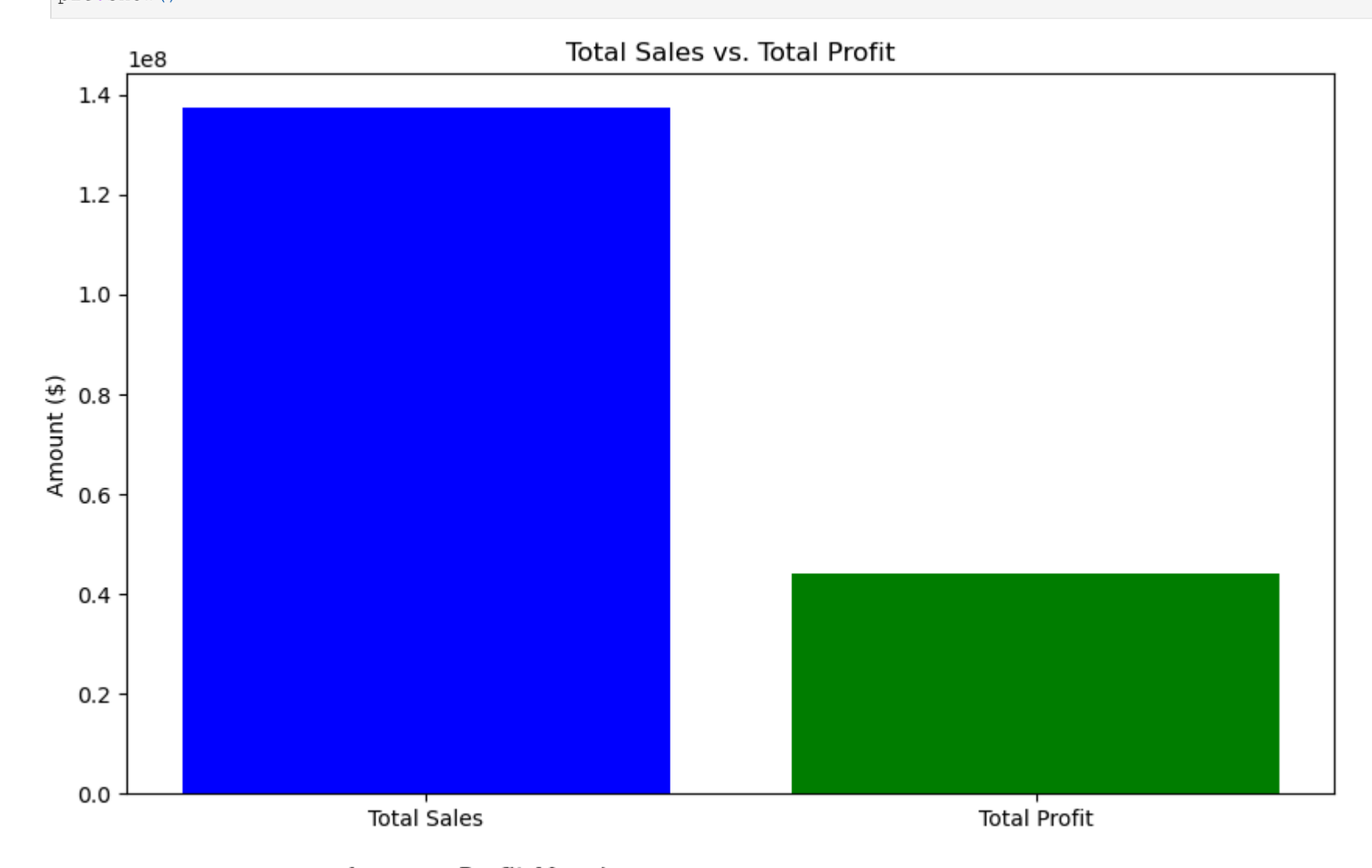
In [1]:

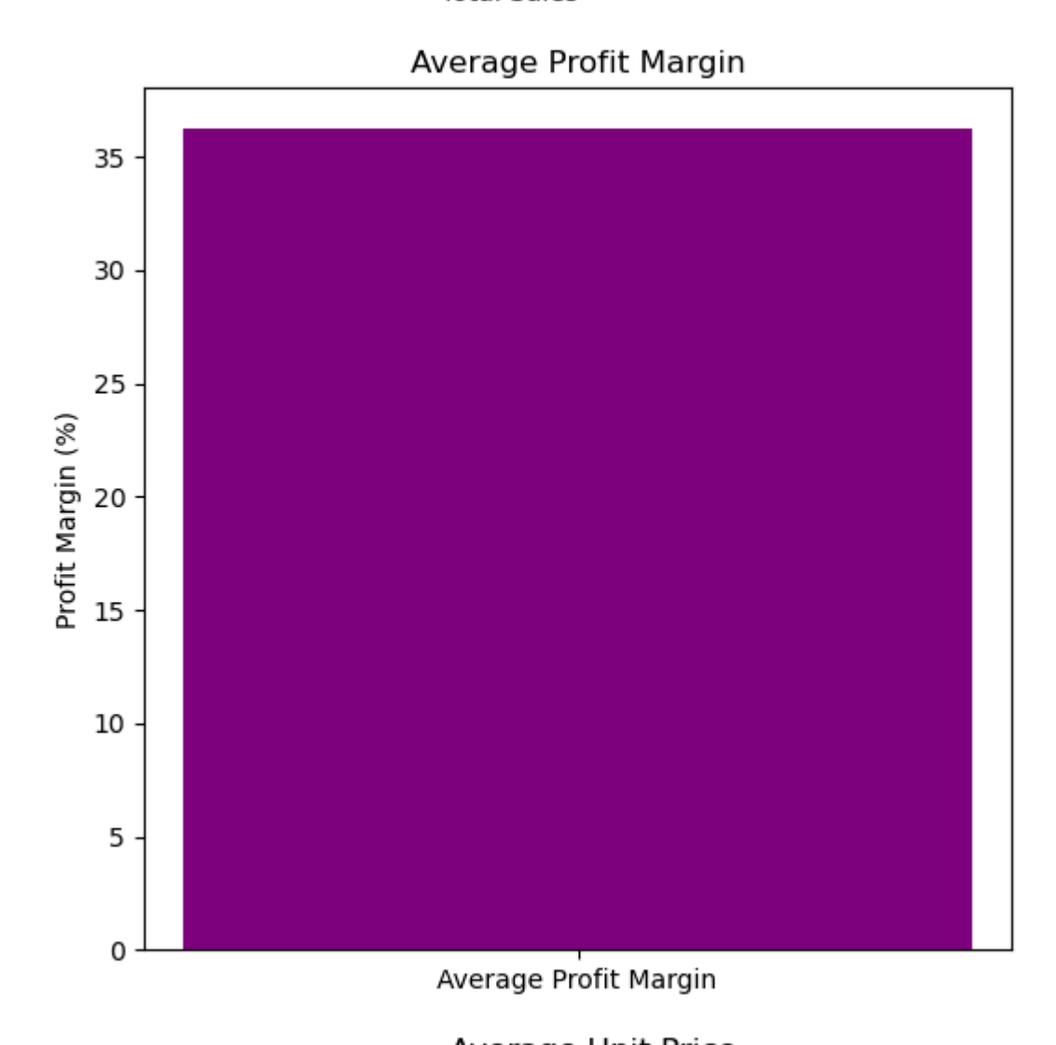
```
import pandas as pd

# Load the dataset
file_path = 'D:\\Internship\\Amazon Data Analytics\\Amazon Sales data.csv'
data = pd.read_csv(file_path)

# Calculate Total Sales
total_sales = data['Total Revenue'].sum()

# Calculate Total Profit
total_profit = data['Total Profit'].sum()

# Average Profit Margin
average_profit_margin = (total_profit / total_sales) * 100

# Average Unit Price
average_unit_price = data['Unit Price'].mean()

# Highest Sales by Order Priority
sales_by_priority = data.groupby('Order Priority')['Total Revenue'].sum()
highest_sales_priority = sales_by_priority.idxmax()

# Bestseller Item Type
sales_by_item_type = data.groupby('Item Type')['Total Revenue'].sum()
bestseller_item_type = sales_by_item_type.idxmax()

# Mode of Channel with Maximum Sales
sales_by_channel = data.groupby('Sales Channel')['Total Revenue'].sum()
max_sales_channel = sales_by_channel.idxmax()

# Yearly Total Sales
data['Order Date'] = pd.to_datetime(data['Order Date'])
data['Year'] = data['Order Date'].dt.year
yearly_sales = data.groupby('Year')['Total Revenue'].sum()

# Analyze Total Sales and Profit by Region
sales_by_region = data.groupby('Region')['Total Revenue'].sum()
profit_by_region = data.groupby('Region')['Total Profit'].sum()

# Print Results
print(f"Total Sales: {total_sales}")
print(f"Total Profit: {total_profit}")
print(f"Average Profit Margin: {average_profit_margin:.2f}%")
print(f"Average Unit Price: {average_unit_price:.2f}")
print(f"Order Priority with Highest Sales: {highest_sales_priority}")
print(f"Bestseller Item Type: {bestseller_item_type}")
print(f"Sales Channel with Maximum Sales: {max_sales_channel}")
print("Yearly Sales:")
print(yearly_sales)
print("Sales by Region:")
print(sales_by_region)
print("Profit by Region:")
print(profit_by_region)
```

