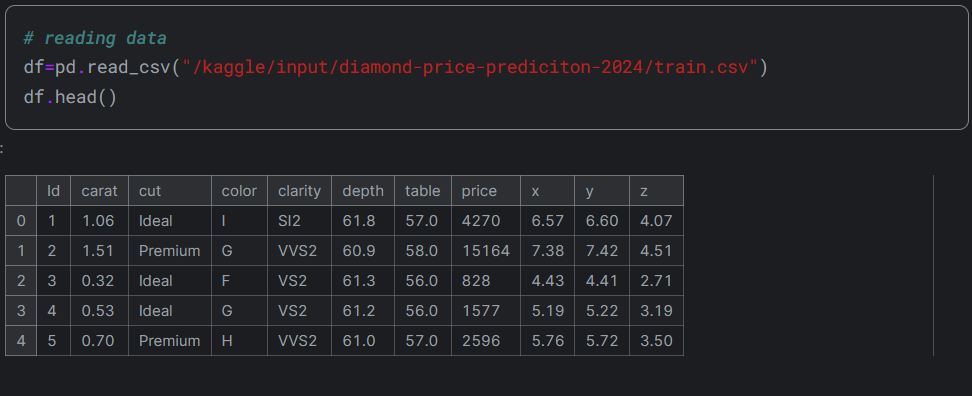
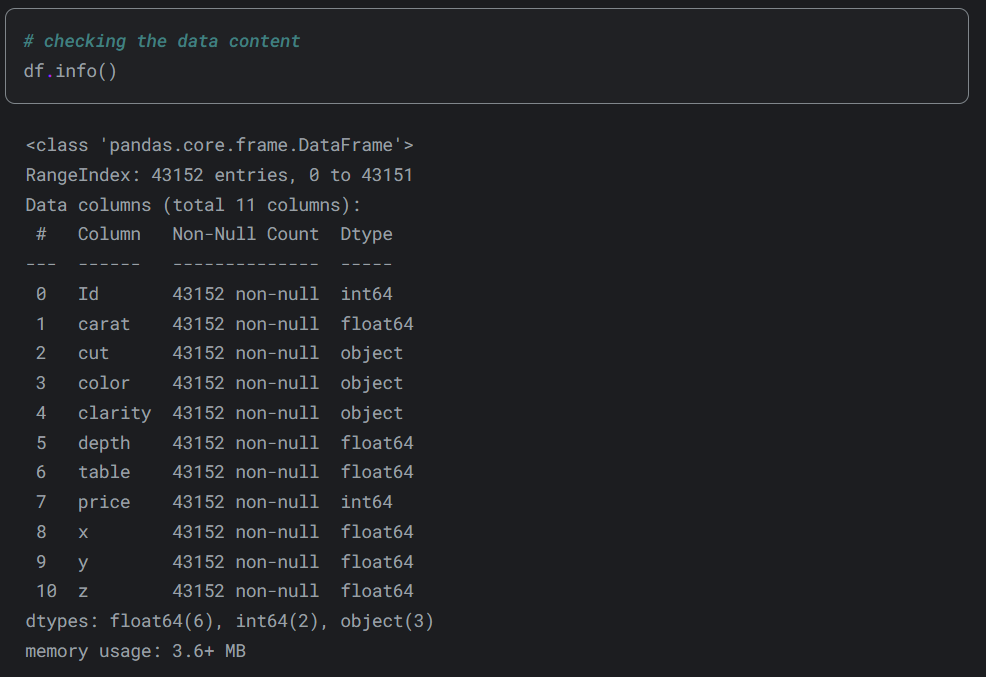
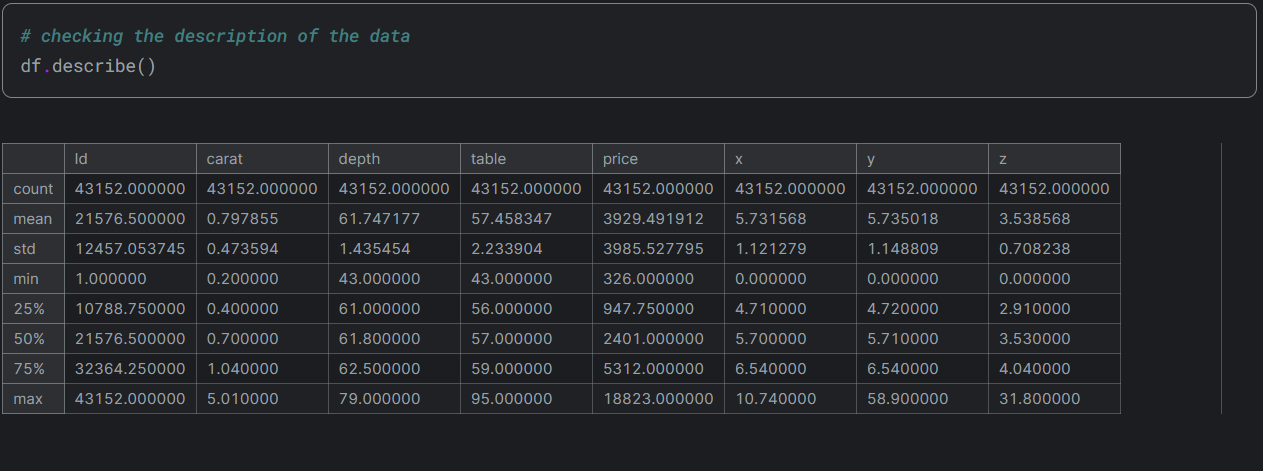


