MIT | Academy of Engineering

(An Autonomous Institute Affiliated to Savitribai Phule Pune University)

Name : Aryan kolpe

Branch : Mechanical(ME1)

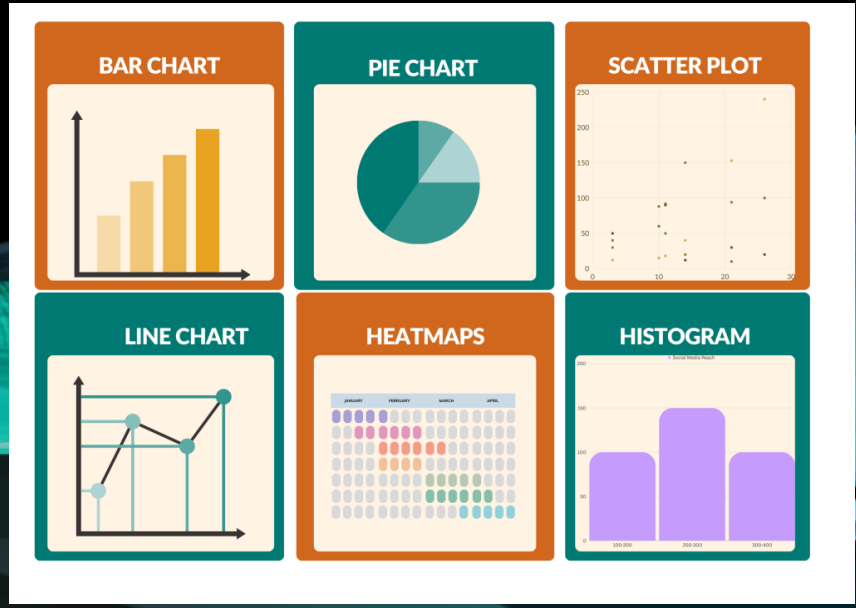
Roll no.: ME8

PRN : 202401090067

Subject : Essential of Data Science (EDS)

Subject Teacher : Prof. Sopan Kshirsagar

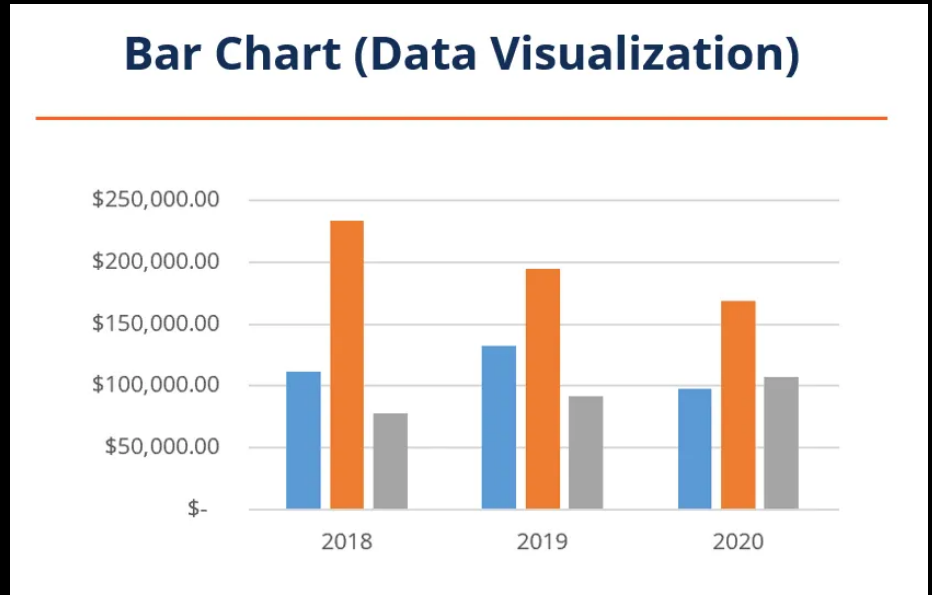
BASIC GRAPHS REQUIRED FOR DATA VISUALIZATION.



