**Problem Statement**

You are working as a business analyst at Amazon, a company currently performing well. The stakeholders wish to maintain this level of performance and seek improvement. For this purpose, they want to devise new strategies. You are part of a team exploring new ways to benefit customers, such as offering more discounts and Prime membership perks. Could you suggest additional methods to identify and reward customers and enhance their shopping experience?

**Objective Questions:**

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

**ANS.: -** Total number of attributes in the customer table = **3** Attributes.

They are Customer ID, Customer Age, Gender

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

**ANS.: -** Here I have Added a new column in Excel File in Order Table named “Age” and used XLOOKUP formula

=XLOOKUP($D2,Customers!$A:$A,Customers!$B:$B)

Where, Lookup Value is specific Customer ID,

Lookup Array is Customer Id Column in Customers Table,

Return Array is Customers Age Column in Customers Table.

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

**ANS.: -** There were Missing cells in Reason Column in Order Table.

Have decided to fill those missing data for Delivered Status as “Delivered” option And Returned Status with “Not Mentioned” respectively.

While uploading the Dataset in Power Bi, there were 5 extra blank columns and 199 null rows were there, have decided to transform the data and remove the null rows from each column and the extra 5 columns also.

Removed those 5 extra columns by Selecting the respective columns and then selecting “Remove” option, and later, selected filtered the Reason Column by unselected “Null” option.

There is one null cell in Product category column also, replaced it with “Health and beauty “as this product category is the highest one among all.

There is one null cell in Customer Gender column also, replaced it with “M “as this gender is the highest one among both the genders.

Now the Data is Complete and Ready for further Analysis.

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

**ANS.: -** For this I have created a new column named Sales **=** Sales = Orders[Order Quantity] \* Orders[Unit Price]

We can calculate total revenue generated by all the sales by adding Measure.

Total revenue = Total Revenue = SUM(Orders[Sales])= **10,72,39,297.20**

**A number of numbers on a white background

Description automatically generated**

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.: -** Total number of unique customers who made purchases each year = Unique customer = DISTINCTCOUNT(Orders[CustomerID]) = **113000** Customers.

A close up of a number

Description automatically generated

Yes, there is an increase in the number over the years.

In 2016, 2017.

A graph of blue rectangular bars with numbers

Description automatically generatedThen there is a decline in 2018, and then again a Raise in 2019, 2020, and there is a huge decline in 2021.

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

**ANS.: -** We can calculate total number of unique products available in the company by adding Measure.

Unique Products = DISTINCTCOUNT(Orders[Product])= **44** Unique Products**.**

A black number on a white background

Description automatically generated

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

**ANS.: -** We can calculate average number of days it takes for products to be delivered specifically for Delivered Orders by adding Measure.

A number with black text

Description automatically generatedAverage Delivery Days = AVERAGEX(FILTER('Orders','Orders'[Status] = "Delivered"), DATEDIFF('Orders'[OrderDate], 'Orders'[Delivery Date],DAY))= **9.41** Days.

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

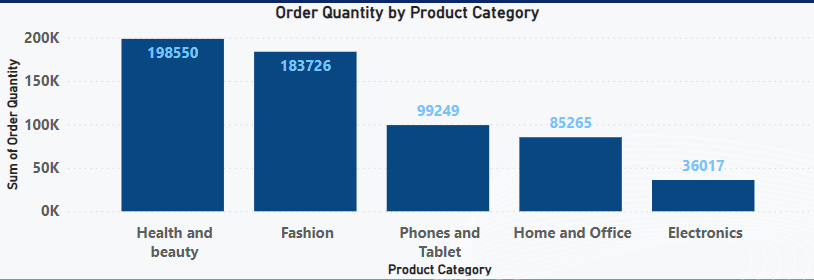
**ANS.: -** Here I have used Field parameter option with Order quantity.

The reason I have selected Order Quantity - it will show me which criteria has highest demand, which is being ordered more.

Most Popular **Product** is Avon Soft Musk Eau de Toilette Spray - 50ml- with total orders of 14444.

