Capstone Project
The Battle of Neighborhoods Presentation

By
Narayanan R

1. Introduction

- **Background:** Safety is a top concern when you are moving to a new area. You must be 100% confident that the place you move in is safe and has got all amenities available

- **Problem:** This project aims to select the safest borough in London based on the total crimes, explore the neighbourhoods of that borough to find the 10 most common venues in each neighbourhood and finally cluster the neighbourhoods using k-mean clustering.

- **Interest:** People 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 neighbourhood

2. Data Collection and Preparation

Data Collection: Data collection for this project is from 3 sources over the internet.

- 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
- The third data source is the list of Neighborhoods in the Royal Borough of Kingston upon Thames as found on a wikipedia page

Data Preparation: The data preparation for the 3 sources above is done separately.

- From the London Crime data only 2016 Crimes are selected. The major categories of crime are pivoted to get total crime per borough 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. We then convert it in to a data frame. This is important because we will be merging the two datasets together using the Borough names.

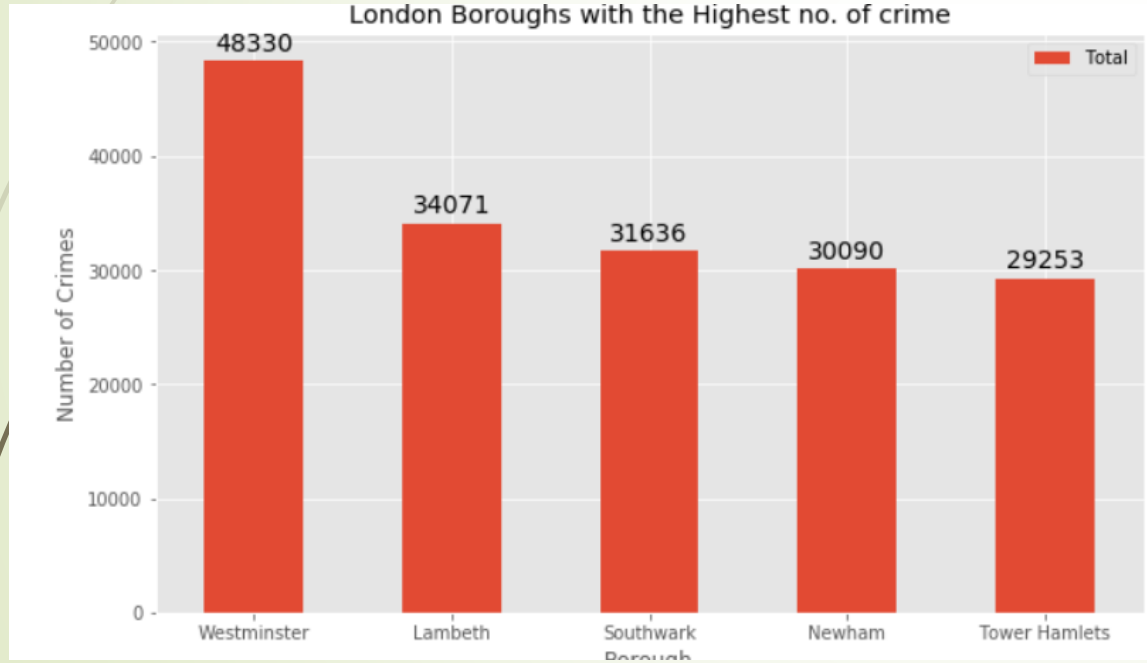
- The third source of data is acquired from the list of neighborhoods in the safest borough on wikipedia. This dataset is created from scratch, the pandas data frame is created with the names of the neighborhoods and the name of the borough with the latitude and longitude left blank

- The coordinates of the neighborhoods is be obtained using Google Maps API geocoding to get the final dataset.

