OPTIMIZING PROPERTY PRICING FOR FALCON REAL ESTATE AGENCIES

Project by: Group 2
Members: Rwenji Murengaitta, Maureen Oketch,
Brian Korir, Mary Mwangi and Peter Kinyanjui

CONTENT

PROJECT OVERVIEW

- Falcon Real Estate agencies, a well-established real estate firm located in King County, USA, is seeking assistance in identifying the primary factors that affect property values within the region. The agency aims to leverage this knowledge to inform their pricing strategies for residential properties.

RESEARCH QUESTIONS

- To identify which factors influence real estate prices?

- To identify which features boost or reduce the property values?

- To identify how to use the predictive model to spot overpriced and underpriced properties?

- To identify the Leverage insights for a more profitable real estate operation?

OBJECTIVES

- **Main Objective**:
- Create an accurate predictive model for competitive property pricing.
- **Specific Objectives**:
1. Identify key price determinants.
2. Discover factors impacting house values.
3. Differentiate overpriced from underpriced properties.
4. Enhance agency revenue through analytical insights and pricing strategy.

DATA SOURCE AND DESCRIPTION

- This dataset contains house sale prices for King County, USA. It includes homes sold between May 2014 and May 2015.
  - The dataset contains 21 columns and 21,597 rows. This means there are 21 different variables each with 21,597 records.
  - There are three main data types in the data; float, integer and object.

DATA CLEANING

- Waterfront, view and year renovated had 2376, 63 and 3842 missing values respectively.

- Dropped the view column

- Replaced the missing values in waterfront with the mode of the column

- Replaced missing values in year renovated with corresponding year in the year built

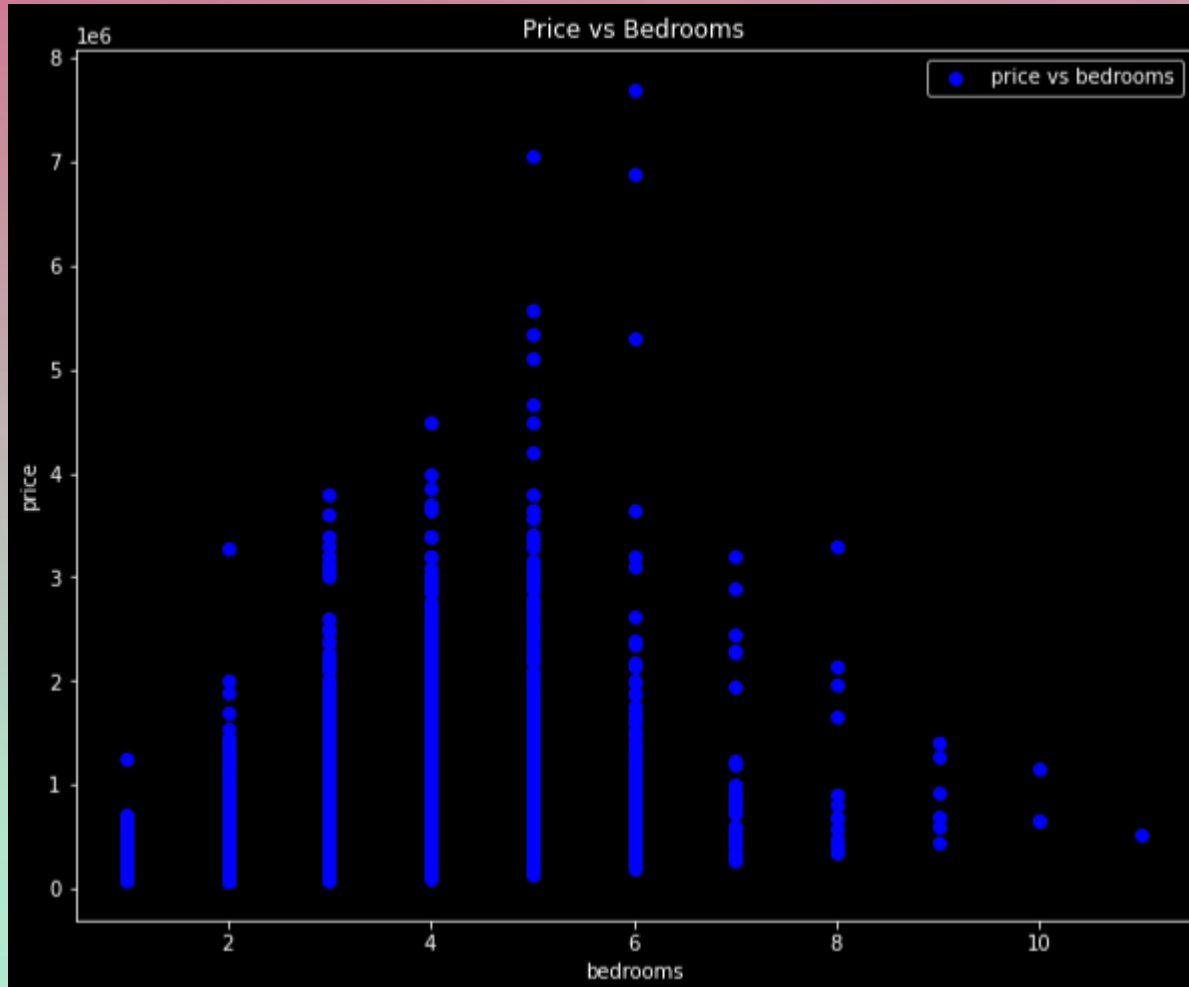- Created new columns 'age'

