



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



FACULTAD DE ADMINISTRACIÓN Y
DIRECCIÓN DE EMPRESAS. UPV

La inteligencia de negocios como una oportunidad clave para las empresas

Grado en Administración y Dirección de Empresas

Autor: José Legido Casanoves

Tutor: Francisco Guijarro Martínez

Curso Académico 2020-2021

Índice

1.Introducción	1
1.1 Resumen.....	2
1.2 Objetivos del trabajo	3
1.3 Metodología	3
2.Antecedentes	5
2.1 ¿Qué es la inteligencia de negocios?.....	6
2.2 Historia de la inteligencia de negocios.....	7
2.3 Características de la inteligencia de negocios	8
2.4 Componentes del business intelligence.....	9
2.5 Situación actual de la inteligencia de negocios	13
2.6 Beneficios de la inteligencia de negocios.....	14
3. El mercado inmobiliario.....	16
3.1 Definición y características	17
3.2 Tipos de vivienda	18
3.3 Evolución del mercado inmobiliario en España.....	20
3.4 Situación actual del mercado inmobiliario y perspectivas futuras	22
3.5 Evolución de los precios en el mercado inmobiliario	22
4. Análisis de datos	26
4.1 Preparación de la base de datos.....	27
4.2 Método no supervisado	29
4.2.1 Modelos jerárquicos	31
4.2.2 Modelos de partición.....	34
4.3 Selección y validación del modelo.....	39
4.4 Interpretación de los resultados.....	40
4.5 Método supervisado	45
4.5.1 Árbol de regresión.....	46
4.5.2 Random Forest	48
4.5.3 Vecino más próximo	51
4.5.4 Máquinas de soporte vectorial.....	52
4.6 Comparación y selección del modelo.....	54
5. Conclusión.....	57
ANEXO.....	58
Bibliografía	59

ÍNDICE DE GRÁFICOS

Gráfico 1: Pronóstico del tamaño del mercado de big data.....	8
Gráfico 2: Gráfico de operaciones de compraventa desde 2007-2015.....	21
Gráfico 3: Gráfico IPV.....	23
Gráfico 4: Gráfico de la evolución del precio del m2 en Madrid desde 2006-actualidad.....	23
Gráfico 5: Gráfico del precio del m2 de los barrios de Madrid (3ºtrimestre de 2020)...	24
Gráfico 6: Gráfico boxplot de las variables escaladas y centradas de la base de datos..	29
Gráfico 7: Mapa de color.....	30
Gráfico 8: Dendograma con el método ward.....	31
Gráfico 9: Dendograma con el método de la media.....	32
Gráfico 10: Gráfico del número óptimo de clusters a través del coef. Silhouette.....	33
Gráfico 11: Gráfico del número óptimo de clusters a través de la suma de cuadrados intracluster.....	33
Gráfico 12: Gráfico scores PCA.....	34
Gráfico 13: Gráfico del número óptimo de clusters a través del coef. Silhouette.....	35
Gráfico 14: Gráfico del número óptimo de clusters a través de la suma de cuadrados intracluster.....	35
Gráfico 15: Gráfico scores PCA.....	36
Gráfico 16: Gráfico de numero de clusters optimo a través del coef. Silhoutte.....	37
Gráfico 17: Gráfico del número óptimo de clusters a través de la suma de cuadrados intracluster.....	37
Gráfico 18: Gráfico scores PCA.....	38
Gráfico 19: Gráfico comparativo de los modelos de clustering.....	39
Gráfico 20: Gráfico de las dimensiones del modelo.....	40
Gráfico 21: Gráfico de la contribución de las variables a la primera dimensión.....	41
Gráfico 22: Gráfico de la contribución de las variables a la segunda dimensión.....	41
Gráfico 23: Gráfico de distribución de los clusters.....	42
Gráfico 24: Gráfico de distribución de las variables.....	42
Gráfico 25: Gráfico del perfil medio de los cluster.....	43

Gráfico 26: Árbol de regresión.....	46
Gráfico 27: Árbol de regresión podado	47
Gráfico 28: Árbol de regresión completo	48
Gráfico 29: Gráfico del MAPE de Random Forest	49
Gráfico 30: Gráfico de comparación de la predicción entre mtry=3 y mtry=6	49
Gráfico 31: Gráfico mean decrease accuracy	50
Gráfico 32: Gráfico de k óptimo.....	51
Gráfico 33: Gráfico de comparación de la predicción entre k=3 y K=10	52
Gráfico 34: Gráfico de la predicción del modelo frente a las observaciones	53
Gráfico 35: Gráfico de la predicción del modelo frente a las observaciones	53
Gráfico 36:Gráfico de las predicciones de todos los modelos frente a las observaciones	54

ÍNDICE DE TABLAS

Tabla 1: Estadístico de Hopkins	30
Tabla 2: Tabla de las aportaciones de cada dimensión al modelo.....	40
Tabla 3: Tabla de las diferencias de las medias entre ambos clusters.....	44
Tabla 4: Tabla de error de los diferentes arboles.....	47
Tabla 5: Tabla de la aportación de las variables al modelo.....	50
Tabla 6: Tabla de k óptimo.....	51
Tabla 7: CART	55
Tabla 8: Vecino más próximo	55
Tabla 9: Máquinas de soporte vectorial(“kernlab”)	55
Tabla 10: Random forest	55
Tabla 11: Máquinas de soporte vectorial(“e1071”).....	55

ÍNDICE DE ILUSTRACIONES

Ilustración 1: Computadora de 1ª generación.....	7
Ilustración 2: Componentes del BI	9
Ilustración 3: Ejemplos de herramientas de BI.....	12

1.INTRODUCCIÓN

1.1 RESUMEN

El presente trabajo final de grado consiste, en mostrar la importancia que tiene la inteligencia de negocios en las empresas, ya que puede ofrecer una gran cantidad de información valiosa que facilite la toma de decisiones, minimizando el error.

La inteligencia de negocios se está convirtiendo en una parte fundamental de las empresas. Cada vez más organizaciones se respaldan en los datos para diseñar sus planes estratégicos, en un mundo tan competitivo donde la diferencia es esencial. La inteligencia de negocios se puede aplicar en todos los sectores y ámbitos de las empresas.

Las organizaciones tienen enormes cantidades de datos, que en si no aportan mucha información, hace falta tratar los datos, para extraer la información deseada. A través del análisis de los datos, las empresas pueden analizar todos sus procesos, ya sean externos o internos, con el objetivo de tener mayor conocimiento del área estudiada y así poder ser más eficientes.

Las empresas actualmente están en un entorno muy competitivo, donde la diferencia marca el destino de la empresa. Por lo tanto, es fundamental tener conocimiento del entorno y saber actuar en los diferentes campos, por lo que, la inteligencia de negocios te facilita esa comprensión del entorno, e incluso puede ayudar a mejorar los procesos internos, por lo tanto, la inteligencia puede llegar a ser una ventaja competitiva.

En el presente trabajo se realizará un análisis inmobiliario, con el objetivo de mostrar la utilidad de la inteligencia de negocios en este sector. Para ello, en primer lugar, se describirá el mercado inmobiliario actual y su evolución. Esto permitirá comprender con mayor facilidad los datos que se trataran en el análisis. Para este análisis se utilizará una base de datos de un portal inmobiliario.

En el análisis, se aplicarán dos métodos de análisis de datos, por un lado, un método no supervisado, que se caracteriza por que no se conoce a priori el objetivo buscado, para ello se utilizará el método conocido como clustering. Se trata de una técnica que clasifica un conjunto heterogéneo de elementos en grupos, en función de las similitudes o diferencias entre ellos, esto permitirá agrupar los individuos de la muestra y ver sus diferencias. El objetivo de utilizar este método es conocer cuál es la tendencia de agrupamiento de las viviendas de la base de datos y que características definen cada grupo, esto permitirá obtener información de las viviendas que se están analizando y por lo tanto un mayor conocimiento de la base de datos.

Por otro lado, se aplicará un método supervisado, que se caracteriza por que se conoce a priori el objetivo buscado, para ello se utilizaran diferentes modelos supervisados y se escogerá el modelo más valido, con el objetivo de predecir el comportamiento de un atributo. El objetivo es desarrollar modelos de predicción que estimen el valor de un inmueble en función de sus características

1.2 OBJETIVOS DEL TRABAJO

Los objetivos fundamentales que se pretenden alcanzar en este trabajo son:

- Convencer de la necesidad e importancia que tiene la inteligencia de negocios para las empresas, ya que es un factor determinante que permite a la empresa optimizar todos sus procesos y puede ayudar a facilitar el camino al que se dirige la empresa.
- Conocer cómo se tratan los datos, cómo aprovechar al máximo la información que obtenemos a través de ellos, y cómo puede ayudar la inteligencia de negocios a las empresas a conseguir sus objetivos.
- Conocer el mercado inmobiliario en España, para ello se explicará su evolución hasta el momento, esto permitirá una mayor comprensión de los datos que posteriormente se analizarán a través de una base de datos de un portal inmobiliario.
- Desarrollar modelos que estimen el valor de un inmueble en función de sus características, lo que puede aportar valor, por ejemplo, a potenciales inversores de este mercado. Podrán comparar el precio de oferta de los inmuebles con el obtenido mediante el modelo, antes de tomar decisiones de inversión o desinversión

1.3 METODOLOGÍA

Para conseguir los objetivos planteados anteriormente, se van a realizar los siguientes pasos:

En primer lugar, para la definición de inteligencia de negocios, se ha planteado la definición de este concepto y sus componentes, así como la historia de la inteligencia de negocios, desde que surgió hasta la actualidad. Para elaborar este apartado se han utilizado diversas fuentes secundarias, tales como libros relacionados con la materia y páginas web oficiales.

En segundo lugar, se van a aplicar técnicas de análisis de datos, para ello me voy a centrar en el sector inmobiliario. En este apartado, antes que nada, se plantea la situación actual del mercado inmobiliario en España y cuál ha sido su evolución hasta el momento. Posteriormente, se analizarán los datos provenientes de una base de datos de un portal inmobiliario. En este análisis, se estudiará la relación de las diferentes variables entre sí y cómo influyen en el valor de la vivienda, también se aplicarán métodos no supervisados y supervisados.

En cuanto al método no supervisado, se caracteriza por qué no se conoce a priori el objetivo buscado, para este método se ha decidido aplicar la técnica de clustering, que consiste en clasificar un conjunto heterogéneo de elementos en grupos, en función de las similitudes o diferencias entre ellos. Posteriormente se validará el modelo y se procederá a la interpretación. El objetivo de utilizar este método es conocer cuál es la tendencia de agrupamiento de las viviendas de la base de datos y que características definen cada

grupo. Esto permitirá obtener información de las viviendas que se están analizando y por lo tanto un mayor conocimiento.

Por último, se aplicará un método supervisado, que se caracteriza por que se conoce a priori el objetivo buscado, a través de este método se pretende desarrollar modelos que estimen el valor de un inmueble en función de sus características. Para ello se utilizarán diferentes modelos supervisados y se escogerá el modelo más válido.

La base de datos que se va a analizar contiene 1000 observaciones y 10 variables, esta base muestra las características de las viviendas ofertadas en la ciudad de Madrid.

Se observan datos anómalos y valores faltantes, para tratar estos datos se eliminarán los datos anómalos, y los datos faltantes se imputarán por regresión, es decir, se predicen los valores ausentes en una variable, a partir de sus relaciones con otras variables de la base de datos.

2.ANTECEDENTES

2.1 ¿QUÉ ES LA INTELIGENCIA DE NEGOCIOS?

La inteligencia de negocios, más conocida por el término anglosajón “Business Intelligence”, hace referencia al proceso de extracción de información de una base de datos, que permite a la empresa obtener información útil para la toma de decisiones minimizando el error, mejorando así la competitividad de la organización.

A través de las tecnologías y los métodos de inteligencia de negocio, se pretende convertir los datos en información y esta información en conocimiento.

Actualmente, existen muchas definiciones para la inteligencia de negocios o business intelligence, una de las más aceptadas es la de (Gartner,2020):

“BI es un proceso interactivo para explorar y analizar información estructurada(datawarehouse) sobre un área, para descubrir tendencias o patrones, a partir de los cuales derivar ideas y extraer conclusiones. El proceso de BI incluye la comunicación de los descubrimientos y efectuar los cambios.”

En el libro “Business Intelligence: Competir con información” (Cano,2007), con el objetivo de entender esta definición de forma más clara, la descompone de la siguiente manera:

- Proceso interactivo: El BI no es un proceso que analiza un momento puntual de la empresa, sino que es un proceso continuo, esto permite observar tendencias y cambios.
- Explorar: Antes de empezar un proyecto, hace falta explorar el negocio con el objetivo de facilitar la interpretación.
- Analizar: Descubrir patrones y tendencias en los datos, para saber cómo actuaran los clientes con características similares.
- Información estructurada (datawarehouse): Los datos que se utilizan en el BI, están almacenados en tablas relacionadas entre ellas. Estas tablas poseen observaciones, las cuales poseen sus respectivos atributos(variable).
- Área de análisis: En los proyectos de BI, se debe concretar el campo que se desea analizar, ya sea el cliente, la empresa, los proveedores, el producto, los costes, etc....
- Comunicar los resultados y efectuar los cambios: Tras descubrir alguna información útil, hace falta comunicarla al responsable para que realice los cambios necesarios y mejorar la competitividad de la empresa.

2.2 HISTORIA DE LA INTELIGENCIA DE NEGOCIOS

Los datos a lo largo de la historia han sido claves para tomar cualquier decisión, han permitido conocer la situación y facilitar la toma de las decisiones.

La historia de la inteligencia de negocios surge a partir de la era de la computación, entre 1930 y 1949, el inicio coincidió con el estallido de la 2ª Guerra mundial. Las primeras computadoras que se diseñaron y construyeron, fueron las denominadas de “primera generación”, estaban fabricadas a mano, y se utilizaban tarjetas perforadas para la entrada y almacenamiento de datos.

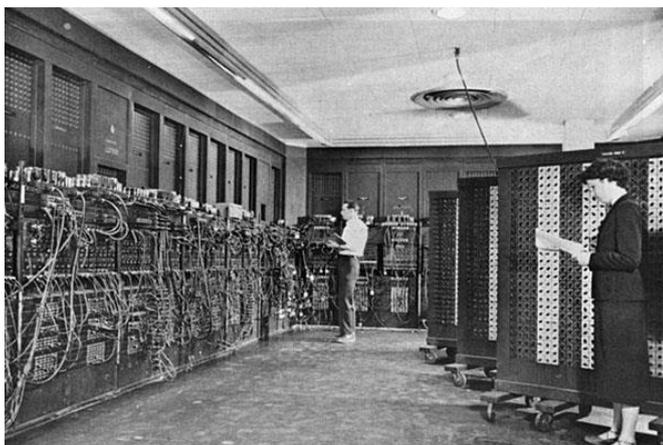


Ilustración 1: Computadora de 1ª generación

Fuente: (lifeder)

A partir del año 1950, empieza la comercialización de la analítica a través de modelos de predicción, se aplicaron modelos predictivos sobre todo en el ámbito de la logística, meteorología y las finanzas.

Posteriormente, en el año 1970, empieza a hacerse popular el análisis de datos, pero el auge vino cuando se empezaron a crear las compañías de Amazon (1994) y eBay (1995), que empezaron a ofrecer a sus clientes una experiencia personalizada basada en sus gustos y preferencias. (Gentleman, Hornik, Parmigiani, 2011)

En la entrada del nuevo milenio, este auge no ha parado de crecer hasta el momento, actualmente se está extendiendo el uso del análisis de datos cada vez en más empresas, y está siendo cada vez más valorado por las organizaciones, ya que les ofrece información muy relevante que les facilita la toma de decisiones.

Este incremento de la popularidad del análisis de datos en las empresas ha propiciado un incremento de la demanda de perfiles con estos conocimientos, esta demanda crece en gran medida año tras año, siendo un campo de estudio con una gran proyección para el futuro.

Se estima que, en 2020, la gestión de datos empleara a más de 10 millones de personas en Europa. A pesar de esto la oferta actual de trabajadores cualificados no puede cubrir la gran demanda de las empresas. (Martin,2019)

Forecast Revenue Big Data Market Worldwide 2011-2027
Big Data Market Size Revenue Forecast Worldwide From 2011 To 2027
(in billion U.S. dollars)

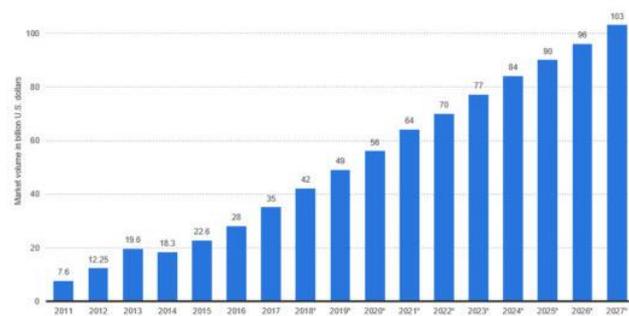


Gráfico 1: Pronóstico del tamaño del mercado de big data

Fuente: Statista

2.3 CARACTERÍSTICAS DE LA INTELIGENCIA DE NEGOCIOS

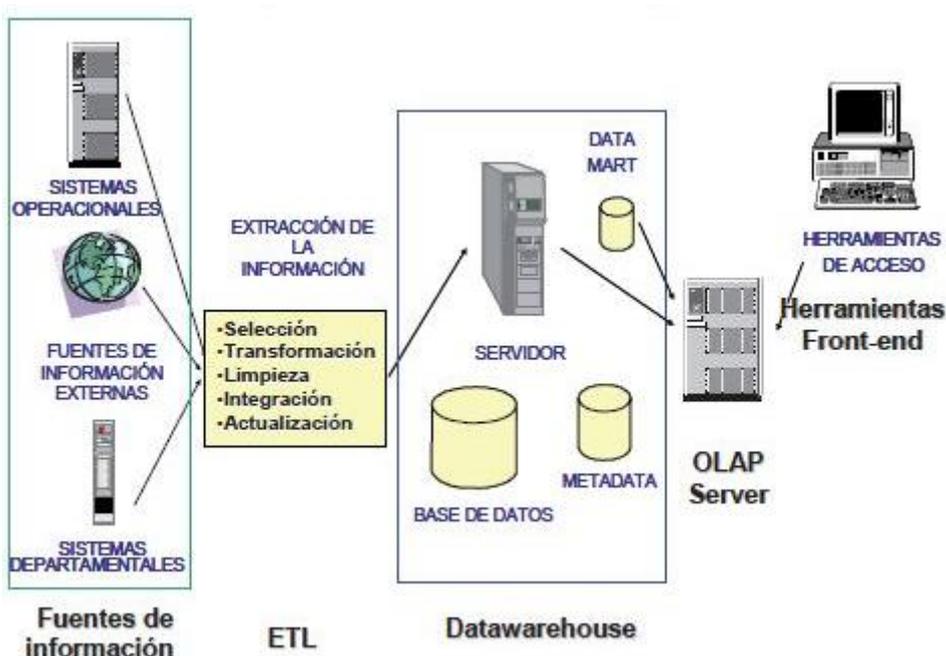
La inteligencia de negocios se basa en las tecnologías, que permiten a la empresa mejorar su competitividad. Por lo tanto, las características de la inteligencia de negocios van muy ligadas a las herramientas que se utilizan para procesar los datos y extraer información, entre las características del BI se encuentran según Cano (2007):

- **Accesibilidad a la información:** Cualquiera herramienta de BI, puede acceder a cualquier base de datos sin importar la procedencia de los datos, es decir, que cualquier usuario es capaz de acceder a los datos que desea, sin tener en cuenta el sistema que los genera.
- **Apoyo en la toma de decisiones:** Permite al usuario manipular los datos de la base de datos, para seleccionar aquellos que sean más oportunos y realizar un análisis personalizado, según el objetivo buscado.
- **Orientación al usuario:** Las herramientas deben ser intuitivas y fáciles de manejar, ya que permitirán al usuario un mejor control de los datos.

2.4 COMPONENTES DEL BUSINESS INTELLIGENCE

En el proceso del business intelligence intervienen varios componentes, desde la parte en la que se recogen los datos, hasta que llegan a las herramientas de BI, la siguiente imagen muestra todos los componentes del BI:

Ilustración 2: Componentes del BI



Fuente: Cano (2007)

Para entender de forma más clara, todo el proceso de BI, hace falta comprender como funcionan todos los componentes que intervienen.

En primer lugar, se encuentran las fuentes de información, que son donde se generan o extraen los datos, las fuentes a las que se pueden acceder son (Cano,2007):

- **Sistemas operacionales:** Los sistemas operacionales ofrecen apoyo a las actividades del sistema de la empresa, estas actividades están relacionadas con la generación, manipulación y tratamiento de la información. Se encuentran aplicaciones como ERP (Enterprise Resources Planning) que ayuda a la planificación de los recursos empresariales, y CRM (Customer Relationship Management) que ofrece soporte en la gestión de las relaciones con los clientes.

- **Sistemas departamentales:** Hace referencia a la información que se puede extraer de cada departamento, ya sea el departamento financiero, de recursos humanos o producción, a través de hojas Excel, previsiones, etc.
- **Fuentes de información externas:** Esta información es la que se obtiene de fuentes ajenas a la empresa, como por ejemplo puede ser comprada a terceros o derivar de estudios de mercado. Además, también se puede obtener información útil para las empresas de forma gratuita a través del INE.

