



GRADO EN INGENIERÍA EN TECNOLOGÍAS INDUSTRIALES

TRABAJO FIN DE GRADO

EXTRACCIÓN DE CONOCIMIENTO DE PUNTOS DE INTERÉS GEORREFERENCIADOS DE UNA CIUDAD UTILIZANDO TÉCNICAS DE APRENDIZAJE AUTOMÁTICO

Autor: Ignacio Jiménez Serrano

Director: Eugenio Francisco Sánchez Úbeda

Madrid

Julio de 2020

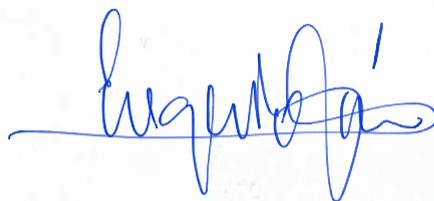
Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título **Extracción de conocimiento de puntos de interés georreferenciados de una ciudad utilizando técnicas de aprendizaje automático** en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el curso académico 2019/2020 es de mi autoría, original e inédito y no ha sido presentado con anterioridad a otros efectos. El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido tomada de otros documentos está debidamente referenciada.



Fdo.: Ignacio Jiménez Serrano

Fecha: 14/ 07/ 2020

Autorizada la entrega del proyecto
EL DIRECTOR DEL PROYECTO



Fdo.: Eugenio Francisco Sánchez Úbeda Fecha: 14/ 07/ 2020



GRADO EN INGENIERÍA EN TECNOLOGÍAS
INDUSTRIALES

TRABAJO FIN DE GRADO

EXTRACCIÓN DE CONOCIMIENTO DE PUNTOS DE
INTERÉS GEORREFERENCIADOS DE UNA CIUDAD
UTILIZANDO TÉCNICAS DE APRENDIZAJE
AUTOMÁTICO

Autor: Ignacio Jiménez Serrano

Director: Eugenio Francisco Sánchez Úbeda

Madrid

Julio de 2020

EXTRACCIÓN DE CONOCIMIENTO DE PUNTOS DE INTERÉS GEORREFERENCIADOS DE UNA CIUDAD UTILIZANDO TÉCNICAS DE APRENDIZAJE AUTOMÁTICO

Autor: Jiménez Serrano, Ignacio.

Director: Sánchez Úbeda, Eugenio Francisco

Entidad Colaboradora:: ICAI – Universidad Pontificia Comillas

RESUMEN DEL PROYECTO

Palabras clave: Funciones de densidad, clustering, visualización, OSM, INE, COVID-19.

1. Introducción

El objetivo principal de este proyecto es desarrollar un modelo basado en técnicas de aprendizaje automático que permita trabajar con datos georreferenciados altamente desestructurados de forma generalista, para extraer conocimiento relevante asociado a la distribución espacial de diferentes aspectos socioeconómicos de una región.

2. Definición del Proyecto

En este proyecto se ha desarrollado un modelo basado en datos georreferenciados para extraer conocimiento sobre la organización espacial de una ciudad. El proyecto se ha dividido en tres partes: descarga y almacenamiento de la información, desarrollo del modelo y aplicación del modelo en dos problemas concretos.

En la primera parte del proyecto se ha desarrollado código específico en Matlab para descargar información de Open Street Map (OSM) y del Instituto Nacional de Estadística (INE). Posteriormente esta información se ha filtrado y estructurado para generar una base de datos propia que ha servido para desarrollar el modelo. En el caso de OSM se ha estudiado tres formas diferentes de descarga para seleccionar la forma que mejor se ajusta a la región de estudio, la Comunidad de Madrid.

En la segunda parte se ha desarrollado un modelo ajustado en cuatro etapas. Para demostrar su utilidad, se ha aplicado en dos problemas diferentes: la clasificación de los barrios de

Madrid y la extracción de conocimiento sobre la evolución de la pandemia del coronavirus en Madrid.

3. Descripción del modelo

El modelo desarrollado combina el ajuste de funciones de densidad bidimensionales con las técnicas de clustering para extraer conocimiento de los datos georreferenciados disponibles para una ciudad como Madrid. En su planteamiento ha sido necesario introducir dos conceptos: dimensión e indicador. Una dimensión es una cualidad que define un barrio de la ciudad, por ejemplo, sanidad, educación, transporte, ocio, etc. Por otro lado, un indicador socioeconómico es una característica que ayuda a definir una dimensión, por ejemplo, si se quiere definir la dimensión educación, se utilizará los indicadores colegio, universidad y academias. En base a los indicadores y dimensiones definidas, es posible ajustar los coeficientes que definen el modelo propuesto. A continuación, se describe las 4 etapas fundamentales para crear el modelo.

La primera etapa consiste en agrupar los datos disponibles georreferenciados para construir los indicadores de aquellas dimensiones que se quiera generar. Esta agrupación tiene lugar en el plano latitud-longitud. No todos los datos se pueden representar en este plano, por eso, para representar algunas dimensiones hay que realizar aproximaciones en algunos indicadores, concretamente en el número de habitantes por secciones censales del INE, en algunos de los datos de OSM que no tienen una única localización espacial, y en el número de contagiados de coronavirus de la CAM.

En la segunda etapa se ajustan funciones de densidad bidimensionales a partir de las dimensiones definidas de la etapa anterior. Los datos de los indicadores que forman estas dimensiones se utilizan para ajustar una función de densidad tipo Kernel usando la función *mvsdensity* de Matlab. Las funciones de densidad obtenidas pueden ser superficies muy complejas, definidas en el espacio longitud-latitud sin una expresión analítica compacta, por lo que es necesario almacenarlas muestreadas. En la tercera etapa se realiza este muestreo, que en el caso estudio de la Comunidad de Madrid, está formado por 250,000 puntos organizados en una rejilla rectangular.

Por último, en la cuarta etapa se utilizan las técnicas de clustering mediante la función *kmeans* de Matlab para encontrar grupos de puntos coherentes en el espacio definido por las diferentes funciones de densidad ajustadas para cada una de las dimensiones consideradas.

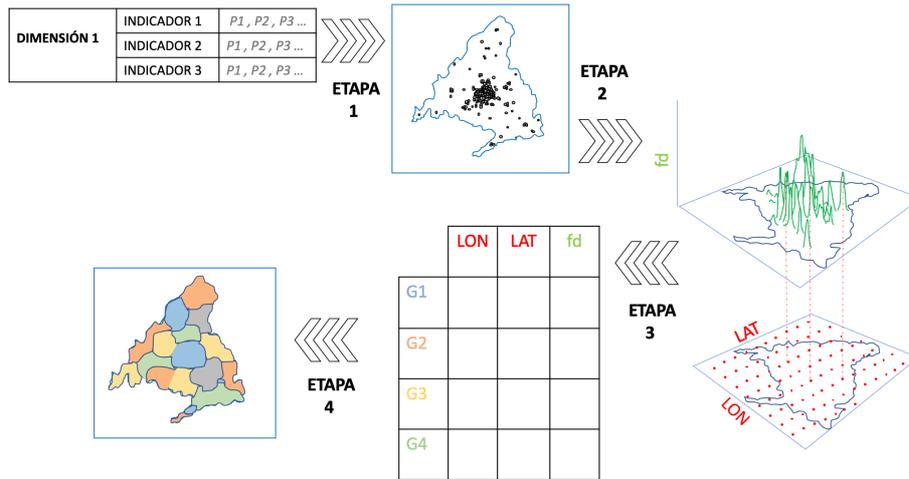


Figura 1: Mapa conceptual del modelo

4. Resultados

El modelo desarrollado se ha aplicado en dos problemas diferentes para demostrar su utilidad. En el primero se han clasificado los diferentes barrios de Madrid atendiendo a diferentes dimensiones. En el segundo, se ha estudiado la evolución espacial de la pandemia COVID-19 en Madrid. A continuación se muestran los principales resultados obtenidos.

4.1. Clasificación de los barrios de Madrid

Se han generado 6 clústeres según 12 dimensiones para analizar las diferencias entre los principales barrios de Madrid, estas dimensiones son: alimentación, restaurantes, educación, financiero, sanidad, cultura, ocio, tiendas, población joven, población adulta y población mayores, renta.

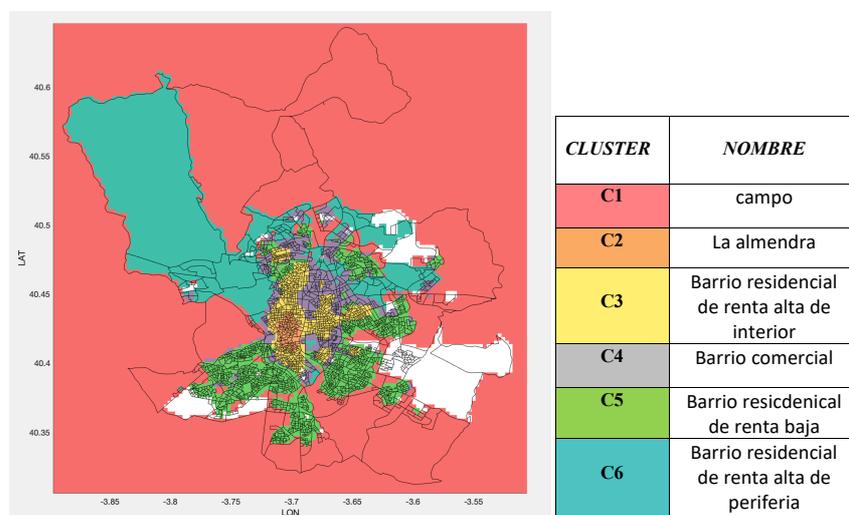


Figura 2: Tipos de barrios atendiendo a las 12 dimensiones consideradas.

4.2. Análisis de la Pandemia COVID-19 en Madrid

Para poder estudiar la evolución de la pandemia se ha descargado información sobre el número de contagiados en la Comunidad de Madrid y las Zonas Básicas de Salud (ZBS).

Por un lado se ha evaluado que ZBS han sido las más afectadas en función de la densidad de contagiados y por otro lado se ha cruzado la información de los contagiados con las densidades de la población total y de los hospitales para extraer conocimiento sobre la evolución de la pandemia, contrastando los resultados obtenidos con la información publicada en los periódicos sobre los hospitales más afectados en Madrid.

5. Conclusiones

El modelo propuesto para extraer conocimiento de información georreferenciada altamente desestructurada ha demostrado ser de gran utilidad al permitir resolver dos problemas concretos. Por un lado, se ha podido determinar automáticamente, a partir de un conjunto de dimensiones socioeconómicas y demográficas, cómo está organizado espacialmente la ciudad de Madrid, determinado que existen 6 tipos fundamentales de barrios diferentes. Además, se ha estudiado la evolución espacial de la pandemia del COVID-19 en la región de Madrid, identificando las zonas más afectadas y su relación con los hospitales que más demanda de atención sanitaria han soportado.

Basándose en el conocimiento extraído con este tipo de modelado se podría facilitar la comprensión de la organización socioeconómica de las zonas que puedan verse afectadas en nuevos rebrotes de COVID-19 o futuras pandemias, ayudando de esta forma a la toma de decisiones para realizar una buena gestión de los casos, tanto desde el punto de vista asistencial en los hospitales implicados, como social de los barrios.

En conclusión, se ha demostrado con los dos casos concretos considerados la flexibilidad del enfoque propuesto y la diversidad de soluciones que pueden aportar estos modelos a posibles problemas que se puedan dar en la sociedad.

KNOWLEDGE EXTRACTION FROM GEOREFERENCED INTEREST POINTS FROM A CITY USING LEARNING TECHNIQUES.

Author: Jiménez Serrano, Ignacio.

Director: Sánchez Úbeda, Eugenio Francisco

Colaborative entity: ICAI – Universidad Pontificia Comillas

ABSTRACT

Keywords: Density functions, clustering, visualization, OSM, INE, COVID-19

1. Introduction

The main objective of this project is to develop a model based on automatic learning techniques that allows working with highly unstructured georeferenced data in a generalist way, to extract relevant knowledge associated with the spatial distribution of different socioeconomic aspects of a region.

2. Project definition

In this project, a model based on georeferenced data has been developed to extract knowledge about the spatial organization of a city. The project has been divided into three parts: download and storage of the information, development of the model and the application of the model for two specific problems.

In the first part of the project, specific Matlab code has been developed to download information from Open Street Map (OSM) and the National Institute of Statistics (INE). This information was then filtered and structured in order to generate the database used to develop the model. For the OSM extraction, three different ways of downloading have been studied in order to select the most appropriate method for the region of study, The Community of Madrid.

In the second part, a model has been developed in four steps. To illustrate its usefulness, it has been applied in two problems: (i) the classification of Madrid's neighborhoods and (ii) the knowledge extraction of the evolution of the coronavirus pandemic in Madrid.

3. Model description

The developed model combines the adjustment of two-dimensional density functions with clustering techniques to extract knowledge from the georeferenced data available for a city such as Madrid. In order to apply these methods, it has been necessary to introduce and define two concepts: dimension and indicator. A dimension is a quality that defines a neighborhood in the city, for example, health, education, transport, leisure, etc. On the other hand a socio-economic indicator is a characteristic that helps define a dimension, for example, in order to define the education dimension, you will use the school, university and academy indicators. Based on the indicators and dimensions defined, it is possible to adjust the coefficient that define the proposed model. The following paragraphs describe the four fundamental steps to create the model.

The first stage consisted of grouping the available georeferenced data to build the indicators of those dimensions that were to be generated. This grouping took place on the latitude-longitude plane. Not all data could be represented on this plane, so in order to represent some dimensions it was necessary to make approximations in some indicators. Specifically, in the number of inhabitants by INE census sections, in some of the OSM data did not have a single spatial location, and in the number of people with the coronavirus in the Community of Madrid.

In the second stage, two-dimensional density functions were adjusted from the dimensions defined in the previous stage. The indicator data that forms these dimensions was used to adjust a Kernel-type density function using Matlab's *mvskdensity* function. The obtained density function can be a very complex surface, defined in the longitude-latitude space without a compact analytical expression, so it is necessary to store them sampled. In the third stage, this sampling is done, which in the case study of the Community of Madrid, is formed by 250,000 points organized in a rectangular grid.

Finally, in the fourth stage clustering techniques were applied using Matlab's *kmeans* function to find groups of coherent points in the space defined by the different density functions adjusted for each of the dimensions considered.

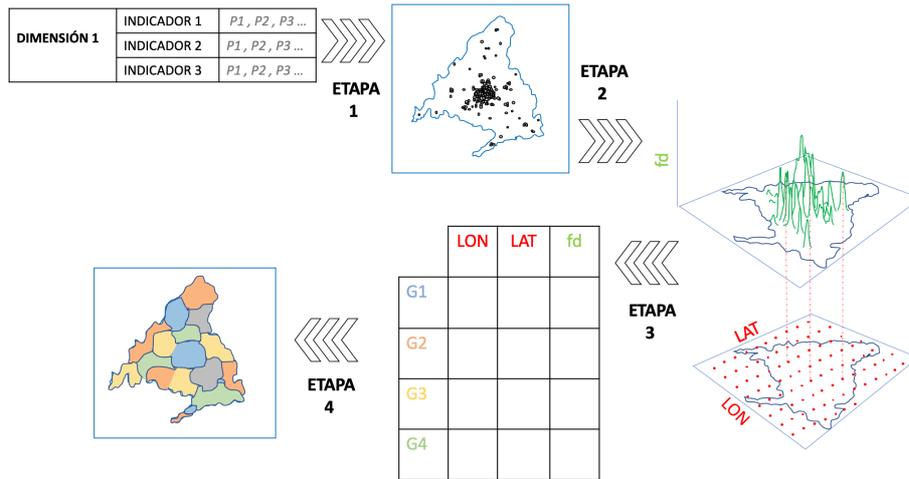


Figure 3: concept map of the model

4. Results

The developed model has been applied in two different problems to demonstrate its usefulness. In the first problem, the different districts of Madrid were classified according to different dimensions. The second problem was the study of the spatial evolution of the COVID-19 pandemic in Madrid. The main results obtained are shown below.

4.1. Madrid's neighbourhood classification

Six clusters, according to 12 dimensions, have been generated to analyse the differences between the main districts of Madrid. These dimensions are: food, restaurants, education, finance, health, culture, leisure, shops, young population, adult population and elderly population.

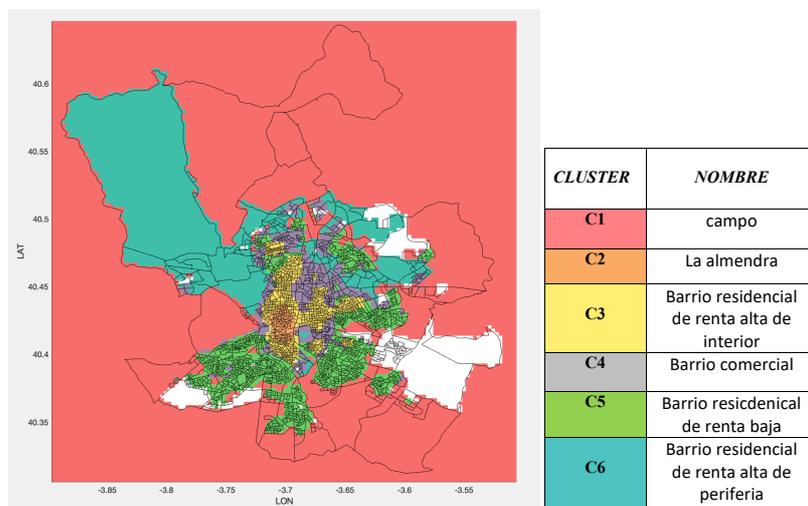


Figure 4: neighbourhood classification according to 12 dimensions.

4.2. COVID-19 pandemic analysis in Madrid

In order to study the evolution of the pandemic, information on the number of people infected in the Community of Madrid and the Basic Health Zones (BHS) was downloaded.

On the one hand, it has been evaluated which of the ZBS have been the most affected according to the density of infected people. On the other hand, the information of infected people was crossed with the densities of the total population and the hospitals in order to extract knowledge about the evolution of the pandemic, contrasting the results obtained with the information published in the newspapers about the most affected hospitals in Madrid.

5. Conclusion

The proposed model for extracting knowledge from highly unstructured georeferenced information has proved to be very useful in solving two specific problems. On the one hand, it has been possible to determine automatically, from a set of socio-economic and demographic dimensions, how the city of Madrid is organized spatially, determining that there are six fundamental types of different neighbourhoods. Furthermore, the spatial evolution of the COVID-19 pandemic in the region of Madrid has been studied, identifying the most affected areas and their relationship with the hospitals that have borne the greatest demand for health care.

Based on the knowledge extracted with this type of modeling, we could facilitate the understanding of the socioeconomic organization of the areas that could be affected in new outbreaks of COVID-19 or future pandemics, thus helping in the decision-making process to carry out good case management, both from the point of view of care in the hospitals involved, and from the social point of view of the neighborhoods.

In conclusion, the two specific cases considered demonstrate the flexibility of the proposed approach and the diversity of solutions that these models can provide to possible problems that may arise in society.

Índice de la memoria

Capítulo 1. Introducción	11
Capítulo 2. Estado de la Cuestión	13
Capítulo 3. Definición del Trabajo	14
3.1 Justificación.....	14
3.2 Objetivos.....	14
3.3 Metodología.....	16
3.4 Planificación	17
Capítulo 4. Modelo Desarrollado.....	19
4.1 Descarga y almacenamiento de datos.....	19
4.1.1 Instituto Nacional de Estadística (INE).....	19
4.1.1.1 Secciones censales.....	20
4.1.1.2 Habitantes.....	21
4.1.1.3 Renta Media	21
4.1.2 Open Street Maps (OSM)	22
4.1.2.1 Consulta mediante bounding box.....	22
4.1.2.2 Consulta mediante nombre.....	24
4.1.2.3 Consulta por polígono	25
4.2 Diseño del modelo.....	29

4.2.1 Etapa 1 – Representación en el plano espacial.....	30
4.2.1.1 Habitantes.....	30
4.2.1.2 Descarga en formato Way.....	33
4.2.2 Etapa 2 – Función de densidad.....	35
4.2.3 Etapa 3 – Estructuración de la información.....	36
4.2.4 Etapa 4 – Clústeres.....	37
Capítulo 5. Aplicación.....	38
5.1 Ciudad de Madrid.....	38
5.1.1 Dimensiones.....	38
5.1.1.1 Alimentación.....	38
5.1.1.2 Restaurantes.....	40
5.1.1.3 Educación.....	41
5.1.1.4 Financiero.....	42
5.1.1.5 Sanidad.....	43
5.1.1.6 Cultura.....	44
5.1.1.7 Ocio.....	45
5.1.1.8 Tiendas.....	46
5.1.1.9 Población.....	47
5.1.1.9.1 Jóvenes.....	47
5.1.1.9.2 Adultos.....	48

INTRODUCCIÓN

5.1.1.9.3 Mayores	49
5.1.1.10 Renta Media	50
5.2 Covid - 19	52
5.2.1 Número de contagiados	52
5.2.2 Zonas Básicas de Salud (ZBS).....	55
5.2.3 Dimensiones.....	63
5.2.3.1 Número de contagiados – Población total.....	63
5.2.3.2 Número de Contagiados – Hospitales	65
Capítulo 6. Análisis de Resultados.....	68
6.1 Comunidad de Madrid.....	68
6.1.1 n° clústeres 2.....	69
6.1.2 n° clústeres 3.....	71
6.1.3 n° clústeres 4.....	73
6.1.4 n° clústeres 5.....	75
6.1.5 n° clústeres 6.....	77
6.2 Covid-19	79
6.2.1 Número de contagiados – Población.....	79
6.2.2 Número de contagiados – Hospitales.....	82
Capítulo 7. Conclusiones.....	87

Capítulo 8. Bibliografía.....	88
ANEXO I – Fuentes de Información.....	92
8.1 Open Street Maps (OSM).....	92
8.1.1 Consulta mediante Bounding Box	93
8.1.2 Consulta mediante nombre.....	95
8.1.3 Consulta por polígono.....	97
8.2 Instituto Nacional de Estadística (INE).....	101
8.2.1 Consulta de las secciones censales	101
8.2.2 Consulta del número de habitantes	103
8.2.3 Consulta de la renta media por hogar.....	104
ANEXO II - METAS DE DESARROLLO SOSTENIBLE.....	106

Índice de figuras

<i>Figura 1: Mapa conceptual del modelo</i>	12
Figura 2: Tipos de barrios atendiendo a las 12 dimensiones consideradas.	12
Figure 3: concept map of the model	16
Figure 4: neighbourhood classification according to 12 dimensions.	16
Figura 5 :Delimitación Comunidad de Madrid para consulta mediante bbox.....	23
Figura 6 : Ejemplo consulta mediante bbox de los POIs 'amenity'	24
Figura 7 : Datos solución Código 1.....	25
Figura 8 : Ejemplo consulta de los POIs 'amenity' por polígonos.....	26
<i>Figura 9 : Mapa conceptual del modelo</i>	30
<i>Figura 11 : Ejemplo estimación del número de habitantes totales en una SC</i>	32
<i>Figura 12 : Zoom de la Figura 11</i>	32
Figura 13: Representación de la simplificación de los hospitales en formato way.....	34
<i>Figura 14: Función de densidad de los habitantes menores de edad de la Comunidad de Madrid</i>	36
<i>Figura 15 : Centroides de los clústeres de la dimensión Alimentación</i>	38
Figura 16 : Representación espacial de los clústeres de la dimensión Alimentación.....	39
<i>Figura 17: Centroides de los clústeres de la dimensión Restaurantes</i>	40

<i>Figura 18: Representación espacial de los clústeres de la dimensión Restaurantes.....</i>	40
<i>Figura 19: Centroides de los clústeres de la dimensión Educación</i>	41
<i>Figura 20: Representación espacial de los clústeres de la dimensión Educación</i>	41
<i>Figura 21: Centroides de los clústeres de la dimensión Financiero</i>	42
<i>Figura 22: Representación espacial de los clústeres de la dimensión Financiero.....</i>	42
<i>Figura 23: Centroides de los clústeres de la dimensión Sanidad.....</i>	43
<i>Figura 24: Representación espacial de los clústeres de la dimensión Sanidad</i>	43
<i>Figura 25: Centroides de los clústeres de la dimensión Cultura.....</i>	44
<i>Figura 26: Representación espacial de los clústeres de la dimensión Cultura</i>	44
<i>Figura 27: Centroides de los clústeres de la dimensión Ocio</i>	45
<i>Figura 28: Representación espacial de los clústeres de la dimensión Ocio.....</i>	45
<i>Figura 29: Centroides de los clústeres de la dimensión Tiendas.....</i>	46
<i>Figura 30: Representación de los clústeres de la dimensión Tiendas</i>	46
<i>Figura 31: Centroides de los clústeres de la dimensión Jóvenes.....</i>	47
<i>Figura 32: Representación espacial de los clústeres de la dimensión Jóvenes</i>	47
<i>Figura 33: Centroides de los clústeres de la dimensión Adulto</i>	48
<i>Figura 34: Representación espacial de los clústeres de la dimensión Adulto</i>	48
<i>Figura 35: Centroides de los clústeres de la dimensión Mayores.....</i>	49
<i>Figura 36: Representación espacial de la dimensión Mayores.....</i>	49

Figura 37: Centroides de los clústeres de la dimensión Renta Media.....	50
Figura 38: Representación gráfica de los clústeres de la dimensión Renta Media.....	51
Figura 39: Web Comunidad de Madrid con las información de los contagiados del COVID19	52
Figura 40: Densidad de contagiados en la Comunidad de Madrid.....	54
Figura 41 : Zoom de Figura 40.....	55
Figura 42: Web Comunidad de Madrid con las información de las ZBS.....	56
Figura 43: Zonas Básicas de Salud de la Comunidad de Madrid.....	57
Figura 44: Comparación ZBS (negro) con las SC (rojo).....	58
Figura 45: Zoom de la Figura 44.....	59
Figura 46: Densidad de contagiados por cada ZBS.....	60
Figura 47: 10 ZBS con más contagiados.....	61
Figura 48: Señalización de las 10 ZBS con más contagiados.....	62
<i>Figura 49: Método del codo - número de contagiados – Población total.....</i>	<i>63</i>
Figura 50: Centroides de los clústeres generados de Número de contagiados – Población total.....	64
Figura 51 : Representación espacial de los clústeres generados de Número de contagiados – Población total.....	64
<i>Figura 52 : Método del codo número de contagiados – Hospitales.....</i>	<i>65</i>
Figura 53 : Centroides de los clúster generados de Número de contagiados – Hospitales .	66

Figura 54 : Representación espacial de los clústeres generados de Número de contagiados - Hospitales	67
Figura 55: Método del codo.....	68
Figura 56: Centroides de 2 clústeres.....	69
Figura 57: División de la ciudad de Madrid en 2 clústeres	70
<i>Figura 58: Centroides de 3 clústeres</i>	<i>71</i>
Figura 59: División de la ciudad de Madrid en 3 clústeres	72
<i>Figura 60: Centroides de 4 clústers</i>	<i>73</i>
Figura 61: División de la ciudad de Madrid en 4 clústeres	74
Figura 62: Centroides de 5 clústeres.....	75
Figura 63: División de la ciudad de Madrid en 5 clústeres	76
Figura 64: Centroides de 6 clústeres.....	77
Figura 65: División de la ciudad de Madrid en 6 clústeres	78
<i>Figura 66: centroides dimensiones Población y Contagiados.....</i>	<i>80</i>
Figura 67 : Representación espacial de las variables Contagiados y Población total.	81
Figura 68: Zoom de la Figura 54	82
Figura 70: Introducción de la localización de los Hospitales en la Figura 68.....	85
Figura 71: Diagrama de Voronoi.....	86
Figura 71: Acotación barrio Salamanca mediante una bbox.....	94

Figura 72: Datos solución de la consulta por bbox a los restaurantes del barrio Salamanca	95
<i>Figura 73: Datos solución de la consulta por nombre a los restaurantes del barrio Salamanca</i>	<i>97</i>
Figura 74: Delimitación barrio Salamanca mediante polígono	98
Figura 75: Datos solución de la consulta por polígono a los restaurantes del barrio Salamanca	100
Figura 76: Web del INE con la información de las secciones censales	102
Figura 77: Web del INE con la información de habitantes	103
Figura 78: Web del INE con la información de renta media por hogar	104
Figura 79: Metas de desarrollo sostenible	107

Índice de tablas

Tabla 1: Comparativa entre realizar la consulta por bbox y por polígono	27
Tabla 2: Puntos de bbox que delimitan el barrio Salamanca.....	94
Tabla 3 : Puntos de acotación polígono barrio Salamanca.....	98
Tabla 4: ODS con los que interactúa el proyecto.	108

Capítulo 1. INTRODUCCIÓN

La sociedad se encuentra actualmente es una Era Digital o Era de la Información, que se caracteriza por las nuevas tecnologías emergentes que están cambiando nuestras costumbres y nuestra forma de vida. Estas tecnologías han revolucionado la forma de transmisión de la información permitiendo aumentar la capacidad de almacenamiento y ha facilitado el acceso a ella. De esta forma ahora cualquier persona puede acceder de forma rápida y sencilla a una cantidad enorme de información. Esto unido a la globalización nos permite alcanzar un mayor público a la hora de compartirla.

La información está formada por datos que se generan continuamente gracias a los dispositivos y máquinas que nos rodean. Estos datos, una vez organizados y estructurados, pueden ser utilizados para mejorar nuestras vidas de muchas formas. El potencial que tienen estos datos es razón por la que las empresas lo están demandando tanto. Hoy en día nadie toma decisiones sin basarse en los datos, sobre todo en los negocios. Los datos son necesarios para que las compañías crezcan y mejoren. Esta gran demanda de datos ha convertido la información en un producto con el que se comercializa. Este producto adquiere valor según cuatro factores: velocidad, volumen, veracidad y variedad.

En este proyecto se va a estudiar la aplicación de un modelo basado en una variedad de datos muy específico, los datos georreferenciados. Este tipo de datos permite definir ubicaciones de forma precisa con unidades cartográficas y están dando soluciones a problemas, sobre todo, en el ámbito de la demografía y de la economía, también está cogiendo fuerza en la estadística. El INE (Instituto Nacional de Estadística) se ha dado cuenta del potencial de este tipo de datos y está realizando varias colaboraciones con el IGN (Instituto Geográfico Nacional).

Es por todo esto, por lo que ha considerado necesario el estudio de un modelo basado en este tipo de datos, capaz de dar soluciones novedosas a problemas que hasta ahora no se han resuelto del todo.

Capítulo 2. ESTADO DE LA CUESTIÓN

Tras una búsqueda de soluciones ya existentes en el ámbito de la organización espacial, se ha encontrado varios artículos sobre proyectos relacionados [19][20][21]. En estos artículos se expone el uso de los sistemas de información geográfica como herramienta del método desarrollo para llevar a cabo el proyecto, también se comentan el gran potencial y la gran utilidad de los datos georreferenciados y como la búsqueda de un modelo basado en estos datos puede traer grandes beneficios para la sociedad.

Entre los artículos encontrados destaca [19] por su similitud con el trabajo realizado. Este artículo expone como gracias a las nuevas tecnologías se están mejorando métodos antiguos utilizados durante décadas, como el método de gravedad, para describir la accesibilidad geográfica de la atención primaria. Algunas de las mejoras expuestas es el método de densidad de Kernel basado en datos georreferenciados, utilizado también en este proyecto. El artículo termina destacando la necesidad de estudiar la sensibilidad de estos nuevos modelos respecto a los antiguos para encontrar el modelo que mejor se ajuste a la realidad.

Como se ha demostrado en proyecto anteriores, el uso de los datos georreferenciados tiene un gran beneficio, por eso, con este proyecto lo que se quiere es seguir visibilizando la necesidad de estos modelos.

Capítulo 3. DEFINICIÓN DEL TRABAJO

3.1 JUSTIFICACIÓN

La pandemia que se está viviendo es una situación sin precedentes que ha pillado a muchos gobiernos por sorpresa. La actuación lenta, la falta de protocolo o los miles de muertes son indicadores de la falta de modelos que ayuden a explicar o predecir la evolución de la pandemia. Por lo tanto, se ha decidido estudiar el desarrollo de un modelo basado en datos georreferenciados para resolver los problemas de organización espacial que estamos viviendo.

El primer problema que se quiere resolver, objetivo inicial del proyecto, es la clasificación de los barrios de Madrid en función de la población y los servicios. El segundo problema es la evolución de la pandemia utilizando los conocimientos extraídos del primer problema y los datos de los contagiados.

3.2 OBJETIVOS

- Estudio de la distribución de la población en una ciudad

Ha sido necesario descargar del INE la información de los habitantes por cada sección censal (SC). Dentro de esta información se tiene el número de habitantes total y número de habitantes dividido por género y rango de edades. Se ha utilizado Matlab para georreferenciar, estructurar y analizar la información de la población de Madrid utilizando las SC como división.

- Estudio de la distribución de los diferentes servicios públicos y privados en una ciudad

Para el estudio de la distribución de los servicios, en vez de descargarnos la información del INE, se ha utilizado Matlab para generar consultas a la API de Open Street Map (OSM). Una

vez descargados los datos georreferenciados de OSM, ha sido necesario depurar y estructurar la información para su posterior análisis.

- Clasificación de los barrios de una ciudad

Los datos almacenados se ha modelado considerando diferentes dimensiones para extraer, mediante técnicas de machine learning, conocimientos sobre la organización espacial de una ciudad.

- Análisis de la pandemia del COVID-19

Utilizando la información extraída anteriormente sobre los servicios y la población, ha sido necesario relacionarla con los datos del número de contagiados y las Zonas Básicas de Salud (ZBS) para analizar la evolución de la pandemia.

3.3 METODOLOGÍA

El proyecto se ha dividido en tres partes: descarga y almacenamiento de la información, desarrollo del modelo y aplicación del modelo en dos problemas concretos.

La primera parte ha consistido en la programación de un código en Matlab capaz de descargar información de dos bases de datos: Open Street Map (OSM) y el Instituto Nacional de Estadística (INE). Esta información se ha filtrado y estructurado para generar una base de datos propia que ha servido para desarrollar el modelo. En el caso de OSM se ha estudiado tres formas diferentes de descarga para seleccionar la forma que mejor se ajusta a la región de estudio, la Comunidad de Madrid.

En la segunda parte se ha generado un modelo formado por 4 etapas que posteriormente, en la tercera parte, se ha aplicado en dos problemas: la clasificación de los barrios de Madrid y la extracción de conocimientos de la evolución espacial de la pandemia.

