HOUSE PRICE PREDICTION

PROJECT.

PRESENTED BY: SAMMY WARAH
GROUP 11

CONTENT

PROBLEM STATEMENT

The goal of this project is to provide valuable insights to a real estate agency, which aims to offer data-driven recommendations to homeowners regarding home renovations and their potential impact on the estimated value of their properties.

OBJECTIVES.

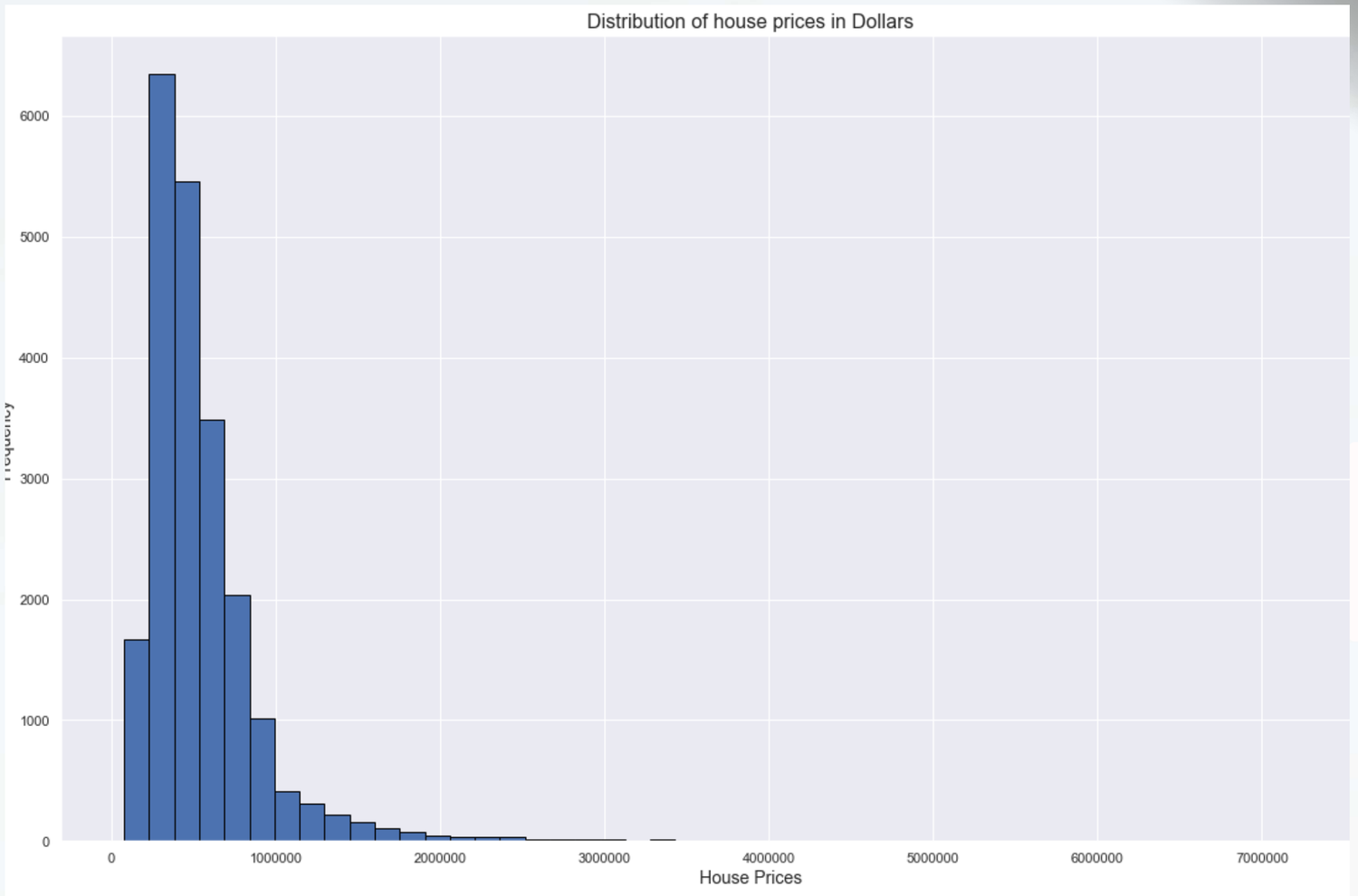
01 Utilize multiple linear regression modeling to analyze the King County House Sales dataset and identify key factors influencing home prices.

