**** analysis

by Yash Kumar

**Objective Q&A:**

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

Every Inconsistent value and missing value has been addressed during data cleaning for analysis.

* Removed all the null values and values with error from each column except ‘Reason’ column.
* Replaced empty values in the Reason column with ‘None’.
* Performed Split to Columns to Reason column into ReasonType and Return Reason attributes.
* Corrected and replaced the ‘Delivey’ to ‘Delivery’ in the ReasonType column.
* Formatted ReasonType column value to UPPERCASE for better readability.
* OrderDate and DeliveryDate were in inconsistent format of numbers.  
  Thus, changed the type of both attributes from Number to Date format.   
  (Dates are in short date format).
* Created Product Category Hierarchy, with Product Category, Subcategory and Products in respective orders.  
  *This will help in the Drill down function for further analysis.*
* Created new column Days to Reach, using NETWORKDAYS function between OrderDate & DeliverDate to get the number of days it took to reach for that particular order.

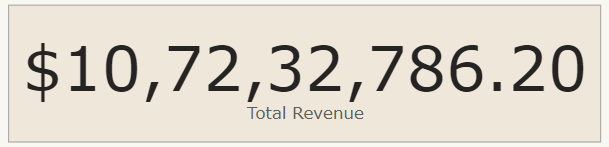
**Days to Reach = NETWORKDAYS(Data[OrderDate],Data[DeliveryDate])**

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

To calculate total revenue generated by all the sales, we will

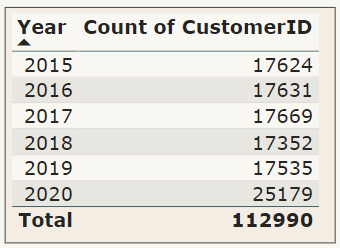
* Create new measure
* In the formula bar, use the aggregate function SUM on the value of Sales.
* Following, use scorecard visualisation to show Total Revenue.

**Total Revenue till now:** $**10,72,32,786.20**



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

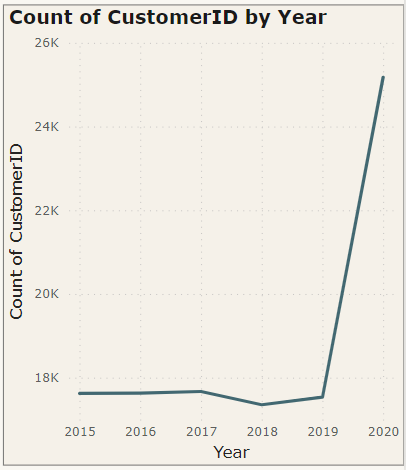
Total number of unique customer per year are as follows:



There haven’t been much significant changes in the number of unique customers over the 2015-2019 period.

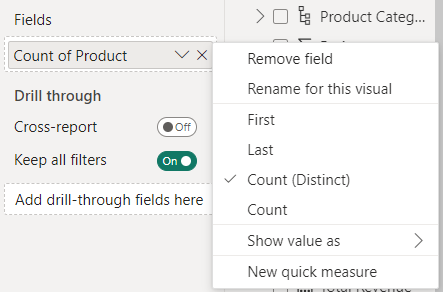
But has a significant increase (of more than 8K) in the year 2019-2020.

***A major reason could be due to Covid and lockdowns around the world in FY 2020*.**



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

* Create scorecard visual
* Place Products in Fields section of visual
* Right click and select Count(Distinct).



**Total distinct product as of now:** **44**

****

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

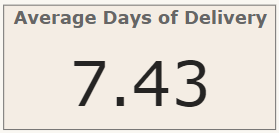
* First create a new measure, named ‘Average Delivery Days’.
* We used formulae

**AverageDays for Delivery = CALCULATE(AVERAGE(Data[Days to Reach]),FILTER(Data,Data[Status]="DELIVERED"))**

*Here we use the* Average *function on* Days to Reach *column value with a* Filter *for* DELIVERED *orders using the* Calculate *function.*

* This will calculate average days for delivery for specifically delivered orders.

