Coursera IBM Data Science Capstone Project

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.