Capstone Project: Finding a location

APPLIED DATA SCIENCE CAPSTONE BY IBM/COURSERA

Table of Contents

Introduction

- In this project we will try to find an optimal location to rent a studio apartment in New York City. Specifically, this report will be addressed to **those students who need to rent a room to access their classes at Columbia University in New York City**.

- As there are many studio apartments in New York City, we will try to detect places considering 3 factors; crime rate, rent cost and finally, proximity to places of preference. We would also prefer locations as close as possible to the university, assuming the first three conditions are met.

- We will use data science to generate some more promising neighborhoods based on this criterion. The advantages of each area will be clearly expressed so that those interested can choose the best possible final location

Data

Based on definition of our problem, factors that will influence our decission are:

- Number of crimes committed in each county of New York City
- Average rental cost of a studio apartment
- Proximity to places of preference

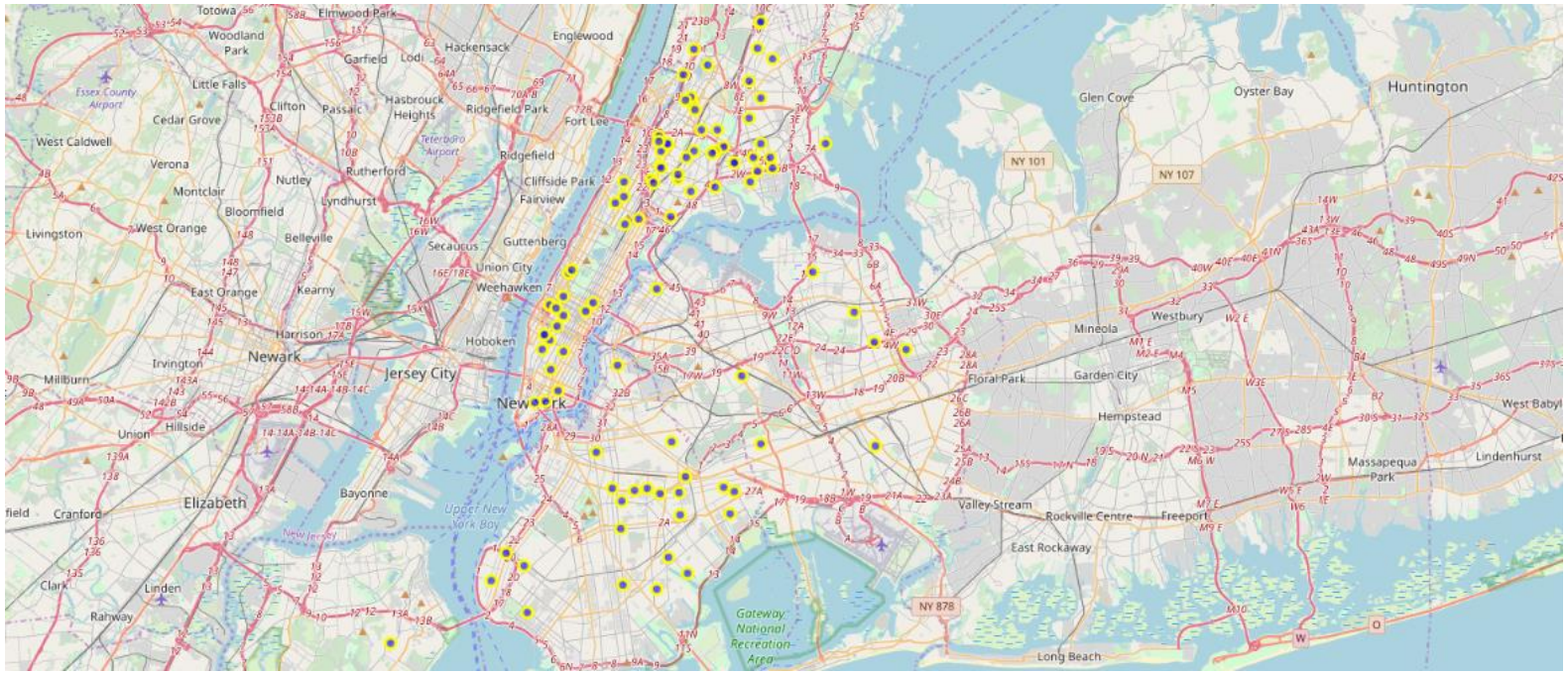
We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

