Applied Data Science Capstone Project: The Battle of Neighborhoods

CARLOS LLANO

JANUARY OF 2021

Table of contents

Introduction/Business Problem

Bogotá is the metropolitan center of Colombia, which is located at 2,600 meters above sea level. This city offers varied artistic expressions such as museum exhibits, dance, theater, music, and splendid cuisine (1). There are also hundreds of places to go to eat and frequent food festivals (2), and consequently, the catering industry could be a very competitive field in this city.

In this project, the data available on restaurants and territorial distribution of Bogotá will be used to analyze and cluster different locations in which a restaurant could be potentially be established. The results obtained in this study could be of interest to businessmen or investors that want to open a restaurant in Bogotá, Colombia.

available at: https://wwwnc.cdc.gov/travel/destinations/traveler/none/colombia

Data acquisition

The necessary data to perform the analysis of the best locations for a restaurant in Bogotá, Colombia will be obtained from the following sources:

- Lists of the boroughs and neighborhoods and their geospatial position will be obtained from www.bogota-laburbano.opendatasoft.com.
- Data about the territorial distribution of Bogotá will also be acquired from www.bogota-laburbano.opendatasoft.com. This web page contains information about tourist, urban and company areas in Bogotá.
- The types and location of each restaurant in all the neighborhoods will be obtained from Foursquare API (www.foursquare.com).

Methodology

1. Data search and acquisition

- All the data needed about the neighborhoods in Bogotá, Colombia was first searched and collected.

2. Boroughs and neighborhoods data exploration

- The data about the boroughs and neighborhoods were cleaned and then analyzed in order to identify the boroughs that could have the most potential customers for the restaurant.

3. Venues retrieving

- Venues data in each neighborhood were obtained from the Foursquare API.

4. Neighborhoods clustering

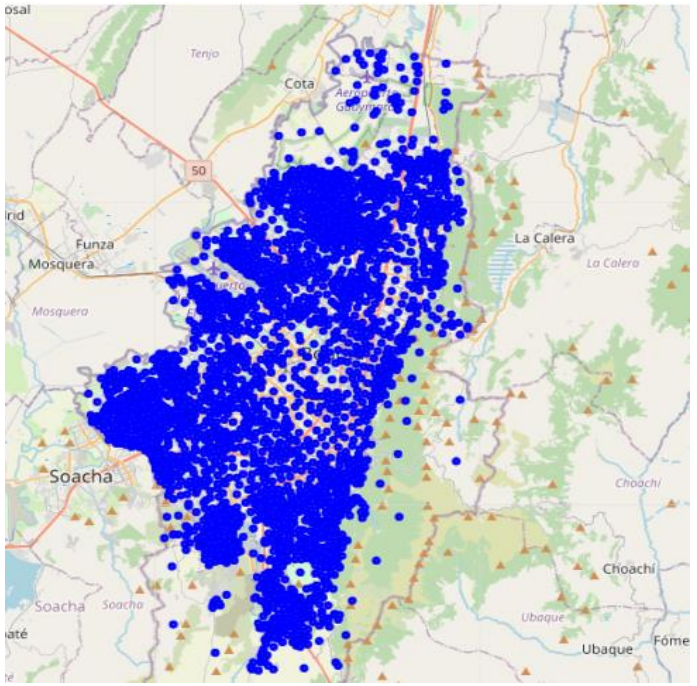
- The clustering was performed using the K-means model based on the most common restaurant categories in each neighborhood.

Analysis: Neighborhoods

The data of the neighborhoods and the boroughs with the corresponding geospatial position were imported and cleaned.

```
Number of rows dropped for not belonging to any borough: 3
Number of rows dropped because the geospatial position was missing: 3
The dataframe of the neighborhoods in Bogotá has 3865 rows and 5 columns.
```

ID_N	ID_B	Neighborhood	Borough	Latitude	Longitude
622	9	S.C. Modelia Occidental	Fontibón	4.663872	-74.126101
623	9	Mallorca	Fontibón	4.667070	-74.129611
631	11	Cantalejo Sector Alejandría	Suba	4.745914	-74.057101
642	1	Montearroyo	Usaquén	4.716169	-74.024873
654	1	Escuela de Caballería II	Usaquén	4.680785	-74.037319



Map of Neighborhoods of Bogotá. The Neighborhoods are represented as blue dots.

