**The Battle of Neighborhoods**

**Staten Island, New York City Analysis**

### **Introduction:**

This Project aims to give people who want to migrate to New York city a clear understanding about where to settle or relocate and start their life in the Boroughs of New York city. This project caters for the following important things which every person looks for in a city when relocating:

1. **Crime Analysis** in New York City and mainly Staten Island.
2. **Real Estate Prices**, Location and House Size in Staten Island.
3. **Schools** Locations in Staten Island.
4. **Hospitals** Locations, Rating and Features in Staten Island.
5. **Most Common Venues**

**Targeted Problems:**

1. Anyone who wants to relocate to a new city. The crime rate of that city must be analyzed. This project does exactly that, Crime Rate of all Boroughs of NYC will be analyzed, and the Borough with Least Crime Rate will be suggested to the user.
2. Real Estate Price is one of the most important things to look for while relocating. Everyone wants a good place to settle which is in a good neighborhood but not very expensive. This project will give the Most and Least Expensive neighborhoods of Staten Island.
3. All parents want their children to get good education. This project suggests users the schools in every neighborhood in Staten Island and School types, Level etc.
4. Hospitals is an overlooked point while moving to a new Location. This project will give the information about nearby Hospitals, their overall Rating, and many Features.
5. Finally, this project will provide an analysis of Most Visited venues like Restaurants, Grocery Shops, Hospitals, Malls Rating etc. in every Neighborhood of Staten Island so relocators know beforehand which venues they value the most so it will become easier for them to settle.

**Data Description**

The following Datasets will be used for this project:

1. **NYPD Arrest Data dataset** of year 2020 available on CityOfNewyork Website. This dataset will be used to analyze the crime rate in the Boroughs of NYC. This dataset includes crime rate, crime type, criminal race and gender and the location where the crime occurred.
2. **Real Estate dataset** of Staten Island, NYC available on nyc.gov website. This dataset includes Real Estate prices, size, type of houses of every neighborhood in Staten Island, NYC.
3. **USA Hospitals dataset** available on Kaggle. This dataset includes Hospital name, location, type etc. and will be further cleaned to include only Staten Island Hospitals.
4. **USA Hospitals Ratings and Features dataset** available on Kaggle. This dataset includes Hospitals Ratings and many Features like Safety of care national comparison, Patient experience national comparison, Effectiveness of care national comparison etc. This dataset will be merged with the above dataset to make one big dataset that will include all required attributes in one dataset.
5. **USA Public Schools dataset** available on Kaggle. This dataset includes details of all public schools in USA and will be further cleaned to only include public schools of Staten Island, NYC.
6. **Staten Island Neighborhoods** available on Wikipedia. The Wikipedia page will provide the names of all Neighborhoods of Staten Island and will be used to get the geographic locations using Geopy and Foursquare API.
7. **Foursquare API** will be used to get the Latitude & Longitude values for Neighborhoods and Venues of Staten Island, NYC.

**Methodology:**

There were three main parts for the project:

1. **Data Engineering**
2. **Machine Learning**
3. **Results**

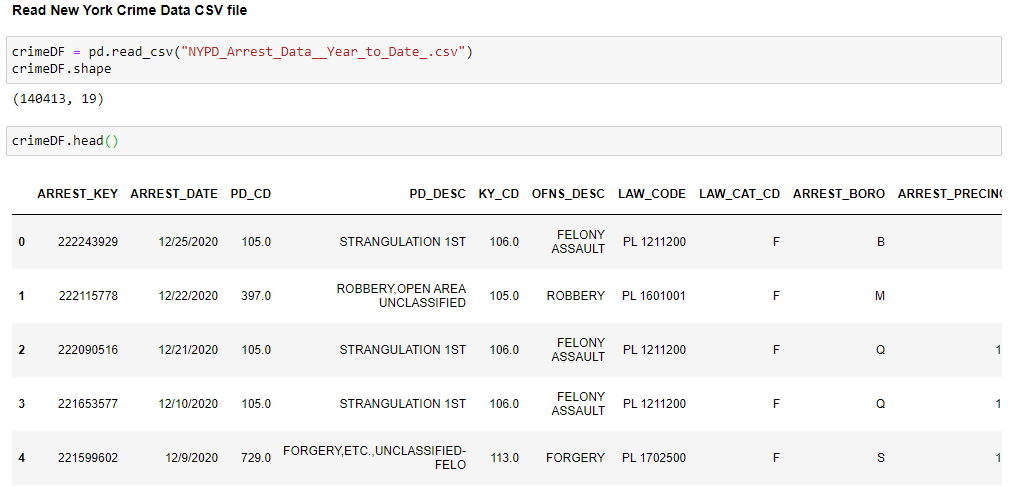
**Data Engineering:**

The Data wrangling, Data cleansing and Data preparation were combined into Data Engineering. There was a total of 6 datasets used in this project and in every Analysis the dataset had to be cleaned and prepared for analysis.

