

# **Coursera Capstone Project**

## **Vienna Coffee House**

### **in Madrid**

---

Sara Stark

# **Table of Content**

INTRODUCTION / BUSINESS UNDERSTANDING	3
DATA REQUIREMENTS	3
METHODOLOGY	3
RESULTS	4
DISCUSSION	4
CONCLUSION	5

# 1. INTRODUCTION / BUSINESS UNDERSTANDING

In Madrid, if someone were looking to open a typical viennese coffee house, the question is, where would you recommend to open it?

The background of the problem is that in order for the coffee house to be profitable, there must be enough customers, and in order to have enough customers, it is not worth setting up one in the immediate proximity of existing other ones.

Let's also make sure that the location where we open the coffee house will have the highest amount of customers possible. The New owner should care about this problem because the location of the new viennese coffee house has a significant impact on the expected returns.

## 2. DATA REQUIREMENTS

To obtain the data from the districts of Madrid, We used a CSV file with data extracted from Wikipedia with the information of each of the 21 districts in Madrid: Population, Latitude and Longitude.

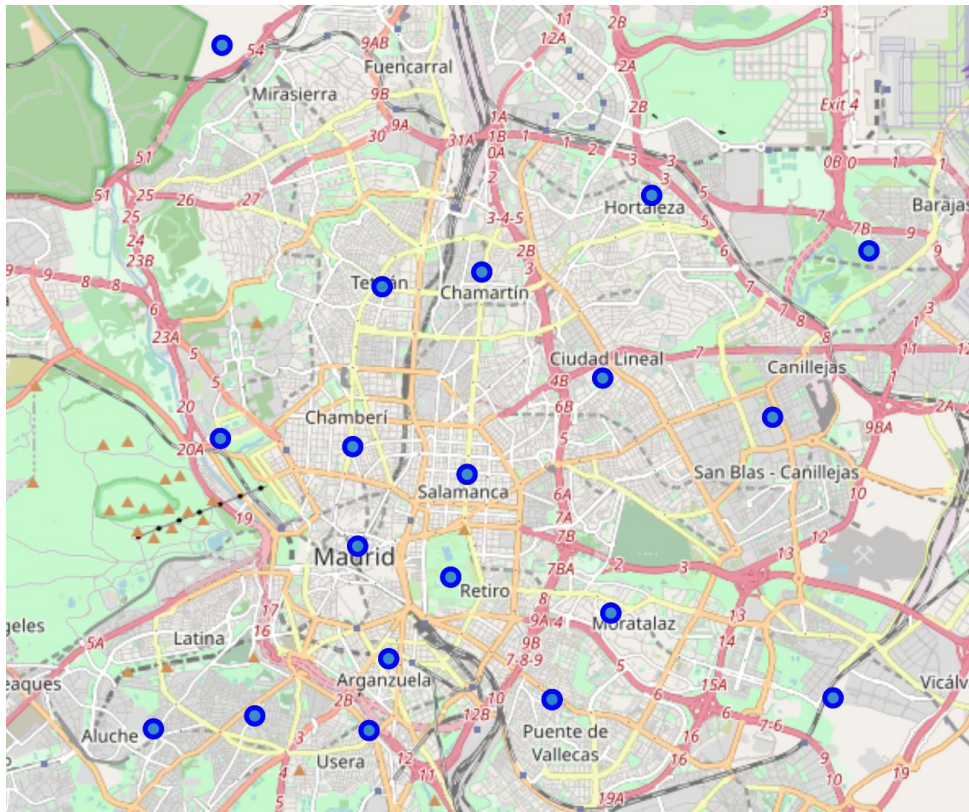
- Source: [https://en.wikipedia.org/wiki/Districts\\_of\\_Madrid](https://en.wikipedia.org/wiki/Districts_of_Madrid)

	District Number	District	latitude	longitude	population
0	1	Centro	40.41831	-3.70275	131928
1	2	Arganzuela	40.40021	-3.69618	151965
2	3	Retiro	40.41317	-3.68307	118516
3	4	Salamanca	40.42972	-3.67975	143800
4	5	Chamartin	40.46206	-3.67660	143424

Then we also calculate a **population score** based on each (District Population) / (Total population \* 100).

