Does proximity to amenities impact the rental price for a property in Aberdeen?

by Alexandra Ungureanu

I. Introduction

1. Background to the problem

I have been renting in Aberdeen, Scotland for nearly 9 years and one thing I could never fully comprehend was the rental price. I would often find two properties in close proximity to each other, and which to me seemed to be fairly similar, but with very different asking prices.

It got me thinking: 'What actually determines the rental price for a property in Aberdeen?'. Is it the location within the city? Maybe the square footage or Energy Performance Certificate band? What about its proximity to amenities such as shops, cafes and restaurants?

There is surprisingly less data publicly available on this matter than compared to that accessible to home buyers.

The purpose of this analysis is to determine whether proximity to amenities impacts the rental price for properties in Aberdeen at city, not council ward, level.

2. Interest

I have decided to tackle this issue as part of my Capstone Project for the IBM Data Science Professional Certificate. The coding and data will be available to developers for both feedback and their own work.

The conclusions of this mini-research are meant to give my fellow renters some insight into rental prices in Aberdeen and what most impacts them.

II. Data Acquisition and Cleaning

1. Data Sources

Below are the data sources used for analysis which have also been made available on my Github repository:

https://github.com/alex-spencer03/Coursera_Capstone.

a. Data of rental properties available for rent

This was scraped from the Aberdeen Solicitors Property Centre's (ASPC) website on the 15th July 2020: https://www.aspc.co.uk/ using the *requests* package.

The data was then loaded into a *pandas* dataframe using a loop function resulting in a 639x12 dataset which contained all properties available for rent in Aberdeenshire.

First_line	Second_line	City	Postcode	Bedrooms	Bathrooms	Lounges	Price	Square_ft	Property_type	Description	Coordinates
2 Albert Den	Aberdeen, AB25 1YX	Aberdeen	AB25 1YX	2	2	1	750.0	75.0	FLAT	Fully furnished 2 Bdrm 1st flr Flat. Hall. Lou...	POINT (-2.117279 57.1461223)
3c Anderson Drive	Aberdeen, AB15 4ST	Aberdeen	AB15 4ST	2	2	1	750.0	0.0	FLAT	Fully furnished 2 Bdrm Grnd flr Flat. Hall. Lo...	POINT (-2.13440599206544 57.1336311153834)

Following cleaning and handling the null values, a new pandas dataframe was then created from this containing only properties listed as being in Aberdeen.

Although the resulting dataframe contained 459 properties, following sanity checks 446 were moved forward into the analysis. This was due to 13 properties being listed in Aberdeen when they were in fact part of Aberdeenshire which is outside the scope of this analysis.

The final dataset, called '*aberdeen_master*', contains the data as described in the below table:

Column name	Description
First_line	First line of address of the property
Postcode	Postcode of the property
Council_ward	The area within the local authority - usually designated for election purposes
Bedrooms	Number of bedrooms
Bathrooms	Number of bathrooms
Lounges	Number of lounges (public rooms)
Price	Monthly asking price
Square_ft	Square footage of the property
Property_type	Flat, house, etc
Latitude	Latitude of the property
Longitude	Longitude of the property
Council_tax_band	Council tax band for 2020-2021
EPC_band	Energy Performance Certificate band
Parking	Access to parking

Please note this dataframe was persisted to DB2 Warehouse all throughout the data collection and cleaning process.

i. Council tax bands clarification

Council tax is a local tax paid by homeowners and renters and is used by the city councils to fund their services each year[1]. It is paid depending on the band assigned to the property by the council and runs from band A to H.

[1] Further details can be found here: https://www.aberdeencity.gov.uk/services/council-tax/how-your-council-tax-spent

The tax bands for the current tax year of 2020-2021 are in the below table and the annual amounts due can be seen under the 'Total charges to be paid' column header.

