Data set

The data set consist of features that could be associated to heart disease Like age , chest pain type and other features you can check data set and its documentation on
https://www.kaggle.com/fedesoriano/heart-failure-prediction

Summary of findings

I was able to build a model to predict heart disease with accuracy 86.5% , Fixing data outliers , Find most effective features on heart disease and proportions of features unique values and heart disease

Key insights

Finding that our target Heart disease is balanced in our data set , however data collected from unbalanced gender

Then we dive into the most effective features and we find that the most effective are ST_slope and Chest Pain Type also we found useless features that is not a factor to heart disease like RestingBP

Also we dive into relation between heart rate and age combined with proportion of heart disease occurring

We build a trusted model with an accuracy of 86.5% to predict Heart disease

We found that heart disease more likely to occur in men than females

Used references

https://www.kaggle.com/andreshg/xgboost-optuna-hyperparametertunning