PPRESENTATION ON

Amazon Sales Analysis

By Kushal Sangwan

PRESENTATION

Table of Contents

Amazon Sales Analysis • Kushal Sangwan

About our data

Region	Country	Item Type	Sales Channel	Order Priority	Order Da	Order ID	Ship Da	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost	Total Profit
Australia and Oceania	Tuvalu	Baby Food	Offline	H	5-28-2010	669165933	6-27-2010	9925	255.28	159.42	2533654	1582243.5	951410.5
Central America and the Caribbean	Grenada	Cereal	Online	C	8-22-2012	963881480	9-15-2012	2804	205.7	117.11	576782.8	328376.44	248406.36
Europe	Russia	Office Supplies	Offline	L	05-02-2014	341417157	05-08-2014	1779	651.21	524.96	1158502.59	933903.84	224598.75
Sub-Saharan Africa	Sao Tome and Principe	Fruits	Online	C	6-20-2014	514321792	07-05-2014	8102	9.33	6.92	75591.66	56065.84	19525.82
Sub-Saharan Africa	Rwanda	Office Supplies	Offline	L	02-01-2013	115456712	02-06-2013	5062	651.21	524.96	3296425.02	2657347.52	639077.5
Australia and Oceania	Solomon Islands	Baby Food	Online	C	02-04-2015	547995746	2-21-2015	2974	255.28	159.42	759202.72	474115.08	285087.64
Sub-Saharan Africa	Angola	Household	Offline	M	4-23-2011	135425221	4-27-2011	4187	668.27	502.54	2798046.49	2104134.98	693911.51
Sub-Saharan Africa	Burkina Faso	Vegetables	Online	H	7-17-2012	871543967	7-27-2012	8082	154.06	90.93	1245112.92	734896.26	510216.66
Sub-Saharan Africa	Republic of the Congo	Personal Care	Offline	M	7-14-2015	770463311	8-25-2015	6070	81.73	56.67	496101.1	343986.9	152114.2
Sub-Saharan Africa	Senegal	Cereal	Online	H	4-18-2014	616607081	5-30-2014	6593	205.7	117.11	1356180.1	772106.23	584073.87
Asia	Kyrgyzstan	Vegetables	Online	H	6-24-2011	814711606	07-12-2011	124	154.06	90.93	19103.44	11275.32	7828.12
Sub-Saharan Africa	Cape Verde	Clothes	Offline	H	08-02-2014	939825713	8-19-2014	4168	109.28	35.84	455479.04	149381.12	306097.92
Asia	Bangladesh	Clothes	Online	L	1-13-2017	187310731	03-01-2017	8263	109.28	35.84	902980.64	296145.92	606834.72
Central America and the Caribbean	Honduras	Household	Offline	H	02-08-2017	522840487	2-13-2017	8974	668.27	502.54	5997054.98	4509793.96	1487261.02
Asia	Mongolia	Personal Care	Offline	C	2-19-2014	832401311	2-23-2014	4901	81.73	56.67	400558.73	277739.67	122819.06
Europe	Bulgaria	Clothes	Online	M	4-23-2012	972292029	06-03-2012	1673	109.28	35.84	182825.44	59960.32	122865.12
Asia	Sri Lanka	Cosmetics	Offline	M	11-19-2016	419123971	12-18-2016	6952	437.2	263.33	3039414.4	1830670.16	1208744.24
Sub-Saharan Africa	Cameroon	Beverages	Offline	C	04-01-2015	519820964	4-18-2015	5430	47.45	31.79	257653.5	172619.7	85033.8
Asia	Turkmenistan	Household	Offline	L	12-30-2010	441619336	1-20-2011	3830	668.27	502.54	2559474.1	1924728.2	634745.9
Australia and Oceania	East Timor	Meat	Online	L	7-31-2012	322067916	09-11-2012	5908	421.89	364.69	2492526.12	2154588.52	337937.6
Europe	Norway	Baby Food	Online	L	5-14-2014	819028031	6-28-2014	7450	255.28	159.42	1901836	1187679	714157
Europe	Portugal	Baby Food	Online	H	7-31-2015	860673511	09-03-2015	1273	255.28	159.42	324971.44	202941.66	122029.78
Central America and the Caribbean	Honduras	Snacks	Online	L	6-30-2016	795490682	7-26-2016	2225	152.58	97.44	339490.5	216804	122686.5
Australia and Oceania	New Zealand	Fruits	Online	H	09-08-2014	142278373	10-04-2014	2187	9.33	6.92	20404.71	15134.04	5270.67
Europe	Moldova	Personal Care	Online	L	05-07-2016	740147912	05-10-2016	5070	81.73	56.67	414371.1	287316.9	127054.2
Europe	France	Cosmetics	Online	H	5-22-2017	898523128	06-05-2017	1815	437.2	263.33	793518	477943.95	315574.05
Australia and Oceania	Kiribati	Fruits	Online	M	10-13-2014	347140347	11-10-2014	5398	9.33	6.92	50363.34	37354.16	13009.18
Sub-Saharan Africa	Mali	Fruits	Online	L	05-07-2010	686048400	05-10-2010	5822	9.33	6.92	54319.26	40288.24	14031.02
Europe	Norway	Beverages	Offline	C	7-18-2014	435608613	7-30-2014	5124	47.45	31.79	243133.8	162891.96	80241.84
Sub-Saharan Africa	The Gambia	Household	Offline	L	5-26-2012	886494815	06-09-2012	2370	668.27	502.54	1583799.9	1191019.8	392780.1
Europe	Switzerland	Cosmetics	Offline	M	9-17-2012	249693334	10-20-2012	8661	437.2	263.33	3786589.2	2280701.13	1505888.07