	District Number	District	latitude	longitude	population	Population Score
10	11	Carabanchel	40.39094	-3.72420	243998	7.667649
7	8	Fuencarral-El Pardo	40.49840	-3.73140	238756	7.502919
9	10	Latina	40.38897	-3.74569	233808	7.347427
12	13	Puente de Vallecas	40.39354	-3.66200	227595	7.152184
14	15	Ciudad Lineal	40.44505	-3.65132	212529	6.678734
15	16	Hortaleza	40.47444	-3.64110	180462	5.671027
19	20	San Blas-Canillejas	40.43893	-3.61537	154357	4.850676
5	6	Tetuan	40.45975	-3.69750	153789	4.832827
1	2	Arganzuela	40.40021	-3.69618	151965	4.775507
3	4	Salamanca	40.42972	-3.67975	143800	4.518922

That score will be used later as a variable to obtain the best location. We use folium to show the location of each district in the map of Madrid:



We use FourSquare data to obtain the 5 most common venues for each district of Madrid.

	name	categories	lat	lng
0	La Pequeña Graná	Tapas Restaurant	40.399574	-3.698550
1	Havana Blues	Cuban Restaurant	40.402050	-3.698488
2	Tres Cerditos	Chinese Restaurant	40.397316	-3.694184
3	Mercado de Santa María de la Cabeza	Farmers Market	40.402860	-3.696819
4	Salón de Té Al Yabal	Cocktail Bar	40.399015	-3.700249

### 3. METHODOLOGY

Once the most common venues for each district are retrieved we calculate two new variables: Coffee Score and Breakfast Spot Score based on if we find these types of restaurants in the top five most common venues for each district. We decided to include the Breakfast Spot Score because it is a similar type of establishment and can be also competition for our new coffee shop, even if they do not specialize in viennese coffees.

We join the results with our Population score. Then we give the following weight to each variable in order to obtain a total score:

- population\_weight = 0.5
- coffee\_score\_weight = 0.3
- breakfast\_spot\_weight = 0.2

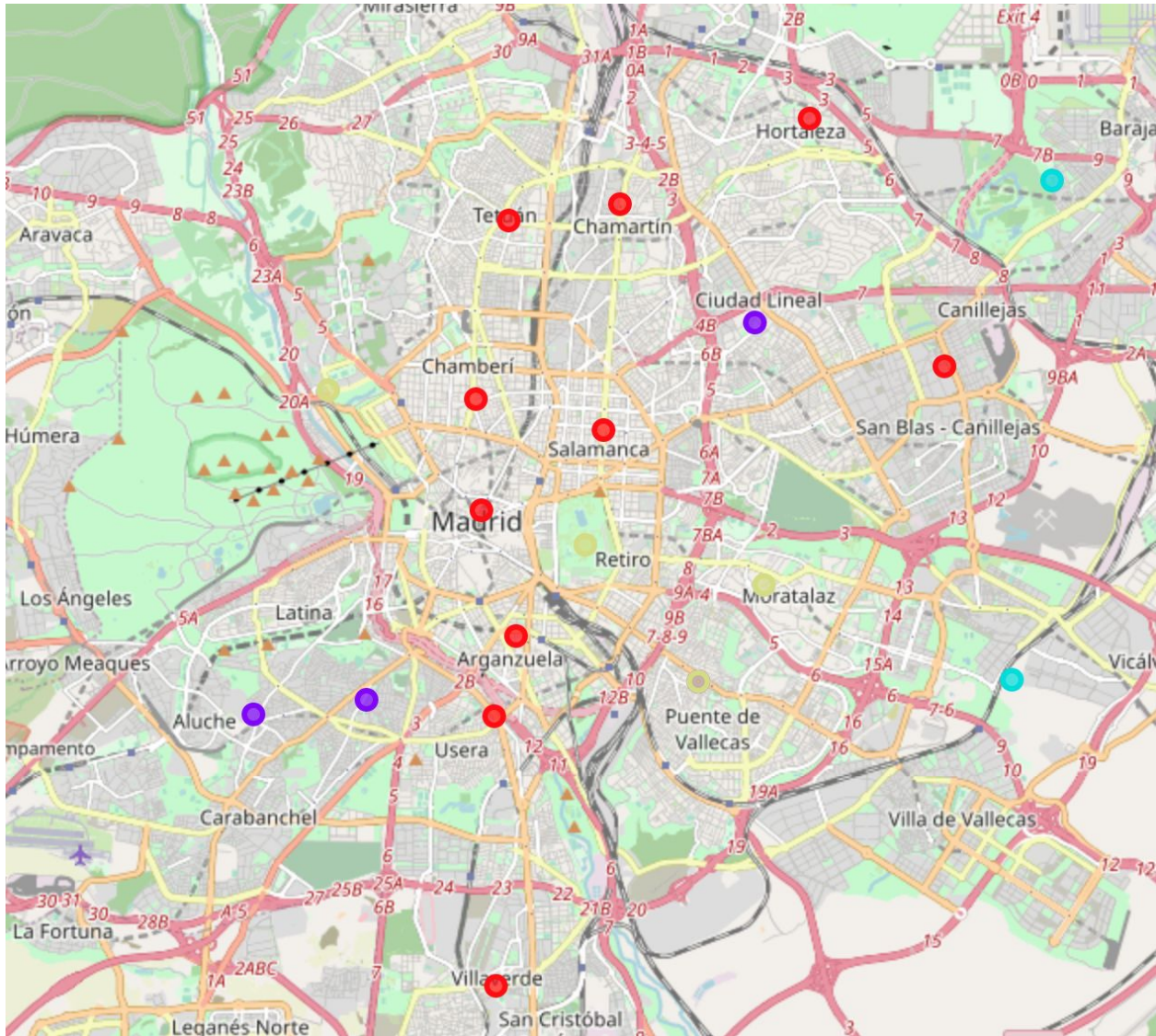
We also use the machine learning technique K-means to segment and cluster these districts so that we can group them together to understand their similarities. This is very important in order to recommend to possible best districts to open the coffee shop and to expand the business in the future.

