CAPSTONE PROJECT - THE BATTLE OF NEIGHBORHOODS REPORT

Dhivakar R

Introduction

Background

The average American moves about eleven times in their lifetime. This brings us to the question: Do people move until they find a place to settle down where they truly feel happy, or do our wants and needs change over time, prompting us to eventually leave a town we once called home for a new area that will bring us satisfaction? Or, do we too often move to a new area without knowing exactly what we're getting into, forcing us to turn tail and run at the first sign of discomfort? To minimize the chances of this happening, we should always do proper research when planning our next move in life. Consider the following factors when picking a new place to live so you don't end up wasting your valuable time and money making a move you'll end up regretting. 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

The crime statistics dataset of London found on Kaggle has crimes in each Boroughs of London from 2008 to 2016. The year 2016 beingthe latest we will be considering the data of that year which is actually old information as of now. The crime rates in each borough may have changed over time. 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.
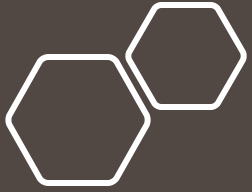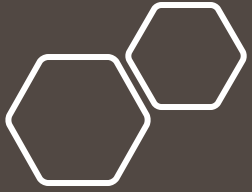
Data Acquisition and Cleaning

2.1 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 soruce of data is scraped from Wikipedia page that contains the list of London Boroughs.

- The Third data source is the list of Neighborhoods in the Royal Borough of the Kingston upon Thomas as found on the Wikipedia Page.

Data Acquisition and Cleaning - continued

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 secord data is scraped from Wikipedia page using the Beautiful Soup Library in python.

The two datasets are merged on the Borough names to form a new dataset.

After visualizing the crime in each borough we can find the borough with the lowest crime rate.

The new dataset is used to generate the 10 most common venues for each neighborhood using the Foresquare API, finally using K mean clustering algorith to lcluster similar neighborhood together.

Methodology

Exploratory Data Analysis

- The count for each of the major categories of crime returns the value 33 which is the number of London borough. 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.

| [65]:	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	397.696970	371.787879	229.363636	87.787879	126.575758	1670.000000	1322.575758	4205.787879
std	141.488623	116.699335	124.123129	37.796956	81.076750	869.599908	468.022504	1675.481139
min	0.000000	1.000000	1.000000	0.000000	2.000000	27.000000	5.000000	36.000000
25%	284.000000	320.000000	151.000000	64.000000	59.000000	1118.000000	1116.000000	3285.000000
50%	415.000000	383.000000	214.000000	84.000000	117.000000	1665.000000	1395.000000	4313.000000
75%	484.000000	442.000000	302.000000	113.000000	178.000000	2004.000000	1647.000000	5213.000000
max	670.000000	608.000000	613.000000	186.000000	311.000000	5145.000000	2071.000000	9099.000000

Boroughs with the highest crime rate

- Comparing the five boroughs with the highest crime rate during the year 2016 it is evident that Westminster has the highest crimes recorded followed b Lambeth, Southwark, Newham and Tower Hamlets. Westminster has a significantly higher crime rate than the other 4 boroughs.

