

# **SNS COLLEGE OF TECHNOLOGY**

**Coimbatore-35**  
**An Autonomous Institution**



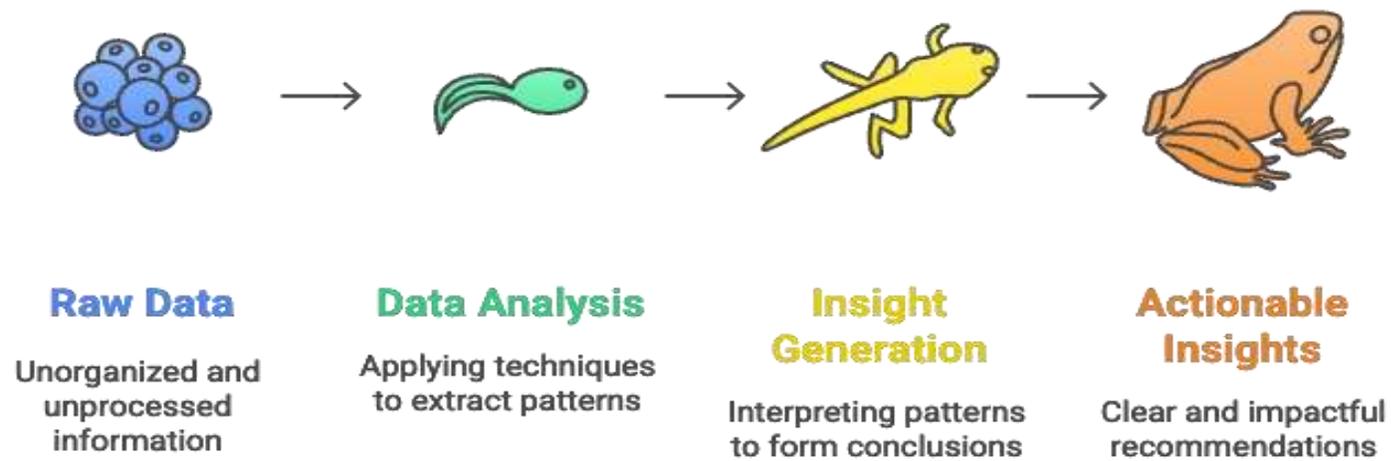
**Department Of Artificial Intelligence and Machine  
Learning**

**23AMO305 – Foundation of Data Science**

**UNIT-1 Data Science: Benefits and uses**

# Recap

## Data to Actionable Insights



Made with  Napkin

# What We'll Learn Today?

## Types of Data

- Structured Data
- Unstructured Data
- Semi-Structured Data

## Sources of Data

- Internal Sources
- External Sources
- Machine-Generated Data
- Human-Generated Data



# Problem Statement

- Data is collected from multiple sources in different formats
- Analysts face challenges such as:
- Handling structured and unstructured data
- Data inconsistency and redundancy
- Choosing the right data source for analysis



Made with Napkin

# Empathize



## Users of data:

- Data analysts

- Data scientists

- Business decision-makers

## User needs:

- Reliable and high-quality data

- Easy access to data sources

- Well-organized data formats

## Pain points:

- Poor data quality

- Missing or incomplete data

- Unclear data structure

# Define

- **Structured Data**
- Data that is **organized in rows and columns**.
- Easy to store in Excel sheets or databases.
- Examples: sales records, customer details, attendance logs.
- **2. Unstructured Data**
- Data with **no fixed format**.
- Harder to analyze directly.
- Examples: emails, images, videos, social media comments.

## 3. Semi-Structured Data

- Not fully structured but contains **tags or markers**.
- Easier to interpret than unstructured data.
- Examples: JSON files, XML files, sensor logs.

