

```
In [3]: import pandas as pd

In [35]: customers = pd.read_csv('C:\Users\yashu\Downloads\Customers.csv')
products = pd.read_csv('C:\Users\yashu\Downloads\Products.csv')
transactions = pd.read_csv('C:\Users\yashu\Downloads\Transactions.csv')

In [37]: customers

Out[37]:
```

	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
...
195	C0196	Laura Watts	Europe	2022-06-07
196	C0197	Christina Harvey	Europe	2023-03-21
197	C0198	Rebecca Ray	Europe	2022-02-27
198	C0199	Andrea Jenkins	Europe	2022-12-03
199	C0200	Kelly Cross	Asia	2023-06-11

200 rows × 4 columns

```
In [39]: products

Out[39]:
```

	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
...
95	P096	SoundWave Headphones	Electronics	307.47
96	P097	BookWorld Cookbook	Books	319.34
97	P098	SoundWave Laptop	Electronics	299.93
98	P099	SoundWave Mystery Book	Books	354.29
99	P100	HomeSense Sweater	Clothing	126.34

100 rows × 4 columns

```
In [41]: transactions

Out[41]:
```

	TransactionID	CustomerID	ProductID	TransactionDate	Quantity	TotalValue	Price
0	T00001	C0199	P067	2024-08-25 12:38:23	1	300.68	300.68
1	T00112	C0146	P067	2024-05-27 22:23:54	1	300.68	300.68
2	T00166	C0127	P067	2024-04-25 07:38:55	1	300.68	300.68
3	T00272	C0087	P067	2024-03-26 22:55:37	2	601.36	300.68
4	T00363	C0070	P067	2024-03-21 15:10:10	3	902.04	300.68
...
995	T00496	C0118	P037	2024-10-24 08:30:27	1	459.86	459.86
996	T00759	C0059	P037	2024-06-04 02:15:24	3	1379.58	459.86
997	T00922	C0018	P037	2024-04-05 13:05:32	4	1839.44	459.86
998	T00959	C0115	P037	2024-09-29 10:16:02	2	919.72	459.86
999	T00992	C0024	P037	2024-04-21 10:52:24	1	459.86	459.86

1000 rows × 7 columns

```
In [43]: customers_info = customers.info()
products_info = products.info()
transactions_info = transactions.info()

<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

<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

<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
```

```
In [45]: customers_head = customers.head()
products_head = products.head()
transactions_head = transactions.head()
```

```
In [47]: customers_info, customers_head, products_info, products_head, transactions_info, transactions_head
```

```
Out[47]: (None,
0 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,
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,
None,
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 )
```

```
In [49]: customers['SignupDate'] = pd.to_datetime(customers['SignupDate'], errors='coerce')
transactions['TransactionDate'] = pd.to_datetime(transactions['TransactionDate'], errors='coerce')
```

```
In [51]: customers_missing = customers.isnull().sum()
products_missing = products.isnull().sum()
transactions_missing = transactions.isnull().sum()
```

```
In [53]: customers_missing, products_missing, transactions_missing
```

```
Out[53]: (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)
```

```
In [55]: import matplotlib.pyplot as plt
import seaborn as sns
```

