GROUP 3 PRESENTATION

HOUSE PRICING ANALYSIS

KING COUNTY HOUSE DATA

Group Members:

Lilian Ngige

Sharon Njogu

Karanja Gakio

John Elvis

Content

PROJECT OVERVIEW

- In this data-driven real estate analysis project, we explore the King County House Sales dataset to assist homeowners in making informed decisions about home renovations.

- Our primary goal is to predict house prices based on a range of features, including square footage, bedroom count, bathroom count, condition, grade, and more.

- By utilizing multiple linear regression modeling, we aim to offer transparent and actionable recommendations to homeowners, helping them understand how specific renovations might impact the estimated value of their homes.

OBJECTIVES

Assist Homebuyers in Making Informed Decisions:

- Empower homebuyers with accurate price estimates so they can make informed decisions and negotiate effectively.

Develop a pricing model:

- The model should incorporate the factors such as living room square feet, number of bathrooms/bedrooms, grade ratings. This should help the agency to accurately price their properties.
- Evaluate the model's performance using appropriate regression metrics.

Identify Key Features

- Determine house which features have the most significant impact on house prices.
- Explore correlations and relationships between different features and the target variable (price)

Improve marketing strategies:

Developing focused marketing initiatives highlighting the key factors influencing prices, such as the number of bathrooms, living space, and condition and grade rate.

DATA UNDERSTANDING

- This data contains house sale prices for King County, USA.
- The data set contains 21,597 rows and 21columns.
- The data set contains 3 main datatypes that is float, integer and object.

DATA CLEANING

- This data contains house sale prices for King County, USA.
- Dropped missing values.
- Converted “grade” and “condition”columns from object to integer data type.
- Created new columns “condition_encoded” and “numerical_grade” by modifying grade and condition column.
- Removed outliers.

DATA MODELLING:BIVARIATE REGRESSION MODEL.

- With ‘sqft_living’ and ‘bathrooms’ as predictor variables:

- ***Model Summary:***

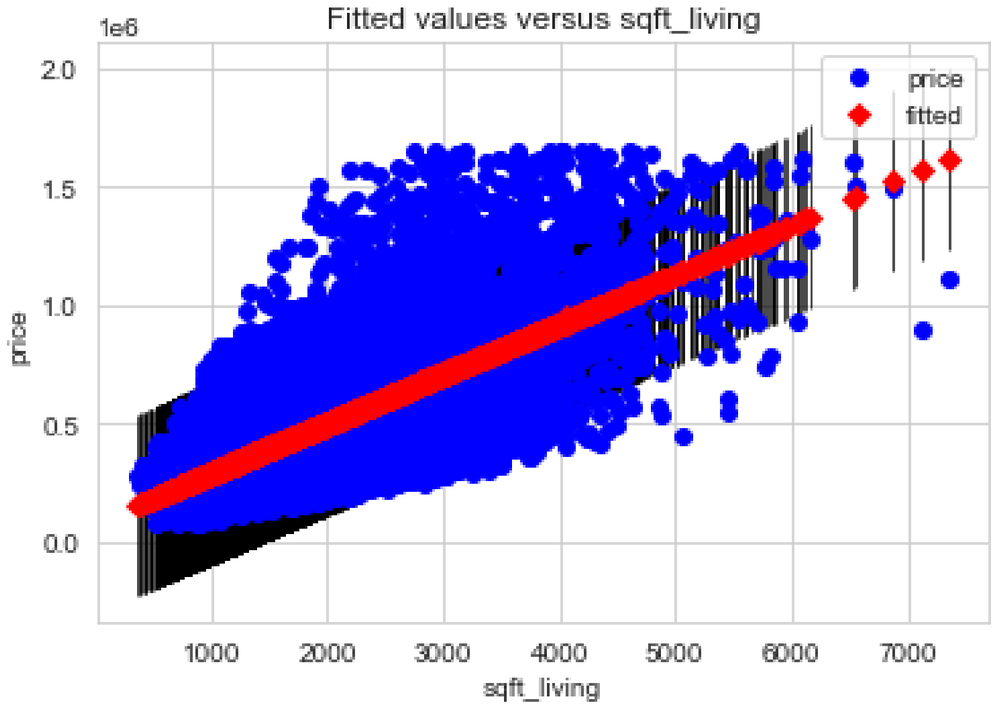
R-squared: 0.44

sqft_living coef: 206.2575

bathrooms coef: 4497.1519

Skew: 0.984

Kurtosis: 5.185



DATA MODELLING: SIMPLE LINEAR REGRESSION MODEL

Fit of the baseline model on the "sqft_living" feature as the predictor variable

Model Summary:

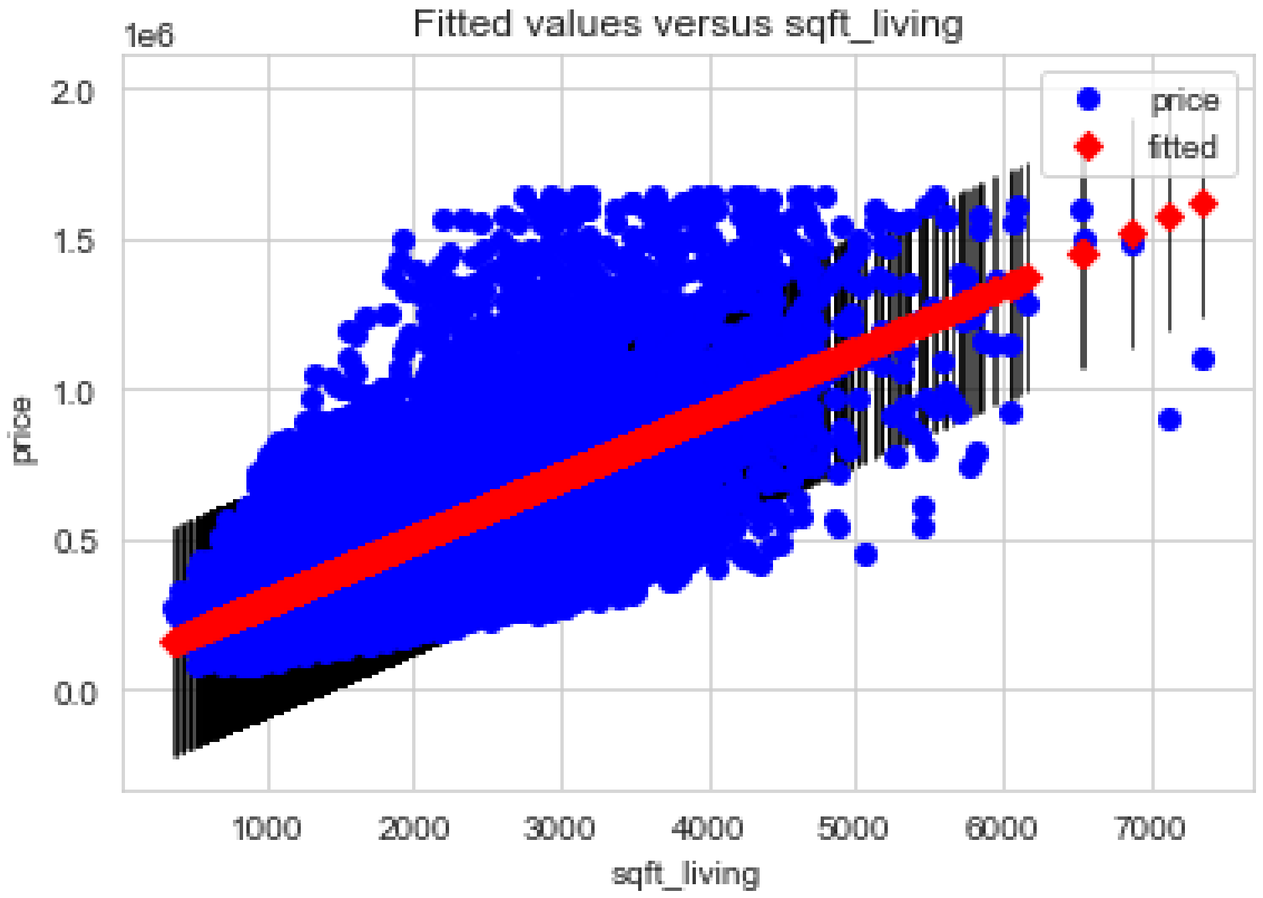
R-Squared: 0.45

P-Value: 0.00

sqft_living coef: 209.1581

Skew: 0.984

Kurtosis: 5.188



DATA MODELLING

Multilinear Regression Model.