02 Offer personalized recommendations to homeowners regarding the types of renovations or improvements that could potentially increase the estimated value of their homes.

OUR DATA.

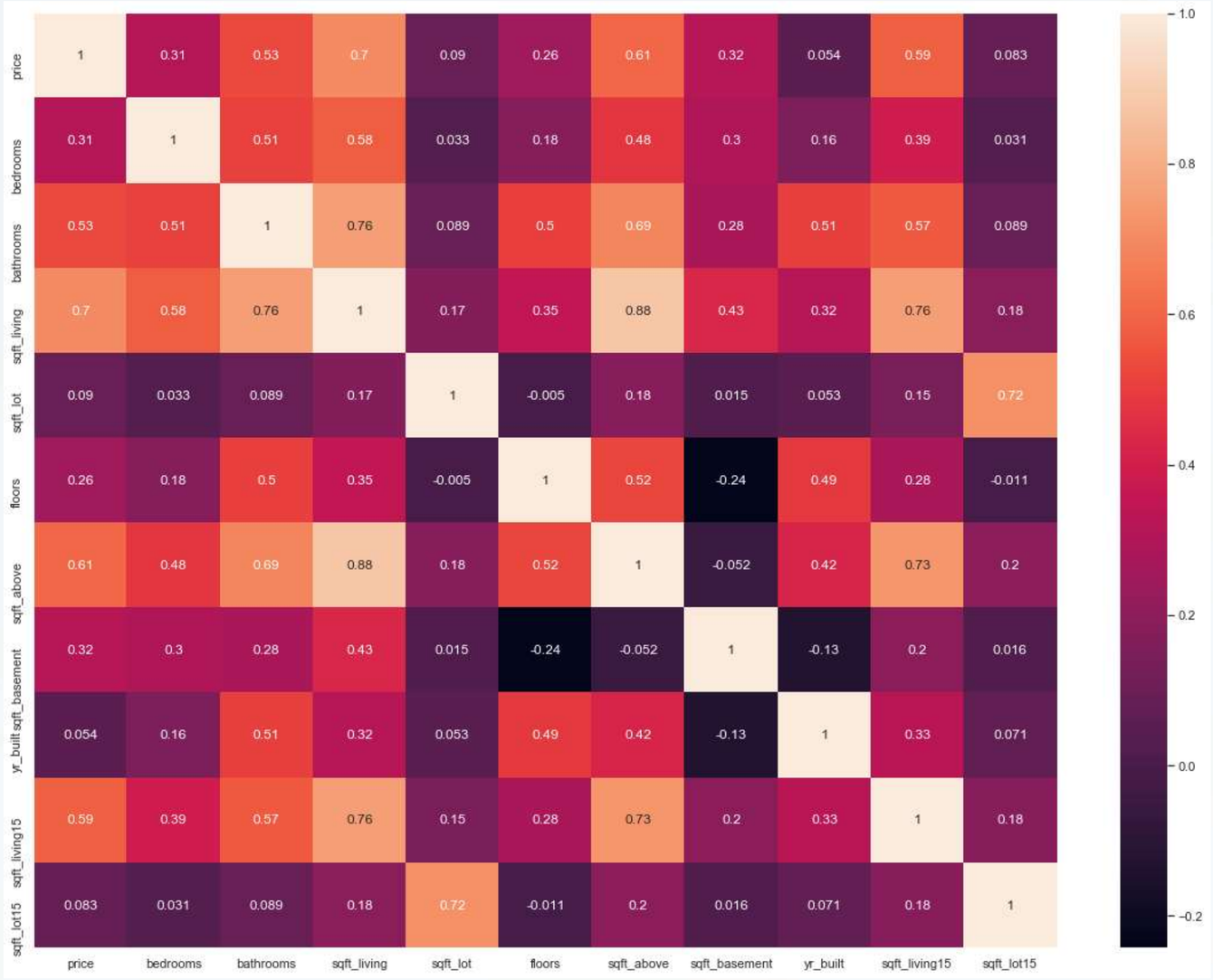
This project uses the King County House Sales dataset

