Coursera Applied Capstone Project
"The Battle of Neighborhoods (Week 2)"

Presentation Prepared By

Shaileshkumar S. Patel

1. Introduction

- **Background:** Saftey is a top concern when moving to a new area. If you feel don't safe in your own home, you are not going to be able to enjoy living there.

- **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.

- **Interest:** Expats who are considering to relocate to London will be interested to identify the safest borough in London and explore its neighbourhoods and common venues around each neighborhood.

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. This page contains additional information about the boroughs.

➢ The third data source is the list of the Neighbourhoods in the Royal Borough of kingston upon Themes as found on the wiki page.

- Data Cleaning: The data cleaning process for each of the three sources of data are 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 data sets are merged on the Borough names to form a new data set. The purpose of this data set 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 data set is created, with the names of the neighborhoods and the name of the borough with the latitude and longitude obtained using Google Maps API geocoding.

- The new data set 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 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 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 the highest crime rates

London Boroughs with the Highest no. of crime
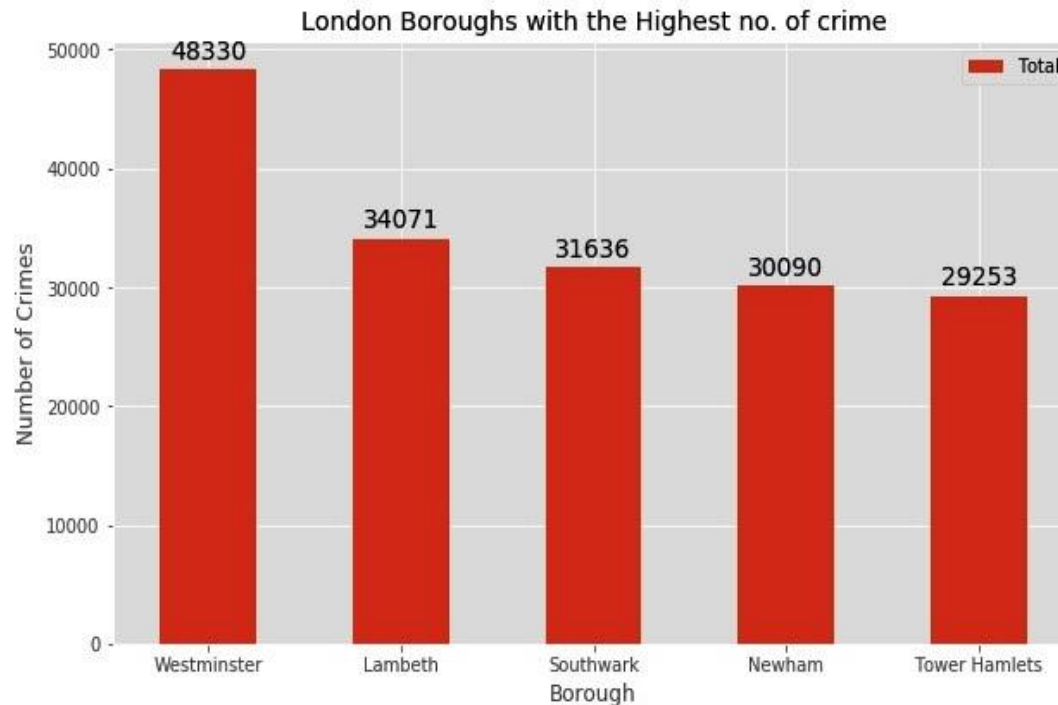
Comparing five boroughs with the highest crime rate during the year 2016 it is evident that Westminster has the highest crimes recorded followed by Lambeth, Southwark, Newham and Tower Hamlets. Westminster has a significantly higher crime rate than the other 4 boroughs.

