Coursera Capstone Project

THE BATTLE OF NEIGHBORHOODS

-ANKIT MITTAL

1. INTRODUCTION

- **Background:** Safety is at the to priority when moving to a new location. If you don't feel safe in your own home and locality, you are not going to enjoy your living.
- **Problem:** This project aims to select the safest borough in London based on the total crimes, explore the neighborhoods of that borough to find the 10 most common venues in each neighborhood and finally cluster the neighborhoods using k-mean clustering. This will let you now the safest borough in London and allows you to explore its neighborhood and common venues around each neighborhood to select your new residing location.

2. DATA ACQUISITION AND CLEANING

- **Data Acquisition:** The data acquired for this project is a combination of data from three sources
  - The first data source of the project uses a London crime data that shows the crime per borough in London.
  - The second source of data is scraped from a Wikipedia page that contains the list of London boroughs
  - The third data source is the list of Neighborhoods in the Royal Borough of Kingston upon Thames as found on a Wikipedia Page.

- **Data Cleaning:** The data preparation for each of the three sources of data is done separately.
  - From the London crime data, the crimes during the most recent year (2016) are only selected. The major categories of crime are pivoted to get the total crimes per the boroughs for each major category.
  - The second data is scraped from a Wikipedia page using the Beautiful Soup library in python. Using this library, we can extract the data in the tabular format as shown in the website.
  - The two datasets are merged on the Borough names to form a new dataset that combines the necessary information in one dataset. The purpose of this dataset is to visualize the crime rates in each borough and identify the borough with the least crimes recorded during the year 2016.
  - After visualizing the crime in each borough, we can find the borough with the lowest crime rate. The third source of data is acquired from the list of neighborhoods in the safest borough on Wikipedia. The coordinates of the neighborhoods is be obtained using Google Maps API geocoding to get the final dataset
  - The new dataset is used to generate the 10 most common venues for each neighborhood using the Foursquare API, finally using k means clustering algorithm to cluster similar neighborhoods together.

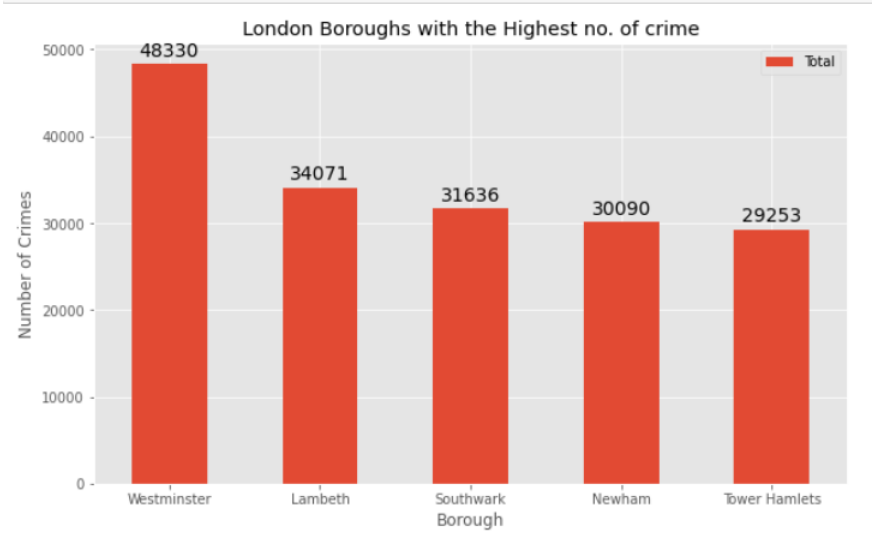
3. Methodology

- **Exploratory Data Analysis**
  - Statistical Summary of Crime Data

	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
count	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000
mean	2069.242424	1941.545455	1179.212121	479.060606	682.666667	8913.121212	7041.848485	22306.696970
std	737.448644	625.207070	586.406416	223.298698	441.425366	4620.565054	2513.601551	8828.228749
min	2.000000	2.000000	10.000000	6.000000	4.000000	129.000000	25.000000	178.000000
25%	1531.000000	1650.000000	743.000000	378.000000	377.000000	5919.000000	5936.000000	16903.000000
50%	2071.000000	1989.000000	1063.000000	490.000000	599.000000	8925.000000	7409.000000	22730.000000
75%	2631.000000	2351.000000	1617.000000	551.000000	936.000000	10789.000000	8832.000000	27174.000000
max	3402.000000	3219.000000	2738.000000	1305.000000	1822.000000	27520.000000	10834.000000	48330.000000

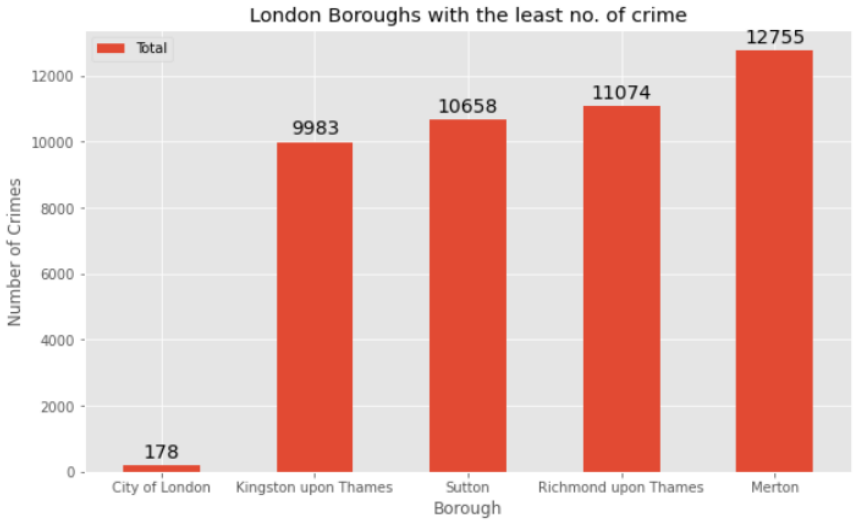
The count for each of the major categories of crime returns the value 33 which is the number of London boroughs. 'Theft and Handling' is the highest reported crime during the year 2016 followed by 'Violence against the person', 'Criminal damage'. The lowest recorded crimes are 'Drugs', 'Robbery' and 'Other Notifiable offenses.