DATA UNDERSTANDING



Based on our data and the Histogram above, House Prices are positively skewed, meaning that: Mode < Median < Mean

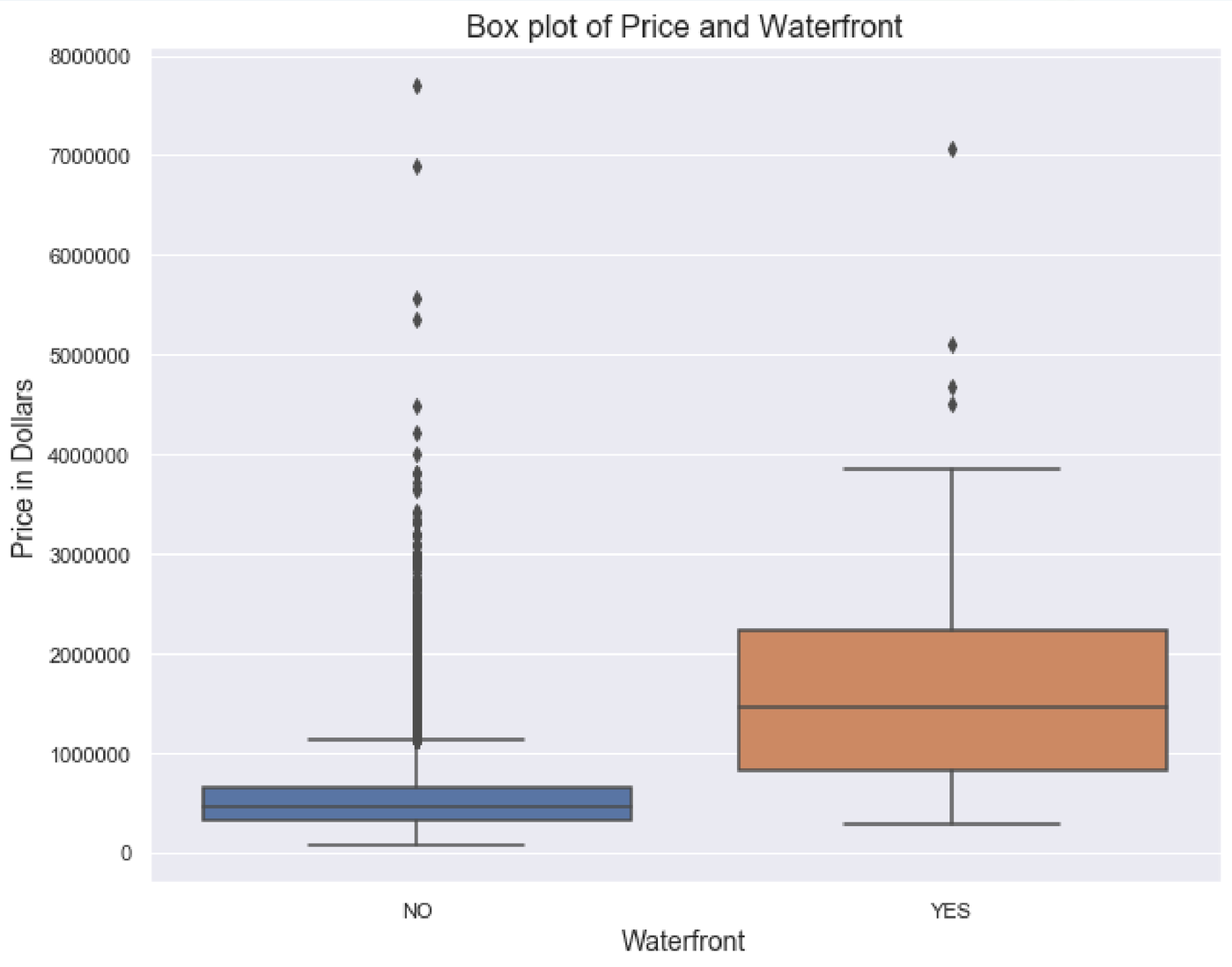
CORRELATION MATRIX.



