Applied Data Science Capstone

Capstone Project - The Battle of Neighborhoods

M. Zimmer, 12.06.2020

1. Introduction

Universities in the UK are among the best worldwide - attracting thousands of students from around the globe per year. Major focus is usually given to bigger and more well-known cities like London, Cambridge or Oxford (Figure 1).

Rank	Institution	Student Satisfaction	Research Quality	Graduate Prospects	Student-Staff Ratio	Overall Score
1	Cambridge	4.09	3.33	86.7	11.2	1000
2	Oxford	4.1	3.34	83.4	10.4	989
3	St Andrews	4.26	3.13	79.7	11.7	944
4	London School of Economics	3.67	3.35	86.1	11.1	934
5	Imperial College London	4.02	3.36	90.4	11.4	933
6	Durham	4.01	3.14	84.8	15.0	915
7	Lancaster	4.14	3.15	89.2	12.6	914
8	Loughborough	4.18	2.95	86.8	13.4	905
9	Bath	4.05	3.17	87.4	15.6	882
10	University College London	3.87	3.22	83.8	10.5	880
11	Exeter	4.08	3.08	85.5	15.5	868
11	Warwick	4.05	3.22	83.2	13.0	868
13	Birmingham	4.01	3.07	85.3	13.9	841
14	Leeds	4.09	3.13	81.2	13.4	838
15	Manchester	3.98	3.16	80.4	13.3	827
16	Edinburgh	3.84	3.18	77.3	11.9	826
16	Bristol	3.95	3.18	80.5	13.5	826
18	Glasgow	4.06	3.1	85.4	13.7	823
19	Nottingham	4.02	3.09	86.0	14.5	814
20	Southampton	4.03	3.15	82.1	13.5	804

Figure 1. University League Table 2020 *[1]*.

But what about other parts of the UK? One region for studying among some of the top universities, is the East Midlands, which is one of the nine official regions of England in the eastern part of central England. It comprises the counties of Derbyshire, Leicestershire, Rutland, Nottinghamshire, Lincolnshire and Northamptonshire and is home to the beautiful Peak District [2].

Having experienced and currently still experiencing university life abroad and here in the UK, the focus of this project will be on providing students in the East Midlands some additional information for their life at university. While there are many things to consider when it comes to beginning your studies, one important aspect is the housing situation. For many, it is the first time living away from home, having to take responsibility for paying bills, etc. As such, we will be looking at places to rent, concentrating on the city of Nottingham. In addition, we will investigate other aspects that might be important for students when thinking about the location of their accommodation, such as entertainment, proximity to university, and public transport.

2. Data

For this project, we will require data about the available properties for rent in Nottingham, geo location data, and information about nearby things to do. Thus, we will use datasets from the following sources:

- Data from Zoopla. Zoopla is one of the UK's largest online real estate portals, where one can find properties for rent and sale. Easy access to these data for further processing are provided through their very own Zoopla API. We use the python wrapper zoopla as it facilitates working with the API. One limitation that is to note here, is that the Zoopla API limits the size for each page of results to a 100. To have a better working dataset, the API call was repeated a few times for different result pages.
- To plot the properties in their appropriate districts, the data for the electoral ward boundaries from 2019 for Nottingham City have downloaded as a json file and used in conjunction with a folium map.
- Using the Foursquare City Guide Developer API provides us with things to do in Nottingham, such as a list of places to eat, shop and visit.

3. Methodology

This section covers the main components of the analysis and is comprised of the following stages:

1. Exploratory Data Analysis
2. Data Pre-processing
3. Visualisation

3.1. Exploratory Data Analysis and Data Pre-processing

Overall, 500 rows of data were obtained. The first five entries of the dataset of the features that were chosen are shown in Figure 2.

	num_bedrooms	num_bathrooms	price	property_type	category	street_name	county	post_town	outcode	latitude	longitude
0	8	8	1320.0	Flat	Residential	Byard Lane	Nottingham	Nottingham	NG1	52.952120	-1.146667
1	1	1	1175.0	Detached house	Residential	Watson avenue	Nottingham	Nottingham	NG3	52.961685	-1.112865
2	5	3	1154.0	Detached house	Residential		Nottingham	Nottingham	NG9	52.939785	-1.220182
3	8	8	1112.0	Flat	Residential	Maid Marian Way	Nottingham	Nottingham	NG1	52.950990	-1.151675
4	8	3	1080.0	Terraced house	Residential		Nottingham	Nottingham	NG1	52.964428	-1.152991

Figure 2. Resulting dataset.

