Exploratory Data Analysis Report

Import Necessary Libraries

The following libraries are imported to facilitate data analysis and visualization:

- pandas
- NumPy
- matplotlib
- seaborn

Data Loading

Three datasets are loaded:

- Customers dataset: Contains information about customers.
- Products dataset: Contains information about products.
- Transactions dataset: Contains transaction details.

Data Merging

The datasets are merged using the following steps:

1. Merging the Transactions dataset with the Customers dataset on CustomerID.
2. Merging the resulting dataset with the Products dataset on ProductID.

Data Quality Checks

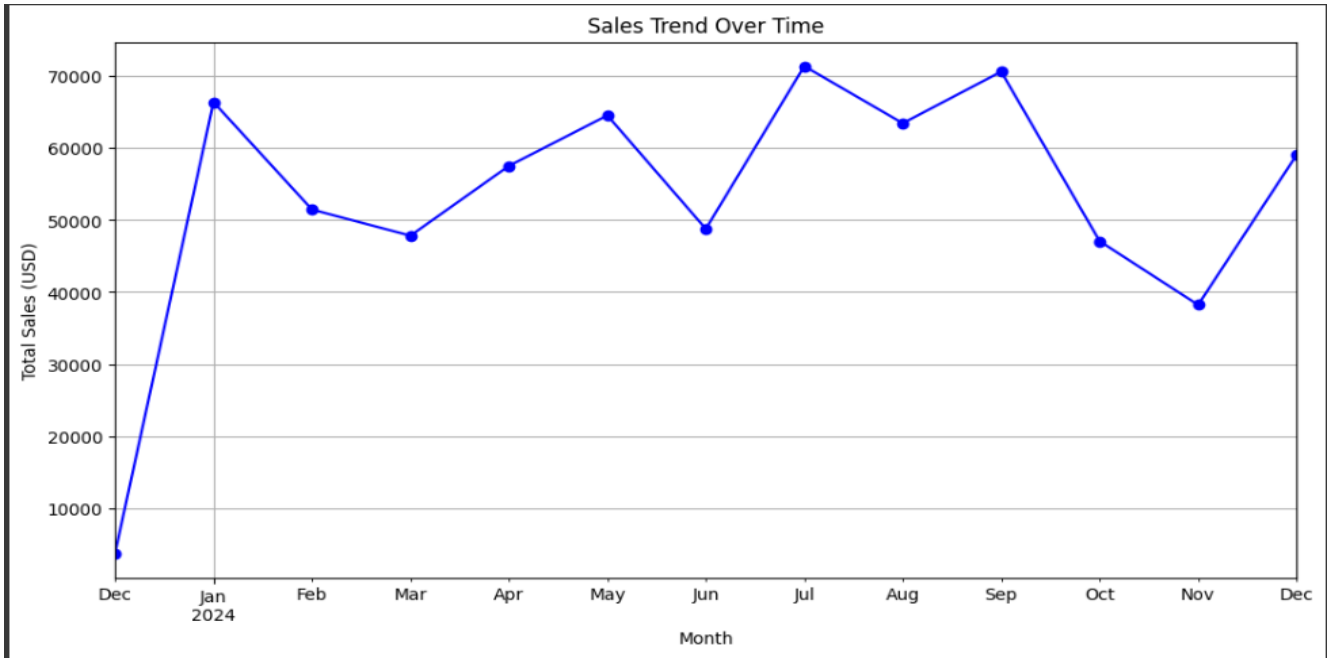
1. **Null Values:** The data is checked for missing values.
2. **Duplicate Records:** Duplicate records are identified and handled.

Analyses and Visualizations

1. Sales Trend Over Time

A line chart shows the total sales trends over time, aggregated monthly.