Model Summary:

R-squared score: 0.616

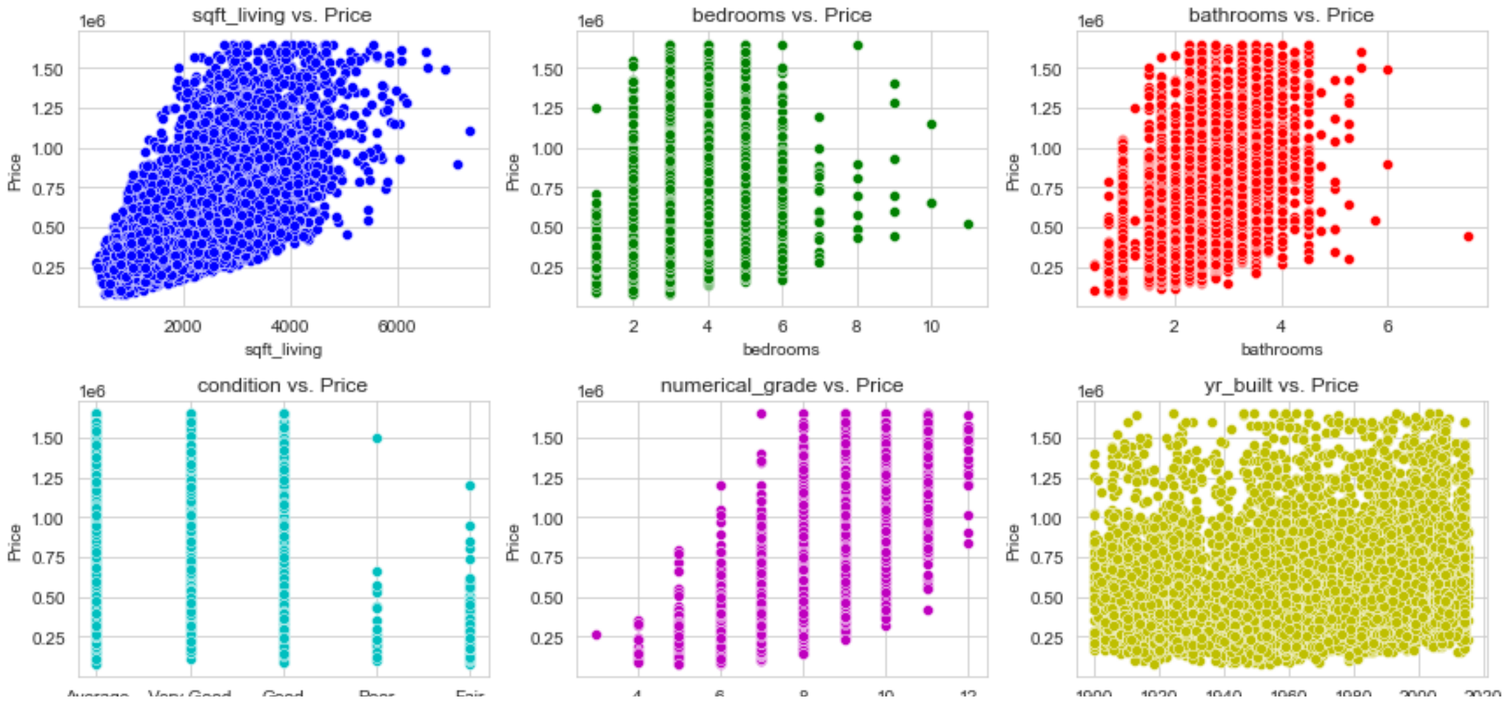
M.A.E - 119,189.93

‘floors’ co-ef: 4.38

‘sqft_living’ co-ef: 139.9

Skewness: 0.896

Multicollinearity visualization



CONCLUSION

- The final model - Multilinear Regression Model - identifies key factors influencing house prices and explains 61% of the price variations.

- Important factors affecting house prices include **living room square footage**, **number of bedrooms**, **bathrooms**,**floors** and **property grade**.

RECOMMENDATIONS

1

Focus on key features analyzed when setting house prices i.e square footage, bathroom renovations, property maintenance, and modernization

2

Consider Multi-Story Design: explore the possibility of adding additional floors or converting existing space into additional stories

3

Consider each property's unique characteristics and consult with real estate professionals for personalized advice based on local market dynamics.


THE
END