PREDICTING HOUSE PRICE DIFFERENCES BETWEEN LONDON BOROUGHS

ALEC WILSON

BACKGROUND

- London has amongst the highest house prices in the world.

- There is a large variation in house prices between 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.

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.

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.

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.

DATA CLEANING

- The data was wrangled from various data sets and was cleaned to make 'borough' the index column.
- The data was then merged into one data frame.

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.

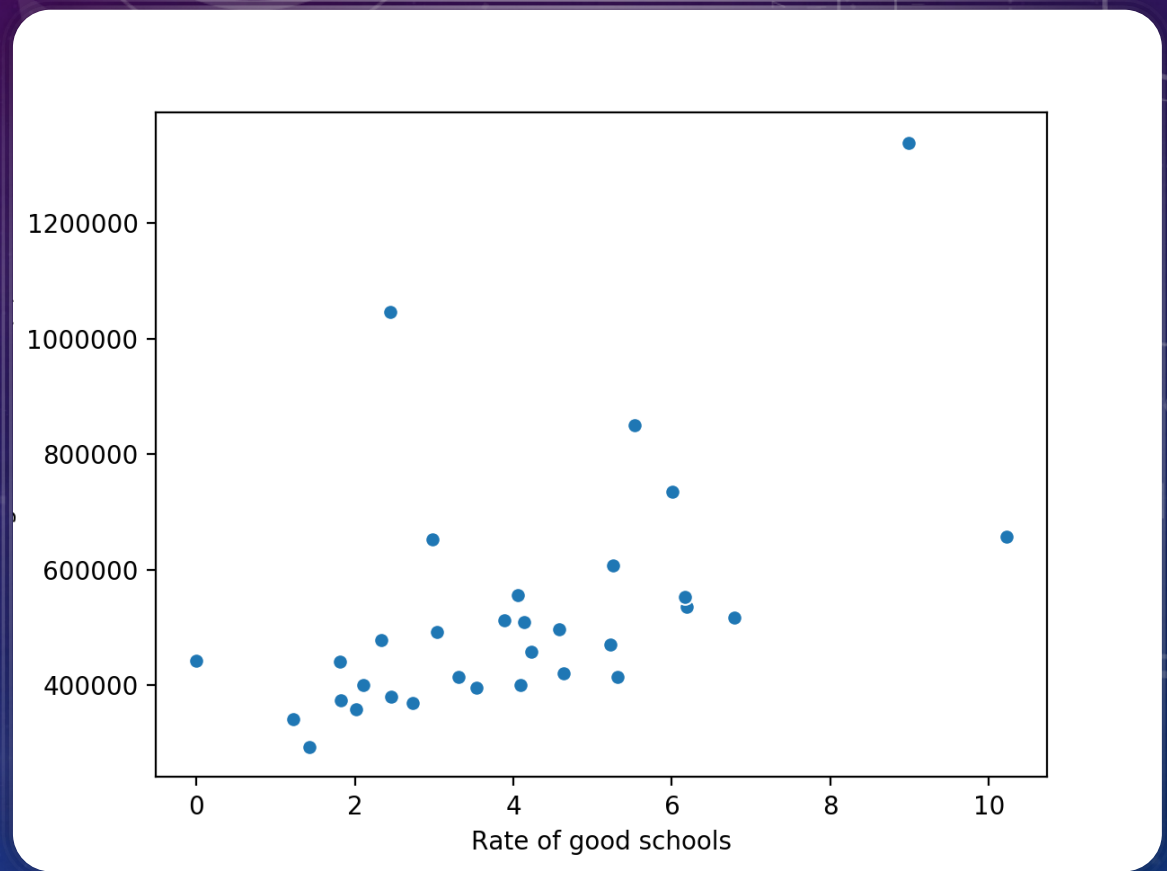- Foursquare API was used in the exploratory analysis of the project.

