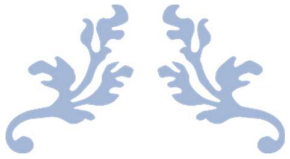


IBM DATA SCIENCE PROFESSIONAL CERTIFICATE

Capstone Project Report - The Battle of Neighbourhoods (London)

AUGUST 15, 2020

UTKARSH SHARMA
London, UK

1. Introduction

A. Context

In the year ending December 2019, 677,000 people migrated into the UK*. The concentration of this migrant population is highest for London*. This is a significant number of new residents and poses a big challenge for the migrants as well as the host country UK. Migrating in itself brings a great deal of anxiety and uncertainty due to an alien environment and the associated cultural shock. If the home in the new place is also not in a safe neighbourhood, the transition can be painful as well. A proper research can reduce the risk of accidently residing in a threatening neighbourhood.

* Source - https://commonslibrary.parliament.uk/research-briefings/sn06077/)

B. Business Problem

The purpose of this project is to analyse the boroughs of London (UK) for safety based upon the criminal records available in public domain. This analysis would be helpful to those planning to move to London for pursuing their ambitions. It will provide them with an unbiased report on where to rent or buy a place to live, assuming that safety is the topmost priority for anyone moving to a new place.

Once the safest borough is identified by the analysis, an attempt will be made to explore the top neighbourhoods in that borough. This will provide more information to a new person for selecting a specific neighbourhood to reside based upon his / her individual preferences and tastes.

C. Target Audience

- Professionals from other countries shifting to London as part of their international assignments.
- Professionals from other countries shifting to London for beginning a new job.
- Overseas students shifting to London for pursuing higher education in one of the many London universities (London School of Business, Imperial College of London, etc.).
- Entrepreneurs from other countries shifting to London to try their luck in the UK business environment.

2. Data Acquisition

A. Description of Sources

The below data will be required to solve the aforementioned business problem: -

I. Criminal records for the city of London
- Source of data - Real world dataset from Kaggle providing details about crimes in London - https://www.kaggle.com/jboysen/london-crime.
- Comprehensive description of the data - The dataset from Kaggle consists of London Crime Data from 2008 to 2016. The csv file has a total 13000 rows of criminal data for all boroughs of London. The columns in the csv file are listed below with description of what each type of heading means: -
  - lsoa_code: Code for Lower Super Output Area in Greater London.
  - borough: Common name for London borough.
  - major_category: High level categorization of crime.
  - minor_category: Low level categorization of crime within major category.
  - value: Monthly reported count of categorical crime in given borough.
  - year: Year of reported counts, 2008-2016.
  - month: Month of reported counts, 1-12.
- Example of what features can be extracted from the data -
  - The latest available data ('year' 2016) can be extracted from the csv file to provide latest crime statistics for London.
  - The 'value' column can be combined borough-wise to give a direct indication of which boroughs have the minimum crimes recorded for the year 2016.
  - The 'major_category' and 'value' columns can be used to generate a borough-wise DataFrame for further analysis.

II. Neighbourhood details for London boroughs
- Source of data (part 1) - Information about London's boroughs from Wikipedia - https://en.wikipedia.org/wiki/List_of_London_boroughs.
- Source of data (part 2) - Information about neighbourhoods from Wikipedia for the safest borough identified from the analysis - https://en.wikipedia.org/wiki/List_of_districts_in_the_Royal_Borough_of_Kingston_upon_Thames
- Comprehensive description of the data - Part 1 data source from Wikipedia provides the details of London's boroughs and local authorities. Part 2 data source provides the neighbourhoods (districts) of a particular borough. The columns in the Wikipedia table (part 1 data source) are listed below with description of what each type of heading means: -
  - Borough: Provides the names of the administrative regions of London.
  - Inner: Lists which boroughs fall within the inner administrative boundary of London.
  - Status: Provides royal or city status of a borough, if any.
  - Local authority: Provides the name of the council responsible for administration of a particular borough.
  - Political control: Provides the name of the political party in control of a particular borough.
  - Headquarters: Provides the address of the council office.
  - Area (sq. mi): Provides the area of the borough in square miles.
  - Population (2013 est.): Provides the population of the borough.
  - Coordinates: Provides the coordinates of the borough.
  - Nr. in map: Provides the designated map number for the borough.

