DSC - 465 PROJECT
CUSTOMER ANALYSIS.

Data Visualization on

Sales Dataset

by

Sidhant Thakur

Parth Patel

Saransh Thakur

Preetham Sai K

Sumera Fatima Khatoon

Under the guidance of: Dr. Eli T. Brown

DePaul University

21st November 2022

Project Summary

Introduction:

The purpose of this report is to visualize the sales dataset so that customer analysis can be performed on it. This report will use appropriate charts to represent revenue based on a variety of variables, as well as evaluate graphs and draw conclusions from them.

Data, Audience, Message:

Data:

Data consists of 286369 rows and 36 Columns
order_id (Numerical)
order _date (Ordinal)
Item_id (Numerical)
Product Name (categorical)
qty_ordered(Numeri)
price (Continuous)
value (Numerical)
Discount_amount (Continuous)
total (Numerical)
category (Categorical)
payment_method
bi_st (Categorical)

cust_id (Categorical)
year (Categorical)
month (Categorical)
ref_num (continuous)
Name Prefix (Categorical)
First Name (Categorical)
Middle Initial (Categorical)
Last Name (Categorical)
Gender (categorical)
Age (continuous)
full_name (Categorical)
E Mail (Categorical)
Customer Since (Ordinal)
SSN (Categorical)
Phone No. (Nominal)
Place Name (Categorical)
County (Categorical)
City (Categorical)
State (Categorical)
Zip (Nominal)
Region (Categorical)
User Name (Categorical)
Discount_Percent (Continuous).

Audience:

The audience could be a marketing team or a research team.

Message:

We used a calculated field to calculate revenue and appropriate charts to show revenue based on a variety of variables, as well as analyzing graphs and coming to a conclusion from them.

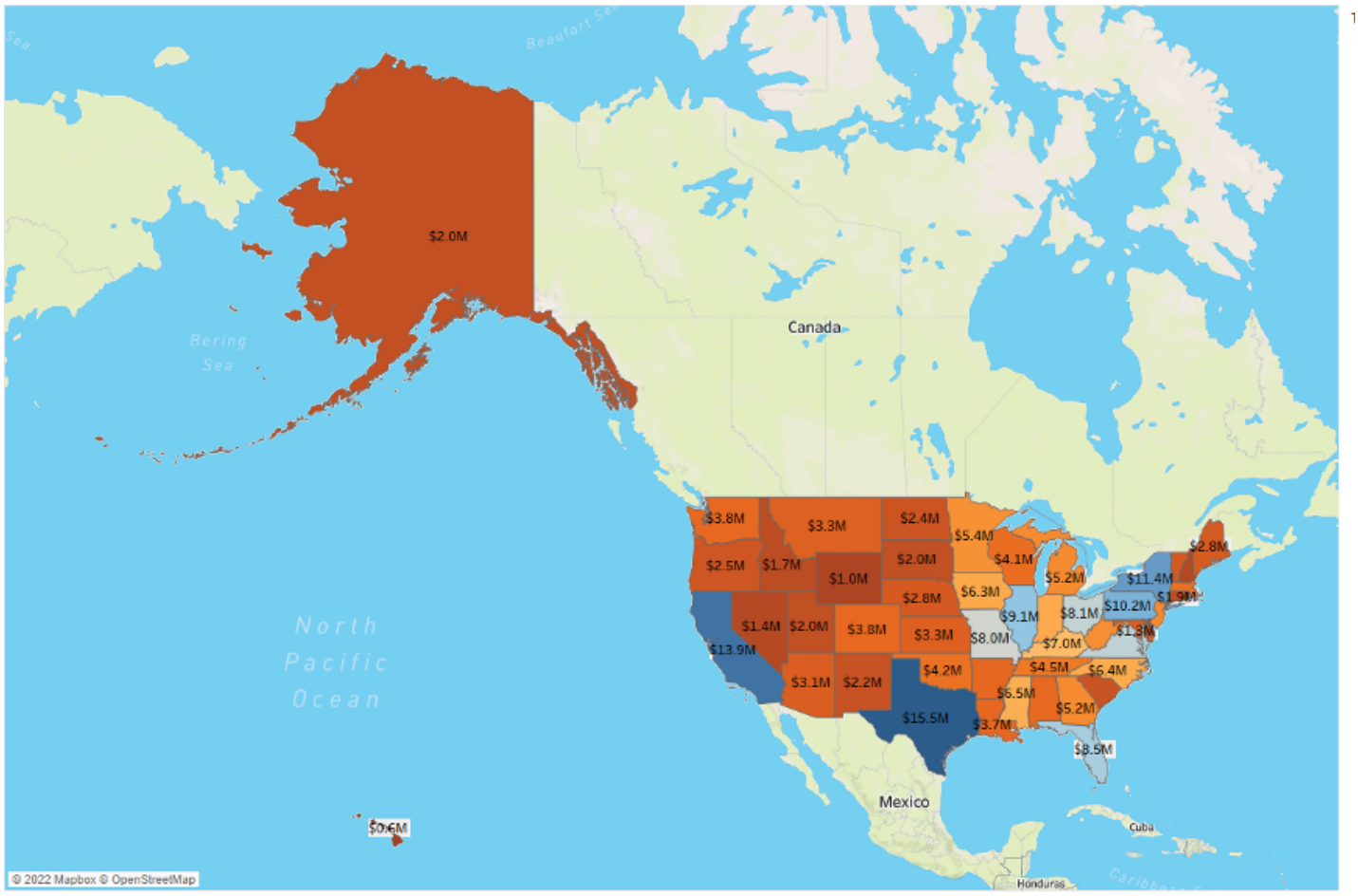
Application of Analysis and Visualizations:

When first looking at the dataset, the first questions raised were what variables influence data sales and how to determine revenue.

After then, decided to work on the following analysis

- Revenue per state.
- Revenue based on month of year.
- Order status based on revenue.
- Revenue based on Age
- Quantity – Discount percentage correlation
- Percentage of revenue per region
- Revenue per category per gender
- Payment Method

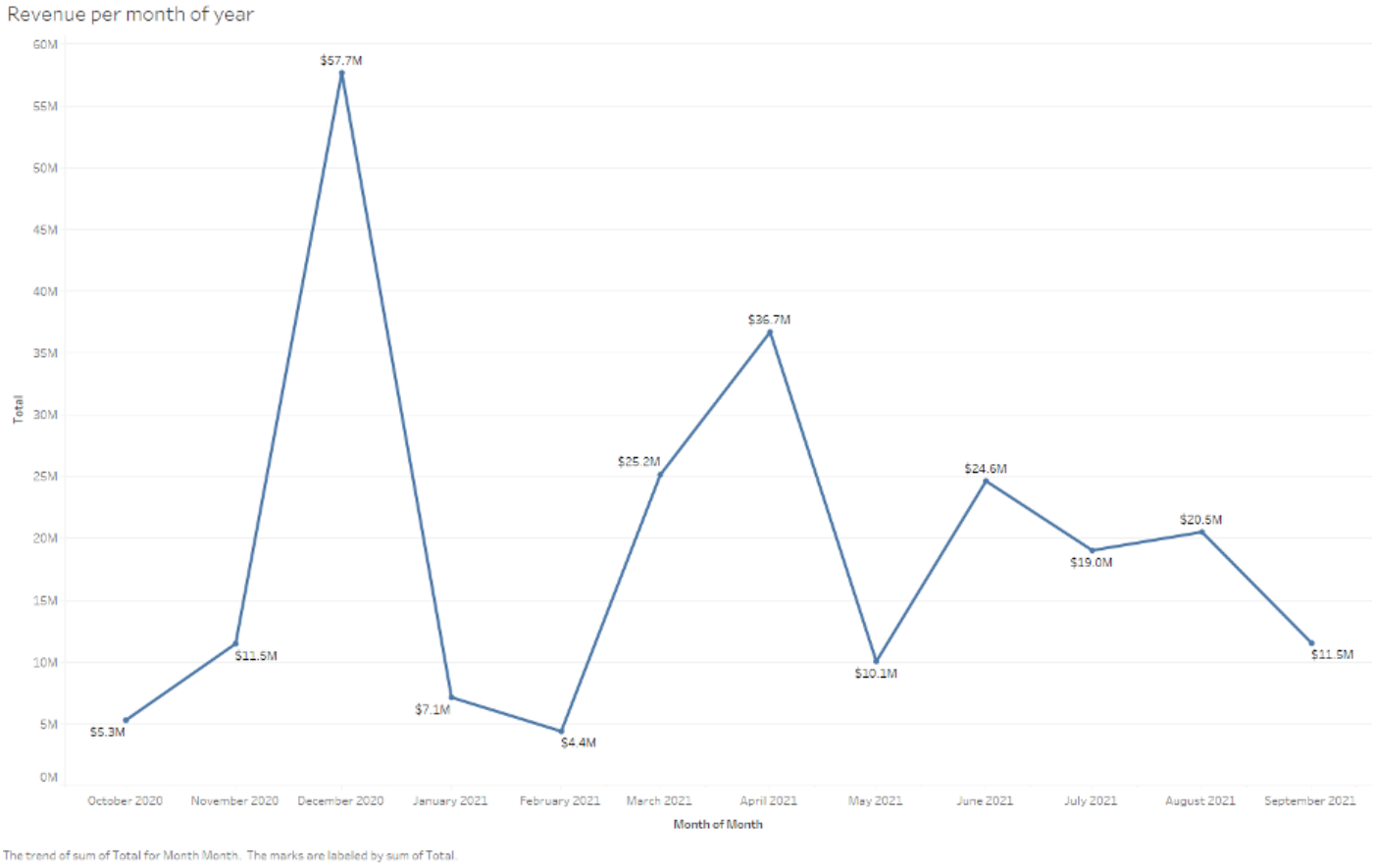
Revenue per state.



