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.

Read New York Crime Data CSV file

```
crimeDF = pd.read_csv("NYPD_Arrest_Data__Year_to_Date_.csv")
crimeDF.shape
```

```
(140413, 19)
```

```
crimeDF.head()
```

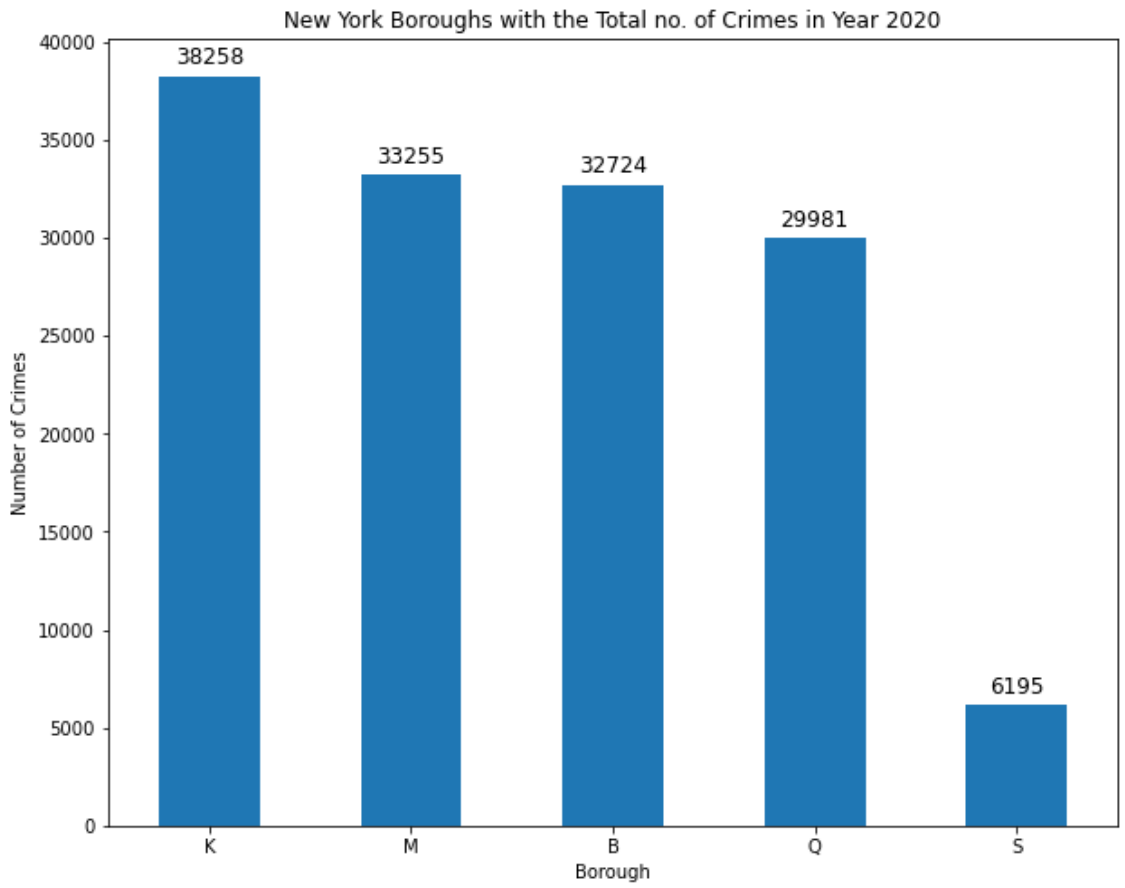
	ARREST_KEY	ARREST_DATE	PD_CD	PD_DESC	KY_CD	OFNS_DESC	LAW_CODE	LAW_CAT_CD	ARREST_BORO	ARREST_PRECIN
0	222243929	12/25/2020	105.0	STRANGULATION 1ST	106.0	FELONY ASSAULT	PL 1211200	F	B	
1	222115778	12/22/2020	397.0	ROBBERY,OPEN AREA UNCLASSIFIED	105.0	ROBBERY	PL 1601001	F	M	
2	222090516	12/21/2020	105.0	STRANGULATION 1ST	106.0	FELONY ASSAULT	PL 1211200	F	Q	1
3	221653577	12/10/2020	105.0	STRANGULATION 1ST	106.0	FELONY ASSAULT	PL 1211200	F	Q	1
4	221599602	12/9/2020	729.0	FORGERY,ETC.,UNCLASSIFIED-FELO	113.0	FORGERY	PL 1702500	F	S	1

