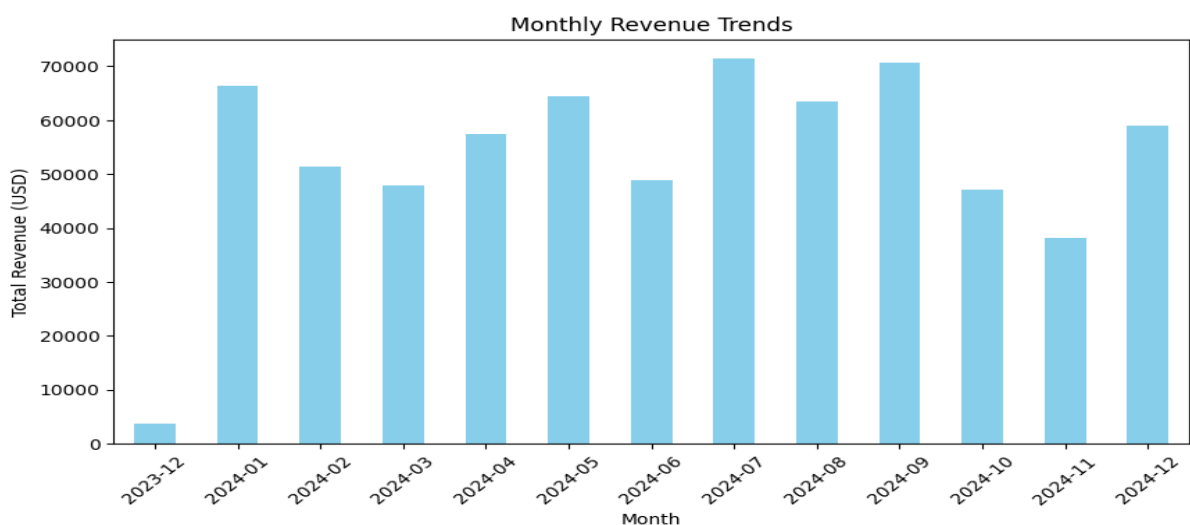


TASK – I

EDA AND BUSINESS INSIGHTS

- The code imports essential libraries (Pandas, NumPy, Matplotlib, Seaborn) for data analysis and visualization.
- It loads and previews the Customers.csv, Products.csv, and Transactions.csv datasets using `.head()` and `.info()`.
- Missing values in all datasets are identified using `.isnull().sum()`.
- The SignupDate and TransactionDate columns are converted to datetime format for proper time-based analysis.
- Datasets are merged using CustomerID and ProductID to create a unified dataset for transactions.

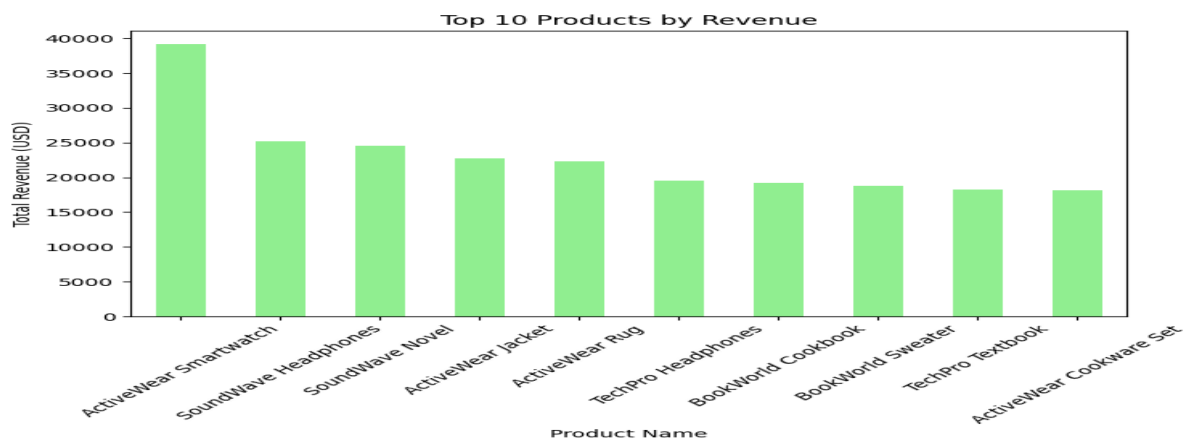
1. Monthly Revenue Trends: A bar chart plots total revenue grouped by transaction month using `.dt.to_period('M')`. It provides a clear view of revenue fluctuations over time, aiding in trend analysis. It is clearly visible that the months from July 2024 to September 2024 have the best revenue generation cycle.



