Predicting house price differences between London Boroughs

Alec Wilson - 15/01/2020

Contents:

1.1 Background

House prices in London are amongst the highest in the world, and, are the highest in the UK compared to other UK regions. Within London, however, there is also a large variation in the average price levels between the 32 London boroughs. Understanding the determinants of these differences is important as it enables firms and households to understand both how to price and value properties for future house building and when planning where to buy. An example of a determinant of house prices may be whether there are good schools nearby or strong transport links.

1.2 Problem

The data which will help solve the problem of determining house prices in various London boroughs will be descriptive data about the individual boroughs, e.g. the crime rate and number of outstanding schools, and, data which helps describe the quality of life in a particular borough.

1.3 Interest

A number of parties would be interested in the determinants for house prices between boroughs. These include commercial interests from sectors such as estate agents and house builders, who, once the determinants are understood, can use the insights to predict future trends.

2.1 Data Sources

The core data for house prices, crime levels, and population were sourced from 3 Kaggle data sets, link. Further data regarding the number of Ofsted (UK schools regulator) rated 'outstanding' primary schools was found from here and a commutability score, which assesses various factors including average cost to commute and average time commuting each way, was sourced from here. The data set for weekly pay was sourced from the ONS, here. Data used for exploratory data analysis was also sourced using the Foursquare API. The co-ordinates for the Boroughs needed to be scraped off Wikipedia.

2.2 Data Cleaning

Initially the datasets were imported from .csv files into the IBM Watson Studio environment. The data regarding the total crimes in each London borough was given for each month since late 2016. I decided to focus on the last 12 months of available data and then aggregated the last 12 months of crimes for each borough. The categories of crime were removed from the data frame, further analysis of the types of crime present would be interesting but were not the focus of this analysis. A crime rate per 1000 people in each borough was then created using the population dataset mentioned below.

The data for house prices included prices since January 1995. I was only interested in the latest 12 months available to begin cross-sectional analysis. The data needed to be transposed as the rows were made up of the dates and the columns individual boroughs. In order to merge the data frames together the boroughs were moved to an individual row and indexed. An average of the last 12 months of house prices was taken.

The population data columns were renamed to 'borough' and 'population' and the boroughs were indexed.

The co-ordinates for each borough were scraped off Wikipedia using Beautiful Soup and the latitudes and longitudes were decimalised in order for them to be used by the Foursquare API.

The data for number of outstanding schools and commutability score were found through the links above and extracted into a .csv file. The .csv file was then uploaded into the environment, the .csv file was already formatted in a consistent way. A 'good school' rate per 100,000 people was then created by dividing the number of outstanding schools by the population.

The Data for weekly pay was loaded in an already consistent format.

All of this data was then merged into one data frame which was now ready for data analysis.

2.3 Feature Selection

The dependent variable of this problem is the average house price of the individual boroughs.

Whilst cleaning the data a number of variables were created which were normalised for the population size of the individual boroughs. The variables for total crimes, number of outstanding schools and population were therefore not used when analysing the data.

The Foursquare API was used in the exploratory data analysis of the problem. An example of this was querying all of the train and underground stations in London and mapping these using Folium in order to learn more about the distribution of certain venues. As you move further into the centre of London for example there are more stations and there begins to be London Underground stations which are much more densely situated. This distribution was not however used in the analysis due to other variables such as the commutability score which takes this into account and provides a deeper insight than this particular query.

3.1 Relationship between average house prices and Crime rate

The relationship that one would expect to see from Crime rate and house prices is a negative one. Where an increase in the crime rate decreases the house prices. However, within the exploratory analysis the opposite was found. This was confusing as it is contradictory to what should be observed. The data was sourced reliably and checked for inconsistencies and none were observed. The reasoning behind this, even when all crimes but violent and sexual crimes were dropped from the data frame, must therefore be, that, the variability of crime rates inside certain boroughs is higher than between boroughs and different boroughs are not consistent with how many affluent neighbourhoods which are skewing house prices nearby there may be. Crime rate, although in general a helpful predictor for house prices, does not help to explain house price differences between boroughs.