Based on the correlation matrix, it can be observed that the Square footage of living space other than the basement has the highest positive correlation to the price of 0.70.Which is a strong correlation.

Additionally, the yr_build variable had the least correlation of 0.054

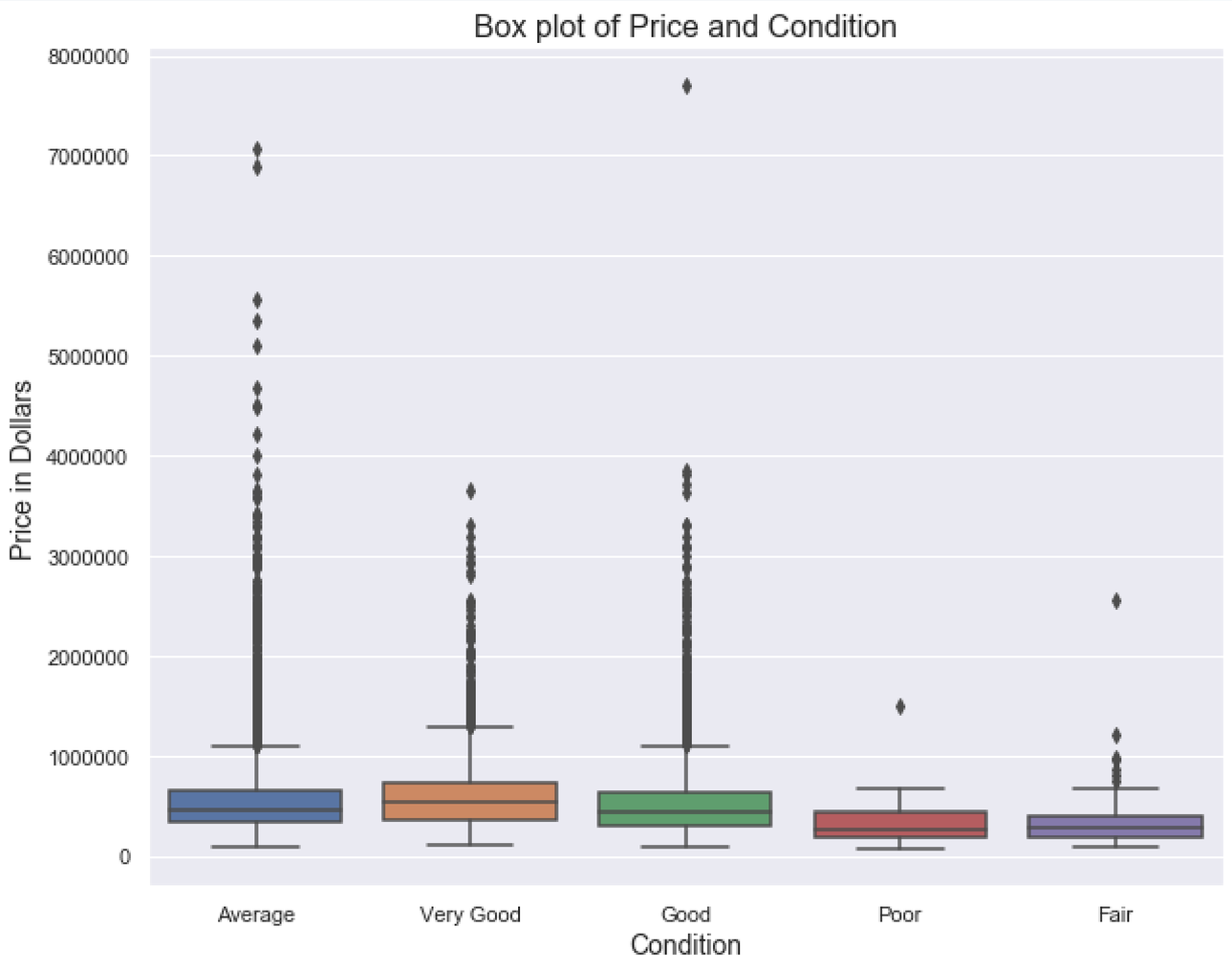
PRICE VS WATERFRONT



From the Box plot above, we can observe that the price of a house tends to be higher when the house has a waterfront.

