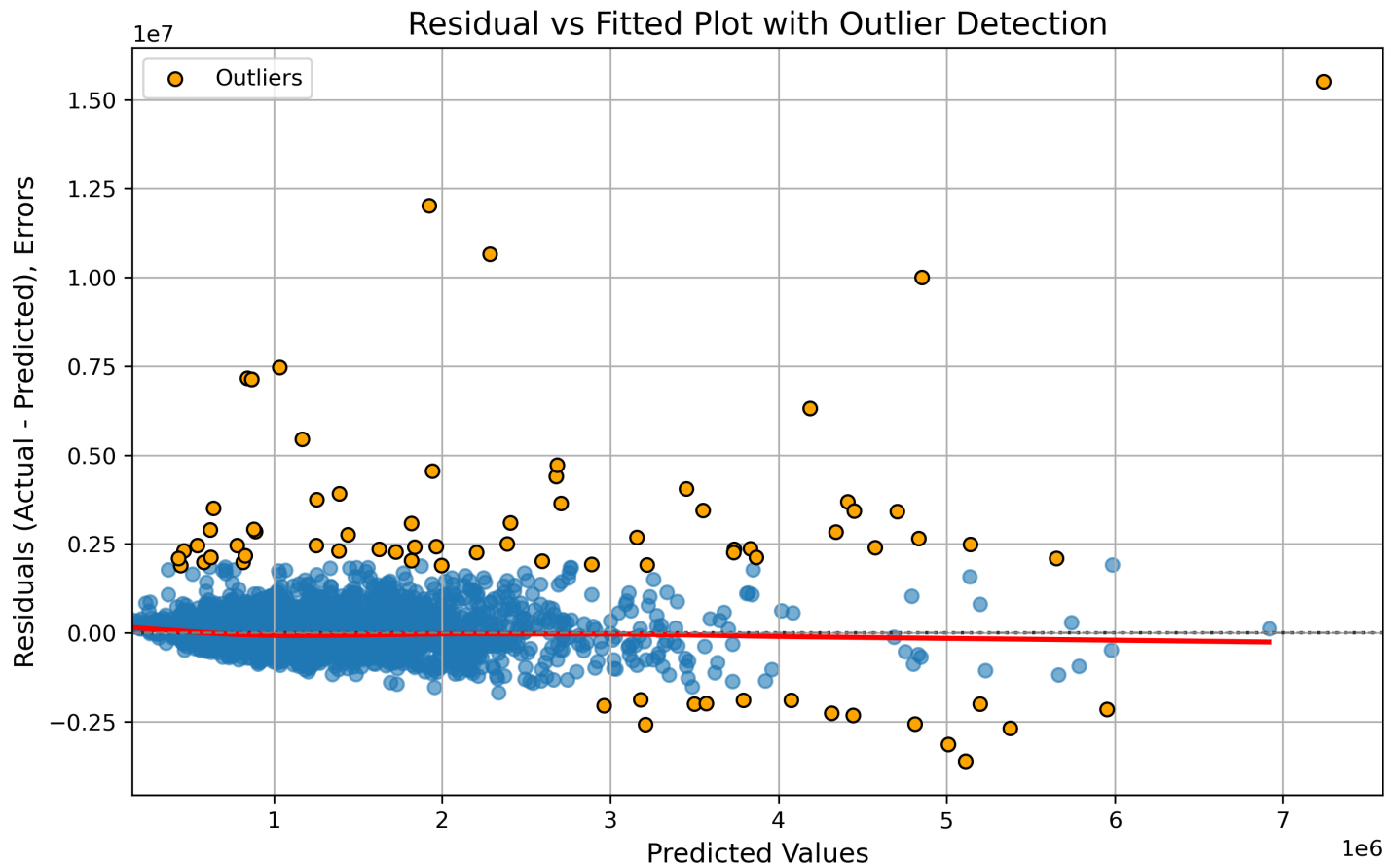


This model and explanation were generated by ModelBot, an agent designed to help non-technical users perform basic machine learning modeling, powered by Llama 3. It is not a replacement for a human data scientist, and there may be discrepancies and inaccuracies within this report.

