The team had previously decided on water quality dataset, this dataset has been discarded as there was no correlation between the features and output. We have chosen the heart attack analysis dataset for further training and modelling. The new dataset is [Heart Attack Analysis & Prediction Dataset | Kaggle](#).

The team has built scatterplots, box plots, corelation matrices, heatmaps and count plots using each feature and the dataset output. This has led us to the conclusion that some of the features in the dataset does not have a significant correlation with the output. Hence, the team has decided to remove one of the features to increase accuracy while training the model. The fasting blood sugar level feature has been deleted.

The dataset was further inspected for null values, data cleaning was performed and there are zero null values. Histograms were built in-order to make sure that the data was in normal form.