King County Housing Model

Jordan Jones
Gabe Warner

Summary

- Analysis of the housing prices in King County to make a predictive model to estimate the cost of a house based off multiple features.

Outline

- Business Problem
- Data & Methods
- Results
- Final Model
- Conclusion

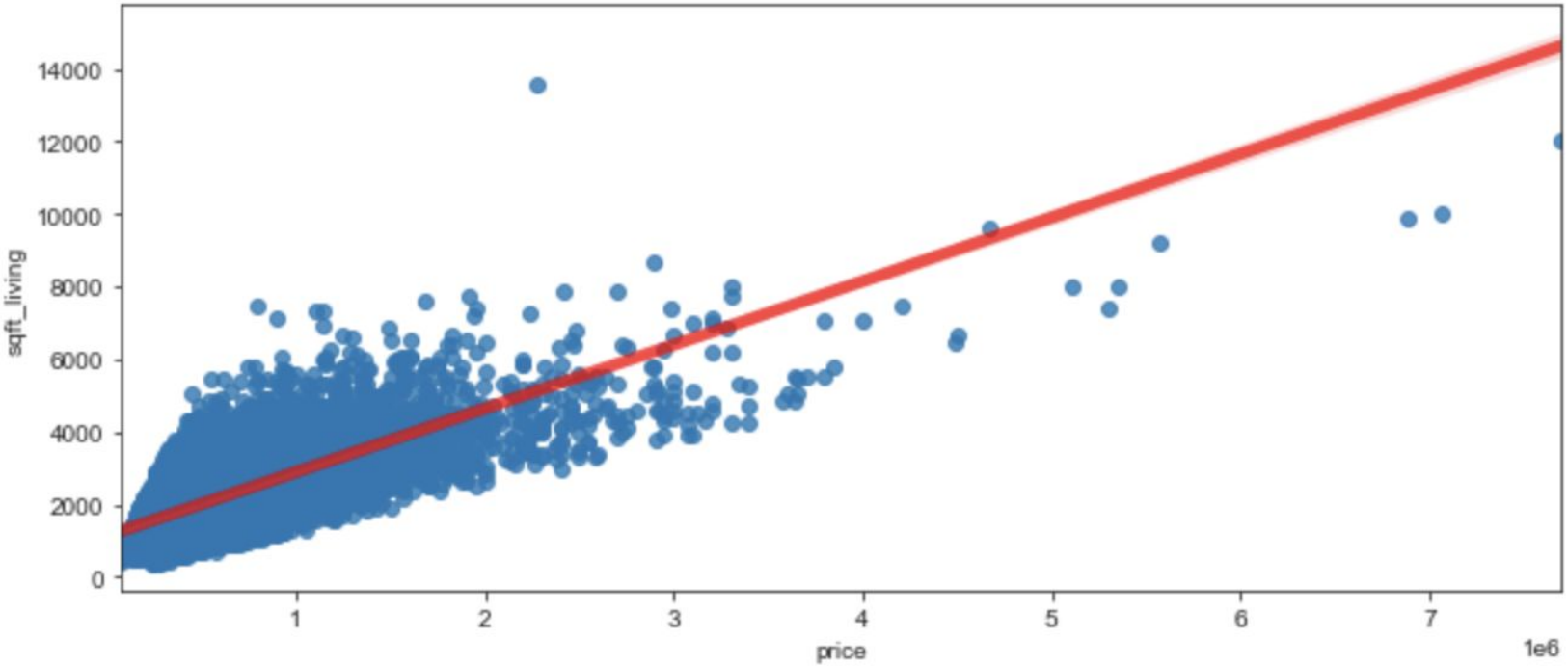
Business Problem

- A Seattle real estate firm would like to create a tool that allows their clients to enter information about their home and receive a prediction for their home's sale price.
- Our goal is to help homeowners visualize and predict the sale price of a home as accurately as possible.

