

# Housing for Students in London

Sanket Gokhale

July 2020

# The problem

- According to Wikipedia, London has one of the largest concentrations of universities and higher education institutions in the world.
- It has 40 higher education institutions and has a student population of more than 400,000.
- According to an analysis of new data from the UK's Higher Education Statistics Agency (HESA) released on Wednesday,
- Indian student numbers in London grew by 34.7 per cent in 2018-19, marking the largest numbers since 2011-12. More than 7000 Indian students move to London for Higher Education every year.



# Data



The above three types of data are used for the analysis in the case study.

Data Cleaning:

- In this dataset, only the entries after 2014 were considered as the type and pattern of crime has changed over the years.
- Like the crime dataset, the monthly prices of only the year 2016 were used to prepare the dataset.
- The table containing data regarding the Boroughs was scraped. String manipulation techniques were used to match the names of the Boroughs as in the other two datasets.

# Exploratory Data Analysis

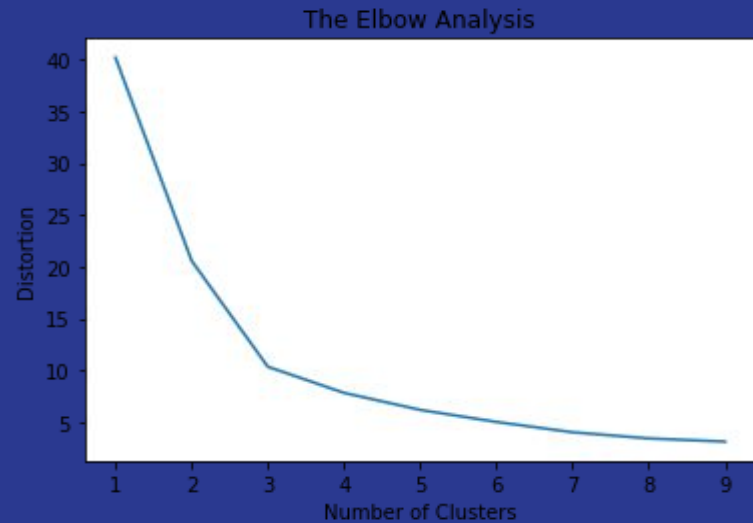
For the Exploratory Data Analysis of the problem, visualization of various features of the Boroughs on maps was done using the Folium library of Python.

As the size and population of each Borough varies considerably, comparing the crime figures on a per capita basis gives a better idea of the likelihood of crime occurring in the Borough. It also gives a more accurate sense of how likely it is that a person staying in any of the given Boroughs is a victim of those crimes.

```
1 Ld_crime["Burglary_per_capita"] = Ld_crime['Burglary']/Ld_crime['Population']
2 Ld_crime["Criminal Damage_per_capita"] = Ld_crime['Criminal Damage']/Ld_crime['Population']
3 Ld_crime["Drugs_per_capita"] = Ld_crime['Drugs']/Ld_crime['Population']
4 Ld_crime["Other Notifiable Offences_per_capita"] = Ld_crime['Other Notifiable Offences']/Ld_crime['Population']
5 Ld_crime["Robbery_per_capita"] = Ld_crime['Robbery']/Ld_crime['Population']
6 Ld_crime["Theft and Handling_per_capita"] = Ld_crime['Theft and Handling']/Ld_crime['Population']
7 Ld_crime["Violence Against the Person_per_capita"] = Ld_crime['Violence Against the Person']/Ld_crime['Population']
8 Ld_crime["Total_per_capita"] = Ld_crime['Total']/Ld_crime['Population']
```



# Modelling and Results



# K-Means Clustering

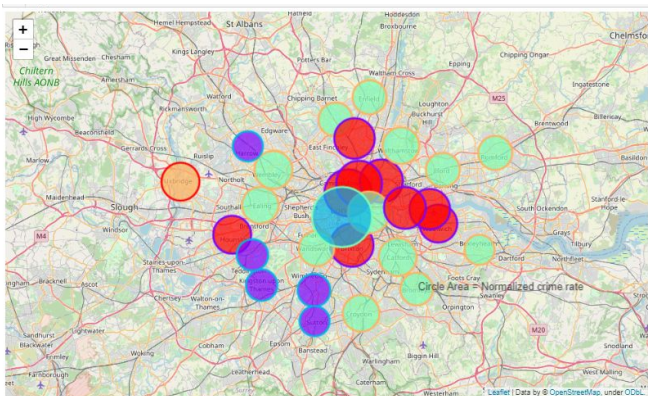
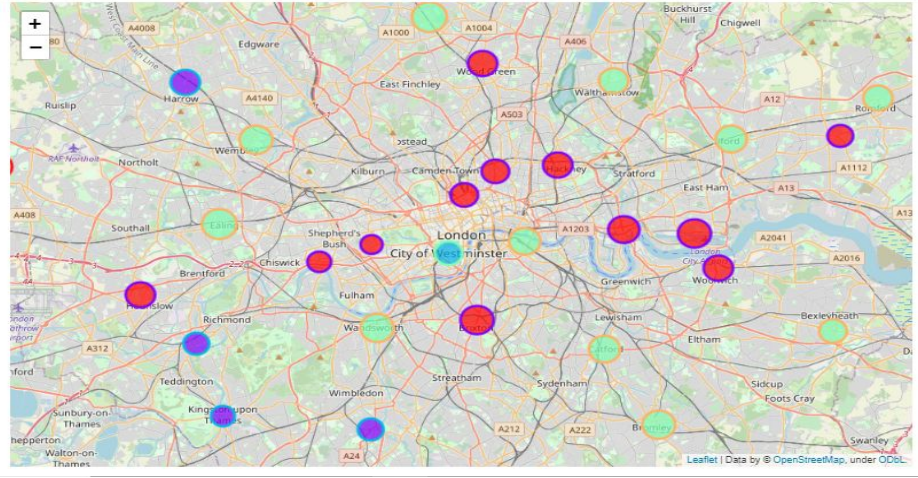
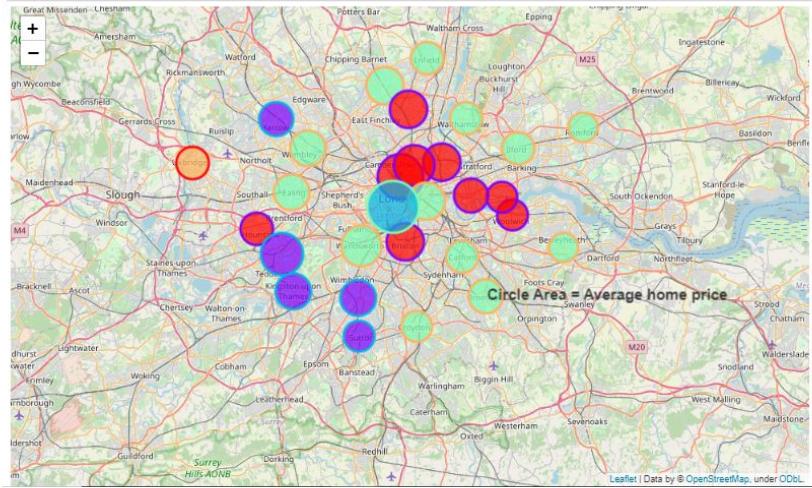
K-means Clustering was used for the process of segmenting the Boroughs. Distortion and elbow analysis was used to finalize the number of clusters at 5.

Number of Clusters		Distortion
0	1	40.172973
1	2	20.559339
2	3	10.333476
3	4	7.815270
4	5	6.172126
5	6	5.008901
6	7	4.011691
7	8	3.416559
8	9	3.112137

Cluster Labels		Number of Boroughs
0	0	9
1	1	5
2	2	1
3	3	13
4	4	1



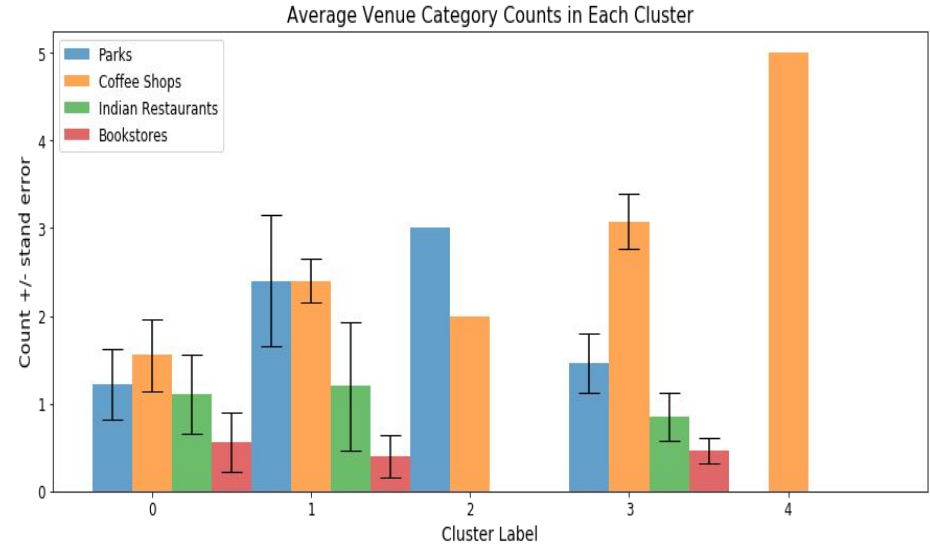
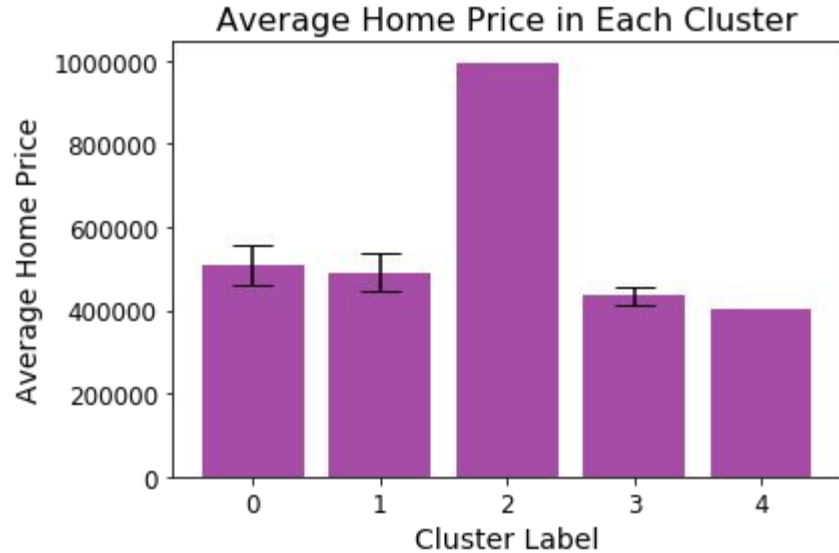
# Map Visualizations



After the clustering, map visualizations were done with features such as population, normalized crime rate, etc to see how the clusters perform with respect to each feature.



## Analysis according to House Prices, Venue Prices and Crime Rates:



# Selection of Clusters and Boroughs

## 1. Home Prices

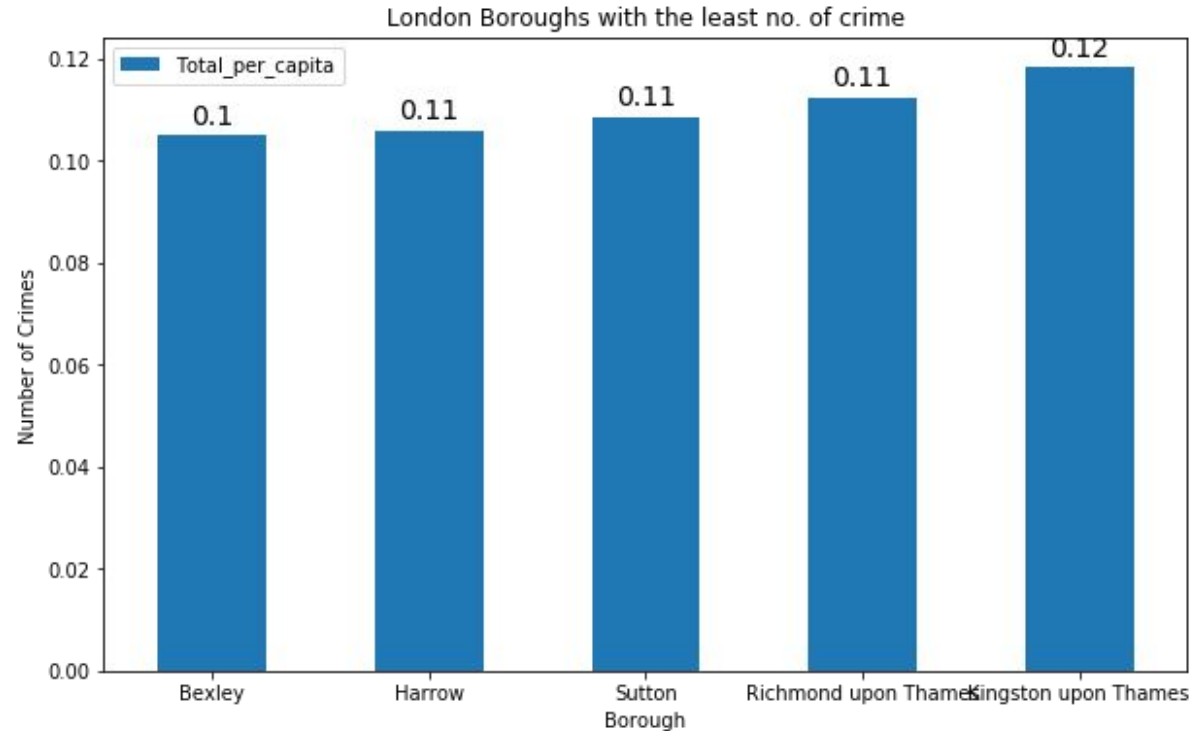
Clusters 0,1,3,4 within range

## 2. Venue Categories

Clusters 0,1 and 3 have all the required venue categories

## 3. Crime Rate

Cluster 1 and 3 have the safest Boroughs



# Conclusion

From the report generated by our analysis, Bexley Borough is the most ideal borough for Indian students arriving in London to stay at. However, Boroughs from Cluster 1 such as Harrow, Sutton, Richmond upon Thames and Kingston upon Thames are also good alternatives and can be considered depending on factors like distance to University, availability of housing and so on. With the help of Foursquare API, other amenities (in addition to Parks, Cafes, Indian Restaurants and Bookstores) can also be used for analysis depending on personal preferences.

