

# Sales Analysis

## Introduction

Sales analysis is a crucial procedure that enables companies to assess how well their goods and services are performing, pinpoint areas for improvement, and make data-driven decisions (Efti, 2023). To stay competitive in today's market, it is more crucial than ever for firms to have a comprehensive grasp of their sales data. Examining different data, including revenue, profit margins, sales volume, and customer demographics, is often part of a sales analysis. Businesses can learn more about the factors influencing their sales success and create plans to improve their operations by analysing these indicators (Efti, 2023). In this situation, a thorough sales study of five distinct product lines can offer insightful information on the advantages and disadvantages of each line as well as trends and patterns that may impact the overall sales performance of the business.

Dataset:

	Variable	Variable Type
1.	order_id: Unique identifier for each order	Unique
2.	order_date: Date the order was placed	Date
3.	status: Current status of the order (e.g. processing, shipped, delivered)	Categorical
4.	item_id: Unique identifier for each item in the order	Unique
5.	qty_ordered: Quantity of the product ordered	Numeric
6.	price: Price of the product	Numeric
7.	value: Total value of the order (price x quantity)	Numeric
8.	discount_amount: Amount of discount applied to the order	Numeric
9.	sales: Total amount of the order after discount	Numeric
10.	category: Category of the product being ordered	Categorical
11.	payment_method: Method of payment for the order	Categorical
12.	bi_st: Billing state for the order	Categorical
13.	cust_id: Unique identifier for each customer	Unique
14.	year: Year of the order	Numeric
15.	month: Month of the order	Date
16.	ref_num: Reference number for the order	Numeric
17.	Name Prefix: Prefix of the customer's name (e.g. Mr., Mrs., Dr.)	Categorical
18.	First Name: First name of the customer	String
19.	Middle Initial: Middle initial of the customer's name	String
20.	Last Name: Last name of the customer	String
21.	Gender: Gender of the customer	Categorical
22.	age: Age of the customer	Numeric
23.	full_name: Full name of the customer	String
24.	E Mail: Email address of the customer	String
25.	Customer Since: Date when the customer first became a customer	Date
26.	SSN: Social Security number of the customer	Numeric
27.	Phone No.: Phone number of the customer	Numeric
28.	Place Name: Name of the location where the customer lives	String
29.	County: County where the customer lives	Category
30.	State: State where the customer lives	Category
31.	Zip: Zip code where the customer lives	Numeric
32.	Region: Region where the customer lives	Category
33.	Username: Username for the customer's account	String
34.	Discount Percent: Percentage of discount applied to the order.	Numeric

Dataset can be found over [here](#).

## Persona and Research Question:

Persona: The user is a “Sales Analyst” who plays a crucial role in a business by analysing and interpreting sales data. They develop dashboards and reports that offer insights into the company's sales performance using business intelligence tools. Sales analysts can spot patterns and trends in sales data because they are adept at data analysis, visualisation, and statistical modelling. Their advice enables companies to optimise their sales performance through data-driven decisions.

Research Questions:

1. How does the sales performance of a particular product category vary across various geographic regions and states, as well as throughout different months of the year and is there any similarities in sales across different states of that product category?

Requirement: In order to provide a comprehensive answer to this inquiry, we must utilize two specific visual aids: a tree map and a map. We may analyse the sales monthly using the tree map, in which each rectangle represents a month, and its size represents the sales revenue for that month and use different shades of colour to represent different regions. On the other hand, using the map will enable us to look at sales on state level and spot any geographic patterns or discrepancies in performance and apply clustering in it to see similarities. Together, these two charts help us better comprehend the sales data and can help us respond to the original query in a more comprehensive way.

2. What is a quantity of product orders across different age groups and gender, and how does the order quantities changes for product category by gender?

Requirement: To address the issue at hand, we require a visual representation in the form of a bar plot. The bar plot will display the total quantity of items ordered on Y – axis by individuals in different age groups on x - axis, with age bins and gender being taken into account. Additionally, a stacked bar chart is necessary to examine how the number of items ordered varies across different product categories by gender. We will plot product categories on y axis and numbers of orders on x axis and most importantly, we will assign colours to bars to differentiate genders. Both bar plots will help to gain insight into the trends and patterns within the data and allow for better analysis of the information.

## Design:

There are 2 stages of designs, first is prototype view and second is final tableau dashboard.

Prototype (Paper Landscape):

