + f_2(X_2) + \dots + f_p(X_p),$$

donde $f_j(X_j)$ es una función suavizadora de la variable X_j .

En este trabajo se utilizó el GAM de la familia gaussian con la función de enlace identidad y algoritmo de optimización de Newton para estimar los parámetros.

2.4. Clasificación

En este trabajo, se identifican segmentos en los que el margen ordinario se comporta de manera diferente, así que surge la necesidad de clasificar los clientes según el comportamiento del margen ordinario. Para ello, en primer lugar se utilizan árboles de clasificación y regresión y posteriormente random forest.

2.4.1. Árboles de clasificación y regresión (CART)

Los árboles de clasificación y regresión (CART, por sus siglas en inglés) son modelos que se basan en establecer una serie de reglas de decisión, cada una en función de valores umbrales para una de las variables, Breiman et al. (1984). De este modo se va fraccionando la muestra y finalmente se puede obtener una alta probabilidad de pertenencia a un determinado grupo.

En la figura 2.2 se muestra un ejemplo de un modelo CART, en el que se dispone de dos variables dependientes *Height* y *Weight*. Por otro lado, la variable respuesta es el género de una persona, y por lo tanto sus valores pueden ser *Male* y *Female*. En la imagen se puede observar cómo se van planteando un serie de condiciones lógicas para las variables explicativas hasta llegar a una predicción.

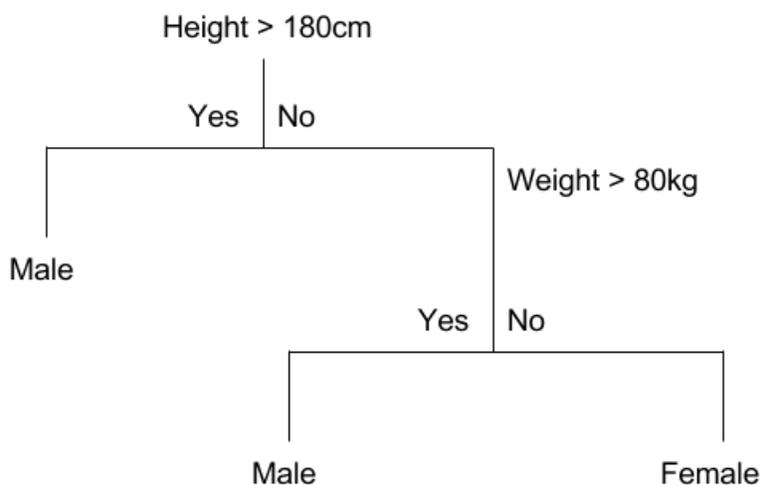


Figura 2.2: Ejemplo de un modelo CART.

Como se puede ver en la imagen, estas reglas van guiando hasta llegar a una predicción. Estas reglas se obtienen mediante el criterio de impureza de Gini. La impureza de Gini se puede calcular sumando la probabilidad de cada elemento siendo elegido multiplicado por la probabilidad de un error en la categorización de ese elemento. Alcanza su mínimo (cero) cuando todos los casos del nodo corresponden a una sola categoría de destino.

Para calcular la impureza de Gini de un conjunto de elementos con m categorías posibles, suponiendo que i toma valores en $\{1, 2, \dots, m\}$, y siendo f_i la fracción de artículos etiquetados con valor i en el conjunto, se tendría lo siguiente:

$$I_G = 1 - \sum_{i=1}^m f_i^2, \quad (2.2)$$

Por otro lado, el modelo CART utiliza varios parámetros ajustables para definir las condiciones lógicas anteriormente mencionadas. Estos son el minsplit, el minbucket y el Cp . El minsplit es el número mínimo de observaciones que deben existir en un nodo para que este exista. El minbucket es el número mínimo de observaciones de cada uno de los terminales del nodo. El Cp es valor mínimo que debe aumentar la complejidad de árbol a cada nodo.

random

2.4.2. Random forest

En los modelos CART se vio que a la hora de ramificar, este punto de corte se obtenía a partir de la impureza de Gini para todas las variables. En los random forest, explicados en Breiman (2001), se va a construir un árbol en el que el punto de corte para cada nodo es calculado sobre un número reducido de variables que son escogidas aleatoriamente, este número de variables se va a expresar como $mTry$.

Este proceso es repetido obteniendo un número $ntree$ de árboles, por lo que para obtener la predicción por random forest se realiza la media aritmética de todas las predicciones de los árboles que componen el random forest. En la figura 2.3 se muestra un ejemplo de una predicción de un random forest.

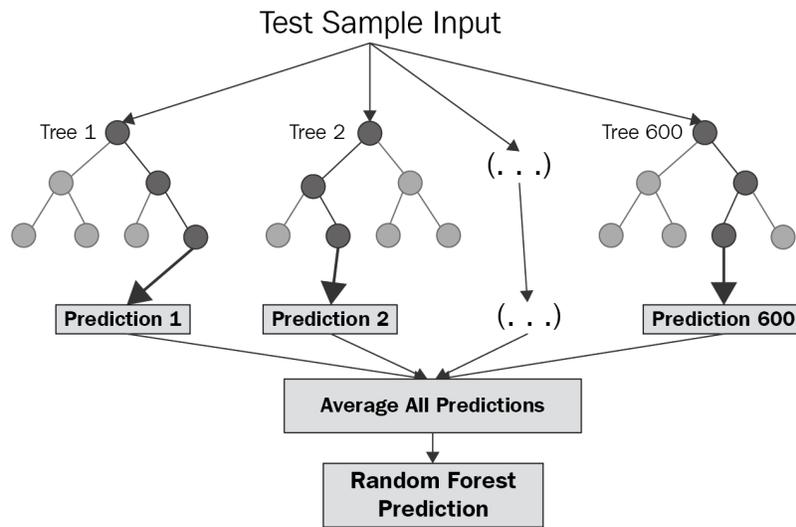


Figura 2.3: Esquema de funcionamiento de un modelo Random Forest.

Capítulo 3

Análisis exploratorio de los datos

Generalmente las empresas que manejan una gran cantidad de datos tienden a diseñar su base de datos reduciendo al mínimo la redundancia, puesto que los datos repetidos producen anomalías semánticas que dificultan tanto el procesamiento automático de los datos como el mantenimiento mismo de la base de datos. Por ello, es habitual guardar la información en tablas interconectadas por medio de claves, pero esto implica que la información esté repartida en numerosas tablas. Esto provoca que a veces sea complicado acceder a todos los datos que se necesitan en un ámbito concreto.

Para solucionar estas dificultades de acceso a los datos de un ámbito concreto se utilizan los data marts. Un data mart es una base de datos centrada en un ámbito que muchas veces es un segmento aislado de un almacén de datos de empresa. El subconjunto de datos contenido en un data mart suele alinearse con una unidad comercial concreta, como Ventas, Finanzas o Marketing. Los data marts aceleran los procesos comerciales al dar acceso a la información de forma más sencilla y rápida. Como un data mart tan solo contiene los datos aplicables a un ámbito comercial concreto, resulta una forma rentable de obtener información explotable rápidamente.

3.1. Data mart de clientes

En este caso, para poder realizar el estudio del Customer Lifetime Value se accede a los datos del data mart de clientes de la entidad. El acceso se realiza a los datos modificados mediante técnicas de seudonimización de datos personales, cumpliendo con la Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal (LOPD). Y de él, se obtienen los datos de clientes que son **personas físicas** que la entidad considera según su criterio que son clientes **con actividad**. Este almacén de datos se creó con el fin de entender cuáles son las dinámicas de la cartera de clientes, tanto las producidas por factores externos (coyuntura económica, competencia, etc.) como por factores internos (políticas de tipos, comisiones, servicios y atención, campañas de captación de clientes, etc.). El data mart se actualiza con periodicidad mensual y a continuación se incluyen dos tablas en las que se resume la información disponible. A continuación, en la tabla 3.1 se muestran las dimensiones de análisis, que son los factores por los que interesa agregar o desglosar los indicadores o métricas de clientes. Aunque pueden variar con el tiempo, normalmente son bastante estables o, al menos, presentan cierto grado de predictibilidad. Y en la tabla 3.2 se tienen las métricas a las que se hace referencia en algún punto de esta memoria. La tabla de métricas completa se añade en la tabla A.3 del apéndice A.8.

Tabla con las dimensiones de análisis

A continuación se muestran las dimensiones en las que se agrupan todas las métricas disponibles junto con una descripción de las mismas.

Dimensión	Descripción
Fecha	Año, Trimestre, Mes
Geografía Cliente	Situación geográfica de la residencia del cliente (Sección Censal, Población, Provincia, Comunidad Autónoma, País, Tipo País)
Carpetización	Distingue a los clientes carpetizados de los que no están asociados a una cartera de clientes de un gestor
Segmento Estratégico	Segmentos de clientes basados en su valor potencial para el banco
Edad	Edad tramificada de los clientes (-18, 18-25, ...)
Antigüedad	Tiempo, tramificado, que lleva el cliente en el banco (-3 meses, 4-6 meses, ...)
Sexo	Sexo del cliente
Geografía Oficina	Situación geográfica de la oficina a la que está asignado el cliente (Centro, ...)
Productos Financieros PT	Familias de productos financieros que tiene contratados el cliente como primer titular a nivel ahorro, inversión, financiación y previsión
Productos Financieros CT	Familias de productos financieros que tiene contratados el cliente con cualquier titularidad a nivel ahorro, inversión, financiación y previsión
Servicios	Tipo de servicios que contratan y usan los clientes (seguros, tarjetas, domiciliaciones, rentas)
Tipo Cliente	Clase de cliente (nacional, extranjero, no residente)
Continúa en la siguiente página	

Cuadro 3.1 – continuado de la página anterior

Dimensión	Descripción
Tipo de Hogar	Clase de hogar (con hijos, parejas jóvenes, ...)
Situación Cliente	Estado de actividad en la que se encuentra el cliente, en función de sus saldos y nº movimientos
Situación Contrato	Relación contractual que mantiene el cliente con el banco
Canal	Canales de contacto con los clientes
Campañas	Campañas dirigidas a clientes
Promociones	Promociones diseñadas para clientes

Cuadro 3.1: Lista de dimensiones de análisis en el data mart de clientes.

Tabla reducida de métricas del data mart

A continuación se muestra una tabla reducida de las métricas disponibles en el data mart. En esta tabla reducida 3.2 se indican solamente las que se mencionan en algún punto de esta memoria, pero se añade en el apéndice A.8 la tabla completa de métricas para reflejar la cantidad de información disponible para el estudio del CLV.

Campo	Descripción
CLIENTE.ID	Año, Trimestre, Mes
FECHA	Fecha de obtención de los datos
IMP_RECIBOS_AGUA_4M_CT	Importe medio mensual en los últimos 4 meses correspondiente a los recibos del agua. Se calcula la media obteniendo la suma de los recibos del agua de los últimos 4 meses y dividiendo por los meses en los que hay dato.
IMP_RECIBOS_TELEFONO_4M_CT	Importe medio mensual en los últimos 4 meses correspondiente a los recibos del teléfono. Se calcula la media obteniendo la suma de los recibos del teléfono de los últimos 4 meses y dividiendo por los meses en los que hay dato.
IMP_RECIBOS_ENERGIA_4M_CT	Importe medio mensual en los últimos 4 meses correspondiente a los recibos de energía. Se calcula la media obteniendo la suma de los recibos de energía de los últimos 4 meses y dividiendo por los meses en los que hay dato.
IMP_RECIBOS_COMUNIDAD_4M_CT	Importe medio mensual en los últimos 4 meses correspondiente a los recibos de la comunidad. Se calcula la media obteniendo la suma de los recibos de la comunidad de los últimos 4 meses y dividiendo por los meses en los
Continúa en la siguiente página	

Cuadro 3.2 – continuado de la página anterior

Campo	Descripción
	que hay dato.
IMP_RECIBOS_SM.4M.CT	Importe medio mensual en los últimos 4 meses correspondiente a los recibos del seguro médico. Se calcula la media obteniendo la suma de los recibos del seguro médico de los últimos 4 meses y dividiendo por los meses en los que hay dato.
IMP_RECIBOS_IMPUESTOS.4M.CT	Importe medio mensual en los últimos 4 meses correspondiente a los recibos de impuestos. Se calcula la media obteniendo la suma de los recibos de impuestos de los últimos 4 meses y dividiendo por los meses en los que hay dato.
IMP_RECIBOS_RESTO.4M.CT	Importe medio mensual en los últimos 4 meses correspondiente a los recibos del resto. Se calcula la media obteniendo la suma de los recibos del resto de los últimos 4 meses y dividiendo por los meses en los que hay dato.
OP_CAJ_ABONO_TJDB.3M	Calcula la media de operaciones de los abonos realizados en cajeros a las tarjetas de débito de los últimos 3 meses contando solamente los meses que tienen operaciones en tarjetas
IMP_PENSION_SUA	Calcula la media de las pensiones de los últimos 5 meses, contando solamente los meses que tienen ingresos
Continúa en la siguiente página	

Cuadro 3.2 – continuado de la página anterior

Campo	Descripción
VINCULACION_TRANSACCIONAL_CT	Vinculación transaccional como cualquier titular
VINCULACION_NEGOCIO_PT	Vinculación de negocio como primer titular
VINCULACION_NEGOCIO_CT	Vinculación de negocio como cualquier titular
NUM_COMISIONES	Número de comisiones del cliente
IMP_COMISIONES	Importe de las comisiones del cliente
SALDO_FINAL_VISTA_CT	Saldo final de los productos a la vista como cualquier titular del cliente
SALDO_MEDIO_VISTA_CT	Saldo medio de los productos a la vista como cualquier titular del cliente
OP_HABER_VISTA_CT	Número de operaciones al haber de productos a la vista como cualquier titular
IMP_HABER_VISTA_CT	Importe de operaciones al haber de productos a la vista como cualquier titular
NUM_RECIBOS_SM_CT	Número de recibos de seguro médico domiciliados como cualquier titular
IMP_RECIBOS_SM_CT	Importe de recibos de seguro médico domiciliados como cualquier titular
IMP_PENSION	Importe de ingresos por pensiones
IMP_CAJ_ABONO_TJDB	Importe de operaciones de abono en cajeros para tarjetas de débito
SALDO_TARJ_CREDITO_PT	Saldo de las tarjetas de crédito como primer titular
SALDO_MEDIO_IPFS_CORTO_CT	Saldo medio de los depósitos a corto plazo del cliente
Continúa en la siguiente página	

Cuadro 3.2 – continuado de la página anterior

Campo	Descripción
	como cualquier titular
SALDO_FINAL_IPFS_CORTO_CT	Saldo final de los depósitos a corto plazo del cliente como cualquier titular
SALDO_PDTE_PTMOS_GP_VIV_PT	Saldo pendiente de los préstamos de garantía personal de viviendas del cliente como primer titular
SALDO_PDTE_PTMOS_GR_VIV_CT	Saldo pendiente de los préstamos de garantía real de viviendas del cliente como cualquier titular
CUOTA_PTMOS_GR_VIV_CT	Cuota de los préstamos de garantía real de viviendas del cliente como cualquier titular
TIPOMEDIO_PTMOS_GR_VIV_CT	Tipo medio de los préstamos de garantía real de viviendas del cliente como cualquier titular
RENTA_ESTIMADA	Renta estimada del cliente en función del segmento al que pertenece
MO_CLIENTE_PT	Margen ordinario del cliente en sus contratos como primer titular
MO_ACTIVO_PT	Margen ordinario del cliente en sus contratos de activo como primer titular
MO_PASIVO_PT	Margen ordinario del cliente en sus contratos de pasivo como primer titular
MO_FB_PT	Margen ordinario del cliente en sus contratos de fuera de balance como primer titular
MO_SEGUROS_PT	Margen ordinario del cliente en sus contratos de seguros
Continúa en la siguiente página	

Cuadro 3.2 – continuado de la página anterior

Campo	Descripción
	como primer titular
MO_TARJETAS_PT	Margen ordinario del cliente en sus contratos de tarjetas como primer titular
MO_CLIENTE_CT	Margen ordinario del cliente en sus contratos como cualquier titular
MO_ACTIVO_CT	Margen ordinario del cliente en sus contratos de activo como cualquier titular
MO_PASIVO_CT	Margen ordinario del cliente en sus contratos de pasivo como cualquier titular
MO_FB_CT	Margen ordinario del cliente en sus contratos de fuera de balance como cualquier titular
MO_SEGUROS_CT	Margen ordinario del cliente en sus contratos de seguros como cualquier titular
MO_TARJETAS_CT	Margen ordinario del cliente en sus contratos de tarjetas como cualquier titular
RF_CLIENTE_PT	Rentabilidad final del cliente como primer titular
MI_CLIENTE_PT	Margen de intermediación del cliente en sus contratos como primer titular
RF_CLIENTE_CT	Rentabilidad final del cliente como cualquier titular
MI_CLIENTE_CT	Margen de intermediación del cliente en sus contratos como cualquier titular
SALDO_MEDIO_ACTIVADO_PT	Saldo medio de los contratos de activo como primer
Continúa en la siguiente página	

Cuadro 3.2 – continuado de la página anterior

Campo	Descripción
	titular
SALDO_MEDIO_PASIVO_PT	Saldo medio de los contratos de pasivo como primer titular
SALDO_MEDIO_TARJETAS_PT	Saldo medio de los contratos de tarjetas como primer titular
SALDO_MEDIO_CLIENTE_CT	Saldo medio de todos los contratos del cliente como cualquier titular
SALDO_MEDIO_ACTIVIVO_CT	Saldo medio de los contratos de activo como cualquier titular
SALDO_MEDIO_PASIVO_CT	Saldo medio de los contratos de pasivo como cualquier titular
SALDO_MEDIO_FB_CT	Saldo medio de los contratos de fuera de balance como cualquier titular
SALDO_MEDIO_SEGUROS_CT	Saldo medio de los contratos de seguros como cualquier titular
SALDO_MEDIO_TARJETAS_CT	Saldo medio de los contratos de tarjetas como cualquier titular

Cuadro 3.2: Tabla reducida de campos disponible en el data mart de clientes.

3.2. Margen ordinario

Como ya se comentó en capítulos anteriores, para este trabajo se utiliza el valor del margen ordinario para el cálculo del CLV. Como se puede comprobar en la figura 3.1, los datos de margen ordinario del data mart de clientes (SUMA_MO_PT) son muy parecidos al de la entidad (SUMA_MO_ABANCA). Las diferencias pueden venir por múltiples motivos, pero en general se puede comprobar que el comportamiento es muy parecido (entre SUMA_MO_PT y SUMA_MO_ABANCA) y también se comprobó que los clientes que están informados en el data mart tienen en su mayoría el MO coincidente con el del

data warehouse. En la figura 3.1, además de los dos márgenes ya comentados, se muestran también los márgenes que conforman SUMA_MO_PT. En SUMA_MO_PT se tiene el margen ordinario del cliente en sus contratos como primer titular, pero este campo es el resultado de la suma del margen ordinario de distintos tipos de contratos. Los distintos tipos de contratos que se tienen en el data mart de clientes son: contratos de activo (SUMA_MO_ACTIVO), contratos de pasivo (SUMA_MO_PASIVO), contratos de fuera de balance (SUMA_MO_FB), contratos de seguros (SUMA_MO_SEGUROS) y contratos de tarjetas (SUMA_MO_TARJETAS). Evidentemente, con la suma del margen ordinario de los distintos tipos de contratos de un cliente ya se obtendría el margen ordinario de ese cliente. Por eso, en el capítulo de resultados, cuando se obtienen las correlaciones de las variables del data mart respecto al margen ordinario (ver tabla 4.2), se indica que algunas de las variables no son utilizables para realizar la predicción. Respecto a la figura 3.1, también habría que comentar que los valores mostrados en el eje de ordenadas no se corresponden con la realidad ya que, por confidencialidad, los valores reales se han multiplicado por un factor. Por lo tanto, en el siguiente gráfico se puede comprobar que en el data mart de clientes, al sumar la variable MO_CLIENTE_PT (SUMA_MO_PT) se obtiene mensualmente una cifra muy parecida al margen ordinario registrado en la cuenta de resultados de la entidad, cuya suma se representa con SUMA_MO_ABANCA. Además, la suma de las variables que conforman MO_CLIENTE_PT, que son MO_ACTIVO_PT, MO_PASIVO_PT, MO_FB_PT, MO_SEGUROS_PT Y MO_TARJETAS_PT se representan con SUMA_MO_ACTIVO, SUMA_MO_PASIVO, SUMA_MO_FB, SUMA_MO_SEGUROS y SUMA_MO_TARJETAS respectivamente.

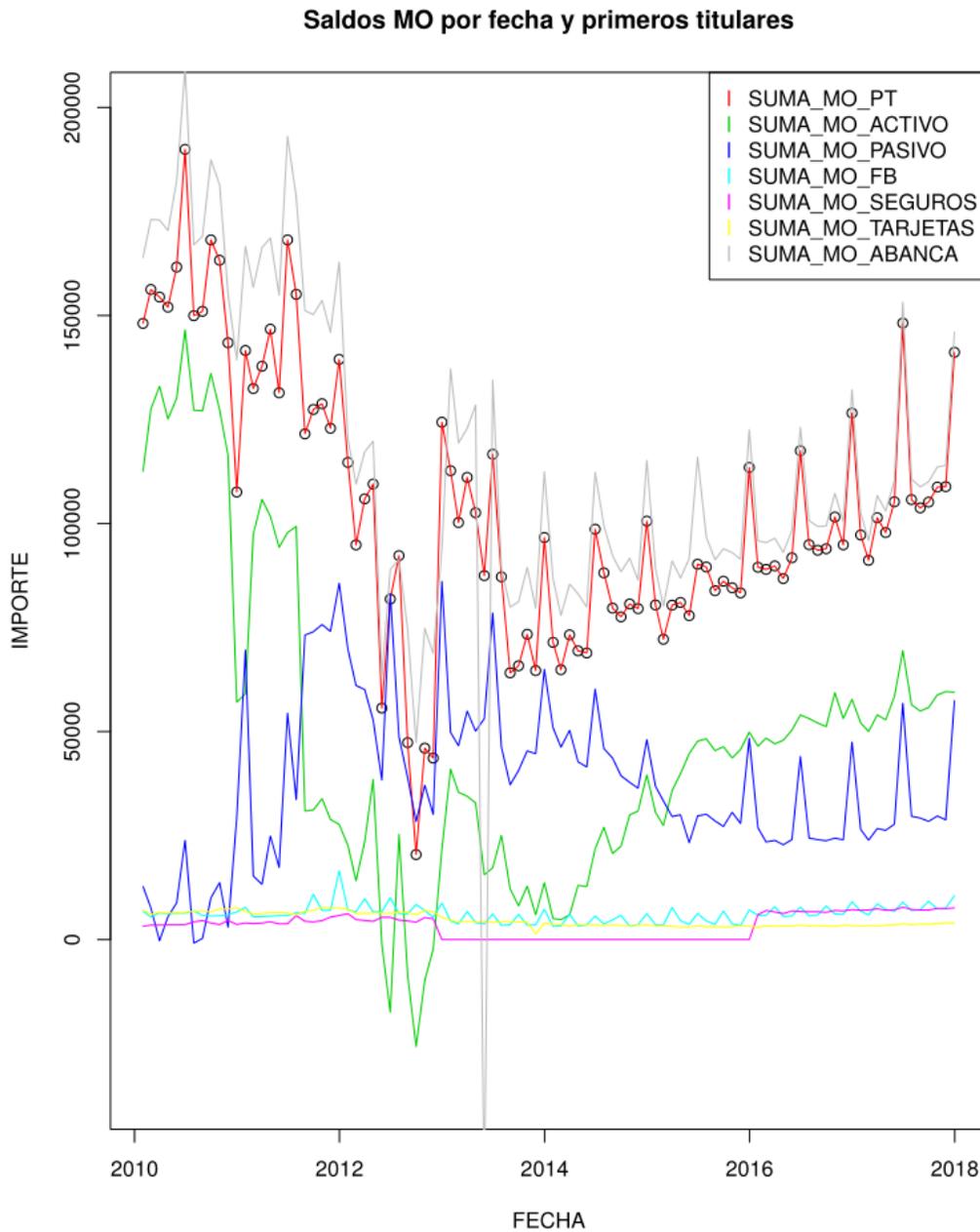


Figura 3.1: Gráfico con la evolución de margen ordinario en el data mart de clientes

3.3. Transformación de las variables

Si se cuentan los puntos de la figura 3.1, se verá que cada dos años se dispone de 24 observaciones y esto se debe a que el data mart de clientes se actualiza mensualmente. Pero en este trabajo se utilizaron datos anuales. Se anualizaron los datos para evitar cambios bruscos del margen ordinario derivados únicamente del mes en el que se realiza el cálculo del margen.

En las dos siguientes figuras (3.2 y 3.3), se tiene la aportación de cada *aplicación*¹ al margen ordinario del banco en dos meses distintos. En ellas, los valores del eje de ordenadas no se corresponden con la realidad ya que se han multiplicado por un factor por confidencialidad.

Como se puede ver, la mayoría de las variables tienen una situación parecida (respecto a las demás variables) en ambas figuras. Pero observando la variable RAJE, se verá que en diciembre pasa a ser claramente la que más margen ordinario aporta a la entidad. Esto se debe a la política de comisiones del banco. En diciembre se produce un cobro semestral de comisiones de las cuentas a la vista y esto provoca un gran cambio en el total del margen ordinario. Como la mayoría de las comisiones liquidan en periodos inferiores al año se decidió anualizar las variables. De esta forma, los cambios que se detecten entre periodos serán derivados de cambios financieros (altas o bajas de contratos, variación del saldo en los contratos, etc.) y no derivados del periodo que se está observado.

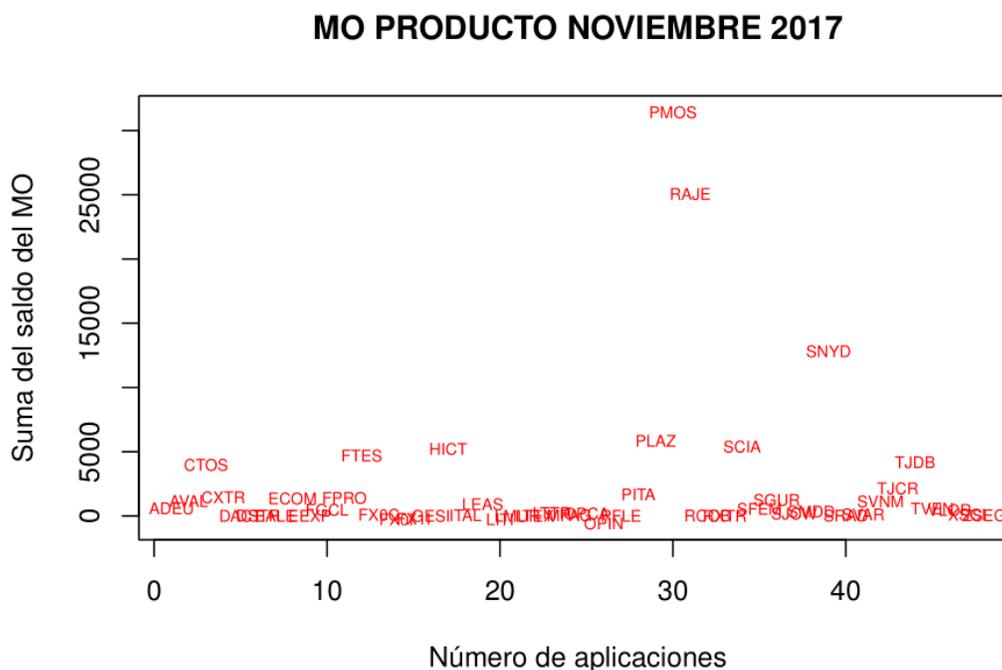


Figura 3.2: Gráfico del impacto de las distintas aplicaciones sobre el MO (NOV '17)

¹La entidad genera los datos de los contratos en distintas **aplicaciones** en función del tipo de producto. Se pueden consultar las distintas aplicaciones que influyen en el margen ordinario en la tabla A.4 del apéndice A.10. Por ejemplo, los datos de préstamos se generan en la aplicación *PMOS*, los datos de cuentas a la vista (también llamados recursos ajenos desde la perspectiva de la entidad) se generan en la aplicación *RAJE*, etc.

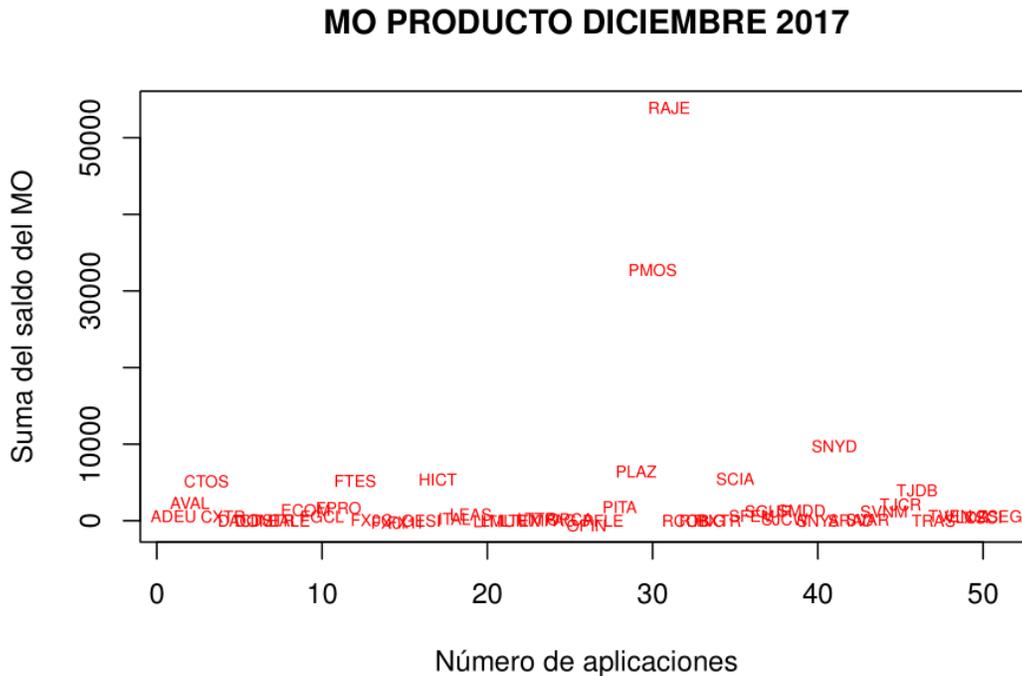


Figura 3.3: Gráfico del impacto de las distintas aplicaciones sobre el MO (DIC '17)

3.4. Variables finales

Como se puede comprobar en el capítulo de Resultados, en los últimos modelos realizados se llega a explicar más de un 85% de la variabilidad de los datos con un modelo GAM de 8 variables. A continuación se incluye un histograma y la función de densidad del margen ordinario de cada una de esas 8 variables para que se pueda observar cómo se comportan.

Como se puede comprobar en la figura A.68 del apéndice A.9, la variable MO tiene clientes con valores en torno a 60.000€ pero debido a que la gran mayoría de los clientes tienen un margen ordinario mucho menor, en la figura A.68 no se aprecia muy bien el comportamiento real de la variable. Por esta razón, a continuación se muestra el histograma respecto a un intervalo de valores de la variable MO en el que se aprecia mejor el comportamiento de la variable. En la figura 3.4 vemos que existen clientes con margen ordinario negativo pero que la mayor densidad se observa entre los 0€ y los 200€ y a partir de ahí la densidad va bajando según va aumentando el margen ordinario.

Hay que recordar que este margen es el obtenido para cada cliente teniendo en cuenta todos los contratos de los que es titular (pero no tiene porqué ser el primer titular). Por ejemplo, se tiene un matrimonio en el que ambos cónyuges tienen una hipoteca, una tarjeta de crédito y una cuenta a la vista en los que el cónyuge 1 es primer titular y el cónyuge 2 es cotitular y además tienen una cuenta a la vista individual para cada uno. El margen ordinario del cónyuge 1 sería el obtenido de los contratos que tienen en común más el margen ordinario de su cuenta a la vista y el margen ordinario del cónyuge 2 sería el obtenido de los contratos que tienen en común más el margen ordinario de la cuenta a la vista individual del cónyuge 2. Esto se hace así porque en un caso como el anterior, en el que el cónyuge 1 es primer titular de todos los contratos comunes del matrimonio, el margen ordinario del cónyuge 1 sería mucho mayor si se consideran solo los contratos de los que cada cliente es primer titular. Porque

podría ocurrir que los dos cónyuges aportasen el mismo saldo mensualmente a la cuenta común para pagar todos los intereses de los productos de activo que tenían contratados y por lo tanto no tendría sentido una desigualdad tan grande.

Esta situación provoca que el mismo margen de un contrato, se esté sumando en más de un cliente y por lo tanto, el margen que se ve en la siguiente gráfica no se corresponde con el margen de la cuenta de resultados de la entidad ya que en esta, el margen ordinario de cada contrato se asocia a su primer titular.

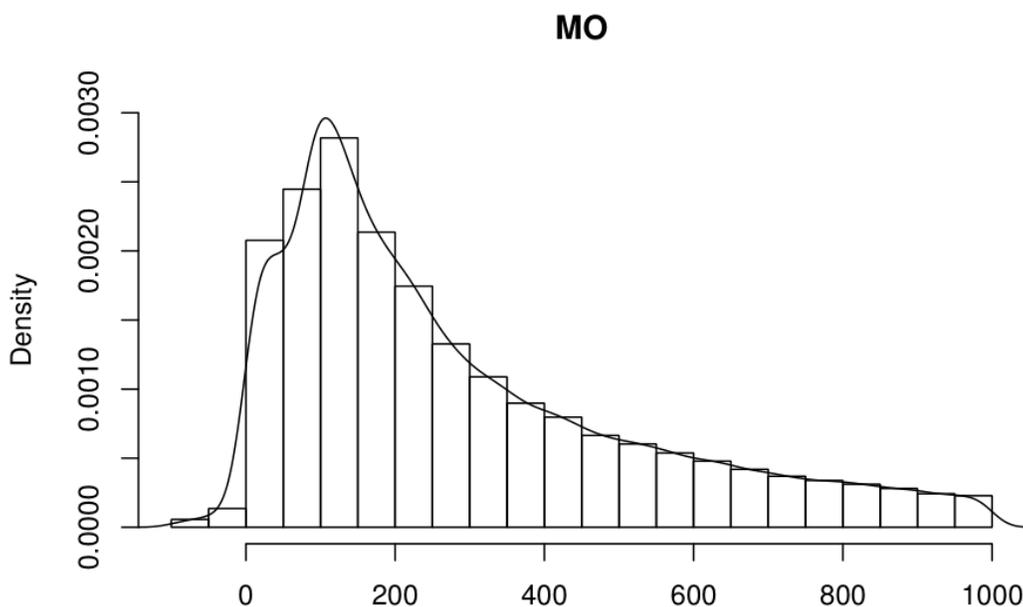


Figura 3.4: Histograma y densidad de la variable MO condicionada a que dicha variable se encuentra en el intervalo $(-100,1.000)$.

En la siguiente figura tenemos el importe en recibos de impuestos de los clientes de la muestra. Como se puede ver en la figura A.69 del apéndice A.9, hay clientes que llegan a tener valores en torno a 140.000€ pero a continuación se muestra la figura 3.5 en donde se aprecia mejor el comportamiento de la variable. Vemos que la mayor densidad se produce en el 0, lo que indica que muchos clientes no tienen domiciliado ningún recibo de impuestos pero se ve que según aumenta el importe en los recibos también descende la densidad de clientes.

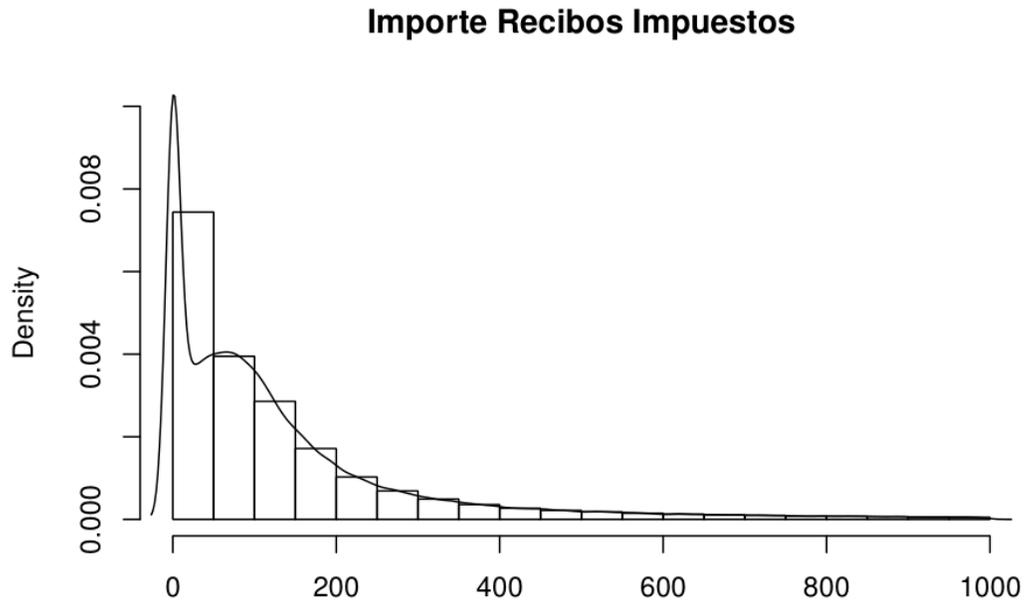


Figura 3.5: Histograma y densidad de la variable del importe de recibos de impuestos condicionada a que dicha variable se encuentra en el intervalo $[0,1.000)$.

En cuanto al saldo medio, también se ven valores muy superiores a los más frecuentes. Vemos en la figura A.70 que existen clientes con un saldo medio en torno a 20.000.000€ pero vemos en la figura 3.6 que los clientes con saldo medio entre 0€ y 5.000€ son más del doble que los clientes con saldo medio entre 5.000€ y 10.000€ y la densidad sigue bajando según aumenta el saldo medio.

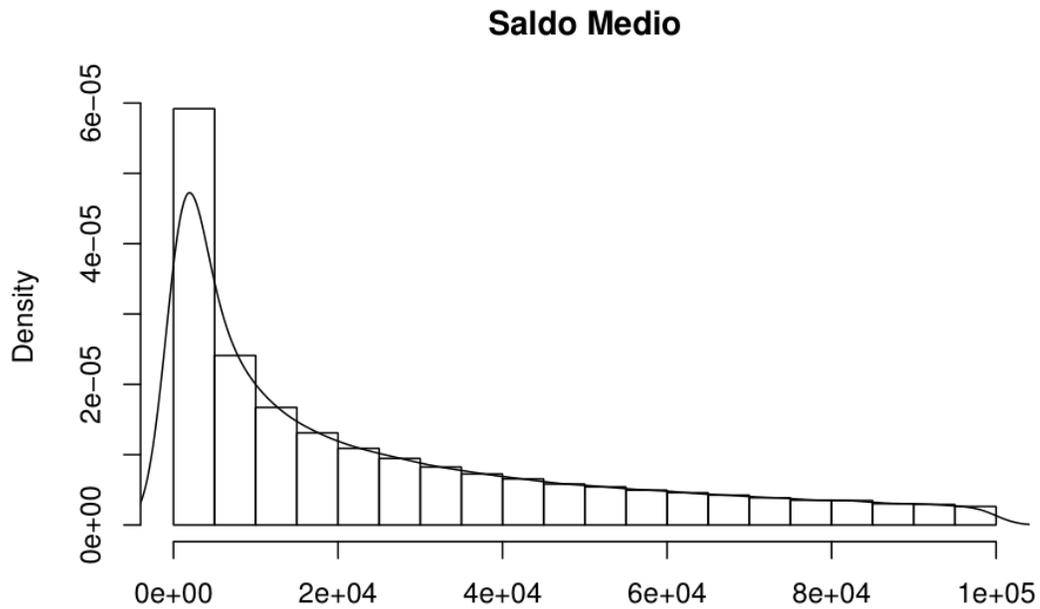


Figura 3.6: Histograma y densidad de la variable saldo medio condicionada a que dicha variable se encuentra en el intervalo $[0,100.000)$.

En la variable con la vinculación de negocio no es necesario un segundo histograma para ver hasta dónde llegan los valores extremos. En la figura 3.7 se puede ver que aunque la vinculación llega a valores cercanos a 7 las mayores frecuencias se encuentran en valores entre 0 y 1 y que a partir de valores superiores a 4,5 apenas se observan frecuencias significativas respecto a los valores más pequeños. También se puede ver que la variable toma ciertos valores concretos con mucha frecuencia. Vemos pequeñas modas debido a que la variable tiene valores exactamente iguales para muchos clientes. Esto se debe a que la vinculación es un cálculo que hace la entidad en función de unos criterios que se repiten entre muchos clientes. Por lo tanto, se trata de una variable con una componente discreta debido a que hay ciertos valores que se toman con probabilidad estrictamente mayor que 0.

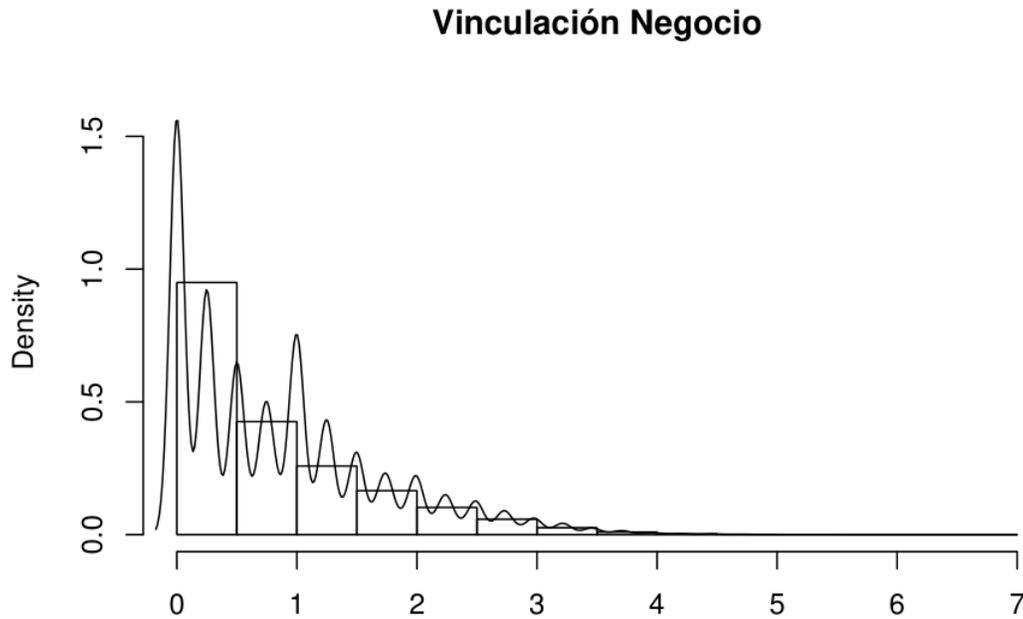


Figura 3.7: Histograma de la variable vinculacion negocio con una muestra de 200.000 clientes.

En la variable con el importe de comisiones ha sido necesario dejar de tener en cuenta los valores iguales a 0 en la generación del histograma porque tenían el valor 0, 129.129 clientes de los 200.000 que tenemos en la muestra y esta situación provocaba que no se apreciara el comportamiento de la variable en el histograma. Por ello, aunque la figura A.71 se hizo con todos los datos, a continuación se muestra la figura 3.8 en donde se puede comprobar que de media, para el año 2019, la variable con el importe de comisiones tomó más frecuentemente valores entre 0€ y 2€. Entre 2 y 6 baja mucho la densidad, pero entre 6 y 8 vuelve a subir la densidad hasta valores similares a los del intervalo (0,2]. A partir de los 8€ vuelve a bajar mucho la densidad y continúa bajando según aumenta el importe de las comisiones.

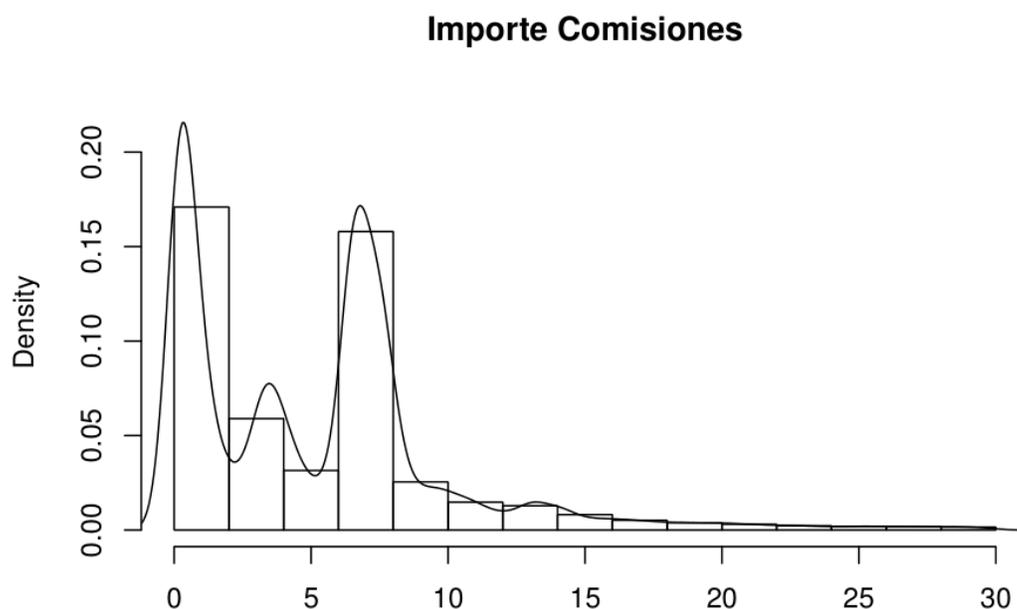


Figura 3.8: Histograma y densidad de la variable importe de comisiones condicionada a que dicha variable se encuentra en el intervalo $(0,30)$.

