2019

Potential Student Accommodation Locations in London

A STUDY OF STUDENT ACCOMMODATION IN LONDON

BENJAMIN FROST

1 CONTENTS

2 INTRODUCTION

The business problem I have chosen to set out to solve for this project is based around the idea of needing to find the best location within London to set up new student accommodation. This is a problem that will be faced increasingly frequently by developers as the numbers of students studying in London is increasing year on year. “In 2015/16 there were 2.3 million students in the UK with about 372,000 or 16% studying in London”. (London Higher, 2017). As this number continues to grow, students will be forced to pay more and more towards rent, to the benefit of private accommodation companies. Now may be the best time in years for investment in London, with property prices stalling. “In London prices fell 1.2 per cent in the year to April – more than any other UK region” (Chapman, 2019).

This project aims to recommend different areas within London where demand may be high for new student accommodation investment. I will be looking at existing accommodation buildings and property prices to help educate an answer.

3 DATA

3.1 DATA SOURCE AND VALIDITY

The data used in this project will be sourced from Wikipedia, and government data collection websites. I have used data from a range of different sources so that I may piece together a picture of different factors affecting the viability of new property developments. The data sources can be found in the appendix, along with being referenced in the notebook, and consist of two types of data accumulation; csv tables and API requests. I have used the csv tables as the basis for my datasets, and the API requests to retrieve location data for the universities and buildings that I have used.

The final dataset used in this project is as follows:

	University	Latitude	Longitude	Distance	Students	Borough	Value
0	Brunel University London	51.532553	-0.473994	24.831126	13130	Hillingdon	415000
1	University of West London	51.489519	-0.313640	13.766399	10390	Hounslow	395000
2	University of East London	51.507273	0.064405	12.673621	13215	Newham	390500
3	Kingston University	51.437255	-0.251575	12.299441	19470	Wandsworth	654000
4	Middlesex University	51.590297	-0.229632	11.829639	19505	Barnet	533500
5	University of Greenwich	51.482542	-0.006653	8.314634	19915	Greenwich	420000
6	King's College London	51.469844	-0.089279	4.881667	30565	Southwark	530000
7	London Metropolitan University	51.548156	-0.106189	4.339511	12145	Islington	615000
8	Imperial College London	51.498871	-0.175608	4.176682	17690	Westminster	1025000
9	University of Westminster	51.518252	-0.141074	1.848184	19650	Westminster	1025000
10	London South Bank University	51.497788	-0.101859	1.753800	17985	Southwark	530000
11	University College London	51.523161	-0.128204	1.637336	37905	Camden	760000
12	City, University of London	51.521767	-0.130191	1.568050	19405	Camden	760000
13	University of the Arts London	51.517642	-0.116389	0.873328	18290	Camden	760000
14	London School of Economics	51.514429	-0.116588	0.518405	11210	Westminster	1025000

3.2 API

The API references are as follows:

- GeoPy was able to return latitude and longitude coordinates given the names of universities, which was essential to helping me to place the universities on the map in bulk. It was also used to calculate the distances between locations given their coordinates.
- Foursquare was used to return nearby student accommodation companies around each university.
- Finally, Postcodes.io was used to return the borough of each location to me in a format that would be reliably read each time it was run.

3.3 CSV

In terms of the csv files, the wiki tables were scraped of the websites using the very useful website 'https://wikitable2csv.ggor.de/'. The data from data.london.gov.uk was easily accessible in the csv format. The files were uploaded to the IBM database and accessed from there into my notebook. The csv file sources were used as follows:

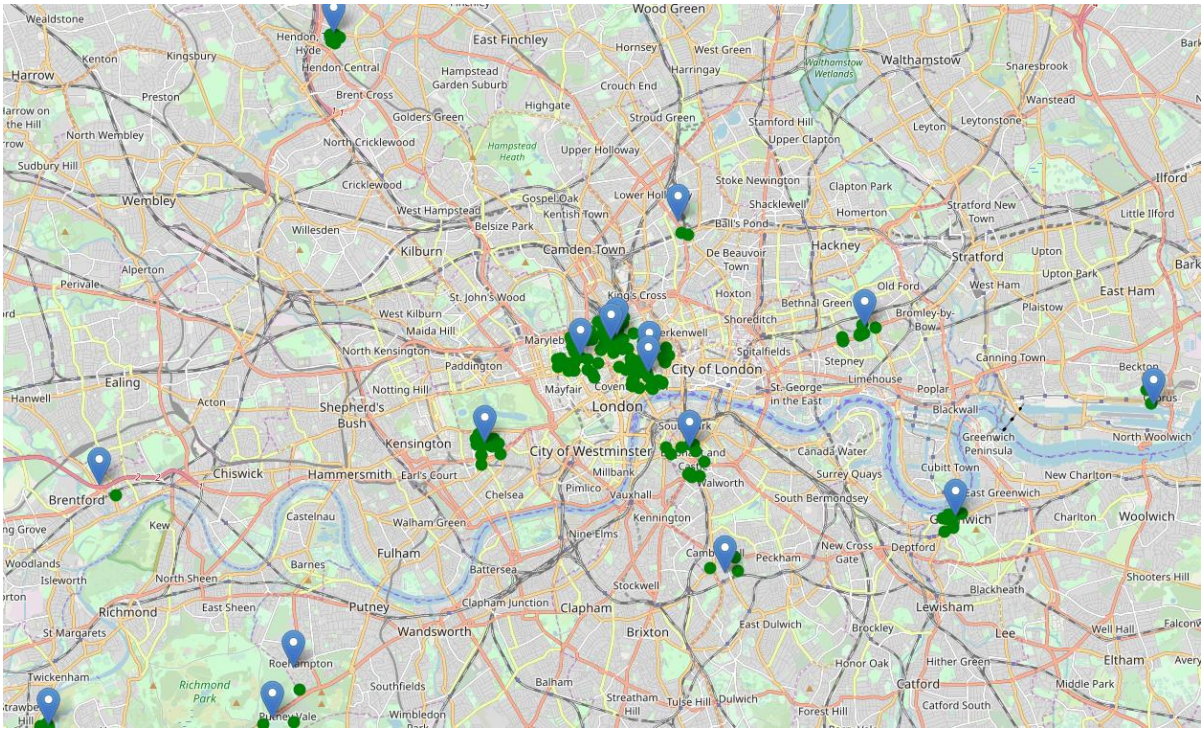
- Wikipedia was used to obtain a list of universities within London. This was an essential first step as it gave me only the universities within the city I was studying. Wikipedia also gave me a list of student numbers for every university within the UK. This table was merged with the list of universities within London.
- The gov.uk website was used to give me housing price data for boroughs within the UK. This dataset was over 12000 rows initially, however after processing this dropped to around 40 rows. I selected the median property prices for London boroughs.
- LatLong was used to give me an accurate value for the centre of London from which I could calculate the distance of universities.

Of the universities given to me by Wikipedia, about 4 of them did not return a value for latitude or longitude by the Foursquare API. I made the decision to remove these universities from my data as they did not represent a large proportion of my data, nor would their absence impact my results enough to bring a different outcome.

4 METHODOLOGY

4.1 UNIVERSITIES AND ACCOMMODATION PROXIMITY

With the goal of finding an area of London with high demand and low property prices, it was intuitive to plot my data on a map to be able to view gaps in the market:

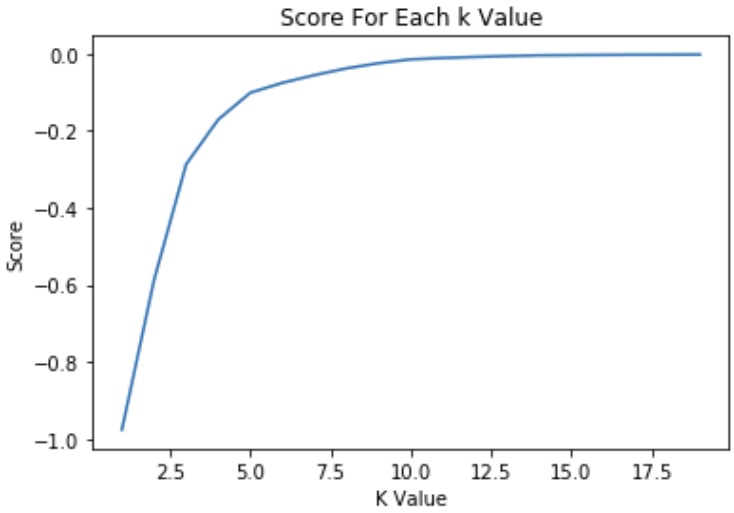


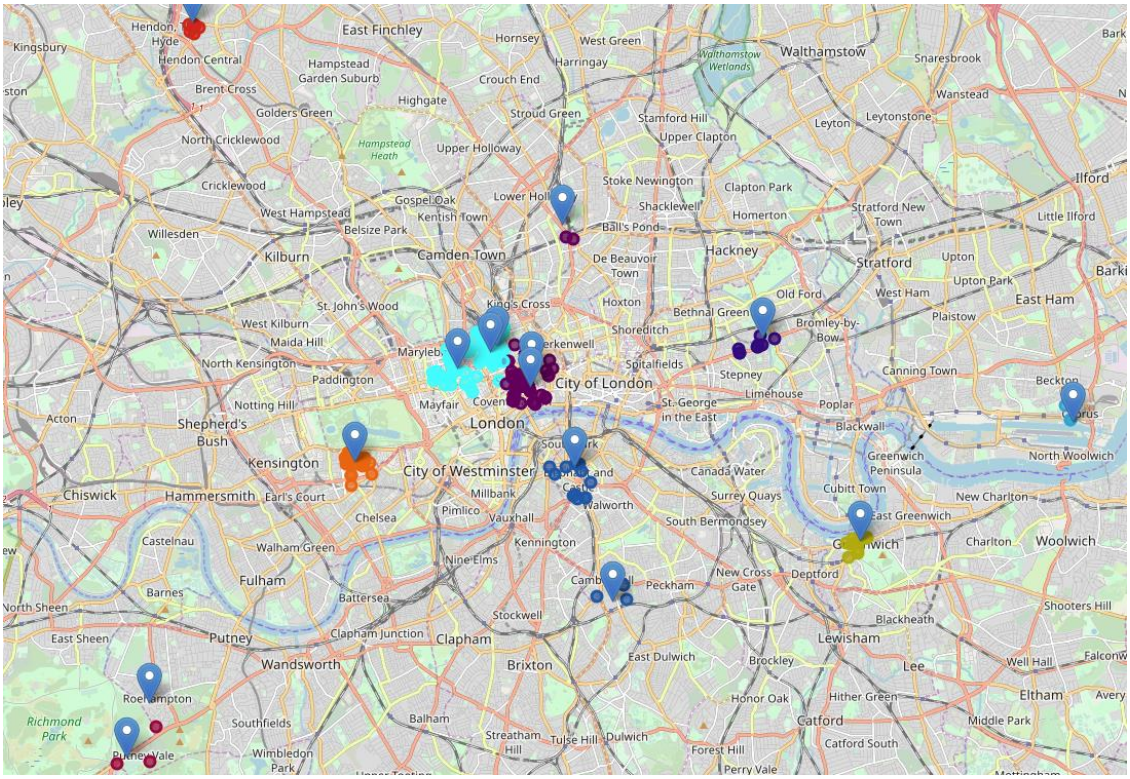
This map shows the universities in my dataset along with their location within London. As you can tell, London universities are spread out across the suburbs with a high concentration in the very centre of the city. The python library Folium was used for all my maps.

The green markers show the student accommodations around each university. As there is such a high concentration of universities in the very centre of the city, there are a very high number of accommodation buildings here.

4.2 K-MEANS CLUSTERING

In order to better visualise the data we have here, I used a k-means clustering algorithm to group together accommodation in the same proximity. To do this I used the elbow method to find the optimum value for k for my algorithm to use. Clearly shown in the graph below is the drop off of change in score after each increment in K value at around 10. Therefore, the k value of 10 was used for this project, and 10 clusters of data were used.



