

CAPSTONE: Battle of the Neighborhoods

Report

Best places in Bogotá to have possible businesses

Table of Contents

Introduction	1
Data.....	1
Methodology.....	2
Results.....	2
Discussion.....	4

Introduction

As the main part of the IBM Data Science certification, is needed a Capstone Project that for this season is called The Battle of Neighborhoods.

Bogotá, the capital of Colombia, is a city with more than 10 million habitants, it is the main economic center of the country. It is the Colombian city with the largest number of foreign companies, one of the factors that position it as the largest labor market in its country. Bogotá is the largest business platform in Colombia, with 21% of the companies registered in the country, and additionally, 67% of Colombian high-impact enterprises are in the city. Bogotá one of the five most attractive Latin American capitals to invest

Despite being a very populated capital, it is one of the few in the world that for the number of habitants it has not a massive transport like the subway. Until this year 2020 begins the first works. That is why in the middle term economic growth is planned at all levels.

Therefore, the problem posed is to identify the areas in which there are more businesses, which have been the most popular and which areas are in which there are more opportunities to open a business.

Data

To perform this analysis, the following main sources of data are available:

- Wikipedia, where we will get the locations in which the city is divided. Bogota is divided in 20 "Localidades". For each "Localidad" we will obtain the postal codes of https://es.wikipedia.org/wiki/Anex:Bogot%C3%A1_Localities.
- Then we will get the geographic coordinates using geopy.
- With the previous data, we started the search of the different venues, the top venues categories in the different areas using the Foursquare API.
- ML tools will also be used as clustering to support the conclusions.
- Finally, we have the tools and data to start with the analysis.

I think this project may interest tourism agencies, tourists themselves and business companies, although the focus is that it serves potential investors as an alternative for expanding their businesses.

Additionally, it serves as an example for data science analyzes to focus on these developing countries.

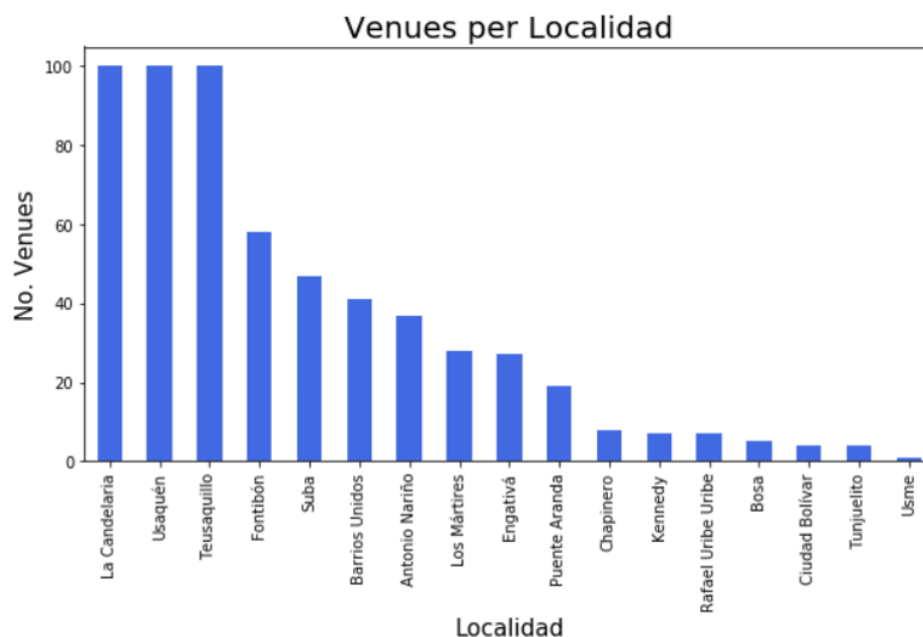
Methodology

The methodology followed to analyze the data was:

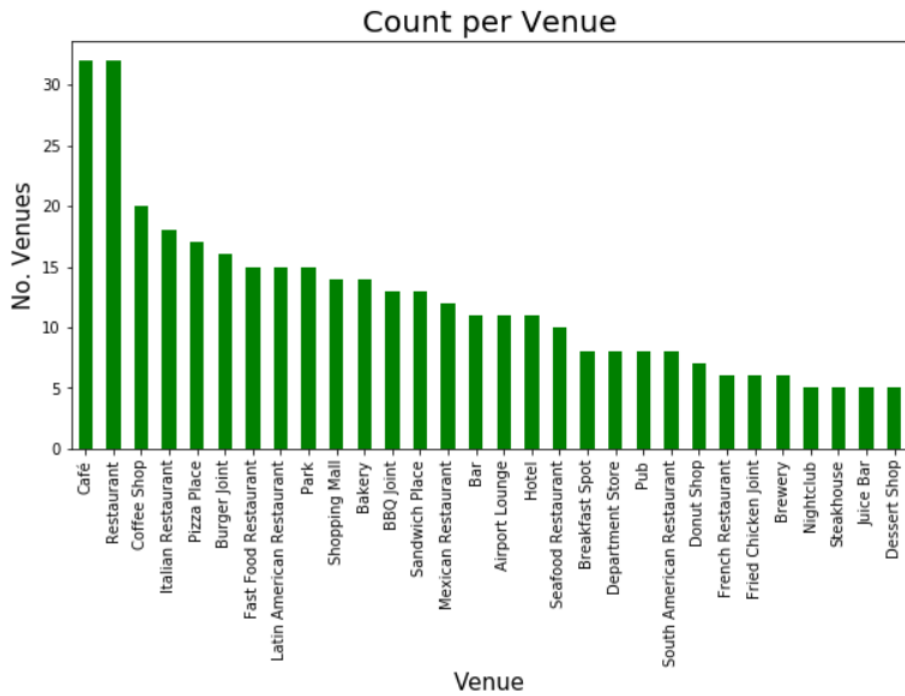
- First, we obtain the data of Bogotá Neighborhoods, called “Localidades”. To create a DataFrame with these data it was necessary some data wrangling and cleaning.
- Next, we obtain the geographical coordinates for each Localidad to add to the DataFrame.
- After that, we create a map using Folium to visualize the “Localidades” in Bogotá.
- Next, we defined the Foursquare credentials to use this tool in order to identify the venues in each Localidad.
- We cluster all neighborhoods and venues correlated. To do the clustering I decided to select Kmeans ML method.
- Data was sorted based.
- Finally, analyzing the data using histograms and tables to have a good visualization and proper understanding.