London Boroughs with the least no. of crime
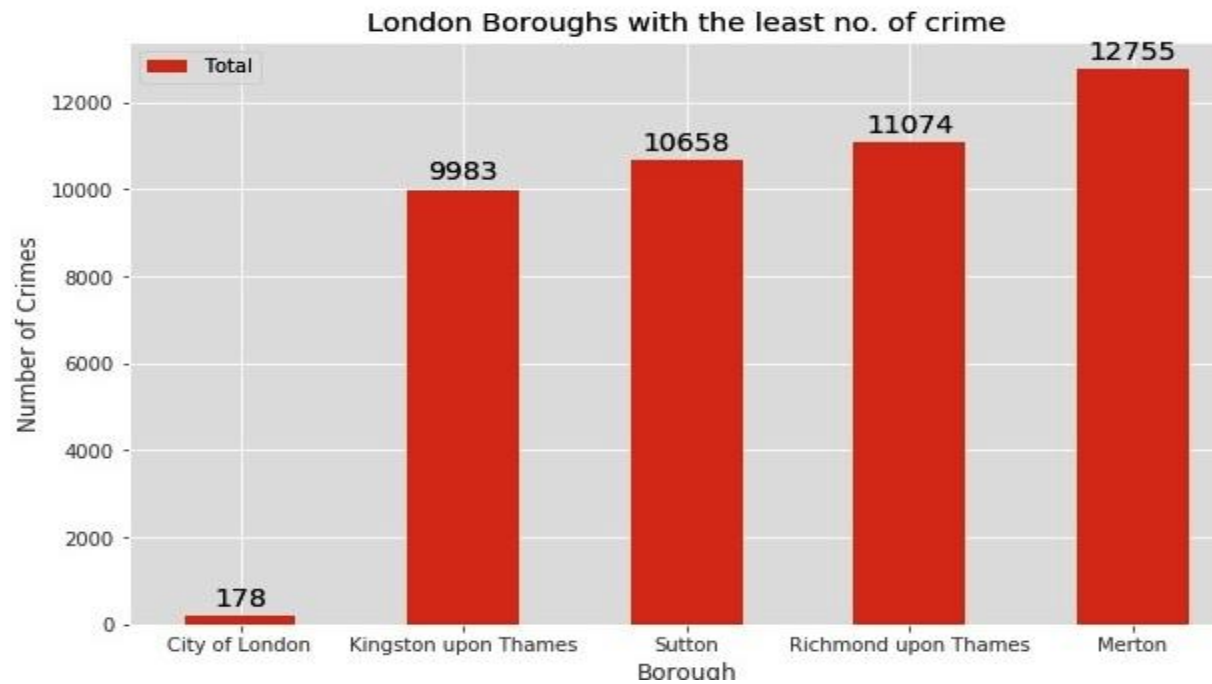
• Comparing five boroughs with the lowest crime rate during the year 2016, City of London has the lowest recorded crimes followed by Kingston upon Thames, Sutton, Richmond upon Thames and Merton.
• We will consider the next borough with the lowest crime rate as the safest borough in London which is Kingston upon Thames.

