



Trends for Increasing Home Sale Price in King County

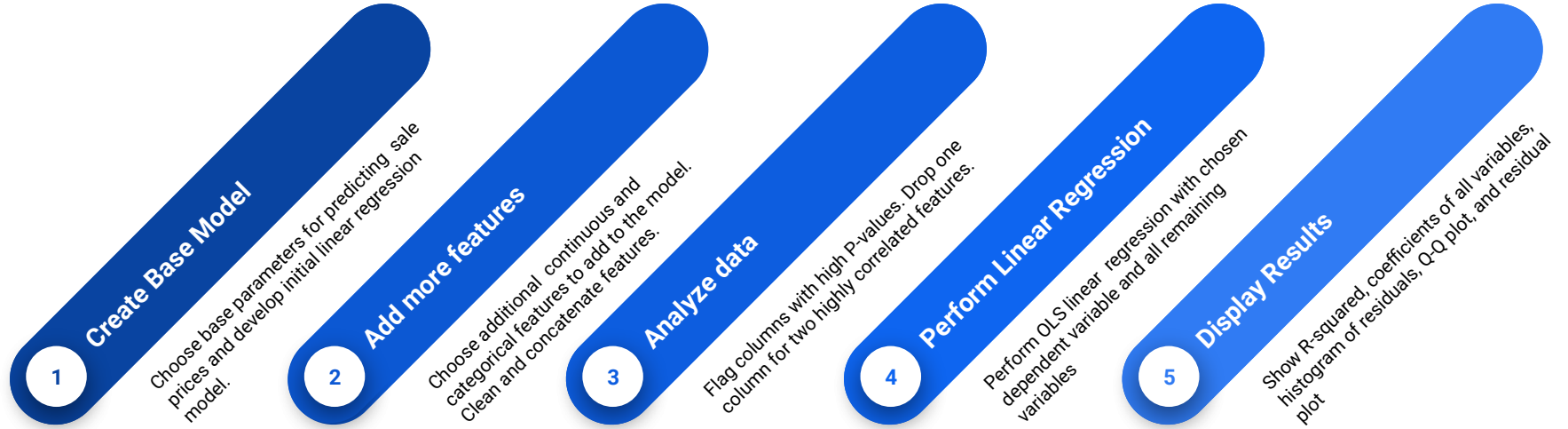
Hunter Ewin, Akash Rama, and Wyatt Sharber



Goals of Study

- Identify features that are related to property sale prices in King Co. Washington in 2018
- Features of special interest are:
 - Total square feet of house
 - Having a porch
 - Having waterfront property
 - Noise-related nuisances
- Build a linear regression model that predicts sale price with these features as explanatory variables

