Seattle Real Estate Analysis

SHORT STUDY ON THE OPPORTUNITIES OF THE REAL ESTATE MARKET IN SEATTLE

Executive Summary

- The objective of the analysis conducted is to understand which are the features that have a bigger impact on the increment of the value of the houses in Seattle and consequently their price.

- I Formulated the following 3 Business Questions that can guide us into the analysis :

1 - As grade and bathrooms show multicollinearity, which is the most significant variable in predicting the price?

2 - Does having a waterfront in the house has a positive impact in the price?

3- How does compare the houses built before 1975 and after 1975 upon the price?

Analysis Timeline

STEP 1 : LOAD THE DATA SET AND IDENTIFY NULL VALUES

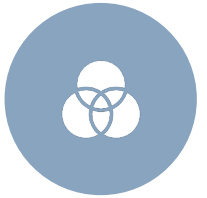
STEP 2: TREATMENT OF NULL VALUES

STEP 3: EXPLORATORY ANALYSIS/ DATA OBSERVATIONS

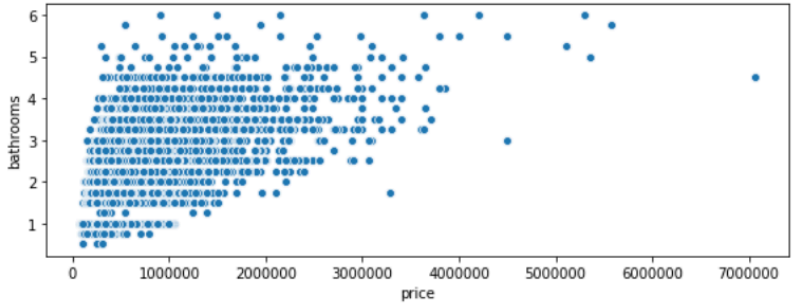
STEP 4: GRAPH RELATIONSHIP ALL THE FEATURES AND THE PRICE

