

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

```
df = pd.read_csv(r'/content/drive/MyDrive/indian liver patient.csv')
```

```
df.head()
```

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alami
0	65	Female	0.7	0.1	187	
1	62	Male	10.9	5.5	699	
2	62	Male	7.3	4.1	490	
3	58	Male	1.0	0.4	182	
4	72	Male	3.9	2.0	195	

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 583 entries, 0 to 582
Data columns (total 11 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Age                                    583 non-null    int64
1   Gender                                583 non-null    object
2   Total_Bilirubin                       583 non-null    float64
3   Direct_Bilirubin                      583 non-null    float64
4   Alkaline_Phosphotase                  583 non-null    int64
5   Alamine_Aminotransferase              583 non-null    int64
6   Aspartate_Aminotransferase            583 non-null    int64
7   Total_Protiens                        583 non-null    float64
8   Albumin                              583 non-null    float64
9   Albumin_and_Globulin_Ratio            579 non-null    float64
10  Dataset                               583 non-null    int64
dtypes: float64(5), int64(5), object(1)
memory usage: 50.2+ KB
```

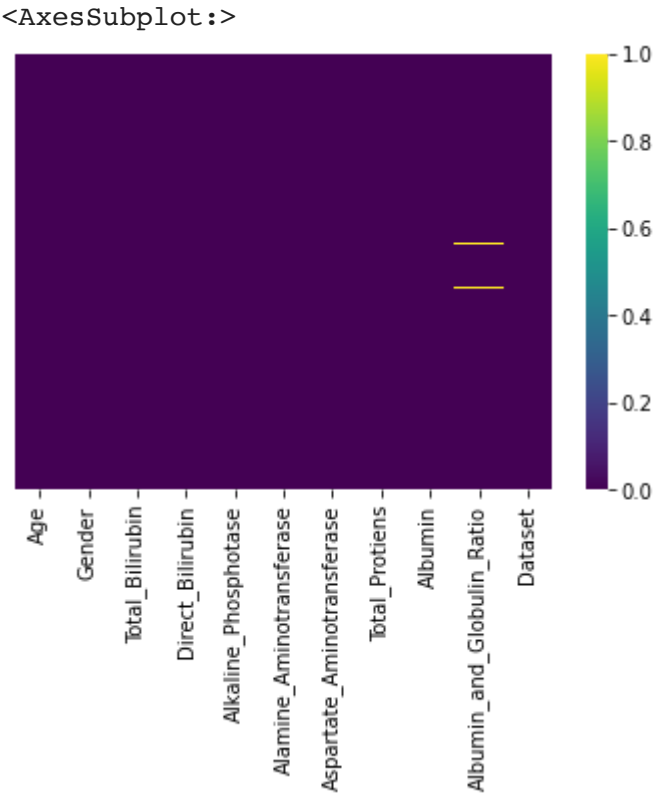
```
df.describe()
```

	Age	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Ala
count	583.000000	583.000000	583.000000	583.000000	
mean	44.746141	3.298799	1.486106	290.576329	
std	16.189833	6.209522	2.808498	242.937989	
min	4.000000	0.400000	0.100000	63.000000	

```
df.isnull().sum()
```

```
Age          0
Gender        0
Total_Bilirubin  0
Direct_Bilirubin  0
Alkaline_Phosphotase  0
Alamine_Aminotransferase  0
Aspartate_Aminotransferase  0
Total_Protiens  0
Albumin       0
Albumin_and_Globulin_Ratio  4
Dataset       0
dtype: int64
```

```
sns.heatmap(df.isnull(), yticklabels=False, cmap='viridis')
```



```
df.shape
```

```
(583, 11)
```

```
df['Gender'].unique()
```

```
array(['Female', 'Male'], dtype=object)
```

```
df['Gender'].nunique()
```

```
2
```

```
df['Gender'] = df['Gender'].map({'Male': 1, 'Female': 0})
```

```
df.head()
```

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alami
0	65	0	0.7	0.1	187	
1	62	1	10.9	5.5	699	
2	62	1	7.3	4.1	490	
3	58	1	1.0	0.4	182	
4	72	1	3.9	2.0	195	

```
df.dropna(inplace=True)
```

```
df.shape
```

```
(579, 11)
```

```
df.head()
```

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alami
0	65	0	0.7	0.1	187	
1	62	1	10.9	5.5	699	
2	62	1	7.3	4.1	490	
3	58	1	1.0	0.4	182	
4	72	1	3.9	2.0	195	