The data seem well organised at first glance. We can, however, see that a few values in the 'street_name' are missing, which were decided to be removed as geolocation data is present

and street name would not add much detail as such. Additionally, inconsistent upper and lowercase spelling in the 'street_names' were identified and resolved by making all street names uppercase. 'County' and 'post_town' seem to be containing the same values, which seems to be in this case redundant. The column 'category' should also be checked for the same reason. And indeed, as further analysis showed, ‘county’ and ‘post_town’ contained both only “Nottingham” as a value and ‘category’ contained solely the values “Residential”. As such, the ‘county’ and ‘category’ columns were dropped.

3.2. Visualisation

In terms of visualisation, the property data was first displayed according to the respective location using a Folium map (Figure 3). Certain clusters are already visible.

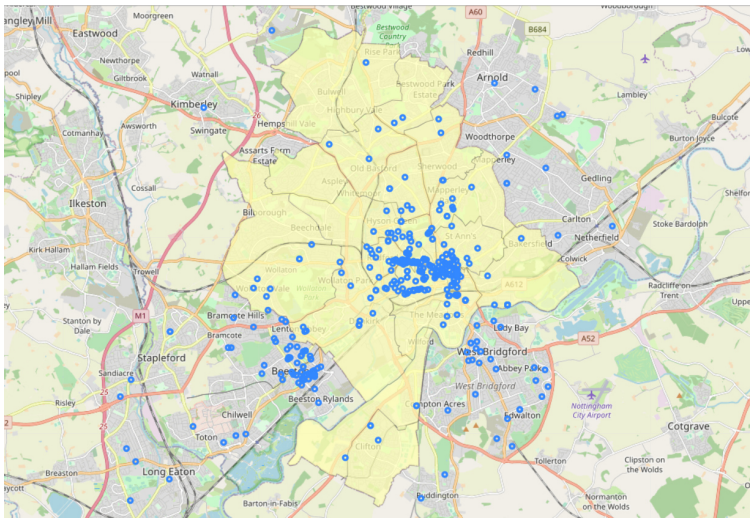

Figure 3. Visualisation of property data.

Next, data from Foursquare were obtained being of the following entertainment options: 'Creperie', 'Bookstore', 'Pub', 'Indian Restaurant', 'Bar', 'Coffee Shop', 'Restaurant', 'Gastropub', 'Candy Store', 'Plaza', 'Burger Joint', 'Indian Restaurant', 'Theater', 'Thai Restaurant', 'Pub', 'Deli / Bodega', 'Movie Theater', 'Coffee Shop', 'BBQ Joint'. When plotting both datasets in one map, the following map is created (Figure 4):

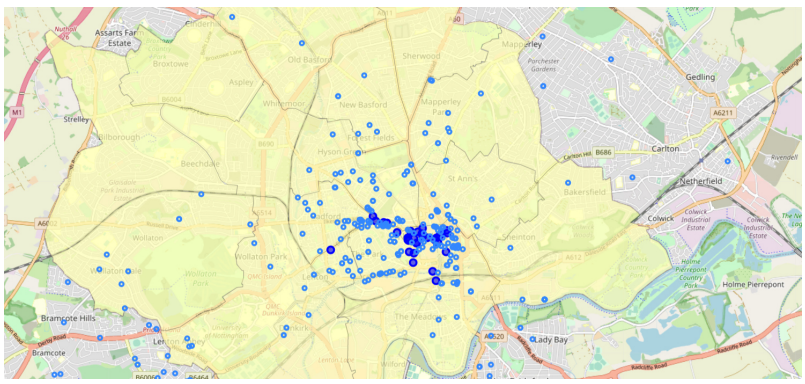

Figure 4. Zoopla and Foursquare data.