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.

```
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

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.

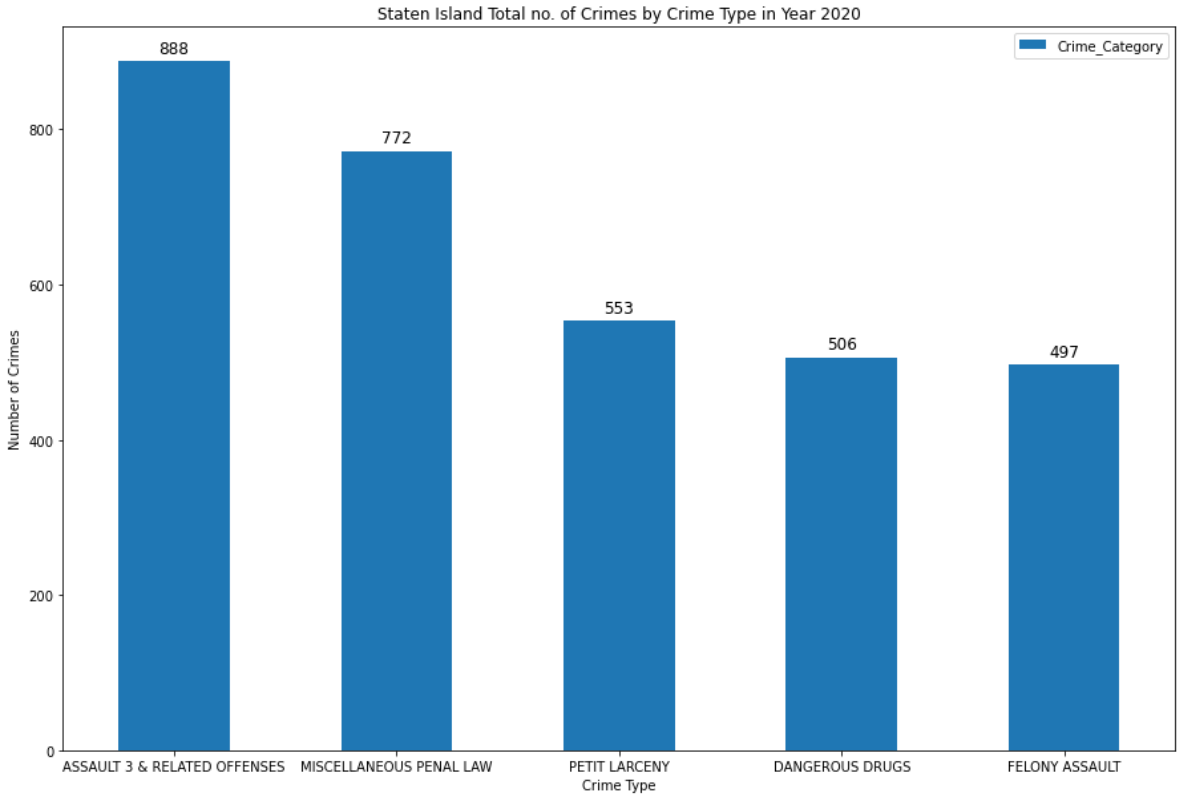
Only Staten Island Crime Record was selected from the dataframes and plotted afterwards.

Staten Island Crime Analysis

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



Real Estate Analysis:

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