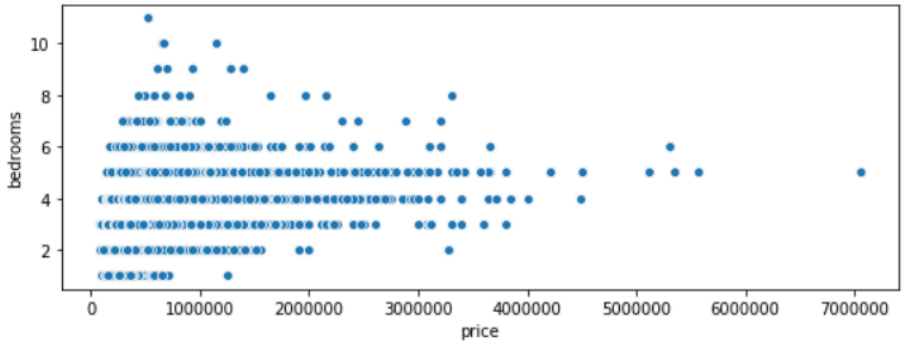
STEP 5: MODEL APPROACH

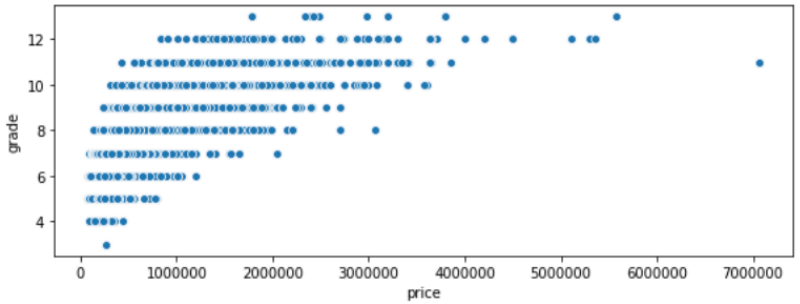
STEP 6: CHOOSE VARIABLES

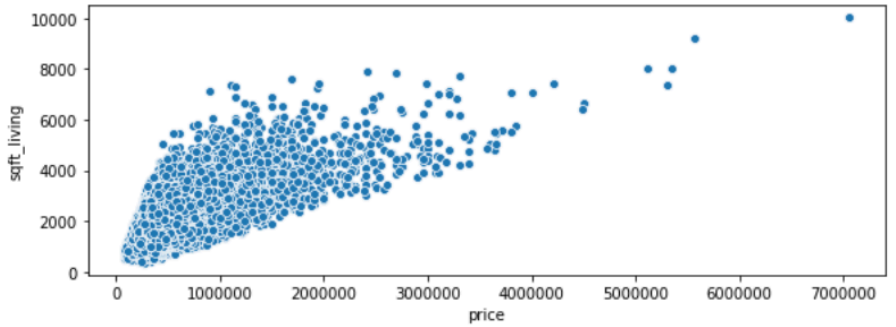


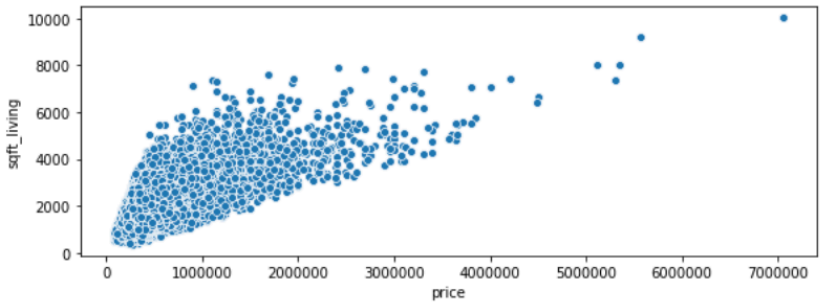
