In [ ]:

```
IMPORTING LIBRARIES
```

In [ ]:

```
import os
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns

import warnings
warnings.filterwarnings('ignore')
```

In [ ]:

```
IMPORTING DATASET
```

In [2]:

```
df=pd.read_csv("C:/Users/ADMIN/Desktop/DATASETS/Blood_samples_dataset_balanced_2(f).csv")
```

In [3]:

```
df.head()
```

Out[3]:

	Glucose	Cholesterol	Hemoglobin	Platelets	White Blood Cells	Red Blood Cells	Hematocrit	Mean Corpuscular Volume	Mean Corpuscular Hemoglobin	Mean Corpuscular Hemoglobin Concentration	...	HbA1c	Chol
0	0.739597	0.650198	0.713631	0.868491	0.687433	0.529895	0.290006	0.631045	0.001328	0.795829	...	0.502665	0.
1	0.121786	0.023058	0.944893	0.905372	0.507711	0.403033	0.164216	0.307553	0.207938	0.505562	...	0.856810	0.
2	0.452539	0.116135	0.544560	0.400640	0.294538	0.382021	0.625267	0.295122	0.868369	0.026808	...	0.466795	0.
3	0.136609	0.015605	0.419957	0.191487	0.081168	0.166214	0.073293	0.668719	0.125447	0.501051	...	0.016256	0.
4	0.176737	0.752220	0.971779	0.785286	0.443880	0.439851	0.894991	0.442159	0.257288	0.805987	...	0.429431	0.

5 rows × 25 columns

In [4]:

```
df.tail()
```

Out[4]:

	Glucose	Cholesterol	Hemoglobin	Platelets	White Blood Cells	Red Blood Cells	Hematocrit	Mean Corpuscular Volume	Mean Corpuscular Hemoglobin	Mean Corpuscular Hemoglobin Concentration	...	HbA1c	C
2346	0.012956	0.336925	0.451218	0.175006	0.734664	0.382770	0.656463	0.177502	0.808162	0.684499	...	0.670665	
2347	0.407101	0.124738	0.983306	0.663867	0.361113	0.663716	0.232516	0.341056	0.847441	0.309766	...	0.491185	
2348	0.344356	0.783918	0.582171	0.996841	0.065363	0.242885	0.658851	0.543017	0.290106	0.838722	...	0.141738	
2349	0.351722	0.014278	0.898615	0.167550	0.727148	0.046091	0.900434	0.136227	0.134361	0.279219	...	0.570553	
2350	0.032726	0.053596	0.102633	0.221356	0.153956	0.216573	0.312577	0.608940	0.486174	0.450700	...	0.188750	