```
estDF = pd.read_csv('rollingsales_statenisland.csv')
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	315,000	10/2/2020
6	NaN	A1	118 ELMBANK STREET	NaN	...	1.0	0.0	1.0	6,200	2,580	1975.0	1	A1	450,000	6/24/2020
31	NaN	A1	7 SANDGAP STREET	NaN	...	1.0	0.0	1.0	2,500	1,222	1925.0	1	A2	525,000	7/2/2020
83	NaN	A1	7 SEGUINE PLACE	NaN	...	1.0	0.0	1.0	3,989	2,300	2004.0	1	A1	720,000	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

The dataset was further cleaned. The saleprice column had all values in string format

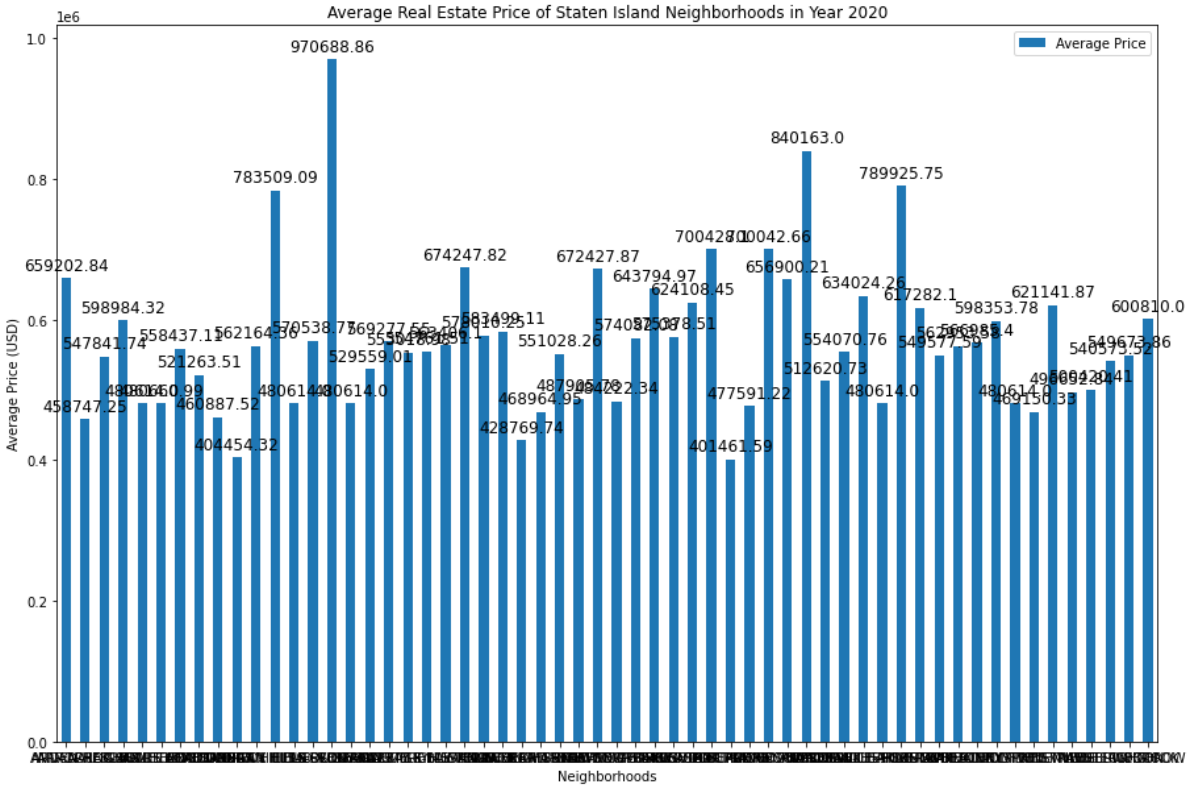
```
estDF['SALE PRICE'] = estDF['SALE PRICE'].str.replace(',', '').astype(float)
estDF['SALE PRICE'] = estDF['SALE PRICE'].astype(int)
estDF.head()
```