About our data

- By looking at it we could easily see it's about sales.
- Observing the columns we could easily relate the data with each other, what to look for and what to analyze and visualize.
- Size of our Data: **Rows** (observations): 101
  Columns (variables): 14
- Type of data: Mix-type (Integers, Characters, Float)

Data Cleaning

- First we look at our source of data.
- Our data is stored in our CSV (Comma Separated file).
- First look at null values if present and then remove. If there are columns which are totally empty and don't add anything, it's better to remove them.
- Every column should have certain type of data. Example, columns related to money like **Profit**, **Cost**, **Revenue**, **Items sold** should be integer.
- Then after checking and converting look for duplicates.
- Columns like **Order ID**, **Product ID**, etc. must have unique ID whether it's mix of alphabets or numbers and must not have duplicates.

DATA CLEANING

- For this I used MS Excel.
- We used **Filters** from the **Data** tab.
- To check the null values and duplicates.
- You can use **Pivot table** too here.
- Try to convert data that makes sense and ease in our later work.

Analysis

- For analysis, I used Jupyter Notebook and Python language. Also some libraries to make our work easy and take advantages of them.

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

- Then import our data.

```
# import csv file
df = pd.read_csv('Amazon Sales data.csv', encoding= 'unicode_escape’)
```

- Use methods to see if the data is imported perfectly.

```
df.shape
```

- (100, 14)
- `df.head()`

	Region	Country	Item Type	Sales Channel	Order Priority	Order Date	Order ID	Ship Date	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost	Total Profit
0	Australia and Oceania	Tuvalu	Baby Food	Offline	H	5/28/2010	669165933	6/27/2010	9925	255.28	159.42	2533654.00	1582243.50	951410.50
1	Central America and the Caribbean	Grenada	Cereal	Online	C	8/22/2012	963881480	9/15/2012	2804	205.70	117.11	576782.80	328376.44	248406.36
2	Europe	Russia	Office Supplies	Offline	L	05-02-2014	341417157	05-08-2014	1779	651.21	524.96	1158502.59	933903.84	224598.75
3	Sub-Saharan Africa	Sao Tome and Principe	Fruits	Online	C	6/20/2014	514321792	07-05-2014	8102	9.33	6.92	75591.66	56065.84	19525.82
4	Sub-Saharan Africa	Rwanda	Office Supplies	Offline	L	02-01-2013	115456712	02-06-2013	5062	651.21	524.96	3296425.02	2657347.52	639077.50

