**Assignment**

1. Use the loan prediction dataset csv file and perform EDA analysis for the loan prediction.

Do list of activities:

1. Read the CSV data using pandas library.

--> import io

from google.colab import files

data = files.upload()

df = pd.read\_csv(io.BytesIO(data['dataloan (1).csv']))

1. Print the full data set.

--> df

1. Print the total number of rows and columns.

--> df.shape

(614, 13)

1. Print number of rows only.

--> df.shape[0]

614

1. Print number of columns only.

--> df.shape[1]

13

1. Print the first five records.

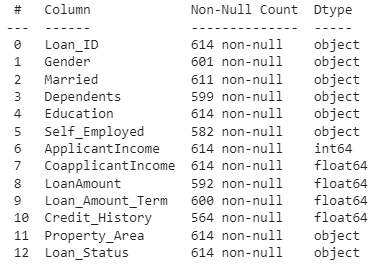
--> df.head()

1. Print the last five records.

--> df.tail()

1. Print the all non-null columns using info().

--> df.info()



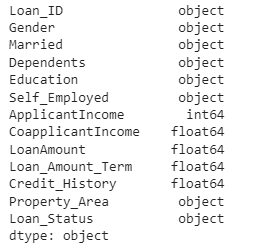
1. Print all column names.

--> df.columns

Index(['Loan\_ID', 'Gender', 'Married', 'Dependents', 'Education', 'Self\_Employed', 'ApplicantIncome', 'CoapplicantIncome', 'LoanAmount', 'Loan\_Amount\_Term', 'Credit\_History', 'Property\_Area', 'Loan\_Status'], dtype='object')

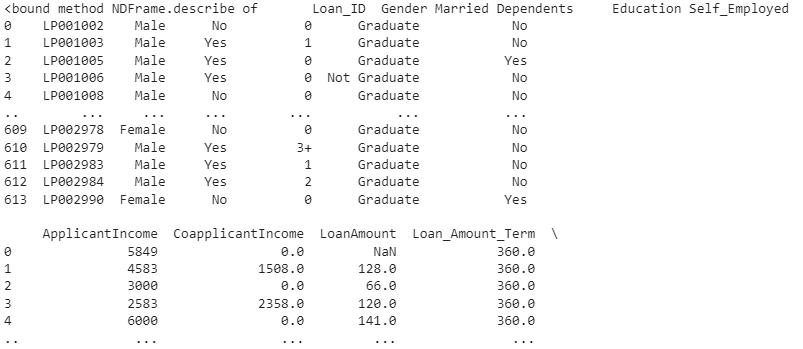
1. Print all datatypes of respective column names.

--> df.dtypes



1. Print the descriptive statistics of data frame.

--> df.describe



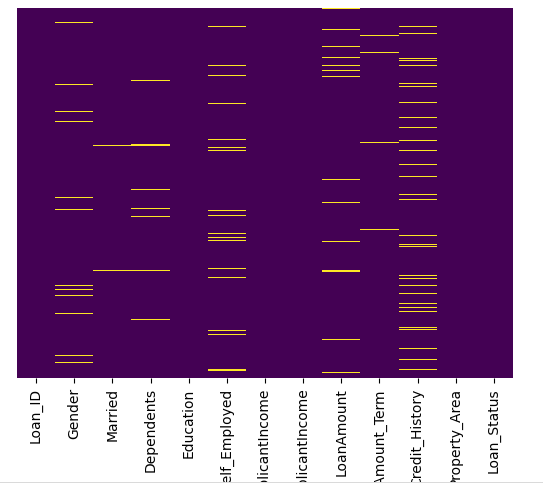
1. Print the sum of NAN values.

--> df.isna().sum().sum()

149

1. Plot the necessary graphs for data visualization like heatmap, countplot.

--> sns.heatmap(df.isnull(),yticklabels=False,cbar=False,cmap='viridis')



1. Fill the NAN values, compute mean of the column, and apply it.

--> mean\_value=df['Loan\_Amount\_Term'].mean()

df['Loan\_Amount\_Term'].fillna(value = mean\_value,inplace=True)

1. Convert the string values into categorical features using pd.get\_dummies()

--> Married=pd.get\_dummies(df['Married'],drop\_first=True)

Loan\_Sts=pd.get\_dummies(df['Loan\_Status'],drop\_first=True)

1. Concatenate the categorical features to the existing dataset.

--> df = pd.concat([df,Married,Loan\_Sts],axis=1)

1. Print the independent and dependent variable.

--> y=df["Loan\_Status"]

x=df.drop("Loan\_Status",axis=1)