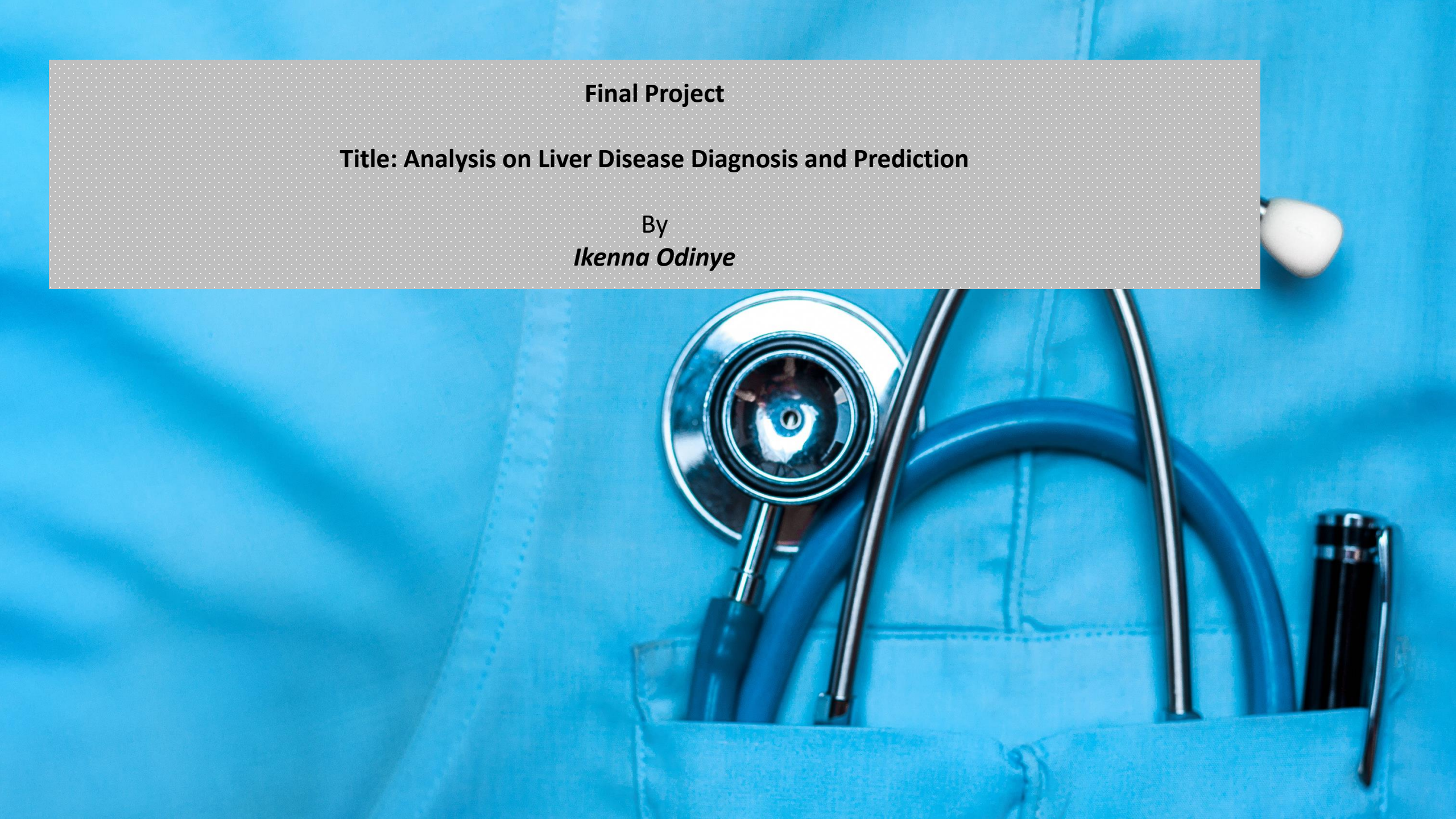
Final Project

Title: Analysis on Liver Disease Diagnosis and Prediction

By
Ikenna Odinye



Background

The liver is a vital visceral organ that plays a crucial role in various metabolic processes, including digestion, detoxification, regulation of blood sugar levels, and production of important proteins. Liver diseases can range from mild to life-threatening illnesses.

Most of the common liver diseases are:

- **Fatty liver disease – associated with excessive alcohol, diabetes and obesity**
- **Hepatitis : inflammation caused by viruses**
- **Cirrhosis : advanced scarring of the liver tissue caused by long-term damage, often due to chronic alcoholism, viral hepatitis, or other factors**
- **Primary liver cancer (hepatocellular carcinoma) originates in the liver cells and is often associated with underlying liver disease.**

Problem Analysis and Motivation

- Problem: Accurate and early diagnosis of liver disease is critical for effective treatment and patient well-being.

- Motivation: Improving diagnosis can lead to better healthcare outcomes, reduced medical costs, and enhanced patient care.

Diagnosis

Blood tests or Liver function tests which is the basis of this project in the major way of diagnosing liver disease measures enzymes, bilirubin, and proteins to assess how well the liver is functioning.

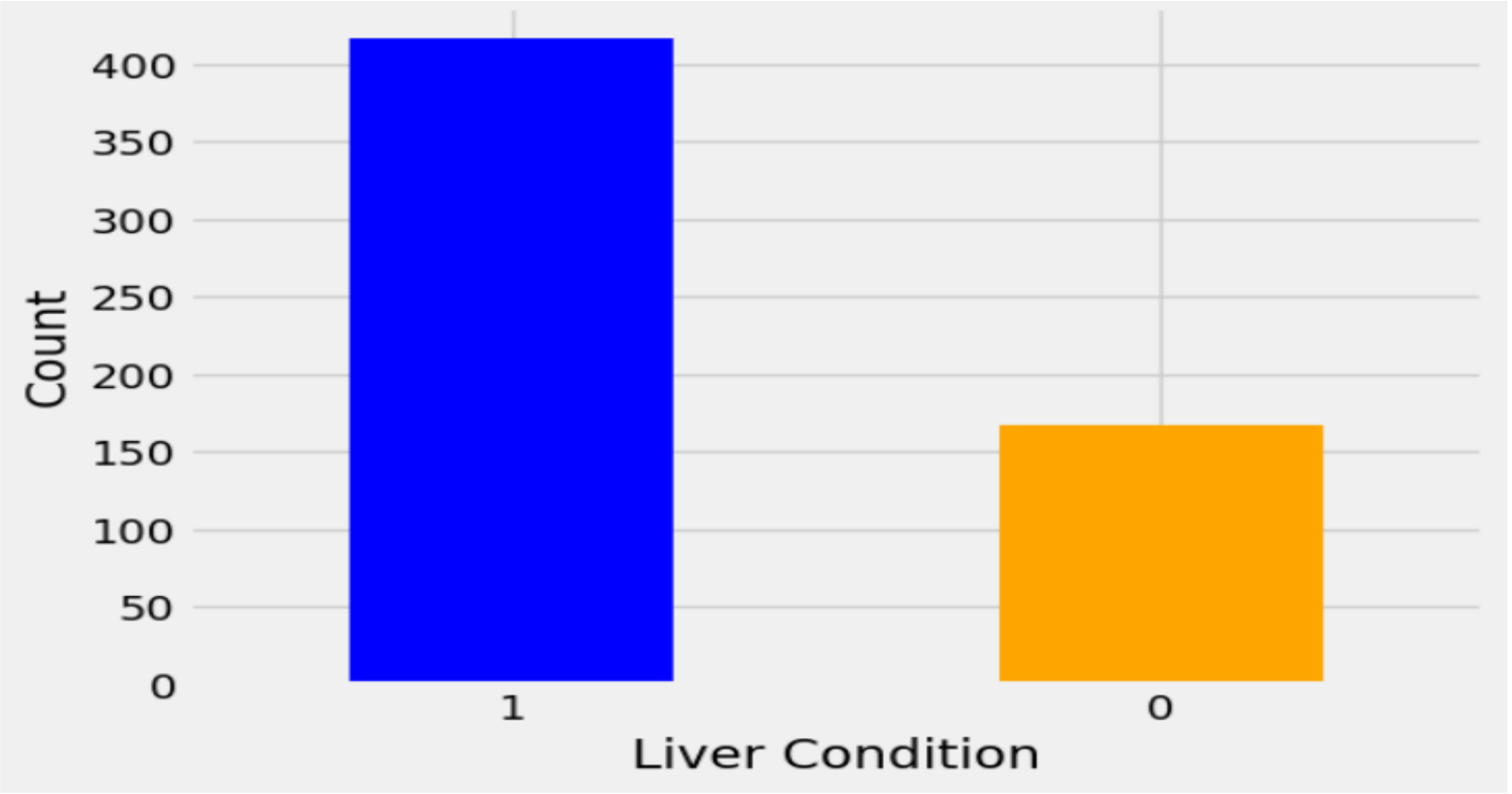
Project goals & objectives

Improving Healthcare service delivery - Contribute to enhanced patient care, reduced medical costs, and improved healthcare outcomes through reliable diagnosis.

