

## **Project Description**

The dataset for this project is a fabricated list of exam grades for a class. The label feature in this dataset is the sum of grades and the IDs of the records, representing the names of students. Each part of the project is explained below.

(Please ensure that the last line of the data file is not a record. This line has been added to assist in post-processing the data.)

### **Preprocessing:**

In this section, you need to perform the following steps:

1. Fill in the missing data with the mode (The mode of a set of data values is the value that appears most often) of the target feature.
2. Identify outliers using IQR method with a threshold of 2.5 IQR.
3. Manage outliers with filling in the data with the median of the target feature.

(You can use the NumPy and Pandas libraries)

### **Normalization:**

In this section, perform necessary actions to shift the mean of each feature to zero and then normalize the standard deviation of each of the features to one.

### **Correlation Matrix:**

In this section, calculate the correlation between all features of the dataset. Then, do the following:

1. Select two features with the highest correlation.
2. Remove the feature with the lower correlation with the label from the dataset.
3. Repeat steps 1 and 2 until a total of two features are removed from the dataset.