**Student: Seif Kungulio**

**Date: 03/09/2025**

**Subject: Project 8**

**Class: DSCI 502**

**Section: 01W**

**Instructor: Sean Yang**

**File Name: Project8\_Kungulio\_Seif.docx**

1. Load the dataset in breast\_cancer\_data.csv into R. Call the loaded data breast\_cancer\_data. Make sure that you have the directory set to the correct location for the data.
2. Define a user defined function BoxplotPredictorOnTarget with two arguments, the target and one predictor to plot the box plot of predictor based on different category of the target. Then use this user defined function to generate the box plot:
   1. area\_mean against Diagnosis
   2. area\_se against Diagnosis
   3. texture\_mean against Diagnosis
3. Build the following logistic models to forecast the Diagnosis and recommend the best model based on McFadden/pseudo R squared to the management.
4. forecast Diagnosis using area\_mean
5. forecast the Diagnosis using area\_mean and area\_se
6. forecast the Diagnosis using area\_mean, area\_se and texture\_mean
7. forecast the Diagnosis using area\_mean, area\_se, texture\_mean and concavity\_worst
8. forecast the Diagnosis using area\_mean, area\_se, texture\_mean, concavity\_worst and concavity\_mean