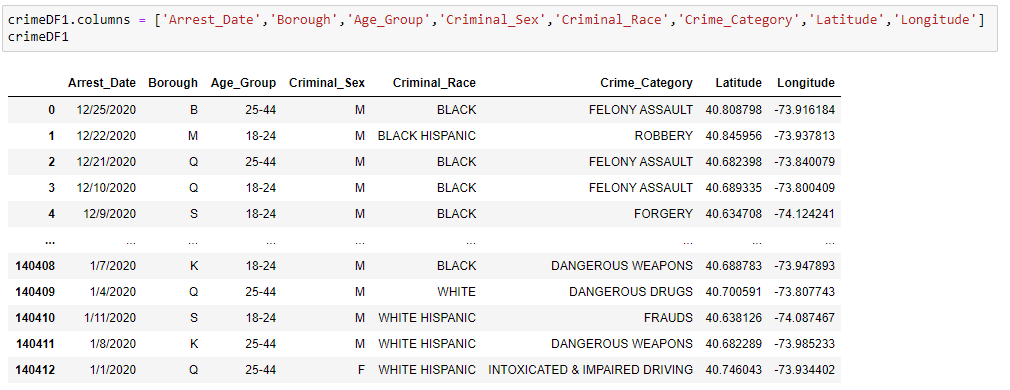
A total of 5 attributes had to be analyzed to decide whether Staten Island is a suitable place to live or not. The points were:

**Crime Analysis:**

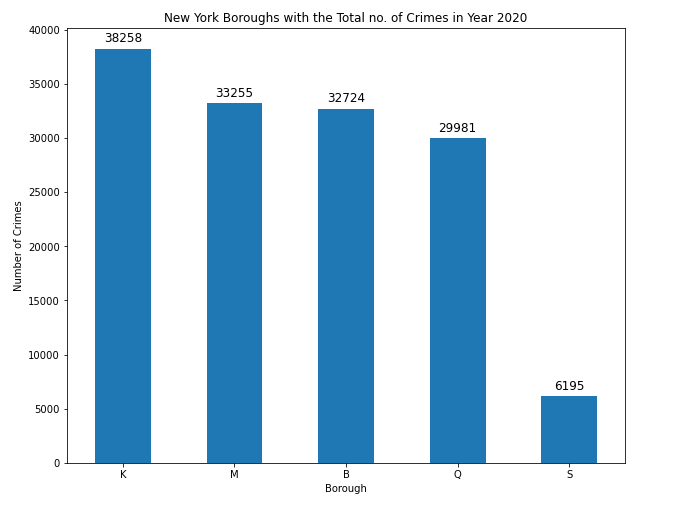
At first the Crime Analysis for all of 5 boroughs of New York City was performed.



The Dataframe was modified to only include required columns like Arrest Date, Arrest location, Crime type etc. Columns names were also modified for ease of use later.