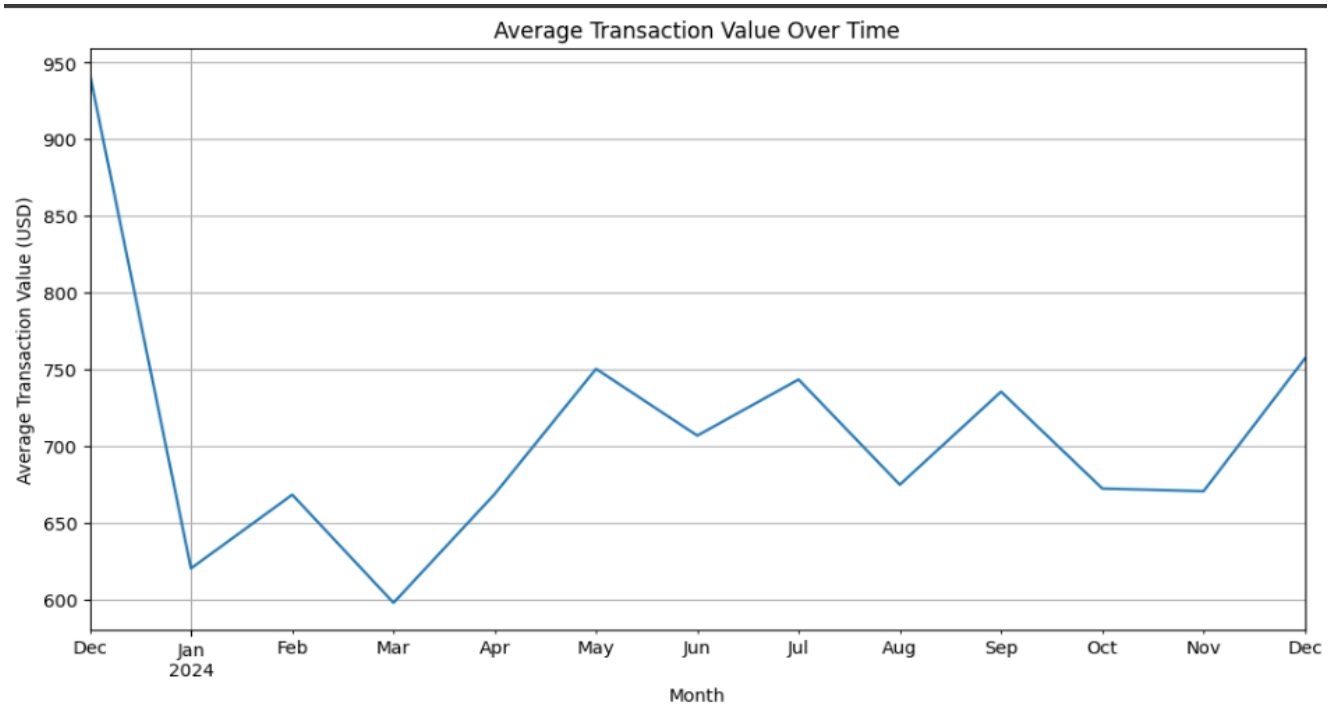
- **Insights:** Patterns and trends in sales performance.



2. Average Transaction Value Over Time

A line chart displays the average transaction values over time, aggregated monthly.

- **Insights:** Trends in customer spending behaviour.



3. Customer Distribution by Region

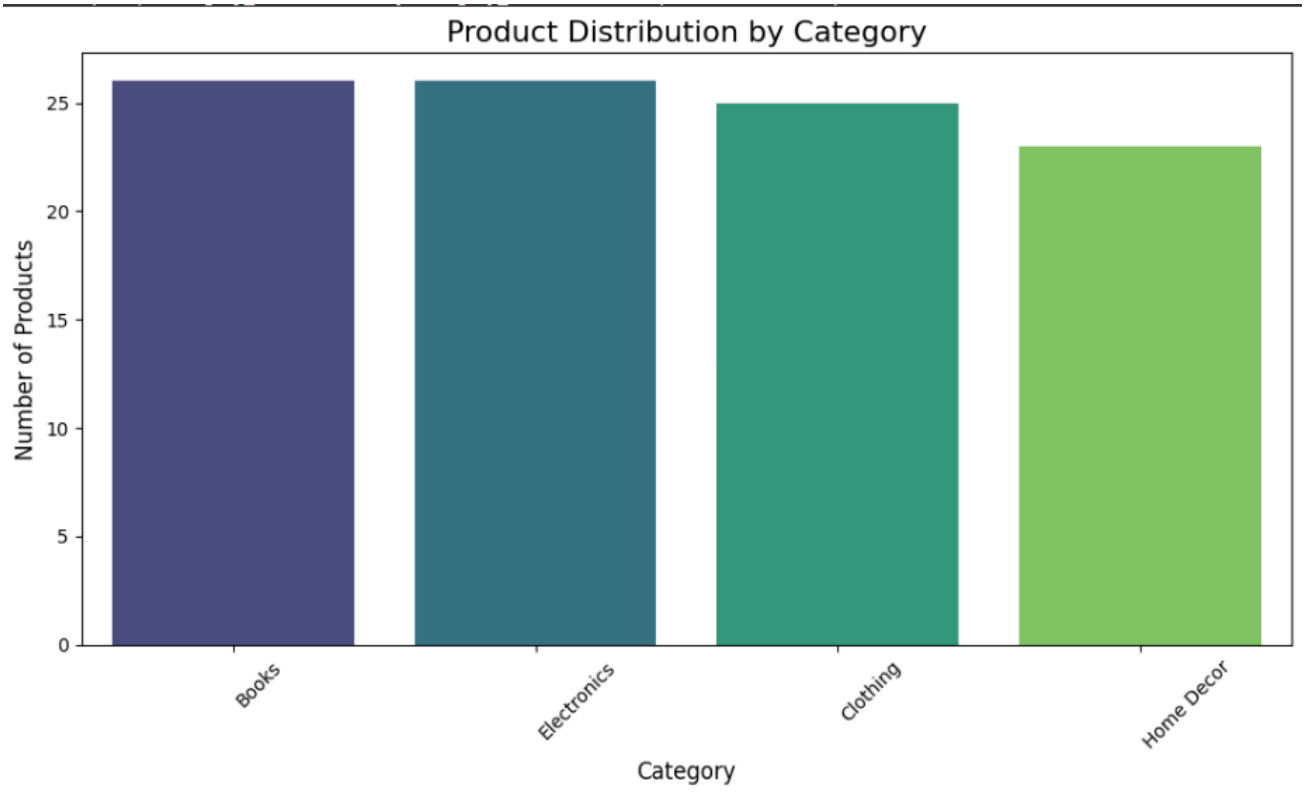
A bar chart visualizes the number of customers in each region.

