

KINGS COUNTY HOUSING REGRESSION MODELLING ANALYSIS

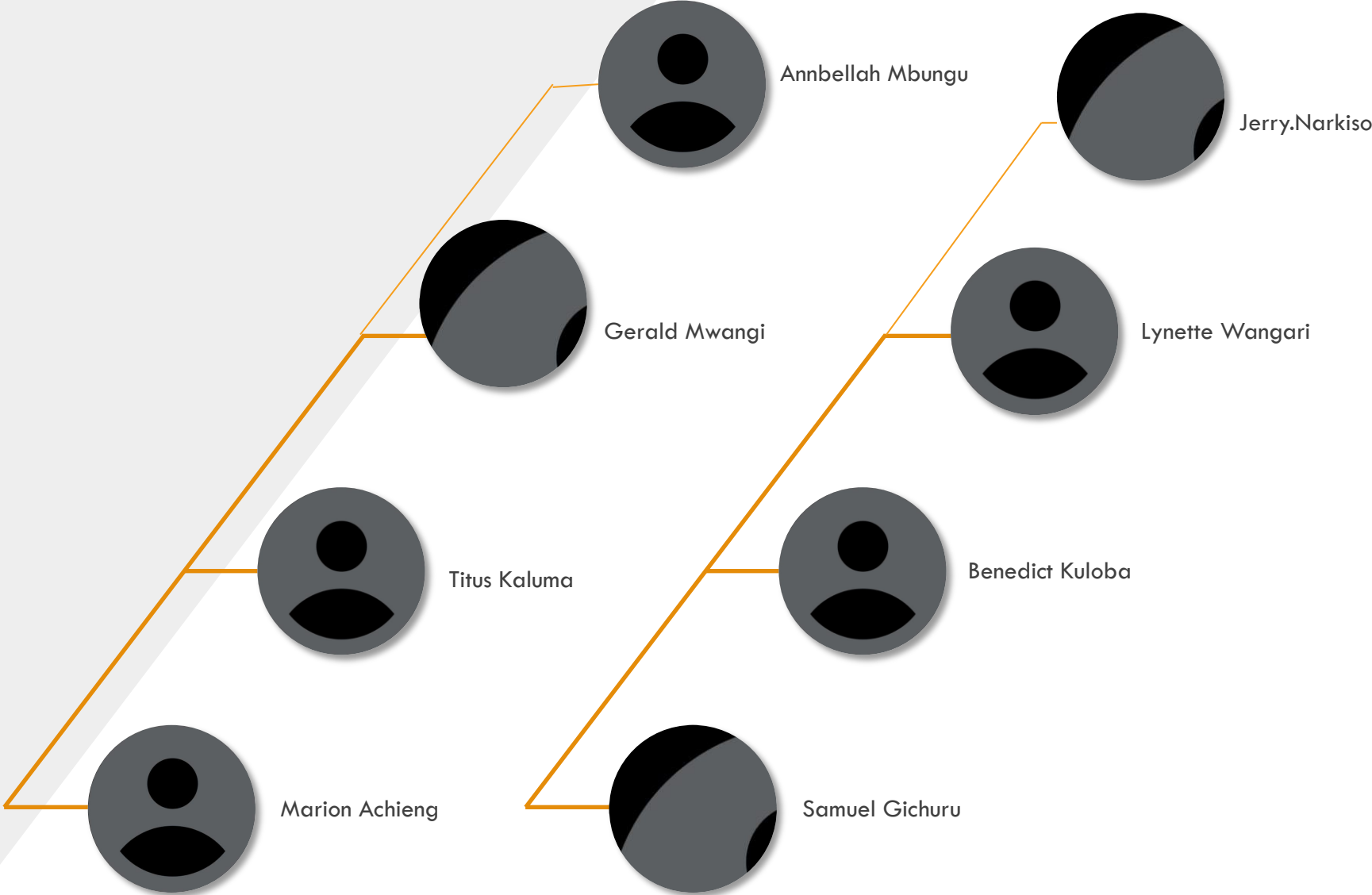


GROUP MEMBERS



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

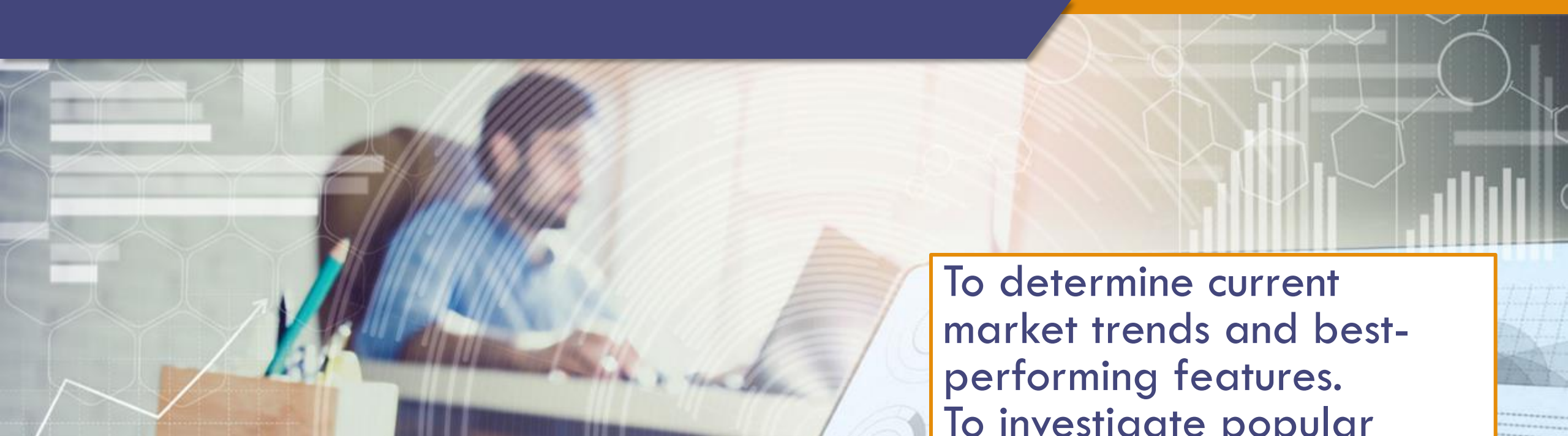


King County's real estate investment market is thriving. Brokers face inquiries about home renovations, waterfront views e.tc