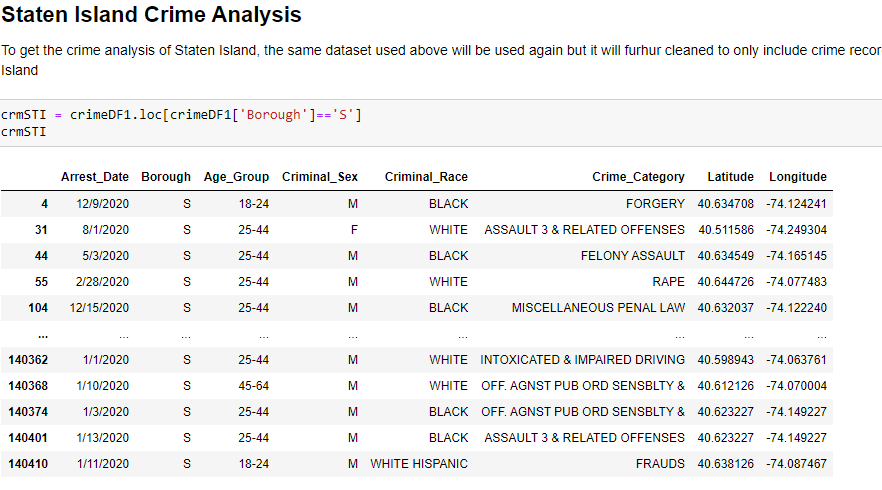


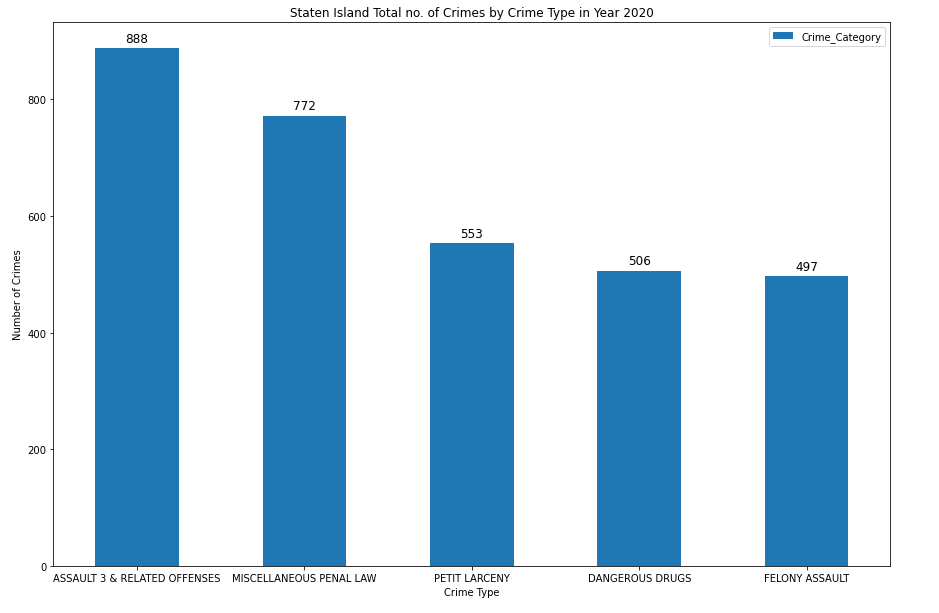
After Dataframe modification, the Crime Records was plotted on bar chart using matplotlib.



After analyzing the chart, Staten Island was chosen as best place to live amongst the five boroughs because of the lowest crime rate.

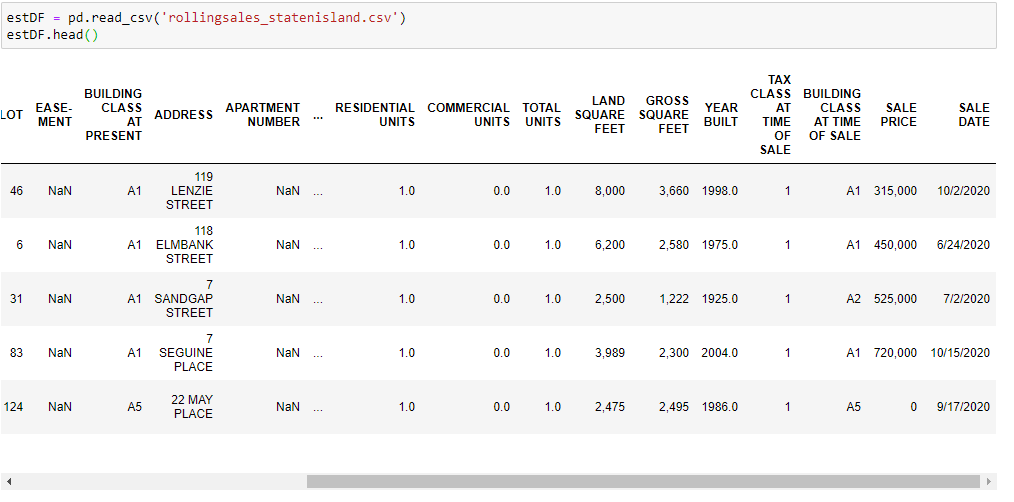
Only Staten Island Crime Record was selected from the dataframes and plotted afterwards.