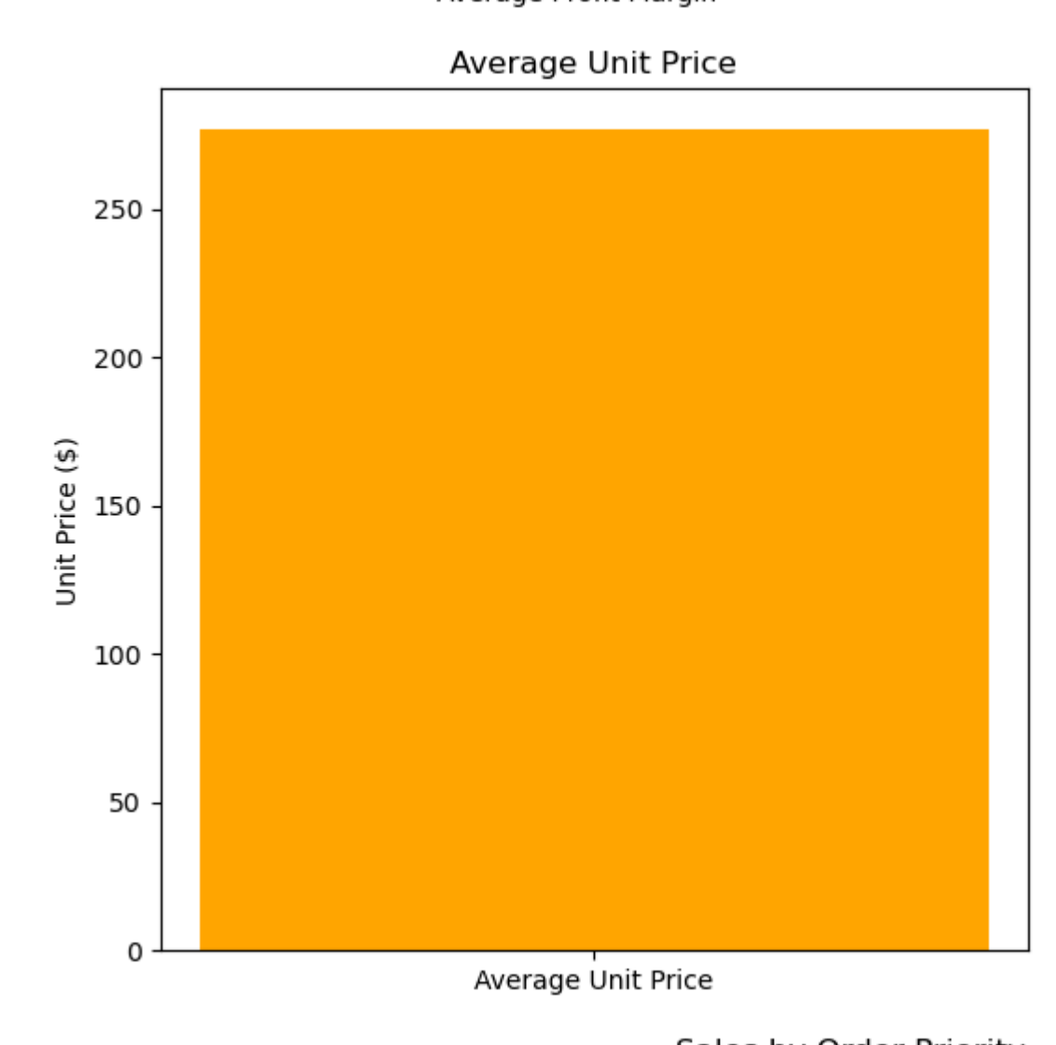
```
Total Sales: 137348768.31
Total Profit: 44168198.39999999
Average Profit Margin: 32.16%
Average Unit Price: 276.76
Order Priority with Highest Sales: H
Bestseller Item Type: Cosmetics
Sales Channel with Maximum Sales: Offline
Yearly Sales:
Year
2010    19186024.92
2011    11129166.07
2012    31898644.52
2013    20330448.66
2014    16630214.43
2015    12427982.86
2016    12372867.22
2017    13373419.63
Name: Total Revenue, dtype: float64
Sales by Region:
Region
Asia                                 21347091.02
Australia and Oceania                14094265.13
Central America and the Caribbean     9170385.49
