

## 1. INTRODUCTION

Madrid is the capital and the most populated city in Spain. The city has almost 3.3 million inhabitants and a metropolitan area population of approximately 6.5 million. It concentrates activities directly related to power and to knowledge and technological innovation. It is one of the largest financial centers in Europe and the largest in Spain.

### BUSSINES POBLEM

Madrid is administratively divided into 21 districts, which in turn are subdivided into 131 neighborhoods.

Are there differences from one district to another? How are businesses distributed in each of your districts? Where would it be more prosperous to open a new restaurant?

In this analysis we will try to find an optimal location for a restaurant in Madrid taking into account the competitors and which group of people with income will be attracted according to the population of the neighborhood. To do this we will use an unsupervised automatic learning algorithm (K-Means) and will take into account demographic data, per capita income by district and housing prices.

## 2. DATA

Based on the definition of our problem. The necessary data sources have been obtained from:

- **Foursquare API:** Collection of restaurant related data (location, distribution, type of food).
- **National Institute of Statistics (INE):** Official site for downloading demographic data, per capita income and housing prices.
- **Madrid City Council open data portal:** Official site for downloading data related to Madrid.

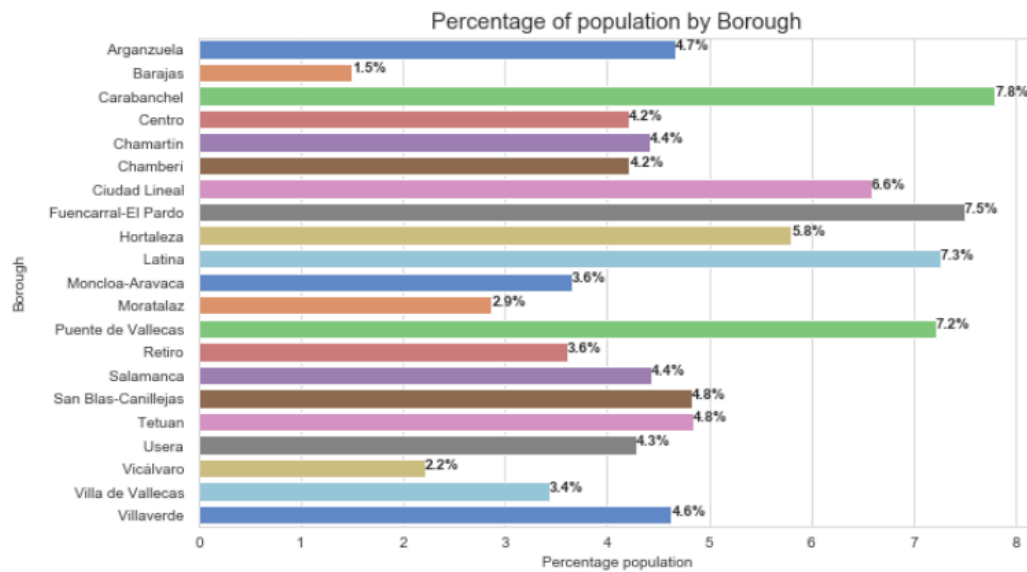