3.4 PLANIFICACIÓN

METODOLOGÍA DE TRABAJO		OCTUBRE	NOVIEMBRE	DICIEMBRE	ENERO	FEBRERO	MARZO	ABRIL	MAYO	JUNIO
PARTE 1 – DESCARGA Y ALMACENAMIENTO										
P1.1	Familiarización con la base de datos OSM.									
P1.2	Investigación de diferentes formas de consulta a OSM.									
P1.3	Descarga de la información del INE.									
P1.4	Descarga de la información de OSM									
PARTE 2 – DESARROLLO DEL MODELO										
P2.1	Etapa 1 del modelo									
P2.2	Etapa 2 y 3 del modelo									
P2.3	Etapa 4 del modelo									
PARTE 3 – APLICACIÓN										
P3.1	Búsqueda y descarga de los datos de los contagiados y de las ZBS.									
P3.2	Clasificación de los barrios de Madrid.									
P3.4	Extracción de conocimientos de la evolución espacial de la pandemia.									
Redacción de la memoria del proyecto										

Capítulo 4. MODELO DESARROLLADO

Este proyecto, como ya se ha comentado antes, está formado por tres partes: descarga y almacenamiento de la información, desarrollo del modelo y aplicación del modelo en dos problemas concretos. Todas las partes se han llevado a cabo con el objetivo principal de visibilizar la utilidad de los datos georreferenciados, por tanto, el modelo se ha desarrollado con el fin de poder representar gráficamente todas las conclusiones en el plano espacial. A continuación se describe con detalle las dos primeras partes del proyecto: descarga y almacenamiento y diseño del modelo.

4.1 DESCARGA Y ALMACENAMIENTO DE DATOS

En este apartado se describe la información descargada del Instituto Nacional de Estadística (INE) y de Open Street Maps (OSM), así como la estructura de almacenamiento utilizada para gestionar estos datos durante el proyecto.

4.1.1 INSTITUTO NACIONAL DE ESTADÍSTICA (INE)

El Instituto Nacional de Estadística (INE) pone al alcance de los ciudadanos información pública que se ha aprovechado para desarrollar este proyecto. En este proyecto se ha utilizado la información de las fronteras de las secciones censales (SC), el número de habitantes y la renta media por hogar.

Las secciones censales (SC) desempeñan dos funciones principalmente. Su primera función consiste en delimitar con exactitud la zona de estudio, en este caso la Comunidad de Madrid (CM), y su segunda función es de apoyo, ha servido de ayuda para entender gráficamente las conclusiones del proyecto.

La información de los habitantes también ha desempeñado dos funciones. Por un lado, se ha utilizado para estimar la densidad de población y por otro lado ha servido para estudiar las tendencias que existen en la organización espacial según edades y género.

Por último, la tercera información descargada del INE es la renta media por hogar, se ha utilizado como indicador económico de los barrios de Madrid.

A continuación se explica para cada tipo de información mencionada su estructuración y almacenamiento. Para más información detallada sobre la consulta de datos ver el ANEXO I – Fuentes de Información - Instituto Nacional de Estadística (INE) .

4.1.1.1 Secciones censales

Siguiendo los pasos mencionados en 8.2.1 se descarga la carpeta *seccionado_2019*, el archivo relevante para este proyecto es el archivo SECC_CE_20190101.shp. En este archivo se almacena la información de tal forma que se tiene por filas la sección censal y por columnas los atributos. Para optimizar la memoria que ha decidido almacenar solo algunos atributos y se ha generado uno nuevo (Polígono).

Dos de esos atributos son X e Y, estos atributos necesitan ser tratados antes de ser almacenados. El INE georreferencia los puntos que delimitan las SC en el sistema de coordenadas universal transversal de Mercator (utm) y es necesario hacer un cambio al sistema de coordenadas geográficas. Como la mayoría de los datos del almacén son de OSM, se ha priorizado su sistema de georreferenciación y por tanto en este proyecto se va a desarrollar con longitudes y latitudes. Se ha descargado la función `utm2deg` [7] para convertir los atributos X e Y al sistema de coordenadas geográficas.

El atributo creado, Polígono, es un vector con la información organizada de X e Y que se utilizará como polígono para descargar la información de OSM, se explica en la sección 4.1.2 Open Street Maps (OSM).

La estructura guardada tiene por filas las SC y por columnas los siguientes atributos:

- X : Vector con las coordenadas longitud de los puntos que delimitan la SC.
- Y: Vector con las coordenadas latitud de los puntos que delimitan la SC.
- CUSEC: El código de identificación de la SC.
- SUPERFICIE_M2: Los metros cuadrados que ocupa la SC.

- NPRO: Nombre de la provincia.
- NCA: Nombre de la comunidad autónoma.
- NMUN: Nombre del municipio.
- POLÍGONO: [X1 Y1 X2 Y2 X3 Y3 ...] Vector que unifica la información de las columnas X e Y.

4.1.1.2 Habitantes

El archivo descargado (leer 8.2.2 para más información) divide la información del número de habitante según edad y género. En las primeras filas del archivo se encuentra el número total de habitantes en grupos quinquenales de edad (columnas) por cada sección censal (filas). En las siguientes filas se encuentra la misma división de edades por cada SC, pero solo para los hombres y finalmente solo de las mujeres.

Esta información se almacena en tres estructuras, una para toda la población, otra para hombres y otra para mujeres. Las estructuras guardadas tienen por filas las SC y por columnas las divisiones por edades:

- IS_SC: El código de identificación de la SC.
- TOTAL: El número total de habitantes en esa sección.
- E_0_4: El número de habitantes entre las edades 0 y 4.
- E_5_9: El número de habitantes entre las edades 5 y 9.
- E_10_14: El número de habitantes entre las edades 10 y 14.
- ...
- E_100YM_s: El número de habitantes con más de 100 años

4.1.1.3 Renta Media

Esta información no ha sido necesario estructurarla para su posterior uso, se ha trabajado directamente con el archivo descargado del INE.

4.1.2 OPEN STREET MAPS (OSM)

Se ha averiguado tres formas diferentes de realizar las consultas: mediante bbox, mediante nombre y por polígonos. A continuación se estudia cuál de los tres tipos de consulta se ajusta mejor a las condiciones del problema, se evalúa las ventajas e inconvenientes y se selecciona la mejor opción. Para más información detallada sobre cómo realizar las consultas leer ANEXO I – 8.1.

4.1.2.1 Consulta mediante bounding box

Lo primero que se necesita para realizar este tipo de consulta es encontrar el bounding box o bbox que mejor delimite la Comunidad de Madrid, este bbox es necesario para indicarle al programa la región que se quiere consultar. Para poder fijarlo, se han utilizado los puntos que delimitan las SC, concretamente, la máxima y la mínima longitud y latitud. Estas cuatro coordenadas dejan definido el bbox (ver Figura 5).

Consultar una región muy grande puede llevar a desborde en la descarga de datos, por eso, una vez se tiene el bbox (rectángulo principal), se ha realizado un programa que lo divide en rectángulos secundarios para realizar las consultas de forma segura. Dividendo el rectángulo principal lo que se consigue es realizar muchas consultas pequeñas asegurando que el programa no desborde. De esta forma si alguna consulta falla, gracias al bloque TRY-CATCH, solo hay que ir al fichero log que se ha generado y buscar aquel rectángulo secundario donde ha fallado la consulta, generar otra vez esta consulta y guardar la estructura con las demás.

El programa realiza una consulta por cada rectángulo secundario y OSM devuelve la información en una estructura. Al final lo que se consigue es almacenar tantas estructuras como consultas realizadas.

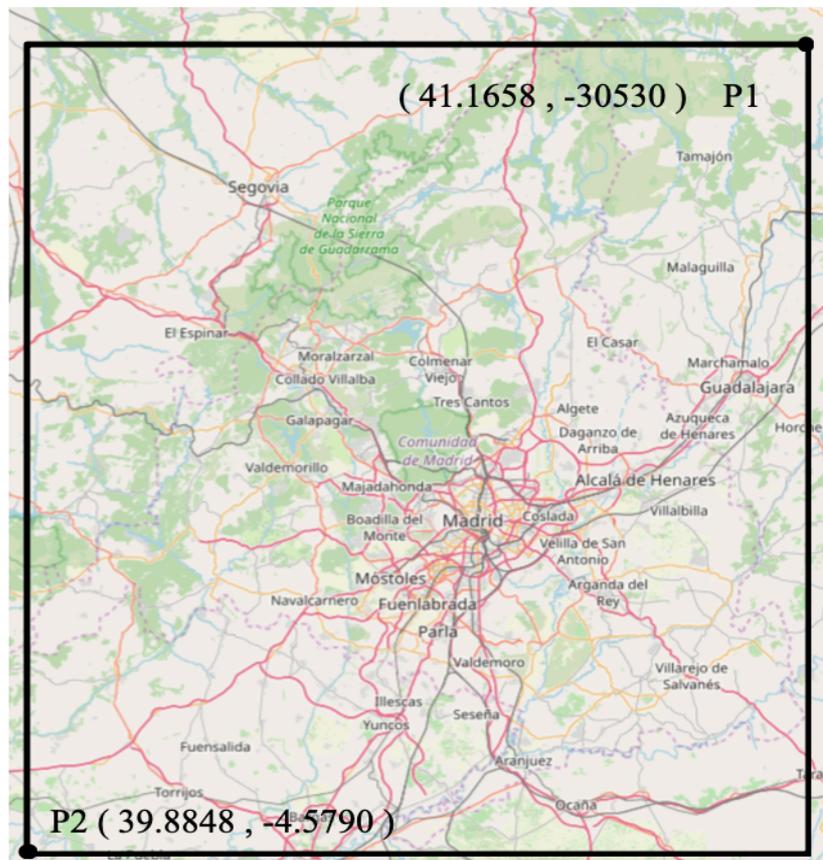


Figura 5 :Delimitación Comunidad de Madrid para consulta mediante bbox

En la Figura 6 se puede observar con diferentes colores cada consulta que se ha realizado. Cada punto representa un POI descargado y cada rectángulo es una división del bbox que define la CM.

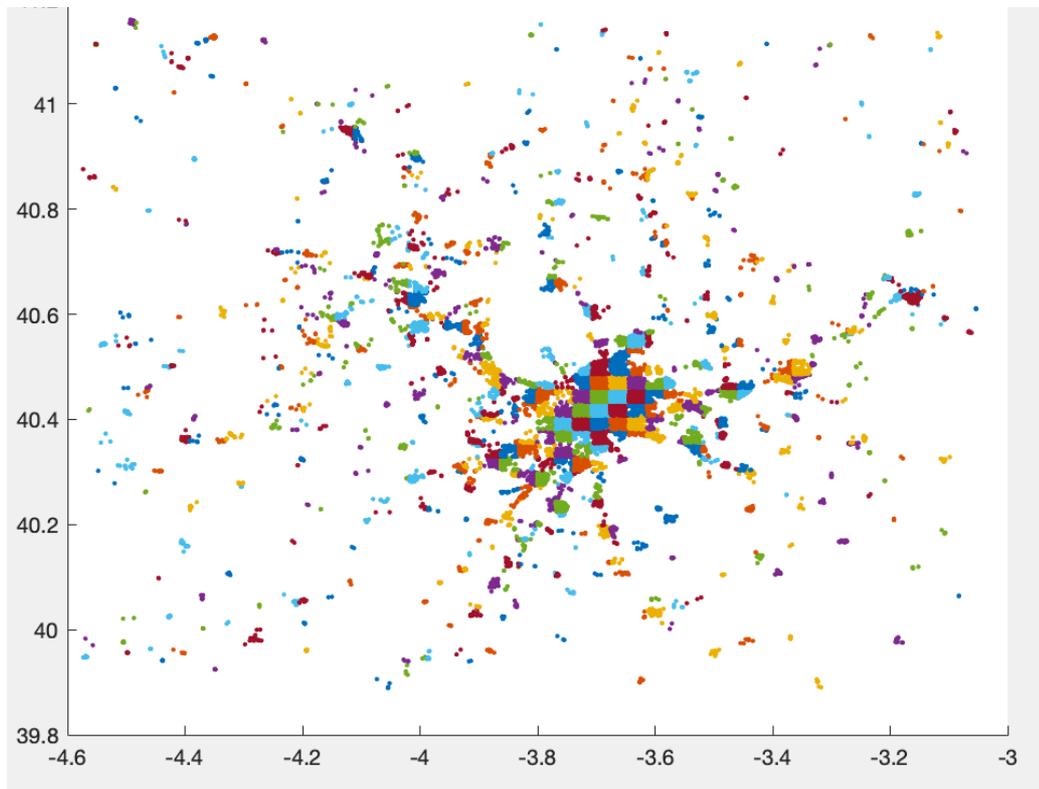


Figura 6 : Ejemplo consulta mediante bbox de los POIs 'amenity'

4.1.2.2 Consulta mediante nombre

La segunda forma de realizar la consulta es utilizando el nombre de los barrios de Madrid. Para ver si este tipo de consulta es factible hay que comprobar que todos los barrios tiene definidas su fronteras en OSM. Se ha utilizado la web Overpass-tubo [3] para consultar esta información. En el

Código 1 se puede ver el código utilizado y en Figura 7 se representa gráficamente los datos solución.

Al analizar la Figura 7 se puede ver cómo en algunos barrios como Canillas o Lavapiés no tienen un punto encima del nombre, esto se debe a que el barrio no está delimitado en la base de datos de OSM. Como la información necesaria para ejecutar esta opción está incompleta se descarta por falta de viabilidad.

```

1  /*
2  This has been generated by the overpass-turbo wizard.
3  */
4  [out:json][timeout:25];
5  area[name="Madrid"]->.kk;
6  (
7  node(area.kk)[admin_level="9"];
8  );
9  // print results
10 out body;
11 >;
12 out skel qt;

```

Código 1 : (web: Overpass-tubo) Busca las fronteras de los barrios de Madrid de OSM

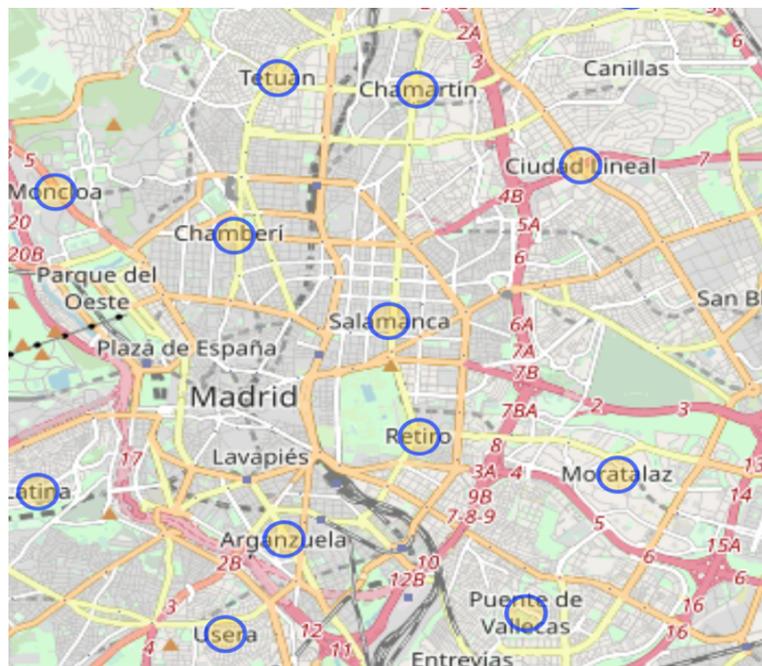


Figura 7 : Datos solución Código 1.

4.1.2.3 Consulta por polígono

En esta forma de consulta se han utilizado las siluetas de las SC como polígonos. Para poder utilizar estas SC previamente se han estructurado los datos de la longitud y latitud en un vector. Este vector se encuentra en el atributo Polígono, ver 4.1.1.1. El programa utiliza esta información y realiza una consulta por cada SC y guarda en el almacén de datos la

información devuelta por OSM. Al final se obtiene tantas estructuras como consultas realizadas, es decir tantas estructuras como SC.

En la Figura 8 se puede observar con diferentes colores cada consulta que se ha realizado.

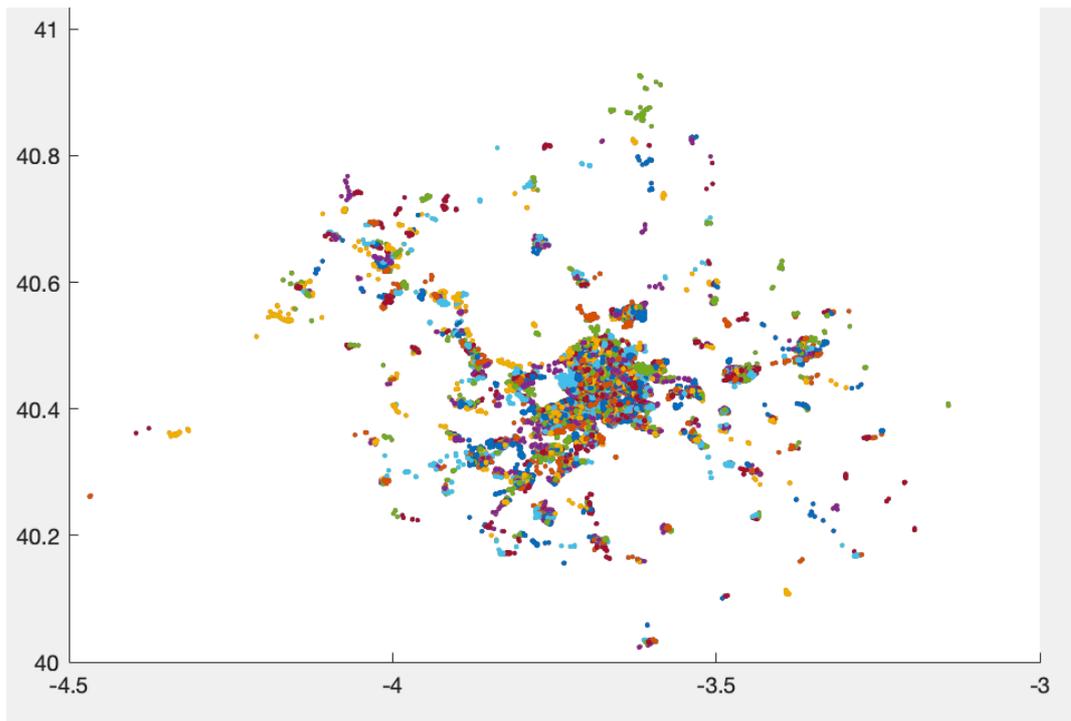


Figura 8 : Ejemplo consulta de los POIs 'amenity' por polígonos

A continuación se estudia cuál de los tres tipos de consulta se ajusta mejor a las condiciones del problema evaluando las ventajas e inconvenientes y seleccionado la mejor opción.

Como se ha explicado anteriormente, la opción de realizar la consulta utilizando el nombre de las regiones no es factible por falta de información en la base de datos OSM.

En la Tabla 1 se ha realizado una comparativa con los aspectos más relevantes a la hora de decidir el tipo de consulta que se debe elegir.

Tabla 1: Comparativa entre realizar la consulta por bbox y por polígono

	<i>BBOX</i>	<i>POLÍGONO</i>
NECESITA DE INFORMACIÓN PREVIA	NO	SÍ
ERROR(%)	10%	35%
NECESITA TRATAR LOS DATOS POSTERIORMENTE	SÍ	NO

Por un lado la opción de realizar la consulta por bbox no necesita necesariamente de información previa, a pesar de haber calculado el bbox a partir de las secciones censales se puede aproximar a ojo, no es necesaria mucha precisión en este primer paso. Donde sí se necesita precisión es a la hora de eliminar los datos sobrantes, como nuestra región es irregular y está siendo acotada por un rectángulo hay datos descargados que se encuentran fuera de la CM. Por tanto, para poder realizar esta filtración con exactitud, se necesita la información de las SC.

Por otro lado, la opción de realizar consulta por polígono sí necesita los datos de las SC como información previa pero no necesita un tratamiento de datos después de la consulta ya que todos los datos descargados se encuentran dentro de CM.

Resumiendo, se tiene que las dos opciones necesitan información exterior a OSM para poder obtener los datos deseados, pero una de ella necesita de un código adicional para un tratamiento posterior a la consulta. En conclusión, si solo se tuviera en cuenta lo comentado hasta ahora la elección idónea sería realizar consulta por polígono, sin embargo, al realizar las consultas se observa que en torno al 35% de las consultas realizadas dan error. Elegir esta opción estaría comprometiendo la veracidad del proyecto y por consiguiente se elige la opción de realizar las consultas por bbox.

En cada consulta OSM devuelve una información que se estructura de la siguiente manera. Cada fila es un POI descargado y cada columna es un atributo de este POI:

- TYPE: Formato del dato: nodo.
- ID: El código de identidad del nodo
- LAT: La latitud del nodo
- LON: La longitud del nodo
- TAGS: Información sobre el nodo como nombre del establecimiento, horarios...
- POI: Especifica el tipo de nodo. Por ejemplo, si se hace la consulta de ‘amenity’ el nodo puede ser muchas cosas como restaurante, universidad, bar, veterinario...Esta columna proporciona esa información más precisa del nodo.

4.2 DISEÑO DEL MODELO

En esta sección se describen las cuatro etapas que forman el modelo. En concreto, para cada una de las partes se explica los datos utilizados, así como su tratamiento y posterior almacenaje. Se describirá con detalle los pasos seguidos y las aproximaciones realizadas para poder reproducir este modelo en otras regiones.

En su planteamiento ha sido necesario introducir dos conceptos: dimensión e indicador. Una dimensión es una cualidad que define un barrio de la ciudad, por ejemplo, sanidad, educación, transporte, ocio, etc. Por otro lado, un indicador socioeconómico es una característica que ayuda a definir una dimensión, por ejemplo, si se quiere definir la dimensión educación, se utilizará los indicadores colegio, universidad y academias. En base a los indicadores y dimensiones definidas, es posible ajustar los coeficientes que definen el modelo propuesto. A continuación, se describe las 4 etapas fundamentales para crear el modelo.

La primera etapa consiste en agrupar los datos disponibles georreferenciados para construir los indicadores de aquellas dimensiones que se quiera generar. Esta agrupación tiene lugar en el plano latitud-longitud. No todos los datos se pueden representar en este plano, por eso, para representar algunas dimensiones hay que realizar aproximaciones en algunos indicadores, concretamente en el número de habitantes por secciones censales del INE, en algunos de los datos de OSM que no tienen una única localización espacial, y en el número de contagiados de coronavirus de la CAM.

En la segunda etapa se ajustan funciones de densidad bidimensionales a partir de las dimensiones definidas de la etapa anterior. Los datos de los indicadores que forman estas dimensiones se utilizan para ajustar una función de densidad tipo Kernel usando la función *mvskdensity* de Matlab. Las funciones de densidad obtenidas pueden ser superficies muy complejas definidas en el espacio longitud-latitud sin una expresión analítica compacta, por lo que es necesario almacenarlas muestreadas. En la tercera etapa se realiza este muestreo, que en el caso estudio de la Comunidad de Madrid, está formado por 250,000 puntos organizados en una rejilla rectangular.

Por último, en la cuarta etapa se utilizan las técnicas de clustering mediante la función *kmeans* de Matlab para encontrar grupos de puntos coherentes en el espacio definido por las diferentes funciones de densidad ajustadas para cada una de las dimensiones consideradas.

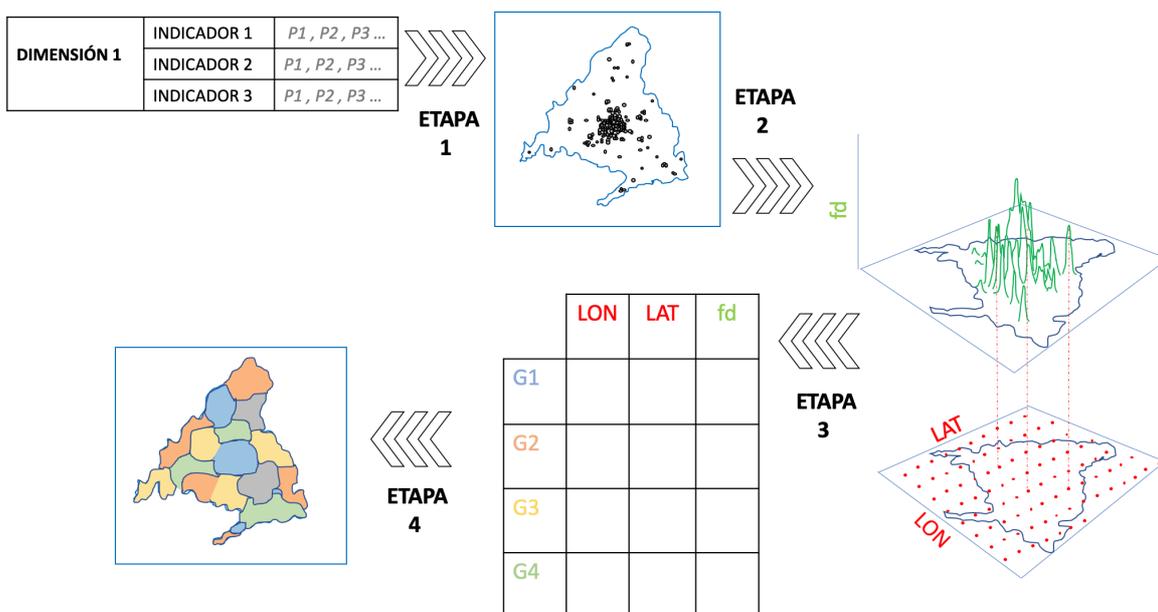


Figura 9 : Mapa conceptual del modelo

4.2.1 ETAPA 1 – REPRESENTACIÓN EN EL PLANO ESPACIAL

Este apartado se explican las hipótesis que se han realizado para poder representar toda la información en el plano espacial.

4.2.1.1 Habitantes

Se parte del número de habitantes por cada sección censal y se necesita representar dicha información en el plano longitud-latitud. Para conseguir este objetivo se ha decidido generar aleatoriamente tantos puntos (longitud-latitud) como habitantes dentro de la cada sección censal. De esta forma cada habitante es representado por un punto situado, más o menos, cerca de su domicilio. Lo que se consigue al final es una estructura que tiene por filas las secciones censales y por columnas:

- ID_SC : Código de identificación de la SC.
- NHOM : Número total de hombres en la SC.
- NMUJ : Número total de mujeres en la SC.
- HOM : Una subestructura cuyas filas son los grupos quinquenales de edades y por columnas:
 - X: Vector con las longitudes de los hombres generados.
 - Y: Vector con las latitudes de los hombres generados.
- MUJ : Una subestructura cuyas filas son los grupos quinquenales de edades y por columnas:
 - X: Vector con las longitudes de las mujeres generadas.
 - Y: Vector con las latitudes de las mujeres generadas.

Esta hipótesis es más precisa en aquellas SC pequeñas con muchos habitantes y menos precisa en las SC grandes con pocos habitantes.

En la Figura 10 se representa un ejemplo de la hipótesis del número de habitantes en la SC 2800201001. Se sabe que hay 1171 hombres y 1044 mujeres, en función de que población se quiere estimar, la total, hombres o mujeres, el código programado generará aleatoriamente en esa SC 1171, 1044 o 2215 puntos aleatoriamente. En la Figura 10 se ha estimado la población total.

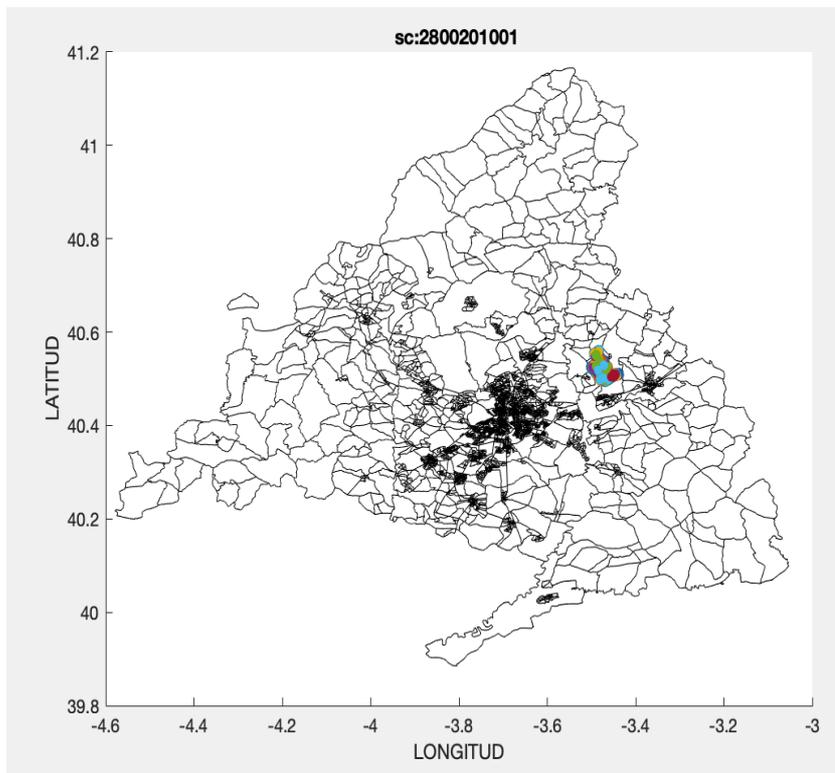


Figura 10 : Ejemplo estimación del número de habitantes totales en una SC

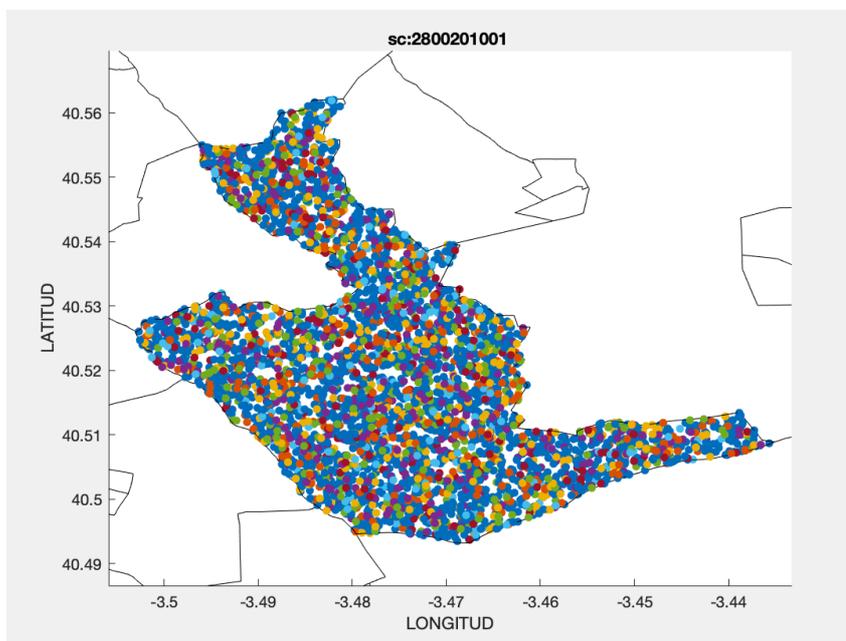


Figura 11 : Zoom de la Figura 10

4.2.1.2 Descarga en formato Way

Hasta ahora el formato que se ha utilizado para descargar la información de OSM ha sido el formato node, que utiliza un único punto georreferenciado para describir un solo POI. En este apartado se presenta el segundo formato que se va a utilizar, el formato way.

Debido a la existencia de POIs que normalmente ocupan edificios enteros de amplias dimensiones, OSM tiene un formato para describir esta información, el formato way. Este formato describe el POI mediante una serie de nodos que forman un polígono que delimita la superficie que ocupa, muy parecido a las SC descargadas del INE. En este proyecto se han descargado en formato way los POIs de los colegios y los hospitales.

La diferencia principal entre el formato node y el formato way es la cantidad de puntos que se utilizan para describir la información. Como en el modelo que se está desarrollando cada POI se debe representar con un único punto, los datos descargados en formato way necesitan una simplificación.

Esta simplificación consiste en calcular la media de las longitudes y latitudes de los puntos que delimitan el polígono, de esta forma se consigue un único punto situado cerca del centro del polígono. En la Figura 12 se puede ver en color rojo el polígono que se ha descargado de OSM y en azul el punto que se ha calculado y que se va a utilizar como entrada en el modelo.

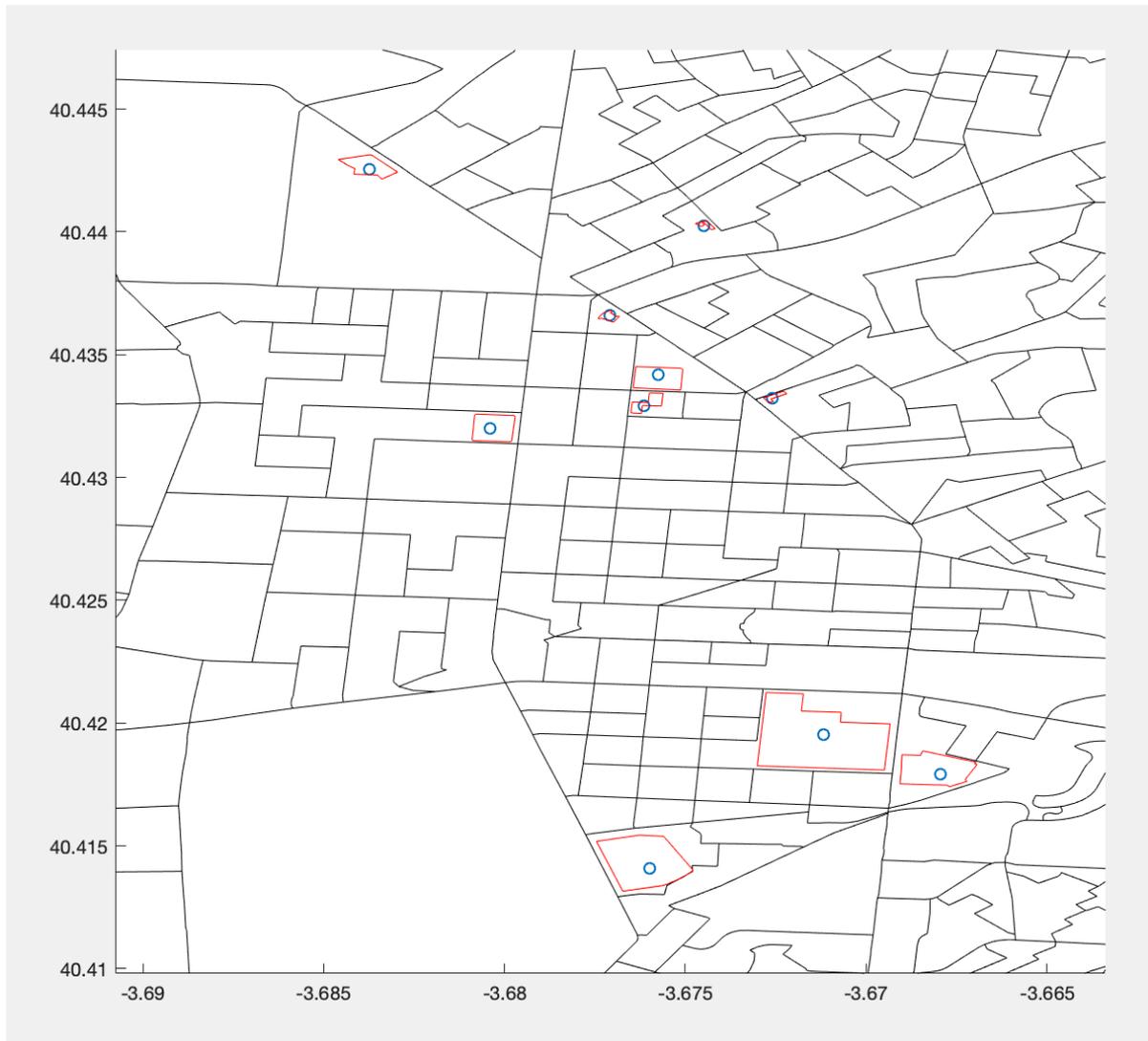


Figura 12: Representación de la simplificación de los hospitales en formato way

Esta información se guarda de la siguiente forma. Se tiene por cada fila un POI y por columnas:

- TYPE: Tipo de formato.
- ID: Código de identidad en la base de datos de OSM.
- TAGS: Información sobre el nodo como nombre del establecimiento, horarios...
- POI: Especifica el tipo de nodo. Por ejemplo, si se hace la consulta de 'amenity' el nodo puede ser muchas cosas como restaurante, universidad, bar, veterinario...Esta columna nos da esa información.

