

The Battle of Neighborhoods

IBM CAPSTONE PROJECT – APPLIED DATA SCIENCE

Pablo de la Fuente Aguilera | 2020

1. Introduction to the Problem

1.1. BACKGROUND

Nowadays, entrepreneurs have a great deal of information before starting their projects. The objective of this project is to use the knowledge learned from Data Science and apply it to a fictitious problem when opening a business in a city.

This will be done using the Python language, which provides numerous tools to acquire data from various sources, process the information and draw conclusions.

The objective is to develop tools that help to make a better analysis of the problem environment and to take an optimal solution.

1.2. PROBLEM

A friend of us who wants to start a new and successful business in the area of car parking and car maintenance service. He's an architect and he is considering the idea to build an underground parking. He needs to choose a location in order to maximize the incomes of this business.

1.3. INTEREST

We are thinking about building an underground parking in Madrid. This could be a good business opportunity but we need to carry out a market research in order to establish a long-term success.

To start with, we will analyse the neighborhoods, their existing hotels and rental services in the different neighborhoods. To do that we will use Foursquare venues and information of the geographical location of every restaurant.

After that we will sort them by neighborhood, in order to identify the best possible location and determine which neighborhood inside Madrid could be the best place. We will cluster the possible candidates in groups so we can address better the problem and its understanding.

At the end, we will identify, based on filtering and ponderation of candidates, the best possible location in Madrid. We will support our decision in a map, to give a easy understanding of the problem.

2. Data Acquisition and cleaning

We will use data from the following sources:

- Govern of Madrid Website <https://datos.comunidad.madrid>
- Website of Idealista (biggest website for renting and selling a house or flat in Spain) <https://www.idealista.com/>
- API Foursquare to ge the venues of the near business

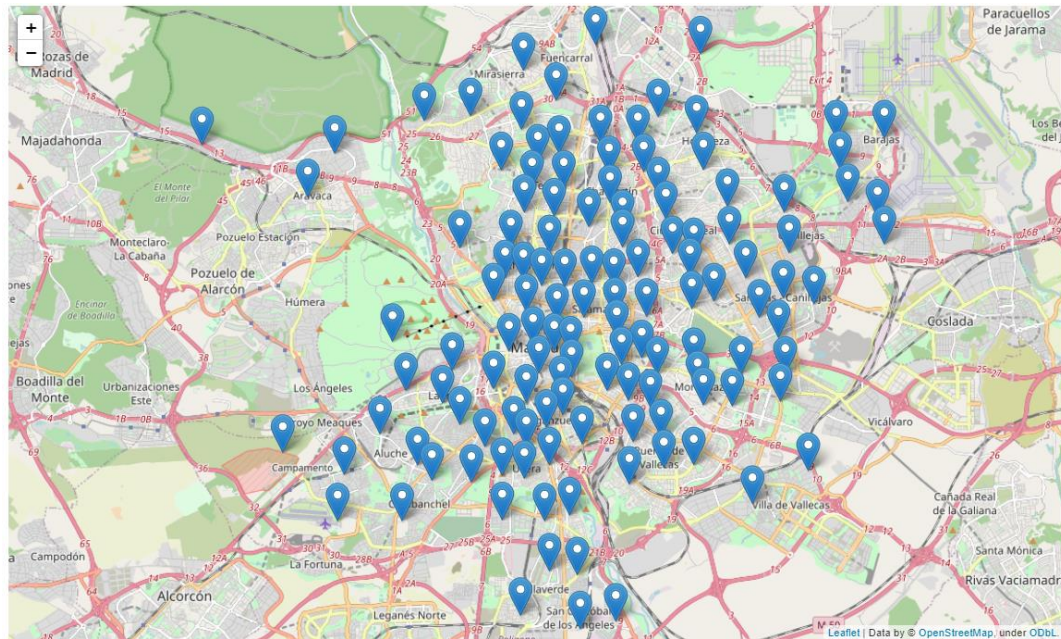
2.1. DATA CLEANING

A dataframe with the following columns is generated with information of the neighborhoods in Madrid City:

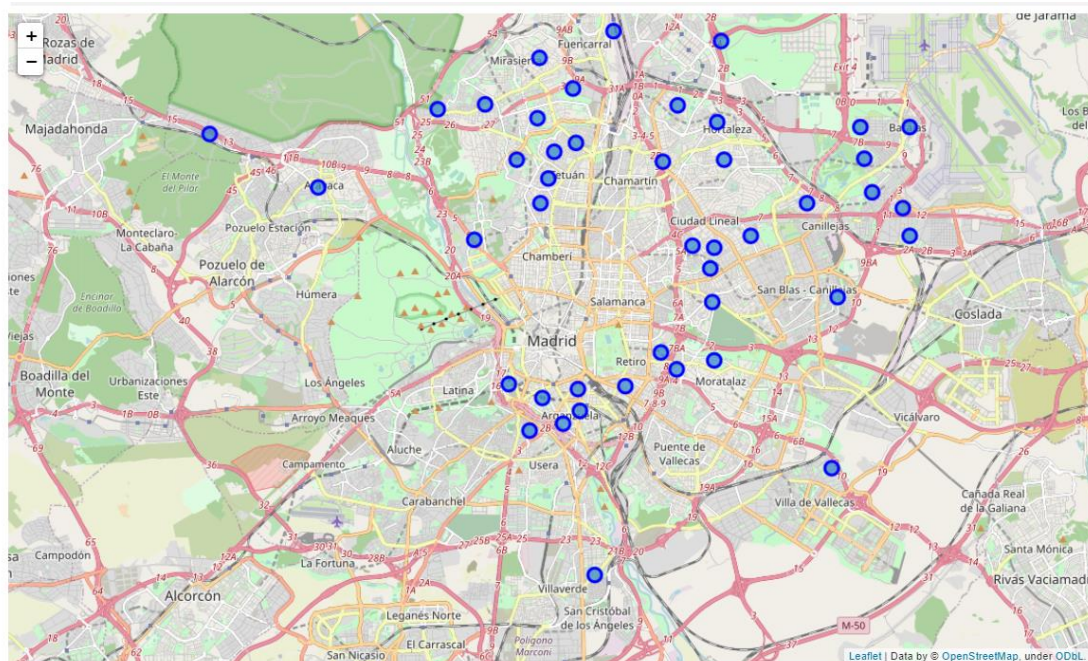
	District	Neighborhood	Latitude	Longitude	Surface (km2)	Density (hab/km2)	Price (€/m2)
0	Centro	Palacio	-3.713134	40.415325	1.46	15323.287671	4852.0
1	Centro	Embajadores	-3.702543	40.409444	1.03	43345.631068	4479.0
2	Centro	Cortes	-3.696785	40.415439	0.59	17850.847458	5272.0
3	Centro	Justicia	-3.695976	40.423497	0.75	21866.666667	5893.0
4	Centro	Universidad	-3.706963	40.426121	0.93	33051.612903	5282.0

This Dataframe has 7 columns and 128 rows in total, one row for each Neighborhood.

Using Folium we can have a visual approach to the location of our data:



Due to budget restrictions, we apply a filter between 2600 and 4200 €/m². With this change we have now only 44 neighborhoods from the initial 128.



We use Data Science K-Means Clustering Procedure to cluster the candidate Neighborhoods into 4 different groups and get with this new approach a better understanding of our location challenge:

Cluster Nr. 0

	Cluster Labels	District	Neighborhood	Latitude	Longitude	Surface (km2)	Density (hab/km2)	Price (€/m2)
12	0	Fuencarral-El Pardo	Peña Grande	-3.726833	40.479017	2.86	15476.923077	3283.0
14	0	Fuencarral-El Pardo	La Paz	-3.696045	40.483378	2.17	15578.801843	3580.0
22	0	Moratalaz	Marroquina	-3.647142	40.411364	1.75	15633.714286	2868.0
23	0	Moratalaz	Media Legua	-3.660231	40.408978	1.02	17466.666667	2664.0
24	0	Ciudad Lineal	Ventas	-3.647657	40.426883	3.22	14700.931677	2665.0
27	0	Ciudad Lineal	San Pascual	-3.654677	40.441785	1.06	17079.245283	3406.0
30	0	Hortaleza	Canillas	-3.643754	40.464324	2.56	15820.312500	3241.0
32	0	Hortaleza	Apostol Santiago	-3.659881	40.478759	1.20	12627.500000	2850.0
35	0	Villa de Vallecas	Santa Eugenia	-3.606203	40.382919	2.04	11822.058824	2695.0
39	0	Barajas	Alameda de Osuna	-3.592139	40.455778	1.98	9795.454545	3273.0
41	0	Barajas	Casco Histórico de Barajas	-3.579222	40.472985	0.64	11415.625000	3104.0

