



# The Battle of Neighborhoods

Staten Island, New York City Analysis

March  
2021

# 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
2. Real Estate Prices
3. Schools
4. Hospitals
5. Most Common Venues



# Location/City:

---

The city of choice for this project is Staten Island, New York City, USA. New York City is the one of the most iconic and biggest city of USA and a bustling metropolis. The city comprises of Five major Boroughs and for this project we'll be focusing on Staten Island. Staten Island is known for its various landmarks like Staten Island Museum, Fort Wadsworth, Historic Richmond Town, Low Crime Rate, way above average Schools, Hospitals etc.





# 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.

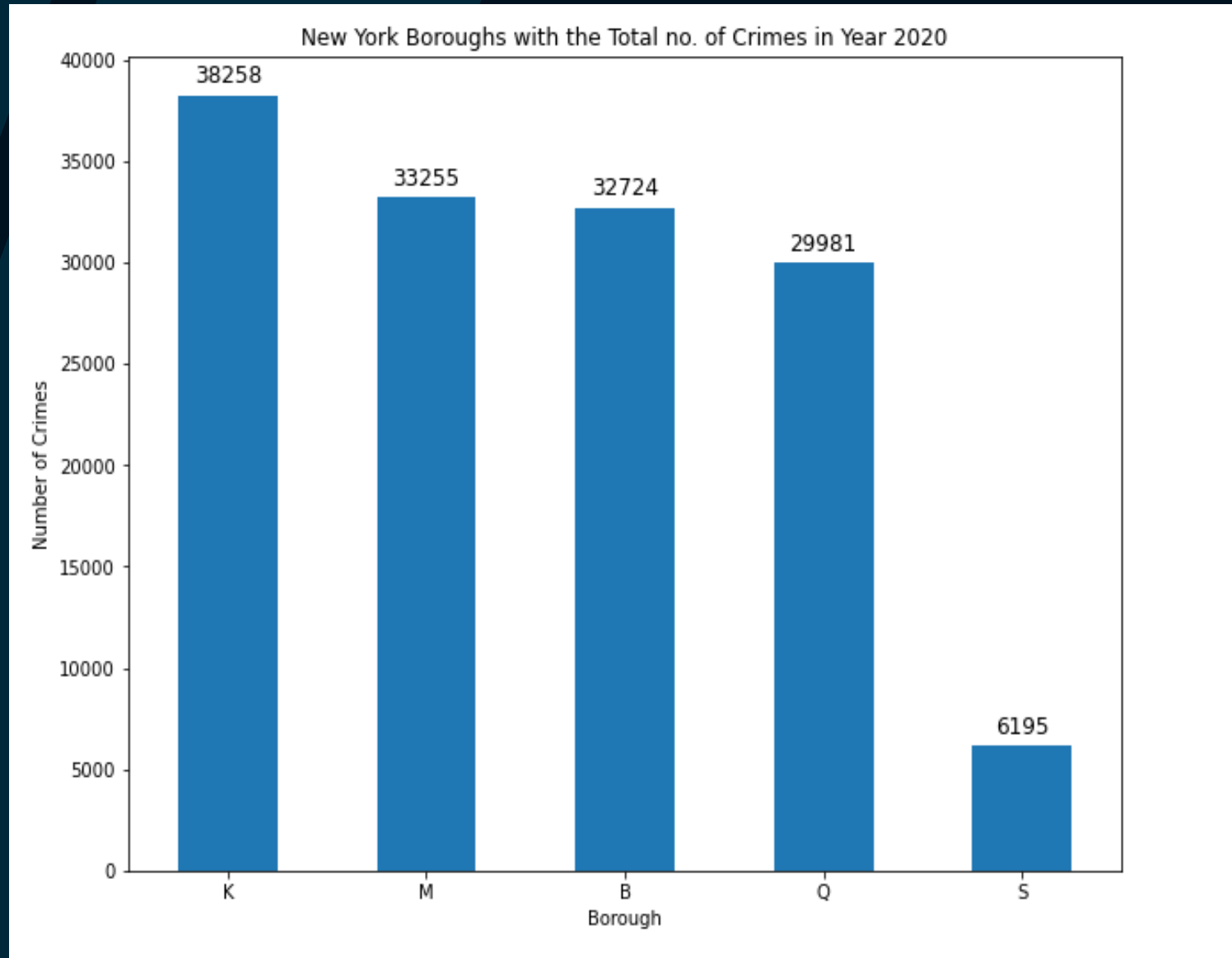
# Crime Analysis:

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

```
crimeDF1.columns = ['Arrest_Date', 'Borough', 'Age_Group', 'Criminal_Sex', 'Criminal_Race', 'Crime_Category', 'Latitude', 'Longitude']  
crimeDF1
```

	Arrest_Date	Borough	Age_Group	Criminal_Sex	Criminal_Race	Crime_Category	Latitude	Longitude
0	12/25/2020	B	25-44	M	BLACK	FELONY ASSAULT	40.808798	-73.916184
1	12/22/2020	M	18-24	M	BLACK HISPANIC	ROBBERY	40.845956	-73.937813
2	12/21/2020	Q	25-44	M	BLACK	FELONY ASSAULT	40.682398	-73.840079
3	12/10/2020	Q	18-24	M	BLACK	FELONY ASSAULT	40.689335	-73.800409
4	12/9/2020	S	18-24	M	BLACK	FORGERY	40.634708	-74.124241
...	...	...	...	...	...	...	...	...
140408	1/7/2020	K	18-24	M	BLACK	DANGEROUS WEAPONS	40.688783	-73.947893
140409	1/4/2020	Q	25-44	M	WHITE	DANGEROUS DRUGS	40.700591	-73.807743
140410	1/11/2020	S	18-24	M	WHITE HISPANIC	FRAUDS	40.638126	-74.087467
140411	1/8/2020	K	25-44	M	WHITE HISPANIC	DANGEROUS WEAPONS	40.682289	-73.985233
140412	1/1/2020	Q	25-44	F	WHITE HISPANIC	INTOXICATED & IMPAIRED DRIVING	40.746043	-73.934402

# Bar Chart for NYC Crime Analysis



# Staten Island Crime Analysis:

Afterwards, only the Crime Analysis for all Staten Island was performed.

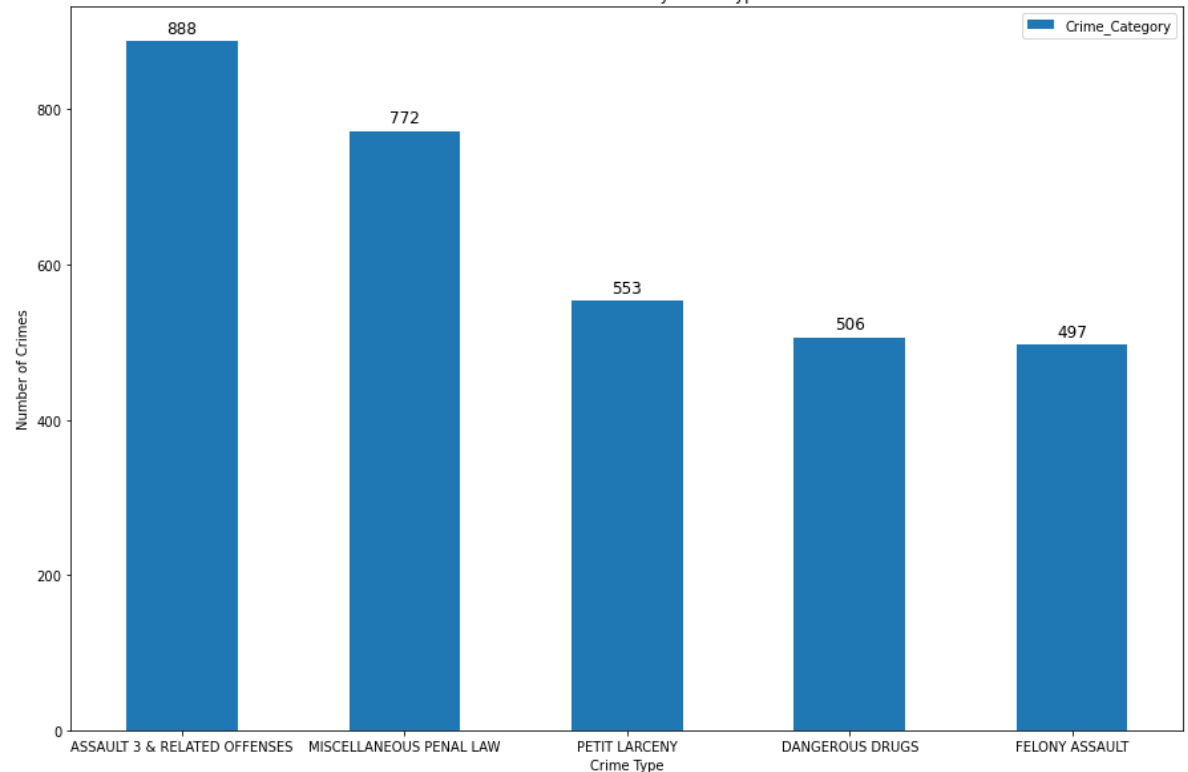
## Staten Island Crime Analysis

To get the crime analysis of Staten Island, the same dataset used above will be used again but it will further cleaned to only include crime records from Staten Island

```
crmSTI = crimeDF1.loc[crimeDF1['Borough']=='S']  
crmSTI
```

