



Sales & Customer Analysis

Project Phase - 2

Group Name: Data Vizards

Members:

Muskan Garg

Vatsal Gohel

Engaging Summary (Elevator Pitch)

Achieving success in the fast-paced world of e-commerce, where online shopping has become the new norm, requires an understanding of customer behavior, which is key to the success of your business. This dataset contains information about orders placed by customers on an e-commerce website. The dataset has a total of 286,392 entries with 36 columns. Through analysis of sales data, we uncovered fascinating insights that shed light on consumer behavior and trends. One of the fascinating aspects of our research is how the age of customers influences their behavior. By analyzing the data, we can come up with patterns and trends that give us an insight into how different age groups engage with e-commerce platforms and determine what their needs are. Understanding these nuances allows businesses to tailor their marketing strategies, user experience, and product offerings to effectively target and engage specific age demographics for their marketing strategies, product offerings, and user experience.

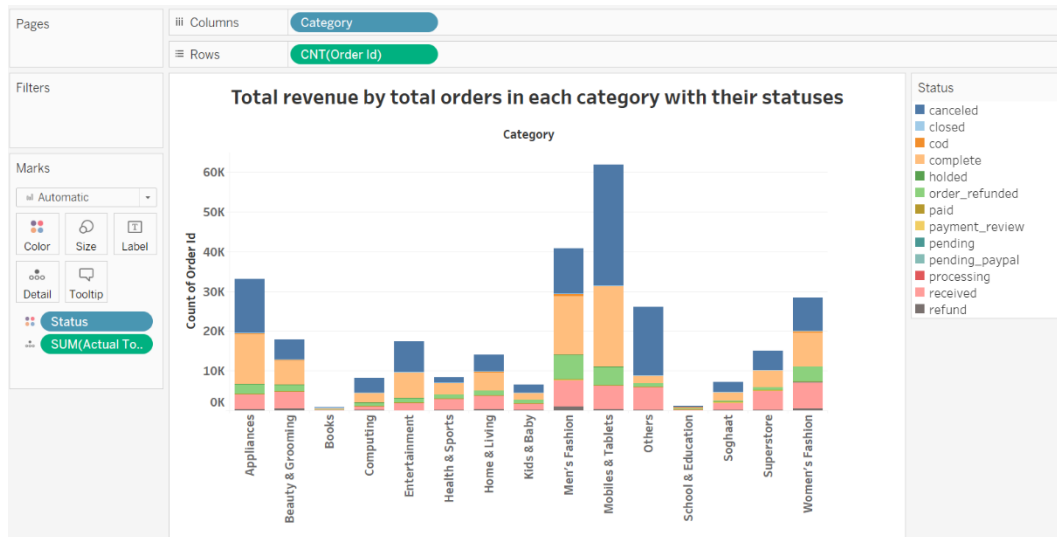
A crucial aspect we explore is the effect of discount percentages on sales. Analyzing the data, we can identify the relationship between discount rates and customer behavior. With this information, businesses can optimize pricing strategies, determine the most effective discount levels, and boost sales by leveraging discounts' psychological impact. The traditional Cash on Delivery (COD) method remains popular despite the digitalization of payments. It challenges the assumption that all customers prefer digital payments. Adapting to the reasons behind this preference for COD, businesses can provide a seamless and trustworthy buying experience, accommodating old-school preferences of some customers while still incorporating modern conveniences.

In summary, our dataset provides insight into age-related customer behavior, discount percentages' impact on sales, and the predominance of Cash on Delivery over other payment methods. These insights help businesses develop targeted marketing strategies, optimize pricing, discounting, and payment options to cater to diverse customer preferences. Take advantage of this valuable knowledge to drive growth, enhance customer satisfaction, and secure a competitive edge.

Data Cleaning

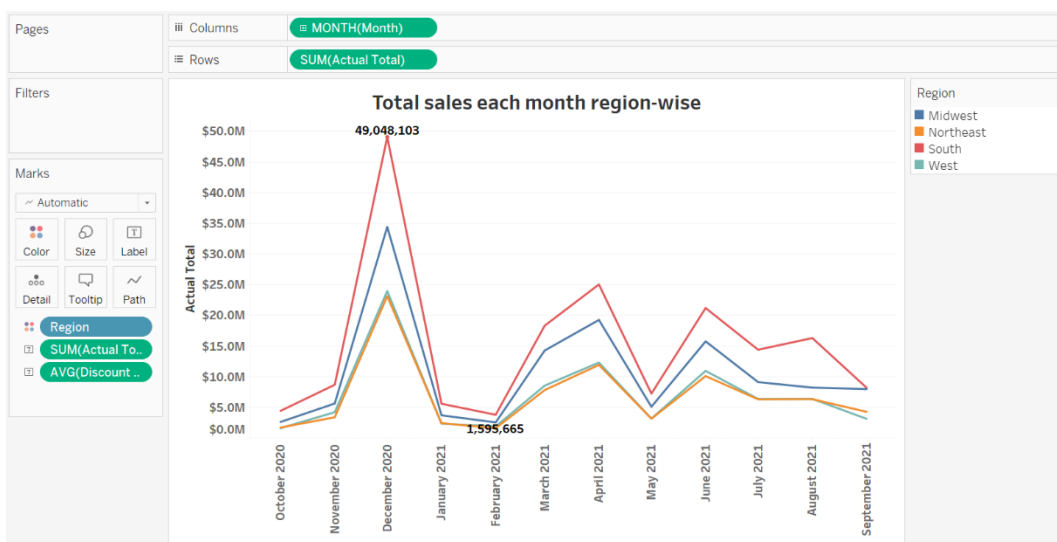
In order to visualize the dataset better and make the cleaning process easier, we loaded the data in a Python file. We performed some descriptive analysis on the data. We imported the necessary libraries to dive deep into the dataset to check for missing values, outliers, and duplications in the dataset. Outliers and missing values did not need to be rectified, so we did not touch the dataset. As the dataset contained sensitive information such as SSN, we removed it. There is an error with the value column as quantity ordered is decreased by a value in all rows. In order to correct the data, we created two calculated fields, "actual value" and "actual total". Likewise, we dropped those columns that were irrelevant and redundant such as value, total, Name Prefix, Middle Initial, User name, Full name, SSN. To analyze customer behavior patterns based on age group, we created a calculated field named age bin using that column.

