Assignment observation:

step 1: Download anaconda software in your local machine

step2: After installing software now creating new env for our assignment **conda create -n habermanenv python=3.7.** after created environment you have to activate environment through this command **activate habermanenv(env\_name).**

Step3: navigate to the assignment folder and download **jupyter notebook, pandas, numpy, matplotlib and seaborn library** by the help of this command

**conda install jupyter notebook, conda install pandas, conda install numpy, conda install matplotlib and conda install seaborn.**

Step4**:** then after launch **jupyter notebook** in the anaconda terminal**.**

Step5: import necessary library library like **import numpy as np, import pandas as pd, import matplotlib.pyplot as plt** and **import seaborn as sns.**

Step6: now, read csv file by the help of pandas library.

Step7: in ou dataset there is no column names then defining column names by the help of **kaggle dataset description**.

Step8: now, check is any null values in our dataset or not by the help of this command **df.isnull().sum().**

Step9: my dependent variable have **datatype int** we have to convert it into **categorical datatype.**

Step10: Now Univariate analysis by the help of Probability density function(PDF), cumulative density function(CDF), Box plot and violin plot.

**Boxplot** is used for check the outliers in our dataset. I am using **Interquartile range(IQR)** technique to handle outliers through our dataset.

**Violinplot** is the combination of box plot and probability density function.

Step11: Now Bi-variate analysis by the help of Scatterplot and Pairplot.

Step12: define **x** and **y** before **train and test split.**

Step13: now its time to train and test split our dataset and pass some argument like **x,** **y , test\_size** and **random\_state** where **x is independent variable , y is dependent variable and test\_size is 30%** and **handle random** value giving **random\_state=10.**

Step14: At last build model by the the help **Logistic regression algorithm** and **check accuracy.**