	Arrest_Date	Borough	Age_Group	Criminal_Sex	Criminal_Race	Crime_Category	Latitude	Longitude
4	12/9/2020	S	18-24	M	BLACK	FORGERY	40.634708	-74.124241
31	8/1/2020	S	25-44	F	WHITE	ASSAULT 3 & RELATED OFFENSES	40.511586	-74.249304
44	5/3/2020	S	25-44	M	BLACK	FELONY ASSAULT	40.634549	-74.165145
55	2/28/2020	S	25-44	M	WHITE	RAPE	40.644726	-74.077483
104	12/15/2020	S	25-44	M	BLACK	MISCELLANEOUS PENAL LAW	40.632037	-74.122240
...	...	...	...	...	...	...	...	...
140362	1/1/2020	S	25-44	M	WHITE	INTOXICATED & IMPAIRED DRIVING	40.598943	-74.063761
140368	1/10/2020	S	45-64	M	WHITE	OFF. AGNST PUB ORD SENSBLTY &	40.612126	-74.070004
140374	1/3/2020	S	25-44	M	BLACK	OFF. AGNST PUB ORD SENSBLTY &	40.623227	-74.149227
140401	1/13/2020	S	25-44	M	BLACK	ASSAULT 3 & RELATED OFFENSES	40.623227	-74.149227
140410	1/11/2020	S	18-24	M	WHITE HISPANIC	FRAUDS	40.638126	-74.087467

Staten Island Total no. of Crimes by Crime Type in Year 2020



# Real Estate Analysis:

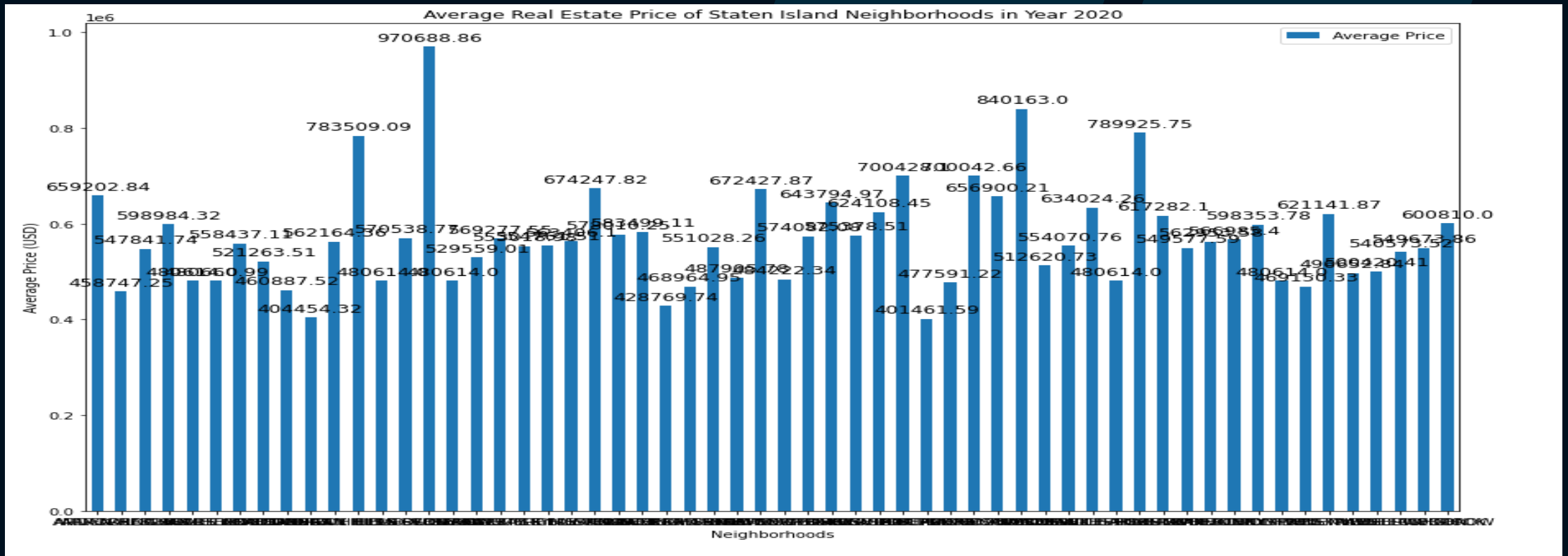
After crime analysis, Staten Island Real Estate was analyzed. The dataset used for this was taken from nyc.gov website.

```
estDF['SALE PRICE'] = estDF['SALE PRICE'].str.replace(',', '').astype(float)
estDF['SALE PRICE'] = estDF['SALE PRICE'].astype(int)
estDF.head()
```