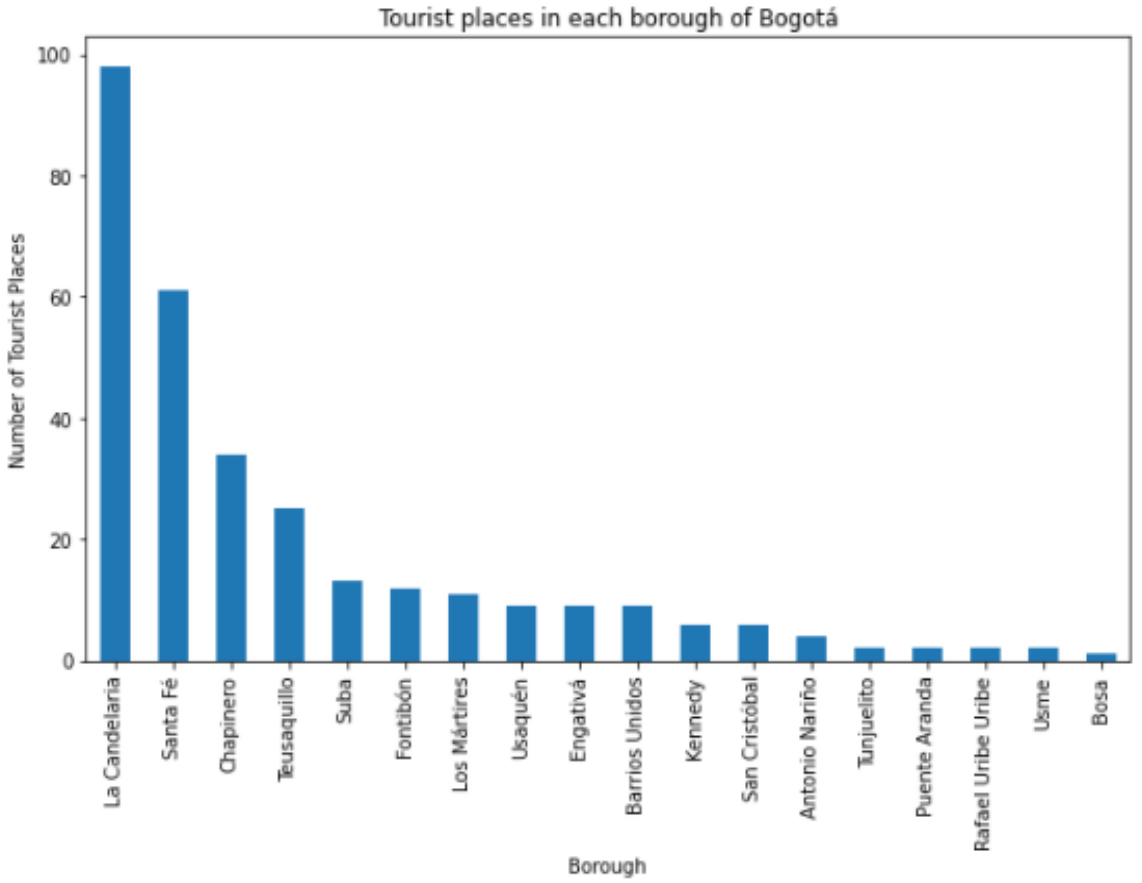
Analysis: Boroughs

- Since the amount of neighborhoods obtained is very high (3865). The boroughs are going to be explored first in order to only consider the most suitable.
- Bogotá is composed of 20 boroughs. The 20th borough (Sumapaz) is mostly rural and is less populated (4). This borough was not had into account in the database initially.

ID_B	Borough	Latitude	Longitude
1	Usaquén	4.734712	-74.031662
2	Chapinero	4.649913	-74.050649
3	Santa Fe	4.595280	-74.069871
4	San Cristóbal	4.557628	-74.086110
5	Usme	4.501817	-74.110217
6	Tunjuelito	4.578853	-74.138583
7	Bosa	4.619518	-74.191555
8	Kennedy	4.628286	-74.155276
9	Fontibón	4.675987	-74.141078
10	Engativá	4.700351	-74.114939
11	Suba	4.738390	-74.079011
12	Barrios Unidos	4.671068	-74.072016
13	Teusaquillo	4.638342	-74.086160
14	Los Mártires	4.606766	-74.088545
15	Antonio Nariño	4.590737	-74.105088
16	Puente Aranda	4.612981	-74.113413
17	Candelaria	4.594495	-74.072441
18	Rafael Uribe	4.562269	-74.114165
19	Ciudad Bolívar	4.555813	-74.153630