- Removed outliers.

DATA SELECTION

- Dependent variable: Price

- Independent variables: sqft_living, bedrooms, bathrooms, floors, age and condition
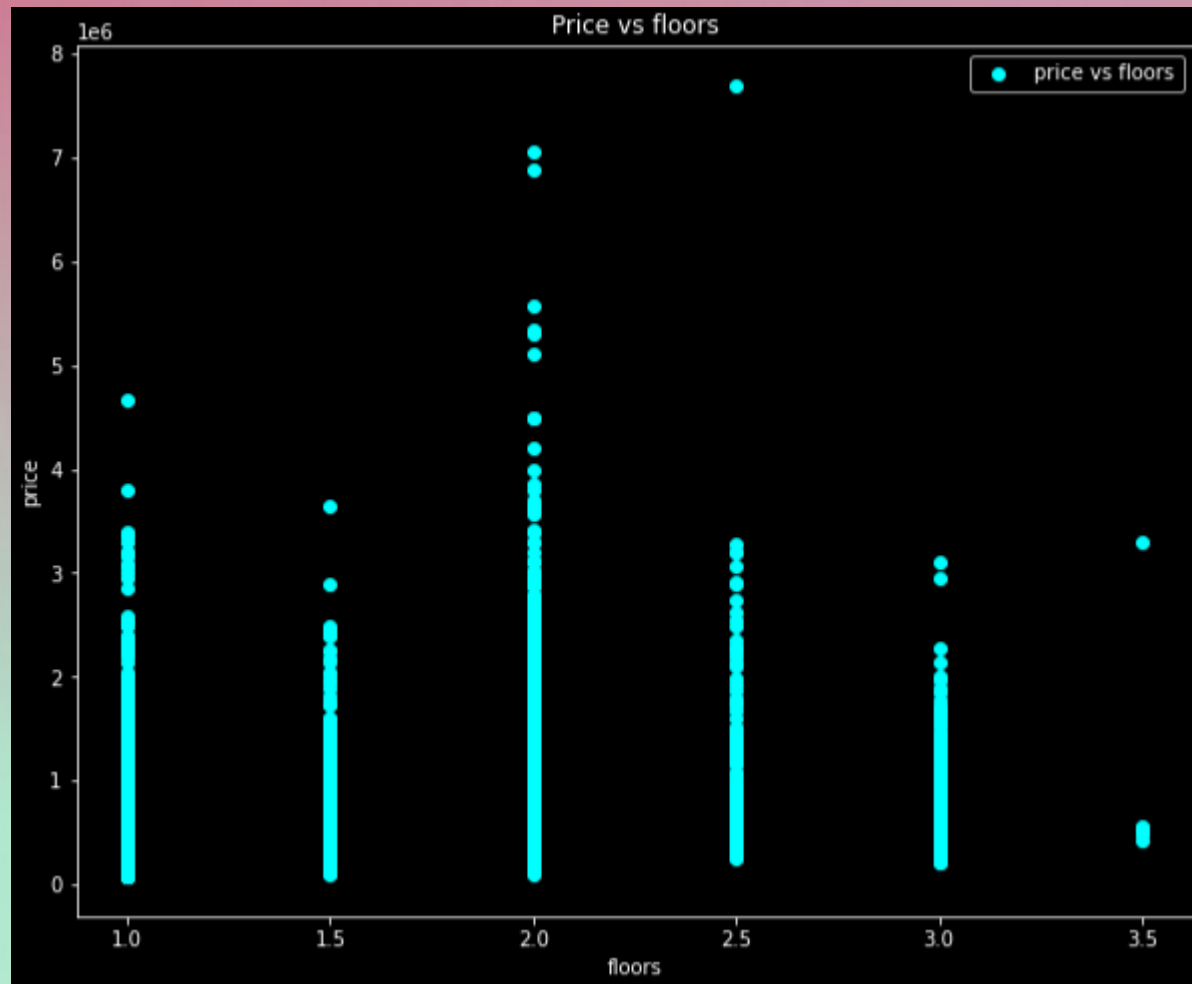
DATA VISUALIZATION