Con la variable saldo activo también ha sido necesario no tener en cuenta los valores 0 para la creación del histograma definitivo de la variable. En el histograma de la figura A.72 se ve que la mayor parte de los clientes tienen un saldo en productos de activo muy inferior al máximo de la muestra. Vemos que existen clientes con un importe en torno a 4.000.000€ pero la mayoría tienen importes inferiores a 400.000€. Además, si nos centramos en los importes más habituales (figura 3.9) se ve que lo más habitual son importes inferiores a 50€. De 50€ a 100€ se produce un cambio muy grande en cuanto a la frecuencia de casos y a partir de aquí va descendiendo poco a poco según va aumentando el importe de saldo activo. En todo esto, hay que tener en cuenta lo comentado anteriormente, ya que al mostrar la variable saldo activo se han dejado fuera los valores 0 porque de no ser así la densidad en el intervalo $(0,50]$ sería aún mucho mayor.

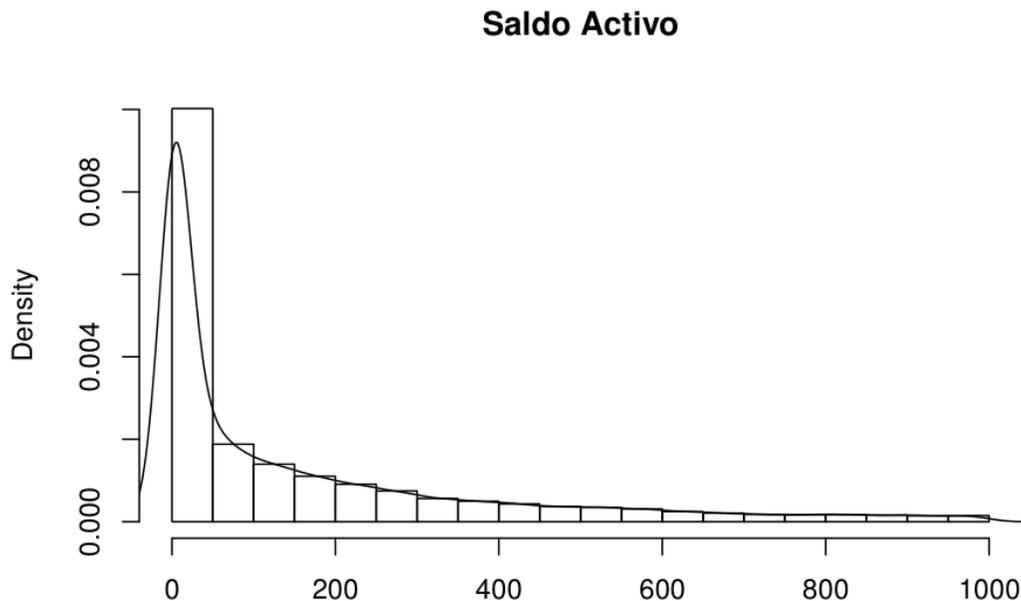


Figura 3.9: Histograma y densidad de la variable saldo activo condicionada a que dicha variable se encuentra en el intervalo $(0, 1.000)$.

En la figura 3.10 se puede ver que los importes más frecuentes en productos de pasivo son los importes entre -49.99€ y 50€ . La razón por la que se incluyen los valores negativos es que al hacer el histograma se incluyó al 0 en el intervalo con los valores negativos, pero aún así los importes entre 0.01 y 50€ tienen una densidad similar. Esto se produce porque el valor 0 no es tan frecuente como en activo. Y como veremos a continuación, también en productos de fuera de balance o tarjetas es muy habitual el valor 0. Igual que con las demás variables, se añade en el apéndice A.9 el histograma con todos los valores de la variable saldo pasivo (A.73)

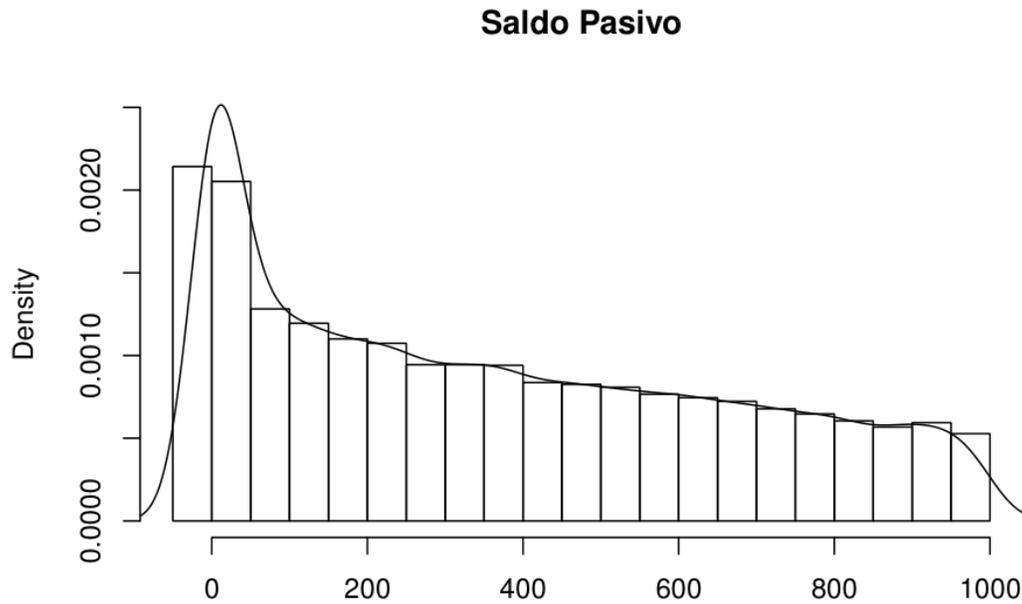


Figura 3.10: Histograma y densidad de la variable saldo pasivo condicionada a que dicha variable se encuentra en el intervalo $[-500, 1.000)$.

En la variable saldo fuera de balance también ha sido necesario no tener en cuenta los valores 0 para obtener un histograma que permitiese ver el comportamiento de la variable. En la figura A.74 se puede ver una densidad muy alta en torno al 0 y es que de los 200.000 clientes de la muestra, más de 160.000 tienen el valor 0 en esta variable. Por ello, al realizar el histograma en el intervalo de valores que se toman con más probabilidad (figura 3.11) vemos una densidad muy alta en los clientes con saldo fuera de balance menor a 5.000€, pero en esta tipología de productos también tienen alta densidad importes bastante mayores, aunque como en la mayoría de las variables, la densidad disminuye según aumenta el importe de la variable.

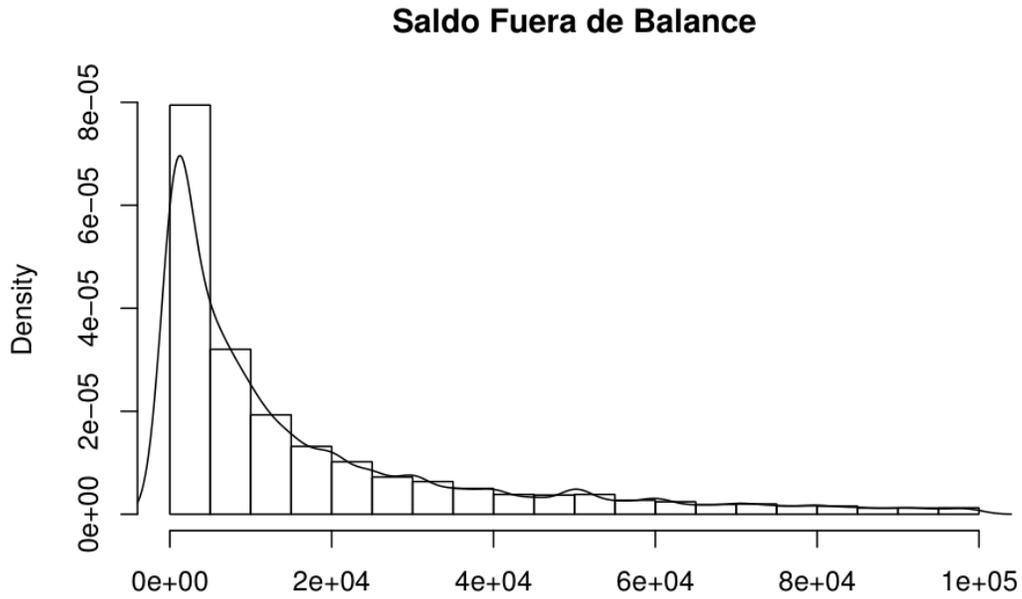


Figura 3.11: Histograma y densidad de la variable saldo fuera de balance condicionada a que dicha variable se encuentra en el intervalo (0, 100.000).

En esta variable tampoco se ha tenido en cuenta el valor 0 porque de los 200.000 clientes de la muestra, casi 150.000 tienen ese valor en la variable. Al no tener en cuenta este valor, en la figura 3.12 se ve que los valores entre 0€ y 100€ son los más frecuentes pero que desde los 200€ en adelante la frecuencia va disminuyendo lentamente según aumenta el saldo de tarjetas. En la figura A.75 se ve que los importes positivos de tarjetas no llegan a ser tan grandes como los de otras tipologías de productos (como en activo o en pasivo por ejemplo) pero los importes con signo negativo sí que son los más grandes.

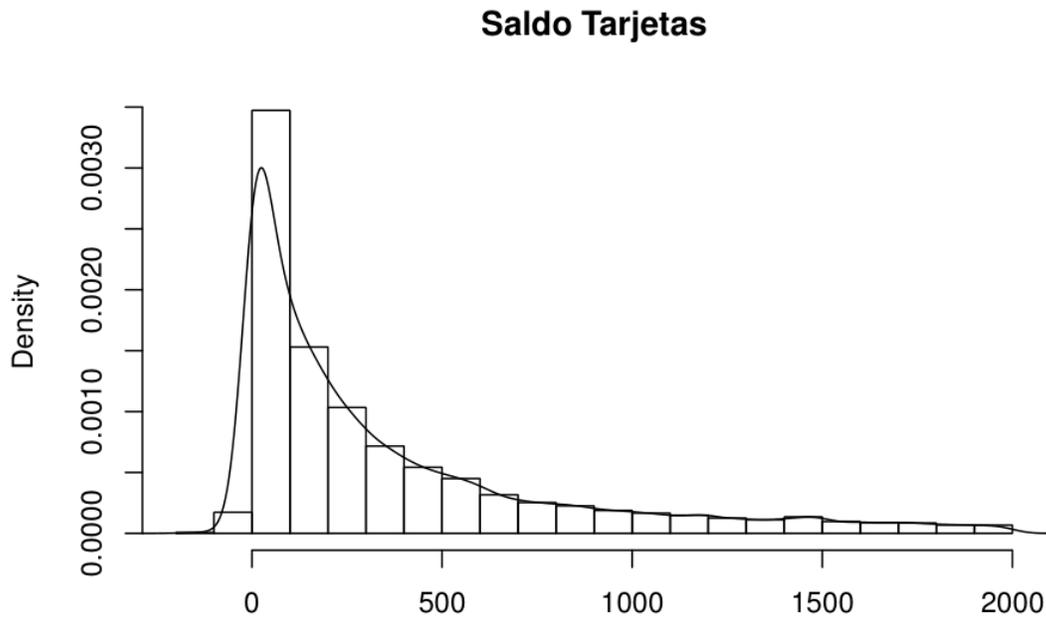


Figura 3.12: Histograma y densidad de la variable saldo de tarjetas condicionada a que dicha variable se encuentra en el intervalo $[-200, 2.000)$.

Capítulo 4

Resultados

Para poder explicar los resultados obtenidos en los principales objetivos de este trabajo, que son la clasificación y la predicción del CLV, es necesario comentar primeramente qué resultados se obtienen en el análisis de componentes principales y en el cálculo de los coeficientes de correlación entre las variables originales y nuestra variable objetivo MO_CLIENTE_CT.

Como ya se comentó en el capítulo 3, el conjunto de datos disponible está formado por una gran cantidad de variables y por lo tanto la reducción de la dimensionalidad podría ser interesante para tener la mayor cantidad de información posible en un conjunto más reducido de variables. Por otra parte, en caso de que esta reducción del número de variables no sea satisfactoria, es necesario calcular una medida de la dependencia lineal de la variable objetivo respecto a las demás variables para poder centrar un poco los esfuerzos a la hora de clasificar y predecir.

Por todo esto, en este capítulo, las dos primeras secciones se dedican a presentar los resultados del análisis de componentes principales y los coeficientes de correlación entre las variables originales y la variable MO_CLIENTE_CT, las dos siguientes secciones explican los resultados obtenidos en la clasificación y la predicción y en la última sección de este capítulo se indican los resultados obtenidos en una combinación de predicción y clasificación que se ha considerado conveniente realizar tras los resultados obtenidos en la sección 4.4.

4.1. Análisis de componentes principales

Primero realizamos un ACP para intentar reducir las más de 200 variables disponibles a unas pocas componentes. Para ello utilizamos la función `prcomp` y se calcula el análisis de componentes principales (ACP) con las correlaciones.

El conjunto de datos consta de 367 variables de las cuales 235 eran variables numéricas. Tras realizar un ACP como se podrá comprobar en la figura 4.1 (la tabla al completo está en el apéndice A.1), vemos que no se puede realizar una reducción a un número de variables pequeño si queremos explicar una cantidad razonable de la variabilidad del conjunto de datos. Como se puede ver en la tabla 4.1, por ejemplo con las primeras 25 componentes principales explicaríamos un 50 % de la variabilidad, con 72 componentes explicaríamos un 80 % y tendríamos que utilizar las primeras 110 componentes principales para explicar un 90 % de la variabilidad de las variables numéricas de nuestro conjunto de datos.

Pasar de 235 a 110 variables es una gran reducción pero a la hora de realizar la segmentación y la predicción siguen siendo demasiadas variables para poder explicar los resultados de los modelos.

Componente	Desviación típica	Proporción de variación explicada	Proporción acumulada
PC1	3,74361	0,05964	0,05964
PC2	2,81076	0,03362	0,09326
PC3	2,68842	0,03076	0,12401
PC4	2,42986	0,02512	0,14914
PC5	2,35042	0,02351	0,17264
PC6	2,20267	0,02065	0,19329
PC7	2,17188	0,02007	0,21336
PC8	2,1295	0,0193	0,23270
PC9	2,09474	0,01867	0,25133
PC10	2,0846	0,01849	0,26982
(...)	(...)	(...)	(...)
PC24	1,7615	0,0132	0,49480
PC25	1,74217	0,01292	0,50775
PC26	1,61451	0,01109	0,51884
(...)	(...)	(...)	(...)
PC71	0,93409	0,00371	0,79694
PC72	0,91627	0,00357	0,80052
PC73	0,9076	0,0035	0,80400
(...)	(...)	(...)	(...)
PC109	0,70476	0,00211	0,89912
PC110	0,69859	0,00208	0,90120
PC111	0,69436	0,00205	0,90325
(...)	(...)	(...)	(...)

Continúa en la siguiente página

Cuadro 4.1 – continuado de la página anterior

Componente	Desviación típica	Proporción de variación explicada	Proporción acumulada
PC233	3,5E-16	0	1
PC234	2,826E-16	0	1
PC235	2,518E-16	0	1

Cuadro 4.1: Resultado ACP reducido (Completo en apéndice).

4.2. Coeficientes de correlación con MO_CLIENCTE_CT

Tras hacer unas pruebas y ver que la ejecución de los siguientes modelos era razonablemente rápida y que por lo tanto el tiempo de CPU no iba a ser un problema, decidimos realizar un estudio de la correlación que tenían todas las variables con nuestra variable objetivo, MO_CLIENCTE_CT. La intención fue utilizar en los modelos directamente las variables de nuestro conjunto de datos y no las componentes principales calculadas, ya que para explicar un porcentaje alto de la variabilidad era necesario una gran cantidad de componentes principales.

En un primer intento, se obtuvieron las correlaciones con las variables sin transformar, pero al ver las gráficas de todas las variables respecto a nuestra variable objetivo MO_CLIENCTE_CT, se vio necesario realizar una transformación a las variables tanto para obtener gráficas como para el cálculo de las correlaciones. Dada la variable $x = MO_CLIENCTE_CT$, para el cálculo de las correlaciones se considera una nueva variable $x' = x/(1 + |x|)$. De esta forma, x' ya no tiene las colas tan pesadas a la derecha y a la izquierda. En cuanto a las demás variables, vamos a considerar una variable y genérica que representa a cada una de las demás variables, y se considera la nueva variable $y' = y/(1 + |y|)$. Al final, se aplicó la misma transformación a todas las variables cuya gráfica se puede observar en la figura 4.1, aunque en un principio se valoró la transformación $y' = \log(y)$, pero esta transformación daba muchos problemas por la gran cantidad de variables que tienen valores negativos.

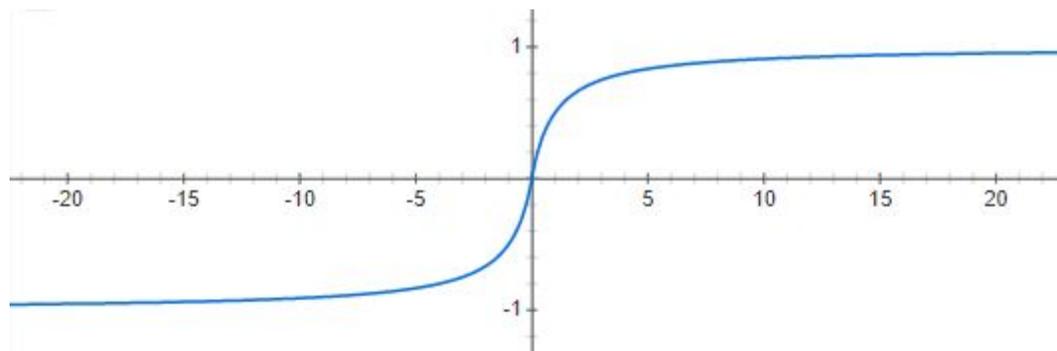


Figura 4.1: Gráfica de la transformación utilizada.

Una vez hechas las transformaciones se realizó un nuevo cálculo de las correlaciones obteniendo los valores de la tabla 4.2. En la tabla 4.2 se muestra un resumen de los resultados obtenidos, indicando las variables con las correlaciones más altas pero se puede ver en el apéndice A.2 la tabla al completo.

Ambas se componen de cuatro columnas que contienen, el nombre de la variable, la correlación obtenida, si es utilizable en los modelos o no y el número del gráfico que se generó para ver un resumen de cada análisis de correlación (disponibles en el apéndice A.3). La columna *Utilizable* que indica si podemos utilizar la variable o no en los modelos, deriva de la naturaleza de cada variable. Como nuestra variable objetivo es la suma del margen ordinario de diferentes tipos de saldo (activo, pasivo, fuera de balance, saldo en tarjetas...), no tendría sentido utilizar estas variables para la predicción ya que la suma de ellas ya nos daría el resultado buscado. Respecto a los gráficos a los que se aluden en la cuarta columna, en cada uno de ellos se muestra una cuadrícula con 4 imágenes. En la imagen de arriba a la izquierda tenemos un histograma de la variable MO_CLIENTE_CT, arriba a la derecha tenemos el valor del coeficiente de correlación entre la variable MO_CLIENTE_CT y la variable a la que se dedica cada gráfico, abajo a la izquierda tenemos un diagrama de dispersión de nuestra variable objetivo y la variable de referencia de cada gráfico, en donde la variable MO_CLIENTE_CT se representa en el eje de las x y la otra variable en el eje de las y y por último en la imagen de abajo a la derecha tenemos un histograma de la variable que se relaciona con nuestra variable objetivo en cada gráfico.

Variable	Correlación	Utilizable	Gráfico
MO_CLIENTE_CT	1	NO	Figura A.1
RF_CLIENTE_CT	0,99	NO	Figura A.2
MO_PASIVO_CT	0,67	NO	Figura A.3
MI_CLIENTE_CT	0,65	NO	Figura A.4
MO_CLIENTE_PT	0,43	NO	Figura A.5
RF_CLIENTE_PT	0,43	NO	Figura A.6
MO_PASIVO_PT	0,32	NO	Figura A.7
MI_CLIENTE_PT	0,32	NO	Figura A.8
IMP_RECIBOS_ENERGIA_4M_CT	0,29	SI	Figura A.9
IMP_RECIBOS_IMPUESTOS_4M_CT	0,29	SI	Figura A.10
MO_ACTIVO_CT	0,27	NO	Figura A.11
SALDO_MEDIO_CLIENTE_CT	0,27	SI	Figura A.12
VINCULACION_NEGOCIO_CT	0,26	SI	Figura A.13
IMP_RECIBOS_SM_4M_CT	0,25	SI	Figura A.14
NUM_RECIBOS_ENERGIA_CT	0,25	SI	Figura A.15
IMP_RECIBOS_ENERGIA_CT	0,25	SI	Figura A.16
Continúa en la siguiente página			

Cuadro 4.2 – continuado de la página anterior

Variable	Correlación	Utilizable	Gráfico
SALDO_MEDIO_ACTIVADO_CT	0,25	SI	Figura A.17
IMP_COMISIONES	0,24	SI	Figura A.18
VINCULACION_TRANSACCIONAL_CT	0,23	SI	Figura A.19
MESES_INCLUIDO_AGENDA	0,23	SI	Figura A.20
MO_ACTIVADO_PT	0,22	NO	Figura A.21
NUM_COMISIONES	0,21	SI	Figura A.22
NUM_RECIBOS_TELEFONO_CT	0,21	SI	Figura A.23
IMP_RECIBOS_TELEFONO_CT	0,21	SI	Figura A.24
IMP_RECIBOS_AGUA_4M_CT	0,2	SI	Figura A.25
IMP_RECIBOS_TELEFONO_4M_CT	0,2	SI	Figura A.26
VINCULACION_NEGOCIO_PT	0,2	SI	Figura A.27
OP_DEBE_VISTA_CT	0,2	SI	Figura A.28
SALDO_MEDIO_ACTIVADO_PT	0,2	SI	Figura A.29
IMP_DEBE_VISTA_CT	0,19	SI	Figura A.30
NUM_RECIBOS_SM_CT	0,19	SI	Figura A.31
IMP_RECIBOS_SM_CT	0,19	SI	Figura A.32
IMP_CAJ_ABONO_TJDB_3M	-0,19	SI	Figura A.33
OP_HABER_VISTA_CT	0,18	SI	Figura A.34
IMP_HABER_VISTA_CT	0,18	SI	Figura A.35
OP_CAJ_ABONO_TJDB_3M	-0,18	SI	Figura A.36
IMP_PENSION_SUA	0,17	SI	Figura A.37
IMP_PENSION	0,17	SI	Figura A.38
SALDO_TARJ_CREDITO_PT	0,17	SI	Figura A.39

Continúa en la siguiente página

Cuadro 4.2 – continuado de la página anterior

Variable	Correlación	Utilizable	Gráfico
RENTA_ESTIMADA	0,17	SI	Figura A.40
MO_SEGUROS_PT	0,17	NO	Figura A.41
MO_SEGUROS_CT	0,17	NO	Figura A.42
SALDO_MIN_VISTA_CT	0,16	SI	Figura A.43
SALDO_PDTE_PT MOS_GR_VIV_CT	0,16	SI	Figura A.44
CUOTA_PT MOS_GR_VIV_CT	0,16	SI	Figura A.45
TIPO MEDIO_PT MOS_GR_VIV_CT	0,16	SI	Figura A.46
IMP_RECIBOS_COMUNIDAD_4M_CT	0,15	SI	Figura A.47
IMP_CAJ_ABONO_TJDB	-0,15	SI	Figura A.48
IMP_RECIBOS_RESTO_4M_CT	0,14	SI	Figura A.49
SALDO_FINAL_VISTA_CT	0,14	SI	Figura A.50
SALDO_MEDIO_VISTA_CT	0,14	SI	Figura A.51
IMP_RECIBOS_COMUNIDAD_CT	0,14	SI	Figura A.52
SALDO_MEDIO_IPFS_CORTO_CT	0,14	SI	Figura A.53
SALDO_FINAL_IPFS_CORTO_CT	0,14	SI	Figura A.54
MO_TARJETAS_PT	0,14	SI	Figura A.55
MO_FB_CT	0,14	SI	Figura A.56
MO_TARJETAS_CT	0,14	SI	Figura A.57
SALDO_MEDIO_TARJETAS_PT	0,14	SI	Figura A.58

Cuadro 4.2: Tabla de correlaciones reducida (completa en apéndice).

4.3. Cálculo de los segmentos

Con las variables con más correlación intentamos obtener segmentos para clasificar a los clientes respecto a su CLV. Seleccionamos las 14 variables con mayor correlación aunque utilizamos distintas

cantidades de variables para la obtención de clusters para observar las diferencias. En principio no se contempló utilizar más variables porque la finalidad de los clusters era particionar los clientes en función del margen ordinario que generan anualmente usando el menor número de variables posible para facilitar la interpretación de los clusters.

4.3.1. Segmentación en 8 clusters (*mclust*)

En una primera prueba se generó un subconjunto de datos con las 4 variables con mayor correlación respecto a MO_CLIENTE_CT. Se mostrará con un poco de más detalle la prueba con las 4 variables y posteriormente se resumirán los resultados obtenidos con las pruebas con más variables.

Las 4 variables con mayor correlación con nuestra variable objetivo son:

- IMP_RECIBOS_ENERGIA_4M_CT

- IMP_RECIBOS_IMPUESTOS_4M_CT

- SALDO_MEDIO_CLIENTE_CT

- VINCULACION_NEGOCIO_CT

Están disponibles las estimaciones de sus funciones de densidad, además de la densidad de la variable objetivo, en el apéndice A.4. También están disponibles en el apéndice A.5 los gráficos de dispersión de cada una de las 4 variables frente a las otras. Estos gráficos se utilizaron para comprobar si se podían intuir grupos a simple vista. En estos gráficos de dispersión se incluyeron unos pocos clientes ya que con todos los datos disponibles, la acumulación de puntos era notable y no se apreciaba cómo se comportaba una variable frente a otra.

La primera prueba se realizó con el paquete *mclust*, para realizar un clustering basado en modelos cuya fundamentación teórica se refleja en el apartado 2.2.1. Tras ajustar el modelo se obtiene que los mejores resultados se obtendrían para 8 clusters como se puede comprobar en la figura 4.2 aunque se ve que hasta 4 clusters el BIC mejora bastante y a partir de 4 comienza a mejorar en menor medida. Por esta razón, mostraremos las pruebas para 4 y 8 clusters aunque se hicieron pruebas con distinto número de clusters. El significado de las abreviaturas de la leyenda de la figura 4.2 se pueden comprobar en el capítulo 2, concretamente en la tabla 2.1 y en la figura 2.1.

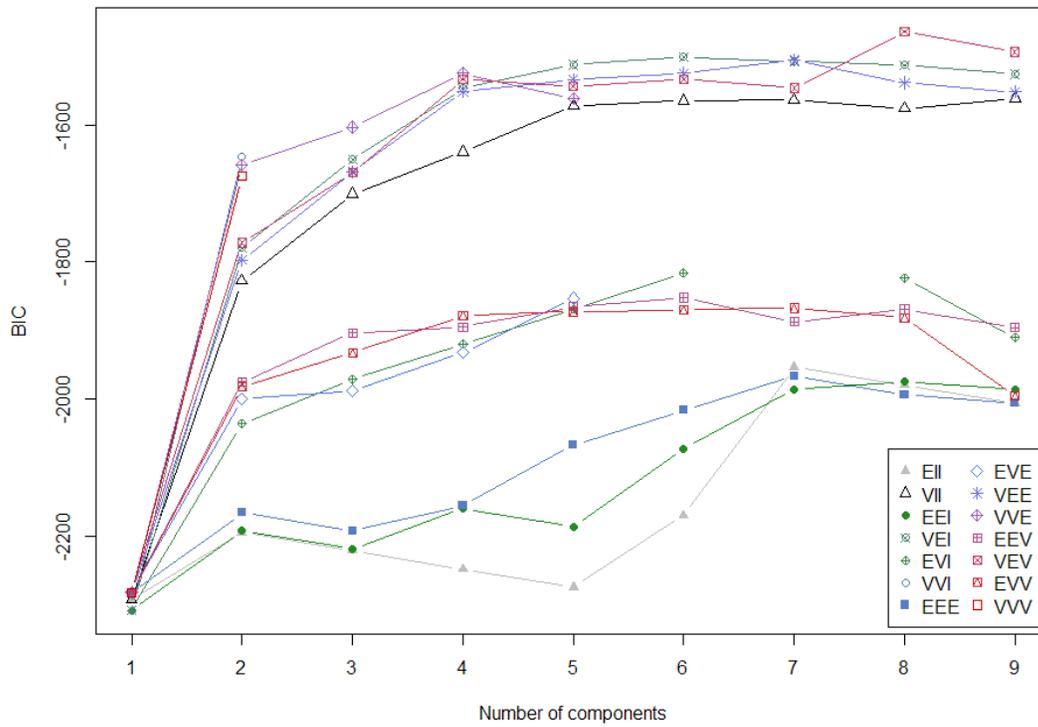


Figura 4.2: Comparación resultados según número de clusters para los modelos incluidos en la tabla 2.1

En esta prueba con 200 clientes y con *mclust*, se obtiene la segmentación de clientes que se puede ver en la tabla 4.3.

Cluster	Número de clientes
1	17
2	18
3	15
4	19
5	51
6	18
7	36
8	26

Cuadro 4.3: Tabla de resultados del *mclust*.

Además se generan los clusters de la variable objetivo, de manera que los clientes de cada cluster tienen un CLV parecido entre sí, pero distinto al de los demás clusters y se obtienen los clusters que se pueden observar en la tabla 4.4.

Cluster	Número de clientes
1	50
2	54
3	13
4	1
5	26
6	8
7	22
8	26

Cuadro 4.4: Clustering de MO_CLIENTE_CT.

Para comprobar si el clustering a partir de las cuatro variables funciona bien desde el punto de vista del CLV, se genera una tabla de contingencia (ver tabla 4.5) con los resultados de las dos anteriores tablas para observar si los clientes de cada cluster de la tabla 4.3 se clasifican también en un único cluster en la tabla 4.4 pero como se puede comprobar en la tabla 4.5 esto no sucede de manera satisfactoria. Vemos por ejemplo en la primera fila que los 50 clientes del primer cluster de la tabla 4.4 aparecen repartidos en todos los clusters de la tabla 4.3.

Cluster MO \ Cluster 4 Variables	C1	C2	C3	C4	C5	C6	C7	C8
	C1	5	2	7	2	4	3	14
C2	8	2	1	16	3	3	10	11
C3	0	4	0	0	4	5	0	0
C4	0	1	0	0	0	0	0	0
C5	3	4	2	0	12	3	2	0
C6	0	3	0	0	3	1	1	0
C7	0	1	2	0	13	2	4	0
Continúa en la siguiente página								

Cuadro 4.5 – continuado de la página anterior

Cluster MO \ Cluster 4 Variables	C1	C2	C3	C4	C5	C6	C7	C8
C8	1	1	3	1	12	1	5	2

Cuadro 4.5: Tabla de contingencia con los resultados de *mclust* para comparar los clusters generados a partir de MO con los generados a partir de las 4 variables más correlacionadas con MO.

4.3.2. Segmentación en 4 clusters (*hclust*)

Además de hacer numerosas pruebas con pocos clientes, como se mostraba en la subsección anterior, también se hicieron pruebas con una gran cantidad de clientes. A continuación se muestran los resultados obtenidos con el paquete *fpc* (Flexible Procedures for Clustering) a través de la función `cluster.magazine`. Esta función permite ejecutar un conjunto de métodos de clasificación con salida unificada para facilitar su comparación. Se realizaron varias pruebas y a continuación se muestran los resultados obtenidos con el método de clasificación de segmentación jerarquizada, cuya fundamentación teórica se expone en la subsección 2.2.2.

Para esta prueba se utilizó un conjunto de datos de 20.000 clientes. No se incluyen los dendrogramas obtenidos porque debido a la cantidad de clientes utilizados es imposible poder observar las distintas ramas. Pero sí se incluyen en el apéndice A.11 los dendrogramas que se obtendrían para una muestra de 100 clientes. En las figuras A.76 y A.77 se muestran los dendrogramas que se obtienen para 100 clientes, realizando la clasificación en función del MO y de las 4 variables más correlacionadas con MO respectivamente, seleccionando 4 clusters. En las figuras A.78 y A.79 se muestran los dendrogramas que se obtienen para 100 clientes, realizando la clasificación en función del MO y de las 4 variables más correlacionadas con MO respectivamente, seleccionando 8 clusters.

Aunque no se muestran los dendrogramas con los 20.000 clientes, a continuación se incluyen las tablas de frecuencias obtenidas. Como se puede comprobar en la tabla 4.6, no se obtienen segmentos balanceados en cuanto al número de clientes con la segmentación jerarquizada para el margen ordinario. Vemos que los 2 primeros segmentos comprenden la inmensa mayoría de los clientes. Pero incluso peores resultados se obtienen si realizamos la clasificación en función de las 4 variables con mayor dependencia respecto a `MO_CLIENCTE_CT` como se puede comprobar en la tabla 4.7. Pese a que los resultados no son buenos, se muestran también las gráficas de las densidades de la variable MO de cada cluster. Se muestran para comprobar que, pese a que los segmentos no se obtienen balanceados, utilizando el MO (figura 4.3) para generar los segmentos se obtienen grupos de clientes distintos en cuanto a MO (las densidades apenas se solapan) mientras que si se utilizan las variables para generar los grupos (figura 4.4), las densidades están muy solapadas.

Cluster	Número de clientes
1	13531
2	6416
Continúa en la siguiente página	

Cuadro 4.6 – continuado de la página anterior

Cluster	Número de clientes
3	47
4	6

Cuadro 4.6: Clustering de MO.CLIENTE.CT con 4 clusters.

Cluster	Número de clientes
1	19994
2	2
3	1
4	3

Cuadro 4.7: Clustering en función de las 4 variables con mayor correlación con MO en 4 clusters.

Cluster MO \ Cluster 4 Variables	Cluster 4 Variables			
	C1	C2	C3	C4
C1	13531	0	0	0
C2	6411	1	1	3
C3	46	1	0	0
C4	6	0	0	0

Cuadro 4.8: Tabla de contingencia para 4 clusters utilizando segmentación jerarquizada.

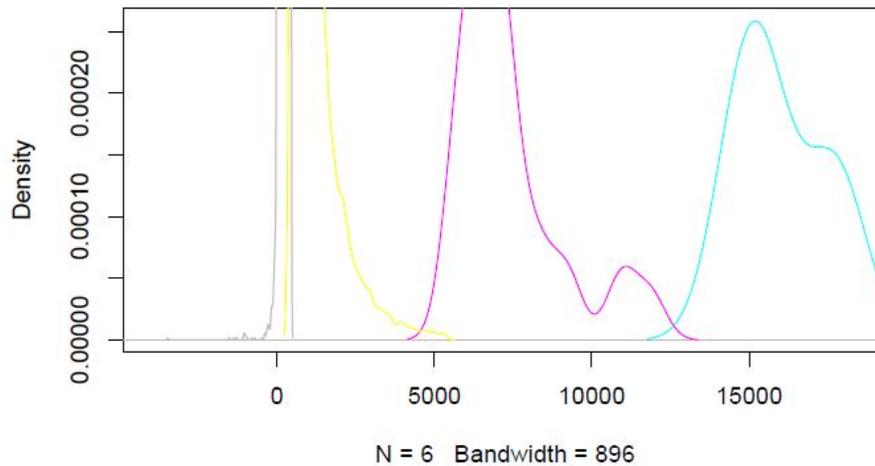


Figura 4.3: Densidades de la variable MO en los 4 clusters usando hclust sobre la variable MO

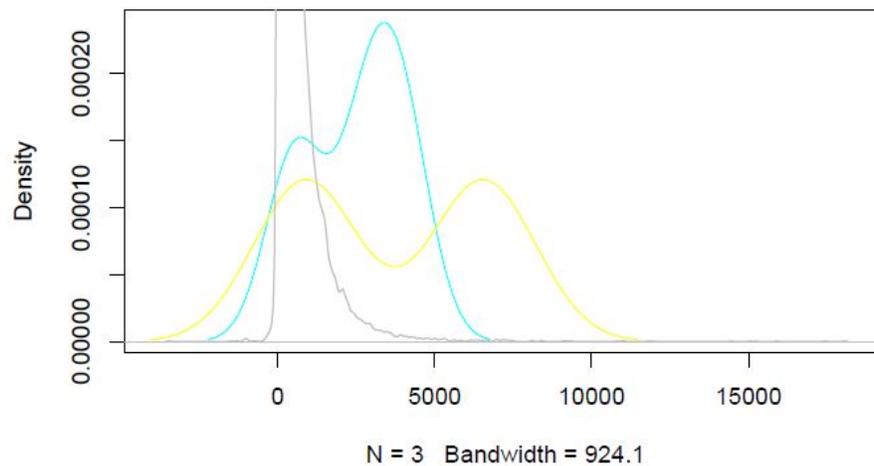


Figura 4.4: Densidades de la variable MO en los 4 clusters usando hclust sobre las cuatro variables en estudio

4.3.3. Segmentación en 8 clusters (*hclust*)

Como la segmentación en 4 clusters no funcionó todo lo bien que cabía esperar, se hicieron varias pruebas con otros números de clusters pero, como se puede comprobar a continuación, los resultados con mayor número de clústers son parecidos. A continuación se muestran los resultados obtenidos para 8 clusters con 20.000 clientes.

Cluster	Número de clientes
1	13530
Continúa en la siguiente página	

Cuadro 4.9 – continuado de la página anterior

Cluster	Número de clientes
2	6321
3	35
4	95
5	5
6	7
7	6
8	1

Cuadro 4.9: Clustering de MO_CLIENCT con 8 clusters.

Cluster	Número de clientes
1	19863
2	99
3	16
4	2
5	16
6	1
7	2
8	1

Cuadro 4.10: Clustering en función de las 4 variables con mayor correlación con MO en 8 clusters.

Cluster MO \ Cluster 4 Variables	C1	C2	C3	C4	C5	C6	C7	C8
C1	13492	38	0	0	0	0	0	0
C2	6243	58	2	1	14	1	1	1
C3	24	2	8	1	0	0	0	0
C4	87	1	4	0	2	0	1	0
C5	5	0	0	0	0	0	0	0
C6	7	0	0	0	0	0	0	0
C7	4	0	2	0	0	0	0	0
C8	1	0	0	0	0	0	0	0

Cuadro 4.11: Tabla de contingencia para 8 clusters utilizando segmentación jerarquizada.

A continuación, se muestran las gráficas de las densidades de la variable MO de cada cluster y se comprueba que ocurre lo mismo que con 4 clusters. Aunque los segmentos no se obtienen balanceados, utilizando el MO (figura 4.5) para generar los segmentos, se obtienen grupos de clientes distintos en cuanto a MO (las densidades apenas se solapan) mientras que si se utilizan las variables más correlacionadas con MO para generar los grupos (figura 4.6), las densidades están muy solapadas.

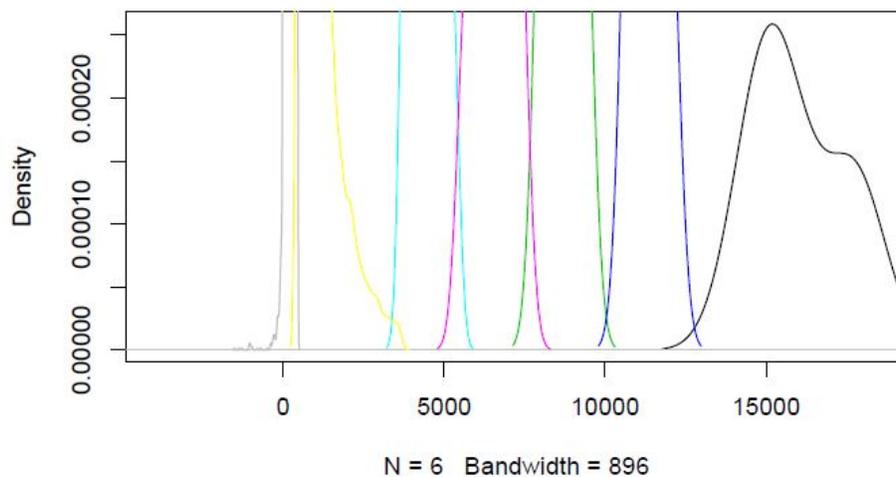


Figura 4.5: Densidades de la variable MO en los 8 clusters formados a partir de MO

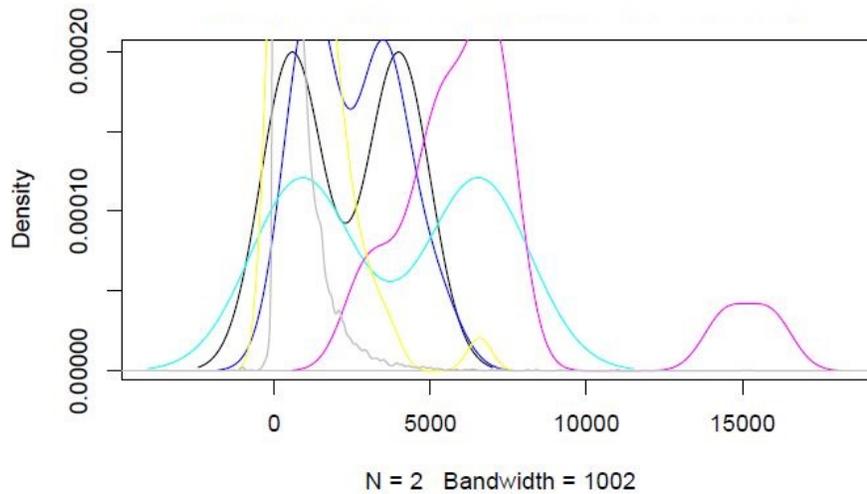


Figura 4.6: Densidades de la variable MO en los 8 clusters formados a partir de las 4 variables en estudio

4.4. Predicción

Al ver que no se podían obtener segmentos que fueran significativamente distintos en cuanto a MO pasamos a abordar la predicción.

4.4.1. Regresión lineal simple

En primer lugar se llevó a cabo un análisis de regresión lineal simple. Se empezó con un número alto de variables (todas las que tenían un coeficiente de correlación superior a 0,2 con MO_CLIENTE_CT) para ver hasta donde se podía llegar en la explicación de la variabilidad de la variable objetivo. Después se fueron eliminando las variables menos significativas para intentar obtener el modelo con el menor número de variables posible que tuviese un R^2 similar al del modelo con todas las variables iniciales.

En la tabla 4.12 se puede ver un resumen de las distintas pruebas realizadas con el número de variables en cada prueba, el R^2 y el R^2 ajustado.

Número de prueba	Número de variables	R^2	R^2 ajustado
1	13	0,4462	0,4461
2	12	0,4459	0,4459
3	11	0,4458	0,4458
4	10	0,4331	0,4331
5	9	0,4331	0,4311

Continúa en la siguiente página

Cuadro 4.12 – continuado de la página anterior

Número de prueba	Número de variables	R^2	R^2 ajustado
6	8	0,4331	0,4311
7	7	0,4326	0,4326

Cuadro 4.12: Tabla resultados pruebas regresion lineal simple.

Como se puede ver, pese a la cantidad de variables, no se llega a explicar más del 45 % de la variabilidad de la variable objetivo. Tras eliminar los coeficientes menos significativos, nos quedan coeficientes asociados a 7 variables dependientes más el intercepto con los que no se llega a explicar el 45 % de la variabilidad de la variable respuesta.

En este último modelo nos quedan las siguientes variables explicativas:

- IMP_RECIBOS_IMPUESTOS_4M_CT
- SALDO_MEDIO_CLIENTE_CT
- VINCULACION_NEGOCIO_CT
- IMP_COMISIONES
- VINCULACION_TRANSACCIONAL_CT
- NUM_COMISIONES
- IMP_RECIBOS_TELEFONO_CT

A continuación se prueba la regresión lineal con las 7 variables y todas las interacciones uno a uno posibles entre las 7 y se obtiene un R^2 de **0,4652** (con un R^2 ajustado igual). También se realiza una prueba con todas las interacciones posibles (no solo las interacciones uno a uno como en el anterior ejemplo comentado) y se obtiene un R^2 de **0,5467** y un R^2 ajustado de 0,5408. Este último es el modelo de regresión lineal simple con mayor R^2 aunque no es muy superior a los anteriores modelos mostrados y sería muy difícil de explicar debido a todas las interacciones incluidas. Se muestra a continuación la tabla 4.13 con los valores comentados anteriormente.

Número de prueba	Número de variables	R^2	R^2 ajustado
1	7 (+ interacciones 1 a 1)	0,4652	0,4652
2	7 (+ todas las interacciones)	0,5467	0,5408

Cuadro 4.13: Tabla resultados pruebas regresion lineal simple con interacciones.

