Applied Data Science Capstone Project Report

The Battle of Neighborhoods - Bronx New York

Name: Tharani Govindarajulu

Table of Contents

1. Introduction:

The Aim of this Project is to find better facilities around neighborhood, It will helppeople to make smart decision in selecting great neighborhood in Bronx ,New York. New York City's demographics show that it is a large and ethnically diverse metropolis. It is the largest city in the United States with a long history of international immigration with attractive job offers.

Lots of people are migrating to Bronx, NY and needed lots of research for good housing prices and reputed schools for their children. This project will help people who are looking for better neighborhoods. For ease of accessing to Cafe, School, Super market, medical shops, grocery shops, mall, theater, hospital, like minded people, etc.

The work focus on create an analysis for people moving to Bronx, NY to search a best neighborhood as a comparative analysis between neighborhoods. The features include median housing price and better school according to ratings, crime rates of that particular area, road connectivity, weather conditions, good management for emergency, water resources 3both fresh and waste water.

2. Data Preparation

New York City data that contains list Boroughs, Neighbourhoods along with their latitude and longitude.

Data source : https://cocl.us/new_york_dataset

This data set contains the required information and we will use this data set to explore various neighborhoods of New York City.

For city Geo Space,

Data source: https://data.cityofnewyork.us/City-Government/BoroughBoundaries/tqmj-j8zm

By using this geospace data we will get the New York Borough boundaries that will help us visualize choropleth map.

3. Foursquare API Data:

The data about different venues in different neighborhoods of that specific borough, in order to gain that information we will use "Foursquare" location information. Foursquare is a location data provider with information about all venues and events within area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform

will be used as the sole data source since all the stated required information can be obtained through the API.

After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 100 meter.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Venue
5. Name of the venue e.g. the name of a store or restaurant
6. Venue Latitude
7. Venue Longitude
8. Venue Category

4. Methodology and Analysis

After cleaning and preparing the data, let us identify the steps, that have to be performed in order to find the most optimal boroughs.

Firstly, we will apply some basic exploratory analysis to our data. For that let's find the location of each borough on the map. Then we can visually inspect some values in our data with the help of bar charts.

Secondly, we have the possibility to reduce the number features in data frame by replacing them with more reasonable data.

Finally, we will perform cluster analysis to find the best cluster of boroughs with meaningful features

4.1 Exploratory Data Analysis

First, it's is quite useful to visualize the center locations of each borough. For that, the map of New York was created with boroughs superimposed on top.

