



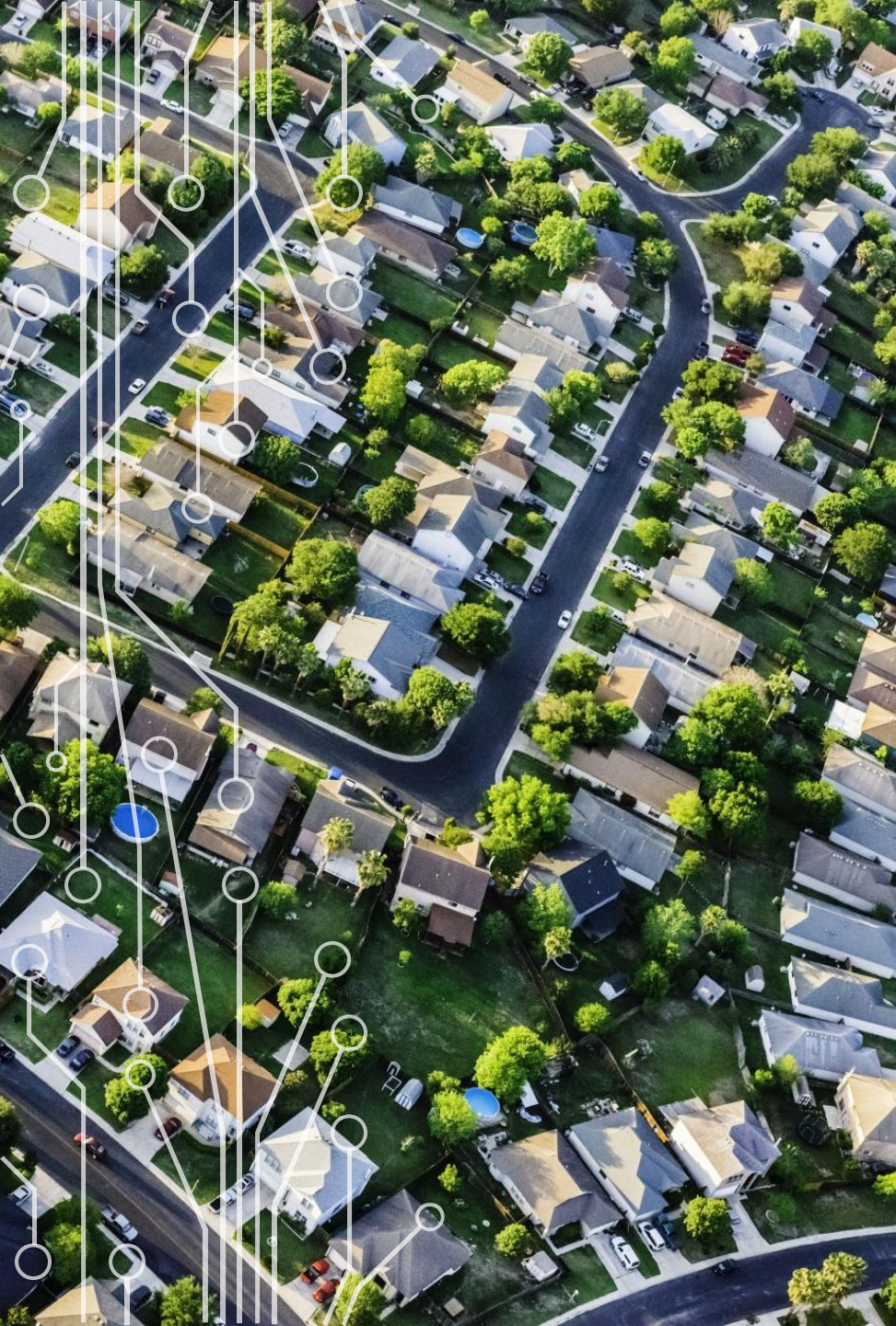
3/25/2023

PREDICTING HOUSE PRICES IN KING COUNTY: A MACHINE LEARNING APPROACH



BACKGROUND

- Housing has been a basic human need for centuries, and over time, the concept of housing has evolved with the changing needs and lifestyles of people. Lately, people have commercialized housing aiming at maximum income generation. Different variables however can significantly impact the overall value of a property in different magnitudes.



PROJECT OVERVIEW

- For this project, I will use regression modelling to analyse house sales in a north-western county.

The background of the slide is a dark blue gradient. On the left side, there are white circuit-like lines with small circles at the ends, resembling a printed circuit board. In the center, there is a large, dense pile of 3D dollar signs (\$).

