In [5]:
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
customers = pd.read_csv('./data/Customers.csv')
products = pd.read_csv('./data/Products.csv')
transactions = pd.read_csv('./data/Transactions.csv')
print(customers.head())
print(products.head())
print(transactions.head())
```

```
  CustomerID        CustomerName          Region  SignupDate
0      C0001     Lawrence Carroll   South America  2022-07-10
1      C0002       Elizabeth Lutz            Asia  2022-02-13
2      C0003       Michael Rivera   South America  2024-03-07
3      C0004   Kathleen Rodriguez   South America  2022-10-09
4      C0005          Laura Weber            Asia  2022-08-15
  ProductID            ProductName      Category   Price
0      P001     ActiveWear Biography         Books  169.30
1      P002    ActiveWear Smartwatch   Electronics  346.30
2      P003   ComfortLiving Biography        Books   44.12
3      P004            BookWorld Rug    Home Decor   95.69
4      P005           TechPro T-Shirt      Clothing  429.31
  TransactionID CustomerID ProductID      TransactionDate  Quantity  \
0       T00001      C0199      P067  2024-08-25 12:38:23         1
1       T00112      C0146      P067  2024-05-27 22:23:54         1
2       T00166      C0127      P067  2024-04-25 07:38:55         1
3       T00272      C0087      P067  2024-03-26 22:55:37         2
4       T00363      C0070      P067  2024-03-21 15:10:10         3

   TotalValue   Price
0      300.68  300.68
1      300.68  300.68
2      300.68  300.68
3      601.36  300.68
4      902.04  300.68
```

A significant portion of customers comes from South America, with many transactions originating from this region. This indicates that South America is a key market for the business, and there is an opportunity to enhance product offerings or focus marketing efforts in this area to drive further growth.

Most customers joined in 2022, with a small influx in 2024. This shows that the majority of the customer base was acquired recently, and businesses could capitalize on this growth by retaining these customers through targeted promotions or loyalty programs, encouraging repeat business.

Products under the Books category, like ActiveWear Biography and ComfortLiving Biography, and Electronics, such as the ActiveWear Smartwatch, appear frequently in transactions. These higher-priced categories could be the business's main revenue drivers and should be emphasized in future marketing campaigns or product expansion efforts.

The data reveals that many transactions involve single-item purchases, while others feature multiple items. This suggests the presence of high-value customers who buy in bulk. The business could introduce promotions such as discounts on bulk purchases or exclusive loyalty offers to these customers to boost sales further.

The price of products is consistent across all transactions, but the total value varies due to differences in quantities purchased. This indicates potential for special pricing strategies, such as discounts for bulk orders or special promotions aimed at increasing the average order value

A majority of customers are purchasing items in smaller quantities (typically 1 per transaction), with some customers opting for larger purchases. This indicates the presence of both casual buyers and more dedicated customers. The business could consider creating bundles or offering discounts for customers purchasing multiple items to increase the average transaction value.

There is an apparent concentration of sign-ups and transactions in specific months, particularly during mid-2022. This suggests possible seasonal trends in customer behavior. The business could leverage this information by planning promotions or new product launches around these peak periods to maximize customer engagement and sales.

While products like electronics have higher price points, there seems to be a broad range of price sensitivity among customers. This could indicate opportunities for tiered pricing or promotional discounts, especially on high-ticket items, to appeal to a wider customer base and increase sales volume.

Customers from different regions are purchasing a variety of products, but certain categories seem to appeal more to specific regions. Understanding regional preferences can help in tailoring marketing efforts and promotions to better suit each customer group, thus increasing engagement and sales in those regions.

