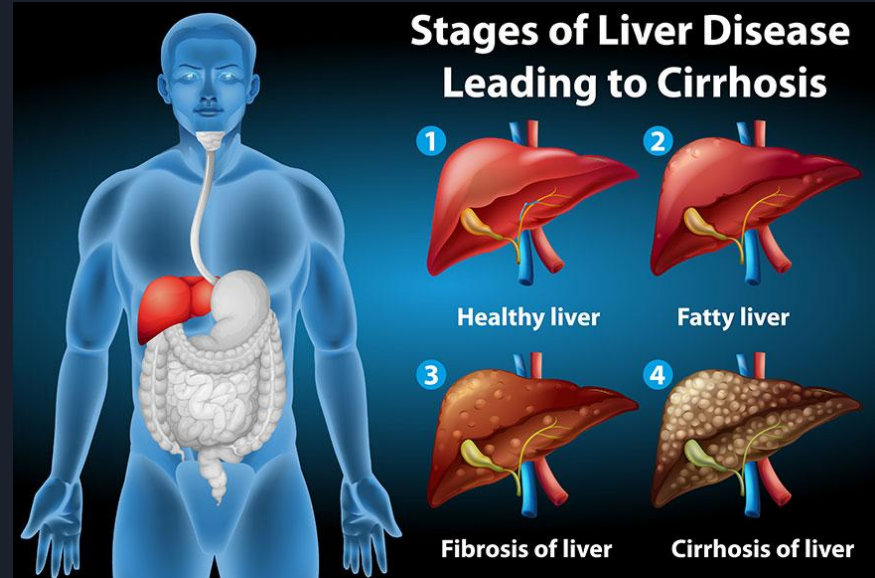# Cirrhosis Prediction

By Ashik Sathiya

Context

The Problem

- Cirrhosis is advanced liver scarring caused by various liver diseases like hepatitis and chronic alcoholism.

- **Goal**: Predict what stage of Cirrhosis a patient is in based on patient data in order to create better treatment plans and improve patient outcome

Potential Benefits

- **Early Intervention and Improved Treatment Planning:**
    - Early detection and treatment can slow down or stop progression of cirrhosis, leading to better outcomes for the patients.

- **Tailored Treatment Approach:**
    - Predicting the stage helps doctors create a more tailored treatment plans to the specific needs and severity of the patient's condition,

- **Resource Allocation:**
    - Hospitals can allocate resources more efficiently based Patients with more advanced stages requiring closer monitoring, and specialized care

Stakeholders

| Stakeholders	
Patients with PBC	Individuals diagnosed with primary biliary cirrhosis
Mayo Clinic	Conducting the trial and collecting data, with an interest in improving treatments for PBC
Medical Researchers and Doctors	Professionals involved in designing, conducting, and analyzing the trial
Drug Manufacturers	Companies producing drugs like D-penicillamine, and are interested in effectiveness of drug and market opportunities

Dataset

- Data is from the Mayo Clinic trial on primary biliary cirrhosis (PBC) conducted from 1974 to 1984

- The trial involved 424 PBC patients eligible for a drug trial on D-penicillamine within a ten-year period

- 312 cases participated in the randomized trial, having comprehensive data.

- 112 cases did not take part in the trial but agreed to have measurements and survival data recorded