- LON: Todas las longitudes de los puntos que delimitan la superficie.
- LAT: Todas las latitudes de los puntos que delimitan la superficie.
- MEANLON: La media del vector LON
- MEANLAT: La media del vector LAT

4.2.2 ETAPA 2 – FUNCIÓN DE DENSIDAD

Como ya se ha comentado antes, en esta etapa se va a ajustar una función de densidad tipo Kernel a los puntos que forman las dimensiones de la anterior etapa. Se ha elegido la función *mvksdensity* de Matlab para generar esta función de densidad, a continuación se explica la línea del código.

```
fd = mvksdensity([x  
y],xi, 'Bandwidth',bw, 'Support',soporte, 'Function', 'pdf', 'Kernel', 'normal');
```

- `fd`: Vector de densidades en los puntos `xi`.
- `[x y]`: Matriz con las longitudes y las latitudes de los puntos a los que se quiere ajustar la función.
- `xi`: Son los puntos del espacio de los cuales se quiere conocer sus densidades. Se ha elegido un grid para seleccionar estos puntos, de esta forma se asegura una repartición equitativa por toda la superficie. En el caso de la CM las dimensiones del grid son 50x50, formando un total de 250000 puntos.
- `'Bandwidth'`: Valor del ancho de banda de la campana de Kernel. Se debe ir ajustando para cada densidad pero con un valor de 0.1 se han visto resultados coherentes en la mayoría de los caso realizados.
- `'Support'`: Matriz de dos filas y dos columnas que indican los límites superiores e inferiores de los ejes x e y, en este caso los límites de la longitud y la latitud de la CM.
- `'Function'`: Función a estimar, en este caso pdf.
- `'Kernel'`: Tipo de Kernel, en este caso normal.

Para entender esta etapa, en la Figura 13

Figura 13 se representa un ejemplo la función de densidad de los habitantes de la Comunidad de Madrid cuyas edades se encuentran entre los 0 y 19 años.

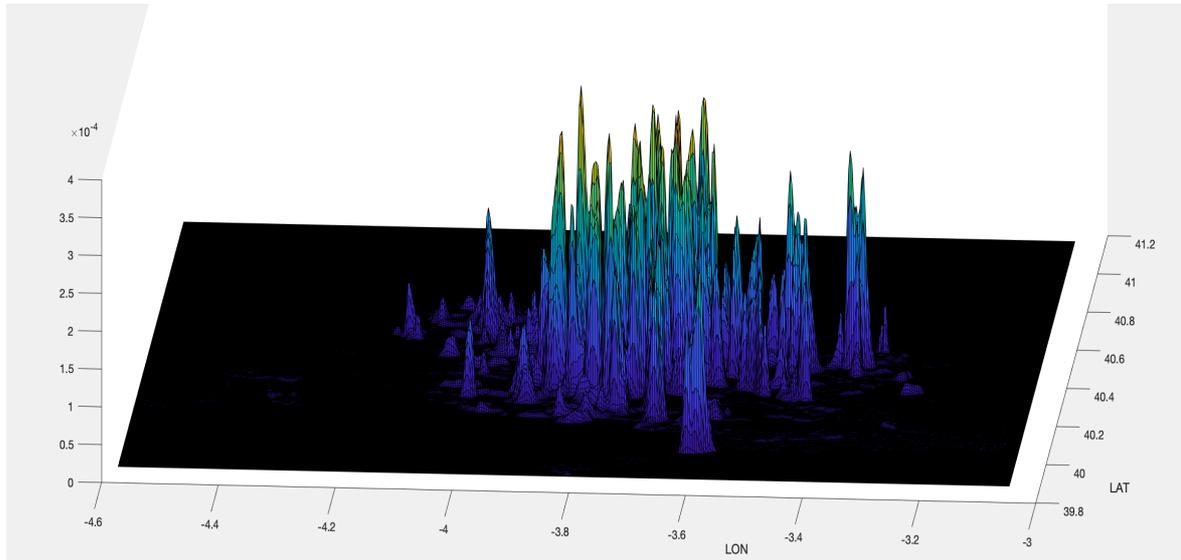


Figura 13: Función de densidad de los habitantes menores de edad de la Comunidad de Madrid

4.2.3 ETAPA 3 – ESTRUCTURACIÓN DE LA INFORMACIÓN

En esta etapa se explica la estructuración de la información que devuelve *mvksdensity* para que pueda ser utilizada por la función *kmean* para generar los clústeres.

Debido a la limitación de memoria y tiempo de ejecución, solo se almacenan 250 000 puntos de toda nuestra superficie. Como ya se ha comentado antes, estos 250 000 puntos son los puntos rojos de la **¡Error! No se encuentra el origen de la referencia.** y el vector ξ de la función *mvksdensity*.

La estructuración es la siguiente. La primera columna es el vector longitudes de la matriz ξ , la segunda columna son las latitudes de ξ y la tercera columna es el vector fd que devuelve la función *mvksdensity*. Al final del todo lo que se obtiene es una estructura de 3 columnas y 250 000 filas por cada dimensión generada. Las dos primeras columnas serán comunes en todas las estructuras y la tercera será la que varíe para cada dimensión.

4.2.4 ETAPA 4 – CLÚSTERES

En esta etapa se generan las agrupaciones en función de las dimensiones seleccionadas, se utiliza la función `kmeans` de Matlab. Se procede a explicar el código.

```
[idx, ctrs, sumd] = kmeans(C, K, 'replicates', 25);
```

- `idx` : vector con el número del clúster al que pertenece cada punto.
- `ctrs`: ubicación del centro del clúster.
- `sumd`: sumas dentro del clúster de distancias punto a centroide.
- `C`: estructura generada en la etapa anterior de la que se quiere generar los clústeres.
- `K`: número de clústeres que se quiere generar.
- `'replicates'`: número de réplicas del proceso de búsqueda que se quiere hacer, en este caso 25.

Capítulo 5. APLICACIÓN

Este apartado se presentan los dos problemas en los que se ha aplicado el modelo diseñado. La primera de ellas es la organización espacial de los barrios de Madrid y la segunda aplicación es el análisis del comportamiento de la pandemia.

Para cada caso se explica las dimensiones utilizadas, así como sus indicadores y las modificaciones que se han realizado en el código para poder llevar a cabo la aplicación.

5.1 CIUDAD DE MADRID

5.1.1 DIMENSIONES

5.1.1.1 Alimentación

Indicadores: supermercados, panaderías, carnicerías, pastelerías, pescaderías, mercados, de conveniencia.

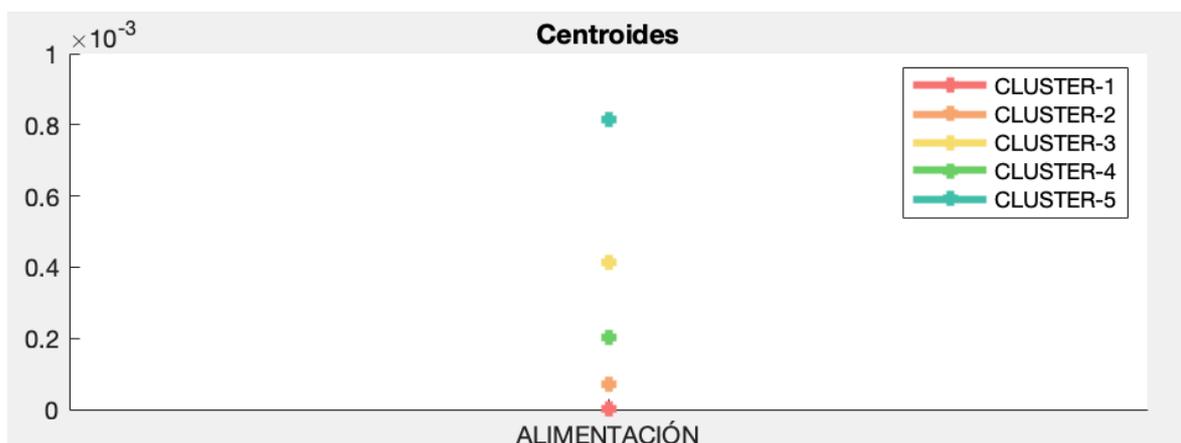


Figura 14 : Centroides de los clústeres de la dimensión Alimentación

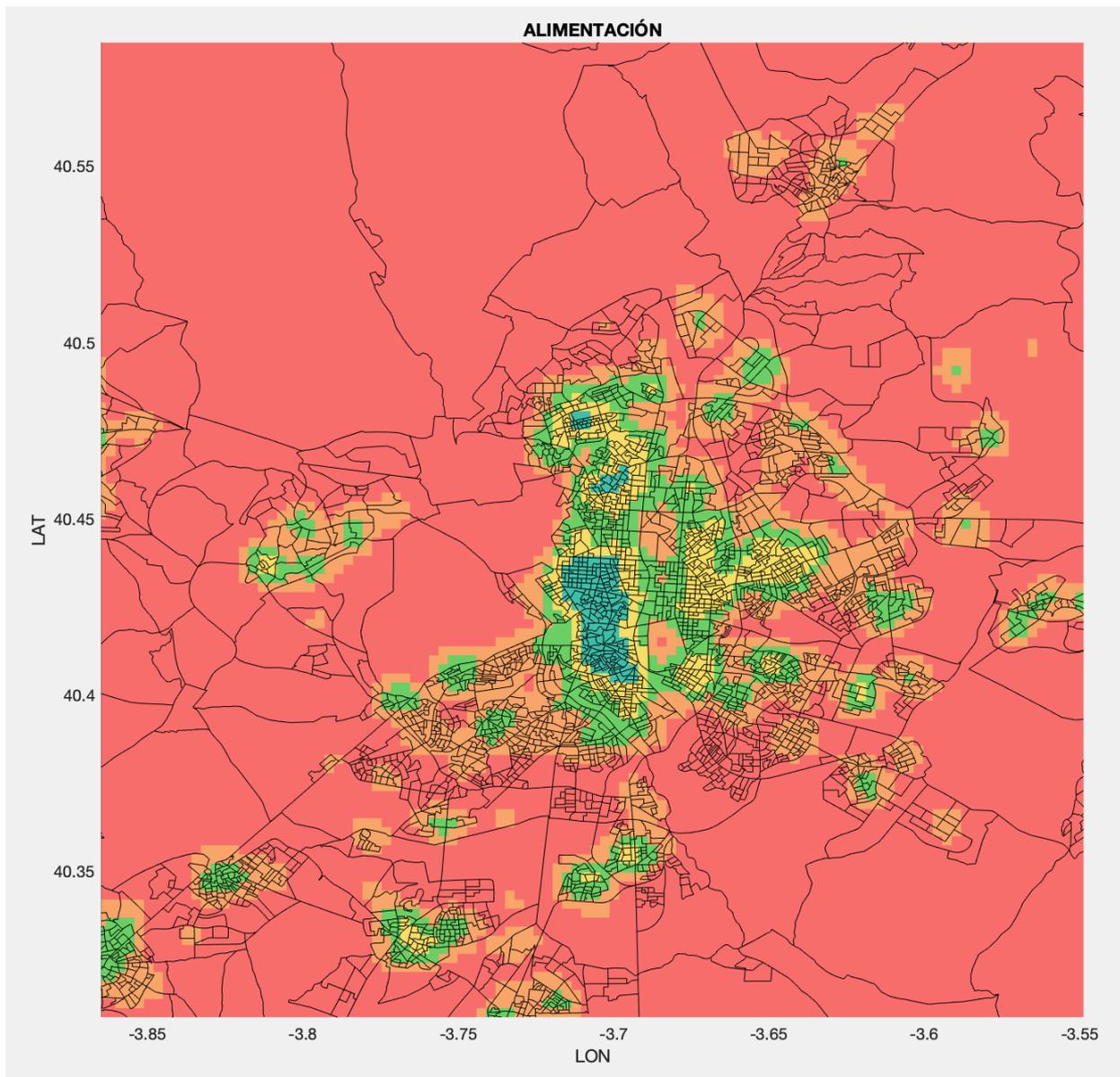


Figura 15 : Representación espacial de los clústeres de la dimensión Alimentación

5.1.1.2 Restaurantes

Indicadores: bares, cafeterías, comida rápida, pubs, restaurantes.

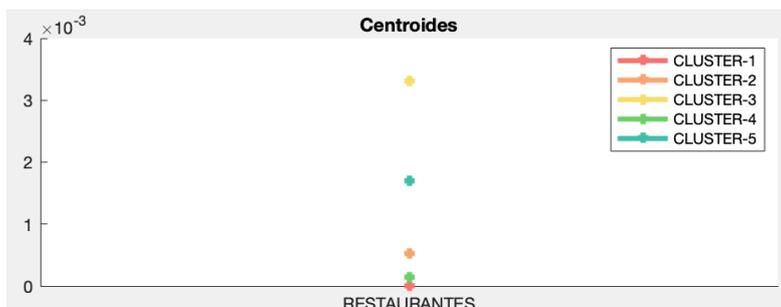


Figura 16: Centroides de los clústeres de la dimensión Restaurantes

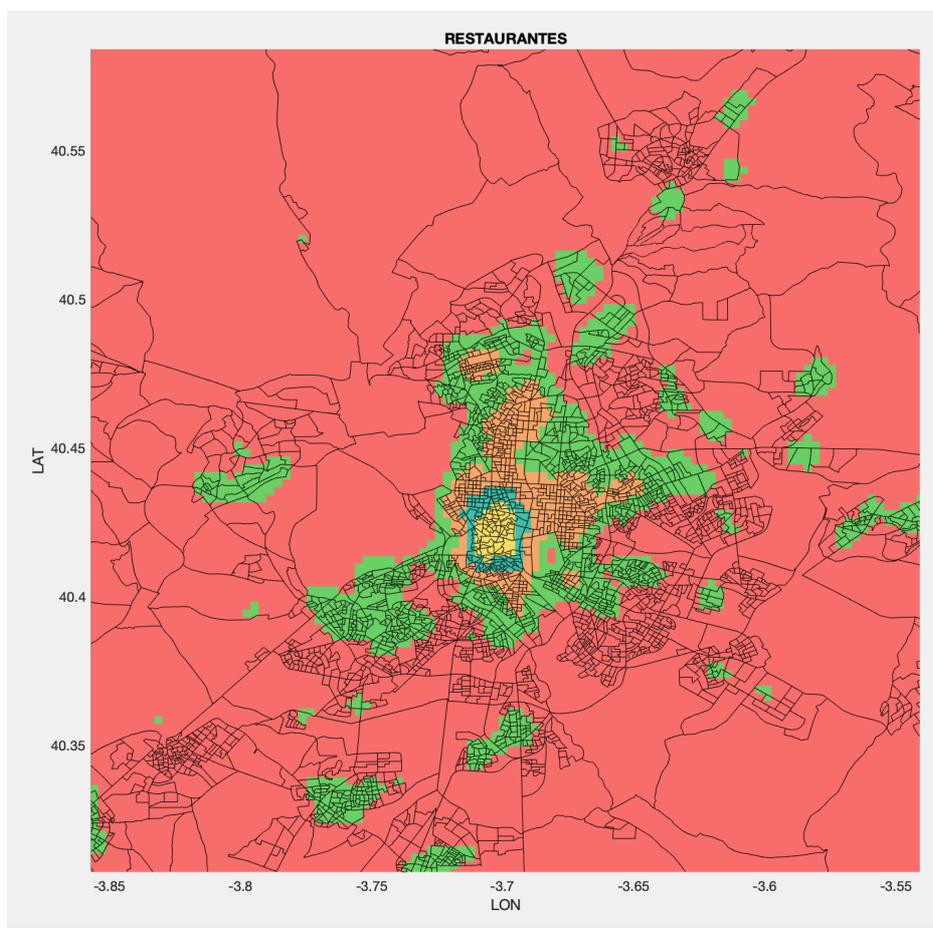


Figura 17: Representación espacial de los clústeres de la dimensión Restaurantes

5.1.1.3 Educación

Indicadores: universidades, guarderías, academias, conservatorios, colegios.

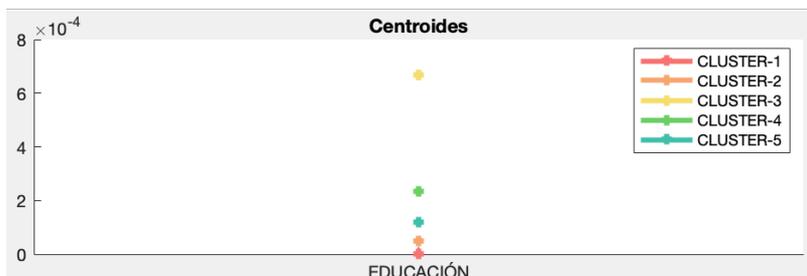


Figura 18: Centroides de los clústeres de la dimensión Educación

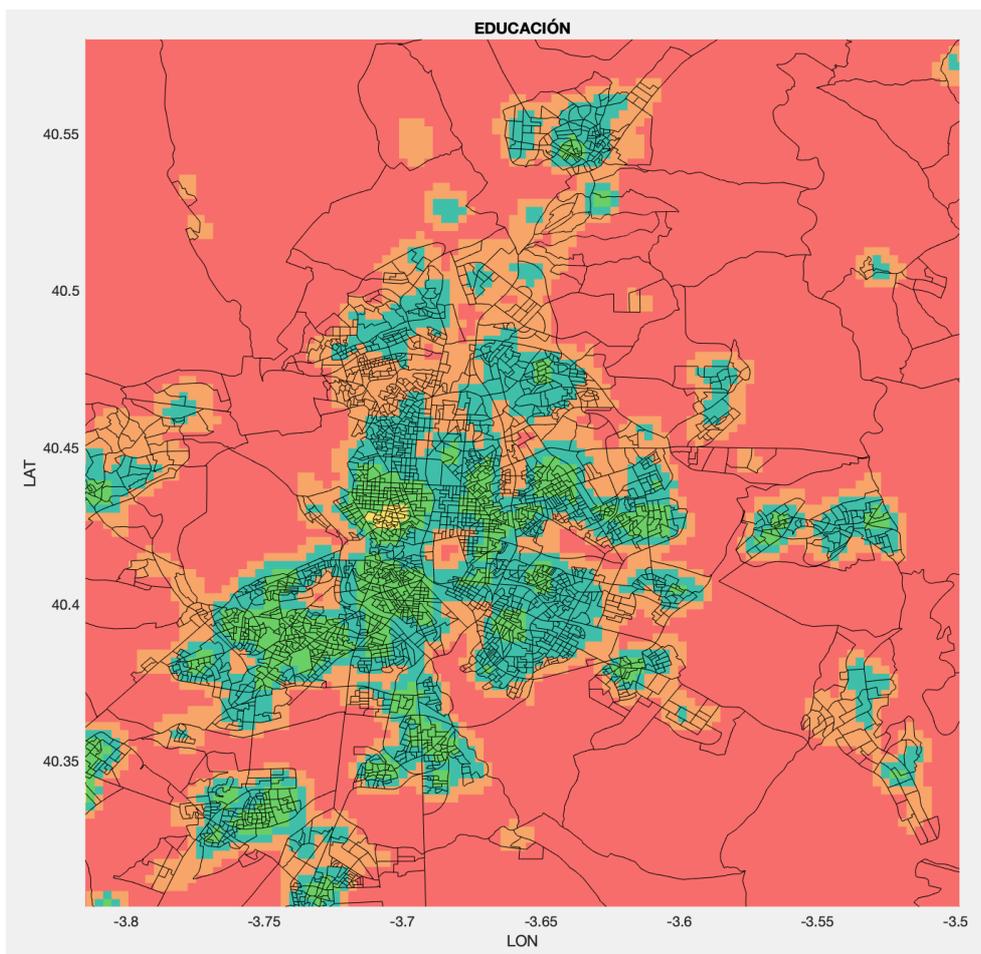


Figura 19: Representación espacial de los clústeres de la dimensión Educación

5.1.1.4 Financiero

Indicador: bancos



Figura 20: Centroides de los clústeres de la dimensión Financiero

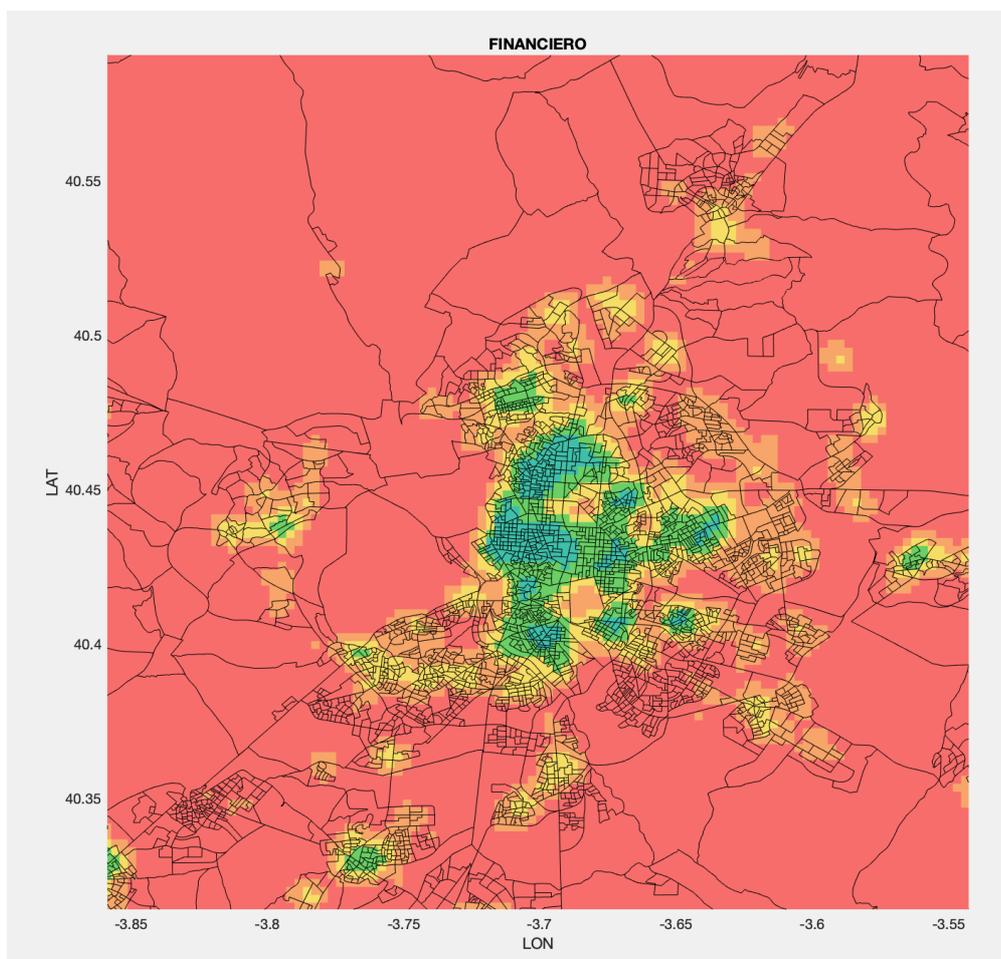


Figura 21: Representación espacial de los clústeres de la dimensión Financiero

5.1.1.5 Sanidad

Indicadores: hospitales, farmacias, consultas, centros de salud.

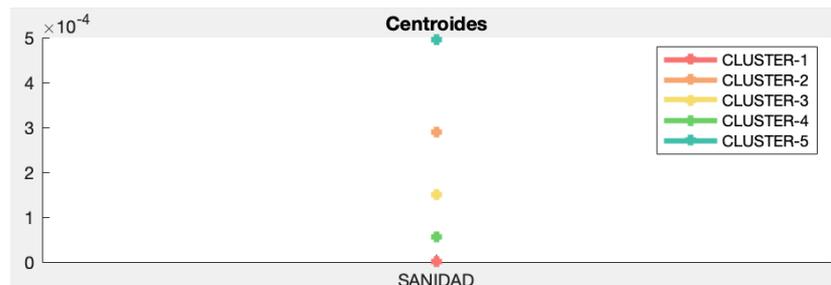


Figura 22: Centroides de los clústeres de la dimensión Sanidad

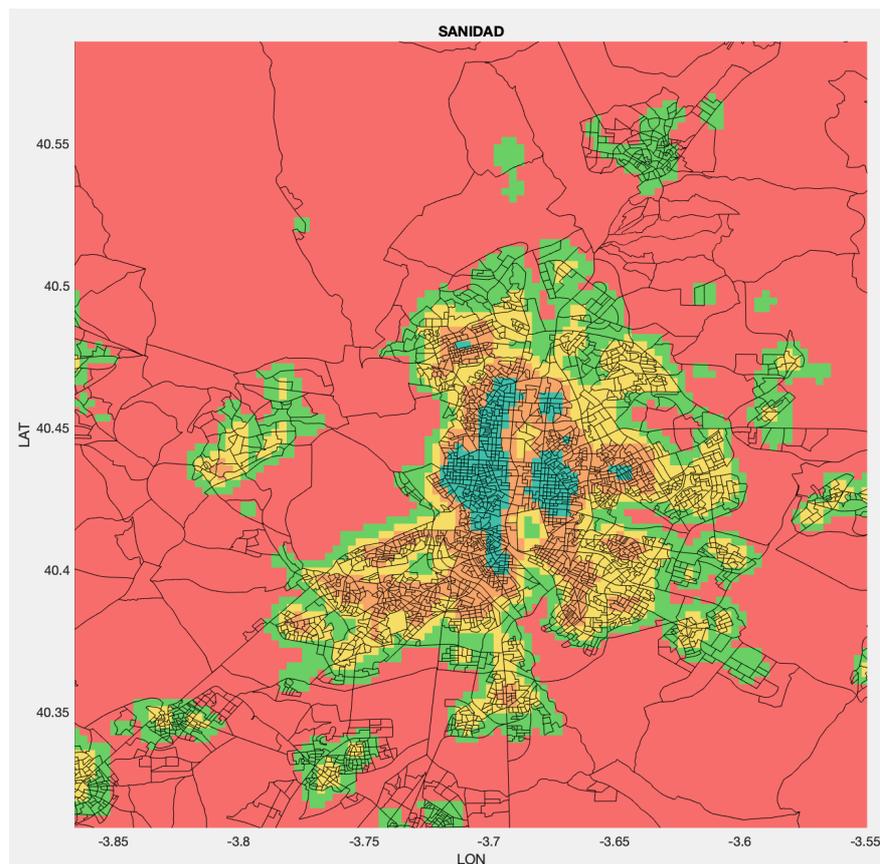


Figura 23: Representación espacial de los clústeres de la dimensión Sanidad

5.1.1.6 Cultura

Indicadores: galerías de arte, cines, centro comunitario, planetarios, teatro, museos.

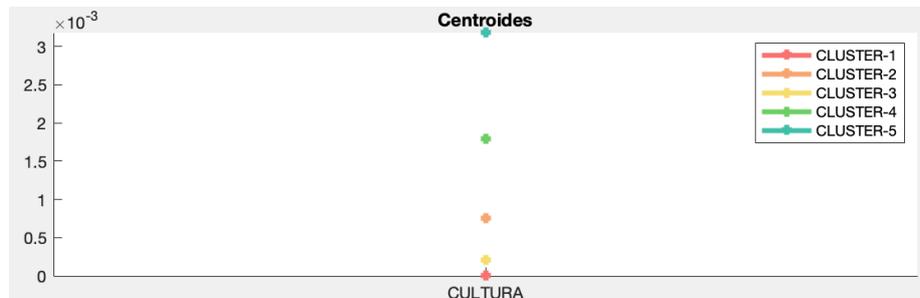


Figura 24: Centroides de los clústeres de la dimensión Cultura

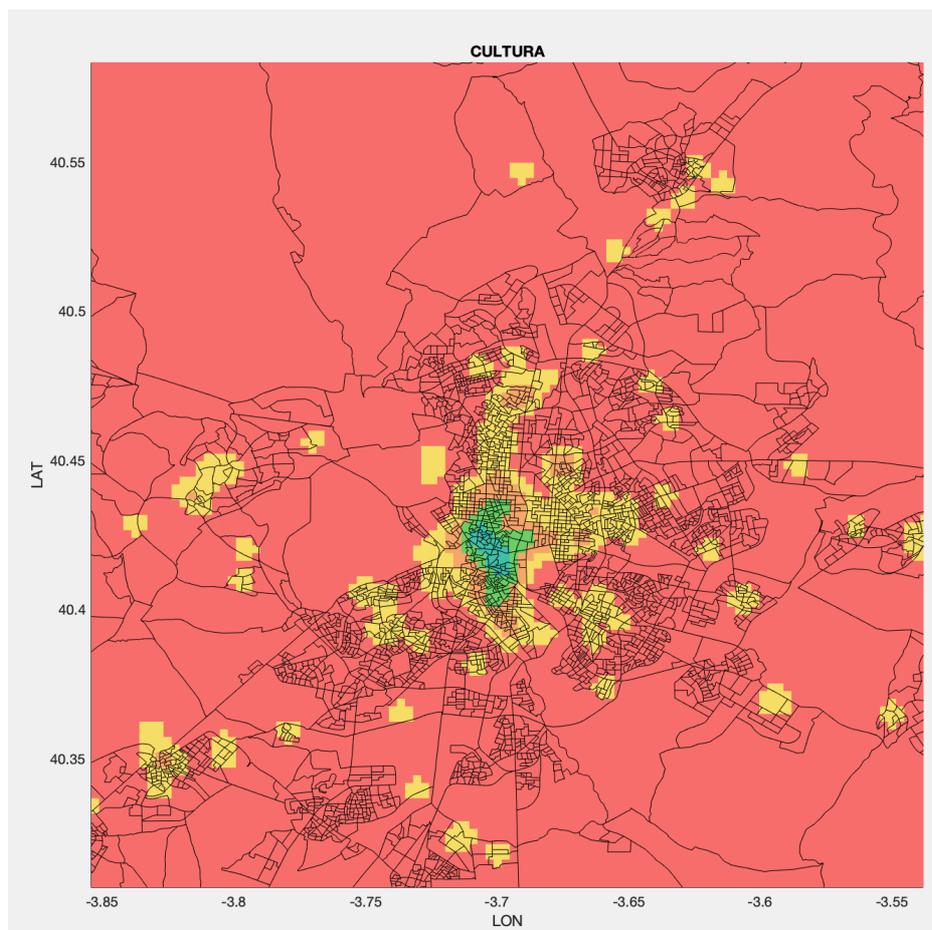


Figura 25: Representación espacial de los clústeres de la dimensión Cultura

5.1.1.7 Ocio

Indicadores: discotecas, bares, restaurantes, teatro, cines, instalaciones donde se practica deporte, casino, galerías de arte, conciertos.

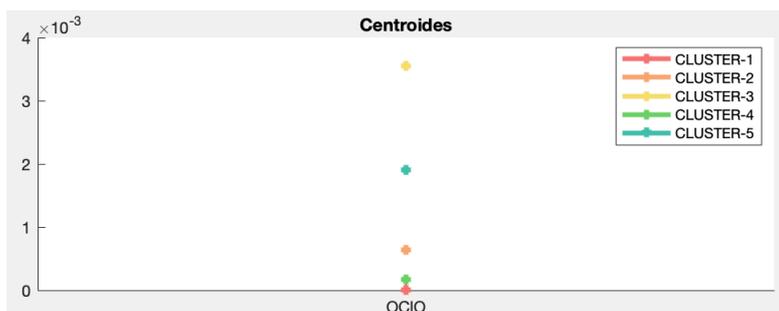


Figura 26: Centroides de los clústeres de la dimensión Ocio



Figura 27: Representación espacial de los clústeres de la dimensión Ocio

5.1.1.8 Tiendas

Indicadores: centros comerciales, ropa, zapatos, accesorios, cosméticas y belleza, herramientas, muebles y decoraciones, electrónicas, deporte, arte, música, hobbies, regalos, libros, plantas y animales.

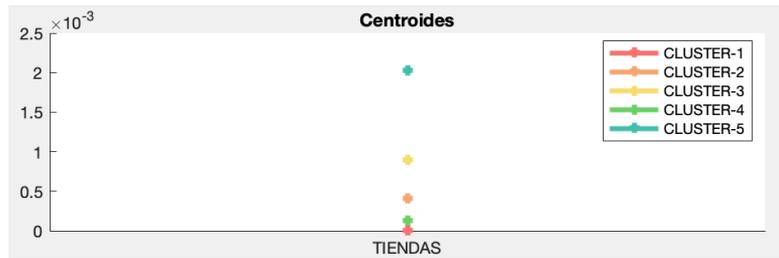


Figura 28: Centroides de los clústeres de la dimensión Tiendas

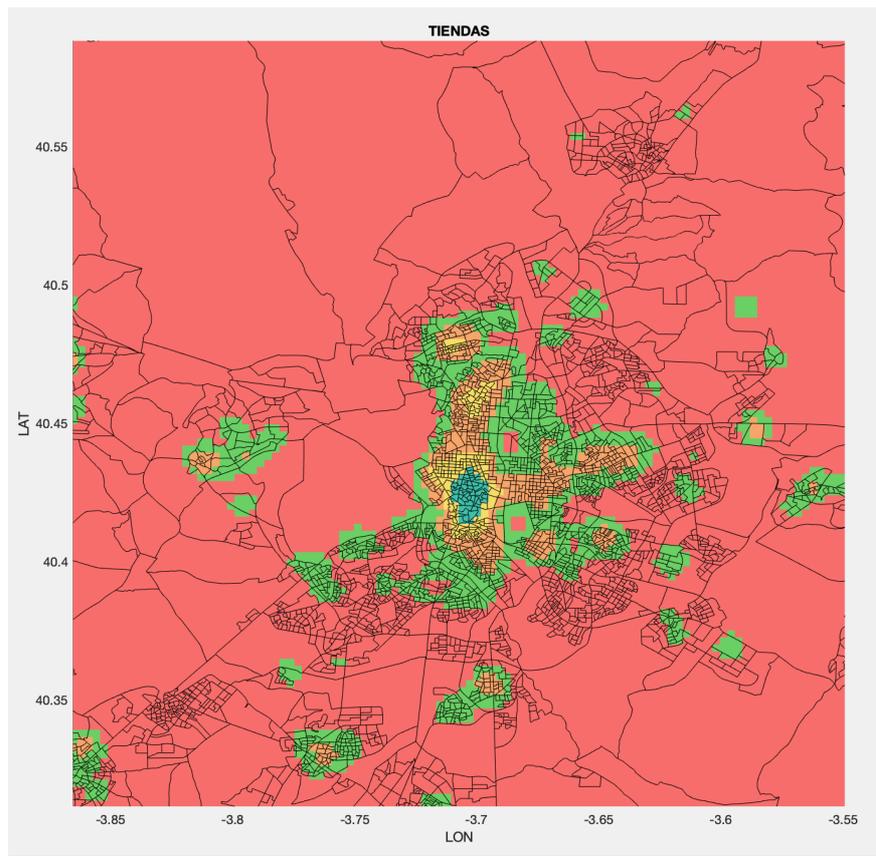


Figura 29: Representación de los clústeres de la dimensión Tiendas

5.1.1.9 Población

Se ha dividido la población en tres dimensiones: jóvenes, adultos, mayores.