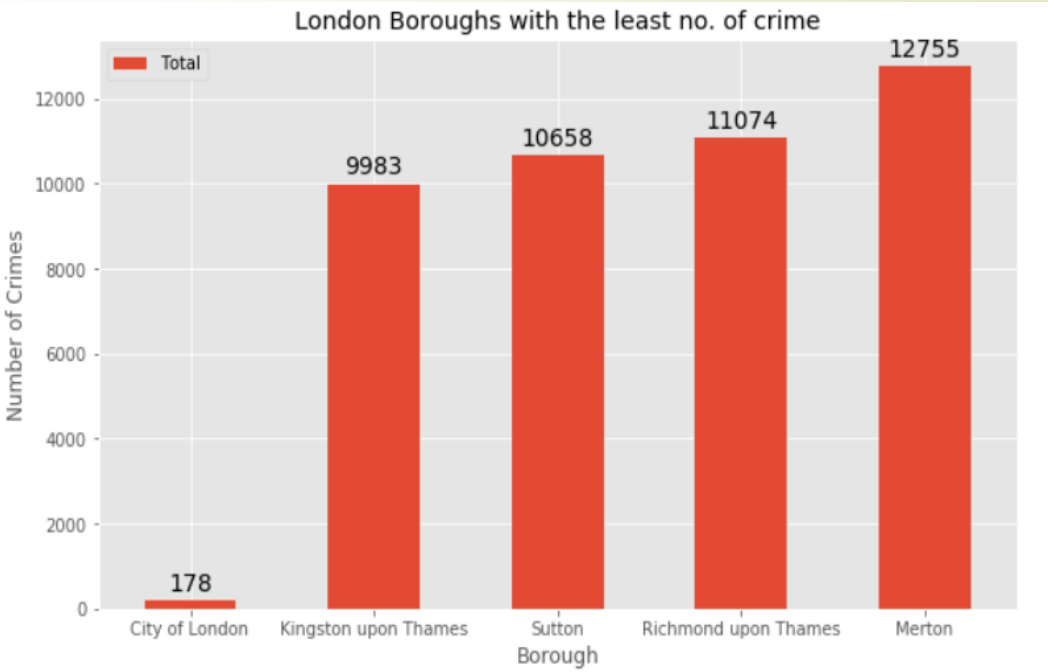
3. Methodology

Exploratory Data Analysis

Boroughs with highest crime rates

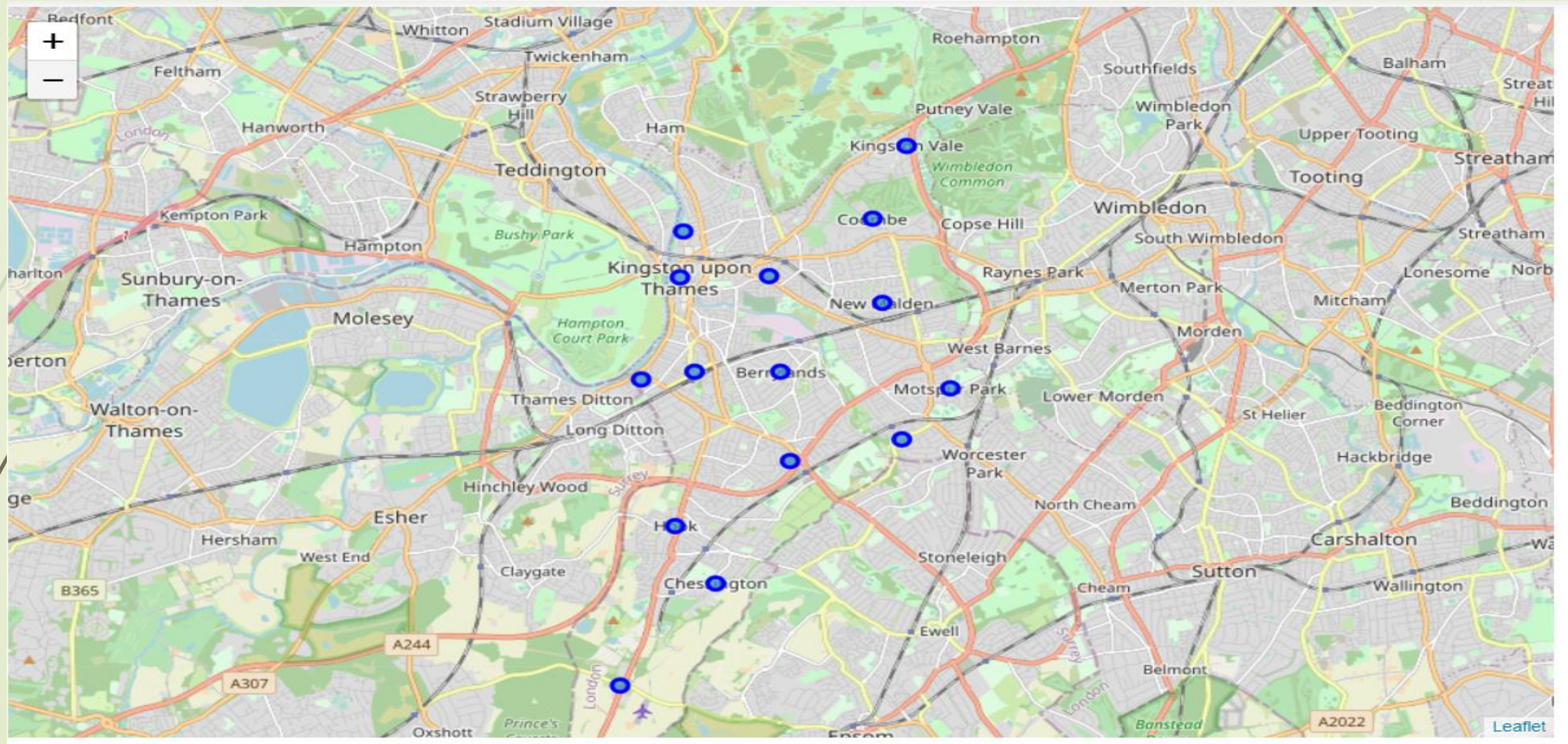


Boroughs with lower Crime rates



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 which suggests that it is a small area. Hence we will consider the next borough with the lowest crime rate as the safest borough in London which is Kingston upon Thames.

