



UNIVERSITAT OBERTA DE CATALUNYA (UOC)
MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS (*Data Science*)

TRABAJO FINAL DE MÁSTER

ÁREA: ÁREA 1. DATA ANALYSIS Y BIG DATA

Análisis de los factores de Gentrificación Gentrificación Turística en Madrid

Autor: José Alonso Ayllón Gutiérrez

Tutor: Anna Muñoz Bolas

Profesor: Albert Solé Ribalta

Entrega: 13 de enero de 2024



Esta obra está sujeta a una licencia de Reconocimiento - NoComercial - SinObraDerivada
3.0 España de Creative Commons.

FICHA DEL TRABAJO FINAL

Título del trabajo:	Análisis de los factores de Gentrificación Gentrificación Turística en Madrid
Nombre del autor:	José Alonso Ayllón Gutiérrez
Nombre del colaborador/a docente:	Anna Muñoz Bolas
Nombre del PRA:	Albert Solé Ribalta
Fecha de entrega (mm/aaaa):	01/2024
Titulación o programa:	Máster Universitario en Ciencia de Datos
Área del Trabajo Final:	Data Analysis y Big Data
Idioma del trabajo:	Español
Palabras clave:	Gentrificación, Turistificación, Geoespacial
Enlace al proyecto:	https://github.com/jayllongu/TFM

Abstract

Gentrification and **Touristification** are two phenomena that are altering large cities. Both processes are modifying urban planning and city life, causing a strong impact on the population of the affected neighborhoods.

Gentrification is the process by which an urban area is transformed into an attractive neighborhood for investment. As a consequence, the expulsion of the original residents occurs, among other factors, due to the increase in housing prices and other costs derived from the gentrified area.

Touristification (a word not yet accepted by the RAE) refers to the impact that tourist massification has on the commercial and social fabric of certain neighborhoods or cities.

Both concepts are attributed, a priori, a negative character in their impact on citizens and traditional neighborhoods, coming to have a rather pejorative meaning.

This work aims to analyze the factors that influence the gentrification of a neighborhood and the possible relationship with the touristification phenomenon. The city of Madrid has been chosen because it is the Spanish city with the greatest movement of tourists and because of the urban transformation it has undergone in recent years.

The first objective is a visualization of the variables that can influence the gentrification of Madrid neighborhoods and try to know the impact of these variables and study the possible influence of tourism on gentrification.

Subsequently, classification algorithms are applied according to the factors of gentrification and tourism to try to predict which urban areas of Madrid may be affected by these phenomena.

Key words: Gentrification, Touristification, Data Science, Visualization, Clustering, Madrid, Master's Final Project

Resumen

Gentrificación y "Turistificación" son dos fenómenos que están alterando las grandes ciudades. Ambos procesos están modificando el urbanismo y la vida de las ciudades, provocando fuerte impacto en la población de los barrios afectados.

Gentrificación es el proceso por el que una zona urbana se transforma en un barrio atractivo para la inversión. Como consecuencia se produce la expulsión de los residentes originales, entre otros factores, por el encarecimiento de la vivienda y otros costes derivados en la zona gentrificada.

Turistificación (palabra aún no aceptada por la RAE) hace referencia al impacto que tiene la masificación turística en el tejido comercial y social de determinados barrios o ciudades.

Ambos conceptos se les atribuye, a priori, un carácter negativo en su afectación a los ciudadanos y los barrios tradicionales, llegando a tener un sentido más bien peyorativo.

Este trabajo pretende un análisis de los factores que influyen en la gentrificación de un barrio y la posible relación con el fenómeno de turistificación. Se ha elegido la ciudad de Madrid por ser la ciudad española con mayor movimiento de turistas y por la transformación urbana que ha sufrido en los últimos años.

El primer objetivo es una visualización de las variables que pueden influir en la gentrificación de los barrios madrileños y tratar de conocer el impacto de estas variables y estudiar la posible influencia de la turistificación sobre la gentrificación.

Posteriormente se aplican algoritmos de clasificación según los factores de gentrificación y turistificación para intentar predecir qué zonas urbanas de Madrid podrán verse afectadas por estos fenómenos.

Palabras clave: Gentrificación, Turistificación, Ciencia de Datos, Visualización, Clustering, Madrid, Trabajo Final de Máster

Índice general

Abstract	III
Resumen	IV
Índice	V
Listado de Figuras	VII
1. Introducción	2
1.1. Descripción general del problema	2
1.2. Motivación	4
1.3. Objetivos	5
1.3.1. Objetivos generales	5
1.3.2. Objetivos específicos	5
1.4. Metodología	6
1.5. Planificación	6
1.6. Medios técnicos	7
2. Estado del Arte	10
2.1. Predicción del riesgo de gentrificación por turistificación. Metodología	13
3. Desarrollo del Trabajo	14
3.1. Fuentes de Datos	15
3.2. Visualizaciones previas	19
3.2.1. Mapas base	20
3.2.2. Datos Socio-Económicos	20
3.3. Implementación	25
3.3.1. Construcción del Conjunto de Datos	26
3.3.2. Procesado del conjunto de datos por Distritos	30
3.3.3. Resultados	35

3.4. Conclusiones y líneas de trabajo futuro	39
Bibliografía	41

Índice de figuras

1.1. Fases de la Gentrificación.	3
1.2. Modelo CRISP-DM. Fuente[1]	6
1.3. Planificación del TFM.	8
1.4. Diagrama de Gantt del TFM.	9
2.1. Diferencias entre gentrificación y turistificación . Fuente[2]	11
2.2. ejemplo de indicadores de vivienda. Fuente[3]	12
2.3. Atlas Turistificación Madrid. Fuente[4]	13
3.1. Entorno QGIS del trabajo	15
3.2. Portal Datos Abiertos Madrid	16
3.3. Visor del Sistema Estatal Índices de Alquiler de Vivienda	17
3.4. Visor de Inside Airbnb, Madrid	18
3.5. ficheros obtenidos de las fuentes de datos	19
3.6. división por distritos	20
3.7. subdivisión por barrios	21
3.8. población por barrios en 2022	21
3.9. población por barrios en 2022	22
3.10. distribución del precio de alquiler por distritos	23
3.11. Proporción de viviendas que no están en buen estado	23
3.12. distribución de colegios y escuelas infantiles	24
3.13. zonas vulnerables de Madrid	24
3.14. defunciones	25
3.15. mapa de detenciones policiales	26
3.16. vista Excel de fichero panel_indicadores_distritos_barrios.csv	28
3.17. movimientos de población	29
3.18. Conjunto de datos para el análisis	31
3.19. <i>BoxPlot</i> de la variable Vivienda_Deficiente para identificar <i>outliers</i>	32

3.20. Correlaciones entre las variables	33
3.21. PCA. Varianza explicada por cada componente	34
3.22. PCA. Contribución de las variables originales en los componentes principales . .	34
3.23. Agrupamiento de las 2 dimensiones principales para k=2	35
3.24. Agrupamiento de las 2 dimensiones principales, k=2 y la variable objetivo	36
3.25. Curva para identificar el número de clústeres	36
3.26. K-means para 2 clústeres	37
3.27. Kmeans para 3 clústeres	37
3.28. K-medoids para 3 clústeres	38
3.29. K-medoids para 2 clústeres	38
3.30. Comparación de agrupamientos para alquiler turístico y calidad de vida	39
3.31. Comparación de agrupamientos para alquiler turístico y percepción de seguridad	40

Capítulo 1

Introducción

1.1. Descripción general del problema

Se puede afirmar que con el siglo XXI comienza una nueva era humana resultado del progreso tecnológico y la penetración de las tecnologías de la información y de las comunicaciones en nuestras vidas, formando parte esencial de nuestro quehacer diario. Es la era de la Globalización[5].

La tecnificación de la sociedad está presente en la forma en que la economía mueve al mundo, llegando los tentáculos de la globalización a todos los aspectos de la sociedad, alcanzado la cultura, la política, el medio ambiente y las relaciones humanas.

Esta globalización está afectando a cómo evolucionan las ciudades en todos sus aspectos, obligando a la transformación del entorno urbano y, por tanto, la forma en cómo vivimos las personas, afectando a la economía doméstica, lo que nos obliga a modificar hábitos y costumbres. La globalización es una realidad que no tiene por qué ser un problema en sí, sino que puede ser una oportunidad de mejora para la calidad de vida de las personas[6].

Una consecuencia de la globalización es el fenómeno de **Gentrificación**[7] que se está produciendo principalmente en las grandes ciudades. La gentrificación se define como el proceso socioeconómico por el que una zona urbana, generalmente de bajo nivel económico y con edificios antiguos o deteriorados, se transforma en un barrio atractivo para la inversión. Como consecuencia se puede producir la expulsión de los residentes originales, que no pueden permitirse los precios más altos de la vivienda y otros costes de vida en el área gentrificada[8].

En el proceso de gentrificación, ilustrado en la figura 1.1, el barrio sufre una mejora significativa desde el punto de vista urbano, con edificios nuevos o restaurados, nuevos comercios, mayor seguridad y limpieza, lo que dota al barrio de un nuevo valor añadido al tiempo que sube el precio del suelo, vivienda, comercios... Sin embargo, esta mejora no suele beneficiar a los residentes originales de la zona, que a menudo se ven obligados a mudarse debido al aumento



Figura 1.1: Fases de la Gentrificación.

del coste de vida en el nuevo barrio. Como consecuencia, puede tener un impacto negativo en el bienestar social y económico, así como en el sentido de pertenencia y comunidad. La gentrificación también puede tener un impacto negativo en la diversidad cultural y étnica de un barrio, y puede dar lugar a la eliminación de la historia y las tradiciones locales.

Si bien la gentrificación tiene muchos aspectos negativos, se puede mirar con una perspectiva positiva. Por ejemplo:

Mejoras en la calidad de vida. A menudo implica una inversión importante en infraestructuras y servicios, como mejoras en calles, aceras, parques, transporte público, tiendas y restaurantes.

Reducción del crimen. Puede atraer a mayor inversión en seguridad, lo que puede disminuir la delincuencia en el barrio.

Creación de empleo. Una consecuencia es la creación de nuevos empleos en la zona, particularmente en el sector servicios.

Mayor diversidad. Aunque la gentrificación a menudo resulta en la expulsión de residentes originales, también puede llevar a una mayor diversidad en la zona, debido a que atrae a nuevos residentes de diferentes orígenes, culturas y estilos de vida.

Mejora en la educación. puede atraer inversión en la construcción de nuevos colegios o la mejora de los existentes.

Aumento del valor de la propiedad. El aumento del valor de la propiedad parece ser algo inherente al proceso de gentrificación, lo que puede resultar beneficioso para los propietarios de inmuebles y los inversores inmobiliarios.

Es importante tener en cuenta que estos aspectos positivos no compensan necesariamente los efectos negativos de la gentrificación, y que se genera un debate sobre si estos beneficios son suficientes para justificar la expulsión de los residentes originales y la eliminación de la historia y la cultura locales.

Otro aspecto a tener en cuenta es el fenómeno de **Turistificación**, palabra aún no aceptada por la RAE, pero que sí está incluida en el diccionario Fundéu (Fundación patrocinada por la Agencia Efe y el BBVA, y asesorada por la RAE) y “alude al impacto que tiene la masificación turística en el tejido comercial y social de determinados barrios o ciudades.” Turistificación es un fenómeno que parece estar influyendo en la gentrificación de algunas ciudades[9].

Con este trabajo se intenta un análisis de los factores que influyen en la gentrificación de un barrio y la posible relación con el fenómeno de turistificación. Se ha elegido la ciudad de Madrid por ser la ciudad española con mayor movimiento de turistas y por la transformación urbana que ha sufrido en los últimos años.

Conocer cómo es el proceso de gentrificación, qué variables influyen y en qué medida ejercen esta influencia es de vital importancia para poder disponer de conocimiento que advierta de cómo puede ser la evolución de un barrio o una ciudad. La planificación urbana de las ciudades debe hacerse para logra la mejora de los ciudadanos, de manera sostenible pero al mismo tiempo sin que obligue a transformar a sus ciudadanos.

La Ciencia de Datos proporciona conocimientos y herramientas útiles para ayudar a mejorar la calidad de vida en las ciudades. Con este trabajo se pretende ayudar en el conocimiento de los fenómenos de gentrificación y turistificación.

1.2. Motivación

Tengo la suerte de vivir en un pueblo en el entorno de la Costa del Sol, Istán, entrada al Parque Nacional Sierra de las Nieves mirando al norte y tiene vistas al Mediterráneo mirando al Sur. Es un pueblo típico andaluz cargado de cultura y tradiciones, con una vida apacible lejos del mundanal ruido, y dotado con los servicios necesarios para una vida placentera. En mi pueblo aún no se sufren los problemas de las grandes urbes, sin embargo, la proximidad a ciudades como Marbella y Málaga, pueden alterar esta normalidad por fenómenos de gentrificación y turistificación. Esta inquietud, junto con el hecho de que hace años trabajo con sistemas de información geográfica en el ayuntamiento de Marbella, me ha motivado para realizar este trabajo final de mis estudios de Máster en Ciencia de Datos. Los estudios de máster me han aportado conocimiento de como se pueden abordar ciertos problemas humanos no sólo con herramientas matemáticas y técnicas, sino también con la aplicación del método científico.

La aplicación de métodos propios de la ciencia de datos para adquirir nueva información e

incluso nuevo conocimiento sobre fenómenos que están afectando a las ciudades puede ser de utilidad para su posterior aplicación a casos concretos que permitan una mejor planificación de las ciudades de cara a la mejora de la calidad de vida de las personas que las habitan.

Este Trabajo de Final de Máster es toda una oportunidad que no debo dejar escapara para realizar un estudio útil no sólo para obtener el título de Máster en Ciencia de Datos, sino también para que pueda mejorar mis habilidades en el puesto de trabajo que actualmente desempeño y pueda servir como aporte positivo a la sociedad en que vivimos.

