King County: How Home Attributes and Location Affect Home Price

SUMMARY

In order for us to accurately inform King County home sellers how much their homes would sell for depending on different attributes, we had to go through the process of making multiple models that would show how price is affected by different attributes of a home. After finding more data, cleaning the data, making multiple models and graphs, we have a better idea on how much different features can affect the price of a home. These results will be shown in multiple graphs.

OUTLINE

➢ Business Problem

➢ Baseline Model

➢ Preprocessing

➢ Advanced Model

➢ Visualizations

➢ Conclusions

BUSINESS PROBLEM

A Seattle Real Estate Firm would like us to create a model that will make predictions about a home's selling price as accurately as possible. Homeowners want to know how components of a home will affect the selling price of their home. The models that we created will tell the owners about how much their home will sell for once they put in all of the components of that home. The selling price of a home changes based on certain features which we will show.

BASELINE MODEL

- All original data
- Size of House
  - sqft_living, sqft_lot, sqft_living15, sqft_lot15, sqft_above
- Features of House
  - grade, condition, bathrooms, bedrooms, yr_built, lat, long, condition, grade
- Target: price
- R^2 value : 0.6749801766510155
- Training RMSE: 205095.986
- Testing RMSE: 211169.6608

PREPROCESSING

- Cleaned data
- Got rid of outliers
- Found more data
- Found which features had highest coefficients
- Tested models

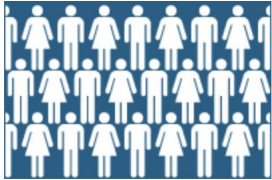# How does logging price improve normalization?
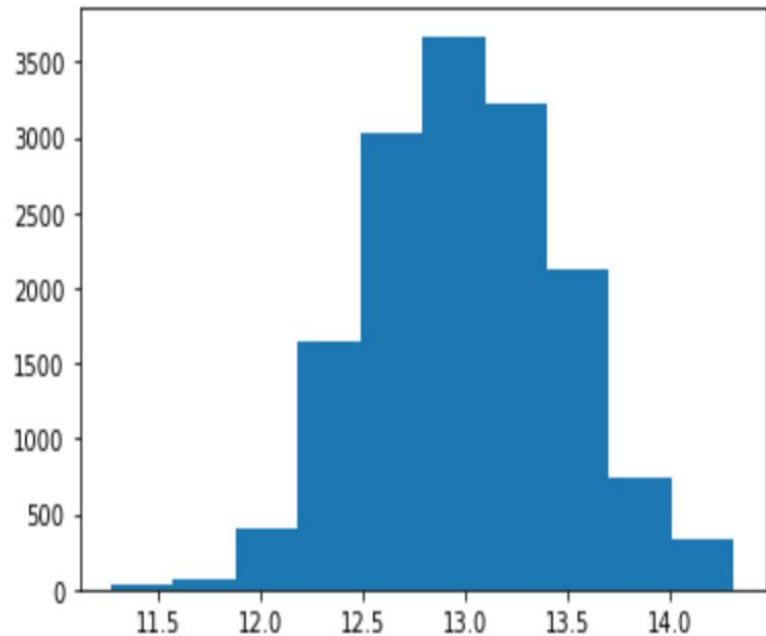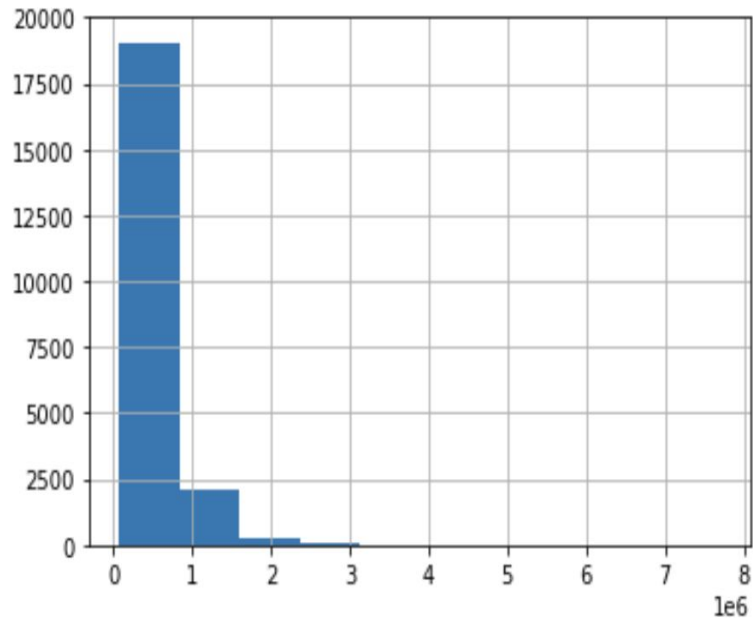
Utilizing Census Data
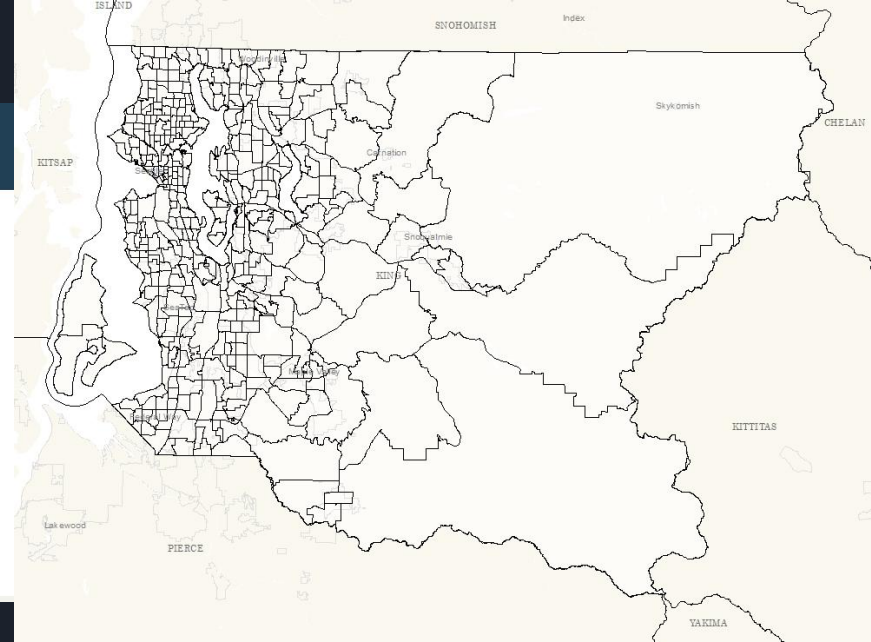
King County GIS Open Data

