



AMAZON SALES DATA – ANALYSIS

DETAILED REPORT

MOHAMMAD ZAHID GUROO

UMIP15259

| PROJECT NAME | AMAZON SALES ANALYSIS |
|---------------------------|-----------------------|
| DOMAIN | E-COMMERCE |
| DIFFICULTY | ADVANCED |
| PROGRAMMING LANGUAGE USED | PYTHON |
| TECHNOLOGY | DATA ANALYSIS |

INTRODUCTION

The primary goal of this project is to analyze Amazon's sales data to gain insights into sales performance across different regions, countries, item types, and other factors. The analysis aims to identify key performance indicators (KPIs) that can drive strategic decision-making. Objective Understand the distribution of sales across regions, countries, and item types. Analyze the impact of different sales channels and order priorities on sales performance. Examine the relationship between units sold, pricing strategies, and profitability. Identify the time patterns in order and shipping dates. Provide actionable insights to improve sales performance and profitability

Variables In The Data-Set

Region: The geographical region where the sale was made.

Country: The specific country within the region.

Item Type: The category of the item sold (e.g., cosmetics, clothing).

Sales Channel: Whether the sale was made online or offline.

Order Priority: The priority assigned to the order (e.g., H- high , M- medium , L- low).

Order Date: The date the order was placed.

Order ID: A unique identifier for each order.

Ship Date: The date the order was shipped.

Units Sold: The number of units sold in the order.

Unit Price: The price per unit of the item.

Unit Cost: The cost per unit of the item.

Total Revenue: The total revenue generated from the order.

Total Cost: The total cost incurred for the order.

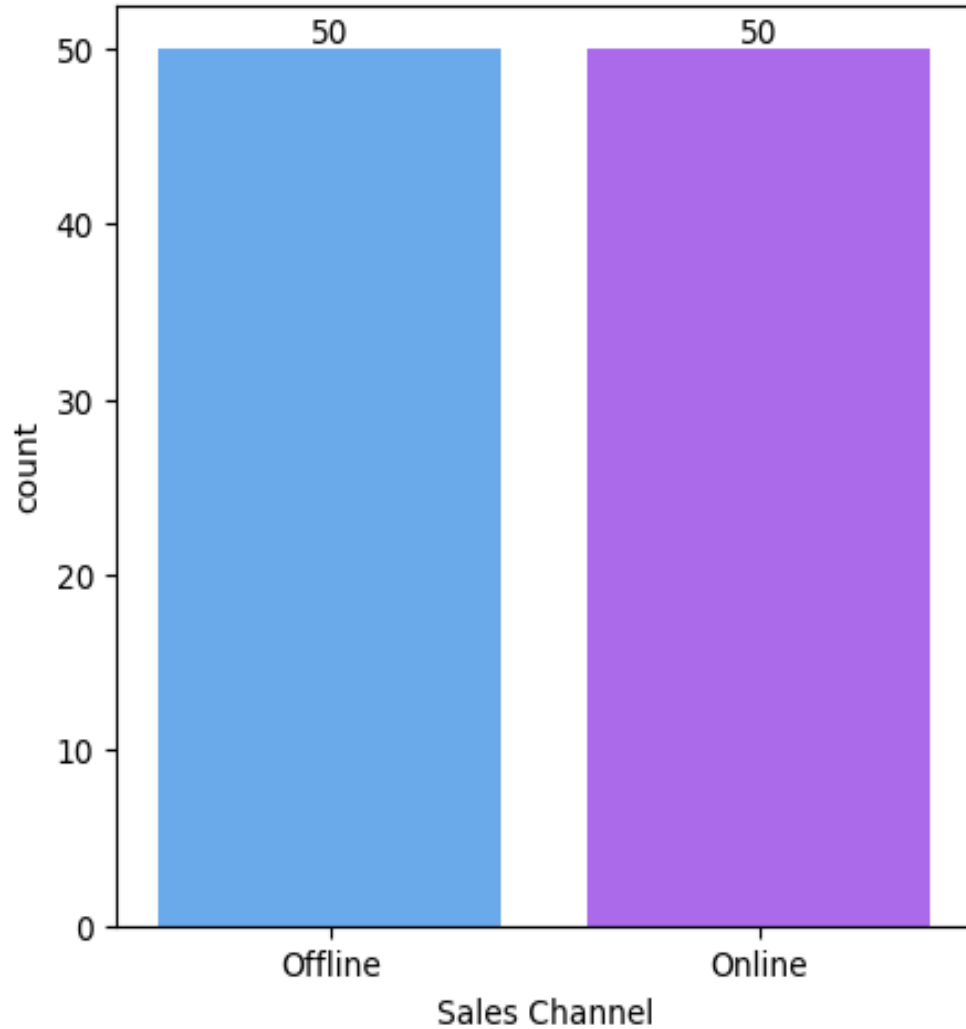
Total Profit: The profit made from the order.

Data Preprocessing:

- Importing necessary libraries for data analysis such as: Pandas, Numpy, Matplotlib & Seaborn etc.
- Using `pd.read_csv()` function stores the data in pandas dataframe named `data`. Using `data.columns` showing columns present in dataframe.
- `info()` function show basic information of dataframe like null value count of each column and their data type
- Changing the data type of different column for model training and analysis
- Using `describe` function on dataframe for getting basic stats of numerical dataset
- Adding extra column to dataframe which contain only month, year.
- Using `isnull().sum()` checking out total null value in all the column of dataframe

