For my last project, I wanted to do a predictive analysis on something that I am not too familiar with but something that I am interested in. The dataset is pretty straightforward. It provides different factors that can affect the liver. It also provides information on the patients, such as age and gender of the patient. Liver disease is pretty common, especially when there is an increase in alcohol consumption, contaminated foods, or the use of drugs. In the exploratory analysis phase, I did see there was some null values in the data, and I have decided to remove them so they would not affect the end results.

The dataset was what I thought it was. The benefits machine learning has in the health industry interests me as I think that it can really help a lot of people. They say that without health, what do you have? Health is very important to maintain and keep up with. Being able to easily detect if someone has a disease or not can really help get them started on the right path. Looking for the signs of the disease before it even happens can help the patients avoid the diseases and know what to look for.

I have not fully flushed out on all the predictive analysis that I am going to do. I am thinking of doing a logistic regression approach. There is some categorical data in the dataset, such as gender of the patient. The overall goal of the project would be able to predict if a patient would have liver disease or not. One of the variables in the dataset, shown as dataset, is if the patient currently has liver disease or not. I will need to train and test my data to make sure everything is working correctly, and I am hoping for a high score of accuracy.

So far, all I have really done is the exploratory data analysis. I have found that there were some null values and I have removed them. In the dataset there are more male patients than there are female patients, because of this the dataset is skewing more than males are more likely to get or have some type of liver disease. I have also found that most would have liver disease in their mid-forties. Currently, I am looking at the different distribution levels of those who have a liver disease with some of the continuous data.

There are not many challenges that I have been facing, other than time. There are only the two weeks left in class, so I need to push myself to get it done. The last weeks of class, I feel are the hardest because you know you are so close and sometimes you start slowing down. Another challenge would be that I am not too familiar with liver disease, so I am not that proficient on the different terminology. I will definitely be relying on Google to help in this area.