# Business Expansion Analysis

IBM Applied Data Science Capstone Project

Luis Jacobo
04/27/2020

# 1. Introduction

## 1.1 Background

In the last years, the city of Santo Domingo has had a great growth of fast food businesses such as food trucks. Many customers are replacing visits to fast food restaurants with visits to foodtrucks.

The main reasons that customers justify the use of food trucks is the availability of "interesting" foods and convenience.

## 1.2 Description of the problem

The objective of this project is to analyze and select the best locations in the city of Santo Domingo, Distrito Nacional to open a Food Truck. Using data science methodology and machine learning techniques like clustering, this capstone project aims to provide solutions to new investors looking to the best location to open a Food Truck.

## 1.3 Target audience

A food truck owner has contacted us to help him find the best location for a second food truck.

We will study the success of the different food trucks in the city, using the information obtained with the Foursquare API, together with the information of the amount of population in each neighborhood and thus decide the best location for the second food truck.

# 2. Data Collection and Cleaning

## 2.1 Data sources

A list of the neighborhoods of Santo Domingo, Distrito Nacional with the population of each one.

Latitude and longitude of each neighborhood. Required in order to plot the map and also to get the venue data.

The Foursquare API will give us the results for the search "Food truck" in each neighborhood of Santo Domingo, Distrito Nacional within a 500m radius.

## 2.2 Extraction of the data

This wikipedia page contains a list of the neighborhoods in Santo Domingo, Distrito Nacional. We will use web scraping techniques to extract the data from Wikipedia.

https://en.wikipedia.org/wiki/Distrito_Nacional#Sectores_(neighborhoods).

The data downloaded from Wikipedia only had the name of the neighborhood and his population, in this case using Python Geocoder package we will get the geographical coordinates of the neighborhoods.

The Foursquare API will give us the results for the search "Food truck" in each neighborhood of Santo Domingo, Distrito Nacional within a 500m radius.

Will create a new dataset as result of merging the Geolocation data with the neighborhoods. Finally we will save the data frames for further visualization and analysis.

# 3. Methodology

After creating a new dataset with the neighborhoods of the city of Santo Domingo, Distrito Nacional with the coordinates of each neighborhood I plotted the neighborhoods on a map using Folium to check if the coordinates were ok.
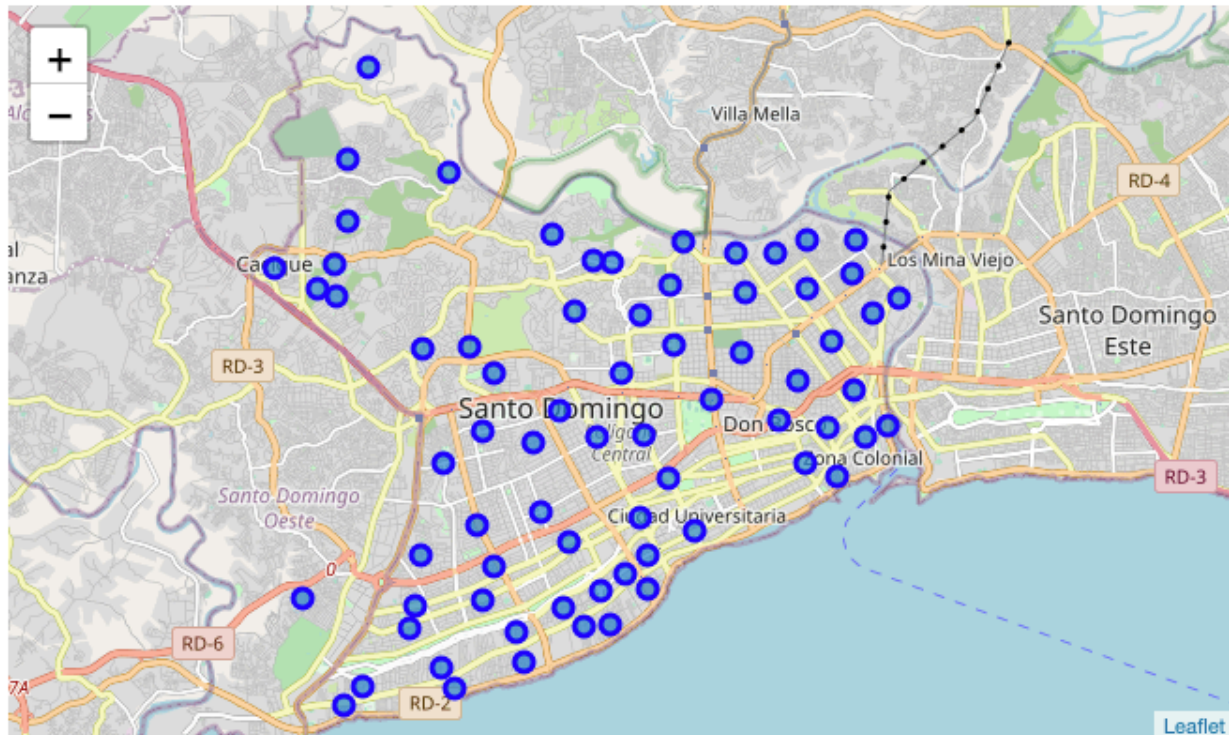


