```
# Import necessary libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the datasets
customers = pd.read_csv("C:/Users/satis/OneDrive/Desktop/793/Customers.csv")
products = pd.read_csv("C:/Users/satis/OneDrive/Desktop/793/Products.csv")
transactions = pd.read_csv("C:/Users/satis/OneDrive/Desktop/793/Transactions (1).csv")

# Check for missing values
print(customers.isnull().sum())
print(products.isnull().sum())
print(transactions.isnull().sum())

# Check for duplicates
print(customers.duplicated().sum())
print(products.duplicated().sum())
print(transactions.duplicated().sum())

# Convert date columns to datetime format
customers['SignupDate'] = pd.to_datetime(customers['SignupDate'])
transactions['TransactionDate'] = pd.to_datetime(transactions['TransactionDate'])
# Exploratory Data Analysis (EDA)
# 1. Customer Distribution by Region
customer_region = customers['Region'].value_counts()
plt.figure(figsize=(8,6))
sns.barplot(x=customer_region.index, y=customer_region.values)
plt.title('Customer Distribution by Region')
plt.xlabel('Region')
plt.ylabel('Number of Customers')
```

```
plt.show()
# 2. Product Price Distribution
plt.figure(figsize=(8,6))
sns.histplot(products['Price'], kde=True)
plt.title('Product Price Distribution')
plt.xlabel('Price (USD)')
plt.ylabel('Frequency')
plt.show()

# 3. Total Sales Over Time
total_sales = transactions.groupby('TransactionDate')['TotalValue'].sum()
plt.figure(figsize=(10,6))
plt.plot(total_sales.index, total_sales.values)
plt.title('Total Sales Over Time')
plt.xlabel('Date')
plt.ylabel('Total Sales (USD)')
plt.show()

# 4. Most Purchased Product Categories
merged = pd.merge(transactions, products, on='ProductID')
category_sales = merged['Category'].value_counts()
plt.figure(figsize=(8,6))
sns.barplot(x=category_sales.index, y=category_sales.values)
plt.title('Most Purchased Product Categories')
plt.xlabel('Category')
plt.ylabel('Number of Transactions')
plt.show()

# 5. Average Transaction Value by Region
transaction_customers = pd.merge(transactions, customers, on='CustomerID')
avg_transaction_value = transaction_customers.groupby('Region')['TotalValue'].mean()
```

```
plt.figure(figsize=(8,6))
sns.barplot(x=avg_transaction_value.index, y=avg_transaction_value.values)
plt.title('Average Transaction Value by Region')
plt.xlabel('Region')
plt.ylabel('Average Transaction Value (USD)')
plt.show()
```