

Opening a Clothing Store in New York City

By Scott Paz

July 14, 2020

1. Introduction

1.1 Background

The total value of the apparel market around the world is worth about 1.78 trillion dollars, where in the United States' stores are worth 368 billion dollars. The best of the best clothing stores bring in almost billions of dollars in sales a year which can amount to millions in revenue. Gucci alone brought in almost 270 million euros in revenue during the 2018 fiscal year. Of course, there are stores that do not succeed as their competitors such as Forever 21, where they filed bankruptcy in 2019. But what can help clothing stores succeed and stay successful? Price? Clothing Lines? Sponsorships? Or how about mere location? If a brand-new clothing company was looking to open up a store, where would they do it?

1.2 Business Problem

The client has a successful clothing business, where she does all her sales online. She wants to expand, make more money, and get her product out as much as possible, but she does not know where to start. She knows she wants to do it in New York City, because that is where her company is based but she does not want to settle anywhere. She wants to have her first store to be open in a populated area surrounded by many other stores so people have more of a tendency to stop by her store. Assuming money is not a problem since she is successful, she wants to know where she should open up shop.

1.3 Interest

This should interest clothing companies that want to open up a physical store in New York City and want it to be in the most populated and busiest shopping areas.

2. Data

2.1 Data Sources

To answer this question to where the best location for a clothing store should open, the client will need 2 specific data sources. The first being crime rates data and the second will be regarding the Foursquare API, to determine popular neighborhoods.

1. New York Population

https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City

This data set will provide population per neighborhood in New York City

2. Foursquare API

This data set will provide a list of all the stores are located in New York City, this will be an indication of popular foot traffic in each city/neighborhood in New York City

3. New York City Data

https://cocl.us/new_york_dataset

This specific data set pulls the latitude and longitude for each neighborhood in each borough

3. Methodology

To analyze the situation and to understand what the client wants, several tasks need to be performed. First, what was needed and secondly, how was it going to be achieved. To first started this project in python, an import of a ton of libraries were needed, this was so they could be referenced throughout the project.

```
import numpy as np # library to handle data in a vectorized manner
import requests # library to handle requests
import pandas as pd # library for data analysis
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

import json # library to handle JSON files

!conda install -c conda-forge geopy --yes # uncomment this line if you haven't completed the Foursquare API lab
from geopy.geocoders import Nominatim # convert an address into latitude and longitude values

# libraries for displaying images
from IPython.display import Image
from IPython.core.display import HTML

import requests # library to handle requests
from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors

# import k-means from clustering stage
from sklearn.cluster import KMeans

!conda install -c conda-forge folium=0.5.0 --yes # uncomment this line if you haven't completed the Foursquare API lab
import folium # map rendering library

print('Libraries imported.')
```

Once the libraries were imported, the data set of New York City was read. This dataset from “https://cocl.us/new_york_dataset” brought in all of New York Cities’ boroughs and neighborhoods along with their coordinates.

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

This dataframe was needed to eventually combine the dataframes at the end. After this was populated, the population of New York Cities' neighborhoods needed to be brought in. During this instance, BeautifulSoup was used to read off the Wikipedia's data set on New York. Below is a snippet of the code.

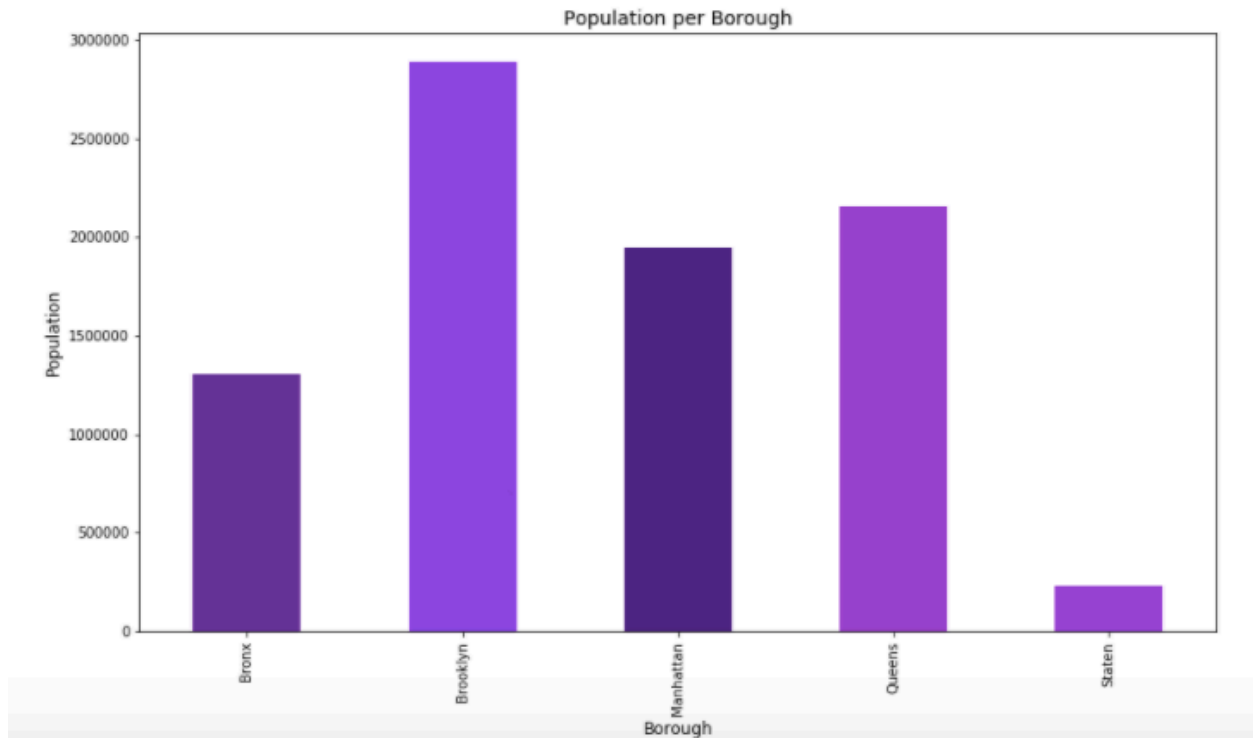
```
from bs4 import BeautifulSoup
#import BeautifulSoup to read wikipedia page

def get_population_per_neighborhood(read_from_csv=False):
    #BeautifulSoup function, to go on wikipedia and obtain population per borough/
    if not read_from_csv:
        WIKI_LINK = "https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City"
        ROOT_WIKI_LINK = "https://en.wikipedia.org"
        page = requests.get(WIKI_LINK)
        soup = BeautifulSoup(page.text, 'html.parser')
        population_list = []
```

