PROJECT-1

REPORT

ON

AMAZON SALES DATA

STARTING FROM EXAMINING THE FEATURES OF DATA USING JUPYTER NOTEBOOK

-IMPORTING LIBRARIES TO CARRU OUT THE FUNCTIONS ASSOCIATED

-LOAD THE DATA INTO KERNEL

```
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

RUN THROUGH THE DATA AND ITS DESCRIPTION

```
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
```

GOING THROUGH EACH FIELD

Now we observe the each features present in the dataset.

Region: The geographical area where the sale was made.
Country: The specific country where the transaction took place.
Item Type: The category or type of item sold.
Sales Channel: The medium through which the sale was conducted (e.g., online, offline).
Order Priority: The priority level assigned to the order (e.g., high, medium, low).
Order Date: The date when the order was placed.
Order ID: A unique identifier for each order.
Ship Date: The date when the order was shipped to the customer.
Units Sold: The quantity of items sold in the order.
Unit Price: The selling price per unit of the item.
Unit Cost: The cost to the company for each unit of the item.
Total Revenue: The total amount of money generated from the sale (calculated as Units Sold * Unit Price).
Total Cost: The total cost incurred for the items sold (calculated as Units Sold * Unit Cost).
Total Profit: The total profit earned from the sale (calculated as Total Revenue - Total Cost).

CHECKING IF DUPLICATED FIEDLS PRESENT

Dropping the duplicate rows

```
data = df.drop_duplicates()
data.to_csv('C:/Users/shamb/Downloads/Amazon Sales data.csv', index=False)

# Save the cleaned dataset to a new file
data
```

	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

	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
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
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
95	Sub-Saharan Africa	Mali	Clothes	Online	M	7/26/2011	512878119	09-03-2011	888	109.28	35.84	97040.64	31825.92	65214.72
96	Asia	Malaysia	Fruits	Offline	L	11-11-2011	810711038	12/28/2011	6267	9.33	6.92	58471.11	43367.64	15103.47
97	Sub-Saharan Africa	Sierra Leone	Vegetables	Offline	C	06-01-2016	728815257	6/29/2016	1485	154.06	90.93	228779.10	135031.05	93748.05
98	North America	Mexico	Personal Care	Offline	M	7/30/2015	559427106	08-08-2015	5767	81.73	56.67	471336.91	326815.89	144521.02
99	Sub-Saharan Africa	Mozambique	Household	Offline	L	02-10-2012	665095412	2/15/2012	5367	668.27	502.54	3586605.09	2697132.18	889472.91

100 rows × 14 columns

No duplicates present as rows and columns are equal to values before we have applied the function.

CHECKING IF ANY VALUE IS MISSING OR NOT

```
# drop rows with any missing values
data1 = data.dropna()
data1.to_csv('C:/Users/shamb/Downloads/Amazon Sales data.csv', index=False)
data1
```

| 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
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
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
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	639077.50
...	...	...	...	...	...	...	...	...	...	...	...	...	...
Sub-Saharan Africa	Mali	Clothes	Online	M	7/26/2011	512878119	09-03-2011	888	109.28	35.84	97040.64	31825.92	65214.72

100 rows × 14 columns

No value is missing.

BUILDING UP THE VISUALIZATION THROUGH TABLEAU

1)Financial Analysis

Revenue over region
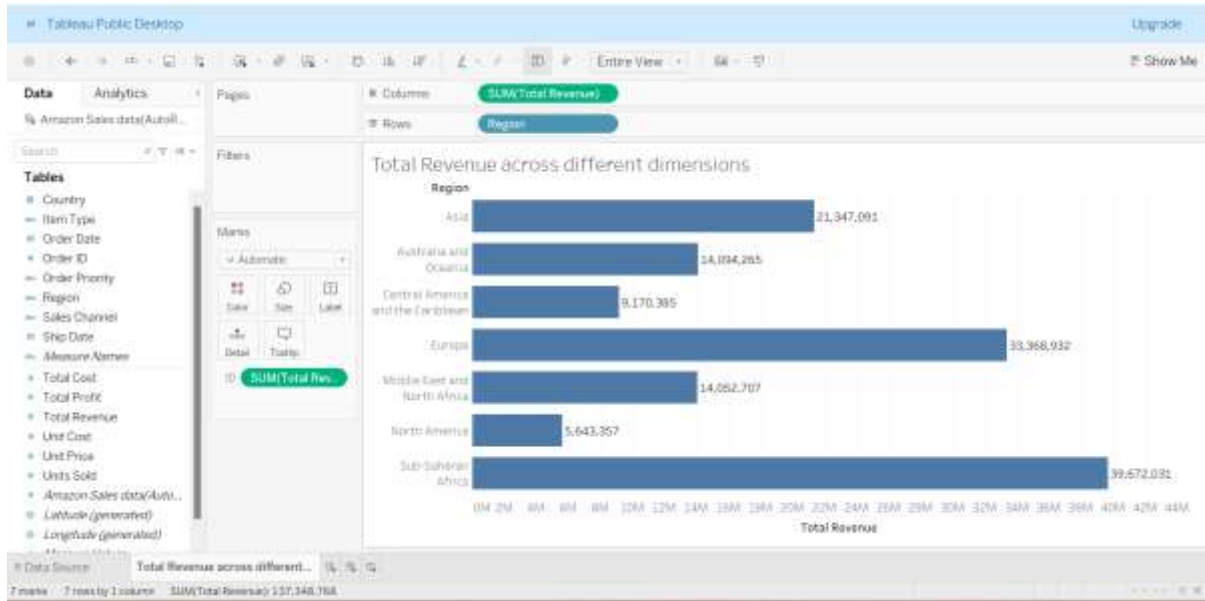
• **Regional Performance Comparison**: You can quickly identify in this format that which regions are generating the highest and lowest revenue. This helps in understanding which markets are performing well and which may need attention or further investment.

• **Revenue Distribution**: Observing the relative size of the bars will show how revenue is distributed across regions. This could highlight if revenue is concentrated in a few regions or spread more evenly. In this we can see sub-saharan Africa has concentrated major amount of revenue.