A blue and white bar chart

Description automatically generated

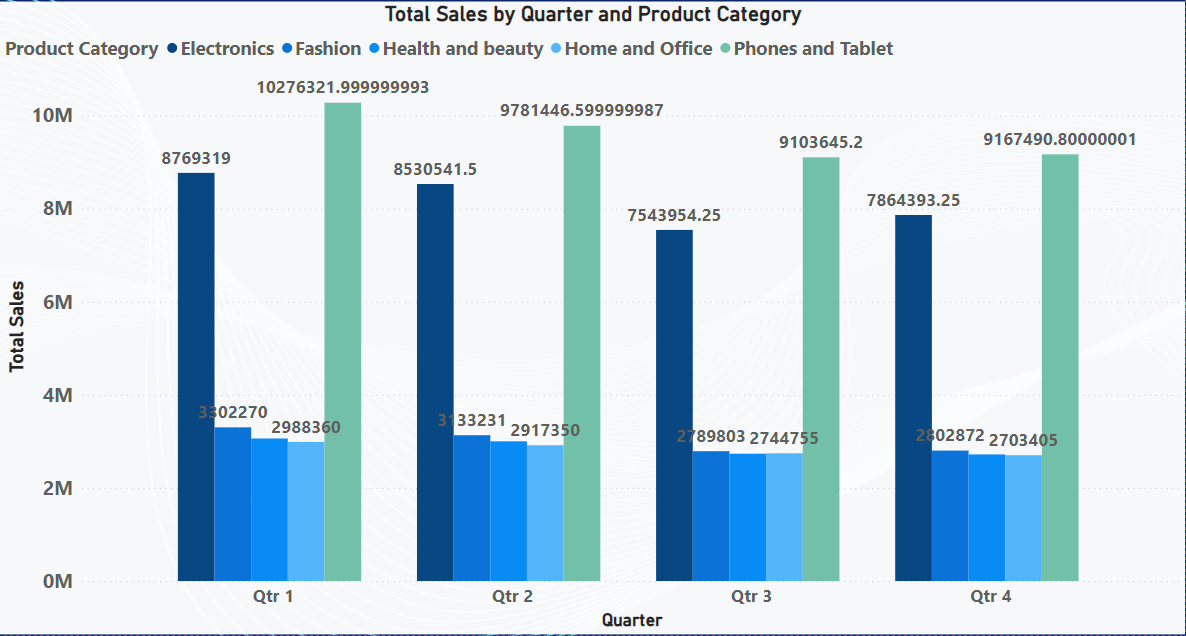
Most Popular **Product Categories** is Health And Beauty – with total orders of 198549.

**A bar graph with blue and purple numbers

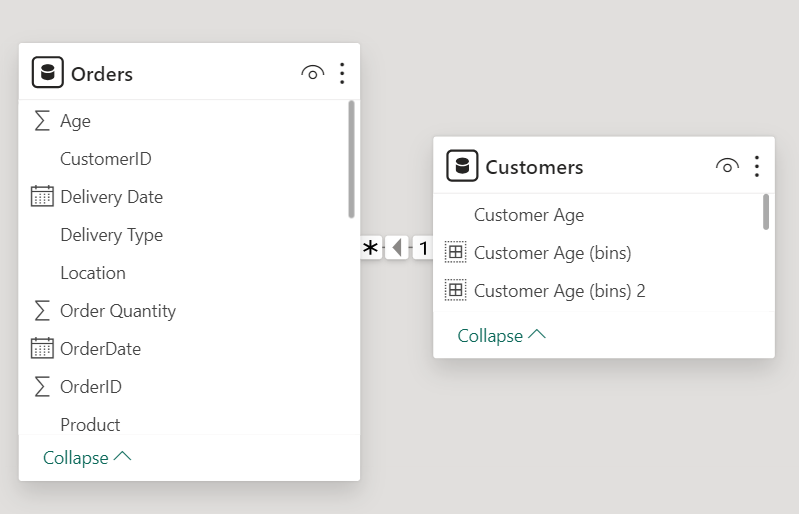
Description automatically generated**Most Popular **Sub Categories** is Vitamins & Dietary Supplements – with total orders of 70553.

