Data and their Types

Data is defined as a systematic record corresponding to a specific quantity. Basically, data can be summarized as a set of facts and figures which can be used to serve a specific usage or purpose. For instance, data can be used as a survey or an analysis. Data in a systematic and organized form is referred to as information. In addition to this, the source of data primary or secondary is also an essential factor.

Different types of data based on their

1.Source: The origin or location of the data, such as where it was collected or obtained from. Data Based on Source can be subdivided into two categories

- **Primary data**: Data that is collected firsthand by the researcher through methods such as surveys, interviews, or experiments.
- **Secondary data**: Data that is obtained from existing sources such as government reports, academic journals, or databases.

2.Organization: The structure or arrangement of the data, which can be structured (organized in a specific way, such as in a database) or unstructured (not organized in any particular way, such as free text). Data based on Organization can be subdivided into two categories:

- **Structured data (Classified Data)**: Data that is organized in a specific way, such as in a table, spreadsheet, or database.
- **Unstructured data (Raw Data)**: Data that is not organized in any particular way, such as free text, images, or audio.

3. Values: The different categories or types of data values that can be measured, such as numerical, categorical, or textual.

- **Qualitative Data**: Qualitative data is used to represent some characteristics or attributes of the data. The facts and figures depicted by the qualitative data cannot be computed. These properties reflect observable attributes. These are non-numerical in nature. The qualitative data characteristics are exploratory on a larger end than being conclusive in nature. For instance, data on attributes such as honesty, loyalty, wisdom, and creativity for a set of persons defined can be considered as qualitative data. Examples: Attitudes of people to a political system, Music and art, Intelligence, Beauty of a person. It can be further divided into 2 categories: Nominal and Ordinal.
- **Nominal Data**: Nominal data is a sub-category belonging to one of the types of qualitative information. Also known as the nominal scale, it is used to label the variables without providing the numerical value for them. Nominal data attributes can't either be ordered or measured. The nominal data can be both qualitative and quantitative in nature. For instance, some of the nominal data attributes are letters, symbols or gender, etc. The examination of the nominal data is based on the usage of the grouping method. This method is based on the principle of the grouping of data into different categories. This is followed by the calculation of the frequency or the percentage of the data. The visualization of this data is done using the pie charts. Examples: Gender (Women, Men), Eye color (Blue, Green, Brown),Hair color (Blonde, Brown, Brunette, Red, etc.), Marital status (Married, Single)
- **Ordinal Data:** Ordinal data/variable is the specific type of data that follows a natural order. The difference between the data values is not determined in the case of nominal data. For instance, ordinal data variable is mostly found in surveys, economics, questionnaires, and finance operations. The examination of the nominal data is based on the usage of visualization tools. The visualization of this data is done using the bar chart. The ordinal data can be expressed in the form of tables which have each row corresponding to the distinct category. Examples: Feedback is recorded in the form of ratings from 1-10. Education level: elementary school, high school, college. Economic status: low, medium, and high. Letter grades: A, B, C, etc. Customer level of satisfaction: very satisfied, satisfied, neutral, dissatisfied, very dissatisfied.

- **Quantitative data** can be measured and is not just observable. The measurement of data is numerically recorded and represented. Calculations and interpretations can then be performed on the obtained results. Numerical data is indicated by quantitative data. For instance, data can be recorded about how many users

found a product satisfactory in terms of the collected rating, and therefore, an overall product review can be generated. Examples: Daily temperature, Price, Weights, Income etc. It can be further divided into 2 categories: Discrete and continuous

- **Discrete data** refers to the data values which can only attain certain specific values. Discrete data can't attain a range of values. Discrete data can be represented using bar charts. For instance, ratings of a product made by the users can only be in discrete numbers. Examples: The number of students in a class, The number of chips in a bag, The number of stars in the sky.

- **Continuous Data** can contain values between a certain range that is within the highest and lowest values. The corresponding difference between the highest and lowest value of these intervals can be termed as the range of data. Continuous data can be tabulated in what is called a frequency distribution. The frequency distribution table can be computed for the range type of data. It can also be depicted using histograms. For example, the heights of the students in the class can be largely varying in nature, therefore, they can be divided into ranges to summarize the data. Examples: Height and weight of a student, Daily temperature recordings of a place, Wind speed measurement

4. Variables: The characteristics or attributes of the data that can be measured or observed, such as age, height, weight, gender, etc.

- **Univariate Data:** Univariate refers to an analysis or data that involves a single variable. In statistics, univariate analysis is used to describe the characteristics of a single variable, such as its distribution, central tendency, variability, and outliers. Univariate analysis is useful for understanding the behavior of a single variable and can help identify patterns or trends in the data.
- **Bivariate Data**: Bivariate refers to an analysis or data that involves two variables. Bivariate analysis is used to explore the relationship between two variables, such as their correlation or causation. It helps to understand how one variable affects the other and can be used to make predictions or to identify potential causes of a phenomenon. Bivariate analysis can be done using different statistical techniques, such as correlation analysis, regression analysis, and t-tests.