Sales Via Sales Channel

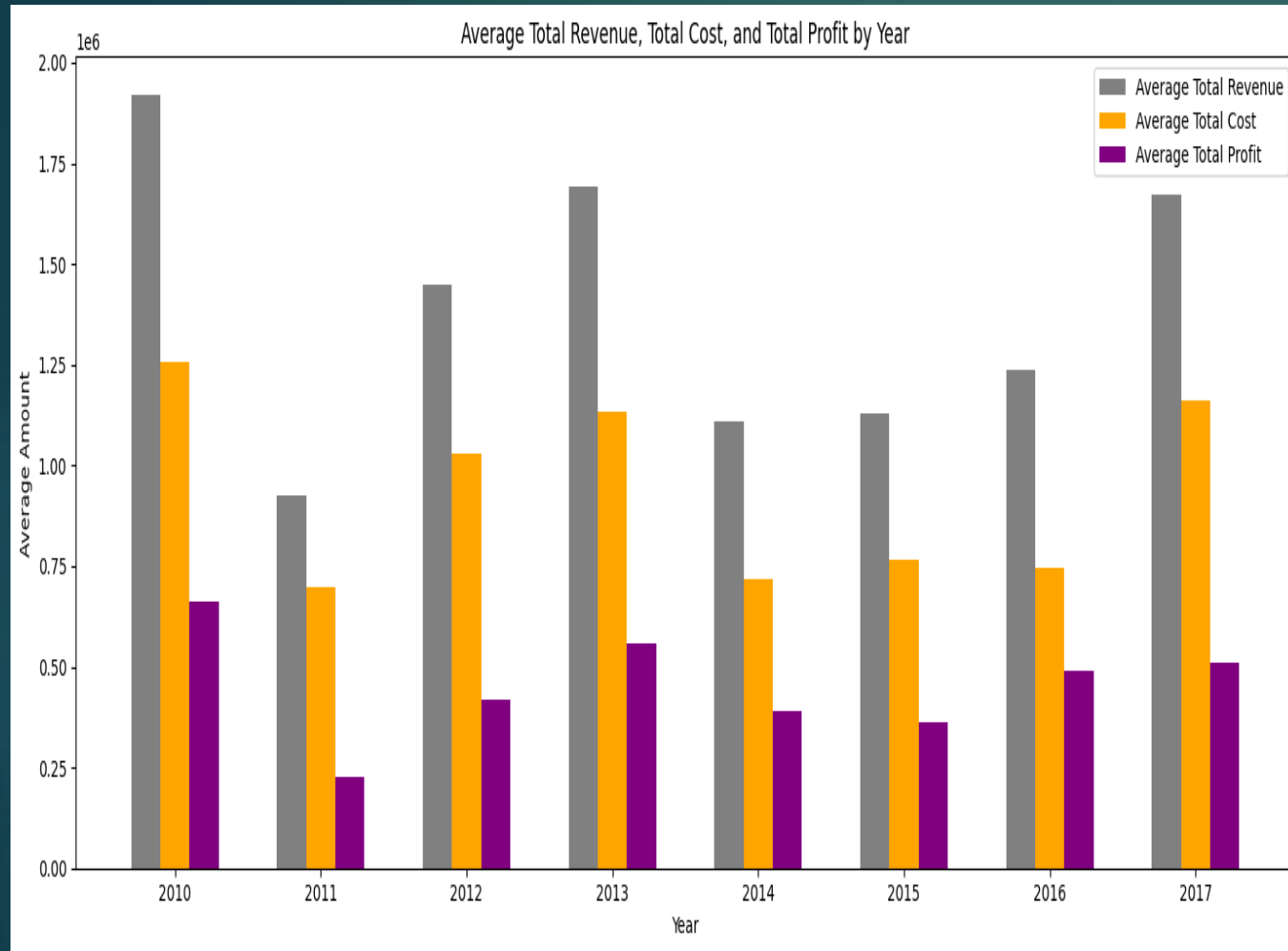
Countplot Displaying the Offline Vs Online Channel Sales



--Insights--

- From the graph we observe that there's been an equal amount of sales via the both channels meaning according to the dataset no channel has been superior compared to other.

Average Revenue, Cost & Profit Over The Years

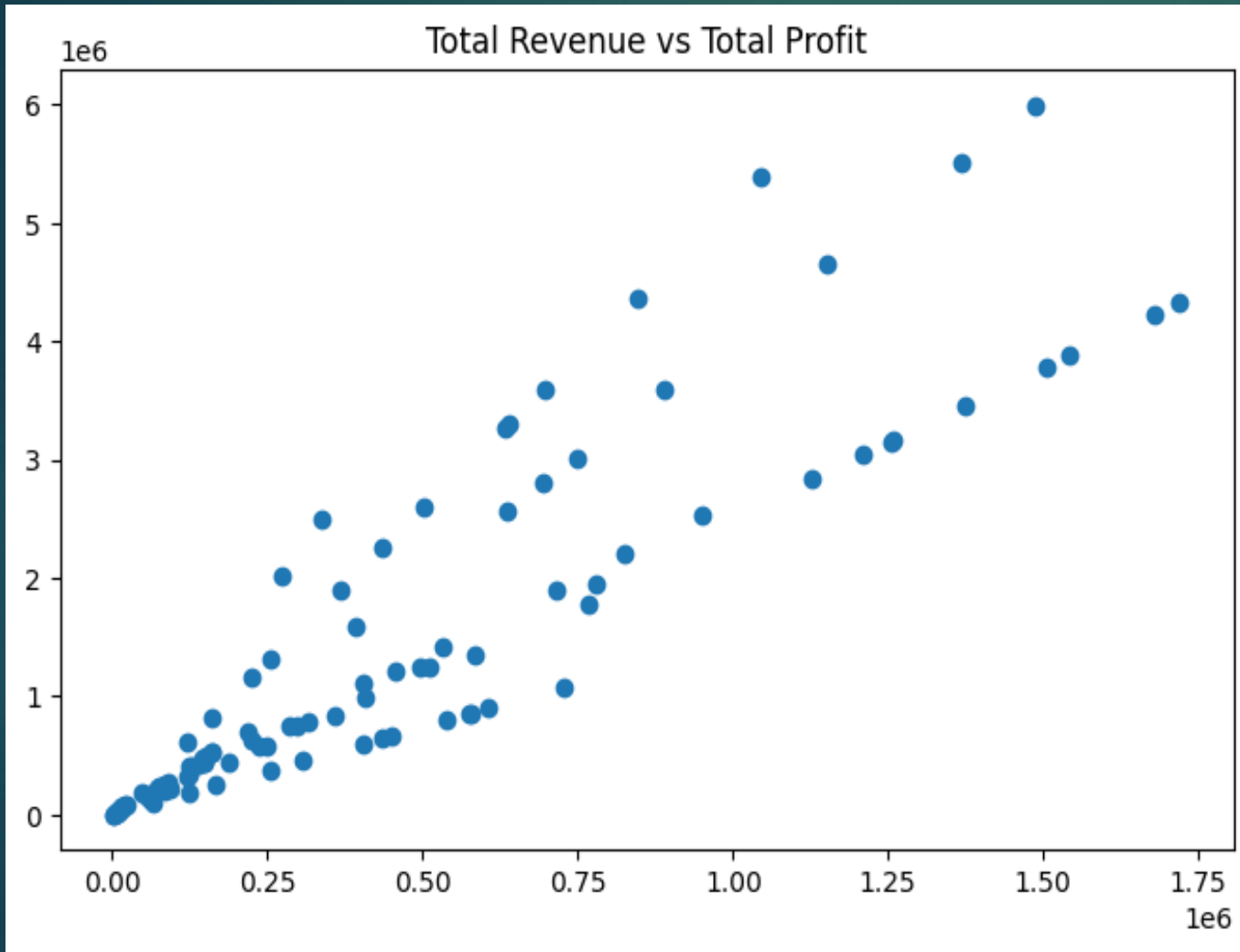


.-Insights—

From the graph we can observe that:

- **2010:** - Total Cost (Gray): The highest among all years, reaching close to 2 million. - Total Revenue (Orange): Also relatively high, indicating significant business activity. - Total Profit (Purple): Despite the high revenue and cost, profit is moderate, suggesting high operational costs.
- **2011:** - Significant drop in revenue, cost, and profit compared to 2010. - Costs are nearly equal to revenue, leaving very little profit.
- **2012:** - Revenue and cost have increased from 2011 levels. - Profit remains low, indicating that costs continue to rise in tandem with revenue.
- **2013:** - A noticeable drop across all metrics, possibly indicating a challenging year. - Profit is still minimal, reflecting low efficiency or high operational costs.
- **2014-2015:** - A gradual increase in revenue, cost, and profit, showing a recovery trend. - Profit margins remain low, but they are gradually improving.
- **2016-2017:** - Continued growth in revenue and costs, with 2017 showing one of the highest revenues and profits. - Although the profit is increasing, it is still relatively small compared to total revenue, highlighting high costs.

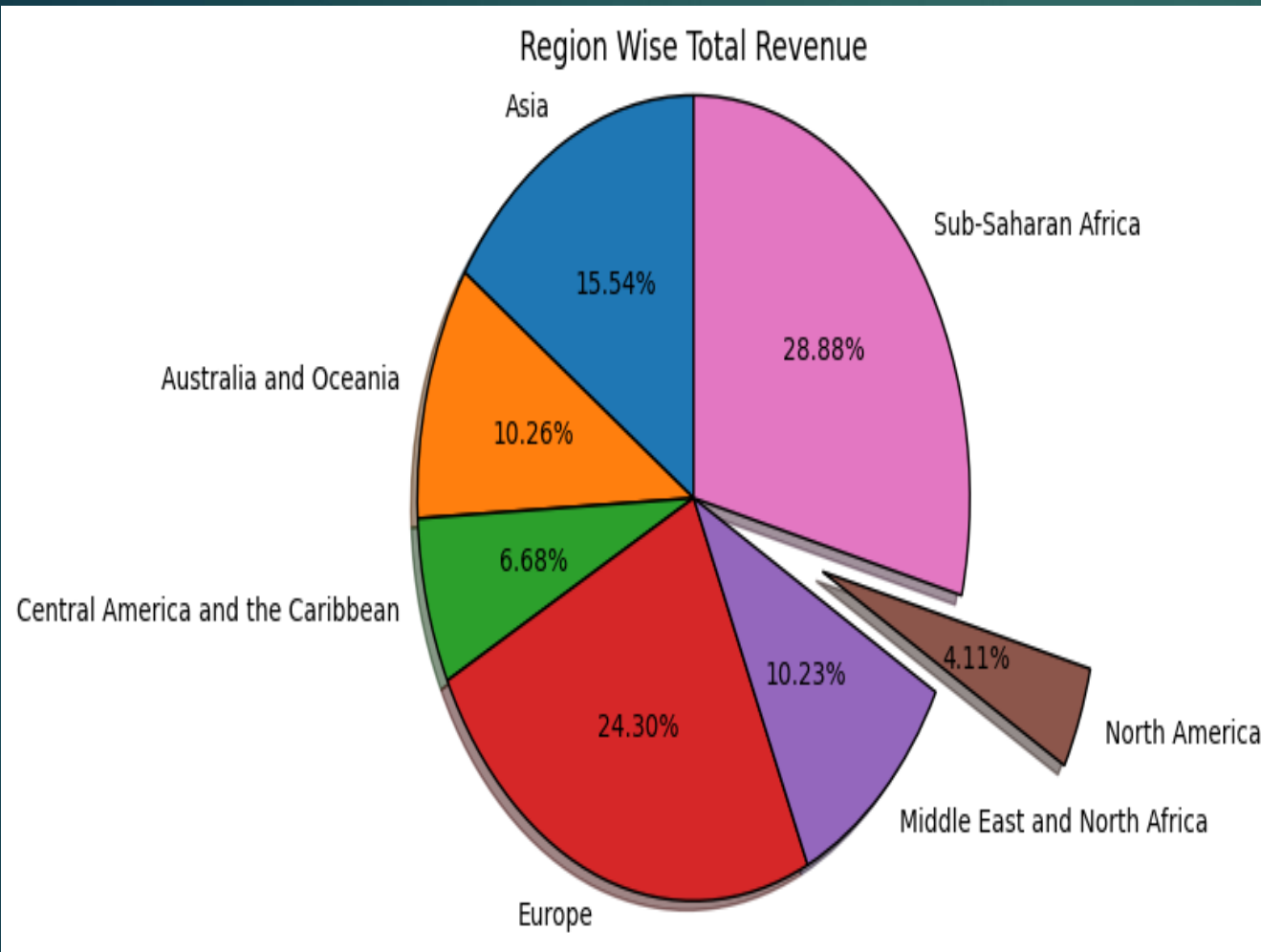
Total Revenue Vs Total Profit



--Insights--

- There appears to be a strong positive correlation between total revenue and total profit. This means that as total revenue increases, total profit tends to increase as well.
- A few data points, particularly those with high total revenue but relatively low profit, might be considered outliers.
- The data points are somewhat clustered in certain regions of the graph. This might suggest that there are groups of data points with similar characteristics or that certain factors are influencing the relationship between revenue and profit.