BUSINESS PROBLEM

- Homeowners often seek ways to increase the value of their property, and renovations can be an effective way to achieve this. Therefore, there is a need for a reliable way to estimate the impact of home renovations on property value to help homeowners make informed decisions.

DATA UNDERSTANDING

The King County House Sales dataset contains information on the sale of houses in King County, Washington, USA between May 2014 and May 2015. The dataset includes 21,597 observations and 19 variables.

The dataset King County House Sales dataset and the data description can be found in, <http://localhost:8888/tree/data>

OBJECTIVES.

MAIN OBJECTIVES

- To provide advice to homeowners about how different variables might increase the estimated value of their homes, and by what amount.

SPECIFIC OBJECTIVES

- To build a predictive model that can accurately estimate the sale price of houses in King County based on their various features such as number of bedrooms, bathrooms, square footage, location, and others.
- To identify the key features that have the most significant impact on the sale price of houses in King County and provide actionable insights to real estate agents, property developers, and homeowners on how to maximize the value of their properties.

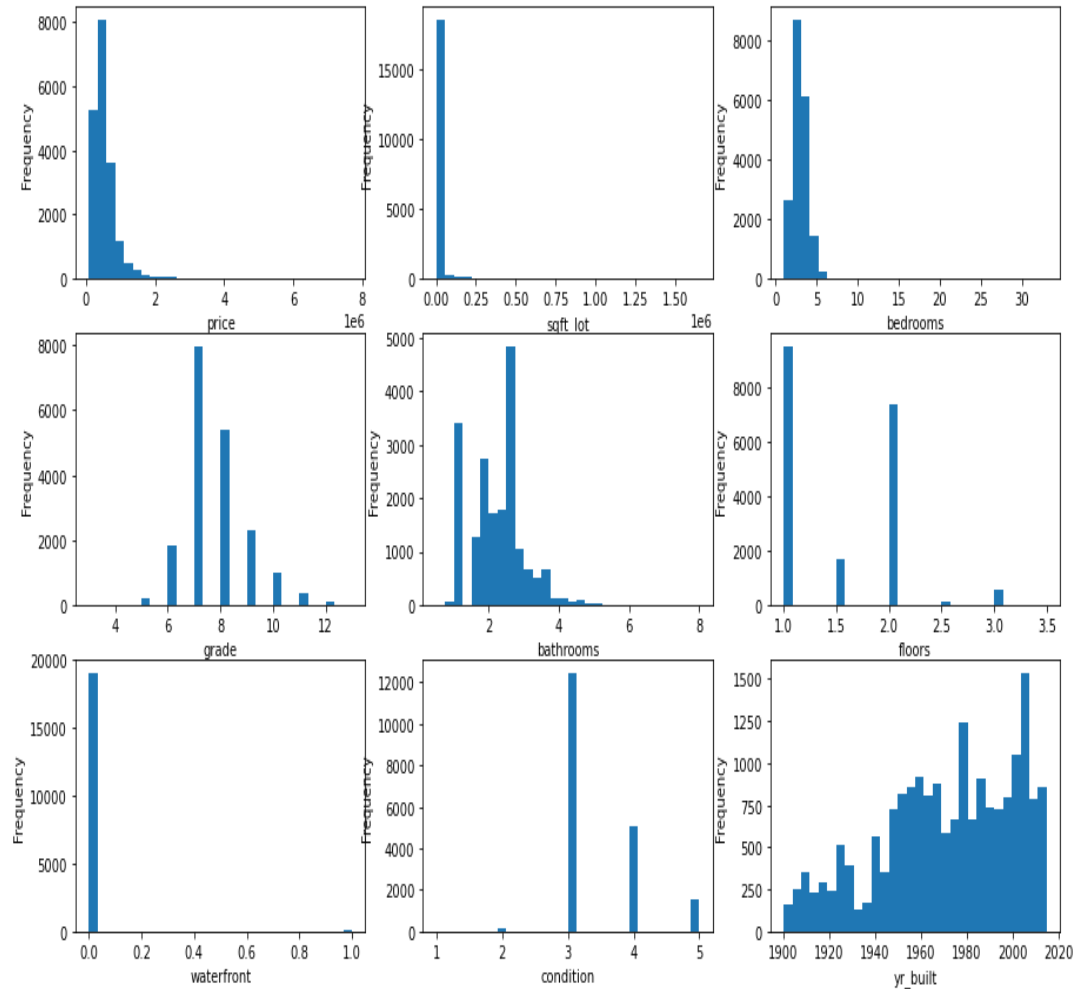
LIBRARIES

- PYTHON- Programming language
- PANDAS- Exploratory data analysis
- SEABORN- Visualization
- NUMPY - Numerical computing and data analysis
- MATPLOTLIB- Visualization
- STATSMODELS-statistical analysis and modeling
- SCIKIT LEARN- machine learning
- SCIPY- scientific computing
- MATH- basic mathematical operations