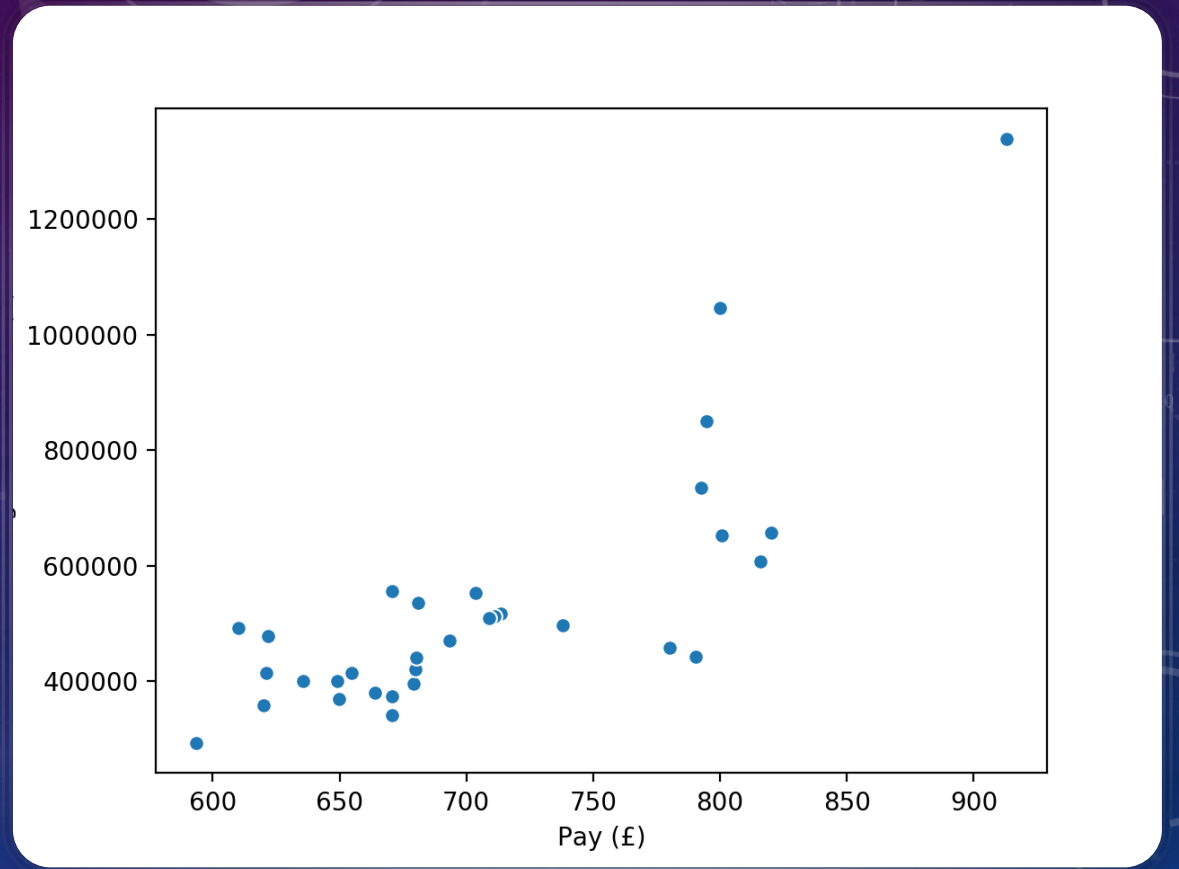RELATIONSHIP BETWEEN AVERAGE HOUSE PRICES AND CRIME RATE
