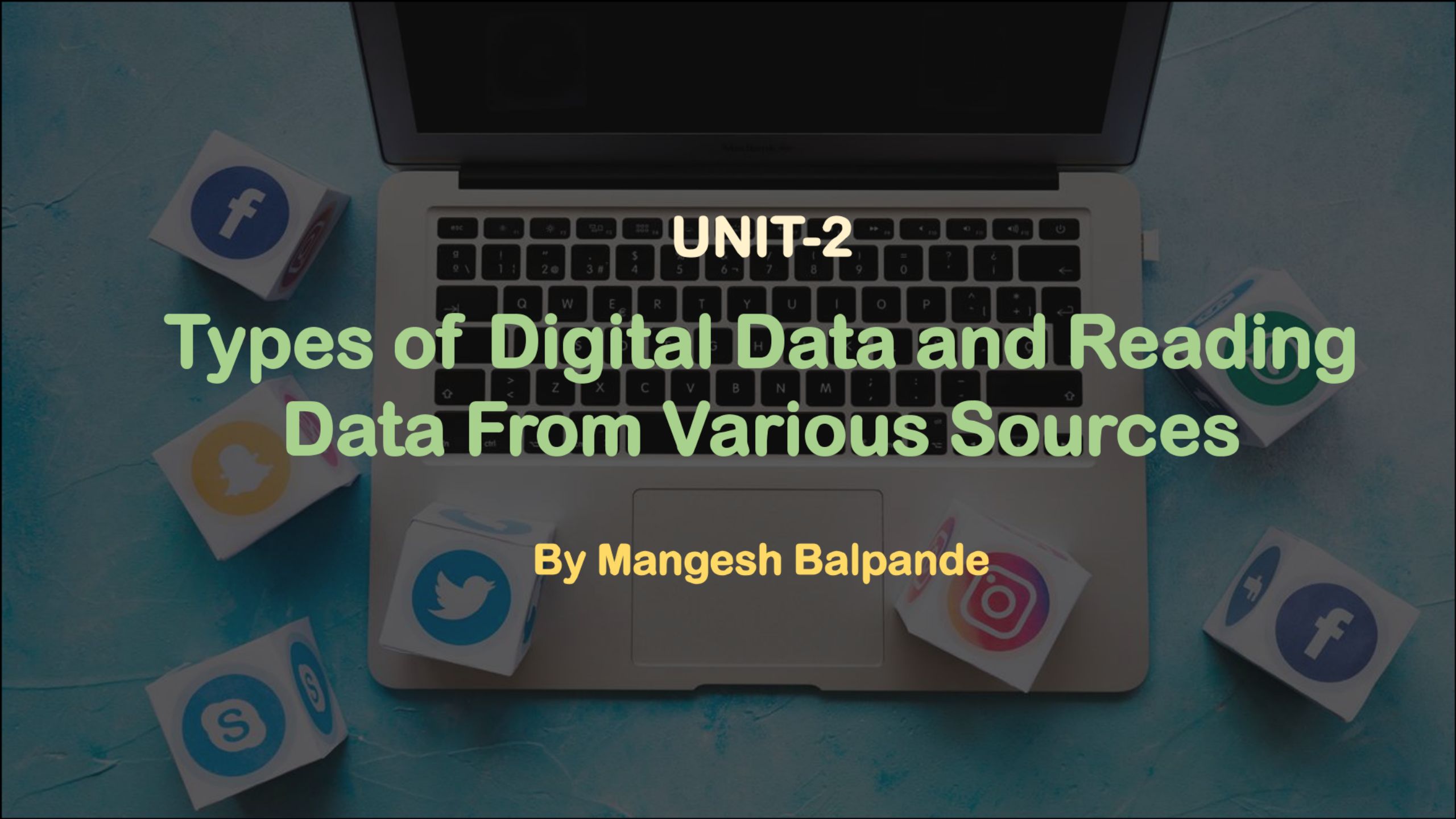


UNIT-2

Types of Digital Data and Reading Data