```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import StandardScaler
from sklearn.metrics.pairwise import cosine_similarity
from sklearn.cluster import KMeans
from sklearn.metrics import davies_bouldin_score
import warnings

warnings.filterwarnings('ignore')

# Load the datasets
customers_df = pd.read_csv('Customers.csv')
products_df = pd.read_csv('Products.csv')
transactions_df = pd.read_csv('Transactions.csv')

# Display the first few rows of each dataset
print(customers_df.head())
print(products_df.head())
print(transactions_df.head())

# Check for missing values
print(customers_df.isnull().sum())
print(products_df.isnull().sum())
print(transactions_df.isnull().sum())

# Basic information about the datasets
print(customers_df.info())
print(products_df.info())
print(transactions_df.info())
```

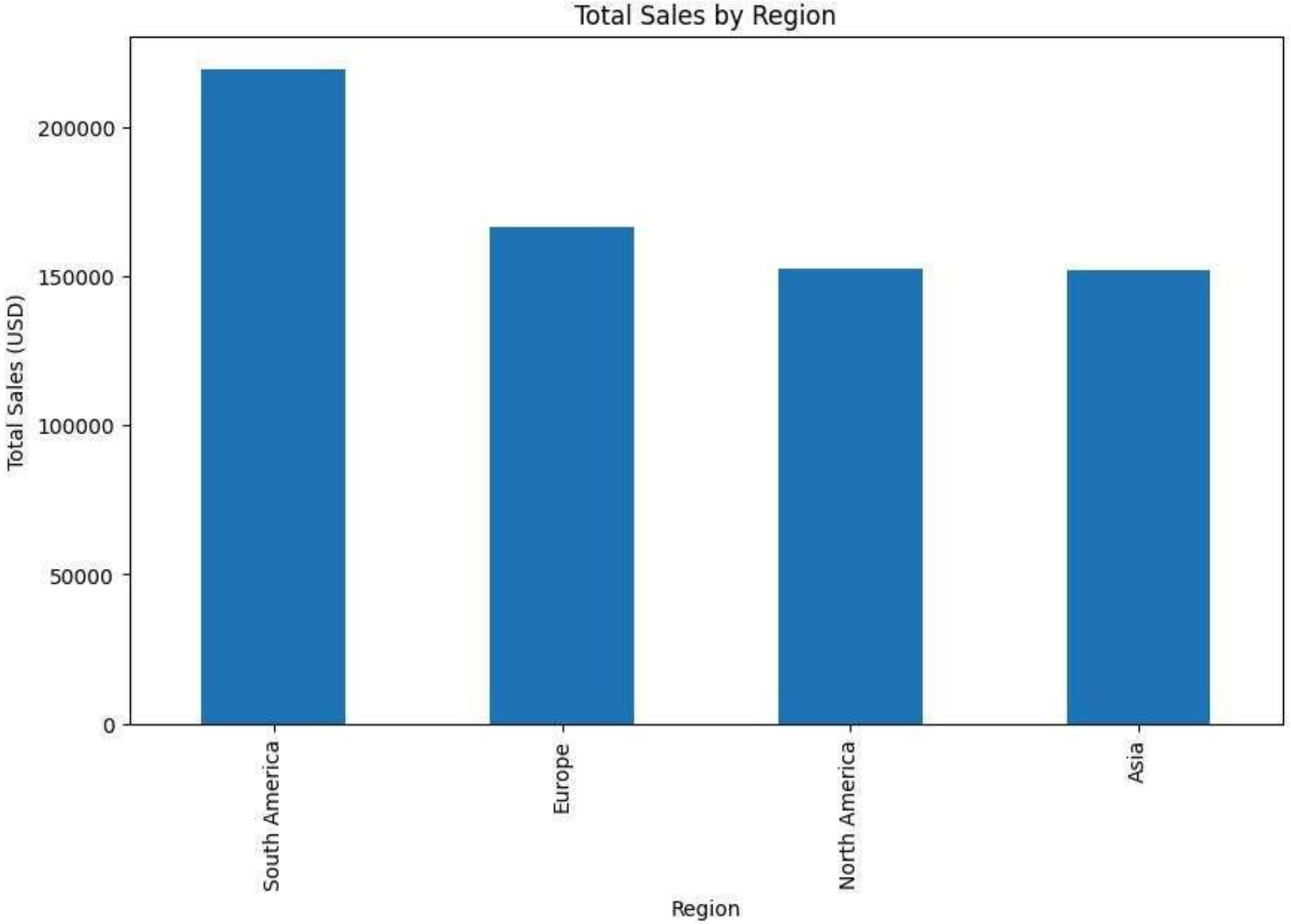
```
Category       0
Price          0
dtype: int64
TransactionID      0
CustomerID         0
ProductID          0
TransactionDate    0
Quantity           0
TotalValue         0
Price              0
dtype: int64
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 4 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   CustomerID    200 non-null    object
 1   CustomerName  200 non-null    object
```

