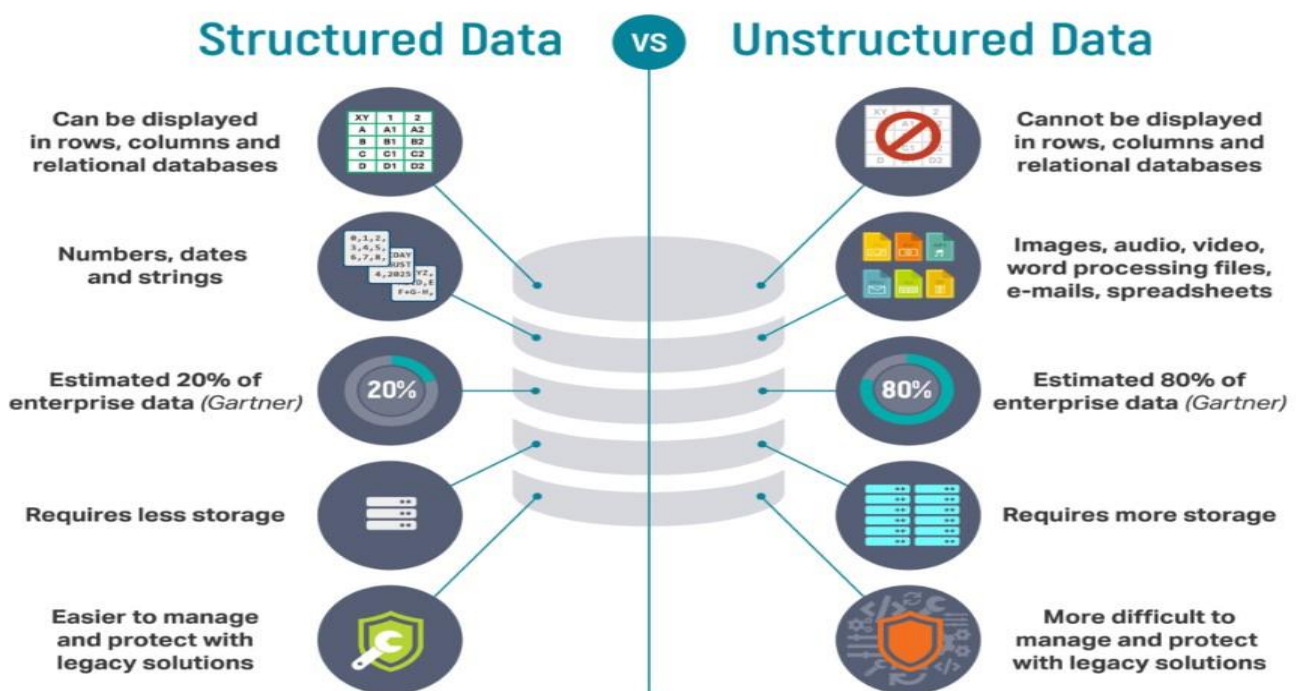# ➢ Differentiate between structured, semi-structured, and unstructured data.

➢ Structured data is highly specific and is stored in a predefined format, such as a table. This makes it easy to organize and analyze. Unstructured data is a compilation of many varied types of data that are stored in their native formats, such as text, audio, and video. This makes it more difficult to organize and analyze.
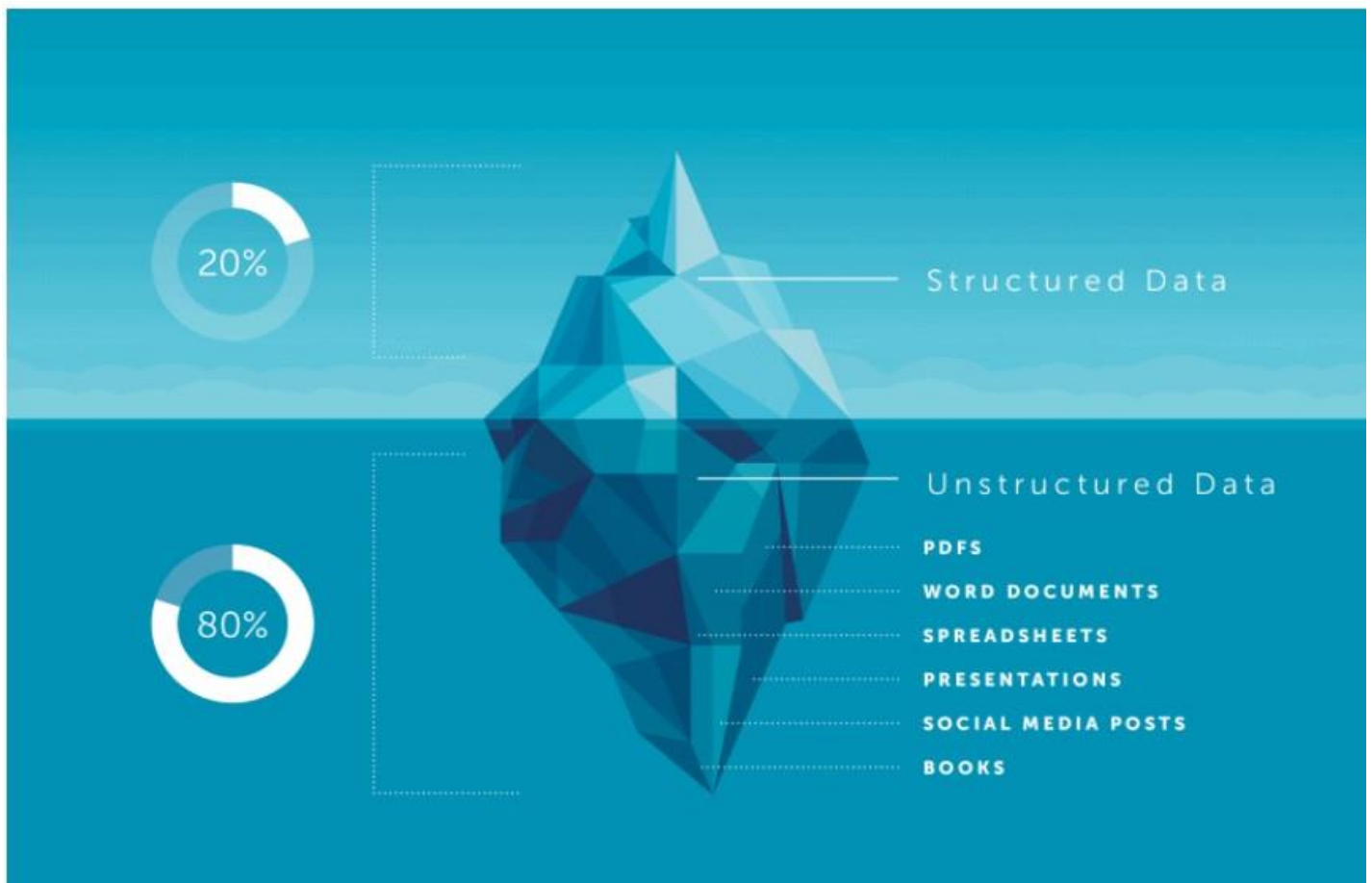
➢ Here is a table comparing the two types of data:

Feature	Structured data	Unstructured data
Format	Predefined	Native
Organization	Easy	Difficult
Analysis	Easy	Difficult
Examples	Tables, spreadsheets, databases	Text, audio, video, images

Structured Data VS **Unstructured Data**

Can be displayed in rows, columns and relational databases — Cannot be displayed in rows, columns and relational databases

Numbers, dates and strings — Images, audio, video, word processing files, e-mails, spreadsheets

Estimated 20% of enterprise data (Gartner) — Estimated 80% of enterprise data (Gartner)

Requires less storage — Requires more storage

Easier to manage and protect with legacy solutions — More difficult to manage and protect with legacy solutions

➢ What Is Data?

- Data is a set of facts such as descriptions, observations, and numbers used in decision making.
- We can classify data as structured, unstructured, or semi-structured data.

1) **What is structured data:**

- **Structured data** is generally tabular data that is represented by columns and rows in a database.
- Databases that hold tables in this form are called *relational databases*.
- The mathematical term "*relation*" specify to a formed set of data held as a table.
- In structured data, all row in a table has the same set of columns.
- SQL (Structured Query Language) programming language used for structured data.

id	name	age
1	Jim	28
2	Pam	26
3	Michael	42

id	subject	Teacher
1	Languages	John Jones
2	Track	Wally West
3	Swimming	Arthur Curry
4	Computers	Victor Stone

student_id	subject_id	grade
2	1	98
1	2	100
1	4	75
3	3	60
2	4	76
3	2	88

2) What is Semi-structured Data:
3)

- **Semi-structured** data is information that doesn't consist of Structured data (relational database) but still has some structure to it.
- Semi-structured data consist of documents held in *JavaScript Object Notation* **(JSON) format**. It also includes *key-value* stores and *graph* databases.

```
## Document 1 ##
{
  "customerID": "103248",
  "name":
  {
    "first": "AAA",
    "last": "BBB"
  },
  "address":
  {
    "street": "Main Street",
    "number": "101",
    "city": "Acity",
    "state": "NY"
  },
  "ccOnFile": "yes",
  "firstOrder": "02/28/2003"
}
```

3)What is Unstructured Data:

- **Unstructured data** is information that either does not organize in a pre-defined manner or not have a pre-defined data model.
- Unstructured information is a set of text-heavy but may contain data such as numbers, dates, and facts as well.
- **Videos, audio, and binary** data files might not have a specific structure. They're assigned to as **unstructured** data.

Structured Data vs Unstructured Data vs Semi-Structured:

Structured data is stored is predefined format and is highly specific; whereas **unstructured data** is a collection of many varied data types which are stored in their native formats; while **semi structured data** that does not follow the tabular data structure models associated with relational databases or other data table forms.

Pros and Cons of Structured Data:

Pros	Cons
Requires less processing in comparison to unstructured data and is easier to manage.	Limited usability because of its pre-defined structure/format
Machine algorithms can easily crawl and use structured data which simplifies querying	Structured data is stored in data warehouses which are built for space saving but are difficult to change and not very scalable/flexible.
As an older format of data, there are several tools available for structured data that simplify usage, management, and analysis	

Pros and Cons of Unstructured Data:

Pros	Cons
A variety of native formats facilitates a greater number of use cases and applications	The greater number of formats makes it equally challenging to analyze and leverage unstructured data.
As there is no need to predefine data, unstructured data is collected quickly and easily.	The large volume and undefined formats make data management a challenge and specialized tools a necessity.
Unstructured data is stored in on-premises or cloud data lakes which are highly scalable.	
Although challenging, the greater volume of unstructured data provides better insights and more opportunities to turn your data into a competitive advantage.	