

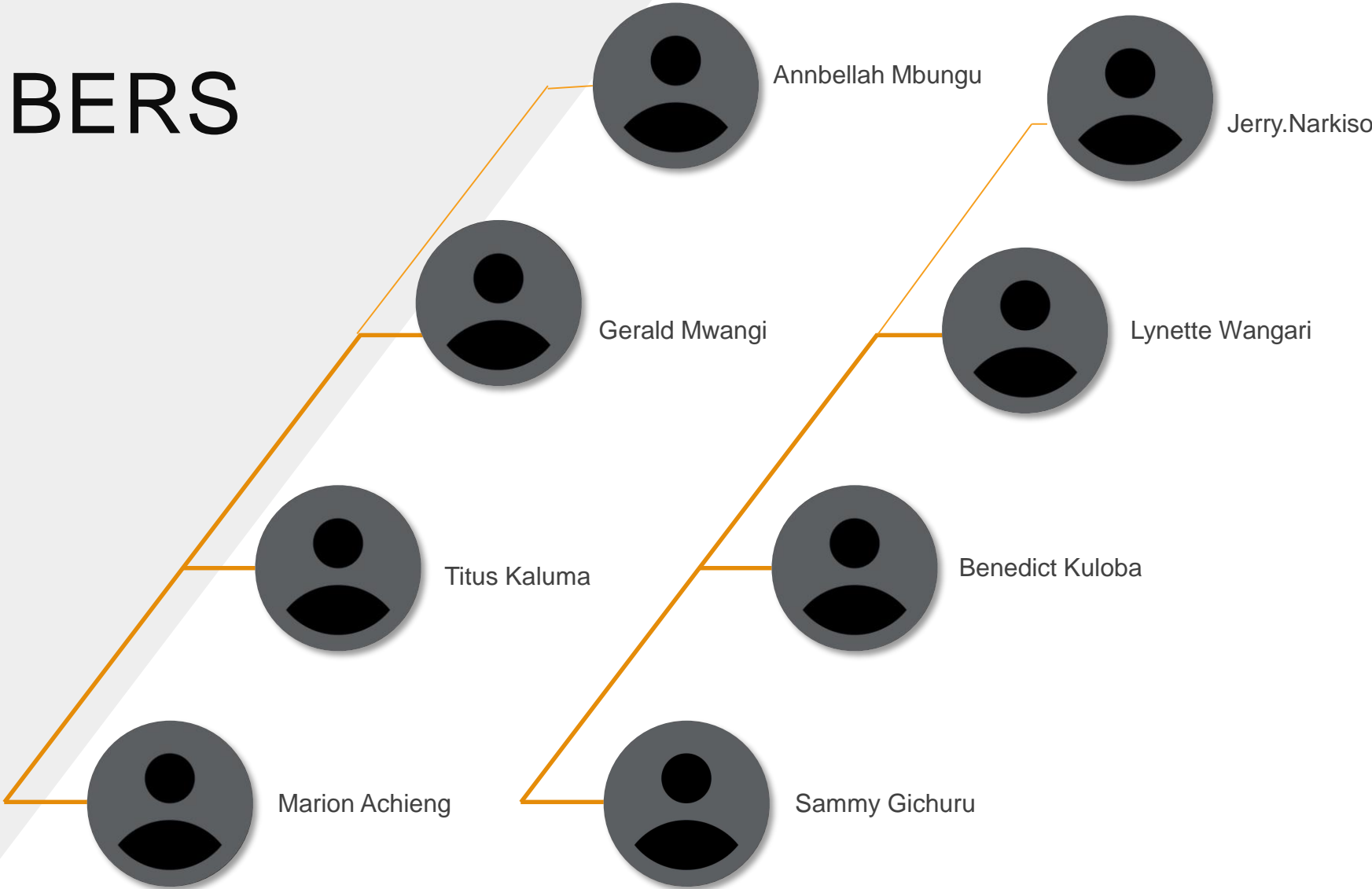
KINGS COUNTY HOUSING REGRESSION MODELLING ANALYSIS



GROUP MEMBERS



GEORGE KARIUKI
Group Leader





BUSINESS UNDERSTANDING

BACKGROUND INFORMATION

King County, Washington, is a diversified area encompassing urban, suburban, and rural areas. The housing market is fiercely competitive, fueled by substantial demand from homebuyers and investors. The county provides a variety of housing alternatives, including single-family houses, condominiums, townhouses, apartments, and luxurious estates. The county is home to exceptional schools and educational organizations, including the University of Washington.

