Task 1: Exploratory Data Analysis (EDA) and Business Insights:

1. **Data Preparation**

- **Load Data**: Read Customers.csv, Products.csv, and Transactions.csv into DataFrames.
- **Inspect Data**: Check for missing values, duplicates, and data types in each file.
- **Handle Missing Values**: Fill, drop, or impute missing values appropriately.
- **Date Conversion**: Convert Signup Date and Transaction Date into datetime format for temporal analysis.

2. **Exploratory Data Analysis (EDA)**

Customer Analysis:

  - Distribution of customers by region.
  - Customer sign-up trends over time.

Product Analysis:

  - Distribution of products across categories.
  - Price range analysis for different categories.

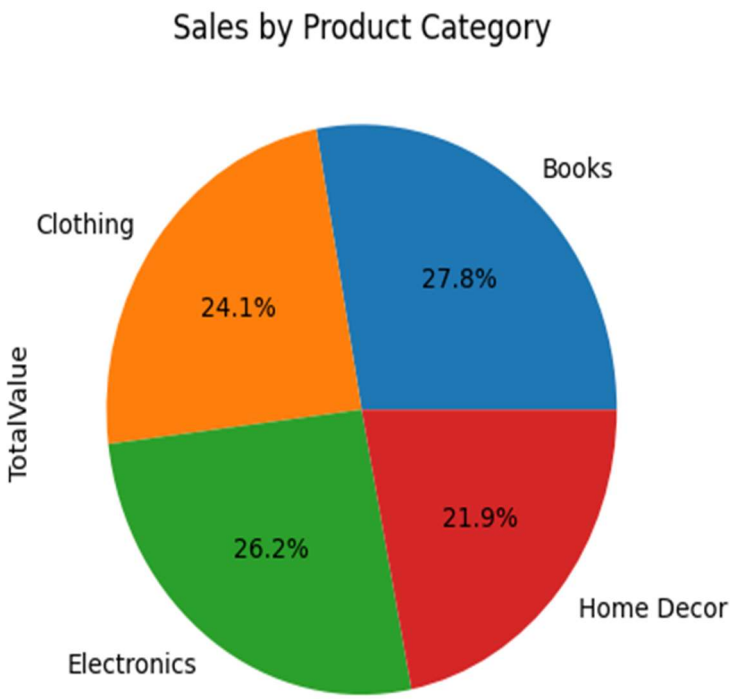
Transaction Analysis:

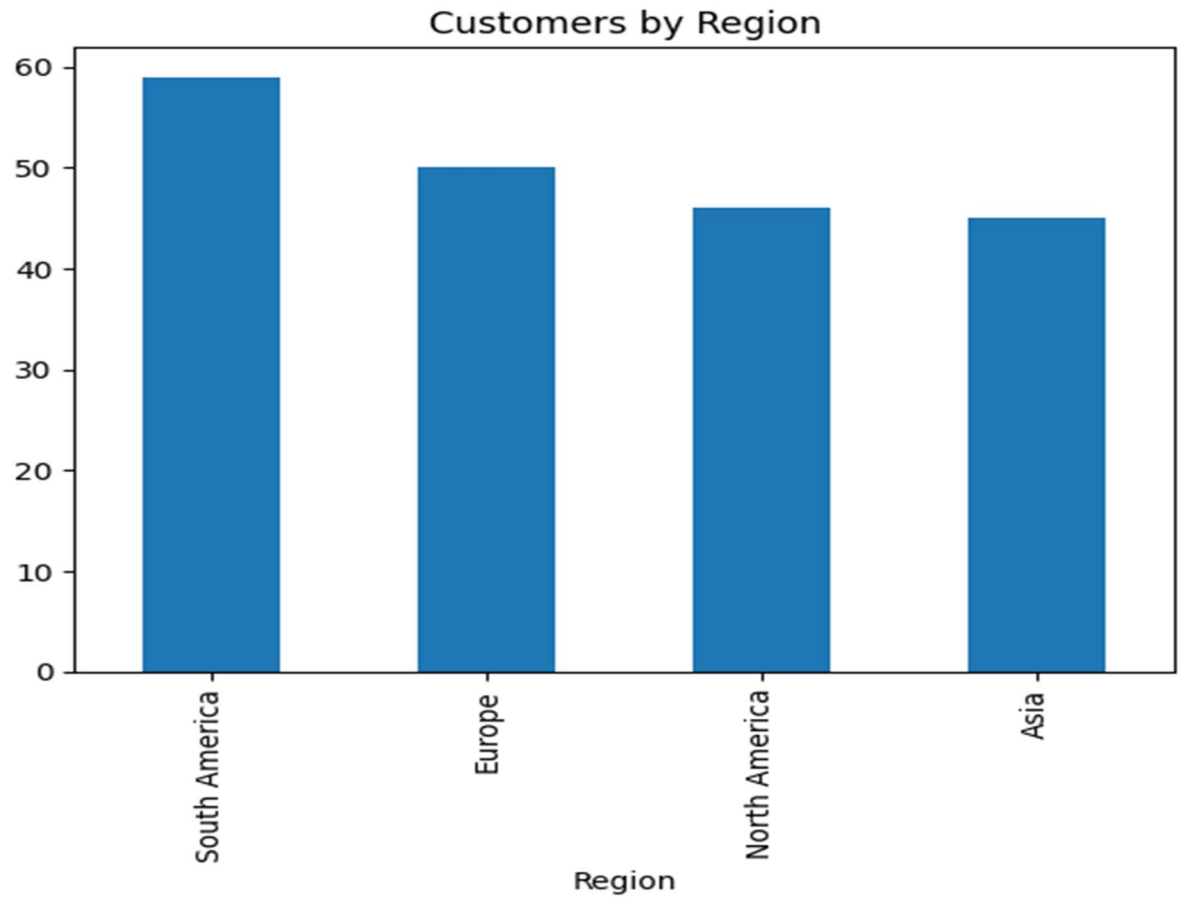
- Monthly/Yearly sales trends.
- Top-performing products based on total sales.
- Analysis of customer purchasing behavior (e.g., average transaction value, frequently bought products).

