

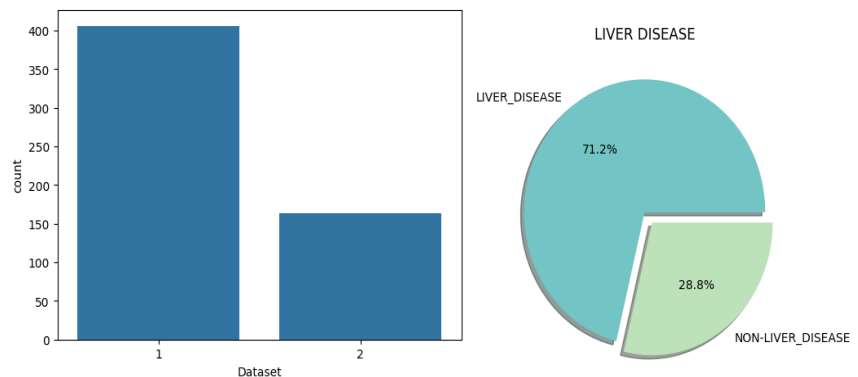
Data Collection and Preprocessing Phase

Date	23 september 2024
Team ID	LTVIP2024TMID25021
Project Title	Prediction and Analysis of liver patient data using machine learning
Maximum Marks	6 Marks

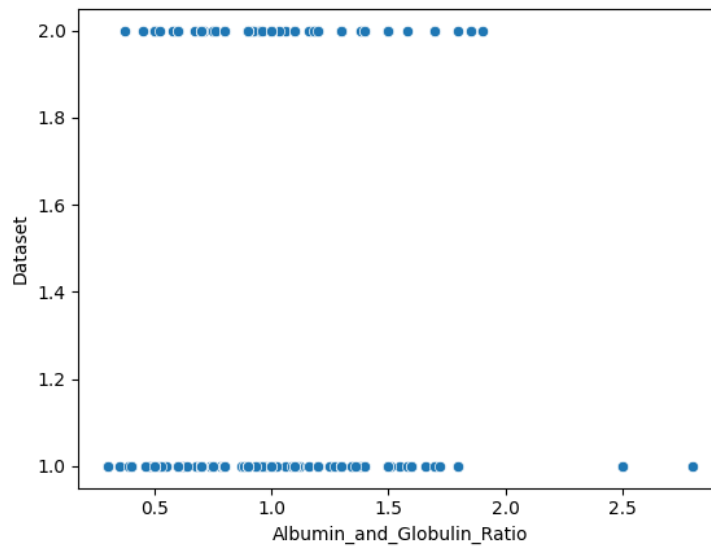
Data Exploration and Preprocessing Report

Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong foundation for insights and predictions.

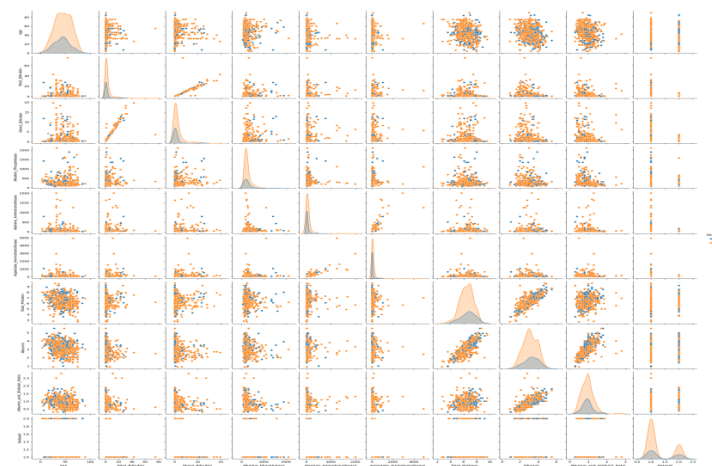
Univariate Analysis



Bivariate Analysis



Multivariate Analysis



Outliers and Anomalies

-

Data Preprocessing Code Screenshots

Loading Data	<pre>df.head(5)</pre> <table><thead><tr><th></th><th>Age</th><th>Gender</th><th>Total_Bilirubin</th><th>Direct_Bilirubin</th><th>Alkaline_Phosphatase</th><th>Alamine_Aminotransferase</th><th>Aspartate_Aminotransferase</th><th>Total_Protiens</th><th>Albumin</th></tr></thead><tbody><tr><td>0</td><td>65</td><td>Female</td><td>0.7</td><td>0.1</td><td>187</td><td>16</td><td>18</td><td>6.8</td><td>3.3</td></tr><tr><td>1</td><td>62</td><td>Male</td><td>10.9</td><td>5.5</td><td>699</td><td>64</td><td>100</td><td>7.5</td><td>3.2</td></tr><tr><td>2</td><td>62</td><td>Male</td><td>7.3</td><td>4.1</td><td>490</td><td>60</td><td>68</td><td>7.0</td><td>3.3</td></tr><tr><td>3</td><td>58</td><td>Male</td><td>1.0</td><td>0.4</td><td>182</td><td>14</td><td>20</td><td>6.8</td><td>3.4</td></tr><tr><td>4</td><td>72</td><td>Male</td><td>3.9</td><td>2.0</td><td>195</td><td>27</td><td>59</td><td>7.3</td><td>2.4</td></tr></tbody></table>		Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphatase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens	Albumin	0	65	Female	0.7	0.1	187	16	18	6.8	3.3	1	62	Male	10.9	5.5	699	64	100	7.5	3.2	2	62	Male	7.3	4.1	490	60	68	7.0	3.3	3	58	Male	1.0	0.4	182	14	20	6.8	3.4	4	72	Male	3.9	2.0	195	27	59	7.3	2.4
	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphatase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens	Albumin																																																				
0	65	Female	0.7	0.1	187	16	18	6.8	3.3																																																				
1	62	Male	10.9	5.5	699	64	100	7.5	3.2																																																				
2	62	Male	7.3	4.1	490	60	68	7.0	3.3																																																				
3	58	Male	1.0	0.4	182	14	20	6.8	3.4																																																				
4	72	Male	3.9	2.0	195	27	59	7.3	2.4																																																				
Handling Null Values	<pre>mode_value = df['Albumin_and_Globulin_Ratio'].mode()[0] df['Albumin_and_Globulin_Ratio'] = df['Albumin_and_Globulin_Ratio'].fillna(mode_value) df.isnull().sum()</pre>																																																												
Data Transformation	<pre>df['Gender']=df['Gender'].map({'Male':1,'Female':0}) df['Gender']=df['Gender'].astype('int64')</pre>																																																												
Feature Engineering	Attached the codes in final submission.																																																												
Save Processed Data	-																																																												