```
df['Dataset'].unique()
```

```
array([1, 2], dtype=int64)
```

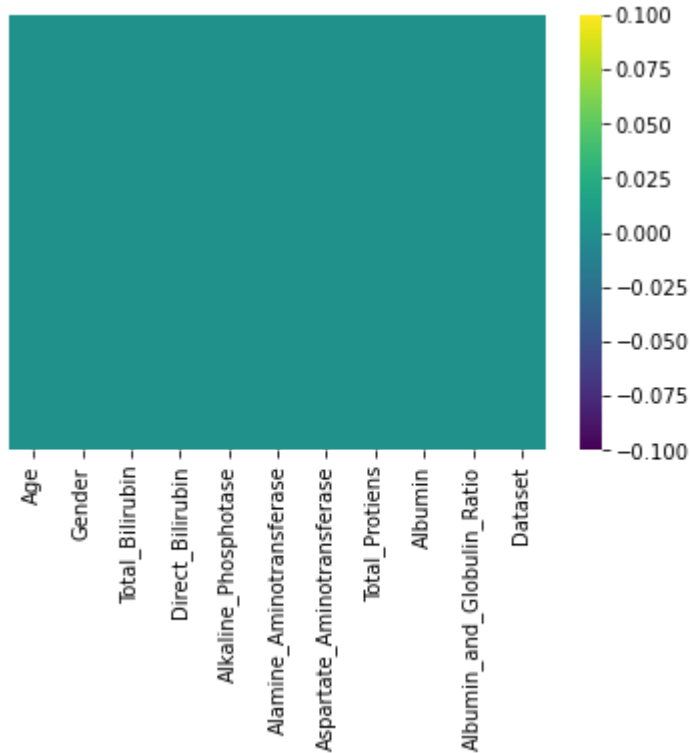
```
df['Dataset'].value_counts()
```

```
1    414
2    165
```

Name: Dataset, dtype: int64

```
sns.heatmap(df.isnull(),yticklabels=False,cmap='viridis')
```

<AxesSubplot:>



```
df.corr()['Dataset']
```

```
Age -0.133164
Gender -0.081349
Total_Bilirubin -0.220218
Direct_Bilirubin -0.246273
Alkaline_Phosphotase -0.183363
Alamine_Aminotransferase -0.163117
Aspartate_Aminotransferase -0.151834
Total_Protiens 0.033614
Albumin 0.159770
Albumin_and_Globulin_Ratio 0.163131
Dataset 1.000000
Name: Dataset, dtype: float64
```

```
!pip install pandas-profiling
```

```
Requirement already satisfied: pandas-profiling in c:\users\wecome\anaconda3\
Requirement already satisfied: pydantic<1.11,>=1.8.1 in c:\users\wecome\anaco:
Requirement already satisfied: scipy<1.10,>=1.4.1 in c:\users\wecome\anaconda
Requirement already satisfied: missingno<0.6,>=0.4.2 in c:\users\wecome\anaco:
Requirement already satisfied: tqdm<4.65,>=4.48.2 in c:\users\wecome\anaconda
Requirement already satisfied: numpy<1.24,>=1.16.0 in c:\users\wecome\anacond
Requirement already satisfied: phik<0.13,>=0.11.1 in c:\users\wecome\anaconda
Requirement already satisfied: visions[type_image_path]==0.7.5 in c:\users\we
Requirement already satisfied: seaborn<0.13,>=0.10.1 in c:\users\wecome\anaco:
Requirement already satisfied: multimethod<1.10,>=1.4 in c:\users\wecome\anac
Requirement already satisfied: statsmodels<0.14,>=0.13.2 in c:\users\wecome\anac
Requirement already satisfied: PyYAML<6.1,>=5.0.0 in c:\users\wecome\anaconda
Requirement already satisfied: requests<2.29,>=2.24.0 in c:\users\wecome\anac
```