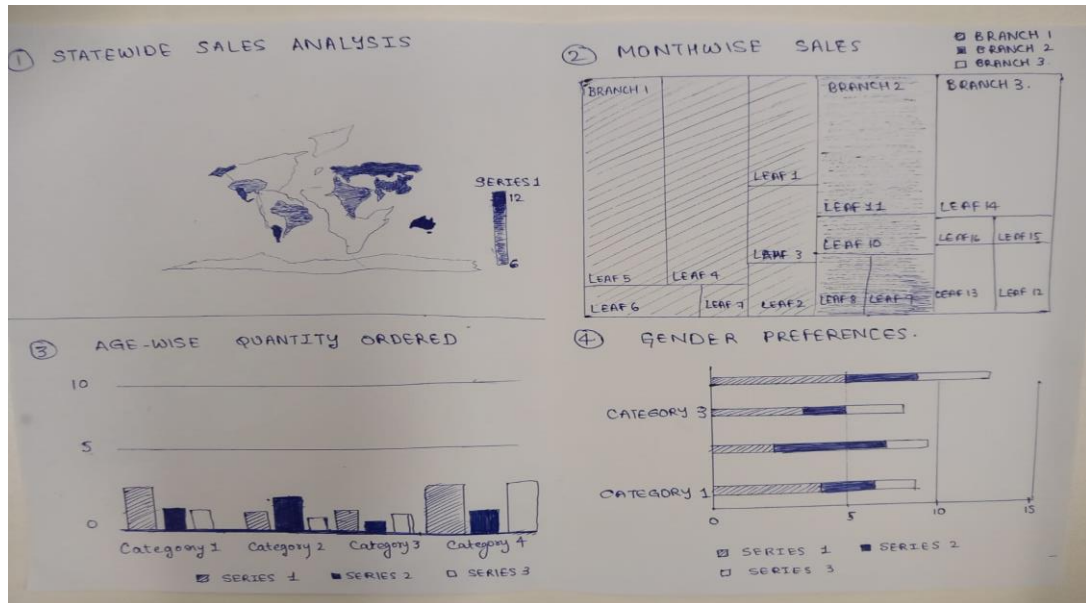
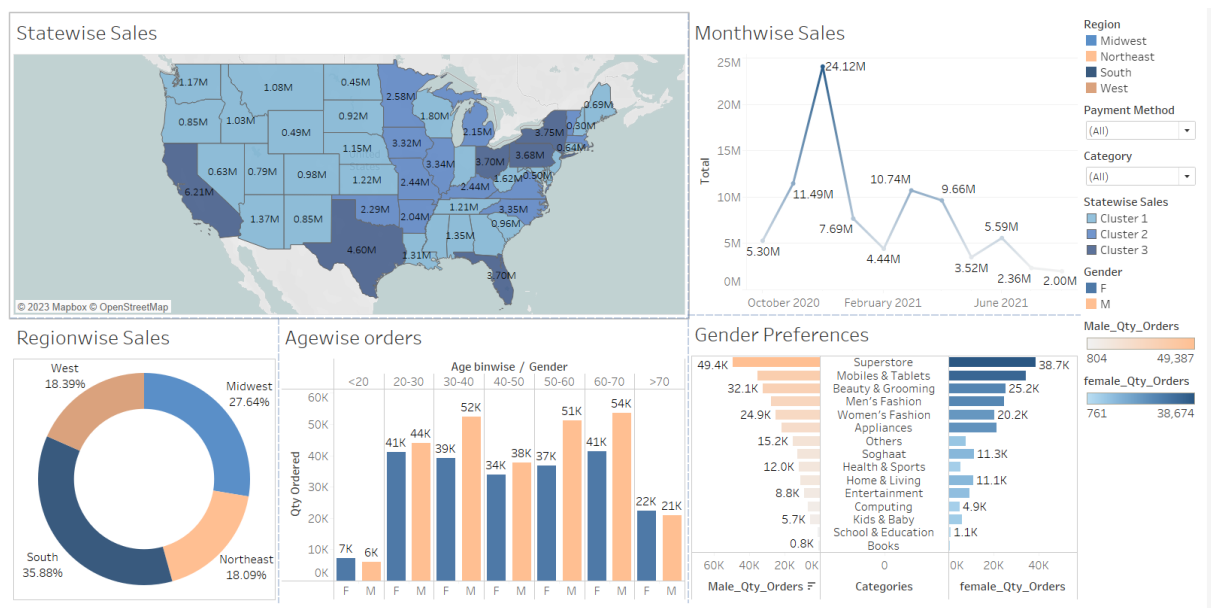


Image 1.

The Paper Landscape in Image 1 shows four charts—a map, a tree map, a bar chart, and a stacked bar chart—that are used to help sales analysts with their request for information. While a tree map will show information about sales made in various months and region, a map chart will show total sales broken down by state. Additionally, a simple bar chart will show the number of orders placed by various AGR groups and gender combinations, and a stacked bar chart will show information about consumer preferences for each category according to gender.

## Tableau Design:

There are a total of five charts in the final dashboard that was created in Tableau: a map, a line chart, a donut chart, a bar chart, and a butterfly chart.



final design in addition to the map. This choice was taken because the donut chart offers a more understandable illustration of sales data based on various regions. With the addition of this new chart, the design has been made more thorough and user-friendly.

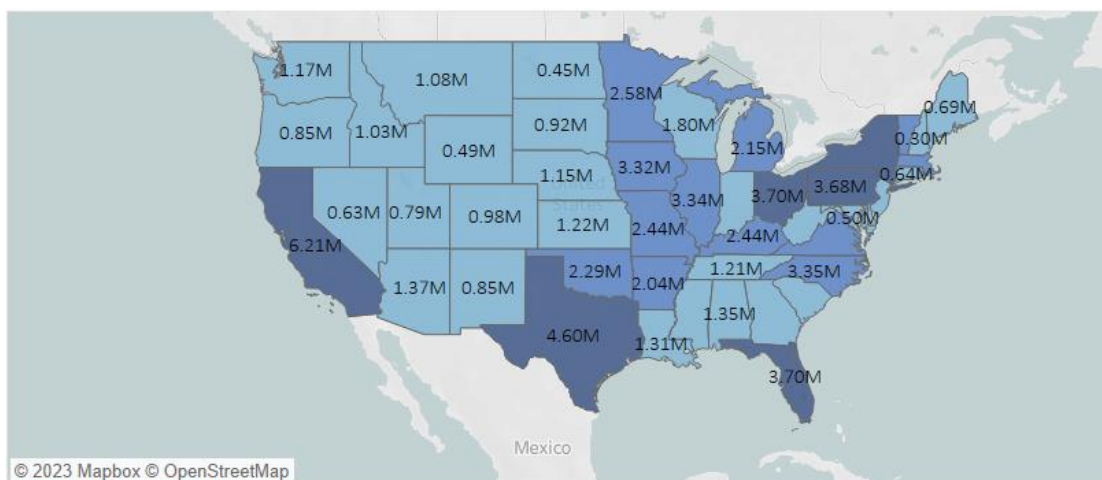
- A tree map was initially used to show monthly sales while developing a prototype. It was soon realised, nevertheless, that this approach wasn't the most practical because it didn't offer a precise temporal analytical perspective. As a result, a line chart was used in the final design to illustrate the sales trend over time, specifically on a monthly basis. This decision was made to provide a more comprehensive and easy-to-read representation of the sales data.
- The prototype first used a stacked bar chart to show the number of orders for each gender's preferred category. However, realising that this strategy was inadequate, a superior replacement was explored. As a result, a butterfly chart was chosen for the dashboard's final layout. Butterfly charts are more appropriate for the demands because they are great for contrasting and comparing two sets of data.

