Battle of Boroughs and Neighborhoods

Osman Muglu

Introduction

- **Background:**
  - New York City (NYC) is one of the largest and wealthiest metropolises in the world
  - Before making a move to NYC it is necessary to know in which neighborhood you want to live

- **Problem/Research questions:**
  - Aiming to select safest borough
  - Explore neighborhoods of safest borough and cluster them

  (1) What is the safest borough in NYC?

  (2)What is the best neighborhood based on venues?

- **Interest**
  - People who are considering moving to NYC will be interested to identify the safest borough and the best neighborhood based on venues.

2. Data Acquisition and Cleaning

- **Data acquisition:**

- The recourses:
- The crime data set: of this project, it contains the crimes per borough in NYC [1]

- The list of Neighborhoods of Staten Island [2]

- The most common venues of safest borough of NYC [3]

2. Data Acquisition and Cleaning

- **Data cleaning:**

- NYC crime dataset:
  - Selected columns: crime description borough
  - Values are counted and pivoted in new table
  - Extra column created with total amount of crimes

- NYC Safest borough:
  - Created from scratch using list from Wikipedia
  - Filled the frame with data obtained by Google Maps API

- Venues data set:
  - Retrieving the data through Foursquare API

3. Methodology

- **Packages:**

  - **requests** : Handle http requests
  - **pandas** : To collect and manipulate data in JSON and HTMl and then data Analysis
  - **matplotlib** : Detailing the generated maps
  - **numpy**: to handle arrays
  - **folium** : Generating maps of London and Paris
  - **sklearn** : To import Kmeans which is the machine learning model that we are using.

3. Methodology

- Visualization by **folium**
- Exploration of neighborhoods by **Foursquare API**
- Dealing with categorical data by **One Hot Coding**
- Model Building by **K-means clustering**