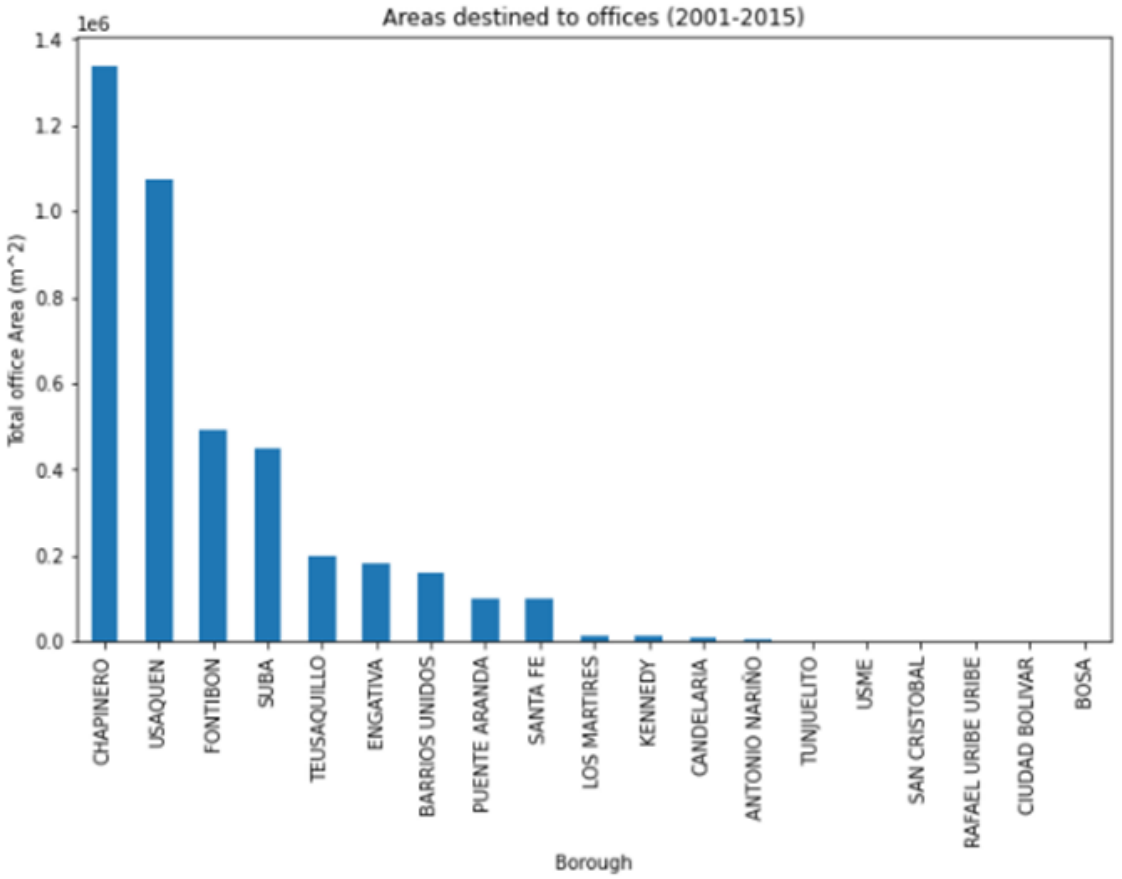
Analysis: Tourist places

- Restaurants located in these places are interesting since could be very attractive for tourists.
- The borough with more tourist places is Candelaria, followed by Santa Fé and Chapinero.



