Comparing Neighborhoods in the Bronx and Brooklyn

Yunjung Joe

August 3, 2019

Part 1. Introduction

1.1 Description of the Problem

Brooklyn is on the rise while The Bronx seems unfortunately in decline. I got a phone call yesterday. One of my colleagues has to move into the Bronx who has been lived in Brooklyn for a long time. For my colleague who needs to choose new neighborhood in the new city, I make up my mind to give a real insight between the two cities: Brooklyn VS the Bronx. Also, Brooklyn and The Bronx have their own unique characteristic. I believe that exploring and comparing these two cities would be interesting. By discovering how similar or dissimilar they are, I will give a real insight for my colleague.

1.2 Background

Brooklyn and The Bronx are very different. They don't mix like oil and water. One reason is that Staten Island aside, which is totally separate and apart, The Bronx and Brooklyn are the two most geographically distant boroughs. Just as a matter of time and distance, residents of the two boroughs rarely mix.

The basic difference is that Brooklyn and The Bronx developed differently. Brooklyn has its own separate and independent identity as a city. For much of its existence is as one of the largest in the country. Even as a borough, it would qualify as the 3rd or 4th largest city in the USA.

The city of Brooklyn developed in parallel to the City of New York, rivaling NYC until they merged, evidenced by Brooklyn's downtown and Brownstone Belt. Also, the outlying areas of Brooklyn (before Brooklyn was consolidated) were well developed and prosperous, such as Flatbush, Bushwich, and the other Dutch towns.

The Bronx was an outlying area of mostly farmland and remained this way for most of its existence, until the subway was extended to the South Bronx area. After which, the Bronx, unlike Brooklyn, developed with dense neighborhoods of mostly apartment buildings in which the masses of the lower east side expunged upon.

1.3 Target Audience

For this, my target audience is my colleague. However, these kinds of researches will provide insights for the clients who need to decide a correct location. For example, as a data scientist for my client, I can recommend to the marketing team which neighborhood will be the best choice for maximizing their marketing activity performance.

Part 2. Data

2.1 Data sources

To consider the problem we can list the data as below:

- Data 1: **new_york_dataset**
  I downloaded a json file. This json contains all the names, the borough, and latitude and longitude of every neighborhood in New York City. I extracted the json file from the internet.

- Data 2: **Foursquare location API** I got the top 100 venues within a 750 meter radius of each neighborhood by using the Foursquare location API.

2.2 Data Preparing

Now that my data is in a Pandas dataframe, it'll be much easier to manipulate and examine. Since I am only interested in neighborhoods in The Bronx and Brooklyn, I simply dropped all the rows that contained neighborhoods in Manhattan, Queens and Staten Island. Yes, Staten Island is still technically a borough of NYC. I know it's easy to forget about. Next I needed to get the top 100 venues within a 750 meter radius of each neighborhood using the Foursquare location API.
Use the geopy library to get the latitude and longitude values of New York City.

Part 3. Methodology

3.1 Exploratory Data Analysis

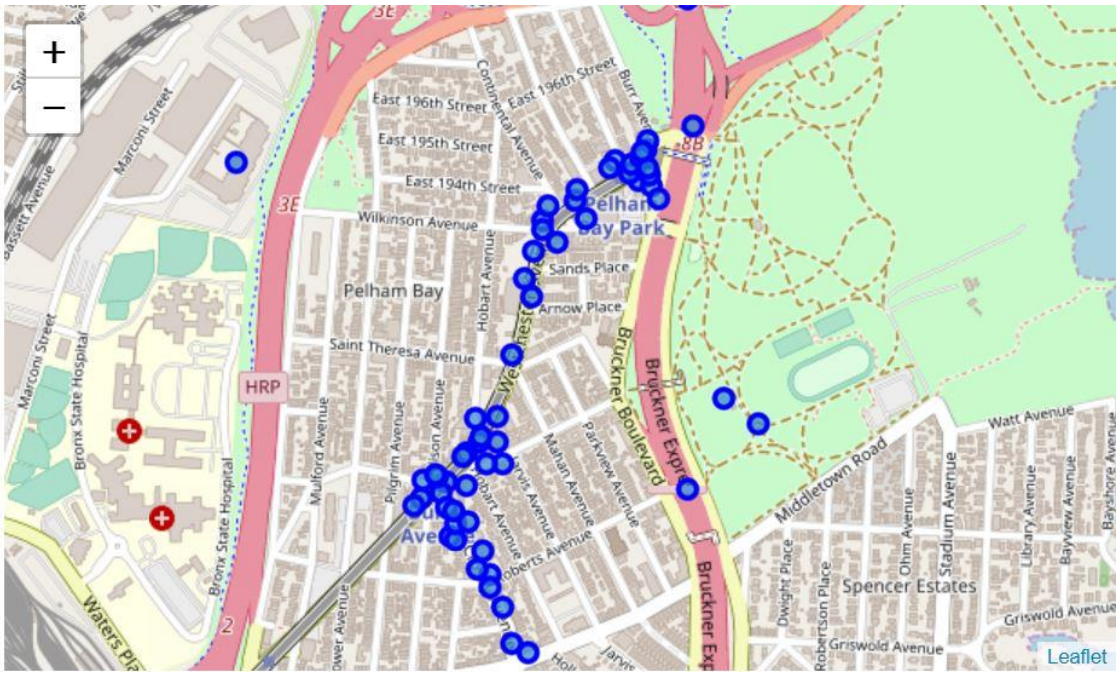
1. Above dataframe which I created is for **Exploratory Data Analysis**.
2. This dataframe contains the geographical coordinates of the two Boroughs.
3. This data will be used to get Venues data by leveraging **Fouresquare**.
4. I want to map out the neighborhoods.
5. Eventually, I will group them into 5 different clusters by using **k-means Clustering**.

