

Capstone Project: Battle of the Neighbourhoods

Alex Lee

03/06/2021

1. Introduction

In this project, neighbourhoods in Madrid, Spain will be analysed to find candidate locations for a new hotel.

The ideal location will be close to the city centre and within a certain distance of the main attractions in Madrid.

Additionally, it will be not too close to other hotels and will also be in an area where real estate prices are not too high.

To accomplish this, an unsupervised clustering algorithm will be used to narrow down to a few candidate locations, from which a final location may be chosen.

2. Data

For this project, the data I will need for performing the analysis will be the following:

- Hotel location data, obtained from the Foursquare API, to find where hotels are located and how many hotels there are in each neighbourhood.
- Madrid main tourist attractions, obtained by web-scraping popular tourist information websites.
- Madrid average real estate prices by district, obtained from www.statista.com.

The first step in acquiring the data was finding the latitude and longitude of the centre of Madrid. From there, a square grid of points was created, equidistant from each other and in an area of 4km by 4km surround the centre of Madrid. The resulting grid can be viewed in Figure 1.

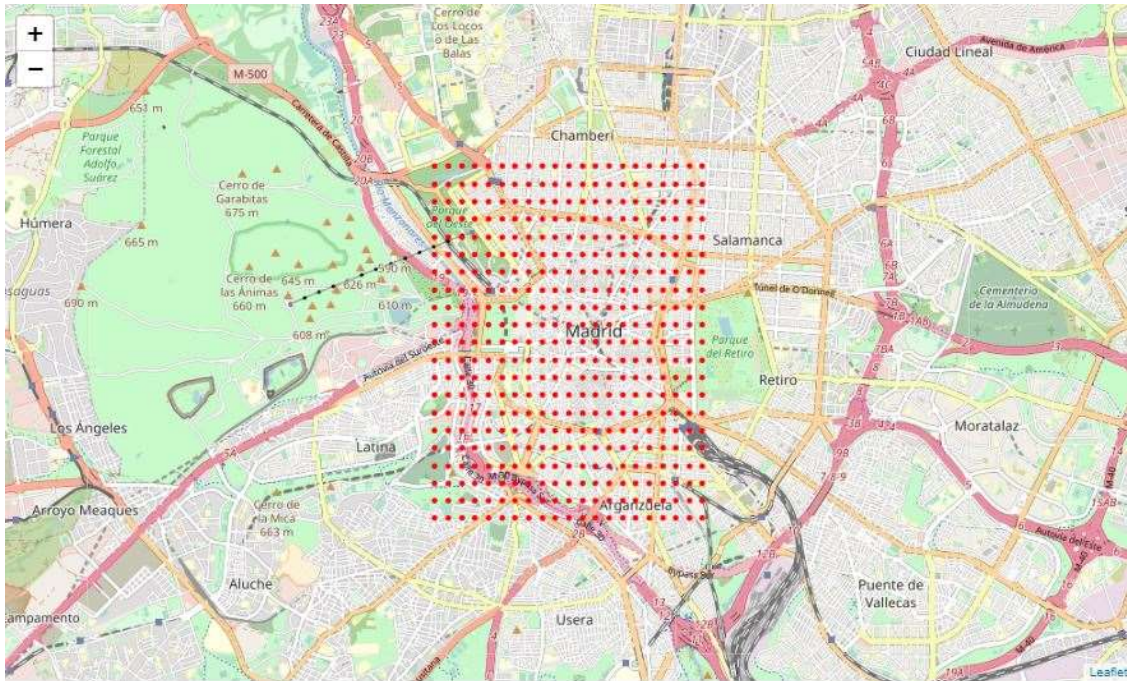


Figure 1: Grid of points in Madrid

For each of these points, a request was sent to the Foursquare API, to obtain information on nearby hotels.

In total, 356 Hotels were found in the central area of Madrid. To visualize the locations of these hotels, a Heat Map was created, which can be viewed in Figure 2.

In addition, a popular tourist information website (TripAdvisor) was scraped to obtain information on the top 10 tourist attractions in Madrid. The locations of these attractions can be viewed in Figure 3.