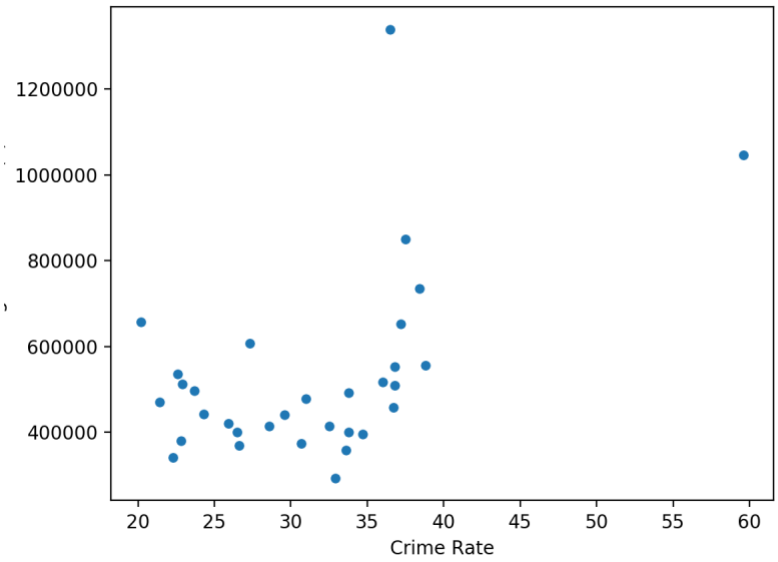

Figure 1 - Average house prices on y-axis

3.2 Relationship between average house prices and Commutability Score

There is a positive relationship between house prices and a borough's commutability score. The boroughs Commutability score is calculated by assessing various factors including time taken to commute into work and satisfaction with commute.

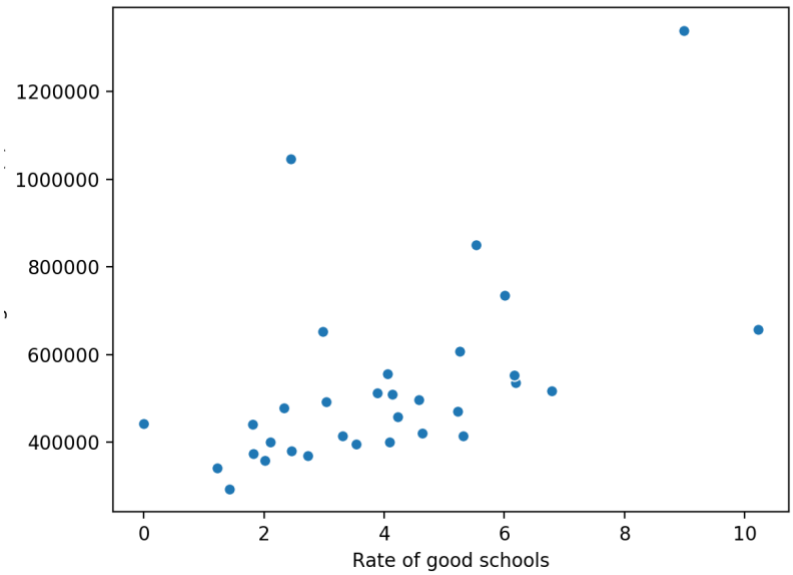

Figure 2 - Average house prices on y-axis

3.3 Relationship between average house prices and Rate of good schools

The relationship between rate of good schools and average house prices is positive. An increase in the rate of good schools leads to an increase in the average house price.

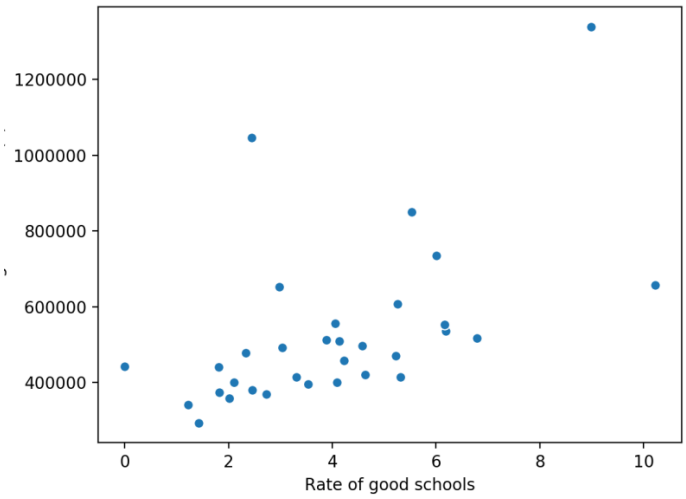

Figure 3 - Average house prices on y-axis

3.4 Relationship between average house prices and Pay

There is a strong positive correlation between average weekly pay and average house prices of boroughs. As can be shown on the scatter plot below.