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

**ANS.: -** All product have seen an increase or 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 customer tables?**

**ANS.: -** The type of relationship between the customer ID of Orders and customer tables is Many to One Relationship.

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

**ANS.: -** There were Missing cells in Reason Column in Order Table.

Have filled the missing data for Delivered Status as “Delivered” option And Returned Status with “Not Mentioned” respectively.

There is one null cell in Product category column also, replaced it with “Health and beauty “as this product category is the highest one among all.

There is one null cell in Customer Gender column also, replaced it with “M “as this gender is the highest one among both the genders.

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

**ANS.: -** Yes there data format issues in the data.

**I will handle them in the below steps :-**

* Data cleaning is the process of finding and fixing mistakes or inconsistencies in the data. Power BI provides number of data cleaning tools, such as duplicate removal, sorting, and filtering.
* Finding any problems with data format is the first step. Which could include text that should be numeric, missing values, inconsistent date formats, or incorrect data types.
* The "Transform Data" view allows to manually modify the data types for every column.
* There are several ways to deal with missing values: remove rows containing missing values, or fill them with default values. To deal with missing values, we can use DAX functions like IF, ISBLANK, or Power Query's "Replace Values" function.
* Power Query can be used to standardize date formats that are not consistent. We to apply date format transformations to date columns, use the "Transform" tab.
* The "Transform" tab in Power Query can be used to convert text to numbers if the numerical data is stored as text. Functions like Text.From, Text.To Number, and Text are also available in Power Query.

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 we add a column in Power Query, the M language code will show the formula which we applied is displayed in the formula bar. The M code generated in the formula bar reflects the steps we used to add the column.

**M - Query**

* Power Query uses a language called M-query, for formula in formula bar. It is a functional language which used for dealing with data, and in Power Query it shows the steps of actions.
* With the help of M-query, we can work on Power Query to perform variety of data operations, cleaning, transforming, and preparing the data for reporting or analysis.

**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.**

**ANS.: -** Here I have Created a Matrix where years are shown as columns, products are shown as rows, and revenue is displayed in the cells.

After Analysing the contribution of different products to annual revenue, I have sorted the matrix by total revenue to identify high-selling and low-selling products.

**A screenshot of a computer

Description automatically generated**Identifying low-selling items by looking at products with lower revenue contributions, I have used conditional formatting to highlight these items. I have used Background colour option with rules <=100000 for low-selling items.

**Suggestions to increase the sales of the low-selling items:-**

* 1. Conducting market research to understand why low-selling items are not performing well and identifying reasons like low demand, high competition, or product quality issues.
  2. Developing targeted marketing campaigns to promote low-selling items which could include discounts, special offers, or advertising which caters to specific customer segments.
  3. Making improvements to low-selling items based on customer feedback or market trends which could include product redesigning, changes in features, or pricing changes.
  4. Engaging with customers to gather feedback and understand their preferences and we can use this information for product offerings and improve the overall customer experience.

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.**

**ANS.: -** We can calculate Returned Products by adding Measure.

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

A number on a white background

Description automatically generatedTotal 30526 products were returned.

**Possible reasons for returns and consider how this metric could indicate improvements in product descriptions or quality control :\_**

* 1. As I observed the data that customers return products due to differences between product descriptions and actual features, if we make improvements which ensure correctness in product descriptions which includes providing more detailed and accurate information about product specifications, materials, and instructions to use it.
  2. As I observed the data that customers return result from mismatched customer expectations regarding product appearance, size, or functionality. After analysing return data which provide details for customer complaints or dissatisfaction, if we make improvements, which can be made to product designs, packaging, or marketing messages to better align with customer expectations.
  3. As I observed the data that customers return products due to order fulfilment errors such as shipping the wrong product or incorrect quantities. Improvements which we can make for order processing to minimize errors, such as implementing barcode scanning systems, improving inventory management systems.
  4. Requesting feedback from consumers who return products will provide detailed information about the causes of returns and important areas for development. This data can be collected through feedback systems like online reviews, customer service interactions, and surveys. Product improvements can be managed and areas for improvement can be prioritized by analyzing this feedback.
  5. Amazon can also adopt focused strategies to improve product descriptions, strengthen quality control procedures, and ultimately improve the overall customer experience by tracking and analyzing the quantity of returned products and determining the original causes for returns.

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?**

**ANS.: -** There is no relationship between the ratings and sales value.

**A graph with numbers and a line

Description automatically generated**Therefore, the rating of the products does not impact the sales value. Visualization is shown below –

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?**

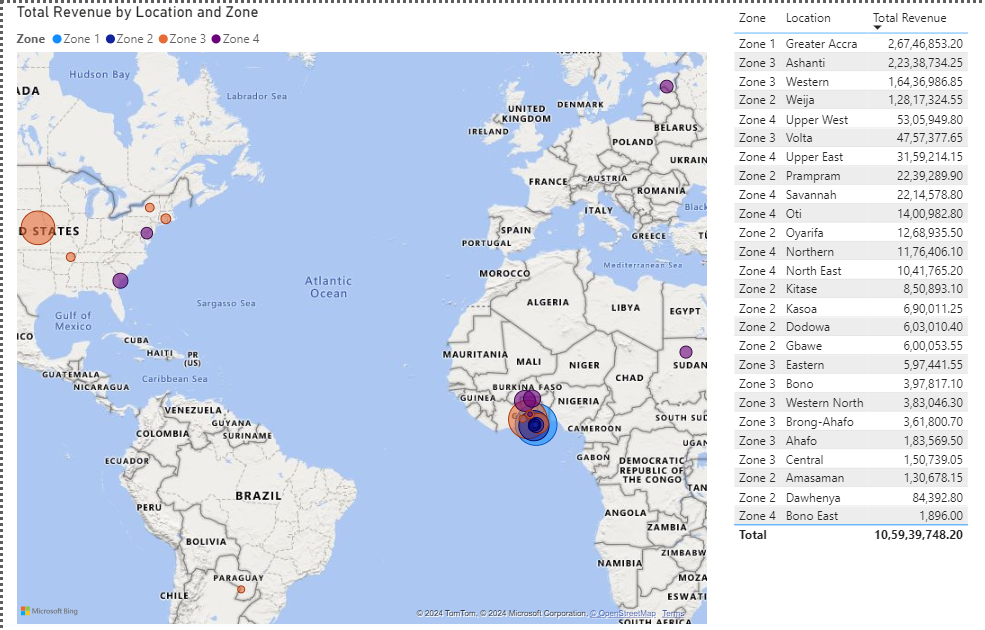
A table with numbers and a number of people

Description automatically generated with medium confidence**ANS.: -** Geographical areas contribute most to sales – Greater Accra with total revenue of 2,70,73,392.20

**Considering the strategic implications of location-based trends for regional marketing and distribution efforts :-**

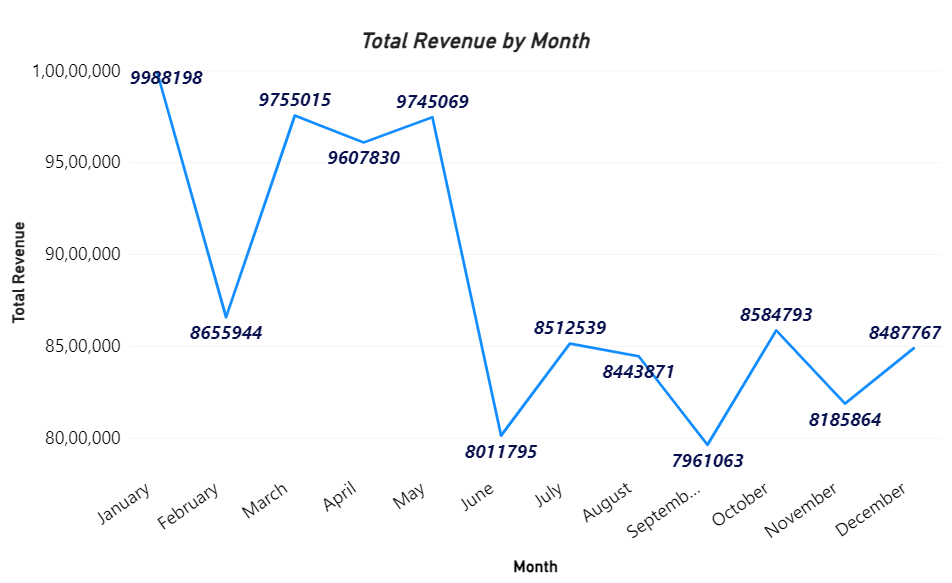
* 1. Creating specialized product offerings and marketing strategies to cater to the needs and preferences of consumers in various geographic areas.
  2. Utilizing the data from the revenue distribution analysis, pinpointing prospects for expansion of new markets or geographical regions.

**Company's market segmentation and resource allocation approach:-**

* + - * 1. Maintaining a close eye on market dynamics and revenue distribution to adjust resource allocation for shifting market conditions.
        2. Determining which markets or areas are performing well for focused marketing initiatives, then allocate resources appropriately.
        3. Considering the elements like the competitive landscape, purchasing patterns, and demographics of the population in each geographic area.
        4. Concentrating on advertising, marketing, and promotional efforts on areas with the greatest potential for revenue contribution.
        5. ****Streamlining supply chain processes and distribution networks to effectively meet demand in strategic regions.

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

**ANS.: -** As per my understanding for the trends in the sales on monthly bases – June, September and November are the months which could benefit from enhanced promotional offers to boost sales.



**Suggesting some targeted marketing strategies :-**

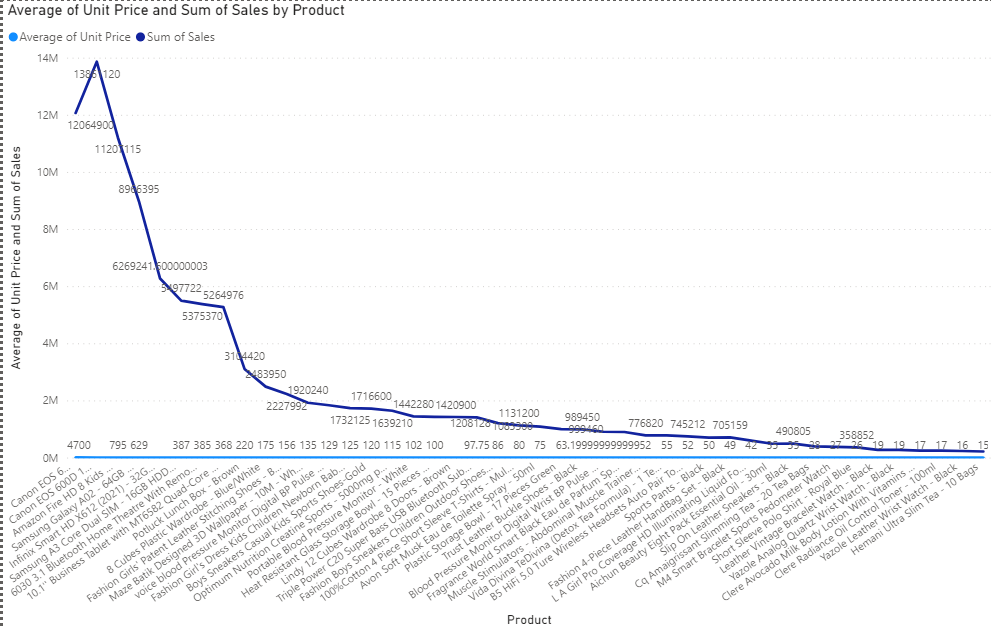
* 1. Implementing specialized social media advertising campaigns to connect with prospective customers by sending them customized marketing messages.
  2. Starting specialized email marketing campaigns to inform current customers and inform about exclusive deals and incentives.
  3. To reward consumers for purchases made during promotional periods and promote repeat business, to improve loyalty programs.
  4. Providing bundled packages, discounts, or promotions in line with holidays or other festival occasions that fall within the specified months (June, September & November).
  5. Using flash sales or limited-time promotions to create a sense of urgency and promote quick purchases.