En segundo lugar, se encuentra el proceso ETL (Extract, Transform and Load), en el cual, tras recibir la información de las diferentes fuentes, se procede a su tratamiento, debido a que no toda la información es válida, y se querrá trabajar con una base de datos de la mayor calidad posible. Este proceso prepara los datos para almacenarlos posteriormente en la base de datos

Este proceso se puede dividir en 3 etapas:

- **Extracción:** En la primera etapa, se recopilan los datos obtenidos por las diferentes fuentes de información, en este paso los datos son estructurados, debido a que cuando se obtienen de diversas fuentes los datos no están estructurados. Por lo tanto, para trabajar con ellos es necesario estructurarlos y convertirlos en el mismo formato. Tras estos pasos, quedan listos para su posterior transformación.
- **Transformación:** En esta etapa, se estructuran los datos de acuerdo con el negocio, es decir, se ordenan, se filtran y se cruzan los datos, para estructurarlos de acuerdo con el tipo de empresa. Además, se limpian los datos, para ello se eliminan los datos duplicados, los nulos o los que no aportan información. Tras este paso, se ha mejorado la calidad de los datos y ya están listos para almacenarlos en la base de datos.
- **Carga de los datos:** En la última etapa, los datos ya se encuentran estructurados y limpiados, por lo que se procede a su carga en la base de datos (data warehouse) donde se registran.

Existen herramientas ETL que facilitan el proceso, como son:

- Apache nifi
- Talend
- SAP
- BI tool
- Streamsets
- Informática powercenter

Las características que deben poseer las herramientas ETL, según (Gartner,2020) son:

- Adaptación: Que sea capaz de utilizar y conectar diferentes tipos de estructuras de datos, es decir, conectar ficheros de datos de diferentes fuentes de información y diferentes formatos.
- Entrega de datos: Ser capaz de facilitar los datos a otras aplicaciones o procesos.
- Transformación de datos: Las herramientas deben poseer librerías que permitan transformar los datos originales en las estructuras de datos deseadas e incluso crear tablas de agregación.
- Gestión del metadata: Poseer un repositorio que almacene y registre el proceso ETL y su ejecución.
- Diseño gráfico: Entorno que proporcione al usuario representar gráficamente objetos en el repositorio (metadata), modelos de datos, flujos, etc.

Tras haber realizado el proceso ETL, los datos ya se encuentran estructurados, y por lo tanto ya se pueden almacenar, el lugar donde se guardan se denomina base de datos (data warehouse).

Existen muchas definiciones que explican que es un data warehouse, una de las más aceptadas es la de Hugh J. Watson (Watson,2008) que define el data warehouse como:

“Un data warehouse es una colección de información creada para soportar las aplicaciones de toma de decisiones”

Cualquier data warehouse debe poseer las siguientes características:

- Integrado: Que permita homogenizar las estructuras de datos, es decir, que todos los datos estén medidos en la misma forma, ya sea, en metros, euros, kilos, litros, etc.
- Temático: La información esta almacenada por contenido, por lo que es fácil de entender para el usuario.
- Tiempo variante: Las bases de datos permiten hacer consultas de registros históricos.
- No es volátil: La información antigua se guarda, no se elimina, por lo que la información es permanente.

Por último, tras tener la información estructurada y homogenizada en el data warehouse, para analizar esta información es necesario disponer de herramientas de análisis.

Según (Cano,2007), las herramientas principales del business intelligence son:

- Generadores de informes: Se utilizan para generar informes estándar para grupos como departamentos o la empresa.
- Herramientas de consultas e informes: Utilizadas por los usuarios finales para generar informes para su uso propio.
- Herramientas OLAP: Facilitan a los usuarios finales tratar la información, se basan en el análisis de grandes cantidades de datos a través de estructuras multidimensionales, que permiten explorar la información desde diferentes perspectivas y periodos de tiempo.
- Herramientas de dashboard y scorecard: Ofrecen al usuario la posibilidad de ver información crítica tan solo presionando un icono, y también permite ver con más detalle los informes.
- Herramientas de planificación, modelización y consolidación: Permiten a los usuarios crear planes de negocio y simulaciones con la información que se posee. Se suelen utilizar estas herramientas para generar presupuestos, previsiones, etc.
- Herramientas datamining: Estas herramientas permiten a los usuarios generar modelos estadísticos de las actividades del negocio. El data mining permite descubrir patrones mediante los cuales resolver problemas de la empresa.



Ilustración 3: Ejemplos de herramientas de BI

Fuente: (cleformacion,2021)

2.5 SITUACIÓN ACTUAL DE LA INTELIGENCIA DE NEGOCIOS

En la actualidad, existe un crecimiento en las empresas que empiezan a darle valor a los datos, estas empresas tienen enormes cantidades de datos que cada día van creciendo por lo que, necesitan bases de datos cada vez más grandes.

En estas bases de datos, las empresas recogen información sobre sus clientes, sus proveedores, sus productos o servicios incluso hasta datos financieros, de producción y demás áreas de la empresa.

Las empresas tienen almacenadas enormes cantidades de datos que en sí no aportan demasiada información. Ahí entra en juego el análisis de datos, que se encarga de tratar dichos datos, con el objetivo de explotar al máximo este recurso, para obtener la mayor cantidad de información posible que facilite la toma de decisiones de la empresa.

Actualmente, la minería de datos es muy útil en empresas en las que su actividad está enfocada al canal online, ya que no tienen un contacto directo con los consumidores, por lo que no pueden conocer su comportamiento. La minería de datos permite analizar su comportamiento, registrando los clics que hacen, esto permite a la empresa hacerse una idea más acertada sobre el perfil del cliente. Esto permite a la empresa ofrecer una experiencia más personalizada a cada cliente, pudiéndoles recomendar algún producto o servicio que se ajuste a sus características, con el propósito de incrementar las ventas. (Hodeghatta y Nayak,2017)

La inteligencia de negocios no solo es aplicable a las empresas online sino también es beneficiosa para las empresas tradicionales, como por ejemplo a través de los datos que poseen pueden extraer información de sus clientes, y enfocarse de manera más precisa al segmento de mercado al que se dirigen sus productos. Incluso conocer cómo podrían mejorar un determinado producto para adaptarlo, la percepción de los consumidores sobre un producto o incluso la marca. La inteligencia de negocio abarca un amplio abanico de posibilidades que ayudan a la empresa a alcanzar los objetivos planteados.

Por último, la inteligencia de negocio también se puede utilizar para predecir comportamientos futuros, basándose en los datos que se poseen, esto permite a la empresa poder anticiparse y por lo tanto estar preparada para ese posible futuro. (Shmueli *et al.*, 2018)

2.6 BENEFICIOS DE LA INTELIGENCIA DE NEGOCIOS

Como se ha visto anteriormente, la inteligencia de negocios aporta muchos beneficios a la empresa, según (Cano,2007) los beneficios que se obtienen se pueden clasificar en:

- **Beneficios tangibles:** Son aquellos beneficios que están relacionados con el aumento de ventas, reducción de costes y también con la reducción de tiempo en los procesos de la empresa.

Algunos ejemplos de beneficios tangibles son:

- Aumentar el control de costes
- Reasignar al personal para realizar los procesos, asignando a cada tarea el personal con las competencias más adecuadas.
- Mejorar y proveer el suministro dentro del plazo, atendiendo a las predicciones de demanda.
- Hacer un seguimiento del proceso de producción, para detectar los factores que influyen que un producto salga defectuoso, ya que permitirá mejorar esos factores y disminuir el número de productos defectuosos, lo que ocasionara una reducción de costes.

- **Beneficios intangibles:** el hecho de que todos los trabajadores de la empresa tengan acceso a la información, hará que mejoren la toma de decisiones, mejorando así la posición competitiva de la empresa.

Algunos ejemplos de beneficios intangibles son:

- Optimizar la atención del cliente, pudiendo ofrecer una atención personalizada.
- Aumentar la satisfacción de los clientes.
- Ofrecer información más actualizada.
- Conseguir ventajas competitivas.

- Beneficios estratégicos: Son aquellos beneficios que facilitan la elaboración del plan estratégico, es decir, a que segmento dirigirse, a que mercado y con qué productos.

Algunos ejemplos de beneficios estratégicos son:

- Mayor habilidad para realizar estrategias de precios.
- Mayor habilidad para atraer clientes potenciales.
- Mejorar la toma de decisiones, al tener más conocimientos.
- Dar soporte a las estrategias.

3. EL MERCADO INMOBILIARIO

En este apartado, se realizará una descripción del mercado inmobiliario en España, con el objetivo de comprender como funciona y, por lo tanto, obtener un mayor conocimiento sobre este sector. Permitiendo entender de una forma más adecuada los datos recogidos en la base de datos, además facilitará la interpretación de los resultados obtenidos y por lo tanto se obtendrá una conclusión más acertada.

De modo que, en primer lugar, se procederá a definir que es el sector inmobiliario, y cuáles son sus características. Esto permitirá tener un conocimiento más profundo del sector, y ver cuáles son los factores que más influyen, ya que habrá que tenerlos en cuenta cuando se realice el análisis.

Posteriormente, se describirá la situación actual del sector inmobiliario en España, con el objetivo de comprender el entorno el que se encuentra este sector en España. También se comentará como ha afectado la crisis sanitaria en este mercado, y cuáles son las predicciones que se esperan.

3.1 DEFINICIÓN Y CARACTERÍSTICAS

El mercado inmobiliario es el lugar en el que se producen las transacciones económicas, por las cuales se adquiere la propiedad o arrendamiento de un inmueble, es decir, la finalidad del mercado inmobiliario es la obtención del uso y disfrute de un bien inmueble, pudiendo ser bienes de naturaleza residencial, comercial, industrial o urbano, entre otros.

En este mercado interactúan compradores y vendedores, que hacen referencia a los ofertantes y demandantes de inmuebles. También existen intermediarios, como son las agencias inmobiliarias que se encargan de poner en contacto a los oferentes y demandantes, a cambio de una comisión que suele ser un tanto por ciento sobre el precio de venta.

En cuanto a las operaciones que más se desarrollan en este mercado, destacan la compraventa y el arrendamiento. Por un lado, el arrendamiento, es un contrato entre el ofertante y el demandante, a través del cual el ofertante cede durante un tiempo determinado el uso y disfrute del inmueble, a cambio de un pago regular, que normalmente es mensual. Por otro lado, la compraventa consiste en un contrato, en el cual el ofertante está dispuesto a ceder la totalidad de la propiedad al demandante de forma indefinida, a cambio de precio acordado por ambas partes.

La diferencia entre la compraventa y el arrendamiento es que en la compraventa se traspasa la propiedad al demandante, en cambio, en el arrendamiento, solo se le cede el uso y disfrute del inmueble por un periodo predeterminado.

Por último, el sector inmobiliario es un pilar fundamental para el desarrollo de una economía sostenible de un país. Además, cabe destacar, que no es un mercado inmóvil, ya que la situación de la economía del país le afecta, es decir, los auges y caídas de la economía de un país pueden llegar a ocasionar cambios en los precios de los inmuebles, que pueden repercutir en un aumento o disminución de su valor. (Realia,2015)

Tras haber definido lo que es mercado inmobiliario y cuáles son las principales operaciones que se desarrollan, también es necesario ver cuáles son las principales características de este mercado, con el objetivo de profundizar en el tema.

Entre las características del mercado inmobiliario destacan:

- No se conoce con exactitud el valor de los activos, por lo que se utilizan referencias próximas, aunque solo tienen carácter orientativo.
- En la valoración de un activo, existe la posibilidad de error, ya que al fijar el precio adecuado y no conocer el precio exacto, se puede cometer un error debido a que la hora de determinar el precio se tiene en cuenta referencias próximas, por lo tanto, no son del todo acertadas ya que se realizan pocas operaciones y las características de cada inmueble varían.
- Los activos con los que opera el mercado inmobiliario carecen de mucha liquidez, debido a que son operaciones con un importe elevado y se necesitan realizar varios trámites legales.
- El mercado inmobiliario se ve afectado por la situación económica del país, por lo que es sensible a las fluctuaciones de los ciclos económicos.

3.2 TIPOS DE VIVIENDA

Para el posterior análisis, será importante conocer los diferentes tipos de vivienda que existen en la actualidad, ya que cada tipo de vivienda posee unas características determinadas, las cuales influyen en el precio de la vivienda. Por lo tanto, conocer estas diferencias será muy útil a la hora de interpretar los resultados del análisis, ya que puede ser un factor esencial que influya a la hora de determinar el precio de la vivienda.

En primer lugar, cabe destacar que los tipos de vivienda se separan en dos grandes grupos, conocidos en el mercado inmobiliario como: viviendas unifamiliares y viviendas plurifamiliares. (Housing, 2020)

Las viviendas unifamiliares se caracterizan porque la totalidad de la construcción se destina a una sola familia, estas viviendas se suelen encontrar más en núcleos de menor densidad de población. Suelen situarse en zonas que se encuentran a las afueras de la ciudad, además también suelen tener un precio más elevado que el otro tipo de viviendas, debido a que su superficie es mucho mayor. Aunque debido a su localización tienen menos infraestructuras que las viviendas más céntricas de la ciudad y por lo tanto un menor tráfico.

Las viviendas plurifamiliares se caracterizan porque varias viviendas comparten el mismo terreno y poseen zonas comunes, en estas viviendas viven varias familias. Suelen ser viviendas más económicas que las unifamiliares, además son más baratas de mantener, están mejor comunicadas con los diferentes servicios de la ciudad y además de más variedad. El mayor inconveniente de este tipo de viviendas es el espacio, ya que las viviendas tienen un espacio muy reducido en comparación al otro tipo de viviendas, y

además suelen ser viviendas menos tranquilas debido a que tienen que convivir con otras familias y también lidiar con la molestia del tráfico de la ciudad.

Entre las viviendas unifamiliares se encuentran:

- Casa independiente: Se trata de una edificación aislada que no linda con otras viviendas ya que está rodeada por un terreno. Suelen ser tranquilas ya que no conviven con otras familias y tienen poco tráfico. Uno de los inconvenientes de este tipo de viviendas es que tienen un elevado coste de mantenimiento y están peor comunicados con los servicios.
- Chalet adosado: Se trata de un grupo de viviendas construidas de forma seguida y pegadas unas a otras a través de una pared. Suelen tener un coste alto de mantenimiento, a diferencia de las viviendas plurifamiliares, este tipo de vivienda tiene su independencia ya que solo hay una vivienda por terreno. En algunas ocasiones los vecinos de este grupo de viviendas comparten zonas comunes.
- Chalet pareado: Se trata una vivienda que cuenta con otra vivienda adosada solo a uno de sus extremos, teniendo el otro extremo libre. Este tipo de viviendas permite contar con un mayor espacio exterior, lo que ofrece a la vivienda una amplia variedad de posibilidades de aprovechar el espacio, como por ejemplo una piscina, o una casita donde poner un asador para organizar comidas con la familia y los amigos. Por lo tanto, este tipo de vivienda tiene un coste elevado de mantenimiento

Por otro lado, entre las viviendas plurifamiliares se encuentran:

- Estudio: Se trata de una vivienda donde en una misma estancia se encuentra la cocina, el salón y el dormitorio. Son viviendas muy baratas, pero requiere que este muy bien organizada para aprovechar el poco espacio.
- Apartamento: Se trata de viviendas que solo disponen de una habitación que se encuentra independientemente del resto de la vivienda. Suelen tener un precio bajo y su principal usuario son parejas sin hijos.
- Piso: Se trata de una vivienda ubicada en un edificio. Suelen tener varios dormitorios y tiene un bajo coste de mantenimiento, el mayor inconveniente es la convivencia con otras familias lo que resta tranquilidad y privacidad.

- **Dúplex:** Se trata de una vivienda de dos plantas que se comunican a través de una escalera. Son más caros de mantener debido a que son más grandes y tienen un precio elevado.
- **Ático:** Se trata de viviendas que se sitúan en el último piso del edificio, son los pisos más demandados debido a la luminosidad y vistas que poseen. Suelen ser las viviendas más caras del edificio.
- **Bajo:** Se trata de las viviendas situadas en la planta baja del edificio, son bastante demandadas. Algunos de estos pisos suelen tener patios privados.
- **Buhardilla:** Se trata de viviendas situadas en el último piso, suelen tener un espacio reducido, y la principal característica que los diferencia es su techo abuhardillado, es decir, con las paredes inclinadas. Tienen las mismas ventajas que los áticos, como son la luminosidad y las vistas, pero son más viviendas más baratas.
- **Loft:** Se trata de un nuevo concepto de vivienda, que se caracteriza por tener un gran espacio y pocas divisiones, el mayor inconveniente es que al tener la vivienda tan abierta y con pocas divisiones se pierde intimidad.

3.3 EVOLUCIÓN DEL MERCADO INMOBILIARIO EN ESPAÑA

El mercado inmobiliario, a lo largo del siglo XXI, ha sufrido grandes oscilaciones, debido a los ciclos económicos y a diversos motivos globales. Desde 1997 hasta 2007, se produjo un auge de la construcción debido a los altos precios de las viviendas en algunas economías de Europa, ocasionados por la mejora de los accesos a la financiación, lo que provocó un gran aumento de la inversión en este sector. Además, gracias a las políticas del BCE, se mantuvo un tipo de interés reducido, mejorando las condiciones de financiación, a lo que hay que añadir que España se encontraba en un periodo de expansión. Todos estos hechos provocaron que las viviendas fueran fácilmente accesibles, lo que produjo un estallido de compraventa de inmuebles.

El mayor impacto que ha sufrido el mercado inmobiliario fue en 2008, donde estallo la burbuja inmobiliaria, que ocasiono una recesión económica. El motivo de esta crisis se debió a la especulación de los precios de las viviendas, sobre todo entre 2005-2007, donde muchas empresas y particulares compraban inmuebles y posteriormente los vendían, obteniendo una gran cantidad de beneficio en un corto plazo. A esto se le añade, la facilidad con la que los bancos concedían créditos tanto a las empresas como a los particulares.

Por otra parte, la elevada inflación en EE. UU., provoco que aumentaran el tipo de interés con el objetivo de controlarla, lo que produjo que las personas no pudieran pagar las deudas al banco, y esto ocasiono problemas de liquidez a los bancos. Todo esto provoco el estallido de la burbuja inmobiliario que se extendió al resto de países.

En España, la crisis provoco una alta tasa de paro lo que ocasiono que los ciudadanos no puedan hacer frente a los pagos, y los banco empezaron a embargar una gran cantidad de inmuebles.

A partir de esta crisis, los precios de las viviendas bajaron de forma brusca, hasta el año 2015 donde gracias a las medidas tomadas por el gobierno hasta la fecha ayudaron a recuperarse al mercado inmobiliario. (Bde,2020)

Gráfico 2: Gráfico de operaciones de compraventa desde 2007-2015



Fuente: (INE,2020)

En el gráfico se puede observar las operaciones de compraventa que se realizaron desde principios de 2007 hasta principios de 2015. Se puede apreciar como la crisis inmobiliaria redujo notoriamente las operaciones de compraventa, pasando de cerca 86000 operaciones en 2007 a 32000 en 2008.

3.4 SITUACIÓN ACTUAL DEL MERCADO INMOBILIARIO Y PERSPECTIVAS FUTURAS

El mercado inmobiliario ha ido creciendo poco a poco, debido a que la economía se iba recuperando de la crisis y los bancos empezaban a ofrecer financiación, así fue hasta 2020. A partir de este año, surgió una crisis sanitaria, causada por el Covid-19, que ha ocasionado una de las mayores crisis de las últimas décadas. Actualmente, el mercado inmobiliario se encuentra en un futuro incierto ya que es uno de los sectores que se está viendo más afectado por esta situación. Esto se debe principalmente a la incertidumbre y al cambio que podría provocar esta pandemia en la forma vivir, trabajar y consumir de las personas.

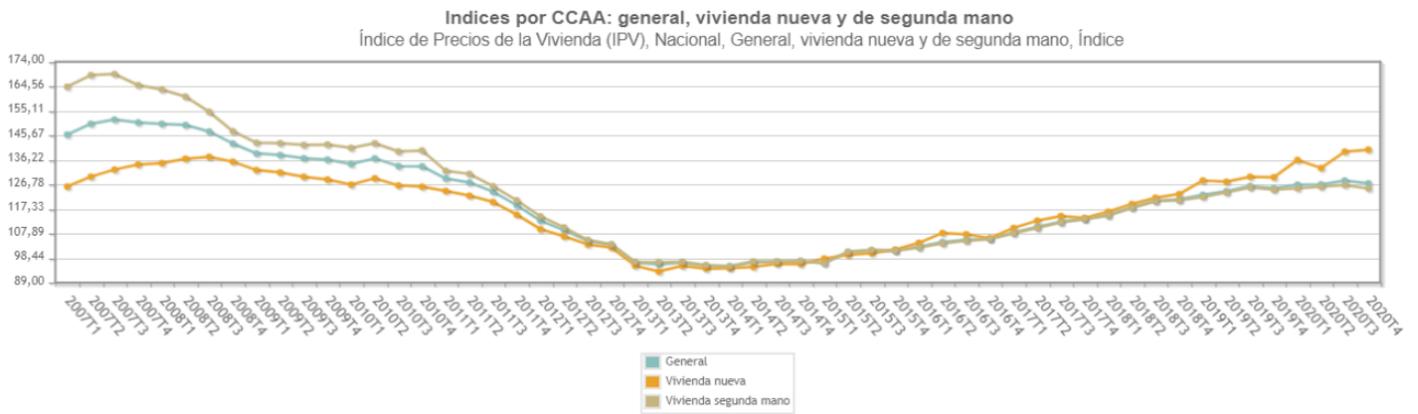
Por último, las predicciones que se esperan para el mercado inmobiliario destacan, que las empresas empezaran a fomentar el teletrabajo. Por lo tanto, se reducirá en gran cantidad el número de personas en las oficinas, por lo que los inmuebles que anteriormente agrupan concentraciones de personas (oficinas, centros comerciales, estadios, teatros) serán los que más sufran el impacto de esta crisis y se verán obligados a innovar para seguir en funcionamiento. Por otro lado, el auge que está teniendo el comercio online ha provocado un incremento en la demanda de los espacios de almacén, debido al aumento de la demanda. A pesar de esto, se han visto reducidas las ventas en los comercios locales y los centros comerciales, por lo que para mantenerse deberán innovar y fomentar modelos de negocio adaptados a la situación.

