

# Best Location for a New Restaurant in Bogotá, Colombia

Andres Ocampo Dávila

January 2021

## 1. Introduction: Business Problem

Through this project we will collect and deliver information that helps our target audience to find the optimal location for a new restaurant in Bogotá. The results of this analysis are intended to be used by entrepreneurs interested in opening a restaurant of any given type in one of the 21 *Localidades* (boroughs) of the Colombia's capital city.

In order to deliver valuable information to our target audience, we will identify those city zones where the restaurant and food service industry are already located and what type of restaurants or food business are running. Also, we will identify locations that have not plenty of restaurants already. This way we will provide information so the most promising neighborhoods and the best possible location can be chosen by the entrepreneur.

Based on our problem definition, as a result of the analysis the information described below will be delivered:

- number of existing restaurants in every neighborhood (any type of restaurant).
- number of existing restaurants in every neighborhood by type/category, if any.
- segmentation of neighborhoods by type/category of restaurant or food business.

## 2. Data Acquisition and Cleaning

### 2.1. Data Sources

The following data sources will be used to generate the required information:

- geographic information (latitude and longitude) of every neighborhood of Bogotá city, will be obtained from Laboratorio Urbano Bogotá API. API documentation [here](#).
- geographic information (latitude and longitude) of every *Localidad* of Bogotá city, will be obtained from Laboratorio Urbano Bogotá API. API documentation [here](#).
- number of restaurants/food business and their type and location in every *Localidad*, will be obtained from Foursquare API.

## 2.2. Data cleaning and wrangling

To begin with, it is necessary to note that Python, running on an instance in IBM Cloud Pak for Data, has been used for all data processing and analysis.

From Laboratorio Urbano Bogotá API we get a JSON file which includes among other features the geographic information (latitude and longitude) for every *Localidad* in Bogotá. From that file the relevant part is extracted and is later transformed it into a Pandas dataframe. We keep only the columns that include *Localidad* name and its location (latitude and longitude) and rename the columns for simplicity. This data is saved into a Pandas dataframe to be used later.

	Localidad	Latitude	Longitude
0	Barrios Unidos	4.666400	-74.084000
1	Engativá	4.707100	-74.107200
2	Sumapaz	4.034746	-74.315224
3	Teusaquillo	4.644800	-74.093800
4	La Candelaria	4.593900	-74.073900

In a similar way, From Laboratorio Urbano Bogotá API we get a JSON file which includes among other features the geographic information (latitude and longitude) for every Neighborhood in Bogotá. From that file the relevant part is extracted and is later transformed it into a Pandas dataframe. We keep only the columns that include Neighborhood name, the corresponding *Localidad* and the neighborhood's location (latitude and longitude). Exploring the data some null values are identified. Around 78% of the data records are complete. Any null value is removed from the data since we need the *Localidad* and the location for every Neighbourhood for further analysis. We rename the columns for simplicity. This data is saved into a Pandas dataframe to be used later.

	Neighbourhood	Localidad	Latitude	Longitude
0	Batavia	Fontibón	4.685055	-74.150638
1	Belmira	Usaquén	4.719588	-74.030167
2	Las_Acacias	Usaquén	4.721538	-74.029576
3	Carmen del Sol I Sector	Rafael Uribe	4.566524	-74.114154
4	Urb. Furatena	Usaquén	4.703728	-74.027762

The top 100 venues that are within a radius of 500 meters from the coordinates of each neighborhood are obtained from the Foursquare API. A Pandas dataframe is created with the features: Neighborhood, *Localidad*, Neighborhood Latitude, Neighborhood Longitude, Venue, Venue Latitude, Venue Longitude and Venue Category.

Since we have a limit for the free calls to the Foursquare API, we will randomly select a sample of the data (a list of neighborhoods) for further analysis.