4.4.2. Modelo lineal generalizado

Dado que nuestra variable objetivo es continua, para las pruebas con los modelos lineales generalizados (GLM por sus siglas en inglés) se utiliza un GLM de la familia gaussian. Se hicieron 4 pruebas distintas para comparar los resultados con la regresión lineal simple y ver si se mejoraban los resultados. Se realizó, en primer lugar un modelo con las cuatro variables con más correlación respecto a `MO_CLIENTE_CT`, a continuación, el mismo modelo realizando en la variable respuesta la misma transformación que se había realizado a la hora de obtener las correlaciones, es decir, se utilizó la función $x' = x/(1 + |x|)$, siendo $x = MO_CLIENTE_CT$, en tercer lugar se entrenó un modelo con las 7 variables que quedaron en el último modelo de regresión lineal simple y por último el mismo modelo anterior de 7 variables eliminando ciertas observaciones que por los resultados del modelo anterior parecían atípicas. Respecto a estas observaciones atípicas, se revisaron y no parecían datos incorrectos pero con este último modelo se pretendía observar si la influencia de esas observaciones provocaban un empeoramiento significativo de los resultados del modelo. Como se puede comprobar en la tabla 4.14, la influencia de estos valores (que los gráficos del modelo con 7 variables mostraba como atípicos) es muy limitada en cuanto a la variabilidad explicada por el modelo, dado que tanto el R^2 como el R^2 ajustado era el mismo (con la precisión de 4 decimales que estamos utilizando).

Número de prueba	Número de variables	R^2	R^2 ajustado
1	4	0.4116	0.4116
2	4 (transformando la respuesta)	0,0223	0,0223
3	7	0.4326	0.4326
4	7 (sin atípicos)	0,4326	0,4326

Cuadro 4.14: Tabla resultados pruebas GLM.

4.4.3. Modelos aditivos generalizados

Tras comprobar que con los GLM no se mejoraban los resultados en cuanto a variabilidad explicada, se pasó a utilizar modelos aditivos generalizados (GAM por sus siglas en inglés). En primer lugar se realizó un modelo con las mismas 7 variables y se vio que sí se mejoraban algo los resultados y tras ver esto, se añadieron algunas variables más al modelo. Con las 7 variables se pasó de un R^2 de 0,4326, que se obtenía con la regresión lineal simple y con los GLM, a un R^2 de 0,502 y al añadir alguna variable más se mejoró hasta un 0,58. Las variables con las que se obtuvo un R^2 de 0,58 son las siguientes:

- `IMP_RECIBOS_IMPUESTOS_4M_CT`
- `SALDO_MEDIO_CLIENTE_CT`
- `VINCULACION_NEGOCIO_CT`
- `IMP_COMISIONES`
- `VINCULACION_TRANSACCIONAL_CT`
- `NUM_COMISIONES`
- `IMP_RECIBOS_TELEFONO_CT`

- SALDO_MEDIO_ACTIVADO_CT
- SALDO_MEDIO_PASIVO_CT
- SALDO_MEDIO_FB_CT
- SALDO_MEDIO_TARJETAS_CT

Sobre este modelo con 11 variables se fueron eliminando variables para ver cuáles eran las más importantes para explicar MO_CLIENTE_CT. También se quería comprobar qué R^2 se podía obtener con un número menor de variables ya que 11 variables eran bastantes y se pretendía poder predecir el margen ordinario de un cliente con el menor número de variables posible. Tras eliminar las variables VINCULACION_TRANSACCIONAL_CT, NUM_COMISIONES y IMP_RECIBOS_TELEFONO_CT llegamos a un modelo con 8 variables y un R^2 de 0,577. A continuación eliminamos IMP_RECIBOS_IMPUESTOS_4M_CT y VINCULACION_NEGOCIO_CT para obtener un modelo con 6 variables y un R^2 de 0,567 y por último, eliminando SALDO_MEDIO_PASIVO_CT, SALDO_MEDIO_FB_CT y SALDO_MEDIO_TARJETAS_CT llegaríamos a un modelo con únicamente 3 variables (SALDO_MEDIO_CLIENTE_CT, IMP_COMISIONES y SALDO_MEDIO_ACTIVADO_CT) y con un R^2 de 0,519. Se puede observar un resumen de los resultados en la tabla 4.15.

Número de prueba	Número de variables	R^2 ajustado
1	11	0,58
2	8	0,577
3	6	0,567
4	3	0,519

Cuadro 4.15: Tabla resultados pruebas GAM.

Se añaden en el apéndice A.6 las funciones estimadas por el modelo GAM de 8 variables y su correspondiente función de densidad. Se incluyen estas, porque se considera el modelo de 8 variables como el mejor modelo GAM ya que en cuanto a R^2 es muy similar al modelo de 11 variables pero se reduce el modelo bastante al eliminar 3 variables del modelo inicial.

4.4.4. Otras pruebas

Revisando los clientes con peores predicciones se intentó mejorar los resultados realizando segmentaciones de clientes. Por ejemplo, parecía que los mayores errores se producían en clientes con alto margen ordinario causado por tener una proporción muy alta del saldo medio en productos de fuera de balance. Por esta razón, se segmentaron los clientes en 2 grupos, clientes con saldo en fuera de balance y clientes sin saldo en fuera de balance.

Tras ajustar el modelo GAM con 11 variables para cada uno de los segmentos, se obtuvo el mismo R^2 ajustado (0,58) para los clientes sin fuera de balance y un R^2 ajustado de 0,51 para los clientes con saldo en fuera de balance por lo tanto no se mejoró la predicción con esta segmentación.

En vista de que esa segmentación no era buena, se decidió segmentar a los clientes en función de lo correcta que era la estimación del modelo. La finalidad era comprobar qué R^2 se obtenía con los

clientes que se estimaban bien y qué R^2 se obtenía con los clientes mal estimados, y posteriormente intentar estudiar las características de los clientes que se estimaban mal para intentar realizar un modelo específico para ellos.

4.4.5. Segmentación según la precisión de la estimación

Como se comentaba, tras ver que algunas segmentaciones no funcionaron se realizó una segmentación según la precisión de la estimación del modelo y esta segmentación sí que generó resultados más satisfactorios que los observados hasta el momento.

Para realizar estas pruebas, el primer paso fue obtener el porcentaje de desviación de cada estimación del modelo respecto de la observación real. Con estos porcentajes se hicieron varias pruebas. En cada una de ellas, se elegía un porcentaje que marcaría el punto de corte entre las estimaciones buenas y las malas y a continuación se generaban 3 segmentos: uno con los clientes para los que se obtenía una estimación con un error porcentual menor al umbral elegido para la prueba (vamos a considerarlo como el segmento con los aciertos), un segundo segmento con los clientes cuya estimación era superior al valor real y excedía el umbral marcado para la prueba (que lo identificaremos como el segmento con desviaciones por encima) y un último segmento con los clientes cuya estimación era inferior al valor real y excedía el umbral marcado para la prueba (que lo identificaremos como el segmento con las desviaciones por debajo).

La primera prueba se realiza fijando un porcentaje de precisión del 25%. Tras observar los porcentajes de desviación de la estimación respecto del valor real de la variable respuesta, se obtiene que, de los 200.000 clientes que se tenían en la muestra, 79.805 se encuentran en el segmento cuyo error es inferior al 25%, 82.306 se encuentran en el segmento cuyo error era superior al margen marcado y el valor estimado era superior al valor real y 37.889 clientes conformaban el tercer segmento con los clientes cuyo error era superior al 25% y el valor estimado por el modelo era inferior al valor real. En la figura 4.7, los 79.805 clientes cuyo error de estimación es inferior al 25% se corresponderían con las observaciones de color negro, los 82.306 clientes estimados por encima del valor real, cometiendo un error superior al 25%, serían las observaciones de color rojo y los 37.889 clientes estimados por debajo del valor real con un error superior al umbral se corresponderían con las observaciones de color verde.

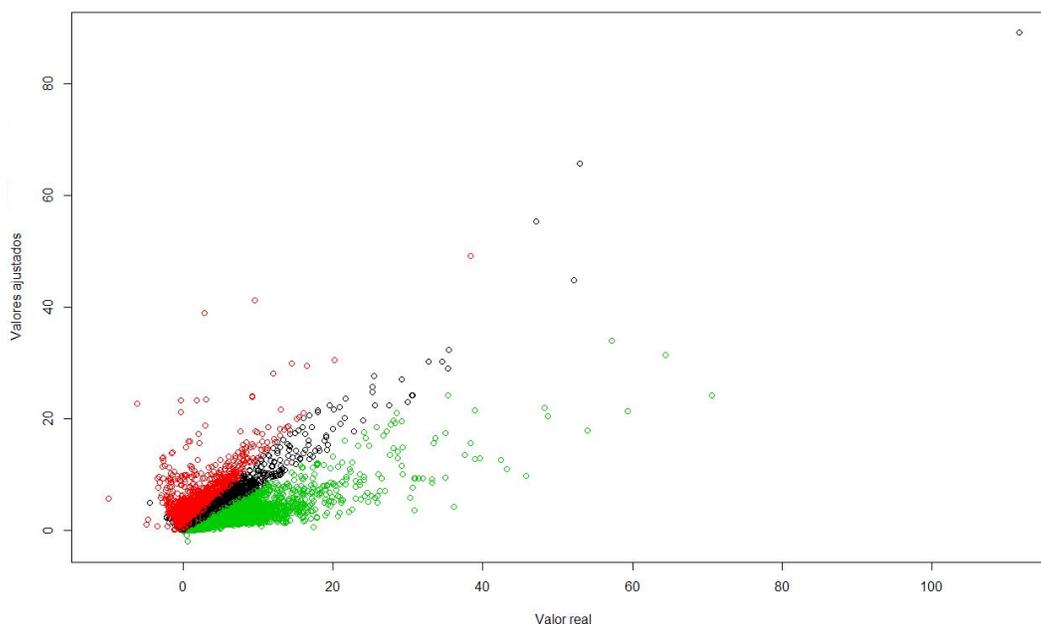


Figura 4.7: Gráfico representativo de los 3 segmentos generados por la desviación del valor ajustado respecto del valor real con un umbral del 25 %.

También se han incluido en el apéndice A.7 las densidades de las variables utilizadas en los modelos GAM para cada uno de los segmentos anteriores, para poder comprobar cómo se comporta cada segmento en cada variable. Por lo tanto en la figura A.59 tenemos la densidad de la variable MO en cada uno de los segmentos anteriores y lo mismo tendríamos en las figuras A.60, A.61, A.62, A.63, A.64, A.65, A.66 y A.67 para cada una de las variables explicativas del modelo GAM.

Una vez obtenidos los subconjuntos pasamos a realizar un modelo GAM para cada uno de ellos. Se utiliza el modelo de 8 variables (que obtenía casi tan buenos resultados como el de 11) que entrenándolo con toda la muestra de clientes se obtenía un R^2 ajustado de 0,577. Al aplicarlo a cada uno de los subconjuntos se observa que para el segmento con las desviaciones por encima el modelo tiene un R^2 ajustado de 0,801, para el segmento con las desviaciones por debajo se obtiene un 0,833 y para el segmento con los aciertos un R^2 ajustado de 0,965. Teniendo en cuenta el número de clientes en cada segmento se obtiene un R^2 ponderado de 0,872.

A continuación se realizaron distintas pruebas variando el umbral que marca el límite con el que considerar si un cliente se va al segmento de los aciertos o a uno de los segmentos con las desviaciones de la estimación del modelo respecto de valor real de la variable respuesta. En la tabla 4.16 se puede ver el número de clientes que se va a cada segmento según el umbral elegido y en la tabla 4.17 vemos el R^2 que se obtiene con cada una de las pruebas.

Umbral	Segmento aciertos	Segmento desv. encima	Segmento desv. debajo
15 %	51.936	95.151	52.913
25 %	79.805	82.306	37.889

Continúa en la siguiente página

Cuadro 4.16 – continuado de la página anterior

Umbral	Segmento aciertos	Segmento desv. encima	Segmento desv. debajo
30 %	90.330	76.667	33.003
40 %	109.448	66.400	24.152
50 %	126.286	57.600	16.114

Cuadro 4.16: Número de clientes en cada segmento en función del umbral.

Como se puede observar, en la tabla 4.16, a partir del umbral del 30 % empezamos a ver un número de clientes en el segmento de aciertos superior al número de clientes que hay en los otros segmentos. También se puede observar que siempre es superior el número de clientes cuya predicción se desvía por encima del valor real de la variable respuesta que los que se desvían por debajo.

Umbral	R^2 S. aciertos	R^2 S. desv. encima	R^2 S. desv. debajo	R^2 Ponderado
15 %	0,9850	0,8258	0,8290	0,8687
25 %	0,9652	0,8334	0,8007	0,8725
30 %	0,9559	0,8459	0,7883	0,8735
40 %	0,9297	0,8746	0,7547	0,8649
50 %	0,8996	0,8983	0,7248	0,8492

Cuadro 4.17: Resumen de los R^2 obtenidos en función del umbral elegido

En cuanto a los R^2 obtenidos con cada umbral vemos que en las pruebas realizadas los mejores resultados se obtienen con los umbrales del 20 % y del 30 %, siendo este último ligeramente superior y el mejor de todos los R^2 observados.

Pero este proceso de predicción tiene un inconveniente respecto al planteamiento inicial del problema. Desde un principio se planteaba la idea de intentar predecir el valor de vida del cliente a partir de unas variables que se pudieran solicitar al cliente en el momento en que este quisiera formalizar una relación contractual con la entidad. Pero en este proceso que acabamos de describir se estaba utilizando la variable MO_CLIENTE_CT para determinar con qué modelo se iba a predecir el margen ordinario del cliente y esta variable no sería susceptible de solicitar a un cliente en estas condiciones comentadas. Por lo tanto, con esta carencia surge un nuevo hito en el desarrollo del proceso de predicción del valor de vida de un cliente: clasificar a los clientes en uno de los 3 segmentos que ayudan a realizar una predicción más precisa.

4.5. Árboles de clasificación

Como vimos en el apartado anterior, si obteníamos 3 segmentos en donde agrupábamos los clientes según la precisión de la predicción inicial respecto al valor real de la variable MO_CLIENTE_CT, se podía llegar a obtener un R^2 de más de 0,85. Para intentar esta clasificación se parte de la clasificación realizada en el anterior apartado y se intenta entrenar un modelo para predecir esta clasificación a partir de otras variables que no sean el MO_CLIENTE_CT.

Para llevar a cabo esta clasificación se eligen dos métodos de clasificación diferentes, **árboles de clasificación y regresión** (CART por sus siglas en inglés) y **random forest**. Pese a que los mejores resultados en cuanto a R^2 se obtenían fijando el umbral de error en el 30 %, se decide intentar la clasificación con varios de los porcentajes por si alguno de ellos resultaba significativamente más efectivo que los otros. De esta forma, aunque con uno de los umbrales no se obtuvieran tan buenos resultados como con el umbral en el 30 %, podía tener mejores resultados en la clasificación del segmento y por esta razón obtener mejores resultados con el R^2 ponderado final.

4.5.1. Árboles de clasificación y regresión

Tras realizar la clasificación para los umbrales de 25 %, 30 % y 50 %, se puede comprobar en la tabla 4.18 que los mejores resultados se obtienen con el umbral del 50 %. Vemos que con este margen de error se predice correctamente el segmento en el que agrupar cada cliente en un 71.90 % de los casos.

Umbral	% Error	% Acierto
25 %	35,53 %	64,47 %
30 %	34,80 %	65,20 %
50 %	28,10 %	71,90 %

Cuadro 4.18: Tabla precisión clasificación CART

4.5.2. Random forest

Ahora se indican los resultados obtenidos tras clasificar a los clientes con random forest. Se indican con algo más de detalle que los resultados obtenidos con CART porque, como se puede comprobar si comparamos los resultados, la precisión obtenida con random forest (4.22) es algo mejor que con CART (4.18). Así que se muestran también los resultados de las matrices de confusión para que se pueda observar cómo se equivoca la clasificación ya que, además de conocer cuánto se equivoca es importante saber en qué clasificaciones se equivoca. Si la equivocación es que un cliente del segmento de aciertos se clasifica en uno de los segmentos de desviación, podría ser causada por estar cerca del umbral, y en estos casos el modelo aunque no fuese el ideal, podría hacer una buena estimación, pero si un cliente es del segmento de desviaciones de los que se estiman por encima del valor real y se clasifica en el otro segmento de desviaciones, este cliente se va a estimar muy mal con el modelo que se le asigna.

En cuanto a la clasificación del Random Forest, se ha elegido una muestra de entrenamiento del 70 % del total de la muestra utilizada, por lo tanto de los 200.000 clientes que estábamos manejando en la muestra, se utilizan 140.000 para el entrenamiento del modelo y 60.000 para realizar el test de clasificación.

A continuación se muestran las matrices de confusión obtenidas con el random forest para los umbrales de 25 %, 30 % y 50 % en las figuras 4.19, 4.20 y 4.21 respectivamente.

Segm. predicción \ Segm. real	Segm. aciertos	Segm. desv. encima	Segm. desv. debajo
Segm. aciertos	16.974	4.072	2.521
Segm. desv. encima	5.928	19.233	4.505
Segm. desv. debajo	1.200	1.356	4.211

Cuadro 4.19: Matriz de confusión con umbral del 25 %.

Segm. predicción \ Segm. real	Segm. aciertos	Segm. desv. encima	Segm. desv. debajo
Segm. aciertos	20.619	4.891	2.459
Segm. desv. encima	5.504	17.002	3.722
Segm. desv. debajo	1.143	1.101	3.559

Cuadro 4.20: Matriz de confusión con umbral del 30 %.

Segm. predicción \ Segm. real	Segm. aciertos	Segm. desv. encima	Segm. desv. debajo
Segm. aciertos	32.321	5.817	1.975
Segm. desv. encima	5.099	11.018	1.388
Segm. desv. debajo	625	341	1.416

Cuadro 4.21: Matriz de confusión con umbral del 50 %.

Y a continuación se muestran los porcentajes de acierto de los modelos que obtuvieron las anteriores matrices de confusión. Vemos que los resultados con el random forest que podemos observar en la tabla 4.22 son mejores que los que obteníamos con CART (tabla 4.18).

Umbral	% Error	% Acierto
25 %	32,64 %	67,36 %
30 %	31,37 %	68,63 %
50 %	25,41 %	74,59 %

Cuadro 4.22: Tabla precisión clasificación Random forest.

4.5.3. Resultados finales tras la clasificación

Con las clasificaciones anteriores se realizaron de nuevo modelos GAM para ver qué R^2 se llegaba a obtener. Se observa que la precisión baja considerablemente, como era de esperar, viendo que en ninguna de las pruebas de clasificación se llega a alcanzar una precisión del 75 %. En la tabla 4.23 vemos que en el mejor de los casos se obtiene un R^2 ponderado de 0,63 lo que es un valor muy lejos del 0,8735 que se obtenía con la segmentación del umbral al 30 %.

Umbral	R^2 S. aciertos	R^2 S. desv. encima	R^2 S. desv. debajo	R^2 Ponderado
25 %	0,8225	0,6625	0,4766	0,6334
30 %	0,7866	0,6713	0,4567	0,6313
50 %	0,6839	0,7227	0,4243	0,6097

Cuadro 4.23: Tabla resumen modelos GAM finales

Desde el punto de vista práctico, estos modelos anteriores serían posibles, por ejemplo, a la hora de la adquisición de una cartera de clientes, ya que se estaba manejando un número alto de clientes. Pero lo más habitual en la práctica sería realizar una predicción para un cliente potencial en un determinado momento. Para ello, en lugar de hacer nuevos modelos como se hace en la sección anterior, habría que realizar predicciones con los modelos ya existentes. Al calcular el R^2 a partir de las predicciones se obtiene que: los clientes clasificados en el segmento de aciertos se predecirían con una precisión del 79,29 %, los clientes clasificados en el segmento con las desviaciones por encima se predecirían con una precisión del 30,06 % y los clientes clasificados en el segmento con las desviaciones por debajo se predecirían con una precisión del 27,18 %.

Como era de esperar, el R^2 para los clientes clasificados en el segmento de los aciertos es relativamente bueno, pero el problema surge con las predicciones de los segmentos de desviaciones. En primer lugar, el modelo del segmento de los aciertos era el que mejor R^2 obtenía y esa es una de las razones por las que las predicciones de este modelo son más precisas. Y en segundo lugar, los errores de clasificación que más van a penalizar serán cuando una observación que debería clasificarse en un segmento de desviaciones, se clasifica en el otro segmento de desviaciones (y no en el segmento de aciertos), ya que los modelos más diferentes entre sí son los segmentos de desviaciones como se puede deducir de la figura 4.7.

Capítulo 5

Conclusiones

Teniendo en cuenta los resultados obtenidos, vemos que las variables disponibles en el data mart de clientes permiten obtener buenas predicciones del margen ordinario, y por lo tanto, del customer lifetime value.

Tras identificar las variables que tienen más correlación con el CLV, se ha comprobado que, mediante el uso de modelos GAM, se puede generar un proceso por el cual se llega a explicar más del 87% de la variabilidad del margen ordinario. Este proceso se puede resumir en los siguientes pasos:

- Ajuste de un primer modelo GAM para comprobar cuánto se desvía cada estimación de su valor real.
- Clasificación de cada cliente en función de la precisión de la estimación. Tras fijar el umbral que marca el límite de error, se realiza la siguiente clasificación: en el primer grupo se incluyen los clientes cuya estimación de margen ordinario ha sido superior al valor real y, que además, el porcentaje de error cometido en la estimación ha sido superior al umbral fijado, en el segundo grupo se incluyen los clientes cuya estimación de margen ordinario ha sido inferior al valor real y cuyo porcentaje de error ha sido superior al umbral fijado y en el último grupo se incluyen los clientes cuyo porcentaje de error en la estimación del margen ordinario no ha superado el umbral fijado.
- Ajuste de un modelo GAM para cada uno de los 3 grupos generados.

Con este proceso, se obtiene para cada uno de los 3 modelos ajustados un R^2 superior a 0,80, que tras ponderación, se corresponde con un R^2 global superior a 0,87. El inconveniente de este proceso es que para poder realizar la clasificación, es necesario tener el valor real de margen ordinario de cada cliente y, aunque en el data mart de clientes está disponible, esta clasificación no podría hacerse a un cliente potencial porque este dato no sería conocido antes de formalizar una relación contractual con la entidad.

Como uno de los objetivos de este trabajo era la predicción del CLV para los clientes potenciales, se han realizado técnicas de clasificación para agrupar los clientes sin utilizar su valor real de margen ordinario pero en este caso los resultados obtenidos no han sido suficientemente satisfactorios.

El problema es que en el conjunto de clientes parecen existir segmentos en los que el margen ordinario se comporta de manera diferente y estos segmentos no se han conseguido identificar de forma precisa con las variables disponibles.

Por todo ello, una de las posibles vías para mejorar los resultados sería profundizar en el estudio de los clientes de cada segmento, porque con una correcta clasificación, la predicción sería más precisa. Además, en caso de saber cómo diferenciar estos clientes, probablemente quizás ya no sea necesario realizar esta segmentación y un único GAM podría modelizar el comportamiento del margen ordinario de todos los clientes.

Apéndice A

Apéndice resultados

En este apéndice se incluye toda la información complementaria que puede ser de ayuda para la comprensión de los distintos capítulos y para completar la información del trabajo.

A.1. Tabla resultados del ACP

Componente	Desviación típica	Proporción de variación explicada	Proporción acumulada
PC1	3,74361	0,05964	0,05964
PC2	2,81076	0,03362	0,09326
PC3	2,68842	0,03076	0,12401
PC4	2,42986	0,02512	0,14914
PC5	2,35042	0,02351	0,17264
PC6	2,20267	0,02065	0,19329
PC7	2,17188	0,02007	0,21336
PC8	2,12950	0,01930	0,23270
PC9	2,09474	0,01867	0,25133
PC10	2,08460	0,01849	0,26982
PC11	2,06405	0,01813	0,28795
PC12	2,05384	0,01795	0,30590
PC13	2,04277	0,01776	0,32366

Continúa en la siguiente página

Cuadro A.1 – continuado de la página anterior

Componente	Desviación típica	Proporción de variación explicada	Proporción acumulada
PC14	2,00983	0,01719	0,34085
PC15	1,98584	0,01678	0,35763
PC16	1,96983	0,01651	0,37414
PC17	1,94742	0,01614	0,39028
PC18	1,94083	0,01603	0,40631
PC19	1,93478	0,01593	0,42224
PC20	1,92149	0,01571	0,43795
PC21	1,89783	0,01533	0,45327
PC22	1,86311	0,01477	0,46805
PC23	1,78658	0,01358	0,48163
PC24	1,76150	0,01320	0,49480
PC25	1,74217	0,01292	0,50775
PC26	1,61451	0,01109	0,51884
PC27	1,60713	0,01099	0,52983
PC28	1,54159	0,01011	0,53994
PC29	1,52264	0,00987	0,54981
PC30	1,48927	0,00944	0,55925
PC31	1,47464	0,00925	0,56850
PC32	1,45981	0,00907	0,57757
PC33	1,41424	0,00851	0,58608
PC34	1,40763	0,00843	0,59451
PC35	1,38283	0,00814	0,60265
PC36	1,34153	0,00766	0,61031

Continúa en la siguiente página

Cuadro A.1 – continuado de la página anterior

Componente	Desviación típica	Proporción de variación explicada	Proporción acumulada
PC37	1,33610	0,00760	0,61790
PC38	1,31376	0,00734	0,62525
PC39	1,30769	0,00728	0,63252
PC40	1,29350	0,00712	0,63964
PC41	1,28334	0,00701	0,64665
PC42	1,26813	0,00684	0,65305
PC43	1,25081	0,00666	0,66015
PC44	1,24616	0,00661	0,66676
PC45	1,23023	0,00644	0,67320
PC46	1,21609	0,00629	0,67950
PC47	1,20092	0,00614	0,68563
PC48	1,19398	0,00607	0,69170
PC49	1,16939	0,00582	0,69752
PC50	1,14554	0,00558	0,70310
PC51	1,13987	0,00553	0,70863
PC52	1,12851	0,00542	0,71405
PC53	1,10249	0,00517	0,71922
PC54	1,09200	0,00507	0,72430
PC55	1,08118	0,00497	0,72927
PC56	1,06610	0,00484	0,73411
PC57	1,05949	0,00478	0,73888
PC58	1,04560	0,00465	0,74354
PC59	1,03608	0,00457	0,74810

Continúa en la siguiente página

Cuadro A.1 – continuado de la página anterior

Componente	Desviación típica	Proporción de variación explicada	Proporción acumulada
PC60	1,02413	0,00446	0,75257
PC61	1,01605	0,00439	0,75696
PC62	1,00366	0,00429	0,76125
PC63	0,99566	0,00422	0,76547
PC64	0,99022	0,00417	0,76964
PC65	0,98614	0,00414	0,77378
PC66	0,97376	0,00403	0,77781
PC67	0,96403	0,00395	0,78177
PC68	0,95918	0,00392	0,78568
PC69	0,94625	0,00381	0,78949
PC70	0,93764	0,00374	0,79323
PC71	0,93409	0,00371	0,79694
PC72	0,91627	0,00357	0,80052
PC73	0,90760	0,00350	0,80400
PC74	0,90327	0,00347	0,80749
PC75	0,89169	0,00338	0,81088
PC76	0,88080	0,00330	0,81420
PC77	0,87504	0,00326	0,81744
PC78	0,85635	0,00312	0,82056
PC79	0,85062	0,00308	0,82364
PC80	0,84180	0,00302	0,82665
PC81	0,83821	0,00299	0,82964
PC82	0,82856	0,00292	0,83256

Continúa en la siguiente página

Cuadro A.1 – continuado de la página anterior

Componente	Desviación típica	Proporción de variación explicada	Proporción acumulada
PC83	0,81936	0,00286	0,83542
PC84	0,81521	0,00283	0,83825
PC85	0,81193	0,00281	0,84105
PC86	0,80632	0,00277	0,84382
PC87	0,79710	0,00270	0,84650
PC88	0,79303	0,00268	0,84920
PC89	0,79031	0,00266	0,85186
PC90	0,78609	0,00263	0,85449
PC91	0,78038	0,00259	0,85708
PC92	0,77447	0,00255	0,85963
PC93	0,77327	0,00254	0,86218
PC94	0,77075	0,00253	0,86470
PC95	0,76446	0,00249	0,86719
PC96	0,75827	0,00245	0,86964
PC97	0,75540	0,00243	0,87206
PC98	0,75140	0,00240	0,87450
PC99	0,74721	0,00238	0,87684
PC100	0,74208	0,00234	0,87919
PC101	0,74062	0,00233	0,88152
PC102	0,73673	0,00231	0,88383
PC103	0,73070	0,00227	0,88610
PC104	0,72265	0,00222	0,88832
PC105	0,71840	0,00220	0,89050

Continúa en la siguiente página

Cuadro A.1 – continuado de la página anterior

Componente	Desviación típica	Proporción de variación explicada	Proporción acumulada
PC106	0,71707	0,00219	0,89271
PC107	0,71338	0,00217	0,89487
PC108	0,70774	0,00213	0,89701
PC109	0,70476	0,00211	0,89912
PC110	0,69859	0,00208	0,90120
PC111	0,69436	0,00205	0,90325
PC112	0,68490	0,00200	0,90520
PC113	0,68359	0,00199	0,90723
PC114	0,67795	0,00196	0,90919
PC115	0,67129	0,00192	0,91111
PC116	0,66810	0,00190	0,91300
PC117	0,66614	0,00189	0,91489
PC118	0,66331	0,00187	0,91677
PC119	0,66064	0,00186	0,91862
PC120	0,65911	0,00185	0,92047
PC121	0,65830	0,00184	0,92232
PC122	0,65682	0,00184	0,92415
PC123	0,65605	0,00183	0,92598
PC124	0,65218	0,00181	0,92779
PC125	0,64803	0,00179	0,92958
PC126	0,64278	0,00176	0,93134
PC127	0,64115	0,00175	0,93309
PC128	0,63734	0,00173	0,93482

Continúa en la siguiente página

Cuadro A.1 – continuado de la página anterior

Componente	Desviación típica	Proporción de variación explicada	Proporción acumulada
PC129	0,63405	0,00171	0,93653
PC130	0,63260	0,00170	0,93820
PC131	0,62832	0,00168	0,93991
PC132	0,61726	0,00162	0,94153
PC133	0,61518	0,00161	0,94314
PC134	0,60694	0,00157	0,94471
PC135	0,60529	0,00156	0,94627
PC136	0,60108	0,00154	0,94781
PC137	0,59261	0,00149	0,94930
PC138	0,58971	0,00148	0,95078
PC139	0,58088	0,00144	0,95222
PC140	0,57877	0,00143	0,95364
PC141	0,56445	0,00136	0,95500
PC142	0,56130	0,00134	0,95634
PC143	0,56007	0,00133	0,95767
PC144	0,55512	0,00131	0,95898
PC145	0,54921	0,00128	0,96027
PC146	0,54766	0,00128	0,96154
PC147	0,53403	0,00121	0,96276
PC148	0,53120	0,00120	0,96400
PC149	0,52930	0,00119	0,96515
PC150	0,51677	0,00114	0,96629
PC151	0,51577	0,00113	0,96742

Continúa en la siguiente página

Cuadro A.1 – continuado de la página anterior

Componente	Desviación típica	Proporción de variación explicada	Proporción acumulada
PC152	0,51234	0,00112	0,96854
PC153	0,51067	0,00111	0,96964
PC154	0,50008	0,00106	0,97071
PC155	0,49641	0,00105	0,97176
PC156	0,49263	0,00103	0,97279
PC157	0,48698	0,00101	0,97380
PC158	0,48400	0,00100	0,97480
PC159	0,47675	0,00097	0,97576
PC160	0,47482	0,00096	0,97672
PC161	0,46921	0,00094	0,97766
PC162	0,45200	0,00087	0,97853
PC163	0,43430	0,00080	0,97930
PC164	0,43048	0,00079	0,98012
PC165	0,41992	0,00075	0,98087
PC166	0,41909	0,00075	0,98162
PC167	0,41577	0,00074	0,98235
PC168	0,41537	0,00073	0,98309
PC169	0,40680	0,00070	0,98380
PC170	0,40470	0,00070	0,98450
PC171	0,39979	0,00068	0,98517
PC172	0,39236	0,00066	0,98582
PC173	0,38199	0,00062	0,98645
PC174	0,38024	0,00062	0,98706

Continúa en la siguiente página

Cuadro A.1 – continuado de la página anterior

Componente	Desviación típica	Proporción de variación explicada	Proporción acumulada
PC175	0,37520	0,00060	0,98770
PC176	0,36976	0,00058	0,98824
PC177	0,36716	0,00057	0,98881
PC178	0,35640	0,00054	0,98936
PC179	0,34440	0,00050	0,98990
PC180	0,33563	0,00048	0,99034
PC181	0,32961	0,00046	0,99080
PC182	0,32907	0,00046	0,99126
PC183	0,32480	0,00045	0,99171
PC184	0,31082	0,00041	0,99212
PC185	0,29815	0,00038	0,99250
PC186	0,29093	0,00036	0,99286
PC187	0,28674	0,00035	0,99321
PC188	0,28126	0,00034	0,99355
PC189	0,27556	0,00032	0,99387
PC190	0,27279	0,00032	0,99419
PC191	0,26859	0,00031	0,99449
PC192	0,26360	0,00030	0,99480
PC193	0,26254	0,00029	0,99508
PC194	0,25202	0,00027	0,99535
PC195	0,25025	0,00027	0,99562
PC196	0,24994	0,00027	0,99589
PC197	0,24535	0,00026	0,99614

Continúa en la siguiente página

Cuadro A.1 – continuado de la página anterior

Componente	Desviación típica	Proporción de variación explicada	Proporción acumulada
PC198	0,24155	0,00025	0,99639
PC199	0,23926	0,00024	0,99663
PC200	0,23818	0,00024	0,99688
PC201	0,23120	0,00023	0,99710
PC202	0,23035	0,00023	0,99733
PC203	0,22799	0,00022	0,99755
PC204	0,22633	0,00022	0,99777
PC205	0,22541	0,00022	0,99798
PC206	0,22340	0,00021	0,99820
PC207	0,21860	0,00020	0,99840
PC208	0,20935	0,00019	0,99859
PC209	0,20459	0,00018	0,99876
PC210	0,19990	0,00017	0,99893
PC211	0,18211	0,00014	0,99908
PC212	0,17598	0,00013	0,99921
PC213	0,15360	0,00010	0,99930
PC214	0,14747	0,00009	0,99940
PC215	0,14329	0,00009	0,99949
PC216	0,14304	0,00009	0,99957
PC217	0,13878	0,00008	0,99966
PC218	0,12504	0,00007	0,99972
PC219	0,11970	0,00006	0,99978
PC220	0,11683	0,00006	0,99984

Continúa en la siguiente página

A.2. TABLA RESULTADOS DEL ANÁLISIS DE CORRELACIONES ENTRE CADA VARIABLE Y MO_CLIENTE_CT

Cuadro A.1 – continuado de la página anterior

Componente	Desviación típica	Proporción de variación explicada	Proporción acumulada
PC221	0,10535	0,00005	0,99989
PC222	0,09986	0,00004	0,99993
PC223	0,07496	0,00002	0,99996
PC224	0,05956	0,00002	0,99997
PC225	0,05167	0,00001	0,99998
PC226	0,04664	0,00001	0,99999
PC227	0,04414	0,00001	1,00000
PC228	0,007208	0,00000	1,00000
PC229	0,0001737	0,00000	1,00000
PC230	0,000000000000003491	0,00000	1,00000
PC231	0,000000000000006153	0,00000	1,00000
PC232	0,000000000000001662	0,00000	1,00000
PC233	0,00000000000000035	0,00000	1,00000
PC234	2,826E-16	0,00000	1,00000
PC235	2,518E-16	0,00000	1,00000

Cuadro A.1: Resultado ACP.

A.2. Tabla resultados del análisis de correlaciones entre cada variable y MO_CLIENTE_CT

Variable	Correlación	ABS(Correlación)	Gráfico
MO_CLIENTE_CT	1	1	Figura A.1
RF_CLIENTE_CT	0,99	0,99	Figura A.2
MO_PASIVO_CT	0,67	0,67	Figura A.3

Continúa en la siguiente página

Cuadro A.2 – continuado de la página anterior

Variable	Correlación	ABS(Correlación)	Gráfico
MI_CLIENTE_CT	0,65	0,65	Figura A.4
MO_CLIENTE_PT	0,43	0,43	Figura A.5
RF_CLIENTE_PT	0,43	0,43	Figura A.6
MO_PASIVO_PT	0,32	0,32	Figura A.7
MI_CLIENTE_PT	0,32	0,32	Figura A.8
IMP_RECIBOS_ENERGIA_4M_CT	0,29	0,29	Figura A.9
IMP_RECIBOS_IMPUESTOS_4M_CT	0,29	0,29	Figura A.10
MO_ACTIVADO_CT	0,27	0,27	Figura A.11
SALDO_MEDIO_CLIENTE_CT	0,27	0,27	Figura A.12
VINCULACION_NEGOCIO_CT	0,26	0,26	Figura A.13
IMP_RECIBOS_SM_4M_CT	0,25	0,25	Figura A.14
NUM_RECIBOS_ENERGIA_CT	0,25	0,25	Figura A.15
IMP_RECIBOS_ENERGIA_CT	0,25	0,25	Figura A.16
SALDO_MEDIO_ACTIVADO_CT	0,25	0,25	Figura A.17
IMP_COMISIONES	0,24	0,24	Figura A.18
VINCULACION_TRANSACCIONAL_CT	0,23	0,23	Figura A.19
MESES_INCLUIDO_AGENDA	0,23	0,23	Figura A.20
MO_ACTIVADO_PT	0,22	0,22	Figura A.21
NUM_COMISIONES	0,21	0,21	Figura A.22
NUM_RECIBOS_TELEFONO_CT	0,21	0,21	Figura A.23
IMP_RECIBOS_TELEFONO_CT	0,21	0,21	Figura A.24
IMP_RECIBOS_AGUA_4M_CT	0,2	0,2	Figura A.25
IMP_RECIBOS_TELEFONO_4M_CT	0,2	0,2	Figura A.26

Continúa en la siguiente página

Cuadro A.2 – continuado de la página anterior

Variable	Correlación	ABS(Correlación)	Gráfico
VINCULACION_NEGOCIO_PT	0,2	0,2	Figura A.27
OP_DEBE_VISTA_CT	0,2	0,2	Figura A.28
SALDO_MEDIO_ACTIVADO_PT	0,2	0,2	Figura A.29
IMP_DEBE_VISTA_CT	0,19	0,19	Figura A.30
NUM_RECIBOS_SM_CT	0,19	0,19	Figura A.31
IMP_RECIBOS_SM_CT	0,19	0,19	Figura A.32
IMP_CAJ_ABONO_TJDB_3M	-0,19	0,19	Figura A.33
OP_HABER_VISTA_CT	0,18	0,18	Figura A.34
IMP_HABER_VISTA_CT	0,18	0,18	Figura A.35
OP_CAJ_ABONO_TJDB_3M	-0,18	0,18	Figura A.36
IMP_PENSION_SUA	0,17	0,17	Figura A.37
IMP_PENSION	0,17	0,17	Figura A.38
SALDO_TARJ_CREDITO_PT	0,17	0,17	Figura A.39
RENTA_ESTIMADA	0,17	0,17	Figura A.40
MO_SEGUROS_PT	0,17	0,17	Figura A.41
MO_SEGUROS_CT	0,17	0,17	Figura A.42
SALDO_MIN_VISTA_CT	0,16	0,16	Figura A.43
SALDO_PDTE_PT MOS_GR_VIV_CT	0,16	0,16	Figura A.44
CUOTA_PT MOS_GR_VIV_CT	0,16	0,16	Figura A.45
TIPOMEDIO_PT MOS_GR_VIV_CT	0,16	0,16	Figura A.46
IMP_RECIBOS_COMUNIDAD_4M_CT	0,15	0,15	Figura A.47
IMP_CAJ_ABONO_TJDB	-0,15	0,15	Figura A.48
IMP_RECIBOS_RESTO_4M_CT	0,14	0,14	Figura A.49

Continúa en la siguiente página

Cuadro A.2 – continuado de la página anterior

Variable	Correlación	ABS(Correlación)	Gráfico
SALDO_FINAL_VISTA_CT	0,14	0,14	Figura A.50
SALDO_MEDIO_VISTA_CT	0,14	0,14	Figura A.51
IMP_RECIBOS_COMUNIDAD_CT	0,14	0,14	Figura A.52
SALDO_MEDIO_IPFS_CORTO_CT	0,14	0,14	Figura A.53
SALDO_FINAL_IPFS_CORTO_CT	0,14	0,14	Figura A.54
MO_TARJETAS_PT	0,14	0,14	Figura A.55
MO_FB_CT	0,14	0,14	Figura A.56
MO_TARJETAS_CT	0,14	0,14	Figura A.57
SALDO_MEDIO_TARJETAS_PT	0,14	0,14	Figura A.58
SALDO_MEDIO_TARJETAS_CT	0,14	0,14	NA
OP_CAJ_ABONO_TJDB	-0,14	0,14	NA
NUM_RECIBOS_COMUNIDAD_CT	0,13	0,13	NA
NUM_RECIBOS_RESTO_CT	0,13	0,13	NA
IMP_RECIBOS_RESTO_CT	0,13	0,13	NA
SALDO_DISP_TARJ_CREDITO_PT	0,13	0,13	NA
TIPOMEDIO_PT MOS_GR_VIV_PT	0,13	0,13	NA
NUM_SEG_VIDA_V	0,13	0,13	NA
PRIMA_SEG_VIDA_V	0,13	0,13	NA
MO_FB_PT	0,13	0,13	NA
SALDO_MEDIO_PASIVO_CT	0,13	0,13	NA
IMP_TPV_IMPOSICION_TJCR_3M	0,12	0,12	NA
NUM_CONTRATA_ACTIVOS	0,12	0,12	NA
NUM_CONTRATA_FBALANCE	0,12	0,12	NA

Continúa en la siguiente página

Cuadro A.2 – continuado de la página anterior

Variable	Correlación	ABS(Correlación)	Gráfico
SALDO_MAX_VISTA_CT	0,12	0,12	NA
NUM_RECIBOS_AGUA_CT	0,12	0,12	NA
NUM_RECIBOS_IMPUESTOS_CT	0,12	0,12	NA
IMP_RECIBOS_AGUA_CT	0,12	0,12	NA
IMP_RECIBOS_IMPUESTOS_CT	0,12	0,12	NA
SALDO_PDTE_PTMOS_GR_VIV_PT	0,12	0,12	NA
CUOTA_PTMOS_GR_VIV_PT	0,12	0,12	NA
SALDO_MEDIO_FB_CT	0,12	0,12	NA
IMP_RECIBOS_EF_4M_CT	0,11	0,11	NA
OP_TPV_IMPOSICION_TJCR_3M	0,11	0,11	NA
SALDO_MEDIO_IPFS_CORTO_PT	0,11	0,11	NA
SALDO_FINAL_IPFS_CORTO_PT	0,11	0,11	NA
NUM_SEG_OTROS	0,11	0,11	NA
PRIMA_SEG_OTROS	0,11	0,11	NA
SALDO_MEDIO_FB_PT	0,11	0,11	NA
OP_TPV_IMPOSICION_TJCR	0,1	0,1	NA
IMP_TPV_IMPOSICION_TJCR	0,1	0,1	NA
OP_TPV_IMPOSICION_TJDB_3M	-0,1	0,1	NA
IMP_TPV_IMPOSICION_TJDB_3M	-0,1	0,1	NA
SALDO_PDTE_PTMOS_GP_CON_CT	0,098	0,098	NA
CUOTA_PTMOS_GP_CON_CT	0,098	0,098	NA
TIPOMEDIO_PTMOS_GP_CON_CT	0,098	0,098	NA
NUM_SEG_HOGAR_V	0,097	0,097	NA

Continúa en la siguiente página

Cuadro A.2 – continuado de la página anterior

Variable	Correlación	ABS(Correlación)	Gráfico
PRIMA_SEG_HOGAR_V	0,097	0,097	NA
IMP_RECIBOS_EF_CT	0,092	0,092	NA
NUM_RECIBOS_EF_CT	0,091	0,091	NA
IMP_NOMINA_PURO_SUA	0,088	0,088	NA
NUM_SEG_VIDA	0,086	0,086	NA
PRIMA_SEG_VIDA	0,086	0,086	NA
IDEAL	0,084	0,084	NA
SALDO_PDTE_PT MOS_GP_CON_PT	0,082	0,082	NA
CUOTA_PT MOS_GP_CON_PT	0,082	0,082	NA
TIPOMEDIO_PT MOS_GP_CON_PT	0,082	0,082	NA
IMP_CAJ_IMPOSICION_TJCR_3M	0,08	0,08	NA
SALDO_MEDIO_FOND_RIESGO_CT	0,08	0,08	NA
SALDO_FINAL_FOND_RIESGO_CT	0,08	0,08	NA
OP_CAJ_IMPOSICION_TJCR_3M	0,079	0,079	NA
IMP_OP_HABER_MOVIL	-0,079	0,079	NA
IMP_NOMINA	0,078	0,078	NA
RECORRIDO	0,078	0,078	NA
IMP_CAJ_IMPOSICION_TJDB_3M	-0,078	0,078	NA
OP_HABER_MOVIL	-0,078	0,078	NA
IMP_TPV_IMPOSICION_TJDB	-0,078	0,078	NA
NUM_RECIBOS_SS_CT	0,077	0,077	NA
NUM_CONTRATA_OT_SERV	-0,077	0,077	NA
IMP_RECIBOS_SS_4M_CT	0,075	0,075	NA