Aunque actualmente el futuro del mercado aun es incierto, se van observando cambios en las tendencias de los consumidores, que pueden provocar la reducción de muchos activos inmobiliarios, que actualmente se encuentran en el mercado. (Deloitte,2020)

3.5 EVOLUCIÓN DE LOS PRECIOS EN EL MERCADO INMOBILIARIO

Desde comienzos del siglo XXI el precio ha ido variando hasta la actualidad. Para comprender mejor como ha sido esta variación, se observará la evolución del índice de precios de venta que refleja la evolución de los precios de la vivienda.

Gráfico 3: Gráfico IPV



Fuente: (INE,2020)

El gráfico muestra la variación de los precios de las viviendas nuevas y de segunda mano desde principios de 2007 hasta finales de 2020. Como se puede observar en el gráfico, a partir de la crisis inmobiliaria de 2008, el precio de la vivienda cambio de tendencia y empezó a decaer hasta 2013, donde toco fondo. A partir de este año, empezó a aumentar debido a que la economía empezaba a recuperarse.

Como el posterior análisis tratara sobre las viviendas en la ciudad de Madrid, sería interesante ver cómo ha variado el precio del m2 hasta la actualidad, y la diferencia que existe entre los precios de los diferentes barrios de la ciudad.

Gráfico 4: Gráfico de la evolución del precio del m2 en Madrid desde 2006-actualidad



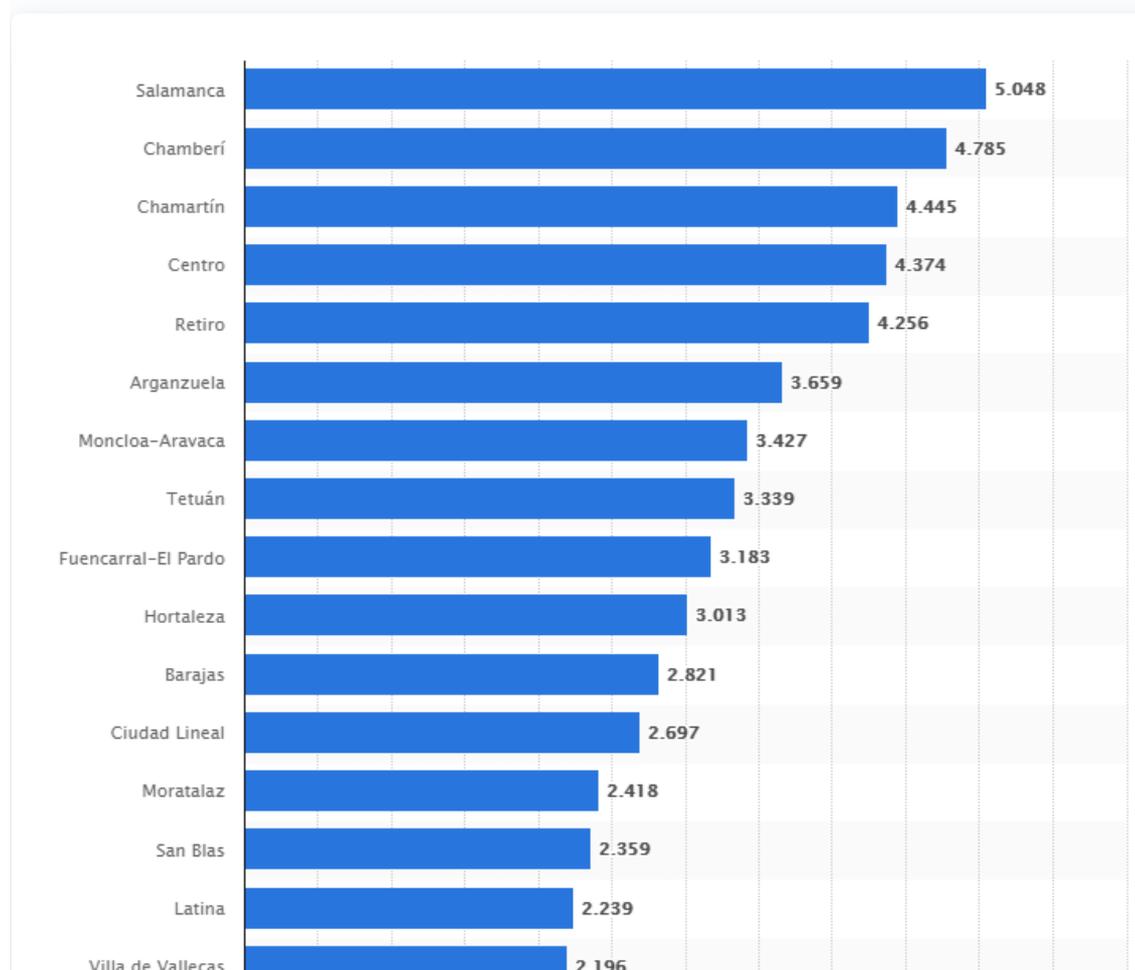
Fuente: (idealista,2020)

El gráfico muestra la variación del precio del m2 en la comunidad de Madrid desde principios de 2006 hasta la actualidad. Se puede apreciar que la crisis inmobiliaria no afectó en gran medida al precio del m2, también se puede observar un ligero descenso desde julio de 2012, que se recuperó a partir de principios de 2018. Otro aspecto relevante, es que el precio del m2 antes de la crisis inmobiliaria se situaba en torno a 2700 €/m2 y en la actualidad, se encuentra a un precio superior que ronda los 3000 €/m2.

A continuación, se mostrará el precio por m2 de los diferentes barrios de Madrid:

Gráfico 5: Gráfico del precio del m2 de los barrios de Madrid (3ºtrimestre de 2020)

(en euros por metro cuadrado)



Fuente: (Statista,2020)

El gráfico muestra el valor por m² de los diferentes barrios de Madrid en el tercer trimestre de 2020. Se aprecia una gran diferencia entre el valor del m² en el barrio de Salamanca que se sitúa en 5048 €/m², en comparación a los 2196 €/m² del barrio de villa de Vallecas.

4. ANÁLISIS DE DATOS

En este apartado, se procederá a analizar la base de datos, que muestra las características de 1000 viviendas ofertadas en la ciudad de Madrid. Para realizar este análisis, en primer lugar, se describirán las variables que contiene la base de datos, con el objetivo de conocer con mayor profundidad a qué se refiere cada variable.

Antes de empezar a aplicar los diferentes métodos de análisis, se debe preparar la base de datos. Para ello, se deben eliminar las variables que no aportan utilidad al modelo y las observaciones anómalas o inconsistentes. También para poder realizar adecuadamente los métodos de análisis, se deberán reemplazar los valores faltantes, entre otras medidas a tomar, antes de empezar a analizar la base de datos.

Posteriormente, cuando la base de datos ya esté preparada, se procederá a analizarla. Para ello, en primer lugar, se aplicará un método no supervisado, que se caracteriza porque no se conoce a priori el objetivo buscado. Para este método se ha decidido aplicar la técnica de clustering. A través de esta técnica se clasifica un conjunto heterogéneo de elementos en grupos, en función de las similitudes o diferencias entre ellos. Posteriormente se validará el modelo y se procederá a la interpretación. Este método permitirá entender más profundamente la base de datos analizada, y observar cuáles son las variables que más influyen en la variabilidad del modelo, así como la relación entre ellas mismas.

El software que se utilizará para realizar todos los procesos pertinentes y los métodos de análisis no supervisado y supervisado será el Rstudio. Se trata de un entorno de desarrollo integrado para el lenguaje de programación R, dedicado a la computación estadística y gráficos. Incluye una consola, editor de sintaxis que apoya la ejecución de código, así como herramientas para el trazado, la depuración y la gestión del espacio de trabajo.

Además, se trata de un software libre, por lo que está a disposición de todos los usuarios, los cuales pueden modificar o crear nuevas funciones para el software y compartirlas con los demás usuarios.

4.1 PREPARACIÓN DE LA BASE DE DATOS

Antes de empezar a preparar la base de datos, se debe entender la base de datos para ello a continuación se describirán las 10 variables:

-X1:id de la vivienda.

-rawPrice: Precio de la vivienda.

-Surface: superficie de la vivienda, medida en metros cuadrados.

-rooms: habitaciones

- bathrooms: baños.
- elevator: ascensor (1 si posee ascensor, 0 en caso contrario)
- parking: zona de aparcamiento (1 si posee aparcamiento, 0 en caso contrario)
- terrace: terraza (1 si posee terraza la vivienda, 0 en caso contrario)
- buildingType: tipo de vivienda.
- buildingSubType: subtipo de vivienda.

A continuación, se empezará a preparar la base de datos, para ello, en primer lugar, se observará si existen variables que no aportan nada al modelo. En esta base de datos se aprecia, que las siguientes variables no aportan valor al modelo: X1(ya que solo es un identificador de la vivienda) y buidingType (debido a que todas las viviendas analizadas son pisos, por lo tanto, esta variable no será significativa).

Tras haber eliminado estas variables, se obtiene una nueva base de datos formada por 1000 observaciones y 8 variables, de las cuales 7 son de tipo numérico y una de tipo categórica, que se trata de buildingSubType.

Por otro lado, en las variables no se ha detectado ningún valor anómalo o inconsistente, en cambio sí que se han detectado valores faltantes, esto puede perjudicar a la hora de aplicar los métodos de análisis, ya que si una variable tiene un elevado porcentaje de valores faltantes el programa los sustituirá automáticamente por la media de la variable, y por lo tanto esto puede no ser adecuado. Para solucionar este problema se ha decidido imputar los valores por regresión, que consiste en predecir los valores faltantes de una variable a partir de sus relaciones con otras variables de la base de datos.

Otro a factor para tener en cuenta, es si hace falta transformar o recodificar alguna variable, se ha decidido por ahora no recodificar ninguna variable, y en el caso de necesitarlo se hará más adelante.

Ahora, que ya se ha limpiado, extraído y procesado los datos de forma que se pueda trabajar con ellos con una mayor calidad, se procederá al análisis, para ello se empezara con el análisis no supervisado.

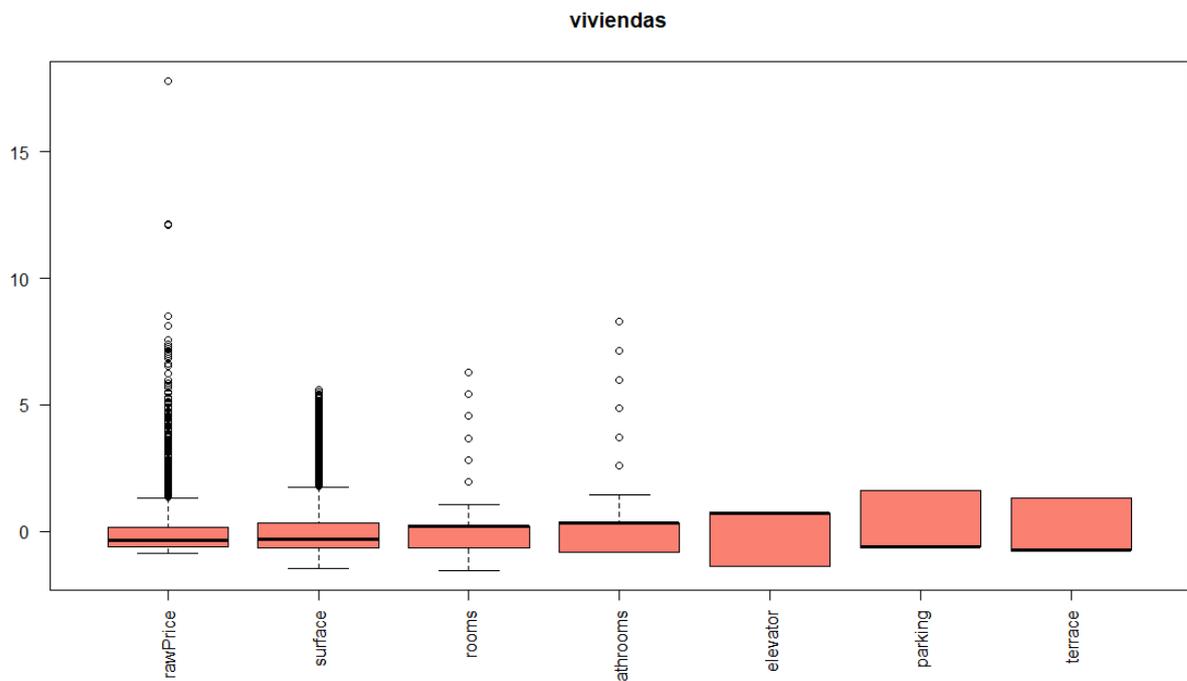
A continuación, se realizará un método de análisis no supervisado, conocido como clustering. El objetivo de utilizar este método es conocer cuál es la tendencia de agrupamiento de las viviendas de la base de datos y que características definen cada grupo. Esto permitirá obtener información de las viviendas que se están analizando y por lo tanto un mayor conocimiento del mercado inmobiliario.

4.2 MÉTODO NO SUPERVISADO

Los métodos no supervisados, se caracterizan por qué no se conoce a priori el objetivo buscado, para este método se ha decidido aplicar la técnica clustering, a través de este modelo se reducirá la dimensión de los datos. Esta técnica clasifica un conjunto heterogéneo de elementos en grupos, en función de las similitudes o diferencias entre ellos, posteriormente se validará el modelo y se procederá a la interpretación.

En primer lugar, para empezar con el clustering, se necesitarán las variables numéricas y la variable categórica se utilizará posteriormente para facilitar la interpretación. Primero, habrá que observar en qué medida están representada las variables, con el objetivo de saber si es necesario escalar y centrar, en este caso, sí que haría falta escalar y centrar los datos, ya que todas las variables no están medidas en la misma escala y también hace falta que la media sea 0.

Gráfico 6: Gráfico boxplot de las variables escaladas y centradas de la base de datos

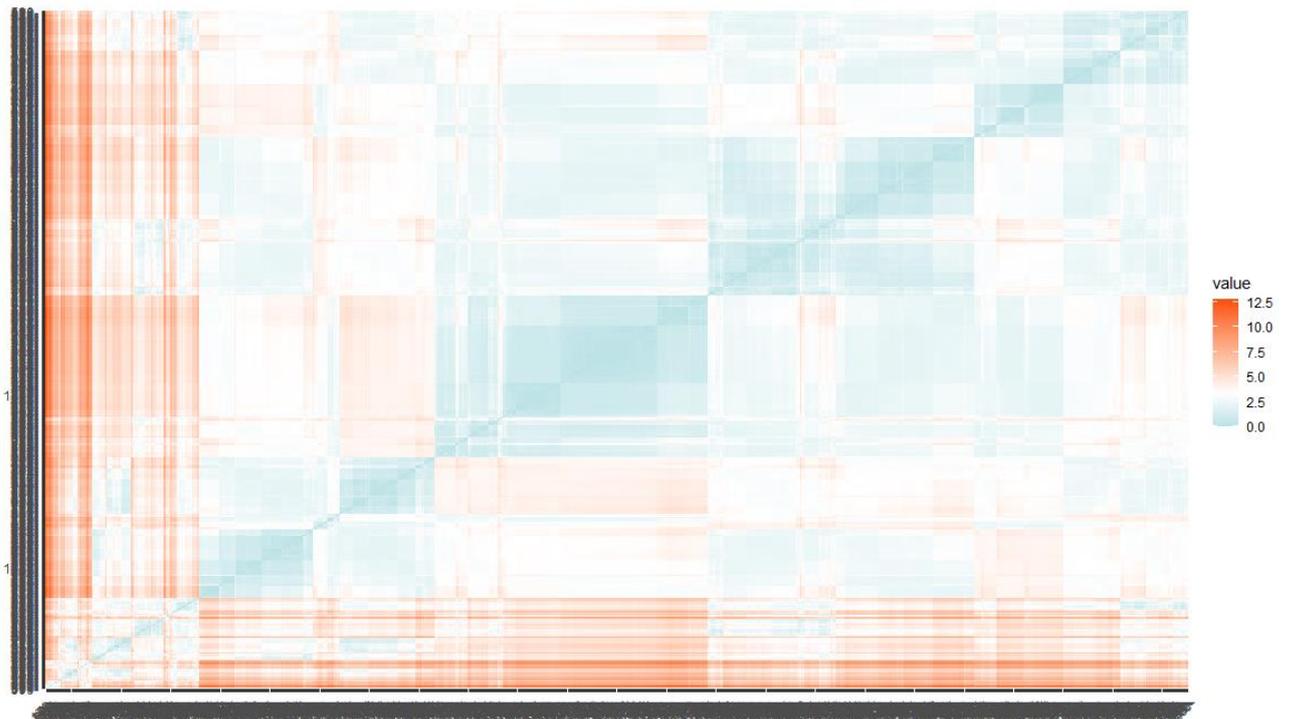


Fuente: elaboración propia

Como se observa en el gráfico boxplot, las variables numéricas, ya han sido escaladas y centradas. También, se observan valores anómalos en algunas variables, pero no se han eliminado debido a que son valores que, a pesar de no ser habituales, se pueden dar en ciertas viviendas.

A continuación, se procederá a generar el mapa de colores, en el cual se ha utilizado la distancia euclídea.

Gráfico 7: Mapa de color



Fuente: elaboración propia

Como se puede observar en el mapa, las viviendas se agrupan mayoritariamente en dos grupos y también hay varios grupos de menor tamaño. Este mapa da una idea del posible número de clusters.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.8587	0.9002	0.9231	0.9174	0.9360	0.9715

Tabla 1: Estadístico de Hopkins

Fuente: elaboración propia

El estadístico de Hopkins selecciona aleatoriamente observaciones, calculando la distancia entre el elemento y su elemento más cercano, como muestran los datos anteriores existe una elevada tendencia de agrupamiento ya que sus valores están entre 0.85 y 0.97.

4.2.1 MODELOS JERÁRQUICOS

En primer lugar, se aplicará el modelo jerárquico, el cual es útil cuando los individuos tienen una clara estructura jerárquica, además, puede ser utilizado como paso previo para la determinación del número de grupos a formar, con un método no jerárquico.

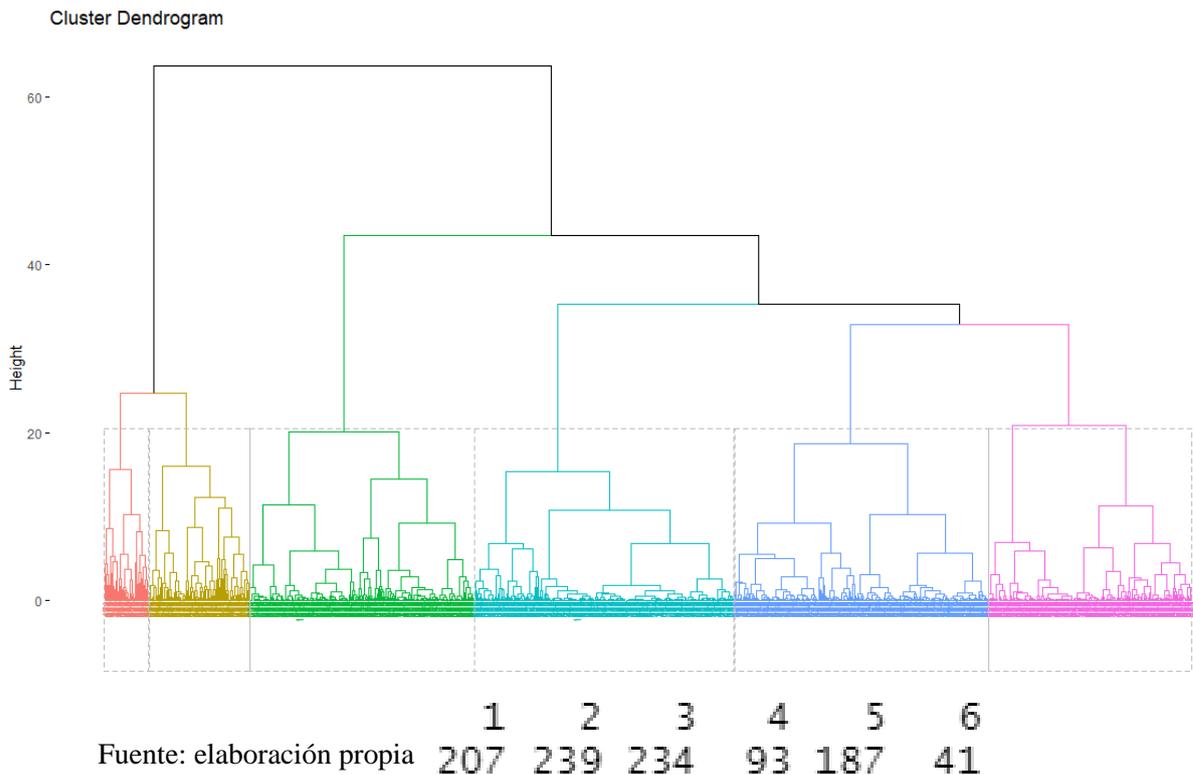
Para ello se aplicarán dos métodos: método ward y método de la media.

