King County Property Analysis

King County

Scott Graham
5th June, 2022

Summary

The goal of this analysis is to review data from King County of property sale prices in 2014 and 2015 and provide insight to prospective buyers and sellers what property features will provide the best sale price.

Outline

- Business Problem and Hypothesis

- Data Understanding

- Data Cleaning

- Regression Analysis

- Conclusion

Business Problem And Hypothesis

Provide insight to buyers and sellers

Property features are most effective to achieve greater sale price

Null Hypothesis - There is no key indicators that any property features will increase sales price
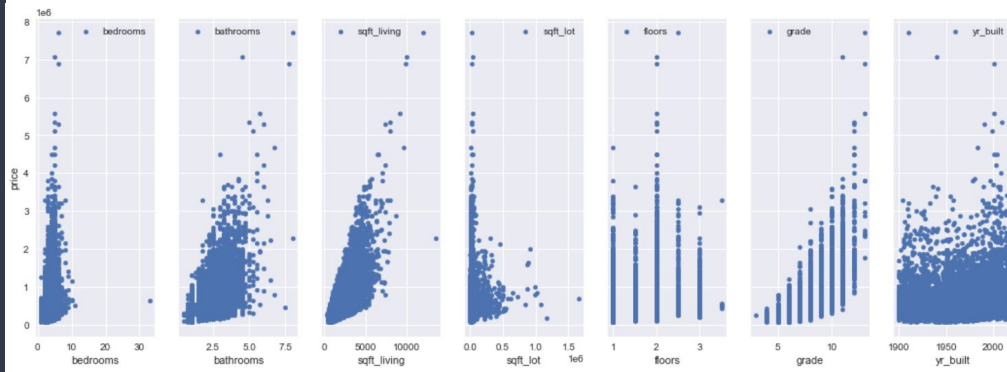
Alternative Hypothesis - There are indicators that certain features of a property will provide a higher chance of increased sales price
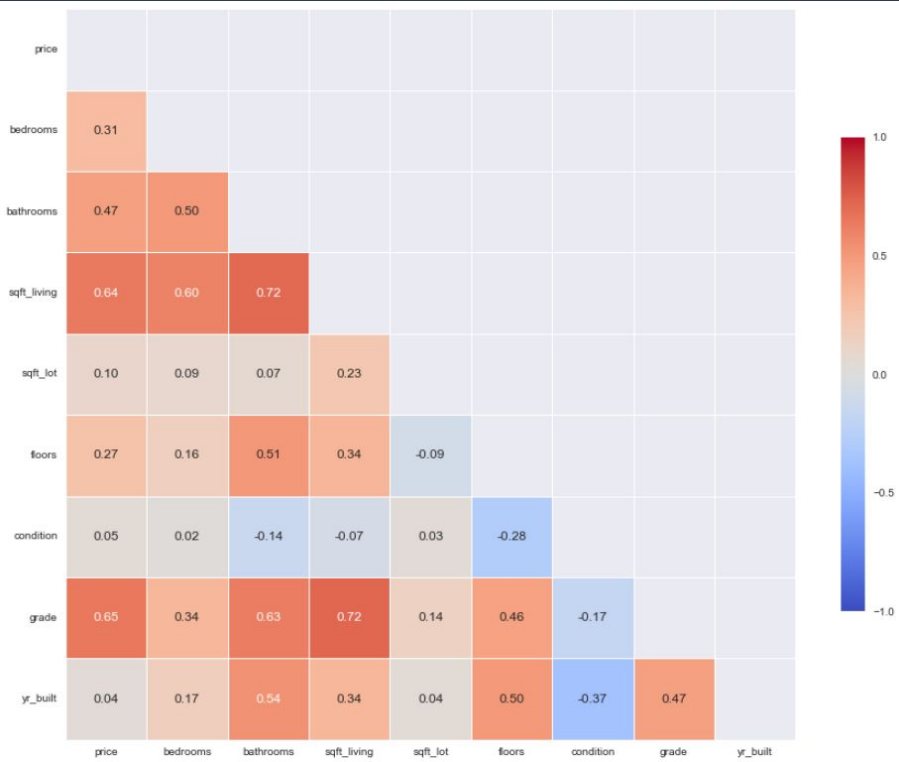
Data Understanding

Dataset including 21,500+ rows and 21 columns

Removal of 10 columns based on the scope of this analysis

Base model assessment using Histograms, scatter plots and KDE

Data Cleaning

Removal of outliers using Z-Score

Perform log function on base model price distribution

Performed Correlation Matrix to further analyse the correlation between datasets and further refine the list
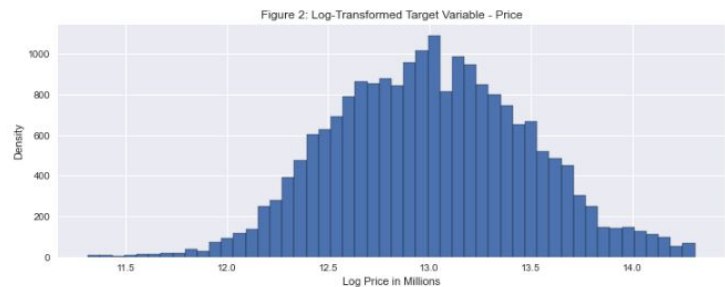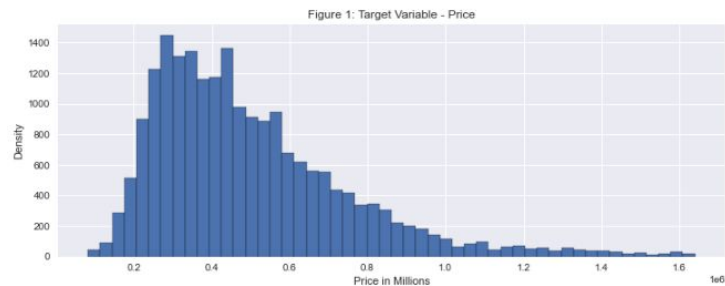
Regression Analysis

Train-Test-Split done with 70/30 split

Training and test residuals calculated

Train and Test Mean Square Error converged

Figure 1: Target Variable - Price

Figure 2: Log-Transformed Target Variable - Price

Conclusions

Bigger is better!

Mean Square Error:

Train Mean Squarred Error: 0.47860656574948161

Test Mean Squarred Error: 0.4852373857492091

Null Hypothesis rejected

Factors such as area of living space, bedrooms, bathrooms, etc play a role in the sale price of property

Bigger is better!

Thank You!

Email: scottgraham14@gmail.com
GitHub: @scottgraham1
LinkedIn: linkedin.com/in/scott-graham/