
Real Estate Market in King County



Data-driven insights and predictive modelling

Business Understanding

- Goal - create a platform that delivers accurate estimates of house prices
- Develop a model that can identify the key factors influencing house prices.
- Requirements - precise and representative data related to the real estate market in king county,

Problem Statement

Advertising Campaign to encourage sales in King County

- Data-driven recommendations
- Model to predict house prices

Analysis

Qn 1: What features are most closely connected with changes in house prices?

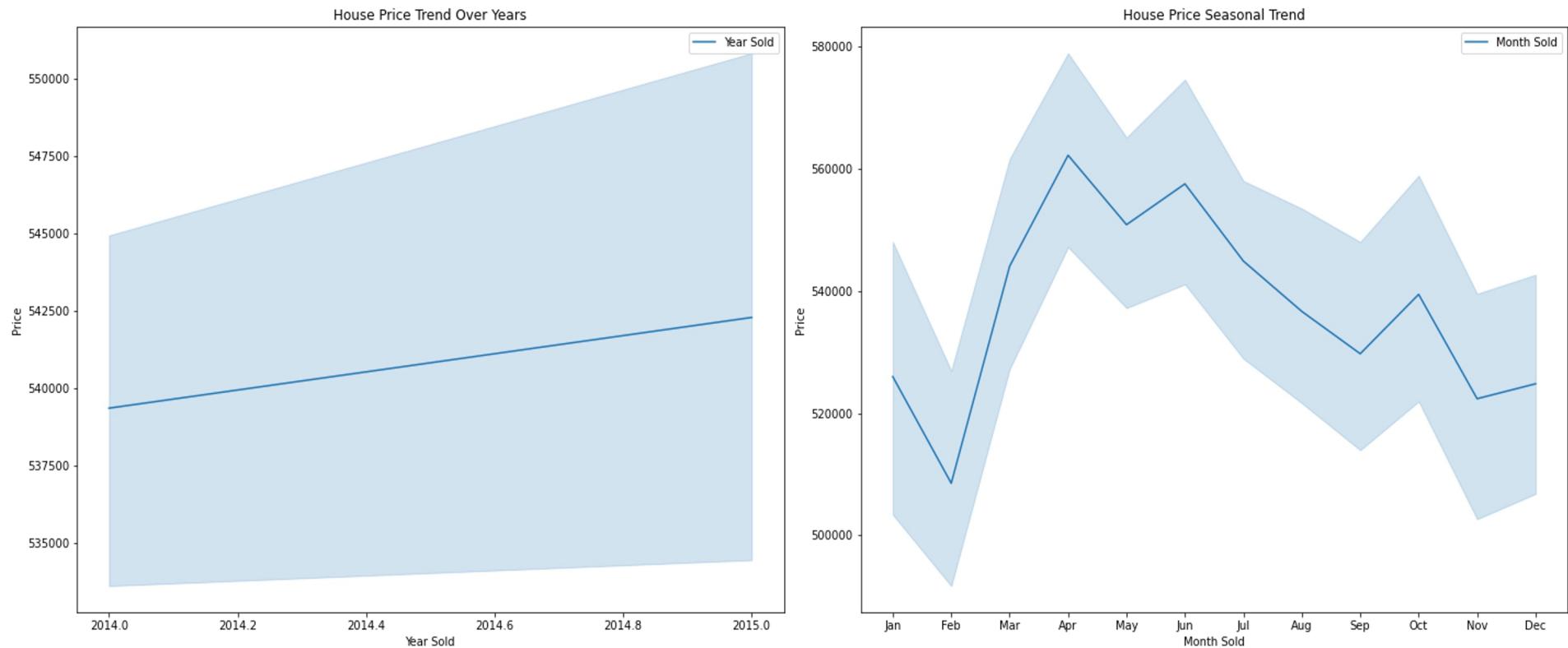
Qn 2: Which features show the strongest connections with other factors that help predict house prices?

Qn 3: What combination of features gives the most accurate predictions for how much a house will cost?

Methodology

- Gather relevant data and perform comprehensive analysis to understand its structure, quality, and potential insights
- Delve into the features of the dataset to uncover meaningful patterns and insights that can inform the prediction model.
- Construct a prediction model using appropriate methodologies, incorporating the insights gained from data analysis and feature investigation.

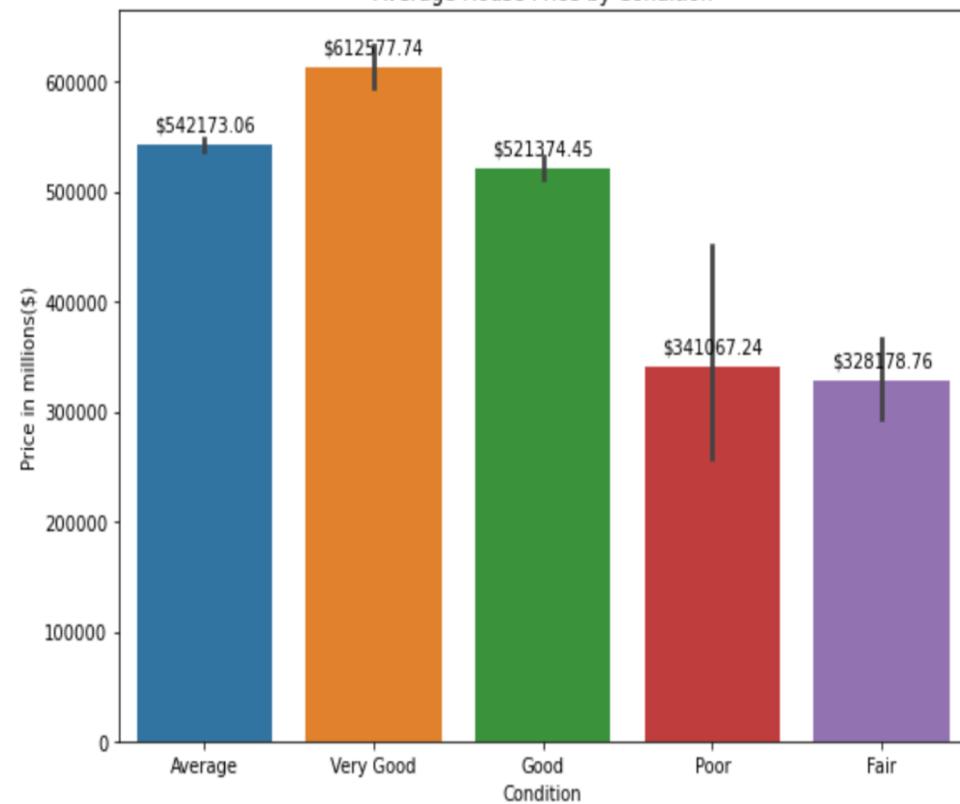
House price trends



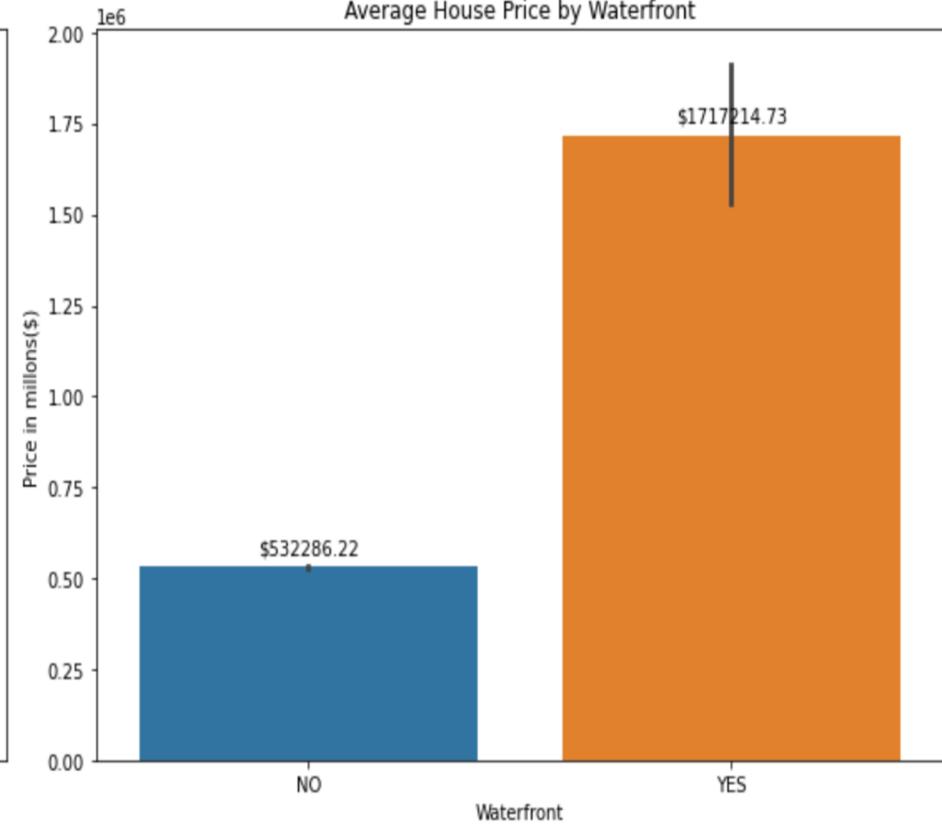
In the graphs below;

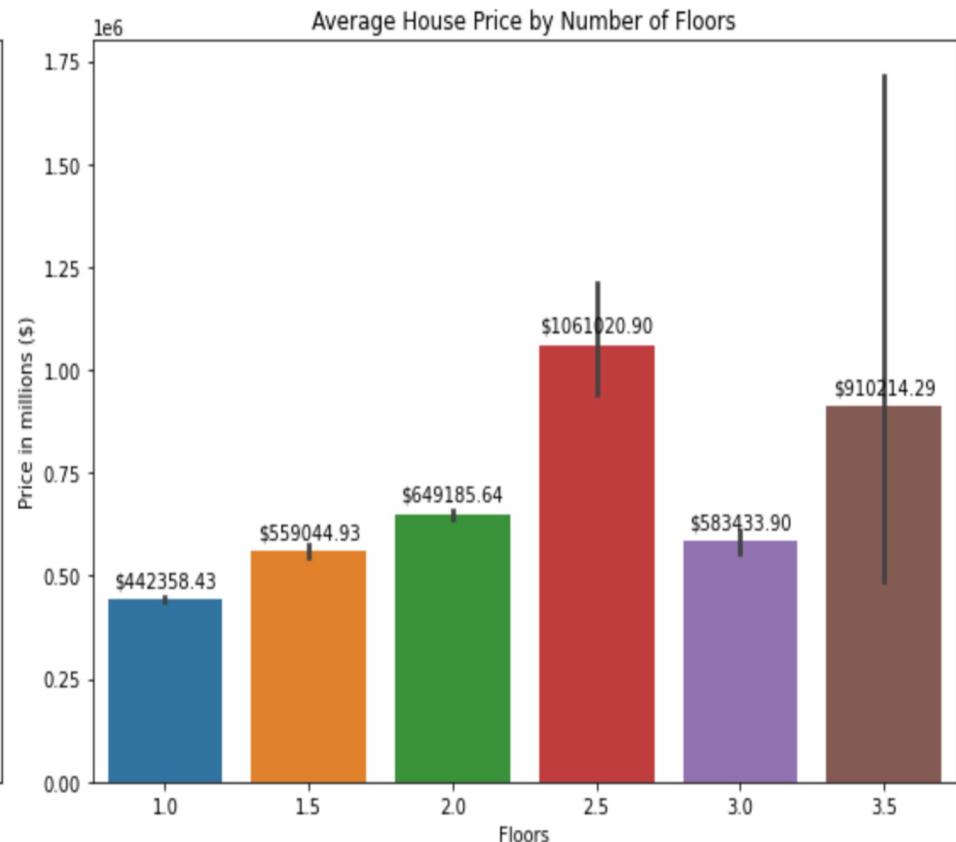
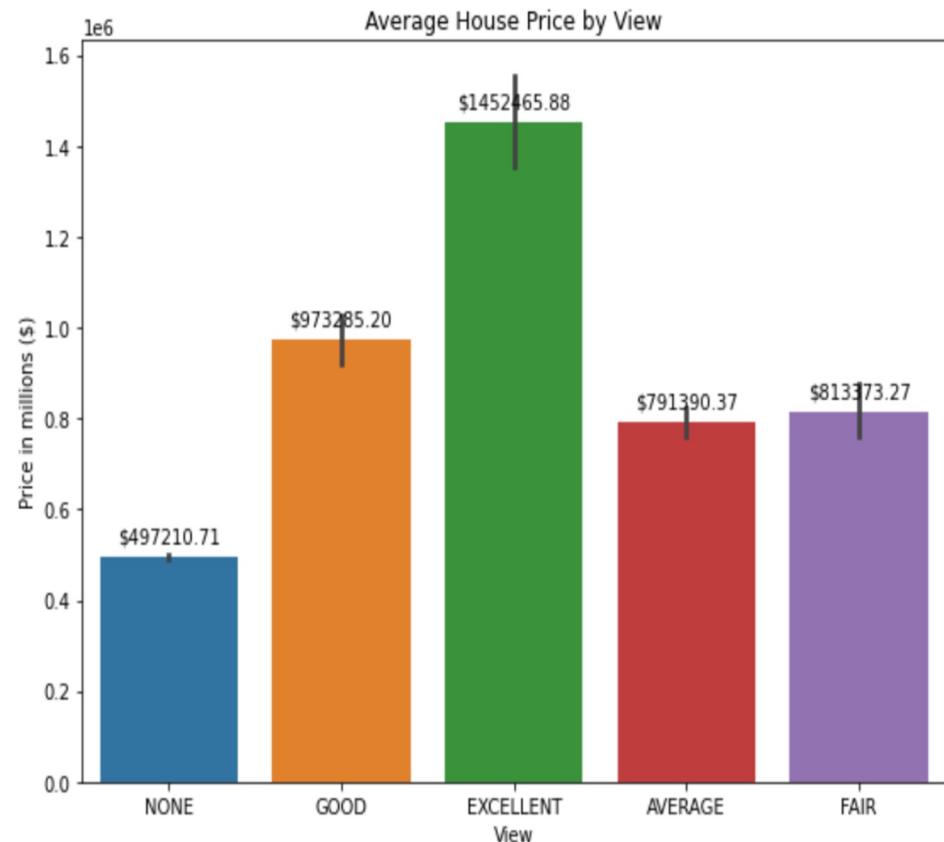
We analyzed the variation of average house prices as per the overall condition of a house, whether a house is located on a waterfront or not, the quality of views from the house and the number of floors/levels in a house.

Average House Price by Condition

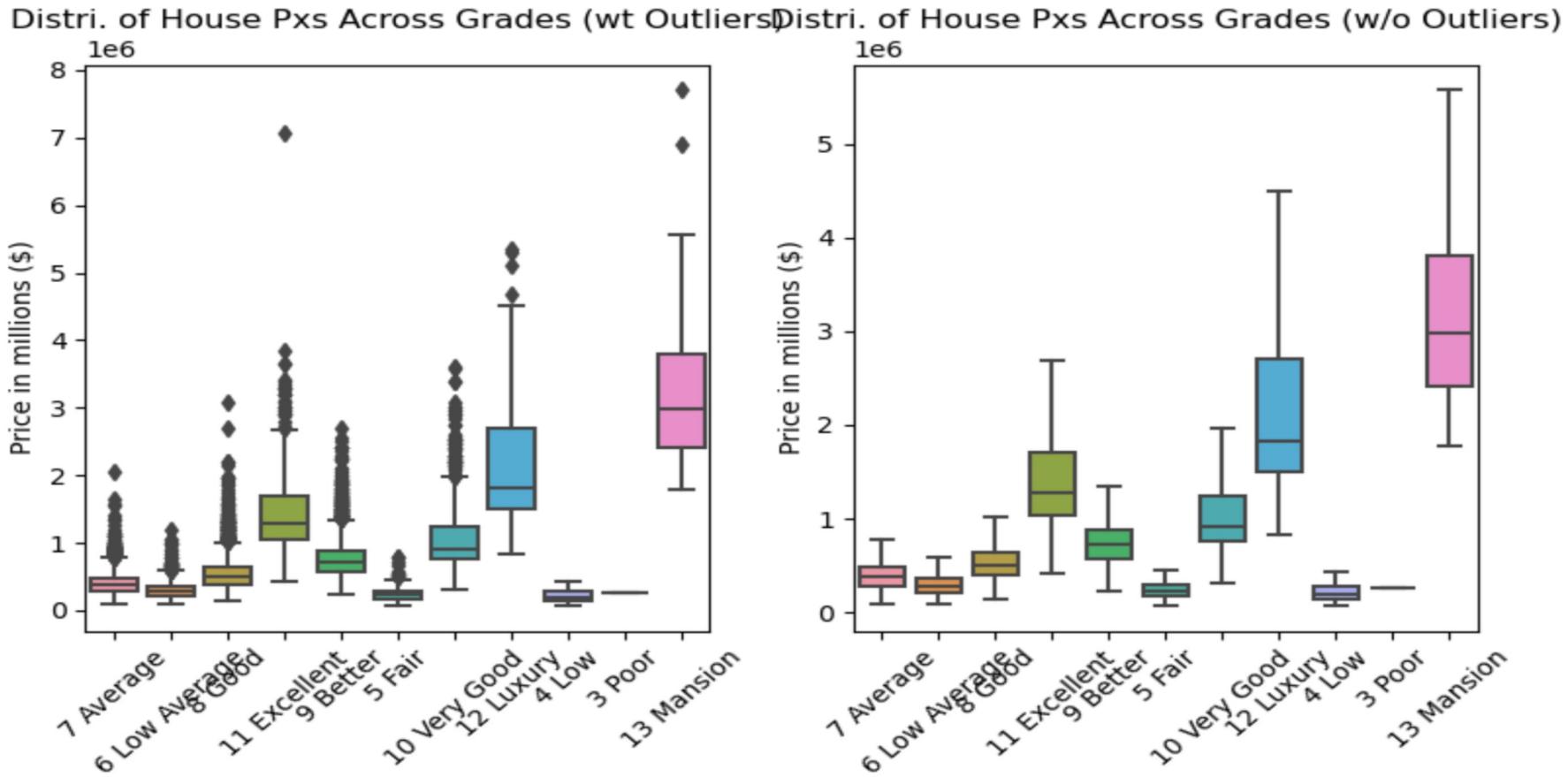


Average House Price by Waterfront

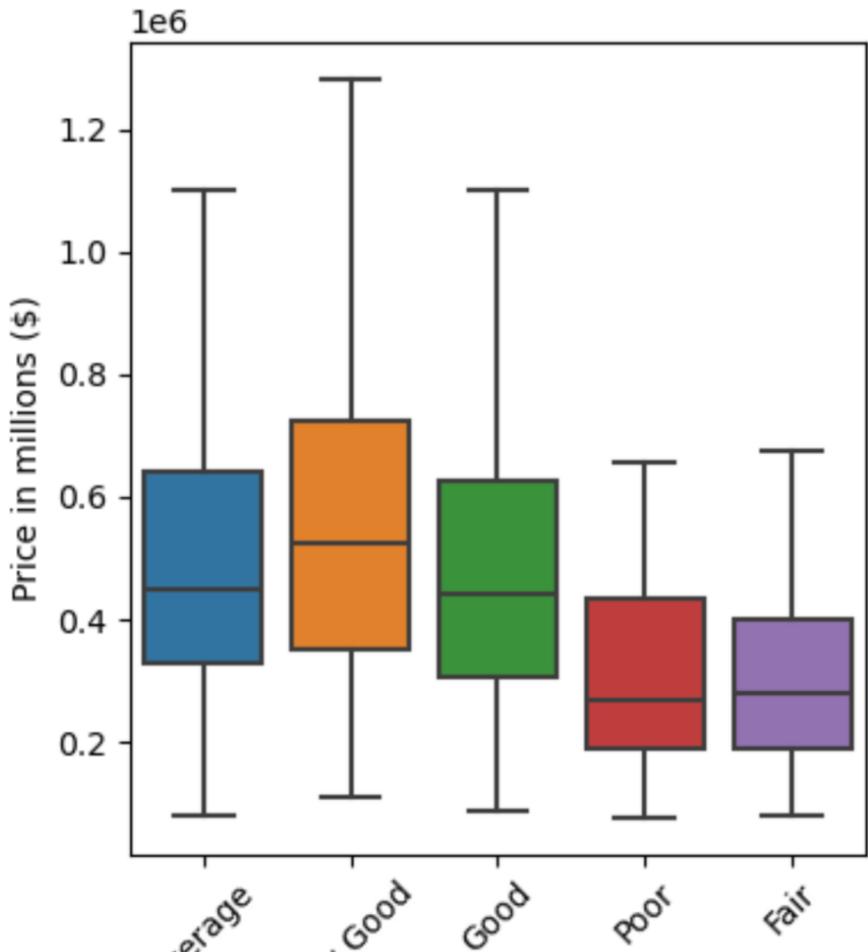
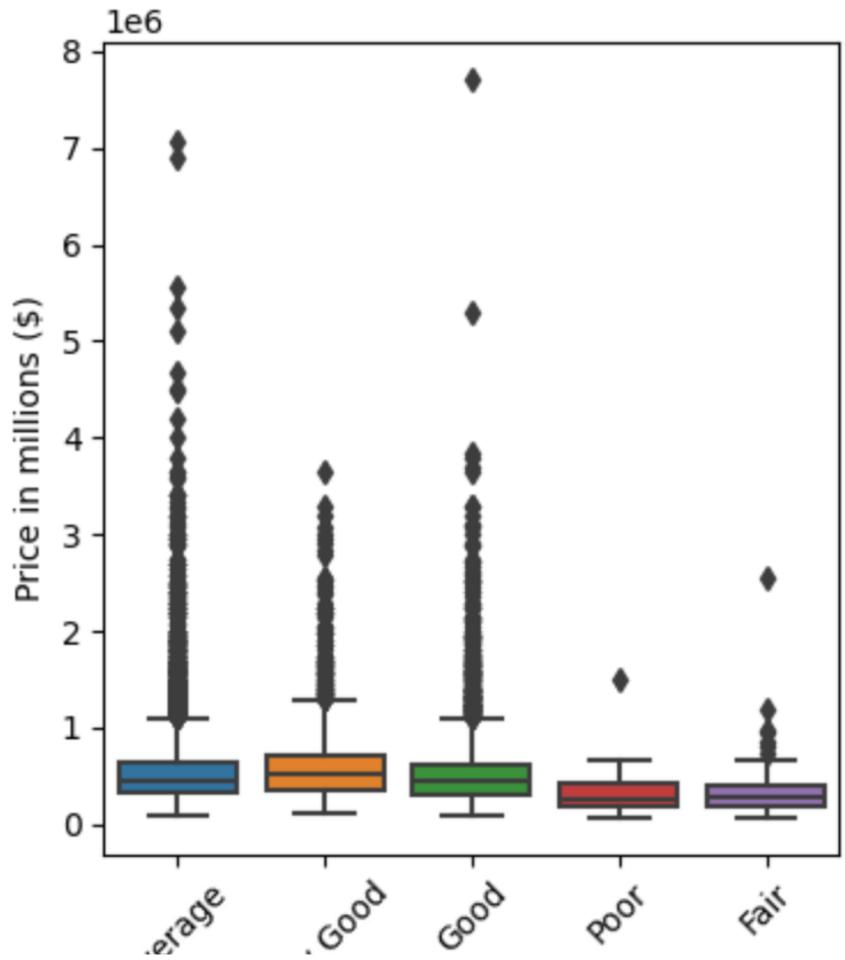


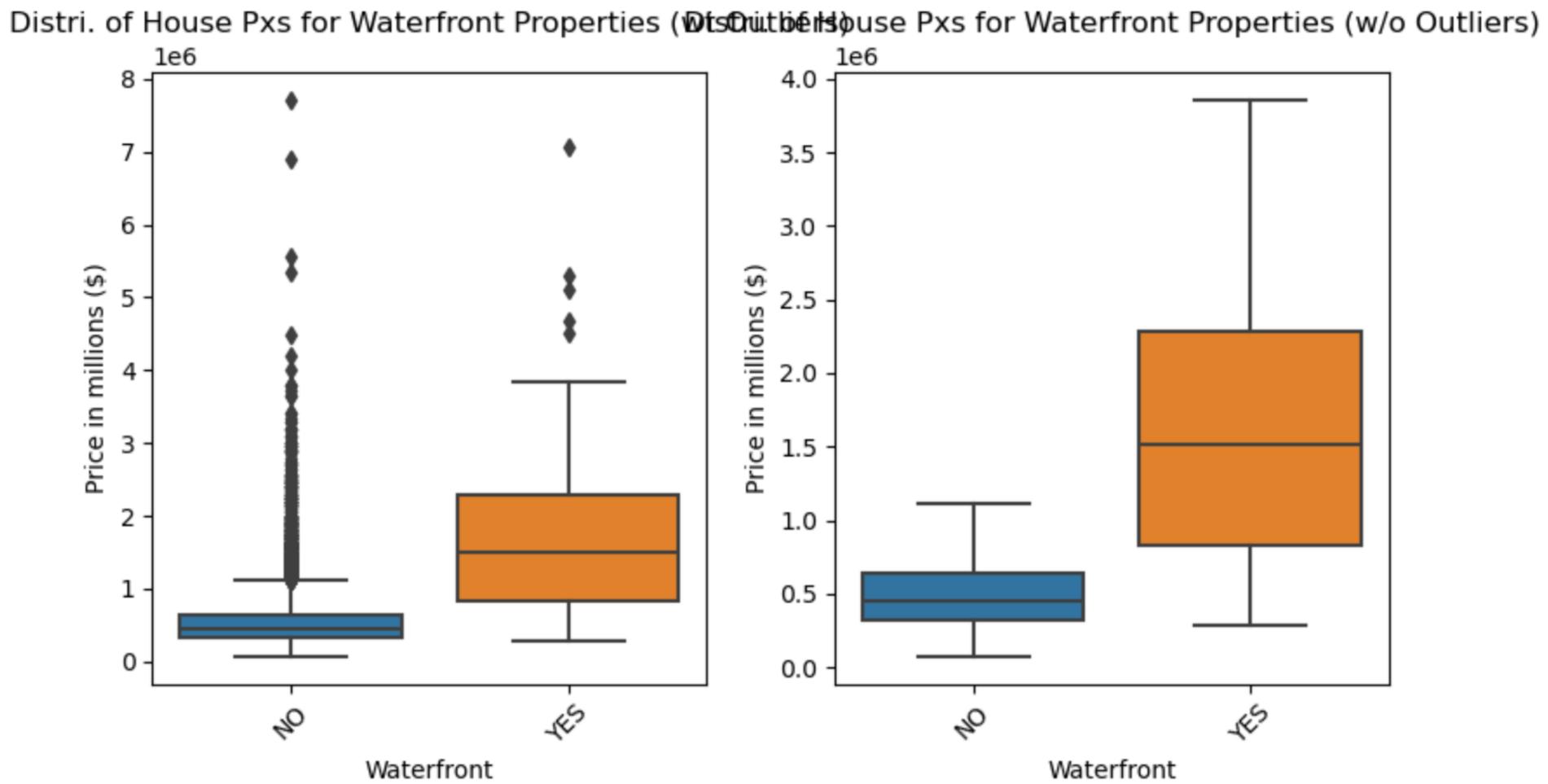


Assessing the change in house price values as per the grade, condition and waterfront location of a house with and without outliers using box plots as shown below.

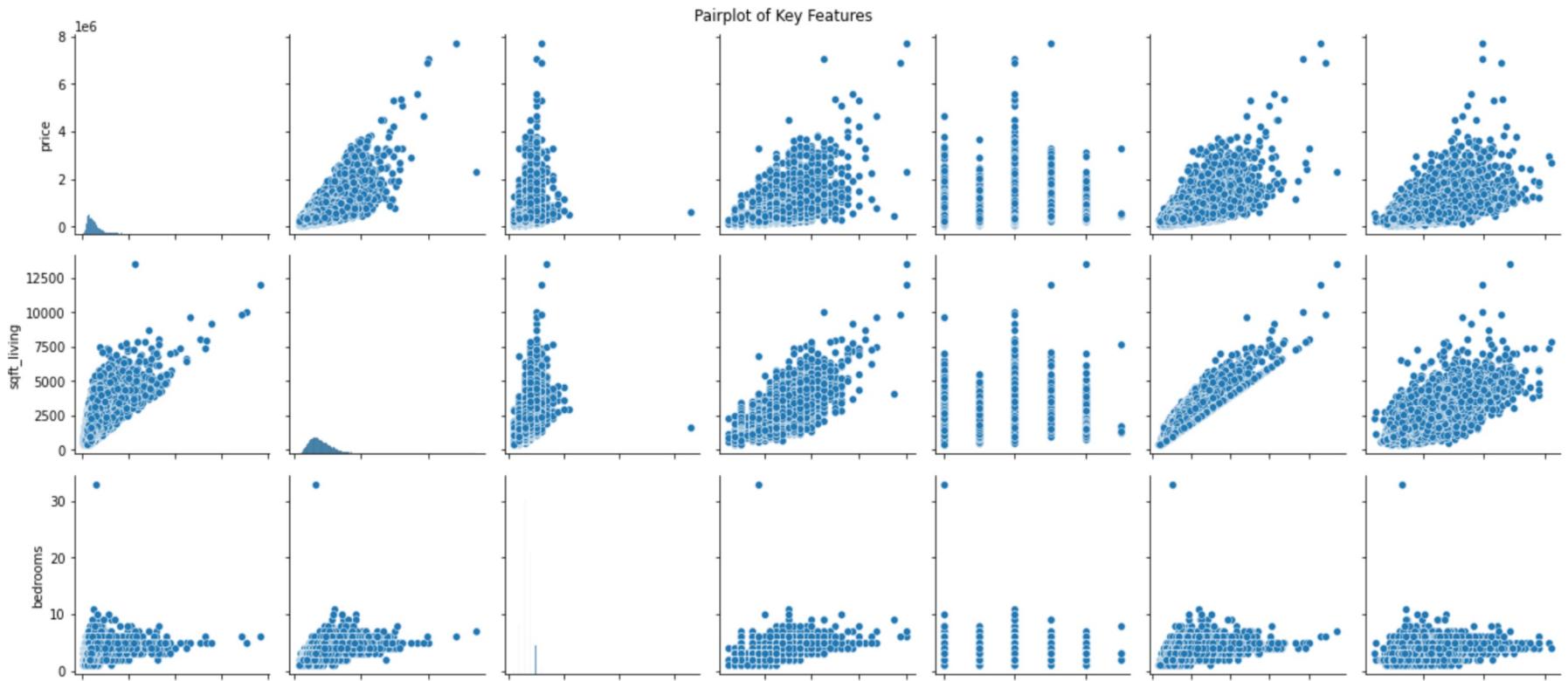


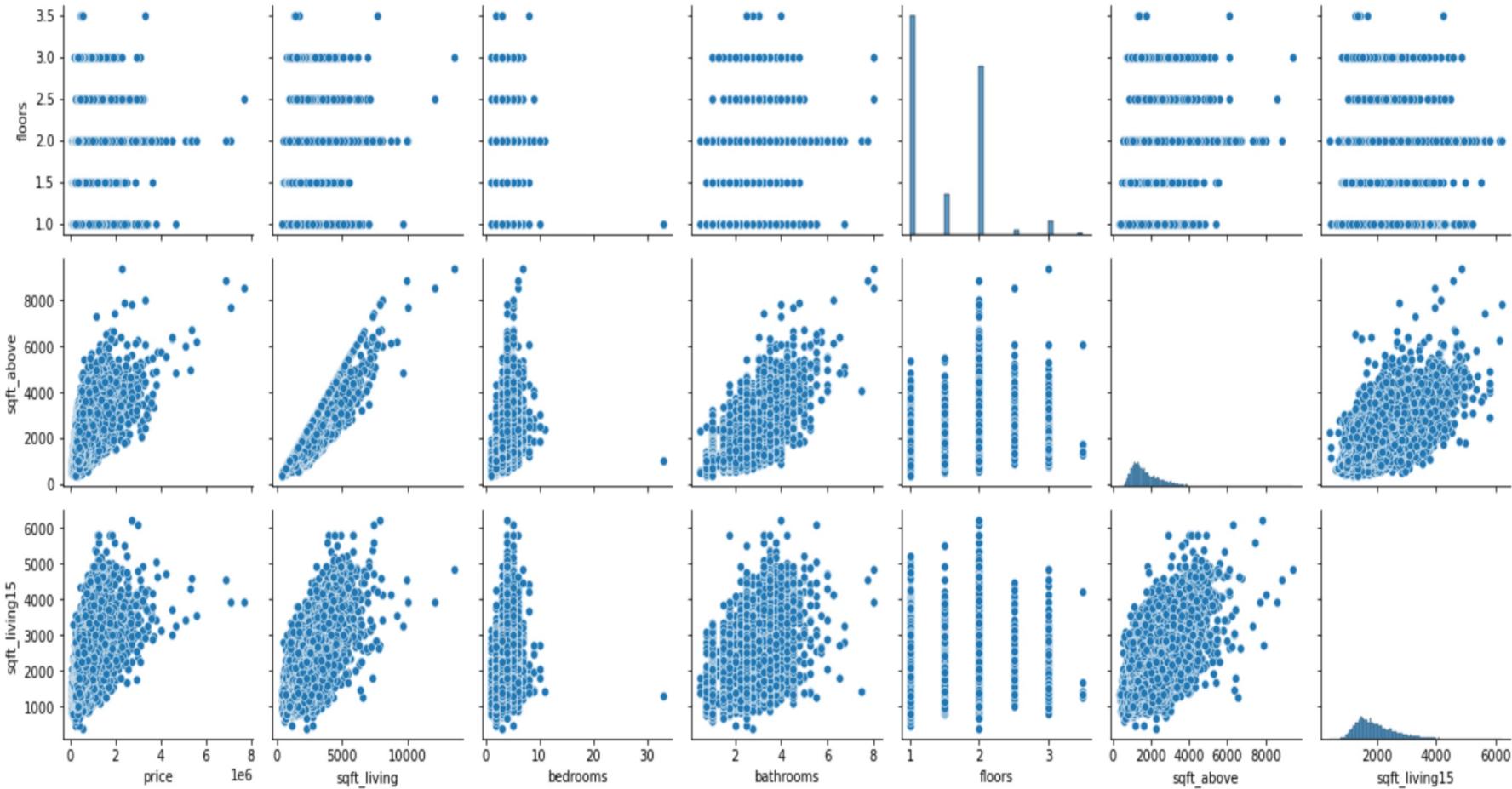
Distri. of House Prices Across Conditions (wt Outliers) of House Prices Across Conditions (w/o Outliers)





Scatterplots to visualise and explore the relationships between multiple variables





General overview on the variables that have a strong correlation with the price

OLS Regression Results

```
=====
Dep. Variable:          price    R-squared:         0.493
Model:                 OLS     Adj. R-squared:      0.493
Method:                Least Squares  F-statistic:      7003.
Date:                 Mon, 29 Apr 2024  Prob (F-statistic): 0.00
Time:                  15:21:28    Log-Likelihood:   -3.0005e+05
No. Observations:      21597    AIC:                  6.001e+05
Df Residuals:          21593    BIC:                  6.001e+05
Df Model:                  3
Covariance Type:        nonrobust
=====
```

	coef	std err	t	P> t	[0.025	0.975]
Intercept	-3.815e+04	5252.893	-7.262	0.000	-4.84e+04	-2.79e+04
sqft_living	297.9354	4.483	66.458	0.000	289.148	306.723
sqft_above	-18.4232	4.479	-4.113	0.000	-27.202	-9.644
bathrooms	-3972.1769	3545.059	-1.120	0.263	-1.09e+04	2976.400

```
=====
Omnibus:                 14748.993  Durbin-Watson:        1.982
Prob(Omnibus):            0.000    Jarque-Bera (JB):  538245.600
Skew:                      2.806    Prob(JB):            0.00
Kurtosis:                  26.804   Cond. No.          9.59e+03
=====
```

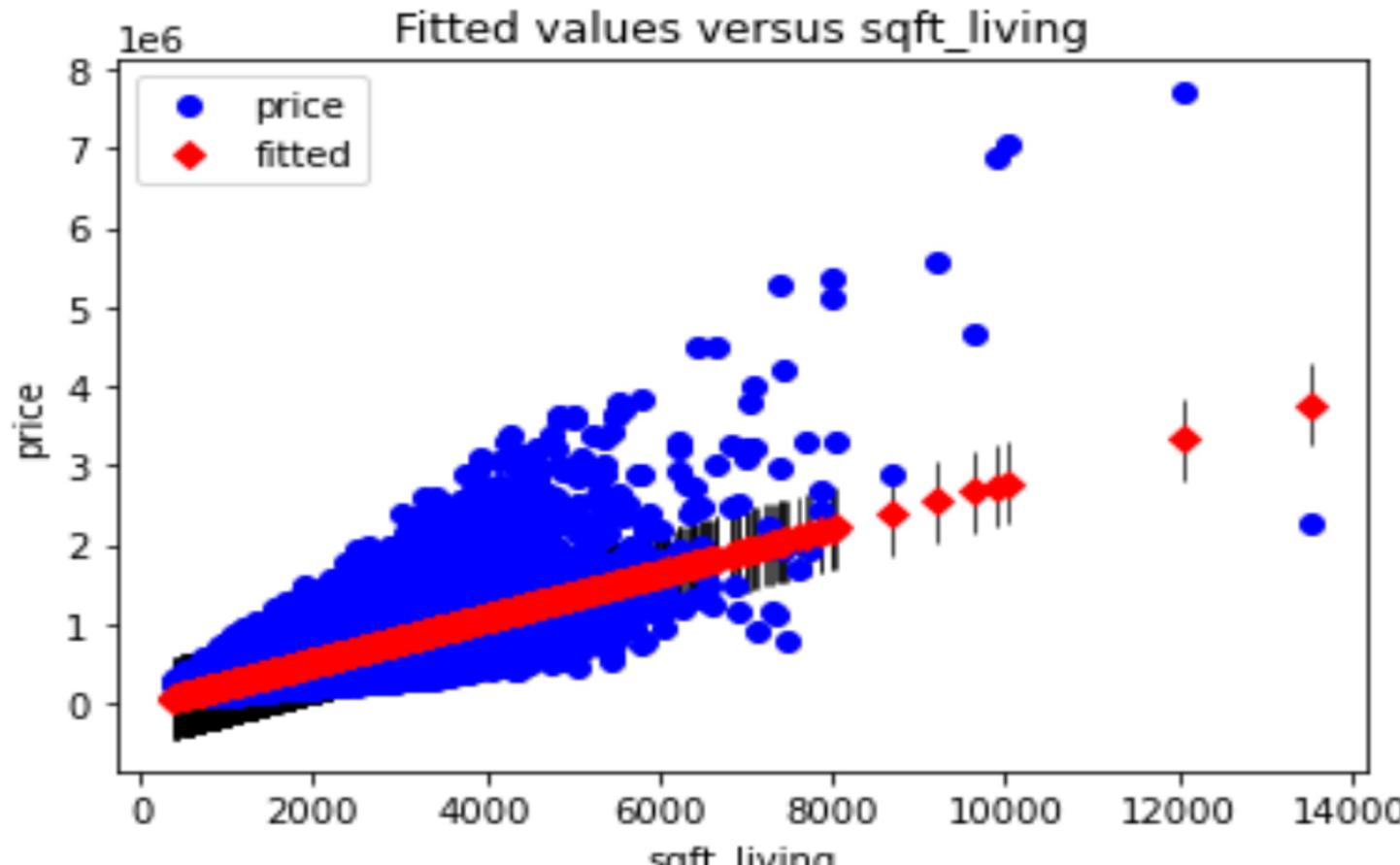
Notes:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The condition number is large, 9.59e+03. This might indicate that there are strong multicollinearity or other numerical problems.

Simple linear regression:

Linear regression model with price as the dependent variable and sqft_living as the independent variable to determine the coefficient and y-intercept

Plotting actual and predicted values



OLS Regression Results

=====

Dep. Variable: price R-squared: 0.493
Model: OLS Adj. R-squared: 0.493
Method: Least Squares F-statistic: 2.097e+04
Date: Mon, 29 Apr 2024 Prob (F-statistic): 0.00
Time: 15:04:43 Log-Likelihood: -3.0006e+05
No. Observations: 21597 AIC: 6.001e+05
Df Residuals: 21595 BIC: 6.001e+05
Df Model: 1
Covariance Type: nonrobust

=====

	coef	std err	t	P> t	[0.025	0.975]
const	-4.399e+04	4410.023	-9.975	0.000	-5.26e+04	-3.53e+04
sqft_living	280.8630	1.939	144.819	0.000	277.062	284.664

=====

Omnibus: 14801.942 Durbin-Watson: 1.982
Prob(Omnibus): 0.000 Jarque-Bera (JB): 542662.604
Skew: 2.820 Prob(JB): 0.00
Kurtosis: 26.901 Cond. No. 5.63e+03

=====

Notes:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The condition number is large, 5.63e+03. This might indicate that there are strong multicollinearity or other numerical problems.

Multiple linear regression:

Incorporating all other features we were interested in to understand their impact on price based on the baseline model.

Results

OLS Regression Results

Dep. Variable:	price	R-squared:	0.592			
Model:	OLS	Adj. R-squared:	0.592			
Method:	Least Squares	F-statistic:	2412.			
Date:	Fri, 03 May 2024	Prob (F-statistic):	0.00			
Time:	08:12:23	Log-Likelihood:	-2.9770e+05			
No. Observations:	21597	AIC:	5.954e+05			
Df Residuals:	21583	BIC:	5.955e+05			
Df Model:	13					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	1.436e+05	6678.106	21.501	0.000	1.3e+05	1.57e+05
sqft_living	210.5141	3.989	52.780	0.000	202.696	218.332
sqft_living15	25.5470	3.841	6.652	0.000	18.019	33.075
sqft_above	-99.0892	4.255	-23.286	0.000	-107.430	-90.749
grade_10 Very Good	4.185e+05	9518.512	43.965	0.000	4e+05	4.37e+05
grade_11 Excellent	7.205e+05	1.47e+04	48.897	0.000	6.92e+05	7.49e+05
grade_12 Luxury	1.265e+06	2.75e+04	46.068	0.000	1.21e+06	1.32e+06
grade_13 Mansion	2.49e+06	6.72e+04	37.073	0.000	2.36e+06	2.62e+06
grade_3 Poor	2.624e+04	2.35e+05	0.112	0.911	-4.34e+05	4.86e+05
grade_4 Low	-4.615e+04	4.53e+04	-1.018	0.308	-1.35e+05	4.27e+04
grade_5 Fair	-4.374e+04	1.54e+04	-2.839	0.005	-7.39e+04	-1.35e+04
grade_6 Low Average	-2.211e+04	5931.389	-3.728	0.000	-3.37e+04	-1.05e+04
grade_8 Good	7.169e+04	4235.107	16.928	0.000	6.34e+04	8e+04
grade_9 Better	2.127e+05	6586.660	32.295	0.000	2e+05	2.26e+05
Omnibus:	14135.657	Durbin-Watson:	1.985			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	538054.998			
Skew:	2.615	Prob(JB):	0.00			
Kurtosis:	26.887	Cond. No.	5.34e+05			

Train – test results

- After Training our data using the (80/20, 70/30, 60/40) ratios, we noticed that the model with the 80/20 had the highest r squared. This conclusion can be inferred since they were all scaled uniformly and they used the same predictor variables all through.
- `Random Forest Regression` of the 80/20 split is the most accurate with `59.31`% of house price variance explained with a Mean Absolute Error of `155053.45818852124` dollars.

Key takeaways

- **Positive Influencers on Price:**

The presence of additional bathrooms increased square footage, higher floors, waterfront access, captivating views, and elevated grade scores positively impact house prices.

Notably, the inclusion of cities like Medina and Mercer Island in the analysis reveals their positive association with higher property values.

- **Negative Influencers on Price:**

The number of bedrooms, certain city affiliations (e.g., Auburn, Federal Way, Kent), in reference to Bellevue, and specific features (e.g., Fall Season, City) exhibit a negative correlation with house prices.

Premiere Property Group should be cognizant of these factors when devising pricing strategies.

- **Seasonal and Unique Factors:**

While some seasonal variables do not significantly impact prices, it's crucial to note that the age of the house and the presence of a basement can influence pricing dynamics.

- **City-Specific Considerations:**

Each city has a unique influence on house prices, emphasizing the need for tailored strategies for different locations

Recommendations



Home Features

- Larger living area (sqft)
- Grade above 10
- Grade better predictor



Location

- Importance of waterfront
- Bellevue and Mercer Island areas



Time of Year

- Constant median house price
- April/May highest sales volume
- Campaign in March/April

Recommendations

- **Market Positioning Strategies:** Sellers should leverage insights into key price-driving factors to strategically position their properties in the market. Emphasize features eg. spacious living areas, recent renovations, and waterfront access to increase property attractiveness and market competitiveness.
- **Investment Considerations:** Investors should carefully evaluate properties based on their potential for value appreciation. Prioritizing investments in neighborhoods with proximity to amenities and waterfront locations could yield higher returns.
- **Buyer Awareness:** Should be cognizant of the factors influencing house prices when making purchasing decisions.

Insights

- **Key Factors Affecting House Prices:** Square footage of living space emerges as the primary driver of house prices. Features like grade, waterfront status and recent renovations correlate with prices, underscoring their impact on property values.
- **Significance of Property Characteristics:** The number of bedrooms and bathrooms also contribute to determining house prices, albeit with varying degrees of influence.
- **Neighborhood and Environmental Factors:** Proximity to amenities positively correlates with house prices implying that location remains a critical factor in property valuation.
- **Modeling and Predictive Insights:** Employing regression models such as Simple Linear Regression and Multiple Linear Regression coupled with iterative modeling and feature engineering techniques, enhances the predictive accuracy of models.

Conclusions

- The dataset's geographic and temporal constraints may restrict the generalizability of findings to other regions or time periods.
- The analysis equips stakeholders, including homeowners, real estate professionals, and potential buyers with actionable insights to navigate the housing market effectively.
- Leveraging knowledge of key price determinants and understanding the nuances of property valuation, stakeholders can make informed decisions regarding pricing strategies, property investments and market participation.
- Continued research and analysis are warranted to refine models and explore additional factors for enhanced predictive accuracy and comprehensive market understanding.

OPPORTUNITIES FOR FURTHER ANALYSIS

- Micro-location Trends
- Analyzing long-term Market Trends
- Economic and Demographic Factors
- Advanced Predictive Modeling
- Impact of Renovations
- Customer Segmentation and Targeting

THANK YOU!