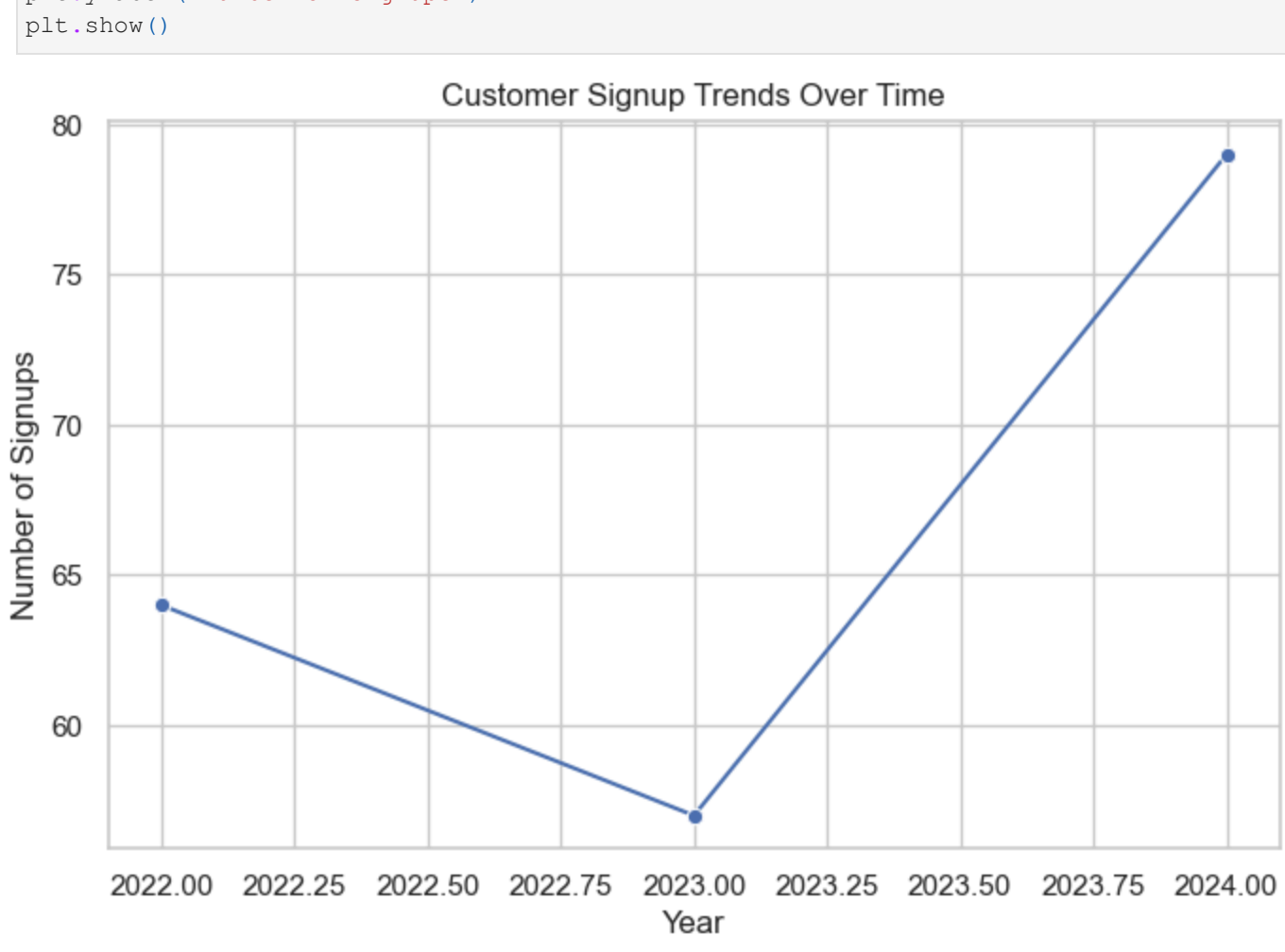
```
In [57]: sns.set_theme(style="whitegrid")
```

```
In [59]: region_distribution = customers['Region'].value_counts()
plt.figure(figsize=(8, 5))
sns.barplot(x=region_distribution.index, y=region_distribution.values, palette="viridis")
plt.title("Customer Distribution by Region")
plt.xlabel("Region")
plt.ylabel("Number of Customers")
plt.show()
```

C:\Users\yashu\AppData\Local\Temp\ipykernel_23972\365805721.py:3: 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.barplot(x=region_distribution.index, y=region_distribution.values, palette="viridis")