Results

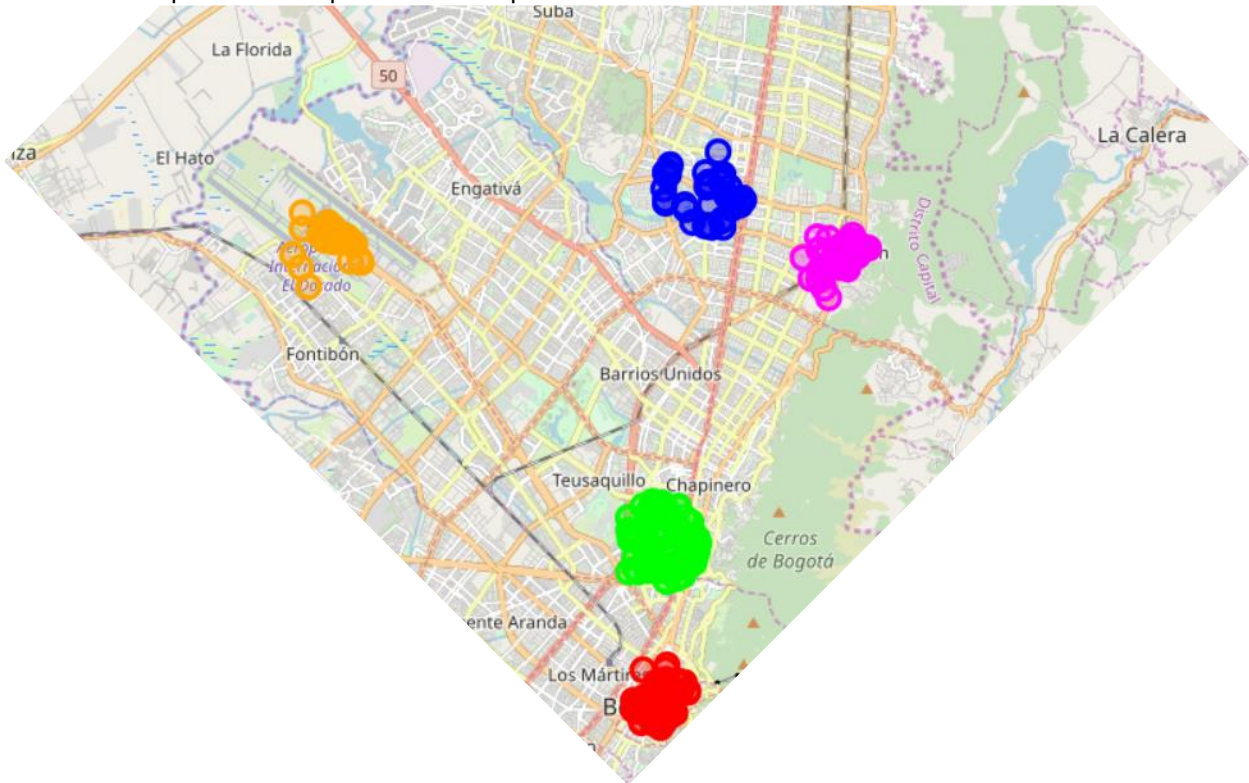
- We selected the top 5 of “Localidades” with highest venues.



- It was found 159 unique categories in the returned venues.
- The most common venues are coffes and Restaurants.



- Parks are in the top five of venues.
- Places where there are highest venues are of touristic interest as Fontibón and La Candelaria, or office places as Usaquén and Teusaquillo or residence sites as Suba.



Discussion

For Foursquare, It was necessary to use a LIMIT of 500 venues and a radius of 1.000 meters because with lower values there is not enough data to do the analysis.

There are many other places, but they aren't registered in Foursquare, so it would be important to look for some way to encourage the use of Foursquare and thus be able to make the use of this data more reliable

Conclusion

- The Localidades with more venues are:
 - 'La Candelaria', 'Teusaquillo', 'Usaquén', 'Suba', 'Fontibón'
- It is interesting to note that these Localidades are not the most densely populated but in those places are main for offices and for tourist interest.
- Although there are many Venues, these sites still have the potential to open new sites such as gyms and multicultural restaurants
- We can recommend analyzing "La Candelaria" because is a in the town and it is a tourist site per excellence in which there is great potential for economic growth, and the future "Metro" will pass through this area.

In general terms. developing all the specialization was something very interesting and the most exciting thing was to be able to work in this Capstone because it is possible to bring everything studied to reality and it materializes in concrete results.