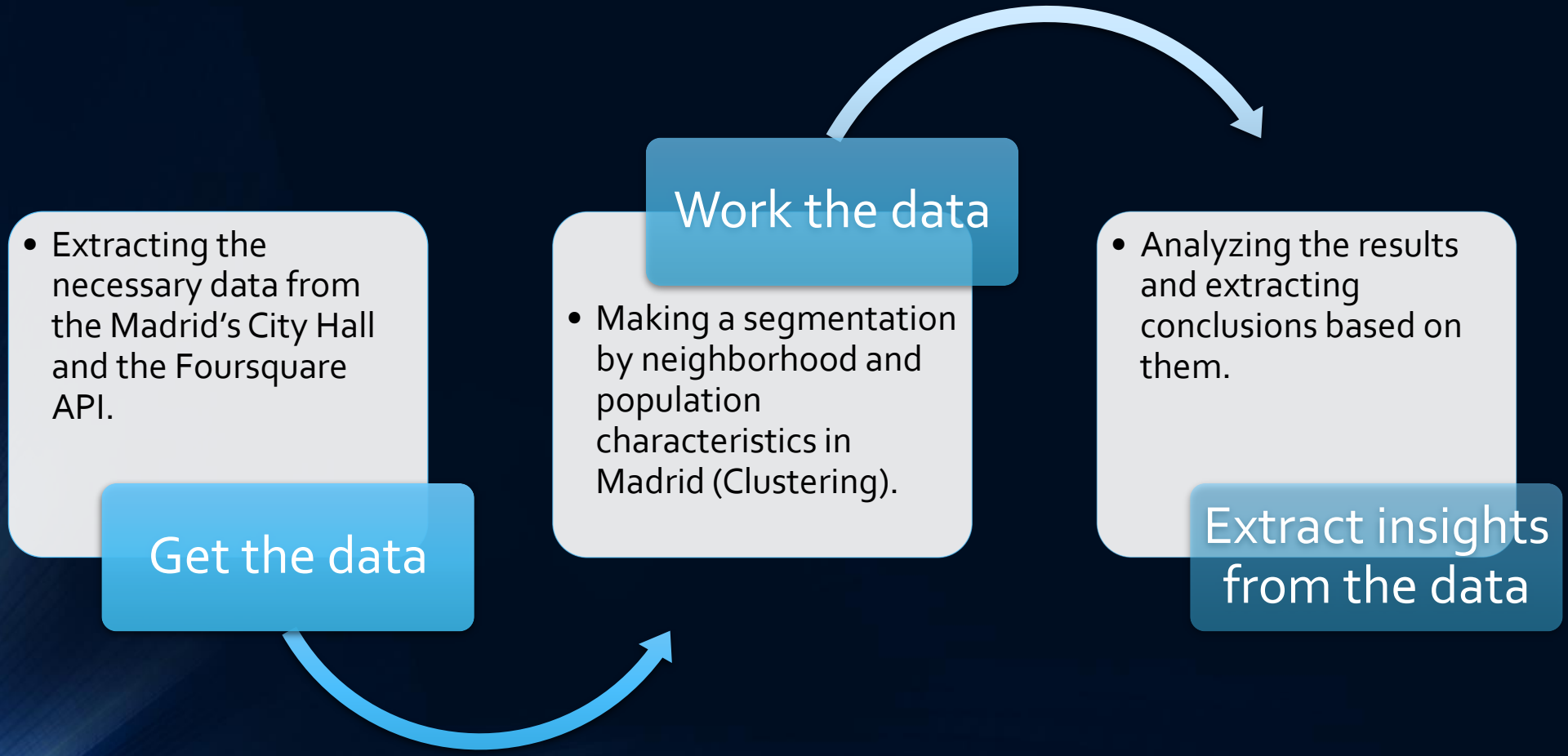
SOLVING A BUSINESS OPTIMAL LOCATION PROBLEM

MIGUEL ÁNGEL BARBERÁN

In this project we will be determing the optimal location of a business in a city area by:



The data that was used contained information about the nationalities of the inhabitants of each neighborhood and the quantity of people by neighborhood:

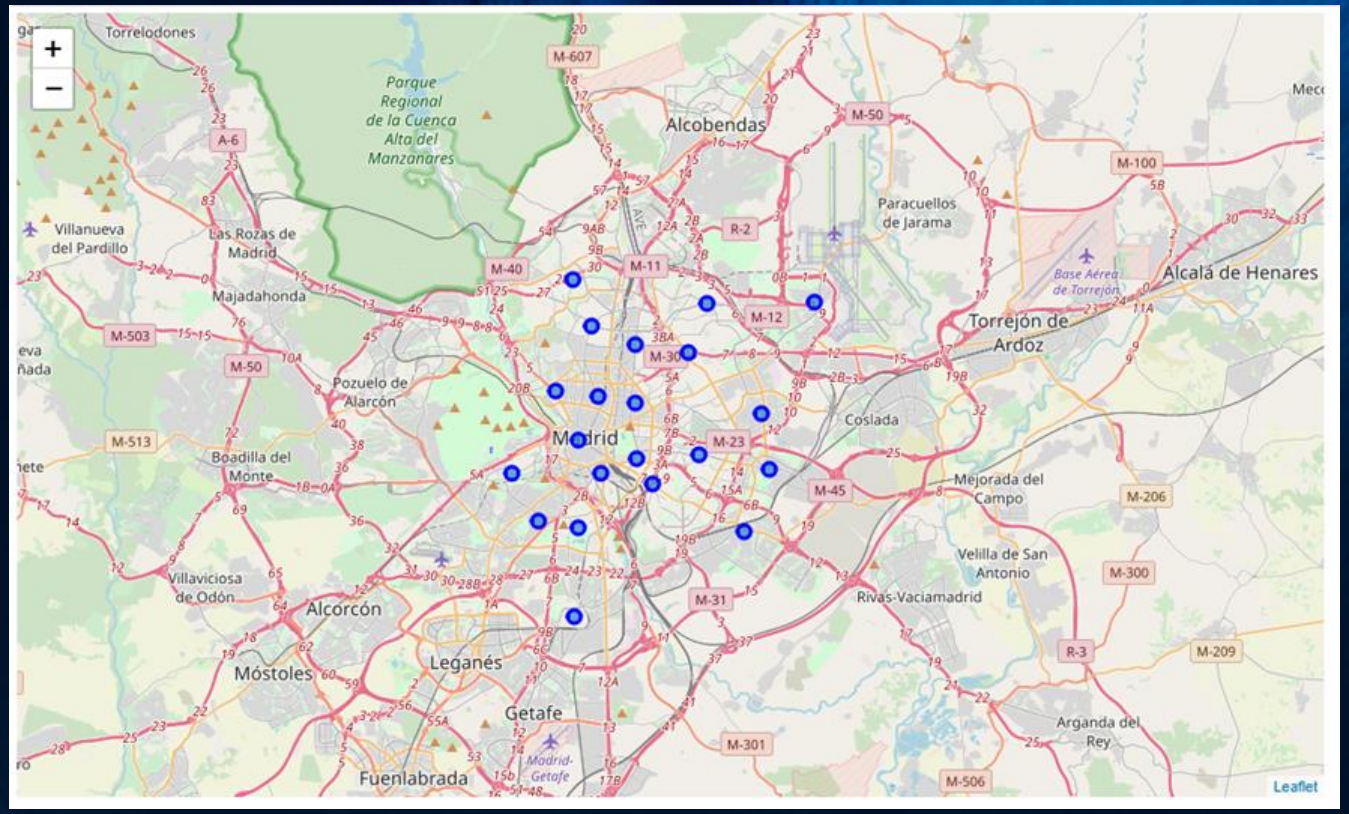
Country of P	Total	Ciudad Centro	Arganzuela	Retiro	Salamanca	Chamartin	Tetuán
Rumanía	450360	8150	7540	4800	7530	6800	14680
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Ecuador	239530	6470	7410	2650	6190	3800	13950
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Colombia	226180	9980	7170	4830	8030	5510	8220
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Italia	203080	30300	12190	8400	18170	10600	11940
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Paraguay	186820	3640	4740	2370	5210	6570	33110
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Honduras	159810	1490	2280	2320	3320	3370	7550

In order to access the Foursquare API data, the raw data had to be transformed this into something useful for the API:

- A second dataframe was created
- The neighborhoods's names were included
- The latitude and longitude values of each neighborhood were added

Neighbor	Latitude	Longitude
Centro	40415347	-3707371
[illegible]	[illegible]	[illegible]
Retiro	40408072	-3676729
[illegible]	[illegible]	[illegible]
Chamartin	40453333	-36775
[illegible]	[illegible]	[illegible]
Chamberí	40432792	-3697186
[illegible]	[illegible]	[illegible]
Moncloa-Ara	40435151	-3718765
[illegible]	[illegible]	[illegible]
Carabanchel	40383669	-3727989
[illegible]	[illegible]	[illegible]
Puente de V	40398204	-3669059
[illegible]	[illegible]	[illegible]
Ciudad Linea	4045	-365

Once the data was obtained, it was possible to draw the neighborhoods on a map:



Then, the nearby venues by neighborhood were extracted along with their frecuencies of occurrence:

VENUES BY NEIGHBORHOOD

Neighborhoo	Neighborhoo	Neighborhoo	Venue	Venue Latitu	Venue Longi	Venue Categ
Centro	40415347	-3707371	Plaza Mayor	4,0415E+16	-3,7076E+16	Plaza
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Centro	40415347	-3707371	La Taberna d	4,0415E+16	-3,7081E+15	Other Nightl
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Centro	40415347	-3707371	Amorino	4,0416E+15	-3,7084E+16	Ice Cream Sh
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Centro	40415347	-3707371	Bar El Cogoll	4,0414E+15	-3,7067E+15	Spanish Rest
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Centro	40415347	-3707371	Pinkleton &	4,0415E+15	-3,7091E+16	Wine Bar