1.3. Objetivos

El objetivo principal de este Trabajo de Final de Máster (TFM) es aplicar los métodos y técnicas estudiados para realizar un estudio sobre los factores que pueden influir en los fenómenos de Gentrificación y Turistificación en la ciudad de Madrid.

1.3.1. Objetivos generales

Los objetivos generales del TFM son los definidos en la propia asignatura TFM de los estudios de Máster en Ciencia de datos de la UOC dentro del Área 1: Data Analysis y Big Data:

La propuesta de este trabajo final de máster consiste en la visualización interactiva de los datos económicos de los barrios de una gran ciudad, como Madrid, para evaluar su evolución en los últimos años, analizar los factores que caracterizan los barrios con alto índice de gentrificación y detectar los barrios en riesgo de gentrificación a corto plazo.

También se pretende abordar el reto de realizar un estudio del impacto del fenómeno de gentrificación turística en la sostenibilidad de los barrios, y tratar de abordar el problema desde una perspectiva de género, con el objetivo de determinar si los factores de tipo género influyen en el problema.

1.3.2. Objetivos específicos

Los objetivos específicos del TFM son los propios de un proyecto de Minería de Datos:

1. Plantear el problema e identificación de fuentes de datos.
2. Procesos ETL sobre los datos
3. Visualización y análisis de estadísticos básicos
4. Implementación de los datos en un sistema SIG.

5. Aplicar técnicas de Clasificación y Agrupación.
6. Extracción de los factores que influyen en los procesos de gentrificación y turistificación.
7. Aplicar técnicas de predicción para identificar las zonas urbanas de Madrid que puedan ser afectadas a corto plazo.

1.4. Metodología

Una vez seleccionada la temática del trabajo, se realiza una serie de búsquedas en Internet para obtener información sobre qué es la gentrificación y los estudios que sobre este tema se pueden encontrar. También se hace una búsqueda dentro del repositorio de publicaciones de la UOC de trabajos anteriores sobre esta temática. Esto me permite familiarizarme con los conceptos que irán apareciendo a lo largo del trabajo.

Para el desarrollo del TFM se aplica la metodología propia de un proyecto de Minería de Datos siguiendo el modelo **CRISP** (Cross Industry Standard Process), cuyas fases son las que se muestran en la figura 1.2.

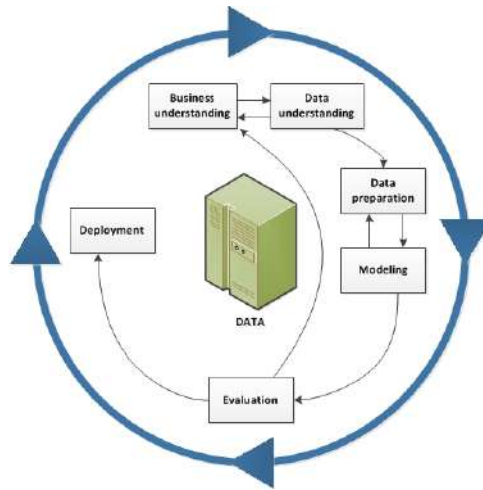


Figura 1.2: Modelo CRISP-DM. Fuente[1]

1.5. Planificación

El desarrollo del Trabajo Final de Máster en Ciencia de Datos se realiza conforme a la planificación del curso, que consta de 5 Pruebas de Evaluación Continuada (PECs) y la preparación y defensa pública del trabajo. Las pruebas de evaluación continuada se concretan en cinco tareas:

M1 - Definición y planificación del trabajo final. Esta primera fase tiene como principal objetivo definir claramente cuál es la temática del trabajo, justificar su interés y/o relevancia, qué se quiere conseguir al finalizar el TFM, definir los objetivos principales y una planificación temporal del proyecto. Además, incluye la motivación personal para hacer este trabajo final.

M2 - Estado del arte. En esta tarea se completará la parte del estado del arte de la memoria, en donde:

1. Explicará qué investigación (o soluciones) se han desarrollado hasta el momento para resolver la problemática en la que se trabajará. Esto es lo que se denomina “Estado del arte” o “Estado de la cuestión”.
 2. Este estudio puede inducir a la ampliación o modificación de la explicación de la temática escogida y de la problemática a resolver, ya que se pueden detectar matices que se desconocían en el momento de escribir la actividad anterior. En este caso no hay problema, siempre es posible corregir o modificar algunos detalles realizados anteriormente.
- M3 - Implementación del trabajo. En esta fase se realizan las tareas definidas en la planificación del trabajo (M1 y M2). Esta etapa puede tener varias entregas parciales antes de la entrega definitiva para garantizar que el trabajo se está haciendo correctamente.
 - M4.1 - Redacción de la memoria (1a Entrega). El objetivo de esta actividad es tener un borrador completo de la memoria del TFM.
 - M4.2 - Redacción de la memoria (Entrega Final). Entrega final de la memoria del TFM.
 - M4.3 - Presentación audiovisual del trabajo.
 - M5.1 - Entrega de la documentación al tribunal.
 - M5.2 - Defensa pública del Trabajo Final de Máster.

1.6. Medios técnicos

Se dispone de los siguientes medios técnicos para la realización del Trabajo:

Hardware:

- PC con procesador Intel Core i7, 16 GB RAM, MS Windows 11

Software:

- R 4.2.0 (2022-04-22 ucrt), entorno de programación en lenguaje R

- RStudio 2023.09.01 Build 494, IDE desarrollo lenguaje R.
- QGis 3.34.1-Prizren, software GIS
- ChatGPT 3.5, plataforma chatbot útil para la ayuda en la generación de código
- MS Office 365, herramienta ofimática
- Overleaf, editor LaTeX online
- GitHub, plataforma de desarrollo colaborativo

Nombre de la tarea	Rango de fechas	Duración
▼ Planificación del trabajo		
✓ M1. Definición	27 sep – Hoy	10 días
✓ M2 Estado del arte	Hoy – 24 oct	11 días
✓ M3 Implementación	24 oct – 19 dic	41 días
✓ M4.1 Borrador Memoria	19 dic de 2023 – 2 ene de 2024	11 días
✓ M4.2 Memoria	2 ene de 2024 – 9 ene de 2024	6 días
✓ M4.3 Presentación	9 ene de 2024 – 16 ene de 2024	6 días
✓ M5.1 Entrega documentación	18 ene de 2024	1 día
✓ M5.2 Defensa Pública	4 feb de 2024	

Figura 1.3: Planificación del TFM.



Figura 1.4: Diagrama de Gantt del TFM.

Capítulo 2

Estado del Arte

El objetivo principal de este trabajo es determinar los factores (variables) que influyen en el proceso de gentrificación, incorporando aquellas variables que adquieren relevancia por los procesos de turistificación.

En la definición clásica de gentrificación, este proceso implica un cambio en la población por el desplazamiento debido a la revalorización de la zona lo que lleva al encarecimiento de la vida, por lo asienta nuevos vecinos y negocios con mayor poder adquisitivo.

Entre los factores que promueven la gentrificación están las mejoras en infraestructuras y servicios públicos como la seguridad ciudadana promovidos por las administraciones, principalmente la local, ayudada por inversiones provenientes de organismos supramunicipales en muchos casos.

Además de los factores tradicionales de la gentrificación, en los últimos tiempos parece que se están incorporando nuevas variables por fenómenos de turistificación.

La gentrificación es un fenómeno que puede afectar a cualquier ciudad en cualquier país, tal y como puede verificarse en los diferentes estudios llevados a cabo en ciudades de todo el mundo[10], encontrándose investigaciones que comparan y contrastan los patrones de gentrificación en diferentes contextos.

La puesta en el mercado de "pisos turísticos" de manera masiva, fomentado por plataformas online que facilitan esta penetración, junto con el alto rendimiento de este tipo de inversión, puede estar modificando el concepto de gentrificación, donde ya no son los habitantes desplazados por nuevos vecinos de mayor poder adquisitivo, sino que son desplazados por visitantes de corta estancia en el barrio[11].

El resultado de la gentrificación por turistificación se podría interpretar como el mismo, es decir, el desplazamiento de los habitantes de un barrio que se ven obligados por las nuevas circunstancias socio-económicas. Sin embargo, la modificación que sufre el barrio no es la misma en un proceso de gentrificación con o sin turistificación.

Las infraestructuras, los negocios, la actividad social, el paisaje del barrio, en definitiva el barrio se puede transformar de distinta forma, atendiendo a las demandas y necesidades de los nuevos vecinos, múltiples turistas que habitan en el barrio por cortos periodos de tiempo.

Aunque aún existe el debate si Gentrificación y Turistificación son conceptos similares que describen un mismo problema, hay estudios que abordan el problema desde el punto de vista sólo de la gentrificación cuando no hay efectos relevantes del turismo, y en caso de que sí exista este impacto del turismo se habla de gentrificación turística[3]. Sin embargo, se pueden apreciar diferencias significativas entre ambos conceptos con consecuencias muy diferentes tras el resultado de ambos fenómenos (figura 2.1), lo que incita a afrontar de forma diferente los estudios de predicción sobre el riesgo de alteración de un barrio según el proceso que pueda darse en el mismo. Así, se puede afirmar que Gentrificación y Turistificación son fenómenos urbanos con un mismo origen pero con resultados diferentes[2].

	Gentrificación	Turistificación
Población	Desplazamiento y sustitución	Flotante y posible desplazamiento
Usos del suelo	Residencial e infraestructura cultural	Turísticos y patrimonial
Comercio	Gastronomía, galerías y ocio	Hoteles, gastronomía y souvenirs
Vivienda	Residencial (larga estancia)	Vacacional (corta estancia)

Figura 2.1: Diferencias entre gentrificación y turistificación . Fuente[2]

Realizar estudios predictivos sobre los fenómenos de gentrificación y turistificación son necesarios como ayuda en la planificación urbana de las ciudades. De producirse estos fenómenos, podrían producirse consecuencias importantes sobre las personas incrementando las desigualdades sociales[12].

Entre las causas de la gentrificación, según puede observarse en la mayoría de publicaciones hasta la fecha, se incluyen principalmente, la demanda de vivienda, la especulación inmobiliaria, la inversión pública en infraestructuras y las políticas de desarrollo urbano.

Aunque en algunos casos los fenómenos de gentrificación pueden resultar en beneficios para las zonas afectadas y sus habitantes, en muchos otros casos se desencadenan consecuencias negativas sobre la población desplazada, y son necesarias medidas que mitiguen estos efectos negativos mediante la aplicación de políticas de vivienda asequible, ayudas a los inquilinos y acciones para preservar la identidad cultural de las comunidades afectadas.

La metodología que se ha observado en los estudios sobre fenómenos de gentrificación o turistificación se basan, fundamentalmente, en análisis de datos y revisiones históricas. Los

datos objeto de análisis son, principalmente, los relativos a movimientos demográficos, actividades comerciales y mercado inmobiliario, más datos geográficos con los que se consigue una visualización del problema.

Las infraestructuras, los negocios, la actividad social, el paisaje del barrio, en definitiva el barrio se puede transformar de distinta forma, atendiendo a las demandas y necesidades de los nuevos vecinos, múltiples turistas que habitan en el barrio por cortos periodos de tiempo.

En la primera parte de este TFM se pretende abordar el problema de identificar los factores sociales y económicos que afectan a los procesos de gentrificación promovidos por fenómenos de turistificación. Por tanto, es necesario conocer cuáles son los indicadores apropiados para este análisis, como por ejemplo, indicadores de vivienda[3].

Fenómeno	Indicador
Incremento del peso de las viviendas secundarias	Incremento %viviendas secundarias (2001-2011)
Infrautilización del parque de viviendas	Incremento %viviendas vacías (2001-2011)
Peso del inquilinato	Incremento %viviendas en alquiler sobre las viviendas principales (2001-2011)
Turistificación de la vivienda	% alojamientos turísticos respecto del total de alojamientos.
Centralización de la vivienda turística	% de viviendas turísticas en el centro histórico sobre el total de la vivienda
Rent Gap turístico	Diferencia entre el precio medio del alquiler convencional y el precio medio de la vivienda turística

Figura 2.2: ejemplo de indicadores de vivienda. Fuente[3]

Entre las fuentes consultadas, ha resultado interesante el trabajo publicado en Kaggle, [Housing Crisis and Tourism in Madrid](#), donde el autor pretende evaluar el impacto que tiene Airbnb en la vivienda en Madrid, obtener información sobre quién, cuándo y cómo alquila en Madrid y hacer consideraciones sobre los flujos turísticos en Madrid. El autor concluye que *“Aunque es bastante normal que cada ciudad tenga una mayor concentración de alojamientos en el centro, el centro de Madrid parece tener un serio problema de gentrificación y especulación con una asombrosa densidad de (aproximadamente) 1750 alojamientos por km2.”*

Otro dato curioso que aporta este estudio es que las ofertas de alquiler vacacional que están en inglés tienen en promedio precios más altos por noche, lo que sugiere que a los turistas extranjeros se les puede cobrar un poco más.

2.1. Predicción del riesgo de gentrificación por turistificación. Metodología

Se utilizará una metodología mixta combinando métodos cuantitativos y cualitativos sobre diferentes ámbitos de análisis (ciudad y barrio). Los datos principales serán los relativos a datos de censos y padrones de habitantes, de forma a similar a la del estudio de **La gentrificación desde una mirada de género: un ejemplo en Barcelona**[13]

Para el trabajo en curso, se intenta utilizar herramientas de visualización geográfica que permitan, de forma visual e intuitiva observar en qué barrios de Madrid se pueden estar produciendo fenómenos de gentrificación por turistificación, de forma similar a la visualización del **Atlas de Turistificación de Madrid**[4].

