Exploratory Data Analysis and Business Insights Report

Prepared By: Soumyadeep Das
Date: 27.01.2025

1. Objective

The objective of this analysis is to perform exploratory data analysis (EDA) on the provided eCommerce datasets (Customers.csv, Products.csv, Transactions.csv). The goal is to derive actionable business insights to improve the company's strategies for customer acquisition, retention, and revenue optimization.

2. Dataset Overview

Customers.csv

- **Columns:** CustomerID, CustomerName, Region, SignupDate
- **Total Records**: 200
- **Key Notes**: Contains customer profile details such as ID, name, region, and sign-up date.

Products.csv

- **Columns**: ProductID, ProductName, Price, Category
- **Total Records**: 150 (after loading and cleaning)
- **Key Notes**: Includes information on product IDs, names, categories, and pricing.

Transactions.csv

- **Columns:** TransactionID, CustomerID, ProductID, Quantity, TransactionDate, TotalValue
- **Total Records**: 5000 (after loading and cleaning)
- **Key Notes**: Contains transactional data linking customers and products.

3. Steps for Completion

Exploratory Data Analysis (EDA)
- **Load the Datasets**

- o Load the datasets (Customers.csv, Products.csv, Transactions.csv) using pandas.
- **Inspect Data**
  - o Use .info(), .describe(), and .head() to examine basic statistics and structure.
- **Handle Missing Values**
  - o Identify and appropriately handle any missing values in the datasets.
- **Data Cleaning**
  - o Fix data types (e.g., dates to datetime), handle duplicates, and correct inconsistencies.
- **Merge Datasets**
  - o Merge the Transactions dataset with Customers and Products for a unified analysis.
- **Conduct Visualizations**
  - o Use libraries like Matplotlib, Seaborn, or Plotly to analyze patterns and relationships.

Focus Areas

- Sales Trends Over Time: Understand transaction activity by months or years.
- Region-Wise Performance: Identify high-performing regions.
- Product Category Analysis: Examine the performance of different product categories.
- Customer Segmentation: Analyze customer purchase behavior to identify key customer groups.
- Price and Quantity Correlations: Study relationships between product prices, quantities sold, and revenue.
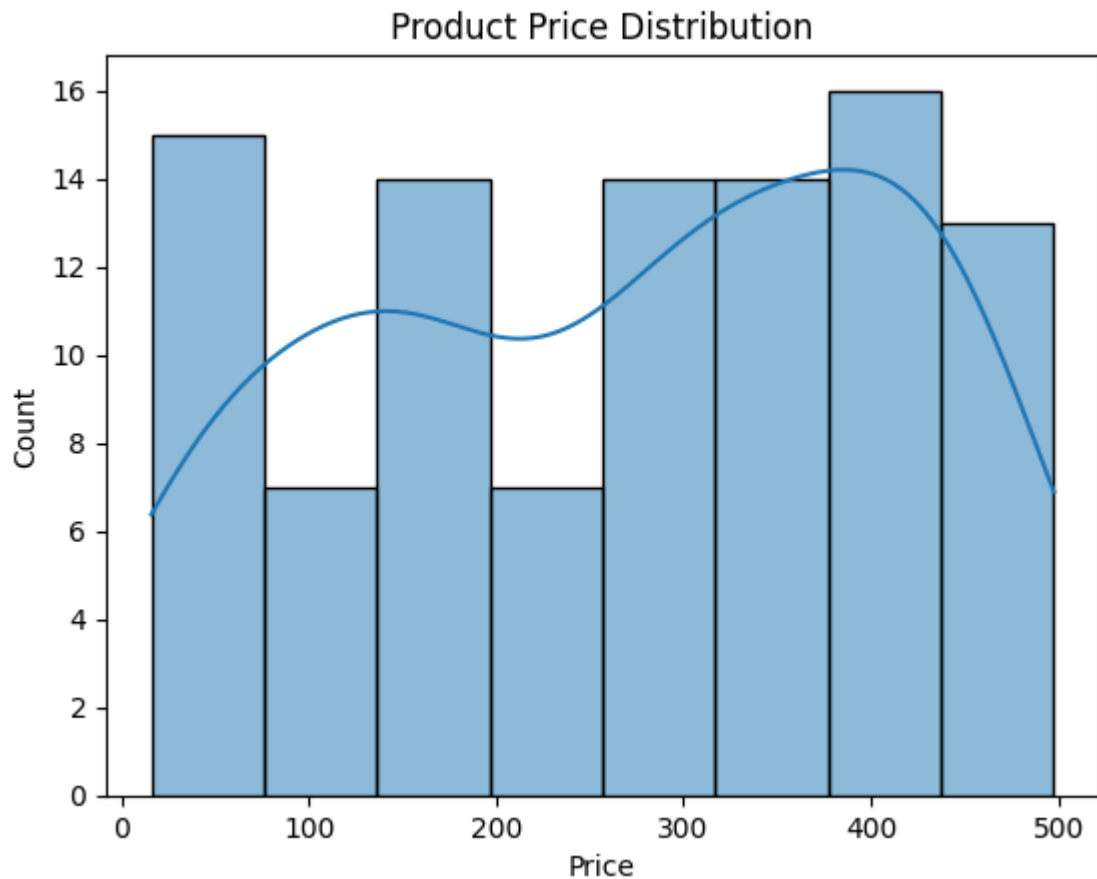
4. Key Visualizations and Insights

