COURSERA IBM DATA SCIENCE CAPSTONE PROJECT

TITLE: OPENING A SUPERMARKET IN THE BRONX, NEW YORK

By: Rafael Perez Roman

March 2020

INTRODUCTION

- Starting a supermarket business is a highly profitable venture. The day to day lives of everyone involves the use of grocery items and other related goods or services. The place where convenience is provided for customers who seek such things daily is the supermarket. This demonstrates to you that starting your own supermarket business is a worthy undertaking because there's an inexhaustible, ever-growing and ever-evolving market. It goes without saying that a poorly located supermarket is the death of the business. Some of the common broad categories of places where supermarkets can be setup are malls, shopping centers, residential areas and standalone places.

BUSINESS PROBLEM

- A client wants to open a supermarket in Bronx, New York. They want to explore the neighborhoods of the Bronx and selected a place to open the business. The client has other business in other states and is the first time in the supermarket industry.
- The objective of this capstone project is to analyze and select the best locations in the New York area specifically in Bronx, to open a supermarket. Through the data science methodology techniques, we want to find the solution the best place to open a new supermarket.

WHO WOULD BE INTERESTED IN THIS PROJECT?

- This project is useful for business developers on the New York area, specifically in the Bronx. The food industry in the area can determine in which area of the Bronx needs more presence of supermarkets. Which means new area of opportunities for new investors or for the investors with presence in the area and want to expand.

DATA

- Description of the data
  1. List of neighborhoods in Bronx, New York.
  2. Neighborhoods Latitude and Longitude.
  3. Venue data of the neighborhoods (Supermarkets).

SOURCE OF THE DATA

- The source of the data for the capstone project was retrieve form the segmenting and clustering neighborhoods in New York city lab. The data is found in the following url: https://geo.nyu.edu/catalog/nyu_2451_34572 (https://geo.nyu.edu/catalog/nyu_2451_34572) or through this other url : https://cocl.us/new_york_dataset. The last url data can be obtained through the wget command. The data contains the list of neighborhoods in the New York area and we extract the Brooklyn neighborhoods data from there.
- We will get the geographical coordinates of the neighborhoods using Python geocoder. Next we will use Foursquare API to get the venue data for the Bronx neighborhoods. The venue data will be retrieving for all the venues in the Bronx area but for this work we will only use the supermarkets venue data. In this project we will work with API, data cleaning, data wrangling, machine learning, and map visualization.