and how they affect the price of a house and by how much.

Business Problem

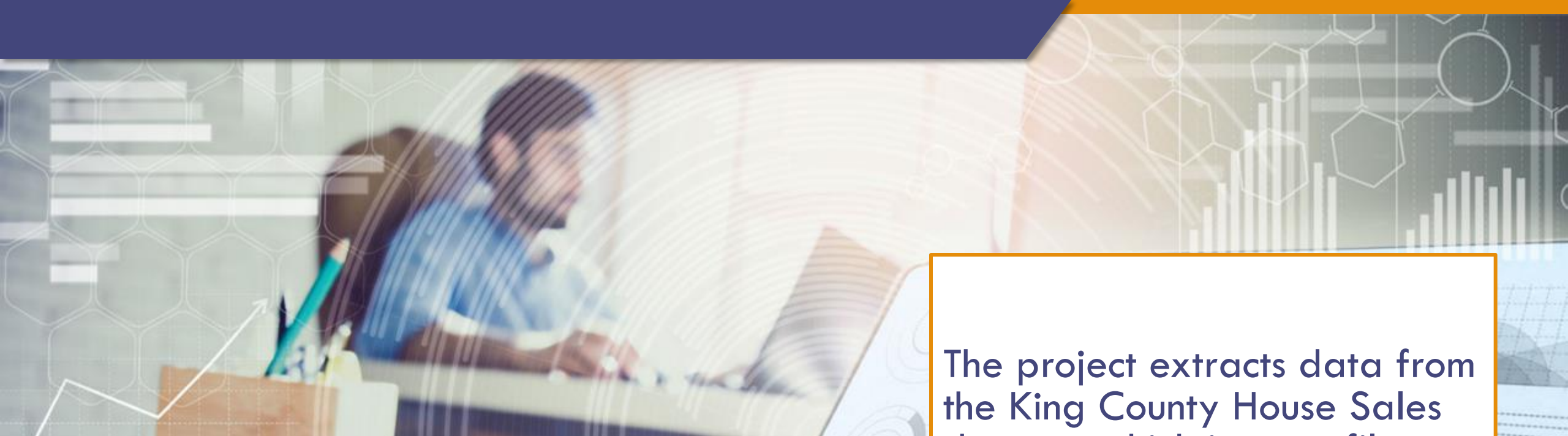


Kings county real estate agency faces a lot of question from prospective clients about which features i.e. home renovation, number of bedrooms waterfront, view etc. will affect value to their property and by how much.