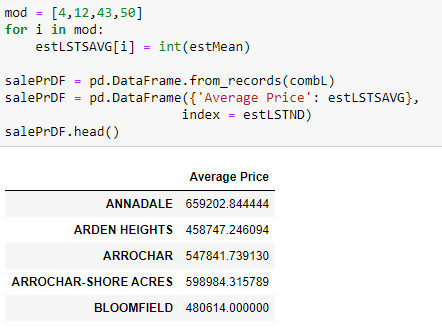
**Real Estate Analysis:**

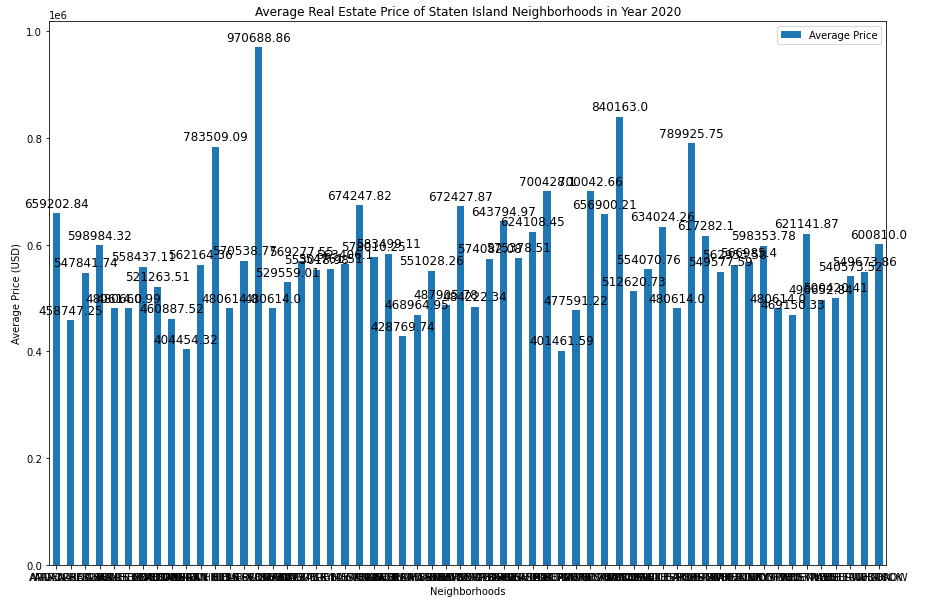
After crime analysis, Staten Island Real Estate was analyzed. The dataset used for this was taken from nyc.gov website.



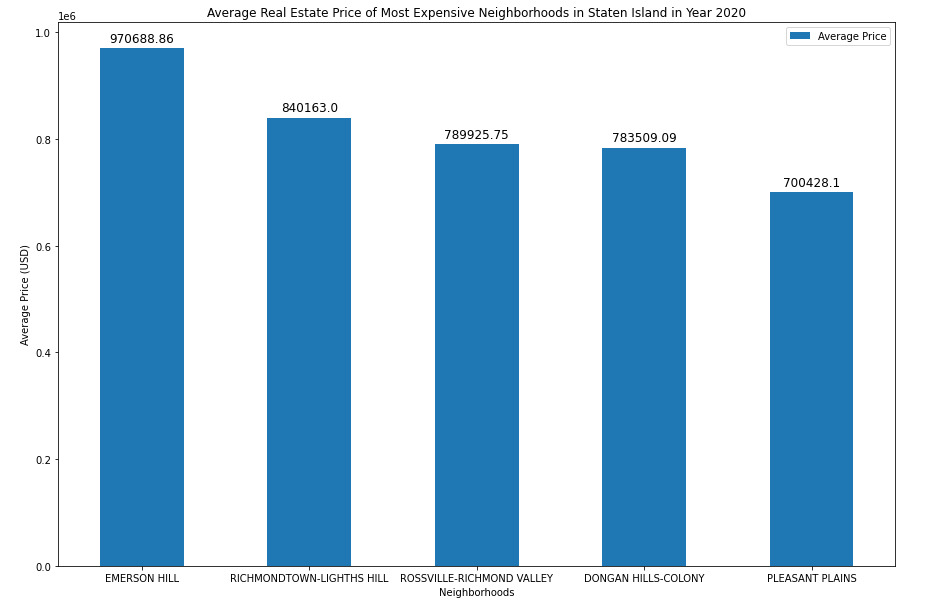
The dataset was further cleaned. The saleprice column had all values in string format which made it in accessible while plotting. So all the string were converted into int. 

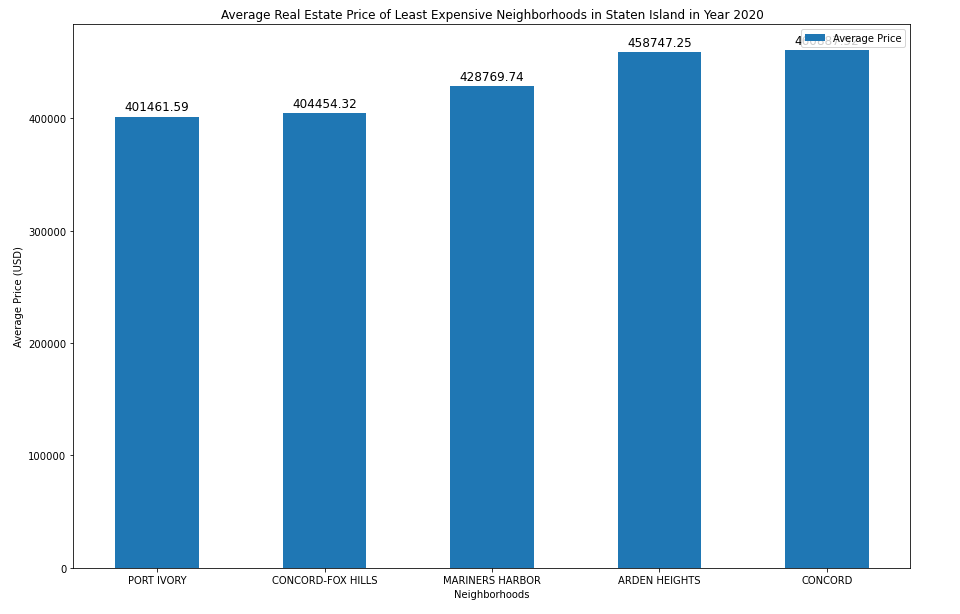
The mean value was needed to get the average price of real estate in every neighborhood which was further used to plot.





