Justin Collins

MODULE 1 PROJECT
KINGS COUNTY HOUSE PRICES

Data File

- Data from Kings County Housing

- Target Variable = Price

- Many Possible Predictors

Data Cleaning

- Transformed ordinal into 1-hot encoding

- Transformed continuous via log or min/max scaling

- Created categorical variables from strings

Features Created

- Zip code binning system by SES of

- Percent of Lot Used predictor

- Age of House predictor

- Categories for Grade, Condition and View

Model Creation

- Utilized Multivariate Linear Regression from the StatsModels library to create a model to predict house price based on predictor variables.

- Slightly different iterations of predictor variables were used between the two models

Model Testing Outcomes

- Model 1:
  - R-squared of .593
  - RMSE-test = $219,759

- Model 2:
  - R-squared of.578
  - RMSE-test = $242,849