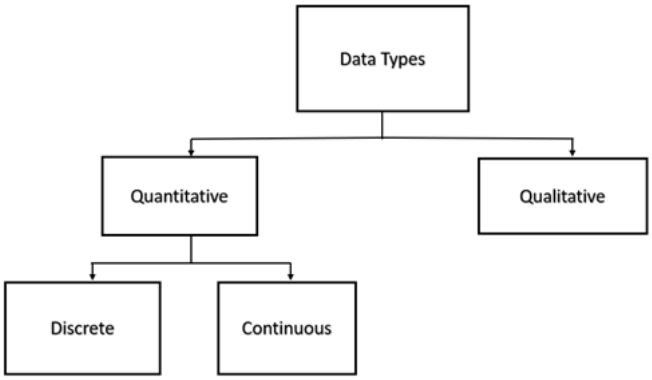
TYPES OF DATA

Data has to be converted into numeric representation so that the machines are able to learn the patterns within data. Understanding the different data types can help us identify correct preprocessing techniques & convert the data appropriately. Furthermore, it will also enable us to perform the best visualizations and uncover hidden knowledge.

Structured data

This type of data is usually composed of numbers or words. They are usually stored in Relational databases and can be easily searched using SQL queries.



Numeric/Quantitative data

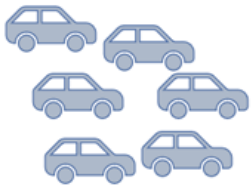
As the name suggests, this encompasses data that can be represented through numbers. Examples of such data are sales price, metric quantities such as temperature, time, length, height & weight of a person, and so on. Numeric data is further divided into two categories, namely discrete and continuous.

Discrete

In this category, the data takes on discrete values or whole numbers i.e numbers without decimal points. Examples are the number of houses in a city, the number of consumers in a grocery store over the last month, the number of Instagram followers that you have, and so on.

E.G.: – No. Of Cars You Have, No. Of Marbles In Containers, Students In A Class, Etc.

No. of Laptops

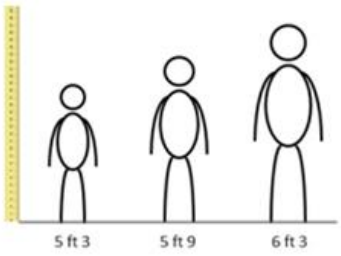


No. of Cars

Continuous

In this category, the data takes on integer values i.e numbers with decimal values. Examples for continuous numeric data are house prices in the city, sale prices of grocery store items, Instagram earnings that you received, and so on.

E.G.: – Height, Weight, Time, Area, Distance, Measurement Of Rainfall, Etc.



Height

Time

Categorical/Qualitative data

As the name suggests, this encompasses data that can be represented through words. It usually defines groups or categories & is therefore known as categorical data. Some examples are the names of all items in a supermarket, movie ratings (good, average, bad), country of birth of individuals & so on.

Ordinal

This type of data has an inherent ordering present within the categories. For instance, if you consider movie ratings with good, average & bad as the different categories, good has a higher

ranking than average which is higher than bad. This needs to be taken into account while converting this type of data into numbers so that the models can learn this ranking as well. There is a fixed, finite number of categories/groups. Examples will be movie ratings, student grades, Employee performance, and so on.

E.G., Likert Rating Scale, Shirt Sizes, Ranks, Grades, Etc.

Fig: Rating (Good, Average, Poor), An Example Of Ordinal Data Type

Nominal

This type of data has categories that don't have any particular order or ranking associated with them. The total number of categories is usually finite in this type of data as well. Examples will be the country of birth of individuals, all items in a supermarket, educational degrees of individuals, and so on.

E.G., Male Or Female (Gender), Race, Country, Etc.

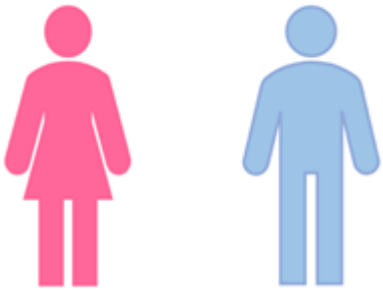


Fig: Gender (Female, Male), An Example Of Nominal Data Type

Interval Data Type:

This Is Numeric Data Which Has Proper Order And The Exact Zero Means The True Absence Of A Value Attached. Here Zero Means Not A Complete Absence But Has Some Value. This Is The Local Scale.

E.G., Temperature Measured In Degree Celsius, Time, Sat Score, Credit Score, PH, Etc. Difference Between Values Is Familiar. In This Case, There Is No Absolute Zero. Absolute

Fig: Temperature, An Example Of Interval Data Type

IV. Ratio Data Type:

This Quantitative Data Type Is The Same As The Interval Data Type But Has The Absolute Zero. Here Zero Means Complete Absence And The Scale Starts From Zero. This Is The Global Scale. E.G., Temperature In Kelvin, Height, Weight, Etc.

Fig: Weight, An Example Of Ratio Data Type

Unstructured data

This type of data is usually composed of everything else including texts, images, videos, speech/audio, time series, and so on. They are usually stored in non-relational databases and cannot be searched easily.

Image

As the name suggests, this type of data usually consists of image files of different types. A key attribute of this data type is the presence of spatial features/relationships within images that need to be understood to extract insightful information from the images. Examples are images of all items in the grocery store, photos of all students in a college, and so on.

Video

This type of data is also pretty self-explanatory as it consists of videos in different formats. A distinguishing factor with this data type is the relationships between different frames in the video with respect to positions, movements of objects/people etc. need to be taken into account to better obtain information from the videos.

Audio/Time series

This type of data has a sequence of ordered data points each having a timestamp. The most salient feature in this data is the relationship between the different data points such as periodic patterns, seasonal behaviors, and so on. For example, if you consider the temperature recorded in a city over last year, looking at the changes over time, we can easily identify that winter months are colder and summer months are hotter. This type of insight is basic but can only be observed if you look at the data points with their timestamps. Figure 2 shows an example visualization of time series data.

Text

This type of data has textual data composed of multiple words occurring such that they make sense as a whole. The most important attribute is understanding the overall context and relationships between different words within a sentence as well as understanding that each word can have multiple meanings as well as associations with other words.