## 3. METHODOLOGY

### GETTING, EXPLORING AND ANALYZING THE DATA

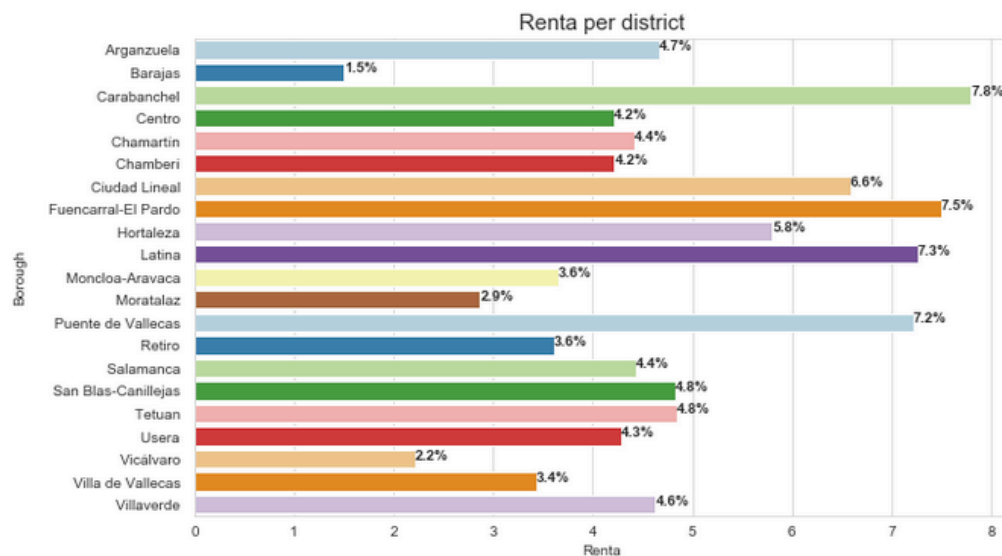
After downloading the data from the sites indicated in the previous section we started reading, exploring and analysing the datasets. This step was completed with a single dataset composed of Borough, Neighborhood, Population, RentalNE, AvgHousingPrice of each of the boroughs.

	Borough	Neighborhood	Population	RentalNE	AvgHousingPrice
0	Centro	Palacio, Embajadores, Co...	140.473	16147	5479.500000
1	Arganzuela	Imperial, Las Acacias, L...	155.660	17306	4547.750000
2	Retiro	Pacífico, Adelfas, Estre...	120.406	21504	5178.416667
3	Salamanca	Recoletos, Goya, Fuente ...	147.854	24433	6352.916667
4	Chamartín	El Viso, Prosperidad, Ci...	147.551	25969	5944.666667
5	Tetuan	Bellas Vistas, Cuatro Caminos, ...	161.313	14970	4333.708333
6	Chamberí	Gaztambide, Arapiles, Tr...	140.866	22499	6040.750000
7	Fuencarral-El Pardo	El Pardo, Fuentelarreina, ...	249.973	18573	4322.333333
8	Moncloa-Aravaca	Casa de Campo, Argüelles, ...	121.683	22152	4676.958333

We used the Seaborn library to view details of population and income data by borough generating the following graphics.



In this graph you can see the population differences from one borough to another. The most populated borough are Carabanchel, Fuencarral- El Pardo, Latina and Puente de Vallecas.



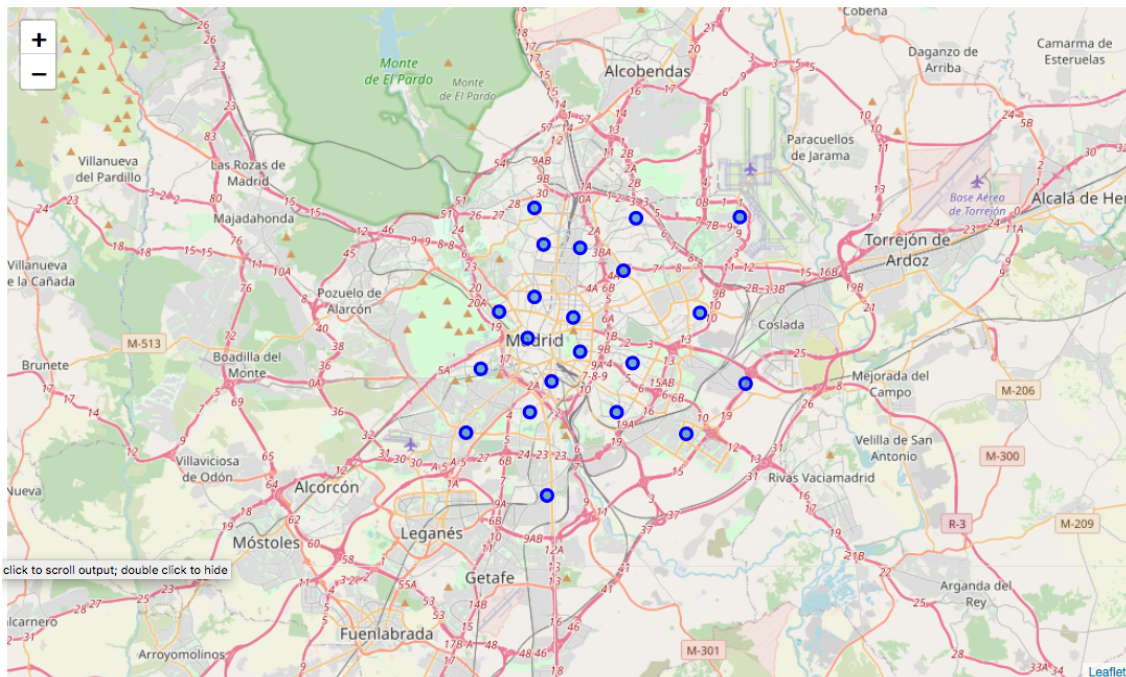
In this graph you can see the differences in income per borough. The borough with the highest income are Chamartin and Salamanca.

