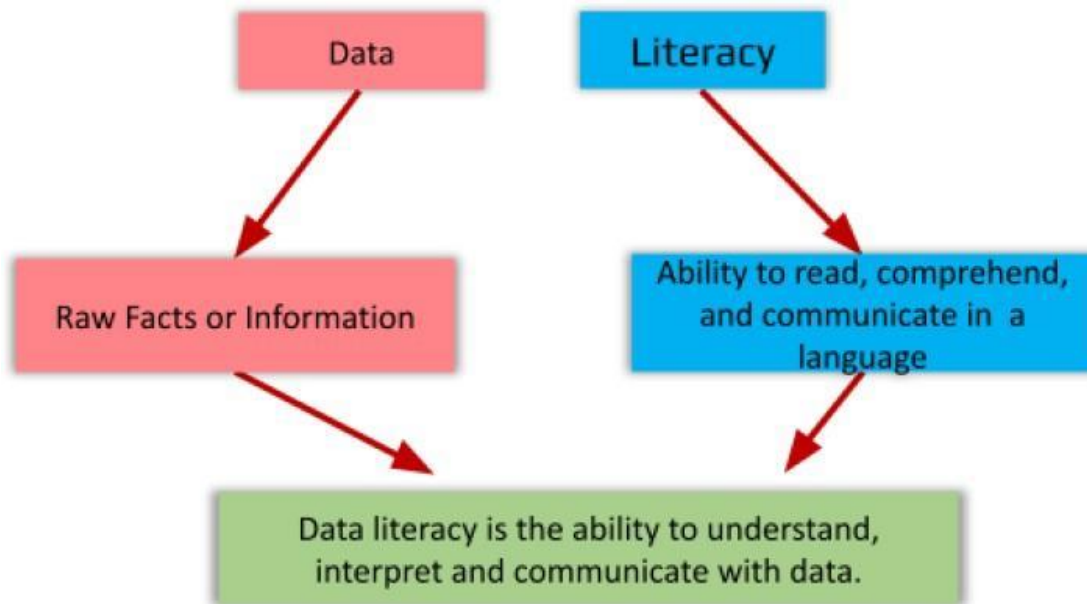

Unit 2: Data Literacy and AI

2.1 - Basics of Data Literacy

2.1.1 Introduction to Data Literacy

Data literacy is about understanding, working with, and communicating data. It involves knowing how to collect, analyze, and present data in meaningful ways. It enables informed decision-making and critical thinking.



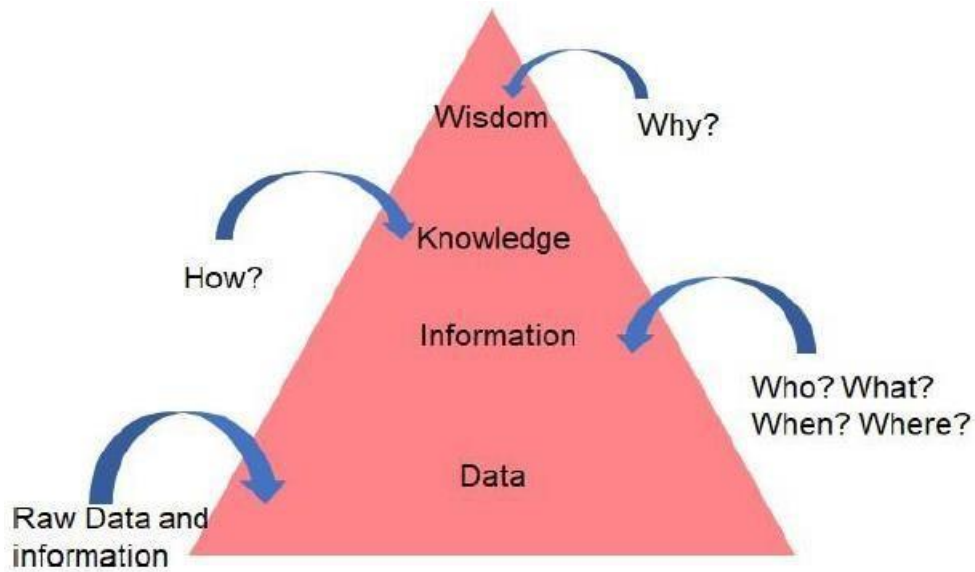
Why is Data Literacy Important?