## Implementation:

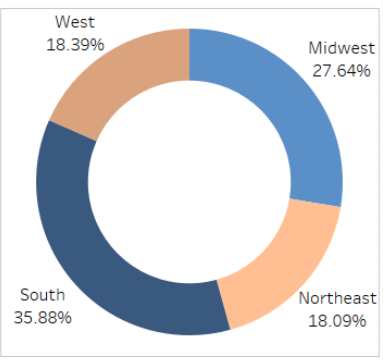
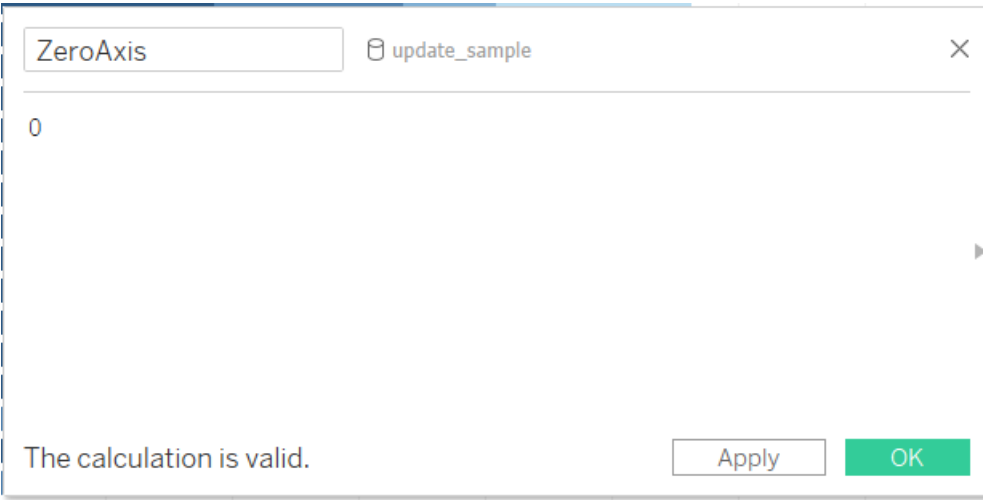
Implementation of final design:

A map, bar chart, butterfly chart, donut chart, and line chart were among the visualisations that were employed to build a visually appealing and educational dashboard. These visualizations were selected based on the specific data being presented and the message that needed to be conveyed. More importantly, dashboard is divided into two parts, first part is about sales analysis. To begin with, the "state" and "total" variables were used to depict sales data from the dataset on a map to gain information on sales by state. This made it quick and simple for the viewer to determine what were the sales of each state. Applying clustering to the map's sales data allowed researchers to better understand how sales were spread around the states and spot any groups of states that displayed consistent sales trends.

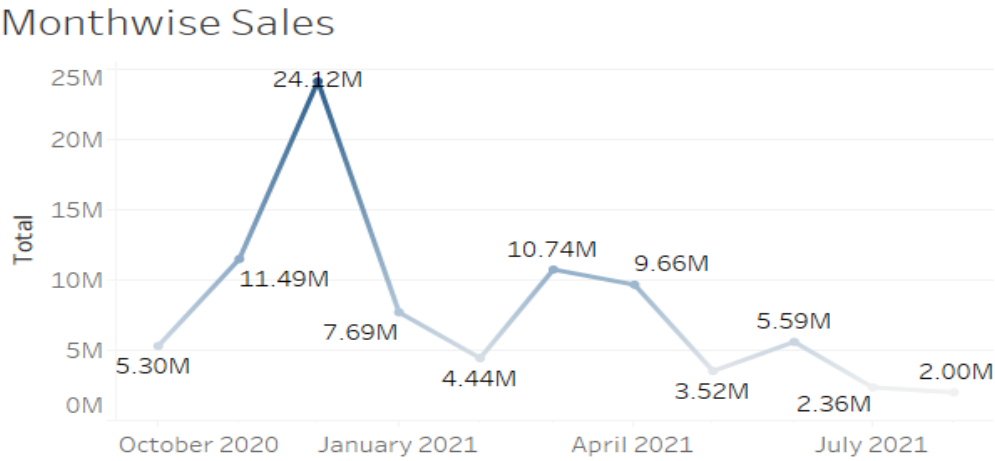
### Statewise Sales



The "region" variable was used in a donut chart to display sales by region in addition to the map. This type of chart is effective for displaying the proportions of a whole and allowed the viewer to quickly see how regions contributed to the overall sales. To make this chart we used zero axis and dual axis concepts which are shown below. So, we created 2 pie charts and combined it to make donut chart.



The "month" variable was then utilised to depict monthly sales data in a line chart. This kind of graph helped the reader understand how sales changed over the course of the year and is useful for showing trends over time. To create this chart, we used sales on y axis and months on x axis and , additionally, we applied color to sum of "total"(sales) to get shade.



The second part of the analysis mainly focuses on the quantity of orders. The goal was to understand how the quantity of orders varies by age and gender and to gain insights into customer preferences for different product categories.

A simple bar chart was also utilised to analyse the overall number of orders by age and gender. This kind of graph let the reader rapidly understand how the number of orders differed by age and gender and how the data was distributed across various categories. To create this, we utilised "age group" on x axis and "gender" as color and quantity of orders on y axis. We created bins of age with 10 years of intervals which is shown below.

Describe Field ×

Age binwise

Role:

Discrete Dimension

Type:

Numeric bin

Bin size:

10.0

Remote column:

[USA\_sales\_2020-21.csv].[age]

Remote type:

Eight-byte, signed integer

Contains NULL:

No

Status:

Valid

Domain (7 members)

