

King County Real Estate: Crafting Competitive Pricing Strategies

Authors: [Felix Mburu Njoroge, Winnie Osolo, Violet Musyoka, Alex Korir, Lydia Masabarakiza]

Project Overview

A commercial real estate agency is interested in investing in the real estate within King County. Their main concern is having an overview of prices of houses within the area and identify some of the attributes that influence the pricing strategy

Business and Data Understanding

A real estate agency wants to come up with a pricing strategy. They want to do this by analyzing the factors that most significantly impact house prices.

In order to achieve this, we are required to come up with a model that is capable of deducing the main factors that influence the house prices. The client seeks to utilize accurate and representative data pertaining to the real estate market in King County. This data contains historical sales, size of the property and other pertinent features

Modelling

Data Preparation

Data Cleaning: Removing any duplicate or irrelevant entries. Handle missing or null values appropriately

Exploratory Data Analysis (EDA)

We used descriptive statistics and visualizations to summarise the main characteristics and examine relationships between the features and our target variable (price). - Measures of Central Tendency and Dispersion - Univariate - Bivariate - correlation - Feature engineering - Categorical Data Encoding - one hot encoding, and label Encoder

Feature Selection

Correlation Analysis was used to identify highly correlated features with the target (price).

Model Selection, Training and evaluation

Baseline Model: Simple linear regression model to establish a baseline performance.

Data Splitting: Split the dataset into training and testing sets (e.g., 80% train, 20% test).

Model Training: Train each selected model on the training data.

Model Evaluation: Evaluate the models using appropriate metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared to measure predictive performance.

Regression results

Model Performance

R-squared: The model explains approximately 58.1% of the variance in house prices based on the selected features. This indicates a moderate level of explanatory power.

F-statistic: The F-statistic is 2660 with a p-value close to 0, suggesting that the model is statistically significant in predicting house prices.

Coefficients Interpretation

Intercept (const): The intercept term is not practically meaningful as it assumes unrealistic scenarios (e.g., zero square footage, bedrooms, bathrooms, etc.).

sqft_living: For every additional square foot of living space, the house price is expected to increase by \$248.26, holding all other variables constant.

bedrooms: Each additional bedroom is associated with a decrease in house price by \$48,620, on average, holding all other variables constant.

bathrooms: Each additional bathroom is associated with an increase in house price by \$38,590, on average, holding all other variables constant.

floors: Each additional floor is associated with an increase in house price by \$6,398.69, on average, holding all other variables constant.

waterfront: Houses with waterfront property have an expected increase in price by \$579,800, holding all other variables constant.

view: Each additional level of view rating is associated with an increase in house price by \$64,210, on average, holding all other variables constant.

condition: Each additional level of condition rating is associated with an increase in house price by \$47,390, on average, holding all other variables constant.

grade: Each additional level of grade rating is associated with a decrease in house price by \$20,270, on average, holding all other variables constant.

Interpretation Notes

The negative coefficient for bedrooms and grade might seem counterintuitive but could be influenced by other factors not included in the model. Factors like waterfront property, better views, and excellent condition tend to increase house prices significantly.

Conclusions and Recommendations

1. **Incorporate identified factors into pricing strategy:** Adjust listing prices based on factors like number of bathrooms, presence of waterfront views, and overall condition of the house to align with market demand and maximize profitability.
2. **Utilize insights for targeted marketing:** Tailor marketing efforts towards properties with desirable features such as waterfront views or higher-grade ratings to attract potential buyers and optimize sales.
3. **Consider local market dynamics:** Recognize the significance of location, as evidenced by the best-priced houses being in the central region of Kings County. Utilize this insight to focus resources on areas with high demand or potential for growth to maximize returns.

Next Steps

1. Develop advanced predictive models to forecast future house prices based on historical data and key predictors identified in the analysis. These models will provide valuable insights for pricing decisions and market forecasting.
2. Implement market segmentation techniques to identify distinct buyer segments based on preferences and demographics. Tailoring marketing strategies and pricing approaches to these segments will enhance targeting and effectiveness.
3. Conduct spatial analysis to identify spatial patterns and trends in house prices across different neighborhoods. Geographic visualization techniques will help pinpoint hotspots and areas of potential investment.

For More Information

See the full analysis in the [Jupyter Notebook](#) or review this [presentation](#).

Repository Structure

|— content |— README.md |— Phase 2 project Group 12.pdf |— Phase 2 project Group 12.ipynb