Assignment: 7

Aim: Visualize the data using R/Python by plotting the graphs for assignment no. 2 and 3 (Group B)

OBJECTIVE:

1. To understand and apply the analytical concept of Big data using Python.
2. To study detailed data visualization techniques in Python programing.

THEORY:

INTRODUCTION TO VISUALIZATION

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

We can use Matplotlib and Seaborn library for data Visualization. Matplotlib is one of the most popular Python packages used for data visualization. Itis across-platform library for making 2D plots from data in arrays. Seaborn is a library for making statistical graphics in Python. It builds on top of matplotlib and integrates closely with pandas data structures. Seaborn helps you explore and understand your data.

LineChart

A line chart is a type of chart that displays information as a series of data points connected by straight line segments. A line chart is a way of visually representing an asset's price history using a single, continuous line.

Barplot

Definition. A barplot (or barchart) is one of the most common types of graphic. It shows the relationship between a numeric and a categoric variable. Each entity of the categoric variable is represented as a bar. The size of the bar represents its numeric value.

Countplot

The countplot is used to represent the occurrence(counts) of the observation present in the Categorical variable. It uses the concept of a bar chart for the visual depiction.

Boxplot

A boxplot is a standardized way of displaying the distribution of data based on its five-number summary ("minimum", first quartile [Q1], median, third quartile [Q3] and "maximum"). Boxplots can tell you about your outliers and their values, if your data is symmetrical, how tightly your data is grouped and if and how your data is skewed.

Histogram

A histogram graph is a type of graph that uses rectangular bars to represent the Frequency of discrete and continuous data. The rectangular bars represent the number of data points that fall within a certain class interval.

Distplot

Seaborn distplot lets you show a histogram with a line on it. This can be shown in all kinds of variations. We use seaborn in combination with matplotlib, the Python plotting module. A distplot plots a univariate distribution of observations. The distplot() function combines the matplotlib hist function with the seaborn kdeplot() and rugplot() functions.

Heatmap

Heatmaps organize data in a grid, with different colors or shades indicating different levels of the data's magnitude. The visual nature of heatmaps allows for immediate recognition of patterns, such as clusters, trends, and anomalies. This makes heatmaps an effective tool for exploratory data analysis.

Scatterplot

A scatter plot is a diagram where each value in the data set is represented by a dot. The Matplotlib module has a method for drawing scatter plots, it needs two arrays of the same length, one for the values of the x-axis, and one for the values of the y-axis.

Pairplot

A pairs plot, also known as a scatterplot matrix, is a grid of scatterplots that displays pairwise relationships between multiple variables in a dataset. Each cell in the grid represents the relationship between two variables, and the diagonal cells display histograms or kernel density plots of individual variables.

Piechart

A pie chart is a type of graph representing data in a circular form, with each slice of the circle representing a fraction or proportionate part of the whole. All slices of the pie add up to make the whole equaling 100 percent and 360 degrees. The pie chart represents data in a pictorial form, making it easier to visualize and understand the proportionate parts or composition of a data set. It's also sometimes referred to as a circle chart.

Conclusion

We can use Matplotlib, Seaborn tools in python for data visualization.