import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

In [2]:

linkcode

data =pd.read\_csv('/kaggle/input/indian-liver-patient-records/indian\_liver\_patient.csv')

In [3]:

data.head

data.shape

data.info()

data.describe().transpose()

data.isnull().sum()

data.duplicated().sum()

data.drop\_duplicates()

data.Dataset.value\_counts()

data.shape

data['Age'].describe()

plt.figure(figsize=[12,3])

sns.boxplot(x = 'Age', data = data,color='maroon')

plt.title('Boxplot for Age Variable')

plt.show()

plt.figure(figsize=[12,4])

sns.histplot(data = data['Age'], kde = True,color='maroon')

plt.title('Histogram for Age Variable')

plt.show()

plt.figure(figsize=[12,4])

sns.countplot(y = data['Gender'],palette='Greens')

plt.title('Histogram for Age Variable')

plt.show()

plt.figure(figsize=[12,3])

sns.boxplot(x = 'Total\_Bilirubin', data = data,color='maroon')

plt.title('Boxplot for Total\_Bilirubin Variable')

plt.show()

plt.figure(figsize=[12,4])

sns.histplot(data = data['Total\_Bilirubin'], kde = True,color='maroon')

plt.title('Histogram for Total\_Bilirubin Variable')

plt.show()

plt.figure(figsize=[12,3])

sns.boxplot(x = 'Direct\_Bilirubin', data = data,color='maroon')

plt.title('Boxplot for Direct\_Bilirubin Variable')

plt.show()

plt.figure(figsize=[12,4])

sns.histplot(data = data['Direct\_Bilirubin'], kde = True,color='maroon')

plt.title('Histogram for Direct\_Bilirubin Variable')

plt.show()

plt.figure(figsize=[12,3])

sns.boxplot(x = 'Alkaline\_Phosphotase', data = data,color='maroon')

plt.title('Boxplot for Alkaline\_Phosphotase Variable')

plt.show()

plt.figure(figsize=[12,4])

sns.histplot(data = data['Alkaline\_Phosphotase'], kde = True,color='maroon')

plt.title('Histogram for Alkaline\_Phosphotase Variable')

plt.show()

plt.figure(figsize=[12,4])

sns.histplot(data = data['Alamine\_Aminotransferase'], kde = True,color='maroon')

plt.title('Histogram for Alamine\_Aminotransferase Variable')

plt.show()

plt.figure(figsize=[12,3])

sns.boxplot(x = 'Alamine\_Aminotransferase', data = data,color='maroon')

plt.title('Boxplot for Alamine\_Aminotransferase Variable')

plt.show()

plt.figure(figsize=[12,4])

sns.histplot(data = data['Aspartate\_Aminotransferase'], kde = True,color='maroon')

plt.title('Histogram for Aspartate\_Aminotransferase Variable')

plt.show()

plt.figure(figsize=[12,3])

sns.boxplot(x = 'Aspartate\_Aminotransferase', data = data,color='maroon')

plt.title('Boxplot for Aspartate\_Aminotransferase Variable')

plt.show()

plt.figure(figsize=[12,4])

sns.histplot(data = data['Total\_Protiens'], kde = True,color='maroon')

plt.title('Histogram for Total\_Protiens Variable')

plt.show()

plt.figure(figsize=[12,3])

sns.boxplot(x = 'Total\_Protiens', data = data,color='maroon')

plt.title('Boxplot for Total\_Protiens Variable')

plt.show()

plt.figure(figsize=[12,4])

sns.histplot(data = data['Albumin'], kde = True,color='maroon')

plt.title('Histogram for Albumin Variable')

plt.show()

plt.figure(figsize=[12,3])

sns.boxplot(x = 'Albumin', data = data,color='maroon')

plt.title('Boxplot for Albumin Variable')

plt.show()

plt.figure(figsize=[12,4])

sns.histplot(data = data['Albumin\_and\_Globulin\_Ratio'], kde = True,color='maroon')

plt.title('Histogram for Albumin\_and\_Globulin\_Ratio Variable')