Data literacy is important in everyday life as it helps people make decisions based on facts, think critically, solve problems, and innovate. It also helps in recognizing data privacy and security concerns.

Key Concepts

1. Quantitative Data: Data that consists of numbers, like scores, measurements, or readings.
2. Qualitative Data: Data made up of words and phrases, like opinions or descriptions.
3. Data Privacy: Protecting sensitive information such as personal data.
4. Data Security: Safeguarding digital data from unauthorized access, corruption, or theft.

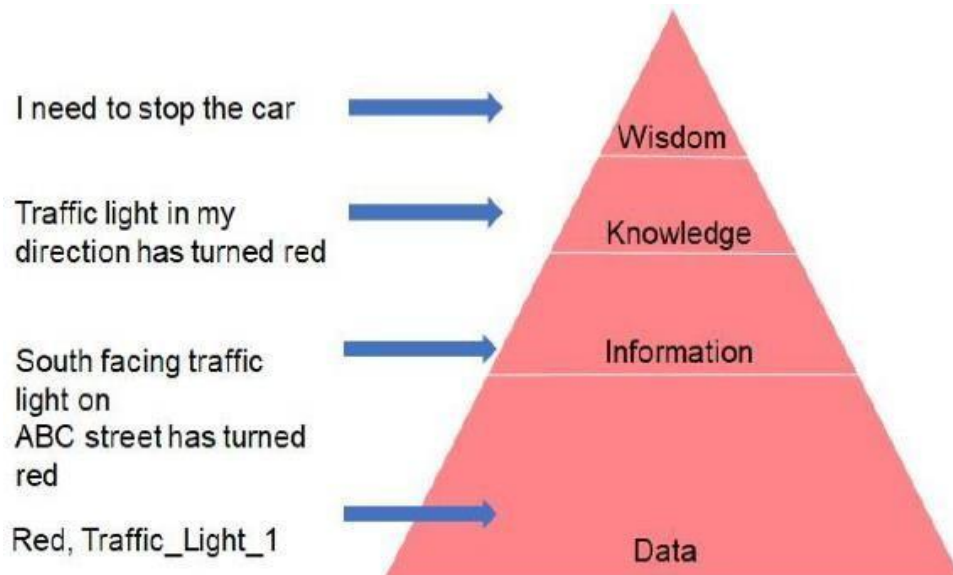
Data Pyramid



The Data Pyramid has four stages:

1. **Data:** Raw information, not very useful on its own.
2. **Information:** Processed data that gives insights.
3. **Knowledge:** Information that leads to understanding.
4. **Wisdom:** Understanding why things happen the way they do.

Let's understand Data Pyramid with a simple Traffic Light example:



2.1.2 Impact of Data Literacy

Activity: Impact of News Articles (Select any trending news) [To be pasted in AI practical file]

Session Preparation Logistics: For a class of 40 Students [Pair Activity]

Materials Required:

ITEM QUANTITY

Online Data Sources Clues NA

Computers 20

Purpose: The purpose of this activity is to engage participants in various scenarios that involve collecting data and analyzing its sources. By understanding how authentic data sources contribute to reliable and unbiased decision-making.

Brief: {Pair Activity} Participants will search the internet for data sources, extracting key information to support their decisions.

Use the given template:

Author of the Source	Weblink to the Source	How was the situation described by the Source	Key figures/points in the source

You have to rank the sources of the news articles from most accurate to least, state reasons for your choice.

Rank	Data Source	Remarks

So, we can conclude that every data tells a story, but we must be careful before believing the story. Data literacy is essential because it enables individuals to make informed decisions, think critically, solve problems, and innovate.

2.1.3 How to Become Data Literate

Being data literate means understanding how to research, filter, and analyze data for everyday tasks like online shopping or reading reviews. It involves using data to make informed decisions and recognizing the differences between reliable and unreliable sources of information.

Scenario: Buying a Video game online

Data literacy helps people research about products while shopping over the Internet

How do you decide the following things when we are shopping online?

- Which is the cheapest product available?
- Which product is liked by the users the most?

