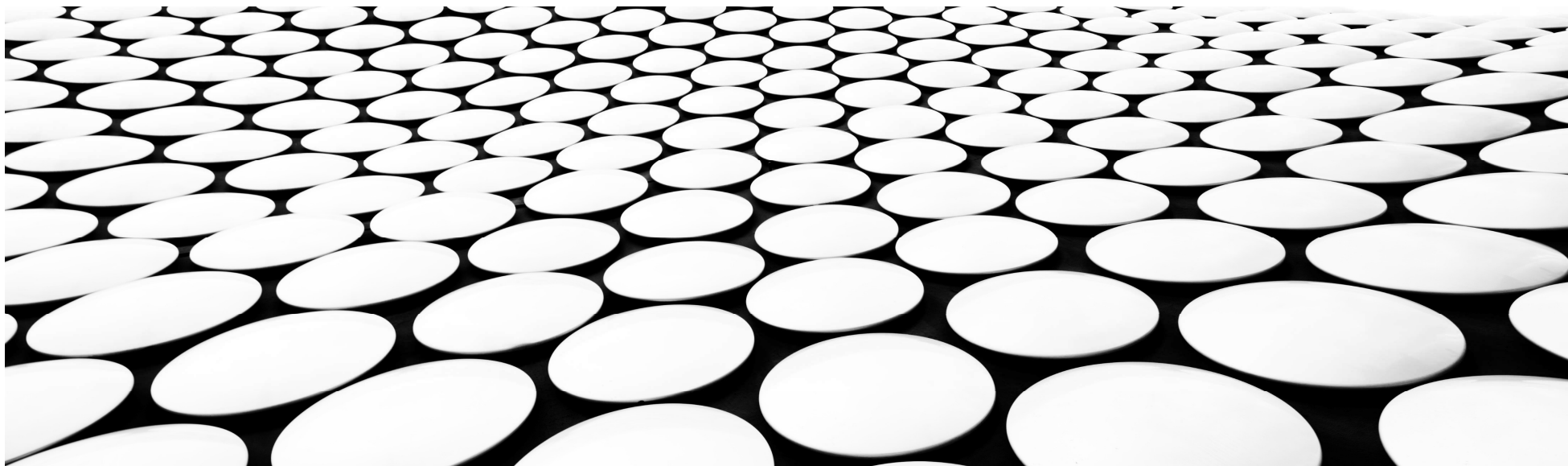# IBM DATA SCIENCE PROFESSIONAL CERTIFICATE

PRESENTATION - THE BATTLE OF NEIGHBOURHOODS (LONDON) BY UTKARSH SHARMA

INTRODUCTION

Context
- In the year ending December 2019, 677,000 people migrated into the UK. The concentration of migrants was maximum in London.
- A proper research can help in reducing the problems associated with moving to a new place.

Business Problem
- Analyse the boroughs of London (UK) for safety based upon the criminal records available in public domain. Once the safest borough is identified, explore it's neighbourhoods.

Target Audience
- Professionals shifting to London as part of their job assignments or for beginning a new job.
- Overseas students shifting to London for pursuing higher education.
- Entrepreneurs shifting to London to try their luck in the UK business environment.

DATA ACQUISITION

The following data was collected for undertaking the analysis: -

1. 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.

2. 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.

3. 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.

METHODOLOGY

Data pre-processing

- The London Crime data was processed only for year 2016 (latest available data) and converted to a pivoted DataFrame using major crime category as columns and rows a borough names.

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

- This above two DataFrames were then merged using the common column 'Borough' for further data analysis.

- After identifying the safest borough in London, the Wikipedia link for neighbourhoods for London boroughs was used to create a DataFrame focussed on neighbourhoods in that borough.

