REAL ESTATE PRICE

ANALYSIS

GROUP IV MEMBERS

- MARYIMMACULATE KARIUKI
- DAVID AMABANI
- JOY WANGUI
- EJIDIAH KANYI
- MARY GACERI
- SIMON NG’ETHE

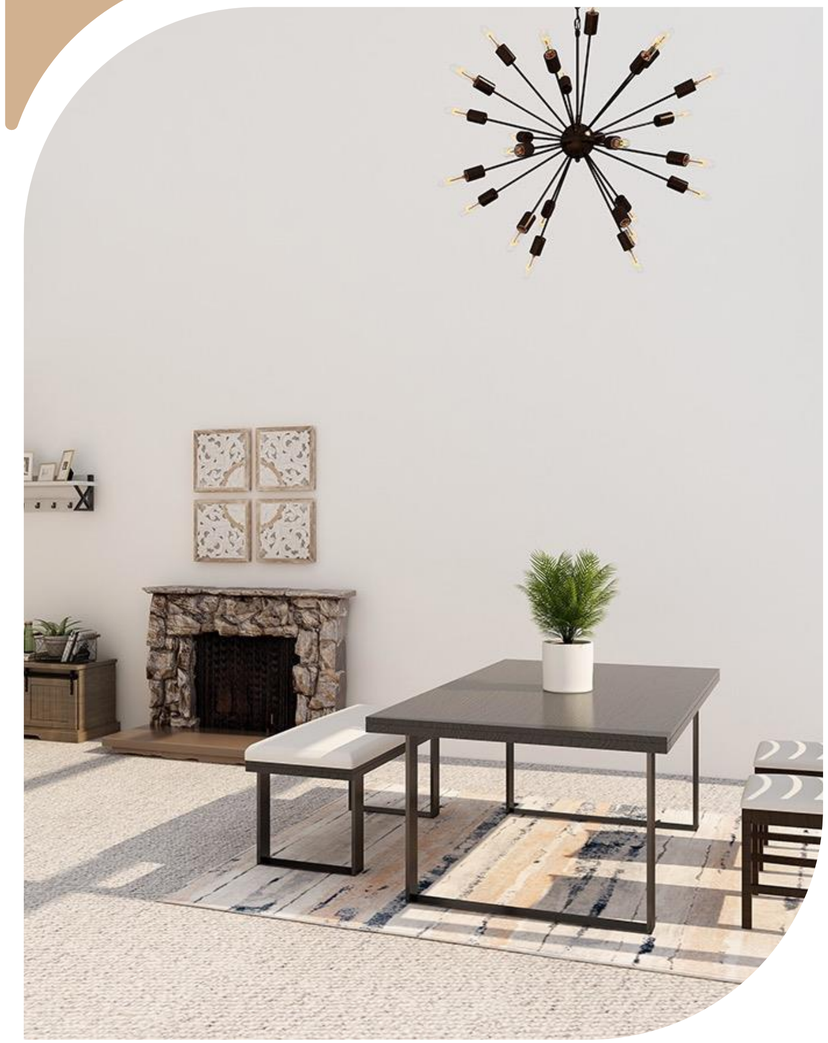
CONTENTS

1.0 OVERVIEW

This project involves a data analysis that focuses on examining housing prices. Its goal is to offer insights into the elements affecting housing prices by using a sample dataset.

TARGET AUDIENCE

- Real Estate Agencies
- Property Investors
- Homeowners
- Anyone keen on gaining insights into the housing market of Northwest County.



2.0 PROBLEM STATEMENT

Understanding house price determinants is vital. Conventional methods lack reliability in predicting price changes. Our approach advocates utilizing multiple regression models on housing sales data, revealing the interplay between various factors, enabling a comprehensive understanding of house price dynamics.

3.0 OBJECTIVES

- ❑To choose relevant data variables for the data regression model by exploring the dataset.
- ❑To create a multiple regression model to predict house prices and check model assumptions to ensure a good fit.
- ❑To analyze the coefficient of independent variables to find the most influential factors affecting house prices and describe their effects.
- ❑To check the model's performance to ensure accuracy and reliability.
- ❑To offer practical insights and recommendations for real estate investors and policymakers to make informed decisions and understand market trends.

