Data Analysis- Logistic Regression

For our logistic regression we started off by initializing a spark session to retrieve our data to read four CSV files to use for our machine learning. Afterwards, we merged them into a DataFrame using the union operation. Once a complete, the next step was to clean the data before modeling. To do this we converted the DataFrame to a pandas Dataframe. Removing unnecessary columns and changing the data types accordingly. Additionally, identifying the categorical columns that would help our machine learning get an accurate score. Then using the OneHotEncoder to convert the categorical columns into a numerical format for our machine learning algorithm. Separating the features from the target to train and scale the data by defining the X, and Y with our data. Then we split the data and scaled it to fit the model. Afterwards, we were ready to proceed creating the logistic regression model. We initiated the model with the solver set to the default, with a parameter of 200. Furthermore, we put in the X and Y values we set from our data to get our final machine learning score. However, our accuracy score was not reaching the 75% threshold required. Although a few modifications were made, there was no drastic changes. Therefore, we decided to change models to see how it would compare with this logistic regression model.