plt.show()

plt.figure(figsize=[12,3])

sns.boxplot(x = 'Albumin\_and\_Globulin\_Ratio', data = data,color='maroon')

plt.title('Boxplot for Albumin\_and\_Globulin\_Ratio Variable')

plt.show()

plt.figure(figsize=[12,4])

sns.countplot(y = data['Dataset'],palette='Greens')

plt.title('countplot for Dataset Variable')

plt.show()

data.corr()

plt.figure(figsize = [20,8])

sns.heatmap(data.corr(),annot=True,cmap='magma', vmin=-1, vmax=1)

plt.figure(figsize=[12,4])

sns.countplot(x = data['Dataset'],hue=data['Gender'],palette='magma')

plt.title('countplot for Dataset Variable')

plt.show()

data.info()

var\_list=['Total\_Bilirubin', 'Direct\_Bilirubin',

'Alkaline\_Phosphotase', 'Alamine\_Aminotransferase',

'Aspartate\_Aminotransferase', 'Total\_Protiens', 'Albumin',

'Albumin\_and\_Globulin\_Ratio']

def draw\_scattterplots(df, variables, n\_rows, n\_cols):

fig=plt.figure(figsize = [20,20])

for i, var\_name **in** enumerate(variables):

ax=fig.add\_subplot(n\_rows,n\_cols,i+1)

sns.scatterplot(x=df[var\_name],y=df[var\_name],ax=ax)

ax.set\_title(var\_name+" Distribution")

fig.tight\_layout() *# Improves appearance a bit.*

plt.show()

*#test = pd.DataFrame(np.random.randn(30, 9), columns=map(str, range(9)))*

draw\_scattterplots(data, data[var\_list],8,4)

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Total\_Bilirubin'],y=data['Direct\_Bilirubin'],hue=data['Dataset'],palette='deep')

plt.title('Billirunin (Direct Vs Total)')

plt.show()

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Total\_Bilirubin'],y=data['Alkaline\_Phosphotase'],hue=data['Dataset'],palette='deep')

plt.title('Billirunin Vs Alkaline\_Phosphotase')

plt.show()

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Total\_Bilirubin'],y=data['Alamine\_Aminotransferase'],hue=data['Dataset'],palette='deep')

plt.title('Billirunin Vs Alamine\_Aminotransferase')

plt.show()

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Total\_Bilirubin'],y=data['Aspartate\_Aminotransferase'],hue=data['Dataset'],palette='deep')

plt.title('Billirunin Vs Aspartate\_Aminotransferase')

plt.show()

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Total\_Bilirubin'],y=data['Total\_Protiens'],hue=data['Dataset'],palette='deep')

plt.title('Billirunin Vs Total\_Protiens')

plt.show()

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Total\_Bilirubin'],y=data['Albumin'],hue=data['Dataset'],palette='deep')

plt.title('Billirunin Vs Albumin')

plt.show()

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Total\_Bilirubin'],y=data['Albumin\_and\_Globulin\_Ratio'],hue=data['Dataset'],palette='deep')

plt.title('Billirunin Vs Albumin\_and\_Globulin\_Ratio')

plt.show()

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Direct\_Bilirubin'],y=data['Alamine\_Aminotransferase'],hue=data['Dataset'],palette='deep')

plt.title('Direct\_Bilirubin Vs Alkaline\_Phosphotase')

plt.show()

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Total\_Bilirubin'],y=data['Alamine\_Aminotransferase'],hue=data['Dataset'],palette='deep')

plt.title('Total\_Bilirubin Vs Alamine\_Aminotransferase')

plt.show()

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Total\_Bilirubin'],y=data['Aspartate\_Aminotransferase'],hue=data['Dataset'],palette='deep')

plt.title('Total\_Bilirubin Vs Aspartate\_Aminotransferase')

plt.show()

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Total\_Bilirubin'],y=data['Albumin'],hue=data['Dataset'],palette='deep')

plt.title('Total\_Bilirubin Vs Albumin')

plt.show()