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

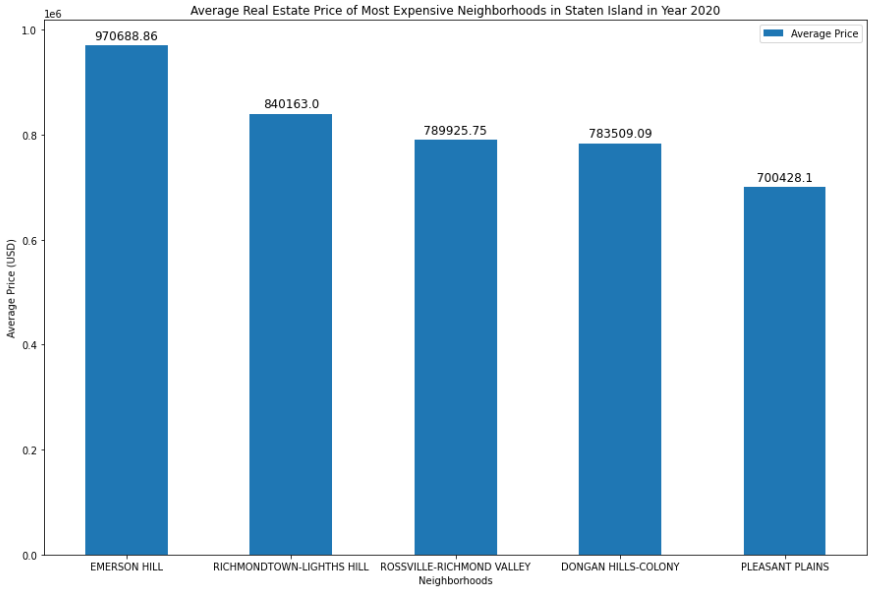
The mean value was needed to get the average price of real estate in every neighborhood which was further used to plot.

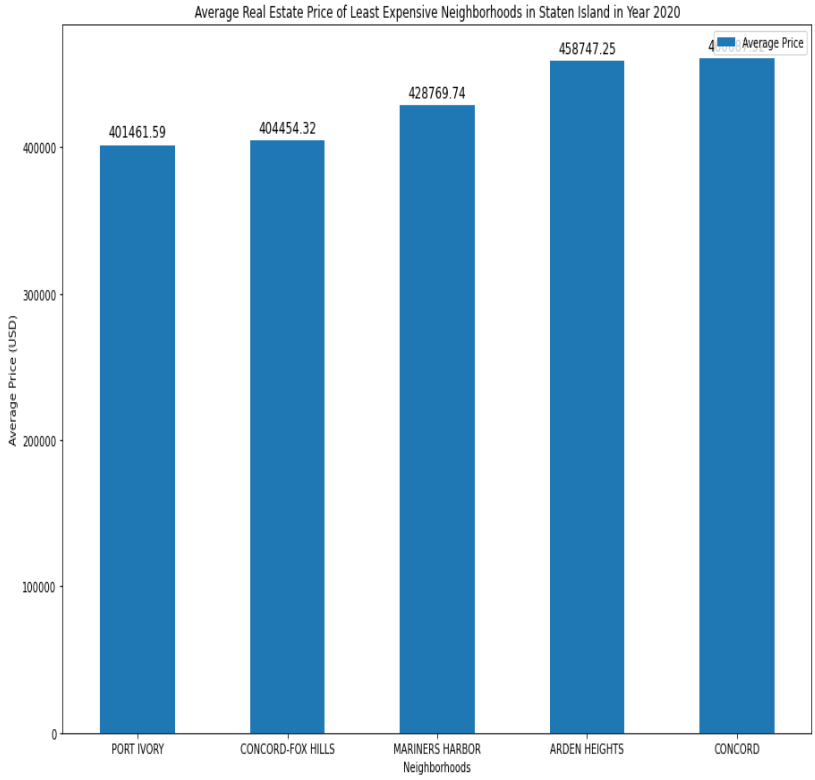
```
mod = [4,12,43,50]
for i in mod:
    estLSTSAVG[i] = int(estMean)

salePrDF = pd.DataFrame.from_records(combL)
salePrDF = pd.DataFrame({'Average Price': estLSTSAVG},
                        index = estLSTND)
salePrDF.head()
```