<20

20-30

30-40

40-50

50-60

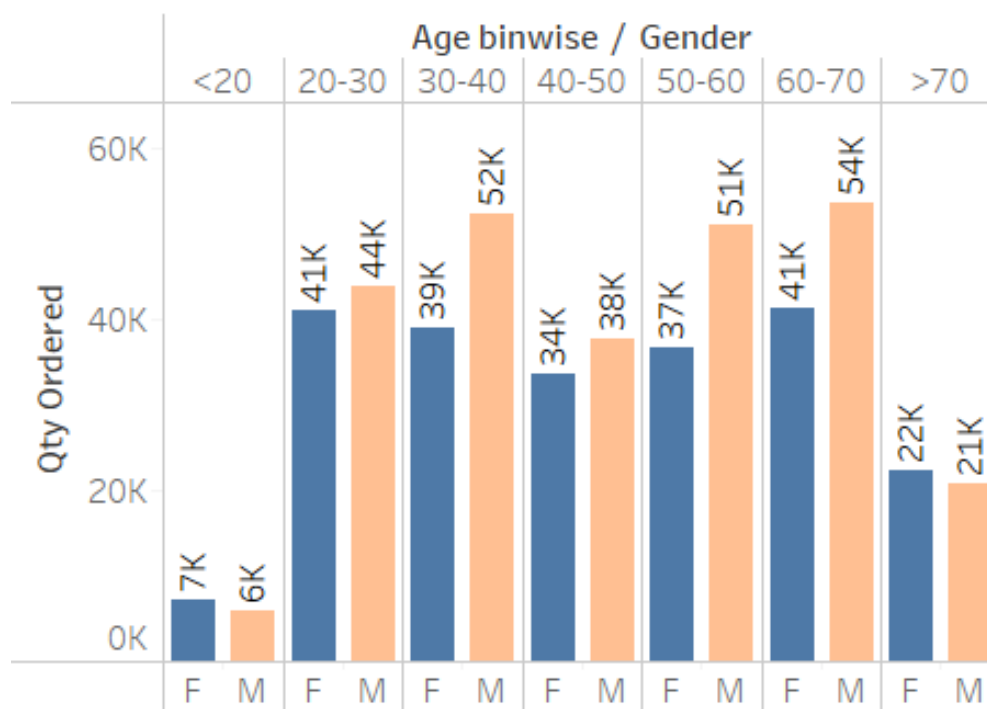
60-70

>70

Load

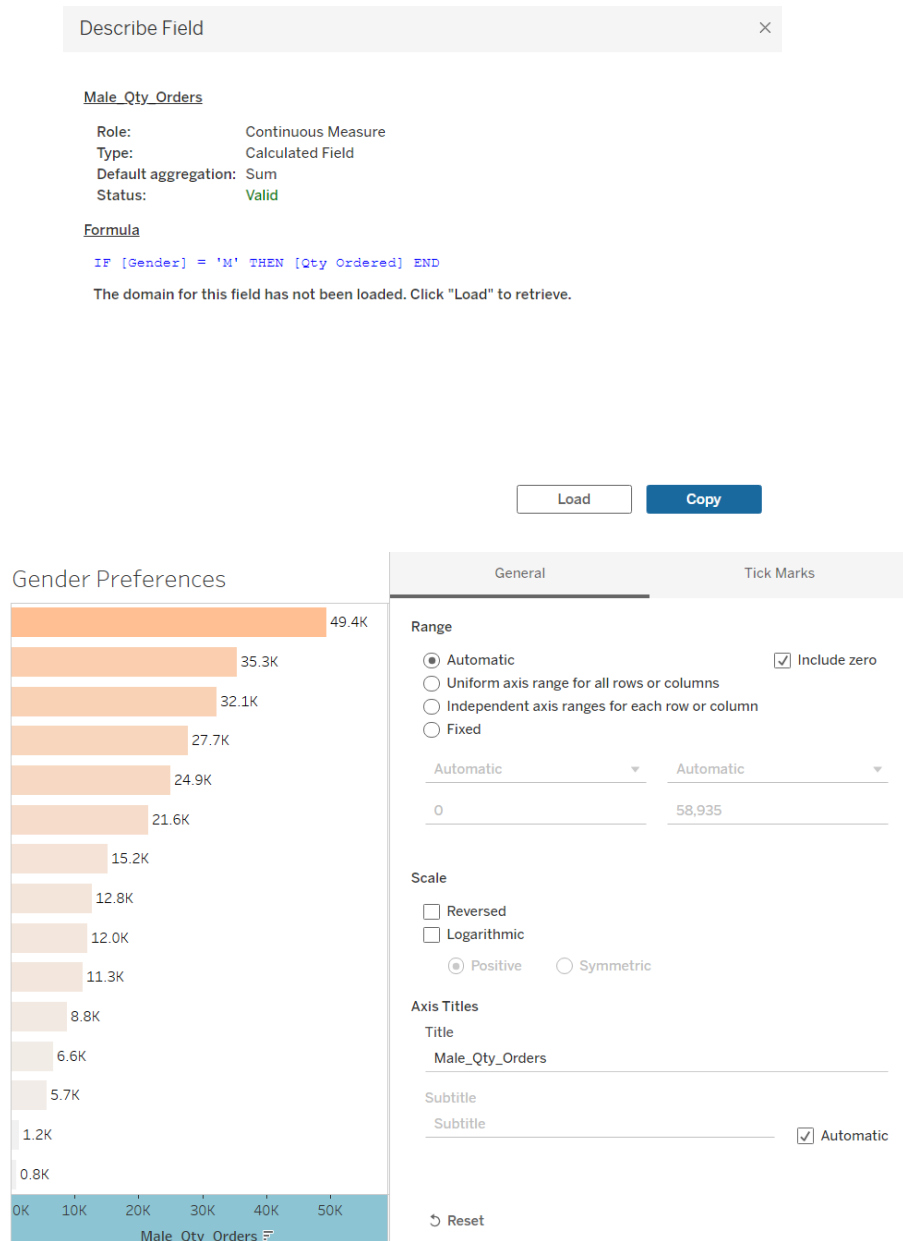
Copy

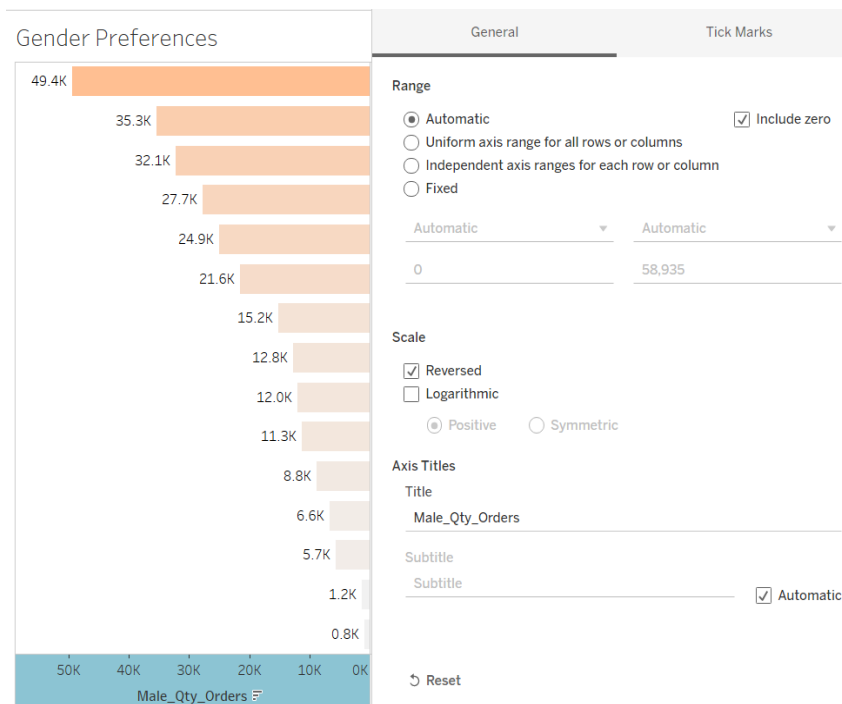
## Agewise orders



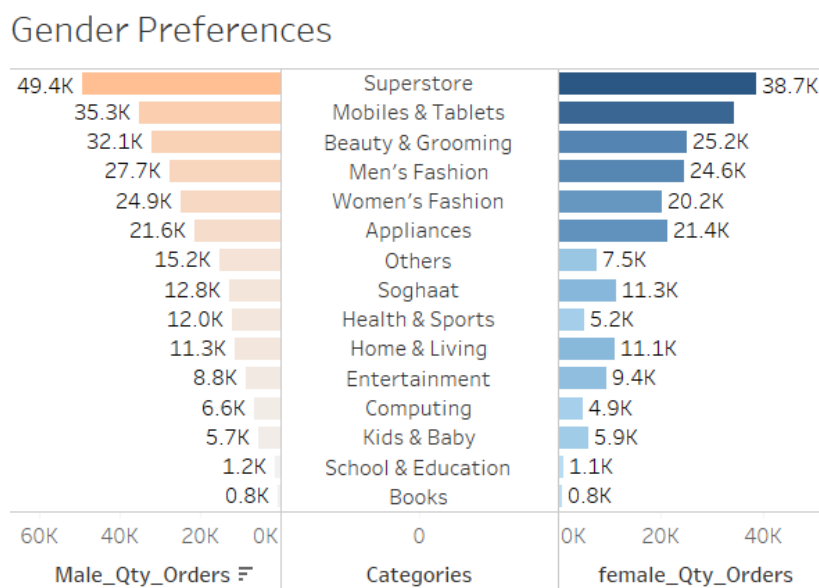
A butterfly chart was used to visualize the product category preferences by gender. We created calculative fields for quantity of orders of each gender called “Male\_Qty\_Orders” and “female\_Qty\_Orders” to assign it to each side of butterfly chart. Additionally, zero axis variable was created using calculating field to separate genders.

Created field using calculated field for male: (Similarly, done for female)





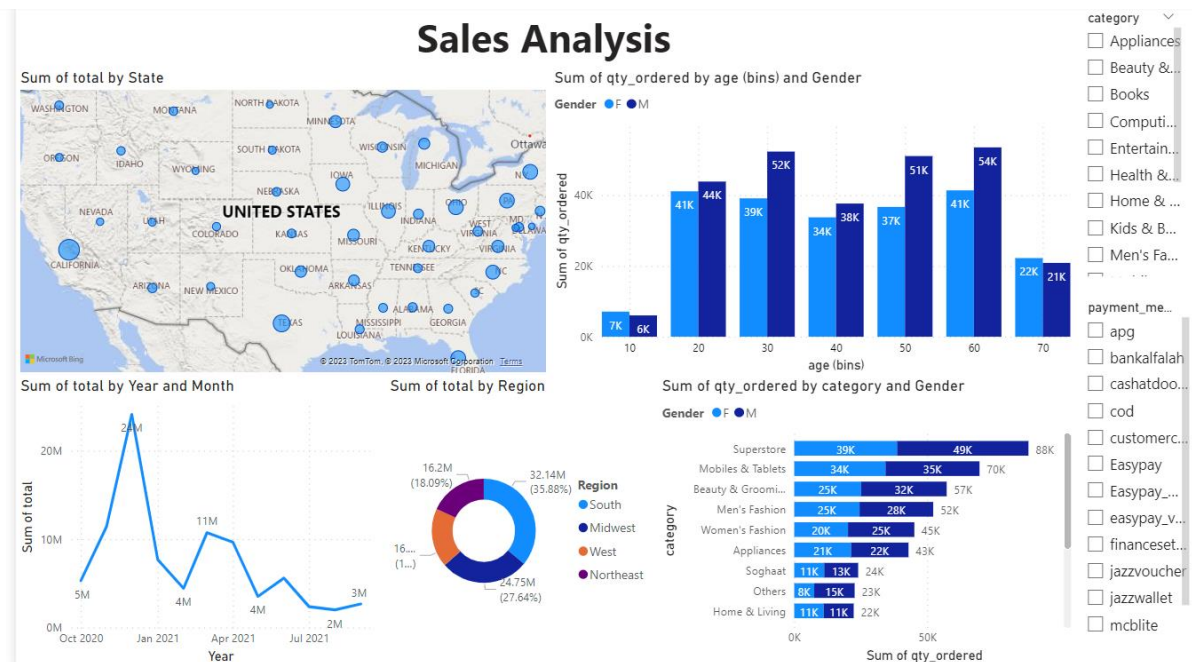
This is how we applied reversed scaling to make it look more efficient.



The "product category" used to create backbone of the chart, allowing the viewer to easily see which product categories were preferred by men and women.



## PowerBI Design:



The design of a Tableau dashboard and a PowerBI dashboard are largely similar with one exception. The butterfly chart, which is a type of bar chart used to show the sensitivity of an outcome to changes in certain variables, is not available in PowerBI. Instead, the tornado chart, also known as a butterfly chart, could have been used, however, I'm facing a problem where I don't have access to premium charts in power bi, which is unfortunate. To work around this, I have resorted to using a stacked bar chart. While it can be used as a butterfly chart or a tornado chart, it doesn't offer the same level of significance as those premium charts would have provided. The stacked bar chart has nearly replaced the need for butterfly or tornado charts by effectively representing the number of orders for each category by gender.

## Walkthrough:

Let's have a walkthrough of how our dashboard answers our questions. To begin with,

1. How does the sales performance of a particular product category vary across various geographic regions and states, as well as throughout different months of the year and is there any similarities in sales across different states of that product category?

To answer this question, Initially, we will choose a category from the category variable, and our attention will be directed towards examining items in the "Mobiles & Tablets" category.