1- Reading the dataset :



2- Checking the dataset :

3- Checking the description :

## At the first glance:

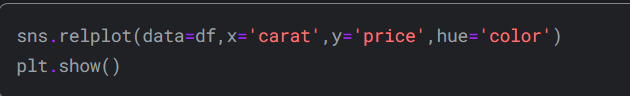
1. x , y, and, z have some 0 values which need to be taken care of.
2. it seems that there are alot of outliers.
3. there are no NaN values which is great.

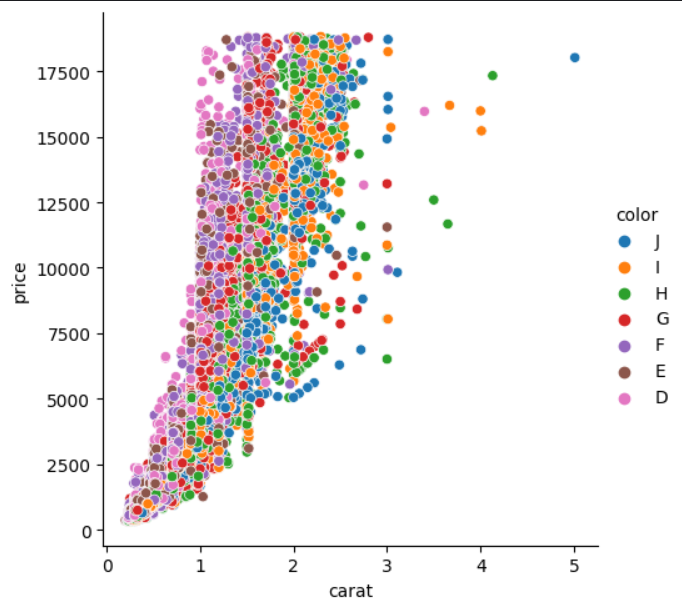
4- handling categorical cols :

