**Reducing travel time between home and work**

**Problem**

At present, Bogotá lives one of the worst vehicle congestion in its history (https://www.bbc.com/mundo/noticias-39045803), this can be verified according to an article from the BBC world, based on this problem we intend to perform a market study to find the best place to establish a business (in this exercise we will install a cafeteria) and at the same time find a place of residence within the selected locality, in order to make viable the undertaking of a new business and reduce the time of displacement between these two places.

**Target audiences**

This project is intended for that entrepreneurial population or who want to be independent, who wants to have a good location between business and housing to reduce travel times and thus greatly improve their quality of life.

**Data**

According to the definition of our problem, the factors that will influence are:

* Concentration of commercial establishments in neighborhoods.
* Types of commercial establishments that do not compete with ours.
* Areas that stand out for their residential environment.
* Distance between the housing location and the commercial location.

The following data sources will be needed to extract and generate the required information.

* Geo-coded locations of the neighborhoods in the city of Bogotá extracted from the Bogotá Urban Laboratory

https://bog ota-laburbano.opendatasoft.com/explore/dataset/barrios\_prueba/table/

* Number of commercial establishments, type and location, which will be obtained from the Foursquare API.
* Distance from residential and commercial areas that will be checked between the data sources of Laboratorio Urbano Bogotá and the Foursquare API.

**Data cleaning and feature selection**

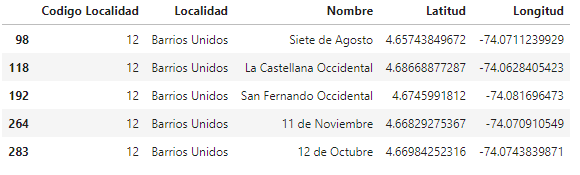
The database of the Bogotá Urban Laboratory is made up of 3871 rows and 10 columns described as follows:

* OBJECTID
* Codigo localidad
* Localidad
* Estado
* Nombre
* Codigo
* SHAPE.AREA
* SHAPE.LEN
* Geo\_shape
* geo\_point\_2d

Of which they will be used:

* Codigo Localidad : Numero con el cual se identifica cada localidad
* Localidad : Nombre de la localidad
* Nombre : Nombre del barrio
* geo\_point\_2d : Ubicación geográfica del barrio

Once these columns have been extracted, we proceed to eliminate the spaces without information, if there is what leaves us 3871 rows and 4 columns, which indicates an acceptable quality of the data, finally it is necessary to divide in the geo\_point\_2d column that it contains in a single string Both latitude and longitude, this procedure is necessary to achieve a correct intake of the data, leaving a size of 3871 rows by 5 columns. The last filter made is the choice of the locality "Barrios Unidos" which leaves us with a matrix of 66 rows and 5 columns.



**Methodology**

In this project we will direct our efforts to find two specific places with different characteristics within the selected locality.

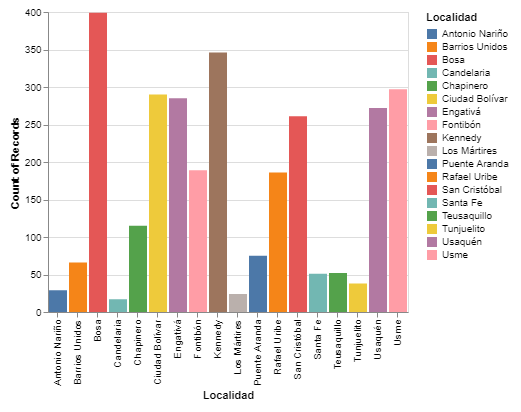
In the first step we will search within the locality which are the neighborhoods that concentrate most of the commerce, since this assures us that in these areas we will have greater probability of success in sales for our store, followed by this analysis we will exclude those neighborhoods that already have cafeteria businesses since the intention is to enter an area that has low competition in this field.

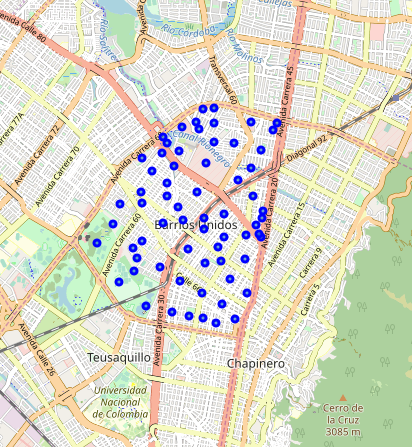
In the second step we will search within the locality for the residential neighborhoods that we will identify according to the recreation or rest areas also that have a low density of trade but enough for small supplies.