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Total\_Bilirubin'],y=data['Alkaline\_Phosphotase'],hue=data['Dataset'],palette='deep')

plt.title('Total\_Bilirubin Vs Alkaline\_Phosphotase')

plt.show()

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Total\_Bilirubin'],y=data['Albumin\_and\_Globulin\_Ratio'],hue=data['Dataset'],palette='deep')

plt.title('Total\_Bilirubin Vs Albumin\_and\_Globulin\_Ratio')

plt.show()

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Alkaline\_Phosphotase'],y=data['Alamine\_Aminotransferase'],hue=data['Dataset'],palette='deep')

plt.title('Alkaline\_Phosphotase Vs Alamine\_Aminotransferase')

plt.show()

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Alkaline\_Phosphotase'],y=data['Aspartate\_Aminotransferase'],hue=data['Dataset'],palette='deep')

plt.title('Alkaline\_Phosphotase Vs Aspartate\_Aminotransferase')

plt.show()

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Alkaline\_Phosphotase'],y=data['Total\_Protiens'],hue=data['Dataset'],palette='deep')

plt.title('Alkaline\_Phosphotase Vs Total\_Protiens')

plt.show()

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Alkaline\_Phosphotase'],y=data['Albumin'],hue=data['Dataset'],palette='deep')

plt.title('Alkaline\_Phosphotase Vs Albumin')

plt.show()

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Alkaline\_Phosphotase'],y=data['Albumin\_and\_Globulin\_Ratio'],hue=data['Dataset'],palette='deep')

plt.title('Alkaline\_Phosphotase Vs Albumin\_and\_Globulin\_Ratio')

plt.show()

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Alamine\_Aminotransferase'],y=data['Aspartate\_Aminotransferase'],hue=data['Dataset'],palette='deep')

plt.title('Alamine\_Aminotransferase Vs Aspartate\_Aminotransferase')

plt.show()

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Alamine\_Aminotransferase'],y=data['Total\_Protiens'],hue=data['Dataset'],palette='deep')

plt.title('Alamine\_Aminotransferase Vs Total\_Protiens')

plt.show()

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Alamine\_Aminotransferase'],y=data['Albumin'],hue=data['Dataset'],palette='deep')

plt.title('Alamine\_Aminotransferase Vs Albumin')

plt.show()

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Alamine\_Aminotransferase'],y=data['Albumin\_and\_Globulin\_Ratio'],hue=data['Dataset'],palette='deep')

plt.title('Alamine\_Aminotransferase Vs Albumin\_and\_Globulin\_Ratio')

plt.show()

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Aspartate\_Aminotransferase'],y=data['Total\_Protiens'],hue=data['Dataset'],palette='deep')

plt.title('Aspartate\_Aminotransferase Vs Total\_Protiens')

plt.show()

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Aspartate\_Aminotransferase'],y=data['Albumin'],hue=data['Dataset'],palette='deep')

plt.title('Aspartate\_Aminotransferase Vs Albumin')

plt.show()

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Aspartate\_Aminotransferase'],y=data['Albumin\_and\_Globulin\_Ratio'],hue=data['Dataset'],palette='deep')

plt.title('Aspartate\_Aminotransferase Vs Albumin\_and\_Globulin\_Ratio')

plt.show()

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Total\_Protiens'],y=data['Albumin'],hue=data['Dataset'],palette='deep')

plt.title('Total\_Protiens Vs Albumin')

plt.show()

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Total\_Protiens'],y=data['Albumin\_and\_Globulin\_Ratio'],hue=data['Dataset'],palette='deep')

plt.title('Total\_Protiens Vs Albumin\_and\_Globulin\_Ratio')

plt.show()

plt.figure(figsize=[12,4])

sns.scatterplot(x = data['Albumin'],y=data['Albumin\_and\_Globulin\_Ratio'],hue=data['Dataset'],palette='deep')

plt.title('Albumin Vs Albumin\_and\_Globulin\_Ratio')

plt.show()