Census business tables

The census thematic category contains census geometry layers, such as Blocks, Block Groups, and Tracts, and business tables (for both decennial census collections and American Community Survey five-year estimates) which can be joined to these geometry layers for display and analysis. We construct and provide only a limited selection of such joined datasets as individually downloadable items in this Open Data site, based on high interest in particular topics. If you wish to download the other census business tables that we offer, you must download all census feature classes or shapefiles as indicated below.
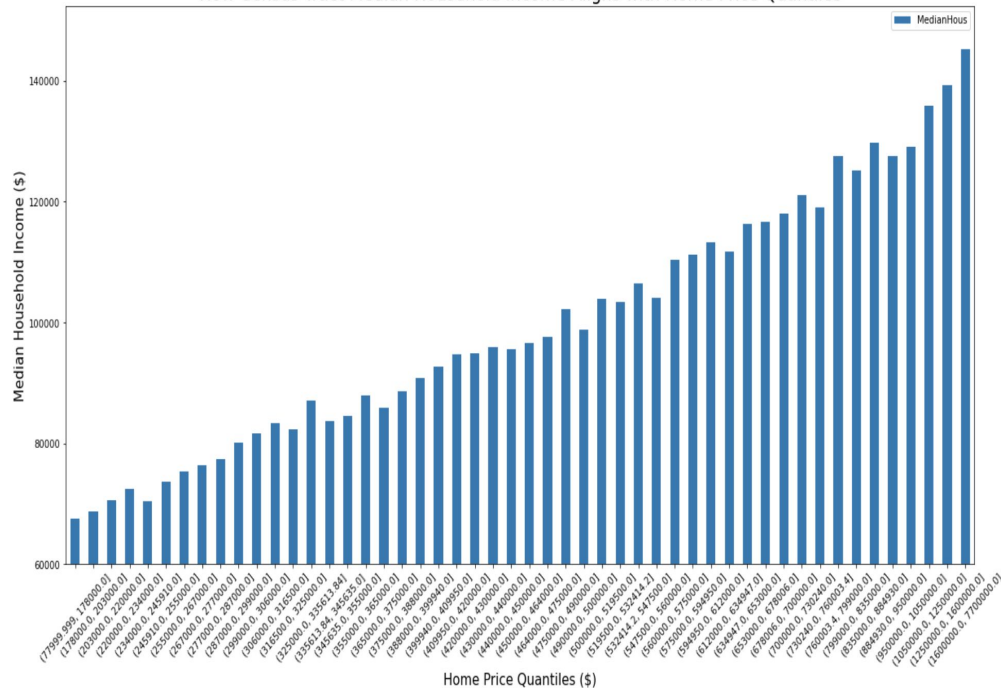
ADVANCED MODEL