Finally, data was also gathered on average prices of real estate by district in Madrid. The data was scraped from www.statista.com, using the BeautifulSoup library.

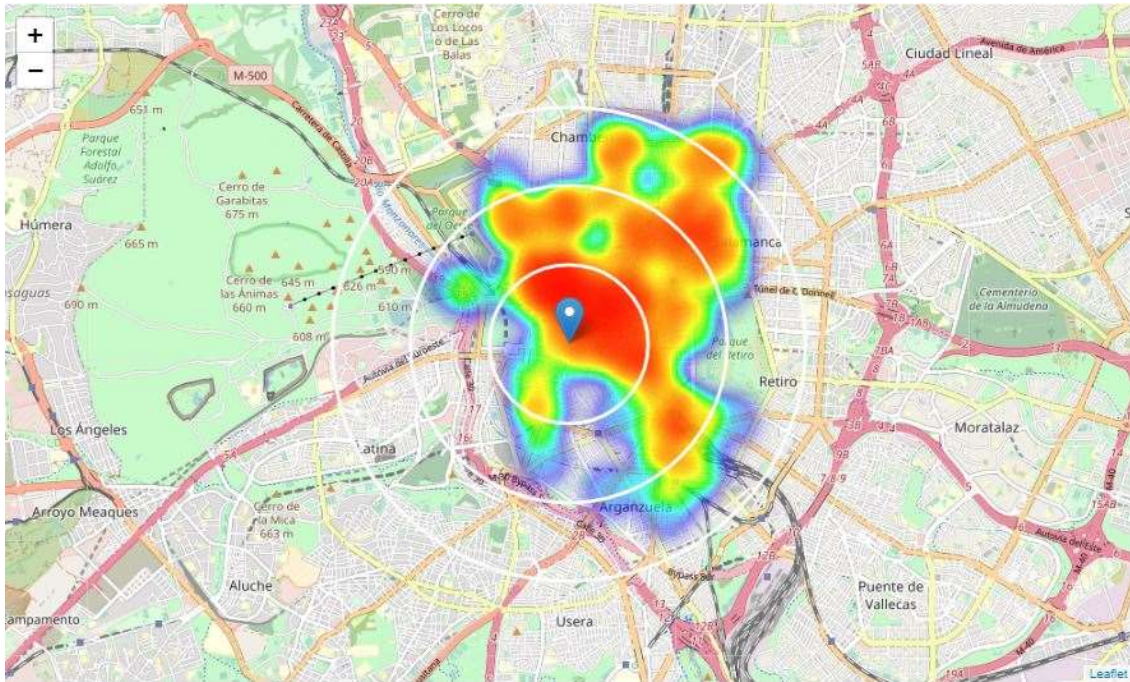


Figure 2: Heat Map of Hotel locations

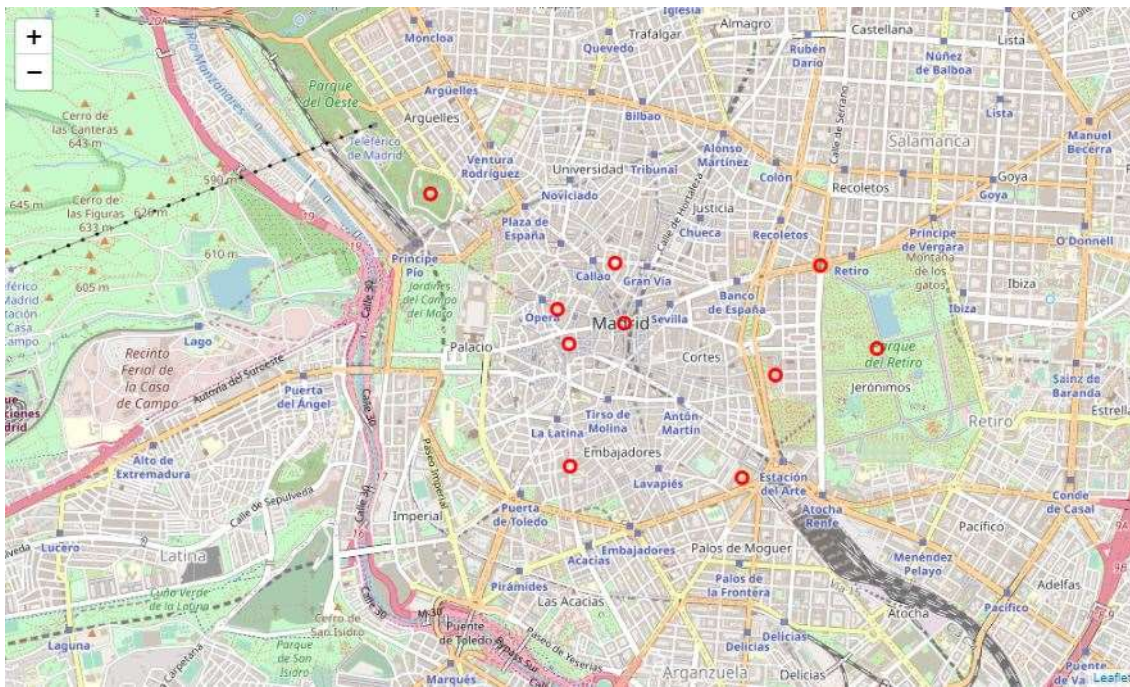


Figure 3: Main tourist attractions in Madrid

3. Methodology

The aim of this project was to find areas of low hotel density, low price but high accessibility in a region about 3km in radius from the centre of Madrid.

The first step was obtaining data needed for the analysis:

- From the Foursquare API: on the location and number of hotels in the neighbourhood of each point in our square grid of points surrounding the centre of Madrid.
- On the locations of the main tourist attractions in Madrid.
- On real estate prices for each of Madrid's districts.

The second step was choosing a metric for each of these three factors to find the main areas of possible locations for a new hotel.

The final step was clustering these possible areas to find cluster centres, which would be ideal locations for a new hotel, which would then be shown to interested stakeholders, for them to perform a final search in those locations.

4. Analysis

With the data gathered as described above, a Pandas Dataframe was created with each of the points in the grid as a row.

The distance from each of these points to the main tourist attractions in Madrid was calculated and appended to the DataFrame.

Then, the number of hotels within a 200m distance was calculated for each point, using the Foursquare API, and stored in the DataFrame.

Finally, using the geopy library, the closest district centre to each of the points was calculated and thus each point was classified into a district. The average price per square metre of real estate was then appended to the DataFrame.

To choose a subset of this grid of points which will contain locations of interest for the new hotel, the following conditions were chosen:

- The distance to all main attractions in Madrid must be less than 3km.
- The number of hotels within 200m must be less than 5.
- The average price per square metre must be less than €4500.

The result is a subset of 121 out of 441 initially chosen points (27.4%) with low hotel density, good location and reasonable price. The resulting points can be viewed in Figure 4.