## Walkthrough

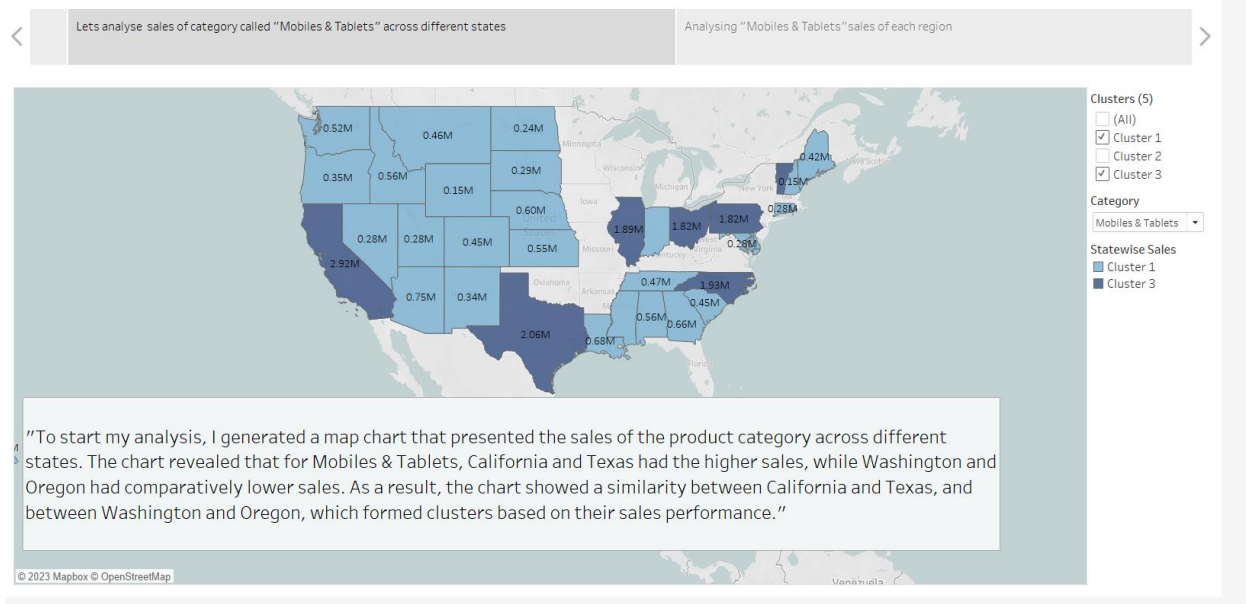


Figure 10.

Figure 10 provides information about sales across different states for category "Mobile & Tablets". California had the largest number of sales, totalling 2.92 million, for the category "Mobile & Tablets," whereas Wyoming had the smallest number of sales, at just 0.15 million. Moreover, California and Texas are regarded as states with equivalent sales patterns, as they both belong to the same cluster called Cluster 3. Sales analyst could use this information to identify the best and worst-performing states in the "Mobile & Tablets" category and plan marketing strategies accordingly. They can also study the sales patterns of states within the same cluster to draw insights on customer behaviour, preferences, and trends, which can inform their decisions on product development, pricing, and distribution.

Now moving forward, now we will see sales across regions.

## Walkthrough

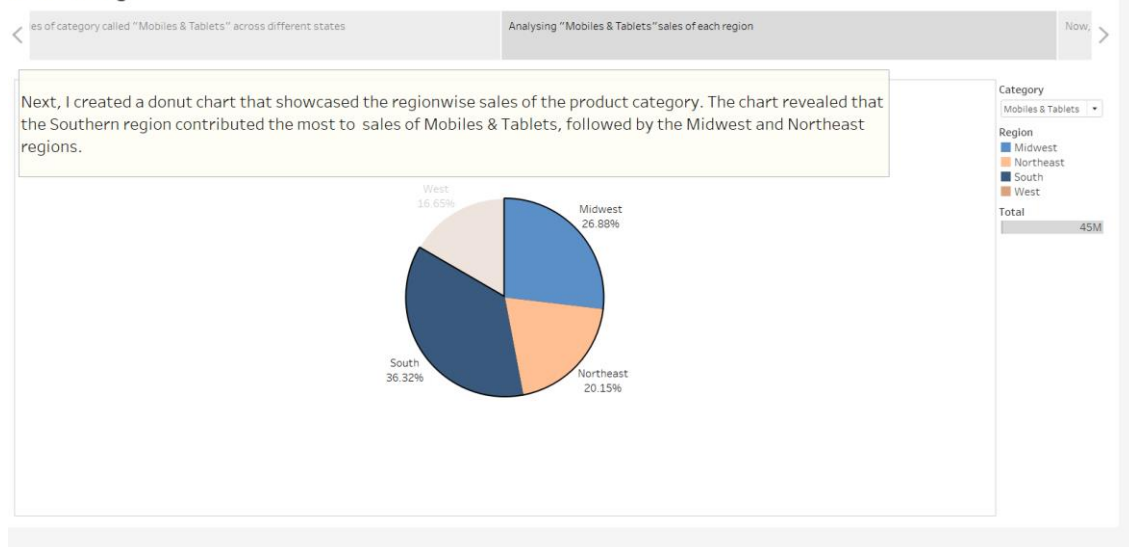


Figure 11.

From figure 11, The sales data indicates the level of contribution by various regions towards the sales of different product categories. The Southern region has the highest contribution of 36.32% towards the sales of the "Mobile & Tablets" category, while the Midwest region comes second with 26.88%. this information can also be used to make informed decisions about inventory and supply chain management, ensuring that the right products are stocked in the right regions to maximize sales potential.

Now, we will look at monthly trends of "Mobile & Tablets" category.

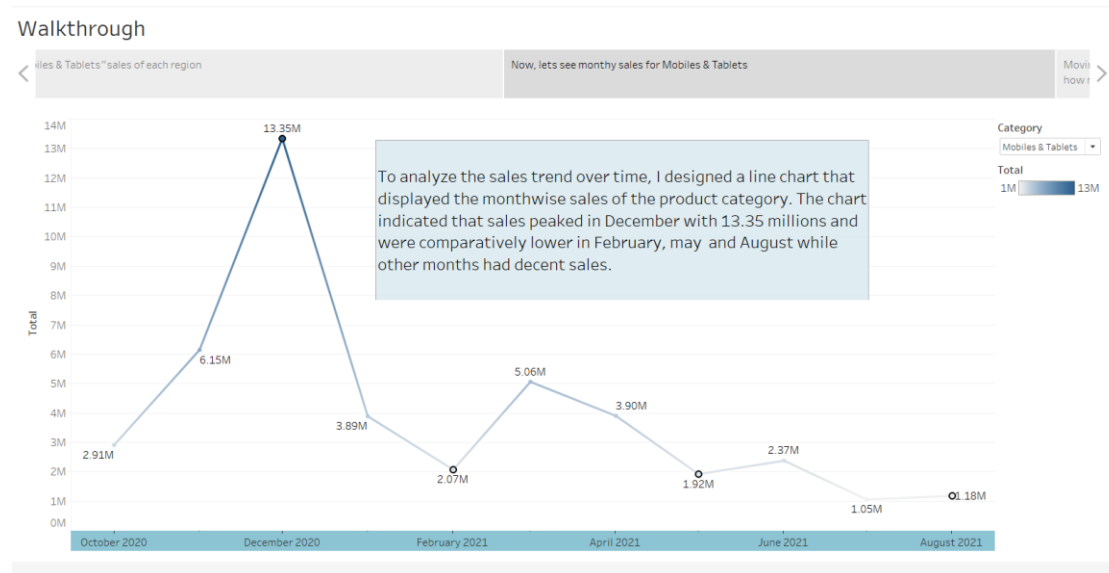


Figure 12.