Also we will use the heat map plugin from Folium, to visualize the existing coffee shops in the best district, this will help us to select one good location for our new coffee shop.



## 4. RESULTS

With the K-Means clustering technique, the top 4 clusters of similar districts appear in the results, each color is a cluster. See below.

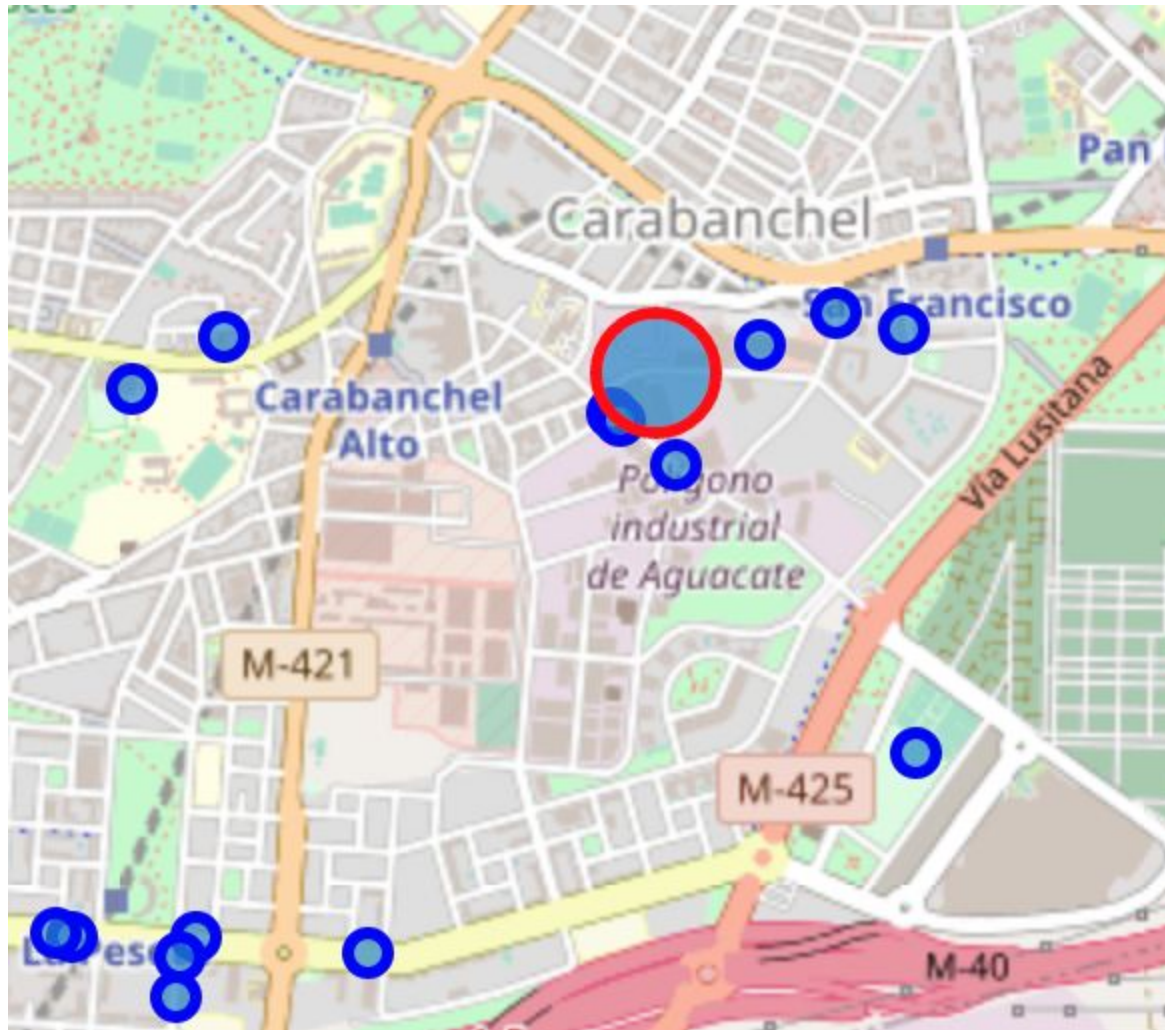


We obtain in the cluster label 2 to be the best districts to locate our viennese coffee shop:

Carabanchel is the best option because it has the perfect mix of all indicators. Other good options can be: Latina, Puente de Vallecas, Ciudad Lineal and Hortaleza.

	District Number	District	latitude	longitude	population	Population Score	Coffee Score	Breakfast Spot Score	Total Score	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	11	Carabanchel	40.39094	-3.72420	243998	7.667649	1.0	0.0	4.133824	1	Coffee Shop	Grocery Store	Fast Food Restaurant	Farmers Market	Supermarket
1	10	Latina	40.38897	-3.74569	233808	7.347427	0.2	0.6	3.853714	1	Grocery Store	Fast Food Restaurant	Breakfast Spot	Comedy Club	Coffee Shop
2	13	Puente de Vallecas	40.39354	-3.66200	227595	7.152184	0.0	0.0	3.576092	1	Gym	Pizza Place	Pub	Chinese Restaurant	Italian Restaurant
3	15	Ciudad Lineal	40.44505	-3.65132	212529	6.678734	0.0	0.0	3.339367	1	Spanish Restaurant	Grocery Store	Gym / Fitness Center	Hotel	Pizza Place
4	16	Hortaleza	40.47444	-3.64110	180462	5.671027	0.0	0.0	2.835513	0	Spanish Restaurant	Food	Tapas Restaurant	Metro Station	Bar

Based on the results we can select a possible good location:





## 5. DISCUSSION

Before opening a new Viennese Coffee Shop, there are some other factors that may be good to take into consideration in order to have a more optimal location.

The usage of FourSquare has increased in Madrid in recent years but Google maps is still the most used one in order to check all the existing restaurants in the city, so it will be good to take that into consideration for a deeper analysis.

For more accurate results it would be good to consider other variables, like age of the population, number of tourists and average income.

To be more specific, this analysis can be done at neighborhood level instead of district level.

## 6. CONCLUSION

In this study we analyzed which would be the best location to open a Viennese Coffee Shop in Madrid, for that we used the location and the population of the 21 districts of Madrid. Also using FourSquare data we got the most common venues for each district. With that data we obtain the districts with the most amount of coffee shops and breakfast spots that will be our competition.

After applying the machine learning K-Means clustering technique, we can conclude that Carabanchel seems to be the better district to establish our new Viennese Coffee Shop.

With the support of a heat map a possible good location in the selected district was found.

To finish we can say that this is a simple study that can be improved by taking more variables into consideration, but it gives us a quick view of the potential benefits of applying the Data Science methodology to make a decision such as the location of a new restaurant.