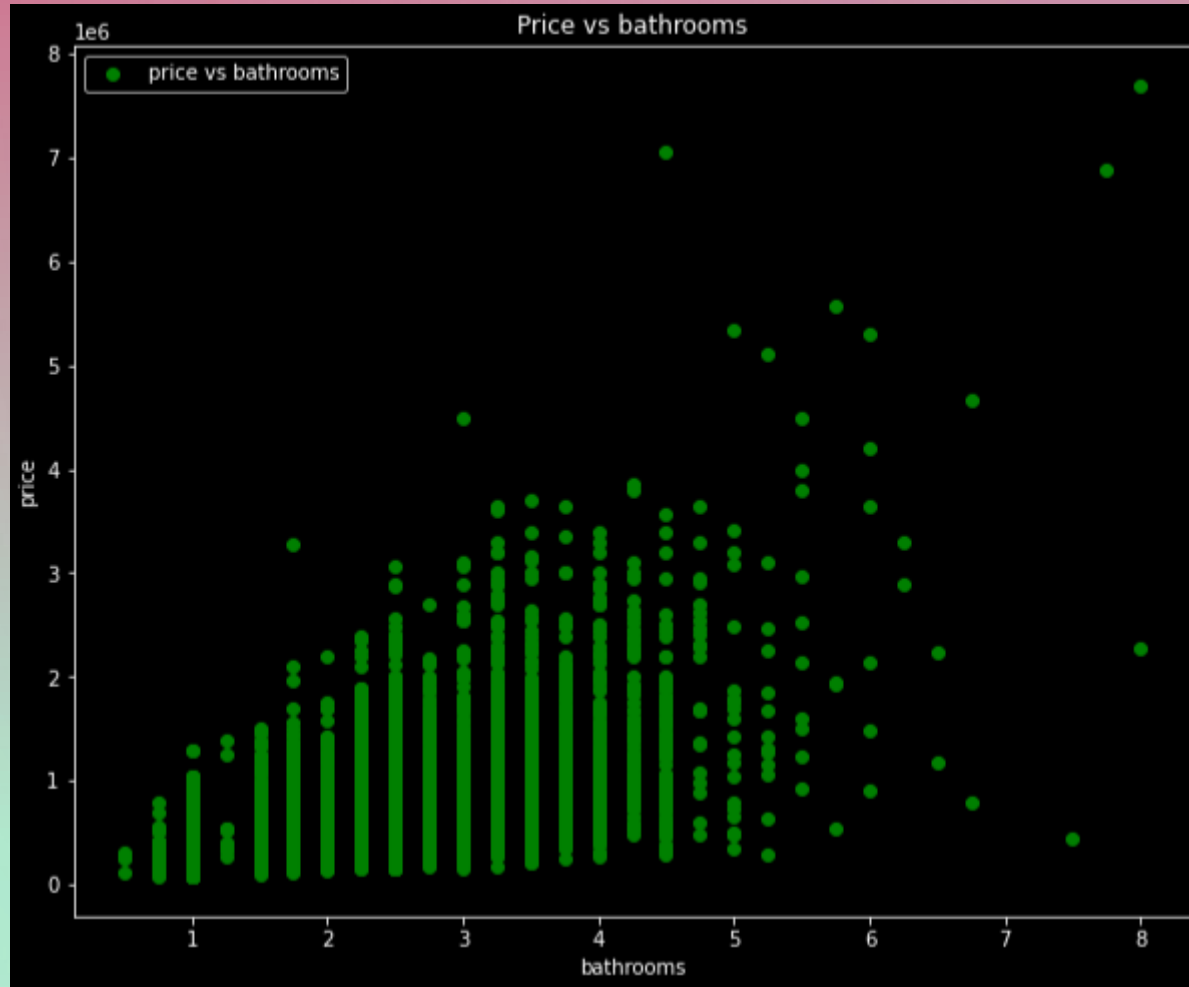- Different visuals prepared to show the findings of the data sets