Our Workflow Function



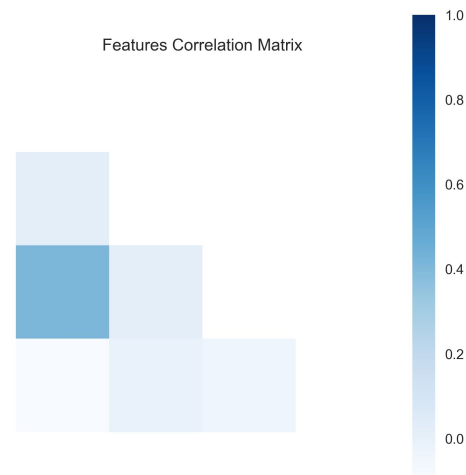
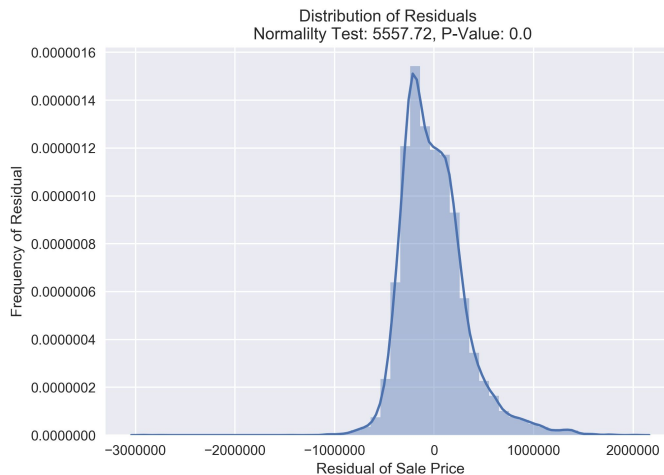
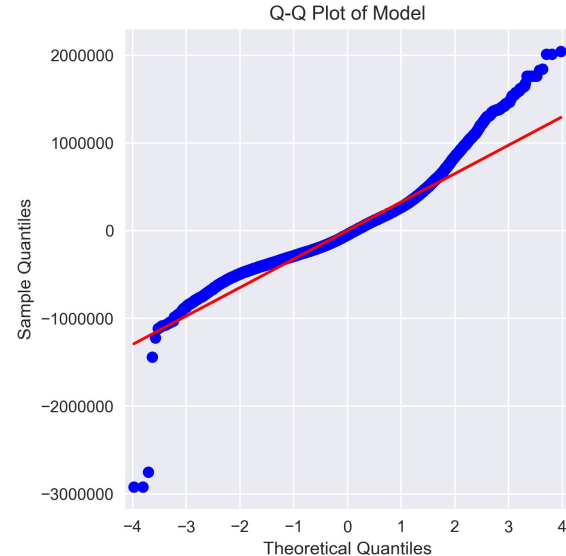
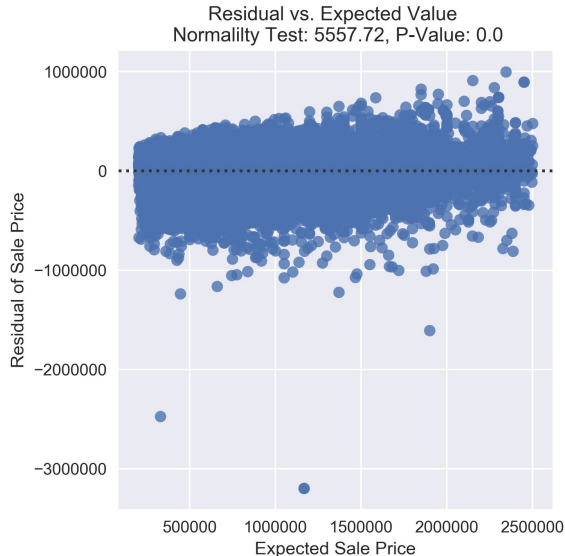
Base Model

Sale Price predicted by 4 features:

- Total Sq. Feet of Living Space
- Waterfront Location
- Presence of Porch
- Presence of Noise Nuisance (e.g., airport, traffic)

Poor fit ($R^2 = 0.24$), but:

- Residuals are mostly normally distributed (although not significant)
- Residuals exhibit homoscedasticity
- Low correlation between features