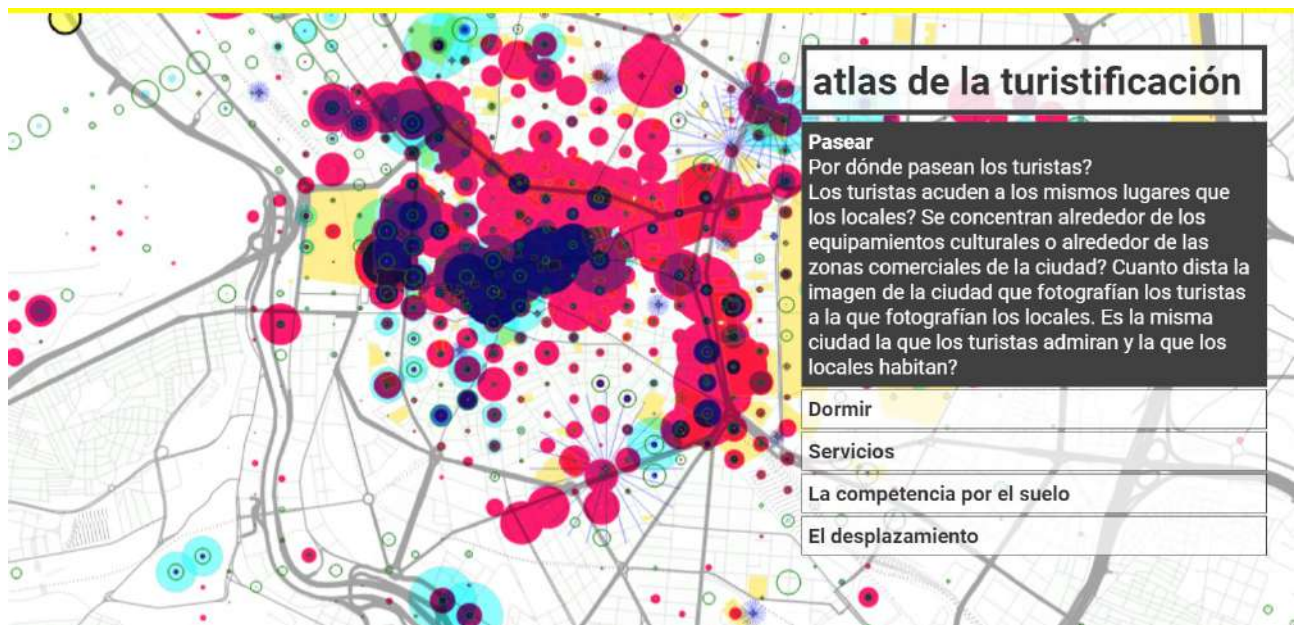


Figura 2.3: Atlas Turistificación Madrid. Fuente[4]

Además de las herramientas de visualización geográfica, se pretende realizar la identificación de los factores (variables) que influyen en la predicción de la probabilidad de gentrificación por turistificación, y conocer en que medida influyen.

Para ello, será necesario implementar técnicas de agrupamiento y clasificación. La aplicación de los algoritmos necesarios precisará de un análisis de los estadísticos básicos de las variables que se consideren adecuadas.

Con las variables adecuadas, se pretende conocer cómo se agrupan los barrios de Madrid en función del grado de gentrificación. Se considera apropiado emplear algoritmos no supervisados, puesto que no se conoce, a priori, si existe una clasificación de las áreas urbanas de Madrid.

Capítulo 3

Desarrollo del Trabajo

Un estudio en el ámbito de la Ciencia de Datos comienza con el planteamiento de un problema, que no es otra cosa sino plantear una serie de cuestiones a las que se quiere dar respuesta usando metodología científica apoyada en la aplicación de técnicas analíticas y algoritmos adecuados al problema en cuestión.

En este trabajo se pretende estudiar cuáles son los factores que influyen en los procesos de Gentrificación por Turistificación en los barrios de Madrid, es decir si los cambios de residencia de vecinos de Madrid están influenciados por aspectos relativos al fomento de una economía basada en el turismo. Se parte del supuesto que efectivamente en Madrid está ocurriendo un movimiento de residentes de unos barrios a otros. Para ver que esto realmente es así, he hecho uso del software QGIS para obtener visualizaciones de estos movimientos de personas y comprobar cómo se están produciendo. En la figura 3.1 se puede ver el entorno completo de QGIS utilizado para el proyecto, incluyendo una visualización temporal de la evolución de la población de Madrid desde 1996 hasta 2020, y las diferentes capas empleadas para las visualizaciones.

Una vez planteado el problema, es necesario disponer de los datos adecuados sobre los que aplicar las técnicas que permitan el tratamiento para obtener información. Los datos utilizados en este trabajo han servido para obtener visualizaciones que ayuden a entender la realidad de los movimientos de personas entre los barrios de Madrid, y realizar un estudio analítico mediante la aplicación de algoritmos con métodos no supervisados.

A partir de las fuentes de datos, se observa que hay indicadores a nivel de distritos y otros más detallados a nivel de barrios. Esto me ha llevado a plantear la cuestión de si tratando los datos en ambos niveles, distritos y barrios por separado, se obtienen resultados similares o, al menos, es posible lograr conclusiones con mayor nivel de precisión sobre los movimientos de personas en el territorio de Madrid.

Por ello, se pretende construir dos conjuntos de datos con el mayor número de indicadores

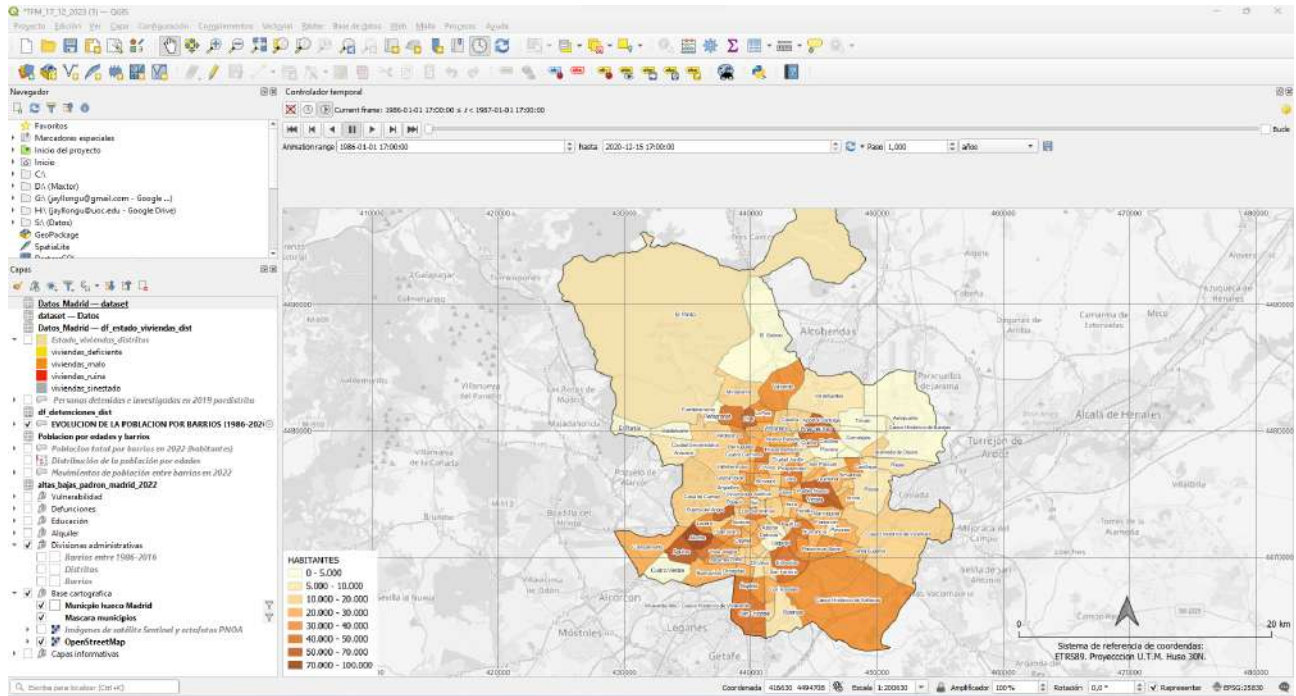


Figura 3.1: Entorno QGIS del trabajo

posible, uno con los datos de los distritos y otro a nivel de barrios.

Finalmente, con la información obtenida, se ha realizado un análisis de los resultados, que se han concretado en las conclusiones de este trabajo.

3.1. Fuentes de Datos

La búsqueda de fuentes de datos adecuadas al estudio de este TFM ha sido la labor más complicada, puesto que se ha dedicado mucho tiempo en realizar numerosas búsquedas en Internet para localizar ficheros con información que podría ser utilizada. Aunque a priori aparecieron muchos ficheros de fuentes de datos abiertos candidatos a ser utilizados en el estudio, al entrar en el detalle de los mismos, se observa que no tienen la calidad que se espera o que se trata más bien de resúmenes de otras fuentes no disponibles o que no se han podido localizar. De las fuentes de datos que sí han resultado útiles, se ha tenido que realizar un tratamiento de los ficheros para adecuarlos al objeto del trabajo.

Para el trabajo, es necesario disponer de fuentes de datos con referencias a los barrios de Madrid. Una dificultad encontrada es que algunas fuentes hacen referencia de los datos sólo por distritos y otras lo hacen por barrios, por lo que no siempre se ha dispuesto del detalle suficiente para abordar el análisis a nivel de barrios y se ha tenido que resolver a nivel superior



Figura 3.2: Portal Datos Abiertos Madrid

de distritos. Se ha recurrido a las siguientes **Fuentes de datos**:

- **Portal datos abiertos de Madrid**

Esta es la principal fuente de datos para este trabajo. El Portal de Datos Abiertos (figura 3.2) es el espacio que el Ayuntamiento de Madrid ofrece para facilitar el acceso a los datos públicos en formato reutilizable relativos a la gestión pública municipal. El Catálogo de datos contiene el listado de todos los conjuntos de datos (datasets) publicados y una canal de API (Application Programming Interface) con operaciones para facilitar el acceso directo a los datos.

- **Sistema Estatal Índices Alquiler de Vivienda**

El **Ministerio de Transportes y Movilidad Sostenible** ofrece un visor geográfico con información de los precios de alquiler de viviendas (figura 3.3). Además dispone para descarga pública de los ficheros con los datos de alquiler con las referencias geográficas. Para este trabajo se ha descargado la capa con los datos agrupados por distritos de España en formato *shape*. El visor del Sistema Estatal Índices de Alquiler de Viviendas es el resultado



Figura 3.3: Visor del Sistema Estatal Índices de Alquiler de Vivienda

de la explotación de fuentes tributarias de los datos sobre arrendamientos de vivienda habitual.

- **Inside Airbnb**

Inside Airbnb es un proyecto impulsado por una comunidad independiente que utiliza datos extraídos de la plataforma de alquileres vacacionales **Airbnb** para analizar y visualizar el impacto de esta plataforma en el mercado de viviendas y la gentrificación en las ciudades más importantes. El proyecto fue creado por Murray Cox con el objetivo de proporcionar información detallada sobre el uso de Airbnb y cómo podría afectar a la disponibilidad de viviendas locales.

Los datos obtenidos por Inside Airbnb contienen información sobre las listas de propiedades en Airbnb, incluyendo la ubicación, la disponibilidad, las tarifas y otros detalles. También incorpora un visor web (figura 3.4) para la representación de manera visual y accesible para que investigadores, periodistas y el público en general puedan explorar y analizar el impacto de Airbnb en los barrios y el mercado inmobiliario local.

- **Housing Crisis and Tourism in Madrid**

Trabajo publicado en **Kaggle**, que estudia el problema de gentrificación en la ciudad de Madrid a partir de los datos de alquileres turísticos con datos de Inside Airbnb.

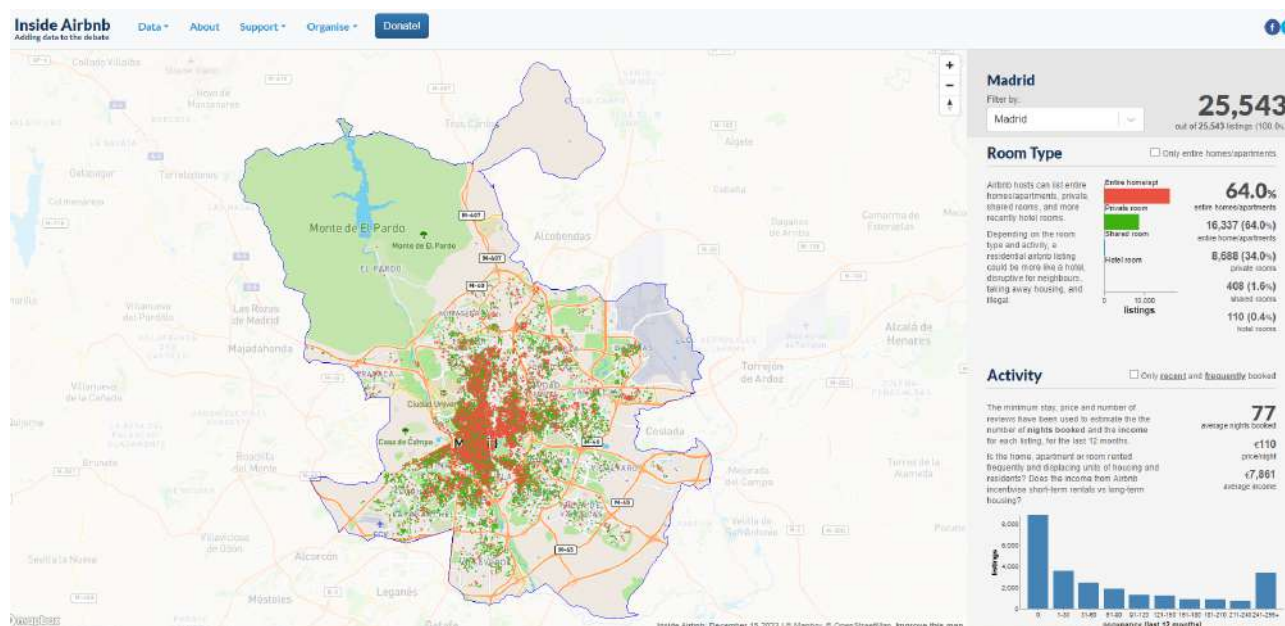


Figura 3.4: Visor de Inside Airbnb, Madrid

3.1.0.1. Listado de ficheros descargados

Se enumeran los ficheros (figura 3.5) que se han descargado de las fuentes de datos y se han utilizado en el desarrollo del trabajo.