Council Tax and Scottish Water charges 2020-21

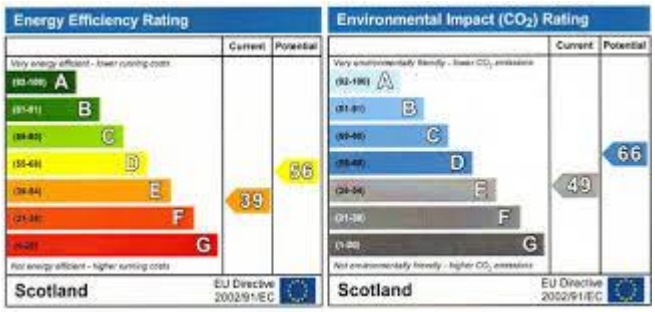
Council Tax band	Council Tax charge	Water supply	Waste water collection	Total charges to be paid
Band A	£918.20	£138.24	£160.50	£1216.94
Band B	£1071.23	£161.28	£187.25	£1419.76
Band C	£1224.27	£184.32	£214.00	£1622.59
Band D	£1377.30	£207.36	£240.75	£1825.41
Band E	£1809.62	£253.44	£294.25	£2357.31
Band F	£2238.11	£299.52	£347.75	£2885.38
Band G	£2697.21	£345.60	£401.25	£3444.06
Band H	£3374.39	£414.72	£481.50	£4270.61

Why is this important?

This is an additional monthly expense for those who rent and will impact their monthly outgoings. Generally, unless you're making bank, you will be looking to rent a property between bands A and C.

ii. EPC band clarification

EPCs, Energy Performance Certificates, are all about the energy efficiency of a property and its impact on the environment. They are mandatory by law in Scotland[2].



Why is this important?

Similar to the council tax bands they run from A to G, with the bands A to C being considered energy efficient, and they will impact a renter's monthly outgoings. The less energy efficient a property is (for example, old electricals) the higher the running costs and the higher the energy bill to be paid by the renter.

iii. Square footage data

As will be further discussed in the 'Data Cleaning' section, 235 out of 446 properties where missing values from the data extracted from ASPC's website and had to be manually retrieved. As this was going to be a very time consuming exercise I had to weigh the value this information would add to the analysis.

Further details of the exploratory analysis are shown and discussed in the below 'Handling null values' section.

[2] Further details can be found here: https://www.epcregister.com/

b. Amenities data from Foursquare

A total of 100 venues were pulled from Foursquare's API on the 6th of August 2020 which were stored in a pandas dataframe named 'ab_venues'. This dataframe is very simple and only contains the name, postcode, latitude and longitude of the amenities in Aberdeen.

2. Data Cleaning and handling null values

a. Data Cleaning

As mentioned above, the ASPC properties data pulled looked like below:

First_line	Second_line	City	Postcode	Bedrooms	Bathrooms	Lounges	Price	Square_ft	Property_type	Description	Coordinates
2 Albert Den	Aberdeen, AB25 1YX	Aberdeen	AB25 1YX	2	2	1	750.0	75.0	FLAT	Fully furnished 2 Bdrm 1st flr Flat. Hall. Lou...	POINT (-2.117279 57.1461223)
3c Anderson Drive	Aberdeen, AB15 4ST	Aberdeen	AB15 4ST	2	2	1	750.0	0.0	FLAT	Fully furnished 2 Bdrm Grnd flr Flat. Hall. Lo...	POINT (-2.13440599206544 57.1336311153834)

The column 'Description' contained the council tax, EPC, garden and parking information, and the 'Coordinates' column contained the latitude and longitude of the properties.

The council tax, EPC, garden and parking data was extracted using the *re* package and columns created for each respectively.

The latitude and longitude data was extracted using the *.strip()* method and columns were created for each respectively.

Following this the 'Description', 'Second-line' and 'Coordinates' columns were dropped as they were no longer required.

The council tax band strings pulled from ASPC did not follow a consistent format and required further cleaning. This was done using the *.replace()* method and the below 'Before and after' table shows the not only the output but also why cleaning was required.