Price vs Bedrooms

- From the graph its clear that the property fetch a higher price as the bedrooms increase

- The houses with floors from 1.0 to 3.0 seem to increase the value of the property

- It's clear that as the number of bathrooms increases the price of the property increases

- The age of the property against price varies this is because they may be some other variables that may be affecting the price

Price vs sqft_living
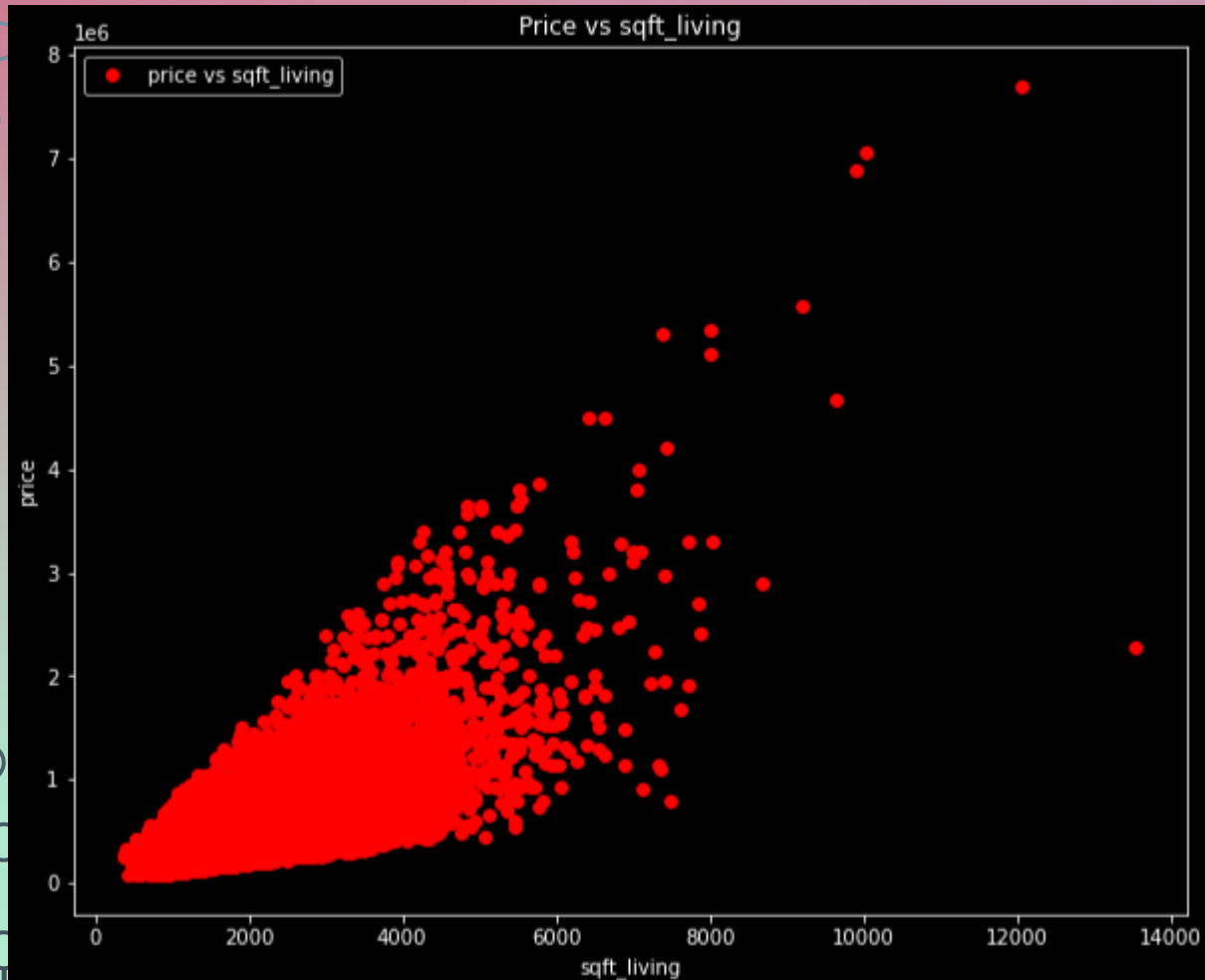
- The graph demonstrates a positive correlation, the sqft_living increases as the price of the property increases