5 rows × 25 columns

In [5]:

```
df.shape
```

Out[5]:

```
(2351, 25)
```

In [6]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2351 entries, 0 to 2350
Data columns (total 25 columns):
 #   Column                                     Non-Null Count  Dtype  
---  ------                                     --------------  -----  
 0   Glucose                                    2351 non-null   float64
 1   Cholesterol                                2351 non-null   float64
 2   Hemoglobin                                 2351 non-null   float64
 3   Platelets                                  2351 non-null   float64
 4   White Blood Cells                          2351 non-null   float64
 5   Red Blood Cells                            2351 non-null   float64
 6   Hematocrit                                 2351 non-null   float64
 7   Mean Corpuscular Volume                    2351 non-null   float64
 8   Mean Corpuscular Hemoglobin                2351 non-null   float64
 9   Mean Corpuscular Hemoglobin Concentration  2351 non-null   float64
 10  Insulin                                    2351 non-null   float64
 11  BMI                                        2351 non-null   float64
 12  Systolic Blood Pressure                    2351 non-null   float64
 13  Diastolic Blood Pressure                   2351 non-null   float64
 14  Triglycerides                              2351 non-null   float64
 15  HbA1c                                      2351 non-null   float64
 16  LDL Cholesterol                            2351 non-null   float64
 17  HDL Cholesterol                            2351 non-null   float64
 18  ALT                                        2351 non-null   float64
 19  AST                                        2351 non-null   float64