Cluster Nr. 1

	Cluster Labels	District	Neighborhood	Latitude	Longitude	Surface (km2)	Density (hab/km2)	Price (€/m2)
1	1	Arganzuela	Las Acacias	-3.706734	40.401523	1.10	33318.181818	4188.0
2	1	Arganzuela	La Chopera	-3.699556	40.394639	0.56	35276.785714	3625.0
4	1	Arganzuela	Palos de Moguer	-3.694600	40.403759	0.65	39372.307692	3943.0
5	1	Retiro	Pacífico	-3.677976	40.404721	0.76	44271.052632	4149.0
7	1	Tetuan	Bellas Vistas	-3.707527	40.453071	0.73	38923.287671	3365.0
10	1	Tetuan	Berruguete	-3.704891	40.459591	0.60	40533.333333	3263.0
13	1	Fuencarral-El Pardo	El Pilar	-3.708572	40.475544	1.37	33451.824818	3303.0
21	1	Carabanchel	Comillas	-3.711391	40.392854	0.67	33000.000000	2702.0
25	1	Ciudad Lineal	Quintana	-3.648602	40.435811	0.71	34056.338028	2828.0

Cluster Nr. 2

	Cluster Labels	District	Neighborhood	Latitude	Longitude	Surface (km2)	Density (hab/km2)	Price (€/m2)
11	2	Fuencarral-El Pardo	Fuente Arreina	-3.743380	40.477849	1.51	2113.907285	3439.0
15	2	Fuencarral-El Pardo	Valverde	-3.682246	40.498379	9.03	6642.857143	3462.0
16	2	Fuencarral-El Pardo	Mirasierra	-3.707996	40.491460	6.91	4465.701881	3645.0
17	2	Moncloa-Aravaca	Ciudad Universitaria	-3.730594	40.443199	14.14	1135.007072	3652.0
19	2	Moncloa-Aravaca	El Plantío	-3.822546	40.471137	3.56	764.887640	3742.0
20	2	Moncloa-Aravaca	Aravaca	-3.784802	40.457048	5.86	4384.300341	3742.0
28	2	Ciudad Lineal	Atalaya	-3.665093	40.463866	0.25	6240.000000	3020.0
29	2	Hortaleza	Palomas	-3.614780	40.452846	1.14	5859.649123	3732.0
33	2	Hortaleza	Valdefuentes	-3.644558	40.495833	15.19	3400.526662	3813.0
36	2	San Blas-Canillejas	Rosas	-3.604156	40.428114	9.26	3403.023758	3121.0
37	2	San Blas-Canillejas	Rejas	-3.579308	40.444413	4.98	3165.662651	3022.0
38	2	San Blas-Canillejas	El Salvador	-3.634497	40.444446	1.86	6112.903226	3074.0
40	2	Barajas	Aeropuerto	-3.581604	40.451745	19.87	89.884248	3203.0
42	2	Barajas	Tiñán	-3.596474	40.472952	16.40	672.012195	3088.0
43	2	Barajas	Corralejos	-3.594800	40.464725	4.67	1577.087794	3476.0

Cluster Nr. 3

	Cluster Labels	District	Neighborhood	Latitude	Longitude	Surface (km2)	Density (hab/km2)	Price (€/m2)
0	3	Arganzuela	Imperial	-3.718656	40.405161	0.98	23105.102041	3922.0
3	3	Arganzuela	Delicias	-3.693955	40.398093	1.07	25485.046729	3904.0
6	3	Retiro	Estrella	-3.665524	40.413454	1.02	22701.960784	4155.0
8	3	Tetuan	Almenara	-3.695287	40.468889	0.99	22232.323232	3419.0
9	3	Tetuan	Valdeacederas	-3.702735	40.466435	1.17	21574.358974	3099.0
18	3	Moncloa-Aravaca	Valdezarza	-3.715617	40.464359	1.44	20317.361111	2989.0
26	3	Ciudad Lineal	Concepción	-3.647228	40.441070	0.90	22828.888889	3259.0
31	3	Hortaleza	Pinar del Rey	-3.646148	40.474450	2.64	19682.954545	2887.0
34	3	Villaverde	Los Rosales	-3.688816	40.354694	1.51	24163.576159	2869.0