METHODOLOGY - CONTINUED

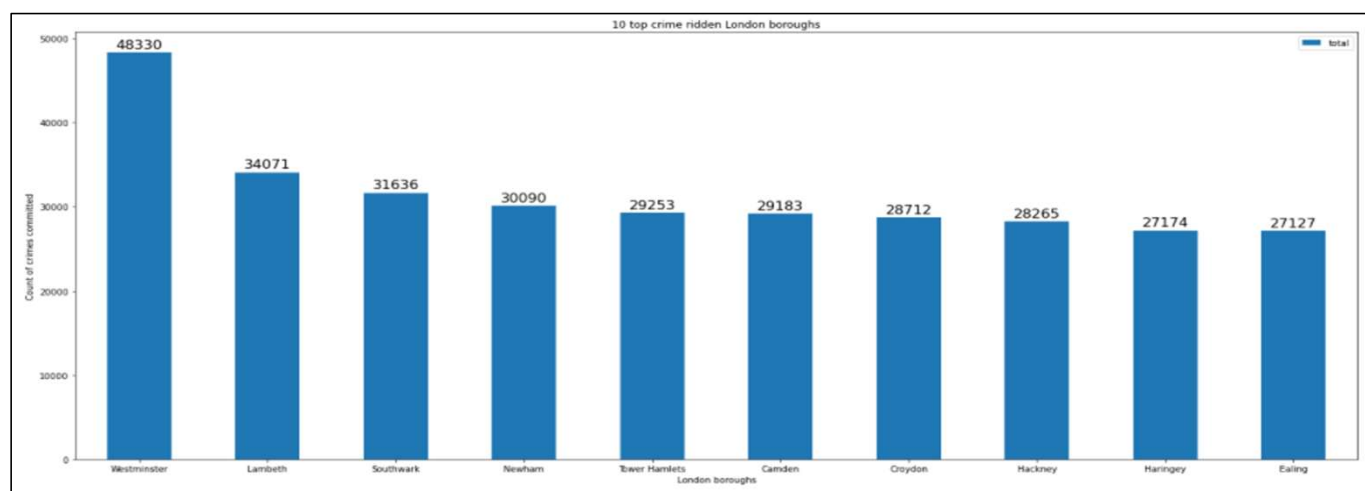Exploratory data analysis
- Statistical summary of London crime data: -

[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	

- The count of 33 represents the total number of boroughs in London.
- The major crime 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).

METHODOLOGY - CONTINUED

Exploratory data analysis

- Boroughs with the highest crime rates: -
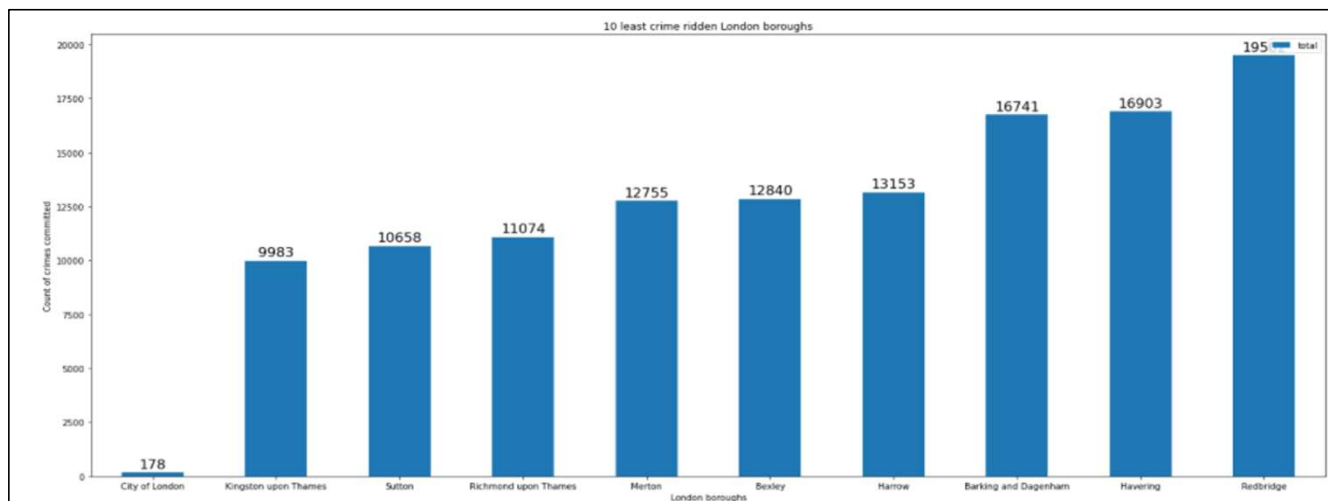

- The 10 most crime ridden boroughs based upon the analysis are listed below:-
  - ➢ Westminster (maximum 48330 crimes in 2016), Lambeth, Southwark, Newham, Tower Hamlets, Camden, Croydon, Hackney, Haringey, Ealing.

