Housing for Students in London

July 2020

Sanket Gokhale

1. Introduction

1.1 Background

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. Among the institutions in London are some of the old and world-famous colleges that today make up the federal University of London, modern universities, as well as a number of smaller and often highly specialised universities and colleges. 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.

1.2 Problem

In a large city and quite expensive city like London, it may become difficult for students looking for a place to stay to find the right neighborhood.his report is targeted towards Indian students who move to London for Higher Education.

2. Data

2.1 Data Sources

Three data sets from open sources were used for the creation of this report. They are:

1. London Crime Data Set - https://www.kaggle.com/jboysen/london-crime
2. London Housing Data Set - https://www.kaggle.com/justinas/housing-in-london
3. List of Boroughs in London Wikipedia Page - https://en.wikipedia.org/wiki/List_of_London_boroughs
   The wikipedia page was scrapped for a list of boroughs in London along with other information like coordinates, area, population, etc.

2.1 Data Cleaning

In this stage, each of the three datasets were handled independently and then merged to form the final dataset.

1. London Crime Data Set -

   In this dataset, only the entries after 2014 were considered as the type and pattern of crime has changed over the years. The data after 2014 ensures that only recent data influences our decision.

	Borough	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
0	Barking and Dagenham	2916	3771	1863	711	1087	10784	11955	33087
1	Barnet	7255	4435	1715	888	1088	18935	14404	48720
2	Bexley	2208	3238	1324	681	380	8557	8442	24830
3	Brent	5332	4408	3904	947	1802	17493	17690	51576
4	Bromley	4738	4497	1598	821	732	15157	13258	40801
5	Camden	5482	3831	3175	928	1929	28543	14772	58660
6	City of London	7	5	16	6	9	233	53	329

2. London Housing Data Set -

   Like the crime dataset, the monthly prices of only the year 2016 were used to prepare the dataset. As we are using the spot prices, and no prediction will be made, only the most recent data is enough to generate the report. The monthly prices were averaged and the mean price of housing in each Borough for the year 2016 was calculated.

	Borough	Average Price 2016
0	Barking & Dagenham	273919.636042
1	Barnet	525939.577300
2	Bexley	321563.508775
3	Brent	489469.418367
4	Bromley	428008.119983

3. List of Boroughs in London Wikipedia Page -

   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. Columns such as ‘Nr in Map’, ‘Political Control’, ‘Local Authority’, etc were dropped as these factors did not affect the topic of the report.

	Borough	Inner	Status	Local authority	Political control	Headquarters	Area (sq mi)	Population (2013 est)[1]	Co-ordinates	Nr. in map
0	Barking and Dagenham [note 1]			Barking and Dagenham London Borough Council	Labour	Town Hall, 1 Town Square	13.93	194,352	51°33′39″N 0°09′21″E / 51.5607°N 0.1557°E /...	25
1	Barnet			Barnet London Borough Council	Conservative	Barnet House, 2 Bristol Avenue, Colindale	33.49	369,088	51°37′31″N 0°09′06″W / 51.6252°N 0.1517°W /...	31
2	Bexley			Bexley London Borough Council	Conservative	Civic Offices, 2 Watling Street	23.38	236,687	51°27′18″N 0°09′02″E / 51.4549°N 0.1505°E /...	23
3	Brent			Brent London Borough Council	Labour	Brent Civic Centre, Engineers Way	16.70	317,264	51°33′32″N 0°16′54″W / 51.5588°N 0.2817°W /...	12
4	Bromley			Bromley London Borough Council	Conservative	Civic Centre, Stockwell Close	57.97	317,899	51°24′14″N 0°01′11″E / 51.4039°N 0.0198°E /...	20

3. 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. Initially, only the population of each Borough was considered and plotted on the map of London, with circles of radius proportional to the population as markers.