Before cleaning	After cleaning
CT band - A 87 CT band - D 66 CT band - B 63 CT band - TBC 57 CT Band - D 49 CT Band - C 47 CT band - C 47 CT Band - B 43 CT band - E 36 CT Band - A 34 CT Band - E 24 CT band - F 21 CT Band - TBC 20 CT Band - F 12 CT band - G 8 CT Band - G 5 CT Band - H 3 CT - band TBC 2 CT band - H 2 CT band A 2 CT band -TBC 2 CT band - TBC 1 CT Band - TBC 1 CT band B 1 CT band -E 1 CT band -F 1	CT Band A 123 CT Band D 115 CT Band B 107 CT Band C 94 CT Band TBC 83 CT Band E 61 CT Band F 34 CT Band G 13 CT Band H 5

b. Handling null values

I firstly checked the entire dataset for any null values using the *.isnull()* method. For the 'Square_ft' data I used a *.loc()* method to extract all rows where the values is 0.

From this I could determine that the number of properties with:

- no council tax entry: 3;
- no EPC entry: 1;
- no square footage entry: 235;
- council tax entry as TBC (To Be Confirmed): 74.

Council tax values:

There were 77 properties with no council tax band data. This information can be found on Aberdeen City Council's website:

https://ecitizen.aberdeencity.gov.uk/publicaccesslive/selfservice/services/counciltax/bandsearch.htm

Square footage values:

There are 235 properties with no square footage data. The only way I have been able to locate this information is to extract it from the individual Energy Performance Certificates which can be accessed on the Scottish EPC Register:

https://www.scottishepcregister.org.uk/CustomerFacingPortal/EPCPostcodeSearch

Unfortunately, the Terms and Conditions of this website explicitly prohibits the use of a web scraper to extract information from it which means it would have to be done manually.

As this would be a very time-consuming process I had to weigh what value having this data would add to the overall analysis.

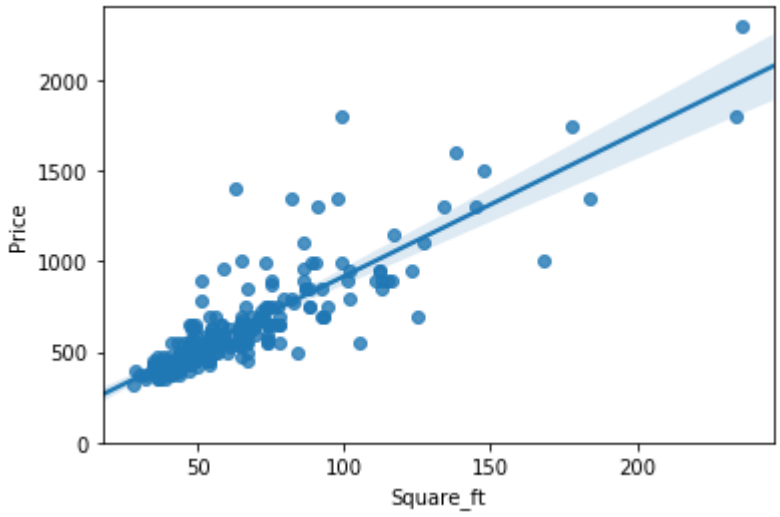
Checking correlation between square footage and price

To determine this I created a separate dataframe containing only the properties that have non-0 values in the 'Square_ft' column – 211 properties.

I then used pandas *.corr()* method between 'Price' and 'Square_ft' which showed a 0.854635 correlation. To put it into perspective, a value of 1 means they are identical.

	Price	Square_ft
Price	1.000000	0.854635
Square_ft	0.854635	1.000000

To visualise the linear relationship I created a scatter plot between the two features using the *matplotlit* and *seaborn* packages. The below graph shows how the vast majority of properties are closely grouped together along an upward line. This indicates a strong relationship between the two and means that not including the square footage will skew the final results of this analysis.



At this stage I downloaded the dataset as a .csv file and performed the additional manual work including adding the council wards for each property. This was done by matching the property postcodes in the dataset with the ones in the council wards data.

