IA2 - AP_DS- 2025/2026 - Tutorial 04- Project

Monthly Sales Analysis (NumPy + Pandas + Matplotlib + Seaborn)

Objective

The goal is to **generate, analyze, and visualize monthly sales data** for **four products over one year**, and extract key business insights using **NumPy, Pandas, Matplotlib, and Seaborn**.

Project Structure

```
project_sales/
│
├─ notebook.ipynb         # Main notebook with all analysis and visualizations
├─ utils.py               # Functions for data generation
├─ data/
│   ├─ initial.csv        # Raw generated dataset
│   ├─ final.csv          # DataFrame with calculated metrics
│   └─ output.csv         # Final results including pivot tables
```

Instructions

1. Data Generation

- Generate a series of monthly dates for one year (e.g., from '2025-01-01' to '2025-12-01').
- Define in utils.py a function generate_random_sales(min_val, max_val, size) that returns a random NumPy array of integers between min_val and max_val for the given size.
- Use this function to generate random monthly sales (integers) for 12 months for four products:
  - Product A: 50–100 units
  - Product B: 30–80 units
  - Product C: 20–60 units
  - Product D: 10–50 units
- Create a DataFrame with columns: Date, Product_A, Product_B, Product_C, Product_D.
- Save this initial dataset as `initial.csv`.

2. Build DataFrame

- Create a Pandas DataFrame with columns: `Month`, `Product_A`, `Product_B`, `Product_C`, `Product_D`.
- Compute monthly metrics:
  - `Total_Sales`: sum of all products per month
  - `Average_Sales`: mean sales per month
  - `Month_over_Month_Growth`: percent change of `Total_Sales` vs previous month
- Assign each month to a quarter:

    - Q1: Jan–Mar, Q2: Apr–Jun, Q3: Jul–Sep, Q4: Oct–Dec
- Add additional columns:
    - `Max_Sales_Product` : product with the highest sales each month
    - `Min_Sales_Product` : product with the lowest sales each month
- Save this updated DataFrame as `final.csv` .

3. Pivot Tables & Summaries

- Compute average sales per quarter for each product and total sales using a pivot table.
- Compute total sales per quarter.
- Save the final output including pivot tables as `output.csv` .

4. Key Insights

- Identify the **best month** (highest total sales)
- Identify the **best product** (highest cumulative annual sales)
- Identify the **best quarter** (highest total sales)

5. Visualizations

- Line chart for each product across months
- Stacked bar chart of total monthly sales by product (annotate best month)
- Seaborn heatmap: monthly sales of each product
- Seaborn boxplot: distribution of sales per product

6. Conclusion Questions

- Which **product** contributes the most to overall sales throughout the year?
- Which **quarter** performs best and why might that be?
- How could this information be used to **improve sales strategy** for the next year?