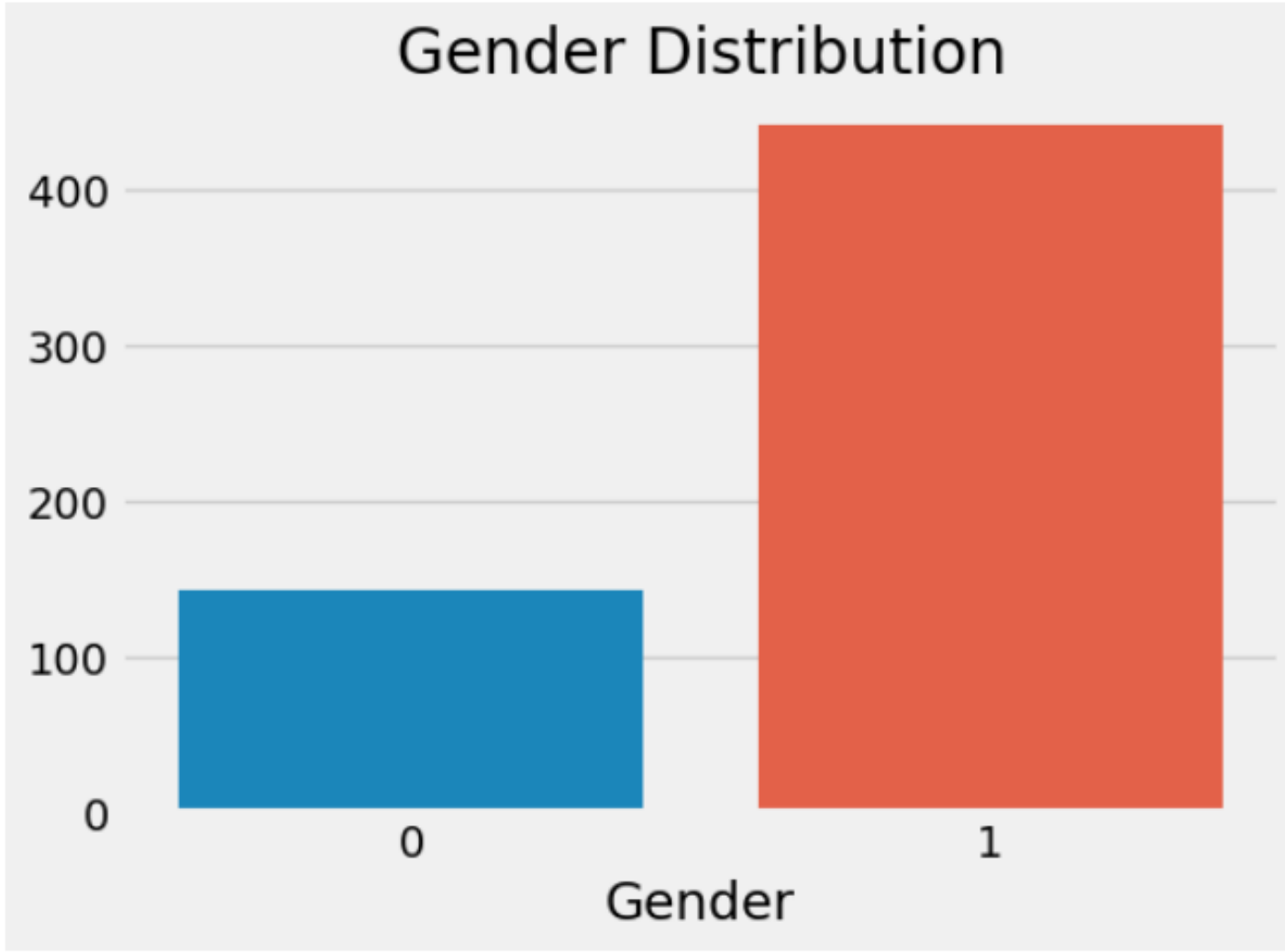
Early Detection - Identify features that contribute to early detection of liver disease, enabling timely medical interventions.

Accurate Diagnosis - Develop predictive models to accurately diagnose liver disease based on patient data.

Dataset Description

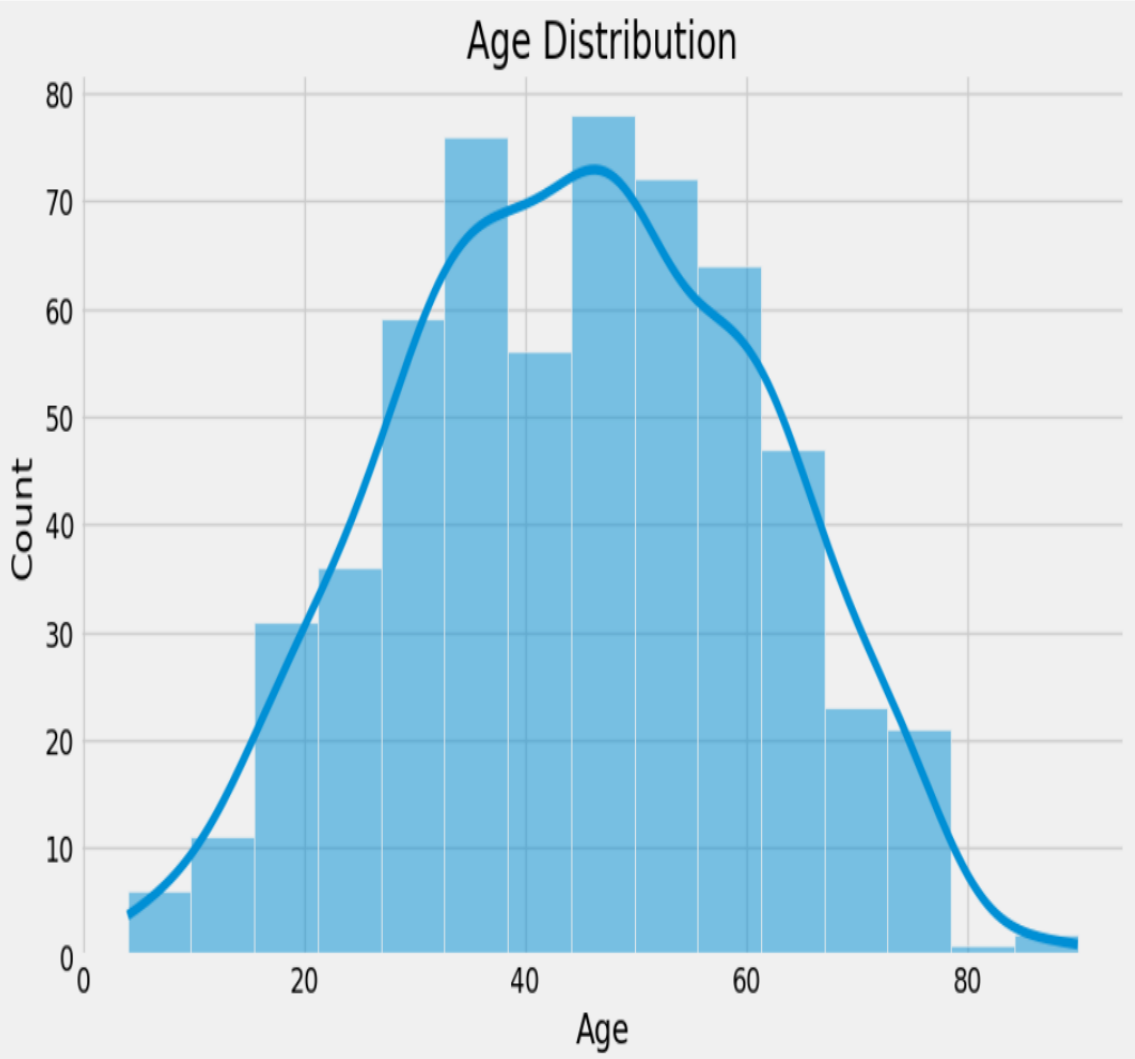



```
Number of patients diagnosed with liver disease:  416
Number of patients with normal liver :  167
```