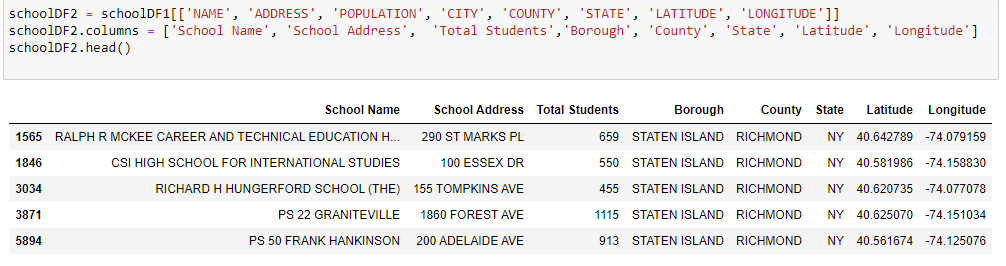
The most expensive and least expensive neighborhoods were plotted on chart in similar way.



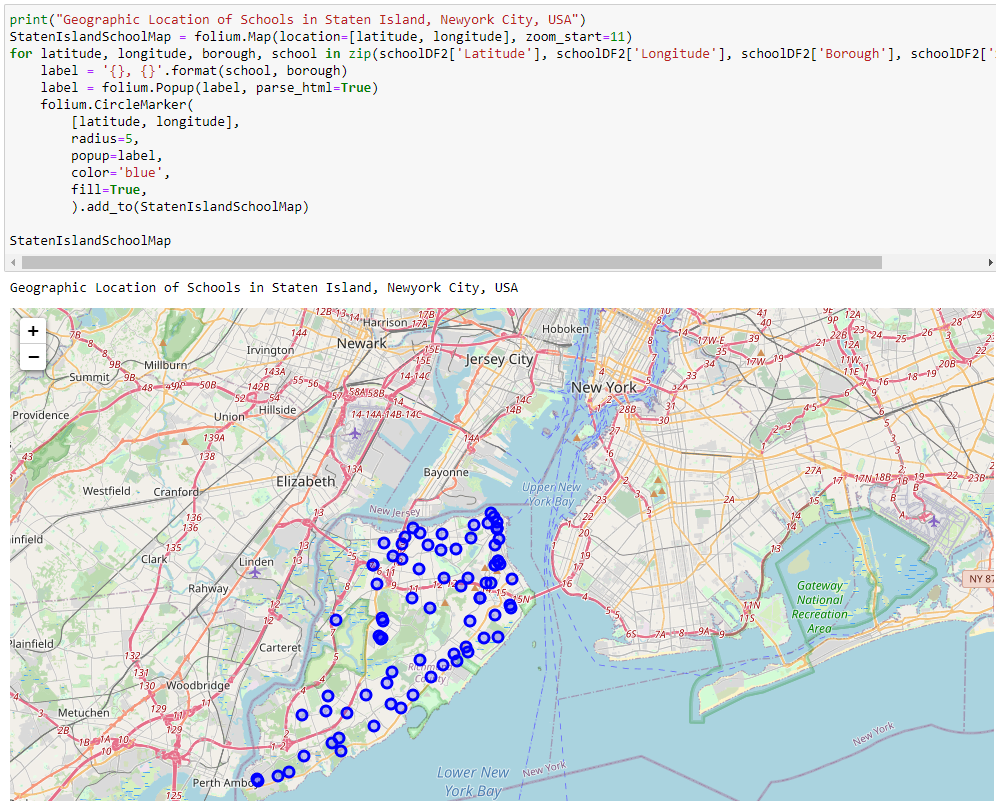


**Schools Analysis:**

After Real Estate, we move down to School Analysis of Staten Island. For this task, USA Public Schools dataset available on Kaggle will be used. The Dataset contains information regarding Schools from all over the USA, but we only need information about Staten Island schools to the dataset will be cleaned further to include only that.