5.1.1.9.1 Jóvenes

Indicadores: Población entre los 0-19 años.

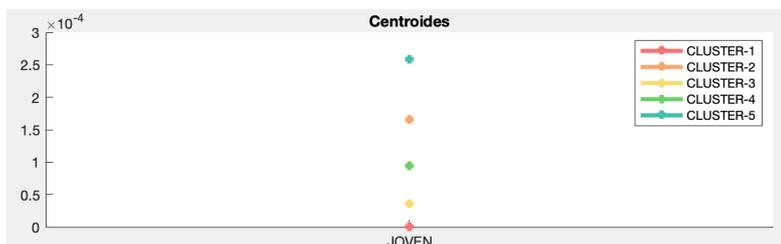


Figura 30: Centroides de los clústeres de la dimensión Jóvenes

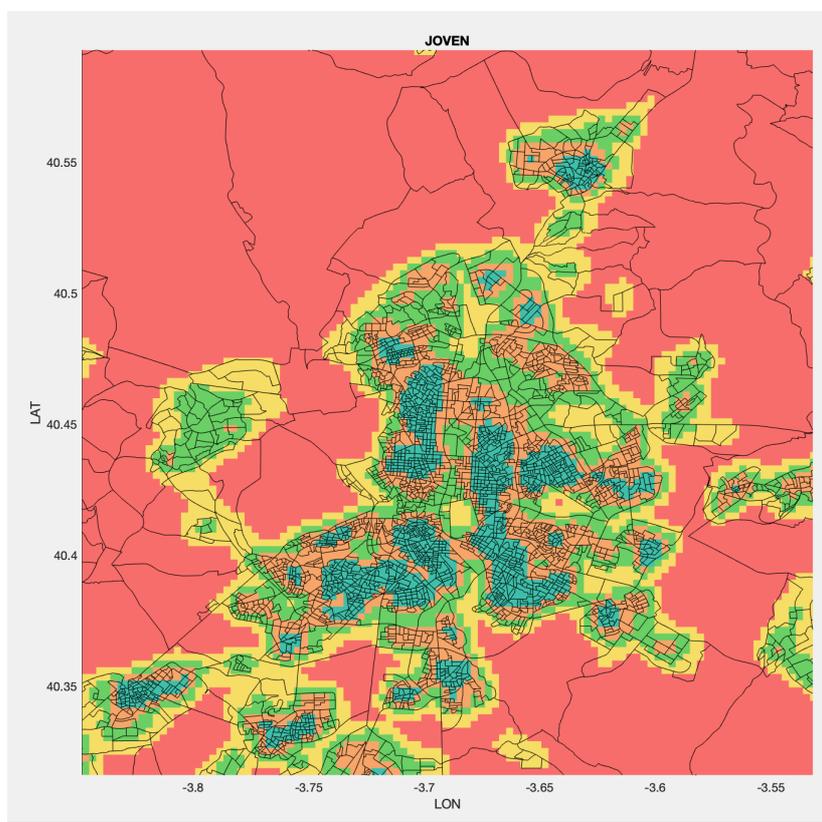


Figura 31: Representación espacial de los clústeres de la dimensión Jóvenes

5.1.1.9.2 Adultos

Indicadores: Población entre los 20-64.

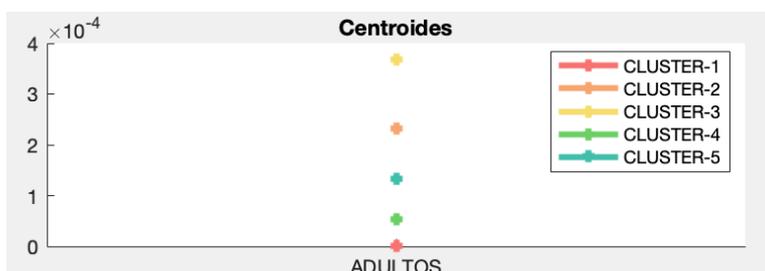


Figura 32: Centroides de los clústeres de la dimensión Adulto

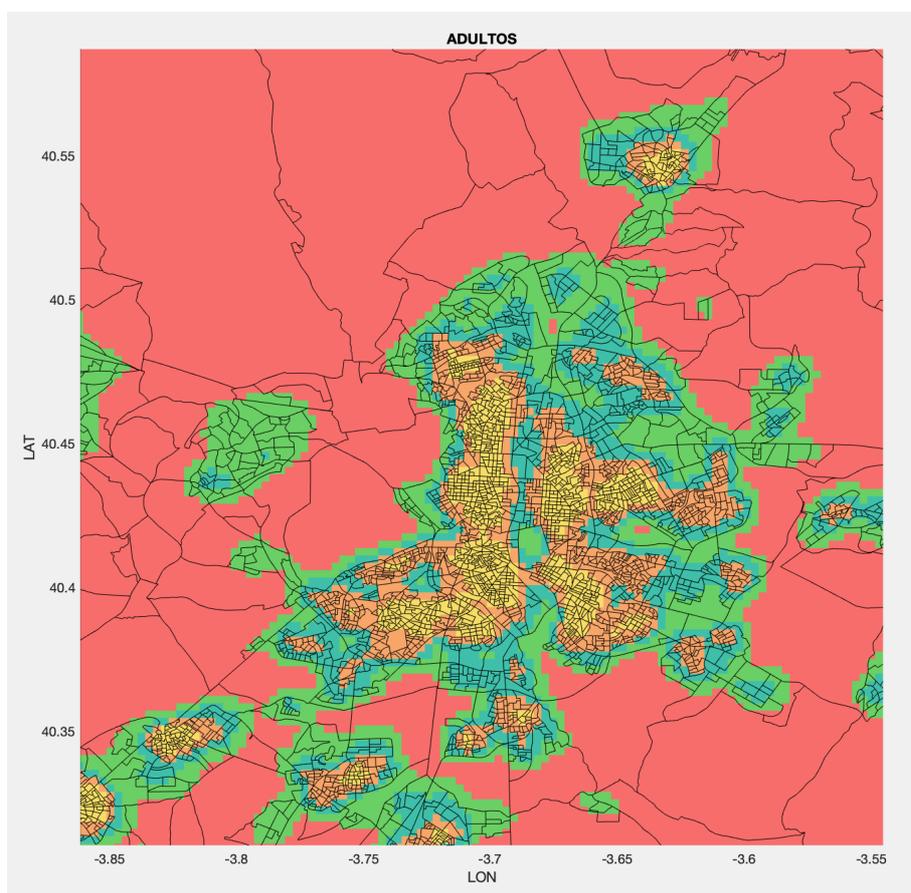


Figura 33: Representación espacial de los clústeres de la dimensión Adulto

5.1.1.9.3 Mayores

Indicadores: Población entre los 65 en adelante.

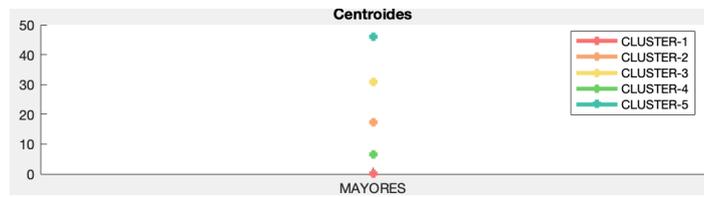


Figura 34: Centroides de los clústeres de la dimensión Mayores

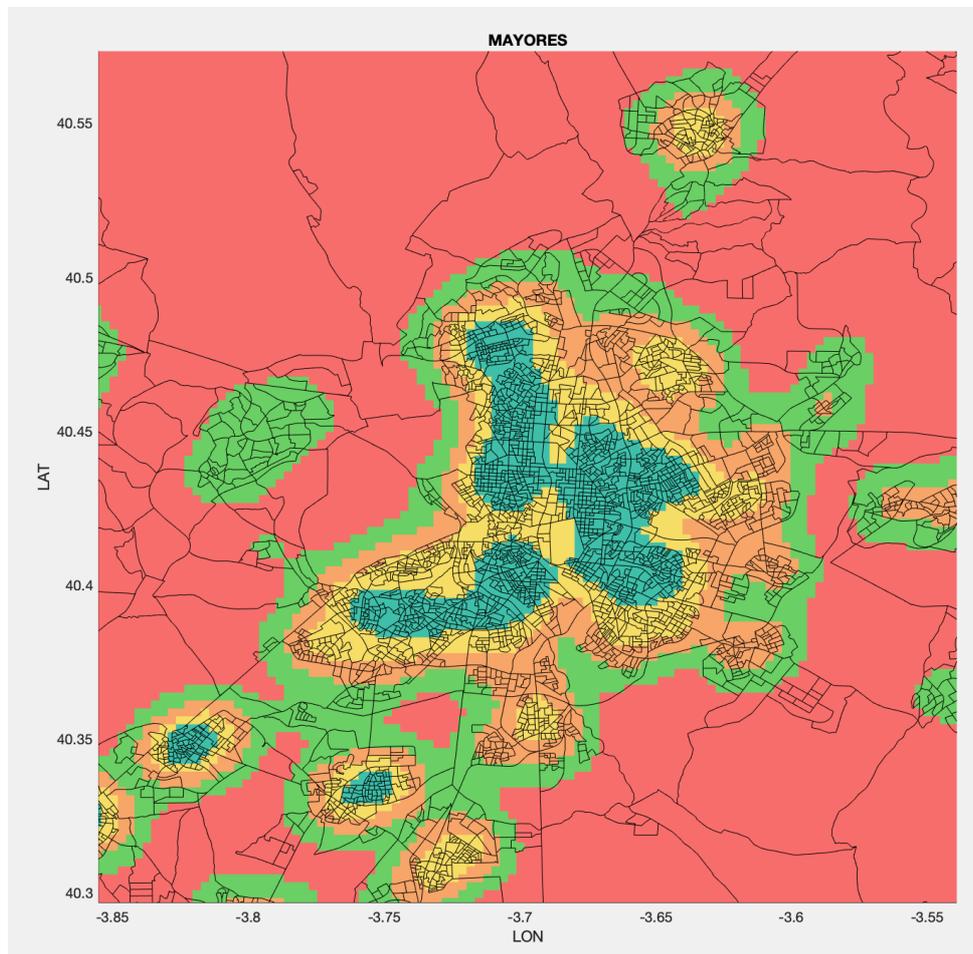


Figura 35: Representación espacial de la dimensión Mayores

5.1.1.10 Renta Media

Esta información, al contrario que el resto, no se ha generado su función de densidad, se ha trabajado directamente con el archivo descargado del INE.

Para introducir esta información en el modelo se ha hecho lo siguiente. Se parte de los 250000 puntos del grid repartidos por nuestra zona de estudio, se ha buscado la SC a la que pertenece cada uno y se le ha asignado a ese punto la renta media de su SC. En el caso de no tener la información de la renta media de su SC se le ha asignado un valor nulo (NaN) y sale de color blanco en las figuras.

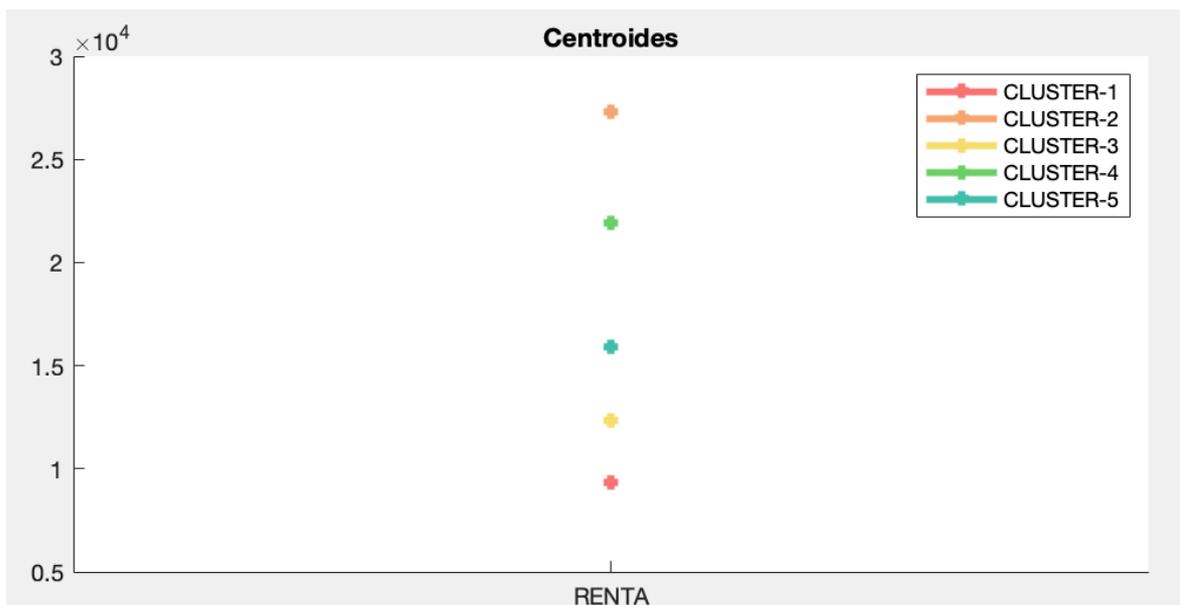


Figura 36: Centroides de los clústeres de la dimensión Renta Media

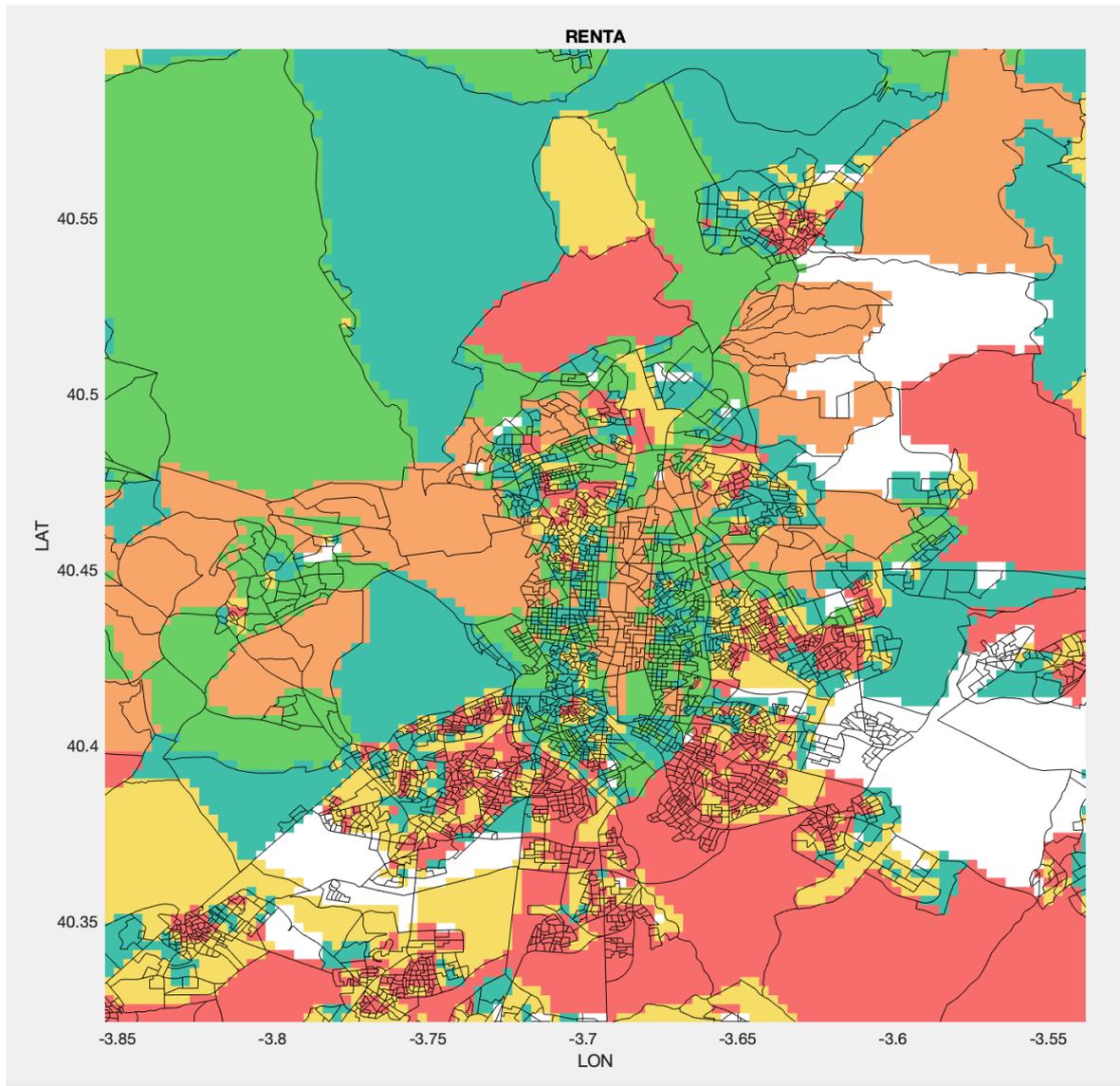


Figura 37: Representación gráfica de los clústeres de la dimensión Renta Media

5.2 COVID - 19

Para el análisis de la pandemia se ha buscado relaciones entre el número de contagiados, la organización de los servicios de sanidad y la población.

A continuación se explica de dónde se han descargado los datos de los contagiados y el procedimiento que se ha seguido para introducirlos en el modelo.

5.2.1 NÚMERO DE CONTAGIADOS

Para obtener estos datos hay que seguir los siguientes pasos: web de la comunidad de Madrid → Datos abiertos → Salud → COVID19 → Tia por Municipio y Distritos de Madrid. [8]



The screenshot shows the 'Datos Abiertos' (Open Data) portal for the Community of Madrid. The URL in the browser is datos.comunidad.madrid/catalogo/dataset/covid19_tia_muni_y_distritos. The page features the 'Datos Abiertos' logo, navigation links for 'Catálogo', 'Transparencia', and 'Participación', and the 'Comunidad de Madrid' logo. The breadcrumb trail is: / Organizaciones / Comunidad de Madrid / Covid 19 -TIA por ...

On the left, there are options to 'Compartir en redes' (Twitter, Facebook, LinkedIn) and 'Licencia' (Creative Commons Attribution, with 'OPEN DATA' and 'DATA' buttons).

The main content area displays the dataset title 'Covid 19 -TIA por Municipios y Distritos de Madrid' and provides a description: 'Información epidemiológica Covid-19: Casos confirmados y tasa de incidencia acumulada (TIA) por municipios y distritos de Madrid. Informe diario con datos de casos confirmados y tasas de incidencia acumulada de los últimos 14 días y desde el inicio de la epidemia (25/02/2020). La fuente de los datos es la Red de Vigilancia Epidemiológica de la Comunidad de Madrid. Cuando se utilicen los datos se debe indicar en qué fecha se ha accedido a los mismos, dada su actualización continua. El conjuntos de datos añade diariamente los registros de datos correspondientes a cada fecha de informe.'

Under 'Datos y Recursos', three data files are listed, each with an 'Explorar' button:

- Covid 19 -TIA por Municipios y Distritos de Madrid** (CSV format)
- Covid 19 -TIA por Municipios y Distritos de Madrid** (JSON format)
- municipios_y_distritos_madrid.zip** (Territorial division in municipalities of the Community of Madrid and districts del...)

Figura 38: Web Comunidad de Madrid con las información de los contagiados del COVID19

La información descargada (se ha trabajado con los contagiados registrados hasta el día 16-abril-2020) indica el número de infectados en cada municipio y en el caso del de Madrid por cada distritos (un municipio es un conjunto de distritos). De este archivo se ha trabajado con las 200 primeras filas y con las columnas: 1 (nombre de los municipios-distritos) y 6 (número total de contagiados). Esta información se almacena en una estructura llamada Ctg.

Esta información tiene el mismo problema que tiene la información de los habitantes del INE, necesita de una hipótesis para que pueda ser introducida en el modelo ya que no puede ser representada en el plano longitud-latitud. Debido a las similitudes de esta información con los datos del número de habitantes, el procedimiento va a ser muy parecido al que se realizó con esta.

A continuación se explica el proceso que se ha seguido para poder introducir esta información en el modelo.

El primer problema que se debe resolver es la acotación de los municipios y distritos, para resolverlo, hay que buscar las SC que los forman. Se ha comparado el nombre que aparece en la primera columna de Ctg con el nombre que aparece en la columna NMUN (de la estructura guardada de las SC), en el caso de los distritos de Madrid se ha comparado con la columna CDIS. Identificadas las SC que forman los municipios y distritos, se conocen los puntos que las delimitan y por tanto se resuelve el problema de la acotación.

Una vez se ha encontrado todas las SC que forman un municipio hay que generar en esta superficie el número de contagiados que indica la estructura Ctg. Para ello se divide el número de contagiado entre el número de SC y se genera por cada SC el cociente de esta división. El resto de la división se guarda para ser representado junto al cociente en la última SC del grupo que forman el municipio.

Cuando se haya realizado este procedimiento con todos los municipio y distritos, se almacena en una estructura el nombre de los municipios con el vector longitudes y el vector latitudes de todos los contagiados. Esta estructura nos permite representar los contagiados en el plano longitud-latitud y de esta forma introducir la información del COVID-19 en el modelo.

A continuación se puede ver en la Figura 39 la función de densidad de los contagiados en la Comunidad de Madrid.

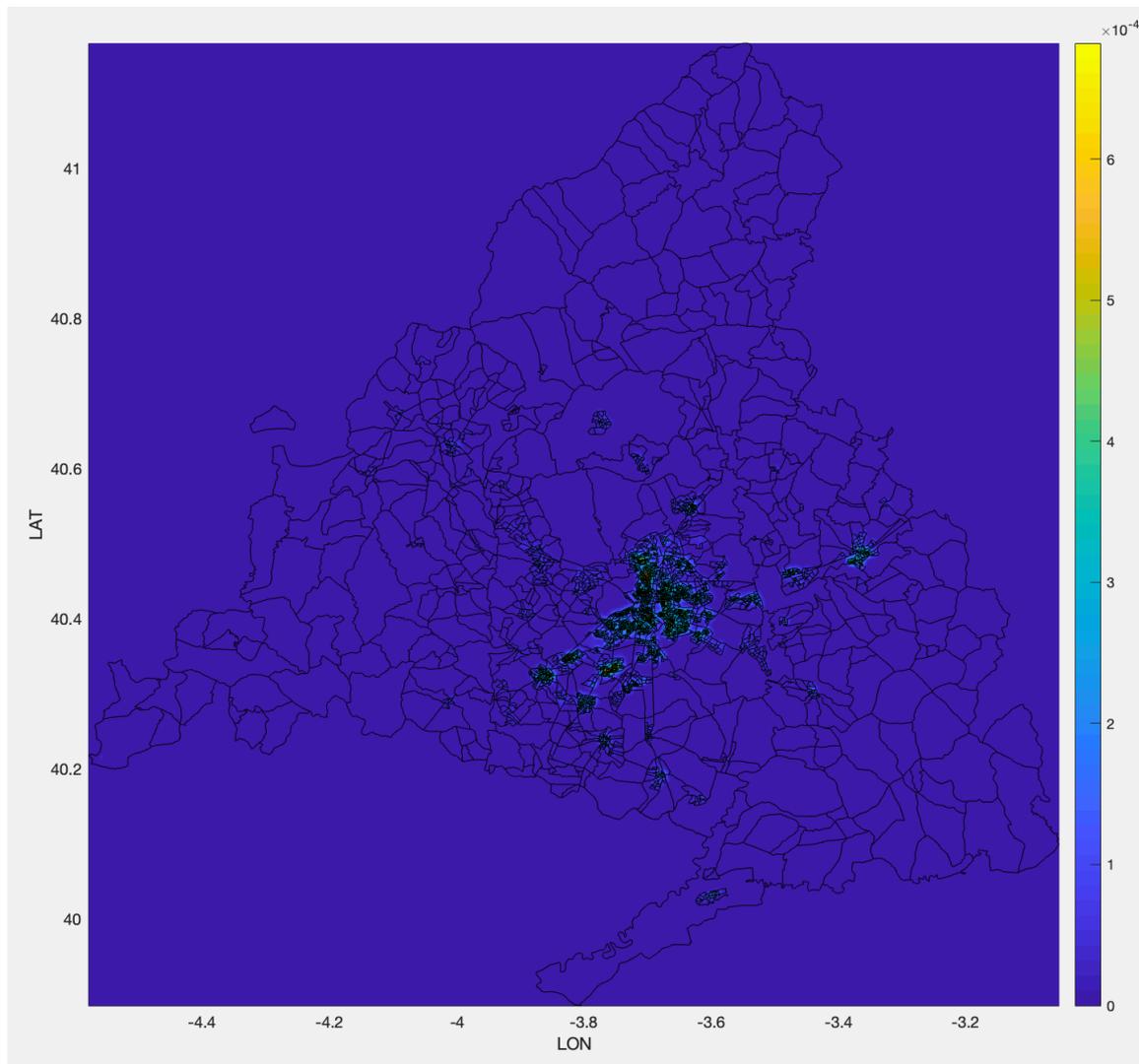


Figura 39: Densidad de contagiados en la Comunidad de Madrid

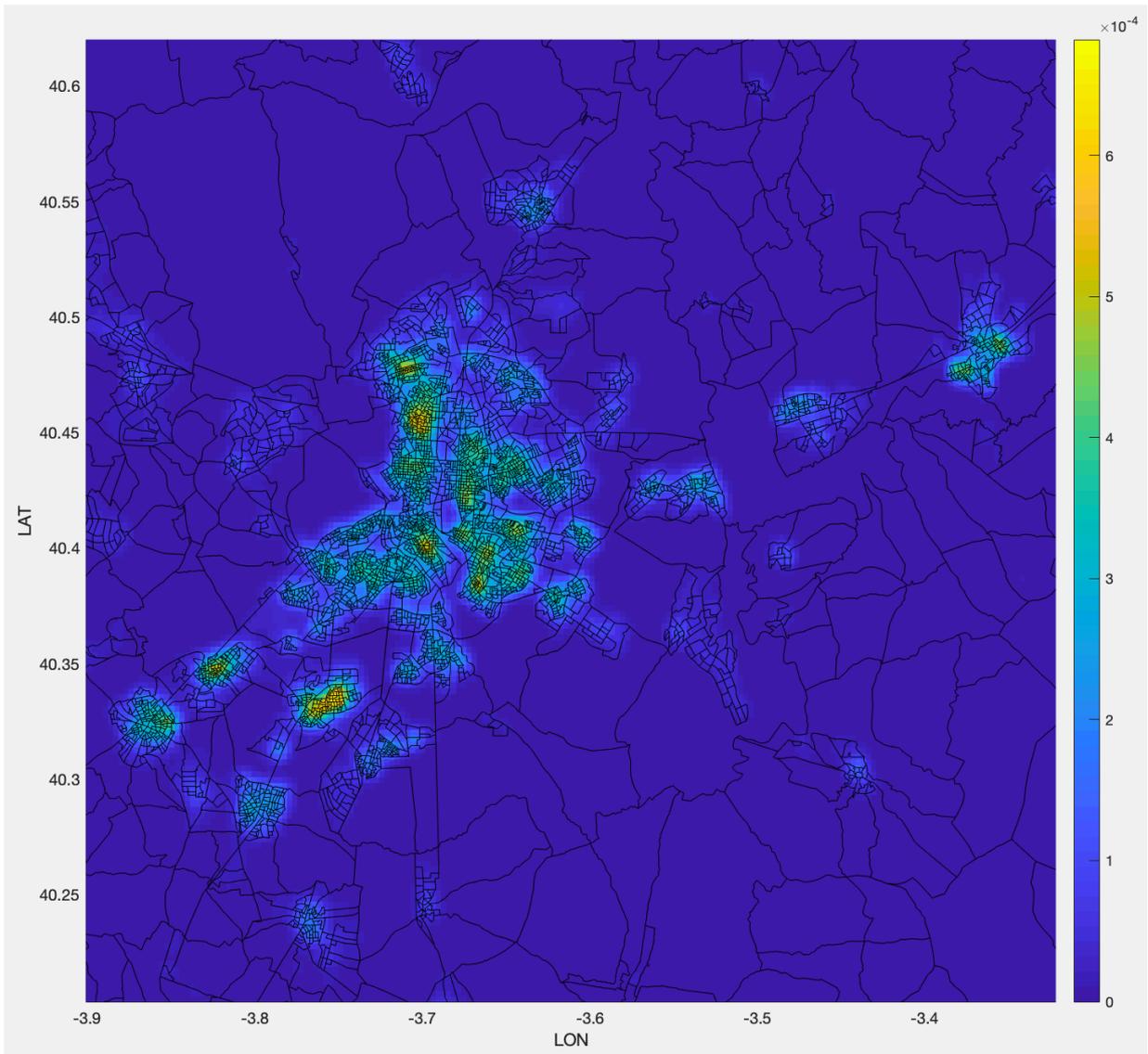


Figura 40 : Zoom de Figura 39

5.2.2 ZONAS BÁSICAS DE SALUD (ZBS)

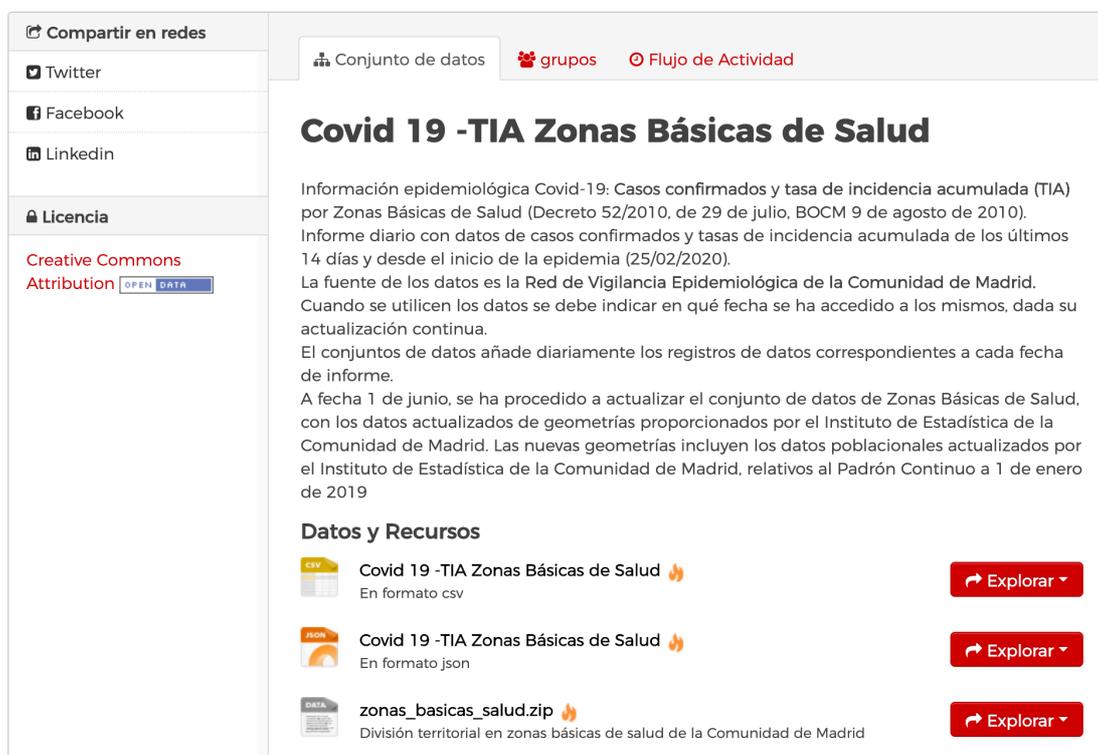
Para poder dar una solución cuantitativa al problema de los contagiados, se ha decidido evaluar la densidad de contagiados en las Zonas Básicas de Salud (ZBS) en vez de las SC.

A continuación se explica la descarga y almacenamiento de las ZBS así como las soluciones obtenidas.

Como uno de los grandes problemas de esta pandemia es el colapso de la red sanitaria, se ha considerado más útil evaluar el número de contagiados en las ZBS. Estas zonas son la mínima división del terreno para optimizar la asistencia médica a los ciudadanos, al conjunto de ZBS (286 en total) se le llama Distrito Sanitario y al conjunto de distritos se le llama Área de Salud (15 en total) [9].

Para obtener estos datos hay que seguir los siguientes pasos: web de la comunidad de Madrid → Datos abiertos → Salud → COVID19 → TIA Zonas Básicas de Salud

🏠 / Organizaciones / Comunidad de Madrid / Covid 19 -TIA Zonas ...



Compartir en redes

- Twitter
- Facebook
- LinkedIn

Licencia

Creative Commons Attribution [OPEN DATA](#)

Conjunto de datos grupos Flujo de Actividad

Covid 19 -TIA Zonas Básicas de Salud

Información epidemiológica Covid-19: Casos confirmados y tasa de incidencia acumulada (TIA) por Zonas Básicas de Salud (Decreto 52/2010, de 29 de julio, BOCM 9 de agosto de 2010). Informe diario con datos de casos confirmados y tasas de incidencia acumulada de los últimos 14 días y desde el inicio de la epidemia (25/02/2020). La fuente de los datos es la Red de Vigilancia Epidemiológica de la Comunidad de Madrid. Cuando se utilicen los datos se debe indicar en qué fecha se ha accedido a los mismos, dada su actualización continua. El conjuntos de datos añade diariamente los registros de datos correspondientes a cada fecha de informe. A fecha 1 de junio, se ha procedido a actualizar el conjunto de datos de Zonas Básicas de Salud, con los datos actualizados de geometrías proporcionados por el Instituto de Estadística de la Comunidad de Madrid. Las nuevas geometrías incluyen los datos poblacionales actualizados por el Instituto de Estadística de la Comunidad de Madrid, relativos al Padrón Continuo a 1 de enero de 2019

Datos y Recursos

	Covid 19 -TIA Zonas Básicas de Salud 🔥 En formato csv	Explorar
	Covid 19 -TIA Zonas Básicas de Salud 🔥 En formato json	Explorar
	zonas_basicas_salud.zip 🔥 División territorial en zonas básicas de salud de la Comunidad de Madrid	Explorar

Figura 41: Web Comunidad de Madrid con las información de las ZBS

De la carpeta *zonas_basicas_salud.zip* descargada, el archivo relevante para este proyecto es el archivo *zonas_basicas_salud.shp*. Tras la lectura de este archivo se almacena la estructura que tiene por filas las ZBS y por columnas los siguientes atributos:

- X: Vector con las coordenadas longitud de los puntos que delimitan la ZBS.
- Y: Vector con las coordenadas latitud de los puntos que delimitan la ZBS.
- DESBDT: Nombre de la ZBS.
- CODIGO GEO: Código de identificación de la ZBS.

En la siguiente figura se puede ver representadas en el plano longitud-latitud las ZBS de la Comunidad de Madrid.

