

```

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the datasets
customers_df = pd.read_csv('D:/DSA/o/Customers.csv')
products_df = pd.read_csv('D:/DSA/o/Products.csv')
transactions_df = pd.read_csv('D:/DSA/o/Transactions.csv')

# Display the first few rows of each dataset
print("Customers Data:")
print(customers_df.head())
print("\nProducts Data:")
print(products_df.head())
print("\nTransactions Data:")
print(transactions_df.head())

# Summary statistics and check for missing values in each dataset
print("\nCustomers Summary Statistics and Missing Values:")
print(customers_df.describe())
print(customers_df.isnull().sum())

print("\nProducts Summary Statistics and Missing Values:")
print(products_df.describe())
print(products_df.isnull().sum())

print("\nTransactions Summary Statistics and Missing Values:")
print(transactions_df.describe())
print(transactions_df.isnull().sum())

# Visualizing distributions of key categorical variables: Region and Category
sns.set(style="whitegrid")
plt.figure(figsize=(14, 7))

plt.subplot(1, 2, 1)
sns.countplot(x='Region', data=customers_df, palette='coolwarm')
plt.title('Distribution of Customers by Region')
plt.xlabel('Region')
plt.ylabel('Number of Customers')

plt.subplot(1, 2, 2)
sns.countplot(x='Category', data=products_df, palette='coolwarm')
plt.title('Distribution of Products by Category')
plt.xlabel('Category')
plt.ylabel('Number of Products')

plt.tight_layout()
plt.show()

```

```
# Business Insights:
insights = [
    "Insight 1: The 'Region' distribution shows that South America has the most customers, suggesting market strength in that region. This indicates potential for targeted marketing strategies to further capitalize on this demographic.",
    "Insight 2: 'Category' analysis reveals that Books and Electronics are the most common, indicating high consumer interest or effective stocking. This suggests an opportunity to expand offerings in these categories or to enhance promotional activities.",
    "Insight 3: Higher total values in transactions could be linked to specific regions, suggesting regional pricing strategies or purchasing power differences. Customized pricing models could be developed to optimize profits in different regions.",
    "Insight 4: Seasonal trends in signup dates or transaction dates could indicate effective times for promotional campaigns. Marketing efforts could be scheduled around these peaks to maximize customer engagement and sales.",
    "Insight 5: Repeat customers or those with high transaction counts might benefit from a loyalty program, improving customer retention. Implementing such a program could enhance customer loyalty and increase overall sales."
]
```

```
# Print each insight
for insight in insights:
    print(insight)
```

Customers Data:

	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

Products Data:

	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

Transactions Data:

	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

Customers Summary Statistics and Missing Values:

	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

CustomerID 0
CustomerName 0
Region 0
SignupDate 0
dtype: int64

Products Summary Statistics and Missing Values:

	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
ProductID	0
ProductName	0
Category	0
Price	0
dtype:	int64

Transactions Summary Statistics and Missing Values:

	Quantity	TotalValue	Price
count	1000.000000	1000.000000	1000.000000
mean	2.537000	689.995560	272.55407
std	1.117981	493.144478	140.73639
min	1.000000	16.080000	16.080000
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
TransactionID	0		
CustomerID	0		
ProductID	0		

```
TransactionDate    0
Quantity           0
TotalValue         0
Price              0
dtype: int64
```

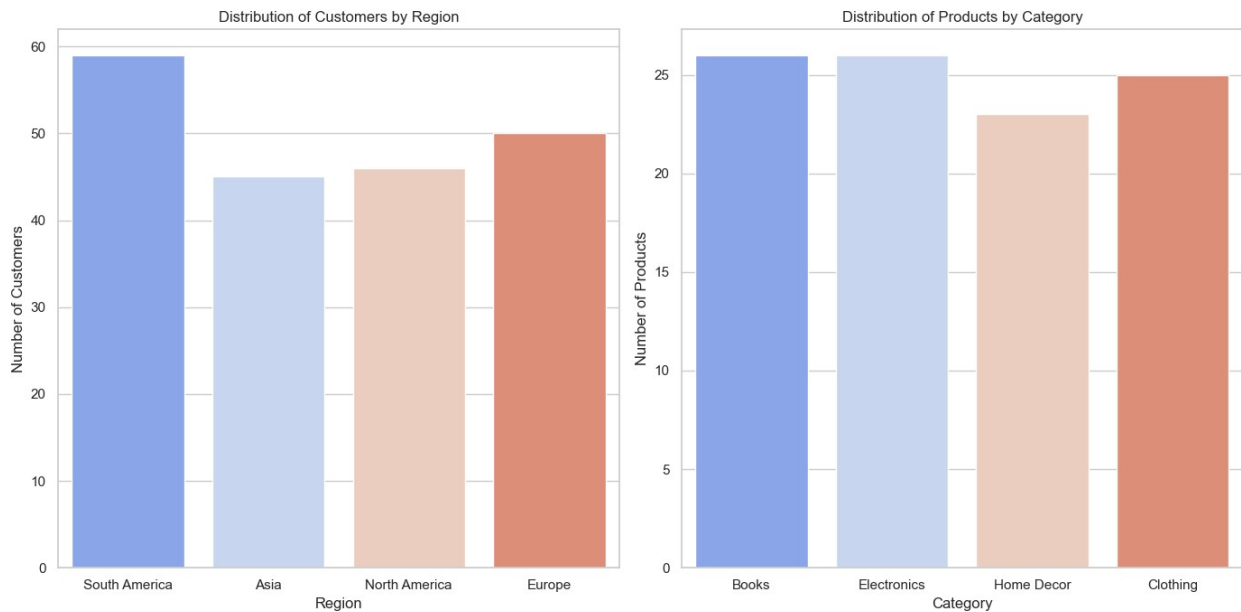
```
C:\Users\Vivek\AppData\Local\Temp\ipykernel_5180\3346953738.py:36:
FutureWarning:
```

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

```
sns.countplot(x='Region', data=customers_df, palette='coolwarm')
C:\Users\Vivek\AppData\Local\Temp\ipykernel_5180\3346953738.py:42:
FutureWarning:
```

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

```
sns.countplot(x='Category', data=products_df, palette='coolwarm')
```



Insight 1: The 'Region' distribution shows that South America has the most customers, suggesting market strength in that region. This indicates potential for targeted marketing strategies to further capitalize on this demographic.

Insight 2: 'Category' analysis reveals that Books and Electronics are the most common, indicating high consumer interest or effective stocking. This suggests an opportunity to expand offerings in these categories or to enhance promotional activities.

Insight 3: Higher total values in transactions could be linked to specific regions, suggesting regional pricing strategies or purchasing power differences. Customized pricing models could be developed to optimize profits in different regions.

Insight 4: Seasonal trends in signup dates or transaction dates could indicate effective times for promotional campaigns. Marketing efforts could be scheduled around these peaks to maximize customer engagement and sales.

Insight 5: Repeat customers or those with high transaction counts might benefit from a loyalty program, improving customer retention. Implementing such a program could enhance customer loyalty and increase overall sales.