Neighborhood in Kingston Upon Thames



4. Modelling

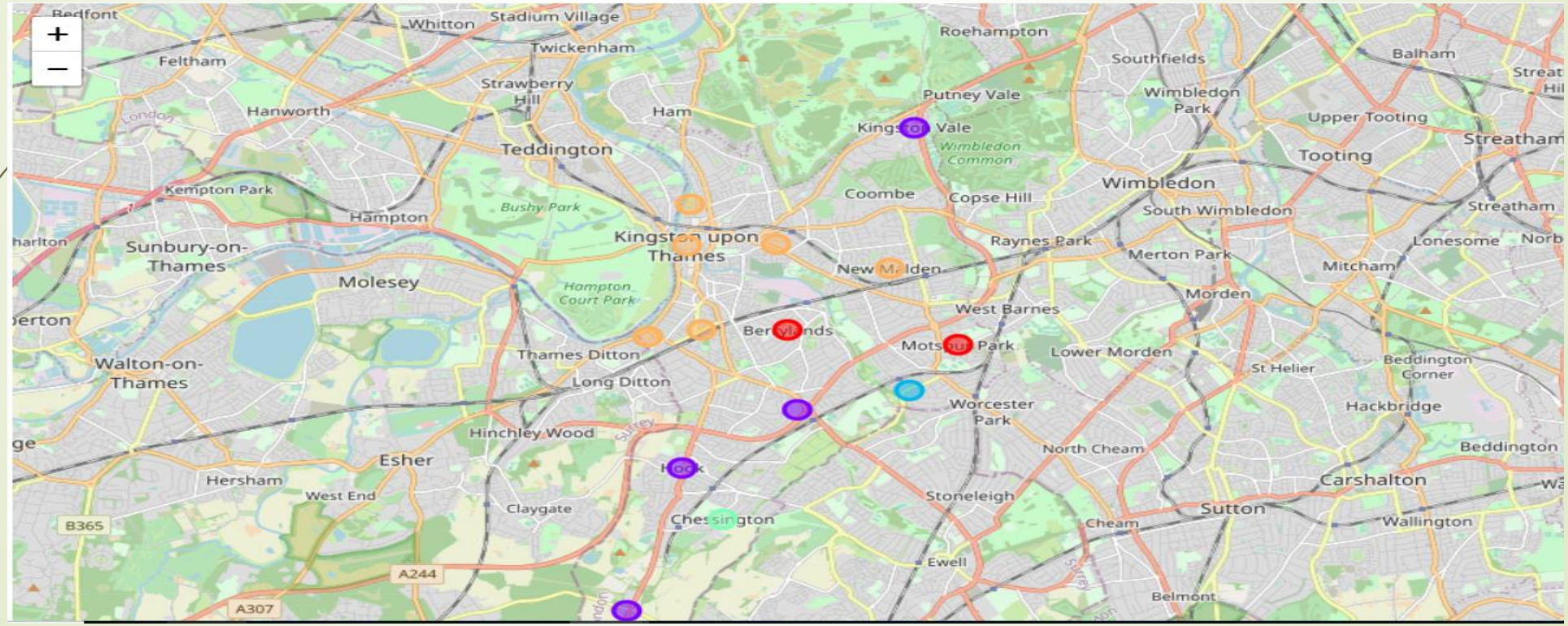
- Using the final dataset containing the neighborhoods in Kingston upon Thames along with the latitude and longitude, we can find all the venues within a 500 meter radius of each neighborhood by connecting to the Foursquare API

	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	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

5. Results

After running the K-means clustering we can access each cluster created to see which neighborhoods were assigned to each of the five clusters. Visualising the clustered neighborhoods on a map using the folium library



6. Discussions

The aim of this project is to help people who want to relocate to the safest borough in London, people 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 has Train stations 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 second cluster and fifth cluster is suitable. For a family the neighborhoods in Cluster 5 is more suitable dues to the common venues in that cluster, these neighborhoods have common venues such as Parks, Gym/Fitness centers, Restaurants, pubs which is ideal for a family.

7. Conslusions

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.