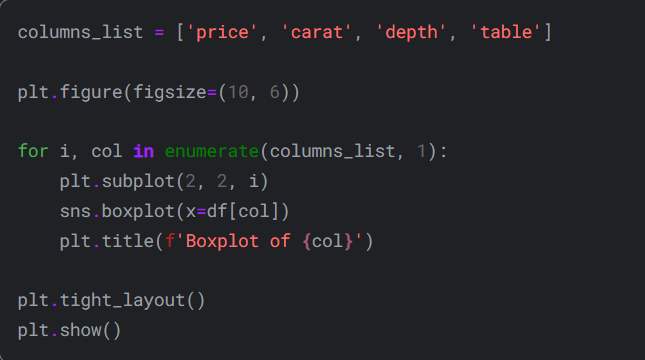


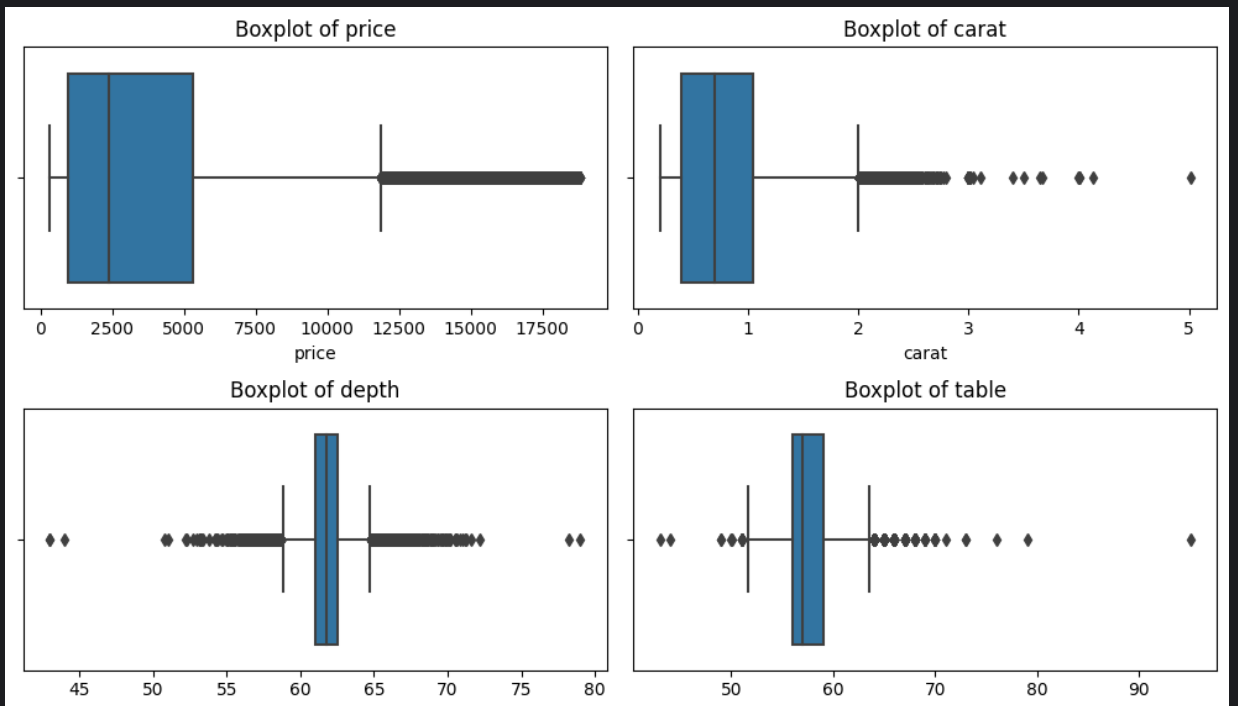
1. to increase readability and optimize the space.
2. order the categories properly.

5- visualizations to understand the data:



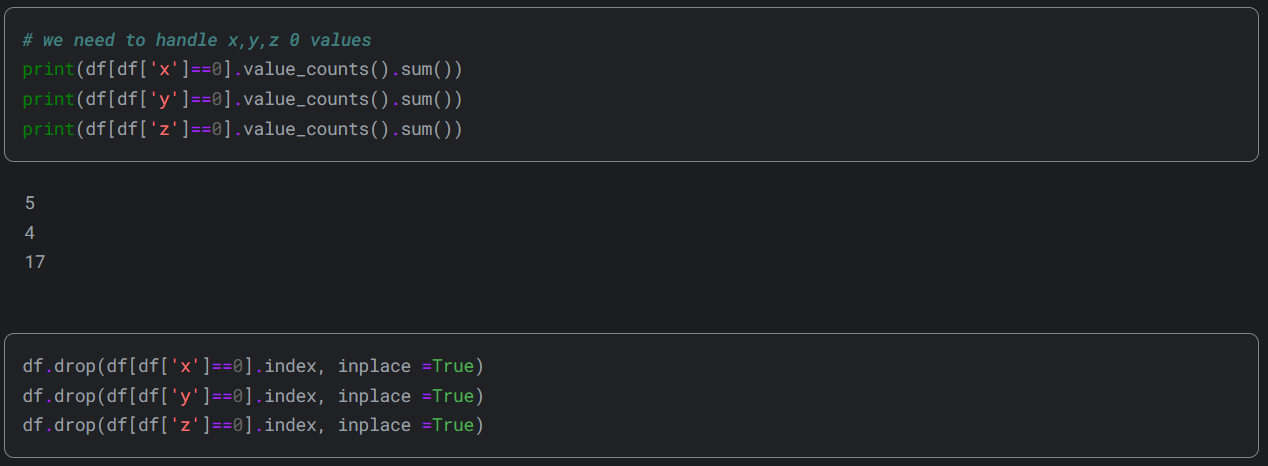




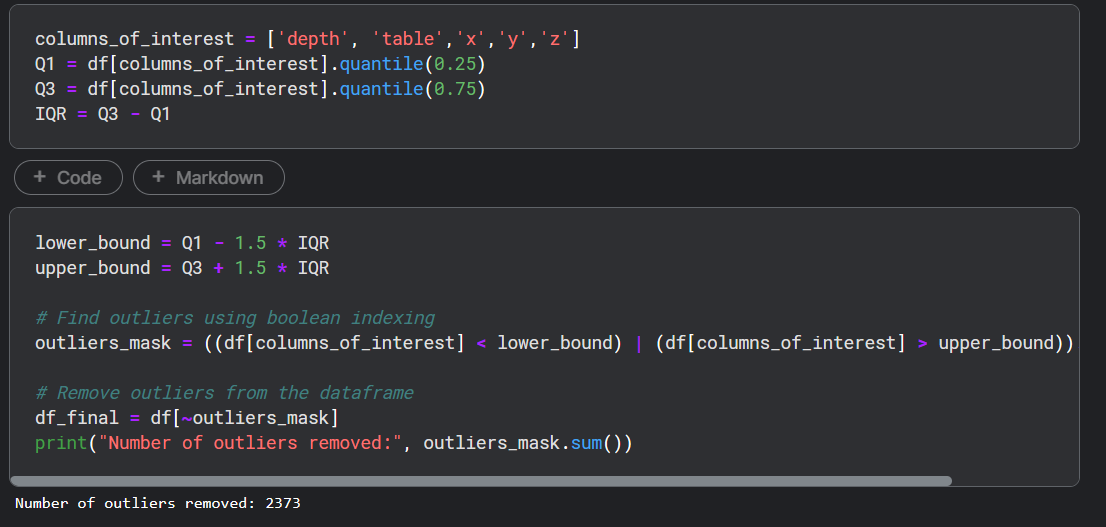


6- cleaning the data:

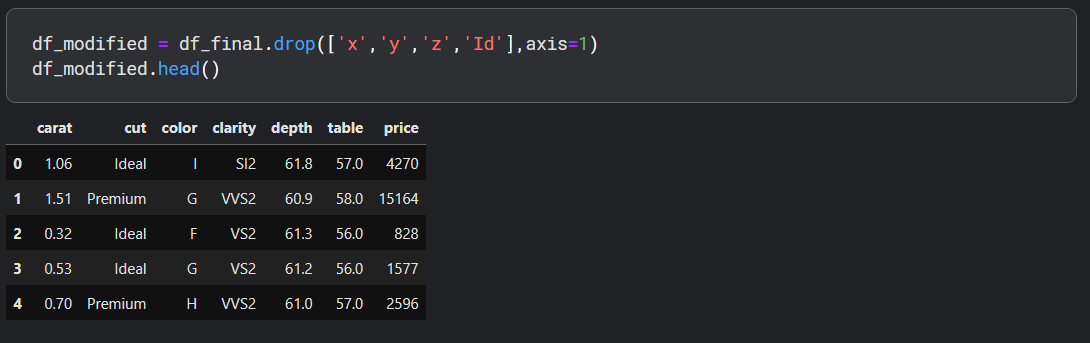
**1- handling the 0 values in the data.**



**2- handling the outliers in the data.**

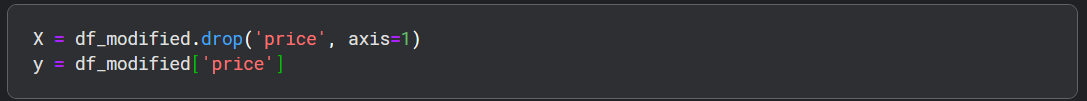


**3- dropping unneeded cols.**

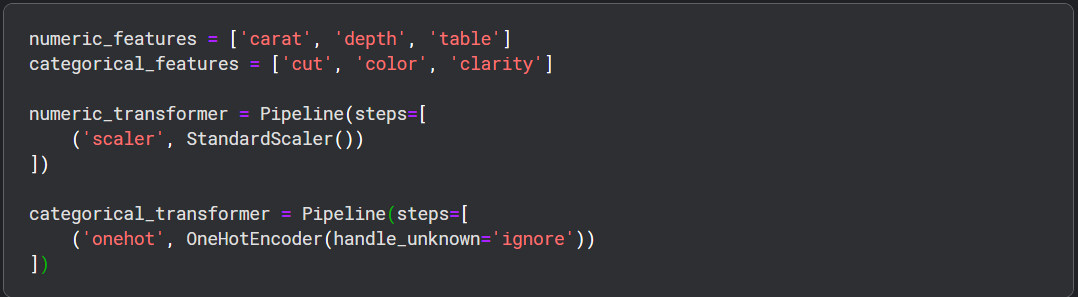
****

7- splitting the data:

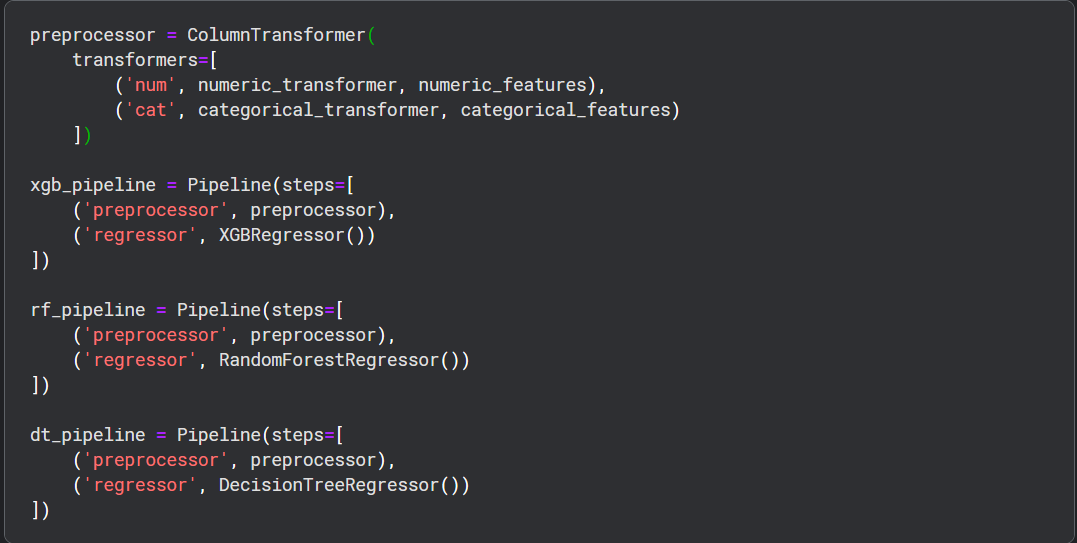
**1- splitting features in data to X and y.**



**2- splitting features into numeric and categorical.**

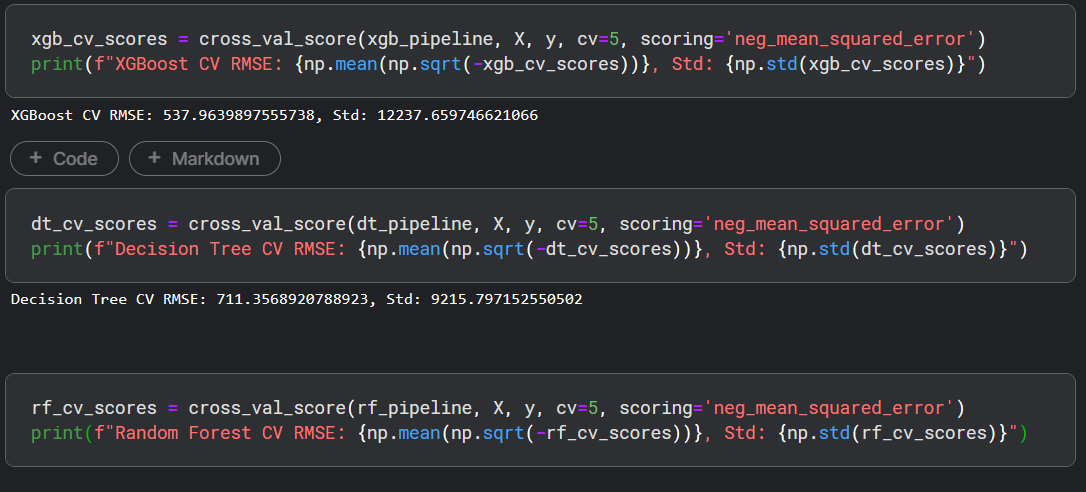


**3- building different pipelines for different models.**

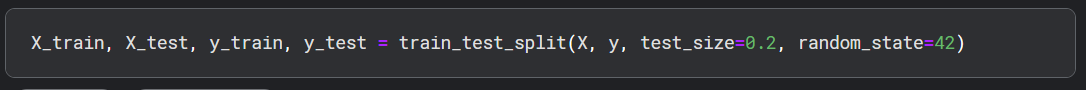


8- Working on the models:

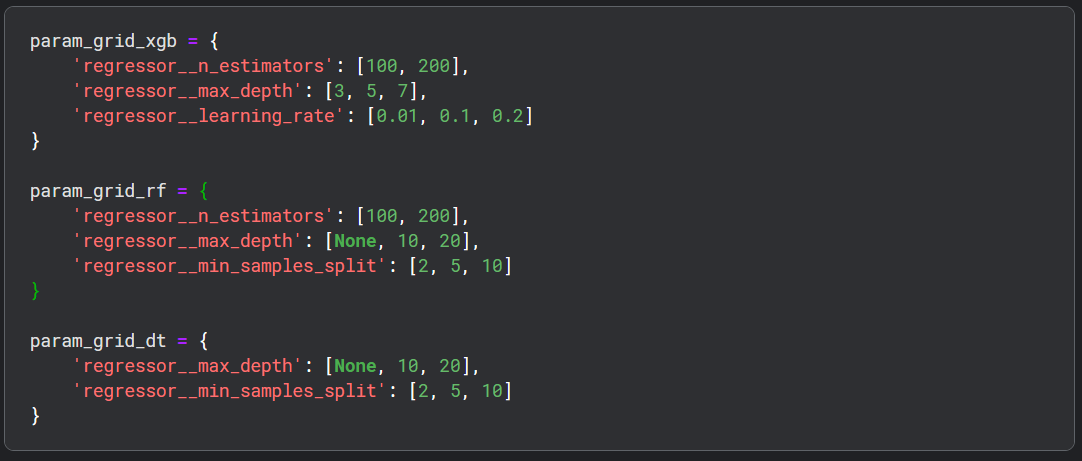
**1- using cross-val across the 3 models**



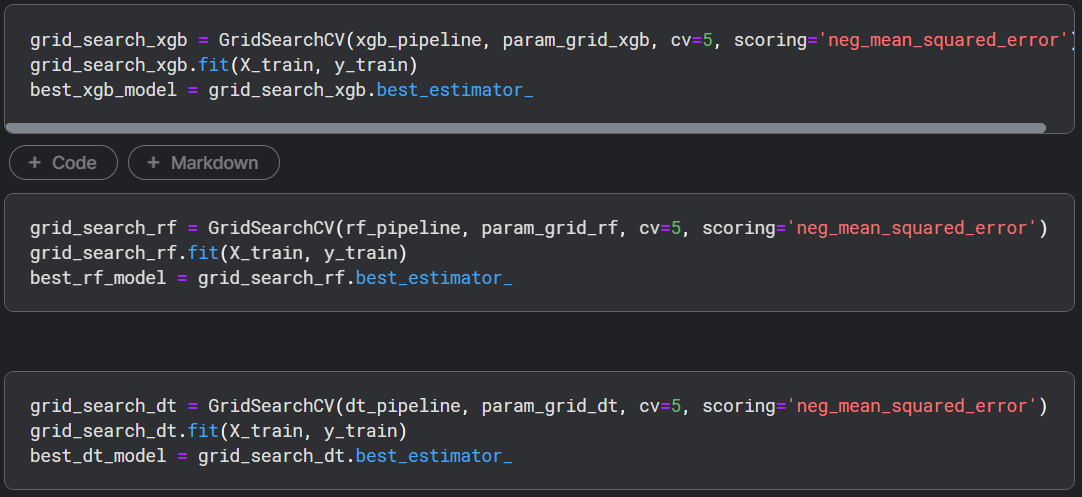
**2- splitting data into train and test.**



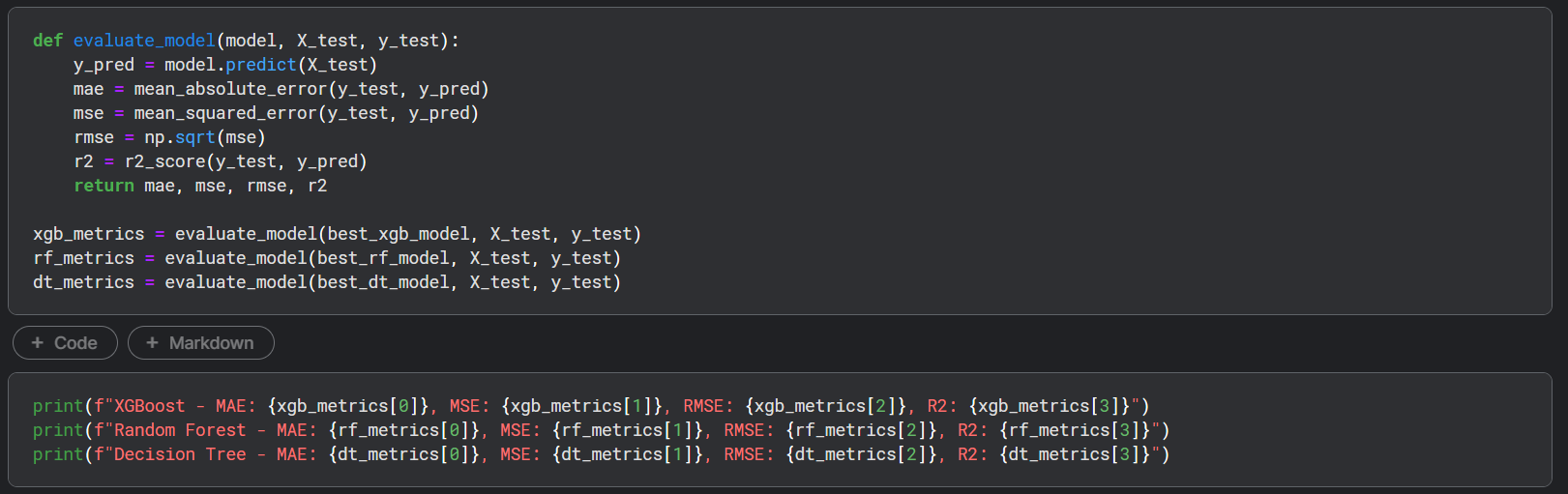
**3- setting up some params for the grid search to finetune the models.**



