What are the most important factors for student housing prices in York?

Introduction

Housing is one of the largest costs for students during their time at university, so the rental price is a major concern in where students chose to live. The aim of this project is to investigate what factors have the largest influence on the price of student rental properties in York, using housing data provided by Student House Finder. These results will give insight to both students and housing providers into what to look for to identify price expectations.

Methodology

1 **Data Preparation**

I first chose the important data which could impact housing price and isolated this into a new data frame. This data included:

- The number of bills included.
- The number of bedrooms.
- The source/agency listing the property.
- The proximity to important city and university landmarks.

2 **Model selection**

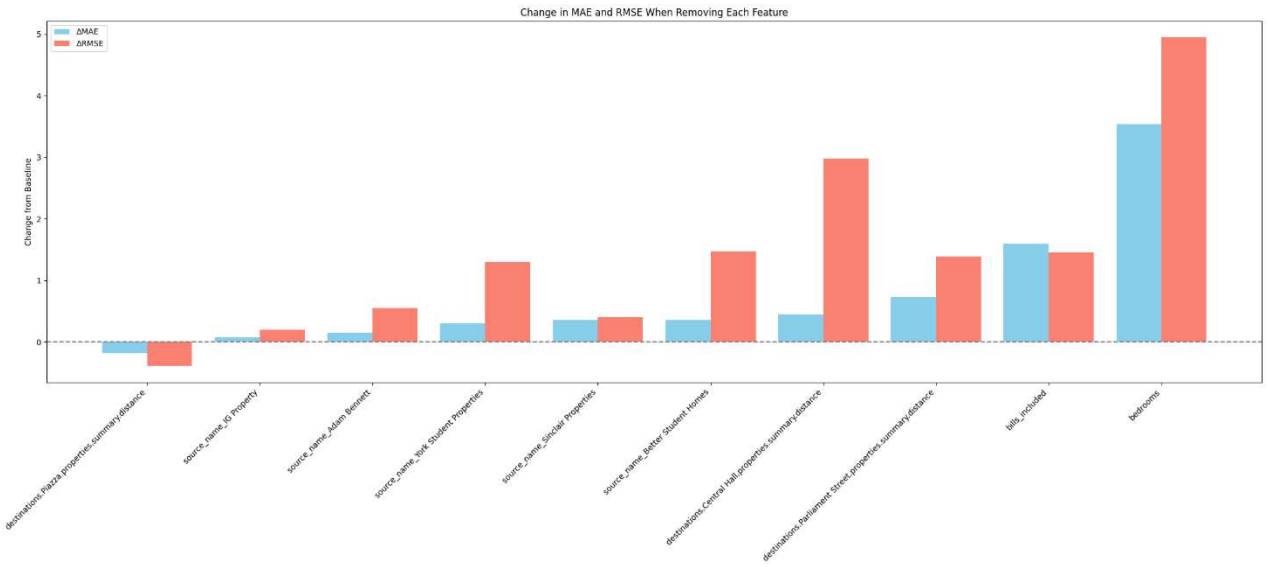
To predict the housing price from these features I used a random forest model due to its good balance of performance and robustness. The base model was trained on all the features and was evaluated using mean absolute error (MAE) and root mean squared error (RMSE), to find performance at both normal and extreme values.

3 **Feature Ablation**

To find the importance of each feature I used a “Leave-One-Feature-Out” approach, where I retrained the model and omitted a different feature each time. Comparing the MAE and RMSE against the base model (figure 1), we see how removing each feature affects the performance of the model and hence, how important it is to predicting the price.

Analysis

Figure 1:



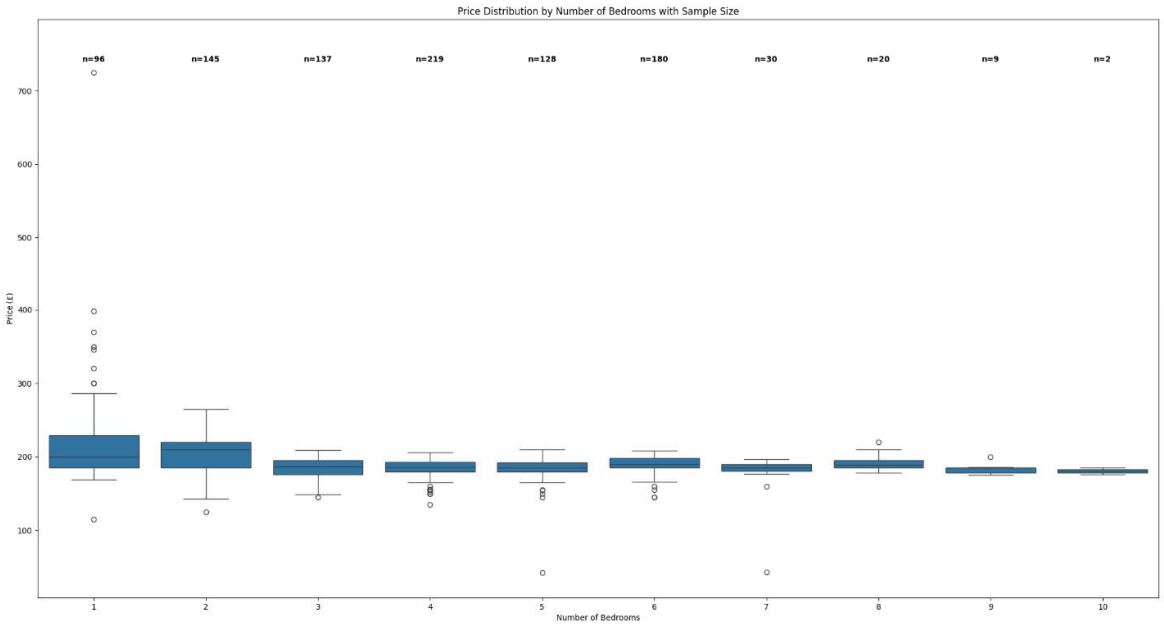
This figure shows the result of feature ablation. The most unexpected result is the negative values for MAE and RMSE in the distance to the Piazza building, which shows that the model improves when this feature is removed. This implies that this distance is completely irrelevant to the prices, so much so that including it only add unnecessary noise, decreasing performance.