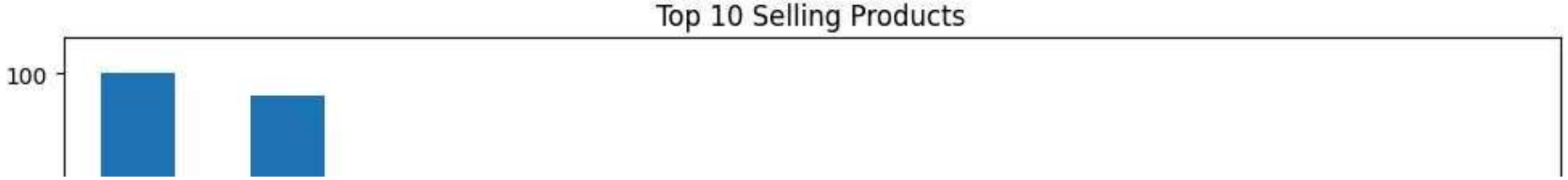
```
 3   TransactionDate  1000 non-null   object
 4   Quantity         1000 non-null   int64
 5   TotalValue       1000 non-null   float64
 6   Price            1000 non-null   float64
dtypes: float64(2), int64(1), object(4)
memory usage: 54.8+ KB
None
```

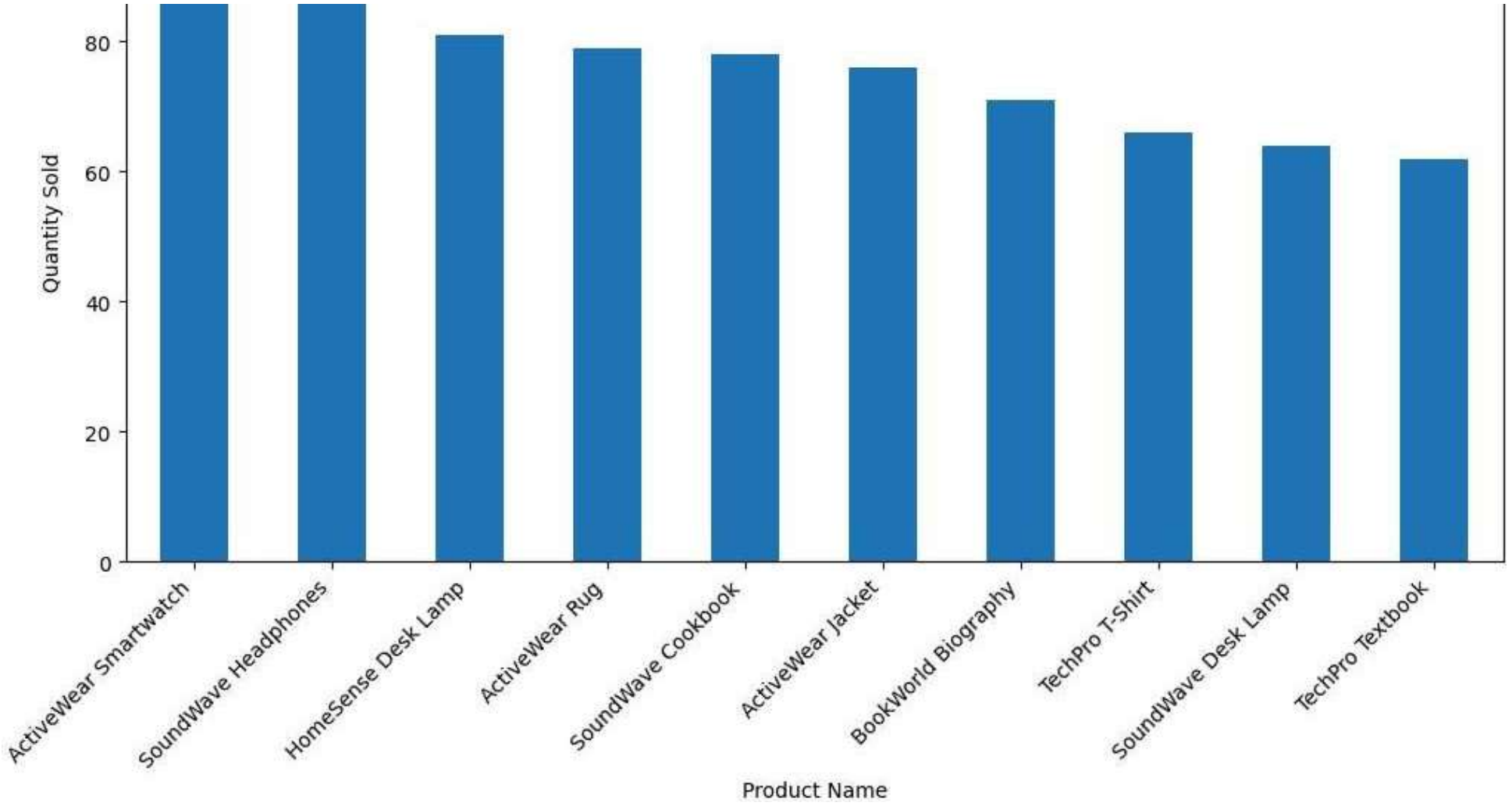
```
merged_df = transactions_df.merge(customers_df, on='CustomerID').merge(products_df, on='ProductID')

# 1. Total sales by region

region_sales = merged_df.groupby('Region')['TotalValue'].sum().sort_values(ascending=False)
plt.figure(figsize=(10, 6))
region_sales.plot(kind='bar')
plt.title('Total Sales by Region')
plt.xlabel('Region')
plt.ylabel('Total Sales (USD)')
plt.show()

# 2. Top 10 selling products
top_products = merged_df.groupby('ProductName')['Quantity'].sum().sort_values(ascending=False).head(10)
plt.figure(figsize=(12, 6))
top_products.plot(kind='bar')
plt.title('Top 10 Selling Products')
plt.xlabel('Product Name')
plt.ylabel('Quantity Sold')
plt.xticks(rotation=45, ha='right')
plt.show()

# 3. Sales trend over time
merged_df['TransactionDate'] = pd.to_datetime(merged_df['TransactionDate'])
daily_sales = merged_df.groupby('TransactionDate')['TotalValue'].sum()
plt.figure(figsize=(12, 6))
daily_sales.plot()
plt.title('Daily Sales Trend')
plt.xlabel('Date')
plt.ylabel('Total Sales (USD)')
plt.show()
```

```
# 4. Customer acquisition over time
customers_df['SignupDate'] = pd.to_datetime(customers_df['SignupDate'])
customer_growth = customers_df.groupby('SignupDate').size().cumsum()
plt.figure(figsize=(12, 6))
customer_growth.plot()
plt.title('Customer Growth Over Time')
plt.xlabel('Date')
plt.ylabel('Total Customers')
plt.show()

# 5. Average transaction value by category
avg_transaction_by_category = merged_df.groupby('Category')['TotalValue'].mean().sort_values(ascending=False)
plt.figure(figsize=(10, 6))
avg_transaction_by_category.plot(kind='bar')
plt.title('Average Transaction Value by Category')
plt.xlabel('Category')
plt.ylabel('Average Transaction Value (USD)')
plt.xticks(rotation=45, ha='right')
plt.show()
```

