

|                        |                            |
|------------------------|----------------------------|
| <b>ASSIGNMENT DATE</b> | <b>22 SEPTEMBER 2022</b>   |
| <b>STUDENT NAME</b>    | <b>MADALA MANU VARDHAN</b> |
| <b>STUDENT ROLL NO</b> | <b>110719104029</b>        |
| <b>MAXIMUM MARKS</b>   | <b>2 MARKS</b>             |

## **ASSIGNMENT : 2**

### **Data Visualization and Pre-processing**

```
liver_data_train['HasDisease'].value_counts()
```

```
yes    416
```

```
no     167
```

```
Name: HasDisease, dtype: int64
```

### **Plotting some columns to gain insights from the data**

```
!conda install -c anaconda seaborn -y
```

First, the relationship between gender and age

```
import seaborn as sns
```

```
bins = np.linspace(liver_data_train.Age.min(), liver_data_train.Age.max(),10)
```

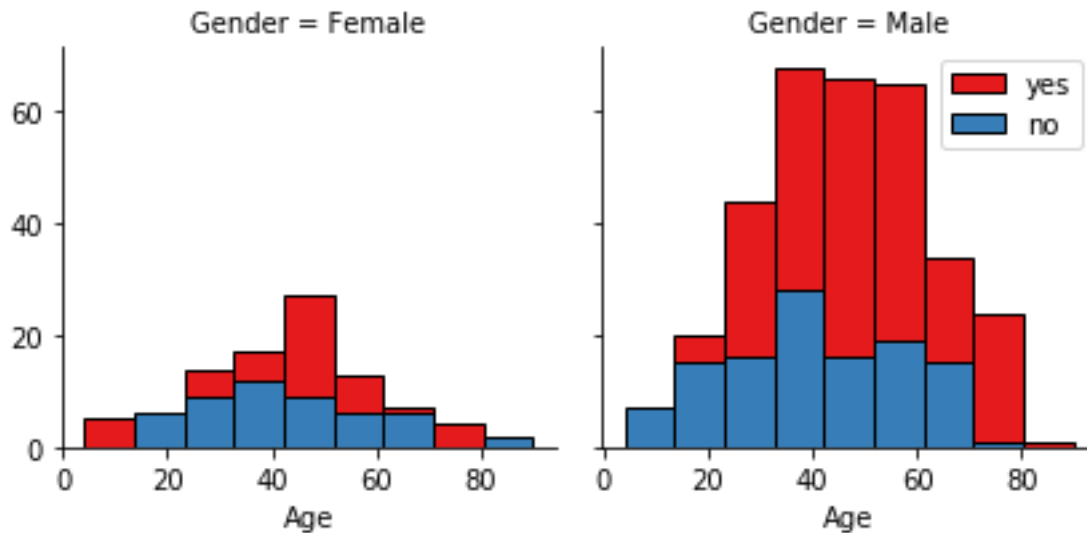
```
grids = sns.FacetGrid(liver_data_train, col="Gender", hue="HasDisease",
```

```
palette="Set1", col_wrap=2)
```

```
grids.map(plt.hist, 'Age', bins=bins, ec="k")
```

```
grids.axes[-1].legend()
```

