CAPSTONE PROJECT – IBM THE BATTLE OF NEIGHBORHOODS

IDDI ABDUL AZIZ

MARCH, 2020

1. INTRODUCTION

- 1.1 BACKGROUND: What are the best and safest places to live in London? This is a question that will be asked by many people as they prepare for their next big move. However, what may be the best place in London for someone else, might not be the best place for you. It is a big decision that requires a lot of research. But we all know that; serene and safety environs is what everyone would think of when locating to a new place. Families, ambitious young professionals, and those seeking an adventure all yearns for a safety, comfortable, and peaceful environments. Safety they say brings first aid to the uninjured.
- In our hunt for an apartment, Edward Coke's dictum "Precaution is better than cure" reminds that safety is deemed prudent and a top concern when relocating to a new place. There is no delight when you don't feel safe at home.
- 1.2 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 ten (10) most common venues in each neighborhood and finally cluster the neighborhoods using k-mean clustering.
- 1.3 INTEREST: This project is would be of interest to families looking for new homes, expats, stakeholders who offer housing to families or city planners who are looking to make neighborhoods safer. Additionally, Police forces and safety officials would benefit from knowing which places have higher crimes so they can implement more resources in those locations to lower crime.

2. DATA

- 2.1 DATA UTILIZED:

For this project we need data about:

- London Crime Rates obtained from Kaggle;
- List of London boroughs; obtained and scraped from a Wikipedia page.
- List of neighborhoods in the Royal Borough of Kingston upon Thames also from Wikipedia page

2.2 DATA CLEANING

- *London Crime Data:*

Data cleaning of the three (3) data sets were done separately; in the London crime data, the latest year (2016), is only selected. The major categories of crime are swiveLled to get the total crimes per the boroughs for each major category.

- *List of London Boroughs (Wikipedia):*

The London boroughs data set taken from Wikipedia is scraped using BeautifulSoup library in python. This library is used in extracting the data in tabular format as in the Wikipedia website. String manipulation is obvious after scraping, to get names of the boroughs in the exact format. Its very essential to as we will be merging the two (2) datasets together using Borough names.

DATA CONTI'

- *Datasets Merged (London Crime Data & List of London Boroughs):*

In other to gain the necessary information we need; the two datasets are merged into one dataset *(Ld_crime)*. In other to visualize the crime rates in each borough and identify the borough with the least crimes recorded during the year 2016.

- *Safest Borough:*

he borough with the lowest crime rate, automatically is the safest borough; after we visualized the crime rate in each borough. The third set of data is created from scratch, with pandas data frame the names of the neighborhoods and the name of the borough with the latitude left blank. Coordinates of the neighborhood is to be gotten from Google Maps API geocoding to get the final dataset.

- The Foursquare API will be used to generate and obtain the 10 most common venues for each neighborhood, K-Mean clustering algorithm will be used finally to cluster similar neighborhoods together.

3. METHODOLOGY

- The methodology in this project consists of two parts:
- 3.1 *Exploratory Data Analysis:* Visualise the crime rates in the London boroughs to identify the safest borough and extract the neighborhoods in that borough to find the 10 most common venues in each neighborhood.

- 3.2 *Modelling:* 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(ML) 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.

3. METHODOLGY CONTI'

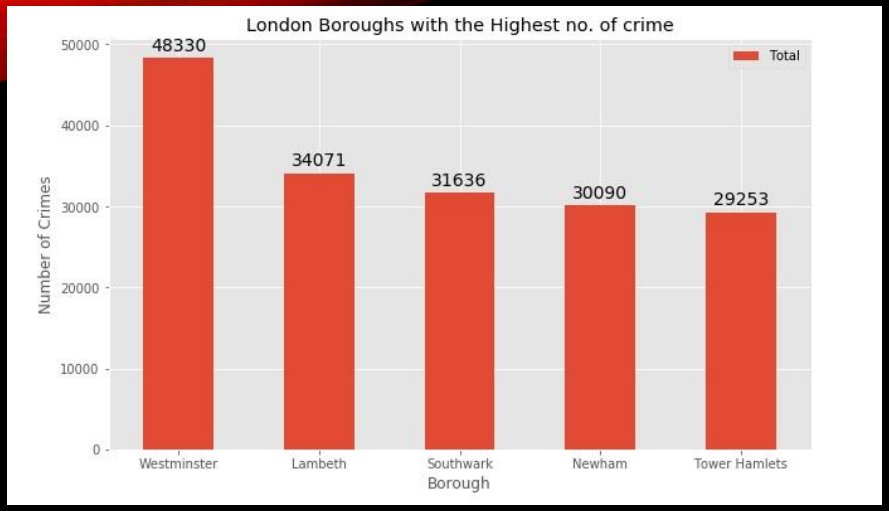
[19]:

	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

- EXPLORATORY DATA ANALYSIS:
- *Statistical Summary of Crimes:*
- To get the statistical information of London crime data, we use the describe function in python. This returns the mean, standard deviation, minimum, maximum, 1st quartile (25%), 2nd quartile (50%), and the 3rd quartile (75%) for each of the major categories of crime.

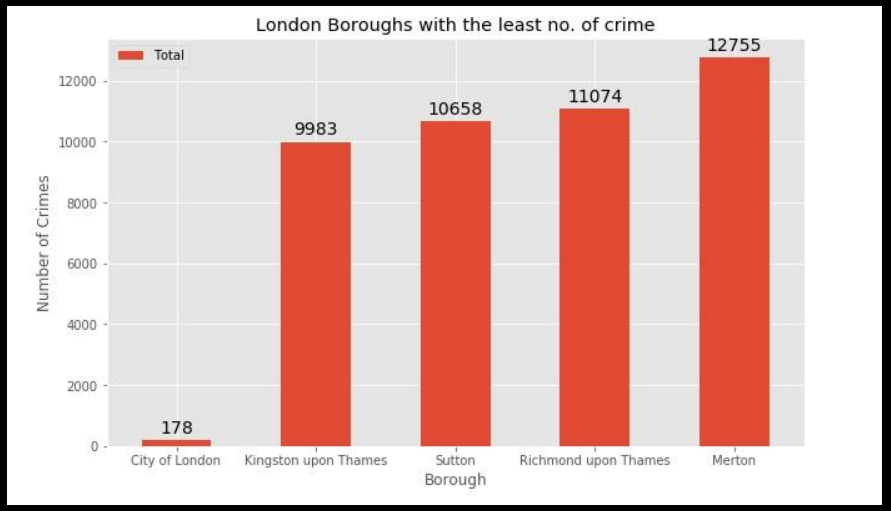
- From the table, the count for each major categories of crime returns the value of 33 which is the number of London boroughs. 'Theft and Handling' is the highest reported crime during the year 2016 followed by 'Violence against the person', 'Criminal damage'. The lowest recorded crimes are 'Drugs', 'Robbery' and 'Other Notifiable offenses'.