*#plot 1:*

plt.subplot(1, 2, 1)

sns.set\_style("whitegrid")

sns.boxplot(x = 'Dataset', y = 'Total\_Bilirubin', data = data)

*#plot 2:*

plt.subplot(1, 2, 2)

sns.set\_style("whitegrid")

sns.boxplot(x = 'Gender', y = 'Total\_Bilirubin', data = data)

plt.tight\_layout()

plt.suptitle("Boxplot for Total\_billirubin")

plt.show()

*#plot 1:*

plt.subplot(1, 2, 1)

sns.set\_style("whitegrid")

sns.boxplot(x = 'Dataset', y = 'Direct\_Bilirubin', data = data)

*#plot 2:*

plt.subplot(1, 2, 2)

sns.set\_style("whitegrid")

sns.boxplot(x = 'Gender', y = 'Direct\_Bilirubin', data = data)

plt.tight\_layout()

plt.suptitle("Boxplot for Direct\_Bilirubin")

plt.show()

*#plot 1:*

plt.subplot(1, 2, 1)

sns.set\_style("whitegrid")

sns.boxplot(x = 'Dataset', y = 'Alkaline\_Phosphotase', data = data)

*#plot 2:*

plt.subplot(1, 2, 2)

sns.set\_style("whitegrid")

sns.boxplot(x = 'Gender', y = 'Alkaline\_Phosphotase', data = data)

plt.tight\_layout()

plt.suptitle("Boxplot for Alkaline\_Phosphotase")

plt.show()

*#plot 1:*

plt.subplot(1, 2, 1)

sns.set\_style("whitegrid")

sns.boxplot(x = 'Dataset', y = 'Alamine\_Aminotransferase', data = data)

*#plot 2:*

plt.subplot(1, 2, 2)

sns.set\_style("whitegrid")

sns.boxplot(x = 'Gender', y = 'Alamine\_Aminotransferase', data = data)

plt.tight\_layout()

plt.suptitle("Boxplot for Alamine\_Aminotransferase")

plt.show()

*#plot 1:*

plt.subplot(1, 2, 1)

sns.set\_style("whitegrid")

sns.boxplot(x = 'Dataset', y = 'Aspartate\_Aminotransferase', data = data)

*#plot 2:*

plt.subplot(1, 2, 2)

sns.set\_style("whitegrid")

sns.boxplot(x = 'Gender', y = 'Aspartate\_Aminotransferase', data = data)

plt.tight\_layout()

plt.suptitle("Boxplot for Aspartate\_Aminotransferase")

plt.show()

*#plot 1:*

plt.subplot(1, 2, 1)

sns.set\_style("whitegrid")

sns.boxplot(x = 'Dataset', y = 'Total\_Protiens', data = data)

*#plot 2:*

plt.subplot(1, 2, 2)

sns.set\_style("whitegrid")

sns.boxplot(x = 'Gender', y = 'Total\_Protiens', data = data)

plt.tight\_layout()

plt.suptitle("Boxplot for Total\_Protiens")

plt.show()

*#plot 1:*

plt.subplot(1, 2, 1)

sns.set\_style("whitegrid")

sns.boxplot(x = 'Dataset', y = 'Albumin', data = data)

*#plot 2:*

plt.subplot(1, 2, 2)

sns.set\_style("whitegrid")

sns.boxplot(x = 'Gender', y = 'Albumin', data = data)

plt.tight\_layout()

plt.suptitle("Boxplot for Albumin")

plt.show()

*#plot 1:*

plt.subplot(1, 2, 1)

sns.set\_style("whitegrid")

sns.boxplot(x = 'Dataset', y = 'Albumin\_and\_Globulin\_Ratio', data = data)

*#plot 2:*

plt.subplot(1, 2, 2)

sns.set\_style("whitegrid")

sns.boxplot(x = 'Gender', y = 'Albumin\_and\_Globulin\_Ratio', data = data)

plt.tight\_layout()

plt.suptitle("Boxplot for Albumin\_and\_Globulin\_Ratio")

plt.show()