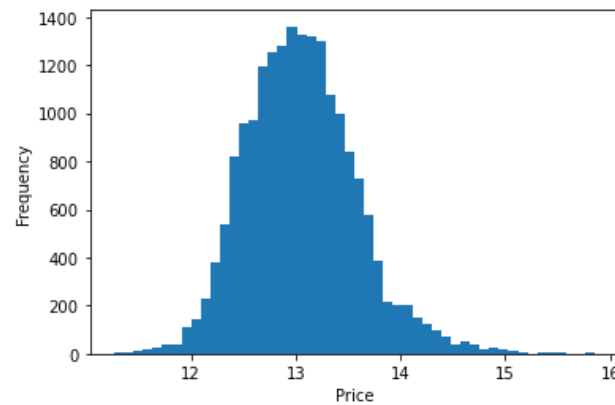
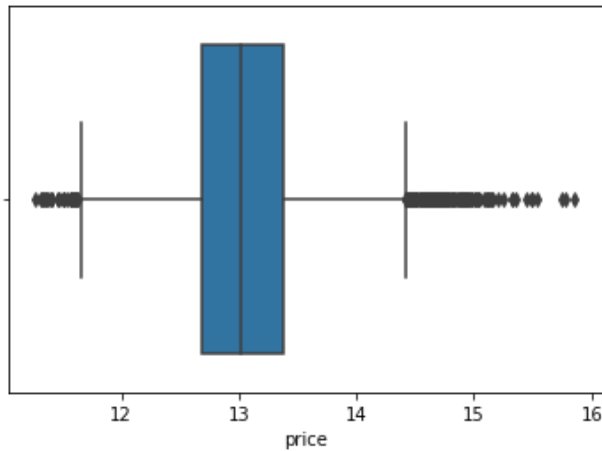


