The Battle of Neighborhoods

Applied Data Science Capstone – Report

Stefano I.

1. Introduction

1.1 Background

NYC is the most populous city in United States, often described as culture, financial, technology, education, tourism capital of Westrern world. Its liguistically most diverse city in the world with millions of foreign-born population of any city in the world.

New york City is a place for migrating people or individuals who arrives to this place with dream of starting there lives away from there comforting families. For any individual or a family to move to a new place will be a hard decision, and moving to a metropolitan like New York City will be even more intimidating.

This projects vision is to make it even more easier for individuals to take decision of which borough to choose to live in the city.

1.2 Business Problem

The objective of this Capstone Project is to analyse and choose the safest borough in the New York City based on the total crimes. This will help the people to buy/rent a home who are newly arriving to NYC. Exploring the neighbourhood and select the best among the five boroughs Brooklyn, Queens, Manhattan, The Bronx and Staten Island it has.

1.3 Who will use it?

The target audience for this problem will be all the individual or the families moving to this new place to make a decision of which location is safe and will be suitable for there preferences.

2. Data Description

2.1 Sources

The data for this project has been gathered from three main sources.
NYC crime data has been collected from the following link.

https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i

The data has then been downloaded and uploaded on the Jupyter notebook. A preview of the derived information can be noticed in Fig(a).

```
 #   Column              Non-Null Count    Dtype
---  ------              --------------    -----
 0   Crime_No            482337 non-null   int64
 1   Crime_DT            482337 non-null   object
 2   Crime_Reported_DT   482337 non-null   object
 3   Classification_Code 482337 non-null   int64
 4   Offence_Desc        482317 non-null   object
 5   Internal_Code       481968 non-null   float64
 6   Level               482337 non-null   object
 7   Borough             481961 non-null   object
 8   Latitude            475612 non-null   float64
 9   Longitude           475612 non-null   float64
10   Lat_Lon             475612 non-null   object
11   No_of_crimes        482337 non-null   int64
dtypes: float64(3), int64(3), object(6)
memory usage: 44.2+ MB
```

Fig: (a)

Subsequently, data is scrapped from the webpage at the link https://www.citypopulation.de/en/usa/newyorkcity/. The dataset included data related to the population within NYC boroughs. A preview of such information is available below, Fig (b).

```
Data columns (total 7 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   Name                         6 non-null      object
 1   Status                       6 non-null      object
 2   PopulationCensus1990-04-01   6 non-null      int64
 3   PopulationCensus2000-04-01   6 non-null      int64
 4   PopulationCensus2010-04-01   6 non-null      int64
 5   PopulationEstimate2019-07-01 6 non-null      int64
 6   Unnamed: 6                   5 non-null      object
dtypes: int64(4), object(3)
memory usage: 464.0+ bytes
```

Fig: (b)

Fig (C) is the representaion of the table I obtained after scrapping the data from the above-mentioned link.

	Borough	Status	Population-1990	Population-2000	Population-2010	Population-2019	Unnamed
0	Bronx	Borough	1203789	1332244	1384580	1418207	→
1	Brooklyn (Kings County)	Borough	2300664	2465689	2504721	2559903	→
2	Manhattan (New York County)	Borough	1487536	1538096	1586381	1628706	→
3	Queens	Borough	1951598	2229394	2230619	2253858	→
4	Staten Island (Richmond County)	Borough	378977	443762	468730	476143	→

Fig (c)

I have utilised only the Population 2019 as the crimes from the first source was only from the 2019 year. Refer Fig(d)

|	Borough	Status	Population-2019
0	Bronx	Borough	1418207
1	Brooklyn (Kings County)	Borough	2559903
2	Manhattan (New York County)	Borough	1628706
3	Queens	Borough	2253858
4	Staten Island (Richmond County)	Borough	476143

Fig(d)

After merging the NYPD crime data with the NYC census the data was Fig(e).

	Borough	Felony	Misdemeanor	Violation	Status	Population-2019	Total
1	BROOKLYN	46631	70504	21247	Borough	2559903	138382
2	MANHATTAN	38903	66785	15862	Borough	1628706	121550
0	BRONX	30356	57102	17367	Borough	1418207	104825
3	QUEENS	31369	49857	15975	Borough	2253858	97201
4	STATEN ISLAND	5059	10727	4217	Borough	476143	20003

Fig(e)

Third source to find the list of the Neighborhood of Staten Island as it is the most safest place according to our data. Fig(f)
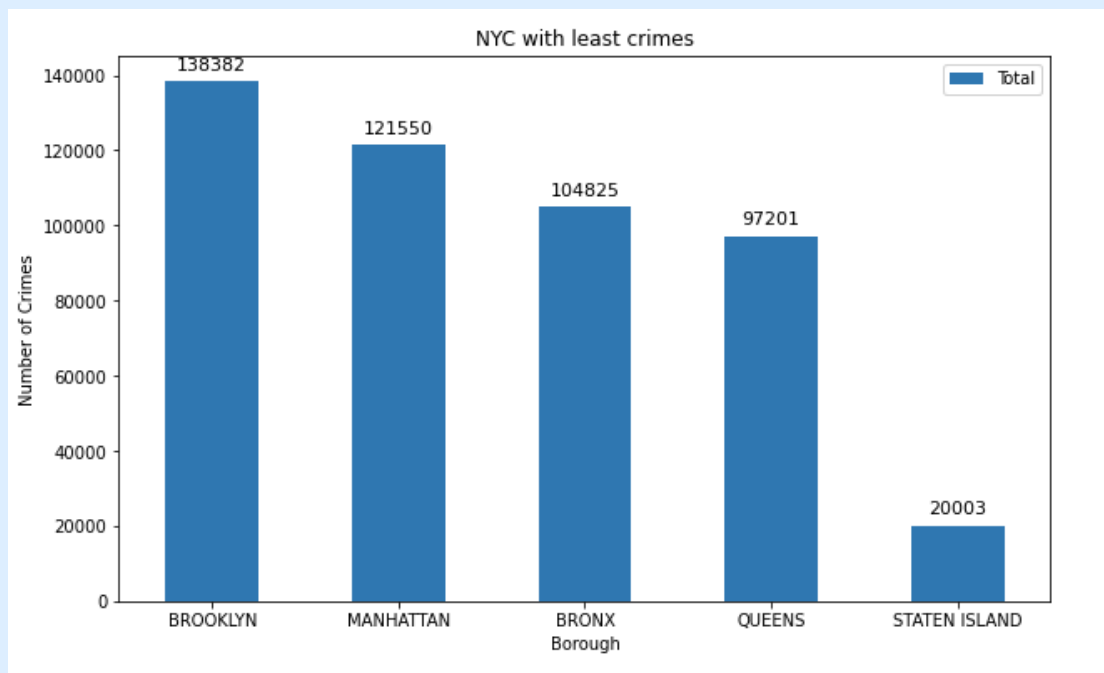
Fig(f)

The link to get the neighbourhood dataset is https://geo.nyu.edu/catalog/nyu_2451_34572

And I downloaded 'newyork_data.json'.
The Coordinates of the neighbourhood are acquired by Geopy library to get the latitude and longitude values of NYC.

The new dataset acquired about neighbourhood is explored and segmented using Foursquare API.