From Various Sources

By Mangesh Balpande

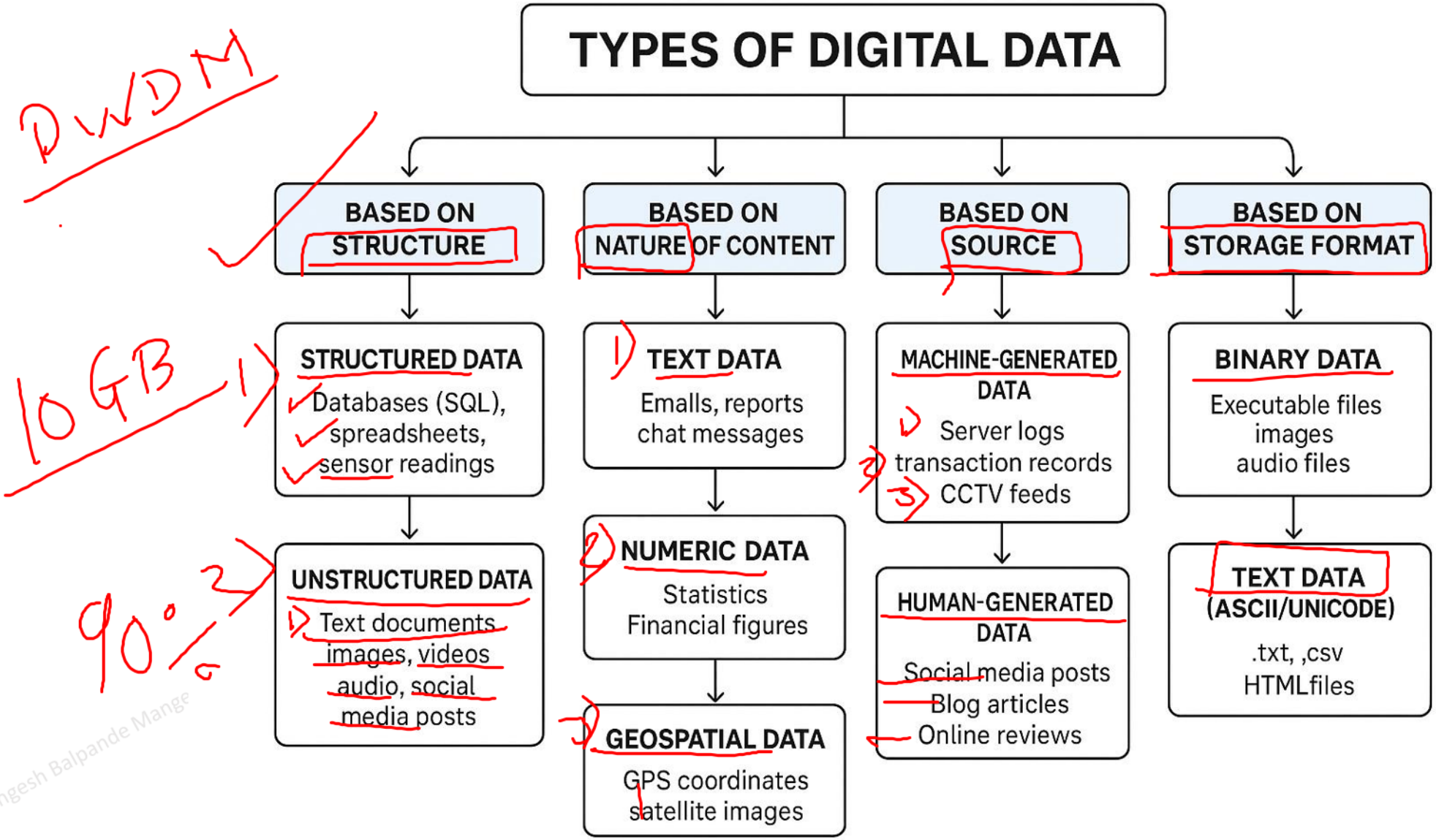
Contents:

- Types of Digital Data
- Read from Excel
- Read Data from .csv File
- Load a Python Dictionary into a Data Frame
- Reading JSON data
- Reading Data from Microsoft Access Database
- Reading Data from .txt File
- Reading Data from .xml file

Practical

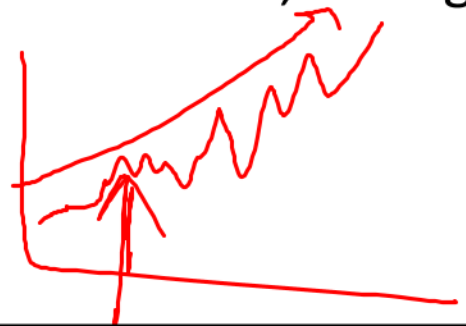
Mangesh Balpande Mangesh Balpande Mangesh Balpande Mangesh Balpande Mangesh Balpande Mangesh Balpande Mangesh Balpande Mangesh Balpande Mangesh Balpande Mangesh Balpande

Types of Digital Data



Quantitative (Numerical) Data

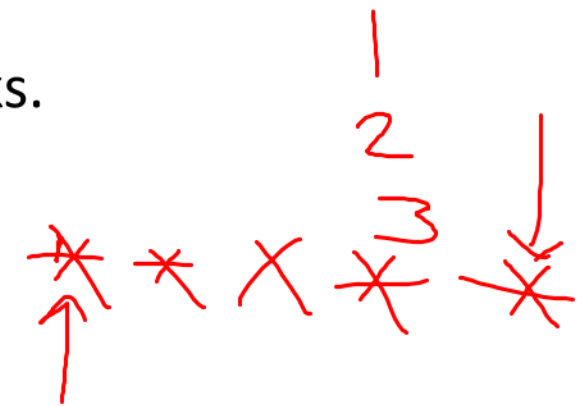
- ***Definition:*** Data that represents measurable quantities and can be expressed in numbers. This is the most common type used in analytics because it supports arithmetic operations.
- Types
  1. Discrete Data: Countable items, often whole numbers.
     - Ex: Number of students in a class, cars in a parking lot, books sold.
     - Visualizations: Bar Chart, Column Chart
  2. Continuous Data: Values that can take any number within a range.
     - Ex: Temperature, height, weight, revenue, rainfall.
     - Visualizations: Line Chart (trends), Area Chart, Histogram, Scatter Plot



Mangesh Balpande

Qualitative (Categorical) Data

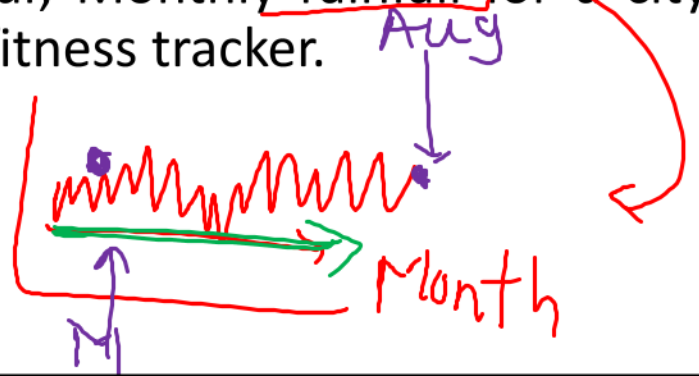
- ***Definition:*** Data that describes categories, labels, or qualities rather than measurements. It is non-numeric (though numbers may be used as labels, like categories).
- Types
  1. Nominal Data: Categories with no inherent order.
     - Ex: Education levels, gender, sport types .
     - Visualizations: Bar Chart, Pie Chart
  2. Ordinal Data: Categories with a natural order, but differences between ranks aren't necessarily uniform.
     - Ex: Education levels Survey responses, Military ranks.
     - Visualizations: Bar Chart, heatmap



