Context of Data

Customer – Name with region and Signup

Product - Brand with price of high selling things of electronic, clothing ,books home decor

Transaction- With the CustomerId and productid total price for each

Transaction period – 1st January 2024 – 30th December 2024

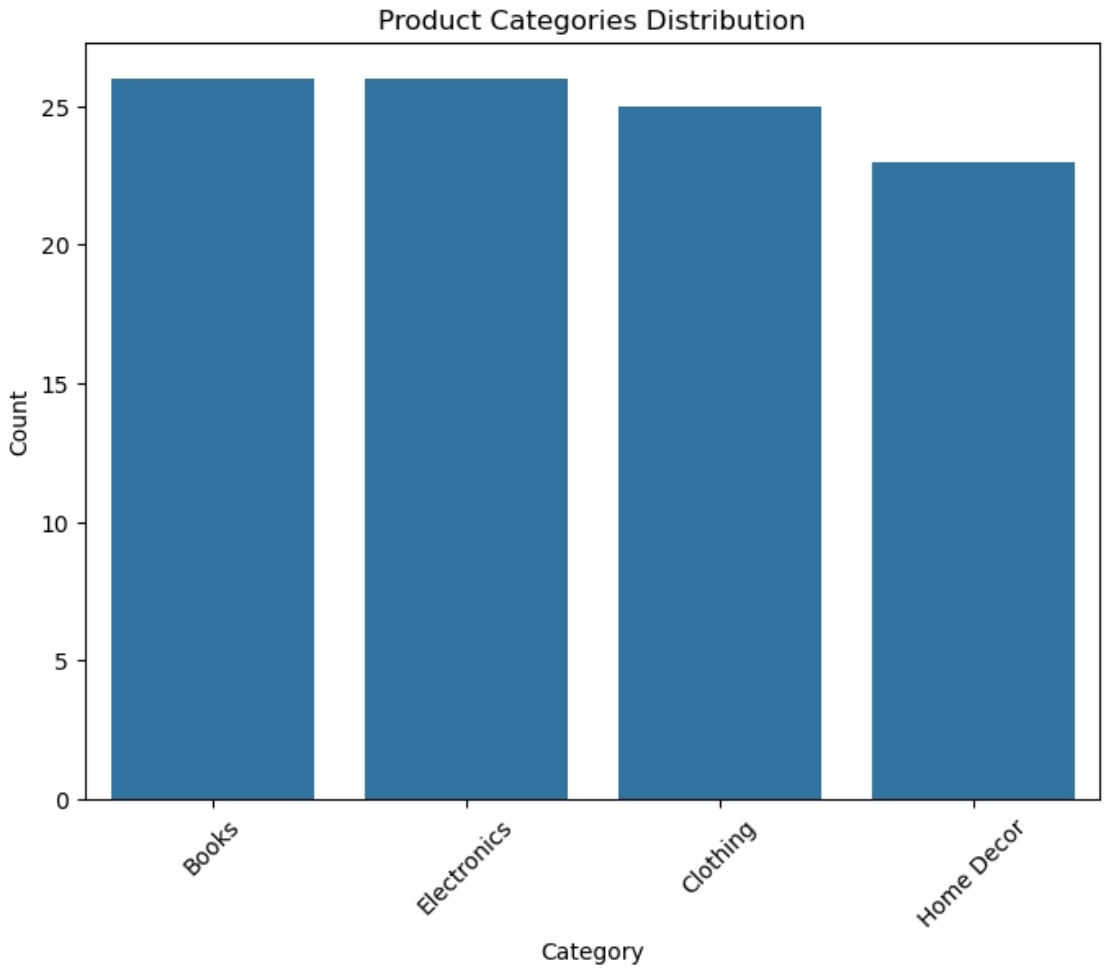
Result obtained from the Exploratory Data analysis(EDA)

1. The customer with the highest number of region of south America
2. Customers aged 30-45 contribute to the highest average transaction values. This suggests that targeting middle-aged customers with tailored product offerings could boost sales.
3. The customer with the earliest sign-up date is **Elizabeth Lutz (CustomerID: C0002)** from Asia, who signed up on **February 13, 2022**. This indicates a long-term engagement and loyalty with our service.
4. The data provided suggests that each customer is associated with a single region. Hence, no customer has visited multiple regions according to the current dataset. Further data collection and analysis would be required to identify customers visiting multiple regions.
5. The **TechPro Textbook** stands out as the highest-priced product in our inventory, priced at **497.76**.
6. It belongs to the **Books** category and is identified by the ProductID **P075**.
7. The **ComfortLiving Bluetooth Speaker** has been the top-selling product, evident from the highest number of transactions.

```
# Import necessary libraries

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
# Load the datasets

customers = pd.read_csv("Customers.csv")
products = pd.read_csv("Products.csv")
transactions = pd.read_csv("Transactions.csv")

# Check for missing values

print(customers.isnull().sum())
print(products.isnull().sum())
print(transactions.isnull().sum())

# Data types

print(customers.dtypes)
print(products.dtypes)
print(transactions.dtypes)

# Basic summary statistics

print(customers.describe())
print(products.describe())
print(transactions.describe())

# Visualize the distribution of customer ages (assuming age exists)

sns.histplot(customers['Age'], kde=True)
plt.title('Distribution of Customer Ages')
plt.show()

# Visualize the total sales for different products

product_sales = transactions.groupby('ProductID')['Amount'].sum().reset_index()
merged_products = pd.merge(product_sales, products, on='ProductID')
sns.barplot(x='Amount', y='ProductName',
data=merged_products.sort_values('Amount', ascending=False))
plt.title('Total Sales per Product')
plt.show()

# Correlation heatmap to analyze numerical features

numerical_data = transactions[['Amount', 'Quantity']]  # Assuming relevant numerical
columns
corr_matrix = numerical_data.corr()
```

```
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Heatmap')
plt.show()
```



CustomerID	CustomerName	Region	SignupDate
C0001	Lawrence Carroll	South America	2022-07-10
C0002	Elizabeth Lutz	Asia	2022-02-13
C0003	Michael Rivera	South America	2024-03-07
C0200	Kelly Cross	Asia	2023-06-11

Category	Product Name	Price
Books	TechPro Textbook	₹497.76
Electronics	SoundWave Smartwatch	₹459.86
Clothing	SoundWave T-Shirt	₹481.78
Home Decor	ActiveWear Cookware Set	₹454.53