**Capstone Project: Moving to New York City**

**1. Business problem**

*“Start spreading the news, I'm leaving today. I want to be a part of it…”* New York City – the city that never sleeps – is still the one destination for many people who want to follow their dreams to succeed. But the city is huge and you would like to move to the area that will meet your needs and wishes. This may mean that you would like to live close enough to the popular spots to improve your chances of success, but at the same time be able to afford the place to live in. You would also like to feel safe and live in the low-crime area of the city. How do you find out which part of the city is the right for you? I will analyse some attributes to help with this problem, such as housing cost, density of social places, and crime rate in the areas of the New York City, to help the people who want to follow their dreams and move to New York City in deciding where to look for the place to live.

**2. Data**

To perform the analysis, the following data will be used:

1. **Rent prices**

* I will use the data available at streeteasy.com to analyse the housing prices and create visualisation comparing the housing prices in different boroughs and neighbourhoods of the New York City. The information on this website offers historical monthly data by borough or neighbourhood on housing price index, asking sale price, asking rent and other data. This information will help the reader in determining the neighbourhoods of the New Your City that would fit the reader’s budget for housing. Link: <https://streeteasy.com/blog/data-dashboard/?agg=Total&metric=Inventory&type=Rentals&bedrooms=Any%20Bedrooms&property=Any%20Property%20Type&minDate=2010-01-01&maxDate=2020-10-01&area=Flatiron,Brooklyn%20Heights>

1. **Crime rates**

* I will use the data at data.cityofnewyork.us site to analyse the crime rates in different areas of New York. I downloaded a .csv file NYPD\_Complaint\_Data\_Historic.csv on historic data including all valid felony, misdemeanour and violation crimes reported to the New York City Police Department for all complete quarters of year 2019. This will help the reader to decide whether the neighbourhood that fits the reader’s budget in terms of housing prices is safe enough. Link: <https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i>
* I downloaded GeoJSON file with Boundaries of Boroughs for New York City. Link: <https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm>

1. **Social places**

* I will use the New York neighbourhood data provided in this course for segmenting and clustering neighbourhoods in New York City containing boroughs and neighbourhoods with the latitude and longitude coordinates, in order to segment the neighbourhoods and explore them.
* I will use Forsquare API to get the most common venues of given Borough of New York City. This will show the most popular places of each neighbourhood of the New York City, which will assist the reader in deciding which areas of the New York City would be the best match to the reader’s needs.

**3. Methodology**

1. **Rent prices**

First, I imported the libraries needed for the analysis.

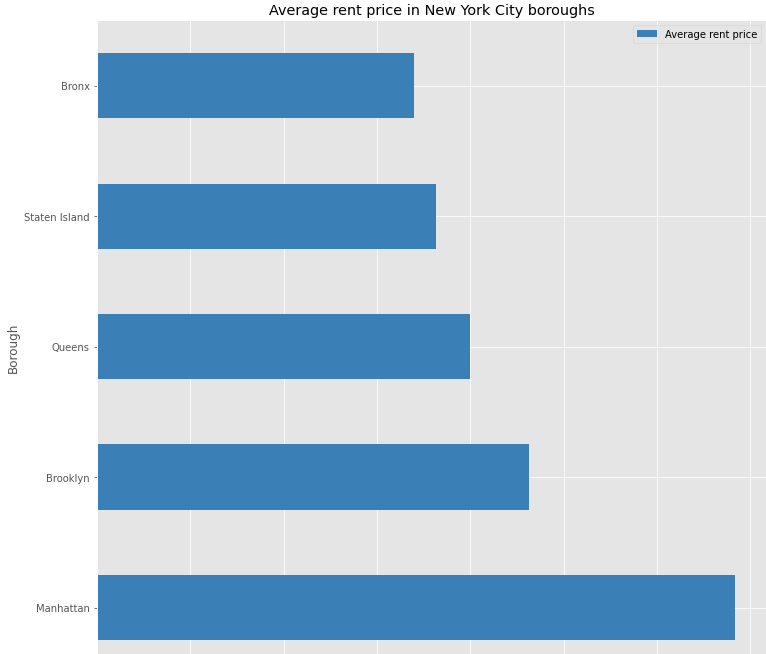
I downloaded a .csv file medianAskingRent\_All.csv on Median asking rent, including historical monthly data on asking rent by borough and neighbourhood on rentals for all types of apartment (studio, one bedroom, two bedroom and 3+ bedrooms). I used the prices of rent versus prices of sales, as I assumed that a person newly moved to New York City would in the first instance rent an apartment versus buying it. I used the rental price data for all types of apartments versus choosing a particular size of apartment, assuming that in a long run, the prices of particular size of apartment would follow the average price of all types of apartments.

I calculated the average asking rent for each area in the given period. I deleted the data for submarkets and deleted the rows that had no data in the given period. As a result, data on average rent price was available for 182 neighbourhoods.

I set the areaName as index.