OBJECTIVES

To determine current market trends and best-performing features.
To investigate popular features and preferences across time.
To Investigate how home renovations affect sale price.
To inspect seasonal trends and optimal time to sell.



DATA UNDERSTANDING

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.



COLUMNS DISTRIBUTION

- **Square footage:** Larger properties generally have higher prices.
- **Number of bedrooms and bathrooms:** More bedrooms and bathrooms often lead to higher prices.
- **Location:** Properties in desirable neighborhoods, close to amenities, schools, or with scenic views tend to have higher prices.
- **Condition and grade:** Higher condition and grade ratings correlate with higher prices.
- **Year built and renovated:** Newer properties or recently renovated ones command higher prices.
- **View and waterfront:** Properties with views or waterfront locations can significantly influence prices.
- **Floors:** Properties with multiple floors or unique architectural features may have higher prices.
- **Neighborhood features:** Size of living spaces and lots of nearby properties can reflect the neighborhood's desirability and impact prices.

FINDINGS

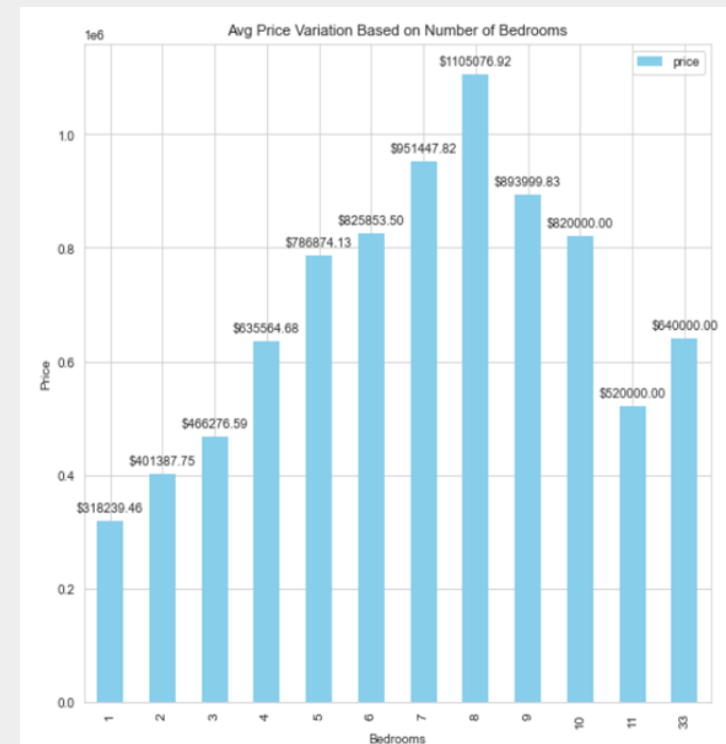
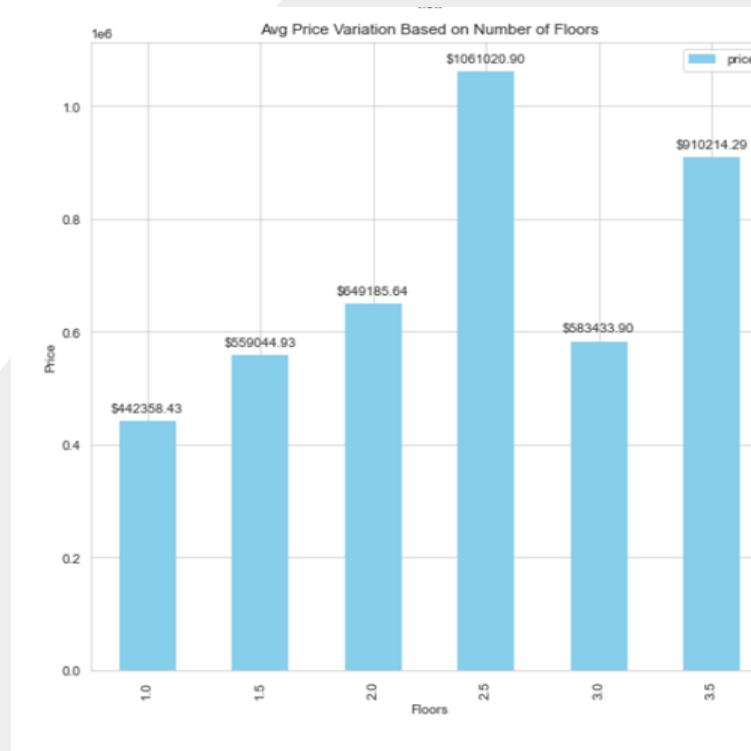
1.TO INVESTIGATE HOW HOME RENOVATIONS AFFECT SALE PRICE:

a.How does adding floors affect the sale price of a home?