```

Requirement already satisfied: matplotlib<3.6,>=3.2 in c:\users\wecome\anaconda3\lib\site-packages\matplotlib-3.5.0-py3.9-win-amd64.exe\matplotlib\
Requirement already satisfied: pandas!=1.4.0,<1.6,>1.1 in c:\users\wecome\anaconda3\lib\site-packages\pandas-1.3.4-py3.9-win-amd64.exe\pandas\
Requirement already satisfied: jinja2<3.2,>=2.11.1 in c:\users\wecome\anaconda3\lib\site-packages\jinja2-3.1.2-py3.9-win-amd64.exe\jinja2\
Requirement already satisfied: htmlmin==0.1.12 in c:\users\wecome\anaconda3\lib\site-packages\htmlmin-0.1.12-py3.9-win-amd64.exe\htmlmin\
Requirement already satisfied: attrs>=19.3.0 in c:\users\wecome\anaconda3\lib\site-packages\attrs-21.4.0-py3.9-win-amd64.exe\attrs\
Requirement already satisfied: networkx>=2.4 in c:\users\wecome\anaconda3\lib\site-packages\networkx-2.6.3-py3.9-win-amd64.exe\networkx\
Requirement already satisfied: tangled-up-in-unicode>=0.0.4 in c:\users\wecome\anaconda3\lib\site-packages\tangled-up-in-unicode-0.0.4-py3.9-win-amd64.exe\tangled-up-in-unicode\
Requirement already satisfied: imagehash in c:\users\wecome\anaconda3\lib\site-packages\imagehash-4.0.0-py3.9-win-amd64.exe\imagehash\
Requirement already satisfied: Pillow in c:\users\wecome\anaconda3\lib\site-packages\Pillow-9.0.1-py3.9-win-amd64.exe\Pillow\
Requirement already satisfied: MarkupSafe>=0.23 in c:\users\wecome\anaconda3\lib\site-packages\MarkupSafe-2.1.1-py3.9-win-amd64.exe\MarkupSafe\
Requirement already satisfied: pyparsing>=2.2.1 in c:\users\wecome\anaconda3\lib\site-packages\pyparsing-3.0.9-py3.9-win-amd64.exe\pyparsing\
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\wecome\anaconda3\lib\site-packages\kiwisolver-1.4.2-py3.9-win-amd64.exe\kiwisolver\
Requirement already satisfied: python-dateutil>=2.7 in c:\users\wecome\anaconda3\lib\site-packages\python-dateutil-2.8.2-py3.9-win-amd64.exe\python-dateutil\
Requirement already satisfied: cycycler>=0.10 in c:\users\wecome\anaconda3\lib\site-packages\cycycler-0.10.0-py3.9-win-amd64.exe\cycycler\
Requirement already satisfied: six in c:\users\wecome\anaconda3\lib\site-packages\six-1.16.0-py3.9-win-amd64.exe\six\
Requirement already satisfied: pytz>=2017.3 in c:\users\wecome\anaconda3\lib\site-packages\pytz-2022.1-py3.9-win-amd64.exe\pytz\
Requirement already satisfied: joblib>=0.14.1 in c:\users\wecome\anaconda3\lib\site-packages\joblib-1.1.0-py3.9-win-amd64.exe\joblib\
Requirement already satisfied: typing-extensions>=4.1.0 in c:\users\wecome\anaconda3\lib\site-packages\typing_extensions-4.1.1-py3.9-win-amd64.exe\typing_extensions\
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\wecome\anaconda3\lib\site-packages\urllib3-1.26.13-py3.9-win-amd64.exe\urllib3\
Requirement already satisfied: charset-normalizer~2.0.0 in c:\users\wecome\anaconda3\lib\site-packages\charset-normalizer-2.0.12-py3.9-win-amd64.exe\charset-normalizer\
Requirement already satisfied: idna<4,>=2.5 in c:\users\wecome\anaconda3\lib\site-packages\idna-3.4-py3.9-win-amd64.exe\idna\
Requirement already satisfied: certifi>=2017.4.17 in c:\users\wecome\anaconda3\lib\site-packages\certifi-2022.9.24-py3.9-win-amd64.exe\certifi\
Requirement already satisfied: patsy>=0.5.2 in c:\users\wecome\anaconda3\lib\site-packages\patsy-0.5.2-py3.9-win-amd64.exe\patsy\
Requirement already satisfied: packaging>=21.3 in c:\users\wecome\anaconda3\lib\site-packages\packaging-21.3-py3.9-win-amd64.exe\packaging\
Requirement already satisfied: colorama in c:\users\wecome\anaconda3\lib\site-packages\colorama-0.4.4-py3.9-win-amd64.exe\colorama\
Requirement already satisfied: PyWavelets in c:\users\wecome\anaconda3\lib\site-packages\PyWavelets-1.3.0-py3.9-win-amd64.exe\PyWavelets\

```

