Regression Modelling

Paul Mbugua

Ronnie Muriithi

Victor Gachie

Mercy Tegekson

Winnie Mauti

Robinson Mumo

Introduction

- The real estate industry is complex and dynamic
- Multiple players like homeowners, buyers, and real estate agents face the challenge of establishing the fair market value of a house
- Inaccurate pricing affects both buyers and sellers
- Utilizing a multilinear regression model can provide a data-driven solution to the challenge
- We seek to provide a tool that assists stakeholders in determining house prices

Challenges

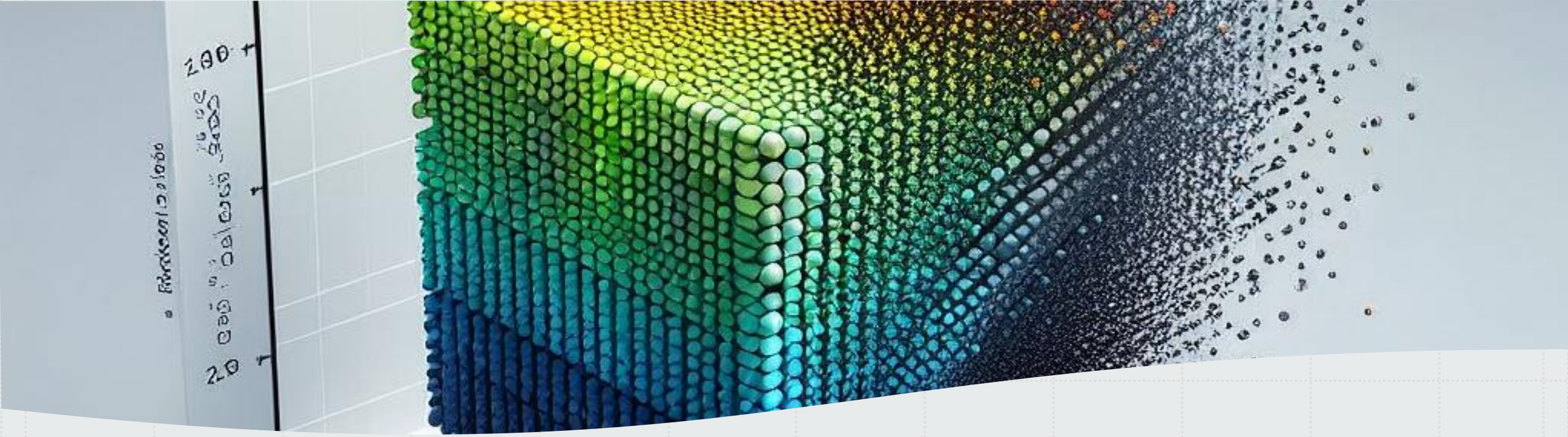
- Overpricing of houses given a lack of a robust housing pricing model
- Overpricing causes the properties to remain unsold for long periods of time
- Prospective buyers face the challenge of establishing a property best suiting their budget

Problem Statement

- To develop a multilinear regression model with the highest possible accuracy to predict the house prices
- To identify the key factors that affect house prices in King County, California
- To analyze the performance of the model using metrics like mean squared error, R-squared, and residual analysis
- To provide actionable recommendations to the Real Estate Agency so that they can improve their profitability and market share

Baseline Model

- The baseline model is a simple linear regression model of the form:
- y = b0 + x*b1; So,
- Price = $100,500 + 0.7434 * Average price per zip code
- The model explains 46.01% of the variation in price
- The root mean squared error, which means that the model's predictions are off by about $124,066 in terms of house prices



Multiple Linear Regression

- The model is of the form
- $y = b0 + b1*X1 + b2*X2 + \ldots + bn*Xn$
- Price = $491,419.84 + 0.2023 * Average price per zip code + 25,647.60 * square feet living + 114.98 * total square feet of the house - 69.62 * square feet of houses for the nearest 15 neighbors + 5783 * bathrooms + 1,292.73 * price per square feet + 7,220.67 * bedrooms - 1,802.29 * floors + 25,444.25 * view - 45.36 * square feet of the basement + 72,121.07 * has basement + 1,116.40 * year renovated - 2,191,957.02 * renovated + 62,300.17 * waterfront + 16,988.52 * condition

Performance of the Model

- The model explained 90.4% variation in price
- The mean absolute error was approximately $124,066 in terms of house prices
- The RMSE in the model is approximately $71295
- Interpretation: Similar to MAE, a lower RMSE is preferred

Interpretation of the Model

Model **Overview**
The model predicts property prices with an intercept of $491,419

Key Factors Affecting **Price**
Living Space: $25,790 increase per square foot.
Bedrooms: $6,716.56 increase per bedroom.
Bathrooms: $5,756.37 increase per bathroom.
Year Renovated: $1,116.40 increase per year renovated

Influential Features
Neighborhood View: $25,444.25 increase per unit.
Basement: $72,121.07 increase for properties with a basement.
Waterfront: $ 62,300.17 increase for properties with waterfront

Recommendations

- The Real Estate Agency should consider increasing the living spaces of the houses to improve their market values
- They should consider investing in houses with a nice view
- They should consider investing in properties with the basement
- They should consider investing in houses with a waterfront

Conclusion

- The real estate market is dynamic and complex
- There is a need for a robust model that predicts the prices of the houses
- We developed a multilinear regression model to explain the relationship between house prices and other variables
- The Real Estate Agency should consider investing in houses with a nice view, large basement, and waterfront