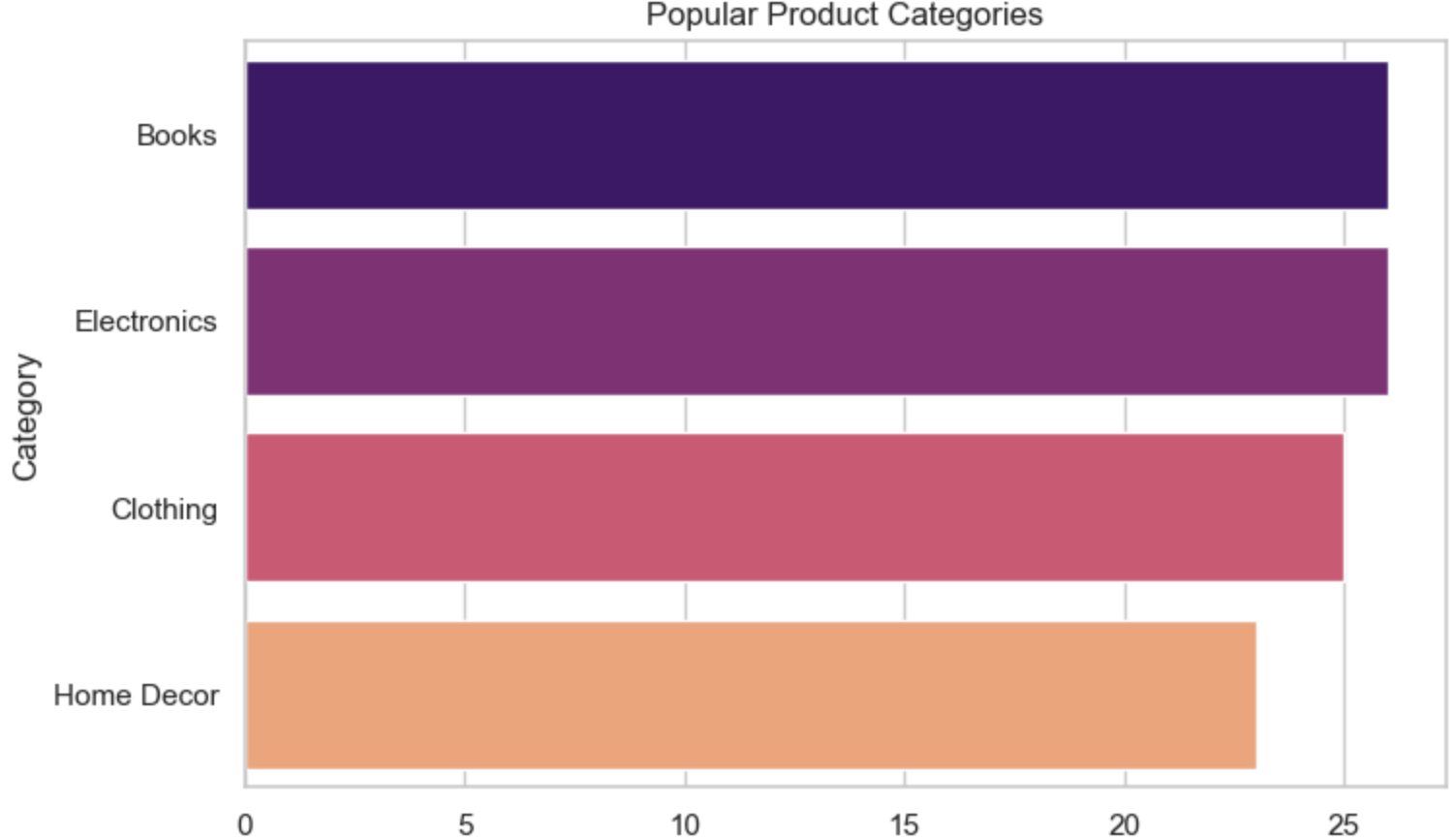


```
In [61]: customers['SignupYear'] = customers['SignupDate'].dt.year
signup_trends = customers['SignupYear'].value_counts().sort_index()
plt.figure(figsize=(8, 5))
sns.lineplot(x=signup_trends.index, y=signup_trends.values, marker='o', color='b')
plt.title("Customer Signup Trends Over Time")
plt.xlabel("Year")
plt.ylabel("Number of Signups")
plt.show()
```