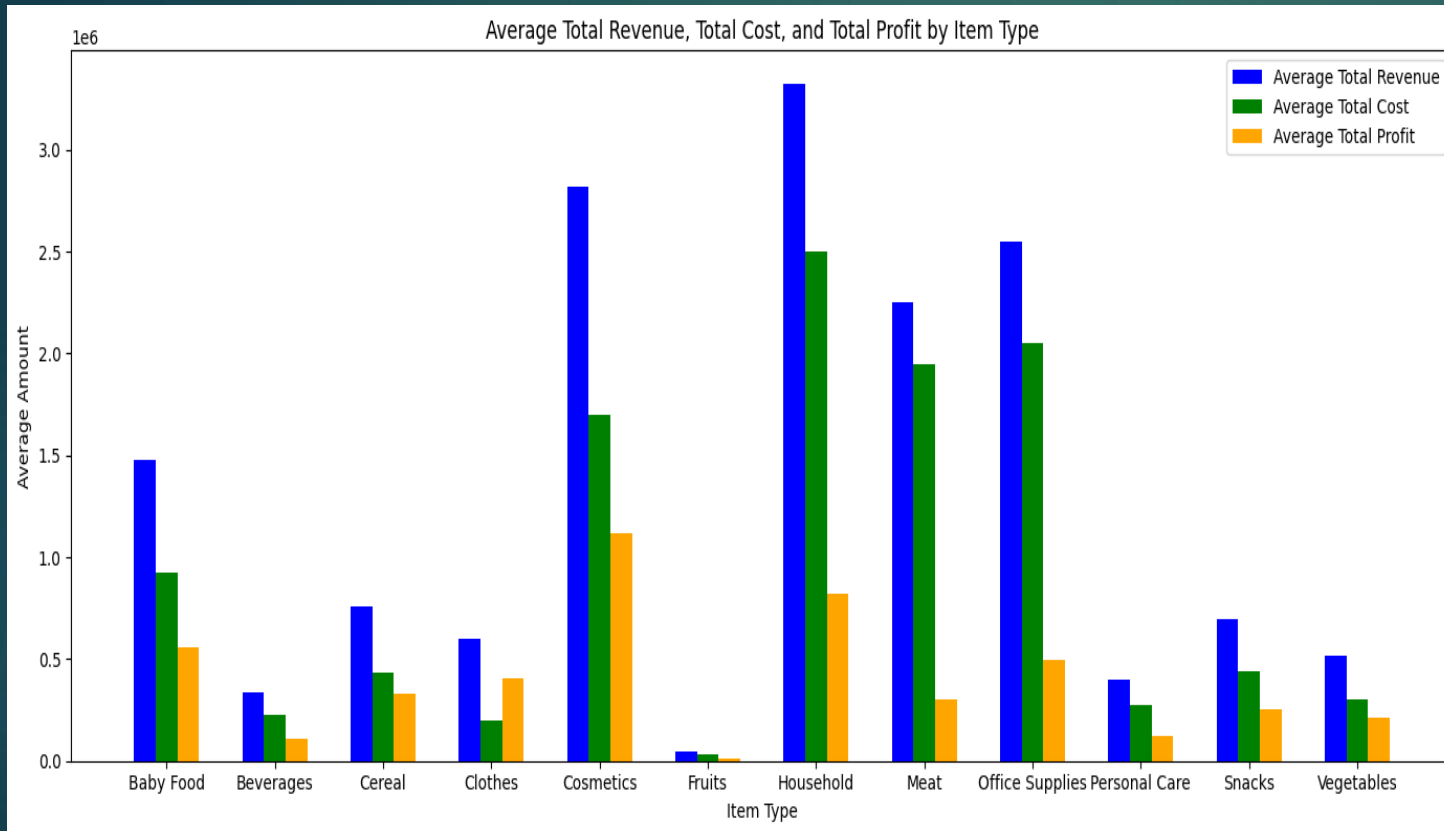
Region Wise Total Revenue



--Insights--

- The largest slice of the pie belongs to Sub-Saharan Africa, indicating that it contributes the most to the overall total revenue.
- The second-largest slice belongs to Europe, suggesting a significant contribution to the total revenue
- Asia accounts for 15.54% of the total revenue.
- Central America and the Caribbean region represents 6.68% of the total revenue.
- The North America contribute 4.11% to the total revenue.
- The Middle East and North Africa contribute 4.11% to the total revenue
- The large share of Sub-Saharan Africa and Europe suggests that the business may have a strong presence or market penetration in these regions.
- Regions with smaller shares, such as North America, Central America and Caribbean , Middle East & North Africa could represent potential growth opportunities for the business.