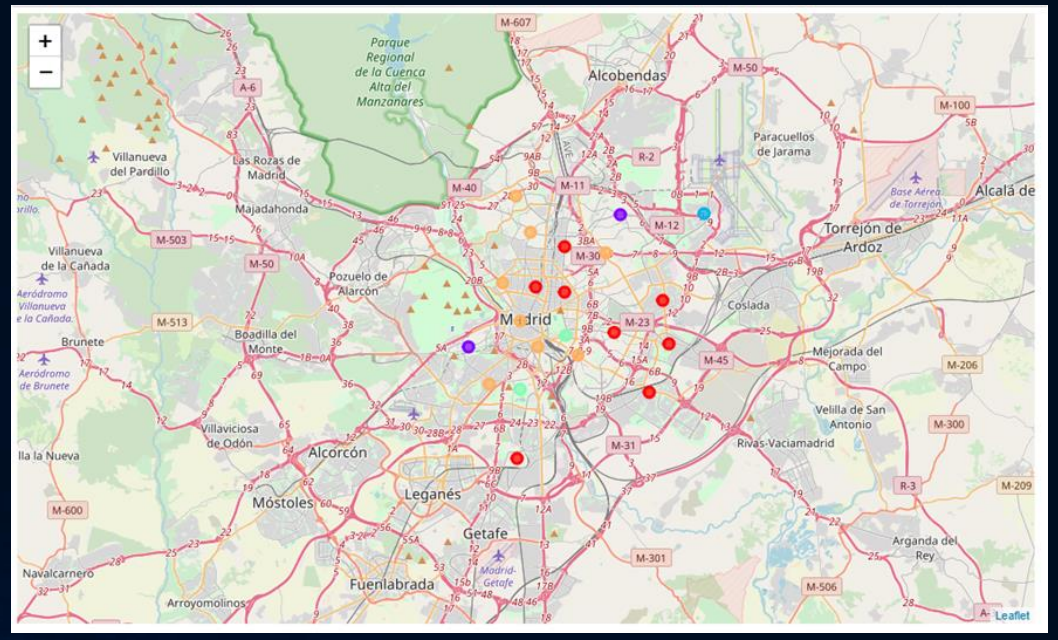
VENUES'S FRECUENCIES OF OCCURRANCE:

- ----Arganzuela----
- venue freq
- 0 Restaurant 0.10
- 1 Spanish Restaurant 0.09
- 2 Tapas Restaurant 0.05
- 3 Bakery 0.05
- 4 Grocery Store 0.05

Finally, a dataframe containing the most common venues by neighborhood was created:

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
Arganzuela	Restaurant	Spanish Restaurant	Bakery
Barajas	Hotel	Spanish Restaurant	Restaurant
Carabanchel	Burger Joint	Fast Food Restaurant	Pizza Place
Centro	Spanish Restaurant	Tapas Restaurant	Plaza
Chamartin	Spanish Restaurant	Restaurant	Pizza Place

After obtaining these data, clusters could be made:



This is a sample of what our final clusters looked like:

Barajas	Neighbor	Cluster Labels	1st Most Common Venue	2nd Most Common Venue
3140	Centro	0	Spanish Restaurant	Tapas Restaurant
740	Villa de Vall	0	Food	Spanish Restaurant
1910	Retiro	0	Spanish Restaurant	Supermarket
3370	Ciudad Linea	0	Spanish Restaurant	Burger Joint
570	VicÃ¡lvaro	0	Spanish Restaurant	Breakfast Spot
2580	Chamartin	0	Spanish Restaurant	Restaurant
920	Usera	0	Seafood Restaurant	Bubble Tea Shop
810	TetuÃ¡n	0	Spanish Restaurant	Brazilian Restaurant

Conclusions

- As far as we can see with this data, there are no Mexican populations registered in Madrid. However, in Cluster 1, it is possible to notice that there's a Mexican restaurant located in the "Centro" neighborhood, which is the town center.

- If a deeper exam is performed into this cluster, it is noticeable that its living population are mostly Latinos, mixed with some other Europeans, but mainly, the people living in this cluster come from south American countries. Apart of this fact, other kinds of Latin restaurants can be found, like Argentinian restaurants, tapas restaurants, and Italian restaurants. So, it is possible to tell that the inhabitants of this area like these kinds of food.

- By following this logic, if we would like to open a new Mexican restaurant in the city or any kind of restaurant in fact, it would only be necessary to find a where are the restaurants similar the one we want to open, study the population in that area, and find similar clusters of population in the city that don't have yet or have very few restaurants like the one we would like to open.

- In this example, clusters 4 and 5 could make a good match for our target population. Looking at the venues in these clusters, it is possible to find one Mexican restaurant, and a good bunch of fast food, Argentinian, and south American restaurants. So, in these clusters, it is possible to state that the existing restaurants matches the population's nationalities and tastes.

- In conclusion and taking into consideration the explanations given above as well as the data, it is highly possible that clusters 4 and five could be a good place to open our Mexican restaurants. As explained above, the same logic could apply to open other kind of restaurant or business in any other area of the city. It is only necessary to examine the existing businesses in our target area, and study the population, then compare these two factors with the same ones in areas where there are existing businesses like the one we want to open, and then verify if the matching is correct.

Miguel Ángel Barberán