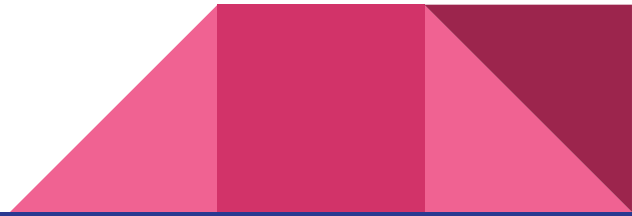
Apllied Data Science Capstone Project (IBM/Coursera)

The Battle of Neighborhoods

By: Aditya Gupta

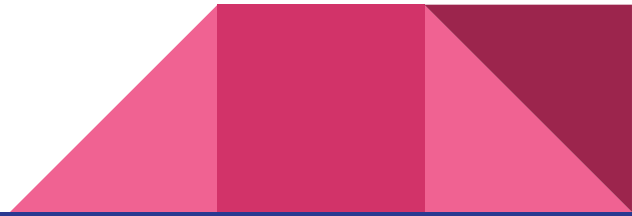
Introduction

The aim of this project is to select the safest borough in London based on the crime rate and to 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.



Interested Audience

This report will be of a great interest to the people who are looking to relocate to London. And in order to find a descent neighborhood or to hunt for an apartment, safety is considered as a major concern. The crime statistics will provide an insight into this issue and help us in solving the problem statement.



Data Acquisition and Cleaning

- The data required for this project is collected from three different sources. The first data source of the project uses the London Crime Data that shows the crime per borough in London.
- The second source of data is scrapped from a wikipedia page that contains the list of London boroughs. This page contains additional information about boroughs.
- The third data source is the list of Neighborhoods in the Royal Borough of Kingston upon Thames as found on the wikipedia page. This dataset is scrapped throught he list of neighborhood available on the site.

Data Cleaning and Scrapping

- Preprocessing a real world data set from Kaggle showing the London Crimes from 2008 to 2016: A dataset consisting of the crime statistics of each borough in London obtained from Kaggle
- Scraping additional information of the different Boroughs in London from a Wikipedia page.: More information regarding the boroughs of London is scraped using the Beautifulsoup library
- Creating a new dataset of the Neighborhoods of the safest borough in London and generating their Co-ordinates.: Co-ordinate of neighborhood will be obtained using Google Maps API geocoding

Methodology

The methodology in this project consists of two parts:

- **Exploratory Data Analysis**
- **Modelling**

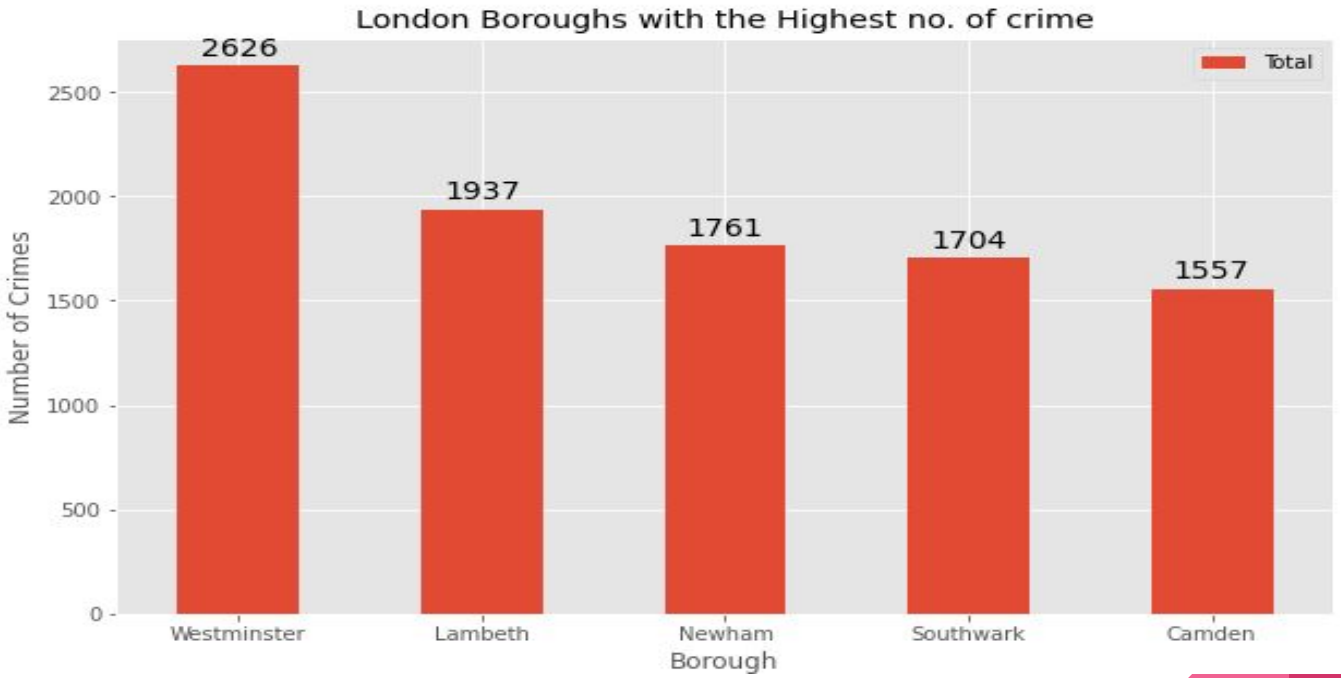
Exploratory Data Analysis

Statistical Summary of Dataset:

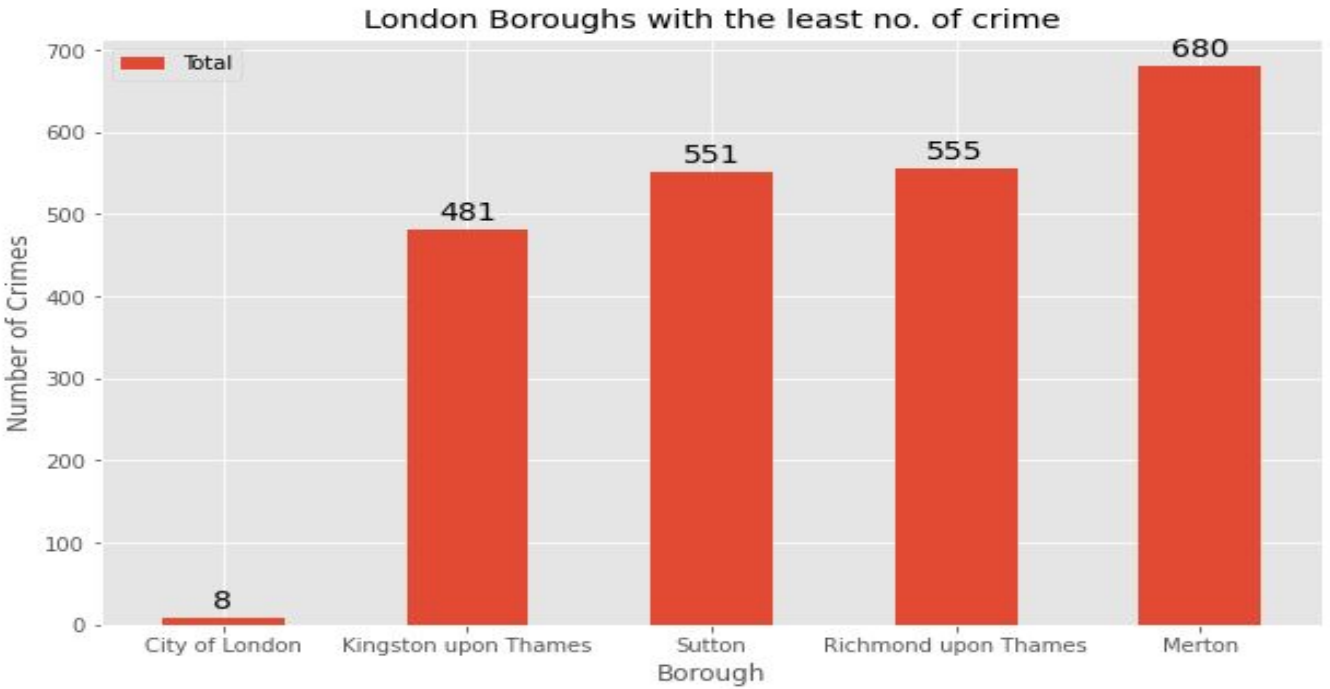
```
[ ]  London_crime.describe()
```

	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	110.121212	104.242424	65.212121	24.757576	33.333333	470.151515	373.727273	1181.545455
std	41.052373	35.411007	38.764318	11.797855	21.548879	251.751753	139.767278	485.989139
min	0.000000	0.000000	1.000000	0.000000	2.000000	4.000000	1.000000	8.000000
25%	82.000000	77.000000	41.000000	18.000000	14.000000	310.000000	309.000000	914.000000
50%	112.000000	107.000000	59.000000	25.000000	27.000000	457.000000	395.000000	1207.000000
75%	133.000000	124.000000	104.000000	30.000000	49.000000	566.000000	467.000000	1457.000000
max	191.000000	186.000000	169.000000	60.000000	73.000000	1446.000000	654.000000	2626.000000