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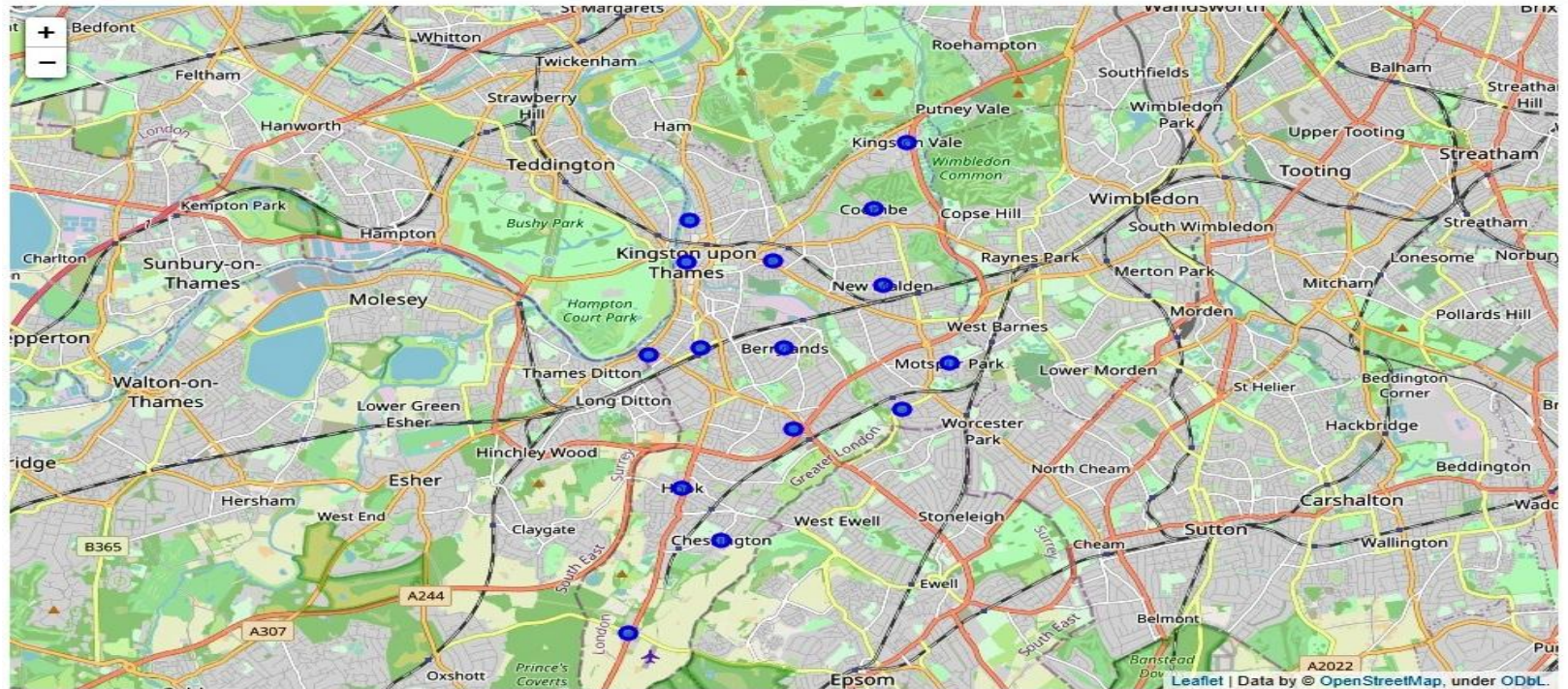
There are 15 neighborhoods in the royal borough of Kingston upon Thames, they are visualised on a map using folium on python.

Modeling

	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	ExactPrint-UK	51.393288	-0.288874	Print Shop
4	Canbury	51.417499	-0.305553	Canbury Gardens	51.417409	-0.305300	Park

One hot encoding is done on the venues data. The Venues data is then grouped by the Neighborhood and the mean of the venues are calculated, finally the 10 common venues are calculated for each of the neighborhoods.
• 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