- **Insights:** Customer base concentration across regions.

4. Product Distribution by Category

A bar chart shows the distribution of products across categories.

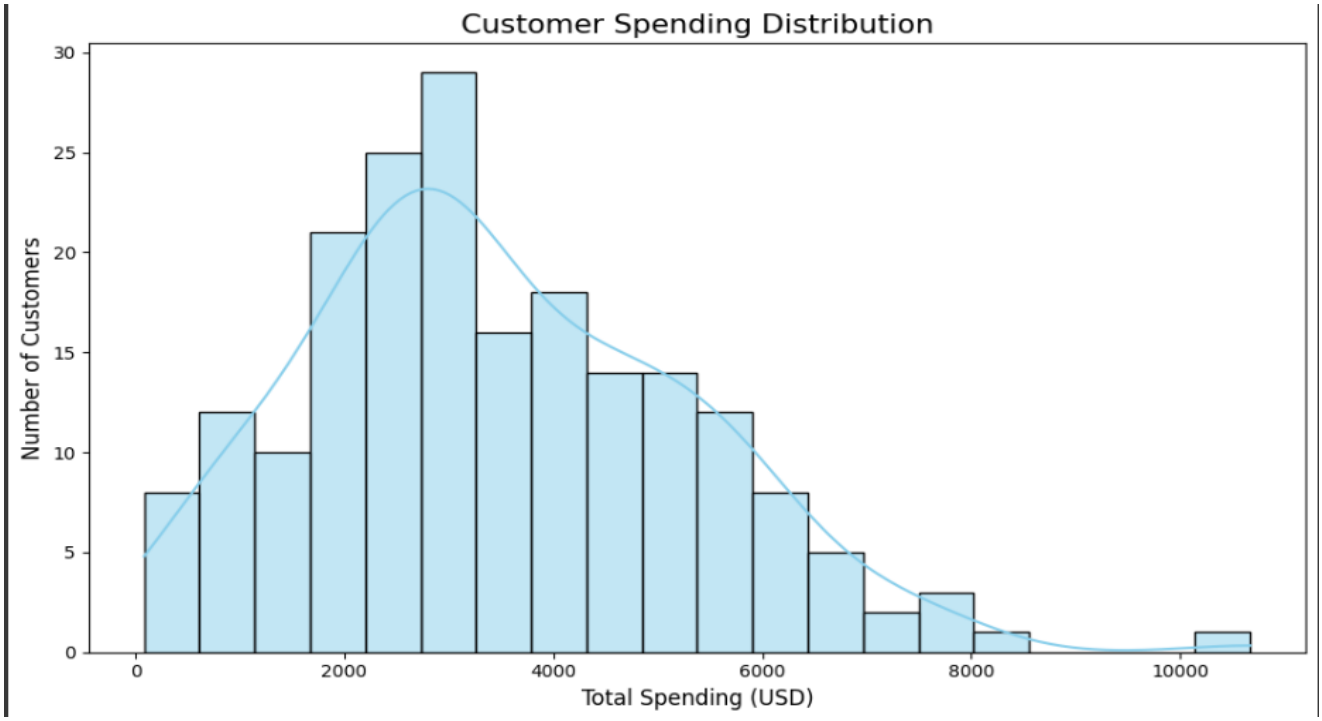
- **Insights:** Product variety and popularity.



5. Customer Spending Distribution

A histogram visualizes the distribution of total spending amounts by customers.