The third and final step will be to focus on the most promising areas of both housing and commercial that meet our criteria, we will compare the locations and define which are the two points that have the least distance since our goal is to reduce the travel time between these two places.

**Exploration and analysis**

First it was necessary to observe the distribution of neighborhoods in the localities to decide which one would work with, due to a limiting technique such as the low number of queries in the Foursquare API it was decided to work with “United Neighborhoods” that has a number of neighborhoods of 66.



****With the help of the folium library we mark the neighborhoods in the chosen locality.

This first impression clearly shows the limits of the locality on the territory and gives us an idea of where the most commercial areas can be located, but to reach a clear concept we must enrich our data even more.

Consult the Foursquare API

Through the use of the Foursquare API we proceed to consult the places that are in the vicinity of our neighborhoods, in this way we discover what are the types of establishment that is linked to each of our points, to further analyze our data is you need to convert them to a one hot array.

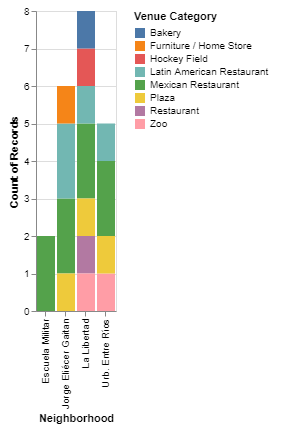
A quick count of the most popular places shows a high number of restaurants, possibly an indicator of a commercial wealth in the area, we also noticed that among our top 10 appear supermarkets, gardens and parks, this is important for our second purpose which is the Search for a place of housing.

****

To delve deeper into our data we use an unsupervised learning model of closterizacion called Agglomerative Clustering, once we get the result we analyze each of the resulting groups.

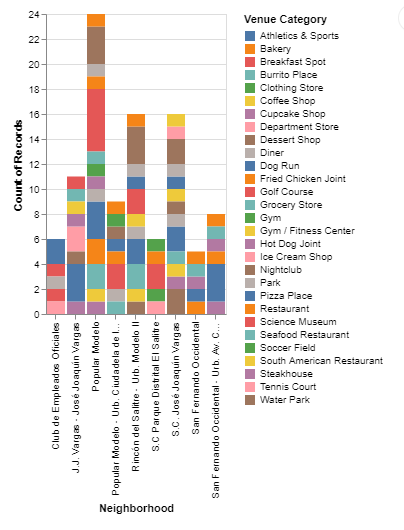
To facilitate the graphic analysis, we mark the neighborhoods in purple and commercial establishments in red on the map.

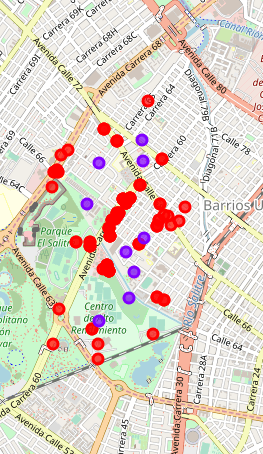
Cluster 0

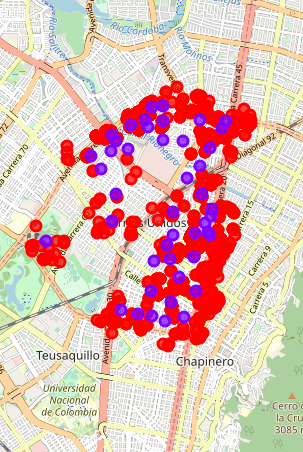
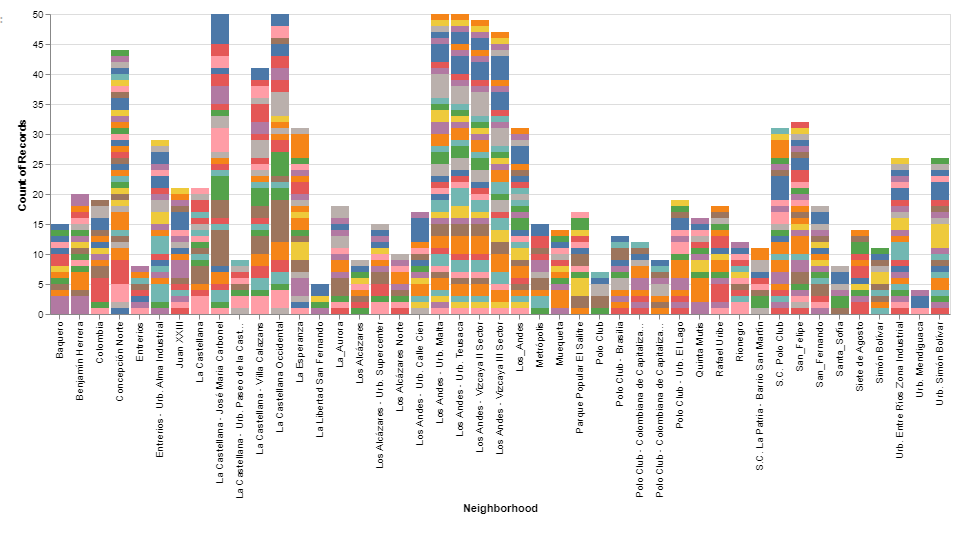
In group 0 we see that the places near the neighborhoods are mostly residential areas, the low volume of trade indicates a good option to find our home.

