

January 24, 2025

Business Report

On

Exploratory Data Analysis (EDA)

Prepared By:

Aditya Kumar Pandey

MTech (AI & Data Science)

Computer Science and Engineering

Indian Institute of Information Technology, Bhagalpur

Exploratory Data Analysis (EDA)

Introduction

The EDA process aims to understand the dataset, identify patterns, and extract business insights that can help in decision-making. The dataset consists of three files:

- 1) Transactions.csv - Contains purchase history details.
- 2) Customers.csv - Contains customer profile information.
- 3) Products.csv - Contains product details.

Data Overview

The data was loaded and inspected using `head()` and `info()` functions to understand the structure, columns, and data types.

Customer Data:

- Top five records was displayed using `customers_df.head()`
- Information about data types and missing values checked with `customers_df.info()`.

Transactions Data:

- Top five records was displayed using `transactions_df.head()`
- Information about data types and missing values checked with `transactions_df.info()`.

Data Merging:

The datasets were merged using `CustomerID` and `ProductID` to create a comprehensive dataset for analysis.

Exploratory Data Analysis

1. Top 5 Customers by Purchase Value

- The top customers were identified based on the total purchase value.

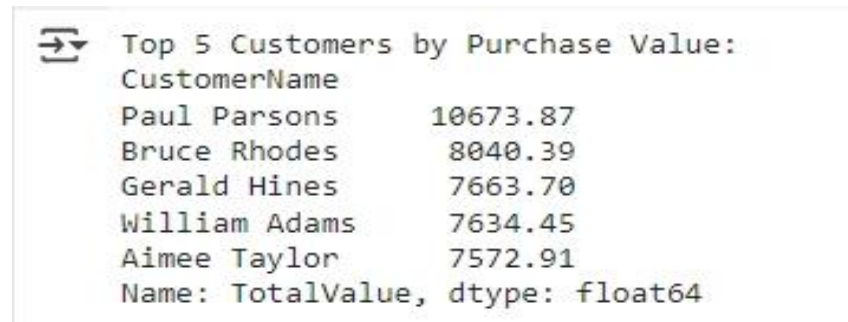


Fig: Top 5 Products by Purchase Value

- Explanation: The chart shows the top five customers who spent the most money. The x-axis shows customer names, and the y-axis shows the total money spent.

2. Top 5 Products by Sales Quantity

- The highest selling products were determined using:



Fig: Top 5 Products by Sales Quantity

Explanation: This chart shows the top five products sold in the highest quantity. The longer the bar, the more units sold.

3. Monthly Sales Trend

- This bar plot shows the monthly sales trend.

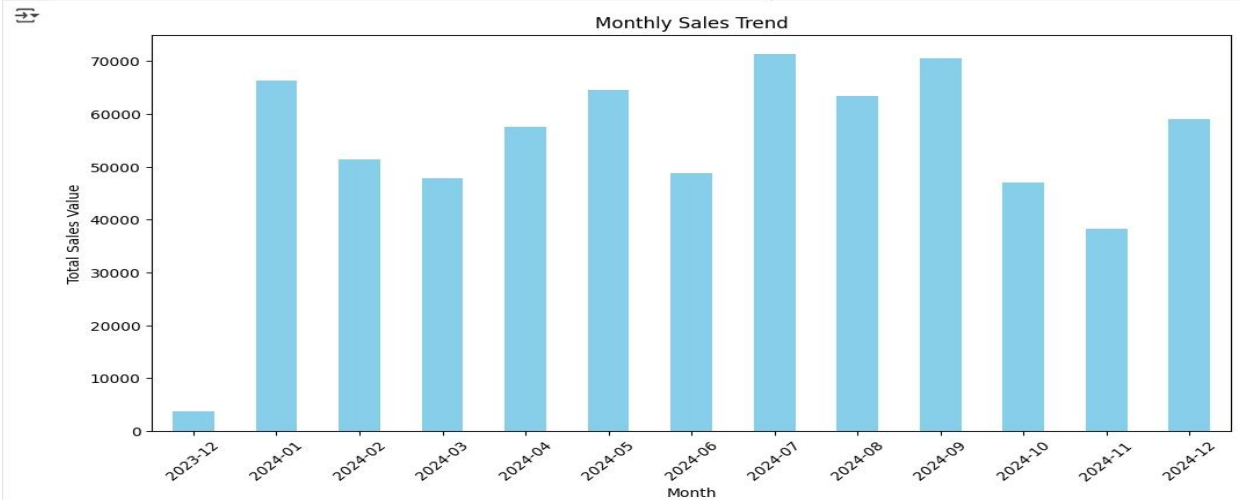


Fig: Monthly Sales Trend

- Explanation: This chart shows total sales for each month. The x-axis has months, and the y-axis has total sales.
- #### 4. Customer Distribution by Region
- This bar plot is showing the distribution of customers across different regions.

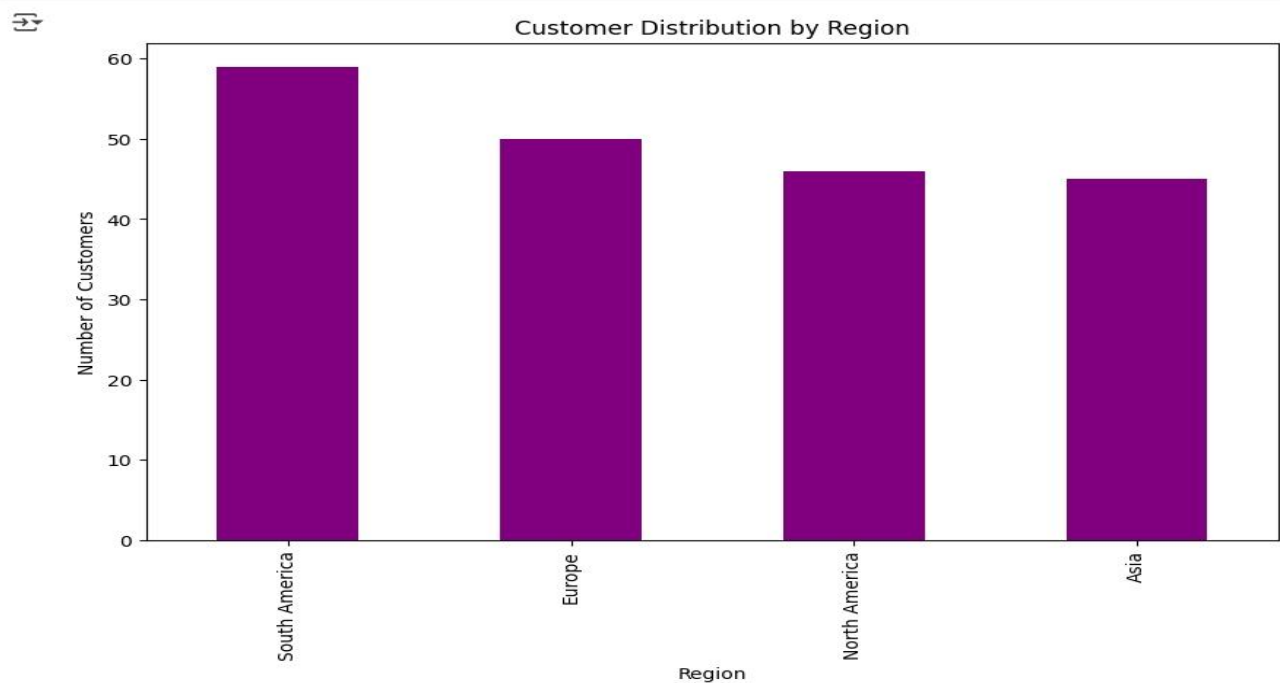


Fig: Customer Distribution by Region

- Explanation: This chart highlights the number of customers in each region, helping identify the most populated areas by customers.

5. Sales Performance by Product Category

- This bar plot is showing the revenue contribution from various product categories.



Fig: Sales Performance by Product Category

- Explanation: This chart helps understand which product categories earn the most revenue.

6. Seasonal Sales Trends

- This line plot is showing sales performance trends across different months.

