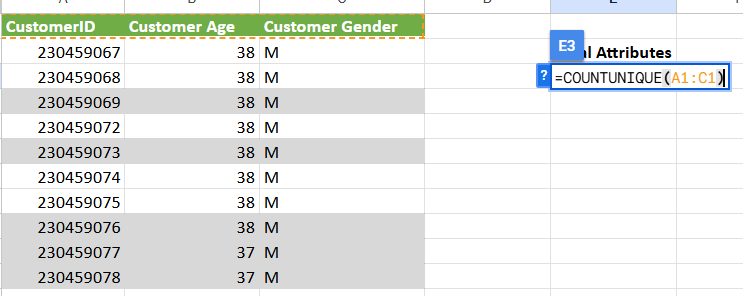
**Objective Questions:**

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

**Coding**—To find the total number of attributes in the customer table, we can use Excel's COUNTUNIQUE function to identify unique columns.

**Visualization -**



**Observation** - There are 3 Attributes in the Customer Table(CustomerID, Customer Age, Customer Gender)

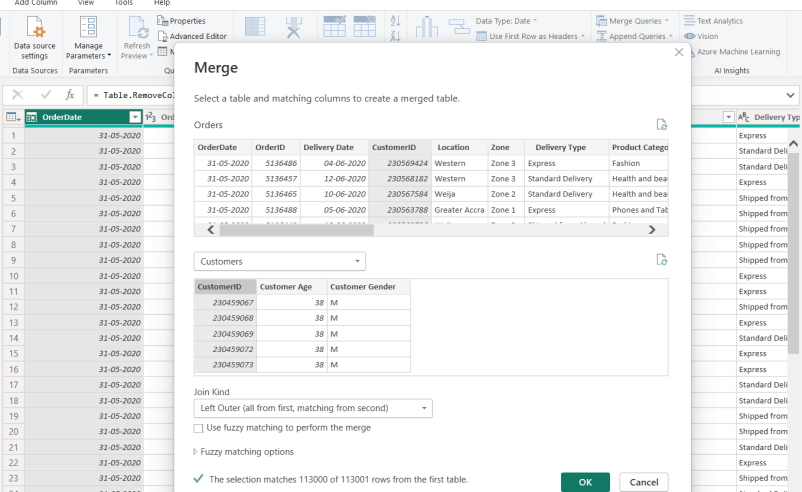
1. How will you get the “Customer’s” ages in the “Order” tables according to customer IDs?

**Coding/Approach -**

In Power BI, this is done by creating a **relationship** between the two tables in the model view, allowing Customer Age to be referenced directly in visualizations involving the Orders table. This is more easier way in terms of creating Visualizations.

Secondly, we can JOIN the Orders Table and Customers Table using the “MERGE QUERIES” option in the POWER QUERY. By selecting common columns between two tables, i.e, CUSTOMERID, finally we can expand the merged column and select the data columns to be included i.e, Customer Age.

**Visualization -**



**Result** - We will have the Customer's Age in the orders table now, “Customers.Customer\_Age.

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

**Approach -**

To prepare or clean the data before loading the same into the Data pane, we can make use of the Power Query editor by choosing/clicking the Transform tab.

Secondly, we get to know the details such as Empty cells, errors, and valid data percentage for each column by hovering over the column header(green line).

**Result -**

In this case, the dataset looks clean with zero null values, while duplicate values were removed for primary keys like Customer ID and Order ID. The Data format remained accurate across all attributes. Finally, there were a few additional columns with null values, which were removed as part of the cleaning process.

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

**Approach/Coding -**

To calculate total revenue generated by all sales in Power BI, we can create a New Measure under the Orders table as -

Visualization -

TotalSalesRevenue = SUM(Orders[Sale Price])

Result -

Total Sales Revenue = 107.24 Million.

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

Approach -

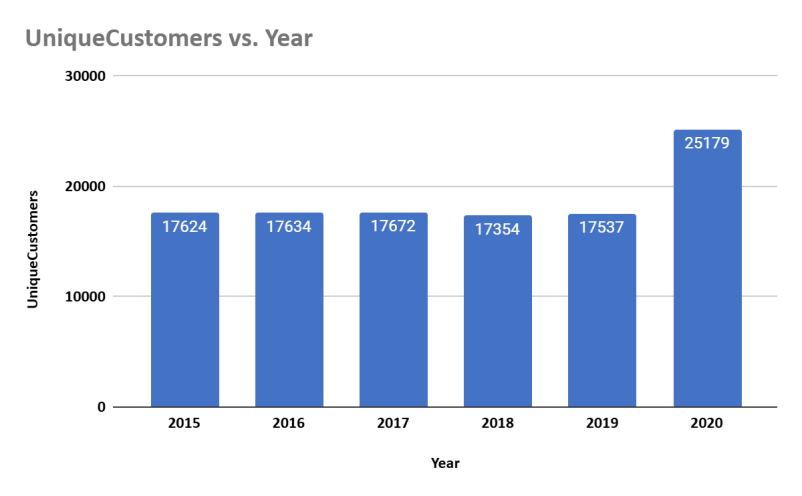
Firstly, we need to create a New column “YEAR” to extract the year from the order date present in the Orders table.

Year = YEAR(Orders![Order date[)

Secondly, we need to create a New Measure, “UniqueCustomers to count the number of customers.

UniqueCustomers = DISTINCTCOUNT(Orders![Customer Id])

Visualization -



Observation/Result -

* There is an increasing trend in the number of customers over the years, apart from the years 2018 and 2019, wherein the number of customers has slightly come down.
* The number of customers has increased gradually in the year 2020, with the highest count of customers across all years.
* Customer orders have dipped in the year 2018, followed by an increasing trend.

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

**Approach -**

To count the total number of unique products, we can create a calculated Measure In Power BI

**Formula -**

TotalUniqueProducts = DISTINCTCOUNT(Orders![Product])

**Result** -

There are 44 Unique Products in the company in total.

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

**Approach -**

Firstly, we need to create a New Column to find the delivery duration/days using the DATEDIFF DAX function.

**Formula -**

DeliveryDays = DATEDIFF(Orders[OrderDate], Orders[Delivery Date], DAY).

Further, we need to calculate a New Measure to find the Average Delivery Days specifically for delivered products. For the same, we can use the CALCULATE DAX function to filter out delivered products.

**Formula -**

Average Delivery Days = CALCULATE(AVERAGE(Orders[DeliveryDays]),

Orders[Status] = "Delivered")

**Result -**

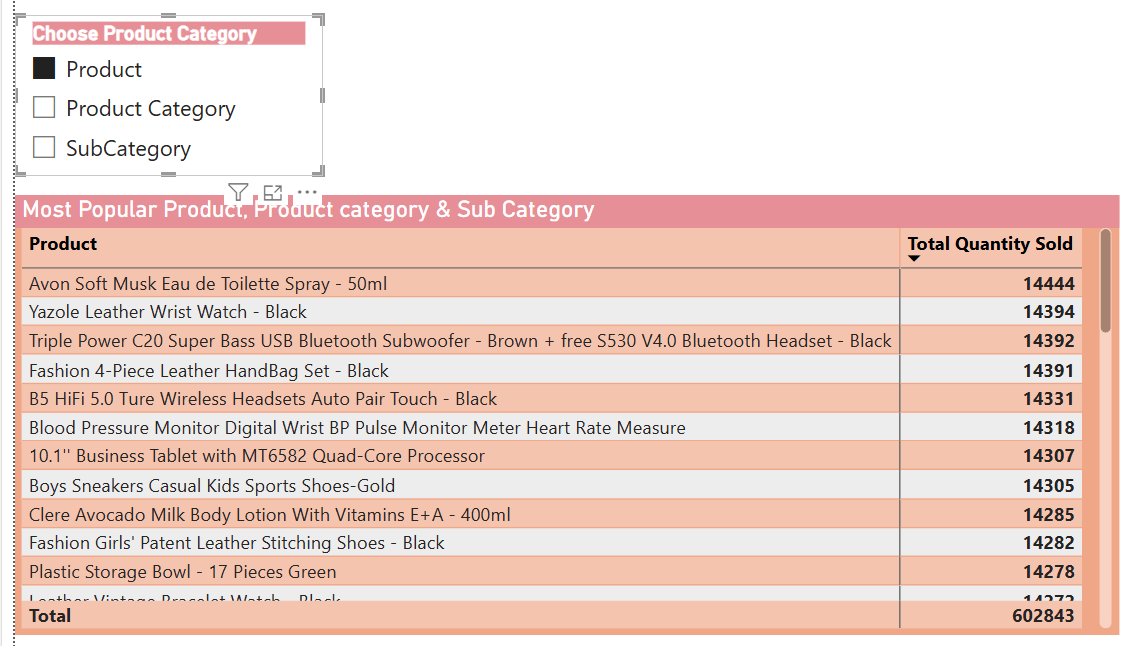
On average, it takes **9.41 days** to deliver the products.

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

**Approach -**

* Firstly, to determine popular products, categories, and sub-categories, we need to define **Popular.** In this context, we can consider the total quantity sold to be a metric for popularity.
* Secondly, to determine the total quantity sold, we can create a New Measure, **Total Quantity Sold = SUM(ORDERS[Order quantity])**
* Thirdly, we need to show the most popular Products, Product categories, and subcategories within one visual for users to navigate.
* In this context, we can make use of the **FIELD PARAMETER** feature of Power BI. Further created a field parameter named “Choose Product Category,” with a slicer on product, product category, and product subcategory.

**Visualization -**



We can navigate between the Product category and subcategory by using the slicer.

