



# Regression Engine

## Model Report

### Overview

In this analysis, multiple linear regression was performed to examine the relationship between the dependent variable and several independent variables. This type of regression was chosen because it allows for the quantification of the individual impact of each predictor on the target variable while controlling for the effects of others. The model assumes a linear relationship between the predictors and the response, and it was trained on a preprocessed dataset with appropriate scaling and handling of missing values. The regression coefficients provide insights into the direction and magnitude of each feature's influence, while statistical metrics such as  $R^2$ , adjusted  $R^2$ , and p-values were used to assess the model's explanatory power and the significance of each predictor.

### Target

- price

### Features

- x
- carat
- color
- cut
- clarity
- y

- z
- depth
- table

## Regression Equation

- 931.500(x)  
+ 11223.461(carat)  
+ 1137.788(color\_E)  
- 120.068(color\_I)  
- 1023.950(color\_J)  
+ 350.656(color\_H)  
+ 1073.286(color\_F)  
+ 863.492(color\_G)  
+ 1352.844(color\_D)  
+ 973.767(cut\_Ideal)  
+ 899.567(cut\_Premium)  
+ 731.924(cut\_Good)  
+ 874.282(cut\_Very Good)  
+ 154.508(cut\_Fair)  
- 647.023(clarity\_SI2)  
+ 309.978(clarity\_SI1)  
+ 1229.447(clarity\_VS1)  
+ 921.915(clarity\_VS2)  
+ 1608.866(clarity\_VVS2)  
+ 1657.116(clarity\_VVS1)  
- 3473.492(clarity\_I1)  
+ 2027.240(clarity\_IF)  
+ 10.532(y)  
- 156.326(z)  
- 58.411(depth)  
- 25.836(table)

# Assumption Check

## 1. Linearity

Test Result: failure

Feature(s) without a significant relationship:

depth

## 2. Independence of Errors

Test Result: success

Durbin Watson Stat: 2.0

Threshold: 1.5 to 2.5

## 3. Normality of Errors

Result: failure

Jarque-Bera Stat: 0.0

Threshold p-value: 0.05

## 4. No Perfect Multicollinearity

Result: failure

High VIF Features: 4

Threshold VIF: between 5 to 10

## 5. Equal Variance of Errors (Homoscedasticity)

Result: failure

Breusch-Pagan Stat: 0.0

Threshold p-value: 0.05

## Model Evaluation

- Mean Absolute Error (MAE): **731.5479** (lower is better)
- Mean Squared Error (MSE): **1214729.4964** (lower is better)
- Root Mean Squared Error (RMSE): **1102.1477** (lower is better)
- R<sup>2</sup> Score: **0.9222** (higher is better)
- Adjusted R<sup>2</sup> Score: **0.9221** (should be close to R<sup>2</sup>)

## Conclusion

Based on the provided metrics, the model demonstrates **excellent** explanatory power, with an R<sup>2</sup> score of **0.9222**.

The prediction errors — Mean Absolute Error (MAE) = **731.5479** and Root Mean Squared Error (RMSE) = **1102.1477** — are considered **significant** based on common thresholds.

The adjusted R<sup>2</sup> score of **0.9221** being close to R<sup>2</sup> indicates that the model includes relevant features and is not overfitting.