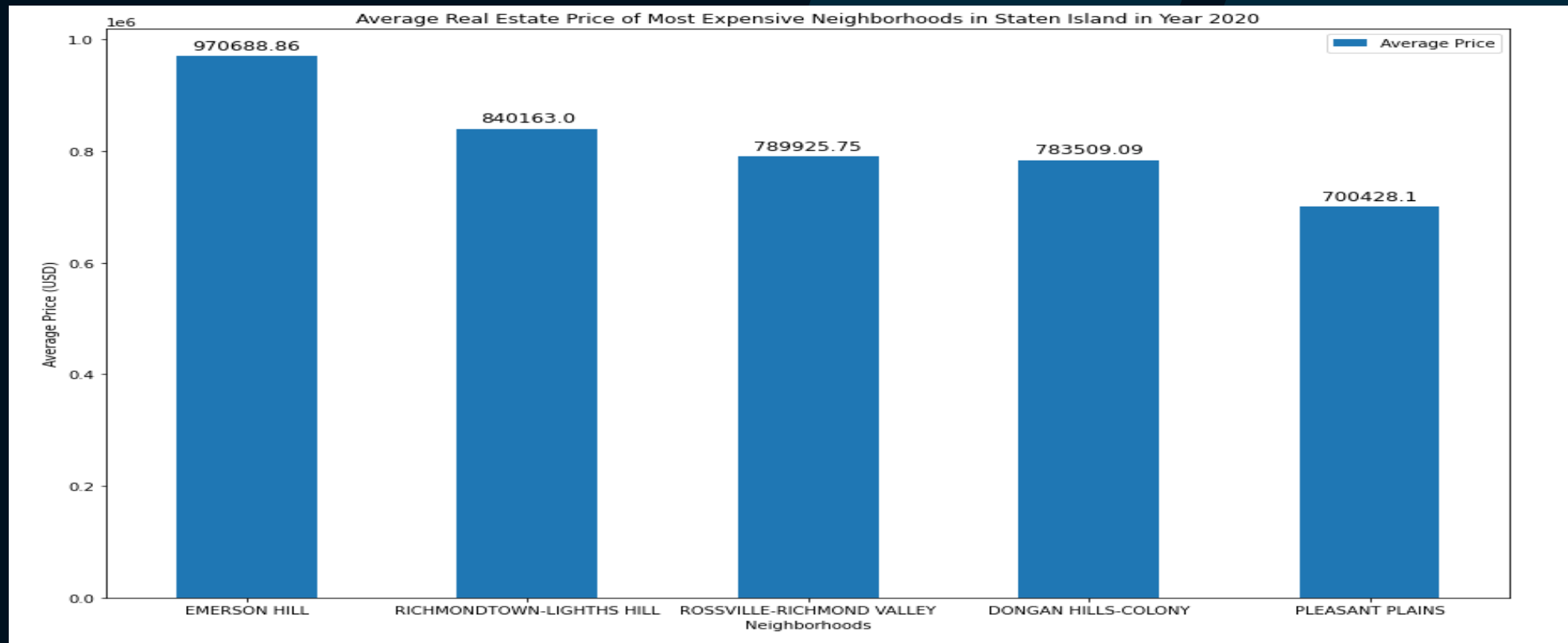
LOT	EASE- MENT	BUILDING CLASS AT PRESENT	ADDRESS	APARTMENT NUMBER	...	RESIDENTIAL UNITS	COMMERCIAL UNITS	TOTAL UNITS	LAND SQUARE FEET	GROSS SQUARE FEET	YEAR BUILT	TAX CLASS AT TIME OF SALE	BUILDING CLASS AT TIME OF SALE	SALE PRICE	SALE DATE
46	NaN	A1	119 LENZIE STREET	NaN	...	1.0	0.0	1.0	8,000	3,660	1998.0	1	A1	315000	10/2/2020
6	NaN	A1	118 ELMBANK STREET	NaN	...	1.0	0.0	1.0	6,200	2,580	1975.0	1	A1	450000	6/24/2020
31	NaN	A1	7 SANDGAP STREET	NaN	...	1.0	0.0	1.0	2,500	1,222	1925.0	1	A2	525000	7/2/2020
83	NaN	A1	7 SEGUINE PLACE	NaN	...	1.0	0.0	1.0	3,989	2,300	2004.0	1	A1	720000	10/15/2020
124	NaN	A5	22 MAY PLACE	NaN	...	1.0	0.0	1.0	2,475	2,495	1986.0	1	A5	0	9/17/2020