**Result -**

* The Most Popular Products are - Avon soft mask Spray, Yazole Leather watch, and Triple Power Brown Bluetooth Headset.
* The most popular Product categories are Health & Beauty, Fashion, followed by Phones and Tablets.
* The most popular product subcategories are Vitamins & Dietary supplements, Medical supplies & equipment, and Men’s Fashion

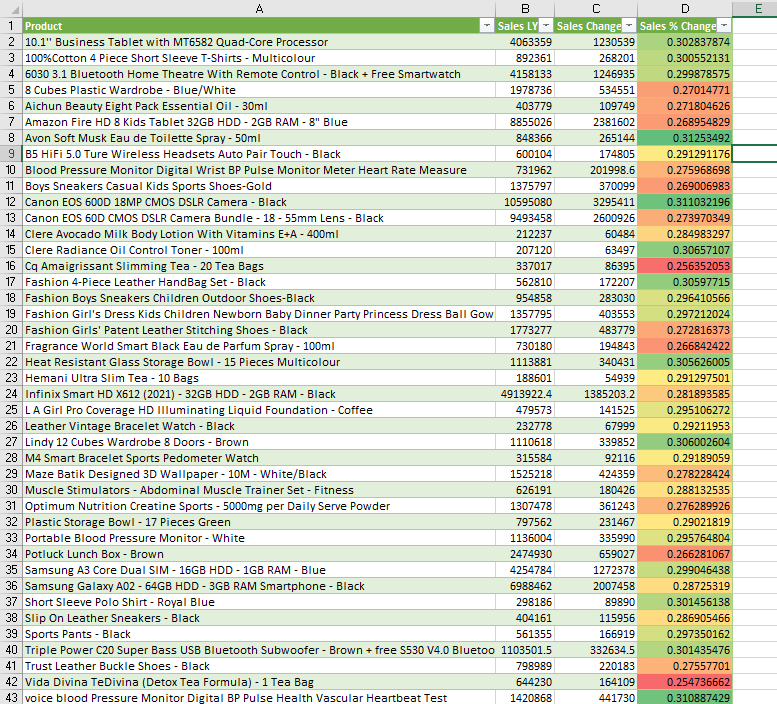
1. Which products have seen an increase or decrease in sales over the year?

To determine which products have seen an increase or decrease in sales over the year, the goal was to compare the total annual sales of each product in a given year (e.g., 2020) with its sales in the previous year (e.g., 2019). The steps followed were

* 1. A dedicated **Date Table** was created using DAX to enable proper time intelligence functions and consistent year-based filtering.
  2. A relationship was established between the DateTable[Date] and Orders[OrderDate].
  3. Separate measures were written to calculate **Sales in 2020** and **Sales in 2019** using CALCULATE and a YEAR() filter on the OrderDate field.
  4. Additional measures were created to calculate:  
     1. **Sales Change** = Difference between 2020 and 2019 sales
     2. **Sales % Change** = Percentage growth or decline over the previous year.

Visualization -

We can use the Matrix visual of Power BI to show the products with an increase or decrease in sales when compared to last year.



**Results/Observation** -

* Products like **Avon Soft Musk Eau de Toilette Spray - 50ml, Canon EOS 600D 18MP CMOS DSLR Camera - Black, and Lindy 12 Cubes Wardrobe 8 Doors - Brown** have witnessed an increase in sales over the year.
* While products like **Cq Amaigrissant Slimming Tea - 20 Tea Bags, Vida Divina TeDivina (Detox Tea Formula) - 1 Tea Bag** have witnessed a decrease in sales over the year.

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

**Approach -**

We can observe the **MODEL View** in Power BI in order to understand the type of relationship between the 2 tables “Customers” and “Orders”.

**Result -**

The relationship between the CustomerID column in the Orders table and the CustomerID column in the Customers table will be a **One-to-Many (1:\*) relationship**, with single directional filtering. Here **Customers** table is the **"one" side**, where each customer appears only once. On the other end, the **Orders** table is the **"many" side**, since a customer can place multiple orders.

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

**Approach-**

* First, used a function like COUNTBLANK to count the missing values or null values
* Secondly, used the Filter option to select only those records with blank cells.
* In the given data set, the **Orders** table had 8 orders with different product categories and subcategories that had missing Unit price and quantity.
* In this case, since each order is important, we cannot delete the orders with missing values rather, we can impute values.
* The missing Unit price and Quantity were imputed based on the values of a similar kind.
* Unit price was imputed by referring to the Unit price of the same kind of product, product category, and subcategory, A Filter was used to filter out specific categories and subcategories.
* While the quantity was imputed by considering the average quantity ordered for that specific category and subcategory.

**Result** -

The orders with Imputed values have been pasted in the following Google sheet

<https://docs.google.com/spreadsheets/d/1begE58QpQ0XojPlZGGppJgesONlOth7i/edit?usp=sharing&ouid=112728103722066366501&rtpof=true&sd=true>

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

**Approach/Result -**

Firstly, there were zero data format issues in the given dataset across two different sheets.

Secondly, assuming there were data format issues, I would have corrected the format by checking the current format available for the column on the left corner of each column. Further, on clicking the format symbol, I would have changed the data formats respectively as per the requirement.

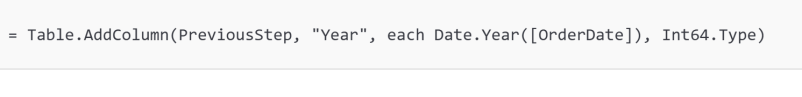
1. When we add a column in Power Query, what’s the code that comes in the M language in the formula bar? What do you know about M-query?

**Approach -**

In the current context, I have created a New Column, [YEAR](http://year.in), in the Orders Table and the M Query is as shown in the visualization.

**DAX - Year = YEAR(Orders![OrderDate])**

**Visualization -**

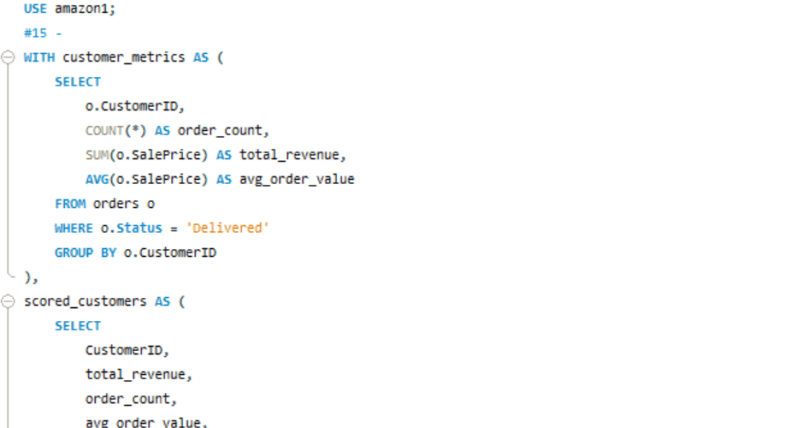
****

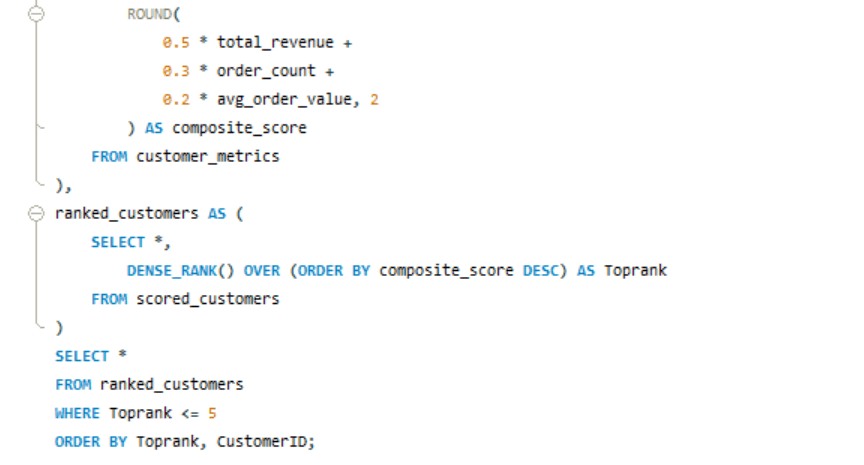
**Result -**

M Query is the **functional language** used in **Power Query** to transform and prepare data before loading it into the Power BI data model.

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.

**Visualization -**





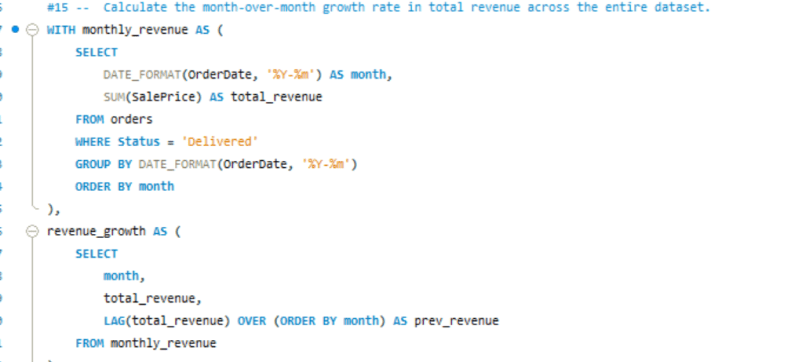
**Result -**

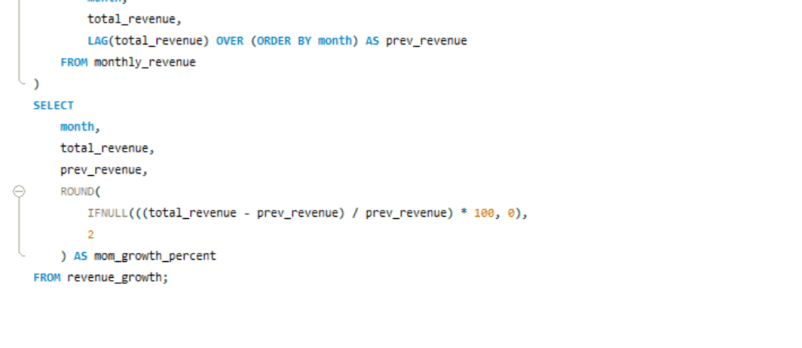
For better Presentation, the Result Data is in the spreadsheet below -

<https://docs.google.com/spreadsheets/d/1begE58QpQ0XojPlZGGppJgesONlOth7i/edit?usp=sharing&ouid=112728103722066366501&rtpof=true&sd=true>