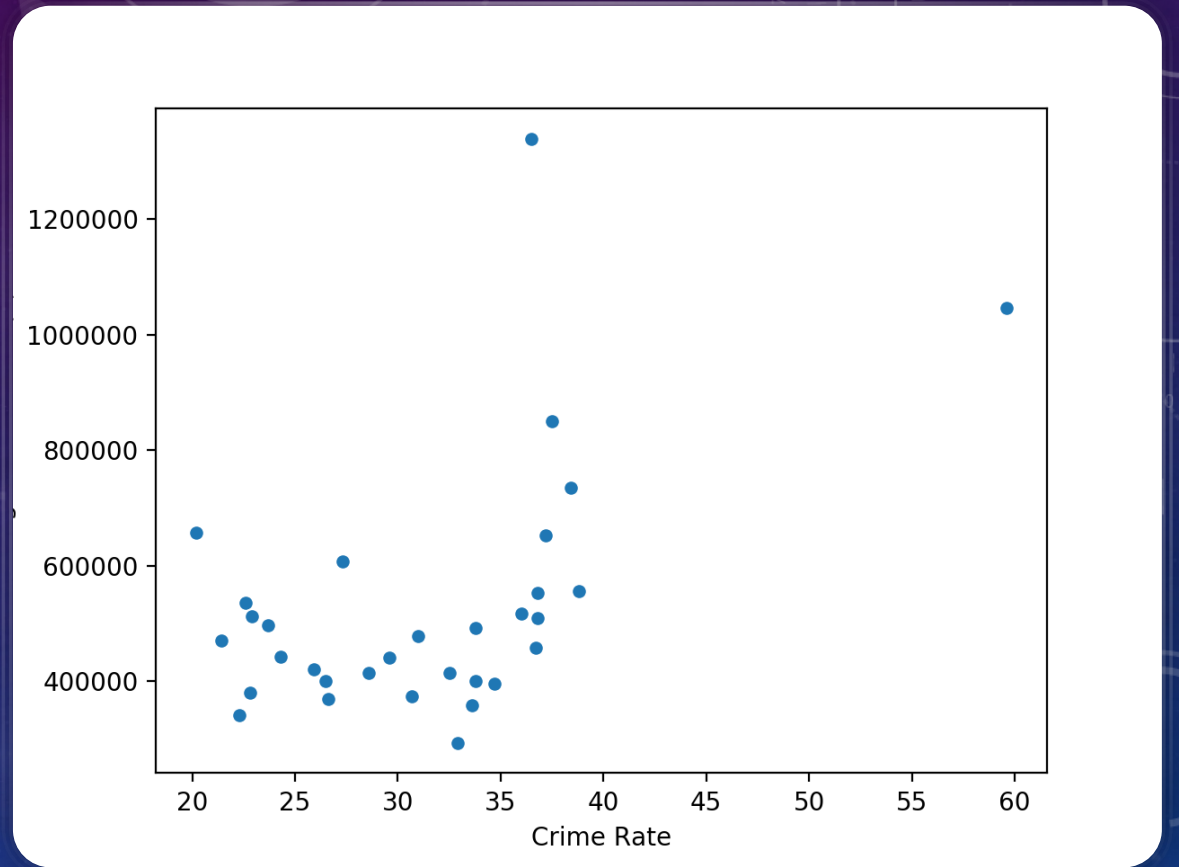
RELATIONSHIP BETWEEN AVERAGE HOUSE PRICES AND COMMUTABILITY SCORE