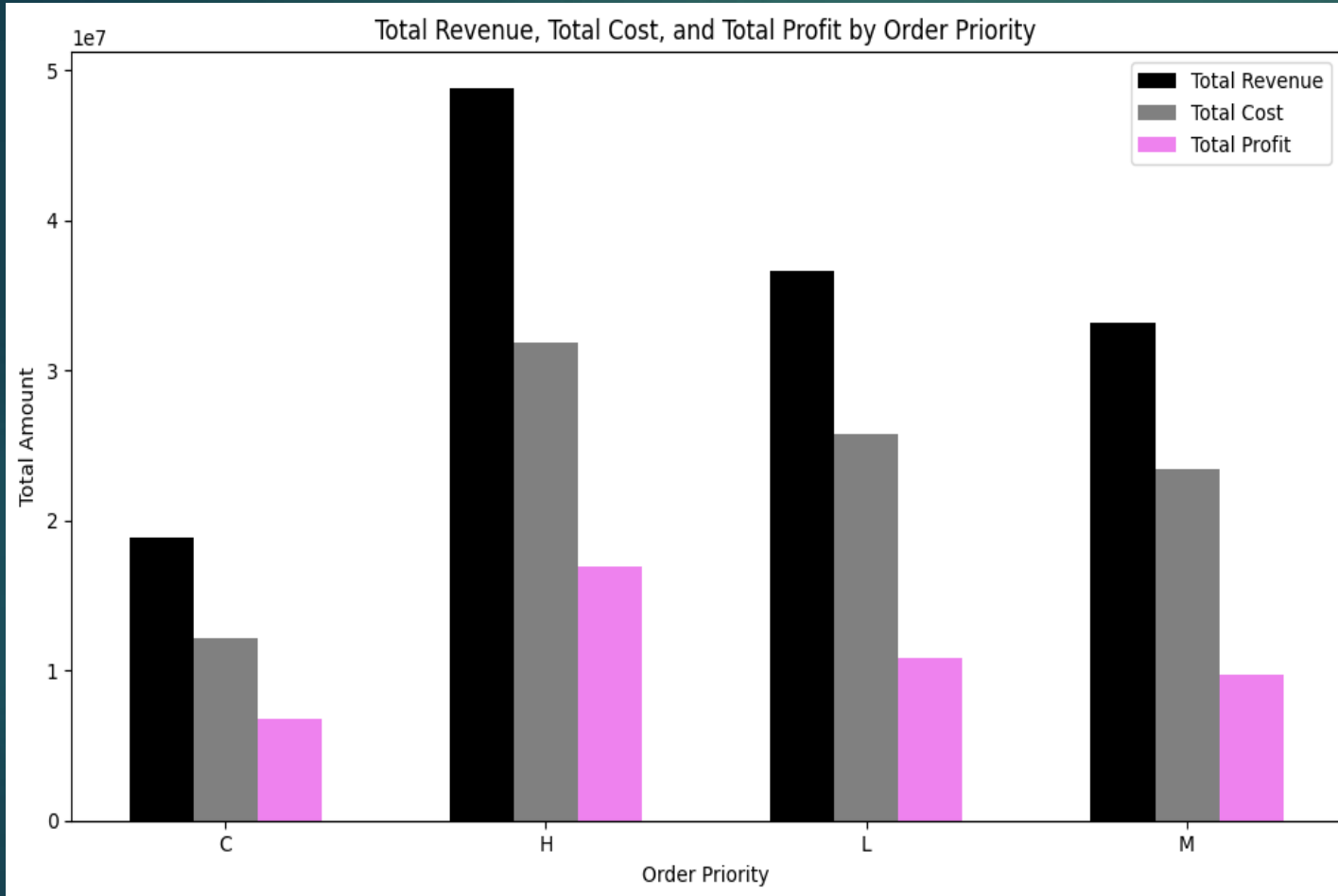
Average Revenue, Cost & Profit by Year



--Insights--

- Household items have the highest average total revenue, total cost, and total profit, indicating they are likely the most popular or high-margin products.
- Office Supplies and Cosmetics products also show relatively high figures, suggesting they are also profitable.
- Fruits and Beverages have the lowest average total revenue, total cost, and total profit, indicating they might be lower-margin or less popular items.
- Amazon should continue to invest in and promote household, office supplies, and cosmetic products.
- Amazon should explore ways to increase the profit margins for snacks, vegetables, beverages and fruits, such as adjusting pricing, reducing costs, or introducing new products.

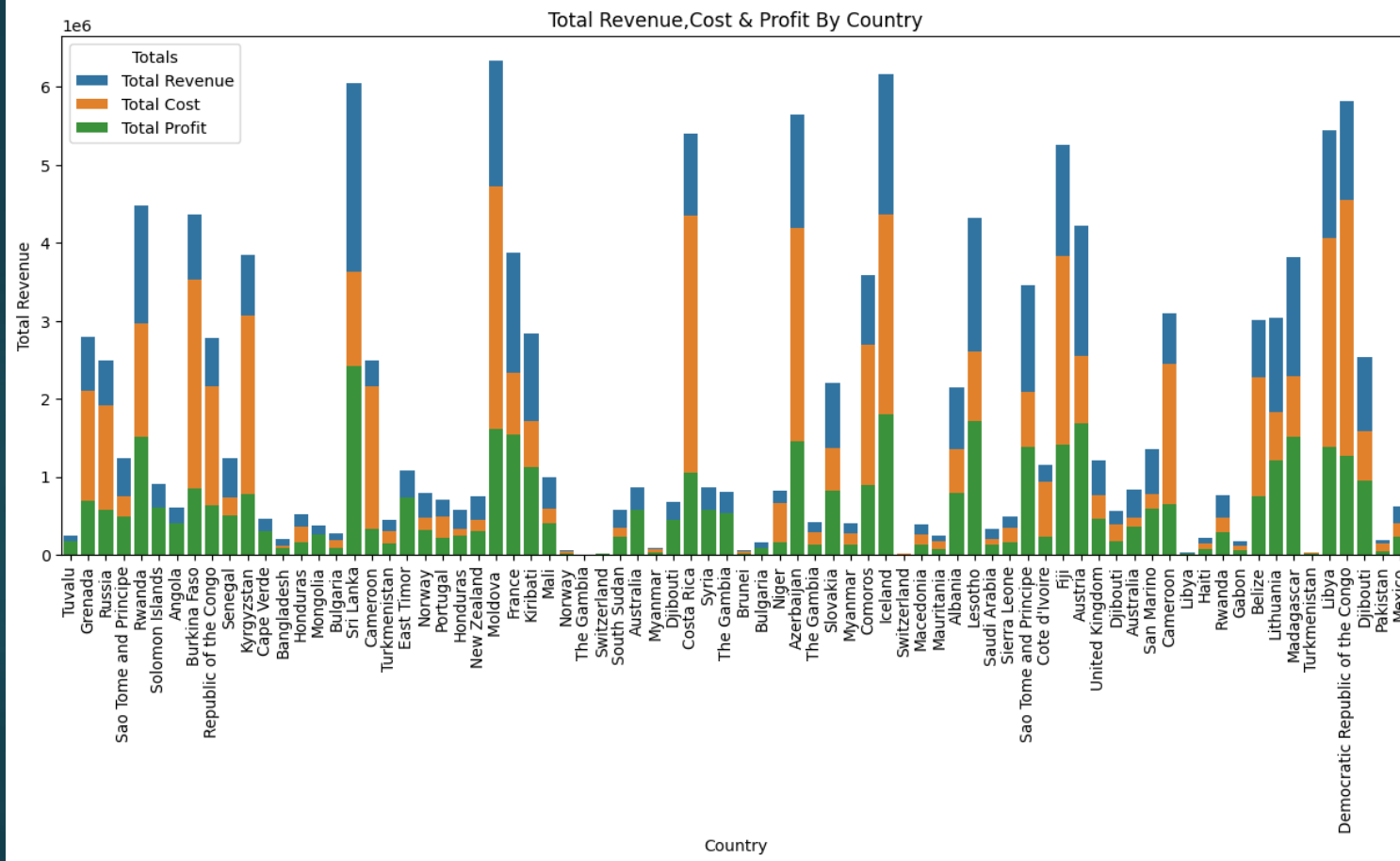
Total Revenue, Cost & Profit By Order Priority



--Insights--

- Order Priority H has the highest total revenue, total cost, and total profit, indicating it might be associated with larger or more complex orders.
- Order Priority C has the lowest total revenue, total cost, and total profit, suggesting it might be related to smaller or less urgent orders.
- Order Priorities M and L show moderate levels of total revenue, total cost, and total profit.