Data & Methods

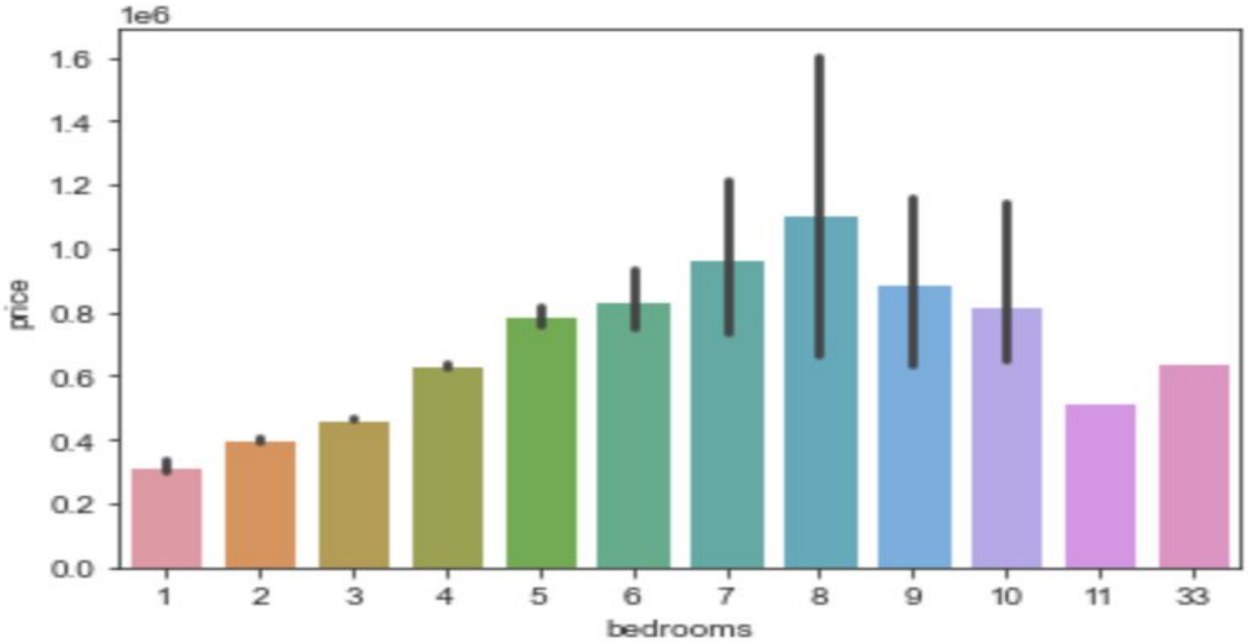
- We used the housing data in a csv file called `kc_house_data.csv`
- Then gathered 12 features related to price containing 21,597 row
- Remove outliers and null values and checking correlation

Correlation between price and square footage



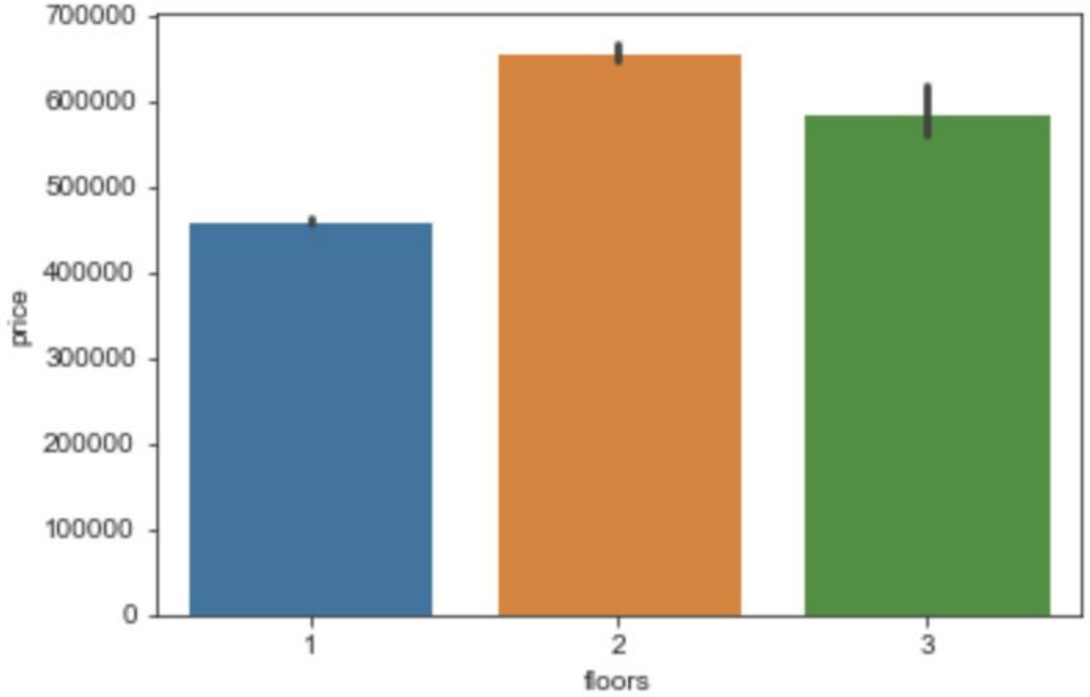
- Square footage of a house was the most strongly correlated feature we found
- Noticed outliers