	Average Price
ANNADALE	659202.844444
ARDEN HEIGHTS	458747.246094
ARROCHAR	547841.739130
ARROCHAR-SHORE ACRES	598984.315789
BLOOMFIELD	480614.000000



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.

```
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

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

```
print("Geographic Location of Schools in Staten Island, Newyork City, USA")
StatenIslandSchoolMap = folium.Map(location=[latitude, longitude], zoom_start=11)
for latitude, longitude, borough, school in zip(schoolDF2['Latitude'], schoolDF2['Longitude'], schoolDF2['Borough'], schoolDF2['
    label = '{}, {}'.format(school, borough)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [latitude, longitude],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        ).add_to(StatenIslandSchoolMap)

StatenIslandSchoolMap
```

Geographic Location of Schools in Staten Island, Newyork City, USA

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.

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

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
```

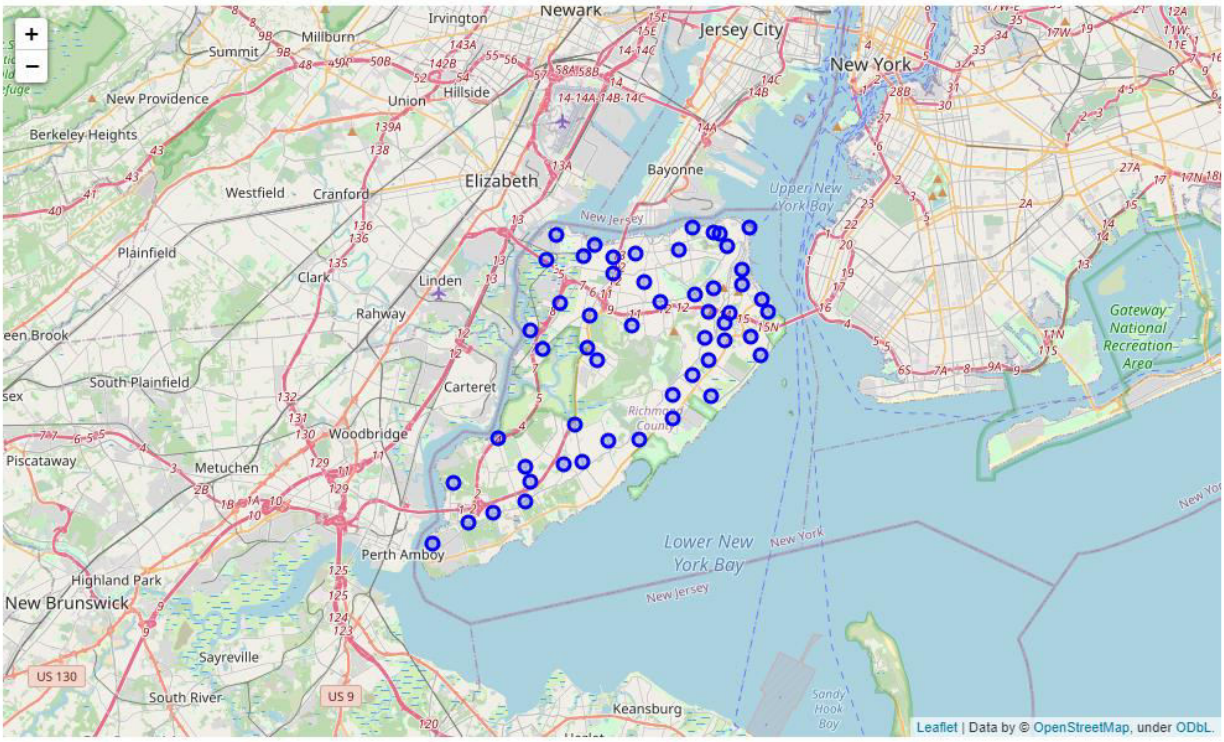
	City	Hospital Type	Emergency Services	Meets criteria for meaningful use of EHRs	Hospital overall rating	Safety of care national comparison	Patient experience national comparison	Effectiveness of care national comparison	Efficient use of medical imaging national comparison
2833	STATEN ISLAND	Acute Care 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