Houses with 2 to 3 floors value increased fetching prices of upto 1 million dollars.

b.How does the addition of bedrooms affect the sale price of a home?

Gradual increase in the number of bedrooms show a significant increase in the price of the house these could be attributed to the number of members in a given family



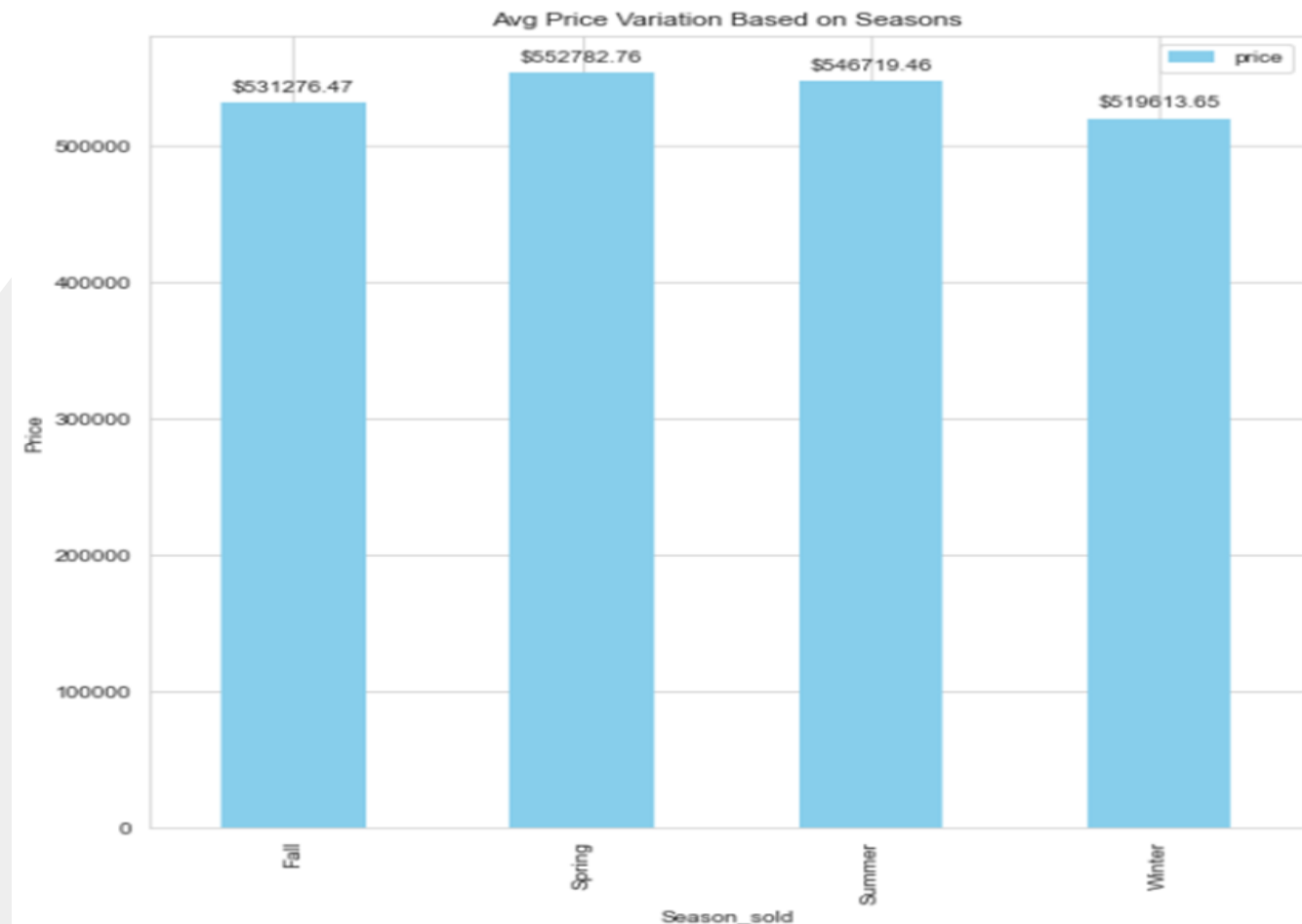
2. TO INSPECT SEASONAL TRENDS AND OPTIMAL TIME TO SELL:

a. Do home sales and prices follow seasonal patterns?