Figure 12 shows that, In December, the sales of "Mobile & Tablets" reached the highest point with 13.5 million, while the sales were lower in February and August. However, the sales in the other months were reasonable. A sales analyst can investigate the causes of decreased sales in August and February and develop plans to boost sales in those months. Finally, they can make plans based on the data's estimate of future sales.

2. What is a quantity of product orders across different age groups and gender, and how does the order quantities changes for product category by gender?

## Walkthrough

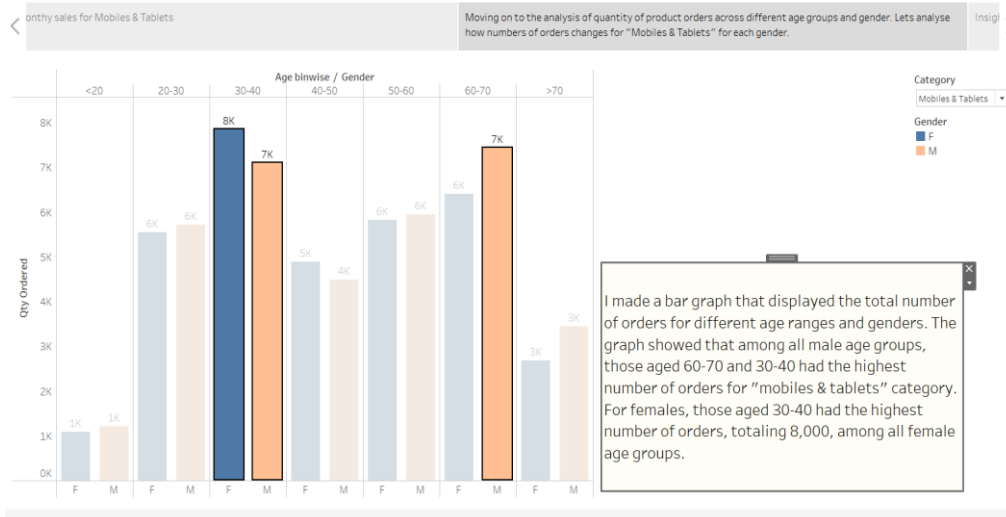


Figure 13.

To illustrate the results of the analysis, we created a bar chart that showed the total quantity of orders for each age group and gender shown in figure 13. The chart showed that the highest number of orders for "Mobile & Tables" were placed by two male's groups, aged 60-70 and aged 30-40, while females aged 30-40 placed the highest number of orders. A sales analyst can focus on promoting the "Mobile & Tablets" category to males aged 30-40 and 60-70 and females aged 30-40 as they appear to be the most interested in making purchases.

Now, lets see how this "Mobile & Tables" by gender:

## Walkthrough

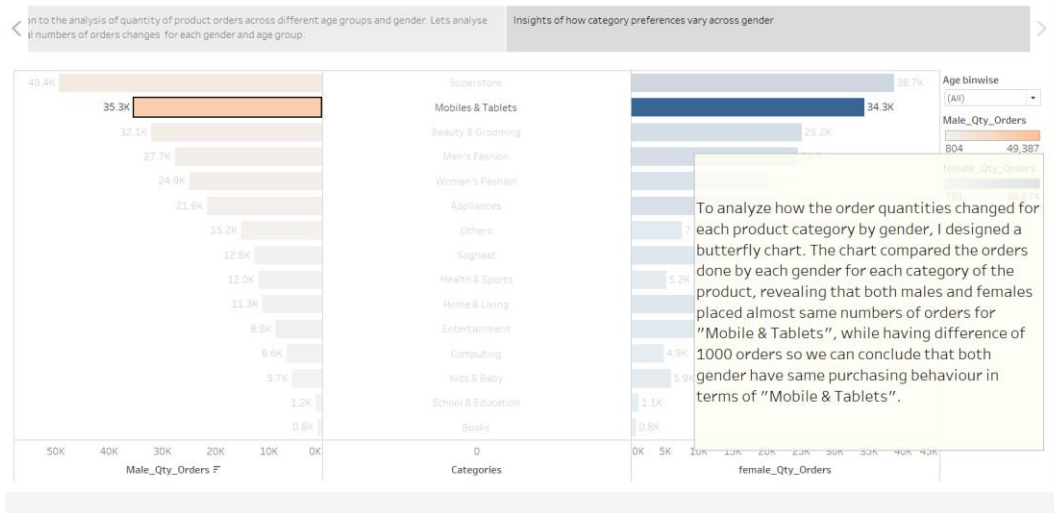


Figure 14.

The chart illustrated the orders made by each gender for each product category, and showed that for "Mobile & Tablets," both males and females placed nearly equal numbers of orders, with a difference of only 1000 orders. This indicates that the purchasing behavior of both genders is similar for "Mobile & Tablets." it can help them to better understand the consumer behavior for this product category, and to tailor marketing and promotional efforts

accordingly. For example, they may choose to run gender-neutral advertising campaigns that appeal to both male and female customers.

Now, let's take another viewpoint, suppose sales analyst wants to analyse patterns of people aged 60-70.

From figure 15, it is evident that the number of orders made by men between the ages of 60 and 70 is 54,000, and for women in the same age range, it is 41,000. These figures are comparatively higher than those for individuals in other age groups.