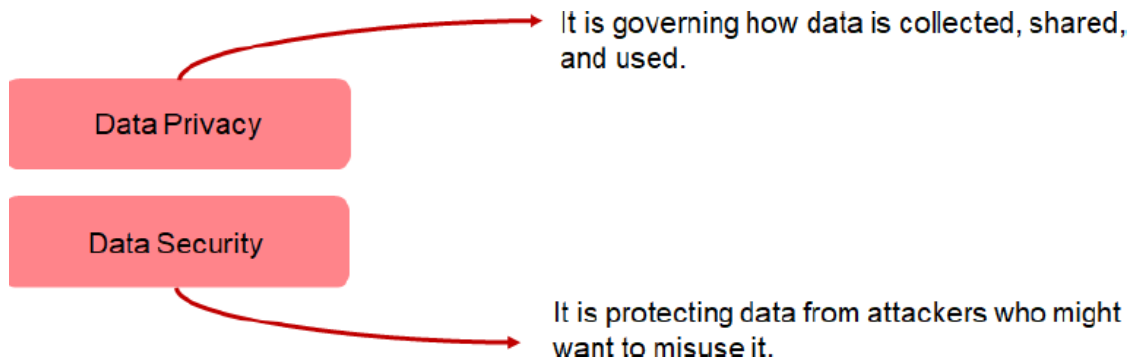
- Does a particular product meet all the requirements?

A data literate person can –

- Filter the category as per the requirement – If the budget is low, select the price filter as low to high
- Check the user ratings of the products
- Check for specific requirements in the product

2.1.4 What are Data Security and Privacy? How are they related to AI?

Data Privacy and Data Security are often used interchangeably but they are different from each other.



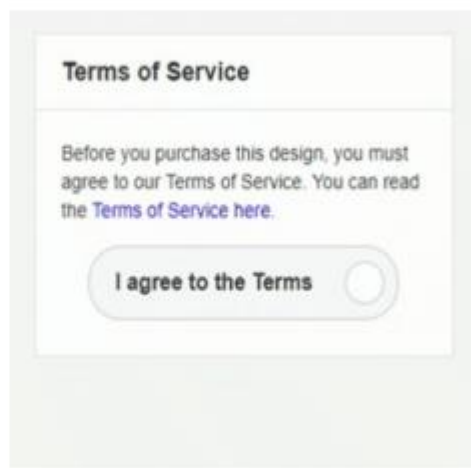
What is Data Privacy?

Data privacy referred to as information privacy is concerned with the proper handling of sensitive data including personal data and other confidential data, such as certain financial data and intellectual property data, to meet regulatory requirements as well as protecting the confidentiality and immutability of the data.

Here are examples of two things which may compromise our data privacy



Downloaded an unverified mobile application



Accepted the Terms of Service without reading

Why is it important?



A data breach at a government agency can put top secret information in the hands of an enemy state.



A breach at a corporation can put proprietary data in the hands of a competitor.



A breach at a hospital can put personal health information in the hands of those who might misuse it.

The following best practices can help you ensure data privacy:

- Understanding what data, you have collected, how it is handled, and where it is stored.
- Necessary data required for a project should only be collected.
- User consent while data collection must be of utmost importance.

What is Data Security?

Data security is the practice of protecting digital information from unauthorized access, corruption, or theft throughout its entire lifecycle.

Why is it important?

Due to the rising amount of data in the cloud there is an increased risk of cyber threats. The most appropriate step for such an amount of traffic being generated is how we control and protect the transfer of sensitive or personal information at every known place.

The most possible reasons why data security is more important now are:

- Cyber-attacks affect all the people
- The fast-technological changes will boom cyber attacks

2.1.5 Best Practices for Data Security

Data Security protects computers, servers, mobile devices, electronic systems, networks and data from harmful attacks.

A few practices are as follows:

1. Use strong passwords with a mix of characters.
2. Enable Two-Factor Authentication (2FA) for extra security.
3. Download files only from trusted sources.
4. Always lock your screen when not in use.
5. Be cautious while sharing personal information online.