Yes, during the spring and summer seasons houses fetched higher prices

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

Summer with a average price of \$600k ,these is because these is the time most new home buyers are looking for house to buy



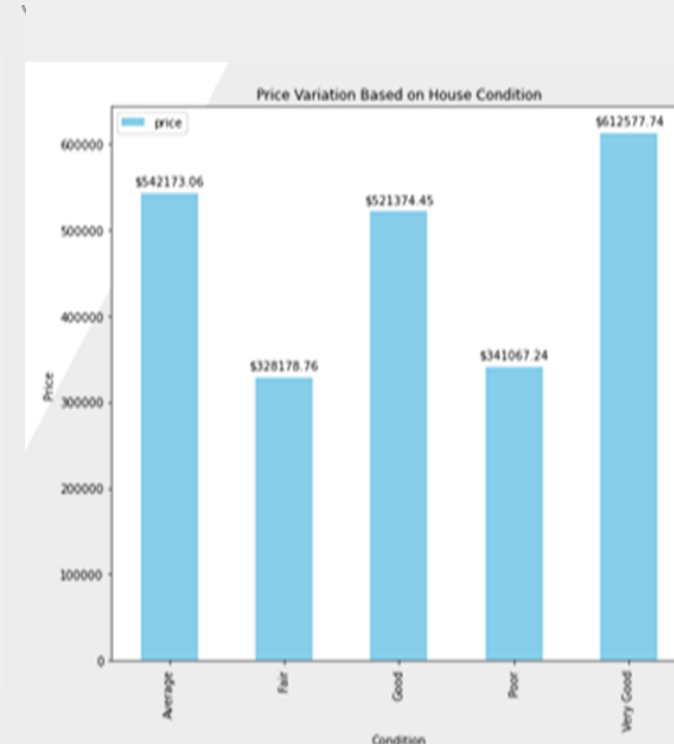
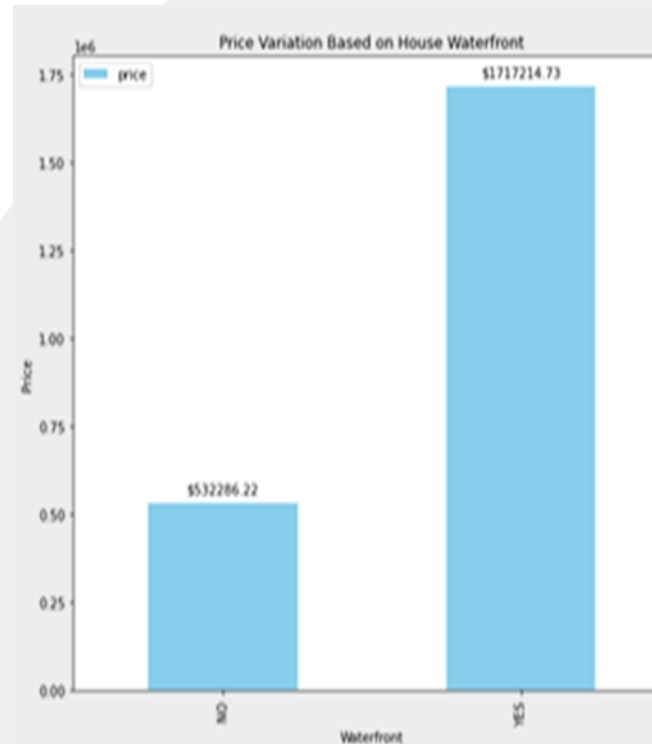
3.TO DETERMINE CURRENT MARKET TRENDS AND BEST-PERFORMING FEATURES

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

Houses in very good condition or have a waterfront fetched high prices ranging from 1.2 to 1.7 million

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

House condition, waterfront ,view and number bedrooms.