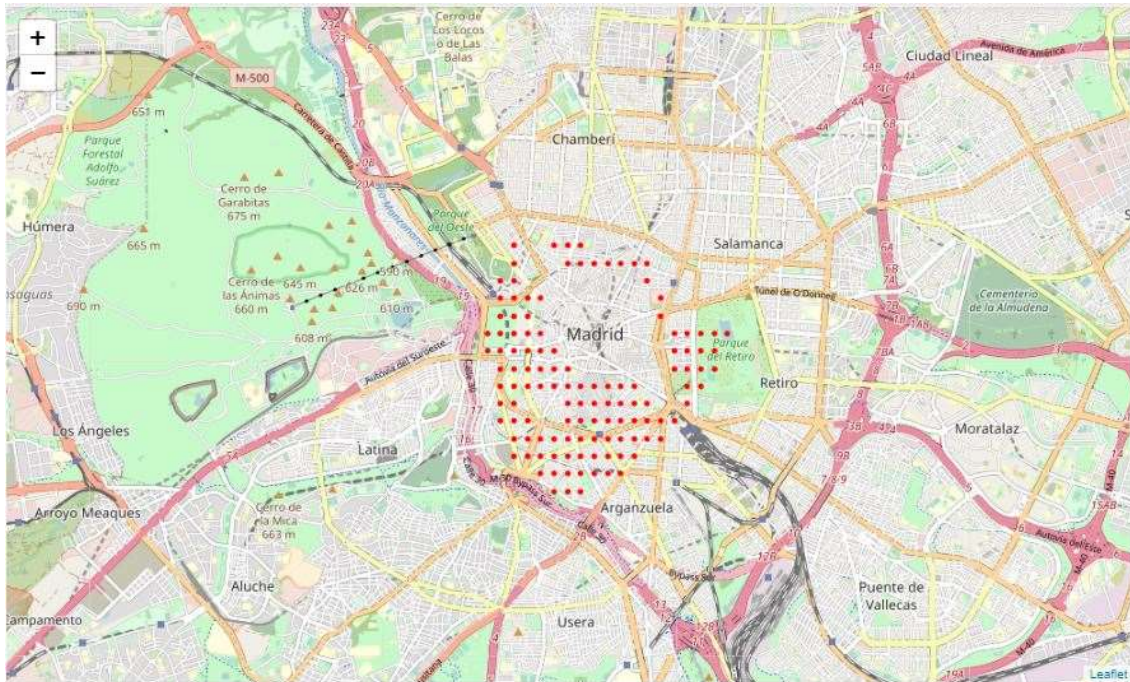


Figure 4: Candidate points remaining after restrictions are applied.

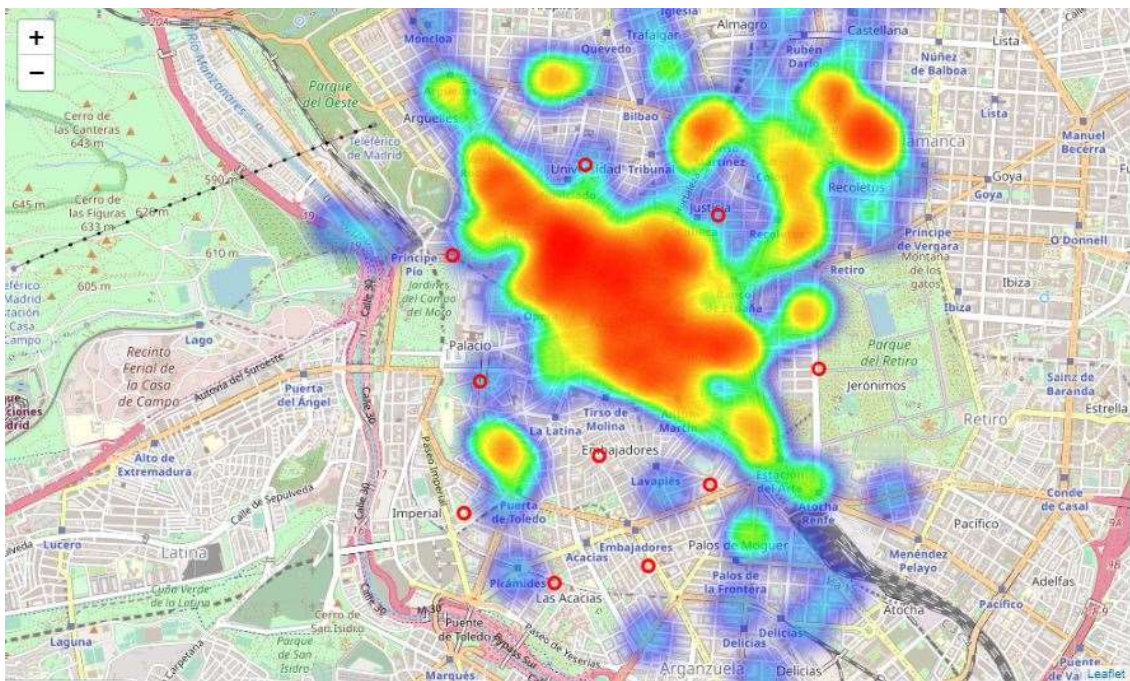


Figure 5: Cluster centres of good potential new hotel locations (in red) in relation to existing hotel locations (given by the Heat Map).

A Clustering Machine Learning algorithm was then applied to find cluster centres of good locations, which will serve as a final candidate list of locations for a new hotel.

