

ASSIGNMENT

20MCA201-DATA SCIENCE AND MACHINE LEARNING

Submitted to:

Ms. Jetty Benjamin
Assistant Professor
AJCE, KERALA

Submitted By:

Arjun A Nair
RMCA – A
Roll No: 24

DATA VISUALIZATION TECHNIQUES

IN DATA SCIENCE

INTRODUCTION

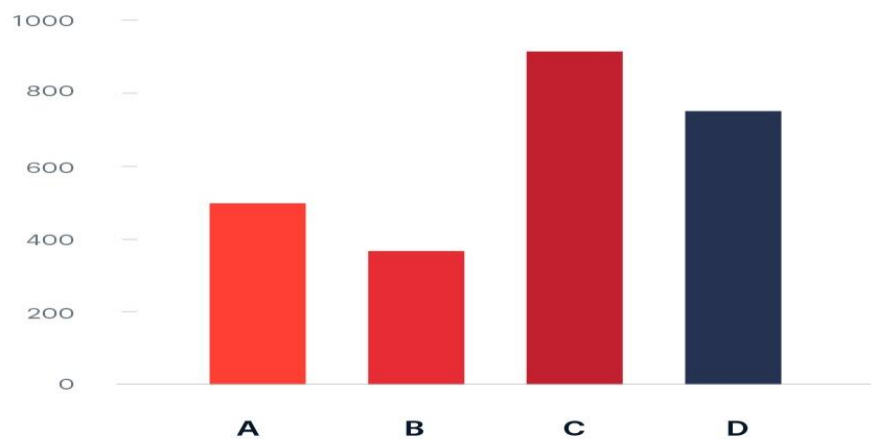
Data visualization is a fundamental aspect of the field of data science, offering a powerful means to transform raw data into meaningful insights. It is the art and science of representing data graphically, allowing for a visual interpretation of complex information. The primary objective of data visualization is to simplify the comprehension of data, making it easier to identify patterns, trends, and outliers, while also facilitating the communication of findings to a wider audience. This assignment delves into the realm of data visualization techniques, equipping you with the skills to harness the potential of visual representations for data analysis.

In today's data-driven world, datasets are increasingly vast and intricate, making data visualization an indispensable tool for data scientists, analysts, and decision-makers. By employing a diverse array of charts, graphs, and plots, data visualization transforms abstract numbers and statistics into visually digestible formats. This assignment will guide you through the process of selecting, preparing, and visualizing a real-world dataset, enabling you to explore the nuances of data representation and interpretation.

As you navigate through this assignment, you will discover the importance of data preprocessing, univariate and bivariate analysis, and the application of multivariate techniques. Additionally, you will recognize the real-world significance of data visualization in various domains, underscoring its role in informing critical decisions and solving complex problems. In an era where data is the lifeblood of modern enterprises and research, mastering data visualization is an essential skill for anyone seeking to unlock the potential hidden within the numbers.

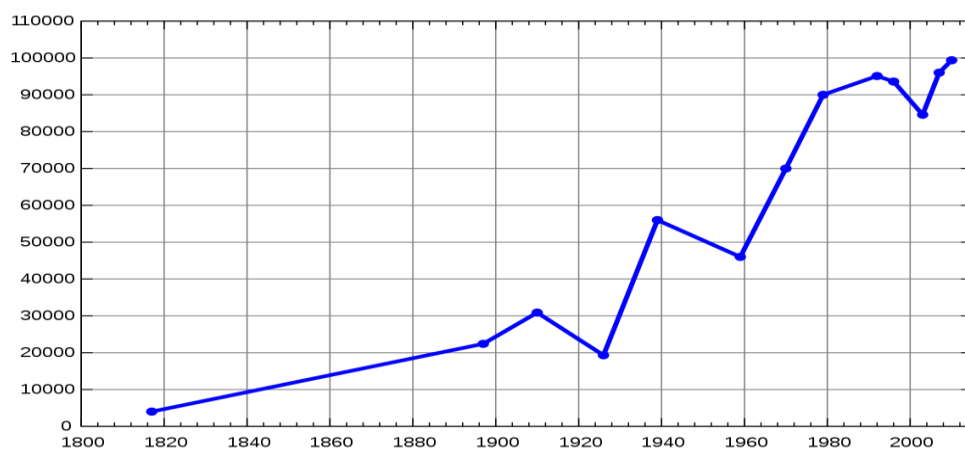
BAR CHART

Bar Charts are a widely used data visualization technique that represents data using rectangular bars of varying lengths. Each bar corresponds to a category or group, and the height or length of the bar is proportional to the value it represents. Bar charts are excellent for comparing data across categories, making it easy to identify trends, differences, and relationships. They are particularly useful for displaying discrete data and can be oriented either horizontally or vertically