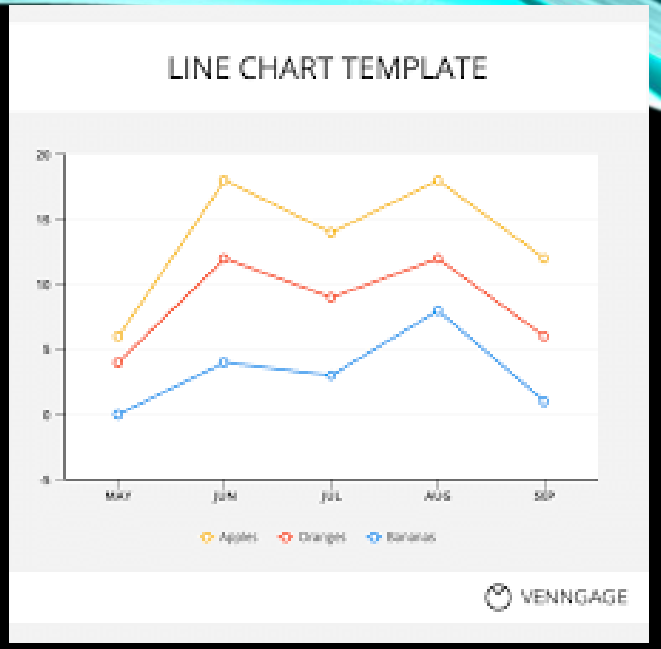
1. BAR CHART (OR COLUMN CHART):

- **Significance:** Bar charts excel at comparing categorical data. The length (or height) of each bar represents the magnitude or frequency of a particular category. They allow for easy visual comparison between distinct groups.

- **Use Cases:** Comparing sales figures across different product lines, showing the number of students in various departments, illustrating survey responses for different options.



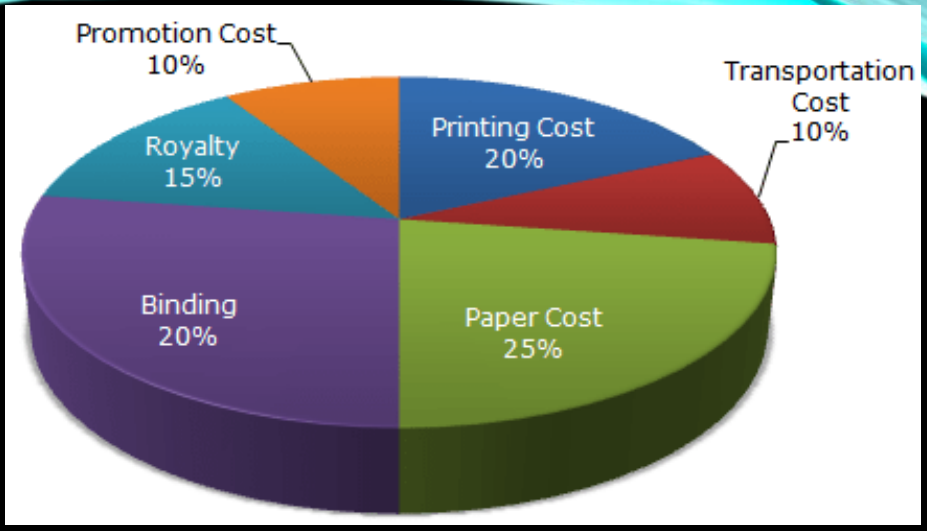
2. LINE CHART:



- **Significance:** Line charts are invaluable for displaying trends and changes over a continuous variable, most commonly time. The connected data points clearly illustrate the progression and fluctuations in the data.

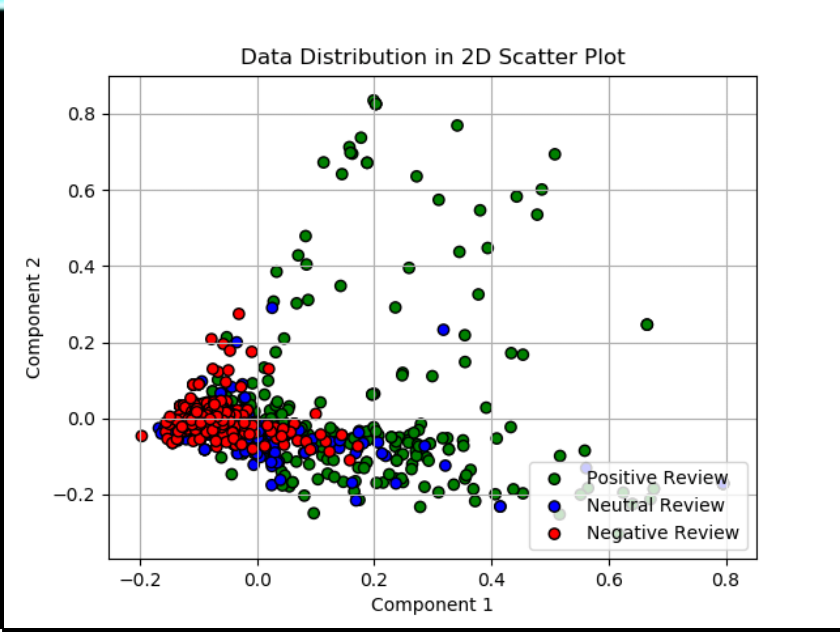
- **Use Cases:** Tracking stock prices over time, showing temperature changes throughout the day, illustrating the growth of a population over several years.