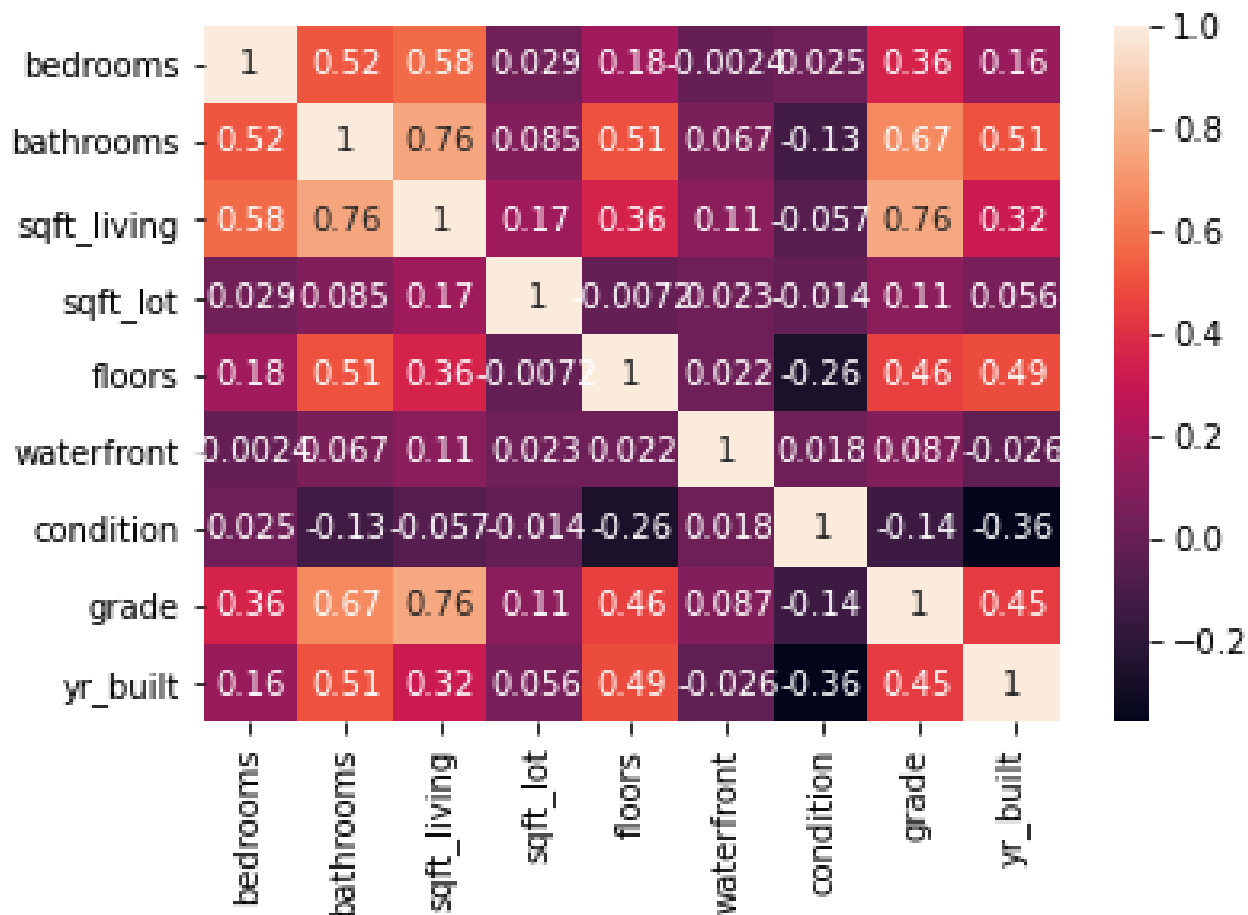
EXPLORATORY DATA ANALYSIS

- The histograms displays the distribution for the variables in the dataset.
- There is positive skewness on price and sqft_lot.
- Presence of categorical variables.

