A dramatic landscape of mountains shrouded in mist at sunset. The sky is a warm, hazy orange and yellow, with the sun low on the right horizon. The mountains are dark and rugged, with thick white mist or clouds filling the valleys and clinging to the slopes. The overall mood is serene yet powerful.

CAPSTONE PROJECT - THE BATTLE OF NEIGHBORHOODS - PRESENTATION

BASU SINGH SEHRAWAT

INTRODUCTION

- **Background:** Safety is a top concern when moving to a new area. If you don't feel safe in your own home, you're 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 neighborhoods and common venues around each neighborhood.

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 Neighborhoods in the Royal Borough of Kingston upon Thames as found on the Wikipedia 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 year 2018 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 BeautifulSoup 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 2018
- After visualizing the crime in each borough we can find the borough with 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-mean clustering algorithm to cluster similar neighborhoods together

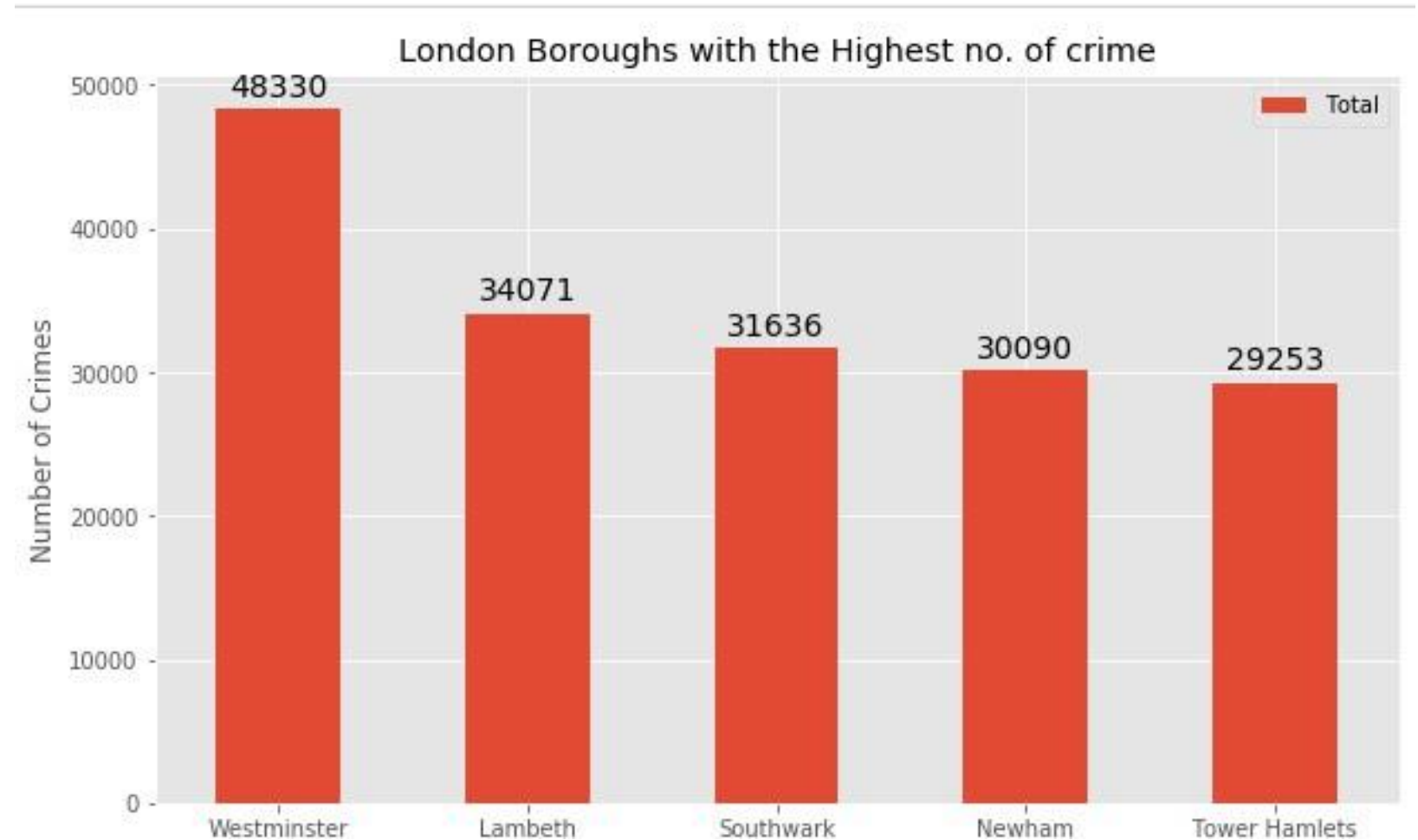
METHODOLOGY

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

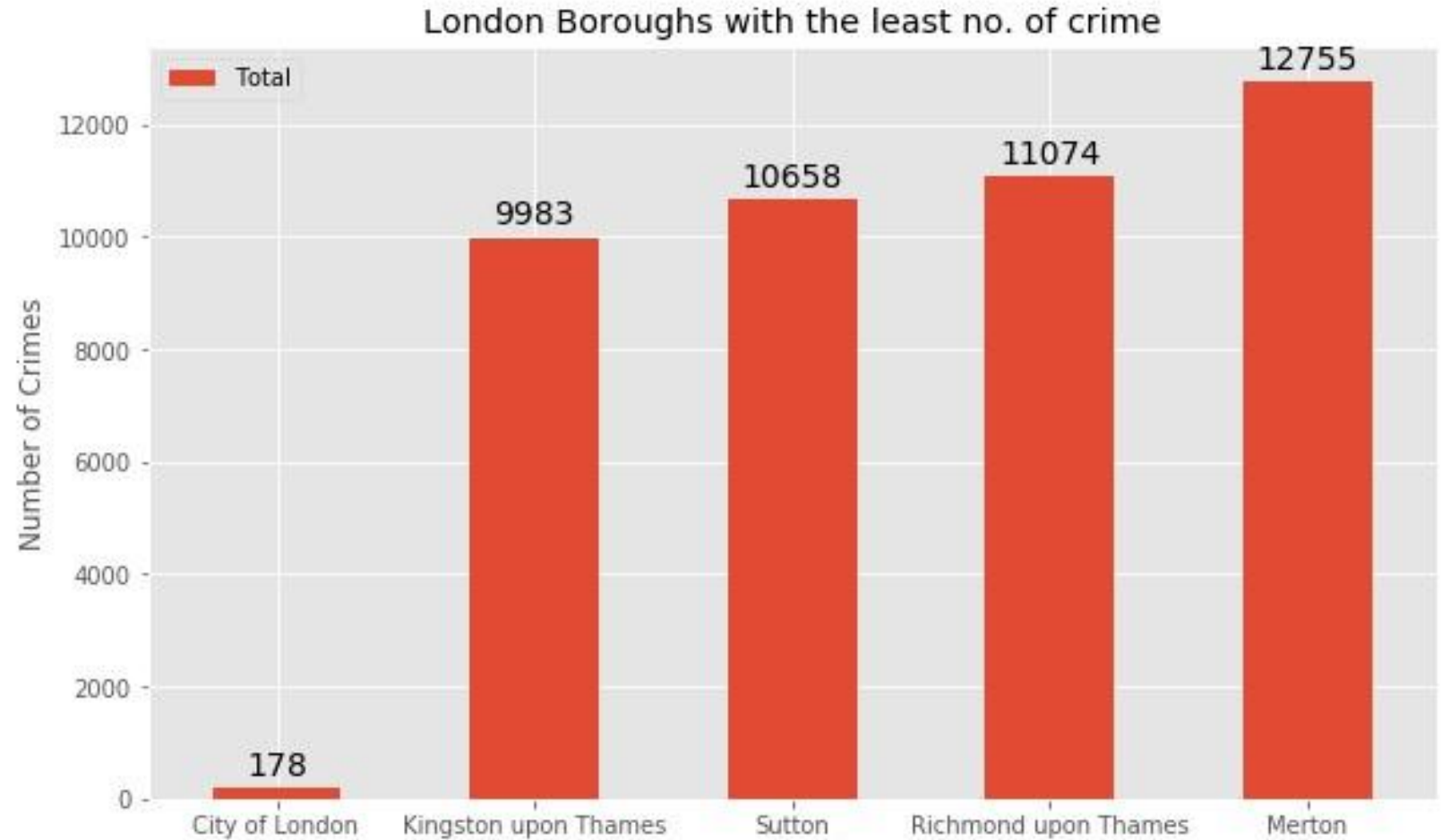
The count for each of the major categories of crime returns the value 33 which is the number of London boroughs. 'Theft and Handling' is the highest reported crime during the year 2018 followed by 'Violence against the person', 'Criminal damage'. The lowest recorded crimes are 'Drugs', 'Robbery' and 'Other Notifiable offenses'

Comparing five boroughs with the highest crime rate during the year 2018 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



Comparing five boroughs with the lowest crime rate during the year 2018, City of London has the lowest recorded crimes followed by Kingston upon Thames, Sutton, Richmond upon Thames and Merton

- City of London has a significantly lower crime rate because it is the 33rd principal division of Greater London but it is not a London borough. It has an area of 1.12 square miles and a population of 7000 as of 2013 that it is a small area
- We will consider the next borough with the lowest crime rate as the safest borough in London which is Kingston upon Thames



Modelling

- Using the final data set 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

	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 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

RESULTS

- After running the K-means clustering we can access each cluster created to see which neighborhoods were assigned to each of the five cluster. Visualizing the clustered neighborhoods on a map using the folium library
- Each cluster is color coded for the ease of presentation; we can see that majority of the neighborhood falls in the red cluster which is the first cluster. Three neighborhoods have their own cluster, these are the clusters 2, 3 & 5. The green cluster consists of two neighborhoods which is the 4th cluster

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 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.
- The preferences of venues may vary from person to person, they can select a neighborhood based on one's priorities

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.