



THE BATTLE OF NEIGHBORHOODS

Coursera Capstone Project
Miguel Navarro Baeza



Where should I go today in Madrid?



Problem Description

- Recently, a client has reached out saying that he/she does not know where to go when he/she wants to do an special activity in the city of Madrid. He/she would like to know which is the most appropriated neighborhood for developing the activity that wants to do that day. For example, lets say that our client wants to go to a park and he/she does not know which is the neighborhood with the biggest park offer of the city. With the solution exposed in this project, he/she would be able to determine where to go.
- Set a methodology for fixing a target and presenting the correct results. We are going to deal with different Foursquare data such as ratings, coordinates, type of venue... The solution must be expressed by giving a type of venue and some coordinates that could be represented in a map. With this analysis, we can give several options to our client to help him/her make the correct decision.



Data Description

- In this project we are going to use the Foursquare API for different purposes. We are going to construct a URL to send a request to the API to search for a specific type of venues, to explore a geographical location, and to get trending venues around a location. Also, we use Folium for creating maps and representing locations.
- Making calls to the Foursquare API will lead us to create json files that can be transformed into Dataframes. These DFs are easily refined and filtered using pandas tools. Data of special interest could range from Ratings, Longitude, Latitude, Type of Food...
- The steps carried out in this process should be:
 - *Data Acquisition*
 - *Data Wrangling and Filtering*
 - *Creating a data analysis model*
 - *Choosing a classification model*
 - *Deciding a solution*

Methodology

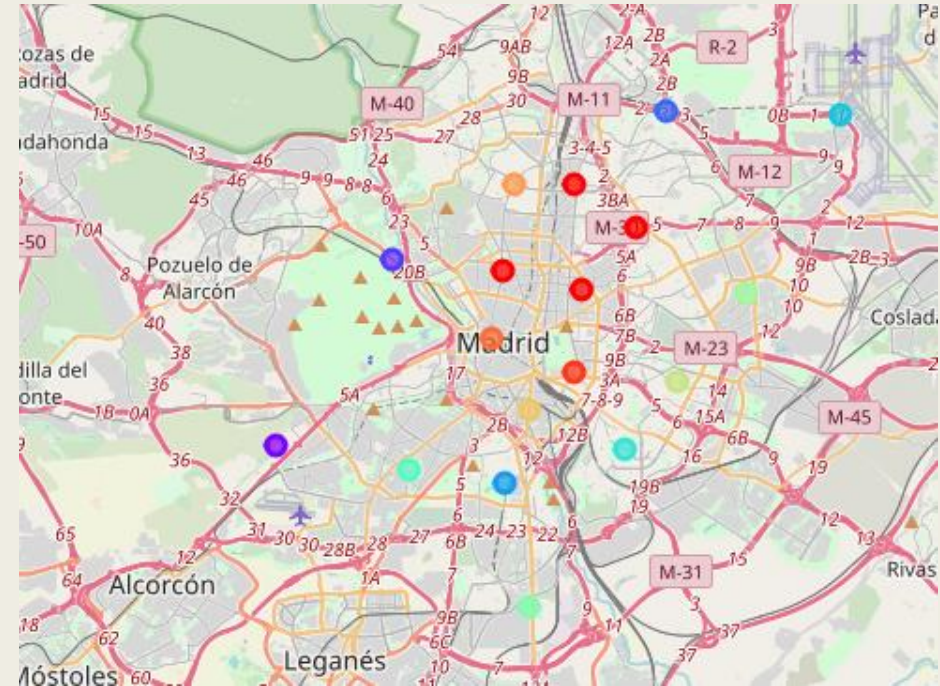
- The first step for data acquisition is going to be scrapping a Wikipedia page that includes the information related to the different neighborhoods in the city of Madrid.
- The Foursquare API permits to extract the information about the venues in each neighborhood determined in the previous web scrapping. After this, an extended data filter is operated. When the data is nicely distributed, a K means classification model is executed in order to obtain the different clusters that contain the different types of venues located in the city of Madrid.

Distrito ↕	Número ↕	Nombre ↕	Superficie (km²) ² ↕	Imagen ↕
	11	Palacio	1,471 km²	
	12	Embajadores	1,032 km²	
	13	Cortes	0,592 km²	
	14	Justicia	0,742 km²	
	15	Universidad	0,947 km²	
	16	Sol	0,445 km²	

FOURSQUARE DEVELOPERS

Results

- The results obtained are represented in a map that permits to visualize the location of each neighborhood and the type of cluster in this specific neighborhood.
- Also, a description of the most common venues in the cluster is stated in order to facilitate the client to find the neighborhood that he/she needs to go for the activity planned.



	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	Salamanca	Spanish Restaurant	Restaurant	Seafood Restaurant	Hotel	Tapas Restaurant	Bar	Japanese Restaurant	Plaza	Coffee Shop	Burger Joint
4	Chamartín	Restaurant	Spanish Restaurant	Tapas Restaurant	Plaza	Dessert Shop	Grocery Store	Gym	Donut Shop	Big Box Store	Indian Restaurant
6	Chamberí	Bar	Tapas Restaurant	Restaurant	Spanish Restaurant	Café	Plaza	Burger Joint	Theater	Coffee Shop	Mexican Restaurant
13	Ciudad Lineal	Spanish Restaurant	Restaurant	Burger Joint	Grocery Store	Molecular Gastronomy Restaurant	Bakery	Pharmacy	Hotel	Pizza Place	Diner

Discussion

- For example, if the client's desire is going to a park, he/she must search for the neighborhood in which the cluster contains a most common venue of a park. In this case, Moncloa is a nice neighborhood where the park offer is high.
- Other option is to select a neighborhood that is close to your current position and see the cluster that belongs to. This can let you see what kind of activity you can develop close to your position.



Conclusion

- In this project, the goal of providing enough information to a client for permitting doing an specific activity.
- I have learned how to gather data, filter it and try to obtain a solution by analyzing the different classification processes that can be done to this data.