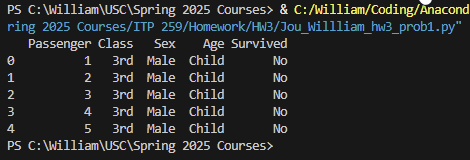
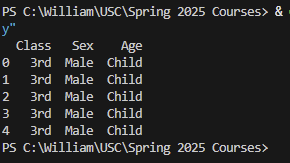
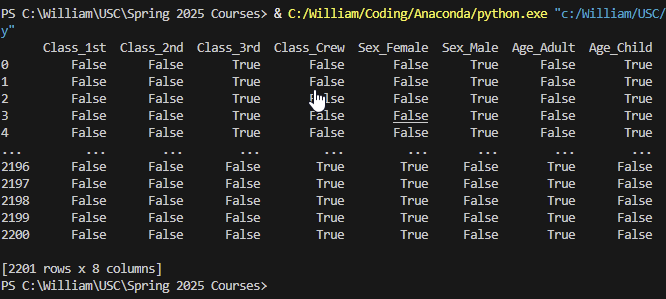
***Problem 1:***

Reading titanic dataset in, and printing the head

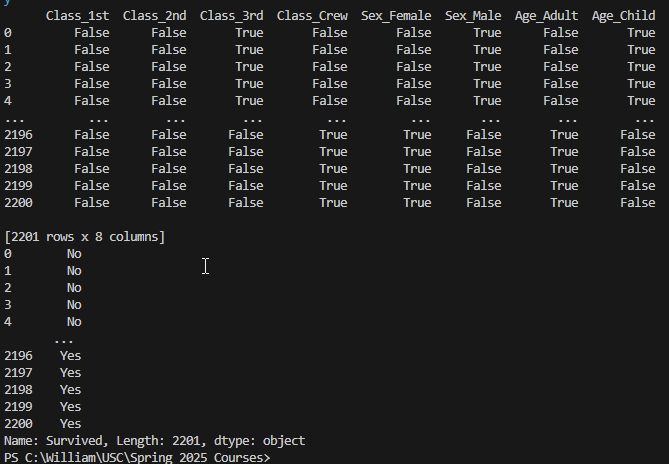
Dropping the irrelevant factors and the target variables and printing the head



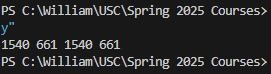
Converting the categorical feature variables into dummy variables, printing the dummy variable data frame head.

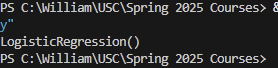


Assigning feature variables to X and target variable vector to y, printing X and y

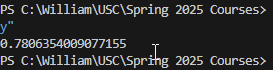


Splitting the data and then printing the shapes of the training/testing data

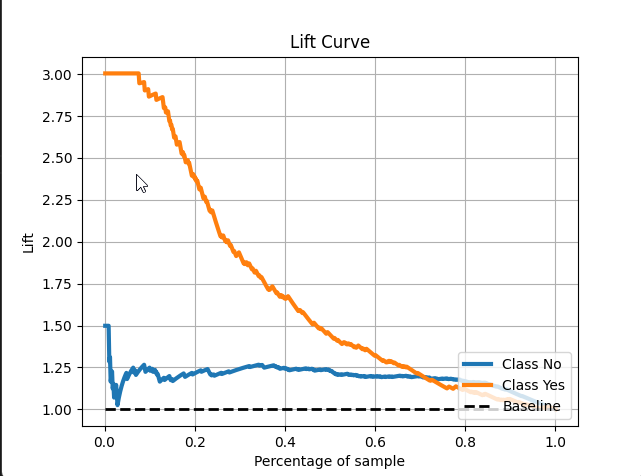


Fitting the training data to a logistic regression model, printing the model  


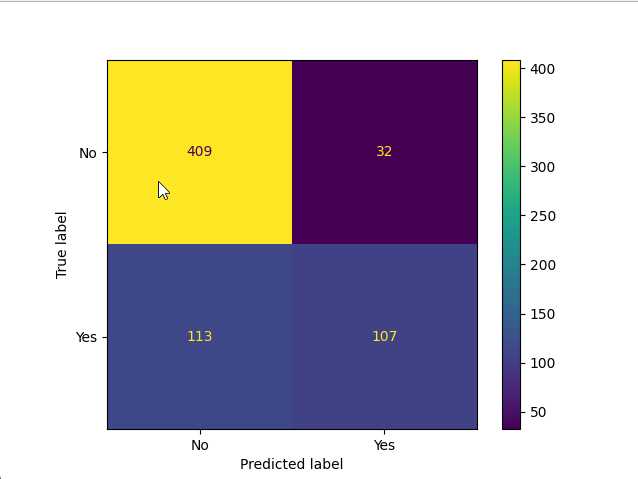
Using the model to predict with the test data, then printing the accuracy score of the model’s predictions



Plotting the lift curve



Plotting the confusion matrix



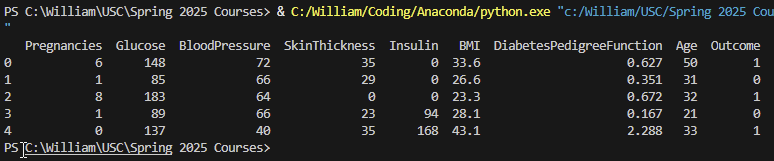
Printing the model’s prediction to see if a male adult passenger in 3rd class would survive

A black screen with white text

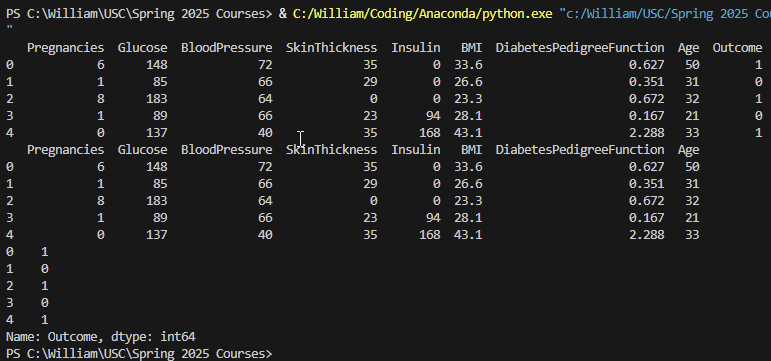
AI-generated content may be incorrect.

***Problem 2:***

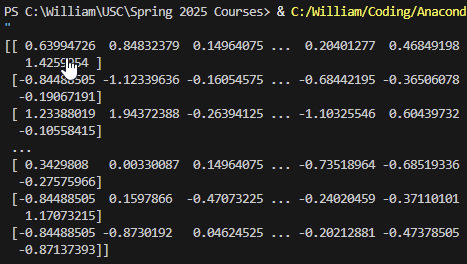
Reading in the diabetes data frame and then printing the head

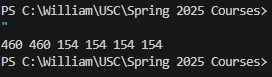


Assigning the feature variable matrix and the target variable vector and printing their heads.

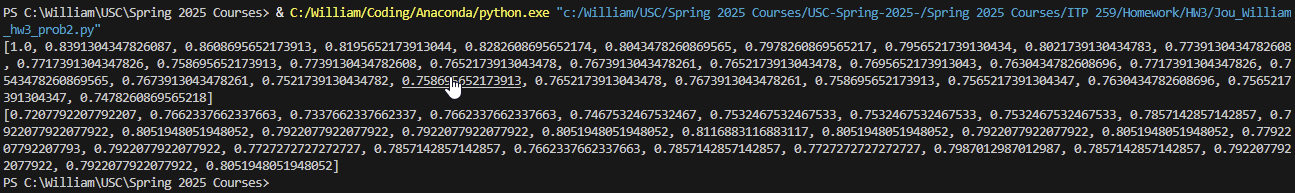


Standardizing the feature data and printing the scaled data

  
Splitting the data into training A data and training B data and test data and then printing the shapes of the data splits



Developing a knn model for number of neighbors ranging from 1-30 and then using it to predict on the training A and training B data. Printing the accuracy scores for training A and training B data (not all of output captured in screenshot because it’s too long).



Plotting training accuracy over different number of ks (you would choose ~10neighbors) A graph of a number of neighbors

AI-generated content may be incorrect.

Creating a new knn model that uses optimal ks (10) and the scores the data using the test

data

A black background with white text

AI-generated content may be incorrect.

Plotting the confusion matrix of the model

A yellow and blue squares with numbers

AI-generated content may be incorrect.

Printing the model’s prediction to see if a person with 2 pregnancies, 150 glucose, 85 blood pressure, 22 skin thickness, 200 insulin, 30 BMI, 0.3 diabetes pedigree, and 55 age would have diabetes.