Europe                               33368932.11
Middle East and North Africa         14052706.58
North America                         5643356.55
Sub-Saharan Africa                   39672031.43
Name: Total Revenue, dtype: float64
Profit by Region:
Region
Asia                                  6113845.87
Australia and Oceania                 4722160.03
Central America and the Caribbean     2846907.85
Europe                               11082938.63
Middle East and North Africa          5761191.86
North America                         1457942.76
Sub-Saharan Africa                   12183211.40
Name: Total Profit, dtype: float64
```

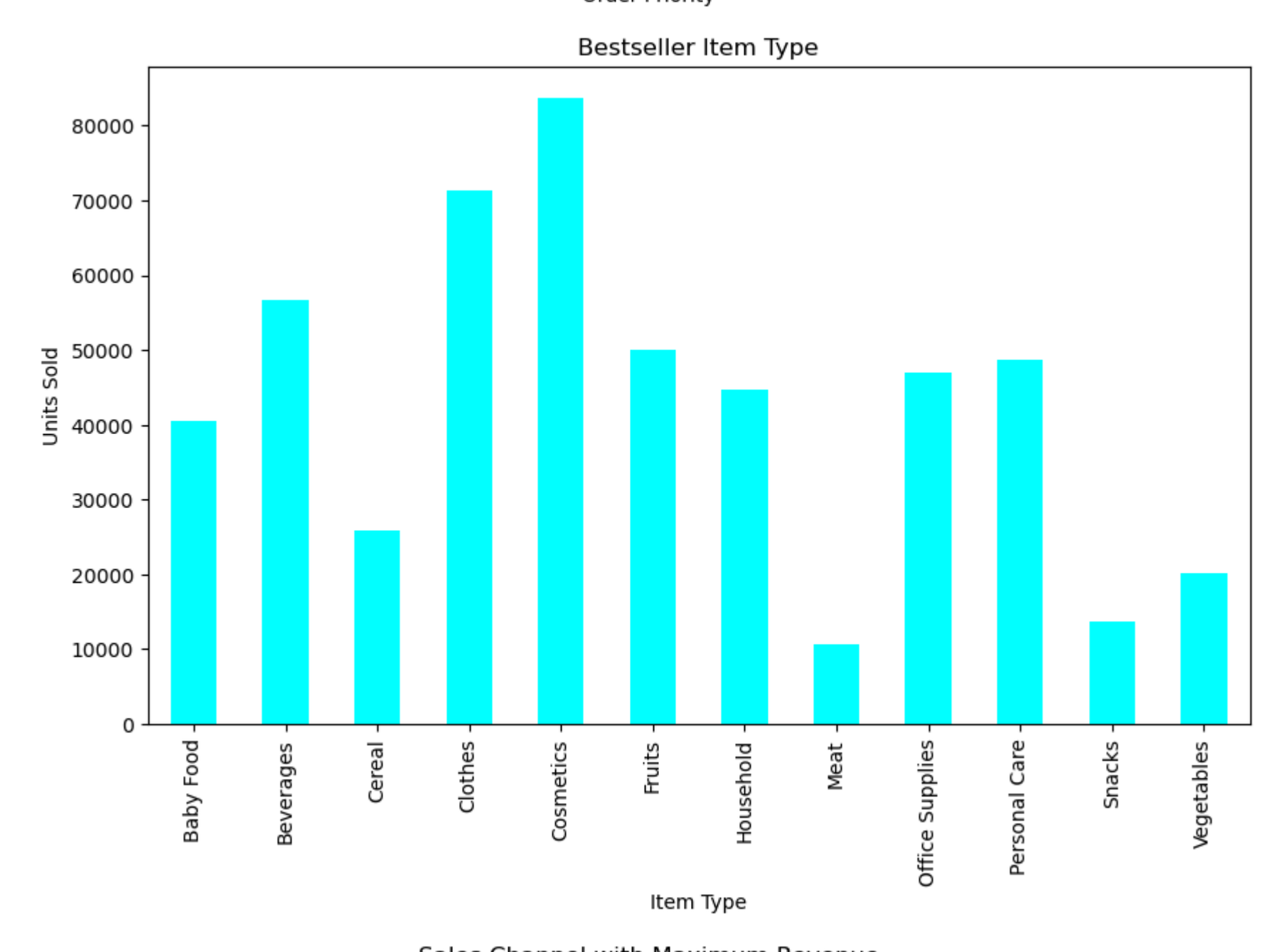
In [3]:

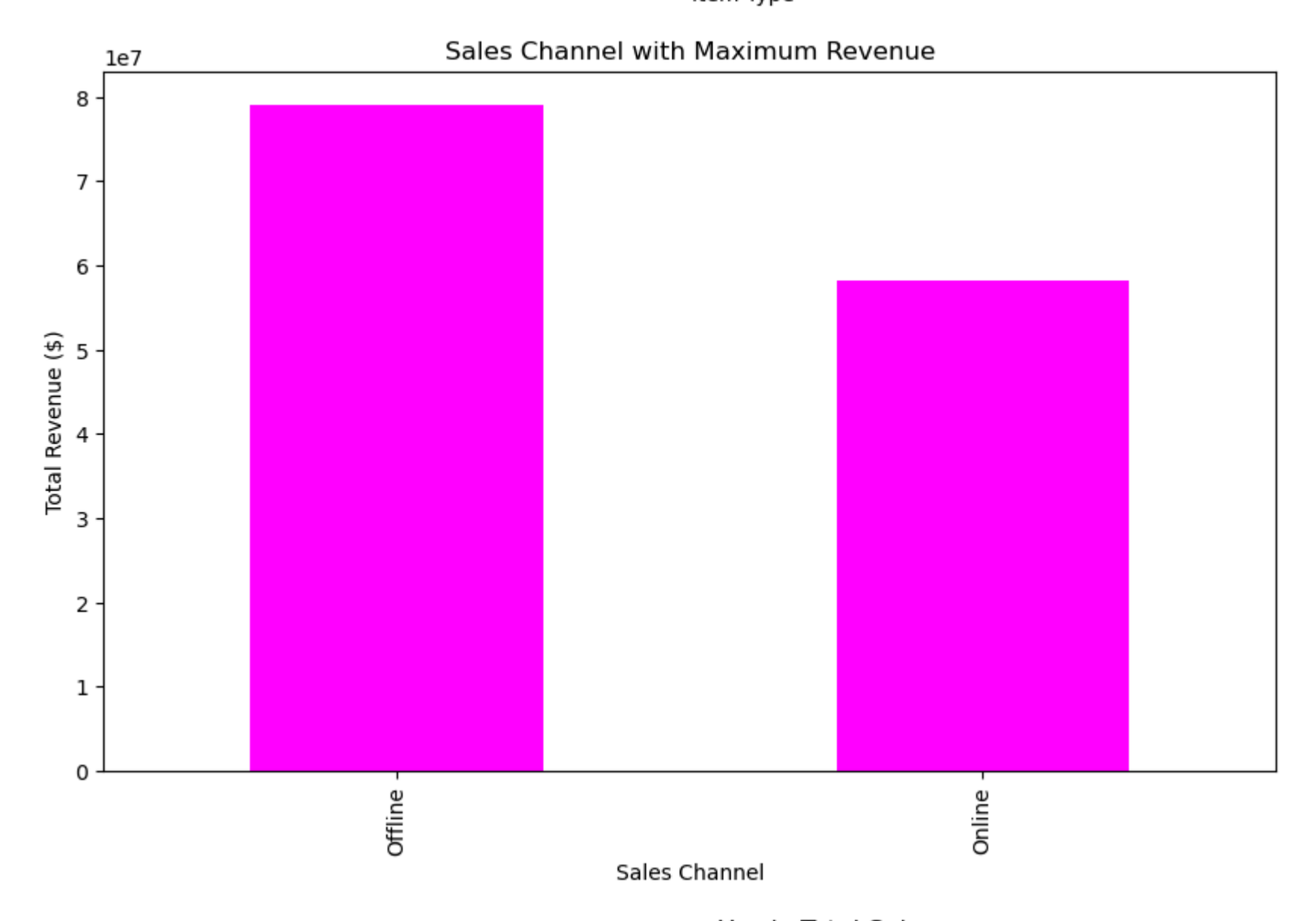
```
import pandas as pd
import matplotlib.pyplot as plt

# Load dataset
file_path = 'D:\\Internship\\Amazon Data Analytics\\Amazon Sales data.csv'
df = pd.read_csv(file_path)

# Calculate Profit Margin
df['Profit Margin'] = (df['Total Profit'] / df['Total Revenue']) * 100

# Calculate metrics
total_sales = df['Total Revenue'].sum()
total_profit = df['Total Profit'].sum()
average_profit_margin = df['Profit Margin'].mean()
average_unit_price = df['Unit Price'].mean()
order_priority_sales = df.groupby('Order Priority')['Total Revenue'].sum().idxmax()
bestseller_item_type = df.groupby('Item Type')['Units Sold'].sum().idxmax()
max_sales_channel = df.groupby('Sales Channel')['Total Revenue'].sum().idxmax()
yearly_total_sales = df.groupby(df['Order Date'].str[:4])['Total Revenue'].sum()
region_sales = df.groupby('Region')['Total Revenue'].sum()

# Plot total sales and profit
plt.figure(figsize=(10, 6))
plt.bar(['Total Sales', 'Total Profit'], [total_sales, total_profit], color=['blue', 'green'])
plt.ylabel('Amount ($)')
plt.title('Total Sales vs. Total Profit')
plt.savefig('D:\\Internship\\sol1\\total_sales_vs_profit.png')
plt.show()

# Plot average profit margin
plt.figure(figsize=(6, 6))
plt.bar(['Average Profit Margin'], [average_profit_margin], color='purple')
plt.ylabel('Profit Margin (%)')
plt.title('Average Profit Margin')
plt.savefig('D:\\Internship\\sol1\\average_profit_margin.png')
plt.show()

# Plot average unit price
plt.figure(figsize=(6, 6))
plt.bar(['Average Unit Price'], [average_unit_price], color='orange')
plt.ylabel('Unit Price ($)')
plt.title('Average Unit Price')
plt.savefig('D:\\Internship\\sol1\\average_unit_price.png')
plt.show()

# Plot sales by order priority
order_priority_sales_df = df.groupby('Order Priority')['Total Revenue'].sum()
plt.figure(figsize=(10, 6))
order_priority_sales_df.plot(kind='bar', color='teal')
plt.ylabel('Total Revenue ($)')
plt.title('Sales by Order Priority')
plt.savefig('D:\\Internship\\sol1\\sales_by_order_priority.png')
plt.show()

# Plot bestseller item type
bestseller_item_type_df = df.groupby('Item Type')['Units Sold'].sum()
plt.figure(figsize=(10, 6))
bestseller_item_type_df.plot(kind='bar', color='cyan')
plt.ylabel('Units Sold')
plt.title('Bestseller Item Type')
plt.savefig('D:\\Internship\\sol1\\bestseller_item_type.png')
plt.show()

# Plot maximum sales channel
max_sales_channel_df = df.groupby('Sales Channel')['Total Revenue'].sum()
plt.figure(figsize=(10, 6))
max_sales_channel_df.plot(kind='bar', color='magenta')
plt.ylabel('Total Revenue ($)')
plt.title('Sales Channel with Maximum Revenue')
plt.savefig('D:\\Internship\\sol1\\max_sales_channel.png')
plt.show()

# Plot yearly total sales
plt.figure(figsize=(12, 8))
yearly_total_sales.plot(kind='line', marker='o', color='red')
plt.ylabel('Total Sales ($)')
plt.title('Yearly Total Sales')
plt.savefig('D:\\Internship\\sol1\\yearly_total_sales.png')
plt.show()

# Plot sales by region
plt.figure(figsize=(12, 8))
region_sales.plot(kind='pie', autopct='%1.1f%%', colors=plt.cm.Paired(range(len(region_sales))))
plt.title('Sales by Region')
plt.ylabel('')
plt.savefig('D:\\Internship\\sol1\\sales_by_region.png')
plt.show()
```