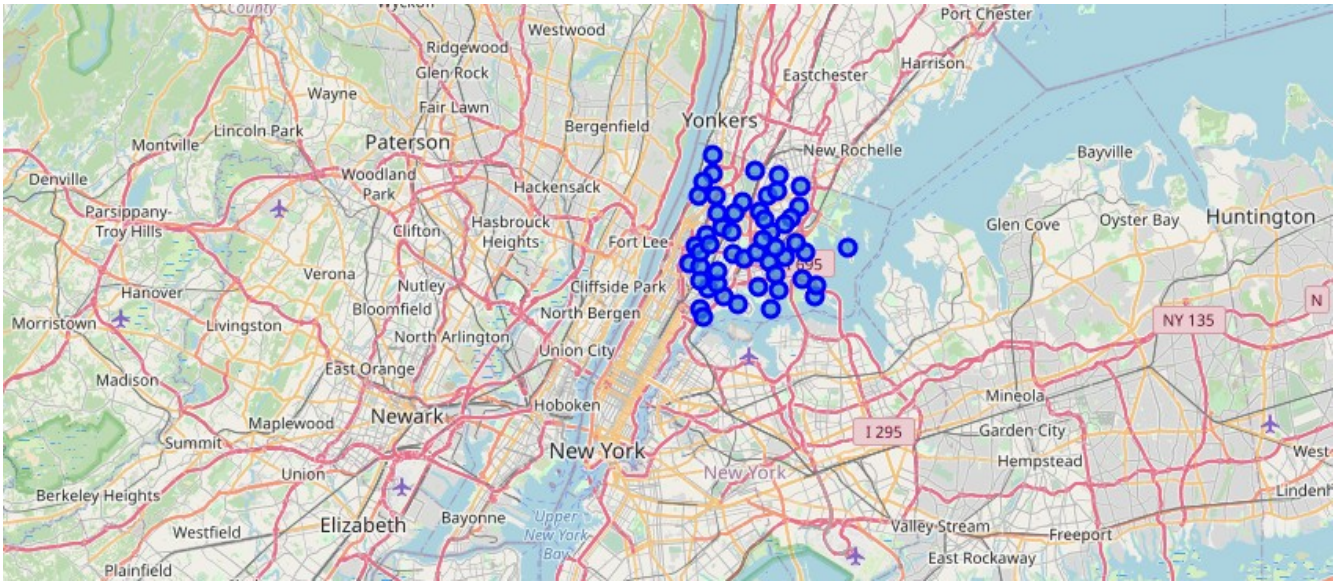

Fig 4.1(a): Map of New York city.

The unique venues in New York city is given below chart.

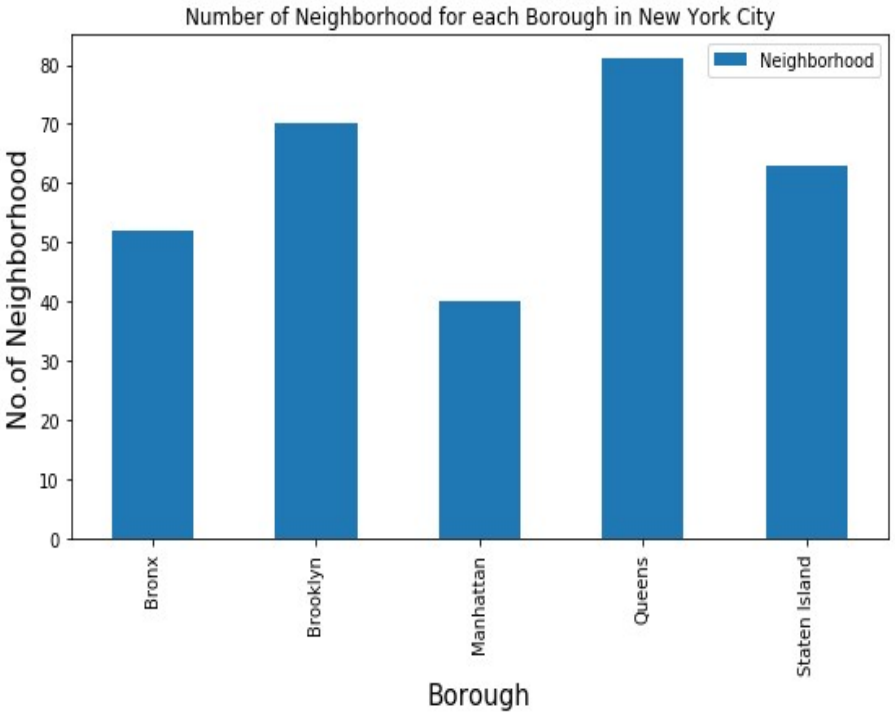

Fig 4.1 (b) – Unique Borough in New York city

Additionally, the several features, that influence our choice, were visualized with respect to the Boroughs.

	name	categories	lat	lng
0	Residence Inn by Marriott New York The Bronx a...	Hotel	40.849917	-73.842152
1	LA Fitness	Gym	40.849739	-73.841949
2	Starbucks	Coffee Shop	40.851371	-73.844087
3	Skyline Bar & Lounge	Lounge	40.852904	-73.842612
4	Starbucks	Coffee Shop	40.847132	-73.844449

Fig 4.1(c) Essential Stores around Bronx

The data frame gives the details about the places that is frequently visited and which is near by housing units

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Allerton	Donut Shop	Pizza Place	Sandwich Place	Bar	Martial Arts School	Fast Food Restaurant	Pharmacy	Supermarket	Chinese Restaurant	Caribbean Restaurant
1	Baychester	Bus Station	Department Store	Donut Shop	Supermarket	Discount Store	Furniture / Home Store	Women's Store	Pet Store	Park	Paper / Office Supplies Store
2	Bedford Park	Pizza Place	Diner	Sandwich Place	Deli / Bodega	Grocery Store	Mexican Restaurant	Pharmacy	Spanish Restaurant	Chinese Restaurant	Donut Shop
3	Belmont	Italian Restaurant	Pizza Place	Deli / Bodega	Bakery	Coffee Shop	Dessert Shop	Plaza	Mexican Restaurant	Café	Spanish Restaurant
4	Bronxdale	Pizza Place	Italian Restaurant	Sandwich Place	Bank	Coffee Shop	Pharmacy	Mobile Phone Shop	Donut Shop	Diner	Ice Cream Shop