**Average number of days for delivery: 7.43**

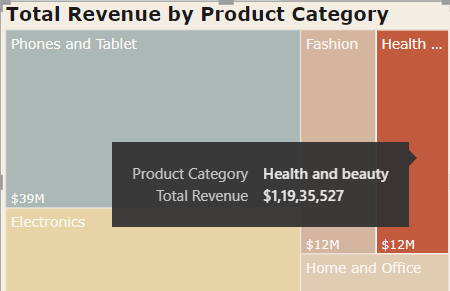
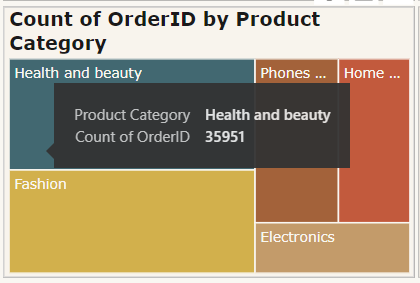
****

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

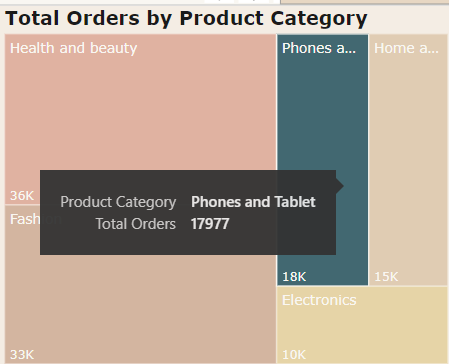
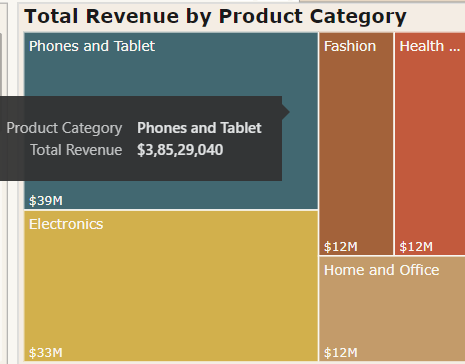
**Note:** *‘****% of Total’ is total with respect to one hierarchical order above***

* **Product Category**

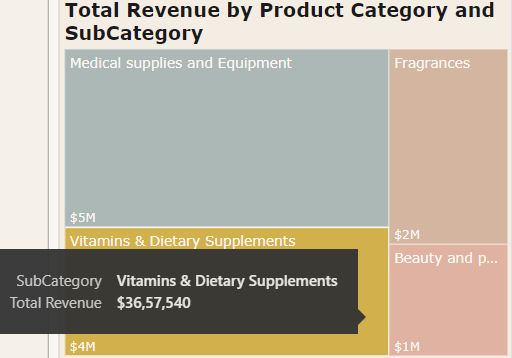
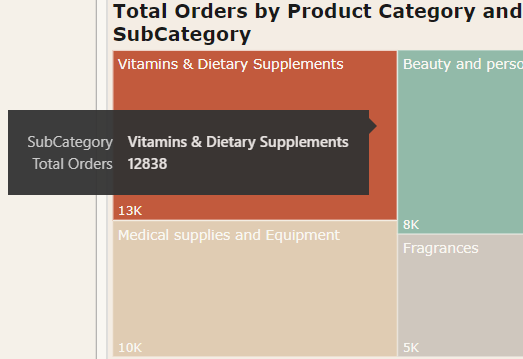
In terms of Orders**:**  
 **Most Popular:** Health & Beauty **Revenue Generated:** $11.9M **(11.1% of Total)  
Count of Orders:** 35.9K **(31.8% of Total)**

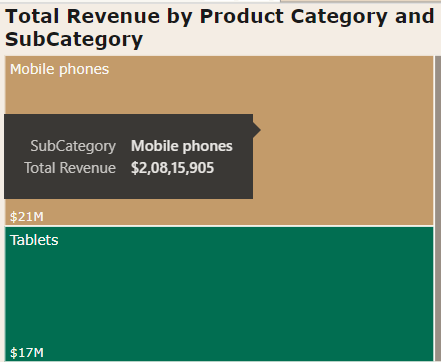
****

In terms of Revenue**:  
  
Most Popular:** Phones & Tablets **Revenue Generated:** $38.5M **(35.9%)  
Count of Orders:** 17.98K **(15.9%)**

