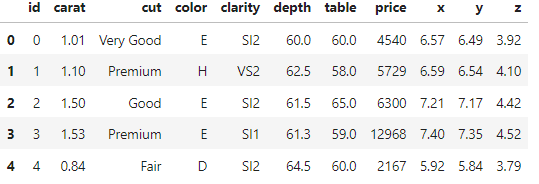
The original training dataset looks as follows:



* We first convert the x, y and z variables which represent the length, width and depth of the diamond respectively into a single variable representing the volume of the variable.
* The cut quality, color and clarity variables can each be converted to ordinal variables:

Cut: ranked from 0 (‘Fair’) to 4 (‘Ideal’)

Color: ranked from 0 (‘J’) to 6 (‘D’)

Clarity: ranked from 0 (‘I1’) to 7 (‘IF’)

The final training dataset looks as follows:

Graphical user interface, application, Teams

Description automatically generated

Using the price variable as our target variable, we can choose the remaining variables as explanatory variables in our predictive model. We use a LightGBM model to predict the price of diamonds.