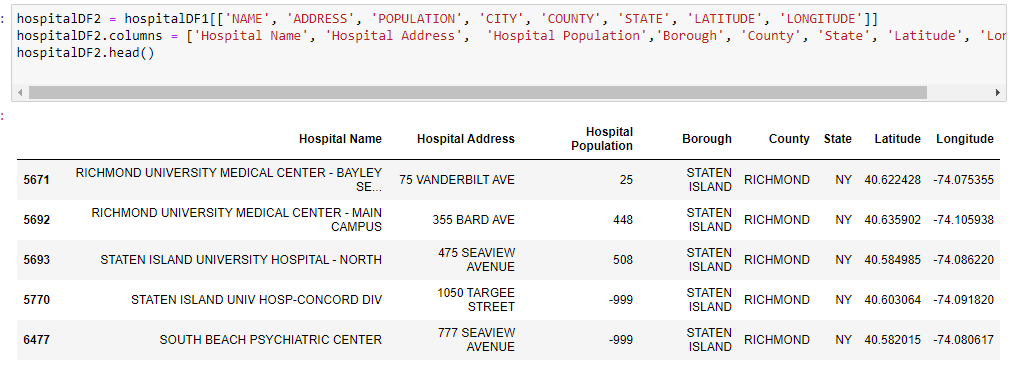


Folium was used to plot the location of schools on map using Latitude and Longitude values.

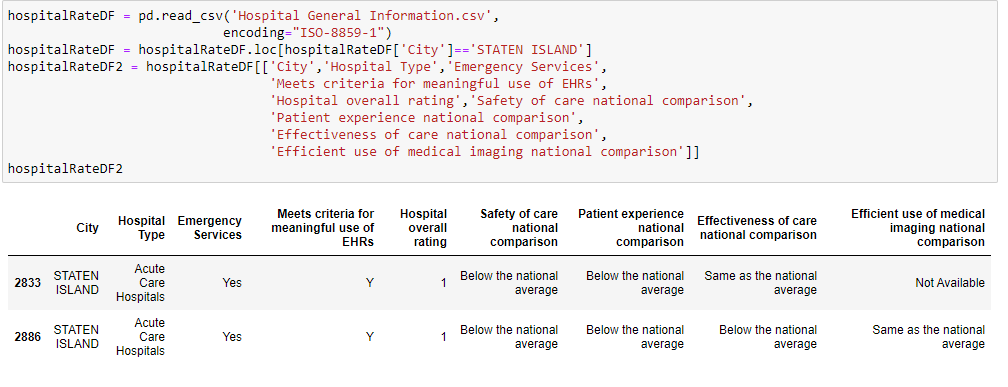


**Hospitals Analysis:**

After School Analysis, we will look into Staten Island Hospitals. For this task, two datasets will be used. First dataset is the USA Hospitals dataset available on Kaggle. This dataset includes Hospital name, location etc. and second dataset is the USA Hospitals Ratings and Features dataset available on Kaggle. This dataset includes Hospitals Ratings and many Features. Both datasets will be merged to include only the required attributes.