BOROUGH WITH THE HIGHEST CRIME RATES



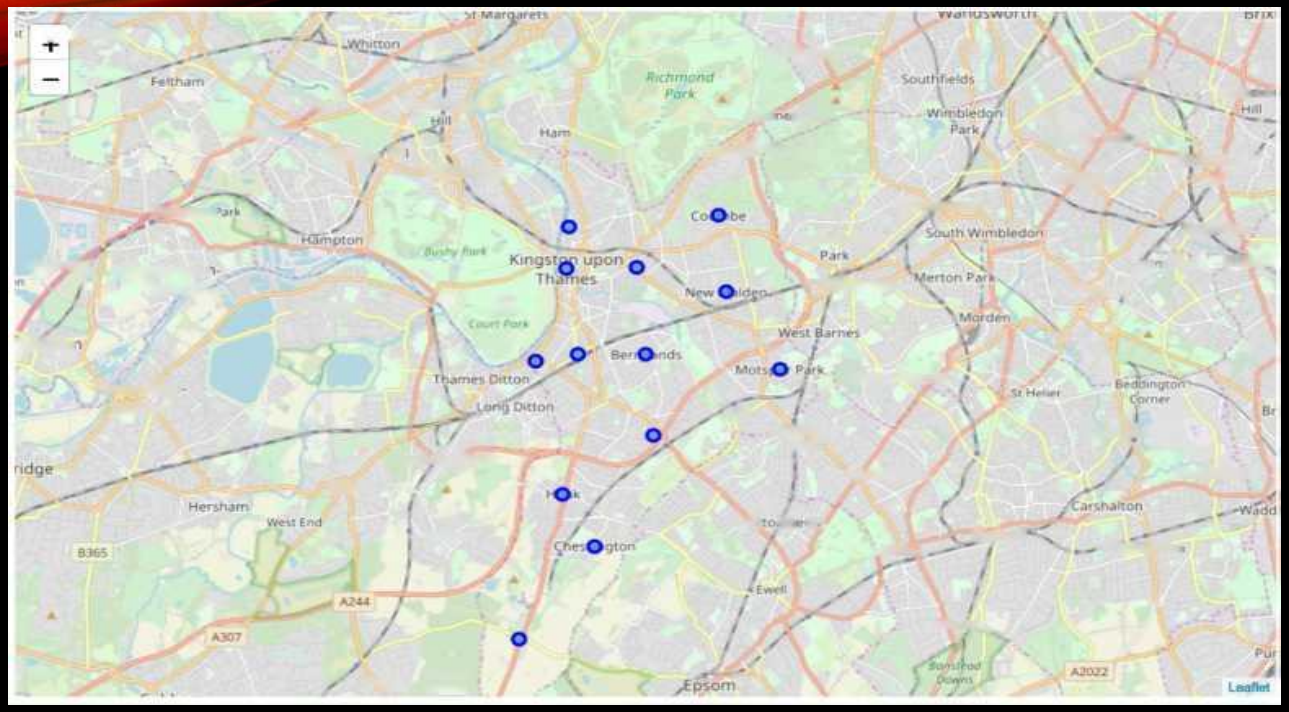
In comparing the five (5) boroughs with highest crime rate during the year (2016) is quite evident that the Westminster has the highest crimes recorded followed by Lambeth, Southwark, Newham and Tower Hamlets. Westminster has significantly higher crimes rate than the other 4 boroughs.

BOROUGH WITH THE HIGHEST CRIME RATES



It is evident that in comparing the five (5) boroughs with 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

KINGSTON UPON THAMES NEIGHBORHOODS



- *Kingston upon Thames Neighborhoods:*

Fifteen (15) neighborhoods are in the royal borough of Kingston upon Thames. Below is a visualized map using folium on python

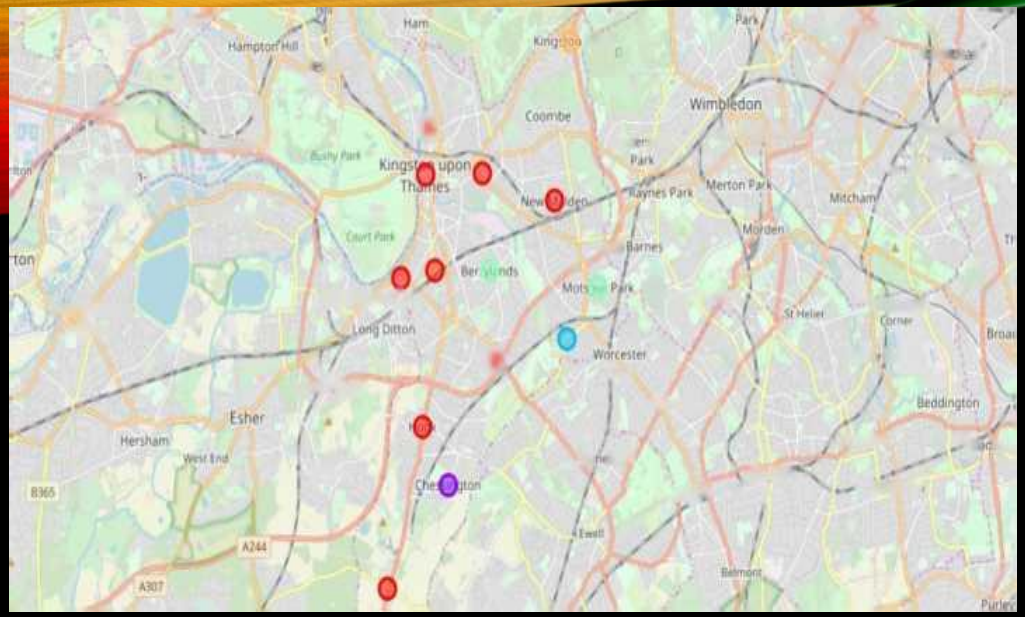
MODELLING

[68]:

	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 Grey Horse	51.414192	-0.300759	Pub

- In helping interested folks find similar neighborhoods in the safest borough we will be clustering similar neighbourhoods using K-means clustering which is form of unsupervised ML 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 five (5) clusters. Reason is to conduct a K-means clustering to cluster neighborhoods with similar venues together so that people can shortlist the area of their interests based on the amenities (parks, playgrounds, Libraries, coffee, etc) around each neighborhood.

- We will be using the final dataset containing the neighborhoods in Kingston upon Thames along with the latitude and longitude, with that we can find all the venues within a 500-meter radius of each neighborhood by connecting to the Foursquare API. This returns a json file containing all the venues in each neighborhood which is converted to a pandas data frame. This data frame contains all the venues along with their coordinates and categories.
- 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.
- A process by which categorical variables are converted into a form that could be provided to Machine Learning algorithms to do a better job in prediction.



- After running the K-means clustering we can access each cluster created to see which neighborhood were assigned to each of the five clusters. Looking into the neighborhoods in the first cluster

RESULTS

- 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 (1st) cluster. Three neighborhoods have their own cluster (Blue, Purple and Yellow), these are clusters two (2), three (3) and five (5). The green cluster consists of two neighborhoods which is the fourth(4th) cluster.

[86]:

	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	9th Most Common Venue	10th Most Common Venue
1	Canbury	Kingston upon Thames	51.417499	-0.305553	0	Pub	Café	Spa	Indian Restaurant	Plaza	Hotel	Shop & Service	Gym / Fitness Center	Supermarket	Fish & Chips Shop
4	Hook	Kingston upon Thames	51.367898	-0.307145	0	Fish & Chips Shop	Indian Restaurant	Bakery	Supermarket	Deli / Bodega	Department Store	Discount Store	Dry Cleaner	Electronics Store	Farmers Market
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	0	Coffee Shop	Café	Pub	Sushi Restaurant	Burger Joint	Department Store	German Restaurant	Furniture / Home Store	French Restaurant	Electronics Store
9	New Malden	Kingston upon Thames	51.405335	-0.263407	0	Gastropub	Chinese Restaurant	Korean Restaurant	Sushi Restaurant	Supermarket	Bar	Gym	Indian Restaurant	Electronics Store	Department Store
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	0	Indian Restaurant	Food	Pub	Italian Restaurant	Fried Chicken Joint	Dry Cleaner	Grocery Store	Hardware Store	Hotel	Japanese Restaurant
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	0	Indian Restaurant	Café	Coffee Shop	Pub	Pet Café	Fish & Chips Shop	Fast Food Restaurant	Golf Course	Chinese Restaurant	Gym
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	0	Coffee Shop	Pub	Pharmacy	Grocery Store	Italian Restaurant	Train Station	French Restaurant	Pizza Place	Breakfast Spot	Deli / Bodega

DISCUSSION

- With more neighborhoods to choose from the clusters, it should be relatively easy for and of interest to families looking for new homes, expats, stakeholders who offer housing to families or city planners who are looking to make neighborhoods safer based on their preferences.
- 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 third (3rd) and fourth (4th) 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(1st) cluster is suitable.
- For a family I feel that the neighborhoods in Cluster fourth (4th) 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 choices of neighborhoods may vary from person to person.
- In areas of high crime rates should be avoided by families. But those in charge of police forces (Police Commissioners) and safety would benefit from implementing more resources into these areas to bring the crime down.

CONCLUSION

- We conclude that there are many safest and family-friendly neighborhoods to choose from. Which is great news as many people believe it is impossible to find safer and family-friendly neighborhoods in Big Cities. Now that you have read about the neighborhoods of London and several of the areas the city has to offer, I hope some of your stress is relieved. Remember to think deeply about what is right for you and your situation, and not get caught up in generalizations. Without a doubt, London is a great place to live, work, and raise a family.

- Before making the big move, 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. Do your research and ask people who have lived there. But, above all remember:

It's not the place. It's the people.

Good luck with your choice!