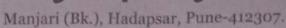


## Pune District Education Association's College Of Engineering





Accredited by NAAC

DSBDAL	Assignment No: 08
*	Aim: - Data Visualization II  1) use the inbuilt dataset titanic! The dataset contains 891 rows and contains information about the passangers who boarded the unfortunate Titanic Ship, use the seaborn library to see if we can find any
	patterns in the data. बहुजन हिताय, बहुजन सुखाय।
	2) Write a code to check how the price of the ticket for each passanger is distributed by plotting a histogram.
	seaborn is a library mostly used for statistical plotting in python. It is built on top of motplot to and providers beautiful default styles of color paletters to make statistical plots more attractive.
	Different categories of plot in seaborn.  1) Relational plots:-  This plot is used to understand the relation between two variables.
2	categorical plots -

This plot deals with categorical variable and hold they can be visualized. This plot is used for examining univariate 3) distribution plots: and biv aviate distribution. 4) Regression plots: The regression plots in seabour are primarily intended to add a visual guide that helps to emphasize patterns in a dataset during exploratory data and lyres. S) matrix prots:-A matrix plot is an array of scatter 6) multi-plot -goids:-It is useful approach is to down multiple instances of the same plot an different subsets of the datasets. The datasets consists of 891 mils of Dpassenger Id 2) syrvived 3) polass 4) Name

## istrict Education Association's College Of Engineering

Manjari (Bk.), Hadapsar, Pune-412307.



Accredited by NAAC

- 7) SibSP
  - 8) paach
    - 9) Ticket
  - (0) fase
  - Embasked

The seaboun library is built on top or matplotlib and offers many advanced data Visualization capabilities.

\* features:-

The titanic dataset has roughly the following types of features:

· categorical Inominal:

variables that can be divided into mytiple categories but having no order for priority.

E.g. Embarked (c=cherbourg; Q= aypen stoldn; S= South ampton)

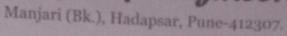
· Bingry :

A subtype of categorical features, where the variable has two categories. E.g. Sex ( male / female)

· ordinal:-They are similar to categorical feature but they have an order (ie. can be started). E.g. polass (1,2,3) · continous :-They can take up any value beth the minimum & maximum values in a column. E.g. Age, fare They represent the count of a variable E.g. - Sibsp , parch. · USE 1 - 55! -They don't contribute to the final of an mi model. · pistaibution plots: pistribution plots, as the name suggests goe type of plot that should the Stastical distaibution of data. The dist plot () shows the histogram distribution of data for a single column. We can see that most of the Ackets have been solved between the 0-50 dollars.



## Pune District Education Association's College Of Engineering





## Accredited by NAAC

The line that you see represents the kemel density estimation you can remove this line by passing false as the parameter for the kde attribute as shown below.

sns. distplot (dotaset ['fare'], kde:fase)

sns. distprot (dataset ['fare'], kde = false, bins = 10)

Here we set the no. of bins to 10. In the olp, you will see data distributed in 10 bins as shown below.

\* Histogram:-

-Historyrams are visualization tools that represents the distribution of set of continuous data.

-In a histogram, the data is divided into a set of Intervals or bin & the count of data points that fall into each bin corresponding to the height of the bar above that bin.

Syntax:-

seaborn histoplot (data x , 1 , hue,

stat, bins, bin width, discorte, Kde, log. scale). \* parameters: · data = input data in the form of pataforms or Humpy good. \*x, y (optional) = key of the data to be positioned on the x & y axes respectively. · hue (optional) - semantic data key which is mapped to determine the color. · stat coptional) = count, frequency, etc. Return: - This method returns the matprotlib 9x2s with plot doalen on