Coursera Capstone Project
Burger Joint in Madrid

Jesús Merino
September 16, 2020

INDEX

INDEX

1. INTRODUCTION / BUSINESS UNDERSTANDING

Burgers although reviled for being considered an icon of fast food, has earned a place in our gastronomy and they are more fashionable than ever.

In Madrid, if someone is looking to open a burger joint, the question is, where would you recommend to open it? The background of the problem is that in order for a burger joint 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 ones.

Let's also make sure that the location to open the local should be a place with the higher possible customers. The new owner should care about this problem because the location of the new burger joint 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

	Number	District	latitude	longitude	population
0	1	Centro	40.415347	-3.707371	131928
1	2	Arganzuela	40.402733	-3.695403	151965
2	3	Retiro	40.413170	-3.683070	118516
3	4	Salamanca	40.430000	-3.677778	143800
4	5	Chamartin	40.453333	-3.677500	143424

Then we also calculate a population score based on each district population / Total population * 100.

	Number	District	latitude	longitude	population	Population Score
10	11	Carabanchel	40.383669	-3.727989	243998	7.667649
7	8	Fuencarral-El Pardo	40.478611	-3.709722	238756	7.502919
9	10	Latina	40.402461	-3.741294	233808	7.347427
12	13	Puente de Vallecas	40.398204	-3.669059	227595	7.152184

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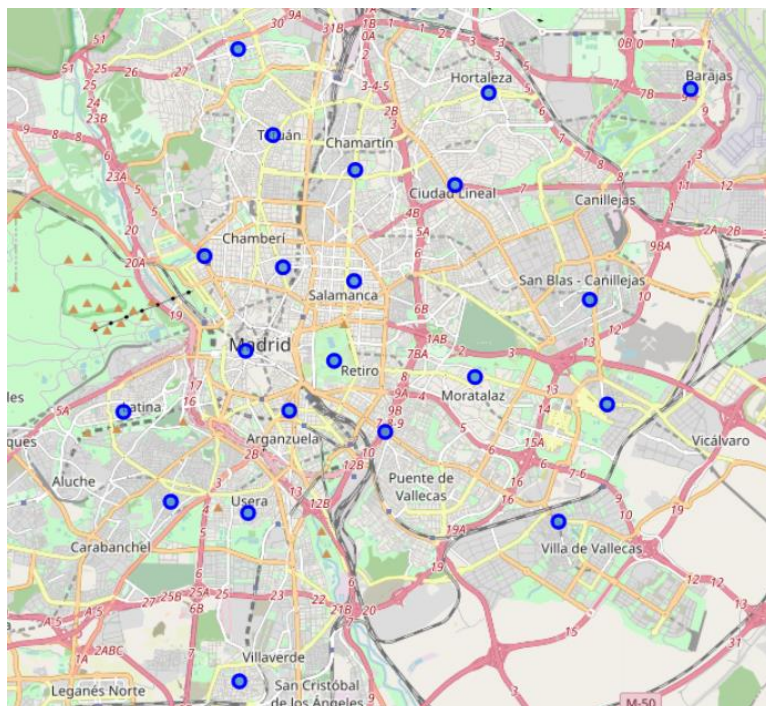
We use FourSquare data to obtain the 5 most common venues for each district of Madrid.

	District	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Arganzuela	Restaurant	Spanish Restaurant	Grocery Store	Bakery	Tapas Restaurant
1	Barajas	Hotel	Restaurant	Spanish Restaurant	Coffee Shop	Tapas Restaurant
2	Carabanchel	Soccer Field	Plaza	Burger Joint	Fast Food Restaurant	Tapas Restaurant
3	Centro	Plaza	Tapas Restaurant	Spanish Restaurant	Hostel	Bistro
4	Chamartin	Restaurant	Spanish Restaurant	Grocery Store	Bakery	Tapas Restaurant
5	Chamberi	Spanish Restaurant	Bar	Restaurant	Brewery	Japanese Restaurant

3. METHODOLOGY

Once the most common venues for each district are retrieved we calculate two new variables: Burger score and Fast food score based on if we find these types of restaurants in the top five most common venues for each district. We decide to include the fast food score because is a similar type of restaurant and can be also competition for our new burger joint, even if they are not specialize in burgers. We join the results with our Population score.

