

Capstone Project - The Battle of Neighborhoods

(Week2) Report

Table of contents:

- Introduction: Business Problem
- Data
- Methodology
- Analysis
- Results and Discussion
- Conclusion

Introduction: Business 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 report will be targeted to people who are looking to relocate to London. In order to finalise a neighborhood to hunt for an apartment, safety is considered as a top concern when moving to a new place. If you don't feel safe in your own home, you're not going to be able to enjoy living there. The crime statistics will provide an insight into this issue. We will focus on the safest borough and explore its neighborhoods and the 10 most common venues in each neighborhood so that the best neighborhood suited to an individual's needs can be selected.

Data:

- Based on definition of our problem, factors that will influence our decision are:
- The total number of crimes committed in each of the borough during the last year.
- The most common venues in each of the neighborhood in the safest borough

selected.

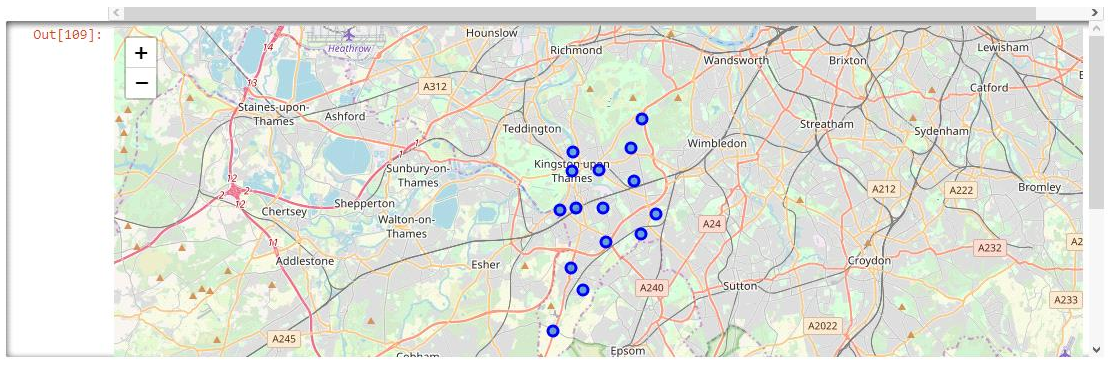
Following data sources will be needed to extract/generate the required information:

- Part 1: 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.

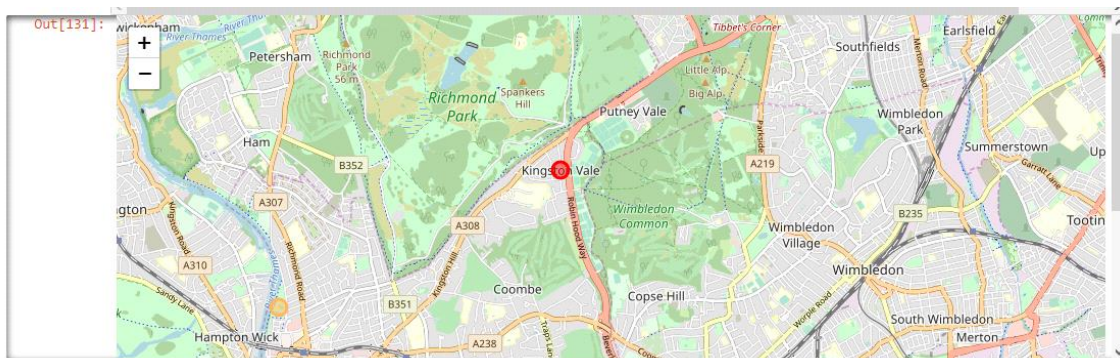
Out[14]:

	Borough	No_of_Crimes							Total
Major_Category		Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	
0	Barking and Dagenham	1287	1949	919	378	534	5607	6067	16741
1	Barnet	3402	2183	906	499	464	9731	7499	24684
2	Bexley	1123	1673	646	294	209	4392	4503	12840
3	Brent	2631	2280	2096	536	919	9026	9205	26693
4	Bromley	2214	2202	728	417	369	7584	6650	20164
5	Camden	2652	1935	1493	490	899	14088	7626	29183
6	City of London	2	2	10	6	4	129	25	178
7	Croydon	2738	3219	1367	718	1139	9229	10302	28712
8	Ealing	2492	2562	1355	613	669	10040	9396	27127
9	Enfield	2541	2136	1063	492	807	8037	7409	22485
10	Greenwich	1780	2476	867	521	486	8010	8590	22730
11	Hackney	2719	1981	1353	499	1030	11851	8832	28265
12	Hammersmith and Fulham	1531	1408	1321	474	397	8925	6118	20174

- Part 2: 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.
- Part 3: 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.