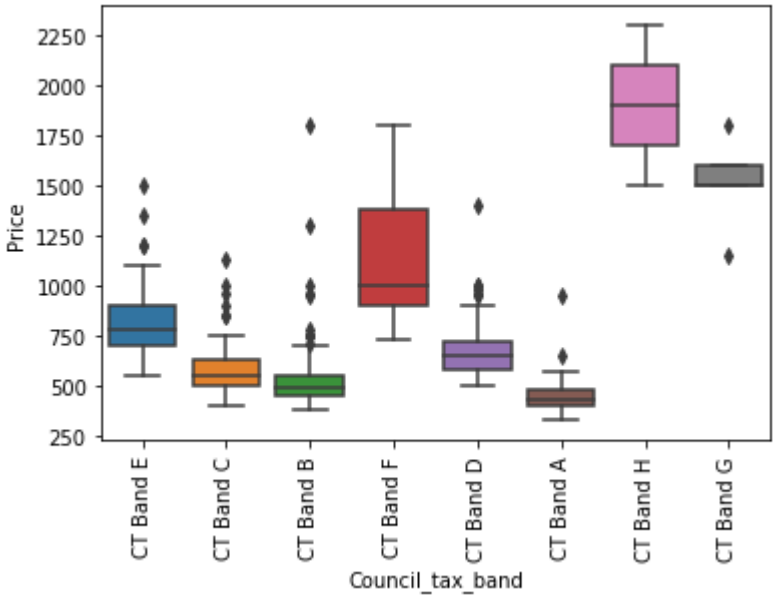
III. Exploratory Data Analysis

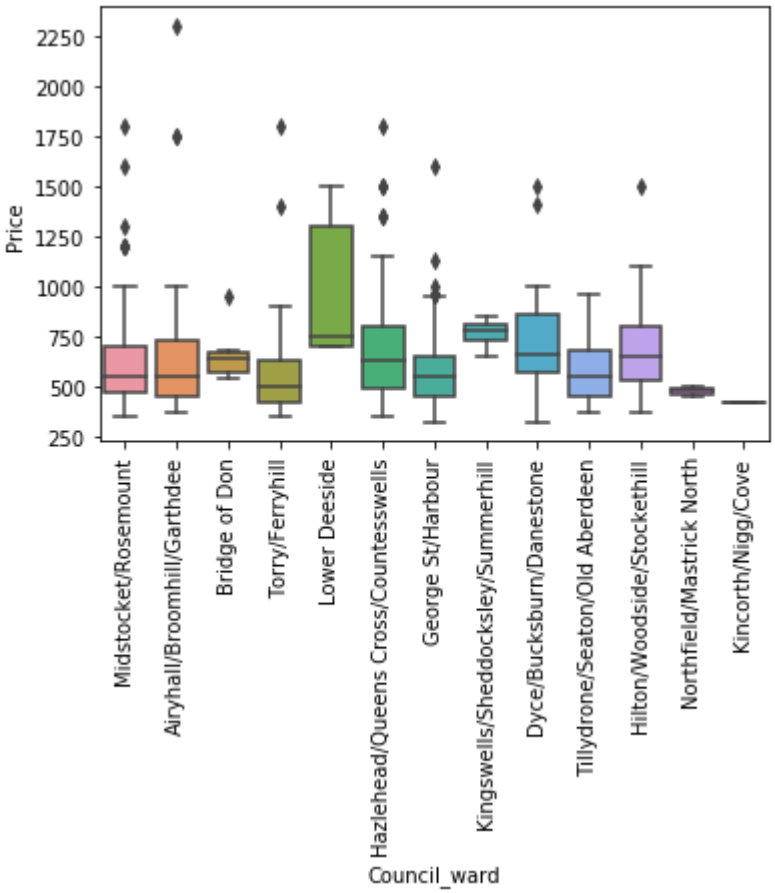
The first step was to use the .describe() method to do a quick exploration of the properties dataset. As shown in the below image, the average rental property in Aberdeen has close to two bedrooms, one bathroom and one lounge with approximately 66 square meters of space and goes for £640 a month.

	Bedrooms	Bathrooms	Lounges	Price	Square_ft	Longitude	Latitude
count	446.000000	446.000000	446.000000	446.000000	446.000000	446.000000	446.000000
mean	1.894619	1.215247	1.065022	639.094170	65.964126	-2.117851	57.151481
std	0.834707	0.443014	0.325384	268.684001	32.615962	0.027320	0.016319
min	1.000000	1.000000	0.000000	325.000000	26.000000	-2.263654	57.096120
25%	1.000000	1.000000	1.000000	475.000000	47.000000	-2.126179	57.140279
50%	2.000000	1.000000	1.000000	550.000000	58.000000	-2.112188	57.148793
75%	2.000000	1.000000	1.000000	700.000000	75.000000	-2.100270	57.158662
max	6.000000	3.000000	4.000000	2300.000000	330.000000	-2.078191	57.212590

The next step was to calculate and visualise the correlation between the numerical variables, categorical variables and price. As can be seen below each feature has a relatively, if not strong, correlation to the price on their own.