Total cost and profit
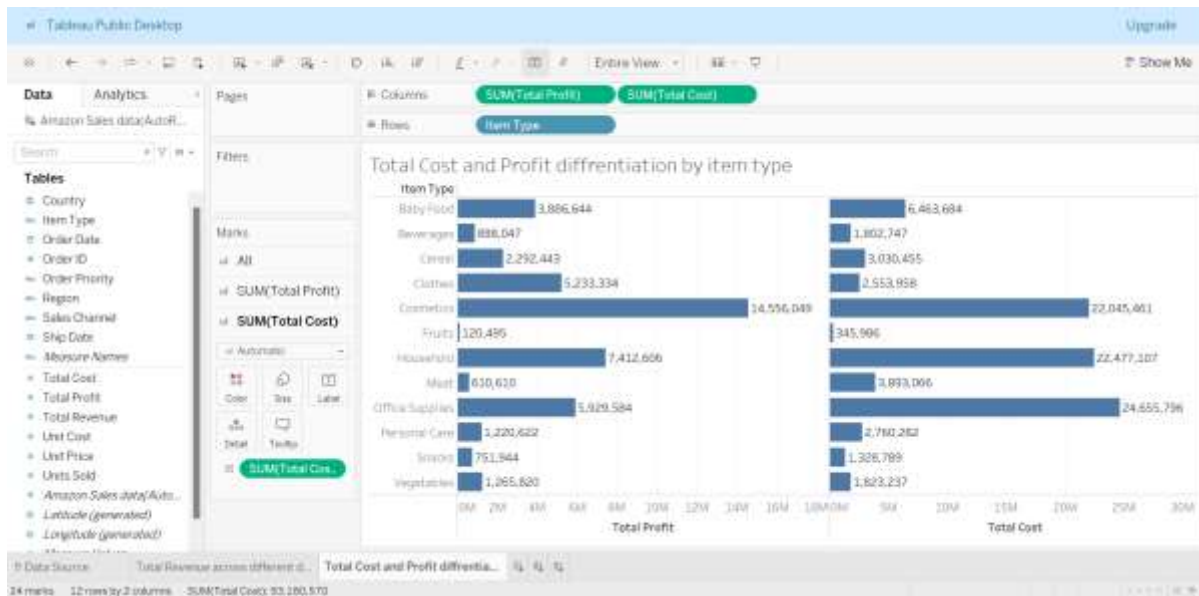
- **Cost vs. Profit Comparison:**

    - By examining the bars, helps to examine understand the profitability of different items.

- **Profitability Analysis:**

    - Determine which item types are generating the most profit relative to their costs. High-profit items with relatively low costs are ideal for focus and expansion.
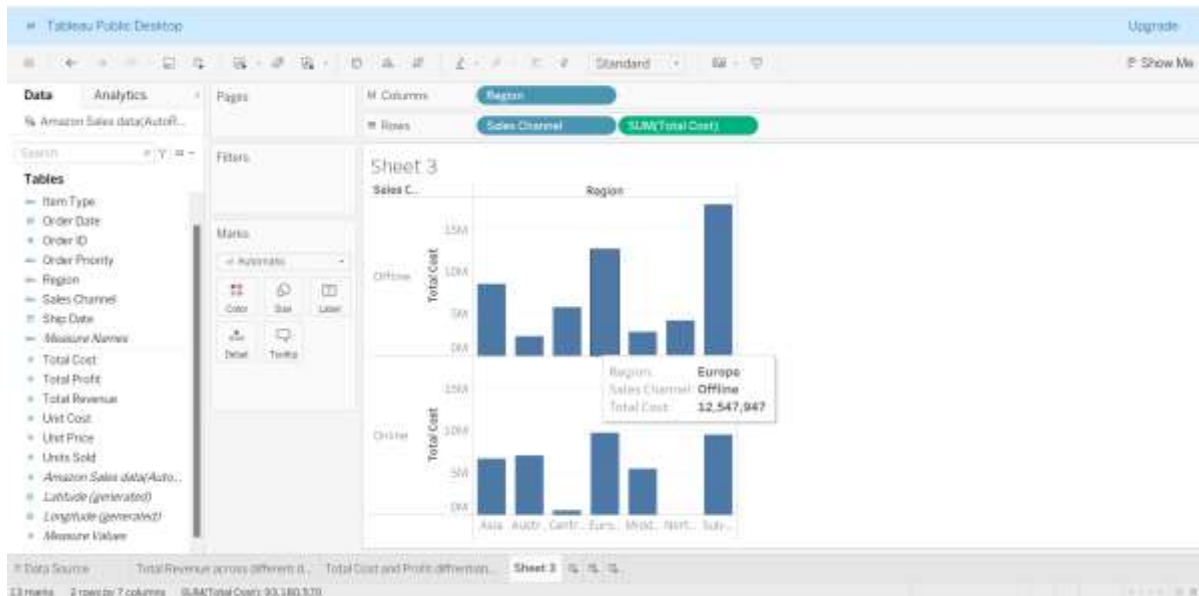
- **Cost Efficiency:**

    - Analyse item types with high costs but low profit. These may be areas where cost control measures or price adjustments are needed.

- **Product Performance:**

    - Identify which item types contribute most to overall profitability. This can inform inventory decisions and marketing strategies.

Sales channel and cost by region

- **Regional Cost Distribution:**

    - Determine which regions have the highest total costs. This can help identify where significant expenditures are occurring and whether these costs align with the expected business performance in those regions.

- **Sales Channel Effectiveness:**

    - Evaluate which sales channels contribute most to total costs in different regions. This insight can help assess the efficiency of each channel and guide improvements or adjustments.

2)PROFIT ANALYSIS

- **Strategic Resource Allocation:**

    - **Focus on High-Profit Areas:** Allocate resources and investment towards sales channels, item types, and regions that show high profitability to maximize returns.
    - **Cost Management in Low-Profit Areas:** Implement cost control measures in areas with lower profitability to improve margins and overall financial health.
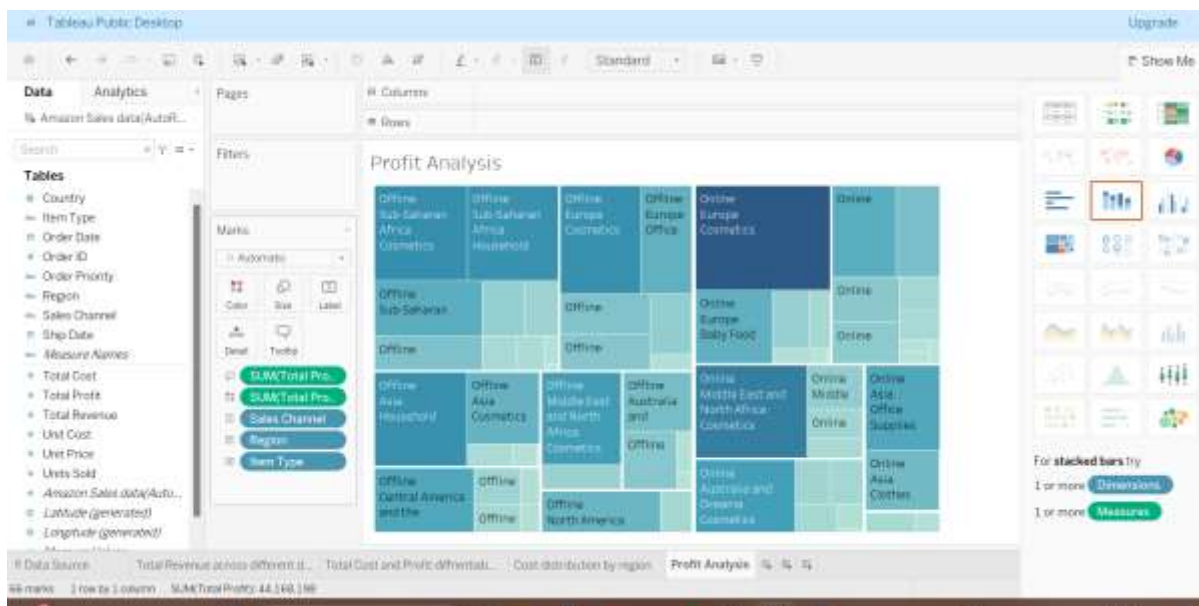
Order priority

- **High Priority Orders:**

    - **Revenue and Profit Analysis:** Determine if high-priority orders generate significantly more revenue and profit compared to lower-priority ones. This can highlight the effectiveness of prioritizing certain orders.
    - **Cost Implications:** Assess whether high-priority orders incur higher costs due to expedited processing or special handling. This insight can help balance between speed and cost-efficiency.

- **Low Priority Orders:**

    - **Revenue Contribution:** Evaluate if low-priority orders still contribute substantial revenue or profit. This helps in understanding if lower-priority orders are profitable despite not being expedited.
    - **Cost Management:** Analyze the cost associated with fulfilling low-priority orders. Lower costs with lower priority might suggest efficiency in handling these orders.

Cost Efficiency Analysis

- **Cost vs. Revenue:** Compare the cost of fulfilling orders with the revenue generated. High-cost orders should ideally generate high revenue to maintain profitability. Assess if the high cost of prioritizing orders is justified by the revenue and profit.
- **Profitability of Different Priorities:** Evaluate the cost-effectiveness of different order priorities. Determine if prioritizing leads to higher or lower overall costs and whether it's balanced by corresponding revenue and profit.
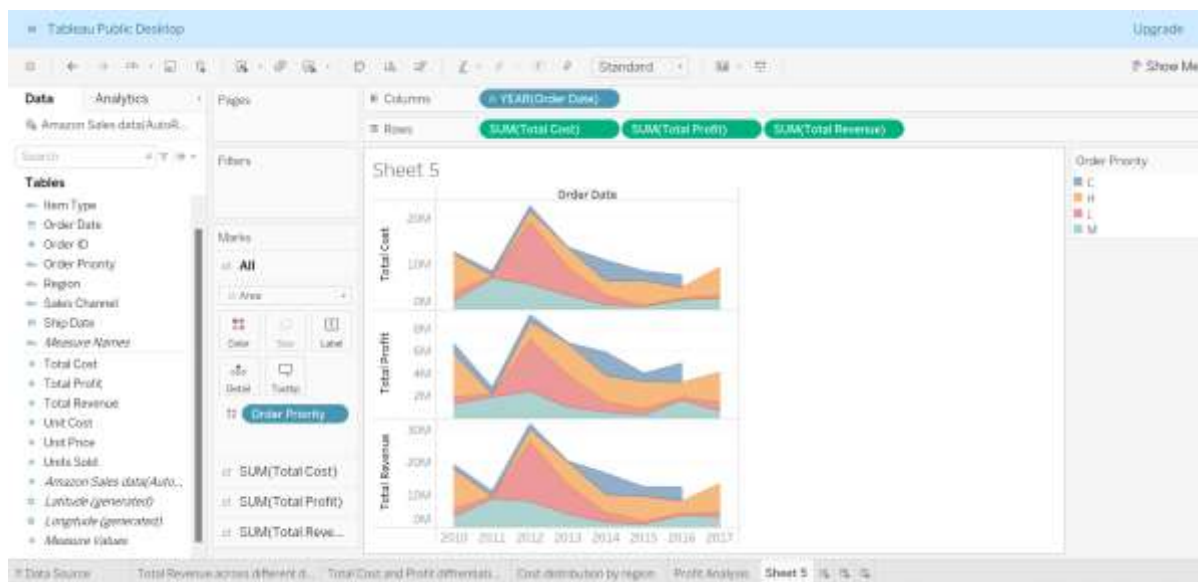
Dashboard
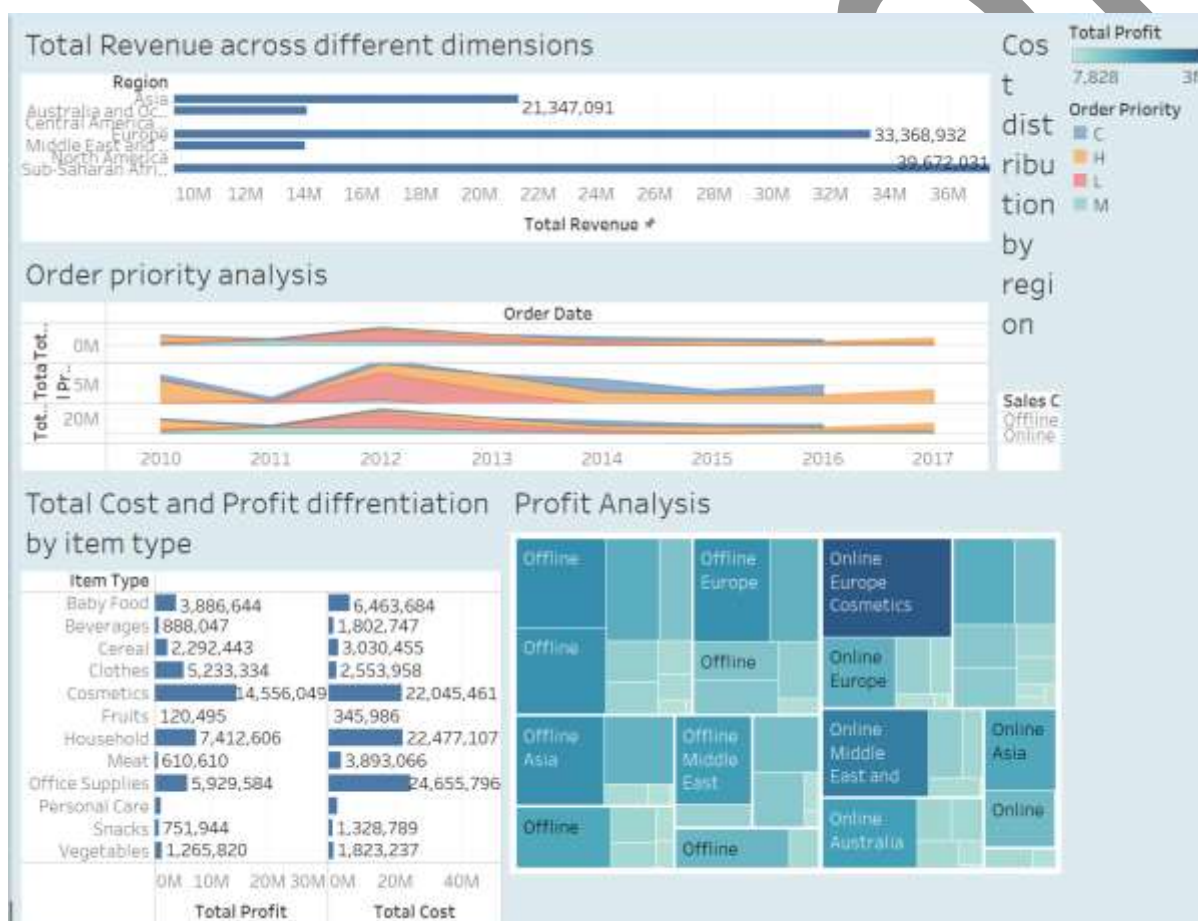
CUSTOMER ANALYSIS

Customer preferences

- **Popular Item Types:**

- **Top Sellers:** Identify which item types are the most popular based on units sold. This helps in understanding customer preferences and demand.
- **Market Trends:** Determine if certain item types are trending or if their popularity is increasing or decreasing over time.

- **Inventory Management:**

    - **Stock Levels:** Adjust inventory levels based on the popularity of item types. Ensure that high-demand items are well-stocked to meet customer needs and avoid stockouts.
    - **Reordering Strategies:** Develop efficient reordering strategies for top-selling items to maintain optimal stock levels and reduce the risk of lost sales.
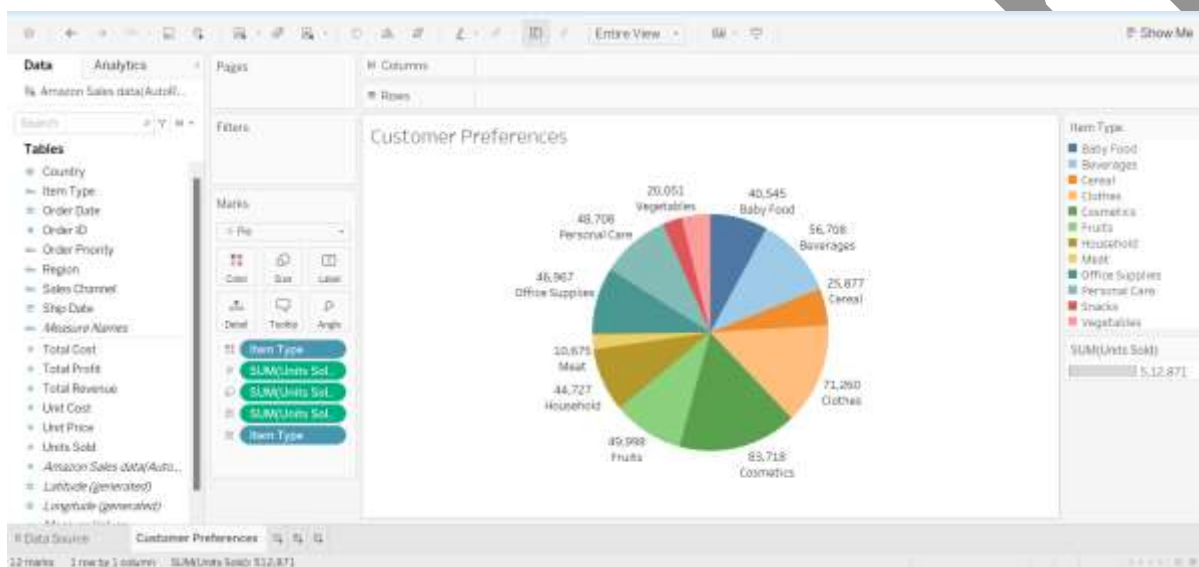
Order profitability impact

- **Market Focus:**

    - **High-Priority Regions:** Recognize regions that generate significant revenue. These areas should be the focus of your sales and marketing efforts to further capitalize on their revenue potential.
    - **Underperforming Regions:** Highlight regions with smaller slices. Investigate why these regions are underperforming and explore opportunities to boost revenue.

- **Resource Allocation:**

    - **Investment Decisions:** Allocate resources effectively by investing more in regions that contribute the most to revenue. This ensures that high-performing regions continue to thrive.
    - **Growth Opportunities:** Identify regions with lower revenue contributions but high growth potential. Allocate resources to develop these markets and increase their revenue share
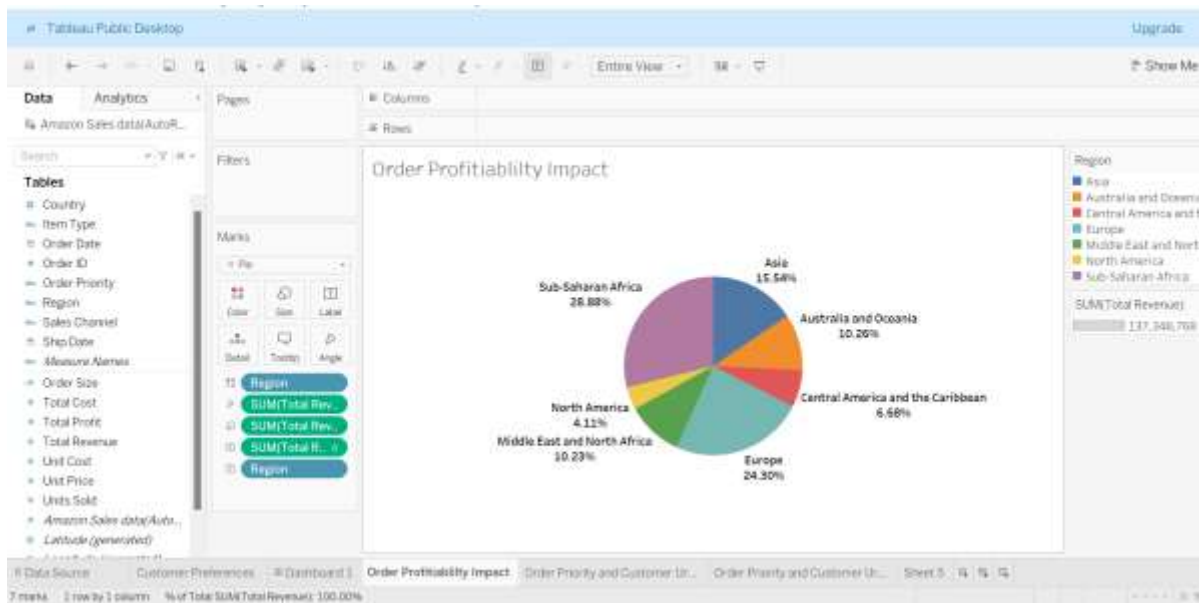
Order profitability and customer urgency

- **Customer Segmentation:**

    - **Segment Preferences:** Understand which customer segments prefer large versus small orders. This helps in tailoring marketing strategies to different customer groups.