 20  Heart Rate                                 2351 non-null   float64
 21  Creatinine                                 2351 non-null   float64
 22  Troponin                                   2351 non-null   float64
 23  C-reactive Protein                         2351 non-null   float64
 24  Disease                                    2351 non-null   object 
dtypes: float64(24), object(1)
memory usage: 459.3+ KB
```

In [7]:
```
df.describe(include='all')
```

Out[7]:

	Glucose	Cholesterol	Hemoglobin	Platelets	White Blood Cells	Red Blood Cells	Hematocrit	Mean Corpuscular Volume	Mean Corpuscular Hemoglobin	Mean Corpuscular Hemoglobin Concentration
count	2351.000000	2351.000000	2351.000000	2351.000000	2351.000000	2351.000000	2351.000000	2351.000000	2351.000000	2351.000000
unique	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
top	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
freq	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
mean	0.362828	0.393648	0.586190	0.504027	0.511086	0.506590	0.507152	0.492200	0.484459	0.562273
std	0.251889	0.239449	0.271498	0.303347	0.277270	0.266565	0.285537	0.275735	0.315618	0.273281
min	0.010994	0.012139	0.003021	0.012594	0.010139	0.044565	0.011772	0.046942	0.000554	0.006947
25%	0.129198	0.195818	0.346092	0.200865	0.259467	0.263589	0.288132	0.287532	0.207938	0.355774
50%	0.351722	0.397083	0.609836	0.533962	0.527381	0.467431	0.493428	0.453052	0.420723	0.603635
75%	0.582278	0.582178	0.791215	0.754841	0.743164	0.743670	0.753657	0.722293	0.778160	0.741381
max	0.968460	0.905026	0.983306	0.999393	0.990786	1.000000	0.977520	0.995263	0.963235	0.975586

11 rows × 25 columns

In [ ]:
```
CHECKING FOR MISSING VALUES
```

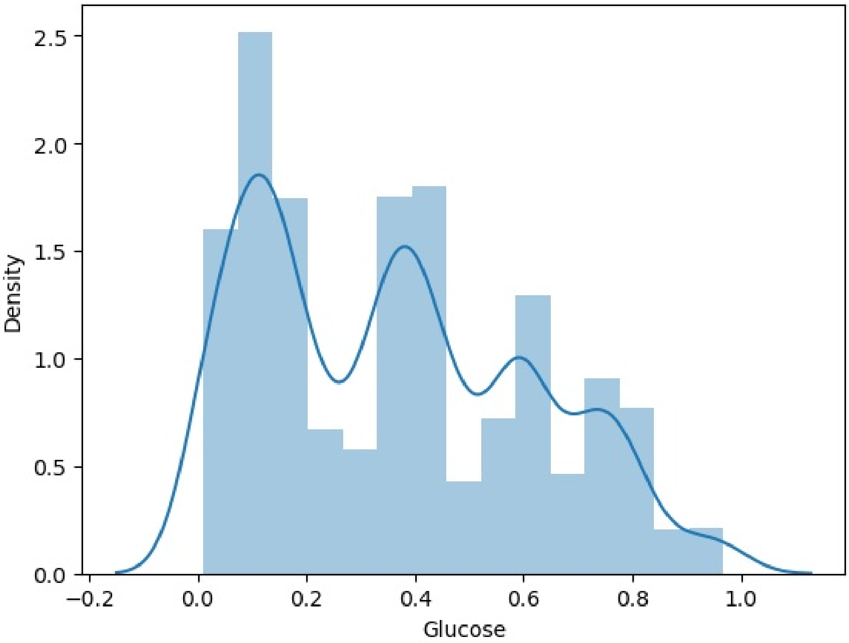
In [8]:
```
df.isnull().sum()
```