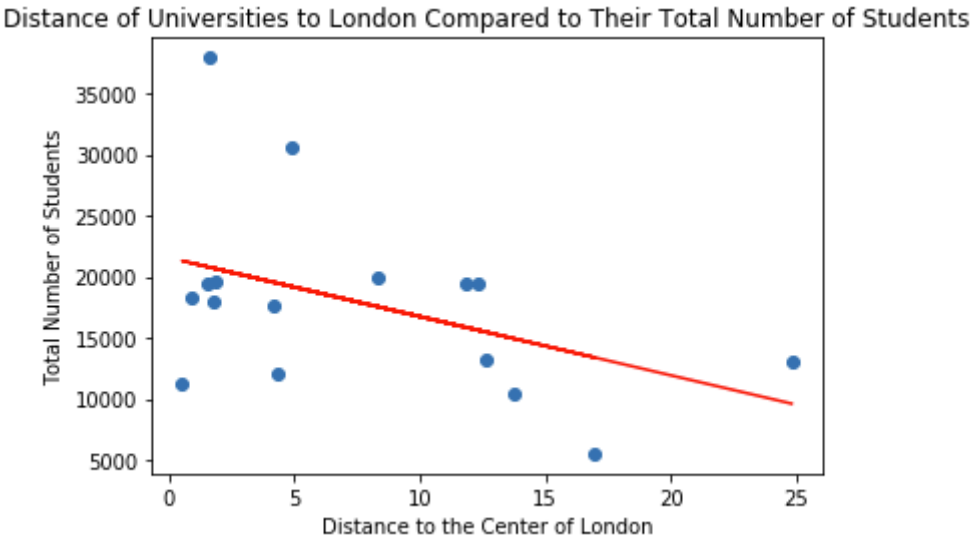


After clustering the data, we ended up with a much clearer picture of where the highest concentration of existing student accommodation buildings are.

Very large clusters are found in the centre of London, with much smaller clusters around universities in the surrounding areas. In particular, there is one area in the centre of London with one very large cluster. This is shown in light blue on the map. The python library Sci-Kit Learn was used for all my machine learning algorithms.

4.3 Linear Regression

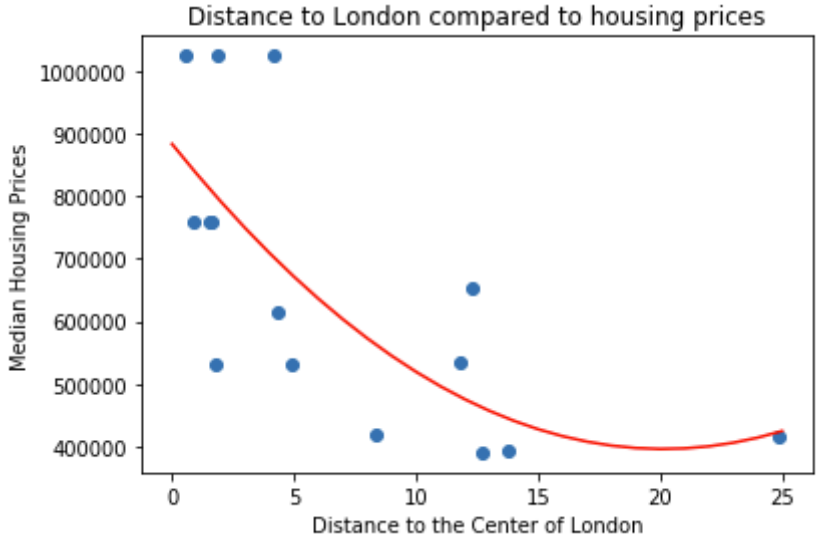
When comparing the total number of students in each university, it was clear that the data followed a linear pattern. As you get closer to the centre of London, the number of students increases by a linear amount, found in my data to be y = -480x + 21545. This reinforces the idea that there is a negative correlation between these columns.



Although there is a negative correlation between this data, it is not strong. In fact, the correlation coefficient of this data is -0.44. So, although universities with a closer distance do have a greater student number, this is not always the case. It is the case however that many of the data points are on the left-hand side of the graph, showing that in the centre of the city there is a higher density of universities.

4.4 Polynomial Regression

Property prices also play a large part in the viability of starting a business in a particular area. It is reasonable to think that as you get closer to London that prices will increase, but by how much will they increase?



As you can tell from this data, the relationship between property prices and distance to London does not follow a linear relationship. The equation of this line is $y = 883784 - 48472x + 1204x^2$. Clearly,

there is a drastic increase in property prices following a quadratic curve. The prices in this dataset range from £395,000 all the way up to £1.03 million. Such a drastic change in property prices may be enough to dissuade property developers with less capital.