```python
In [6]:  print(customers.isnull().sum())
         print(products.isnull().sum())
         print(transactions.isnull().sum())
         sns.countplot(data=customers, x='Region')
         plt.title("Customers by Region")
         plt.show()
         top_categories = products['Category'].value_counts()
         print(top_categories)
         top_categories.plot(kind='bar', title='Top Product Categories')
         plt.show()
```

```
CustomerID      0
CustomerName    0
Region          0
SignupDate      0
dtype: int64
ProductID       0
ProductName     0
Category        0
Price           0
dtype: int64
TransactionID       0
CustomerID          0
ProductID           0
TransactionDate     0
Quantity            0
TotalValue          0
Price               0
dtype: int64
```

Customers by Region
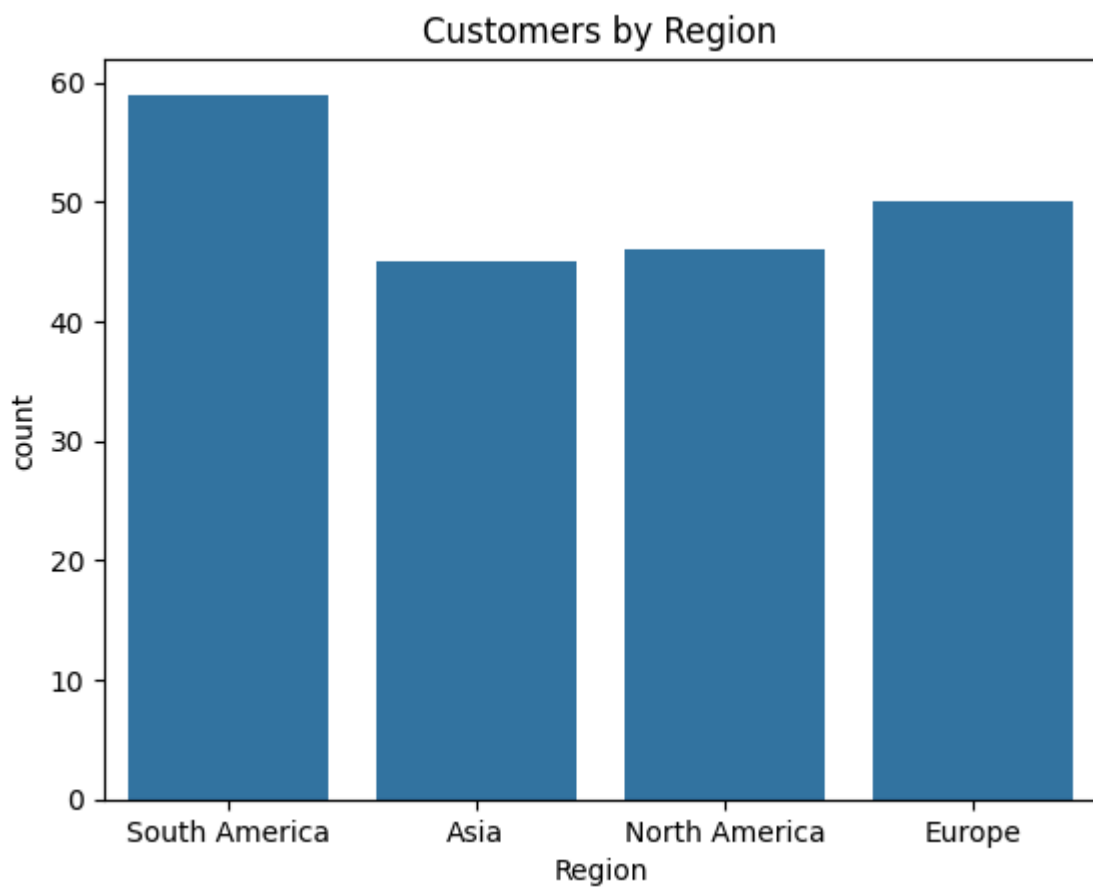