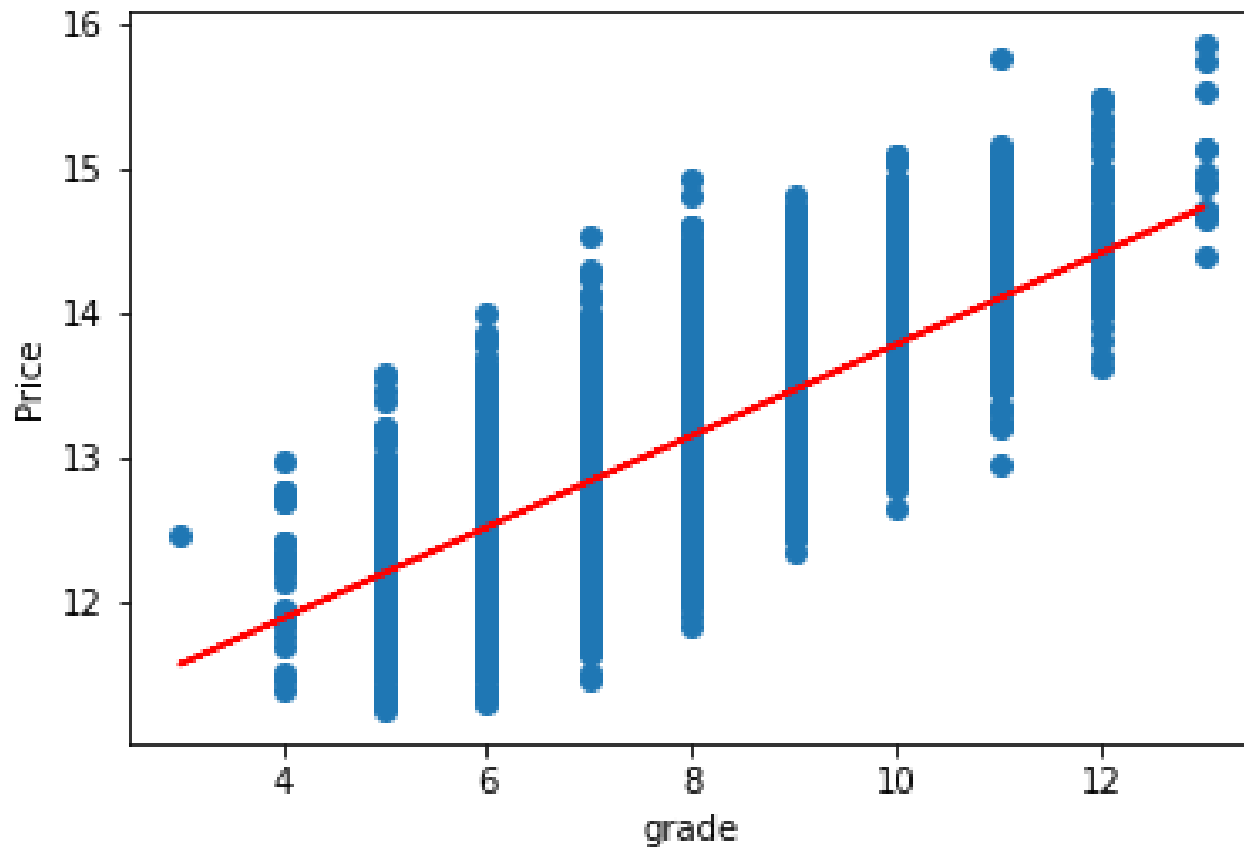


The distribution of price



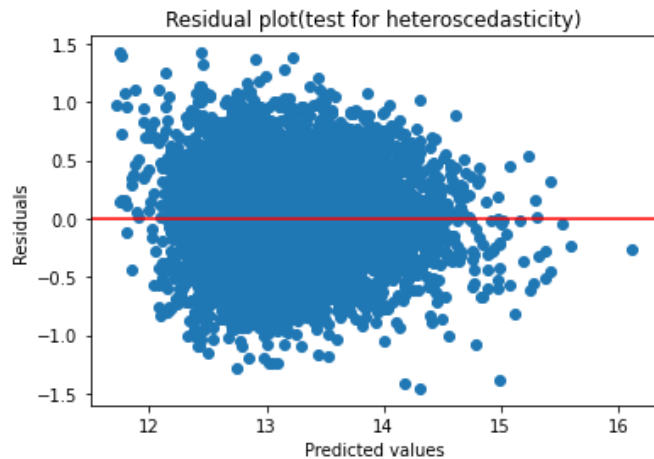


- The heatmap showing the correlation other variables and themselves. bathroom, sqft_living and grade are highly correlated with a value above .75. this could affect our model.

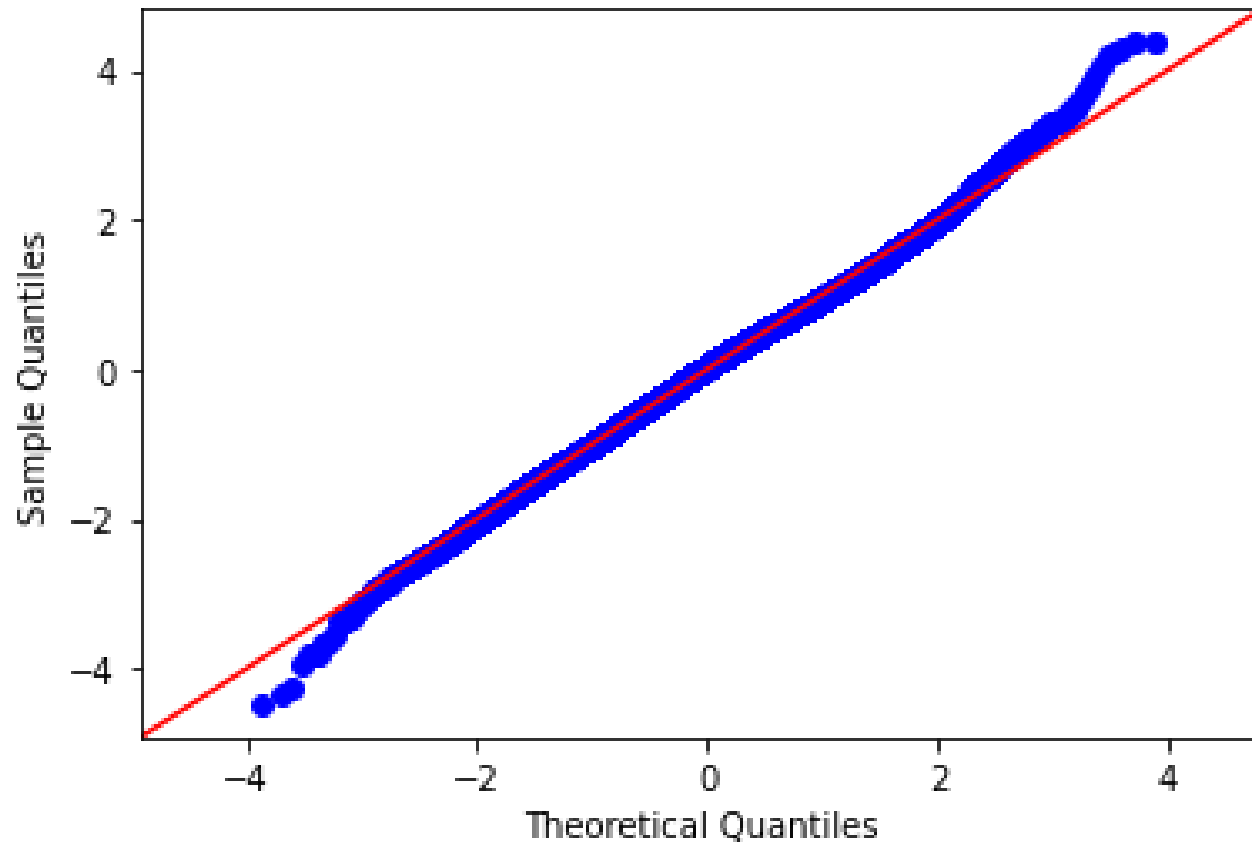


model one :
simple linear
regression
(price and grade)

The regression equation is $\text{Price} = 10.63 + 0.32 * \text{Grade}$



- Test for heteroskedasticity. The plot also shows homoscedasticity in our dataset.



