

THE BATTLE OF NEIGHBORHOODS

Capstone Project provide by IBM on Coursera

By:

Sreerag Rajeevan



INTRODUCTION

- Migration has become a very common phenomenon. This could be due to various reasons.
- The aim of this project is to help people shifting to London, find the optimal location. Keeping Safety of localities the main feature, we will be segregating boroughs within London.
- We will explore the safest neighborhood in detail using K-Means clustering algorithm.
- People moving into the London from various places will a clear picture about the safety of various boroughs in London.

DATA EXTRACTION

- We will be extract data from various sources.
- The main dataset, explaining the various crimes committed in London is extracted from Kaggle.
- The next dataset contains the coordinates, area, population, etc. of various boroughs of London. The data set is extracted from a Wikipedia page.
- Finally, in the later stage of the project, once we identify the safest borough in London, coordinates of the neighborhoods within that borough is extracted from a Wikipedia page.

CLEANING DATA

- Data extracted from the internet is bound to have lots of impurities.
- Cleaning data is a very important step to obtain perfect results.
- Impure data could mingle with our final results and going through the entire process is time consuming.
- After basic cleaning, we will merge the Kaggle data and the Wikipedia data, so that, we can go ahead with visualization.
- We will be using this table to find the top venues around a particular location using Foursquare API.