# Ideate

## Methods to handle different data types:

Data preprocessing and cleaning

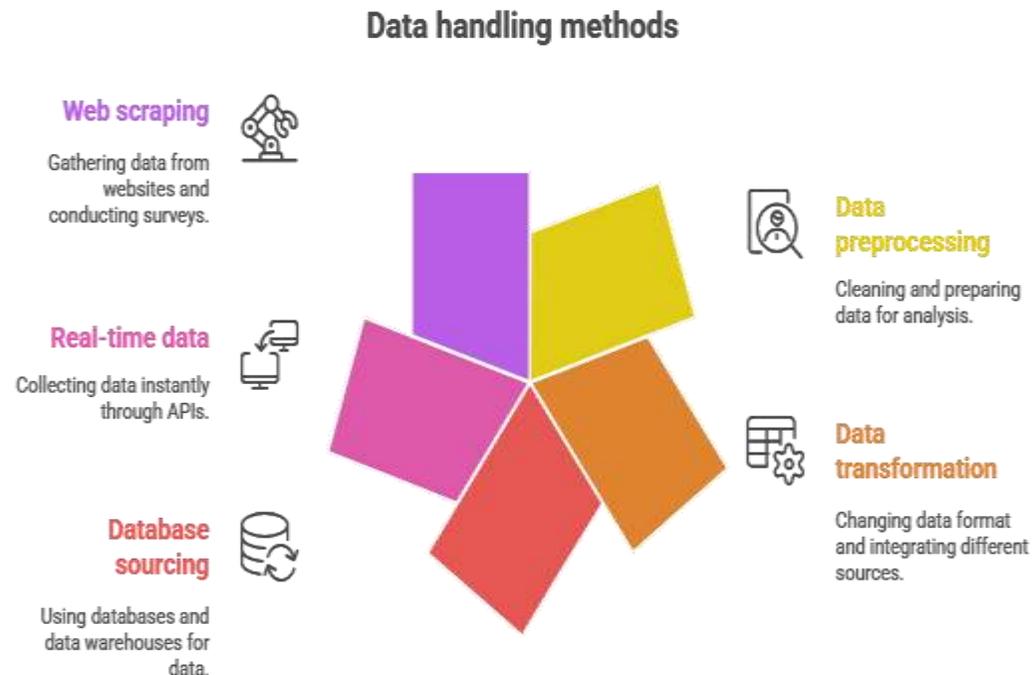
Data transformation and integration

## Strategies for data sourcing:

Using databases and data warehouses

Collecting real-time data via APIs

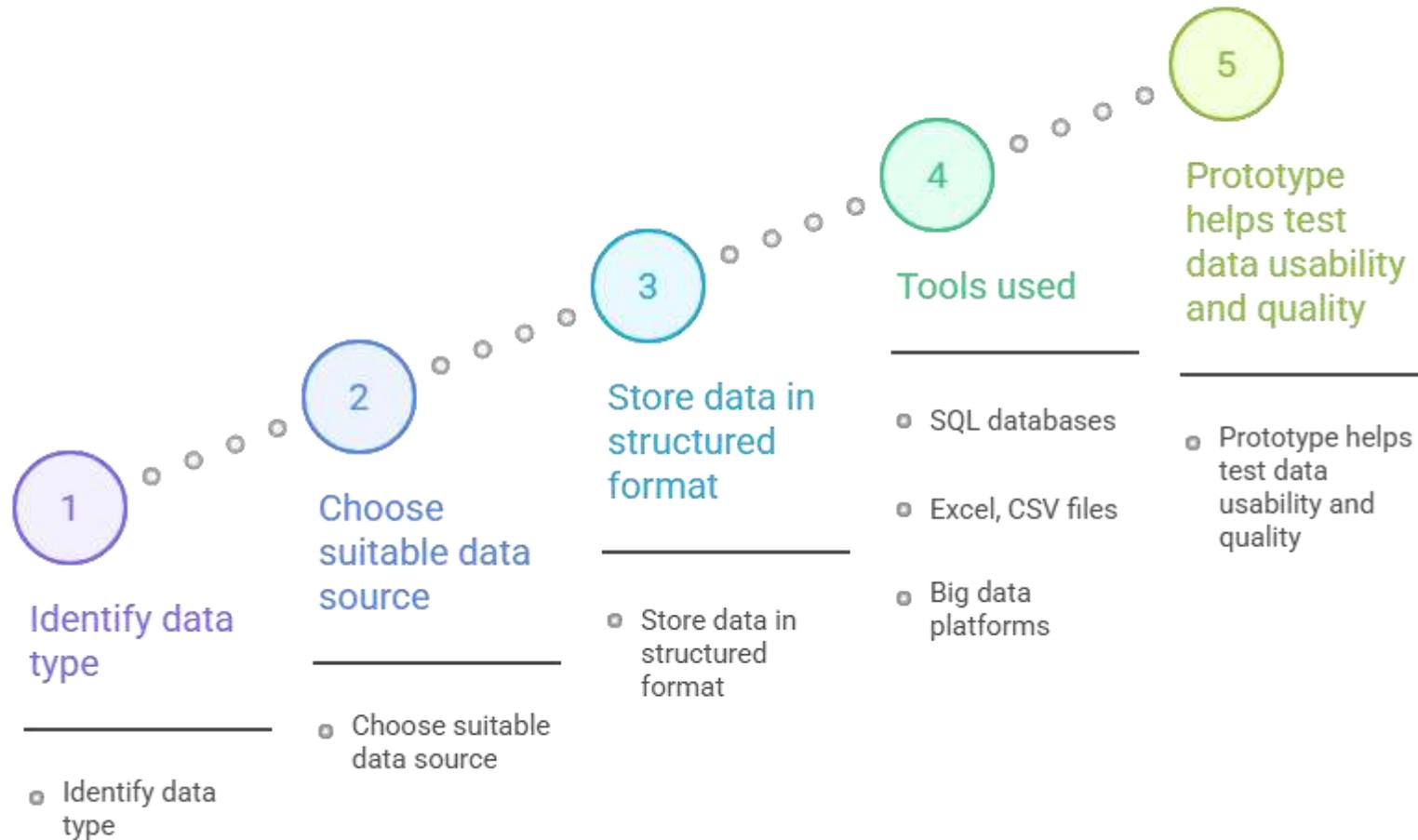
Web scraping and surveys



Made with  Napkin

# Prototype

## Build sample datasets from multiple sources



Made with  Napkin

# Data Type Comparison

## Data Types Comparison

Characteristic	Structured Data	Unstructured Data	Semi-Structured Data
 <b>Format</b>	Organized in rows and columns	No fixed format	Contains tags or markers