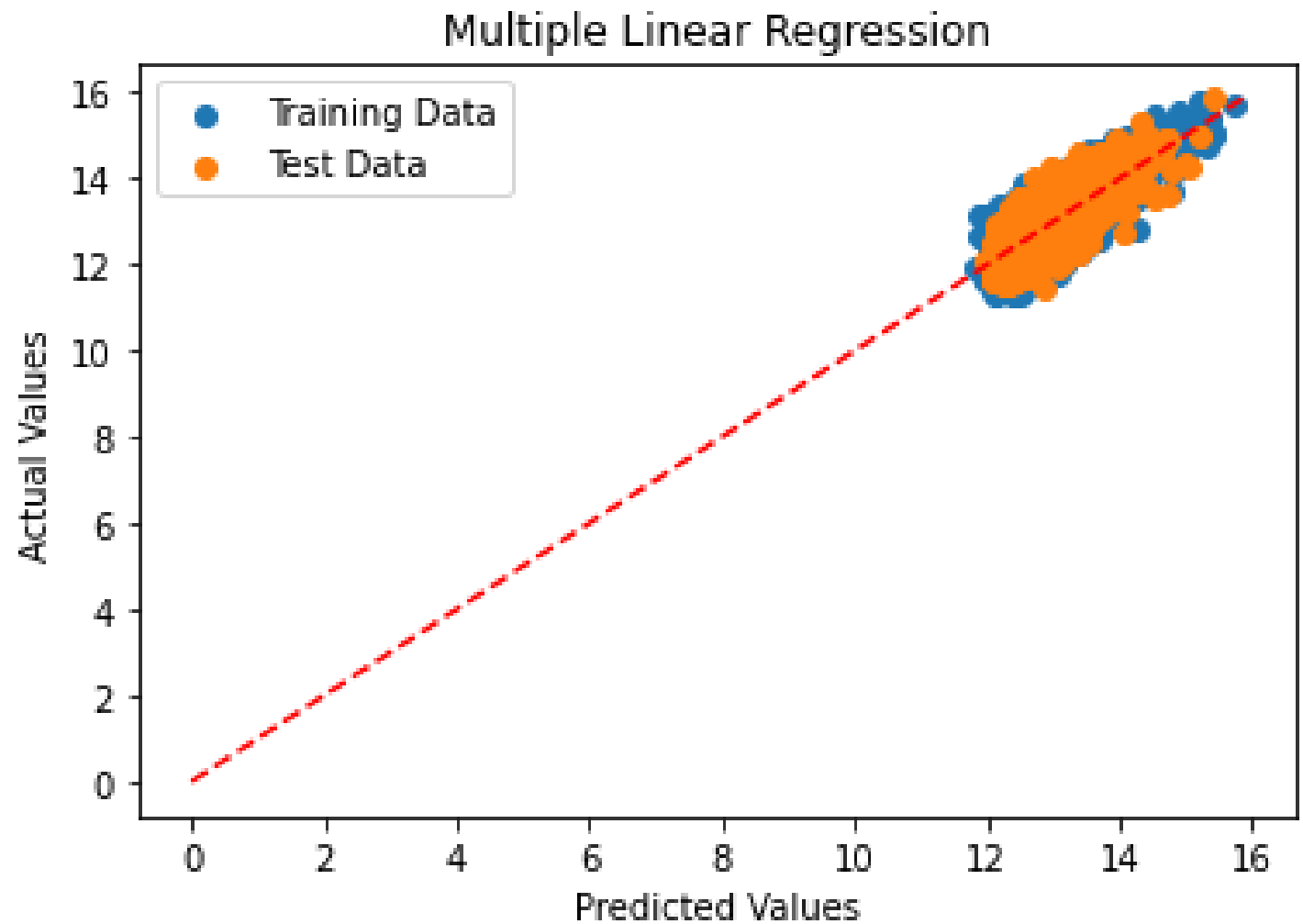
Q-Q plot to test for
linearity assumption