Método Ward

Este método forma clusters maximizando la homogeneidad intra-clusters.

A través de un dendrograma, se representará gráficamente en forma de árbol el proceso de agrupamiento con los 6 clusters que se han apreciado en el mapa de color, y se cortara el dendrograma a un determinado nivel, para obtener esa clasificación de los elementos en cada grupo.

Gráfico 8: Dendrograma con el método ward

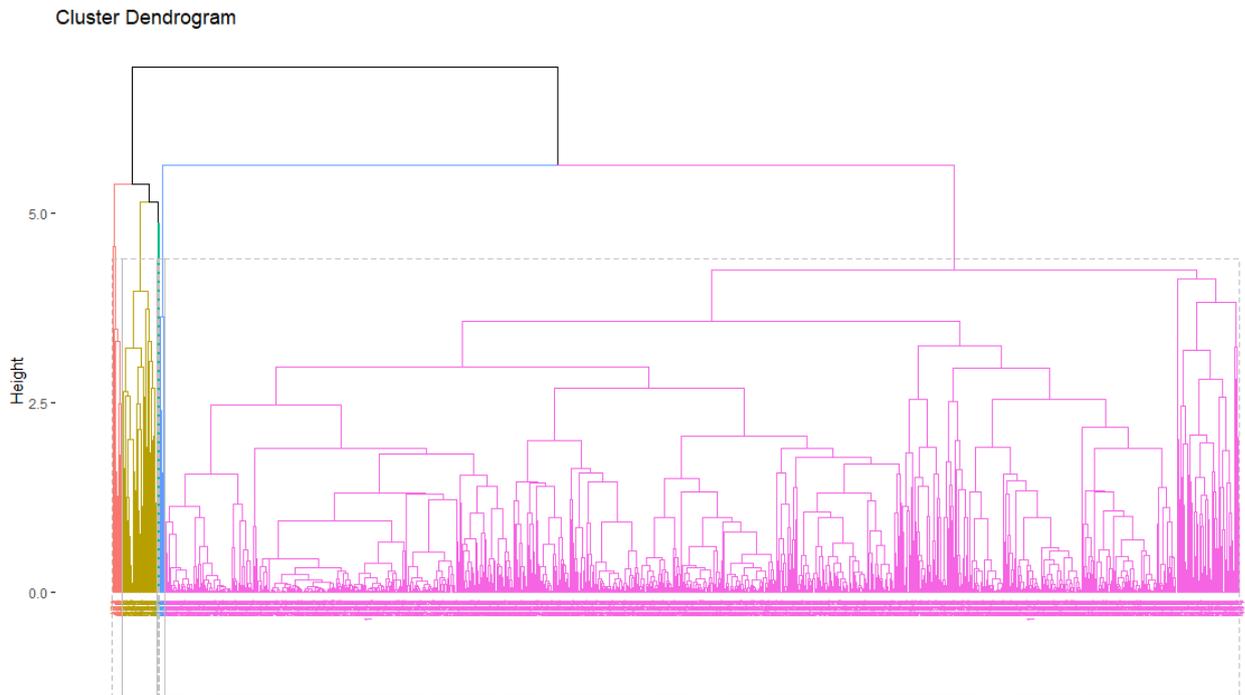


En el dendrograma se observa que el cluster 2 (239) es el que más viviendas agrupa, seguido del cluster 3 (234), el cluster 1 (207), el cluster 5 (187), el cluster 4 (93) y por último el cluster 6 (41).

Método de la media

Este método define la distancia entre clusters como la media de las distancias entre todas las parejas de elementos que la componen.

Gráfico 9: Dendograma con el método de la media



Fuente: elaboración propia

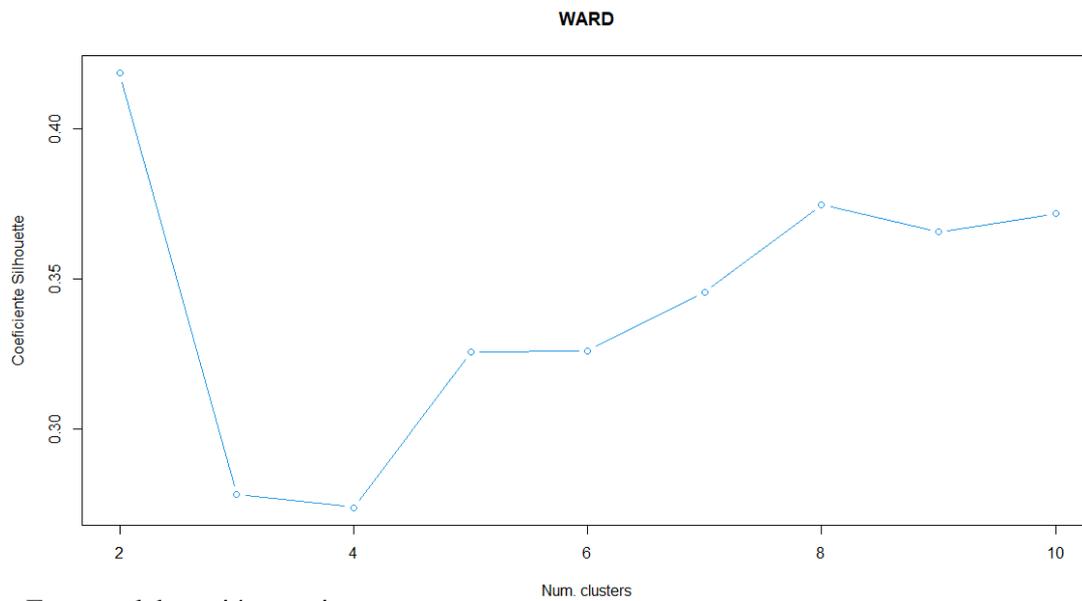
	1	2	3	4	5	6
	954	31	9	5	1	1

Como se observa en el dendograma, el grupo 1 (954) agrupa a la gran mayoría de las viviendas, mientras que los otros grupos poseen un menor número de viviendas, en concreto el grupo 2 (31), el grupo 3 (9), el grupo 4 (5), el grupo 5 (1) y el grupo 6 (1).

Tras haber visto los dos métodos del modelo jerárquico, el método Ward parece tener más sentido y ajustarse mejor a lo observado en el mapa de color, por lo que se utilizara el método Ward.

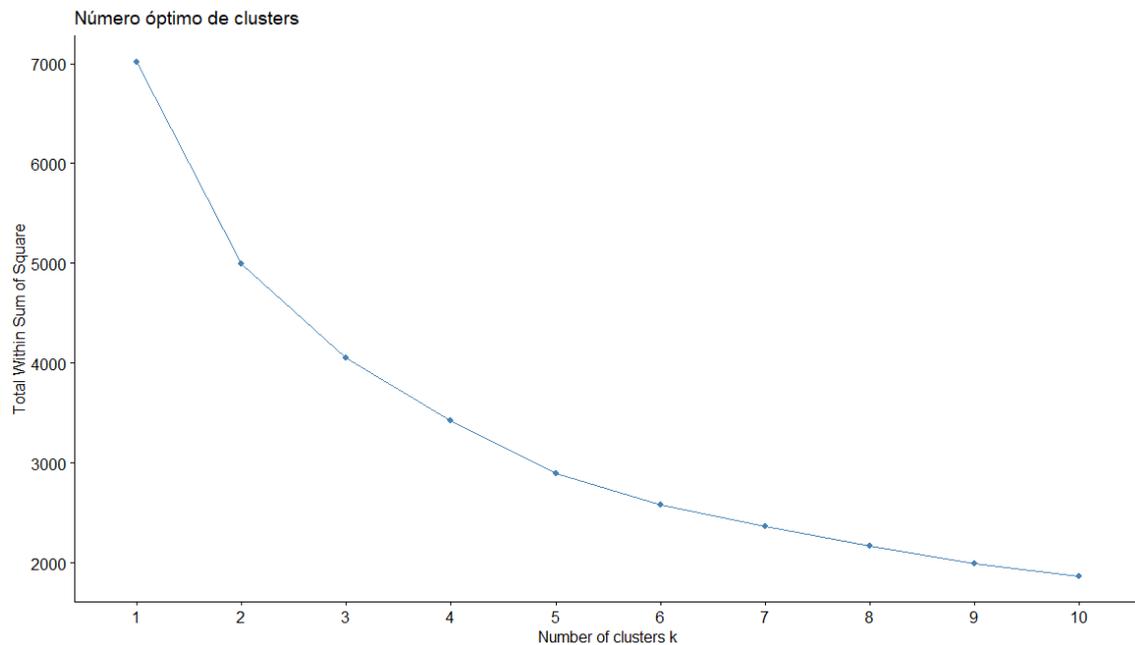
Para obtener el número de clusters óptimo para este algoritmo, se pueden aplicar distintos criterios como se observa a continuación, donde se aplicará el coeficiente de silhouette y el método de la suma de cuadrados intracluster.

Gráfico 10: Gráfico del número óptimo de clusters a través del coef. Silhouette.



Fuente: elaboración propia

Gráfico 11: Gráfico del número óptimo de clusters a través de la suma de cuadrados intracluster.



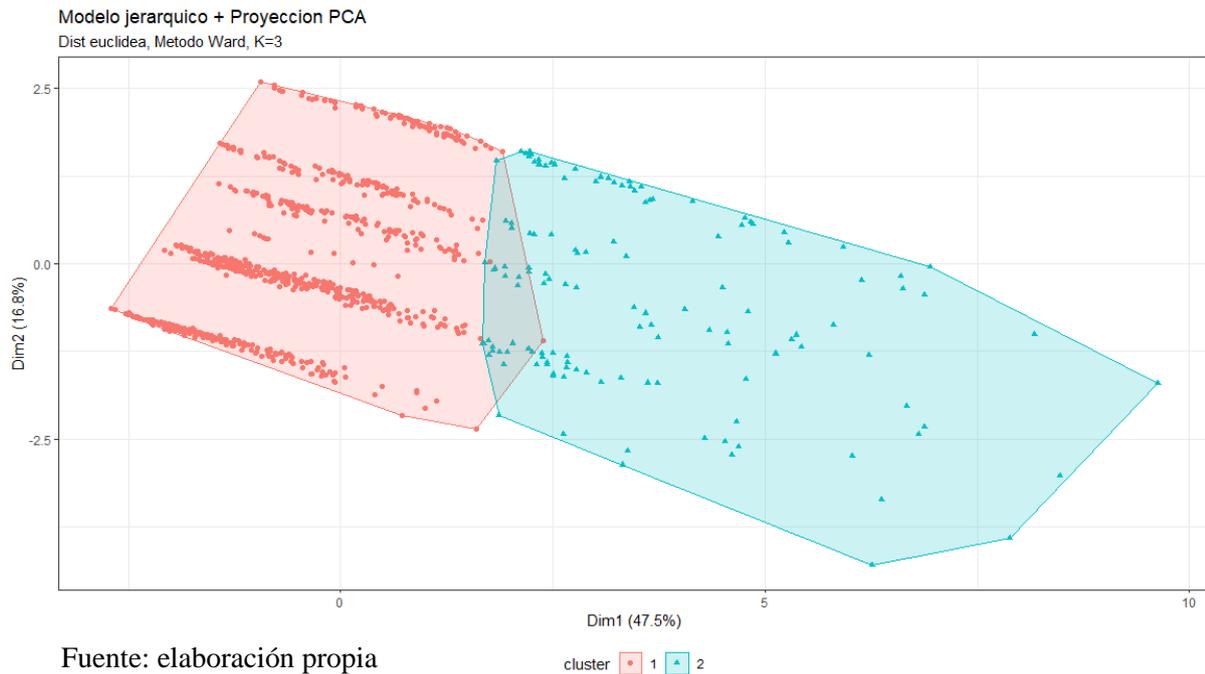
Fuente: elaboración propia

Como se observa en el primer gráfico, el mayor coeficiente de Silhouette indica que el número óptimo de clusters es de 2, en cambio en el método de la suma de cuadrados intraclusters no se ve claramente el codo bien diferenciado.

A continuación, se proyectará el PCA aplicándolo a los datos y los cluster obtenidos a través del método Ward que es el que más se ajustaba a lo observado en el mapa de

colores, se observa que la primera dimensión explica el 47,5% de toda la variabilidad del modelo y la segunda dimensión el 16,8%.

Gráfico 12: Gráfico scores PCA



Como se observa en el gráfico PCA, tras utilizar 2 cluster tal y como indicaba el coeficiente de silhouete, se observa en el gráfico de scores del PCA, que ambos clusters se solapan entre si. Esto puede ser debido a que se diferencian en algunas variables que no están correlacionadas con las dos primeras componentes principales.

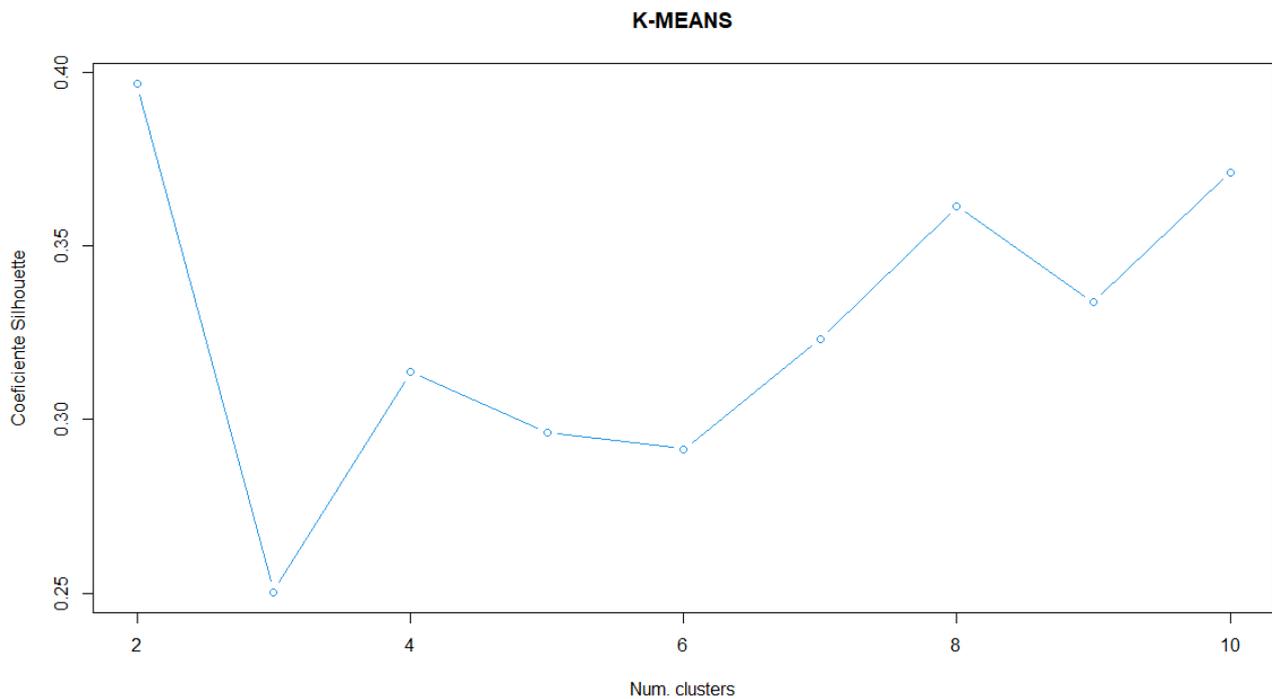
4.2.2 MODELOS DE PARTICIÓN

El objetivo de los modelos de partición es obtener una partición de los individuos en clusters, de tal forma que todos los individuos pertenezcan a uno de los posibles clusters y que estos clusters sean diferentes.

Método k-medias

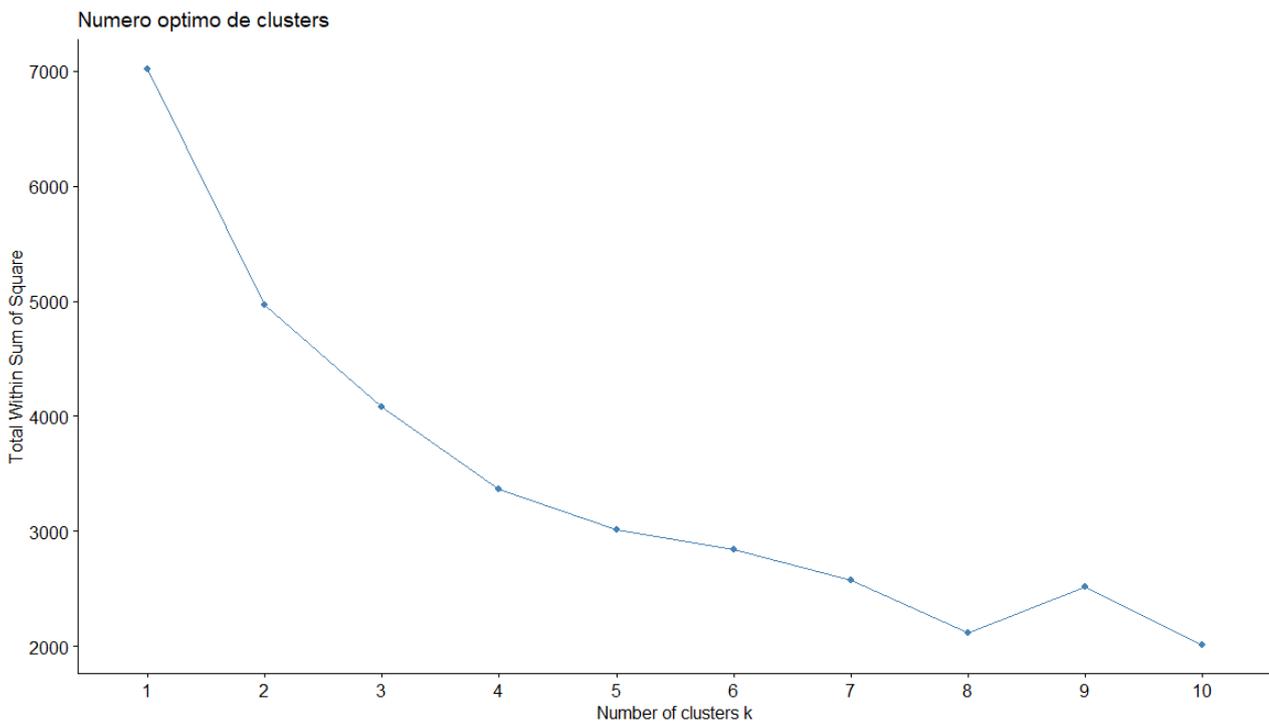
En el método k-medias, en primer lugar, es necesario determinar el número de clusters necesarios con los que se trabajara. Para ello, se seleccionan como centroides los clusters iniciales y calculando la distancia euclídea de cada elemento a estos, se asignan al cluster que tenga su centroide más próximo.

Gráfico 13: Gráfico del número óptimo de clusters a través del coef. Silhouette.



Fuente: elaboración propia

Gráfico 14: Gráfico del número óptimo de clusters a través de la suma de cuadrados intracluster.

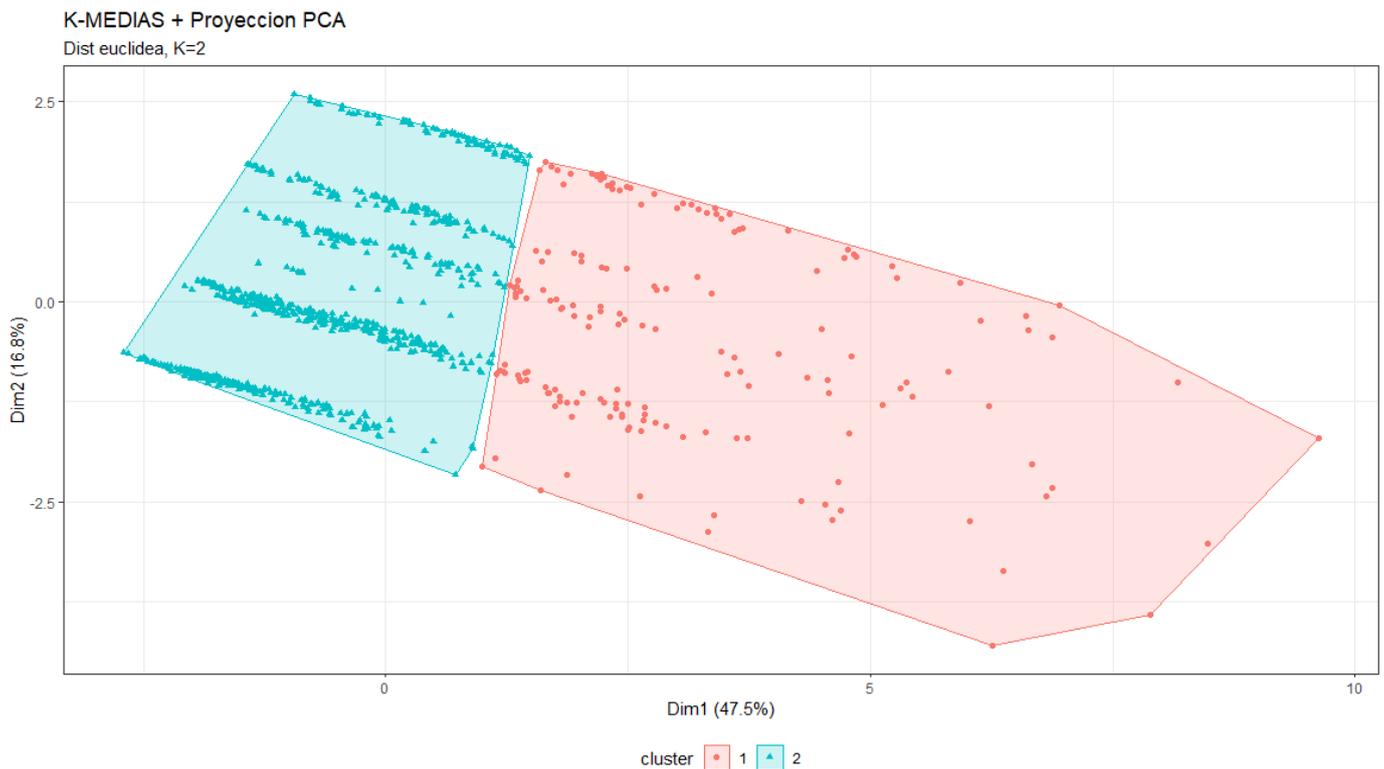


Fuente: elaboración propia

En el primer gráfico, se ha aplicado el coeficiente de silhouette para establecer el número óptimo de clusters. Este coeficiente indica la calidad de agrupamiento de los individuos, por lo tanto, contra más cerca este de 1, esto indicara que los individuos se han asignado al cluster correcto. Como se aprecia en el gráfico, el coeficiente de silhouette indica que el número de clusters optimo son 2, con un coeficiente de 0.40.