4.0 BUSINESS UNDERSTANDING

Understanding the variables that affect home prices is one way that real estate agents can profit from regression analysis. They can find important aspects to emphasize in real estate listings, give their clients more precise price estimates, and advise sellers and buyers based on well-informed information by using the regression model. We might also take into account sellers and homeowners who can assess the prospective value of their homes using the regression model. By taking into account the influencing elements They can determine the effect of certain upgrades or renovations on the property's price and make well-informed decisions about pricing and marketing tactics by taking into account the significant characteristics found in the regression study.

5.0 DATA UNDERSTANDING

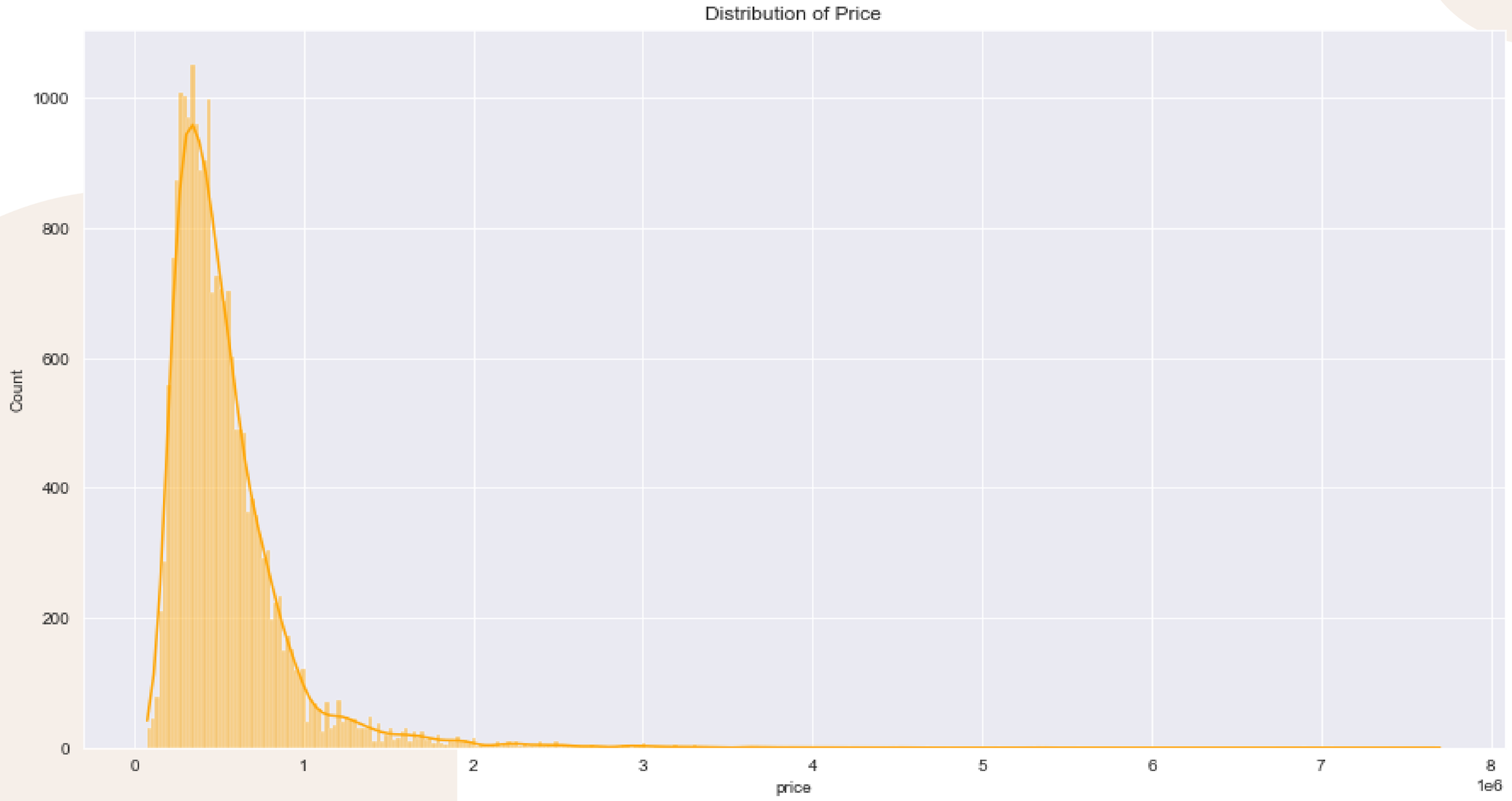
The dataset in this project has info on factors affecting housing prices like date, sqft_above, view, and sqft_basement. We have used data analysis to understand it, including handling missing data, checking types, finding outliers, and selecting important features.

