Capstone Project: The Battle of Neighbourhoods

By
Atharva Kulkarni

Introduction

- Background:
  - Before moving to new place, safety and access to essential places is important factor to be considered.
  - It is important that the chosen place should be crime-free and nearer to different places
- Problem:
  - The project aims to solve this scenario by identifying the safest place in London based on crime data of London by exploring neighbourhoods and identifying 10 famous venues.
- Solution:
  - Applying k-means clustering algorithm for identifying venue
  - using Foursquare API for location data.
  - Visualized clusters using Folium

Data Acquisition

- There are 3 data sources

1. London crime data from Kaggle Dataset
2. Web scrapping of Wikipedia page having list of boroughs in London
3. Web scrapping Wikipedia page having list of neighbours of Royal borough

Data Cleaning

- Crime data consists of each crime occurred in London, so data is grouped based on the type of crime and locality
- For web scrapping “beautifulsoup” library used to scape html table element, which is then converted into dataframe

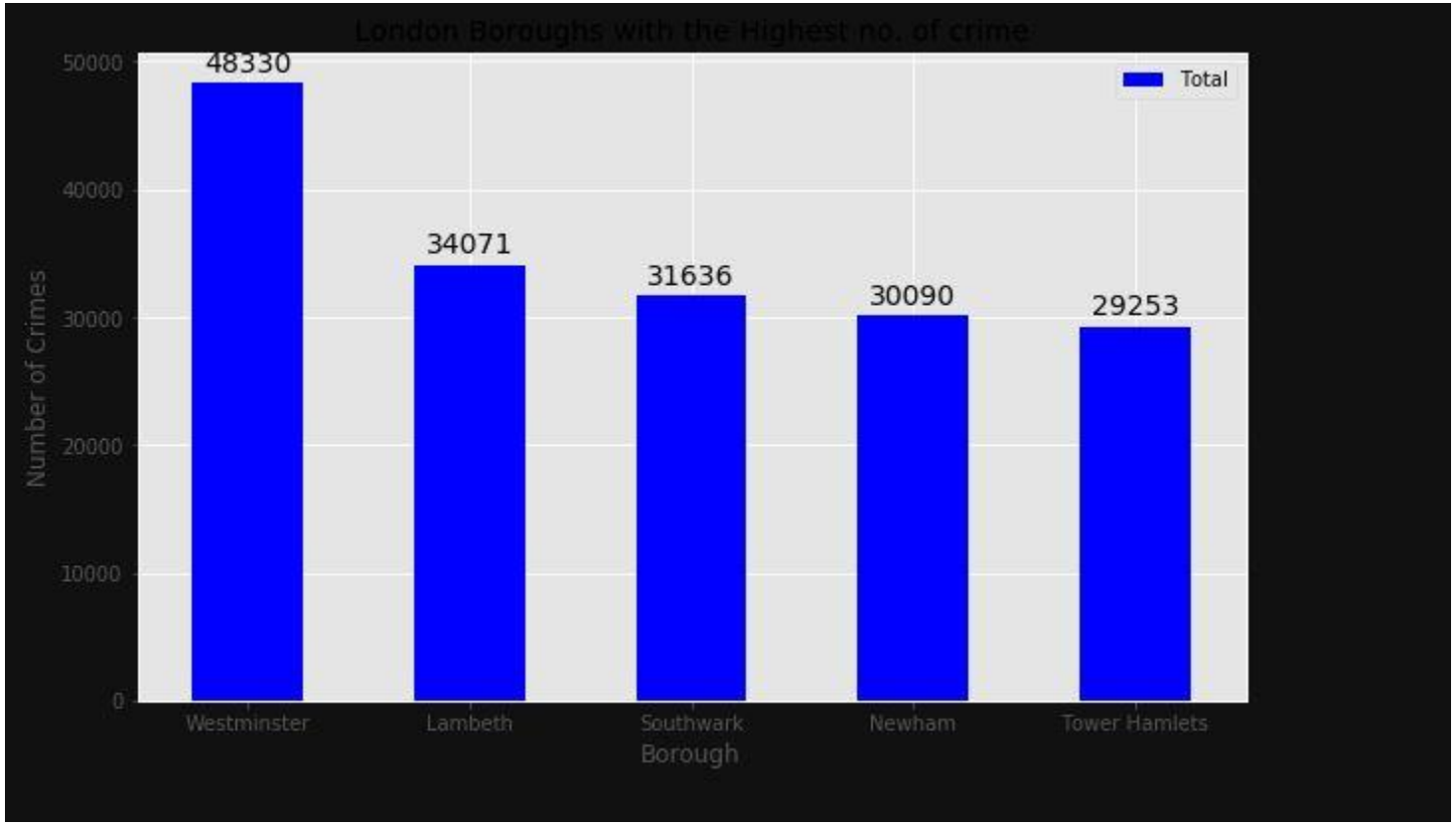
Insights of Data

- Exploratory Data Analysis
  - Statistical Summery of crime data in London

	No_of_CrimesBurglary	No_of_CrimesCriminal Damage	No_of_CrimesDrugs	No_of_CrimesOther Notifiable Offences	No_of_CrimesRobbery	No_of_CrimesTheft and Handling	No_of_CrimesViolence 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