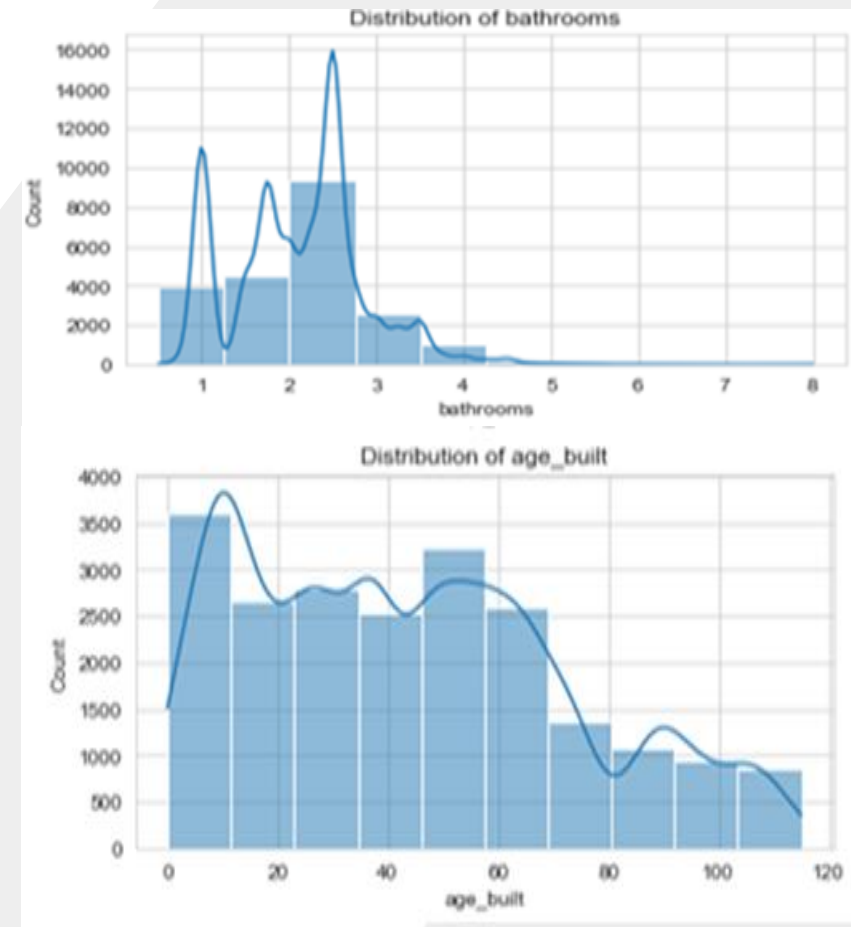
4. TO INVESTIGATE POPULAR FEATURES AND PREFERENCES ACROSS TIME

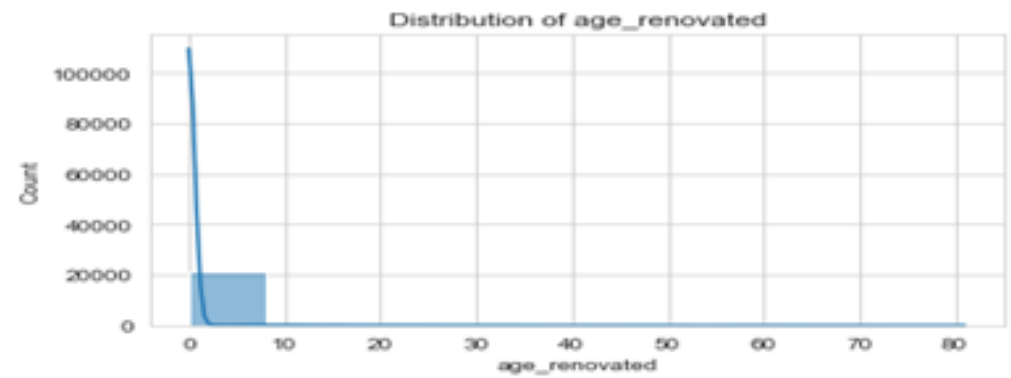
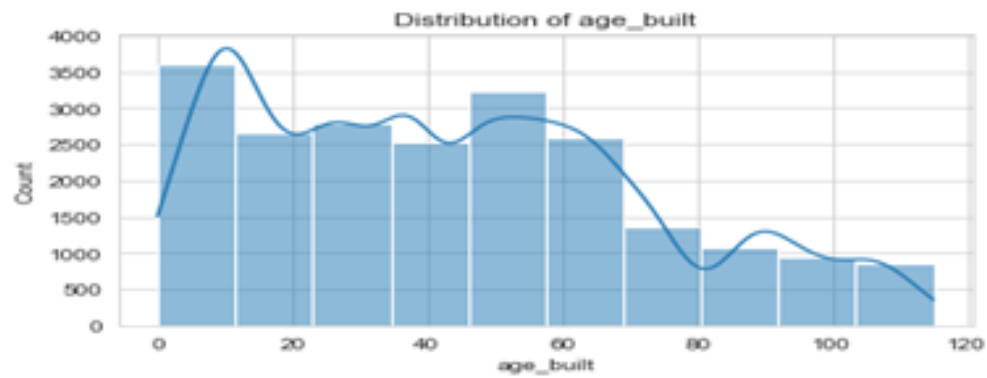
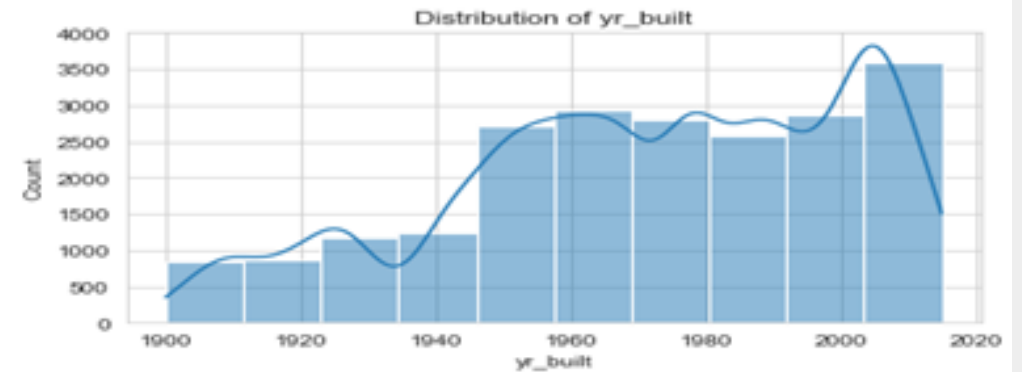