From this Foursquare dataset, we identify those venues whose categories are related to restaurants and food businesses, selecting only the venue categories of interest for our analysis. This way we obtain a dataframe with the information (Name, Category and Location) of the restaurants and food businesses in the neighborhoods of Bogotá to be used later.

	Neighbourhood	Localidad	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Nueva Villemar	Fontibón	4.671135	-74.136278	Punta Camaron	4.673152	-74.139024	Seafood Restaurant
1	Nueva Villemar	Fontibón	4.671135	-74.136278	Juan Valdez Café	4.667768	-74.134086	Coffee Shop
2	Nueva Villemar	Fontibón	4.671135	-74.136278	panadera zona gourmet	4.671727	-74.137035	Bakery
3	Nueva Villemar	Fontibón	4.671135	-74.136278	Witt Company Pizza Y Alitas Bbq	4.669465	-74.140160	Pizza Place
4	Urb. Nueva Granada	Chapinero	4.648046	-74.054591	Amor Perfecto	4.648939	-74.055450	Coffee Shop

### 3. Methodology

In order to deliver information that helps our target audience finding the optimal location for a new restaurant in Bogotá, and based on the data described in the previous section, we are going to:

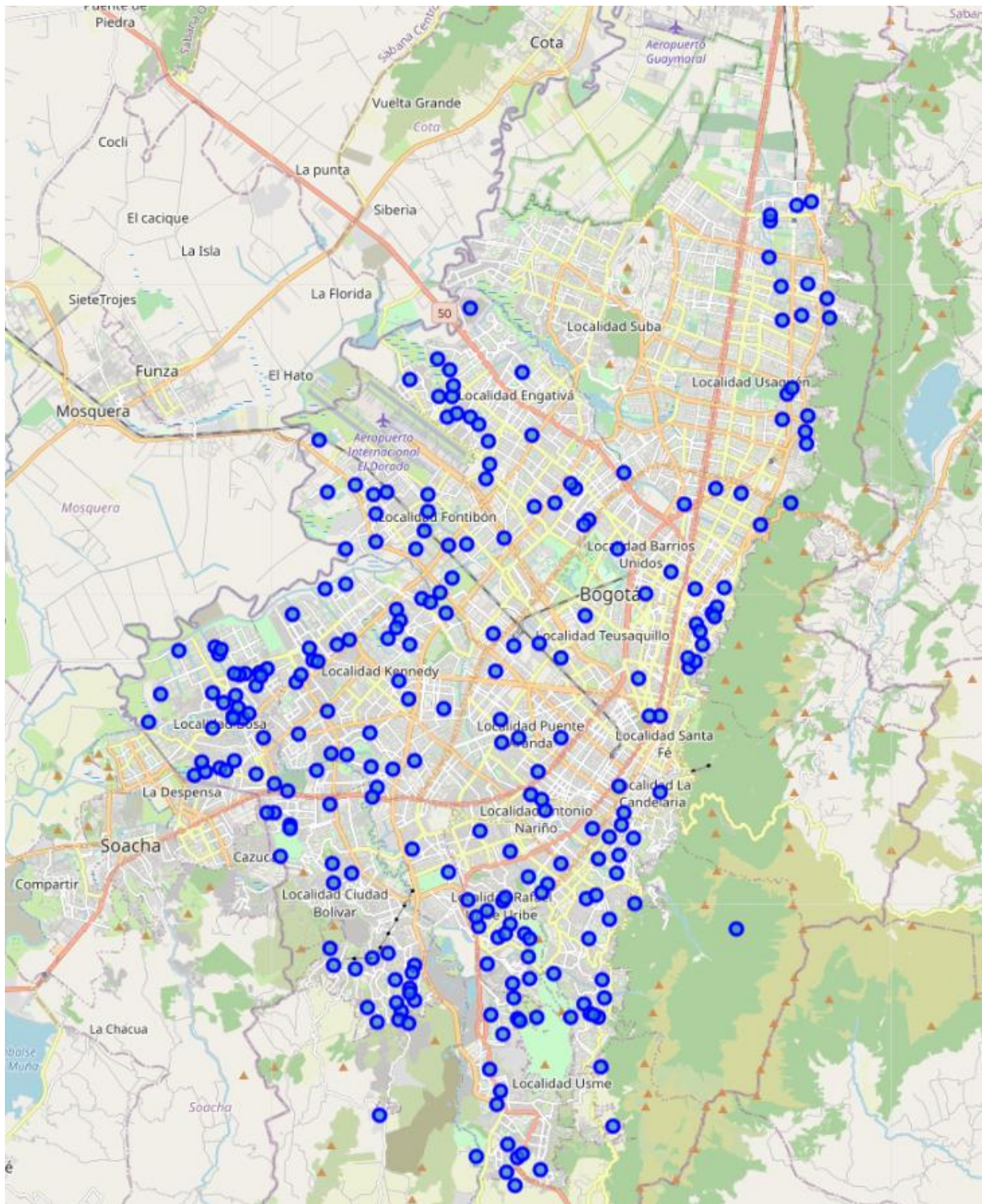
- Explore and arrange the data to identify the Neighborhoods with the highest and lowest number of existing restaurants (any type of restaurant) and later use visualization tools (maps) with this data for better understanding.
- Explore and arrange the data to identify the number of existing restaurants in every *Localidad* and later use visualization tools (maps) with this data for better understanding.
- Explore and arrange the data to identify the number of existing restaurants by type/category.
- Segmentate the Neighborhood by type/category of restaurant or food business using k-means method.