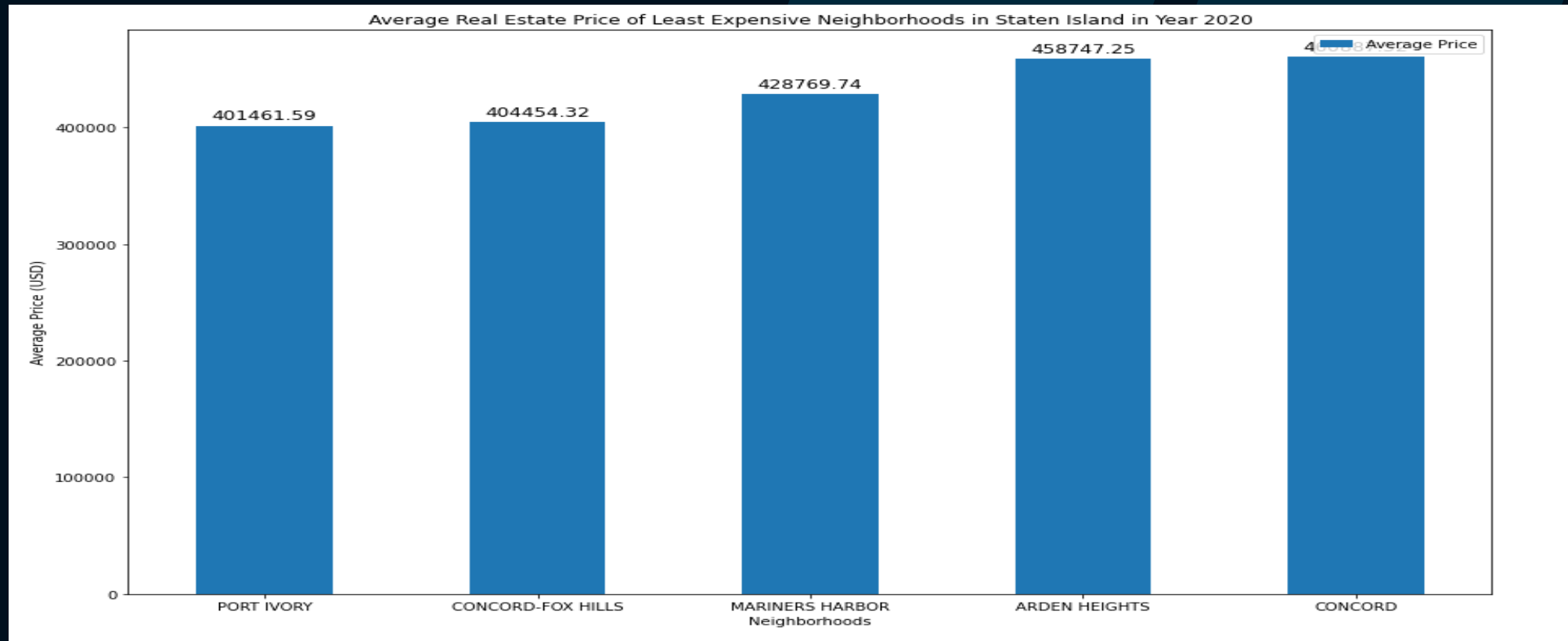
# Average Price of Real Estate in Staten Island



# Most Expensive Real Estate in Staten Island



# Least Expensive Real Estate in Staten Island



# Schools Analysis:

After Real Estate, we moved down to School Analysis of Staten Island. For this task, USA Public Schools dataset available on Kaggle was used. The Dataset contains information regarding Schools from all over the USA.

```
schoolDF2 = schoolDF1[['NAME', 'ADDRESS', 'POPULATION', 'CITY', 'COUNTY', 'STATE', 'LATITUDE', 'LONGITUDE']]
schoolDF2.columns = ['School Name', 'School Address', 'Total Students', 'Borough', 'County', 'State', 'Latitude', 'Longitude']
schoolDF2.head()
```