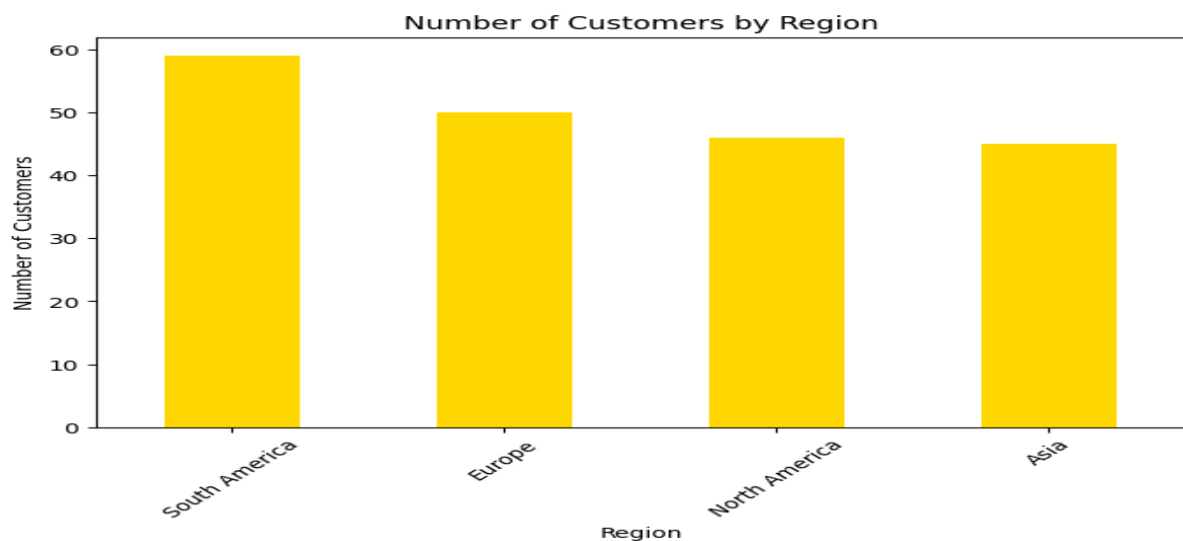
2. Revenue by Region: The code aggregates revenue by Region and visualizes it in descending order using a bar chart, helping to identify high-performing regions. It is visible that South America is the continent where the services are utilized the most with the highest amount of revenue generated. . Targeted marketing campaigns in underperforming regions could help increase revenue.



3. Top 10 Products by Revenue: Products are ranked by total revenue contribution, and the top 10 are visualized using a bar chart to highlight the most profitable products. It is clear that the customers prefer activewear products for than other products as shown by the chart. These top-performing products should be prioritized for inventory management, promotions, and cross-selling opportunities.



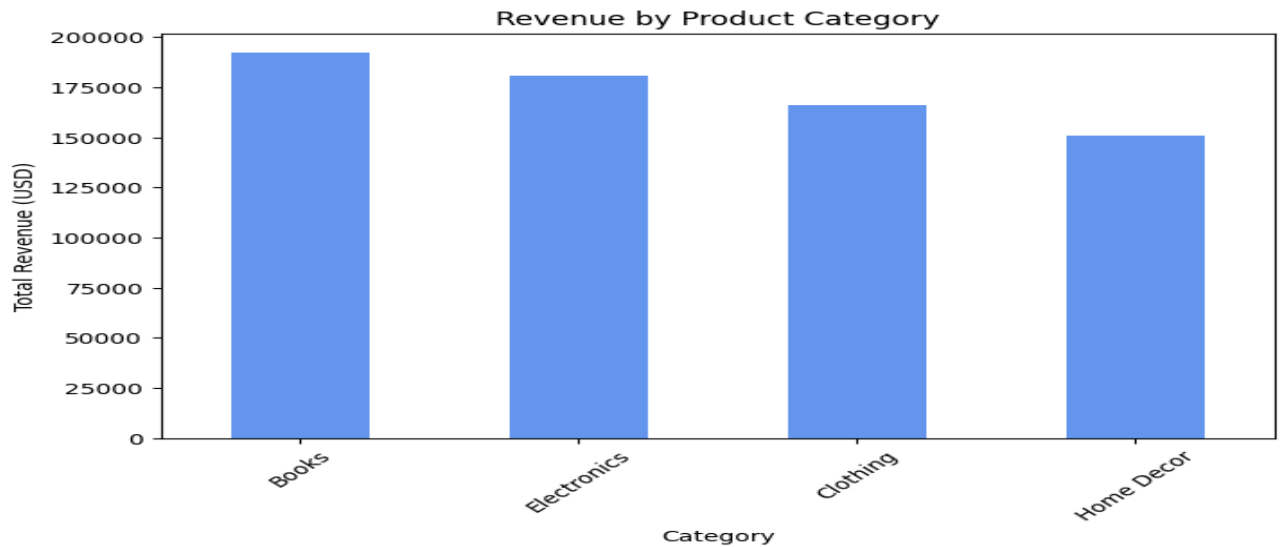
4. Customer Segmentation by Region: The Region column is analyzed to count customers per region, displayed in a bar chart for an understanding of customer distribution. It is also evident that most customers are from South America and the western hemisphere which makes it a very good place for the HQ of the company or create warehouses to reduce transportation costs.