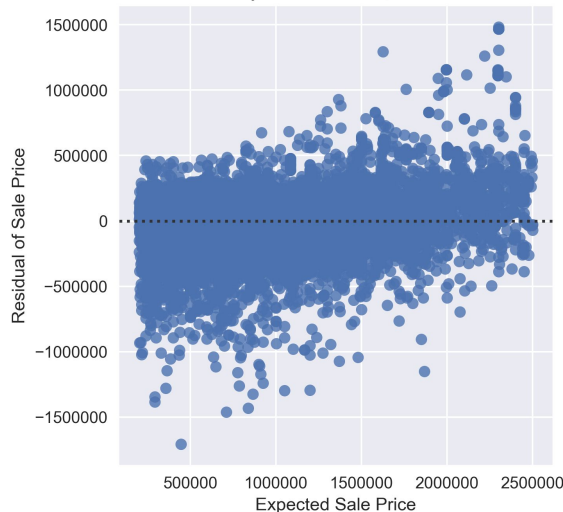
Improved Model

Sale Price predicted by **634** features

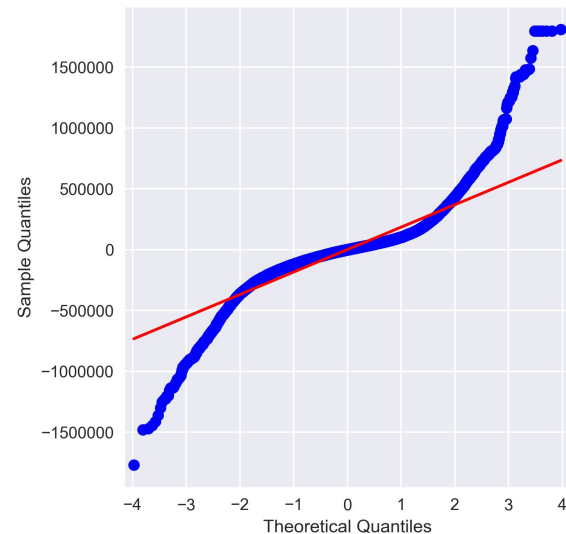
Good fit ($R^2 = 0.788$, Adj. $R^2 = 0.783$),
but:

- Residuals are not normally distributed
- Residuals exhibit heteroscedasticity
- Still low correlation between features

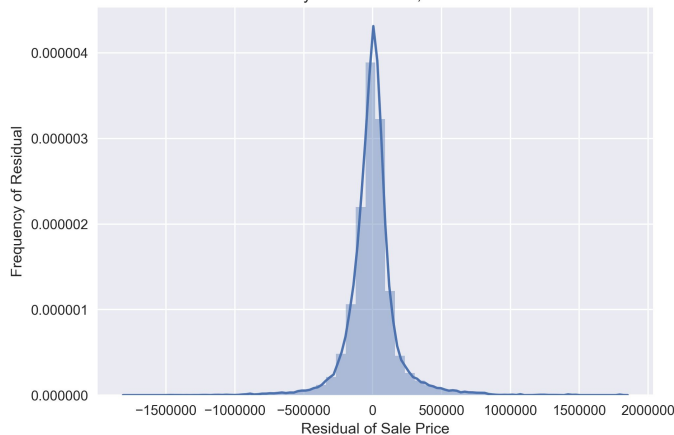
Residual vs. Expected Value
Normality Test: 7378.98, P-Value: 0.0



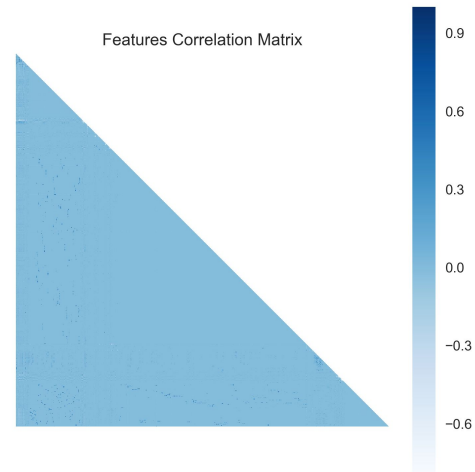
Q-Q Plot of Model



Distribution of Residuals
Normality Test: 7378.98, P-Value: 0.0



Features Correlation Matrix



Features of Interest are significantly correlated with Property Sale Price

- Total Sq. Feet of Living Space is positively correlated with Sale Price
 - 101833.2, $P = 6.9e-272$
- Waterfront Location is positively correlated with Sale Price
 - 14000.03, $P = 6.6e-10$
- Presence of Porch is positively correlated with Sale Price
 - 15155.58, $P = 1.0e-26$
- Presence of Noise Nuisance (e.g., airport, traffic) is negatively correlated with Sale Price
 - -5414.88, $P = 0.000005$

Other Important Features in Improved Model

- Heating-related features (heat source, heating system)
- Location-based features (area, zip code)
- Property use and condition features (condition, building grade, property type, present use)

Model Improvements

- Our model does not perform well at extreme valued homes
 - Outlier detection and handling
 - Correct for non-normal error distribution
- Detailed data validation
- Feature engineering
- Explore other modeling techniques for nonlinear predictions