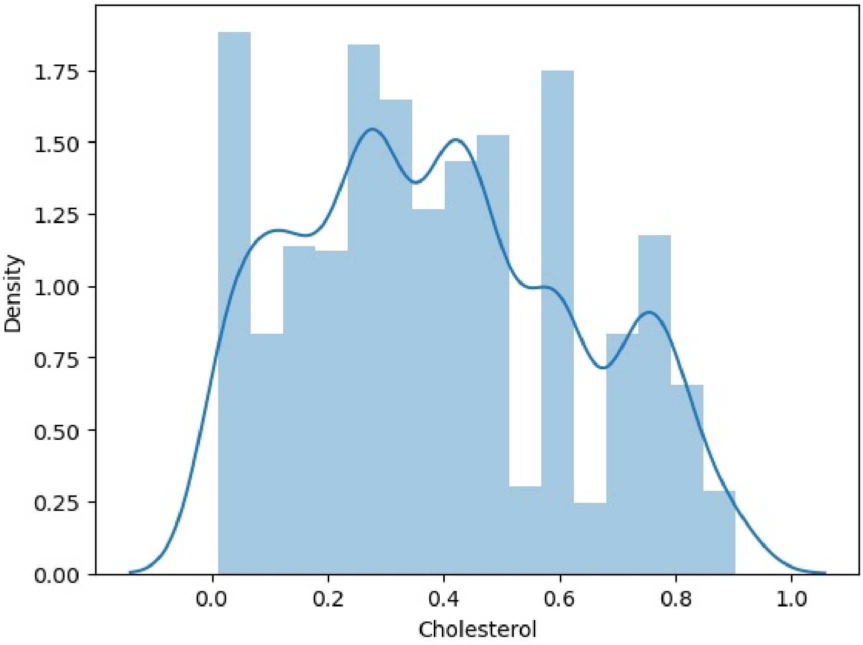
```
Out[8]: Glucose                                         0
        Cholesterol                                     0
        Hemoglobin                                      0
        Platelets                                       0
        White Blood Cells                               0
        Red Blood Cells                                 0
        Hematocrit                                      0
        Mean Corpuscular Volume                         0
        Mean Corpuscular Hemoglobin                     0
        Mean Corpuscular Hemoglobin Concentration       0
        Insulin                                         0
        BMI                                             0
        Systolic Blood Pressure                         0
        Diastolic Blood Pressure                        0
        Triglycerides                                   0
        HbA1c                                           0
        LDL Cholesterol                                 0
        HDL Cholesterol                                 0
        ALT                                             0
        AST                                             0
        Heart Rate                                      0
        Creatinine                                      0
        Troponin                                        0
        C-reactive Protein                              0
        Disease                                         0
        dtype: int64
```

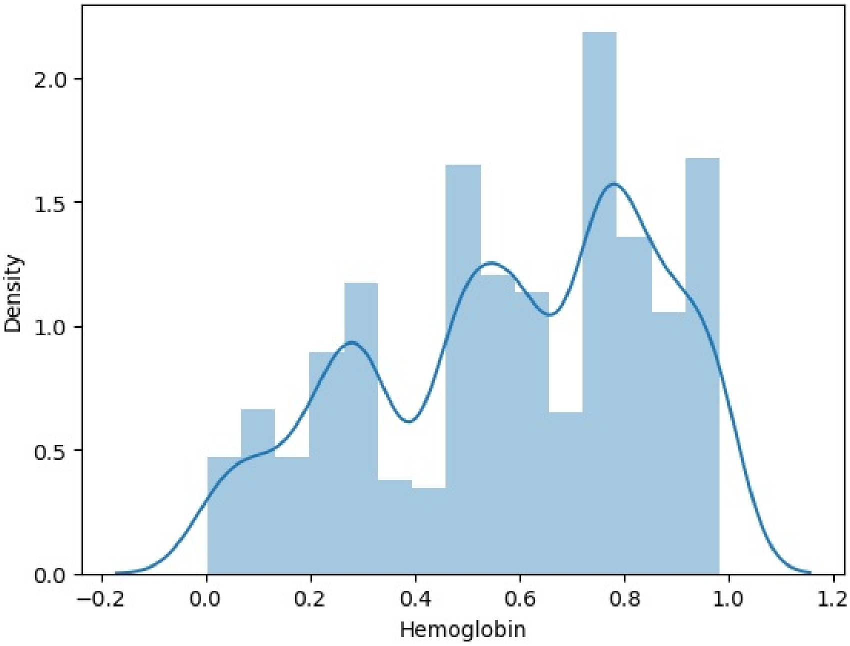
```
In [ ]: VISUALIZING THE DISTRIBUTION
```

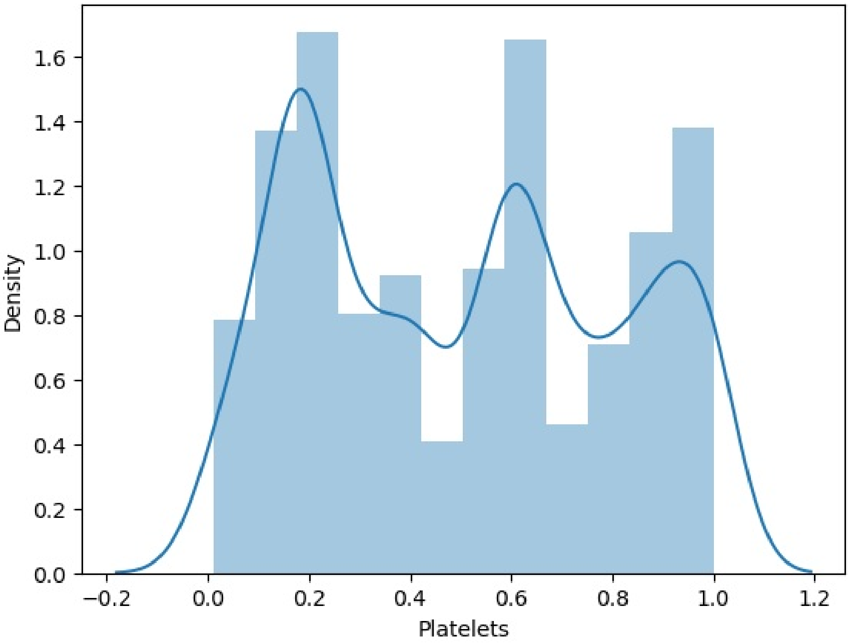
```
In [9]: def distplots(col):
            sns.distplot(df[col])
            plt.show()
```

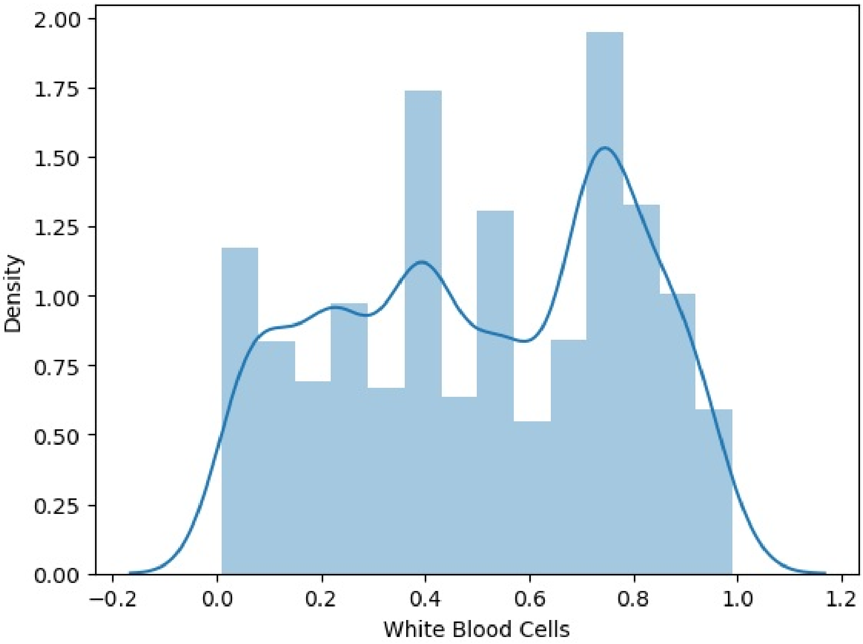