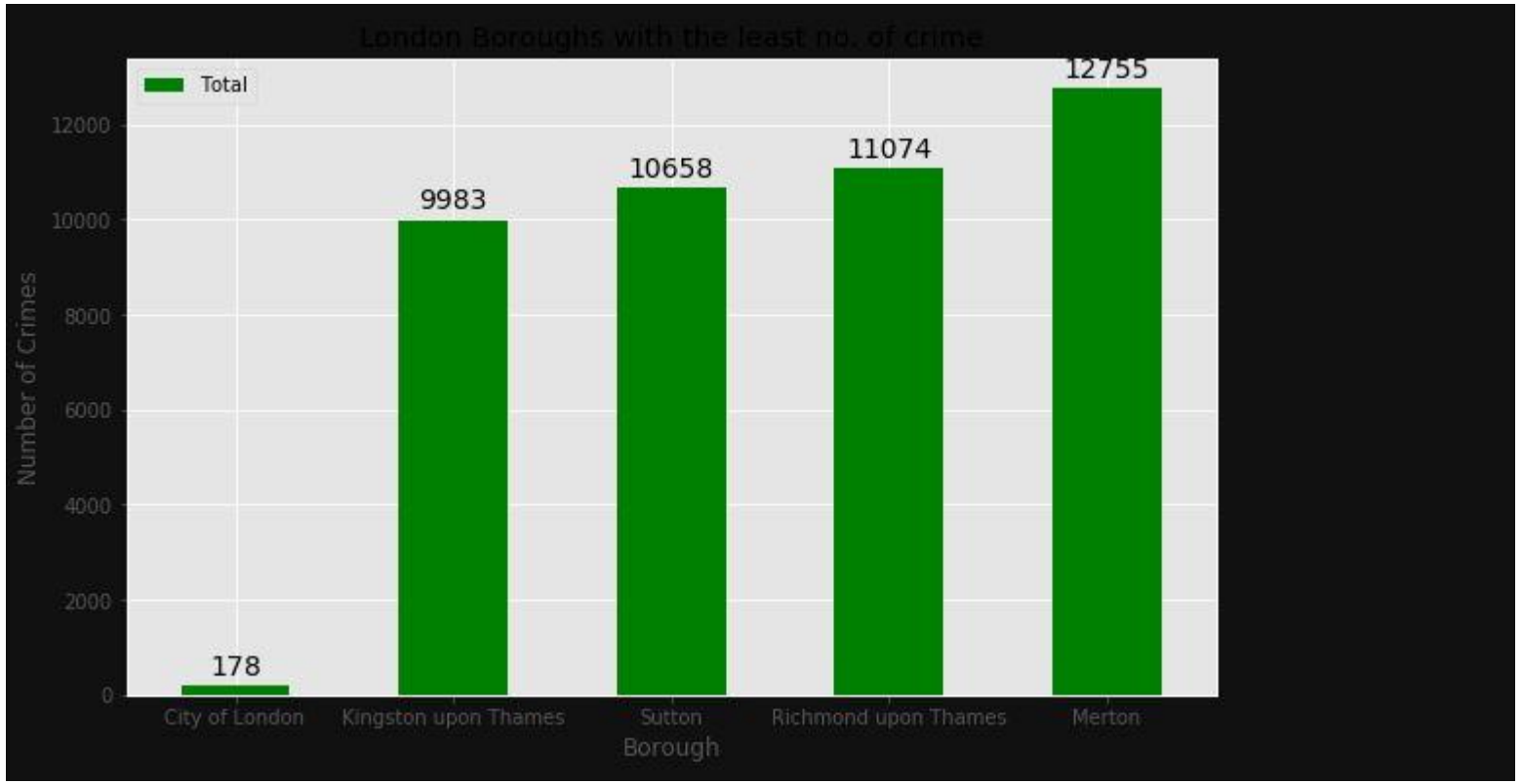
Insights of Data

- Observation of crime rate against boroughs in London(top 5)