	Price	Bedrooms	Bathrooms	Lounges	Square_ft
Price	1.000000	0.791809	0.647835	0.530821	0.785406
Bedrooms	0.791809	1.000000	0.511176	0.480349	0.728791
Bathrooms	0.647835	0.511176	1.000000	0.308011	0.601473
Lounges	0.530821	0.480349	0.308011	1.000000	0.616610
Square_ft	0.785406	0.728791	0.601473	0.616610	1.000000





Additional exploratory analysis was performed by looking at how combinations of property features, including its location in a specific council ward, would affect the price. I felt the dataset was too small to obtain an in-depth, clear output.

Following this the properties where clustered using KMeans method from sklearn package into 3 clusters. The first cluster contains 332 properties with an average of 1 bedroom, 1 bathroom and 1 lounge, 52 square meters and with a monthly asking price of approximately £505. In terms of features, the properties go from 1-3 bedrooms, 1-2 bathrooms, 0-2 lounges and 26 to 121 square meters within a price bracket of £325-£650.

	Cluster Labels	Bedrooms	Bathrooms	Lounges	Square_ft	Price	Latitude	Longitude
count	332.0	332.000000	332.000000	332.00000	332.000000	332.000000	332.000000	332.000000
mean	0.0	1.521084	1.030120	1.00000	52.448795	504.765060	57.151543	-2.113819
std	0.0	0.518108	0.171177	0.10993	13.372892	83.356144	0.015918	0.022576
min	0.0	1.000000	1.000000	0.00000	26.000000	325.000000	57.096701	-2.190493
25%	0.0	1.000000	1.000000	1.00000	43.000000	445.000000	57.140361	-2.119906
50%	0.0	2.000000	1.000000	1.00000	50.000000	500.000000	57.149014	-2.110062
75%	0.0	2.000000	1.000000	1.00000	60.000000	571.250000	57.157350	-2.099129
max	0.0	3.000000	2.000000	2.00000	121.000000	650.000000	57.212590	-2.078191

The second cluster was quite interesting in comparison to the first one. It contains 129 properties averaging 2 and half bedrooms, 1 and half bathrooms, 1 lounge and 88 square meters at approximately £831 per month.

When I looked at the features themselves they go from 1-4 bedrooms, 1-3 bathrooms, 0-2 lounges, 34-200 square meters within a price bracket of £675-£1200. The reason this is interesting is because when compared

to the first cluster's minimums which start at the same 1 bed, 1 bath and no lounge, on the face of it, the only difference is the square footage which is an extra 8 square metres of space for £350 more per month!

	Cluster Labels	Bedrooms	Bathrooms	Lounges	Square_ft	Price	Latitude	Longitude
count	129.0	129.000000	129.000000	129.000000	129.000000	129.000000	129.000000	129.000000
mean	1.0	2.488372	1.550388	1.116279	88.620155	831.093023	57.155683	-2.129136
std	0.0	0.674494	0.529759	0.367166	29.516996	131.538103	0.019341	0.036295
min	1.0	1.000000	1.000000	0.000000	34.000000	675.000000	57.096120	-2.263654
25%	1.0	2.000000	1.000000	1.000000	72.000000	725.000000	57.142875	-2.138420
50%	1.0	2.000000	2.000000	1.000000	83.000000	795.000000	57.150886	-2.123921
75%	1.0	3.000000	2.000000	1.000000	97.000000	900.000000	57.165151	-2.103305
max	1.0	4.000000	3.000000	2.000000	200.000000	1200.000000	57.212590	-2.081247

The same interesting observations were noted when comparing the third cluster with the second. The third cluster contains 19 properties averaging close to 4 bedrooms, 2 bathrooms, 1 and half lounges and 154 square meters for approximately £1580 a month.