Fig 3.1(d) Essential Stores around Bronx

3.2 Cluster Analysis

To identify groups (clusters) with similar characteristics, the unsupervised learning method to our data, namely K-Means algorithm, was applied to data. To compare the similarities of two cities, exploring neighborhoods, segment them, and group them into clusters to find similar neighborhoods in a big city like New York and Toronto.

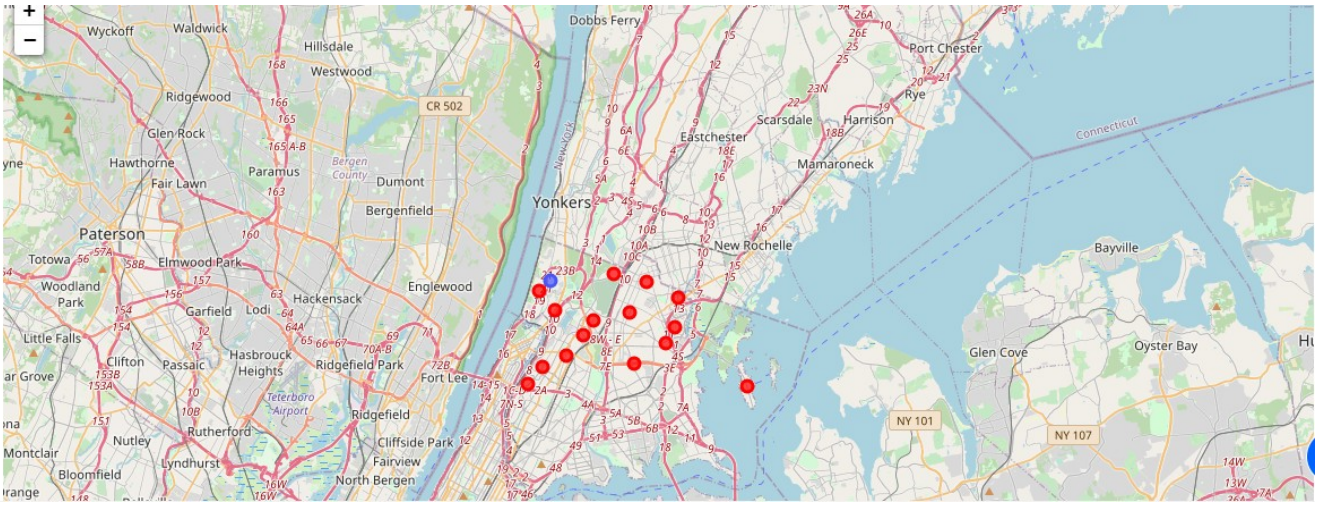

Fig 3.2(a)Clusters in Bronx

The average home price and school rating data frame is listed below for the Bronx area and explored using barchart

Neighborhood	Average_Housing_Price
Wakefield	335000.0
Co-op City	286600.0
Eastchester	335000.0
Fieldston	325900.0
Riverdale	419400.0

