

King County Housing Price Modeling Project

DSF-PT7
Phase 2 Group 19



Overview

Context: Homeowners aim to maximize property value through renovations in a competitive real estate market.

Aim: Leverage linear regression modeling to understand how different home renovations impact property value.

Approach:

- Analyze historical sales data.
- Apply regression techniques to quantify the impact of specific renovation projects on home prices.

Outcome:

- Robust Predictive Model: R^2 : 0.57 on the test set, explaining 57% of the variance in housing prices.
- Strategic renovations can substantially increase property value.



Outline

- Business Understanding
- Data Understanding
- Modeling & Validation
- Conclusions



Business Understanding

Provide actionable insights to a real estate agency assisting homeowners in buying and selling properties.

Key Objectives:

1. Develop a Predictive Model:

- Create and validate a linear regression model to predict the increase in home value based on renovations.

2. Quantify the Impact of Renovations:

- Analyze historical sales data to determine how different renovations affect property value.



Data Understanding

Dataset: King County House Sales (kc_house_data.csv)

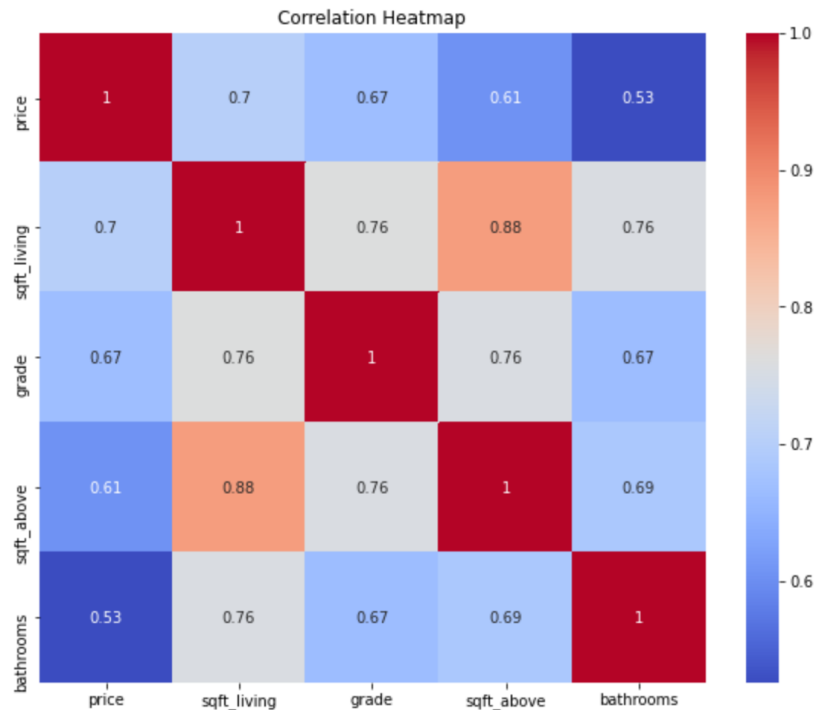
Features (some):

- Square footage
- Number of bedrooms and bathrooms
- Waterfront presence
- View quality
- Year built
- Renovation status
- Sale price (dependent variable)

Details: Column descriptions are in the column_names.md file.

Aim: Use this dataset to understand and quantify the impact of various home renovations on property values.

Data Understanding



Features Selected for Modeling:

Chosen based on correlation analysis and domain knowledge for their impact on price and relevance in real estate valuation.

1. **sqft_living** (Square Footage of Living Area):
2. **grade** (Overall Grade Given to the Housing Unit):
3. **sqft_above** (Square Footage Above Ground):
4. **bathrooms** (Number of Bathrooms):



Modeling & Validation

Approach: Iterative evaluation to improve model performance based on MAE, MSE, and R^2 .

Initial Models:

- Built four models without addressing outliers, multicollinearity, or scaling.
- Best initial model (Model 4) included features: sqft_living, grade, bathrooms, sqft_above.

Refinements: a) Addressed outliers by capping, improving metrics significantly;
b) Multicollinearity addressed by dropping sqft_above but it didn't enhance performance;
c) Scaling predictors had no significant impact.