The county's robust transportation system effects home demand and pricing. Hiking, skiing, boating, and fishing are all outdoor recreational activities. King County homes provide a desirable balance of natural beauty, economic prospects, and quality of life amenities, making them a popular choice for both homeowners and investors. However, successfully navigating the housing market involves careful study, planning, and consideration of individual requirements and preferences.

Overview



Kings county real estate agency tasked us a project to analyses home prices and market trends using their King County House Sales dataset.

We employed data analytics and predictive modelling whose key purpose was to develop insights that will assist homeowners in understanding how different predictor features affect the projected value of their property.

Challenges



King County's real estate investment market is thriving. Brokers face inquiries about home renovations, waterfront views e.tc and how they affect the price of a house and by how much.

Business Problem



Kings county real estate agency faces a lot of question from prospective clients about which features i.e. home renovation, number of bedrooms waterfront, view etc. will affect value to their property and by how much.



OBJECTIVES

- 1.To determine Current Market Trends and Best-Performing Features
- 2.To Investigate Popular Features and Preferences Across Time
- 3.To Investigate How Home Renovations affect Sale Price
- 4.To Inspect Seasonal Trends & Optimal Time to Sell



DATA UNDERSTANDING

DESCRIPTION OF DATA

The project extracts data from the King County House Sales dataset, which is a csv file. It contains 21 columns, 21,597 rows.

The dataset provides King County's residential sales values from May 2014 to May 2015.



OUR VARIABLES

Our target variable is 'house Price'.

Our predictor variables are ; sales date, renovation year, number of bedrooms and bathrooms, number of floors, square footage of the properties (including living and lot), and available view of the waterfront with addition of the grade and condition of each house as one of our predictor variables.



DATA TYPES

OBJECTS

Date
Sqft basement
Waterfront
View
Condition
Grade

OTHER VARIABLES

Integers or
Floats

COLUMNS DISTRIBUTION

ID: An integer representing some unique identifier.
Date: An object (likely a string) representing dates.
Price: A floating-point number representing the price of a property.
Bedrooms: An integer representing the number of bedrooms in a property.
Bathrooms: A floating-point number representing the number of bathrooms in a property.
Sqft_living: An integer representing the square footage of living space in a property.
Sqft_lot: An integer representing the square footage of the lot on which the property sits.
Floors: A floating-point number representing the number of floors in the property.
Waterfront: An object (likely a string) indicating whether the property has a waterfront view.
View: An object (likely a string) representing the level of view from the property.
Condition: An object (likely a string) representing the condition of the property.
grade: An object (likely a string) representing the grade given to the property.
Sqft_above: An integer representing the square footage of the property apart from the basement.
Sqft_basement: An object (likely a string) representing the square footage of the basement.
Yr_built: An integer representing the year the property was built.
Yr_renovated: A floating-point number representing the year the property was renovated.
zip code: An integer representing the zip code of the property location.
Lat: A floating-point number representing the latitude of the property location.
Long: A floating-point number representing the longitude of the property location.
Sqft_living15: An integer representing the square footage of interior housing living space for the nearest 15 neighbors.
Sqft_lot15: An integer representing the square footage of the land lots of the nearest 15 neighbors.