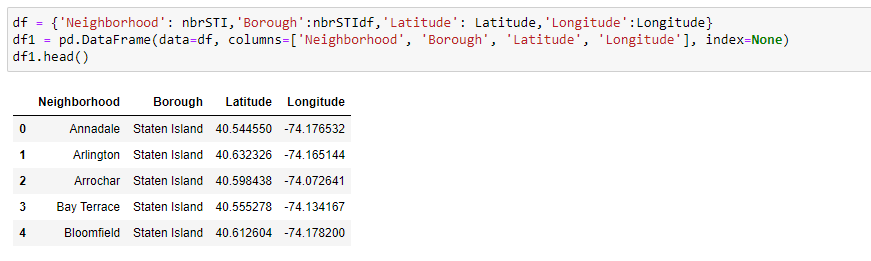


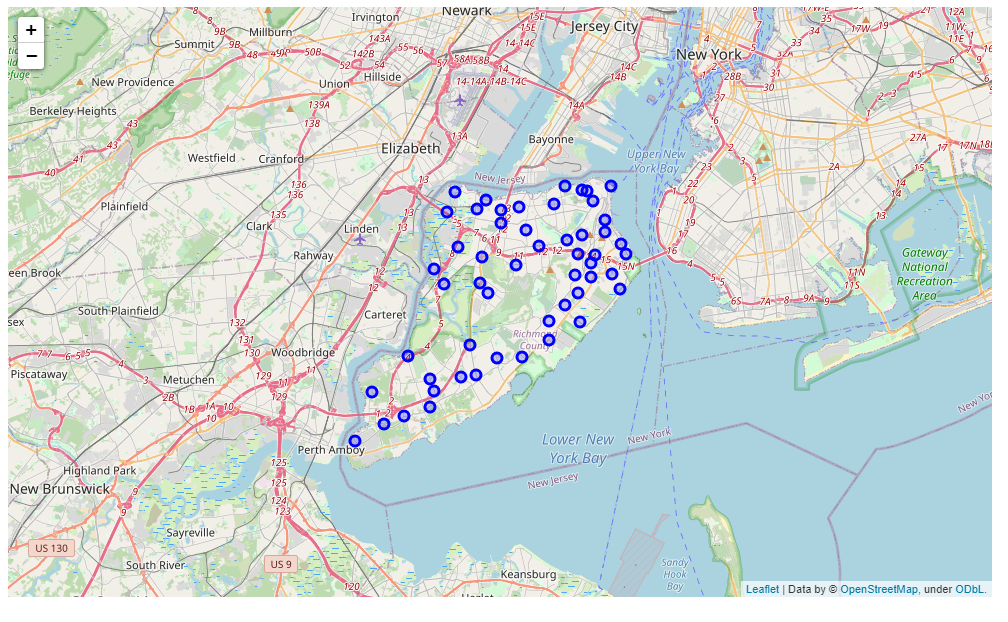
Two of the hospitals meet the criteria for meaningful use of EHRs. Safety of care and Patient experience is below the national comparison level and the Effectiveness of care when compared with national comparison, one of them is below that.



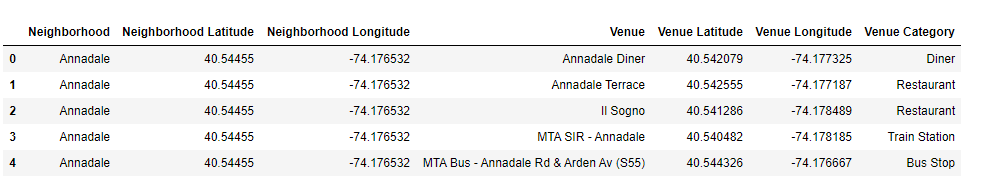
**Most Common Venues:**

All the neighborhoods in Staten Island were gathered together. The names of the neighborhoods are available on Wikipedia.The names for the all the Neighborhoods were used to get the latitude and longitude values for all the Neighborhoods using **Geocode**.

****

****

FourSquare API wil be used to get the nearby venus of every neighborhood in Staten Island.

