

EXPERIMENT No. 09

Aim:- Data visualization II

1. Use the input dataset 'titanic' as used in above problem plot a box for distribution of age with respect to each gender age with information about whether they survived or not.
2. Write observations on the inference from above statistics.

Objective:- To perform Data Visualization.

THEORY:-

Data Visualization play a very important role in Data mining various data scientists spent their time exploring data through visualization. To accelerate the process we need to explore the plots & libraries.

1. Seaborn library:-

seaborn is python data visualization library based on matplotlib. It provides ~~at~~ high level inference for drawing attractive

for installation of seaborn

`pip install`

`seaborn` conda

`install seaborn`

to import seaborn:

02

Import seaborn as sns.

2. Matplotlib library:-

- Matplotlib popular python library used for creating static, animated & interactive visualizations in data analysis. It provides a variety of plotting tools, making it easy to generate graphs, charts & plots.

Import matplotlib.pyplot as plt.

3. Know your data.

Import seaborn as sns

Import matplotlib.pyplot as plt.

to load data

df = sns.load_dataset('titanic')

df

dataset contains 891 rows & 15 columns & contain information about passengers who boarded the unfortunate titanic ship. The original task is to predict whether or not passenger survived or not.

Box plot

- Box plot is also known as box & whisker plot.
- A box plot is graphical representation of distribution of dataset.

The smallest data point excluding outliers
- first quartile (Q_1).

ii. First Quartile (Q_1)
- The median of lower half.
- 25th percentile.

iii. Median (Q_2)
- The middle value of dataset
- 50th percentile.

iv. Third Quartile (Q_3)
- 75th percentile.
- The median of upper half of dataset.

v. Maximum :-
- The largest data points excluding outliers.

- Box plots are useful for identify the spread, central tendency & outliers.

- How to create?
`plt.boxplot(df)`
`plt.title('Boxplot')`

→ To show.

`plt.show()`.

What is outliers?

- Outliers are the data points that are significantly different from rest of data in dataset.

Applications-

- I. Visualization of distribution data.
- II. Detect & understand outliers.

Conclusion:-

Seaborn is advanced data visualization library built on top of matplotlib library.

EXPERIMENT No:- 10

Aim:- Data Visualization III

Download Iris flower dataset or any other dataset into a dataframe. Scan the dataset & give inference as:-

- I. List down the features of their types available in dataset.
- II. Create a histogram for such features in this dataset to illustrate features distribution.
- III. Create boxplot for each feature in dataset.

Objective :-

To perform Data Visualization.

Theory:-

Data Visualization plays important role in Data mining various data scientist spent their time.

1. Seaborn Library:-

— Seaborn is python data visualization library based on matplotlib.

import seaborn as sns.

2. Matplotlib Library :-

- Import matplotlib.

3. Know data.

- The dataset contains 100 rows & columns. that contains sepal width, sepal length, petal width.
- List of datatypes of each features.
The dtype method or attribute is used in the context of pandas & numpy where it refers to datatype.

Syntax -

print (df.dtypes).

- Histograms:-

- A histogram is graphical representation of distribution of dataset.

- How to plot histogram?

```
plt.hist (data, bins=5, edgecolor='r')
plt.title ('Histogram Example')
plt.xlabel ('Data Volumes')
plt.ylabel ('Frequency')
```


→ To show plot:-

~~plt. hist~~ plt. show().

* Box plot:-

A Boxplot also known as box-and-whisker plot.

A Boxplot is graphical representation of the distribution of dataset.

* Outliers:-

- Outliers are data points that are significantly different from the rest.
- These points are usually high or low compared.

* Application:-

1. Visualize distribution of data.
11. Detect and understand outliers.

* Conclusion:-

seaborn is advanced data visualization library built on top on matplotlib library.