We explore the different categories of the Venues around our Neighborhoods using the API of Foursquare. Foursquare lets users search for restaurants, nightlife spots, shops and other places in a location. The app displays personalized recommendations based on factors that include the time of day, a user's check-in history, their "Tastes" and their venue ratings.

	name	id
0	Arts & Entertainment	4d4b7104d754a06370d81259
1	Amphitheater	56aa371be4b08b9a8d5734db
2	Aquarium	4fceea171983d5d06c3e9823
3	Arcade	4bf58dd8d48988d1e1931735
4	Art Gallery	4bf58dd8d48988d1e2931735

	name	id
0	Bike Rental / Bike Share	4e4c9077bd41f78e849722f9
1	Hotel	4bf58dd8d48988d1fa931735
2	Rental Car Location	4bf58dd8d48988d1ef941735
3	Rental Service	56aa371be4b08b9a8d573552
4	Taxi	4bf58dd8d48988d130951735
5	Taxi Stand	53fca564498e1a175f32528b

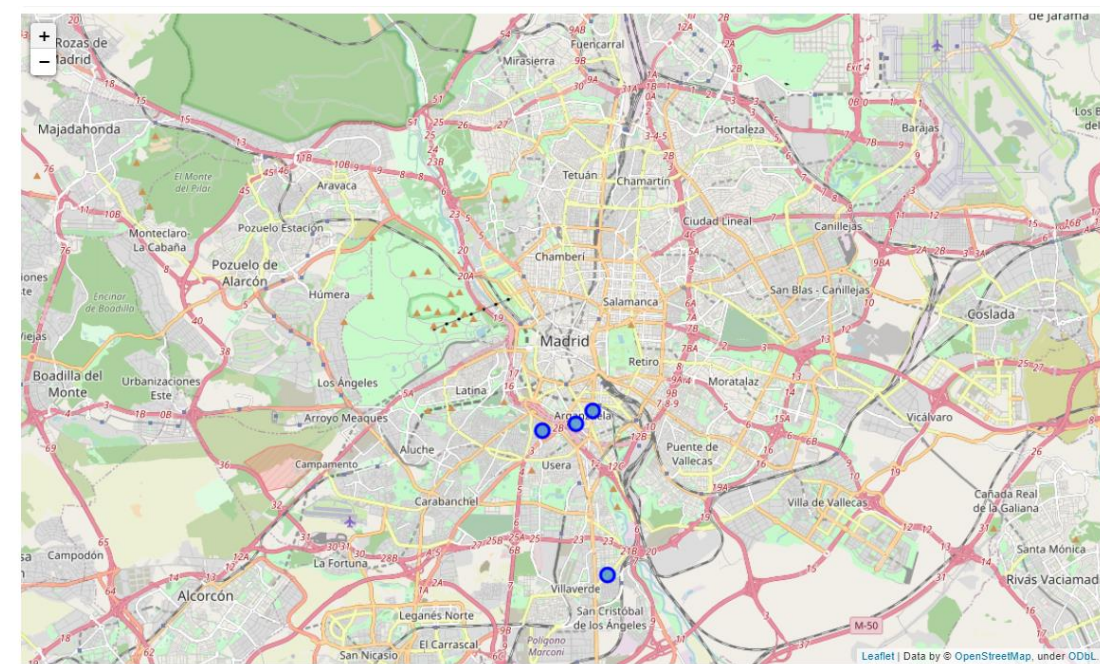
We establish a classification based on a point system which gives points according to their near services and we sort the resulting dataframe based on these points:

	Hotel	Rental Car	Location	Points
Neighborhood				
La Chopera	1		2	3
Delicias	1		2	3
Los Rosales	1		2	3
Comillas	1		2	3
Santa Eugenia	1		1	2
Rosas	1		1	2
Rejas	1		1	2
Palos de Moguer	0		2	2
Palomas	1		1	2
Marroquina	1		1	2
Las Acacias	0		2	2
Alameda de Osuna	1		1	2
Imperial	0		2	2

4. Results

As a result of our analysis we have four major candidates to be the location of our business. We create a new dataframe taking all the information of those four candidates.