En el segundo gráfico, se ha aplicado el método de la suma de cuadrados intracluster, para determinar el número óptimo de clusters, para ello se ha utilizado el método del codo, a pesar de esto no se puede ver un codo bien diferenciado.

Gráfico 15: Gráfico scores PCA



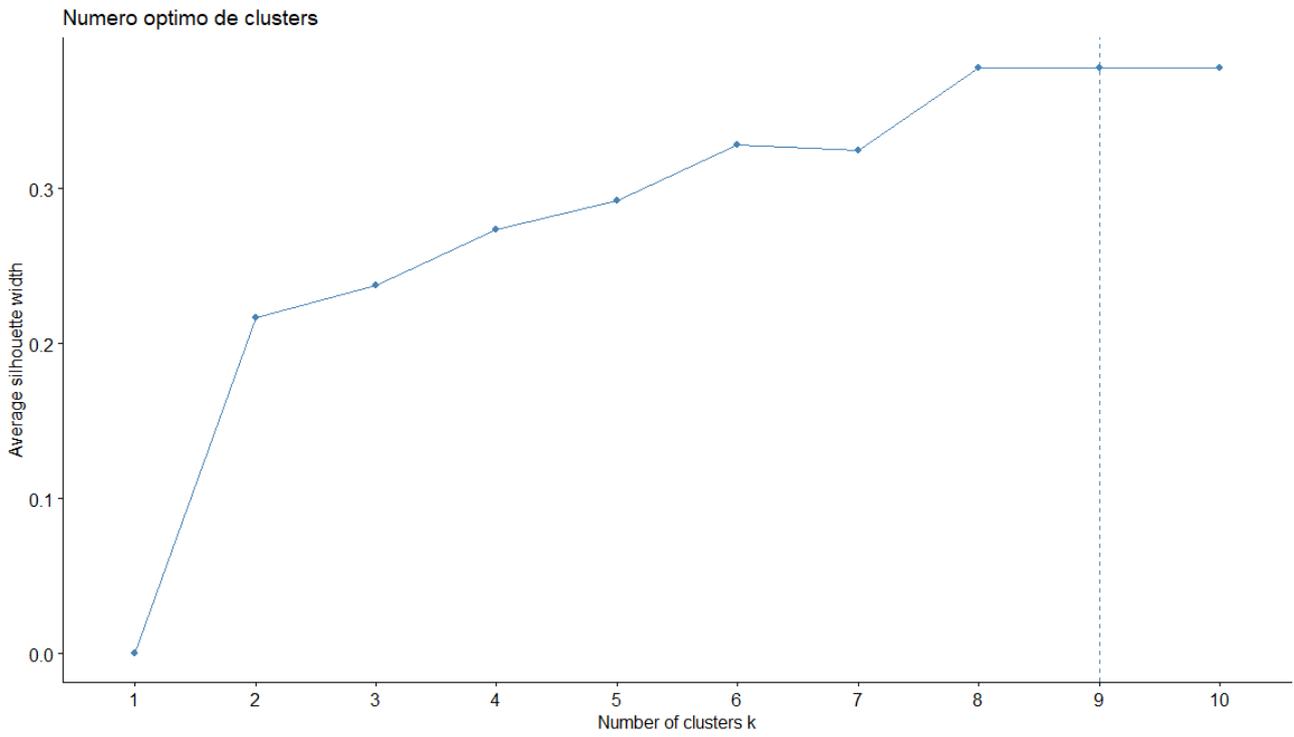
Fuente: elaboración propia

En el gráfico de scores del PCA, tras utilizar 2 clusters tal y como indicaba el coeficiente de silhouette, se observa en el gráfico que ambos clusters no se solapan entre sí, a pesar de estar muy cerca el uno del otro, por lo que ambos clusters están completamente diferenciados.

Método k-medoides

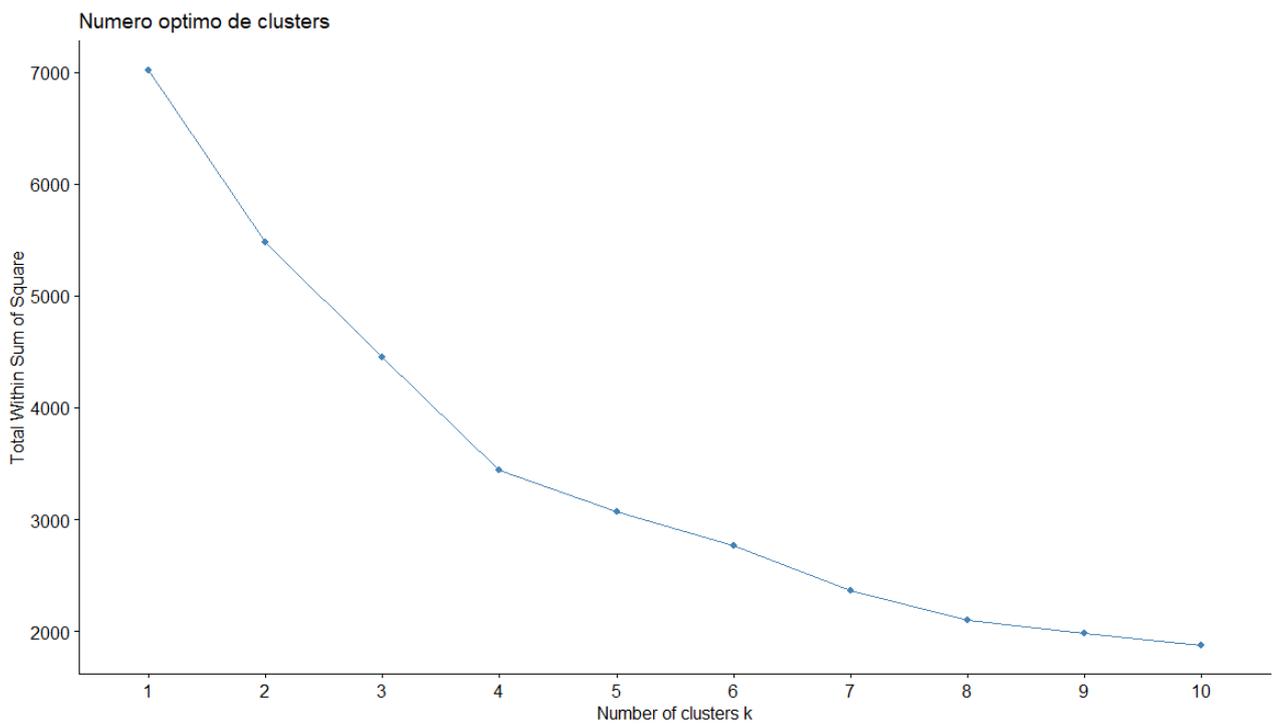
A continuación, se utilizará el método de los k-medoides que, en teoría, debería ser más robusto frente a los valores atípicos. En este algoritmo también es necesario determinar a priori el número de clusters.

Gráfico 16: Gráfico de numero de clusters optimo a través del coef. Silhoutte.



Fuente: elaboración propia

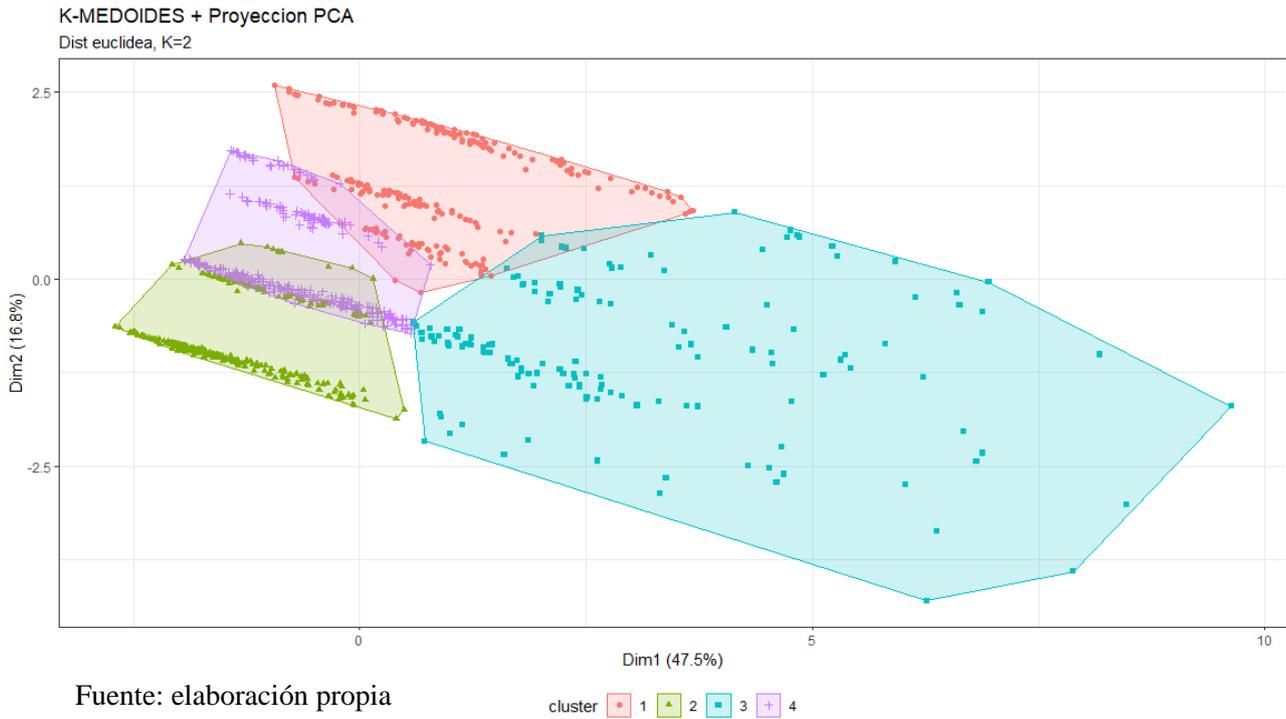
Gráfico 17: Gráfico del número óptimo de clusters a través de la suma de cuadrados intracluster.



Fuente: elaboración propia

Como se observa en el gráfico, el coeficiente de silhouette indica que el número de cluster óptimo son 9, en cambio, como se puede apreciar en el segundo gráfico, en el cual se ha utilizado el método de la suma de cuadrados intracluster, parece indicar que el número de clusters óptimo son 4. Por lo tanto, se optará a utilizar 4 clusters, ya que 9 clusters son demasiados y seguramente se solapan.

Gráfico 18: Gráfico scores PCA

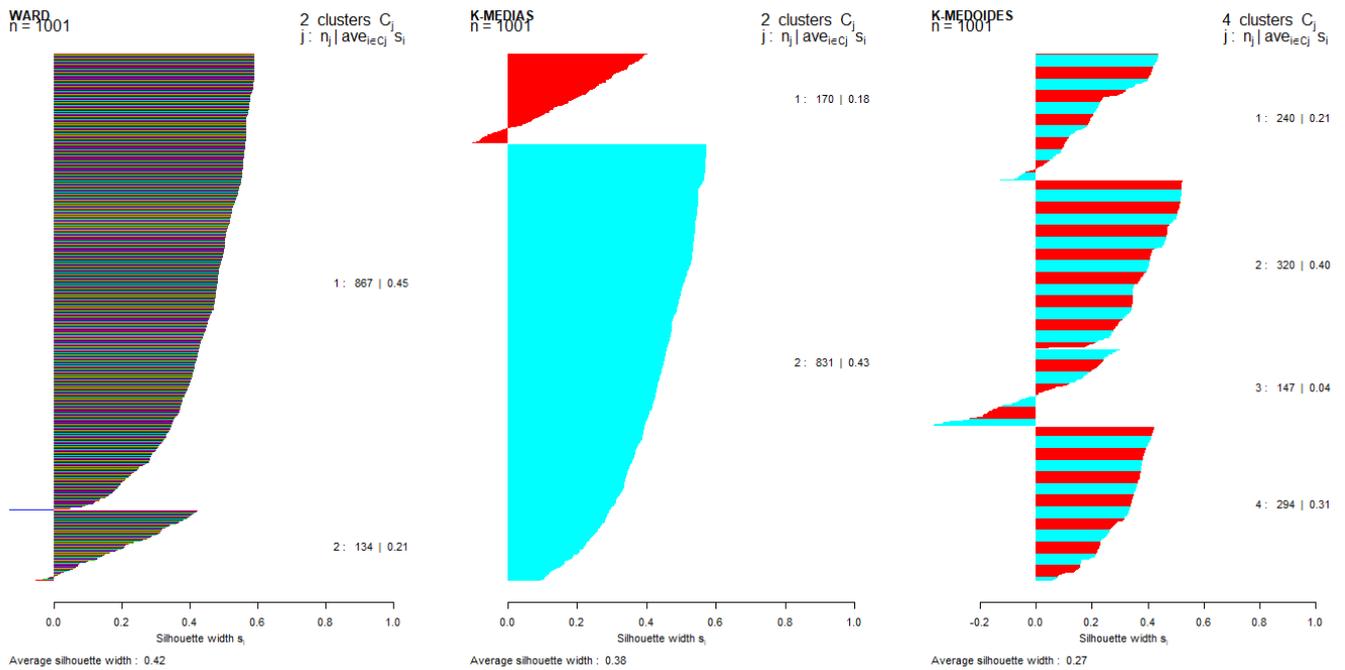


Como se observa en el gráfico scores PCA, existen clusters que se solapan entre sí, como el cluster 4, que está solapado por el cluster 1 y 2, como se ha explicado anteriormente, esto puede deberse a que se diferencian en algunas variables que no están correlacionadas con las dos primeras componentes principales. Por lo tanto, a partir del PCA se podría ver la posibilidad de reducir el número de clusters con el objetivo de que se reagrupan los clusters que se solapan.

4.3 SELECCIÓN Y VALIDACIÓN DEL MODELO

Tras haber generado los diferentes modelos jerárquicos y de partición, y observado los resultados, es difícil decantarse por un modelo, pero en principio el criterio k-medias y el Ward son los más fiables por los resultados, por lo tanto, se utilizará el coeficiente de silhouette para la selección y validación del modelo.

Gráfico 19: Gráfico comparativo de los modelos de clustering



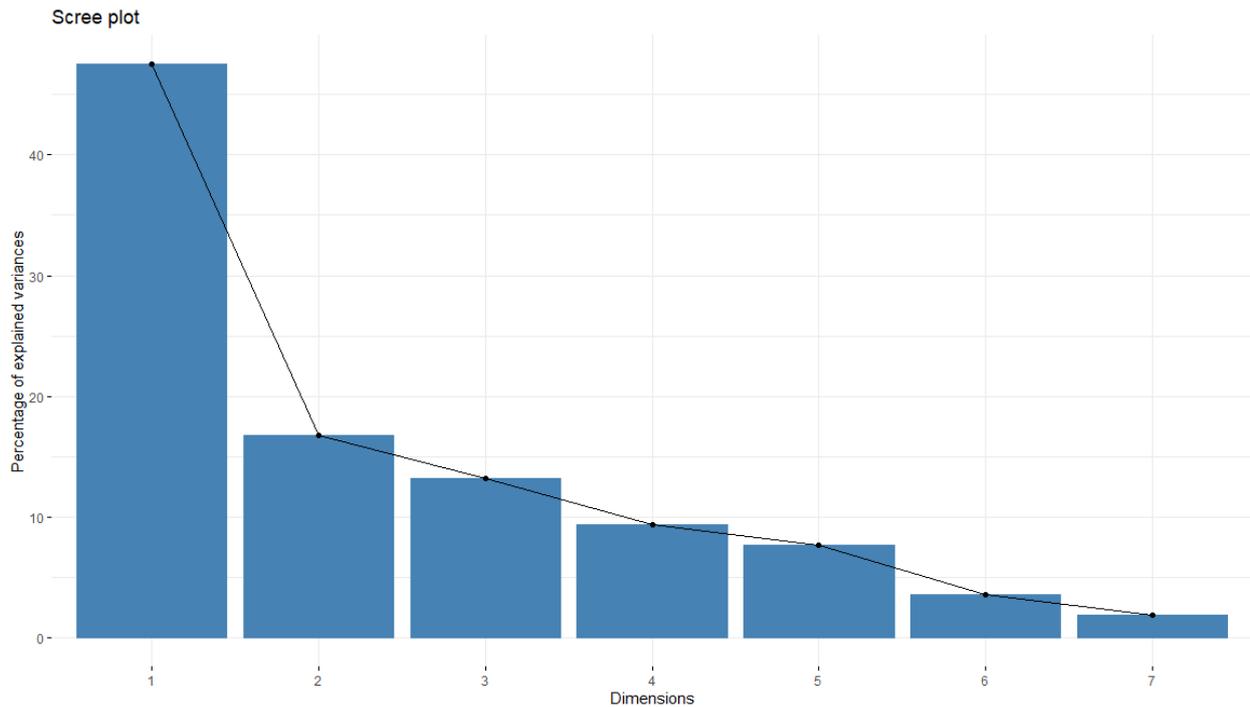
Fuente: elaboración propia

Tras ver los resultados de los diferentes modelos, se observa que el mejor modelo es el método ward, ya que es el modelo que presenta un mayor coeficiente de silhouette (0.42) frente al algoritmo k-medias (0.38) y el método k-medoides (0.27), además el método ward también es el que tiene un menor número de viviendas mal clasificadas.

4.4 INTERPRETACIÓN DE LOS RESULTADOS

En primer lugar, se va a utilizar el gráfico PCA, para observar cuales de las variables utilizadas en el análisis han tenido mayor contribución en la determinación de los clusters obtenidos a través del método Ward.

Gráfico 20: Gráfico de las dimensiones del modelo



Fuente: elaboración propia

Tabla 2: Tabla de las aportaciones de cada dimensión al modelo

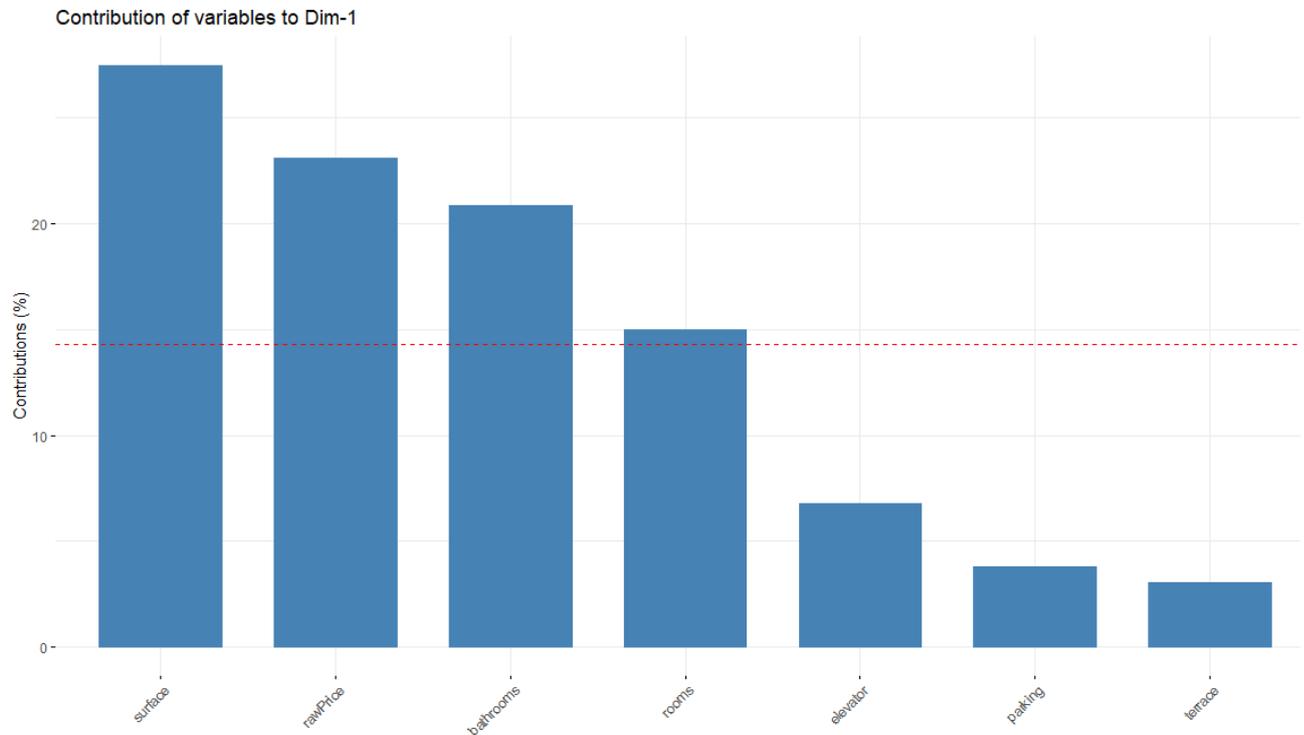
	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	3.3322483	47.527998	47.52800
Dim.2	1.1786368	16.810946	64.33894
Dim.3	0.9243919	13.184641	77.52358
Dim.4	0.6569532	9.370151	86.89374
Dim.5	0.5369397	7.658394	94.55213
Dim.6	0.2509062	3.578685	98.13082
Dim.7	0.1310509	1.869185	100.00000

Fuente: elaboración propia

Como se observa en el gráfico PCA, existen 7 dimensiones que explican el 100% de la variabilidad del modelo. La primera dimensión es la que más variabilidad explica con un 47,5%, seguido de la segunda dimensión que explica el 16,8%.