6.0 DATA ANALYSIS

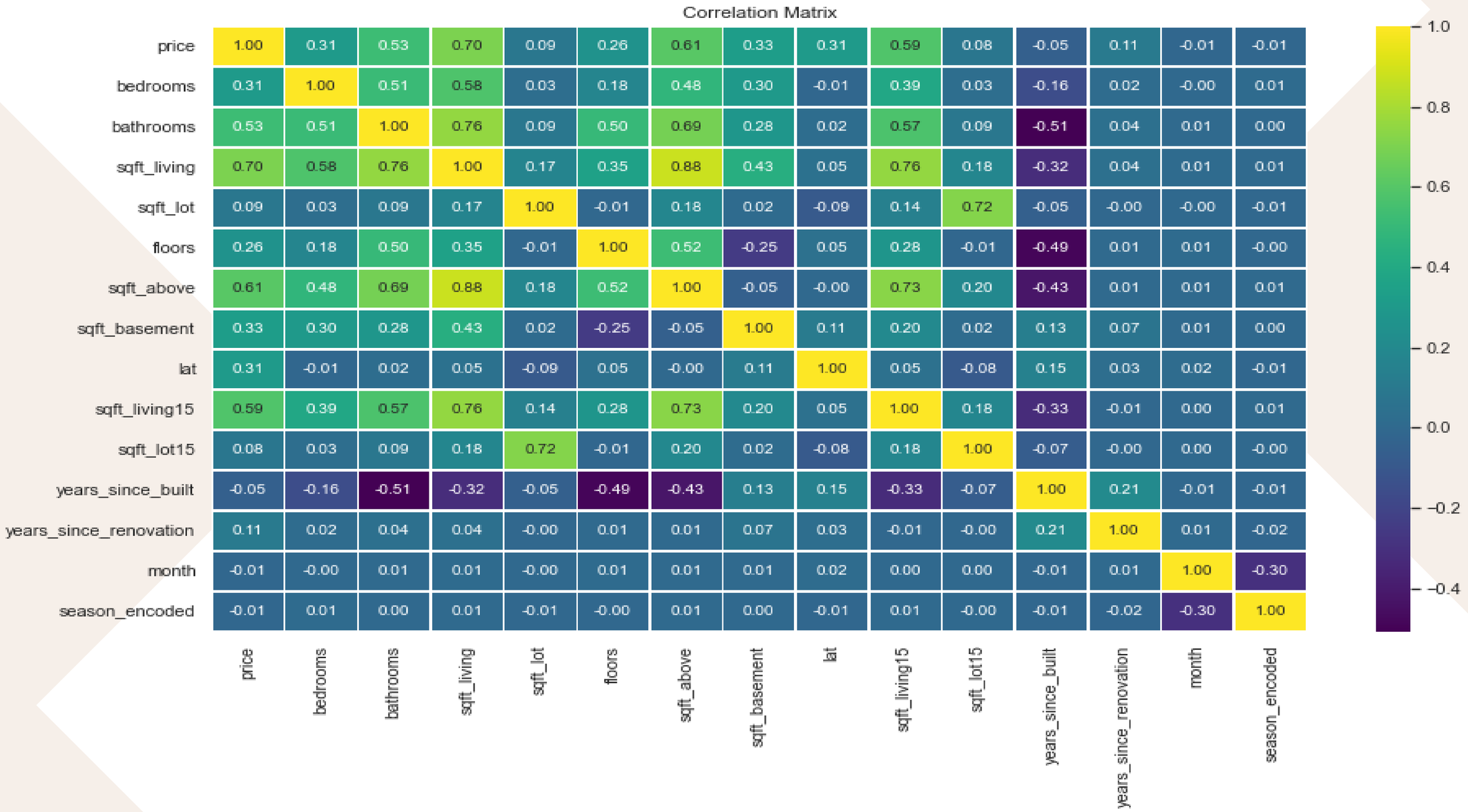
6.1 Exploratory Data Analysis

The EDA in this project involved checking dataset shape and info, exploring descriptive stats, visualizing price distribution, and examining correlations. Categorical data was encoded, and relevant columns were added for years since construction and renovation.