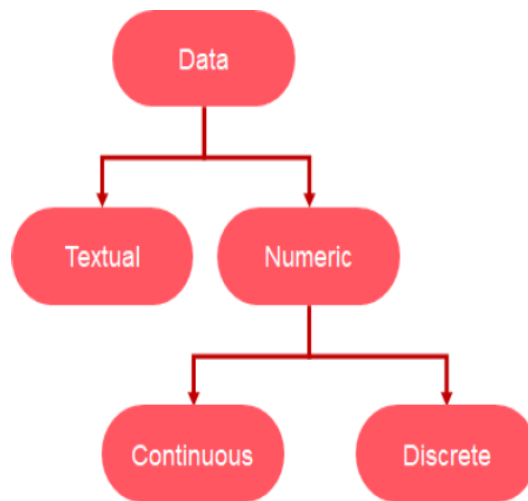
Activity - Exploring Data Online [To be pasted in AI practical file]

1. Choose a product to research online.
2. Compare products based on price, ratings, and reviews.
3. Write down how the data helps you make a decision.

2.2 Acquiring Data, Processing, and Interpreting

2.2.1 Data Types of data

Artificial Intelligence is crucial, with data serving as its foundation. We come across different types of information every day. Some common types of data include:

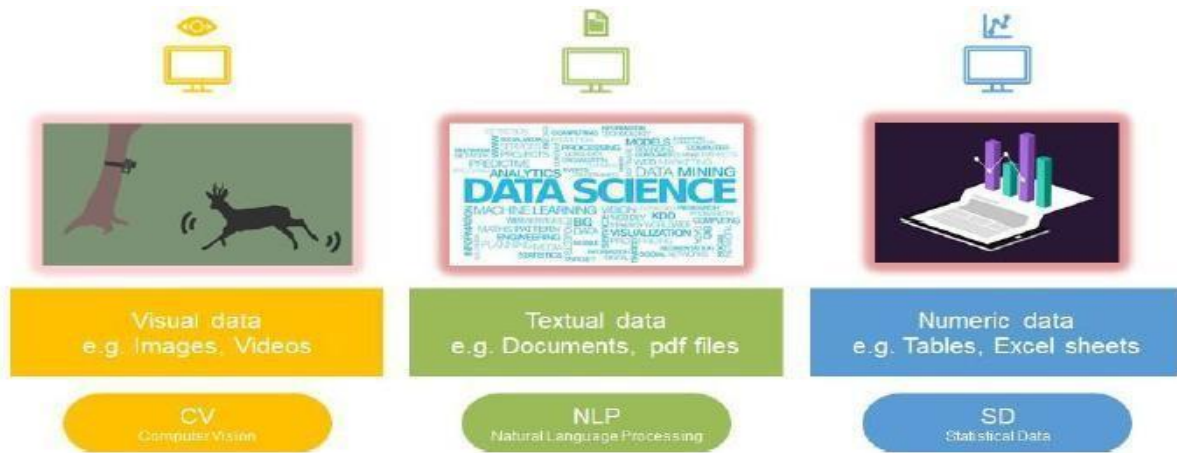


Textual Data (Qualitative Data)	Numeric Data (Quantitative Data)
<ul style="list-style-type: none"> • It is made up of words and phrases • It is used for Natural Language Processing (NLP) • Search queries on the internet are an example of textual data • Example: “Which is a good park nearby?” 	<ul style="list-style-type: none"> • It is made up of numbers • It is used for Statistical Data • Any measurements, readings, or values would count as numeric data • Example: Cricket Score, Restaurant Bill

Numeric Data is further classified as:

- **Continuous data** is numeric data that is continuous. E.g., height, weight, temperature, voltage
- **Discrete data** is numeric data that contains only whole numbers and cannot be fractional. E.g. the number of students in the class – it can only be a whole number, not in decimals

Types of Data used in three domains of AI:



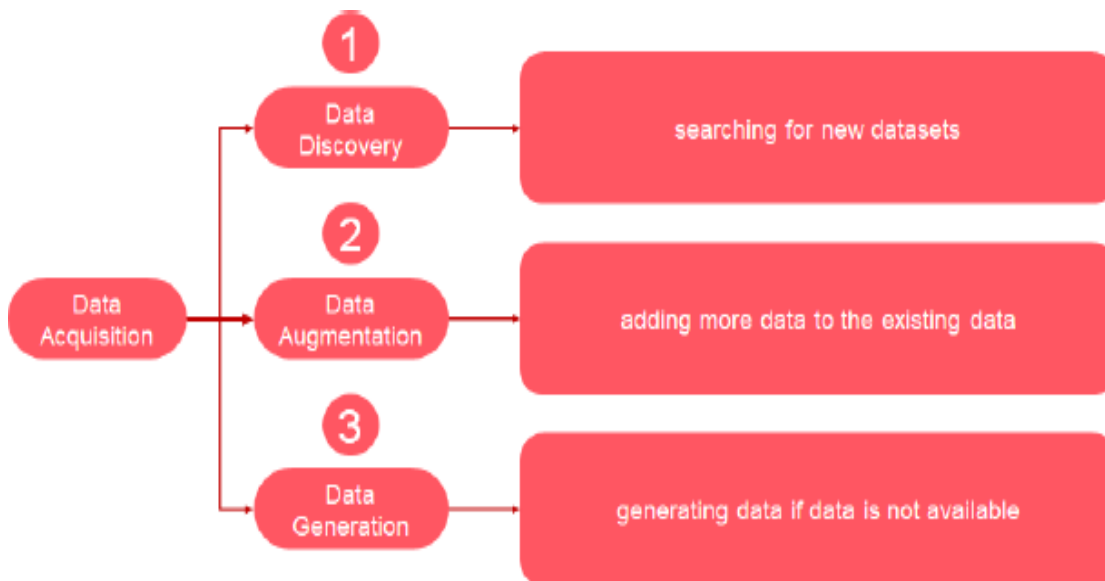
Pick and Choose (Quantitative or Qualitative?)



2.2.2 Data Acquisition/Acquiring Data

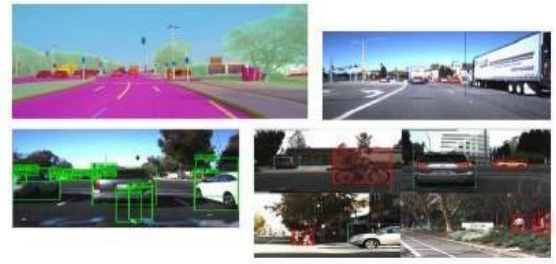
Data Acquisition also known as Acquiring data, refers to the procedure of gathering data. It involves searching of datasets suitable for training AI models.

The process typically comprises 3 key steps:



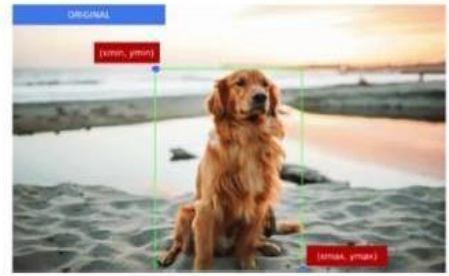
Acquiring Data – Sample Data Discovery

- Let's say we want to collect data for making a CV model for a self-driving car
- We will require pictures of roads and the objects on roads
- We can search and download this data from the internet
- This process is called **data discovery**



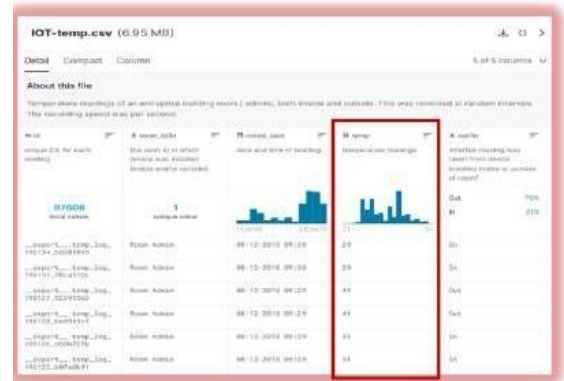
Acquiring Data – Sample Data Augmentation

- Data augmentation means increasing the amount of data by adding copies of existing data with small changes
- The image given here does not change, but we get data on the image by changing different parameters like color and brightness
- New data is added by slightly changing the existing data



