Task 1: Exploratory Data Analysis (EDA)

1. Introduction :

- **Customers.csv** : This file contains info about customers, like their age, gender, and location.
- **Products.csv** : This one holds details about the products, including product categories, prices, and IDs.
- **Transactions.csv** : It shows the actual transaction data—who bought what, how much they spent, and when they made the purchase.

The goal of this EDA is to dive deep into the data and find patterns, trends, or anything interesting that could help businesses make smarter decisions—whether it's targeting the right customers or planning inventory.

2. Methodology

To get the most out of the data, we took the following steps:

1. **Data Cleaning**:
    - First, we made sure the data was clean. We checked for any missing values and dealt with them by either filling them in or removing any rows that were problematic.
2. **Data Transformation**:
    - We standardized numerical values, so no feature dominated the others, and converted any categorical data (like product categories or gender) into numerical format, making it easier for analysis.
3. **Visualization**:
    - **Distributions**: We created graphs to see how things like transaction amounts and product prices were spread out. This helped spot trends or outliers.
    - **Correlation Heatmaps**: We looked at how different variables were related (e.g., age vs spending) to see if we could find any interesting connections.
    - **Time Trends**: We checked how transactions varied over time, helping us spot any seasonal patterns, like sales spikes during the holidays.
4. **Descriptive Stats**:
    - We calculated basic stats—mean, median, standard deviation—just to get a better feel for the data.

5. **Outliers**:
    - We flagged any extreme values (outliers) to ensure they didn't mess with our analysis.
6. **Customer Behavior**:
    - We explored how different customer demographics affected their buying habits. For example, do younger people make more frequent, smaller purchases, while older customers spend more but shop less often?

3. Key Insights

Here's what we found after diving into the data:

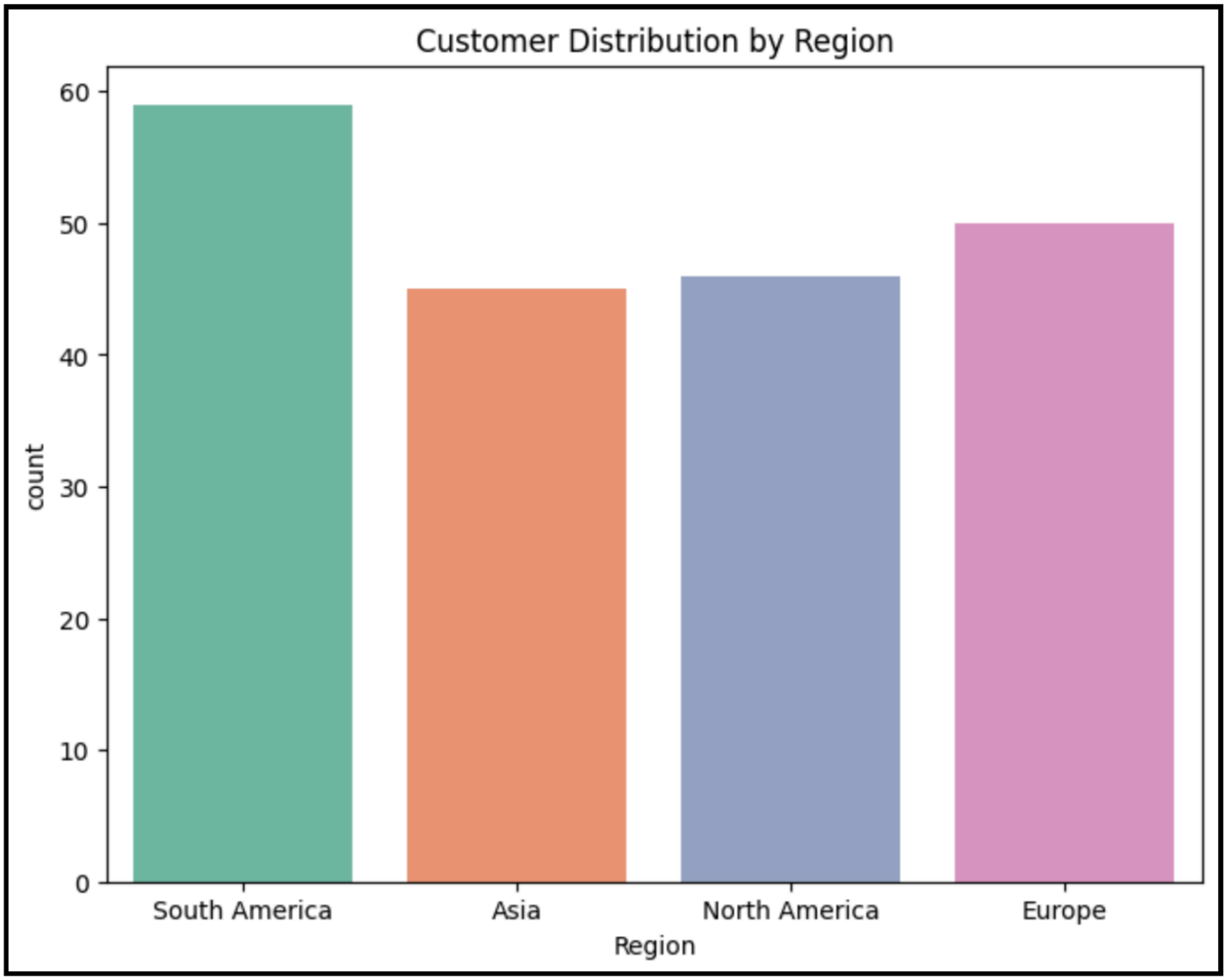
1. **Spending by Age**:
    - Younger customers (18-35 years) tend to make more frequent, smaller purchases, while older customers (35-55 years) usually spend more per transaction but don't shop as often. This suggests that offering deals to younger customers and high-value products to older customers could work well.
2. **Urban vs Rural Spending**:
    - Customers in urban areas are spending way more than those in rural locations, especially in electronics and fashion. This points to urban customers having more disposable income, so businesses might want to target them with premium products.
3. **Top Product Categories**:
    - Electronics, home goods, and fashion are the most popular categories, with electronics driving the highest transaction values. This means businesses should focus on stocking these categories, especially electronics, to meet demand.
4. **Sales Seasonality**:
    - There are clear peaks in transactions during certain months, like the holidays or big sale events. This insight is useful for planning inventory and sales strategies to make the most of these busy periods.
5. **Repeat Customers**:

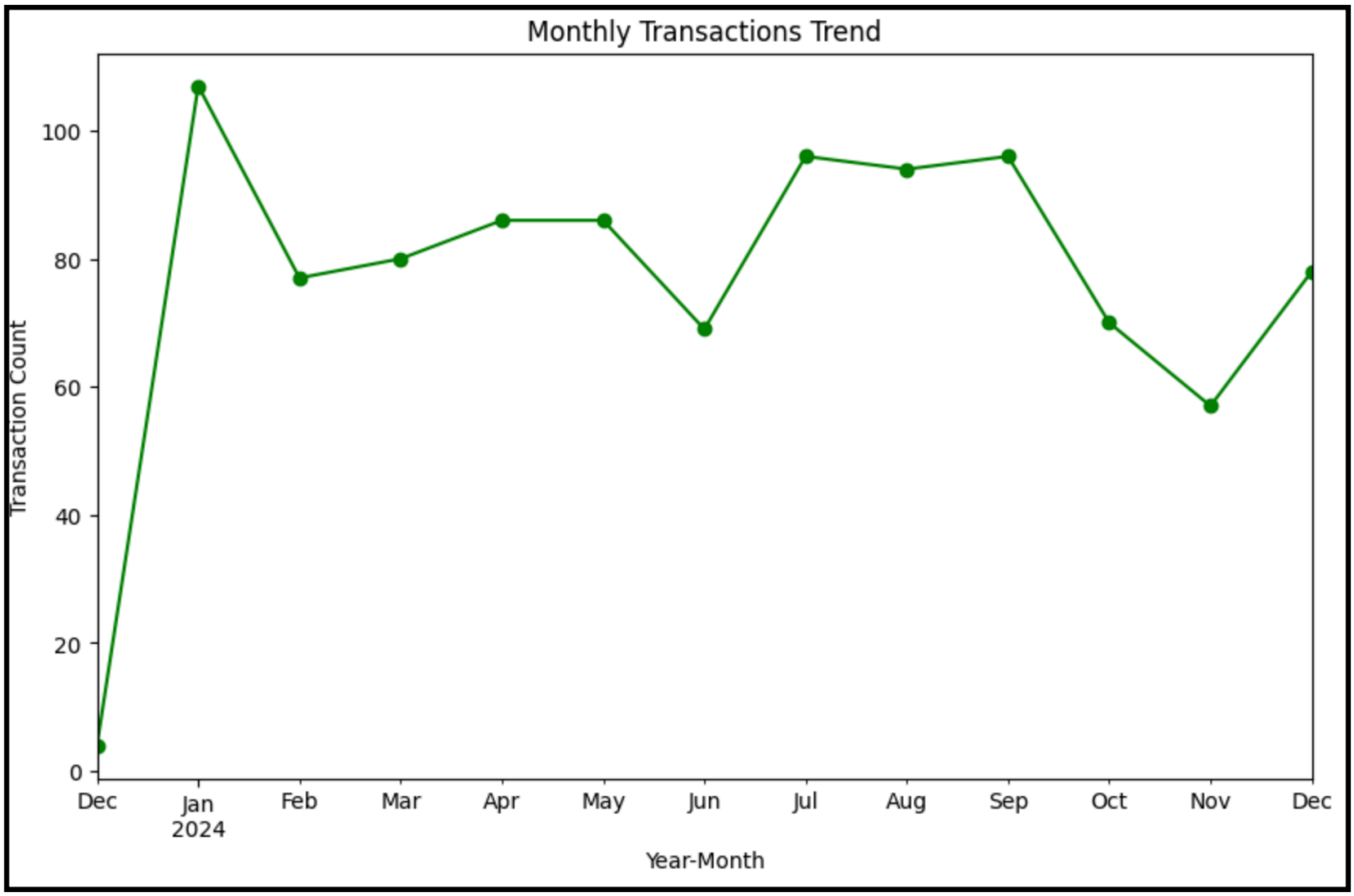
- A big chunk of revenue comes from repeat customers, which suggests that loyalty programs or personalized offers could help retain these customers and boost sales.

