**ASSIGNMENT #1:**

Use the titanicTrain.csv dataset for this first assignment

**NOTE: reference assignment1sample.py for code guidance**

1. Read the dataset into a dataframe
2. Determine what is the target variable
3. Determine and remove the variable(s) that are not likely to be relevant for logistic regression. (You will only need the following variables: Survived, Pclass, Sex, Age)
4. Make sure there are no missing values
5. Plot the histograms of all the variables in a 2x2 figure
6. Convert all categorical feature variables into dummy variables
7. Fit the training data tot a logistic regression model
8. Using the Test data, calculate the accuracy of your predictions for the target variable
9. Plot the confusion matrix along with the labels (Yes, No). Include in the title of this plot the accuracy number you calculated from the step above
10. Display the predicted value of the survivability of a 30 year old male passenger traveling in 3rd class.

**SAMPLE OUTPUTs SHOULD LOOK LIKE: Chart, histogram

Description automatically generated**

**Chart

Description automatically generated**

**ASSIGNMENT #2:**

Use the wineQualityReds.csv dataset for this first assignment

**NOTE: reference assignment2sample.py for code guidance**

1. Read the dataset into a dataframe
2. Drop Wine from the Dataframe
3. Extract Quality and store it in a separate dataframe
4. Drop Quality from the DataFrame
5. Normalize all columns of the Dataframe. Use *StandardScaler*
6. Create a range of k values from 1:11 for KMeans clustering. Iterate on the k values and store the inertia\_ for each clustering in a list
7. Plot the chart of inertia vs number of clusters k
8. What K (number of clusters) would you pick for KMeans?
9. Now cluster the wines into that same number of K clusters. Use random\_state = 2021 when you instantiate the KMeans model. Assign the respective cluster number to each wine
10. Add the quality back to the dataframe
11. Now print a crosstab (from Pandas) of cluster number vs quality
12. Do the clusters represent the quality of wine? Why or why not?

**SAMPLE OUTPUT SHOULD LOOK LIKE:**Chart, line chart

Description automatically generated