REGRESSION MODELLING PROJECT ON PREDICTING HOUSE PRICES IN KING COUNTY.

PROJECT DONE AND SUBMITTED BY:

1. AKOKO JIM.
2. STEPHEN KAHUTHU.
3. CAMILLA LUMWAJI.
4. GLORIA TISNANGA.
5. PETER NJUGU.

IN FULFILMENT OF DATA SCIENCE PHASE TWO COURSE, MORINGA SCHOOL.

22nd JULY. 2024.

INTRODUCTION.

- Welcome to our data driven exploration of the King County House Sales dataset!
- In this project, we'll delve into fascinating world of real estate, armed with regression modelling techniques.
- Our mission is to unravel the hidden patterns, predict housing prices and provide valuable insights to stakeholders.

PROBLEM STATEMENT.

- In King County, a bustling real estate agency aims to empower homeowners by providing essential insights into how renovations impact property values.
- The agency seeks to determine the financial implications of popular renovations such as kitchen remodels, bedroom additions, and backyard transformations on the asking prices of homes.
- By understanding these dynamics, homeowners can make informed decisions to enhance the market value of their properties effectively.

OBJECTIVES.

- To develop a Predictive Model for Homebuyers to estimate housing prices based on key features.
- To develop an Investment Strategy for Sellers and Investors by identifying features that significantly impact housing prices.
- To analyze neighborhood level trends and disparities to get insights for urban planners.

DATA UNDERSTANDING.

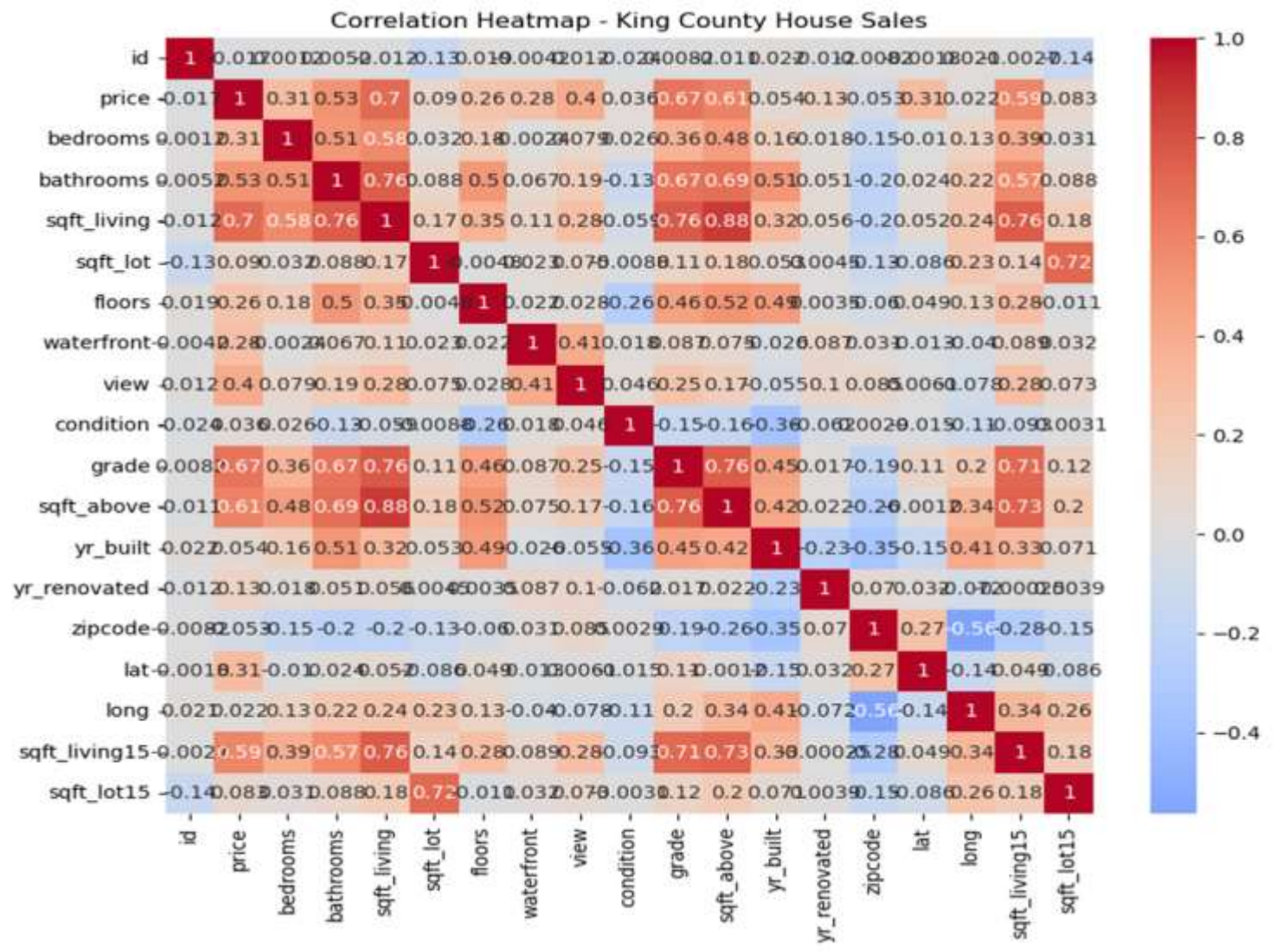
- **Data Sources:**
  - Comprehensive real estate transaction data from King County.
  - Details include property characteristics, sale prices, renovation specifics, and geographic locations.
- **Relevance to Agency Objectives:**
  - **Empowering Homeowners:** Provides insights into the impact of renovations (kitchen remodels, bedroom additions, backyard transformations) on home values.
  - **Strategic Decision-Making:** Enables data-driven recommendations to enhance property marketability and value.
  - **Market Insights:** Facilitates understanding of local market dynamics and trends in King County real estate.
- **Data Attributes:**
  - Property features: Size, age, condition.
  - Renovation details: Type, cost, timing
- **Analytical Approach:**
  - Statistical analysis: Regression models to quantify the effect of renovations on home prices.
  - Data visualization: Charts and graphs to illustrate trends and correlation
- **Expected Outcomes:**
  - Clear understanding of how specific renovations impact home values in King County.
  - Insights to guide homeowners in optimizing property investments through strategic renovations.
  - Enhanced ability to advise clients on maximizing property marketability and profitability.

