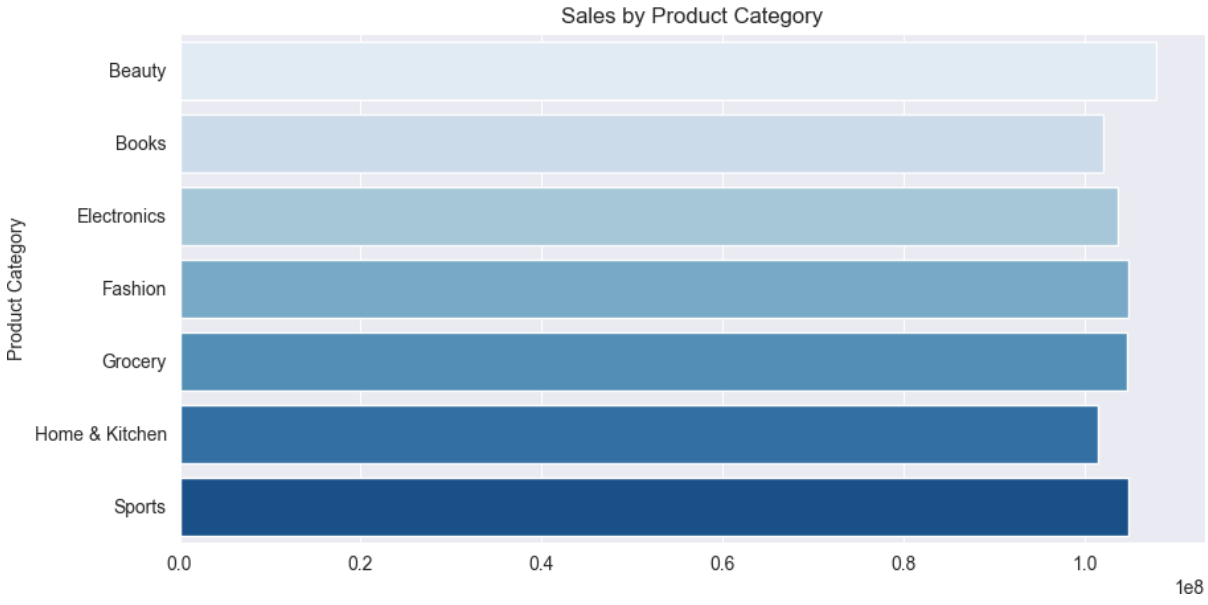
In [25]:

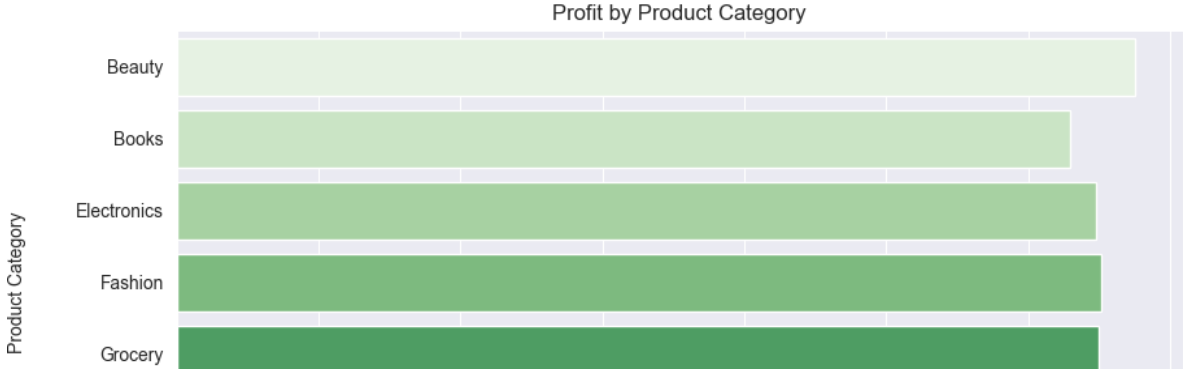
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# LOAD DATA
df = pd.read_csv("E-commerce_sales_data.csv")

print(df.head())
print("\nDataset Info:\n")
print(df.info())