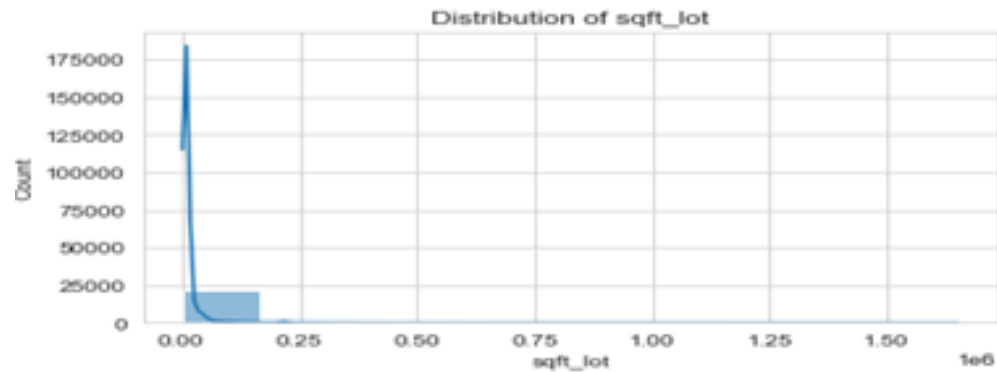
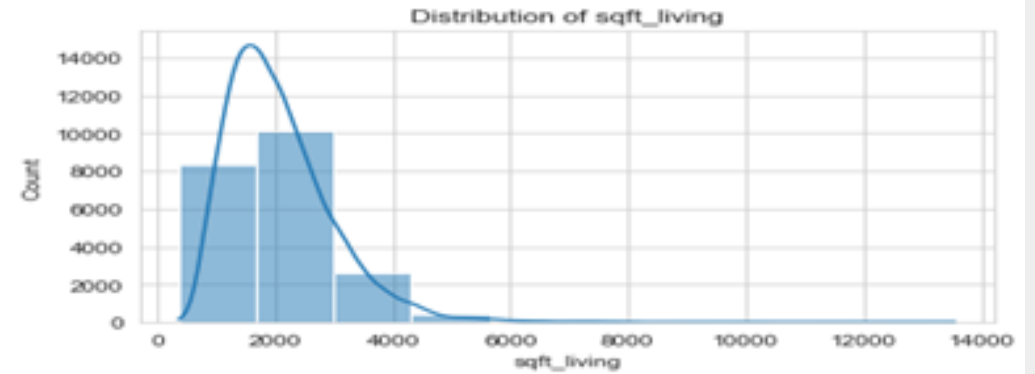
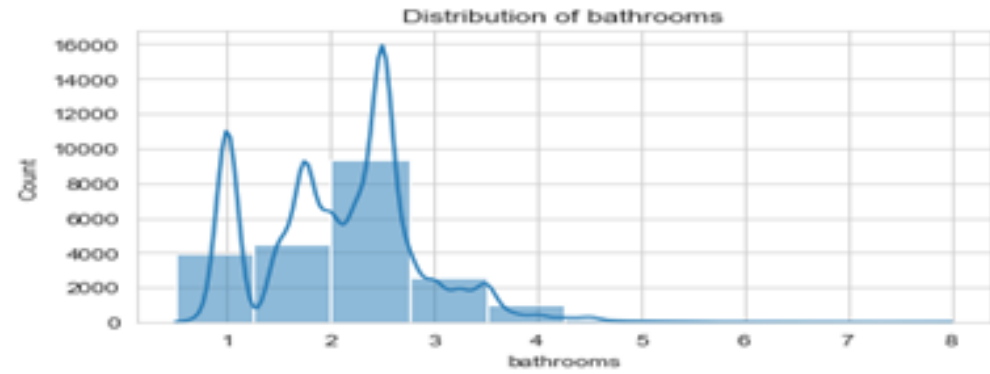
a. Which housing features and attributes have proven most popular and successful throughout time?

