Objective Questions:

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

Customer Table has3 Attributes -
***CustomerID** : Unique Identifier for a Customer.*
***Customer Age** : Age of Customer in Years.*
***Customer Gender** : Gender code for Male & Female namely M, F respectively.*

CustomerID	Customer Age	Customer Gender
230459067	38	M
230459068	38	M
230459069	38	M
230459072	38	M
230459073	38	M
230459074	38	M

—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

2. How will you get the "Customer's" ages in the "Order" tables according to customer IDs?

*The **RELATED** function in DAX is used to fetch a value from a related table when a relationship exists between tables. It works in a **many-to-one** relationship, retrieving values from the **one-side** of the relationship.*
Here, Orders table has a Many to one relation with Customers table.
Hence, using Related Function we can bring Age into Orders Table.

DAX Formula to Add Customer Age Column in Orders Table -

```
Customer Age = RELATED(Customers[Customer Age])
```

—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

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

Data was Cleaned and Transformed before Analysis, which involves -

- *Removing Duplicate records by selecting the whole table and then using Remove Duplicates in the Home tab.*
- *Removing whitespaces, selecting the whole table & then Transform > Format > Trim.*
- *Changing Data Types to most appropriate ones using Detect Data Type & then Manually for incorrect data types.*
  Example- CustomerID & OrderID data type was changed from Numeric to Text.
- *Finding Missing Values in Customers[Gender] & Orders[Category] using Column Quality option and replacing them with Appropriate Values using Replace Values.*
- *Created Calculated Column for Gender by Replacing M = Male & F = Female and*
- *Age Group Column using Add Columns from Examples option.*
- *Formatting Date columns to DD/MM/YYYY Format.*

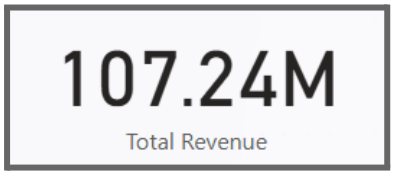
—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

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

Using calculate function & finding sum of Sale Price ignoring All the filters will give total revenue generated by all the sales.

DAX formula -

```
Total Revenue = CALCULATE(SUM(Orders[Sale Price]),ALL(Orders))
```

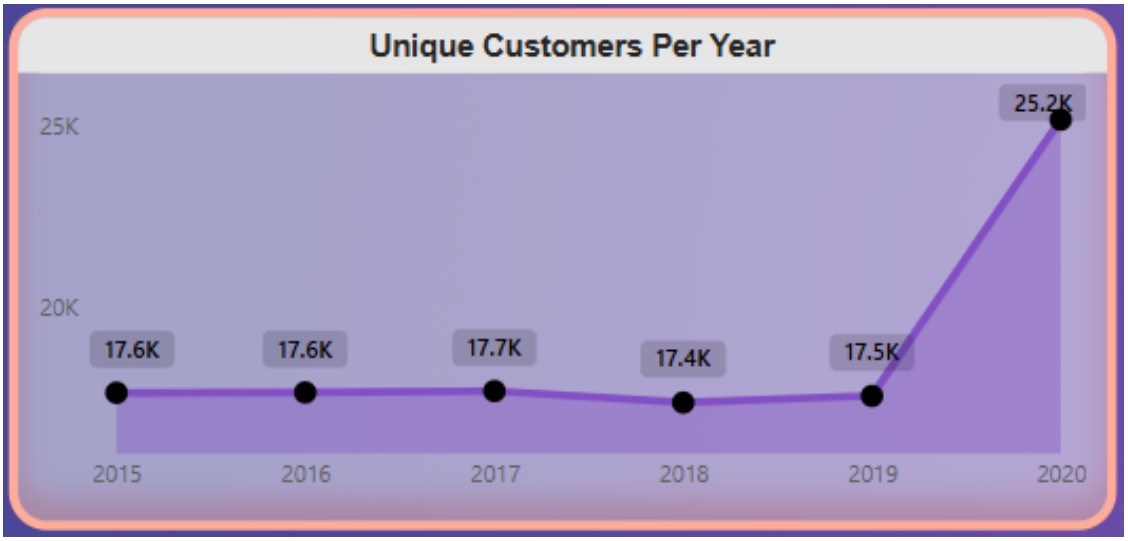


—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

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

This can be determined using Line Chart by Plotting Years of Orders[OrderDate] on X-axis & Distinct Count of Orders[CustomerID] on Y-axis.

Number of customers has a very marginal increase from 2015 to 2017, then a minor dip in 2018, then slight increase in 2019 but a very significant increase in numbers is observed for the year 2020.



—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

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

We can create a Dax formula using the DistinctCount function on Orders[Product].

DAX formula -

```
Total Products = DISTINCTCOUNT(Orders[Product])
```

There are a total 44 unique products available in the company.

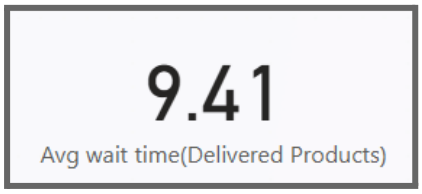
—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

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

This can be achieved using a combination of Calculate, AverageX, datediff & Status="Delivered".

DAX formula -

```
Avg wait time(Delivered Products) = CALCULATE(AVERAGEX(Orders, DATEDIFF(Orders[OrderDate], Orders[Delivery Date],
DAY)),Orders[Status]="Delivered")
```



Average wait time for delivered products is 9.41 days.

—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

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

Considering Total Revenue generated as the Primary parameter for popularity. Created a matrix showing Total Revenue, Total Qty & Total Orders. For Product Category > Sub-Category > Products.
Applied Conditional Formatting For Visual Analysis, Here Darker shade means higher Value.

Product Category	Sum of Sale Price	Sum of Order Quantity	Count of OrderID
⊞ Phones and Tablet	₹ 3,85,35,343	99249	17978
⊞ Electronics	₹ 3,28,26,079	36017	10273
⊞ Fashion	₹ 1,24,11,565	183726	33388
⊞ Health and beauty	₹ 1,19,35,540	198549	35952
⊞ Home and Office	₹ 1,15,30,632	85265	15408
⊞ Not Specified	₹ 139	1	1
Total	**₹ 10,72,39,297**	**602807**	**113000**

SubCategory	Sum of Sale Price	Sum of Order Quantity	Count of OrderID
Digital Cameras	₹ 2,59,84,875	7663	5137
Mobile phones	₹ 2,08,22,208	42505	7703
Tablets	₹ 1,65,30,526	28404	5138
Home Audio	₹ 68,41,204	28354	5136
Kitchen and dinning	₹ 56,17,298	42638	7703

Product	Sum of Sale Price	Sum of Order Quantity	Count of OrderID
Canon EOS 600D 18MP CMOS DSLR Camera - Black	₹ 1,38,90,491	5096	2569
Canon EOS 60D CMOS DSLR Camera Bundle - 18 - 55mm Lens - Black	₹ 1,20,94,384	2567	2568
Amazon Fire HD 8 Kids Tablet 32GB HDD - 2GB RAM - 8" Blue	₹ 1,12,36,628	14097	2569
Samsung Galaxy A02 - 64GB HDD - 3GB RAM Smartphone - Black	₹ 89,95,920	14255	2569
Infinix Smart HD X612 (2021) - 32GB HDD - 2GB RAM - Black	₹ 62,99,126	14044	2566
Samsung A3 Core Dual SIM - 16GB HDD - 1GB RAM - Blue	₹ 55,27,162	14206	2568

Conclusion -

- *Phones and Tablets is the most popular Category followed by Electronics and Fashion.*
- *Digital Cameras is the most popular Sub-Category followed by Mobile Phones Tablets.*

—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

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

Created a Visual displaying Total Sales for each Product along with a sparkline for analysing YOY Sales trend. Applied Conditional Formatting on Sales value where dark shade of Grey represents high sale Value.