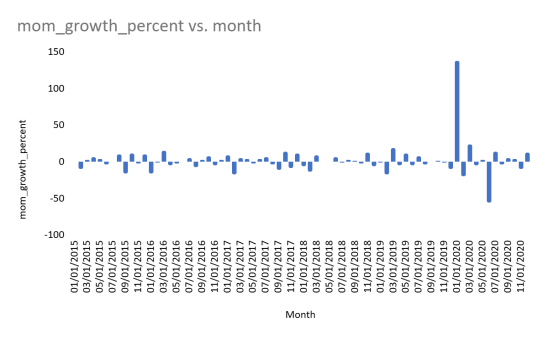
1. Calculate the month-over-month growth rate in total revenue across the entire dataset. (SQL)

**Visualization -**





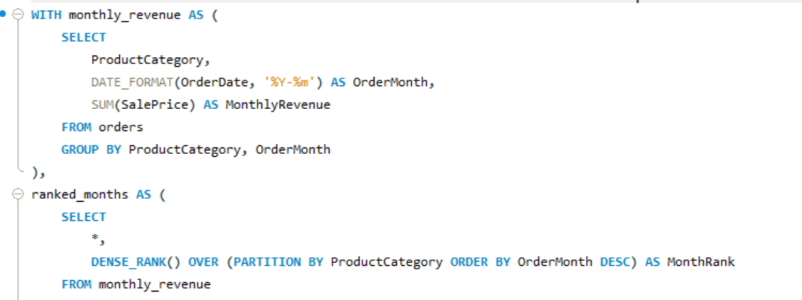
**Result -**

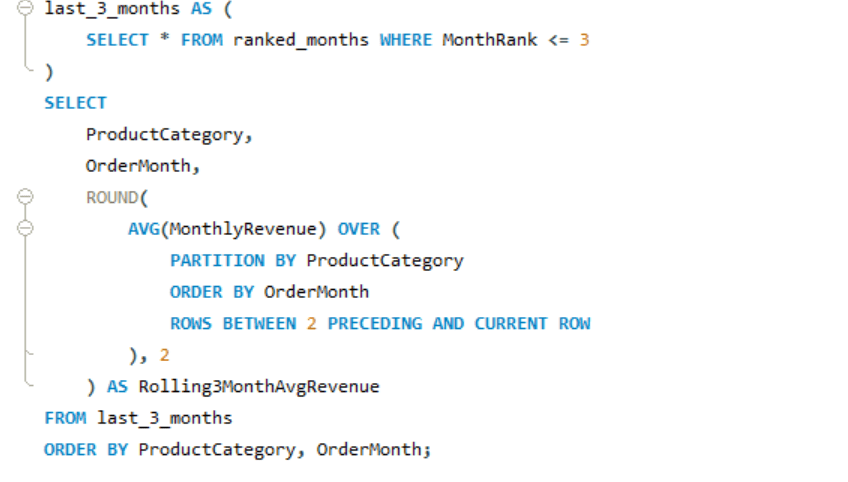


<https://docs.google.com/spreadsheets/d/1begE58QpQ0XojPlZGGppJgesONlOth7i/edit?usp=sharing&ouid=112728103722066366501&rtpof=true&sd=true>

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

**Visualization -**





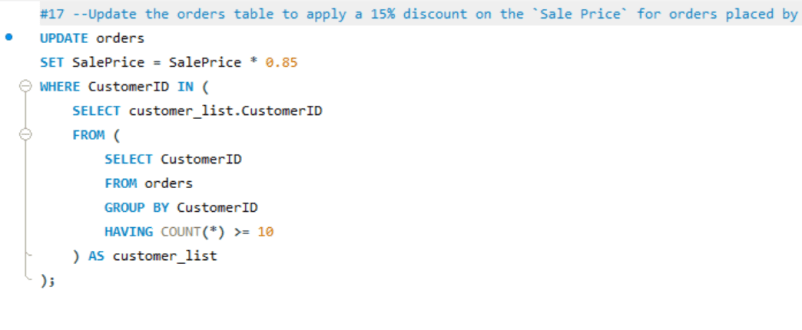
**Result -**

For better presentation, the result has been included in the following Sheet

<https://docs.google.com/spreadsheets/d/1begE58QpQ0XojPlZGGppJgesONlOth7i/edit?usp=sharing&ouid=112728103722066366501&rtpof=true&sd=true>

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)

**Visualization -**

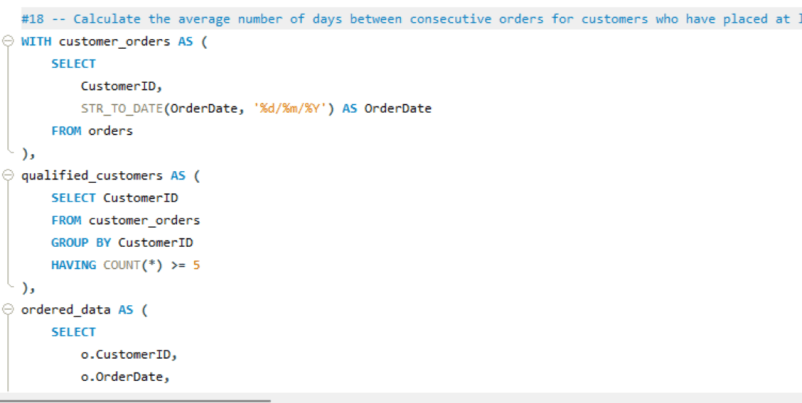


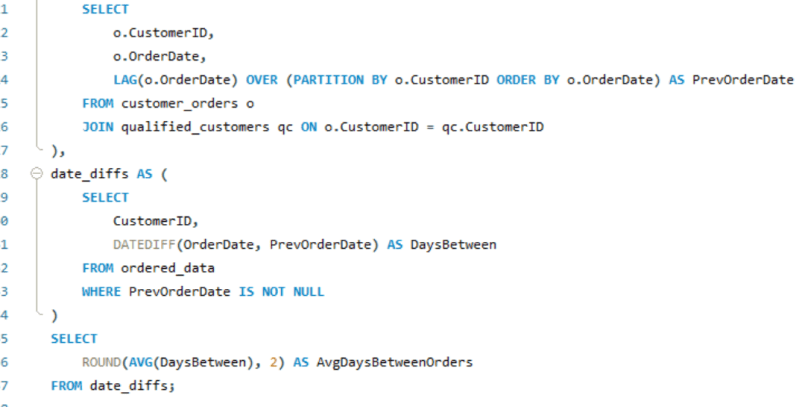
**Result -**

All the data records in the dataset are unique and none of the customers have placed more than or even 10 orders.

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

**Visualization -**



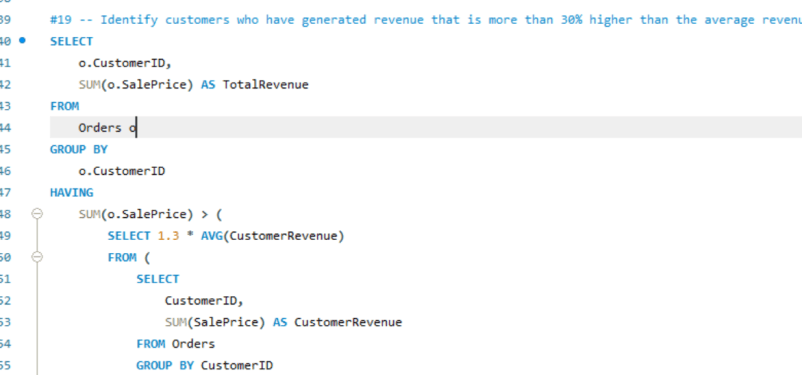


**Result** -

All the data records in the dataset are unique and none of the customers have placed more than or even 5 orders. Thus the result is Null

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

Visualization -



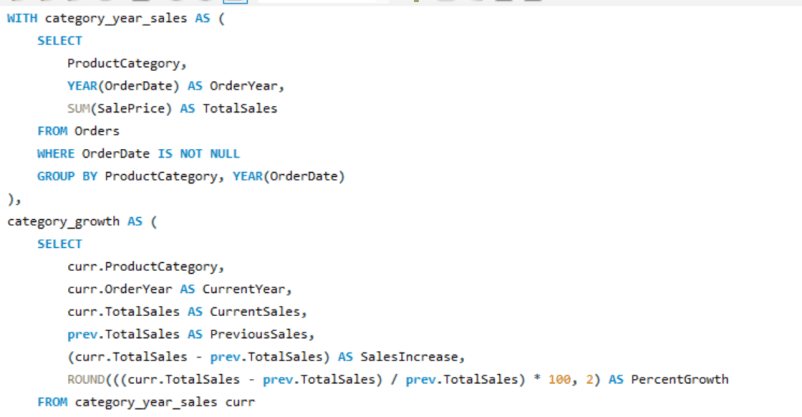
**Result** -

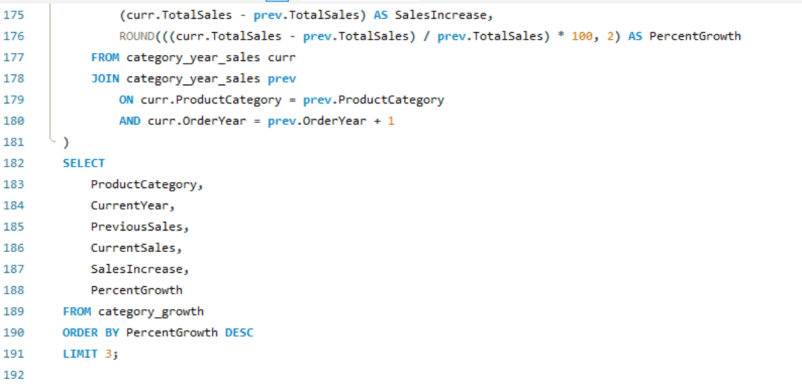
For better Presentation, the Result Data is in the spreadsheet below -

<https://docs.google.com/spreadsheets/d/1begE58QpQ0XojPlZGGppJgesONlOth7i/edit?usp=sharing&ouid=112728103722066366501&rtpof=true&sd=true>

