Report

Objective Questions:

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

Ans There are total 3 attributes in the customer table that is customer id, customer age and customer gender.

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

Ans To get customer ages in the order table we need to join customer id that is unique id of customer table to join customer id of order table.

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

Ans In orders table most of the row reason column is empty so I removed reason column from order tables by following steps

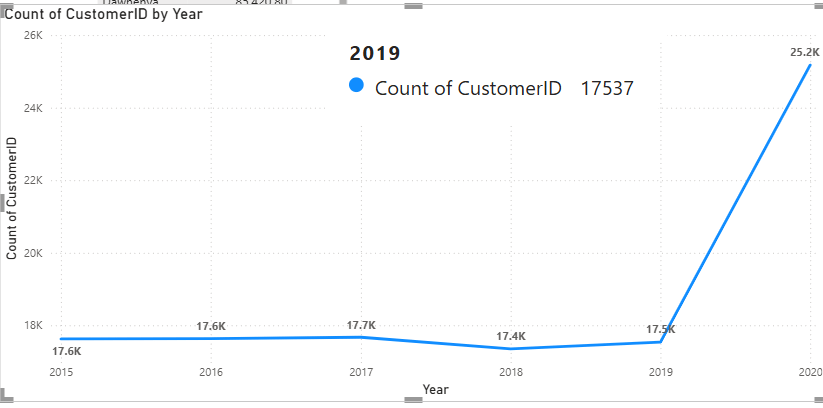
1. How can we calculate the total revenue generated by all the sales?

Ans To calculate revenue generated by all the sales we need to sum sales prices column



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

Ans Between 2015 and 2019, there was an average of approximately 17,600 unique customers making purchases each year. However, from 2019 to 2020, we experienced a significant surge, with the number of customers increasing to 25,000 annually.



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

Ans By count distinct product we can find total number of unique product

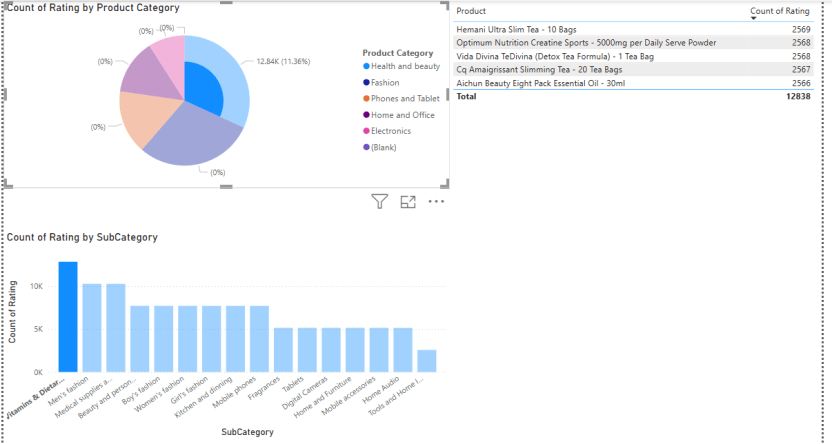


1. What is the average number of days it takes for products to be delivered, get the metric for only the delivered orders.

Ans By calculating the difference between the order date and the delivery date, we obtained the total number of days for each order's delivery. After applying a filter to include only those orders that were delivered, we found that the average delivery time is 9 days.

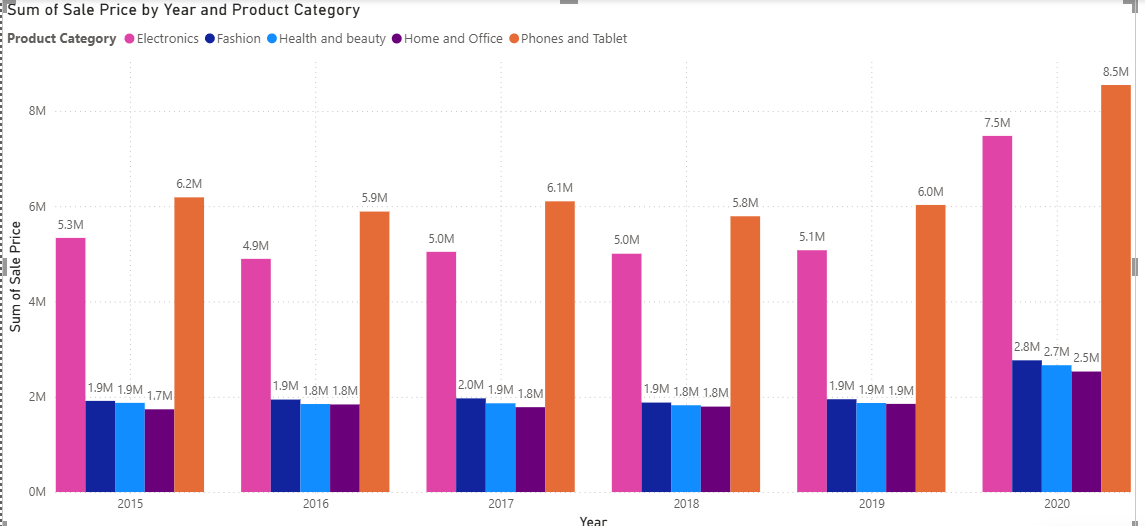
1. Which products, categories, and subcategories are the most popular?

Ans Beginning with the product categories, the most highly-rated category is Health and Beauty. Within this category, the top-rated subcategory is Vitamins and Dietary Supplements. The most popular product in this subcategory is Hermani Ultra Slim Tea.



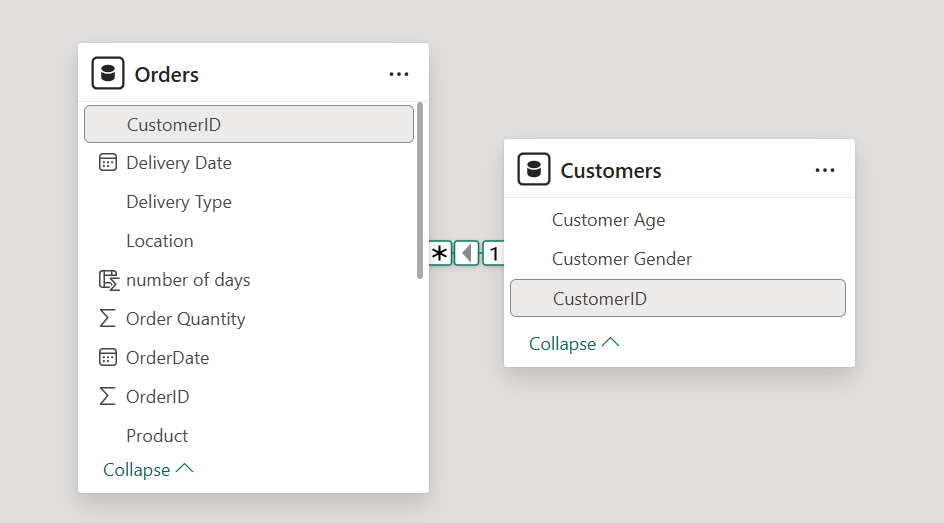
1. Which products have seen an increase or decrease in sales over the year?

Ans Using a clustered bar chart, we can analyze the increase in sales over the year. The chart clearly shows that Electronics, as well as Phones and Tablets, experienced a significant surge in sales during the years 2019 and 2020.



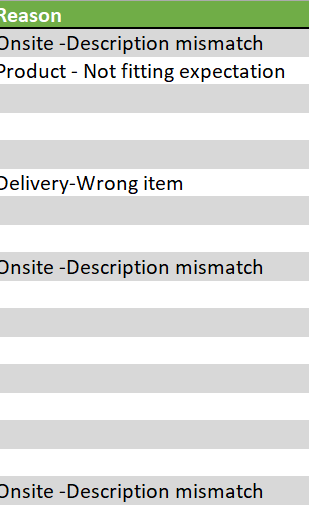
1. While modeling the data relationships, what will be the type of relationship between the customer ID of Orders and customer tables?

Ans The relationship between the Orders and Customers tables is one-to-one, based on the CustomerID. Each CustomerID in the Customers table corresponds to at most one CustomerID in the Orders table, and vice-versa. This means a single customer can have only one order, and each order belongs to only one customer.



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

Ans There is a column named "Reason" in the given table that contains null values. I handled these null values by following the steps below.



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

Ans Yes, there were null values in the Reasons table, so I deleted it.

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

Ans When you add a column in Power Query, the M language code typically appears as:

`= Table.AddColumn(Source, "NewColumnName", each [Column1] + [Column2])`

This is the common syntax that appear in formula bar.

M-query, also known as Power Query Formula Language, is a functional and case-sensitive language used for data transformation, including filtering, grouping, merging, and aggregating data.

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.

Ans Sql query to solve the question is

select Customer\_ID, (sum(Sale\_Price)\*0.5 + Order\_Quantity\*0.3 + avg(Sale\_Price)\*0.2) composite\_score from orders

group by Customer\_ID,Order\_Quantity

order by composite\_score desc

limit 5

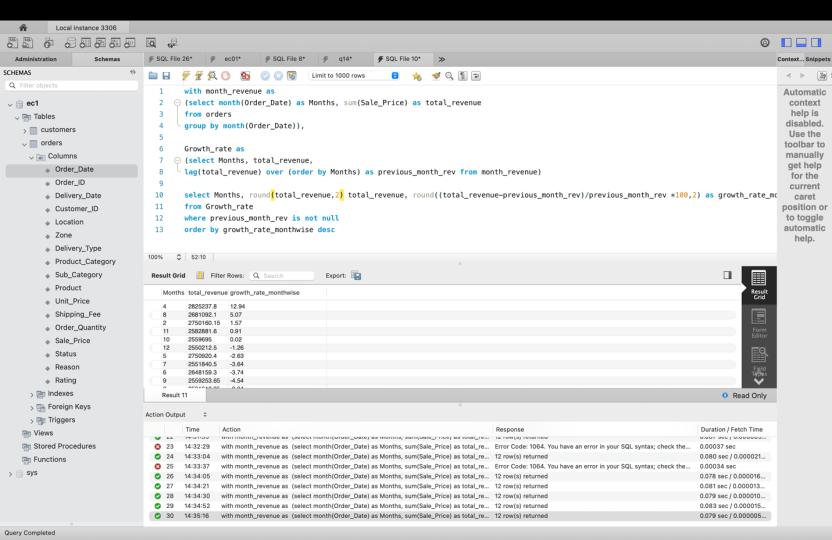
A screenshot of a black screen

AI-generated content may be incorrect.

This are the top 5 customer who are in top in all there criteria that is total revenue,

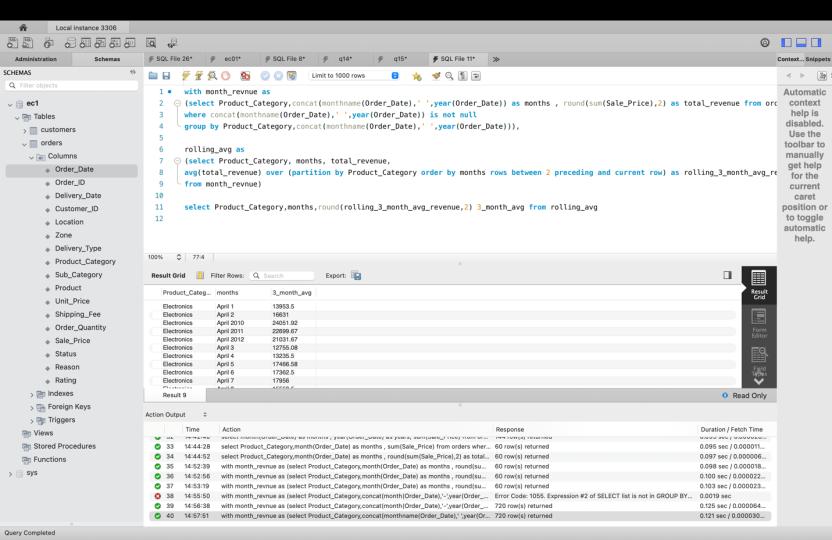
Order frequency, average order value

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

Ans month over month growth rate in total revenue

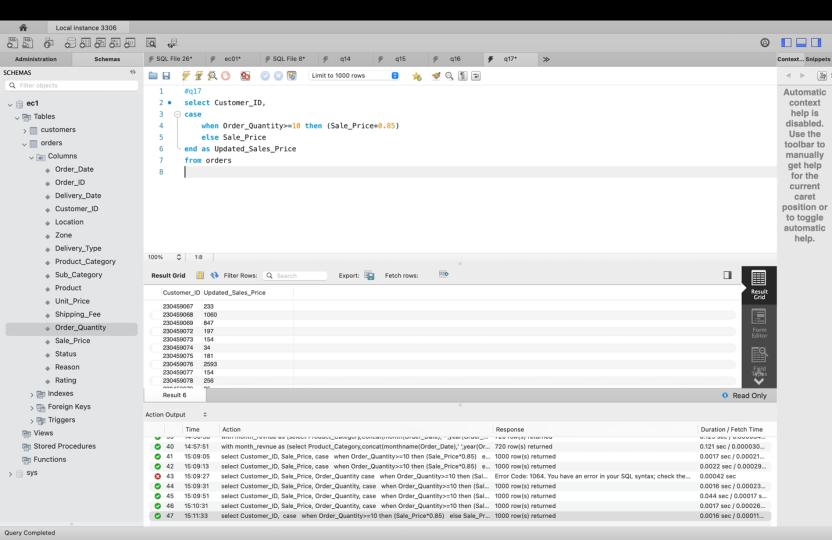
1. Calculate the rolling 3-month average revenue for each product category. (SQL)

Ans Three month rolling average is calculated as followed



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)

Ans updated the orders tables are in below image



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

Ans To calculate average numbers of days I used following the steps. Output is null because every customer is unique customers the count of order id and count of customer id is same.

A screenshot of a computer

AI-generated content may be incorrect.

A close up of a number

AI-generated content may be incorrect.

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

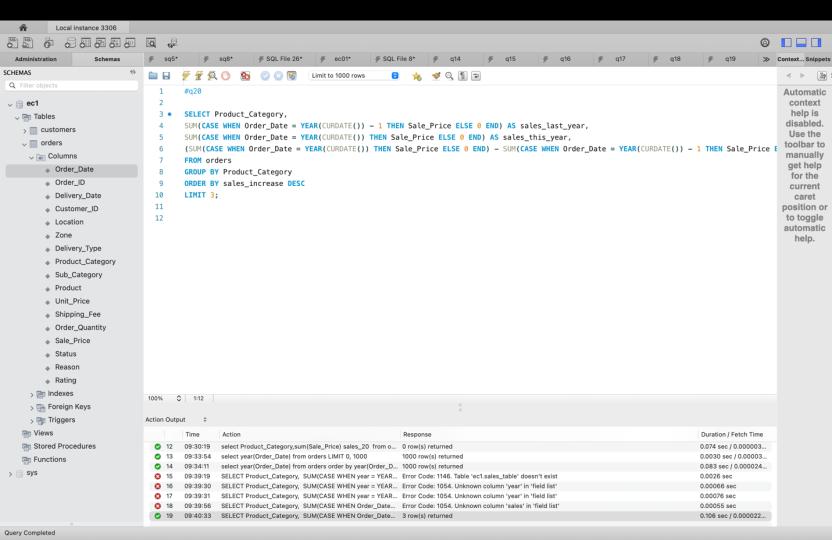
Ans Below image represent the customer who generated revenue that is more than 30% the average revenue per custome.

A screenshot of a computer program

AI-generated content may be incorrect.

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)

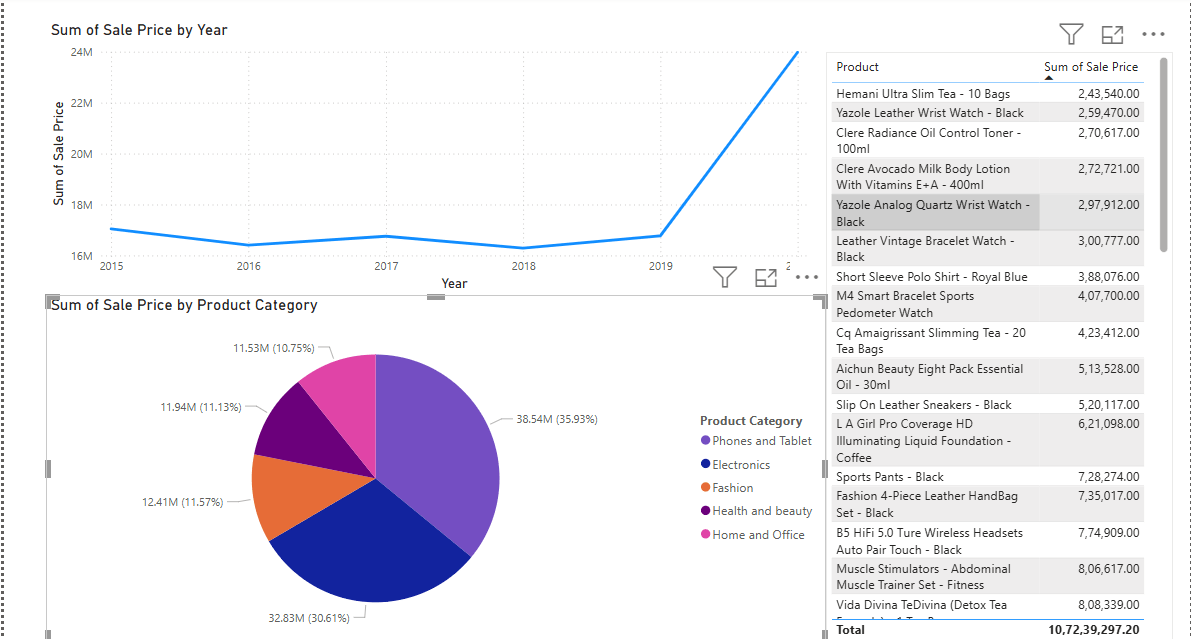
Ans Below image describe top 3 category that shown highest increase in sales over the past year



**Subjective Question:**

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

**Approach**: We used a line chart with order date by year and sales price to understand the yearly revenue breakdown. For product revenue breakdown, we analyzed sales by product category and individual products to easily identify trends in sales increases and decreases.



**Insights**: From 2015 to 2019, revenue exhibited minor fluctuations with a slight increase and decrease. However, there was a significant surge in revenue from 2019 to 2020 within a single year.

When analyzing product category data, it is evident that phones and tablets contribute the highest share at 36%, followed by the electronics category at 30.61%. Other categories hover around 10% to 11%.

Delving further into product details, it is observed that items such as Hemani Ultra Slim Tea and Yazole Leather Wrist Watch, which fall under the health and beauty category, generated the lowest sales revenue.

**Suggestion**: There are several strategies to increase the sales of these items:

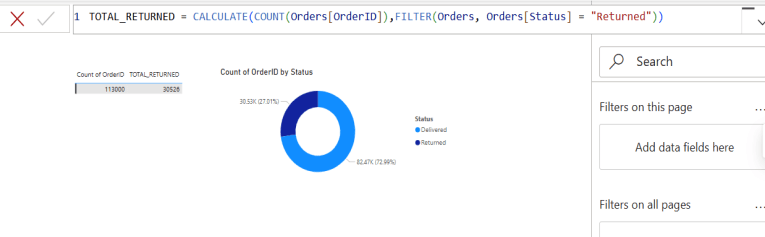
* We can bundle the low-selling items with more popular items, offering them at a discounted price when purchased together.
* We should gather customer feedback to understand why these items are not performing well in the market.
* We should enhance product packaging and descriptions, and offer more free trial opportunities, especially since these products fall under health and beauty, where customers prefer to try before purchasing.

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 identify the potential reasons for returns, we used a DAX function to determine the total number of returned orders. The formula is:

TOTAL\_RETURNED = CALCULATE(COUNT(Orders[OrderID]),

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



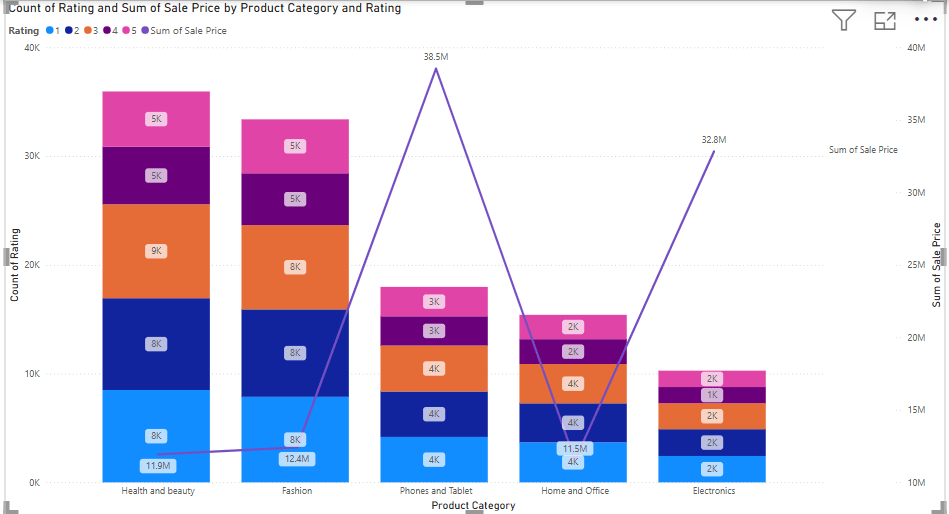
**Insights**: The image shows that a total of 113,000 orders were placed, with 30,526 orders (27%) returned and 73% successfully delivered. The gauge chart provides a clear view of the number of orders returned and delivered.

**Suggestion**:

* Ensure that product descriptions are accurate and detailed. This can help set the right expectations for customers and reduce the likelihood of returns.
* Implement stricter quality control measures to ensure that products meet customer expectations. This can help in reducing the number of defective items being shipped.
* Collect and analyze customer feedback on returned items. This can provide insights into common issues and areas for improvement.
* Review and possibly revise the return policy to make it more customer-friendly while also protecting the business. Clear communication of the return policy can also help in managing customer expectations.

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**: To determine whether product ratings influence sales, we used line and stacked graphs. The parameters are product category, sum of sale prices, and count of ratings. The x-axis represents the product category, the column y-axis represents the count of ratings, and the line y-axis represents the sum of sales prices.



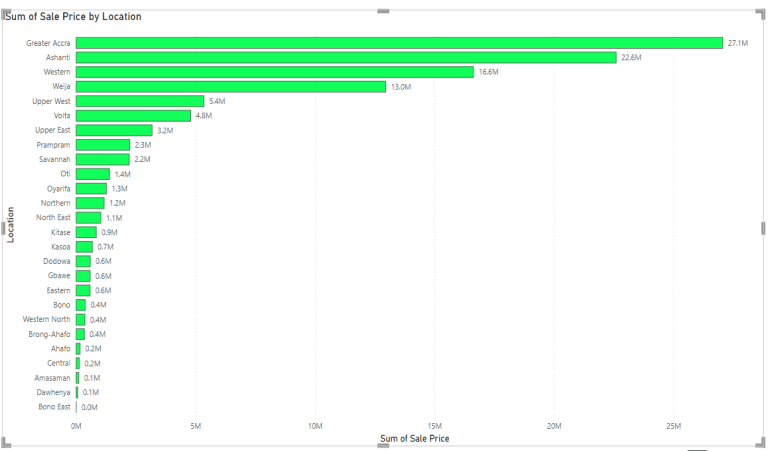
**Insights**: The graph clearly illustrates that the Health and Beauty category has the highest number of ratings, yet its sales revenue is the lowest compared to all other categories. Within this category, only 5,000 customers gave a 5-star rating, while 8,000 customers provided a lower 1-star rating. Approximately 9,000 customers, which constitutes the majority, rated it with 3 stars. This data suggests that there is no direct correlation between ratings and sales; despite an increase in ratings, sales have not increased proportionately.

**Suggestion**:

* Dive deeper into the customer feedback for the Health and Beauty category. Understanding the reasons behind the low ratings can provide valuable insights into areas that need improvement.
* Based on the feedback, work on improving the quality of the products in this category. This could involve better sourcing of materials, improved manufacturing processes, or enhanced packaging.
* Consider revising the marketing strategies for these products. Highlighting the improvements made based on customer feedback can help regain customer trust and boost sales.

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**: To analyze revenue distributions across various locations, we utilized a cluster bar chart with the sum of sales price on the x-axis and location on the y-axis.



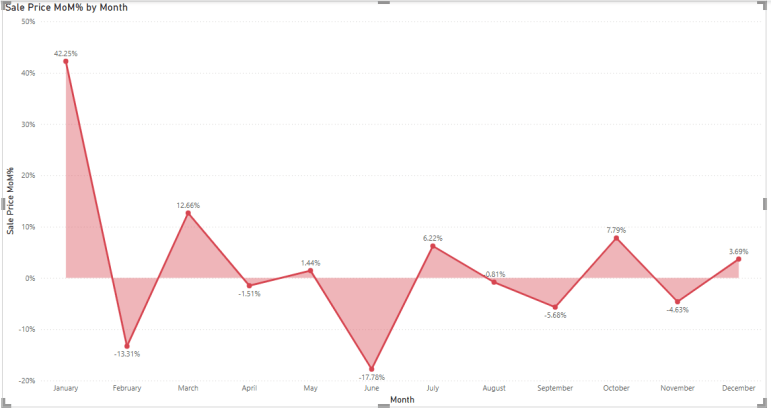
**Insights**: Utilizing the graph, it is evident that among the 27 locations, the top five with the highest sales are Greater Accra with sales of 27.1M, which is the highest, followed by Ashanti with sales of 22.6M, Western with 16.6M, Weija with 13.0M, and Upper West with 5.4M. The remaining locations have sales below the 5M mark. Locations such as Bono East and Dawhenya have not yet begun generating revenue.

**Suggestion**:

* Since regions like Greater Accra, Ashanti, Western are driving the majority of your sales, it makes sense to concentrate your marketing efforts and budget in these areas. Don't spread your resources too thin across other regions that are underperforming.
* Within each region, identify specific areas or communities where your products are most popular. This allows for hyperlocal targeting, where you can tailor your messaging and promotions to the specific needs and preferences of those communities.
* Analyze which marketing channels are most effective in each region. For example, if social media is highly effective in Greater Accra, but radio is more popular in Ashanti, adjust your channel mix accordingly.
* Develop marketing campaigns that resonate with the cultural nuances and local trends in each region. This could involve using local languages, featuring local influencers, or sponsoring local events.
* Ensure that your warehouses and distribution centers serving these high-performing regions have sufficient inventory of your best-selling products. Analyze sales data to predict demand and adjust inventory levels accordingly.
* Evaluate the delivery infrastructure in each region. If delivery times are slow in certain areas, consider partnering with local logistics providers or establishing new distribution points to improve efficiency.
* Invest in customer service resources in these regions to handle the higher volume of sales and customer inquiries. This could involve hiring local customer service representatives or establishing dedicated phone lines or online support channels.

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

**Approach**: To understand the month wise sales in given we use time intelligence function in quick measures that is month on month sales. Than we connected month of month sales to order date indicating only month. We use line chart to understand in which month our sales are going down.



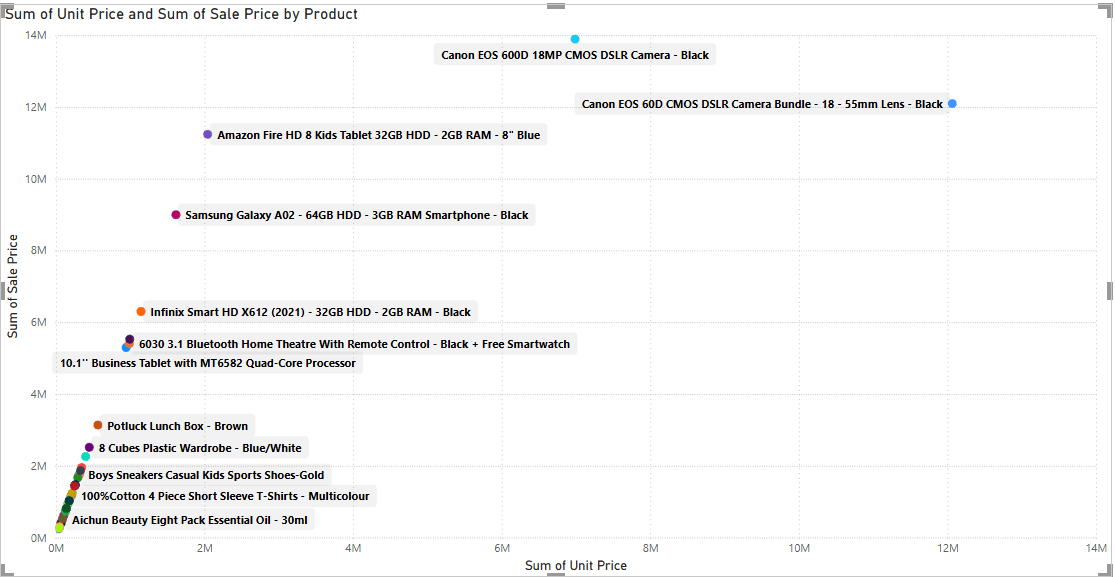
**Insights**: The graph indicates that January has the highest sales percentage at approximately 42.25%. February shows a decline in sales by 13.1%, and June experiences a significant decline of 17.78%. This is a major concern. Sales for other months fluctuate between 12% and -5%.

**Suggestion**:

* Introduce seasonal promotions and discounts for February and June, such as Valentine's Day promotions in February and summer sales in June to attract more customers.
* Create a sense of urgency with limited time offers and flash sales to encourage customers to make quick purchases to take advantage of the deals.
* Offer product bundles at a discounted price to help increase the average order value and make the deals more appealing to customers.
* Implement or enhance loyalty programs to reward repeat customers. Offering double points or exclusive discounts during these months can incentivize more purchases.
* Use targeted advertising on social media and other platforms to reach potential customers and highlight the special promotions and discounts available in February and June.
* Partner with influencers or other brands to promote products during these months. Collaborations can help reach a wider audience and boost sales.

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

**Approach:** To gain a comprehensive understanding of the market performance of our products, we employed a multi-faceted analysis using three key parameters: the product itself, the sum of unit prices, and the sum of sales prices. By integrating these parameters into a scatter plot graph, we were able to visually represent the relationship between product pricing and their respective contributions to overall sales. This approach allowed us to identify trends, outliers, and potential areas for improvement in our pricing strategy and sales performance.



**Insights:** By utilizing a scatter plot, we can observe that the majority of products are priced reasonably. However, there are two notable exceptions: the Canon EOS 60D CMOS and the Canon EOS 600D 18MP. These products are priced significantly higher, yet their sales figures do not reflect a corresponding increase in demand. This discrepancy suggests that despite their higher price points, these products are not performing as well in terms of sales.

**Suggestion**:

Marketing and Sales Optimization:

* Invest in dedicated marketing efforts for the Canon EOS 60D CMOS and 600D 18MP. These campaigns should prominently showcase their unique features, benefits, and any value propositions that support their higher price point.
* Analyze sales performance across all channels to identify any underperforming areas. Weak performance in a specific channel may point to distribution or presentation problems.
* Gather customer feedback through surveys, reviews, and other means to understand their perceptions of these cameras, especially concerning price. This will provide valuable insights into customer needs and expectations.

Product and Strategy Review:

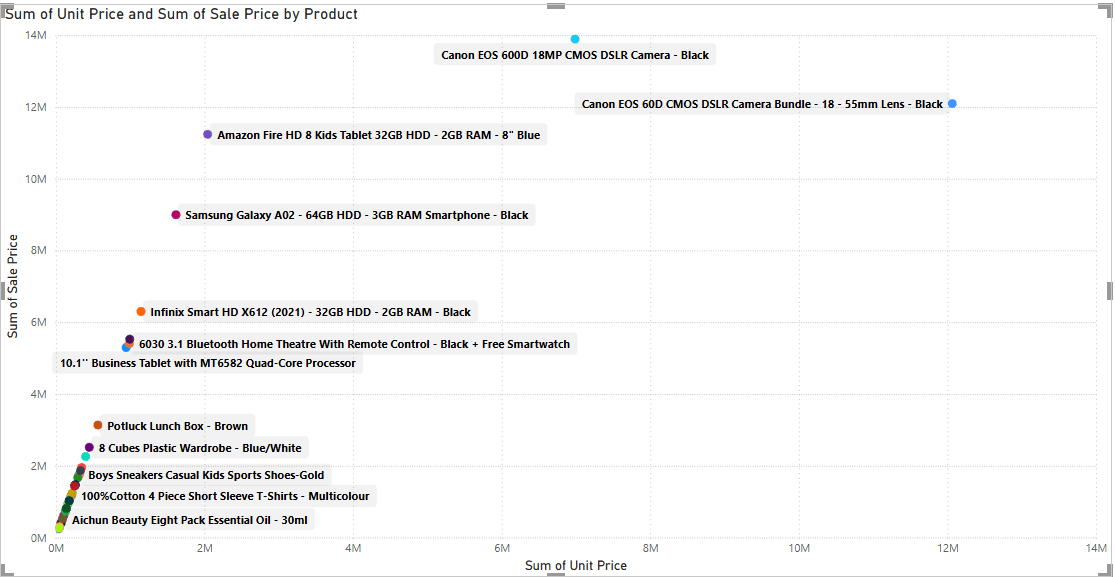
* If price adjustments and marketing efforts are insufficient, re-evaluate the product positioning. Ensure the target audience is clearly defined and that the camera's features are genuinely differentiated and valued by that audience. Consider whether the cameras are being marketed to the appropriate customer segment.
* If possible, consider adding or enhancing features to better justify the current price. This could include software updates, bundle accessories, or highlighting unique capabilities that distinguish these models from competitors.

Inventory and Sales Strategies:

* If sales remain slow, carefully manage inventory levels to prevent overstocking and minimize potential losses.
* Explore alternative sales strategies such as offering refurbished models at a reduced price, creating bundles with complementary accessories, or targeting niche markets with specialized needs.

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

**Approach**: To identify products suitable for discounts, we analyzed the relationship between product price and sales performance. A scatter plot was generated, visualizing the total unit price against total sales price for each product. This allowed us to pinpoint high-priced items with low sales, indicating potential discount candidates.

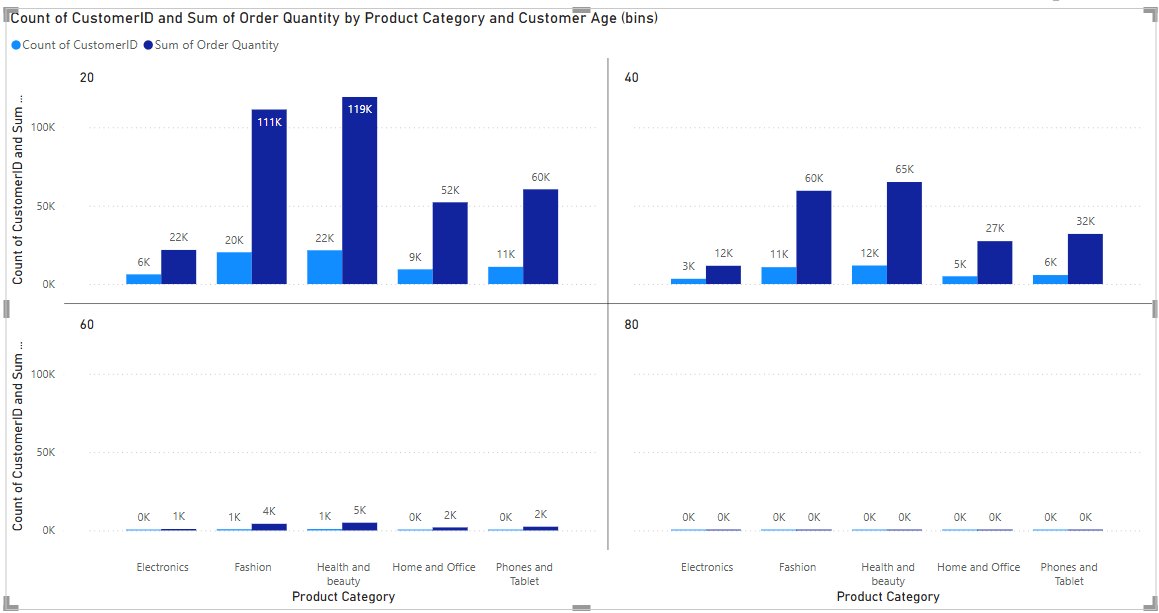


**Insights**: A scatter plot analysis shows that the majority of products have appropriate price points. However, the Canon EOS 60D CMOS and 600D 18MP stand out as outliers with considerably higher prices. Importantly, their sales volumes do not match these elevated prices, suggesting weaker market performance.

**Suggestion**:

* + Implement a substantial percentage discount (e.g., 15-25%) on both the Canon EOS 60D CMOS and 600D 18MP to immediately stimulate sales and bring their prices in line with market demand.
  + Reduce the price of the Canon EOS 60D CMOS and 600D 18MP by a fixed dollar amount, calculated to make them more competitive within their product category.
  + Run a limited-time promotional discount (e.g., a 'flash sale' or 'weekend special') on these two models to create a sense of urgency and drive immediate sales.
  + Offer a bundle deal that includes the Canon EOS 60D CMOS or 600D 18MP with complementary accessories (e.g., lenses, memory cards, camera bags) at a discounted package price.
  + Introduce a trade-in program, allowing customers to trade in their old camera equipment for a discount on the Canon EOS 60D CMOS or 600D 18MP.
  + Offer an extended warranty or free camera servicing with the purchase of either of these models to enhance their perceived value.
  + Offer exclusive discounts on these models to existing loyalty program members.
  + Provide special discounts to students or educators to tap into a specific market segment.

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**:

Parameters Used:

* Customer ID Count: To track the number of unique customers.
* Sum of Order Quantity: To understand the total quantity purchased.
* Product Category: To identify the range and type of categories customers are purchasing from.
* Customer Age: To group customers into age brackets for further analysis.

Bucketing Process:

* Customers were segmented based on the count of unique customer IDs and the sum of order quantities, grouped according to the product categories they purchased from. Age segmentation was applied, grouping customers into 20-year intervals (e.g., 0–20 years, 21–40 years, etc.) for a demographic perspective.

**Insights**: As per age groups:-

* The "Health and beauty" category has the highest "Sum of Order Quantity" at 119K. This suggests that this age group orders the largest number of items in this category.
* The "Fashion" category has the second-highest "Sum of Order Quantity" at 111K. This indicates a strong interest in fashion items within this age group.
* The "Electronics" category shows the lowest "Sum of Order Quantity" at 6K. This implies that this age group orders the fewest number of electronic items compared to other categories.
* Notice that while "Health and beauty" has the highest order quantity (119K), the "Count of CustomerID" (22K) is similar to "Fashion" (20K). This suggests that while a similar number of customers are buying from both categories, those buying "Health and beauty" are likely purchasing a larger quantity of items per order on average.
* "Home and Office" and "Phones and Tablet" have relatively lower order quantities (52K and 60K respectively) compared to "Health and beauty" and "Fashion". The customer counts for these categories (9K and 11K respectively) are also lower.
* A similar trend in terms of order quantity is observed in the "40" age group. "Health and beauty" has the highest "Sum of Order Quantity" at 65K. "Fashion" is the second highest at 60K. "Electronics" has the lowest at 3K.
* Lower Overall Order Quantities: It's important to note that the overall order quantities across all categories appear to be lower in the "40" age group compared to the "20" age group. This could indicate differences in purchasing power, needs, or shopping habits between these age groups.
* Customer Count Observations: Similar to the "20" age group, the customer counts ("Health and beauty": 12K, "Fashion": 11K) are relatively close despite the difference in order quantities. This again suggests a higher average number of items per order in the "Health and beauty" category.

**Suggestion**:

* + Tiered Beauty/Fashion Insider: Offer points per purchase in these categories, with increasing benefits at higher tiers (e.g., early access to new arrivals, exclusive discounts, free samples, personalized style consultations).
  + Beauty/Fashion Subscription Box Perks: If you offer subscription boxes, provide loyalty points or discounts for subscribers who maintain their subscriptions or refer friends.
  + Social Sharing Rewards: Encourage user-generated content (reviews, photos with purchases) related to beauty and fashion with loyalty points or contest entries.
  + Birthday/Milestone Rewards: Offer special discounts or gifts on birthdays related to beauty or fashion items.
  + Tech/Home Upgrade Program: Offer discounts on future purchases in these categories after a certain number of purchases or a specific spending threshold.
  + Bundle & Save Rewards: Incentivize purchasing related items together (e.g., phone accessories with a phone).
  + Educational Discounts: If applicable, offer discounts for students on tech or home Mature & Home/Tech Interested: Show some engagement in "Home and Office" Loyalty Program Suggestions:
  + Home & Tech Loyalty Tiers: Similar to the Beauty/Fashion tiers, offer increasing benefits for repeat purchases in these categories (e.g., extended warranties, installation discounts, smart home consultations).
  + Trade-In Programs: For electronics, offer discounts for trading in older devices.

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?

Ans

* **Approach**: Used Data Analysis Expressions (DAX) to calculate total sales in the fashion category with overseas delivery. DAX, a formula language for Power BI, consists of functions, operators, and constants for data analysis and calculations.



other DAX function I used to find total returned which helped me identify how many products are returning and whats the reason behind returning.



A blue bar graph with black text

AI-generated content may be incorrect.

**Insights**:

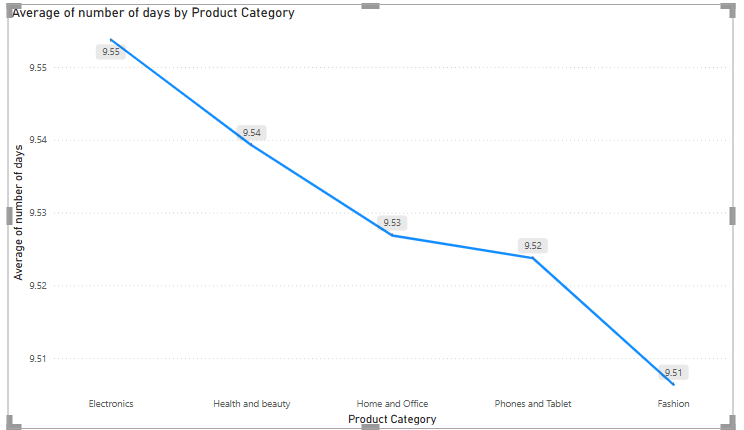
* The use of this DAX function proved beneficial in filtering and sorting total sales according to the product category "Fashion" and the delivery type "Shipped from Abroad." Through this analysis, it was determined that the total sales for the fashion category from overseas deliveries amounted to 4.14 million.
* This comprehensive filtering mechanism provides valuable insights into the performance and distribution of products delivered from abroad, enabling more informed decision-making and strategic planning for future sales initiatives. Furthermore, understanding the sales dynamics specific to overseas shipments assists in identifying potential areas for optimization and growth within the global market.

**Suggestion**:

* + Develop marketing campaigns specifically targeted at the high-performing international markets, considering their cultural and preferences.
  + Consider collaborations with local influencers, retailers, or distributors to expand reach and build trust.
  + Ensure your website is localized for key international markets (language, currency, payment methods) to improve the customer journey.
  + Since this segment generated a significant 4.14 million in total sales, it's a strong area to continue investing in and potentially expand.
  + The analysis highlights the importance of understanding sales dynamics for overseas shipments. This suggests a need to further investigate other product categories or delivery types in the global market to identify potential areas for improvement or expansion.

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

**Approach**: This analysis aims to optimize scheduling and staffing by examining the variation in average wait times across different product categories. The methodology involves a two-parameter approach, considering average delivery days and product categories. A line graph was utilized to visualize the average delivery time to end customers for each category, facilitating the identification of correlations between delivery duration and product type.



**Insights**: The line graph reveals that the average delivery times across the five product categories (Electronics, Health and beauty, Home and Office, Phones and Tablet, and Fashion) are remarkably similar, ranging from a high of 9.55 days for Electronics to a low of 9.51 days for Fashion. This suggests that, on average, the delivery timeframe for all product categories is consistently around 9.5 days, with minimal variation between them.

**Suggestion**:

* **Standardized Processes:** This could indicate that logistics and fulfillment processes are fairly standardized across different product categories, leading to consistent delivery times.
* **Limited Category-Specific Issues:** It might suggest that there aren't significant category-specific bottlenecks or efficiencies that drastically impact delivery times.
* **Focus on Overall Improvement:** Instead of focusing on optimizing individual product category delivery times, our efforts might be more impactful if directed towards improving the overall delivery process for all categories. Even small improvements to the general process could benefit all product lines.
* **Customer Expectation Management:** we can likely set similar delivery expectations for customers across all product categories, simplifying communication.

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

**Approach**: To gain comprehensive insights into the relationship between **delivery type** and customer experience, a multi-parameter analysis was conducted. The primary parameters under investigation are the **delivery type**, the **average waiting time**(quantified as the number of days between order placement and product delivery), and the **count of customer IDs**associated with each delivery type (representing the volume of orders). A **line and clustered column chart** was employed for visualization. The **clustered columns** represent the **average waiting time** for each delivery type, allowing for direct comparison. An **overlaid line** displays the **count of customer IDs** for each corresponding delivery type, providing insight into the popularity or usage of each method. This analysis aims to identify potential inefficiencies or areas for optimization by simultaneously examining delivery speed and adoption rates across different delivery processes.

A graph of a number of different types of goods

AI-generated content may be incorrect.

**Insights**

* **Significant Discrepancy in Delivery Speed:** The data reveals a substantial variation in average delivery times across the different methods. "Shipped from Abroad" stands out with a significantly longer average of 15 days, contrasting sharply with the rapid 3-day average for "Express" delivery. "Standard Delivery" occupies a middle ground at 10 days. This highlights fundamentally different logistical processes and potentially geographical distances involved in each option.
* **Relationship Between Speed and Volume at the Extremes:** An intriguing observation is the inverse correlation at the extremes of the delivery speed spectrum. The delivery method with the longest average waiting time ("Shipped from Abroad") commands the largest customer base. Conversely, the method offering the quickest delivery ("Express") attracts the fewest customers. This suggests that factors beyond just the speed of delivery are primary drivers of customer choice for a significant portion of your clientele.
* **Popularity of the Middle Ground:** "Standard Delivery" demonstrates a strong appeal, capturing a customer volume nearly equivalent to the most popular (but slowest) option. This indicates that a considerable segment of your customers finds the 10-day delivery timeframe acceptable, potentially representing a balance between cost and speed expectations.
* **Relatively Even Customer Distribution Despite Varying Speeds:** While "Express" has a noticeably lower customer count, the overall distribution of customers across the three delivery types is surprisingly balanced. This suggests that different segments of your customer base have distinct needs and priorities when it comes to receiving their orders, leading to a fairly even spread across the available options, even the one with a significantly longer wait.
* **Potential Latent Demand or Unmet Need:** The low adoption of "Express" delivery, despite its speed advantage, could indicate a latent demand that isn't being fully met or understood. It raises questions about whether the current pricing, visibility, or perceived value proposition of express aligns with customer needs.

**Suggestion**:

**Focused on "Shipped from Abroad" (Longest Wait, Highest Demand):**

* **Investigate and Optimize the Logistics Chain:** Conduct a thorough analysis of the "Shipped from Abroad" process to identify potential bottlenecks and areas for optimization. Even small reductions in the 15-day average could significantly improve customer satisfaction for the largest customer segment. This could involve exploring different carriers, streamlining customs procedures (if applicable from Indore), or optimizing the initial dispatch process.
* **Explore Tiered "Shipped from Abroad" Options (if feasible):** If logistical constraints allow, investigate the possibility of introducing a slightly faster, albeit potentially more expensive, tier within the "Shipped from Abroad" category to cater to customers willing to pay a bit more for quicker delivery of these specific items.

**Focused on "Express" (Shortest Wait, Lowest Demand):**

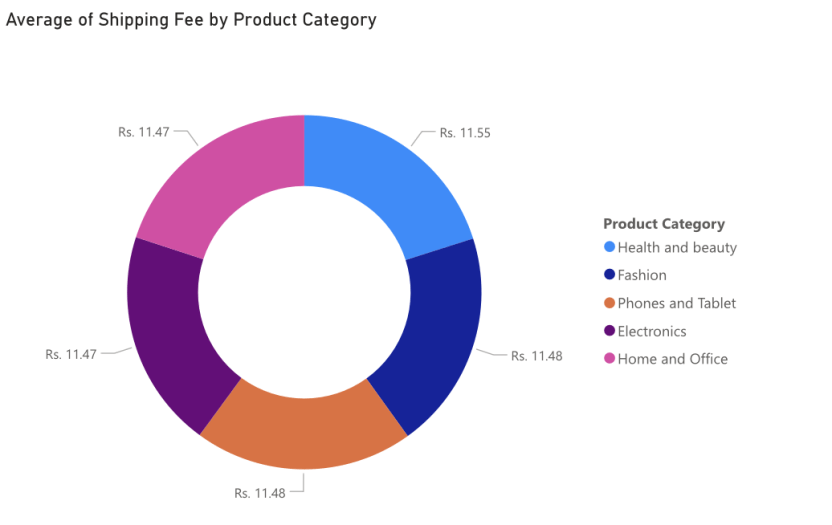
* **Re-evaluate Pricing Strategy for "Express":** Analyze the current pricing of "Express" delivery in the context of your target market in Indore and surrounding areas.
* **Increase Visibility and Promotion of "Express":** Ensure "Express" delivery is prominently displayed during the checkout process and in marketing materials. Highlight the significant speed advantage (3 days) and target customers who might have urgent needs.
* **Bundle "Express" with Certain Product Categories:** If specific product types tend to have more urgent needs, consider strategically bundling or offering discounted express shipping for those items.

**Focused on "Standard Delivery" (Moderate Wait, High Demand):**

* **Maintain and Monitor Performance:** Continue to ensure the reliability and consistency of the 10-day average for standard delivery, as it caters to a large portion of your customer base. Regularly monitor its performance and address any potential slowdowns proactively.
* **Gather Feedback on "Standard Delivery":** Collect customer feedback specifically on their experience with standard delivery to identify any areas for potential improvement, even if the current timeframe is generally acceptable.

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

**Approach**: To gain insights into how shipping charges vary across different product categories, a two-parameter analysis was conducted focusing on **product type** and the **average shipping charges** associated with each category. A gauge chart was selected as the primary visualization tool to effectively represent the average shipping charges for each product type in relation to a predefined scale or target range. This approach aims to identify product categories with notably higher or lower average shipping costs, potentially revealing underlying factors influencing these charges and informing pricing and logistical strategies.



**Insights**:

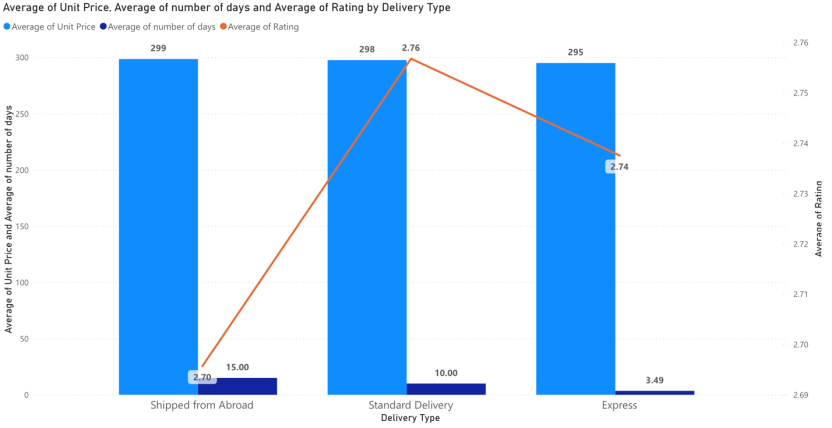
* **Minimal Variation in Average Shipping Fees:** The most striking insight is the extremely small difference in the average shipping fees across all the product categories displayed. The highest average shipping fee is Rs. 11.55 (for Health and Beauty), and the lowest is Rs. 11.47 (for Electronics and Home and Office), resulting in a negligible difference of only Rs. 0.08.
* **Consistent Average Shipping Cost:** This near uniformity suggests that, on average, the cost to ship an item is practically the same regardless of whether it falls into the Health and Beauty, Fashion, Phones and Tablet, Electronics, or Home and Office category.
* **Lack of Category-Specific Shipping Cost Differences :** The data indicates that, when looking at the average, there are no significant cost drivers related to the product category that are influencing the shipping fee. Factors that typically differentiate shipping costs between product types (like significant differences in weight, size, or special handling needs) appear to be either absent or averaged out across these categories.

**Suggestion**:

* **Simplify Shipping Communication:** Communicate a clear and straightforward standard shipping fee that applies to most or all product categories. This makes it easier for customers to understand the total cost upfront and can improve the user experience during checkout.
* **Streamline Internal Processes:** Consistent shipping costs can simplify internal processes related to order fulfillment, invoicing, and customer service, as there's less need for complex calculations based on product type.
* **Review Underlying Cost:** Periodically analyze the underlying cost components (packaging, carrier rates, handling) for each product category to ensure the average consistency isn't masking significant individual product-level variations that could impact profitability.

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

**Approach**: To identify key factors potentially influencing low order ratings, a multi-parameter analysis was conducted. The primary parameters under investigation are **waiting time** (measured in the number of days for order delivery), **shipping type**, and **unit price**. A **cluster line and column chart** was employed for visualization. The **clustered columns** are used to represent the average waiting time and/or unit price for orders associated with different shipping types, allowing for direct comparison. An **overlaid line** is used to represent the proportion or count of low-rated orders within each shipping type category. This approach aims to visually correlate these factors with the occurrence of low order ratings, thereby highlighting potential areas for improvement in service and product offerings.



**Insights**:

* **Strong Inverse Relationship Between Delivery Time and Customer Rating:** The most prominent deep insight is the clear negative correlation between the average number of days for delivery and the average customer rating. The delivery method with the longest average transit time ("Shipped from Abroad" at 15 days) corresponds to the lowest average customer rating (approximately 2.70). Conversely, the delivery methods with shorter average transit times ("Standard Delivery" at 10 days and "Express" at 3.49 days) exhibit higher average customer ratings (approximately 2.76 and 2.74, respectively). This strongly suggests that delivery speed is a significant driver of customer satisfaction.
* **Average Unit Price as a Non-Contributing Factor to Rating Variance (in this dataset):** The remarkably consistent average unit price across all three delivery types (ranging narrowly from approximately 295 to 299) suggests that the product's average price point is not a primary differentiating factor influencing customer ratings related to the delivery experience in this context. The variability in ratings appears to be more strongly tied to the logistics of the delivery itself.
* The data subtly hints at customer tolerance thresholds for delivery times. The noticeable drop in average rating for "Shipped from Abroad" suggests that the 15-day waiting period significantly impacts customer satisfaction. The relatively close ratings for "Standard Delivery" and "Express" might indicate that once delivery is within a certain timeframe (around 3-10 days), the impact on average rating plateaus or becomes influenced by other factors more than just the incremental speed difference.

**Suggestion**:

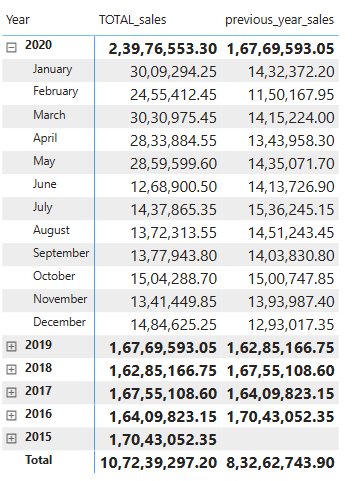
* **Prioritize Reducing Longest Delivery Time:** The strong negative correlation between the longest average delivery time and the lowest customer ratings indicates this is a critical area for improvement. Investigate the logistics chain for this slowest shipping method.
* **Leverage the Success of Mid-Tier Delivery:** Understand the factors contributing to the higher average rating for the delivery option with a moderate timeframe. Identify best practices in this process – carrier reliability, packaging, communication – and see if any of these can be applied to improve the experience of the slowest option.
* **Communicate Realistic Expectations:** Ensure clear and transparent communication of delivery timelines for each shipping option, especially for the slowest one. Setting realistic expectations upfront can mitigate dissatisfaction even with longer delivery times. Consider providing more detailed tracking information.
* **Gather Specific Feedback on Delivery:** Implement post-purchase surveys that specifically ask customers about their satisfaction with the delivery process, broken down by the shipping method they chose. This qualitative data can provide valuable context for the quantitative insights from the chart.
* **Consider Local Logistics Partnerships:** Explore partnerships with reliable local courier services for your standard and faster delivery options to potentially further optimize delivery times and customer satisfaction within your primary service area.

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

**Approach**: To understand the year-over-year sales performance for the current month, a time intelligence analysis was conducted utilizing Data Analysis Expressions (DAX). Specifically, the SAMEPERIODLASTYEAR function was employed to identify the corresponding period from the previous year. The results of this calculation were then presented in a table, enabling a direct comparison of the **total sales for the current year's current month** against the **total sales for the previous year's corresponding month**. This approach facilitates a clear and immediate understanding of growth or decline in sales performance on a monthly basis compared to the prior year.







**Insights**:

* **Significant 2020 Growth with Nuances:** While 2020 shows a substantial overall sales increase compared to 2019, the monthly breakdown reveals this growth wasn't uniform. The strong early performance in the first half of 2020 significantly outweighs the slight dips observed in July, August, and September. This suggests that the drivers of growth were potentially more impactful in the initial months of the year, warranting investigation into what those factors were and why their impact might have lessened or reversed mid-year.
* **Potential Mid-Year Shift in 2020:** The decline in sales during July-September 2020, despite the overall yearly growth, could indicate a significant shift in market conditions, customer behavior, or internal factors during that period. This warrants a focused analysis on the events or changes that occurred specifically in those months to understand the cause of this temporary downturn within a growth year.

**Suggestion**:

* Analyze marketing, product launches, market conditions, sales strategies, and customer behavior that contributed to the significant early sales surge.
* Conduct a thorough analysis of external (market, economic, seasonal) and internal (operational, marketing, product) factors causing the July-September decline.
* Review strategic shifts, key events, and investment levels in years prior to 2020 to understand drivers of both growth and decline.
* Extract lessons from pre-2020 performance to build more stable and predictable future growth strategies.
* Schedule regular meetings to analyze sales performance and implement necessary actions.

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

Ans The Power BI Gateway is a secure bridge enabling Microsoft cloud services like Power BI, Power Automate, and Azure Analysis Services to access on-premises data sources. It allows organizations to connect their local data securely to the cloud for analysis and automation.

Some of its use cases are:

1. Automated Data Refresh: Keeps Power BI datasets current by automatically connecting to and updating from on-premises data.
2. Live Data Connectivity: Enables real-time querying of on-premises data for dynamic dashboards and immediate insights.
3. Hybrid Data Integration: Seamlessly combines on-premises and cloud data for comprehensive analysis.
4. Secure Data Access: Provides encrypted and authenticated communication between local systems and cloud services.
5. Centralized Enterprise Reporting: Facilitates unified reporting across large organizations with diverse data locations.
6. How would you approach this problem, if the objective and subjective questions weren't given?

Ans To approach this problem I will start from understanding the data like

* **Customer Table Attributes:** The customer table contains 3 attributes: customer ID, age, and gender.
* **Joining Tables:** Customer ages can be retrieved in the Orders table by joining it with the Customer table using the Customer ID.
* **Data Cleaning:** The "Reason" column with null values was removed from the Orders table.
* **Revenue Calculation:** Total revenue is calculated by summing the sales prices.
* **Unique Customers:** The number of unique customers making purchases annually averaged around 17,600 between 2015 and 2019, and then increased to 25,000 in 2020.
* **Unique Products:** The total number of unique products is determined by counting distinct products.
* **Average Delivery Time:** The average delivery time for delivered orders is 9 days.
* **Popular Items:** The most popular product category is Health and Beauty, the subcategory is Vitamins and Dietary Supplements, and the top product is Hermani Ultra Slim Tea.
* **Sales Increase:** Electronics, Phones, and Tablets showed a significant sales increase in 2019 and 2020.
* **Table Relationship:** The relationship between the Orders and Customers tables based on CustomerID is one-to-one.
* **Null Value Handling:** Null values in the "Reason" column were handled by removing the column.
* **Power Query M-Code:** The M-code to add a column in Power Query is typically: = Table.AddColumn(Source, "NewColumnName", each [Column1] + [Column2]). M-query is a case-sensitive language used for data transformation.
* **Top 5 Customers (SQL):** SQL was used to identify the top 5 most valuable customers based on a composite score of total revenue (50%), order frequency (30%), and average order value (20%).
* **Revenue Growth (SQL):** Month-over-month revenue growth rate was calculated using SQL.
* **Rolling Average Revenue (SQL):** A SQL query was used to calculate the rolling 3-month average revenue for each product category.
* **Revenue Breakdown:** Revenue was analyzed by year and product. Phones and tablets contribute the most to revenue, while Health and Beauty items had lower sales.
* **Product Returns:** A DAX function was used to calculate returned orders. 27% of orders were returned.
* **Impact of Ratings:** Analysis showed no direct correlation between product ratings and sales.
* **Revenue by Location:** Revenue distribution varies across different locations, with Greater Accra, Ashanti, and Western contributing the most to sales.
* **Monthly Sales Analysis:** January has the highest sales, while February and June show declines.
* **Underperforming Products:** Canon EOS 60D CMOS and Canon EOS 600D 18MP have high prices but underperform in sales.
* **Product Discounts:** Products with high prices and low sales may require discounts.
* **Customer Segmentation:** Customers were segmented by age and purchasing behavior to develop a loyalty program.
* **DAX Functions:** DAX functions were used to calculate total sales for the "Fashion" category and "Shipped from Abroad" delivery type, and to find total returned orders.
* **Average Wait Times:** Average wait times are similar across product categories.
* **Delivery Type vs. Waiting Time:** "Shipped from Abroad" has the longest delivery time but the highest customer volume.
* **Shipping Charges:** Shipping charges are consistent across product types.
* **Low Rating Orders:** Low ratings are strongly correlated with longer delivery times.
* **Year-Over-Year Sales:** 2020 showed overall sales growth, but with fluctuations throughout the year.