Product	Total Sales	Total Sales by Year
10.1'' Business Tablet with MT6582 Quad-Core Processor	52,93,898.00	
100%Cotton 4 Piece Short Sleeve T-Shirts - Multicolour	11,60,562.00	
6030 3.1 Bluetooth Home Theatre With Remote Control - Black + Free Smartwatch	54,05,068.00	
8 Cubes Plastic Wardrobe - Blue/White	25,13,287.00	
Aichun Beauty Eight Pack Essential Oil - 30ml	5,13,528.00	
Amazon Fire HD 8 Kids Tablet 32GB HDD - 2GB RAM - 8" Blue	1,12,36,628.00	
Avon Soft Musk Eau de Toilette Spray - 50ml	11,13,510.00	
B5 HiFi 5.0 Ture Wireless Headsets Auto Pair Touch - Black	7,74,909.00	
Blood Pressure Monitor Digital Wrist BP Pulse Monitor Meter Heart Rate Measure	9,33,960.60	
Boys Sneakers Casual Kids Sports Shoes-Gold	17,45,896.00	
Canon EOS 600D 18MP CMOS DSLR Camera - Black	1,38,90,491.00	
Canon EOS 60D CMOS DSLR Camera Bundle - 18 - 55mm Lens - Black	1,20,94,384.00	
Clere Avocado Milk Body Lotion With Vitamins E+A - 400ml	2,72,721.00	
Clere Radiance Oil Control Toner - 100ml	2,70,617.00	
Cq Amaigrissant Slimming Tea - 20 Tea Bags	4,23,412.00	
Fashion 4-Piece Leather HandBag Set - Black	7,35,017.00	
Fashion Boys Sneakers Children Outdoor Shoes-Black	12,37,888.00	
Fashion Girl's Dress Kids Children Newborn Baby Dinner Party Princess Dress Ball Gown	17,61,348.00	
Fashion Girls' Patent Leather Stitching Shoes - Black	22,57,056.00	
Fragrance World Smart Black Eau de Parfum Spray - 100ml	9,25,023.00	
Heat Resistant Glass Storage Bowl - 15 Pieces Multicolour	14,54,312.00	

Heat Resistant Glass Storage Bowl - 15 Pieces Multicolour	14,54,312.00
Hemani Ultra Slim Tea - 10 Bags	2,43,540.00
Infinix Smart HD X612 (2021) - 32GB HDD - 2GB RAM - Black	62,99,125.60
L A Girl Pro Coverage HD Illuminating Liquid Foundation - Coffee	6,21,098.00
Leather Vintage Bracelet Watch - Black	3,00,777.00
Lindy 12 Cubes Wardrobe 8 Doors - Brown	14,50,470.00
M4 Smart Bracelet Sports Pedometer Watch	4,07,700.00
Maze Batik Designed 3D Wallpaper - 10M - White/Black	19,49,577.00
Muscle Stimulators - Abdominal Muscle Trainer Set - Fitness	8,06,617.00
Optimum Nutrition Creatine Sports - 5000mg per Daily Serve Powder	16,68,721.00
Plastic Storage Bowl - 17 Pieces Green	10,29,029.00
Portable Blood Pressure Monitor - White	14,71,994.00
Potluck Lunch Box - Brown	31,33,957.00
Samsung A3 Core Dual SIM - 16GB HDD - 1GB RAM - Blue	55,27,162.00
Samsung Galaxy A02 - 64GB HDD - 3GB RAM Smartphone - Black	89,95,920.00
Short Sleeve Polo Shirt - Royal Blue	3,88,076.00
Slip On Leather Sneakers - Black	5,20,117.00
Sports Pants - Black	7,28,274.00
Triple Power C20 Super Bass USB Bluetooth Subwoofer - Brown + free S530 V4.0 Bluetooth Headset - Black	14,36,136.00
Trust Leather Buckle Shoes - Black	10,19,172.00
Vida Divina TeDivina (Detox Tea Formula) - 1 Tea Bag	8,08,339.00
voice blood Pressure Monitor Digital BP Pulse Health Vascular Heartbeat Test	18,62,598.00
Yazole Analog Quartz Wrist Watch - Black	2,97,912.00
Yazole Leather Wrist Watch - Black	2,59,470.00

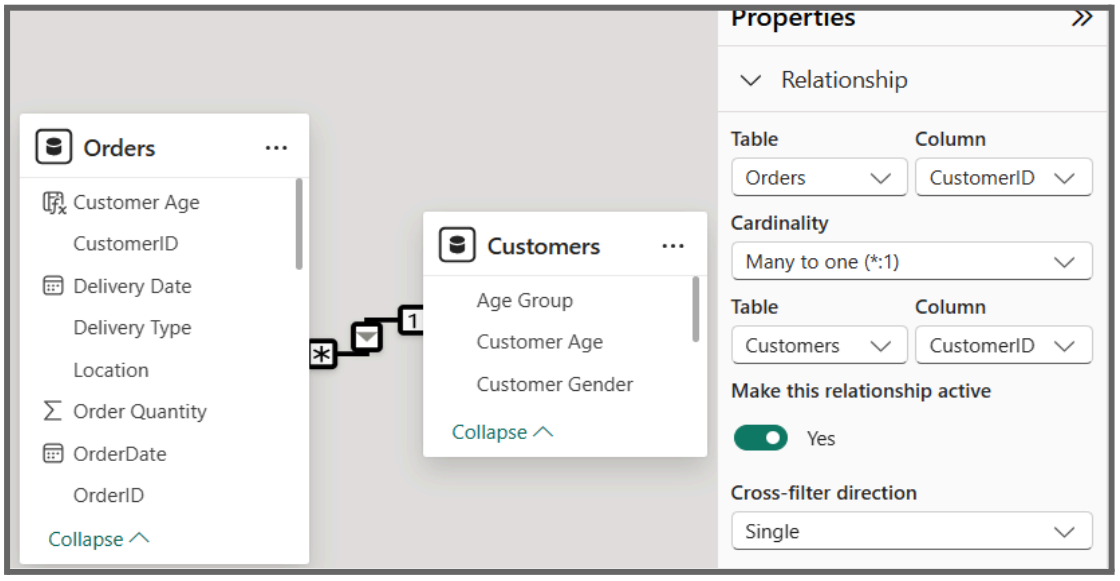
Conclusion-

As the Chart shows,
Every Single Product has produced better sales in 2020 compared to 2015.
It is confirmed that all the Products have seen increase in Sales over years.

—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

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

Orders Table has many to one Cardinality with Customers on the basis of CustomerIDs.



—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

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

Missing Values were found using the Column Quality tool available in the View Tab of Power Query Editor. Then using the replace Values option.

Blank and Nulls were found in Orders[Product Category] & Customers[Customer Gender] and replaced with "Not Specified".

—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

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

Yes, there were Certain Data Type & Format issues in the dataset, I fixed them using below mentioned Steps-

- *Changing Data Types to most appropriate ones using Detect Data Type & then Manually for incorrect data types.*
- *CustomerID & OrderID data type was changed from Numeric to Text.*
- *Formatting of OrderDate & Delivery Date was changed to DD/MM/YYYY from DD MMMM YYYY format.*

—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

13. 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?

This code appeared in the formula bar when I added a conditional Column named Gender.

```
= Table.AddColumn(#"Removed Columns", "Custom", each if [Customer Gender] = "M" then "Male"
    else if [Customer Gender] = "F" then "Female" else "Not Specified")
```

An M-query in Power BI is the code written in the M language, which is used within Power Query Editor for data transformation tasks. M is case-sensitive and supports functional programming. It allows users to perform a wide range of data manipulations, including data import, cleaning, filtering, and reshaping.This enables users to prepare and shape their data before it's loaded into the Power BI model for further analysis and visualization.

Benefits of M-Queries

- ***Customizable:*** *You can manually write or tweak the query for specific requirements.*
- ***Reusable:*** *Copy and adapt M-code for similar data transformations.*
- ***Powerful:*** *Enables complex data transformation and integration tasks.*

—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

14. Identify the top 5 most valuable customers using a composite score that combines three key metrics: (SQL)

a. Total Revenue (50% weight): The total amount of money spent by the customer.

b. Order Frequency (30% weight): The number of orders placed by the customer, indicating their loyalty and engagement.

c. Average Order Value (20% weight): The average value of each order placed by the customer, reflecting the typical transaction size.

Approach -

Used group by to aggregate values at individual level.
Created Composite_Score Column by Multiplying each value with their weight % and summing them.
Ordered output based on Composite_Score in descending and used Limit 5.

Query -

```
select CustomerID, sum(SalePrice) as Total_Revenue, count(*) as Order_Frequency,
round(avg(SalePrice),1) as Avg_Order_Value,
round(sum(SalePrice)*0.5 + count(*)*0.3 + avg(SalePrice)*0.2,1) as Composite_Score
from orders
group by CustomerID
order by Composite_Score desc
limit 5;
```

Output -

CustomerID	Total_Revenue	Order_Frequency	Avg_Order_Value	Composite_Score
230484390	8180	1	8180.0	5726.3
230482911	8180	1	8180.0	5726.3
230461607	8180	1	8180.0	5726.3
230471593	8180	1	8180.0	5726.3
230481217	8180	1	8180.0	5726.3

—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

15. Calculate the month-over-month growth rate in total revenue across the entire dataset. (SQL)

Approach -

Extracted Month & Year from OrderDate, then concatenated then together for Find Month_Year to calculate Revenue on MoM basis.
Calculated total sales for each Month_Year using sum on SalePrice and aggregating them on Month_Year level.
Used Lag function on SalePrice to find Previous_Month_Year_sale.
*Then using Formula = (Current_month_Sale - Previous_Month_sale)*100/ Previous_Month_sale, Calculated Percentage Growth.*

Query -

```sql
with cte1 as (select concat(right(OrderDate, 4), mid(OrderDate, 3, 3)) as Month_Year,
sum(SalePrice) as Curr_Month_Rev
from orders
group by Month_Year
),
cte2 as (select Month_Year, Curr_Month_Rev,
lag(Curr_Month_Rev) over(order by Month_Year) as Prev_Month_Rev
from cte1)

select Month_Year, Curr_Month_Rev, Prev_Month_Rev,
round((Curr_Month_Rev - Prev_Month_Rev)*100/Prev_Month_Rev,1) as Percentage_Growth
from cte2
order by Month_Year;
```

Output -

Month_Year	Curr_Month_Rev	Prev_Month_Rev	Percentage_Growth
2015-01	1484379	NULL	NULL
2015-02	1343872	1484379	-9.5
2015-03	1328538	1343872	-1.1
2015-04	1446957	1328538	8.9
2015-05	1517289	1446957	4.9
2015-06	1378116	1517289	-9.2
2015-07	1369656	1378116	-0.6
2015-08	1520699	1369656	11.0
2015-09	1281762	1520699	-15.7
2015-10	1507930	1281762	17.6

—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

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

Approach -

Extracted Month & Year from OrderDate, then concatenated then together for Find Month_Year to calculate 3 month rolling average Sale.
Used Group by on Month_Year to find Sales Average on Monthly Basis.
Used Concept of Frames in window function to find average of SalePrice For Past 3 months.