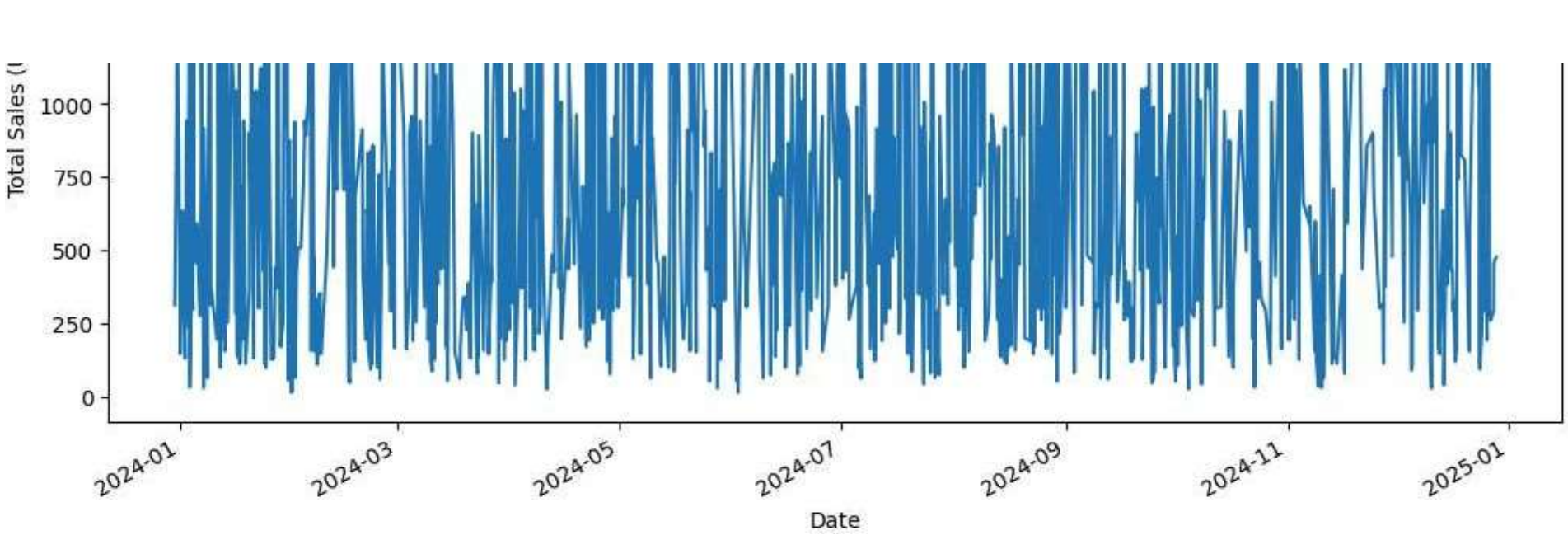
[ ]