- Imported Census Data From King County GIS
- Features used:
  - Size of House
    - sqft_living, sqft_lot, sqft_living15, sqft_lot15, sqft_above
  - Features of House
    - grade, condition, bathrooms, bedrooms, yr_built
  - New Columns
    - empty_lot_space = sqft_living-sqft_lot
    - Compared_to_neighbors = sqft_living/sqft_living15
    - Median Household Income from King County GIS
  - Categorical Features
    - Census Tract
- Scaler: MinMaxScaler
- Target: price
- R^2 value: 0.880499
- Training RMSE: 137156.26
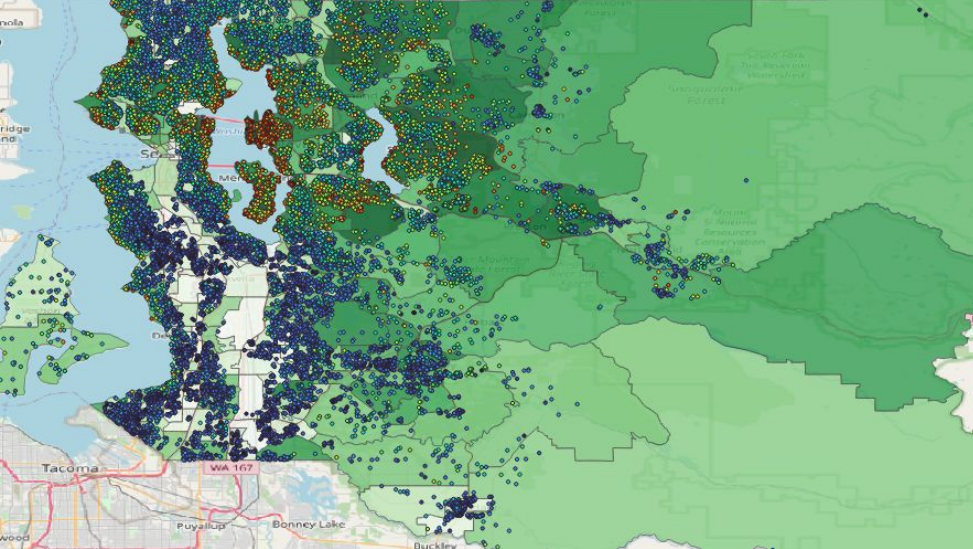- Testing RMSE: 137393.085

Do our features have a linear relationship with price?
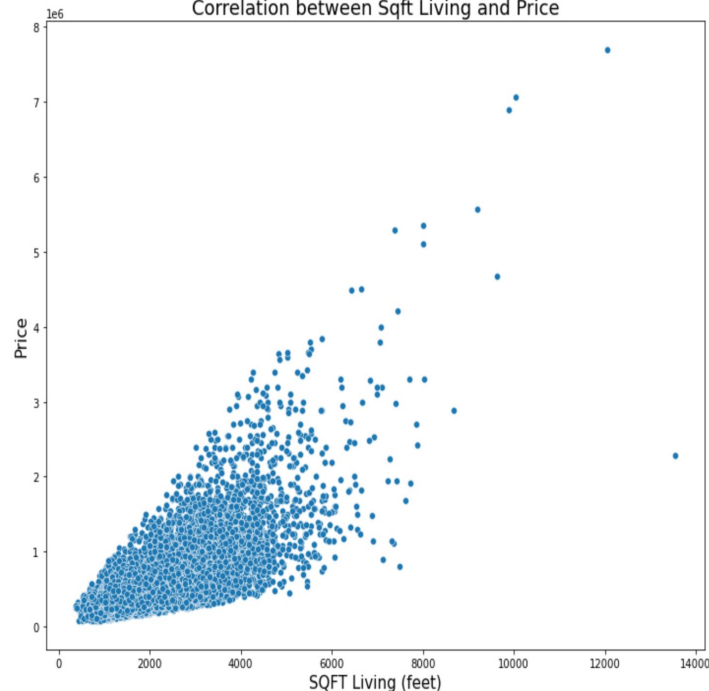
Median Household Income by Census Tract