**4- using grid search for all the 3 models.**

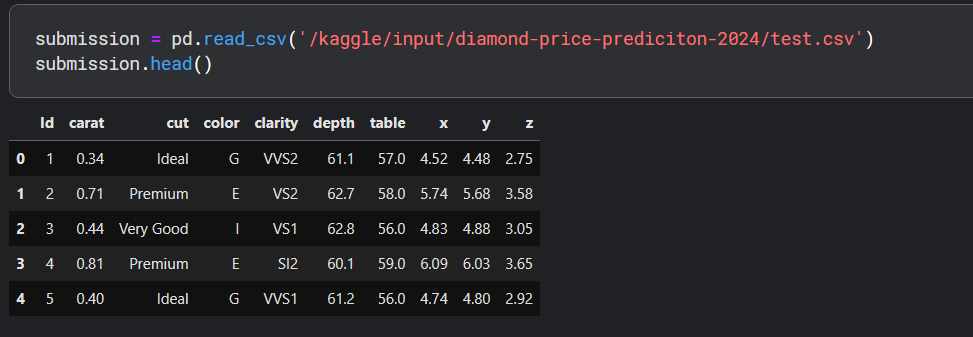


**5- evaluating the models.**

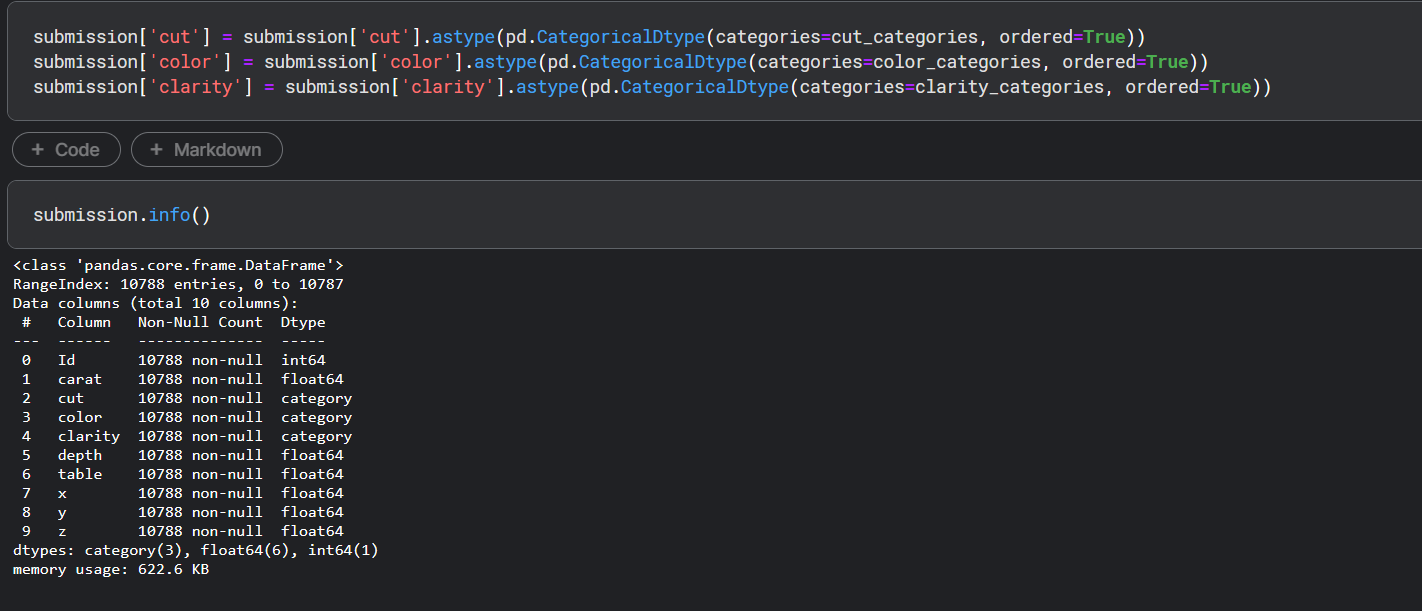


9- Predicting the test data:

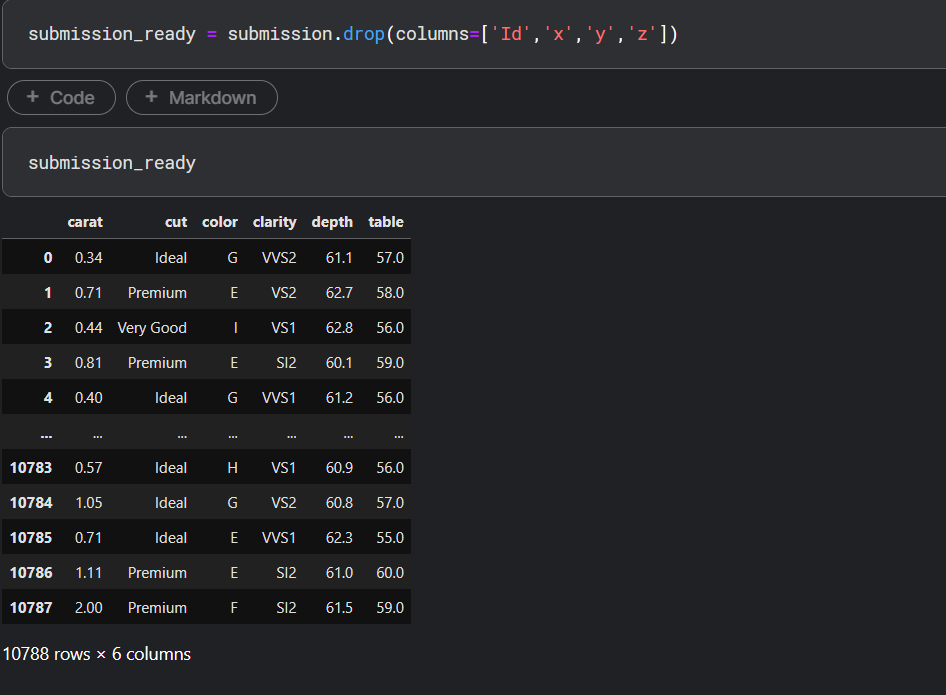
**1- reading the test data.**



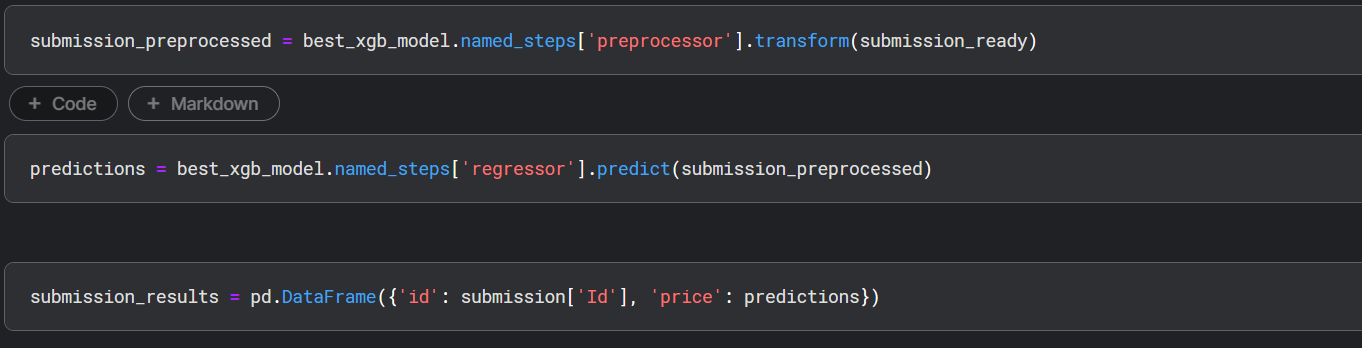
**2- preparing the data.**



**3- reading the test data.**



**4- predicting the prices for the diamonds in the test data.**



**5- checking and saving the predictions in a csv file format.**