Cluster 1



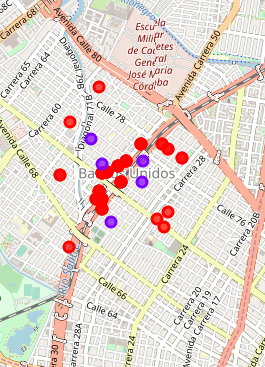
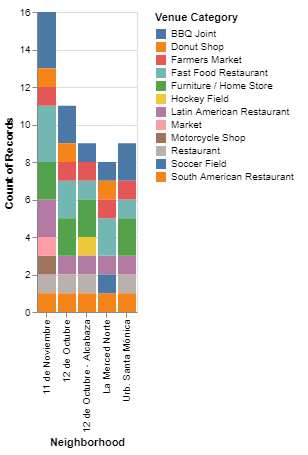


Cluster 1, although it has a high number of commercial stores, also highlights social recreation places such as tennis courts, water parks, gyms and soccer fields, emerging as housing neighborhoods.

Cluster 2

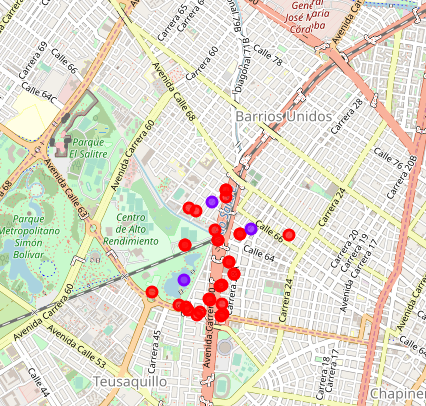
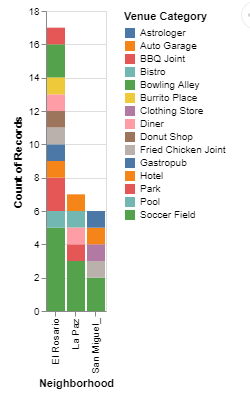
Cluster 2 stands out for being the one that concentrates the largest number of commercial premises, it also shows us that these establishments border the boundaries of the town, coinciding with the main avenues, these characteristics ensure a high flow of people and increase our probability of Have a high number of sales in business.

Cluster 3

This cluster is a small commercial area that is built around the avenue that runs through the town from south to north, just like cluster 2, the high traffic of people on these roads guarantees greater demand for products and therefore higher sales.

Cluster 4

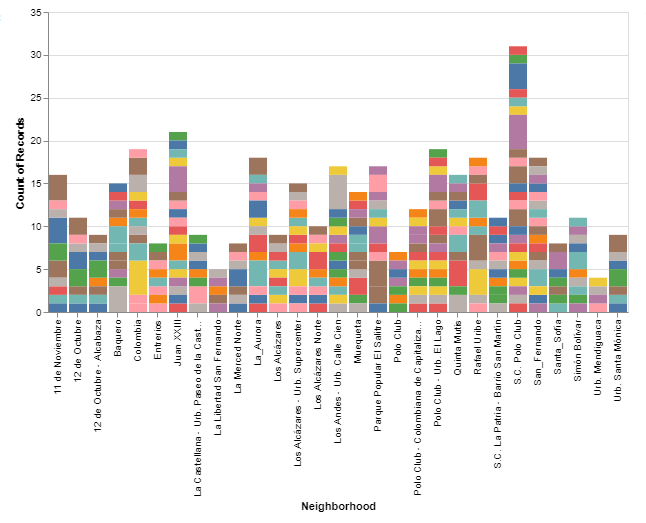
 

The places that make up this group are bowling alleys, soccer fields, parks, swimming pools and garages, this indicates that it can be a winery area, so we will discard this cluster for our search.