- Six cases were lost to follow-up after diagnosis, leaving 106 cases along with the 312 trial participants in the dataset
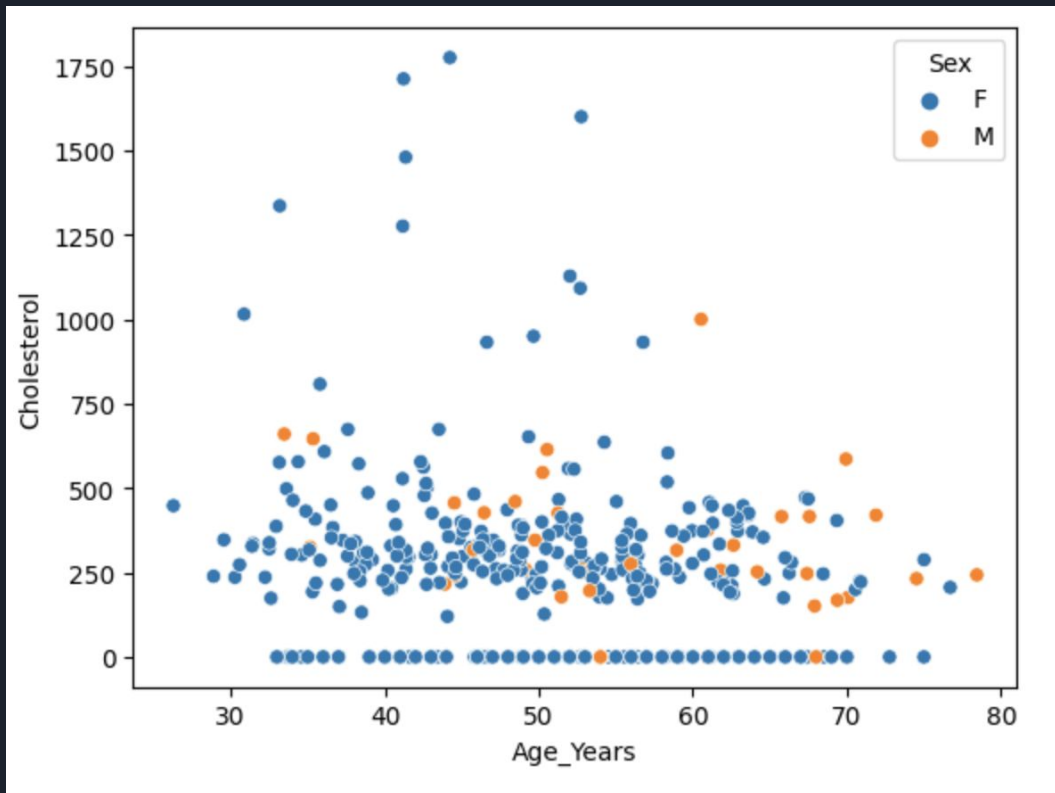
Insights
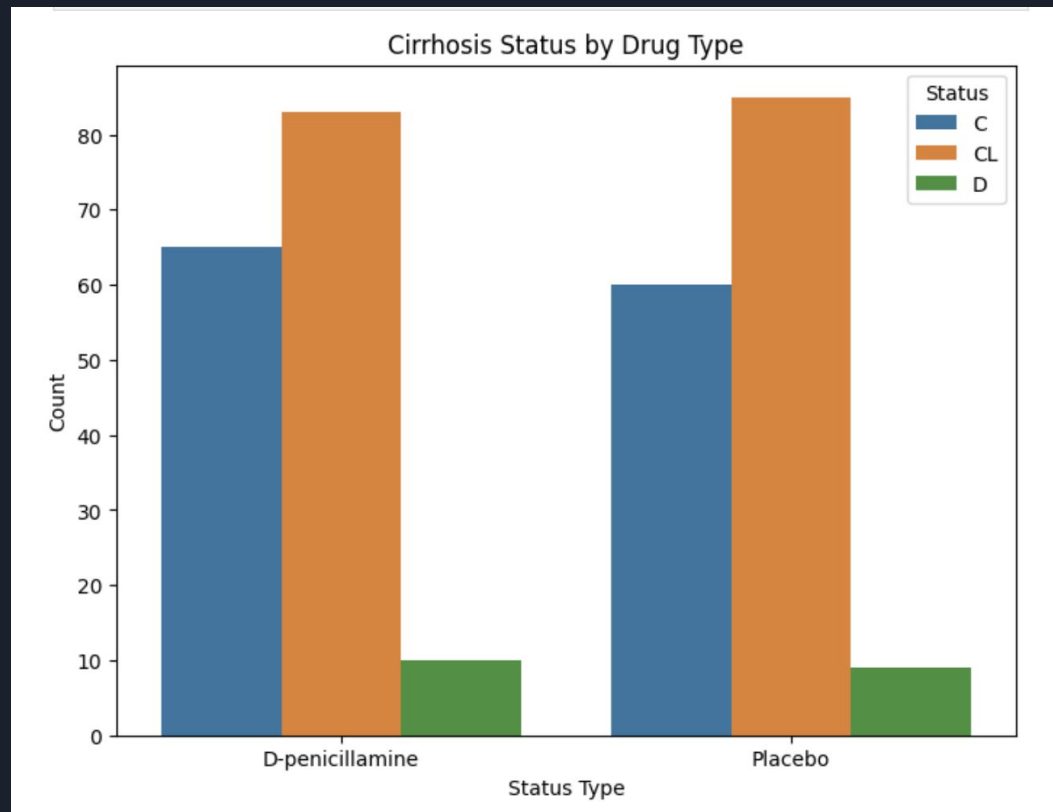
Status: status of the patient C (censored), CL (censored due to liver tx), or D (death)