Continúa en la siguiente página

Cuadro A.2 – continuado de la página anterior

Variable	Correlación	ABS(Correlación)	Gráfico
IMP_RECIBOS_SS_CT	0,075	0,075	NA
OP_TPV_IMPOSICION_TJDB	-0,074	0,074	NA
CUOTA_PTAMOS_GP_INV_CT	0,073	0,073	NA
OP_CAJ_IMPOSICION_TJDB.3M	-0,073	0,073	NA
SALDO_PDTE_PTAMOS_GP_INV_CT	0,072	0,072	NA
TIPOMEDIO_PTAMOS_GP_INV_CT	0,072	0,072	NA
SALDO_MEDIO_PLANES_PT	0,071	0,071	NA
SALDO_FINAL_PLANES_PT	0,071	0,071	NA
SALDO_MEDIO_PLANES_CT	0,071	0,071	NA
SALDO_FINAL_PLANES_CT	0,071	0,071	NA
IMP_OP_DEBE_MOVIL	-0,068	0,068	NA
OP_TPV_ABONO_TJCR.3M	0,066	0,066	NA
SALDO_MEDIO_FOND_RIESGO_PT	0,063	0,063	NA
SALDO_FINAL_FOND_RIESGO_PT	0,063	0,063	NA
TIPO_MEDIO_IPFS_CORTO_PT	0,063	0,063	NA
SALDO_MEDIO_IPFS_LARGO_CT	0,063	0,063	NA
SALDO_PDTE_PTAMOS_GR_INV_CT	0,063	0,063	NA
CUOTA_PTAMOS_GR_INV_CT	0,063	0,063	NA
TIPOMEDIO_PTAMOS_GR_INV_CT	0,063	0,063	NA
IMP_TPV_ABONO_TJCR.3M	0,062	0,062	NA
TIPO_MEDIO_IPFS_CORTO_CT	0,062	0,062	NA
IMP_NOMINA_SUA	0,061	0,061	NA
OP_CAJ_IMPOSICION_TJCR	0,061	0,061	NA

Continúa en la siguiente página

Cuadro A.2 – continuado de la página anterior

Variable	Correlación	ABS(Correlación)	Gráfico
IMP_CAJ_IMPOSICION_TJCR	0,061	0,061	NA
SALDO_FINAL_IPFS_LARGO_CT	0,061	0,061	NA
OP_DEBE_MOVIL	-0,061	0,061	NA
SALDO_MEDIO_VALORES_CT	0,059	0,059	NA
SALDO_FINAL_VALORES_CT	0,059	0,059	NA
NUM_SEG_HOGAR	0,059	0,059	NA
PRIMA_SEG_HOGAR	0,059	0,059	NA
IMP_OP_DEBE_OFICINA	0,057	0,057	NA
SALDO_PDTE_PTMOS_GP_INV_PT	0,056	0,056	NA
CUOTA_PTMOS_GP_INV_PT	0,056	0,056	NA
TIPOMEDIO_PTMOS_GP_INV_PT	0,056	0,056	NA
SALDO_MEDIO_VALORES_PT	0,054	0,054	NA
SALDO_PDTE_PTMOS_GP_NEG_CT	0,054	0,054	NA
TIPOMEDIO_PTMOS_GP_NEG_CT	0,054	0,054	NA
OP_DEBE_OFICINA	0,053	0,053	NA
SALDO_FINAL_VALORES_PT	0,053	0,053	NA
OP_TPV_ABONO_TJDB_3M	-0,053	0,053	NA
SALDO_MEDIO_SEGUROS_AHORRO_PT	0,052	0,052	NA
SALDO_FINAL_SEGUROS_AHORRO_PT	0,052	0,052	NA
SALDO_MEDIO_SEGUROS_AHORRO_CT	0,052	0,052	NA
SALDO_FINAL_SEGUROS_AHORRO_CT	0,052	0,052	NA
CUOTA_PTMOS_GP_NEG_CT	0,052	0,052	NA
IMP_TPV_ABONO_TJDB_3M	-0,052	0,052	NA

Continúa en la siguiente página

Cuadro A.2 – continuado de la página anterior

Variable	Correlación	ABS(Correlación)	Gráfico
SALDO_MAX_VISTA_PT	-0,052	0,052	NA
IMP_TRANSFERENCIAS_SUA	0,05	0,05	NA
OP_TPV_ABONO_TJCR	0,05	0,05	NA
NUM_SEG_COCHE	0,049	0,049	NA
IMP_CAJ_ABONO_TJCR_3M	0,048	0,048	NA
SALDO_MEDIO_IPFS_LARGO_PT	0,048	0,048	NA
PRIMA_SEG_COCHE	0,048	0,048	NA
OP_CAJ_ABONO_TJCR_3M	0,047	0,047	NA
IMP_TPV_ABONO_TJCR	0,047	0,047	NA
SALDO_PDTE_PTMOS_GP_VIV_CT	0,047	0,047	NA
CUOTA_PTMOS_GP_VIV_CT	0,047	0,047	NA
NUM_CONTRATA_CAPTACION	-0,047	0,047	NA
SALDO_FINAL_IPFS_LARGO_PT	0,046	0,046	NA
TIPOMEDIO_PTMOS_GP_VIV_CT	0,046	0,046	NA
SALDO_PDTE_PTMOS_GR_INV_PT	0,046	0,046	NA
CUOTA_PTMOS_GR_INV_PT	0,046	0,046	NA
TIPOMEDIO_PTMOS_GR_INV_PT	0,046	0,046	NA
SALDO_MEDIO_VISTA_PT	-0,046	0,046	NA
NUM_ENTRADAS_DESCUBIERTO	0,045	0,045	NA
IMP_CAJ_IMPOSICION_TJDB	-0,045	0,045	NA
SALDO_MEDIO_PASIVO_PT	-0,044	0,044	NA
SALDO_PDTE_PTMOS_GP_NEG_PT	0,042	0,042	NA
OP_HABER_VISTA_PT	0,041	0,041	NA

Continúa en la siguiente página

Cuadro A.2 – continuado de la página anterior

Variable	Correlación	ABS(Correlación)	Gráfico
TIPOMEDIO_PTMOS_GP_NEG_PT	0,041	0,041	NA
SALDO_PDTE_PTMOS_GR_NEG_CT	0,041	0,041	NA
CUOTA_PTMOS_GR_NEG_CT	0,041	0,041	NA
TIPOMEDIO_PTMOS_GR_NEG_CT	0,041	0,041	NA
SALDO_DISPUESTO_CREDITO_CT	0,041	0,041	NA
LIMITE_CREDITO_CT	0,041	0,041	NA
TIPO_MEDIO_CREDITO_CT	0,041	0,041	NA
CUOTA_PTMOS_GP_NEG_PT	0,04	0,04	NA
OP_CAJ_IMPOSICION_TJDB	-0,039	0,039	NA
OP_DEBE_VISTA_PT	0,038	0,038	NA
IMP_CAJ_ABONO_TJCR	0,038	0,038	NA
IMP_OP_HABER_OFICINA	-0,038	0,038	NA
OP_CAJ_ABONO_TJCR	0,037	0,037	NA
SALDO_PDTE_PTMOS_GP_VIV_PT	0,037	0,037	NA
CUOTA_PTMOS_GP_VIV_PT	0,037	0,037	NA
IMP_HABER_VISTA_PT	0,036	0,036	NA
TIPOMEDIO_PTMOS_GP_VIV_PT	0,036	0,036	NA
OP_HABER_OFICINA	-0,036	0,036	NA
IMP_RECIBOS_COLEGIOS_4M_CT	0,033	0,033	NA
NUM_RECIBOS_COLEGIOS_CT	0,033	0,033	NA
IMP_RECIBOS_COLEGIOS_CT	0,033	0,033	NA
SALDO_DISPUESTO_CREDITO_PT	0,032	0,032	NA
LIMITE_CREDITO_PT	0,032	0,032	NA

Continúa en la siguiente página

Cuadro A.2 – continuado de la página anterior

Variable	Correlación	ABS(Correlación)	Gráfico
TIPO_MEDIO_CREDITO_PT	0,032	0,032	NA
DIAS_DESCUBIERTO	0,032	0,032	NA
TIPO_MEDIO_IPFS_LARGO_CT	0,03	0,03	NA
SALDO_PDTE_PTMOS_GR_NEG_PT	0,03	0,03	NA
CUOTA_PTMOS_GR_NEG_PT	0,03	0,03	NA
TIPOMEDIO_PTMOS_GR_NEG_PT	0,03	0,03	NA
SALDO_FINAL_VISTA_PT	-0,03	0,03	NA
IMP_TPV_ABONO_TJDB	-0,03	0,03	NA
SALDO_MEDIO_FOND_CONS_CT	0,029	0,029	NA
SALDO_FINAL_FOND_CONS_CT	0,029	0,029	NA
NUM_CONTRATA_TARJETAS	-0,029	0,029	NA
OP_TPV_ABONO_TJDB	-0,029	0,029	NA
SALDO_PDTE_PTMOS_GR_CON_CT	0,028	0,028	NA
CUOTA_PTMOS_GR_CON_CT	0,028	0,028	NA
TIPOMEDIO_PTMOS_GR_CON_CT	0,028	0,028	NA
IMP_DEBE_VISTA_PT	0,027	0,027	NA
OP_DEBE_ON	0,024	0,024	NA
TIPO_MEDIO_IPFS_LARGO_PT	0,023	0,023	NA
IMP_OP_DEBE_ON	0,022	0,022	NA
SALDO_MEDIO_FOND_CONS_PT	0,022	0,022	NA
SALDO_FINAL_FOND_CONS_PT	0,022	0,022	NA
SALDO_PDTE_PTMOS_GR_CON_PT	0,021	0,021	NA
CUOTA_PTMOS_GR_CON_PT	0,021	0,021	NA

Continúa en la siguiente página

Cuadro A.2 – continuado de la página anterior

Variable	Correlación	ABS(Correlación)	Gráfico
TIPOMEDIO_PT MOS_GR_CON_PT	0,021	0,021	NA
SALDO_MEDIO_RESTO_PASIVO_CT	0,02	0,02	NA
SALDO_FINAL_RESTO_PASIVO_CT	0,02	0,02	NA
SALDO_DISPUESTO_OTROS_ACT_CT	0,02	0,02	NA
IMP_ORD_EXTRANJERO_SUA	0,017	0,017	NA
SALDO_MEDIO_RESTO_PASIVO_PT	0,016	0,016	NA
LIMITE_DTO_CT	0,016	0,016	NA
IMP_OP_HABER_ON	-0,016	0,016	NA
SALDO_FINAL_RESTO_PASIVO_PT	0,015	0,015	NA
SALDO_DISPUESTO_OTROS_ACT_PT	0,015	0,015	NA
OP_HABER_ON	-0,015	0,015	NA
NUM_TARIFA_PLANA	0,014	0,014	NA
LIMITE_DTO_PT	0,013	0,013	NA
SALDO_DISPUESTO_DTO_CT	0,013	0,013	NA
TIPO_MEDIO_DTO_CT	0,013	0,013	NA
TIPO_MEDIO_OTROS_ACT_CT	0,013	0,013	NA
IMP_RECIBOS_ALQUILER_4M_CT	0,011	0,011	NA
SALDO_MIN_VISTA_PT	-0,011	0,011	NA
IMP_DESEMPLEO	-0,011	0,011	NA
SALDO_DISPUESTO_DTO_PT	0,01	0,01	NA
LIMITE_OTROS_ACT_CT	0,01	0,01	NA
NUM_RECIBOS_ALQUILER_CT	0,0095	0,0095	NA
IMP_RECIBOS_ALQUILER_CT	0,0094	0,0094	NA

Continúa en la siguiente página

A.2. TABLA RESULTADOS DEL ANÁLISIS DE CORRELACIONES ENTRE CADA VARIABLE Y MO_CLIENTE.CT

Cuadro A.2 – continuado de la página anterior

Variable	Correlación	ABS(Correlación)	Gráfico
TIPO_MEDIO_RESTO_PASIVO_CT	0,0094	0,0094	NA
TIPO_MEDIO.DTO_PT	0,0094	0,0094	NA
LIMITE_OTROS_ACT_PT	0,0093	0,0093	NA
TIPO_MEDIO_OTROS_ACT_PT	0,009	0,009	NA
IMP_INCAPACIDAD_SUA	0,0086	0,0086	NA
SALDO_PRINCIPAL_AVALES_CT	0,0083	0,0083	NA
IMP_TRANS_PERIODICAS	-0,0074	0,0074	NA
TIPO_MEDIO_RESTO_PASIVO_PT	0,0073	0,0073	NA
IMP_INEM_PURO_SUA	-0,007	0,007	NA
OFLULT_CONTRATO_PT	-0,007	0,007	NA
IMP_INCAPACIDAD	0,0069	0,0069	NA
SALDO_PRINCIPAL_AVALES_PT	0,0062	0,0062	NA
OFLULT_CONTRATO_CT	-0,006	0,006	NA
DEUDA_PDTE	-0,0055	0,0055	NA
OP_DEBE_TELF	0,0039	0,0039	NA
IMP_OP_DEBE_TELF	0,0039	0,0039	NA
COMISIONES_MES_AVALES_CT	0,0039	0,0039	NA
IMP_ORD_EXTRANJERO	0,0035	0,0035	NA
COMISIONES_MES_AVALES_PT	0,0024	0,0024	NA
SALDO_MEDIO_SEGUROS_PT	0,0019	0,0019	NA
SALDO_MEDIO_SEGUROS_CT	0,0019	0,0019	NA
IMP_TARIFA_PLANA	0,0015	0,0015	NA
OP_HABER_TELF	0,0011	0,0011	NA

Continúa en la siguiente página

Cuadro A.2 – continuado de la página anterior

Variable	Correlación	ABS(Correlación)	Gráfico
IMP_OP_HABER_TELF	0,0011	0,0011	NA
SALDO_PDTE_PTMOS_GR_OTR_CT	0,00093	0,00093	NA
CUOTA_PTMOS_GR_OTR_CT	0,00093	0,00093	NA
TIPOMEDIO_PTMOS_GR_OTR_CT	0,00088	0,00088	NA
SALDO_PDTE_PTMOS_GR_OTR_PT	0,0007	0,0007	NA
CUOTA_PTMOS_GR_OTR_PT	0,0007	0,0007	NA
TIPOMEDIO_PTMOS_GR_OTR_PT	0,00065	0,00065	NA
CLIENTE_ID	NA	NA	NA
FECHA	NA	NA	NA
CARPETA	NA	NA	NA
SEGMENTO_ID_DINAMICO	NA	NA	NA
SEGMENTO_ID_ESTATICO	NA	NA	NA
SEXO	NA	NA	NA
EDAD	NA	NA	NA
AYUNTAMIENTO_ID	NA	NA	NA
PROVINCIA_ID	NA	NA	NA
PAIS_RESIDENCIA	NA	NA	NA
PAIS_NACIONALIDAD	NA	NA	NA
RELACION_CONTRACTUAL_ID	NA	NA	NA
PRODUCTO_FINANCIERO_ID_PT	NA	NA	NA
PRODUCTO_FINANCIERO_ID_CT	NA	NA	NA
SERVICIO_ID	NA	NA	NA
MESES_ANTIGUEDAD_CLIENTE	NA	NA	NA

Continúa en la siguiente página

Cuadro A.2 – continuado de la página anterior

Variable	Correlación	ABS(Correlación)	Gráfico
MESES_ANTIGUEDAD_ACTIVO	NA	NA	NA
MESES_ANTIGUEDAD_PASIVO	NA	NA	NA
MESES_ANTIGUEDAD_FBALANCE	NA	NA	NA
MESES_ANTIGUEDAD_SERVICIOS	NA	NA	NA
MESES_ANTIGUEDAD_INGRESOS	NA	NA	NA
ANTIGUEDAD_CLIENTE	NA	NA	NA
ANTIGUEDAD_ACTIVO	NA	NA	NA
ANTIGUEDAD_PASIVO	NA	NA	NA
ANTIGUEDAD_FBALANCE	NA	NA	NA
ANTIGUEDAD_SERVICIOS	NA	NA	NA
ANTIGUEDAD_INGRESOS	NA	NA	NA
ANTIGUEDAD	NA	NA	NA
TRANSICION_ESTADO_ID	NA	NA	NA
ESTADO_ANTERIOR_CLIENTE_ID	NA	NA	NA
ESTADO_ACTUAL_CLIENTE_ID	NA	NA	NA
OFICINA_1_CT	NA	NA	NA
TARIFA_PLANA	NA	NA	NA
OFICINA_2_CT	NA	NA	NA
OFICINA_3_CT	NA	NA	NA
OFICINA_OPERA_PT	NA	NA	NA
OFICINA_OPERA_CT	NA	NA	NA
TIPO_PERSONA	NA	NA	NA
FECHA_ALTA	NA	NA	NA

Continúa en la siguiente página

Cuadro A.2 – continuado de la página anterior

Variable	Correlación	ABS(Correlación)	Gráfico
FECHA_NACIMIENTO	NA	NA	NA
FECHA_ANTIGUEDAD_CLIENTE	NA	NA	NA
FECHA_ANTIGUEDAD_ACTIVADO	NA	NA	NA
FECHA_ANTIGUEDAD_PASIVO	NA	NA	NA
FECHA_ANTIGUEDAD_FBALANCE	NA	NA	NA
FECHA_ANTIGUEDAD_SERVICIOS	NA	NA	NA
FECHA_ANTIGUEDAD_INGRESOS	NA	NA	NA
FC_SITUACION	NA	NA	NA
FC_ACTIVACION	NA	NA	NA
FC_CAPTACION	NA	NA	NA
FC_BAJA	NA	NA	NA
IMP_RECIBOS_IVA_4M_CT	NA	NA	NA
NUM_QUEJAS	NA	NA	NA
NUM_RECLAMACIONES	NA	NA	NA
NUM_SESIONES_ON	NA	NA	NA
NUM_CONTACTOS_TELEFONICO	NA	NA	NA
NUM_CONTACTOS_MOVIL	NA	NA	NA
NUM_VISITAS_OFICINA	NA	NA	NA
NUM_RECIBOS_IVA_CT	NA	NA	NA
IMP_RECIBOS_IVA_CT	NA	NA	NA
SALDO_MEDIO_AHORRO_VIV_PT	NA	NA	NA
SALDO_FINAL_AHORRO_VIV_PT	NA	NA	NA
TIPO_MEDIO_AHORRO_VIV_PT	NA	NA	NA

Continúa en la siguiente página

A.2. TABLA RESULTADOS DEL ANÁLISIS DE CORRELACIONES ENTRE CADA VARIABLE Y MO_CLIENTE.CT

Cuadro A.2 – continuado de la página anterior

Variable	Correlación	ABS(Correlación)	Gráfico
SALDO_MEDIO_AHORRO_VIV_CT	NA	NA	NA
SALDO_FINAL_AHORRO_VIV_CT	NA	NA	NA
TIPO_MEDIO_AHORRO_VIV_CT	NA	NA	NA
SALDO_PDTE_PTMOS_GP_OTR_PT	NA	NA	NA
CUOTA_PTMOS_GP_OTR_PT	NA	NA	NA
TIPOMEDIO_PTMOS_GP_OTR_PT	NA	NA	NA
SALDO_PDTE_PTMOS_GP_OTR_CT	NA	NA	NA
CUOTA_PTMOS_GP_OTR_CT	NA	NA	NA
TIPOMEDIO_PTMOS_GP_OTR_CT	NA	NA	NA
FC_ULT_AGENDA	NA	NA	NA
FECHA_ULT_CONTRATO_PT	NA	NA	NA
FECHA_ULT_CONTRATO_CT	NA	NA	NA
REC_CONTRATA_CAPTACION	NA	NA	NA
REC_CONTRATA_ACTIVOS	NA	NA	NA
REC_CONTRATA_FBALANCE	NA	NA	NA
REC_CONTRATA_TARJETAS	NA	NA	NA
REC_CONTRATA_OT_SERV	NA	NA	NA
REC_CONTRATA_ING_REGULARES	NA	NA	NA
FC_ULT_DESCUBIERTO	NA	NA	NA

Cuadro A.2: Tabla de correlaciones.

A.3. Gráficos de dispersión entre las distintas variables y MO_CLIENTE_CT

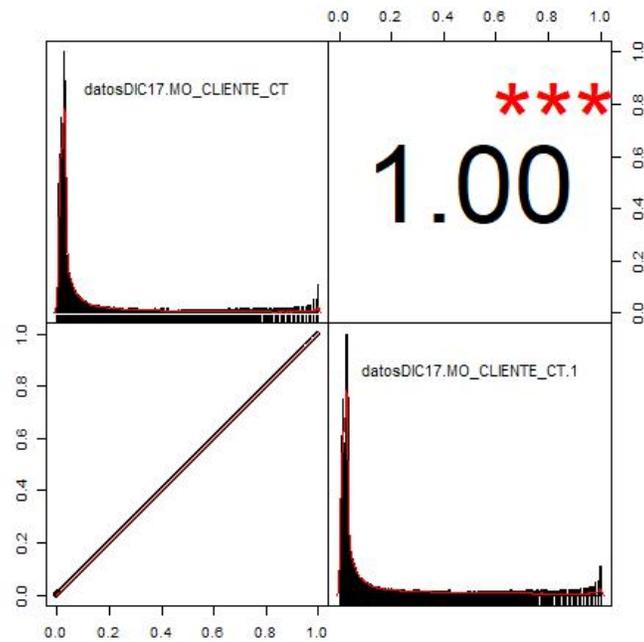


Figura A.1: Gráfico de la variable MO_CLIENTE_CT

A.3. GRÁFICOS DE DISPERSIÓN ENTRE LAS DISTINTAS VARIABLES Y MO_CLIENTE_CT89

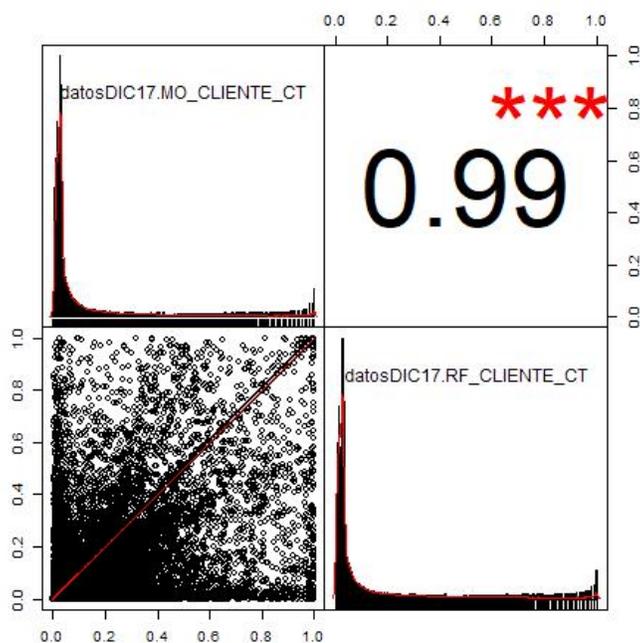


Figura A.2: Gráfico de la variable RF_CLIENTE_CT

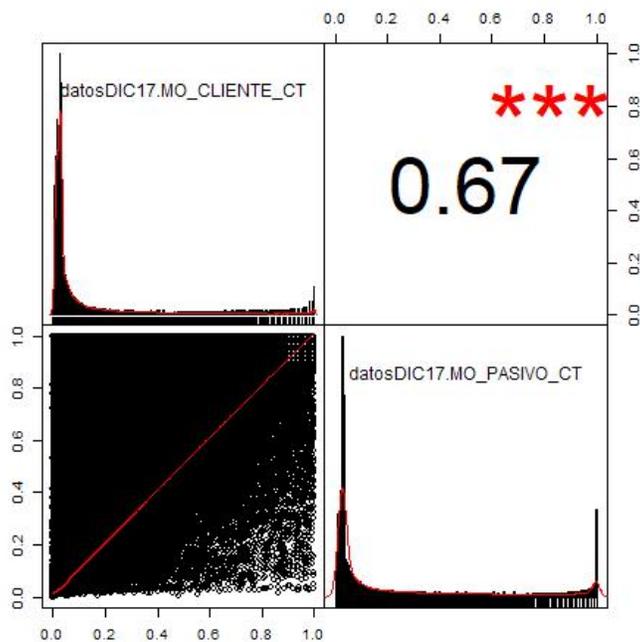


Figura A.3: Gráfico de la variable MO_PASIVO_CT

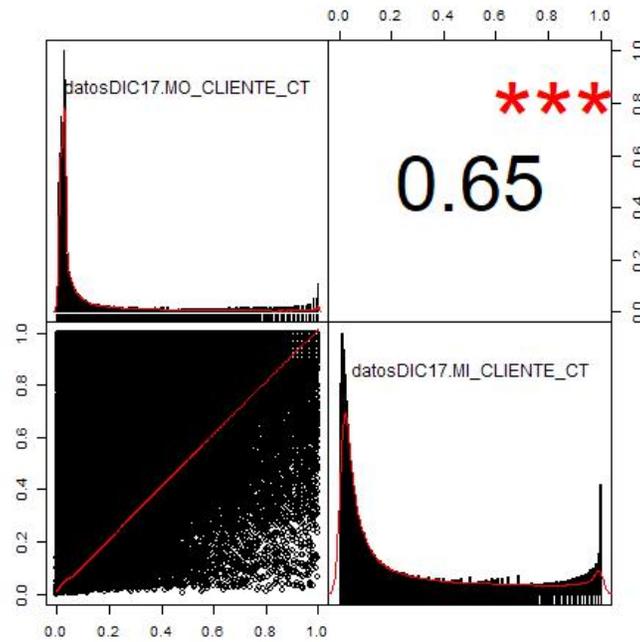


Figura A.4: Gráfico de la variable MI_CLIENTE_CT

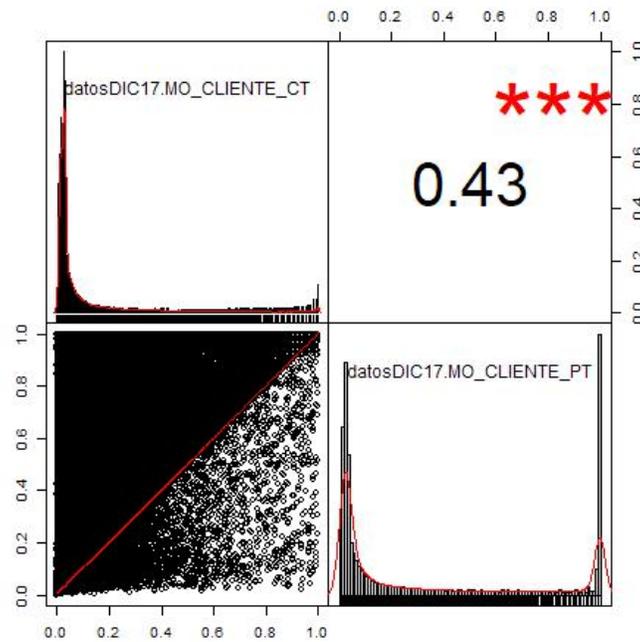


Figura A.5: Gráfico de la variable MO_CLIENTE_PT

A.3. GRÁFICOS DE DISPERSIÓN ENTRE LAS DISTINTAS VARIABLES Y MO_CLIENTE_CT91

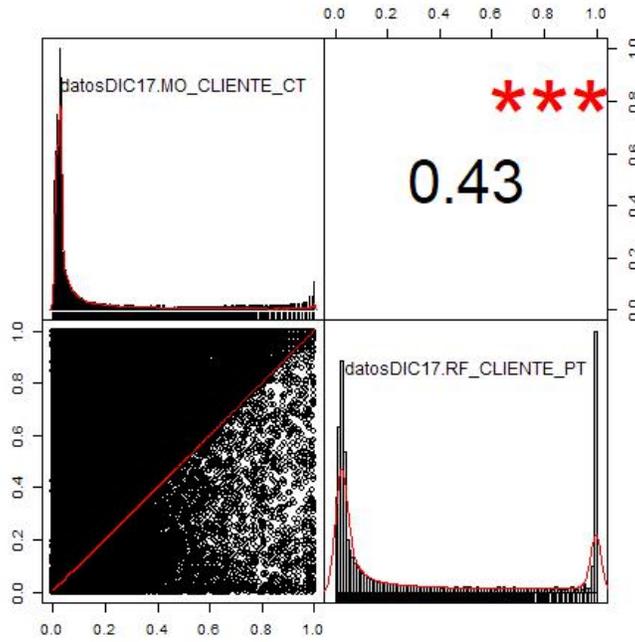


Figura A.6: Gráfico de la variable RF_CLIENTE_PT

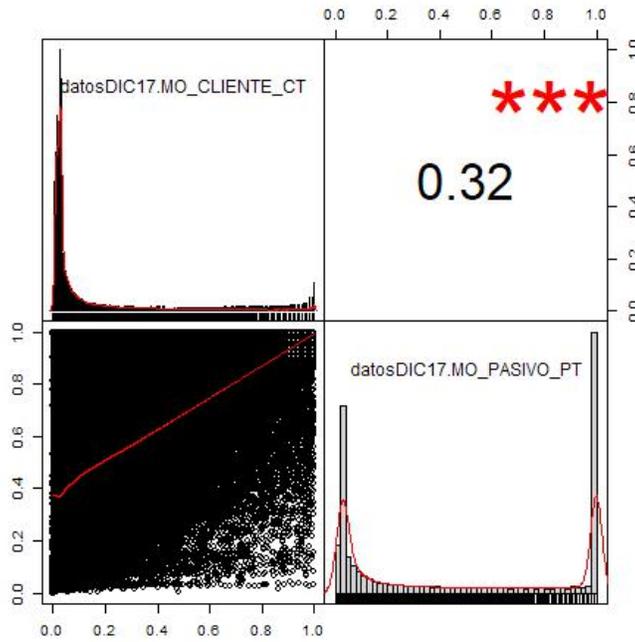
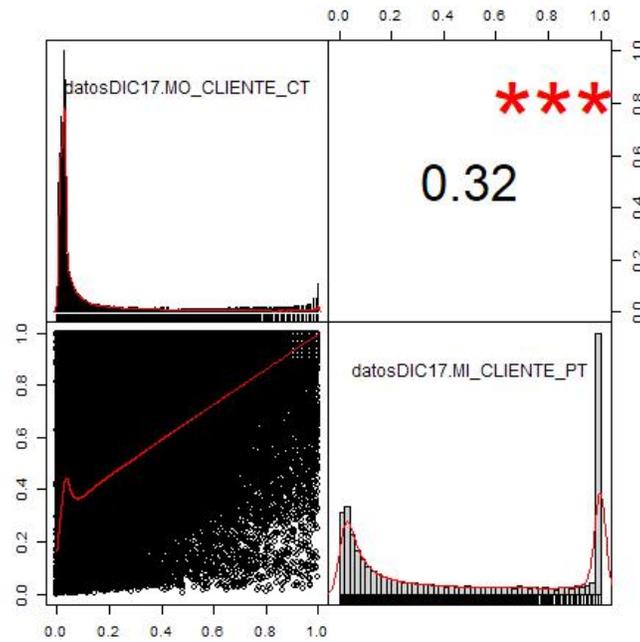
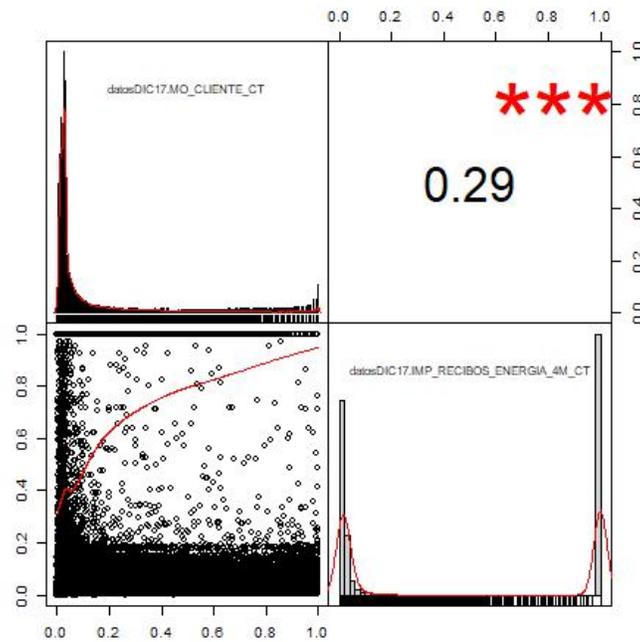


Figura A.7: Gráfico de la variable MO_PASIVO_PT

Figura A.8: Gráfico de la variable `MI_CLIENTE_PT`Figura A.9: Gráfico de la variable `IMP_RECIBOS_ENERGIA_4M_CT`

A.3. GRÁFICOS DE DISPERSIÓN ENTRE LAS DISTINTAS VARIABLES Y MO_CLIENTE_CT93

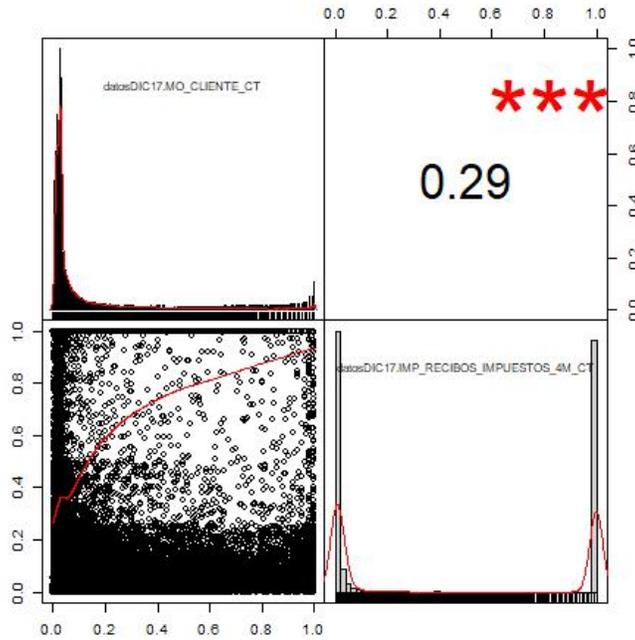


Figura A.10: Gráfico de la variable IMP_RECIBOS_IMPUESTOS_4M_CT

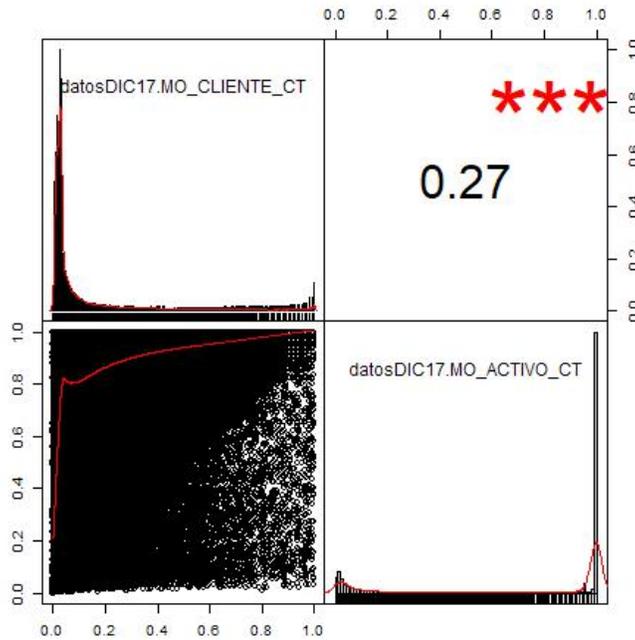


Figura A.11: Gráfico de la variable MO_ACTIVO_CT

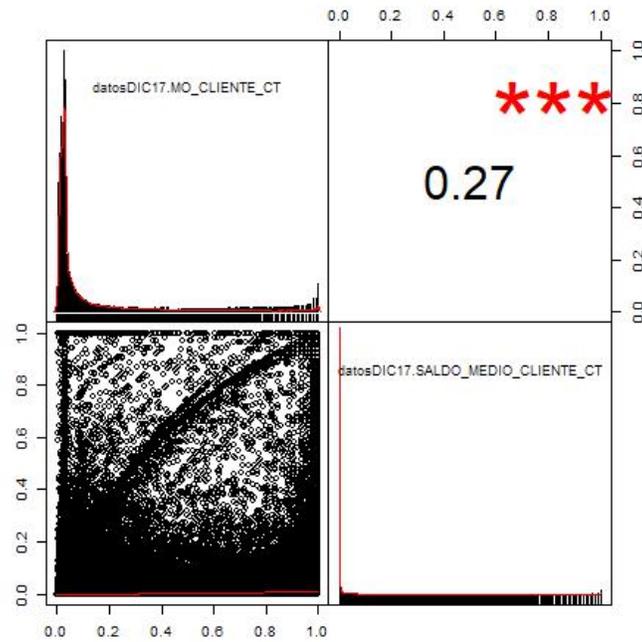


Figura A.12: Gráfico de la variable SALDO_MEDIO_CLIENTE_CT

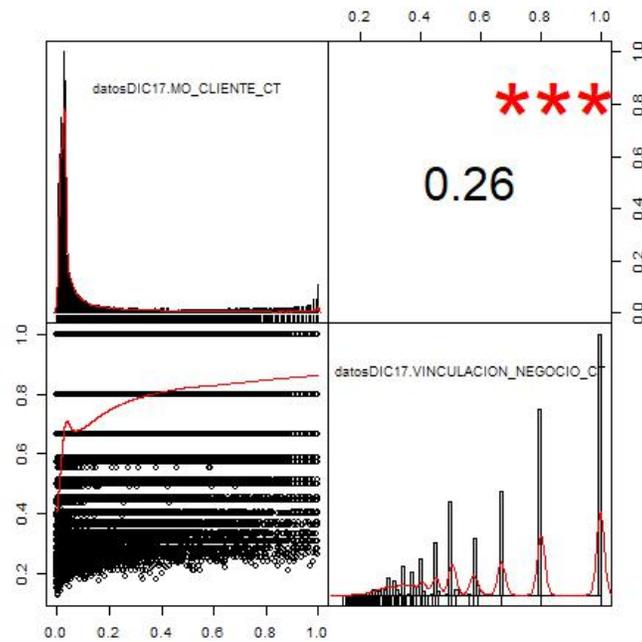


Figura A.13: Gráfico de la variable VINCULACION_NEGOCIO_CT

A.3. GRÁFICOS DE DISPERSIÓN ENTRE LAS DISTINTAS VARIABLES Y MO_CLIENTE_CT95

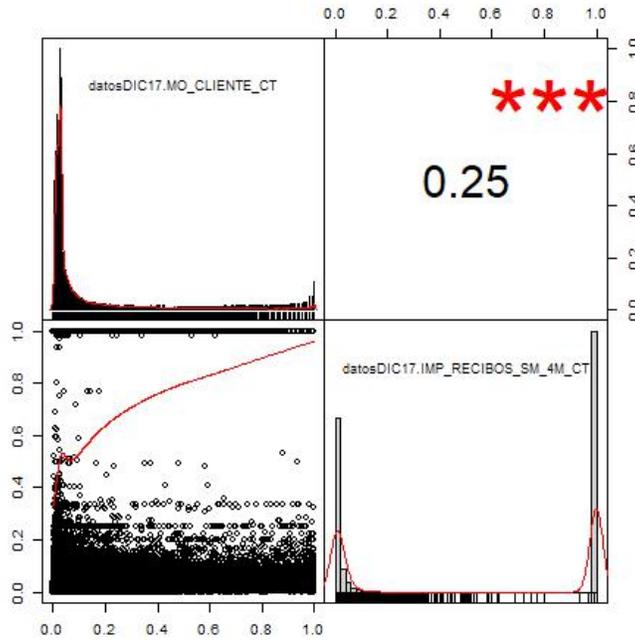


Figura A.14: Gráfico de la variable IMP_RECIBOS_SM_4M_CT

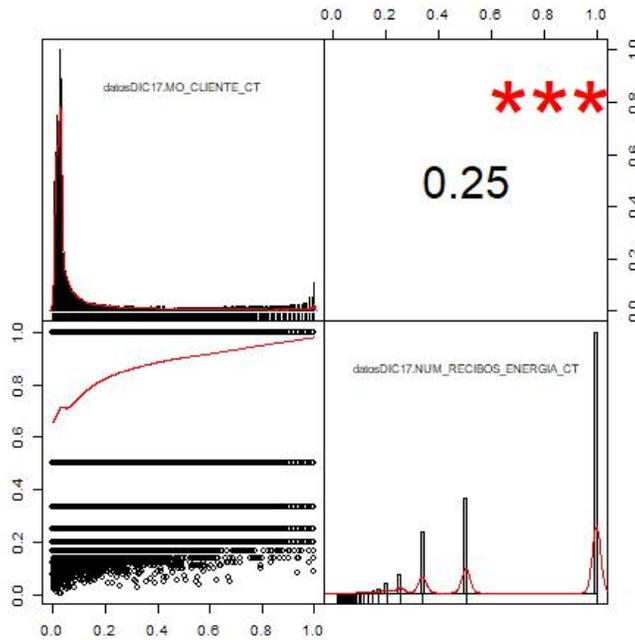
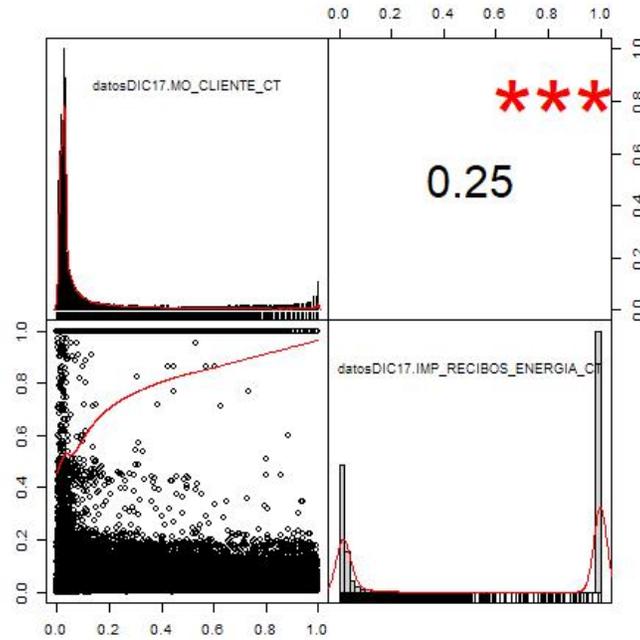
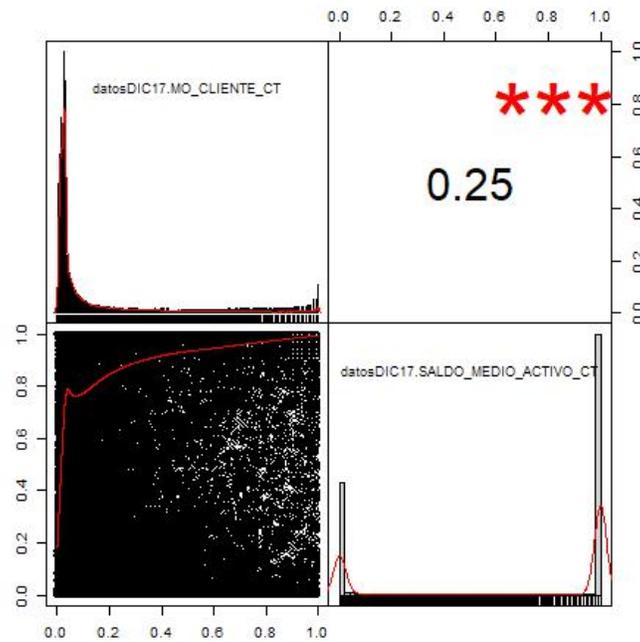


Figura A.15: Gráfico de la variable NUM_RECIBOS_ENERGIA_CT

Figura A.16: Gráfico de la variable `IMP_RECIBOS_ENERGIA_CT`Figura A.17: Gráfico de la variable `SALDO_MEDIO_ACTIVO_CT`

A.3. GRÁFICOS DE DISPERSIÓN ENTRE LAS DISTINTAS VARIABLES Y MO_CLIENTE_CT97

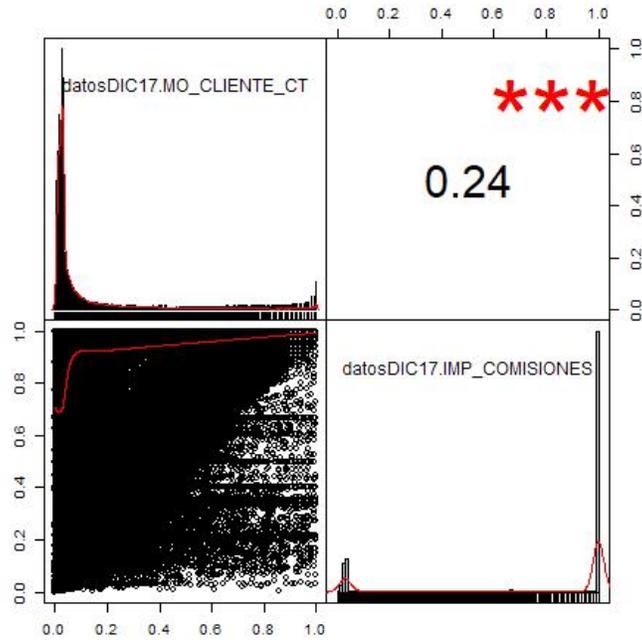


Figura A.18: Gráfico de la variable IMP_COMISIONES

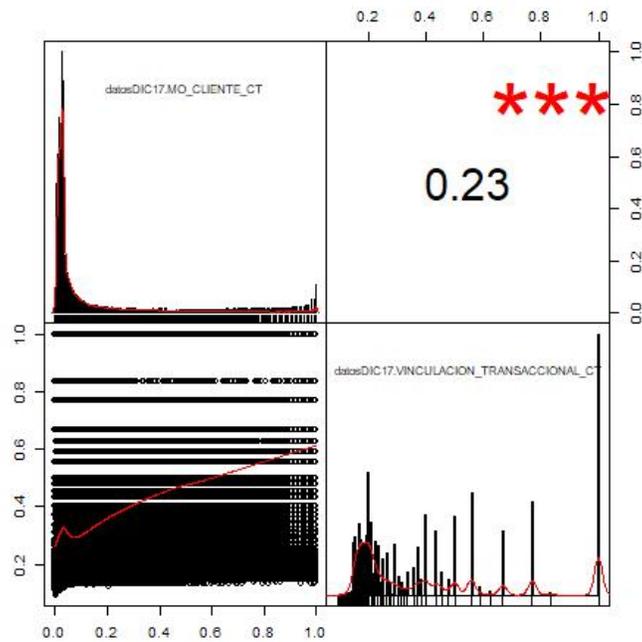


Figura A.19: Gráfico de la variable VINCULACION_TRANSACCIONAL_CT

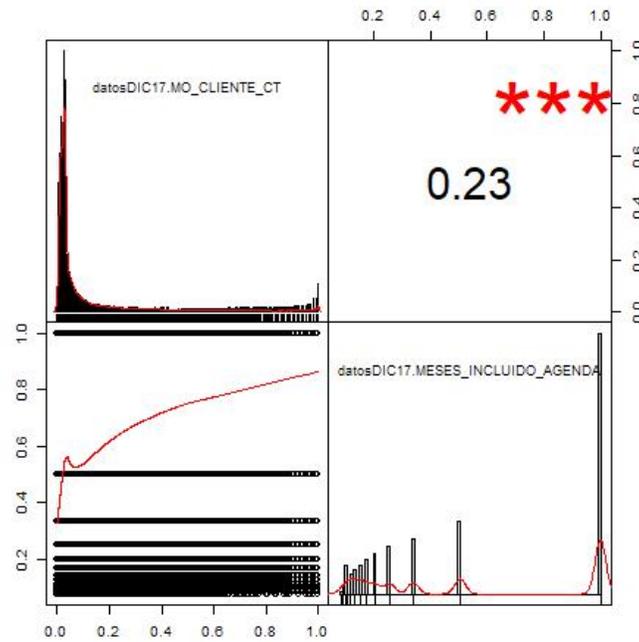


Figura A.20: Gráfico de la variable MESES_INCLUIDO_AGENDA

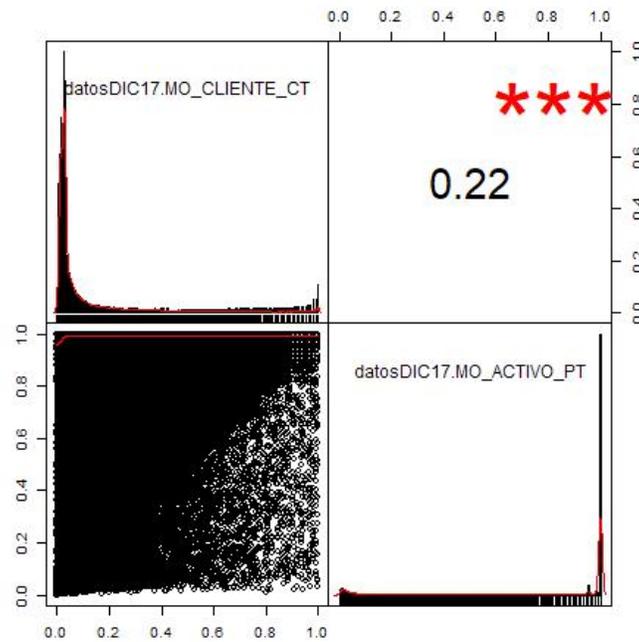


Figura A.21: Gráfico de la variable MO_ACTIVADO_PT

A.3. GRÁFICOS DE DISPERSIÓN ENTRE LAS DISTINTAS VARIABLES Y MO_CLIENTE_CT99

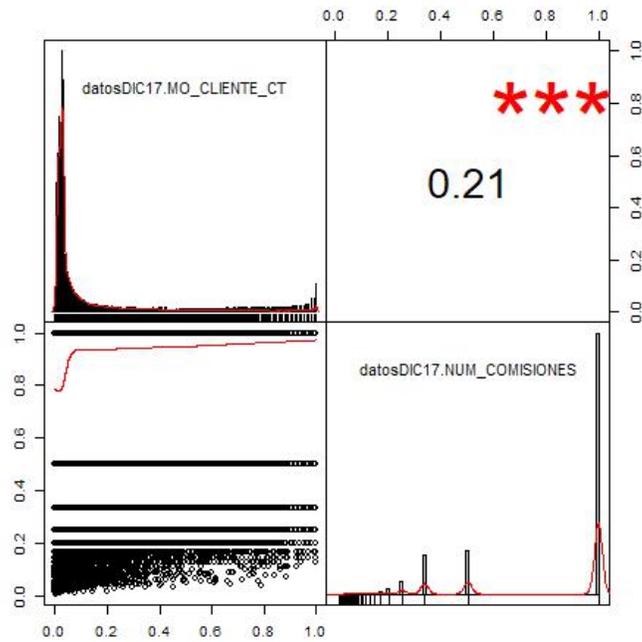


Figura A.22: Gráfico de la variable NUM_COMISIONES

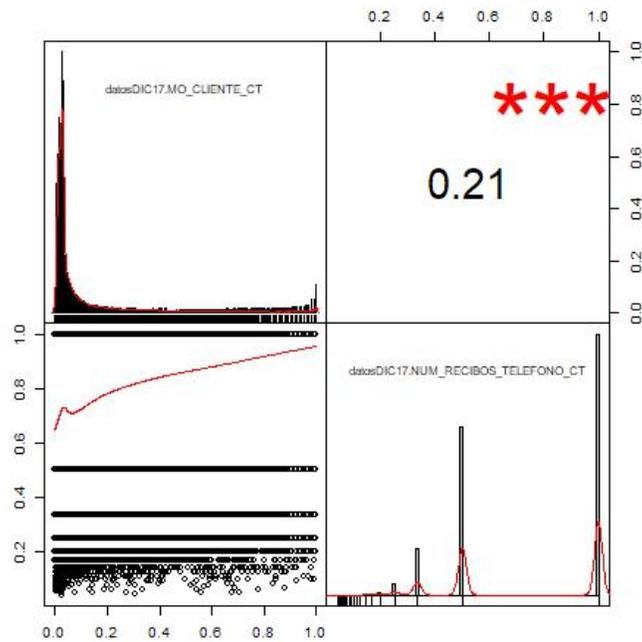
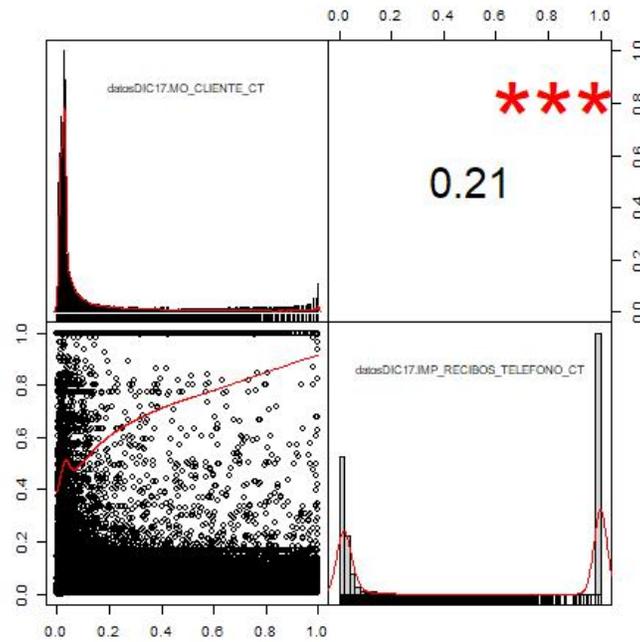
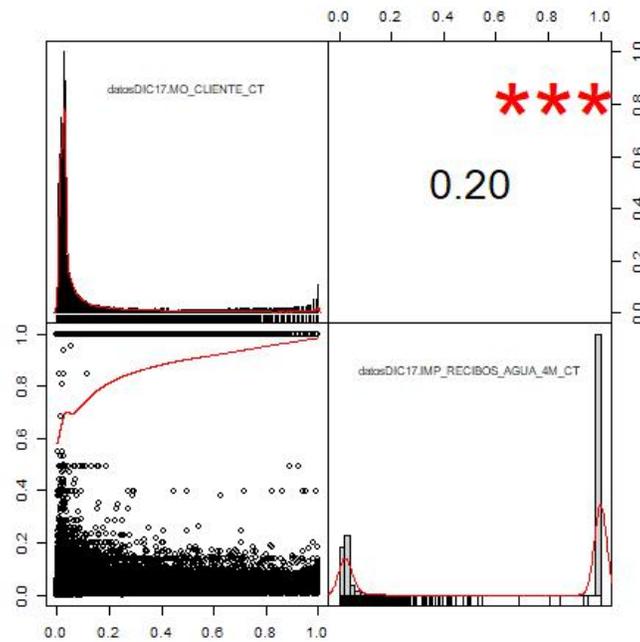


Figura A.23: Gráfico de la variable NUM_RECIBOS_TELEFONO_CT

Figura A.24: Gráfico de la variable `IMP_RECIBOS_TELEFONO_CT`Figura A.25: Gráfico de la variable `IMP_RECIBOS_AGUA_4M_CT`

A.3. GRÁFICOS DE DISPERSIÓN ENTRE LAS DISTINTAS VARIABLES Y MO_CLIENTE_CT101

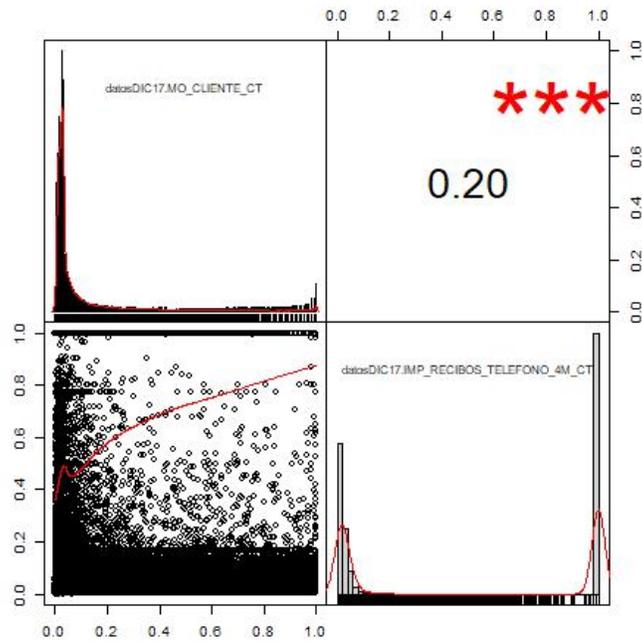


Figura A.26: Gráfico de la variable IMP_RECIBOS_TELEFONO_4M_CT

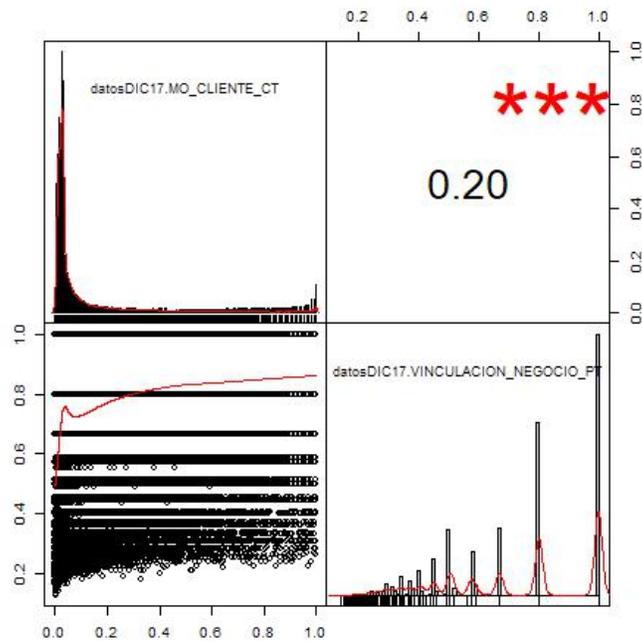


Figura A.27: Gráfico de la variable VINCULACION_NEGOCIO_PT

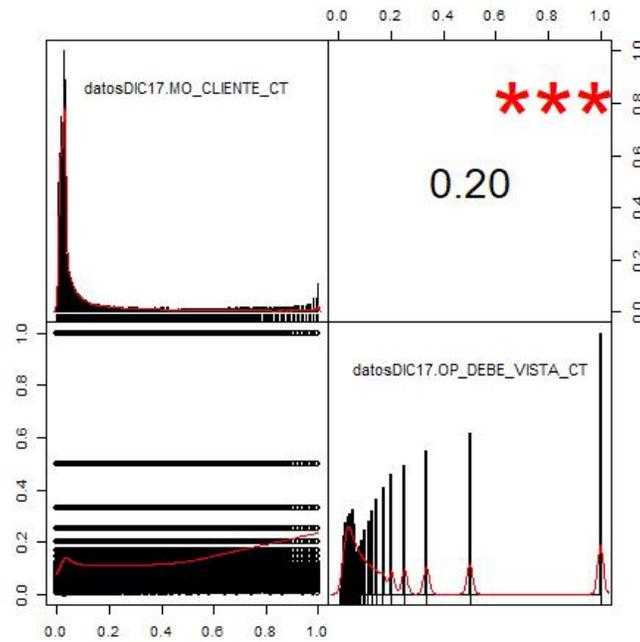


Figura A.28: Gráfico de la variable OP_DEBE_VISTA_CT

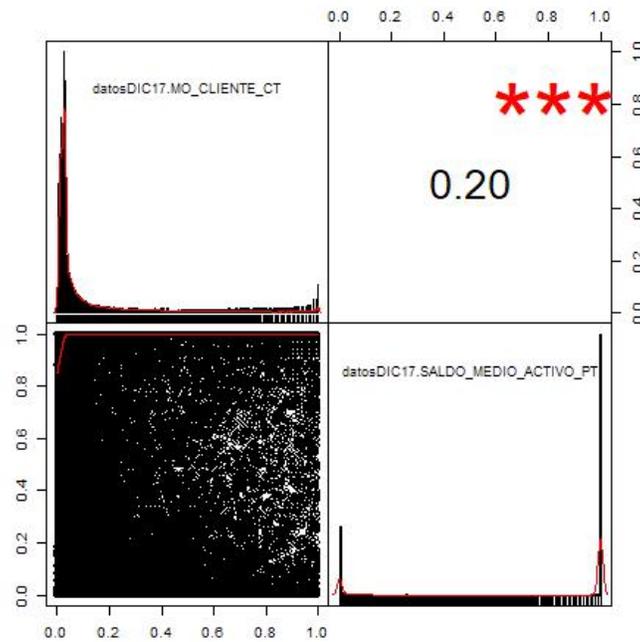


Figura A.29: Gráfico de la variable SALDO_MEDIO_ACTIVO_PT

A.3. GRÁFICOS DE DISPERSIÓN ENTRE LAS DISTINTAS VARIABLES Y MO_CLIENTE_CT103

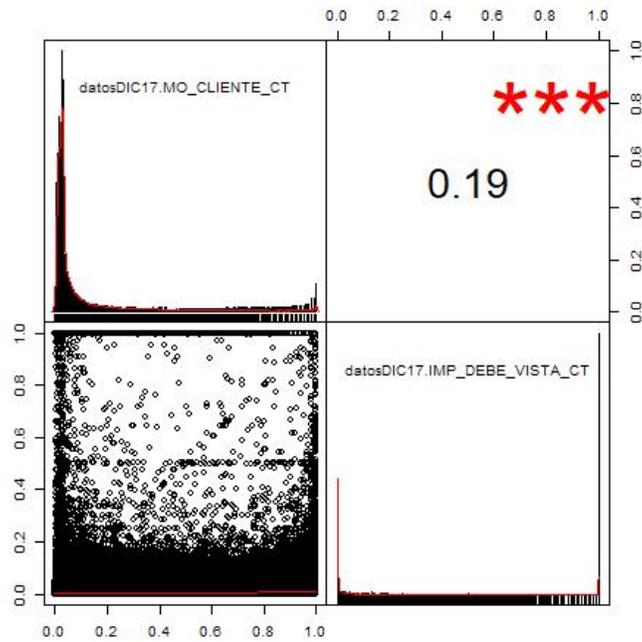


Figura A.30: Gráfico de la variable IMP_DEBE_VISTA_CT

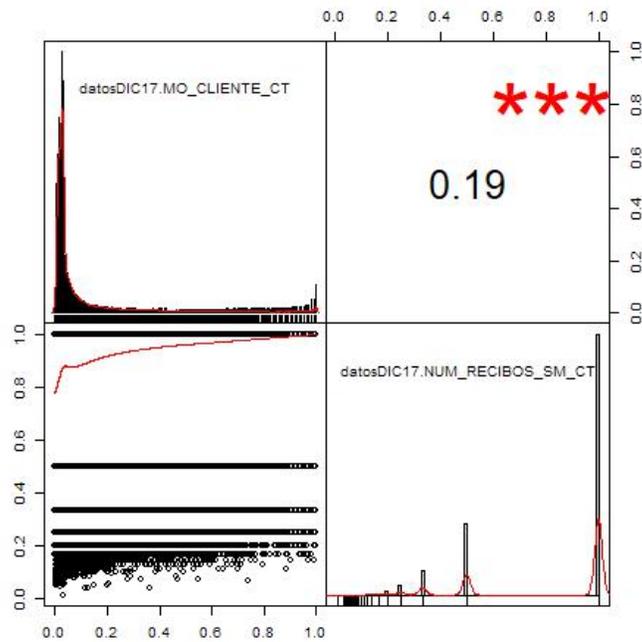


Figura A.31: Gráfico de la variable NUM_RECIBOS_SM_CT

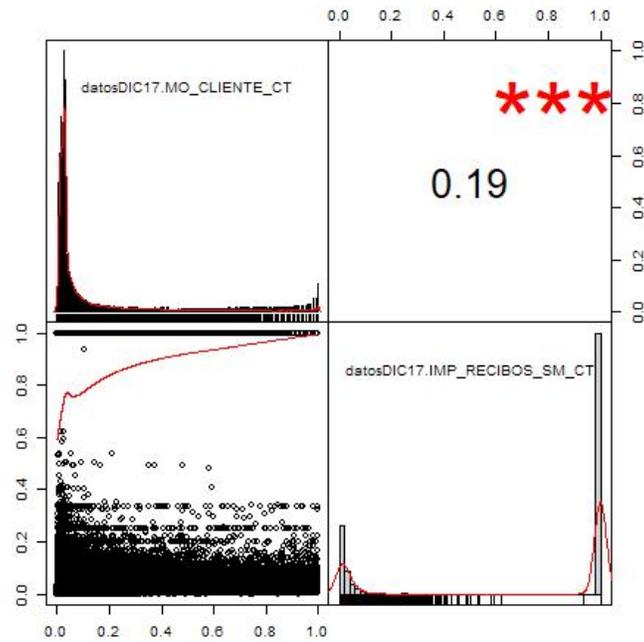


Figura A.32: Gráfico de la variable IMP_RECIBOS_SM_CT

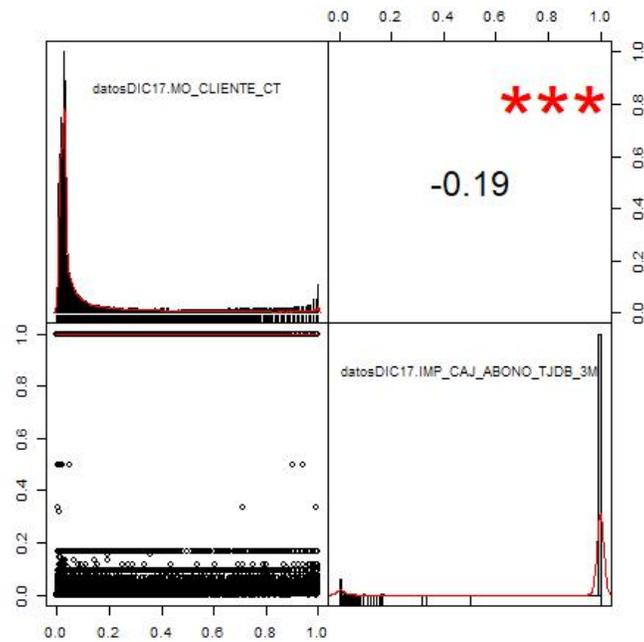


Figura A.33: Gráfico de la variable IMP_CAJ_ABONO_TJDB_3M

A.3. GRÁFICOS DE DISPERSIÓN ENTRE LAS DISTINTAS VARIABLES Y MO_CLIENTE_CT105

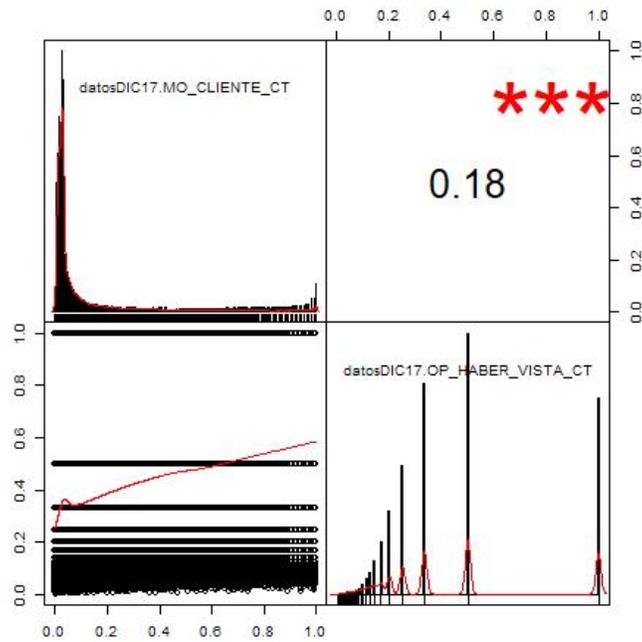


Figura A.34: Gráfico de la variable OP_HABER_VISTA_CT

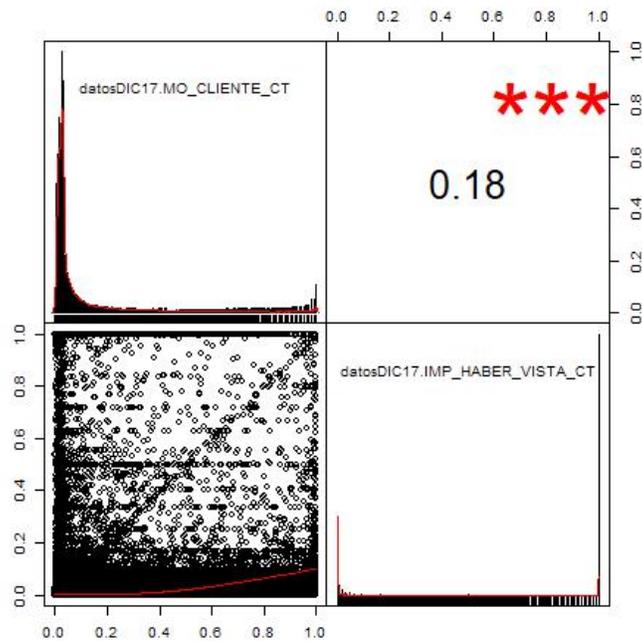


Figura A.35: Gráfico de la variable IMP_HABER_VISTA_CT

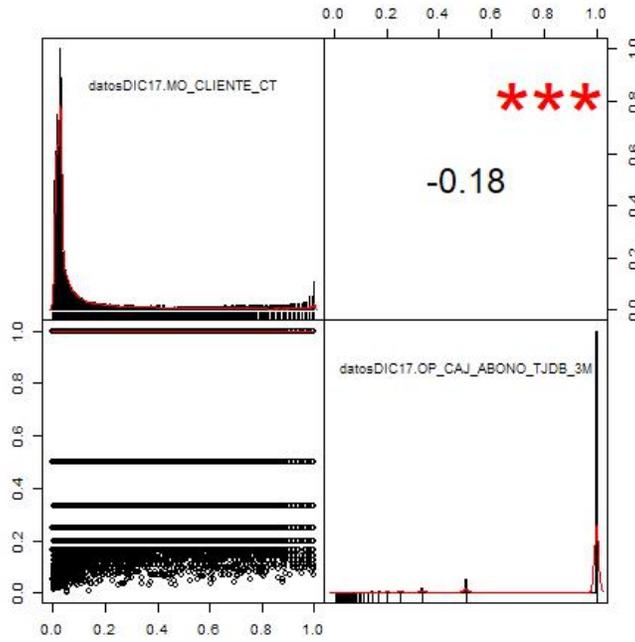


Figura A.36: Gráfico de la variable `OP_CAJ_ABONO_TJDB_3M`

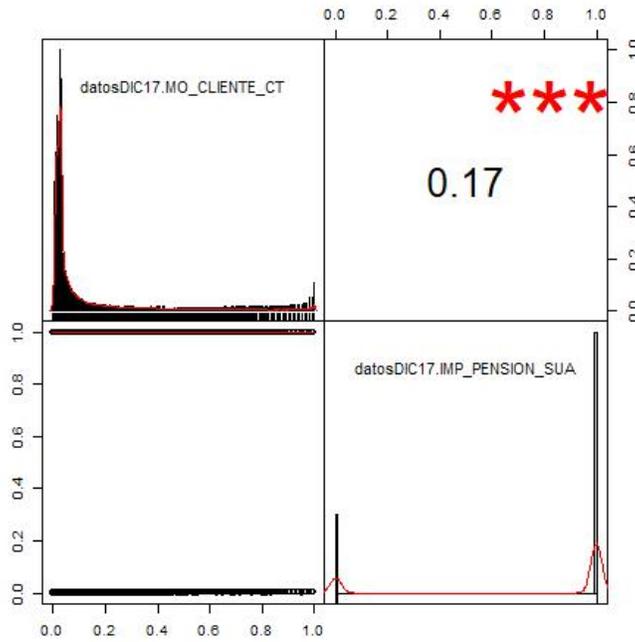


Figura A.37: Gráfico de la variable `IMP_PENSION_SUA`

A.3. GRÁFICOS DE DISPERSIÓN ENTRE LAS DISTINTAS VARIABLES Y MO_CLIENTE_CT107

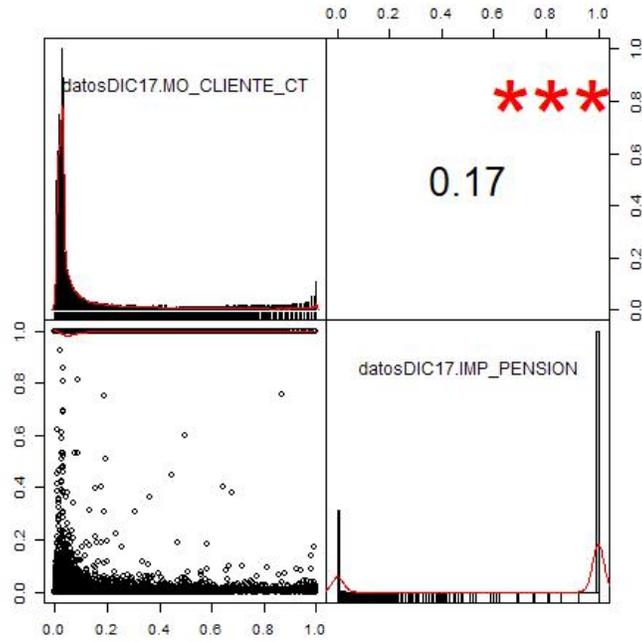


Figura A.38: Gráfico de la variable IMP_PENSION

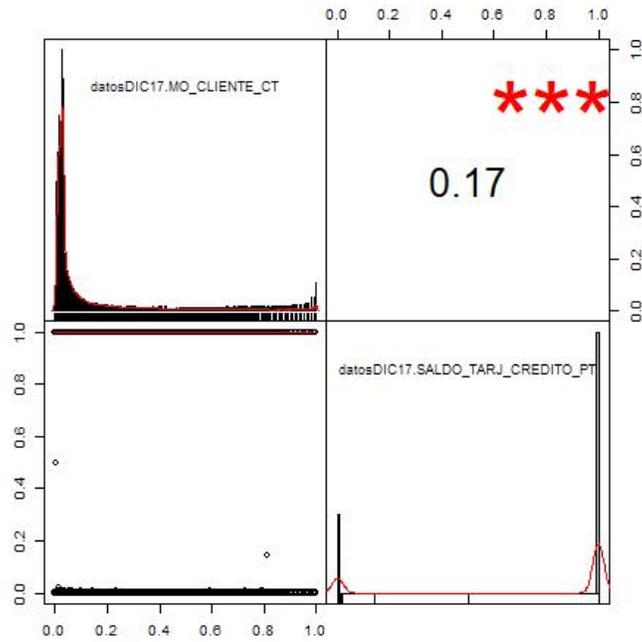


Figura A.39: Gráfico de la variable SALDO_TARJ_CREDITO_PT

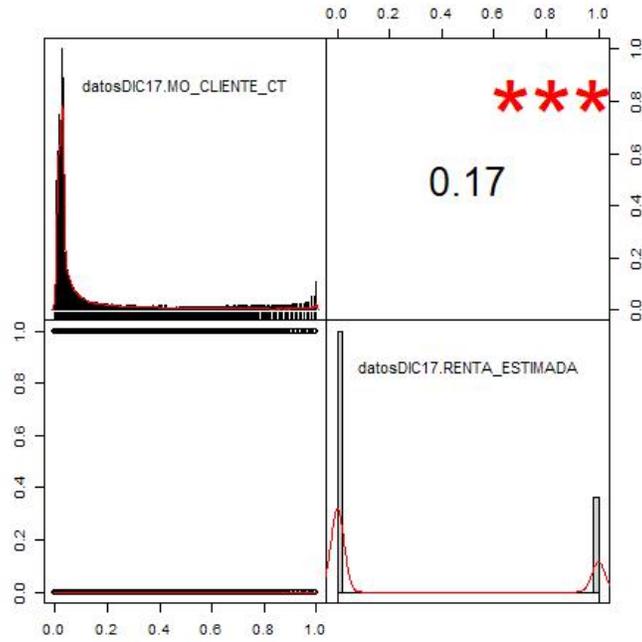


Figura A.40: Gráfico de la variable RENTA_ESTIMADA

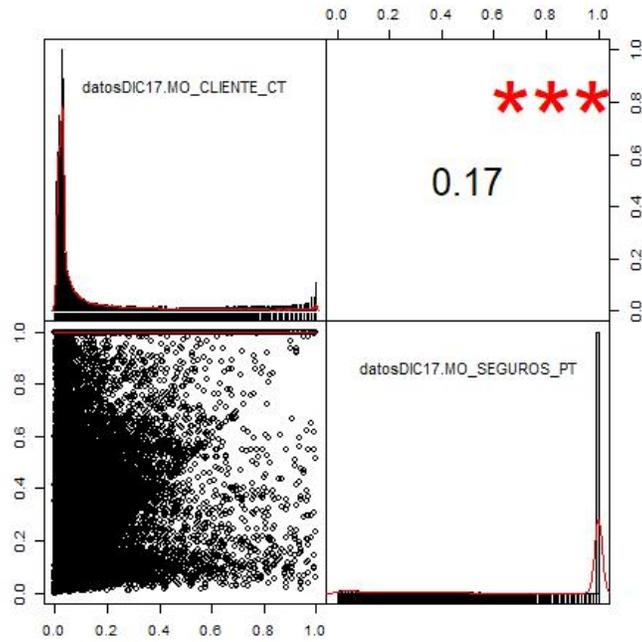


Figura A.41: Gráfico de la variable MO_SEGUROS_PT

A.3. GRÁFICOS DE DISPERSIÓN ENTRE LAS DISTINTAS VARIABLES Y MO_CLIENTE_CT109

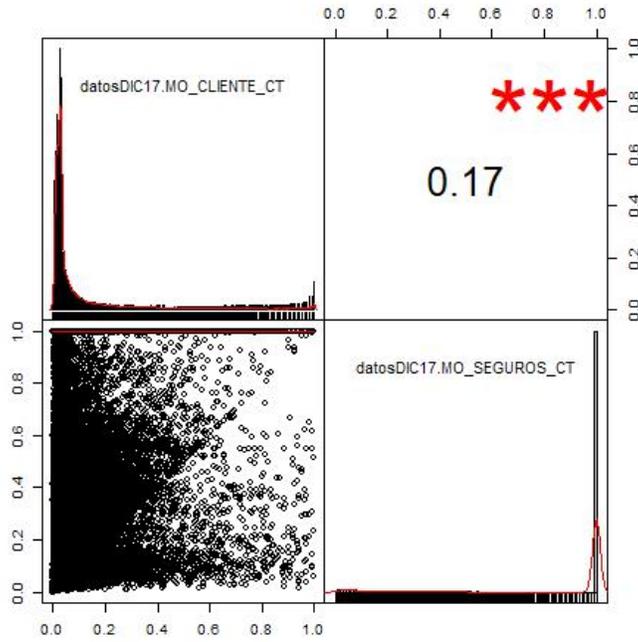


Figura A.42: Gráfico de la variable MO_SEGUROS_CT

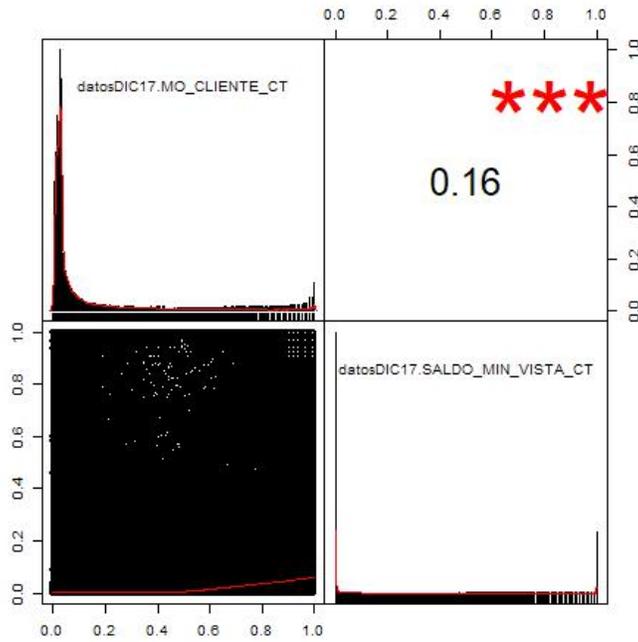


Figura A.43: Gráfico de la variable SALDO_MIN_VISTA_CT

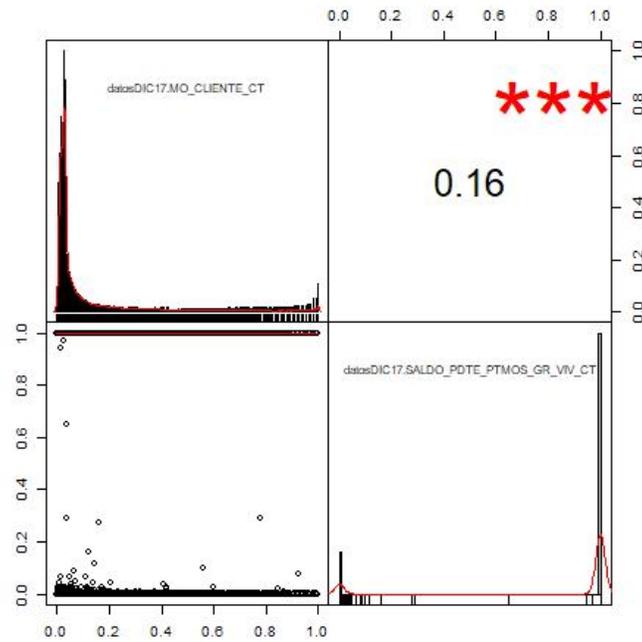


Figura A.44: Gráfico de la variable SALDO_PDTE_PTMOS_GR_VIV_CT

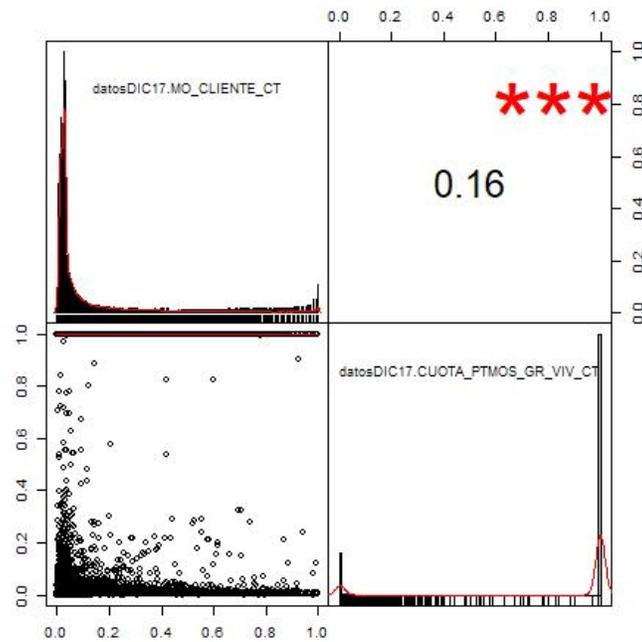


Figura A.45: Gráfico de la variable CUOTA_PTMOS_GR_VIV_CT

A.3. GRÁFICOS DE DISPERSIÓN ENTRE LAS DISTINTAS VARIABLES Y MO_CLIENTE_CT111

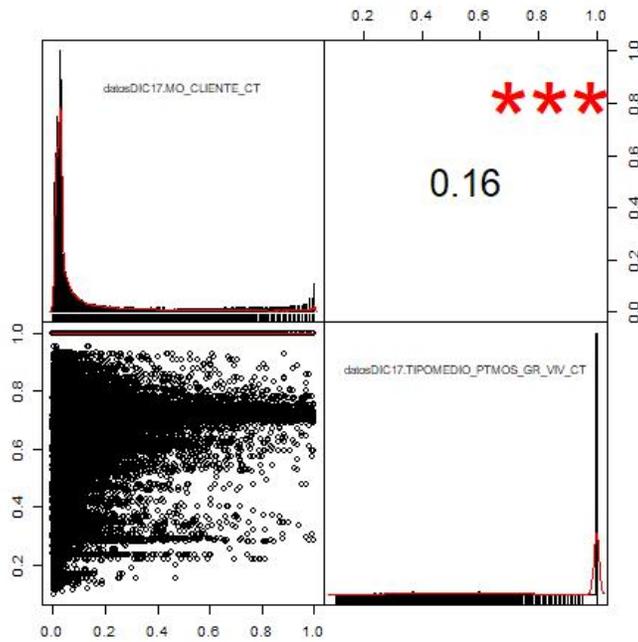


Figura A.46: Gráfico de la variable TIPOMEDIO_PTAMOS_GR_VIV_CT

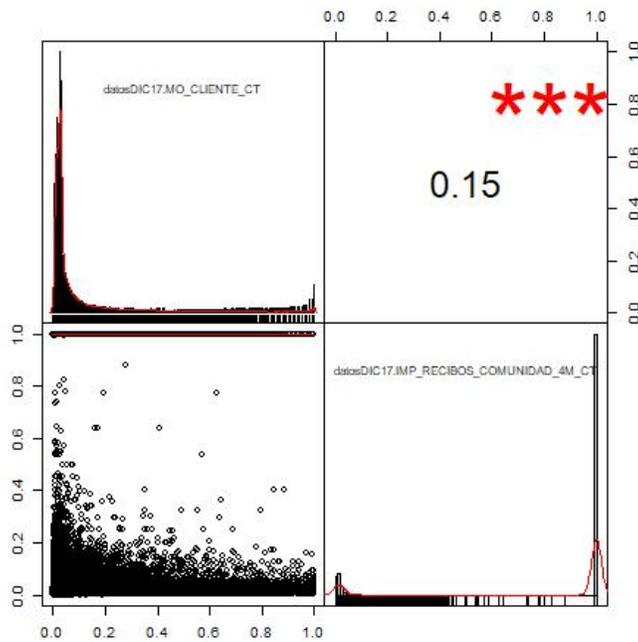


Figura A.47: Gráfico de la variable IMP_RECIBOS_COMUNIDAD_4M_CT

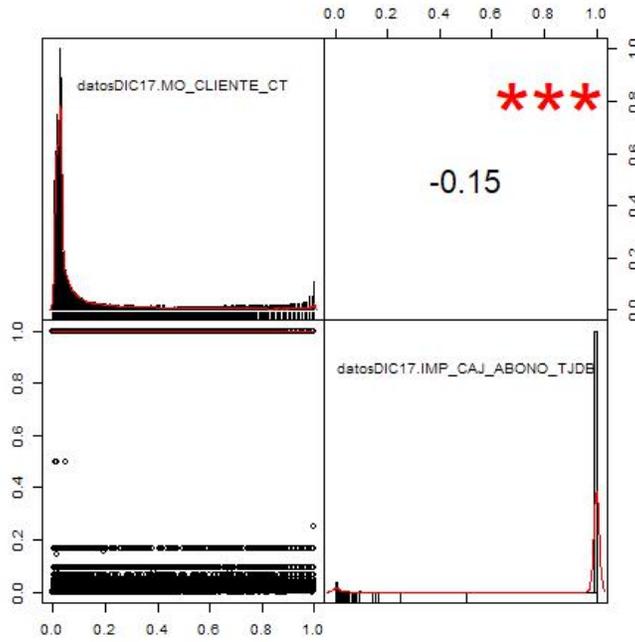


Figura A.48: Gráfico de la variable IMP_CAJ_ABONO_TJDB

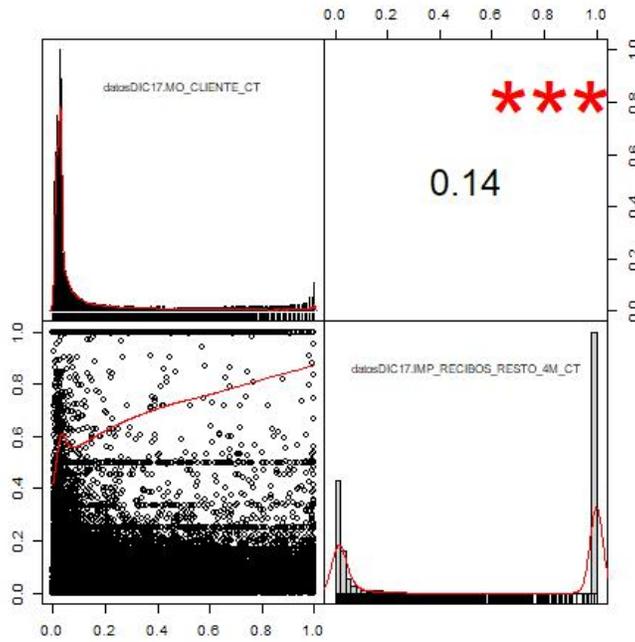


Figura A.49: Gráfico de la variable IMP_RECIBOS_RESTO_4M_CT

A.3. GRÁFICOS DE DISPERSIÓN ENTRE LAS DISTINTAS VARIABLES Y MO_CLIENTE_CT113

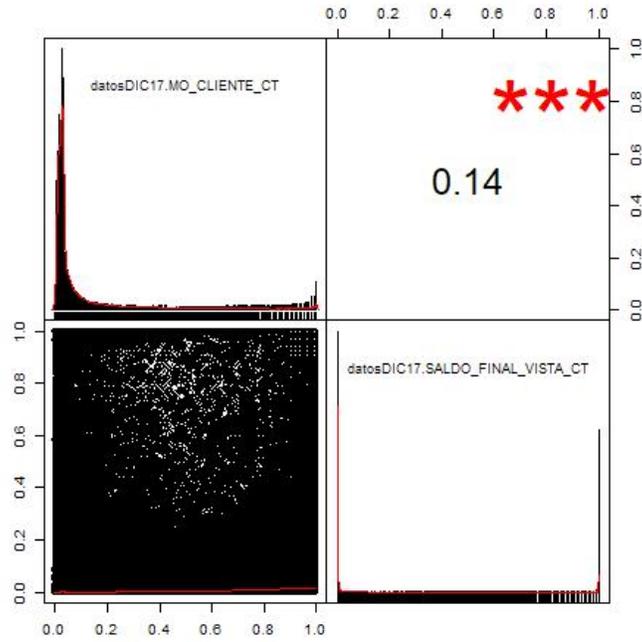


Figura A.50: Gráfico de la variable SALDO_FINAL_VISTA_CT

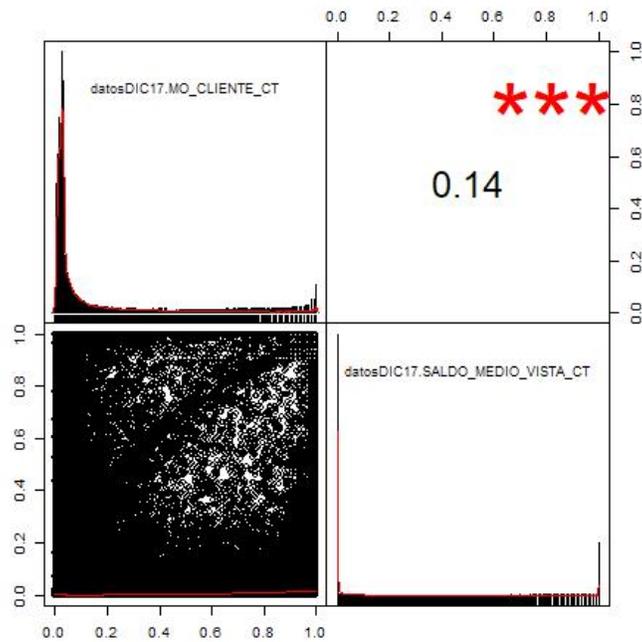


Figura A.51: Gráfico de la variable SALDO_MEDIO_VISTA_CT

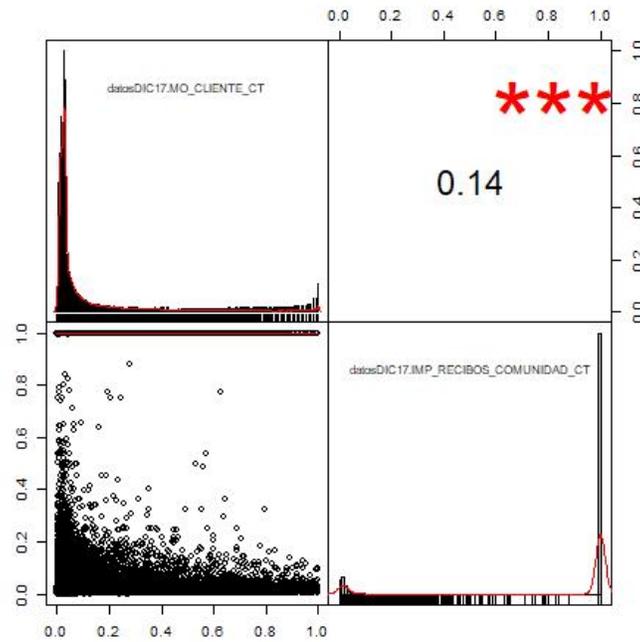


Figura A.52: Gráfico de la variable IMP_RECIBOS_COMUNIDAD_CT

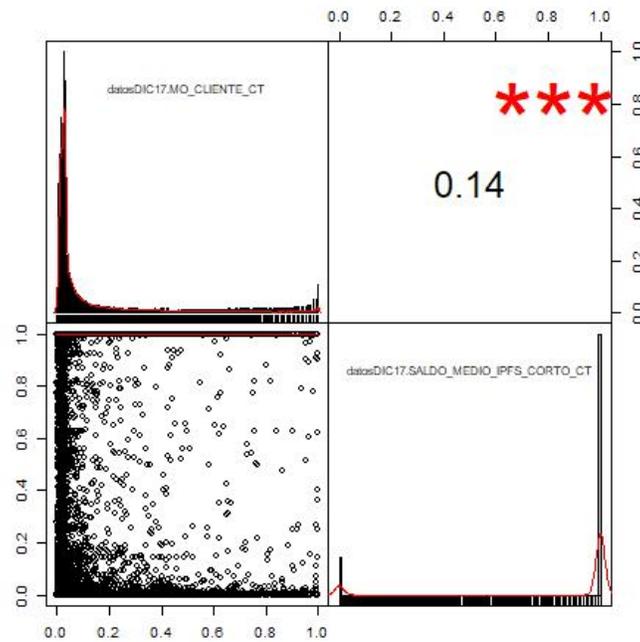


Figura A.53: Gráfico de la variable SALDO_MEDIO_IPFS_CORTO_CT

A.3. GRÁFICOS DE DISPERSIÓN ENTRE LAS DISTINTAS VARIABLES Y MO_CLIENTE_CT115

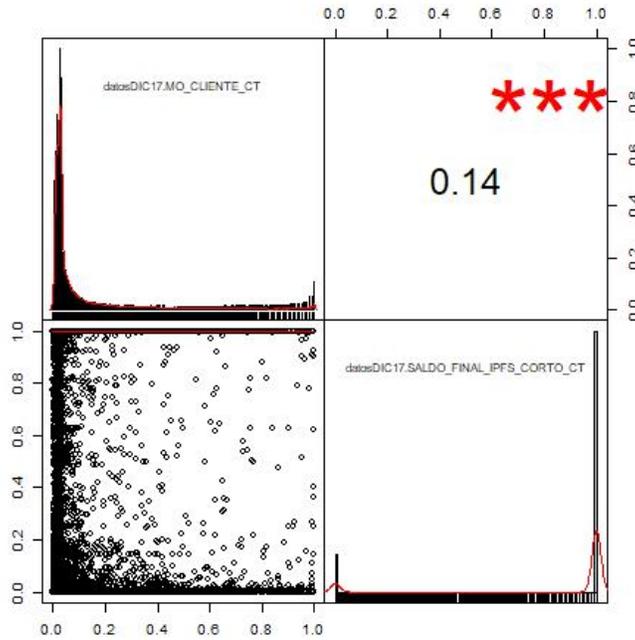


Figura A.54: Gráfico de la variable SALDO_FINAL_IPFS.CORTO_CT

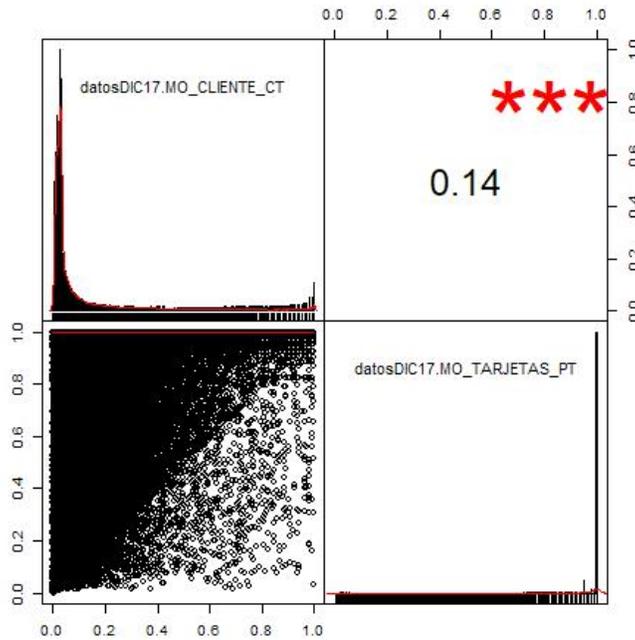


Figura A.55: Gráfico de la variable MO_TARJETAS_PT

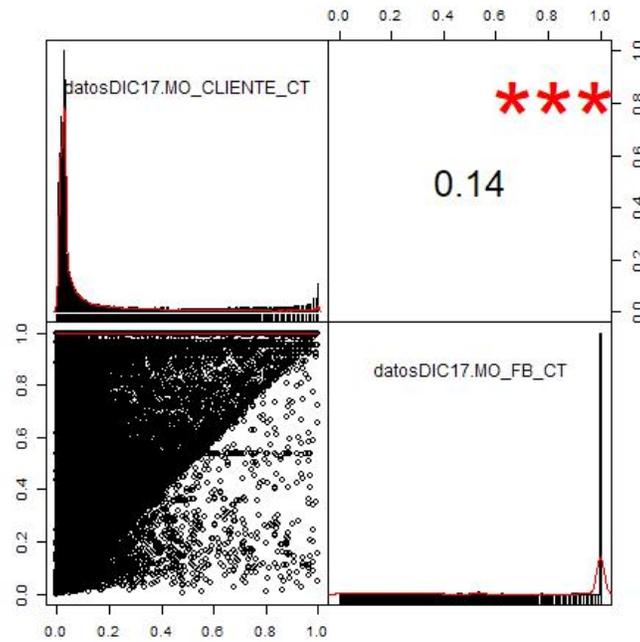


Figura A.56: Gráfico de la variable MO_FB_CT

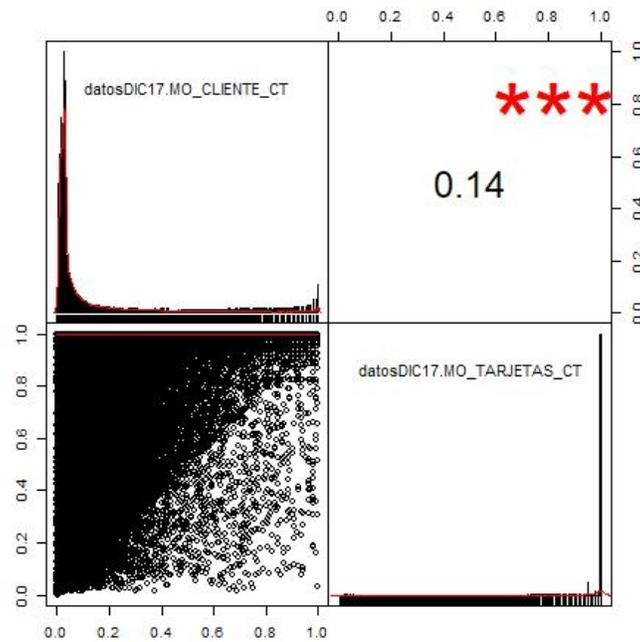


Figura A.57: Gráfico de la variable MO_TARJETAS_CT

A.4. GRÁFICOS DE DENSIDADES DE LAS VARIABLES UTILIZADAS PARA LOS CLUSTER117

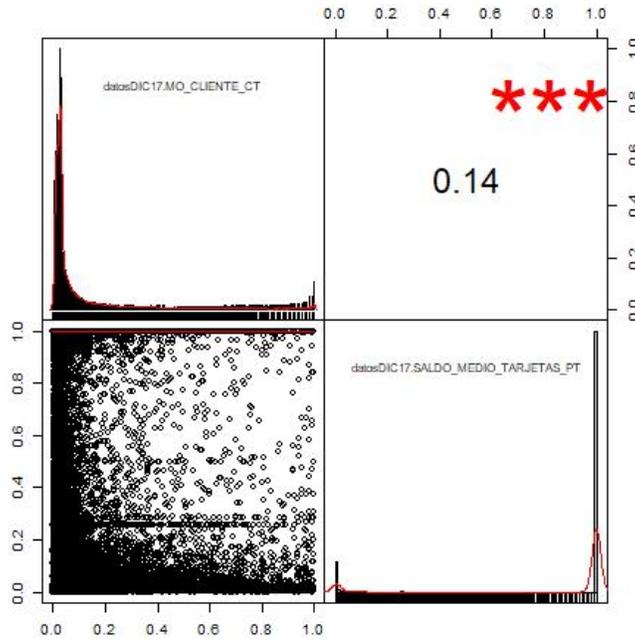
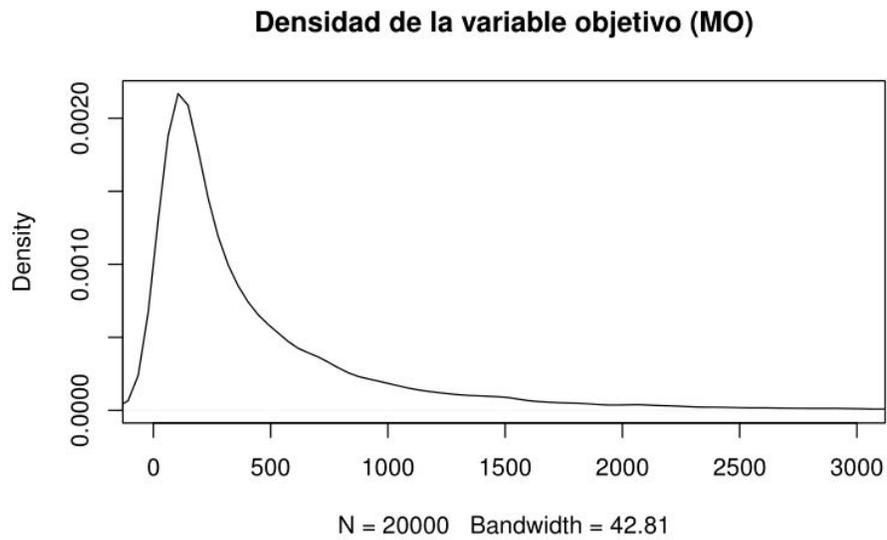
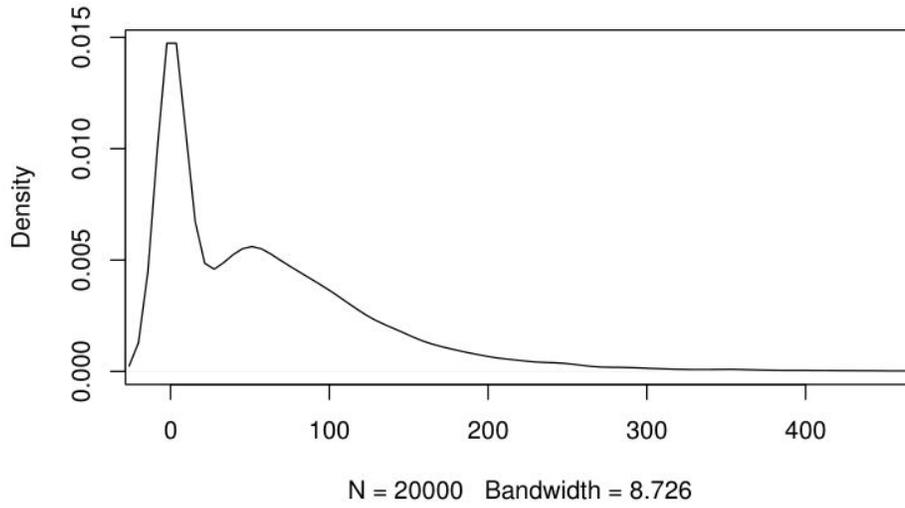


Figura A.58: Gráfico de la variable SALDO_MEDIO_TARJETAS_PT

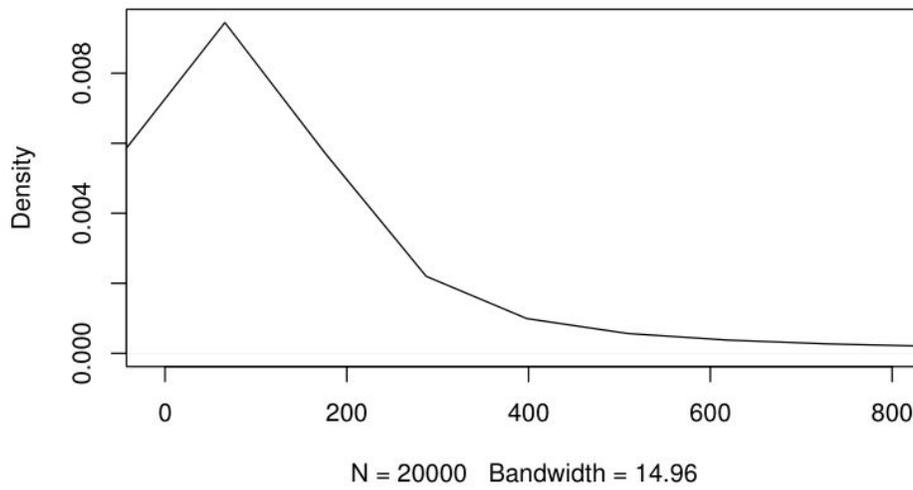
A.4. Gráficos de densidades de las variables utilizadas para los cluster



Densidad de la variable de recibos de energia

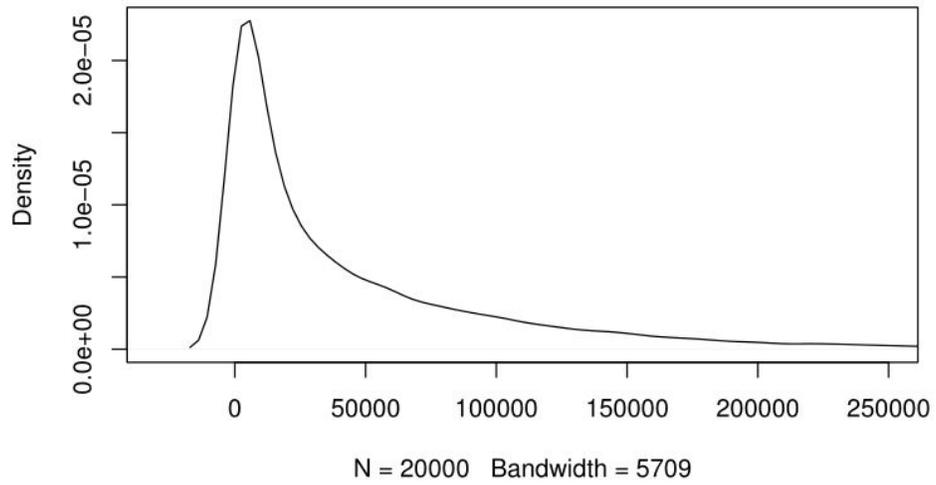


Densidad de la variable de recibos de impuestos

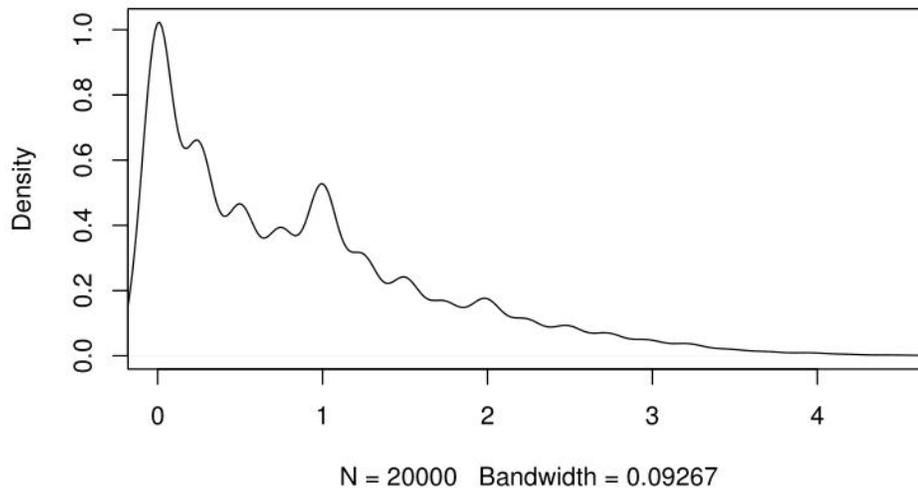


A.4. GRÁFICOS DE DENSIDADES DE LAS VARIABLES UTILIZADAS PARA LOS CLUSTER119

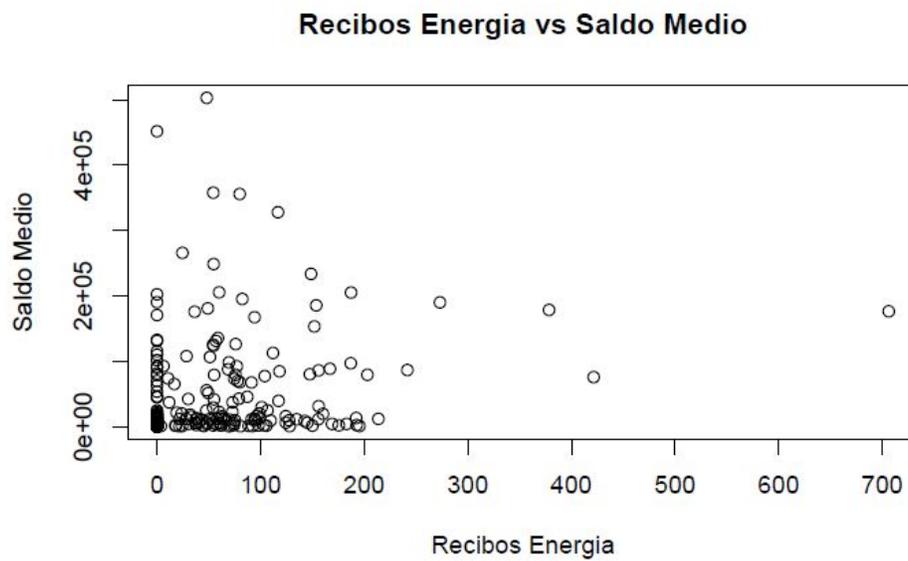
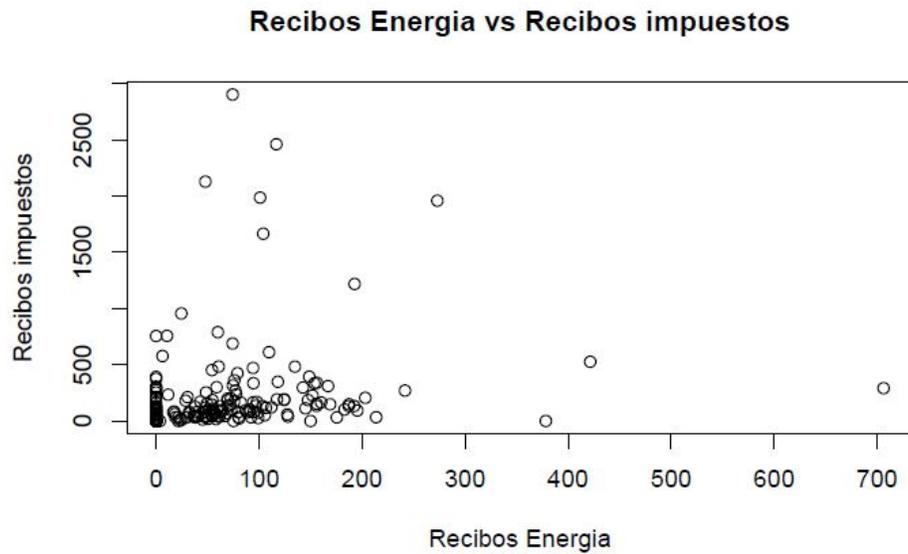
Densidad de la variable del saldo medio



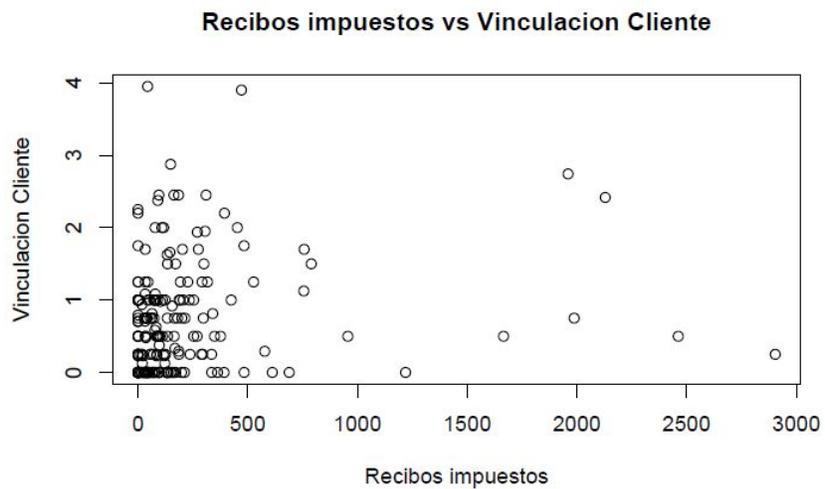
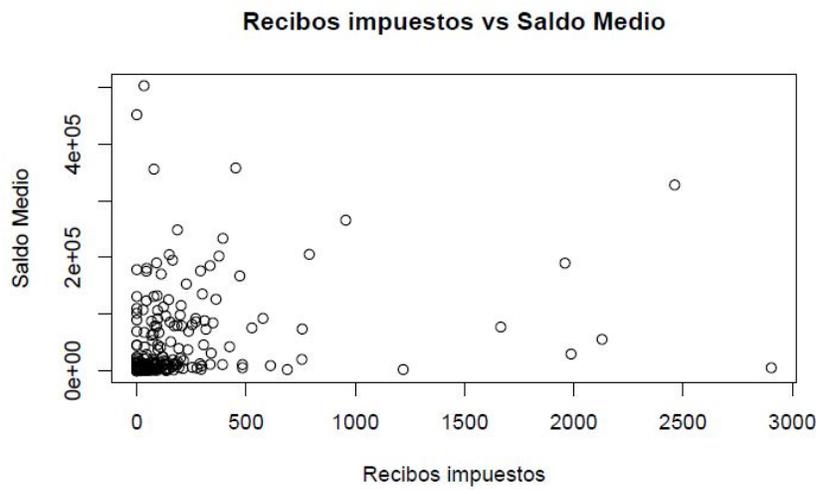
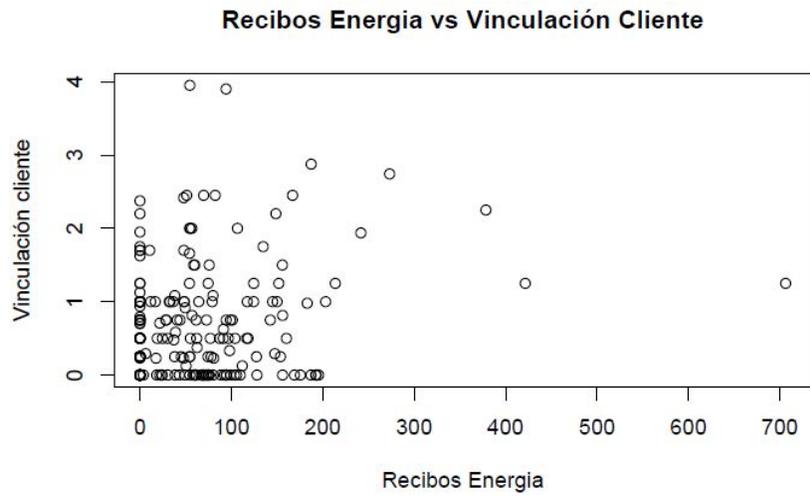
Densidad de la variable de la vinculación del cliente

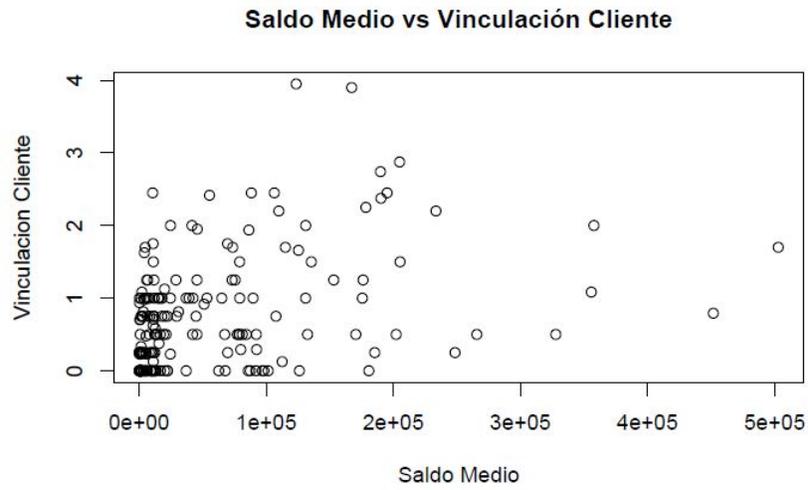


A.5. Gráficos de dispersión de las variables utilizadas para los cluster

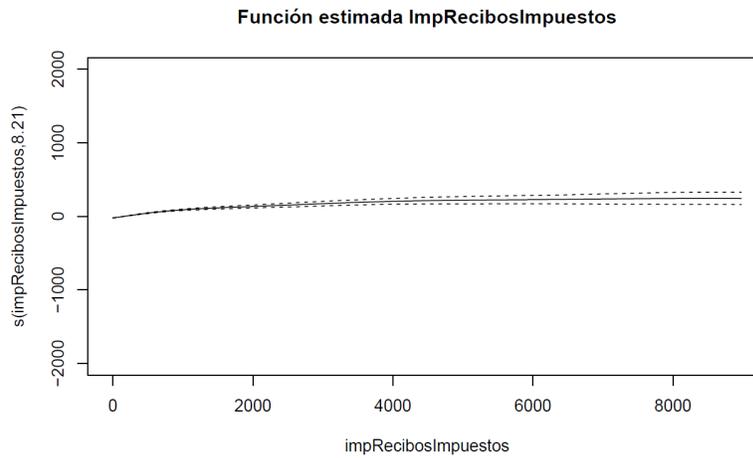


A.5. GRÁFICOS DE DISPERSIÓN DE LAS VARIABLES UTILIZADAS PARA LOS CLUSTER121

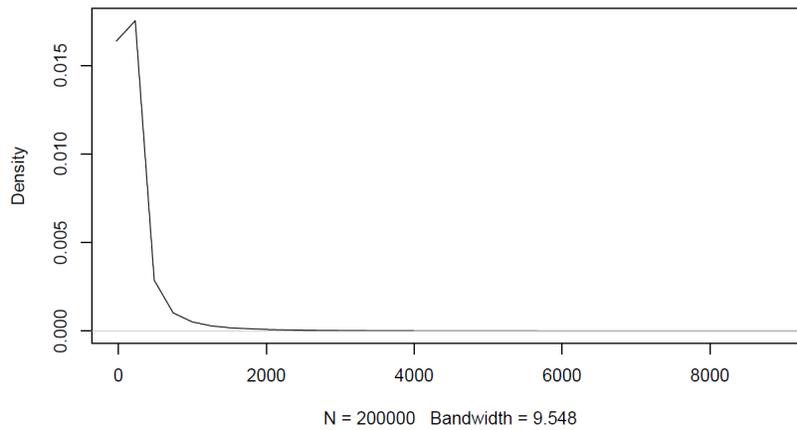




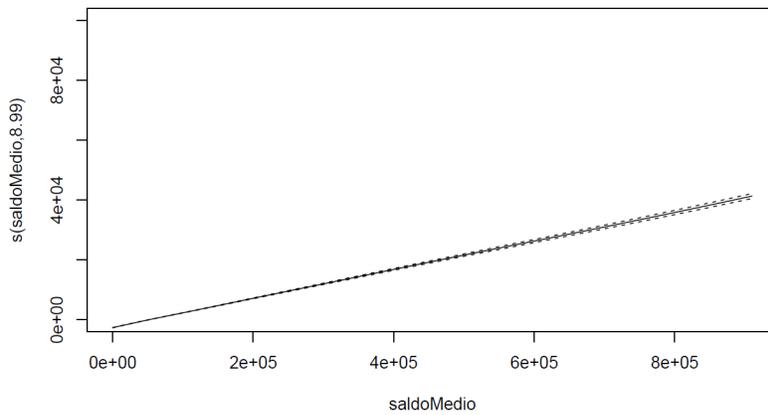
A.6. GAM. Funciones estimadas y funciones de densidad



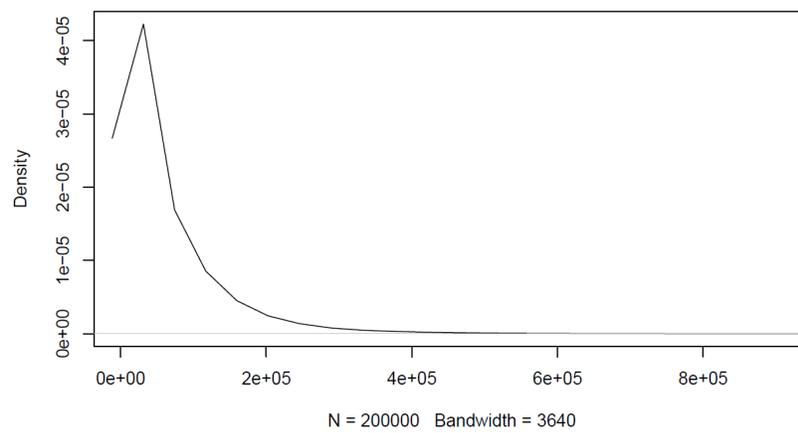
Función de densidad de ImpRecibosImpuestos

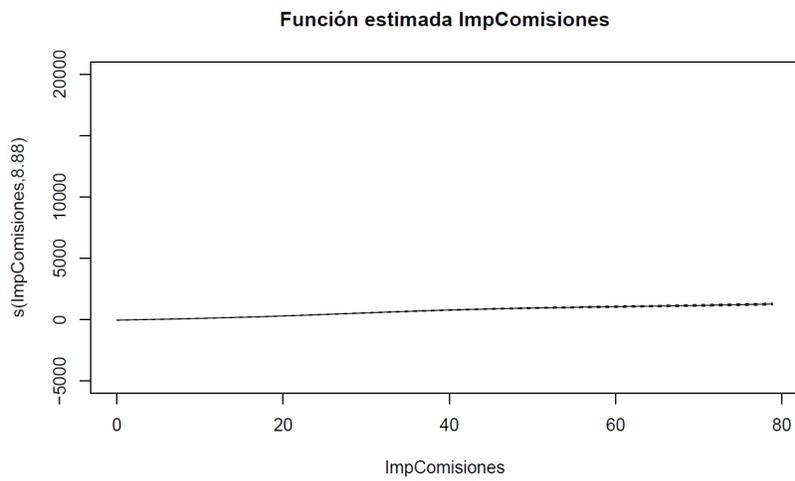
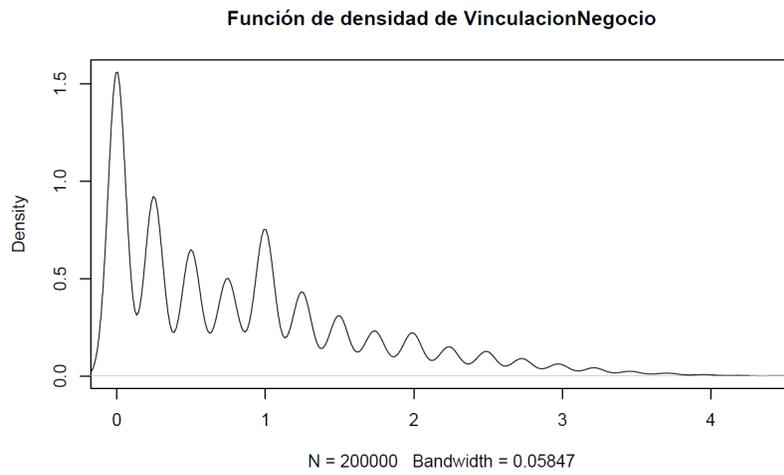
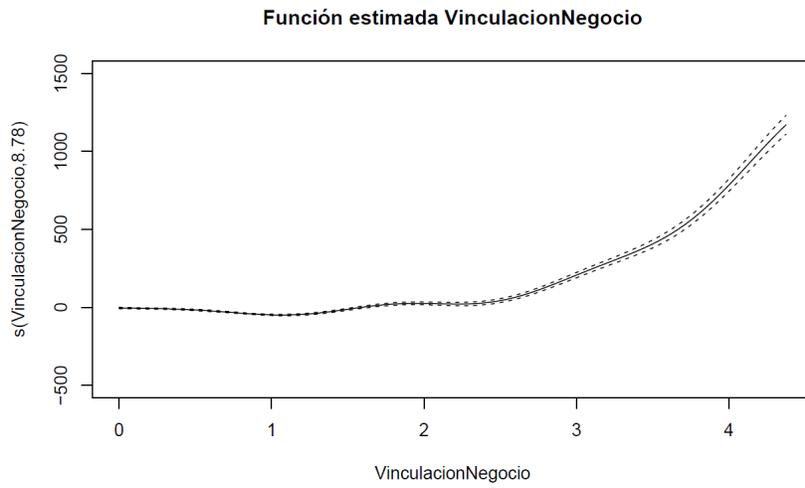


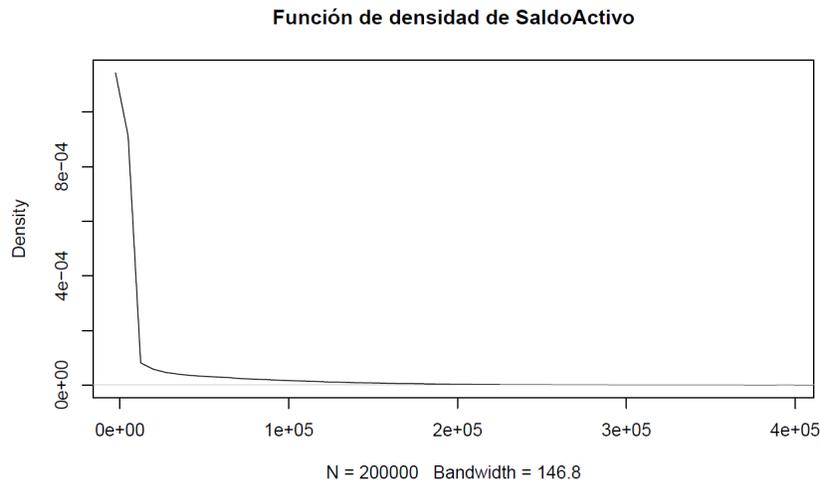
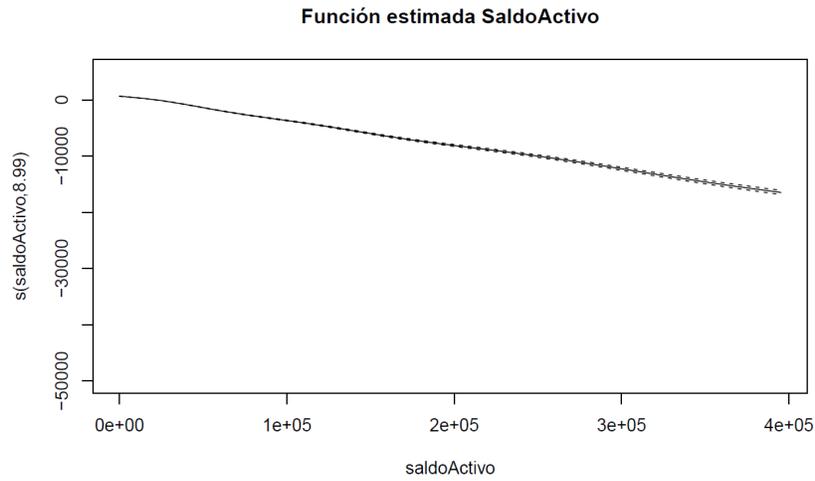
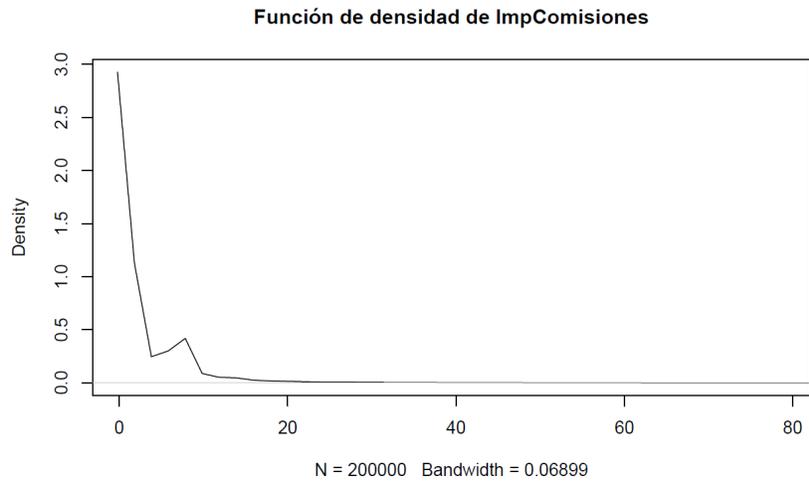
Función estimada SaldoMedio



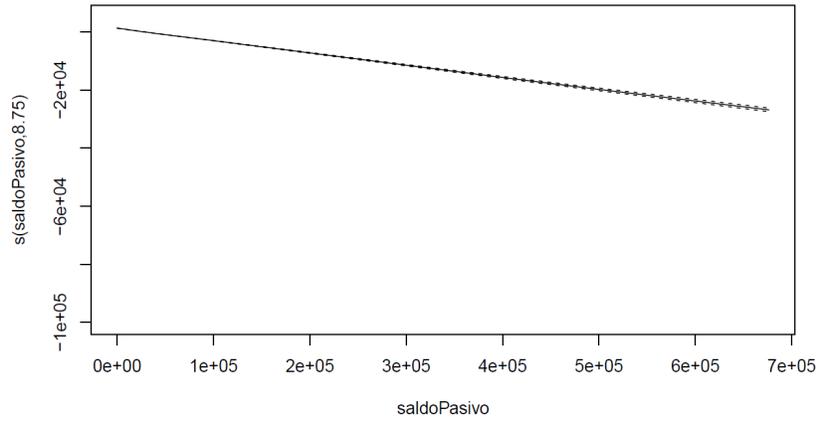
Función de densidad de SaldoMedio



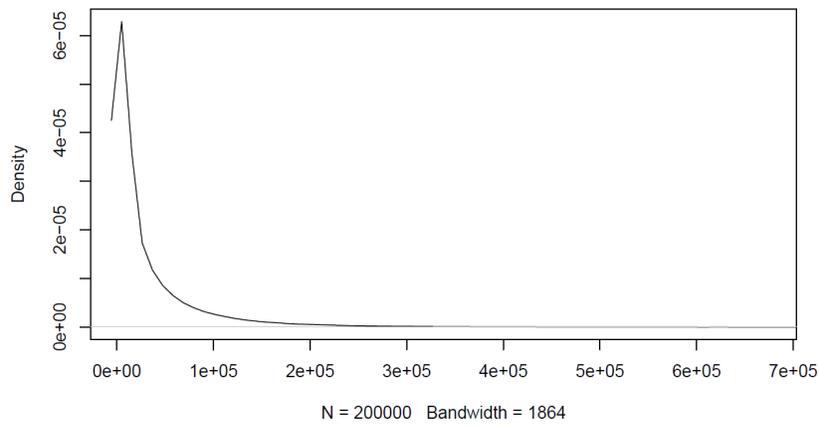




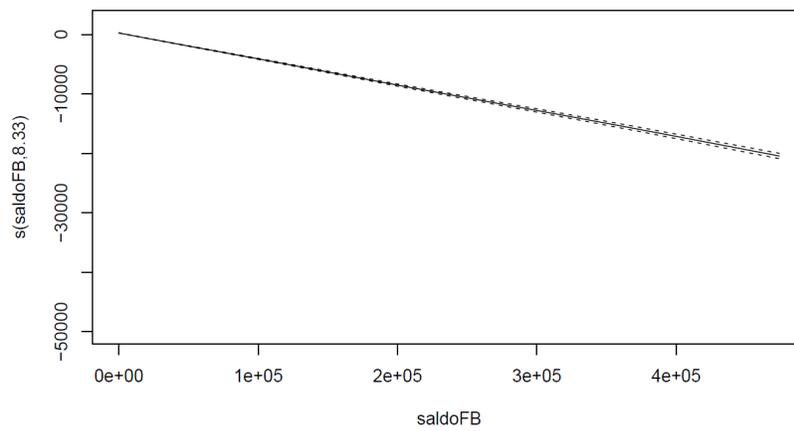
Función estimada SaldoPasivo



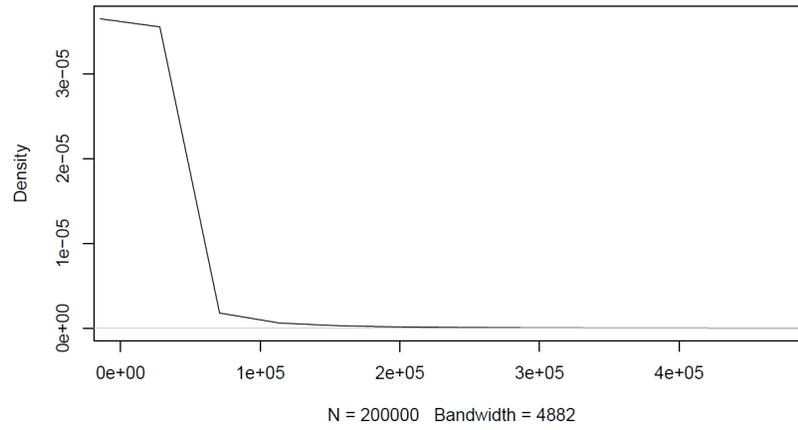
Función de densidad de SaldoPasivo



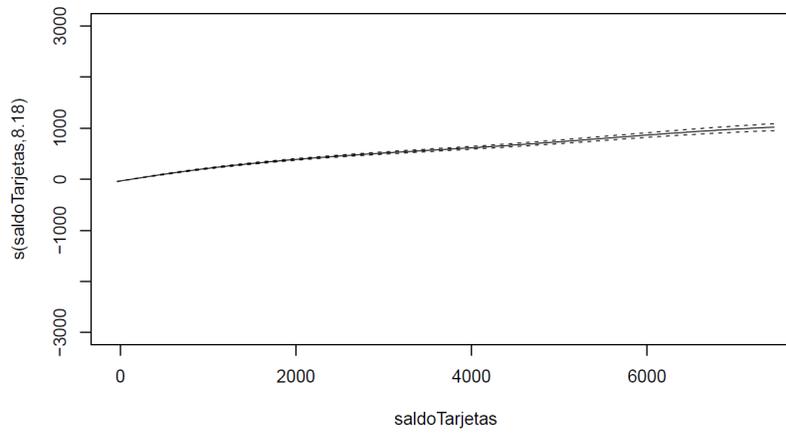
Función estimada SaldoFB



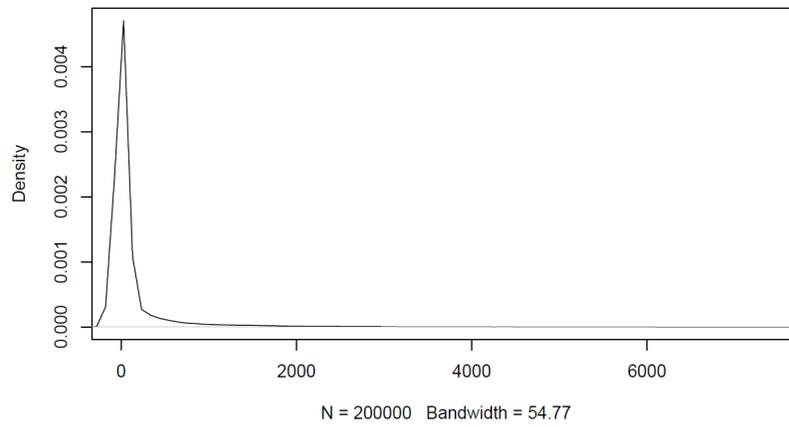
Función de densidad de SaldoFB



Función estimada SaldoTarjetas



Función de densidad de SaldoTarjetas



A.7. Gráficas de la densidades de las variables diferenciando segmentos

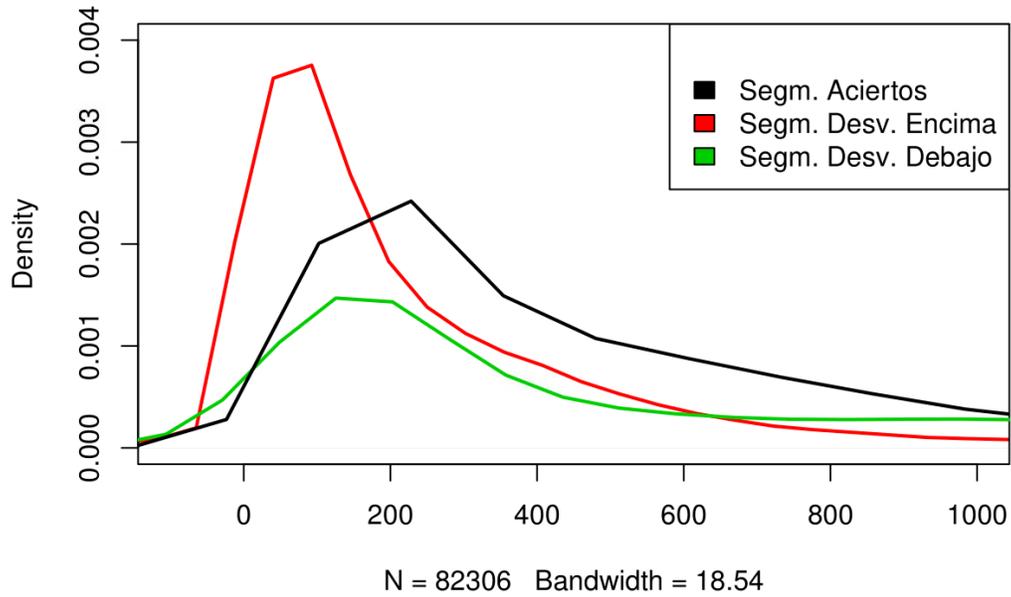


Figura A.59: Gráfico de la función de densidad de MO segmentada

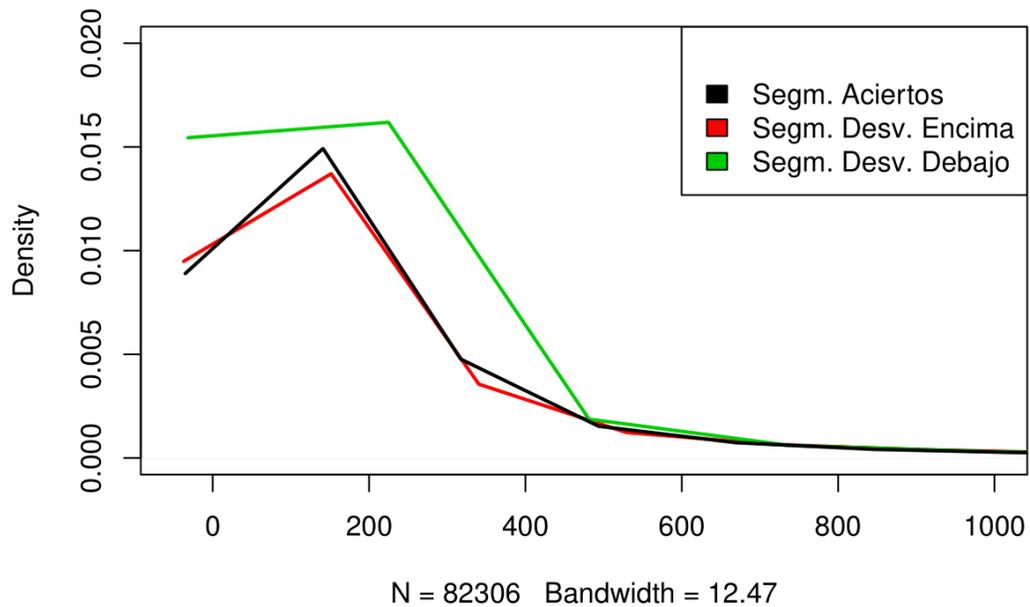


Figura A.60: Gráfico de la función de densidad de IMP_RECIBOS_IMPUESTOS_4M_CT segmentada

