For numerical features of the diamonds, such as weight, they will be encoded as is. However, for qualitative values such as color and cut, we would need to use one hot encoding to put the values into numerical format. For example, diamond cut can have one of the qualities

- Fair
- Good
- Very good
- Ideal
- Premium

For a diamond with qualities, say

- Carat = 0.23
- Cut = Ideal
- Depth = 61.1
- Price = 231 …

We can encode these features as

$[0.23\ 0, 0, 0, 1, 0, 61.1, 231 \ldots]^T$

This says

- Carat = 0.23
- Fair = 0
- Good = 0
- Very good =0
- Ideal = 1
- Premium = 0
- Depth = 61.1
- Price = 231 …

This treats each option for diamond cut as a separate characteristic in the feature vector. For this to work on the dataset, the ordering of attributes must be strictly followed. After this is done, we can apply some algorithm to the encoded features of the dataset.