METHODOLOGY - CONTINUED

Exploratory data analysis
- Boroughs with the lowest crime rates: -

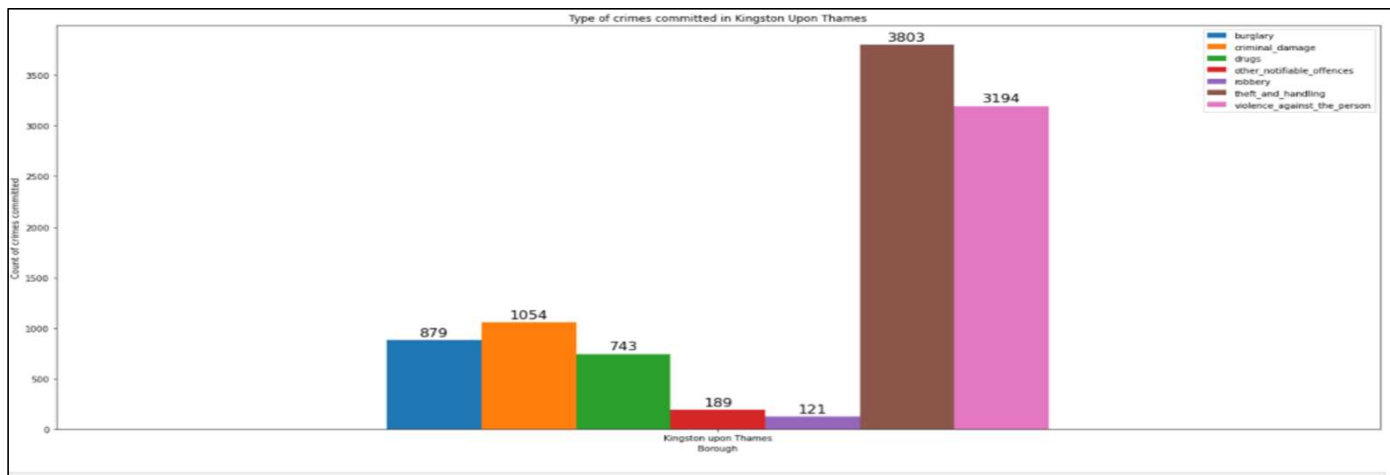- The 10 least crime ridden boroughs based upon the analysis are listed below:-
  - ➢ City of London (minimum 178 crimes in 2016) – However, it is not a borough so not analysed further.
  - ➢ Kingston upon Thames (next least 9983 crimes in 2016) - considered for further analysis.
  - ➢ Rest of the 8 boroughs - Sutton, Richmond upon Thames, Merton, Bexley, Harrow, Barking & Dagenham, Havering, Redbridge.