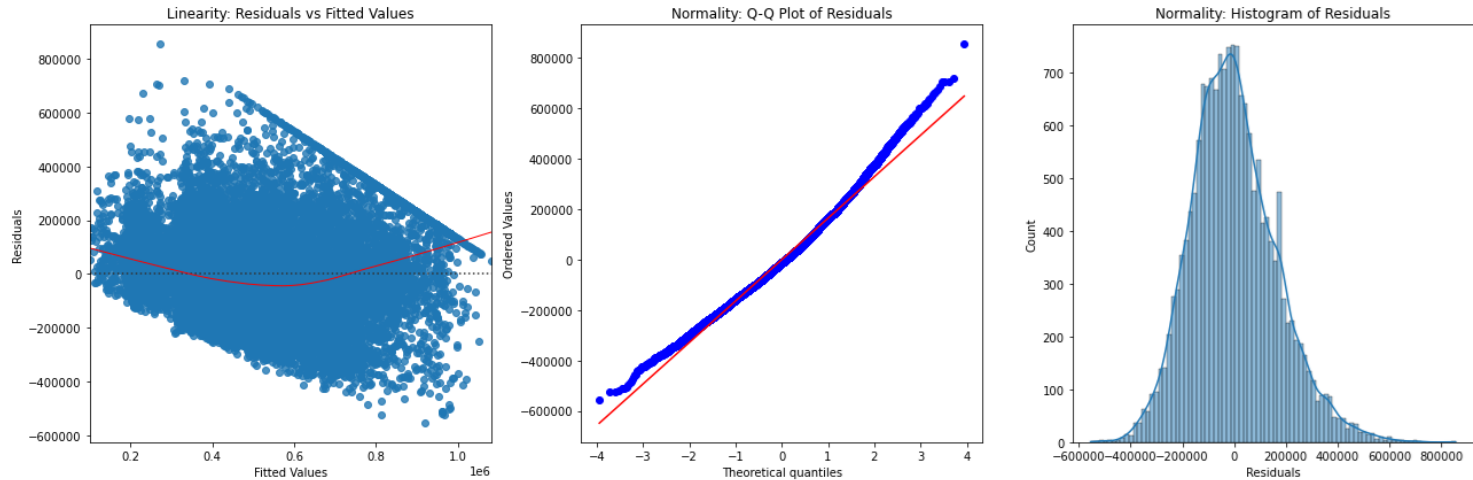
Metrics for Model 4:

- MAE: 129,160
- MSE: 27,242,617,861
- R^2 : 0.560



Modeling & Validation

Model 4 Diagnosis



Residuals show a curved pattern, violating linearity. Q-Q plot and histogram indicate deviations from normality, suggesting outliers or skewness.



Modeling & Validation

Model 5:

To improve Model 4, we log-transformed the dependent variable, scaled the predictors, removed outliers, and included bedrooms (correlation with price: 0.31) as additional feature.

Metrics for Model 5:

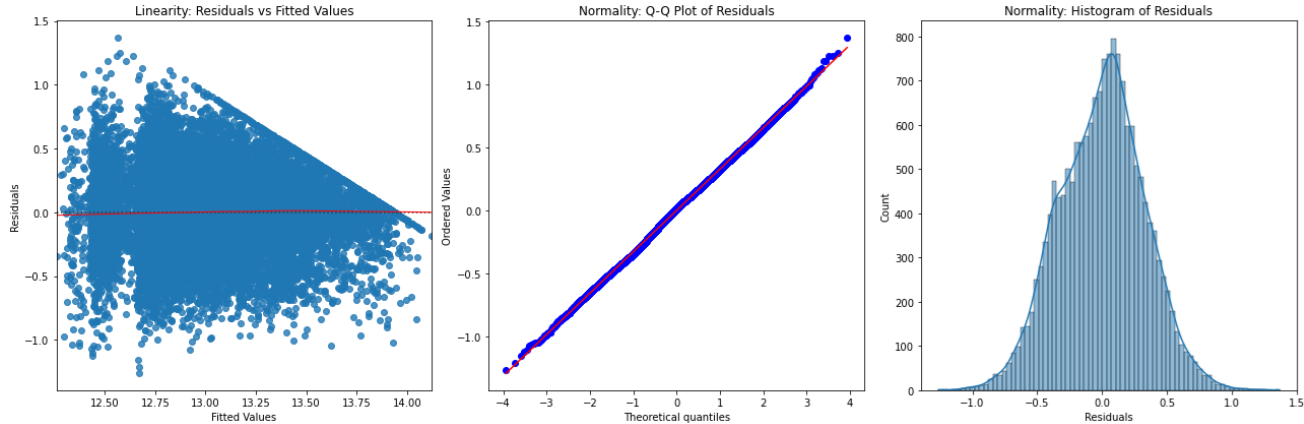
- MAE: 123,967
- MSE: 26,571,843,532
- R^2 : 0.568

Model 5 demonstrated the **highest accuracy** (lowest MAE & MSE) and **highest explanatory power** (R^2), ensuring a more robust and reliable linear regression model.



Modeling & Validation

Model 5 Diagnosis



Residuals are more randomly scattered, indicating improved linearity. Q-Q plot shows residuals closely following the red line with minor tail deviations, and the histogram is bell-shaped with reduced skewness and kurtosis, indicating a better approximation of normality.



Conclusions

Robust Predictive Model:

- Developed Model 5 predicts home value increase based on key features: sqft_living, grade, sqft_above, bathrooms, and bedrooms.
- R^2 : 0.57 on the test set, explaining 57% of the variance in housing prices.
- Strategic renovations can substantially increase property value.

Recommendations:

Use Model to:

1. Focus on High-Impact Renovations.
2. Provide data-driven Client Consultations.
3. Develop Property Marketing Strategies.

Questions?

- **Linet Maz'susa**
- **Tabitha Kariuki**
- **Brian Amani**
- **Collins Ouko**
- **Joseph Ngige**

Thank You!