Capstone Project - The Battle of Neighborhoods

ERIM SEZER

Introduction

- **Background:** Safety is a top concern when moving to a new area. Not feeling safe will result in not enjoying life.
- **Problem:** The aim of this project is to select the safest area in london based on the total criminal rates , to explore the neighborhoods of that area to find the 10 most common venues in each neighborhood by using k-means clustering
- **Interest:** Business man, expats, who ever is cosnidering to move to London will be interested in this analysis.

Data acquisition and cleaning

- The project is based on three data sources
  - London crime data → Shows the crime per area in London
  - Wikipedia → Shows the list of London boroughs.
  - Wikipedia → List of Niegborhoods in the Royal Borough of Kingston

Data cleaning

- The data of each data source is cleaned sepereately
  - London crime data: most recent yeras are selected (starting from 2016). The major categories of crime are pivoted to get the total crimes per the boroughs for each major category
  - Data from wikipedia is retrieved by using the Python library beautiful soup
  - The data sources are merged and visualized in order to identify the boroughs with the least crimes
  - Once identified: The thir data sources is used to identify the names of the neighborhoods with their latitude and longitude using Google Maps API
  - The new data set shows the top 10 most common venues for each neighborhood using the Foursquare API and finally using k-means clustering algorithms to cluster similar neighborhoods together