3. PIE CHART:



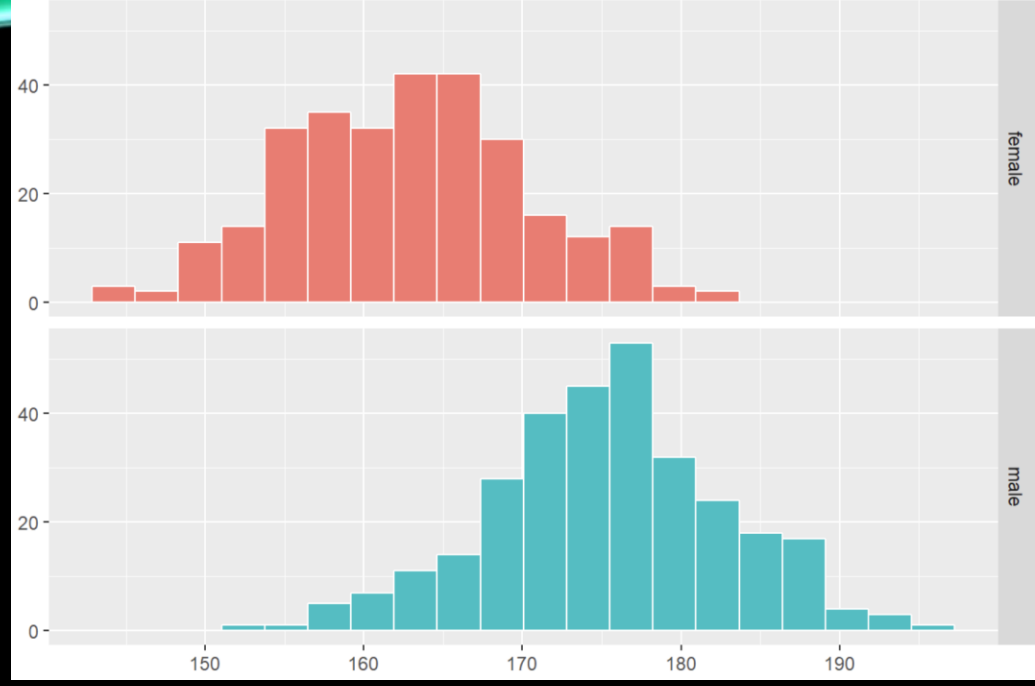
- **Significance:** Pie charts are used to show the proportion or percentage of different parts that make up a whole. Each slice represents a category, and the size of the slice is proportional to its contribution to the total.
- **Use Cases:** Showing the market share of different companies, illustrating the breakdown of a budget into different expense categories, representing the distribution of votes in an election.
- **Caution:** Pie charts can become difficult to interpret when there are many categories or when the proportions are similar. Bar charts are often a better alternative in such cases

4. SCATTER PLOT:



- **Significance:** Scatter plots are powerful for visualizing the relationship between two quantitative variables. Each point on the plot represents a pair of values, and the pattern of these points can reveal correlations, clusters, or outliers.
- **Use Cases:** Examining the relationship between advertising spend and sales revenue, exploring the correlation between height and weight, identifying potential outliers in a dataset

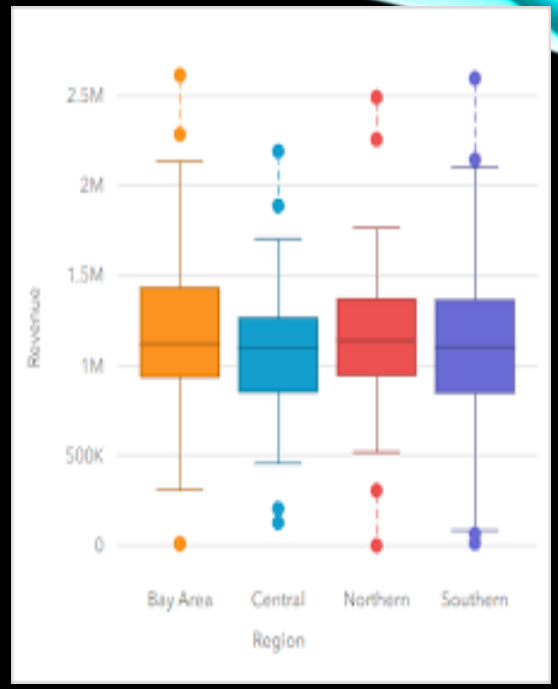
5. HISTOGRAM:



- **Significance:** Histograms display the distribution of a single quantitative variable. The data is divided into bins (intervals), and the height of each bar represents the frequency (or count) of data points falling within that bin. They help understand the shape, center, and spread of the data.

- **Use Cases:** Showing the distribution of test scores in a class, illustrating the frequency of different income levels in a population, analyzing the distribution of product weights in a batch.

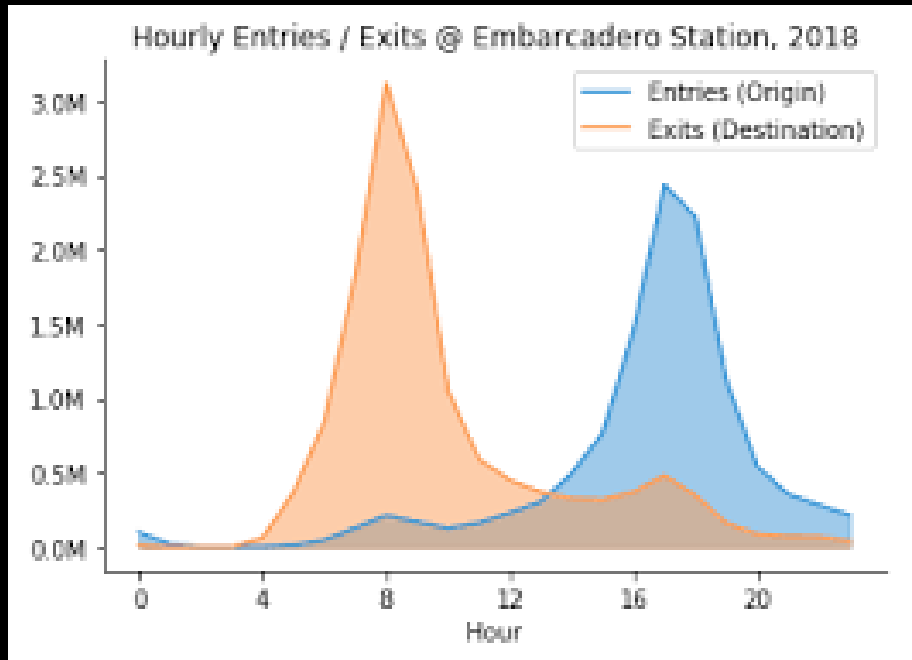
6. BOX PLOT (OR BOX AND WHISKER PLOT):



- **Significance:** Box plots provide a concise summary of the distribution of a quantitative variable, highlighting key statistics such as the median, quartiles, and potential outliers. The "box" represents the interquartile range (IQR), the line inside the box represents the median, and the "whiskers" extend to [1] show the range of the data (excluding outliers).
- **Use Cases:** Comparing the distribution of salaries across different departments, visualizing the spread of customer ages for different product segments, identifying variability in manufacturing process measurements.

7. AREA CHART:

- **Significance:** Area charts are similar to line charts but fill the area below the line with color. They are useful for emphasizing the magnitude of change over time and can also be used to compare the cumulative values of different categories
- **Use Cases:** Showing the cumulative sales of different product lines over time, illustrating the total energy consumption from various sources over a period.



8.HEATMAP:

- **Significance:** Visually represents data magnitude across two dimensions using color intensity, quickly revealing patterns and correlations.
- **Use Cases:** Correlation matrices, website activity, geographic data, gene expression, financial trends, customer segmentation, missing data.