PITAO, ANGELICA JAYNE L.

To be able to create a proper Logistic Regression Model, I first split the test data into two parts, 80% to train the model and 20% to test its accuracy. I separated the variables by creating descriptive dummy variables to properly classify each data. Then, I also transformed the TotalCharges to numeric data to make it usable for the creation of the Logistic Regression Model, and I made sure to remove possible missing variables.

After this, I was able to create a Logistic Regression Model with a training accuracy of 0.8010 and a test accuracy of 0.8194 (rounded off to four decimal places). The model would have changed/improved if I had the proper knowledge of how to adjust the threshold, but errors kept coming up so I didn't have the chance to see the possible outcomes. Therefore, I used my model for the testing dataset.

The test set had a similar format with the original train set. However, after cleaning up the data by converting TotalCharges to numeric and setting the dummy variables, I noticed that the customerIDs 4075-WKNIU and 9003-CPATH (highlighted in the excel file) had blanks for the TotalCharges which didn't make them qualified to be tested using the regression model. Therefore, I redid the process earlier and created another model to remove the TotalCharges. This model had a training accuracy of 0.8072 and a test accuracy of 0.7968 (rounder off to four decimal places). I used this model only for the two customerIDs stated earlier.

On the other hand, I used my original model for the rest of the customerIDs to obtain my final prediction. To summarize, my predictions resulted to 106 yeses and 394 noes. Looking into the customers that are predicted to churn, there are some common behaviors based on the data. Majority of them do not have partners and dependents, they use FiberOptic, do not avail online security, online backup, device protection and tech support, subscribe to paperless billing, and pay using electronic checks. Services such as online security, online backup and device protection are somehow related so if a customer does not prioritize security, then most likely he/she would avail none/less of these offers. I would recommend for the company to offer more variety to their services so that they are able to attract people of different interest or increase trending campaigns in relation to phone security and data privacy to make their market highly informed of the need for protection.