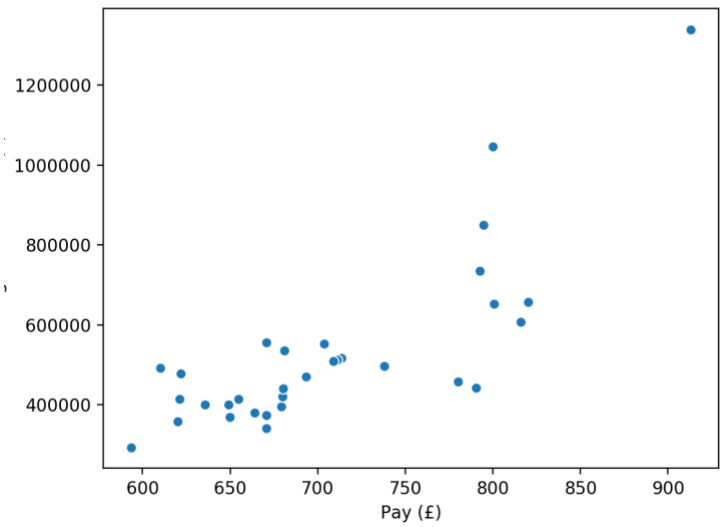

Figure 4 - Average house prices on y-axis

3.5 Foursquare API call results

A foursquare query for each of the 32 boroughs was used to help to understand the distribution of transport links in the City. The Map below clearly shows that as one moves further into the centre of the city the frequency of these transport links also increases. This may be due to the number of London terminals in the centre in addition to the large increase in London Underground services in the city centre. The south of London for example only has a handful of tube stations, of these most are situated very near the centre of the city.

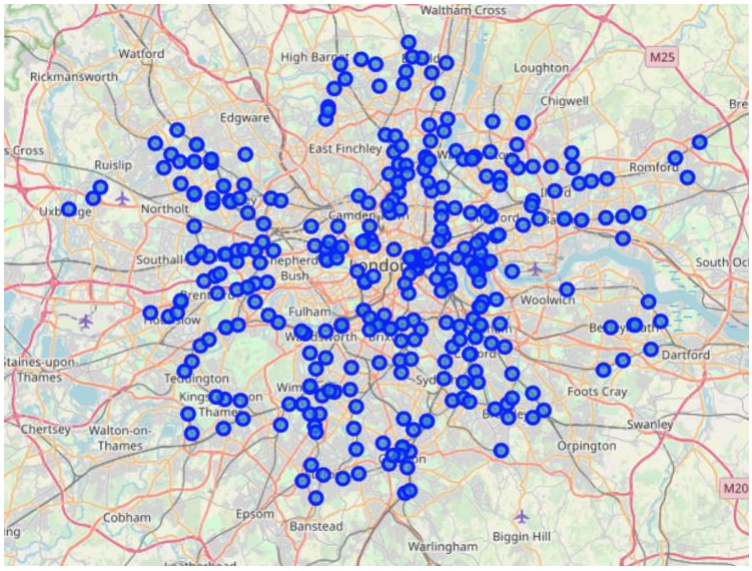

Figure 5 - Each marker shows one Train/Tube Station

3.6 Correlation Matrix

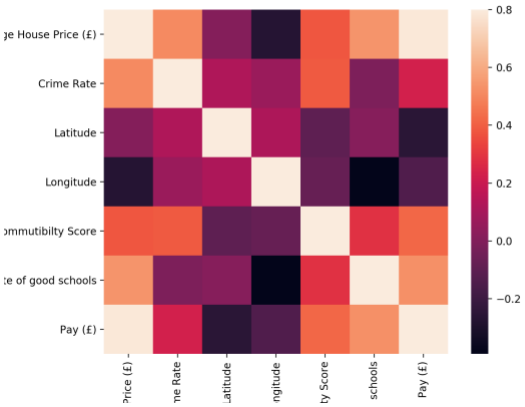

Figure 6

4.1 Regression Model

The multiple regression model used had a dependant variable of the average house price of the individual London boroughs. The independent variables used were: Commutability Score, Rate of good schools and weekly average pay (£). The Crime rate was not used as from the exploratory analysis it was evident that it was not a good predictor of average house prices between boroughs.