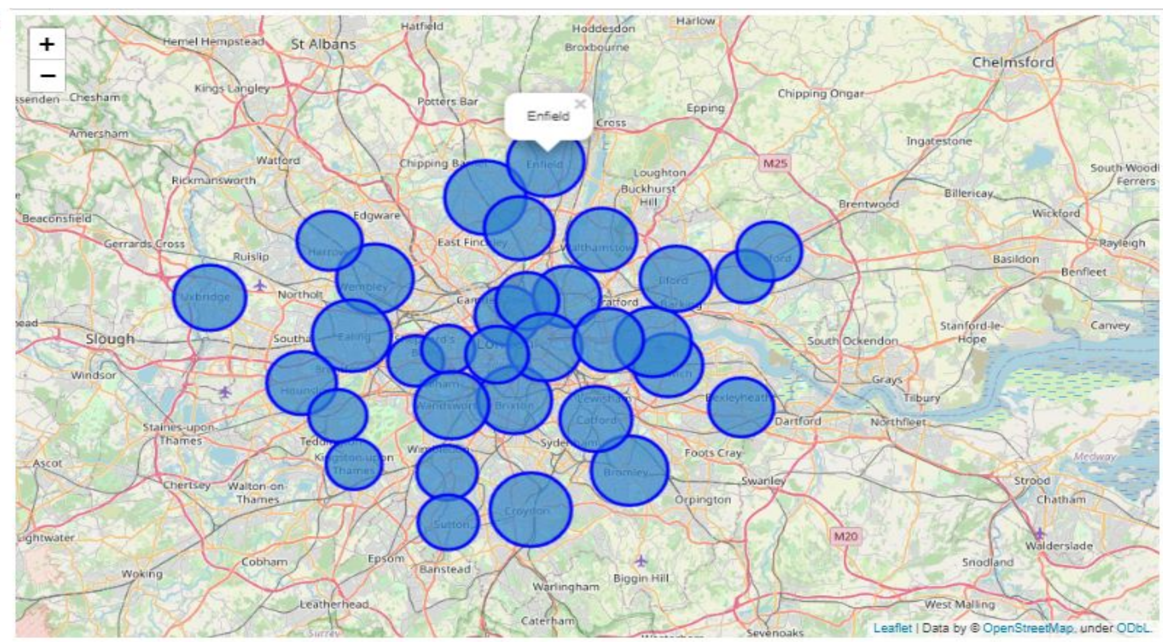


Figure 1: Plotting the Population of each Borough of London

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. Hence, the data was calculated on a per capita basis and added to the dataset.

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

After this, various features such as Robberies per capita in each Borough were plotted to get an idea of the type and amount of crime occurring in each Borough.

Figure 2: Plotting Robberies per capita in each Borough of London

4. Modelling and Results

4.1 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.

In [70]:
```
X_norm = X_norm*weights
```

In [71]:
```
distortions = []
ks = []
for kclusters in range (1,10):
    kmeans = KMeans(n_clusters=kclusters, random_state=21).fit(X_norm)
    ks.append(kclusters)
    distortions.append(kmeans.inertia_)
    ks_distortions = pd.DataFrame({'Number of Clusters': ks, 'Distortion': distortions})
ks_distortions
```

Out[71]:

	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

The cluster label of each Borough was added to the dataset for further analysis using those labels.

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

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.