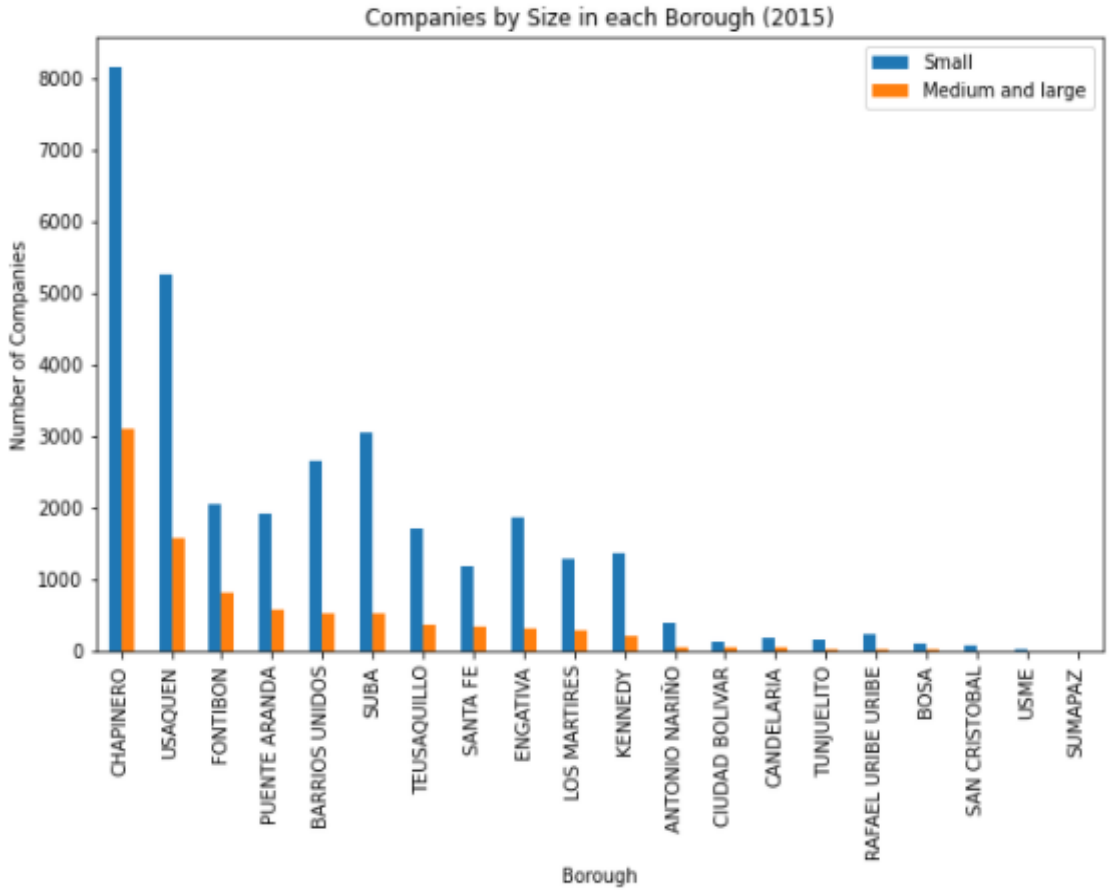
Analysis: Areas destined to Offices

- The areas destined to offices between the years 2001 and 2015 were summed to get an approximation of the total areas that are currently used for offices.
- The higher amounts of areas used for offices are in the boroughs Chapinero and Usaquén.



Analysis: Companies

- Data with information of companies subscribed to the chamber of Trade and Industry in 2015 were used.
- The boroughs Chapinero and Usaquén have again the higher number of Large and small companies.
- This information could be useful to find administrative and corporate zones, where many people work and need a restaurant close for meals.

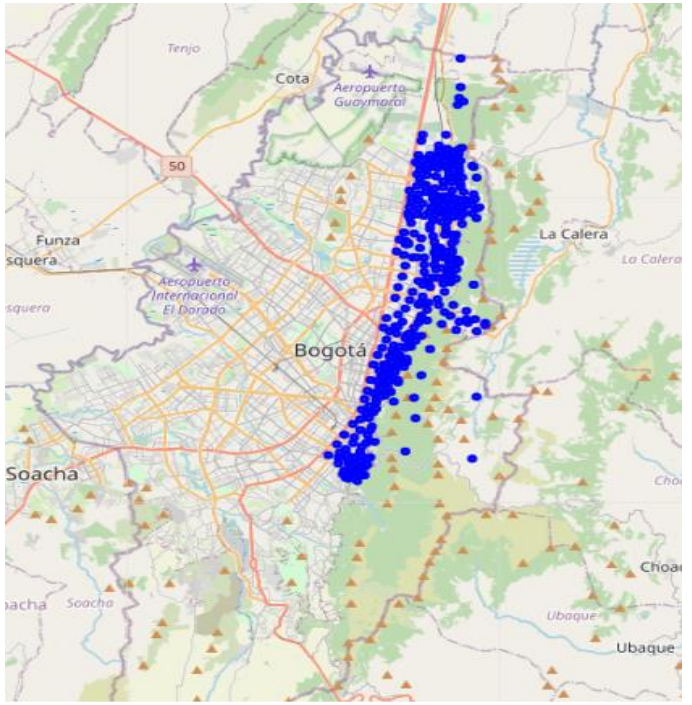


Analysis: Selection of Boroughs

- Based on the previuos analyzis only four bouroghs were having into account: Candelaria, Santa Fé, Chapinero and Usaquén.
- The selected zone corresponds to the east of the city.

The neighborhoods dataframe of the selected boroughs has 455 rows and 5 columns.

ID_N	ID_B	Neighborhood	Borough	Latitude	Longitude
642	1	Montearroyo	Usaquén	4.716169	-74.024873
654	1	Escuela de Caballería II	Usaquén	4.680785	-74.037319
656	1	La Glorieta	Usaquén	4.696209	-74.027298
663	1	Cerros de Santa Bárbara	Usaquén	4.692080	-74.026955
736	3	Cartagena	Santa Fe	4.579915	-74.076817



Map of Neighborhoods in the selected boroughs.

