UK Housing price variation analysis & Advice for investors to raise the value of their capital

DATA 7001 Group 3

Team members:

Weiquan Zeng (s4406382)
Yan Liao (s4738406)
Feng Jiaxun (s4694975)
Yangbo Liao (s4690577)
Taisheng Song (s4713214)

The University of Queensland

We DO NOT give consent for this to be used as a teaching resource

Executive summary

In this project, we explore what factors will affect the changes in house prices by analysing different factors that may have an impact on UK house prices. Among many factors, we have selected several factors, including population growth, locals' educational level, birth and death rate of enterprises, locals' income, the quantity of affordable housing and quality of life to analyse their impact on housing prices.

After sorting and analysing the data, we found that income and population will have a greater impact on changes in housing prices, especially income. In addition, other factors such as local people's educational background are not enough to establish a relationship. Therefore, to understand housing price changes in a city, one needs to focus on local demographic changes, such as birth and death rates, population inflows and outflows; and the local income levels, like minimum wage and median income.

Table of content

Problem solving with data (design thinking)

Background

Nowadays, an increasing number of people are becoming more concerned with housing issues. The house is one of the most important properties that people are concerned. People need a house to provide them security, control, belonging, identity and private space to satisfy their psychological needs (Moeller, 2022). Furthermore, the house property is usually contributed the most expenditure proportion in people's life. Thus, the potential house buyers should have a basic understanding of the potential variables that may affect the house price. According to Nguyen (2021) stated that there are four main key factors that have the most significant impact on the house price, including demographic, economy, interest rate and policy. A good decision making on the house prices could significantly save/increase the house owner's wealth.

In this project, we will primarily take the UK as an example to analyse what factors are mainly affected by today's housing prices, so as to help common house buyers or real estate investors make better choices that meet their needs.

Problems investigation

The house property usually contributes the biggest proportion of people's expenditure in their life. In the meantime, the house property price will directly affect the house owner's equity (Foote, 2016). For example, people intend to purchase assets when they drop to the lowest price, however, if the house price decreased dramatically in a short period after people purchased. The house owner may not be able to pay the fees since the property is not worth this price anymore and the buyer can also suffer the financial punishment because of violating the contract. Therefore, it is critical to have an understanding of the potential factors that may affect the house price before the investor/buyer purchases the house.

The purpose of the investigation is to find out the potential variables that have a strong relationship with the UK house price and investigate the outliner (special areas that are not in the normal range have dramatically increased) and provide a reference for the stakeholders.

Stakeholders

House buyers and House Investors

One of the major stakeholders of this project is the potential house buyer and investor who would like to purchase house property in the United Kingdom. The buyer/investor can refer to this project to obtain an understanding of the potential variables that may have a strong correlation with UK house prices. This project would also provide a logical methodology based on the data science process and recommendations based on the results for the potential UK house buyers/investors to improve their decision-making ability before purchasing the house.

Good decision making may save the buyer's costs or increase their wealth or even both.

Project process

The team will conduct this project based on the data science process (figure 1). The process is not a one-way process, it is allowed the data scientist to iterate/go back one of the steps during this project.

Figure 1: Data science process

Getting the data our need

Primary data source

The master data for the project comes from the UK House Price Index (UK HPI), which is an official statistic that reflects changes in the value of UK residential properties.

UK HPI is calculated by the Office for National Statistics in Northern Ireland and the Office of Land and Property Services. Data for the UK House Price Index is provided by the HM Land Registry, the Registers of Scotland, Land & Property Services Northern Ireland and the Office of Valuation.

We intercept the data set according to the needs, and combine with other relevant data sets required to analyse the influence of various factors on housing prices. Categories in the dataset include but are not limited to the following:

Categories	Explanation
Date	The year and month to which the monthly statistics apply
RegionName	Name of geography (Country, Regional, County/Unitary/District Authority and London Borough)
AreaCode	Code of geography (Country, Regional, County/Unitary/District Authority and London Borough)
Average Price	Average house price for geography in a particular period
12m%change	The percentage change in the Average Price compared to the same period twelve months earlier.
Sales Volume	Number of registered transactions for geography in a particular period

Table 1: UK House Price Index (UK HPI) dataset

This data comes from government websites with high data reliability, which is the main reason why we choose it as the primary dataset. Besides, the data recorded in this data has a large time span, and its categories are also detailed enough for us to analyse. What's more, the data has been collated by professionals, so there is no need for overly complex data patching or cleaning.

Related data sources

The selection of related dataset is based on the direction (in problem formulation) of the analysis factors as well as can be connected with the primary dataset and has analytical value. Whenever possible, we choose datasets from governments, educational institutions or large organisations to ensure the credibility of the data source.

Below are the related datasets based on the analysis direction:

UK population density in each local authority - (Demographics)

People's education qualifications level in each UK local authority - (Demographics)

UK people's earnings - (Economy)

Enterprise birth/death rate in each UK local authority - (Economy)

Affordable housing supply in each UK local authority - (Policy/Economy/demographics)

Making the data fit for use

Data integration

Aside from the main dataset which contains the change of housing price, several other datasets are used to study the relationship between the housing price and some potential independent variables. The area code and time are used to limit the range of data for further comparison and prediction. The figure displays how the datasets are linked to each other.

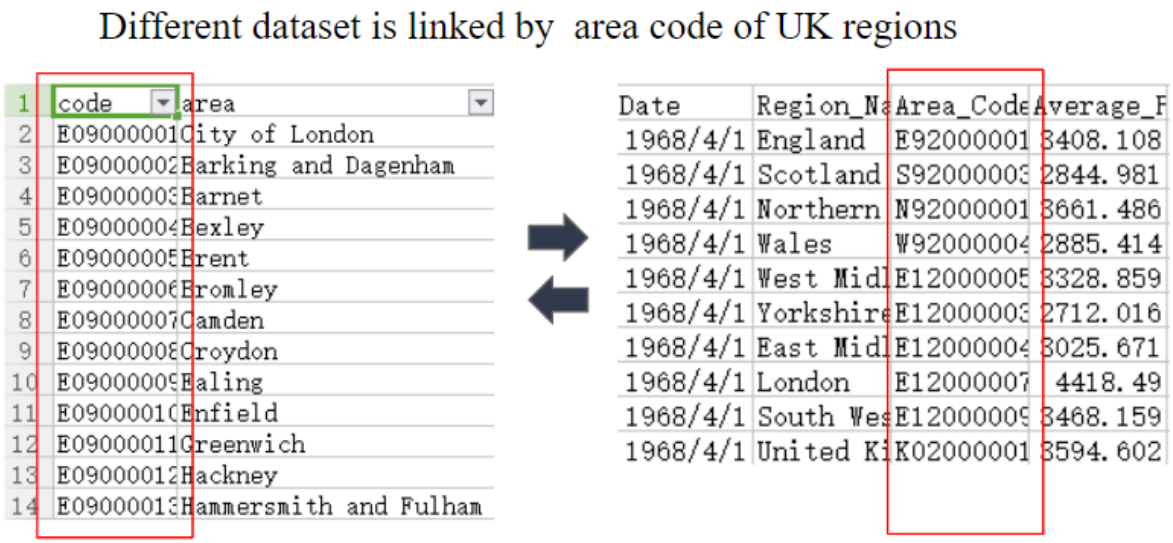

Figure 2: link main dataset with new other datasets

Data quality issues & Data cleaning method

Missing data

Several methods are used to deal with the missing data in all of the datasets used in this project. These datasets contain large quantities of data with a time span of over five decades, and there is massive missing data from the early years. For example, the annual housing price change data in the 1960s and 1970s is missing, while the data from 2000 to 2021 is more specific. To solve this problem we redetermine the time range of the data we are going to use and discard unnecessary ancient data.

We also try to replenish some important but missing data by making a prediction using the data we already have but we find it did not work well because the prediction curve is not accurate enough and this method will also have a negative influence on the validity of the data.

Distinct data range

Not all areas of England have published all the data we need about house prices and other variables we are going to study, so we choose to continue the research on a case-by-case basis

- using the completed and specific data of several typical and representative regions and cities to find the relationship between variables we are interested in. Besides, when studying the relationship between the main dependent variable and potential independent variables, different time and areas are chosen to specify the relationship and find whether the relationship is statistically significant.

Rows with incomplete data are deleted and columns with a high proportion of missing data will not be studied in the project. For example, we choose to study annual house price change instead of monthly house price change because less than 50 percent of the cities have data of monthly price change and for a short time period.

Input error & Outliers

Data with obvious type errors were corrected manually. Few data which contain extremely weird values of variables are deleted as outliers and were not used in the study.

Some of the cities, for example, the city of London, may have a very significant house price fluctuation in a certain period of time and we lack a method to find out whether it is true or is just a data fluctuation exception. We decide to ignore this in some situations to focus on finding the general trend of house price change and potential influences after discussion.

Inconsistent data format

Some datasets, by their nature, maybe quarterly or annual and non-continuous. In order to meet analytical needs or make datasets usable, we need to merge multiple related but fragmented datasets into a new dataset. The method we use is to expand additional categories. For example, Deutsche Bank's statistics are presented on an annual basis. In order to get data for a certain time period (such as 2015 to 2020) we need to combine 6 independent data sets for these 6 years. However, if it is directly spliced, it will make the data of different years lose their uniqueness and lead to the loss of availability of the data.

To this end, we added a new category "year" to the annual dataset to differentiate and concatenate the data of all different years. In addition, we use "Inner Join" when splicing, so as to avoid the possibility of blank data.

Making the data confess and Storytelling with Data

Data exploration

The United Kingdom house price trend

The data exploration provided an overview of the United Kingdom house price trend from 1968 to 2021. Because the UK house price has been recorded in each local authority, thus, to view the UK overall house price the team has decided to calculate the average house price yearly for the United Kingdom to view the trend.

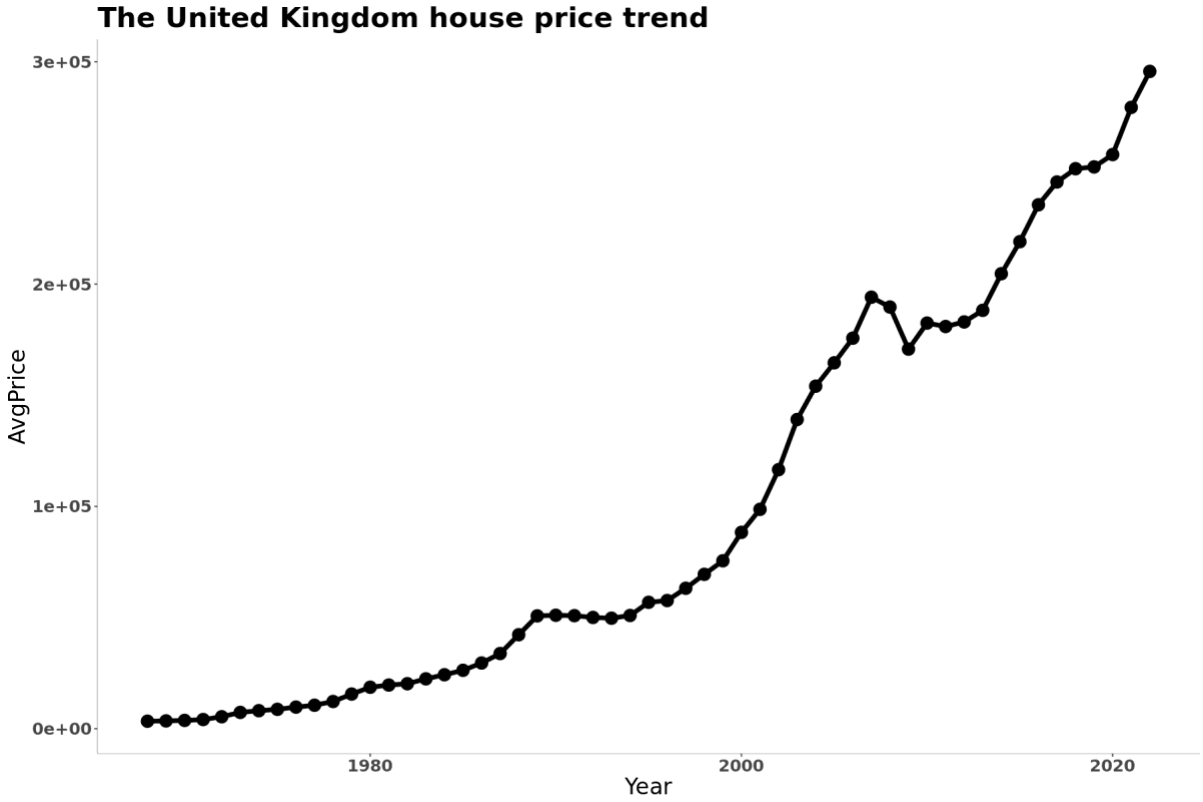

Figure 3: United Kingdom yearly average house price from 1968-2021

As we can see the UK house price is increasing most of the time, especially after 2000 the house price increased faster than before. This trend gives us the insight to analyse the possible factors that lead to the UK house price increase. Furthermore, the house price decreased between 2007 and 2008 which is the only period that the UK house price decreased. The reason behind this may be caused by the global financial crisis from 2007 to 2008 that led to the people's property depreciating significantly ("Financial crisis of 2007-2008 - Wikipedia", 2022). Since the UK house price is growing most of the time except for the financial crisis period, and the time is in the short period and affected globally, thus, we do not consider this outliner in our analysis.