- Example of what features can be extracted from the data -
  - The borough-wise table can be converted to a DataFrame.
  - This DataFrame can then be combined with the Kaggle DataFrame for further analysis.
  - The 'Coordinates' column can be used to analyse the criminal record.

III. Venues' information from Foursquare for neighbourhoods

- Source of data - Information from Foursquare for venues in the neighbourhoods of the safest borough identified from the analysis.
- Comprehensive description of the data - The Foursquare database can be used to generate a list for venues, their latitudes, and longitudes as well as venue categories. This can then be used to generate clusters of neighbourhoods which have common venues, which can further help a new person in shortlisting a neighbourhood to reside within the safest council. A list of possible venues that can be generated is provided below: -
  - Pub
  - Train station
  - Supermarkets
  - Stables
  - Parks
  - Restaurants
  - Grocery store
  - Bus station
  - Coffee shops
- Example of what features can be extracted from the data -
  - Clusters of boroughs can be generated which are closest to parks.
  - Clusters of boroughs can be generated which are closest to bus or train stations.
  - Clusters of boroughs can be generated which are closest to coffee shops, pubs, or restaurants.

3. Methodology

A. Pre-processing

The aforementioned datasets were pre-processed before attempting the analysis of the data: -

I. Criminal records for the city of London

The Kaggle London Crime data was read from the csv file and converted into a Pandas DataFrame: -

https://www.kaggle.com/jboysen/london-crime

[2]:

	lsoa_code	borough	major_category	minor_category	value	year	month
0	E01001116	Croydon	Burglary	Burglary in Other Buildings	0	2016	11
1	E01001646	Greenwich	Violence Against the Person	Other violence	0	2016	11
2	E01000677	Bromley	Violence Against the Person	Other violence	0	2015	5
3	E01003774	Redbridge	Burglary	Burglary in Other Buildings	0	2016	3
4	E01004563	Wandsworth	Robbery	Personal Property	0	2008	6

Snapshot of the DataFrame created from the Kaggle London Crime data (2008-2016)

The DataFrame was then processed to extract only the details of 2016 crimes (latest available data used for analysis):

[7]:

	lsoa_code	borough	major_category	minor_category	value	year	month
0	E01004177	Sutton	Theft and Handling	Theft/Taking of Pedal Cycle	1	2016	8
1	E01000733	Bromley	Criminal Damage	Criminal Damage To Motor Vehicle	1	2016	4
2	E01003989	Southwark	Theft and Handling	Theft From Shops	4	2016	8
3	E01002276	Havering	Burglary	Burglary in a Dwelling	1	2016	8
4	E01003674	Redbridge	Drugs	Possession Of Drugs	2	2016	11

Snapshot of the processed Kaggle London Crime data (2016)

The above processed DataFrame was then used to create a pivoted DataFrame with rows a crime numbers for all London boroughs and columns as major crimes' categories:

[17]:

	Borough	burglary	criminal_damage	drugs	other_notifiable_offences	robbery	theft_and_handling	violence_against_the_person	total
0	Barking and Dagenham	1287	1949	919	378	534	5607	6067	16741
1	Barnet	3402	2183	906	499	464	9731	7499	24684
2	Bexley	1123	1673	646	294	209	4392	4503	12840
3	Brent	2631	2280	2096	536	919	9026	9205	26693
4	Bromley	2214	2202	728	417	369	7584	6650	20164

Snapshot of the DataFrame created by using the Pandas pivot table function

II. Neighbourhood details for London boroughs

Part 1 - The Wikipedia link for information on London boroughs was scraped and converted to a DataFrame by using the **Beautifulsoup** library: -

https://en.wikipedia.org/wiki/List_of_London_boroughs

[26]:

	Borough	Inner	Status	Local authority	Political control	Headquarters	Area (sq mi)	Population (2013 est) [1]	Co-ordinates	Nr. in map
0	Barking and Dagenham	NaN	NaN	Barking and Dagenham London Borough Council	Labour	Town Hall, 1 Town Square	13.93	194352	51°33'39"N 0°09'21"E / 51.5607°N 0.1557°E	25
1	Barnet	NaN	NaN	Barnet London Borough Council	Conservative	Barnet House, 2 Bristol Avenue, Colindale	33.49	369088	51°37'31"N 0°09'06"W / 51.6252°N 0.1517°W	31
2	Bexley	NaN	NaN	Bexley London Borough Council	Conservative	Civic Offices, 2 Watling Street	23.38	236687	51°27'18"N 0°09'02"E / 51.4549°N 0.1505°E	23

Snapshot of the DataFrame created by scraping the Wikipedia page

This DataFrame was then merged with the Kaggle crime DataFrame using the common column 'Borough' for further data analysis.