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)

**Visualization -**





**Result -**



**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.

Approach -

* Firstly, we will create a New Measure, **Total Revenue**

Total Revenue = SUM(Orders[Sale Price])

* Secondly, we will use the New Column **YEAR,** which we created earlier.
* Thirdly, we will create a New Measure **Rank for each product per year** based on revenue to determine the low-selling items for each year

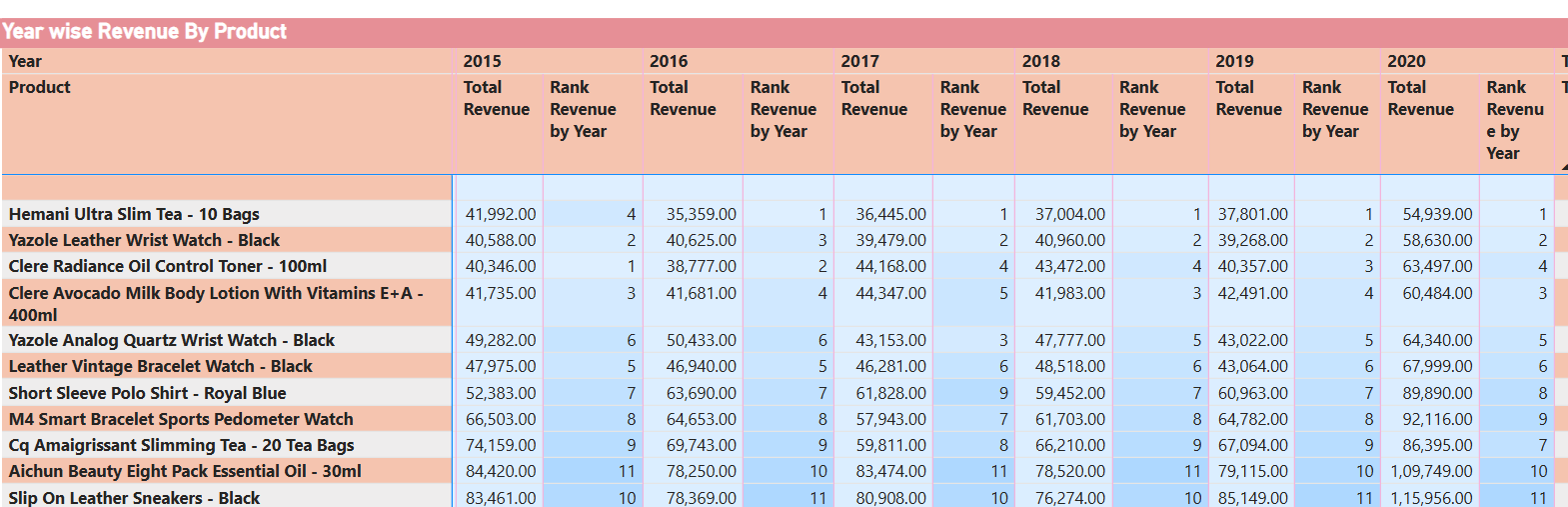
RRank Revenue by Year =

RANKX(FILTER(ALL(Orders[Product], Orders[Year]),

Orders[Year] = MAX(Orders[Year])),[Total Revenue],,ASC, )

* Further, we will use a Matrix visual to present the revenue breakdown for each product per year.

Visualization -



Result/Observation -

Over the years, some of the low-revenue-generating products are as follows

* Hemani Ultra Slim Tea = 10 bags
* Yazole Leather Wrist Watch - Black
* Cleare Radiance Oil Control Toner
* Cleare Avacado Milk Body lotion with vitamins
* Leather Vintage Bracelet Watch - Black

**Recommendations -**

1. **Enhanced Product Visibility & Merchandising**

* **Create themed collections** that include these low-performing products alongside popular complementary items. For example, feature the Yazole Leather Wrist Watch in a "Affordable Fashion Accessories" collection.
* **Implement seasonal promotions** targeting specific items - the Hemani Ultra Slim Tea could be featured in a "New Year Health Reset" campaign in January.
* **Optimize product placement** by ensuring these items appear in relevant "Frequently Bought Together" or "You May Also Like" sections.

1. **Pricing & Listing strategies -**

* **Create value bundles** - Pair the Cleare skincare products together at a slight discount to increase average order value while moving slow inventory.
* **Introduce tiered discounts** (e.g., buy 2 Hemani Tea packages, get 15% off) to encourage bulk purchases
* **Enhance product descriptions** with more compelling benefit-focused language and clear use cases.

1. **Targeted Marketing Approach**

* **Run targeted advertising campaigns** using Amazon's advertising platform focused on the most likely customer segments for each product.
* **Implement influencer marketing** for beauty and fashion items (Cleare products and watches).

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.

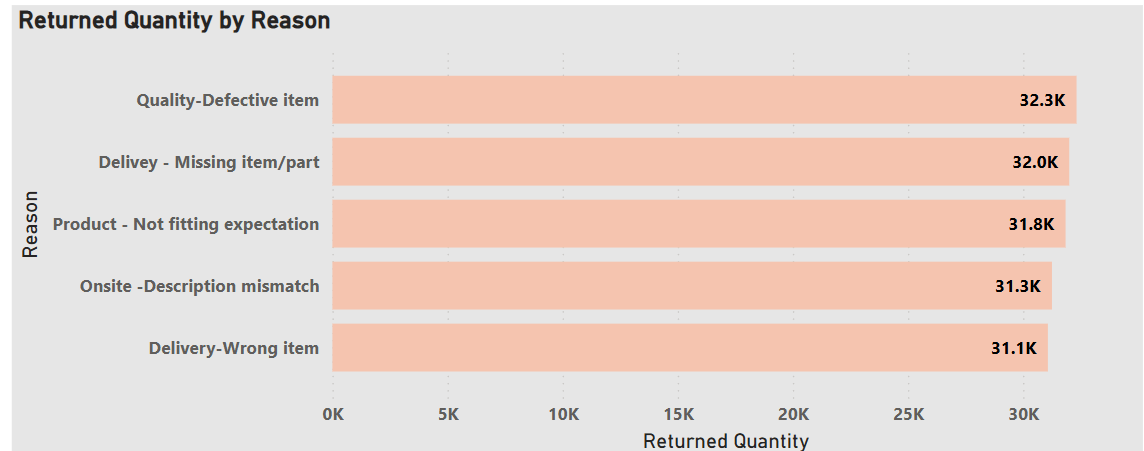
Approach -

* To determine the total quantity of products returned, we can create a New Measure using a DAX query as follows

Returned Quantity = CALCULATE (SUM(Orders[Order Quantity]),Orders[Status] = "Returned")

* Secondly, we can create a clustered bar chart to show the possible reasons for return and identify the most frequent reasons for product return.

**Visualization -**



**Result/Observation -**

* The total products returned was 163 thousand.
* Quality -Defective item was the most frequent reason for returning products, followed by Delivery - Missing item/part.
* Further products were returned due to Product - Not fitting Expectation. Onsite-Description mismatch and Delivery - Wrong Item.

**Recommendations -**

* Based on the visualization, it is clear that Amazon should **implement more rigorous supplier quality standards** with clear penalties for failing to meet benchmarks to ensure quality.
* **Enhance packing verification systems** with computer vision technology to confirm complete orders, thereby reducing the chances of Missing items.
* Overhaul product detail pages with more comprehensive specifications, measurements, and comparison tools to match product expectations
* **Audit product listings** with high return rates to identify and correct misleading descriptions to avoid description mismatch
* **Standardize product attribute terminology** across similar items
* **Enhance warehouse picking validation processes** with barcode scanning confirmation at multiple points to avoid delivering 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?

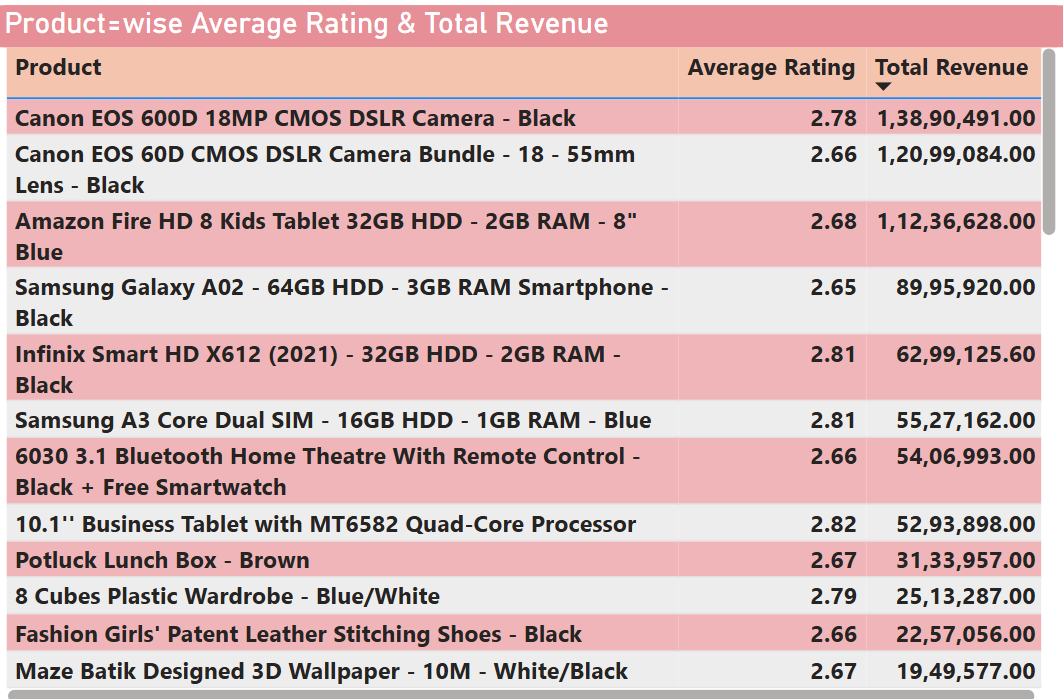
**Approach -**

* In this context, we need to verify whether is a correlation between product rating and sales. To determine the correlation, we need to create a New Measure **Average Rating**

Average Rating = AVERAGE(Orders[Rating])

* Secondly, we need to create a New Measure **Total Revenue,** which is already done previously.
* Thirdly, we can create a Table visual to show products, Average rating and Total revenue per product, to see whether there is a correlation between rating and sales revenue. In this context, since the products list is too much, we will use Table visual rather than Scatter chart.

