Real-Estate Pricing Predictors

Justin Grisanti

Objectives

Obtain a Business Understanding

Understand the Data

Prepare Data for Modeling

Create Models

Generate Regression Results

Business Understanding

- Stakeholders:
  - Home Improve Inc.
  - Homeowners
  - Home buyers
  - Real-Estate Agents
- Business Problem:
  - To help customers understand and improve the value of their home
- Predictive and Inferential Model

Data Understanding

- Homes in Kings County, Washington
- Important Housing Information:
  - Price
  - Bedrooms
  - Sq Feet
  - Grade
  - Floors
  - Etc.

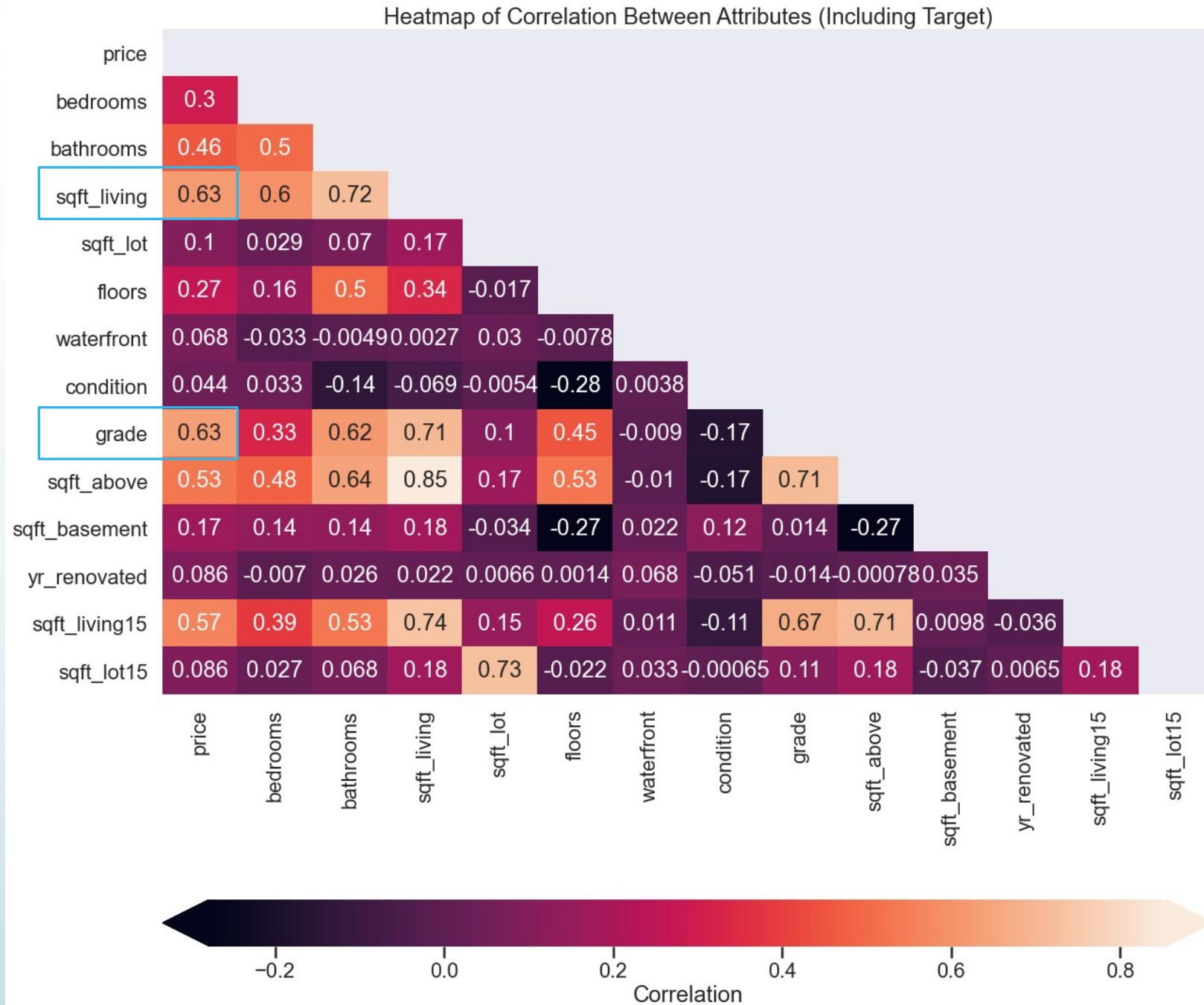# Data Preparation

- Train and test data:
    - Remove outliers
    - Fill empty data
    - Check for erroneous data
- Clean all columns to ensure better modeling