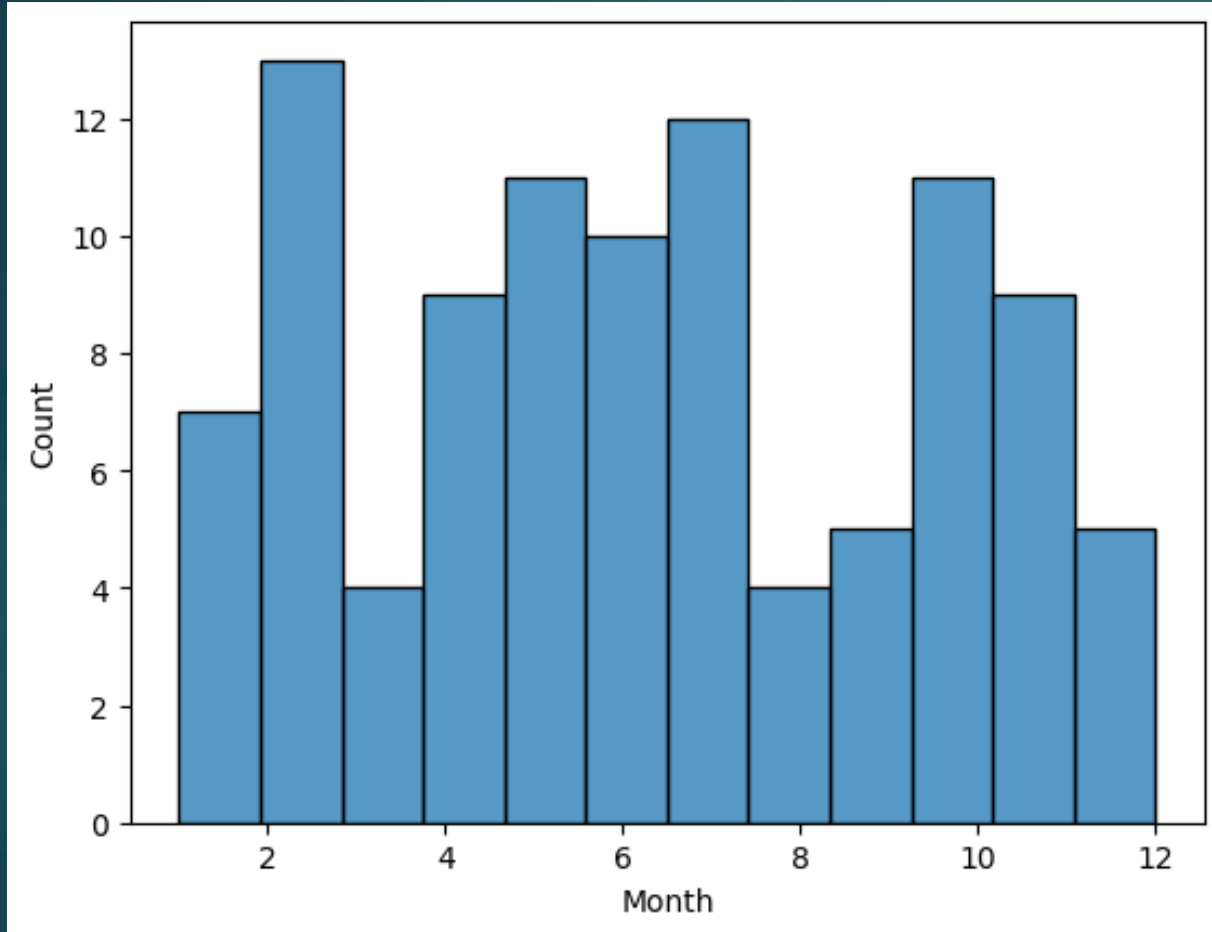
Total Revenue, Cost & Profit By Country



--Insights--

- Each country has three corresponding bars: blue for total revenue, orange for total cost, and green for total profit.
- Several countries, like the ones in the middle of the chart, have significantly higher total revenues and profits compared to others. The countries with the tallest blue bars (total revenue) and green bars (total profit) indicate the best financial performance.
- In some countries, the orange bars (representing total costs) are quite high relative to revenue and profit. This suggests that operating or product costs are a major factor in these markets, potentially impacting profitability.
- The green bars (total profit) are generally lower than the blue (revenue) and orange (cost) bars, indicating that while revenue is high, profit margins may not be substantial in many countries. The height of the green bar relative to the orange and blue bars varies, reflecting differences in profitability.
- A few countries have very high revenues and profits compared to others, suggesting these could be key markets driving overall performance.

No of Orders Per Month

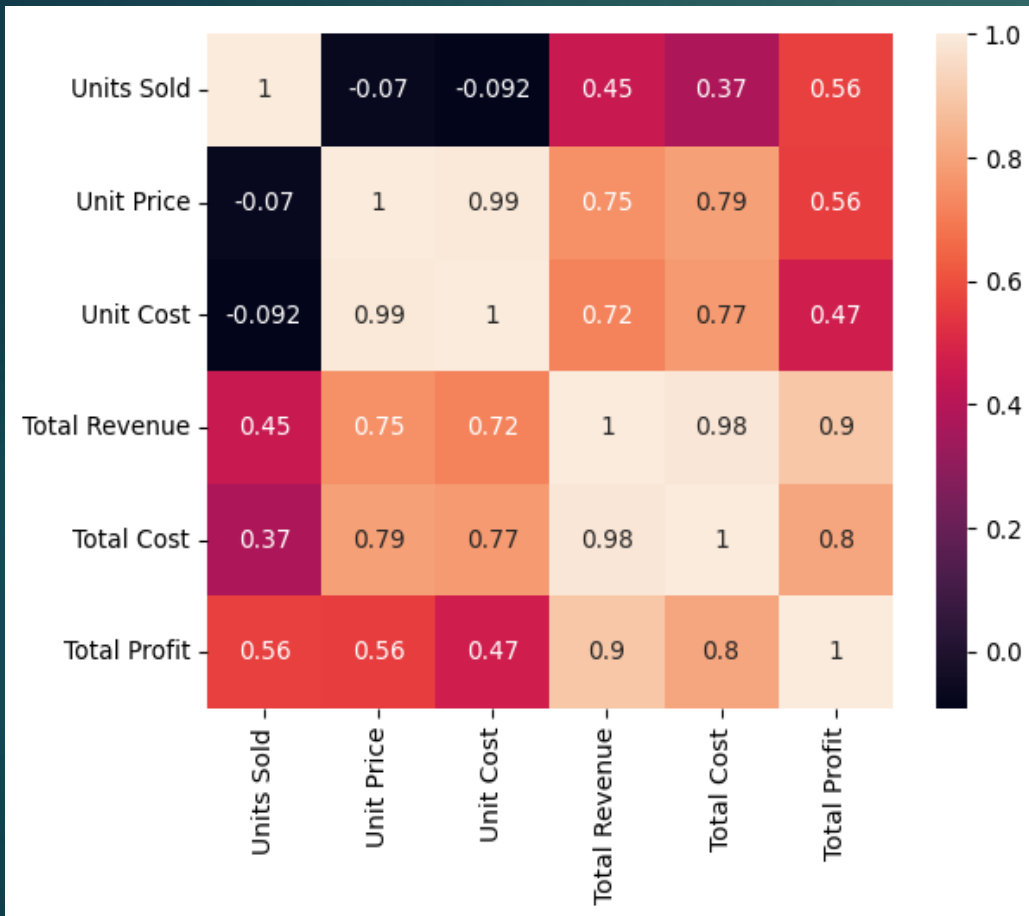


--Insights--

Months such as February, July and May have had the most amount of orders placed

Months such as August, March have had the least amount of orders placed as these are the bars with least frequency.

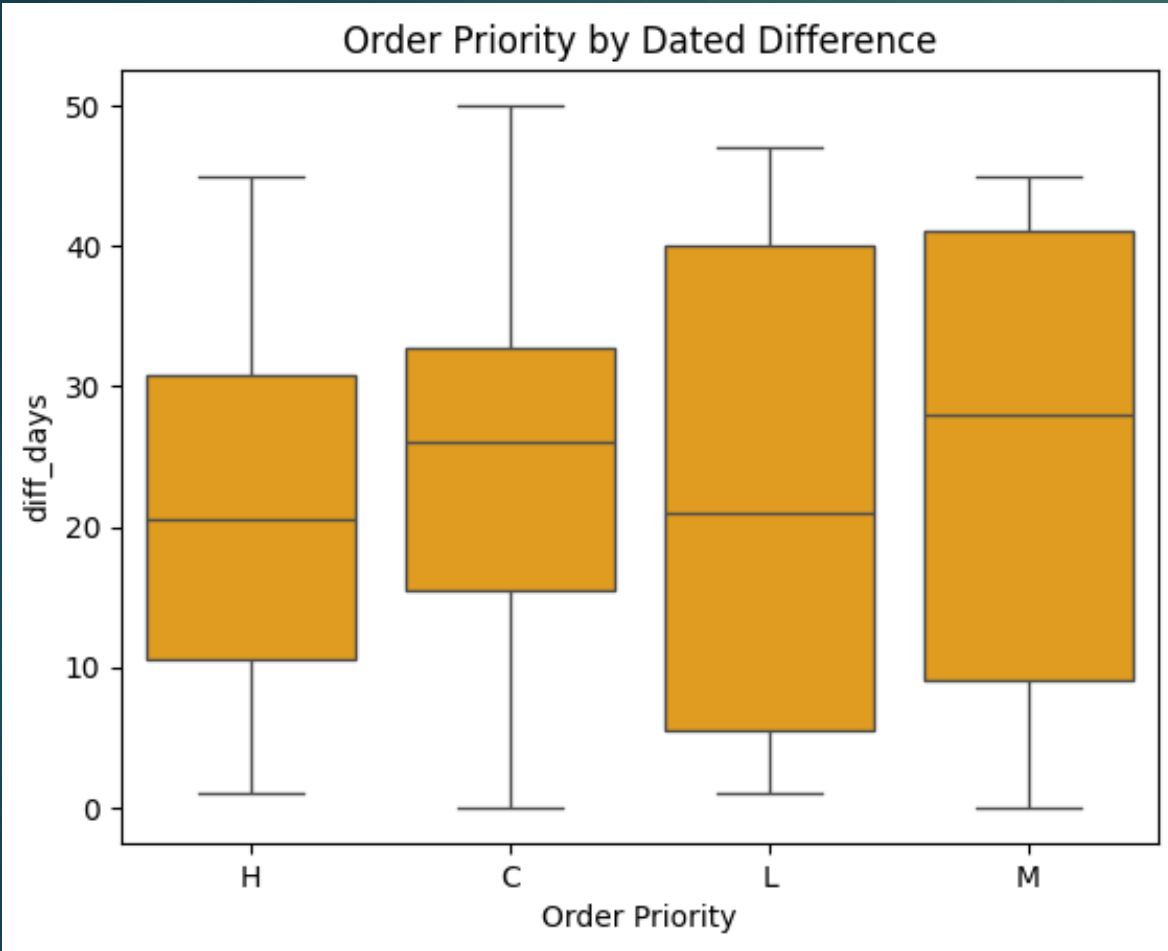
Correlation b/w all the Numerical Variables



--Insights--

- **Units Sold:** - Negatively correlated with Unit Price (-0.07) and Unit Cost (-0.092), though the correlations are weak. - Positively correlated with Total Revenue (0.45) and Total Profit (0.56), indicating that more units sold generally lead to higher revenue and profit.
- **Unit Price:** - Strongly positively correlated with Unit Cost (0.99), suggesting that as the unit price increases, the unit cost also increases almost proportionally. - Moderately positively correlated with Total Revenue (0.75), indicating that higher unit prices contribute to higher revenue. - Positively correlated with Total Profit (0.56), meaning higher unit prices generally lead to more profit.
- **Unit Cost:** - Similar correlations as Unit Price, with a strong positive correlation with Total Cost (0.98). - Moderately positively correlated with Total Revenue (0.72) and Total Profit (0.47), indicating that unit cost influences both revenue and profit.
- **Total Revenue:** - Strong positive correlations with Total Cost (0.98) and Total Profit (0.9), showing that revenue is a significant driver of both costs and profit. - Moderately positively correlated with Units Sold (0.45), meaning that selling more units generally increases revenue.
- **Total Cost:** - Strong positive correlation with Total Profit (0.8), indicating that as costs increase, profits still tend to increase, likely due to the scale of operations.
- **Total Profit:** - Positively correlated with all other variables, especially with Total Revenue (0.9) and Units Sold (0.56). This suggests that both higher revenue and more units sold significantly boost profits.