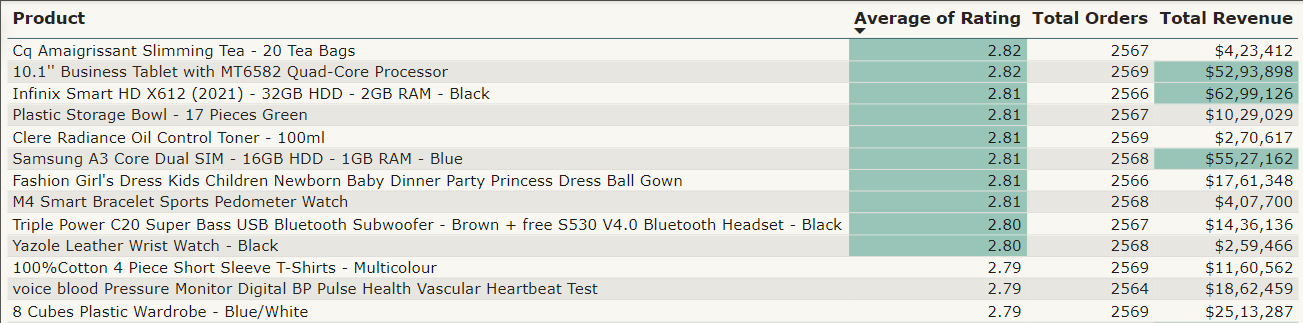
****

* **Product Sub-Category**In terms of Orders**:**  
   **Most Popular:** Vitamins & Dietary Supplements **Revenue Generated:** $3.6M **(30.6%)  
  Count of Orders:** 12.8K **(35.7%)**

****In terms of Revenue**:**  
  
**Most Popular:** Mobile Phones **Revenue Generated:** $20.8M **(54%)  
Count of Orders:** 7.7K **(42.8%)**

****

* **Products**In terms of **Orders:** Mostly all products have the same number of orders. So it’ll be futile to compare from that criteria.  
  In terms of **Ratings** (**> 2.8**)**:**

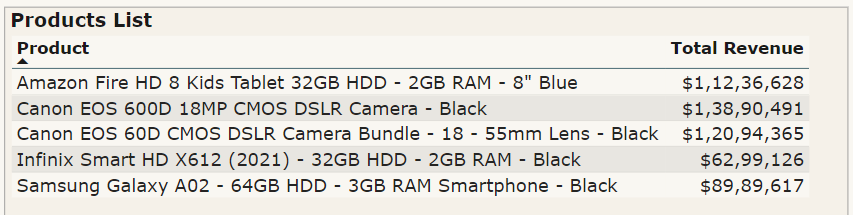
****In terms of **Revenue** (**> $5M**)**:**