The Price Distribution of Houses



Correlation Matrix between Price and Other Variables



6.2 MODELLING

6.2.1 Multi Linear Regression

For this project, we created a multiple linear regression model that forecasted housing prices according to a number of different factors. The house sales dataset was used to train the model, and we used the sci-kit-learn library. Bedrooms, waterfront, location, property grade, and sqft_living were the primary independent factors in the model. We handled missing values, outliers, and feature scaling by performing data cleaning and preprocessing procedures before to training the model.

6.3 Multi Linear Regression Results

Mean Squared Error (MSE): 0.01008288564268929

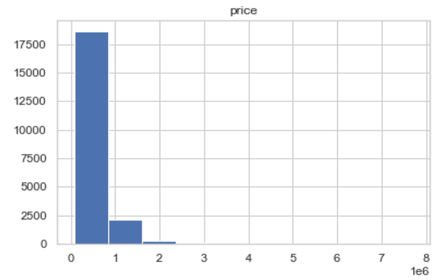
Root Mean Squared Error (RMSE): 0.1004135730003135

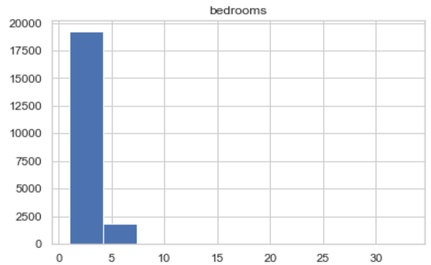
R-squared Score: 0.8204189430500062

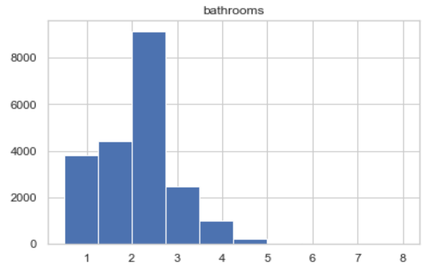
6.3.1 Multi Linear Regression Results

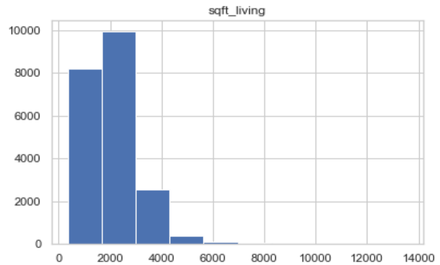
MSE measures the average squared difference between predicted and actual values. Our model's MSE is 0.0101, indicating relatively low prediction errors. RMSE is the square root of MSE, providing a more interpretable scale. Our model's RMSE is 0.1004, suggesting that, on average, our predictions are off by about 10% in price. Our model achieved an R^2 score of 0.8204, indicating that it explains approximately 82% of the variance in house prices. Our regression model performs well, with low MSE and RMSE. The R^2 score of 0.8204 shows that our model captures a significant portion of the variance in house prices.