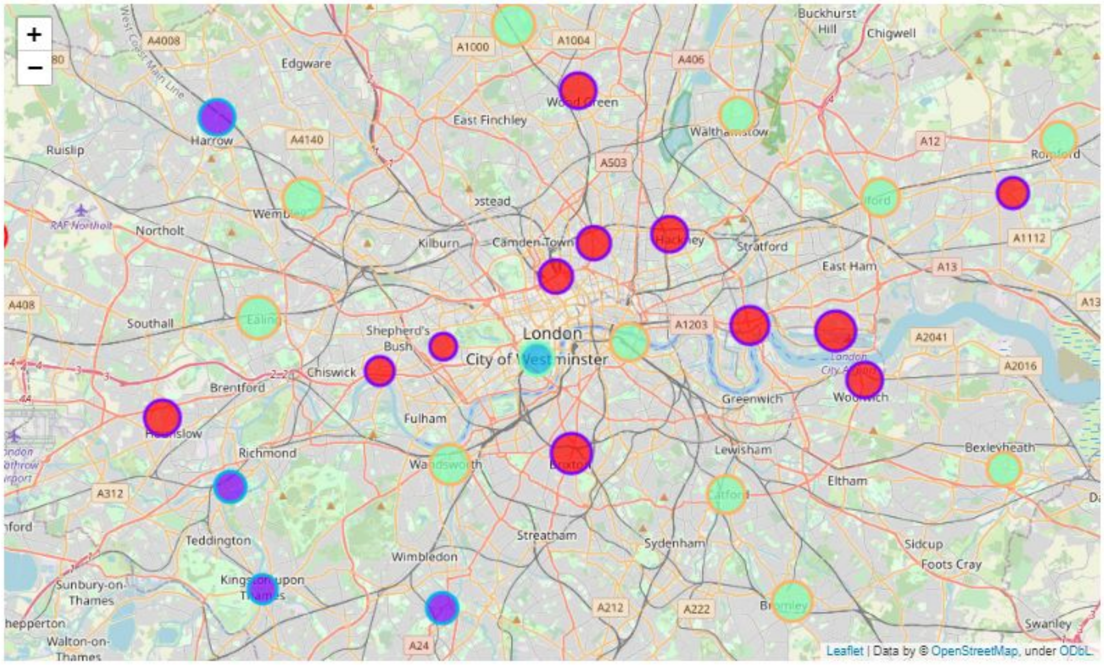


Figure 3: Plotting Population according to clusters in each Borough of London

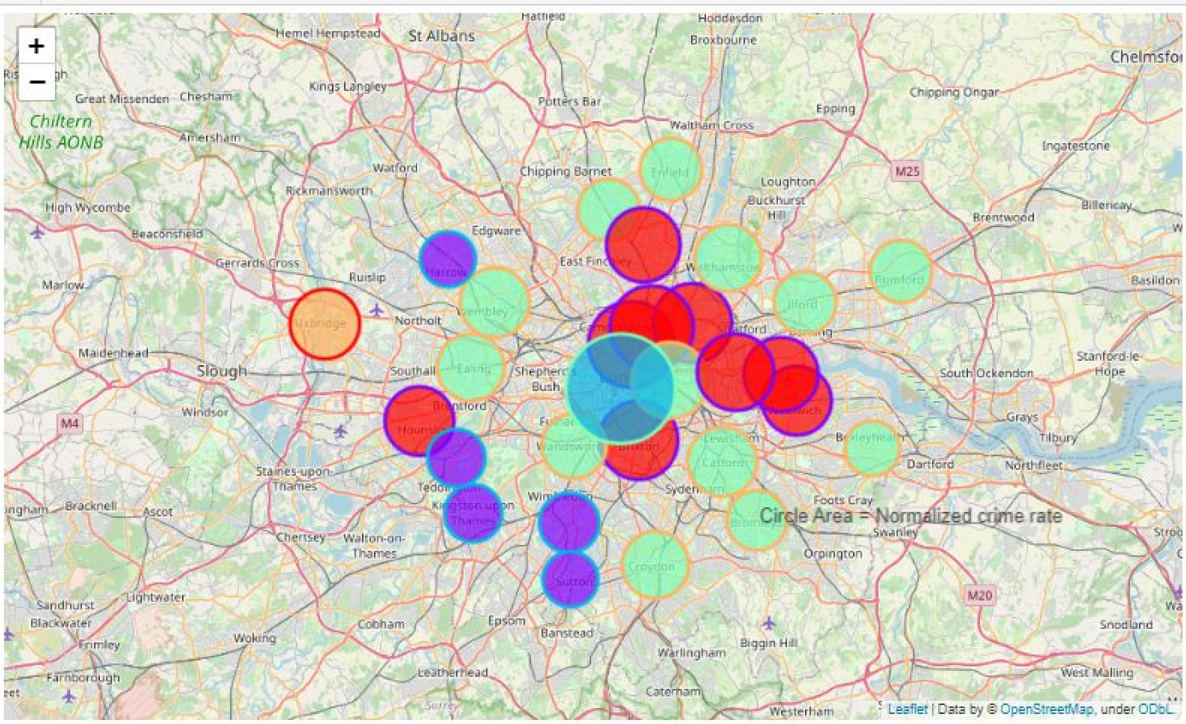


Figure 4: Plotting Normalized Crime Rate according to clusters in each Borough of London