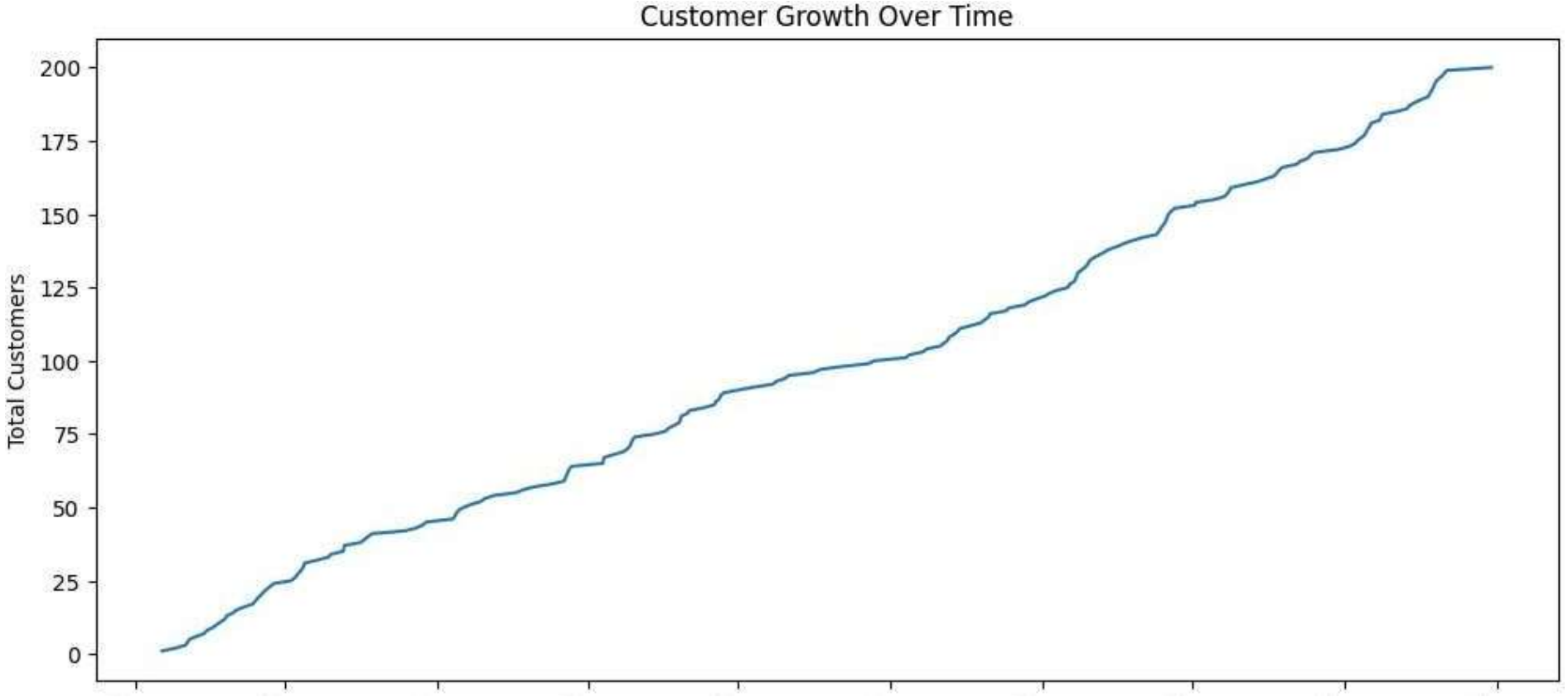

Total Sales by Region
Total Sales (USD)
200000
150000
100000
50000
0
South America
Europe
North America
Asia
Region