METHODOLOGY

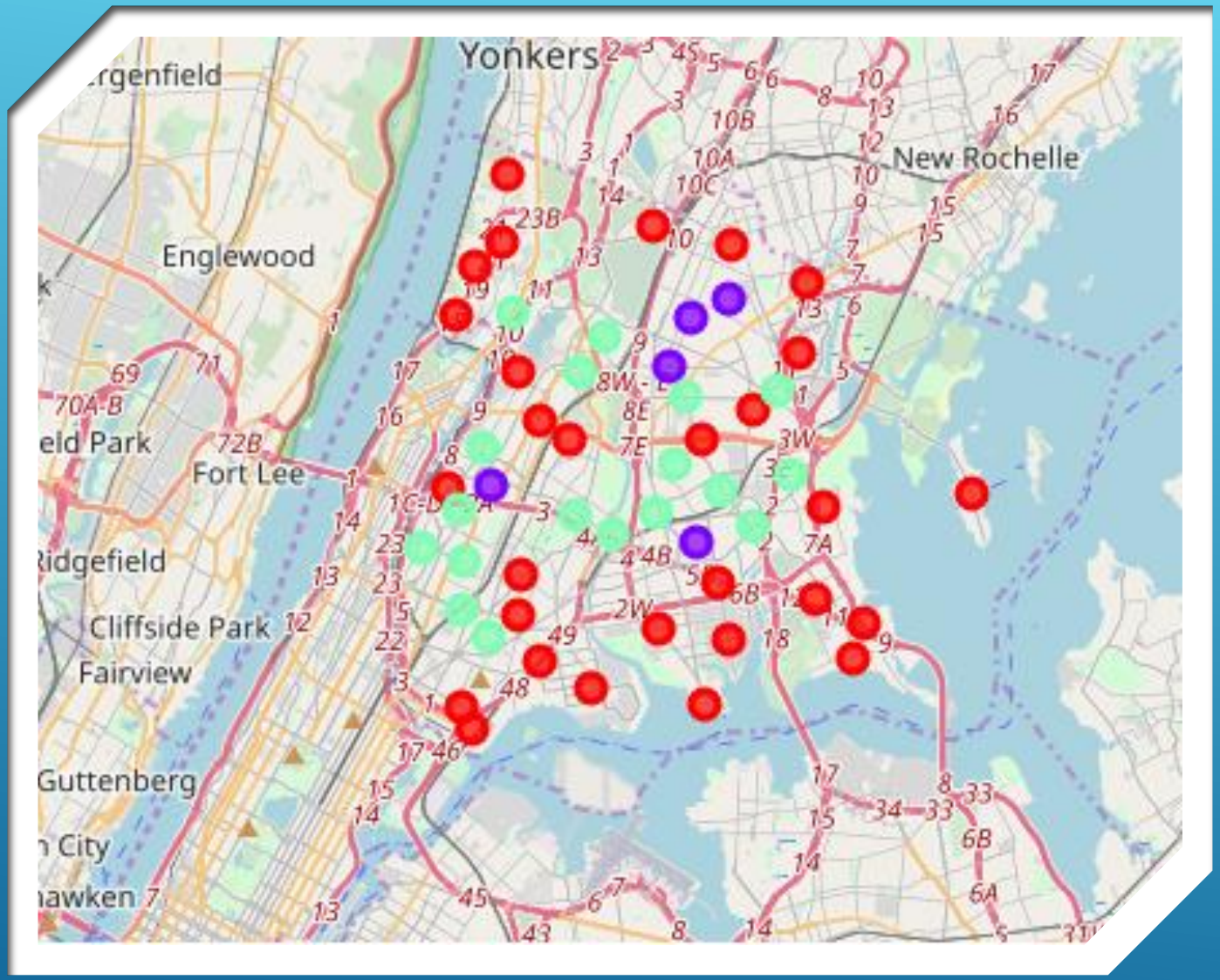
- We get the neighborhood data from the following url: https://cocl.us/new_york_dataset.
- Used wget command to obtain all the Burroughs and neighborhoods of New York.
- From the data we obtain the neighborhoods for the Bronx area.
- Geocoder to obtain latitude and longitude for the Bronx.
- Create a DataFrame for Bronx neighborhoods.
- Create a map using folium package with the neighborhoods.
- Foursquare Client ID, Client Secret and version.
- Used Foursquare to get the top 100 venues within a radius of 500 meters.
- Extract venue name, venue category, venue latitude and longitude.
- Cluster the data with k-means.
- Cluster the neighborhoods into three clusters based on the frequency of occurrence for “Supermarket”.
- Analyze the data of the clusters.

RESULTS

- The results from the k-means clustering are divided into three clusters, depending of the occurrence for “Supermarket”.

The clusters show the following data:

- Cluster 0 (red) = 29 neighborhoods
- Cluster 1 (purple) = 5 neighborhoods
- Cluster 2 (green) = 18 neighborhoods



RESULTS

Cluster map of the Supermarkets in Bronx, New York

- Cluster 0 (red)
- Cluster 1 (purple)
- Cluster 2 (green)

DISCUSSION

- The result will allow us to identify which neighborhood have higher concentration of supermarkets while which neighborhood have fewer number. Based on the occurrence of supermarkets in the Bronx neighborhoods, we will determine the ideal place to open the business. Most of the supermarkets are concentrated in the Cluster 2.
- Cluster 1 had less occurrence of supermarkets than the Cluster 2, but for our recommendation we prefer to not recommend these two clusters. Between the three clusters the one with less occurrence of supermarkets in their neighborhoods are the components of the Clusters 0.

	Neighborhood
25	Morris Heights
22	Kingsbridge Heights
23	Longwood
27	Morrisania
28	Mott Haven
31	North Riverdale
36	Pelham Gardens
37	Pelham Parkway
38	Port Morris
39	Riverdale
40	Schuylerville
41	Soundview
42	Spuyten Duyvil
43	Throgs Neck
44	Unionport
47	Wakefield
20	Hunts Point
18	Fordham
51	Woodlawn
5	Castle Hill
16	Edgewater Park
7	Claremont Village
8	Clason Point
14	Eastchester
9	Co-op City

- In this capstone project we analyze the occurrence of supermarkets in the Brooklyn area in New York. After analyzing the results obtained, we can recommend to our customer the best place to open a new supermarket are the places in the cluster 0. We recommend opening a new supermarket in one of the following neighborhoods presented in the left table.
- These neighborhoods have the less occurrence of supermarkets in the Brooklyn area

CONCLUSION