- Boroughs with High Crime rates



Comparing five boroughs with the highest crime rate during the year 2016 it is evident that Westminster has the highest crimes recorded followed by Lambeth, Southwark, Newham and Tower Hamlets. Westminster has a significantly higher crime rate than the other 4 boroughs

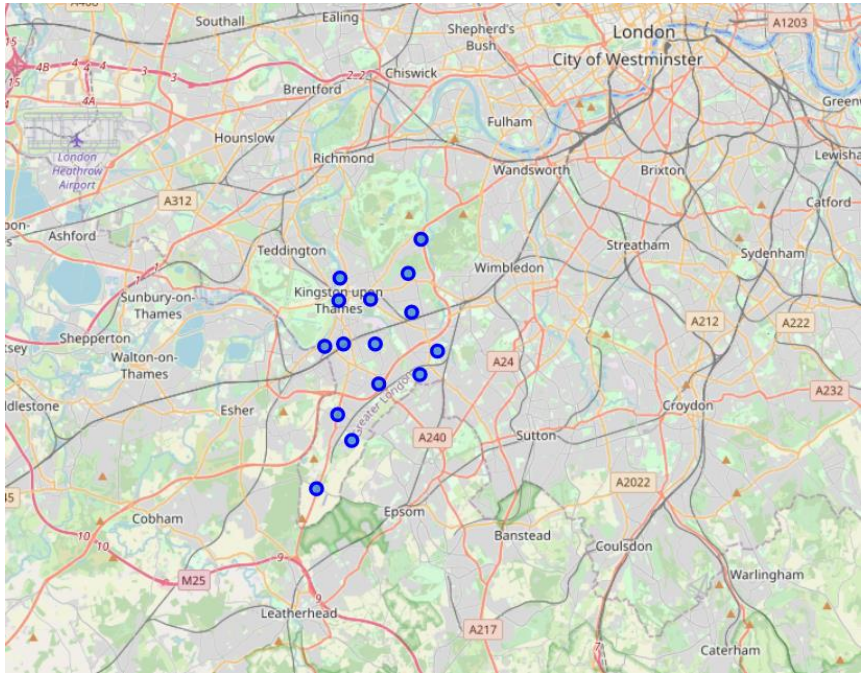
- Boroughs with Lowest Crime rates



Comparing five boroughs with the lowest crime rate during the year 2016, City of London has the lowest recorded crimes followed by Kingston upon Thames, Sutton, Richmond upon Thames and Merton. we will consider the next borough with the lowest crime rate as the safest borough in London which is Kingston upon Thames.

- Neighborhoods in Kingston upon Thames:

  There are 15 neighborhoods in the royal borough of Kingston upon Thames, they are visualised on a map using folium on python.



Modelling:

- Using the final dataset containing the neighborhoods in Kingston upon Thames along with the latitude and longitude, we can find all the venues within a 500-meter radius of each neighborhood by connecting to the Foursquare API. This data frame contains all the venues along with their coordinates and category

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Berrylands	51.393781	-0.284802	Surbiton Racket & Fitness Club	51.392676	-0.290224	Gym / Fitness Center
1	Berrylands	51.393781	-0.284802	K2 Bus Stop	51.392302	-0.281534	Bus Stop
2	Berrylands	51.393781	-0.284802	Cafe Rosa	51.390175	-0.282490	Café
3	Berrylands	51.393781	-0.284802	Kamala Food and Wine	51.397810	-0.284045	Wine Shop
4	Canbury	51.417499	-0.305553	Canbury Gardens	51.417409	-0.305300	Park

- To help people find similar neighborhoods in the safest borough we will be clustering similar neighborhoods using K - means clustering which is a form of unsupervised machine learning algorithm that clusters data based on predefined cluster size. We will use a cluster size of 5 for this project that will cluster the 15 neighborhoods into 5 clusters.
- The reason to conduct a K- means clustering is to cluster neighborhoods with similar venues together so that people can shortlist the area of their interests based on the venues/amenities around each neighborhood.

4. Result and Discussion

- The aim of this project is to help people who want to relocate to the safest borough in London, person can choose the neighborhoods to which they want to relocate based on the most common venues in it.
- Clusters are made based on common venues which allows a person to select the best suitable location based on the personal required neighborhoods and priorities.

5. Conclusion

- This project helps a person get a better understanding of the neighborhoods with respect to the most common venues in that neighborhood.
- It is always helpful to make use of technology to stay one step ahead i.e. finding out more about places before moving into a neighborhood.
- We have just taken safety as a primary concern to shortlist the safest borough of London. The future of this project includes taking other factors such as cost of living in the areas into consideration to shortlist the borough, such as filtering areas based on a predefined budget.