STEP 7: INTERPRETATION OF THE MODELS

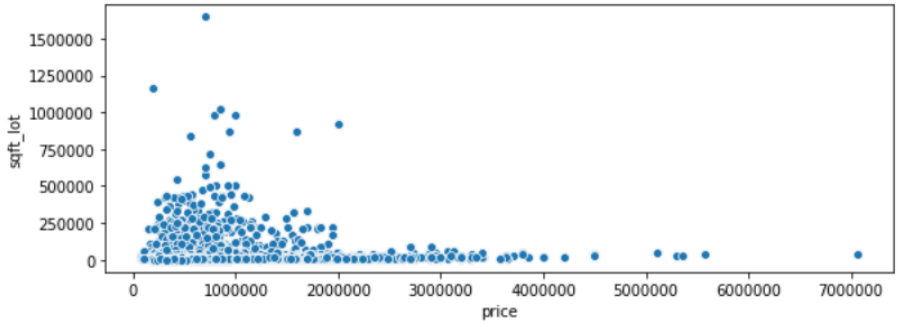






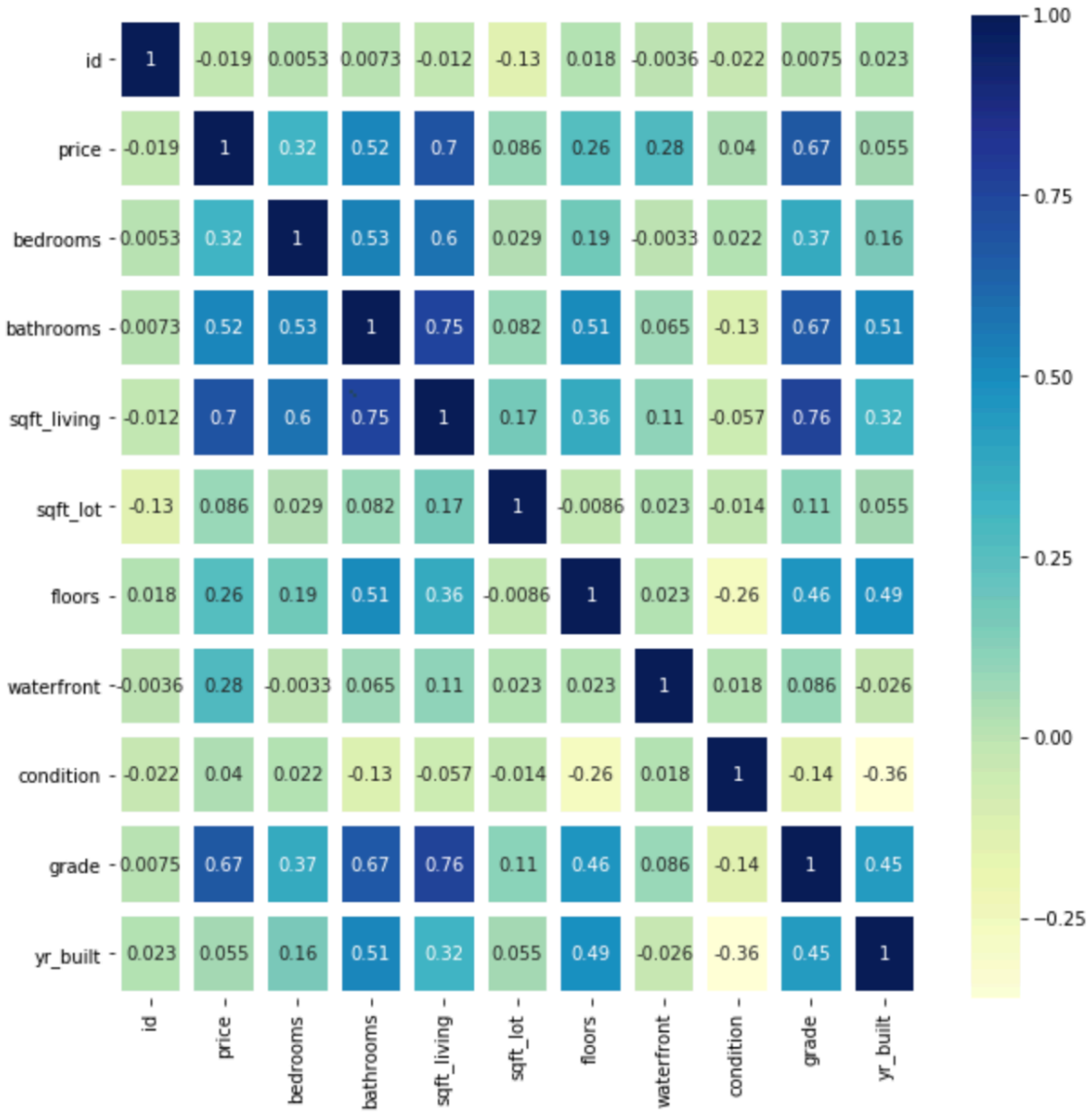






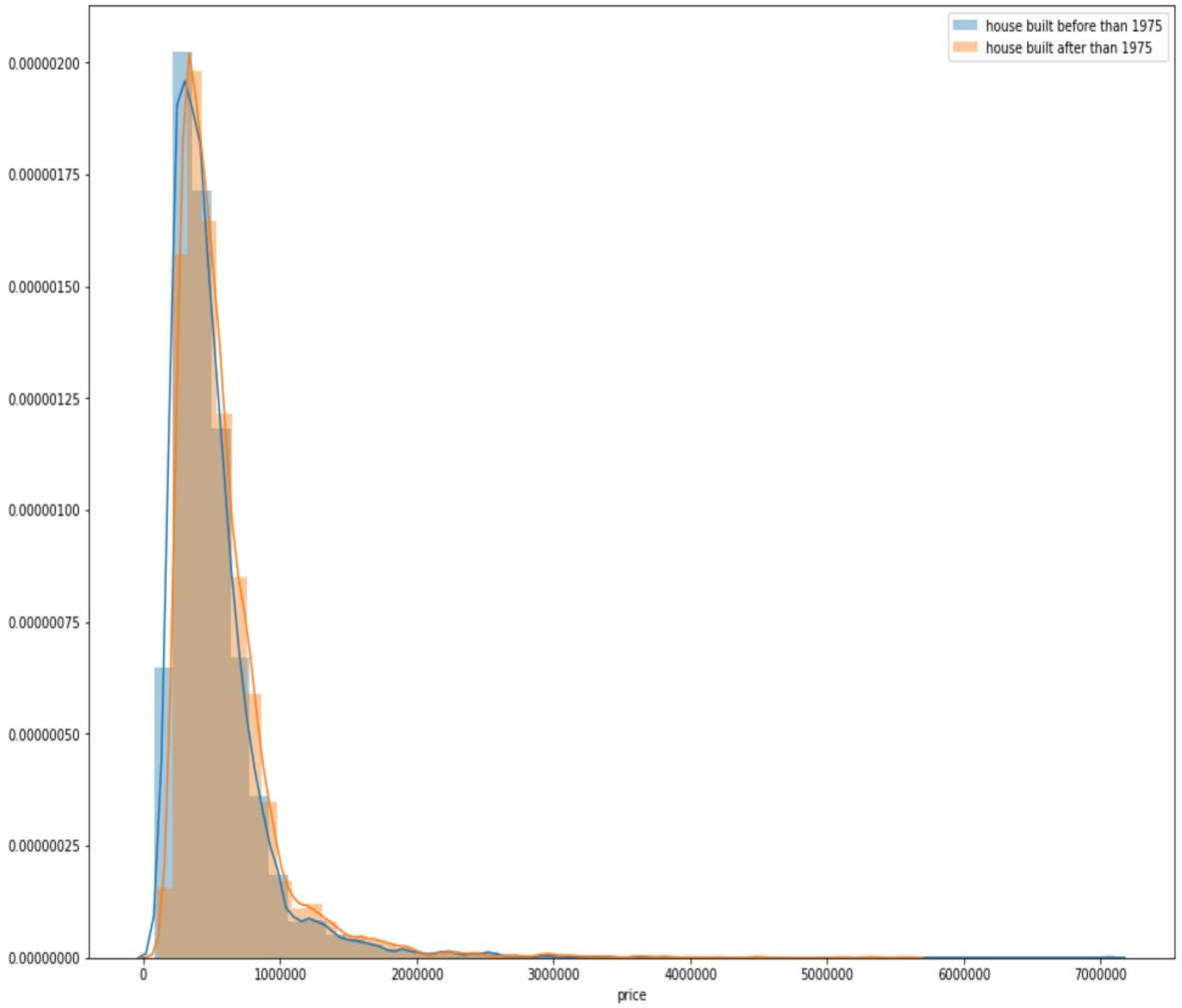
Graph relationship between all relevant variables and price