Modeling

- What are the most correlated variables?
  - Grade & living space

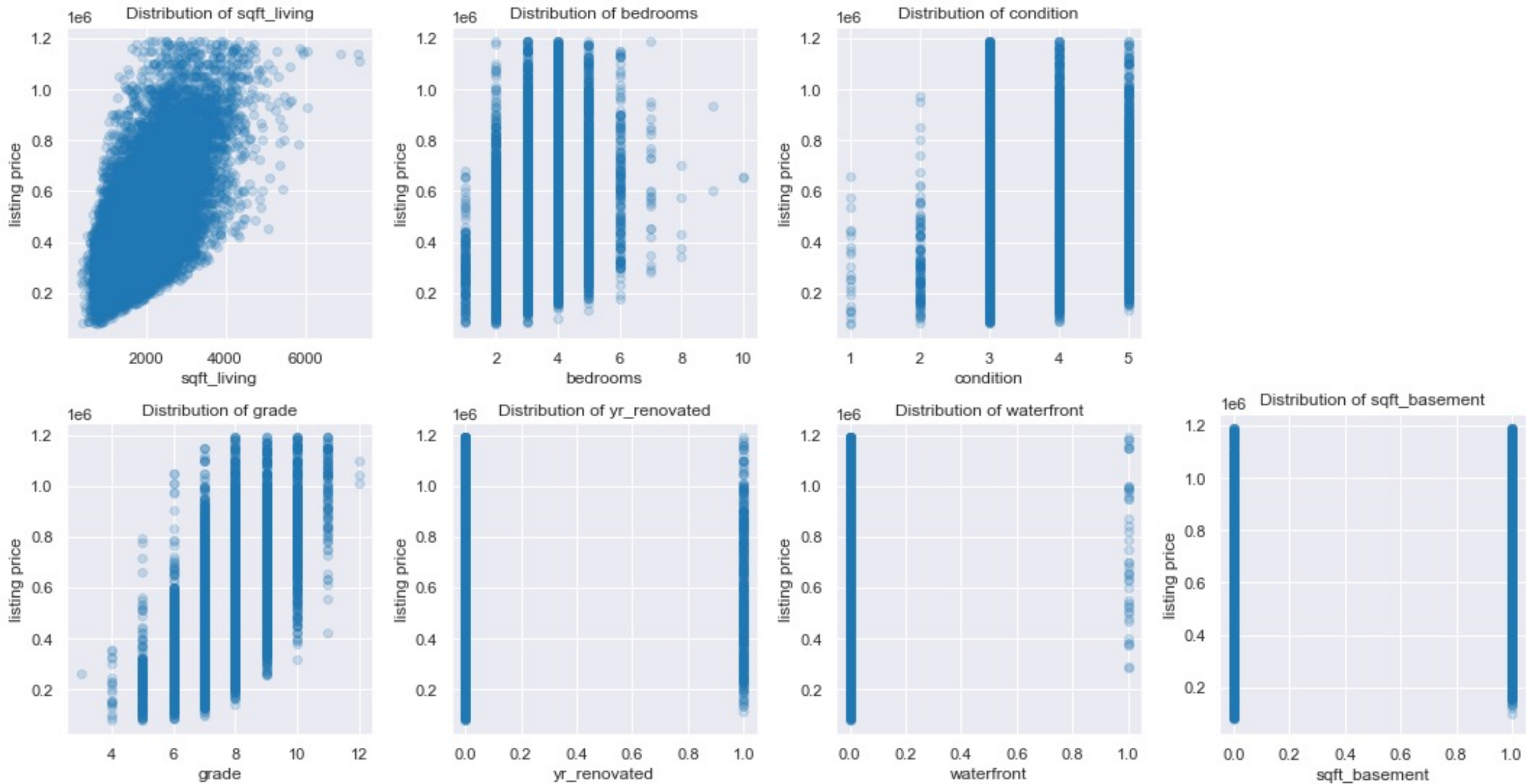Heatmap of Correlation Between Attributes (Including Target)