London Boroughs with the Highest no. of crime
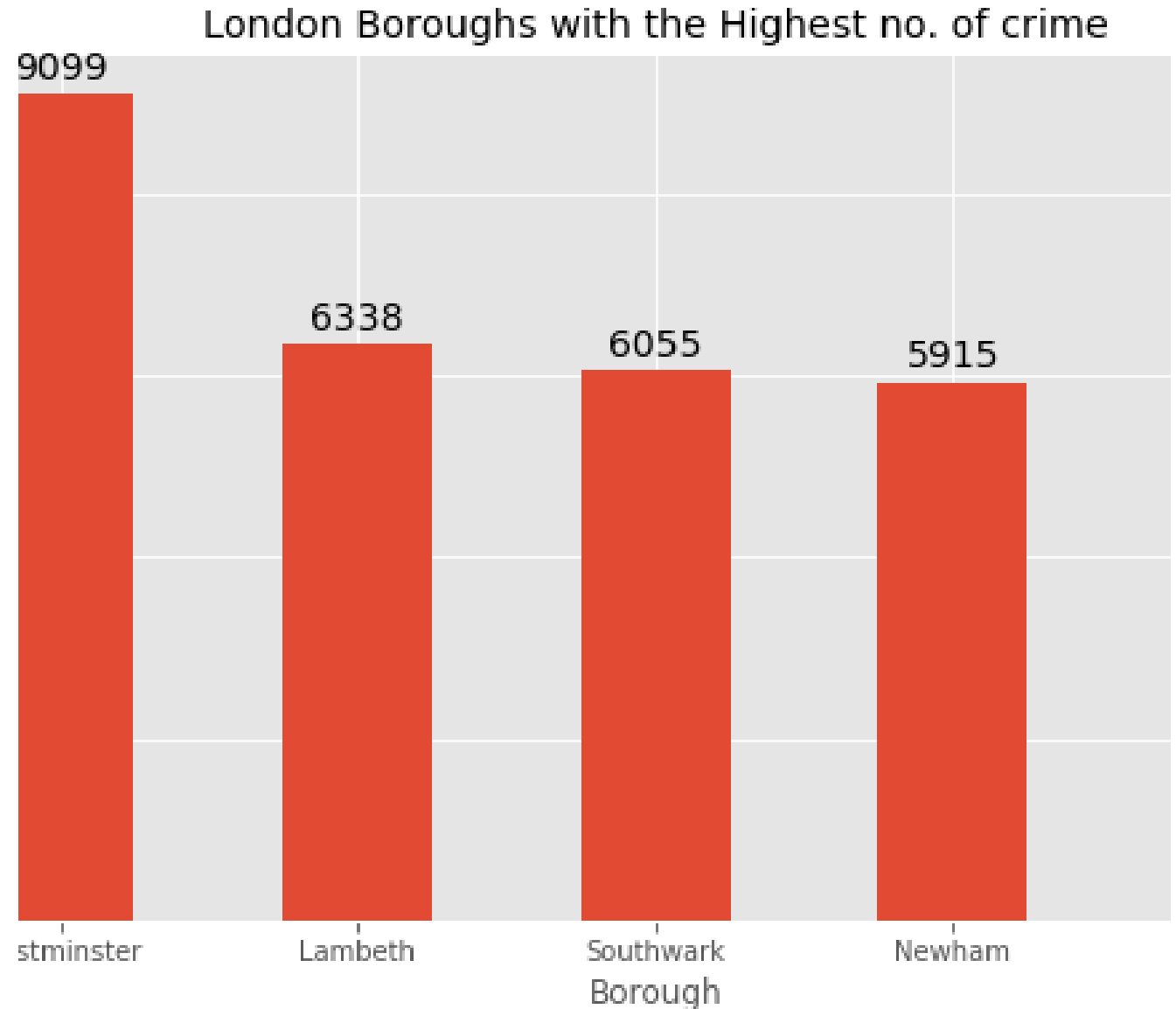
Boroughs with the lowest crime rate

- Comparing the 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.

- City of London has 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.

- We will consider the next borough with the lowest crime rate as the safest borough in London which is Kingston upon Thames.