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**.

```
df = {'Neighborhood': nbrSTI,'Borough':nbrSTIdf,'Latitude': Latitude,'Longitude':Longitude}
df1 = pd.DataFrame(data=df, columns=['Neighborhood', 'Borough', 'Latitude', 'Longitude'], index=None)
df1.head()
```

	Neighborhood	Borough	Latitude	Longitude
0	Annadale	Staten Island	40.544550	-74.176532
1	Arlington	Staten Island	40.632326	-74.165144
2	Arrochar	Staten Island	40.598438	-74.072641
3	Bay Terrace	Staten Island	40.555278	-74.134167
4	Bloomfield	Staten Island	40.612604	-74.178200



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

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Annadale	40.54455	-74.176532	Annadale Diner	40.542079	-74.177325	Diner
1	Annadale	40.54455	-74.176532	Annadale Terrace	40.542555	-74.177187	Restaurant
2	Annadale	40.54455	-74.176532	Il Sogno	40.541286	-74.178489	Restaurant
3	Annadale	40.54455	-74.176532	MTA SIR - Annadale	40.540482	-74.178185	Train Station
4	Annadale	40.54455	-74.176532	MTA Bus - Annadale Rd & Arden Av (S55)	40.544326	-74.176667	Bus Stop

The most common venues of all neighborhoods will be returned.

In [265]:
```
def return_most_common_venues(row, num_top_venues):
    row_categories = row.iloc[1:]
    row_categories_sorted = row_categories.sort_values(ascending=False)

    return row_categories_sorted.index.values[0:num_top_venues]
```

In [266]:
```
num_top_venues = 10
indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['Neighborhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

# create a new dataframe
mergedVen = pd.DataFrame(columns=columns)
mergedVen['Neighborhood'] = STI_grouped['Neighborhood']

for ind in np.arange(STI_grouped.shape[0]):
    mergedVen.iloc[ind, 1:] = return_most_common_venues(STI_grouped.iloc[ind, :], num_top_venues)
mergedVen
```

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:

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 previously mentioned 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.