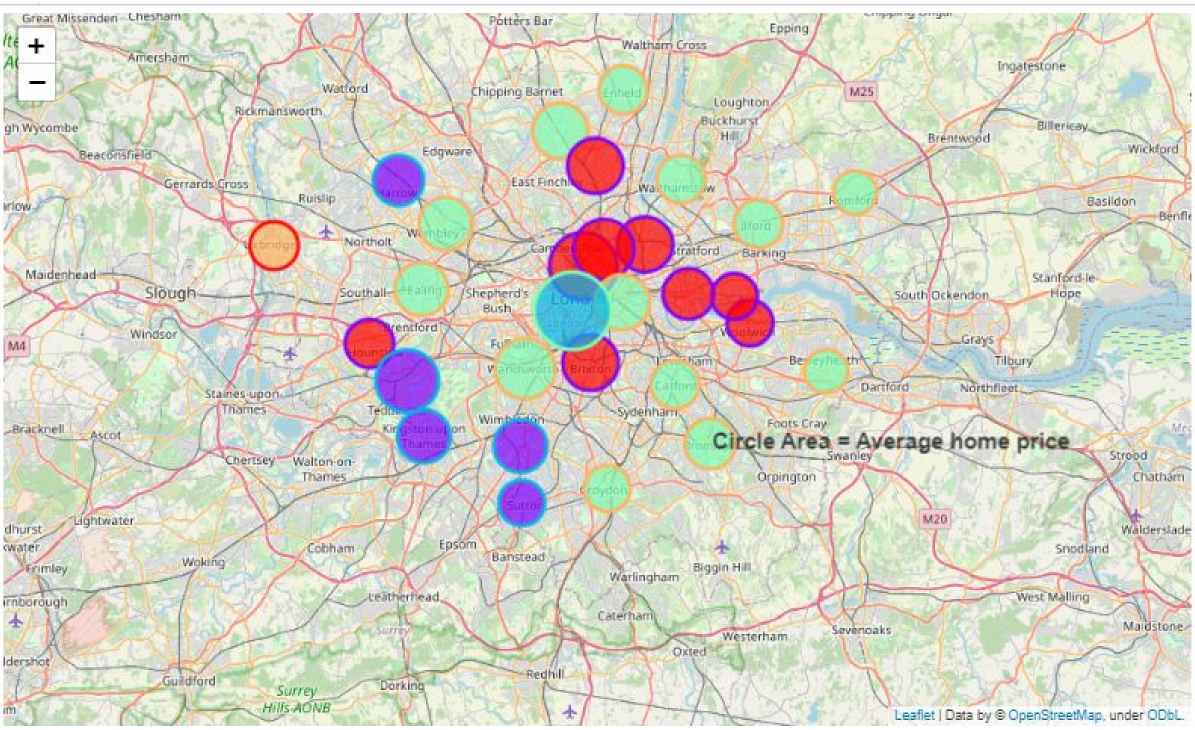
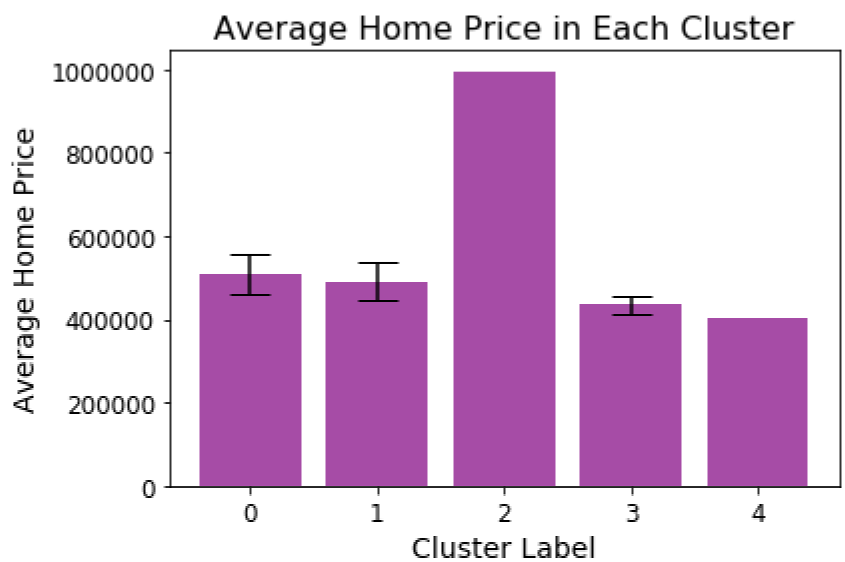


Figure 5: Plotting Average House Price according to clusters in each Borough of London