- 01_ArboladoZonasVerdesDistritosCalles_2022.xlsx
- 201747-0-bibliobuses-bibliotecas.csv
- 20220101_defunciones.zip
- 202311-0-colegios-publicos.geo
- 202318-0-escuelas-infantiles.geo
- airbnb_listings.csv
- altas_bajas_padron_madrid_2022.xlsx
- AREAS-INFANTILES_2022.csv
- Barrios.kml
- catalogo.csv
- Distritos.zip

Nombre fichero	Tipo	Descripción
01_ArboladoZonasVerdesDistritosCalles_2022.xlsx	xlsx	ubicación arbolado en los doistritos
201747-0-bibliobuses-bibliotecas.csv	csv	ubicaciones bibliotecas
20220101_defunciones.zip	zip	datos defunciones por distritos
202311-0-colegios-publicos.geo	geo	ubicaciones colegios
202318-0-escuelas-infantiles.geo	geo	ubicaciones escuelas infantiles
airbnb_listings.csv	csv	listado de alquileres turísticos Inside Airbnb
altas_bajas_padron_madrid_2022.xlsx	xlsx	movimientos población entre barrios
AREAS_INFANTILES_2022.csv	csv	ubiaciones zonas infantiles
Barrios.kml	kml	limites geográficos de barrios
catalogo.csv	csv	catálogo del Portal de Datos Abiertos de Madrid
Distritos.zip	zip	limites geográficos de distritos. Shapefile
Informe_PMH 2021.doc	doc	Informe padrón Municipal Habitantes de Madrid
panel_indicadores_distritos_barrios.csv	csv	Panel de Indicadores de distritos y barrios
poblacion por edades y barrios.xlsx	xlsx	distribución de la población de Madrid
Ranking_Vulnerabilidad_Barrios_original.csv	csv	rankin de Vulnerabilidad por barrios (IVR)
USOS_DE_SUELO.zip	zip	ordenamiento urbano. Shapefile

Figura 3.5: ficheros obtenidos de las fuentes de datos

- Informe_PMH 2021.doc
- panel_indicadores_distritos_barrios.csv
- poblacion por edades y barrios.xlsx
- Ranking_Vulnerabilidad_Barrios_original.csv
- USOS_DE_SUELO.zip

3.2. Visualizaciones previas

Este trabajo tiene una carga importante de datos geoespaciales, y ha sido interesante la visualización sobre un mapa de la ciudad de Madrid a partir de los datos disponibles.

La herramienta utilizada para este propósito es **QGIS**. Se ha cargado un mapa base (**OpenStreetMap**), los límites geográficos de los distritos y barrios de Madrid, y se han creado una serie de capas espaciales mediante el uso de técnicas como mapas de calor o densidad de puntos, para conseguir una visualización de información.

Se ha generado un proyecto que contiene los ficheros necesarios para reproducir todas la visualizaciones. El sistema de referencia utilizado es ETRS89/UTM zone 30N en coordenadas geográficas con código EPSG 25830 (European Petroleum Survey Group).



Figura 3.6: división por distritos

3.2.1. Mapas base

La ciudad de Madrid, administrativamente, está dividida en distritos, y estos a su vez están subdivididos en barrios, tal y como muestran las figuras 3.6 y 3.7. Estos mapas constituyen la base de las visualizaciones geográficas, junto con el mapa base OpenStreetMap.

3.2.2. Datos Socio-Económicos

El primer dato de interés es conocer cómo está distribuida la población de Madrid. El mapa representado en la figura 3.8 muestra la distribución de la población de los barrios de Madrid incluyendo gráficas por grupos de edad: menores de 16 años, entre 16 y 64 (población activa) y mayores de 65, mostrando de forma clara que el grupo predominante es el de población en edad de trabajar. El mapa de la figura 3.9 muestra esta información para la subdivisión del territorio en distritos, mediante un diagrama de puntos, lo que obliga a hacer zoom sobre la imagen para visualizar los puntos de los grupos de edad de la población no activa.

En el proyecto QGIS se puede hacer zoom mostrando gráficas de valores de población por tramos de edad en cada barrio.

Los datos revelan que las zonas más pobladas corresponden a los barrios de la periferia del sur de Madrid, donde además, las gráficas de edades muestran que hay una mayor proporción de personas en edad de trabajar frente a menores o personas en edad de jubilación.



Figura 3.7: subdivisión por barrios

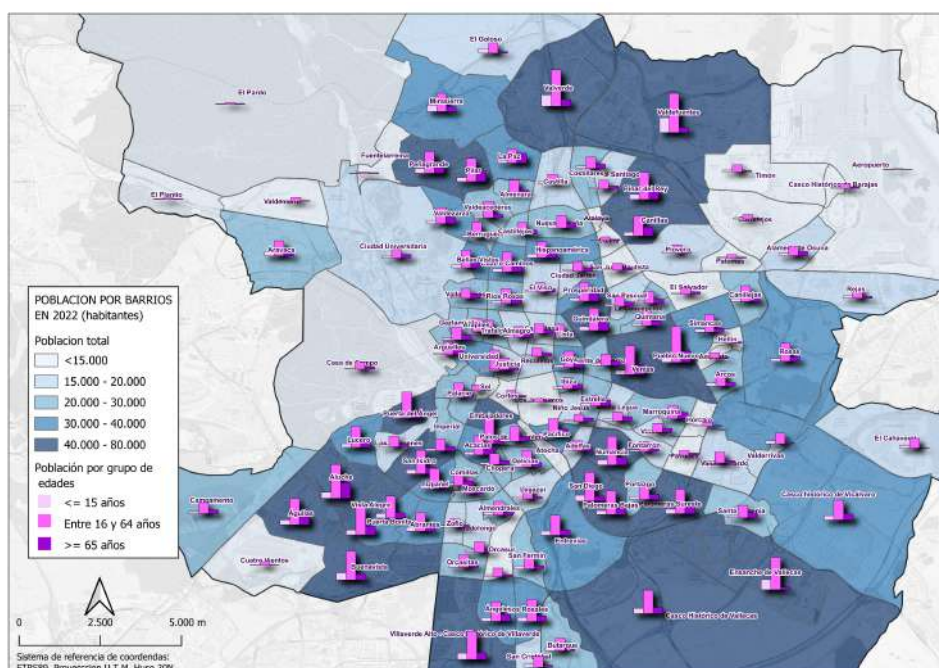


Figura 3.8: población por barrios en 2022

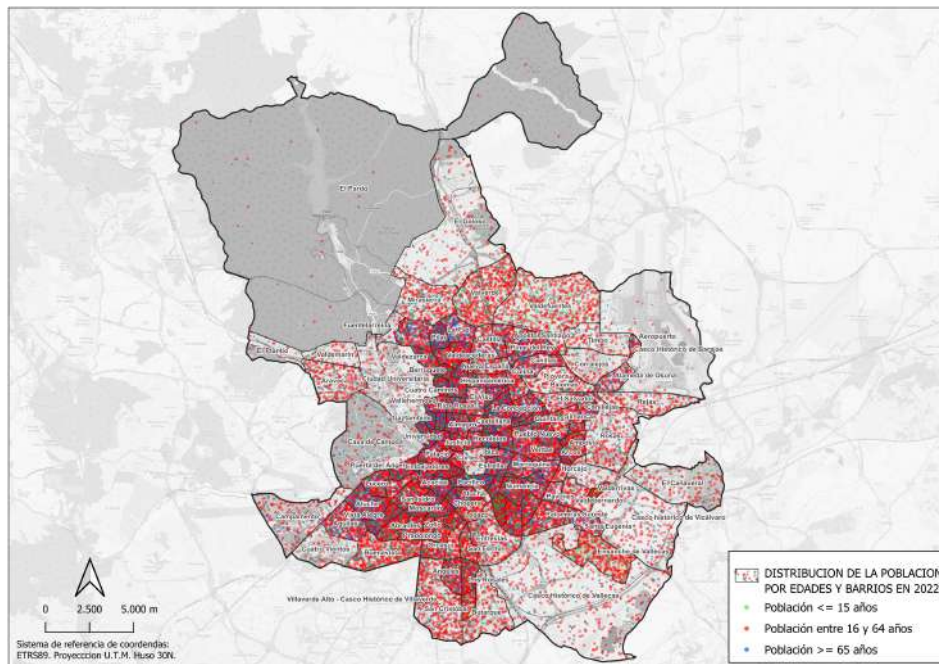


Figura 3.9: población por barrios en 2022

Uno de los indicadores socio-económicos más relevantes es el precio de la vivienda en alquiler. La figura 3.10 representa cómo está distribuido este parámetro en los distritos de Madrid, y se puede observar cómo las rentas más altas corresponden con los distritos centrales y zona norte.

Junto con el precio de la vivienda, es importante realizar alguna comparación con el estado de las viviendas. La gráfica 3.11 muestra la proporción por distritos de viviendas que no están en buen estado.

Otro parámetro importante es la distribución de colegios y escuelas infantiles, que da una idea por un lado de dónde están las zonas residenciales, y por otro la necesidad de desplazamiento. La figura 3.12 muestra un mapa de calor con esta distribución, donde se puede observar la escasa presencia de escuelas infantiles y colegios en la zona norte, mientras que esta concentración es mayor en la zona centro y sur-este de Madrid.

Un indicador muy apropiado para este trabajo es el Índice de Vulnerabilidad Territorial de los distritos y barrios de Madrid (IVR), que establece un ranking de vulnerabilidad por distritos y barrios de Madrid. Se trata de información sobre los riesgos de exclusión y la aparición de zonas especialmente vulnerables. La figura 3.13 muestra cómo son los distritos del sur de Madrid los más vulnerables frente a los riesgos identificados por el IVR.

La figura 3.14 muestra la distribución de defunciones en los distritos de Madrid durante 2022. Este mapa muestra como la concentración de defunciones es mayor en la zona centro y