```
import pandas_profiling as pp
```

```
profile = pp.ProfileReport(df)
```

```
profile
```

```
Summarize dataset: 0%|          | 0/5 [00:00<?, ?it/s]
Generate report structure: 0%|          | 0/1 [00:00<?, ?it/s]
Render HTML: 0%|          | 0/1 [00:00<?, ?it/s]
```

Overview

Dataset statistics

Number of variables	12
Number of observations	579
Missing cells	0
Missing cells (%)	0.0%

EDA -EXPLORATORY DATA ANALYSIS

total size in memory

54.4 KiB

▸ Univariate Analysis

[] ↪ 7 cells hidden

Total Bilirubin is highly correlated with Direct Bilirubin

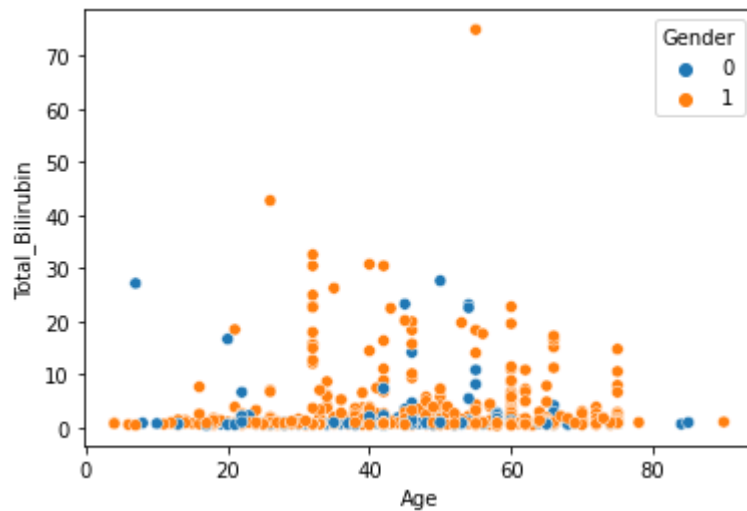
▼ Bi-Variate Analysis

```
df.head()
```

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alami
0	65	0	0.7	0.1	187	
1	62	1	10.9	5.5	699	
2	62	1	7.3	4.1	490	
3	58	1	1.0	0.4	182	
4	72	1	3.9	2.0	195	

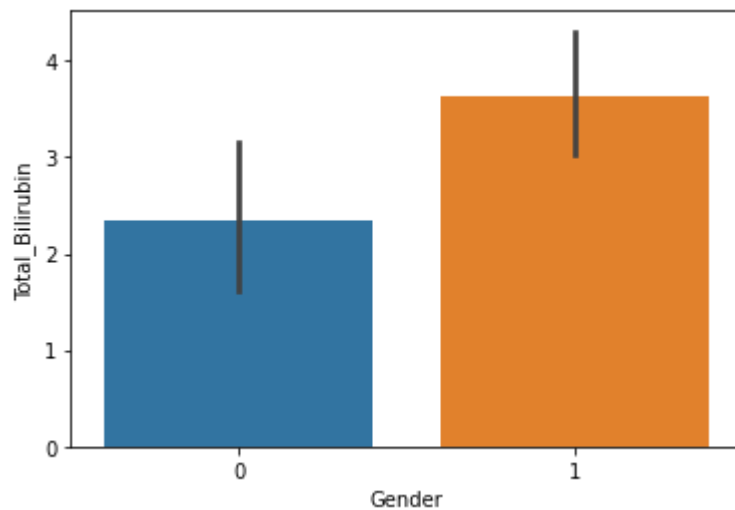
```
sns.scatterplot(x='Age',y='Total_Bilirubin',data=df,hue='Gender')
```

```
<AxesSubplot:xlabel='Age', ylabel='Total_Bilirubin'>
```



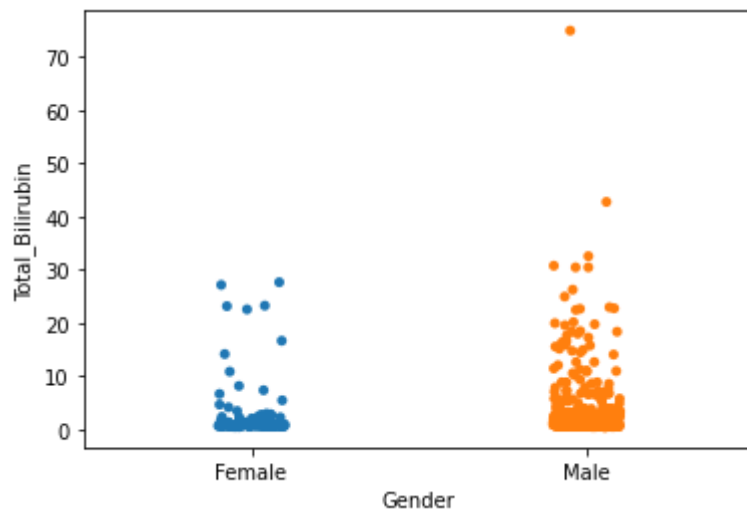
```
sns.barplot(x='Gender',y='Total_Bilirubin',data=df)
```

