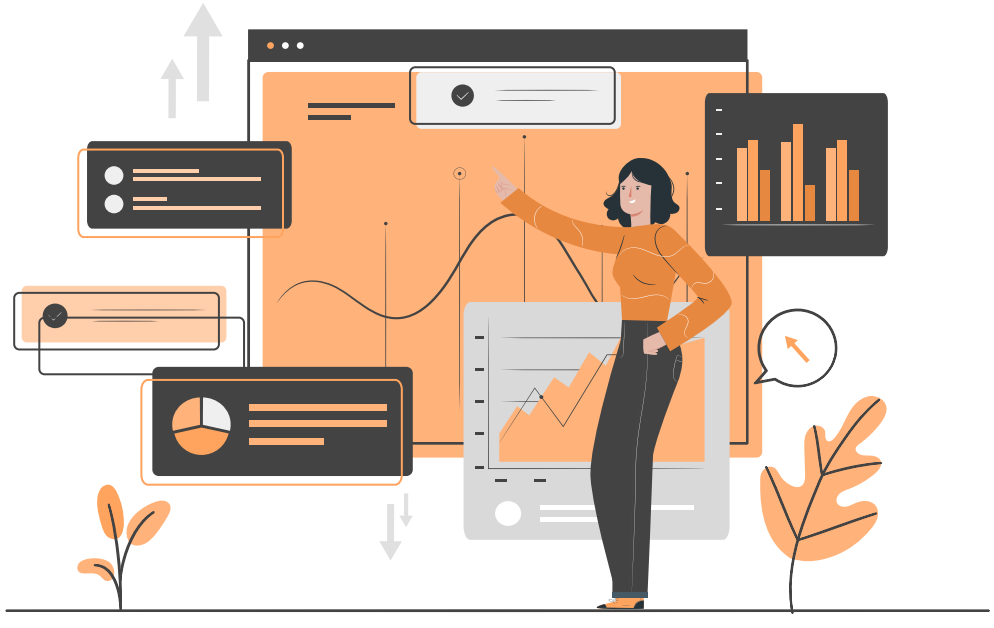


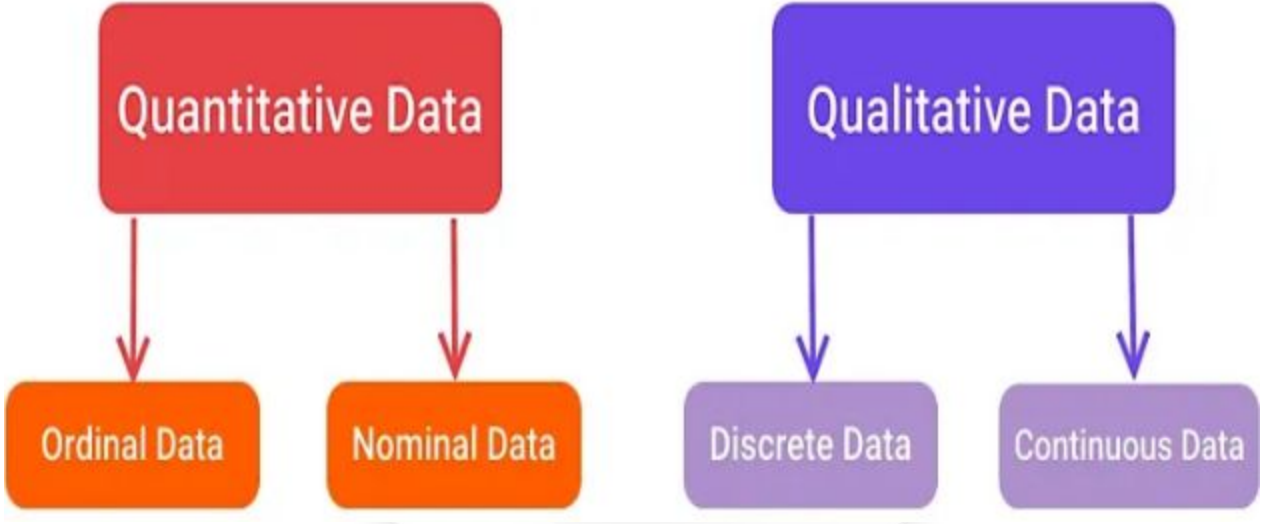
Data Visualization – Part 1

By Fireblaze AI School

Data

- “Data is the new oil.” Today data is everywhere in every field.
- Whether you are a data scientist, marketer, businessman, data analyst, researcher, or you are in any other profession, you need to play or experiment with raw or structured data.
- Data has become an important asset for an organization and data science job roles are on the uptrend.

Types of Data



Types of Data

- **Quantitative or Categorical Data** is data that can't be measured or counted in the form of numbers.
- These types of data are sorted by category, not by number. That's why it is also known as Categorical Data.
- These data consist of audio, images, symbols, or text. The gender of a person, i.e., male, female, or others, is qualitative data.

Types of Data

Examples:

- What language do you speak,
- Favourite holiday destination,
- Opinion on something (agree, disagree, or neutral), Colours

Types of Data

The Qualitative data are further classified into two parts :

Nominal Data

Nominal Data is used to label variables without any order or quantitative value. The colour of hair can be considered nominal data, as one colour can’t be compared with another colour.