A value of 10 was chosen to obtain a list of 10 candidate areas of high density of good locations. These can be viewed (in red circles) in Figure 5, where the Heat Map of existing hotel locations is superimposed to visualize where the cluster centres are in relation to existing hotels.

The addresses for these cluster centres were then found by reversing the latitude and longitude values and are as follows:

- Paseo Doctor Vallejo Nájera 25, 25, Paseo de Juan Antonio Vallejo-Nájera Botas, Arganzuela, Imperial, Madrid, Área metropolitana de Madrid y Corredor del Henares, Comunidad de Madrid, 28005, España
- Paseo del Marqués de Pontejos, Jerónimos, Retiro, Madrid, Área metropolitana de Madrid y Corredor del Henares, Comunidad de Madrid, 28009, España
- Viaducto de Segovia, Calle de Bailén, Palacio, Madrid, Área metropolitana de Madrid y Corredor del Henares, Comunidad de Madrid, 28005, España
- 39, Calle del Espíritu Santo, Universidad, Centro, Madrid, Área metropolitana de Madrid y Corredor del Henares, Comunidad de Madrid, 28004, España
- 16, Calle de los Cabestreros, Embajadores, Madrid, Área metropolitana de Madrid y Corredor del Henares, Comunidad de Madrid, 28012, España
- Colegio Público Joaquín Costa, Travesía de Gil Imón, Arganzuela, Imperial, Madrid, Área metropolitana de Madrid y Corredor del Henares, Comunidad de Madrid, 28005, España
- 26, Cuesta de San Vicente, Argüelles, Moncloa-Aravaca, Madrid, Área metropolitana de Madrid y Corredor del Henares, Comunidad de Madrid, 28008, España
- 23, Calle de Martín de Vargas, Arganzuela, Las Acacias, Madrid, Área metropolitana de Madrid y Corredor del Henares, Comunidad de Madrid, 28005, España
- 23, Calle de Gravina, Justicia, Madrid, Área metropolitana de Madrid y Corredor del Henares, Comunidad de Madrid, 28004, España
- La canival, 28, Calle de Argumosa, Lavapiés, Embajadores, Madrid, Área metropolitana de Madrid y Corredor del Henares, Comunidad de Madrid, 28012, España

Finally, to give an idea of how much initial investment it would take to purchase real estate in one of these locations, a table was created with different numbers of rooms:

| ROOMS | INITIAL INVESTMENT |
|-------|--------------------|
| 10 | €829,540.00 |
| 20 | €1,659,080.00 |
| 30 | €2,488,620.00 |
| 40 | €3,318,160.00 |
| 50 | €4,147,700.00 |
| 60 | €4,977,240.00 |
| 70 | €5,806,780.00 |
| 80 | €6,636,320.00 |
| 90 | €7,465,860.00 |
| 100 | €8,295,400.00 |

5. Results and Discussion

The result of the analysis is that, while there are many hotels in the centre of Madrid, there are plenty of areas of low hotel density within a 3km region of the centre.

These regions were then clustered using K-Means, to obtain a list of 10 possible areas for a location for a new hotel, which are all areas of low hotel density, are within 3km distance of all major Madrid tourist attractions and are reasonably priced (less than €4500 per square metre).

These locations are not necessarily good locations for a hotel - there might not be any real estate available there, for example - but are excellent starting points for anybody interested in opening a hotel in Madrid.

Then, an estimation of initial investment required was made based on real estate price data and average hotel size. For example, a 100-room hotel in one of these locations would require an initial investment of approximately €8 million to purchase the property.

6. Conclusion

The aim of this project was to identify some candidate locations for a new hotel in Madrid, based on three main ideas: not being too close to existing hotels, being close to main tourist attractions and the price of the real estate not being too high.

Using the data acquired, many locations were found that satisfied all the criteria. A clustering algorithm was then used to create 10 cluster centres based on those locations.

This list of 10 locations will be passed to interested parties, as a guide, so they can make informed decisions on which neighbourhoods to look for properties in. To make their final decision on a hotel location, however, they will consider other factors which were not included in this study, such as availability of property, attractiveness of locations, available capital, etc.