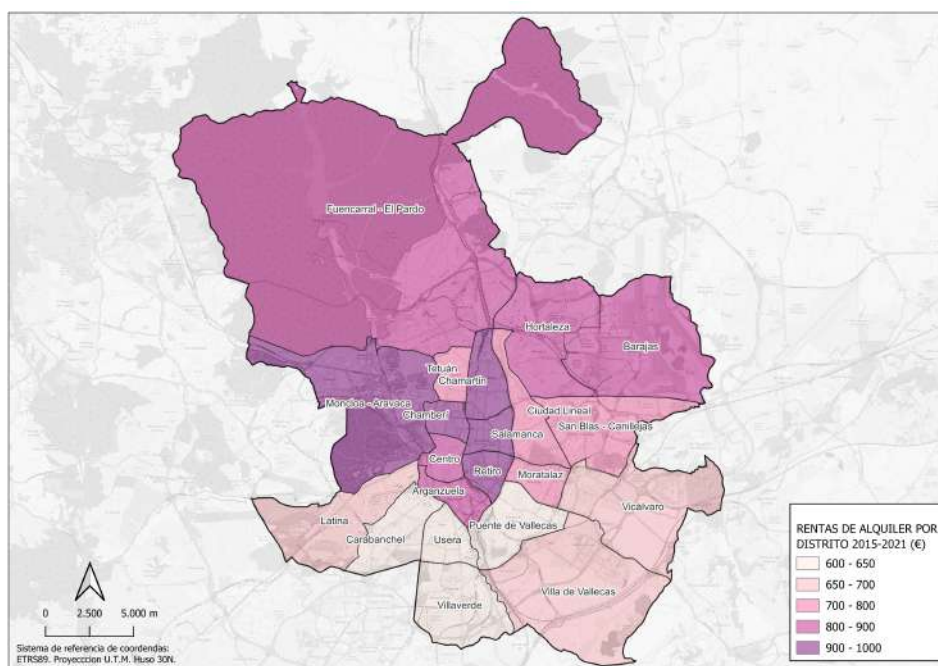


Figura 3.10: distribución del precio de alquiler por distritos

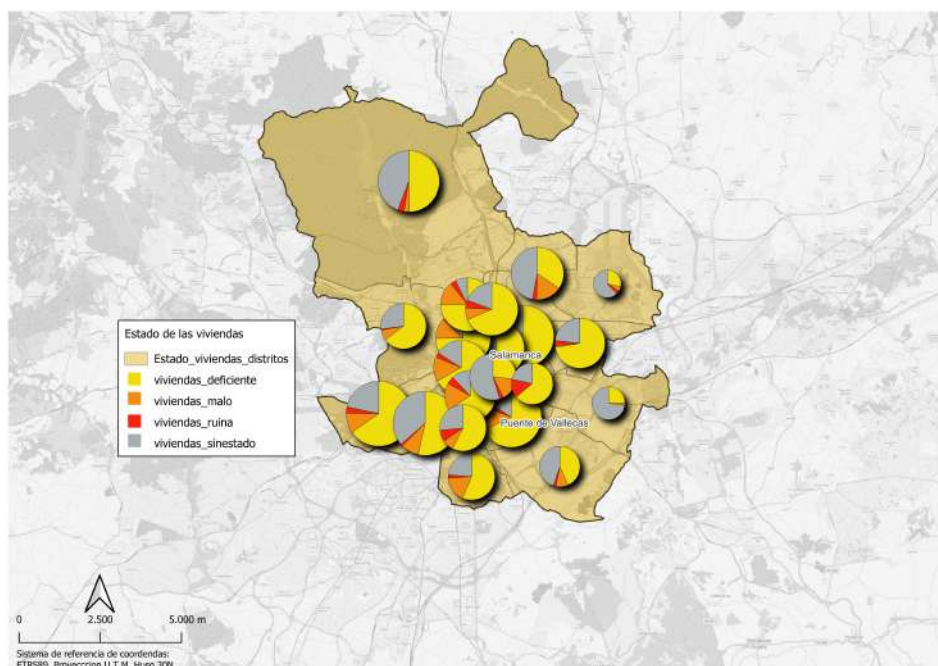


Figura 3.11: Proporción de viviendas que no están en buen estado

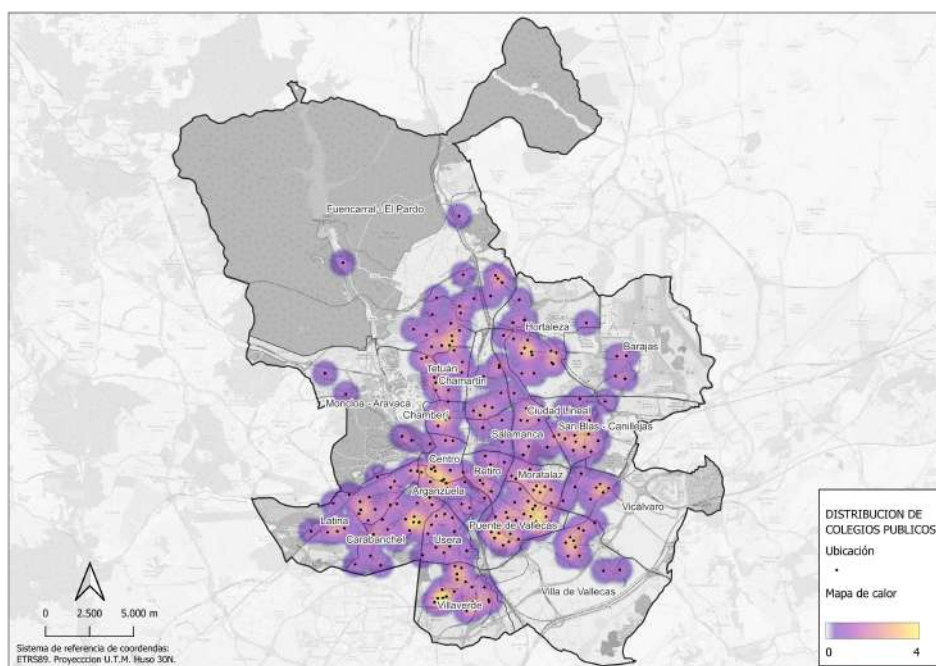


Figura 3.12: distribución de colegios y escuelas infantiles

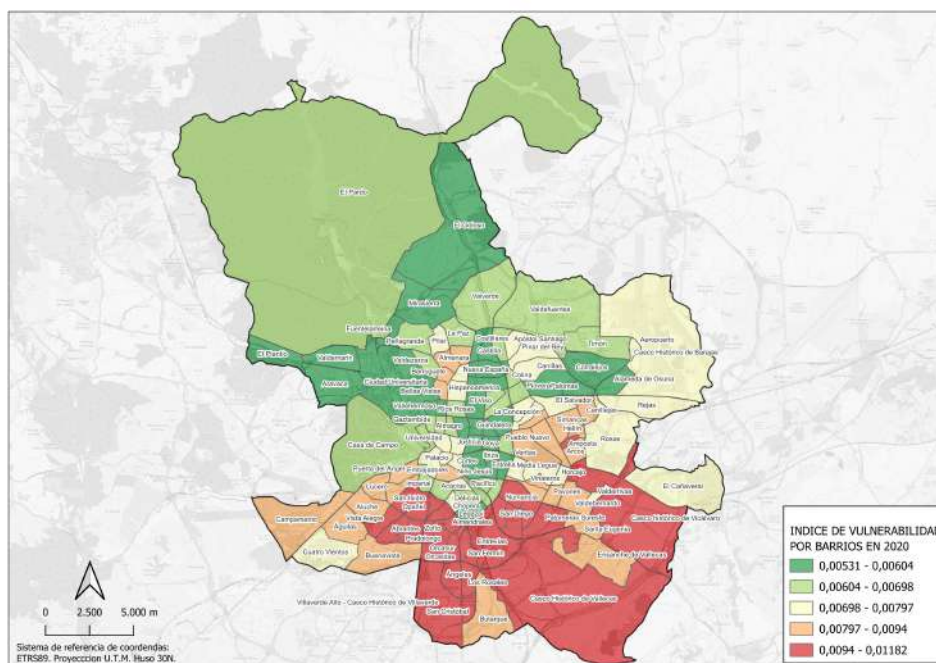


Figura 3.13: zonas vulnerables de Madrid

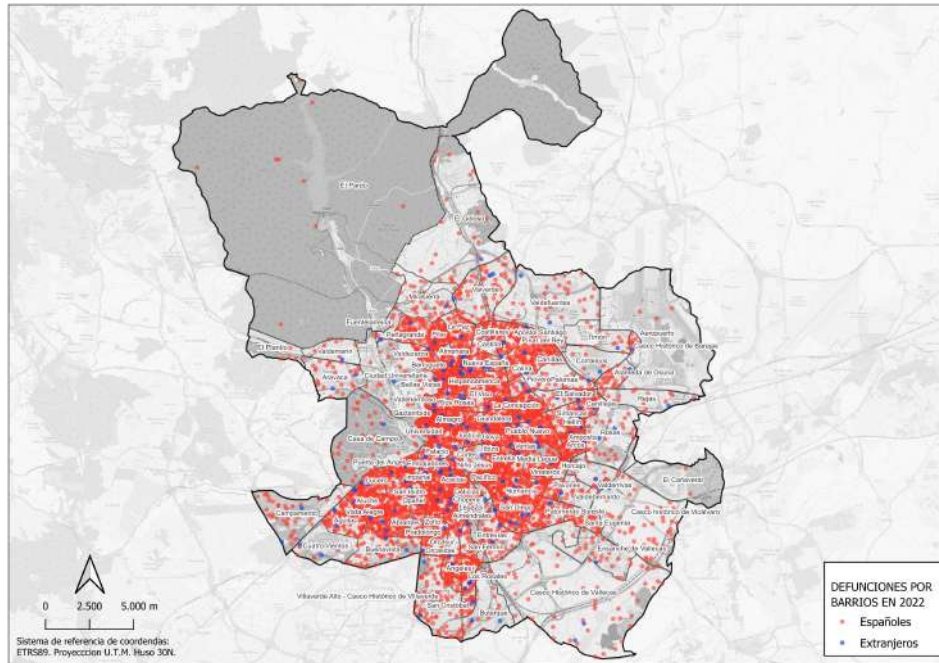


Figura 3.14: defunciones

sur-oeste, coincidiendo con las zonas de mayor concentración de personas y donde residen las de mayor edad.

Otro parámetro importante es la seguridad ciudadana. La figura 3.15 es un mapa de la distribución del número de personas detenidas en los distritos de Madrid, lo que puede ser un indicador del estado de la seguridad en los distritos, frente a la percepción de la seguridad, que es una de las variables a incluir en el estudio.

3.3. Implementación

La implementación del trabajo consta de una serie de pasos previos necesarios con los que se preparan los datos y se preprocesan para adecuarlos a la ejecución del algoritmo no supervisado **K-Means**. Todo el código se ha construido en un documento *R Markdown* y el resultado se ha generado en un documento *html*, ambos disponibles en el repositorio de [GitHub](#).

Los pasos a seguir son los siguientes:

- construcción del Conjunto de Datos objeto del estudio
- Revisar valores ausentes o anómalos
- Normalización de las variables

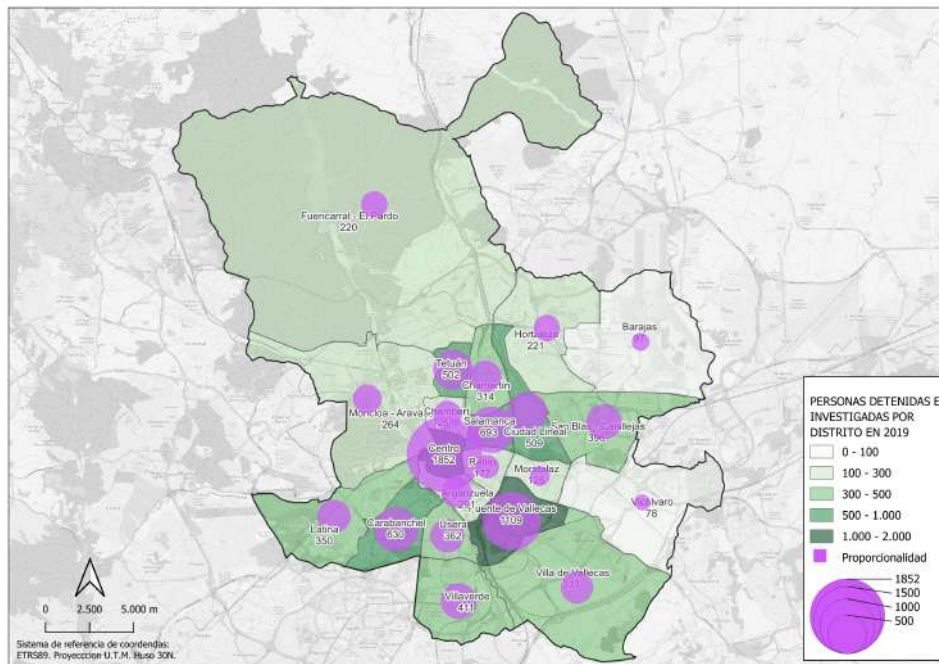


Figura 3.15: mapa de detenciones policiales

- Análisis de Componentes Principales (PCA)
- Determinar el valor óptimo de K (número de clústeres)
- Ejecución de los algoritmos K-means y K-medoids

3.3.1. Construcción del Conjunto de Datos

A partir de los ficheros obtenidos en las fuentes de datos, se hace una preselección de aquellos que pueden contener datos útiles.

Como se ha mencionado en el desarrollo del Trabajo, la idea es construir dos conjuntos de datos que faciliten las tareas, uno a nivel de distritos y otro a nivel de barrios.

3.3.1.1. Datos socio-económicos. Panel de Indicadores de Madrid

El conjunto de datos más interesante se ha obtenido del **Panel de indicadores de distritos y barrios de Madrid**. Se trata de un estudio socio-demográfico, que ofrece una visión territorial de las variables socio-económicas, de salud, demográficas, educativas, calidad de vida, vivienda, medio ambiente, equipamientos municipales, participación ciudadana y presupuesto de los distritos de Madrid. también incluye parte de la información a nivel general de la ciudad y a nivel barrio para algunos de los indicadores disponibles. Se trata de un fichero en formato

XLS (lo he tratado como csv para la implementación) con una gran cantidad de indicadores de la ciudad de Madrid, organizados en las siguientes columnas:

- Orden: indica el número de orden en el fichero;
- Periodo panel: año de publicación;
- ciudad: Ciudad de Madrid;
- cod distrito: código identificador del distrito, del 1 al 21;
- distrito: nombre del distrito;
- cod barrio: código identificador del barrio. Sigue el patrón código distrito + número barrio;
- barrio: nombre del barrio;
- año: año al que corresponden los datos;
- fecha indicador: fecha del indicador;
- fuente indicador: origen del indicador;
- categoría1: categorización de los datos;
- categoría2: sub-categorización de los datos;
- indicador nivel1: primer nivel de clasificación;
- indicador nivel2 segundo nivel de clasificación;
- indicador nivel3 tercer nivel de clasificación;
- unidad indicador: unidad de medida del indicador;
- indicador completo: descripción del indicador;
- valor indicador: valor numérico del indicador;

Lo primero ha sido una visualización del fichero con Excel, y aplicar una serie de filtros para familiarizarme con el contenido del fichero. Esto ha permitido elegir una serie de indicadores que he considerado útiles para el estudio. La figura [3.16](#) muestra el fichero abierto con Excel aplicando filtros sobre los datos.

