**Report**

**I. Dataset Selection**

The dataset "House Sales in King County, USA" contains data on house sales in King County, Washington. It includes various features such as price, number of bedrooms, location and square footage.

**II. Research Question and Hypothesis**

Research Question: “How do location (zipcode, waterfront), size (sqft\_living, sqft\_lot), and quality (condition, grade) impact house prices in King County, and which features are the most significant predictors of house price variability?”

Hypothesis: “Larger houses with better conditions, located near water, built or renovated recently, tend to have higher sale prices.”

The goal of the case study is to explore the dataset, identify key pricing factors and provide actionable recommendations.

**III. Data Preprocessing and EDA**

The dataset had no missing values, so no imputation was needed. The date was converted to a datetime format, and features like year and month were extracted. Outliers in price, square footage, and bedrooms were addressed. Key insights include:

* Size and Quality are the most influential factors on price.
* Larger homes with higher quality (higher grade and condition) tend to be more expensive.
* Location also plays a crucial role, with zip codes and proximity to water significantly affecting house prices.

Visualizations:

* Scatter plot: sqft\_living vs. price.
* Boxplot: price distribution across condition, grade, and zipcodes.
* Histogram: Distribution of price and sqft\_living.
* Interactive map: Price by latitude and longitude.
* Time-series analysis of price trends over time.

Insights:

* Larger homes typically have higher prices, as shown in the scatter plot and histogram.
* Houses in better condition or with higher grades are more expensive (boxplot).
* Waterfront and urban center properties command higher prices, as indicated by the map.
* Price trends show seasonal fluctuations, with higher prices generally in summer months.

**IV. Methodology**

1. Model Selection

Applied Linear Regression and Random Forest models. The Random Forest model outperformed Linear Regression in terms of both RMSE (Root Mean Squared Error) and R² score.

2. Feature Engineering

Created derived features such as house age (current year - yr\_built), renovation status and price per sqft. Categorical variables like waterfront and zipcode were one-hot encoded.

3. Evaluation Metrics

Random Forest outperformed Linear Regression due to its ability to capture non-linear relationships and interactions between features. While Linear Regression is sensitive to outliers and assumes a linear relationship, Random Forest provides better predictive accuracy.

4. Visualization

Residual plots and a feature importance plot for Random Forest were used to check model assumptions.

**V. Interpretation and Recommendations**

Findings

* House prices generally rise over time with seasonal fluctuations, influenced by economic factors.
* Proximity to desirable amenities (waterfront, urban areas) and larger properties result in higher prices.
* Homes with better conditions and higher quality (higher grades) tend to have higher prices.
* Random Forest outperforms Linear Regression due to its ability to model complex, non-linear relationships.

Recommendations

* For buyers: Focus on location, size, and condition when making decisions. Consider timing sales based on seasonal trends.
* For sellers: Emphasize location, size, and quality when presenting properties.
* Invest in areas with high demand, like waterfront or urban proximity.
* Promote policies that support quality construction and home maintenance.

**VI. Conclusion**

This analysis confirms that location, size, and condition are key drivers of property prices. Proximity to water and urban areas, along with larger properties and better construction quality, significantly influence house prices. Machine learning models, particularly Random Forest, provide effective predictions by capturing complex relationships within the data. Understanding these factors is crucial for making informed property investment and sale decisions.