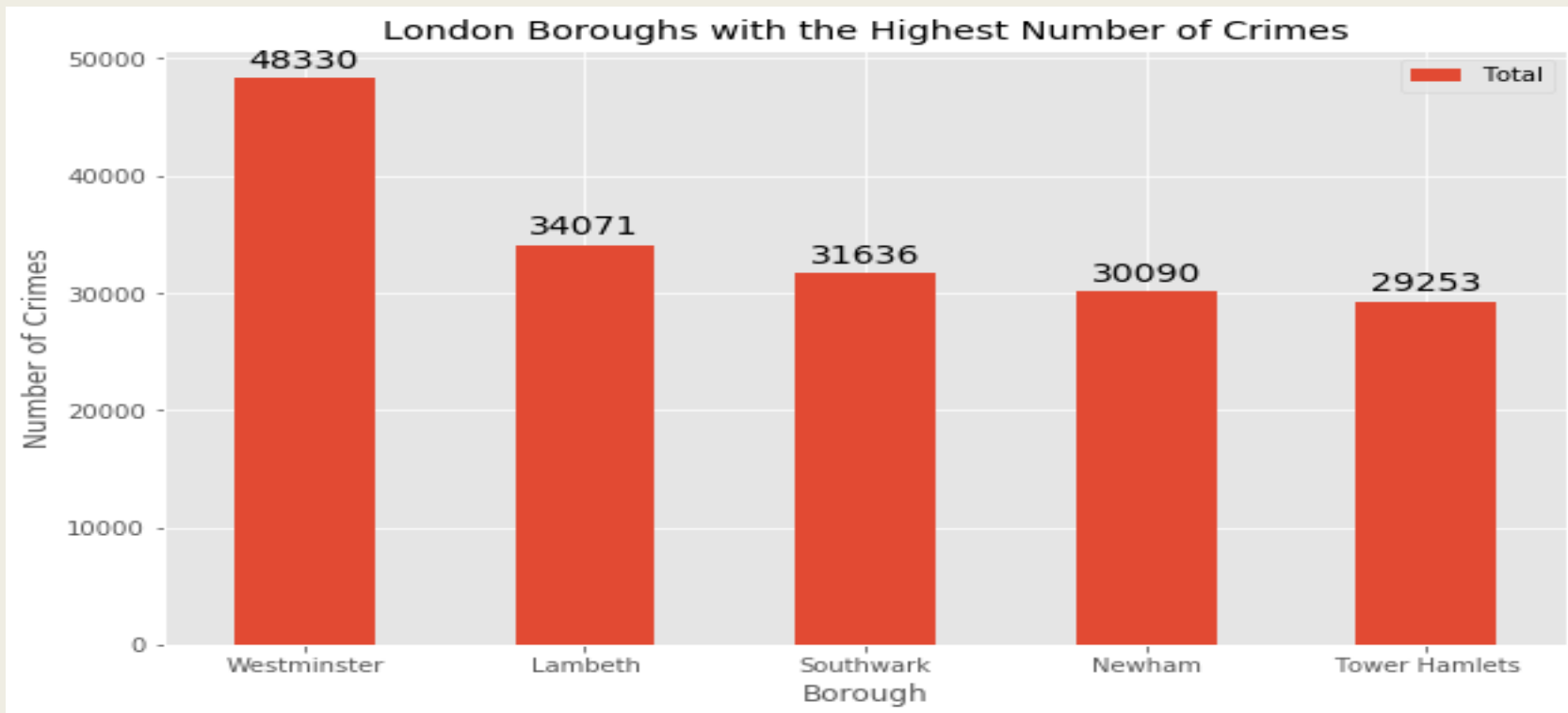
METHODOLOGY

We will be using various methods to obtain at our final result. First, we have to know what type of crime is being committed in the boroughs of London. This is obtained using the describe function in Python.

	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

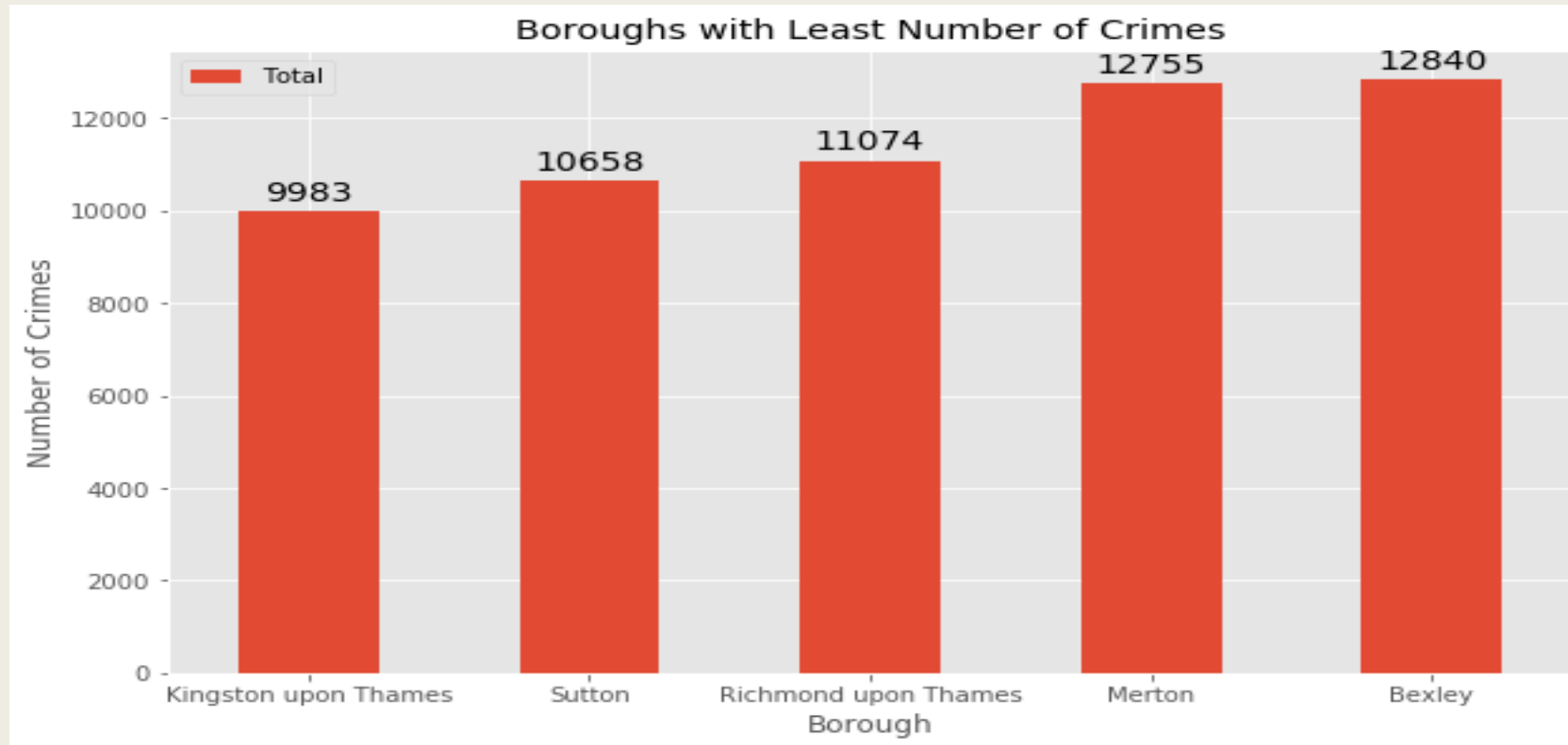
METHODOLOGY(Contd..)

