Regional Sales Performance Analysis

A retail company wants to analyze its sales performance across different regions, product categories, and months. The goal is to uncover insights into sales trends, identify the best-performing regions, and understand the relationship between sales and product categories.

You are hired as a Python developer to perform this analysis using **NumPy**, **Pandas**, **Matplotlib**, and **Seaborn**.

Assignment

Scenario

The dataset (sales_data.csv) contains the following columns:

- Date: The date of the transaction.
- Region: The region where the sale occurred (e.g., North, South, East, West).
- Category: The product category (e.g., Electronics, Furniture, Clothing).
- Units_Sold: The number of units sold.
- Revenue: The total revenue generated for that transaction.
- Profit: The profit earned from the transaction.

Your Tasks

Step 1: Data Preparation

1. Load the dataset using Pandas.
2. Perform exploratory data analysis:
    - Check the structure of the dataset, data types, and summary statistics.
    - Identify and handle any missing or inconsistent data.
3. Create new columns:
    - Month: Extract the month from the Date column.

Step 2: Analysis Using Pandas and NumPy

1. Calculate the following:
    - Total revenue and profit for each region.
    - Total units sold for each product category.
    - Monthly sales trends (revenue) for the entire company.
2. Identify the best and worst-performing regions based on:
    - Total revenue.
    - Total profit.

Step 3: Visualizations Using Matplotlib and Seaborn

1. Create a bar chart showing total revenue by region.
2. Create a line plot showing monthly sales trends (revenue) across all regions.
3. Use a grouped bar chart to compare total revenue for each category across regions.
4. Create a heatmap to show the correlation between revenue, units sold, and profit.
5. Use a pie chart to display the proportion of total revenue contributed by each region.

Deliverables

1. A Python script (`sales_analysis.py`) containing:
    - Code for data analysis and visualization.
    - Proper comments explaining each step.
2. A PDF report with:
    - The visualizations.
    - A summary of insights and recommendations.

Bonus Task

- Use Seaborn's `FacetGrid` to create small multiples of revenue trends by category for each region.