The visualization displays the revenue allocation by state, here we have used geographical plot and have used divergent color so that our graph has pre attentive qualities that means visual signals that are easily visible and do not require processing to recognize.

TX has the greatest revenue of $15.5 million, followed by California with $13.9 million. DC has the lowest revenue of $1.3 million.

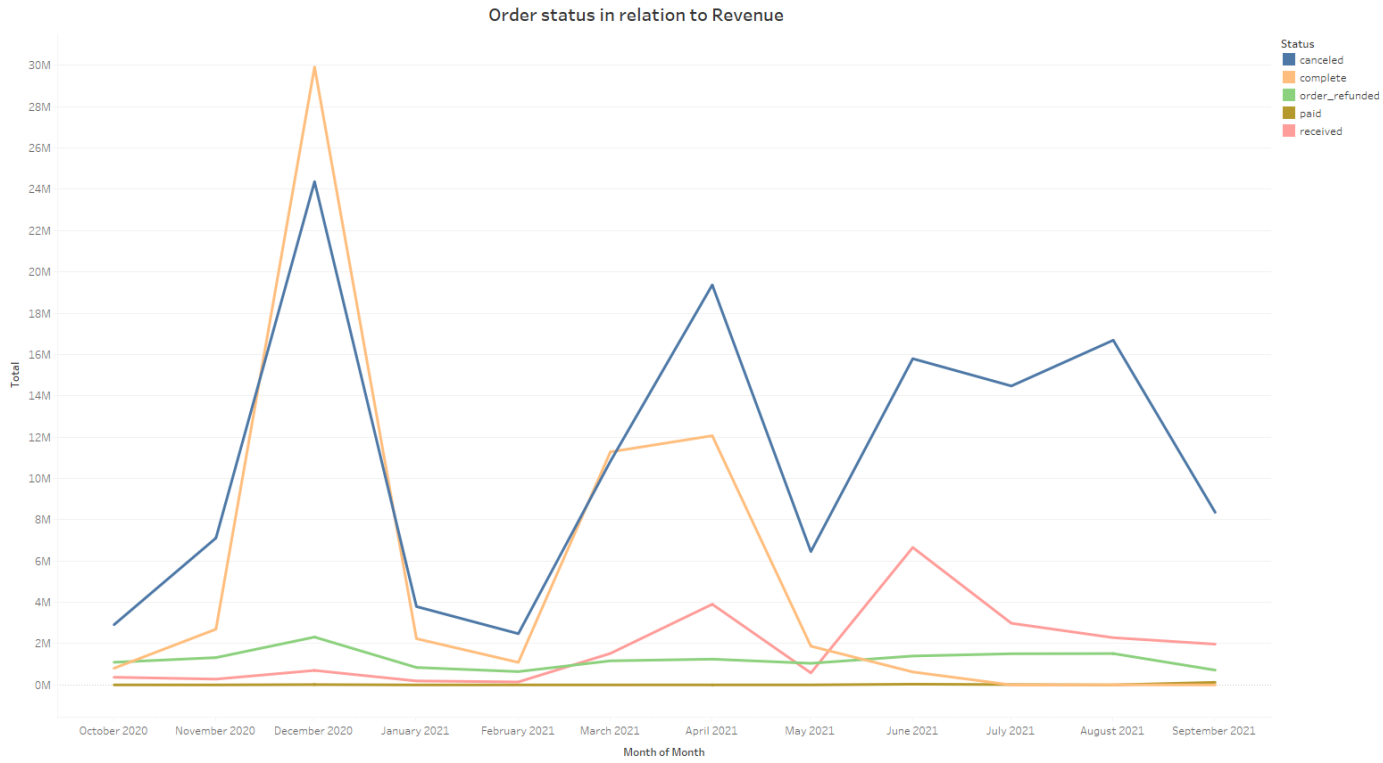
Revenue based on month of year.