Drill down the analysis range by data exploration

After we have the overview of the entire UK house price change from 1968 to 2021 since the UK housing index dataset contains 135,251 records since the data recorded over 300 local authorities' monthly house prices. Since the UK house price is increasing most of the time, to evaluate the potential reasons that lead to the house price increase, it will be better to narrow down the analysis range from the entire country to the particular local authorities that the house price increased dramatically than others to see the correlation behind. To do this, we have calculated the **Average Difference** for each local authority's house price from 1968 to 2021. And we also constructed an interactive web application to provide a clear view of the UK local authority house price changes.

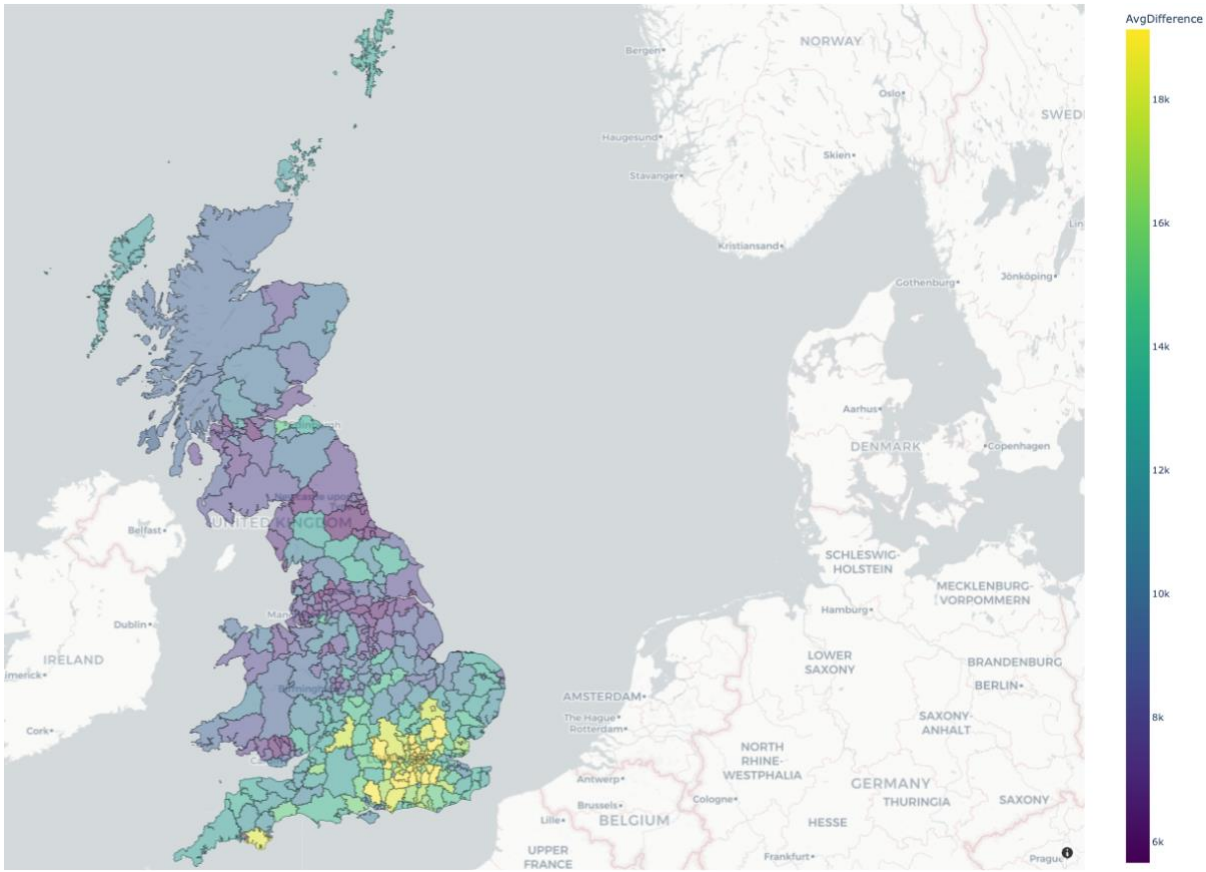

Figure 4: The United Kingdom local authorities' house price changed from 1968 to 2021

As we can see the UK house price increase is uneven. Most local authorities in the south increase significantly than in the north. To view more detail of the house price change for each local authority, we can zoom in on the map.

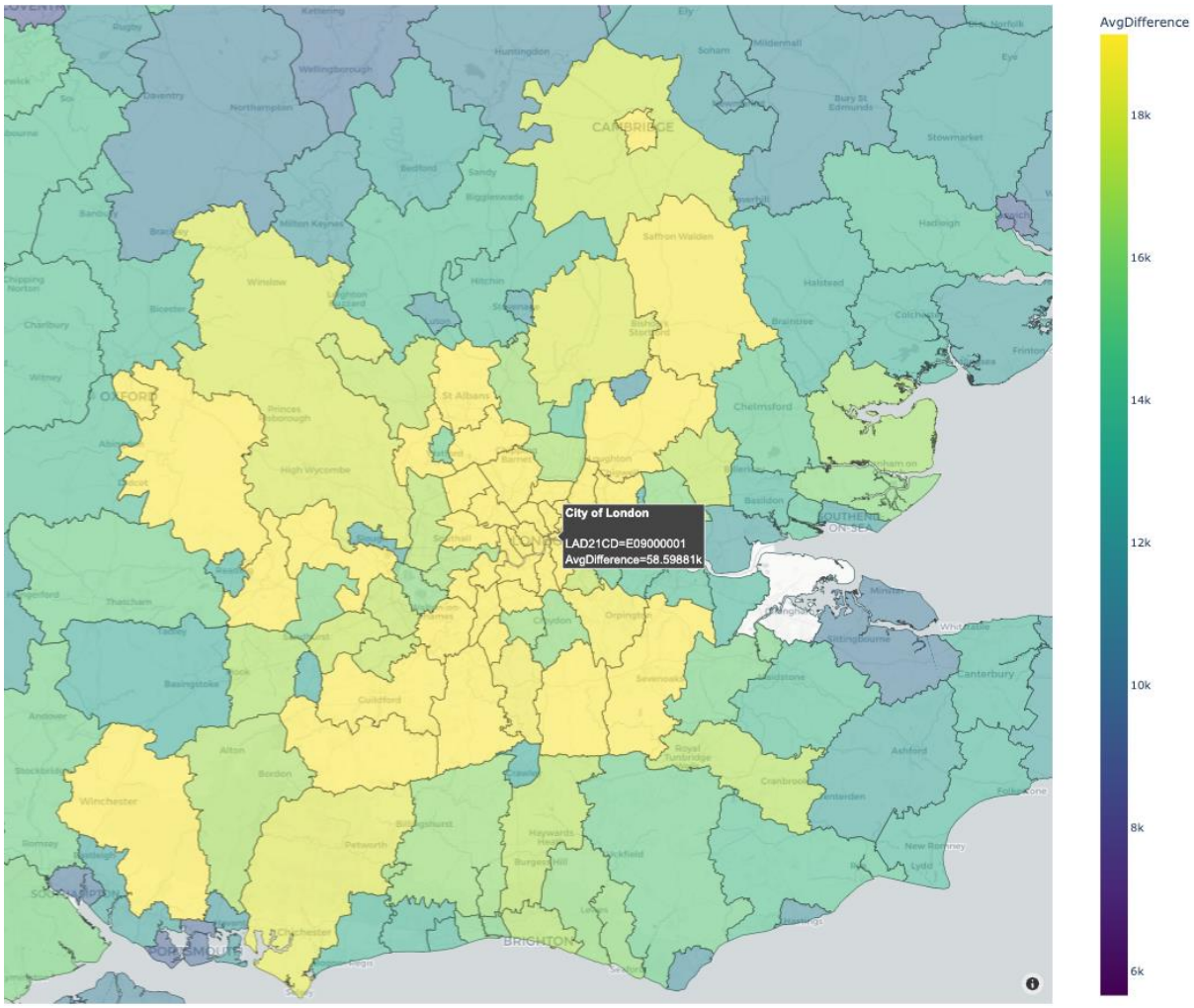

Figure 5: The UK regions that have the most housing price increase

According to the figure 5, we can see most cities/regions that have dramatical house price increases are around the city of London. Because those areas have much increased on the house price compared with other regions. Thus, we have decided to analyse the potential variables that may affect the house price in those areas that have house prices increased significantly to make the data analysis more logically.

Population growth

Population is one of the most important elements that affect the house price since the house is sold when people intend to live in a specific area. According to M. David (2012) stated that if the population density trend is increasing, the increasing people's earnings and population will lead to house price increases and diminishing rises in the stock of housing. Therefore, when an area's population density is increasing, the demand for housing is also increasing at the same time. The increasing demand for houses stimulates the house price increase because of decreasing stock of housing.

To analyse the correlation between population and house price in the UK, we would like to see the overall trend of each local authority population trend in the last 10 years.