Part 2 - The Wikipedia link for neighbourhoods in Kingston Upon Thames borough was used to create a DataFrame focussed on neighbourhoods: -

https://en.wikipedia.org/wiki/List_of_districts_in_the_Royal_Borough_of_Kingston_upon_Thames

[45]:

	Neighbourhood	Borough	Latitudes	Longitudes
0	Berrylands	Kingston upon Thames	51.393781	-0.284802
1	Canbury	Kingston upon Thames	51.417499	-0.305553
2	Chessington	Kingston upon Thames	51.358336	-0.298622
3	Coombe	Kingston upon Thames	51.419450	-0.265398
4	Hook	Kingston upon Thames	51.367898	-0.307145

Snapshot of the DataFrame created by using information from the Wikipedia page

B. Exploratory Data Analysis

I. Summary of crime statistics

The describe function of Pandas DataFrame was used to obtain the following statistics of the London crime data: -

- Count
- Mean
- Standard Deviation
- Minimum
- 1st Quartile
- 2nd Quartile
- 3rd Quartile
- Maximum

[36]:

	burglary	criminal_damage	drugs	other_notifiable_offences	robbery	theft_and_handling	violence_against_the_person	total	area_sq_mi	pc
count	33.000000	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	18.408485	
std	737.448644	625.207070	586.406416	223.298698	441.425366	4620.565054	2513.601551	8828.228749	12.645367	
min	2.000000	2.000000	10.000000	6.000000	4.000000	129.000000	25.000000	178.000000	1.120000	
25%	1531.000000	1650.000000	743.000000	378.000000	377.000000	5919.000000	5936.000000	16903.000000	10.360000	
50%	2071.000000	1989.000000	1063.000000	490.000000	599.000000	8925.000000	7409.000000	22730.000000	14.520000	
75%	2631.000000	2351.000000	1617.000000	551.000000	936.000000	10789.000000	8832.000000	27174.000000	21.780000	
max	3402.000000	3219.000000	2738.000000	1305.000000	1822.000000	27520.000000	10834.000000	48330.000000	57.970000	

Snapshot of the statistical criminal summary of London

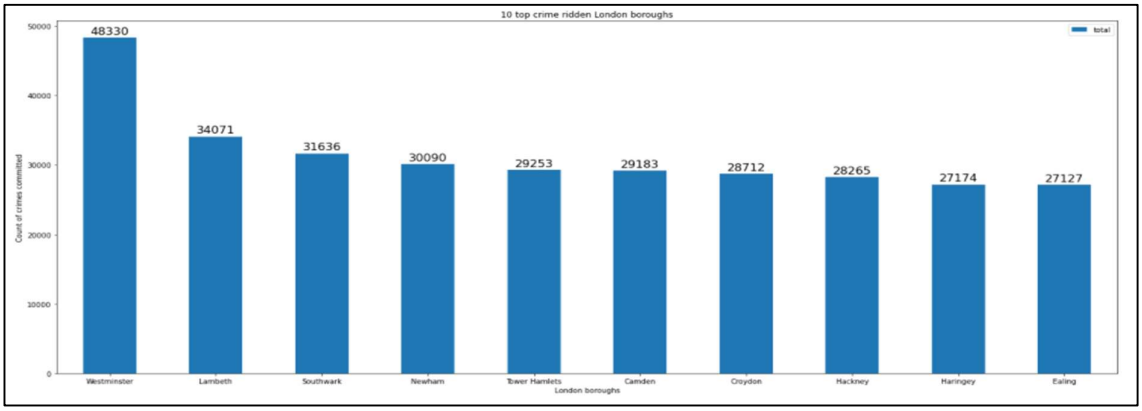
The following points can be deduced from the above table: -