A continuación, se mostrará más detalladamente la contribución de las variables a las dos primeras dimensiones, que entre ambas explican el 64,3% de la variabilidad total del modelo.

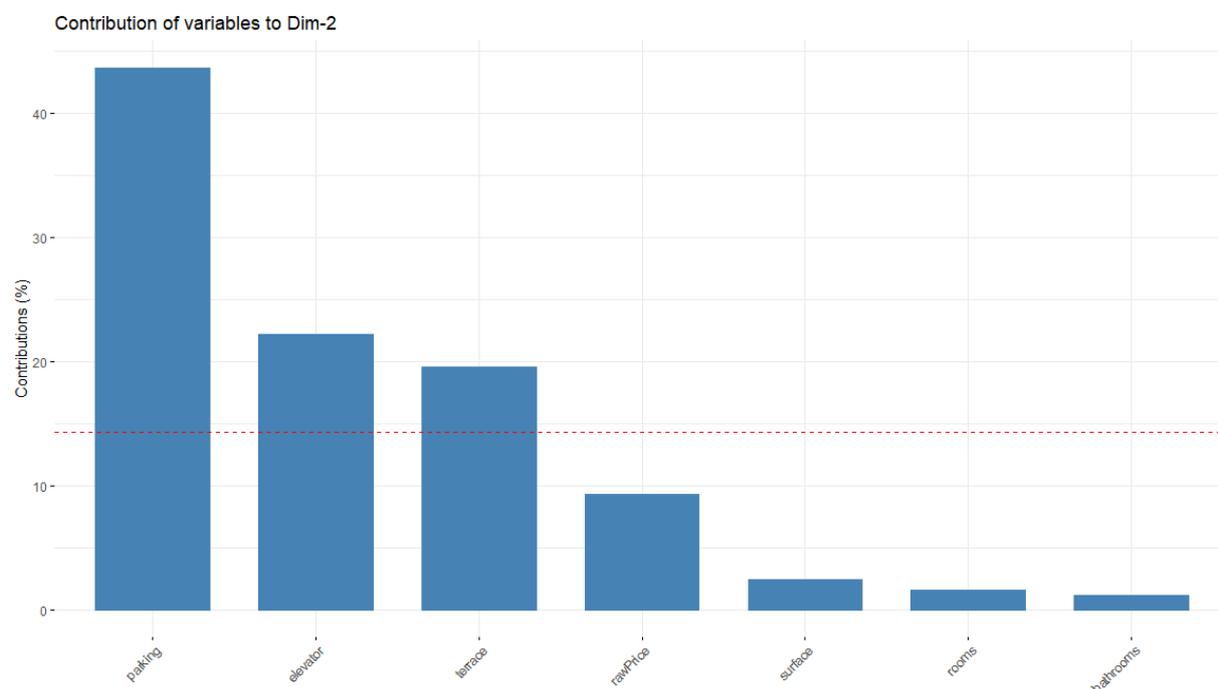
Gráfico 21: Gráfico de la contribución de las variables a la primera dimensión



Fuente: elaboración propia

Este gráfico muestra, la primera dimensión, la cual explica el 47,5% del total de la variabilidad del modelo, entre las variables, las que más influyen son la superficie, los baños, el precio y las habitaciones, parece ser que esta dimensión hace referencia al tamaño de las viviendas.

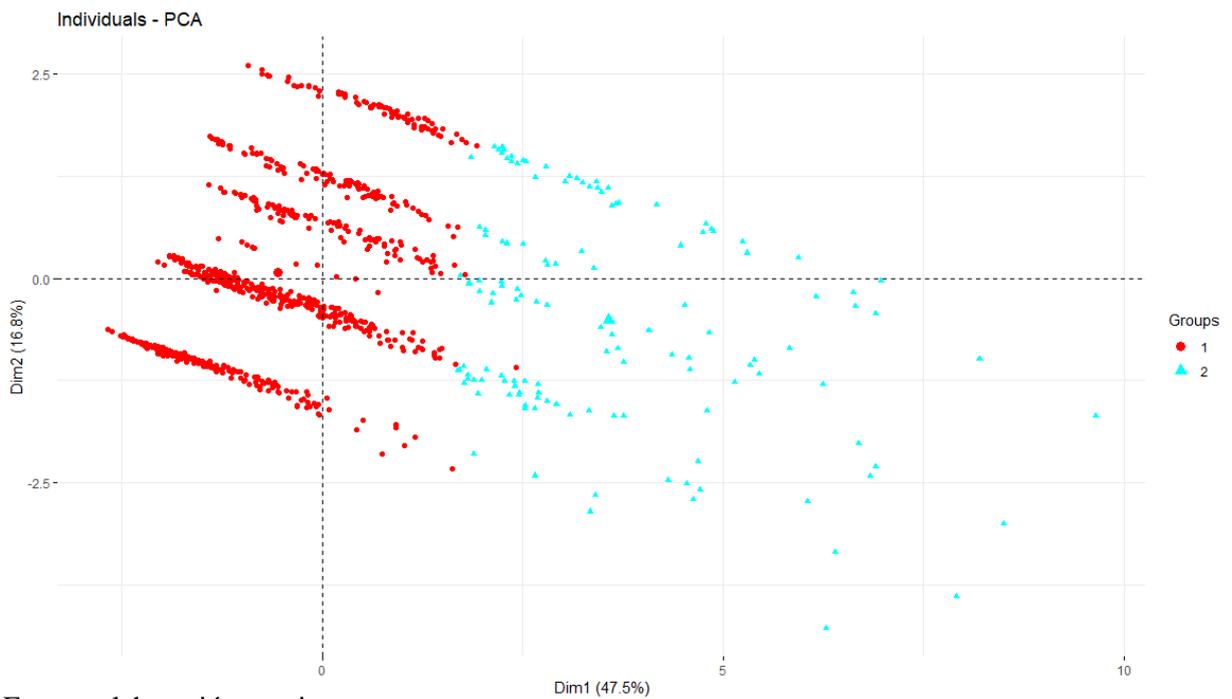
Gráfico 22: Gráfico de la contribución de las variables a la segunda dimensión



Fuente: elaboración propia

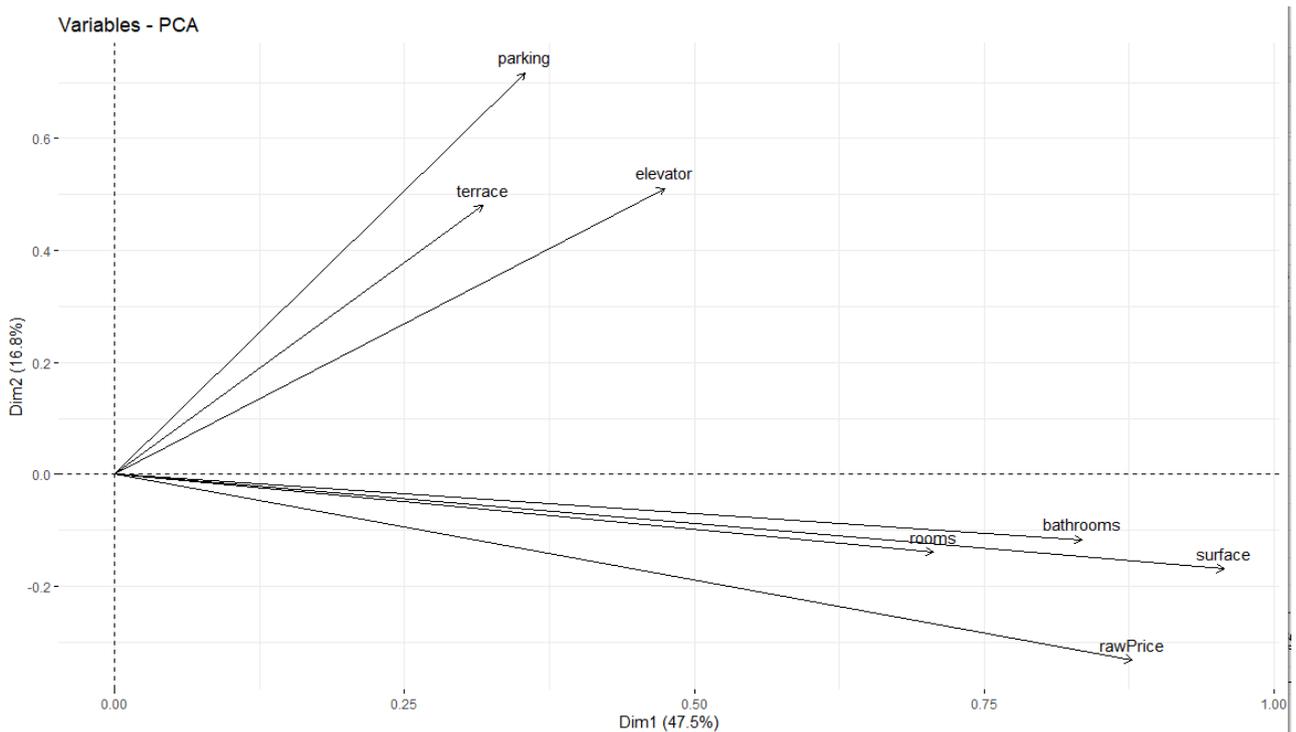
El anterior gráfico muestra, la segunda dimensión, la cual explica el 16,8% de la variabilidad total del modelo, entre las variables que más influyen se encuentran el parking, la terraza y el ascensor, esta dimensión hace referencia a los extra que trae consigo la vivienda

Gráfico 23: Gráfico de distribución de los clusters



Fuente: elaboración propia

Gráfico 24: Gráfico de distribución de las variables



Fuente: elaboración propia

Tras utilizar el método Ward, que es el que ofrecía un mayor coeficiente de silhouette, las viviendas se han agrupado en 2 cluster, el primer cluster está formado por 867 viviendas y el segundo por 134.

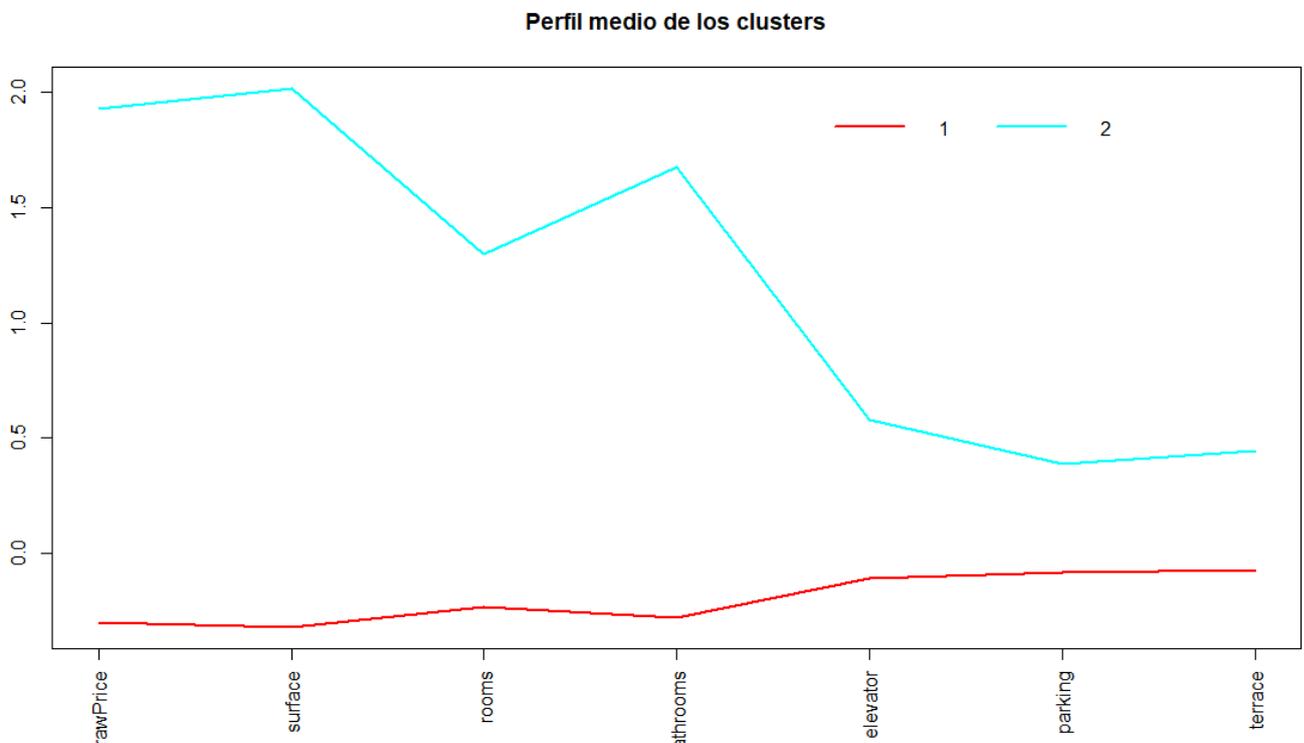
Al observar ambos gráficos, se puede apreciar que las viviendas del grupo 2 tienen mayor superficie, más habitaciones, más baños y por lo tanto tienen un mayor precio.

A simple vista, se puede apreciar que las viviendas del grupo 1 pertenecen a las personas con un menor poder adquisitivo en comparación a las personas del grupo 2, debido a que las viviendas que poseen tienen una menor superficie y precio.

En cuanto al aparcamiento, la terraza y el ascensor no son una característica distintiva de un grupo particular, sino que ambos grupos poseen viviendas con algunas o todas estas características.

A continuación, se va a realizar un gráfico descriptivo del perfil de ambos cluster con el objetivo de ver las diferencias entre ellos. Para ello, se calculará la media de cada variable para cada cluster.

Gráfico 25: Gráfico del perfil medio de los cluster



Fuente: elaboración propia

Tabla 3: Tabla de las diferencias de las medias entre ambos clusters

	c1	c2
rawPrice	-0.30	1.93
surface	-0.31	2.01
rooms	-0.23	1.30
bathrooms	-0.28	1.67
elevator	-0.11	0.58
parking	-0.08	0.39
terrace	-0.07	0.45

Fuente: elaboración propia

Entre el perfil medio de cada cluster vemos una notable diferencia entre ambos. En el segundo cluster las viviendas tienen un mayor precio, una mayor superficie, más habitaciones y baños, y se diferencian mucho en estos aspectos en comparación a las viviendas del primer grupo. En cambio, no existe tanta diferencia entre las viviendas de ambos grupos cuando se trata de si la vivienda posee ascensor, terraza o parking, ya que es una característica que poseen determinadas viviendas de ambos grupos, aunque más las viviendas del grupo 2.

Uno de los mayores problemas en el mercado inmobiliario, es que no se conoce con exactitud el valor de los activos, por lo que se utilizan referencias próximas, aunque solo tienen carácter orientativo. Por lo tanto, en la valoración de un activo, existe la posibilidad de error, ya que al fijar el precio adecuado y no conocer el precio exacto, se puede cometer un error, ya que cuando se fija el precio se tiene en cuenta referencias próximas, por lo tanto, no son del todo acertadas ya que se realizan pocas operaciones y las características de cada inmueble varían.

A continuación, se realizará un método de análisis supervisado de regresión, con el objetivo de desarrollar modelos que estimen el valor de un inmueble en función de sus características, lo que podrá aportar valor a potenciales inversores de este mercado. Podrán comparar el precio de oferta de los inmuebles con el obtenido mediante el modelo, antes de tomar decisiones de inversión o desinversión.

Además, este modelo también podrá ser útil, para las personas que hayan decidido poner en venta su inmueble, y quieran tener un precio estimado de la vivienda en base a sus características.

4.5 MÉTODO SUPERVISADO

El método supervisado se caracteriza por que se conoce a priori el objetivo buscado, es decir, para los individuos conoces alguna variable respuesta asociada a ellos. Para realizar este método, se utilizarán diferentes modelos y se escogerá el modelo más válido, con el objetivo de predecir el comportamiento de un atributo.

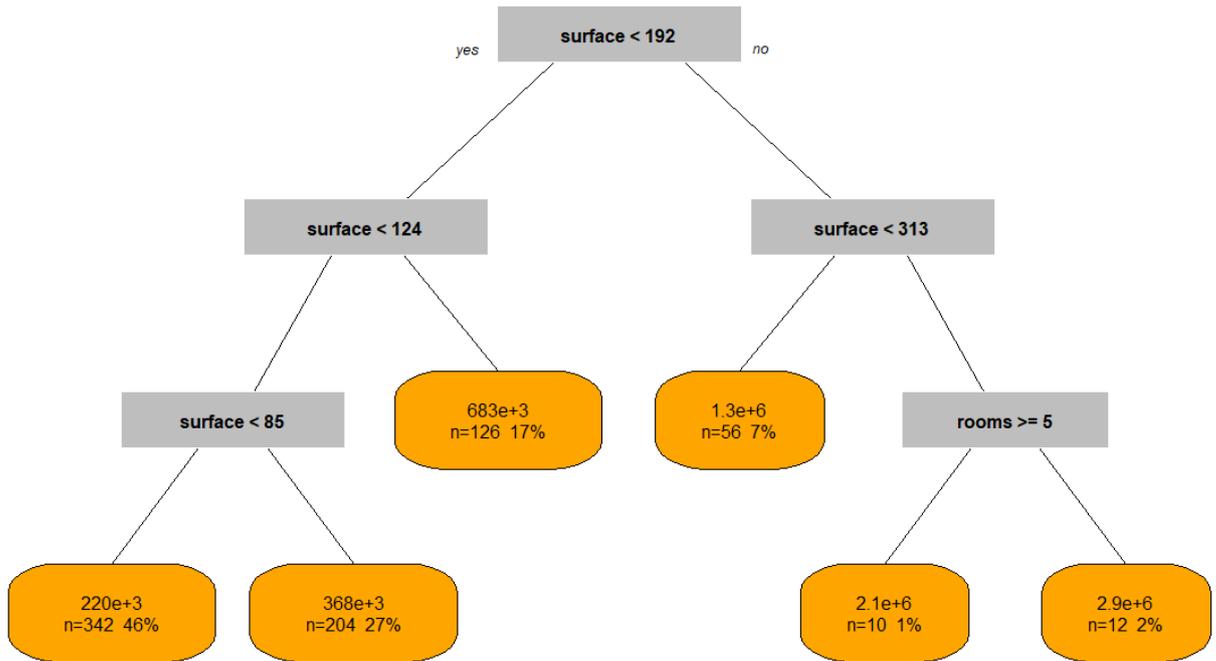
En primer lugar, se debería explorar, limpiar y procesar los datos, en este caso no será necesario ya que los datos que se van a utilizar ya han sido preparados anteriormente. También, sería primordial reducir la dimensión de los datos, pero al solo tener 8 variables esto no será necesario.

En segundo lugar, se realizará la partición de los datos, para ello se dividirán los datos en dos particiones: entrenamiento y validación. Los datos de entrenamiento se utilizarán para construir el modelo y los de validación se usarán para comprobar cómo funciona el modelo cuando se aplica a nuevos datos. Para la partición se dividirá aleatoriamente los datos, correspondiendo el 75% del total a entrenamiento y el 25% restante a validación.

4.5.1 ÁRBOL DE REGRESIÓN

A continuación, se realizará el árbol de regresión para predecir el precio de la vivienda, para ello se ajustará el árbol, teniendo en cuenta la variable “Rawprice”, como dependiente, y el resto de las variables serán independientes.

Gráfico 26: Árbol de regresión



Fuente: elaboración propia

Un árbol de regresión es una jerarquía de pruebas lógicas sobre algunas de las variables explicativas, por lo tanto, los modelos basados en árbol seleccionan automáticamente las variables más relevantes, y por esta razón, no todas las variables necesitan aparecer en el árbol final.