```
In [10]: for i in list(df.select_dtypes(exclude=['object']).columns):
             distplots(i)
```

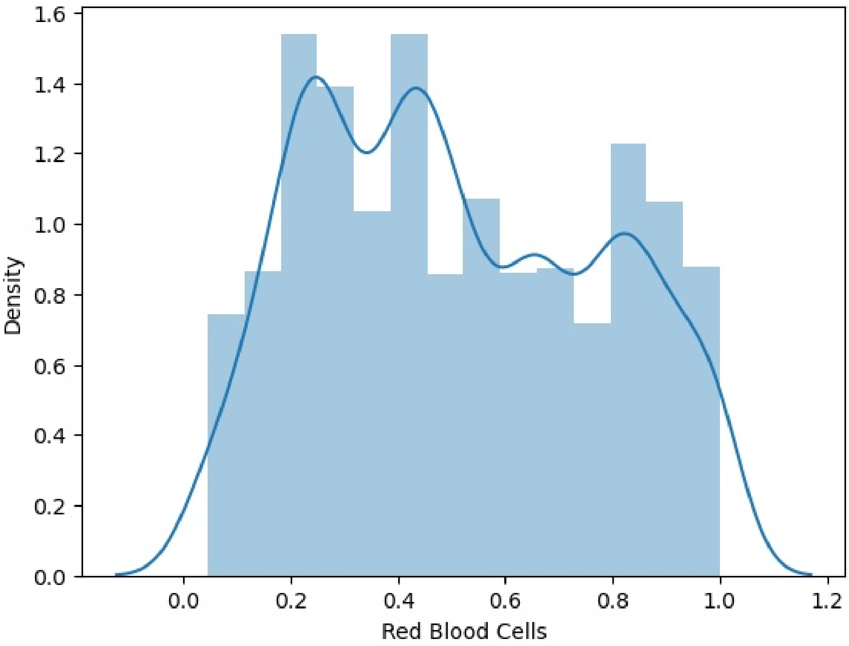



Density
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
0.0
0.2
0.4
0.6
0.8
1.0
Cholesterol

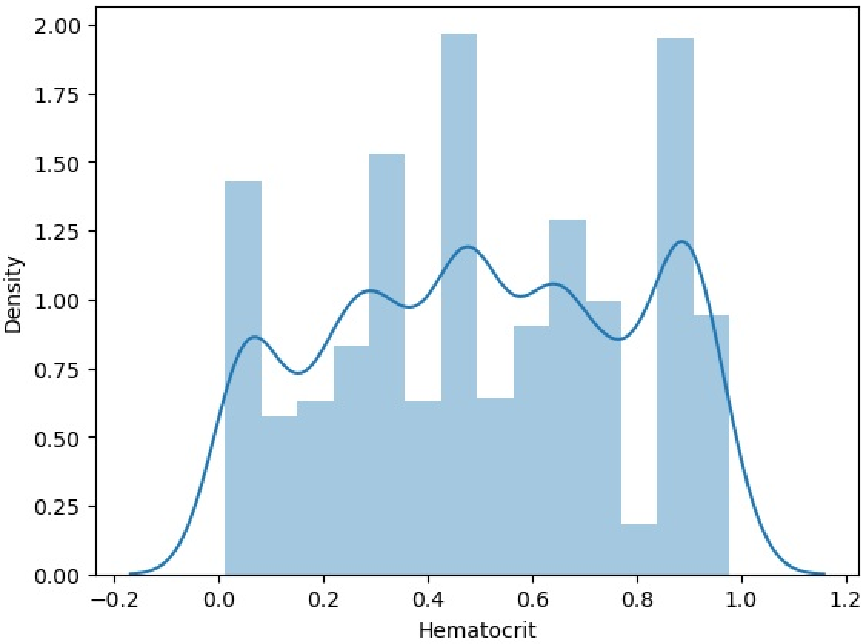

Density
0.0
0.5
1.0
1.5
2.0
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Hemoglobin


Density
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Platelets

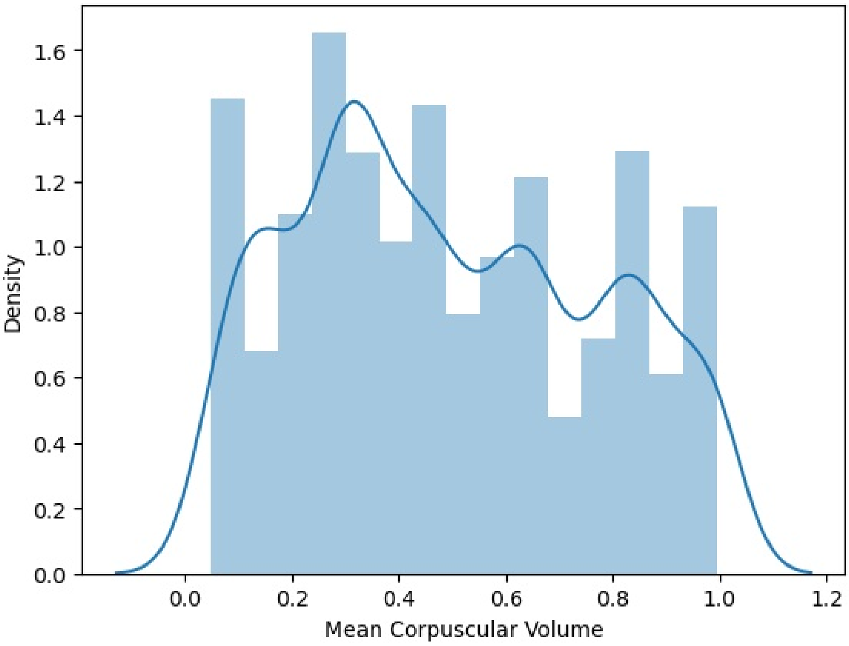

Density
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
White Blood Cells

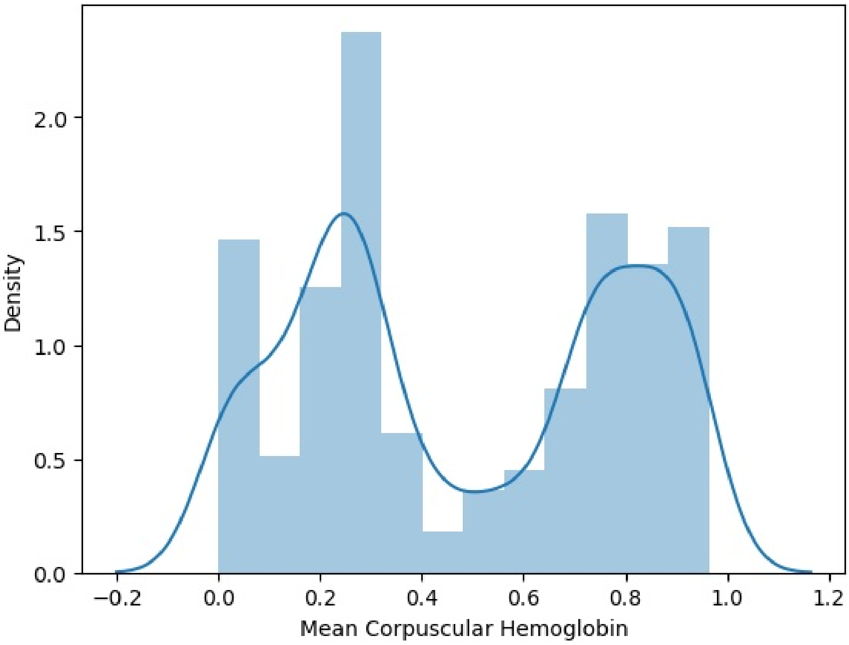


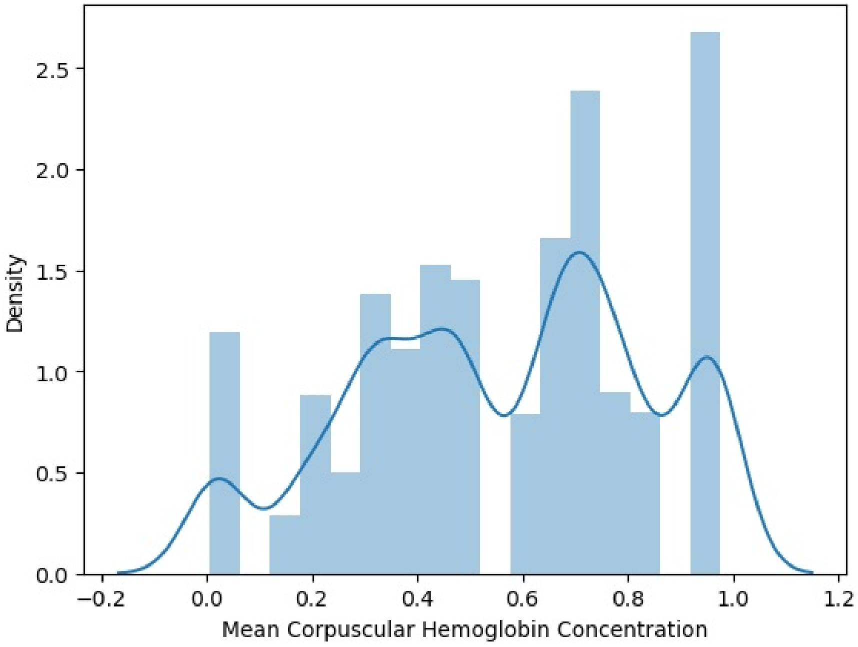
Density
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Red Blood Cells

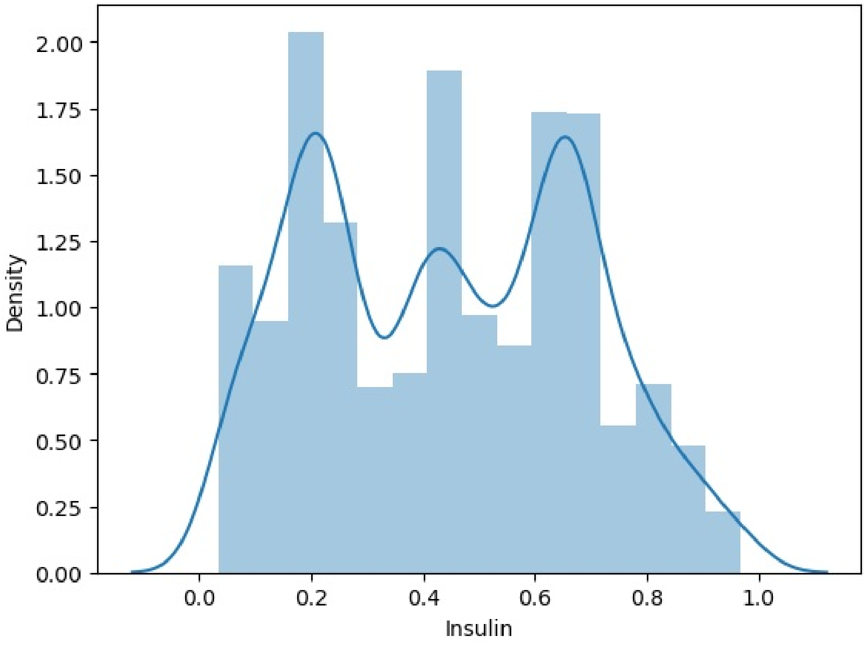


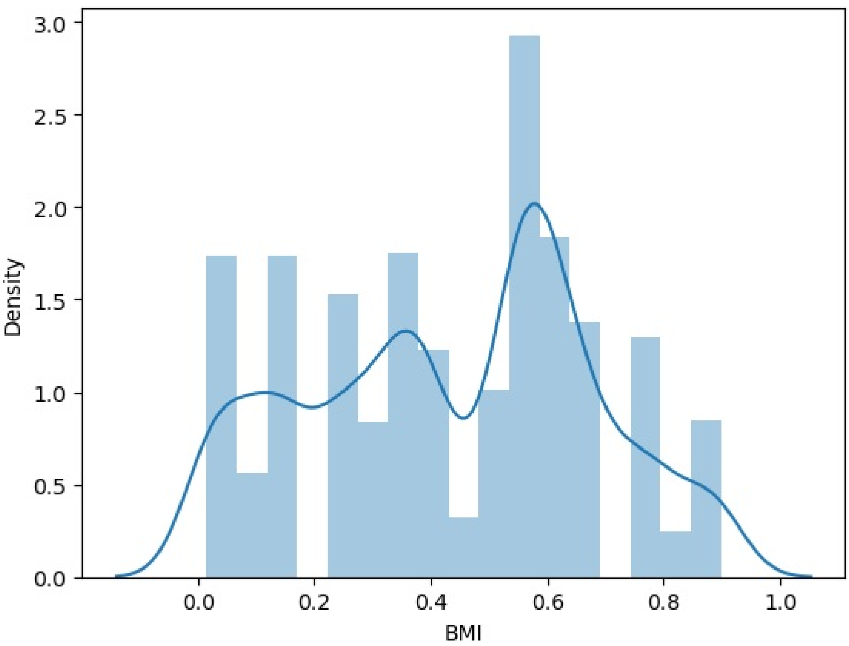
Density
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Hematocrit

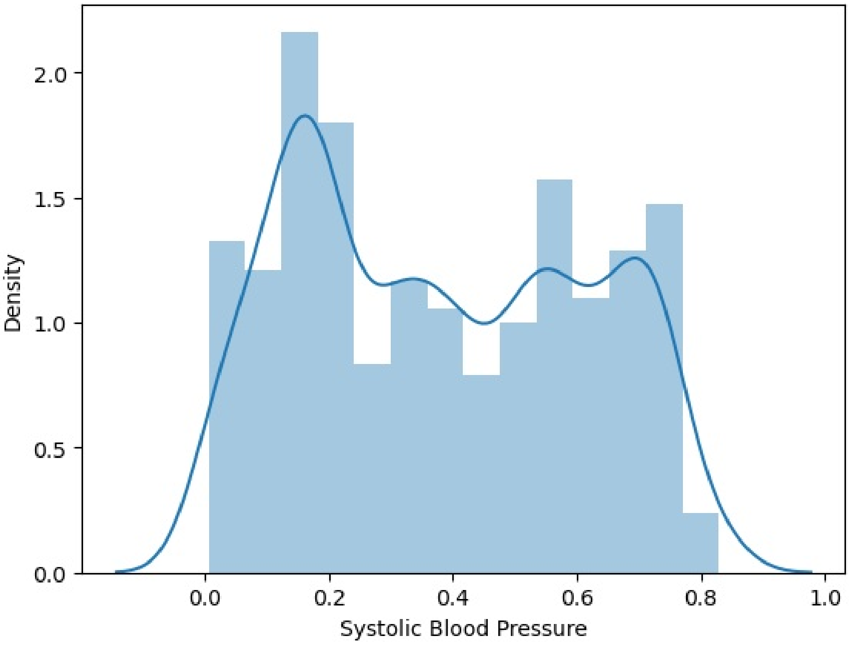

1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Density
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Mean Corpuscular Volume

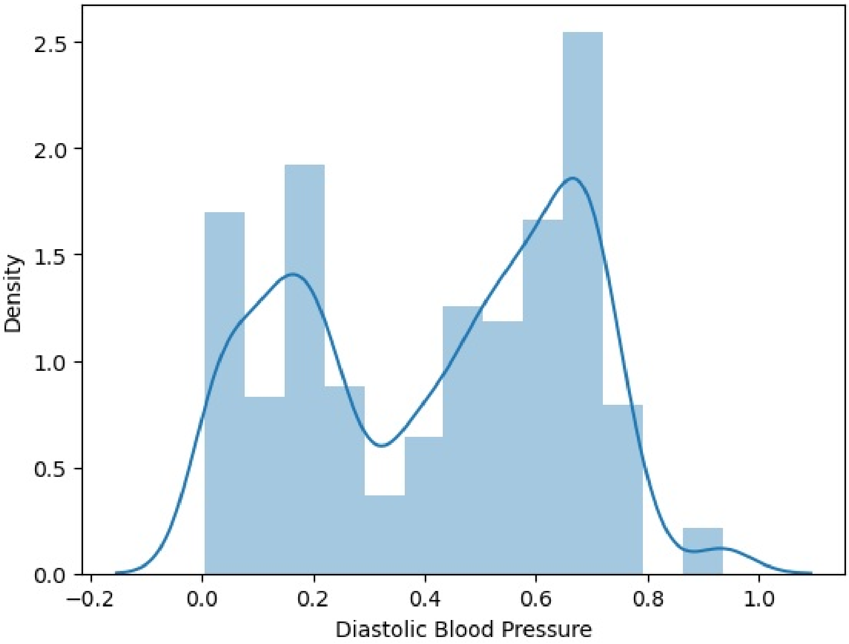

2.0
1.5
1.0
0.5
0.0
Density
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Mean Corpuscular Hemoglobin

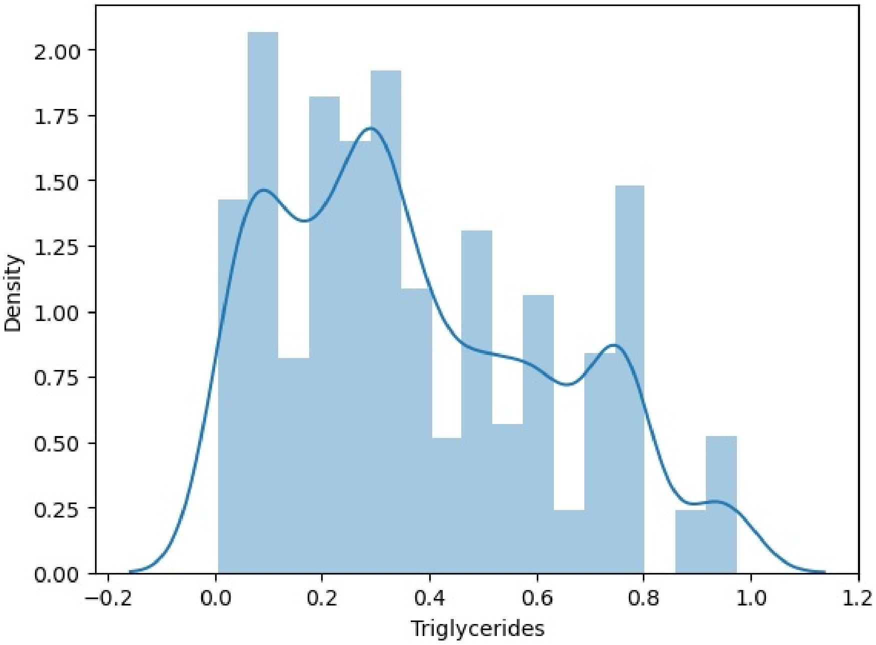

Density
0.0
0.5
1.0
1.5
2.0
2.5
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Mean Corpuscular Hemoglobin Concentration

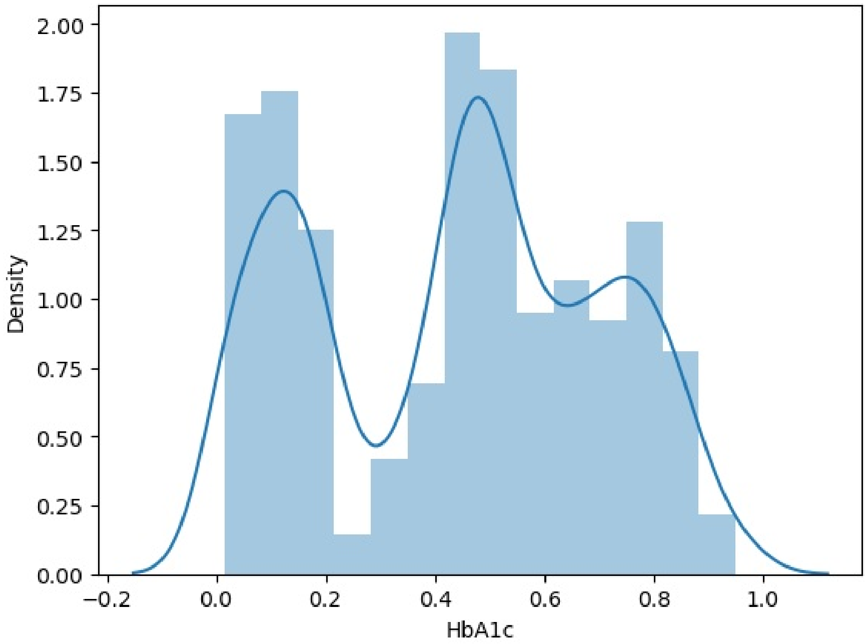

Density
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
0.0
0.2
0.4
0.6
0.8
1.0
Insulin


3.0
2.5
2.0
1.5
1.0
0.5
0.0
Density
0.0
0.2
0.4
0.6
0.8
1.0
BMI

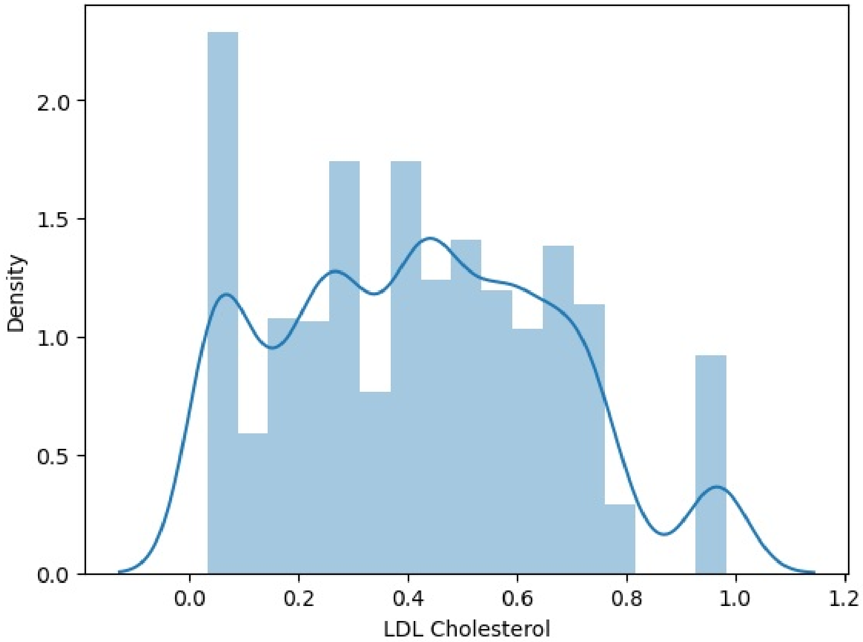

2.0
1.5
1.0
0.5
0.0
Density
0.0
0.2
0.4
0.6
0.8
1.0
Systolic Blood Pressure


Density
0.0
0.5
1.0
1.5
2.0
2.5