Modeling

- Baseline Model (Grade and Sq. Feet Living):
  - Train: 0.403
  - Validation: 0.400
- Second Model (Numeric Fields):
  - Train: 0.488
  - Validation: 0.485
- Final Model:
  - Train: 0.505
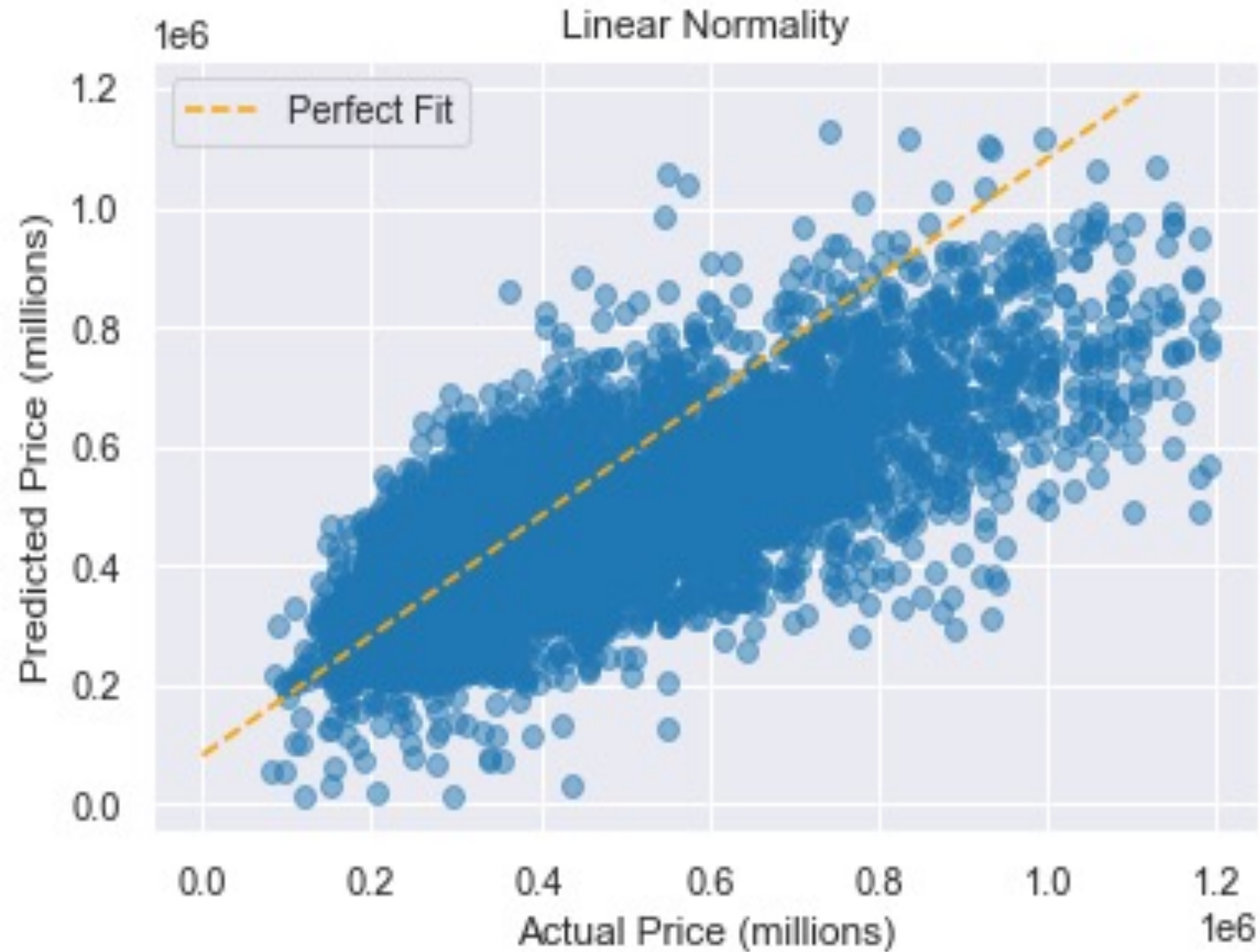 - Validation: 0.498

