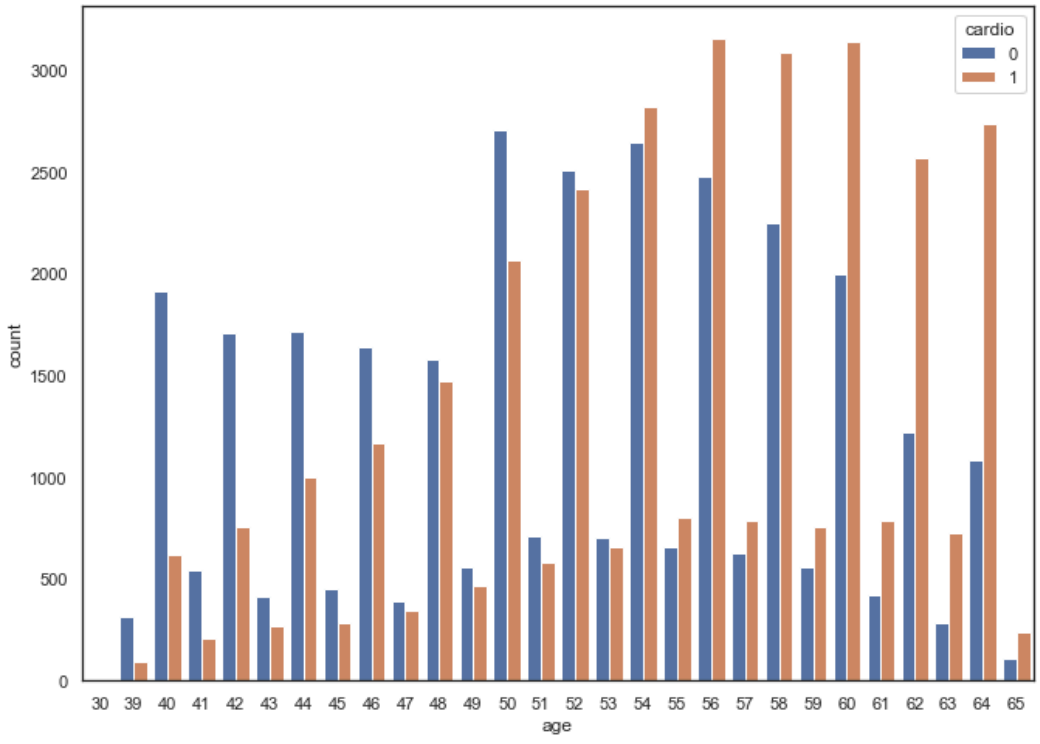
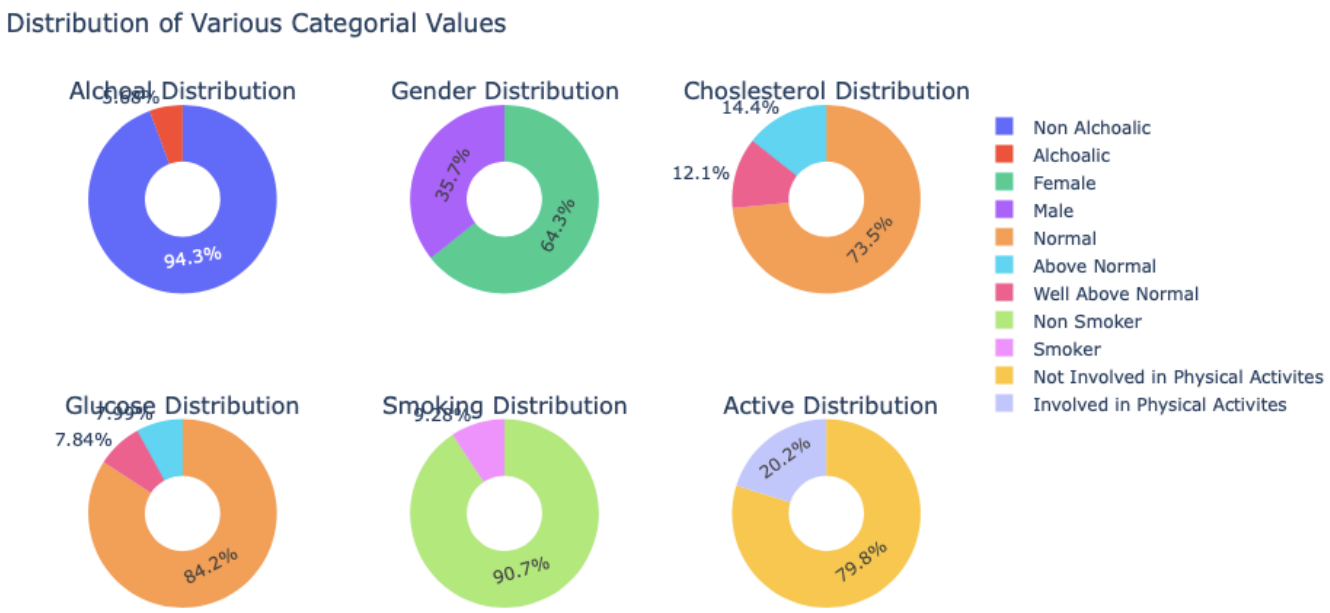
Cardiovascular Disease - Classification MVP

The goal of this project is to predict the likelihood of a person developing cardiovascular disease. by a classification model.

To start the classification, work a baseline model was needed to compare the affect of any modification in the data and the model. After cleaning the data by dropping duplicated rows and the outliers, the baseline model has been chosen to be Logistic Regression model and it gave 0.717training accuracy and 0.719 for the test





Modeling workflow:

KNN	KNN Accuracy = 71.73 KNN Precision = 75.58 KNN Recall = 66.49 KNN F1 score= 70.75
Decision Tree Classifier	Decision Tree Accuracy = 60.85 Decision Tree Precision = 62.10 Decision Tree Recall = 61.25 Decision Tree F1 score= 61.67
RandomForest	Random Forest Accuracy = 66.79 Random Forest Precision = 67.41 Random Forest Recall = 68.56 Random Forest F1 =0.679