Using The King County Data to Predict house prices

DAVID MWITI

2022

Process

Data cleaning and Processing

Data exploration

Building Linear Models

Interpreting the results

Conclusion

Understanding Our DATA

In its entirety, the king county data set has many features about the different houses in the area.

For instance, number of bedrooms, bathrooms, square footage of the home, square footage of the lot, floors (levels) in house, overall grade given to the housing unit, based on King County grading system, year house was built, among other features of the house.

After cleaning Our data we are left with 21534 house sales records with 20 Features.

Correlation

▶ The heatmap shows the correlation of our different variables in our data set.

▶ In our case we'll mostly focus on the correlation of our dependent variable (price) and the other features (Independent variables)
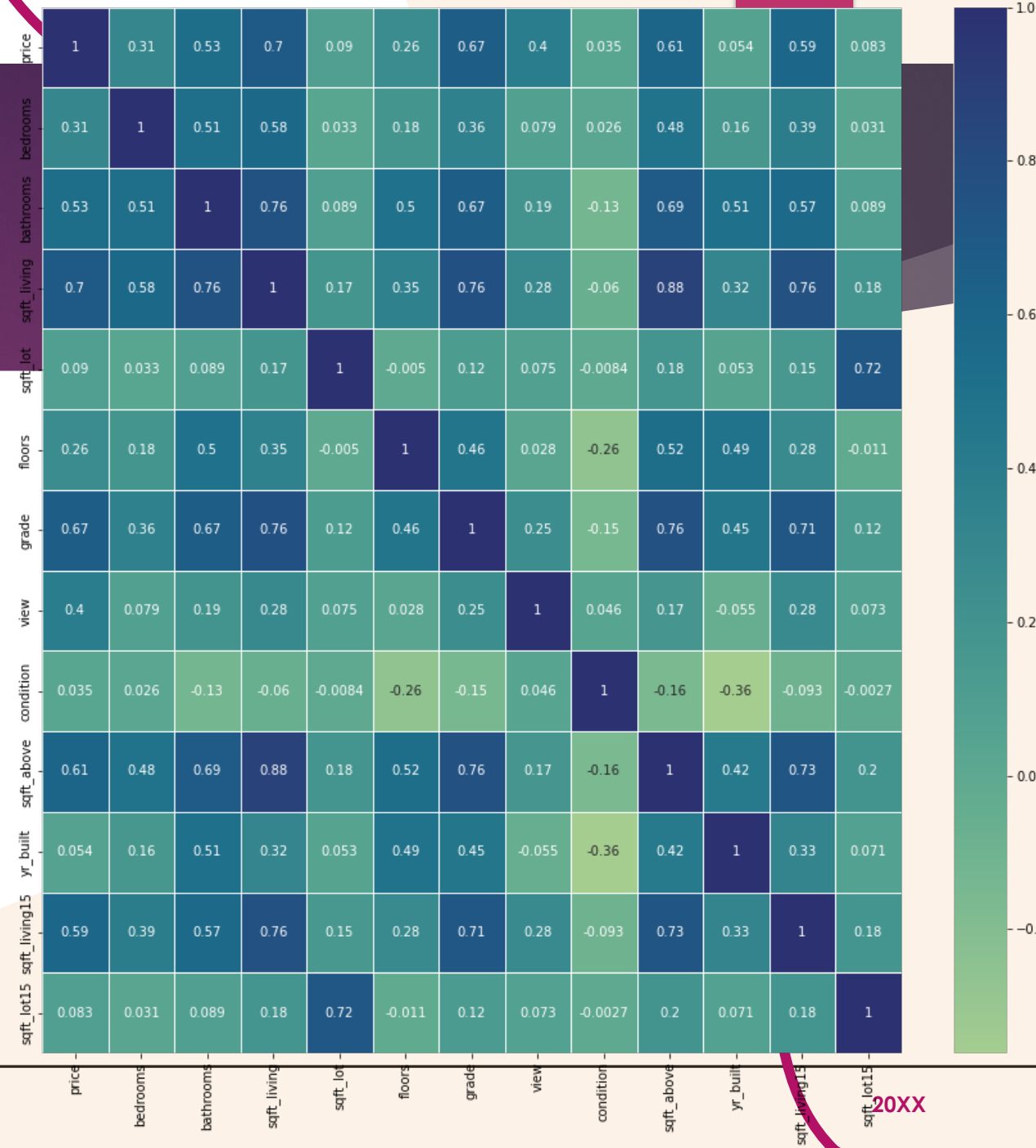
FINAL MODEM AND RESULTS

The R-squared value is another measure of the quality of a predictive model. It represents the proportion of the variance in the dependent variable that is explained by the model. A value of 0.54 for the train data and 0.55 for the test data indicates that the model is only explaining about 54-55% of the variance in the dependent variable. This is not a particularly strong model, however it gets the job done.

Conclusion

THE THREE FACTORS THAT TOGETHER BEST PREDICT A HOME'S PRICE IN KING COUNTY ARE SQUARE FOOTAGE, GRADE, AND BATHROOMS. HOMEOWNERS WHO WANT TO SELL THEIR PROPERTIES FOR MORE MONEY SHOULD CONCENTRATE ON INCREASING SQUARE FOOTAGE AND RAISING THE STANDARD OF CONSTRUCTION. HOMEOWNERS SHOULD THINK ABOUT ADDING MORE BATHROOMS WHILE INCREASING THEIR SQUARE FOOTAGE BECAUSE, ACCORDING TO THIS DATA, THE NUMBER OF BATHROOMS IS POSITIVELY CORRELATED WITH PRICING.

thank you

DAVID MWITI

MMWITIDAVEVEN@GMAIL.COM

WWW.CONTOSO.COM