Visualize the Neighborhood of Kingston upon Thames Boroug

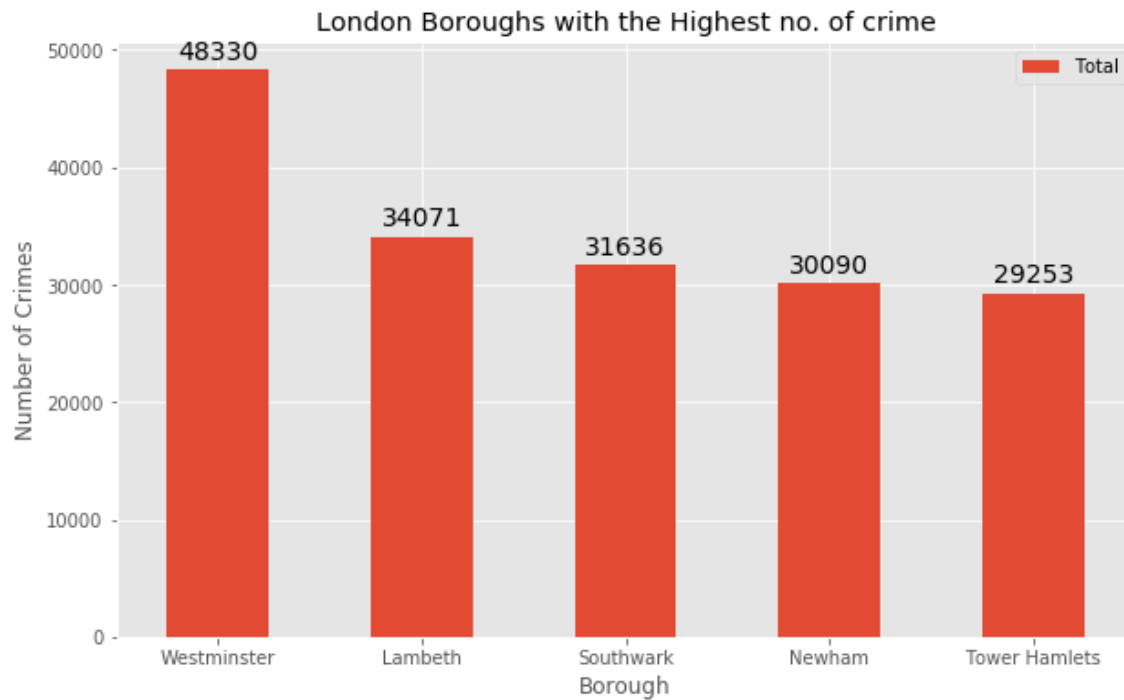


Visualize the clusters

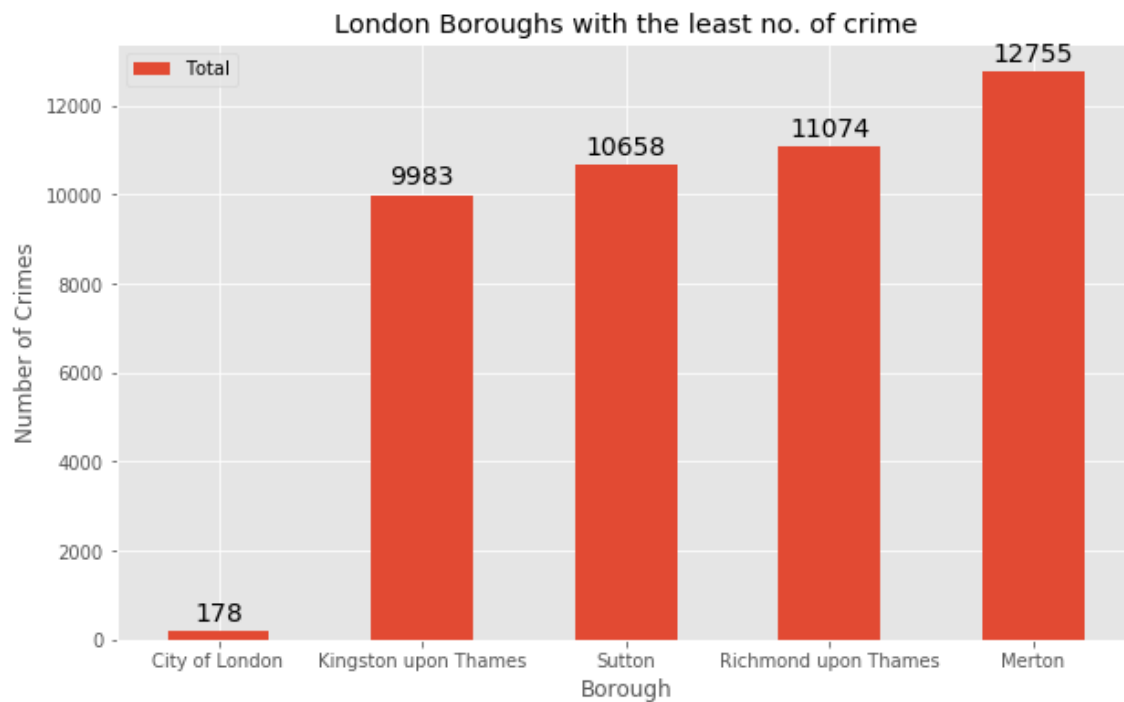
Methodology:

The methodology in this project consists of two parts:

- Exploratory Data Analysis: Visualise the crime rates in the London boroughs to identify the safest borough and extract the neighborhoods in that borough to find the 10 most common venues in each neighborhood.

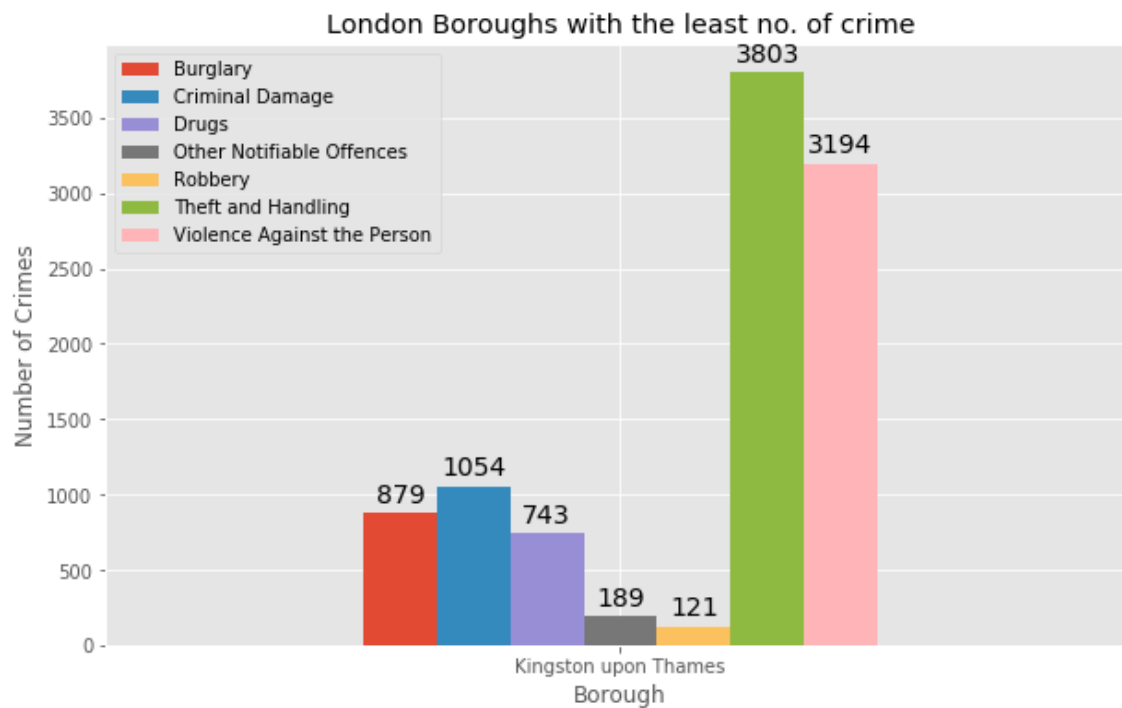


Visualize the five boroughs with the highest number of crimes



Visualize the five boroughs with the least number of crimes

- **Modelling:** 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.



Visualizing different types of crimes in the borough 'Kingston upon Thames'

Analysis:

Analyse each of the clusters to identify the characteristics of each cluster and the neighborhoods in them. We can conclude that Kingston upon Thames is the safest borough when compared to the other boroughs in London.

Results and Discussion:

The aim of this project is to help people who want to relocate to the safest borough in London, expats can choose 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 due 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:

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 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.