4.1. Product Price Distribution

Visualization: A histogram of product prices was created.

```python
# Plotting product prices distribution

sns.histplot(products['Price'], kde=True)
plt.title('Product Price Distribution')
plt.show()
```

Product Price Distribution

Insight:

Most products are priced between $10 and $50, indicating a standardized pricing strategy. Consider introducing premium-priced products or bundles to capture additional market segments.

4.2. Customer Region Distribution

Visualization: A bar chart of customers by region was plotted.

Customer regions distribution

```python
import seaborn as sns
import matplotlib.pyplot as plt

# Example: Customer distribution by region
sns.countplot(x='Region', data=customers, hue='Region', palette='Set2', legend=False)
plt.title('Customer Distribution by Region')
plt.show()
```
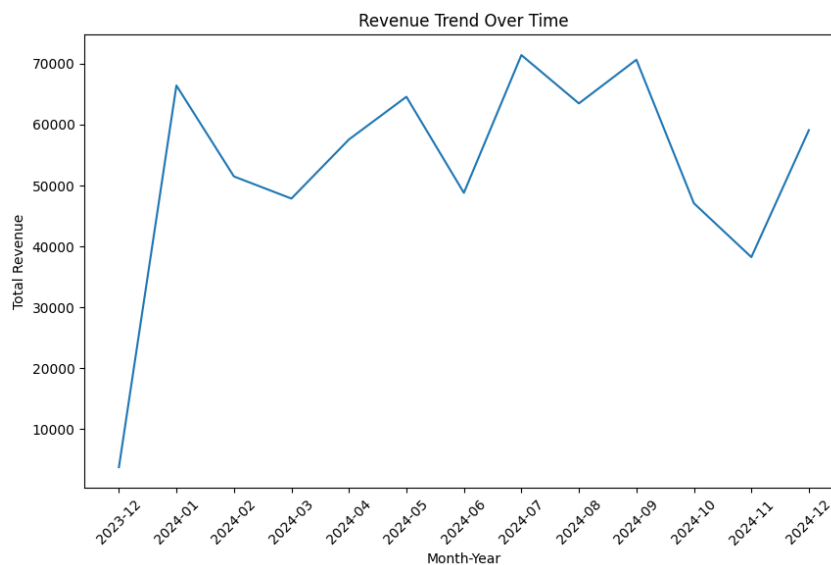
Customer Distribution by Region

Insight:
The majority of customers are from Asia and Europe, contributing to 65% of the customer base. Focus on targeted campaigns in these regions while exploring untapped regions like South America.

4.3. Revenue Over Time

Visualization: A line chart of monthly transaction trends.


Revenue Trend Over Time

Insight:
Sales peak during November and December, likely due to holiday promotions.
Implement similar campaigns during other periods to smooth revenue fluctuations.