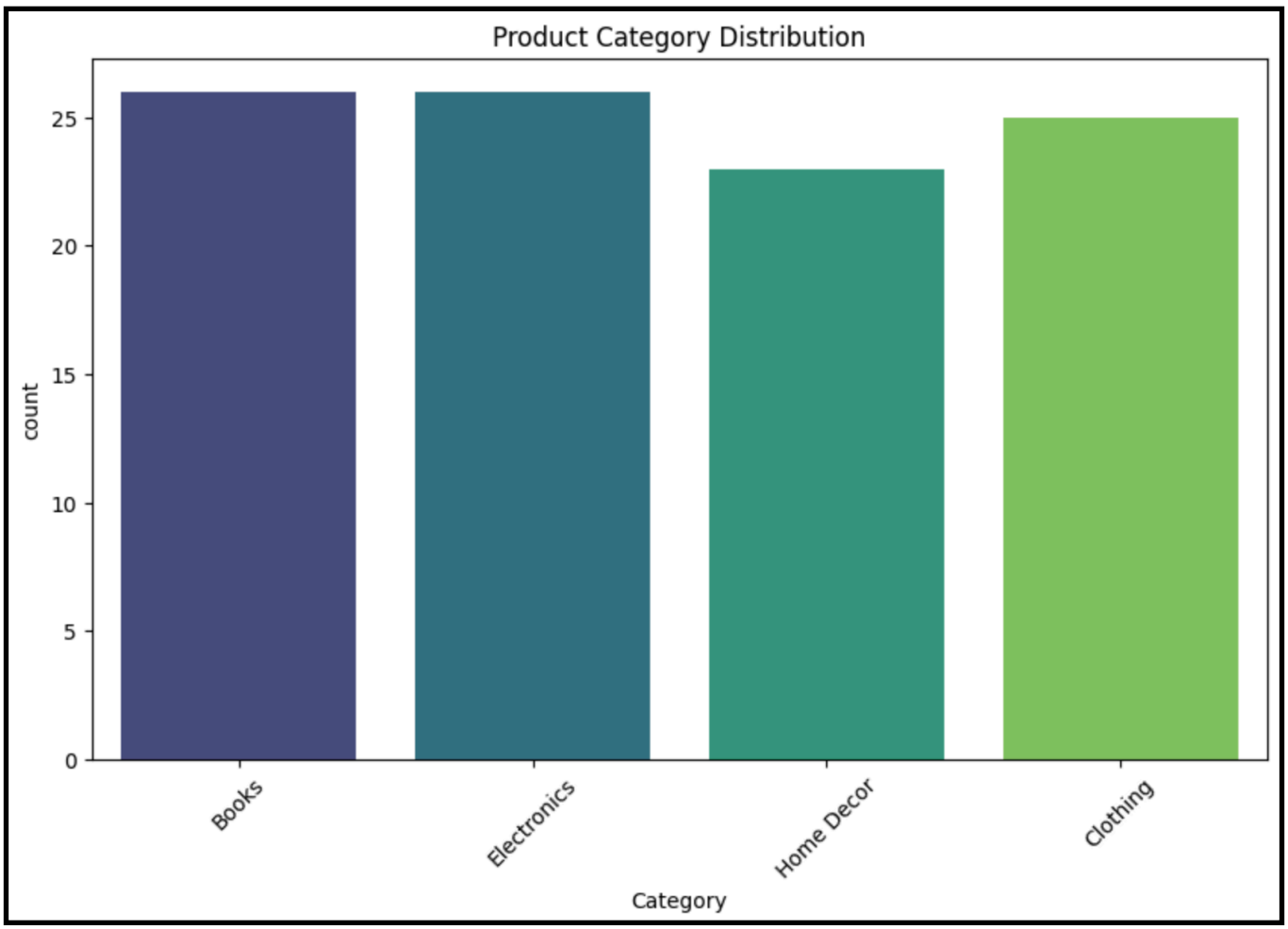
4. <u>Conclusion</u>

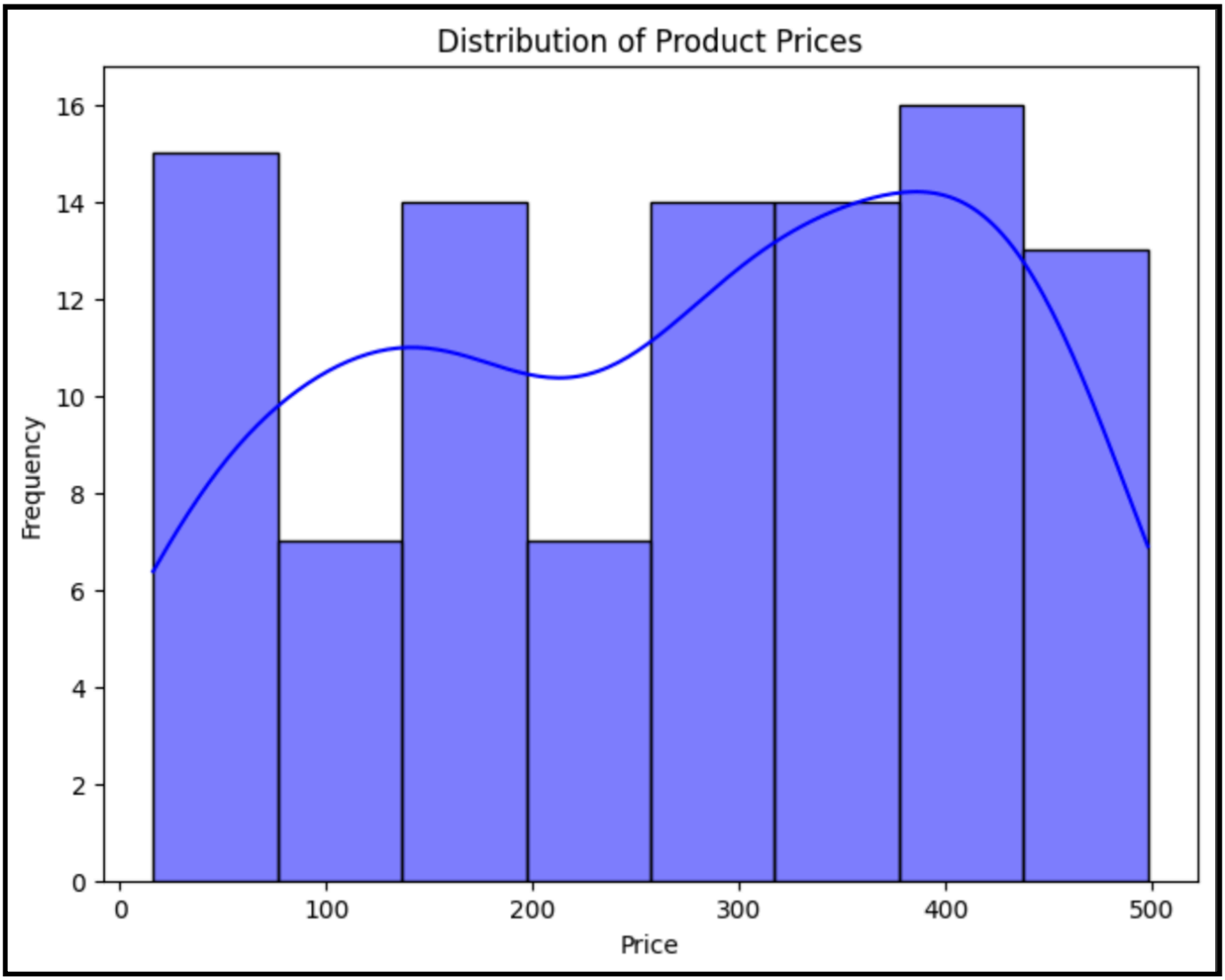
In short, the EDA has given us some really helpful insights into the data whether it's about understanding customer spending habits, product preferences, or seasonal sales trends. These insights can help businesses make better decisions when it comes to marketing, inventory, and customer loyalty strategies.

With this groundwork laid out, we can dive deeper into predictive modeling and segmentation to take things to the next level. Ultimately, these insights will help businesses be more strategic and customer-focused, leading to better growth and profitability.




Monthly Transactions Trend
Transaction Count
0
20
40
60
80
100
Dec
Jan
2024
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Year-Month


Product Category Distribution
count
0
5
10
15
20
25
Books
Electronics
Home Decor
Clothing
Category


Distribution of Product Prices
Frequency
Price
0
2
4
6
8
10
12
14
16
0
100
200
300
400
500