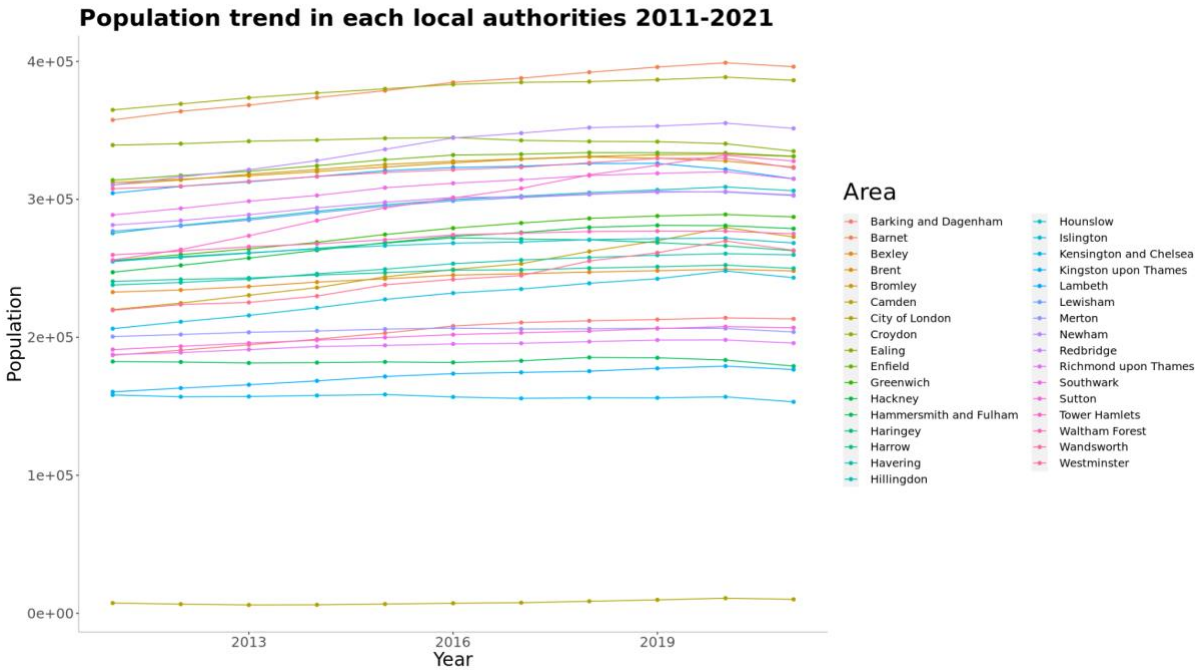

Figure 6: UK local authority regions population trends from 2011 - 2021

As we can see, most local authority regions' populations have been slowly increasing in the last decade. There are only a few areas where population densities have decreased in the last decade. We also can see the city of London has the lowest population, and the city of London has a significant increase in the house price. Based on the data exploration of figure 5, we can see the city of London's geographical area is very small, so it may be the main reason that the city of London has a low population density. According to figure 6, it looks like the population has similarities to the UK house price trend (also increasing) in the last decade. Then we decided to combine the area's population and its average house price in the last decade and calculate each region's average population and house price's growth rate to see whether the population growth and the price are also growing. To calculate the average growth we decided to use second-year population/price minus first-year population to get the growth rate. And to calculate the average growth rate, we add all the growth rates together and divide them by the number of years that have been calculated.

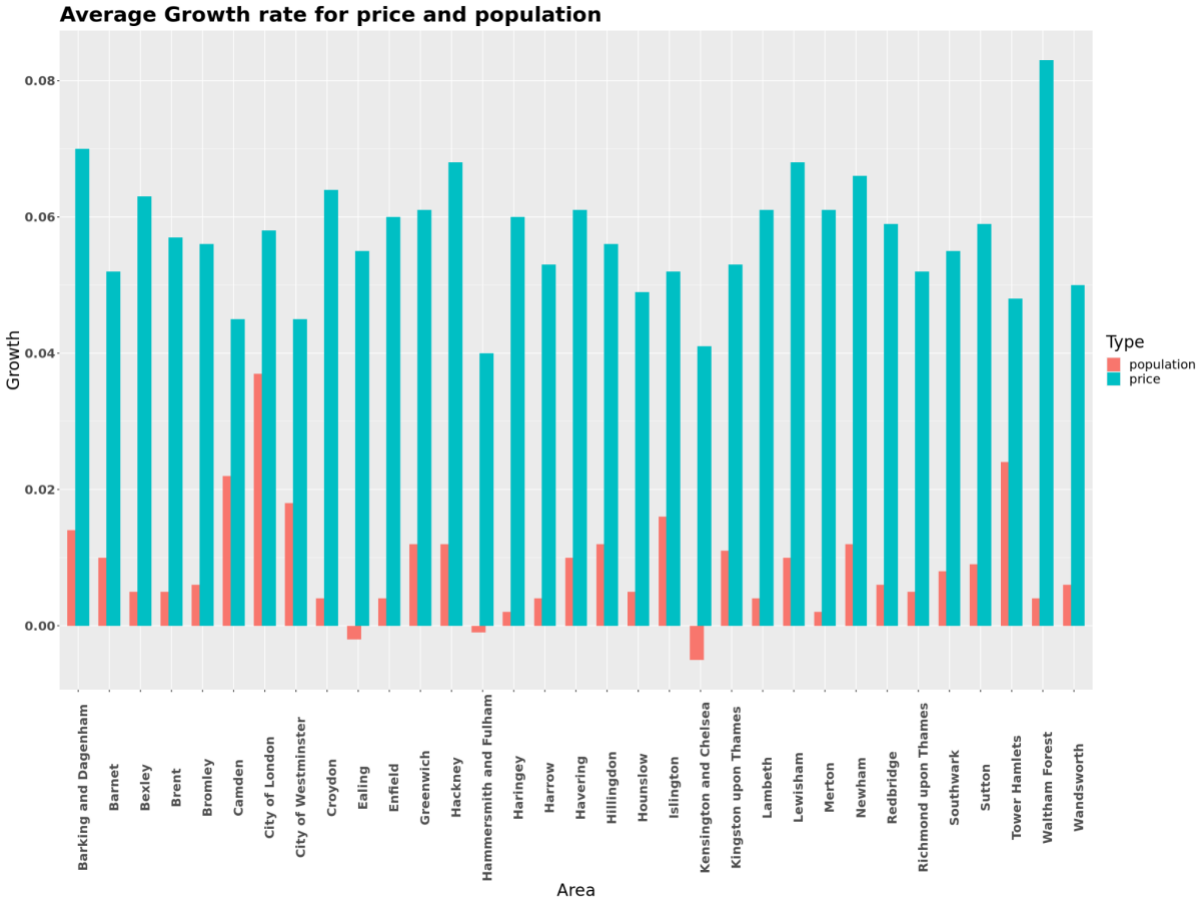

Figure 7: UK local authorities' average population and house price growth rate

According to figure 7, we are unable to see the clear correlation between population and housing price that we desire. The average growth plot demonstrates the average housing price in the UK is all positive which means the housing price is mostly increasing in the last 10 years. On the other hand, the population average growth rate does not indicate the same growth as the price. There are 3 situations that we can observe in this plot. Firstly, we can see the city of London's population and price growth have both increased significantly in the last 10 years. However, Kensington and Chelsea had negative population growth in the last 10 years but its house price keeps increasing. Finally, we are able to see most local authorities' population grew much lower than the house price in the UK. Therefore, the average growth rate for the UK housing price and population in the last 10 years plot did not provide a clear correlation between price and population, thus, the team has decided to construct a linear regression model to evaluate the relationship between them.

```
Call:
lm(formula = AvgPrice ~ Population)

Residuals:
    Min      1Q  Median      3Q     Max
-339234 -120360  -57635  101253  467968

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.202e+06  7.661e+04  15.688  < 2e-16 ***
Population  -1.960e+00  3.002e-01  -6.529 2.95e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 191900 on 98 degrees of freedom
Multiple R-squared:  0.3031,    Adjusted R-squared:  0.296
F-statistic: 42.63 on 1 and 98 DF,  p-value: 2.95e-09
```

Figure 8: Linear regression model summary with Top 100 regions house price raised and its population from 2011 - 2021 in the UK

From this summary (figure 8) we can see the model is above satisfaction. Since the P-value of the model is smaller enough (lower than 0.05). Furthermore, the R-square of this model is acceptable. Then, we can visualise the model to view the correlation between population and housing price.

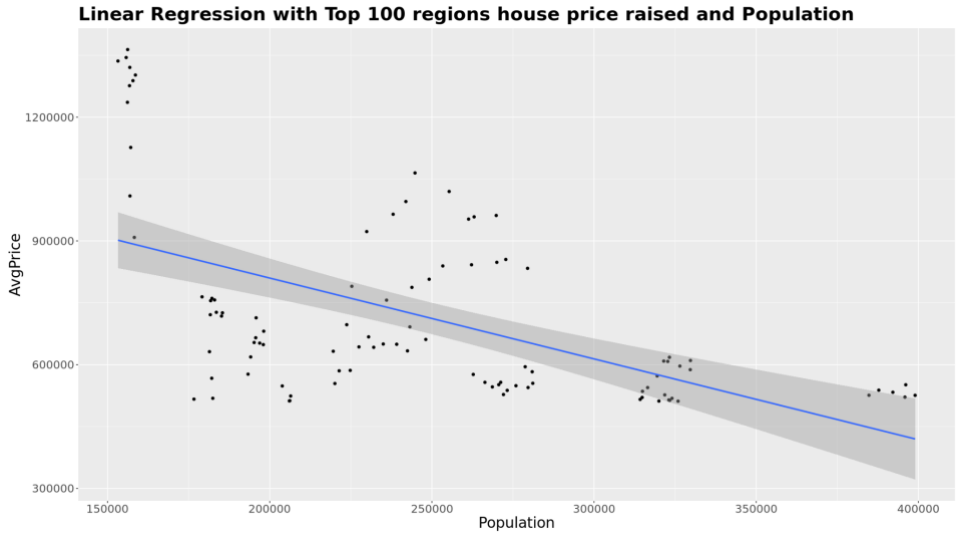

Figure 9: Linear regression of Top 100 regions house price raised and their population in the last decade in the UK

Based on figure 9, interestingly we noticed that when the population increases but the average house price decreases which violate the research we mentioned above. And we also can see some outliers in this linear regression indicating that even though the population is quite low, the average house price in that region is extremely high compared with other areas that have more populations.

Our team decided to conduct some research to find out the factors behind this result since it violates general rules “Higher population means higher demand on the house which will

increase the house price". According to York's (2022) analysis stated that one of the most critical truths that keep the UK houses increasing is the demand is still far outpacing the supply which means the demand for the house is keep increasing, however, the stock of the house for sales is far behind of the demand. Thus, to explain the result of the population analysis. Increasing population lead to increasing demand for some regions but the stock of houses in that region did not increase. Hence, the population in that area did not grow but house prices did. Furthermore, people who cannot afford the house price may choose other areas with low house prices to live which causes the population to increase. And based on the data exploration (figure x), the larger region could have more space to provide houses for people which may be less affected by the demand increase. Since the analysis is not comprehensively considered the factors related to the population such as region size, hence, we may conclude that a small region with a high demand for houses leads to house prices increase and the population in that area may also decrease because fewer people can afford it.

People's educational qualifications level

To keep drilling down the geographic factor, the team decided to look at people's education levels in the population. People spend a lot of time at university to obtain higher educational qualifications because they want to have a great job after graduation. According to B2B marketing (2022) stated that "there is a clear correlation between education level and salary. Generally, the more educated you are, the greater salary you becomes". Furthermore, the Office of National Statistics (B2B marketing, 2022) also indicates that people who have more advanced education and higher earnings throughout their life in the UK.

To begin with, the education qualification has been separated into 5 different levels, shown below table:

No qualifications	'No qualifications': No academic or professional qualifications.
Level 1	'1-4 GCSEs or equivalent (Level 1 qualifications): 1-4 O Levels/CSE/GCSEs (any grades), Entry Level, Foundation Diploma, NVQ level 1, Foundation GNVQ, Basic/Essential Skills.
Level 2	'5+ GCSEs or equivalent' (Level 2 qualifications): 5+ O Level (Passes)/CSEs (Grade 1)/GCSEs (Grades A*-C), School Certificate, 1 A Level/ 2-3 AS Levels/VCEs, Intermediate/Higher Diploma, Welsh Baccalaureate Intermediate Diploma, NVQ level 2, Intermediate GNVQ, City and Guilds Craft, BTEC First/General Diploma, RSA Diploma.
Trade Apprenticeship	'Apprenticeship': Apprenticeship
Level 3	'2+ A-levels or equivalent' (Level 3 qualifications): 2+ A

	Levels/VCEs, 4+ AS Levels, Higher School Certificate, Progression/Advanced Diploma, Welsh Baccalaureate Advanced Diploma, NVQ Level 3; Advanced GNVQ, City and Guilds Advanced Craft, ONC, OND, BTEC National, RSA Advanced Diploma
Level 4 +	'Degree level or above' (Level 4 qualifications and above): Degree (for example BA, BSc), Higher Degree (for example MA, PhD, PGCE), NVQ Level 4-5, HNC, HND, RSA Higher Diploma, BTEC Higher level, Foundation degree (NI), Professional qualifications (for example teaching, nursing, accountancy)
Other qualifications	'Other qualifications': Vocational/Work-related Qualifications, Foreign Qualifications/Qualifications gained outside the UK (NI) (Not stated/level unknown)

Table 2: Dictionary of education qualification level

The proportion for each educational qualification level in the sample of local authorities

To view the trends of the number of each education qualification can give us a brief view for each level qualification trends from 2004 to 2021.

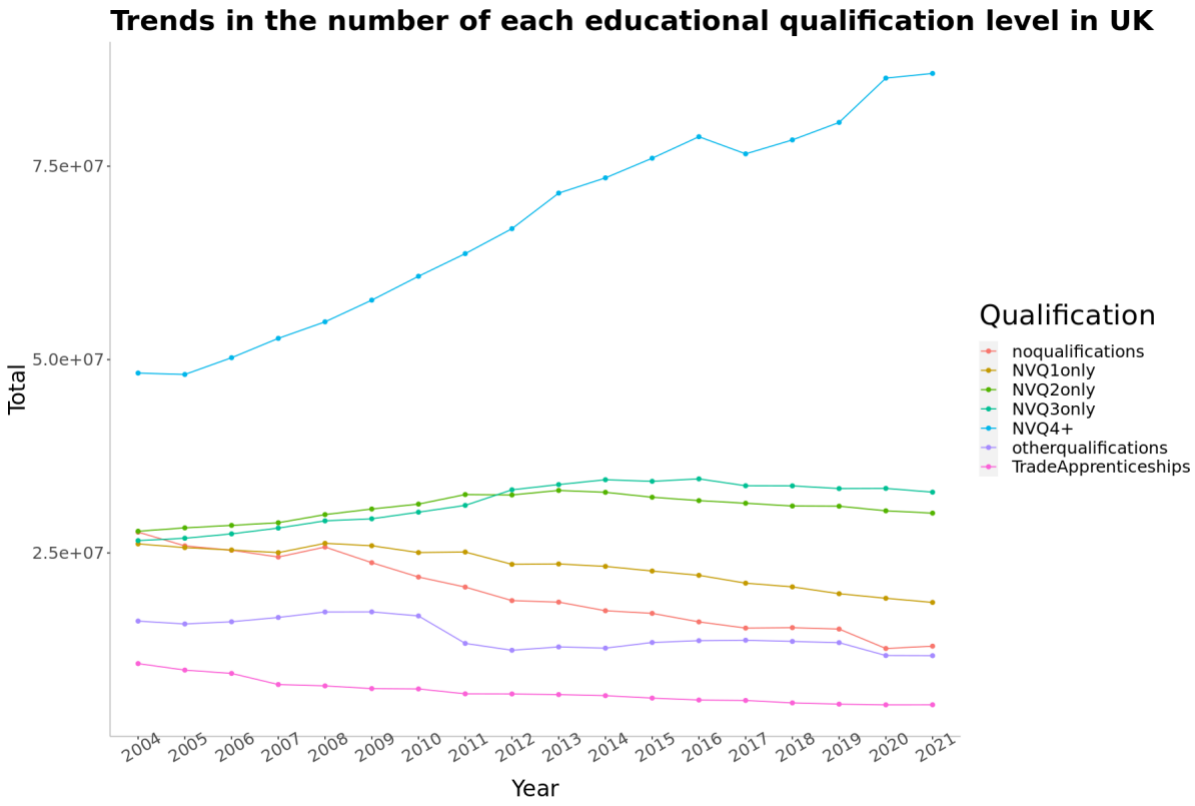

Figure 10: Trends of the number of each education qualification from 2004 to 2021 in the UK

According to this plot, we can see that the level 4 education qualification people have the largest proportion and keep increasing from 2004 to 2021 in the UK. Moreover, the other levels' qualification distribution does not have an obvious change in period. To combine the house price and the qualifications to do the analysis. The team has decided to narrow down the range of regions and the top 3 regions that have the highest and lowest housing prices changed in this time period to see each qualification's proportion in those areas.

The volume trend of high educational level people and the trend of house price

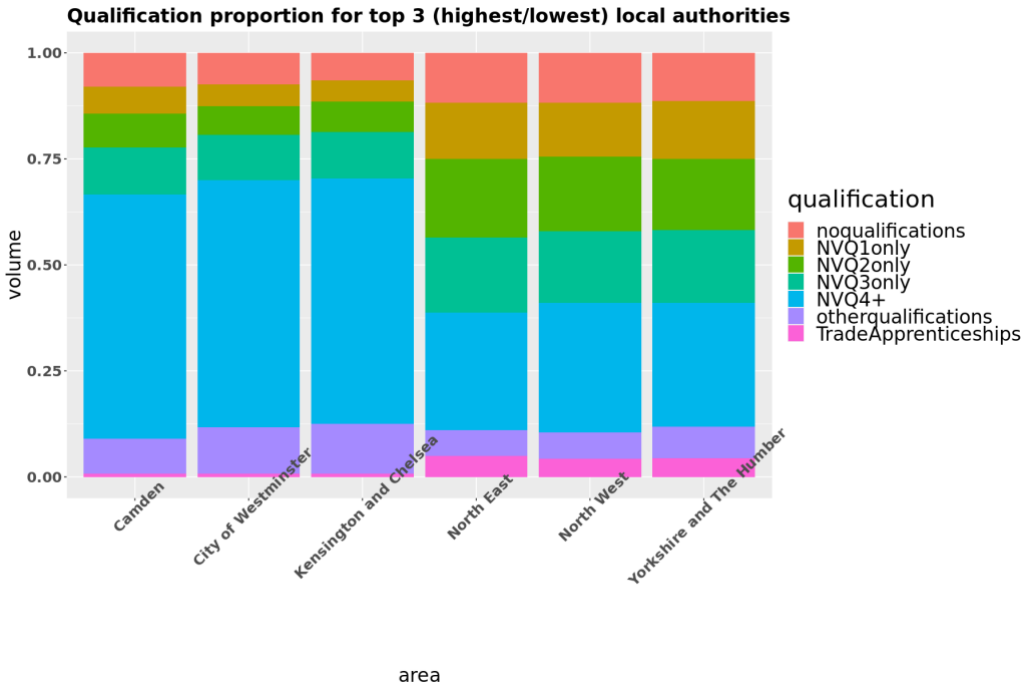

Figure 11: The proportion of educational qualifications in the highest(Left 3) and lowest (Right 3) house prices changed from 2004 - 2021 in the UK

According to figure 11, as we can see the regions which have increased house prices much more than others have a larger proportion of higher education qualification people. On the other hand, the low housing prices changed region does not indicate obvious features in this analysis. Therefore, this proves our research at the beginning that higher education people gain more salaries, and the higher incomes may increase the house prices. Thus, we assume the number of level 4 qualified people has linear regression with the house price. To prove this assumption, we have built a linear regression model.

Linear relationship evaluation

```
Call:
lm(formula = AvgPrice ~ number)

Residuals:
    Min      1Q  Median      3Q     Max
-150967 -109884  -33392   86674  329538

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.777e+05  1.698e+04  39.906   <2e-16 ***
number      -1.104e-01  6.893e-02  -1.602    0.113
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 128900 on 82 degrees of freedom
Multiple R-squared:  0.03035,   Adjusted R-squared:  0.01853
F-statistic: 2.567 on 1 and 82 DF,  p-value: 0.113
```

Figure 12: linear regression model summary of the number of higher education qualification people and house price

Figure 13: Linear regression of Top 100 regions house price raised and the number of high education qualification people in its region from 2004 - 2021 in the UK

The linear regression model shows there is no linear relationship between the number of higher education qualification people and house price. Furthermore, we can see the model summary shows that the P-value is over 0.05 which means there is weak evidence between these two variables. Furthermore, we can also observe the model R-square is very low which means the number of higher education people does not explain much to the house price.

Based on the result of people's education qualifications and house prices, we concluded that there is no linear relation between them which violates the research and assumption we made before. The reasons behind this could be that higher-income people may not intend to buy a house at a high price and the high-income people can also have low education levels. Since the analysis only selected a few areas to make an assumption, the house price may more correlate with the income rather than the qualification because higher qualification brings higher income.

Earnings

House price

The following is the median house price (existing dwellings) by country and region, England and Wales, year ending September 1997 to year ending September 2021 (£), the data is retrieved from UK Gov Housing Official Website, and shown as Figure 14 below.

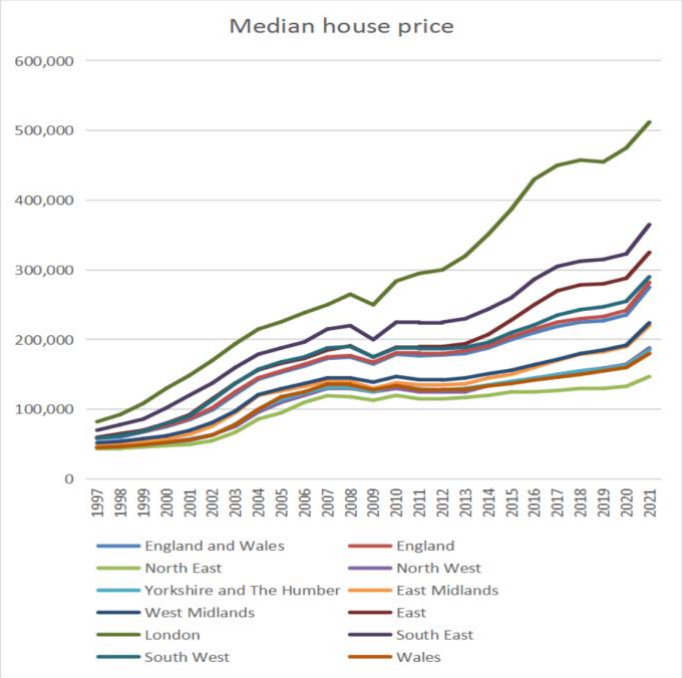

Figure 14. House prices in England and Wales

Here, we can find, from 1997 to 2021, the median house price in England and Wales, is rising in general. It was an exception in 2009. As we further analyse the reason for this sudden fall, we look back in time and find something that might have a very close relation to such a result. According to "*The Fall of the Market in the Fall" of 2008*. (2021, December 1). Investopedia due to the subprime economical crisis which took place in the US, the housing price went down significantly across the world, especially in the North American continent.

As for the UK, since the outbreak after-maths only reached Europe roughly a year later, we see the curve went down significantly all of a sudden in 2009, resulting in a sunken point in our plot. From 2009 to 2019, again, all curves rise rapidly, without much fast-changing. According to the *UK Housing Review 2019* (2018, December 20), during 2009 to 2019, the UK economy rose steadily in pace, thus resulting in relatively smooth growth of the housing price in the real estate market.

In 2020, and 2021, the rising price moves much faster for some reason. According to *Housing amid Covid-19: Policy responses and challenges*. (n.d.). (2020, OECD), in most countries, including the UK, the outbreak of covid-19 results in a rapid change in supply and need a

balance of the housing market, directly affecting housing prices to go up as a result. Therefore, due to the Covid-19 pandemic, we see even faster growth in housing prices in the UK market.

On the other hand, From north to south of England, the house price obviously goes up. In particular, London's housing price is always significantly higher than the rest of the regions due to its popularity and prosperity.

Earnings

The following is the Median gross annual (where available) workplace-based earnings by country and region, England and Wales, 1997 to 2021 (£), the datafrom UK Gov Official Website, shown as Figure 15.

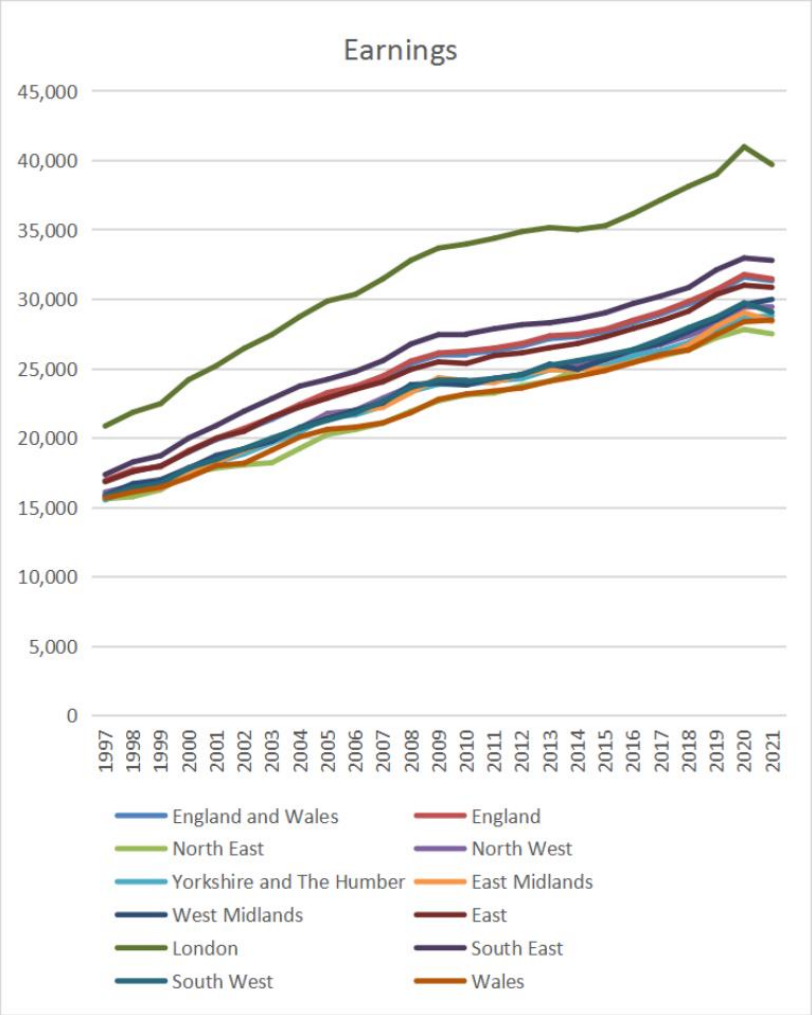

Figure 15: Annual earnings in England and Wales

To further analyse our case, we can analyse this visual result by timeline change and geographical their locations as demonstrated below:

We can find, from 1997 to 2021, the earning curve is always rising smoothly, without rapid changes. By country and region, London's earning curve has significantly higher value

compared to the rest of the regions. The earnings gap of the other regions is not big in England and Wales. According to the *2011 UK census* (2011, OCBC), London has the highest average earning among all regions across the UK due to its high degree of urbanisation. Hence, the exceptionally high earning curve of London is as expected.

Housing price to earnings ratio

The following is the Ratio of median house price (existing dwellings) to median gross annual (where available) workplace-based earnings by country and region, England and Wales, 1997 to 2021, the data from the UK Gov Official Website, shown as Figure 16 below.

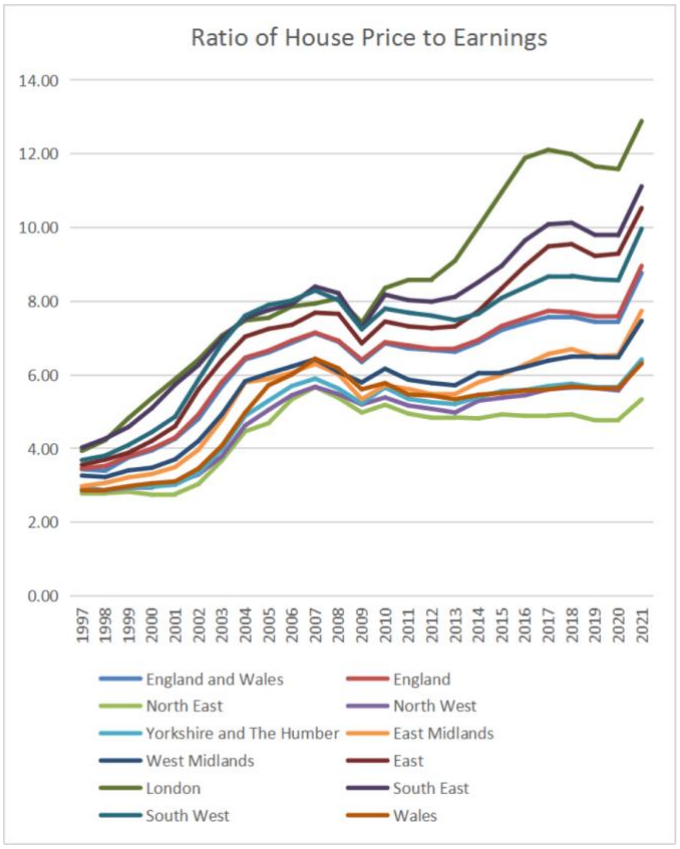

Figure 16: Ratio of house price to annual earnings in England and Wales

Here, we can find the curve of ratio of house price to earnings moves just like House price from 1997 to 2021.

From 1997 to 2021, the overall change trend of housing price to income ratio is basically consistent with the change trend of housing prices. Before 2003, the price to income ratio was relatively low. From 2003 to 2013, the ratio was basically between 5 to 8. After 2013, the ratio rose to 5 to 12, the highest in London, reaching more than 12.

From the above analysis, we notice that the curve is generally very smooth from 2008 onwards, considering the impact after the global economic crisis. This strongly suggests that the housing price could potentially have a positive linear relationship to income.

Linear correlation analysis

From the above house price to earnings ratio figure, we can see that house prices in England and Wales are related to earnings, and linear regression is used to study their correlation. First, look at England, out of the nine statistical regions, each of their linear regression is shown in Figure 17.

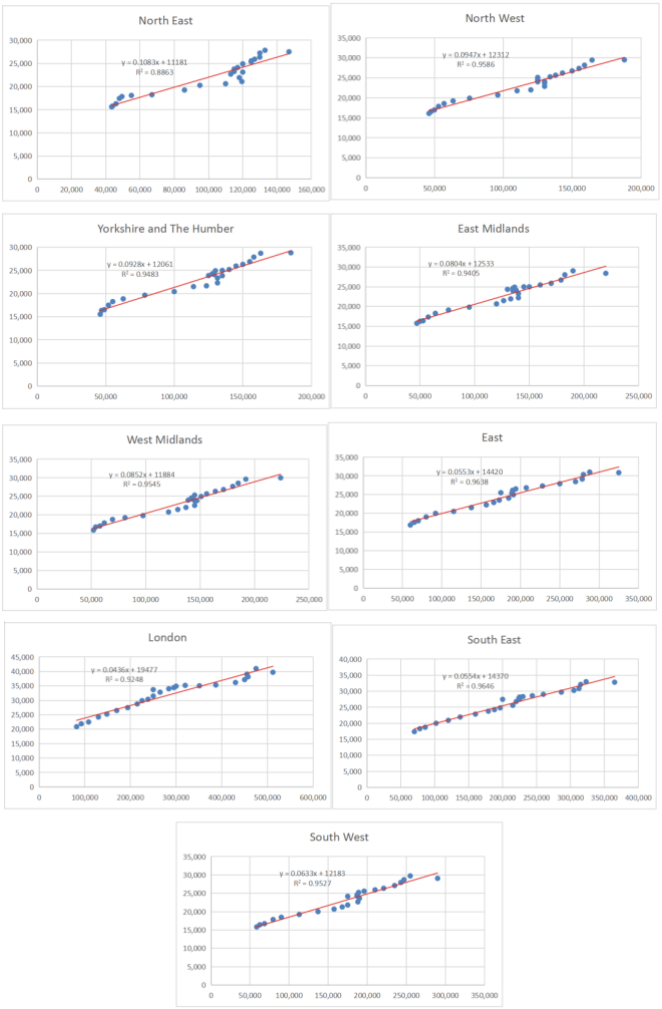

Figure 17: Linear regression of nine regions in England

From the whole of England (including the nine zones above), the linear regression is shown in Figure 18.

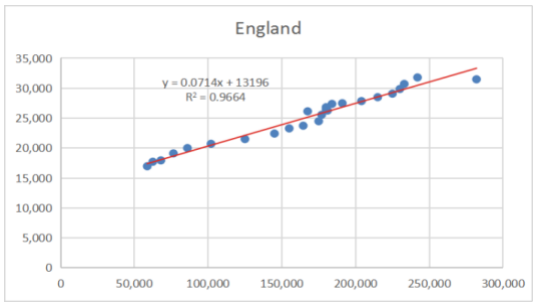

Figure 18: Linear regression of England

Then look at Wales alone, similarly, we also have linear regression as shown in Figure 19.

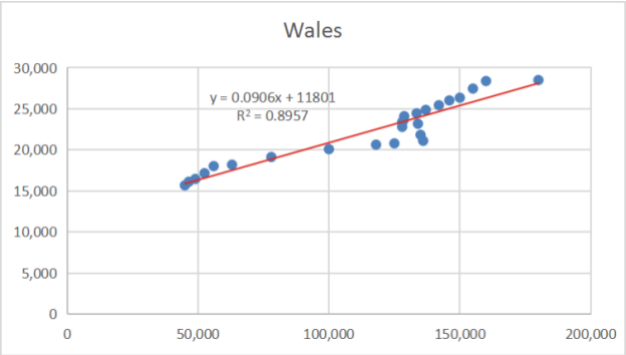

Figure 19: Linear regression of Wales

Finally, look at the average situation between England and Wales, whose linear regression is shown in Figure 20.

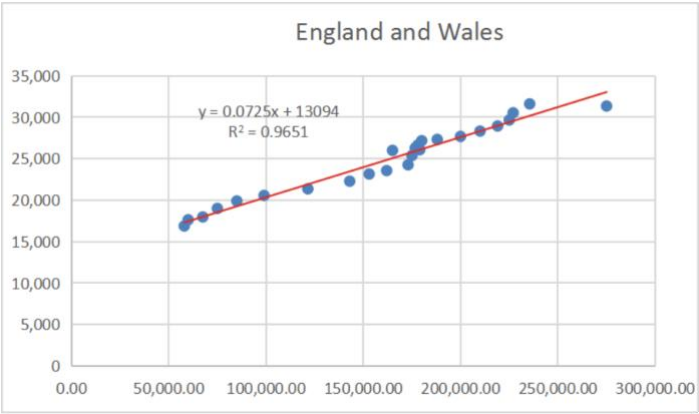

Figure 20: Linear regression of England and Wales

From the above graphics, we can all find a good correlation between house price and earnings.

Besides, according to Investor, T. (2012, March 15), the housing price alone in UK the housing price to earnings ratio in the UK is relatively unchanged since 1997 till 2011's census, meaning although many factors do affect housing prices other than earning, if we consider people's average earnings alone, it does have strong effects on the fluctuation of housing prices, as they both rise up or fall down in the same rhythm across time. Thus, with the findings and evidence we list above, we have good reasons to suggest that housing prices have a positive correlation with people's average earnings in general.

Linear model and Validation

In order to examine our findings before, we now take the P-value for Housing Price vs Average Earnings (e.g. England and Wales):

Ttest_indResult (statistic = 7.098, P-value = 1.25)

Figure 21: P-value for Housing Price vs Average Earnings

Similarly, we apply the same method to find P-values in other regions as well as the corresponding correlation coefficient R^2. Shown in Table 1 and Figure 26 below:

Table 1. Correlation between House Price and Annual Earnings

Name	Correlation coefficient R^2	P-value (Using Python)
England and Wales	0.9651	1.25*e-5
England	0.9664	1.0*e-5
North East	0.8863	1.26*e-5
North West	0.9586	0.86*e-5
Yorkshire and The Humber	0.9483	1.25*e-5
East Midlands	0.9405	1.34*e-5
West Midlands	0.9545	0.9*e-5
East	0.9638	1.31*e-5
London	0.9248	1.32*e-5
South East	0.9646	0.7*e-5
South West	0.9527	0.86*e-5
Wales	0.8957	1.52*e-4

Table 3: Correlation coefficient R^2 by country and region, England and Wales

Here, we notice that since the p values are all small enough (close to 0), indicating statistically significant results. And the correlation coefficient R^2 is more than 0.88. Thus, we reject the null hypothesis and conclude that Average earnings does have a strong correlation and a **linear relationship** with Price.

Enterprise Birth/Death rate

House prices are not only closely related to personal living standards, but also affected by local economic conditions and the number of enterprises. According to the House Money and Entrepreneurship (2017) point out that high housing prices may discourage entrepreneurial activities, which may lead to slow economic development and finally affect the local house price.

As a result, we assume that there may exist a relationship between the birth/death rate of local enterprises and house price changing trends. Areas that have a high rate of new enterprises may also have a high land usage. Line charts and linear regression models are used to find out the relationship between the birth/death rate of local enterprises and the change in house prices.

These two graphs in the following show the time series of birth rate and death rate of enterprises in the two areas, which have the highest (the red one) and lowest (the green one) annual average house price change.

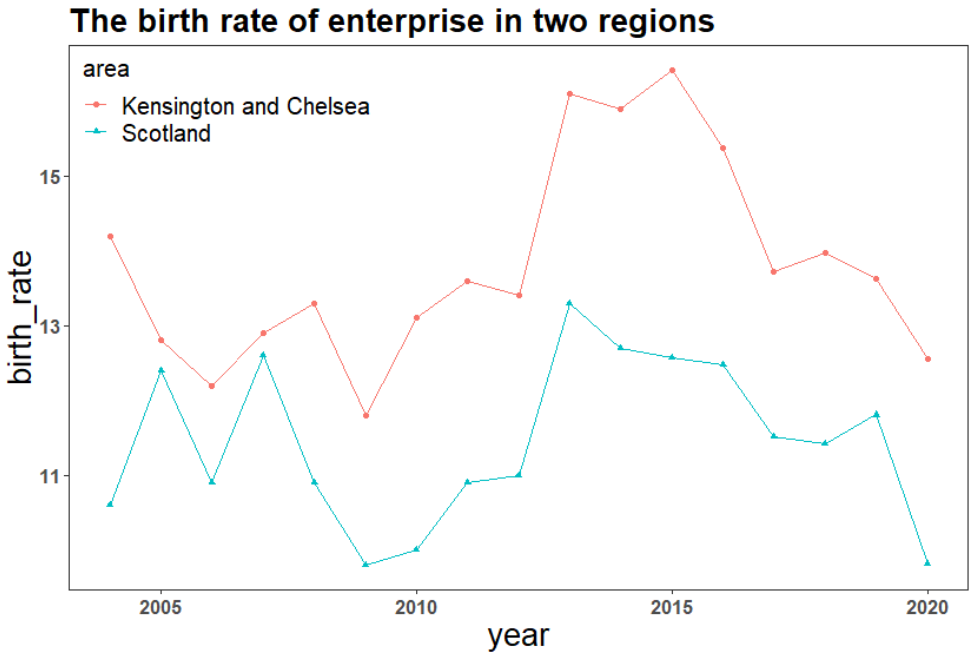

Figure 21: Birth rate of Enterprise versus Local Housing Price Change

Based on figure 21, we can clearly find that a high birth rate of enterprises may lead to a significant change in house prices. The red line stays above the green line for the whole period of time, which implies a positive relationship between these two variables. However, it is only a rough conclusion and further statistical study is needed to approve this assumption.

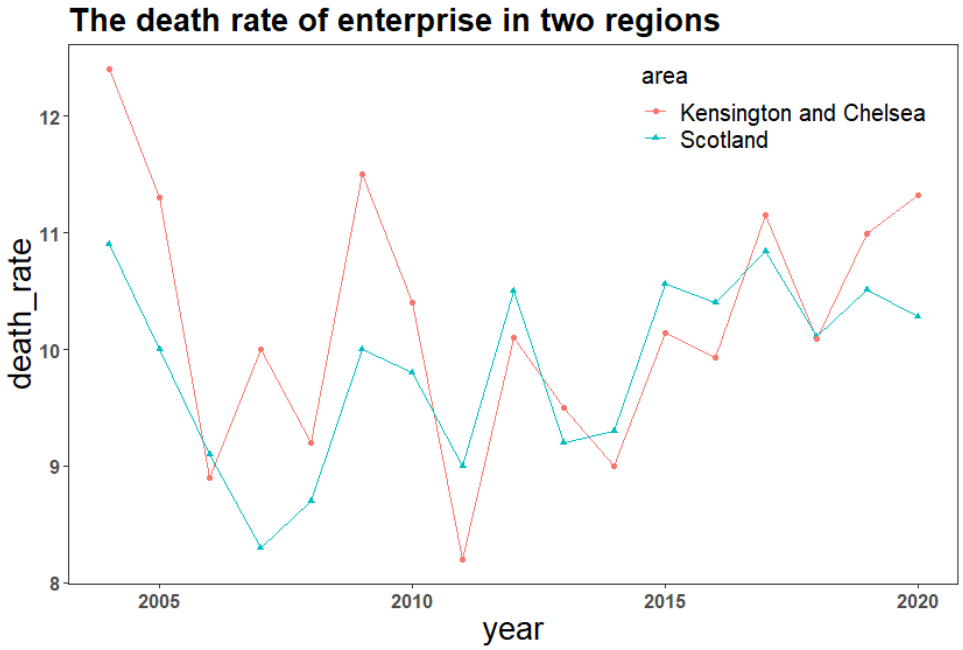

Figure 22: Birth rate of Enterprise versus Local Housing Price Change

In contrast, the death rate may have no connection with the change in house price. The result stays the same when line charts of more cities and regions are created. In order to find out whether there is a statistical linear relationship between these variables, R software is used to build the regression model and check the relationship.

```
Call:
lm(formula = Annual_Change ~ birth_rate, data = all)

Residuals:
     Min       1Q   Median       3Q      Max
-16.1759  -4.6932  -1.6327   0.8864  28.4378

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.4986     5.6043   0.267    0.790
birth_rate    0.2837     0.4516   0.628    0.531

Residual standard error: 8.819 on 100 degrees of freedom
Multiple R-squared:  0.003931,  Adjusted R-squared:  -0.00603
F-statistic: 0.3946 on 1 and 100 DF,  p-value: 0.5313
```

Figure 23: Linear regression model summary of the birth rate of enterprise per year versus local annual average house change

In the summary of the model (figure 23), the p-value is over 0.5, which means that the relationship between birth rate and the annual average change in house price is not statistically significant. The scatter plot with a regression curve also implies we cannot only use the enterprise birth rate to predict the change rate of the house price.

Besides referring to the summary of the model, the prediction interval, which can be visualised by drawing diagrams, can also be used to interpret the relationship between the birth rate of enterprises and the change in housing prices. Half of the points are outside of the intervals and from which we can contribute to the same result as the model summary in the following.

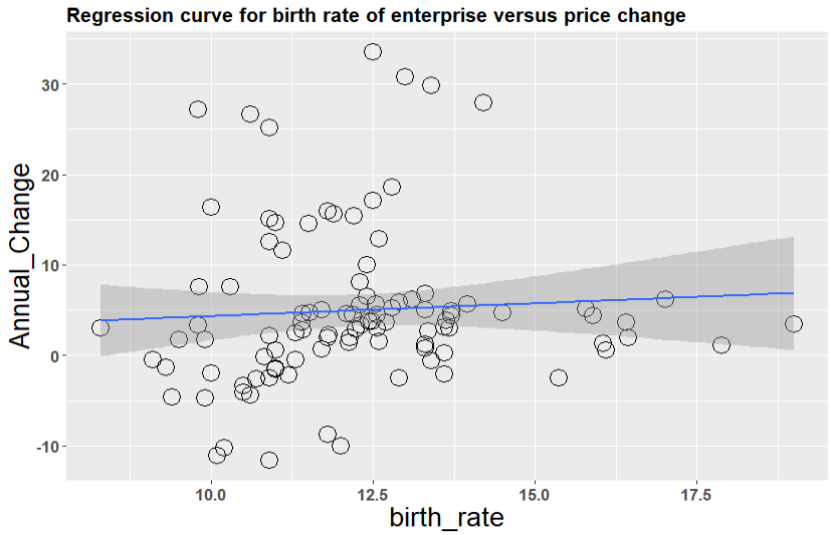

Figure 24: Prediction interval of the birth rate of enterprise versus price change

In conclusion, the birth and death rates of enterprises are supposed to be regarded as a response variable instead of an independent variable which may lead to the difference in house price annual change. More variables should be taken into consideration for building a prediction model for the change in housing prices.

The study also indicates that there is no linear relationship between the housing price change and neither birth rate or the death rate of enterprises. The birth rate of enterprise is more suitable to be regarded as a response variable instead of the independent variable and may be affected by some other objective variables, for example, the current economic condition or some sudden events, such as the covid-19 disease.

Affordable housing supply

Another possible angle emerges when we think about the factors that influence house prices, namely affordable housing. According to housingforwardva, affordable housing is important to families and economic vitality of communities. Today affordable housing as a broad category of property is an important part of the overall property market in the UK and is likely to come within the reach of investors in the future, just like other types of property products. So we consider that there may be a correlation between the number of affordable housing units available and the price of housing.

The trend in the number of affordable houses

To evaluate the correlation between the number of affordable houses and house prices. The overall trend of the affordable housing supply should be presented to give a brief view of the trend and compare it with the house price trend. To do this, we first made a graph of the number of affordable housing units from 2004 to 2021. The number of affordable housing units here refers to the sum of the number of all types of affordable housing units in that year.

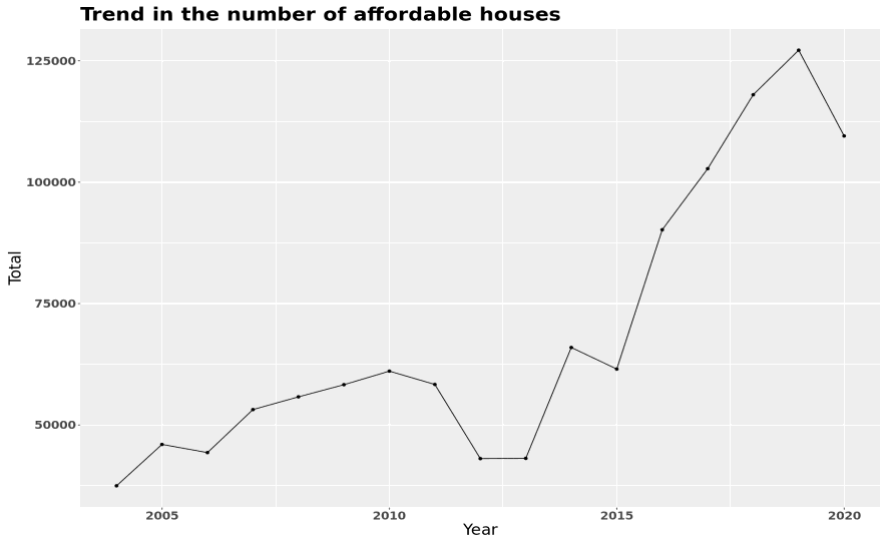

Figure 25: Trend in the number of affordable houses

In figure 25 we can see the similarity between the trend of the number of affordable housing units over time and the trend of house prices over time, so we speculate that the factor of the number of affordable housing units deserves to be discussed. Therefore, in order to more accurately use the data to determine the strength of the correlation between the number of

affordable houses and house prices, our group chose to use linear regression model as a method of inquiry.

Linear relationship evaluation

```
Call:
lm(formula = AvgPrice ~ Units)

Residuals:
    Min      1Q  Median      3Q     Max
-157540  -70721  -25609   36054 1161429

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 202185.74    2452.89  82.428  < 2e-16 ***
Units           93.60      11.58   8.085 7.74e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 113100 on 5026 degrees of freedom
Multiple R-squared:  0.01284,   Adjusted R-squared:  0.01264
F-statistic: 65.36 on 1 and 5026 DF,  p-value: 7.742e-16
```

Figure 26: Linear regression model summary with the number of affordable housing units in each city on the average house price in each city.

From this summary, we can see the model is not particularly good. Since the P-value of the model is smaller enough (lower than 0.05). However, the R-square of this model is small, even close to 0, proving that the correlation between the two is needed for further discussion. Then, we can visualise the model to view the correlation between the number of affordable houses and housing prices.

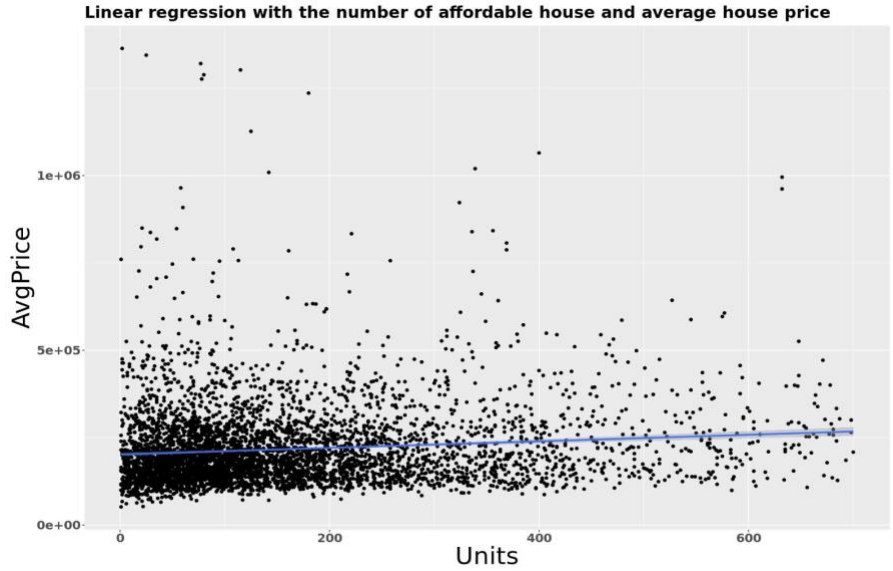

Figure 27: Linear regression of the number of affordable housing units in each city on the average house price in each city.

The linear regression model shows that in a region, when the number of affordable houses increased the house price increased as well. So we can dig deeper to see if the number of certain specific types of affordable housing correlates with housing prices.

Selecting the proper type of affordable housing to discuss

According to the linear regression model from figure 3, we know when the number of affordable houses increases the house price in that region increases as well. Then the affordable house has been separated into 7 different types. Therefore, to drill down the analysis, we would like to observe the trend in different types of affordable houses. The trend can provide an overview of whether the number of a particular type of affordable house may significantly affect the house price.

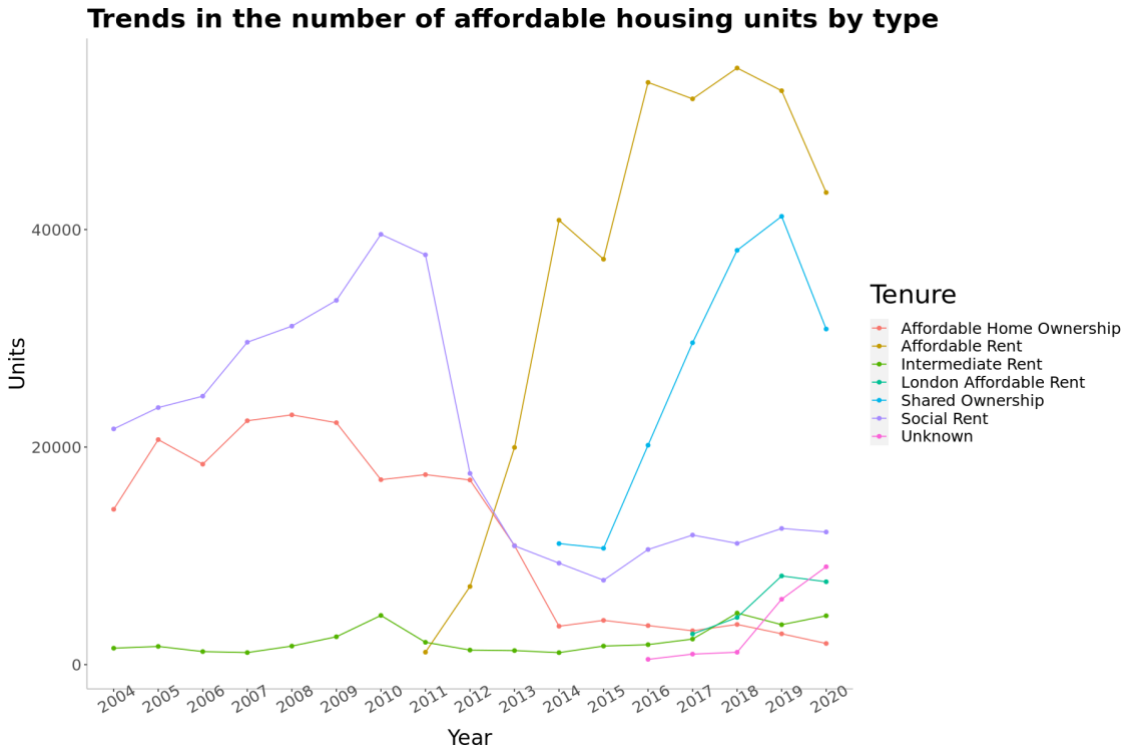

Figure 28: Trends in the number of affordable housing units by type

According to figure 28, we can see that the two lines Affordable Rent and shared ownership have similar trends to house prices, but given the relatively short time period of Shared Ownership, we choose the type Affordable Rent to discuss the relationship with house prices. We still use the linear regression model to see the correlation between the two.

Linear relationship evaluation

We did a simple linear regression of the number of affordable housing units by city for the type Affordable Rent on the average house price by city.

```
Call:
lm(formula = AvgPrice ~ Units)

Residuals:
    Min      1Q  Median      3Q     Max
-177254  -92541  -33109   49400 1120812

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 242935.87    3558.31  68.273   <2e-16 ***
Units           27.05      18.09   1.495    0.135
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 135800 on 2683 degrees of freedom
Multiple R-squared:  0.0008326, Adjusted R-squared:  0.0004602
F-statistic: 2.236 on 1 and 2683 DF,  p-value: 0.135
```

Figure 29: Linear regression model summary with the number of affordable housing units for the type of Affordable Rent house in each city on the average house price in each city.

From this summary (Figure 29), we can see that the model fit is not particularly ideal. The P-value is relatively large (over than 0.05), which means the test results are not significant. And the value of R-squared is even smaller than the first model, which means the correlation between the number of Affordable Rent houses and average house price is weaker than the correlation between the number of affordable house units and average house price in each city.

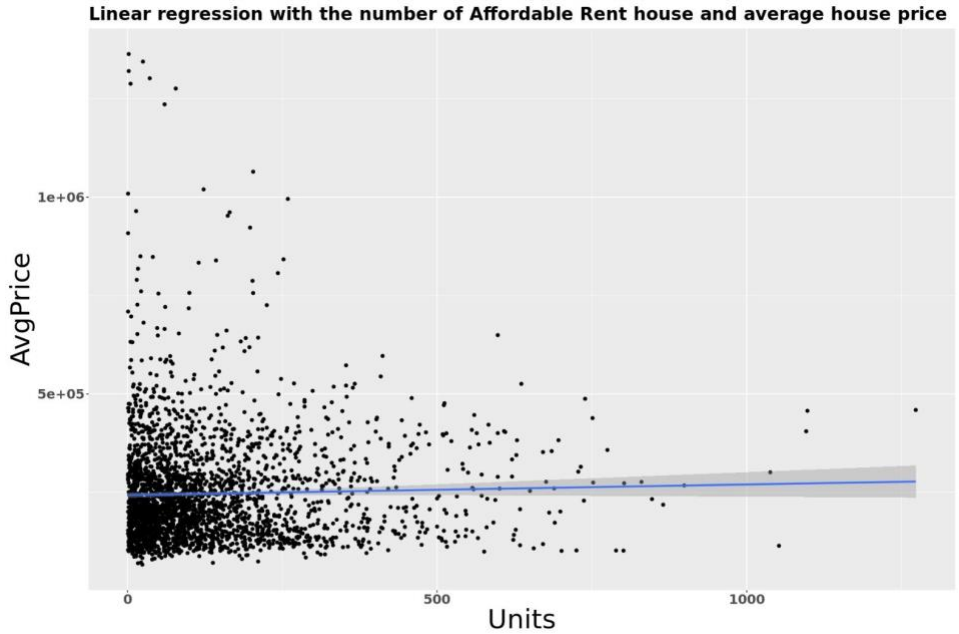

Figure 30: Linear regression with the number of Affordable Rent house and average house price

The linear regression model (Figure 30) shows that most of the data are clustered in one area and do not have a good linear regression relationship. So, there is no linear relation between the number of particular affordable house types and house price.

Based on the results, we found that there is a linear relationship between the number of affordable houses and house prices, which means when the number of affordable houses increases the house price increases as well. However, there is no obvious linear relationship between the number of particular types of affordable houses and house price. We think the possible reason for the weak correlation in figure 26 is that we need to clear some extreme data, because from the dataset, there are cities where affordable housing is not available at certain time periods. For example, areas like London, Inner London, etc., which happen to be cities with relatively high housing prices. Therefore, the presence of these data is likely to affect the correlation of the regression model between the number of affordable housing units and the house prices to some extent.

People living space quality

According to the UK house price fluctuation map, most of the cities with large house price changes are in areas centred on London. As the capital and economic centre of the United Kingdom, the development of London has driven the surrounding housing prices. Outside of this area, housing prices are generally not as high as those in the inner circle of London. But there are exceptions, one of the more volatile cities is Edinburgh.

Clustered by year, according to the primary dataset, plot the trend line of increase and decrease of house prices in Edinburgh and the United Kingdom. From the line chart Figure 31, it can be concluded that there is not much difference in the trend of increase and decrease between the two. In other words, Edinburgh has no major special events that cause house price volatility to be different from other cities, and its house price trend is in line with the overall house price change in the UK.

However, the fluctuation of house prices in Edinburgh is relatively large. In the following analysis we will explore what makes Edinburgh special. And to find out what are the advantages of its house prices that have changed significantly in recent years.

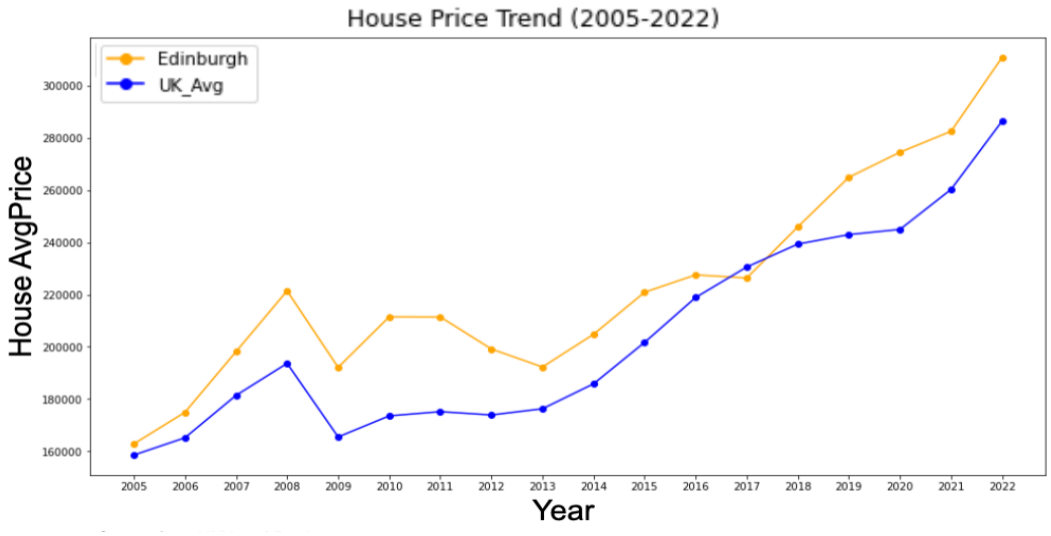

Figure 31: House price trend of Edinburgh and the UK

Edinburgh has topped the list of Scotland's most dynamic regions, according to the latest Vibrant Economics Index, published by consultancy Grant Thornton, which measures the social environment and the economy. In addition, Edinburgh topped the list in Britain's best place to live and work in a competition organised by the Royal Mail. This may suggest that the high volatility of Edinburgh house prices may be related to quality of life We need to find out by analysing the data. According to Deutsche Bank's statistics on the cities of quality of life in the world, Edinburgh's liveability index has been consistently high among European cities (shown in Figure 32). The statistic measures the quality of life in a city based on multiple indicators such as purchasing power, health care, pollution, cost of living, etc.

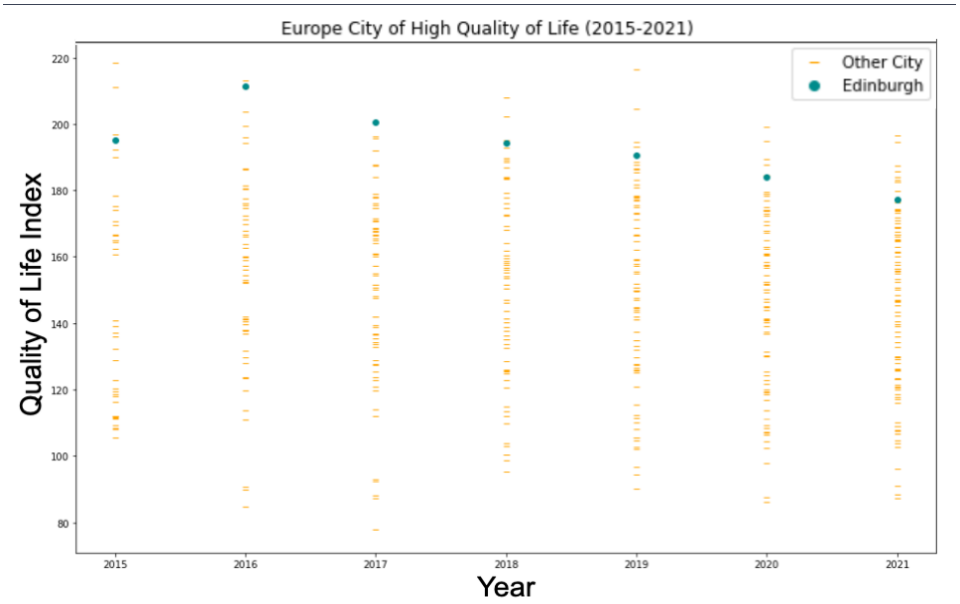

Figure 32: Europe City of High Quality of Life (2015-2021)

However, this still does not mean that quality of life can be the main factor affecting house prices, as Edinburgh may be a special case that is affected by other factors. Let's filter out all the UK cities on the list and bring back the original dataset to see how these cities are rated as liveable cities, which included Belfast, Birmingham, Bristol, Cambridge, Glasgow, Leeds, Liverpool, London Edinburgh and Manchester. It can be found that many of these cities are located outside the "high house price growth ring" around London, such as Belfast, Leeds. This may provide evidence that the quality of life has improved housing prices. However we still need evidence, so we need to compare house price fluctuations in these cities with other cities to see if they can support this conclusion. For this reason, we need a scatterplot that includes all UK cities in the dataset, and highlights the ranked cities to see if they outperform others.

However, the results were unexpected. It turned out not to be the case. According to Figure 33, it can be found that even the housing prices in cities rated as having a high quality of life are not significantly different from other cities, but mixed in with other cities.

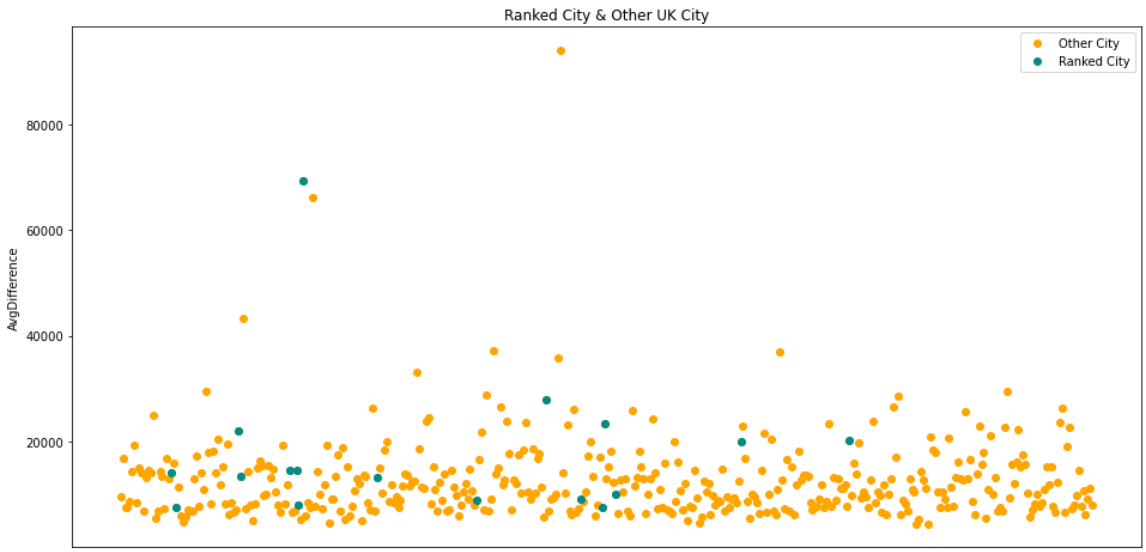

Figure 33: Average differences of ranked cities and other UK cities

This cannot be conclusive evidence that quality of life can affect city house prices. In turn, it only provides that the quality of life is not a strong correlation factor for UK housing prices.

There is no doubt that people prefer to live in places with a high quality of life. However, this does not directly mean that the fluctuation of housing prices in places with a high-quality living environment will be very significant. Our analysis also proves this, quality of life does not decisively affect the fluctuation of house prices.

Changes in housing prices are based on many factors, such as local labour costs (related to people's income), land fees (related to urban density and vacant areas), local policies and population (related to local housing demand), famous attractions and high Educational institutions (the demand of foreigners to buy houses), etc. Perhaps under the influence of these various factors, the impact on the quality of life is negligible. In order to grasp the changes in housing prices, there are still many factors to consider.

Conclusion

The data exploration of the main dataset demonstrated that the United Kingdom house prices increased most of the time excluding the short period affected by the global financial crisis between 2007 and 2008. And the UK house price growth is uneven and Most regions have significant house price increases around the city of London. And there are some special outliers that are not around the city of London but also have dramatic house price increases such as Edinburgh.

Depending on the analysis results, we found that the UK house price has a linear correlation with the local earnings and the number of affordable houses. This means when a local people's earnings increase and the number of affordable houses increase, the related region's house price will increase as well. Furthermore, we also found that a small area with high demand for houses will increase the house price significantly. This led to the population analysis indicating the high house price increases regions contain less population.

The project also found that the UK house price does not have a linear correlation with the number of educational qualifications, the number of particular types of affordable houses and the enterprise birth/death rate. Moreover, analysis demonstrates the UK house price is not affected by the quality of living space.

Recommendations

Based on the results conducted of this project, we recommend the potential house buyer/investor purchase property when the local authority announces an increase in the affordable housing supply. Furthermore, if people's earnings trends are increasing in the region we would recommend the buyer purchase the property since there is a strong linear relation between people's earnings and house prices. Although the analysis found that the higher house price area contains less population and the region that has more population will decrease the house price. However, the model does not comprehensively demonstrate the correlation between population and house price such as region size. Thus, we will recommend the buyer/investor comprehensively consider the factors that may affect the house price such as the demand for the house and local policy.

Reference List

B2B marketing. (2022). *What Effect Does Education Level Have on Wealth?*. QS. https://www.qs.com/what-effect-does-education-level-have-on-wealth/#:~:text=There's%20a%20clear%20correlation%20between,the%20greater%20your%20salary%20becomes.&text=Focusing%20on%20the%20UK%20specifically,your%20salary%20throughout%20your%20lifetime.

Financial crisis of 2007–2008 - Wikipedia. En.wikipedia.org. (2022). Wikipedia. https://en.wikipedia.org/wiki/Financial_crisis_of_2007%E2%80%932008.

Foote. A. (2016). The effects of negative house price changes on migration: Evidence across U.S. housing downturns. Regional Science and Urban Economics, 60, 292–299. https://doi.org/10.1016/j.regsciurbeco.2016.08.001

Housing amid Covid-19: Policy responses and challenges. (n.d.). OECD. https://www.oecd.org/coronavirus/policy-responses/housing-amid-covid-19-policy-responses-and-challenges-cfdc08a8/

Housing forward. (n.d.). *Why is affordable housing important? Is rental or homeownership more important?* https://housingforwardva.org/toolkits/affordable-housing-101/why-is-affordable-housing-important-is-rental-or-homeownership-more-important/

I. (n.d.). *UK Housing Review 2019*. Investopedia. https://www.ukhousingreview.org.uk/ukhr19/index.html

Investor, T. (2012, March 15). *House prices versus earnings*. Monevator. https://monevator.com/house-price-to-earnings-ratio-2012/

Kosakowski, P. (2021) *House Money and Entrepreneurship* KKN-Nov17.pdf (columbia.edu)

Mail, R. (2018). *Britain's 'BEST' cities ranked as the most attractive places to live and work* dailymail.co.uk. https://www.dailymail.co.uk/news/article-5504695/Britains-BEST-cities-live-work-in.html

Miles. (2012). Population Density, House Prices and Mortgage Design. Scottish Journal of Political Economy, 59(5), 444–466. https://doi.org/10.1111/j.1467-9485.2012.00589.x

Numbeo. (2022). *Cost of Living.* numbeo.com https://www.numbeo.com/cost-of-living/

Moeller, P. (2022). *Why Our Homes Make Us Happy*. U.S.News. Retrieved 2 June 2022, from https://money.usnews.com/money/personal-finance/articles/2012/04/05/why-our-homes-make-us-happy#:~:text=Home%20provides%20security%2C%20control%2C%20belonging,which%20we%20return%20each%20evening.%22.

The Fall of the Market in the Fall of 2008. (2021, December 1). Investopedia. https://www.investopedia.com/articles/economics/09/subprime-market-2008.asp#:%7E:text=The%20stock%20market%20and%20housing%20crash%20of%202008,and%20a%20higher%20risk%20of%20defaulting%20on%20loans.

Thornton, G.(2016). *Vibrant Economy Index.* Grantthornton.co.uk. https://www.grantthornton.co.uk/globalassets/1.-member-firms/united-kingdom/pdf/documents/vibrant-economy-index-summary.pdf

York, M. (2022). *Why do UK house prices keep rising?*. Thetimes.co.uk. https://www.thetimes.co.uk/article/why-do-uk-house-prices-keep-rising-62kn9pxfs.