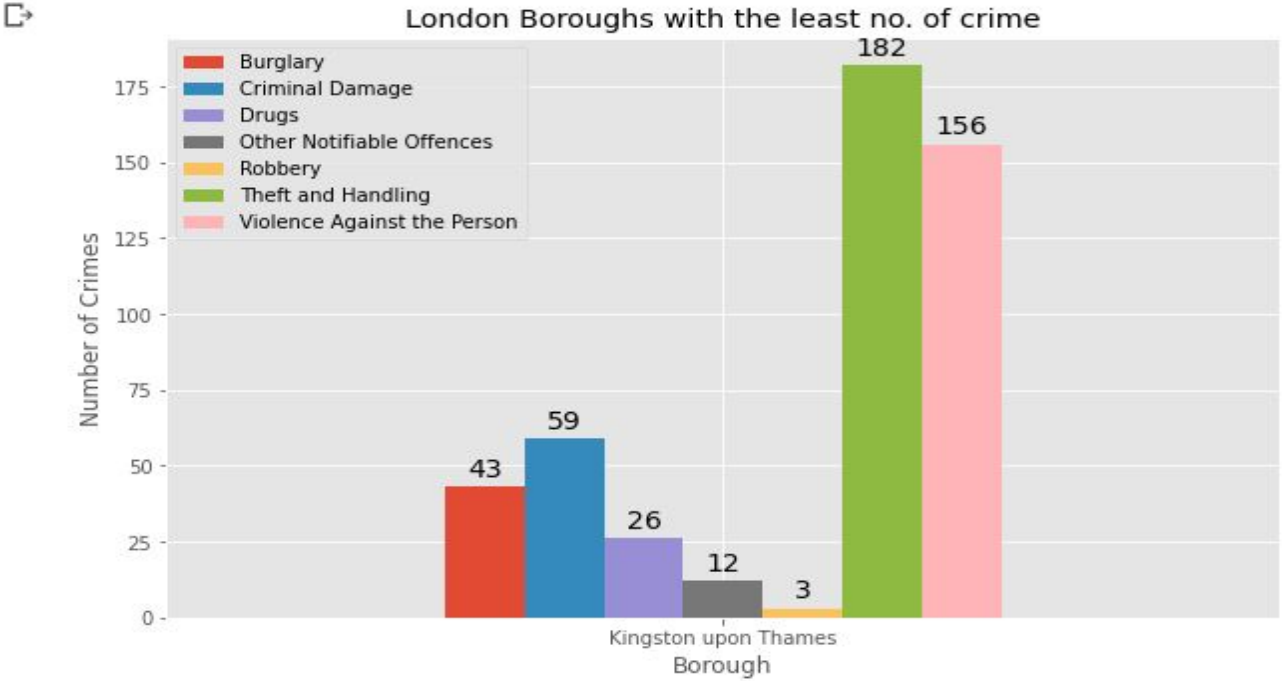
Boroughs with Highest Crime Rate



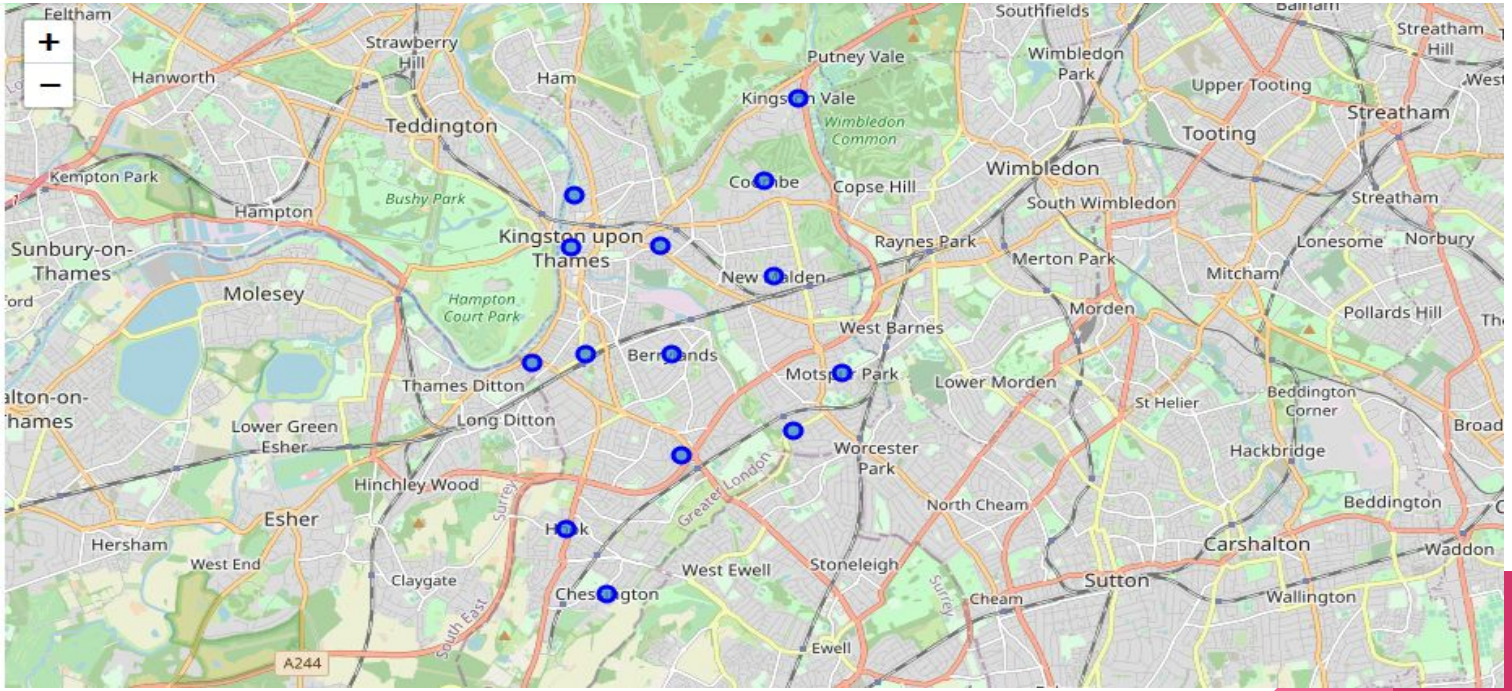
Boroughs with Lowest Crime Rate



Different Types of Crime in the borough: ‘Kingston upon Thames’

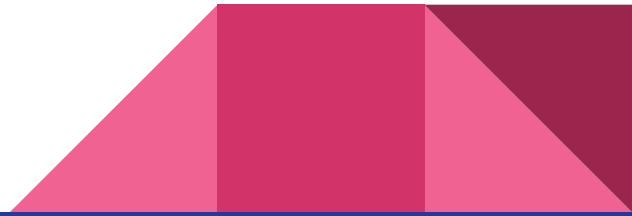


Neighborhood in Kingston upon Thames



Modelling

- Finding all the venues within a 500 meter radius of each neighborhood.
- Perform one hot ecoding on the venues data.
- Grouping the venues by the neighborhood and calculating their mean.
- Performing a K-means clustering (Defining K = 5)