## 3.1 Venue categories for each neighborhood

After cleaning all the data and preparing the final datasets, I created a "One Hot Encoding" for each category for each neighborhood based on the categories Foursquare returned and Neighborhood column added. I added this column in order to be able to group by the neighborhood with the category mean.

Then I created a new dataframe *neighborhood_venues_sorted* with the most common venues for each neighborhood.
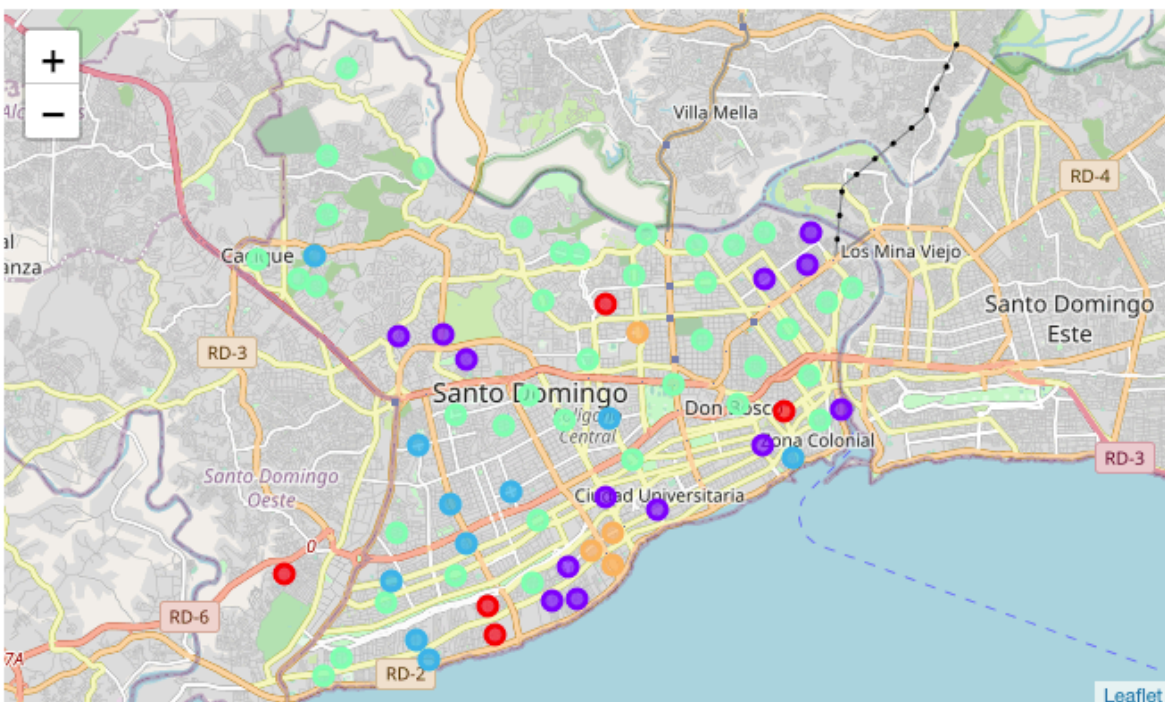
## 3.2 Cluster Analysis