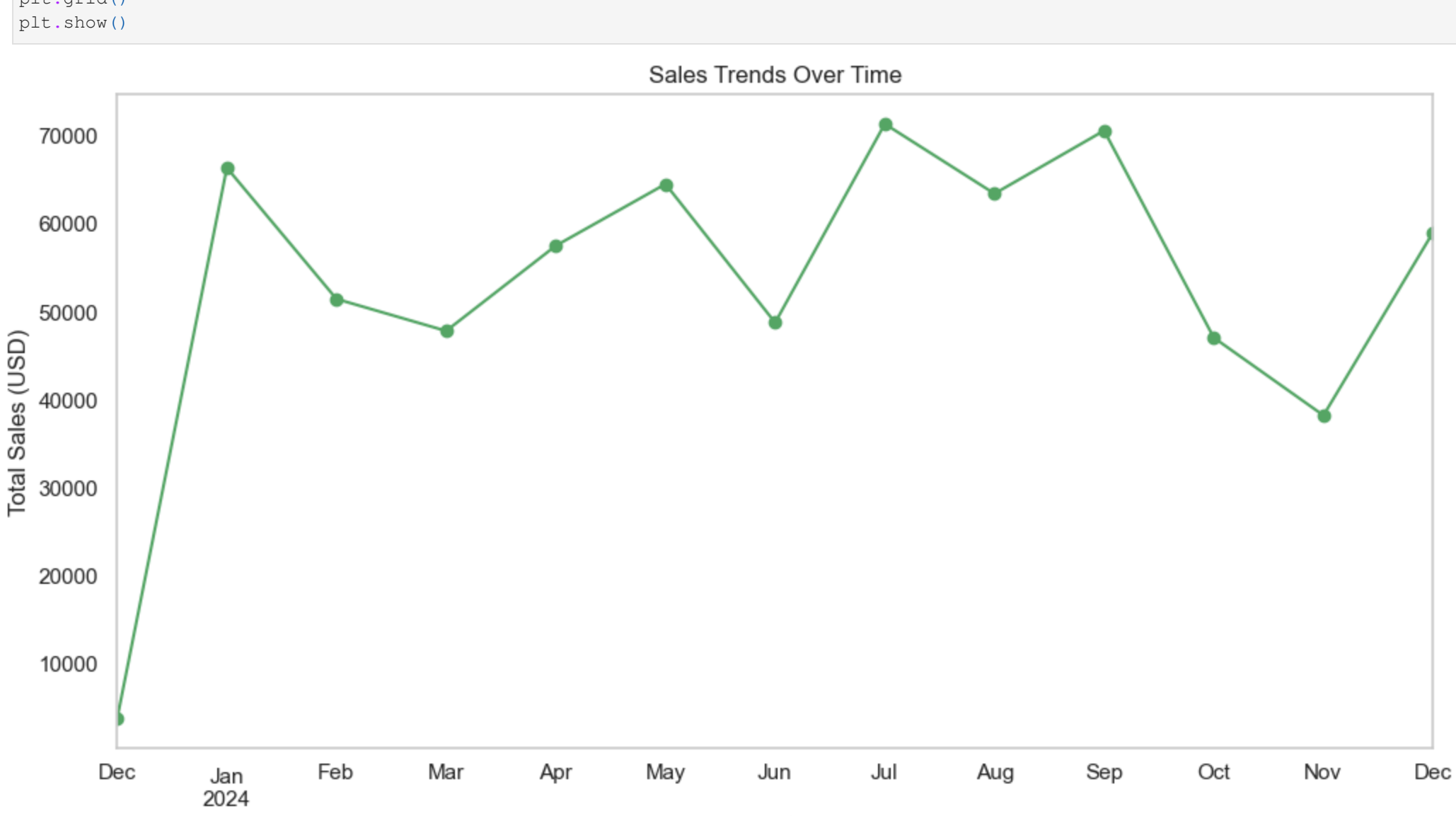


```
In [63]: category_distribution = products['Category'].value_counts()
plt.figure(figsize=(8, 5))
sns.barplot(x=category_distribution.index, y=category_distribution.values, palette="magma")
plt.title("Popular Product Categories")
plt.xlabel("Number of Products")
plt.ylabel("Category")
plt.show()
```

C:\Users\yashu\AppData\Local\Temp\ipykernel_23972\1293399035.py:3: FutureWarning:
Passing 'palette' without assigning 'hue' is deprecated and will be removed in v0.14.0. Assign the 'y' variable to 'hue' and set 'legend=False' for the same effect.
sns.barplot(x=category_distribution.index, y=category_distribution.values, palette="magma")



```
In [65]: transactions['TransactionMonth'] = transactions['TransactionDate'].dt.to_period('M')
sales_trends = transactions.groupby('TransactionMonth')['TotalValue'].sum()
plt.figure(figsize=(12, 6))
sales_trends.plot(kind='line', marker='o', color='g')
plt.title("Sales Trends Over Time")
plt.xlabel("Month")
plt.ylabel("Total Sales (USD)")
plt.grid()
plt.show()
```



```
In [67]: top_products = transactions.groupby('ProductID')['TotalValue'].sum().nlargest(10)
top_products = top_products.reset_index().merge(products, on='ProductID', how='left')
plt.figure(figsize=(10, 6))
sns.barplot(x=top_products['TotalValue'], y=top_products['ProductName'], palette="coolwarm")
plt.title("Top 10 Performing Products by Total Sales")
plt.xlabel("Total Sales (USD)")
plt.ylabel("Product Name")
plt.show()
```

C:\Users\yashu\AppData\Local\Temp\ipykernel_23972\3635085470.py:4: FutureWarning:
Passing 'palette' without assigning 'hue' is deprecated and will be removed in v0.14.0. Assign the 'y' variable to 'hue' and set 'legend=False' for the same effect.
sns.barplot(x=top_products['TotalValue'], y=top_products['ProductName'], palette="coolwarm")