    - **Order Size Trends:** Analyze trends in order sizes to identify shifts in customer purchasing behavior over time.
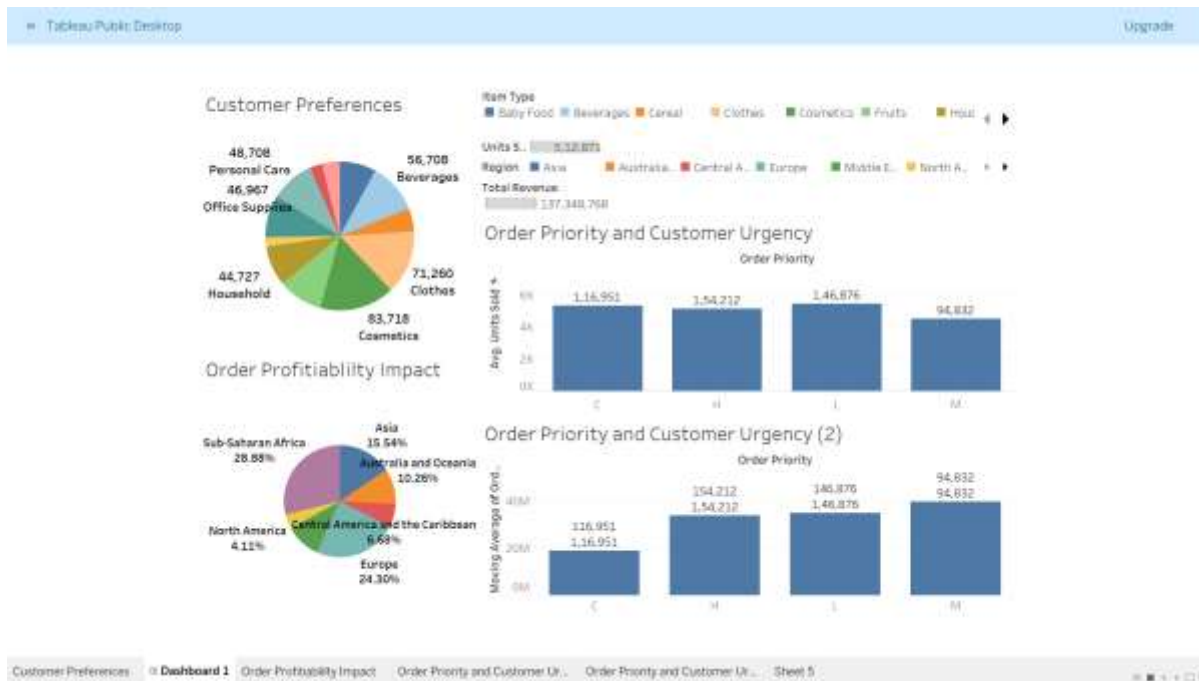
- **Inventory and Supply Chain Management:**

    - **Stock Management:** Ensure inventory levels are sufficient to meet the demand for large orders without causing stockouts.
    - **Order Fulfillment:** Optimize supply chain processes to handle both large and small orders efficiently, minimizing delays and maximizing revenue.
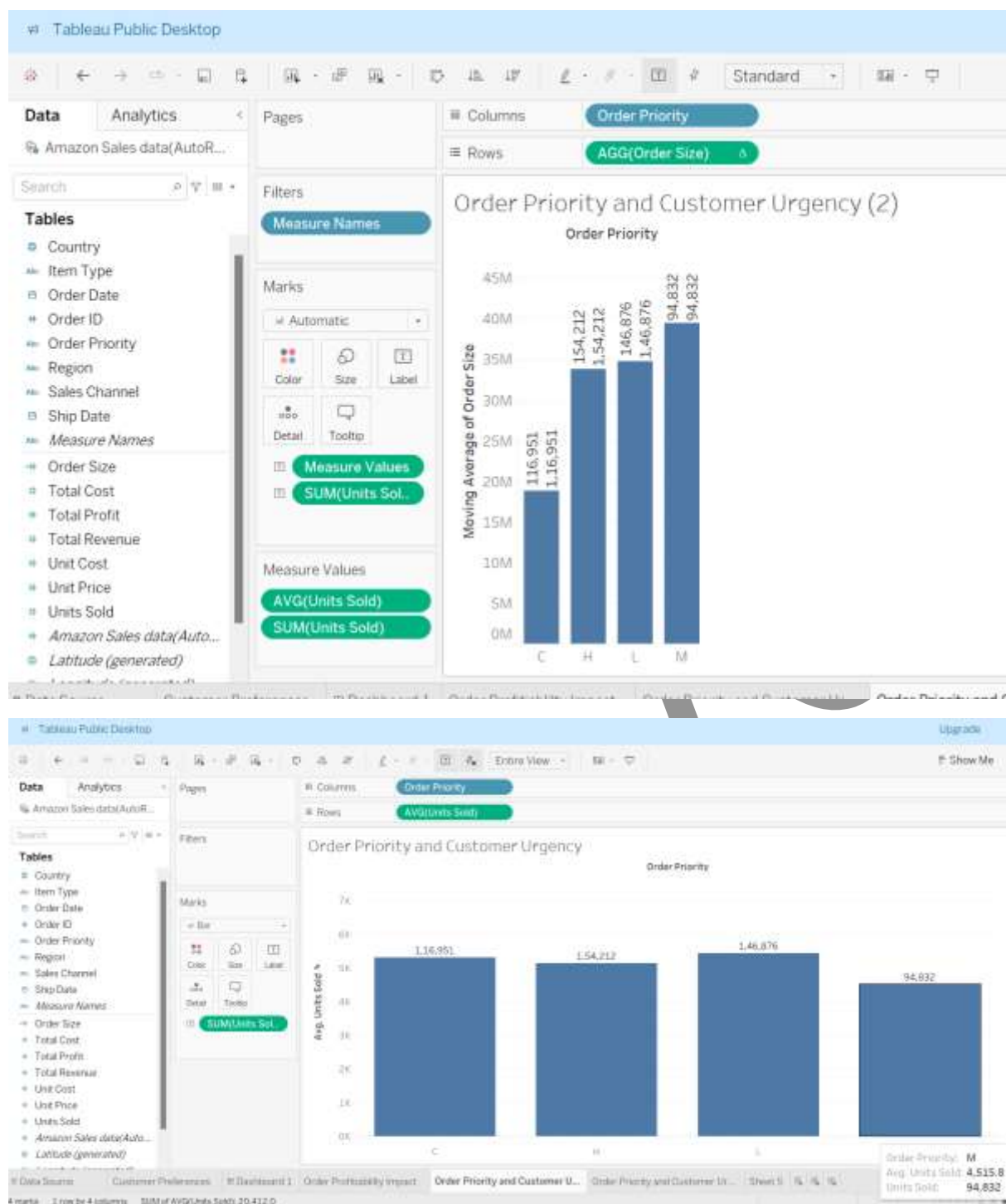
Dashboard

Sales Analysis

Sales by region

- **Proportional Differences: (Example)**

  - **Region A:**
    - o **Category 1:** 5000 sales
    - o **Category 2:** 7000 sales
    - o **Insight:** Region A shows a significant increase in sales in Category 2 compared to Category 1. This indicates growing market potential or increased demand in this region.
  - **Region B:**
    - o **Category 1:** 3000 sales
    - o **Category 2:** 2000 sales
    - o **Insight:** Region B has a decrease in sales in Category 2 compared to Category 1, suggesting a decline in performance. This region might need targeted marketing or sales strategies to address the drop.
  - **Region C:**
    - o **Category 1:** 2000 sales
    - o **Category 2:** 4000 sales
    - o **Insight:** Region C shows a substantial increase in sales in Category 2, indicating strong growth. This region could be a new focus area for expanding market efforts.
  - **Region D:**
    - o **Category 1:** 4000 sales
    - o **Category 2:** 3000 sales
    - o **Insight:** Region D sees a decrease in sales in Category 2. Understanding the reasons behind this decline will be crucial for future planning.