	District	Neighborhood	Latitude	Longitude	Surface (km2)	Density (hab/km2)	Price (€/m2)
8	Arganzuela	La Chopera	-3.699556	40.394639	0.56	35276.785714	3625.0
10	Arganzuela	Delicias	-3.693955	40.398093	1.07	25485.046729	3904.0
65	Carabanchel	Comillas	-3.711391	40.392854	0.67	33000.000000	2702.0
109	Villaverde	Los Rosales	-3.688816	40.354694	1.51	24163.576159	2869.0



Presenting them in a Map with Folium it is to appreciate that one of them, Los Rosales, is further away from the city center as the other candidates.

It should be noted that the location with respect to the city center is always an important factor in our choice, since this distance determines, to a large extent, the number of people who circulate in the nearby streets.

Since our business has as one of its main clients: tourists, this aspect takes on great importance.

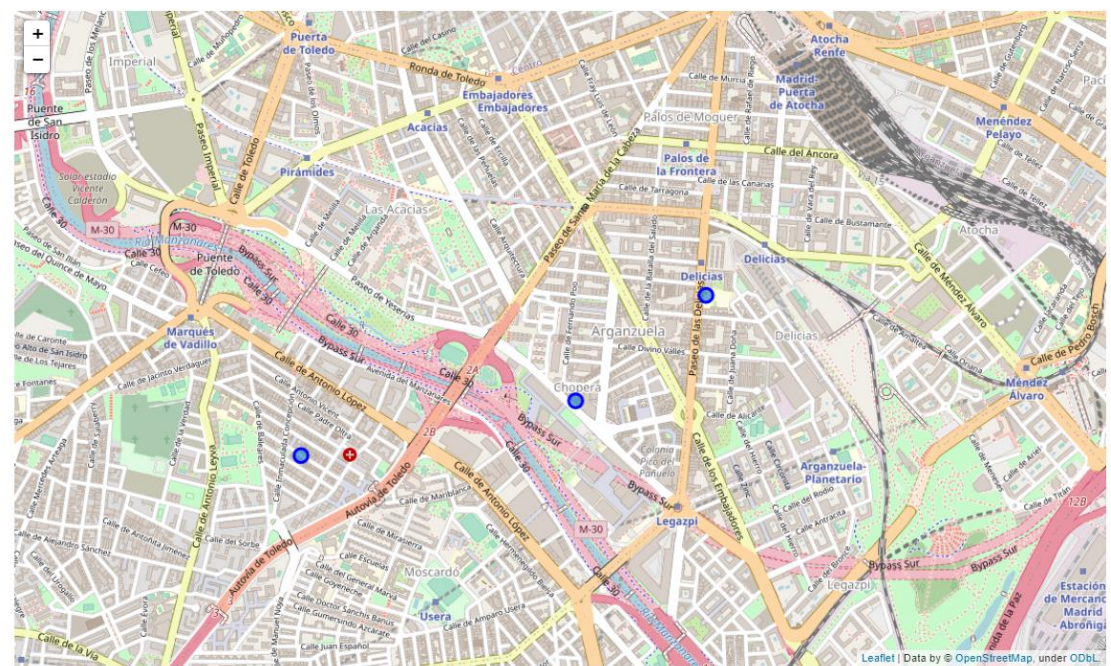
Thus, Los Rosales is discarded for future deliberations.

5. Conclusion

At the end, we have three possible locations.

	District	Neighborhood	Latitude	Longitude	Surface (km2)	Density (hab/km2)	Price (€/m2)
8	Arganzuela	La Chopera	-3.699556	40.394639	0.56	35276.785714	3625.0
10	Arganzuela	Delicias	-3.693955	40.398093	1.07	25485.046729	3904.0
65	Carabanchel	Comillas	-3.711391	40.392854	0.67	33000.000000	2702.0

Given the proximity of the three neighborhoods, it is determined that the characteristics of population density will not be taken into account since they share practically the same audience. We can see this on the map:



In the map we can see that Delicias and Chopera have a really good location, between the highway Calle 30 (which surrounds with a circular ring all the center of the city) and Atocha Train Station (the biggest Train Station in Spain).

As a result, we will build our parking in the marked area in red:

