EDA (Exploratory Data Analysis)

Exploratory Data Analysis (EDA) is the process of analyzing and summarizing datasets to uncover patterns, detect anomalies, and check assumptions using statistical and graphical methods. It is a crucial step in the data analysis pipeline, providing insights and understanding before applying machine learning models or making data-driven decisions.

Steps in EDA

1. **Data Understanding**:
    - Understanding the structure and characteristics of the data (columns, types, shape,etc.).
2. **Data Cleaning**:
    - Checking for missing values, duplicates, and outliers and Handling inconsistencies in data.
3. **Statistical Summary**:
    - Generating basic statistics like mean, median, mode, count, etc., for numerical data.
4. **Data Visualization**:
    - Creating charts and plots (e.g., histograms, bar plots, scatter plots) to identify trends and relationships in the data.
5. **Checking Relationships**:
    - Analyzing relationships between variables using correlation and cross-tabulations

Key business insights that can be derived from exploratory data analysis (EDA) on the provided datasets:

1. **Top-Selling Products and Categories**: Identify the products and categories generating the highest revenue by analyzing the total sales value and transaction frequency in "Transactions.csv ". This can help prioritize inventory and marketing efforts.
2. **Regional Performance Analysis**: Compare total sales and customer distribution across regions ("Customers.csv" and "Transactions.csv") to identify high-performing areas and regions requiring targeted promotions.
3. **Customer Behavior Trends**: Analyze purchase frequency and average transaction value per customer ("Customers.csv" and "Transactions.csv") to segment customers into high-value and low-value groups for personalized offers.
4. **Seasonal Demand Patterns**: Examine transaction dates in "Transaction.csv" to detect seasonal trends in product demand, aiding in inventory planning and promotional campaigns.
5. **New Customer Acquisition Impact**: Assess the relationship between signup dates ("Customers.csv") and transaction volumes to measure how quickly new customers make their first purchases and contribute to revenue.