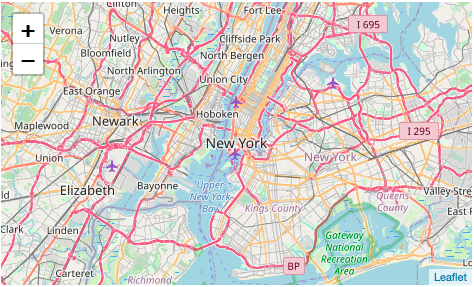


Using Matplotlib, I visualised the data in horizontal bar plot for each available neighbourhood and for each borough as presented below.

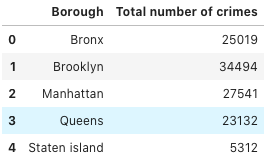


1. **Crime rates**

First, I imported the libraries needed for the analysis. I installed Folium which will be used to create the maps.



I downloaded the dataset and cleaned it. I grouped the data by Borough with the total number of crimes occurred in each borough in the presented period.



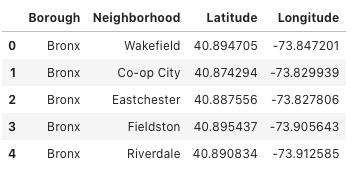
I downloaded GeoJSON file and reviewed the data. I changed the font case in my dataset to match the borough names in the GeoJSON file. Based on this I created a choropleth map using the total number of crimes per borough.



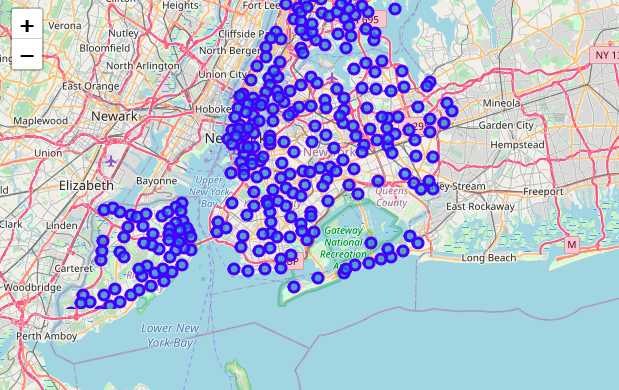
1. **Social places**

First, I imported all the libraries needed for the analysis.

I loaded the data in the .json file and transformed it into a pandas dataframe with the following columns: Borough, Neighbourhood, Latitude and Longitude. This resulted in a dataframe with 5 boroughs and 306 neighbourhoods.



I used geopy library to get the latitude and longitude values of New York City. The geographical coordinate of New York City are 40.7127281, -74.0060152. Based on this I created a map of New York City with neighbourhoods superimposed on top, using Folium.



Utilizing Foursquare API to explore neighbourhoods and segment them, I sliced the original data and created a new dataframe for each borough.



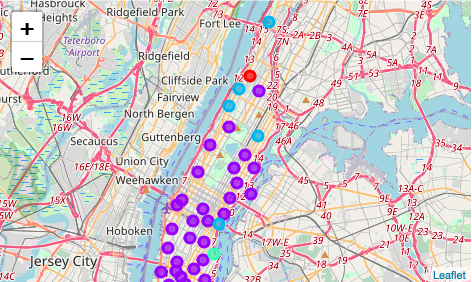
I obtained geographical coordinates of each borough and visualised each borough and the neighbourhoods in it.



I obtained the top 100 venues in each neighbourhood within a radius of 500 meters and cleaned the json and structured it into pandas dataframe and displayed the top 10 venues for each neighbourhood.



I clustered each borough into 5 clusters and created a new dataframe for each borough that includes the cluster as well as the top 10 venues for each neighbourhood. I visualized the clusters in a map for each borough.



I performed an initial examination of each cluster of each borough to determine some discriminating venue categories that distinguish each cluster. However, the detailed analysis of each cluster is left to the reader who will analyse the clusters based on their personal preferences.

**4. Results**

As expected, rent prices are the highest in Manhattan, followed by Brooklyn and Queens. The lowest rent prices are in Bronx. The highest crime rate is in Brooklyn, followed by Manhattan and Bronx. The lowest crime rate is in Staten Island.

While the rent is high in the more popular boroughs of the New York City, the crime rates are also the highest in these boroughs, as expected. Bronx has the lowest average rent prices, and lower crime rates than Brooklyn and Manhattan.

Depending on the reader’s personal preference, the analysis of the most popular social venues each borough and neighbourhood may factor in the decision where to move. The person preferring outdoor life may prefer location close to parks. The person preferring night life may prefer location close to bars and restaurants. The person with pets may prefer location close to pet stores, and so on.

**5. Discussion**

The data on rent prices seem to be missing data for several neighbourhoods. With more data it would be interesting to analyse the rent price per each neighbourhood versus borough, to get more details on the preferred areas. The same applies for crime rates, which were analysed using boroughs versus neighbourhoods. More data would allow a more detailed analysis.

Depending on the reader’s personal preferences, a more detailed analysis of preferred social venues could be made . Moreover, additional data could be analysed depending on the personal preferences, such as school districts for families for example.

**6. Conclusion**

The presented analysis should provide a good initial insight of the data of the New York City. With more data and additional analysis, the preferred locations could be narrowed down to a greater extent.