ANALYSIS

- If all the output matches to our original data then it means we can work on it.
- Look for another information of our data.
- `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 14 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   Region          100 non-null    object 
 1   Country         100 non-null    object 
 2   Item Type       100 non-null    object 
 3   Sales Channel   100 non-null    object 
 4   Order Priority  100 non-null    object 
 5   Order Date      100 non-null    object 
 6   Order ID        100 non-null    int64  
 7   Ship Date       100 non-null    object 
 8   Units Sold      100 non-null    int64  
 9   Unit Price      100 non-null    float64
 10  Unit Cost       100 non-null    float64
 11  Total Revenue   100 non-null    float64
 12  Total Cost      100 non-null    float64
 13  Total Profit    100 non-null    float64
dtypes: float64(5), int64(2), object(7)
memory usage: 11.1+ KB
```

ANALYSIS

- `df.describe()`
- See the mean, mode, count, etc. to apply mathematical approach and derive insights from it.

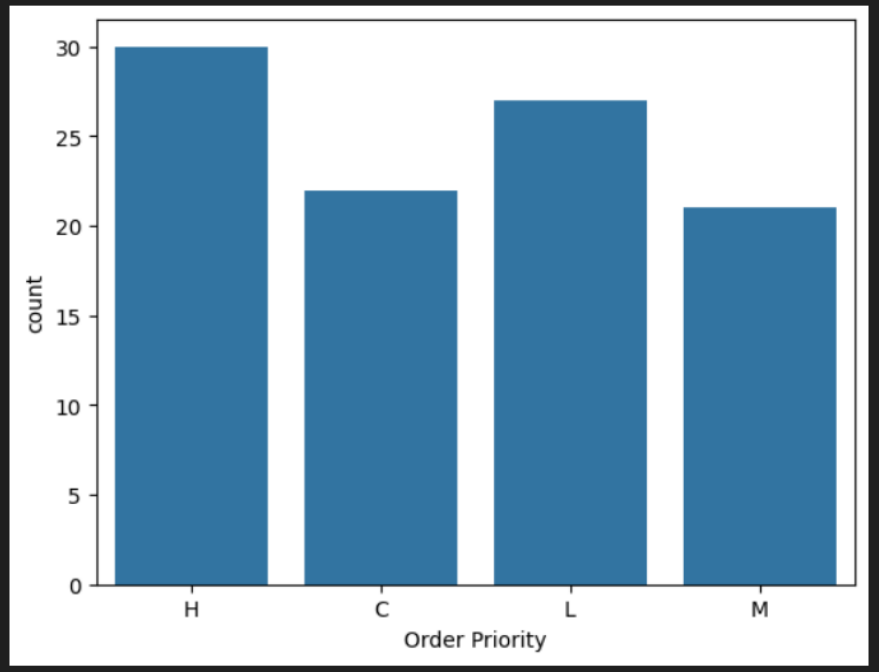
	Order ID	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost	Total Profit
count	1.000000e+02	100.000000	100.000000	100.000000	1.000000e+02	1.000000e+02	1.000000e+02
mean	5.550204e+08	5128.710000	276.761300	191.048000	1.373488e+06	9.318057e+05	4.416820e+05
std	2.606153e+08	2794.484562	235.592241	188.208181	1.460029e+06	1.083938e+06	4.385379e+05
min	1.146066e+08	124.000000	9.330000	6.920000	4.870260e+03	3.612240e+03	1.258020e+03
25%	3.389225e+08	2836.250000	81.730000	35.840000	2.687212e+05	1.688680e+05	1.214436e+05
50%	5.577086e+08	5382.500000	179.880000	107.275000	7.523144e+05	3.635664e+05	2.907680e+05
75%	7.907551e+08	7369.000000	437.200000	263.330000	2.212045e+06	1.613870e+06	6.358288e+05
max	9.940222e+08	9925.000000	668.270000	524.960000	5.997055e+06	4.509794e+06	1.719922e+06

Visualization

- Use `Matplotlib` and `Seaborn` libraries to generate graph to see patterns and derive insights and information from our data.

