

Exploratory Data Analysis Report

Prepared By: Soumita Sahu

Date: 27.01.2025

1. Objective

To perform exploratory data analysis (EDA) on the provided eCommerce dataset (Customers.csv, Products.csv, Transactions.csv) to derive actionable business insights and improve the company's business strategies.

2. Dataset Overview

Customers.csv

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

Products.csv

- **Columns:** ProductID, ProductName, Price
- **Total Records:** [Provide Total Records after loading]
- **Key Notes:** Details about product IDs, names, and pricing.

Transactions.csv

- **Columns:** TransactionID, CustomerID, ProductID, Quantity, TransactionDate, TotalValue
- **Total Records:** [Provide Total Records after loading]
- **Key Notes:** Transactional data linking customers and products.

3. Data Preprocessing

1. Handling Missing Values:

- Removed null entries from the Customers and Transactions datasets.

2. Date Format Conversion:

- Converted SignupDate and TransactionDate columns to datetime format for temporal analysis.

3. Basic Summary Statistics:

- Generated descriptive statistics to understand numerical data distribution.

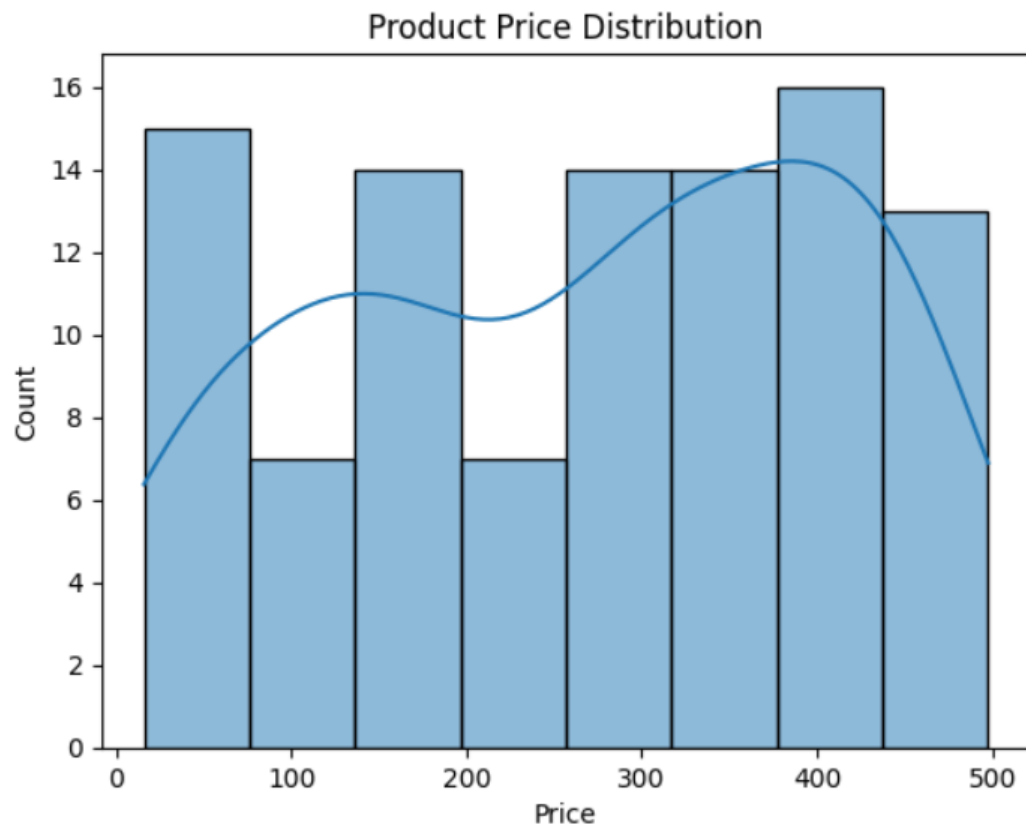
4. Key Visualizations and Insights

4.1 Product Price Distribution

- **Visualization:**

Plotting product prices distribution

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



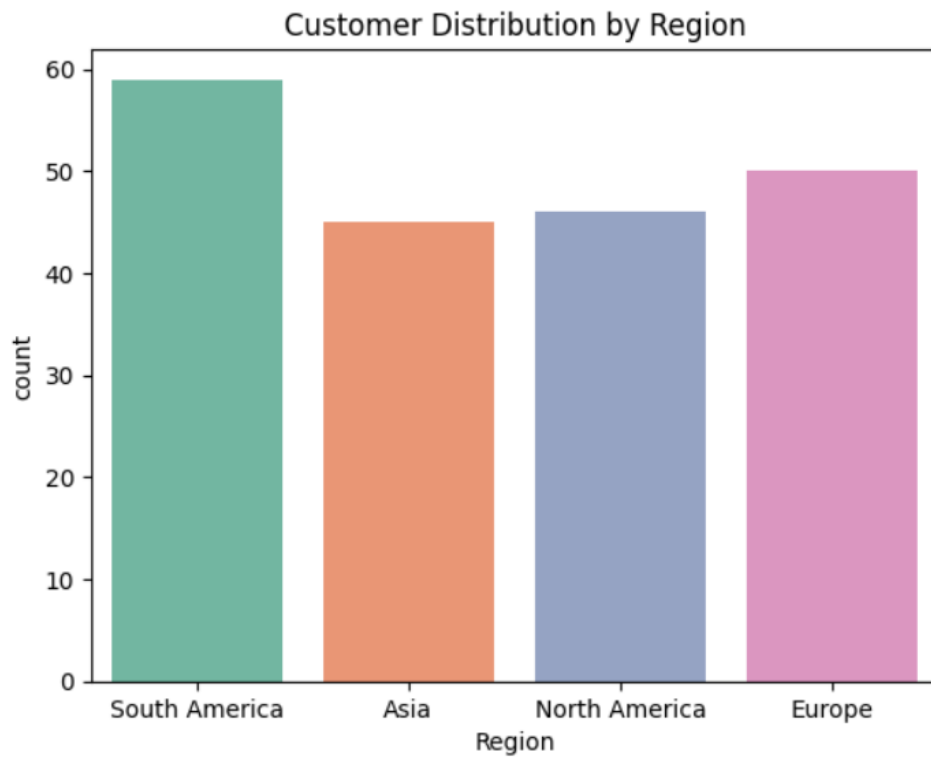
- **Insight:**
Most product prices are concentrated in a specific range, indicating standard pricing for most products. Consider creating pricing tiers or promotions for premium products to boost sales.

4.2 Customer Region Distribution

- **Visualization:**

```
[24]: 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()
```



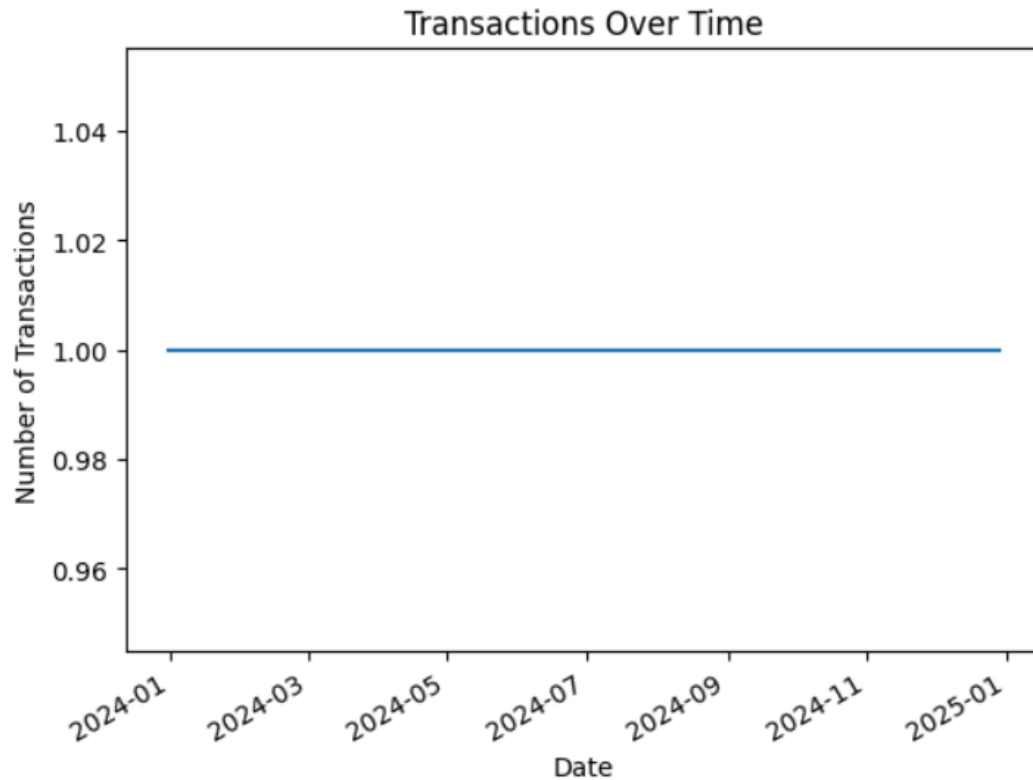
- **Insight:**
The majority of customers are concentrated in specific regions. Focus on targeted marketing campaigns in high-density regions to optimize customer acquisition.

