The second phase of Microsoft Student Accelerator consists of several parts. In what follows, I will give a schematic picture of the processes and methods that I was deployed. The three stages consist of preprocessing, modeling, and deep learning.

Part one:

Regarding preprocessing and cleaning stage, I used the three data sets given. I grasped a general look of the columns, types, the number of rows, and state of missing data. I formed, using merge, a main data set to work with. A general look, using describe, was obtained. I evaluated the density of each column, whether categorical or continuous. The dependent variable and independent variables were recognized. The correlation of the variables was investigated using different plots. Categorical data were encoded. The missing data was eliminated. Further, newly needed columns were added. The data was standardized, and outliers were eliminated.

Part two:

Regarding the modeling, I used the first part to eliminate those columns which do not seem to be influential. I used a pipeline to build several regression models. In this part, for a better modeling, I used cross validation. I chose the model and further evaluated it and used it for forecasting.

Part three:

Regarding phase three, deep learning, after basic preparation of data, I created three different sets of train, validation, and test, from the training data set. I built a Convolutional Neural Network and using a standard approach, I trained my model. Using several different metrics and the confusion matrix, I assessed the performance of my model.