Name: Siddharth Mohanty
Branch: EXTC
UID: 2019120038
Batch: B
Course: Data Analytics Lab
Experiment: 1

AIM: To perform EDA on a dataset which is analyzing the data features such as count of different values, analyzing the mean, min and max of the values, seeing the correlation and plotting the heatmap for the same using the seaborn library and then finally making conclusions about the dataset.

Problem Statement: Choose any of the given datasets and apply EDA on it

Code: https://github.com/51stDimension/EDAExperiments/blob/main/Experiment%201/EXP1.ipynb
Dataset: https://github.com/51stDimension/EDAExperiments/blob/main/Experiment%201/E1Data.csv

Conclusion:

After implementing EDA on the dataset, I made the following conclusions about the dataset:
The dataset is about Student's performance. The attributes in the dataset are
Gender: The gender of each student
Race: The race of the student
Parental level of education: About the education of their parents
Lunch: Lunch option opted by the student in school
Test prep course: If it has been completed or opted
Math/Reading/Writing score: The score obtained in respective exams

The count of females(>500) is more than the count of males(>400 and <500) in the dataset. So we can conclude that more females appeared for the test compared to the number of males.

We have analyzed the count of parental level degrees and we can conclude that most of the parents have completed college degrees and then the second most popular degree was the associate's degree.

We also saw how reading score and writing score were highly correlated which indicated that people who are good at reading are also good at writing.

Then we also saw that the data is normally distributed.

We also confirmed on our conclusion that the reading score and writing score have a linear relation by plotting the relational plot of the two.