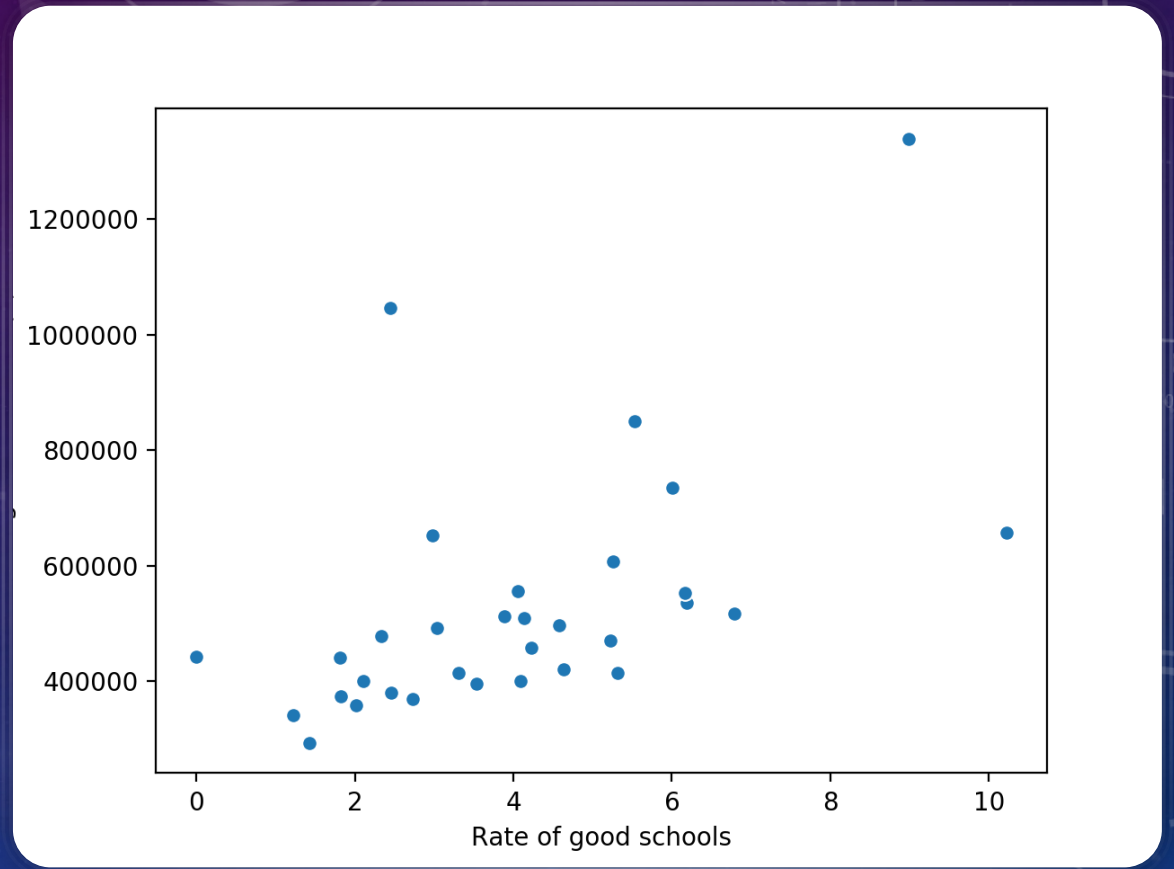
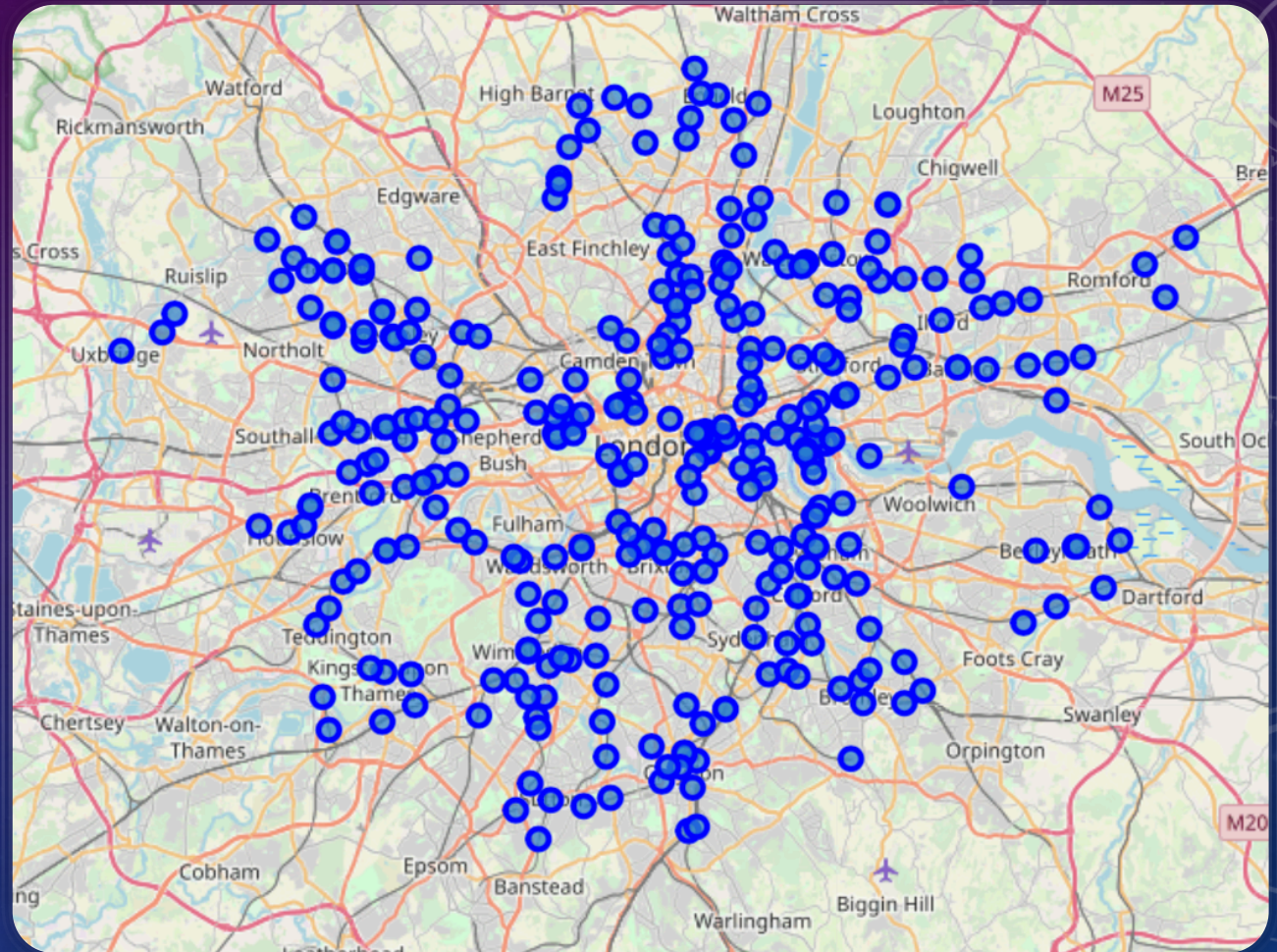