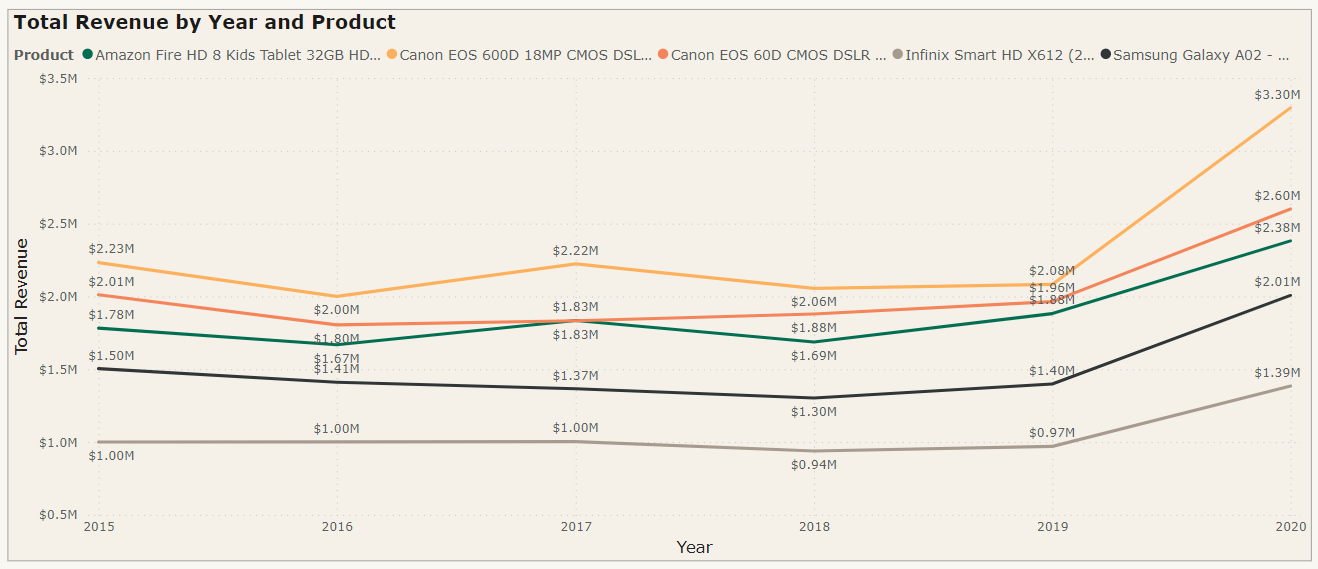


1. Which products have seen an increase or decrease in sales over the year?

Here I chose to show **Top 5** products in terms of revenue with their increase and decrease over the years.

**This has been done using the Top N filtering option in Filter Pane.**





**Subjective Q&A:**

1. How does revenue break down 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.

Here’s Revenue Breakdown by Year and Product**:**

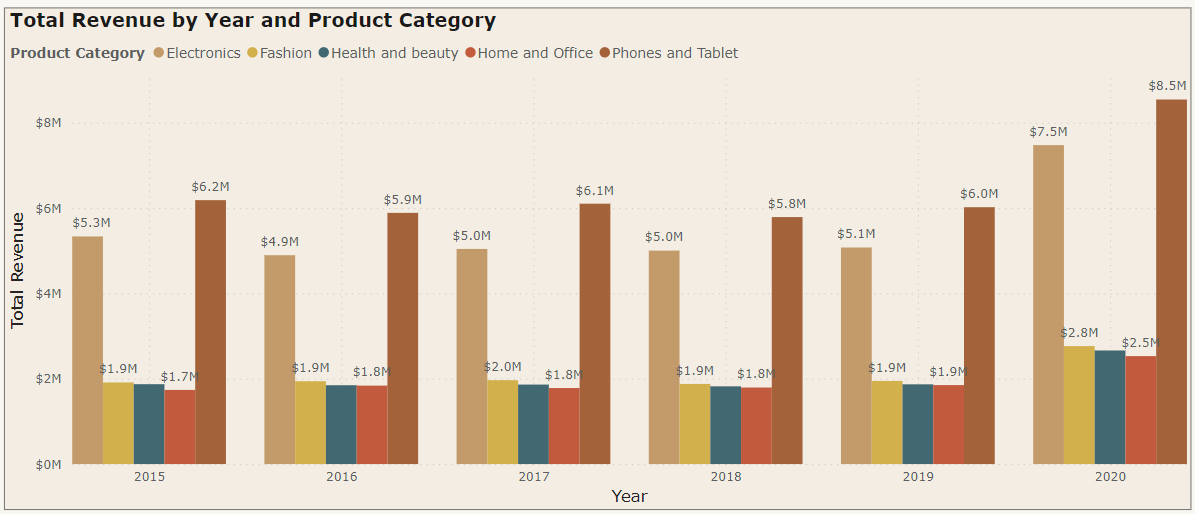
**Yearly Breakdown:**

| **YEAR** | **Roundup REVENUE ($)** |
| --- | --- |
| 2015 | 1,70,43,100 |
| 2016 | 1,64,09,800 |
| 2017 | 1,67,55,000 |
| 2018 | 1,62,85,200 |
| 2019 | 1,67,63,300 |
| 2020 | 2,39,76,600 |

**Year & Product-wise Breakdown:**

Since we have three attributes to show, visualising and presenting it with a clustered column chart would be ideal.

We can drill down specific product categories to get sub categories and individual product info, in power BI.

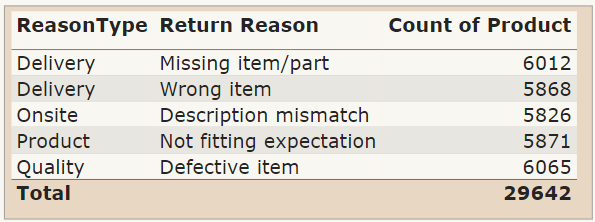


1. How many products were returned? Examine the possible reasons for returns and consider how this metric could inform improvements in product descriptions or quality control.

Total products that were returned**:** **29642**

****

There are currently 5 reasons for returns (with count of products) shown in the chart below.



**Practical Solutions for Return Reasons:**

| **Reason Type** | **Return Reason** | **Suggestions** |
| --- | --- | --- |
| Delivery | Missing Item/Part | * Enhance order fulfilment processes to ensure accurate picking and packing. * Implement barcode scanning or RFID technology for better inventory tracking. |
| Delivery | Wrong Item | * Conduct regular training for warehouse staff on order verification procedures. * Implement a double-check system before shipping to reduce the chances of sending the wrong items. |
| Onsite | Description Mismatched | * Regularly update and review product descriptions to ensure accuracy. * Encourage customers to provide detailed reviews to address potential discrepancies. |
| Product | Not Fitting Expectation | * Improve product sizing information on the website. * Implement a user-friendly size guide and include customer reviews on sizing experiences. |
| Quality | Defective Item | * Strengthen quality control measures, including rigorous inspections. * Consider partnering with suppliers to improve the overall quality of received goods. |

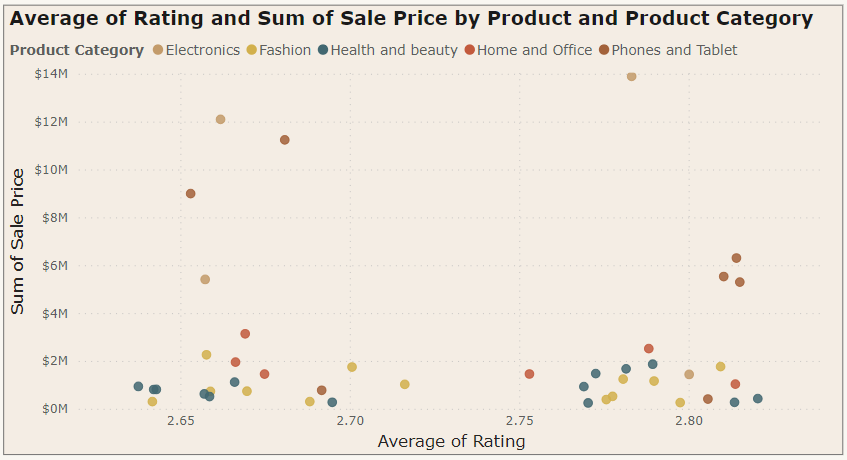
* *These suggestions aim to address each specific reason for returns.*
* *Implementing these improvements can contribute to a reduction in return rates and enhance overall customer satisfaction.*

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

**No, it’s not necessary for the product to be highly rated only to have a big number of sales. As we can see in the scatter plot below, it’s not for sure if the product is highly rated so it must have high sales.**

To analyse the impact of product ratings on sales value, I created the following visualisations using Power BI:

* **Scatter Plot: Ratings vs Sales Price**

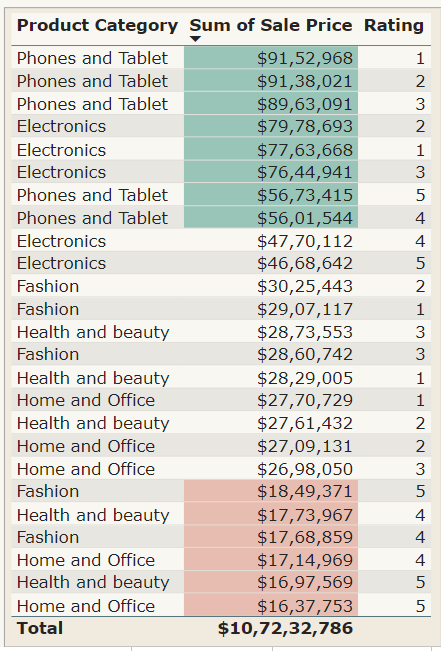
****

In this scatter plot, each data point represents a product, with the:

* x-axis representing the product ratings.
* y-axis representing the corresponding sum of sales price.