Linear Regression Results						
Dependent Variable:	price	R-squared:	0.519			
Model:	linear	Adjusted R-squared:	0.519			
No. Observations:	23632	F-statistic:	821.784			
Df Residuals:	23607	Prob (F-statistic):	1.11e-16			
Df Model:	25	AIC:	695,311.908			
RMSE:	591,812.072	BIC:	695,570.159			
	coef	std err	t	P> t	0.025	0.975
grade_12 Luxury	2e+06	7.1e+04	28.77	0	1.9e+06	2.2e+06
Intercept	1.6e+06	2.9e+04	55.88	0	1.6e+06	1.7e+06
grade_11 Excellent	8.3e+05	3.8e+04	21.56	0	7.5e+05	9e+05
waterfront_1	6.7e+05	3.7e+04	18.20	0	6e+05	7.5e+05
view_EXCELLENT	5.9e+05	3.7e+04	15.82	0	5.1e+05	6.6e+05
sewer_system_PUBLIC	2.1e+05	1.3e+04	16.95	0	1.9e+05	2.4e+05
sqft_above	1.5e+05	1.6e+04	9.15	0	1.2e+05	1.8e+05
bathrooms	8.9e+04	7e+03	12.75	0	7.5e+04	1e+05
sqft_patio	4.3e+04	4.4e+03	9.69	0	3.4e+04	5.1e+04
sqft_garage	-4.7e+04	5.5e+03	-8.62	0	-5.8e+04	-3.6e+04
bedrooms	-5.4e+04	5.5e+03	-9.90	0	-6.5e+04	-4.3e+04
yr_built	-1e+05	6.2e+03	-16.46	0	-1.1e+05	-8.9e+04
grade_9 Better	-4.5e+05	2.2e+04	-20.31	0	-4.9e+05	-4e+05
grade_8 Good	-7.7e+05	2.2e+04	-34.95	0	-8.1e+05	-7.3e+05
grade_7 Average	-9.4e+05	2.4e+04	-38.91	0	-9.9e+05	-8.9e+05
grade_6 Low Average	-1e+06	2.9e+04	-35.63	0	-1.1e+06	-9.8e+05
grade_5 Fair	-1e+06	4.4e+04	-23.67	0	-1.1e+06	-9.5e+05
sqft_living	1.3e+05	1.8e+04	7.21	5.9e-13	9.2e+04	1.6e+05
condition_Very Good	8.4e+04	1.4e+04	6.13	8.8e-10	5.7e+04	1.1e+05
sqft_basement	4.9e+04	8.1e+03	6.05	1.5e-09	3.3e+04	6.5e+04
floors	-3.3e+04	5.7e+03	-5.74	9.6e-09	-4.4e+04	-2.2e+04
nuisance_YES	5.8e+04	1e+04	5.54	3e-08	3.7e+04	7.8e+04
sqft_lot	2.2e+04	4.2e+03	5.26	1.5e-07	1.4e+04	3e+04
view_FAIR	1.2e+05	4.7e+04	2.58	0.0098	2.9e+04	2.1e+05
condition_Good	2.5e+04	9.7e+03	2.54	0.011	5.7e+03	4.4e+04



## ***PREDICTION TARGET***

The model is trying to predict the price of a property based on various features such as grade, waterfront, view, and more. The target variable price represents the actual price of the property, which is an important metric for real estate professionals, investors, and homeowners to forecast accurately.

## ***METRIC EXPLANATIONS***

$R^2$  (coefficient of determination) measures the proportion of the variance in the target variable that is explained by the independent variables. A high  $R^2$  value (close to 1) indicates that the model is able to explain most of the variation in the target variable, while a low  $R^2$  value (close to 0) indicates that the model is not able to explain much of the variation.

RMSE (root mean squared error) measures the average magnitude of the errors in the model's predictions. A low RMSE value indicates that the model is making accurate predictions, while a high RMSE value indicates that the model is making large errors.

Overfitting occurs when a model is too complex and is able to fit the noise in the training data, resulting in poor performance on new, unseen data. Underfitting occurs when a model is too simple and is unable to capture the underlying relationships in the data, resulting in poor performance on both training and new data.

## **OVERFITTING OR UNDERFITTING?**

The training  $R^2$  is 0.52, and the testing  $R^2$  is 0.50. The training RMSE is  $5.9e+05$ , and the testing RMSE is  $6.3e+05$ . The model is not overfitting, as the testing  $R^2$  is still relatively high and the testing RMSE is only slightly higher than the training RMSE. However, the model may be slightly underfitting, as the testing  $R^2$  is lower than the training  $R^2$ .

## **ANOVA TABLE OVERVIEW**

The ANOVA table provides a summary of the model's performance. The F-statistic measures the ratio of the variance explained by the model to the variance not explained by the model. The p-value measures the probability that the observed F-statistic could occur by chance. A low p-value indicates that the model is significant and that the independent variables are able to explain a significant portion of the variation in the target variable.

## **SIGNIFICANT FEATURES ANALYSIS**

The top 5 most significant features are:

- grade\_12 Luxury (coefficient =  $2.042669e+06$ , p-value = 0.0): This feature is highly significant and indicates that properties with a luxury grade have a much higher price.
- Intercept (coefficient =  $1.642747e+06$ , p-value = 0.0): The intercept term is highly significant and indicates that the model is able to capture a significant portion of the variation in the target variable.
- grade\_11 Excellent (coefficient =  $8.293505e+05$ , p-value = 0.0): This feature is highly significant and indicates that properties with an excellent grade have a higher price.
- waterfront\_1 (coefficient =  $6.738919e+05$ , p-value = 0.0): This feature is highly significant and indicates that properties with a waterfront location have a higher price.
- view\_EXCELLENT (coefficient =  $5.859039e+05$ , p-value = 0.0): This feature is highly significant and indicates that properties with an excellent view have a higher price.

## **RESIDUAL PLOT EVALUATION**

The residual plot shows that the residuals are randomly scattered around zero, indicating that the model is a good fit. There are no obvious patterns or outliers in the residuals. The residuals appear to have constant variance, indicating that the model is able to capture the underlying relationships in the data.

## **KEY INSIGHTS FROM RESIDUALS AND PREDICTIONS**

The residuals are randomly distributed around zero, indicating that the model is a good fit. There are no significant outliers in the residuals that could indicate problematic predictions or data points. The residuals appear to have constant variance, indicating that the model is able to capture the underlying relationships in the data. The predictions are close to the actual values, indicating good prediction accuracy.

## **MODEL RATING**

I would give this model a rating of 8/10. The model is able to capture a significant portion of the variation in the target variable and is able to make accurate predictions. However, the model may be slightly underfitting, and the residuals could be improved by adding more features or adjusting the

model's complexity.