Analysis: Importing venues from Foursquare API

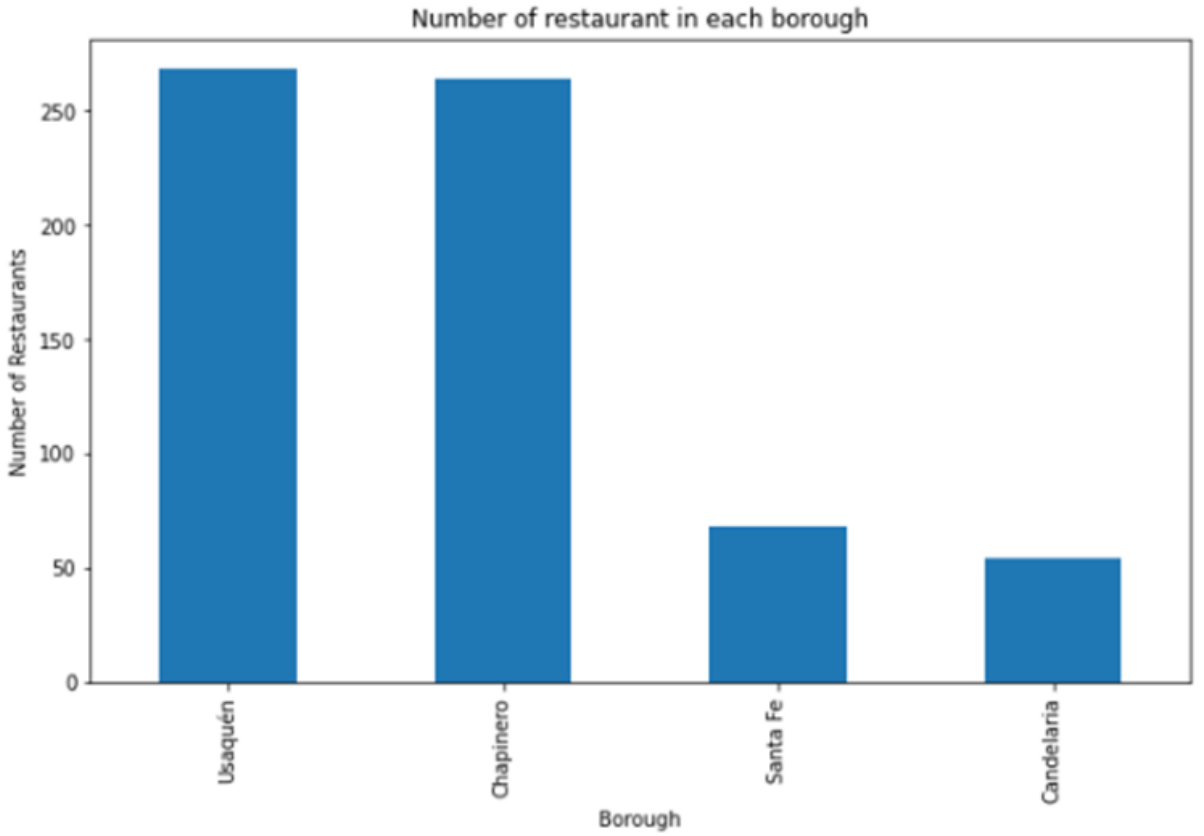
- The Foursquare API is used to retrieve the most popular venues in each location.
- A limit of 100 venues in a radius of 250 meters is selected for each neighborhood.

```
The total number of venues obtained is 2356.
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Escuela de Caballería II	4.680785	-74.037319	OMA Capital Tower	4.679658	-74.038817	Breakfast Spot
1	Escuela de Caballería II	4.680785	-74.037319	La Española	4.681002	-74.039371	Spanish Restaurant
2	Escuela de Caballería II	4.680785	-74.037319	Sopas de Mama y Postres de la Abuela	4.680778	-74.038372	Latin American Restaurant
3	Escuela de Caballería II	4.680785	-74.037319	Restaurante el Fogón Casero	4.680882	-74.039000	Restaurant
4	La Glorieta	4.696209	-74.027298	Catación Pública	4.695898	-74.028142	Coffee Shop

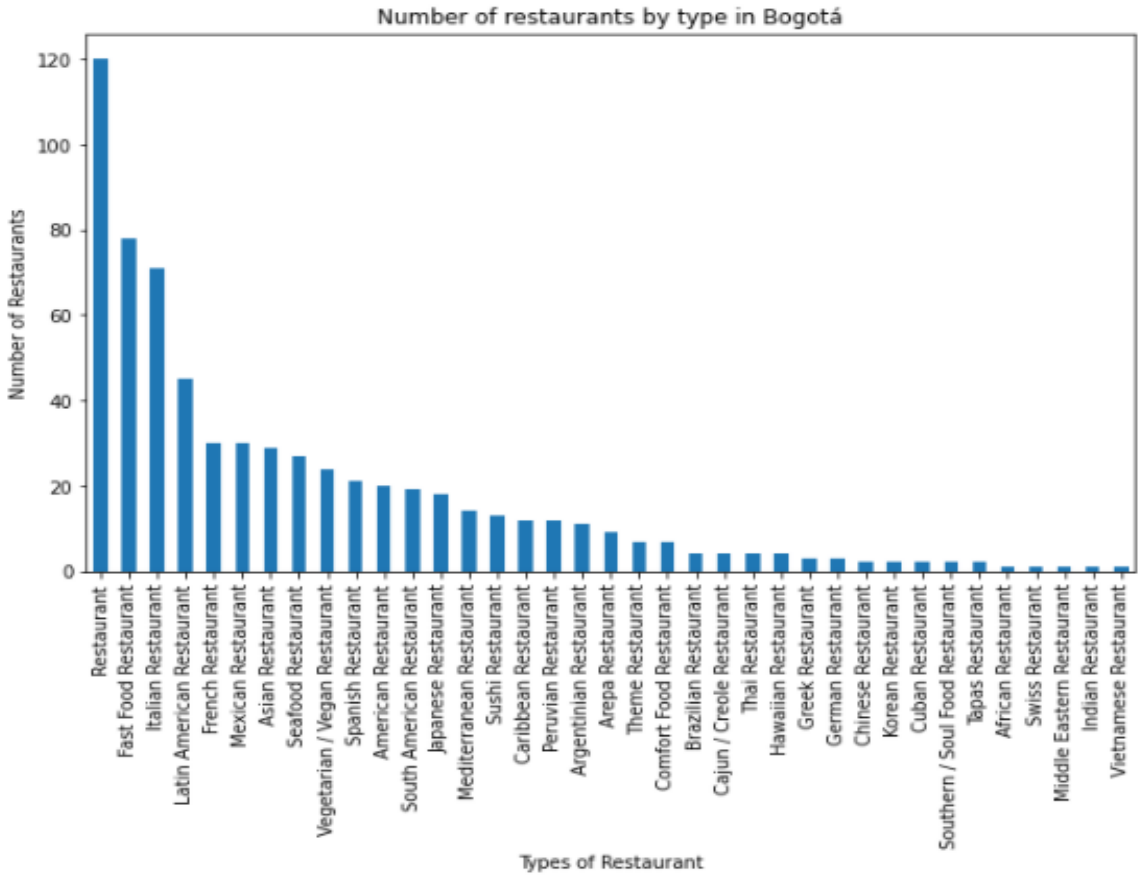
Analysis: Number of restaurants in each borough

- All the venues that do not contain the word restaurant in the venue category column are dropped.
- The boroughs Usaquén and Chapinero have a considerable difference in comparison with the two last ones (Santa Fé and La Candelaria).



Analysis: Restaurant category venues

- A total number of 654 restaurant venues is obtained.
- Most of the venues are only categorized as restaurants, therefore these do not belong to any specific category.