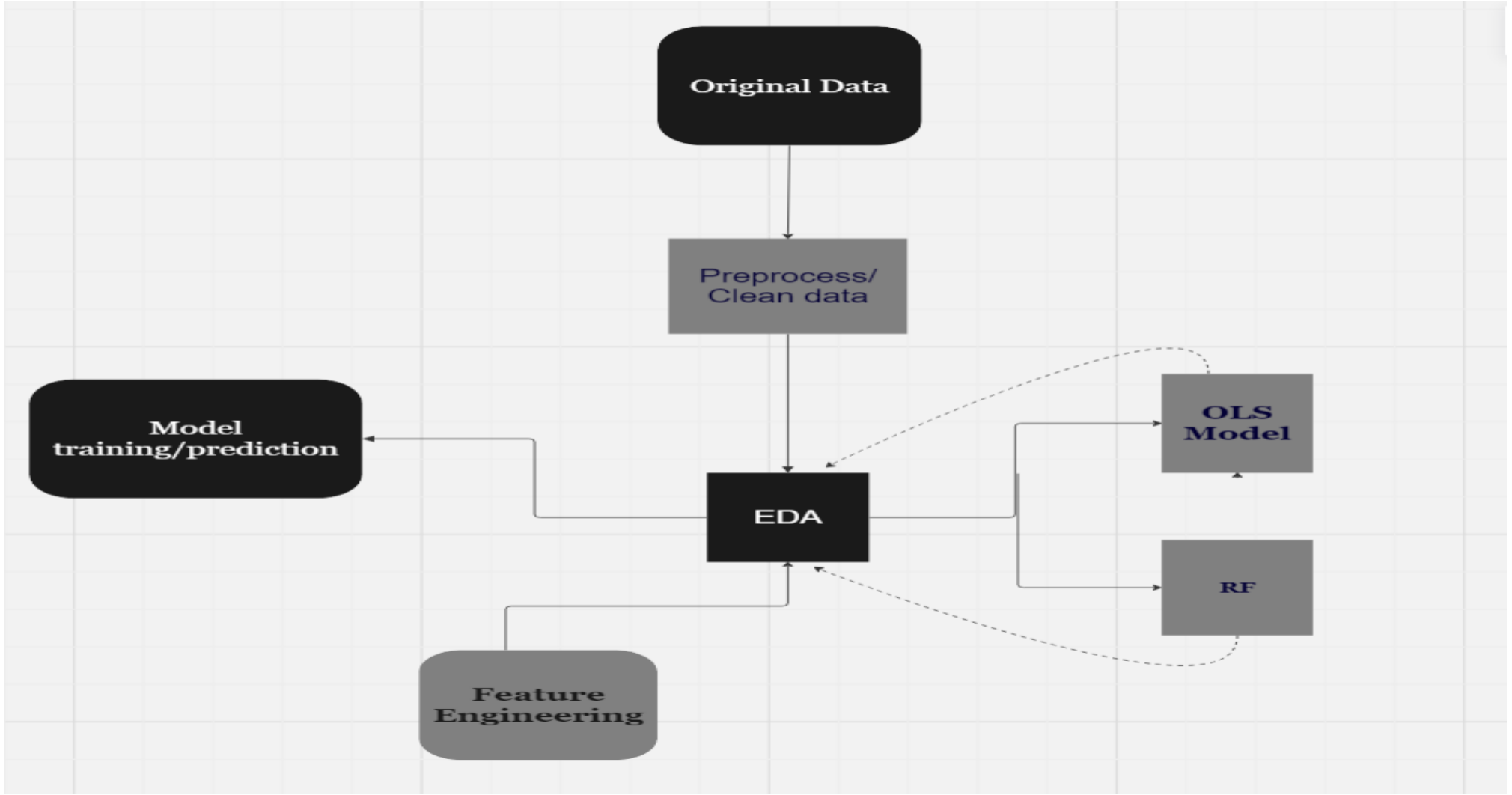

```
Number of patients that are male:  441
Number of patients that are female:  142
```

Age distribution



	LiverCondition	Gender	Mean_Age
0	0	0	42.740000
1	0	1	40.598291
2	1	0	43.347826
3	1	1	46.950617

Approach


Original Data
Preprocess/
Clean data
Model
training/prediction
EDA
OLS
Model
RF
Feature
Engineering

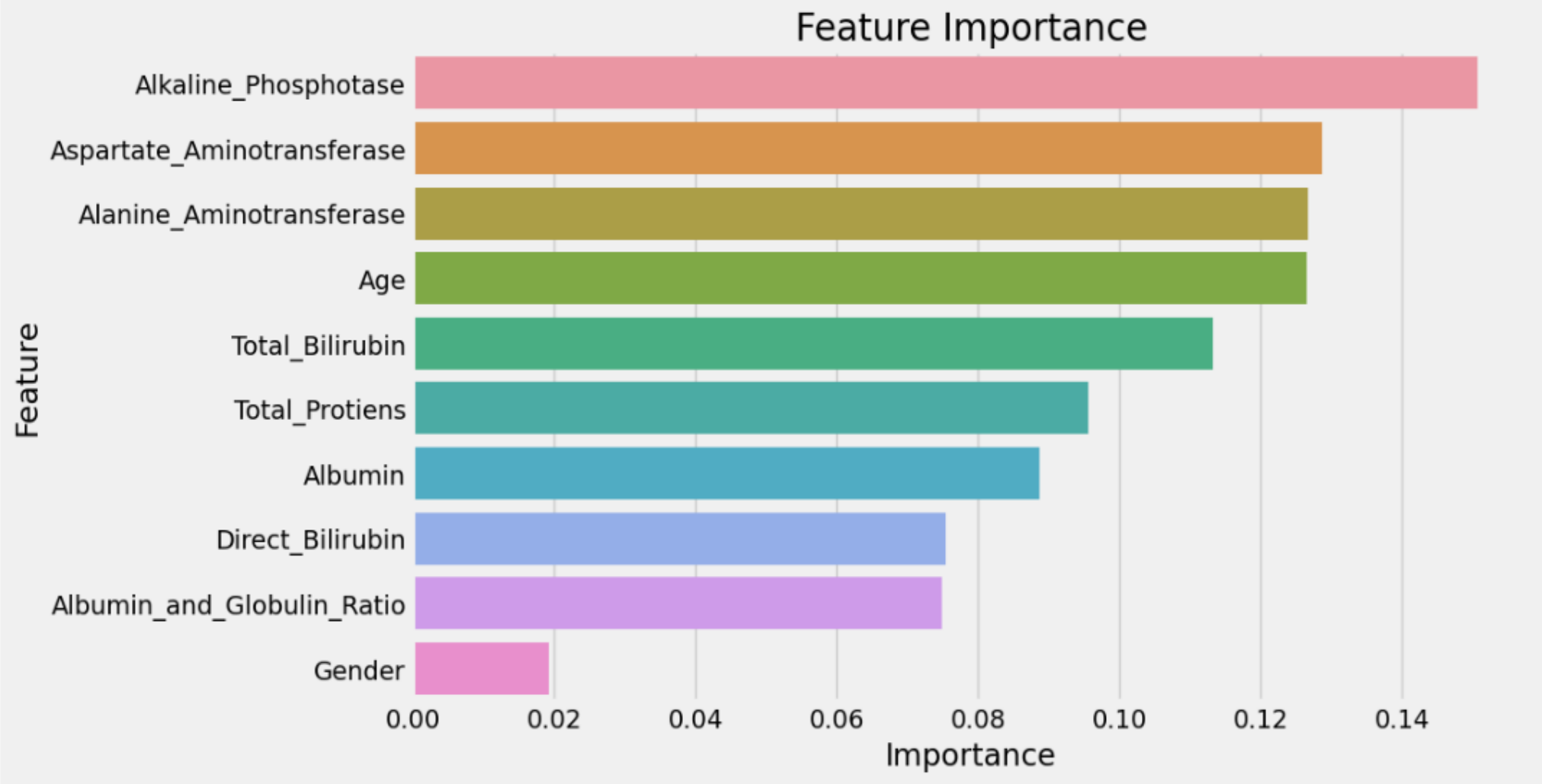
OLS Regression Results

Dep. Variable:	LiverCondition	R-squared:	0.120
Model:	OLS	Adj. R-squared:	0.105
Method:	Least Squares	F-statistic:	7.812
Date:	Sat, 12 Aug 2023	Prob (F-statistic):	8.38e-12
Time:	12:14:18	Log-Likelihood:	-327.11
No. Observations:	583	AIC:	676.2
Df Residuals:	572	BIC:	724.3
Df Model:	10		
Covariance Type:	nonrobust		

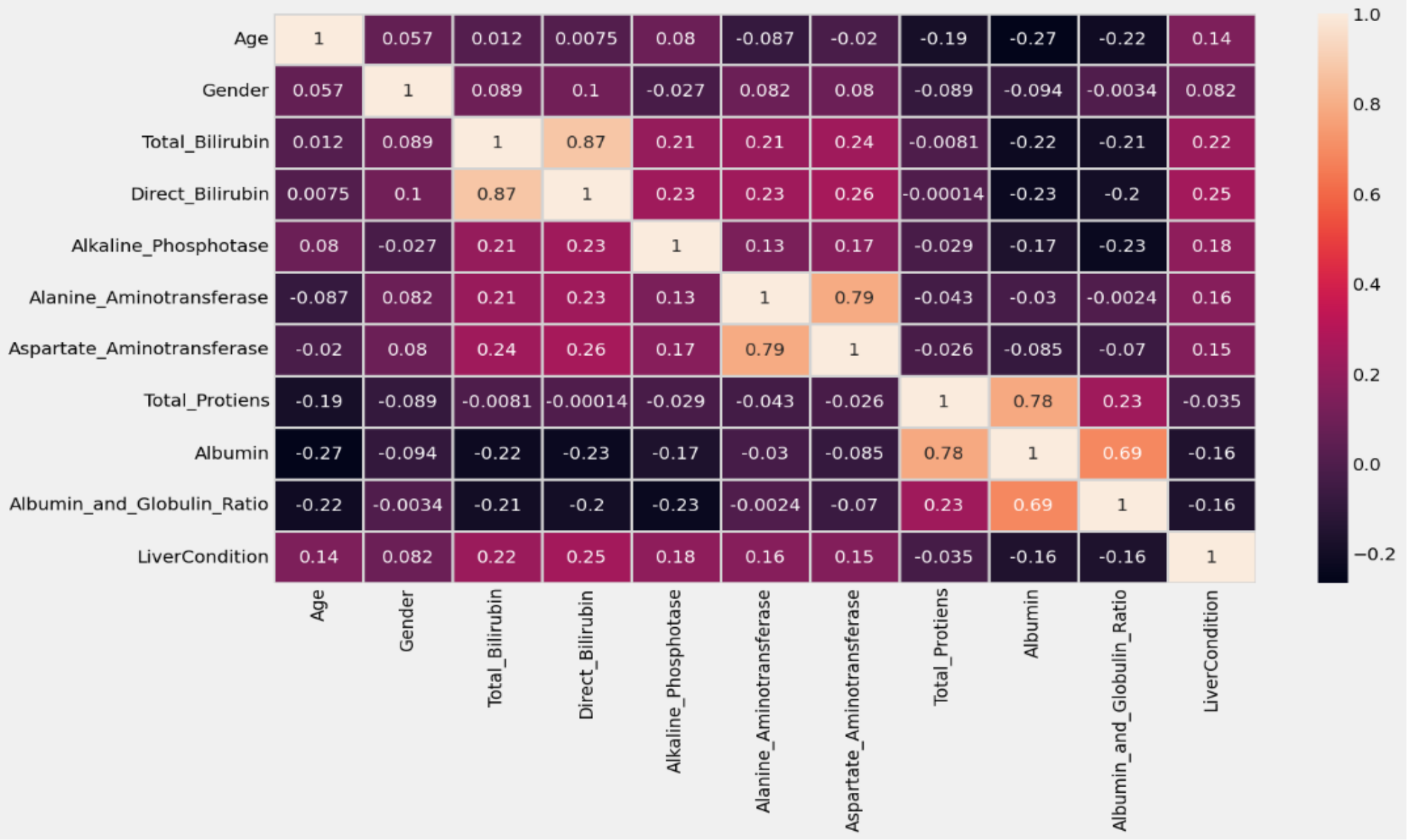
	coef	std err	t	P>|t|	[0.025	0.975]
const	0.3088	0.162	1.903	0.058	-0.010	0.628
Age	0.0032	0.001	2.804	0.005	0.001	0.005
Gender	0.0510	0.042	1.214	0.225	-0.032	0.133
Total_Bilirubin	0.0001	0.006	0.023	0.982	-0.011	0.012
Direct_Bilirubin	0.0225	0.013	1.677	0.094	-0.004	0.049
Alkaline_Phosphotase	0.0002	7.74e-05	2.517	0.012	4.28e-05	0.000
Alanine_Aminotransferase	0.0004	0.000	2.367	0.018	6.57e-05	0.001
Aspartate_Aminotransferase	-6.355e-05	0.000	-0.617	0.537	-0.000	0.000
Total_Protiens	0.0760	0.039	1.972	0.049	0.000	0.152
Albumin	-0.1391	0.071	-1.959	0.051	-0.279	0.000
Albumin_and_Globulin_Ratio	0.0545	0.107	0.509	0.611	-0.156	0.265