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
Diastolic Blood Pressure

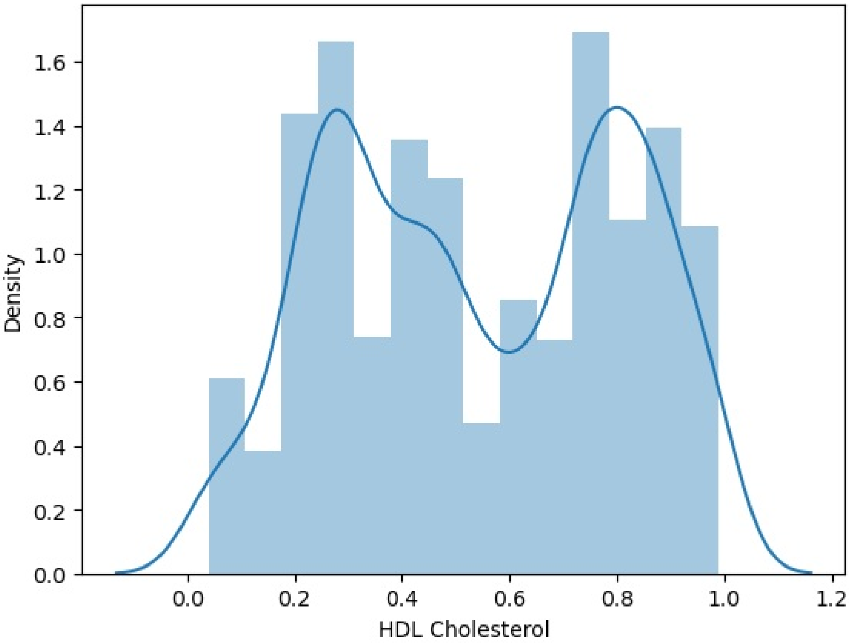

Density
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Triglycerides

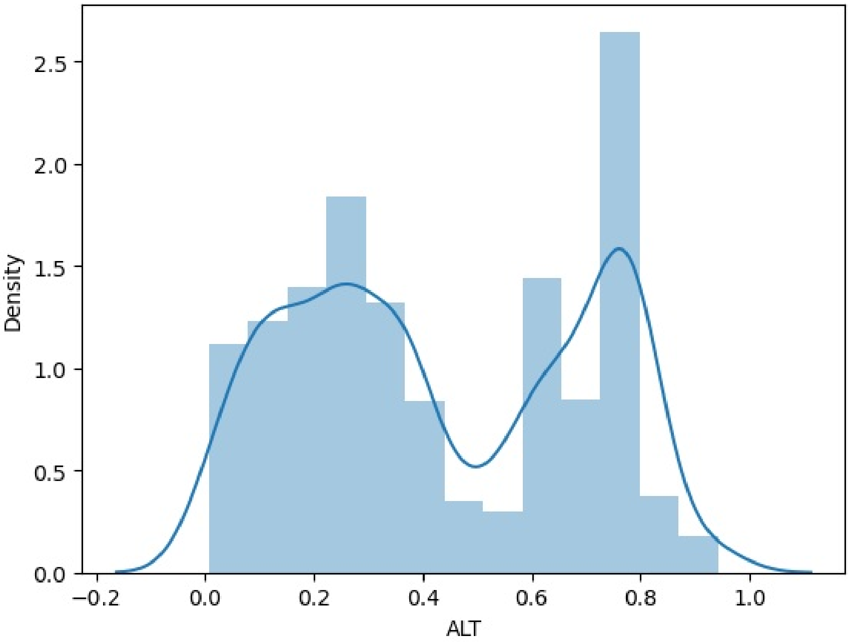


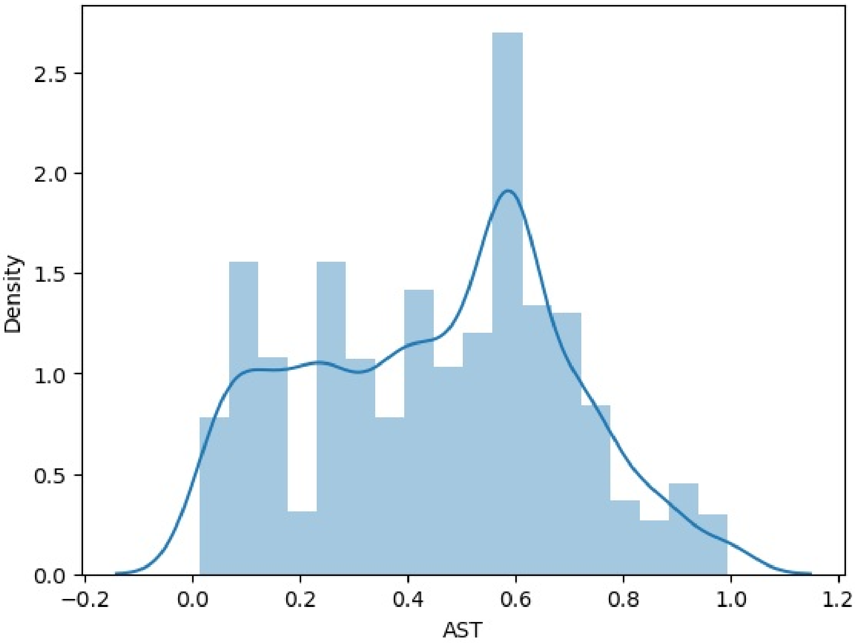
Density
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
HbA1c

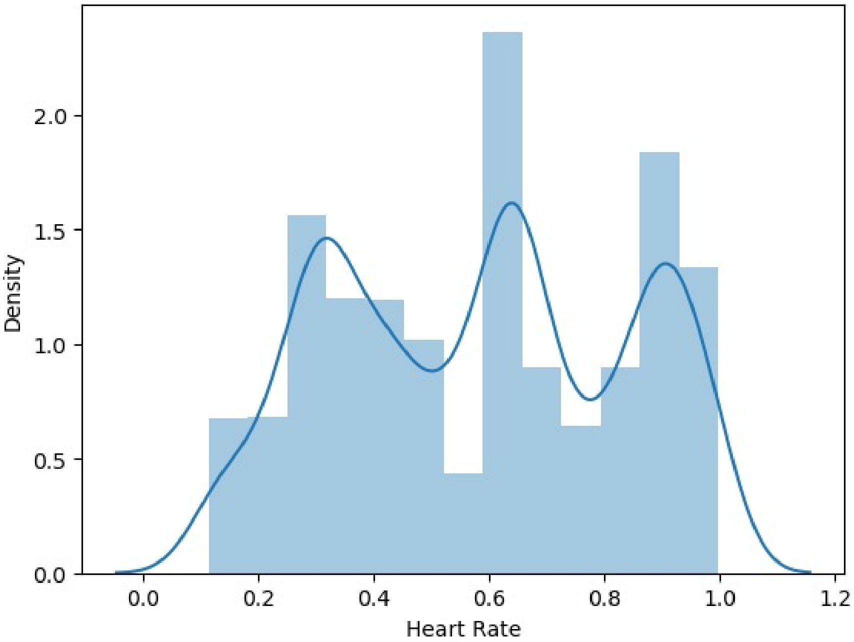


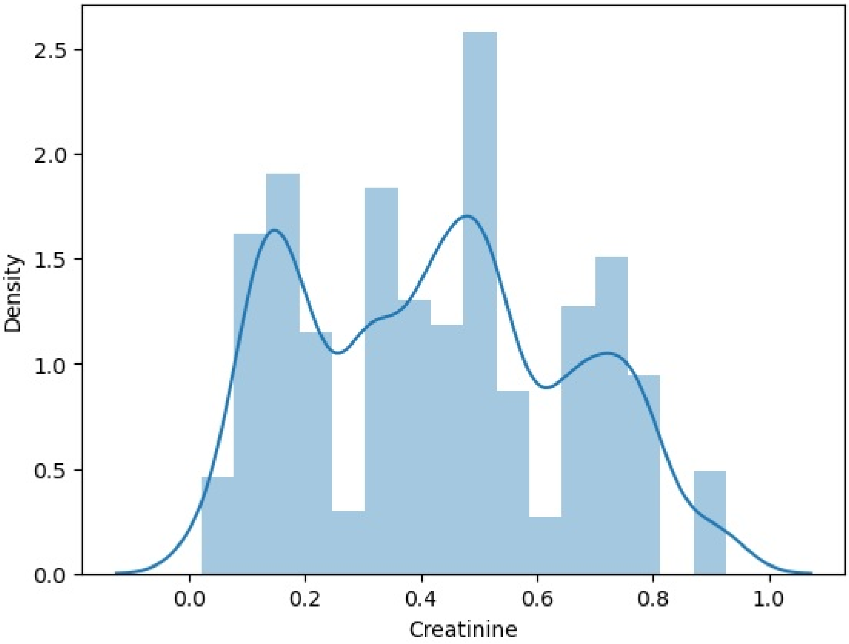
Density
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0
1.2
LDL Cholesterol


Density
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
1.2
HDL Cholesterol

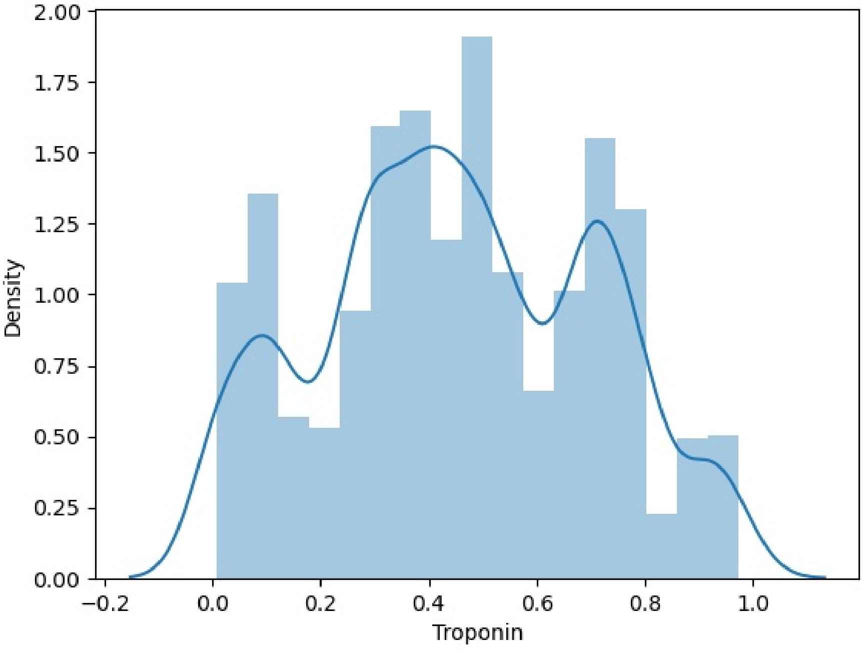

Density
2.5
2.0
1.5
1.0
0.5
0.0
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
ALT


Density
0.0
0.5
1.0
1.5
2.0
2.5
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
AST

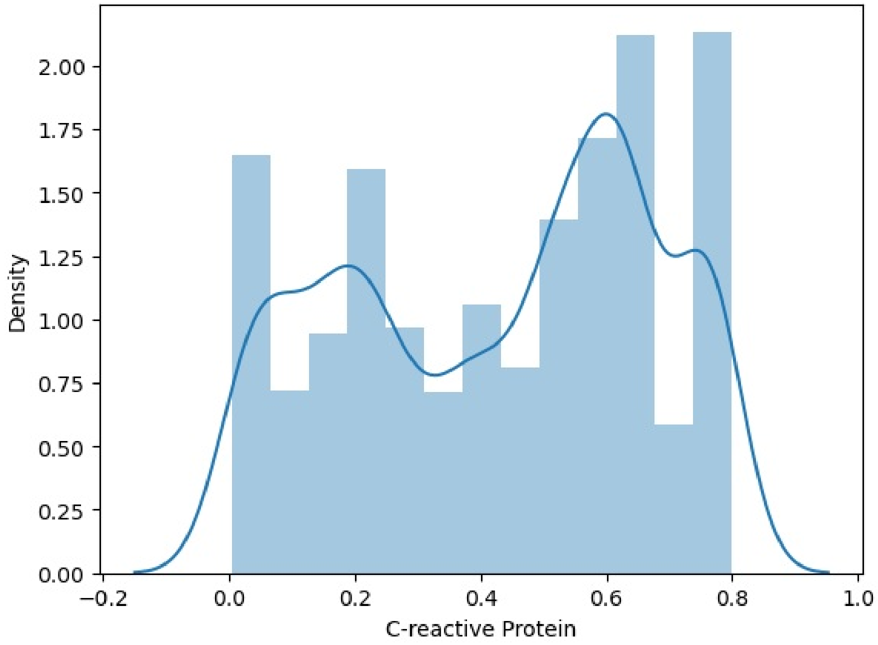

Density
0.0
0.5
1.0
1.5
2.0
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Heart Rate



2.5
2.0
1.5
1.0
0.5
0.0
Density
0.0
0.2
0.4
0.6
0.8
1.0
Creatinine



2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Density
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
Troponin



In [11]:
```
from sklearn.preprocessing import LabelEncoder
encoder=LabelEncoder()
df['Disease']=encoder.fit_transform(df['Disease'])
```

In [16]:
```
df.head()
```

Out[16]:

	Glucose	Cholesterol	Hemoglobin	Platelets	White Blood Cells	Red Blood Cells	Hematocrit	Mean Corpuscular Volume	Mean Corpuscular Hemoglobin	Mean Corpuscular Hemoglobin Concentration	...	HbA1c	Chol
0	0.739597	0.650198	0.713631	0.868491	0.687433	0.529895	0.290006	0.631045	0.001328	0.795829	...	0.502665	0.
1	0.121786	0.023058	0.944893	0.905372	0.507711	0.403033	0.164216	0.307553	0.207938	0.505562	...	0.856810	0.
