GROUP 3 PRESENTATION
PHASE 2 PROJECT

KING COUNTY HOUSE PRICE PREDICTION ANALYSIS

GROUP MEMBERS

ANDREW BARAKA

BRIAN MAINA

JOAN MAINA

MARGARET NJENGA

PURITY KIBAKI

PROJECT OVERVIEW

For this project, we will use linear regression modeling to analyze house sales in King county.

GOAL

To provide valuable insights and recommendations that will aid in informed investment decisions.

AUDIENCE

Diverse Buyers: Homebuyers such as Families, Young Professionals, and Retirees

DATASET

King County House Sales dataset

BUSINESS UNDERSTANDING

Informed Decision-Making: Our predictive model uses various house features, and locational data to empower homebuyers with insights for informed purchasing decisions.

Key Influencing Factors: Understanding market trends, structral impacts, and locational variables is essential for informed homebuying decisions.

Maximizing Investments: Thorough research, predictive modeling, and guidance from professionals are crucial for optimizing investment potential in real estate.

Objectives

What features are the best predictors of house prices, and how might they impact their decision-making?

How can they optimize their investment potential in the real estate market?

DATA UNDERSTANDING

Meaning of the data in different columns

Date:
Date a house was sold
The data house prices was collected in 2014 and 2015

Bathrooms:
Number of Bathrooms and from range from 1 to 5 bathrooms

Price:
Sales Price
The average selling price of a house is $ 540, 297

Bedrooms:
Number of Bedrooms and they range from 1 to 6 bedrooms

Year Built:
The Year the house was built

Condition:
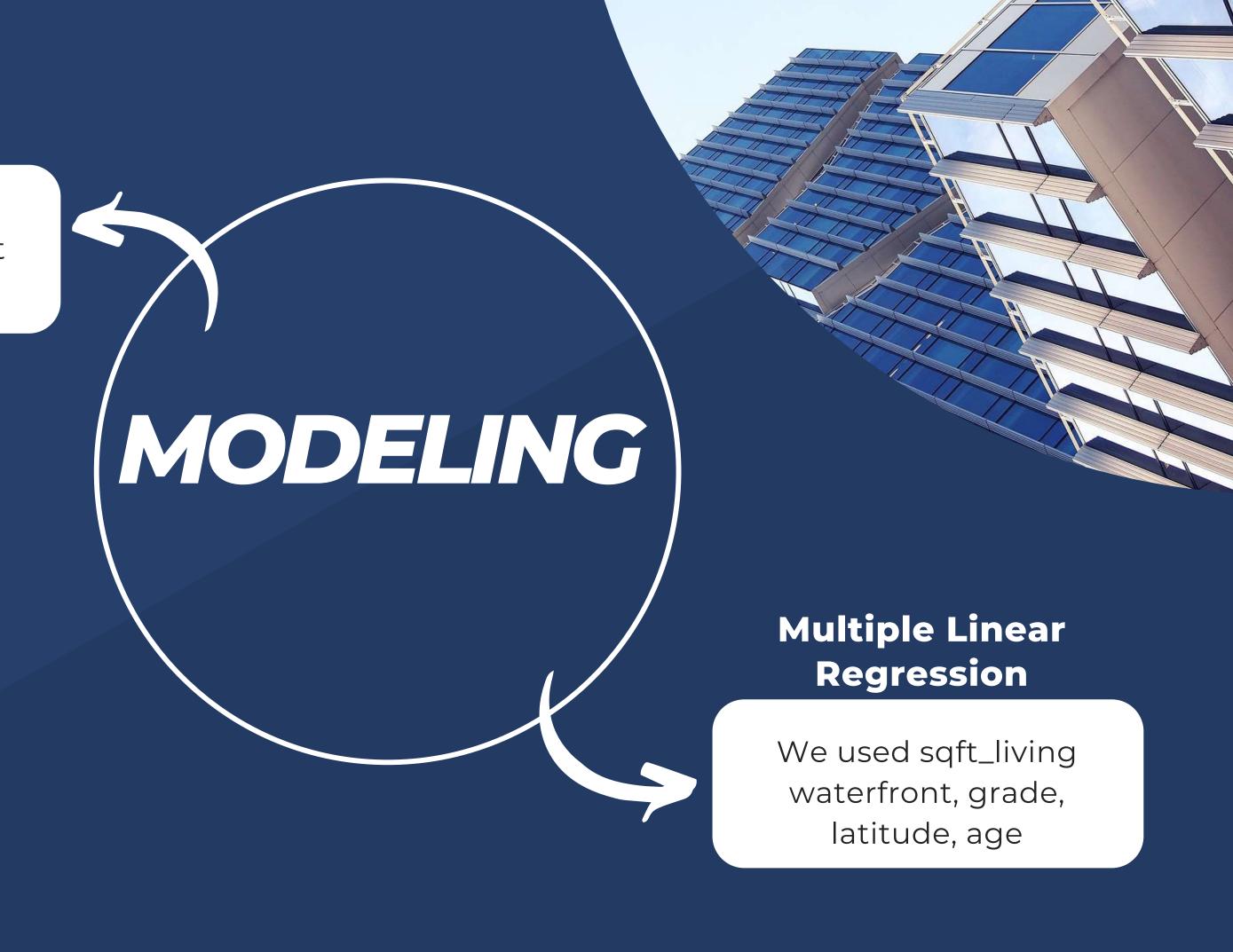Overall Condition of the House i.e
Average, Very Good, Good, Poor and Fair

Sqft_Living/Lot/Above:
Square footage of house part

Waterfront:
Whether the house is on a waterfront

Simple Linear Regression

The analysis indicates that the simple linear regression employed demonstrates that 1 unit increase in Sqft_living raises the of value price by $281 according

The r-squared value, 0.493, shows that the model can account for about 49.3% of the variance of price
The null hypothesis for the simple linear regression is t hat there is no relationship between Sqft_living and price.
The p-values is less than alpha, which means we can reject the null hypothesis

MODEL RESULTS

Multiple Linear Regression

We created a model to predict price using sqft_living + waterfront + sqft_living + waterfront + grade + latitude + Age

From the model we find that:
- Larger living spaces correlate with higher prices.
- Properties with a waterfront command premium prices.
- Higher property grades signify better quality construction and higher prices.
- Location plays a significant role; certain locations have higher demand and prices.
- Newer properties generally have higher prices compared to older ones.

MODEL RESULTS

Polynomial Regression

We did a Polynomial Regression up to the 4th degree.

The 3-degree polynomial regression model performs best in predicting house prices.

This model explains 77% change in house price.

MODEL RESULTS

CONCLUSION

The polynomial model did the best in predicting house prices.

The square footage of a house is significantly related to its price, as evidenced by the linear regression model which accounts for 49.3% of the variance in housing prices, highlighting its importance as a factor in the housing market.

In addition, when buying a house, the buyer should consider the grade, location, age, and if the property has a waterfront.

RECOMMENDATIONS

Potential buyers should prioritize square footage when considering a house purchase. However, it's essential to weigh other factors such as waterfront access, views, and the age of the property, as demonstrated by the multiple linear regression model.

NEXT STEP

We need to explore alternative regression techniques to get improve the significance of the models and provide deeper insights from the data

Thank You

Connect with us.

🌐 https://github.com/Zircornn-Black/Group3