MAXIMIZING HOME VALUE

(INSIGHTS ON THE IMPACT OF RENOVATIONS IN KING COUNTY)

MORINGA
Discover · Grow · Transform

Group 12

AUTHORS

- KEZIAH GICHEHA
- DAVE OMONDI
- PAMELA OKINYO
- CHARLES NDEGWA
- BRIAN KIPNG'ENO

Purpose:

This presentation will explore how specific home renovations can increase home value in King County.

AGENDA:

- OVERVIEW OF THE REAL ESTATE MARKET IN KING COUNTY.
- IDENTIFY KEY RENOVATIONS THAT OFFER THE BEST RETURN ON INVESTMENT.
- PRICE VS. VALUE ANALYSIS
- Q&A SESSION: AN OPPORTUNITY FOR PERSONALIZED ADVICE.

Business Problem:

Impact of renovation on pricing of houses

- Real estate agencies face the challenge of providing reliable advice to homeowners regarding renovations.
- Homeowners are often uncertain about which renovations will yield the highest return on investment.

Objective:

- Our objective is to conduct a comprehensive data analysis to pinpoint the renovations that offer the best return on investment in King County.
- This analysis will help real estate agencies provide informed advice and help homeowners make smart investment decisions.

Methodology

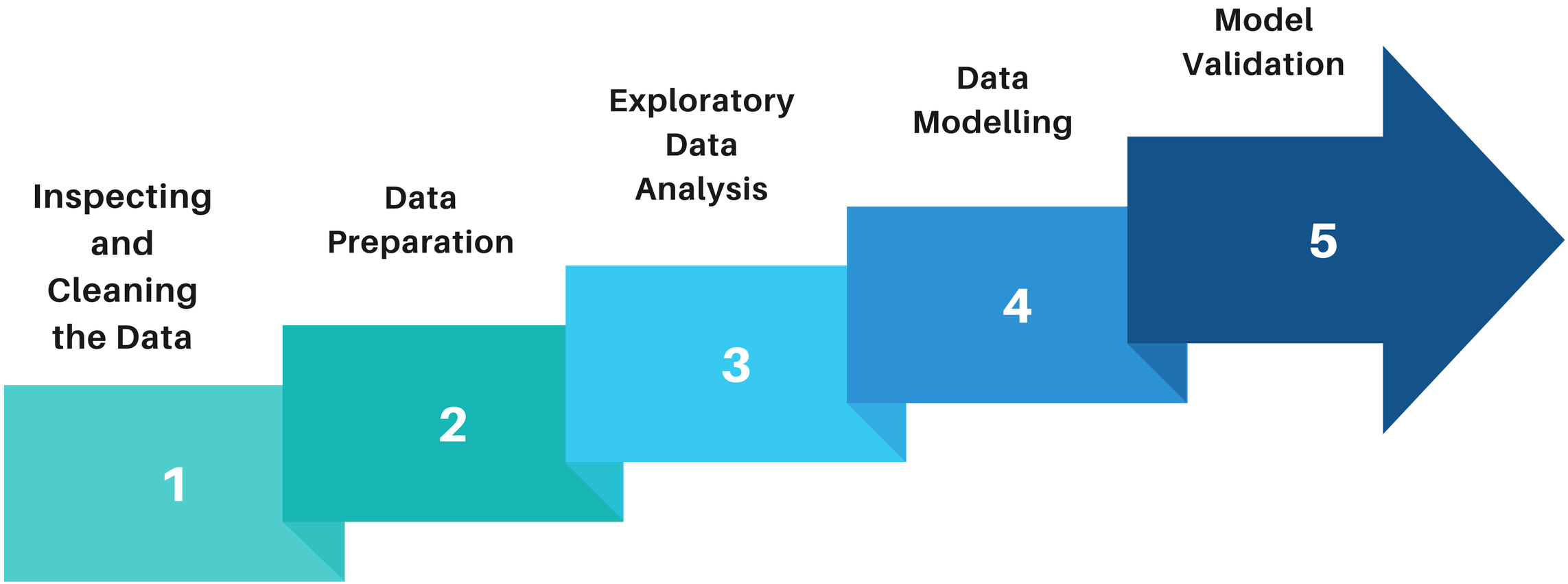
Data and Tools

- Dataset: King County House Sales dataset, which includes a comprehensive set of variables such as (price, bedrooms, bathrooms, sqft_living, sqft_lot, floors, waterfront, view, condition, grade, year_built, and zipcode.)
- The dataset covers a significant number of records, providing a robust foundation for our analysis

Tools Used:

- **Python:** Our primary tool for data processing and analysis.
- **Pandas**: Used for data manipulation, cleaning, and preparation.
- **Seaborn:** Employed for data visualization to uncover trends and insights.
- **Statsmodels**: Utilized for statistical modeling and analysis.
- **Scikit-learn:** Applied for machine learning model development and evaluation

STEPS


Inspecting and Cleaning the Data
1
Data Preparation
2
Exploratory Data Analysis
3
Data Modelling
4
Model Validation
5

Data Preparation

Cleaning:

- We cleaned the dataset by handling missing values through imputation or removal, and correcting data types where necessary
- Outliers were identified and addressed to ensure the data's integrity.

Normalization:

- Numerical features were standardized to a common scale to ensure uniformity and improve model performance.

Outliers:

- Outliers were identified using statistical methods and were treated appropriately to minimize their impact on the analysis.

Transformations:

- Log transformations were applied to skewed data to normalize distributions and improve model accuracy.

