Note: Solution for the exercises will be on GitHub.

Day 8: Types of Data & File Formats

- ☐ Types of data: structured, semi-structured, unstructured
- ☐ Common file formats used in analytics
- ☐ Reading CSV and Excel files using **Pandas**
- ☐ Quick data overview: .head(), .info(), .describe()

🔹 **1. Types of Data**

Type	Description	Example
Structured	Organized in rows and columns	Excel, CSV, SQL databases
Semi-Structured	Partially organized but not in strict schema	JSON, XML
Unstructured	No specific format	Images, videos, audio, emails

---

🔹 **2. Common File Formats in Data Analytics**

Format	Use Case	Tool/Library Used
CSV	Clean tabular data	pandas.read_csv()
Excel	Business data	pandas.read_excel()
JSON	API data, configs	pandas.read_json()
SQL	Databases	pandas.read_sql() + sqlite3

🔸 **3. Code Examples (Python)**

👉 **Read a CSV file**

```
import pandas as pd


data = pd.read_csv("sample.csv")
print(data.head())       # First 5 rows
print(data.info())       # Structure and types
print(data.describe())   # Stats for numeric columns
```

👉 **Read an Excel file**

Note: Solution for the exercises will be on GitHub.

```
data = pd.read_excel("sample.xlsx")

print(data.head())
```

📝 You can use .to_csv() or .to_excel() to **save** files as well.

🧠 Mini Exercise

Download or create a small CSV file like:

Name, Age, Department

Alice, 30, HR

Bob, 24, Finance

Clara, 29, Marketing

Task:

- Load the CSV using Pandas
- Print the first 2 rows
- Print the average age

```
import pandas as pd

df = pd.read_csv("employees.csv")

print(df.head(2))

print("Average age:", df["Age"].mean())
```

🧰 Tools Setup (Optional for today)

If you're not using Jupyter yet, try running:

```
pip install notebook

jupyter notebook
```