Query -

```
with cte1 as (select ProductCategory, concat(right(OrderDate, 4), mid(OrderDate, 3, 3)) as Month_Year,
sum(SalePrice) as Monthly_Revenue
from orders
where ProductCategory is not null
group by ProductCategory, Month_Year
order by ProductCategory, Month_Year)

select ProductCategory, Month_Year, Monthly_Revenue,
round(avg(Monthly_Revenue) over(partition by ProductCategory order by Month_Year
      rows between 2 preceding and current row),1) as 3_Month_Rolling_Average
from cte1;
```

Output -

ProductCategory	Month_Year	Monthly_Revenue	3_Month_Rolling_Average
Electronics	2015-01	527631	527631.0
Electronics	2015-02	484900	506265.5
Electronics	2015-03	443102	485211.0
Electronics	2015-04	458487	462163.0
Electronics	2015-05	448536	450041.7
Electronics	2015-06	383569	430197.3
Electronics	2015-07	447846	426650.3

----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----------

17. 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)

Ans.

Query -

```
update orders
set SalePrice = SalePrice*0.85
where CustomerID in(select CustomerID from orders
                    group by CustomerID having count(OrderID) >= 10);
```

----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----------

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

Approach -

Created CTE to find Current Order's OrderDate and Previous order's OrderDate in the same row using lag function For each customer by using partition by on customerID and using order by on OrderDate.

Then calculated average days by taking the average difference between two order dates. And using a group by CustomerID for aggregation.

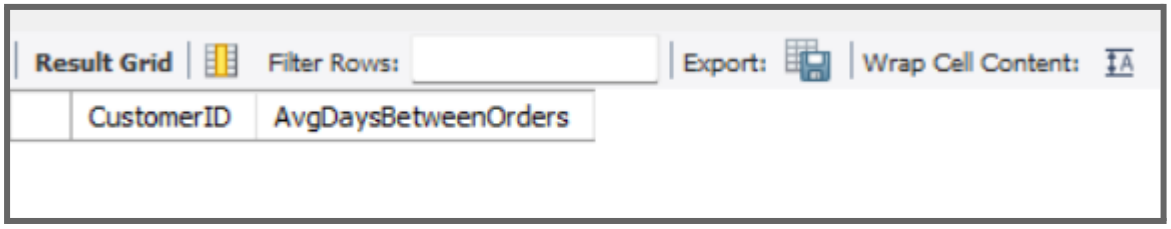
Query -

```
with cte1 as (select CustomerID, OrderDate,
lead(OrderDate) over(partition by CustomerID order by OrderDate) as NextOrderDate
from orders)

select CustomerID, avg(datediff(OrderDate,NextOrderDate)) AvgDaysBetweenOrders
from cte1
where NextOrderDate is not null
group by CustomerID
having count(*)>=5
order by CustomerID;
```

Output -

No output, Since no Customer has made more than one order.



—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

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

Approach -

*Created CTE to find Revenue on Customer level by using group by on CustomerID. Using CTE as table in main Query, Filtered out only those CustomerID's who has Revenue > Avg(Revenue)*1.3.*

Query -

```
with cte1 as (select CustomerID, sum(SalePrice) as TotalRevenue
            from orders group by CustomerID)

Select CustomerID from cte1
where TotalRevenue > (select avg(TotalRevenue)*1.3 from cte1)
order by CustomerID;
```

Output -

CustomerID
230459076
230459099
230459104
230459106
230459108
230459113
230459121

—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

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

Query -

```
with cte1 as ( select ProductCategory, year(STR_TO_DATE(OrderDate, '%d-%m-%Y')) as Years,
sum(SalePrice) as TotalRevenue
from orders
where ProductCategory is not null
group by ProductCategory,Years
order by ProductCategory,Years ),

cte2 as ( select ProductCategory, Years, TotalRevenue,
lag(TotalRevenue) over(partition by ProductCategory order by Years) as PrevYearRevenue,
row_number() over(partition by ProductCategory order by Years desc) as ranking
from cte1 )

select ProductCategory, Years, TotalRevenue, PrevYearRevenue,
round((TotalRevenue-PrevYearRevenue)*100/PrevYearRevenue,1) as PercentageGrowth
from cte2
where ranking =1
order by PercentageGrowth desc
limit 3;
```

Output -

ProductCategory	Years	TotalRevenue	PrevYearRevenue	PercentageGrowth
Electronics	2020	7476017	5075567	47.3
Health and beauty	2020	2662065	1868938	42.4
Fashion	2020	2764786	1948038	41.9

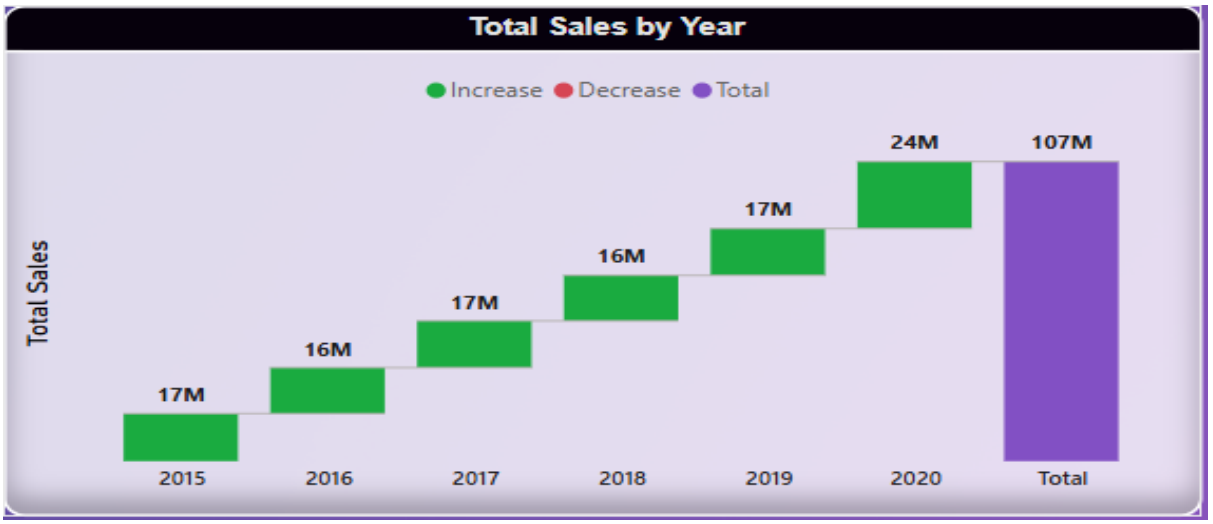
—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

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.**

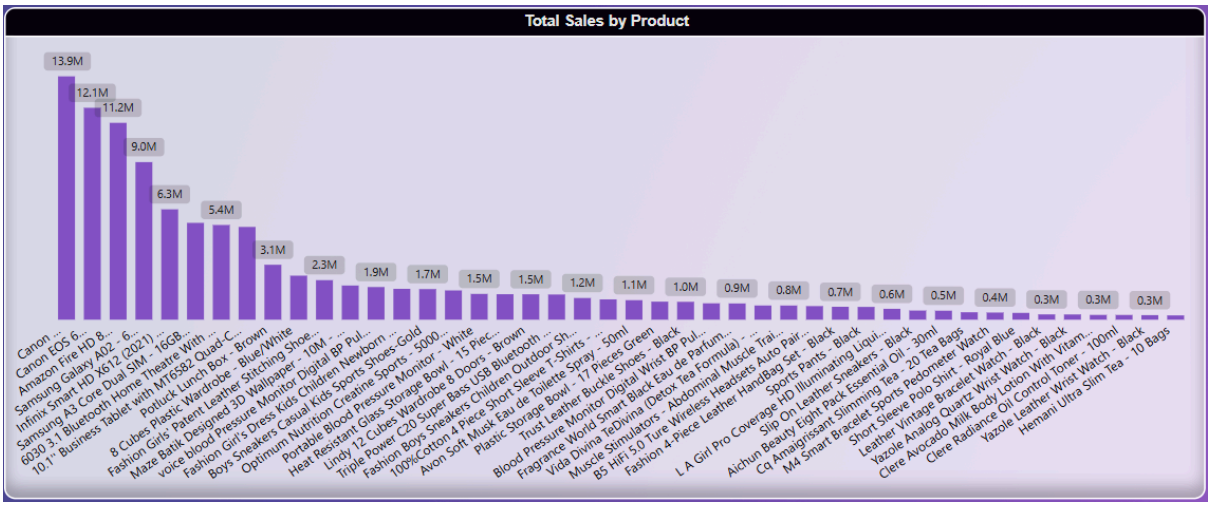
Revenue Breakdown by Year -

- *2015 starts with $17M, and by 2020, sales reach $24M.*
- *Experienced minor dip in 2016 & 2018.*
- *2020 shows the highest sales ($24M), indicating strong performance.*
- *The total sales over the six years sum up to $107M.*



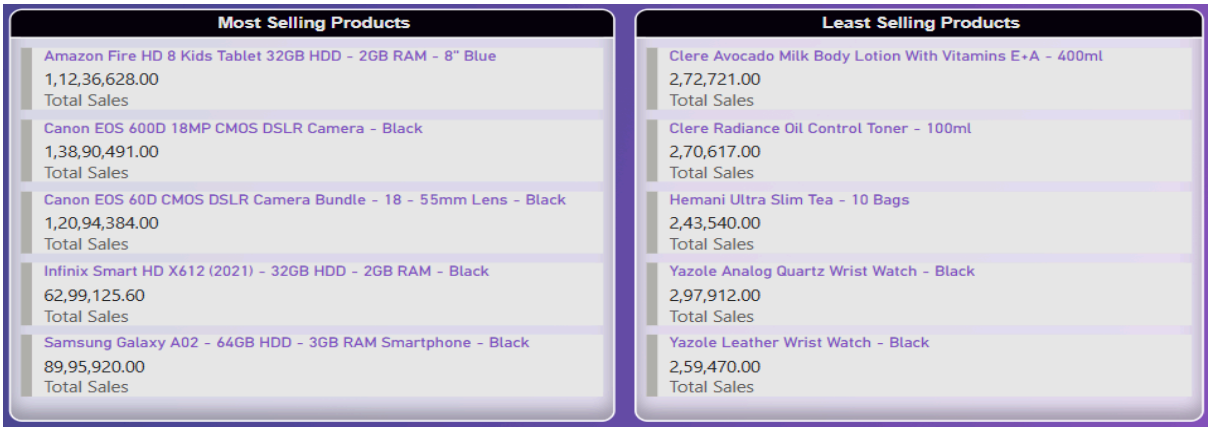
Revenue Breakup by Product -

- *Total 44 unique products made sales of $107M in the span of 6 years.*
- *Top 7 Products contribute to a total of ~65% of overall sales Amount.*



———————XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX———————

Top 5 & Bottom 5 Products on the bases of revenue Generated are -



Suggestions to Boost Low-Selling Items -

1. Improve Marketing & Promotions

- *Increase advertising spend on least selling products*
- *Offer bundling (e.g., "Buy A + Get 20% Off on C").*
- *Run seasonal promotions and discounts.*

2. Expand Distribution Channels

- *Explore e-commerce platforms to increase visibility.*
- *Partner with retailers or distributors to reach more customers.*

3. Price Optimization

- *Consider discounted pricing for bulk orders.*
- *Offer subscription models or loyalty discounts.*

—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

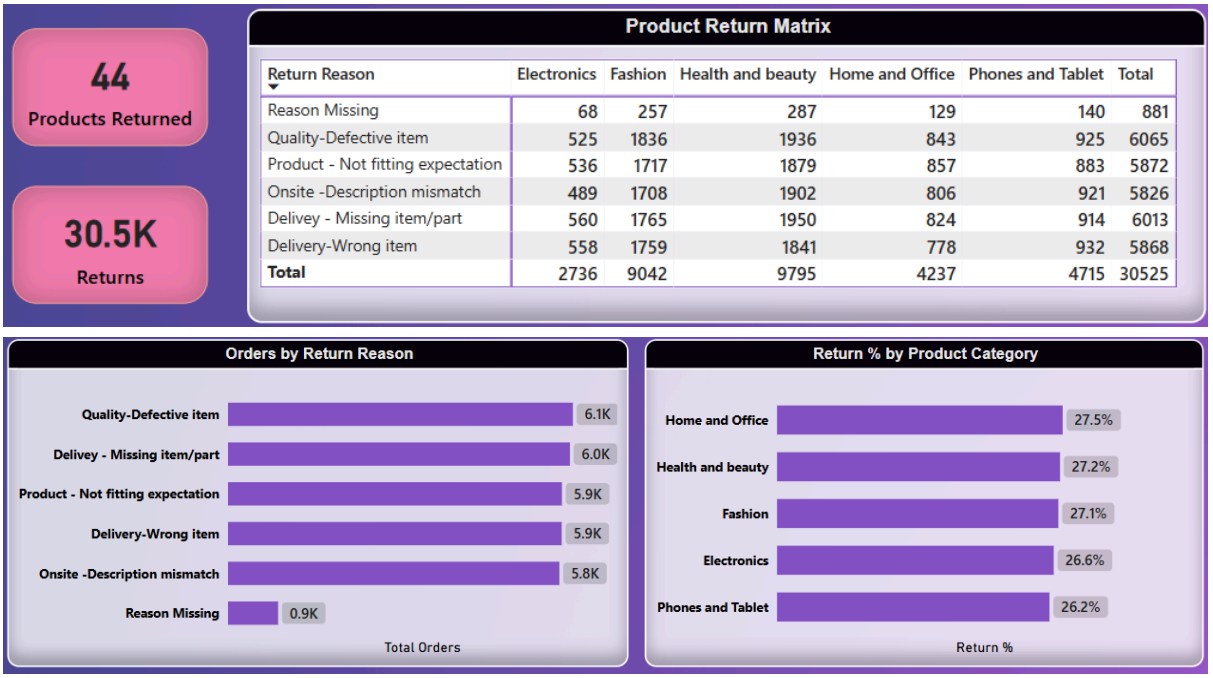
2. **How many products were returned? Use a DAX function to get this metric. Examine the possible reasons for returns and consider how this metric could indicate improvements in product descriptions or quality control.**

DAX Formula to find total orders returned -

```
Returns = CALCULATE(COUNTA(Orders[Status]),Orders[Status]="Returned")
```

Insights -

- *Total 44 distinct products were returned.*
- *Total 30.5k orders out of 113K orders were returned.*
- *Return percentage for each product category lies between 26.2% - 27.5%*
- *Total orders returned lies between 5.8K to 6.2K for each return reason, while for 0.9K returns, reasons are missing.*
- *Most Products returned from Health and Beauty(9795) & Least from Electronics(2736).*



Return Reason	Electronics	Fashion	Health and beauty	Home and Office	Phones and Tablet	Total
Reason Missing	68	257	287	129	140	881
Quality-Defective item	525	1836	1936	843	925	6065
Product - Not fitting expectation	536	1717	1879	857	883	5872
Onsite -Description mismatch	489	1708	1902	806	921	5826
Delivey - Missing item/part	560	1765	1950	824	914	6013
Delivery-Wrong item	558	1759	1841	778	932	5868
Total	2736	9042	9795	4237	4715	30525

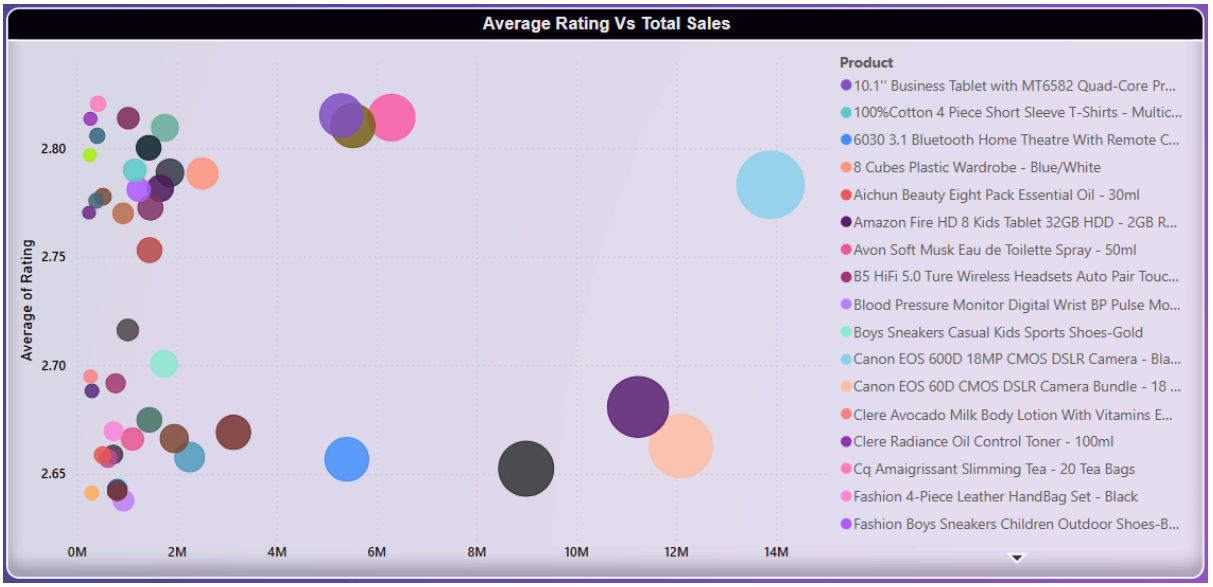
Recommendations -

- *Improve Quality Control – Reduce defective items through better manufacturing and inspection processes.*
- *Enhance Order Accuracy – Streamline warehouse and logistics operations to prevent shipping errors.*
- *Refine Product Descriptions – Ensure accurate and clear product details to manage customer expectations.*
- *Address Customer Expectations – Collect feedback to refine products and reduce dissatisfaction.*
- *A focused approach to quality, logistics, and customer communication can help reduce return rates and improve customer satisfaction.*

—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

3. **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?**

This scatter plot displays the relationship between average product ratings (Y-axis) and total sales volume (X-axis), with each bubble representing a different product.



Insights -

No Strong Correlation Between Rating and Sales:

- *Some products with low ratings (~2.65-2.75) have high sales (8M-14M).*
- *Several highly rated products (above 2.80) have low sales.*
- *This suggests that higher ratings do not necessarily drive higher sales.*

Few High-Sales, High-Rating Products Exist:

- *Only a few bubbles in the higher rating and high sales range, suggesting that high ratings alone don't guarantee high sales.*

Conclusion -

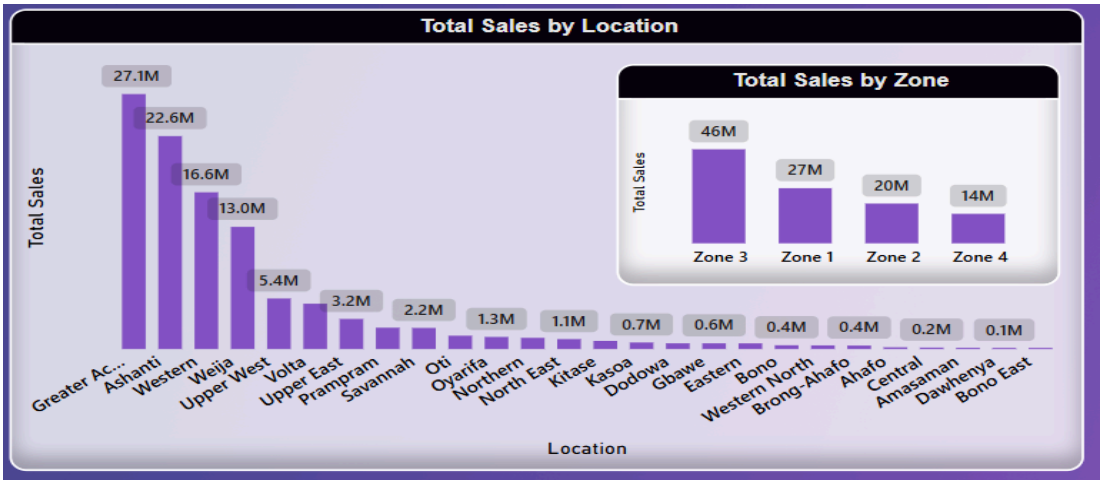
- *Some low-rated products have high sales, possibly due to strong marketing, brand recognition, or competitive pricing.*
- *Well-rated products with lower sales might indicate poor visibility, higher pricing, or less market demand.*

—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

4. **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?**

Insights -

- ***Top-Contributing Locations:*** *The highest sales come from Greater Accra ($27.1M), followed by Ashanti ($22.6M) and Western ($16.6M).*
- ***Low-Contributing Locations:*** *Areas like Dawhenya ($0.2M) and Bono East ($0.1M) contribute minimally.*
- ***Zone-Wise Breakdown:*** *Zone 3 dominates ($46M), followed by Zone 1 ($27M), Zone 2 ($20M), and Zone 4 ($14M).*



Strategic Implications for Marketing & Distribution -

Focus on High-Revenue Locations (Greater Accra, Ashanti, Western)

- *Invest in regional marketing campaigns & promotional offers to reinforce strong sales trends.*
- *Ensure supply chain efficiency to sustain demand.*

Expand in Underperforming Regions (Bono East, Dawhenya, etc.)

- *Assess barriers to sales (e.g., poor logistics, weaker customer base).*
- *Consider localized promotions, partnerships with local businesses, or improved distribution networks.*

Zone-Based Segmentation Strategy:

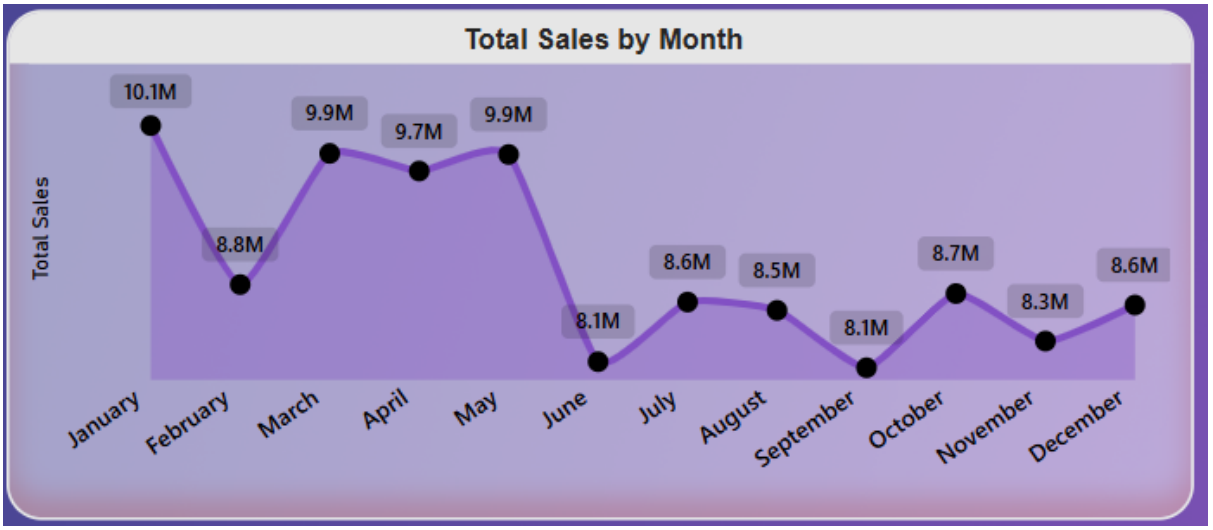
- *Zone 3 (High Sales): Focus on premium products & fast-moving goods.*
- *Zone 4 (Lowest Sales): Experiment with affordable options, discounts, and e-commerce penetration to drive demand.*

—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

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

Key Observations -

- *July (8.1M) and September (8.1M) recorded the lowest sales.*
- *February (8.8M) and December (8.6M) also saw dips compared to peak months.*



Targeted Marketing Strategies for Low Sales Months -

July & September: Mid-Year Sales & Discounts

- *Introduce Mid-Year Sales Campaigns (Flash sales, limited-time discounts).*
- *Run Back-to-School promotions targeting parents and students.*

February: Post-Holiday Engagement

- *Offer Valentine's Day deals to attract shoppers.*
- *Use loyalty rewards to re-engage customers after the holiday shopping slump.*

December: Festive & Holiday Shopping Boost

- *Leverage holiday sales, bundled deals, and free shipping offers.*
- *Increase digital ads and social media campaigns for last-minute shoppers.*

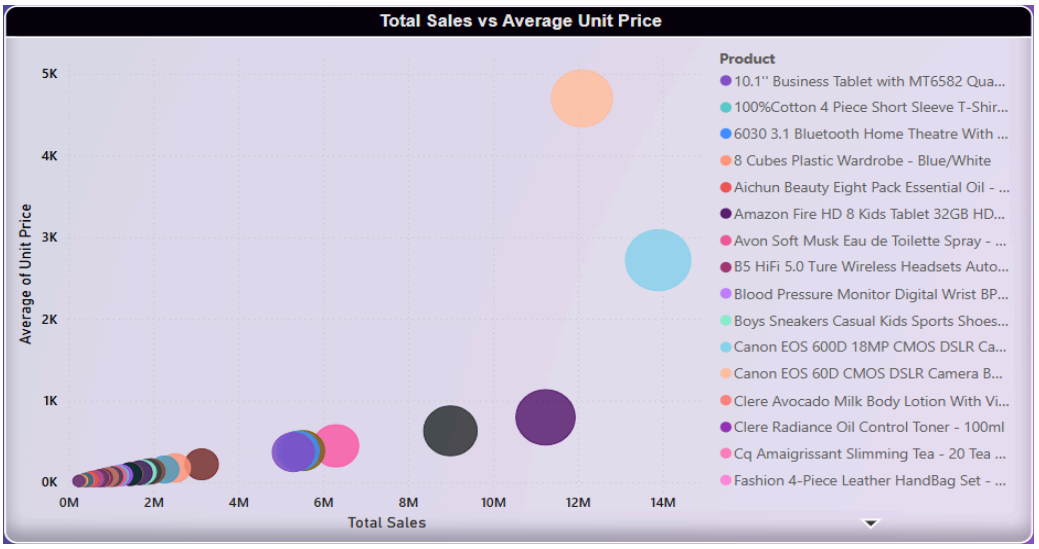
Conclusion -

By implementing seasonal promotions, targeted discounts, and strategic marketing efforts, the company can boost sales in weak months while capitalizing on high-demand periods.