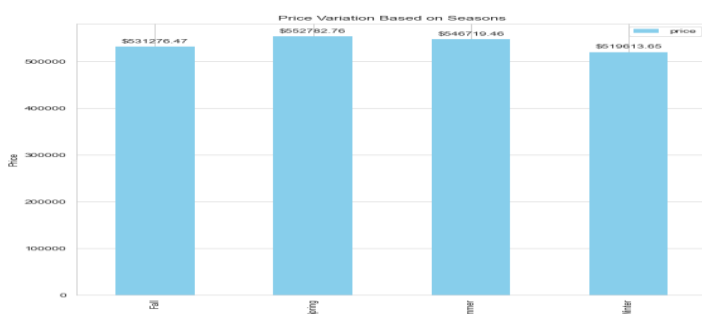
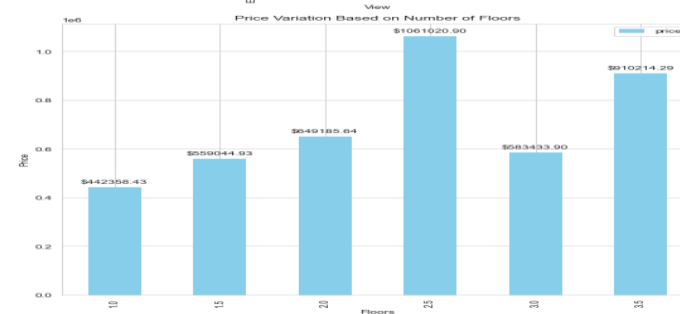
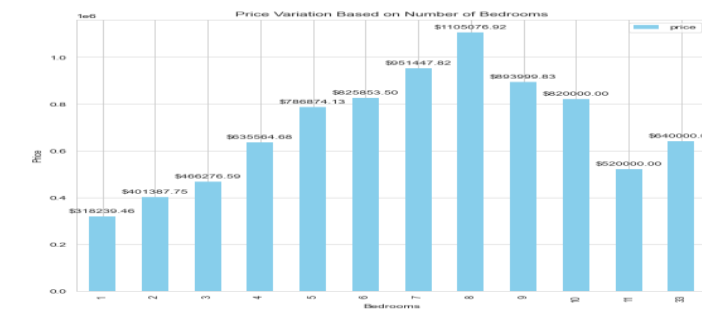
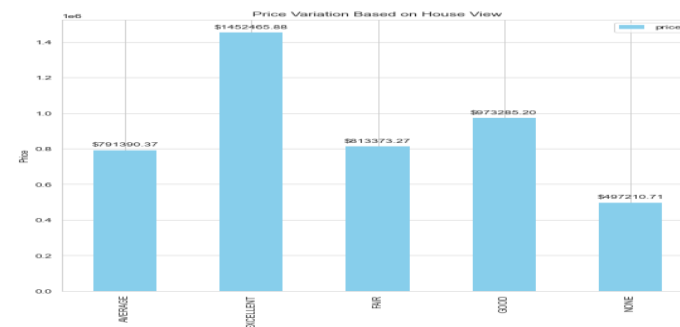
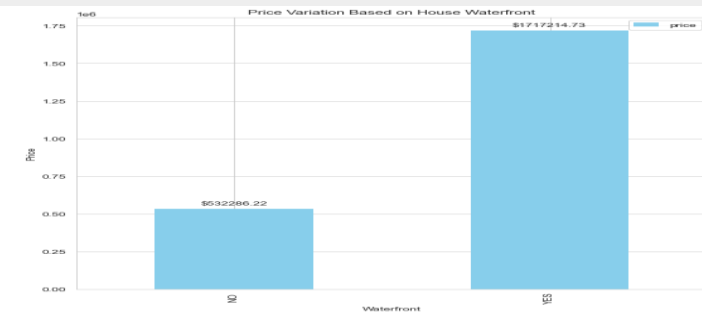
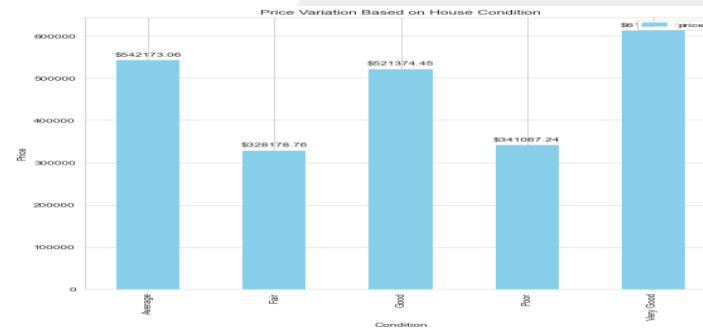
Visualizing the data

1. TO INVESTIGATE HOW HOME RENOVATIONS AFFECT SALE PRICE:

a. How does adding floors affect the sale price of a home? 2.5 and 3.5 floors increased the house prices up to 1 million dollars

b. How does the addition of bedrooms affect the sale price of a home? No since increase in number of bedrooms does not affect the price of a house

c. What is the predicted rise in the selling price of a home after an addition to its living space? 2000 to 3000 square feet of living space increased house prices up to 282 dollars per square feet