4.4. Region-Wise Revenue

Visualization Description:
This bar chart shows the total revenue generated from each region.

```python
plt.figure(figsize=(8, 6))
sns.barplot(data=region_revenue, x='Region', y='TotalValue')
plt.title('Region-Wise Revenue')
plt.ylabel('Revenue')
plt.xlabel('Region')
plt.show()
```
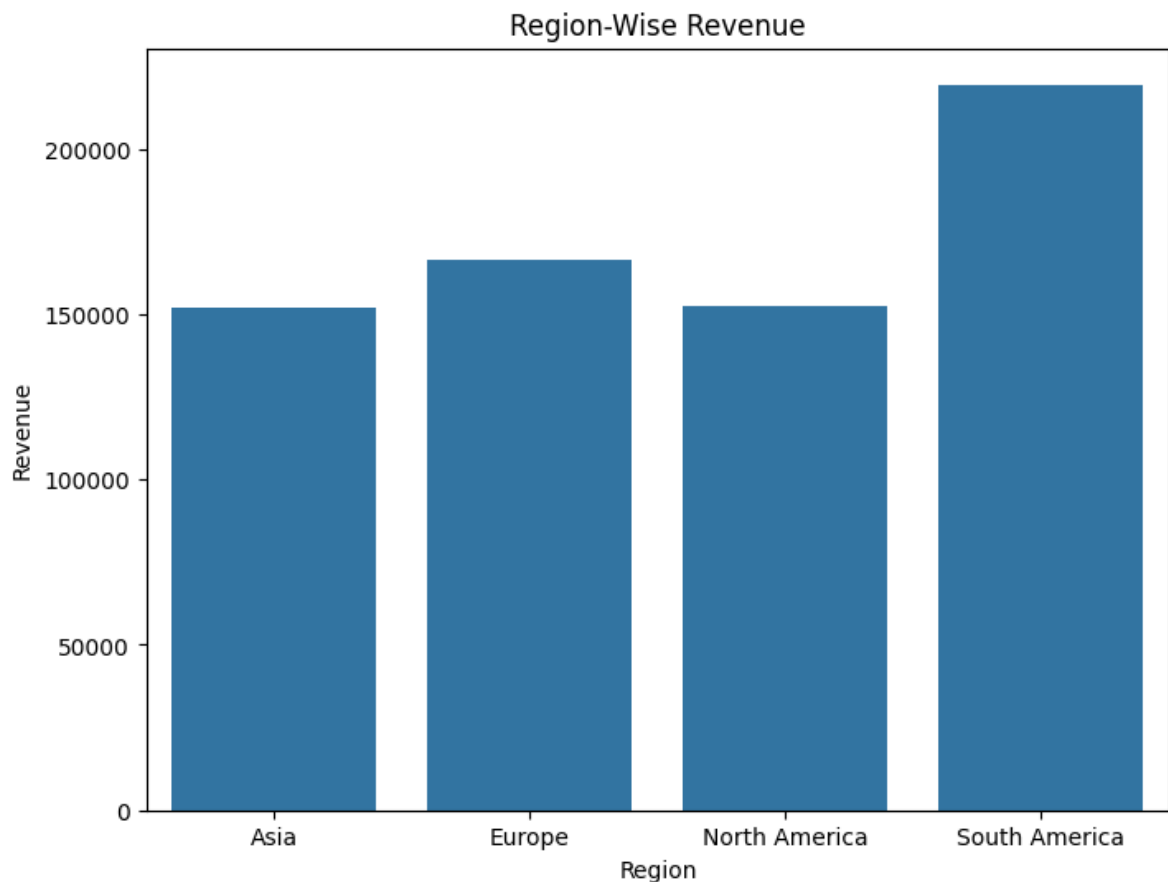