METHODOLOGY - CONTINUED

Exploratory data analysis

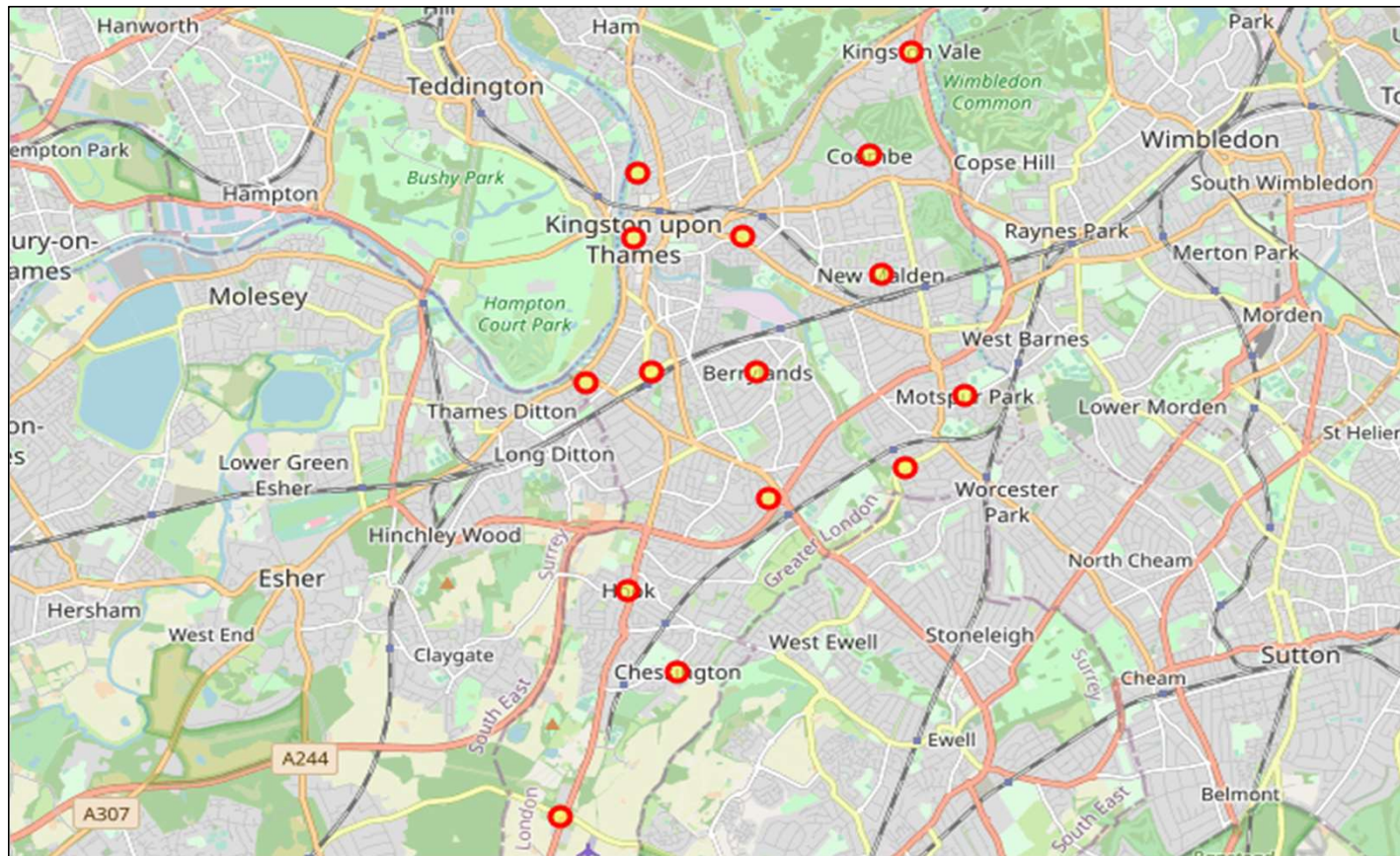- Categorical distribution of crimes in Kingston upon Thames (borough with lowest crimes): -

- The top 3 crime categories in Kingston upon Thames with decreasing number of occurrences are listed below: -
  - ➤ Theft and handling (highest - 3803 crimes)
  - ➤ Violence against the person
  - ➤ Criminal damage

METHODOLOGY - CONTINUED

Distribution map of the 15 neighbourhoods of Kingston upon Thames: -

METHODOLOGY - CONTINUED

KMeans Cluster Modelling: -

- Used Foursquare API to generate DataFrame of venues within 1000m of each neighbourhood.
- This DataFrame was then transformed into another DataFrame, suited to modelling, using one-hot encoding.
- The transformed DataFrame was finally used to generate the final DataFrame for attempting a KMeans clustering model. This final DataFrame listing neighbourhoods against the 15 most respective common venues in each of them is provided below (showing top 12 venues only): -

	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

METHODOLOGY - CONTINUED

KMeans Cluster Modelling: -

- 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 and generation of results: -

	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

RESULTS

- Kingston Upon Thames borough was identified as the London borough with the minimum number of crimes in 2016.
- Within Kingston upon Thames, KMeans clustering was carried out on the DataFrame created earlier to cluster neighbourhoods based upon the most common venues. The clusters are as below (showing top 5 venues only): -
  - ➢ 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