- Our next step is analyzing the boroughs with highest number of crimes committed. This will help us in knowing the boroughs that can be avoided by people moving in from other places. From the graph it is clear that Westminster, Lambeth, Southwark, Newham, and Tower Hamlets are the boroughs to be avoided.



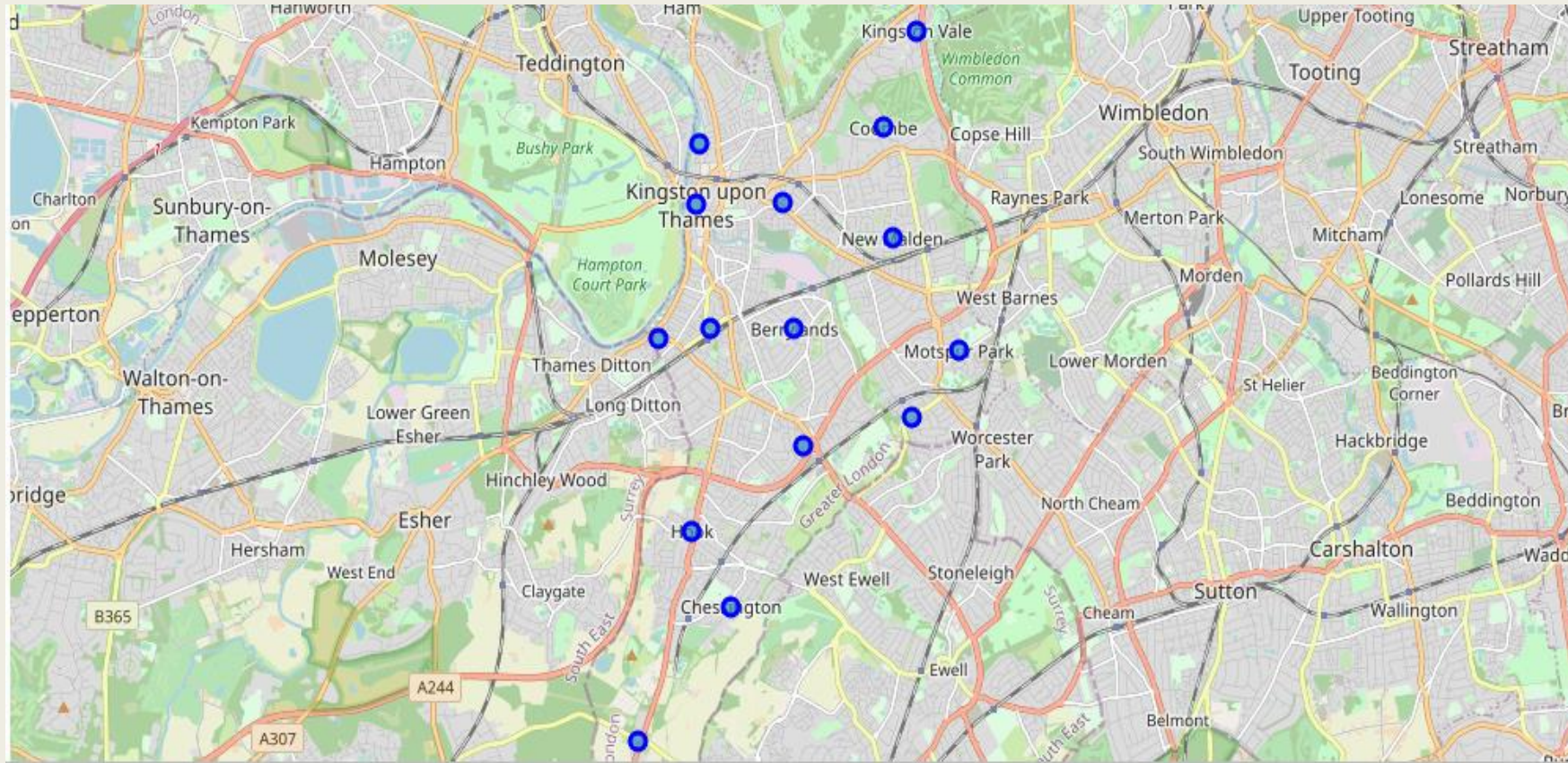
METHODOLOGY(Contd..)