5 RESULTS

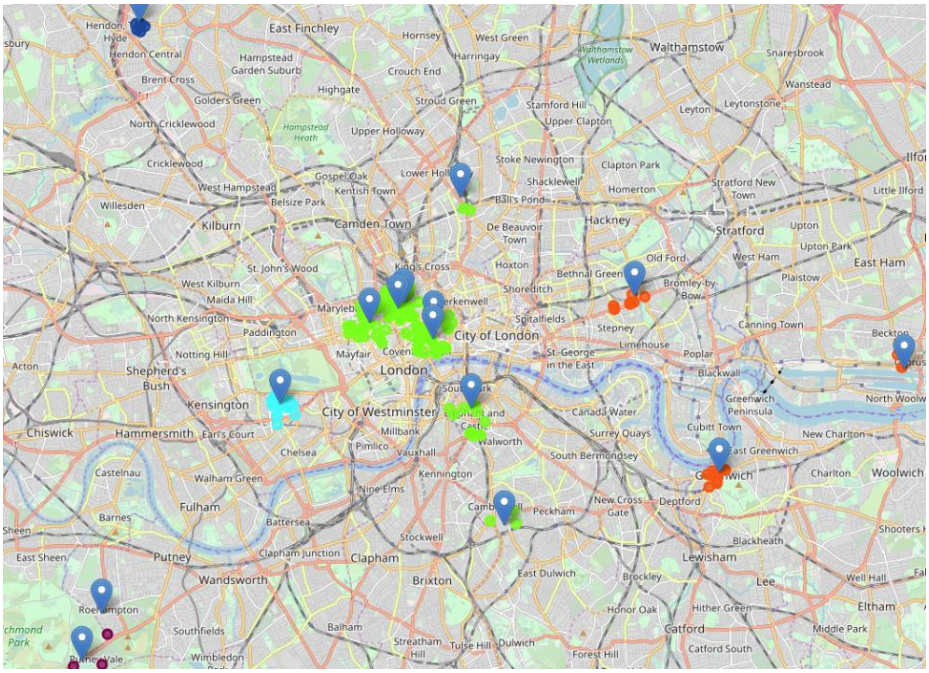
5.1 FOURSQUARE API

The results found by my research and analysis were very helpful in understanding and drawing conclusions from my data, although they had some limitations. The Foursquare API was an invaluable tool for retrieving nearby accommodation locations, but it required trial and error to find the search term that would retrieve mostly student accommodation. Even with the correct search term, about 10% of the returned locations were irrelevant. Such locations had names like 'Students' Union' or 'Student Central', which although matched my search term, were not accommodation buildings. Despite this, I still believe that my results hold water and are a good general indication of the trends in existing accommodation.

5.2 CLUSTERING

The elbow method graph in section 4.2 was a good indication of which value to use to cluster my data. Although the value I used was 10, the graph is not obvious and could be interpreted in many ways. Another answer from the graph would be to use a value of 5, where the score stops dropping off as heavily. I recognise that this value would have been more accurate given the method I was using, however given how many universities were in my dataset, it didn't make sense to have a small number of clusters. Especially in the centre of London, I didn't want all those buildings to be clustered together. The value that I chose gives a good level of detail and is more useful in identifying clusters of buildings that are very close together and of interest.

Indeed, below is the same map but generated with a k value of 5. Far less detail is seen and in my opinion it lacks the results that were intended from the clustering exercise.



The conclusions to be drawn from the final version of the map are a great help in identifying areas of interest for further research. Within the very centre of the city, five universities share between them two very large and close together clusters of student accommodation, shown in light blue and purple. I would recommend research into if this area is saturated with student accommodation, or if the demand is so high from the nearby universities that the existing accommodation is not enough to meet demand.

5.3 Linear Regression

As mentioned in this section, the correlation coefficient for this data was low enough to cast doubt into the relevancy of my findings. There is a correlation, but I don't think there is enough to draw meaningful conclusions from. If I was working with a larger dataset then perhaps it would give me more confidence to say that there is absolutely a correlation here, but in my case I would not recommend drawing conclusions from this graph with any certainty.