- The count of 33 represents the total number of boroughs in London
- The major categories with decreasing number of occurrences are listed below: -
  - **Theft and handling (highest)**
  - Violence against the person
  - Burglary
  - Criminal damage
  - Drugs
  - Robbery
  - **Other notable offences (lowest)**

II. Boroughs with the highest crime rates

The 10 most crime ridden boroughs based upon the analysis are listed below, followed by a bar chart:-

- **Westminster (maximum 48330 crimes in 2016)**
- Lambeth
- Southwark
- Newham
- Tower Hamlets
- Camden
- Croydon
- Hackney
- Haringey
- Ealing

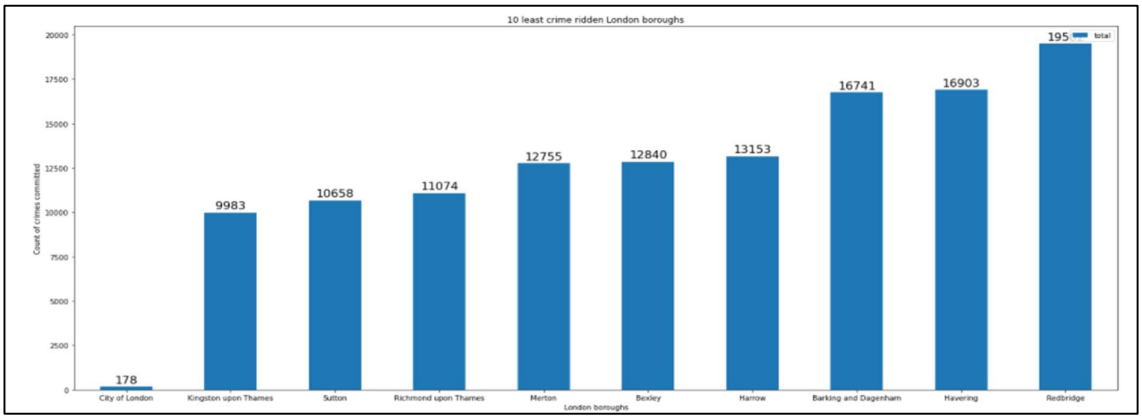


Snapshot of 10 most crime ridden boroughs

III. Boroughs with the lowest crime rates

The 10 least crime ridden boroughs based upon the analysis are listed below, followed by a bar chart:-

- **City of London (minimum 178 crimes in 2016) - principal division and not a London borough as per Wikipedia page -** https://en.wikipedia.org/wiki/List_of_London_boroughs
- **Kingston upon Thames (next least 9983 crimes in 2016) - It is a borough and hence considered in further analysis**
- Sutton
- Richmond upon Thames
- Merton
- Bexley
- Harrow
- Barking and Dagenham
- Havering
- Redbridge

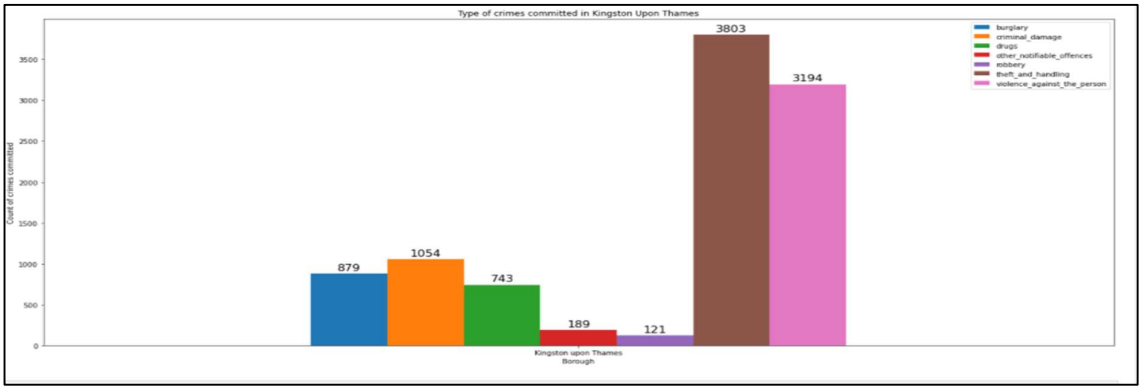


Snapshot of 10 least crime ridden boroughs

IV. Categorical distribution of crimes in Kingston upon Thames (lowest crimes borough)

The top 3 categories with decreasing number of occurrences are listed below: -

- **Theft and handling (highest - 3803 crimes)**
- Violence against the person
- Criminal damage



