Lab assignment 4

Group B

→ Title : Visualize the data using python libraries matplotlib, seaborn by plotting the various types of graphs.

→ Objectives : To understand and apply analytical concept of big data.

→ Theory :

a) R - Pie charts -

- In R, the pie chart is created using the pie() function which takes positive numbers as a vector input.
- The additional parameters are used to control labels, color, title, etc..

Syntax -

pie (x, labels, radius, main, col, clockwise)

b) R - Bar charts -

- A bar chart represents data in rectangular bars with length of the bar proportional to the value of the variable.
- R uses the function barplot() to create bar charts.
- R can draw both vertical and horizontal bars in bar chart.
- In bar chart, each of bar can be given different colors.

Syntax -

barplot (H, xlab, main, names.arg, col)

where,

names.arg is vector of names appearing under each bar

col is used to give colors to bars in graph.

c) R - boxplots -

- Boxplots are a measure of how well distributed is the data.
- It divides the dataset into three quartiles.
- It represents the minimum, maximum, median, first quartile and third quartile in the dataset.
- It is also useful in comparing distribution of data across datasets.

Syntax -

boxplot (x, data, notch, varwidth, names, main)

d) R - histograms -

- A histogram represents the frequencies of values of variable bucketed into ranges.
- Histogram is similar to bar chart but difference is it groups the values into continuous range.
- Each bar represents the height of number of values present in that range.
- R creates histogram using hist() function.

Syntax -

hist (v, main, xlab, xlim, ylim, breaks, col, border)

e) R - line graphs -

- A line chart is graph that connects a series of points by drawing line segments between them.
- These points are ordered in one of their co-ordinate value.
- Line charts are usually used in identifying trends in data.
- The plot function in R is used to create line graph.

Syntax -

plot (v, type, col, xlab, ylab)

f) R - scatterplots -

- Scatterplots show many points plotted in the Cartesian plane.
- Each point represents values of two variables.
- One variable is chosen in horizontal axis and other in vertical axis.
- The simple scatterplot is created using plot() function.

Syntax -

plot (x, y, main, xlab, ylab, xlim, ylim, axes)

g) Scatterplot matrices -

- When we have more than two variables and we want to find correlation between one variable versus remaining ones we use scatterplot matrix.
- We use pairs() function to create matrices of scatterplot.

Syntax -

pairs (formula, data)

→ Conclusion : Thus we have learnt visualization of data using R / python by plotting various types of graphs.