	School Name	School Address	Total Students	Borough	County	State	Latitude	Longitude
1565	RALPH R MCKEE CAREER AND TECHNICAL EDUCATION H...	290 ST MARKS PL	659	STATEN ISLAND	RICHMOND	NY	40.642789	-74.079159
1846	CSI HIGH SCHOOL FOR INTERNATIONAL STUDIES	100 ESSEX DR	550	STATEN ISLAND	RICHMOND	NY	40.581986	-74.158830
3034	RICHARD H HUNGERFORD SCHOOL (THE)	155 TOMPKINS AVE	455	STATEN ISLAND	RICHMOND	NY	40.620735	-74.077078
3871	PS 22 GRANITEVILLE	1860 FOREST AVE	1115	STATEN ISLAND	RICHMOND	NY	40.625070	-74.151034
5894	PS 50 FRANK HANKINSON	200 ADELAIDE AVE	913	STATEN ISLAND	RICHMOND	NY	40.561674	-74.125076



# Schools Analysis:

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



# Hospitals Analysis:

After School Analysis, we looked 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.

```
hospitalDF2 = hospitalDF1[['NAME', 'ADDRESS', 'POPULATION', 'CITY', 'COUNTY', 'STATE', 'LATITUDE', 'LONGITUDE']]
hospitalDF2.columns = ['Hospital Name', 'Hospital Address', 'Hospital Population', 'Borough', 'County', 'State', 'Latitude', 'Longitude']
hospitalDF2.head()
```

	Hospital Name	Hospital Address	Hospital Population	Borough	County	State	Latitude	Longitude
5671	RICHMOND UNIVERSITY MEDICAL CENTER - BAYLEY SE...	75 VANDERBILT AVE	25	STATEN ISLAND	RICHMOND	NY	40.622428	-74.075355