### 4. Data Analysis

#### 4.1. Number of restaurants and food businesses by Neighborhood in Bogotá

Here, we count the number of restaurants and food businesses, grouped by the Neighborhood where they are located. From there, we can identify the neighborhoods with the higher number of restaurants and those with the lower number of restaurants as well.

Below, we have a Bogotá map with all the neighborhoods that will be used for the analysis marked on it.

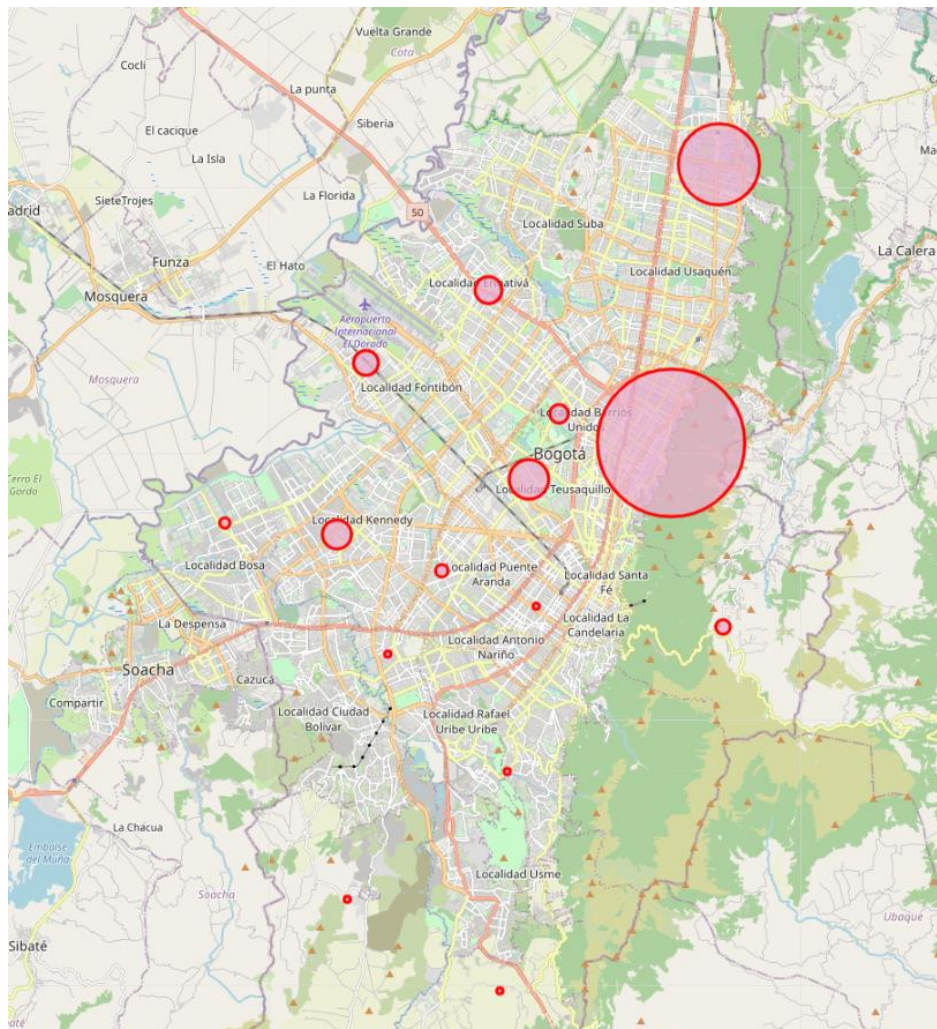
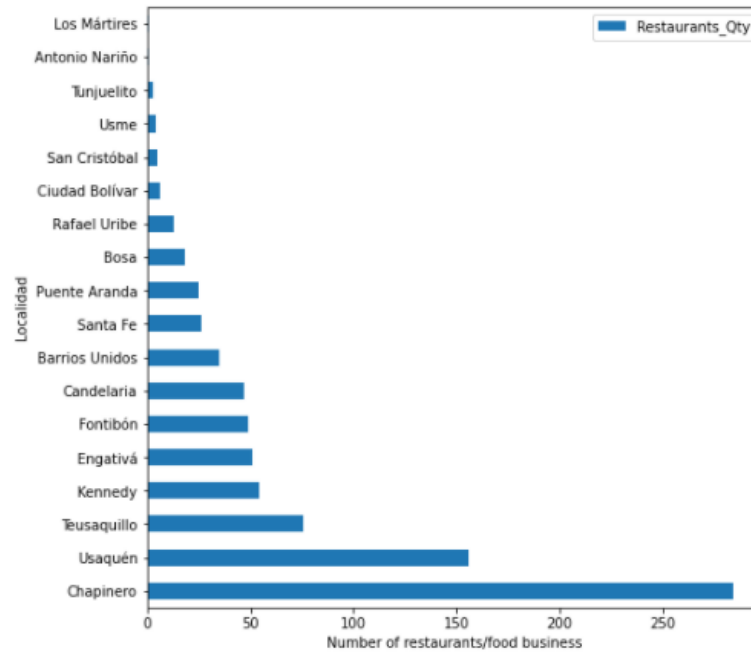


Since we have over 140 neighborhoods included in the analysis, we selected the top 10 neighborhoods by the number of restaurants and food businesses, and the list of the 10 neighborhoods with lowest number to be considered.

The following maps shows:



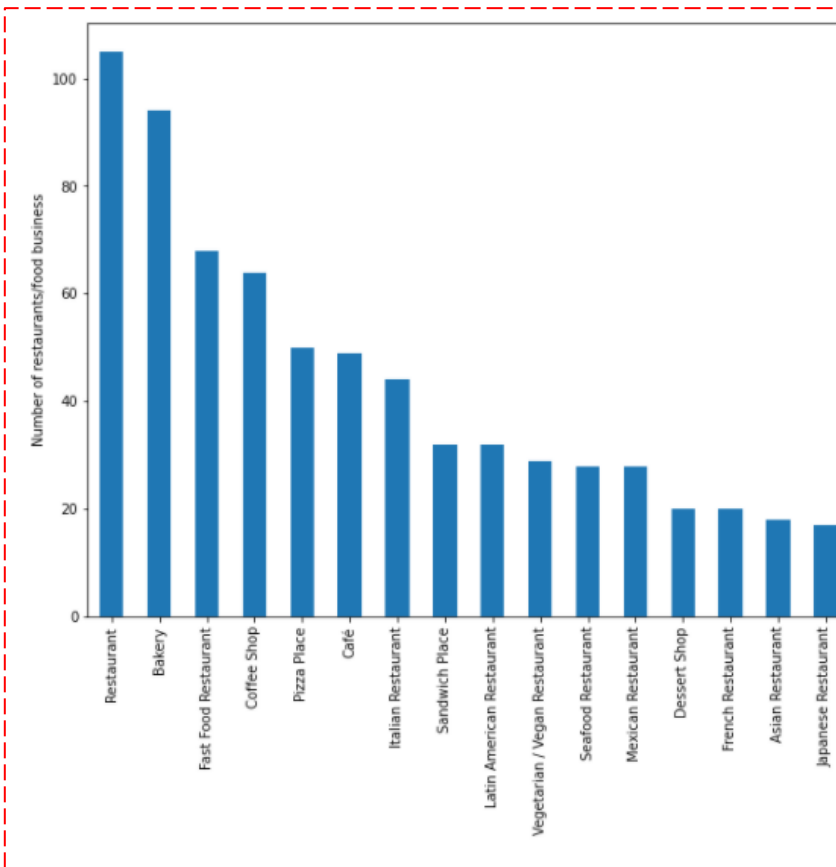
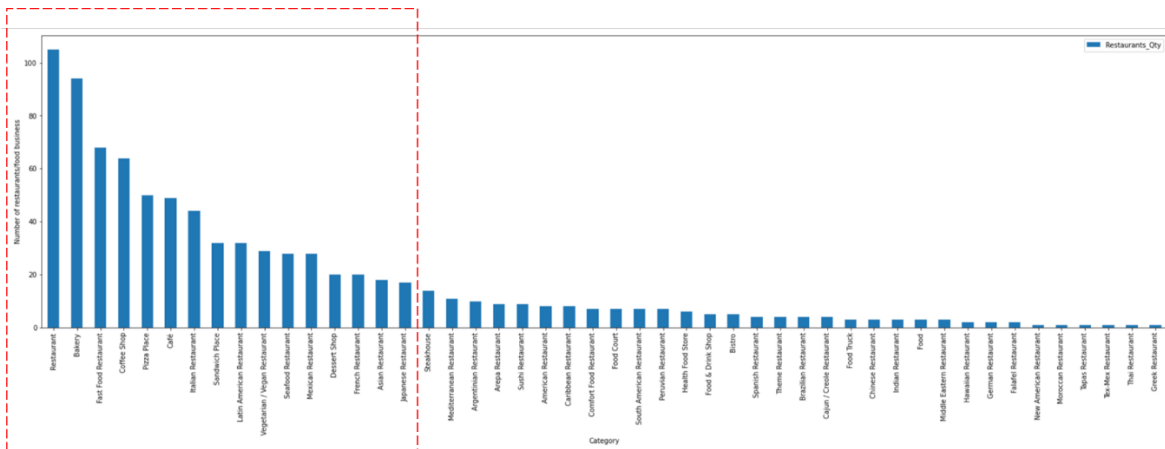




### 4.3. Number of restaurants and food businesses by category in Bogotá

Here, we count the number of restaurants and food businesses, grouped by category or type, for the venues included in the sample.

In the following graph, we can see the number of restaurants and food businesses by Category in Bogotá. It is shown that bakeries, fast food restaurants, cafes/coffee shops, Italian restaurants and pizza places are the most common categories/types of restaurants in the city.