```
<AxesSubplot:xlabel='Gender', ylabel='Total_Bilirubin'>
```



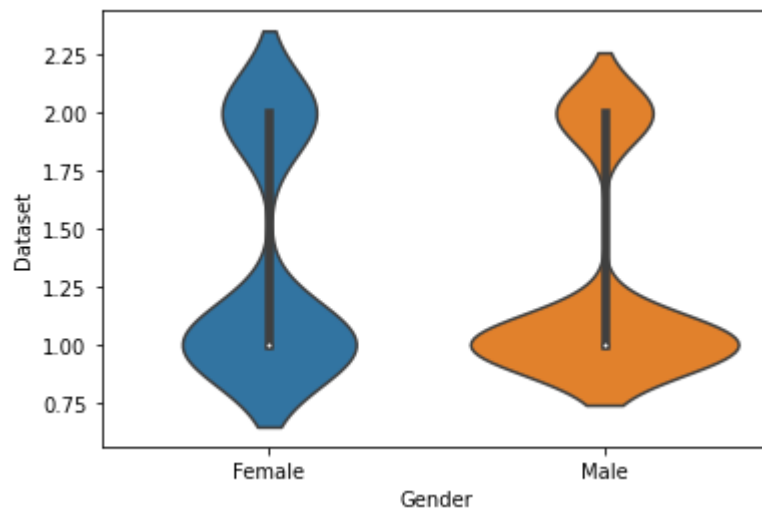
```
sns.stripplot(x='Gender',y='Total_Bilirubin',data=df)
```

```
<AxesSubplot:xlabel='Gender', ylabel='Total_Bilirubin'>
```



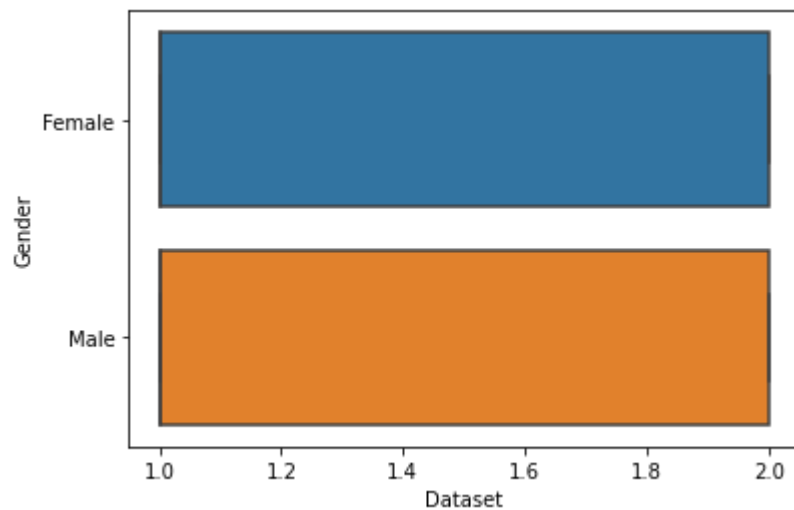
```
sns.violinplot(x='Gender',y='Dataset',data=df)
```

```
<AxesSubplot:xlabel='Gender', ylabel='Dataset'>
```



```
sns.boxplot(x='Dataset', y='Gender', data=df)
```

```
<AxesSubplot:xlabel='Dataset', ylabel='Gender'>
```



▼ Multi-Variate Analysis

```
df.head()
```

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alami
0	65	0	0.7	0.1	187	
1	62	1	10.9	5.5	699	
2	62	1	7.3	4.1	490	
3	58	1	1.0	0.4	182	
4	72	1	3.9	2.0	195	


```
sns.pairplot(data=df,hue='Gender')
```

<seaborn.axisgrid.PairGrid at 0x2536a222550>



✓ 19s completed at 18:01

● ✕