Visualization -



**Result/observation -**

* The data contradicts the assumption that higher ratings directly drive higher sales. Multiple products with ratings below 2.8 generate substantial revenue.
* Most products cluster around 2.6-2.8 rating range regardless of revenue performance, suggesting ratings aren't the primary purchase driver.
* **Category-Specific Behavior:** Electronics (tablets, phones) show strong sales despite poor ratings, while fashion/lifestyle items have similar ratings but lower revenues.
* As a whole, the data does not verify that customers primarily filter by the highest-rated products. Instead, it reveals that factors like product necessity, brand recognition, price competitiveness, and category demand may be stronger sales drivers than ratings alone.

**Recommendations -**

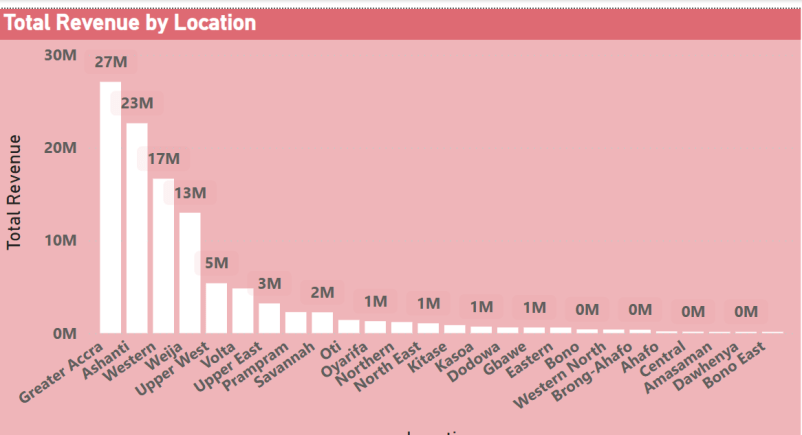
* **Focus on improving ratings for high-revenue products** - Products like Amazon Fire HD tablet (2.68 rating, $1.12M revenue) have massive improvement potential.
* **Prioritize quality improvements** for Samsung Galaxy A02 and Amazon Fire HD tablets - these have low ratings but generate substantial revenue
* **Create feedback loops** between customer reviews and product development teams for consistent underperformers.
* **Address fundamental product issues** causing low ratings rather than just marketing better-rated products
* **Implement proactive customer service** for products with ratings below 2.7 to prevent negative reviews

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?

**Approach -**

* Firstly, we can consider having a Measure for **Total revenue.**
* Secondly, we can create a clustered Bar chart to show a comparison of total revenues across different locations.

**Visualization -**



**Result/Observations -**

* **Extreme Market Concentration** - The top 3 locations (Greater Accra: $27M, Ashanti: $23M, Western: $17M) generate $67M of total revenue, representing a highly concentrated market distribution.
* Greater Accra appears to be the primary market, followed by Ashanti and Western.
* Revenue drops dramatically after the top 5 locations, with most regions generating $2M or less, indicating significant market penetration disparities.
* 15+ locations generate minimal revenue ($0-1M), suggesting either market saturation issues or insufficient market development.

**Recommendations -**

1. **Market Segmentation Approach**

* **Tier 1 Markets** (Greater Accra, Ashanti, Western): Focus on premium product mix and customer retention
* **Tier 2 Markets** (Volta, Upper East): Target market share expansion with competitive pricing
* **Tier 3 Markets** (All others): Implement market development strategies with localized approaches

2. **Regional Marketing Optimization**

* **Prioritize supply chain investments** in top 3 revenue locations
* **Implement location-specific promotional campaigns** to boost underperforming markets
* **Develop partnerships with local retailers** in emerging markets to increase market penetration

**3. Resource Allocation**

* Allocate 80% of resources to the top 5 locations generating the majority of revenue, while reserving 20% for developing high-potential underperforming markets
* **Logistics optimization -** Centralize distribution in the Greater Accra/Ashanti corridor to serve surrounding regions cost-effectively

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

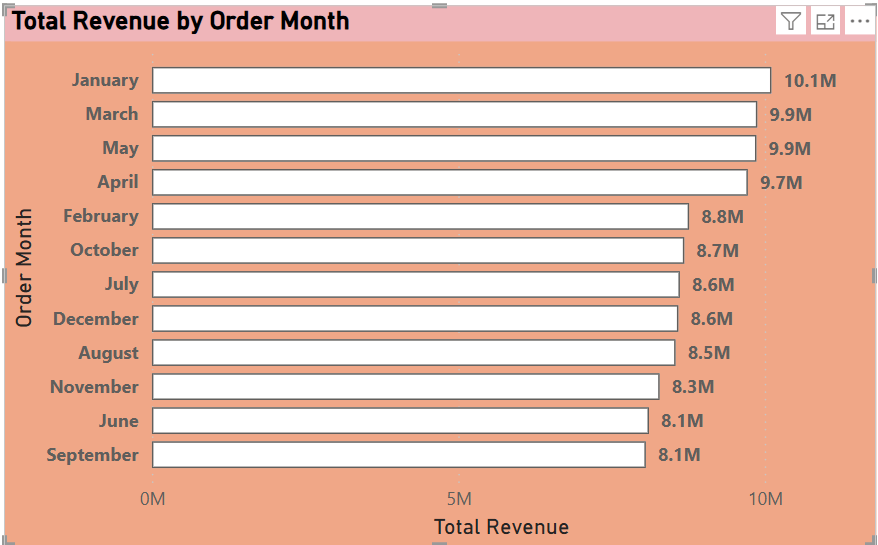
**Approach -**

* To determine sales revenue by month, we need to create a new Column **Order by Month**

Order Month = FORMAT(Orders[OrderDate], "MMMM")

* Secondly, we can create a clustered bar chart with Total revenue and Order month to show the distribution of revenue across months.

**Visualization -**



**Result/observation -**

* **June and September: $8.1M, November: $8.3M (third lowest), and August: $8.5M (fourth lowest) appear to be the low-revenue** months.
* **January: $10.1M** (highest revenue month - likely post-holiday shopping/New Year purchases) & **March & May: $9.9M each** (consistently strong performers) appear to be high revenue months.
* **Revenue Gap:** $2.0M difference between peak (January) and lowest months, representing 25% revenue variance.

**Recommendations -**

### The company should come up with targeted marketing strategies for September & June, months that show the greatest opportunity for revenue generation.

1. Targeted marketing strategies, such as the September Promotional campaigns, which include -

* **"Back-to-School/Work Boost"** - Target electronics, office supplies, and professional items
* **Flash weekend sales** to capitalize on end-of-summer shopping behavior
* **Bundle deals** on technology products as people prepare for new academic/work seasons

### Targeted June Mid-Year Campaign like -

* **"Mid-Year Mega Sale"** positioned as preparation for summer activities
* **Summer essentials bundles** targeting seasonal product categories.

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

**Approach -**

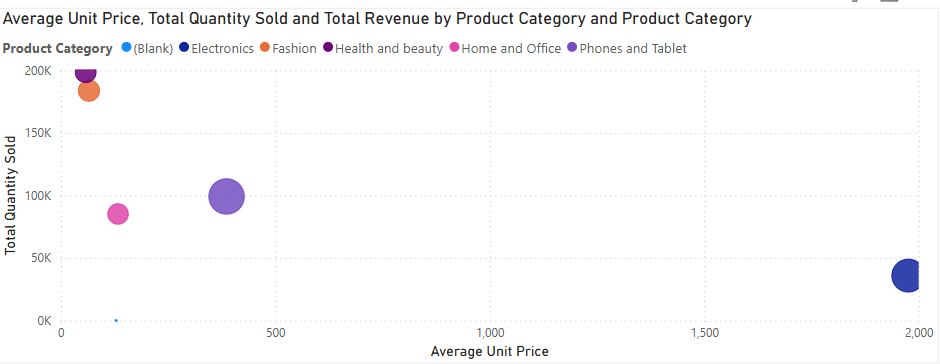
To determine products with high prices and low sales, we need to create a New Measure, **Average Unit Price**, in addition to Total Quantity & Total Revenue

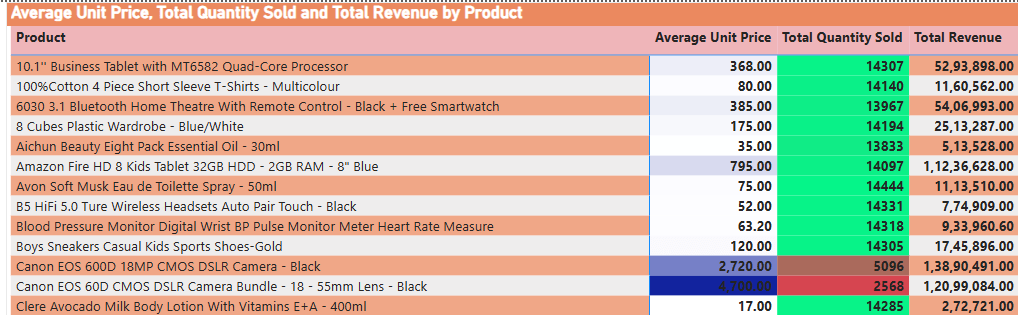
Average Unit Price = AVERAGE(Orders[Unit Price])

Secondly, we can create a scatter chart to show the Average price per unit, total quantity per product category, and which product category needs additional marketing efforts.

Thirdly, we can also create a Matrix table to particularly show which products need additional marketing.

**Visualization -**





Result/observation -

* It is evident that the Electronics Product category has the highest average price per unit but has less quantity been sold.
* **Canon EOS 60D CMOS DSLR Camera Bundle ($4,100 price, 2,568 quantity sold)** is the highest-priced item with relatively low sales volume. On the other end, Revenue: $1.2M, but quantity suggests market resistance
* **Canon EOS 600D 18MP CMOS DSLR Camera ($2,720 price, 5,096 quantity sold)** is the second-highest price point with moderate sales. On the other end, Revenue: $1.38M, but could perform better given the market demand for cameras.

**Recommendations -**

Amazon should come up with targeted marketing strategies to increase the sales of Camera products such as

* **Partner with photography influencers** for authentic product demonstrations
* **Implement bundle pricing strategies** with accessories to improve perceived value
* **Target professional and semi-professional photographers** through specialized channels

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

**Approach -**

To determine the product that needs to be discounted, we can consider the average price per unit, Total Quantity, and Average Rating.

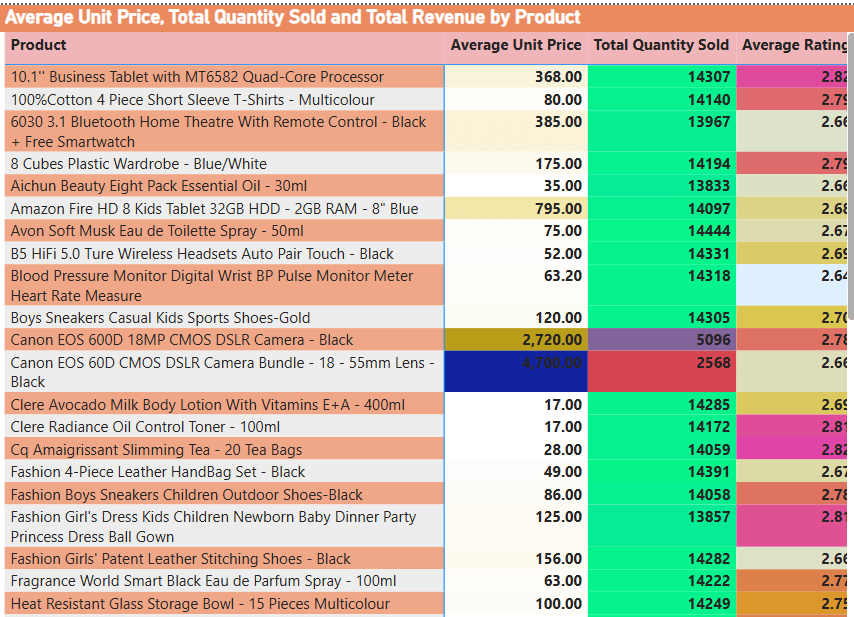
**Discounted Products =**

**Products with the Highest Average price per unit**

**Least Quantity Sold**

**Decent Average Rating >2.65**

**Visualization -**

****

**Result/observation -**

* The Product **Canon EOS 60D CMOS DSLR Camera Bundle ($4,100 price, 2,568 quantity sold), Canon EOS 600D 18MP CMOS DSLR Camera ($2,720 price, 5,096 quantity sold)** have a decent rating of above 2.65.
* These products need to be discounted to boost their sales.

**Recommendations**

* The above products should be given targeted incentives such as   
  **20-25% limited-time discount** to break psychological $4,000 barrier
* **Trade-in program** offering a $300-500 credit for older camera equipment
* **Professional photographer discount** (15%) with portfolio verification
* **First-time buyer incentive** offering an 18-month warranty extension.
* Other Targeted incentives, like giving customer loyalty points, Prime member exclusive discounts, and Referral bonuses, can also boost sales for the camera products.

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.

Approach -

* To devise a loyalty program, we need to bucket customers into different categories based on the total spending by each customer, total quantity ordered by each customer. For the same, we need to calculate some New Measures such as

Customer Orders = CALCULATE(COUNT(Orders[OrderID]), ALLEXCEPT(Orders, Orders[CustomerID]))

Customer Spend = CALCULATE(SUM(Orders[Sale Price]), ALLEXCEPT(Orders, Orders[CustomerID]))

Customer Avg Rating = CALCULATE(AVERAGE(Orders[Rating]), ALLEXCEPT(Orders, Orders[CustomerID]))

Customer Quantity = CALCULATE(SUM(Orders[Order Quantity]), ALLEXCEPT(Orders, Orders[CustomerID]))

* Secondly, we need to define a criterion to bucket the customers under different categories. The criteria here can be as follows

Loyalty Tier =

SWITCH(TRUE(),

[Customer Spend] > 5000 && [Customer Quantity] > 5, "Premium",

[Customer Spend] >= 3000 && [Customer Spend] <= 5000 &&

[Customer Quantity] >= 3 && [Customer Quantity] <= 5, "Silver",

[Customer Spend] >= 1000 && [Customer Spend] < 3000 &&

[Customer Quantity] >= 2 && [Customer Quantity] < 3, "Bronze",

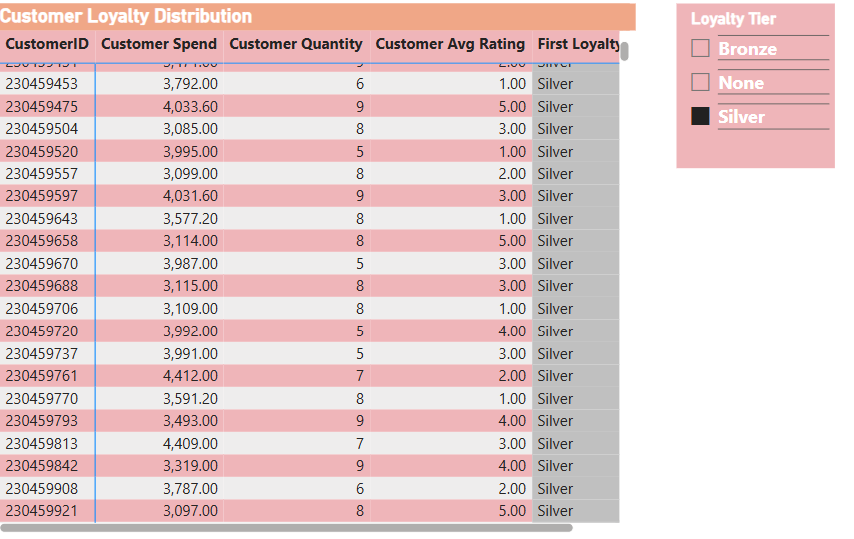
"None"

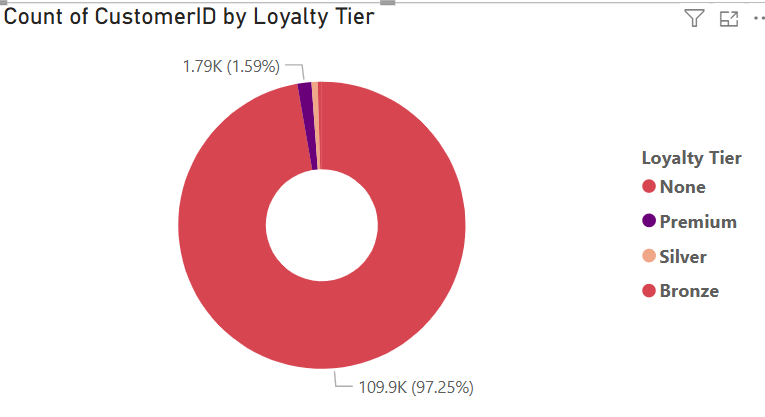
)

Here we are categorizing customers into Premium, Silver, and Bronze categories to provide different forms of loyalty benefits.

* Thirdly, we can create a matrix table to show different loyalty tiers for each customer. Also we can add a slicer to navigate between different loyalty tiers.

Visualization -





**Results/Observations -**

* **97.25% (None tier)** - Vast majority lack loyalty engagement
* **1.59% (Premium tier)** - Small but high-value customer segment
* **0.69% (Silver tier)** - Mid-tier engaged customers
* **0.47% (Bronze tier)** - Entry-level loyal customers

**Recommendation -**

Amazon should provide loyalty benefits as shown below to reward its loyal customers

#### **Bronze Tier (Mass Market Entry)**

* 1% cashback on purchases
* Free shipping on $35+ orders
* Monthly exclusive deals access
* Welcome bonus: $10 credit.

#### **Silver Tier (Growth Focus)**

* 3% cashback on purchases
* Free shipping on all orders
* Early sale access (24 hours)
* Birthday month 15% discount.

#### **Premium Tier (Elite Retention)**

* 7% cashback on all purchases
* Free same-day delivery
* Personal shopping concierge
* Exclusive product launches
* Annual VIP gift ($100+ value)
* Dedicated premium customer service line

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?