Se constata que la ciudad de Madrid está dividida en 21 Distritos que contienen un total de 131 Barrios. Los datos hacen referencia entre los años 2019 y 2022, y no siempre se tiene datos

D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
d_distri	distrito	cod_barri	barrio	año	fecha_indica	fuentes_in	categoria_1	categoria_2	indicador_nivel1	indicador_nivel2	indicador_nivel3	unidad_indicador	indicador_comple	valor_indi
rid				2019	2019	Portal de daSeguridad		Intervención de la Policía Municipi	Número de intervencione	Seguridad	Población delitos relaci	Intervenciones	Intervenciones de la	5.455
rid				2019	2019	Portal de daSeguridad		Intervención de la Policía Municipi	Número de intervencione	Seguridad	Tenencia ilícita de arm	Intervenciones	Intervenciones de la	1.362
rid				2019	2019	Portal de daSeguridad		Intervención de la Policía Municipi	Número de intervencione	Seguridad	Patrimonio	Intervenciones	Intervenciones de la	6.984
rid				2019	2019	Portal de daSeguridad		Intervención de la Policía Municipi	Número de intervencione	Seguridad	Tenencia y consumo de	Intervenciones	Intervenciones de la	8.346
	1 Centro			2019	2019	Portal de daSeguridad		Intervención de la Policía Municipi	Número de intervencione	Seguridad	Población delitos relaci	Intervenciones	Intervenciones de la	943
	1 Centro			2019	2019	Portal de daSeguridad		Intervención de la Policía Municipi	Número de intervencione	Seguridad	Tenencia ilícita de arm	Intervenciones	Intervenciones de la	716
	1 Centro			2019	2019	Portal de daSeguridad		Intervención de la Policía Municipi	Número de intervencione	Seguridad	Patrimonio	Intervenciones	Intervenciones de la	1.439
	1 Centro			2019	2019	Portal de daSeguridad		Intervención de la Policía Municipi	Número de intervencione	Seguridad	Tenencia y consumo de	Intervenciones	Intervenciones de la	2.155
	2 Arganzuela			2019	2019	Portal de daSeguridad		Intervención de la Policía Municipi	Número de intervencione	Seguridad	Población delitos relaci	Intervenciones	Intervenciones de la	199
	2 Arganzuela			2019	2019	Portal de daSeguridad		Intervención de la Policía Municipi	Número de intervencione	Seguridad	Tenencia ilícita de arm	Intervenciones	Intervenciones de la	19
	2 Arganzuela			2019	2019	Portal de daSeguridad		Intervención de la Policía Municipi	Número de intervencione	Seguridad	Patrimonio	Intervenciones	Intervenciones de la	287
	2 Arganzuela			2019	2019	Portal de daSeguridad		Intervención de la Policía Municipi	Número de intervencione	Seguridad	Tenencia y consumo de	Intervenciones	Intervenciones de la	306
	3 Retiro			2019	2019	Portal de daSeguridad		Intervención de la Policía Municipi	Número de intervencione	Seguridad	Población delitos relaci	Intervenciones	Intervenciones de la	70
	3 Retiro			2019	2019	Portal de daSeguridad		Intervención de la Policía Municipi	Número de intervencione	Seguridad	Tenencia ilícita de arm	Intervenciones	Intervenciones de la	20
	3 Retiro			2019	2019	Portal de daSeguridad		Intervención de la Policía Municipi	Número de intervencione	Seguridad	Patrimonio	Intervenciones	Intervenciones de la	86
	3 Retiro			2019	2019	Portal de daSeguridad		Intervención de la Policía Municipi	Número de intervencione	Seguridad	Tenencia y consumo de	Intervenciones	Intervenciones de la	106
	4 Salamanca			2019	2019	Portal de daSeguridad		Intervención de la Policía Municipi	Número de intervencione	Seguridad	Población delitos relaci	Intervenciones	Intervenciones de la	113
	4 Salamanca			2019	2019	Portal de daSeguridad		Intervención de la Policía Municipi	Número de intervencione	Seguridad	Tenencia ilícita de arm	Intervenciones	Intervenciones de la	18
	4 Salamanca			2019	2019	Portal de daSeguridad		Intervención de la Policía Municipi	Número de intervencione	Seguridad	Patrimonio	Intervenciones	Intervenciones de la	438

Figura 3.16: vista Excel de fichero panel_indicadores_distritos_barrios.csv

de todos los periodos, sólo en año concreto. Para el estudio que se quiere realizar se pueden tomar como válidos sin dar un peso importante al año de referencia, puesto que se acepta que no se producen cambios significativos dentro de este periodo que puedan alterar los resultados de forma significativa.

Otro fichero interesante es el **Informe Padrón Municipal de Habitantes**, disponible en la página web municipal de Madrid, en la sección [Estadística](#). Se trata de un informe referido a 2021 en el que, entre otros datos, se da cuenta de los movimientos de población entre distritos y barrios de Madrid, incluyendo datos sobre género, edad, nacionalidad... Con los datos de este informe y los ficheros de población con información geográfica se ha podido construir la capa en QGIS con la evolución temporal de los movimientos de población (figura 3.1), y las gráficas de distribución de la población por barrios (figura 3.8)

3.3.1.2. Generación del conjunto de datos por Distritos

El fichero con los indicadores por distritos y barrios de Madrid se ha cargado en el entorno de desarrollo **RStudio**, procesándose los siguientes indicadores por Distritos y Barrios (cuando la información está disponible):

- Estado de las viviendas. Las viviendas están catalogadas según su estado (Buen estado, deficiente, malo, ruinoso o no consta el estado)
- Tasa de desempleo.
- Renta media.
- Edad media.
- Densidad de población.
- Seguridad ciudadana. Intervenciones de la Policía Local.

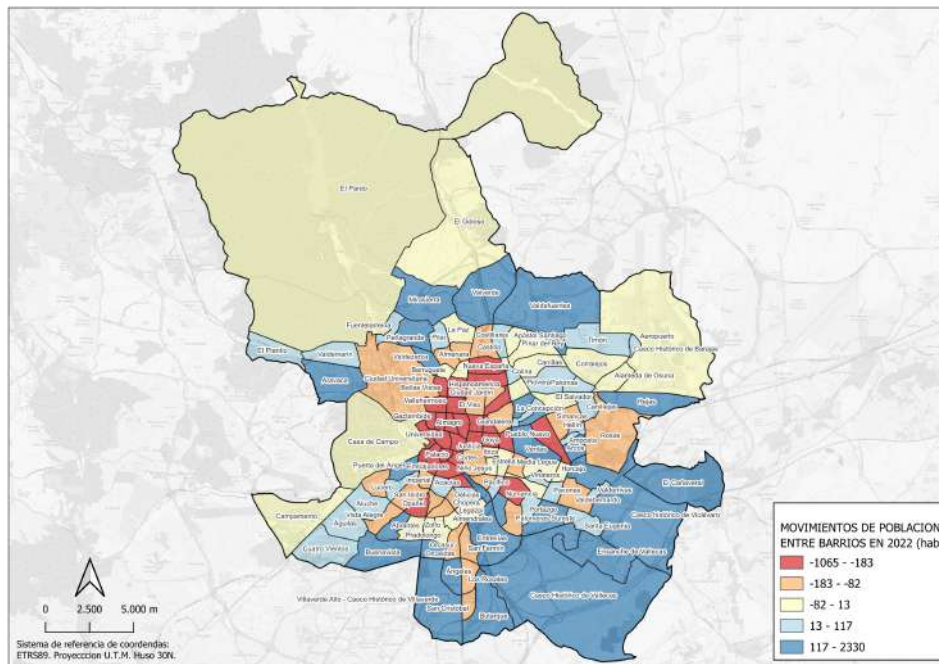


Figura 3.17: movimientos de población

- Seguridad ciudadana. Personas detenidas.
- Seguridad ciudadana. Percepción de la seguridad.
- Valor catastral de los inmuebles. Personas físicas y jurídicas.
- Tasas de desempleo. Por rango de edad y género.
- Indicador de calidad de vida.

Como el objetivo del estudio es tratar de conocer qué factores incluyen en los movimientos de población dentro de Madrid, se ha creado una variable adicional de tipo binaria, *Target*, que indica si el distrito (o barrio) ha tenido un movimiento de población positivo o negativo (ha ganado habitantes o los ha perdido), siguiendo el criterio de:

- $Target = 0$ si el distrito pierde habitantes
- $Target = 1$ si el distrito no pierde habitantes contenidos...

Esta información se muestra en el mapa de movimientos de población (figura 3.17) que refleja los barrios que han perdido o ganado habitantes mediante un semáforo de colores, crecimiento negativo en rojo, naranja, amarillo, y crecimiento positivo en celeste y azul.

3.3.1.3. Alquileres turísticos en Madrid. Inside Airbnb

Otro fichero con información interesante es `airbnb_listings.csv` que contiene un listado de los inmuebles en disponibles en la plataforma airbnb como alquiler turístico en los distritos y barrios de Madrid. El fichero original contiene algunas diferencias en cómo se nombran los distritos y barrios con respecto a los ficheros del Portal de Datos Abiertos de Madrid, lo que se ha solucionado de forma manual con un editor de texto usando la función reemplazar.

3.3.2. Procesado del conjunto de datos por Distritos

El procesamiento anterior finaliza con la obtención del conjunto de datos que se va a utilizar para el estudio con algoritmos no supervisados, K-Means y K-Medoids. El conjunto de datos contiene 21 registros, uno por cada distrito, y 27 atributos de tipo numérico y el nombre del distrito de tipo carácter, tal y como se muestra en la figura 3.18.

La elección de algoritmos no supervisados para este trabajo se justifica, por un lado en que, tal y como se ha enfocado el estudio por distritos, el número de registros disponibles es muy escaso (21 distritos o 131 barrios), lo que es insuficiente para poder realizar el entrenamiento necesario en un algoritmo de tipo supervisado. Por otro lado, se conoce a priori el resultado de la clasificación al tener información de los distritos que han ganado habitantes en detrimento de los que los han perdido, reflejado en el variable *Target*, lo que sería adecuado para un análisis por métodos supervisados, pero teniendo en cuenta que el interés de este estudio no es tanto predecir el número de clústeres (clasificación de los distritos según si sufren gentrificación o no), sino verificar el comportamiento y el grado de validez de las variables utilizadas en el modelo para la predicción, por lo que es más adecuado el uso de métodos no supervisados.

Una vez tomada esta decisión, se ha optado por utilizar dos algoritmos, *K-means* y *K-medoids*, para ver el comportamiento de ambos utilizando distintas métricas y así disponer de mayor información para contrastar los resultados.

Previo a la ejecución de los algoritmos de predicción, se hace una comprobación de que los valores obtenidos están dentro de lo admisible y no aparecen valores anormales. Mediante una representación de tipo *caja de bigotes* se visualizan posibles candidatos a valores anormales, y se comprueba la ausencia de *outliers*, como el ejemplo que se muestra en la figura 3.19, donde se puede observar que hay un candidato a valor anormal pero, una vez estudiados los datos de cada distrito, en realidad se trata de un valor muy por encima de los valores del resto de distritos, pero que toma un valor probable dentro del rango de los datos, además hay que tener en cuenta que estos datos proceden de fuentes oficiales por lo que el carácter de veracidad se le supone.

También es necesario una normalización de las variables, todas numéricas, para poder rea-

```

'data.frame':  21 obs. of  28 variables:
 $ X                : int  1 2 3 4 5 6
 $ Cod_Distrito     : int  1 10 11 12 :
 $ Distrito        : chr  "Centro" "L
 $ Renta_Media      : num  36.1 31.9 29
 $ Edad_Media       : num  44.1 46.5 44
 $ Densidad_Pob     : num  269 95 185 :
 $ Num_Detenciones  : num  1.85 350 630
 $ Calidad_Vida     : num  73.6 68.6 60
 $ Percepcion_seguridad : num  69.8 54.4 49
 $ Tasa_Desem_H_16_24 : num  3.68 5.61 6.
 $ Tasa_Desem_H_25_44 : num  6 7.98 8.49
 $ Tasa_Desem_H_45_64 : num  11.1 10.6 11
 $ Tasa_Desem_M_16_24 : num  3.22 5.39 6.
 $ Tasa_Desem_M_25_44 : num  6.81 9.74 10
 $ Tasa_Desem_M_45_64 : num  10.8 11.6 11
 $ Num_Delitos_Perso : int  943 157 376
 $ Num_Delitos_Pat   : int  1439 185 290
 $ Num_Delitos_Armas  : int  716 29 49 21
 $ Num_Delitos_Drogas : int  2155 214 341
 $ Valor_Cat_Fisicas  : num  117 61.5 52.
 $ Valor_Cat_Juridicas : int  593 232 142
 $ Vivienda_NoConsta : int  1875 1735 41
 $ Vivienda_Bueno    : int  74970 107451
 $ Vivienda_Deficiente : int  7895 5245 60
 $ Vivienda_Malo     : num  1725 810 100
 $ Vivienda_Ruina     : num  330 250 125
 $ Num_Alquileres_Airbnb: int  10607 720 81
 $ Target             : int  0 1 1 0 1 1

```

Figura 3.18: Conjunto de datos para el análisis

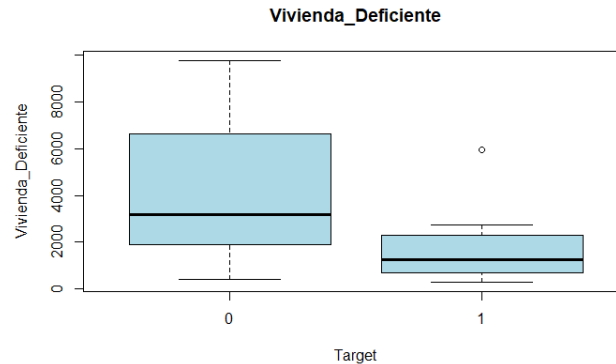


Figura 3.19: *BoxPlot* de la variable Vivienda_Deficiente para identificar *outliers*

lizar las comparaciones de variables con diferentes escalas o sistemas de unidades. La normalización es un tratamiento necesario puesto que posteriormente se necesita realizar un análisis de componentes principales (PCA) que es sensible al cambio de escala en las variables.

3.3.2.1. Correlaciones

Antes de seguir con las predicciones, es interesante conocer la correlación entre las variables objeto de estudio. El que haya variables correlacionadas entre sí quiere decir que “comparte información” por lo que es posible aplicar alguna técnica de reducción de dimensiones como *PCA*. La figura 3.20 muestra el grado de correlación entre las variables, correspondiendo el color rojo a las correlaciones positivas y el azul a las negativas, y la densidad del color representa el grado de correlación.

En principio, parece existir una fuerte correlación entre el mayor número de alquileres turísticos disponibles y el número de intervenciones policiales, mientras que esta correlación del número de alquileres turísticos es muy baja frente a las tasas de desempleo.

También es llamativo el que la percepción de mayor seguridad y la calidad de vida están muy correlacionadas con el valor catastral de los inmuebles, mientras que apenas hay correlación con el resto de variables.

3.3.2.2. PCA

El Análisis de Componentes Principales (PCA) es una técnica necesaria para reducir las dimensiones del problema y evitar lo que se conoce como *la maldición de la dimensionalidad* [14] que no es otra cosa que el problema de que al aumentar la dimensión de información se puede incrementar la dificultad de su interpretación, por lo que esta técnica ayuda a simplificar las dimensiones con la menor pérdida de información.

