(img)

Coursera Capstone Project - Battle of Neighborhoods

IBM Data Science Professional Certificate

TITLE: Finding oportunities for opening a delivery-based supermarket in New York City, considering a post-COVID scenario

(img)

Author: Tiago de Moraes

June 2021

Introduction

This project aims to answer a very concerning question that most countries around the world are questionig themselves: "how will the world be like after the Covid-19 crisis is behind us?". In other words, how will the economy be shaped after the social distancing, the obligatory use of masks and constant fear of going outside is over? Many specialists believe that humanity will eventually move on and resume their pre-Covid routines, such as going in restaurants and enjoying events in closed spaces.

** But what new habits acquired during the pandemic will persist? **

That is exactly the question that many entrepreuners are asking themselves right now. They want to find in first-hand opportunities to invest in in business that caters to new acquired habits that will carry on after the pandemic is over. One of such habits that most people agree that will persist is the delivery of groceries and general goods, specially in big cities.

Two key aspects may explain why that might be the case:

- During 2020, many business had to quickly adapt their supply and delivery chains in order to cater to new demands of a world hit by storm by a new decease that forced people to stay inside their homes for many months. That stablished a new sales and after-sales channels that had to mature in a record time. By 2021, that investment will probably not be lost and will continue to be used by many businesses.
- Costumers in general got used to order many things online during 2020, with online shopping tooking an exponential growth in a matters of weeks. This type of convenience is very appealing and even people that aren't normally very keen to shop online (middle-aged people) are already familiarized with this system. In 2021 this type of convenience might as well persist as a major delivery sales channel for many businesses and will no longer be a diffential, but a mandatory revenue stream for any new business to come.

Considering that scenario, this final project explores the best locations for a delivery-focused market throughout New York City. As with any business, opening a new market requires throughout market study considering many key aspects in particular. The location of the market is one of the most important factors that will affect the outcome of the business. Hence, this project will attempt to find optimals locations to install a new delivery-focused market in New York, considering its neighborhoods profile and local venues locations.

Objective

Using Data Science methodology and instruments such as Data Analysis and Visualization, the main goal of this Capstone project is to analyze and select the best locations in the city of New York for the deployment of a delivery-focused market. This project aims to provide solutions to answer a core business question: * **Where in New York City should the investor consider to install a delivery-based supermarket? * *

Data

To make such analysis, the following data will be querried:

- Geospatial data of the city of New York, containing the neighborhoods and boroughs information;
- Latitude and longitude coordinates of those neighborhoods forplot purposes and to get local venue data.

- Venue data, particularly data related to nearby markets, in order to to perform further analysis of the neighborhoods.

This project will require using of many data science skills, from web scrapping (open source dataset), working with API (Foursquare), data cleaning, data wrangling, to map visualization (Folium), along with extensive exploratory data analysis and statistical testing for model assessment using machine learning techniques.

Methodology

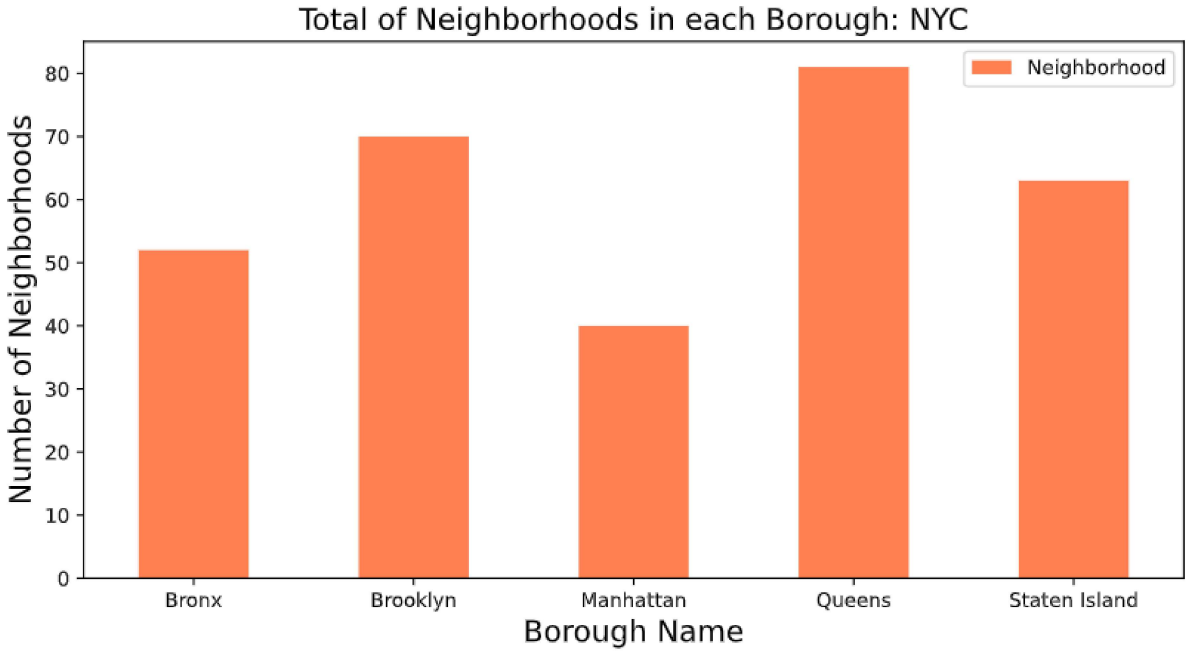
- Data will be collected from https://cocl.us/new_york_dataset (https://cocl.us/new_york_dataset), wrangled and then processed into a Pandas dataframe.
- FourSquare API will be used to locate all venues and then filtered by markets in New York. Ratings, tips, and likes by users will be counted and added to the dataframe, and then data will be sorted based on rankings.
- Data be will be visually assessed using Python libraries.
- Maps will be also generated for a more pleaseant visual analysis
- k-means algorithms will be used for cluesterization of spatial data