 <b>Storage</b>	Easy to store in Excel/databases	Harder to analyze directly	Easier to interpret than unstructured
 <b>Examples</b>	Sales records, customer details	Emails, images, videos	JSON files, XML files

Made with  Napkin

# Sources of Data

- **1. Internal Sources**

- Data generated **within the organization.**

- Examples: sales databases, HR records, inventory data, transaction logs.

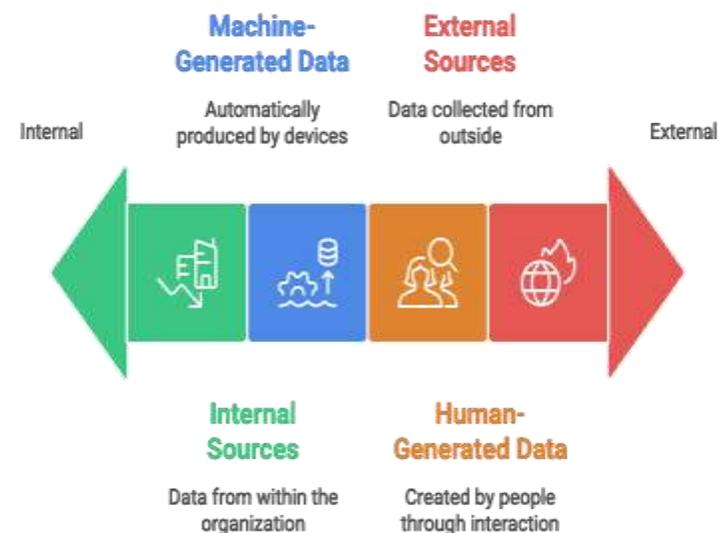
- **2. External Sources**

- Data collected **from outside the organization.**

- Examples: market reports, government data, social media analytics, public datasets.