*The plot allows us to observe any correlation between higher ratings and increased sales.*

* **Table: Ratings and Sales Value Statistics**

****

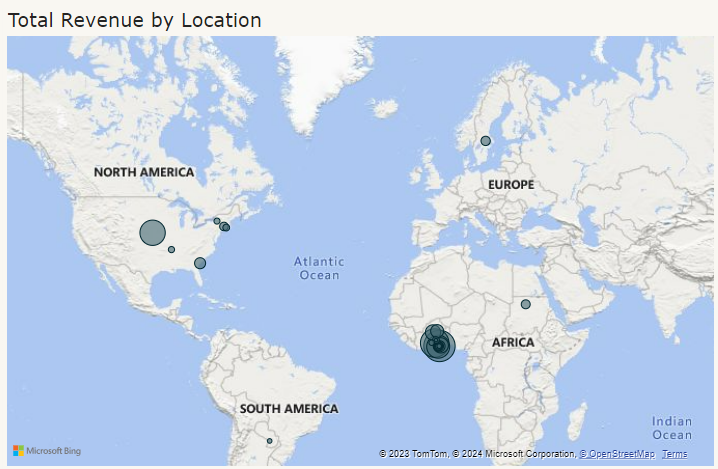
We can see from this table that some of the highly sold products don’t have a high rating, and rather some of the low sold products were rated 4 or 5.

**Thus we CAN’T make a strong correlation between rating and sales value.**

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 approach to market segmentation and resource allocation?

Using Map visualisation we can see, Top 3 countries in revenue generating are: **Ghana**, **United States** and **Sweden**.

| **Countries** |
| --- |
| Ghana |
| United States |
| Sweden |



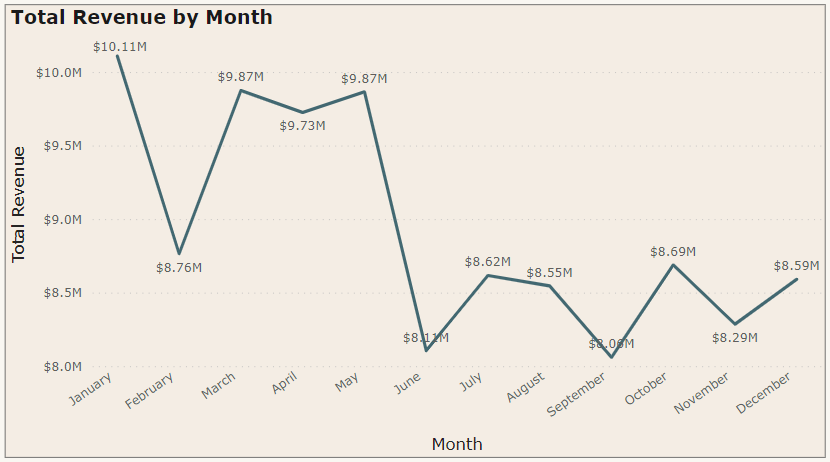
1. Determine which month could benefit from enhanced promotional offers to boost sales. Can you suggest some targeted marketing strategies here?

According to data, we have the lowest revenue generated after May end or June start.

Three lowest performing months are June, September and November.

Some strategies to target these months**:**

* **June: Summer Sensations**
* Feature products associated with summer activities, travel, and outdoor adventures.
* Introduce exclusive discounts on items like beach accessories, outdoor gear, and summer fashion.
* Encourage users to share their summer plans for a chance to win special vacation-themed goodies.
* **September: Back-to-School Extravaganza**
* Showcase a range of products catering to back-to-school needs.
* Offer special deals on stationery, tech accessories, and study essentials.
* Implement a "Smart Shopper Challenge" where customers can share their creative study setups for a chance to win a mystery box.
* **November: Festive Finds Fiesta**
* Highlight festive and gift-worthy products ahead of the holiday season.
* Create mystery boxes with a mix of holiday-themed items and exclusive discounts on festive decor.
* Encourage customers to share their favourite holiday traditions or decor setups for a chance to win a grand festive prize. With the unique characteristics and expectations of each month.