$10,521 $ 50,000 $89,117 $100,000 $210,491

Home Sale Price

$78,000 $350,000 $539,325 $1m $2m $ 7.7m
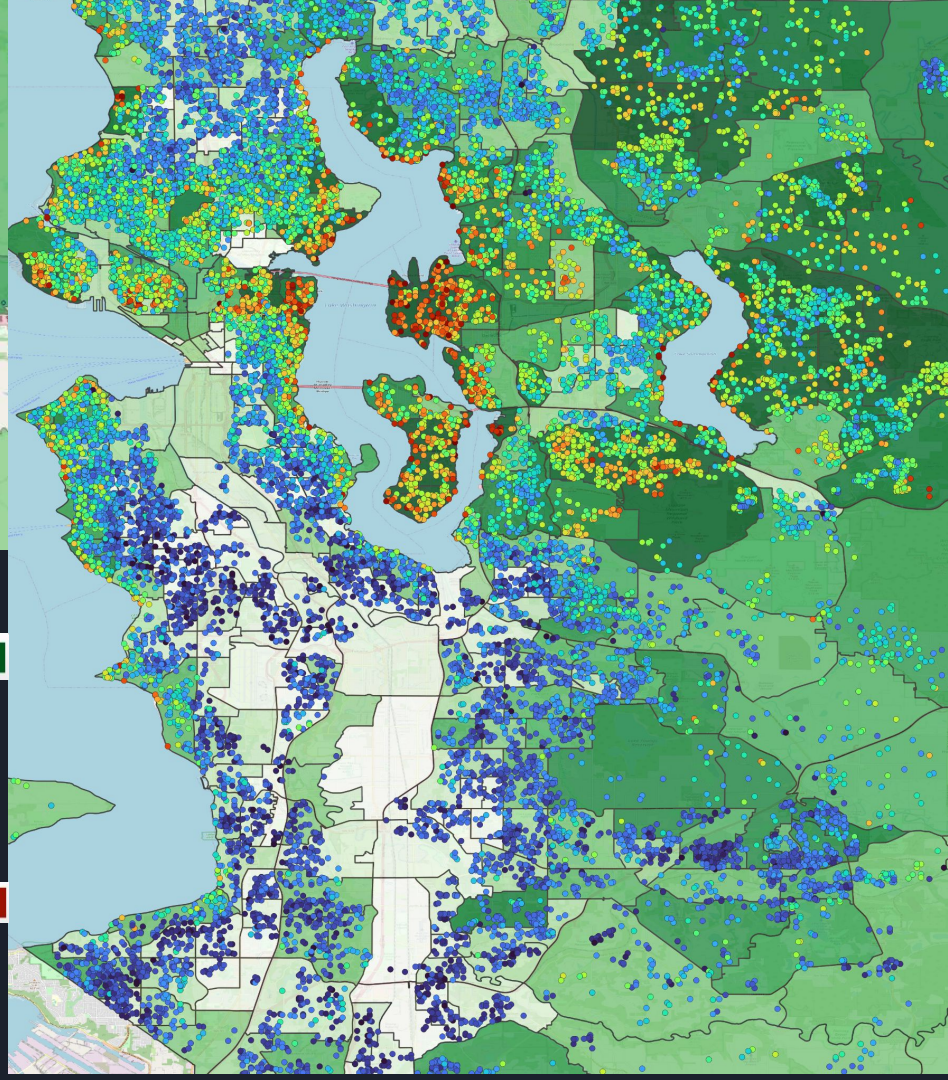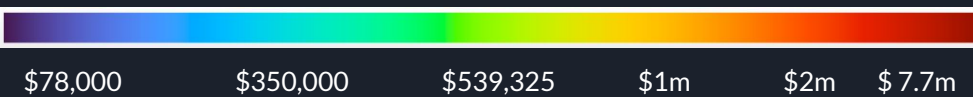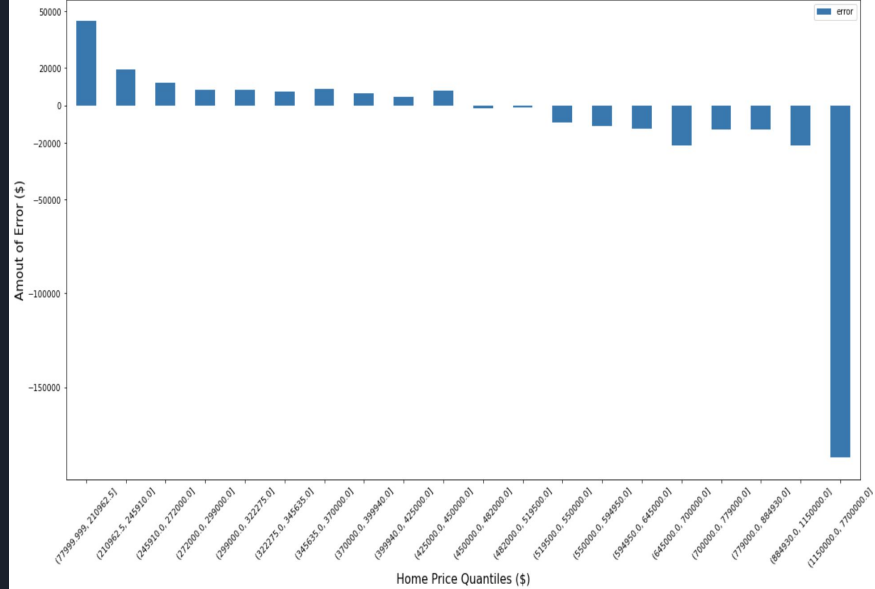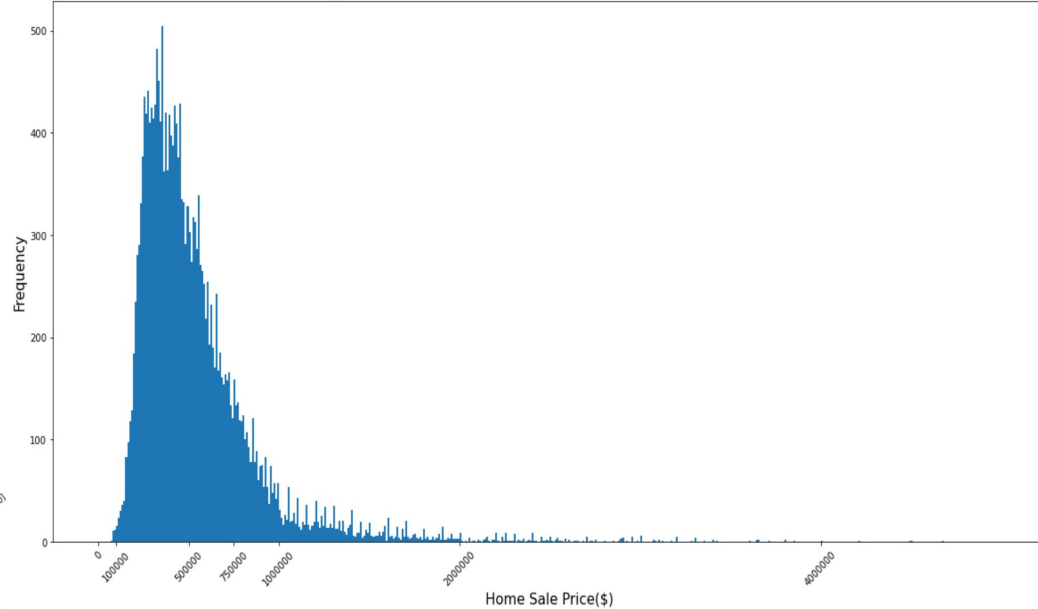
Where our error is found

How Error is Dispersed Among Home Price Quantiles

Histogram of Sale Price or How Much Most Homes Sold For

CONCLUSION

In conclusion, we found a couple attributes of a house that have large impact on the selling price of a home. In particular, the features, Grade, Empty Lot Space, Square Foot Living, Square Foot Lot, and Median Household Income have a large effect on a homes selling price when these features are compared to neighboring houses. For example, if the house the owners want to sell has a square foot living that is larger than the neighboring houses, that house would sell for more.

THANK YOU

For any additional questions, please contact us via email:
mallorye1103@gmail.com
charliefountain122@gmail.com
johncsimmons99@gmail.com