**Lab Objective:**

This lab reinforces the importance of feature engineering concept and maneuvers. The students are to work in pair programming mode to execute the following maneuvers with the given data:  
  
1) sample data: ‘raw\_diabetes\_string included.csv’

This dataset is about a group of females who are being tested for diabetes. It is to be used to develop a prediction model using a ML model.

1. Deal with missing or null values – check these columns to see if there is any missing or null values (Glucose, BloodPressure, SkinThickness, Insulin, BMI, vital\_function, Age) and execute.
2. String to integer conversion – For the column ‘Outcome’, you will need to convert.
3. Data normalization – Before you run the data normalization, make sure to leave out the ‘Outcome’ column as it is NOT part of the features.
4. Column manipulation – move any columns (2X) around and rename a few column headers (2X).

2) Find **two** more csv datasets (eg. kaggle.com or some are on our Canvas) that are similar to above and execute the similar maneuvers.

**What you must turn in:**

Your choice of **two** data and **three** completed .ipynb files from above. From your pair programming, each of you must turn in in your own Canvas submission.

**Due date/hr:**

Before the start of next Monday class, 2pm.