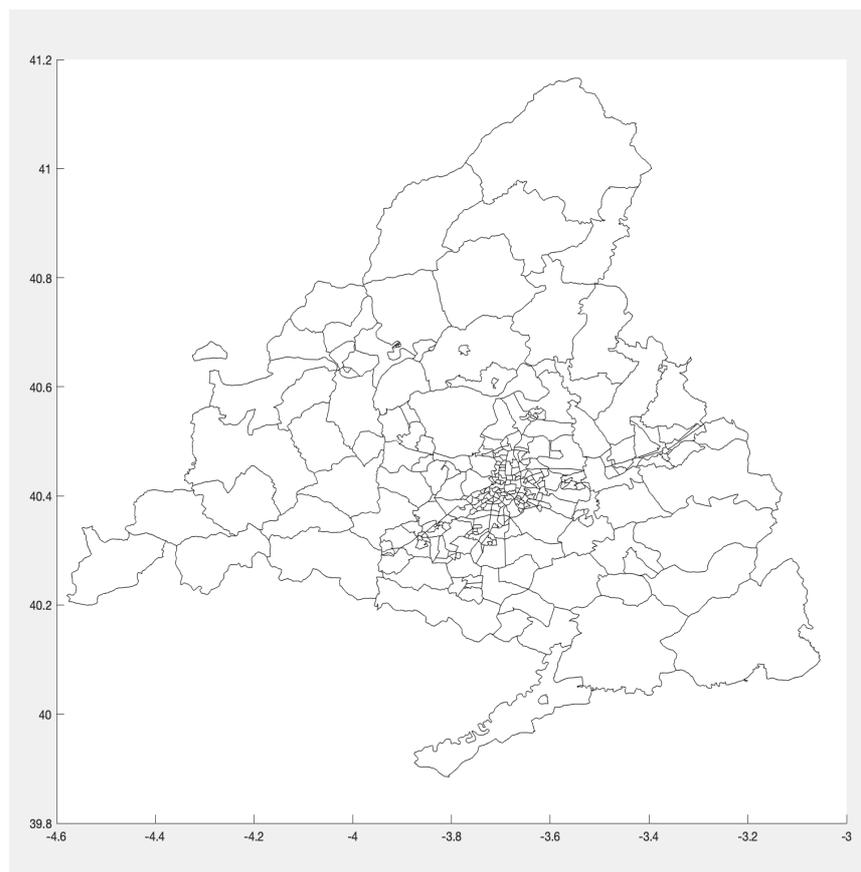


Figura 42: Zonas Básicas de Salud de la Comunidad de Madrid

Para que se vea la diferencia de división territorial de las ZBS y de las SC se presenta la Figura 43. En negro se puede ver las ZBS y en rojo las SC.

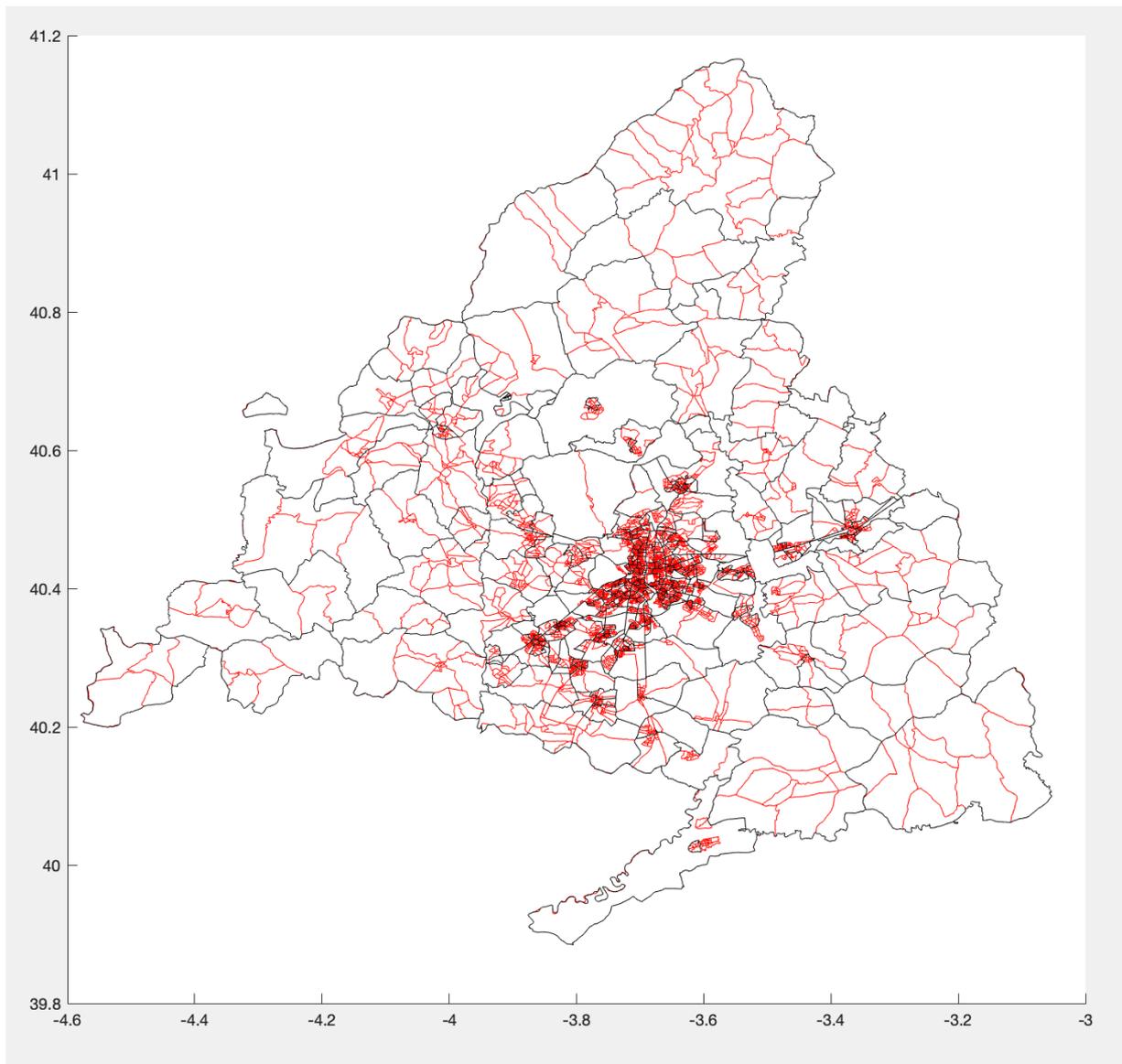


Figura 43: Comparación ZBS (negro) con las SC (rojo)

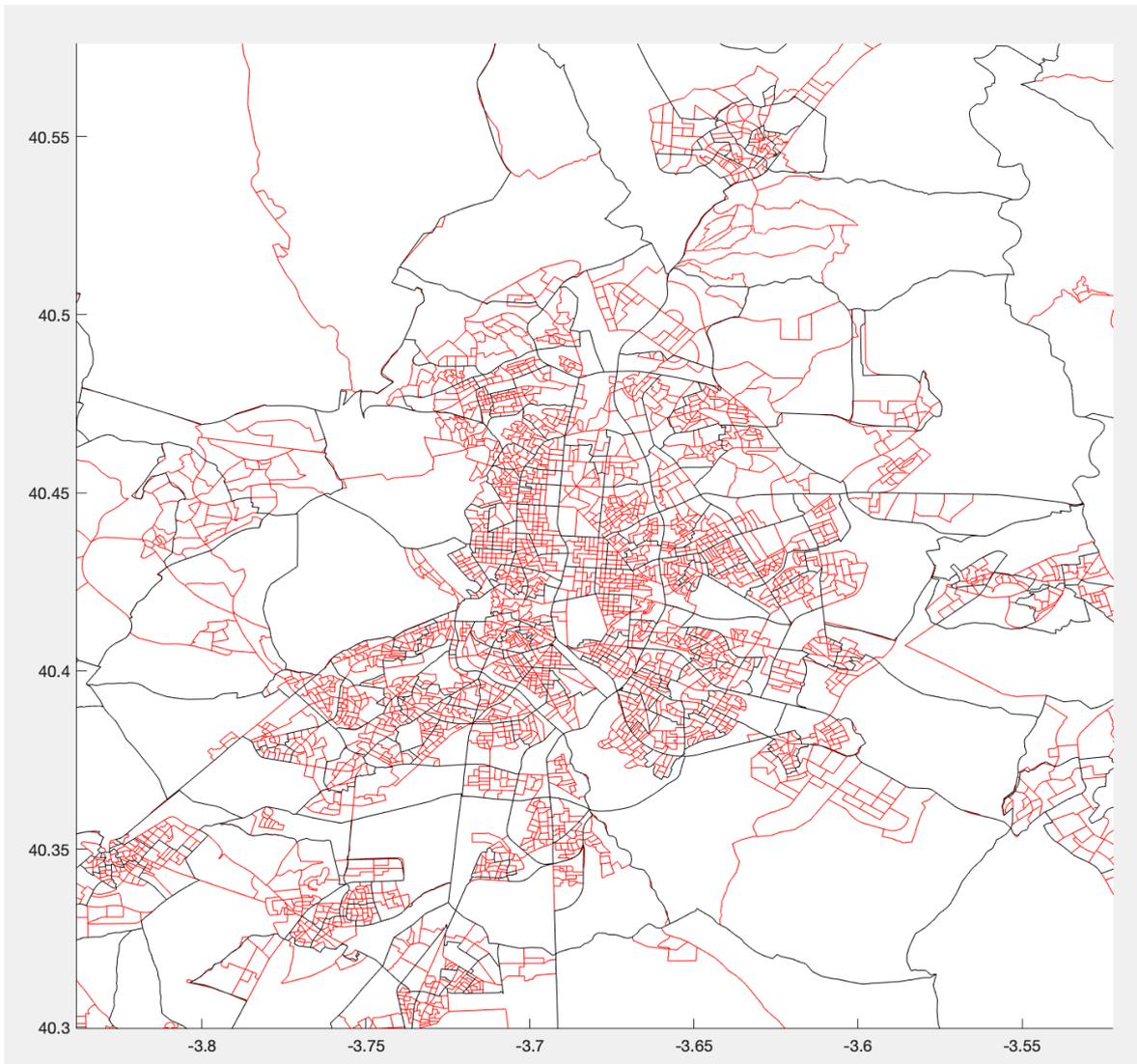


Figura 44: Zoom de la Figura 43

Ahora se evalúa la función de densidad generada de los contagiados (Figura 39) sobre las ZBS. Para ello, se ha realizado la media de los valores de las densidades por cada ZBS y se le ha asignado a toda esa zona el valor de la media.

En la Figura 45 se puede ver las densidades de los contagiados por cada ZBS.

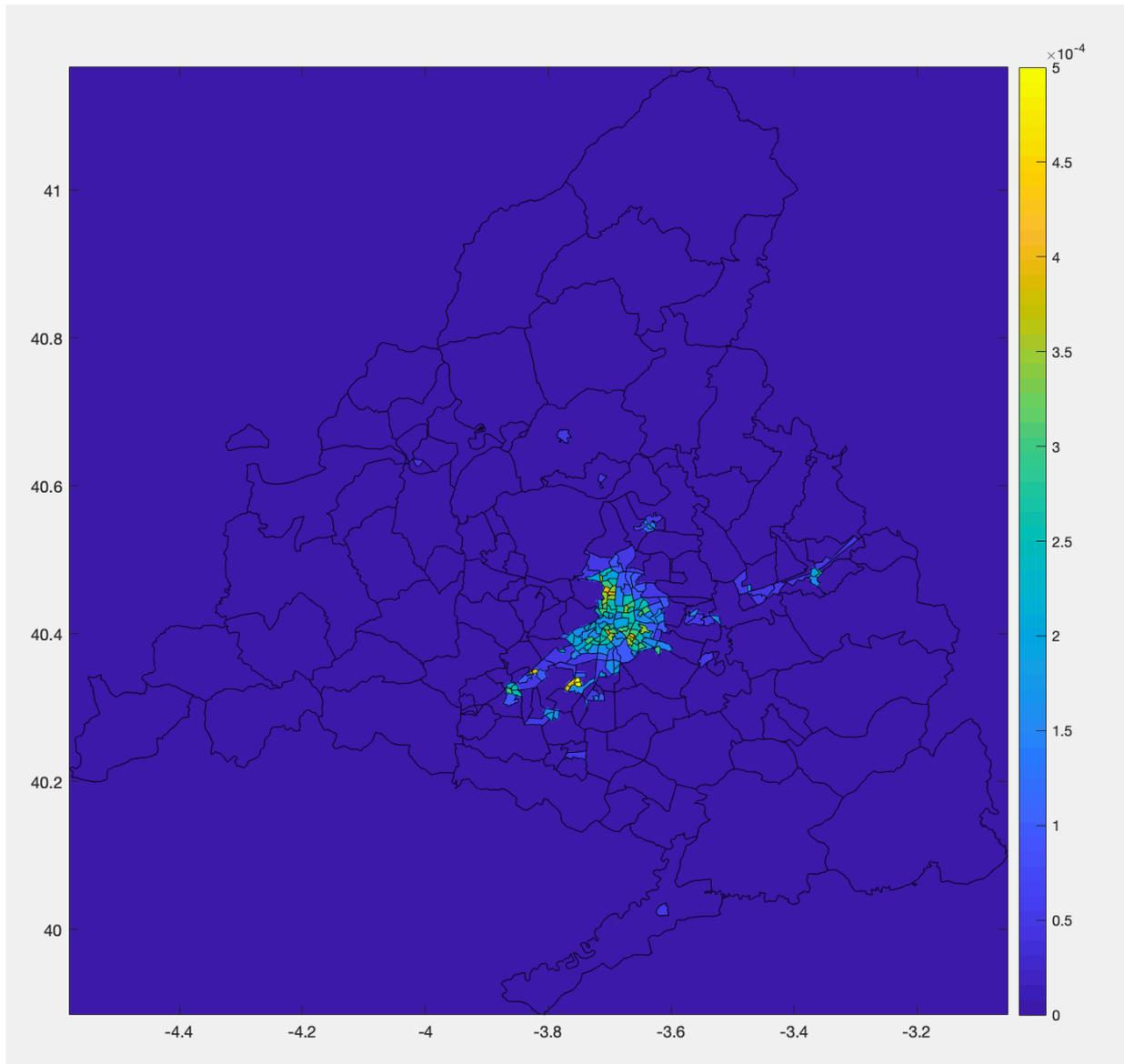


Figura 45: Densidad de contagiados por cada ZBS

Para ver que zonas son las más afectadas se realiza un gráfico de barras y se escoge las 10 ZBS con más densidad de contagiados.

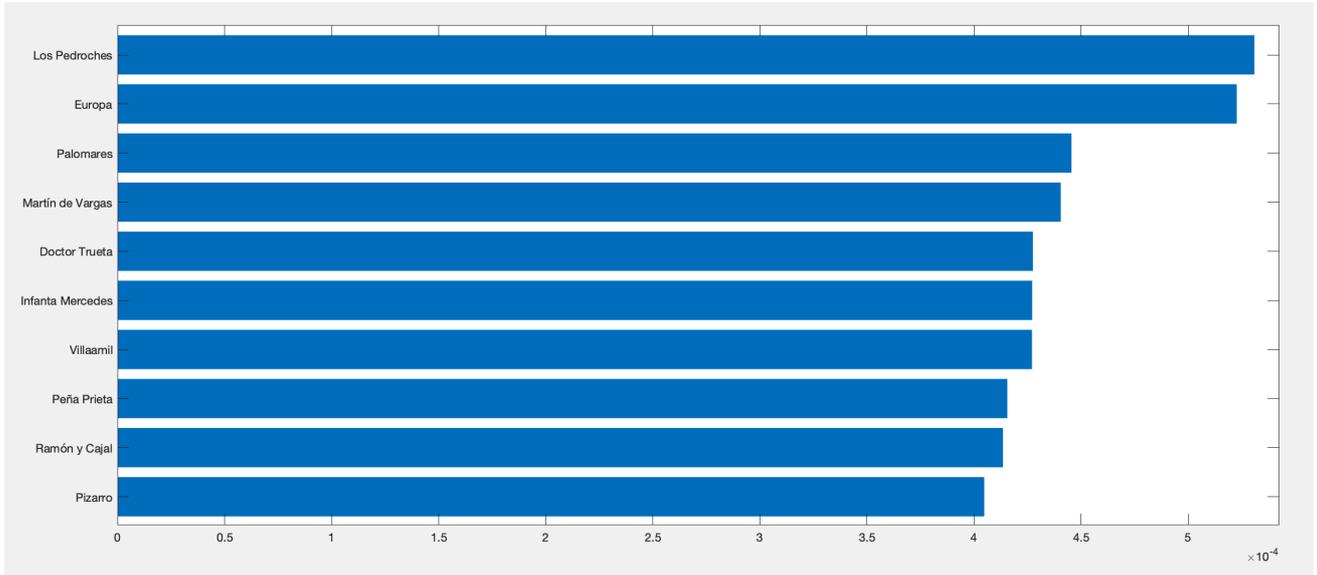


Figura 46: 10 ZBS con más contagiados

Para poder ver con claridad las zonas con más contagiados se ha representado en la Figura 47 las 10 localidades destacadas en la Figura 46.

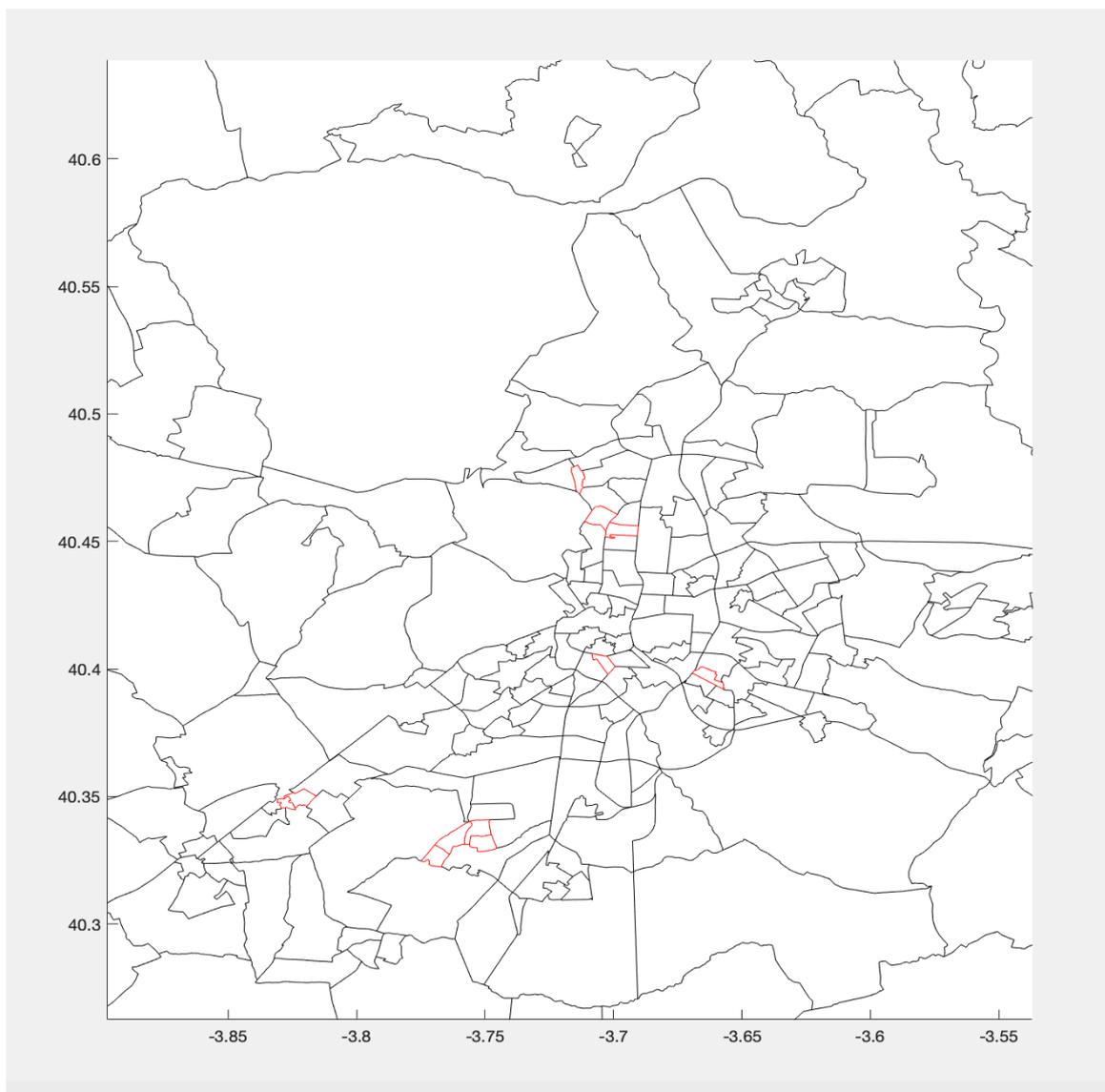


Figura 47: Señalización de las 10 ZBS con más contagiados

5.2.3 DIMENSIONES

A continuación se presenta los clústeres generados para diferentes conjuntos de dimensiones. Para cada grupo de dimensiones elegido se presenta las gráficas del método del codo, seguido de los centroides y finalmente los clústeres. Los análisis de los resultados se han realizado en el Capítulo 6.

5.2.3.1 Número de contagiados – Población total

En la Figura 48 se representa los valores del error de cuantización obtenidos tras aplicar K-means con diferentes números de clústeres (K). El punto en el que se observa un cambio brusco indicará el número óptimo de clústeres que se debe seleccionar [12]. En este caso se selecciona 4 grupos.

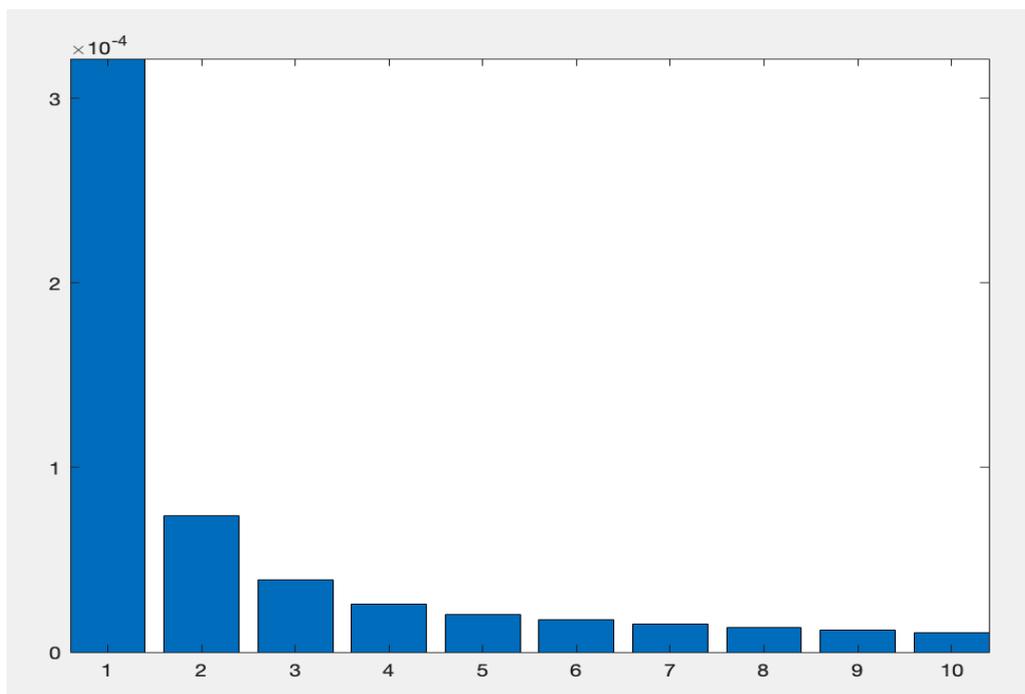


Figura 48: Método del codo - número de contagiados – Población total

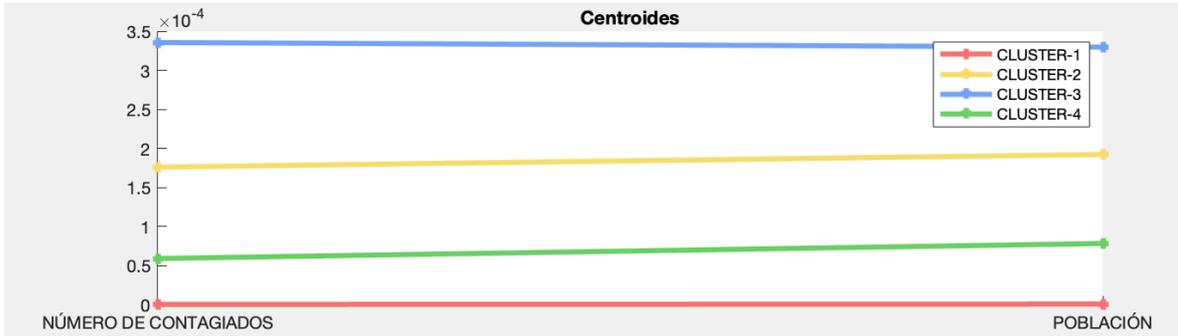


Figura 49: Centroides de los clústeres generados de Número de contagiados – Población total

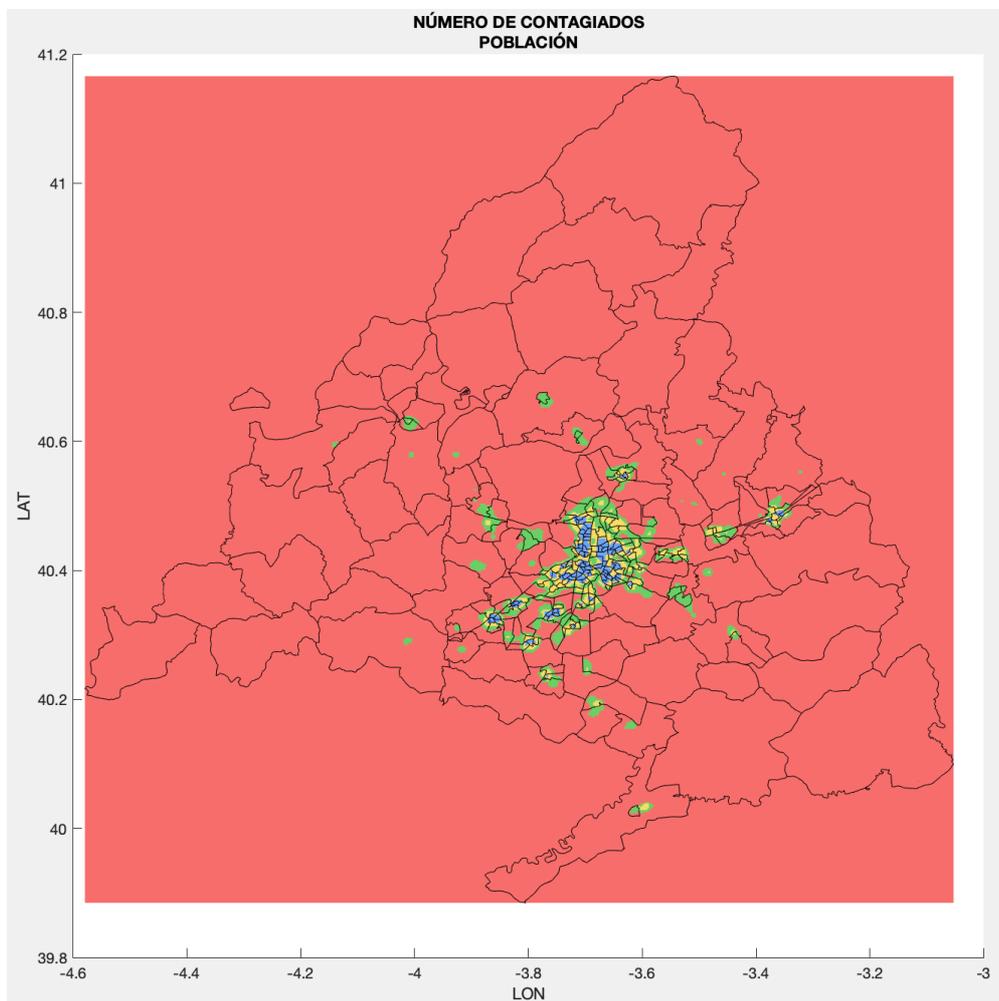


Figura 50 : Representación espacial de los clústeres generados de Número de contagiados – Población total

5.2.3.2 Número de Contagiados – Hospitales

Como ya se ha comentado antes, en la Figura 51 se puede observar la gráfica del método del codo, en este caso no hay un cambio brusco que nos indique un número óptimo de clústeres, se ha elegido 6 grupos para poder analizar con detalle la relación entre hospitales y contagiados.

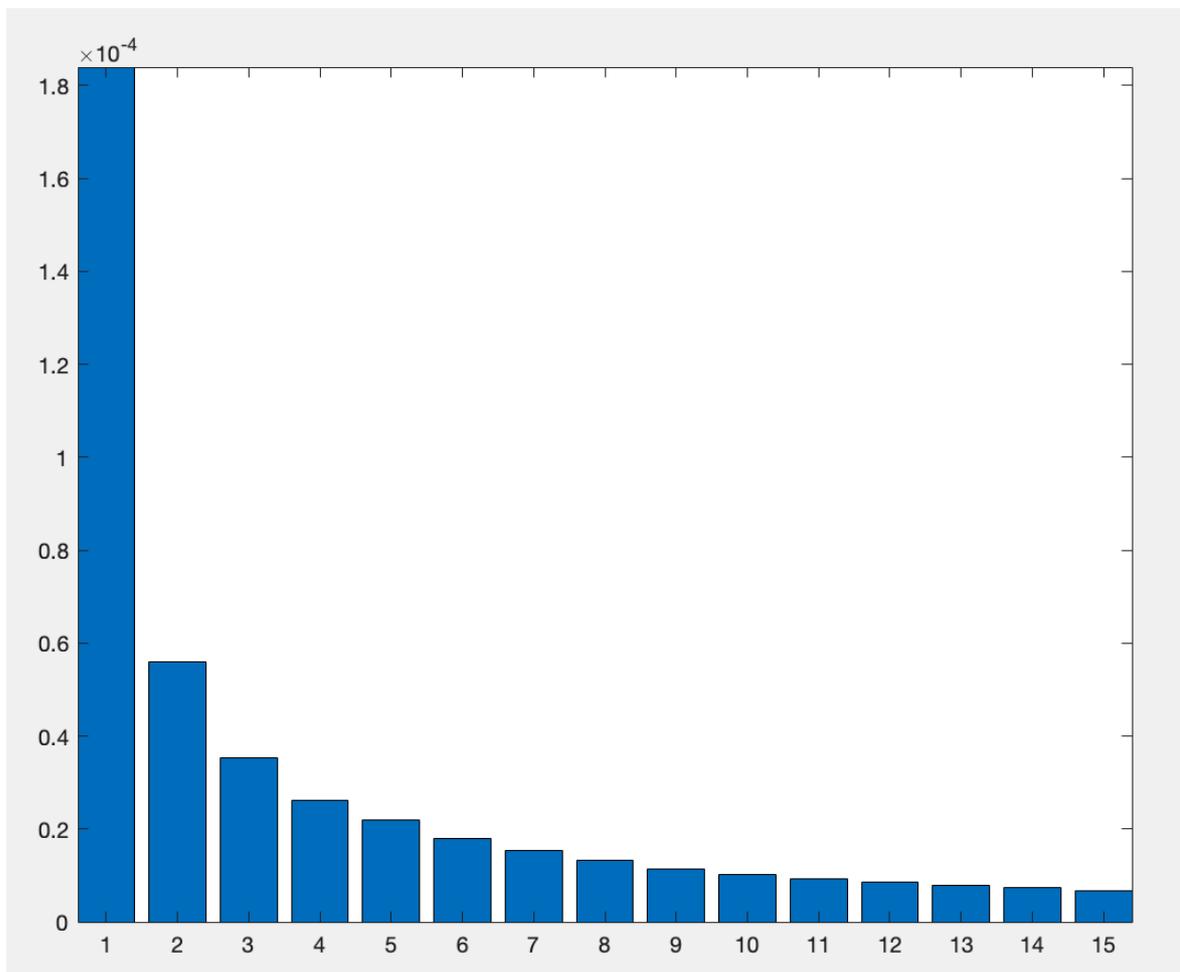


Figura 51 : Método del codo número de contagiados – Hospitales

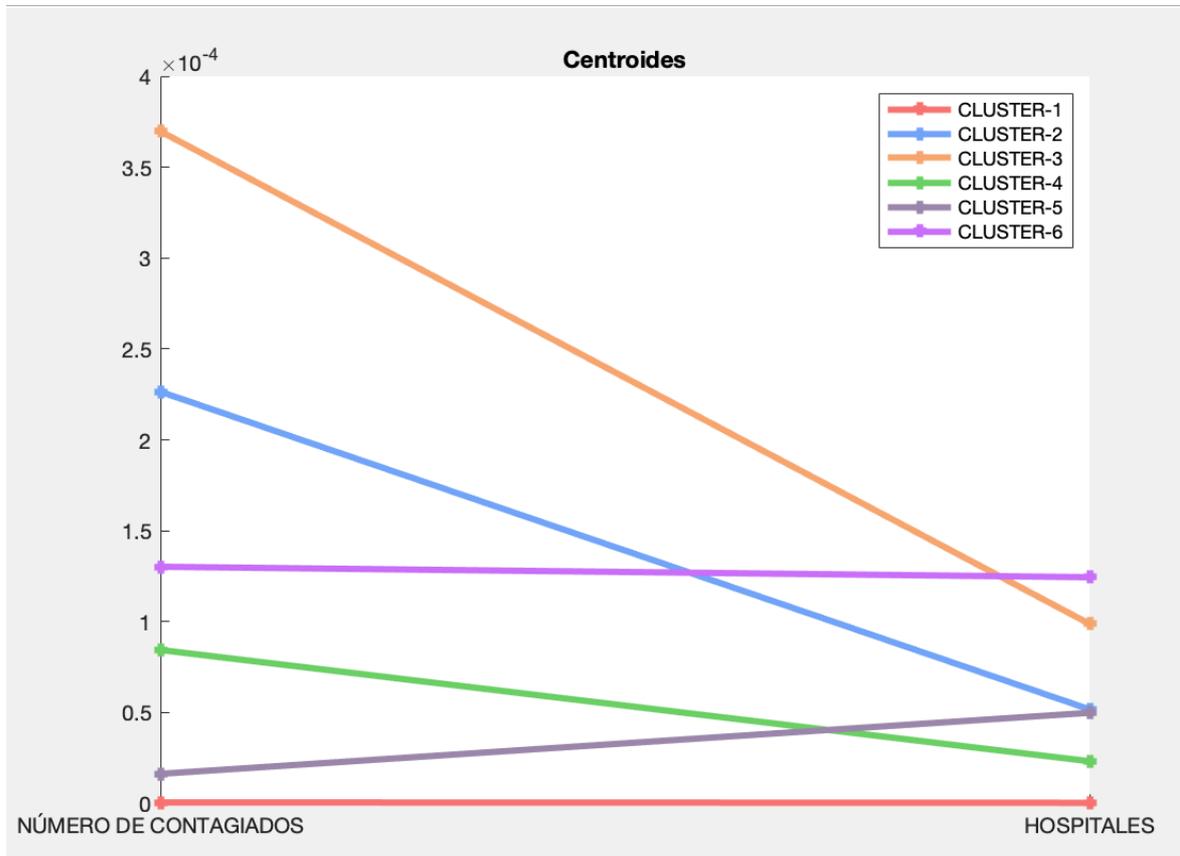


Figura 52 : Centroides de los clúster generados de Número de contagiados – Hospitales

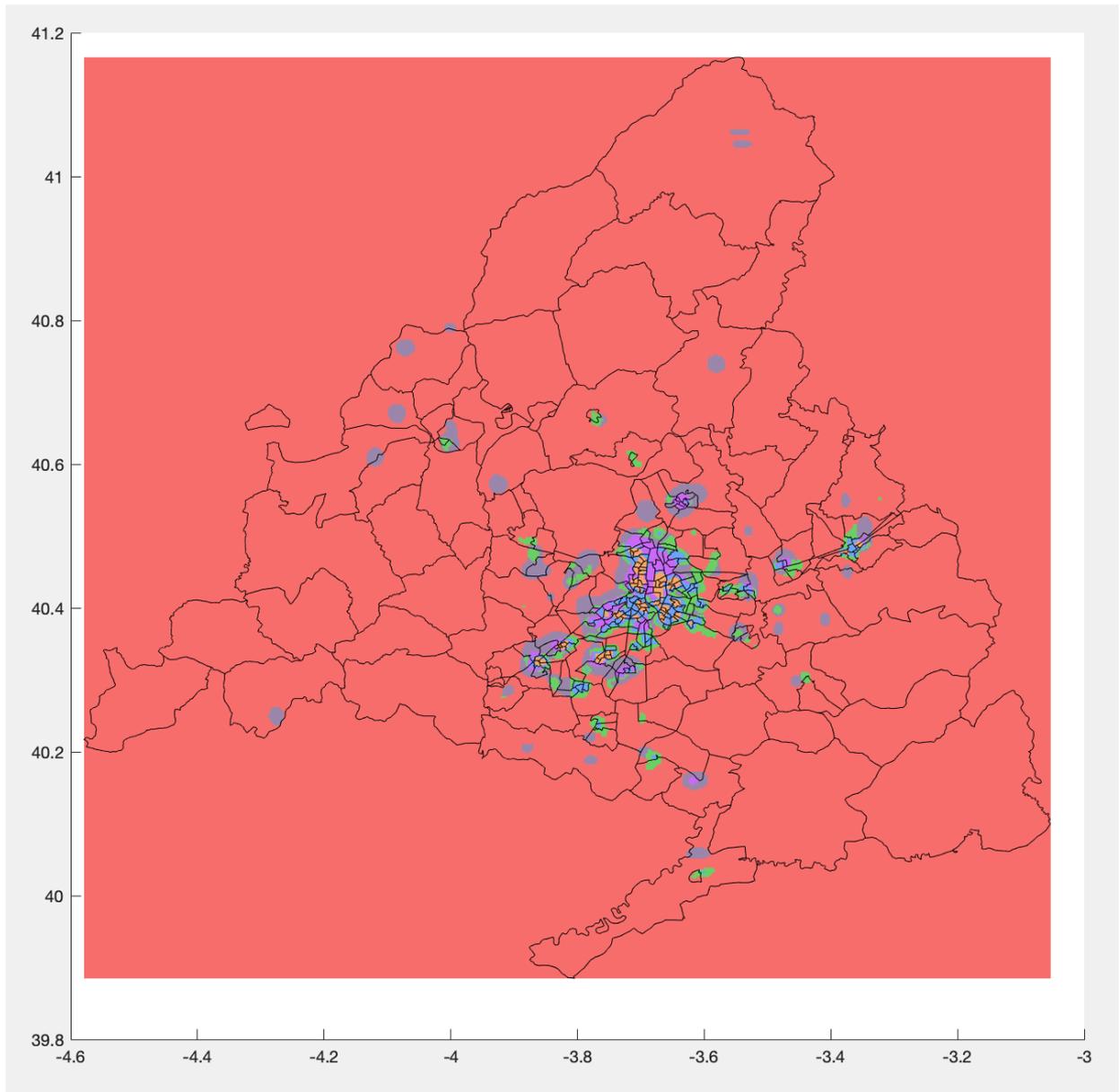


Figura 53 : Representación espacial de los clústeres generados de Número de contagiados - Hospitales

Capítulo 6. ANÁLISIS DE RESULTADOS

En este capítulo se analiza los resultados obtenidos en el anterior capítulo. Por un lado, se presenta la clasificación de los barrios de Madrid para diferentes números de clústeres y por otro se destaca los datos más relevantes obtenidos en el análisis de la pandemia.