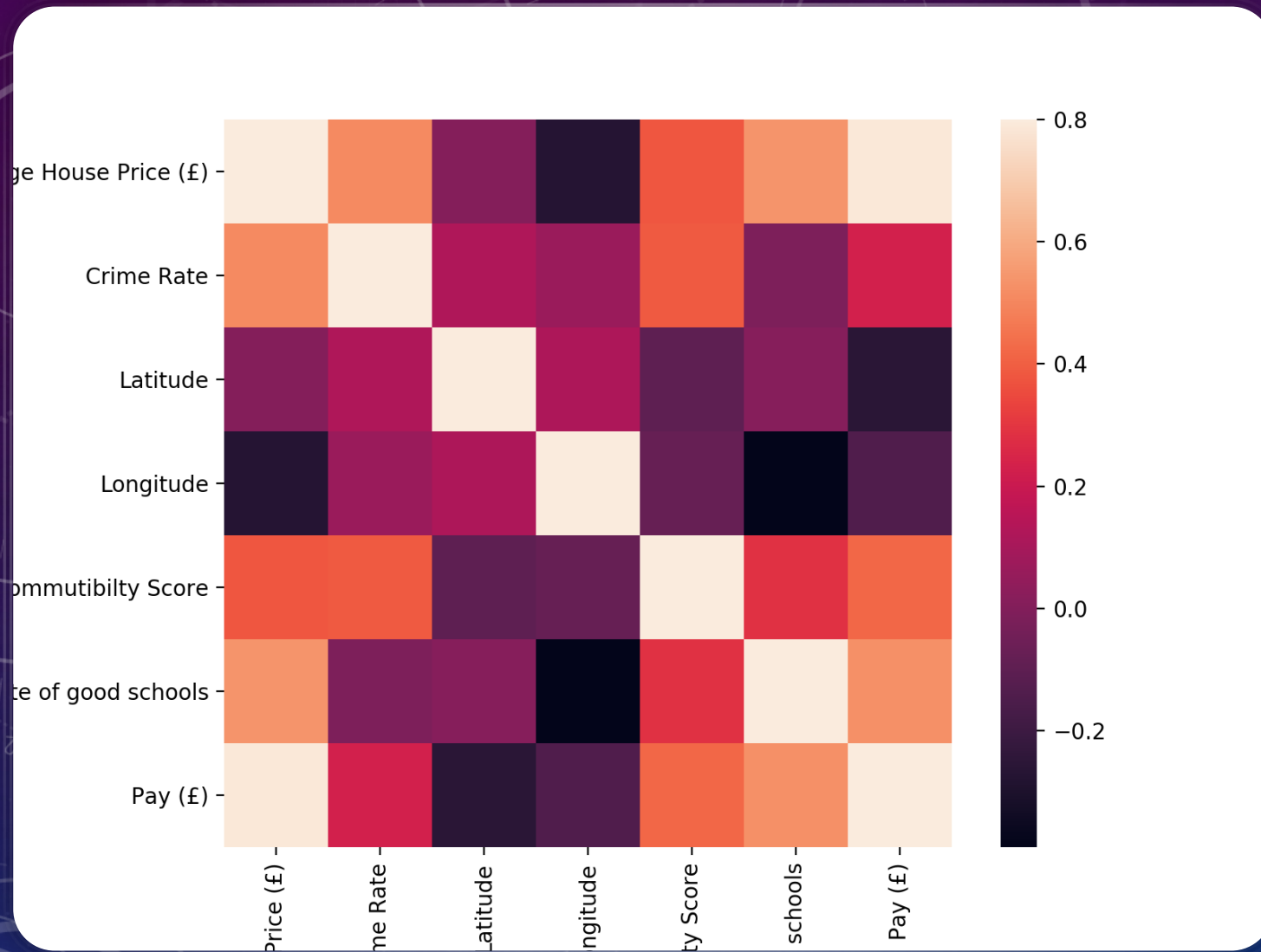RELATIONSHIP BETWEEN AVERAGE HOUSE PRICES AND RATE OF GOOD SCHOOLS