- I tested most variables in order to have a prelaminar idea about their relationship with the price and identify the relevant ones for the model testing



Build correlation table and matrix to get an overview of the relationship between the variables.

- From the heatmap above I can see that this are the features with the highest correlation : grade, waterfront, floors,sqft_living,bathrooms, bedrooms
- By checking the correlation matrix I see that the sqfliving has a strong collinearity with bedrooms, bathrooms and grade
- I can seen as well some multicollinearity(0,67) between grade and bathrooms



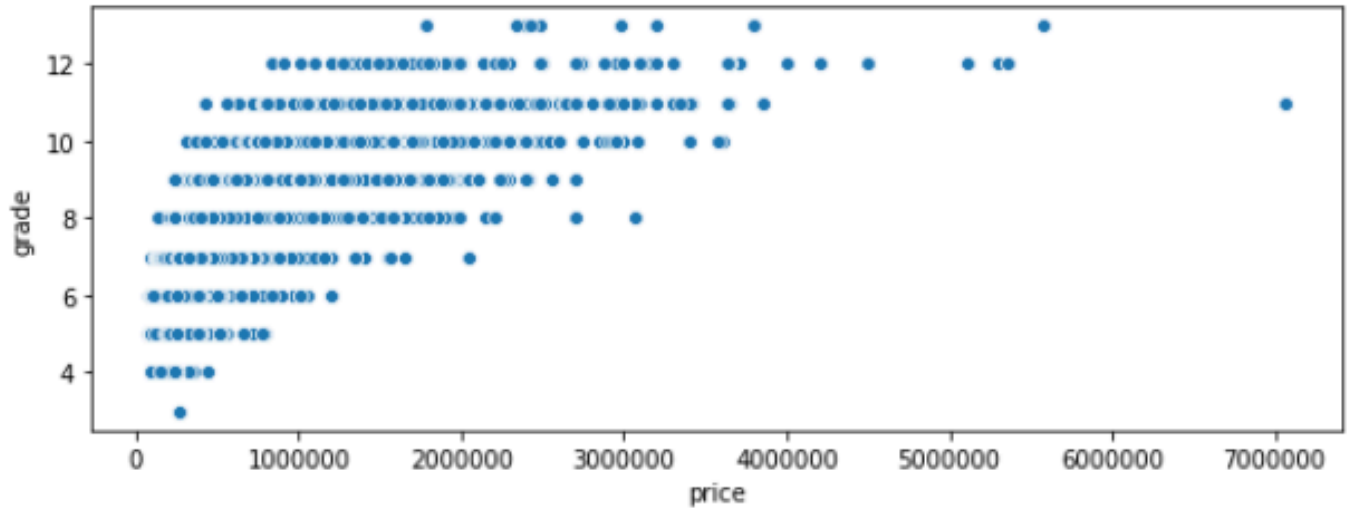
1- How does compare the houses built before 1975 and after 1975 upon the price?

