# DATA 30

# WEEK 2

---

## Day 8: Pandas – Basics

### 📌 Topics to Cover

- **Introduction** – What is Pandas? Why it's essential for data handling and analysis.
- **Installing Pandas** – pip install pandas.
- **Data Structures** – Series (1-D) and DataFrame (2-D).
- **Creating DataFrames** – From lists, dicts, NumPy arrays, and CSV files.
- **Reading and Writing Data** – read_csv(), to_csv(), and file operations.
- **Inspecting Data** – head(), tail(), info(), describe().
- **Selecting Data** – Indexing and slicing with loc[] and iloc[].

### 👉 Task:

Import a CSV file containing student marks. Display basic info, top 5 rows, and summary statistics.

---

## Day 9: Pandas – Data Cleaning & Transformation

### 📌 Topics to Cover

- **Handling Missing Data** – isnull(), dropna(), fillna().
- **Data Filtering** – Conditional selection with boolean indexing.
- **Sorting & Ranking** – sort_values(), rank().
- **String Functions** – Using str methods on columns.
- **Data Transformation** – apply(), map(), lambda functions.
- **Aggregation & Grouping** – groupby() and aggregation methods (mean, sum, count).

👉 **Task:**

Clean a dataset with missing values and group it by a specific column to find the average of another column.

---

## Day 10: Matplotlib – Data Visualization Basics

📌 **Topics to Cover**

- **Introduction** – Purpose of data visualization.
- **Installing Matplotlib** – pip install matplotlib.
- **Basic Plots** – Line, Bar, Scatter, and Histogram using plt.plot() and plt.bar().
- **Plot Customization** – Titles, labels, legends, colors, and styles.
- **Multiple Plots** – Using subplot() for side-by-side graphs.

👉 **Task:**

Visualize the growth of sales over months using a line chart and customize titles and axes.

---

## Day 11: Seaborn – Advanced Visualization

### 📌 Topics to Cover

- **What is Seaborn?** – A statistical visualization library built on Matplotlib.
- **Installing Seaborn** – pip install seaborn.
- **Dataset Functions** – load_dataset(), sns.get_dataset_names().
- **Common Plots** – barplot(), countplot(), boxplot(), violinplot(), pairplot().
- **Styling** – Themes and color palettes (set_style(), set_palette()).
- **Integration with Pandas** – Directly plot from DataFrames.

### 👉 Task:

Load the *tips* dataset and create a boxplot showing the distribution of tips by day.

---

## Day 12: Combined Data Analysis Project



### 📌 Topics to Cover

- **Reading Real Data** (CSV or Excel).
- **Data Cleaning and Transformation** using Pandas.
- **Data Visualization** using Matplotlib & Seaborn.
- **Analyzing Insights** – Trends, correlations, and patterns.

### 👉 Task:

Perform EDA (Exploratory Data Analysis) on a real dataset (e.g., Iris or Sales Data).

Include data cleaning, summary statistics, and visualizations.

---

## Day 13: Introduction to Other Important Libraries

### 📌 Topics to Cover

- **NumPy Advanced** – Broadcasting, vectorization, matrix operations.
- **SciPy** – Statistical functions and scientific computations.
- **Statsmodels** – Intro to statistical modeling.
- **Sklearn Basics** – Introduction to datasets and preprocessing modules.

### 👉 Task:

Use NumPy and SciPy to perform a simple statistical test (mean comparison or correlation).

---

## Day 14: Revision + Quiz/Test

### 📌 Topics to Cover

- **Revision of Week 2 Concepts:**

– Pandas (Basics, Cleaning, Transformation)

– Matplotlib & Seaborn (Visualization)

– Other Libraries Overview

### 👉 Activity:

A quiz/test covering key Week 2 concepts with practical and theoretical questions.

---

⚡ **End Goal:**

After Week 2, you'll be confident in handling, cleaning, and visualizing data using Python libraries like Pandas, Matplotlib, and Seaborn — the core skills for data analysis and machine learning.

---

**Reference Links:**

https://youtu.be/vmEHCJofslg

https://youtu.be/UB3DE5Bgfx4