Omnibus:	343.370	Durbin-Watson:	1.943
Prob(Omnibus):	0.000	Jarque-Bera (JB):	79.234
Skew:	-0.689	Prob(JB):	6.23e-18
Kurtosis:	1.833	Cond. No.	4.51e+03

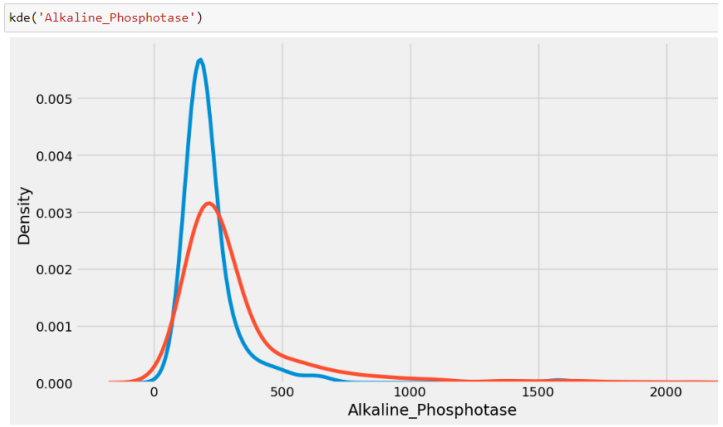
RandomForest

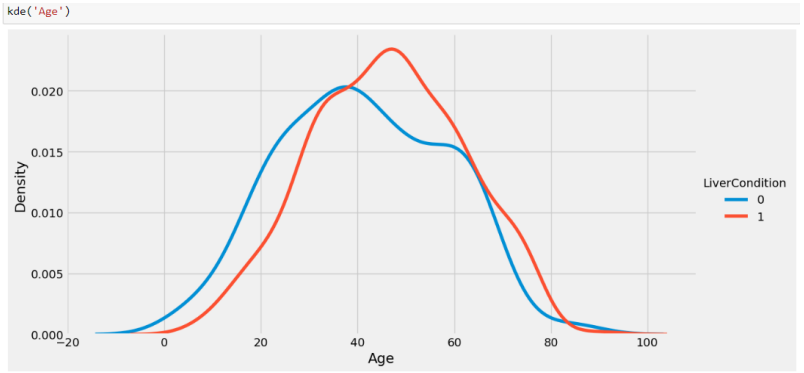


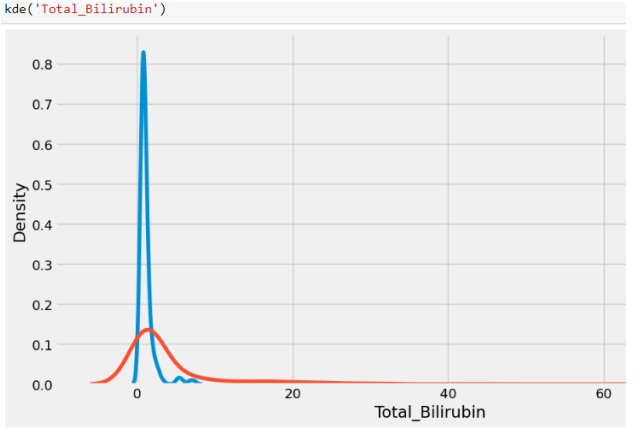
Heatmap

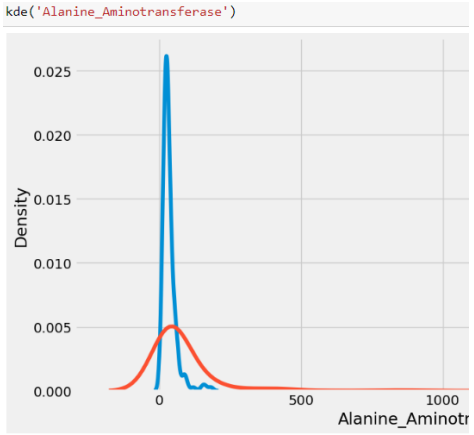

Age 1 0.057 0.012 0.0075 0.08 -0.087 -0.02 -0.19 -0.27 -0.22 0.14
Gender 0.057 1 0.089 0.1 -0.027 0.082 0.08 -0.089 -0.094 -0.0034 0.082
Total_Bilirubin 0.012 0.089 1 0.87 0.21 0.21 0.24 -0.0081 -0.22 -0.21 0.22
Direct_Bilirubin 0.0075 0.1 0.87 1 0.23 0.23 0.26 -0.00014 -0.23 -0.2 0.25
Alkaline_Phosphotase 0.08 -0.027 0.21 0.23 1 0.13 0.17 -0.029 -0.17 -0.23 0.18
Alanine_Aminotransferase -0.087 0.082 0.21 0.23 0.13 1 0.79 -0.043 -0.03 -0.0024 0.16
Aspartate_Aminotransferase -0.02 0.08 0.24 0.26 0.17 0.79 1 -0.026 -0.085 -0.07 0.15
Total_Protiens -0.19 -0.089 -0.0081 -0.00014 -0.029 -0.043 -0.026 1 0.78 0.23 -0.035
Albumin -0.27 -0.094 -0.22 -0.23 -0.17 -0.03 -0.085 0.78 1 0.69 -0.16
Albumin_and_Globulin_Ratio -0.22 -0.0034 -0.21 -0.2 -0.23 -0.0024 -0.07 0.23 0.69 1 -0.16
LiverCondition 0.14 0.082 0.22 0.25 0.18 0.16 0.15 -0.035 -0.16 -0.16 1
Age
Gender
Total_Bilirubin
Direct_Bilirubin
Alkaline_Phosphotase
Alanine_Aminotransferase
Aspartate_Aminotransferase
Total_Protiens
Albumin
Albumin_and_Globulin_Ratio
LiverCondition
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