London Boroughs with the least no. of crime

Borough	Value
Kingston upon Thames	1754
Richmond upon Thames	2016
Sutton	2043
Harrow	2352
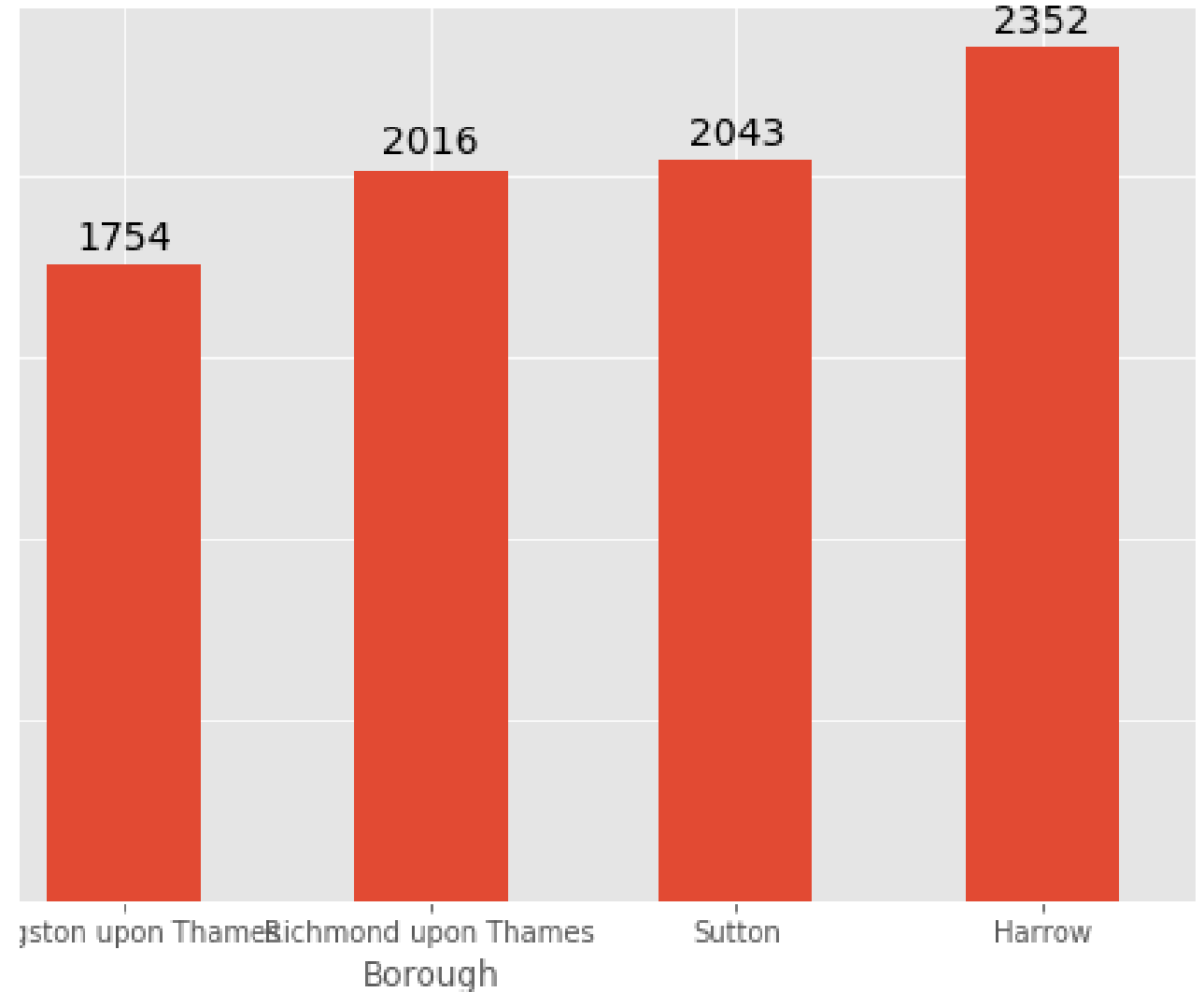
Neighborhoods in Kingston upon Thames
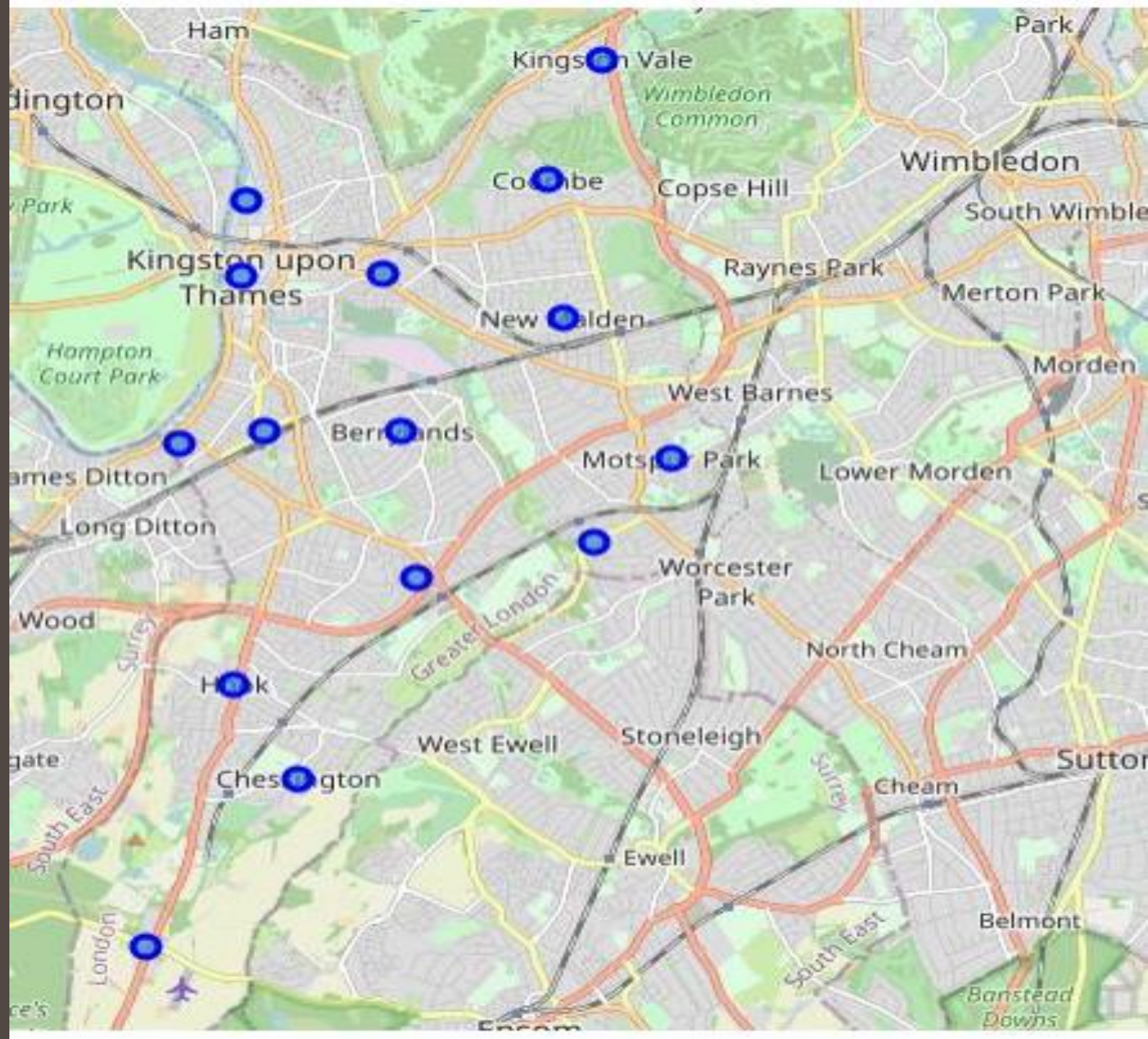
Modelling