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

**ANS.:** - As per my understanding about data – Product (Canon EOS 60D CMOS DSLR Camera Bundle – 18 – 55mm lens - Black) require increased marketing efforts, because the cost of product if high but the sales are less as compared to others.

**Items have high prices yet underperform in sales are :-**

A screenshot of a computer

Description automatically generatedHere I have used Conditional Formatting to highlight the items which have high price but not performing in sales. Top 8 Products (excluding the 2nd products) are listed below:-

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

**ANS.: -** 1.Considering the present inventory levels of every product. Discounts can be helpful in clearing out excess inventory and preventing inventory-outs for products with high inventory levels or slow turnover.

2. The Top 4 Products excluding the 2nd Product, should have discounts, because people always look for discounts which have higher cost.

A screenshot of a computer

Description automatically generated

**Targeted incentives drive sales and customer loyalty for specific products:-**

* 1. For encouraging purchases, offering bundle discounts, buy-one-get-one (BOGO) offers, or percentage discounts on a limited number of products.
  2. Offering flash sales or limited-time promotions on particular products to generate urgency and encourage quick purchases.
  3. Rewarding to consumers and members of loyalty programs with special discounts or early access to new products.
  4. Incorporating reward programs or point systems where customers accumulate points for their purchases, which they can later use to receive free or discounted 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.**

**ANS.: - Loyalty program to benefit the company’s customers:-**

* 1. The goals of loyalty program is to boost sales, retaining customers, or obtaining customer information for targeted advertising.
  2. Creating several membership tiers according to the amount of money or level of interaction from customers. Bronze, Silver, Gold, and Platinum are typical tiers, with benefits growing as users advance through the ranks.
  3. Establish the requirements that customers must meet in order to be eligible for each tier, such as annual spending, frequency of purchases, or duration of membership.

**Strategies to bucket Customers and provide benefits under different loyalty programs:-**

* + - * 1. Making sure that existing customers are aware of the advantages and rewards that come with each tier. Utilizing social media, email marketing, and in-store signs to advertise and motivate people to sign up for the loyalty program.
        2. Giving early access to sales events, special promotions, or discounts based on tier.
        3. Giving points for each purchase Customer make, which can be exchanged for deals, free goodies, or other incentives.
        4. Offering individual assistance, committed customer service, or first access to customer support channels.
        5. Customers can be invited as guest to members-only sales, product launches, or exclusive events.
        6. Providing customers with free or subsidized shipping, particularly for higher tiers.
        7. Offering discounts, presents, or bonus points on Customers birthday to honor customers' special days.

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.: -** Total sales for Fashion & shipped from abroad = CALCULATE(SUM(Orders[Sale Price]), 'Orders'[Product Category] = "Fashion", 'Orders'[Delivery Type] = "Shipped from Abroad") = **41,35,434**.

A black text on a white background

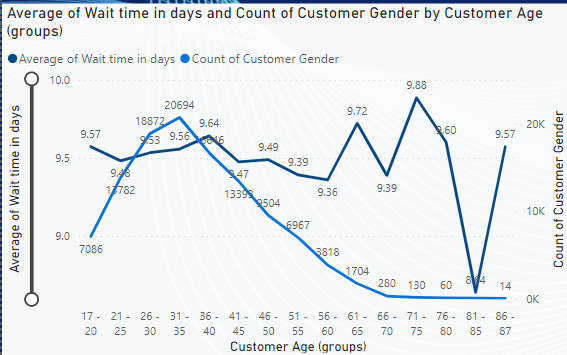
Description automatically generated

**Other types of DAX functions i have used in the project:-**

* + - 1. SUM
      2. DISTINCTCOUNT
      3. AVERAGEX
      4. FILTER
      5. DATEDIFF

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

**ANS.: -** I have created a new column named **=** Wait time in days = DATEDIFF(Orders[OrderDate],Orders[Delivery Date],DAY).

A graph with a line

Description automatically generatedWait Times Correlated with Demographics (Age) & (Gender) -

A graph showing the number of people

Description automatically generated

**Strategies to optimize scheduling and staffing to reduce wait times:-**

* 1. For optimizing delivery scheduling, using of logistical software and sophisticated routing algorithms will help. Orders are grouped according to location in order to reduce travel time and guarantee effective delivery scheduling.
  2. Placing flexible staffing models,which can demand using part-time workers to accommodate varying delivery demands or recruiting more delivery drivers during busy times.
  3. Working with outside logistics companies to increase delivery reach and capacity. Amazon can expand its delivery areas and shorten customer delivery times by utilizing a network of delivery partners.
  4. Faster delivery times can be guaranteed by Amazon by prioritizing the orders and providing expedited shipping options at an extra cost for orders that must be completed quickly or for premium delivery services, using priority scheduling.
  5. To find problems and areas for improvement, tracking and evaluating delivery performance metrics on a constant basis. Placing feedback mechanisms in place to get feedback from customers and deal with delivery speed and reliability issues.
  6. To accommodate higher order volumes and quicker delivery times, making investments in modernizing transportation infrastructure and growing the capacity of inventory centres. This could entail investing in transportation technology like drones or constructing new inventory centres in key locations.

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

**ANS.: -** Yes, there is a positive relationship between the Delivery type and waiting time.

A graph showing a line

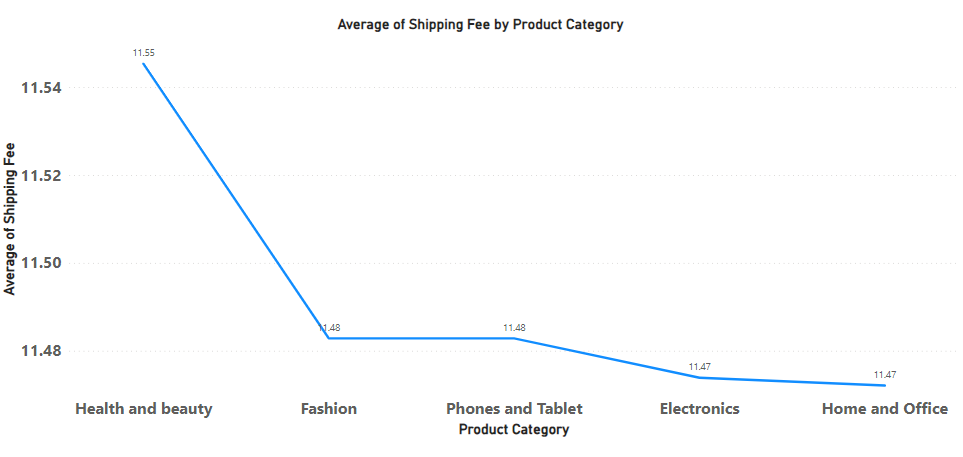
Description automatically generated with medium confidenceWaiting time is high for “shipping from abroad” as it must be in usual scenarios, and less for “Express” which makes sense.