Another interesting result is the extreme importance of the number of bedrooms show in figure 1. To check how this feature affects the price I tested a linear regression model against the price.

Metric	Value	Interpretation
Coefficient	-4.06	For each additional bedroom the price decreases by £4.06.
P-value	0.000	This shows the relationship is statistically significant as $p < 0.05$, so unlikely due to chance.
R-squared	0.067	This shows that the regression is very weak, and other features may play a larger role, counter to what's implied by the feature ablation test.

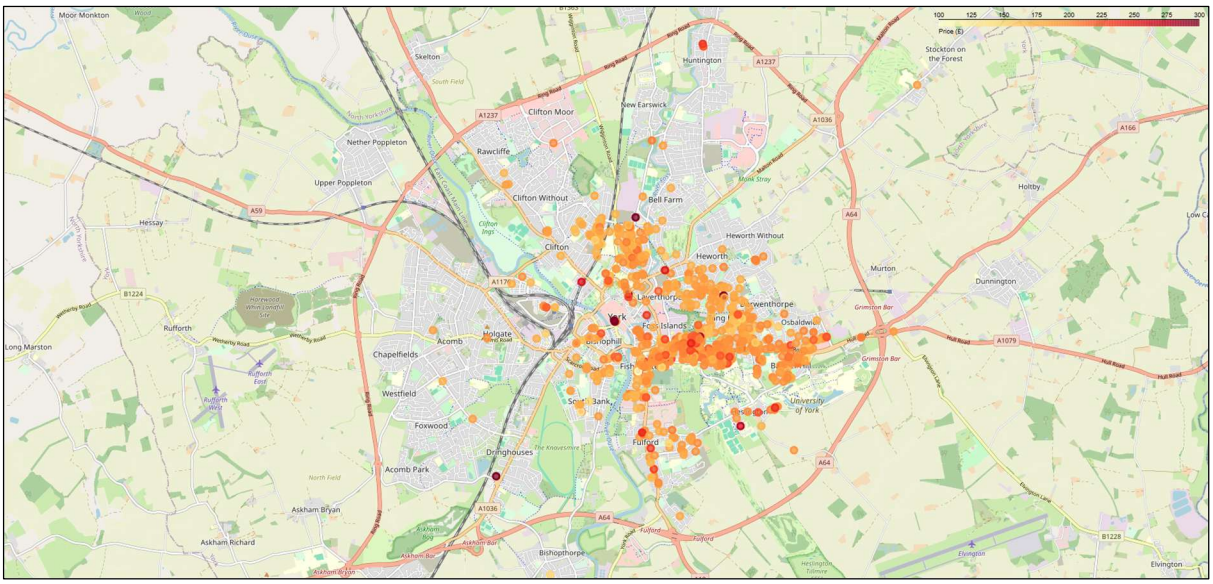
To determine why the feature ablation shows that the number of bedrooms is so important, we can use box plots for each number of bedrooms (figure 2). This shows how both the inter-quartile and extreme ranges of price decrease as the bedrooms increases. So, with larger numbers of bedrooms, the price is much more predictable, while at lower numbers, there are other features that are much more important.

Figure 2:



The next most significant feature is the distance to both Parliament Street and Central Hall, shown to be extremely important on the extremes. We can visualise this better by mapping the house prices in the city (figure 3).

Figure 3:



(note: I had to set a fixed range for prices as the few outliers, 100- or 300+, made points all one colour)

As expected, the properties near parliament street (town centre) are shown to be much more expensive. Additionally, we see that the density of highly priced houses is situated about halfway between Parliament Street and Central Hall. Outside this area, prices are more consistent, outside of a few exceptions.

Conclusion

The most important feature in predicting the price of houses is found the be the number of bedrooms. However, this is only relevant for houses with larger numbers of bedrooms, where other features are much more important in houses with less. The feature that shows the largest effect in this regard is the distance to Parliament Street and Central Hall.

As a result of these findings, students looking at lower house prices should firstly look for properties with more rooms. The next thing they should decide is what they want to prioritize most with location. Students whose subjects are located on east campus are in luck, as finding a property close by does not affect price at all. Otherwise, they need to look at houses which are further away from Parliament Street or Central Hall.

(I should also note that in figure 1 whether the bills are included is also shown to be an important feature of price. This is expected as the prices listed would not only include the price of the rent itself but the bills as well.)