Data Science Capstone Project

Introduction:

The project is about in-depth study of the city of the New York. It is mainly focused on improving someone's travel or visit to the NYC. It considers the different tourist places, famous landmarks, amazing restaurants and many more.

It is mainly focused on finding how crime in NYC affects the tourism industry. This project also provide deeper understanding on how much a person spend on a particular item like hotel stays, best food/restaurants, visiting landmarks or museums, etc.

This topic is very interesting since few decades. And from the famous media companies, the city of New York also hired few data analyst and scientist to study this area. Lot of research are being made in this area. Personally, I thought that it would be great for me to just jump into the topic and discover little more than being discovered already.

Aim Statement:

To find the most enjoyable, safe and cheapest option for your next trip to NYC.

Data Source:

The data is fetched from the Foursquare API, which is best known for its location accuracy of the different landmarks, buildings, restaurants, parks, etc.

The developer version of the Foursquare is used to access premium calls.

The crime data is retrieved from NYPD crime statistics, available at:
https://www1.nyc.gov/site/nypd/stats/crime-statistics/citywide-crime-stats.page

Methodologies used:

- K-shift Clustering

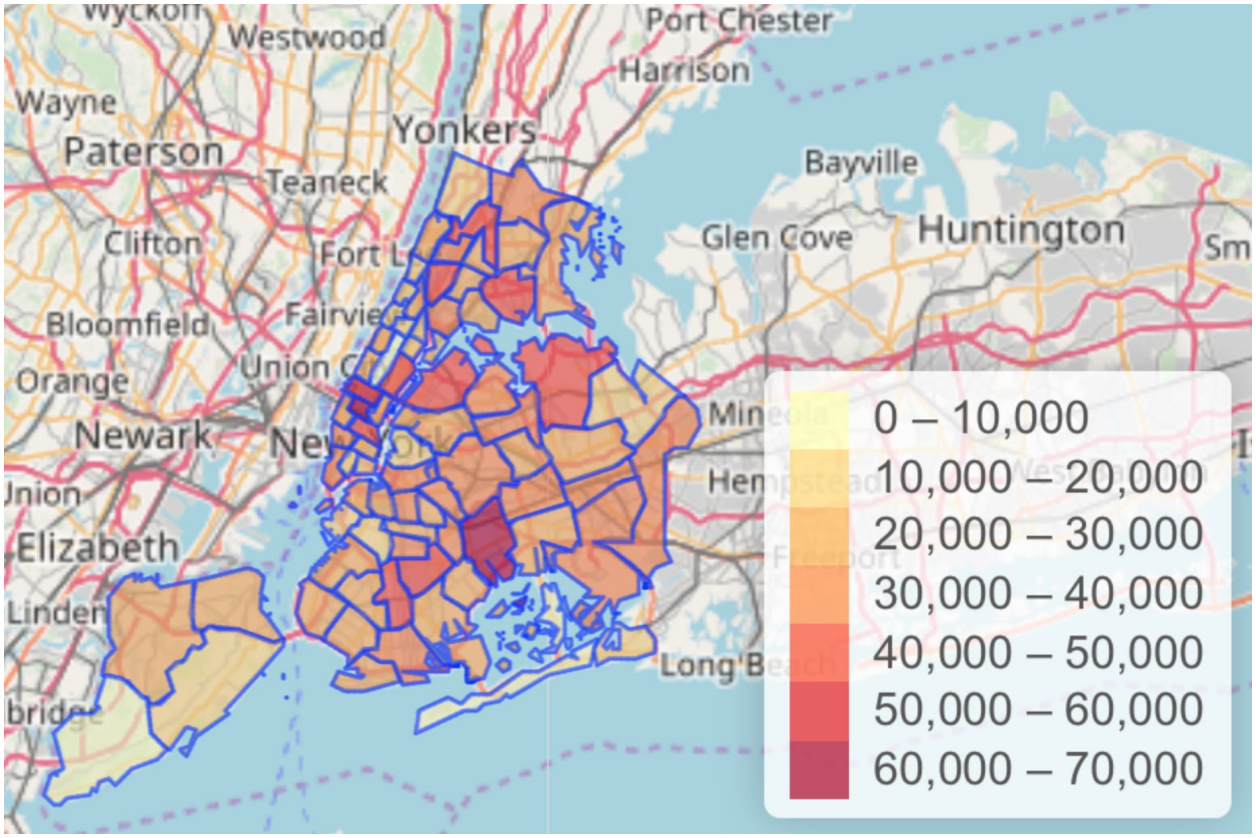
  To get to know the famous landmarks, building and parks based on the number of visits made by a tourist. Then I was able to cluster all this information together for further analysis.