Following data sources will be needed to extract/generate the required information:

- Number of total crimes per county, for which the database of **Open Data** will be used "NYPD Complaint Data Current (Year To Date)" 'https://data.cityofnewyork.us/api/views/5uac-w243/rows.csv'
- The exact middle asking rent among all rental listings available on **StreetEasy** at any point during the month/quarter/year. In general, median values are more accurate than average values, which may be skewed by price outliers (a few rentals that are extremely expensive or extremely inexpensive). https://streeteasy-market-data-download.s3.amazonaws.com/rentals/Studio/medianAskingRent_Studio.zip
- Number of preference places and their type and location in every neighborhood will be obtained using **Foursquare API**

Methodology: Crimes in NYC

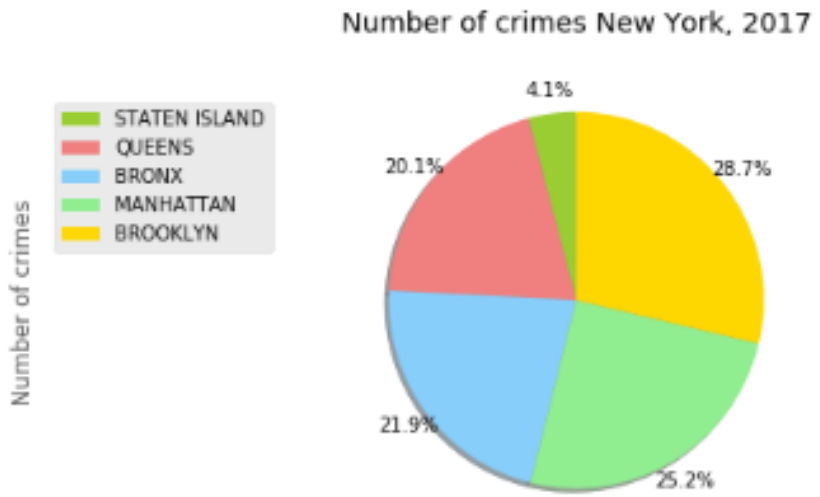
Borough	Id	Latitude	Longitude
BRONX	100994	100994	100994
BROOKLYN	132445	132445	132445
MANHATTAN	116352	116352	116352
QUEENS	92575	92575	92575
STATEN ISLAND	19019	19019	19019

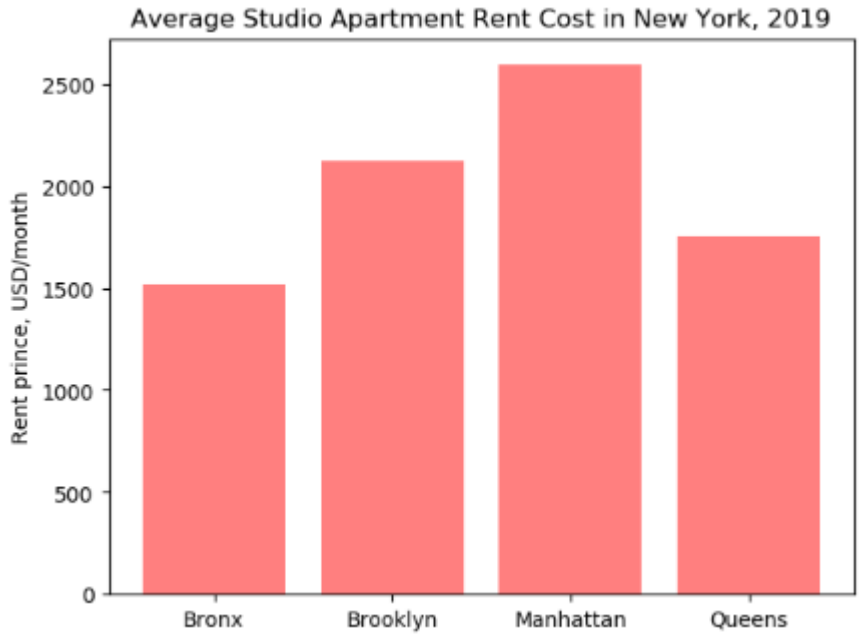


Methodology: Apartment rental prices

Borough	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Average
Bronx	1420.5	1500.0	1462.5	1436.5	1544.0	1500.0	1544.0	1562.5	1575.0	1556.5	1550.0	1550.0	1516.8
Brooklyn	2046.4	2044.8	2064.4	2091.0	2155.4	2150.2	2162.6	2159.2	2166.7	2149.7	2121.1	2160.4	2122.7
Manhattan	2514.7	2551.8	2531.7	2563.5	2579.5	2600.1	2615.3	2599.6	2619.2	2649.6	2638.4	2639.2	2591.9
Queens	1681.6	1687.6	1692.5	1745.4	1747.0	1795.0	1783.7	1783.5	1794.9	1782.0	1768.8	1778.5	1753.4