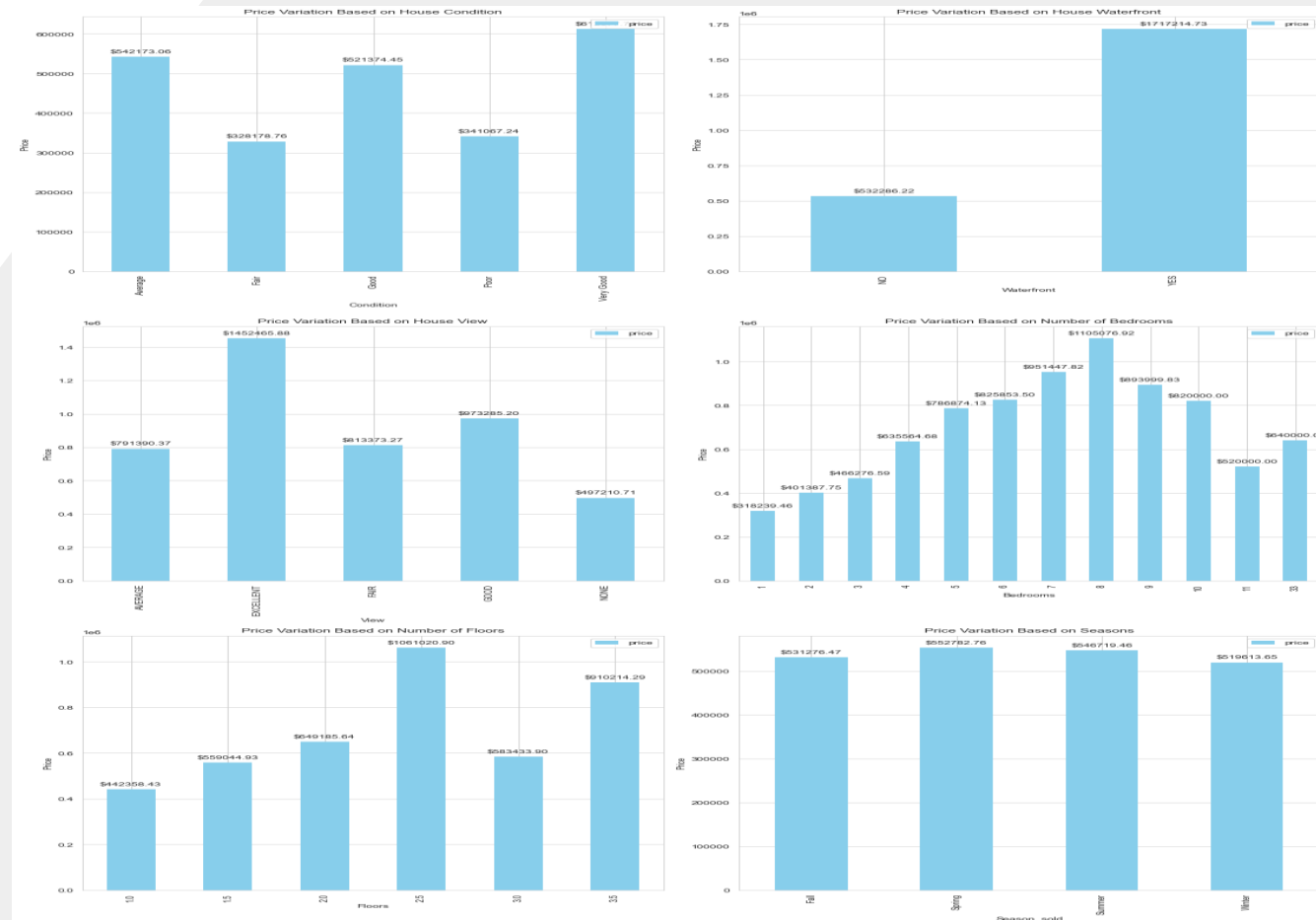
2. TO INSPECT SEASONAL TRENDS & OPTIMAL TIME TO SELL:

a. Do home sales and prices follow seasonal patterns?

Yes, spring and summer had the highest sales

b. When is the greatest time to sell a home for the highest possible price?