# ---------- BASIC METRICS ----------

print("QUESTION 1: TOTAL SALES")
total_sales = df['Amount'].sum()
print("Total Sales:", total_sales)

print("QUESTION 2: TOTAL PROFIT")
total_profit = df['Profit'].sum()
print("Total Profit:", total_profit)

print("QUESTION 3: OVERALL PROFIT MARGIN")
profit_margin = (total_profit / total_sales) * 100
print("Profit Margin %:", profit_margin)

print("QUESTION 4: AVERAGE ORDER VALUE")
avg_order_value = df['Amount'].mean()
print("Average Order Value:", avg_order_value)

# ---------- CATEGORY ANALYSIS ----------

print("QUESTION 5: SALES BY CATEGORY")
sales_by_category = df.groupby('Product Category')['Amount'].sum()
print(sales_by_category.sort_values(ascending=False))

print("QUESTION 6: TOP 10 PRODUCTS BY SALES")
top_sales_products = df.groupby('Product Name')['Amount'].sum().sort_values(ascending=False).head(10)
print(top_sales_products)

print("QUESTION 7: TOP 10 PRODUCTS BY PROFIT")
top_profit_products = df.groupby('Product Name')['Profit'].sum().sort_values(ascending=False).head(10)
print(top_profit_products)

print("QUESTION 8: LOSS-MAKING PRODUCTS")
loss_products = df.groupby('Product Name')['Profit'].sum()
print(loss_products[loss_products < 0])

# ---------- REGION ANALYSIS ----------

print("QUESTION 9: SALES BY REGION")
region_sales = df.groupby('Region')['Amount'].sum()
print(region_sales.sort_values(ascending=False))

print("QUESTION 10: LOSS-MAKING REGIONS")
region_loss = df.groupby('Region')['Profit'].sum()
print(region_loss[region_loss < 0])

# ---------- DEMOGRAPHIC ----------

print("QUESTION 11: SALES BY GENDER")
print(df.groupby('Gender')['Amount'].sum())

print("QUESTION 12: PROFIT BY GENDER")
print(df.groupby('Gender')['Profit'].sum())

print("QUESTION 13: PAYMENT METHOD USAGE")
print(df['Payment Method'].value_counts())

print("QUESTION 14: PROFIT BY PAYMENT METHOD")
print(df.groupby('Payment Method')['Profit'].sum())

print("QUESTION 15: AVERAGE QUANTITY PER ORDER")
print(df['Quantity'].mean())

# ---------- VISUALIZATION ----------

# 1 Sales by Category
plt.figure(figsize=(10,5))
sns.barplot(x=sales_by_category.values, y=sales_by_category.index, hue=sales_by_category.index, palette="Blues", legend=False)
plt.title("Sales by Product Category")
plt.show()

# 2 Profit by Category
profit_by_category = df.groupby('Product Category')['Profit'].sum()
plt.figure(figsize=(10,5))
sns.barplot(x=profit_by_category.values, y=profit_by_category.index, hue=profit_by_category.index, palette="Greens", legend=Fal
plt.title("Profit by Product Category")
plt.show()

# 3 Sales vs Profit Scatter
plt.figure(figsize=(6,5))
sns.scatterplot(x=df['Amount'], y=df['Profit'])
plt.title("Sales vs Profit Relationship")
plt.show()
```

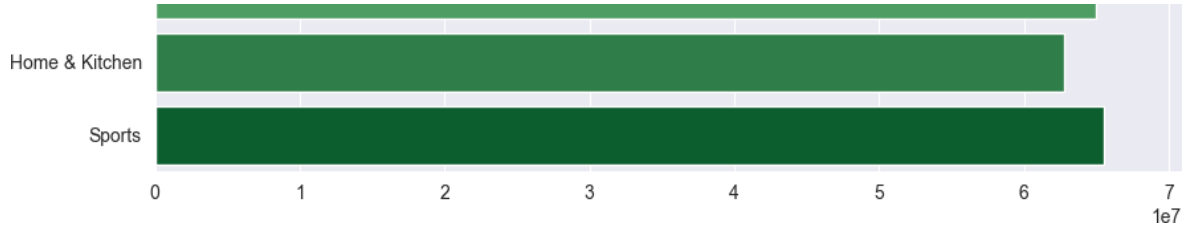
```
# 4 Sales by Region
plt.figure(figsize=(8,5))
sns.barplot(x=region_sales.index, y=region_sales.values, hue=region_sales.index, palette="Paired", legend=False)
plt.title("Sales by Region")
plt.show()

