Python Script Containing EDA Code

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

# Load data
customers = pd.read_csv("Customers.csv")
products = pd.read_csv("Products.csv")
transactions = pd.read_csv("Transactions.csv")

# Data cleaning and inspection
print(customers.info())
print(products.info())
print(transactions.info())

# Convert dates to datetime
customers['SignupDate'] = pd.to_datetime(customers['SignupDate'])
transactions['TransactionDate'] = pd.to_datetime(transactions['TransactionDate'])

# Merge datasets
merged_data = transactions.merge(customers, on="CustomerID").merge(products, on="ProductID")

# Summary statistics
print(merged_data.describe())

# EDA visualizations
plt.figure(figsize=(10, 6))
sns.barplot(data=merged_data.groupby('Region')['TotalValue'].sum().reset_index(),
            x='Region', y='TotalValue')
plt.title("Revenue by Region")
plt.show()
```

```
plt.figure(figsize=(10, 6))
sns.countplot(data=products, y='Category', order=products['Category'].value_counts().index)
plt.title("Product Distribution by Category")
plt.show()

# Additional visualizations for insights
# Monthly revenue trend
merged_data['Month'] = merged_data['TransactionDate'].dt.to_period('M')
monthly_revenue = merged_data.groupby('Month')['TotalValue'].sum()

plt.figure(figsize=(12, 6))
monthly_revenue.plot(kind='line', marker='o', title="Monthly Revenue Trend")
plt.xlabel("Month")
plt.ylabel("Total Revenue")
plt.show()

# Business insights
# 1. Identify high-revenue customers
top_customers = merged_data.groupby('CustomerID')['TotalValue'].sum().nlargest(5)

# 2. Popular product categories
popular_categories = merged_data.groupby('Category')['Quantity'].sum().nlargest(5)

# Save EDA results
merged_data.to_csv("MergedData.csv", index=False)
```