When comparing its minimum of 2 beds, 1 bath and 1 lounge to its equivalent in the second cluster you're actually getting less square footage, 9 square meters less, for £575 a month more.

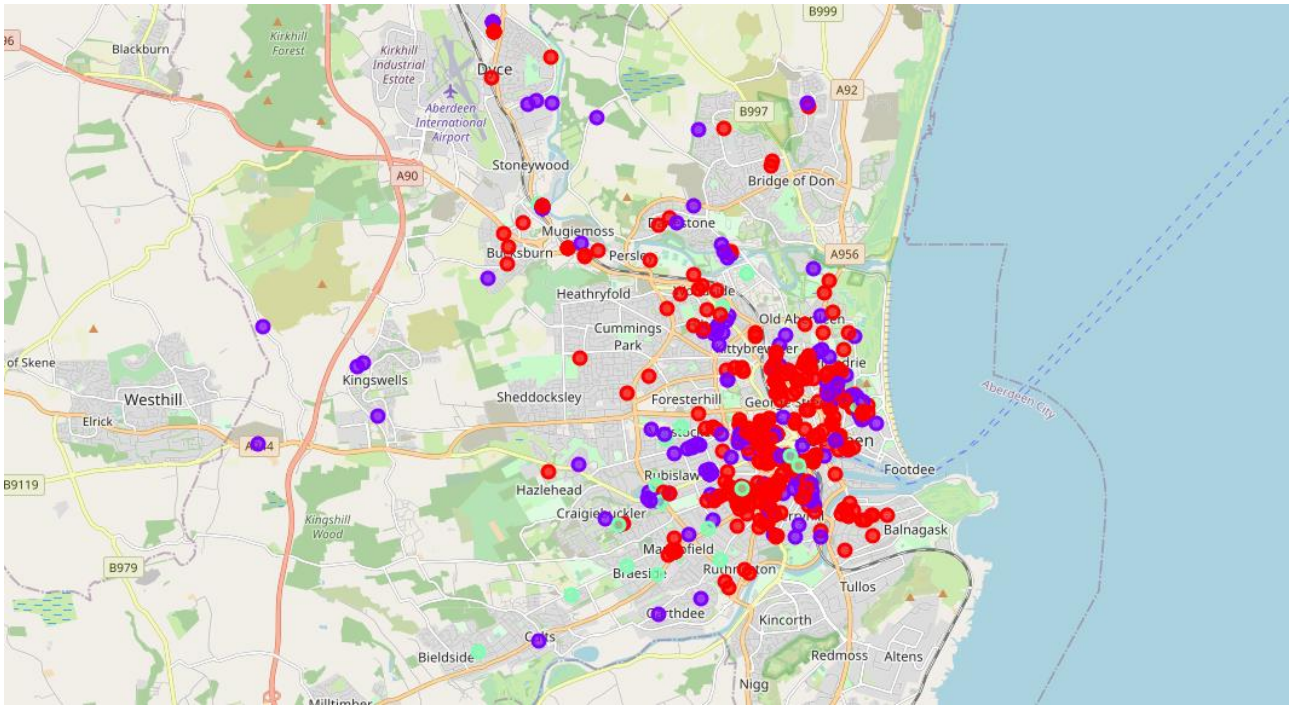
	Cluster Labels	Bedrooms	Bathrooms	Lounges	Square_ft	Price	Latitude	Longitude
count	19.0	19.000000	19.000000	19.000000	19.000000	19.000000	19.000000	19.000000
mean	2.0	3.789474	2.105263	1.736842	153.947368	1579.210526	57.143674	-2.135142
std	0.0	1.031662	0.567131	0.933459	69.193347	244.927785	0.017350	0.027098
min	2.0	2.000000	1.000000	1.000000	63.000000	1300.000000	57.115645	-2.194117
25%	2.0	3.000000	2.000000	1.000000	100.000000	1402.500000	57.132068	-2.147878
50%	2.0	4.000000	2.000000	1.000000	148.000000	1500.000000	57.141288	-2.136580
75%	2.0	4.500000	2.000000	2.000000	166.500000	1750.000000	57.149294	-2.118591
max	2.0	6.000000	3.000000	4.000000	330.000000	2300.000000	57.184614	-2.086557