6.1 COMUNIDAD DE MADRID

Utilizando las dimensiones especificadas en el capítulo anterior, se procede a introducirlas todas en un mismo análisis. Antes de realizar los clústeres se debe normalizar todas las dimensiones para que tenga sentido comparar diferentes unidades (densidades y euros).

Se ha empezado analizando los barrios de Madrid con 2 clústeres y se irá aumentando hasta 6 para aumentar la profundidad del análisis.

Los huecos blancos de las figuras se deben a falta de información en la renta de algunas SC.

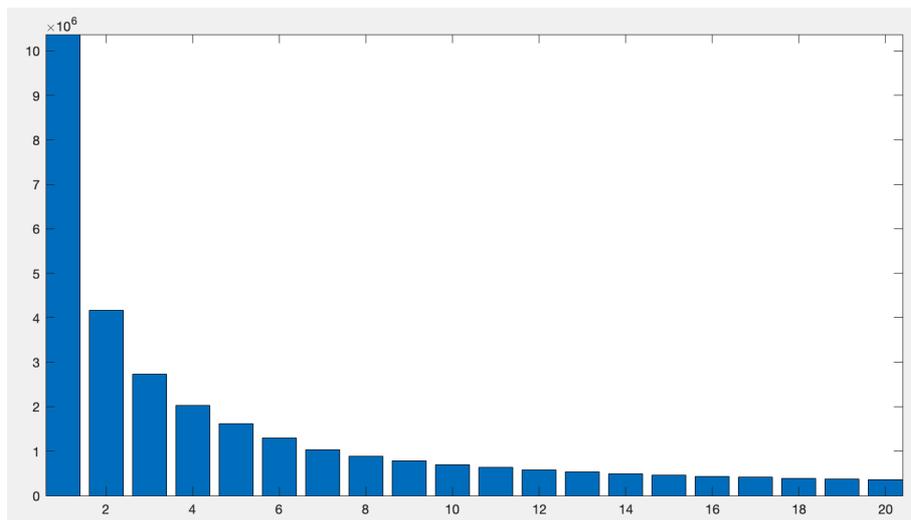


Figura 54: Método del codo

6.1.1 N° CLÚSTERES 2

Dividendo la ciudad de Madrid en 2 grupos, se puede ver como el programa ha separado la parte residencial de la parte comercial de la ciudad. Por tanto analizando la Figura 56 se puede concluir que casi toda la periferia de Madrid se puede etiquetar como residencial (clúster 1) mientras que el interior de la M30 se puede etiquetar como comercial (clúster 2).

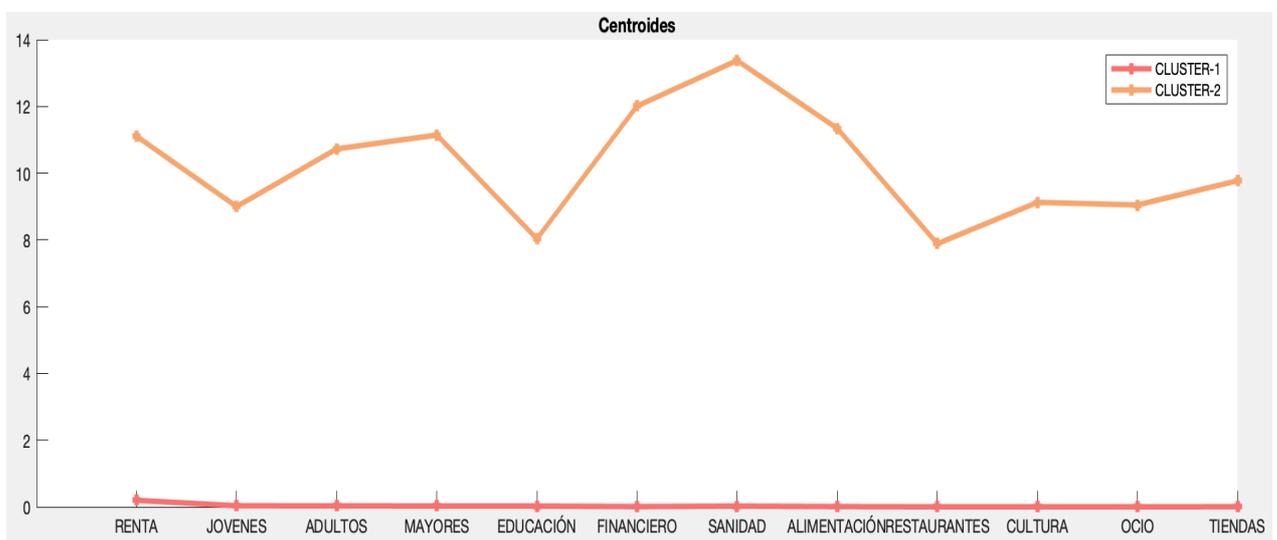


Figura 55: Centroides de 2 clústeres

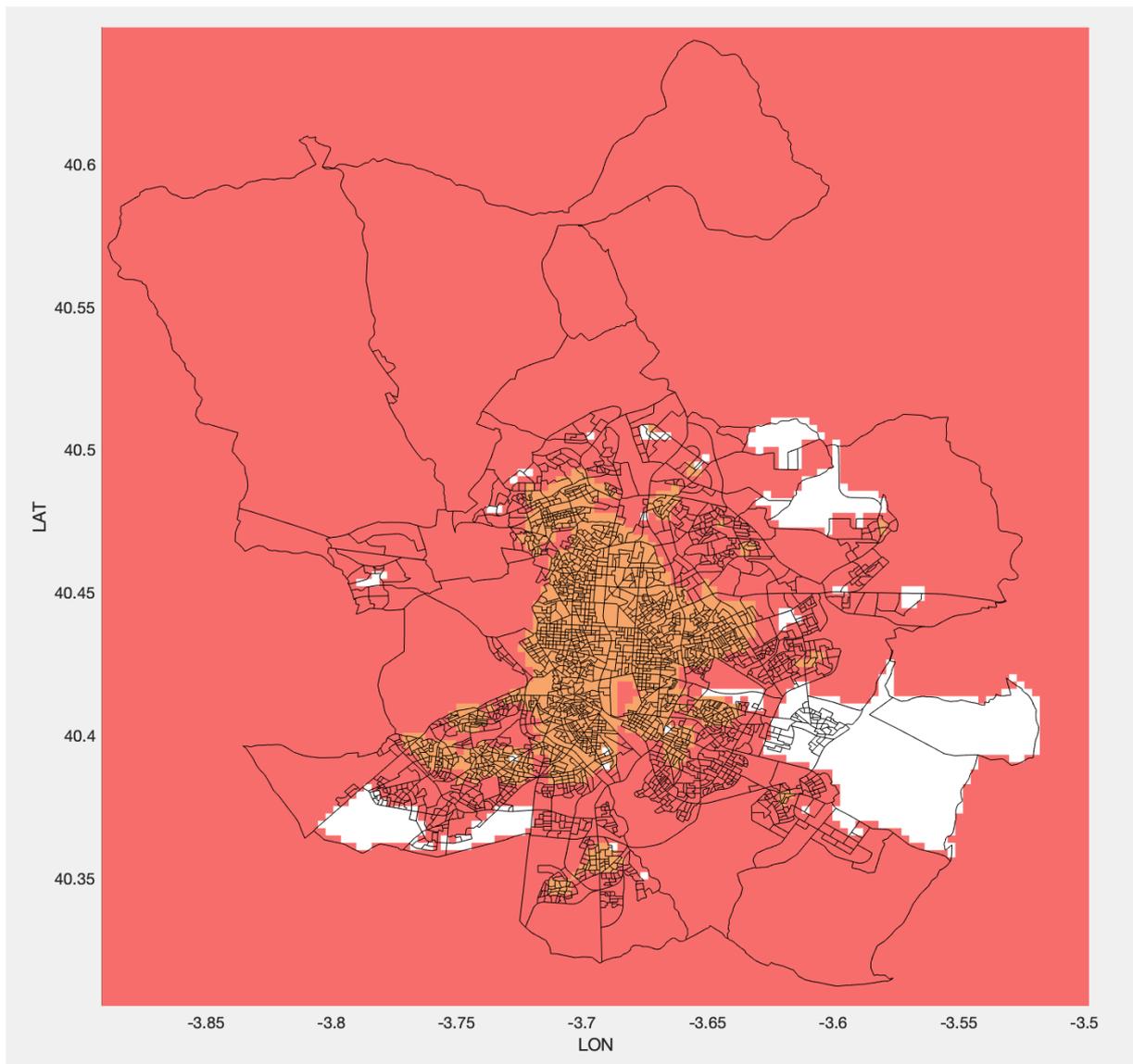


Figura 56: División de la ciudad de Madrid en 2 clústeres

6.1.2 N° CLÚSTERES 3

Al introducir un clúster más en el análisis, el programa ha separado la parte residencial en dos grupos diferenciados por la renta. El norte de Madrid destaca en la Figura 58 por tener una renta mucho más elevada que la zona residencial del sur de la ciudad. Por tanto, una de las posibles opciones de etiquetar los barrios es:

- Clúster 1- Barrio residencial con renta baja.
- Clúster 2 - Barrio residencial con renta alta.
- Clúster 3 - Barrio comercial.

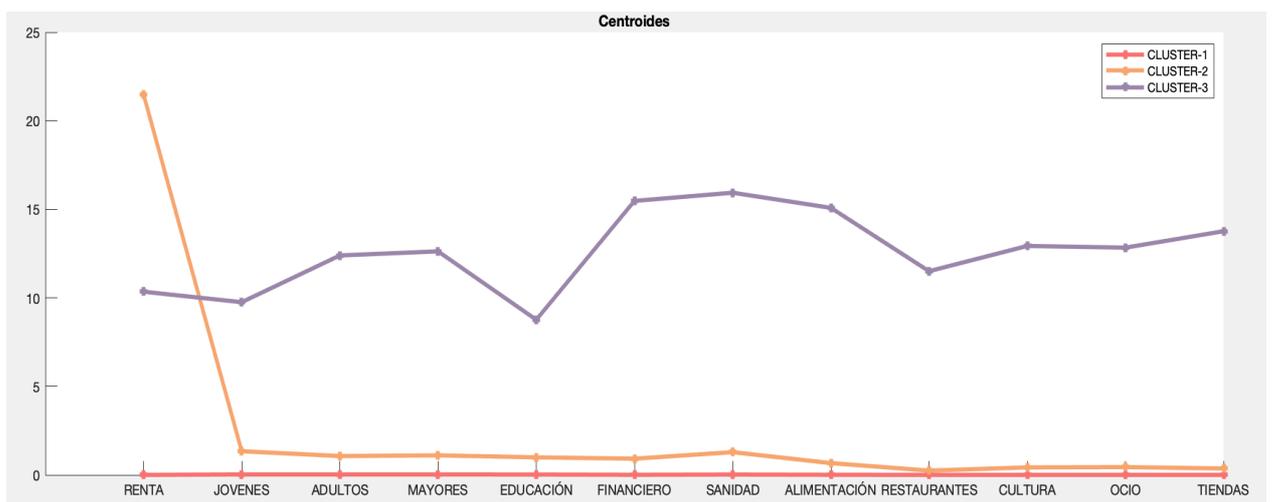


Figura 57: Centroides de 3 clústeres

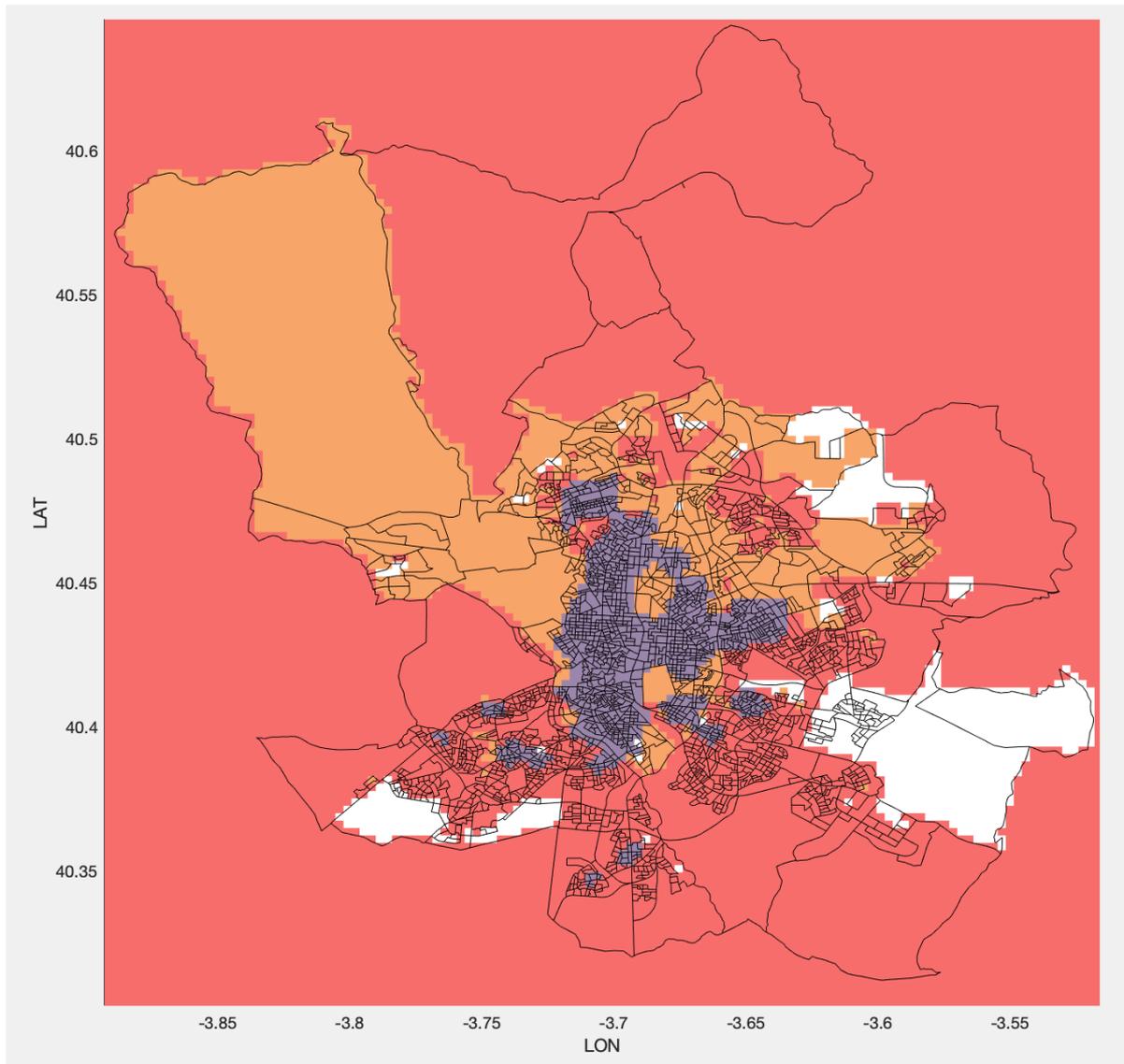


Figura 58: División de la ciudad de Madrid en 3 clústeres

6.1.3 N° CLÚSTERES 4

Comparando esta división con la anterior, se puede ver como los valores de los centroides de los clústers 1 y 2 no han variado, en cambio en la Figura 60 se observa como el clúster 1 en vez de ser zona residencial se podría etiquetar, en esta división, como campo. También destaca que lo que antes era zona comercial y barrio residencial de renta baja, se han unido en una nueva zona a la que se ha etiquetado como barrio residencial renta media. En esta nueva división la zona comercial ha perdido mucha superficie y ahora queda definida en La Almendra de Madrid. La alta densidad del clúster 3 es coherente con la organización urbanística de la almendra de Madrid, son calles estrechas con muchos edificios y negocios pero pocos espacios abiertos y pocas zonas verdes. Otra característica de la almendra de Madrid es el tipo de vivienda, en general son pequeñas y caras, esto también puede ser un motivo de la baja densidad de niño y alta densidad de adulto respecto al clúster 4. Por tanto, se puede etiquetar los clústeres de la siguiente forma:

- Clúster 1: Campo.
- Clúster 2: Barrio residencial con renta alta.
- Clúster 3: La Almendra.
- Clúster 4: Barrio residencial con renta media.

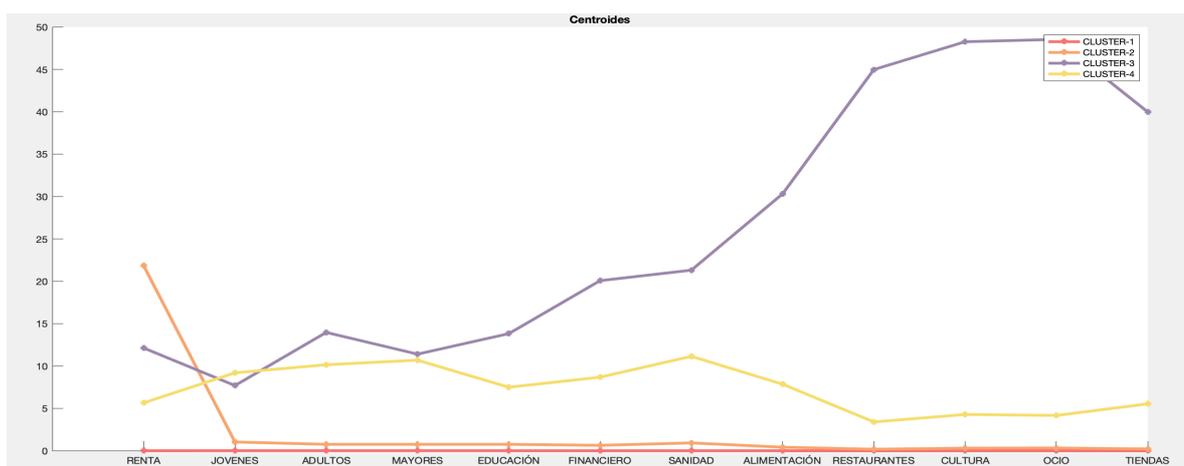


Figura 59: Centroides de 4 clústers

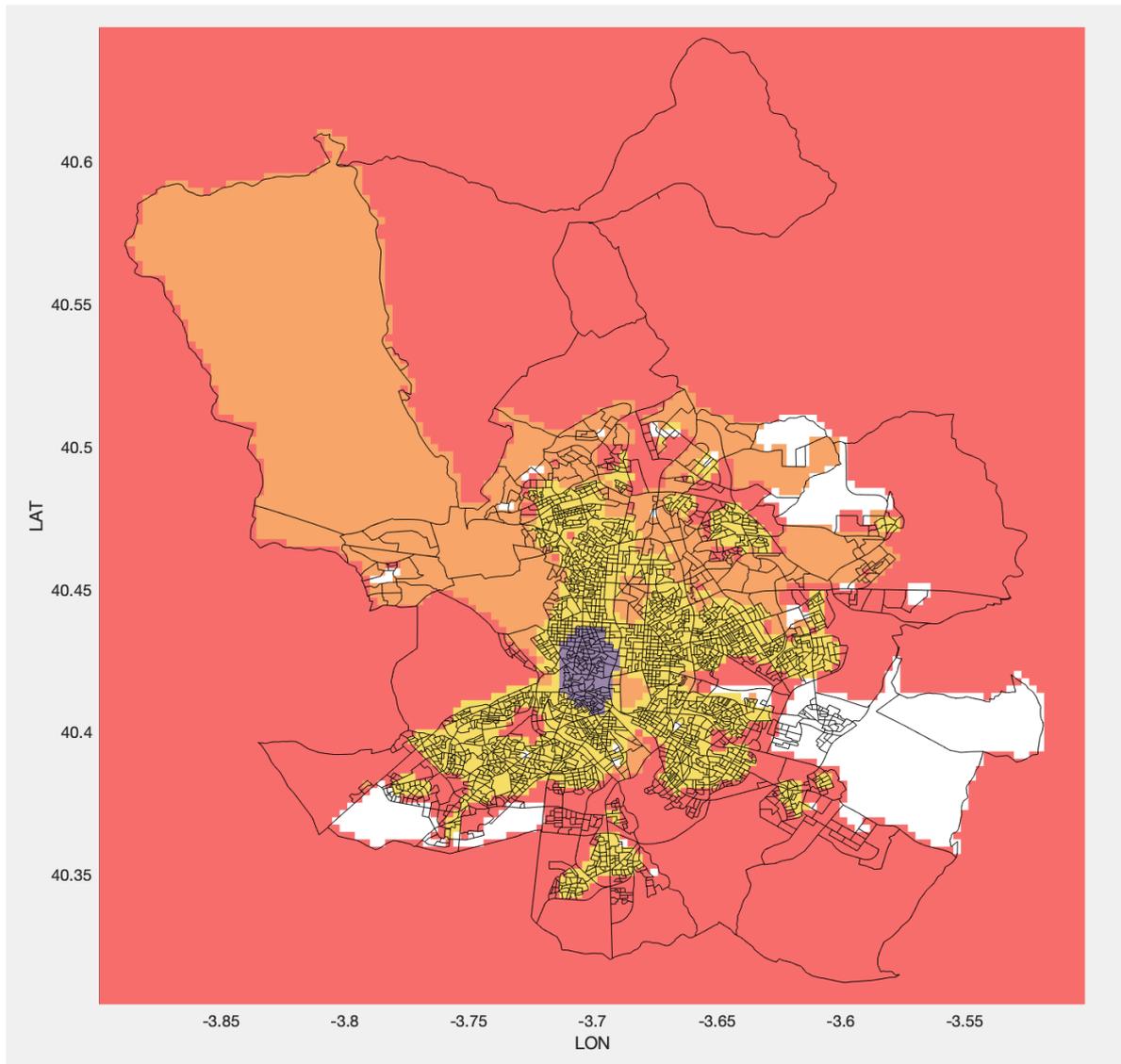


Figura 60: División de la ciudad de Madrid en 4 clústeres

6.1.4 N° CLÚSTERES 5

Se puede clasificar estos 5 grupos en 3 subcategorías: residencial, comercial y campo. Los dos grupos residenciales se diferencian por la renta y los dos grupos comerciales se diferencian por la densidad de servicios.

- Clúster 1: Campo.
- Clúster 2: Barrio comercial.
- Clúster 3: Barrio residencial renta alta.
- Clúster 4: La Almendra.
- Clúster 5: Barrio residencial renta baja.

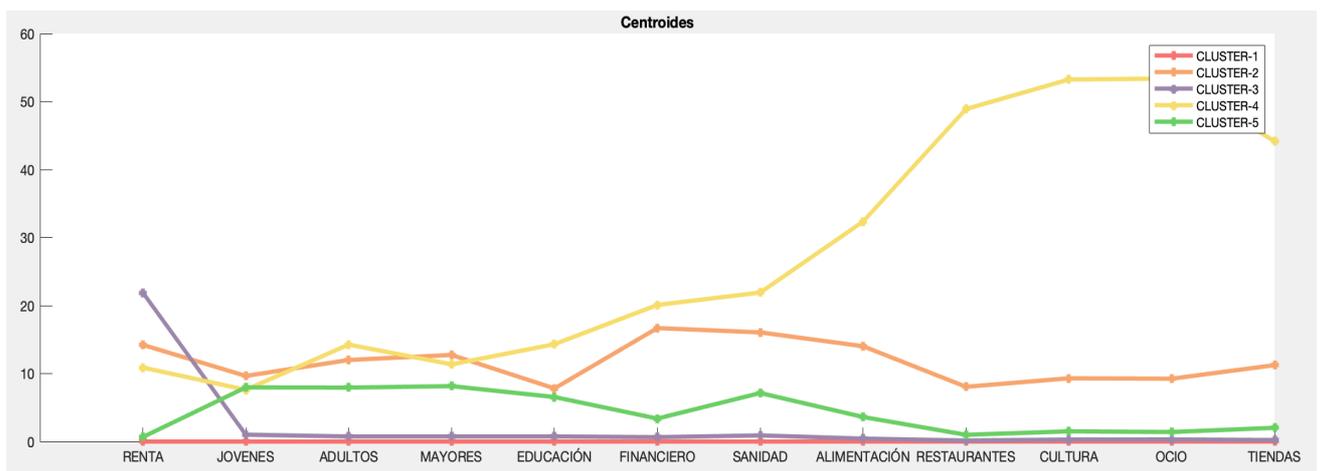


Figura 61: Centroides de 5 clústeres

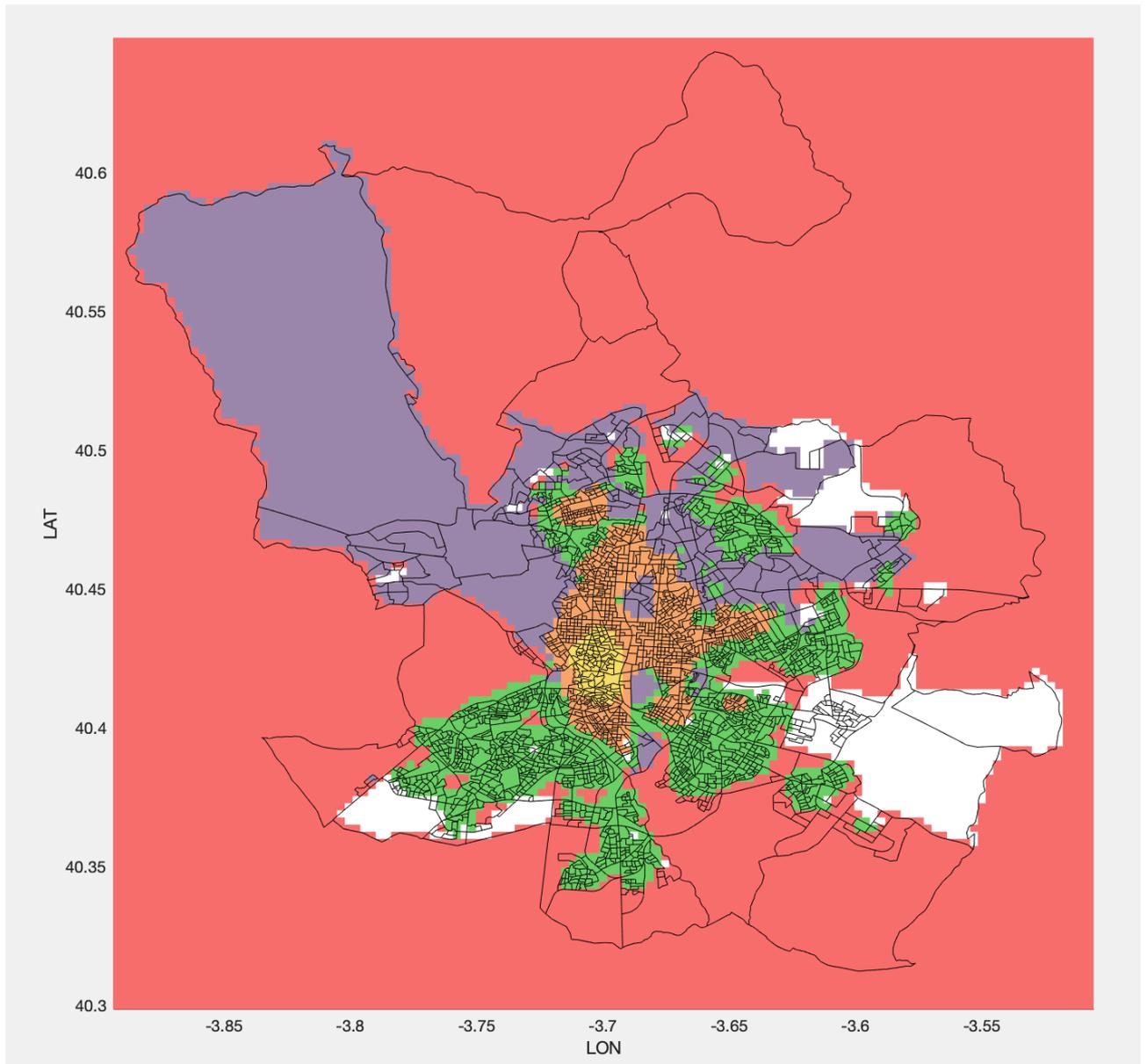


Figura 62: División de la ciudad de Madrid en 5 clústeres

6.1.5 N° CLÚSTERES 6

Con esta nueva clasificación la zona comercial se divide en una zona más residencial (clúster 3) de otra más activa (clúster 4). Un dato a destacar es el parecido entre la densidad de servicios del clúster 3 y 4 pero una gran diferencia en la renta media. Las etiquetas quedan:

- Clúster 1: Campo.
- Clúster 2: La Almendra.
- Clúster 3: Barrio residencial renta alta de interior.
- Clúster 4: Barrio comercial.
- Clúster 5: Barrio residencial renta baja.
- Clúster 6: Barrio residencial renta alta de periferia.