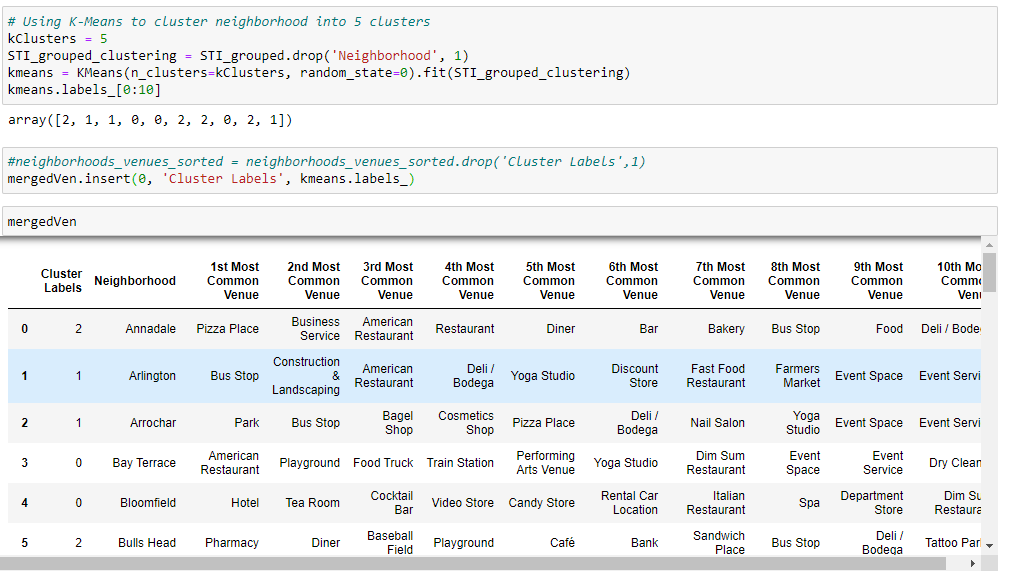
****

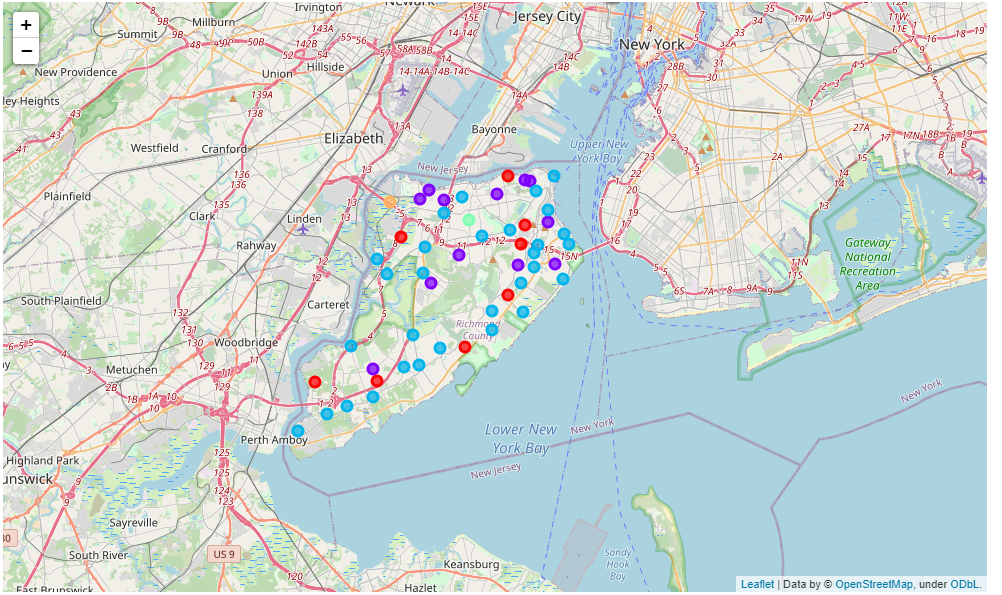
The most common venues of all neighborhoods will be returned.

****

**Machine Learning:**

The second part of the project was the machine learning phase. In this phase the data returned by the foursquare API was used for comparison between all the aforementioned places and getting the most visited venues in every neighborhood of Staten Island clustered together, an efficient algorithm will be used. For this project, **K-Means clustering** will be used as the primary **unsupervised machine learning** approach-Means clustering has vector quantization method and uses **Squared Euclidean Distance** to cluster neighborhoods based on most visits and other similarities.





**Results**

After Clustering the following are the inferred results from the clustered dataframes:

In **Cluser 1**, The most common venues are Construction & Landscaping.

In **Cluser 2**, The most common venue is Bus Stop.

**Cluser 3** is the biggest cluster so it has multiple most common venues, the most common venues are Pizza Places and restaurants.

**Cluser 4 & 5** are the smallest clusters and the most common venues are Harbor/Marina and Arcade.

**Conclusion:**

After performing all of the steps required to perform in a Data science Pipeline. We end up with a refined and clear understanding as to whether one should settle or not in Staten Island, If yes then where to settle can also be found out easily. Now We'll look at the main attributes separately to understand better from the results:

**1. Crime Rate:** First main point was Crime Rate. We can easily conclude now from the above results that the Crime Rate for Staten Island is the lowest and is relatively peaceful to safer to live in when compared with its neighboring boroughs. So, if safety is the top priority then Staten Island is the best.

**2. Real Estate:** Next comes the Real Estate. Staten Island has both expensive and cheap, big and small, clean and average neighborhoods. One can find an average sized Real Estate at affordable rate in Staten Island with some effort. If one has wealth to spend then Emerson Hills, Richmond Town and Roseville valley are the top 3 options. If budget and quality are on average then Grant City, Clove Lakes and Bulls Head are some of the best option. If budget is low, then Port Ivory and Concord-Fox hills should be the one to go for.

**3. Schools:** In Staten Island, the number of schools are high, and their quality is high as well. Staten Island has above average schools with high rating and most of the schools provide education till high school level.

**4. Hospitals:** Staten Island only lacks in terms of hospitals. The Borough only has 6 hospitals and the rating for two of them is 1 which is low when compared with hospitals of other boroughs. Both hospitals meet the criteria for meaningful use of EHRs. Safety of care and Patient experience is below the national comparison level and the Effectiveness of care when compared with national comparison, one of them is below that. In conclusion, Hospitals in Staten Islands are not the best in the nation, but they do provide what most do not.

**5. Venues** At last, all the venues of every neighborhood were clustered. In the Results section, one can clearly see the most visited venues of every neighborhood, so it depends on the person that what they favors the most.

Finally, we can conclude that Staten Island is an above average place to settle. If someone values safety, affordable real estate, good schools then Staten Island is the city to settle down.