Correlation between bedrooms and price



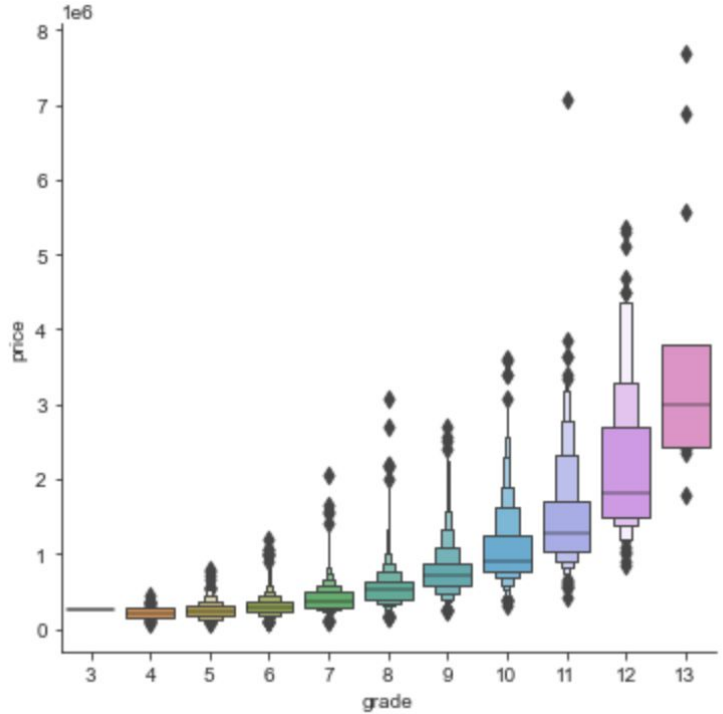
- Price tends to increase as the number of bedrooms increase.
- There are some outliers in the data with a house with 33 bedrooms

Correlation between floors and price



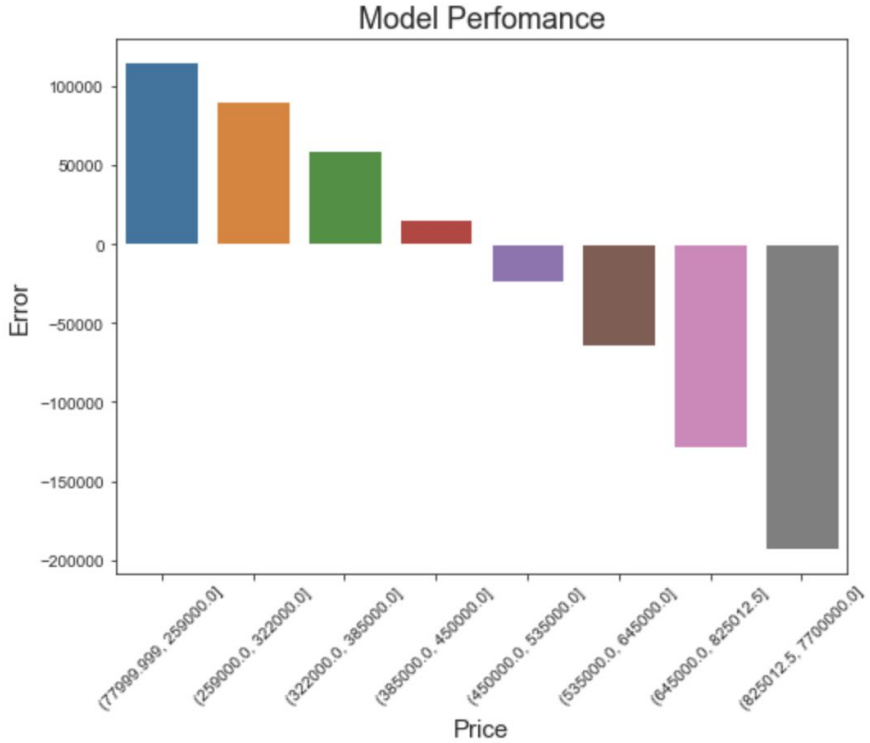
- The amount of floors play a role in how much a house costs.
- Houses with 2 floors usually cost more

Correlation between grade and price



- As the overall grade of a house increases the price increases

Model Performance



- The final model performance shows that it is performing well in the middle

Conclusion

- The model performs well in prices from 385,000 - 535,000 dollars but has high error with prices from 77,999 - 322,000 and 645,000 - 7,700,000 dollars.
- The features that are driving our model is sqft_living, bedrooms, and grade.
- We recommend to not use this model because of the high error in the lower and higher price ranges.