The next step was to capture the geographical coordinates of each of the borough from Geocoder. The data obtained has been incorporated into the main dataset.

	Borough	Neighborhood	Population	RentalNE	AvgHousingPrice	Latitude	Longitude
0	Centro	Palacio, Embajadores, Co...	140.473	16147	5479.500000	40.417653	-3.707914
1	Arganzuela	Imperial, Las Acacias, L...	155.660	17306	4547.750000	40.398068	-3.693734
2	Retiro	Pacífico, Adelfas, Estre...	120.406	21504	5178.416667	40.411150	-3.676057
3	Salamanca	Recoletos, Goya, Fuente ...	147.854	24433	6352.916667	40.427045	-3.680602
4	Chamartín	El Viso, Prosperidad, Ci...	147.551	25969	5944.666667	40.458987	-3.676129

## FOLIUM

We used python **folium** library to visualize geographic details of Madrid and we created a map of Madrid with boroughs superimposed on top. We used latitude and longitude values to get the visual as below.



## FOURSQUARE API

Foursquare API allowed us to explore the boroughs and segment them. We designed the limit as 200 venue and the radius 2000 meter for each borough from their given latitude and longitude informations. Here is a head of the list Venues name, category, latitude and longitude informations from Forsquare API.

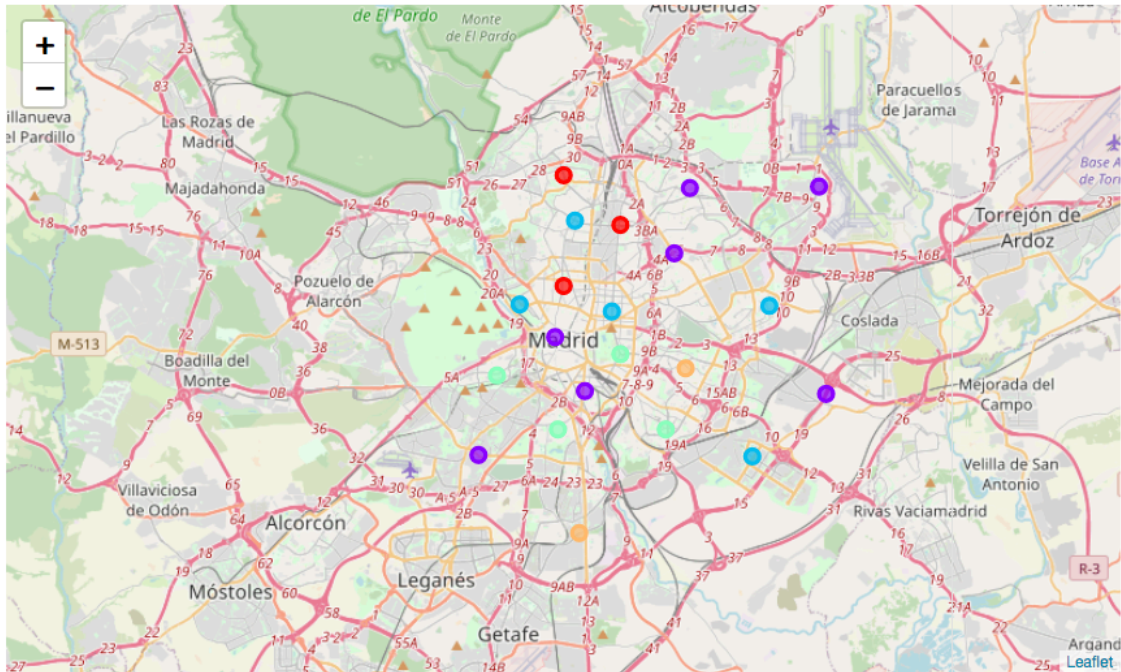
	Borough	Borough Latitude	Borough Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Centro	40.417653	-3.707914	Chocolatería San Ginés	40.416754	-3.707079	Chocolate Shop
1	Centro	40.417653	-3.707914	Churrería Los Artesanos 1902	40.417581	-3.707029	Coffee Shop
2	Centro	40.417653	-3.707914	Five Guys	40.420526	-3.705540	Burger Joint
3	Centro	40.417653	-3.707914	Starbucks	40.424975	-3.684107	Coffee Shop
4	Centro	40.417653	-3.707914	Mallorca	40.421205	-3.688435	Bakery

In summary of this data 998 venues and 89 unique categories were returned by Foursquare.

```
Venue Category
American Restaurant      8
Argentinian Restaurant   10
Asian Restaurant         13
BBQ Joint                4
Bagel Shop               1
..
Udon Restaurant          1
Vegetarian / Vegan Restaurant 6