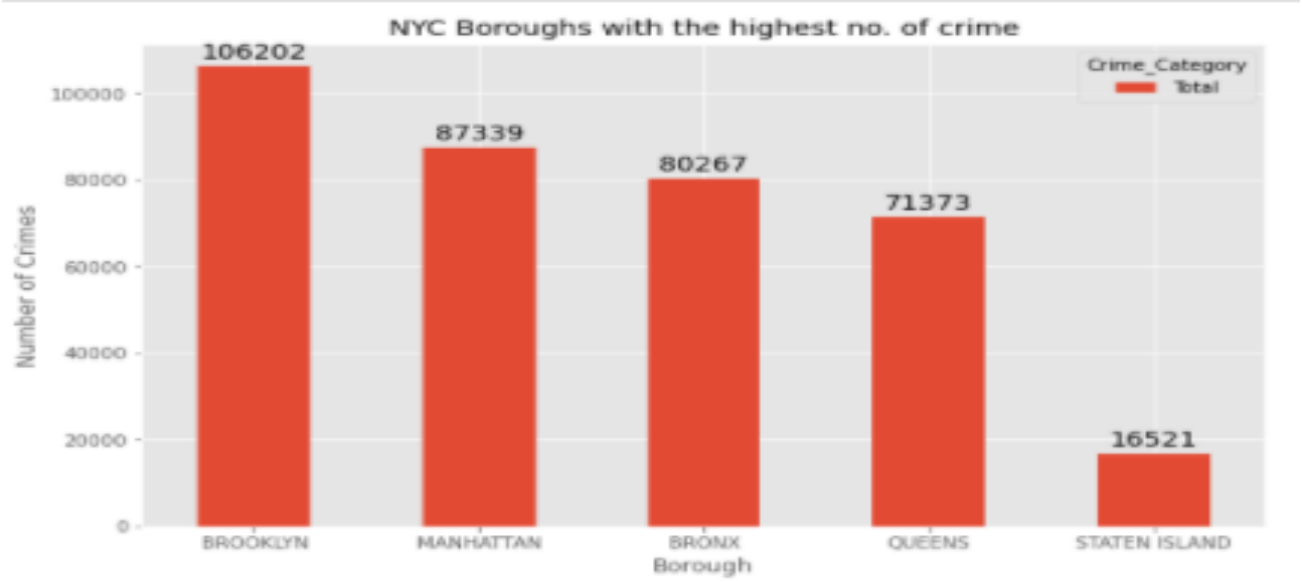
4. Results – Data Analysis

- **Statistical summary of crimes:**

ıe Category	ADMINISTRATIVE CODE	ADMINISTRATIVE CODES	AGRICULTURE & MRKTS LAW-UNCLASSIFIED	ALCOHOLIC BEVERAGE CONTROL LAW	ANTICIPATORY OFFENSES	ARSON	ASSAULT 3 & RELATED OFFENSES	BURGLAR'S TOOLS	BURGLARY	CHILD ABANDONMENT/NON SUPPORT	...	PROSTITUTION & RELATED OFFENSES	RAPE	ROBBERY	SEX CRIMES	THEFT OF SERVICES	THEFT-FRAUD	UNAUTHORIZED USE OF A VEHICLE	UNLAWFUL POSS. WEAP. ON SCHOOL	VEHICLE AND TRAFFIC LAWS	Total
count	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	...	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.00000	5.00000
mean	151.800000	0.200000	21.600000	15.000000	0.400000	122.600000	7995.400000	41.000000	1927.000000	3.200000	...	20.400000	221.400000	2294.400000	947.800000	65.400000	740.600000	267.600000	0.400000	990.00000	72340.40000
std	108.936679	0.447214	9.659193	16.643317	0.894427	54.123008	3964.934716	28.714108	1006.508569	2.167948	...	27.327642	112.251058	1269.217397	514.771017	79.415364	352.273757	205.563372	0.547723	574.44669	33731.06218
min	37.000000	0.000000	6.000000	0.000000	0.000000	36.000000	1661.000000	2.000000	364.000000	0.000000	...	3.000000	36.000000	298.000000	141.000000	0.000000	197.000000	85.000000	0.000000	133.00000	16521.00000
25%	64.000000	0.000000	21.000000	2.000000	0.000000	109.000000	7903.000000	27.000000	1824.000000	3.000000	...	5.000000	220.000000	2185.000000	868.000000	9.000000	644.000000	129.000000	0.000000	815.00000	71373.00000
50%	170.000000	0.000000	22.000000	7.000000	0.000000	135.000000	7942.000000	45.000000	1947.000000	3.000000	...	7.000000	249.000000	2237.000000	938.000000	14.000000	780.000000	142.000000	0.000000	1030.00000	80267.00000
75%	176.000000	0.000000	28.000000	32.000000	0.000000	163.000000	10300.000000	52.000000	2401.000000	4.000000	...	19.000000	265.000000	3149.000000	1376.000000	144.000000	972.000000	470.000000	1.000000	1311.00000	87339.00000
max	312.000000	1.000000	31.000000	34.000000	2.000000	170.000000	12171.000000	79.000000	3099.000000	6.000000	...	68.000000	337.000000	3603.000000	1416.000000	160.000000	1110.000000	512.000000	1.000000	1661.00000	106202.00000