summer with a average price of \$600k



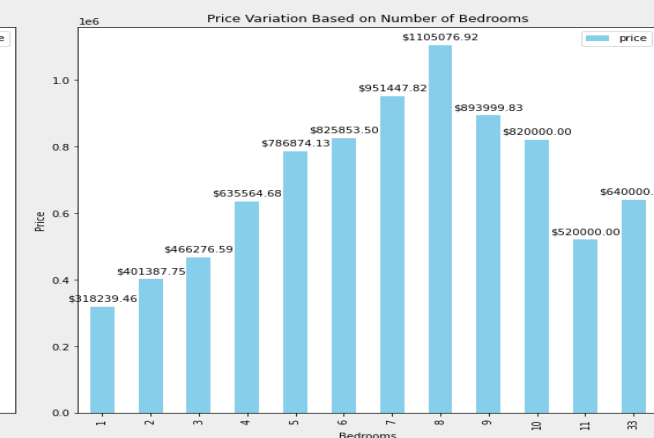
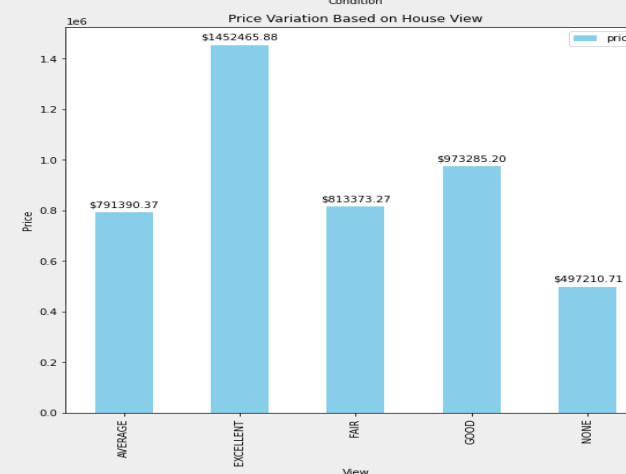
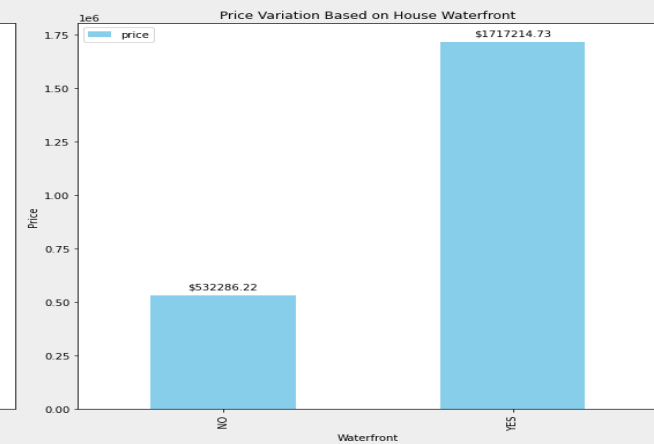
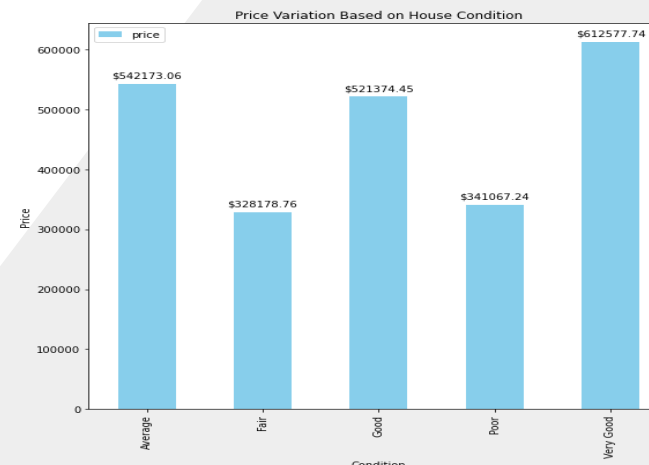
3.TO DETERMINE CURRENT MARKET TRENDS AND BEST-PERFORMING FEATURES

a.What factors distinguish top-performing properties in terms of sale price?

Very good condition, presence of a waterfront and excellent view

b. Which variables (e.g. bedrooms and floors) of the residences raise sale prices?