- Our aim is to obtain the borough with lowest crimes committed. We will be plotting the graph of top five boroughs with lowest crime rates. We see that Kingston upon Thames has the lowest crime rate. Our next step is to analyze this borough in detail.



METHODOLOGY(Contd...)

There are a total of fifteen neighborhoods in Kingston upon Thames. This is shown below.

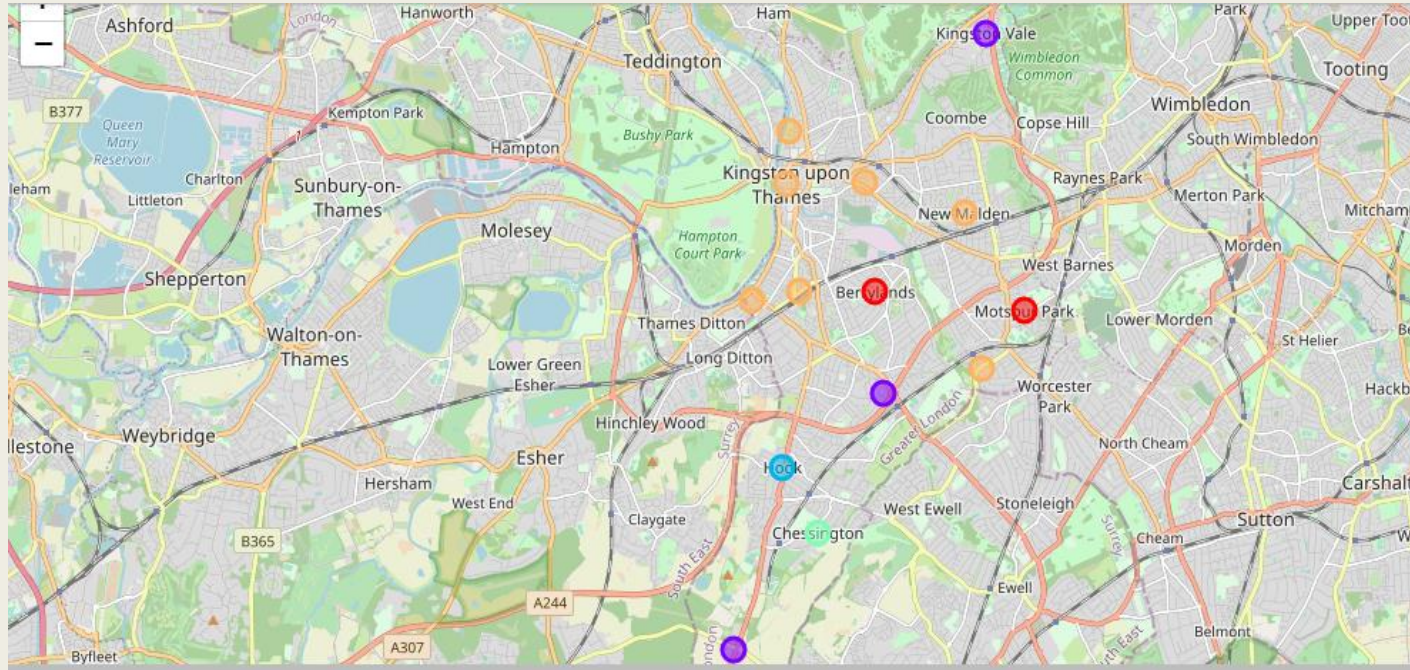


METHODOLOGY(Contd...)

- Now that we have mapped the neighborhoods of Kingston upon Thames, we will explore the neighborhoods using Foursquare API.
- Before applying Foursquare, we will convert dataset into a form that would be easy for Machine Learning algorithm to process. This is done by converting the data using Onehot encoding process.
- Using Foursquare, we will identify the ten most common venues around a neighborhood.
- Then using this data, we will cluster neighborhoods using K-Means clustering algorithm.
- Our next step is to explore the clustered neighborhoods and find the best neighborhood.