- Mean shift Clustering

  To study the total number of visit made by an tourist over the period of time. It answered the question: What is the total number of visits over 10 years for particular landmark?

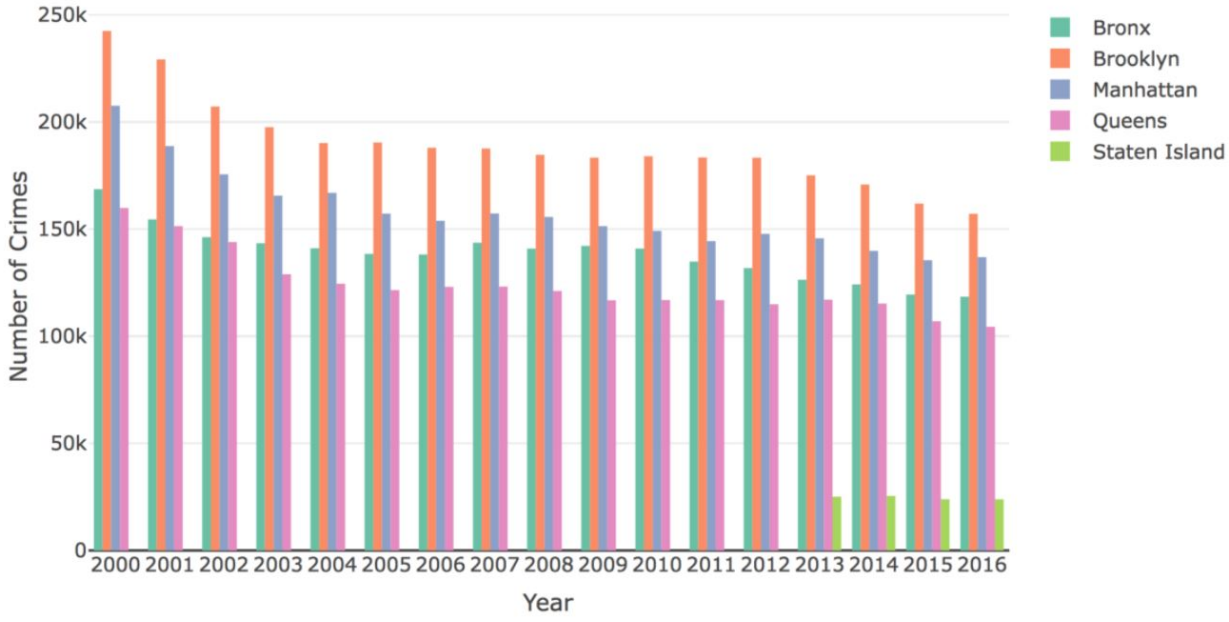
  The particular reason of using this type of Machine Learning concepts is that it suits perfectly to this type of data. I also tried using other concepts in my analysis, but at the end they were not making any sense, so I thought being stick to the basics.

Results:

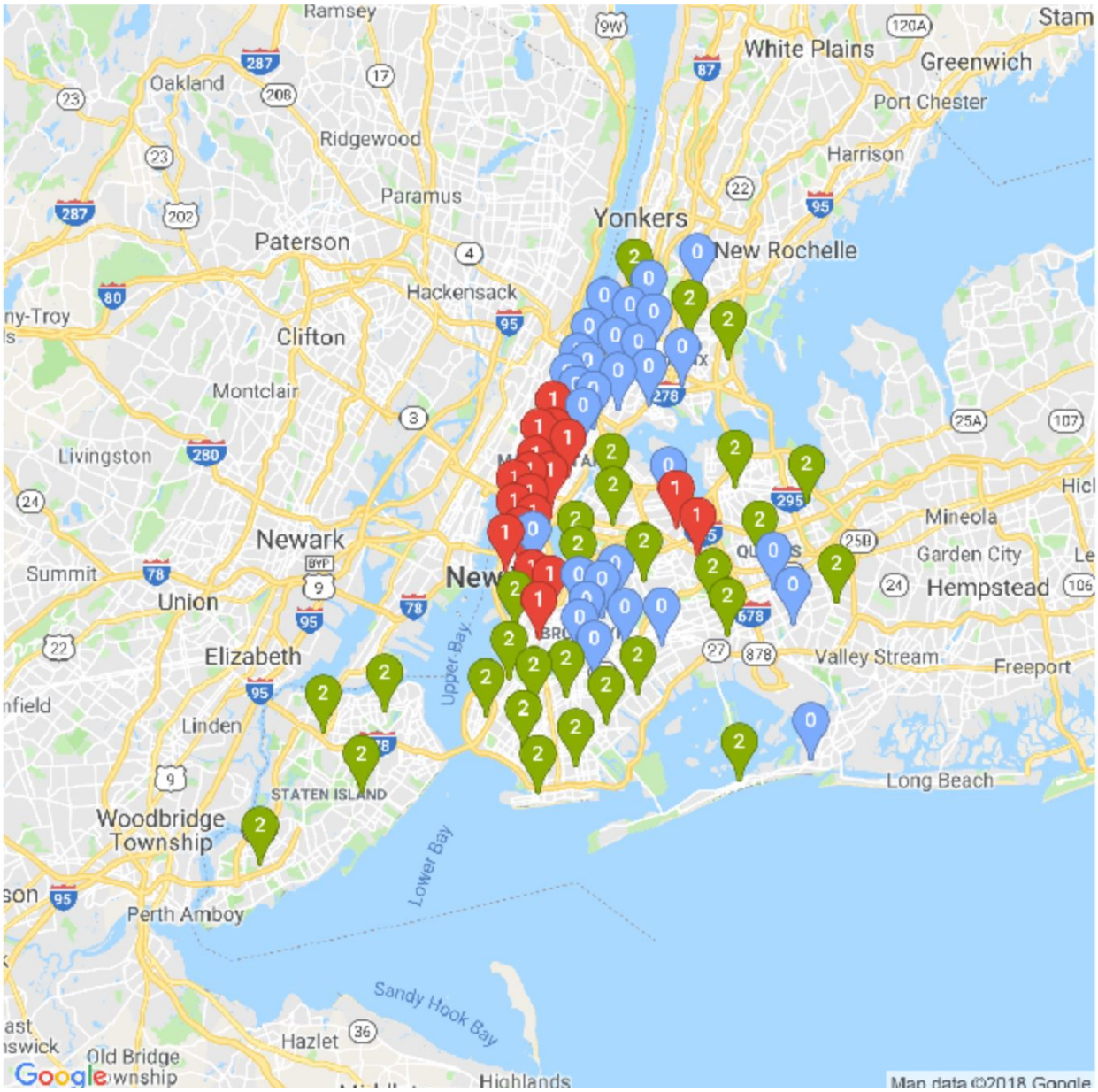


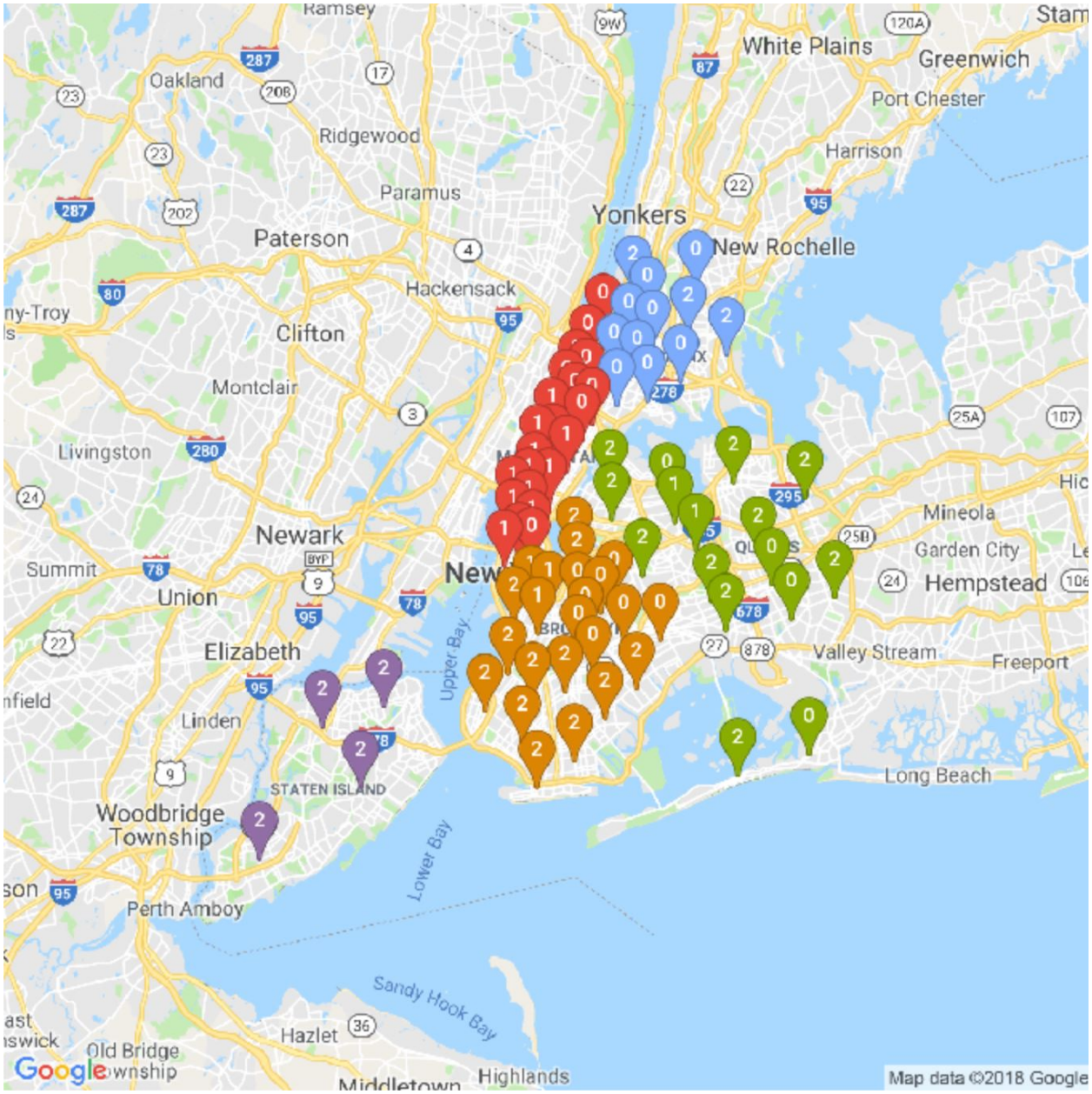
This is a choropleth map showing the highest crime affected areas of the NYC. You can see the highest crime rate are in the Brooklyn, followed by Manhattan and then few pop ups in Bronx and Queens.

The next graph is a focused more specifically on which areas/borough are most affected with crime. Obviously it has some historical data for further inference, but the main conclusion here will be that Brooklyn is always the highest crime affected borough at least from more than a decade. Followed by Manhattan, Bronx, Queens and Staten Island. Now we are going to compare the price index in all of these five borough.