	Number	District	latitude	longitude	population	Population Score	Burger Score	Fast Food Score
0	11	Carabanchel	40.383669	-3.727989	243998	7.667649	0.6	0.4
1	8	Fuencarral-El Pardo	40.478611	-3.709722	238756	7.502919	0.6	0.2
2	10	Latina	40.402461	-3.741294	233808	7.347427	0.0	0.4
3	13	Puente de Vallecas	40.398204	-3.669059	227595	7.152184	0.0	1.0
4	15	Ciudad Lineal	40.450000	-3.650000	212529	6.678734	0.6	0.0
5	16	Hortaleza	40.469457	-3.640482	180462	5.671027	0.0	0.0
6	20	San Blas-Canillejas	40.426001	-3.612764	154357	4.850676	0.0	0.0
7	6	Tetuan	40.460556	-3.700000	153789	4.832827	0.0	0.0
8	2	Arganzuela	40.402733	-3.695403	151965	4.775507	0.0	0.0
9	4	Salamanca	40.430000	-3.677778	143800	4.518922	0.2	0.0

Then we give the following weight to each variable in order to obtain a total score:
- population weight = 0.5
- burger weight = 0.3
- fast food 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 burger joint and to expand the business in the future.

latitude	longitude	population	Population Score	Burger Score	Fast Food Score	Total Score
40.383669	-3.727989	243998	7.667649	0.6	0.4	4.093824
40.478611	-3.709722	238756	7.502919	0.6	0.2	3.971459
40.398204	-3.669059	227595	7.152184	0.0	1.0	3.776092
40.402461	-3.741294	233808	7.347427	0.0	0.4	3.753714
40.450000	-3.650000	212529	6.678734	0.6	0.0	3.519367
40.469457	-3.640482	180462	5.671027	0.0	0.0	2.835513
40.426001	-3.612764	154357	4.850676	0.0	0.0	2.425338
40.460556	-3.700000	153789	4.832827	0.0	0.0	2.416413
40.402733	-3.695403	151965	4.775507	0.0	0.0	2.387754
40.430000	-3.677778	143800	4.518922	0.2	0.0	2.319461

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

4. RESULTS

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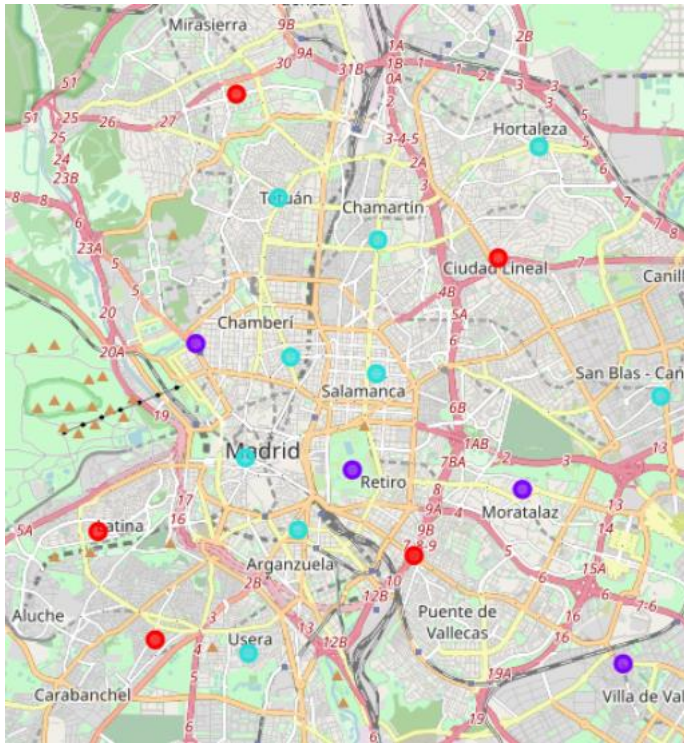

We obtain in the cluster label 2 the best districts to locate our burger joint:

	District	latitude	Population Score	Burger Score	Fast Food 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
5	Hortaleza	40.469457	5.671027	0.0	0.0	2.835513	2	Breakfast Spot	Supermarket	Pizza Place	Spanish Restaurant	Donut Shop
6	San Blas-Canillejas	40.426001	4.850676	0.0	0.0	2.425338	2	Metro Station	Supermarket	Snack Place	Gym	Grocery Shop
7	Tetuan	40.460556	4.832827	0.0	0.0	2.416413	2	Spanish Restaurant	Supermarket	Bakery	Grocery Store	Brazilian Restaurant
8	Arganzuela	40.402733	4.775507	0.0	0.0	2.387754	2	Restaurant	Spanish Restaurant	Grocery Store	Bakery	Tapas Restaurant
9	Salamanca	40.430000	4.518922	0.2	0.0	2.319461	2	Spanish Restaurant	Restaurant	Coffee Shop	Bakery	Burger Joint

Hortaleza is the best option because it has the higher Population score and less burger joints / fast food restaurant competition. Other good options can be: Canillejas, Tetuan, Arganzuela and Salamanca.