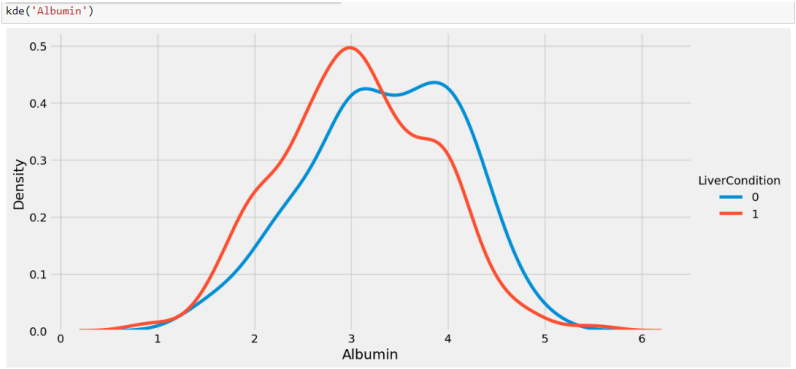
Kernel density

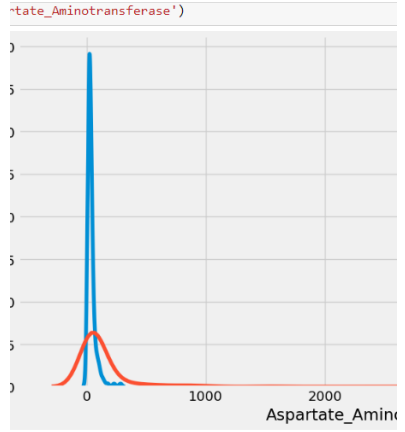




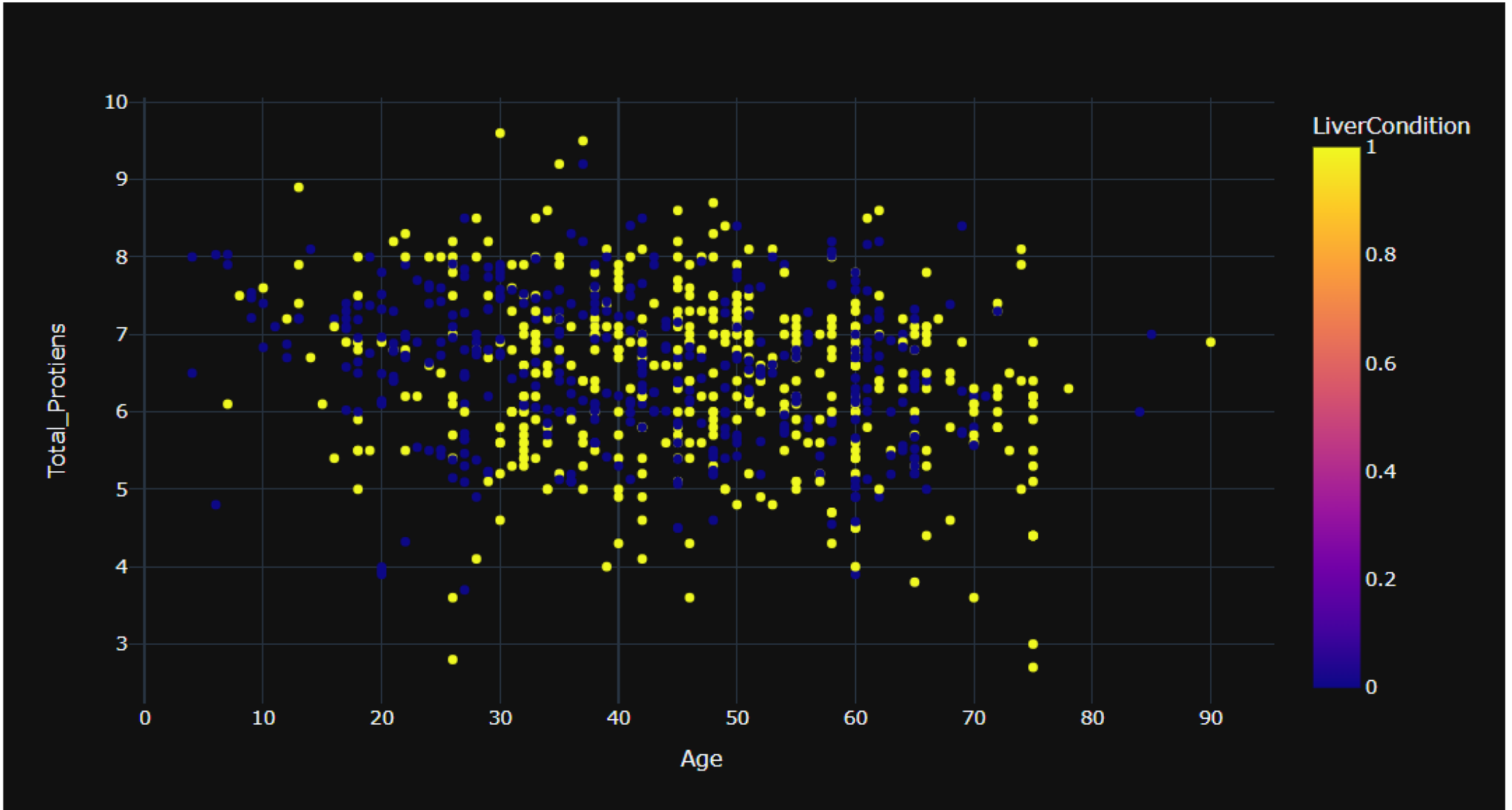


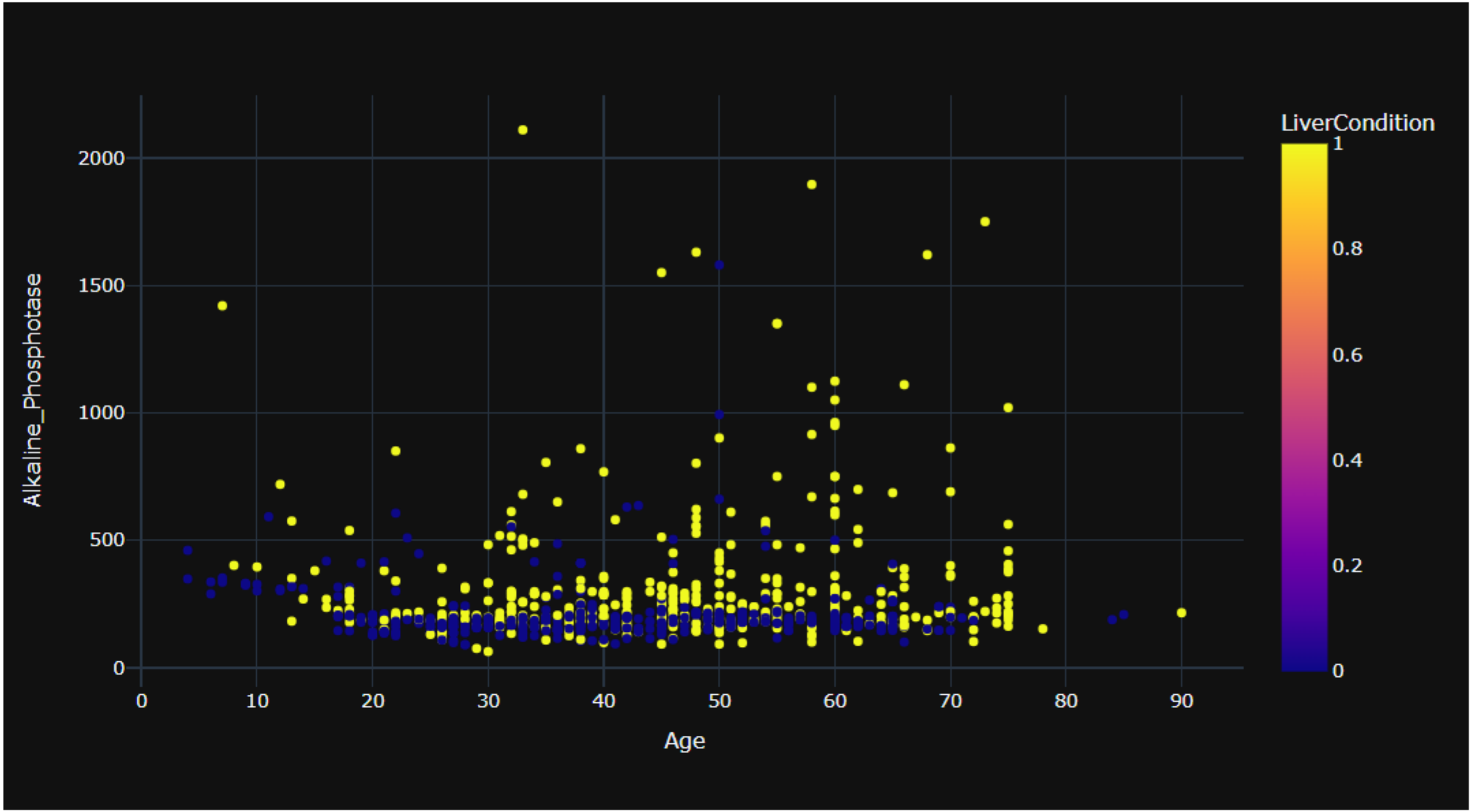