Fig4.2(b) Data Frame for Bronx neighborhood and their average housing price

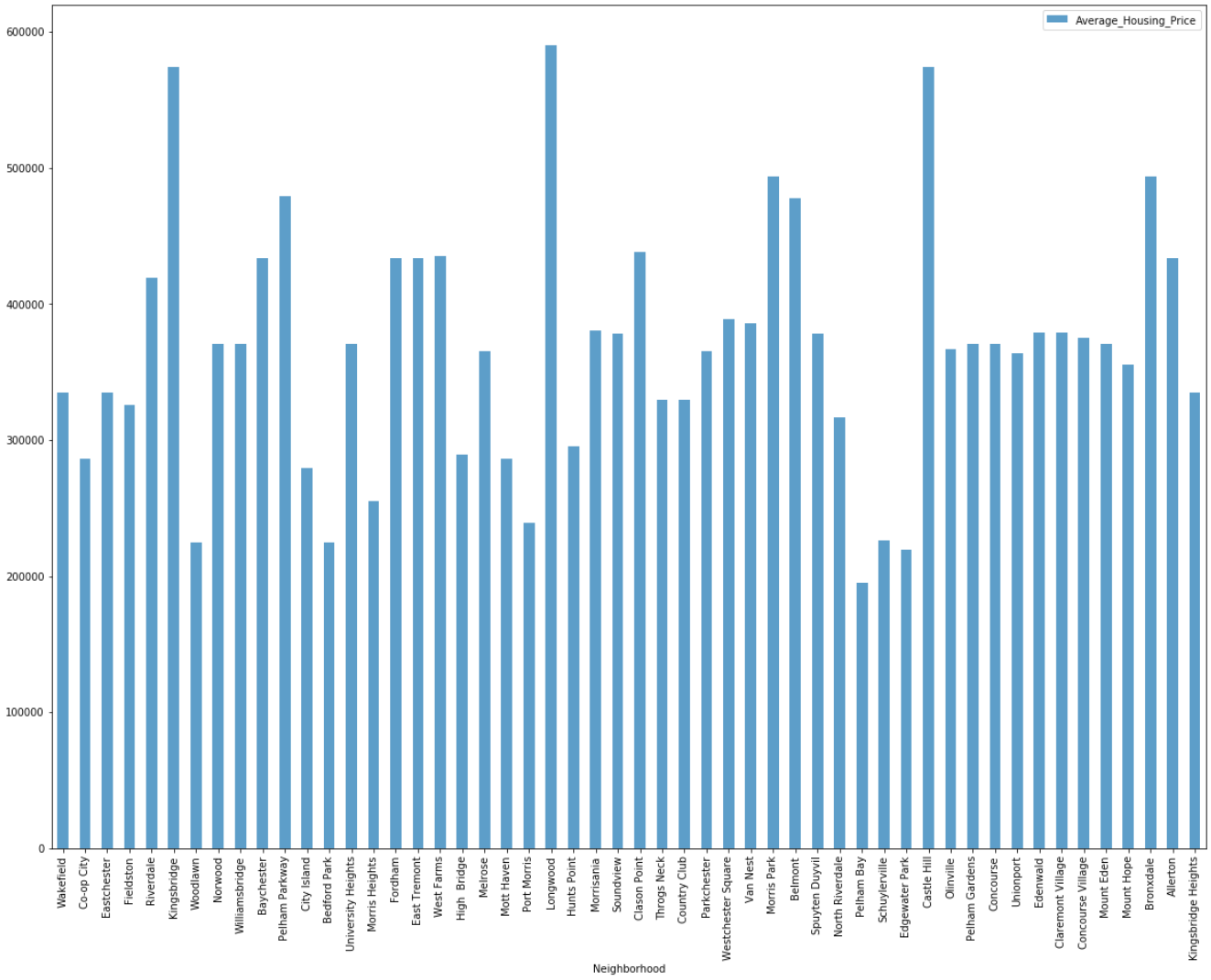

Fig 4.2(c) Bar chart - Bronx neighborhood Vs average housing price

Neighborhood	Top School Rating
Wakefield	7
Co-op City	9
Eastchester	5
Fieldston	8
Riverdale	10

Fig 4.2(d) Data Frame for Bronx neighborhood and school rating.