Types of Data

Examples:

- Colour of hair (Blonde, red, Brown, Black, etc.),
- Marital status (Single, Widowed, Married),
- Nationality (Indian, German, American)
- Gender (Male, Female, Others)
- Eye Color (Black, Brown, etc.)

Types of Data

Ordinal Data

- Ordinal data have natural ordering where a number is present in some kind of order by their position on the scale.
- These data are used for observation like customer satisfaction, happiness, etc., but we can’t do any arithmetic tasks on them.

Types of Data

Examples:

- Letter grades in the exam (A, B, C, D, etc.)
- Ranking of peoples in a competition (First, Second, Third, etc.)
- Economic Status (High, Medium, and Low)
- Education Level (Higher, Secondary, Primary)

Types of Data

Quantitative Data

- Quantitative data can be expressed in numerical values, which makes it countable and includes statistical data analysis.
- These kinds of data are also known as Numerical data.
- It answers the questions like, “how much,” “how many,” and “how often.” For example, the price of a phone, the computer’s ram, the height or weight of a person, etc., falls under the quantitative data.

Types of Data

The Quantitative data are further classified into two parts :

Discrete Data

The term discrete means distinct or separate. The discrete data contain the values that fall under integers or whole numbers. The total number of students in a class is an example of discrete data.

These data can’t be broken into decimal or fraction values.

The discrete data are countable and have finite values; their subdivision is not possible.

Types of Data

Examples of Discrete Data :

- Total numbers of students present in a class
- Cost of a cell phone
- Numbers of employees in a company
- The total number of players who participated in a competition
- Days in a week

Types of Data

Continuous Data

- Continuous data are in the form of fractional numbers. It can be the version of an android phone, the height of a person, the length of an object, etc. Continuous data represents information that can be divided into smaller levels. The continuous variable can take any value within a range.

Types of Data

- The key difference between discrete and continuous data is that discrete data contains the integer or whole number. Still, continuous data stores the fractional numbers to record different types of data such as temperature, height, width, time, speed, etc.

Types of Data

Examples of Continuous Data :

- Height of a person
- Speed of a vehicle
- “Time-taken” to finish the work
- Wi-Fi Frequency
- Market share price

Why data important?

- The principal purpose of Data Science is to find patterns within data.
- It uses various statistical techniques to analyze and draw insights from the data.
- From data extraction, wrangling and pre-processing, a Data Scientist must scrutinize the data thoroughly.

Data Visualization

- Data visualization is the practice of translating information into a visual context, such as a map or graph, to make data easier for the human brain to understand and pull insights from.
- The main goal of data visualization is to make it easier to identify patterns, trends and outliers in large data sets. The term is often used interchangeably with others, including information graphics, information visualization and statistical graphics.

Data Visualization

- Data visualization is one of the steps of the data science process, which states that after data has been collected, processed and modeled, it must be visualized for conclusions to be made.
- Data visualization is important for almost every career. It can be used by teachers to display student test results, by computer scientists exploring advancements in artificial intelligence (AI) or by executives looking to share information with stakeholders

Benefits of data Visualization

- the ability to absorb information quickly, improve insights and make faster decisions;
- an increased understanding of the next steps that must be taken to improve the organization;
- an improved ability to maintain the audience's interest with information they can understand;
- an easy distribution of information that increases the opportunity to share insights with everyone involved;
- eliminate the need for data scientists since data is more accessible and understandable; and
- an increased ability to act on findings quickly and, therefore, achieve success with greater speed and less mistakes.

Exploratory Data Analysis

- The preliminary analysis of data to discover relationships between measures in the data and to gain an insight on the trends, patterns, and relationships among various entities present in the data set with the help of statistics and visualization tools is called **Exploratory Data Analysis (EDA**).
- Exploratory data analysis is cross-classified in two different ways where each method is either graphical or non-graphical. And then, each method is either **univariate, bivariate or multivariate.**

Univariate Analysis

- Uni means one and variate means variable, so in univariate analysis, there is only one dependable variable.
- The objective of univariate analysis is to derive the data, define and summarize it, and analyze the pattern present in it.
- In a dataset, it explores each variable separately. It is possible for two kinds of variables- Categorical and Numerical.

Univariate Analysis

- Some patterns that can be easily identified with univariate analysis are Central Tendency (mean, mode and median), Dispersion (range, variance), Quartiles (interquartile range), and Standard deviation.

Univariate Analysis

Univariate data can be described through:

- Frequency Distribution Tables
- Bar Charts
- Histograms
- Pie Charts

Bivariate Analysis

Bi means two and variate means variable, so here there are two variables. The analysis is related to cause and the relationship between the two variables. **Bivariate Analysis of two Numerical Variables (Numerical-Numerical)**

- **Scatter plots**
- **Correlation Heatmaps**

Mulitivariate Analysis

- Multivariate analysis is required when more than two variables have to be analyzed simultaneously.
- It is a tremendously hard task for the human brain to visualize a relationship among 4 variables in a graph and thus multivariate analysis is used to study more complex sets of data.
- Types of Multivariate Analysis include Cluster Analysis, Factor Analysis, Multiple Regression Analysis, Principal Component Analysis, etc.

Mulitivariate Analysis

- More than 20 different ways to perform multivariate analysis exist and which one to choose depends upon the type of data and the end goal to achieve.


The End