La figura 3.21 muestra la varianza explicada por componentes como resultado de aplicar

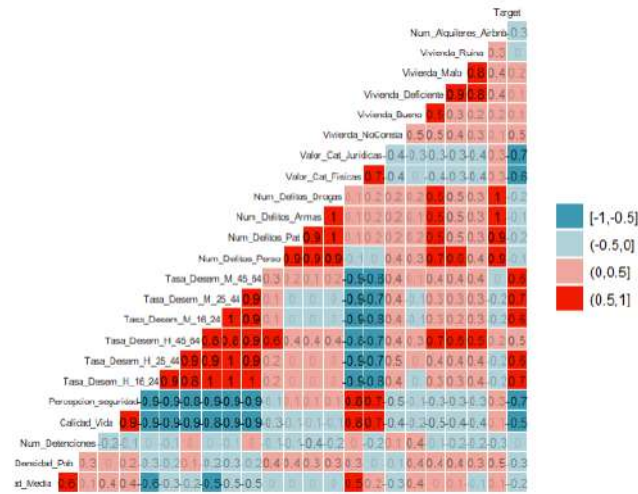


Figura 3.20: Correlaciones entre las variables

PCA al conjunto de datos, mientras que en la figura 3.22 se visualiza cómo contribuye cada variable en las componentes principales. esta gráfica muestra que son las dos primeras componentes las que tienen mayor relevancia y que a partir de la tercera componente no hay saltos en el porcentaje de variabilidad.

3.3.2.3. *K-means*

El primer algoritmo no supervisado que se quiere implementar es *K-means* basado en métrica euclídea, donde el centroide de un grupo se calcula como la media de los puntos de datos del grupo. Para ello, es necesario identificar el número de *clústeres* que se quieren usar, lo que se averigua aplicando el *método del codo* mediante la función *fviz_nbclust*. La figura 3.25 ayuda en la determinación del número de clústeres óptimo, valor que se aproxima a la zona donde la gráfica se suaviza, siendo los valores 3 y 4 los candidatos a número de clústeres adecuado.

Parece más adecuado utilizar un valor de 2 para el número de clústeres, puesto que la variable objetivo (*Target*), de tipo binaria, representa si el distrito ha ganado o perdido habitantes. Sin embargo, se puede entender que el agrupamiento no es sólo de tipo Sí/No, sino que puede representar una probabilidad de ganar o perder habitantes, o una medida que cuantifica si el grado de ganancia o pérdida de habitantes, y por tanto el grado de gentrificación.

Se aplica K-means para ambos valores para, con los resultados obtenidos analizar las posibles conclusiones. Las figuras 3.26 y 3.27 muestran los agrupamientos obtenidos con este algoritmo. en el caso de $K = 2$ agrupamientos, se han obtenido 9 distritos en el primer clúster y 12 en el segundo.

La gráfica 3.23 representa cómo se agrupan las dos dimensiones principales obtenidas con la técnica *PCA* según el resultado de clasificar con *K-means* con $K = 2$.

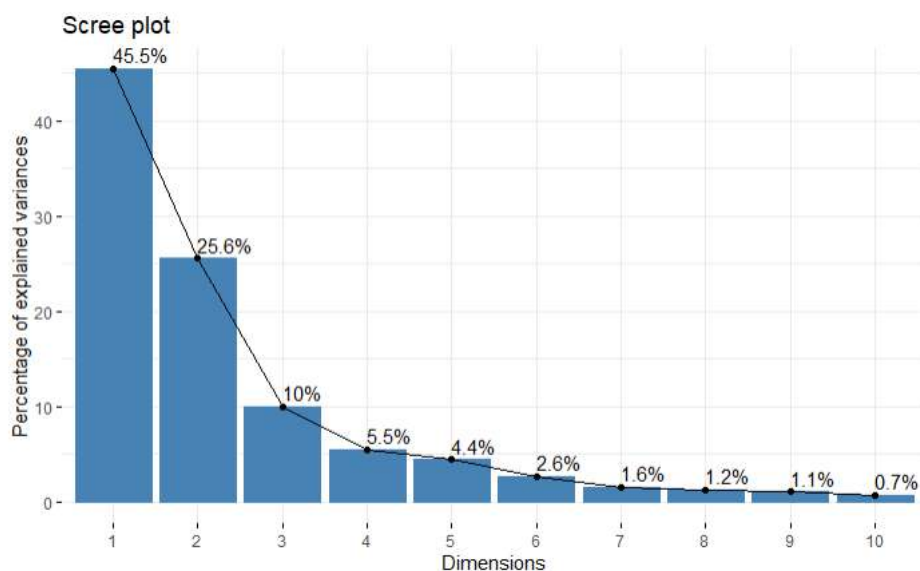


Figura 3.21: PCA. Varianza explicada por cada componente

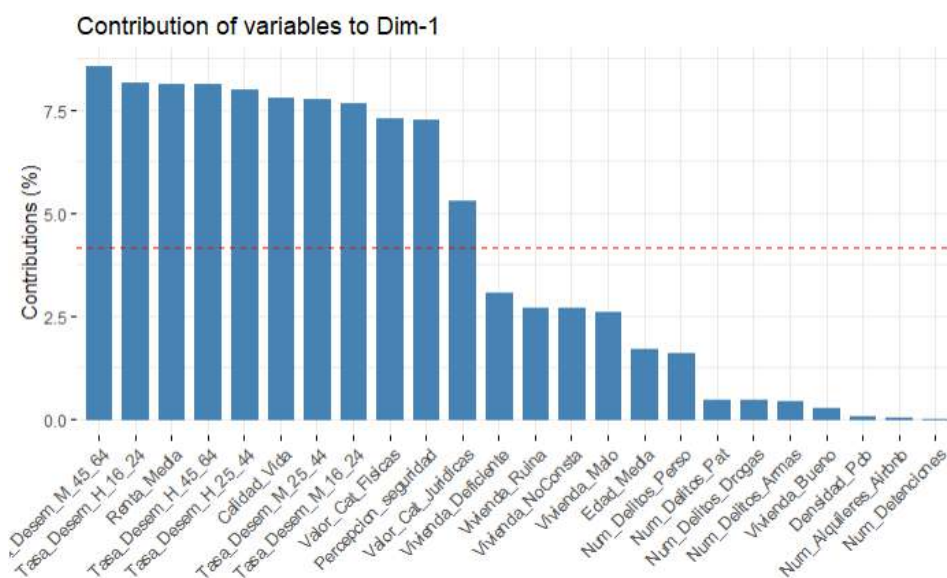


Figura 3.22: PCA. Contribución de las variables originales en los componentes principales

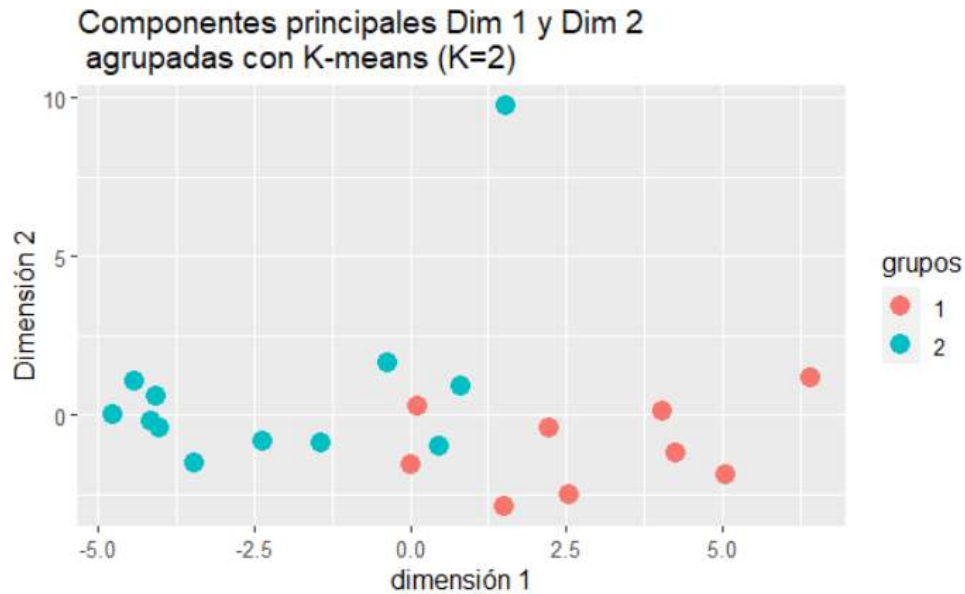


Figura 3.23: Agrupamiento de las 2 dimensiones principales para $k=2$

En la gráfica 3.24 se superpone a la anterior la clasificación de las observaciones con respecto a la variable *Target*. Se puede comprobar que hay un alto grado de coincidencias, por lo que el método de agrupamiento para predecir qué distritos están en riesgo de gentrificación es muy alto.

3.3.2.4. *K-medoids*

Otro algoritmo no supervisado es *K-medoids* con métrica *manhattan*, que es la suma de las diferencias absolutas de las coordenadas, y a diferencia de *k-means*, el centroide de un grupo es uno de los puntos de datos reales en el grupo. Al igual que en el algoritmo *k-means*, es necesario indicar el número de clústeres inicial. Se implementa, como es lógico, con los mismos valores de clústeres que en el caso anterior.

Las figuras 3.28 y 3.29 muestran los resultados obtenidos con éste algoritmo.

3.3.3. Resultados

La aplicación de los algoritmos *K-means* y *K-medoids* han generado resultados muy similares, y próximos a los esperados. La figura 3.27 muestra dos agrupamientos bien definidos, lo que indica que el conjunto de datos seleccionado para este estudio contiene variables adecuadas para la predicción de fenómenos de gentrificación por turistificación en la ciudad de Madrid.

Incluso el caso de 3 clústeres muestra un agrupamiento consistente con lo esperado, donde el agrupamiento del centro de la figura 3.26 coincide con los distritos con barrios con baja pérdida

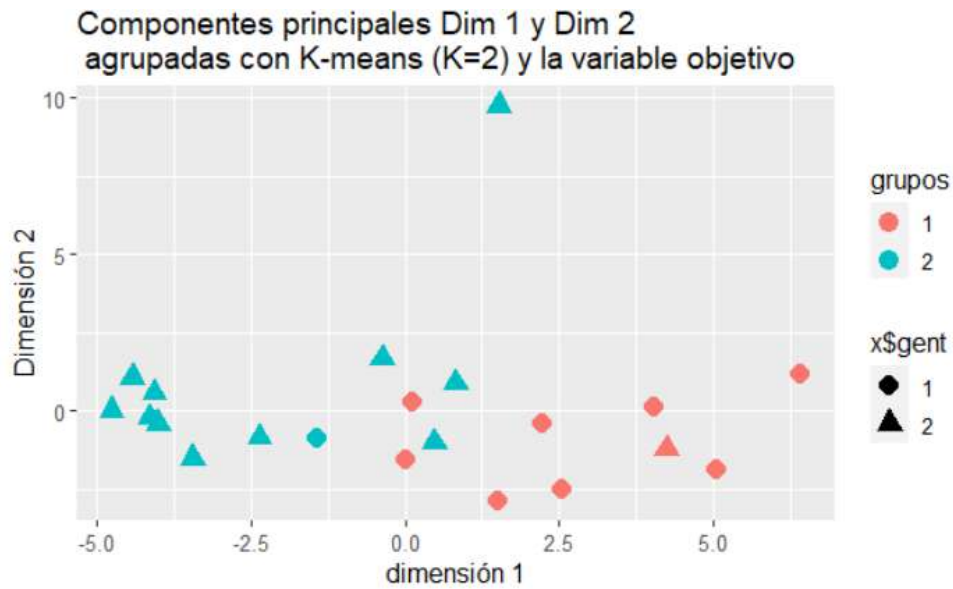


Figura 3.24: Agrupamiento de las 2 dimensiones principales, k=2 y la variable objetivo