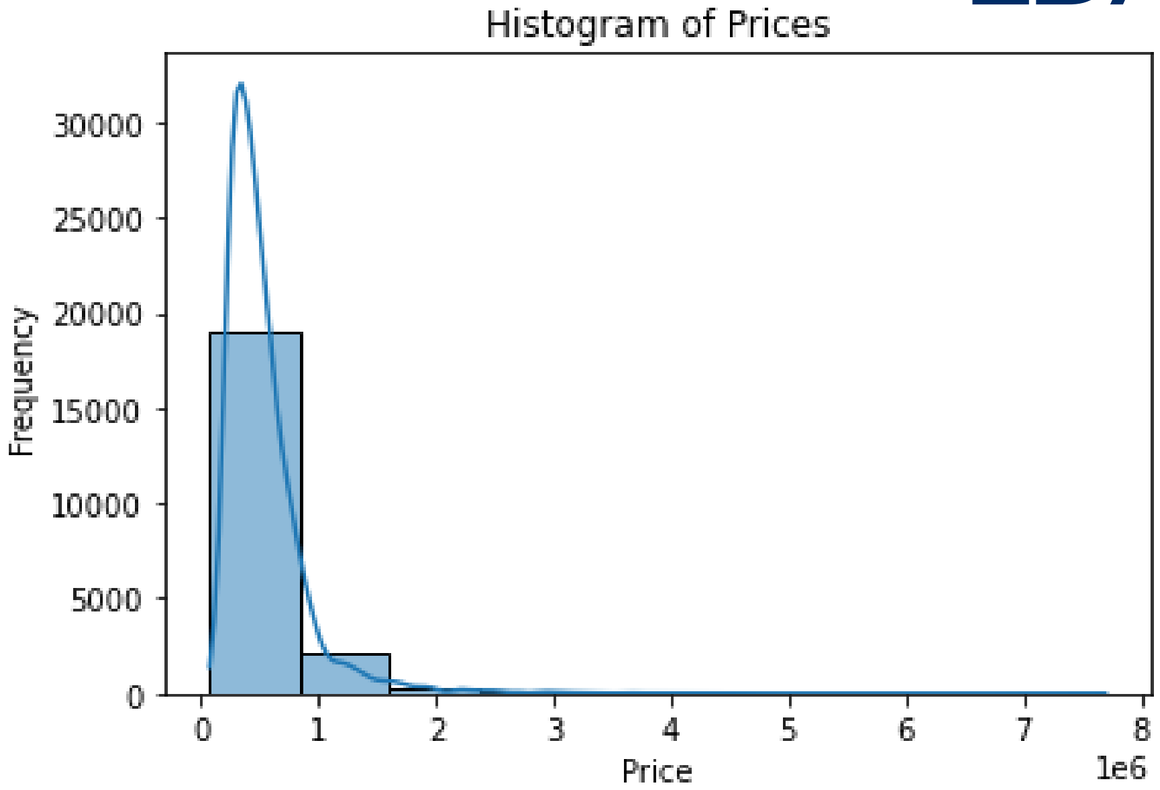
UNDERSTANDING OF DATA

Statistic	id	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	condition	grade	sqft_above	sqft_basement	yr_built	yr_renovated	zipcode	lat	long	sqft_living15	sqft_lot15
Count	21,597	21,597	21,597	21,597	21,597	21,597	21,597	21,597	21,597	21,597	21,597	21,597	21,597	21,597	21,597	21,597	21,597	21,597	21,597	21,597
Mean	4,580,474,000	540,296.60	3.37	2.12	2,080.32	15,099.41	1.49	0.007	0.233	3.41	7.66	1,788.60	285.72	1,971	68.76	98077	47.56	-122.2	1,986.62	12,758.28
Std Dev	2,876,736,000	367,368.10	0.93	0.77	918.11	41,412.64	0.54	0.082	0.765	0.65	1.17	827.76	439.82	29.38	364.04	53.51	0.14	0.14	685.23	27,274.44
Min	1,000,102	78,000	1	0.5	370	520	1	0	0	1	3	370	0	1,900	0	98001	47.16	-122.5	399	651
25%	2,123,049,000	322,000	3	1.75	1,430	5,040	1	0	0	3	7	1,190	0	1,951	0	98033	47.47	-122.3	1,490	5,100
50%	3,904,930,000	450,000	3	2.25	1,910	7,618	1.5	0	0	3	7	1,560	0	1,975	0	98065	47.57	-122.2	1,840	7,620
75%	7,308,900,000	645,000	4	2.5	2,550	10,685	2	0	0	4	8	2,210	550	1,997	0	98118	47.68	-122.1	2,360	10,083
Max	9,900,000,000	7,700,000	33	8	13,540	1,651,359	3.5	1	4	5	13	9,410	4,820	2,015	2,015	98199	47.78	-121.3	6,210	871,200

From above; we can confirm that

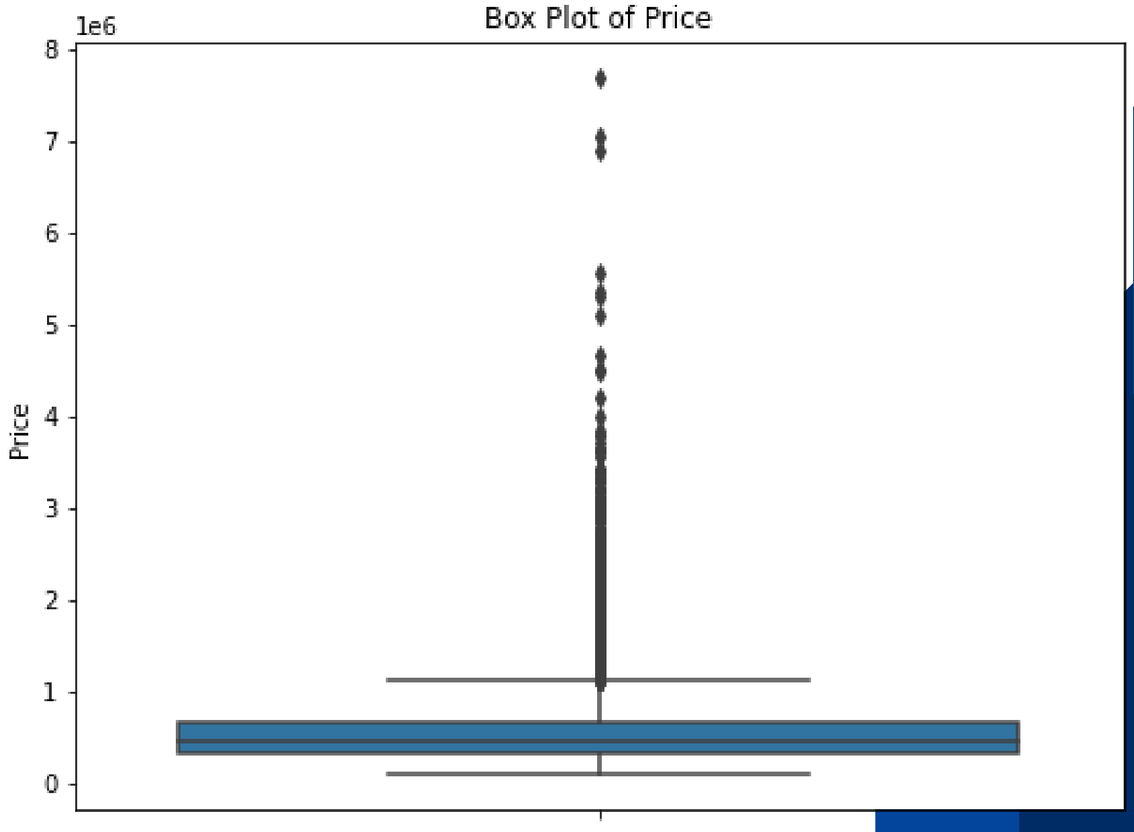
1. Home prices range from $78,000 to $7,700,000.
2. Most homes are priced between $322,000 and $645,000.
3. The average home has 3.3 bedrooms and 2.1 bathrooms, with approximately 2,080 square feet of living space.
4. All homes have between 1 and 3.5 floors.
5. We observed a listing for a home with 33 bedrooms. This might be an extreme outlier.

EDA (Univarite Analysis)