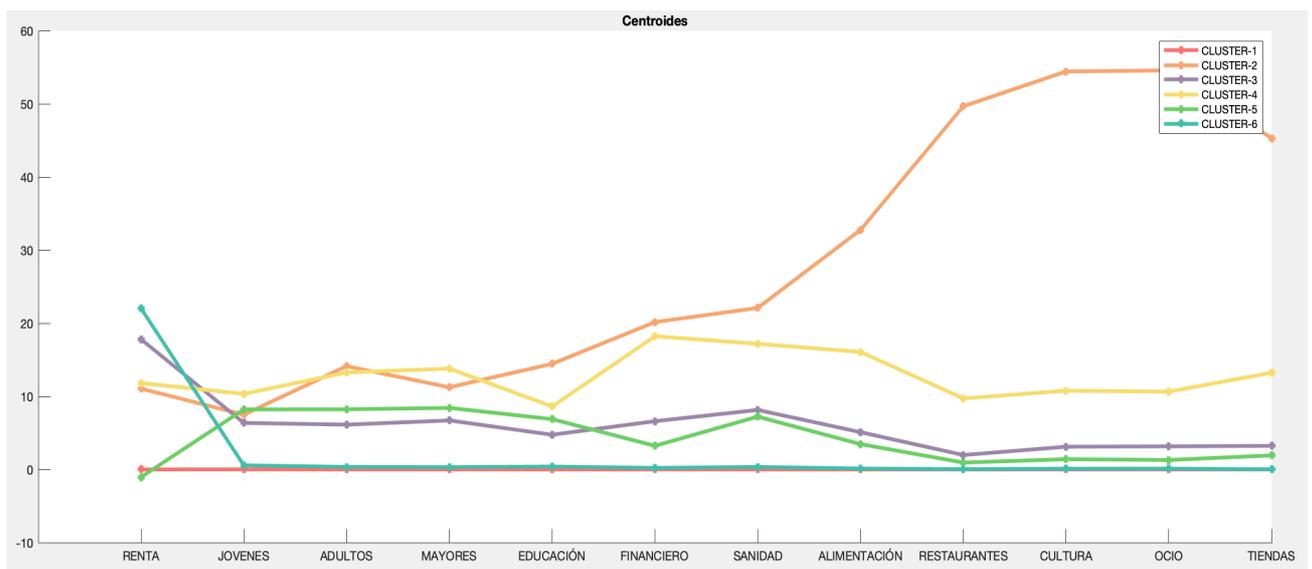


Figura 63: Centroides de 6 clústeres

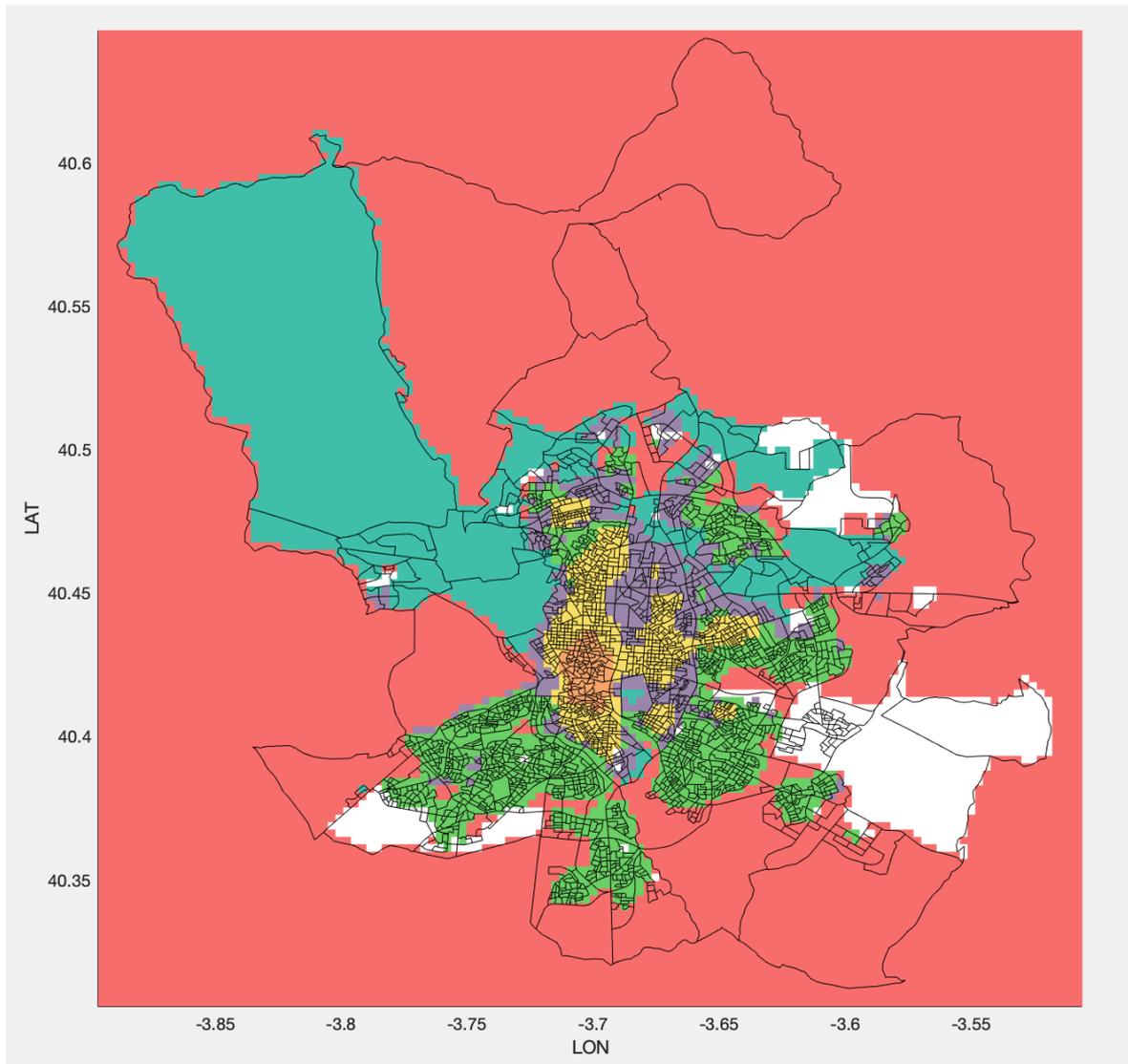


Figura 64: División de la ciudad de Madrid en 6 clústeres

6.2 COVID-19

6.2.1 NÚMERO DE CONTAGIADOS – POBLACIÓN

Observando las rectas de los clústeres generados a partir de la dimensión del número de contagiados y la dimensión población (Figura 49) se puede ver una relación directa, a más población, más contagiados. Para analizar con más detalle, se realiza la clasificación para cada dimensión por separado (Figura 66). Al comparar los clústeres de cada dimensión se puede percibir dos siluetas muy parecidas, esto confirma la relación directa detectada anteriormente.

De la Figura 66 destaca el barrio de Vallecas y Alcalá de Henares por apreciarse más clúster amarillo en la gráfica de los contagiados que en la de población, esto significa que la relación contagiados-población es más elevada que el resto de regiones. Por el contrario, destacan los ayuntamientos de Alcobendas, San Sebastián y Getafe. En general, se puede concluir que la ciudad Madrid no ha conseguido aislar la propagación del virus, todas las zonas han sido afectadas de igual forma en relación a su población.

A la vista del resultado anterior, uno de los temas que se debe reforzar en el protocolo de actuación, para posibles repuntes del número de contagiados en el futuro, sería la detección y aislamiento de los pacientes cero para intentar aislar aquella población afectada evitando la propagación a todas sus regiones cercanas.

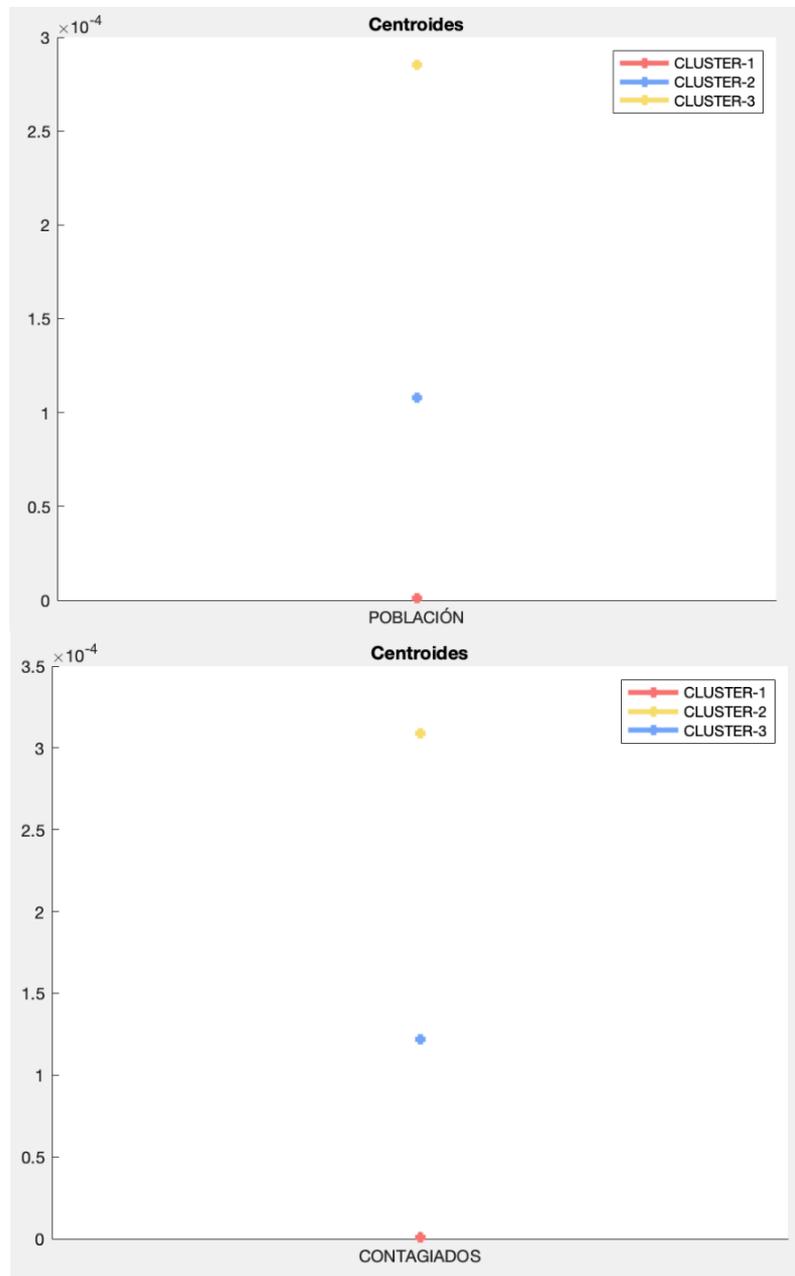


Figura 65: centroides dimensiones Población y Contagiados

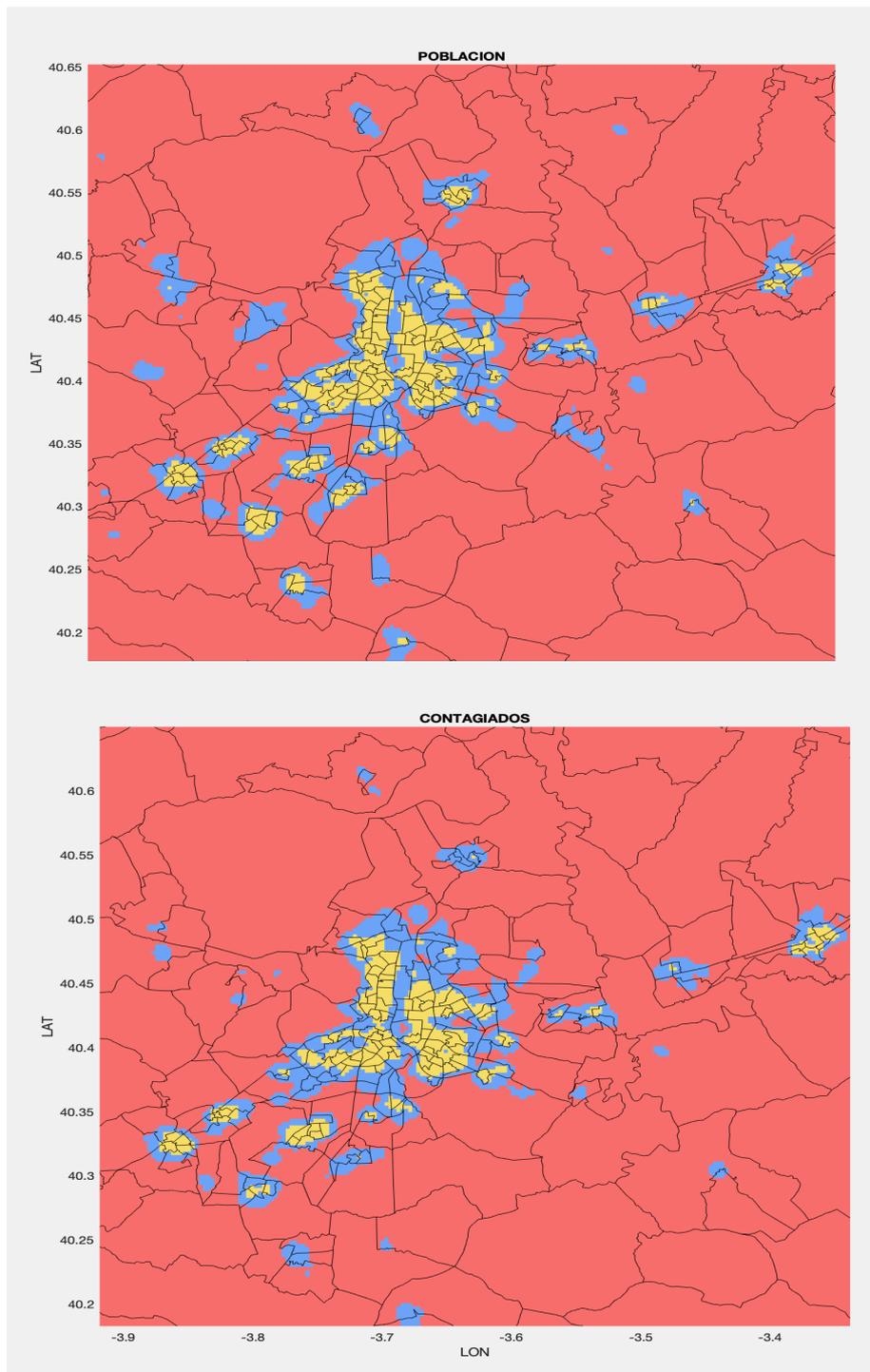


Figura 66 : Representación espacial de las variables Contagiados y Población total.

6.2.2 NÚMERO DE CONTAGIADOS – HOSPITALES

Si se analiza las rectas de los clústeres de la Figura 52 se puede decir que aquellas que tengan una pendiente negativa, es decir, más densidad de contagiados que de hospitales, serán las zonas que tengan más difícil superar exitosamente la pandemia. En consecuencia, aquellos hospitales de referencia en esas zonas serán los más afectados. Esos clústeres son el 2 (azul), 3 (naranja), 4 (verde).

En la Figura 67 se puede observar con más detalle la representación espacial de los clústeres.

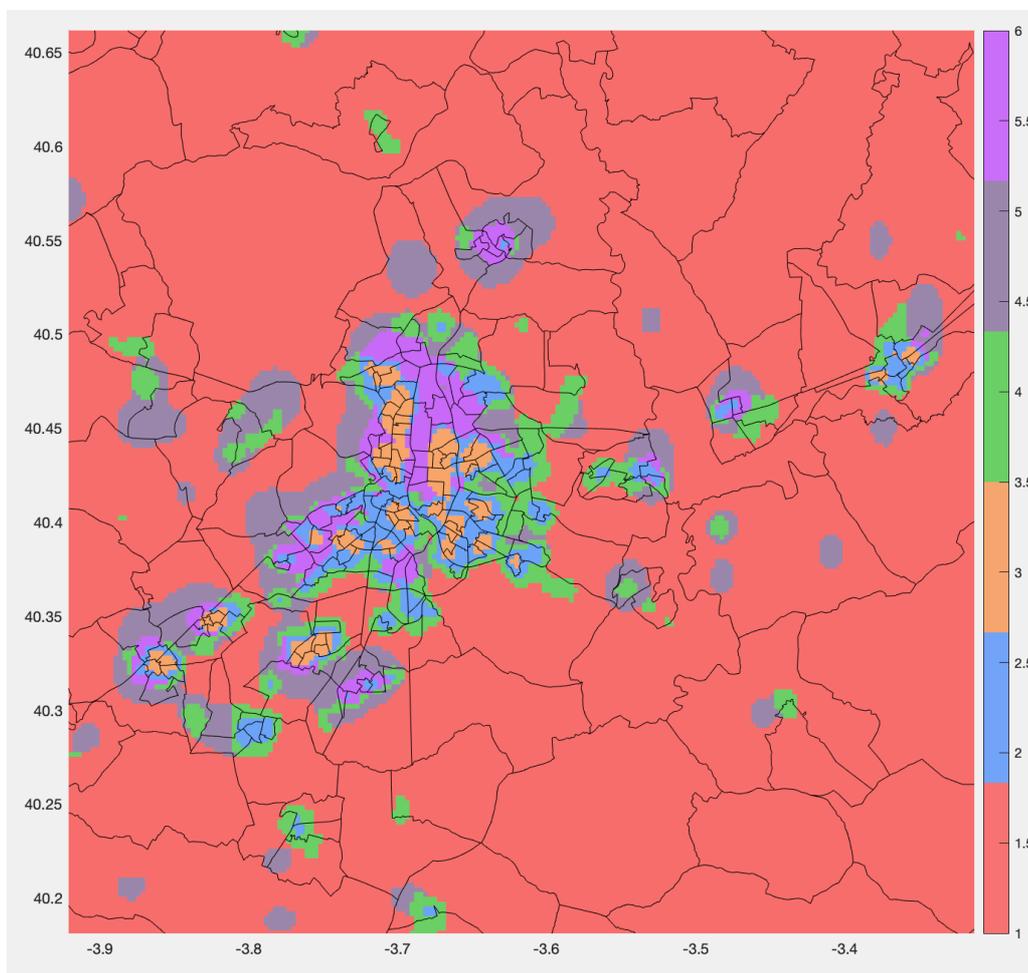


Figura 67: Zoom de la Figura 53

En la Figura 67 destacan los barrios de Usera, Vallecas, Chamberí, Tetuán, El Pilar, Ciudad Lineal y los municipios de Leganés, Móstoles y Alcorcón por ser zonas críticas. Por el contrario destacan los barrios de Chamartín, Arturo Soria y Aravaca por abundar en esas zonas los clúster 5 (gris) y 6 (morado), estos clústeres, tiene más densidad de hospitales que de contagiados, es decir pendiente positiva de la Figura 52.

A continuación se realiza una búsqueda entre los periódicos sobre los hospitales más afectados para representarlos posteriormente junto a los clústeres para que se vea dónde se necesita más apoyo sanitario.

Los hospitales de Madrid bordean el colapso por la llegada de casos y la escasez de material para atajar el coronavirus

Muchos centros están al límite de su capacidad: en Alcalá de Henares, en Leganés, Fuenlabrada, 12 de Octubre, Puerta de Hierro, La Paz... los sanitarios describen situaciones de colapso. "Estamos en una situación dramática. En urgencias había este miércoles 200 personas a las que no se podía atender. La gente llorando", relata un facultativo con mucha experiencia en el hospital Severo Ochoa de Leganés. En el Príncipe de Asturias de Alcalá de Henares se repite el calificativo: "Es dramático, hay pacientes hasta en las camillas de las consultas". [13]

Los hospitales de Madrid, desbordados hace unos días, empiezan a notar la caída de la presión en Urgencias

El Severo Ochoa, en Leganés [...] .Desde el Hospital Puerta de Hierro, en Majadahonda [...] El 12 de octubre [...] Fuentes del Gregorio Marañón [...] En La Paz [...] [14]

Colapso: Tres hospitales de Madrid ya no admiten nuevos infectados por saturación

Severo Ochoa de Leganés, Vallecas y Coslada registran ya tantos enfermos en urgencias en espera de ingreso como hospitalizados en sus instalaciones. [15]

Estos son los 10 hospitales de Madrid con más casos de coronavirus

La mayoría de pacientes hospitalizados se encuentran en observación, sin embargo, cada día aumentan las muertes por coronavirus. [16]

Recolectando la información de las anteriores fuentes, se ha seleccionado los siguientes hospitales como algunos de los más afectados:

1. Hospital Universitario La Paz
2. Hospital General Universitario Gregorio Marañón
3. Hospital Universitario 12 de Octubre
4. Hospital Universitario Príncipe de Asturias
5. Hospital Virgen de la Torre
6. Hospital Clínico San Carlos
7. Hospital Ramón Y Cajal
8. Hospital Universitario Puerta De Hierro
9. Hospital Universitario de Torrejón
10. Hospital Universitario Severo Ochoa

Para que se vea gráficamente, se representa en la **¡Error! No se encuentra el origen de la referencia.** la ubicación los hospitales junto con la Figura 67. Para estimar la región a la que atiende cada hospital se ha representado el diagrama de Voronoi (Figura 69). Este método divide la región de estudio en secciones en las que se incluye todos los puntos más cercanos a cada hospital que a cualquier otro del conjunto [17]. Esta división no se corresponde con la región a la que atiende cada hospital, es simplemente una hipótesis basada en distancia euclídea. Examinando estas divisiones, se aprecia que los colores que más abundan en todas las regiones (excluyendo el rojo) son los colores naranja, verde y azul. Estos colores representan a los clústeres con más densidad de contagiados que de hospitales y por tanto a las zonas más críticas.

Recopilando toda esta información, se podría utilizar para generar los protocolos de actuación. Centrándose en aquellas zonas más críticas y buscando soluciones personalizadas para cada región y para cada hospital en función de la población y de los servicios que tengan a su alcance.

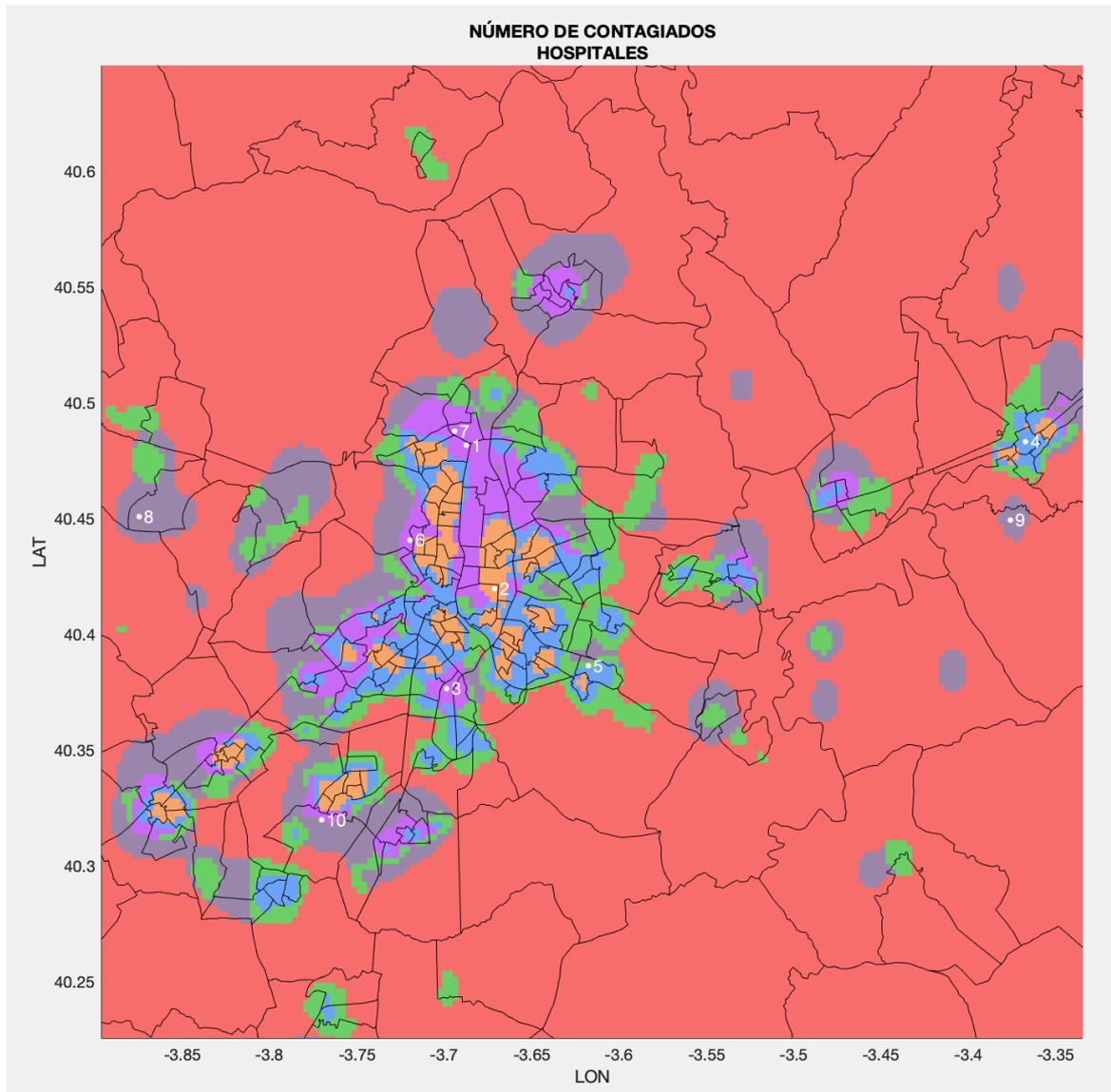


Figura 68: Introducción de la localización de los Hospitales en la Figura 67

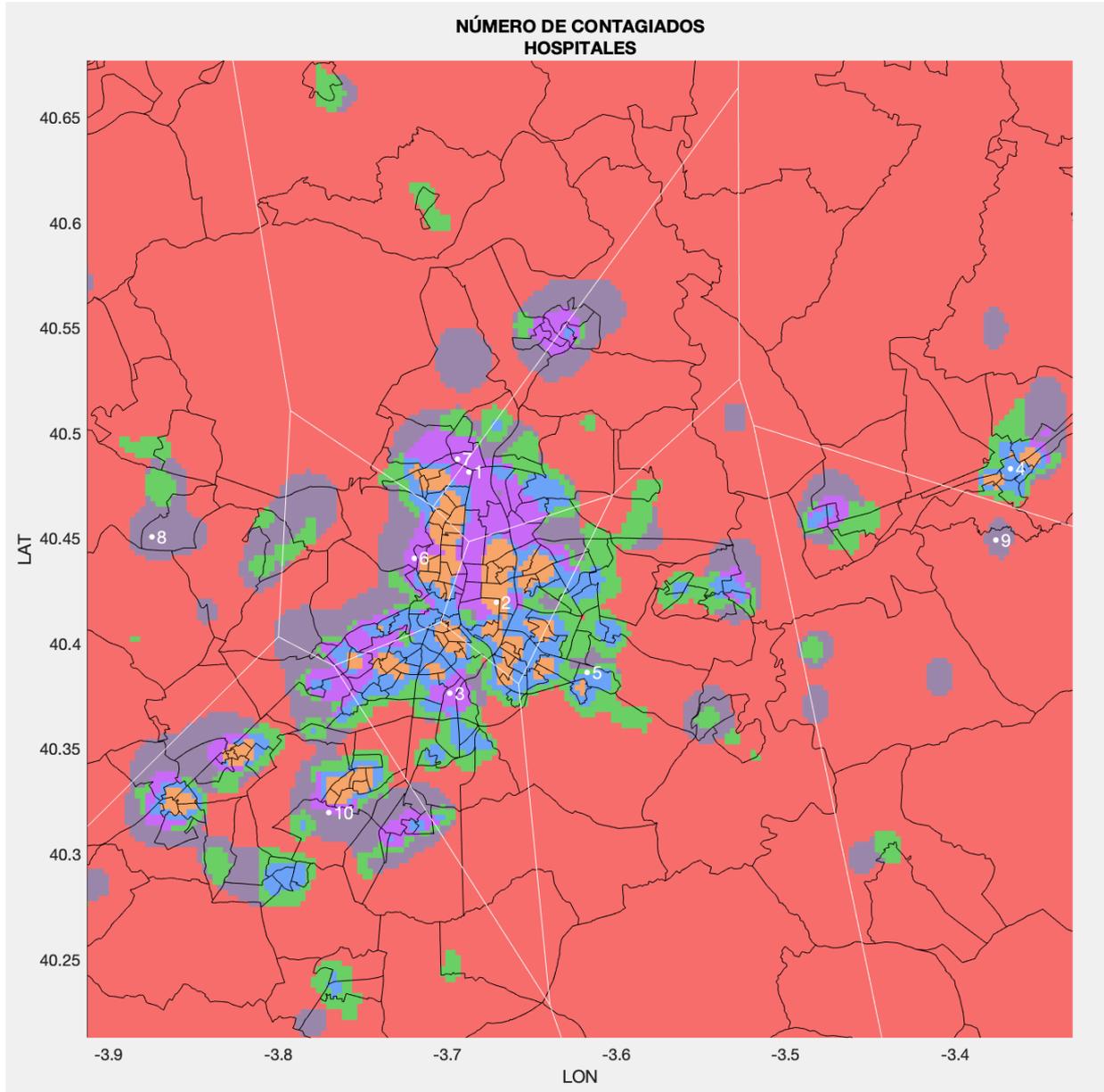


Figura 69: Diagrama de Voronoi

Capítulo 7. CONCLUSIONES

El modelo propuesto para extraer conocimiento de información georreferenciada altamente desestructurada ha demostrado ser de gran utilidad al permitir resolver dos problemas concretos. Por un lado, se ha podido determinar automáticamente, a partir de un conjunto de dimensiones socioeconómicas y demográficas, cómo está organizado espacialmente la ciudad de Madrid, determinado que existen 6 tipos fundamentales de barrios diferentes. Además, se ha estudiado la evolución espacial de la pandemia del COVID-19 en la región de Madrid, identificando las zonas más afectadas y su relación con los hospitales que más demanda de atención sanitaria han soportado.

Basándose en el conocimiento extraído con este tipo de modelado se podría facilitar la comprensión de la organización socioeconómica de las zonas que puedan verse afectadas en nuevos rebrotes de COVID-19 o futuras pandemias, ayudando de esta forma a la toma de decisiones para realizar una buena gestión de los casos, tanto desde el punto de vista asistencial en los hospitales implicados, como social de los barrios.

En conclusión, se ha demostrado con los dos casos concretos considerados la flexibilidad del enfoque propuesto diversidad de soluciones que pueden aportar estos modelos a posibles problemas que se puedan dar en la sociedad.

Capítulo 8. BIBLIOGRAFÍA

- [1] **Blog OpenAlfa.** *Último acceso:* 25 junio 2020_ <https://blog.openalfa.com/como-consultar-openstreetmap-con-overpass-api>
- [2] **Wiki OSM.** *Último acceso:* 25 junio 2020
https://wiki.openstreetmap.org/wiki/Map_Features
- [3] **Overpass turbo.** *Último acceso:* 25 junio 2020
<https://overpass-turbo.eu/>
- [4] **Overpass _QL.** *Último acceso:* 25 junio 2020
https://wiki.openstreetmap.org/wiki/Overpass_API/Overpass_QL
- [5] **Instituto Nacional de Estadística.** *Último acceso:* 25 junio 2020
<https://www.ine.es>
- [6] **Atlas de distribución de renta de los hogares.** Proyecto técnico. *Último acceso:* 25 junio 2020 https://www.ine.es/experimental/atlas/exp_atlas_proyecto.pdf
- [7] **Mathworks.** *Último acceso:* 25 junio 2020
<https://es.mathworks.com/matlabcentral/fileexchange/10914-utm2deg>
- [8] **Comunidad de Madrid.** *Último acceso:* 25 junio 2020
https://datos.comunidad.madrid/catalogo/dataset/covid19_tia_muni_y_distritos
- [9] **ZBS .** *Último acceso:* 25 junio 2020
https://datos.comunidad.madrid/catalogo/dataset/covid19_tia_zonas_basicas_salud
- [10] **ARISTA SUR.** *Último acceso:* 25 junio 2020
<https://www.aristasur.com/contenido/openstreetmap-mapas-libres-y-abiertos>
-

[11] **Wikipedia.** *Último acceso:* 25 junio 2020

[https://es.wikipedia.org/wiki/Instituto_Nacional_de_Estadística_\(España\)](https://es.wikipedia.org/wiki/Instituto_Nacional_de_Estadística_(España))

[12] **Jarroba.** *Autor:* Ricardo Mora. *Fecha:* 12-sep-2016 *Último acceso:* 25 junio 2020

[https://jarroba.com/seleccion-del-numero-optimo-clusters/#:~:text=M%20todo%20del%20codo%20\(Elbow%20Method\),-Este%20m%20todo%20utiliza&text=El%20punto%20en%20el%20que,Clusters%20para%20ese%20data%20set.](https://jarroba.com/seleccion-del-numero-optimo-clusters/#:~:text=M%20todo%20del%20codo%20(Elbow%20Method),-Este%20m%20todo%20utiliza&text=El%20punto%20en%20el%20que,Clusters%20para%20ese%20data%20set.)

[13] **El Diario.** *Fecha:* 18-marzo-2020 *Autores:* Raúl Rejón / Ana Requena Aguilar / Sofía Pérez Mendoza *Último acceso:* 25 junio 2020

https://www.eldiario.es/sociedad/hospitales-Madrid-bordean-material-coronavirus_0_1007200253.html

[14] **El Mundo.** *Fecha:* 8-abril-2020 *Autores:* VIRGINIA GÓMEZ / ROBERTO BÉCARES *Último acceso:* 25 junio 2020

<https://www.elmundo.es/madrid/2020/04/07/5e8b6a4bfdddf8fae8b45b7.html>

[15] **La Razón.** *Fecha:* 21-03-2020 *Autor:* Sergio Alonso *Último acceso:* 25 junio 2020

<https://www.larazon.es/salud/20200321/g5157qtidbc67lqsberaivz7ry.html>

[16] **El Español.** *Fecha:* 19-marzo-2020 *Autores:* Juan Rodríguez de Rivera *Último acceso:* 25 junio 2020

<https://www.lespanol.com/espana/madrid/20200319/hospitales-madrid-casos-coronavirus/47595373>

[17] **Diagrama Voronoi.** *Último acceso:* 25 junio 2020

<https://es.mathworks.com/help/matlab/ref/voronoi.html#bvmagic>

[18] **Sustainable development.** *Último acceso:* 25 junio 2020

<https://sustainabledevelopment.un.org/sdgs>

[19] *Artículo:* Spatial accessibility of primary care: concepts, methods and challenges *Autor:* M.F. Guagliardo. *Publicado:* International Journal of Health Geographics 2004

[20] *Artículo*: A two-stage spatial allocation model for elderly healthcare facilities in large-scale affordable housing communities: a case study in Nanjing City *Autores*: (1)School of Civil Engineering, Southeast University, Jiangning District, Nanjing (2) 211189, China. Lyles School of Civil Engineering, Purdue University, West (3)Lafayette, IN 47907, USA. School of Civil Engineering, Nanjing University of Technology, Nanjing 211816, China. *Publicado*: International Journal for Equity in Health (2018)

[21] *Artículo*: Using GIS for Determining Variations in Health Access in Jeddah City, Saudi Arabia *Autor*: Abdulkader Murad *Published*: International Journal of Geo-Information 2018

ANEXO I – FUENTES DE INFORMACIÓN

En este anexo se describen las principales fuentes de información utilizadas en este proyecto. En concreto, para cada una de ellas se indica qué datos se han descargado y de dónde, describiéndose todos los detalles técnicos necesarios para poder reproducir el proceso fácilmente.

8.1 OPEN STREET MAPS (OSM)

Open Street Maps es un proyecto para generar un mapa mundial de libre acceso [10]. Este proyecto está formado por una comunidad de voluntarios que introducen información georreferenciada de todo tipo bajo una licencia abierta (ODbl), lo que permite poder descargarse los datos libremente siempre que se indique de donde se han descargado (de esta forma se da reconocimiento a la comunidad de OSM). Dado que esta licencia no tiene ninguna limitación en el número y volumen de las descargas realizadas, se ha elegido OSM para realizar este proyecto frente a otras como Google Maps. A pesar de esta gran ventaja, la descarga gratuita e ilimitada de información, OSM tiene una limitación que puede suponer un problema: la información introducida por los usuarios no se valida. Esta desventaja obliga a invertir tiempo adicional para verificar la calidad de la información obtenida.