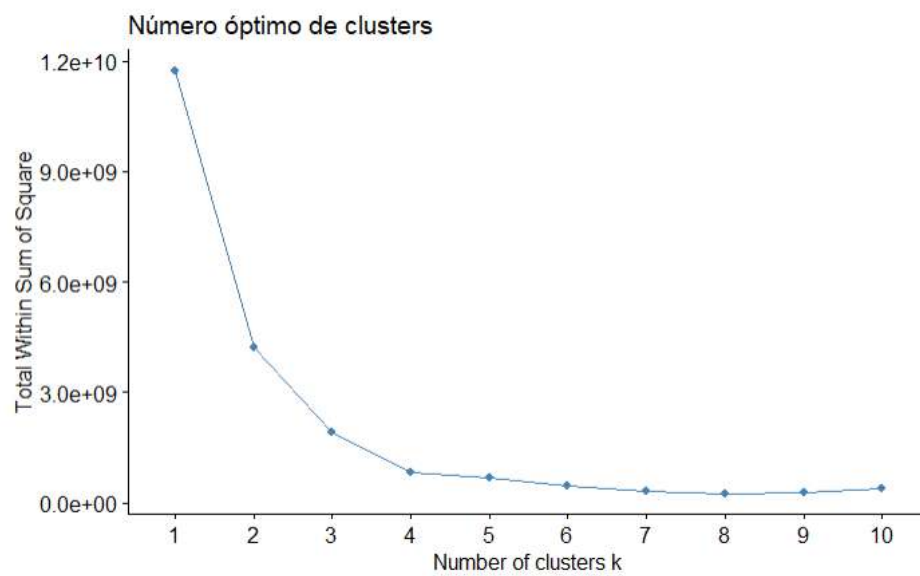


Figura 3.25: Curva para identificar el número de clústeres



Figura 3.26: K-means para 2 clústeres

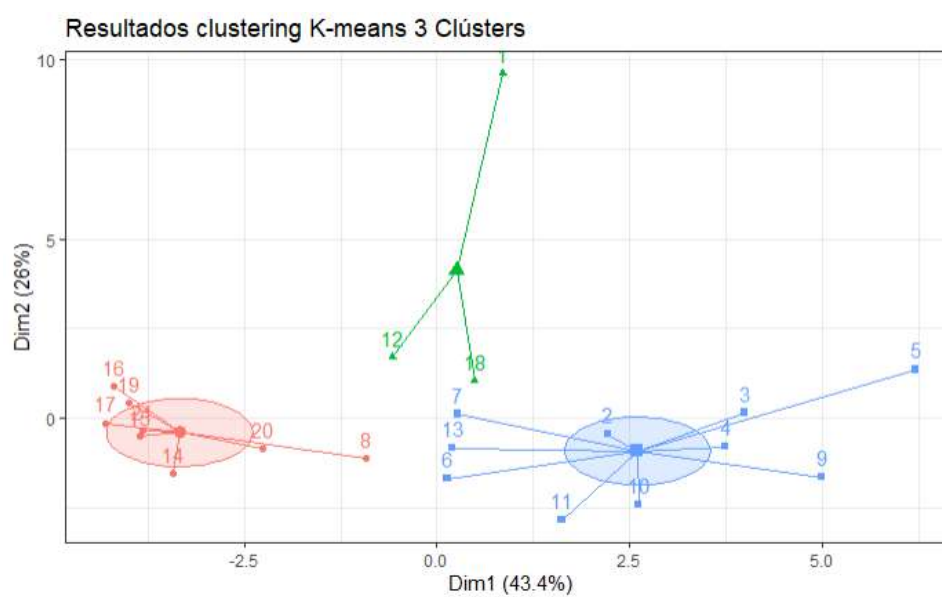
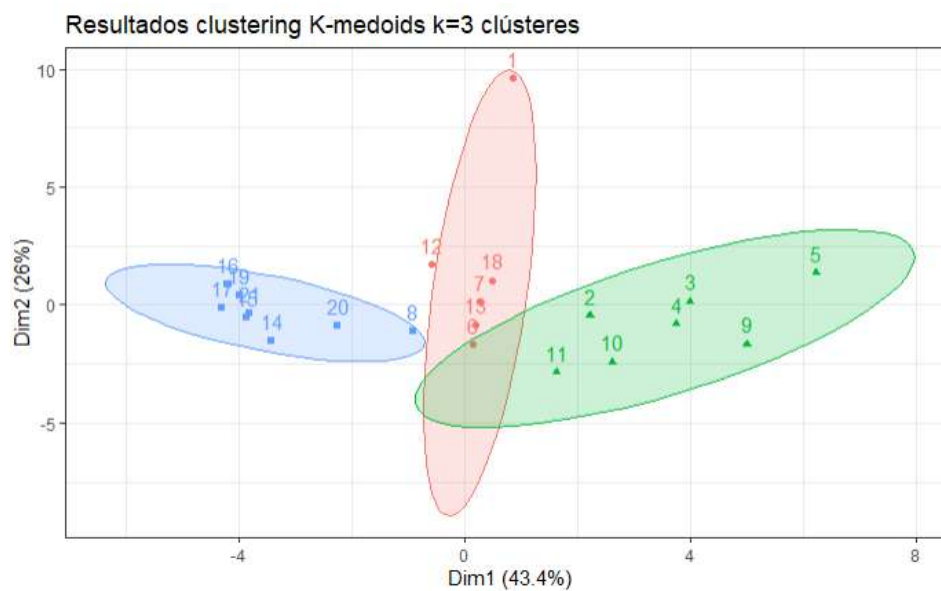


Figura 3.27: Kmeans para 3 clústeres



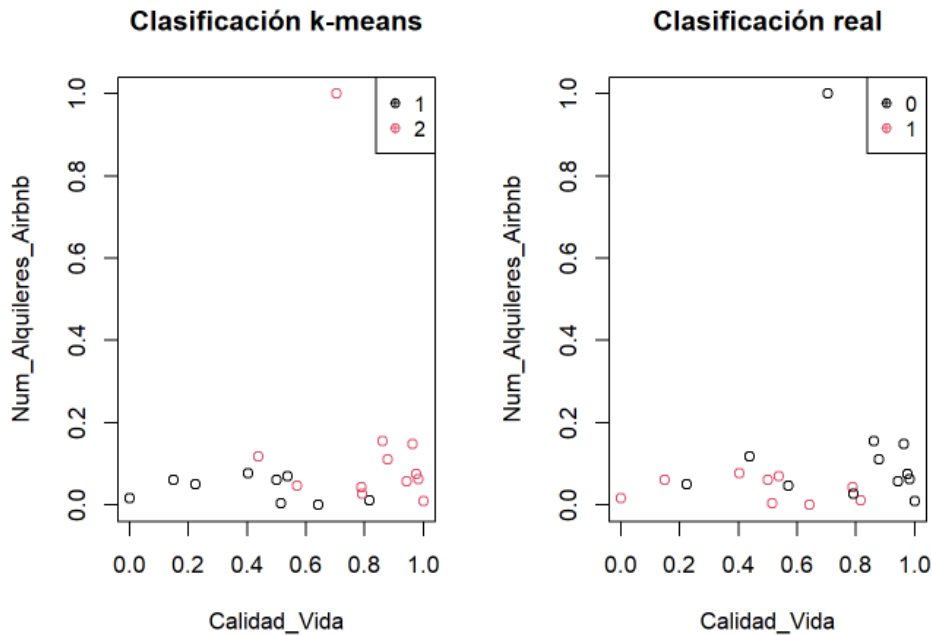


Figura 3.30: Comparación de agrupamientos para alquiler turístico y calidad de vida

de habitantes.

En el caso de *K-medoids*, las gráficas 3.29 y 3.28 coinciden con las predicciones de *k-means*, pero en el caso de 2 clústeres, se muestra una intersección de los agrupamientos, lo que es consistente con la forma en que *K-medoids* realiza el agrupamiento.

Las gráficas comparativas del agrupamiento obtenido y el esperado (figuras 3.30 y 3.31), muestran que efectivamente el agrupamiento predicho por el algoritmo coincide con los datos conocidos a priori. Se muestra la variable *Calidad de Vida* que según la fuente de datos (panel de indicadores de Madrid) representa el porcentaje de satisfacción de los ciudadanos con la calidad de vida en su barrio. Después de la normalización, el porcentaje de satisfacción está entre 0 (muy baja calidad de vida) y 1 (muy alta calidad de vida).

3.4. Conclusiones y líneas de trabajo futuro

En primer lugar, cabe concluir que las visualizaciones realizadas con *QGIS* han facilitado la comprensión del problema, además de ser una herramienta muy práctica para ayudar a determinar variables candidatas al análisis numérico posterior. Los mapas generados a partir de los datos disponibles han permitido comprobar que existe un desplazamiento de la población de Madrid desde el centro hacia los territorios del sur y oeste de la ciudad.

En segundo lugar, los dos algoritmos no supervisados, *K-means* y *K-medoids*, junto con la técnica de reducción de dimensiones mediante Análisis de Componentes Principales (PCA),

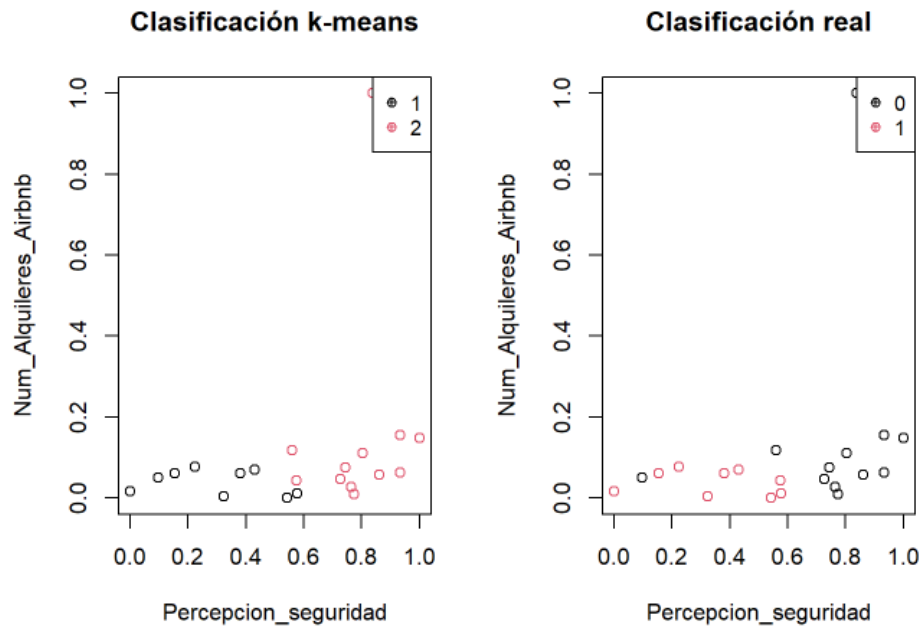


Figura 3.31: Comparación de agrupamientos para alquiler turístico y percepción de seguridad

utilizados en este trabajo han resultado una herramienta útil para analizar el impacto de las variables seleccionadas en el conjunto de datos en los movimientos de personas hacia los barrios de la periferia de una gran ciudad como es Madrid.

Los resultados obtenidos sugieren que las variables con mayor correlación entre ellas, que corresponden con las relativas a la seguridad ciudadana, la conservación del barrio y la penetración de alquileres de corta estancia pueden tener efectos negativos en la población original de los barrios e inducen desplazamientos hacia la periferia sur de Madrid, incrementando la segregación social, es decir, provocando un proceso de gentrificación acentuado por el nuevo fenómeno de turistificación que se añade a la anterior.

La aplicación de los dos algoritmos no supervisados han producido resultado casi idénticos por lo que no es determinante la decisión entre un tipo de métrica (euclídea o manhattan). En principio se podría haber pensado que hubiese sido más apropiado utilizar el método *K-medoids* por que trabaja mejor con densidades de puntos, sin embargo, a la vista de los resultados obtenidos para el enfoque del problema, no ha sido determinante y cualquiera de los dos métodos ha resultado igualmente válido.

Hay que notar que los resultados obtenidos son específicos para el conjunto de datos utilizados y aunque no se debe generalizar a otros contextos, es cierto que las variables utilizadas son comunes a cualquier otra gran población.

Como se indicaba en la motivación de este trabajo, considero interesante proponer como **futura línea de investigación** el estudio de gentrificación turística aplicando los métodos de

este trabajo, al análisis de los fenómenos de migración del lugar de residencia desde ciudades que son núcleo económico hacia las pequeñas poblaciones de una comarca, como el caso que ocurre en la Costa del Sol donde se está produciendo un importante desplazamiento hacia los pequeños pueblos del interior, lo que produce importantes cambios en la forma de vida, pudiendo alterar el tipo de vida y costumbres tradicionales de estos pueblos. sin embargo, durante el proceso inicial de búsqueda previa de fuentes de datos adecuadas para abordar este tipo de problemas se ha hecho patente que, si bien en el caso de la ciudad de Madrid sí se han localizado fuentes de datos con carácter oficial con muchos indicadores adecuados para este estudio, no es así en la mayoría de grandes ciudades españolas, lo que puede condicionar trasladar este estudio a otras ciudades y más aún para el caso de comarcas.

También considero interesante incluir este tipo de análisis en los planes estratégicos de grandes ciudades y provincias, sobre todo, para ayudar en la elaboración de planes urbanísticos y tener presente el impacto sobre el modelo de ciudad que puede tener el fenómeno de la turistificación.

Bibliografía

- [1] IBM Corporation. Conceptos básicos de ayuda de crisp-dm, 2017.
- [2] A. Hernández. Gentrificación y turistificación: origen común, efectos diferentes. *Dimensiones Turísticas* 5(9), 128-137., 2021.
- [3] Julio José Parralejo Sánchez and Ibán Díaz Parra. Gentrificación y turistificación en las áreas urbanas centrales de sevilla y cádiz. *La reconfiguración capitalista de los espacios urbanos: transformaciones y desigualdades / Juan Manuel Parreño Castellano y Claudio Jesús Moreno Medina (Coords.), p. 293-306*, 2021.
- [4] Pablo Martínez Mar Santamaría. atlas de la turistificación, 2019.
- [5] R. Atkinson and G. Bridge. *Gentrification in a Global Context*. Housing and Society Series. Taylor & Francis, 2004.
- [6] Tim Verlaan and Cody Hochstenbach. Gentrification through the ages. *City*, 26(2-3):439–449, 2022.
- [7] Wikipedia. Gentrification.
- [8] Lance Freeman. Displacement or succession?: Residential mobility in gentrifying neighborhoods. *Urban Affairs Review - URBAN AFF REV*, 40:463–491, 03 2005.
- [9] Agustin Cocola-Gant. Tourism gentrification. *Handbook of Gentrification Studies*, 2018.
- [10] David Moriente and Ordinas Botella. Procesos de turistificación y gentrificación en américa latina y españa: una panorámica. *hist.mem.*, N^o. 26. Año 2023, pp. 11 - 18, 2023.
- [11] Pablo Martínez Galíndez. La expansión del turismo hacia la periferia a través de airbnb: el caso de villaverde. *Handbook of Gentrification Studies*, 2020.
- [12] Álvaro Mazorra Rodríguez. Tendencias de gentrificación y turistificación en el centro histórico de madrid. el caso de lavapiés. *Historia y Memoria, ISSN-e 2027-5137, N^o. 26, 2023*, 2022.

-
- [13] Anna Ortiz-Guitart, Antonio López-Gay and Joan Sales-Favà, and Miguel Solana-Solana. La gentrificación desde una mirada de género: un ejemplo en barcelona. *CIUDAD Y TERRITORIO, ESTUDIOS TERRITORIALES*, Vol. LIII, N^o 210, Págs. 945-962, 2021.
- [14] Francesc Pozo Montero and Nuria Parés Maríné. Descomposiciónen valores singulares: introducción y aplicaciones. 2020.