In [48]:
```
import pandas as pd
import matplotlib.pyplot as plt
customers=pd.read_csv(r"J:\Downloads\Customers.csv")
products=pd.read_csv(r"J:\Downloads\Products.csv")
transactions=pd.read_csv(r"J:\Downloads\Transactions.csv")
```

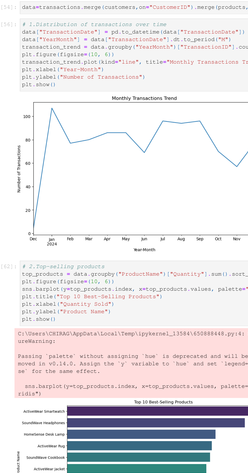
In [52]:
```
print("Customers Dataset Overview:")
print(customers.head(), customers.info(), customers.describe())

print("Products Dataset Overview:")
print(products.head(), products.info(), products.describe())

print("Transactions Dataset Overview:")
print(transactions.head(), transactions.info(), transactions.describe
```

```
Customers Dataset Overview:
<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
 2   Region      200 non-null    object
 3   SignupDate  200 non-null    object
dtypes: object(4)
memory usage: 6.4 KB
  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 None
CustomerID     CustomerName     Region  SignupDate
count         200              200         200         200
unique        200              200           4         179
top         C0001  Lawrence Carroll  South America  2024-11-11
freq            1                1          59           3
Products Dataset Overview:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 4 columns):
 #   Column      Non-Null Count  Dtype 
---  ------      --------------  ----- 
 0   ProductID   100 non-null    object
 1   ProductName  100 non-null    object
 2   Category    100 non-null    object
 3   Price       100 non-null    float64
dtypes: float64(1), object(3)
memory usage: 3.3+ KB
  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         BookNook Rug  Home Decor   95.69
4      P005         TechPro T-Shirt     Clothing  429.31 None
Price
count  100.000000
mean   267.551700
std    143.219383
min    16.080000
25%    147.767500
50%    292.875000
75%    397.090000
max    497.760000
Transactions Dataset Overview:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 7 columns):
 #   Column         Non-Null Count  Dtype 
---  ------         --------------  ----- 
 0   TransactionID  1000 non-null   object
 1   CustomerID     1000 non-null   object
 2   ProductID      1000 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
  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        T00146      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  None          Quantity   TotalValue       Pri
ce
count  1000.000000  1000.000000  1000.00000
mean      2.537000   689.995560   272.55407
std       1.117981   493.144478   140.73639
min       1.000000    16.080000    16.08000
25%       2.000000   295.295000   147.95000
50%       3.000000   588.880000   299.93000
75%       4.000000  1011.660000   404.40000
max       4.000000  1991.040000   497.76000
```

In [ ]:
```
# Task 1: Exploratory Data Analysis (EDA) and Business Insights
```

In [58]:
```
data=transactions.merge(customers,on="CustomerID").merge(products,on=
```

In [56]:
```
# 1.Distribution of transactions over time
data["TransactionDate"] = pd.to_datetime(data["TransactionDate"])
data["YearMonth"] = data["TransactionDate"].dt.to_period("M")
transactions_trend = data.groupby("YearMonth")["TransactionID"].count()
plt.figure(figsize=(10, 6))
transaction_trend.plot(kind="line", title="Monthly Transactions Trend"
plt.xlabel("Year-Month")
plt.ylabel("Number of Transactions")
plt.show()
```



In [60]:
```
# 2.Top-selling products
top_products = data.groupby("ProductName")["Quantity"].sum().sort_valu
plt.figure(figsize=(10, 6))
sns.barplot(y=top_products.index, x=top_products.values, palette="viri
plt.title("Top 10 Best-Selling Products")
plt.xlabel("Quantity Sold")
plt.ylabel("Product Name")
plt.show()
```

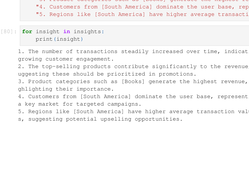
```
C:\Users\CHIRAG\AppData\Local\Temp\ipykernel_13584\650888448.py:4: Fut
ureWarning:

Passing `palette` without assigning `hue` is deprecated and will be re
moved in v0.14.0. Assign the `y` variable to `hue` and set `legend=fal
se` for the same effect.

  sns.barplot(y=top_products.index, x=top_products.values, palette="vi
ridis")
```

In [64]:
```
# 3.Revenue contribution by category
category_revenue = data.groupby("Category")["TotalValue"].sum().sort_v
plt.figure(figsize=(10, 6))
sns.barplot(x=category_revenue.index, y=category_revenue.values, palet
plt.title("Revenue Contribution by Product Category")
plt.xlabel("Product Category")
plt.ylabel("Total Revenue")
plt.xticks(rotation=45)
plt.show()
```

```
C:\Users\CHIRAG\AppData\Local\Temp\ipykernel_13584\3415018261.py:4: Fu
tureWarning:

Passing `palette` without assigning `hue` is deprecated and will be re
moved in v0.14.0. Assign the `x` variable to `hue` and set `legend=fal
se` for the same effect.

  sns.barplot(x=category_revenue.index, y=category_revenue.values, pal
ette="coolwarm")
```

In [68]:
```
# 4. Region-wise customer distribution
region_distribution = customers["Region"].value_counts()
plt.figure(figsize=(8, 6))
sns.barplot(x=region_distribution.index, y=region_distribution.values,
plt.title("Customer Distribution by Region")
plt.xlabel("Region")
plt.ylabel("Number of Customers")
plt.show()
```

```
C:\Users\CHIRAG\AppData\Local\Temp\ipykernel_13584\3690725065.py:4: Fu
tureWarning:

Passing `palette` without assigning `hue` is deprecated and will be re
moved in v0.14.0. Assign the `x` variable to `hue` and set `legend=fal
se` for the same effect.

  sns.barplot(x=region_distribution.index, y=region_distribution.value
s, palette="Pastel1")
```



In [70]:
```
# 5. Average transaction value per region
region_avg_transaction = data.groupby("Region")["TotalValue"].mean().s
plt.figure(figsize=(8, 6))
sns.barplot(x=region_avg_transaction.index, y=region_avg_transaction.v
plt.title("Average Transaction Value by Region")
plt.xlabel("Region")
plt.ylabel("Average Transaction Value (USD)")
plt.show()
```

```
C:\Users\CHIRAG\AppData\Local\Temp\ipykernel_13584\3190929798.py:4: Fu
tureWarning:

Passing `palette` without assigning `hue` is deprecated and will be re
moved in v0.14.0. Assign the `x` variable to `hue` and set `legend=fal
se` for the same effect.

  sns.barplot(x=region_avg_transaction.index, y=region_avg_transactio
n.values, palette="rocket")
```


In [78]:
```
# Business Insights
insights = [
    "1. The number of transactions steadily increased over time, indic
    "2. The top-selling products contribute significantly to the rever
    "3. Product categories such as [Books] generate the highest reven
    "4. Customers from [South America] dominate the user base, represe
    "5. Regions like [South America] have higher average transaction v
]

for insight in insights:
    print(insight)
```

```
1. The number of transactions steadily increased over time, indicating
growing customer engagement.
2. The top-selling products contribute significantly to the revenue, a
uggesting these should be prioritized in promotions.
3. Product categories such as [Books] generate the highest revenue, hi
ghlighting their importance.
4. Customers from [South America] dominate the user base, representing
a key market for targeted campaigns.
5. Regions like [South America] have higher average transaction value
x, suggesting potential upselling opportunities.
```