Venezuelan Restaurant    1
Vietnamese Restaurant    3
Wine Bar                 2
Name: Borough, Length: 89, dtype: int64
```

## CLUSTERING

After a proces of One hot encoding, grouping by borough, adding datas and transforming the dataset by scaling each feature, we run K-Means algorith to cluster the boroughs in 5 groups. The variables that took into account for de algorithm were: Population, AvgHousePrice and all venues characteristics.



### 4. ANALYSIS BY CLUSTERS

#### CLUSTER 1

	Borough	Neighborhood	Population	RentalNE	AvgHousingPrice	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue
4	Chamartín	El Viso, Prosperidad, Gí...	147.551	25969	5944.666667	40.458987	-3.676129	0	Restaurant	Spanish Restaurant
6	Chamberí	Gaztambide, Arapiles, Tr...	140.866	22499	6040.750000	40.436247	-3.703830	0	Coffee Shop	Bakery
7	Fuencarral-El Pardo	El Pardo, Fuentelarreina, ...	249.973	18573	4322.333333	40.477133	-3.703582	0	Spanish Restaurant	Coffee Shop

- This cluster contains 3 boroughs.
- Boroughs to the north and centre of the city
- With high average house price
- And high income

## CLUSTER 2

	Borough	Neighborhood	Population	RentalNE	AvgHousingPrice	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue
1	Arganzuela	Imperial, Las Acacias, L...	155.660	17306	4547.750000	40.398068	-3.693734	1	Coffee Shop	Tapas Restaurant
20	Barajas	Alameda de Osuna, Aeropuerto, ...	50.010	17641	3770.916667	40.473318	-3.579845	1	Spanish Restaurant	Coffee Shop
10	Carabanchel	Comillas, Opañel, San Is...	260.196	10872	2537.500000	40.374211	-3.744676	1	Fast Food Restaurant	Pizza Place
0	Centro	Palacio, Embajadores, Co...	140.473	16147	5479.500000	40.417653	-3.707914	1	Coffee Shop	Bakery
14	Ciudad Lineal	Ventas, Pueblo Nuevo, Qu...	219.867	15408	3539.333333	40.448431	-3.650495	1	Spanish Restaurant	Restaurant
15	Hortaleza	Palomas, Piovera, Canill...	193.264	18277	4218.958333	40.472549	-3.642552	1	Bakery	Restaurant
18	Vicálvaro	Casco Histórico de Vicálvaro, A...	74.048	11695	2682.333333	40.396584	-3.576622	1	Café	Spanish Restaurant