A.7. GRÁFICAS DE LA DENSIDADES DE LAS VARIABLES DIFERENCIANDO SEGMENTOS129

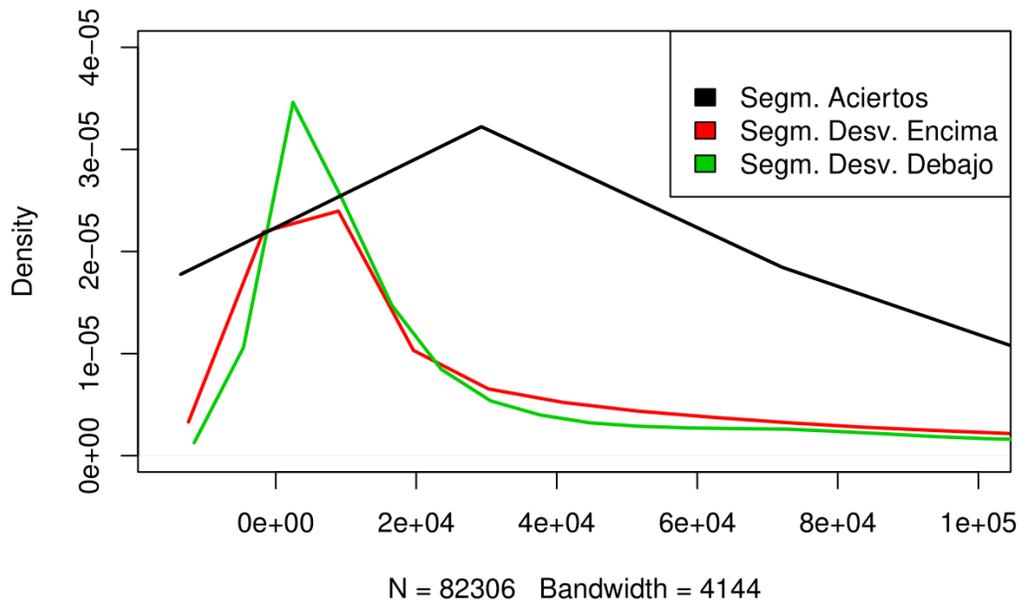


Figura A.61: Gráfico de la función de densidad de SALDO_MEDIO_CLIENTE_CT segmentada

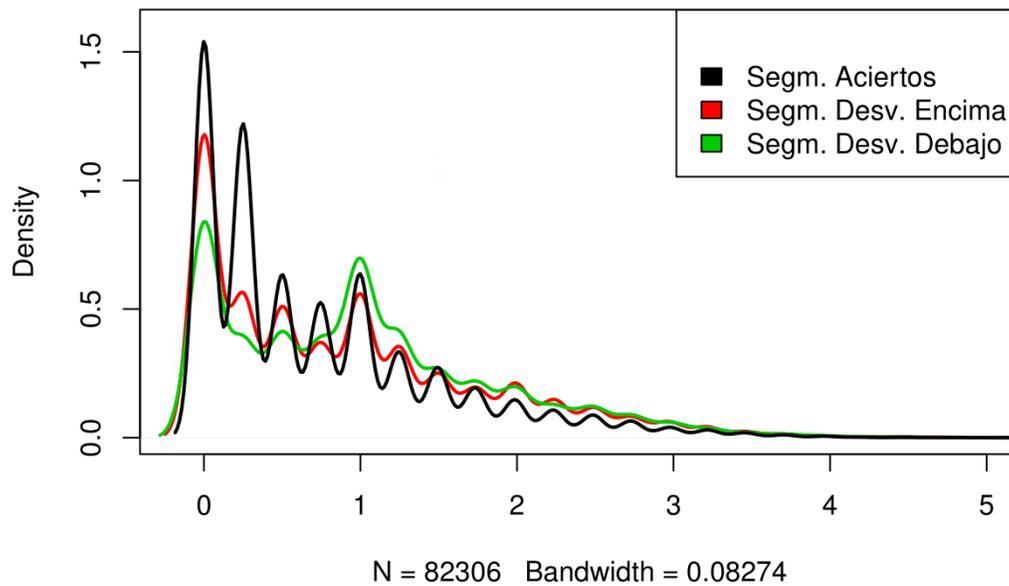


Figura A.62: Gráfico de la función de densidad de VINCULACION_NEGOCIO_CT segmentada

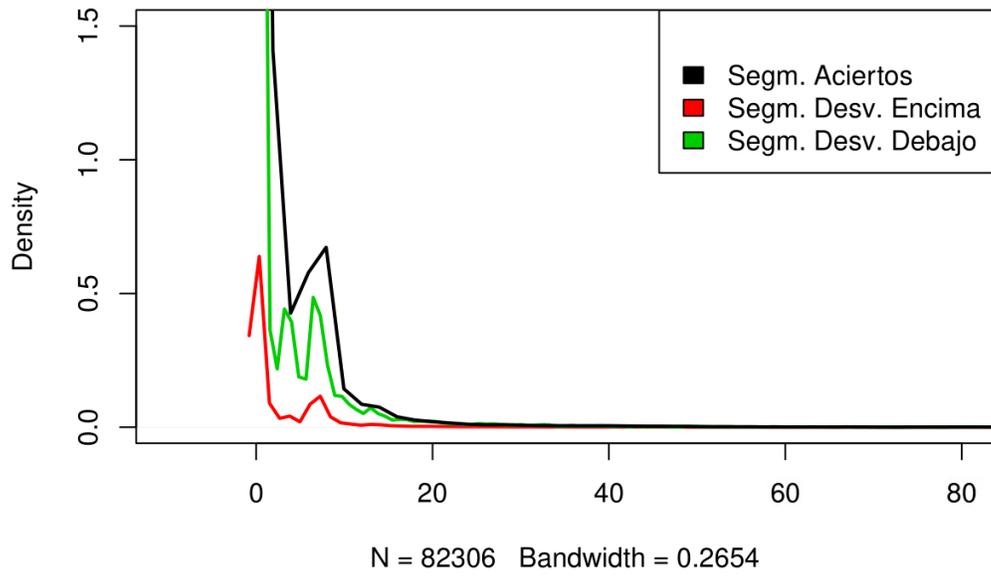


Figura A.63: Gráfico de la función de densidad de IMP_COMISIONES segmentada

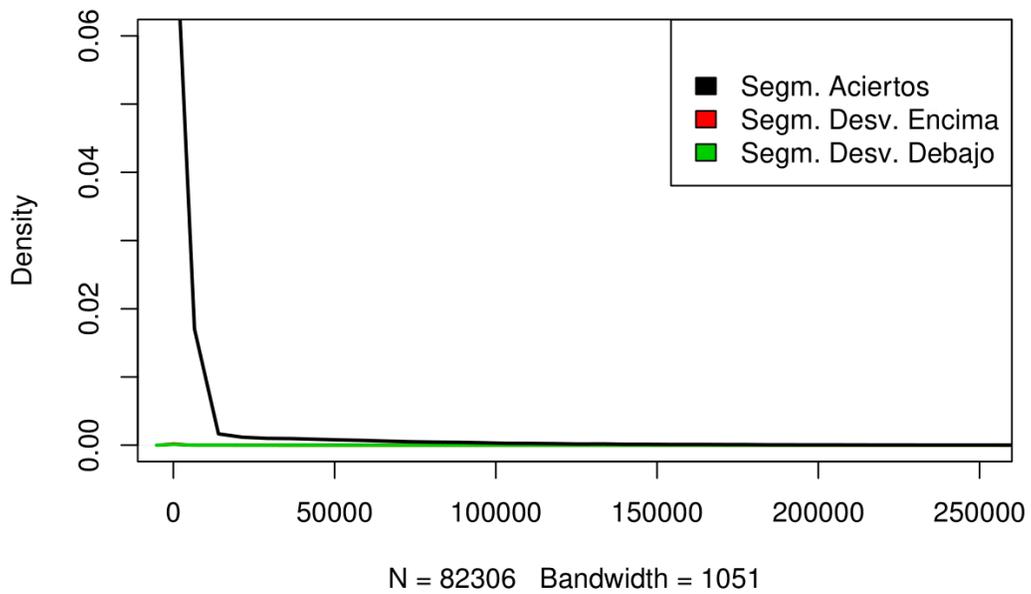


Figura A.64: Gráfico de la función de densidad de SALDO_MEDIO_ACTIVADO_CT segmentada

A.7. GRÁFICAS DE LA DENSIDADES DE LAS VARIABLES DIFERENCIANDO SEGMENTOS131

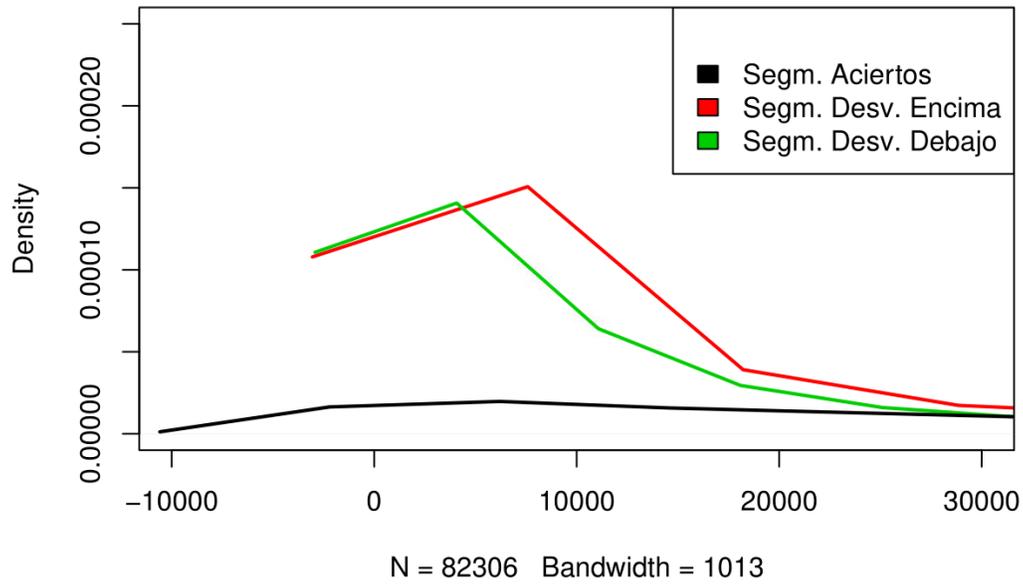


Figura A.65: Gráfico de la función de densidad de SALDO_MEDIO_pasivo_CT segmentada

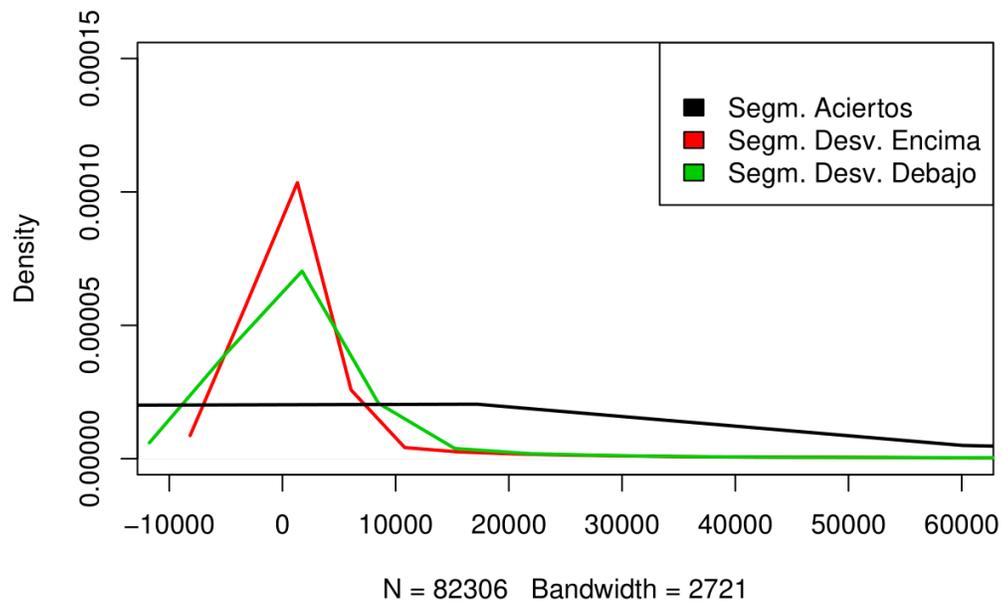


Figura A.66: Gráfico de la función de densidad de SALDO_MEDIO_fb_CT segmentada

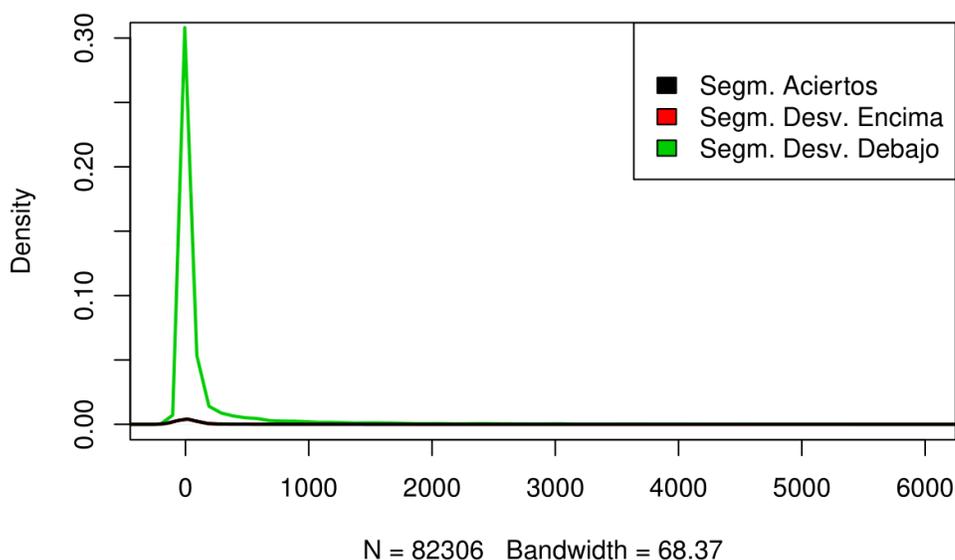


Figura A.67: Gráfico de la función de densidad de SALDO_MEDIO_TARJETAS_CT segmentada

A.8. Métricas del data mart de clientes

Campo	Descripción
CLIENTE.ID	Año, Trimestre, Mes
FECHA	Fecha de obtención de los datos
CARPETA	Identificador de la cartera a la que pertenece el cliente
SEGMENTO.ID.DINAMICO	Identificador del segmento estrategico dinamico al que pertenece el cliente
SEGMENTO.ID.ESTATICO	Identificador del segmento estrategico estatico al que pertenece el cliente
SEXO	Sexo del cliente
EDAD	Edad del cliente
AYUNTAMIENTO.ID	Identificador del ayuntamiento del cliente

Continúa en la siguiente página

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
PROVINCIA.ID	Identificador de la provincia del cliente
PAIS.RESIDENCIA	Pais de residencia del cliente
PAIS.NACIONALIDAD	Nacionalidad de cliente
RELACION.CONTRACTUAL.ID	Identificador de las relaciones contractuales que tiene el cliente con los contratos.
PRODUCTO.FINANCIERO.ID.PT	Muestra los productos financieros contratados por el cliente como PT.
PRODUCTO.FINANCIERO.ID.CT	Muestra los productos financieros contratados por el cliente como CT.
SERVICIO.ID	Muestra los servicios contratados por el cliente.
MESES.ANTIGUEDAD.CLIENTE	Número de meses de antigüedad del cliente
MESES.ANTIGUEDAD.ACTIVO	Número de meses de antigüedad del primer contrato de activo del cliente
MESES.ANTIGUEDAD.PASIVO	Número de meses de antigüedad del primer contrato de pasivo del cliente
MESES.ANTIGUEDAD.FBALANCE	Número de meses de antigüedad del primer contrato fuera de balance del cliente
MESES.ANTIGUEDAD.SERVICIOS	Número de meses de antigüedad del primer contrato de servicios del cliente
MESES.ANTIGUEDAD.INGRESOS	Número de meses desde que el cliente tiene ingresos regulares
ANTIGUEDAD.CLIENTE	Trameado de antigüedad del cliente
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
ANTIGUEDAD.ACTIVO	Trameado de antigüedad del primer contrato de activo del cliente
ANTIGUEDAD.PASIVO	Trameado de antigüedad del primer contrato de pasivo del cliente
ANTIGUEDAD.FBALANCE	Trameado de antigüedad del primer contrato de fuera de balance del cliente
ANTIGUEDAD.SERVICIOS	Trameado de antigüedad del primer contrato de servicios del cliente
ANTIGUEDAD.INGRESOS	Trameado de antigüedad del primer mes del cliente con ingresos
ANTIGUEDAD	Indica las distintas antigüedades del cliente
TRANSICION_ESTADO_ID	Transición del estado del cliente respecto al mes anterior
ESTADO_ANTERIOR_CLIENTE_ID	Estado del cliente en el mes anterior
ESTADO_ACTUAL_CLIENTE_ID	Estado actual del cliente
OFICINA_1_CT	Primer centro del cliente asignado por marketing
TARIFA_PLANA	Indicador de si el cliente tiene contratado algún producto de tarifa plana.
OFICINA_2_CT	Segundo centro del cliente asignado por marketing
OFICINA_3_CT	Tercer centro del cliente asignado por marketing
OFICINA_OPERA_PT	Oficina en la que el cliente mas opera con sus contratos como primer titular
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
OFICINA_OPERA_CT	Oficina en la que el cliente mas opera con sus contratos como cualquier titular
TIPO_PERSONA	Código identificador del tipo de persona
FECHA_ALTA	Fecha de alta del cliente
FECHA_NACIMIENTO	Fecha de nacimiento del cliente
FECHA_ANTIGUEDAD_CLIENTE	Fecha de antigüedad del cliente
FECHA_ANTIGUEDAD_ACTIVO	Fecha de alta del primer contrato de activo del cliente
FECHA_ANTIGUEDAD_PASIVO	Fecha de alta del primer contrato de pasivo del cliente
FECHA_ANTIGUEDAD_FBALANCE	Fecha de alta del primer contrato fuera de balance del cliente
FECHA_ANTIGUEDAD_SERVICIOS	Fecha de alta del primer contrato de servicios del cliente
FECHA_ANTIGUEDAD_INGRESOS	Fecha del primer mes con ingresos regulares del cliente
FC_SITUACION	Fecha del estado actual del cliente
FC_ACTIVACION	Fecha en el que el cliente pasó a estado con actividad
FC_CAPTACION	Fecha en el que se captó al cliente en la entidad
FC_BAJA	Fecha de baja del cliente en la entidad
IMP_RECIBOS_AGUA_4M_CT	Importe medio mensual en los últimos 4 meses correspondiente a los recibos del agua. Se calcula la media obteniendo la suma de los recibos del agua de los últimos 4 meses y dividiendo por los meses en los que hay dato.
IMP_RECIBOS_TELEFONO_4M_CT	Importe medio mensual en los últimos 4 meses correspondiente a los recibos del teléfono.
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
	diente a los recibos del teléfono. Se calcula la media obteniendo la suma de los recibos del teléfono de los últimos 4 meses y diviendo por los meses en los que hay dato.
IMP_RECIBOS_ENERGIA_4M_CT	Importe medio mensual en los últimos 4 meses correspondiente a los recibos de energía. Se calcula la media obteniendo la suma de los recibos de energía de los últimos 4 meses y dividiendo por los meses en los que hay dato.
IMP_RECIBOS_COMUNIDAD_4M_CT	Importe medio mensual en los últimos 4 meses correspondiente a los recibos de la comunidad. Se calcula la media obteniendo la suma de los recibos del a comunidad de los últimos 4 meses y diviendo por los meses en los que hay dato.
IMP_RECIBOS_ALQUILER_4M_CT	Importe medio mensual en los últimos 4 meses correspondiente a los recibos del alquiler. Se calcula la media obteniendo la suma de los recibos del alquiler de los últimos 4 meses y diviendo por los meses en los que hay dato.
IMP_RECIBOS_COLEGIOS_4M_CT	Importe medio mensual en los últimos 4 meses correspondiente a los recibos del colegio. Se calcula la media obteniendo la suma de los recibos del colegio de los últimos 4 meses y diviendo por los meses en los que
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
	hay dato.
IMP_RECIBOS_SS_4M_CT	Importe medio mensual en los últimos 4 meses correspondiente a los recibos de la SS. Se calcula la media obteniendo la suma de los recibos de la SS de los últimos 4 meses y dividiendo por los meses en los que hay dato.
IMP_RECIBOS_IVA_4M_CT	Importe medio mensual en los últimos 4 meses correspondiente a los recibos del IVA. Se calcula la media obteniendo la suma de los recibos del IVA de los últimos 4 meses y dividiendo por los meses en los que hay dato.
IMP_RECIBOS_SM_4M_CT	Importe medio mensual en los últimos 4 meses correspondiente a los recibos del seguro médico. Se calcula la media obteniendo la suma de los recibos del seguro médico de los últimos 4 meses y dividiendo por los meses en los que hay dato.
IMP_RECIBOS_EF_4M_CT	Importe medio mensual en los últimos 4 meses correspondiente a los recibos de otras entidades financieras. Se calcula la media obteniendo la suma de los recibos de otras entidades financieras de los últimos 4 meses y dividiendo por los meses en los que hay dato.
IMP_RECIBOS_IMPUESTOS_4M_CT	Importe medio mensual en los últimos 4 meses correspondiente a los recibos de impuestos. Se calcula la media
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
	obteniendo la suma de los recibos de impuestos de los últimos 4 meses y diviendo por los meses en los que hay dato.
IMP_RECIBOS_RESTO_4M_CT	Importe medio mensual en los últimos 4 meses correspondiente a los recibos del resto. Se calcula la media obteniendo la suma de los recibos del resto de los últimos 4 meses y diviendo por los meses en los que hay dato.
OP_TPV_IMPOSICION_TJCR_3M	Calcula la media de las imposiciones realizadas en TPV a las tarjetas de crédito de los últimos 3 meses, contando solamente los meses que tienen operaciones en tarjetas
OP_CAJ_IMPOSICION_TJCR_3M	Calcula la media de operaciones de los abonos realizados en TPV a las tarjetas de crédito de los últimos 3 meses, contando solamente los meses que tienen operaciones en tarjetas
OP_TPV_ABONO_TJCR_3M	Calcula la media de operaciones de las imposiciones realizadas en cajeros a las tarjetas de crédito de los últimos 3 meses, contando solamente los meses que tienen operaciones en tarjetas
OP_CAJ_ABONO_TJCR_3M	Calcula la media de operaciones de los abonos realizados en cajeros a las tarjetas de crédito de los últimos 3 meses, contando solamente los meses que tienen
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
	operaciones en tarjetas
OP_TPV_IMPOSICION_TJDB_3M	Calcula la media de operaciones de las imposiciones realizadas en TPVs a las tarjetas de débito de los últimos 3 meses, contando solamente los meses que tienen operaciones en tarjetas
OP_TPV_ABONO_TJDB_3M	Calcula la media de operaciones de los abonos realizados en TPVs a las tarjetas de débito de los últimos 3 meses , contando solamente los meses que tienen operaciones en tarjetas
OP_CAJ_IMPOSICION_TJDB_3M	Calcula la media de operaciones de las imposiciones realizadas en cajeros a las tarjetas de débito de los últimos 3 meses, contando solamente los meses que tienen operaciones en tarjetas
OP_CAJ_ABONO_TJDB_3M	Calcula la media de operaciones de los abonos realizados en cajeros a las tarjetas de débito de los últimos 3 meses contando solamente los meses que tienen operaciones en tarjetas
IMP_TPV_IMPOSICION_TJCR_3M	Calcula el importe medio de las imposiciones realizadas en TPVs a las tarjetas de crédito de los últimos 3 meses, contando solamente los meses que tienen operaciones en tarjetas
IMP_CAJ_IMPOSICION_TJCR_3M	Calcula el importe medio de los abonos realizados en TPVs a las tarjetas de crédito de los últimos 3 meses
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
	, contando solamente los meses que tienen operaciones en tarjetas
IMP_TPV_ABONO_TJCR_3M	Calcula el importe medio de las imposiciones realizadas en cajeros a las tarjetas de crédito de los últimos 3 meses, contando solamente los meses que tienen operaciones en tarjetas
IMP_CAJ_ABONO_TJCR_3M	Calcula el importe medio de los abonos realizados en cajeros a las tarjetas de crédito de los últimos 3 meses, contando solamente los meses que tienen operaciones en tarjetas
IMP_TPV_IMPOSICION_TJDB_3M	Calcula el importe medio de las imposiciones realizadas en TPV a las tarjetas de débito de los últimos 3 meses , contando solamente los meses que tienen operaciones en tarjetas
IMP_TPV_ABONO_TJDB_3M	Calcula el importe medio de los abonos realizados en TPVs a las tarjetas de débito de los últimos 3 meses, contando solamente los meses que tienen operaciones en tarjetas
IMP_CAJ_IMPOSICION_TJDB_3M	Calcula el importe medio de las imposiciones realizadas en cajeros a las tarjetas de déb. de los últimos 3 meses, contando solamente los meses que tienen operaciones en tarjetas
IMP_CAJ_ABONO_TJDB_3M	Calcula el importe medio de los abonos realizados en
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
	cajeros a las tarjetas de débito de los últimos 3 meses, contando solamente los meses que tienen operaciones en tarjetas
IMP_NOMINA_SUA	Calcula la media de las nominas de los últimos 5 meses , contando solamente los meses en los que la suma de los importes de refer. 03, 65, 97 y 01-18 supere el umbral de 300€siempre y cuando se cumpla en el último mes el indicador de nómina
IMP_NOMINA_PURO_SUA	Calcula la media de las nominas de los últimos 5 meses , contando solamente los meses en que la suma de los importes de referencias 03, 65, 97 y 01/18 supere el umbral de 300€siempre y cuando se cumpla en el último mes el indicador de nómina pura
IMP_PENSION_SUA	Calcula la media de las pensiones de los últimos 5 meses , contando solamente los meses que tienen ingresos
IMP_INEM_PURO_SUA	Calcula la media de los ingresos del INEM de los últimos 5 meses, contando solamente los meses en los que la suma de los importes de referencia 65 supere el umbral de 300€, siempre y cuando se cumpla en el último mes el indicador de desempleo puro
IMP_TRANSFERENCIAS_SUA	Calcula la media de las transferencias recibidas de los últimos 5 meses, contando solamente los meses en que la suma de importes de referencia 19/22 supere
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
	el umbral de 600€, siempre y cuando se cumpla en el último mes el indicador de transferencias
IMP_INCAPACIDAD_SUA	Calcula la media de las transferencias recibidas por incapacidad de los últimos 5 meses, contando solamente los meses en los que la suma de los importes de referencia 97 supere el umbral de 200€, siempre y cuando se cumpla en el último mes el indicador de incapacidad
IMP_ORD_EXTRANJERO_SUA	Calcula la media de las transferencias recibidas por ordenes del extranjero de los últimos 5 meses, contando solamente los meses en los que la suma de los importes de referencia 06 supere el umbral de 600€
VINCULACION_TRANSACCIONAL_CT	Vinculación transaccional como cualquier titular
VINCULACION_NEGOCIO_PT	Vinculación de negocio como primer titular
VINCULACION_NEGOCIO_CT	Vinculación de negocio como cualquier titular
NUM_QUEJAS	Número de quejas realizadas por el cliente
NUM_RECLAMACIONES	Número de reclamaciones realizadas por el cliente
MESES_INCLUIDO_AGENDA	Meses que lleva el cliente en una agenda
FC_ULT_AGENDA	Fecha de la última entrada del cliente en una agenda
OFI_ULT_CONTRATO_PT	Oficina asociada al último contrato del cliente como PT
FECHA_ULT_CONTRATO_PT	Fecha del último contrato del cliente como PT (El más reciente)
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
OFLULT_CONTRATO_CT	Oficina asociada al último contrato del cliente como CT
FECHA_ULT_CONTRATO_CT	Fecha del último contrato del cliente como CT (El más reciente)
NUM_CONTRATA_CAPTACION	Número de contratos de pasivo del cliente
REC_CONTRATA_CAPTACION	Fecha de contratación de ultimo contrato de pasivo del cliente
NUM_CONTRATA_ACTIVIVO	Número de contratos de activo del cliente
REC_CONTRATA_ACTIVIVO	Fecha de contratación de ultimo contrato de activo del cliente
NUM_CONTRATA_FBALANCE	Número de contratos de fuera de balance del cliente
REC_CONTRATA_FBALANCE	Fecha de contratación de ultimo contrato de fuera de balance del cliente
NUM_CONTRATA_TARJETAS	Número de contratos de tarjetas del cliente
REC_CONTRATA_TARJETAS	Fecha de contratación de ultimo contrato de tarjeta del cliente
NUM_CONTRATA_OT_SERV	Número de contratos de otros servicios del cliente
REC_CONTRATA_OT_SERV	Fecha de contratación de ultimo contrato de otros servicios del cliente
REC_CONTRATA_ING_REGULARES	Información ingresos regulares
NUM_SESIONES_ON	Número de sesiones en la banca electrónica
OP_HABER_ON	Nº de operaciones al haber realizadas en la banca
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
	electrónica
OP_DEBE_ON	Nº de operaciones al debe realizadas en la banca electrónica
IMP_OP_HABER_ON	Importe de las operaciones al haber realizadas en la banca electrónica
IMP_OP_DEBE_ON	Importe de las operaciones al debe realizadas en la banca electrónica
NUM_CONTACTOS_TELEFONICO	Número de contactos telefónicos del cliente
OP_HABER_TELF	Número de operaciones al haber realizadas en la banca telefónica
OP_DEBE_TELF	Número de operaciones al debe realizadas en la banca telefónica
IMP_OP_HABER_TELF	Importe de las operaciones al haber realizadas en la banca telefónica
IMP_OP_DEBE_TELF	Importe de las operaciones al debe realizadas en la banca telefónica
NUM_CONTACTOS_MOVIL	Número de contactos móvil del cliente
OP_HABER_MOVIL	Número de operaciones al haber realizadas en la banca móvil
OP_DEBE_MOVIL	Número de operaciones al debe realizadas en la banca móvil
IMP_OP_HABER_MOVIL	Importe de operaciones al haber realizadas en la banca
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
	móvil
IMP_OP_DEBE_MOVIL	Importe de operaciones al debe realizadas en la banca móvil
NUM_VISITAS_OFICINA	Número de visitas a la oficina realizadas por el cliente
OP_HABER_OFICINA	Número de operaciones al haber realizadas en la oficina
OP_DEBE_OFICINA	Número de operaciones al debe realizadas en la oficina
IMP_OP_HABER_OFICINA	Importe de operaciones al haber realizadas en la oficina
IMP_OP_DEBE_OFICINA	Importe de operaciones al debe realizadas en la oficina
NUM_COMISIONES	Número de comisiones del cliente
IMP_COMISIONES	Importe de las comisiones del cliente
SALDO_FINAL_VISTA_PT	Saldo final de los productos a la vista como primer titular del cliente
SALDO_MEDIO_VISTA_PT	Saldo medio de los productos a la vista como primer titular del cliente
SALDO_MAX_VISTA_PT	Saldo máximo de los productos a la vista como primer titular del cliente
SALDO_MIN_VISTA_PT	Saldo mínimo de los productos a la vista como primer titular del cliente
OP_DEBE_VISTA_PT	Número de operaciones al debe de productos a la vista como primer titular
OP_HABER_VISTA_PT	Número de operaciones al haber de productos a la vista como primer titular
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
IMP_DEBE_VISTA_PT	Importe de operaciones al debe de productos a la vista como primer titular
IMP_HABER_VISTA_PT	Importe de operaciones al haber de productos a la vista como primer titular
SALDO_FINAL_VISTA_CT	Saldo final de los productos a la vista como cualquier titular del cliente
SALDO_MEDIO_VISTA_CT	Saldo medio de los productos a la vista como cualquier titular del cliente
SALDO_MAX_VISTA_CT	Saldo máximo de los productos a la vista como cualquier titular del cliente
SALDO_MIN_VISTA_CT	Saldo mínimo de los productos a la vista como cualquier titular del cliente
OP_DEBE_VISTA_CT	Número de operaciones al debe de productos a la vista como cualquier titular
OP_HABER_VISTA_CT	Número de operaciones al haber de productos a la vista como cualquier titular
IMP_DEBE_VISTA_CT	Importe de operaciones al debe de productos a la vista como cualquier titular
IMP_HABER_VISTA_CT	Importe de operaciones al haber de productos a la vista como cualquier titular
NUM_RECIBOS_AGUA_CT	Número de recibos de agua domiciliados como cualquier titular
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
NUM_RECIBOS_TELEFONO_CT	Número de recibos de teléfono domiciliados como cualquier titular
NUM_RECIBOS_ENERGIA_CT	Número de recibos de energía domiciliados como cualquier titular
NUM_RECIBOS_COMUNIDAD_CT	Número de recibos de comunidad domiciliados como cualquier titular
NUM_RECIBOS_ALQUILER_CT	Número de recibos de alquiler domiciliados como cualquier titular
NUM_RECIBOS_COLEGIOS_CT	Número de recibos de colegios domiciliados como cualquier titular
NUM_RECIBOS_SS_CT	Número de recibos de seguridad social domiciliados como cualquier titular
NUM_RECIBOS_IVA_CT	Número de recibos de iva domiciliados como cualquier titular
NUM_RECIBOS_SM_CT	Número de recibos de seguro médico domiciliados como cualquier titular
NUM_RECIBOS_EF_CT	Número de recibos de otras entidades financieras domiciliados como cualquier titular
NUM_RECIBOS_IMPUESTOS_CT	Número de recibos de impuestos domiciliados como cualquier titular
NUM_RECIBOS_RESTO_CT	Número de recibos del resto domiciliados como cualquier titular
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
IMP_RECIBOS_AGUA_CT	Importe de recibos de agua domiciliados como cualquier titular
IMP_RECIBOS_TELEFONO_CT	Importe de recibos de teléfono domiciliados como cualquier titular
IMP_RECIBOS_ENERGIA_CT	Importe de recibos de energía domiciliados como cualquier titular
IMP_RECIBOS_COMUNIDAD_CT	Importe de recibos de comunidad domiciliados como cualquier titular
IMP_RECIBOS_ALQUILER_CT	Importe de recibos de alquiler domiciliados como cualquier titular
IMP_RECIBOS_COLEGIOS_CT	Importe de recibos de colegios domiciliados como cualquier titular
IMP_RECIBOS_SS_CT	Importe de recibos de seguridad social domiciliados como cualquier titular
IMP_RECIBOS_IVA_CT	Importe de recibos de iva domiciliados como cualquier titular
IMP_RECIBOS_SM_CT	Importe de recibos de seguro médico domiciliados como cualquier titular
IMP_RECIBOS_EF_CT	Importe de recibos de otras entidades financieras domiciliados como cualquier titular
IMP_RECIBOS_IMPUESTOS_CT	Importe de recibos de impuestos domiciliados como cualquier titular
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
IMP_RECIBOS_RESTO_CT	Importe de recibos del resto domiciliados como cualquier titular
IMP_NOMINA	Importe de ingresos por nóminas
IMP_PENSION	Importe de ingresos por pensiones
IMP_DESEMPLEO	Importe de ingresos por desempleo
IMP_TRANS_PERIODICAS	Importe de ingresos por transferencias periódicas
IMP_INCAPACIDAD	Importe de ingresos por incapacidad
IMP_ORD_EXTRANJERO	Importe de ingresos por ordenes del extranjero
OP_TPV_IMPOSICION_TJCR	Número de operaciones de imposición en TPV para tarjetas de crédito
OP_CAJ_IMPOSICION_TJCR	Número de operaciones de abono en TPV para tarjetas de crédito
OP_TPV_ABONO_TJCR	Número de operaciones de imposición en cajeros para tarjetas de crédito
OP_CAJ_ABONO_TJCR	Número de operaciones de abono en cajeros para tarjetas de crédito
IMP_TPV_IMPOSICION_TJCR	Importe de operaciones de imposición en TPV para tarjetas de crédito
IMP_CAJ_IMPOSICION_TJCR	Importe de operaciones de abono en TPV para tarjetas de crédito
IMP_TPV_ABONO_TJCR	Importe de operaciones de imposición en cajeros para tarjetas de crédito
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
IMP_CAJ_ABONO_TJCR	Importe de operaciones de abono en cajeros para tarjetas de crédito
OP_TPV_IMPOSICION_TJDB	Número de operaciones de imposición en TPV para tarjetas de débito
OP_TPV_ABONO_TJDB	Número de operaciones de abono en TPV para tarjetas de débito
OP_CAJ_IMPOSICION_TJDB	Número de operaciones de imposición en cajeros para tarjetas de débito
OP_CAJ_ABONO_TJDB	Número de operaciones de abono en cajeros para tarjetas de débito
IMP_TPV_IMPOSICION_TJDB	Importe de operaciones de imposición en TPV para tarjetas de débito
IMP_TPV_ABONO_TJDB	Importe de operaciones de abono en TPV para tarjetas de débito
IMP_CAJ_IMPOSICION_TJDB	Importe de operaciones de imposición en cajeros para tarjetas de débito
IMP_CAJ_ABONO_TJDB	Importe de operaciones de abono en cajeros para tarjetas de débito
SALDO_TARJ_CREDITO_PT	Saldo de las tarjetas de crédito como primer titular
SALDO_DISP_TARJ_CREDITO_PT	Saldo disponible de las tarjetas de crédito como primer titular
SALDO_MEDIO_FOND_CONS_PT	Saldo medio de los fondos conservadores del cliente
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
	como primer titular
SALDO_FINAL_FOND_CONS_PT	Saldo final de los fondos conservadores del cliente como primer titular
SALDO_MEDIO_FOND_CONS_CT	Saldo medio de los fondos conservadores del cliente como cualquier titular
SALDO_FINAL_FOND_CONS_CT	Saldo final de los fondos conservadores del cliente como cualquier titular
SALDO_MEDIO_FOND_RIESGO_PT	Saldo medio de los fondos de riesgo del cliente como primer titular
SALDO_FINAL_FOND_RIESGO_PT	Saldo final de los fondos de riesgo del cliente como primer titular
SALDO_MEDIO_FOND_RIESGO_CT	Saldo medio de los fondos de riesgo del cliente como cualquier titular
SALDO_FINAL_FOND_RIESGO_CT	Saldo final de los fondos de riesgo del cliente como cualquier titular
SALDO_MEDIO_VALORES_PT	Saldo medio de los valores del cliente como primer titular
SALDO_FINAL_VALORES_PT	Saldo final de los valores del cliente como primer titular
SALDO_MEDIO_VALORES_CT	Saldo medio de los valores del cliente como cualquier titular
SALDO_FINAL_VALORES_CT	Saldo final de los valores del cliente como cualquier
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
	titular
SALDO_MEDIO_PLANES_PT	Saldo medio de los planes del cliente como primer titular
SALDO_FINAL_PLANES_PT	Saldo final de los planes del cliente como primer titular
SALDO_MEDIO_PLANES_CT	Saldo medio de los planes del cliente como cualquier titular
SALDO_FINAL_PLANES_CT	Saldo final de los planes del cliente como cualquier titular
SALDO_MEDIO_SEGUROS_AHORRO_PT	Saldo medio de los seguros de ahorro del cliente como primer titular
SALDO_FINAL_SEGUROS_AHORRO_PT	Saldo final de los seguros de ahorro del cliente como primer titular
SALDO_MEDIO_SEGUROS_AHORRO_CT	Saldo medio de los seguros de ahorro del cliente como cualquier titular
SALDO_FINAL_SEGUROS_AHORRO_CT	Saldo final de los seguros de ahorro del cliente como cualquier titular
SALDO_MEDIO_IPFS_CORTO_PT	Saldo medio de los depositos a corto plazo del cliente como primer titular
SALDO_FINAL_IPFS_CORTO_PT	Saldo final de los depositos a corto plazo del cliente como primer titular
TIPO_MEDIO_IPFS_CORTO_PT	Tipo medio de los depositos a corto plazo del cliente
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
	como primer titular
SALDO_MEDIO_IPFS_CORTO_CT	Saldo medio de los depositos a corto plazo del cliente como cualquier titular
SALDO_FINAL_IPFS_CORTO_CT	Saldo final de los depositos a corto plazo del cliente como cualquier titular
TIPO_MEDIO_IPFS_CORTO_CT	Tipo medio de los depositos a corto plazo del cliente como cualquier titular
SALDO_MEDIO_IPFS_LARGO_PT	Saldo medio de los depositos a largo plazo del cliente como primer titular
SALDO_FINAL_IPFS_LARGO_PT	Saldo final de los depositos a largo plazo del cliente como primer titular
TIPO_MEDIO_IPFS_LARGO_PT	Tipo medio de los depositos a largo plazo del cliente como primer titular
SALDO_MEDIO_IPFS_LARGO_CT	Saldo medio de los depositos a largo plazo del cliente como cualquier titular
SALDO_FINAL_IPFS_LARGO_CT	Saldo final de los depositos a largo plazo del cliente como cualquier titular
TIPO_MEDIO_IPFS_LARGO_CT	Tipo medio de los depositos a largo plazo del cliente como cualquier titular
SALDO_MEDIO_AHORRO_VIV_PT	Saldo medio de los depositos ahorro vivienda del cliente como primer titular
SALDO_FINAL_AHORRO_VIV_PT	Saldo final de los depositos ahorro vivienda del cliente
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
	como primer titular
TIPO_MEDIO_AHORRO_VIV_PT	Tipo medio de los depositos ahorro vivienda del cliente como primer titular
SALDO_MEDIO_AHORRO_VIV_CT	Saldo medio de los depositos ahorro vivienda del cliente como cualquier titular
SALDO_FINAL_AHORRO_VIV_CT	Saldo final de los depositos ahorro vivienda del cliente como cualquier titular
TIPO_MEDIO_AHORRO_VIV_CT	Tipo medio de los depositos ahorro vivienda del cliente como cualquier titular
SALDO_MEDIO_RESTO_PASIVO_PT	Saldo medio del resto de pasivo del cliente como primer titular
SALDO_FINAL_RESTO_PASIVO_PT	Saldo final del resto de pasivo del cliente como primer titular
TIPO_MEDIO_RESTO_PASIVO_PT	Tipo medio del resto de pasivo del cliente como primer titular
SALDO_MEDIO_RESTO_PASIVO_CT	Saldo medio del resto de pasivo del cliente como cualquier titular
SALDO_FINAL_RESTO_PASIVO_CT	Saldo final del resto de pasivo del cliente como cualquier titular
TIPO_MEDIO_RESTO_PASIVO_CT	Tipo medio del resto de pasivo del cliente como cualquier titular
SALDO_PDTE_PTMOS_GP_CON_PT	Saldo pendiente de los préstamos de garantía personal
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
	de consumo del cliente como primer titular
CUOTA_PTMOS_GP_CON_PT	Cuota de los préstamos de garantía personal de consumo del cliente como primer titular
TIPOMEDIO_PTMOS_GP_CON_PT	Tipo medio de los préstamos de garantía personal de consumo del cliente como primer titular
SALDO_PDTE_PTMOS_GP_VIV_PT	Saldo pendiente de los préstamos de garantía personal de viviendas del cliente como primer titular
CUOTA_PTMOS_GP_VIV_PT	Cuota de los préstamos de garantía personal de viviendas del cliente como primer titular
TIPOMEDIO_PTMOS_GP_VIV_PT	Tipo medio de los préstamos de garantía personal de viviendas del cliente como primer titular
SALDO_PDTE_PTMOS_GP_INV_PT	Saldo pendiente de los préstamos de garantía personal de inversión del cliente como primer titular
CUOTA_PTMOS_GP_INV_PT	Cuota de los préstamos de garantía personal de inversión del cliente como primer titular
TIPOMEDIO_PTMOS_GP_INV_PT	Tipo medio de los préstamos de garantía personal de inversión del cliente como primer titular
SALDO_PDTE_PTMOS_GP_NEG_PT	Saldo pendiente de los préstamos de garantía personal de negocio del cliente como primer titular
CUOTA_PTMOS_GP_NEG_PT	Cuota de los préstamos de garantía personal de negocio del cliente como primer titular
TIPOMEDIO_PTMOS_GP_NEG_PT	Tipo medio de los préstamos de garantía personal de
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
	negocio del cliente como primer titular
SALDO_PDTE_PTMOS_GP_OTR_PT	Saldo pendiente de los préstamos de garantía personal de <i>otros</i> como primer titular
CUOTA_PTMOS_GP_OTR_PT	Cuota de los préstamos de garantía personal de <i>otros</i> del cliente como primer titular
TIPOMEDIO_PTMOS_GP_OTR_PT	Tipo medio de los préstamos de garantía personal de <i>otros</i> del cliente como primer titular
SALDO_PDTE_PTMOS_GP_CON_CT	Saldo pendiente de los préstamos de garantía personal de consumo del cliente como cualquier titular
CUOTA_PTMOS_GP_CON_CT	Cuota de los préstamos de garantía personal de consumo del cliente como cualquier titular
TIPOMEDIO_PTMOS_GP_CON_CT	Tipo medio de los préstamos de garantía personal de consumo del cliente como cualquier titular
SALDO_PDTE_PTMOS_GP_VIV_CT	Saldo pendiente de los préstamos de garantía personal de viviendas del cliente como cualquier titular
CUOTA_PTMOS_GP_VIV_CT	Cuota de los préstamos de garantía personal de viviendas del cliente como cualquier titular
TIPOMEDIO_PTMOS_GP_VIV_CT	Tipo medio de los préstamos de garantía personal de viviendas del cliente como cualquier titular
SALDO_PDTE_PTMOS_GP_INV_CT	Saldo pendiente de los préstamos de garantía personal de inversión del cliente como cualquier titular
CUOTA_PTMOS_GP_INV_CT	Cuota de los préstamos de garantía personal de inversión
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
	del cliente como cualquier titular
TIPOMEDIO_PTMOS_GP_INV_CT	Tipo medio de los préstamos de garantía personal de inversión del cliente como cualquier titular
SALDO_PDTE_PTMOS_GP_NEG_CT	Saldo pendiente de los préstamos de garantía personal de negocio del cliente como cualquier titular
CUOTA_PTMOS_GP_NEG_CT	Cuota de los préstamos de garantía personal de negocio del cliente como cualquier titular
TIPOMEDIO_PTMOS_GP_NEG_CT	Tipo medio de los préstamos de garantía personal de negocio del cliente como cualquier titular
SALDO_PDTE_PTMOS_GP_OTR_CT	Saldo pendiente de los préstamos de garantía personal de <i>otros</i> como cualquier titular
CUOTA_PTMOS_GP_OTR_CT	Cuota de los préstamos de garantía personal de <i>otros</i> del cliente como cualquier titular
TIPOMEDIO_PTMOS_GP_OTR_CT	Tipo medio de los préstamos de garantía personal de <i>otros</i> del cliente como cualquier titular
SALDO_PDTE_PTMOS_GR_CON_PT	Saldo pendiente de los préstamos de garantía real de consumo del cliente como primer titular
CUOTA_PTMOS_GR_CON_PT	Cuota de los préstamos de garantía real de consumo del cliente como primer titular
TIPOMEDIO_PTMOS_GR_CON_PT	Tipo medio de los préstamos de garantía real de consumo del cliente como primer titular
SALDO_PDTE_PTMOS_GR_VIV_PT	Saldo pendiente de los préstamos de garantía real de
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
	viviendas del cliente como primer titular
CUOTA_PTMOS_GR_VIV_PT	Cuota de los préstamos de garantía real de viviendas del cliente como primer titular
TIPOMEDIO_PTMOS_GR_VIV_PT	Tipo medio de los préstamos de garantía real de viviendas del cliente como primer titular
SALDO_PDTE_PTMOS_GR_INV_PT	Saldo pendiente de los préstamos de garantía real de inversión del cliente como primer titular
CUOTA_PTMOS_GR_INV_PT	Cuota de los préstamos de garantía real de inversión del cliente como primer titular
TIPOMEDIO_PTMOS_GR_INV_PT	Tipo medio de los préstamos de garantía real de inversión del cliente como primer titular
SALDO_PDTE_PTMOS_GR_NEG_PT	Saldo pendiente de los préstamos de garantía real de negocio del cliente como primer titular
CUOTA_PTMOS_GR_NEG_PT	Cuota de los préstamos de garantía real de negocio del cliente como primer titular
TIPOMEDIO_PTMOS_GR_NEG_PT	Tipo medio de los préstamos de garantía real de negocio del cliente como primer titular
SALDO_PDTE_PTMOS_GR_OTR_PT	Saldo pendiente de los préstamos de garantía real de <i>otros</i> como primer titular
CUOTA_PTMOS_GR_OTR_PT	Cuota de los préstamos de garantía real de <i>otros</i> del cliente como primer titular
TIPOMEDIO_PTMOS_GR_OTR_PT	Tipo medio de los préstamos de garantía real de <i>otros</i>
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
	del cliente como primer titular
SALDO_PDTE_PTMOS_GR_CON_CT	Saldo pendiente de los préstamos de garantía real de consumo del cliente como cualquier titular
CUOTA_PTMOS_GR_CON_CT	Cuota de los préstamos de garantía real de consumo del cliente como cualquier titular
TIPOMEDIO_PTMOS_GR_CON_CT	Tipo medio de los préstamos de garantía real de consumo del cliente como cualquier titular
SALDO_PDTE_PTMOS_GR_VIV_CT	Saldo pendiente de los préstamos de garantía real de viviendas del cliente como cualquier titular
CUOTA_PTMOS_GR_VIV_CT	Cuota de los préstamos de garantía real de viviendas del cliente como cualquier titular
TIPOMEDIO_PTMOS_GR_VIV_CT	Tipo medio de los préstamos de garantía real de viviendas del cliente como cualquier titular
SALDO_PDTE_PTMOS_GR_INV_CT	Saldo pendiente de los préstamos de garantía real de inversión del cliente como cualquier titular
CUOTA_PTMOS_GR_INV_CT	Cuota de los préstamos de garantía real de inversión del cliente como cualquier titular
TIPOMEDIO_PTMOS_GR_INV_CT	Tipo medio de los préstamos de garantía real de inversión del cliente como cualquier titular
SALDO_PDTE_PTMOS_GR_NEG_CT	Saldo pendiente de los préstamos de garantía real de negocio del cliente como cualquier titular
CUOTA_PTMOS_GR_NEG_CT	Cuota de los préstamos de garantía real de negocio del
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
	cliente como cualquier titular
TIPOMEDIO_PTMOS_GR_NEG_CT	Tipo medio de los préstamos de garantía real de negocio del cliente como cualquier titular
SALDO_PDTE_PTMOS_GR_OTR_CT	Saldo pendiente de los préstamos de garantía real de <i>otros</i> como cualquier titular
CUOTA_PTMOS_GR_OTR_CT	Cuota de los préstamos de garantía real de <i>otros</i> del cliente como cualquier titular
TIPOMEDIO_PTMOS_GR_OTR_CT	Tipo medio de los préstamos de garantía real de <i>otros</i> del cliente como cualquier titular
SALDO_DISPUESTO_CREDITO_PT	Saldo dispuesto para los créditos del cliente como primer titular
LIMITE_CREDITO_PT	Límite para los créditos del cliente como primer titular
TIPO_MEDIO_CREDITO_PT	Tipo medio para los créditos del cliente como primer titular
SALDO_DISPUESTO_CREDITO_CT	Saldo dispuesto para los créditos del cliente como cualquier titular
LIMITE_CREDITO_CT	Límite para los créditos del cliente como cualquier titular
TIPO_MEDIO_CREDITO_CT	Tipo medio para los créditos del cliente como cualquier titular
SALDO_DISPUESTO_DTO_PT	Saldo dispuesto para los descuentos del cliente como primer titular
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
LIMITE_DTO_PT	Límite para los descuentos del cliente como primer titular
TIPO_MEDIO_DTO_PT	Tipo medio para los descuentos del cliente como primer titular
SALDO_DISPUESTO_DTO_CT	Saldo dispuesto para los descuentos del cliente como cualquier titular
LIMITE_DTO_CT	Límite para los descuentos del cliente como cualquier titular
TIPO_MEDIO_DTO_CT	Tipo medio para los descuentos del cliente como cualquier titular
SALDO_DISPUESTO_OTROS_ACT_PT	Saldo dispuesto para el resto de activo del cliente como primer titular
LIMITE_OTROS_ACT_PT	Límite para el resto de activo del cliente como primer titular
TIPO_MEDIO_OTROS_ACT_PT	Tipo medio para el resto de activo del cliente como primer titular
SALDO_DISPUESTO_OTROS_ACT_CT	Saldo dispuesto para el resto de activo del cliente como cualquier titular
LIMITE_OTROS_ACT_CT	Límite para el resto de activo del cliente como cualquier titular
TIPO_MEDIO_OTROS_ACT_CT	Tipo medio para el resto de activo del cliente como cualquier titular
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
SALDO_PRINCIPAL_AVALES_PT	Saldo principal de los avales del cliente como primer titular
COMISIONES_MES_AVALES_PT	Comisiones mensuales de los avales del cliente como primer titular
SALDO_PRINCIPAL_AVALES_CT	Saldo principal de los avales del cliente como culaquier titular
COMISIONES_MES_AVALES_CT	Comisiones mensuales de los avales del cliente como culaquier titular
NUM_SEG_HOGAR	Número de seguros del hogar del cliente
PRIMA_SEG_HOGAR	Prima de los seguros del hogar del cliente
NUM_SEG_HOGAR_V	Número de seguros del hogar vinculados al cliente
PRIMA_SEG_HOGAR_V	Prima de los seguros del hogar vinculados al cliente
NUM_SEG_VIDA	Número de seguros de vida del cliente
PRIMA_SEG_VIDA	Prima de los seguros de vida del cliente
NUM_SEG_VIDA_V	Número de seguros de vida vinculados al cliente
PRIMA_SEG_VIDA_v	Prima de los seguros de vida vinculados al cliente
NUM_SEG_COCHE	Número de seguros de coche del cliente
PRIMA_SEG_COCHE	Prima de los seguros de coche del cliente
NUM_SEG_OTROS	Número de otros seguros del cliente
PRIMA_SEG_OTROS	Prima de de otros seguros del cliente
NUM_TARIFA_PLANA	Número de productos de tarifa plana que tiene contratados cada cliente en el mes de carga actual
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
IMP_TARIFA_PLANA	Importe de productos de tarifa plana que tiene contratados cada cliente en el mes de carga actual
DEUDA_PDTE	Se calcula la deuda vencida a último día de mes para todos los contratos vigentes que tenga ese cliente
NUM_ENTRADAS_DESCUBIERTO	Número de entradas de descubiertos de RAJE (Quitando créditos) del cliente
DIAS_DESCUBIERTO	Días que lleva con descubiertos de RAJE (Quitando créditos) del cliente
FC_ULT_DESCUBIERTO	Fecha del último descubierto de RAJE (Quitando créditos) del cliente
RENTA_ESTIMADA	Renta estimada del cliente en función del segmento al que pertenece
RECORRIDO	Recorrido del cliente en función del segmento al que pertenece
IDEAL	Ideal del cliente en función del segmento al que pertenece
MO_CLIENTE_PT	Margen Ordinario del cliente en sus contratos como primer titular
MO_ACTIVO_PT	Margen Ordinario del cliente en sus contratos de activo como primer titular
MO_PASIVO_PT	Margen Ordinario del cliente en sus contratos de pasivo como primer titular
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
MO_FB_PT	Margen Ordinario del cliente en sus contratos de fuera de balance como primer titular
MO_SEGUROS_PT	Margen Ordinario del cliente en sus contratos de seguros como primer titular
MO_TARJETAS_PT	Margen Ordinario del cliente en sus contratos de tarjetas como primer titular
MO_CLIENTE_CT	Margen Ordinario del cliente en sus contratos como cualquier titular
MO_ACTIVO_CT	Margen Ordinario del cliente en sus contratos de activo como cualquier titular
MO_PASIVO_CT	Margen Ordinario del cliente en sus contratos de pasivo como cualquier titular
MO_FB_CT	Margen Ordinario del cliente en sus contratos de fuera de balance como cualquier titular
MO_SEGUROS_CT	Margen Ordinario del cliente en sus contratos de seguros como cualquier titular
MO_TARJETAS_CT	Margen Ordinario del cliente en sus contratos de tarjetas como cualquier titular
RF_CLIENTE_PT	Rentabilidad final del cliente como primer titular
MI_CLIENTE_PT	Margen de intermediación del cliente en sus contratos como primer titular
RF_CLIENTE_CT	Rentabilidad final del cliente como cualquier titular
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
MI_CLIENTE_CT	Margen de intermediación del cliente en sus contratos como cualquier titular
SALDO_MEDIO_ACTIVADO_PT	Saldo medio de los contratos de activo como primer titular
SALDO_MEDIO_PASIVO_PT	Saldo medio de los contratos de pasivo como primer titular
SALDO_MEDIO_FB_PT	Saldo medio de los contratos de fuera de balance como primer titular
SALDO_MEDIO_SEGUROS_PT	Saldo medio de los contratos de seguros como primer titular
SALDO_MEDIO_TARJETAS_PT	Saldo medio de los contratos de tarjetas como primer titular
SALDO_MEDIO_CLIENTE_CT	Saldo medio de todos los contratos del cliente como cualquier titular
SALDO_MEDIO_ACTIVADO_CT	Saldo medio de los contratos de activo como cualquier titular
SALDO_MEDIO_PASIVO_CT	Saldo medio de los contratos de pasivo como cualquier titular
SALDO_MEDIO_FB_CT	Saldo medio de los contratos de fuera de balance como cualquier titular
SALDO_MEDIO_SEGUROS_CT	Saldo medio de los contratos de seguros como cualquier titular
Continúa en la siguiente página	