5.4 Polynomial Regression

The graph in 4.4 Polynomial Regression was of specific interest to me as it shows how drastically the price of properties change as one gets closer to the centre of town. Not only is there a relationship between these sets of data, but a quadratic relationship exists that shows that as one moves closer to the centre, the price of housing will skyrocket. To me the results from this graph can be drawn with far more certainty than those from section 4.3.

The aspect of this graph to note is the fact there is a noticeable dip between the penultimate and final data points. I believe this to be a product of having no data between these points and therefore leaving a large gap. In reality, I would be very surprised to see a dip in housing prices in this area.

A limit of the data I was using was the fact that I did not have a large number of samples in my dataset. I was working with 15 universities, which gave me results to draw conclusions from, but left gaps in my data like those found in 4.4. Although this does not invalidate my findings, I would have liked to have more data and improve confidence in my results.

6 Discussion

6.1 Existing Accommodation

It is clear from the data that the majority of accommodation in London is very close to the centre of the city. I would recommend further research into the characteristics of the clusters of data that I found in the very centre of London, since this area seems to be a popular spot for property developers to set up new accommodation buildings.

It could be argued that a location in the outskirts of the city may be better suited to new accommodation since there are fewer competitors.

6.2 Student Numbers

Although weak, there is a trend towards having larger universities towards the centre of town; Opening accommodation in this area has the added benefit of being able to service more students and generate more revenue. The centre of the city also has a higher density of universities compared to the universities further away that may be the only institution for miles. This means that not only

can more students be clients in the centre, but there are more universities to build relationships with and create long term business partnerships.

6.3 Property Prices

Property prices is another factor that may dissuade developers from setting up businesses in the centre of the city. The closer one gets to London the more they will have to pay for land. However, looking at the graph, this does not seem to have deterred companies in the past.

One solution to the property prices in the very centre of town is to develop properties in the suburbs and then encourage students to commute into their universities. For example, a new accommodation building was developed by the University of London in Stratford, and students are being encouraged to use transport links. In return for the longer commute, students can find slightly cheaper accommodation, along with a slightly calmer neighbourhood atmosphere. (University of London, 2019)

7 Conclusion

Looking at the analysis that's been achieved on the data that's available, a promising conclusion can be drawn from the data. It seems lucrative to be in the centre of London among the highest density of students and universities in the city. The elements to consider from doing this is that there will be far more competition, and initial set up costs will be significantly higher due to property prices.

My recommendation would be to focus research in the very centre of town to find out why there is such a high density of student accommodation, and if there is a gap in the market for a new building.

This project was completed as part of the IBM Data Science Professional Certificate, and has helped me immensely to reinforce my knowledge of data science and machine learning. Given the opportunity to do it again, I would focus my research on a more specific area. Only once I was well into my project did I realise that it would have been more beneficial to specify my research in central London student accommodation. I would conduct exploratory analysis of the market and make an informed decision on where to research.

The notebook created for this project can be found on my GitHub repository:
https://github.com/BenMarkFrost/Coursera_Capstone

8 Bibliography

Chapman, B., 2019. *UK housing market slows as prices fall in London and South East England.* [Online]
Available at: https://www.independent.co.uk/news/business/news/uk-house-prices-london-south-east-property-market-slowdown-a8965021.html
[Accessed 21 06 2019].

London Higher, 2017. *STUDENT NUMBERS IN LONDON.* [Online]
Available at: https://www.londonhigher.ac.uk/ceo-blog/student-numbers-in-london/
[Accessed 21 06 2019].

University of London, 2019. *Eleanor Rosa House.* [Online]
Available at: https://halls.london.ac.uk/eleanor-rosa
[Accessed 21 06 2019].

8.1 DATA SOURCES:

- https://en.wikipedia.org/wiki/List_of_universities_and_higher_education_colleges_in_London
- https://www.latlong.net/place/london-uk-14153.html
- https://en.wikipedia.org/wiki/List_of_universities_in_the_United_Kingdom_by_enrolment
- https://data.london.gov.uk/dataset/average-house-prices?resource=b1b0079e-698c-4c0b-b8c7-aa6189590ca4
- Foursquare.com
- GeoPy
- Postcodes.io