3.2 Create a Map of New York City with the neighborhoods plotted onto it.

Firstly I just wanted to map out the different neighborhoods and make sure my map looked good. So the first thing I did was plot out the neighborhoods of the Bronx and Brooklyn using the Python **Folium** library to visualize geographic details of two boroughs.

3.3 Neighborhood, Pelham Bay

Everything looks great so far. Now let's look around my colleague's neighborhood, Pelham Bay. The map is visualized below:



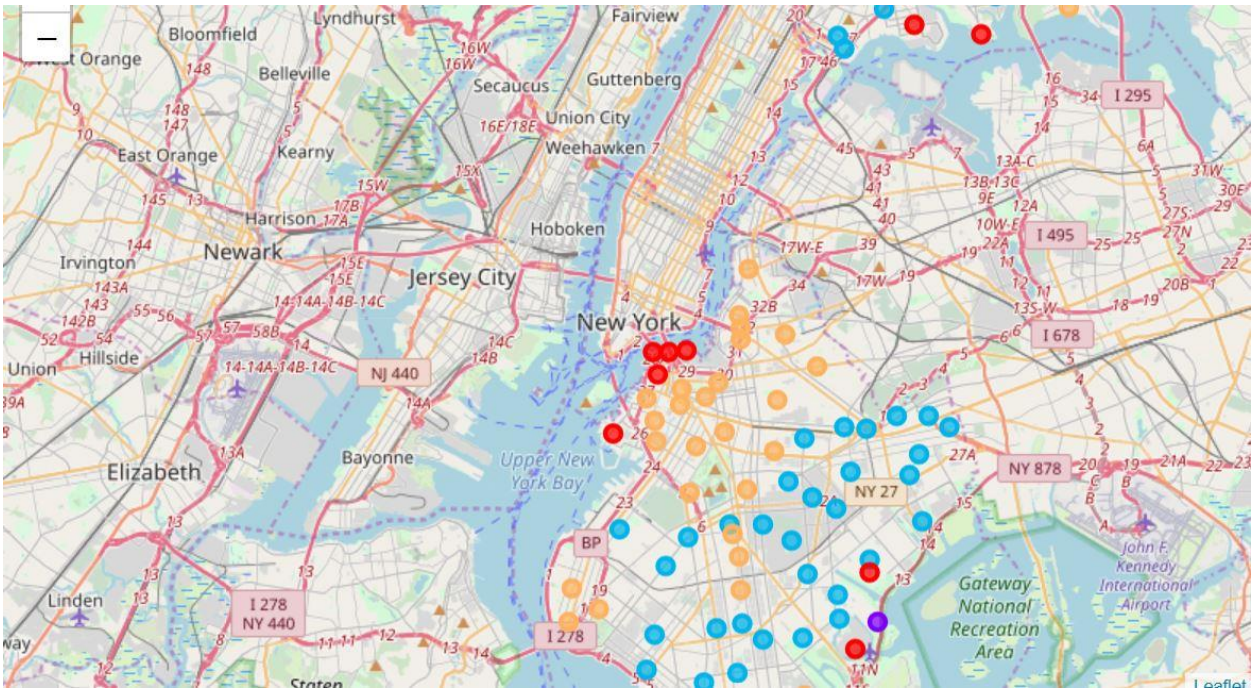
3.4 Print out each neighborhood with the top 5 most common venues.

I was ready to group all the categories of the different venues in each neighborhood in The Bronx and Brooklyn. There were 368 unique venue categories ranging everywhere from Adult Boutiques to Speakeasies to Zoos and of course Pizza Parlors. I then had to encode the data and get dummy variables of all the different categories. Once I did that I was able to break down the frequency of the different categories of each neighborhood. The 5 most frequent venues for my neighborhood are listed below:

3.5 Cluster Neighborhoods.

Run *k*-means to cluster the neighborhoods into 5 clusters.*

Once that was done I could run a k-means clustering algorithm on the data and assign a cluster to every neighborhood that was similar to one another. From there I was able to look at my results and compare the neighborhoods.



Part 5. Conclusion

- My colleague's neighborhood, Pelham Bay was located in the largest cluster (**Cluster 3**) of all of them. Apparently this neighborhood isn't too unique.
- This neighborhood is much more like most other Bronx neighborhoods and southern and eastern Brooklyn neighborhoods.

- I recommended to move to the north eastern part of Brooklyn (the green cluster). Apparently it's not as similar to current neighborhood. However, the gentrification of those Brooklyn neighborhoods and hipster influence on this decision.

Some further examination to be done:

- Add more clusters.
- Look at features such as crime rate, demographics, and median age of each neighborhood.
- Use the top 20 or 30 venues next time instead of just the top 10.