—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

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

This scatter plot displays the relationship between average Unit Price (Y-axis) and total sales volume (X-axis), with each bubble representing a different product.



Insights -

- *Large bubbles at the top left (high price, low sales) indicate products that are expensive but not selling well.*
- *These products are -*
  - Canon EOS 60D CMOS DSLR Camera Bundle - 18 - 55mm Lens - Black.
  - Canon EOS 600D 18MP CMOS DSLR Camera - Black.

Recommendations to Boost Sales -

- *For expensive products, emphasize their quality, features, and exclusive offers.*
- *Offering limited-time promotions, bundles, or loyalty discounts may incentivize buyers to try the product despite the price tag.*
- *Revisit the pricing strategy. If this product is priced similarly to competitors but lacks distinct advantages, adjusting the price slightly or improving its features may help boost sales.*

—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

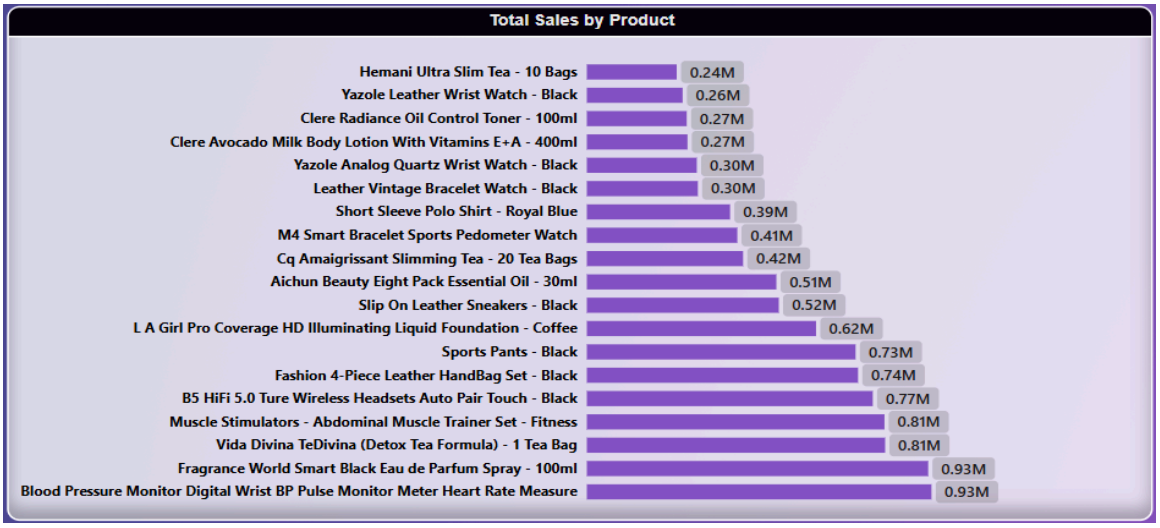
7. **Assess which products should have discounts. How can targeted incentives drive sales and customer loyalty for specific products?**

Products can be identified as underperforming on the basis of -

1. *Generating low sales revenue.*
2. *Less number of Quantities sold.*

This bar chart presents underperforming Products in terms of Revenue.

Overall Average sales value per product is 2.4M, but these products have generated revenue less than 1M, hence are underperforming.



These two products are underperforming wrt quantities sold.

- *Canon EOS 60D CMOS DSLR Camera Bundle - 18 - 55mm Lens - Black.*
- *Canon EOS 600D 18MP CMOS DSLR Camera - Black.*

Recommendations -

- *Targeted incentives, like time-limited offers or exclusive deals, create a sense of urgency, motivating customers to buy now.*
- *Offering discounts to customer preferences or purchase history can encourage hesitant buyers to make a purchase.*
- *Discounts or rewards on specific products can draw attention to items that may not be selling well, increasing their visibility and sales.*

—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

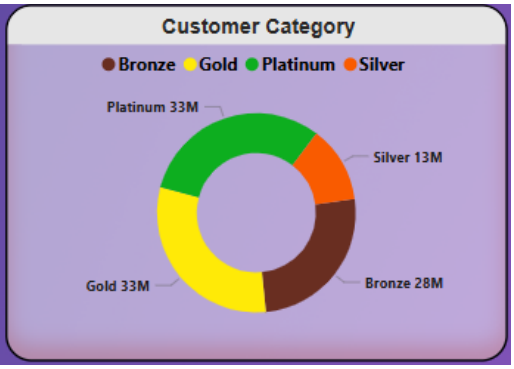
8. **Come up with a loyalty program to benefit the company's customers. From the available lot of customers come up with strategies to bucket them and provide benefits under different loyalty programs.**

Customers are categorized into four segments as -

- ***Platinum (Top 5%):*** *5,755 customers contributing ₹33.49M in total sales.*
- ***Gold (Next 10%):*** *11,217 customers contributing ₹32.87M in total sales.*
- ***Silver (Next 10%):*** *11,284 customers contributing ₹13.31M in total sales.*
- ***Bronze (Remaining 75%):*** *84,744 customers contributing ₹27.54M in total sales.*

Customer Matrix

Index	Customer_Bucket	Description	Customer Count	Total Sales
1	Platinum	Top 5%	5755	3,34,91,508
2	Gold	Between 5-15%	11217	3,28,75,958
3	Silver	Between 15-25%	11284	1,33,17,633
4	Bronze	Rest 75%	84744	2,75,54,199
Total			113000	10,72,39,297



Customer Loyalty Benefits and strategic Recommendations -

Increase Retention & Rewards for Platinum & Gold Customers:

- *Offer exclusive loyalty programs or early access to products.*
- *Provide personalized discounts or premium support to retain them.*

Upsell & Cross-Sell to Silver & Bronze Customers:

- *Design targeted promotions to encourage higher spending.*
- *Offer bundled discounts to increase basket size.*

Convert Gold to Platinum:

- *Gold customers are close in sales volume to Platinum; targeted campaigns can encourage them to spend more.*
- *Provide tiered benefits (e.g., free shipping on high-value orders, exclusive discounts).*

Encourage More Frequent Purchases from Bronze Customers:

- *Introduce gamification techniques (e.g., earn points for higher purchases).*
- *Limited-time discounts/subscription-based models to increase purchase frequency.*

Conclusion -

- *80% of revenue comes from just 25% of customers (Platinum, Gold, and Silver).*
- *A strategic focus on Gold & Platinum retention + increasing Bronze & Silver spending can boost overall sales.*

—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

9. **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?**

DAX function to calculate total sales for Product Category “Fashion” and delivery type “Shipped from Abroad”.

```
F&SFA = CALCULATE(SUMX(Orders, Orders[Sale Price]), Orders[Product Category]
="Fashion",Orders[Delivery Type]="Shipped from Abroad")
```

Other DAX Functions used are -

- *RELATED*
- *RANKX*
- *DIVIDE*
- *COUNTA*
- *DISTINCTCOUNTNOBLANK*
- *ALL*
- *IF*
- *AND*
- *SUMX*
- *DISTINCTCOUNT*
- *DATEDIFF*
- *CALCULATE*
- *AVERAGE*
- *SAMEPERIODLASTYEAR*
- *AVERAGEX*

Measures & Calculated columns created using these functions are -

```
Customer Age = RELATED(Customers[Customer Age])
```

```
Previous YM Sale = CALCULATE(SUM(Orders[Sale Price]),SAMEPERIODLASTYEAR
(Orders[OrderDate].[Date]) )
```

```
Product Rank =
RANKX(ALL(Orders[Product]), [Total Sales], , DESC, Dense)
```

```
Products Returned = CALCULATE(DISTINCTCOUNTNOBLANK(Orders[Product]),Orders
[Status]="Returned")
```

```
Return % = divide(CALCULATE(COUNTA(Orders[OrderID]),Orders[Status]
="Returned"),COUNTA(Orders[OrderID]))
```

```
Return Reason = IF(AND(Orders[Status]="Returned",Orders[Reason]=""),"Reason
Missing",IF(Orders[Status]="Delivered","Not Applicable",Orders[Reason]))
```

```
Returns = CALCULATE(COUNTA(Orders[Status]),Orders[Status]="Returned")
```

```
Sales Difference = [Total Sales] - [Previous YM Sale]
```

```
Total Products = DISTINCTCOUNT(Orders[Product])
```

```
Total Returned = CALCULATE(SUM(Orders[Sale Price]),Orders[Status]="Returned")
```

```
Total Sales = SUM(Orders[Sale Price])
```

```
Wait Time = DATEDIFF(Orders[OrderDate],Orders[Delivery Date],DAY)
```

```
Avg wait time(Delivered Products) = CALCULATE(AVERAGEX(Orders, DATEDIFF
(Orders[OrderDate], Orders[Delivery Date],DAY)),Orders[Status]="Delivered")
```

—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

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

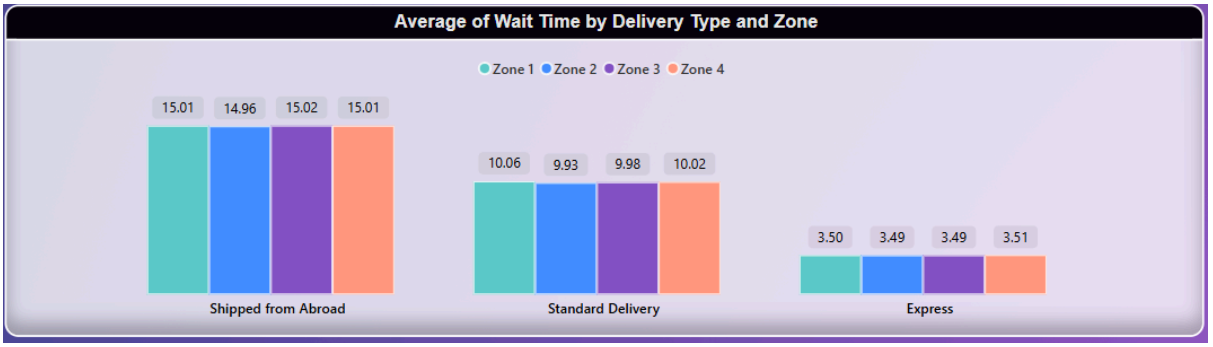
Insights -

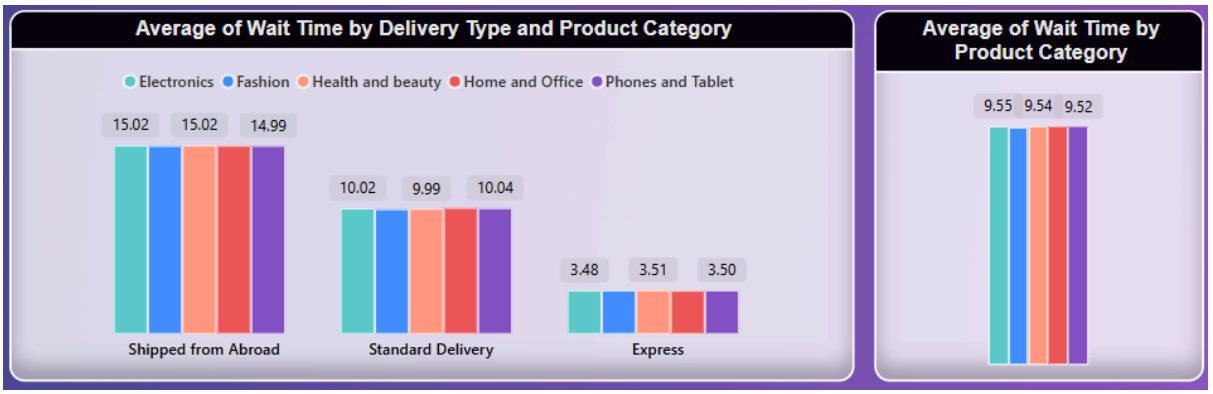
Clear Relationship Between Delivery Type and Wait Time

- ***Shipped from Abroad:*** *Highest wait time (~15 days) due to international logistics, customs processing, and longer transit times.*
- ***Standard Delivery:*** *Moderate wait time (~10 days), indicating local or regional shipping with potential delays in fulfillment or last-mile delivery.*
- ***Express Delivery:*** *Fastest option (~3.5 days), highlighting efficient handling and prioritization in logistics.*

Consistent Wait Times Across Product Categories

- *Products like Electronics, Fashion, Health & Beauty, and Home & Office all show similar wait times within each delivery type, implying that logistics efficiency is more dependent on delivery method rather than product category.*





Strategies to optimize scheduling and staffing -

- *Allocating more staff or optimising fulfilment processes to reduce delays in this Electronics category.*
- *Ensuring the current staff levels are maintained but be prepared to scale during peak demand periods.*
- *By leveraging data to forecast high-demand periods (e.g., holidays) and adjust staffing levels to ensure timely order fulfilment across all categories.*

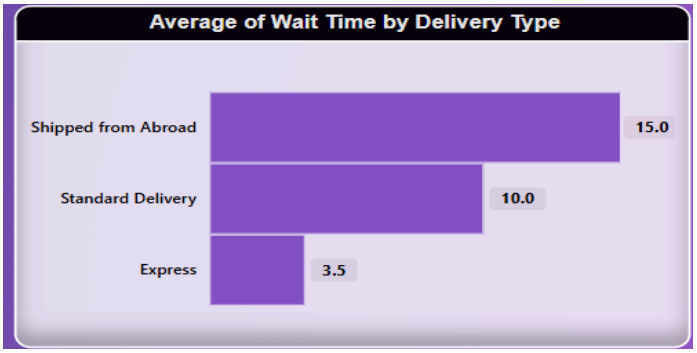
—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

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

Insights -

Clear Relationship Between Delivery Type and Wait Time

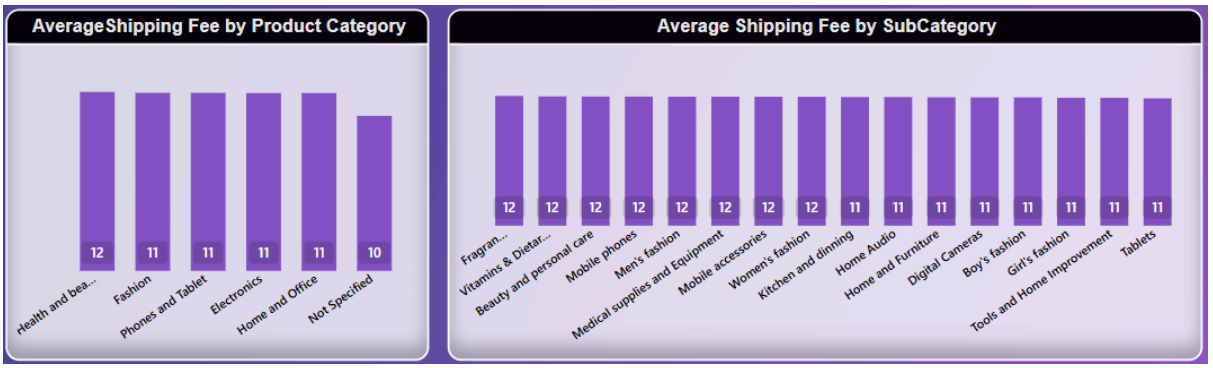
- ***Shipped from Abroad:*** *Highest wait time (~15 days) due to international logistics, customs processing, and longer transit times.*
- ***Standard Delivery:*** *Moderate wait time (~10 days), indicating local or regional shipping with potential delays in fulfillment or last-mile delivery.*
- ***Express Delivery:*** *Fastest option (~3.5 days), highlighting efficient handling and prioritization in logistics.*



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

Insights -

- *Most categories have similar shipping fees, with values ranging between 10-12 units.*
- *Health & Beauty and Fashion have the highest average shipping fee (12).*
- *Home & Office and Not Specified have slightly lower shipping fees (10-11).*
- *Majority of subcategories have a uniform shipping fee of 11-12 units, indicating no major variation based on product type.*
- *Fragrances, Vitamins, and Medical Supplies have the highest shipping charges (12), likely due to packaging, fragility, or special handling needs.*
- *Tablets, Digital Cameras, and Fashion Items maintain a standard 11-unit fee.*



—----------*XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX*—----------

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

This Table shows avg. wait time, avg. unit price & return % for low rated products (rating <2.7)

Product Matrix

Product	Average of Rating	Average of Wait Time	Average of Unit Price	Return %
Blood Pressure Monitor Digital Wrist BP Pulse Monitor Meter Heart Rate Measure	2.64	9.48	63	35.23%
Yazole Analog Quartz Wrist Watch - Black	2.64	9.37	19	33.75%
Muscle Stimulators - Abdominal Muscle Trainer Set - Fitness	2.64	9.59	55	31.66%
Vida Divina TeDivina (Detox Tea Formula) - 1 Tea Bag	2.64	9.52	55	32.87%
Samsung Galaxy A02 - 64GB HDD - 3GB RAM Smartphone - Black	2.65	9.59	629	33.90%
6030 3.1 Bluetooth Home Theatre With Remote Control - Black + Free Smartwatch	2.66	9.54	385	32.19%
L A Girl Pro Coverage HD Illuminating Liquid Foundation - Coffee	2.66	9.48	42	33.90%
Fashion Girls' Patent Leather Stitching Shoes - Black	2.66	9.40	156	32.17%
Aichun Beauty Eight Pack Essential Oil - 30ml	2.66	9.62	35	32.97%
Sports Pants - Black	2.66	9.56	50	33.72%
Canon EOS 60D CMOS DSLR Camera Bundle - 18 - 55mm Lens - Black	2.66	9.58	4700	32.75%
Avon Soft Musk Eau de Toilette Spray - 50ml	2.67	9.52	75	33.48%
Maze Batik Designed 3D Wallpaper - 10M - White/Black	2.67	9.60	135	32.40%
Potluck Lunch Box - Brown	2.67	9.51	220	33.67%
Fashion 4-Piece Leather HandBag Set - Black	2.67	9.36	49	33.54%
Lindy 12 Cubes Wardrobe 8 Doors - Brown	2.67	9.59	100	33.10%
Amazon Fire HD 8 Kids Tablet 32GB HDD - 2GB RAM - 8" Blue	2.68	9.49	795	32.74%
Leather Vintage Bracelet Watch - Black	2.69	9.56	19	32.58%
B5 HiFi 5.0 Ture Wireless Headsets Auto Pair Touch - Black	2.69	9.36	52	32.74%
Clere Avocado Milk Body Lotion With Vitamins E+A - 400ml	2.69	9.67	17	33.32%
Total	2.66	9.52	382	33.13%