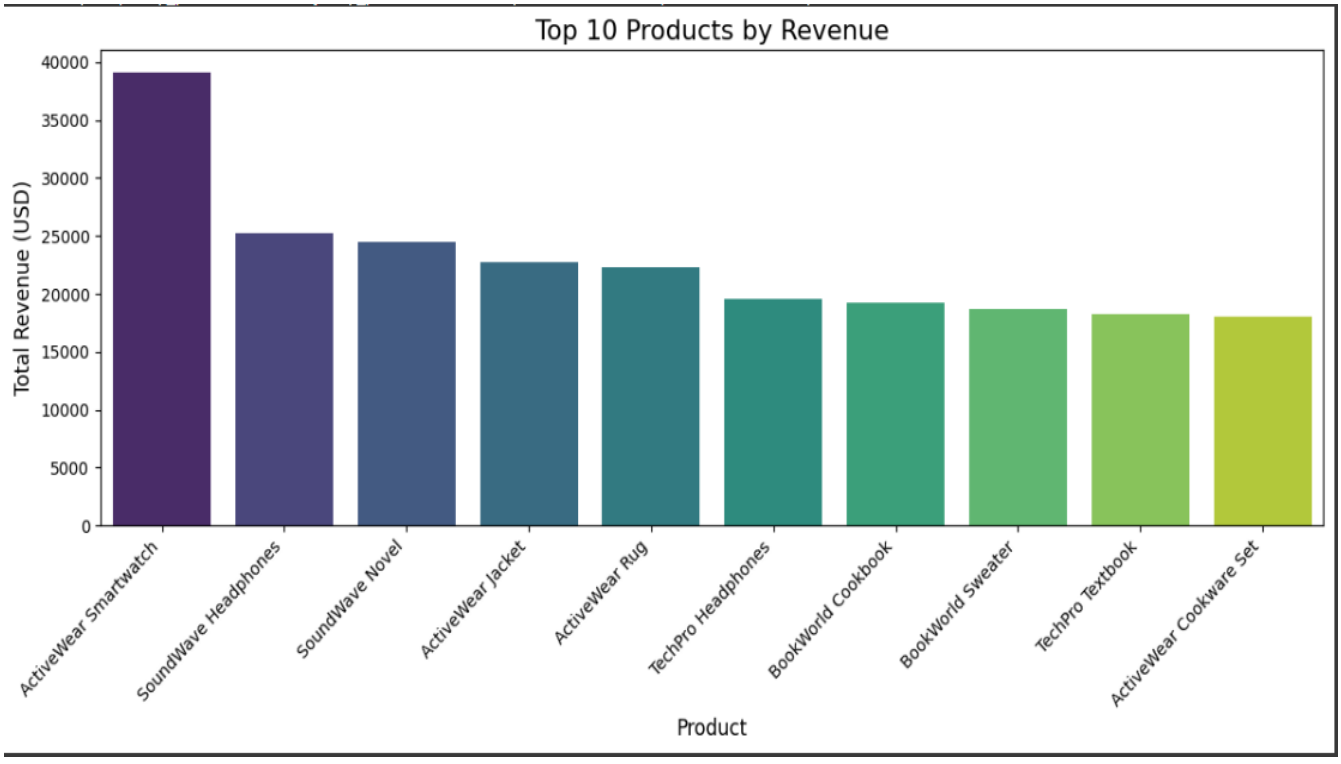
- **Insights:** Spending patterns and customer segments.



6. Top 10 Products by Revenue

A bar chart highlights the top 10 products generating the highest revenue.

- **Insights:** Identification of best-performing products.



Observations and Recommendations

Sales Trends

Observation: There are regular spikes in sales at certain periods, possibly due to holidays, festivals, or promotional campaigns.

Recommendation:

1. Investigate the exact reasons for these spikes (e.g., seasonal demand, holiday seasons, marketing campaigns).
2. Use this insight to plan targeted promotional campaigns during these high-demand periods to maximize sales.
3. Ensure adequate inventory and staffing to handle increased demand during these times.

Customer Insights

Observation: A significant number of customers are located in specific regions, indicating that these areas are core markets.

Recommendation:

1. Conduct a deeper analysis of customer demographics and preferences in these regions.
2. Tailor region-specific marketing campaigns to better address the preferences and needs of customers in these key areas.
3. Expand infrastructure or services (like delivery options) in high-potential regions to further penetrate these markets.

Product Performance

Observation: Certain product categories and specific products consistently generate higher revenue than others.

Recommendation:

1. Focus marketing efforts on promoting the top-performing products to further boost sales.
2. Maintain adequate stock of these products to avoid stockouts during peak demand periods.
3. Use insights from high-performing products to inform new product development or expansion in similar categories.

Conclusion

The analysis provides clear insights into customer behaviour, sales trends, and product performance. Future work could expand on predictive modelling and interactive dashboards for more actionable outcomes.