House condition, waterfront ,view and number bedrooms



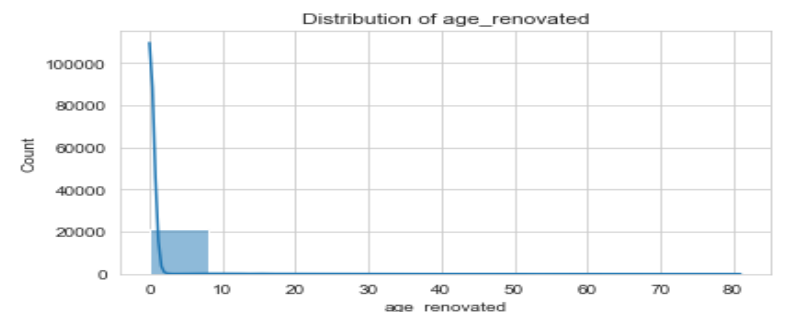
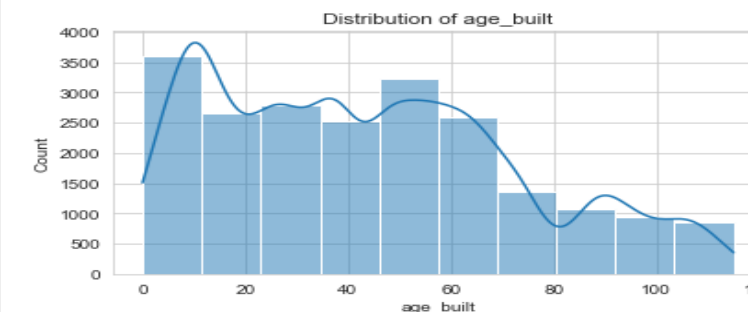
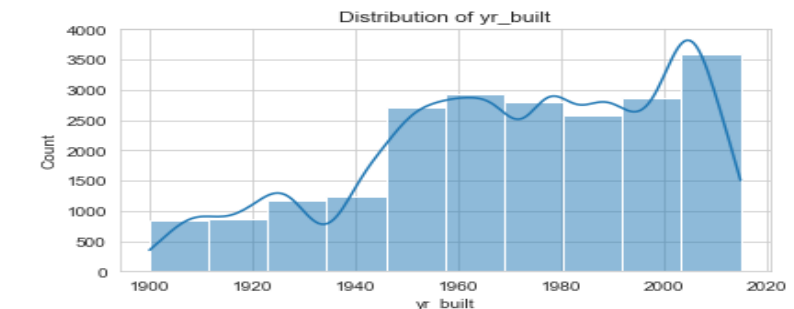
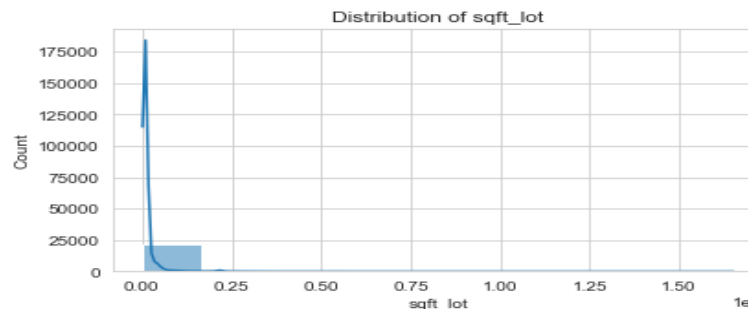
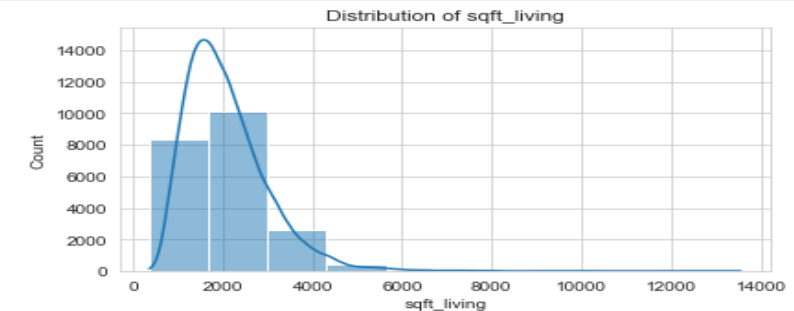
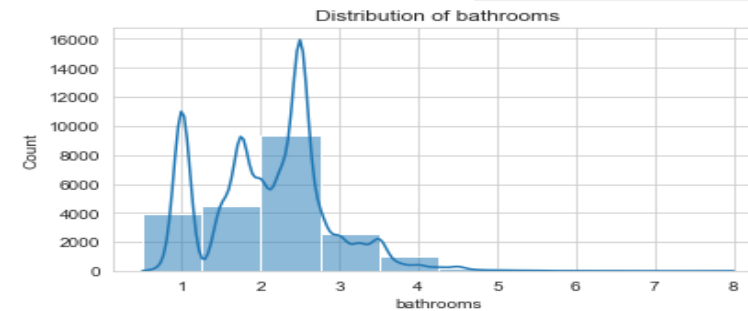
4. TO INVESTIGATE POPULAR FEATURES AND PREFERENCES ACROSS TIME

a. Which housing features and attributes have proven most popular and successful throughout time?

2 to 3 number of bathrooms and 2000 to 3000 square feet of living space,

b. What types of home features appeal to potential buyers?