5692	RICHMOND UNIVERSITY MEDICAL CENTER - MAIN CAMPUS	355 BARD AVE	448	STATEN ISLAND	RICHMOND	NY	40.635902	-74.105938
5693	STATEN ISLAND UNIVERSITY HOSPITAL - NORTH	475 SEAVIEW AVENUE	508	STATEN ISLAND	RICHMOND	NY	40.584985	-74.086220
5770	STATEN ISLAND UNIV HOSP-CONCORD DIV	1050 TARGEE STREET	-999	STATEN ISLAND	RICHMOND	NY	40.603064	-74.091820
6477	SOUTH BEACH PSYCHIATRIC CENTER	777 SEAVIEW AVENUE	-999	STATEN ISLAND	RICHMOND	NY	40.582015	-74.080617

# Hospitals Features:

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.

```
hospitalRateDF = pd.read_csv('Hospital General Information.csv',
                             encoding="ISO-8859-1")
hospitalRateDF = hospitalRateDF.loc[hospitalRateDF['City']=='STATEN ISLAND']
hospitalRateDF2 = hospitalRateDF[['City','Hospital Type','Emergency Services',
                                   'Meets criteria for meaningful use of EHRs',
                                   'Hospital overall rating','Safety of care national comparison',
                                   'Patient experience national comparison',
                                   'Effectiveness of care national comparison',
                                   'Efficient use of medical imaging national comparison']]

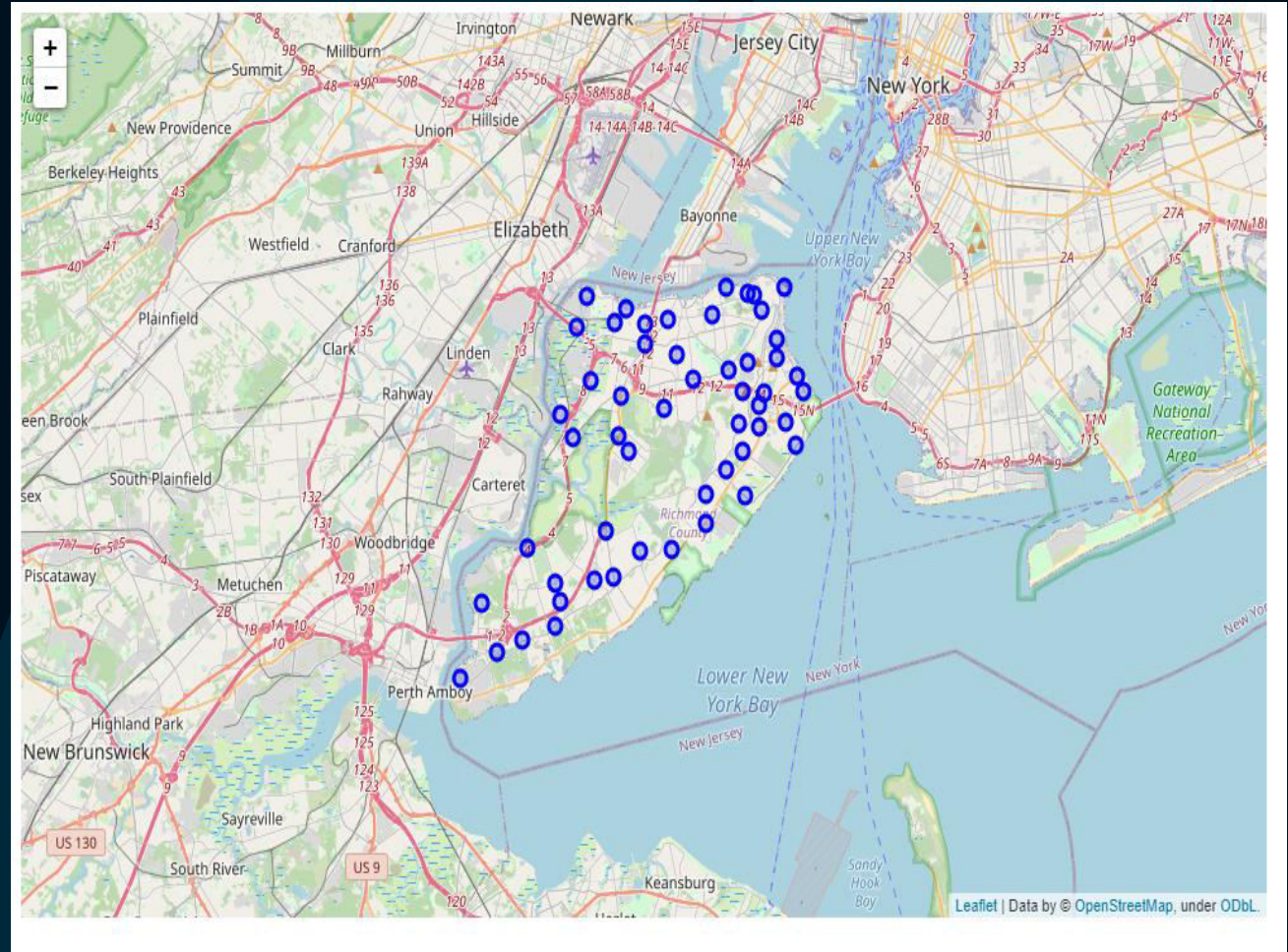
hospitalRateDF2
```