PRICE VS CONDITION

- It is clear that the condition of the property does not have a significant impact on price

SIMPLE LINEAR REGRESSION

OLS Regression Results

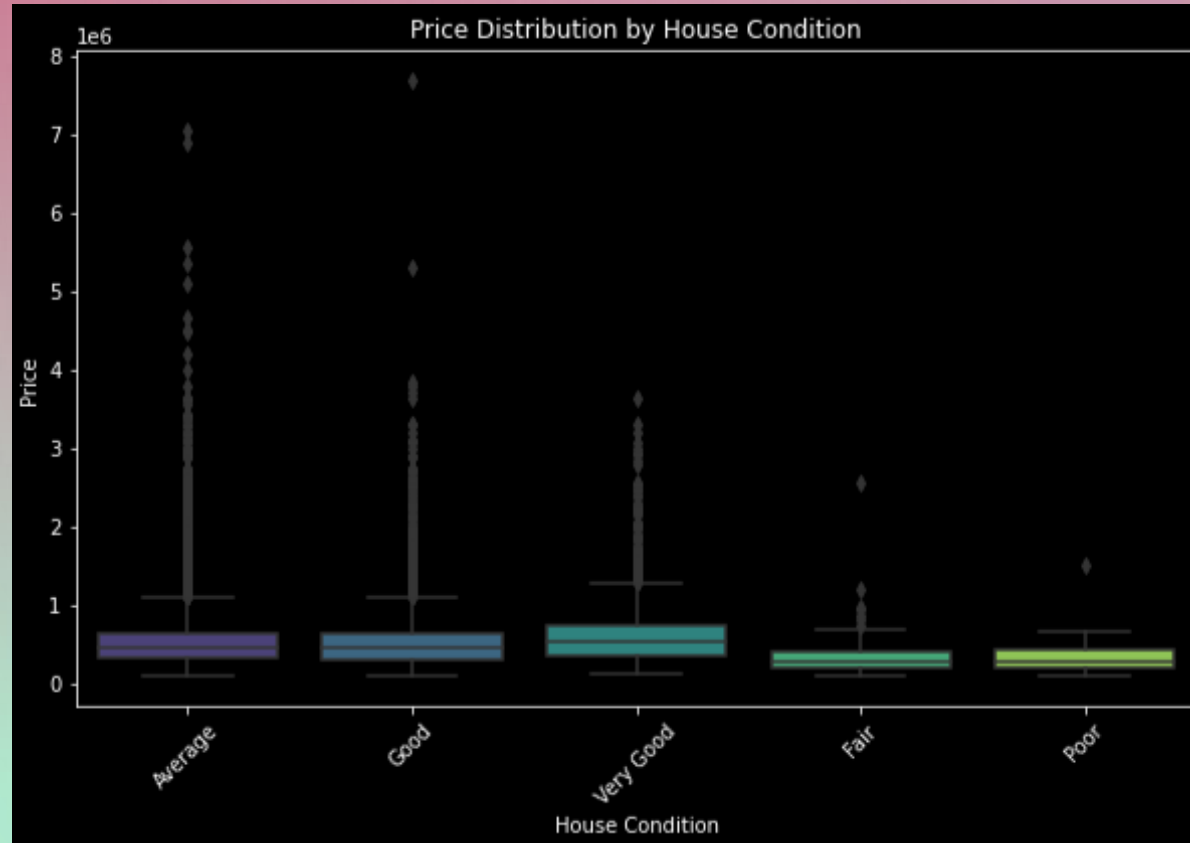|			
Dep. Variable:	price	R-squared:	0.492
Model:	OLS	Adj. R-squared:	0.492
Method:	Least Squares	F-statistic:	2.075e+04
Date:	Sat, 04 Nov 2023	Prob (F-statistic):	0.00
Time:	00:53:29	Log-Likelihood:	-2.9762e+05
No. Observations:	21419	AIC:	5.953e+05
Df Residuals:	21417	BIC:	5.953e+05
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-4.363e+04	4437.771	-9.831	0.000	-5.23e+04	-3.49e+04
sqft_living	280.8010	1.949	144.064	0.000	276.981	284.622