Analysis: Visualization





Analysis: Foursquare

Cluster Neighborhoods

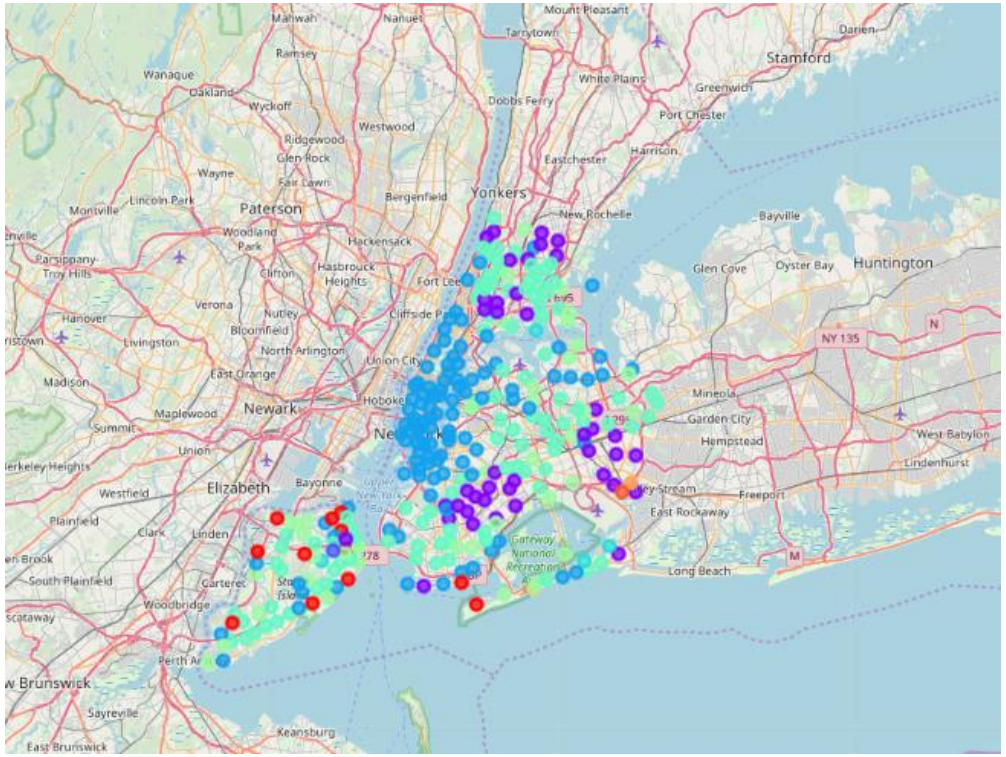
Run *k*-means to cluster the neighborhood into 10 clusters

```
# set number of clusters
kclusters = 10

nyc_grouped_clustering = nyc_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(nyc_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```



Discussion

- After analyzing the data corresponding to each study factor, the following can be commented:
- The 5 boroughs analyzed show crime rates, of which Staten Island stands out, with an indicator of less than 5%. The rest of the boroughs have a similar percentage of crimes.
- Considering that Staten Island is far from the study point, it can be discarded from the analysis.
- Based on the rental cost of a study department, the lowest costs can be found in Bronx and Queens (1518.6 and 1753.4 USD / month respectively).
- Adding the factor of the availability of places to carry out different activities, it is possible to observe that in Manhattan it would be probable to find one kind of activities (bars), while the rest of the options offer greater type of variety, between food and parks.
- It could be expected that the best decision for a student looking to rent an apartment and attend classes at Columbia University is to live in Brooklyn.

Conclusion

It can be concluded that this method of analysis allows to visualize the existing options when making a decision.

It is important to mention that the quality of the dataset have a very important role in the output information.

In these times of rapid changes, a better approximation could be obtained if the information corresponded to at least the last 2 years. External factors such as personal tastes and particular opinions were not considered in this analysis, so the result will depend on the preferences of each student.