Snapshot of categorical distribution of crimes in Kingston upon Thames

V. Distribution map of the 15 neighbourhoods of Kingston upon Thames

The 15 neighbourhoods of Kingston upon Thames (borough with the least number of crimes) were identified from the Wikipedia page: -

https://en.wikipedia.org/wiki/List_of_districts_in_the_Royal_Borough_of_Kingston_upon_Thames

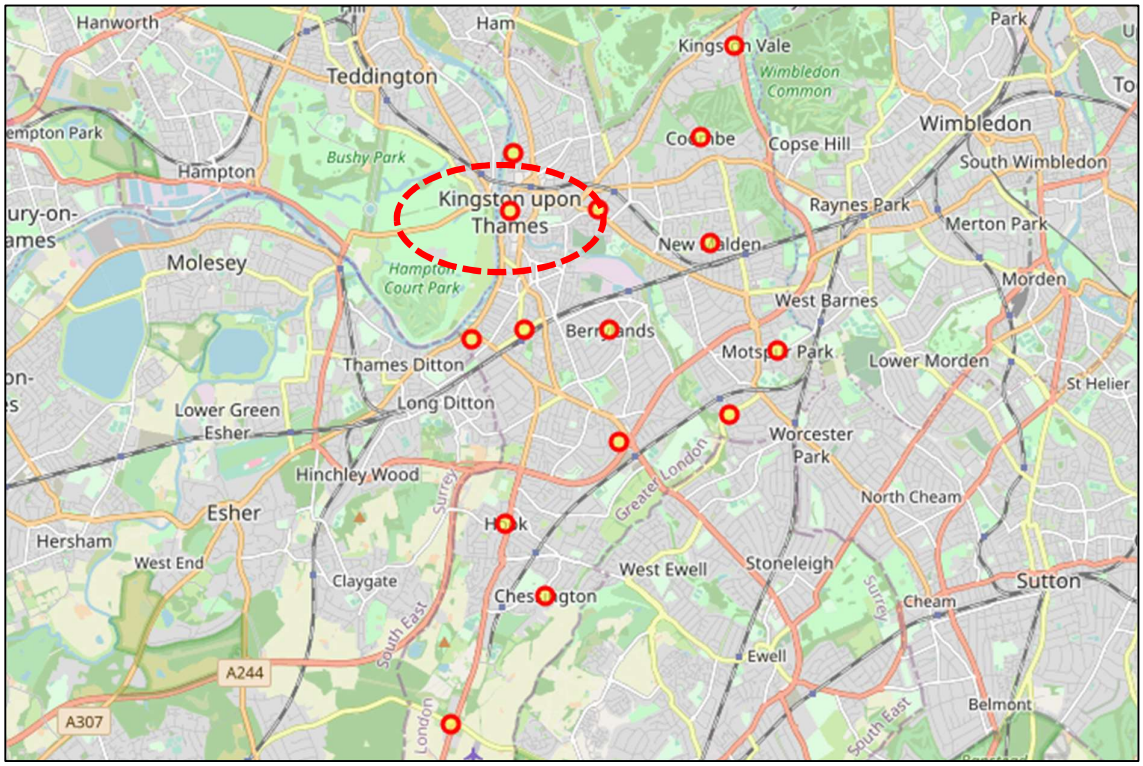
They were then converted into a DataFrame that included their coordinates' data: -

[45]:

	Neighbourhood	Borough	Latitudes	Longitudes
0	Berrylands	Kingston upon Thames	51.393781	-0.284802
1	Canbury	Kingston upon Thames	51.417499	-0.305553
2	Chessington	Kingston upon Thames	51.358336	-0.298622
3	Coombe	Kingston upon Thames	51.419450	-0.265398
4	Hook	Kingston upon Thames	51.367898	-0.307145

Snapshot of the neighbourhoods of Kingston upon Thames borough

This DataFrame was then used to make a map with the help of Folium: -



Map of Kingston upon Thames (borough is marked within a dotted oval) neighbourhood

C. KMeans clustering model

KMeans clustering was the best modelling approach to make clusters of neighbourhoods based upon the venues' information from Foursquare. The process followed is as below: -