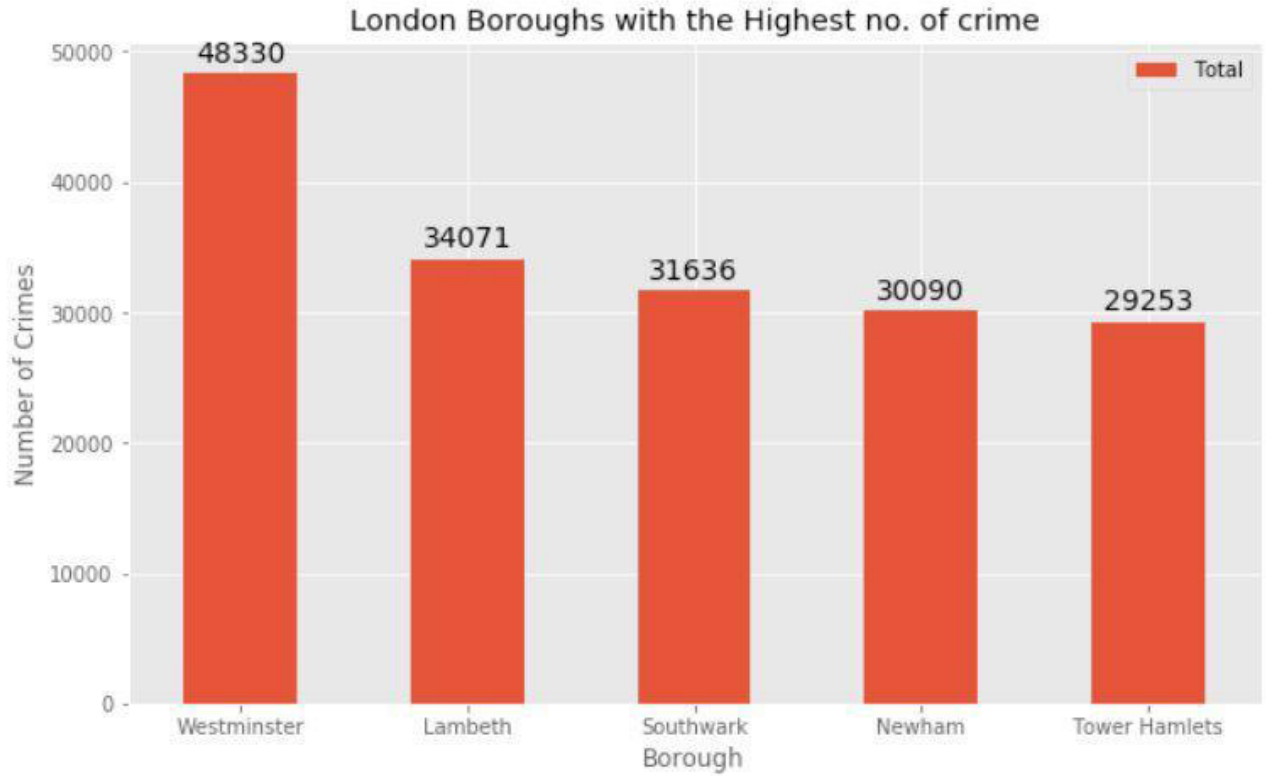
Methodology

Exploratory Data Analysis

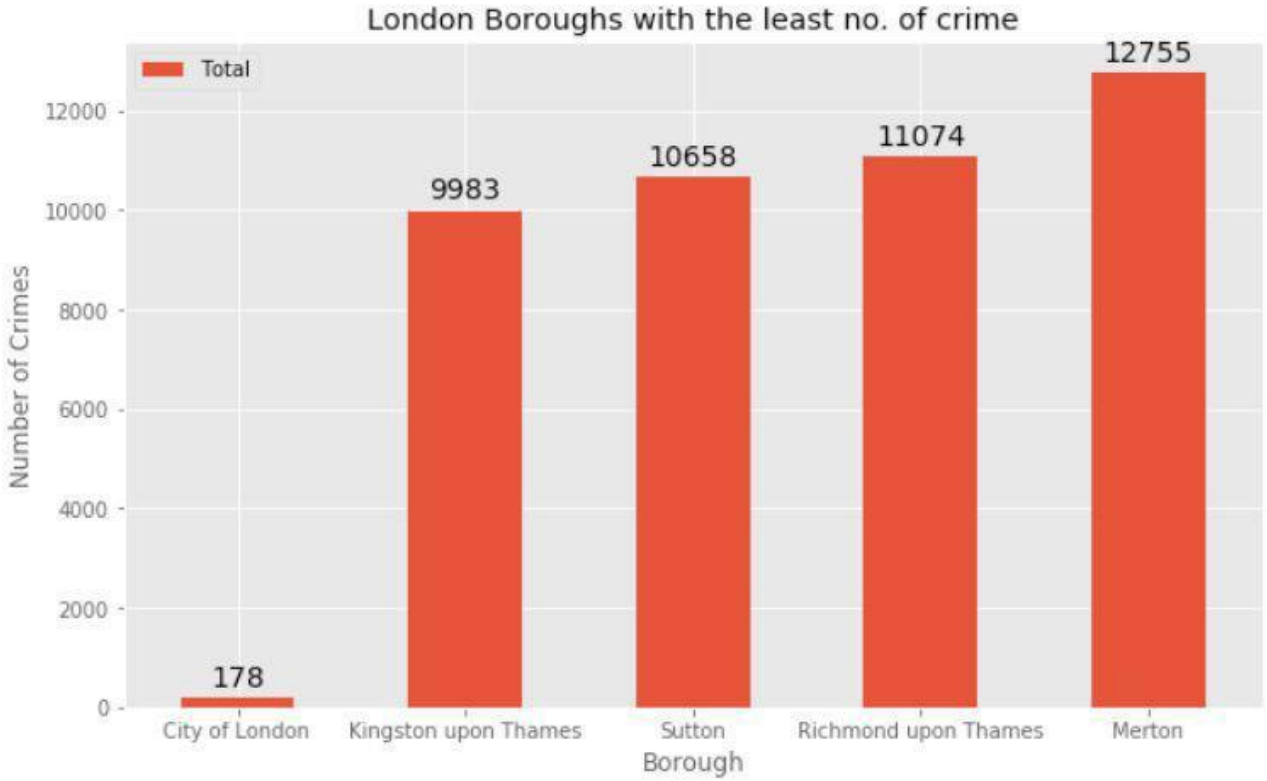
Summary of crimes

	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

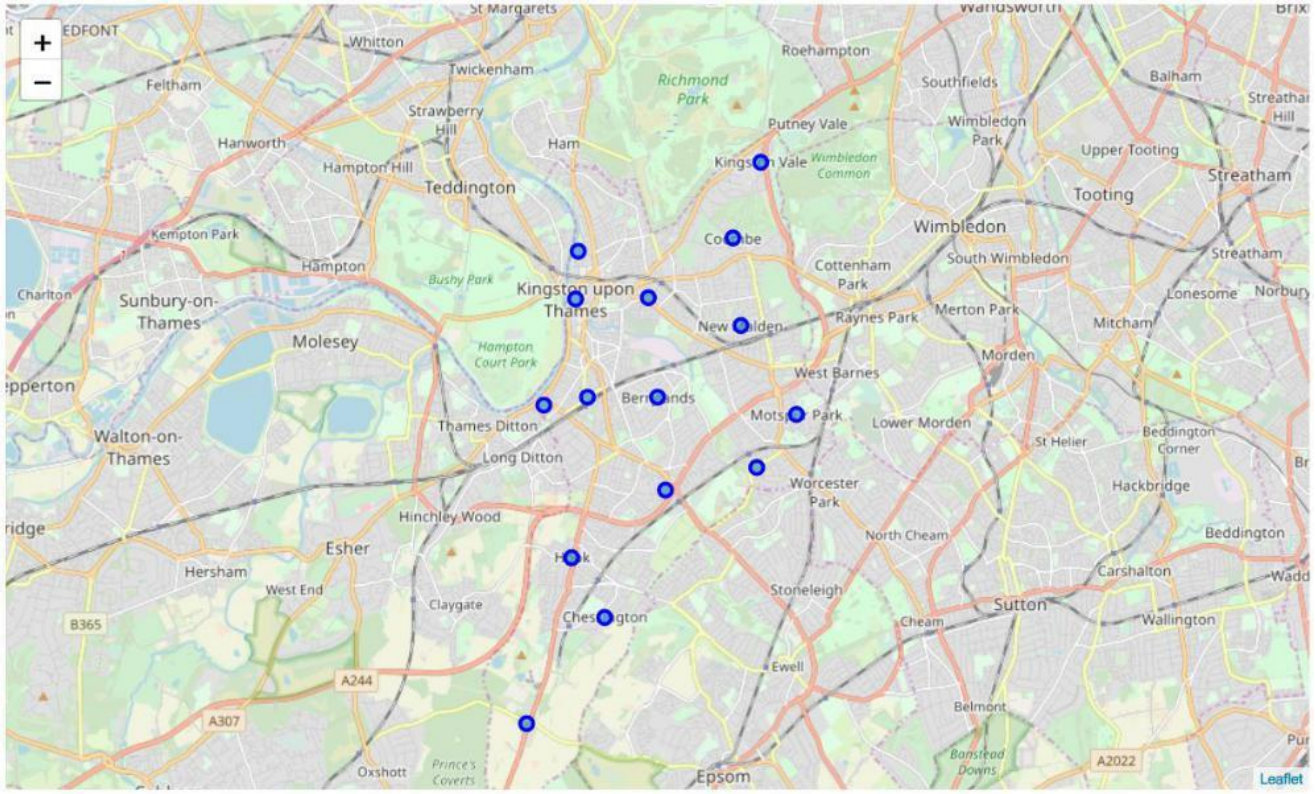
The count for each major category of crime is 33, which is the number of boroughs. „Theft and Handling“ is the highest reported crime during the year 2016 followed by „Violence against a person“, „criminal damage“. The lowest recorded crimes are „Drugs“, „Robbery“ and „Other notifiable offenses“