This was completed using the sklearn linear regression packages. A machine learning training and test model was used. The data in the model was normalised using the sklearn StandardScaler() function. The coefficients of the 3 independent variables are shown in the table below:

Commutability Score	Rate of good schools	Weekly average pay (£)
0.05146474	0.15506678	0.66172619

4.2 Model Evaluation

Average house prices within London boroughs were most responsive to an increase in weekly average pay than to the rate of good schools and the commutability score, although all of the variables showed a positive correlation.

The residual sum of squares for the regression was 0.04, a small value.

The explained variance score which helps identify how well the model fits the actual values was 0.92 with 1.00 representing a perfect prediction. This is a high value and shows that the model does well at predicting the actual values. This below graph visualises this. We can observe that the fitted values fit the actual values well.

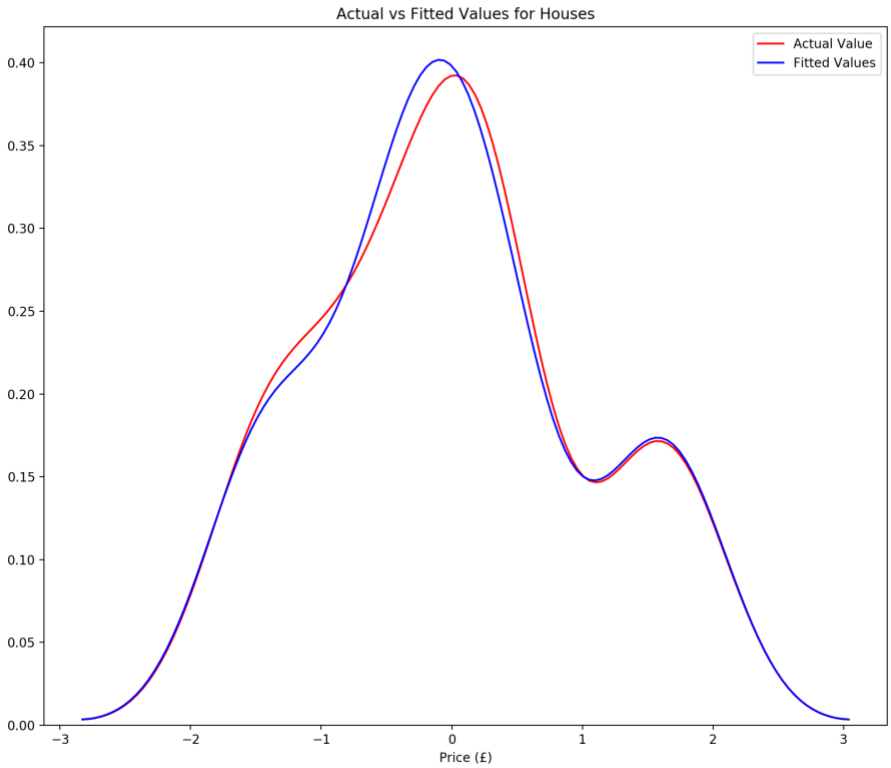

Figure 7

5. Conclusion

In conclusion, an increase in the commutability score and the rate of good schools in a particular borough does have a positive impact on house prices. However, in comparison, the average weekly pay within a borough was much more prevalent than the other independent variables at influencing the average house price. Therefore, future trends where we see a borough's average pay increase e.g. a number of young professionals moving into an area and setting down roots, especially in areas which were not previously populated with higher earners is likely to lead to future rises in house prices of the area.

6. Future Directions

In order to further look into this question an idea would be to break the areas focused down further, not just at the borough level but areas within boroughs. This would reduce the

variation within the areas that are being investigated and allow a more critical assessment of the features of the defined areas. There are however some problems with availability of data at this level and would require a large amount of data cleaning and wrangling. This would however load to more potent insights.