Newer houses less than 60 years old since it was built



MODELLING SIMPLE LINEAR REGRESSIONS

The Scatter Plot Displays the relationship between square footage of living space and house prices. The model has a lower MSE and a higher R-squared value (49.3%), indicating that Sqft_living is a stronger predictor of house prices compared to bathrooms or bedrooms

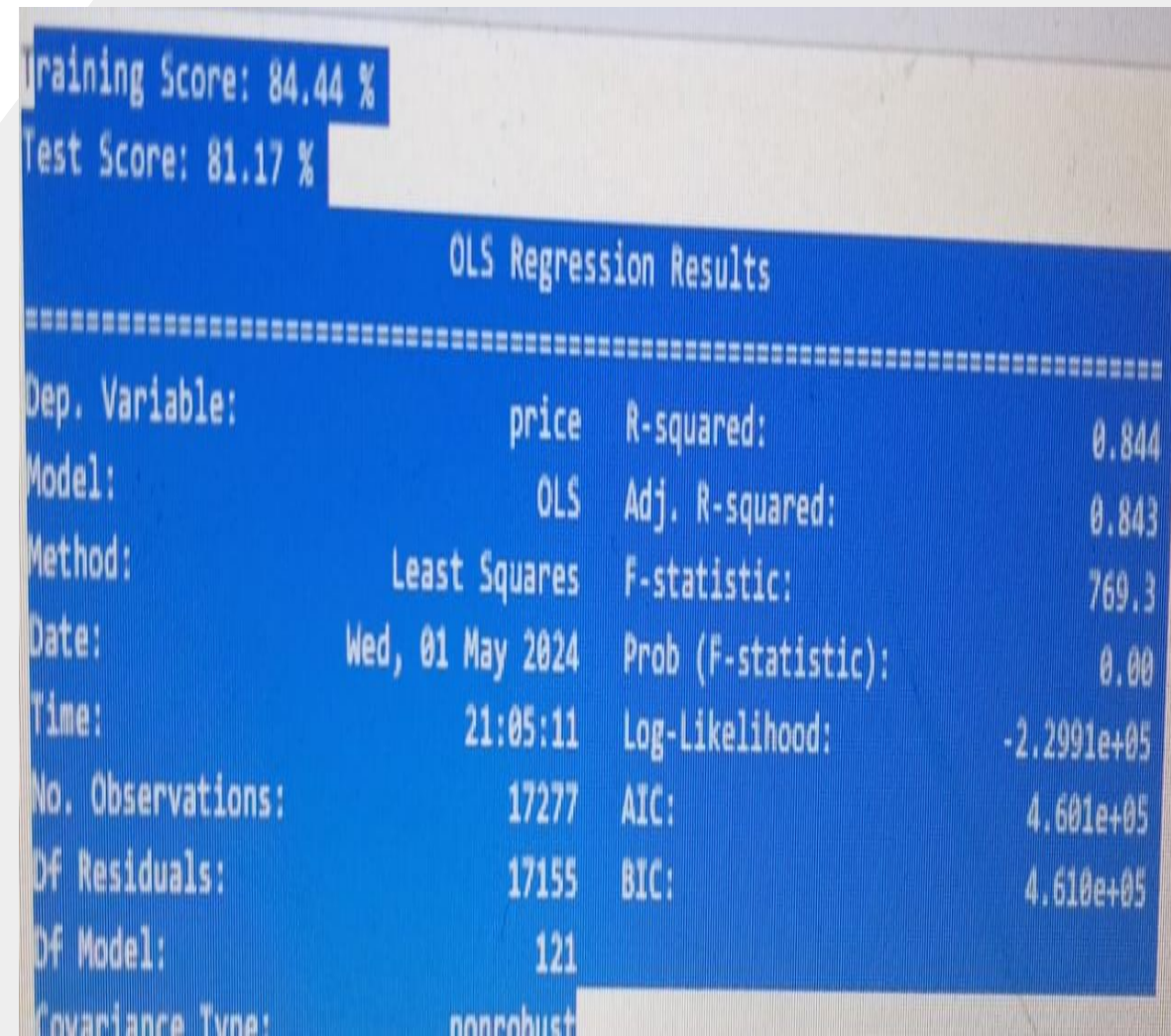


MULTILINEAR REGRESSION MODEL

The R-squared value (0.844) indicates the proportion of the variance in the dependent variable that is predictable from the independent variables. In this case, about 84.4% of the variance in the price can be explained by the independent variables.

The F-statistic tests the overall significance of the model. A low p-value (below 0.05) suggests that the model is statistically significant, meaning at least one independent variable has a non-zero coefficient. The associated p-value (0.00) indicates that the model is statistically significant.

The p-value associated with each coefficient indicates the statistical significance of that feature. A low p-value (typically < 0.05) suggests that the feature is statistically significant in predicting the target variable.