RESULTS - CONTINUED

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

	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

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

	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

RESULTS - CONTINUED

➤ 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

➤ 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

RESULTS - CONTINUED

➢ Cluster 6 - Most common venues: Hotel and golf course: -

	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

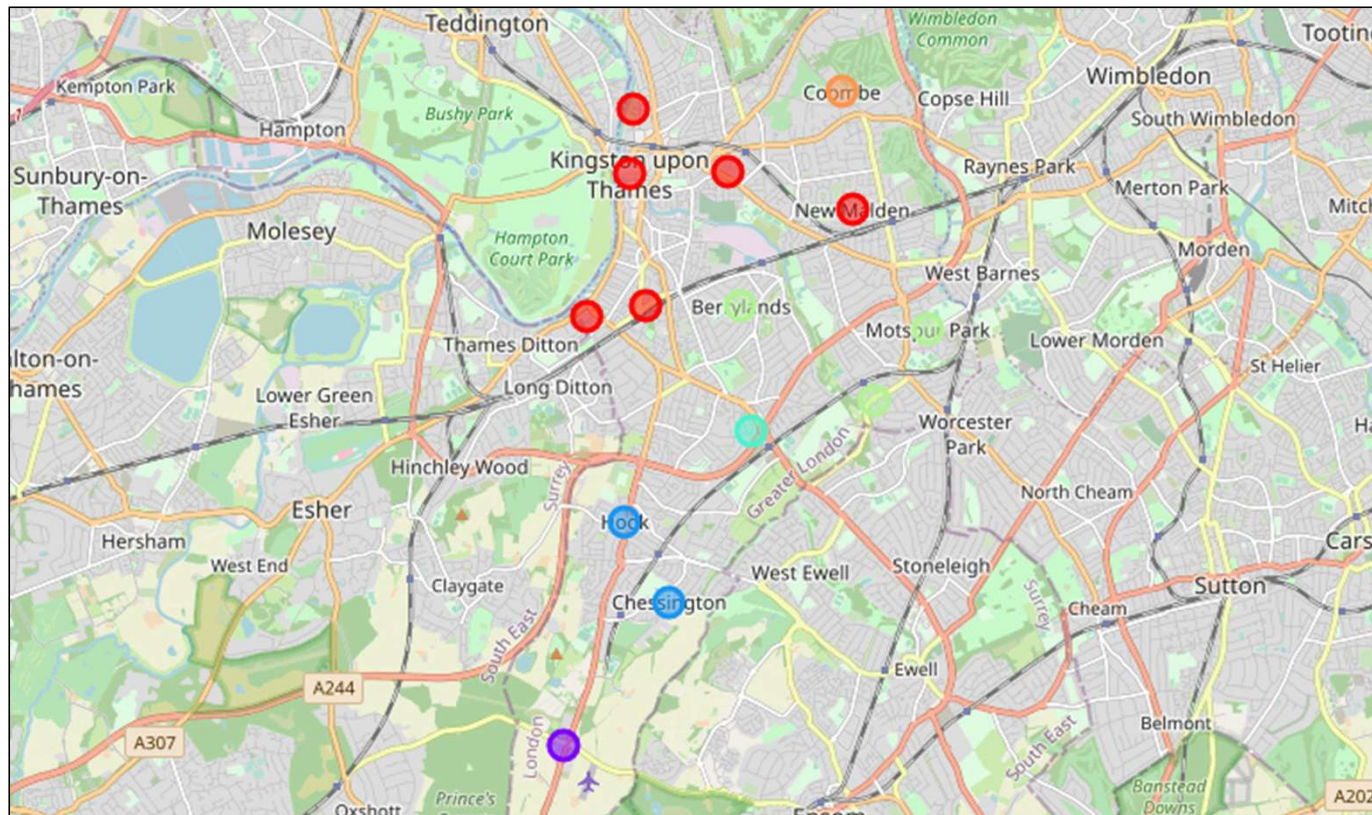# DISCUSSION

- 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.

DISCUSSION - CONTINUED

- A map of the aforementioned clusters is provided below: -

CONCLUSION

- This project has provided a well-researched and 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.

- Individuals and families can use this analysis to make a data backed choice in selecting a house to rent or buy in the borough of Kingston upon Thames in London.