Order Priority: Mostly orders are of high priority

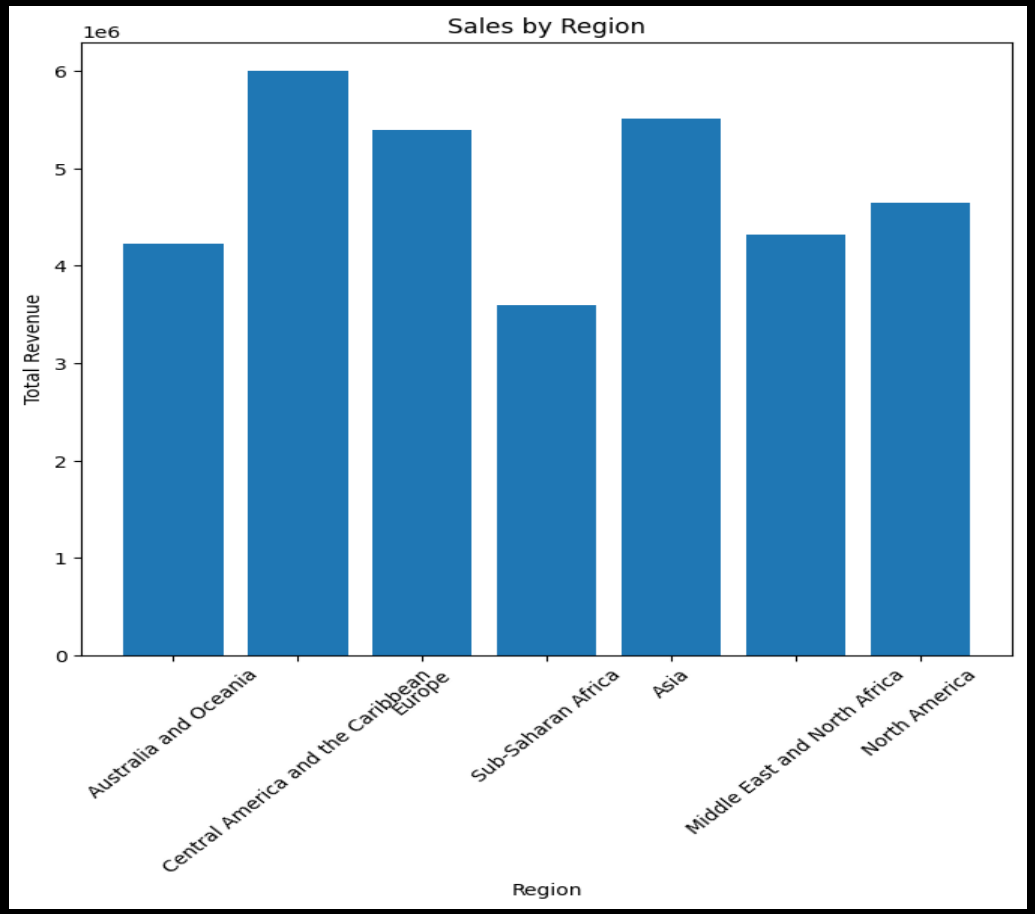
```
ax=sns.countplot(x='Order Priority',data=df)
```



Visualization

- Sales by Region

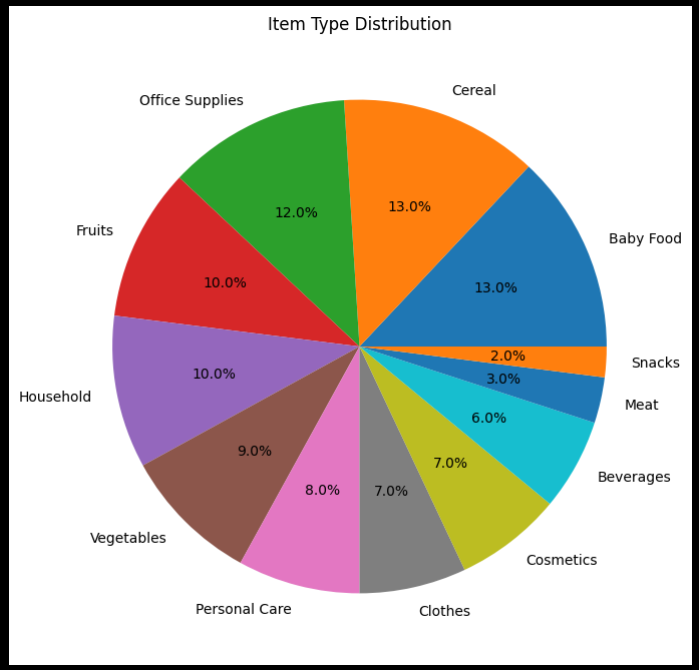
```
plt.figure(figsize=(8, 7))
plt.bar(df['Region'], df['Total Revenue'])
plt.xlabel('Region')
plt.ylabel('Total Revenue')
plt.title('Sales by Region')
plt.xticks(rotation=45)  # Rotate the x-axis labels
plt.show()
```



Visualization

- Item Distribution

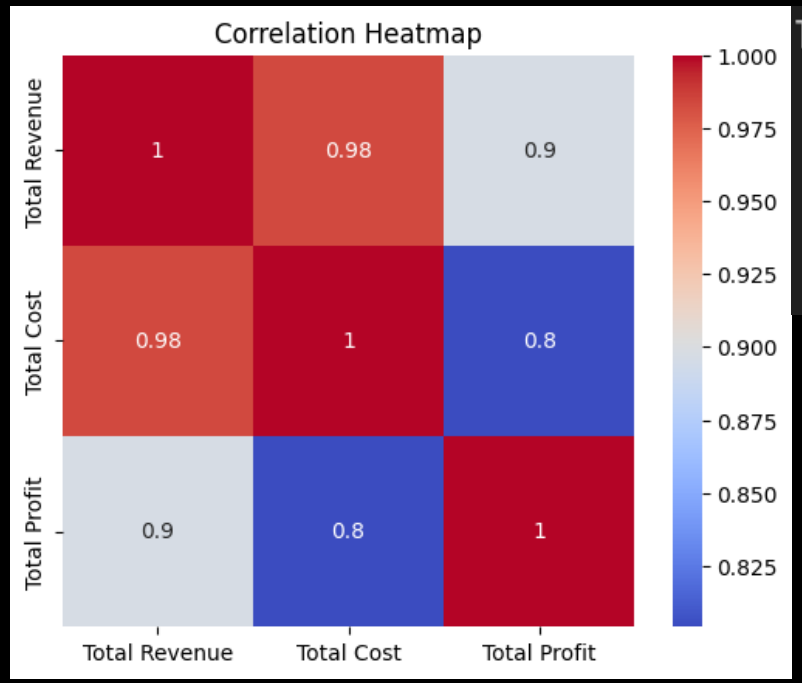
```
plt.figure(figsize=(8, 8))
plt.pie(df['Item Type'].value_counts(), labels=df['Item Type'].unique(),
autopct='%1.1f%%')
plt.title('Item Type Distribution')
plt.show()
```



Visualization

- Correlation Heatmap

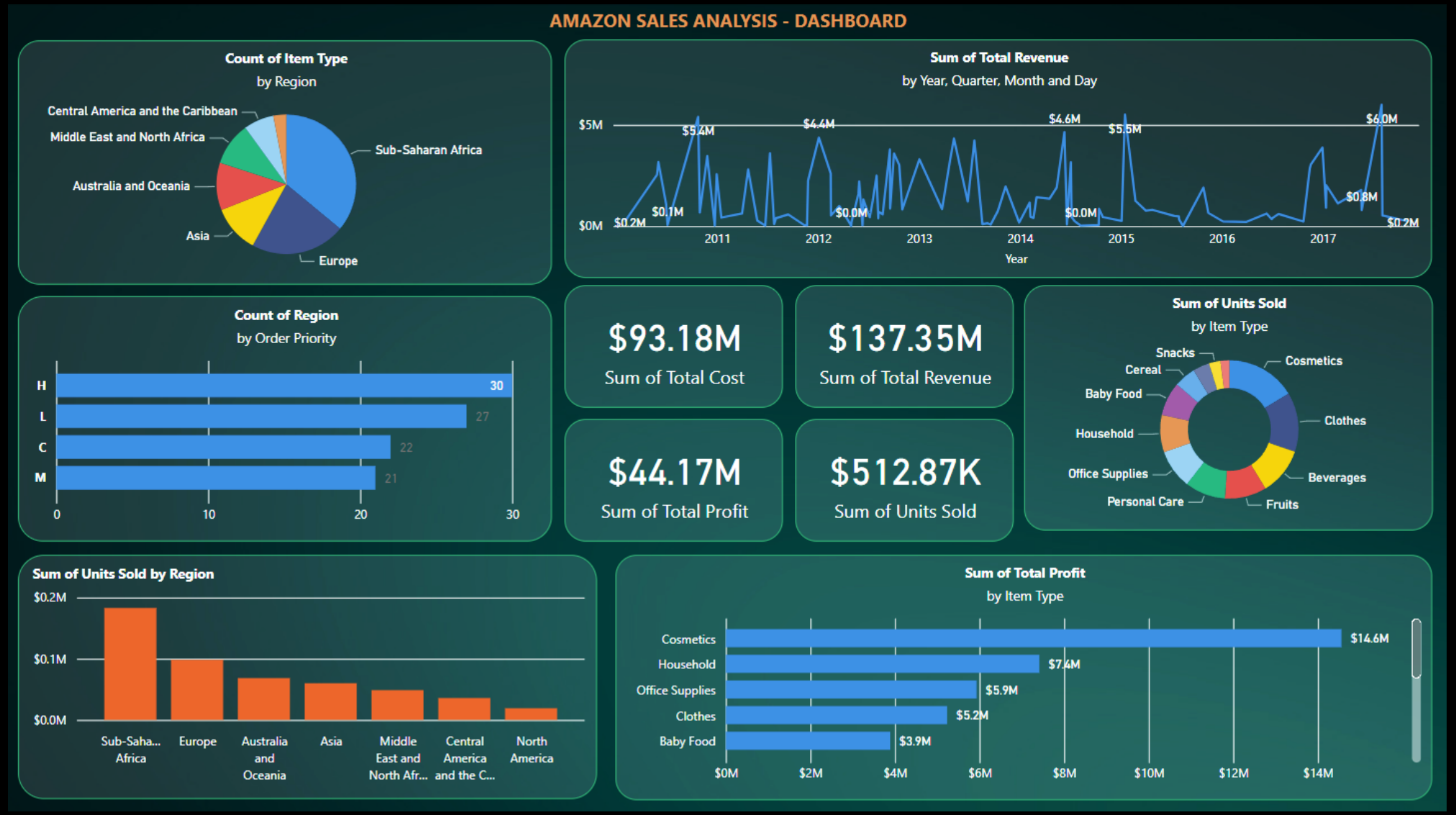
```
corr_matrix = df[['Total Revenue', 'Total Cost', 'Total Profit']].corr()
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', square=True)
plt.title('Correlation Heatmap')
plt.show()
```



The squares adjacent to this diagonal represent correlations between:

- **Total Revenue** and **Total Cost**, marked as *0.98* (a very strong positive relationship).
- **Total Cost** and **Total Profit**, also marked as *0.98* (another strong positive relationship).
- **Total Revenue** and **Total Profit**, marked as *0.9* (a strong positive relationship, slightly less than the other two pairs).

Dashboard (Power BI)



Observations & Insights

- Sum total of Profit created: 44.17 Million
- Total Revenue (the income generated from sales) increases, we expect that Total Cost (the expenses incurred to produce and sell goods) also increases.
- Sum of Total Revenue: 137.35 Million
- Order Priority: Mostly orders are of high priority
- We saw mostly sales are from 'Central America and the Caribbean’.
- Insights:
  Seasonal Trends: Analyze peaks and dips to identify seasonal patterns.
  Strategy Evaluation: Investigate factors behind high-revenue months.

Tools Used

Microsoft Excel

Jupyter Notebook

Power BI

THANK YOU