- **Trend Identification:**

- **Growth Regions:**
  - o **Regions A and C** have shown growth from Category 1 to Category 2, indicating positive trends and successful strategies in these regions.
- **Declining Regions:**
  - o **Regions B and D** have shown a decline, indicating potential issues such as market saturation, increased competition, or ineffective strategies that need to be addressed.

- **Strategic Focus:**

  - **High-Performing Regions:**
    - o **Region A**: Consistent high performance with growth, suggesting it's a strong market with increasing demand.
    - o **Region C**: Significant growth suggests potential for further investment and focus.
  - **Underperforming Regions:**
    - o **Region B**: Declining sales call for a reassessment of strategies, potential market challenges, or shifting consumer preferences.
    - o **Region D**: Despite being a strong performer in Category 1, the decline in Category 2 sales needs to be investigated and addressed.

- **Marketing and Sales Strategies:**

  - **Targeted Campaigns:**
    - o **Region A**: Continue and possibly intensify successful marketing strategies that led to the increase in sales.
    - o **Region C**: Focus on expanding presence and capitalizing on growth with targeted promotions and marketing campaigns.
  - **Resource Allocation:**
   - o **Region B and D**: Allocate resources to understand and address the decline i
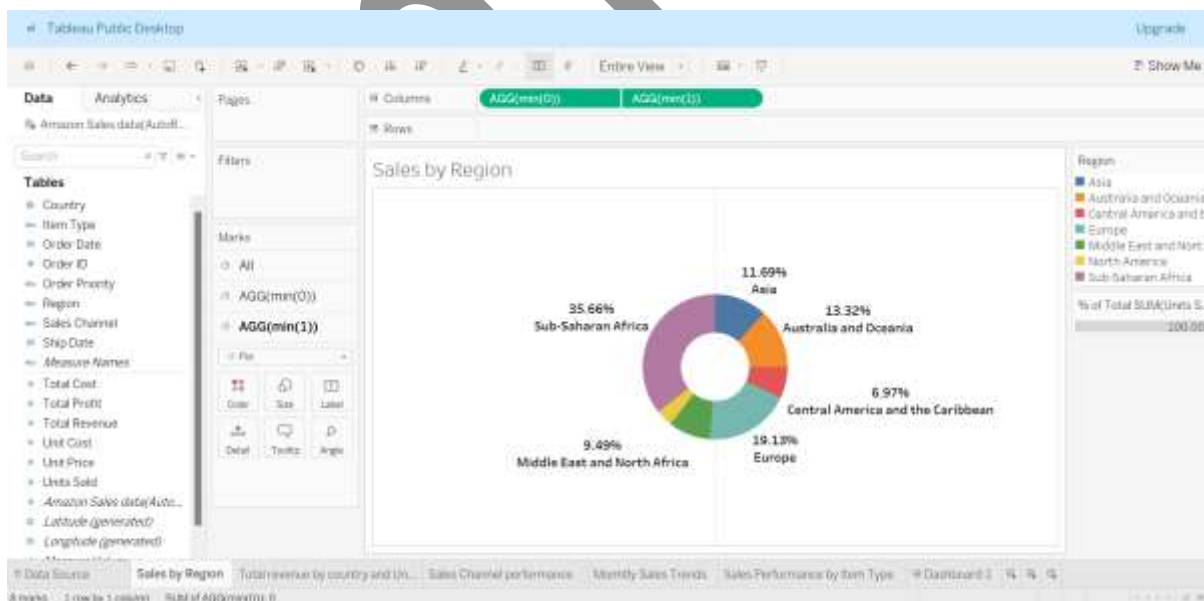
Total revenue by country and units sold

- **Revenue Distribution:**

- **Major Contributors:** Identify countries that contribute the most to total revenue. Larger slices in the pie chart represent these key markets.