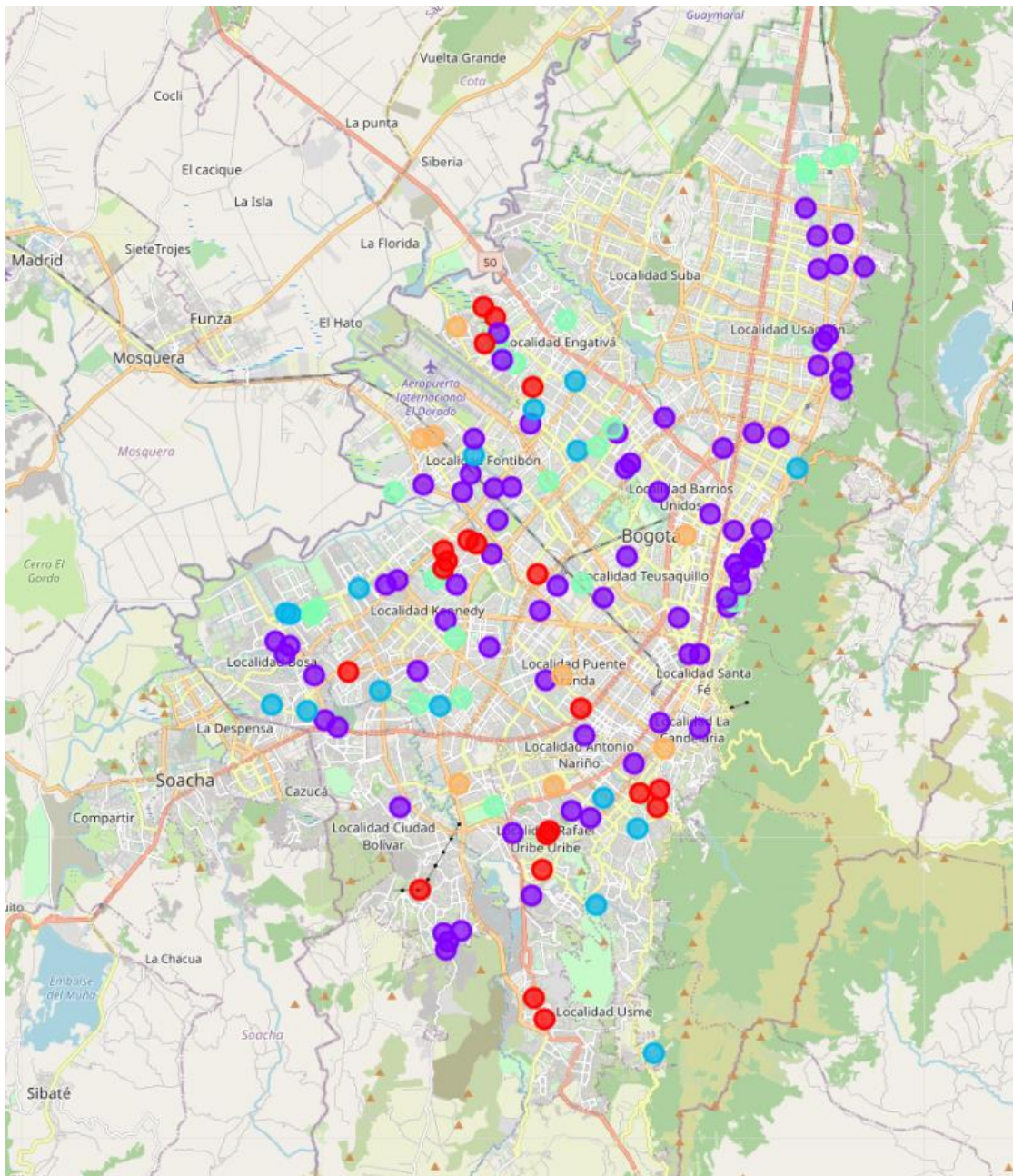




#### 4.4. Segmentation of the Neighborhoods by categories of restaurants using k-means method.

Here, we cluster the neighborhoods included in the sample, based on the mean of the frequency of occurrence of each category/type of restaurant or food business.