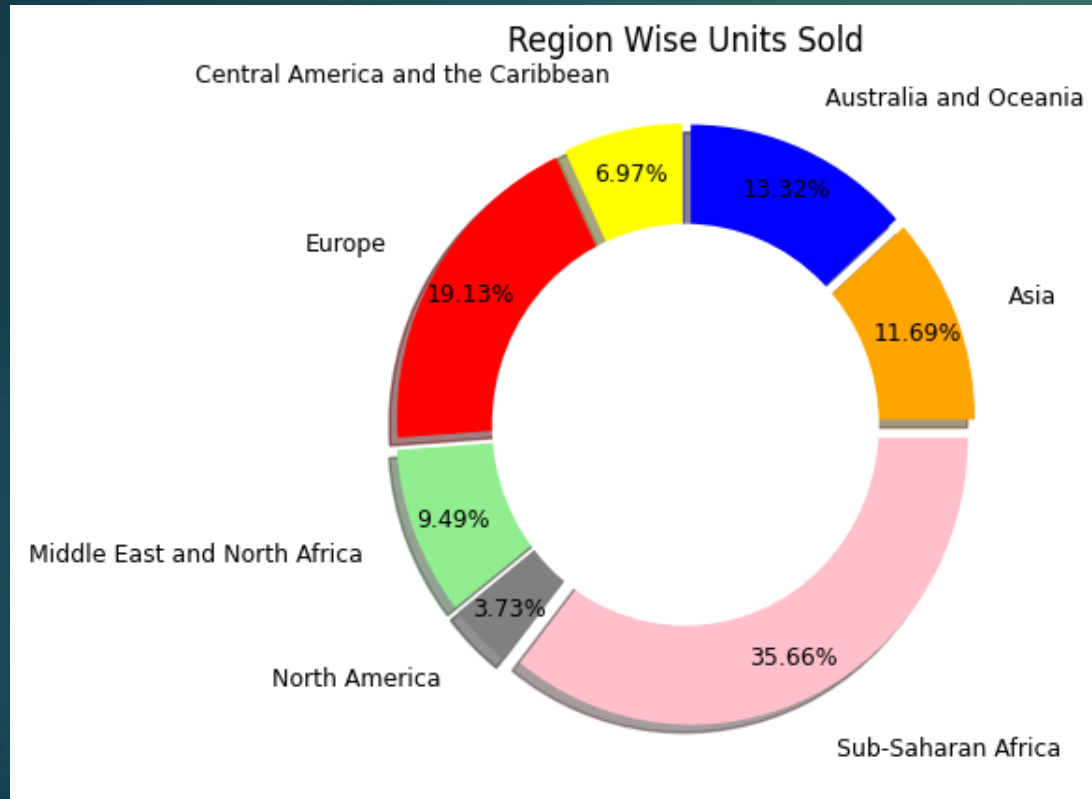
Order Priority By Dated Difference



--Insights--

- High priority orders are generally processed more quickly and consistently than orders with lower priorities as High Priority has the lowest median difference and smallest IQR.
- Low priority orders have the longest processing times and the greatest variability in processing times as Low Priority has the highest median difference and highest IQR.
- Current and Medium priorities fall somewhere in between, with Current being slightly faster.

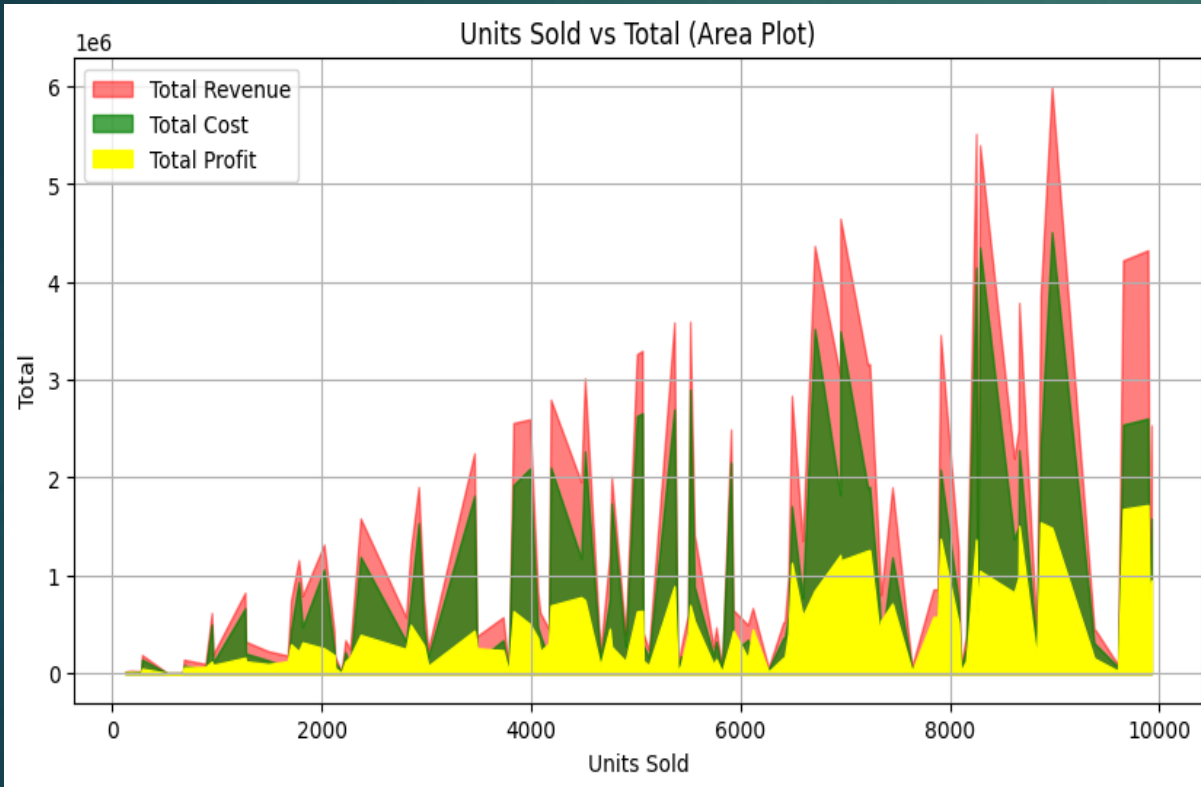
Region Wise Units Sold



--Insights--

- **Sub-Saharan Africa dominates:** The region accounts for the largest share of units sold at 35.66%, indicating a strong market presence and demand.
- **Europe follows:** With 11.69% of the total sales, Europe emerges as the second-largest market. This suggests significant growth potential and opportunities for expansion.
- **Asia and Australia/Oceania:** These regions contribute relatively similar shares to the overall sales, each accounting for around 11-13%.
- **Emerging Markets:** Central America and the Caribbean, Middle East and North Africa, and North America collectively represent a smaller portion of the sales, suggesting untapped potential and future growth opportunities.
- The dominance of Sub-Saharan Africa and Europe highlights the importance of focusing on these regions for maximizing sales.
- The emerging markets, although smaller in size, offer significant growth potential as they develop and increase their purchasing power.

Units Sold vs Total Revenue, Cost & Profit



--Insights--

- There's a clear positive correlation between units sold and total revenue, as expected. As the number of units sold increases, the total revenue also rises.
- Total costs generally increase with the number of units sold, suggesting a linear relationship between production costs and sales volume.
- While revenue and costs are increasing, profit fluctuates significantly
- In some areas, the gap between revenue and costs is relatively small, suggesting lower profit margins. In other areas, the gap is wider, indicating higher profit margins.

Conclusions

- According to Data-Set the Sub-Saharan Region has been the biggest contributor to the Total Revenue, Total Profit & Total Cost meaning the management is taking the right decisions in this region.
- However the regions like North America , Middle East needs better decision-making and marketing strategies.
- Various Item Types Like Household ,Cosmetic items have performed outstandingly , but some item types like Fruits, Personal care need to be either discounted or made better efforts at marketing them .
- Order Priority H has been overall benefittial to the company but Order priority C has been the least benefitting.
- February , July , May have been the months with most orders placed whereas
- August & March have been the least meaning some sort of discounts should be given in these months to increase the amount of orders.
- There has been many fluctuations in revenue generated over the years meaning the cause of fluctuations need to be studied and respective counter-measures need to be taken.