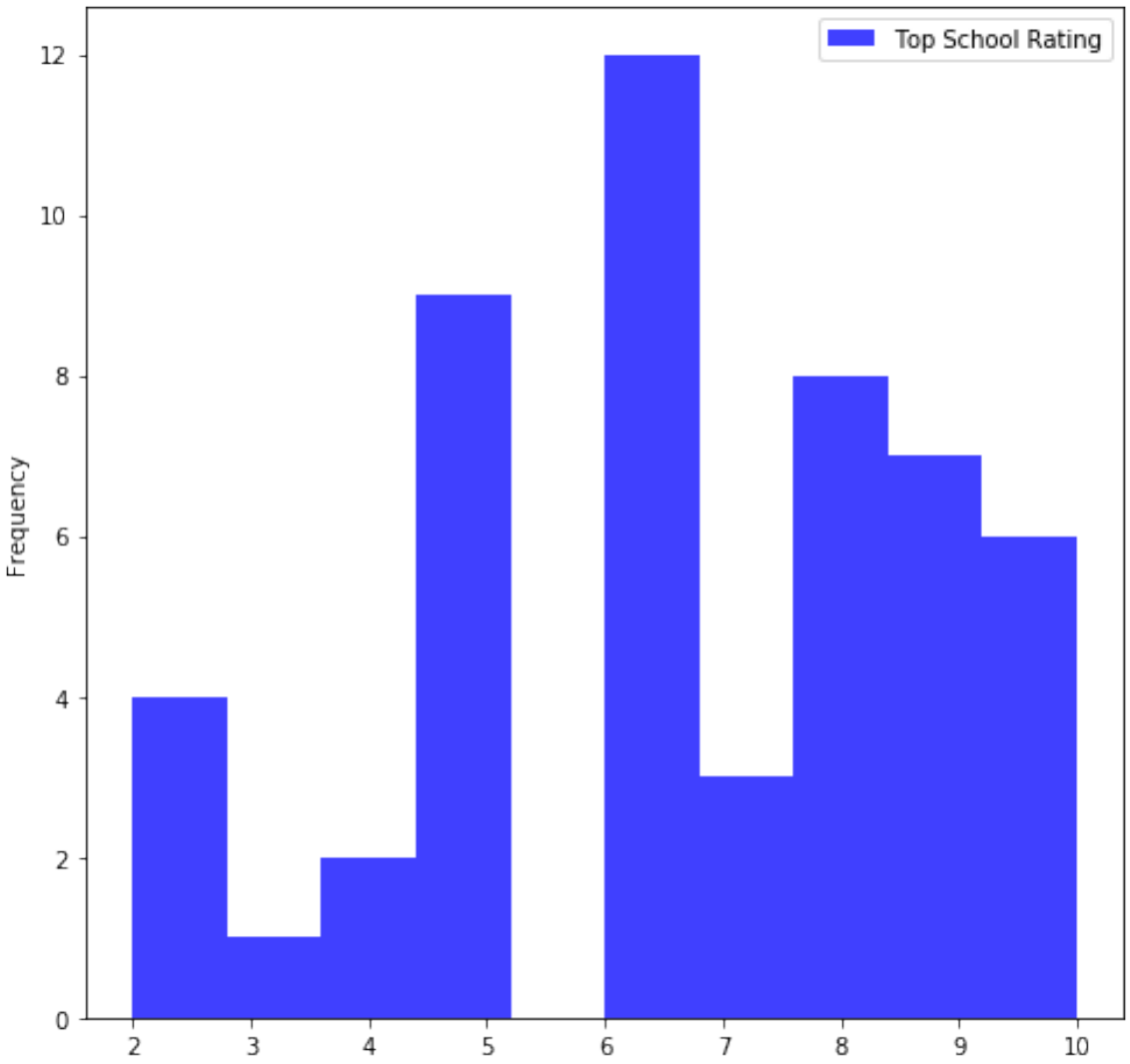

Fig 4.2(e) Histogram for Bronx neighborhood vs school rating

4. Results and discussion

During the analysis, clusters were defined. Bronx is only rare area , there are the many attractive options in terms of distances to the center of their New York cluster and relatively high value of income per person. However, one can perform further analysis of this particular cluster with additional features, such as distance to the center of city or to the center of cluster. After defining a borough, one can

perform deeper analysis to find the best exact location for buying home taking into account factors such as number of parking places in the vicinity of the spot or distances to the main streets. The median price of home in New York city is too high, the future scope is finding the best home price outside city could be the best choice. Same analysis can be done for place outside city considering traffic and distance as important factor.

5. Conclusion

To conclude, the basic data analysis was performed to identify the most optimal boroughs for finding home price and school rating in the city of Bronx, NY. During the analysis, several important statistical features of the boroughs were explored and visualized. Furthermore, clustering helped to highlight the group of areas.

6. References

1. New York city -Wikipedia
2. Housing Sales Prices of Each Borough from www.zillow.com
3. For School ratings: www.greatschools.com
4. Forsquare API
5. **https://data.cityofnewyork.us/City-Government/BoroughBoundaries/tqmj-j8zm**