Acquiring Data – Sample Data Generation

- Data generation refers to generating or recording data using sensors
- Recording temperature readings of a building is an example of data generation
- Recorded data is stored in a computer in a suitable form



Sources of Data

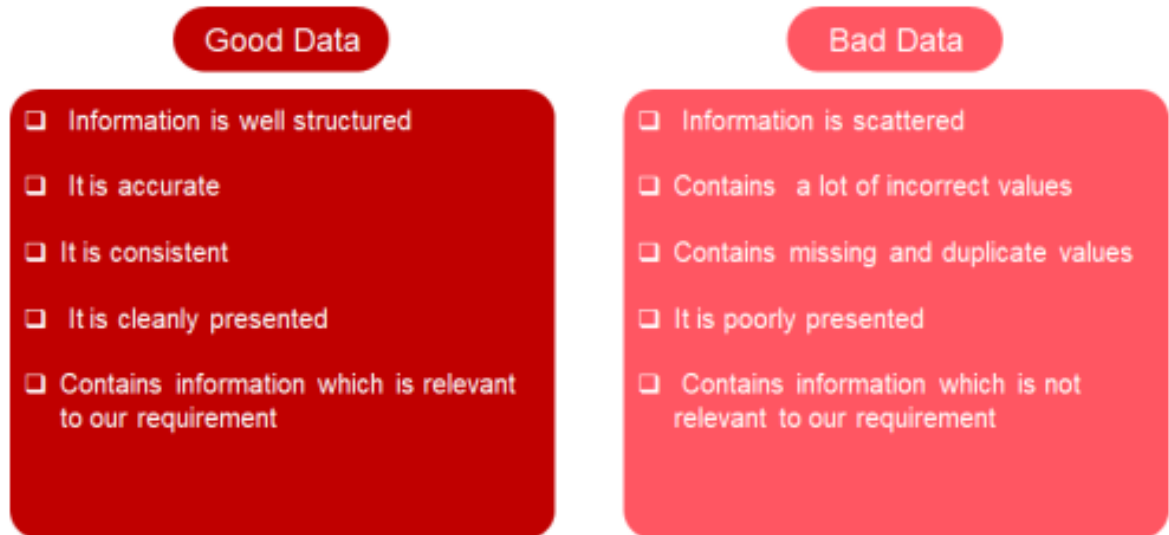
Various Sources for Acquiring Data:

- **Primary Data Sources** — Some of the sources for primary data include surveys, interviews, experiments, etc. The data generated from the experiment is an example of primary data.
- **Secondary Data Sources**—Secondary data collection obtains information from external sources, rather than generating it personally. Some sources for secondary data collection include:

- Kaggle**: Kaggle is an online community of data scientists where you can access different types of data
- .gov datasets**: Countries like Australia, EU, India, New Zealand, and Singapore are openly sharing datasets on various portals
- Google Dataset Search**: This is a toolbox by Google that can search for data by name
- UCI Machine Learning Repository**: UCI is a collection of databases, domain theories, and data generators in collaboration with the University of Massachusetts

2.2.3 Best Practices for Acquiring Data

Checklist of factors that make data good or bad



Data acquisition from Websites:

The process of collecting data from websites using software is called Web Scraping.

WEB SCRAPING



Ethical concerns in data acquisition

While gathering data and choosing datasets, certain ethical issues can be addressed before they occur.

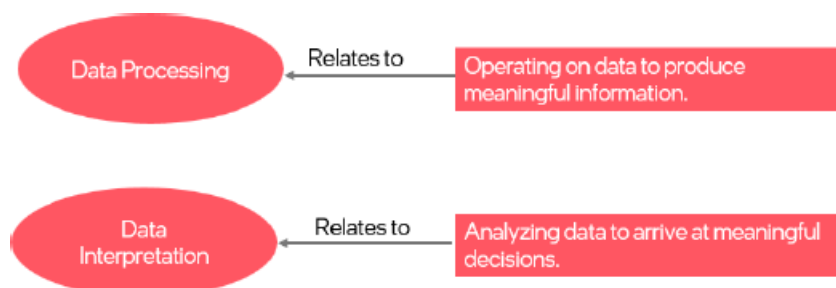


2.2.4 Data Processing and Data Interpretation

Data processing and interpretation have become very important in today's world

Can you answer this?