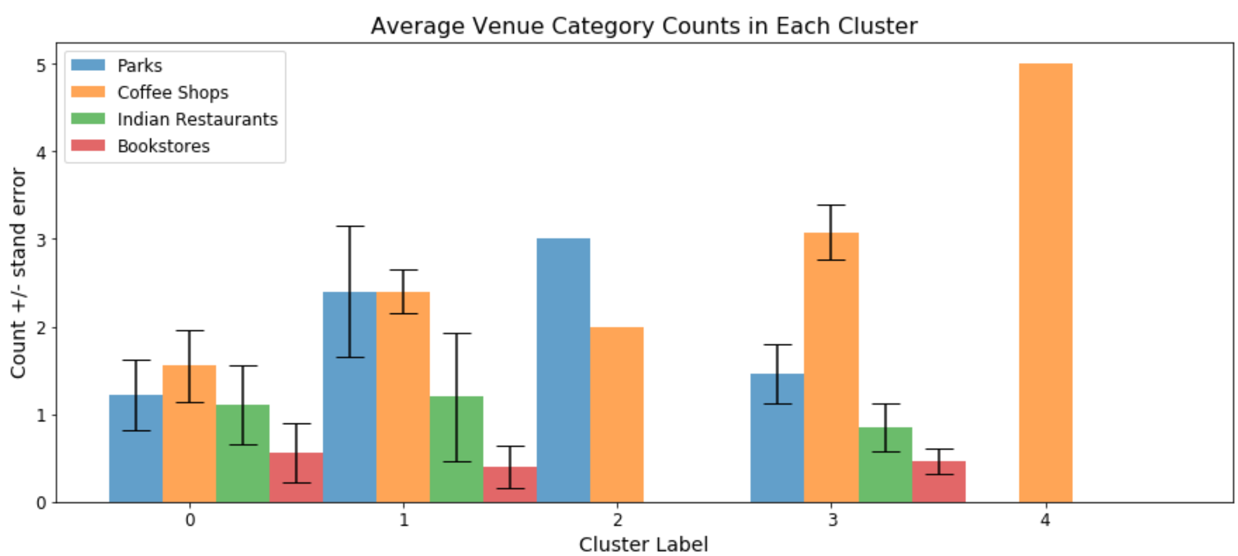
While House Prices and Crime Rate were the primary filters used for differentiating the clusters and Boroughs, Foursquare API provides us with nearby Points of Interest for each of the Borough. Querying the Foursquare API, a list of nearby venues was found for each of the Borough. The report, primarily targeted for students from India coming to London for Higher Education, focussed on amenities that might be deemed helpful for the mentioned demographic. As a result, Parks, Coffee Shops, Indian Restaurants and Bookstores were chosen to help find the ideal Borough(s) for students.

Cluster Labels								Average Price 2016		...	Bookstore					Average Price 2016	Park	Coffee Shop	Indian Restaurant	Bookstore
count	mean	std	min	25%	50%	75%	max	count	mean	...	min	25%	50%	75%	max	error	error	error	error	error
9.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.0	510391.963694	...	0.0	0.0	0.0	1.0	3.0	48590.903797	0.400617	0.412011	0.454742	0.337931
5.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0	5.0	491305.425955	...	0.0	0.0	0.0	1.0	1.0	46805.015678	0.748331	0.244949	0.734847	0.24494
1.0	2.0	NaN	2.0	2.0	2.0	2.0	2.0	1.0	995543.362333	...	0.0	0.0	0.0	0.0	0.0	NaN	NaN	NaN	NaN	NaN
13.0	3.0	0.0	3.0	3.0	3.0	3.0	3.0	13.0	434713.024188	...	0.0	0.0	0.0	1.0	1.0	22999.399986	0.332346	0.309291	0.273771	0.14391
1.0	4.0	NaN	4.0	4.0	4.0	4.0	4.0	1.0	400791.486383	...	0.0	0.0	0.0	0.0	0.0	NaN	NaN	NaN	NaN	NaN

4.2 Results



The plot above shows that the cluster 2 has the highest average home price and hence boroughs in this cluster will not be affordable for students. All other clusters have an average price which is in a very narrow range and hence we need further analysis to differentiate within these clusters.



Using the above plot, we can see that Clusters 0, 1 and 3 have an abundance of amenities that we have deemed important for students. Hence cluster 4 can be ruled out too.