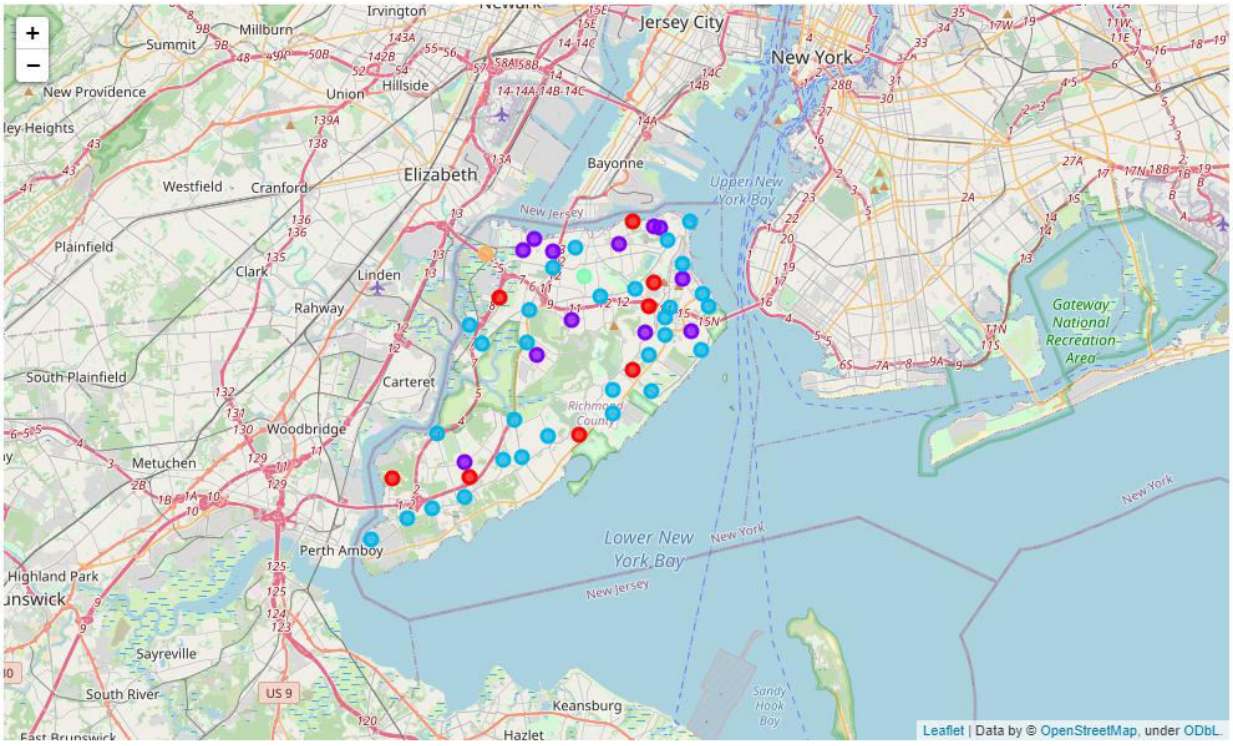
```
# Using K-Means to cluster neighborhood into 5 clusters
kClusters = 5
STI_grouped_clustering = STI_grouped.drop('Neighborhood', 1)
kmeans = KMeans(n_clusters=kClusters, random_state=0).fit(STI_grouped_clustering)
kmeans.labels_[0:10]
```

```
array([2, 1, 1, 0, 0, 2, 2, 0, 2, 1])
```

```
#neighborhoods_venues_sorted = neighborhoods_venues_sorted.drop('Cluster Labels',1)
mergedVen.insert(0, 'Cluster Labels', kmeans.labels_)
```

```
mergedVen
```

	Cluster Labels	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 Mo Comm Veni
0	2	Annadale	Pizza Place	Business Service	American Restaurant	Restaurant	Diner	Bar	Bakery	Bus Stop	Food	Deli / Bode
1	1	Arlington	Bus Stop	Construction & Landscaping	American Restaurant	Deli / Bodega	Yoga Studio	Discount Store	Fast Food Restaurant	Farmers Market	Event Space	Event Servi
2	1	Arrochar	Park	Bus Stop	Bagel Shop	Cosmetics Shop	Pizza Place	Deli / Bodega	Nail Salon	Yoga Studio	Event Space	Event Servi
3	0	Bay Terrace	American Restaurant	Playground	Food Truck	Train Station	Performing Arts Venue	Yoga Studio	Dim Sum Restaurant	Event Space	Event Service	Dry Clean
4	0	Bloomfield	Hotel	Tea Room	Cocktail Bar	Video Store	Candy Store	Rental Car Location	Italian Restaurant	Spa	Department Store	Dim Su Restaura
5	2	Bulls Head	Pharmacy	Diner	Baseball Field	Playground	Café	Bank	Sandwich Place	Bus Stop	Deli / Bodega	Tattoo Par



Results

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

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

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

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

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

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. 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 favor 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.

References:

https://data.cityofnewyork.us/Public-Safety/NYPD-Arrest-Data-Year-to-Date-/uip8-fykc

https://www1.nyc.gov/site/finance/taxes/property-rolling-sales-data.page

https://www.kaggle.com/carlosaguayo/usa-hospitals/version/1

https://www.kaggle.com/center-for-medicare-and-medicaid/hospital-ratings

https://www.kaggle.com/carlosaguayo/usa-public-schools?select=Public_Schools.csv

https://en.wikipedia.org/wiki/List_of_Staten_Island_neighborhoods