Assignment 8 – Stats

Q1. Import the attached CSV files (Diamond.csv) and answer the following questions:

1. Create 2 dataframes out of this dataframe – 1 with all numerical variables and other with all categorical variables.
2. Calculate the measure of central tendency of numerical variables using Pandas and statistics libraries and check if the calculated values are different between these 2 libraries.
3. Check the skewness of all numeric variables. Mention against each variable if its highly skewed/light skewed/ Moderately skwewed.
4. Use the different transformation techniques to convert skewed data found in previous question into normal distribution.
5. Create a user defined function in python to check the outliers using IQR method. Then pass all numeric variables in that function to check outliers.
6. Convert categorical variables into numerical variables using LabelEncoder technique.
7. Use both the feature scaling techniques (standardscaler/min max scaler) on all the variables.
8. Create the Histogram for all numeric variables and draw the KDE plot on that.
9. Check the correlation between all the numeric variables using HeatMap and try to draw some conclusion about the data.

Q2. Explain Gradient descent in detail. How changing the values of learning rate can impact the convergence in Gradient Descent.

