**Objective Questions**

1. What is the total number of attributes in the customer table?

There are **3** attributes in the customer table.

1. How will you get the “Customer’s” ages in the “Order” tables according to customer IDs?

We can use **New Column** in Table Tools to create new column for customer’s age in orders table and use formula

Customer's Age = LOOKUPVALUE(Customers[CustomerAge],Customers[CustomerID],Orders[CustomerID])

this will get customer’s age from customers table using Customer ID

1. In analyzing the dataset with Power BI, ensure data cleaning to address inconsistencies and missing values before further analysis.

To address inconsistencies and missing values in data. I first removed some null columns which were present in Orders Table using **Remove Column**. Then there were some missing values in different columns of orders table that were addressed by using **Remove Empty**.

1. How can we calculate the total revenue generated by all the sales?

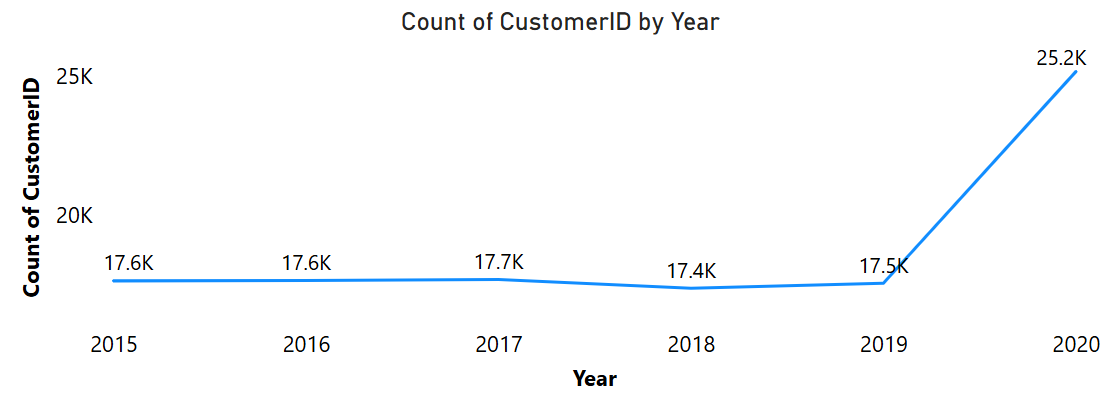
To calculate the Total Revenue we can use New Measure. Using

Total Revenue = SUMX(Orders, Orders[Order Quantity]\*Orders[Unit Price])

Total Revenue was **105.94M.**

1. What is the total number of unique customers who made purchases each year? Is there any increase in the number over the years?

Total number of customers were almost same from 2015 to 2019 by from there it increase significantly from **17.5k to 25.2k.**



1. How can we determine the total number of unique products available in the company?

To count the total number of unique products in the company we can simply use **card visualization** we can place product in value and use count distinct to calculate the total unique products from orders table, which is **44.**

We can also create new measure for same.

1. What is the average number of days it takes for products to be delivered, get the metric for only the delivered orders.

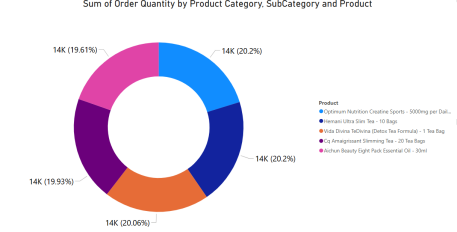
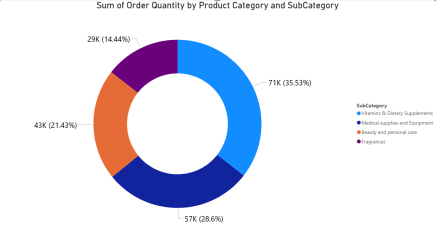
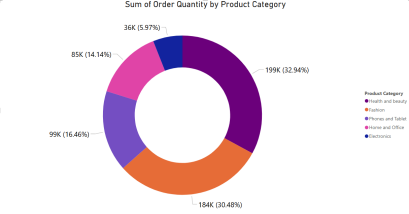
To calculate the average number of delivery days we can create a measure using formula

Average Delivery Time = CALCULATE(AVERAGEX(Orders, Orders[Delivery Date] - Orders[OrderDate]), Orders[Status] = "Delivered")

and we can use card visualization to visualize the number which is **9.41 Days.**

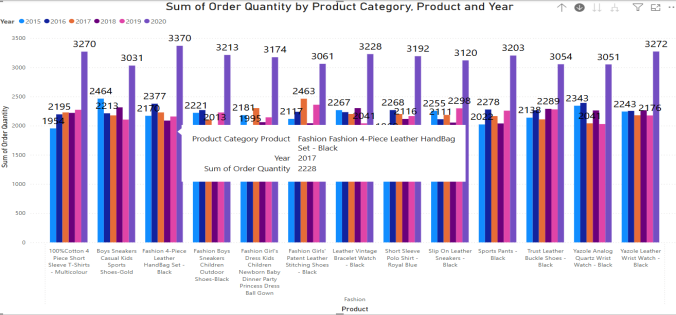
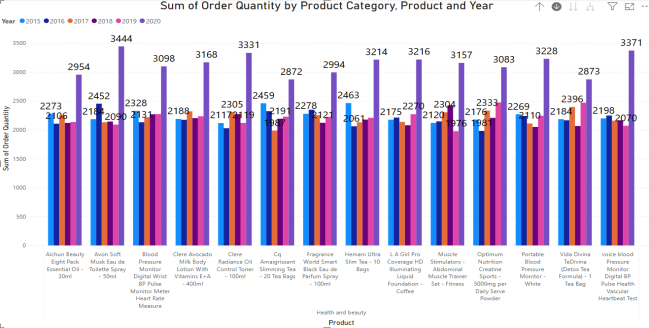
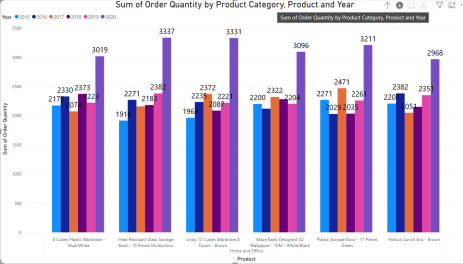
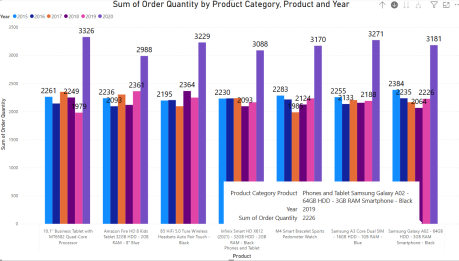
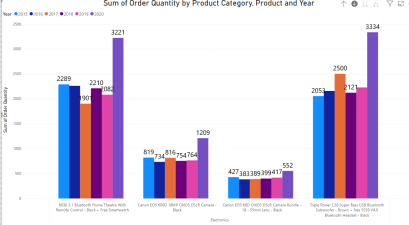
1. Which products, categories, and subcategories are the most popular?

To find out which product category, sub category, product are most popular, we will first create hierarchy product category -> sub category -> product. Then we will use sum of quantity as value and create a pie chart. Then we will use drill through and find out the most popular popular. We find out that **Health and Beauty** is most popular among popular categories with **32.94%.** Inside Health and Beauty Category **Vitamins and Dietary Supplements** is the most popular sub-category with **35.53%.** and inside this subcategory most popular products are **Optimum Nutrition Creatine Sports** and **Hemani Ultra Slim Tea** with **20.2% each.**



1. Which products have seen an increase or decrease in sales over the year?

From year 2015 to 2019 the sales have not changed much but in **2020 All the Product have increased sales**.



1. While modeling the data relationships, what will be the type of relationship between the customer ID of Orders and customer tables?

Relationship type between the customer ID of Orders and customer tables is **many to one.**

1. How have you handled the null values in the data?

To address missing values in data. I first removed some null columns which were present in Orders Table using **Remove Column**. Then there were some missing values in different columns of orders table that were addressed by using **Remove Empty**. There were also very few missing values in gender column in customers table that was handled using **Fill Down** in transform data

1. Were there any data format issues in the data, and if there were/are how you would handle them?

There were not any issues with data format in the data.

1. When we add a column in Power Query what’s the code that comes in M language in the formula bar? What do you know about M-query?

= Table.AddColumn(#"Previous Step", "Custom", each [Unit Price] \* [Order Quantity])

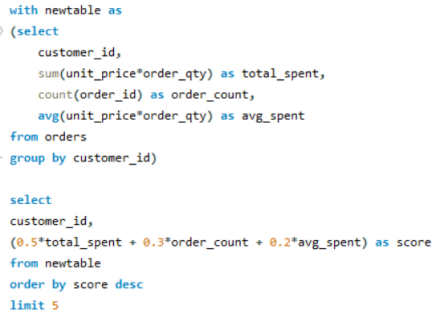
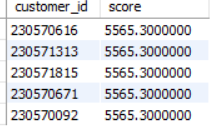
This adds a column named Custom by multiplying **Unit Price** and **Order Quantity**.

**About M-query:**

* M is the language used in Power Query for data transformation.
* It is case-sensitive and follows a step-by-step structure.
* Used to clean, shape, and prepare data before loading into Power BI.

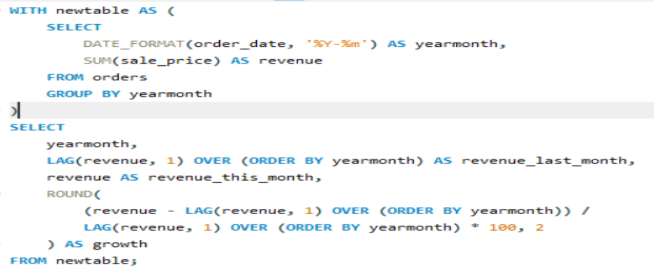
1. Identify the top 5 most valuable customers using a composite score that combines three key metrics:(SQL)
   1. Total Revenue (50% weight): The total amount of money spent by the customer.
   2. Order Frequency (30% weight): The number of orders placed by the customer, indicating their loyalty and engagement.
   3. Average Order Value (20% weight): The average value of each order placed by the customer, reflecting the typical transaction size.

Used SQL query to find the top 5 customers using the weighted conditions provided.

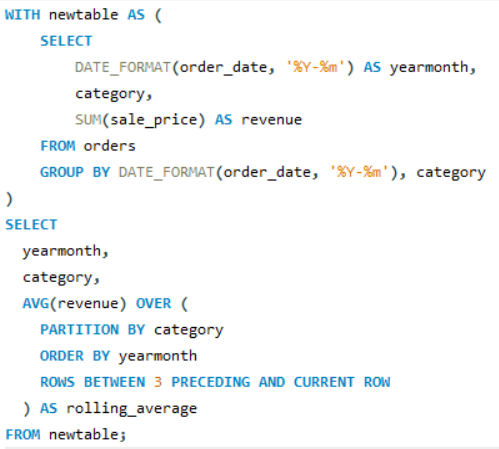
1. Calculate the month-over-month growth rate in total revenue across the entire data-set. (SQL)

Used SQL query to calculate month on month revenue growth. Maximum percentage growth was in Jan 2020 which is **118%** and lowest percentage growth was in June 2020 **-56%**.

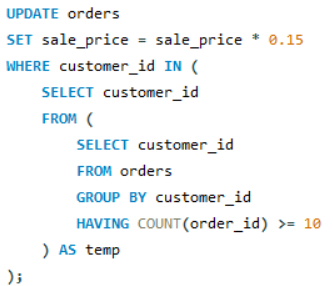
1. Calculate the rolling 3-month average revenue for each product category. (SQL)

Used SQL query to calculate 3 month rolling average product category wise.

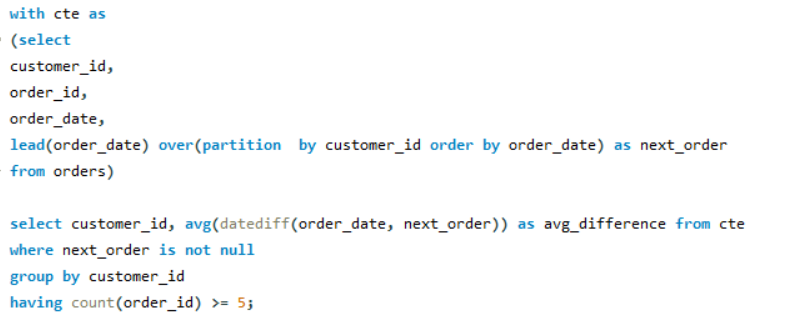
1. Update the orders table to apply a 15% discount on the `Sale Price` for orders placed by customers who have made at least 10 orders. (SQL)

Used SQL query to update the order table sale\_\_price column with 15% discount where order\_count is at least 10.



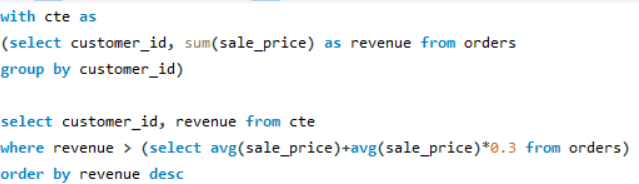
1. Calculate the average number of days between consecutive orders for customers who have placed at least five orders. (SQL)

Used SQL query to find the average number of days between consecutive orders from customers where order count is at least 5. There was no such entry found.



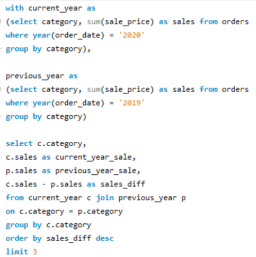
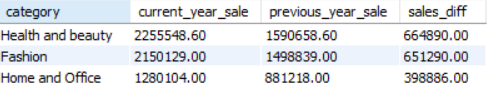
1. Identify customers who have generated revenue that is more than 30% higher than the average revenue per customer. (SQL)

Used SQL query to identify the customer where revenue is 30% higher than average revenue.

1. Determine the top 3 product categories that have shown the highest increase in sales over the past year compared to the previous year. (SQL)

Used SQL query to find out the top 3 product categories that have shown the highest increase in sales over the past year compared to the previous year.

**Subjective Question:**

1. Explain the revenue breakdown by year and by-product. Evaluate how different products contribute to annual revenue and come up with suggestions to increase the sales of the low-selling items.

I created a table with columns product category, product, year and revenue contribution in total revenue. Sorted by revenue contribution column in descending order. The Top 2 product with most contribution in revenue was found to be

1. **Electronics - Canon EOS 600D 18MP CMOS DSLR Camera - Black - 2020 -3.10%**
2. **Electronics - Canon EOS 60D CMOS DSLR Camera Bundle - 18 - 55mm Lens - Black -** **2020 - 2.45%**

And bottom 2 products with least contribution in revenue was found to be

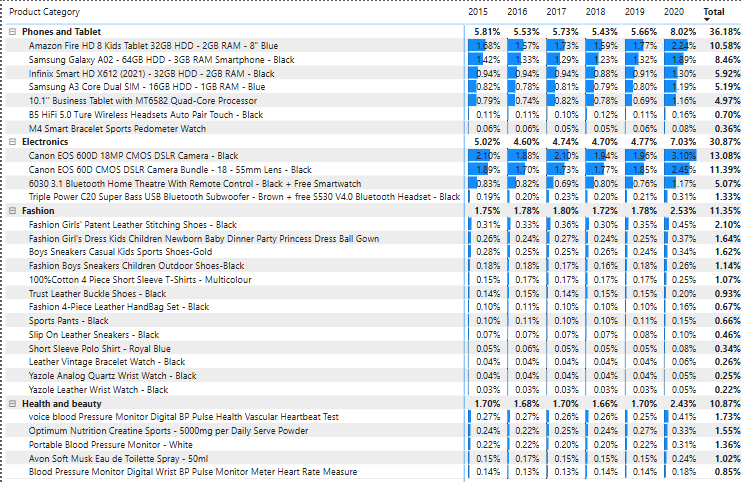
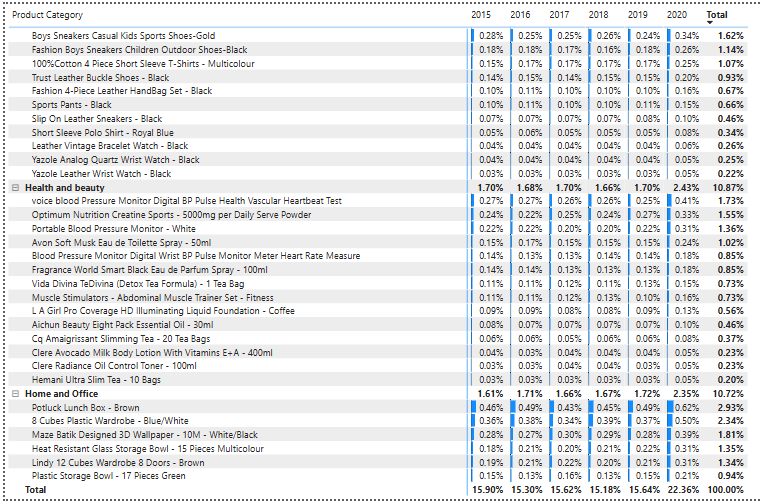
1. **Health and beauty - Hemani Ultra Slim Tea - 10 Bags - 2017 - 0.03%**
2. **Health and beauty - Hemani Ultra Slim Tea - 10 Bags - 2016 - 0.03%**

After evaluating the table I found out the products contributing the most in revenue belong to electronics category and in year 2020 but the products contributing the least in revenue belong to Health and Beauty category and year is 2016 and 2017.

When I looked at overall revenue breakdown products belonging to **Health and Beauty**, **Fashion**  are usually contributing the least in revenue.

The possible reason for that could be there are 14 and 13 products in those categories, which is very much compared to other categories. Managing inventory with these much number of products could be tough.

To increase the sales of these low selling products, we can either remove some products from these categories to make sure the number of products in these categories is also around 4-5 like other good performing categories or we can divide these products into different categories so these categories can be managed easily.

1. How many products were returned? Use a DAX function to get this metric. Examine the possible reasons for returns and consider how this metric could indicate improvements in product descriptions or quality control.

Total products returned over these years is **30526**. DAX function used is

Returned Orders = COUNTROWS(FILTER(Orders, Orders[Status] = "Returned"))

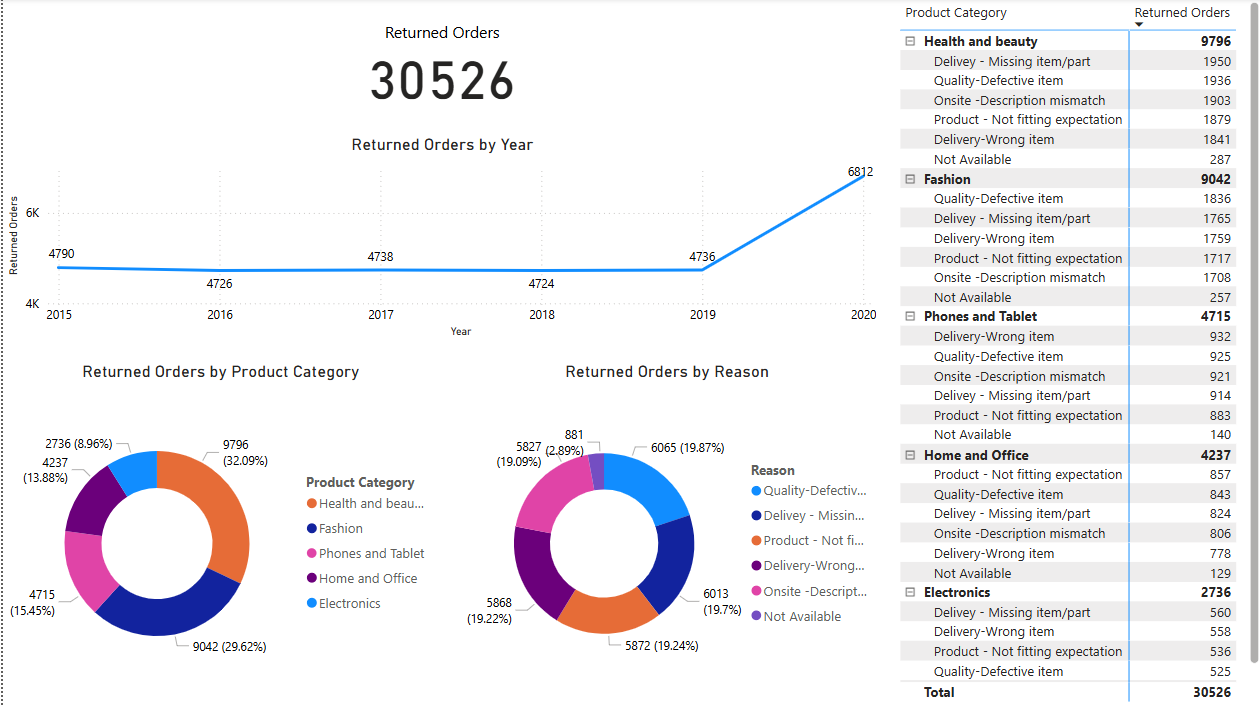
**Health and Beauty** Category has the most number of returns. Reason with most number of return is **Delivery - Missing item/part** in this particular category.

If we look at the reason wise return distribution there is not much different as such. The number of returns has increased significantly in **year 2020.**

To improve this we need to closely identify the reasons associated with most number of returns in each category. There might be specific reasons associated with most returns in specific categories,

for example:- in Fashion category the most reason for return is Quality- Defective item, second most reason for returns is Delivery- Missing item/part and next Delivery- Wrong item.

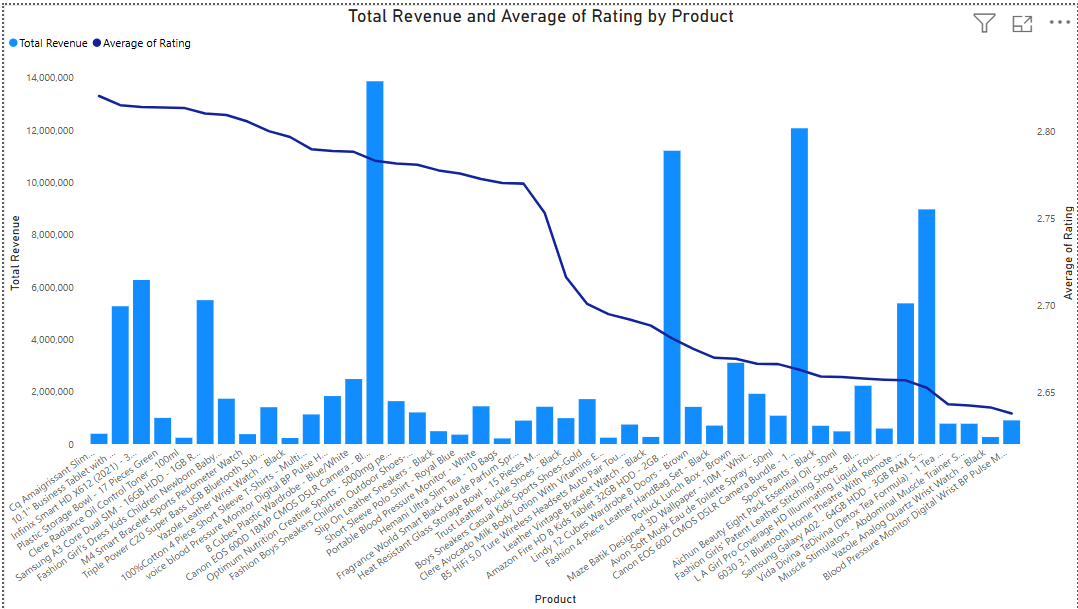
To improve the condition first we will try to improve quality related issues which might be done by clearly mentioning the description of the products and assigning quality check and then to handle delivery related returns we can assign item check before packaging which will solve the problem of Missing item/part and wrong item.



1. Whenever a customer goes to Amazon, they’ll filter the most rated products to buy the better category. Can you verify this using any visualization or table that the ratings of products impact their sales value?

I created a **Line a Clustered Column chart** using total revenue and average rating of each product.

By looking at this chart we can say that most of the products with top revenue were associated with lesser average rating whereas most products with minimum revenues are associated with maximum average rating.



We can rely on this chart as almost all products are having similar count of ratings.

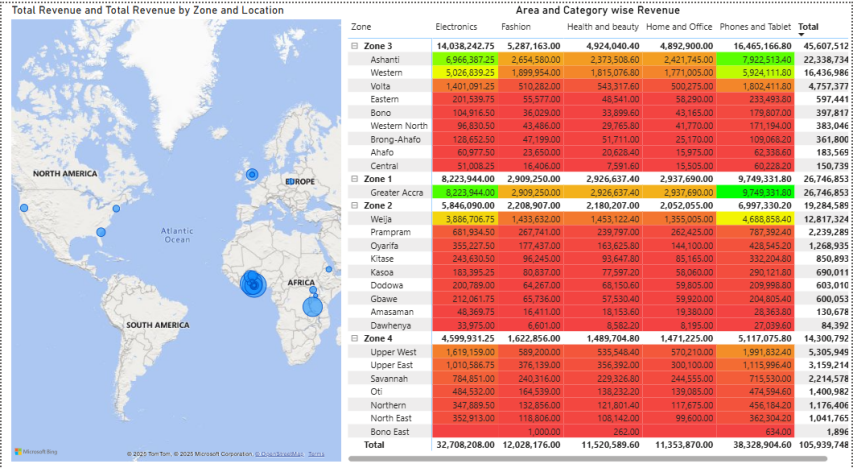
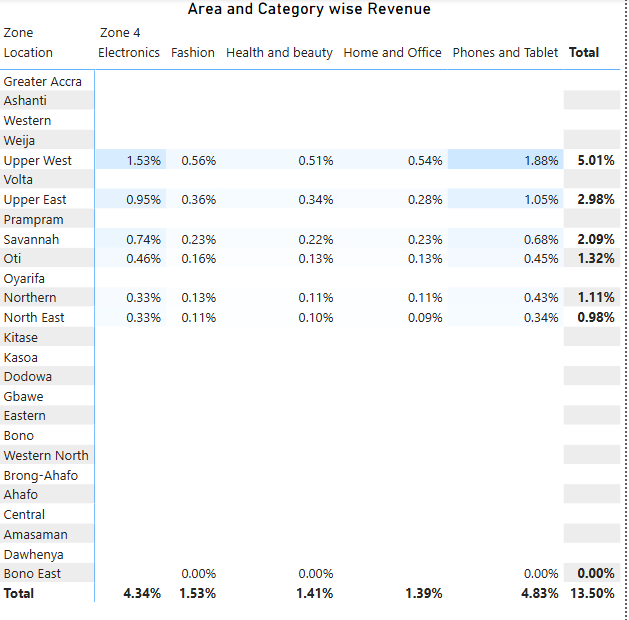
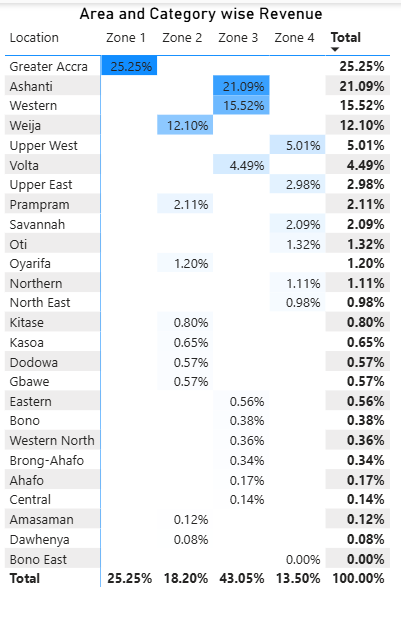
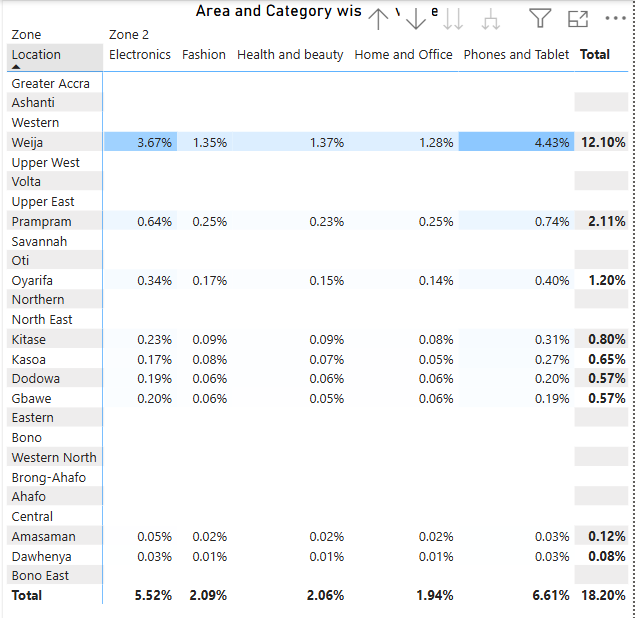
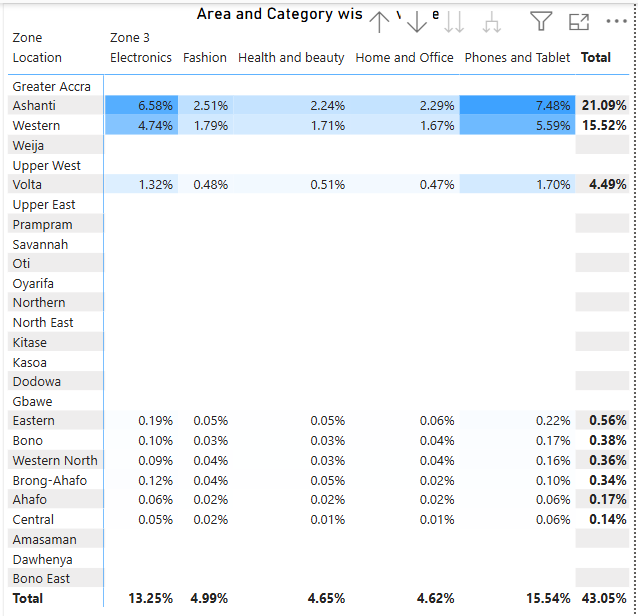
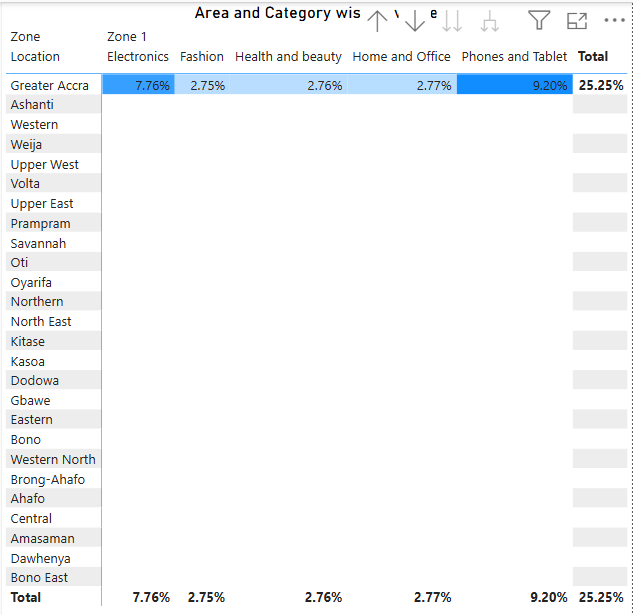
Overall we can say that **higher product rating is not always associated with higher sales revenue.**

1. Investigate how revenue distribution varies across different locations. Explore which geographical areas contribute most to sales and consider the strategic implications for regional marketing and distribution efforts. How might location-based trends inform the company's market segmentation and resource allocation approach?

I created a **Map chart and a table** using Zone, Location, Category and Revenue.

By looking at the map chart we can say most revenue is coming from **Zone 3 followed by Zone 1** and Categories contributing the most in revenue are **Phone and Tablet and Electronics** and least contributing categories in revenue are **Home and Office, Health and Beauty and Fashion.**

In all the Zones revenue from top 2 categories are contributing more than double of revenue from bottom 3 categories combined.

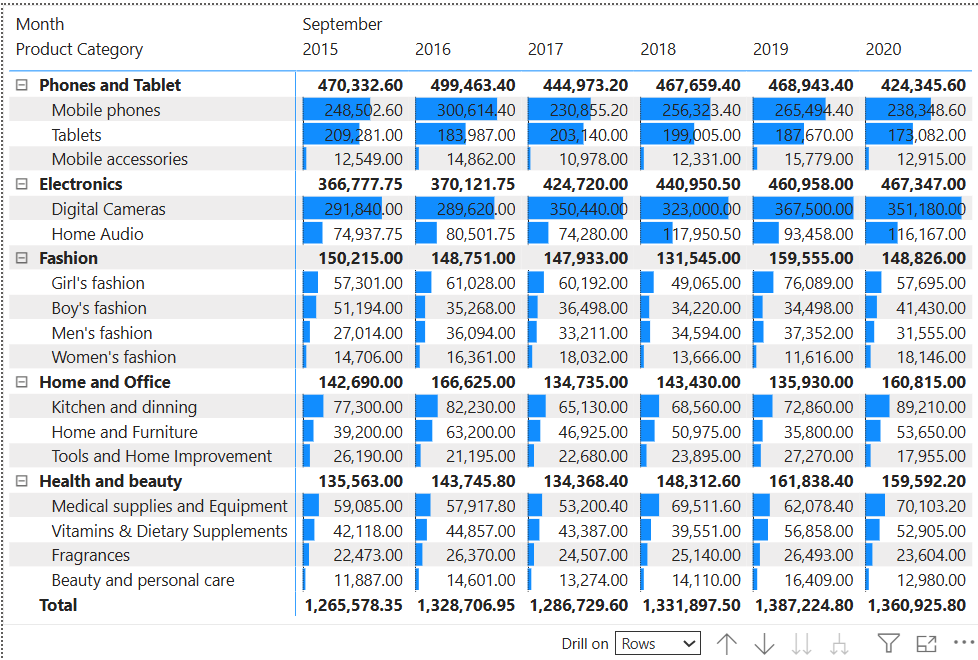
By looking at zone and category wise revenue distribution we can say that revenue distribution across categories is similar in all regions.

There are certain areas in different zones that are not really contributing in the revenue, in fact there are some areas which are contributing even less than **1%** all categories combined.

1. Determine which month could benefit from enhanced promotional offers to boost sales. Can you suggest some targeted marketing strategies here?

**September** month is contributing the least in revenue so enhanced promotional offers can help boost sales for this month.

Looking at the revenue distribution category wise for **September** month over the years we can say that targeted marketing strategies dedicated to **Fashion** category can help increase the revenue of September month as this is the only category where the revenue each year is decreasing contentiously.

1. Identify which products may require increased marketing efforts. Which items have high prices yet underperform in sales?

There are several products which have contributed the least in revenue. Although the average unit sales is similar to the average of other products but due to the unit price which is least. The sales revenue of these product is also least compared to other products.

**Health and beauty - Hemani Ultra Slim Tea - 10 Bags - 5.55 - 15.00 - 94.80**

**Health and beauty - Clere Radiance Oil Control Toner - 100ml - 5.52 - 17.00 -105.34**

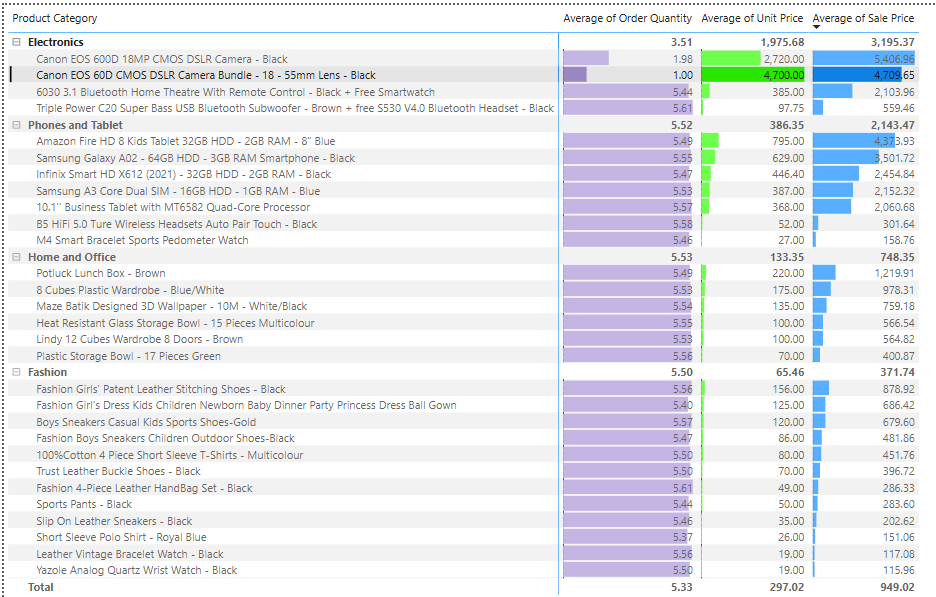
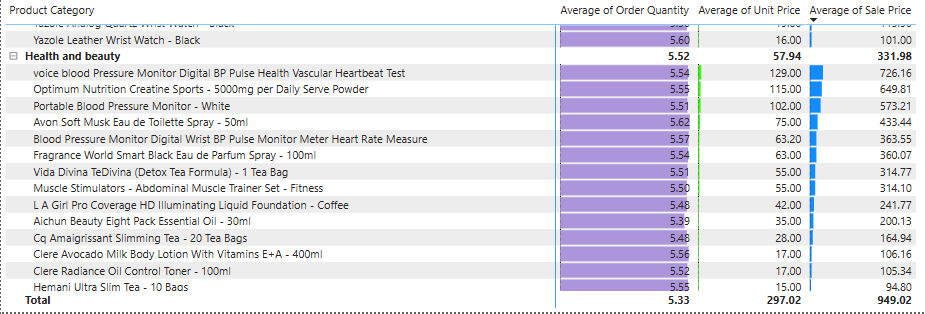
Increased marketing efforts for these products might also help increase the sales revenue and sales.

There are some products where in spite of having highest unit price the contribution in sales is not the highest like

**Electronics - Canon EOS 60D CMOS DSLR Camera Bundle - 18 - 55mm Lens - Black - 1.00 - 4,700.00 - 4,709.65**

This is mainly because average quantity sold is very low compared to other products sold.

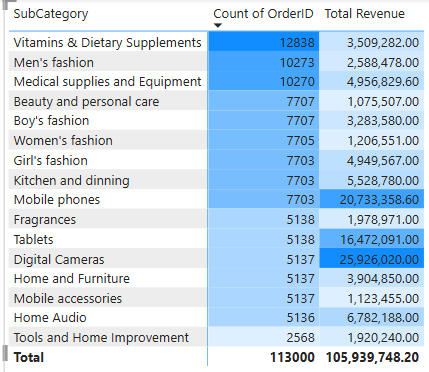
If we can apply selective marketing strategies on these products and increase the average units sold the revenue contribution of these product can increase significantly.

1. Assess which products should have discounts. How can targeted incentives drive sales and customer loyalty for specific products?

There are some products which contributes most in the revenue mostly these products belong to sub-category **Digital Cameras, Mobile phones** but when it comes to quantity sold these subcategories are contributing very less.

If products belonging to these subcategories have discount there might be increase in quantity sold for that product which will ultimately increase the revenue for those products.



To drive sales and customer loyalty for specific products we need to figure out some products where repeat order value is high top 3 sub categories where order quantity is high **Vitamins & Dietary Supplements, Men's fashion, Medical supplies and Equipment**. If we can achieve some targeted incentives the sales and customer loyalty for these products will increase.

1. Come up with a loyalty program to benefit the company’s customers. From the available lot of customers come up with strategies to bucket them and provide benefits under different loyalty programs.

In our current data-set there is no customer with repeat orders. If we want to build loyalty program, we can achieve this using different approaches.

First approach can be, If a customer is placing repeated orders we can put that customer into different tiers like Platinum, Gold, Silver etc these tiers can be identified using number of orders placed and revenue generated.

Different discounts can be provided in different tiers on repeated orders and repeated orders can be rewarded to encourage customer loyalty.

Another approach can be, If we bucket customers based on the categories they are purchasing in, so we can reward customers based on the categories. Different discounts can be implemented based on the tier inside the category.

Also cross category promotion can also be done where achieving a particular tier in a category will unlock some discounts or rewards in another category to encourage customer loyalty.

1. Using the DAX functions Calculate and a row iteration DAX function calculate the total sales for the Product Category “Fashion” and delivery type “Shipped from Abroad”. What are the other types of DAX functions you have used in the project?

SalesDAX = CALCULATE(

    SUMX(

        Orders,

        Orders[Sale Price]

    ),

    Orders[Product Category] = "Fashion",

    Orders[Delivery Type] = "Shipped from Abroad"

)

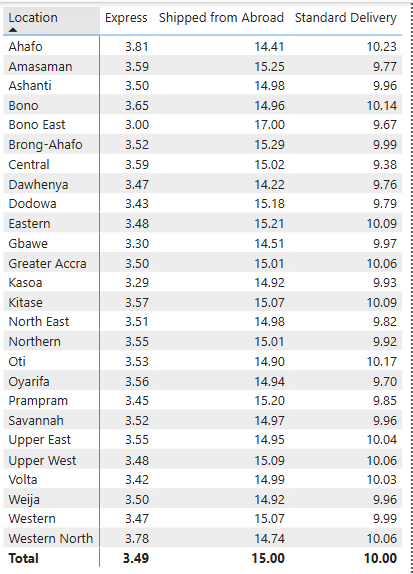
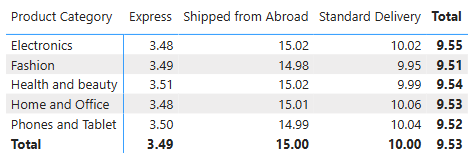
By using the DAX function mentioned above the amount of was **Rs 41,35,434.**

I have used different DAX functions in this project like **New Measure** for calculation of Total Revenue, Number of Orders Returned, Average Delivery Days and **New Column** for Customer Age using Customer ID.

1. Wait Times Correlated with Demographics and Care: Explore how average wait times vary across different product categories to optimize scheduling and staffing.

Average Wait Days for different locations and for different delivery types is calculated using table.

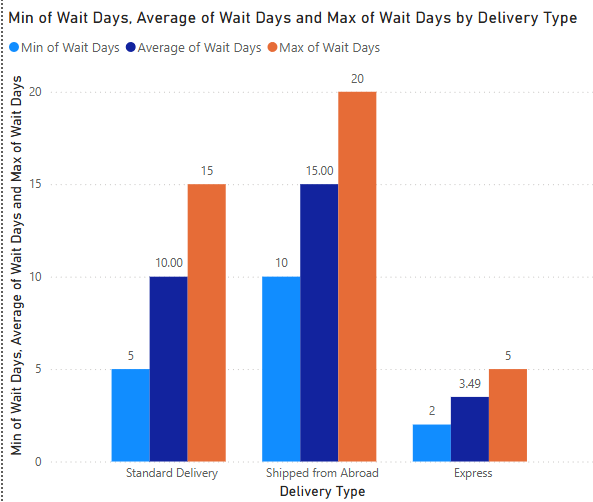
For Express delivery type the average wait days is 3.49 almost same for all locations. In case of Standard Delivery the average wait days is 10 almost same for all locations but in case of Shipped from Abroad the average wait days is 15 which is not same for all locations. There is difference of 1-3 days for different locations.

But looking at average wait days for different categories across different delivery types is similar across all product categories and average wait days across different categories for all delivery types is almost **9.5** days.

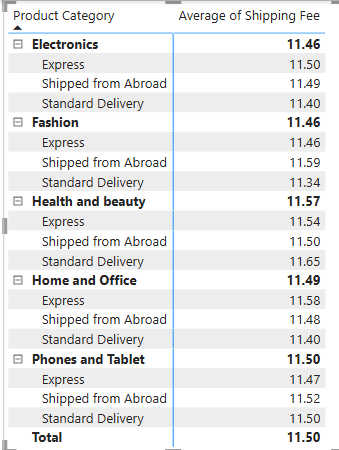
1. Explore if there is any relationship between the Delivery type and waiting time between ordering and receiving an item.

There is a direct relation between Delivery Type and Waiting Time. The least number of waiting time is associated with **Express Delivery Type - Min 2 and Max 5** followed by **Standard Delivery Type - Min 5 and Max 15** and least number of waiting time is associated with **Shipped from Abroad Delivery Type - Min 10 and Max 20.**



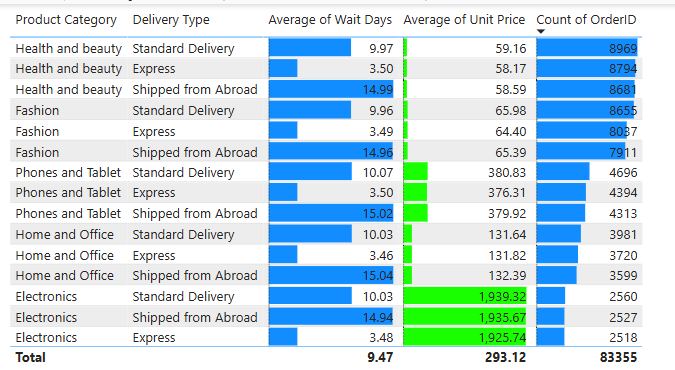
1. Is there any relationship between shipping charges and product type?

There is no relationship between shipping charges and product type. Average shipping price for different product categories and different delivery type is similar which indicates that shipping price is not associated with product categories.



1. Come up with strategies to decrease the low rating orders after analyzing different factors like waiting time, shipping type, unit price, etc.

There is 83355 orders with below average rating. By looking at the table we made from categories, delivery types, wait days and unit price we can see that there is a direct connection between low rating orders and categories and unit price.



Maximum number of below average rating orders are associated with Health and Beauty category followed by Fashion category. Also the relation between below average rating and unit price is disproportional, lower the unit price higher count of below average rating orders.

Now we understand the relation between low rating order and unit price and categories. To come up with strategies to reduce the number of low rating orders-

We can identify the products which are contributing the most number of low rating orders across these categories. Also we can improve our rating systems, we can either add a comment section in rating where costumer can describe reason behind his/her rating or we can make rating based on different parameters like quality, price, delivery experience etc.

Based on this rating system we can find out the rating for different parameters of the product which will help remove the problem and increase the customer satisfaction and ultimately reducing the number of low rating orders.

1. Using the time intelligence DAX function, create a table to compare each month’s sales with the previous year’s same month’s total sales. So there will be four columns in the output year, month, total sales, previous\_years\_sales.

Sales\_Comparison\_Table =

ADDCOLUMNS(

    SUMMARIZE(

        'OrderDate',

        'OrderDate'[Year],OrderDate[Month No]

    ),

    "Total\_Sales",

        CALCULATE(

            SUMX(

                Orders,

                Orders[Order Quantity] \* Orders[Unit Price]

            ),

            Orders[Reason] = "Not Available"

        ),

    "Previous\_Years\_Sales",

        CALCULATE(

            SUMX(

                Orders,

                Orders[Order Quantity] \* Orders[Unit Price]

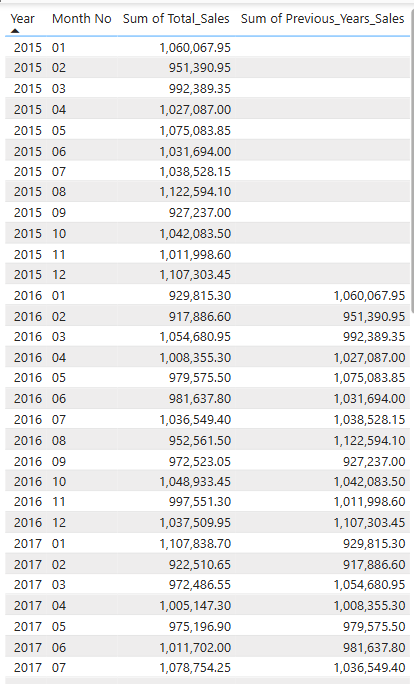
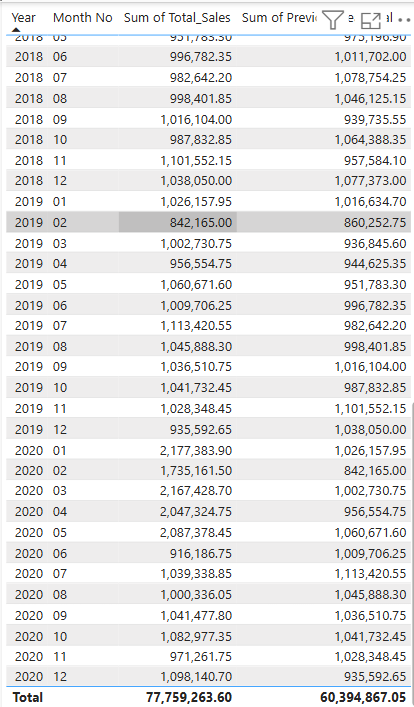
            ),

            SAMEPERIODLASTYEAR('OrderDate'[OrderDate]),

            Orders[Reason] = "Not Available"

        )

)

1. What do you understand by PowerBI gateway? What are its use cases?

**Power BI Gateway:** Power BI Gateway acts as a secure bridge between your on-premises data (like SQL Server, Excel, or Oracle) and the Power BI cloud service. It lets you keep your data where it is, while still refreshing or querying it from Power BI online.

**Use:**

* Scheduled refresh of on-premises data in Power BI reports
* Live or DirectQuery connections for real-time dashboards
* Works with Power BI, Power Apps, Power Automate, and more
* Helps maintain data privacy without needing to move data to the cloud

**Types:**

* Personal Gateway – for individual use; only works with Power BI
* Standard Gateway – for teams; supports multiple data sources and services

**Common Use Cases:**

* Refreshing reports from on-prem databases
* Connecting to SQL Server Analysis Services (SSAS) live
* Keeping sensitive data on-site while analyzing in the cloud
* Enterprise-wide reporting across mixed data environments

1. How would you approach this problem, if the objective and subjective questions weren't given?

If questions weren’t given, I would:

* Understand business goals – Identify what the stakeholders want (e.g., increase sales, improve delivery, reduce returns).
* Explore the data – Check tables, columns, relationships, and clean the data.
* Define key metrics – Like Total Sales, Revenue, Customer Satisfaction, Returns, Ratings, etc.
* Create meaningful visuals – Build dashboards with slicers (e.g., product, location, date) to find trends, outliers, and insights.
* Draw conclusions – Use data to recommend actions like discounts, product focus, or loyalty programs.