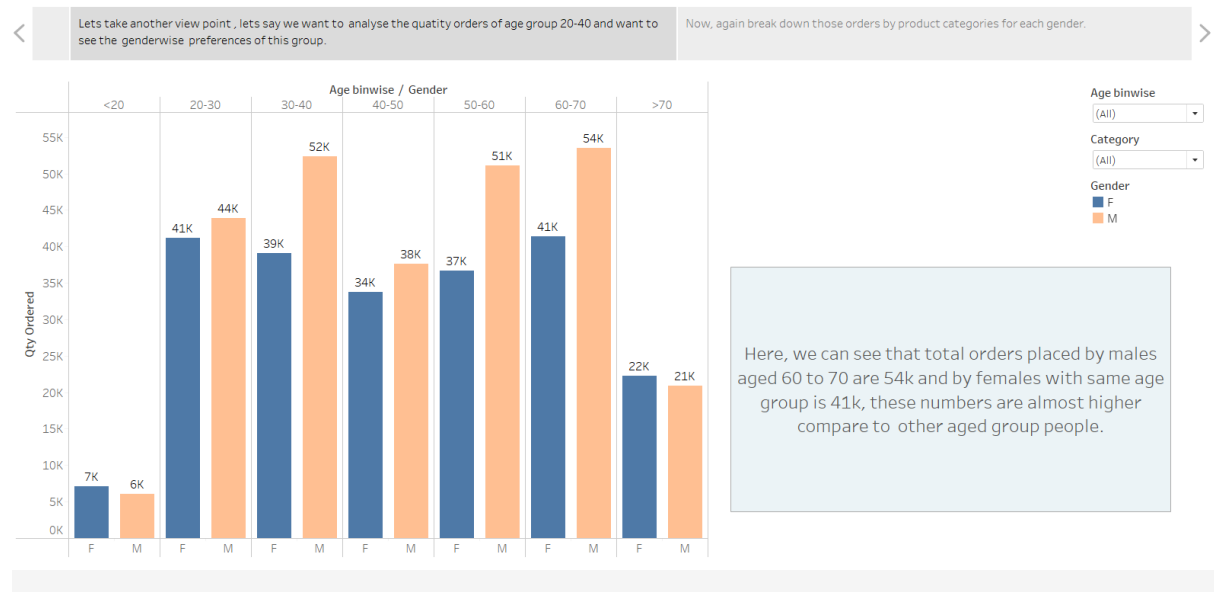


Figure 15.

Figure 16 reveals that the superstore is the most popular choice for individuals in the 60-70 age group to place orders. Moreover, it is worth noting that men in this age bracket have a higher tendency to order from superstores than women.

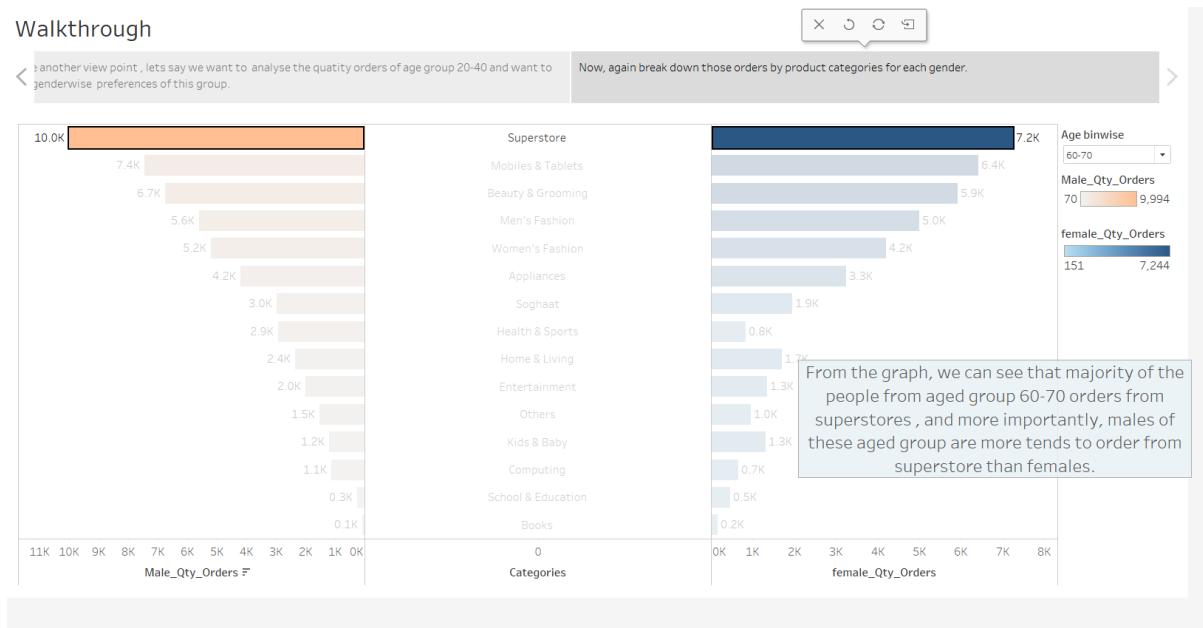


Figure 16.

This data can be used by a sales analyst in a number of ways to guide their sales approach. For instance, to boost sales, they can develop focused marketing efforts that are directed at people between the ages of 60 and 70, particularly both male and female persons. In addition, based on this group's purchasing patterns, businesses can concentrate on stocking goods that are more likely to appeal to them. Further investigation can be done by the sales analyst to ascertain the causes of the higher buy rates in this age group, and they can then use this knowledge to improve their sales and marketing strategy. Additionally, the sales analyst can use data about gender preferences in a variety of ways to improve their sales strategy. For instance, they can think about developing specialised discounts or specials for superstore goods that are more likely to appeal to men between the ages of 60 and 70.

## Reflective Discussion

Tableau proved to be an effective tool for analysing the sales performance of a specific product category over multiple geographic locations, states, and months of the year, as well as the volume of product orders across various age groups and gender. Its user-friendly interface made it simple to visualise complex data sets, giving insights that would have been challenging to find otherwise. One of Tableau's advantages is its capacity to produce interactive dashboards, which let users examine data from many perspectives. When analysing a product category's sales performance across various states and geographical regions, this tool proved helpful. It was simple to spot patterns and trends in the data by using filters and highlighting tools. Tableau's capacity to manage huge data sets without compromising performance is another asset. This was crucial when examining the volume of product orders across various age and gender groupings. It was simple to spot trends in the data and reach insightful conclusions because of the ability to filter and group data fast. Tableau's price is one drawback, though, as it might be prohibitively expensive for individuals or small enterprises. Additionally, there is a high learning curve when using Tableau for the first time because it necessitates a solid grasp of ideas in data visualisation and analysis.

Overall, this project was successful in achieving its objectives and providing insights into the sales performance of a particular product category across different geographic regions, states, and months of the year, as well as the quantity of product orders across various age groups and gender. Through this project, I have learned valuable skills in data visualization and analysis, and I look forward to applying these skills to future projects.

One personal learning objective that I want to achieve soon is to improve my data storytelling skills. While Tableau provided excellent visualizations and insights, the ability to effectively communicate these findings to stakeholders is critical. I plan to work on improving my communication skills and finding ways to effectively tell a story with data.

## References

Efti, S., 2023. *How to Perform a Sales Analysis (Step-by-Step): Methods & Metrics*. [Online].