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

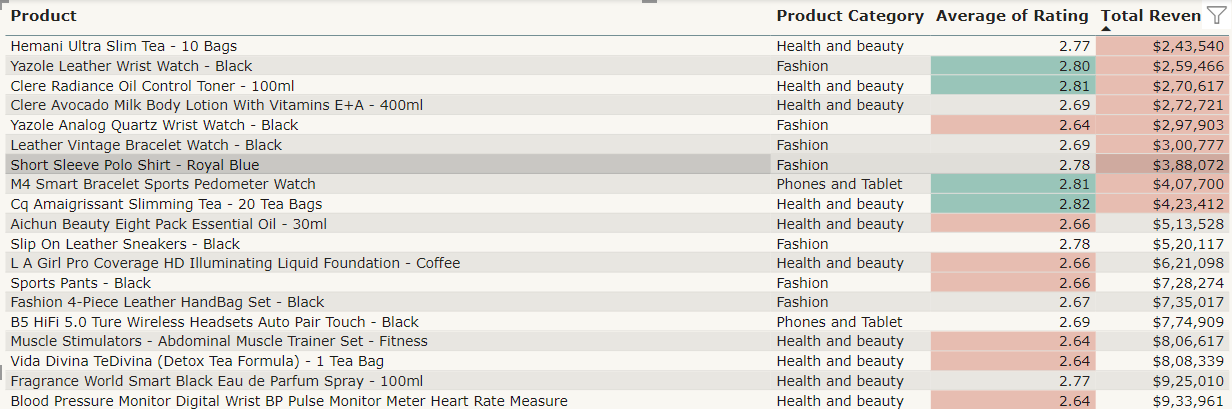
As we can see from the table below, there are NOT any products who are terribly performing compared to their unit price.

We can see a correlation that products with high unit price also have high sales.



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

* For this we will analyse for each product category by year and revenue
* Products revenue with less than $1M are listed below which also had declining or stagnant revenue over the years.
* Products highlighted **red** must be focused majorly.

****

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.

* **Customer Segment Strategy:**

| **Customer Segment** | **Strategy** | **Benefits** |
| --- | --- | --- |
| Frequent Shoppers | Exclusive early access, members-only discounts | Reward loyalty with exclusive perks. |
| High Value Customers | Premium service, personalised assistance | Provide personalised service for high-spending customers. |
| Brand Advocates | Referral bonuses, exclusive previews | Incentivize referrals and offer exclusive product previews. |
| Occasional Shoppers | Limited-time discounts, surprise gifts | Encourage occasional shoppers with special offers. |

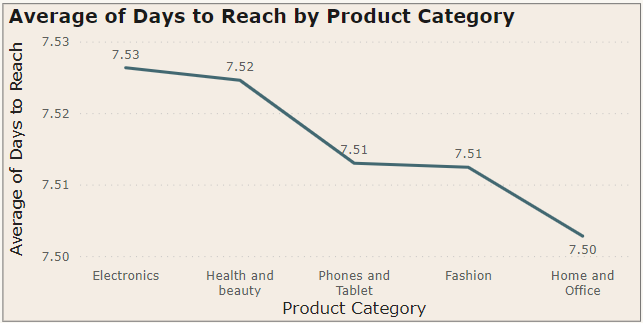
* **Tiered Loyalty Program:**

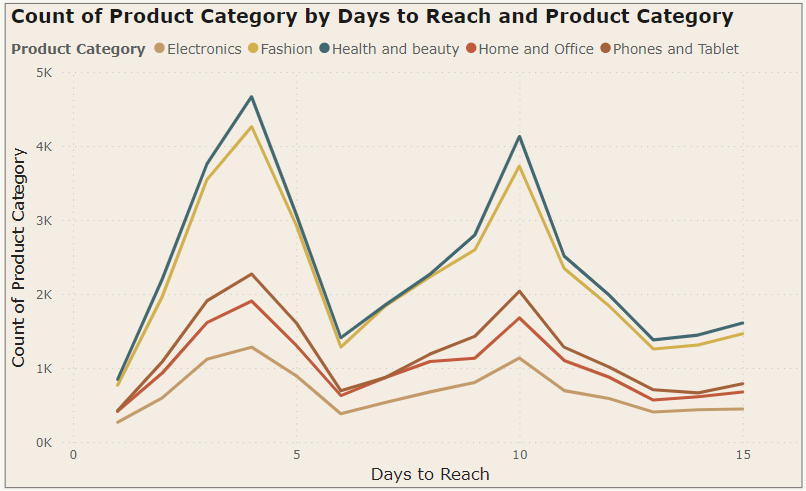
| **Loyalty Level** | **Strategy** | **Benefits** |
| --- | --- | --- |
| Basic Level | Discounts, early access to sales | Attract new customers and create a start. |
| Premium Level | Increased discounts, occasional freebies | Encourage consistent purchases and engagement. |
| VIP Level | VIP access, personalised recommendations | Recognize and reward the most loyal customers. |

* **Communication & Evaluation:**

Clearly communicate the loyalty program and evaluate engagement regularly for continuous improvement.

