



# Prediction of Cirrhosis



By: Loraine Rodriguez





# Data Source



Cirrhosis Prediction Dataset:

<https://www.kaggle.com/datasets/fedesoriano/cirrhosis-prediction-dataset>



“The following data contains the information collected from the Mayo Clinic trial in Primary Biliary Cirrhosis (PBC) of the liver conducted between 1974 and 1984.”

Source [KAGGLE](#)

- This dataset consists of 312 rows, and 19 columns.
- The rows represent 424 observations, and the columns represent 19 features and 1 target variable.

# Data Features

Data	Description
<b>N_Days</b>	Number of days between registration and the earlier of death, transplantation, or study analysis time in July 1986
<b>Staus</b>	Status of the patient C (censored), CL (censored due to liver tx), or D (death)
<b>Drug</b>	type of drug D-penicillamine or placebo
<b>Age</b>	Age in [days]
<b>Sex</b>	M (male) or F (female)
<b>Ascites</b>	Presence of ascites N (No) or Y (Yes)
<b>Hepatomegaly</b>	Presence of hepatomegaly N (No) or Y (Yes)
<b>Spiders</b>	Presence of spiders N (No) or Y (Yes)
<b>Edema</b>	Presence of edema N (no edema and no diuretic therapy for edema), S (edema present without diuretics, or edema resolved by diuretics), or Y (edema despite diuretic therapy)

Data	Description
<b>Bilirubin</b>	Serum bilirubin in [mg/dl]
<b>Cholesterol</b>	Serum cholesterol in [mg/dl]
<b>Albumin</b>	Albumin in [gm/dl]
<b>Copper</b>	Urine copper in [ug/day]
<b>SGOT</b>	SGOT in [U/ml]
<b>Triglycerides</b>	Triglycerides in [mg/dl]
<b>Platelets</b>	Platelets per cubic [ml/1000]
<b>Prothrombin</b>	Prothrombin time in seconds [s]
<b>Stage</b>	Stage: histologic stage of disease (1, 2, 3, or 4)

# Goal

---

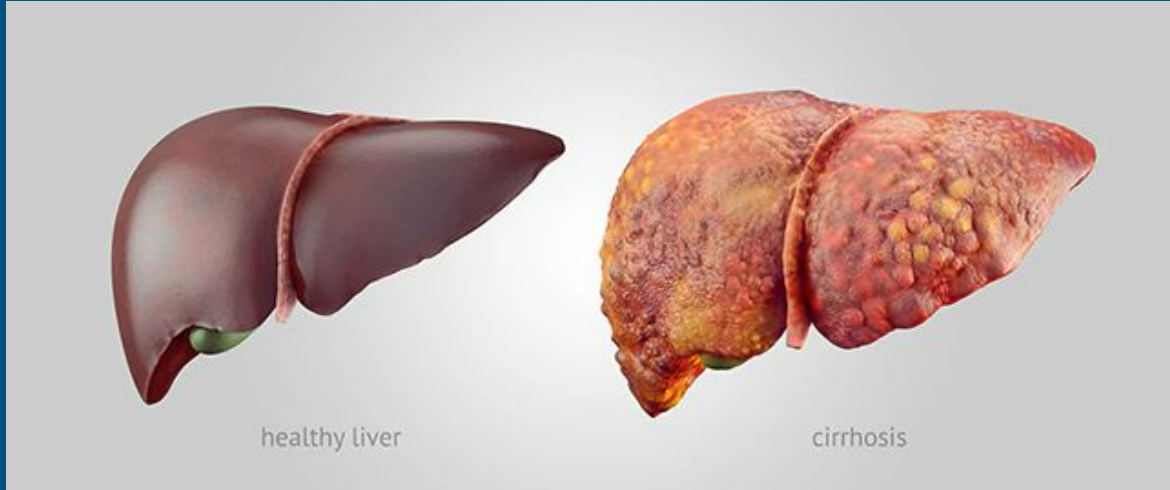
The goal of this project is to evaluate and predict the effects of the trial drug D-penicillamine on features that are known to contribute to Cirrhosis of the liver.

## What is Cirrhosis

Cirrhosis is a late stage of scarring (fibrosis) of the liver caused by many forms of liver diseases and conditions, such as hepatitis and chronic alcoholism.

# Information

---



*"Researchers estimate that about 1 in 400 adults in the United States have cirrhosis." Picture above shows healthy liver and cirrhosis liver with scar tissue.*

- Source - [NIH](#)

# Information on data to be analyzed

---

- **Bilirubin.** Healthcare providers often give **bilirubin tests** to check on the health of your liver. The bilirubin test is one of a comprehensive panel of liver function tests that measure different liver products in your blood. If these values are high or low, they might indicate that your liver is struggling in some way. Source [Cleveland clinic](#)
- **Cholesterol.** High cholesterol can turn fatty liver disease (steatosis) into a more serious and sometimes fatal condition known as nonalcoholic steatohepatitis (NASH). When fatty liver disease turns into NASH, it can lead to other liver problems including: **Liver inflammation Scarring** (cirrhosis)[WebMD](#)