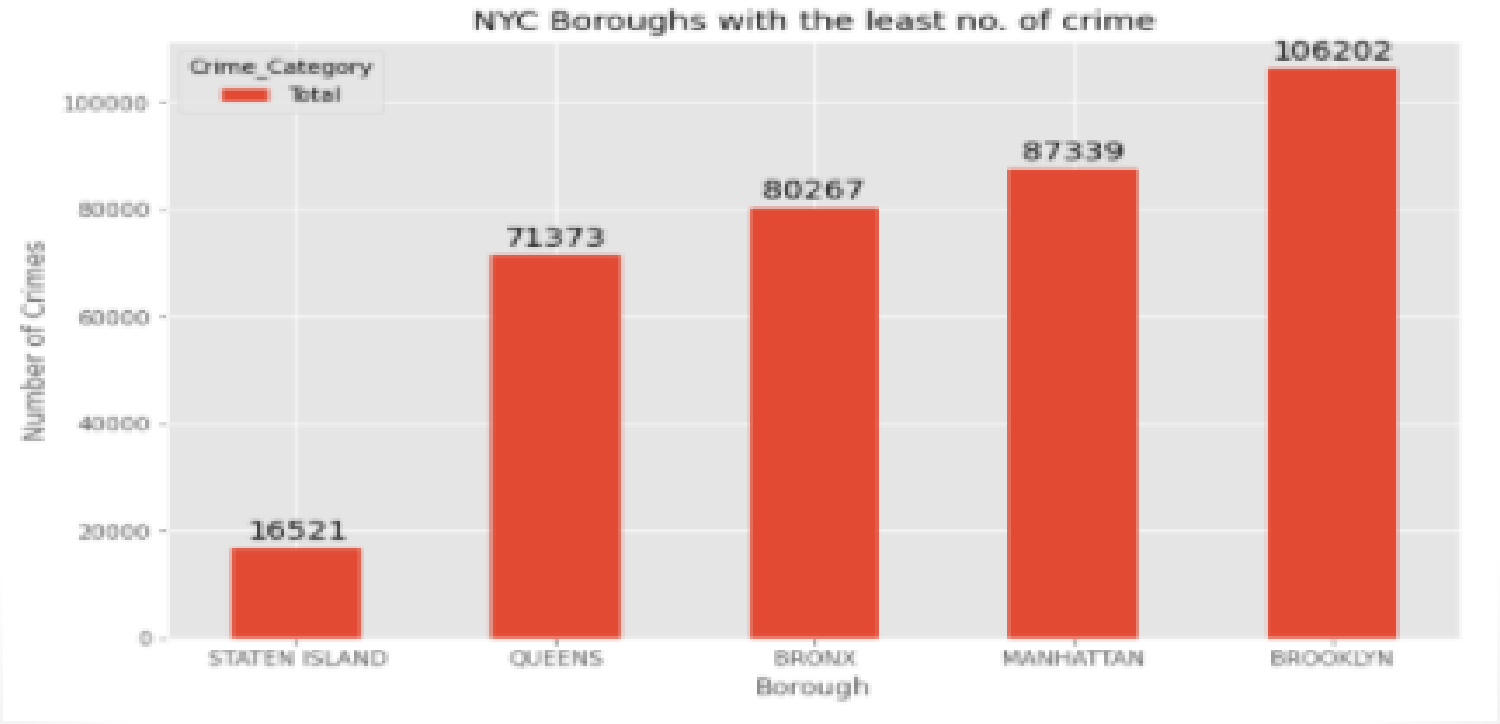
4. Results – Data Analysis

- **Boroughs with the highest crime rates:**
- When we compare the five boroughs we see which one has the highest crime rate during 2016. Brooklyn has the highest crimes recorded followed by Manhattan, Bronx, Queens and Staten Island. Brooklyn has a significantly higher crime rate than the other 4 boroughs.



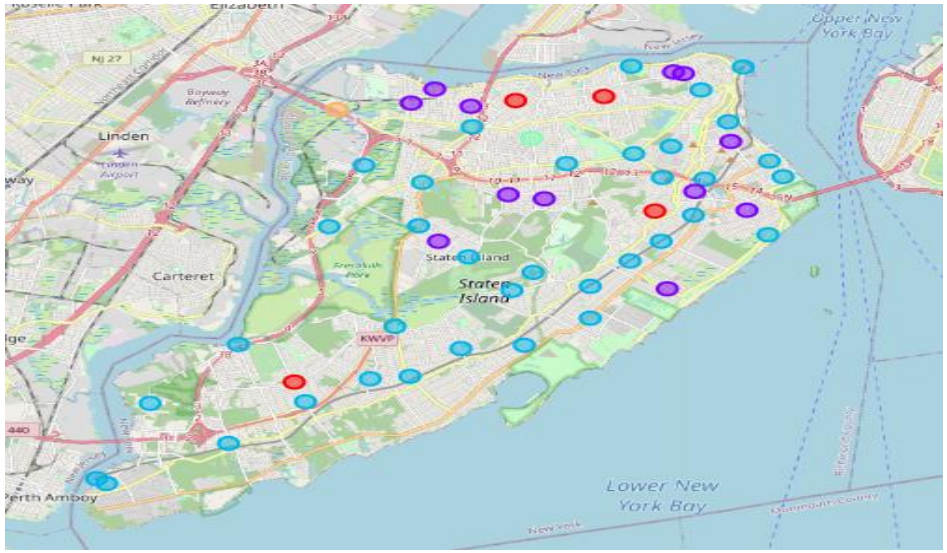
4. Results – Data Analysis

- **Boroughs with the lowest crime rates:**
- We already could the borough with the lowest crime rates in the previous figure. However, we visualized a new chart but in a new order. In this new figure we see that Staten Island borough has the lowest crime rates .



4. Results – Data Analysis

- **Clustering:**
- After running the K-mean cluster we can visualize a map of Staten Island with the clustered neighborhoods. Each cluster has it unique color for the ease of the presentation. We see that the majority of the neighborhoods fall is the blue cluster, which is the third cluster. The colors of the other clusters are as follows: the first cluster has a red color, the second is purple, the fourth cluster is green and the last cluster is yellow.



4. Results – Data Analysis

- **Top most common venues per cluster:**

- **Cluster 1**: bus stops, supermarkets and restaurant serving food from different cultural cuisines.
- **Cluster 2**: Bus stops and bodega's
- **Cluster 3**: pizza places, cafes, restaurants and parks. (biggest)
- **Cluster 4**: arcade, yoga studio and discount store (one neighborhood)
- **Cluster 5**: harbor, toll plaza and Yoga studio (one neighborhood)

5. Discussion

- *What is the safest borough in NYC?*
  - Safest borough: Staten Island
- What is the best neoghborhoods based on venues?
  - **A subjective choise:**
  - **Cluster 1**: for families who are looking for quiter places
  - **Cluster 2**: Good connectivity
  - **Cluster 3**: Offers restaurant and other places where food is served

6. Conclusion

- This project helps people find out which borough in NYC is the safest. It also gives a picture of the neighborhoods, based on venues, within the safest borough.

- Future projects can focus on adding more data to this research. By example including factors such as average house prices or cost of living.

7. References

- [1] NYC Crime data set 2016
- [2] Neighborhoods of Staten Island
- [3] Foursquare API