Synopsis

Title: **Global Superstore Sales Analysis**
Name: **Aryan Verma**
Student ID: **AF0417100**

The dataset used for this analysis contains detailed information about sales transactions from a global superstore. Each row represents an individual sales order, and the columns provide various details, including:

1. **Row ID:** Unique identifier for each record.
2. **Order ID:** Unique ID of the order.
3. **Order Date:** Date when the order was placed.
4. **Ship Date:** Date when the order was shipped.
5. **Ship Mode:** Mode of shipping (e.g., Standard Class, Second Class).
6. **Customer ID:** Unique identifier for customers.
7. **Customer Name:** Name of the customer.
8. **Segment:** Customer segment (e.g., Consumer, Corporate, Home Office).

9. **City, State, Country, Market, Region:** Geographical details of the orders.
10. **Product ID:** Unique ID for each product.
11. **Category:** Main category of the product (e.g., Furniture, Office Supplies).
12. **Sub-Category:** Subcategory of the product.
13. **Product Name:** Name of the product.
14. **Sales:** Revenue generated from the order.
15. **Quantity:** Quantity of items purchased.
16. **Discount:** Discount applied to the order.
17. **Profit:** Profit earned on the order.
18. **Shipping Cost:** Cost incurred for shipping.
19. **Order Priority:** Priority of the order (e.g., High, Low).

Problem Statement

1. **Which product category generates the highest sales?**
2. **Which product category generates the highest profit?**
3. **Which sub-category contributes the most to sales and profit?**
4. **What are the total sales and profit for each region?**
5. **What is the trend of sales over time (monthly or yearly)?**
6. **Who are the top customers contributing the most revenue?**
7. **What are the top 5 countries by sales and profit?**
8. **Which segment has the most orders placed and sales?**
9. **Which market contributes the least profit and sales?**
10. **What are the top 5 products with the highest quantity sold (demand)?**
11. **What are the top 5 products sold with a discount above 10%?**
12. **Which products have sales greater than 10,000 in a specific year (e.g., 2013)?**
13. **Which product categories and sub-categories have total sales between 5,000 and 30,000?**

Data Preprocessing Steps

1. **Data Cleaning:**
    - Handle missing values in columns like Profit, Discount, or Shipping Cost.
    - Ensure consistency in categorical values (e.g., consistent naming for categories).
2. **Normalization:**
    - Scale numeric features like Sales and Profit if necessary for better comparison.
3. **Encoding Categorical Variables:**
    - Convert features like Segment, Ship Mode, and Order Priority into numeric form using label or one-hot encoding.

Implementation Process

1. **Data Ingestion:**
    - Load the dataset into a data analysis environment (e.g., Python, Power BI).
2. **Preprocessing:**
    - Clean and prepare the data for analysis.
3. **Exploratory Data Analysis (EDA):**
    - Uncover patterns in sales and profit based on variables like region, category, or discount.
4. **Visualization:**
    - Use **Power BI** to create visualizations such as:
        - Sales and profit by product category and sub-category.
        - Trend analysis for monthly or yearly sales.
        - Regional and customer segment breakdowns.
        - Top-performing products and countries.

5. **Reporting:**
    - Summarize findings in interactive dashboards or reports, focusing on actionable insights to improve sales and profitability.

Dataset

- **Global Superstore Sales**

Technologies

- **Pandas:** For data manipulation.
- **Power BI:** For interactive dashboards and reporting.

Software Requirements

- **Operating Systems:** Windows, Linux, macOS.
- **IDE:** Jupyter Notebook (for Python) or Power BI (for visualizations).

Hardware Requirements

- **RAM:** Minimum 8GB (for Power BI), recommended 16GB.
- **Processor:** Minimum Intel i5, recommended Intel i7 for faster data processing.
- **Storage:** SSD recommended, at least 256GB for smooth handling of large datasets.