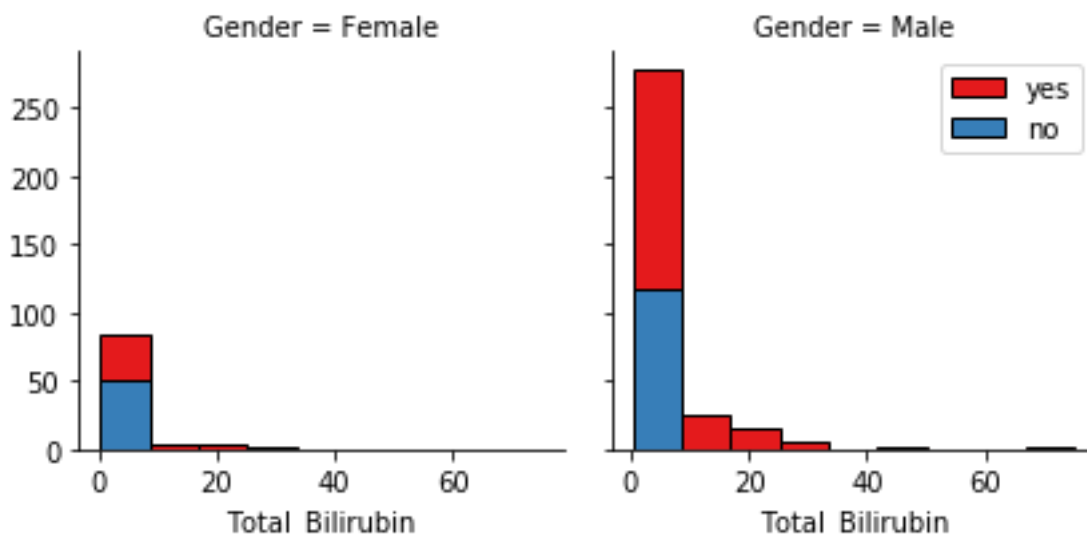
```
plt.show()
```



Amount of bilirubin based on gender

```
bins = np.linspace(liver_data_train.Total_Bilirubin.min(),
liver_data_train.Total_Bilirubin.max(),10)
grids = sns.FacetGrid(liver_data_train, col="Gender", hue="HasDisease",
palette="Set1", col_wrap=2)
grids.map(plt.hist, 'Total_Bilirubin', bins=bins, ec="k")
```

```
grids.axes[-1].legend()
plt.show()
```



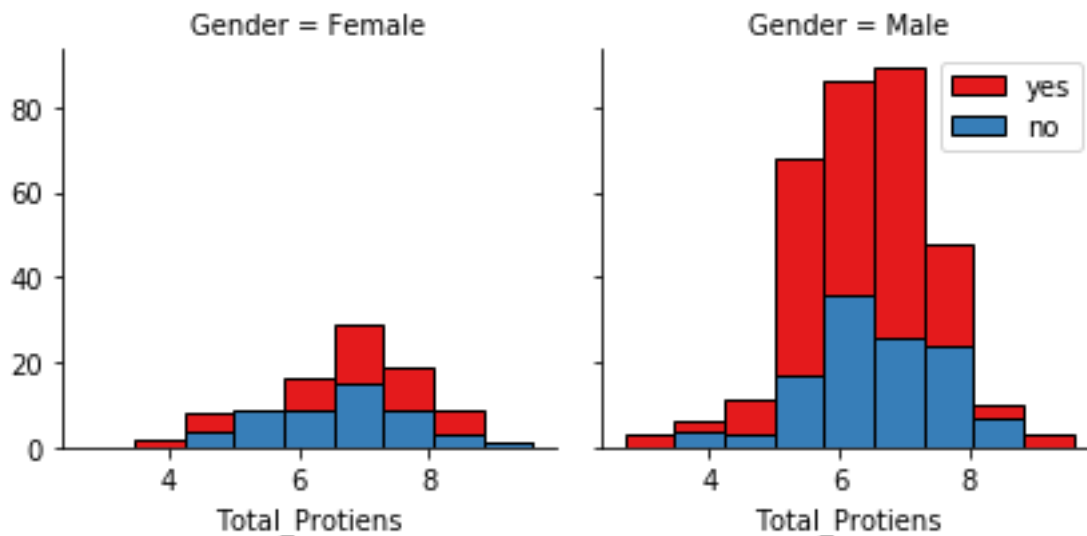
Total protiens play a major role in liver disease. So plotting it gender wise

```
bins = np.linspace(liver_data_train.Total_Protiens.min(),
liver_data_train.Total_Protiens.max(),10)
grids = sns.FacetGrid(liver_data_train, col="Gender", hue="HasDisease",
palette="Set1", col_wrap=2)
grids.map(plt.hist, 'Total_Protiens', bins=bins, ec="k")
```

```

grids.axes[-1].legend()
plt.show()

```



**convert males and females to 0 and 1**

```

liver_data_train['Gender'].replace(to_replace=['Male','Female'],
value=[0,1],inplace=True)

```

**Check if any column has null values and replace with mean**

```

liver_data_train.isna().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
HasDisease      0

```

```

dtype: int64

```

```

liver_data_train['Albumin_and_Globulin_Ratio'].fillna((liver_data_train['Albumin_and_Globulin_Ratio'].mean()), inplace=True)

```