Result - DAX query to calculate Fashion sales from abroad

DAX query = Total Fashion Abroad Sales =

CALCULATE(

SUMX(

Orders,

IF(

Orders[Product Category] = "Fashion" &&

Orders[Delivery Type] = "Shipped from Abroad",

Orders[Sale Price],0)))

Visualization -



Result - The other type of DAX functions utilized are -

## **1. Aggregation Functions**

* **SUM()** - Used to calculate total revenue: TotalSalesRevenue = SUM(Orders[Sale Price])
* **AVERAGE()** - Used for average ratings and delivery days: Average Rating = AVERAGE(Orders[Rating])
* **COUNT()** - Used to count customer orders: Customer Orders = CALCULATE(COUNT(Orders[OrderID]))
* **DISTINCTCOUNT()** - Used to count unique customers and products: UniqueCustomers = DISTINCTCOUNT(Orders![Customer Id])

## **2. Filter Functions**

* **CALCULATE()** - Used extensively for filtered calculations: Average Delivery Days = CALCULATE(AVERAGE(Orders[DeliveryDays]), Orders[Status] = "Delivered")
* **FILTER()** - Used in ranking calculations: RANKX(FILTER(ALL(Orders[Product], Orders[Year]), Orders[Year] = MAX(Orders[Year])))

## **3. Date and Time Functions**

* **YEAR()** - Used to extract year from dates: Year = YEAR(Orders![OrderDate])
* **DATEDIFF()** - Used to calculate delivery days: DeliveryDays = DATEDIFF(Orders[OrderDate], Orders[Delivery Date], DAY)
* **FORMAT()** - Used to format dates: Order Month = FORMAT(Orders[OrderDate], "MMMM")

## **4. Time Intelligence Functions**

* **SAMEPERIODLASTYEAR()** - Used for year-over-year comparisons: Previous Year Revenue = CALCULATE([Total Revenue], SAMEPERIODLASTYEAR(DateTable[Date]))

## **5. Logical Functions**

**SWITCH()** - Used for creating loyalty tiers and age groups:

**6. Relationship Functions**

RELATED() - Used to pull data from related tables: RELATED(Customers[Customer Age])

**7. Table Functions**

CALENDAR() - Used to create date tables:

CALENDAR (MIN(Orders[OrderDate]), MAX(Orders[OrderDate]))

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

**Approach -**

* To determine the correlation between wait time across different demographics, firstly, we need to create a New Column **Wait Time (Days)**

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

* Secondly, we need to create a New Measure of **Average Wait Time**

**Average Wait Time = AVERAGE(Orders[Wait Time (Days)])**

* Thirdly, we can create a New Column **Age group** by placing customers into different age group buckets

Age Group =

SWITCH(

TRUE(),

RELATED(Customers[Customer Age]) < 25, "Under 25",

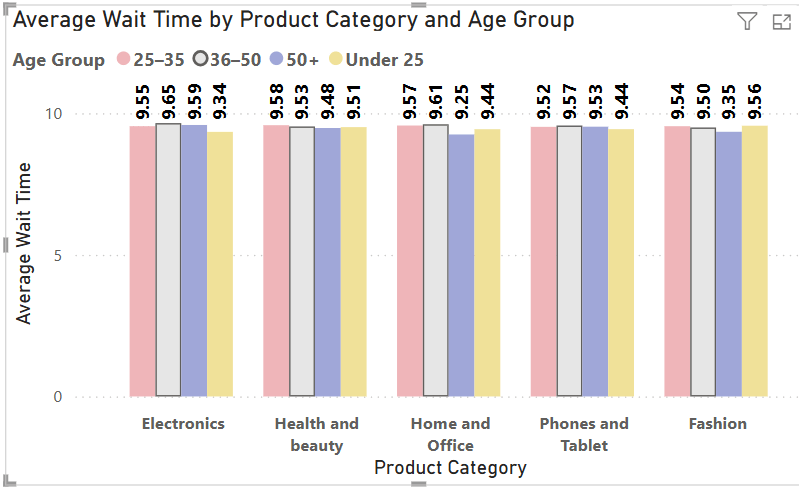
RELATED(Customers[Customer Age]) <= 35, "25–35",

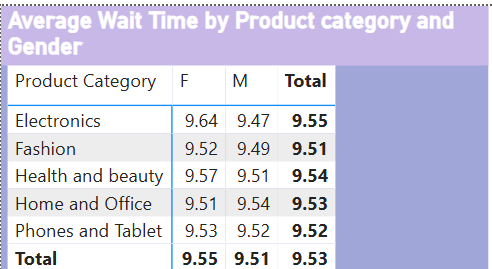
RELATED(Customers[Customer Age]) <= 50, "36–50",

"50+")

* Further, we can create a clustered column chart to show Average wait time by product category and Age group, followed by a Matrix to show Average wait time by product category and gender(Demographics).

**Visualization -**





**Result/observation** -

### **1. Age Group Patterns**

* **The 25-35 age group** consistently shows the **highest wait times** across all product categories (on average 9.55-9.58 days)
* **The 36-50 age group** has the **lowest wait times** across most categories (on average 9.48-9.53 days)
* **Under 25** and **50+** groups fall in between, with relatively similar patterns

### **2. Product Category Analysis**

* **Electronics** has the highest average wait time (9.55 days)
* **Health and Beauty** follow closely (9.54 days)
* **Home and Office** has the lowest average wait time (9.53 days)
* **Fashion** and **Phones, and Tablets** are in the middle range (9.51-9.52 days)

### **3. Gender Differences**

* Minimal gender-based variation in wait times
* Females are slightly higher in most categories (9.52-9.64 days)
* Males consistently around 9.47-9.54 days)

**Recommendations -**

1. **Staffing Optimization**

* **Peak Demand Staffing**: Allocate more customer service representatives during periods when 25-35 age group is most active
* **Category-Specific Resources**: Increase staffing for Electronics and Health & Beauty departments
* **Cross-Training**: Train staff across multiple categories to handle overflow from high-demand areas

### **Scheduling Strategy**

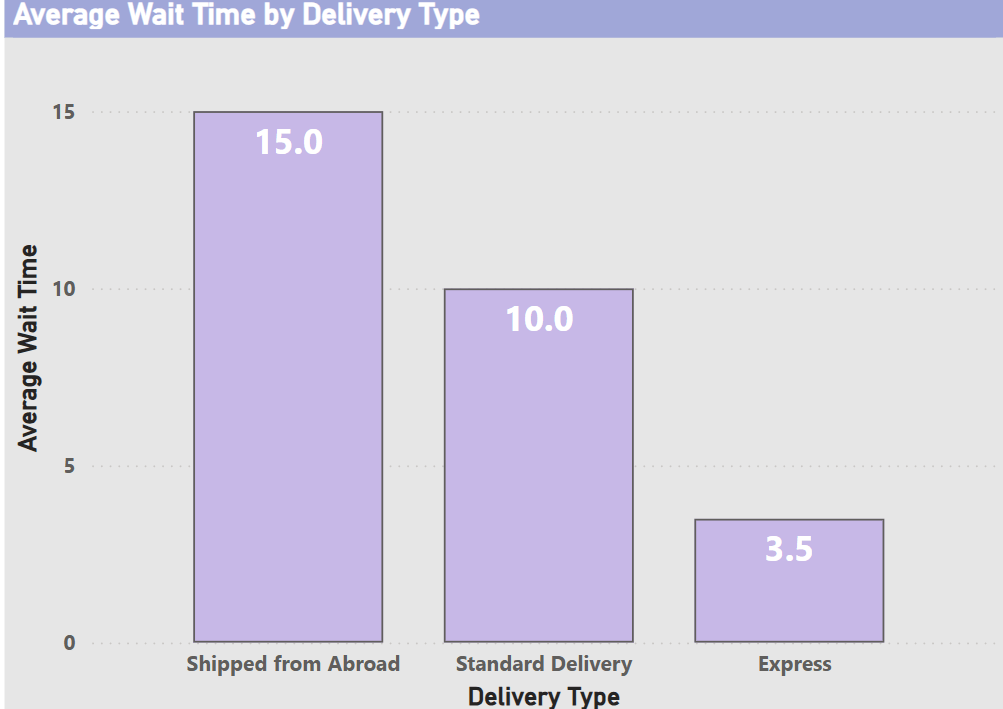
* **Age-Based Scheduling**: Consider the 25-35 demographic's shopping patterns for peak staffing times
* **Category Rotation**: Rotate experienced staff to Electronics during high-traffic periods
* **Efficiency Focus**: Study the 36-50 age group interactions to understand why their wait times are lower.

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

**Approach -**

* To determine the relationship between Delivery type and waiting time, we can use the Measure Average Wait Time and the delivery type.
* We can create a clustered column chart with Delivery type and Average Wait time across different delivery types.

**Visualization** -

**Result/Observations -**

* **There is an inverse relationship between delivery type and average customer wait time.**
* **Faster delivery options result in less customer wait time, while in the case of orders requiring shipments takes a much longer time.**
* **Shipped from Abroad has the highest average wait time at 15 days. This suggests potential delays in international logistics, customs processing, or sourcing from non-local warehouses.**
* **Standard Delivery averages 10 days, which is still significant - This could reflect longer domestic transit routes or non-prioritized fulfillment.**
* **Express Delivery has the lowest average wait time at 3.5 days, making it the fastest and most efficient option.**

**Recommendations -**

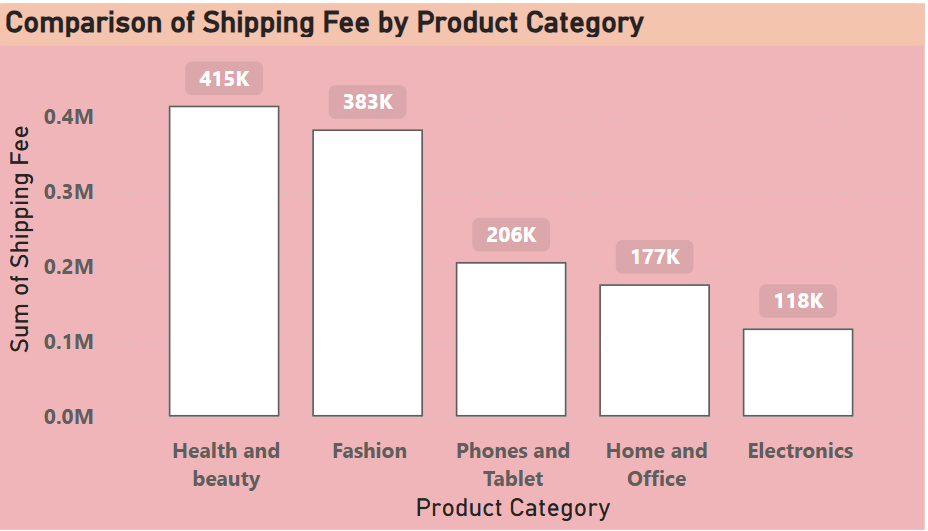
* Long wait times for "Shipped from Abroad" might negatively impact satisfaction. Amazon should consider warning customers of longer delivery windows at checkout.
* **Logistics Optimization**: Evaluate the feasibility of regional warehousing or bulk importing to reduce the average wait time of overseas deliveries.
* **Promotional Strategy**: Amazon should consider offering Express Delivery as a loyalty perk or upsell during checkout to boost revenue and satisfaction.
* **Tiered Pricing Strategy**: Introduce dynamic pricing where customers can opt for faster delivery at a premium (e.g., prioritize Express for high-value orders)

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

**Approach -**

* To determine the relationship between shipping charges and product type, we can create a clustered column chart to show a comparison of shipping fees across different product categories.