- Using the final dataset containing the neighborhoods in Kingston upon Thames along with the latitude and longitude, we can find all the venues with in 500 meters radius of each neighborhood.

- One hot encoding is done on the venues data. The venues data is then grouped by the neighborhood and the mean of the venues is calculated, finally the 10 common values are calculated for each of the neighborhood.

	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	Canbury	51.417499	-0.305553	Canbury Gardens	51.417409	-0.305300	Park
4	Canbury	51.417499	-0.305553	The Boater's Inn	51.418546	-0.305915	Pub

Results

- After running the K-means clustering we can access each cluster to see which neighborhood were assigned to each other of the five clusters.

- Visualizing the clustered neighborhood on a map using folium library.

- Each cluster is color coded for the ease of population, we can see neighborhood failling in red cluster.
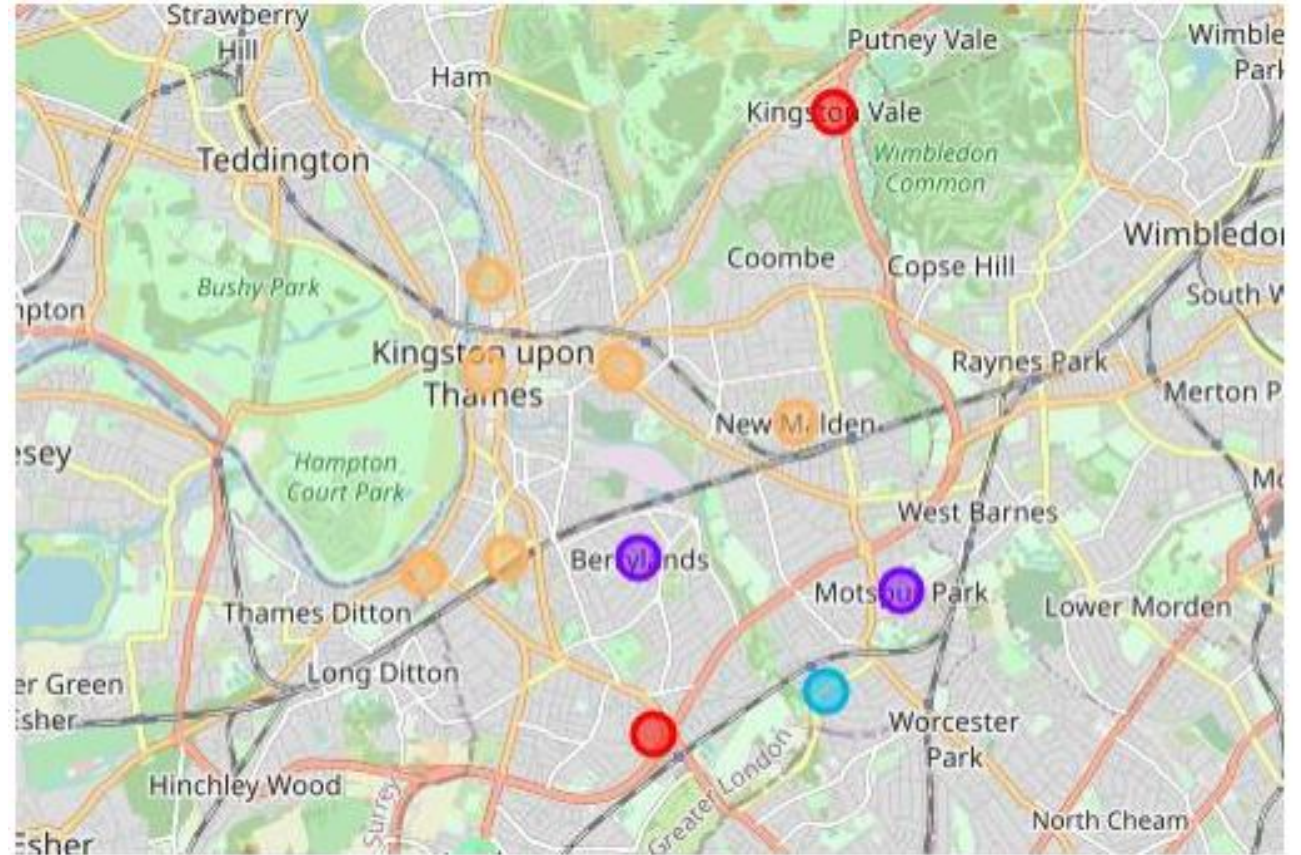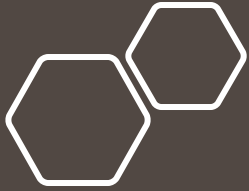
Discussion

- The first data source of the project uses a London crime data that shows the crime per 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 venue may vary from person to person, they can select a neighborhood based on one's preference.

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.