The above map (Figure 4) is to be treated with caution: the data retrieved from Foursquare are rather sparse. Nonetheless, we can still see some overlapping between the properties and the points of interest. If we now look for the postcode of this area, we find the following:

	num_bedrooms	num_bathrooms	price	property_type	street_name	post_town	outcode	latitude	longitude
93	2	1	387.0	Flat	THE PARK	Nottingham	NG7	52.947887	-1.158614

Figure 5. Postcode within area of interest.

Retrieving the other properties with the same postcode results in the following data set (Figure 6 ([excerpt]).

	num_bedrooms	num_bathrooms	price	property_type	street_name	post_town	outcode	latitude	longitude
9	8	2	920.0	Terraced house	WELLINGTON SQUARE LENTON	Nottingham	NG7	52.955970	-1.166062
10	8	2	880.0	Terraced house	WELLINGTON SQUARE	Nottingham	NG7	52.955960	-1.166002
14	7	2	770.0	Terraced house	WELLINGTON SQUARE LENTON	Nottingham	NG7	52.955944	-1.166211
16	6	2	750.0	Terraced house	SWENSON AVENUE	Nottingham	NG7	52.945435	-1.180416
17	7	2	735.0	Detached house	ALBERT SQUARE, CHURCH STREET	Nottingham	NG7	52.947502	-1.177880
18	7	2	700.0	Terraced house	BEESTON ROAD	Nottingham	NG7	52.938236	-1.185582
19	7	4	700.0	End terrace house	BEESTON ROAD	Nottingham	NG7	52.938190	-1.185657
20	1	1	695.0	Detached house	GRIMSTON ROAD	Nottingham	NG7	52.963600	-1.179093
24	6	2	660.0	End terrace house	KIMBOLTON AVENUE LENTON	Nottingham	NG7	52.955530	-1.171994
25	6	2	660.0	Semi-detached house	SEELY ROAD LENTON	Nottingham	NG7	52.955334	-1.167473

Figure 6. Properties within NG7.

4. Results

Based on initial insights from the previous section, we assume a link between the features 'number of bedrooms' / 'number of bathrooms' and 'price'.

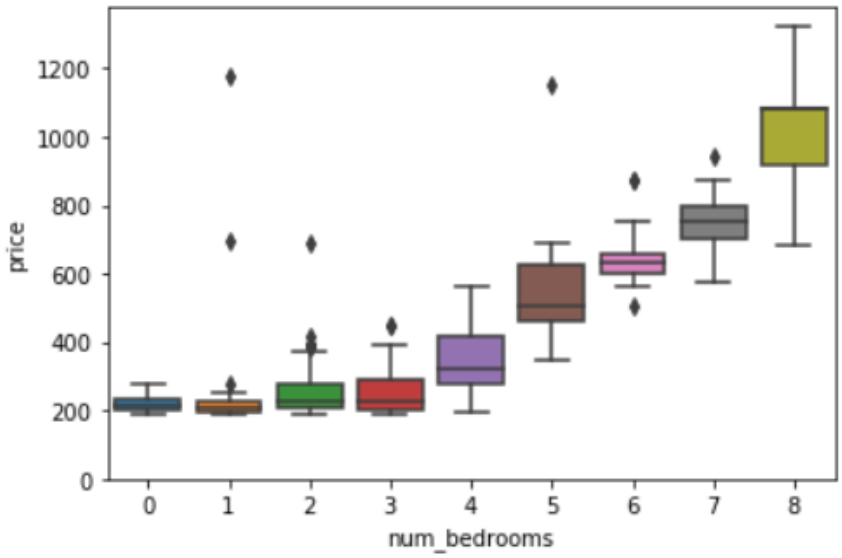

Figure 7. Number of bedrooms in relation to the price per month [GBP].