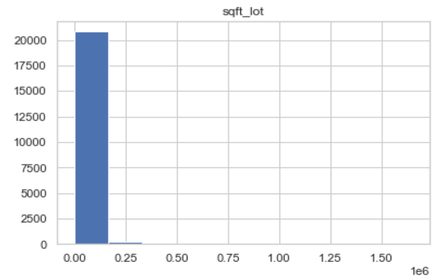
Distribution Of Predictor Variables

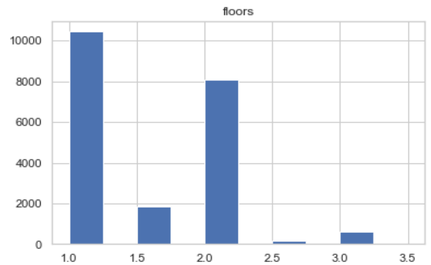


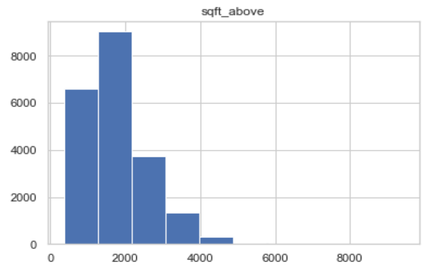


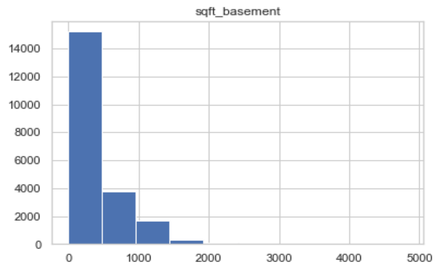


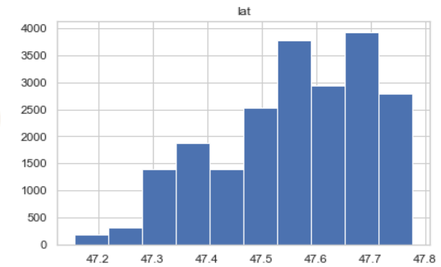


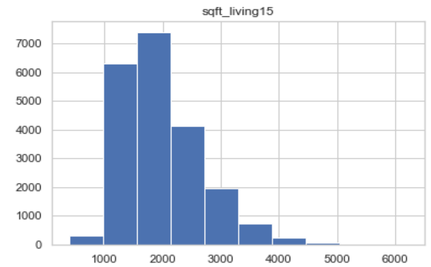


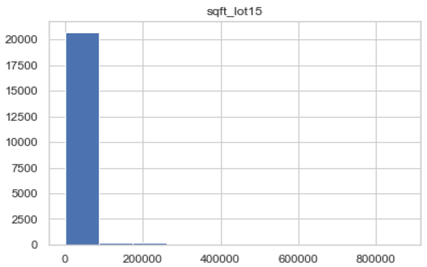


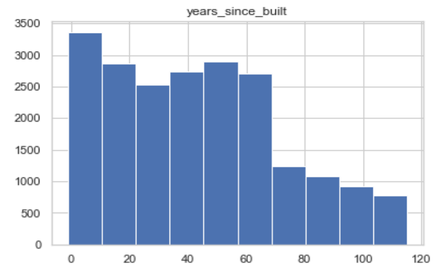


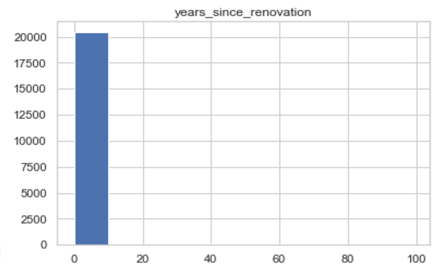


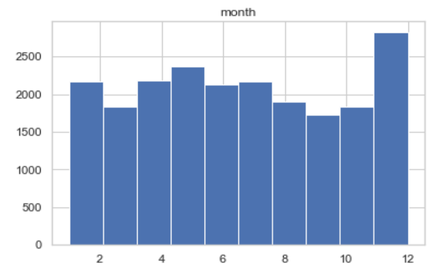


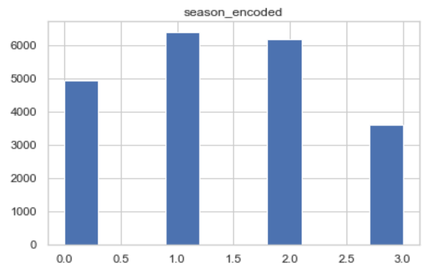












7.0 CONCLUSIONS

- Model Accuracy: The low MSE of 0.0101 shows the model makes accurate predictions, close to the actual values.
- Prediction Consistency: The RMSE of 0.1005 means the model's predictions usually have a 0.1005 unit error. This reliability is crucial for decision-making.
- High Explained Variance: An R2 score of 0.8201 means the model explains around 82.01% of the target variable's variation, showing a strong fit to the data.
- Model Effectiveness: The model is good for accurate predictions (low MSE and RMSE) and explains a lot of the variance (high R2), making it suitable for prediction.
- Room for Improvement: The model is strong but leaves 17.99% unexplained variance. This means we can make it better with refinements and more feature analysis to understand the target variable factors.

8.0 NEXT STEPS

- Utilize the model's accuracy to enhance property valuations and guide price recommendations for sellers.
- Leverage the prediction consistency to offer reliable insights and assist clients in making informed decisions.
- Identify potential areas for improvement highlighted by the unexplained variance to refine investment strategies.
- Gain insights into potential property value through the model's effectiveness.
- Use identified influential factors to evaluate home improvements for maximizing property value.

Thank You !!!