Mangesh Balpande

Time Series Data

- ***Definition:*** Data collected and recorded in chronological order, where time is a key variable. It shows how data changes over seconds, minutes, hours, days, months, or years.
- Always has a time component (date/time stamp).
- The order matters — unlike other data types, shuffling time-series data breaks meaning.
- Often used for trend analysis, forecasting, and anomaly detection.
- Ex: Stock market closing prices per day, website visits per hour, Monthly rainfall for a city, Yearly GDP growth rates, Heart rate readings per second from a fitness tracker.



Algo trading

Mangesh Balpande

Geospatial Data

- ***Definition:*** Data that is tied to a specific location or geographic position on Earth. It usually contains latitude & longitude or other spatial identifiers like addresses, postal codes, or region names.

- Often used to analyze where events happen.

- Can be combined with other data types (quantitative or qualitative) for spatial insights.

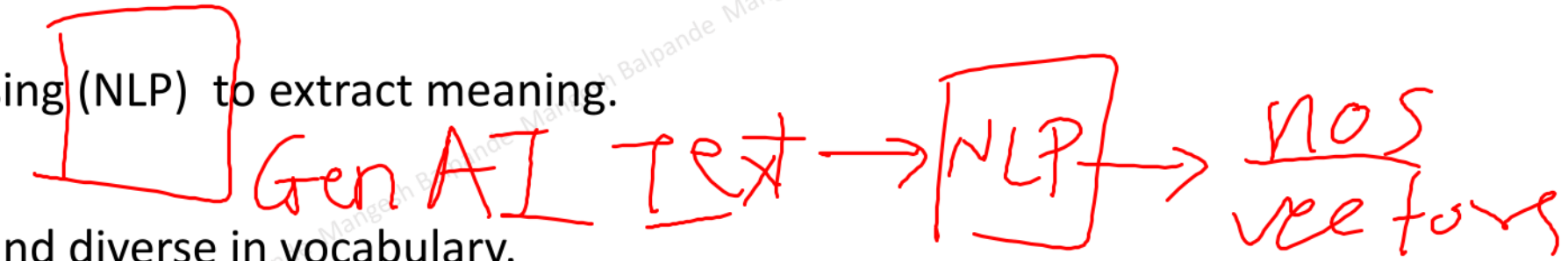
- Requires special visualization tools like GIS (Geographic Information Systems) for complex mapping.

- Ex: GPS coordinates from a delivery vehicle, Earthquake locations with magnitude readings, Sales performance by state or country, Population density by city, Weather radar maps.

Mangesh Balpande

Text Data

- ***Definition:*** Data made up of words, sentences, or documents — often unstructured or semi-structured. It's not numerical by default but can be processed into numbers for analysis.
- May come from human communication or written content.
- Needs text processing (NLP) to extract meaning.
- Can be very large and diverse in vocabulary.
- Ex: Social media posts, Customer reviews on Amazon or Google, News articles, Email messages and chat logs.



Mangesh Balpande

Structured Vs Unstructured Data

Structured data	Unstructured data
It is Organized, typically in tables or databases	It is having No predefined format
It Follows a predefined schema	It have No fixed schema, flexible
It is Easily stored in databases or spreadsheets	It Requires specialized storage solutions
It is Simple and straightforward	It Often requires advanced search algorithms
It is Well-suited for quantitative analysis	It Requires specialized techniques (NLP, etc.)
It is Easily processed using traditional methods	It Requires advanced processing techniques
It is having Low complexity due to structured format	It has High complexity due to lack of structure
It is Typically smaller in size	It Can be larger due to multimedia content
It can be Easily managed with traditional DBMS	It May require NoSQL or other specialized DBMS
It is Highly searchable using SQL queries	It is Less searchable, often relies on metadata
It can be stored in Databases, spreadsheets, CSV files	It can be stored in Text documents, emails, images, videos, Audio