On the following map we can see the geographical distribution of the restaurants/food business, colored according with the assigned cluster (*red = cluster 0, purple = cluster 1, blue = cluster 2, light green = cluster 3, orange = cluster 4*).



Supporting this information, we create a new dataset which contains the top 5 Restaurant and Food Business Categories for each Neighbourhood. This way we can identify the specific Restaurant and Food Business Categories that determine any given



cluster. As an example, on the following table, we can see the categories that define the cluster "0", been compared with the cluster "4".

Cluster "0"		Cluster "2"	
Bakery	21	Restaurant	16
Vegetarian / Vegan Restaurant	19	French Restaurant	15
French Restaurant	19	Vegetarian / Vegan Restaurant	15
Comfort Food Restaurant	18	Coffee Shop	14
German Restaurant	17	Food Truck	10
Coffee Shop	3	Seafood Restaurant	2
Pizza Place	2	Mediterranean Restaurant	1
Dessert Shop	2	Pizza Place	1
Seafood Restaurant	1	Mexican Restaurant	1
Sandwich Place	1	Comfort Food Restaurant	1
Mexican Restaurant	1	Food & Drink Shop	1
Food Truck	1	Steakhouse	1
		Bakery	1
		Chinese Restaurant	1

## 5. Results

In order to answer the questions given by our business problem, the analysis performed based on the information available on Foursquare, has shown that in Bogotá the number of restaurants and food businesses vary widely from one neighborhood to another, having some neighborhoods with scarce or no restaurants running, and others with a high amount of restaurants or food businesses operating. The situation is very similar for the *Localidades* (boroughs).

We have identified that a high number of restaurants and food service businesses are in neighborhoods located on the east and northeast of the city. A similar trend is observed when we analyze the distribution among *Localidades*: The highest number of restaurants are located on the center, east and northeast parts of the city.

Those are the areas where the restaurant and food service industry are already located, so can be identified as potentially interesting Neighborhoods and Localidades for our entrepreneurs to start a new business, since those are zones where the potential costumers usually go searching for a place to eat.

On the other hand, the neighborhoods and Localities on the south of the city have a much lower number of restaurants on service, which can be a business opportunity for our entrepreneurs, since there may be a demand not being attended yet.

Regarding of the type of restaurants and food businesses in service, our analysis has shown that bakeries, fast food restaurants, cafes/coffee shops, Italian restaurants and pizza places are the most common categories/types of restaurants in the city. There are also a considerable number of restaurants without a classification (around 15%), and they are shows simply as restaurants, so they cannot be included in this part of the analysis.

As a result of the clustering performed based on the restaurant type, we have five groups of neighborhoods/locations, where we can see that bakeries, coffee shops,

vegetarian/vegan restaurants are spread all around the city, while fast food restaurants are very common on the center and the south parts of the city.

This information, all together, could be used to evaluate for a given location the current offer of restaurants, in terms of quantity and type, so the most promising neighborhoods and the best possible location can be chosen.

## **6. Conclusions**

The objective of this project was to collect and deliver information about the most promising neighborhoods and the best possible location that can be chosen by those interested in opening a restaurant in Colombia's capital city, Bogotá.

For this purpose, we had brought together the geographic information of the Bogotá's neighborhoods and *Localidades*, with the Foursquare data about the distribution of restaurants and food businesses around the city. From there, through an exploratory analysis, we were able to identify the number of existing restaurants in every neighborhood and *Localidad*, the number of existing restaurants by type/category, and we could segmentate the locations into five groups based on the type/category of restaurant or food business.

With this valuable information our entrepreneurs can make a better decision, based on the potential demand for their products, considering if the neighborhood is a location already established as a restaurant/food service zone or if there is a demand to be supplied, and considering what kind of offer exists in the area, regarding to the type of restaurants on service.