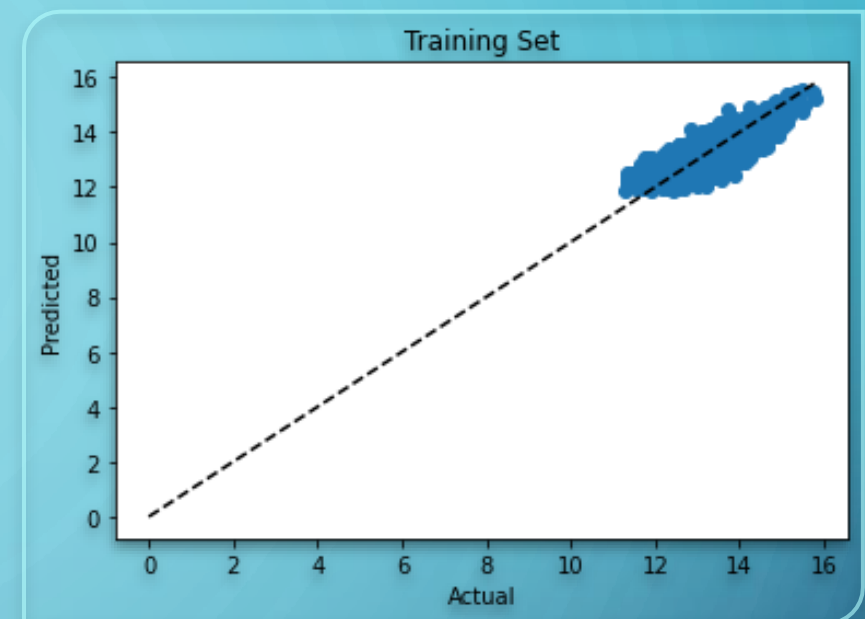
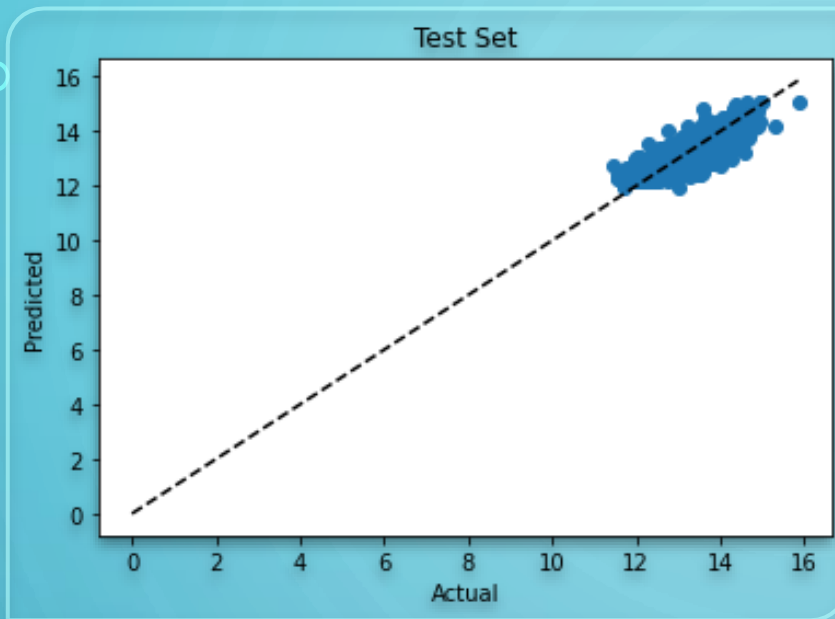
price	1.000000
Grade	0.704877
bathrooms	0.552219
bedrooms	0.353289
floors	0.313838
waterfront	0.180529
sqft_lot	0.137762
yr_built	0.081958
condition	0.039044

Correlation between price and the predictor variables Grade and bathrooms are the most related to the dependent variable "price." due to their high correlation value.