- This cluster contains 7 boroughs
- Boroughs located in the north-western and south-western part of the city forming a diagonal except for Vicalvaro
- With medium population except for Barajas
- And medium average house price and medium income

## CLUSTER 3

	Borough	Neighborhood	Population	RentalNE	AvgHousingPrice	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue
8	Moncloa-Aravaca	Casa de Campo, Argüelles, ...	121.683	22152	4676.958333	40.429696	-3.725188	2	Coffee Shop	Restaurant
3	Salamanca	Recoletos, Goya, Fuente ...	147.854	24433	6352.916667	40.427045	-3.680602	2	Coffee Shop	Restaurant
19	San Blas-Canillejas	Simancas, Hellín, Ampost...	161.222	13404	2969.708333	40.428919	-3.604002	2	Spanish Restaurant	Tapas Restaurant
5	Tetuan	Bellas Vistas, Cuatro Caminos, ...	161.313	14970	4333.708333	40.460578	-3.698281	2	Spanish Restaurant	Restaurant
17	Villa de Vallecas	Casco Histórico de Vallecas, Sa...	114.512	11925	2702.416667	40.373958	-3.612163	2	Coffee Shop	Restaurant

- This cluster contains 5 boroughs
- Boroughs located in the center and east-southeast of the city
- With medium population
- Medium average house price and medium income except for Moncloa y Salamanca
- Moncloa y Salamanca has a high average house price and high income



## CLUSTER 4

	Borough	Neighborhood	Population	RentalNE	AvgHousingPrice	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue
9	Latina	Los Cármenes, Puerta del Angel,...	242.139	12232	2711.583333	40.403532	-3.736152	3	Bakery	Restaurant
12	Puente de Vallecas	Entrevías, San Diego, Pa...	240.867	9545	2249.000000	40.383553	-3.654535	3	Café	Tapas Restaurant
2	Retiro	Pacífico, Adelfas, Estre...	120.406	21504	5178.416667	40.411150	-3.676057	3	Restaurant	Coffee Shop
11	Usera	Orcasitas, Orcasur, San ...	142.894	9395	2324.625000	40.383894	-3.706446	3	Restaurant	Tapas Restaurant

- This cluster contains 4 boroughs
- Boroughs located in the center and south of the city
- With low population except Latina
- Low average house price and low income except for Retiro

## CLUSTER 5

	Borough	Neighborhood	Population	RentalNE	AvgHousingPrice	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue
13	Moratalaz	Pavones, Horcajo, Marroq...	95.614	13944	2900.000000	40.405933	-3.644874	4	Café	Coffee Shop
16	Villaverde	San Andrés, San Cristobal, ...	154.318	9756	2049.583333	40.345610	-3.695956	4	Café	Restaurant

- This cluster contains only 2 boroughs
- Boroughs located in the south south-east of the city
- Medium to low population
- With low average house price
- Medium to low income
- And Café as the most common venue

## 5. CONCLUSION

This project offers a superficial overview for an investor who wants to open a restaurant/food venue in Madrid. It gives some guidelines to decide the type of premises and the most suitable neighbourhood to open it.

In order to carry out the project we have taken into account socio-economic variables of the different borough of Madrid, per capita income, the average house price and some specific characteristics of the most common food premises in each borough.

The results obtained show that for an investor with money, the most suitable place to set up a restaurant is in group 1. Boroughs in the northern part of the city with a medium population level and a high income. Areas where premises have high purchase prices and therefore a high initial investment would be necessary.

These results are very superficial but could be the starting point to analyse new variables depending on the interests and possibilities of each investor.