# Methods

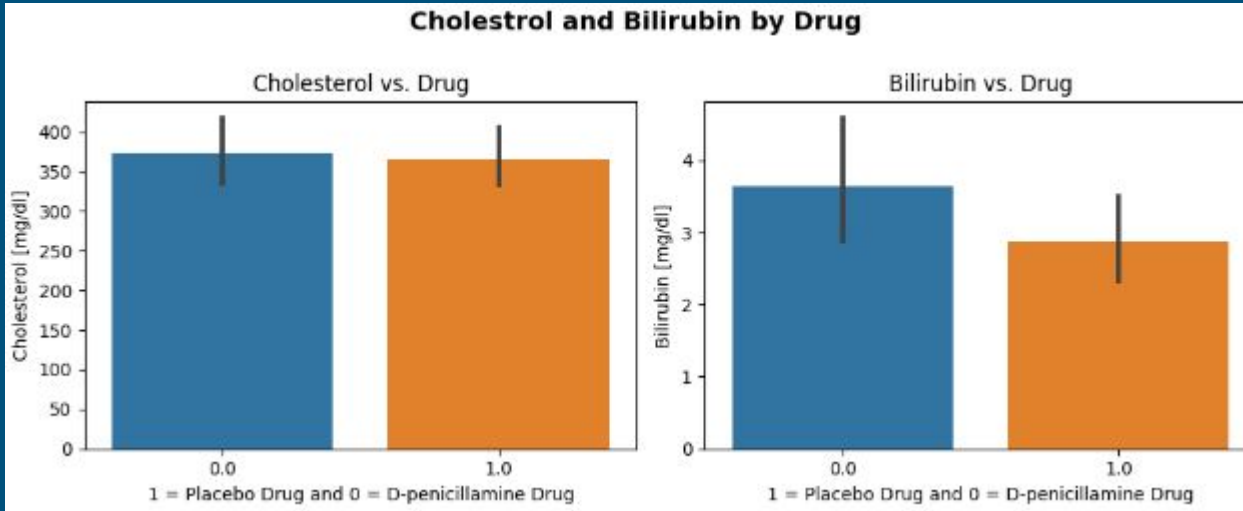
With the data provided, we were able to create various different machine learning models.

Models for review:

- Visual #1 -Side by Side Barplots
  - 1. Cholesterol vs 2 Drugs (Placebo vs D-penicillamine)
  - 2. Bilirubin vs 2 Drugs (Placebo vs D-penicillamine)
- Visual #2 - Barplot (Stages of Cirrhosis vs. Cholesterol Level)
- PCA Model



# Visual #1- Barplot (Stages of Cholesterol vs 2 Drugs)



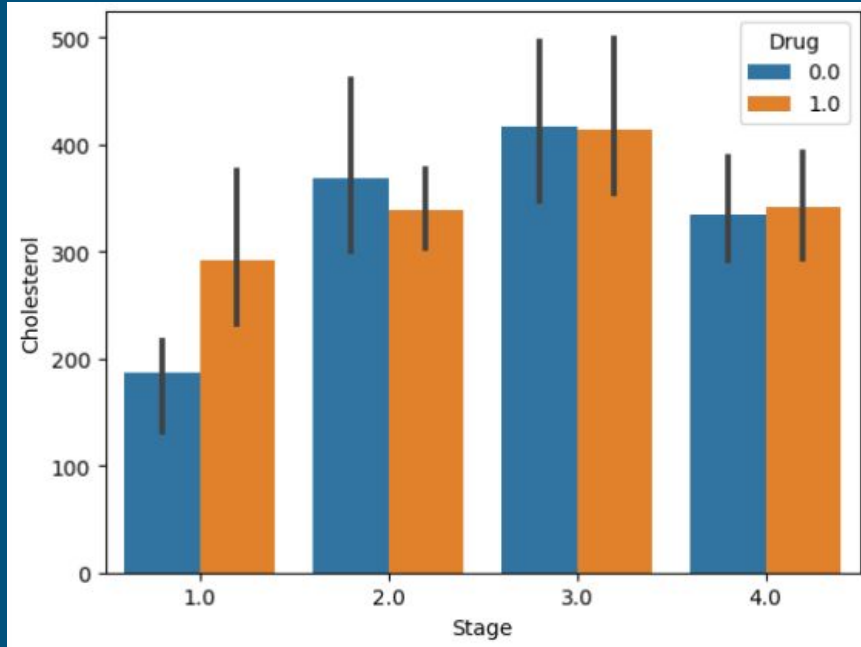
## Cholesterol

1. High Cholesterol is over 200
2. No big significance difference between 2 drugs when comparing Cholesterol
3. D-penicillamine didn't affect the cholesterol levels

## Bilirubin

1. The approximate normal range of bilirubin in the blood is less than 1.0 milligram per deciliter (mg/dL).
2. Slight decrease in Bilirubin levels for those that took D-penicillamine
3. D-penicillamine does slightly lower Bilirubin levels

## Visual #2 - Barplot (Stages vs. Cholesterol Level)



- Blue = (0.0) Placebo
- Orange = (1.0) D-penicillamine

- With exception to Stage 1, other stages had cholesterol levels over 200.
- In Stage 1 the D-penicillamine has higher level of Cholesterol
- In Stage 3 and Stage 4 performed neck and neck, so this tells me D-penicillamine had no affect.
- In Stage 2 D-penicillamine had a slight improvement on the cholesterol levels.
- D-penicillamine does not seem to affect the Cholesterol level by Stages.

# Best Model: PCA

---

- This model lowered the Type II (False Negatives) which are the best metrics to concentrate on when dealing with a medical disease.
- We want to have the least amount possible of False negatives because this means patients would be told they don't have the disease and they really do and they will not get treated.



# Final Recommendations



- The data obtained indicates that the drug D-penicillamine was slightly correlated with lowered Bilirubin levels and was not effective on Cholesterol levels. Based off the overall data, the drug D-penicillamine trial didn't do a significant change in the data/features gathered from the patients that are known to contribute to cirrhosis.
- 