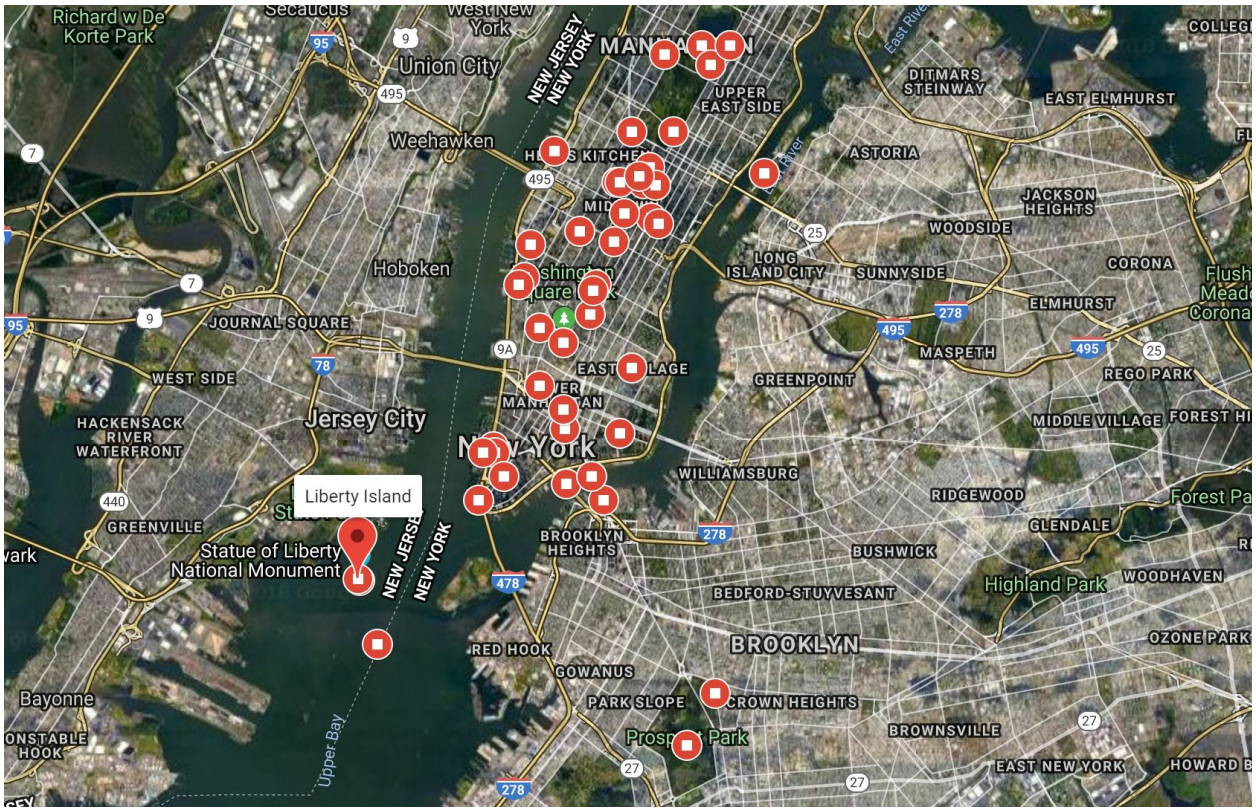


Now comes the clustering of the different borough for further study of the crime rate.




Ramsey
Oakland
Ridgewood
Paramus
Paterson
Hackensack
Clifton
Montclair
Livingston
Newark
Summit
Union
Elizabeth
Linden
STATEN ISLAND
Woodbridge Township
Perth Amboy
Old Bridge
Hazlet
Middletown
Highlands
Sandy Hook Bay
Lower Bay
Upper Bay
Yonkers
New Rochelle
White Plains
Greenwich
Port Chester
Harrison
Mineola
Garden City
Hempstead
Valley Stream
Freeport
Long Beach
Google
Map data ©2018 Google

The places that any ideal tourist is likely to visit is given by:



Discussion:

- Manhattan being the top tourist destination (having the more number of places to visit) has the second highest crime rate out of the five borough.
- Brooklyn has the highest crime rate and has hardly any nice place to visit. So ideal traveller may hardly visit and stay in Brooklyn. Even the data says that it is lot more cheaper to stay in Brooklyn, less tourist are likely to come because of the crime rate.
- Analysis is focused on Manhattan, since it is one of the most famous destination tourist are likely to visit.
- This project analyzed that an ideal tourist must stay and live close to Staten Island for saving maximum amount of money and time. By choosing this option, one may be close to major attractions, hotels, amenities, etc. at a cheaper rate as compared to the rest of the borough.

Conclusion:

- This project analyzed that an ideal tourist must stay and live close to Staten Island for saving maximum amount of money and time. By choosing this option, one may be close to major attractions, hotels, amenities, etc. at a cheaper rate as compared to the rest of the borough.