DATA PREPARATION.

To get better output form our data it was necessary to clean the data for better output here are some ways we dealt with this:

- **Data Collection:** Comprehensive real estate dataset sourced from King County records. Includes property attributes (e.g., size, condition), transaction details (e.g., sale price, date), and neighborhood characteristics (e.g., schools, amenities).
- **Data Cleaning:** Address missing values in critical fields (e.g., price, property size) using appropriate imputation techniques. Standardize data formats and units across all variables for consistency.
- **Feature Engineering:** Create new features
- **Predictive Modeling for Homebuyers:** Focus on key predictors like square footage, number of bedrooms, and neighborhood indicators.
- **Investment Strategy for Sellers and Investors:** Identify impactful features such as lot size, proximity to amenities, and property condition.
- **Neighborhood Insights for Urban Planners:** Aggregate data to assess neighborhood-level trends in socioeconomic factors, school quality, and community amenities.

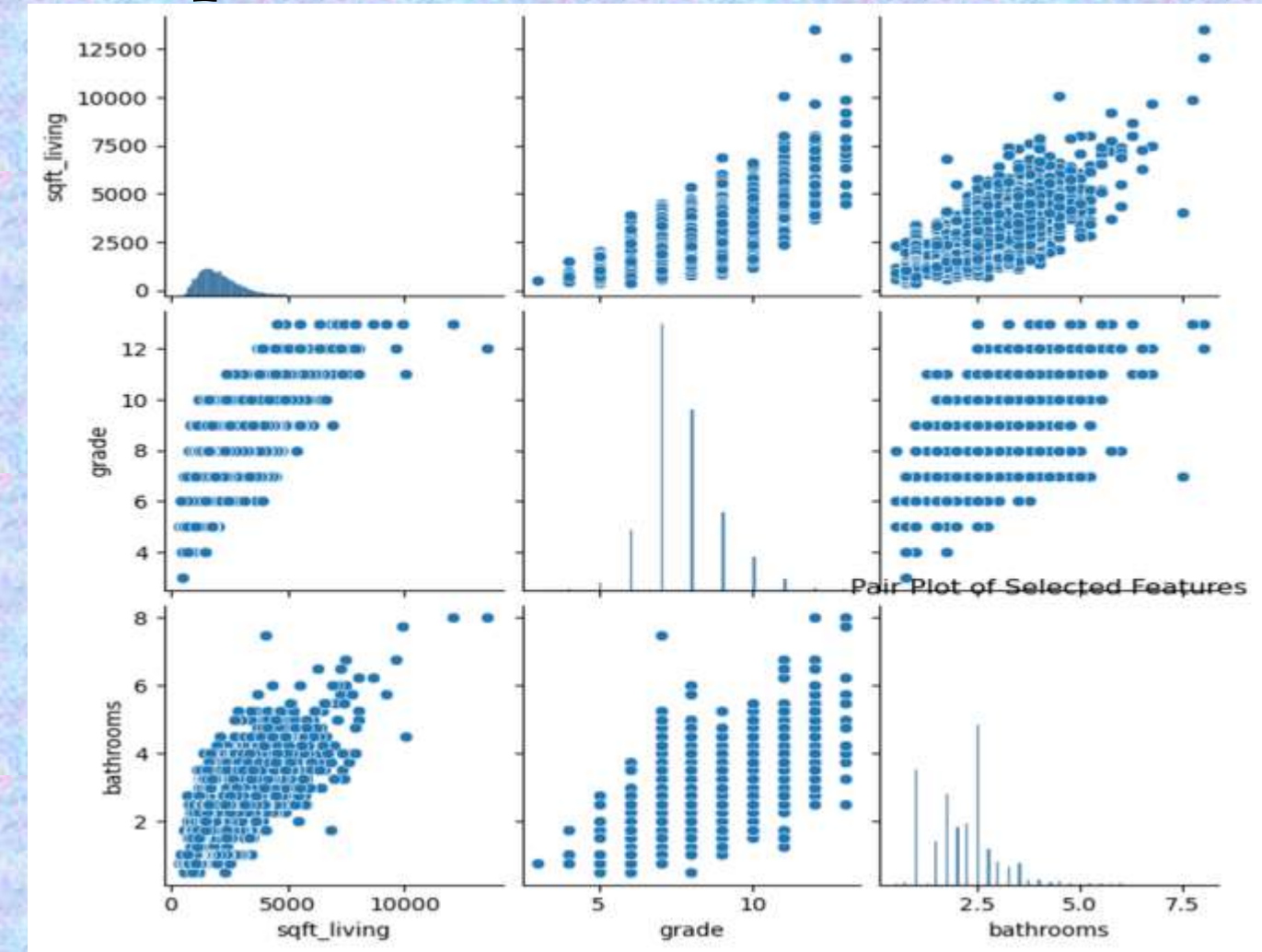
Correlation heatmap.



This graphical tool displays the correlation between various variables in the King County House Sales dataset to establish key features affecting price.

From this heatmap we are able to determine the features with the most correlation to price as;
sqft_living=0.7,
bathrooms=0.53,
grade=067,
sqft_0.61
and
sqft_living15=0.59

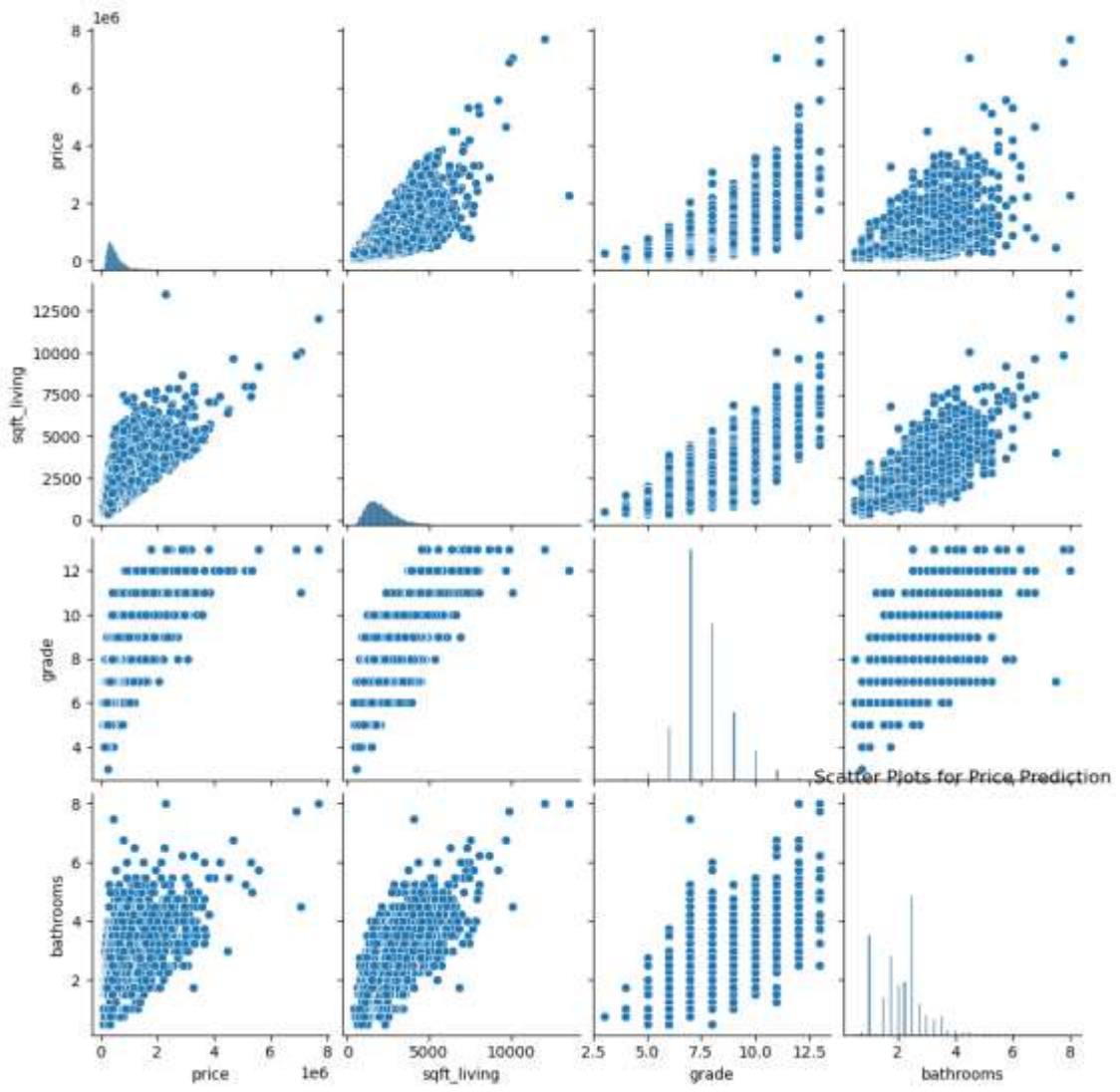
Pair plot for the selected features.