2	0.452539	0.116135	0.544560	0.400640	0.294538	0.382021	0.625267	0.295122	0.868369	0.026808	...	0.466795	0.
3	0.136609	0.015605	0.419957	0.191487	0.081168	0.166214	0.073293	0.668719	0.125447	0.501051	...	0.016256	0.
4	0.176737	0.752220	0.971779	0.785286	0.443880	0.439851	0.894991	0.442159	0.257288	0.805987	...	0.429431	0.

5 rows × 25 columns

In [12]:
```
x=df.iloc[:,0:25]
```

In [14]:
```
y=df.iloc[:,24]
```

In [17]:
```
x
```

Out[17]:

	Glucose	Cholesterol	Hemoglobin	Platelets	White Blood Cells	Red Blood Cells	Hematocrit	Mean Corpuscular Volume	Mean Corpuscular Hemoglobin	Mean Corpuscular Hemoglobin Concentration	...	HbA1c	C
0	0.739597	0.650198	0.713631	0.868491	0.687433	0.529895	0.290006	0.631045	0.001328	0.795829	...	0.502665	
1	0.121786	0.023058	0.944893	0.905372	0.507711	0.403033	0.164216	0.307553	0.207938	0.505562	...	0.856810	
2	0.452539	0.116135	0.544560	0.400640	0.294538	0.382021	0.625267	0.295122	0.868369	0.026808	...	0.466795	
3	0.136609	0.015605	0.419957	0.191487	0.081168	0.166214	0.073293	0.668719	0.125447	0.501051	...	0.016256	
4	0.176737	0.752220	0.971779	0.785286	0.443880	0.439851	0.894991	0.442159	0.257288	0.805987	...	0.429431	
...	...	...	...	...	...	...	...	...	...	...	...	...	
2346	0.012956	0.336925	0.451218	0.175006	0.734664	0.382770	0.656463	0.177502	0.808162	0.684499	...	0.670665	
2347	0.407101	0.124738	0.983306	0.663867	0.361113	0.663716	0.232516	0.341056	0.847441	0.309766	...	0.491185	
2348	0.344356	0.783918	0.582171	0.996841	0.065363	0.242885	0.658851	0.543017	0.290106	0.838722	...	0.141738	
2349	0.351722	0.014278	0.898615	0.167550	0.727148	0.046091	0.900434	0.136227	0.134361	0.279219	...	0.570553	
2350	0.032726	0.053596	0.102633	0.221356	0.153956	0.216573	0.312577	0.608940	0.486174	0.450700	...	0.188750	

2351 rows × 25 columns

In [18]:

```
y
```

Out[18]:

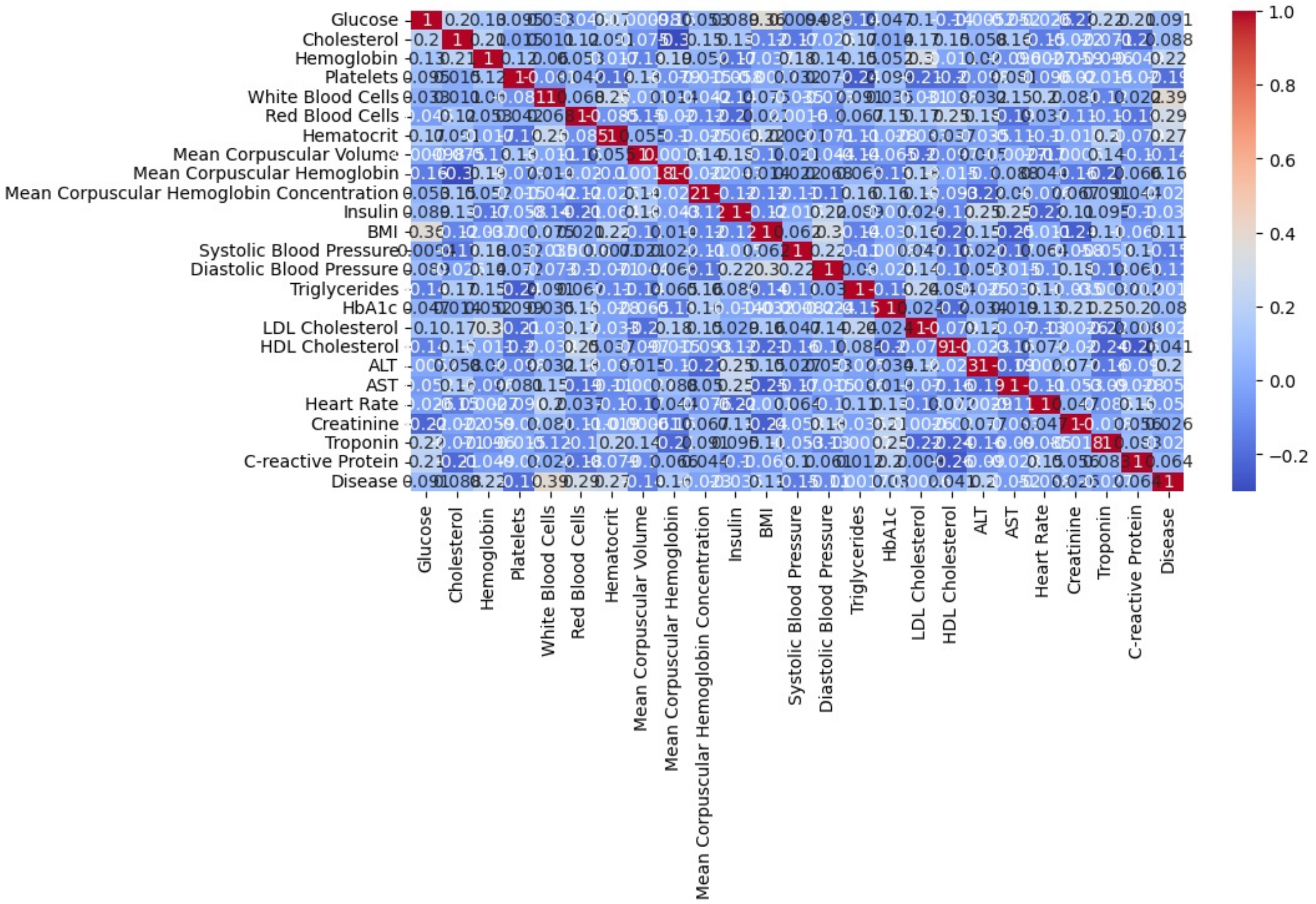
```
0       2
1       1
2       3
3       0
4       3
       ..
2346    1
2347    3
2348    0
2349    1
2350    0
Name: Disease, Length: 2351, dtype: int32
```

In [ ]:

```
CHECKING THE CORRELATION OF THE DATASET
```

In [21]:

```
plt.figure(figsize=(10,5))
sns.heatmap(df.corr(),annot=True,cmap='coolwarm')
plt.show()
```



In [ ]:

```
SPLIT DATASET INTO TRAIN AND TEST DATA
```

In [22]:

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=4)
```

In [24]:

```
from sklearn.preprocessing import StandardScaler
scaler=StandardScaler()
x_train=scaler.fit_transform(x_train)
x_test=scaler.fit_transform(x_test)
```

In [ ]:

```
BUILDING LOGISTIC MODEL
```

In [26]:

```
from sklearn.linear_model import LogisticRegression
lr=LogisticRegression()
lr.fit(x_train,y_train)
```

Out[26]:

```
▾ LogisticRegression
LogisticRegression()
```

In [ ]:

```
PREDICTING AND CHECKING ACCURACY OF THE MODEL
```

In [27]:

```
from sklearn.metrics import accuracy_score
yhat=lr.predict(x_test)
print("Accuracy:",accuracy_score(y_test,yhat))
```

```
Accuracy: 1.0
```

In [ ]:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js