Houses with 2 to 3 bathrooms and houses with 2000 to 3000 square feet of living space. Were more popular among home users

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

Newer houses less than 60 years old since it was built were best preferred by buyers.





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 (0.706) indicates the proportion of the variance in the dependent variable that is predictable from the independent variables. In this case, about 70.6% 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.

OLS Regression Results						
=====						
Dep. Variable:	price	R-squared:	0.706			
Model:	OLS	Adj. R-squared:	0.705			
Method:	Least Squares	F-statistic:	1107.			
Date:	Thu, 02 May 2024	Prob (F-statistic):	0.00			
Time:	17:48:15	Log-Likelihood:	-2.2058e+05			
No. Observations:	16197	AIC:	4.412e+05			
Df Residuals:	16161	BIC:	4.415e+05			
Df Model:	35					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	-1.161e+08	1.44e+07	-8.082	0.000	-1.44e+08	-8.79e+07
bedrooms	1.783e+04	2105.527	8.469	0.000	1.37e+04	2.2e+04
bathrooms	6.521e+04	3245.403	20.093	0.000	5.88e+04	7.16e+04
sqft_lot	0.2873	0.055	5.241	0.000	0.180	0.395
floors	1413.2324	3649.039	0.387	0.699	-5739.287	8565.752
yr_renovated	106.5961	6.826	15.617	0.000	93.217	119.976
zipcode	-385.7128	37.579	-10.264	0.000	-459.371	-312.055
lat	6.73e+05	1.2e+04	55.977	0.000	6.49e+05	6.97e+05
long	-2.881e+05	1.43e+04	-20.081	0.000	-3.16e+05	-2.6e+05
sqft_living15	84.4411	3.689	22.890	0.000	77.210	91.672
sqft_lot15	-0.3009	0.084	-3.571	0.000	-0.466	-0.136
month_sold	3639.1426	987.701	3.684	0.000	1703.139	5575.147

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` = \$97.81 for every additional square foot of living space.
- "Bathrooms" = \$43650 for each additional bathroom.
- "Waterfront YES" = \$547600 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.

CONCLUSION

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

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.

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.

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.

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.

A hand is shown from the bottom right, holding a glowing blue sphere. The sphere is filled with various business and technology icons, including bar charts, gears, a globe, a line graph, a head with gears, and a hierarchical chart. The background is dark blue with a subtle grid pattern. There are orange decorative elements: a vertical bar on the left and a diagonal bar on the bottom right.

THANK YOU