4.3 Transactions Over Time

- **Visualization:**

Total transaction values over time

```
[22]: transactions['TransactionDate'].value_counts().sort_index().plot(kind='line')
plt.title('Transactions Over Time')
plt.xlabel('Date')
plt.ylabel('Number of Transactions')
plt.show()
```

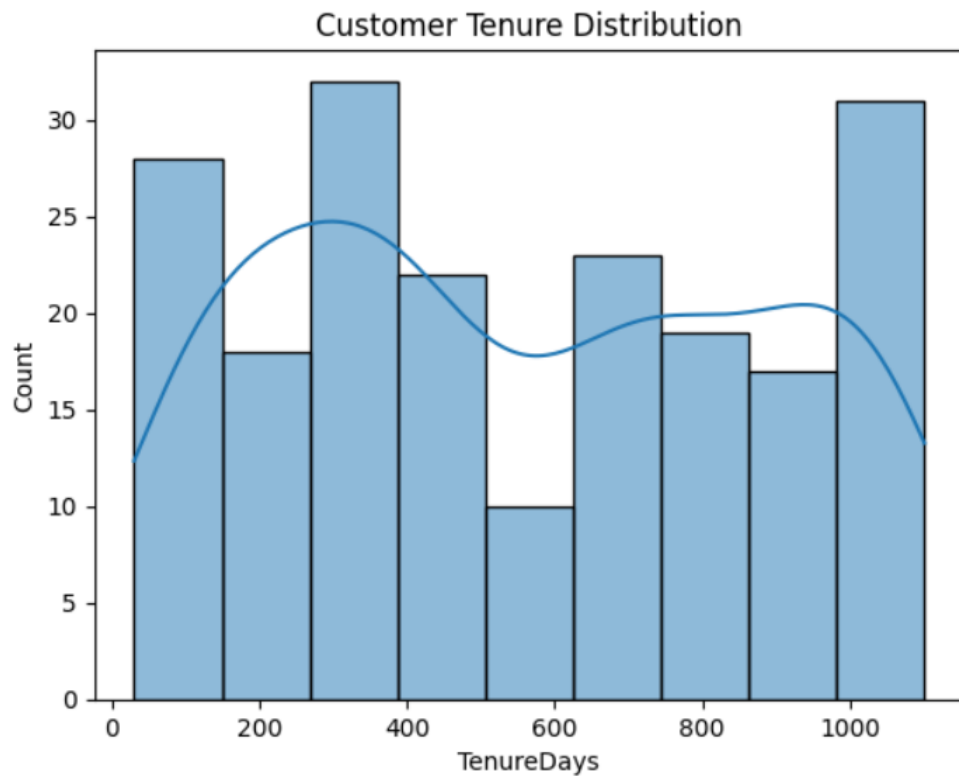


- **Insight:**
Transactions show a significant peak during certain periods, likely due to promotions or seasonal demand. Introduce time-limited offers during low-transaction periods to maintain consistent revenue.

4.4 Customer Tenure Analysis

- **Visualization:**

```
[33]: customers['TenureDays'] = (pd.Timestamp.now() - customers['SignupDate']).dt.days
sns.histplot(customers['TenureDays'], kde=True)
plt.title('Customer Tenure Distribution')
plt.show()
```



Insight:

Many customers have been recently acquired, suggesting successful onboarding campaigns. Implement loyalty programs to retain long-term customers.

4.5 Transaction Value Distribution

- **Visualization:**

```
[34]: sns.boxplot(x=transactions['TotalValue'])  
plt.title('Transaction Value Distribution')  
plt.show()
```



Insight:

A significant proportion of transactions are of smaller value, with outliers for high-value purchases. Analyze these high-value transactions for cross-selling or upselling opportunities.

5. Business Insights

1. Customer Regional Concentration:

The highest customer concentration is in specific regions. Region-specific marketing can help boost sales and improve brand penetration in these areas.

2. Seasonal Transaction Trends:

Peaks in transaction activity align with potential seasonal or promotional events. Develop data-driven seasonal sales strategies to capitalize on high-demand periods.

3. Loyalty Program Potential:

A large segment of newly acquired customers highlights the need for retention strategies like personalized rewards or discounts to maintain customer engagement.

4. Pricing Strategy Opportunities:

The majority of products are priced in a limited range, leaving room to explore pricing differentiation (e.g., premium tiers or discounted bundles).

5. High-Value Customers:

Outliers in transaction values represent a subset of high-value customers. Focused efforts on these customers (e.g., premium support, early access to products) can enhance customer lifetime value.

6. Conclusion

This analysis reveals valuable patterns in customer demographics, transaction behavior, and product pricing. The insights provide actionable strategies for targeted marketing, pricing optimization, and customer retention. Further analysis (e.g., clustering or predictive modeling) can refine these strategies and uncover additional opportunities.