RESULTS

- Once we are done with K-means clustering, we can use the folium library available on Python to visualize the clusters.
- On the map we can see color coded representation of the clusters.
- Same color indicates, similar clusters.



RESULTS(Contd...)

- Cluster ONE: We see that this cluster contains all the ideal places that a family would be looking for like, supermarkets, Train Station, Pharmacies, Restaurants, etc.

	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
1	Canbury	Kingston upon Thames	51.417499	-0.305553	0	Pub	Café	Plaza	Fish & Chips Shop	Supermarket	Spa	Shop & Service	Park
4	Hook	Kingston upon Thames	51.367898	-0.307145	0	Bakery	Convenience Store	Indian Restaurant	Fish & Chips Shop	Wine Shop	Food	Electronics Store	Farmers Market
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	0	Coffee Shop	Café	Burger Joint	Sushi Restaurant	Pub	Record Shop	Cosmetics Shop	Market
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	0	Convenience Store	Pub	Garden Center	Restaurant	Fast Food Restaurant	Discount Store	Dry Cleaner	Electronics Store
9	New Malden	Kingston upon Thames	51.405335	-0.263407	0	Gastropub	Gym	Sushi Restaurant	Supermarket	Korean Restaurant	Indian Restaurant	Fish & Chips Shop	Dry Cleaner
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	0	Indian Restaurant	Pub	Food	Italian Restaurant	Platform	Grocery Store	Farmers Market	Dry Cleaner
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	0	Indian Restaurant	Coffee Shop	Italian Restaurant	Pub	Café	Wine Shop	Fast Food Restaurant	Chinese Restaurant
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	0	Coffee Shop	Pub	Supermarket	Breakfast Spot	Grocery Store	Gastropub	French Restaurant	Train Station
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	0	Grocery Store	Pharmacy	Furniture / Home Store	Train Station	Pizza Place	Discount Store	Coffee Shop	Bus Stop

RESULTS(Contd...)

- Clusters TWO, THREE & FIVE are similar with only one neighborhood. The fact that these neighborhoods are in Kingston upon Thames is an ideal thing, but, cluster ONE is better as a whole.

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1 Most Common Venue	2 Most Common Venue	3 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
3	Coombe	Kingston upon Thames	51.41945	-0.265398	1	Health & Beauty Service	Wine Shop	Food	Department Store	Dry Cleaner	Electronics Store	Farmers Market	Fast Food Restaurant	Fish & Chips Shop
	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1 Most Common Venue	2 Most Common Venue	3 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
11	Old Malden	Kingston upon Thames	51.382484	-0.25909	2	Grocery Store	Food	Construction & Landscaping	Train Station	German Restaurant	Gastropub	Garden Center	Furniture / Home Store	Fried Chicken Joint
	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1 Most Common Venue	2 Most Common Venue	3 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
6	Kingston Vale	Kingston upon Thames	51.43185	-0.258138	4	Sandwich Place	Grocery Store	Bar	Soccer Field	Wine Shop	Dry Cleaner	Electronics Store	Farmers Market	Fast Food Restaurant

RESULTS(Contd...)

- Cluster FOUR: This cluster would be an ideal second choice in case cluster ONE does not work out.

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1 Most Common Venue	2 Most Common Venue	3 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
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	3	Gym / Fitness Center	Park	Bus Stop	Wine Shop	Food	Dry Cleaner	Electronics Store	Farmers Market	Fast Food Restaurant
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	3	Gym	Park	Bus Stop	Soccer Field	Restaurant	Fast Food Restaurant	Department Store	Dry Cleaner	Electronics Store

DISCUSSIONS

- This project provides an in-depth understanding about the boroughs in London.
- Safety is the main feature that we had taken into consideration.
- Anyone relocating to London would get a clear picture regarding the safe borough in the beautiful city.
- According to the analysis that were conducted, I would like to suggest neighborhoods falling in cluster one to be the ideal choice.

CONCLUSION

- I would like to conclude by saying that even though this project gives an understanding regarding the safety of the boroughs in London, safety alone is not what people look before relocating to any place.
- This project should be explore on another level, taking the expense of each locality, i.e., the living expense, rent to be paid, etc. being taken into consideration.