- The Italian and fast-food restaurants would be the type of restaurants more common in the chosen boroughs.



Analysis: Clustering

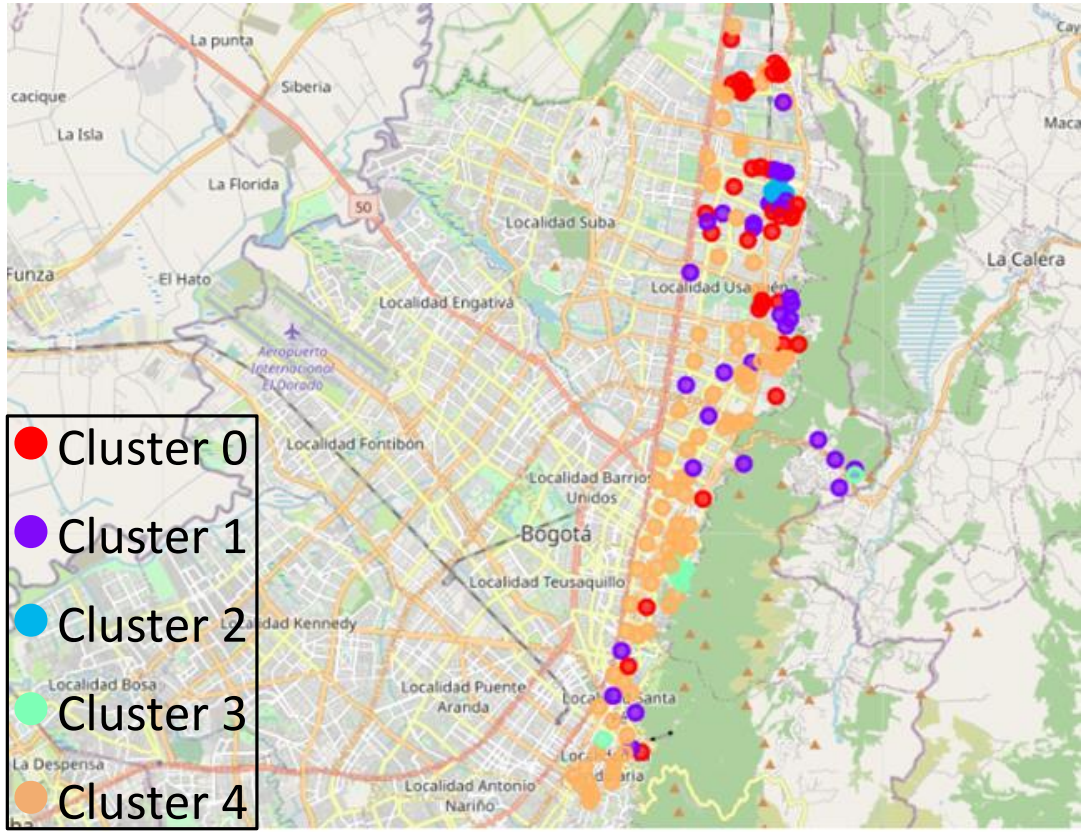
Before segmenting the Neighborhoods in clusters, the data should be arranged:

- The dummy function was used to create a binary column for each category
- The column with the neighborhoods was also inserted to this dataframe.
- The rows were grouped by neighborhood and the mean frequency is taken for each category.

	ID_B	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	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
0	1	Escuela de Caballería II	Usaquén	4.680785	-74.037319	4	Restaurant	Latin American Restaurant	Spanish Restaurant	Vietnamese Restaurant	Greek Restaurant	German Restaurant	French Restaurant	Fast Food Restaurant
1	1	La Glorieta	Usaquén	4.696209	-74.027298	4	Italian Restaurant	Vegetarian / Vegan Restaurant	Hawaiian Restaurant	Latin American Restaurant	Brazilian Restaurant	Cuban Restaurant	Greek Restaurant	German Restaurant
2	1	Los Sauces Norte	Usaquén	4.735974	-74.027742	2	Spanish Restaurant	Vietnamese Restaurant	Chinese Restaurant	Hawaiian Restaurant	Greek Restaurant	German Restaurant	French Restaurant	Fast Food Restaurant
3	17	La Catedral	Candelaria	4.599540	-74.073553	4	Latin American Restaurant	Restaurant	Seafood Restaurant	Comfort Food Restaurant	Italian Restaurant	Arepa Restaurant	Argentinian Restaurant	Vegetarian / Vegan Restaurant