Cuadro A.3 – continuado de la página anterior

Campo	Descripción
SALDO_MEDIO_TARJETAS_CT	Saldo medio de los contratos de tarjetas como cualquier titular

Cuadro A.3: Lista de campos disponible en el data mart de clientes.

A.9. Gráficos de las variables finales

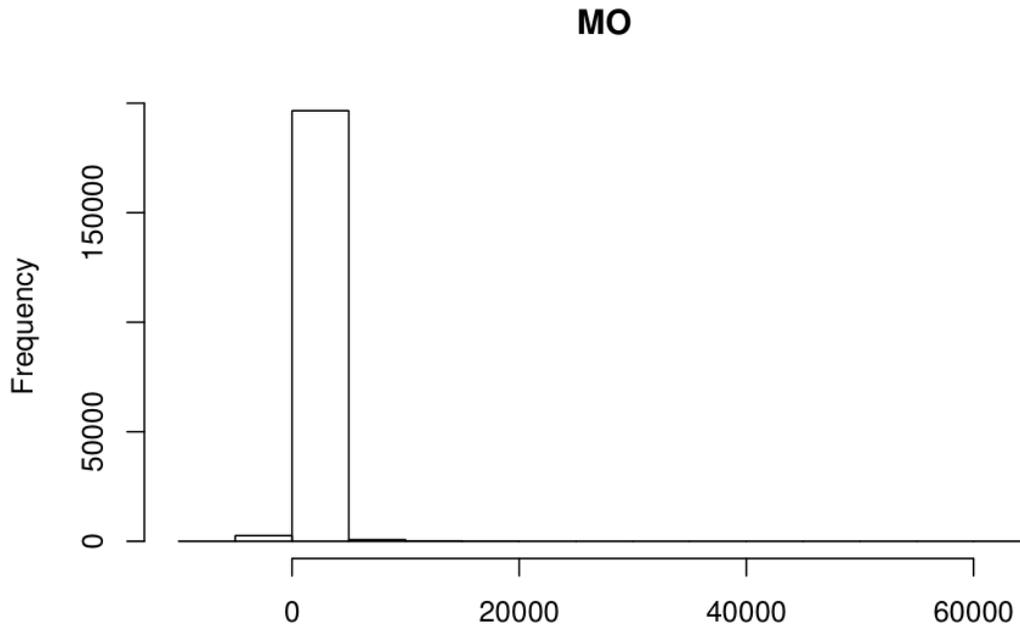


Figura A.68: Histograma de la variable MO con una muestra de 200.000 clientes.

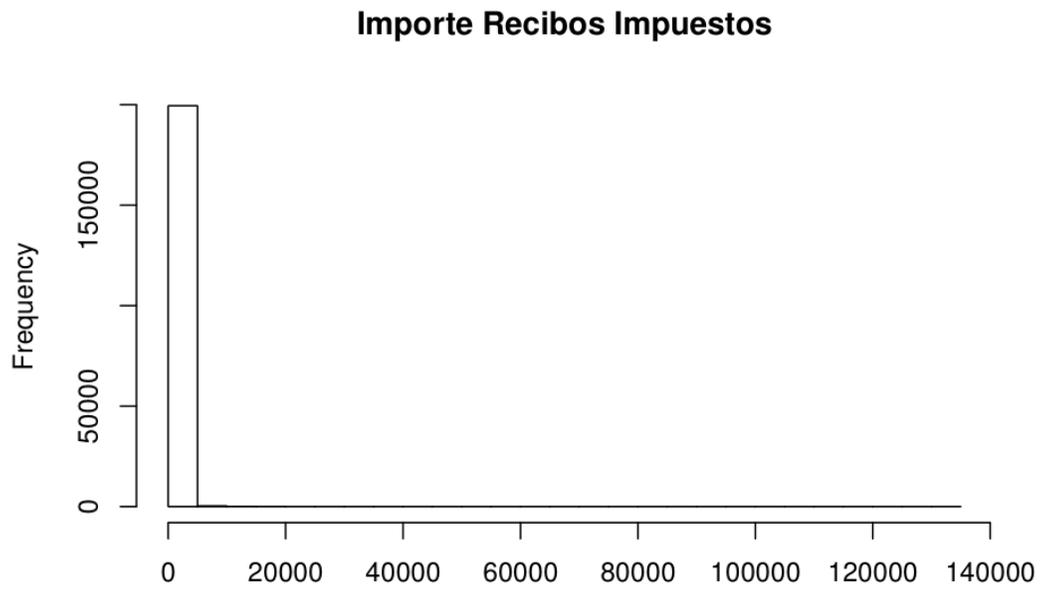


Figura A.69: Histograma de la variable `impRecibImpuestos` con una muestra de 200.000 clientes.

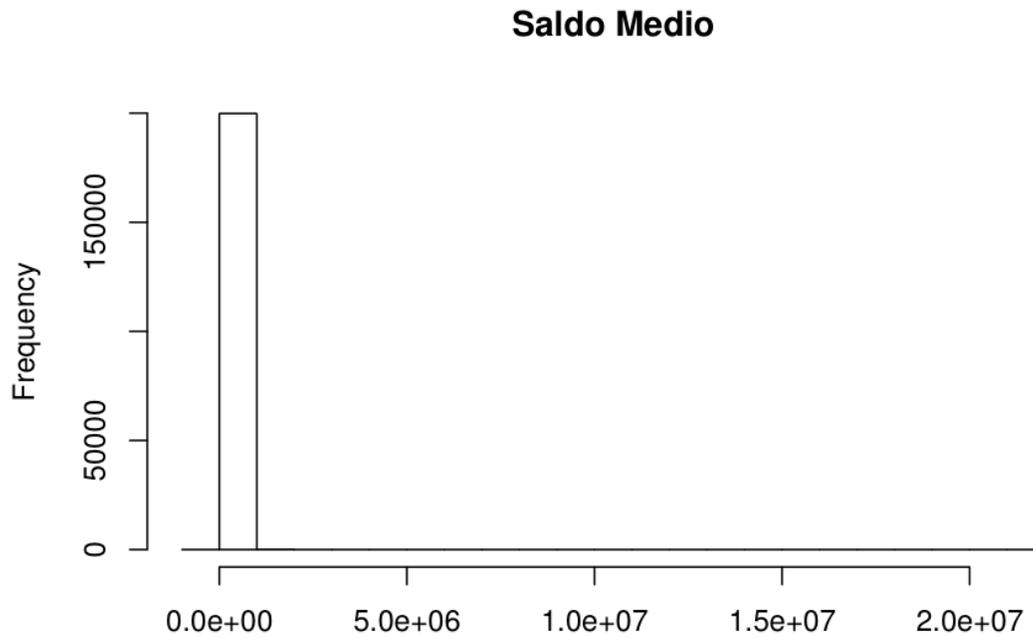


Figura A.70: Histograma de la variable `SaldoMedio` con una muestra de 200.000 clientes.

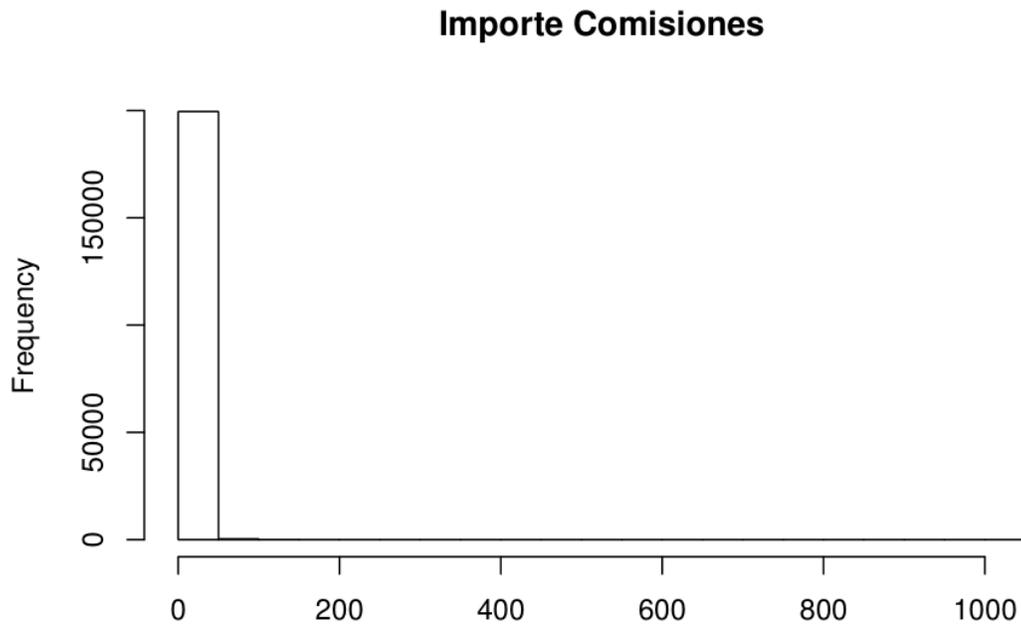


Figura A.71: Histograma de la variable ImpComisiones con una muestra de 200.000 clientes.

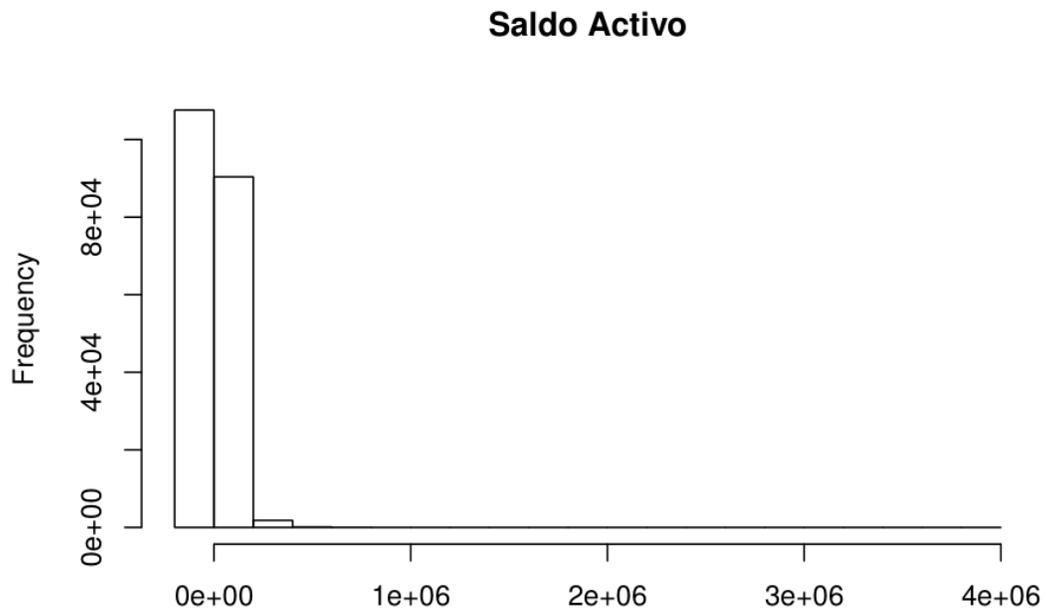


Figura A.72: Histograma de la variable SaldoActivo con una muestra de 200.000 clientes.

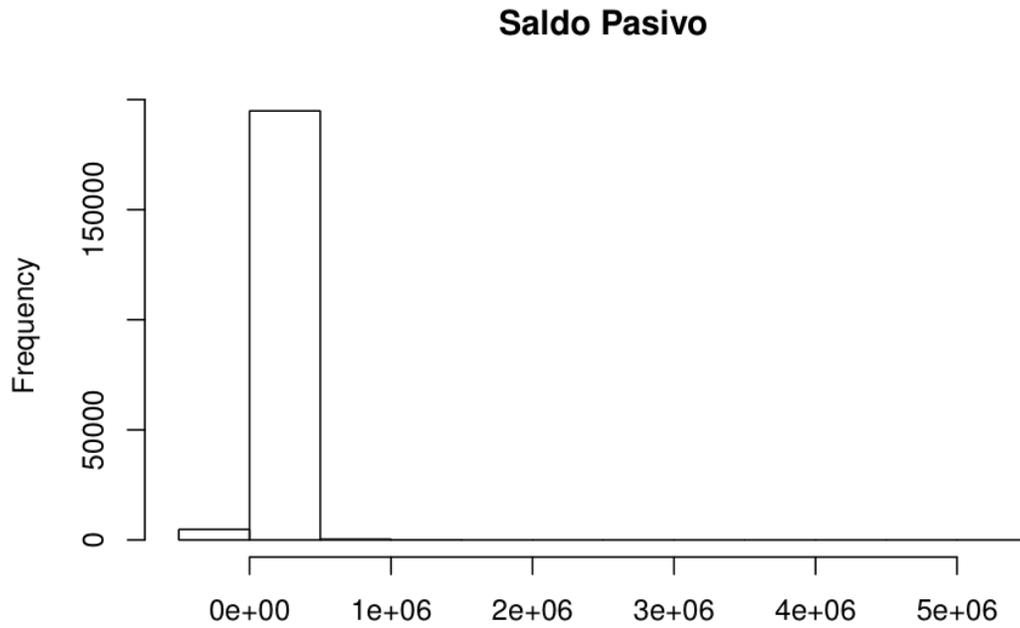


Figura A.73: Histograma de la variable SaldoPasivo con una muestra de 200.000 clientes.

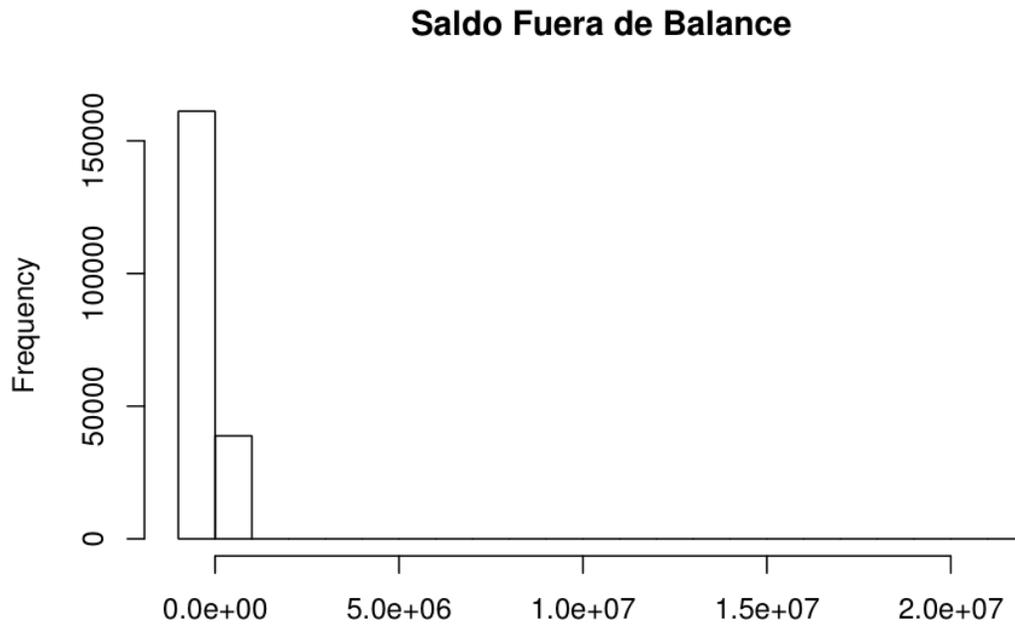


Figura A.74: Histograma de la variable SaldoFB con una muestra de 200.000 clientes.

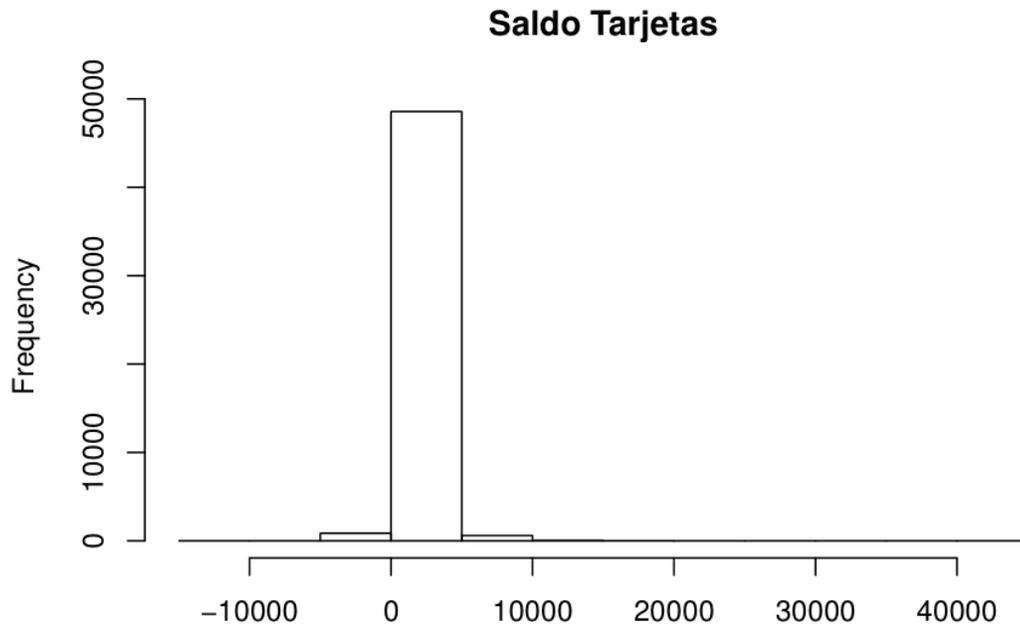


Figura A.75: Histograma de la variable SaldoTarjetas con una muestra de 200.000 clientes.

A.10. Aplicaciones que influyen en el margen ordinario de la entidad

Aplicación	Descripción
ADEU	DOMICILIACIONES DE ADEUDOS
AVAL	AVALES
CTOS	CREDITOS
CXTR	COMERCIO EXTERIOR
DACT	DEUDA ANOTACION EN CUENTA
DOMI	DOMICILIACIONES (NUEVA)
EALE	ENVIO ALERTAS Y SERVICIOS MOVILES
ECOM	LINEAS DE DESCUENTO - EFECTOS COMERCIALES
EEXP	EXPEDIENTE ELECTRONICO

Continúa en la siguiente página

Cuadro A.4 – continuado de la página anterior

Aplicación	Descripción
FGCL	FACTORING A CLIENTES
FPRO	FACTORING A PROVEEDORES
FTES	FONDOS INVERSIÓN
FX0C	PROD: FXCU - GINEBRA CAPTACION
FX0I	PROD: FXCU - GINEBRA INVERSION
FX1I	PROD: FXCU - MIAMI INVERSION
FX2I	PROD: FXCU - MIAMI 2 INVERSION
GESI	BANCA ELECTRÓNICA DE EMPRESAS
HICT	CEDULAS HIPOTECARIAS
ITAL	INTERNACIONAL
LEAS	LEASING
LITI	ASUNTOS EN LITIGIO
LMUR	LINEAS MULTIRIESGO
LTEX	CREDITOS DOCUMENTARIOS EXPORTACION
LTTR	CREDITOS DOCUMENTARIOS MULTIPLATAFORMA
MPAG	MEDIOS DE PAGO - TARJETAS
OPCA	OPERATORIA PORTUGAL
OPIN	OPERATORIA PORTUGAL
PFLE	PRODUCTOS FLEXIBLES
PITA	PRESTAMOS INTERNACIONAL
PJNN	PAGARES REAL DECRETO 874
PLAZ	PLAZO
PMOS	PRESTAMOS
Continúa en la siguiente página	

Cuadro A.4 – continuado de la página anterior

Aplicación	Descripción
RAJE	RECURSOS AJENOS
RCOB	REMESAS GESTION COBRO EXTRANJERO
RTNG	RENTING DE INTERMEDIACIÓN
RXTR	REMESAS EXTRANJERAS DE IMPORTACION
SCIA	SEGUROS BIAGALICIA
SFEM	SEGUROS MEDIACIÓN
SGUR	SEGUROS
SJCW	SEGUROS CASER TPNET (WEB)
SMDD	GESTION DE COBERTURAS DE TIPO DE INTERÉS
SNYA	AVALES SINDICADOS
SNYD	PRESTAMOS SINDICADOS ODS
SRAD	CAJAS DE SEGURIDAD
SVAR	SEGUROS VARIOS
SVNM	SEGUROS VARIOS
TGAL	TASACIONES BIENES
TJCR	TARJETAS DE CREDITO
TJDB	TARJETAS
TVEN	TERMINALES DE VENTA
VLOR	VALORES
XSCI	XESTION DE CARTEIRAS DE INVERSIÓN
ZSEG	COBERTURAS DE TIPO DE CAMBIO

Cuadro A.4: Lista de aplicaciones que influyen en el margen ordinario de la entidad.

A.11. Dendrogramas obtenidos en la sección 4.3 con una muestra de 100 clientes

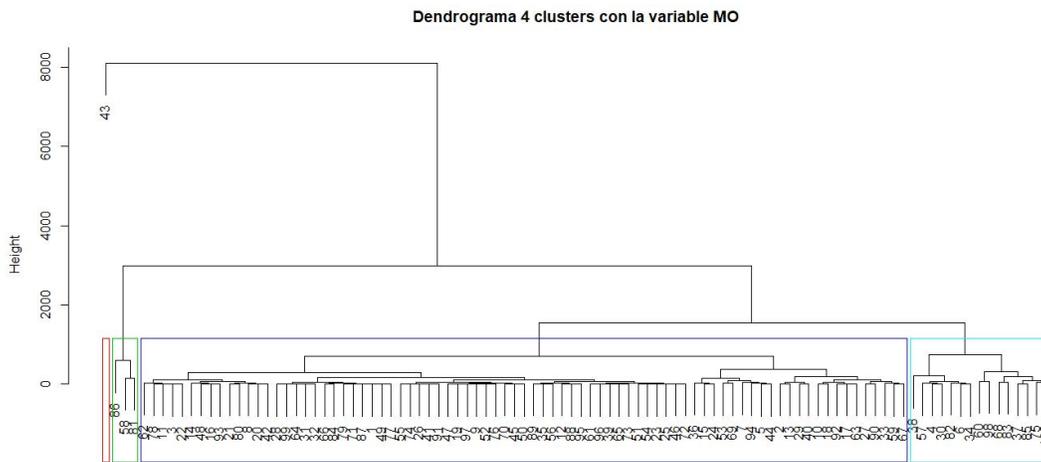


Figura A.76: Dendrograma obtenido con una muestra de 100 clientes a partir de la variable MO seleccionando 4 clusters.

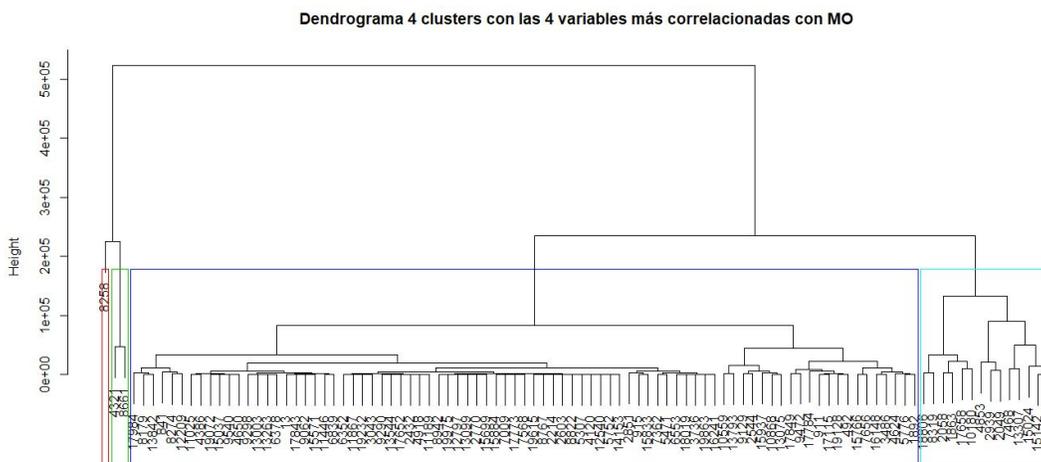


Figura A.77: Dendrograma obtenido con una muestra de 100 clientes a partir de las 4 variables más correlacionadas con MO seleccionando 4 clusters.

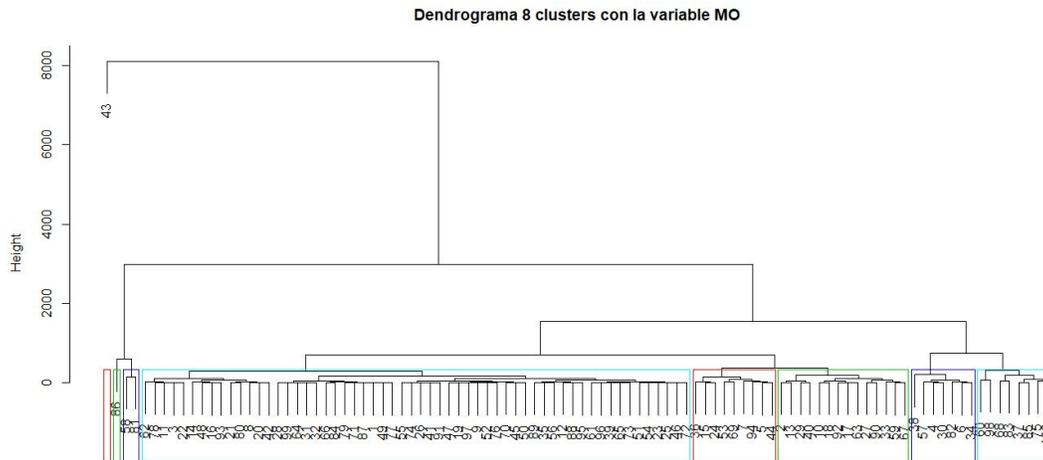


Figura A.78: Dendrograma obtenido con una muestra de 100 clientes a partir de la variable MO seleccionando 8 clusters.

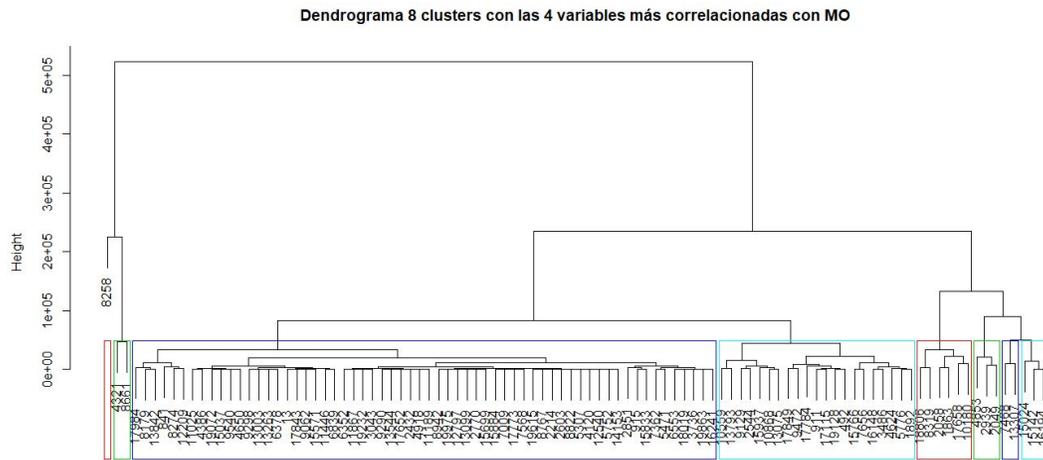


Figura A.79: Dendrograma obtenido con una muestra de 100 clientes a partir de las 4 variables más correlacionadas con MO seleccionando 8 clusters.

Bibliografía

- [1] Banfield J, Raftery AE (1993) Model-based Gaussian and non-Gaussian clustering. *Biometrics*. 49: 803-821.
- [2] Breiman, L (2001) Random Forests. *Machine Learning* 45, 5-32.
- [3] Breiman L, Friedman J, Olshen R, Stone C (1984) *Classification and Regression Trees*. Wadsworth.
- [4] Celeux G, Govaert G (1995) Gaussian parsimonious clustering models. *Pattern Recognition*. 28: 781-793.
- [5] Dempster AP, Laird N, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39 (1): 1-38.
- [6] Edwards AL (1976) The Correlation Coefficient. Ch. 4 in *An Introduction to Linear Regression and Correlation*. San Francisco, CA: W. H. Freeman, pp. 33-46.
- [7] Ekinçi Y, Ulengin F, Uray N, Ulengin B (2014) Analysis of customer lifetime value and marketing expenditure decisions through a Markovian-based model. *European Journal of Operational Research*, 237:278-288.
- [8] Glady N, Baesens B, C. Croux (2009) Modeling churn using customer lifetime value. *European Journal of Operational Research*, 197:402-411.
- [9] Haenlein M, Kaplan AM, Beeser AJ (2007) A model to determine customer lifetime value in a retail banking context. *European Management Journal*, 25(3):221-234.
- [10] Hastie TJ, Tibshirani RJ (1990) *Generalized Additive Models*. Chapman-Hall.
- [11] Mardia KV, Kent JT, Bibby JM (1979) *Multivariate Analysis*. Academic Press.
- [12] McCullagh P, Nelder JA (1989) *Generalized Linear Models*. Chapman and Hall.
- [13] McLachlan G, Peel D (2000) *Finite Mixture Models*. Wiley; New York.
- [14] Seber GAF (1977) *Linear Regression Analysis*. New York: John Wiley & Sons.