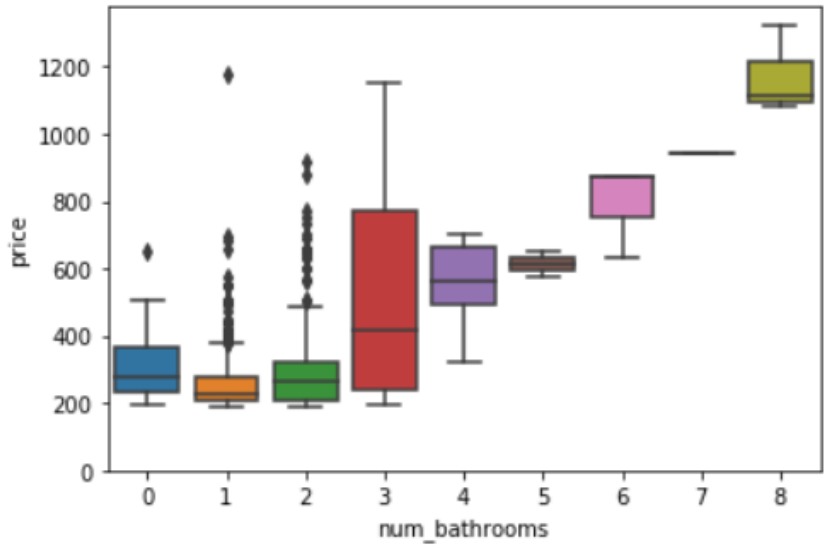

Figure 8. Number of bathrooms in relation to the price per month [GBP].

We can see a similar trend in Figure 8 as in Figure 7before, but it can also be seen that there is a wide spread in price for 1- and 2-bedrooms. One possible conclusion to draw from the box plots shown in Figures 7 and 8, is that (1) the price increases with the number of rooms, which is less surprising, and (2) that the rents for studios and 1-bedroom properties as well as for 2- and 3-bedrooms is relatively similar, respectively.

Looking specifically at the properties in NG7, we can obtain the following information (Figures 9 and 10):

num_bedrooms	num_bathrooms	price	latitude	longitude
0	0.923077	212.076923	52.948250	-1.160962
1	1.000000	291.666667	52.956083	-1.165701
2	1.215686	244.764706	52.956566	-1.166813
3	1.375000	275.625000	52.955976	-1.165223
4	1.666667	390.833333	52.951531	-1.171121
5	1.285714	496.000000	52.954720	-1.169403
6	2.000000	637.142857	52.953202	-1.173344
7	2.500000	726.250000	52.944968	-1.178832
8	2.000000	900.000000	52.955965	-1.166032

Figure 9. Statistical description of properties in NG7 (number of bedrooms).

property_type	num_bedrooms	num_bathrooms	price	latitude	longitude
Detached house	4.800000	1.400000	581.200000	52.951643	-1.174670
End terrace house	5.000000	2.250000	539.000000	52.950382	-1.174797
Flat	2.102041	1.224490	251.734694	52.956488	-1.165101
Maisonette	2.000000	1.000000	257.000000	52.949177	-1.165949
Semi-detached house	3.428571	1.571429	381.857143	52.950154	-1.166745
Studio	0.000000	0.923077	212.076923	52.948250	-1.160962
Terraced house	4.260870	1.434783	457.130435	52.956564	-1.173017
Town house	4.000000	4.000000	322.000000	52.947530	-1.155853

Figure 10. Statistical description of properties in NG7 (property type).

5. Discussion

A (late) disclaimer: this work is far from being perfect and completed. However, it allowed to gain some first insights into the properties for rent market in Nottingham, UK. While we managed to extract property data from Zoopla, which seemed very exhaustive and relatively complete, data from Foursquare is still a bit lacking. This is most likely due to the fact that Foursquare appears not to be well-known in Europe and people still tend to use Google as their preferred means of information gathering. The work can be enhanced by conducting a more thorough data analysis and apply some machine learning tools to the data.

6. Conclusion

This report presented the work that has been conducted for the Coursera course 'Applied Data Science Capstone' week 5. We looked at property data from Nottingham, UK and entertainments options with a particular focus on students. We covered some initial data pre-processing and analysis on data obtained through the Zoopla Developer API. The data were visualised using Python's Folium library and The Foursquare City Guide API was used to obtain some additional information for things to do in Nottingham especially when living in the postcode area around NG7.

References

[1] “The Complete University Guide,” [Online]. Available: https://www.thecompleteuniversityguide.co.uk/league-tables/rankings.

[2] “Wiki Voyage,” [Online]. Available: https://en.wikivoyage.org/wiki/East_Midlands.