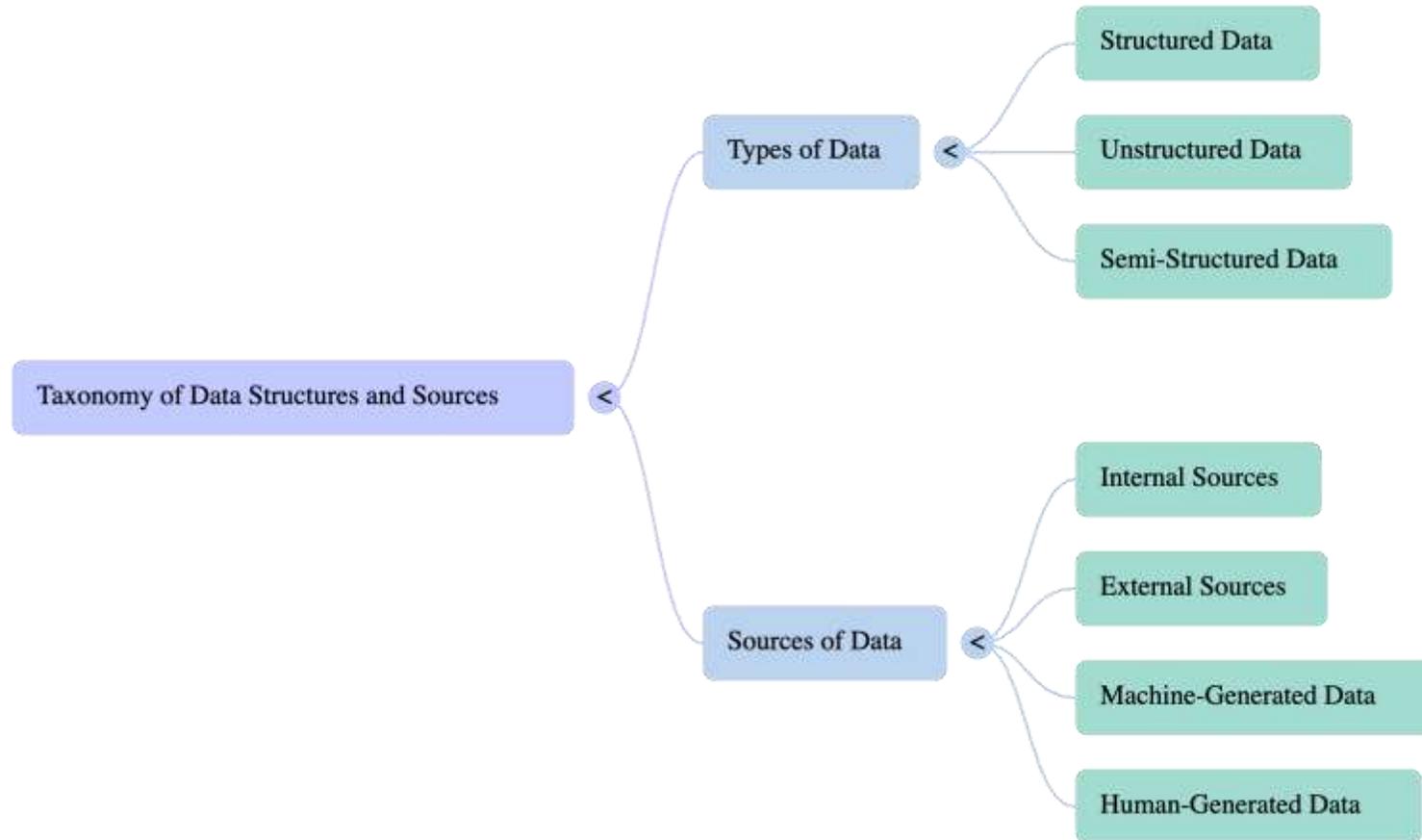
- **3. Machine-Generated Data**
- Automatically produced by devices or systems.
- Examples: sensor data, IoT devices, server logs, web tracking.
- **4. Human-Generated Data**
- Created by people through interaction.
- Examples: surveys, feedback forms, reviews, emails.

Data sources range from internal to external generation.



Made with  Napkin

# Mind Map



# The Data Detective Case!!!

- A data analyst walks into a college campus and collects the following clues:

**1 Clue A:**

“Rows, columns, and roll numbers — I love discipline.”

**2 Clue B:**

“I’m a mess of selfies, voice notes, and memes.”

**3 Clue C:**

“I look messy, but I do have tags and brackets.”

**4 Clue D:**

“I come from the college office database.”

**5 Clue E:**

“I come from Instagram and weather websites.”

**6 Clue F:**

“I never sleep — sensors and machines create me.”

**7 Clue G:**

“I’m created by humans... mostly while complaining or posting.”

# Conclusion

- Data exists in many forms—structured, unstructured, and semi-structured.
- Organizations rely on multiple data sources to understand trends and make decisions.
- Knowing the type and source of data helps analysts choose the right tools and methods for analysis.

Thank  
you!