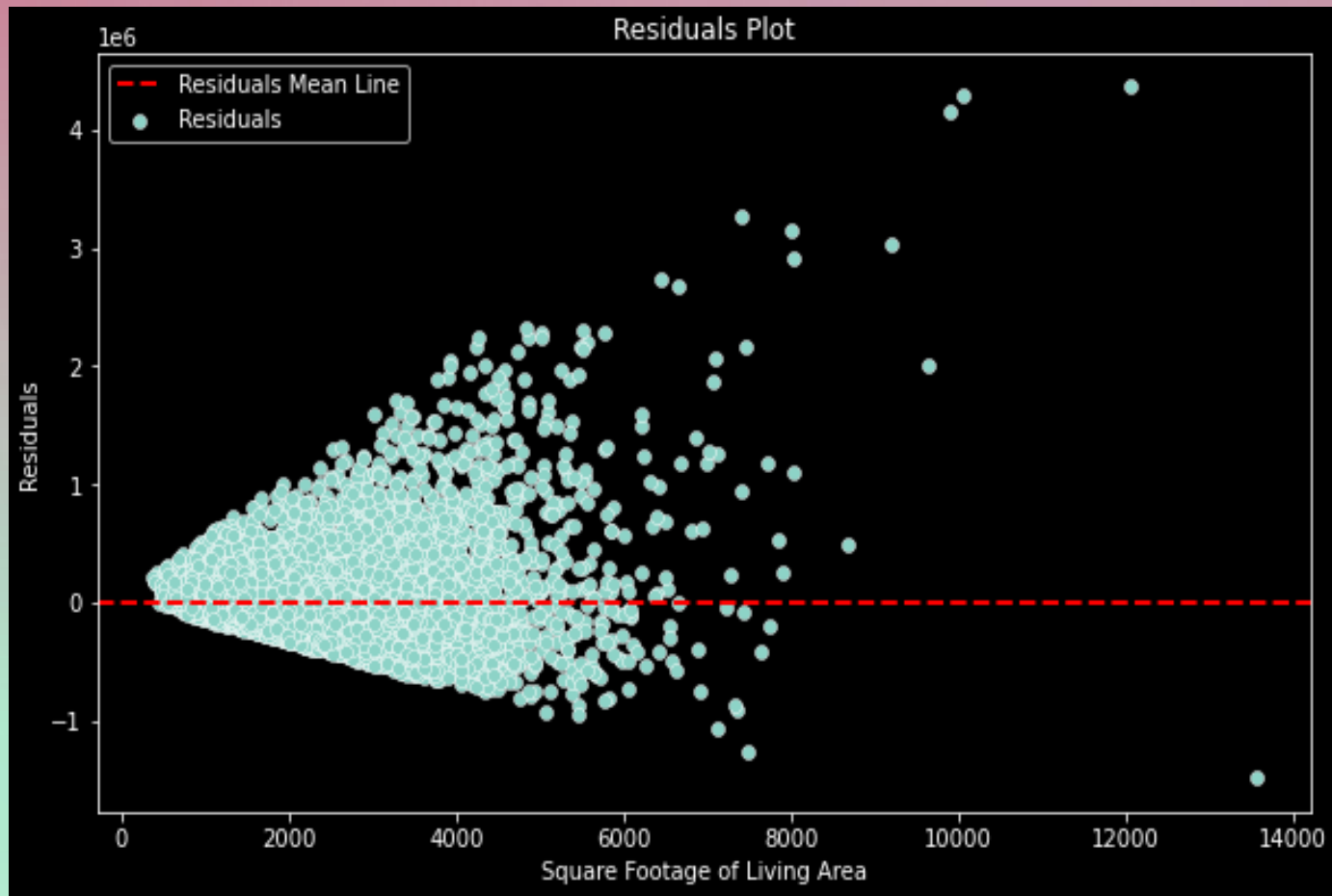|			
Omnibus:	14693.177	Durbin-Watson:	1.042
Prob(Omnibus):	0.000	Jarque-Bera (JB):	539257.422
Skew:	2.823	Prob(JB):	0.00
Kurtosis:	26.924	Cond. No.	5.64e+03