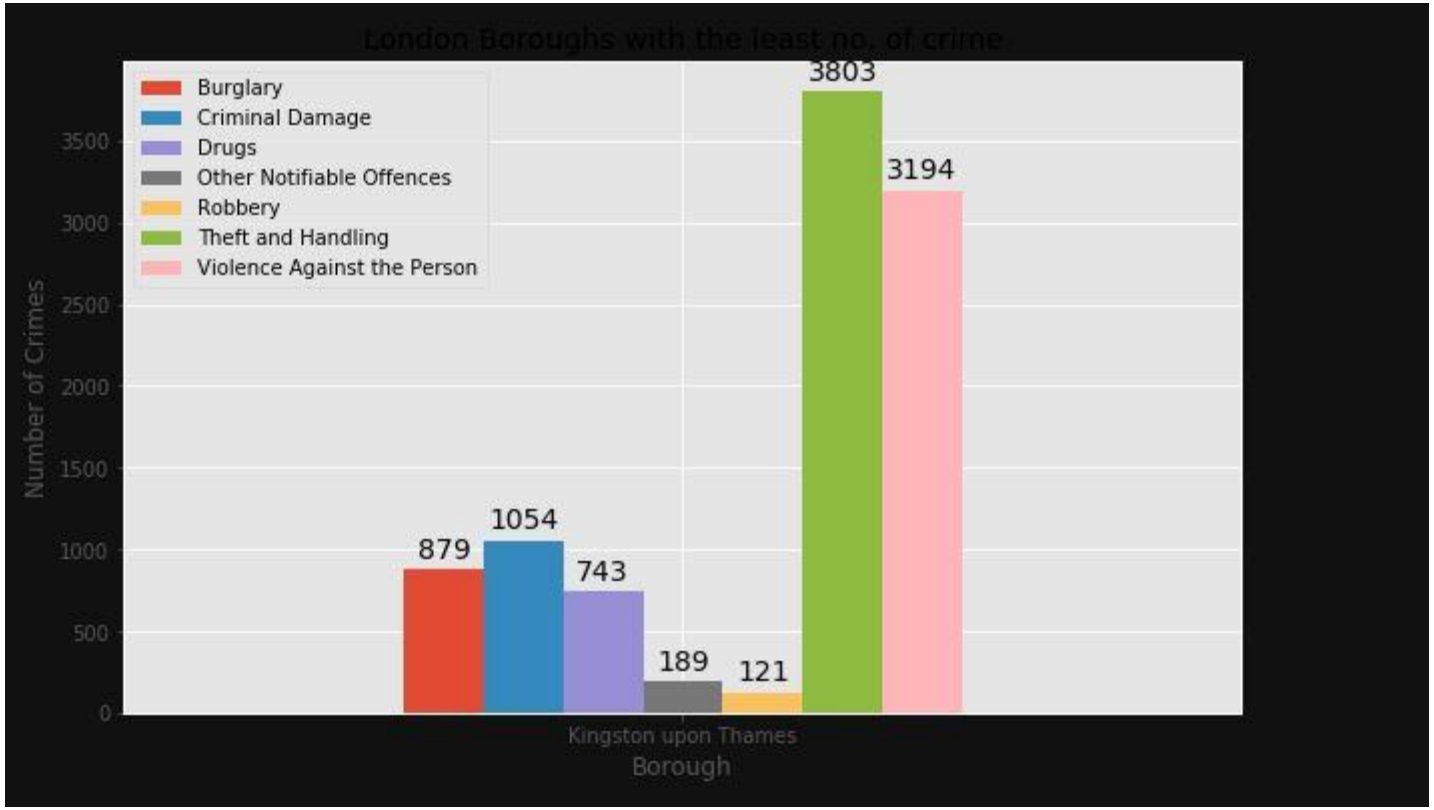
Insights of Data

- Observation of crime rate against boroughs in London(last 5)



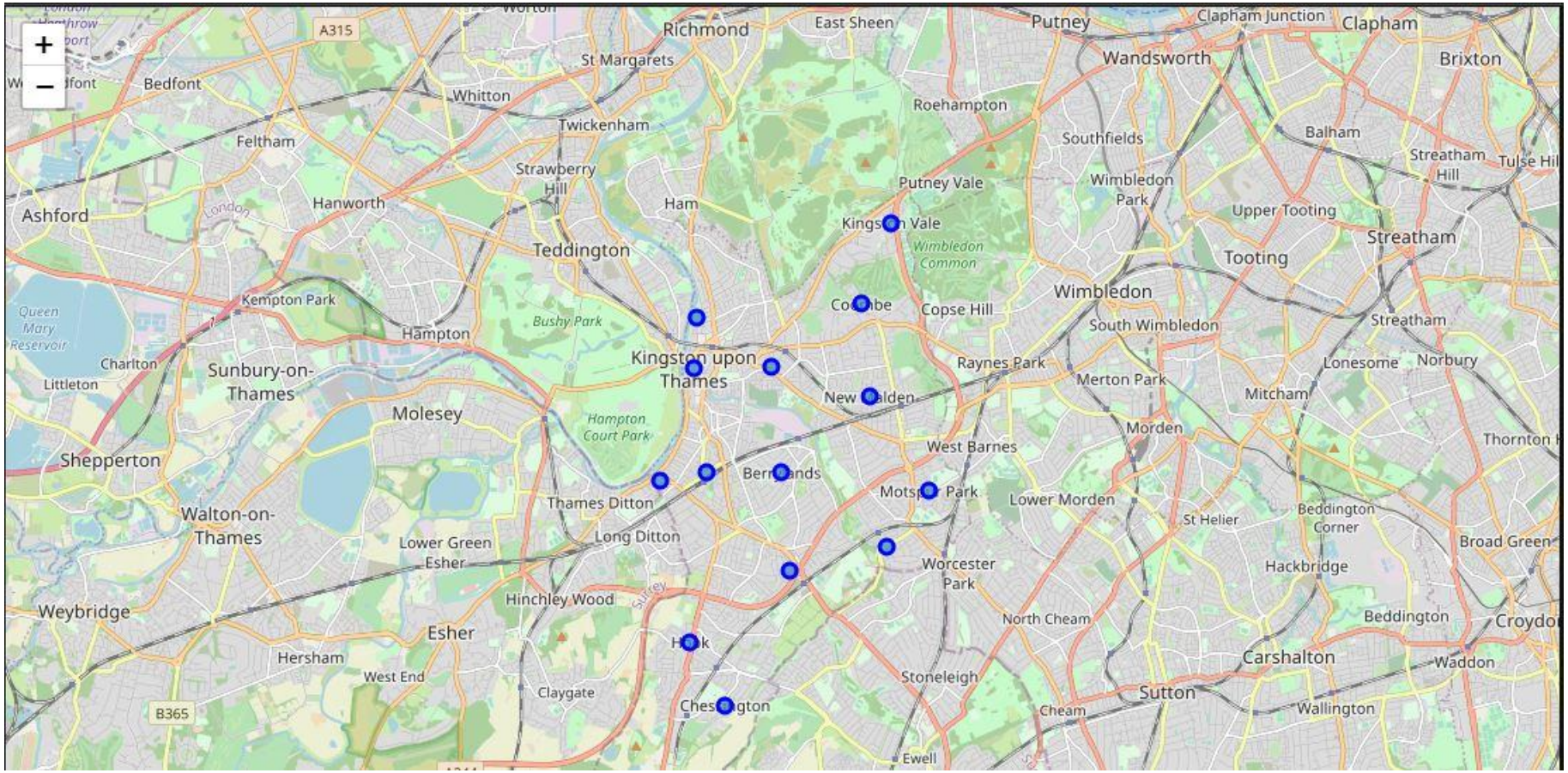
Insights of Data

- Different types of crime occurred in London



Insights of Data

- Neighbourhoods in Kingston upon Thames

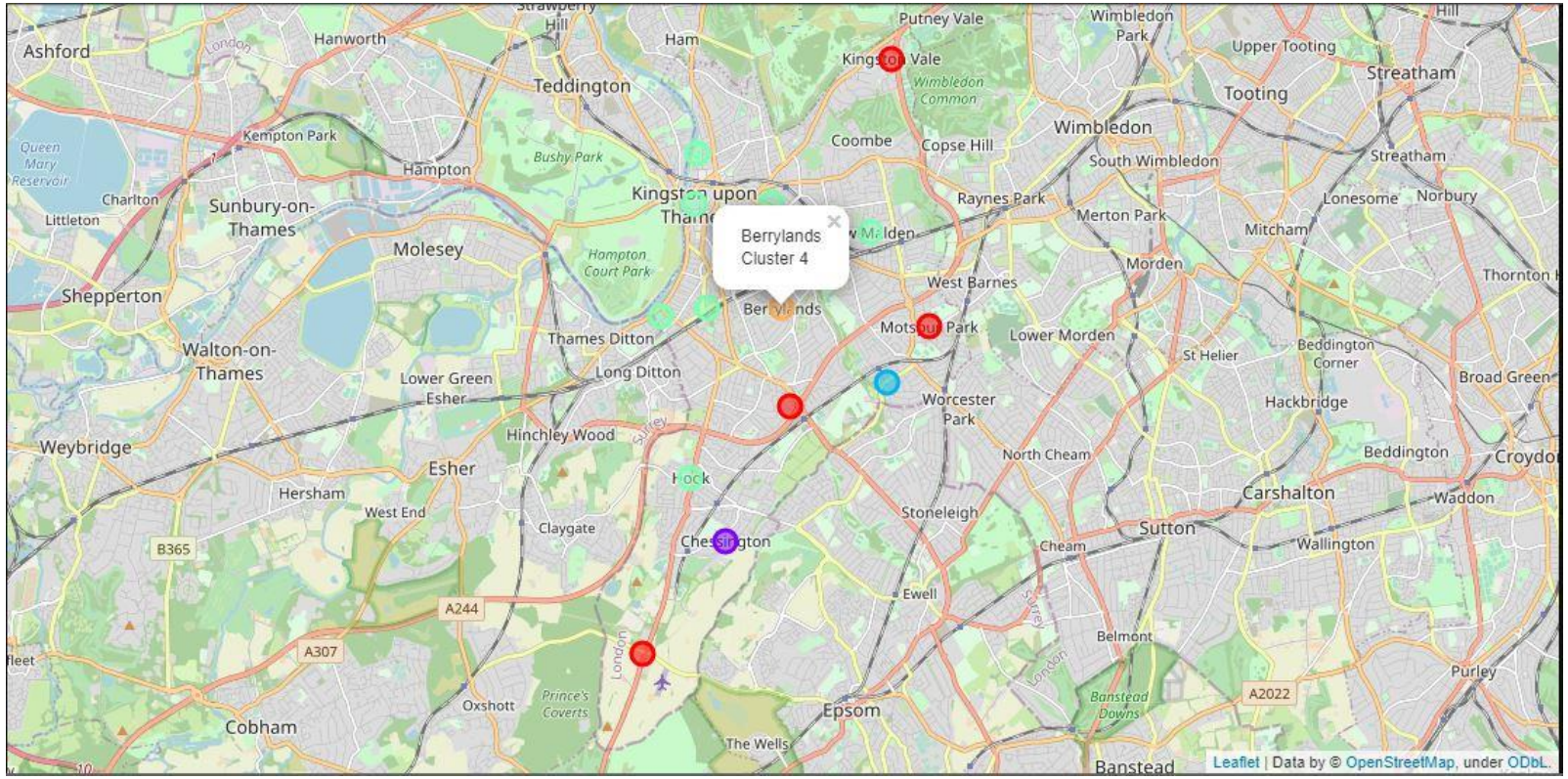


Modelling

- Using final dataset containing neighbourhoods, along with latitude, longitude, we can find all venues within 500 m radius using Foursquare API
- Then venues data is grouped by the neighbours and mean of of the neighbours is calculated
- Used K-mean clustering for grouping the venues
- We set cluster size of 5 for this dataset

Results

- Result after clustering is done
- Each cluster is represented in different colours



Results

- Cluster 1: neighbourhoods in cluster 1

```
[139]: kut_merged[kut_merged['Cluster Labels'] == 0]
```

[139]:

	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
6	Kingston Vale	Kingston upon Thames	51.431850	-0.258138	0	Grocery Store	Sandwich Place	Bar	Soccer Field	Department Store	Discount Store	Dry Cleaner	Electronics Store	Farmers Market	Fast Food Restaurant
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	0	Grocery Store	Garden Center	Pub	Restaurant	Farmers Market	Deli / Bodega	Department Store	Discount Store	Dry Cleaner	Electronics Store
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	0	Park	Gym	Soccer Field	Restaurant	Wine Shop	Farmers Market	Department Store	Discount Store	Dry Cleaner	Electronics Store
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	0	Grocery Store	Restaurant	Train Station	Hotel	Indian Restaurant	Italian Restaurant	Discount Store	Coffee Shop	Furniture / Home Store	Pizza Place

Results

- Cluster 2

```
[140]: kut_merged[kut_merged['Cluster Labels'] == 1]
```

[140]:

	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
2	Chessington	Kingston upon Thames	51.358336	-0.298622	1	Construction & Landscaping	Fish & Chips Shop	Department Store	Discount Store	Dry Cleaner	Electronics Store	Farmers Market	Fast Food Restaurant	Food	Cosmetics Shop

- Cluster 3

```
[141]: kut_merged[kut_merged['Cluster Labels'] == 2]
```

[141]:

	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
11	Old Malden	Kingston upon Thames	51.382484	-0.25909	2	Construction & Landscaping	Pub	Food	Train Station	Bakery	Bar	German Restaurant	Gastropub	Garden Center	Furniture / Home Store

Result

- Cluster 4:

```
[142]: kut_merged[kut_merged['Cluster Labels'] == 3]
```

[142]:

	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	3	Pub	Park	Fish & Chips Shop	Supermarket	Spa	Gym / Fitness Center	Shop & Service	Plaza	Hotel	Indian Restaurant
4	Hook	Kingston upon Thames	51.367898	-0.307145	3	Indian Restaurant	Fish & Chips Shop	Bakery	Supermarket	Department Store	Discount Store	Dry Cleaner	Electronics Store	Farmers Market	Fast Food Restaurant
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	3	Coffee Shop	Café	Sushi Restaurant	Burger Joint	Pub	Asian Restaurant	Portuguese Restaurant	French Restaurant	German Restaurant	Electronics Store
9	New Malden	Kingston upon Thames	51.405335	-0.263407	3	Office	Gastropub	Sushi Restaurant	Supermarket	Bar	Chinese Restaurant	Korean Restaurant	Indian Restaurant	Wine Shop	Electronics Store
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	3	Indian Restaurant	Food	Pub	Italian Restaurant	Fried Chicken Joint	Japanese Restaurant	Hotel	Hardware Store	Wine Shop	Pizza Place
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	3	Indian Restaurant	Coffee Shop	Pub	Café	Italian Restaurant	Restaurant	Hotel	Fast Food Restaurant	Harbor / Marina	Gym / Fitness Center
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	3	Coffee Shop	Pub	Grocery Store	Italian Restaurant	Pharmacy	Thai Restaurant	Tea Room	Gastropub	Train Station	Gym / Fitness Center

Results

- Cluster 5:

```
kut_merged[kut_merged['Cluster Labels'] == 4]
```

	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
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	4	Gym / Fitness Center	Park	Bus Stop	Wine Shop	Fast Food Restaurant	Discount Store	Dry Cleaner	Electronics Store	Farmers Market	Fish & Chips Shop

Conclusion

- This project gives better understanding of clustering algorithm and its application to solve real world problems.
- Grouping of neighbours can be implemented based on crime rate.
- Project has covered several areas of data science including data wrangling, Exploratory Data Analysis, Data preparation, Data visualization as well as applying Machine learning algorithm(K-means Clustering)