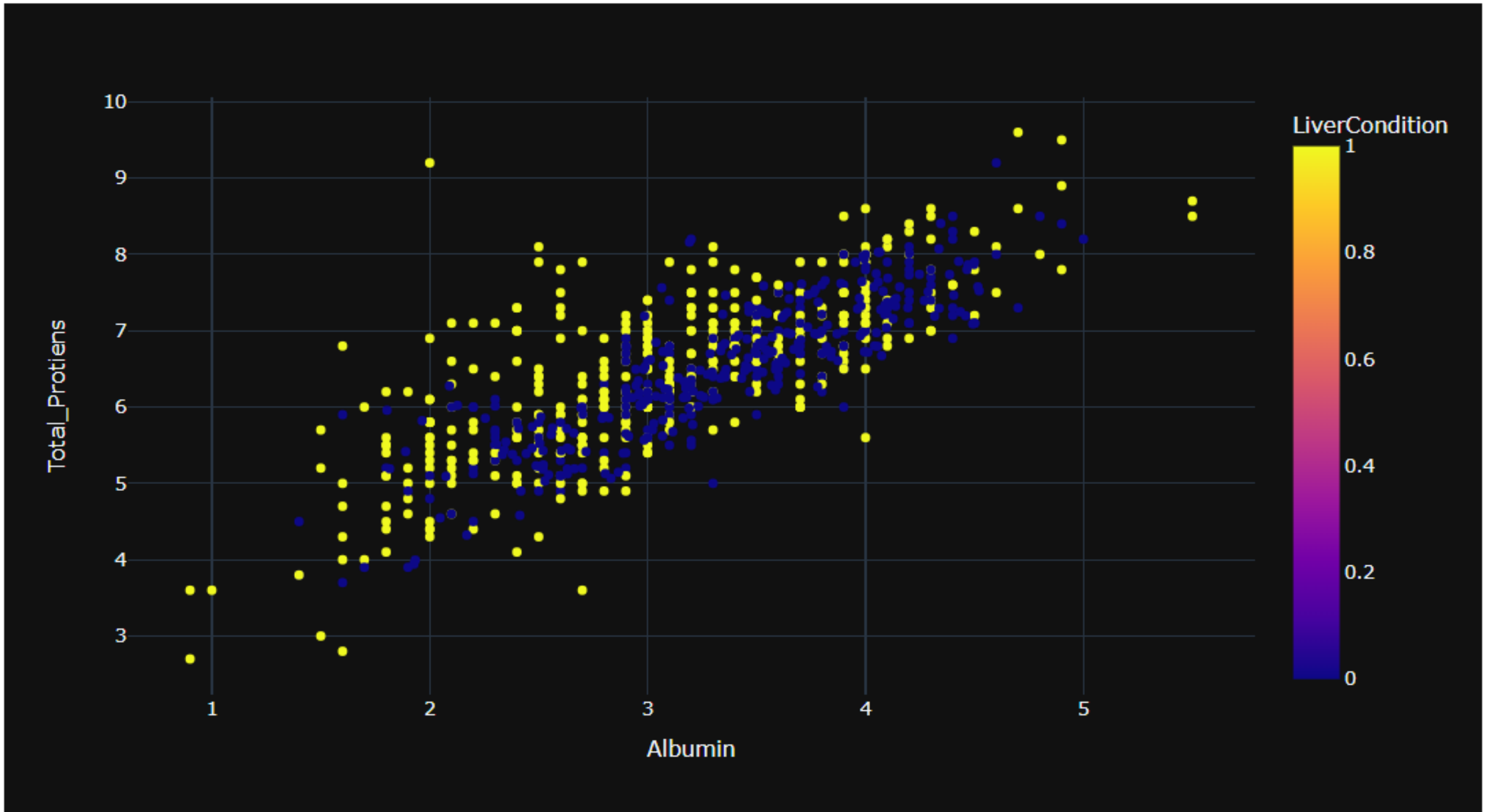


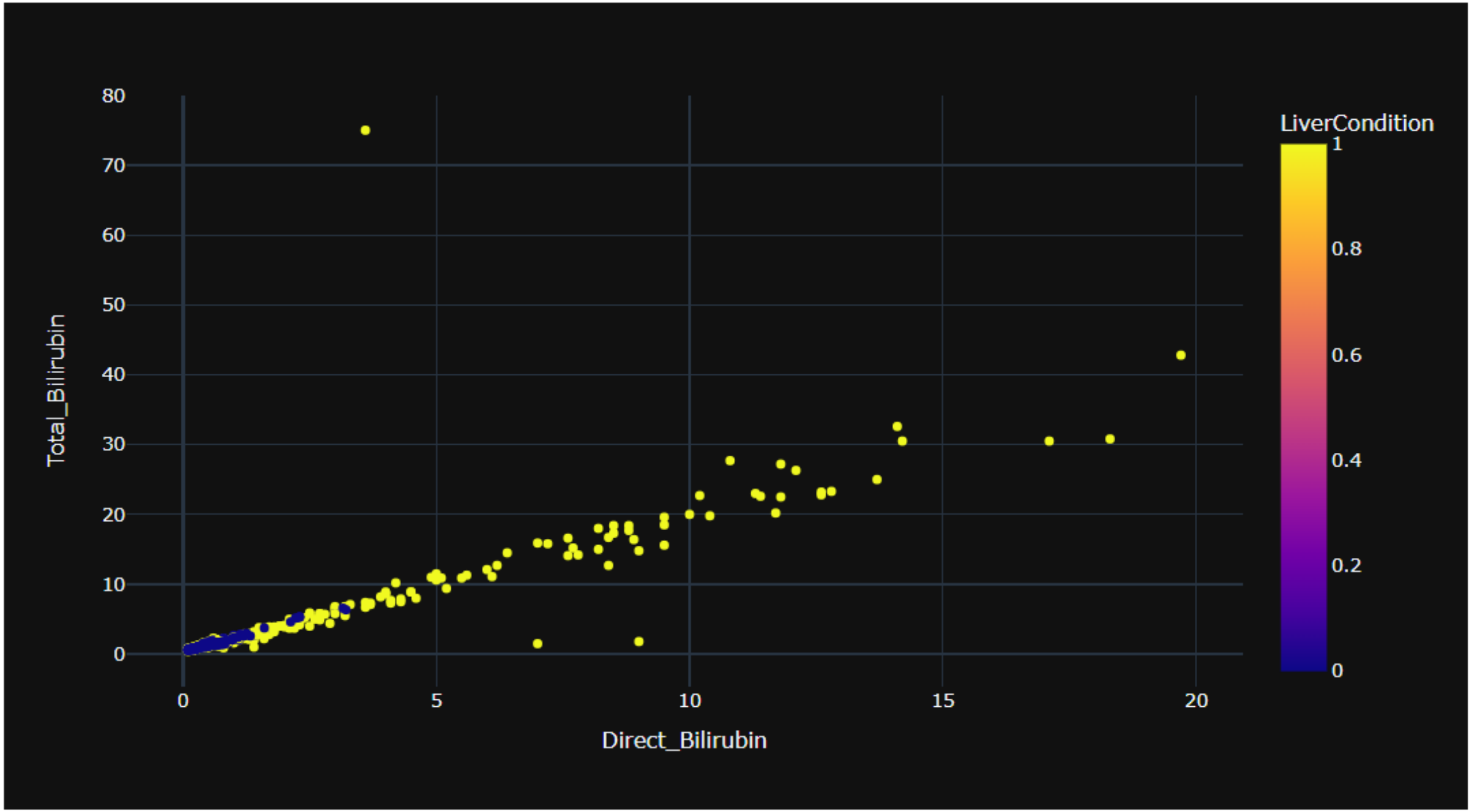


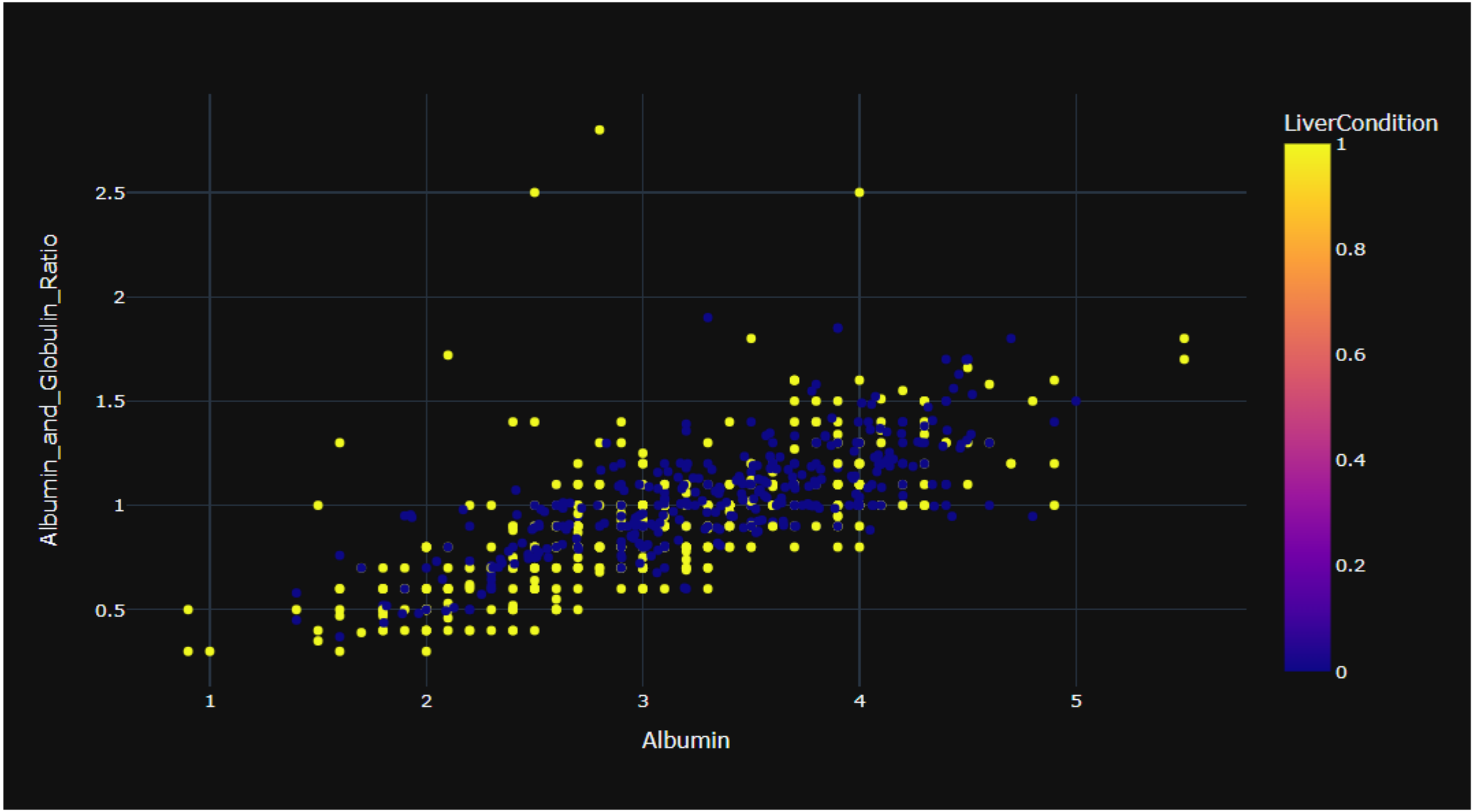
```
scatter('Age', 'Total_Protiens')
```

