**DIAMOND DATA ANALYSIS**



**ABOUT**

A diamond is one of the most expensive stones. The price of diamonds varies irrespective of the size because of the factors affecting the price of a diamond. To analyse the price of diamonds according to their attributes, we first need to have a dataset containing diamond prices based on their features.

**DATASET**

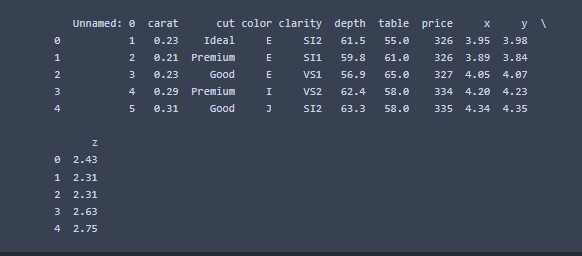
This classic dataset contains the prices and other attributes of almost 54,000 diamonds. It is a great dataset for beginners learning to work with data analysis and visualization.

* **Price:** price in US dollars (\$326--\$18,823)
* **Carat:** weight of the diamond (0.2--5.01)
* **Cut:** quality of the cut (Fair, Good, Very Good, Premium, Ideal)
* **Colour:** diamond colour, from J (worst) to D (best)
* **Clarity:** a measurement of how clear the diamond is (I1 (worst), SI2, SI1, VS2, VS1, VVS2, VVS1, IF (best))
* **X:** length in mm (0--10.74)
* **Y:** width in mm (0--58.9)
* **Z:** depth in mm (0--31.8)
* **Depth:** total depth percentage = z / mean (x, y) = 2 \* z / (x + y) (43--79)
* **Table:** width of top of diamond relative to widest point (43--95)

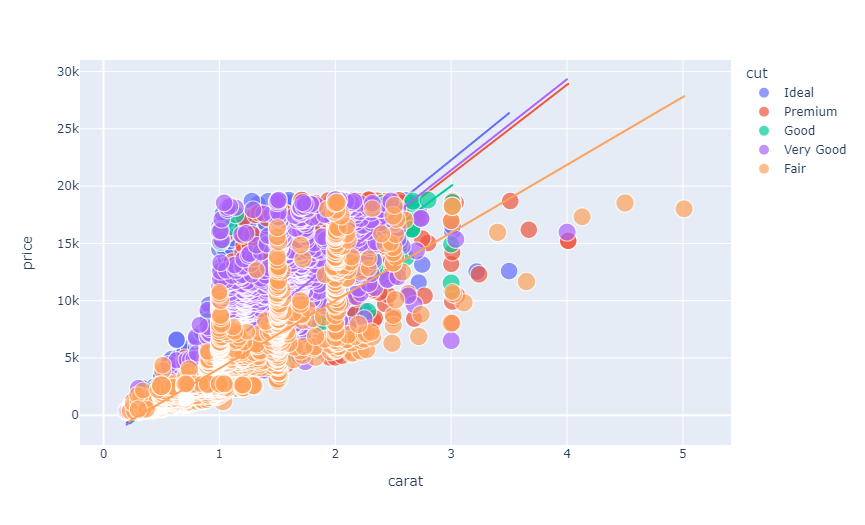


**IMPLEMENTATION**

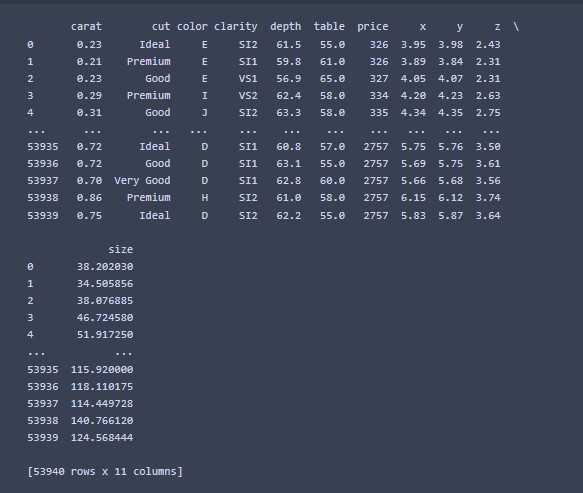
1. Let us start the task of diamond price analysis by importing the necessary Python libraries and the [dataset](https://www.kaggle.com/datasets/shivam2503/diamonds)



1. This dataset contains an Unnamed column. I will delete this column before moving further.
2. Now let us start analysing diamond prices. I will first analyse the relationship between the carat and the price of the diamond to see how the number of carats affects the price of a diamond. We can see a linear relationship between the number of carats and the price of a diamond. It means higher carats result in higher prices.