Once the information was set into the notebook, the dataframe was manipulated to bring in the columns, Borough, Neighborhood, and Population. This only shows the first five rows in the Bronx, because there is a plethora of neighborhoods in New York City.

	Borough	Neighborhood	Population
0	Bronx	Melrose	24913
25	Bronx	Bruckner	38557
26	Bronx	Castle Hill	38557
27	Bronx	Clason Point	9136
28	Bronx	Harding Park	9136

Since the client wanted her store to be in one of the more popular areas, the population of each borough was graph. Below is a representation of Borough's population.



As the graph shows, Brooklyn comes out on top, then Queens, then Brooklyn, second to last is the Bronx, then in last it is Staten Island. As it can be seen Brooklyn has the most people in its area, but this only solves one answer. Which moves to the next point. Bringing in Foursquare's API the number of stores are able to be determined in each Borough and in each Neighborhood, for now though, the Borough will be looked at. Below is a snippet of the code to get the information from the Foursquare API.

```
# Get information for Retail Stores
def get_data(lat, lng, borough, neighborhood):

    radius = 1000
    LIMIT = 100
    VERSION = '20180605'
    FS_CLIENT_ID = "X5STHTOJ5ALK42QMTYDXRXWTKQ20FB4OOWKA5IQZ4PE1D5ZI"
    FS_CLIENT_SECRET = "VPEBQYOI35SAJ5EOP5L34DXJTRQIKHNT5ERK051GFSSQKNWU"
    #https://developer.foursquare.com/docs/build-with-foursquare/categories/
    FS_CLOTHING_KEY = "4bf58dd8d48988d103951735"
    url = 'https://api.foursquare.com/v2/venues/search?&client_id={}&client_secret={}&version={}&lat={}&lng={}&radius={}&limit={}&query={}&fs_clothing_key={}'
    url = url.format(FS_CLIENT_ID,
                    FS_CLIENT_SECRET,
                    VERSION,
                    lat,
                    lng,
                    radius,
                    LIMIT,
                    FS_CLOTHING_KEY)
```