I. Connecting to Foursquare API to generate a list of venues within 1000m of each neighbourhood

The information from Foursquare was converted to the below DataFrame: -

[51]:

	Neighbourhood	Latitudes_neighbourhood	Longitudes_neighbourhood	Venue	Latitudes_venue	Longitudes_venue	Category_venue
0	Berrylands	51.393781	-0.284802	Surbiton Racket & Fitness Club	51.392676	-0.290224	Gym / Fitness Center
1	Berrylands	51.393781	-0.284802	Jean's Cafe	51.393214	-0.296070	Coffee Shop
2	Berrylands	51.393781	-0.284802	The Berrylands Pub	51.398169	-0.280365	Pub
3	Berrylands	51.393781	-0.284802	Prince of Wales	51.389642	-0.297580	Pub
4	Berrylands	51.393781	-0.284802	K2 Bus Stop	51.392302	-0.281534	Bus Stop

Snapshot of the details generated from Foursquare

This DataFrame was then transformed into another DataFrame using one-hot encoding. The transformed DataFrame is shown below: -

[52]:

	Neighbourhood	Aquarium	Arts & Crafts Store	Asian Restaurant	Bakery	Bar	Bistro	Bookstore	Bowling Alley	Breakfast Spot	...	Tapas Restaurant	Tea Room	Tennis Court	Thai Restaurant	Theater	Theme Park
0	Berrylands	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.000000	0.0	0.0
1	Canbury	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.066667	0.0	0.0
2	Chessington	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.142857	...	0.0	0.0	0.0	0.000000	0.0	0.0
3	Coombe	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.000000	0.0	0.0
4	Hook	0.0	0.0	0.0	0.090909	0.0	0.0	0.0	0.0	0.090909	...	0.0	0.0	0.0	0.000000	0.0	0.0

Snapshot of the transformed DataFrame

Further, the transformed DataFrame was used to generate the final DataFrame for attempting a KMeans clustering model. This final DataFrame listed neighbourhoods against the 15 most respective common venues in each of them: -

[55]:

	Neighbourhood	Top venue number 1	Top venue number 2	Top venue number 3	Top venue number 4	Top venue number 5	Top venue number 6	Top venue number 7	Top venue number 8	Top venue number 9	Top venue number 10	Top venue number 11	Top venue number 12
0	Berrylands	Pub	Coffee Shop	Park	Bus Stop	Train Station	Gym / Fitness Center	Convenience Store	College Soccer Field	Platform	Exhibit	Discount Store	Donut Shop
1	Canbury	Café	Pub	Thai Restaurant	Coffee Shop	Sushi Restaurant	Italian Restaurant	Park	Market	Gastropub	Electronics Store	Record Shop	Chinese Restaurant
2	Chessington	Train Station	Supermarket	Fast Food Restaurant	Golf Course	Platform	Breakfast Spot	Department Store	Discount Store	Donut Shop	Electronics Store	English Restaurant	Exhibit
3	Coombe	Hotel	Golf Course	Bus Stop	Zoo Exhibit	Fish & Chips Shop	Discount Store	Donut Shop	Electronics Store	English Restaurant	Exhibit	Farmers Market	Fast Food Restaurant
4	Hook	Supermarket	Fish & Chips Shop	Train Station	Convenience Store	Grocery Store	Bakery	Park	Breakfast Spot	Platform	Fast Food Restaurant	Exhibit	Discount Store
5	Kingston Vale	Grocery Store	Bar	Coffee Shop	Outdoors & Recreation	Sandwich Place	Bus Stop	Soccer Field	Stables	Zoo Exhibit	Exhibit	Discount Store	Donut Shop
6	Kingston upon Thames	Coffee Shop	Pub	Café	Sushi Restaurant	Burger Joint	Furniture / Home Store	Turkish Restaurant	French Restaurant	Electronics Store	Latin American Restaurant	Market	Department Store

Snapshot of the final DataFrame prior to performing KMeans clustering

II. Performing KMeans clustering

KMeans clustering was carried out with a cluster value of 6. The resulting labels array was added to the above DataFrame as below for further analysis: -

[57]:

	Neighbourhood	Borough	Latitudes	Longitudes	Labels	Top venue number 1	Top venue number 2	Top venue number 3	Top venue number 4	Top venue number 5	Top venue number 6	Top venue number 7	Top venue number 8
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	4	Pub	Coffee Shop	Park	Bus Stop	Train Station	Gym / Fitness Center	Convenience Store	College Soccer Field
1	Canbury	Kingston upon Thames	51.417499	-0.305553	0	Café	Pub	Thai Restaurant	Coffee Shop	Sushi Restaurant	Italian Restaurant	Park	Market
2	Chessington	Kingston upon Thames	51.358336	-0.298622	2	Train Station	Supermarket	Fast Food Restaurant	Golf Course	Platform	Breakfast Spot	Department Store	Discount Store
3	Coombe	Kingston upon Thames	51.419450	-0.265398	5	Hotel	Golf Course	Bus Stop	Zoo Exhibit	Fish & Chips Shop	Discount Store	Donut Shop	Electronics Store
4	Hook	Kingston upon Thames	51.367898	-0.307145	2	Supermarket	Fish & Chips Shop	Train Station	Convenience Store	Grocery Store	Bakery	Park	Breakfast Spot

Snapshot of the DataFrame with labels from KMeans clustering added

4. Results

A. Least crime ridden borough

Kingston Upon Thames borough was identified as the London borough with the minimum number of crimes in 2016. Therefore, the target audience mentioned in the introduction section of this report (page 2) must give preference to this borough for selecting their home after arriving for the first time in London.

B. Kingston upon Thames neighbourhoods' clusters

Within Kingston upon Thames, the safest borough in London, KMeans clustering was carried out on the DataFrame created earlier to cluster neighbourhoods based upon the most common venues. This exercise would help new people coming to London in selecting the most suitable neighbourhood based upon his / her individual tastes and preferences of venues. The following is the selection guide for the 6 clusters created: -

I. Cluster 1 - Most common venues: Pubs and coffee joints.

[60]:

	Neighbourhood	Borough	Latitudes	Longitudes	Labels	Top venue number 1	Top venue number 2	Top venue number 3	Top venue number 4	Top venue number 5	Top venue number 6
1	Canbury	Kingston upon Thames	51.417499	-0.305553	0	Café	Pub	Thai Restaurant	Sushi Restaurant	Coffee Shop	Hotel
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	0	Café	Pub	Coffee Shop	Sushi Restaurant	Burger Joint	Record Shop
9	New Malden	Kingston upon Thames	51.405335	-0.263407	0	Korean Restaurant	Supermarket	Grocery Store	Coffee Shop	Gym Pool	Café
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	0	Pub	Thai Restaurant	Italian Restaurant	Indian Restaurant	Gastropub	Bar
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	0	Coffee Shop	Pub	Gastropub	Indian Restaurant	French Restaurant	Park
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	0	Coffee Shop	Pub	Park	Chinese Restaurant	Grocery Store	Gym

II. Cluster 2 - Most common venues: Theme parks / attractions and pubs.

[61]:

	Neighbourhood	Borough	Latitudes	Longitudes	Labels	Top venue number 1	Top venue number 2	Top venue number 3	Top venue number 4	Top venue number 5	Top venue number 6
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	1	Theme Park Ride / Attraction	Pub	Zoo Exhibit	Restaurant	Fast Food Restaurant	Fried Chicken Joint

III. Cluster 3 - Most common venues: Supermarkets, stores, and train station.

[62]:

	Neighbourhood	Borough	Latitudes	Longitudes	Labels	Top venue number 1	Top venue number 2	Top venue number 3	Top venue number 4	Top venue number 5	Top venue number 6
2	Chessington	Kingston upon Thames	51.358336	-0.298622	2	Train Station	Supermarket	Breakfast Spot	Platform	Fast Food Restaurant	Building
4	Hook	Kingston upon Thames	51.367898	-0.307145	2	Supermarket	Convenience Store	Breakfast Spot	Park	Chinese Restaurant	Fast Food Restaurant

IV. Cluster 4 - Most common venues: Stores, coffee shop, bus stop, and soccer field.

[63]:

	Neighbourhood	Borough	Latitudes	Longitudes	Labels	Top venue number 1	Top venue number 2	Top venue number 3	Top venue number 4	Top venue number 5	Top venue number 6
6	Kingston Vale	Kingston upon Thames	51.431850	-0.258138	3	Coffee Shop	Bus Stop	Outdoors & Recreation	Sandwich Place	Soccer Field	Stables
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	3	Grocery Store	Soccer Field	Pharmacy	Garden Center	Bowling Alley	Climbing Gym

V. Cluster 5 - Most common venues: Park, pub, bus stop, and platform.

[64]:

	Neighbourhood	Borough	Latitudes	Longitudes	Labels	Top venue number 1	Top venue number 2	Top venue number 3	Top venue number 4	Top venue number 5	Top venue number 6
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	4	Coffee Shop	Park	Pub	Gym / Fitness Center	Train Station	Platform
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	4	Park	Pub	Bus Stop	Steakhouse	Mediterranean Restaurant	Korean Restaurant
11	Old Malden	Kingston upon Thames	51.382484	-0.259090	4	Train Station	Park	Platform	Gym / Fitness Center	Grocery Store	Bakery

VI. Cluster 6 - Most common venues: Hotel and golf course.

[65]:

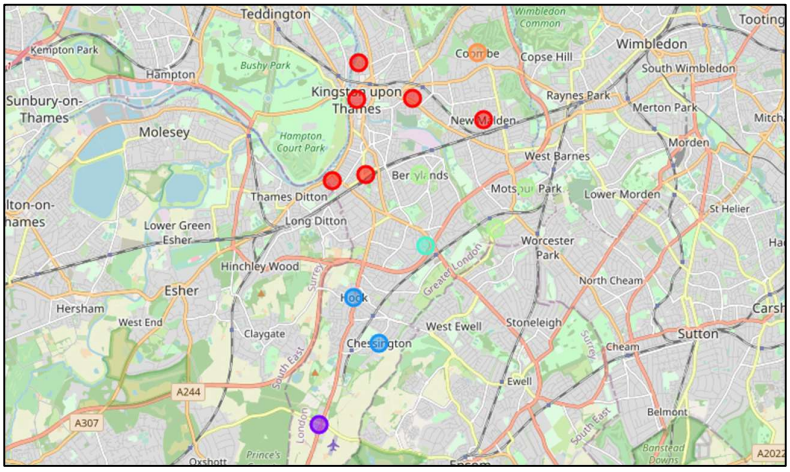
	Neighbourhood	Borough	Latitudes	Longitudes	Labels	Top venue number 1	Top venue number 2	Top venue number 3	Top venue number 4	Top venue number 5	Top venue number 6
3	Coombe	Kingston upon Thames	51.41945	-0.265398	5	Hotel	Golf Course	Stables	Zoo Exhibit	Fish & Chips Shop	Deli / Bodega

5. Discussion

The first part of the project identified the safest borough in London i.e. Kingston upon Thames. Thereafter, the project clustered the neighbourhoods based upon the most common venues in each of the neighbourhoods. This analysis can be used as a guideline for the below type of individuals / families to select a home to either rent or buy in Kingston upon Thames: -

- For families with kids
  - Cluster 2, cluster 3, or cluster 5 which have theme parks, supermarkets, and parks as the most common venues.
- For singles and couples
  - Cluster 1 which has pubs and coffee joints as the most common venues.
- For sports freaks
  - Cluster 4 and 6 which have soccer field and golf course as common venues.
- For the daily commuters
  - Cluster 3, cluster 4, and cluster 5 which have train station, bus stop, and a platform as common venues.

There is also a scope for people who like to eat out, or need a gym, or like horse-riding if the 3rd and the 4th most common venues are also considered before selecting a neighbourhood. In summary, there are many possibilities depending upon the priorities of the people choosing to reside in Kingston upon Thames borough of London. A map of the aforementioned clusters is provided below: -



Map depicting a few of the neighbourhood clusters in Kingston upon Thames

6. Conclusion

This project has provided a well-researched, scientific solution for the business problem that was defined in the introduction section of this report. It has used publicly available data for crimes in London as well as Foursquare API to provide the best possible borough to reside in London, considering safety to the priority for any new person or family moving to London. Furthermore, to aid the selection of the most suitable neighbourhood within Kingston upon Thames, this project provides an eclectic mix of the most common venues within a neighbourhood to make a data backed choice in selecting a house to rent or buy in the borough of Kingston upon Thames in London.