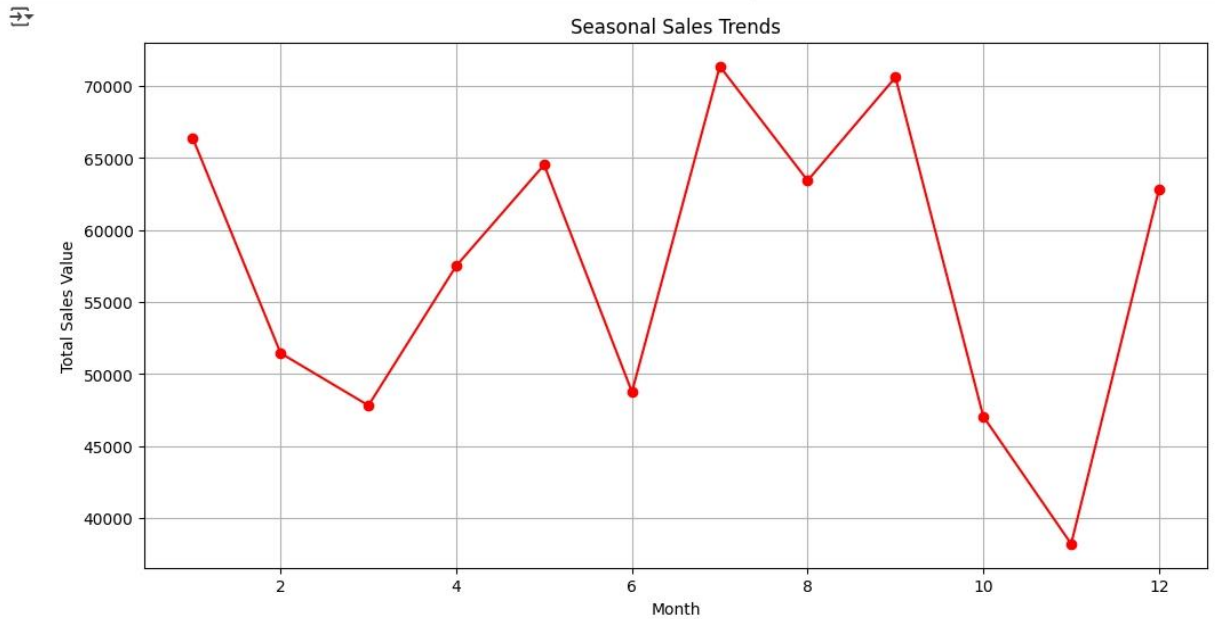


Fig: Seasonal Sales Trends

- Explanation: The line plot shows fluctuations in sales value across different months, indicating potential seasonal patterns and peak sales periods.

7. Top Customers Contribution to Revenue

- This pie chart is illustrating the revenue contribution from top customers.