Boroughs with highest crime rate



Boroughs with lowest crime rate



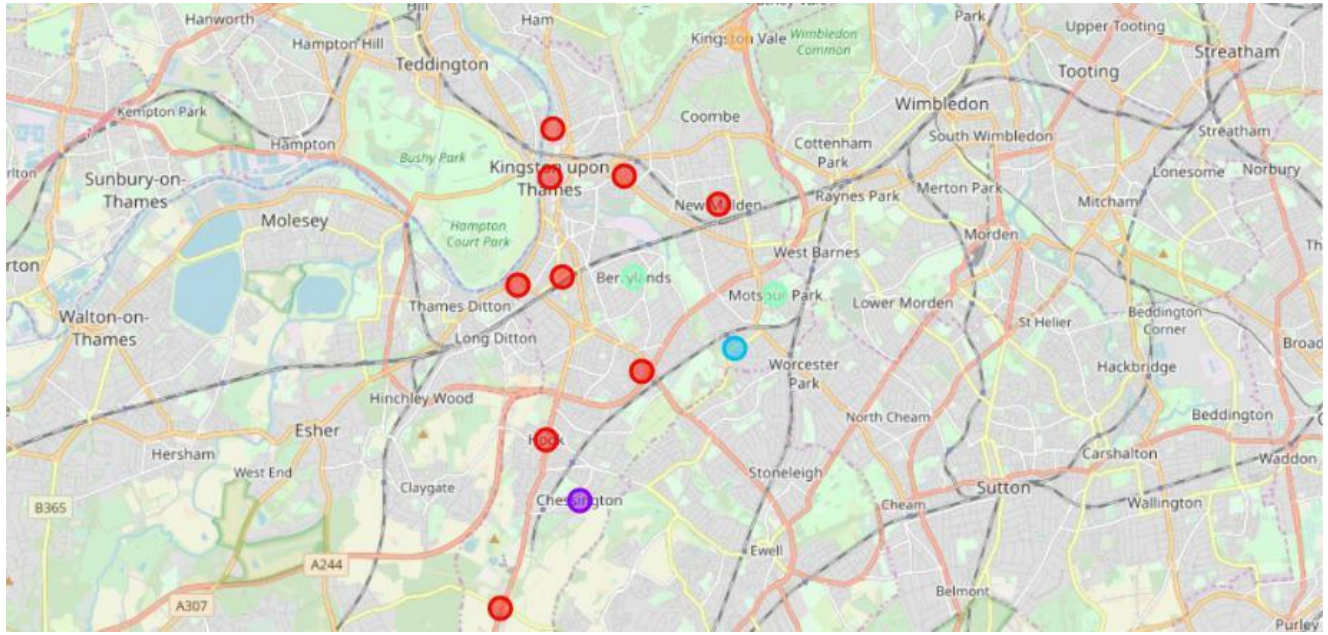
Neighborhoods in Kingston upton Thames

Modelling

The final data set includes neighborhoods in Kingston upon Thames with latitude and longtitude and we can find all venues within 500 meter radius by connecting to Foursquare API.

	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

- One hot encoding → grouped by neighborhood and the mean of the venues are calculated. 10 common venues for each neighborhood is calculated
- Clustering similar beighborhoods using k-means
- Cluster size of 5, which will custer the 15 neighborhoods



Results

Each cluster is color coded → Majority of the neighborhood falls in the red cluster which is the first cluster. Three neighborhoods have their own color coding (blue, pruple, and yellow). Green cluster consist of two neighborhoods.

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 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.
- The choices of neighborhoods may vary from person to person.

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.