From our simple regression model, we got a y-intercept of $-43628.22, Then we can say that from every increase of 1 square foot above ground the price increases by $280.8

Simple Linear Regression: Price vs. Sqft Living Area

This plot shows the actual values vs the predicted ones

CREATING A NEW PLOT FOR THE RESIDUALS

- The R-squared value stands at 0.492, signifying that around 49.2% of the variability in 'price' can be accounted for by the linear association with 'sqft_living'.

- The model's coefficients (const and sqft_living) are both statistically significant, with p-values for their t-statistics comfortably below 0.05.

- the simple linear regression equation, derived from this output, is as follows:
  price = -43,630 + 280.801 * sqft_living

- This equation delineates the connection between a house's 'price' and its living space's square footage ('sqft_living')

MULTIPLE LINEAR REGRESSION

OLS Regression Results

Dep. Variable:	price	R-squared:	0.558
Model:	OLS	Adj. R-squared:	0.558
Method:	Least Squares	F-statistic:	3003.
Date:	Sat, 04 Nov 2023	Prob (F-statistic):	0.00
Time:	00:53:57	Log-Likelihood:	-2.9614e+05
No. Observations:	21419	AIC:	5.923e+05
Df Residuals:	21409	BIC:	5.924e+05
Df Model:	9		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-1.915e+05	1.17e+04	-16.320	0.000	-2.15e+05	-1.69e+05
bedrooms	-7.47e+04	2352.760	-31.751	0.000	-7.93e+04	-7.01e+04
sqft_living	302.3099	2.973	101.698	0.000	296.483	308.137
age	3304.6411	73.993	44.662	0.000	3159.610	3449.672
floors	5.797e+04	3834.261	15.119	0.000	5.05e+04	6.55e+04
bathrooms	6.942e+04	3883.020	17.879	0.000	6.18e+04	7.7e+04
condition_Average	-2.703e+04	8860.422	-3.050	0.002	-4.44e+04	-9658.693
condition_Fair	-8.255e+04	1.79e+04	-4.623	0.000	-1.18e+05	-4.76e+04
condition_Good	-1.604e+04	8992.428	-1.784	0.074	-3.37e+04	1582.235
condition_Poor	-7.737e+04	3.87e+04	-1.997	0.046	-1.53e+05	-1434.579
condition_Very_Good	1.144e+04	1e+04	1.139	0.255	-8239.978	3.11e+04

Omnibus:	14166.395	Durbin-Watson:	1.236
Prob(Omnibus):	0.000	Jarque-Bera (JB):	522975.847
Skew:	2.668	Prob(JB):	0.00
Kurtosis:	26.612	Cond. No.	2.00e+19