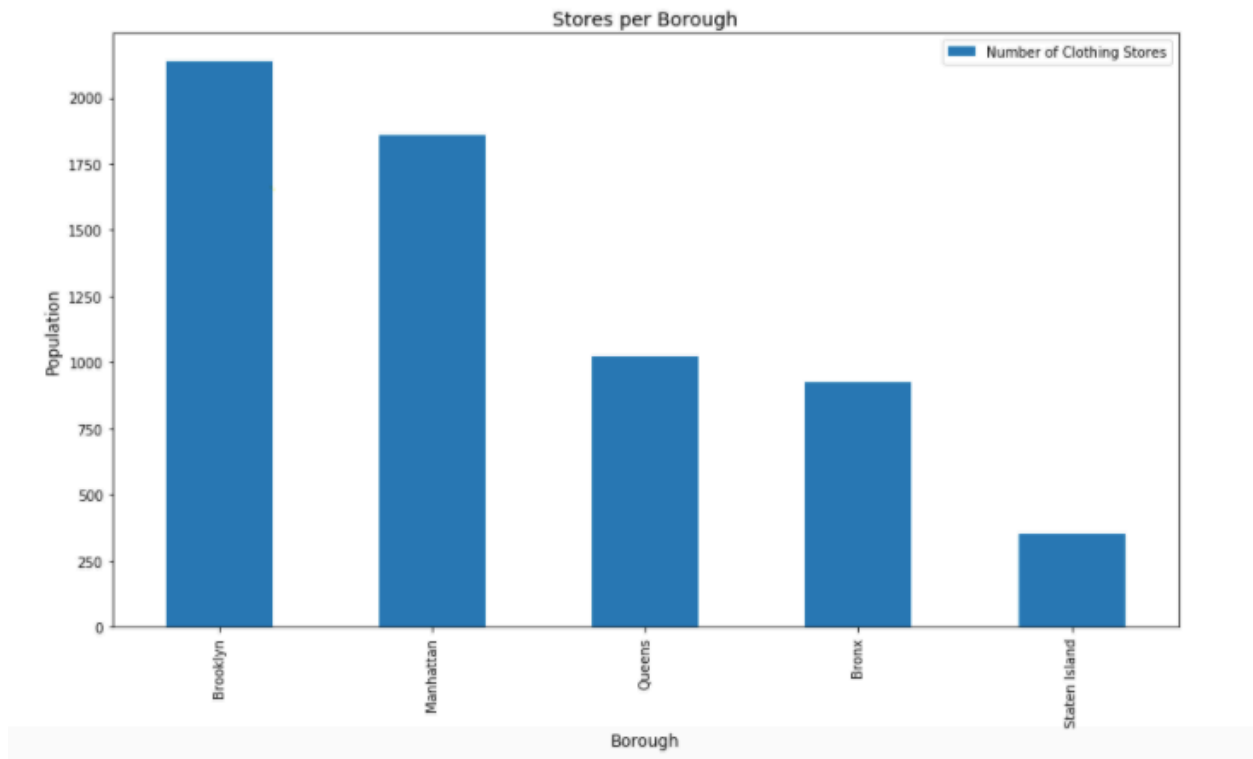
Once this information was manipulated and pass through, the Names of all the stores along with their Latitude, Longitude, Borough and Neighborhood were passed through as well. This can be seen below.

	Name	Latitude	Longitude	Borough	Neighborhood
0	Rainbow Shops	40.898166	-73.855052	Bronx	Wakefield
1	Blinking Genii	40.896970	-73.855026	Bronx	Wakefield
2	Jerseyman Cap USA	40.898116	-73.854388	Bronx	Wakefield
3	B&R Uniforms	40.891888	-73.858560	Bronx	Wakefield
4	Jimmy Jazz	40.873286	-73.824370	Bronx	Co-op City

This Foursquare API dataframe, gave us enough information to understand what Borough has the most stores, again with a bit of manipulation, the Number of Clothing Stores were shown.

Number of Clothing Stores	
Borough	
Brooklyn	2137
Manhattan	1856
Queens	1021
Bronx	922
Staten Island	350

Along with this dataframe, is a bar graph to show how each Borough ranks up against each other.



As can be seen above, the ranking goes as such: Brooklyn, Manhattan, Queens, Bronx then Staten Island.

4. Results

When merging the data together, “Borough, Number of Clothing Stores, Population” it will give us a better look at where the store should be located.

Borough	Number of Clothing Stores	Population
Brooklyn	2137	2889582
Manhattan	1856	1947540
Queens	1021	2157937
Bronx	922	1305720

So, as it can be seen Brooklyn has the most amount of Clothing Stores and a higher Population, but does this mean it may be the most popular? If you look at the second row, at Manhattan there are still a high amount of Clothing Stores and a lower Population than Manhattan. Ultimately though, the client wanted to have her store in the most popular area with the most number of stores, which means her best bet, would be in Brooklyn.

5. Discussion

Some observations that were noticed was first, Staten Island did not have many stores and half the population of Bronx, so it was omitted in the final results table. Also, it can be seen that Brooklyn would be the best fit for the client's request. This is because it fills both her requests of the most popular retail stores and popularity. But what if this was able to be broken down more? Specifically, what if the crime was taken into account? If crime was taken into account though it may get a bit trickier because crime rate in New York City does not separate their crime by neighborhood, it separates it by Borough and Precinct. Which means there can be multiple neighborhoods in one precinct and that could get a bit ugly when trying to merge all the data. Possibly taking into account rental space for each borough/neighborhood could be helpful as well for client, but it is assumed she has a lot of money, but every penny counts and could be saved for her.

6. Conclusions

During this study, I was able to analyze the population of each borough and how many stores are located in each borough. Using Foursquare API, Beautiful Soup (Wiki), and NY Data, it shows that there is a lot of data everywhere, and even one dataset can tell a story. The graphs and information that were produced may be able to help people determine where they want

their next retail store to be or store in general, whether it is somewhere popular with a ton of stores such as Brooklyn, or a popular place with not that many stores such as the Queens. It has enough information for a vague question.