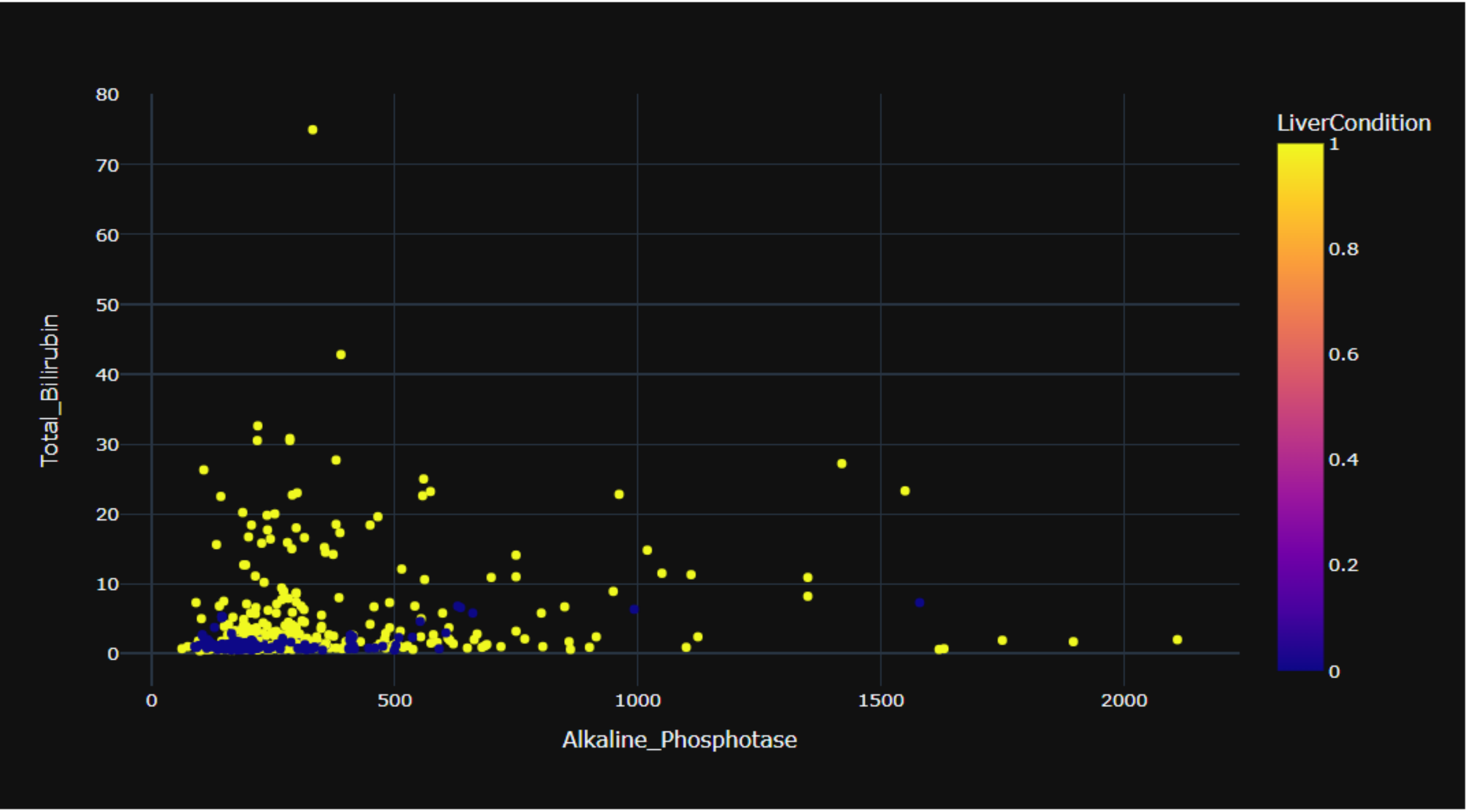



LiverCondition
1
0.8
0.6
0.4
0.2
0
Alkaline_Phosphotase
2000
1500
1000
500
0
0
10
20
30
40
50
60
70
80
90
Age

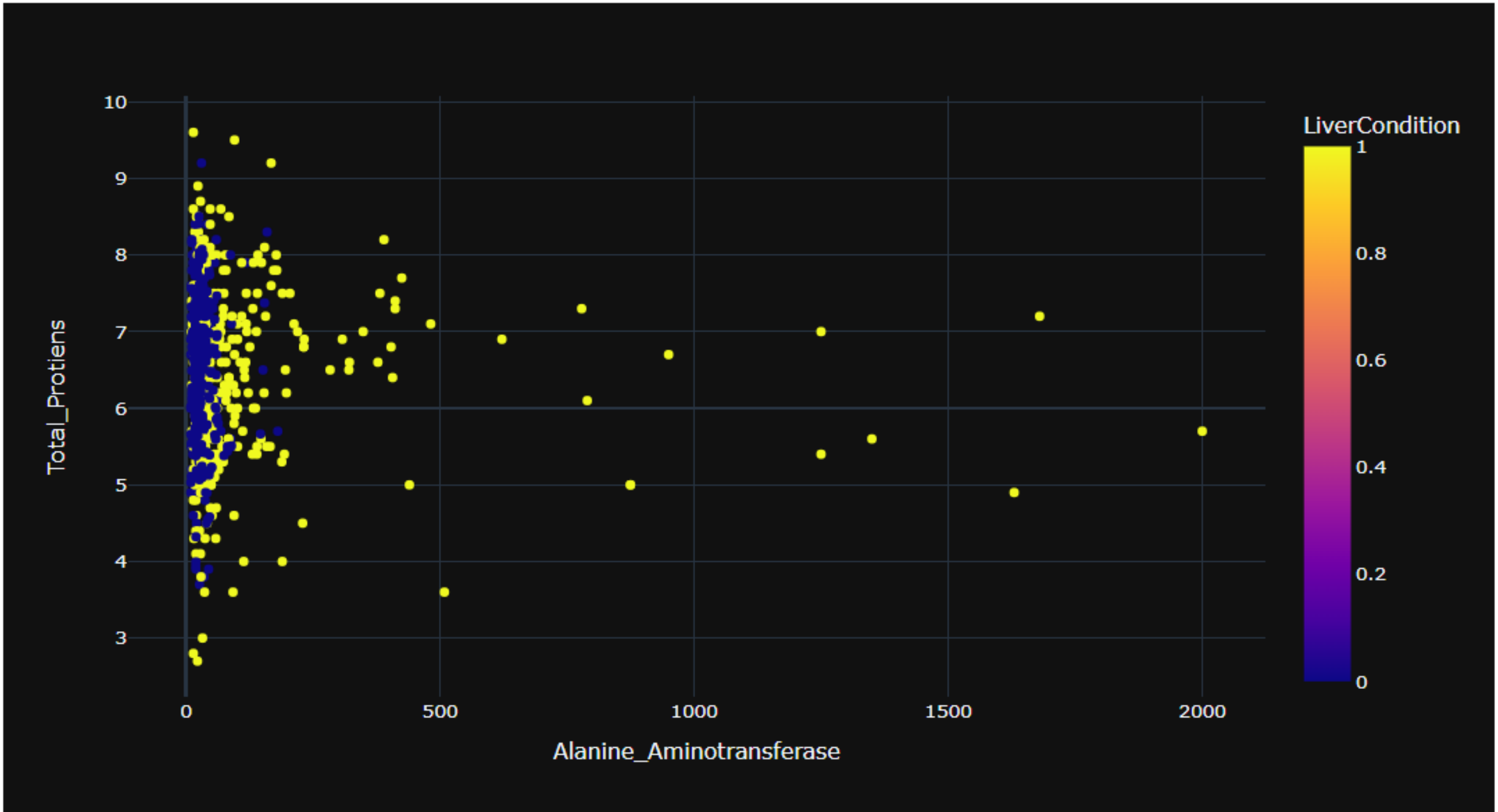

LiverCondition
1
0.8
0.6
0.4
0.2
0
Total_Protiens
10
9
8
7
6
5
4
3
Albumin
1
2
3
4
5


LiverCondition
1
0.8
0.6
0.4
0.2
0
Total_Bilirubin
80
70
60
50
40
30
20
10
0
Direct_Bilirubin
0
5
10
15
20


LiverCondition
1
0.8
0.6
0.4
0.2
0
Albumin_and_Globulin_Ratio
2.5
2
1.5
1
0.5
1
2
3
4
5
Albumin


80
70
60
50
40
30
20
10
0
Total_Bilirubin
0
500
1000
1500
2000
Alkaline_Phosphotase
LiverCondition
1
0.8
0.6
0.4
0.2
0

```
scatter('Alanine_Aminotransferase', 'Total_Protiens')
```