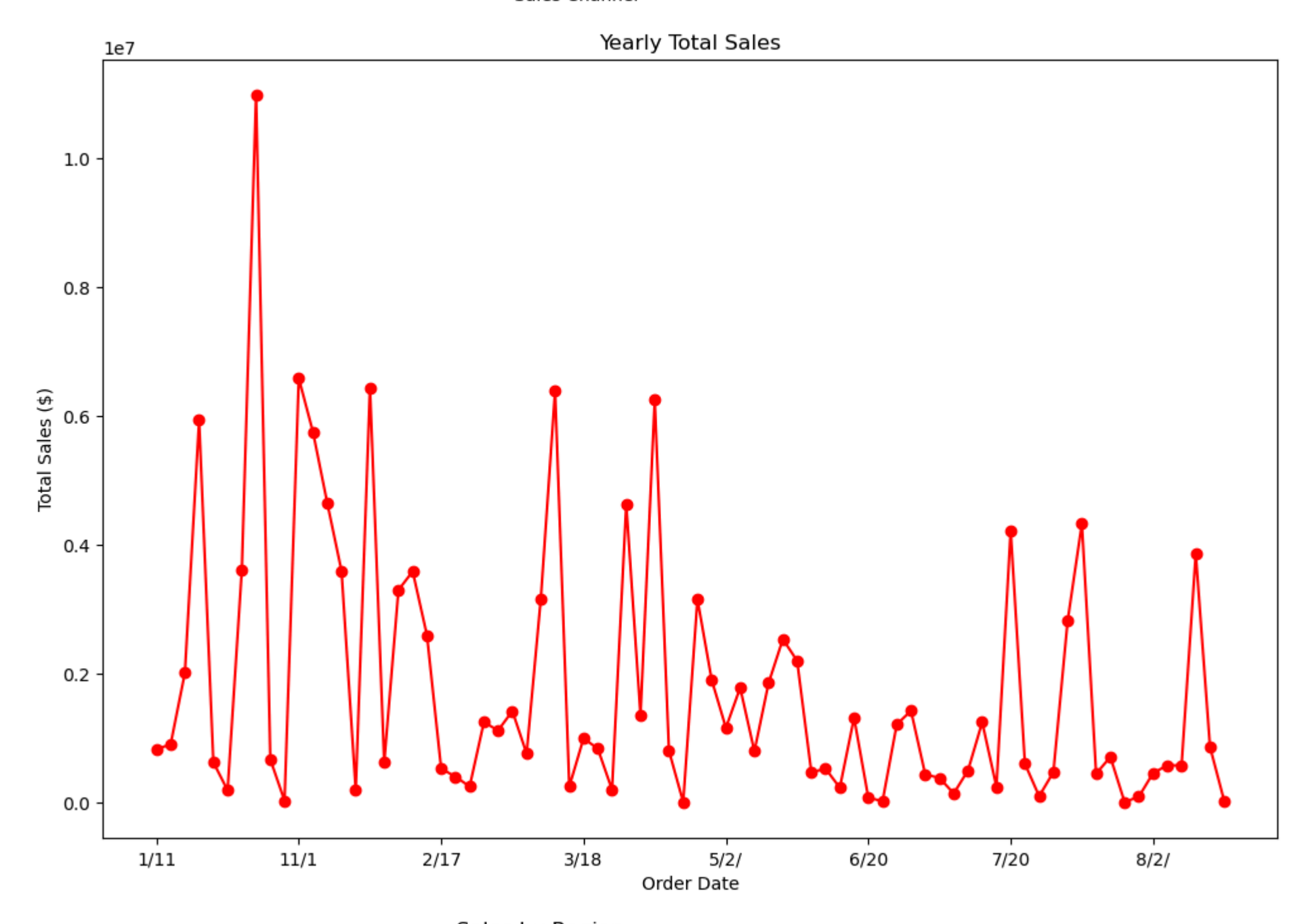


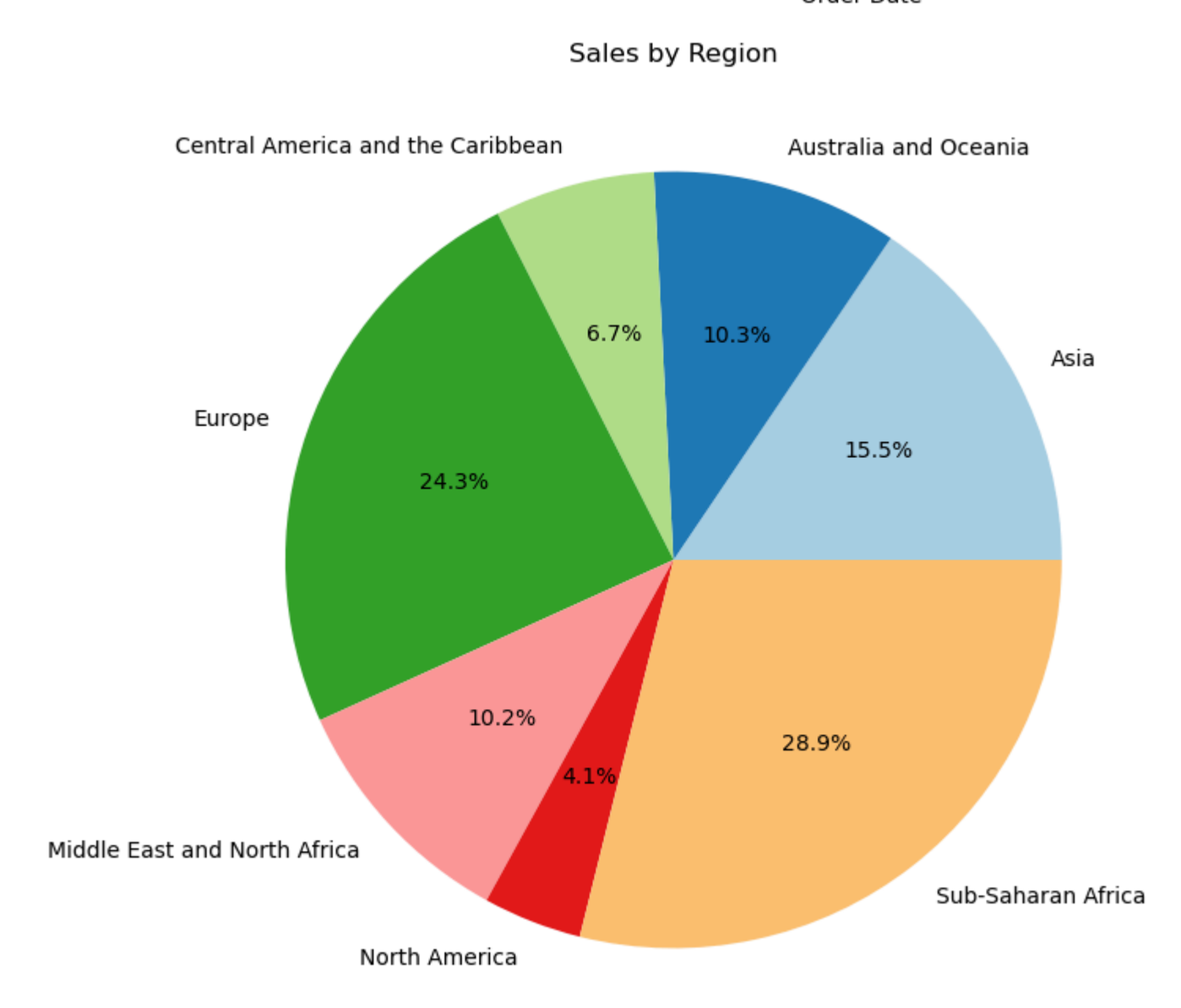








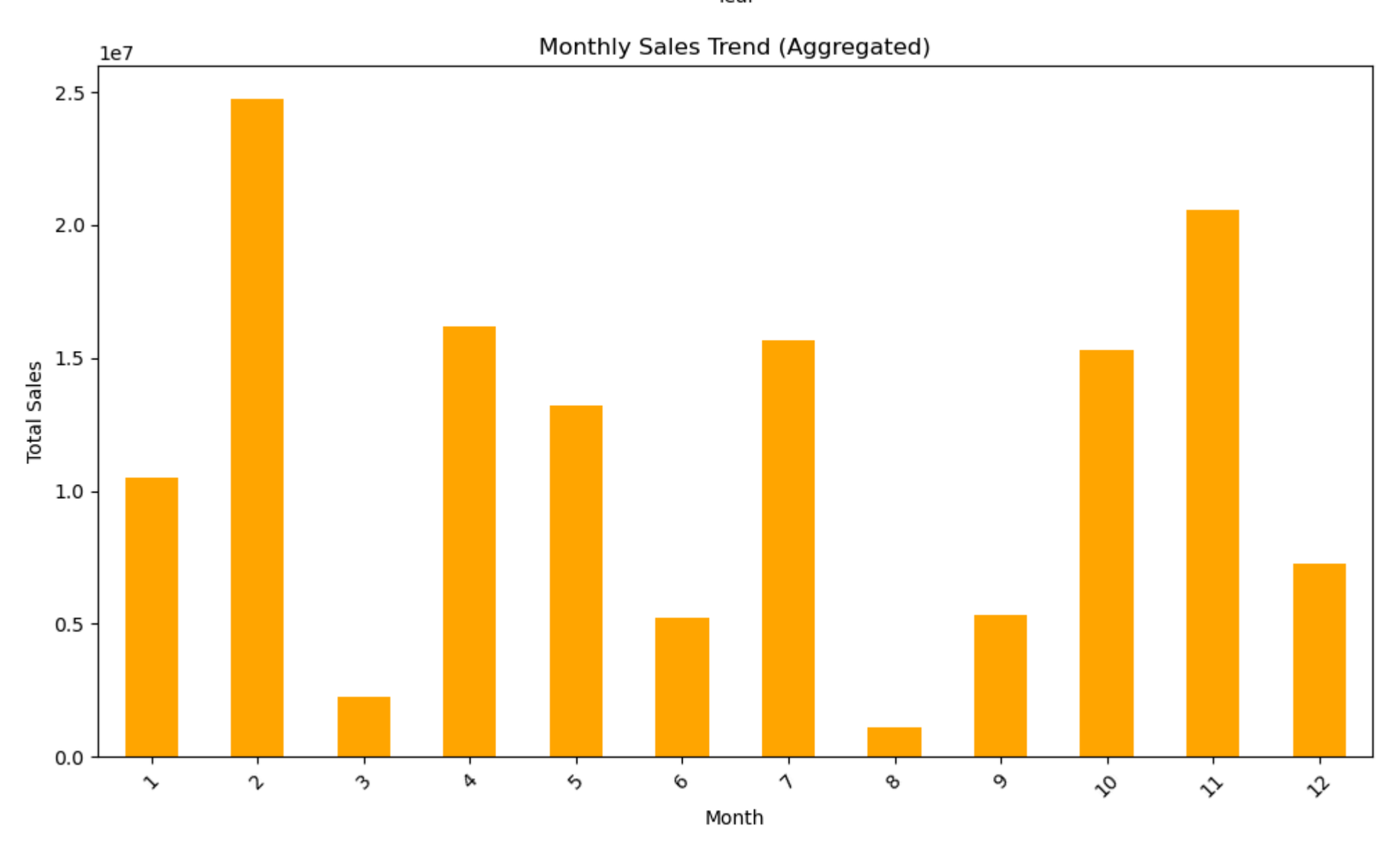


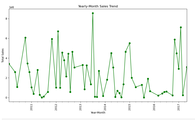


In [4]:

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

# Load the dataset
file_path = 'D:\\Internship\\Amazon Data Analytics\\Amazon Sales data.csv'
data = pd.read_csv(file_path)

# Convert 'Order Date' to datetime
data['Order Date'] = pd.to_datetime(data['Order Date'])

# Extract year, month, and year-month for trend analysis
data['Year'] = data['Order Date'].dt.year
data['Month'] = data['Order Date'].dt.month
data['Year_Month'] = data['Order Date'].dt.to_period('M')

# Calculate Total Sales
total_sales = data['Total Revenue'].sum()

# Calculate Total Profit
total_profit = data['Total Profit'].sum()

# Average Profit Margin
average_profit_margin = (total_profit / total_sales) * 100

# Average Unit Price
average_unit_price = data['Unit Price'].mean()

# Highest Sales by Order Priority
sales_by_priority = data.groupby('Order Priority')['Total Revenue'].sum()
highest_sales_priority = sales_by_priority.idxmax()

# Bestseller Item Type
sales_by_item_type = data.groupby('Item Type')['Total Revenue'].sum()
bestseller_item_type = sales_by_item_type.idxmax()

# Mode of Channel with Maximum Sales
sales_by_channel = data.groupby('Sales Channel')['Total Revenue'].sum()
max_sales_channel = sales_by_channel.idxmax()

# Yearly Total Sales
yearly_sales = data.groupby('Year')['Total Revenue'].sum()

# Sales-trend -> month-wise, year-wise, yearly_month-wise
# Month-wise Sales Trend
monthly_sales = data.groupby('Month')['Total Revenue'].sum()