After running the K-means clustering we can access each cluster created to see which neighborhoods were assigned to each of the five clusters. Visualizing the clustered neighborhoods on a map using the folium library.
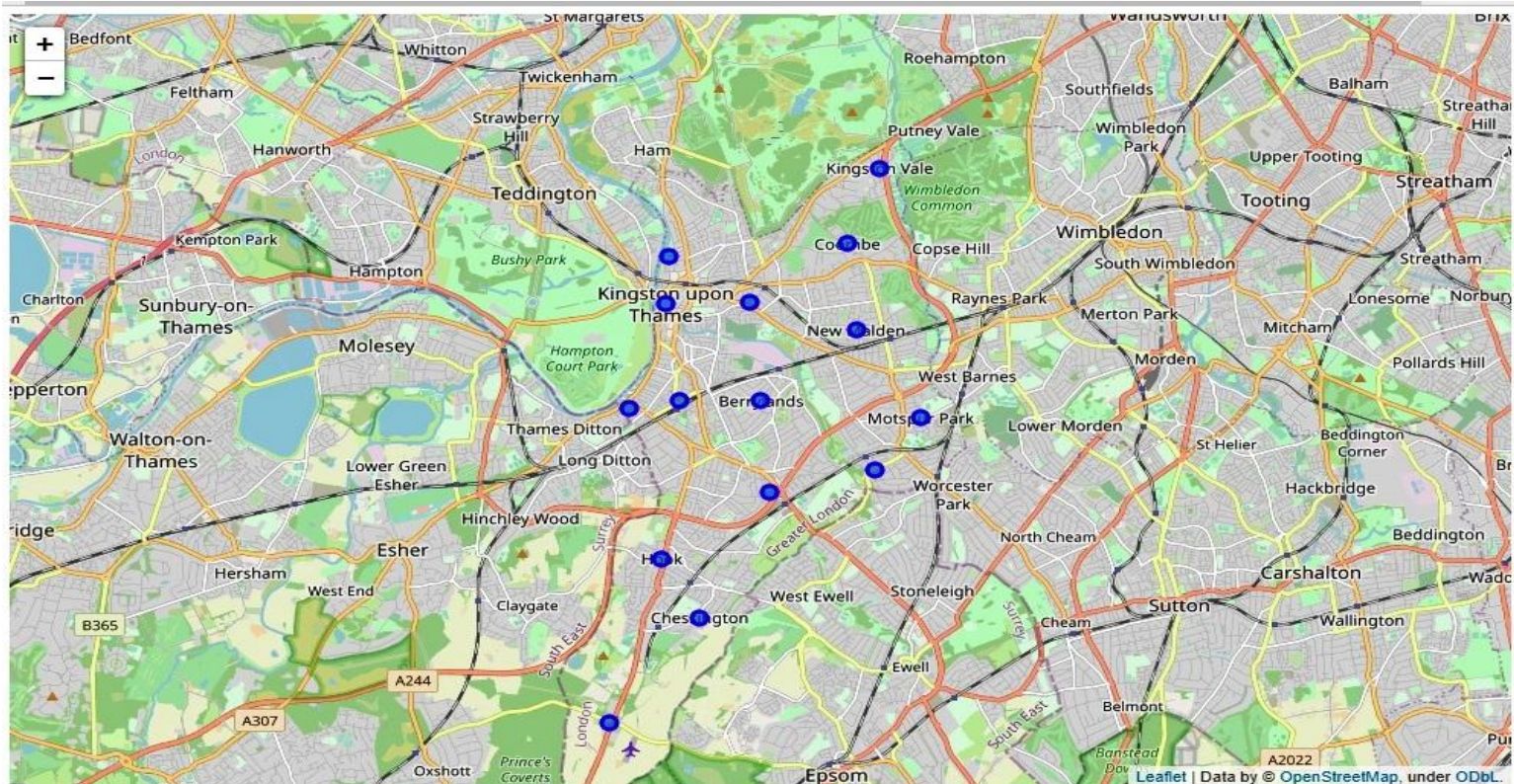
Cluster 1: Looking into the neighborhoods in the first cluster

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
1	Canbury	Kingston upon Thames	51.417499	-0.305553	1	Pub	Café	Supermarket	Spa	Plaza	Shop & Service	Gym / Fitness Center	Fish & Chips Shop
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	1	Café	Coffee Shop	Sushi Restaurant	Pub	Burger Joint	Asian Restaurant	German Restaurant	Gift Shop
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	1	Grocery Store	Garden Center	Pub	Restaurant	Fish & Chips Shop	Discount Store	Dry Cleaner	Electronics Store
9	New Malden	Kingston upon Thames	51.405335	-0.263407	1	Indian Restaurant	Gastropub	Sushi Restaurant	Supermarket	Bar	Chinese Restaurant	Korean Restaurant	Fish & Chips Shop
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	1	Italian Restaurant	Indian Restaurant	Food	Platform	Pub	Japanese Restaurant	Coffee Shop	Hotel
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	1	Indian Restaurant	Coffee Shop	Pub	Gym	Playground	Chinese Restaurant	Fast Food Restaurant	Café
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	1	Coffee Shop	Pub	Grocery Store	Italian Restaurant	Pharmacy	Breakfast Spot	French Restaurant	Gym / Fitness Center
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	1	Grocery Store	Pharmacy	Train Station	Bowling Alley	Furniture / Home Store	Hotel	Italian Restaurant	Discount Store

The cluster one is the biggest cluster with 9 of the 15 neighborhoods in the borough Kingston upon Thames. Upon closely examining these neighborhoods we can see that the most common venues in these neighborhoods are Restaurants, Pubs, Cafe, Supermarkets, and stores.

5. Conclusion

- The aim of this project is 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.

- The preference of venues may vary from person to person, they can select a neighborhood based on ones priorities.

6. 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 projectincludes 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.