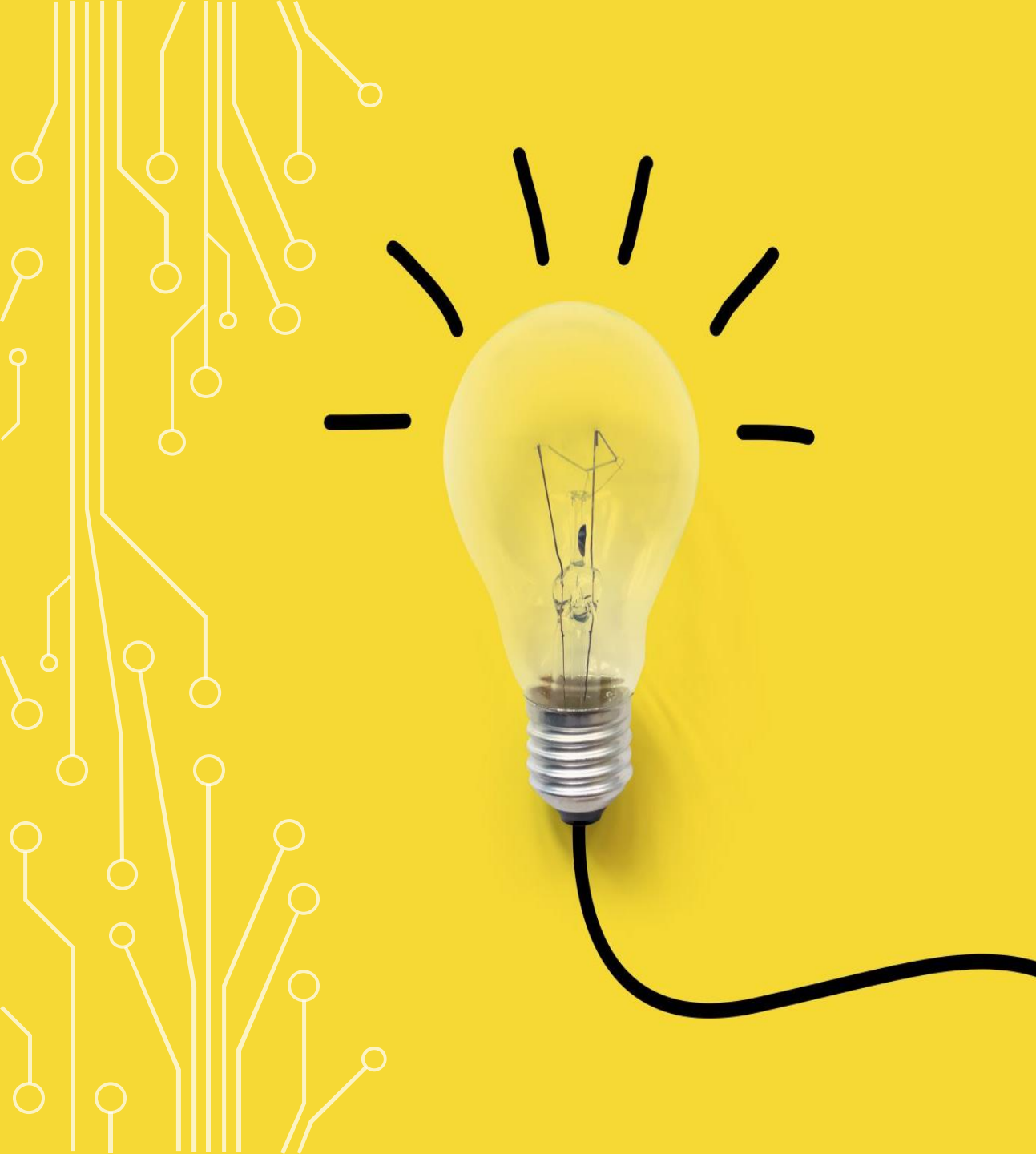
MODEL FOUR : MULTIPLE REGRESSION

- Train R-squared:
0.6430815463918367
- Test R-squared:
0.6449536912791446
- Test set performed better than train.
approximately 64.5% of the variance in the price can be explained by the predictor variables.





- Model five polynomial regression plot
 - Train set r-squared (training set): 0.6499
 - Test set r-squared (test set): 0.6527
 - Test set performed better than train. approximately 64.5% of the variance in the price can be explained by the predictor variables.



Impact of the model to the real-world problems

- The model can help sellers and buyers set realistic expectations and identify key features that have the most significant impact on the sale price.
- To incorporate machine learning techniques to build a more robust and accurate model that can identify nonlinear relationships between variables.

FINDINGS



Model five is the best model for price prediction



Larger and more spacious homes tend to be valued more highly than smaller and less spacious homes.



Although relatively weaker than other features, waterfront properties may be considered desirable by some buyers.



While newer or better-maintained homes may command slightly higher prices than older or more poorly-maintained homes, these factors are not as important to buyers as other features such as size, quality, and location.



RECOMMENDATIONS

- Adopting the model from the notebook to predict price using the predictor variables.
- Focus on improving the quality of the house - Given that the grade of the house has the highest correlation with the sale price
- Maximize the size and number of bedrooms and bathrooms
- Keep the house well-maintained - Although the condition of the house has a relatively weak correlation with the sale price, it is still important to keep the home well-maintained to maximize its value.



WHAT NEXT?

- Develop an action plan - develop a detailed action plan that outlines specific steps to be taken to implement each recommendation.
- Prioritize recommendations - prioritize the recommendations in terms of their potential impact and feasibility.
- Monitor and adjust the plan - monitor the results and adjust as needed.
- Evaluate the effectiveness of the recommendations.