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

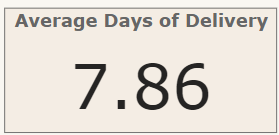
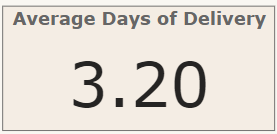




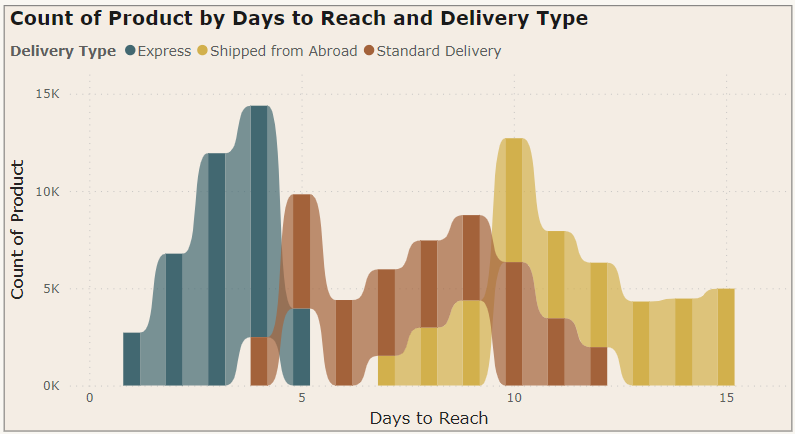
1. Explore if there is any relationship between the Delivery type and waiting time between ordering and receiving an item.

To check average waiting time (days to reach) for each delivery type:

* Open PowerBI File
* Open Main Tab
* Click on Slicer of Delivery Type
* From its dropdown menu, select type of delivery
* Waiting time will be shown in Average Days of Delivery scorecard



| **Delivery Type** | **Average Waiting Time (in days)** |
| --- | --- |
| Express Delivery | 3.2 |
| Standard Delivery | 11.4 |
| Shipped from Abroad | 7.8 |

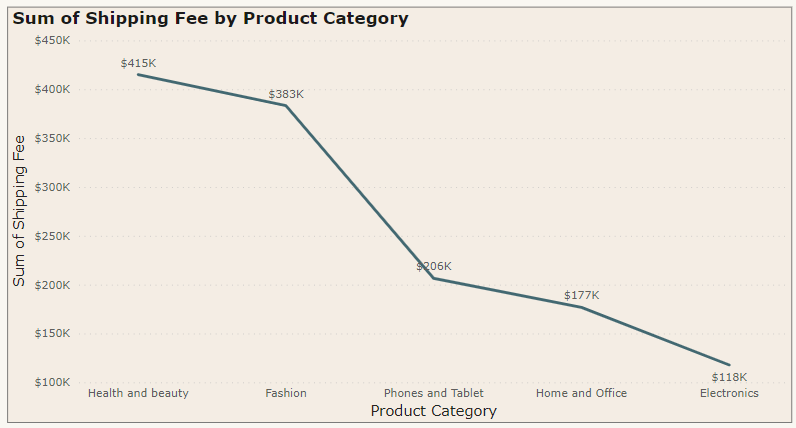


1. Is there any relationship between shipping charges and product type?

We can see the total sum of shipping fees for each product category.

‘Sum’ of shipping fee is used rather than ‘Average’ because mostly all products have nearly the same shipping fee. So it would be futile to use comparison with average aggregation.

We can determine that**: the category with higher sum of shipping fees has higher order quantity.**



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

First to analyse products through rating, created field parameters with:

* Ratings
* Delivery Type
* Days to Reach
* Unit Price.

We can access it in power BI and analyse the products accordingly and provide strategy.