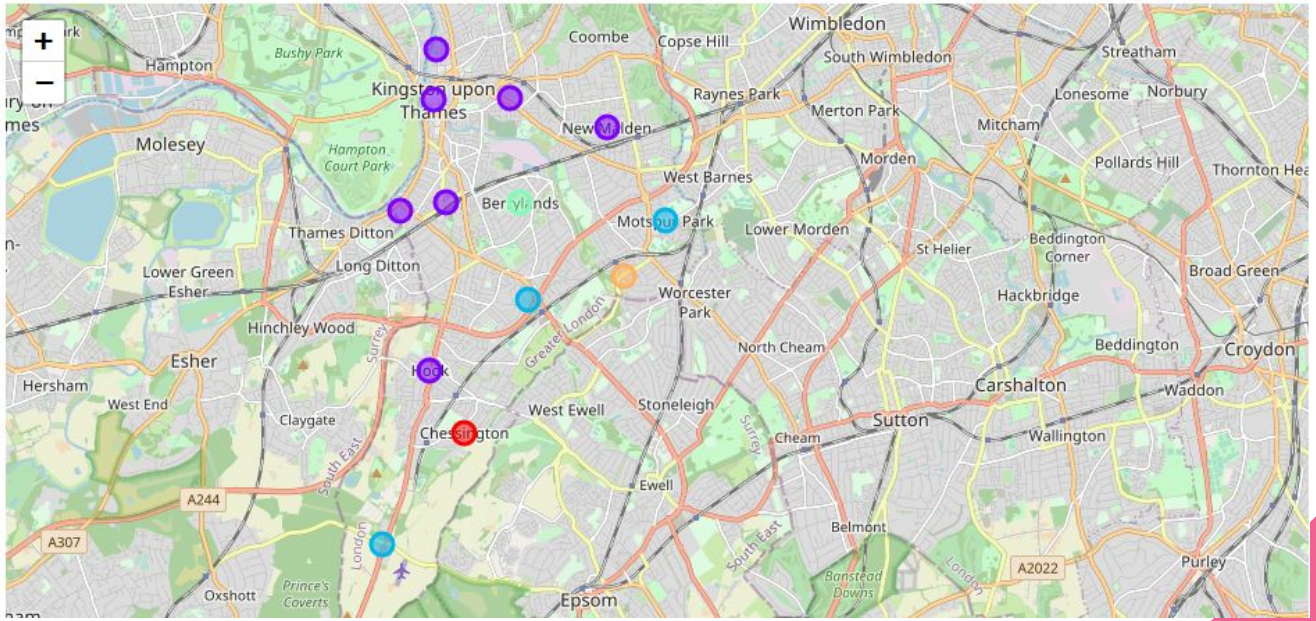


Venue Details of Each Neighborhood

	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	Alexandra Park	51.394230	-0.281206	Park
2	Berrylands	51.393781	-0.284802	K2 Bus Stop	51.392302	-0.281534	Bus Stop
3	Berrylands	51.393781	-0.284802	Cafe Rosa	51.390175	-0.282490	Café
4	Canbury	51.417499	-0.305553	The Boater's Inn	51.418546	-0.305915	Pub

Results

Visualizing the Clusters



Discussions

The project aims to help people who want to relocate to the safest borough in London, expats can chose the neighborhoods to which they want to relocate based on the most common venues in it. For example if a person is looking for a neighborhood with good connectivity and public transportation we can see that Clusters 3 and 4 have Train stations and Bus stops as the most common venues. If a person is looking for a neighborhood with stores and restaurants in a close proximity then the neighborhoods in the first cluster is suitable. For a family I feel that the neighborhoods in Cluster 4 are more suitable dues to the common venues in that cluster, these neighborhoods have common venues such as Parks, Gym/Fitness centers, Bus Stops, Restaurants, Electronics Stores and Soccer fields which is ideal for a family.

Conclusion

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 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 based on safety and a predefined budget.

Thank You!