**Visualization -**

****

**Result/observation -**

* The **Health and Beauty** category records the **highest total shipping fee at ₹415K**, indicating either a high volume of orders or the need for special handling, packaging, or international shipping. This suggests customers frequently purchase these items, possibly from overseas sellers, or that the nature of the products (e.g., liquids, cosmetics) drives up shipping costs.
* **Fashion** products contribute a substantial ₹383K to the overall shipping fees, positioning them just behind Health and Beauty. This might be due to 2 reasons

1. Consumers making frequent purchases.
2. preference for international styles shipped from abroad.

* **Phones and Tablets**: ₹206K & **Home and Office**: ₹177K - These categories show moderate shipping costs. Phones might be expensive but compact, hence lower in total shipping impact.
* **Electronics** have the **lowest total shipping fee** at **₹118K**  due to reasons -

1. Lower order volume
2. Local availability
3. Efficient packaging/shipping mechanisms
4. Come up with strategies to decrease the low rating orders after analyzing different factors like waiting time, shipping type, unit price, etc.

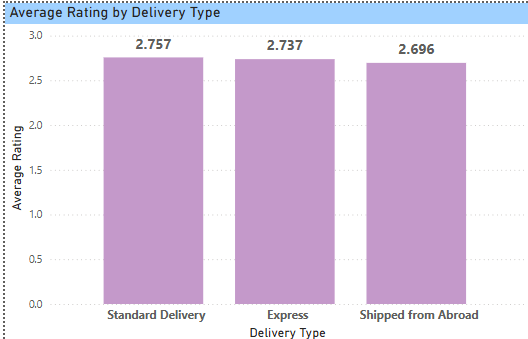
**Approach -**

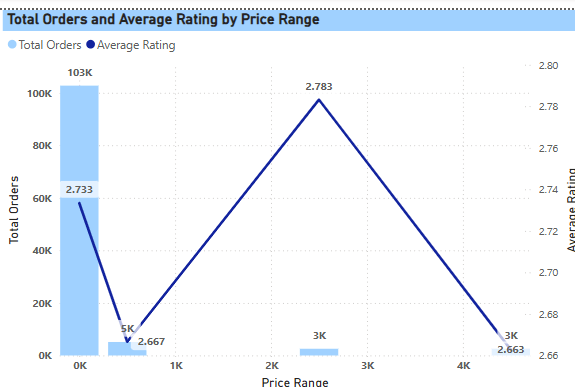
* To devise strategies for low rating orders, firstly, we need to analyze each factor, i.e.,

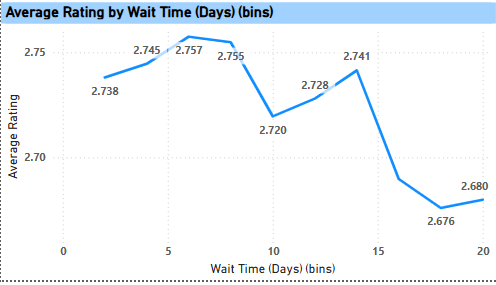
1. Wait time v/s Average Rating.
2. Shipping/Delivery type v/s Average Rating.
3. Unit price v/s Average Rating.

* Secondly, create a Line Clustered column chart for each of these factors.
* Thirdly, in case of Wait time, we can create wait time bins for a better visual experience. In addition to the Unit price bin as well.

**Visualization -**







**Results/Observation** -

* The lowest-priced products (0–1K) have the highest number of orders (103K) but a moderate rating (2.733).
* Mid-priced products (2K–3K) receive the highest average rating (2.783) despite low order volume (3K).
* The highest-priced bracket (4K+) shows lower ratings (2.663) and low order volumes.
* Low-priced products, although popular, might have quality or expectation mismatch issues, affecting ratings.
* Average rating **peaks at around 5–7 days** of wait time (≈ 2.757).
* Ratings **decline significantly after 10 days**, especially **beyond 15 days** (2.676–2.680).
* There is a **negative correlation between wait time and satisfaction** beyond a reasonable delay.
* Wait time exceeding 10 days appears to **negatively impact customer satisfaction**.
* **Standard Delivery** yields the **highest ratings (2.757)**.
* **Express Delivery** surprisingly has slightly **lower ratings (2.737)**.
* **Shipped from Abroad** has the **lowest ratings (2.696)**.
* Despite its name, **Express Delivery may not be meeting expectations**, possibly due to failed promises or high costs.
* International orders are **prone to delays**, causing lower satisfaction.

**Recommendations** - To decrease the Low Rating orders, Amazon needs to devise a few strategies as follows -

* **Improve the quality perception** of low-cost items by enhancing product descriptions, showcasing more customer reviews, and offering better packaging.
* **Bundle low-cost items** with complementary products or discounts to increase perceived value.
* Consider setting a **minimum quality benchmark** for products listed in the lowest price tier.
* Implement a **smart routing algorithm** to prioritize deliveries for high-demand locations using the nearest warehouses.
* Flag orders estimated to take **over 12 days**, and **notify customers proactively** with expected delays or compensation (e.g., coupons or discounts).
* **Reduce international shipments** by identifying popular overseas products and **stocking them locally**.
* Add **delivery-time disclaimers** and “Ships from Abroad” tags with realistic expectations.
* Offer **incentives** (e.g., free returns or loyalty points) for customers who choose overseas shipping despite delays.

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, sales.

**Approach** -

* Firstly, we can create a New Date Table for convenience

DateTable =

ADDCOLUMNS (

CALENDAR (MIN(Orders[OrderDate]), MAX(Orders[OrderDate])),

"Year", YEAR([Date]),

"Month", FORMAT([Date], "MMMM"),

"MonthNumber", MONTH([Date]),

"YearMonth", FORMAT([Date], "YYYY-MM"))

* Secondly, we can create a relationship between the DateTable date & Orders Table order date.
* Thirdly, we can create a New Measure using the Time Intelligence Function to calculate previous year sales.

Previous Year Revenue =

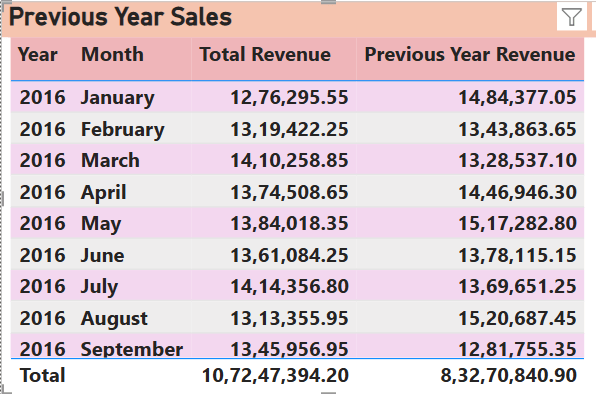
CALCULATE(

[Total Revenue],

SAMEPERIODLASTYEAR(DateTable[Date]))

* Further, we can use a Table Visual to show the Year, Month, Total Revenue, Previous Year sales

**Visualization -**



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

A **Power BI Gateway** is a bridge that connects on-premises data sources (like SQL Server, Excel, SharePoint, etc.) with Power BI services in the cloud. It enables secure data transfer between an organization's internal network and the Microsoft cloud environment, allowing real-time or scheduled data refreshes in Power BI reports and dashboards.

**Some of the Use Cases of Power BI Gateway -**

* **Scheduled Data Refresh**Automates the process of updating reports and dashboards with the latest data from on-premises sources without manual uploads.
* **Live/DirectQuery Connections**Enables real-time data querying directly from on-premises databases without storing data in the Power BI service.
* **Hybrid Data Architecture**Supports organizations using both cloud and on-premises data sources, ensuring seamless integration and reporting.
* **Data Security and Compliance**

Ensures that sensitive data remains within the organization’s firewall while allowing secure access for reporting purposes.

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

In the absence of predefined objective or subjective questions, firstly I would understand the business context and objectives. Here as a data analyst at Amazon the goal is to help maintain strong performance and identify new strategies to enhance the customer experience, increase engagement, and build loyalty.

1. **Business Objectives**

In this course, the first step would be to internalize the broader business goals:

* Sustain Amazon's current high performance
* Improve customer experience and satisfaction
* Identify and reward valuable customers
* Develop new strategies, such as offering discounts and enhancing Prime membership benefits.

1. **Data Exploration and Cleaning**

Secondly, I would conduct a thorough examination of the dataset:

* Understand the structure and fields in the Orders and Customers tables
* Identify key variables such as product details, subcategory, order value, Order date, return status, customer demographics, loyalty tier, and rating.
* Check for missing values, data quality issues, and ensure data consistency.

1. **Develop Key Metrics**

Using the dataset, I would generate several performance and customer-related metrics:

* Total revenue, number of orders, and average order value
* Monthly and seasonal trends in sales
* Product return rates and associated revenue loss
* Average product ratings and review sentiment
* Customer purchase frequency and recency
* Delivery timelines and delays.

1. **Segment Customers for Deeper Insights**

Customer segmentation would be a core component of the analysis. I would classify customers based on:

* Age group
* Region or delivery zone
* Loyalty tier and Prime membership status
* Purchase frequency and average spend
* Return behavior and product ratings given

1. **Identify Opportunities for Improvement and Rewards**

From the insights gathered, I would look for actionable opportunities, such as:

* Rewarding high-spending and high-frequency customers with exclusive offers or loyalty perks
* Encouraging customers who give high ratings with early access to new products
* Identifying frequent returners and addressing their issues through personalized support or product recommendations
* Offering incentives to customers in regions with lower engagement or delayed deliveries.

1. **Develop and Visualize Strategic Insights through PowerBI**

Finally, I would present a set of practical, data-backed recommendations to stakeholders, such as:

* Launch loyalty campaigns based on customer segments
* Improve product listings or supplier quality for frequently returned items
* Expand Prime benefits in regions with slower delivery times