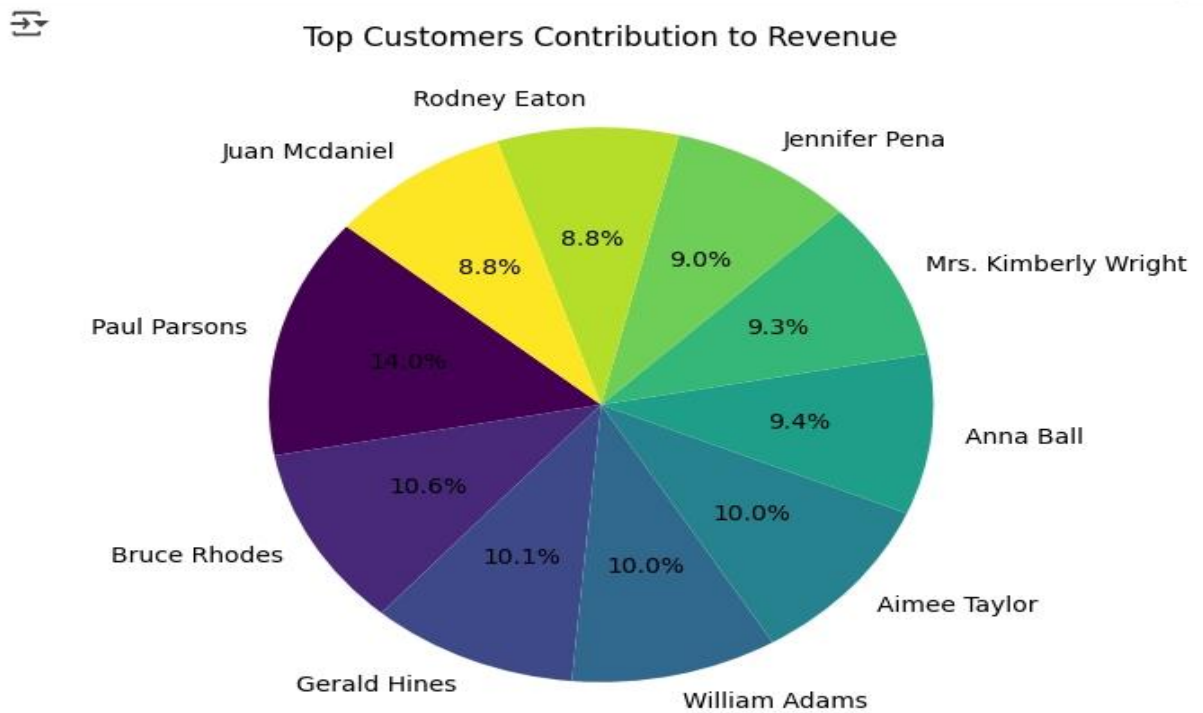


Fig: Top Customers Contribution to Revenue

Explanation: This pie chart visualizes the percentage contribution of the top customers to the total revenue.

8. High-Value vs Low-Value Product Comparison

- This bar plot is highlighting the contribution of high-value products.

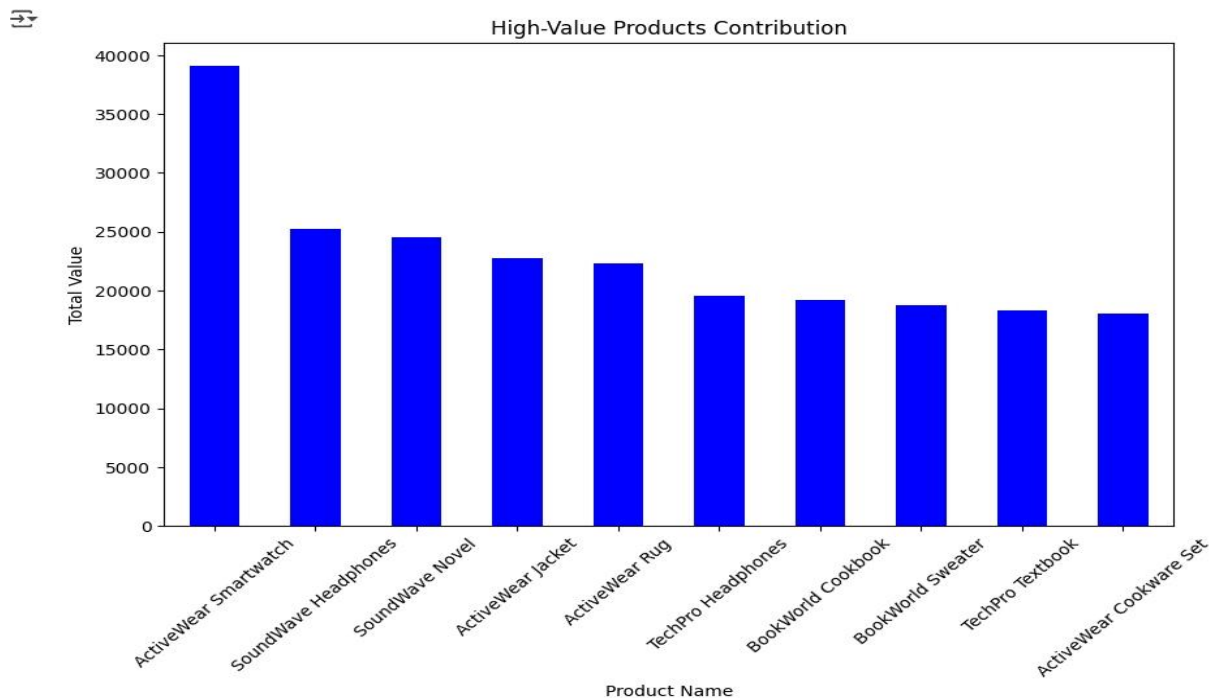


Fig: High-Value vs Low-Value Product Comparison

- Explanation: The chart shows how high-value products contribute more revenue despite selling in lower quantities.

Business Insights

- The highest number of customers come from South America.
- Electronics and Clothing are the top-selling product categories.
- Sales show a peak during the months of March and August.
- A small group of customers contribute to a large percentage of total sales.
- High-value products contribute significantly to revenue despite lower quantities sold.