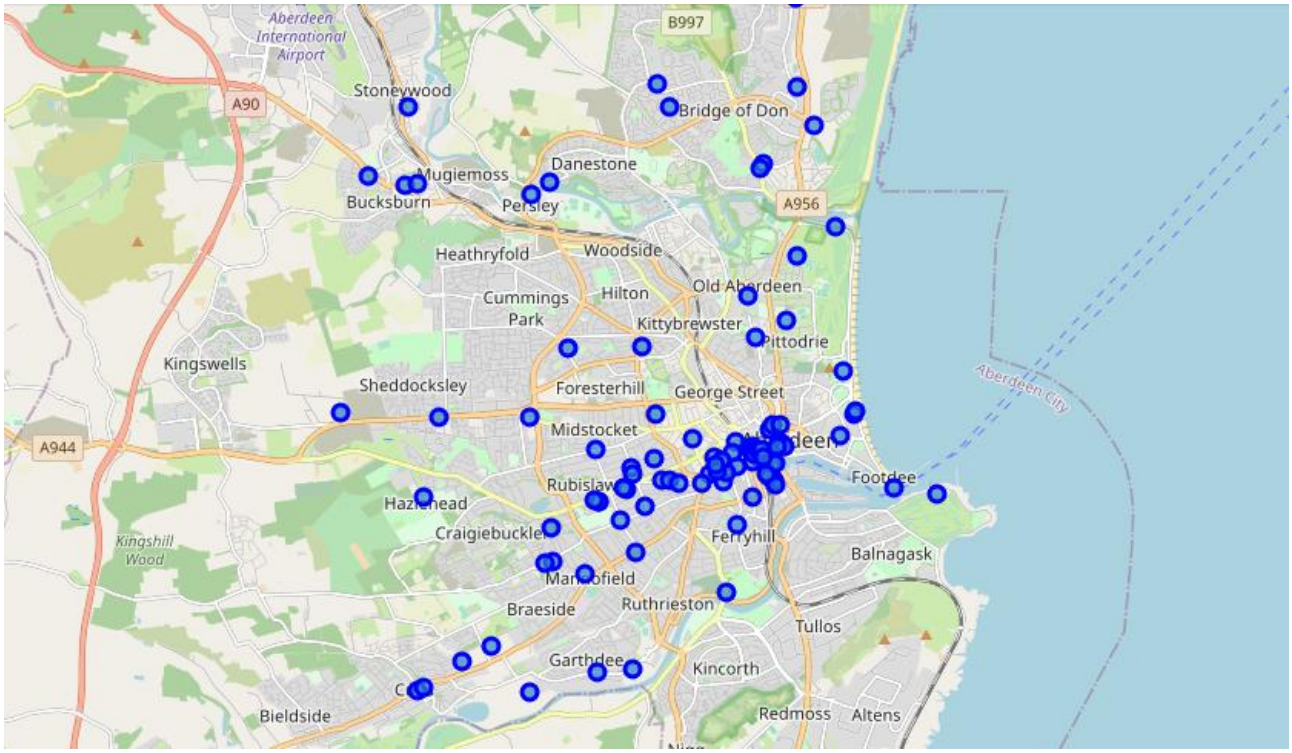
The above observations are very interesting and required additional insight. To do this I created a map of the clusters using the folium package. The coordinates for Aberdeen were obtained using the geopy library.

The below first map is a visualisation of the clusters and the second image is a map of the amenities pulled from Foursquare.

As can be seen, the third cluster (light green) properties tend to be spread towards the South – West and North. For anyone who has lived in Aberdeen this makes sense as those locations are considered to be very desirable residential areas.

When looking at the concentration of amenities and their spread across Aberdeen it can be noticed that the vast majority are located in the city centre. At the same time the first map shows a concentration of lower value properties (in red) in the same area. The higher value properties (third cluster – light green) tend to move away from the amenities.





IV. Conclusion

Based on the above analysis and maps it can be concluded that proximity to amenities does not have weight in determining the rental price in Aberdeen. If anything, it seems to negatively impact it as properties within the concentrated amenities area (the city centre) are on the lower value end of the market.