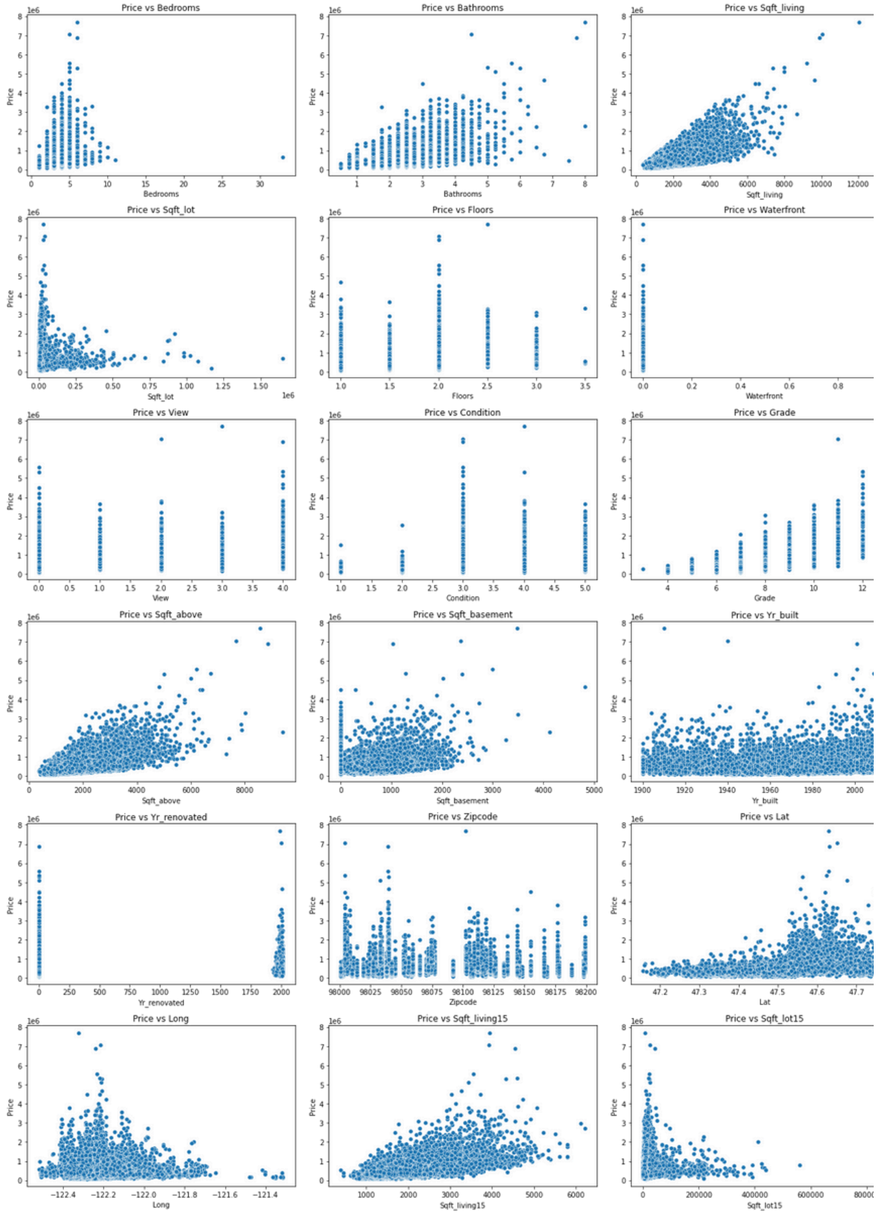


THE DISTRIBUTION OF HOUSE PRICES IS RIGHT-SKEWED (POSITIVELY SKEWED), INDICATING THAT MOST HOUSES ARE RELATIVELY INEXPENSIVE, WHILE A SMALL NUMBER OF HOUSES ARE VERY EXPENSIVE. MOST HOUSES PRICED BELOW $1,000,000.

BASED ON THE BOX PLOT, THERE ARE OUTLIERS PRESENT, BUT WE CHOSE TO RETAIN THEM, ASSUMING THEY ACCURATELY REFLECT THE REAL-WORLD DATASET.



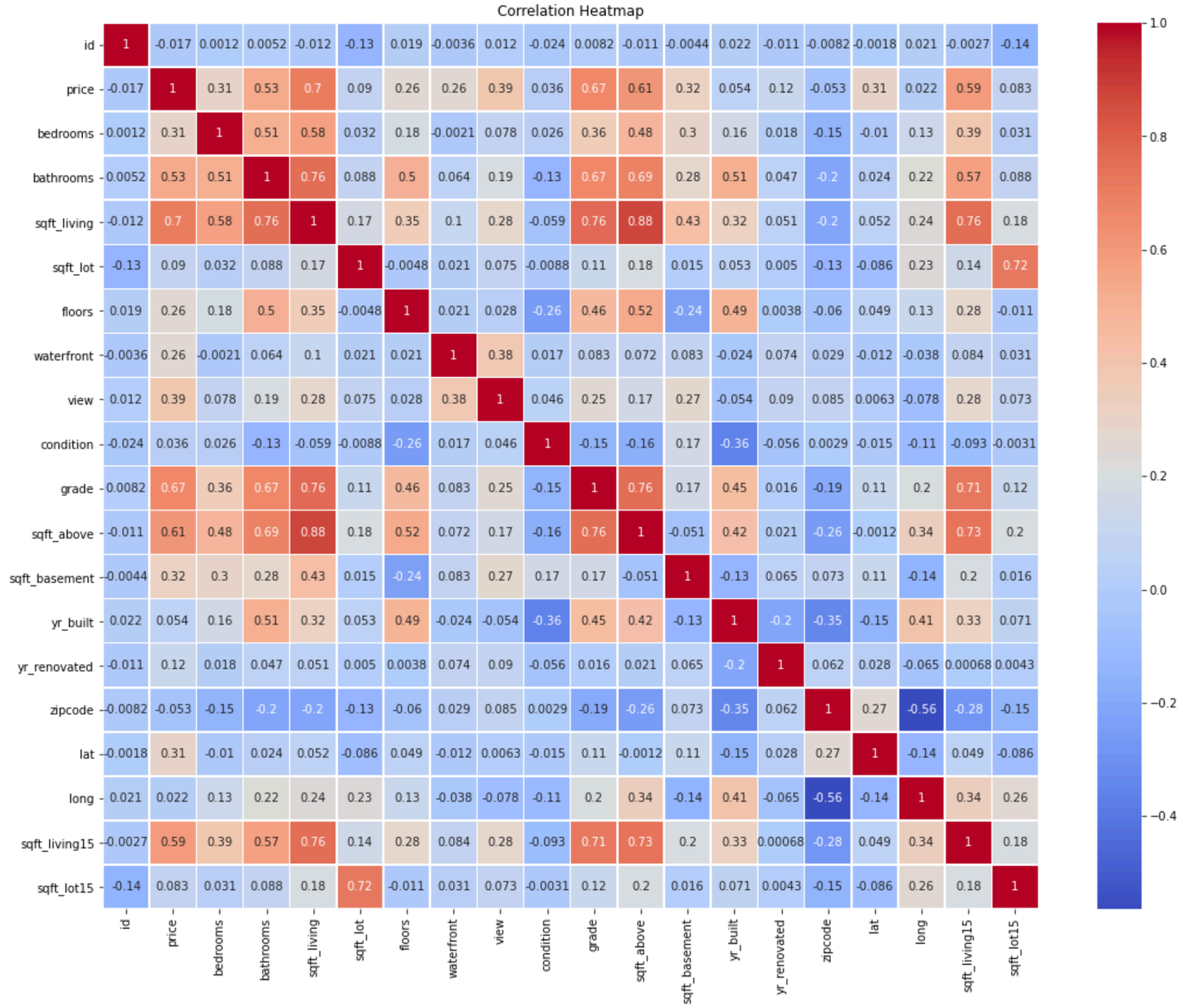
EDA (Bi-varite Analysis)



THIS IS A DISPLAY OF THE RELATIONSHIP BETWEEN PRICE OF A HOUSE AND ALL OTHER VARIABLES

EDA (Bi-varite Analysis)

A HEATMAP OF ALL VARIABLES AND THEIR RELATIONSHIPS



*From this we can observe that;

the size of the living space has a high impact on the price of the house. corr = 76%

* the grade of thehouse also affects how to price the house with a correlation of 67%

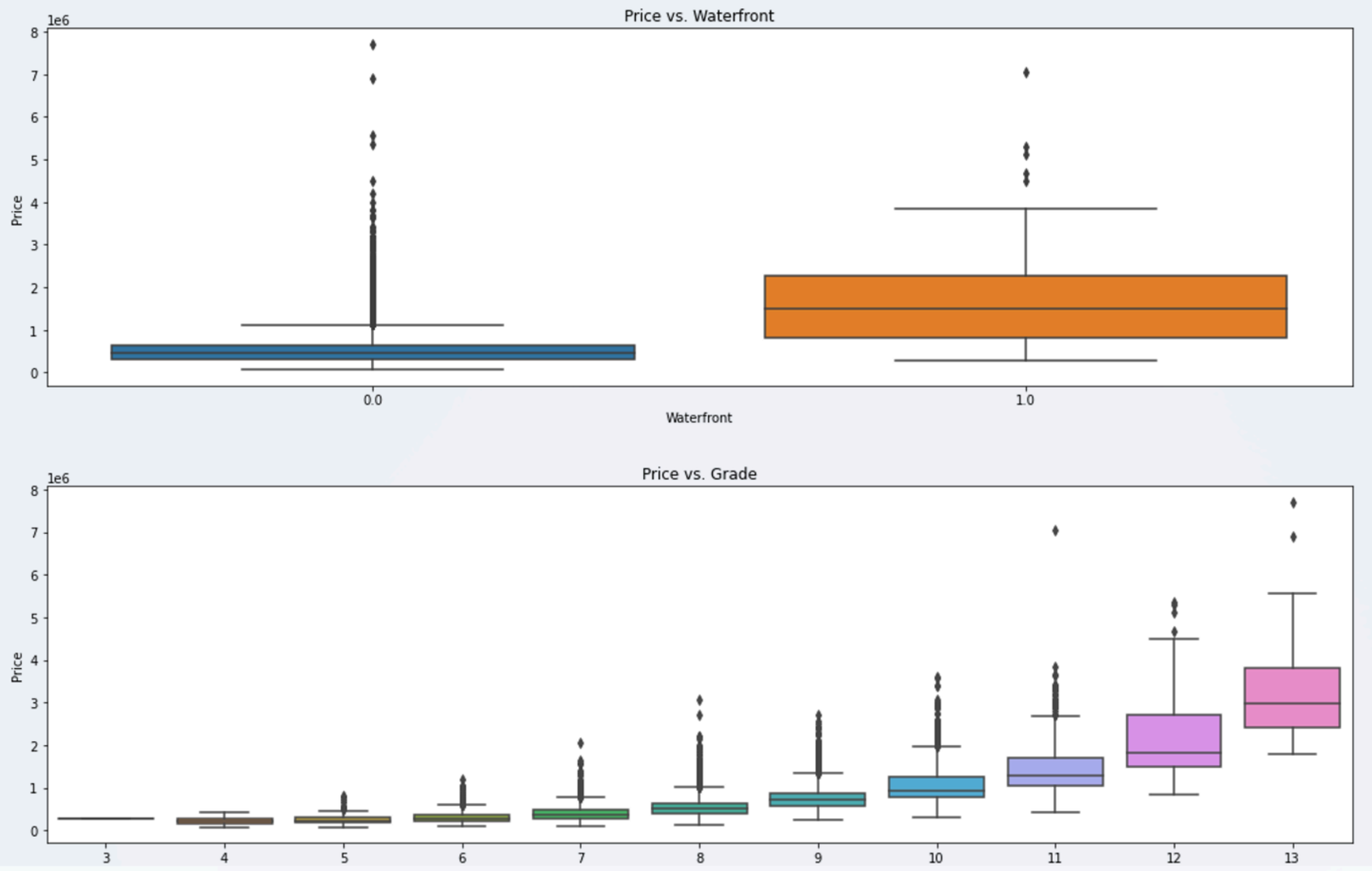
* the size of the house excluding the basement might slightly affect the way you price the house and so does the number of bathrooms with 61% and 53% respectively).

* the prices of the square footage of interior housing living space within the nearest 15 neighbours is is slightly similar with a correlation of 59%

EDA (Bi-varite Analysis)

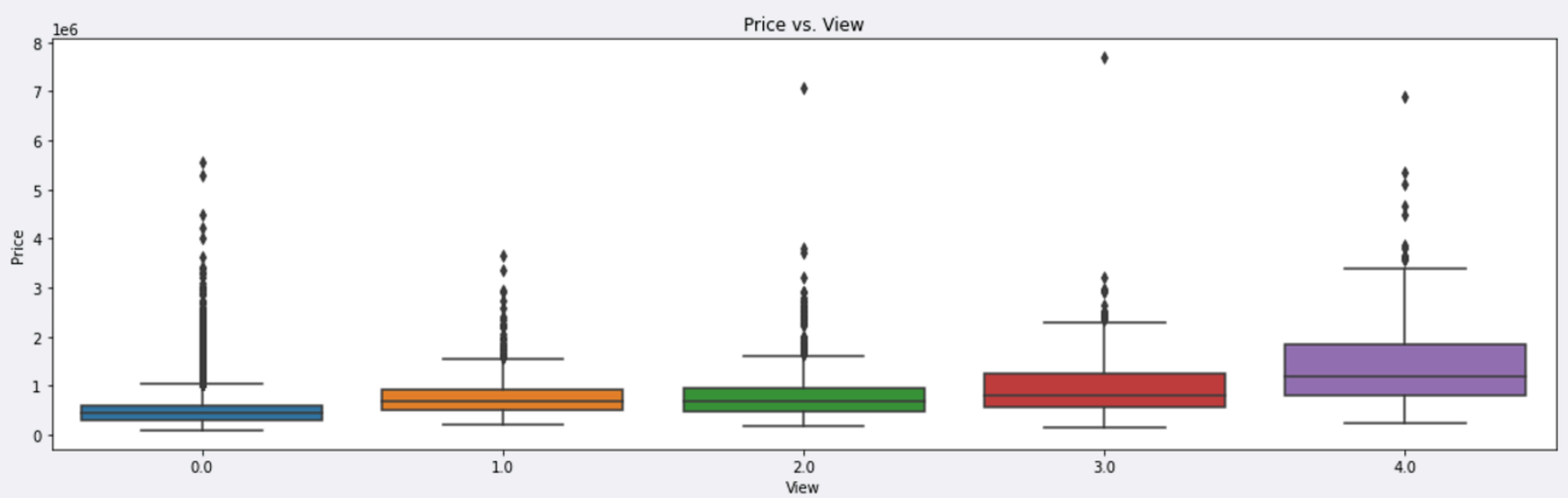
IMPACT OF VARIABLES ON PRICES OF THE HOUSES IN KING COUNTY

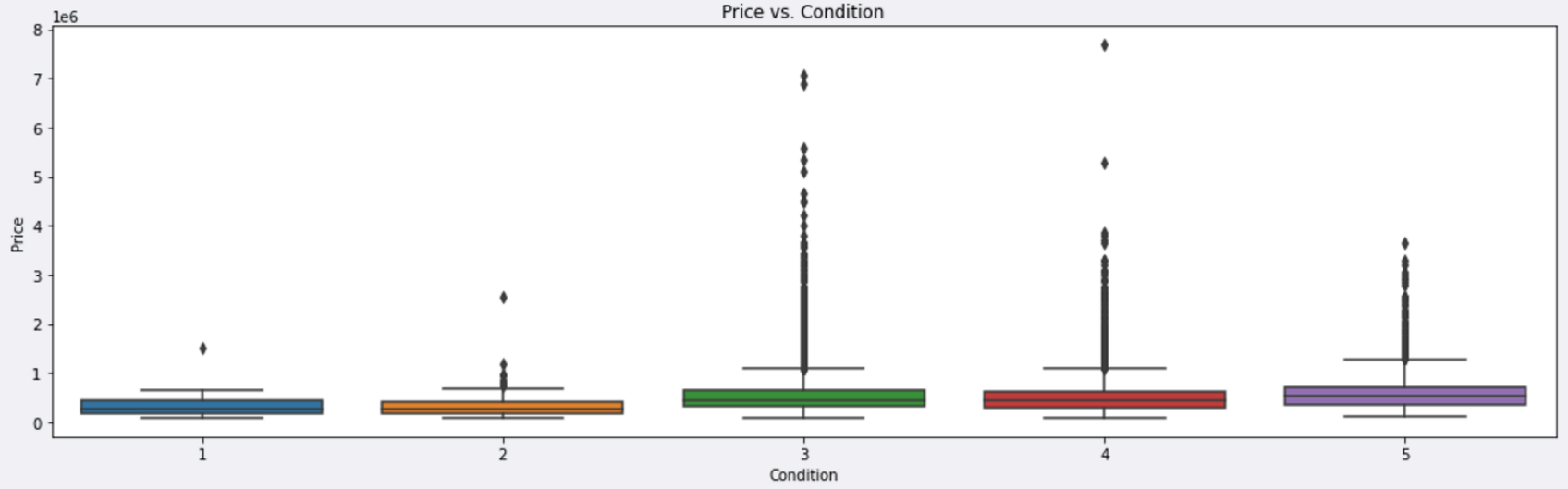
Houses with better views , higher grades(ratings), with the presence of a waterfront, better conditions tend to be priced higher, and the zip code also has an influence on the price. The house price increases steadily as the number of floors increases only to a certain limit (3 floors) then decreases when the floors become greater than 3

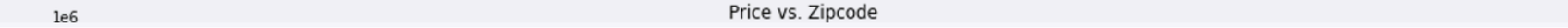


EDA (Bi-varite Analysis)

IMPACT OF VARIABLES ON PRICES OF THE HOUSES IN KING COUNTY

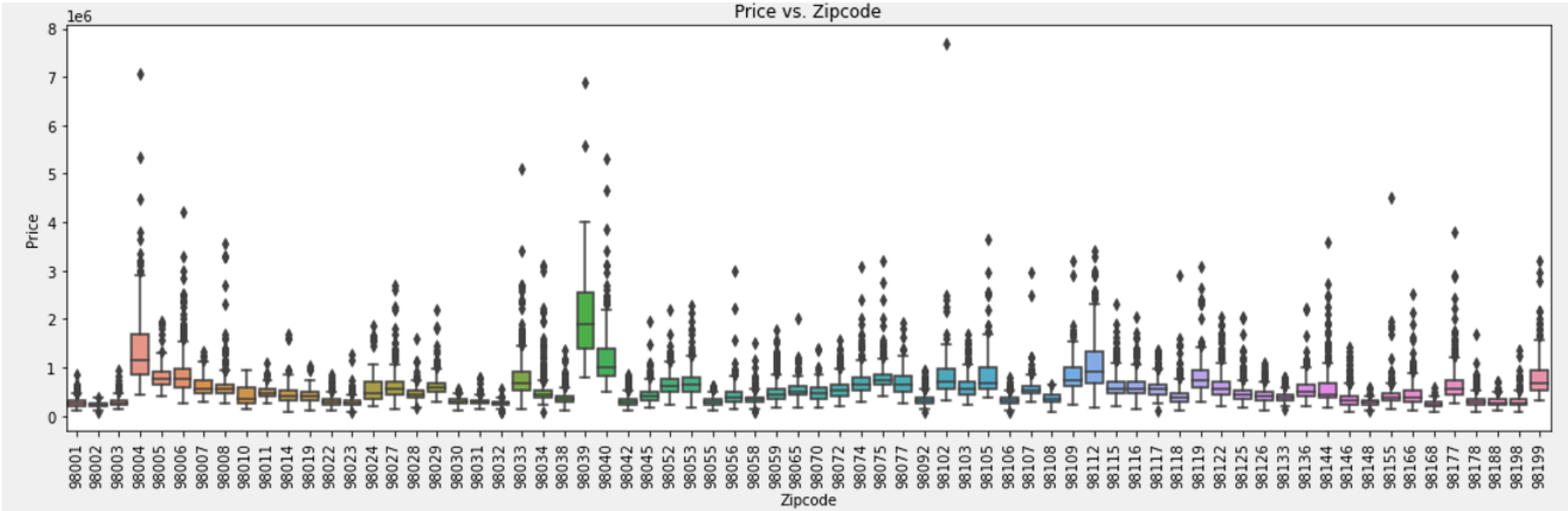


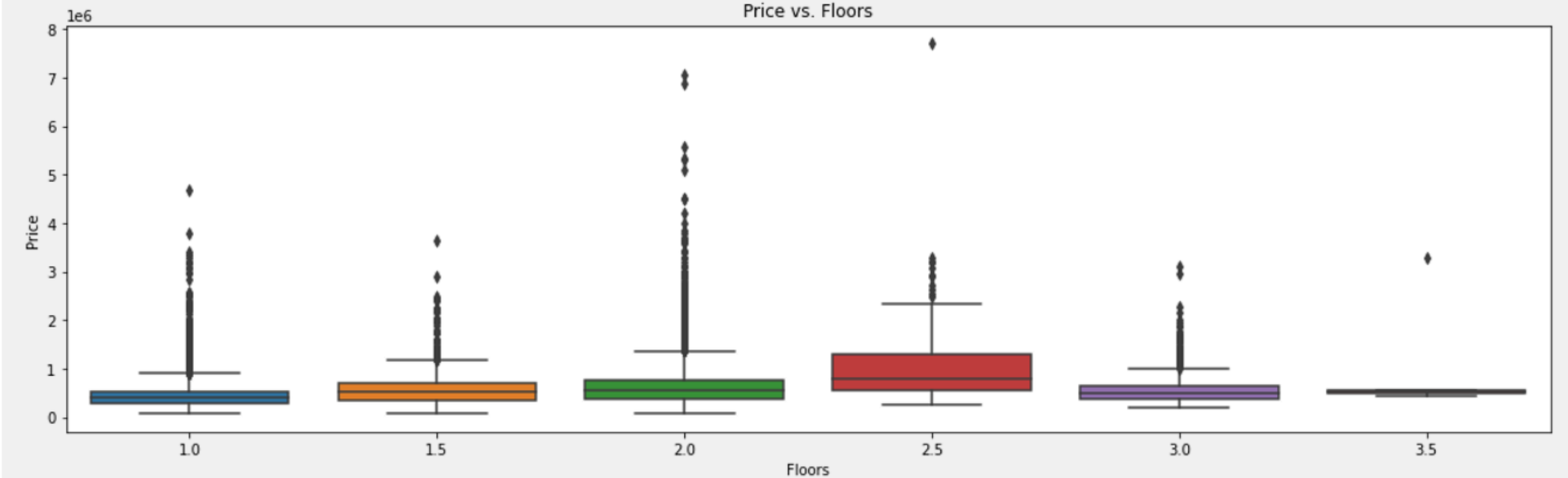




EDA (Bi-varite Analysis)

IMPACT OF VARIABLES ON PRICES OF THE HOUSES IN KING COUNTY





DATA MODELLING

Baseline Model/Initial Model/First Model

After using simple linear regression, using the most highly correlated feature. Results are as shown.

OLS Regression Results	
Dep. Variable:	price
R-squared:	0.49
Model:	OLS
Adj. R-squared:	0.49
Method:	Least Squares
F-statistic:	16760.00
Date:	Wed, 17 Jul 2024
Prob (F-statistic):	0.00
Time:	0.55
Log-Likelihood:	-240120.00
No. Observations:	17277.00
AIC:	480200.00
Df Residuals:	17275.00
BIC:	480300.00
Df Model:	1.00
Covariance Type:	nonrobust
	coef
---	---
const	-46450.00
sqft_living	282.20
Omnibus:	11495.54
Durbin-Watson:	2.00
Prob(Omnibus):	0.00
Jarque-Bera (JB):	371098.39
Skew:	2.74
Prob(JB):	0.00
Kurtosis:	25.04

EXPLANATION OF THE MODEL

- R-squared: Our model can explain about half of why house prices change based on the size of the living space.
- F-statistic: There is a strong link between living space size and house prices, where adding one square foot increases the price by around $282.
- RMSE: Our model's predictions are quite off, so we might need more information or a better model to get more accurate results.

In conclusion, while the model shows that larger living spaces generally lead to higher house prices, it doesn't fully explain the prices, and its predictions are not very accurate. We might need to consider more factors or use a more advanced model to better predict house prices.

Model Refinement

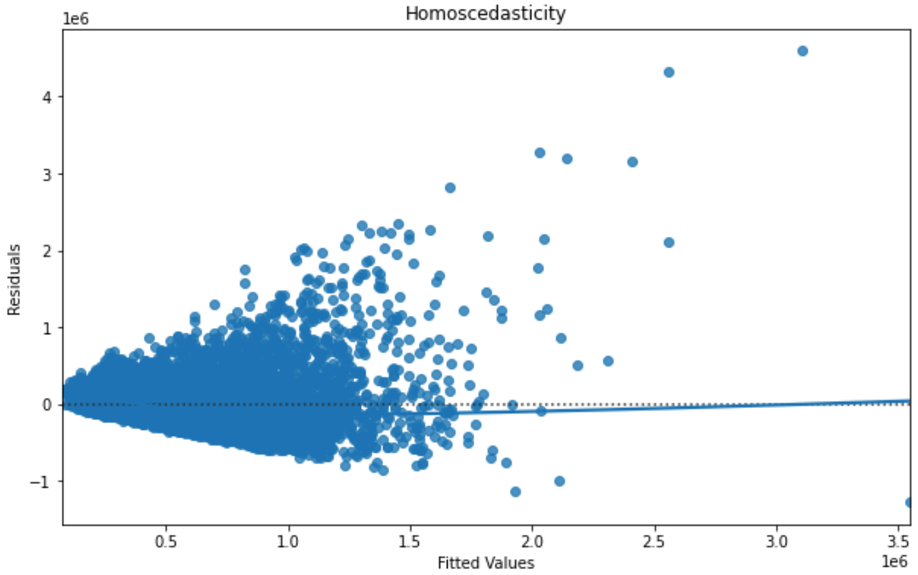
First Iteration (Testing for Homoscedasticity

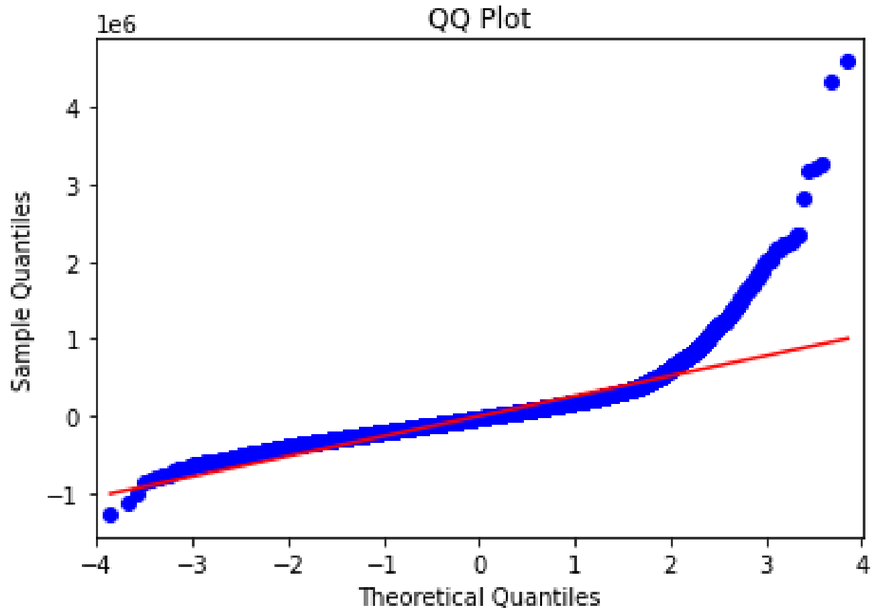
The following features have been added:

sqft_above

sqft_living15

bathrooms





OBSERVATIONS

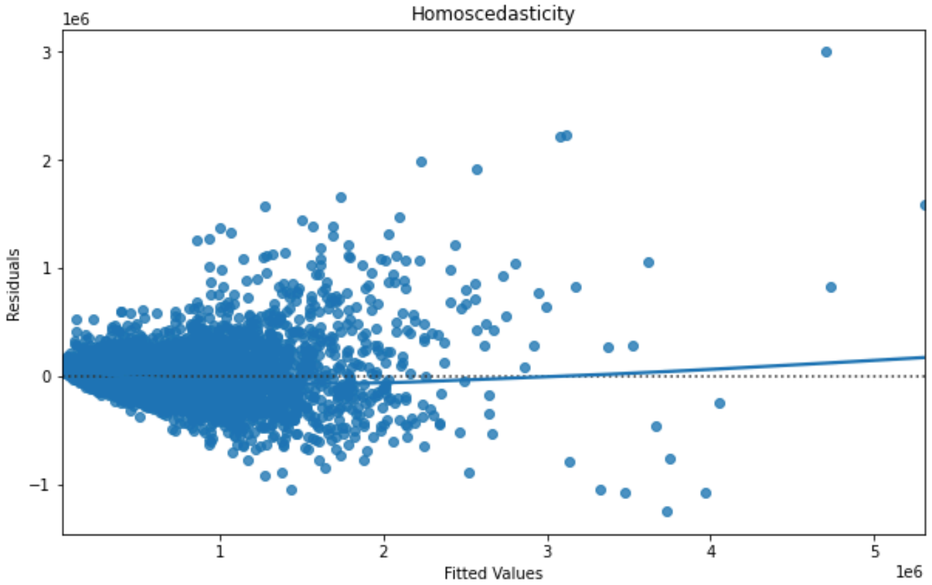
- **R-squared of 0.50: Our model can explain or predict about 50% of the variation in house prices. It's like being correct half of the time when guessing something.**
- **Heteroscedasticity: The errors in our model's predictions are not evenly spread out. This unevenness suggests issues like outliers or skewed data.**
- **Non-normal Residuals: The errors don't follow a normal, bell-shaped pattern, especially at the extremes, indicating that our model's mistakes are not distributed normally.**

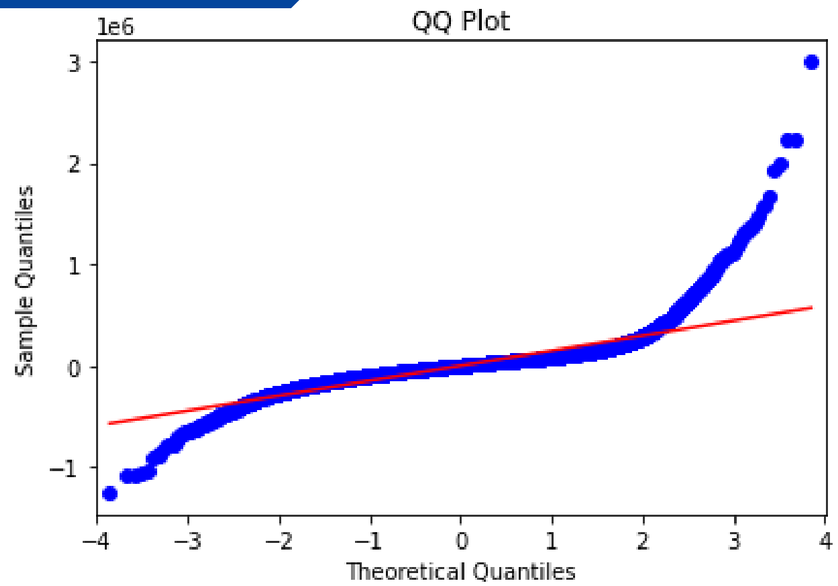
Findings

- Our model is only halfway good at predicting prices.
- The errors our model makes are kind of messy and uneven.
- The errors don't follow a nice, normal pattern.

Second Iteration

The following features have been added:
1.waterfront, view condition, grade, zipcode, floors





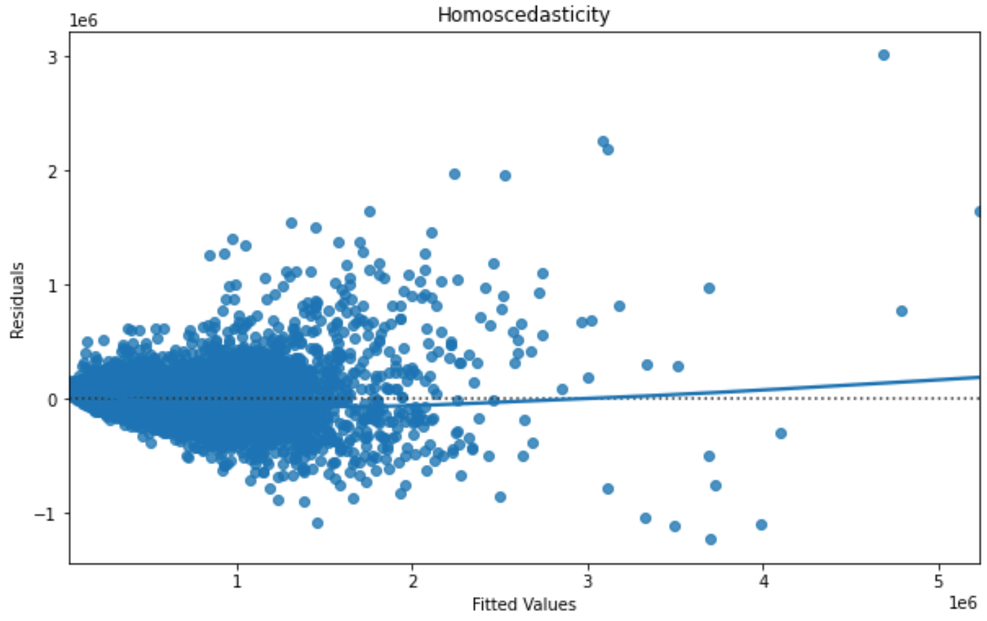
OBSERVATIONS

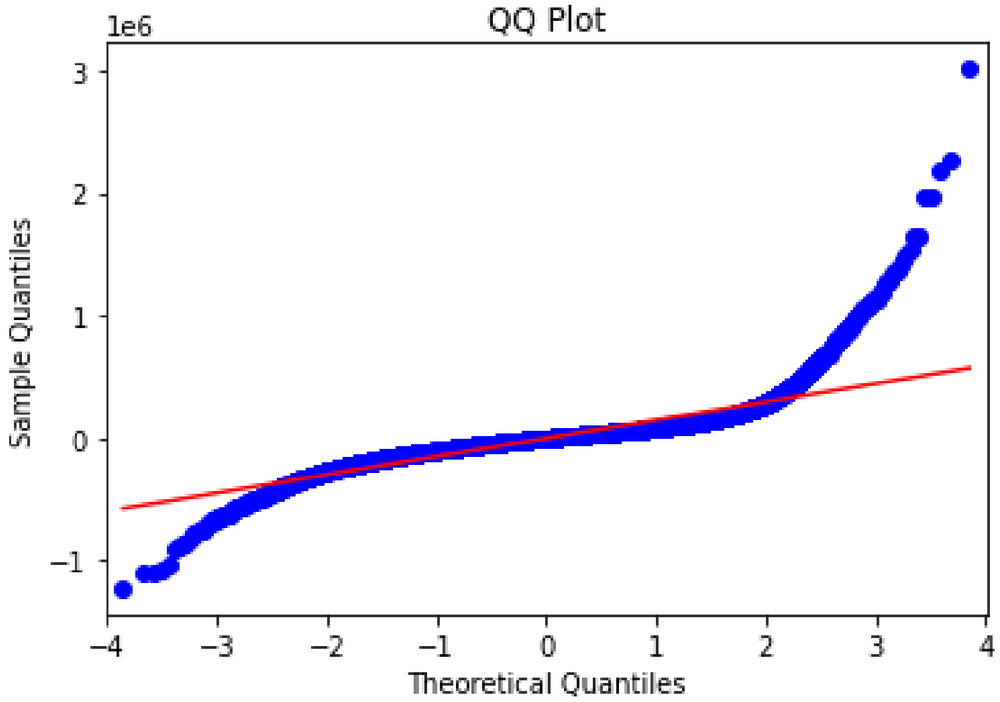
R-squared of 0.839: Our model can explain or predict about 83% of the variation in house prices. It's like being correct most of the time when guessing something.
Heteroscedasticity: The errors in our model's predictions are not evenly spread out. This unevenness suggests issues like outliers or skewed data.
Non-normal Residuals: The errors don't follow a normal, bell-shaped pattern, especially at the extremes, indicating that our model's mistakes are not distributed normally.

In conclusion;
While our model demonstrates strong explanatory power with a high R-squared value, the presence of heteroscedasticity and non-normal residuals indicates areas where further refinement or consideration of data preprocessing techniques may be necessary to improve the robustness and reliability of our predictions across all scenarios.

Third Iteration

In this iteration, we identify and eliminate predictors that exhibit high multicollinearity.





Observation

1. The R-quared of 0.838 indicates that approximately 83% of the variance in the dependent variable (price) can be explained by the model's independent variables.

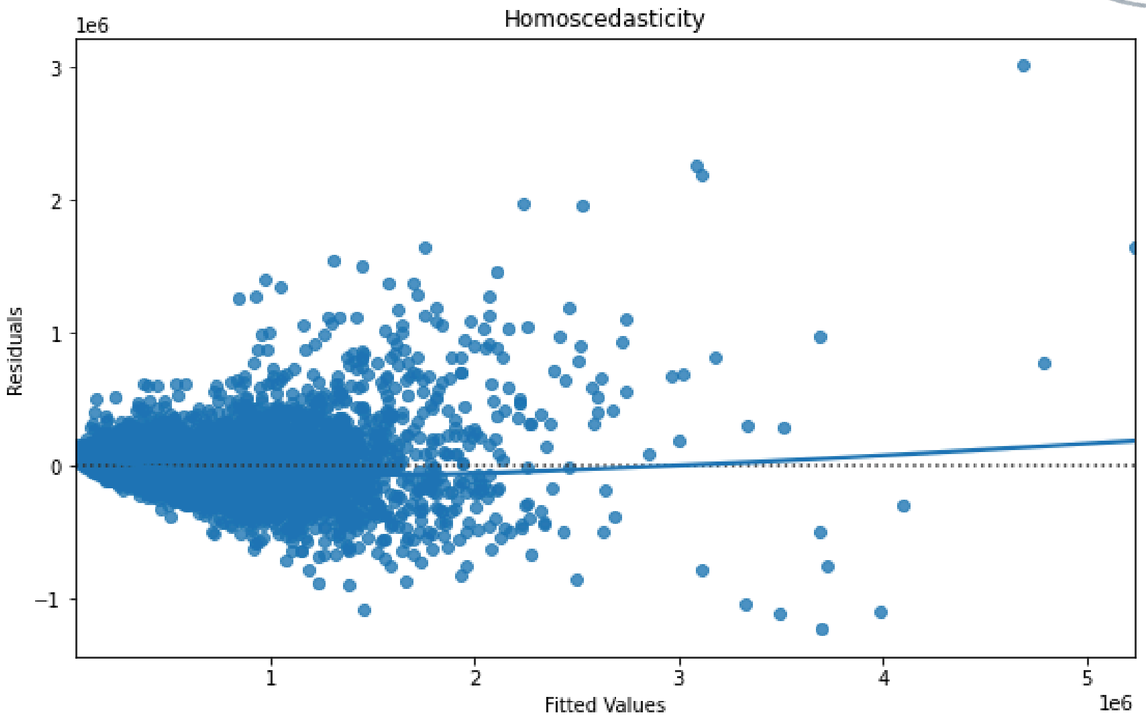
2. The model performed better after dropping sqft_living

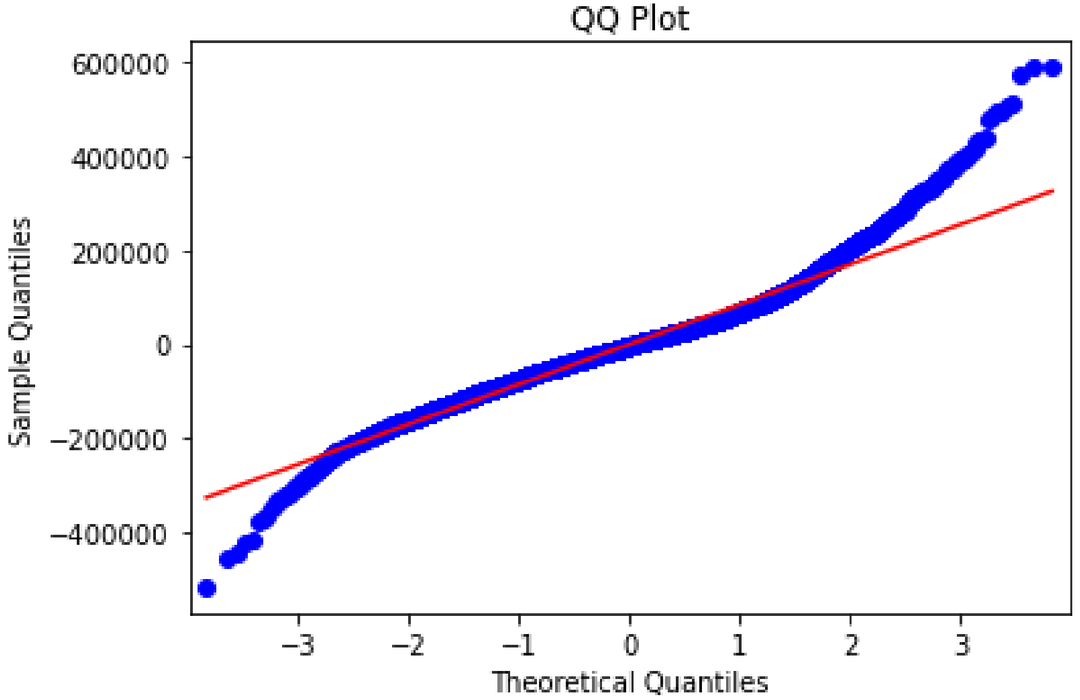
In conclusion;

- Improved performance
- Model Fit
- Removing sqft_living improve the performance

Fourth Iteration

In this iteration, we attempted to eliminate outliers from our dataset to assess their effect on our model's performance





Observation

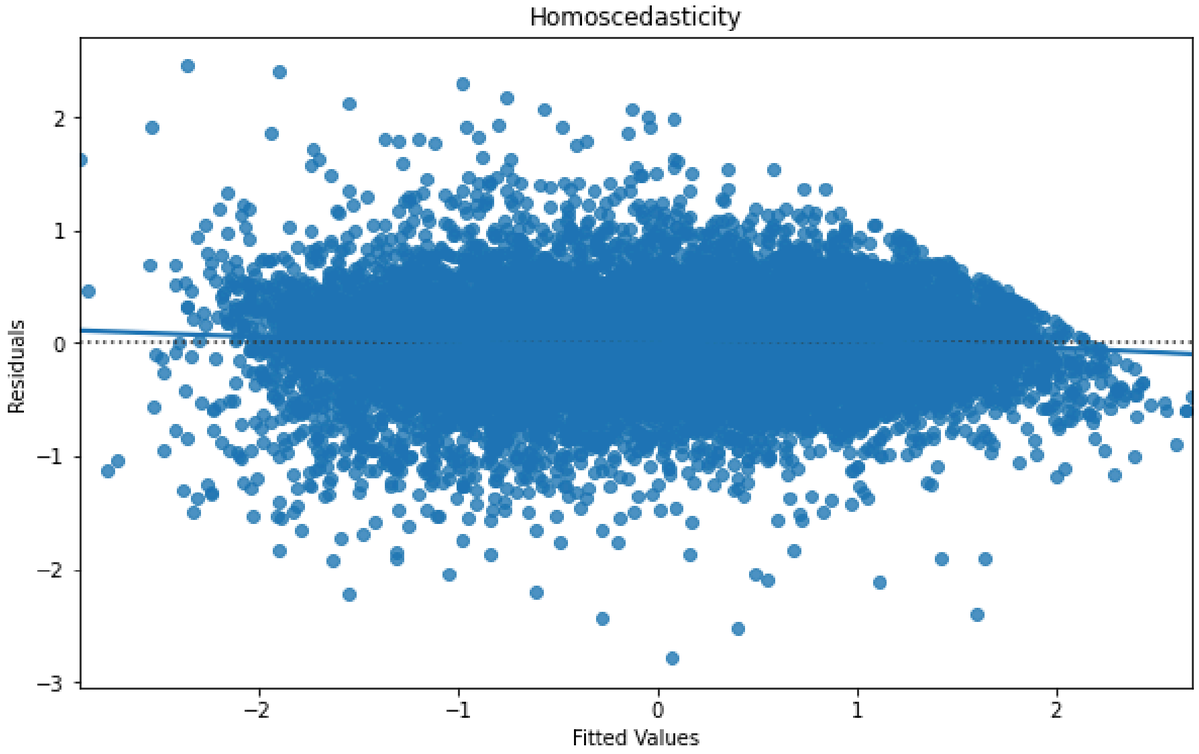
1. The R-quared of 0.838 indicates that approximately 83% of the variance in the dependent variable (price) can be explained by the model's independent variables.
2. The model meets the test for homoscedasticity

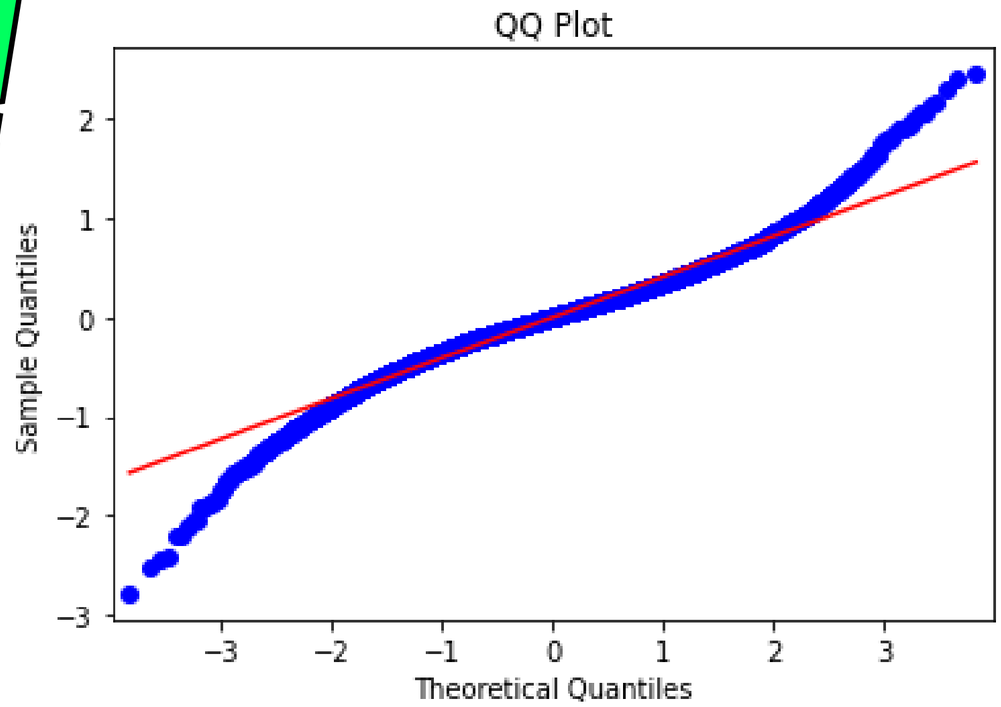
<u>Interpretation</u>:

- Stable variance in residuals

Final Model

In this iteration, we apply normalization and log-transformations to the data. These steps will reduce the impact of outliers, making the dataset more robust and enhancing the linear relationship between the target variable (price) and the features.





Observation

1. The R-quared of 0.838 indicates that approximately 83% of the variance in the dependent variable (price) can be explained by the model's independent variables.
2. The model meets the test for homoscedasticity

Interpretation:

- Model is ready for implementation

Model Validation

Methods Used

1. Residual Analysis
2. Root Mean Square Error
3. R-Squared and Adjusted - Squared
4. Train-Test Split

Conclusion

After several iterations of refining the model, the final model showed significant improvements:

Final Model Performance - The R-Squared value improved to 0.83, indicating that 83% of the variance in house prices can be explained by the model's features. The RMSE of the final model (0.7995) was significantly lower than the initial baseline model (256860.6115), indicating better predictive performance

Recommendations

1. Increase Living Area (sqft_living)- This feature has the highest positive impact on house prices. Each additional square foot significantly increases the home's value by approximately 0.33 x mean value. The agency should consider adding extensions or converting unused spaces into livable areas to increase the total living area of their homes.

2. Improve Condition - The overall condition of the house has a significant impact on its value. Houses in better condition contribute considerably to the price. The agency should ensure regular maintenance and upgrades to improve the house's condition are essential.

3. Upgrade Grade - The grade of the house shows substantial contributions to the house price. Higher grades are associated with higher values.

4. Focus on View Quality - The presence of a good view significantly boosts house prices. We recommend enhancing or creating better views.

5. Waterfront Properties - The presence of a waterfront view (waterfront_1.0) significantly increases house prices.


Thank You


Q & A