Additionally, houses that do not have a waterfront tend to have more outliers than those with waterfronts

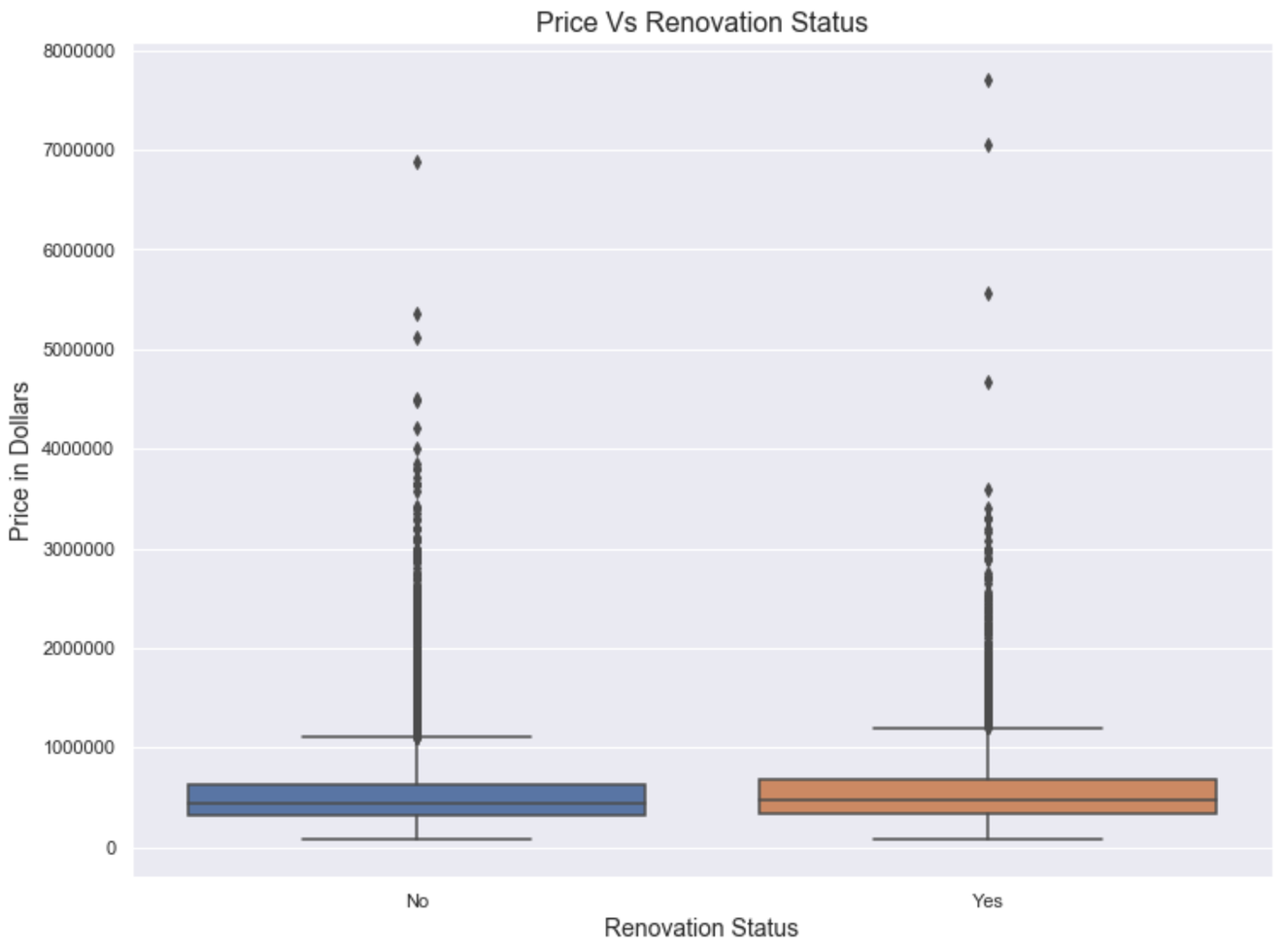
PRICE VS CONDITION.