5. Average Transaction Value: The mean value of Total Value is calculated and displayed, offering insights into typical transaction sizes across the dataset. The average transaction value is 690 dollars .

6. Revenue by Product Category: Categories are analyzed by summing revenue, and a bar chart displays their contributions, revealing top-

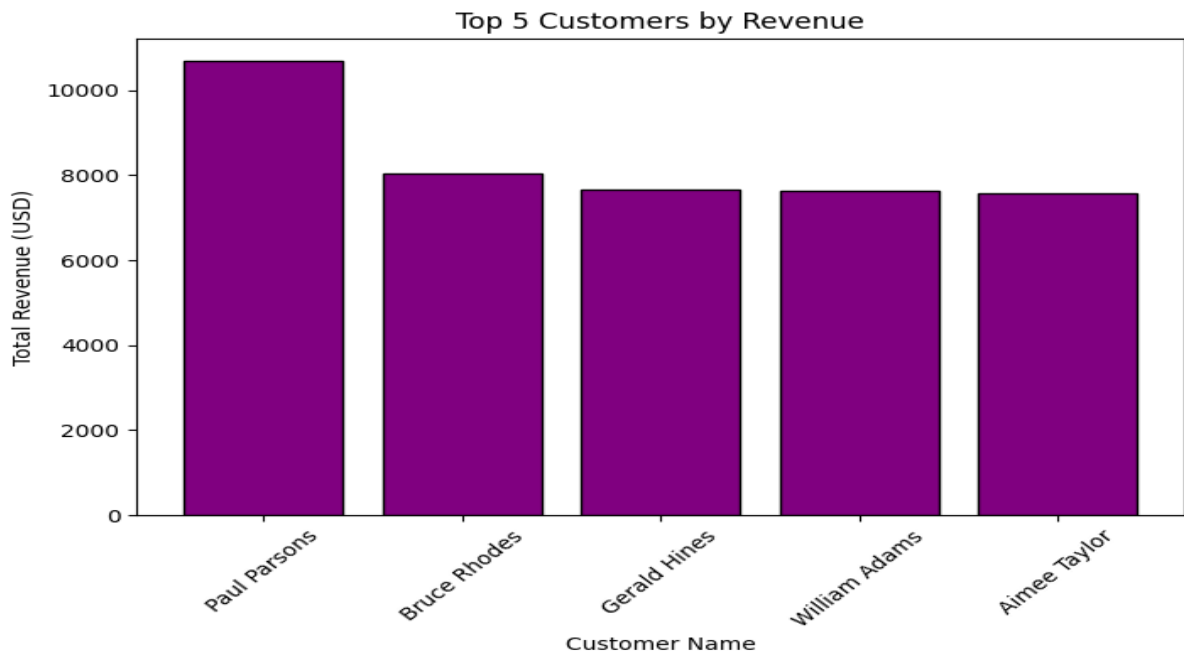
performing categories. It is clear to see that Books and Electronic gadgets products have the most revenue generated.



7. **Customer Retention Analysis:** The time between the first and last transactions is calculated for each customer as `RetentionDays`, providing insights into customer loyalty. The higher the number better the retention rate. Implementing loyalty programs or incentives for returning customers can improve retention and increase lifetime value.



8. **Most Valued Customers:** The top 5 customers account for a significant portion of the revenue. Personalized offers or exclusive perks for these high-value customers could enhance satisfaction and loyalty.



9. Data Export: The cleaned and merged dataset, crucial for future analysis, is saved as Merged_eCommerce_Data.csv.