Chart 1: - Total revenue by total orders in each category with their statuses



From this graph we are finding the total count of orders across different category and to visualize the finding we have used the Stacked bar chart and different order status are seen across each category. From the graph, we can visualize that the greatest number of orders are from the Mobile & Tablets category followed by the Appliances whereas least number of orders are done for the book category. While we discuss about the status of the orders the most canceled as well as complete is for the Mobiles & Tablets category. The Interesting thing in the graph is that due to covid period the customers are not buying the Books and School & Education items instead they are preferring to buy the Appliances and clothing. Additionally, we can see that in Men's & Women's clothing there are a greater number of orders which are returned and it is refunded back to the customers followed by Mobiles & Tablets.

Chart 2: - Total sales each month region-wise



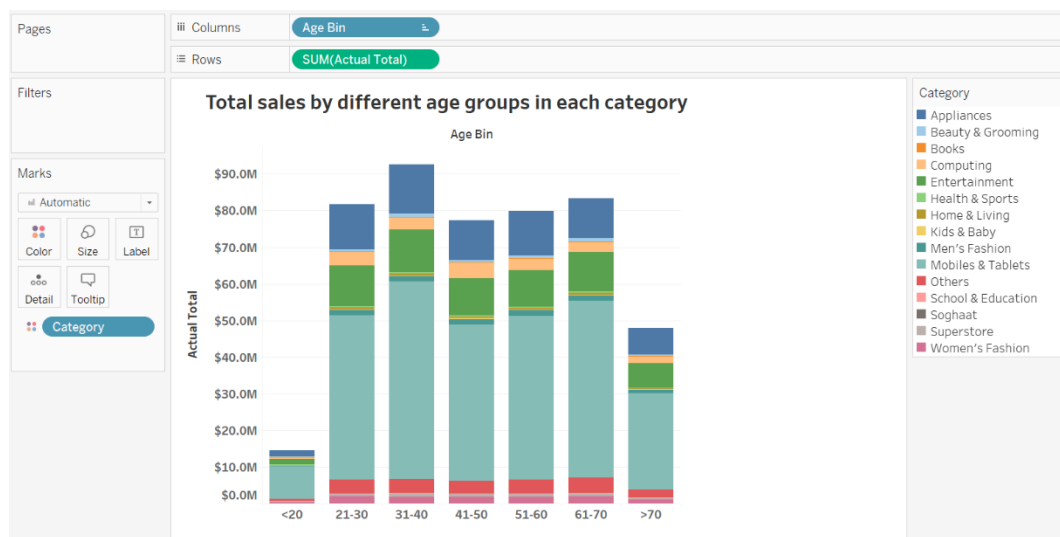
This graph shows the sum of total sales across the region over a period. We can see that among all regions the south region sales is highest followed by the Midwest region over a timeframe. We can see that all four regions follow the same pattern whenever the sales are increased or decreased for a specific month it is almost same for all. Similar pattern is also seen in average discount. In the Month of December, the sales are at Peak because due to the Christmas and new year sales there are various attractive offer through which customers get attracted and buy the items. Whereas in the month of February the sales for all the regions get declined because there is no festival offer going on and generally customers prefer to order if there is some exciting discount or offers available.

Chart 3: - Average Discount for each category



This Chart showcase the average and overall discount trends across all the categories, so to understand and answer one of our objective questions is there an impact on the total revenues generated in each category based on the discounts given. Here, we can clearly see “Soghaat” category has the maximum average discount offers of 31.50. Although, from our analysis we noticed it only has sales of \$0.85M. Whereas, “Mobiles and Tablets” category has an average discount of 14.12, still it managed to generated the maximum revenue of \$270.06M.

Chart 4: - Total sales by different age groups in each category



This graph depicts the value of total sales across different age group. Furthermore, to get more detailing the category is stacked into it. The greatest number of sales come from the age bracket of 31-40 followed by 21-30 because this age group are generally a college student or working professionals so they are much active to buy new items as well as they mostly get attracted by seeing others and wish to buy things. In each age bin we can see that most sales come from the Mobiles & Tablets category because nowadays this thing are necessity for everyone daily usage. These days the mind refreshing and enjoyment is also important so from graph we can say that all age group are spending their money more for their entertainment. There are some categories in which all age group spent less are like Books, Soghaat, Superstore and Home & Living.

This is an initial analysis of an E-Commerce dataset. Nevertheless, we are looking to examine further findings such as state-wise total revenues in order to understand market trends. The most popular categories and whether people are satisfied with them. Analyzing the categories with the most returned orders and the reasons for them. Our goal is to improve customer experience and inventory requirements by region, state, and customer type.