Based on the correlation matrix, it can be observed that the Square footage of living space other than the basement has the highest positive correlation to the price of 0.70.Which is a strong correlation.

Additionally, the yr_build variable had the least correlation of 0.054

PRICE VS RENOVATION STATUS



Based on the correlation matrix, it can be observed that the Square footage of living space other than the basement has the highest positive correlation to the price of 0.70.Which is a strong correlation.

Additionally, the yr_build variable had the least correlation of 0.054

REGRESSION MODELING

Simple Linear Regression

We model our price against the independent variable with the higest correlation

Multiple Linear Regression

Through an itterative proces, we include as many predictive variables to the model.

SIMPLE LINEAR MODEL

OLS Regression Results

Dep. Variable:	price	R-squared:	0.492
Model:	OLS	Adj. R-squared:	0.492
Method:	Least Squares	F-statistic:	2.087e+04
Date:	Tue, 09 Apr 2024	Prob (F-statistic):	0.00
Time:	19:29:40	Log-Likelihood:	-2.9912e+05
No. Observations:	21534	AIC:	5.982e+05
Df Residuals:	21532	BIC:	5.983e+05
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-4.215e+04	4404.521	-9.570	0.000	-5.08e+04	-3.35e+04
sqft_living	279.9321	1.938	144.473	0.000	276.134	283.730

Omnibus:	14582.265	Durbin-Watson:	1.981
Prob(Omnibus):	0.000	Jarque-Bera (JB):	516142.289
Skew:	2.781	Prob(JB):	0.00
Kurtosis:	26.331	Cond. No.	5.63e+03

Comments.

- Our Model and coefficient are statistically significant because the F-value is less than our assumed alpha of 0.05.
- Our Adjusted R Squared is 0.493, hence the model explains 49.2% of the variance in price, the target variable.
- For a square foot living of 0, our model would predict a price of -0.0004399 dollars. An increase of 1 square-feet living, would increase the price by 280.