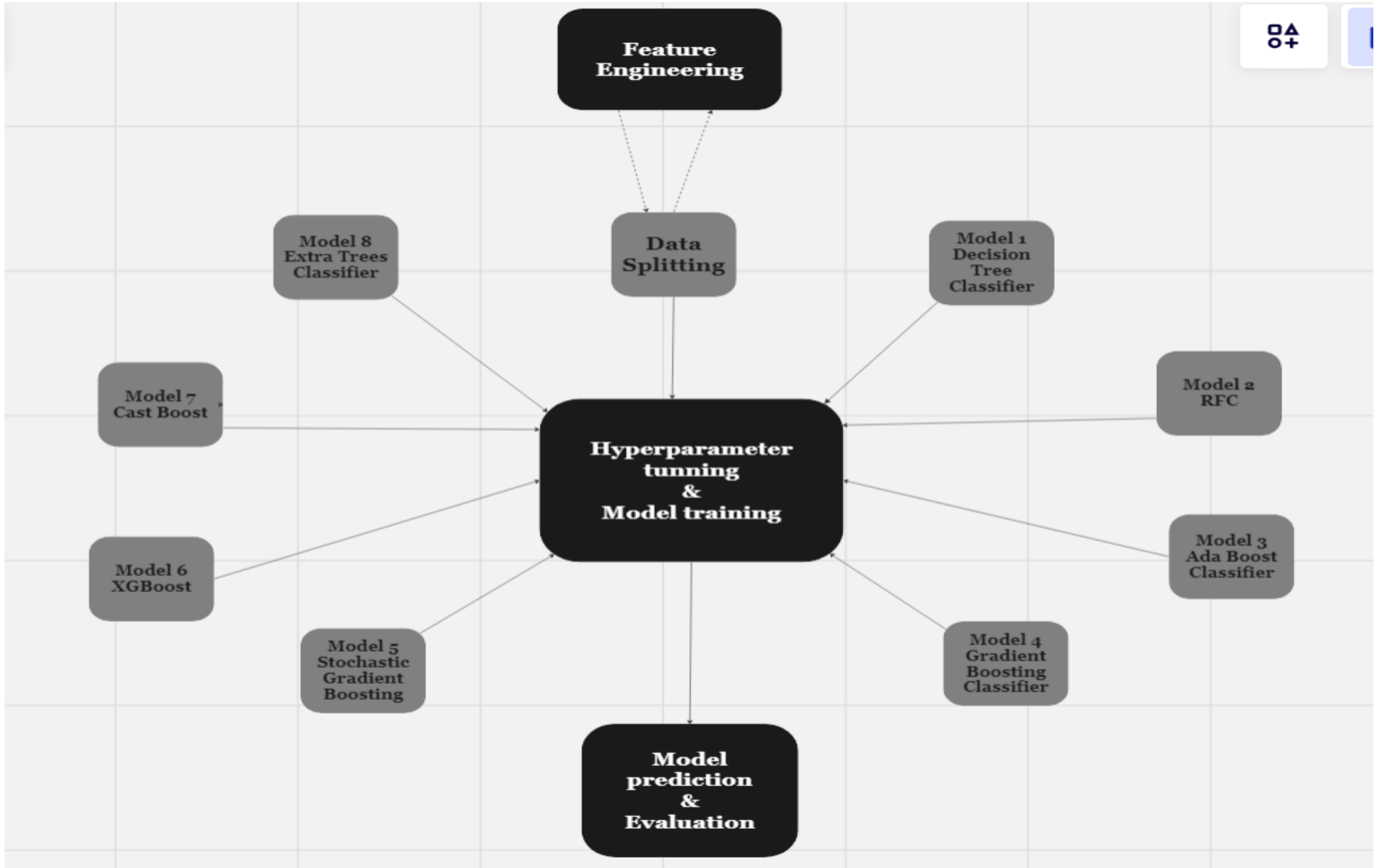
T-test Results

S/N	Features	Liver conditionCategory	Result(P-value)	Conclusion
1	Alkaline_Phosphotase	Normal + Diseased	**7.03e-06 (<0.05)**	Reject null: Significant difference
2	Alanine_Aminotransfera	Normal + Diseased	**7.37e-05 (<0.05)**	Reject null: Significant difference
3	Aspartate_Aminotransfe	Normal + Diseased	**7.80e-08 (<0.05)**	Reject null: Significant difference
4	Total_Bilirubin	Normal + Diseased	**9.07e-05 (<0.05)**	Reject null: Significant difference
5	Albumin	Normal + Diseased	**8.84e-04 (<0.05)**	Reject null: Significant difference
6	Age	Normal + Diseased	**2.31e-04 (<0.05)**	Reject null: Significant difference

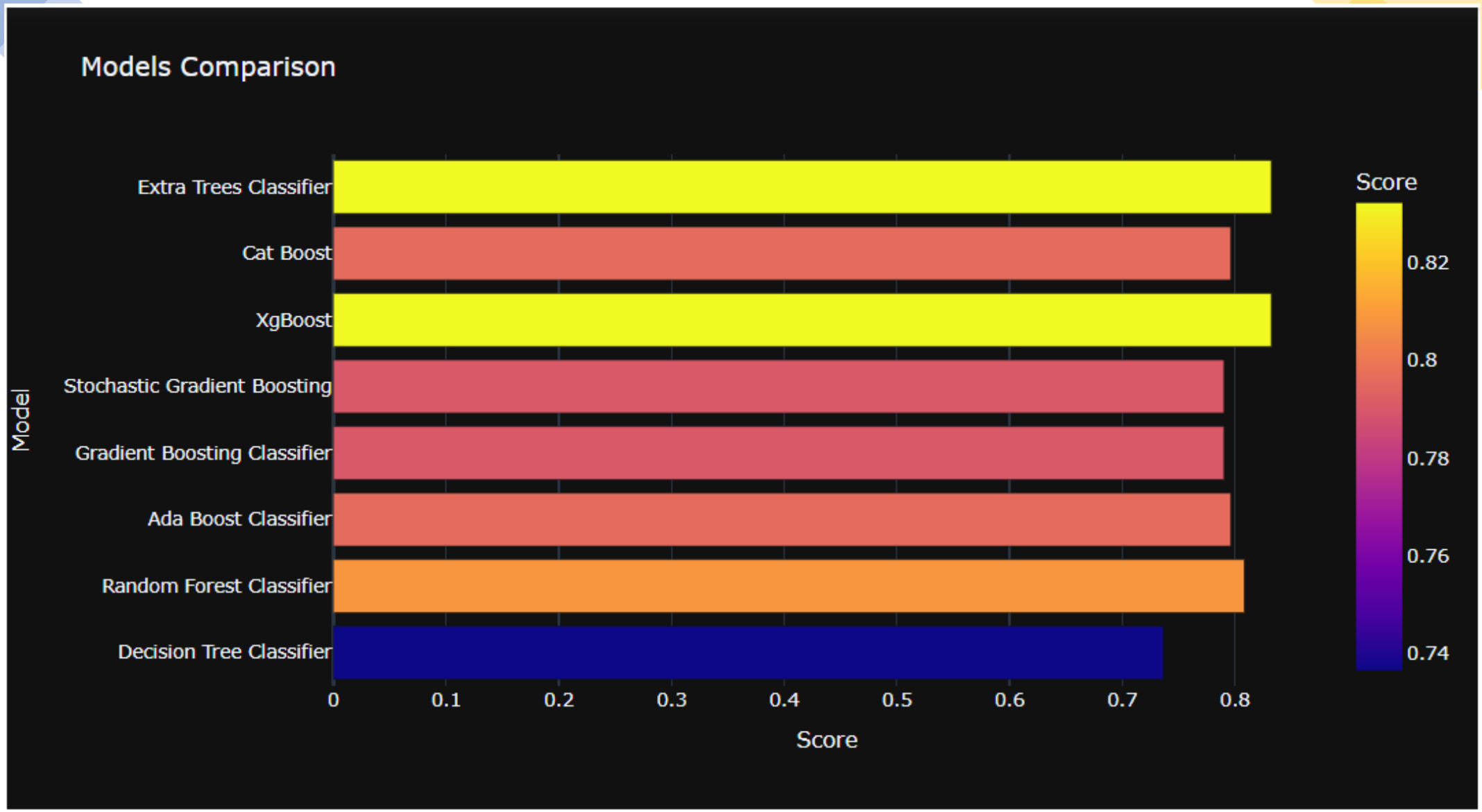
Observations:

- Certain features exhibit direct relationships among themselves, such as:
- **Protein features** - Albumin and Total Proteins; Albumin and Albumin-to-Globulin Ratio.
- **Enzyme features -** Aspartate Aminotransferase and Alanine Aminotransferase also show a direct relationship.
- **Bilirubin features** - Direct Bilirubin and Total Bilirubin

Model Schemas



Model Name	Test Accuracy
XGBoost	83.2%
Extra Trees Classifier	83.2%
Random Forest Classifier	80.8%
Ada Boost Classifier	79.6%
Cat Boost	79.6%
Gradient Boosting Classifier	79.0%
Stochastic Gradient Boosting	79.0%
Decision Tree Classifier	73.7%


Models Comparison
Extra Trees Classifier
Cat Boost
XgBoost
Stochastic Gradient Boosting
Gradient Boosting Classifier
Ada Boost Classifier
Random Forest Classifier
Decision Tree Classifier
Model
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Score
Score
0.82
0.8
0.78
0.76
0.74

What Worked: Certain features, such as enzyme(Alkaline Phosphatase & Alanine Aminotransferase), Total Protein and Total Bilirubin levels, showed significant impact on liver disease prediction.

What Didn't: Some models performed better than others; Decision Tree had comparatively lower accuracy.

Benefit: Accurate prediction models can aid in timely diagnosis and treatment decisions, potentially improving patient outcomes.

Conclusion:

- Based on the analysis, healthcare providers should focus on certain features that play a significant role in diagnosing liver disease for early diagnosis and treatment.

- For accurate predictions, we recommend using machine learning models such as XGBoost or Extra Trees Classifier.

Next Steps:

- Deployment: Planned possible collaboration with diagnostic centers to deploy the best predictive model using their patient dataset.

Thank You