	City	Hospital Type	Emergency Services	Meets criteria for meaningful use of EHRs	Hospital overall rating	Safety of care national comparison	Patient experience national comparison	Effectiveness of care national comparison	Efficient use of medical imaging national comparison
2833	STATEN ISLAND	Acute Care Hospitals	Yes	Y	1	Below the national average	Below the national average	Same as the national average	Not Available
2886	STATEN ISLAND	Acute Care Hospitals	Yes	Y	1	Below the national average	Below the national average	Below the national average	Same as the national average



# Neighborhood Analysis

All the neighborhoods in Staten Island were gathered and plotted using geopy and folium.





# Most Common Venues:

FourSquare API will be used to get the nearby venues of every neighborhood in Staten Island.

Out[266]:

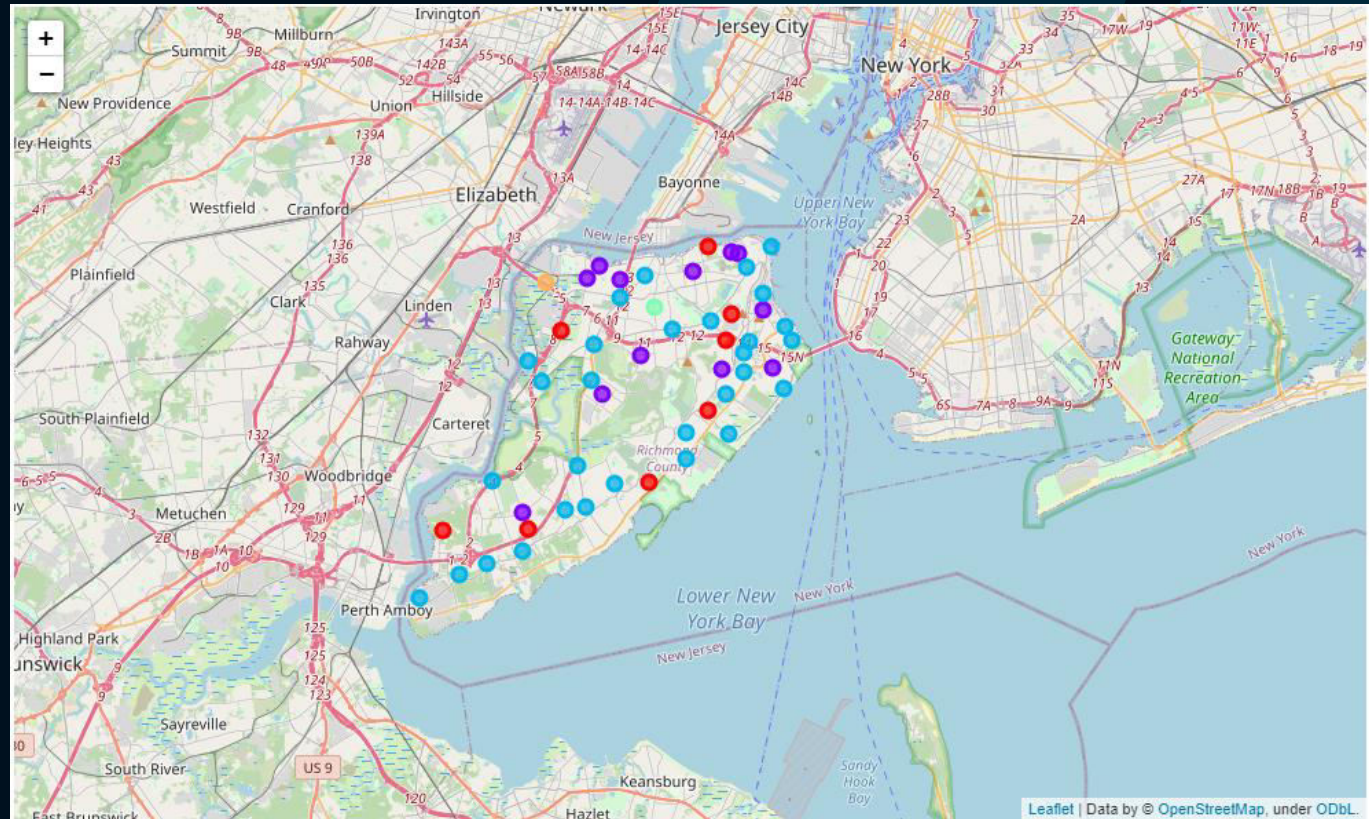
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Annadale	Pizza Place	Business Service	American Restaurant	Restaurant	Diner	Bar	Bakery	Bus Stop	Food	Deli / Bodega
1	Arlington	Bus Stop	Construction & Landscaping	American Restaurant	Deli / Bodega	Yoga Studio	Discount Store	Fast Food Restaurant	Farmers Market	Event Space	Event Service
2	Arrochar	Park	Bus Stop	Bagel Shop	Cosmetics Shop	Pizza Place	Deli / Bodega	Nail Salon	Yoga Studio	Event Space	Event Service
3	Bay Terrace	American Restaurant	Playground	Food Truck	Train Station	Performing Arts Venue	Yoga Studio	Dim Sum Restaurant	Event Space	Event Service	Dry Cleaner
4	Bloomfield	Hotel	Tea Room	Cocktail Bar	Video Store	Candy Store	Rental Car Location	Italian Restaurant	Spa	Department Store	Dim Sum Restaurant
5	Bulls Head	Pharmacy	Diner	Baseball Field	Playground	Café	Bank	Sandwich Place	Bus Stop	Deli / Bodega	Tattoo Parlor

# Machine Learning:

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.

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	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
0	Annadale	Staten Island	40.544550	-74.176532	2.0	Pizza Place	Business Service	American Restaurant	Restaurant	Diner	Bar	Bakery
1	Arlington	Staten Island	40.632326	-74.165144	1.0	Bus Stop	Construction & Landscaping	American Restaurant	Deli / Bodega	Yoga Studio	Discount Store	Fast Food Restaurant
2	Arrochar	Staten Island	40.598438	-74.072641	1.0	Park	Bus Stop	Bagel Shop	Cosmetics Shop	Pizza Place	Deli / Bodega	Nail Salon
3	Bay Terrace	Staten Island	40.555278	-74.134167	0.0	American Restaurant	Playground	Food Truck	Train Station	Performing Arts Venue	Yoga Studio	Dim Sum Restaurant
4	Bloomfield	Staten Island	40.612604	-74.178200	0.0	Hotel	Tea Room	Cocktail Bar	Video Store	Candy Store	Rental Car Location	Italian Restaurant
5	Bulls Head	Staten Island	40.607048	-74.162088	2.0	Pharmacy	Diner	Baseball Field	Playground	Café	Bank	Sandwich Place

# Clustered Neighborhoods:



# Conclusion:

---

After performing all 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.

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.





# THANK YOU!

---

Ahmed Hamza Bin Asif

Email  
hak47660@gmail.com