El árbol está formado por diferentes nodos, los cuales tienen dos ramas que están sujetas a si la vivienda cumple una condición de las variables predictoras. Cuando se llega al final de una rama y esta no tiene un nodo de decisión, es que se ha llegado a un nodo terminal. Se llega a este nodo cuando los subgrupos tienen un tamaño mínimo o ninguna mejora es posible.

Como el árbol anterior se ha formado para ajustarse perfectamente a los datos introducidos, haría falta podarlo para evitar el sobreajuste, debido a que si se utiliza este árbol para nuevos datos dará un bajo desempeño. Por lo tanto, con la poda del árbol lo que se consigue es obtener el árbol óptimo, para facilitar su interpretabilidad y evitar problemas de sobreajuste.

Para la poda del árbol se utilizará la regla X-SE, para ello se generará un árbol con el menor xerror:

Tabla 4: Tabla de error de los diferentes arboles

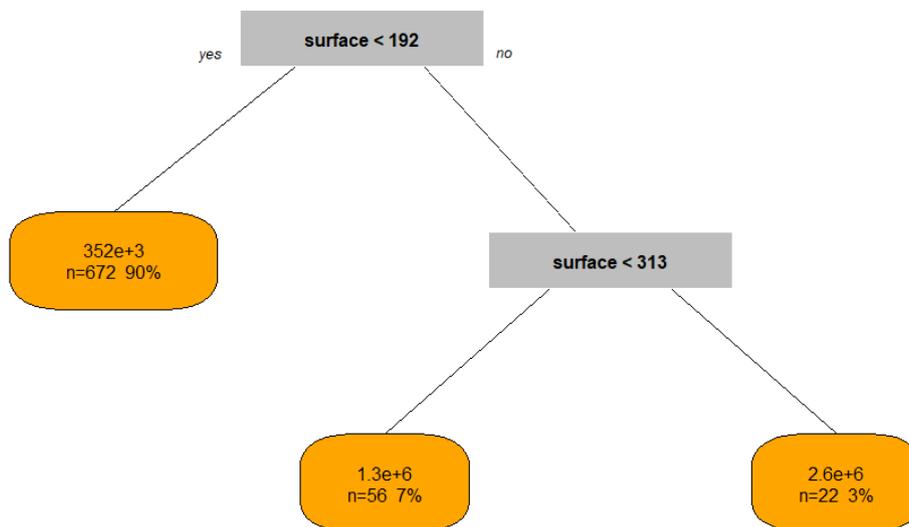
```

n= 750
      CP nsplit rel error  xerror  xstd
1 0.521889    0  1.00000 1.00328 0.154593
2 0.097568    1  0.47811 0.51918 0.073791
3 0.071244    2  0.38054 0.43275 0.063666
4 0.013773    3  0.30930 0.37024 0.060809
5 0.011629    4  0.29553 0.38622 0.064635
6 0.010000    5  0.28390 0.37976 0.064559
  
```

Fuente: elaboración propia

Como se observa en la tabla, el árbol 4 es el que presenta un menor xerror (0.37024), con un cp=0.0137, por lo tanto, el árbol podado sería el siguiente:

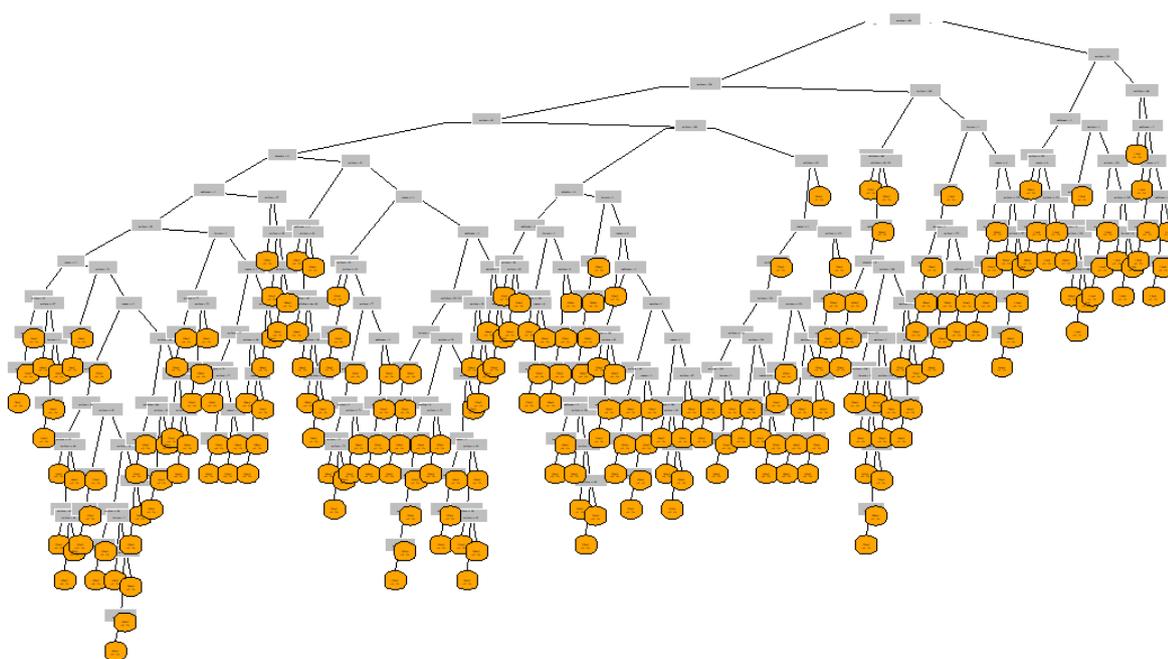
Gráfico 27: Árbol de regresión podado



Fuente: elaboración propia

Para construir los anteriores arboles se ha utilizado el árbol más grande:

Gráfico 28: Árbol de regresión completo



Fuente: elaboración propia

4.5.2 RANDOM FOREST

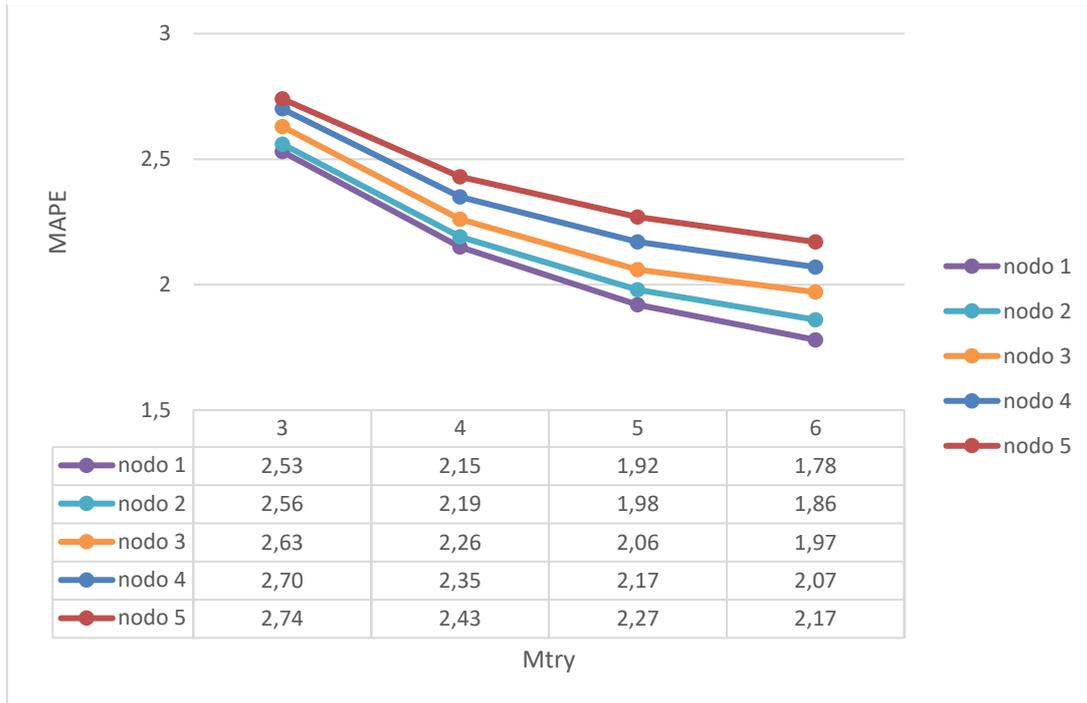
El bosque aleatorio añade una nueva capa de aleatoriedad al proceso bagging. Además, en la construcción de cada árbol la división de las ramas de cada nodo se lleva a cabo, no a partir del conjunto total de predictores, sino a través de un subconjunto seleccionado al azar.

Con este algoritmo se realizará una combinación de diferentes árboles de decisión para obtener modelos más estables y con menos propensión al sobreajuste. Este modelo también proporcionará una medida de la importancia de las variables predictoras y la proximidad de unos datos a otros.

Para que funcione este algoritmo la base de datos no debe poseer ningún dato faltante.

En primer lugar, es necesario determinar la muestra aleatoria de variables que van a intervenir en la división de un nodo, en teoría el valor de m_{try} que ofrece mejores predicciones en regresión es el número de variables dividido por tres, en este caso m_{try} tomaría el valor de 3. Pero como este valor es orientativo, también se realizará con otros valores y se seleccionará el que presente mejores resultados de predicción. Para ello se utilizará la medida de bondad de ajuste conocida como $mape$, que es el error porcentual absoluto medio. Además, también será útil determinar el parámetro $nodesize$ donde se especifica el tamaño mínimo de los nodos terminales de los árboles. Contra mayor sea este parámetro, se cultivarán árboles más pequeños.

Gráfico 29: Gráfico del MAPE de Random Forest

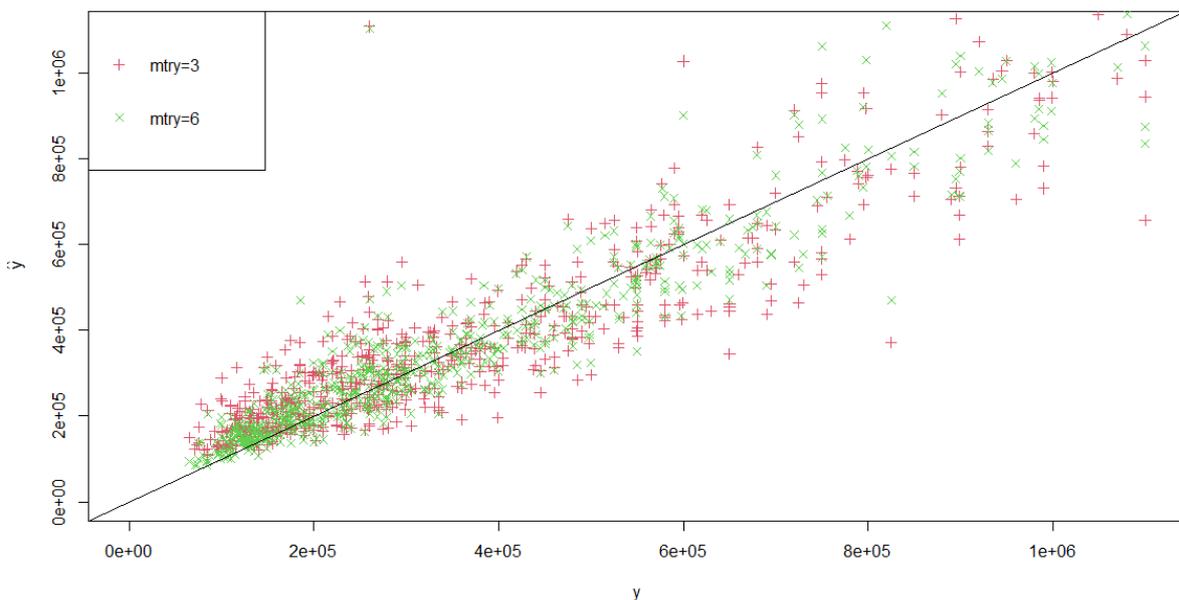


Fuente: elaboración propia

Como se puede observar en el gráfico y en la tabla, el resultado que presenta un mejor Mape se obtiene a través de utilizar un mtry=6 y nodesize=1, consiguiendo un Mape del 17,8%, lo que indica que la predicción esta errada en un 17,8%.

A continuación, se realizará un gráfico para observar la mejoría del modelo al utilizar un mtrv=6. en vez de un mtrv=3.

Gráfico 30: Gráfico de comparación de la predicción entre mtry=3 y mtry=6



Fuente: elaboración propia

Tabla 5: Tabla de la aportación de las variables al modelo

```

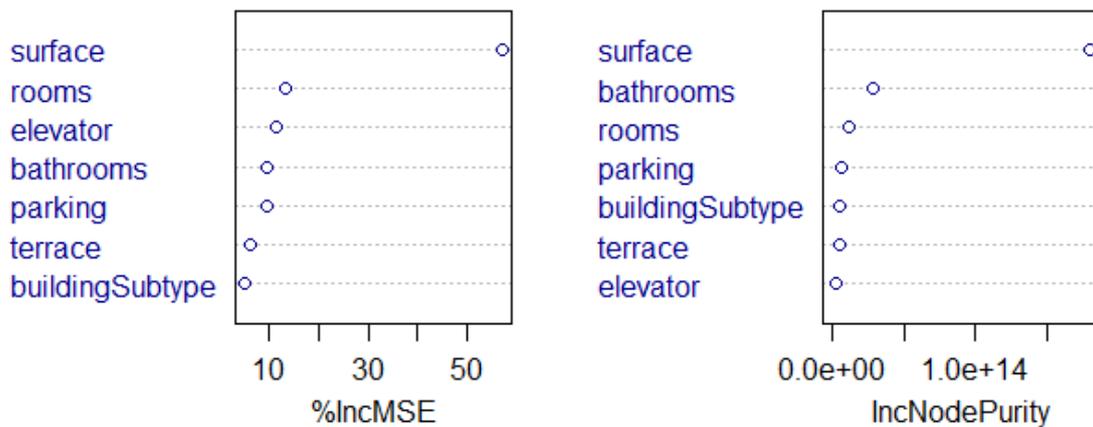
Type of random forest: regression
Number of trees: 500
No. of variables tried at each split: 6

Mean of squared residuals: 106150546313
% Var explained: 66.7
    
```

	%IncMSE	IncNodePurity
surface	57.007164	1.797840e+14
rooms	13.308635	1.156714e+13
bathrooms	9.622282	2.787334e+13
elevator	11.395998	2.506295e+12
parking	9.601528	5.874229e+12
terrace	6.164398	4.634975e+12
buildingSubtype	5.198866	4.943113e+12

Fuente: elaboración propia

Gráfico 31: Gráfico mean decrease accuracy



Fuente: elaboración propia

Observando el gráfico de mean decrease accuracy (%incMSE) y la anterior tabla, las variables que más influyen a la hora de determinar el precio de una vivienda son, en gran medida la superficie, también influyen, aunque bastante menos las habitaciones y el ascensor, y con un menor peso el resto de las variables.

4.5.3 VECINO MÁS PRÓXIMO

Este modelo predice el valor de una observación basándose en el valor de las observaciones más cercanas, se trata de un modelo simple e intuitivo que se suele utilizar para la predicción. Para este modelo todas las variables deben ser numéricas, además no puede haber valores faltantes.

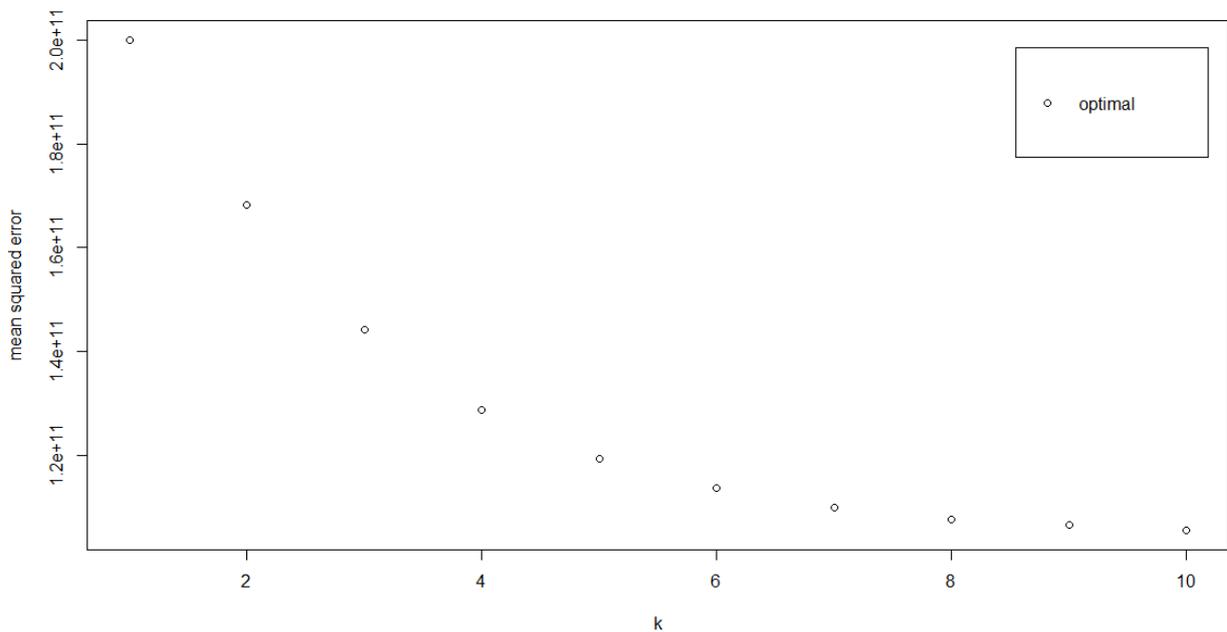
En primer lugar, se obtendrá el valor k , el cual indica el número de vecinos, este valor se puede obtener de dos formas, haciendo la raíz cuadrada del número de variables (este método indica que $K=3$) o a través del paquete `kknn` que obtiene el k óptimo a través de un algoritmo.

Tabla 6: Tabla de k óptimo

```
Type of response variable: continuous  
minimal mean absolute error: 171143.2  
Minimal mean squared error: 1.05761e+11  
Best kernel: optimal  
Best k: 10
```

Fuente: elaboración propia

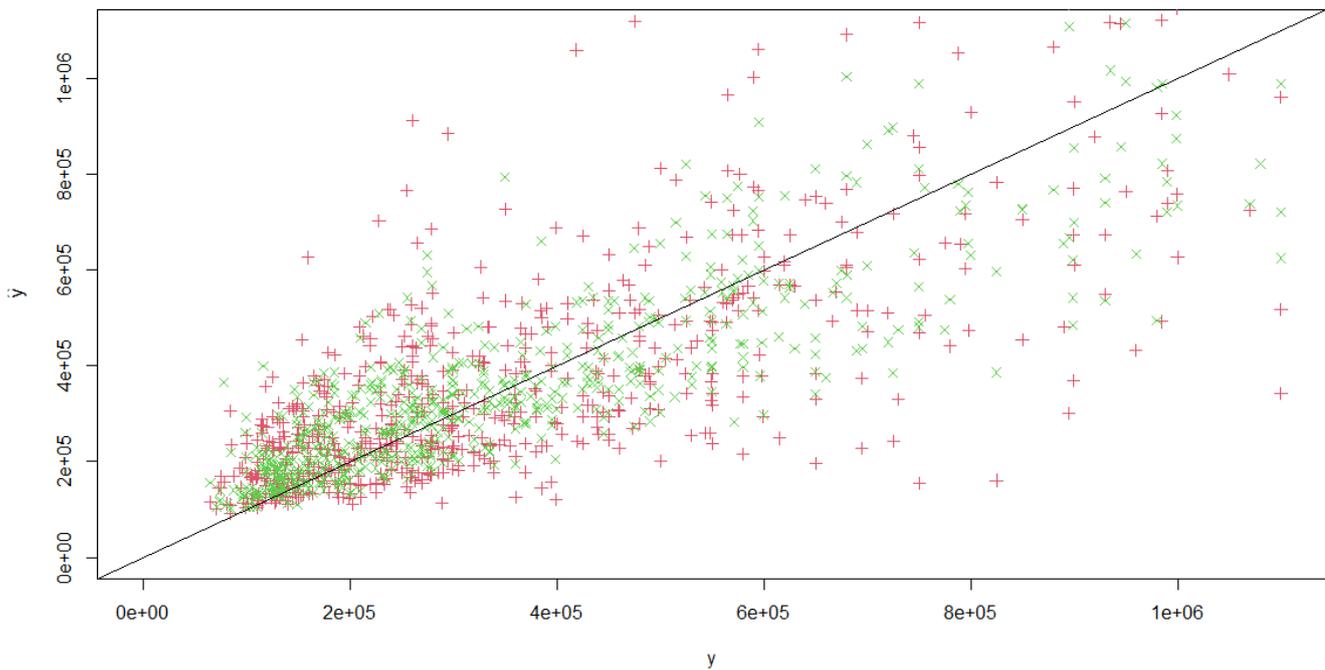
Gráfico 32: Gráfico de k óptimo



Fuente: elaboración propia

Este gráfico muestra el error para cada k , por lo tanto, el valor óptimo de k son 10, ya que es el que presenta un menor error.

Gráfico 33: Gráfico de comparación de la predicción entre k=3 y K=10



Fuente: elaboración propia

Como se observa en gráfico, al utilizar $k=10$ (puntos verdes) lo cual indicaba el paquete `kknn` que era el número de óptimo de k , ha mejorado el modelo respecto a utilizar $k=3$ (puntos rojos), que es la raíz cuadrada del número de variables, por lo que se utilizara $k=10$ para la predicción del vecino más próximo.

