# **Analysis of the Location of a Coffee Shop Branch**

## Problem

**Café Tango** is a popular coffee shop located in the neighborhood of Palermo, Buenos Aires, Argentina. Due to the extraordinary results gotten during the last year, its owners decided to open a new branch in other zone of the city but they don't have a clear idea of what places is the best option.

## Objectives

The objective of this project was to determine what neighborhood of Buenos Aires is the best option to locate the branch.

To get this we established the following sub-objectives:

* Determine what group neighborhoods is more similar to Palermo
* Determine what neighborhood of this similar group is the best option to install the new branch.

## Methodology

The process of determine what is the best place to locate the new branch consisted of two parts. The first part of the project consisted on determine what neighborhoods are the most similar to the neighborhood where coffee shop is located (Palermo).

In order to do this, **clustering** machine learning technique was used to group the neighborhood based on the **similarity of their most popular venues**. The clustering method to get this was **K-Means method**.

The second part of the project consisted on determine what of those similar neighborhoods represent the best opportunity to a new branch. The **customer satisfaction level** of the local cafeterias was used to get this answer.

The customer level of satisfaction indicates the relationship between what clients hope to receive and what they really receive. A low customer satisfaction implies the customers are not getting what they hope to get and it represents an opportunity to Cafe Tango.

**The specific indicator used to estimate the customer satisfaction it's the average rating of the most popular coffee venues of the neighborhood**. The neighborhood with the low average rating will be considerated the best option to located the new branch.

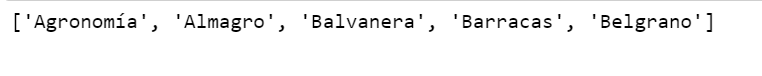
## Data requirements

In order to get this objective, the data used was:

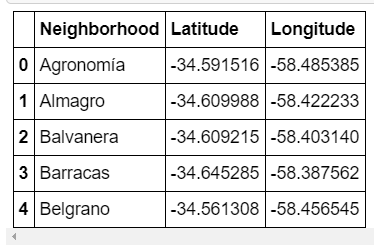
* Buenos Aires' neighborhoods and its location coordinates
* The most common venues for each neighborhood
* The rating of the most popular coffee shop of the neighborhood

## Data collection

The list of Buenos Aires' neighborhood was taken from Wikipedia through a scrapping process of following [page](https://es.wikipedia.org/wiki/Anexo:Barrios_de_la_ciudad_de_Buenos_Aires'). A portion of resulted list with the first 5 neighborhoods could be appreciated below



To get the geolocation of each neighborhood, the **Geopy** library was used. Using a loop, the coordinates were found using the name of each neighborhood and a geolocation agent. Then all the information was gathered to create a data frame as the shown below.



The information of the most popular venues of each neighborhood was collected from Foursquare. The API was used with the endpoint “explore”, what returns the most popular venues near the given geolocation.

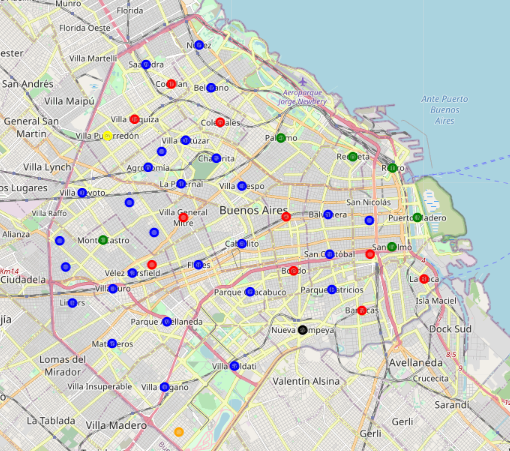
For the last part of the project, the rating of the most popular coffee shop venues was obtained using Foursquare too. In this case the venues id gotten in the previous step were used.

## Result

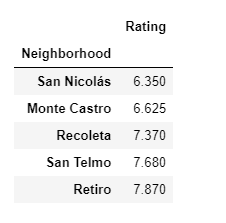
During the clustering process, six cluster were conformed with the following distribution

|  |  |
| --- | --- |
| **Cluster** | **N° of neighborhoods** |
| Cluster 1 | 7 |
| Cluster 2 | 10 |
| Cluster 3 | 23 |
| Cluster 4 | 2 |
| Cluster 5 | 3 |
| Cluster 6 | 3 |

The reference neighborhood, **Palermo**, belong to cluster 1 with **Retiro, San Nicolas, Puerto Madero, Recoleta, Monte Castro, San Telmo.** They location are denoted with green marks in the following map.



After evaluate the average rating of the most popular coffee shops of each of this group of six (6) neighborhoods, we could conclude that general performances of the neighborhoods’ coffee shops are not good. No neighborhood exceeds the 8 point in a scale of 10 for its coffee shops.The neighborhood withthe lowest rating average was **San Nicolas** with a rating of 6.3/10.



## Conclusion

The neighborhood where the new branch must be located is **San Nicolas**

## Discussions

An important fact is that the price of the locations in the neighborhoods wasn't taken into account. This factor may be included in posterior analysis

An interested data to dive in a posterior project is the fact that most of the neighborhood in the **cluster 1** are geographical neighbors and they are located near the sea. This may indicate that the geography has some influence over the type of venues, a thing that not happen on with the others clusters.