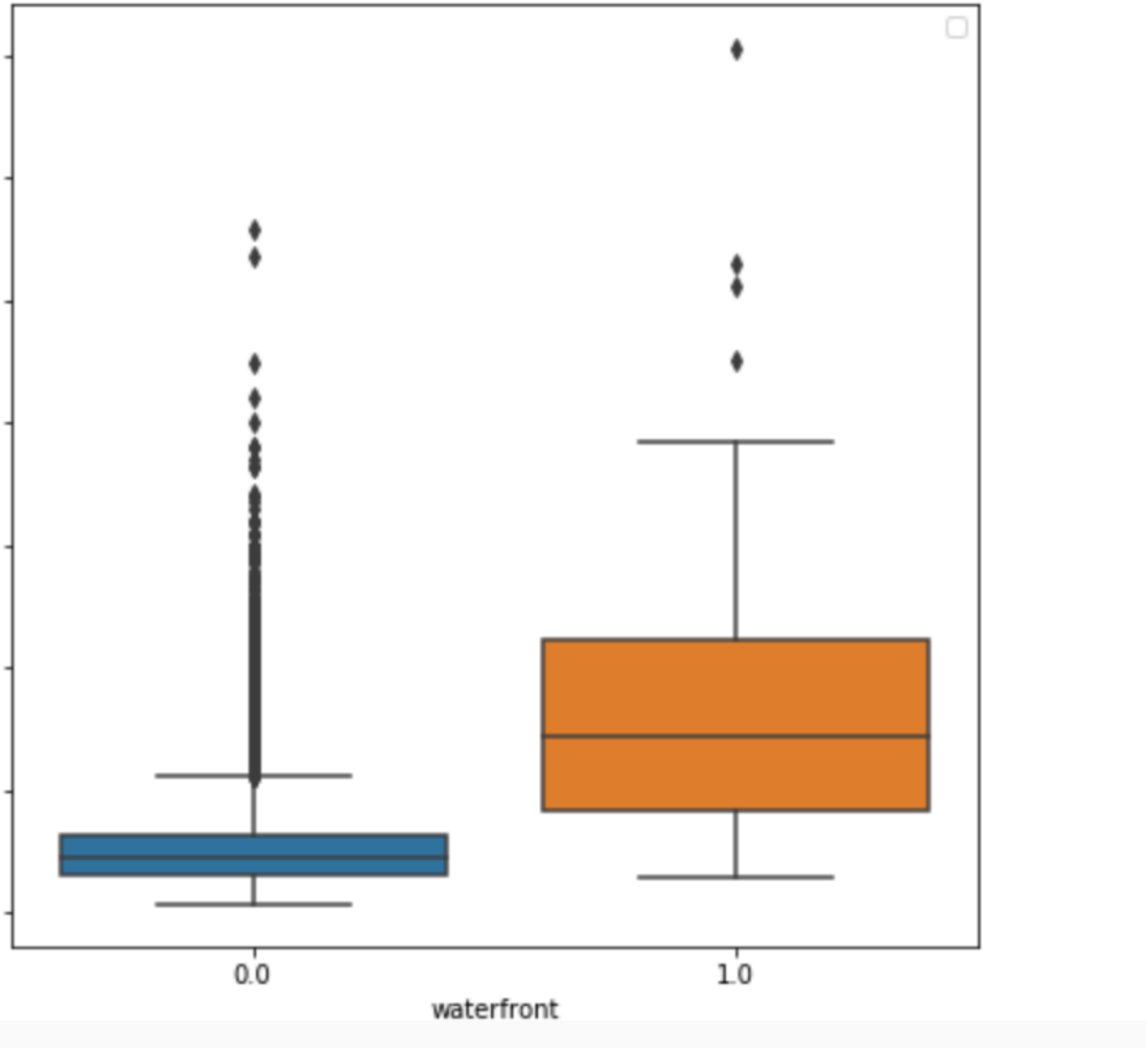
As the graph is showing, it seems there the houses built after 1975 are slightly more expensive but the difference is not much.

The reason of this difference being so small could be due to the **survival bias** on the houses built before 1975. This because the ones that are still in use are really likely to be the best ones built before 1975 by surviving after 50 years.

2- As grade and bathrooms show multicollinearity, which is the most significant variable in predicting the price?

- Following my statistical analysis I find out that the most significant variable in predicting the price is the variable GRADE.





3- Does having a waterfront in the house has a positive impact in the price?

THE ANSWER OF THIS QUESTION IS YES, AS YOU CAN SEE CLEARLY FROM THE BOXPLOT, HAVING A WATERFRONT IN THE HOUSE DOES HAVE A A POSITIVE IMPACT ON THE PRICE.

Conclusions

- Develop more business questions the help the real estate market to maximize the profit
- Find new variables that can be significant for an increasement in the price of the houses
- Check the goodness of the models used and try to improve it

Thank you for listening!