**Case Study Assignment – Data Mining**

A healthcare organization together with a couple of government hospitals in a city has collected information about the vitals that would reveal if the person might have a coronary heart disease in the next ten years or not. This study is useful in early identification of disease and have medical intervention if necessary. This would help not only in improving the health conditions but also the economy as it has been identified that health performance and economic performance are interlinked.

As a data scientist, you are required to construct a classification model based on the available data and evaluate its efficacy. Your activities should include - performing various activities pertaining to the data such as, preparing the dataset for analysis; checking for any correlations; creating a model; evaluating the performance of the classification model. Visualizations would be a value add.

Demonstrate Data Mining process with following activities:

* Problem statement
* Perform exploratory data analysis
* Preprocess the data
* Select Training data, test data
* Train the model
* Test the model (Predictions and reporting)
* Evaluate the model performance
* Suggest ways of improving the model
* Any interesting observations
* Challenges faced and how you mitigated the challenges
* Assumptions if any

Important Points to be considered

* Missing Data may be represented by either NAs, Blanks or values such as -999/-99 etc. Please check for various possibilities.
* Results can be varying from team to team. But, you should be able to justify your result.

**Solution**

Observations

Only A2 has 5.08 percent of NAs in the data