Cholesterol By Age Grouped by Sex

Target

Our Target Value is the 'Stage' column.

- Initially the Value Counts within Stage were a bit Unbalanced
- Merged Stage 0 to Stage 1

Before

After

Best Performing Model

KNN Model Tuned After Applying PCA


```
-----------------------------------------------------------------
Classification Metrics: Training Data
-----------------------------------------------------------------
              precision    recall    f1-score    support

          0       1.00       1.00       1.00          4
          1       1.00       1.00       1.00         18
          2       1.00       1.00       1.00         76
          3       1.00       1.00       1.00        123
          4       1.00       1.00       1.00        113

   accuracy                              1.00        334
  macro avg       1.00       1.00       1.00        334
weighted avg      1.00       1.00       1.00        334
```
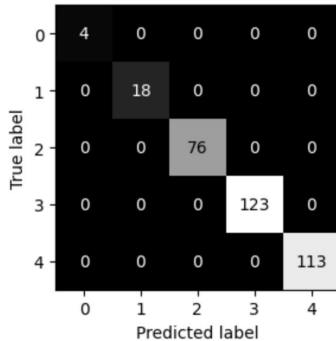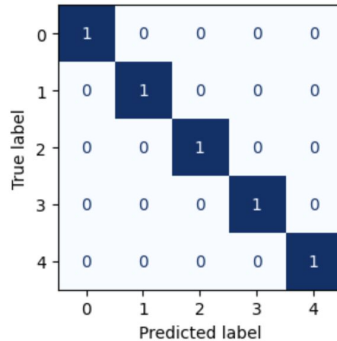
```
-----------------------------------------------------------------
Classification Metrics: Test Data
-----------------------------------------------------------------
              precision    recall    f1-score    support

          0       0.00       0.00       0.00          2
          1       0.00       0.00       0.00          3
          2       0.36       0.25       0.30         16
          3       0.43       0.59       0.50         32
          4       0.66       0.61       0.63         31

   accuracy                              0.50         84
  macro avg       0.29       0.29       0.29         84
weighted avg      0.48       0.50       0.48         84
```
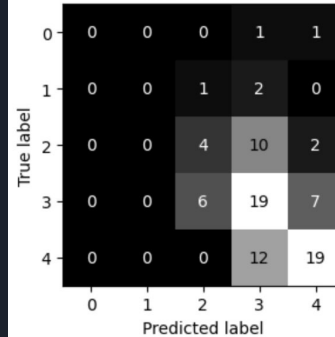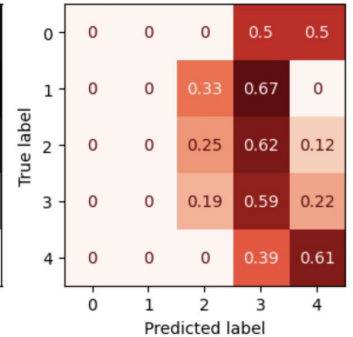
Final Recommendation

- **Although this is a good start it is clear that sampling of target values is severely affecting model performance.**

- **Future Steps:**

    - Applying Resampling methods (Over or Under Sampling) May Help Yield Better Results

    - Deep Learning Models

    - Further Feature Engineering