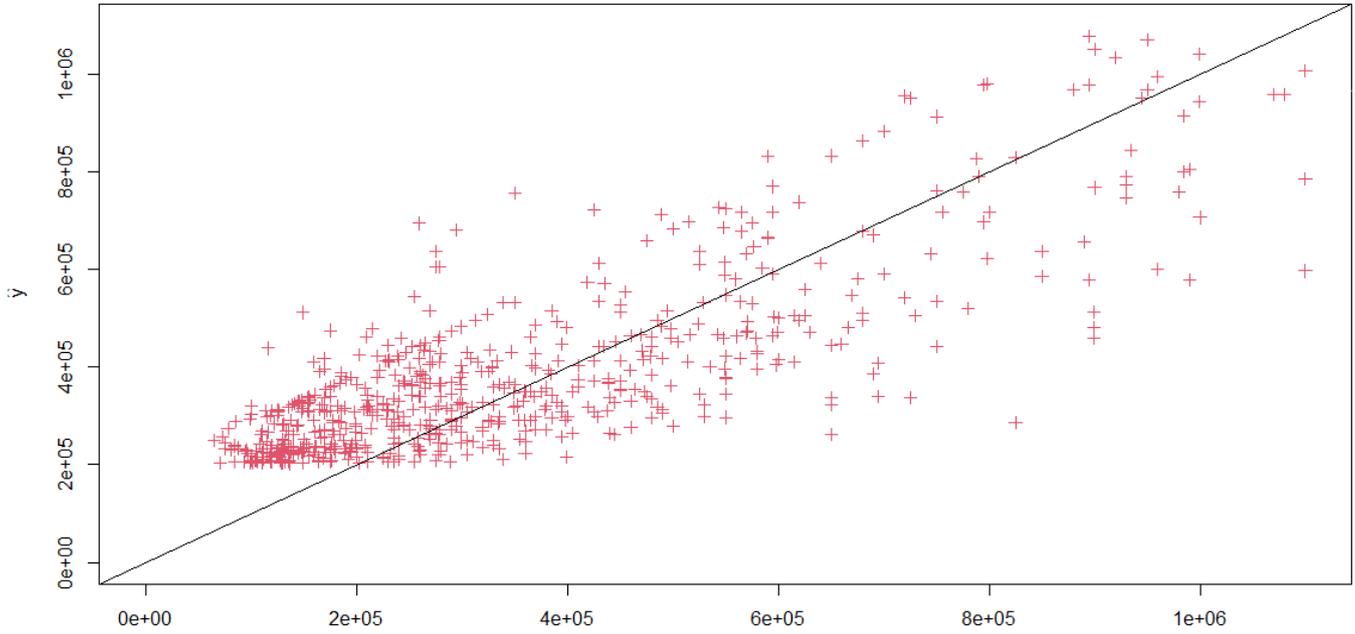
4.5.4 MÁQUINAS DE SOPORTE VECTORIAL

Las máquinas de soporte vectorial son un conjunto de algoritmos que buscan un hiperplano que separa de forma óptima los puntos de una clase de la de otra, que eventualmente han podido ser previamente proyectados a un espacio de dimensionalidad superior. Su nombre es debido al vector formado por los puntos más cercanos al hiperplano de separación que se denomina vector de soporte

Para la realización de este modelo, existen dos librerías en R: `kernlab` y `e1071`, por lo tanto, se realizará el modelo en las dos librerías y se seleccionará el modelo que presente mejores resultados.

Librería “kernlab”:

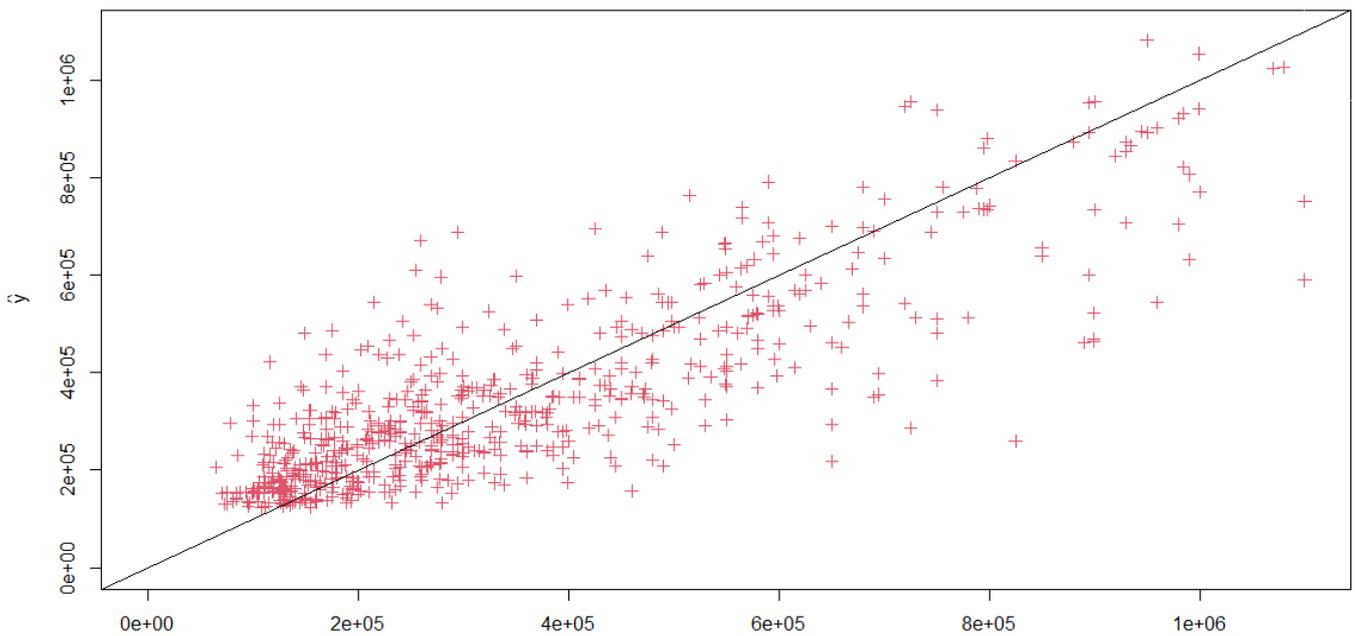
Gráfico 34: Gráfico de la predicción del modelo frente a las observaciones



Fuente: elaboración propia

Librería “e1071”:

Gráfico 35: Gráfico de la predicción del modelo frente a las observaciones



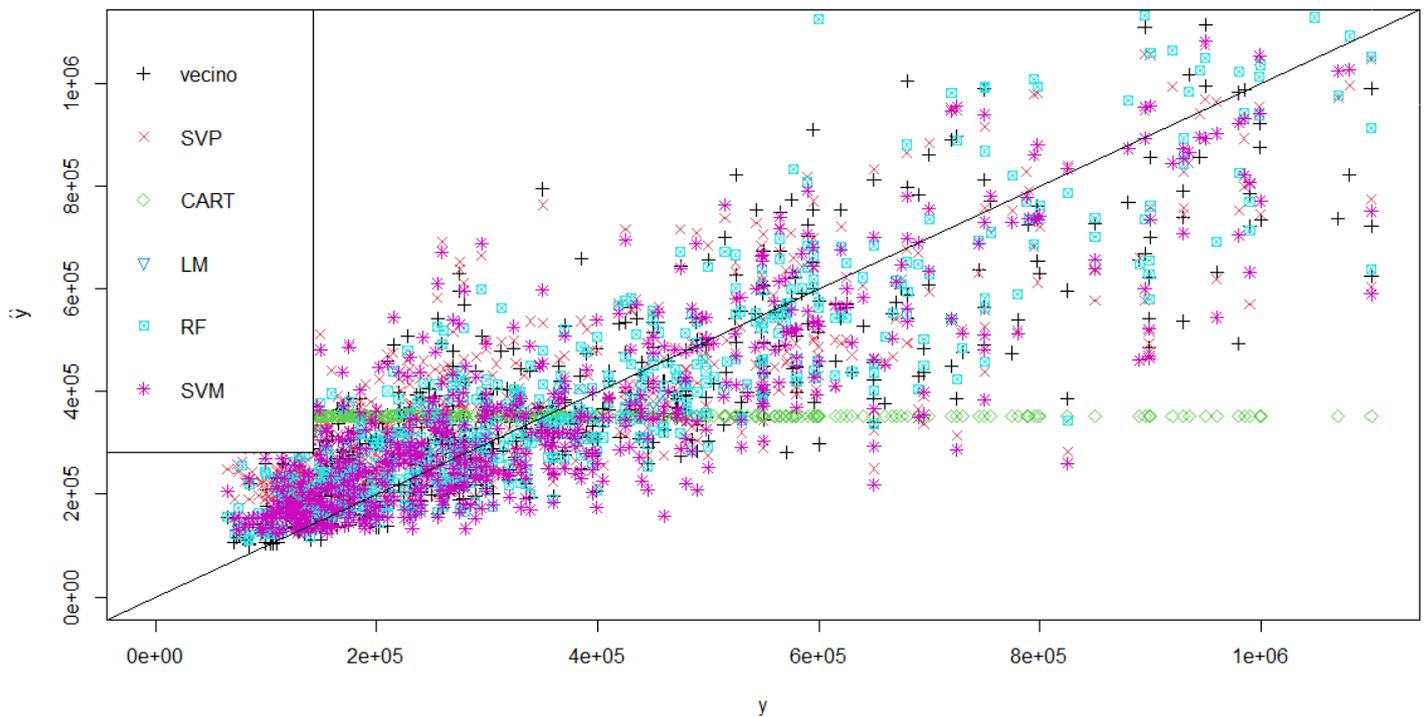
Fuente: elaboración propia

Al observar ambos gráficos y ver las predicciones de las dos librerías, se puede apreciar que la librería “e1071” obtiene una mejor predicción ya que los valores se acercan más a la diagonal. Por lo que se utilizara el modelo de esta librería para la predicción.

4.6 COMPARACIÓN Y SELECCIÓN DEL MODELO

A continuación, se procederá a seleccionar el modelo de predicción que mejor resultado ha aportado, para ello se representará gráficamente las predicciones de los modelos frente a las observaciones, por lo tanto, el modelo que más se acerque a la diagonal será el mejor modelo.

Gráfico 36: Gráfico de las predicciones de todos los modelos frente a las observaciones



Fuente: elaboración propia

En el gráfico no se aprecia claramente el modelo que más se ajusta a la diagonal, por lo que, se procederá a calcular las distancias de cada modelo a las observaciones, a través de varias medidas de bondad de ajuste.

Medidas de bondad de ajuste relativas y absolutas de todos los modelos predictivos.

Tabla 7: CART

mae	mse	rmse	mape	nmse	nmae
2.248843e+05	1.213119e+11	3.482986e+05	7.246647e-01	3.805429e-01	6.493574e-01

Tabla 8: Vecino más próximo

mae	mse	rmse	mape	nmse	nmae
1.292433e+05	5.793759e+10	2.407023e+05	3.031964e-01	1.817442e-01	3.731923e-01

Tabla 9: Máquinas de soporte vectorial(“kernlab”)

mae	mse	rmse	mape	nmse	nmae
1.590317e+05	8.539411e+10	2.922227e+05	4.650482e-01	2.678725e-01	4.592069e-01

Tabla 10: Random forest

mae	mse	rmse	mape	nmse	nmae
7.009516e+04	1.513719e+10	1.230333e+05	1.784410e-01	4.748382e-02	2.024010e-01

Tabla 11: Máquinas de soporte vectorial(“e1071”)

mae	mse	rmse	mape	nmse	nmae
1.261496e+05	5.771840e+10	2.402465e+05	3.343812e-01	1.810567e-01	3.642592e-01

Fuente: elaboración propia

Siendo:

MAE: es la diferencia absoluta entre el valor verdadero y el valor predicho por el modelo.

MSE: es la suma de los cuadrados de la distancia entre el valor predicho y el valor verdadero

RMSE: es el error cuadrático medio, y representa a la raíz cuadrada de la distancia cuadrada promedio entre el valor real y el valor predicho.

MAPE: es el promedio del error absoluto o diferencia entre el valor real y el predicho, expresado como un porcentaje de los valores reales

Se han obtenido varias medidas de bondad de ajuste, tanto relativas como absolutas. Un medidor útil para saber si un modelo es bueno, es el Mape, el cual determina el error porcentual absoluto medio, expresando la exactitud como un porcentaje del error, por lo tanto, contra menor sea el valor Mape, menor será el error de la predicción y mejor será el resultado del modelo.

Teniendo como referencia el Mape, el modelo que presenta mejores resultados es el Random Forest, ya que tiene un Mape de 17,8%, lo que indica que la predicción esta errada en un 17,8%.

Además, como se ha podido observar en el análisis, las características de la vivienda que más influyen en el precio son sobre todo la superficie, y en menor medida el número de habitaciones, si posee ascensor, también el número de baños y si el edificio posee aparcamiento.

5. CONCLUSIÓN

Como se ha visto en el trabajo, la inteligencia de negocios se está convirtiendo en una parte fundamental de las empresas. Cada vez más organizaciones se respaldan en los datos para tomar las decisiones. Por lo tanto, en un mundo tan competitivo donde la diferencia es esencial, la inteligencia de negocios puede marcar la diferencia.

La inteligencia de negocios se puede aplicar en todos los sectores y ámbitos de las empresas. Por lo tanto, la inteligencia de negocios puede llegar a ser un gran apoyo para las empresas, aumentando así su eficiencia, y por lo tanto su competitividad.

El presente trabajo se ha centrado en el mercado inmobiliario, donde se ha tratado uno de los mayores problemas de este mercado, que es el precio.

Por lo tanto, para abordar este problema, se han desarrollado varios modelos que estimen el valor de un inmueble en función de sus características.

El mejor resultado de predicción se ha obtenido a través del algoritmo Random forest, con un Mape del 17,8%, indicando que la predicción de esta técnica esta errada en un 17,8%.

Además, como se ha podido observar en el análisis, las características de la vivienda que más influyen en el precio son sobre todo la superficie, y en menor medida el número de habitaciones, si posee ascensor, también el número de baños y si el edificio posee aparcamiento.

Uno de los mayores problemas a la hora de intentar predecir el valor de las viviendas, ha sido la falta de la variable localización, ya que se trata de un factor incluso más influyente que la superficie. Este ha sido el mayor problema en los resultados obtenidos, ya que, de tener esta variable, los resultados habrían mejorado considerablemente.

Debido a la falta de la localización, como se ha observado en la base de datos, existen bastantes viviendas con un elevado precio y poca superficie. Por lo tanto, como la variable que más influía, con gran diferencia, en la determinación del precio de la vivienda, era la superficie, se producía un mayor error en la predicción.

En un futuro trabajo sería interesante analizar una gran base de datos de una ciudad, que poseyera todas las características de la vivienda, no solo las relacionadas con el tamaño, como en este caso. A pesar de que, junto a la localización, son las variables que más influyen. Sería interesante disponer de variables como: año de construcción, reformada, tipo de vivienda (unifamiliar o plurifamiliar), número de pisos, servicios cercanos, jardín, piscina, etc... Debido a que cada vivienda tiene determinadas características que las diferencia de las demás. Y por lo tanto se obtendría una predicción del precio con un menor error.

ANEXO

ANEXO

OBJETIVOS DE DESARROLLO SOSTENIBLE



Reflexión sobre la relación del TFG con los ODS en general y con el/los ODS más relacionados.

Los objetivos de desarrollo sostenible (ODS) fueron instaurados en 2015 por todos los estados miembros de la ONU, a través de la aprobación de la Agenda 2030 sobre el desarrollo sostenible.

El objetivo de los ODS es poner fin a la pobreza, proteger el planeta y garantizar que todas las personas gocen de paz y prosperidad, entre otros, para el año 2030. Para conseguir estos objetivos, la agenda cuenta con 17 Objetivos de desarrollo sostenible.

Como se ha visto en el trabajo, la inteligencia de negocios puede ser muy útil para la toma de decisiones y el diseño de estrategias, para conseguir los objetivos de la ONU.

El trabajo de fin de grado se podría relacionar con la mayoría de los ODS, pero en este caso, se ha decidido elegir el Objetivo de Desarrollo Sostenible número 13: Acción por el clima, se ha escogido este objetivo, ya que entre los 17 ODS es el que mejor se ajusta al TFG. Además, otro factor que ha influido en su elección es que se trata de un ODS que influye en los otros ODS, y por lo tanto es de gran importancia, ya que, de no conseguir este objetivo, no podrán cumplirse otros objetivos planteados.

El mundo actualmente sufre una situación medioambiental crítica, que está afectando gravemente al clima y alterando los ecosistemas, provocando un aumento de los desastres naturales y propiciando las enfermedades, esto se debe principalmente a la contaminación del ser humano.

Desde la etapa preindustrial hasta la actualidad, la temperatura promedio de la tierra ha aumentado en 1,1 grados Celsius, siendo la década de 2011-2020, la más cálida desde que se tienen registros. Además, se estima que la temperatura promedio aumenta en 0,2 grados Celsius cada década. La ONU al desarrollar el objetivo número 13: Acción por el clima, estableció el objetivo de no superar la temperatura promedio de 2 grados Celsius, que ocasionaría enormes catástrofes medioambientales, y limitar la temperatura promedio a un máximo de 1,5 grados Celsius.

Con el objetivo de reducir e intentar solventar este problema, la mayoría de los países, están aplicando políticas medioambientales para reducir la contaminación, el análisis de datos permitiría analizar cuáles serían las medidas más efectivas para alcanzar el objetivo, así como hacer un seguimiento de estas políticas y comprobar si están consiguiendo el efecto esperado.

El cambio climático es un hecho irreversible, que no se puede cambiar, pero si se puede detener, por lo tanto, es hora de cambiar los hábitos e intentar reducir la contaminación.

BIBLIOGRAFÍA

- BDE (2020). El mercado de la vivienda en España entre 2014 y 2019. Banco de España.
<https://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesSerias/DocumentosOcasionales/20/Fich/do2013.pdf>
(Consulta: 05/02/2021)
- Cano, J. (2007). *Business intelligence: Competir con información*. ESADE
- Cleformación(2014). Tendencias en Business Intelligence.
<https://cleformacion.com/tic-tek/-/blogs/tendencias-en-business-intelligence>
(Consulta: 05/12/2021)
- Deloitte. Impacto del covid-19 en el mercado inmobiliario. *Deloitte*.
<https://www2.deloitte.com/content/dam/Deloitte/ec/Documents/strategy/Impacto-COVID19-mercado-inmobiliario.pdf> .
(Consulta: 14/12/2020)
- Gartner. Business intelligence (BI) Services. Gartner.
<https://www.gartner.com/en/information-technology/glossary/business-intelligence-bi-services>
(Consulta: 22/01/2021)
- Gentleman, R., Hornik, K., Parmigiani, G. (2011) *Business analytics for managers*. Springer
- Hodeghatta, U., Nayak, U.(2017).*Business Analytics Using R - A Practical Approach*. Dr. Umesh R. Hodeghatta and Umesha Nayak

- Housage (2020). Tipos de vivienda: ventajas e inconvenientes de cada una. *Housage*. <https://housage.es/articulos/consejos-comprar-casa-nueva/tipos-viviendas-caracteristicas> .
(Consulta: 25/03/2021)
- Idealista (2020). Evolución del precio de la vivienda en venta en España, Madrid. <https://www.idealista.com/sala-de-prensa/informes-precio-vivienda/>
(Consulta: 20/01/2021)
- INE (2020). Índice de precios de la vivienda. INE. <https://www.ine.es/jaxiT3/Datos.htm?t=6150#!tabs-gráfico>.
(Consulta: 12/02/2021)
- INE (2020). Estadística de transmisión de derechos de la propiedad. INE. https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736152838&menu=ultiDatos&idp=1254735976607.
(Consulta: 23/02/2021)
- Lifeder. Primera generación de computadoras. <https://www.lifeder.com/primera-generacion-computadoras/>
(Consulta: 28/11/2021)
- Martin, L. (25 de marzo, 2019). La gestión de datos empleará a 10 millones de europeos en el 2020. *La Vanguardia*. <https://www.lavanguardia.com/economia/20190325/461210893263/gestion-datos-trabajadores-digitales-big-data-empleo.html>
(Consulta: 15/12/2020)
- Realia. (2015). ¿Qué es el mercado inmobiliario? *Realia*. <https://www.realia.es/que-es-mercado-inmobiliario>
(Consulta: 17/12/2020)

- Rstudio (2009). Rstudio (Version1.4.1103). Rstudio.com
- (Shmueli, G., Bruce, P.,Yahav,I.,Patel,N.,Lichtendahl,K).(2018) *Data mining for business analytics : concepts, techniques, and applications in R* .John Wiley & sons
- Statista (2020). Big data market size revenue forecast worldwide from 2011to2027.
<https://www.statista.com/statistics/254266/global-big-data-market-forecast/>
(Consulta: 05/11/2021)
- Statista (2020). Precio del metro cuadrado de las viviendas de Madrid por distrito. Statista. <https://es.statista.com/estadisticas/578401/precio-del-metro-cuadrado-de-la-vivienda-en-madrid-por-distrito>.
(Consulta: 05/03/2021)
- Watson.H(2008). *Decision Support in the Data Warehouse*. The Data Warehousing Institute Series.