9.5
Average of Wait Time

297
Average of Unit Price

27%
Return %

This matrix chart shows avg rating based on delivery type for low rated products (rating <2.7)

Product Rating (Delivery Type)

Product	Express	Shipped from Abroad	Standard Delivery	Total
Blood Pressure Monitor Digital Wrist BP Pulse Monitor Meter Heart Rate Measure	2.63	2.57	2.71	2.64
Yazole Analog Quartz Wrist Watch - Black	2.63	2.57	2.72	2.64
Muscle Stimulators - Abdominal Muscle Trainer Set - Fitness	2.69	2.61	2.63	2.64
Vida Divina TeDivina (Detox Tea Formula) - 1 Tea Bag	2.66	2.62	2.64	2.64
Samsung Galaxy A02 - 64GB HDD - 3GB RAM Smartphone - Black	2.68	2.57	2.71	2.65
6030 3.1 Bluetooth Home Theatre With Remote Control - Black + Free Smartwatch	2.67	2.63	2.67	2.66
L A Girl Pro Coverage HD Illuminating Liquid Foundation - Coffee	2.72	2.59	2.65	2.66
Fashion Girls' Patent Leather Stitching Shoes - Black	2.74	2.62	2.62	2.66
Aichun Beauty Eight Pack Essential Oil - 30ml	2.64	2.65	2.69	2.66
Sports Pants - Black	2.69	2.64	2.64	2.66
Canon EOS 60D CMOS DSLR Camera Bundle - 18 - 55mm Lens - Black	2.66	2.62	2.71	2.66
Avon Soft Musk Eau de Toilette Spray - 50ml	2.72	2.61	2.67	2.67
Maze Batik Designed 3D Wallpaper - 10M - White/Black	2.65	2.72	2.63	2.67
Potluck Lunch Box - Brown	2.62	2.70	2.69	2.67
Fashion 4-Piece Leather HandBag Set - Black	2.69	2.60	2.72	2.67
Lindy 12 Cubes Wardrobe 8 Doors - Brown	2.68	2.64	2.71	2.67
Amazon Fire HD 8 Kids Tablet 32GB HDD - 2GB RAM - 8" Blue	2.74	2.60	2.70	2.68
Leather Vintage Bracelet Watch - Black	2.69	2.62	2.75	2.69
B5 HiFi 5.0 Ture Wireless Headsets Auto Pair Touch - Black	2.65	2.62	2.81	2.69
Clere Avocado Milk Body Lotion With Vitamins E+A - 400ml	2.74	2.66	2.69	2.69
Total	2.68	2.62	2.69	2.66

Insights -

- *Average wait time does not affect rating of products as it is almost the same for all the products.*
- *Unit price is also not the factor as no trend is observed, both low to high rate products are present in the list.*
- *High return percent for low rated products is noticed, meaning factors like missing item/part, wrong description, defective item delivered or wrong item delivered lead to customer dissatisfaction and low rating.*
- *Product rating is generally low for delivery type = “Shipped from abroad” as its avg wait time is ~15 days.*

Strategies to decrease the low rating orders -

Reduce Delivery Delays from “shipped from abroad” delivery type -

- Improve order processing efficiency to reduce handling time.
- Implement real-time tracking to keep customers informed.

Improved Process and Quality based on return reasons -

*- **Improve Quality Control :** Reduce defective items through better manufacturing and inspection processes.*
*- **Enhance Order Accuracy** : Streamline warehouse and logistics operations to prevent shipping errors.*
*- **Refine Product Descriptions** : Ensure accurate and clear product details to manage customer expectations.*
*- **Address Customer Expectations** : Collect feedback to refine products and reduce dissatisfaction.*

Conclusion -

By reducing delivery times, optimizing shipping methods, improving product quality, and enhancing customer support, low-rating orders can be significantly reduced, leading to improved customer satisfaction and retention

—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

14. 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.

DAX Formula to compare each month's sales with the previous year's same month's total sales -

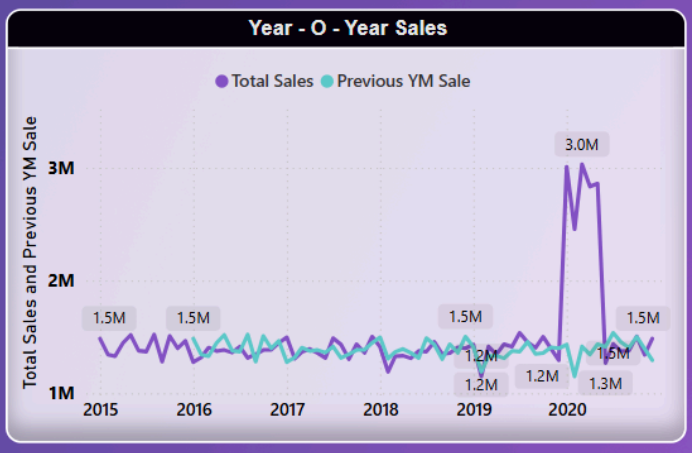
```
Previous YM Sale = CALCULATE(SUM(Orders[Sale Price]),SAMEPERIODLASTYEAR
(Orders[OrderDate].[Date]) )
```

Insights -

- *The graph indicates steady sales performance from 2015 to 2019, with total sales closely aligning with previous year's sales.*
- *A significant sales spike occurred in 2020, reaching up to 3M in a single month, followed by a sharp decline.*

Sales Comparison

Year	Month	Total Sales	Previous YM Sale
2015	January	14,84,377.05	
2015	February	13,43,863.65	
2015	March	13,28,537.10	
2015	April	14,46,946.30	
2015	May	15,17,282.80	
2015	June	13,78,115.15	
2015	July	13,69,651.25	
2015	August	15,20,687.45	
2015	September	12,81,755.35	
2015	October	15,07,927.60	
2015	November	13,98,724.10	
2015	December	14,65,184.55	
2016	January	12,76,295.55	14,84,377.05
2016	February	13,17,497.25	13,43,863.65
2016	March	14,05,136.85	13,28,537.10
2016	April	13,74,508.65	14,46,946.30
2016	May	13,84,018.35	15,17,282.80
Total		10,72,39,297.20	8,32,62,743.90



—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

15. What do you understand about the PowerBI gateway? What are its use cases?

Power BI Gateway -

It is a tool that enables secure data transfer between on-premises data sources and the Power BI cloud service, allowing you to refresh and access data stored locally in reports and dashboards.

Use Cases -

- ***Data Refresh:*** *Automatically refreshes on-premises data for up-to-date reports.*
- ***Secure Data Transfer:*** *Ensures encrypted, secure data transfer between on-premises sources and Power BI.*
- ***Access On-Premises Data:*** *Allows on-premises data (e.g., SQL Server, Excel) to be used in Power BI.*
- ***Hybrid Data Integration:*** *Combines on-premises and cloud data for unified reports.*
- ***Direct Query:*** *Enables real-time querying of live data sources in Power BI.*

Types of Gateways *-*

- ***Personal Gateway:*** *For individual use, refreshing personal data.*
- ***Enterprise Gateway:*** *For organizations, supporting multiple users and data sources.*

Some benefits of Power BI gateway are -

- *The connection between the data source and cloud is very secure and reliable which helps in protection of sensitive data.*
- *Data can be accessed directly without moving it to an expensive cloud storage solution hence saving money.*
- *Dashboards can be refreshed using Power BI gateway with a connection to the on-premise data source.*

—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------

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

If objective and subjective questions weren't provided in the given problem statement, here's how I would approach the task:

Step 1 (Data Exploration) -

I will start by exploring the dataset. Check the summary statistics to understand key trends. Look for missing values and handle them appropriately. Clean the data by standardizing categories and correcting any inconsistencies.

Step2 (Business Objective) -

Focusing on the objectives like boosting sales, optimizing delivery times, improving customer satisfaction, or enhancing product offerings and then identify the areas where business can improve its performance.

Step 3 (Key Metrics) -

Now by focusing on some key metrics like total sales, customer satisfaction (ratings), product performance, and delivery efficiency that will help overall business and identify areas for improvement.

Step 4 (Dashboard) -

Finally, creating visualizations that allow stakeholders to interact with the data.
By designing the dashboards that provide insights into sales by product, region, and time.
This will help the team drill down into the data and make informed decisions.

—----------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX—----------