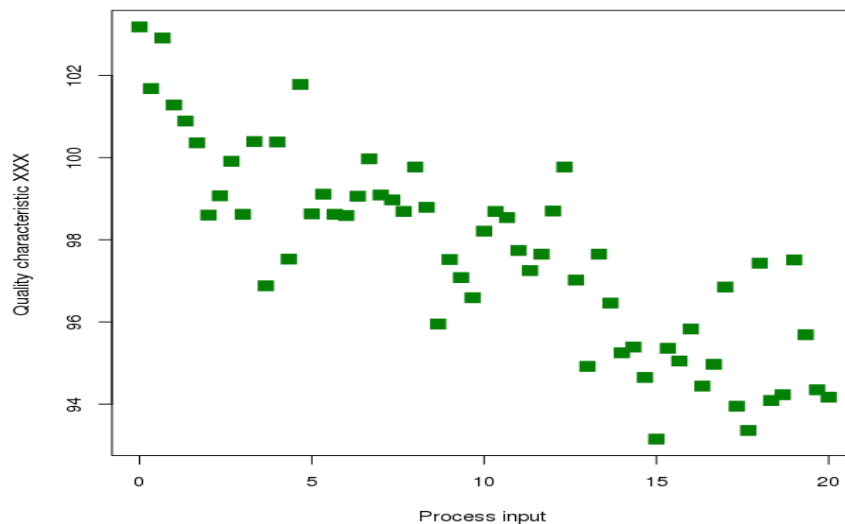
LINE CHART

Line Charts, also known as line graphs, are a fundamental data visualization technique. They display data points connected by straight lines, typically used to visualize trends and changes over time. Each data point represents a value at a specific time or along a continuous scale. Line charts are effective for showing the relationship between two or more variables and revealing patterns, fluctuations, or correlations in the data. They are commonly used in fields such as economics, finance, and scientific research to track changes and make predictions.



SCATTER PLOT

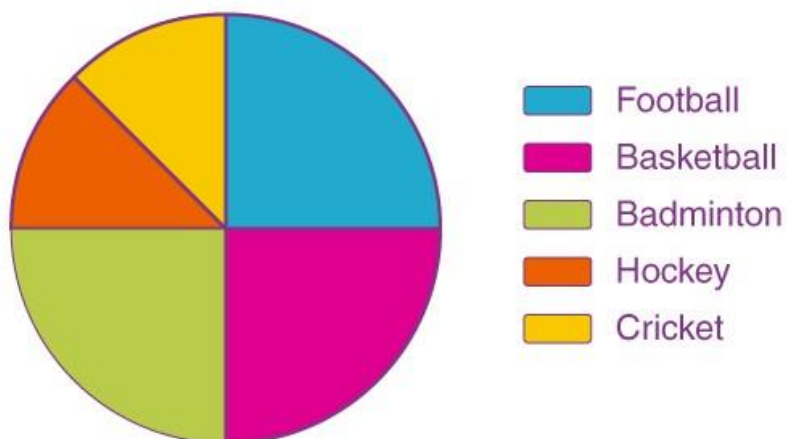
Scatter Plots are a powerful data visualization technique that displays data points as individual dots or markers on a two-dimensional plane. Each dot represents a single data point, and the position of the dot is determined by the values of two variables. Scatter plots are used to explore and visualize the relationships or correlations between these two variables.



PIE CHART

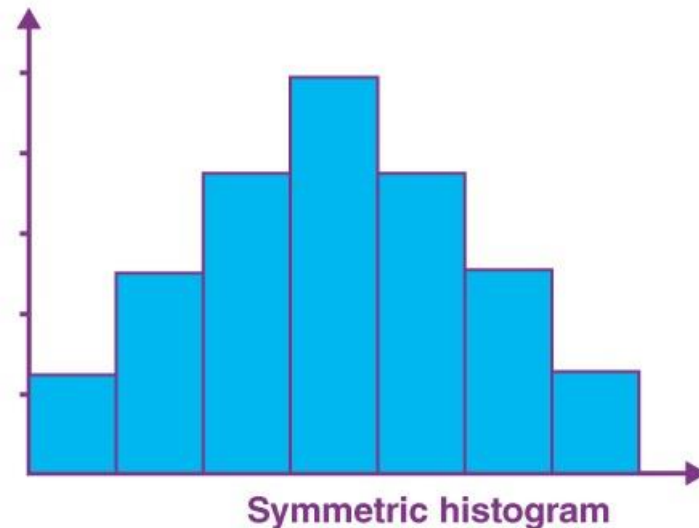
Pie Charts are a common data visualization technique that presents data in a circular form, resembling a pie. The entire "pie" represents 100% of the data, and it is divided into segments or "slices" that correspond to different categories or parts. The size of each segment is proportional to the value it represents relative to the total. Pie charts are ideal for displaying data where you want to show the relationship between parts and the whole, emphasizing the composition of a whole entity.

Favourite Sports Percentage



HISTOGRAM

Histograms are a vital data visualization technique that provides a graphical representation of the distribution of a single variable. They work by dividing the data into intervals or "bins" and then plotting the frequency or count of data points falling into each bin. The bars in a histogram are typically adjacent and have no gaps between them, as they represent continuous or discrete data. Histograms are particularly useful for understanding the shape of a dataset, identifying patterns, and assessing the distribution, including whether it is normal, skewed, or multimodal.



BUBBLE CHART

Bubble Charts are a data visualization technique that builds upon the concept of scatter plots but introduces a third dimension to represent data. In a bubble chart, data points are represented as circles (bubbles) instead of simple points, and the size of each bubble corresponds to a third variable, typically a numeric value.

Bubble Chart