Results and Discussion

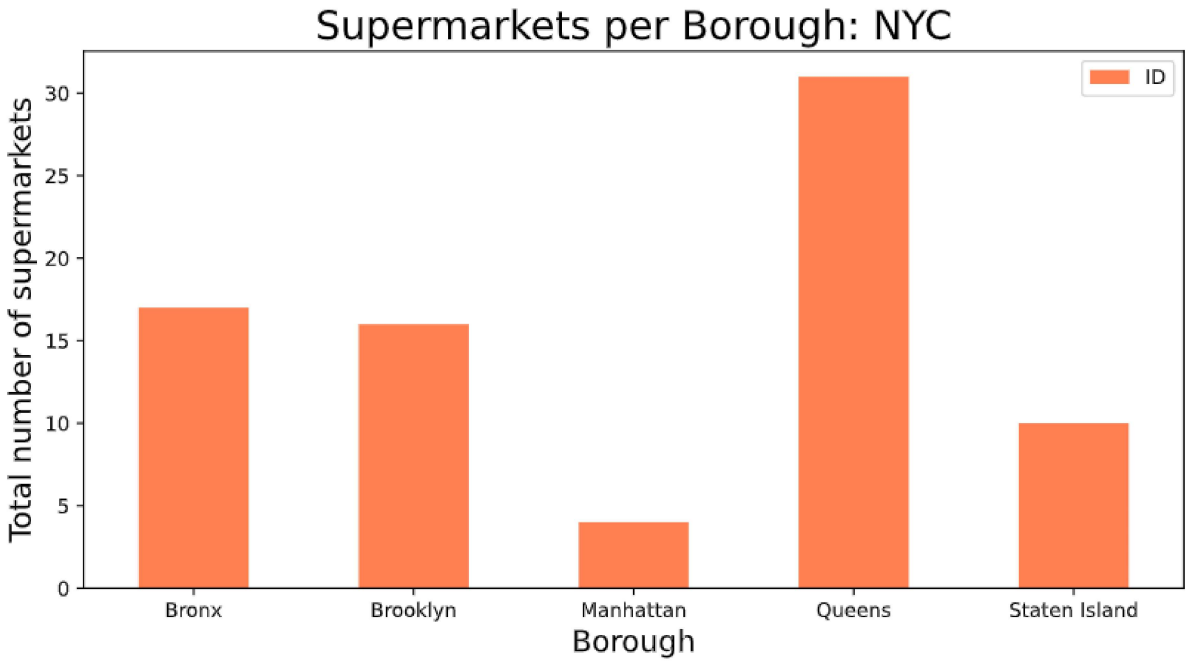
After querrying through the dataset and using the Foursquare API to obtain a valuable amount of spatial dataset, it's possible to drawn some interesting insights that are not clear and trivial at first glance:

Graphical analysis

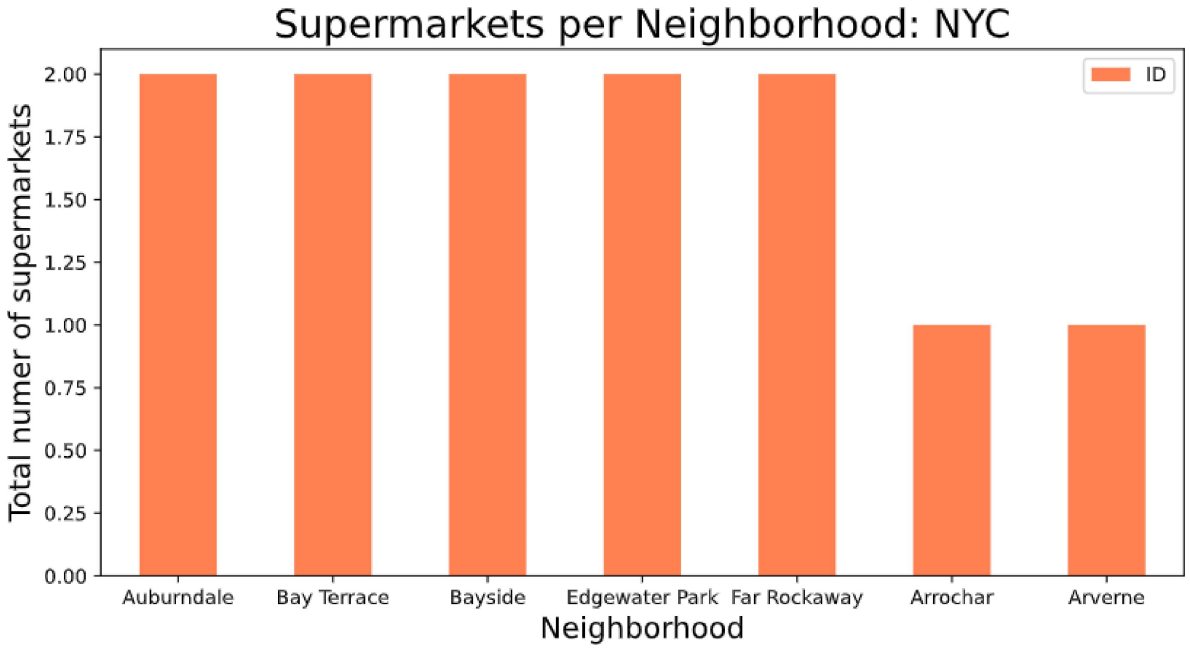
The first insight drawn can be seen from the below graphic, which shows that Queens is the the neighborhood with the most neighborhoods in New York City.



Now, when plotting the number of supermarkets available in each Borough, it becomes clearer that there is a huge discrepancy between the number of supermarkets in Queens and others Boroughs, such as Manhattan, Brooklyn and Bronx



Auburndale, Bay Terrace, Bayside, Egdewaterpark and Far Rockaway are the neighborhoods with the highest number of supermarkets. The most interesting piece from this data shows that supermarkets are far away from each other, showing that even in the Borough with the most number of Supermarkets (Queens) and its neighborhood most populated with markets (Auburndale), you will only find 2 supermarkets nearby.



Main insights from data

Based on the analysis results, it's possible to infere that Brooklyn and Bronx are the best candidates for the installation of the proposed delivery-based market. As those two areas have a strong residential profile and a less abundant supermaket distribuition, in comparisson to Queens, for instance, these two areas presents the lesser risk for investors. The data shows that both Brooklyn and Bronx has multiple neighborhoods with less than one market nearby. Another interesting insight drawn from the data is that even in Queens, which is the Borough with highest number o supermarkets, when looking to its neighborhood that has the most markets in it, we see that this number isn't bigger than 2, which shows how descentralized the supermarket distribuition is in New York. This corroborates the the idea to invest in a more spread type of markets, focused more on delivery model, where costumers make orders through apps and websites.

Suggestions for future researches

For spatial data querries, the Foursquare API was used to get the calls done throughout the project. Despite being an excellent souce for this type of data, it's worth note that it has its share of shortcomings. The first is, as an third part data administrator company, the quality of the data depends entirely on how well the company curates the data, leaving the quality and accuracy of the data entirely on Fousquare's hand. Another point to note is that this project qas carried out using a free Sandbox Tier Developer Account, which means that Foursquare limits the number of API calls possible in 24 hours. When querring a huge amount of data, such as dataset from entire states or contries, the limitation imposed by this type of account can severly hinders the pace of the project, making it necessary the purchase of professional accounts.
For deeper results, it's recommended for future research work that more comprehensive analysis could be carried out using a paid account to bypass the API calls limitations as well as incorporating data from other external databases and adding other variables to be queried, such as the population in each area or commute data.

Conclusions

Considering the post-Covid scenario, this final project explored the best locations for a delivery-focused market throughout New York City. As with any business, opening a new market requires throughout market study consideration, looking through many key aspects of several areas. For that matter, using applied data science methods and routines, this work attempted to tackle this challenge and provide an precise answer for investors interested in finding optimals locations to install a new delivery-focused market in New York, considering its neighborhoods profile and local venues locations. The data showed what was already suspected, that central areas, such as Manhattan, which have a less residential profile, don't have as many supermarkets available as other areas, such as Queend and Brooklyn. The main takeaway from this project is that there is still much more room for many residential areas, in special Queens and Brooklyn, to expand to a more descentralized market model, much more focused in delivery its products than in big store supermarkets, as those areas are highly occupied and have less than 1 supermarket per neighborhood.

In [ ]: