Capstone Project 1

Milestone Report

1   Background

The project examines the movement of house prices for 62 major cities in the UK in the period from 2003/4 to 2015/16 (the start and end years vary depending on availability of data) and which factors influence house price dynamics.

The data set includes information on indicators for each city, including average house prices, housing stock, population, employment, average wages and other data. Each of these factors was correlated to house prices across the 62 cities, in order to identify the factors where there is a relationship of statistical significance.

2   Problem and client

Understanding the behavior of house prices is of huge interest to banks, real estate investors, property developers, and the buying public. A typical client for this analysis would be a UK-based property developer with a portfolio of projects across the country. The results of this exercise will help answer the following questions:

- What key factors influence the movement of house prices in major cities the UK? What makes house prices certain cities more expensive than in others?

- Which towns and cities should we commit our resources on? Can we use any factors to predict the selling price of our developments e.g. one year from now?

- How is the overall level of house prices in the UK related to the FTSE100 stock market index?

The answers obtained from the exercise should ultimately help to define an investment strategy.

3   Data source and steps to clean data

The data set used is published by the Centre for Cities (www.centreforcities.org). The site has a data tool where data can be downloaded in the form of a .csv file.

In order to arrive at a clean data set, the following actions were performed:

- Importing: reading the .csv file into a pandas data frame and exploring content of 63 rows (one row for each city) and 400+ columns (indicators)

- Cleaning: removing data from the footer, removing columns that contain no data and one row with missing price data, converting text rows into floating point numbers

- Manipulation: extracting slices of the data set to obtain specific annual data of interest (house prices, population, wages, employment, housing stock, etc.)
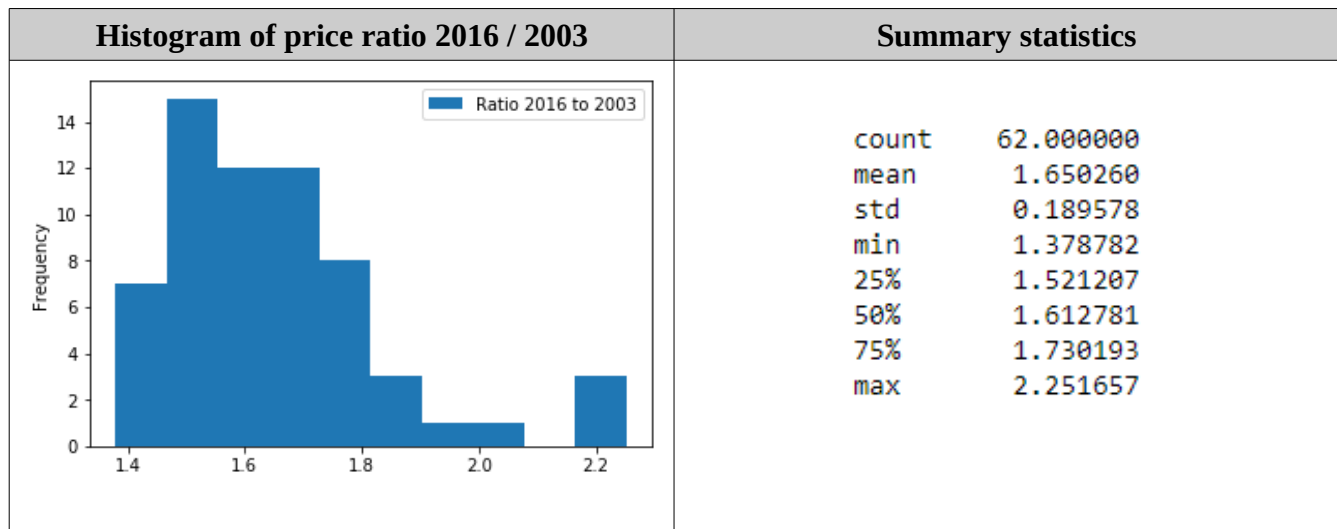
The final data set contains 62 rows and 400 columns, of which one column represents the city name and the remaining 399 columns are indicators of data type float.

For the correlation of FTSE100 data with average UK house prices, the House Price Index data set was used from www.gov.uk and merged with monthly data on FTSE100. The two data sets were merged together using dates as the joining parameter.

4   Exploratory data analysis

4.1   Prices shift from 2003 to 2016

The preliminary analysis involved examining the movement of house prices for the cities and understanding their progression over the period 2003-2016.

Histogram of price ratio 2016 / 2003	Summary statistics
	count 62.000000 mean 1.650260 std 0.189578 min 1.378782 25% 1.521207 50% 1.612781 75% 1.730193 max 2.251657

The analysis shows that prices have increased in all of the cities from 2003 to 2016. The highest increase was a factor of 2.25x, the lowest 1.38x and the mean 1.65x.

4.2   Affordability

Although house prices have risen by an average of 60% from 2004 to 2016, wages have on average remained nearly constant. As a result, it has become increasingly harder for an average worker to afford fa house. Considering the 62 cities, on average 5.3 annual salaries were needed to buy a house in 2004, whereas by 2016 this figure has has increased to 8.0 years. The standard deviation of prices (normalized by the mean price in each year) has also increased, meaning that the variation between cities has become more extreme.

| Average wage ratio 2016 to 2004	Average no. of annual wages needed to buy a house
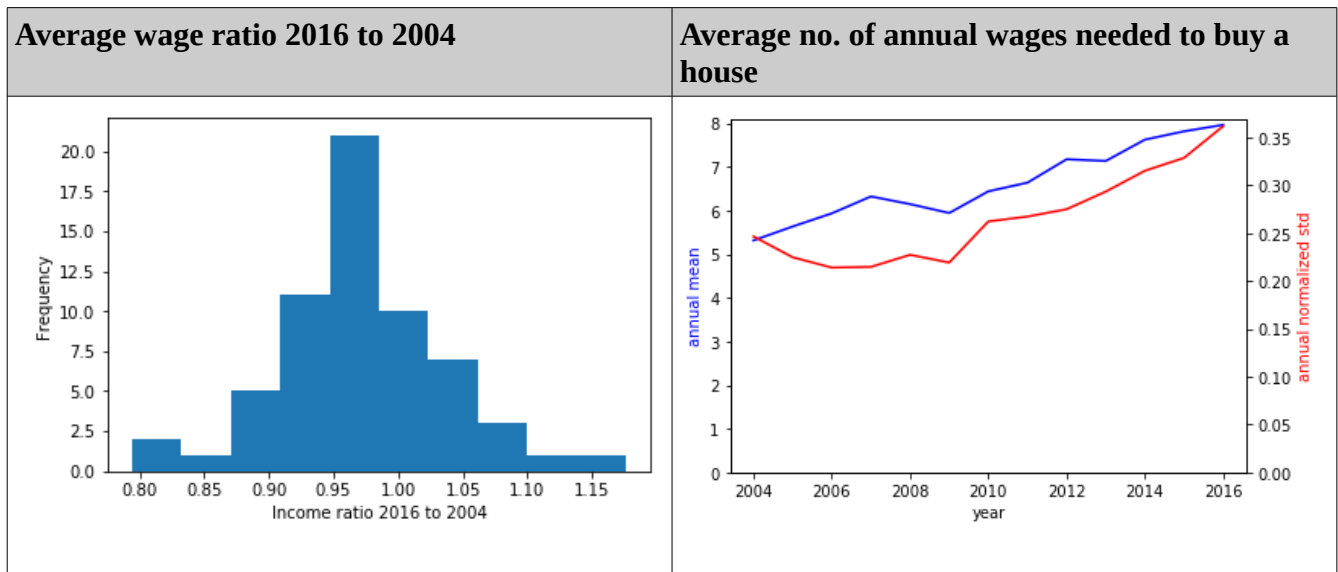
4.3   Location-based analysis

Location-based prices for the year 2016 from the Centre for Cities website:
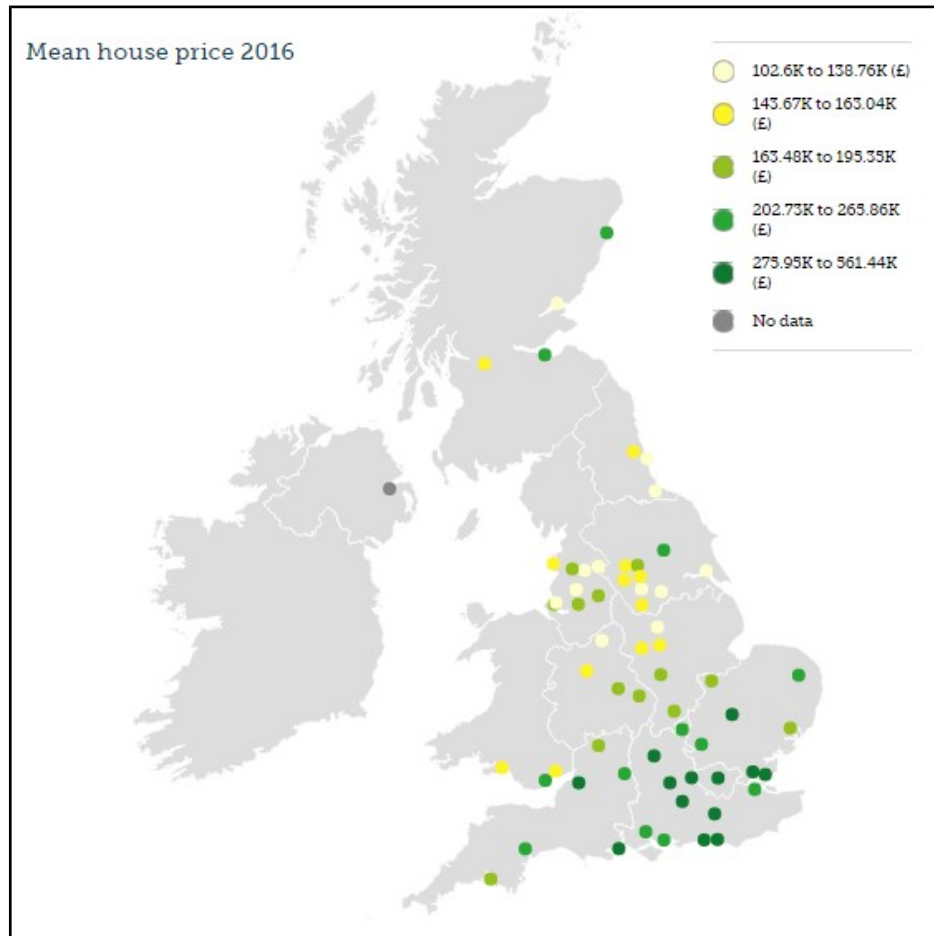
This map clearly shows the geographical distribution of prices, noting that the Southeast is the priciest (green dots), followed by a few cities on the South coast and the West. Cities get progressively cheaper the further north they are in the Midlands, while the northern industrial belt has considerable variation. A small number of towns in Scotland and the North have prices that match those in the Southeast.

The price movement of some specific cities are shown below:

Most expensive cities

Low price and low employment

North

Midlands
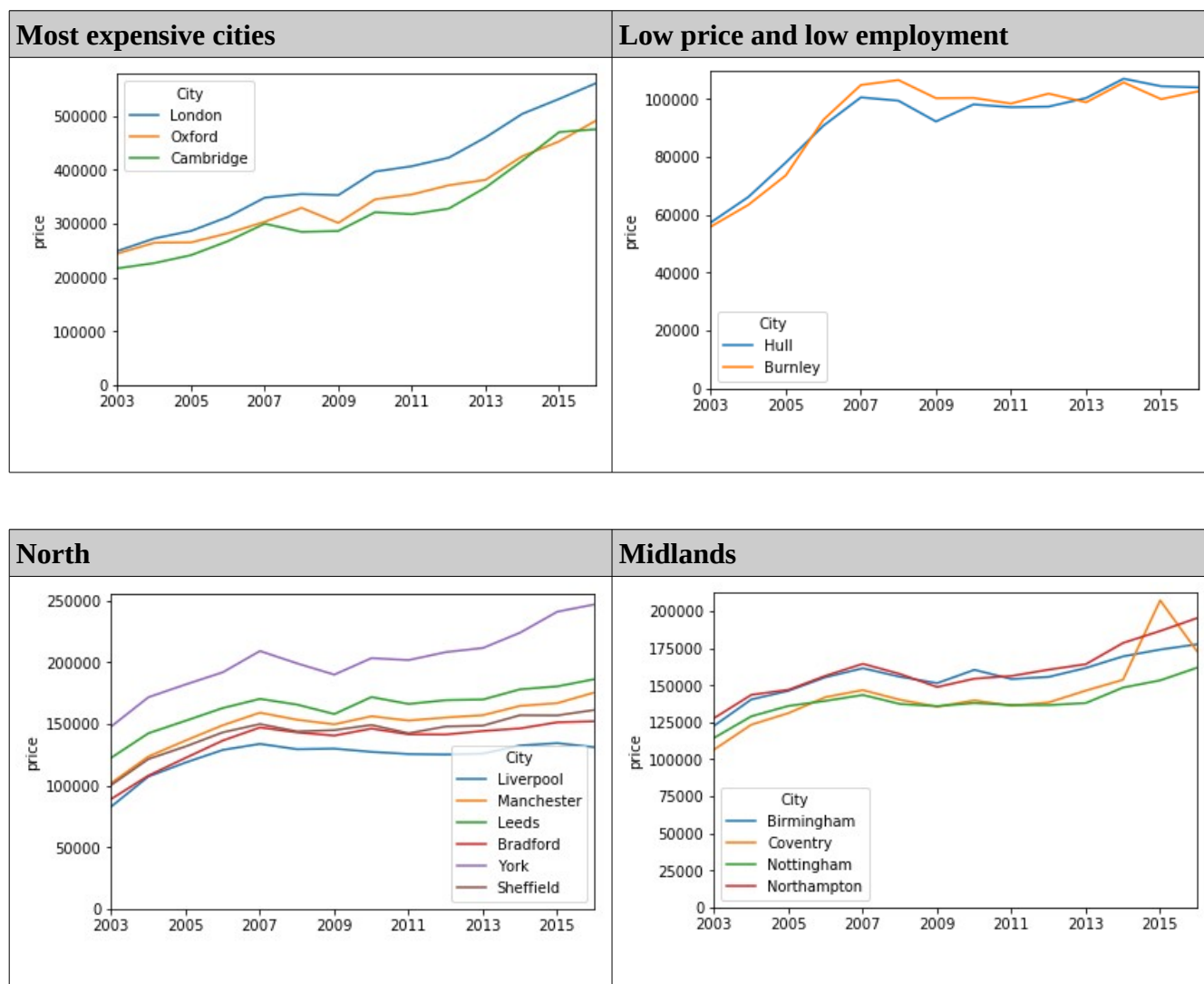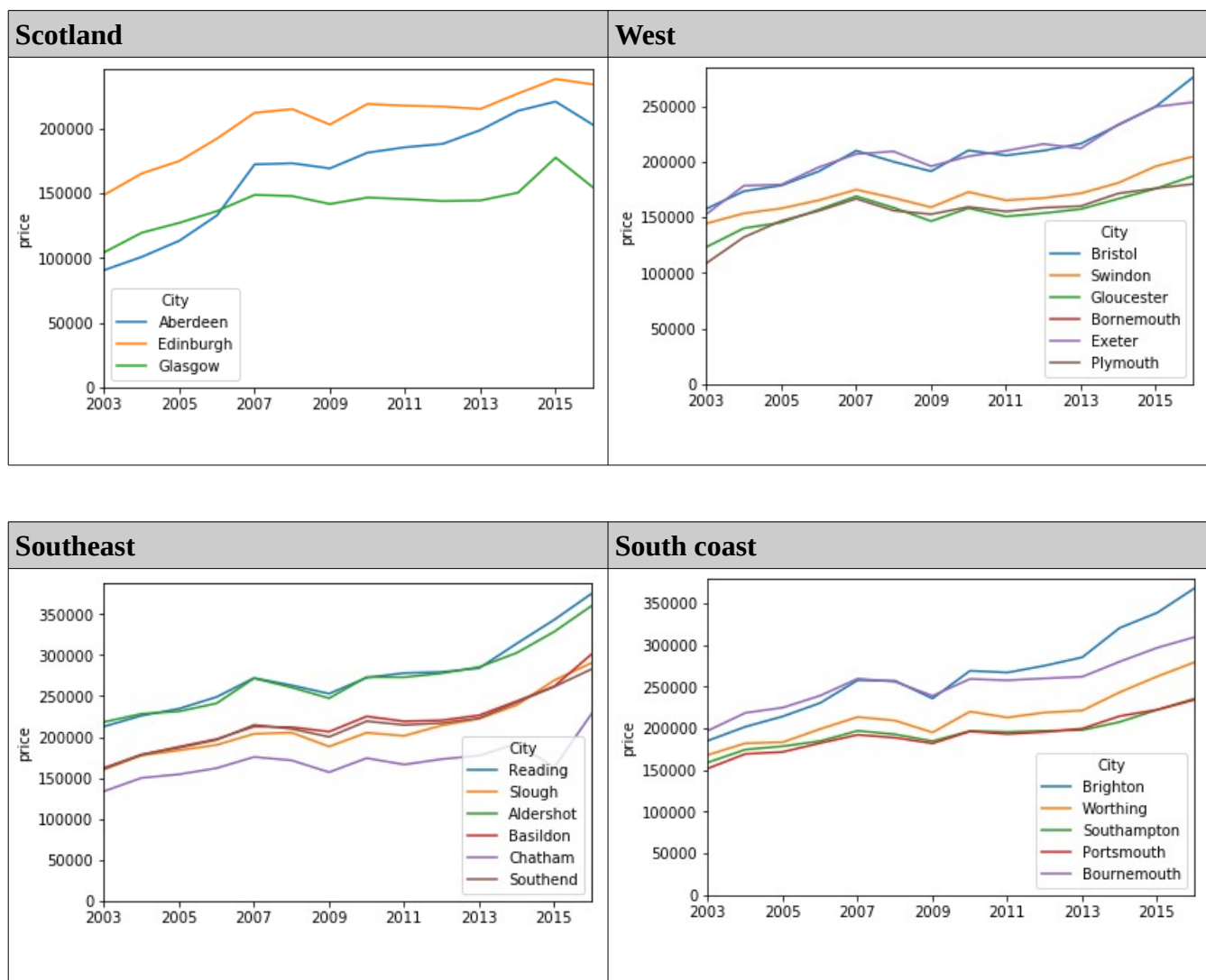
The purpose of this project is look beyond location parameters and explore city data in order to identify factors that have influenced observed house prices in the given period.

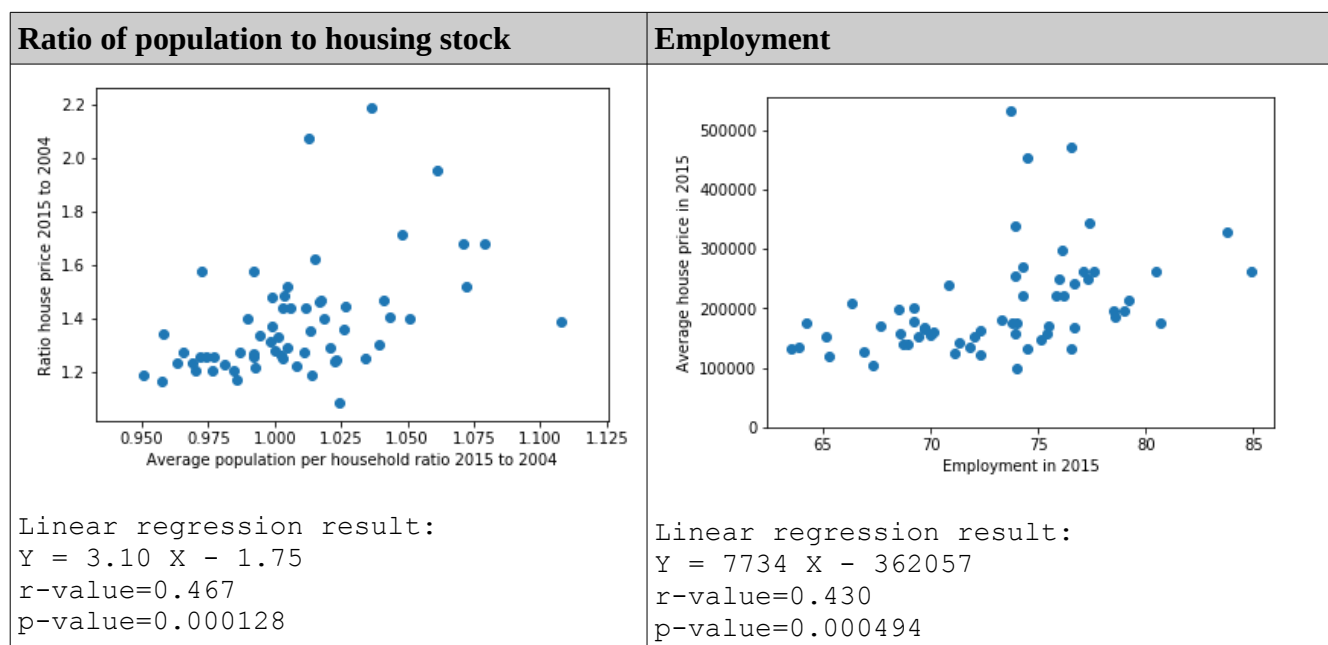# 5   Influence of economic / demographic factors

The influence of economic and demographic factors on house prices yields the following results:

- Wages do not have a statistically significant correlation to house prices

- The employment rate of a city correlates positively with house prices

- The ratio of population per unit of housing stock (

As seen from the chart below, there is no statistically significant relationship between wages and house prices, with a p-value of 0.92. Hence we can accept the null hypothesis and conclude that wages also do not have an effect on prices.
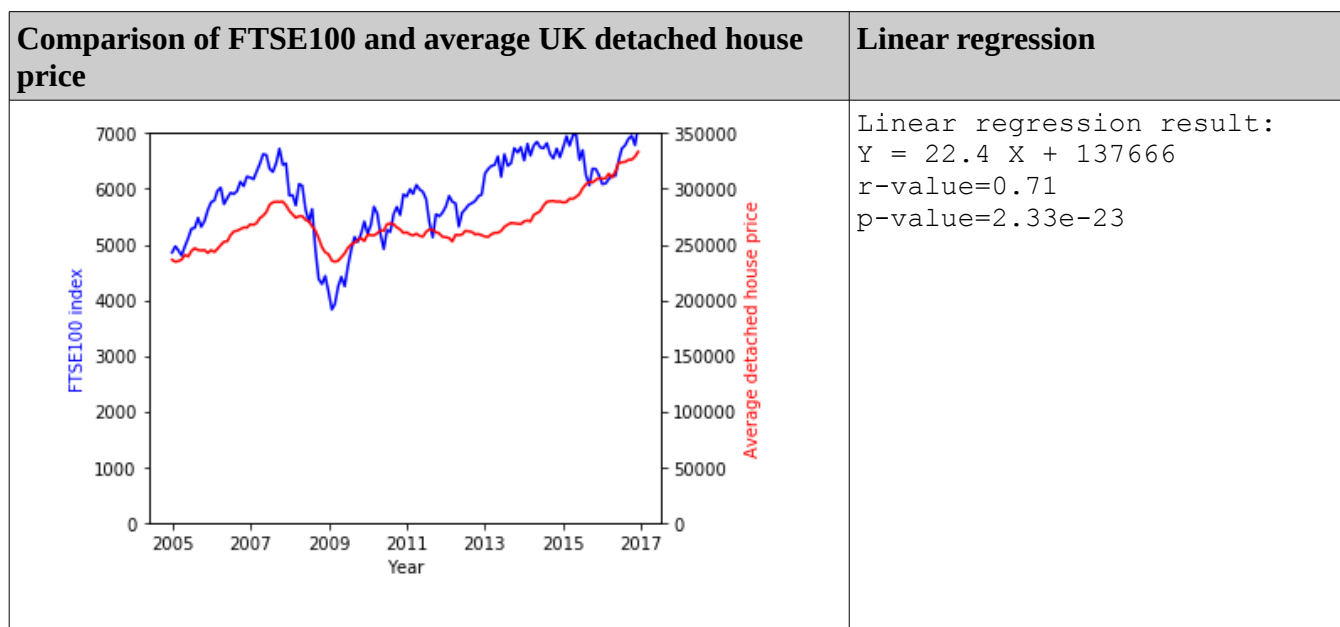
The results show a positive correlation, with a p-value of 0.0001, hence statistically significant to lead us to reject the null hypothesis. However the causality can be argued both ways, i.e. that a shortage of housing stock has the effect of pushing up prices, or alternatively that an increase in prices has forced a higher number of people to share a house on average.

The results for 2015 show that there is a positive correlation, with a p-value of 0.0005, hence statistically significant and we can reject the null hypothesis. The slope of the regression line corresponds to an average price increase of GBP 7,733 for each percentage point of higher employment. Apart from the three outliers that have average prices far above the average (London, Cambridge and Oxford), for the rest of the cities the employment rate is a significant factor in determining house prices.

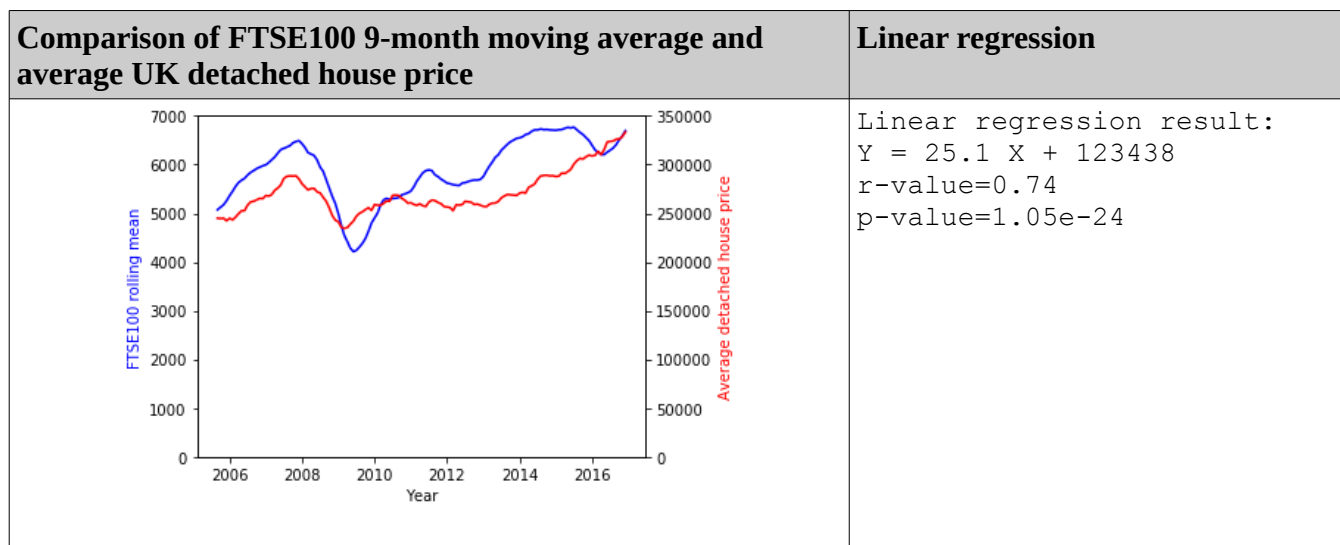| Ratio of population to housing stock	Employment
Linear regression result: Y = 3.10 X - 1.75 r-value=0.467 p-value=0.000128	Linear regression result: Y = 7734 X - 362057 r-value=0.430 p-value=0.000494

6   Correlation with the stock market

Further information on UK house prices is available from www.gov.uk, including the House Price Index (HPI) which tracks prices for different regions and different types of property (detached house, semi-detached house, terraced house and flat). This data provides an overall UK average, for each property type and the detached house UK average was used to compare prices with the FTSE100 index (noting that the four property types have 98-99% correlation between them so any of the types can be used).

Comparison of FTSE100 and average UK detached house price	Linear regression
	Linear regression result: Y = 22.4 X + 137666 r-value=0.71 p-value=2.33e-23

Next, a moving average of the FTSE100 was compared with house prices as house prices seem to react slower than the stock market index.

Comparison of FTSE100 9-month moving average and average UK detached house price	Linear regression
	Linear regression result: Y = 25.1 X + 123438 r-value=0.74 p-value=1.05e-24

Indeed, the FTSE100 9-month moving average provides a slightly closer correlation with the house prices, with a p-value of 0.74 vs. 0.71 before the moving average was taken into account.

7   Potential for further analysis

Additional factors may be considered for correlation with house prices, namely:

- Correlating with other investment assets including bond yields and the GBP exchange rate

- Influence of mortgage lending conditions

- Further exploration of the HPI data set which includes information on sales volume and cash vs. mortgage price

In addition, this data set includes cities and counties and therefore covers the entire UK territory.

8   Conclusion

The analysis leads to the following key take-away remarks:

1. House prices across the 62 major cities in the UK have risen by an average of 65% from 2003 to 2016.

2. As wages have on average remained at the same level, owning a house has become increasingly difficult to afford for the average worker, needing 8.0 annual wages to buy a house compared to 5.3 annual wages in 2003.

3. Population on its own does not correlate to house prices. However, population taken in combination with the housing stock in a city does have a statistically significant correlation with house prices.

4. Employment is an important factor in determining house prices in a city.

5. The 3 most expensive cities (London, Cambridge and Oxford) can be considered outliers as their prices are not explained by the factors covered in this analysis.

6. In the UK as a whole, house prices are strongly correlated with the FTSE100 stock market index and follow the same pattern, but with less intense reactions and with a lag behind the FTSE100 index. A 9-month rolling average of the FTSE100 index provides the closest correlation with the average house prices.