RELATIONSHIP BETWEEN AVERAGE HOUSE PRICES AND PAY

FOURSQUARE API

The Map clearly shows that as one moves further into the centre of the city the frequency of these transport links also increases.

CORRELATION MATRIX

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

MODEL EVALUATION

- 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.
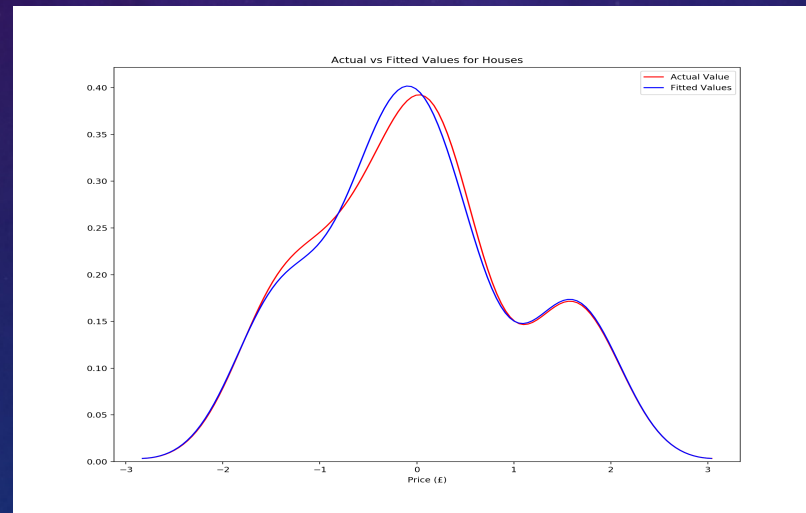
CONCLUSION

- 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.

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.