IBM Competition

London Neighborhood Analysis

Shiyou Li April 2020

1. Introduction

- **Backgroud:** Safety is a top concern when moving to a new area.

- **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 relocating to London will be interested to identify the safest borough in London and explore its neighborhoods 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 different sources

- The first data source of the project uses London crime data that show the crime records per borough in London
- The second data source is scraped from a Wikipedia page that contains the list of London borough
- The third data source is the list of Neighborhoods in the Royal Borough of Kingston upon Thames as found on the webpage

Data cleaning:

- From the London crime data, the major categories of crimes are pivoted to get the total crimes per boroughs for each major category
- The second data is scraped from Wikipedia page using the Beautiful Soup library in python. Using this library we can extract the data in table format as shown in the website
- The two above datasets are merged together on Borough. The purpose of this dataset is to visualize the crime rate in each borough and identify the borough with the least crimes recorded during the year 2008 to year 2016
- The third dataset is created with the names of the neighborhood and the names of the safest borough with the latitude and longitude obtained using Google Maps API geocoding

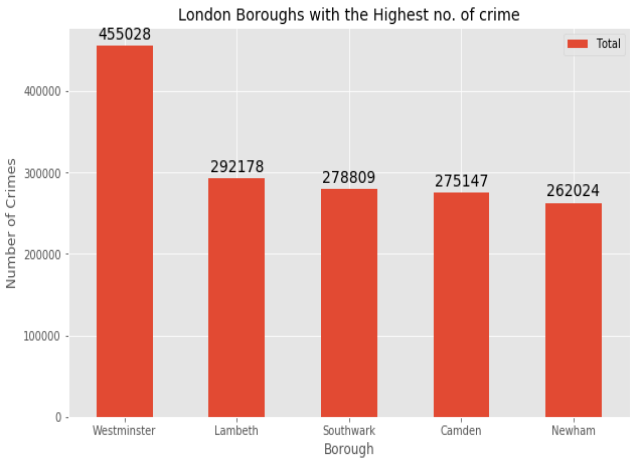
3. Methodology

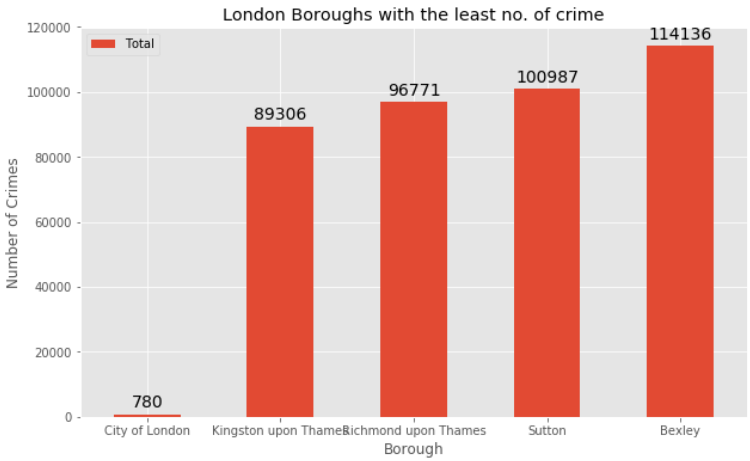
- Exploratory Data Analysis

	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.

- Exploratory Data Analysis

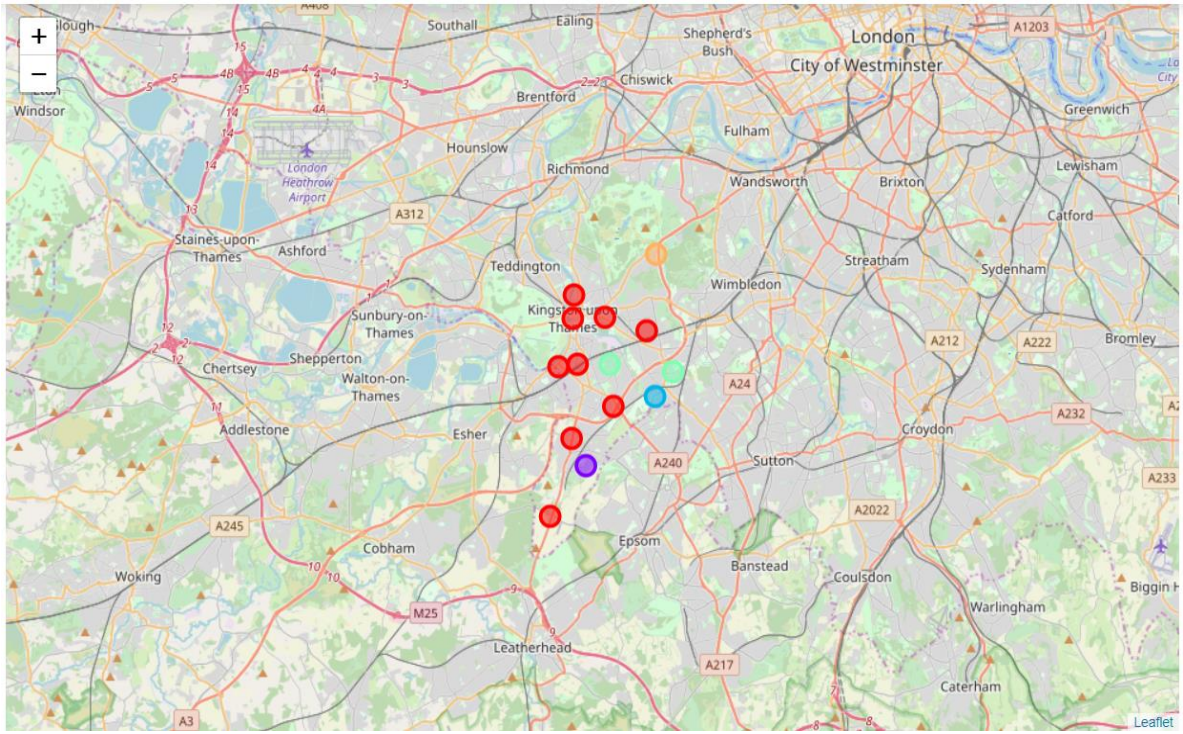




City of London has a significantly lower crime rate because it i is the 33rd principal division of Greater London but it is not a London borough. It has an area of 1.12 square miles and a population of 7000 as of 2013 which suggests that it is a small area (*see fig 3.1.3.1*). Hence, we will consider the next borough with the lowest crime rate as the safest borough in London which is Kingston upon Thames.

- Modeling

Basically, we use k-means to cluster the 15 neighborhoods to 5 clusters based on the most common venues in the neighborhood



Five different colors represent for different clusters

4.Reuslts

- 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.
- The second cluster has one neighborhood which consists of Venues such as Restaurants, Golf courses, and wine shops.
- The third cluster has one neighborhood which consists of Venues such as Train stations, Restaurants, and Furniture shops.
- The fourth cluster has two neighborhoods in it, these neighborhoods have common venues such as Parks, Gym/Fitness centers, Bus Stops, Restaurants, Electronics Stores and Soccer fields
- The fifth cluster has one neighborhood which consists of Venues such as Grocery shops, Bars, Restaurants, Furniture shops, and Department stores.

5. Recommendation

- 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.