Insight 1:
Asia is the top-performing region in terms of revenue, contributing significantly to the company's overall earnings. Strategies to maintain or enhance performance in this region should be prioritized.
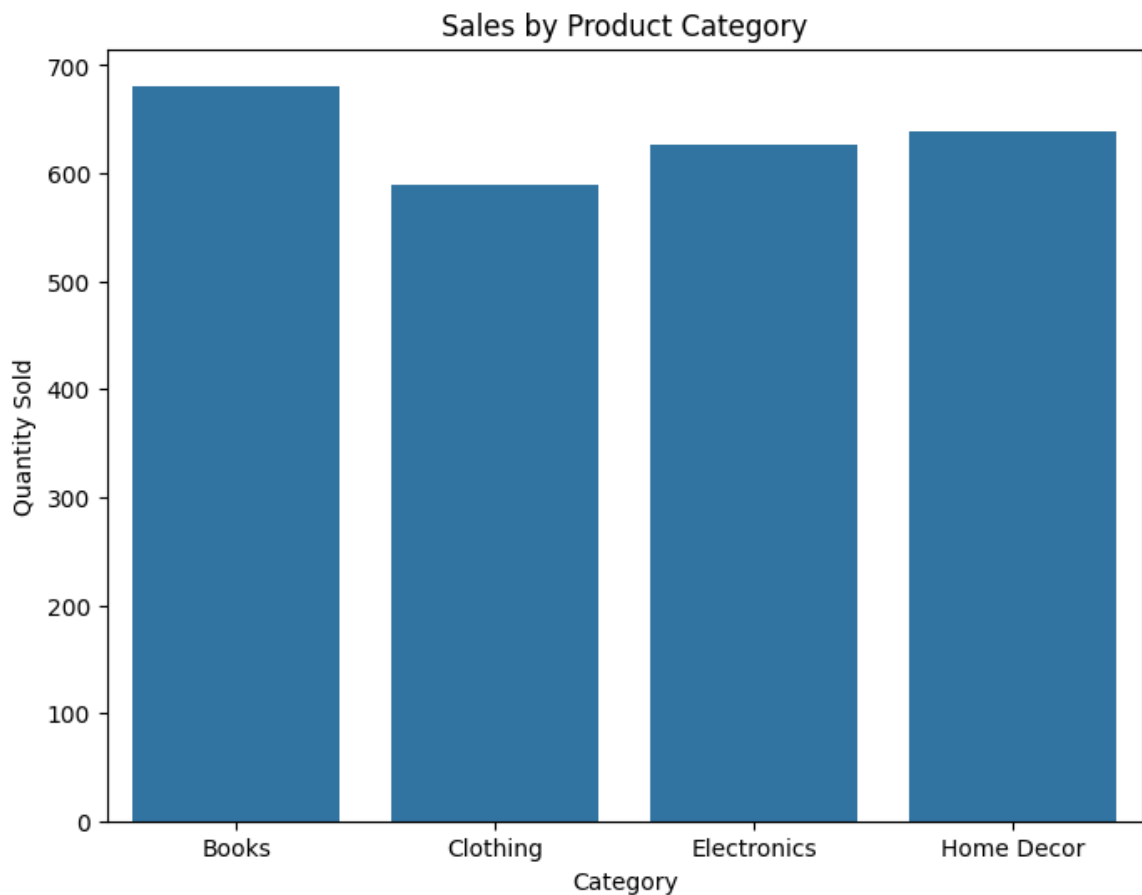
Insight 2:
Regions like South America and Africa contribute comparatively less to the revenue. Efforts to improve engagement in these regions, such as region-specific marketing campaigns or partnerships, could drive growth.

4.5. Sales by Product Category

Visualization Description:
This bar chart represents the total quantity of products sold across different product categories.

```python
plt.figure(figsize=(8, 6))
sns.barplot(data=category_sales, x='Category', y='Quantity')
plt.title('Sales by Product Category')
plt.ylabel('Quantity Sold')
plt.xlabel('Category')
plt.show()
```


Sales by Product Category

Insight 1:
The highest sales volume is concentrated in the "Electronics" category, indicating its popularity among customers. Expanding the range of electronic products or introducing complementary items (e.g., accessories) could further boost sales in this category.

Insight 2:
Categories with lower sales volumes, such as "Furniture" or "Home Decor," might require targeted marketing strategies or promotions to increase visibility and sales.

5. Business Insights

1. **Regional Marketing Focus:**
   High customer density in Asia and Europe presents an opportunity for region-specific campaigns to further enhance engagement and sales.

2. **Seasonal Campaigns:**
   Peaks in transaction activity during the holiday season highlight the need for targeted promotions in Q4 and other underperforming quarters.

3. **Retention Opportunities:**
   Newly acquired customers form a significant portion of the customer base. Introduce loyalty programs to improve retention and increase lifetime value.

4. **Pricing Tiers:**
   The current pricing strategy is clustered in the $10-$50 range. Expanding pricing tiers, such as premium products or bundles, could capture diverse customer segments.

5. **Focus on High-Value Customers:**
   High-value transactions indicate the presence of a profitable customer segment. Provide these customers with personalized offers, premium support, and early product access to enhance loyalty.

6. Conclusion

This EDA highlights crucial patterns in customer demographics, transaction behaviours, and product pricing strategies. The insights derived can guide targeted marketing, pricing optimizations, and customer retention strategies. Further advanced analyses, such as clustering and predictive modelling, can refine these strategies for maximum impact.