- **Minor Contributors:** Highlight countries with smaller revenue contributions. These might represent markets with potential for growth or needing strategic adjustments.

- **Market Focus:**

  - **High-Priority Markets:** Recognize countries generating significant revenue. Focus sales and marketing efforts on these markets to sustain and enhance performance.
  - **Underperforming Markets:** Identify countries with lower revenue. Investigate reasons for underperformance and explore opportunities to boost revenue.

- **Resource Allocation:**

  - **Investment Decisions:** Allocate resources efficiently by investing more in high-revenue countries. Ensure these markets receive adequate attention and support.
  - **Growth Potential:** Allocate resources to develop underperforming markets with high growth potential.

- **Sales and Marketing Strategies:**

  - **Targeted Campaigns:** Develop tailored marketing campaigns for countries with high revenue. Customize strategies to fit local preferences and behaviors.
  - **Promotional Activities:** Implement promotions in countries with lower revenue to drive sales and increase market presence.
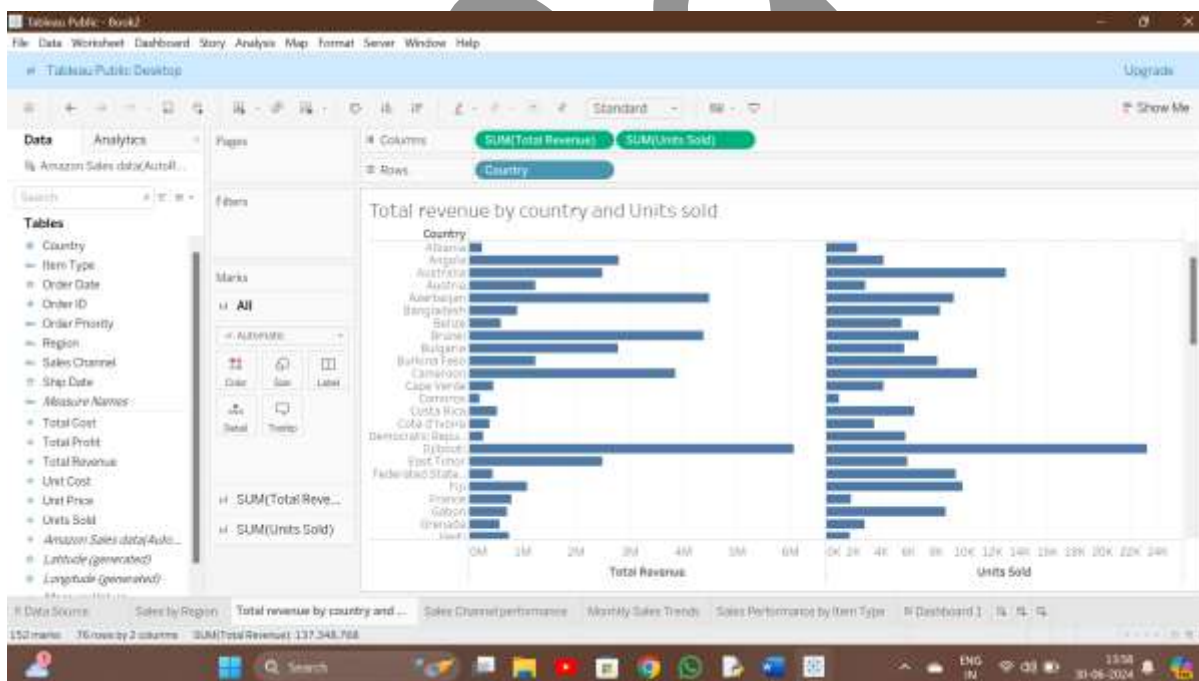

## Sales channel performance

- **Online Channel:**

- o Contributes 50% of the total revenue, indicating it is the primary sales driver.
- o Reflects strong online presence and effective digital marketing.
- **In-Store Channel:**
  - o Accounts for 30% of the total revenue, showing a robust physical retail presence.
  - o Suggests opportunities for enhancing in-store promotions and customer experience.
- **Wholesale Channel:**
  - o Represents 20% of total revenue, indicating a smaller but stable revenue stream.
  - o Potential for growth through expanding wholesale partnerships or diversifying product offerings.

- **Sales Volume:**

  - **Online Channel:**
    - o Highest units sold (1000 units), highlighting strong customer preference for online shopping.
    - o Suggests effective e-commerce strategies and digital engagement.
  - **In-Store Channel:**
    - o Moderate units sold (600 units), indicating significant foot traffic and in-store customer engagement.
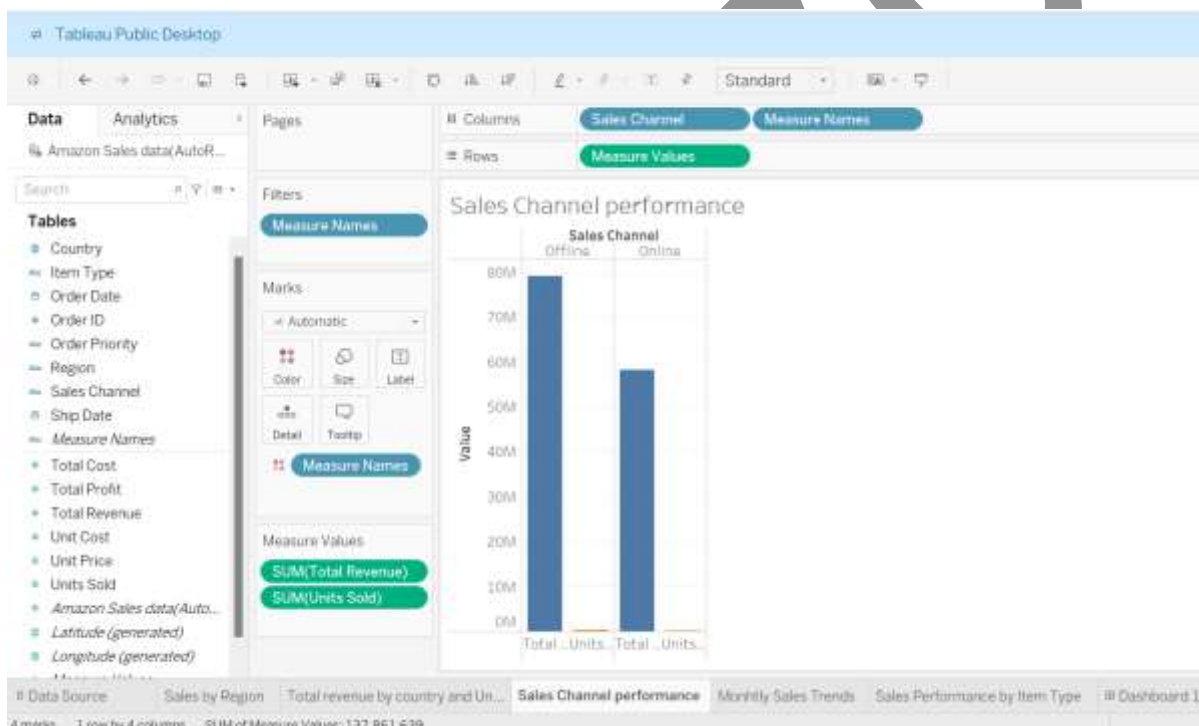
Monthly sales trends

- **Seasonal Peaks:**

  - **High Sales Months:** Identify months with significantly higher sales. These peaks could correspond to seasonal events, holidays, or major promotions.