4	1	Santa Bárbara Central III Sector	Usaquén	4.694677	-74.035788	1	Restaurant	Vietnamese Restaurant	Comfort Food Restaurant	Hawaiian Restaurant	Greek Restaurant	German Restaurant	French Restaurant	Fast Food Restaurant

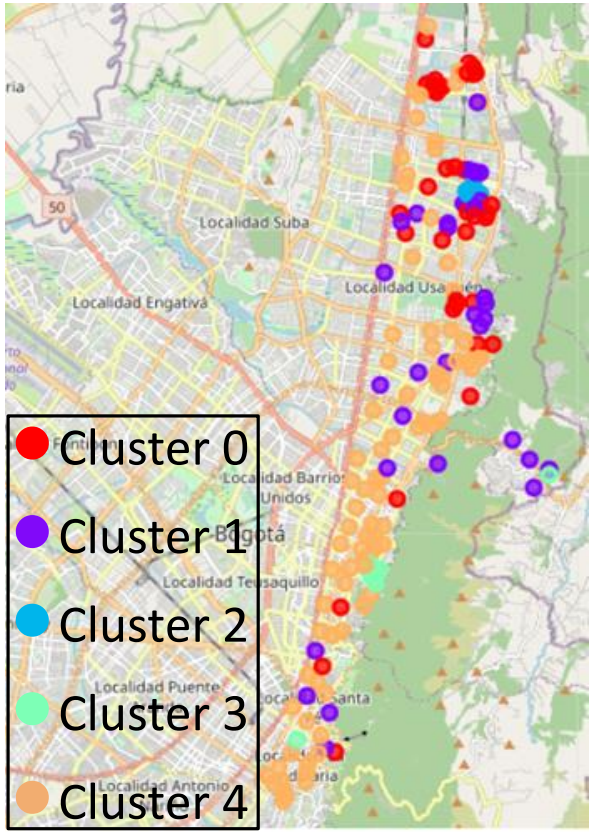
Analysis: Clustering

- Some neighborhoods had to be dropped since for them, data were not obtained from the Foursquare API.
- The neighborhoods were clustered in 5 groups using the K-means model.



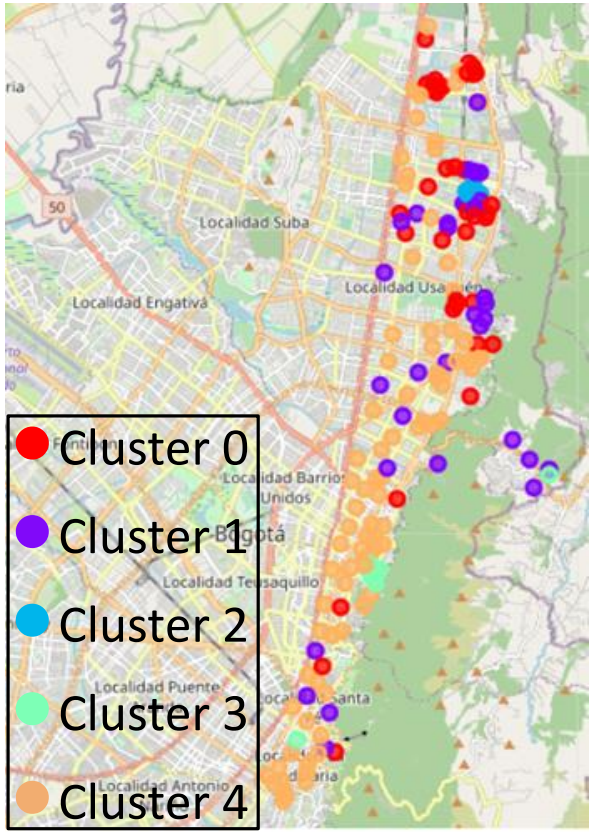
Results: Clusters

- **Cluster 0:** In this cluster, the fast-food restaurants are the most common restaurant types. These neighborhoods are mostly located in Usaquen.
- **Cluster 1**: This cluster corresponds mainly to neighborhoods with restaurants without a defined category assigned and other international food restaurants. The neighborhoods of this cluster are located mostly north in the boroughs of Chapinero and Usaquén.
- **Cluster 2**: The Spanish restaurants are the most common in these areas. All the neighborhoods belonging to this cluster are located in Usaquén.



Results: Clusters

- **Cluster 3**: The Italian and other international restaurants are the most common in the neighborhoods belonging to this cluster. The neighborhoods that belong to this cluster are located in the boroughs Chapinero and Santa Fé.
- **Cluster 4**: In this cluster, the most frequent are the Latin American/southamerican restaurants. These clusters are more common in all the boroughs explored but mainly in the south.



Conclusions

- In this project the data available of the neighborhoods in Bogotá, Colombia was used to explore and cluster them in order to obtain insights into the distributions of restaurants in the city. It is expected that this analysis could help investors from the catering industry who are planning to start a business in one of Bogotá neighborhoods.
- A deeper exploration is needed in order to obtain a better description that allows determining the boroughs that have the best characteristics or have a higher possible number of customers for a restaurant. In order to do this, more data would be needed, as for example information about urban zones, malls, trading points, etc.
- The neighborhoods were clustered in five groups based on the most common restaurant categories, in each one of them. This allowed finding areas where there are similar types of restaurants.

References

(1) Visit Colombia feel the rhythm (2021/01) Tourism in Bogota: a city for experiencing culture. https://colombia.travel/en/blog/tourism-bogota-city-experiencing-culture

(2) Tripadvisor. (2021/01) Explora Bogotá. https://www.tripadvisor.co/Tourism-g294074-Bogota-Vacations.html

(3) Laboratorio Urbano Bogotá´. (2015/01) https://bogota-laburbano.opendatasoft.com/explore.

(4) Alcaldía de Bogotá D.C. (2021/01) https://bogota.gov.co/mi-ciudad/localidades/sumapaz