R-squared (R^2): The R-squared value is 0.558, signifying that approximately 55.8% of the variation in 'price' can be accounted for by the linear relationship with the ten independent variables ('sqft_living', 'bathrooms', 'bedrooms', 'floors', 'age','condition_Average''condition_Fair', 'condition_Good', 'condition_Poor', 'condition_Very Good')
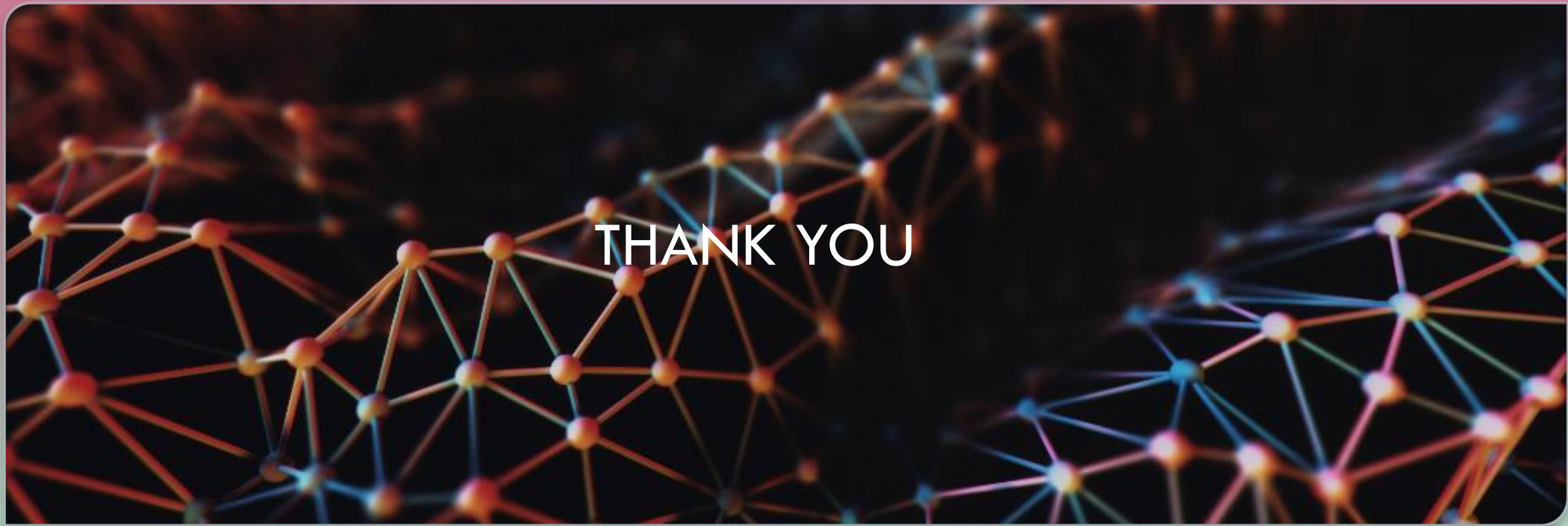
INTERPRETATION OF THE RESULTS

- The model demonstrates overall statistical significance, with an F-statistic p-value significantly below 0.05.

- Among the independent variables, 'const', 'sqft_living', 'bathrooms', 'bedrooms', 'floors', 'condition_Avarage ', and 'condition_Fair' exhibit statistical significance, with p-values below our significance level (alpha of 0.05). This indicates that these variables have a substantial impact on 'price'.

- In contrast, 'condition_Good', 'condition_Poor'and condition_Very_Good' have p-values exceeding our alpha of 0.05, at 0.074 ,0.046 and 0.255, respectively, implying that they are not 'statistically significant in predicting 'price

CONCLUSION

- Important factors affecting the house prices include; bedrooms, sqft_living, bathrooms, floors, age and condition
- The best model out of the two was the multiple linear regression
- The model is statistically significant, as evidenced by the F-statistic of 3,003 and a corresponding p-value of 0.00
- An increase with one bedroom decreases the house sale by $ 0.002.
- An increase with one bathroom increases the house price by $ 0.0725
- there is no significant increase/decrease in the house price with the condition of the house
- An increase in the square footage of the home by one square foot increases the price of the house by $ 0.0002
- An increase in floors by one increases price by $0.0657

RECOMMENDATIONS

- majority of buyers are looking for houses with fewer bedrooms, it may be beneficial to focus on properties with a lower bedroom count
- An increase in the number of bathrooms tends to have a positive impact on the house price. If feasible, consider adding or upgrading bathrooms in properties to attract buyers who value this feature
- Houses with multiple floors tend to have higher prices. If feasible, explore opportunities to add or emphasize multiple floors in properties to increase their perceived value
- Increasing the size of the living space generally increases the house price, while larger lot sizes may have a slight negative impact. It's good to consider the preferences of the target market.
- The condition and the age of the property does not have a significant impact on the price

THANK YOU