From the analysis performed it follows:

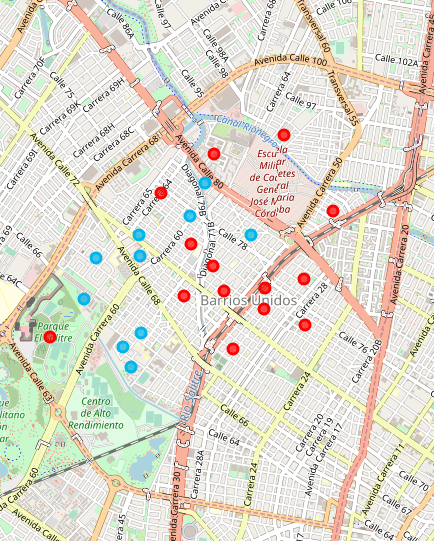
* Clusters 0 and 1 will be used to find our place of housing
* Clusters 2 and 3 will be used to find the best place to locate our business.

 Having decided this, we filter from clusters 2 and 3 based on the criterion of eliminating those neighborhoods that already have coffee establishments, we went from having 50 possible neighborhoods to 29 potential neighborhoods where to install our cafeteria.



For housing groups we proceeded to join clusters 0 and 1 and discard those places that are not neighborhoods such as the Official Employees Club, Military School, SC El Salitre District Park, since these would not fall into this category, this process we leave a total of 10 neighborhoods to work with to find our home.

Finally, the distances between the neighborhoods of the clusters with high commerce and those classified as housing options have been calculated, the distances between these two types of places were filtered, taking as a criterion that a person travels on average 5 km on foot at a time of 60 minutes, it was decided to find the places that are less than 15 minutes away which gives us an approximate one kilometer between them, which does not leave 24 places distributed in 10 housing neighborhoods and 14 commercial neighborhoods, in this way We concentrate our efforts on locating those 2 points that reduce our journey time between our home and our workplace.



The result of this last filter shows us on the blue map those places that we consider suitable for housing, and in red those places that have a high commercial volume, most of them with distances no longer than 15 minutes away on foot , although between some extreme points of the groups the route can be greater.

It is important to highlight that the “last mile” of this result would be based on going to the points we determine as our best option, thanks to technology we can make a virtual visit to two of these places that in my opinion are good options.

Housing point

<https://www.google.com/maps/@4.674412,-74.0721533,3a,75y,237.47h,92.78t/data=!3m6!1e1!3m4!1syPCgfN_T7JHALjeu3y1oDA!2e0!7i13312!8i6656>

Trading point

<https://www.google.com/maps/@4.6703408,-74.0729676,3a,75y,32.43h,102.34t/data=!3m6!1e1!3m4!1sXSXGtGGfZtANo1vqH3jt-w!2e0!7i13312!8i6656>

**Results and Discussion**

During the process a closterization algorithm was implemented finding 5 groups with different trends in terms of place density and types of trade, the number of groups is debatable since after segmentation it was necessary to join clusters 2 and 3 to have A total of 1045 establishments, in these groups were the largest concentration of commercial premises. which indicated a high flow of people, which in theory can improve the demand for our product, we cover a radius of 500 meters to find the largest number of objective businesses, in this case the coffee shops and thus make sure we have the places that possibly are more isolated within the groups, we also leave a maximum of 50 places per point, these parameters can be adjusted to improve the accuracy of our search or in those cases where the population density is not as high as in small cities or towns.

The neighborhoods with fewer commercial establishments had characteristics such as being near parks, cinemas, squares, theaters, schools and shopping centers, of course this scenario was presented as a good option to locate our home, those clusters with this trend were the 0 and 1, having to be united, reconsiders the idea of ​​using fewer groups in the scikit-learn model.

The purpose of this analysis was to find a practical way to reduce the time we spend on city traffic or in the best case avoid it, which is why the distance between the points was filtered, because in Bogotá traffic has a high congestion was chosen because the route was on foot, hence the distance chosen was so low.

**Conclusion**

Although the objective of this analysis was simply to try to reduce or, at best, avoid traffic during the movement between housing and work in one of the cities considered to have a high mobility deficit, several analyzes had to be done that indirectly They were taking us to the goal.

 The problem can be addressed in many other ways such as those who already have an established workplace and only wish to find a better housing option, also those who are in a state of vulnerability such as disabled people, older adults or head-of-home mothers who, due to their situation, are the ones who should undertake the most because societies tend to segregate them solely because of their condition, this application can improve their living conditions by reducing travel time and in turn assuring them the best place to develop their Commercial activity.

Although the information collected thanks to the Foursquare API was sufficient in this case, it can be further enriched by making use of data such as the value of leases in different locations since the different social strata that are distributed over the territory largely determine the cost of living in each sector, finally the interested person will make the decision based on their personal tastes, the characteristics of the neighborhoods, social and economic dynamics, this application will provide a much clearer picture about this issue.