This pair plot is a matrix of graphs that enables the visualization of the relationship between each pair of selected variables in the dataset.

DATA ANALYSIS: 1. Predictive Modelling for Homebuyers.

Linear regression for Price prediction to estimate housing prices based on selected features.



Trend Observation;

- ✓As the square footage increases, the price tends to increase.
- ✓Higher-grade houses tend to have higher prices.
- ✓The relationship between price and the number of bathrooms shows that houses with more bathrooms tend to have higher prices.

Predictive Modelling.

Comparison of Outcomes

Iteration 1: Initial attempt with polynomial features resulted in an OLS R-squared of 0.565, indicating a moderate fit.Iteration

Iteration 2: Using cross-validation to find the optimal polynomial degree improved the R-squared to 0.587.Iteration

Iteration 3: The inclusion of one-hot encoded zipcode and standardization of features led to a substantial improvement, achieving an OLS R-squared of 0.739.

The progression from a fixed polynomial degree to an optimized polynomial degree and finally to including categorical variables with one-hot encoding and standardization has shown clear improvements in model performance. Each iteration's enhancements, particularly the feature engineering step in the final iteration, have contributed to a more robust model with a significantly higher R-squared value.

CONCLUSION.

- Square footage of living area is a strong predictor for house pricing. Larger houses are generally valued more.
- Grade, which reflects overall quality or construction standards, is also an important factor in determining price.
- While the number of bathrooms contributes to price, it may not be as significant as sqft_living or grade.

RECOMMENDATIONS.

1. For Homebuyers:

- Prioritize Square Footage: When evaluating properties, pay close attention to the square footage (sqft_living). Larger living spaces often correlate with higher prices.
- Consider Trade-offs: Balance other features (e.g., grade, condition) against square footage. Sometimes a smaller but well-designed home might be more appealing.

2. For Sellers and Investors:

- Quality Matters: Invest in improving the overall quality (grade) of your property. High-quality finishes, materials, and design can positively impact sale prices.
- Highlight Living Space: Emphasize the square footage in your property listings. Showcase spacious living areas to attract potential buyers.

3. For Urban Planners:

- Neighborhood Development: Prioritize neighborhoods with larger living spaces and higher-quality homes. These areas tend to attract stable communities and contribute positively to the city's image.
- Mixed-Use Planning: Encourage mixed-use developments that incorporate both residential and commercial spaces. A vibrant neighborhood with amenities benefits everyone.

4. Data-Driven Decision-Making:

- Monitor Trends: Continuously analyze housing market data to identify shifts in preferences. Adapt planning strategies accordingly.
- Community Engagement: Involve residents in planning discussions. Their insights can guide equitable development.