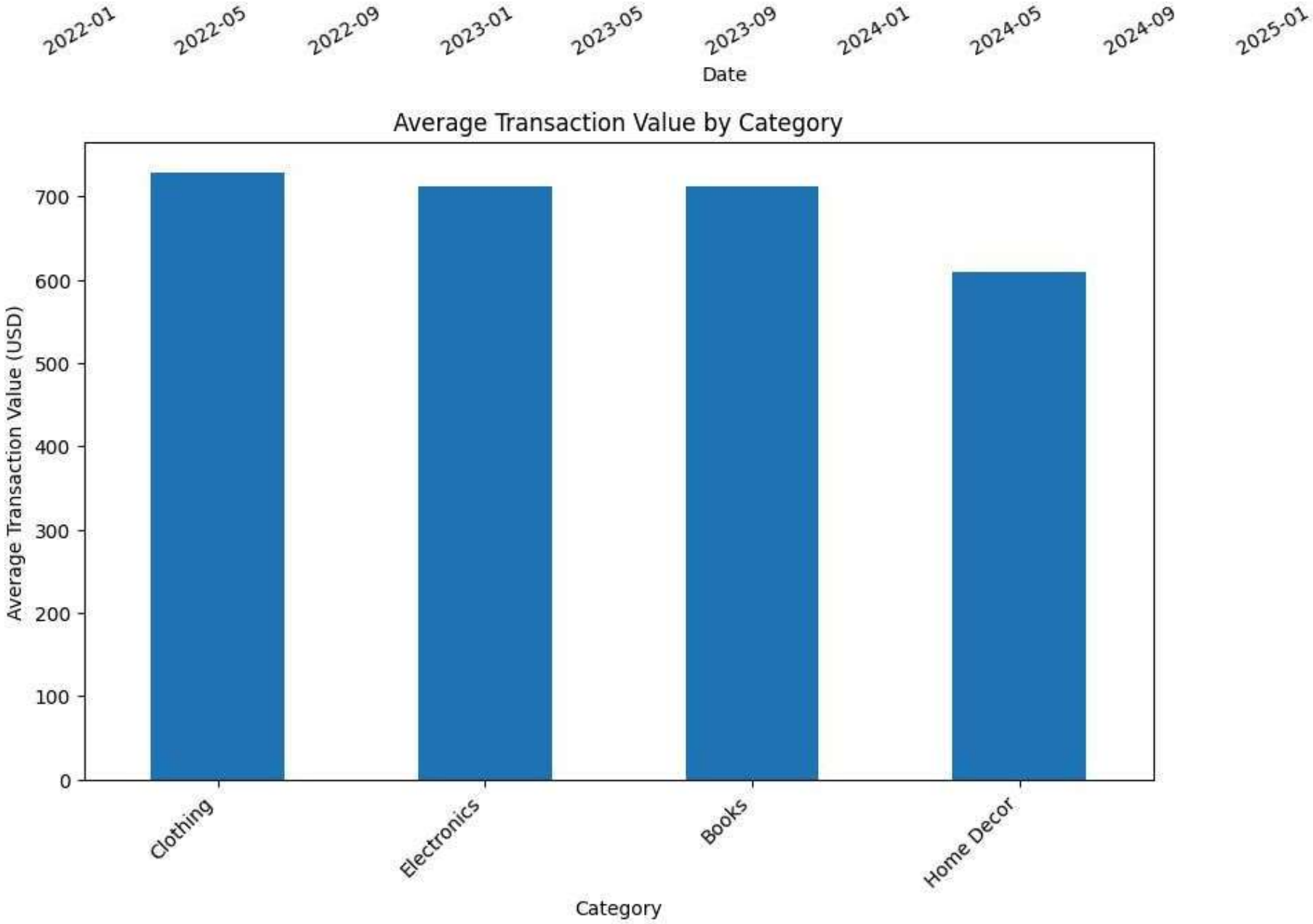

Top 10 Selling Products
100


Quantity Sold
0
20
40
60
80
ActiveWear Smartwatch
SoundWave Headphones
HomeSense Desk Lamp
ActiveWear Rug
SoundWave Cookbook
ActiveWear Jacket
BookWorld Biography
TechPro T-Shirt
SoundWave Desk Lamp
TechPro Textbook
Product Name


Daily Sales Trend
2000
1750
1500
1250
USD)


Total Sales (
1000
750
500
250
0
2024-01
2024-03
2024-05
2024-07
2024-09
2024-11
2025-01
Date


Customer Growth Over Time
Total Customers
200
175
150
125
100
75
50
25
0


2022-01
2022-05
2022-09
2023-01
2023-05
2023-09
2024-01
2024-05
2024-09
2025-01
Date
Average Transaction Value by Category
Average Transaction Value (USD)
0
100
200
300
400
500
600
700
Clothing
Electronics
Books
Home Decor
Category

Based on this exploratory data analysis, here are five business insights:

1. **Regional Sales Distribution:** There are significant differences in total sales across regions. This suggests that the company should focus on understanding the factors contributing to higher sales in top-performing regions and apply these insights to boost sales in underperforming areas.
2. **Product Popularity:** The top 10 selling products analysis reveals which items are driving the most volume. This information can be used to optimize inventory management, focus marketing efforts, and potentially develop similar products to capitalize on customer preferences.
3. **Sales Seasonality:** The daily sales trend shows clear patterns of peaks and troughs. This information can be used to adjust inventory levels, plan marketing campaigns, and allocate resources more effectively during high-sales periods.
4. **Customer Growth Rate:** The customer acquisition graph shows the rate at which the company is gaining new customers. Any changes in this growth rate could indicate the effectiveness of marketing strategies or changes in market conditions, providing valuable feedback for customer acquisition efforts.
5. **Category Performance:** The average transaction value by category highlights which product categories are generating the highest value per transaction. This insight can guide pricing strategies, product development, and marketing focus to maximize revenue.