I used K-Means Cluster Analysis with K=5 and Hierarchical Cluster Analysis and got similar results.

I created a Dendrogram with complete linkage method:
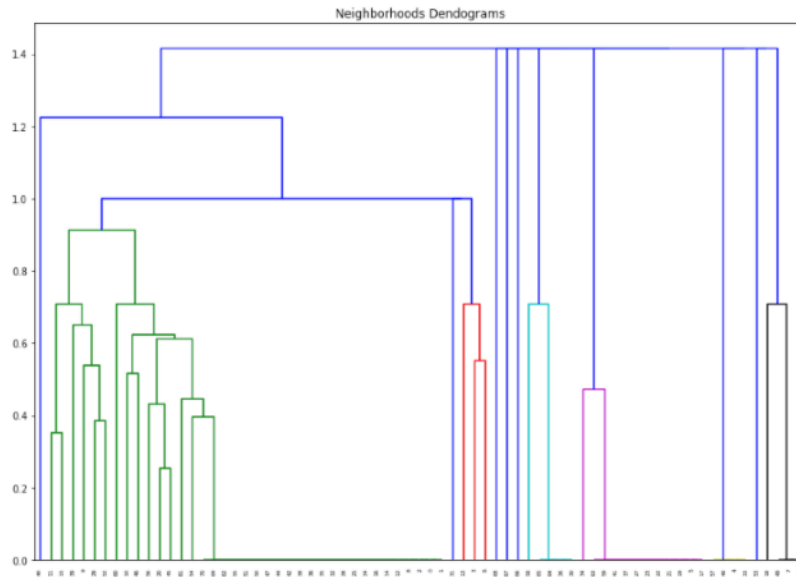


And plotted a Folium map with colors of each cluster:

## 3.3 Comparing clusters

For K-Means Analysis I got the following results:

| Cluster | # of Businesses |
|---------|-----------------|
| Cluster 0 | 5 |
| Cluster 1 | 13 |
| Cluster 2 | 10 |
| Cluster 3 | 39 |
| Cluster 4 | 4 |

The results show that we can categorize the neighborhoods into 5 clusters based on the frequency of occurrence for "Food Truck".

# 4. Discussion

In this project I analyzed every neighborhood of Santo Domingo, Distrito Nacional based on the venues I got from Foursquare API.These neighborhoods clusters can be very helpful for future investors looking to the perfect location to open a new business.

As observations noted from the map and the K-Means Report, most of the fast food / food trucks businesses are concentrated in the Cluster 3, which is the north area of Santo Domingo, Distrito Nacional and a moderated number in Cluster 0 and Cluster 4. As much as Cluster 0 and Cluster 4 represents a great opportunity to open a new location as there is no intense competition due to high concentration of other fast food businesses.

# 5. Conclusion

Therefore, this project recommends the investor to capitalize on these findings to open a new location in Cluster 0 or Cluster 4.

Lastly, the investor is advised to avoid Cluster 3 which already have high concentrations of business alike and intense competition.

# References

Neighborhoods in Santo Domingo, Distrito Nacional. Retrieved from:
 *https://en.wikipedia.org/wiki/Distrito_Nacional#Sectores_(neighborhoods)*

Foursquare Developers Documentation.
*https://developer.foursquare.com/docs/*

Google Cloud Platform for Maps API.
*https://cloud.google.com/docs/?hl=es_419*

GeoPy.
*https://geopy.readthedocs.io/en/stable/*