MULTIPLE LINEAR MODEL

OLS Regression Results

Dep. Variable:	price	R-squared:	0.676
Model:	OLS	Adj. R-squared:	0.675
Method:	Least Squares	F-statistic:	1867.
Date:	Tue, 09 Apr 2024	Prob (F-statistic):	0.00
Time:	19:29:41	Log-Likelihood:	-2.9430e+05
No. Observations:	21534	AIC:	5.886e+05
Df Residuals:	21509	BIC:	5.888e+05
Df Model:	24		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	7.138e+06	1.31e+05	54.692	0.000	6.88e+06	7.39e+06
bedrooms	-2.803e+04	2003.856	-13.987	0.000	-3.2e+04	-2.41e+04
bathrooms	5.024e+04	3383.670	14.848	0.000	4.36e+04	5.69e+04
sqft_living	113.9033	3.916	29.089	0.000	106.228	121.578
floors	4.771e+04	3704.509	12.879	0.000	4.04e+04	5.5e+04
sqft_basement	48.5041	4.359	11.126	0.000	39.959	57.049
yr_built	-3394.7148	66.588	-50.981	0.000	-3525.233	-3264.196
sqft_living15	40.2369	3.481	11.559	0.000	33.414	47.060
sqft_lot15	-0.5212	0.054	-9.623	0.000	-0.627	-0.415
waterfront_YES	7.098e+05	1.76e+04	40.273	0.000	6.75e+05	7.44e+05
condition_Fair	-3.051e+04	1.63e+04	-1.874	0.061	-6.24e+04	1405.416
condition_Good	1.709e+04	3541.776	4.826	0.000	1.01e+04	2.4e+04
condition_Poor	-4.659e+04	3.91e+04	-1.191	0.234	-1.23e+05	3.01e+04
condition_Very Good	5.669e+04	5715.896	9.918	0.000	4.55e+04	6.79e+04
grade_11 Excellent	2.749e+05	1.24e+04	22.120	0.000	2.5e+05	2.99e+05
grade_12 Luxury	7.384e+05	2.37e+04	31.099	0.000	6.92e+05	7.85e+05
grade_13 Mansion	1.993e+06	5.93e+04	33.626	0.000	1.88e+06	2.11e+06
grade_3 Poor	-5.795e+05	2.09e+05	-2.773	0.006	-9.89e+05	-1.7e+05
grade_4 Low	-5.393e+05	4.15e+04	-12.998	0.000	-6.21e+05	-4.58e+05
grade_5 Fair	-5.463e+05	1.67e+04	-32.712	0.000	-5.79e+05	-5.14e+05
grade_6 Low Average	-4.917e+05	1.06e+04	-46.574	0.000	-5.12e+05	-4.71e+05
grade_7 Average	-4.148e+05	8740.474	-47.460	0.000	-4.32e+05	-3.98e+05
grade_8 Good	-3.275e+05	7890.605	-41.510	0.000	-3.43e+05	-3.12e+05
grade_9 Better	-1.824e+05	7682.676	-23.738	0.000	-1.97e+05	-1.67e+05
renovated_Yes	1.18e+04	3516.495	3.355	0.001	4905.392	1.87e+04

Omnibus:	12581.221	Durbin-Watson:	1.976
Prob(Omnibus):	0.000	Jarque-Bera (JB):	446759.374
Skew:	2.224	Prob(JB):	0.00
Kurtosis:	24.866	Cond. No.	4.43e+06

Comments.

1. Our Model and coefficient are statistically significant because the F-value is less than our assumed alpha of 0.05.

2. Our Adjusted R Squared is 0.676, hence the model explains 67.6% of the variance in price, target variable.

OBSERVATIONS

01 R-Squared Value. The R-squared value of 0.676 indicates that the model explains approximately 67.6% of the variance in the target variable (price). This suggests that the model has a moderate level of explanatory power.

02 Significance of Predictors. Several predictors have statistically significant coefficients. That is the calculated P-values are less than the assumed alpha of 0.05 indicating that they have a significant impact on the price of houses

03 Coefficient Interpretation: The coefficients represent the change in the target variable (price) for a one-unit change in the predictor variable, holding all other variables constant.

04 Intercept: The intercept term (const) represents the expected price of a house when all predictor variables are zero.

OBSERVATIONS

05 Positive/Negative Coefficients: Positive coefficients indicate a positive relationship between the predictor variable and the price of houses, while negative coefficients indicate a negative relationship.

06 Waterfront Variable: The coefficient for the waterfront variable is particularly large, suggesting that houses with waterfront views have a substantial impact on price.

07 Grade Variable: The grade variable includes multiple categories, each representing different levels of house quality. Higher-grade categories have positive coefficients, indicating that houses with higher quality grades tend to have higher prices.

08 Renovated Variable: The coefficient for the renovated variable (renovated_Yes) is positive and statistically significant ($p < 0.05$), suggesting that renovated houses tend to have higher prices compared to non-renovated houses.

RECOMMENDATIONS

01

Consider Waterfront Properties.

Given the significant impact of waterfront properties on price, buyers interested in premium properties may prioritize houses with waterfront views.