Final Model Attributes

Regression Statistics
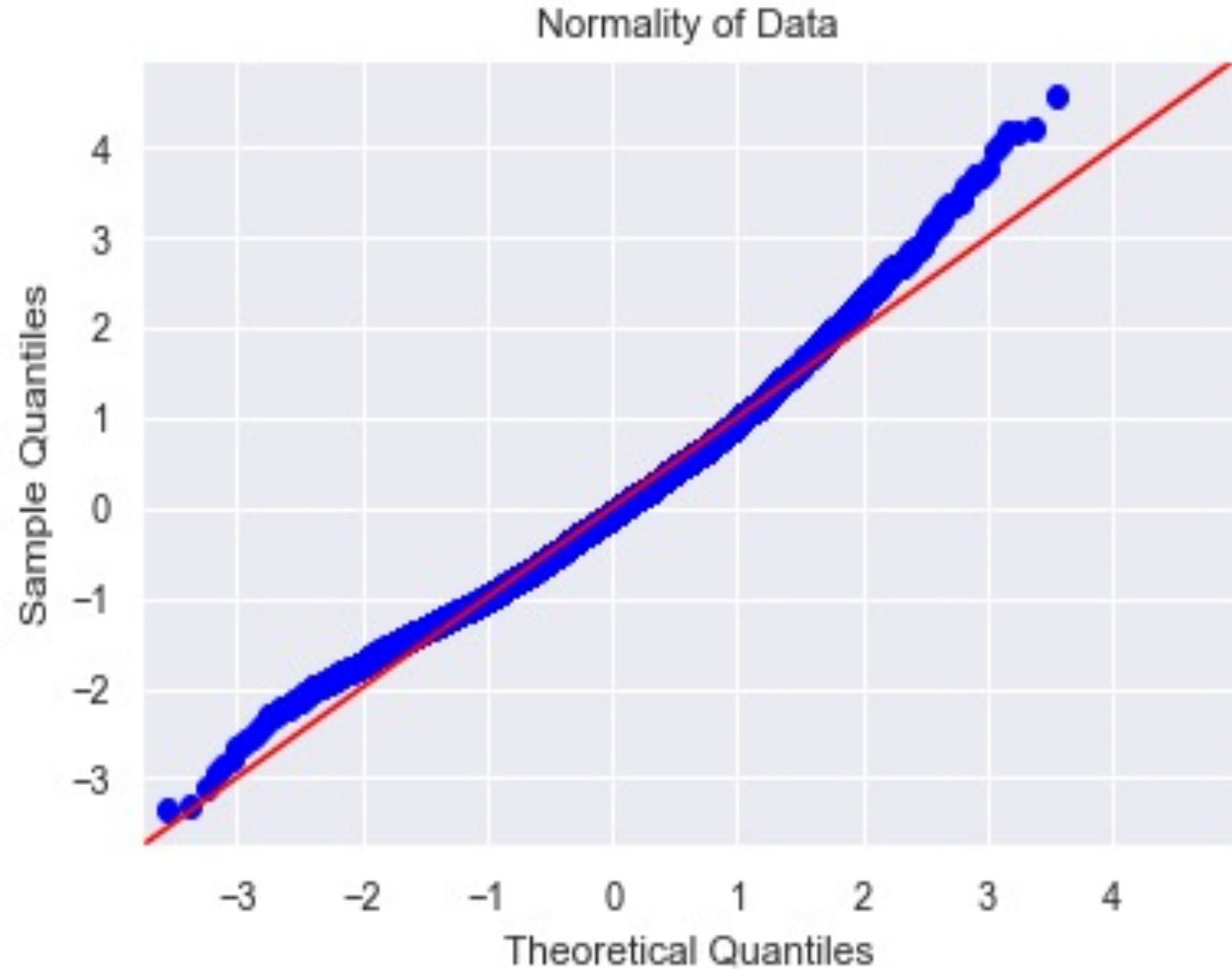
- Mean Absolute Error: $118k