The image shows a screenshot of a software interface displaying OLS Regression Results. At the top, two summary statistics are shown: Training Score: 84.44 % and Test Score: 81.17 %. Below these, the title 'OLS Regression Results' is centered. The main body of the results is presented in a table-like format with two columns. The left column lists various model statistics, and the right column shows their corresponding values. The statistics include the dependent variable (price), the model type (OLS), the method (Least Squares), the date and time of the analysis, the number of observations (17277), the degrees of freedom for residuals (17155) and the model (121), and the covariance type (nonrobust). The right column provides the R-squared value (0.844), the adjusted R-squared value (0.843), the F-statistic (769.3), the probability of the F-statistic (0.00), the log-likelihood (-2.2991e+05), the Akaike Information Criterion (AIC: 4.601e+05), and the Bayesian Information Criterion (BIC: 4.610e+05).

OLS Regression Results	
Dep. Variable:	price
Model:	OLS
Method:	Least Squares
Date:	Wed, 01 May 2024
Time:	21:05:11
No. Observations:	17277
Df Residuals:	17155
Df Model:	121
Covariance Type:	nonrobust
R-squared:	0.844
Adj. R-squared:	0.843
F-statistic:	769.3
Prob (F-statistic):	0.00
Log-Likelihood:	-2.2991e+05
AIC:	4.601e+05
BIC:	4.610e+05

REGRESSION RESULTS

Coefficient Analysis of House Price

- Positive coefficients indicate price increases with increase in predictor variable.
- Negative coefficients suggest price decreases with increase in predictor variable.
- Example: sqft_living = \$89.29 for every additional square foot of living space.
- "Bathrooms" = \$23620 for each additional bathroom.
- "Waterfront_YES" = \$588700 for houses with waterfront view.
- Positive coefficients: houses with more bathrooms are more expensive.
- Positive coefficients: sqft_living and living area are positively related.
- Negative coefficients: older houses are cheaper.

CONCLUSIONS

King County Real Estate Market Analysis

- Pricing data was found to be significantly biased, with outliers distorting the overall picture.
- Average and median prices indicate that housing is generally within the reach of middle-class families.
- A noticeable difference exists between the average cost and the highest price for homes.
- Correlations were found between pricing and other numerical independent variables, such as number of bedrooms, baths, living area square footage, waterfront access, and view quality.
- Homes with better condition ratings, waterfront access, outstanding vistas, and more bedrooms tend to fetch higher prices.
- Multicollinearity showed the importance of considering several variables when evaluating housing costs.
- The study provides valuable insights for real estate agency and prospective purchasers.

RECOMMENDATIONS

1. **Constant Monitoring:** Both the age of the property and any renovations should be carefully evaluated. Buyers should inquire about the age of key components such as the roof, system, and appliances. Sellers should consider investing in renovations to increase the value of their property.
2. **Focus on Key Features:** Buyers and sellers should pay attention to key features such as waterfront views, property condition, and grade when assessing house prices. These features have significant impacts on the value of the property.
3. **Consider number of floors:** By analyzing the relationship between the number of floors and sale prices comprehensively, we can see an upwards trend from the 2.5 mark.
4. **Seasonal Trends and Optimal** Consider purchasing a house during winter. With the lowest average house price compared to other seasons, winter presents an opportunity to potentially find a more affordable home. While spring and summer may be popular seasons for buying, exploring the market during winter could result in cost savings. Additionally, with fewer buyers in the market during winter, there may be less competition for available properties. However, ensure to thoroughly assess any seasonal challenges, such as weather conditions, and work closely with a real estate professional to navigate the market effectively.

NEXT STEPS

Further Exploration of Outliers: Despite the fact that we tackled outliers via normalizing the data, we hypothesized that a deeper assessment of these outliers may reveal any specific market dynamics or extraordinary features of the property that ought to be taken into account.

Feature Engineering: To better deal with the subtleties of the housing market, find out what it would take to develop new features or integrate current features. Square footage may be a more reliable and consistent hint we can use along with other factors like lot size or facility accessibility when determining home prices.

Market Segmentation: In order to make marketing campaigns and product offerings match with a specific customer group, these groups are created using demographic features such as family size or income level. That might lead to the better used resources and better serving buyers of different segments

Predictive Modelling: Utilize machine learning techniques and build prediction models to estimate house values based on various indicators. That then allows real estate companies to price properties based on the data and supply clients with very accurate price forecasts.



YOU

THANK