- Niki has 7 candies, and Ruchi has 4 candies
- How many candies do Niki and Ruchi have in total?
- We can answer this question using data processing
- Who should get more candies so that both Niki and Ruchi have an equal number of candies?
- How many candies should they get?
- We can answer this question using data interpretation



Data Processing

- Data processing helps computers understand raw data.
- Use of computers to perform different operations on data is included under data processing.

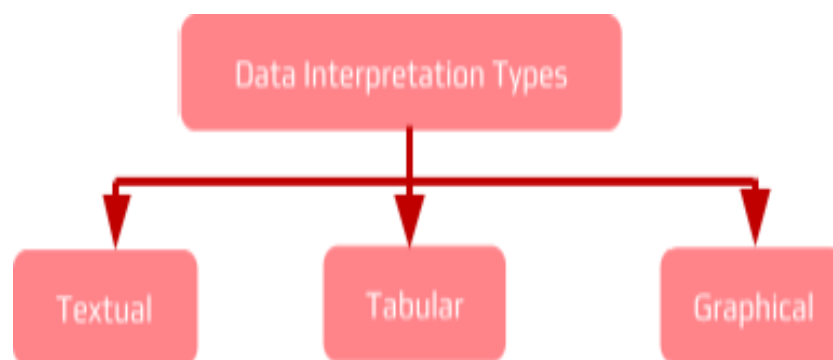
Data Interpretation

- It is the process of making sense out of data that has been processed.
- The interpretation of data helps us answer critical questions using data.



Types of Data Interpretation

There are three ways in which data can be presented:



Textual DI

- The data is mentioned in the text form, usually in a paragraph.
- Used when the data is not large and can be easily comprehended by reading.
- Textual presentation is not suitable for large data.

• Example:

In the Science Olympiad class of 45 Students, 3 students obtained the perfect score of 50. 10 students got a score of 45 and above, 15 students got a score of 40 and above, 8 students got a score of 30 and above, 6 students got a score of 20 and above and 3 got 19 and below.

More than 60% of students scored more than 80% Marks in Olympiad!

Tabular DI

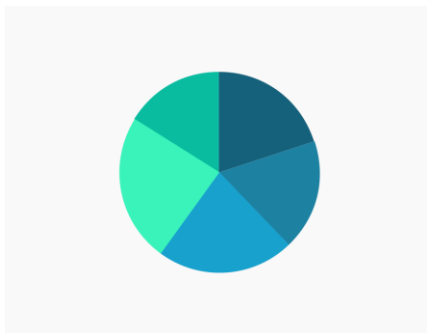
- Data is represented systematically in the form of rows and columns.
- Title of the Table (Item of Expenditure) contains the description of the table content.
- Column Headings (Year; Salary; Fuel and Transport; Bonus; Interest on Loans; Taxes) contains the description of information contained in columns.

Year	Item of Expenditure				
	Salary	Fuel and Transport	Bonus	Interest on Loans	Taxes
1998	288	98	3.00	23.4	83
1999	342	112	2.52	32.5	108
2000	324	101	3.84	41.6	74
2001	336	133	3.68	36.4	88
2002	420	142	3.96	49.4	98

Graphical DI

Bar Graphs

In a Bar Graph, data is represented using vertical and horizontal bar.



Pie Charts

- Pie Charts have the shape of a pie and each slice of the pie represents the portion of the entire pie allocated to each category
- It is a circular chart divided into various sections (think of a cake cut into slices)
- Each section of the pie chart is proportional to the corresponding value

Assignment

1. What is Data Literacy?
2. Differentiate between Data Privacy and Data Security.

In what ways are data privacy and data security different from each other?
3. Explain the Data Pyramid. Describe the four stages of the data pyramid and their significance.
4. What is Quantitative Data? Give an example of quantitative data and explain how it is used.
5. What is Qualitative Data? Provide an example of qualitative data and its relevance in data analysis.
6. List Two Best Practices for Data Security.
7. What is Data Processing? Explain the role of data processing in understanding raw data.
8. How does Data Interpretation help us? Describe how data interpretation helps us answer critical questions.
9. What are the Types of Data Interpretation?