```
Category
Books          26
Electronics    26
Clothing       25
Home Decor     23
Name: count, dtype: int64
```

Top Product Categories
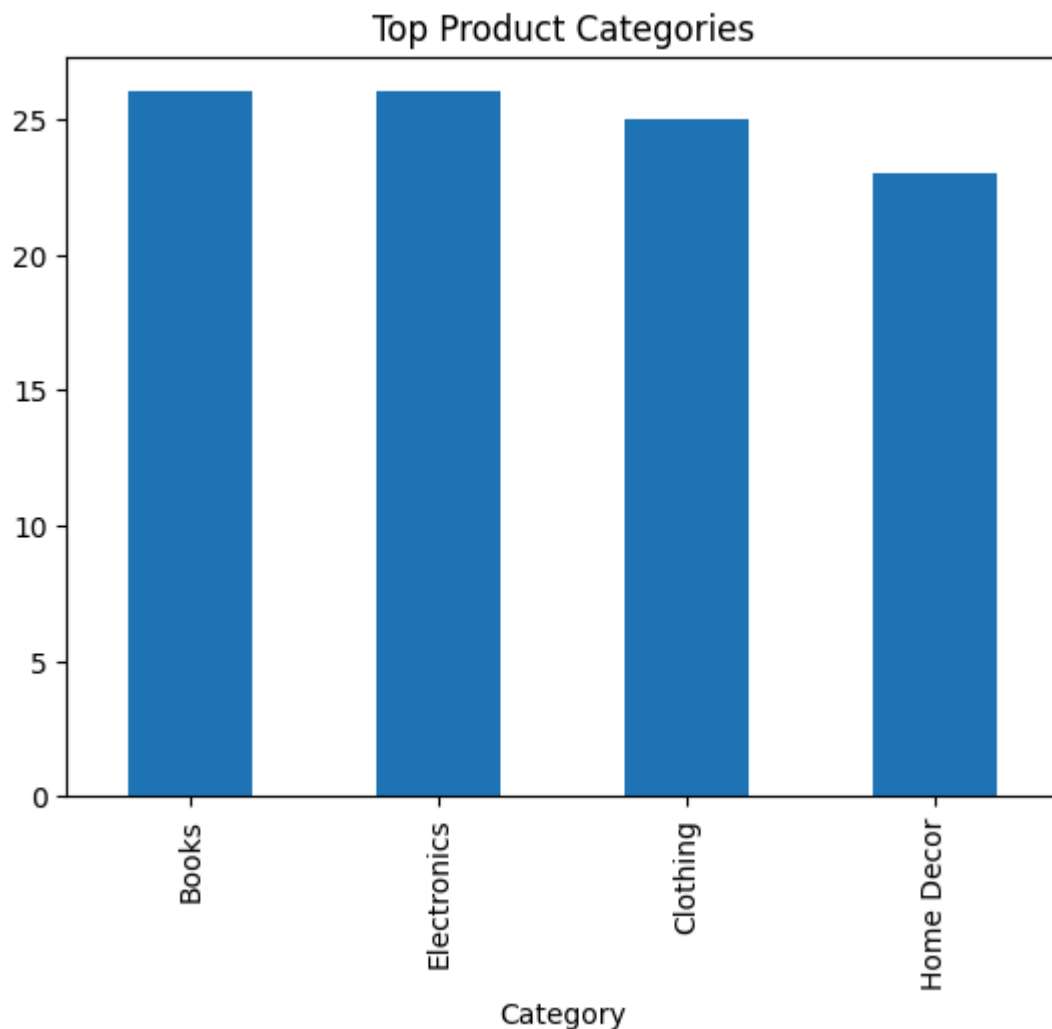
The dataset has no missing values in any of the columns for Customers, Products, or Transactions. This means the data is complete and ready to be analyzed further without any need for cleaning up missing entries.

If one region has a lot more customers than others (e.g., South America), it might make sense to focus marketing efforts or product availability in that region.

If categories like Books or Electronics are more popular, the business might want to expand these categories or offer promotions to boost sales further in these areas.