- Mean House Price: $481k

- Coefficients:
  - Sq Feet Living Space: $98
  - Bedrooms: -$17,333
  - Condition: $45,169
  - Basement: $38,708
  - Grade: $88,275
  - Renovated: $102,962
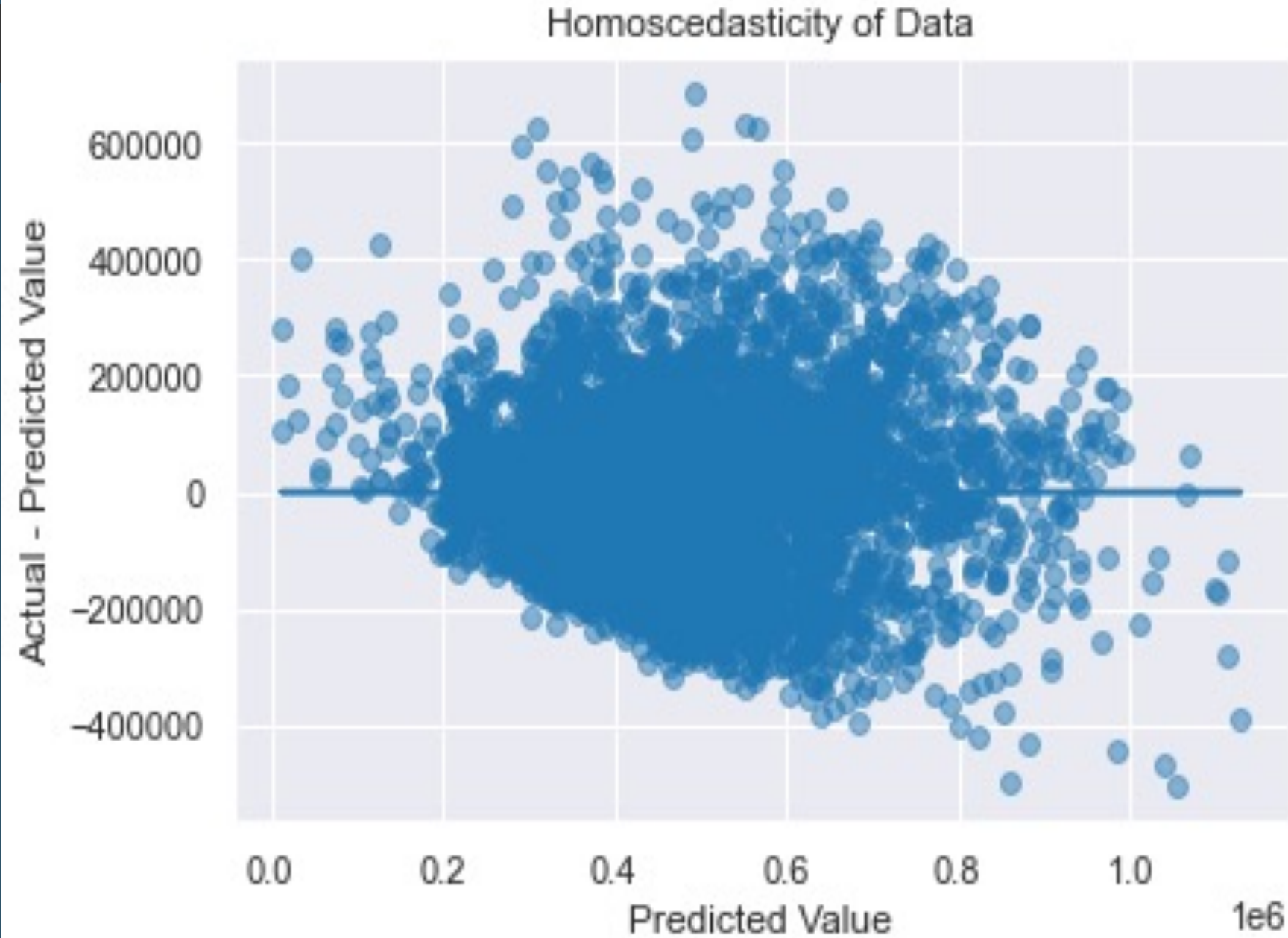 - Waterfront View: $242,268

Normality Assumptions - Linearity

- Normal trend with r^2 statistic of .51

Normality Assumptions – Q-Q

Normality of Data

- ➨ Seems like a normal distribution, with few outliers at higher prices

Homoscedasticity

Homoscedasticity of Data

- Residuals seem even, however at higher prices the model tends to overpredict price

Next Steps – Improve Model

- Add more features to model:
  - House color
  - Location
  - Proximity of Police Station
- Add more data for higher priced houses

Thank you for your time!