**Diamond Analysis Report**

**Data Description**

Data set is about diamonds what are the features that determine the diamond price

Which has five columns and 54,000 rows with no null values

**Features Titles**:

carat, cut, color, clarity, depth, table, price, x, y, z

where x, y ,z are measuring the diamond dimension



**Data Preprocessing:**

No nulls were found , unnamed column was dropped

**objective:**

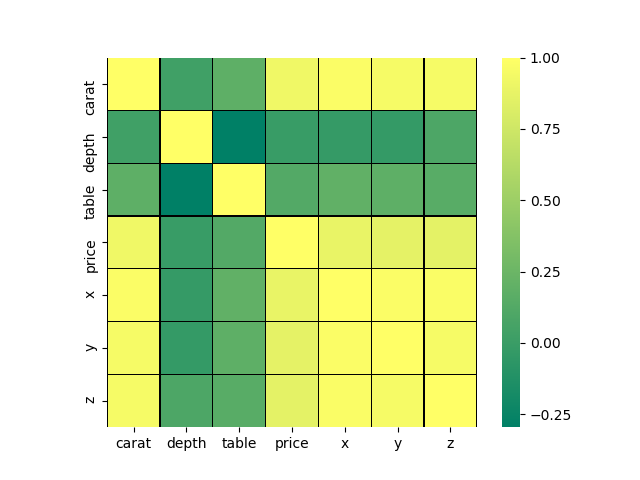
predicting the price of the diamond depending of their features carat, cut, color,

clarity, depth, table using different models and chose the best accurate model

Feature with P-values more than 0.05 were dropped



Correlation between feature:



**Models Evaluation And conclusion :**

**Linear:**

R-squared: Between Test And Predict 0.9209268693223643

R-squared: Between Train And Predict 0.9192587702706269

Mean Absolute Error: 0.18551101366375644

Mean Squared Error: 0.07912802142411524

Root Mean Squared Error: 0.2812970341544952

RMSE is 0.28 which is very less and the algorithm did an excellent of prediction

**KNN:**

With best prams of neighbour of 3 scored: 0.941664883093943

**SDGRegresser:**

With best prams of 'alpha': 0.001, 'eta0': 0.01, 'learning rate': 'constant

Scored: 0.9183967446859295

**AdaBoostRegressor with DecisionTreeRegressor:**

With Best Prams: max\_depth:3), learning\_rate = 0.02, n\_estimators = 110

Scored: 0.8772176690021309

**RandomForestRegressor:**

With Best Prams:max\_depth': 9, 'n\_estimators 180' scored: 0.9523561513872565

SVR:

With Best Prams :'kernel': 'poly' scored :0.9698689252989894

Most of the models have same around of score except the decision Tree has less

SVR had most highest score