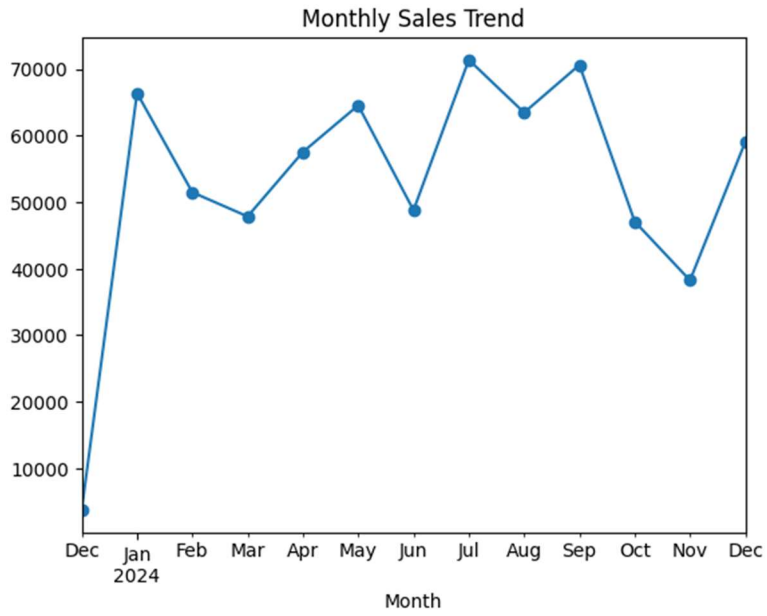
4. Python Script:

```
import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns


# Load datasets

customers = pd.read_csv('Customers.csv')

products = pd.read_csv('Products.csv')

transactions = pd.read_csv('Transactions.csv')
```

```
# Data Overview
print(customers.info())
print(products.info())
print(transactions.info())

# Handle missing values (example)
customers.fillna('Unknown', inplace=True)
products.dropna(subset=['Price'], inplace=True)

# Convert date columns
customers['SignupDate'] = pd.to_datetime(customers['SignupDate'])
transactions['TransactionDate'] = pd.to_datetime(transactions['TransactionDate'])

# EDA: Customers by Region
region_counts = customers['Region'].value_counts()
region_counts.plot(kind='bar', title='Customers by Region')
plt.show()

# EDA: Product Categories
category_sales = products.merge(transactions,
on='ProductID').groupby('Category')['TotalValue'].sum()
category_sales.plot(kind='pie', autopct='%1.1f%%', title='Sales by Product Category')
plt.show()

# EDA: Monthly Sales Trend
transactions['Month'] = transactions['TransactionDate'].dt.to_period('M')
monthly_sales = transactions.groupby('Month')['TotalValue'].sum()
monthly_sales.plot(title='Monthly Sales Trend', marker='o')
plt.show()
```


Sales by Product Category
Books
Clothing
27.8%
24.1%
TotalValue
21.9%
26.2%
Home Decor
Electronics


Customers by Region
60
50
40
30
20
10
0
South America
Europe
North America
Asia
Region



5. Business Insights :

Here are five business insights derived from the exploratory data analysis (EDA) visualizations:

1. **Product Category Sales Contribution:**
   Books contribute the highest sales value (27.8%), followed by Electronics (26.2%), Clothing (24.1%), and Home Decor (21.9%). This indicates that Books and Electronics are the most profitable categories.

2. **Regional Customer Distribution:**
   South America has the largest customer base, accounting for nearly 30% of the customers, while Asia has the least. This highlights an opportunity to expand marketing efforts in Asia to balance regional engagement.

3. **Monthly Sales Trend:**
   Sales peak in January, July, and September, with significant drops in October and February. Businesses can introduce promotional offers in low-performing months to stabilize sales performance.
4. **Customer Preferences by Region:**
   South America and Europe lead in customer numbers, showing strong regional preferences. Marketing strategies in these regions should focus on popular categories like Books and Electronics.
5. **Seasonal Sales Patterns:**
   Sales exhibit a cyclic pattern with fluctuations across months, suggesting seasonality in customer purchasing behavior. Analyzing further can help forecast inventory needs and ensure stock availability.