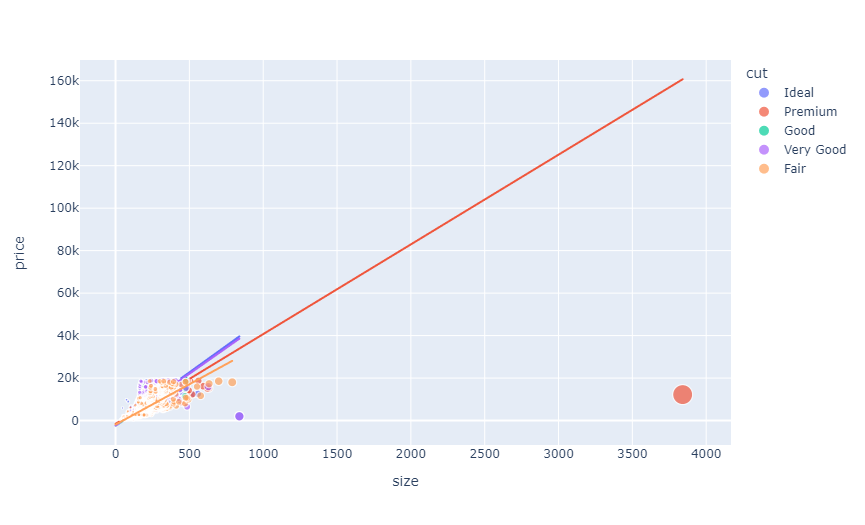


1. Now I will add a new column to this dataset by calculating the size (length x width x depth) of the diamond.



1. Now let us have a look at the relationship between the size of a diamond and its price. The above figure concludes two features of diamonds:

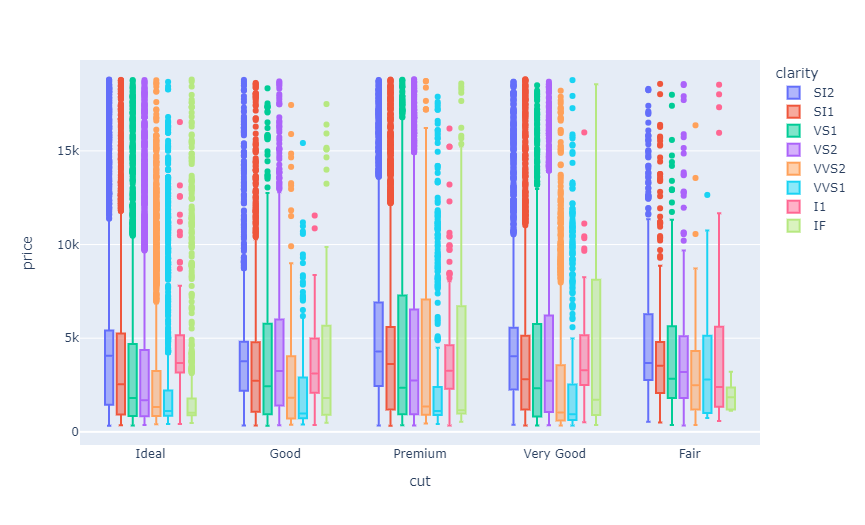
* Premium cut diamonds are relatively large than other diamonds
* There is a linear relationship between the size of all types of diamonds and their prices.



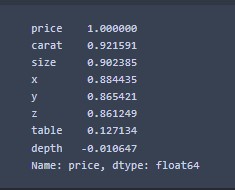
1. Now let us have a look at the prices of all the types of diamonds based on their colour.



1. Now let us have a look at the prices of all the types of diamonds based on their clarity.

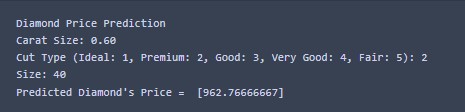


1. Now let us have a look at the correlation between diamond prices and other features in the dataset.



1. Now, I will move to the task of predicting diamond prices by using all the necessary information from the diamond price analysis done above. Before moving forward, I will convert the values of the cut column as the cut type of diamonds is a valuable feature to predict the price of a diamond. To use this column, we need to convert its categorical values into numerical values.

Now, let us split the data into training and test sets. Now I will train a machine learning model for the task of diamond price prediction. Now below is how we can use our machine learning model to predict the price of a diamond.



**SUMMARY**

So, this is how you can use your Data Science skills for the task of diamond price analysis and prediction using the Python programming language. According to the diamond price analysis, we can say that the price and size of premium diamonds are higher than other types of diamonds.