02

Focus on Grade

Buyers seeking high-quality houses should pay attention to the grade variable, as higher-grade categories tend to command higher prices

03

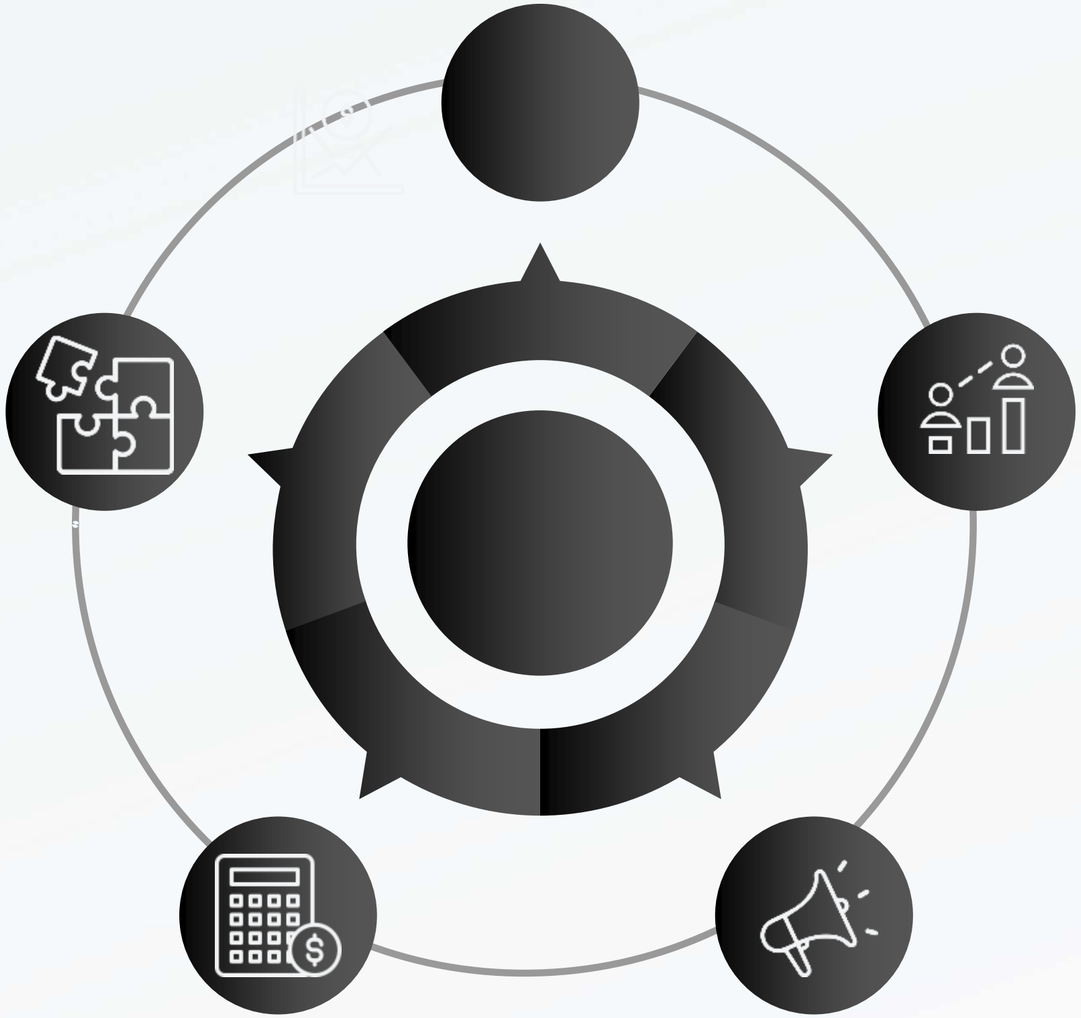
Evaluate Renovated Properties

Sellers may benefit from renovating their properties, as renovated houses tend to have higher prices. Buyers should also consider renovated properties when looking for houses with increased value potential.

04

Further Analysis

Conduct further analysis to explore potential interactions between predictor variables and identify any additional factors that may influence house prices.



OUR TEAM

Sammy Warah
sammy.warah@student.moringaschool.com

James Kimani
james.irungu1@student.moringaschool.com

Dorothy Chomba
dorothy.chomba@student.moringaschool.com

Wilfred Likishorumingi
wilfred.lekishorumongi@student.moringaschool.com


THANK'S FOR
WATCHING