# 5 Correlation Heatmap
plt.figure(figsize=(6,5))
sns.heatmap(df[['Amount','Profit','Purchase Cost','Quantity']].corr(), annot=True, cmap='coolwarm')
plt.title("Correlation Matrix")
plt.show()
```

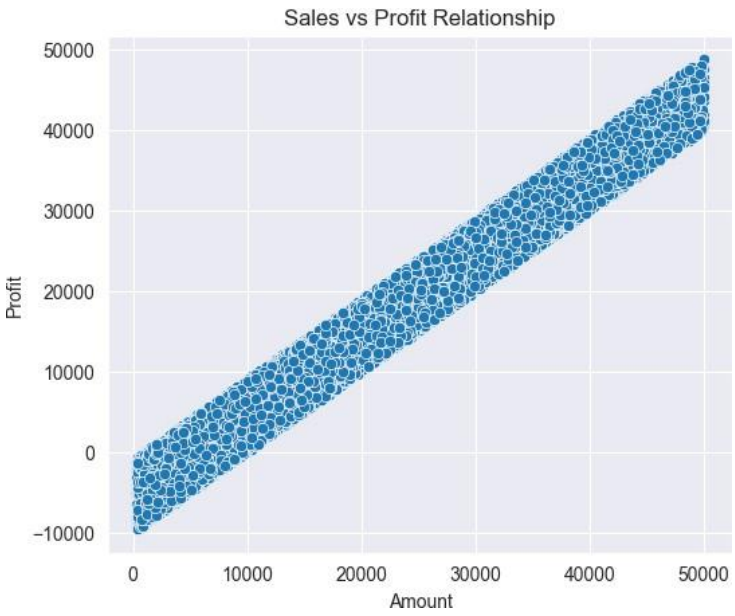
```
   S No Customer Name Customer ID Product Category Product Name    Region  \
0     1         Meena   CUST31122            Books         Rice  Central
1     2         Ankit   CUST22917           Beauty     Lipstick     East
2     3          Amit   CUST70643          Grocery        Shoes    South
3     4         Vikas   CUST48075           Sports        Mixer     West
4     5         Priya   CUST71853           Sports         Rice  Central

        State       City  Gender Payment Method  Quantity  Rate  \
0          UP  Ahmedabad  Female    Net Banking        10  1818
1  Tamil Nadu     Jaipur  Female    Net Banking         8  3735
2  Tamil Nadu  Ahmedabad    Male         Wallet         7   850
3   Rajasthan   Amritsar  Female            COD         4  2220
4          UP  Bengaluru    Male    Credit Card         1  1625

   Purchase Cost  Amount  Profit  Order Date
0           8685   18180    9495   6/21/2023
1           8579   29880   21301  10/14/2022
2           1652    5950    4298    4/1/2024
3           7052    8880    1828   7/16/2023
4           5394    1625   -3769   3/26/2023

Dataset Info:

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50000 entries, 0 to 49999
Data columns (total 16 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   S No              50000 non-null  int64
 1   Customer Name     50000 non-null  object
 2   Customer ID       50000 non-null  object
 3   Product Category  50000 non-null  object
 4   Product Name      50000 non-null  object
 5   Region            50000 non-null  object
 6   State             50000 non-null  object
 7   City              50000 non-null  object
 8   Gender            50000 non-null  object
 9   Payment Method    50000 non-null  object
 10  Quantity          50000 non-null  int64
 11  Rate              50000 non-null  int64
 12  Purchase Cost     50000 non-null  int64
 13  Amount            50000 non-null  int64
 14  Profit            50000 non-null  int64
 15  Order Date        50000 non-null  object
dtypes: int64(6), object(10)
memory usage: 6.1+ MB
None
QUESTION 1: TOTAL SALES
Total Sales: 729148839
QUESTION 2: TOTAL PROFIT
Total Profit: 453696288
QUESTION 3: OVERALL PROFIT MARGIN
Profit Margin %: 62.22272651798051
QUESTION 4: AVERAGE ORDER VALUE
Average Order Value: 14582.97678
QUESTION 5: SALES BY CATEGORY
Product Category
Beauty            107803755
Sports            104810777
Fashion           104784725
Grocery           104596304
Electronics       103686188
Books             102083929
Home & Kitchen    101383161
Name: Amount, dtype: int64
QUESTION 6: TOP 10 PRODUCTS BY SALES
Product Name
Blender       68609147
Perfume       68119344
Jacket        66735026
Headphones    66052307
Shoes         66028552
Bat           65998121
Notebook      65870157
Mixer         65769317
Mobile        65488308
Rice          65252505
Name: Amount, dtype: int64
QUESTION 7: TOP 10 PRODUCTS BY PROFIT
Product Name
Blender       43112800
Perfume       42976952
```

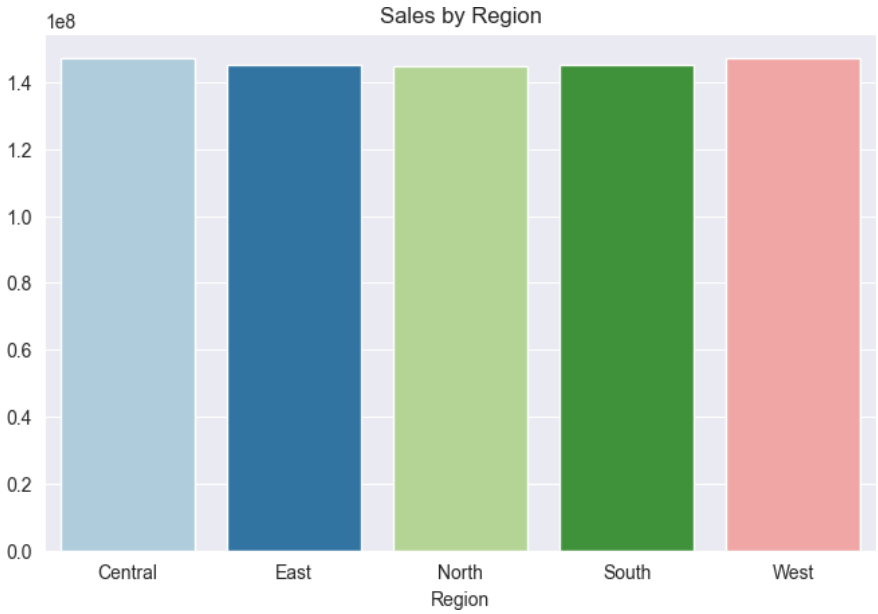
```
Jacket        41735470
Mixer         41490489
Bat           41176481
Shoes         40916257
Lipstick      40853554
Headphones    40715497
Rice          40444432
Mobile        40140915
Name: Profit, dtype: int64
QUESTION 8: LOSS-MAKING PRODUCTS
Series([], Name: Profit, dtype: int64)
QUESTION 9: SALES BY REGION
Region
Central    147152751
West       146988319
South      145082341
East       145061650
North      144863778
Name: Amount, dtype: int64
QUESTION 10: LOSS-MAKING REGIONS
Series([], Name: Profit, dtype: int64)
QUESTION 11: SALES BY GENDER
Gender
Female    363872814
Male      365276025
Name: Amount, dtype: int64
QUESTION 12: PROFIT BY GENDER
Gender
Female    226110938
Male      227585350
Name: Profit, dtype: int64
QUESTION 13: PAYMENT METHOD USAGE
Payment Method
Debit Card     8430
COD            8373
Credit Card    8318
UPI            8314
Net Banking    8293
Wallet         8272
Name: count, dtype: int64
QUESTION 14: PROFIT BY PAYMENT METHOD
Payment Method
COD            74770497
Credit Card    75542095
Debit Card     77067378
Net Banking    77644397
UPI            73669105
Wallet         75002816
Name: Profit, dtype: int64
QUESTION 15: AVERAGE QUANTITY PER ORDER
5.5
```






Home & Kitchen
Sports
0
1
2
3
4
5
6
7
1e7


Sales vs Profit Relationship
Profit
Amount
50000
40000
30000
20000
10000
0
−10000
0
10000
20000
30000
40000
50000


Sales by Region
1e8
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Central
East
North
South
West
Region


Correlation Matrix
Amount
Profit
Purchase Cost
Quantity
1
0.97
-0.0033
0.67
0.97
1
-0.23
0.66
-0.0033
-0.23
1
-0.0081
0.67
0.66
-0.0081
1
Amount
Profit
Purchase Cost
Quantity
1.0
0.8
0.6
0.4
0.2
0.0
−0.2