**Report Analysis**

The data is about, grades on three exams and overall course grades for 233 students during several years for a statistics course at a university.

The data set is taken from the link

https://vincentarelbundock.github.io/Rdatasets/datasets.html

There are 6 dimensions of which 5 are independent and 1 is dependent. In my task, I use all the dimensions

At first step:

Identifying the data type by using “type(‘data’)” is also done manually.

In the second step:

Find missing values and fill them with standard division.

In the third step:

Find the duplicates and handle them by using the drop duplicates method

In the fourth step:

Check the skewness of the data

In firth step:

Identify and remove the outliers by plotting boxplot

In the sixth step:

Normalize the data by using the Sklearn function

In the seventh step:

Smooth the data by using the Min\_Max method

After all the preprocessing steps apply multilinear regression because we have multiple independent variables and only 1 dependent variable.

In the end,t is difficult to visualize the data with so many variables so I represent the data only in tabular form.