La información que se puede descargar de OSM tiene cuatro formatos principales: nodos, vías, relaciones y etiquetas. En este proyecto se han utilizado principalmente los nodos, los cuales, se descargan y estructuran en función de sus etiquetas. La comunidad de OSM ha acordado una serie de etiquetas como las más importantes y frecuentes, pero las etiquetas son ilimitadas, cualquier usuario puede crear una nueva. Por tanto es preferible utilizar alguna etiqueta especificada por la comunidad (como ‘amenity’) a la hora de descargarse los datos, y una vez se tienen descargados, filtrarlos con etiquetas más específicas (como ‘restaurant’). Las etiquetas más importantes se encuentran explicadas en [2]

Las descarga de datos se realiza a través de la API Overpass de OSM [1] [4]. Esta API permite la comunicación entre el programa que se vaya a utilizar para tratar la información y la base de datos de OSM. Las descargas se realizarán a través de peticiones a la url que ofrece la API mediante lenguaje Overpass_QL.

En este proyecto se ha estudiado tres formas diferentes de realizar las consultas según la forma de delimitar la región de estudio: por bbox, por nombre y por polígono. A continuación se describe cada método.

8.1.1 CONSULTA MEDIANTE BOUNDING BOX

Se denomina Bbox (*bounding box*) al cuadrado que delimita la región de estudio. Este rectángulo se acota con la longitud y latitud de la esquina inferior izquierda y con la longitud y latitud de la esquina superior derecha. Este tipo de consulta tiene la ventaja de necesitar poca información para delimitar la región, pero tiene la desventaja de no ser muy útil para zonas de estudio irregulares.

La url que realiza la consulta tiene tres partes, una raíz que será común a todos los tipos de consultas, una petición que irá cambiando en función del tipo de la consulta y un cierre también común a todas:

- Raíz: `'http://overpassapi.de/api/interpreter?data=[out:json][timeout:200];'`
- Petición: Formato + Etiqueta + BBOX.
- Cierre: `';out;'`

Por ejemplo si se quiere descargar los restaurantes del barrio Salamanca de Madrid, la url sería:

- Raíz: `'http://overpassapi.de/api/interpreter?data=[out:json][timeout:200];'`
- Formato: `'node'`
- Etiqueta: `'["restaurant"]'`
- BBOX: `'(40.419754,-3.693152,40.437984,-3.671427)'`

- Cierre: `';out;'`

En la Figura 71 y en la Tabla 2, se puede ver el procedimiento a seguir para acotar la BBOX de la url.

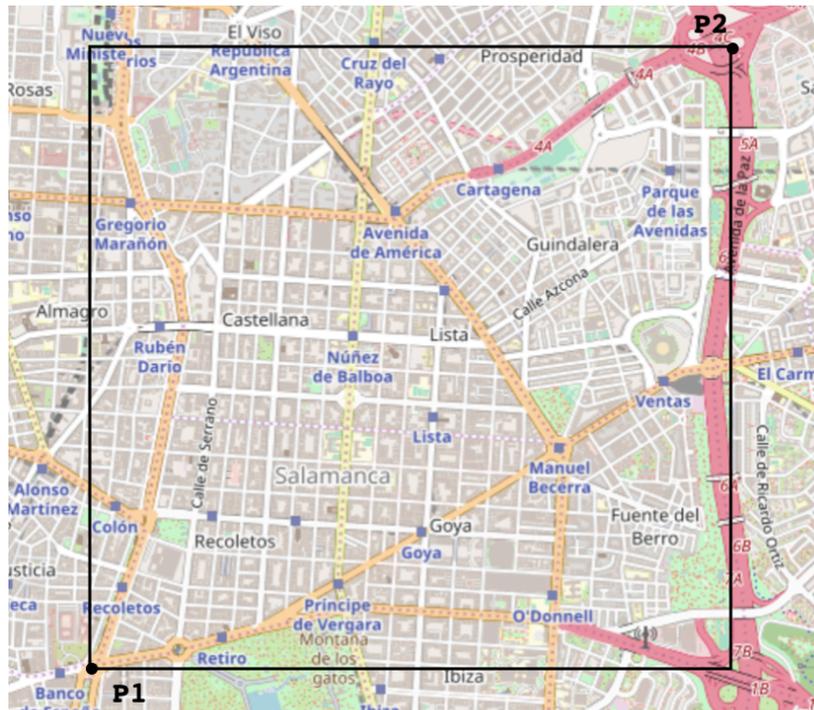


Figura 70: Acotación barrio Salamanca mediante una bbox

Tabla 2: Puntos de bbox que delimitan el barrio Salamanca

Punto	Coordenadas (longitud, latitud)
P1	(40.419456, -3.692797)
P2	(40.444799, -3.661669)

Con la ayuda de *Overpass turbo* se representa en la Figura 71 el resultado obtenido para ejemplo anterior.

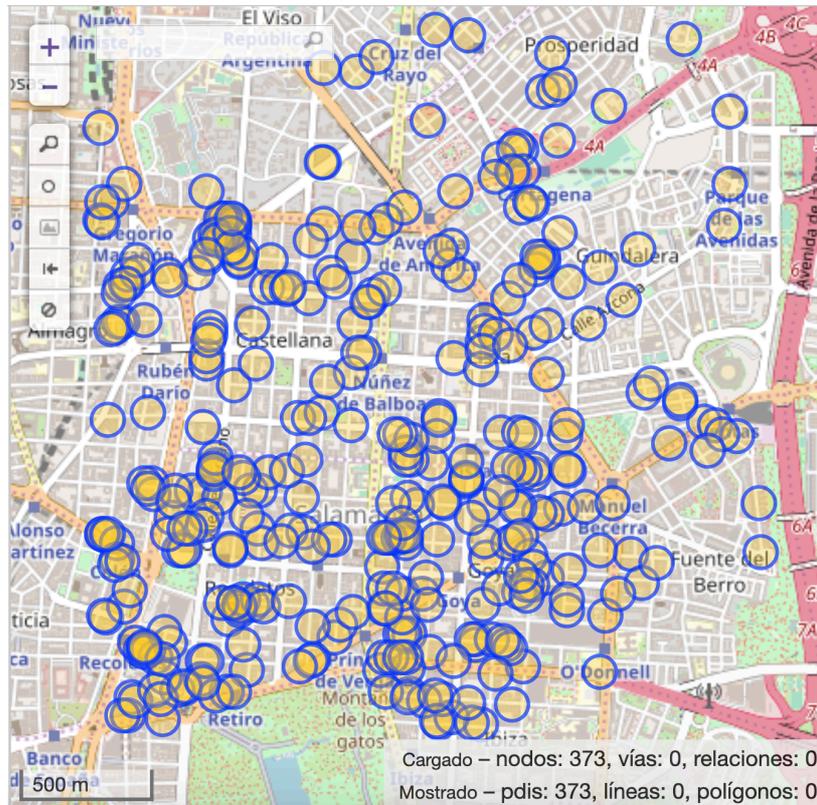


Figura 71: Datos solución de la consulta por bbox a los restaurantes del barrio Salamanca

8.1.2 CONSULTA MEDIANTE NOMBRE

En este caso, en vez de delimitar la región de estudio con un rectángulo, se delimita con el nombre de la región. Este método de consulta tiene la ventaja de ser muy rápida, no es necesario ninguna información de las coordenadas, solo el nombre de la región de estudio. Sin embargo, tiene dos desventajas. La primera es la falta de delimitación de las áreas (no todas las zonas que aparecen en OSM con nombre está delimitados en su base de datos). La segunda es la posible ambigüedad en el uso de un nombre para referirse a diferentes regiones (por ejemplo, el nombre Salamanca podría referirse a la ciudad o al barrio de Madrid). Por consiguiente, se recomienda utilizar este método a la hora de realizar consultas puntuales, no es muy fiable si se quiere estudiar zonas amplias con muchos nombres.

Al igual que en el caso anterior se va a dividir la url en tres partes.

- Raíz: `'http://overpassapi.de/api/interpreter?data=[out:json][timeout:200];'`
- Petición: Área +Formato +Zona+ Etiqueta.
- Cierre: `';out;'`

Utilizando el ejemplo anterior (descarga de los restaurantes del barrio Salamanca de Madrid), la url sería:

- Raíz: `'http://overpassapi.de/api/interpreter?data=[out:json][timeout:200];'`
- Área: `'area[name="Salamanca"]["admin_level"]->.barrio;area[name="Madrid"]["admin_level"]->.ciudad;'`
- Formato: `'node'`
- Zona: `'(area.barrio)(area.ciudad)'`
- Etiqueta: `'["restaurant"]'`
- Cierre: `';out;'`

Con la ayuda de *Overpass turbo* se representa en la

Figura 72 el resultado obtenido para el ejemplo anterior.

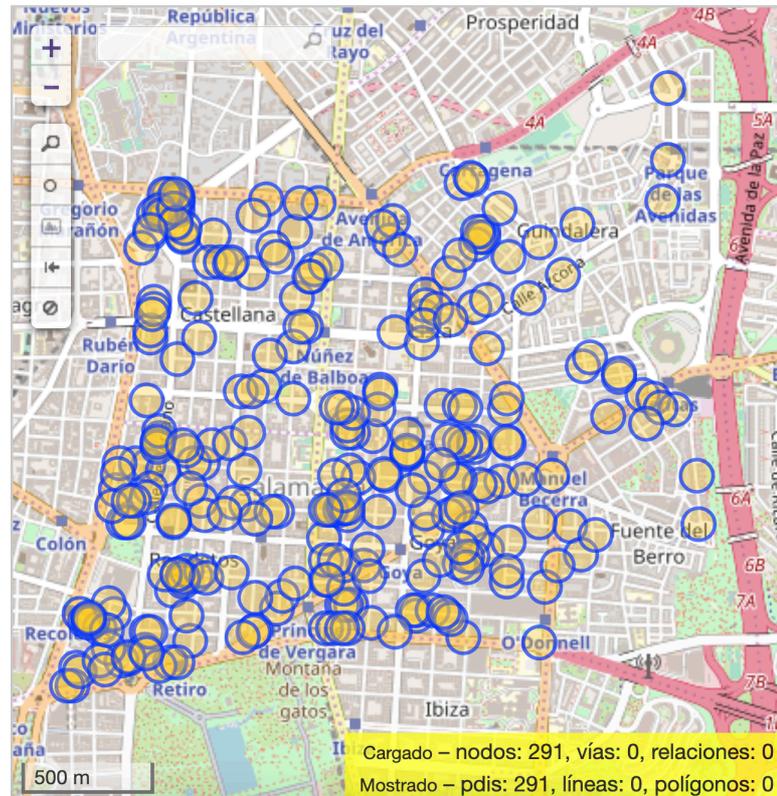


Figura 72: Datos solución de la consulta por nombre a los restaurantes del barrio Salamanca

8.1.3 CONSULTA POR POLÍGONO

Esta opción es muy útil para zonas de estudio irregulares ya que la forma en la que se acota la región es utilizando el polígono de su silueta. El mayor inconveniente es conseguir los puntos que delimitan la región, un trabajo laborioso si se hace a mano. Es por esto por lo que se aconseja utilizar esta opción para estudios que necesiten de mucha precisión a la hora de delimitar las zonas.

Al igual que los dos casos anteriores la url se va a dividir en tres partes:

- Raíz: '[http://overpassapi.de/api/interpreter?data=\[out:json\]\[time
out:200\];](http://overpassapi.de/api/interpreter?data=[out:json][timeout:200];)'
- Petición: Formato + Polígono + Etiqueta.
- Cierre: `';out;`

Se va a utilizar el ejemplo anterior de los restaurante del barrio Salamanca para ilustrar esta forma de consulta.

- Raíz: '[http://overpassapi.de/api/interpreter?data=\[out:json\]\[timeout:200\];](http://overpassapi.de/api/interpreter?data=[out:json][timeout:200];)'
- Formato: 'node'
- Polígono: (poly:"40.438082 -3.690419 40.435387 -3.688886 40.433014 -3.688463 40.425209 -3.690736 40.419456 -3.692797 40.420462 -3.687618 40.421548 -3.679902 40.421267 -3.669121 40.419295 -3.659661 40.430922 -3.6613 40.444799 -3.661669 40.439289 -3.672873 40.437439 -3.677894")
- Etiqueta: '["restaurant"]'
- Cierre: ';out;

En la Figura 73 y en la Tabla 3 se puede ver el procedimiento que se debe seguir para rellenar el apartado Polígono de la url.

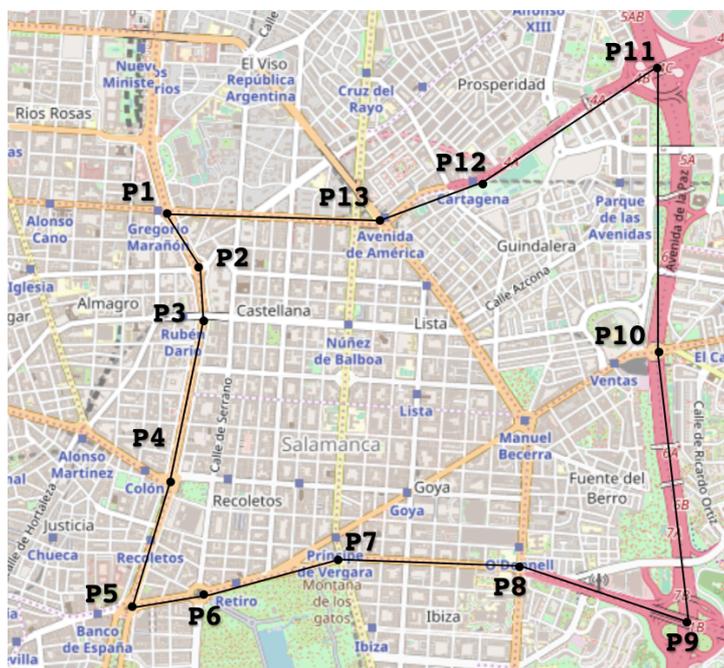


Figura 73: Delimitación barrio Salamanca mediante polígono

Tabla 3 : Puntos de acotación polígono barrio Salamanca

Puntos	Coordenadas (longitud, latitud)
P1	(40.438082, -3.690419)
P2	(40.435387, -3.688886)
P3	(40.433014, -3.688463)
P4	(40.425209, -3.690736)
P5	(40.419456, -3.692797)
P6	(40.420462, -3.687618)
P7	(40.421548, -3.679902)
P8	(40.421267, -3.669121)
P9	(40.419295, -3.659661)
P10	(40.430922, -3.6613)
P11	(40.444799, -3.661669)
P12	(40.439289, -3.672873)
P13	(40.437439 -3.677894)

Con la ayuda de Overpass turbo se representa los datos solución del ejemplo.

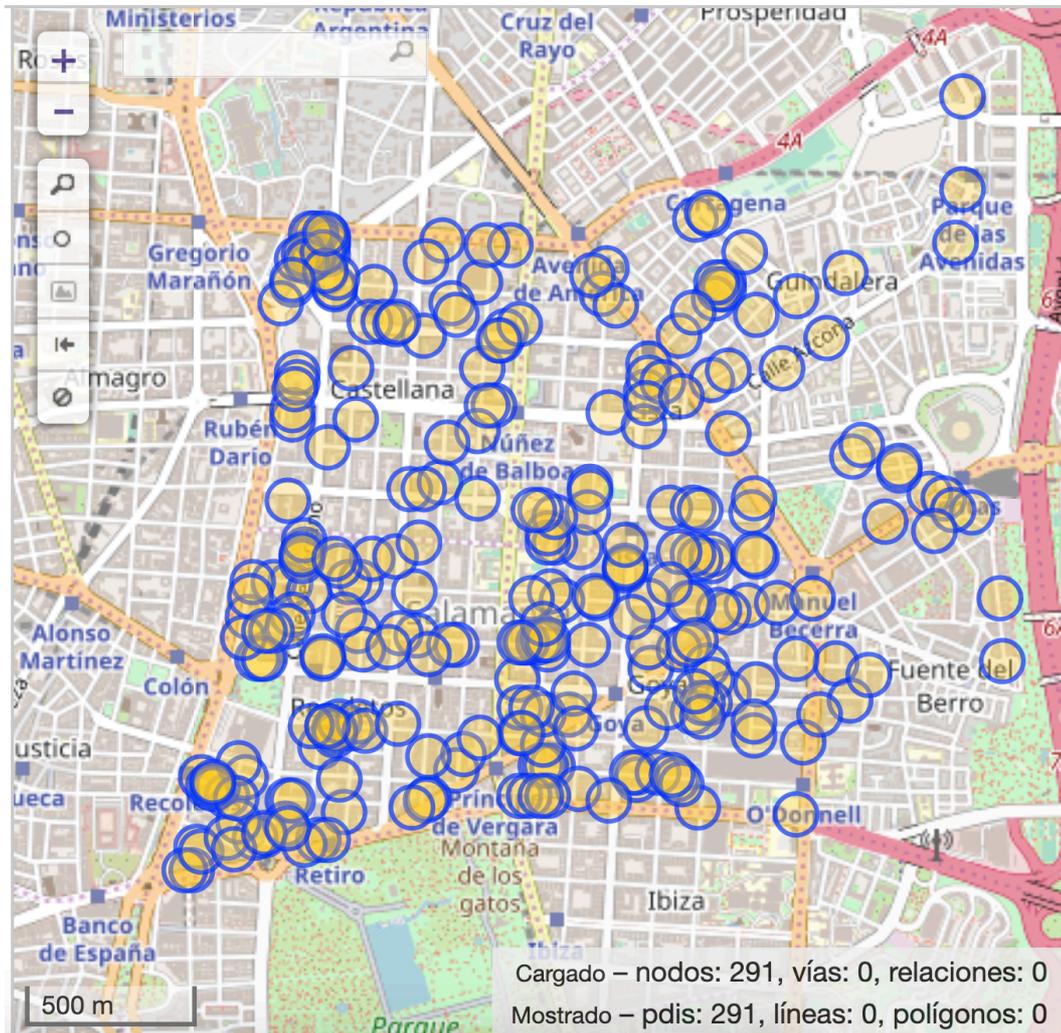


Figura 74: Datos solución de la consulta por polígono a los restaurantes del barrio Salamanca

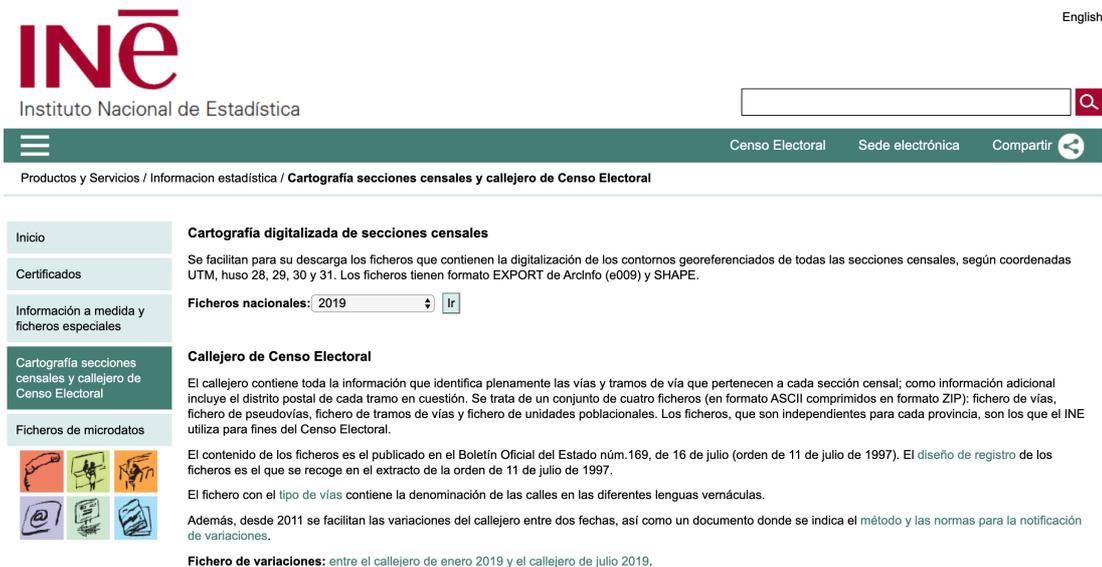
8.2 INSTITUTO NACIONAL DE ESTADÍSTICA (INE)

El Instituto Nacional de Estadística (INE) es un organismo autónomo dedicado a realizar los trabajos estadísticos de la Administración General del Estado, entre los campos que tratan destacan la demografía, economía y sociedad [11]. Este Instituto pone al alcance de los ciudadanos información actualizada sobre los servicios que realizan generando una base de datos pública que se ha aprovechado en este proyecto para analizar la organización espacial algunas variables como el número de habitantes. Estos datos son fácilmente descargables y totalmente gratuitos, solo hay que ir a la web y buscar la información que interese.

En este proyecto se ha utilizado la información de las fronteras de las secciones censales (SC), el número de habitantes y la renta media por hogar. A continuación se describe cada uno de estos conceptos.

8.2.1 CONSULTA DE LAS SECCIONES CENSALES

Para descargar esta información hay que ir a la web del INE y seguir los siguientes pasos *INE → Productos y Servicio → Información estadística → Cartografía secciones censales y callejero de Censo Electoral*. Una vez se ha llegado a esa sección hay dos partes, una llamada *Cartografía digitalizada de secciones censales* y otra con el nombre de *Callejero de Censo Electoral*. Hay que dirigirse a la primera parte y seleccionar en *Ficheros nacionales* el año que se deseé (ver Figura 75).



English

INE
Instituto Nacional de Estadística

Censo Electoral Sede electrónica Compartir

Productos y Servicios / Información estadística / Cartografía secciones censales y callejero de Censo Electoral

Inicio **Cartografía digitalizada de secciones censales**

Certificados Se facilitan para su descarga los ficheros que contienen la digitalización de los contornos georeferenciados de todas las secciones censales, según coordenadas UTM, huso 28, 29, 30 y 31. Los ficheros tienen formato EXPORT de ArcInfo (e009) y SHAPE.

Información a medida y ficheros especiales **Ficheros nacionales:** 2019 Ir

Cartografía secciones censales y callejero de Censo Electoral **Callejero de Censo Electoral**

El callejero contiene toda la información que identifica plenamente las vías y tramos de vía que pertenecen a cada sección censal; como información adicional incluye el distrito postal de cada tramo en cuestión. Se trata de un conjunto de cuatro ficheros (en formato ASCII comprimidos en formato ZIP): fichero de vías, fichero de pseudovías, fichero de tramos de vías y fichero de unidades poblacionales. Los ficheros, que son independientes para cada provincia, son los que el INE utiliza para fines del Censo Electoral.

Ficheros de microdatos El contenido de los ficheros es el publicado en el Boletín Oficial del Estado núm.169, de 16 de julio (orden de 11 de julio de 1997). El diseño de registro de los ficheros es el que se recoge en el extracto de la orden de 11 de julio de 1997.

El fichero con el **lipo de vías** contiene la denominación de las calles en las diferentes lenguas vernáculas.

Además, desde 2011 se facilitan las variaciones del callejero entre dos fechas, así como un documento donde se indica el método y las normas para la notificación de variaciones.

Fichero de variaciones: entre el callejero de enero 2019 y el callejero de julio 2019.

Figura 75: Web del INE con la información de las secciones censales

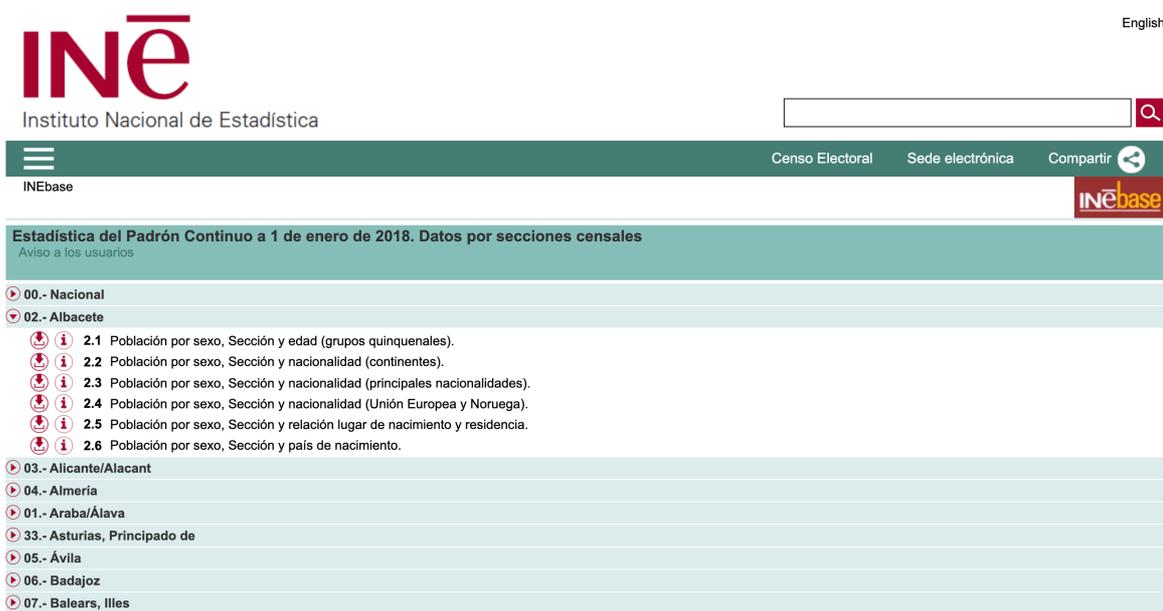
La carpeta que proporciona el INE se llama *seccionado_2019* y tiene 7 archivos en su interior. El archivo relevante para este proyecto es el archivo SECC_CE_20190101.shp. En este archivo se almacena la información de tal forma que se tiene por filas las sección censal y por columnas: Geometry, BoundingBox, X, Y, OBJECTID, CUSEC, CUMUN, CSEC, CDIS, CMUN, CPRO, CCA, CUDIS, CLAU2, NPRO, NCA, CNUT0, CNUT1, CNUT2, CNUT3, ESTADO, OBS, Shape_Leng, Shape_area, Shape_len, SUPERF_M2, NMUN.

Para optimizar los recursos de almacenamiento disponibles, se ha decidido almacenar únicamente los siguientes atributos:

- X: Vector con las coordenadas longitud de los puntos que delimitan la sc.
- Y: Vector con las coordenadas latitud de los puntos que delimitan la sc.
- CUSEC: El código de identificación de la sc.
- SUPERFICIE_M2: Los metros cuadrados que ocupa la sc.
- NPRO: Nombre de la provincia.
- NCA: Nombre de la comunidad autónoma.
- NMUN: Nombre del municipio.

8.2.2 CONSULTA DEL NÚMERO DE HABITANTES

Esta información se puede encontrar introduciendo en el buscador: “Estadística del Padrón Continuo a 1 de enero de 2018 Datos por secciones censales”. Una vez se haya llegado a la página del INE donde aparece esta información hay que seleccionar la provincia que se deseé (ver Figura 76). De los archivos que ofrecen seleccionar *Población por sexo, Sección y edad* (grupo quinquenales) y en la forma de descarga CSV separado por “;”.



The screenshot shows the INE website interface. At the top, there is a search bar and navigation links for 'Censo Electoral', 'Sede electrónica', and 'Compartir'. The main content area displays the title 'Estadística del Padrón Continuo a 1 de enero de 2018. Datos por secciones censales' and a list of available data files. The files are organized by province, with '02.- Albacete' selected. Under '02.- Albacete', several data files are listed, including '2.1 Población por sexo, Sección y edad (grupos quinquenales)', '2.2 Población por sexo, Sección y nacionalidad (continentes)', '2.3 Población por sexo, Sección y nacionalidad (principales nacionalidades)', '2.4 Población por sexo, Sección y nacionalidad (Unión Europea y Noruega)', '2.5 Población por sexo, Sección y relación lugar de nacimiento y residencia', and '2.6 Población por sexo, Sección y país de nacimiento'. Other provinces listed include Alicante/Alacant, Almería, Araba/Álava, Asturias, Principado de, Ávila, Badajoz, and Balears, Illes.

Figura 76: Web del INE con la información de habitantes

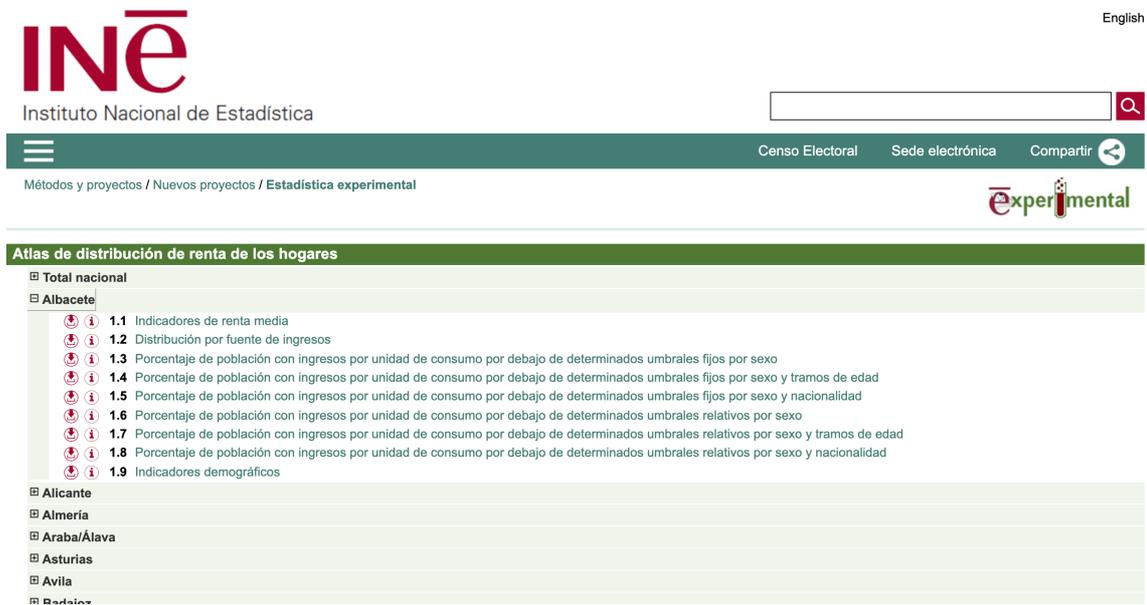
El archivo descargado contiene el número de los habitantes, en grupos quinquenales de edad (columnas), para cada sección censal (filas). Además de la división por edades, la información también está dividida por género. Primero se tiene toda la información de los habitantes (hombres y mujeres), en las siguientes filas la de los hombres y finalmente la de las mujeres. Para que se pueda leer fácilmente este archivo desde Matlab se debe hacer una serie de cambios en el fichero descargado antes de leerlo:

- Eliminar las 5 primeras líneas.
- Introducir en la línea 6 el nombre de la primera columna, ID_SC.
- Comentar con un % las líneas 7, 4424, 8839.

- Eliminar las últimas líneas del archivo.

8.2.3 CONSULTA DE LA RENTA MEDIA POR HOGAR

Esta información se encuentra en la página *web del INE* → *Métodos y proyectos* → *Nuevos proyectos* → *Estadística experimental* → *Atlas de distribución de renta de los hogares* → *más información* → *tabla de resultados* (ver Figura 77). Una vez se ha llegado a la tabla de resultado hay que seleccionar el archivo *Indicadores de renta media* y descargar en formato CSV separado por “;”.



The screenshot shows the INE website interface. At the top left is the INE logo and the text 'Instituto Nacional de Estadística'. On the right, there is a search bar and a language selector set to 'English'. Below the header is a navigation menu with options like 'Censo Electoral', 'Sede electrónica', and 'Compartir'. The main content area is titled 'Atlas de distribución de renta de los hogares' and lists various indicators under different categories like 'Total nacional' and 'Albacete'. The indicators include 'Indicadores de renta media', 'Distribución por fuente de ingresos', and 'Porcentaje de población con ingresos por unidad de consumo por debajo de determinados umbrales fijos por sexo y tramos de edad'.

Figura 77: Web del INE con la información de renta media por hogar

Este archivo contiene la información de la renta media por hogar y por persona de los años 2016 y 2015. Esta información está disponible para cada sección censal, distrito y municipio. En este proyecto se ha utilizado la renta media por hogar del año 2016 en cada sección censal como indicador económico de los barrios. Para más información sobre cómo se ha obtenido los datos leer el proyecto técnico [6].

Hay que realizar una serie de cambios en el archivo para que se pueda leer por Matlab.

- Introducir el nombre de las columnas (línea 6): ID_SC, Renta media por persona 2016, Renta media por persona 2015, Renta media por hogar 2016, Renta media por hogar 2015.

- Comentar con un % la línea 7.
- Eliminar las últimas líneas del archivo.

ANEXO II - METAS DE DESARROLLO SOSTENIBLE

“ La Agenda 2030 para el Desarrollo Sostenible, adoptada por todos los Estados Miembros de las Naciones Unidas en 2015, proporciona un plan compartido para la paz y la prosperidad de las personas y el planeta, ahora y en el futuro. En su corazón se encuentran los 17 Objetivos de Desarrollo Sostenible (ODS), que son un llamado urgente a la acción de todos los países, desarrollados y en desarrollo, en una asociación global. Reconocen que terminar con la pobreza y otras privaciones debe ir de la mano con estrategias que mejoren la salud y la educación, reduzcan la desigualdad y estimulen el crecimiento económico, todo mientras se aborda el cambio climático y se trabaja para preservar nuestros océanos y bosques.

Hoy, la División de Objetivos de Desarrollo Sostenible (DSDG) en el Departamento de Asuntos Económicos y Sociales de las Naciones Unidas (UNDESA) brinda apoyo sustantivo y desarrollo de capacidades para los ODS y sus temas temáticos relacionados, incluyendo agua , energía , clima , océanos , urbanización , transporte , ciencia y tecnología , el Informe Global de Desarrollo Sostenible (GSDR) , asociaciones y pequeños Estados insulares en desarrollo. El DSDG desempeña un papel clave en la evaluación de la implementación de la Agenda 2030 en todo el sistema de las Naciones Unidas y en las actividades de promoción y divulgación relacionadas con los ODS. Para hacer realidad la Agenda 2030, la propiedad general de los ODS debe traducirse en un fuerte compromiso de todas las partes interesadas para implementar los objetivos globales. DSDG tiene como objetivo ayudar a facilitar este compromiso. ” [18]

Metas de desarrollo sostenible



Figura 78: Metas de desarrollo sostenible

El modelo desarrollado en este proyecto busca extraer conocimiento relevante asociado a la distribución espacial de diferentes aspectos socioeconómicos de una ciudad para poder aportar soluciones a problemas que se puedan dar en la sociedad, algunos de estos problemas son los 17 objetivos del Desarrollo Sostenible (ODS). Debido a que el modelo desarrollado está enfocado al estudio de una ciudad, este proyecto está directamente relacionado con el objetivo 11- Sustainable cities and communities. También por la aplicación del modelo a la evolución espacial de la pandemia se podría relacionar con el objetivo 3 – Good Health and well-being.

Tabla 4: ODS con los que interactúa el proyecto.

DIMENSIÓN DE ODS	ODS	TFG
SOCIEDAD	ODS 11 - Hacer que las ciudades y los asentamientos sean inclusivos, seguros, resistentes y sostenibles.	Aportando soluciones creativas a posibles problemas que puedan surgir en la sociedad.
SOCIEDAD	ODS 3 - Garantizar vidas saludables y promover el bienestar para todos en todas las edades.	Estudiar la evolución espacial de una enfermedad para ayudar a generar un protocolo que asegure un reparto de los recursos sanitarios para garantizar el bienestar de la población.