

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.metrics.pairwise import cosine_similarity

# Load the datasets
customer_data = pd.read_csv('Customers.csv')
product_data = pd.read_csv('Products.csv')
transaction_data = pd.read_csv('Transactions.csv')
# Merge the datasets to get a comprehensive dataset
data = transactions.merge(customers, on='CustomerID').merge(products, on='ProductID')

# Basic data exploration
print(data.info())
print(data.describe())
print(data.isnull().sum())

# Total Revenue by Region
revenue_by_region = data.groupby('Region')['TotalValue'].sum().reset_index()
print(revenue_by_region)

# Top Products by Quantity Sold
top_products = data.groupby('ProductName')['Quantity'].sum().sort_values(ascending=False).head(10).reset_index()
print(top_products)

# Revenue Generation by Region (Visualization)
plt.figure(figsize=(10,6))
sns.barplot(x='Region', y='TotalValue', data=revenue_by_region)
plt.title('Revenue Generation by Region')
plt.xticks(rotation=45)
plt.show()

# Visualize Top Products by Quantity Sold
plt.figure(figsize=(10,6))
sns.barplot(x='ProductName', y='Quantity', data=top_products)
plt.title('Top 10 Products by Quantity Sold')
plt.xticks(rotation=45)
plt.show()

# Customer Signup Trends over Time
data['SignupDate'] = pd.to_datetime(data['SignupDate'])
signup_trends = data.groupby(data['SignupDate'].dt.year)['CustomerID'].nunique().reset_index()
plt.figure(figsize=(10,6))
sns.lineplot(x='SignupDate', y='CustomerID', data=signup_trends)
plt.title('Customer Signup Trends')
plt.xlabel('Year')
plt.ylabel('Number of New Signups')
plt.show()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 13 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_x             1000 non-null   float64  
 7   CustomerName        1000 non-null   object  
 8   Region              1000 non-null   object  
 9   SignupDate          1000 non-null   object  
10   ProductName         1000 non-null   object  
11   Category            1000 non-null   object  
12   Price_y             1000 non-null   float64  
dtypes: float64(3), int64(1), object(9)
memory usage: 101.7+ KB
None

```

```

      Quantity  TotalValue  Price_x  Price_y
count  1000.000000  1000.000000  1000.000000  1000.000000
mean      2.537000    689.995560    272.55407    272.55407
std       1.117981    493.144478    140.73639    140.73639
min       1.000000     16.080000     16.08000     16.08000
25%       2.000000     295.295000    147.95000    147.95000
50%       3.000000     588.880000    299.93000    299.93000
75%       4.000000    1011.660000    404.40000    404.40000
max       4.000000    1991.040000    497.76000    497.76000

```

```

TransactionID  0
CustomerID    0
ProductID     0
TransactionDate 0
Quantity      0
TotalValue    0
Price_x       0
CustomerName  0
Region        0
SignupDate    0
ProductName    0
Category      0
Price_y       0

```

```
dtype: int64
```

```

      Region  TotalValue
0      Asia   152074.97
1    Europe   166254.63
2 North America 152313.40
3 South America 219352.56

```

```

      ProductName  Quantity
0  ActiveWear Smartwatch    100
1  SoundWave Headphones     97
2   HomeSense Desk Lamp     81
3   ActiveWear Rug          79
4   SoundWave Cookbook      78
5   ActiveWear Jacket       76
6  BookWorld Biography      71
7   TechPro T-Shirt        66
8  SoundWave Desk Lamp     64
9   TechPro Textbook       62

```

Revenue Generation by Region