Using heat map in folium we can see where the fast food restaurants are located in the district:
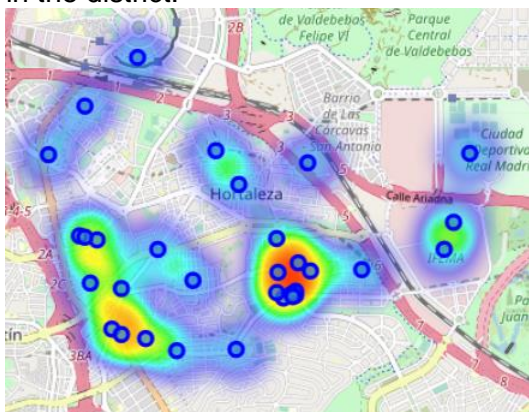
Based on the results we can select a possible good location:
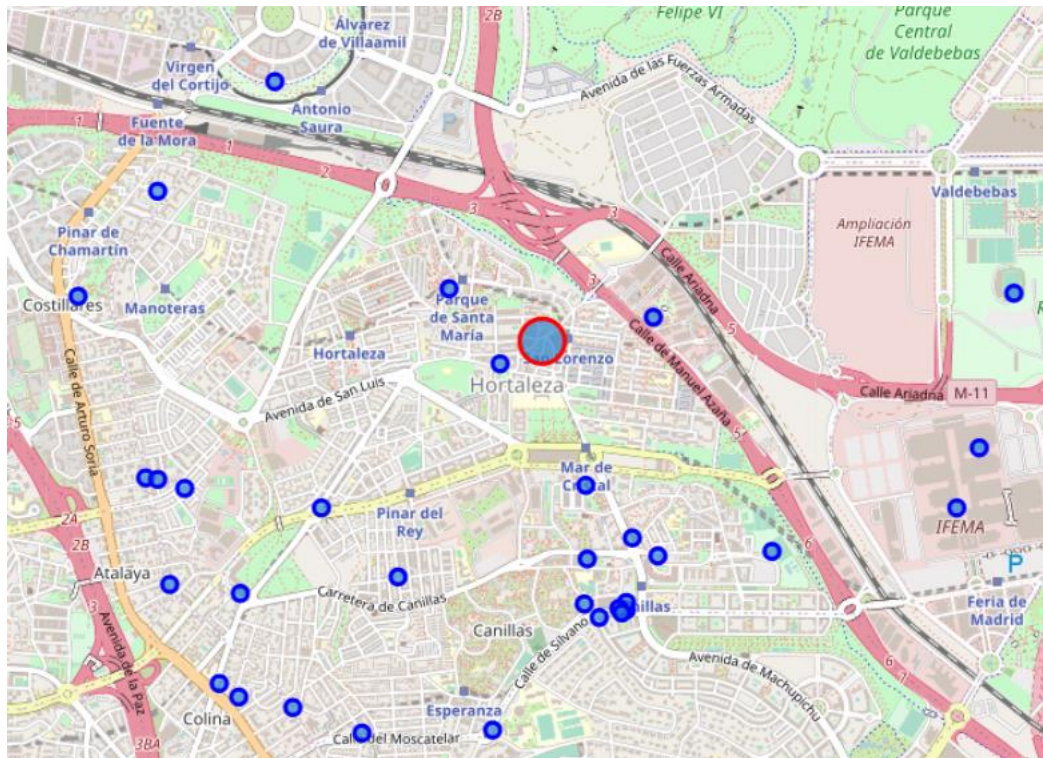
5. DISCUSSION

Before starting a new Burger Joint, 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 new burger joint 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 burger joints and fast food restaurants that will be our competition.

After applying the machine learning K-Means clustering technique, we can conclude that Hortaleza seems to be the better district to stablish our new burger joint.

With the support of a heat map a possible good location in the selected district would be in crossroad of Gregorio Sanchez Herraez Street and Mar de Japon Street.

To finish we can said that this is a simple study that can be improve taking more variables into consideration, but it give 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.