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

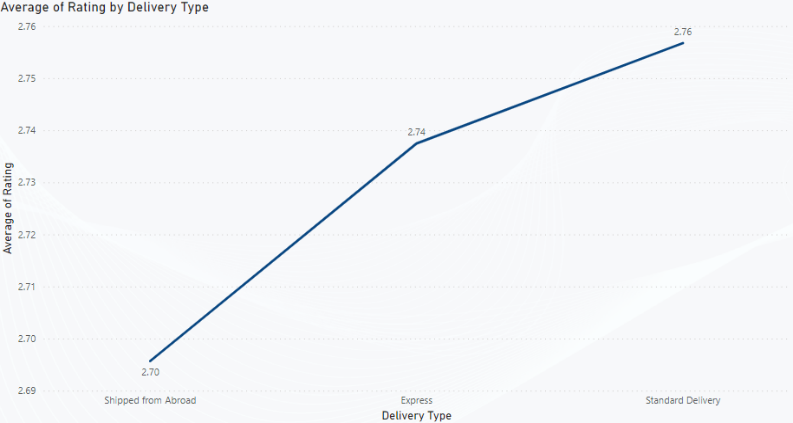
**ANS.: -** Yes, there is a relationship between shipping charges and product type because of one hidden factor named “Maintenance” during the delivery time.

If we observe the cart carefully, “Health and beauty” product category is having high shipping costs as compared to all other product categories.

As “Health and beauty” product category required some sort of “Maintenance” factor like storage, packaging, handling the order, etc., in the whole delivery process.

****

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

**ANS.: - Average rating with Delivery type**

**A graph showing a line

Description automatically generated with medium confidenceAverage rating with Unit Price**

**Average rating with A graph with a line

Description automatically generatedWait time in days**

**Strategies to decrease the low rating orders after analyzing different factors like waiting time, shipping type, unit price, etc. :-**

* 1. Providing customers who need their orders delivered faster with options for expedited shipping.
  2. Customers should be aware of shipping options and related delivery times when they are checking out.
  3. To encourage quicker delivery, providing rewards or discounts for selected premium shipping options.
  4. To reduce customer dissatisfaction, monitoring and quickly addressing issues related to product quality.
  5. Giving customers systematic product reviews and specifications so they can make well-informed purchases.
  6. Utilize tools like chatbots to help customers right away.
  7. Reach out to unhappy customers to resolve their issues and avoid bad reviews or ratings.

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.**

**ANS.: - Here I have created a new measure named Previous year same month total sales =**

Previous year same month total sales = CALCULATE([Total Sales],SAMEPERIODLASTYEAR(Orders[OrderDate]),ALLEXCEPT(Orders,Orders[OrderDate]))

Where - CALCULATE([Total Sales], ...) is calculating the total sales,

SAMEPERIODLASTYEAR(Orders[OrderDate]) is finding the same period last year to the OrderDate,

ALLEXCEPT(Orders, Orders[OrderDate]) is removeing all filters from the Orders table except for the OrderDate column.

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

**ANS.: -** In order to securely connecting and integrating Power BI cloud services with on-premises data infrastructure and providing a variety of data analysis, reporting, and visualization capabilities, Power BI Gateway is essential.

**Its use cases:-**

* 1. By applying Power BI Gateway, organizations can connect to their on-premises databases and data sources to create enterprise-level reports and dashboards.
  2. By enabling direct queries and live connections to on-premises data sources, it supports real-time analytics by giving users access to the most recent data for analysis.
  3. With support for several data sources and formats, Power BI Gateway allows smooth data transfer between on-premises and cloud-based systems, enabling data integration.
  4. Using secure connections to on-premises data sources and access manages for sensitive data, it assists organizations in maintaining compliance and governance standards.

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

**ANS.: -** By the provided dataset, stakeholders wanted to seek improvements. For this they want us to come up with new strategies.

So most of my points of observation for analysis are covered under the provided objective and subjective questions. Like -

* 1. Which Customer is loyal (repetitive orders),
  2. Time gap between ordering the delivery, strategies to reduce time,
  3. Which demographic has more purchases, taking action /strategies for not performing demography,
  4. Customers feedback (ratings), strategies to improve ratings,
  5. Strategies to improve sales,
  6. Strategies for being competitive in market by keeping competitive prices for products,
  7. Providing discounts/offers, etc.