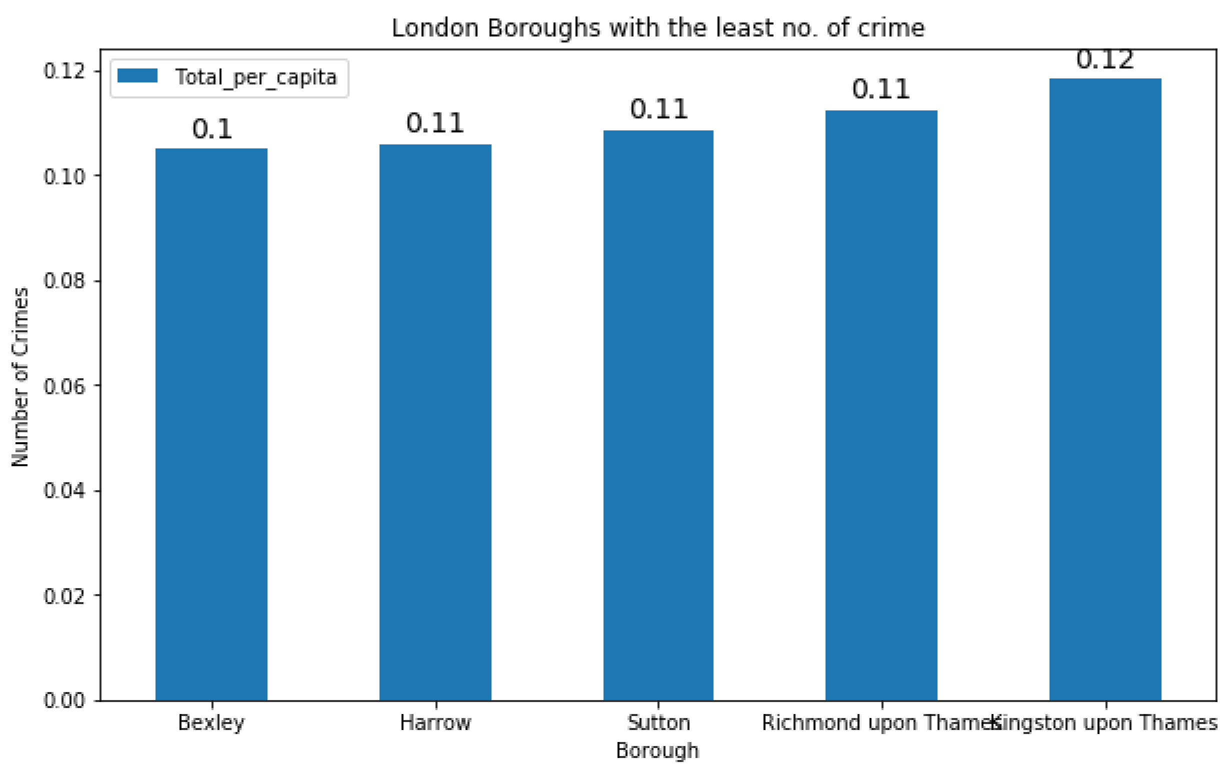
```
London_df.sort_values(['Total_per_capita'], ascending = True, axis = 0, inplace = True )

df_bot5 = London_df.head()
df_bot5
```

	Cluster Labels	Borough	Area	Population	Latitude	Longitude	Burglary_per_capita	Criminal Damage_per_capita	Drugs_per_capita	Other Notifiable Offences_per_capita	...	Turkish Restaurant
1	3	Bexley	23.38	236687.0	51.4549	0.1505	0.009329	0.013681	0.005594	0.002877	...	0
11	1	Harrow	19.49	243372.0	51.5898	-0.3346	0.015059	0.009829	0.004503	0.001948	...	0
24	1	Sutton	16.93	195914.0	51.3618	-0.1945	0.012822	0.013460	0.004905	0.002220	...	0
22	1	Richmond upon Thames	22.17	191365.0	51.4479	-0.3260	0.014684	0.011611	0.003742	0.002268	...	0
16	1	Kingston upon Thames	14.38	166793.0	51.4085	-0.3064	0.010828	0.012483	0.009857	0.002338	...	1

5 rows × 196 columns

Sorting the Boroughs according to the Total Crime per capita, we find that 4 out of the top 5 Safest Boroughs are in cluster 1 and 1 Borough is from Cluster 3.



5. 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.