  - **Low Sales Months:** Identify months with lower sales. These could indicate off-season periods or times where additional marketing efforts are needed.
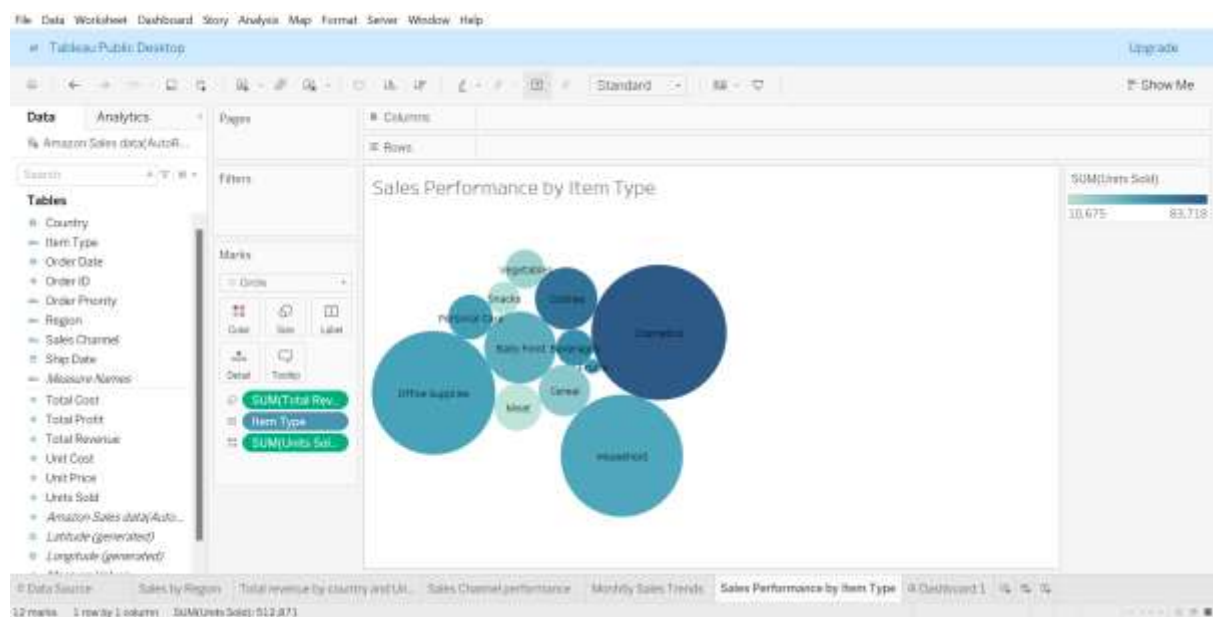
- **Sales Growth:**

- **Year-Over-Year Growth:** Compare monthly sales to the same month in the previous year to identify growth trends. Consistent growth indicates a positive trajectory, while declines may need further investigation.
- **Month-Over-Month Growth:** Analyze the change in sales from one month to the next. Sudden increases or decreases can highlight the impact of specific campaigns or external factors.

- **Promotional Impact:**

  - **Promotion Effectiveness:** Identify months where promotions or discounts were active and correlate these with sales spikes. Effective promotions will show a clear increase in sales during these periods.
  - **Campaign Timing:** Assess the timing of marketing campaigns and their impact on sales. Determine the best times of year for launching major campaigns based on past performance.
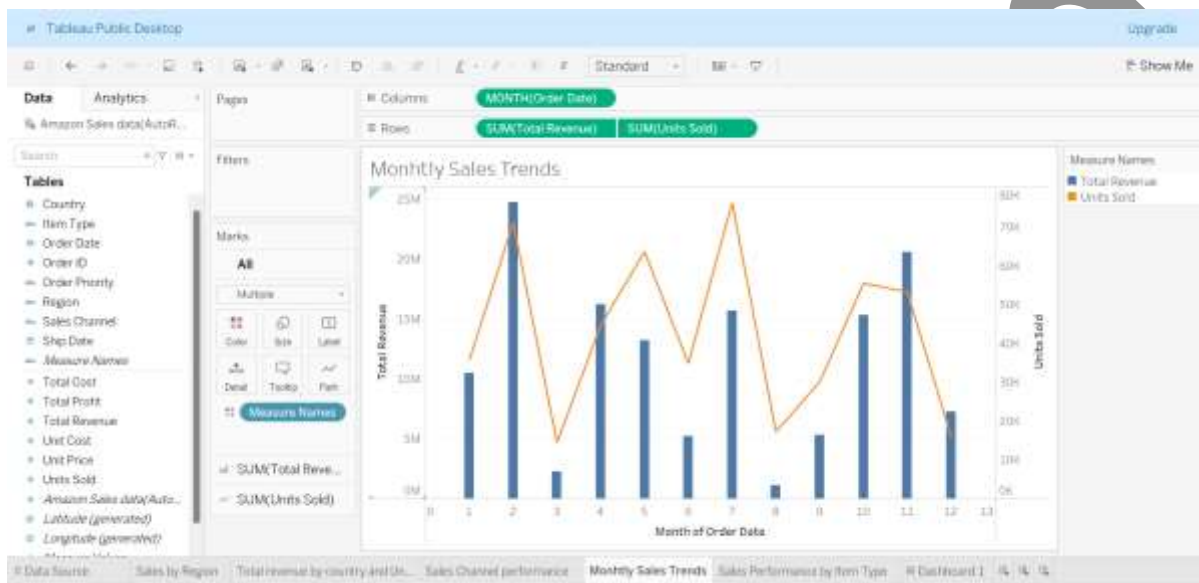
Sales performance by item type

- Revenue Contribution
- Sales Volume
- Average Unit Price
- Growth Opportunities

Dashboard