# Yearly-Month-wise Sales Trend
yearly_month_sales = data.groupby('Year_Month')['Total Revenue'].sum()

# Analyze Total Sales and Profit by Region
sales_by_region = data.groupby('Region')['Total Revenue'].sum()
profit_by_region = data.groupby('Region')['Total Profit'].sum()

# Print Results
print(f"Total Sales: {total_sales}")
print(f"Total Profit: {total_profit}")
print(f"Average Profit Margin: {average_profit_margin:.2f}%")
print(f"Average Unit Price: {average_unit_price:.2f}")
print(f"Order Priority with Highest Sales: {highest_sales_priority}")
print(f"Bestseller Item Type: {bestseller_item_type}")
print(f"Sales Channel with Maximum Sales: {max_sales_channel}")
print("Yearly Sales:")
print(yearly_sales)
print("Sales by Region:")
print(sales_by_region)
print("Profit by Region:")
print(profit_by_region)
print("Monthly Sales Trend:")
print(monthly_sales)
print("Yearly-Month Sales Trend:")
print(yearly_month_sales)

# Plotting the trends

# 1. Yearly Sales Trend
plt.figure(figsize=(10, 6))
yearly_sales.plot(kind='bar', color='skyblue')
plt.title('Yearly Sales Trend')
plt.xlabel('Year')
plt.ylabel('Total Sales')
plt.xticks(rotation=45)
plt.tight_layout()
plt.savefig('D:\\Internship\\sol1\\Yearly_Sales_Trend.png')
plt.show()

# 2. Monthly Sales Trend (Aggregated across all years)
plt.figure(figsize=(10, 6))
monthly_sales.plot(kind='bar', color='orange')
plt.title('Monthly Sales Trend (Aggregated)')
plt.xlabel('Month')
plt.ylabel('Total Sales')
plt.xticks(rotation=45)
plt.tight_layout()
plt.savefig('D:\\Internship\\sol1\\Monthly_Sales_Trend.png')
plt.show()

# 3. Yearly-Month Sales Trend (Sales per Year-Month)
plt.figure(figsize=(10, 6))
yearly_month_sales.plot(kind='line', marker='o', color='green')
plt.title('Yearly-Month Sales Trend')
plt.xlabel('Year-Month')
plt.ylabel('Total Sales')
plt.xticks(rotation=90)
plt.tight_layout()
plt.savefig('D:\\Internship\\sol1\\Yearly_Month_Sales_Trend.png')
plt.show()
```

```
Total Sales: 137348768.31
Total Profit: 44168198.39999999
Average Profit Margin: 32.16%
Average Unit Price: 276.76
Order Priority with Highest Sales: H
Bestseller Item Type: Cosmetics
Sales Channel with Maximum Sales: Offline
Yearly Sales:
Year
2010    19186024.92
2011    11129166.07
2012    31898644.52
2013    20330448.66
2014    16630214.43
2015    12427982.86
2016    12372867.22
2017    13373419.63
Name: Total Revenue, dtype: float64
Sales by Region:
Region
Asia                                 21347091.02
Australia and Oceania                14094265.13
Central America and the Caribbean     9170385.49
Europe                               33368932.11
Middle East and North Africa         14052706.58
North America                         5643356.55
Sub-Saharan Africa                   39672031.43
Name: Total Revenue, dtype: float64
Profit by Region:
Region
Asia                                  6113845.87
Australia and Oceania                 4722160.03
Central America and the Caribbean     2846907.85
Europe                               11082938.63
Middle East and North Africa          5761191.86
North America                         1457942.76
Sub-Saharan Africa                   12183211.40
Name: Total Profit, dtype: float64
Monthly Sales Trend:
Month
1     10482467.12
2     24740517.77
3      2274823.87
4     16187186.33
5     13215739.99
6      5230325.77
7     15669518.50
8      1128164.91
9      5314762.56
10    15287576.61
11    20568222.76
12     7249462.12
Name: Total Revenue, dtype: float64
Yearly-Month Sales Trend:
Year_Month
2010-02    3410661.12
2010-05    2587973.26
2010-06    1082418.40
2010-10    6064933.75
2010-11    3458252.00
2010-12    2581786.39
2011-01    1042225.35
2011-02     387002.20
2011-04    2798046.49
2011-05     272410.45
2011-06      19103.44
2011-07      97040.64
2011-09     574951.92
2011-11    5938385.58
2012-01    1012884.00
2012-02    6707849.42
2012-03     994765.42
2012-04    4556012.38
2012-05    3782781.82
2012-06    2132075.27
2012-07    4445093.92
2012-08     576782.80
2012-09    4648152.72
2012-10    3042246.77
2013-02    3296425.02
2013-03     835759.10
2013-04    3262562.10
2013-06    1352867.40
2013-07    8545511.20
2013-08      89623.98
2013-09      71253.21
2013-10    2702770.40
2013-12     173676.25
2014-02    1819660.25
2014-04    4510578.10
2014-05    3060338.59
2014-06      75591.66
2014-07     688641.85
2014-08     455479.04
2014-09      20404.71
2014-10    1352370.65
2014-11    4647149.58
2015-01    5513227.50
2015-02    2003911.12
2015-04    1059987.26
2015-07    1292409.45
2015-08       6279.09
2015-10    1904138.04
2015-11     648030.40
2016-03     197883.40
2016-05     414371.10
2016-06     568269.60
2016-07     600821.44
2016-10     221117.00
2016-11    5876405.20
2016-12    4493999.48
2017-01    2914130.27
2017-02    7115008.64
2017-03     246415.95
2017-05    3097864.77
Freq: M, Name: Total Revenue, dtype: float64
```




Yearly-Month Sales Trend
1e6
Sales
Year-Month