The month of December generates the most revenue ($57.7 million), while February generates the least revenue ($4.4 million).

Revenue jumped by nearly 400% in November but fell by nearly 87% in January. Furthermore, from June to September, there is a downward trend.

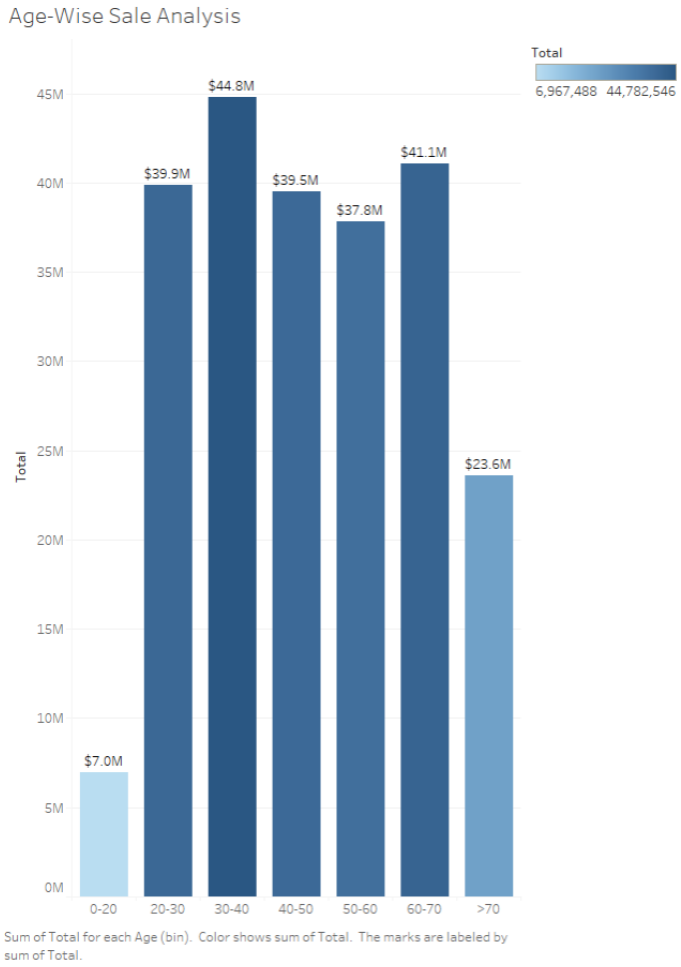
Order status based on revenue.



Since November to December have the most of the orders completed, December has the most revenue.

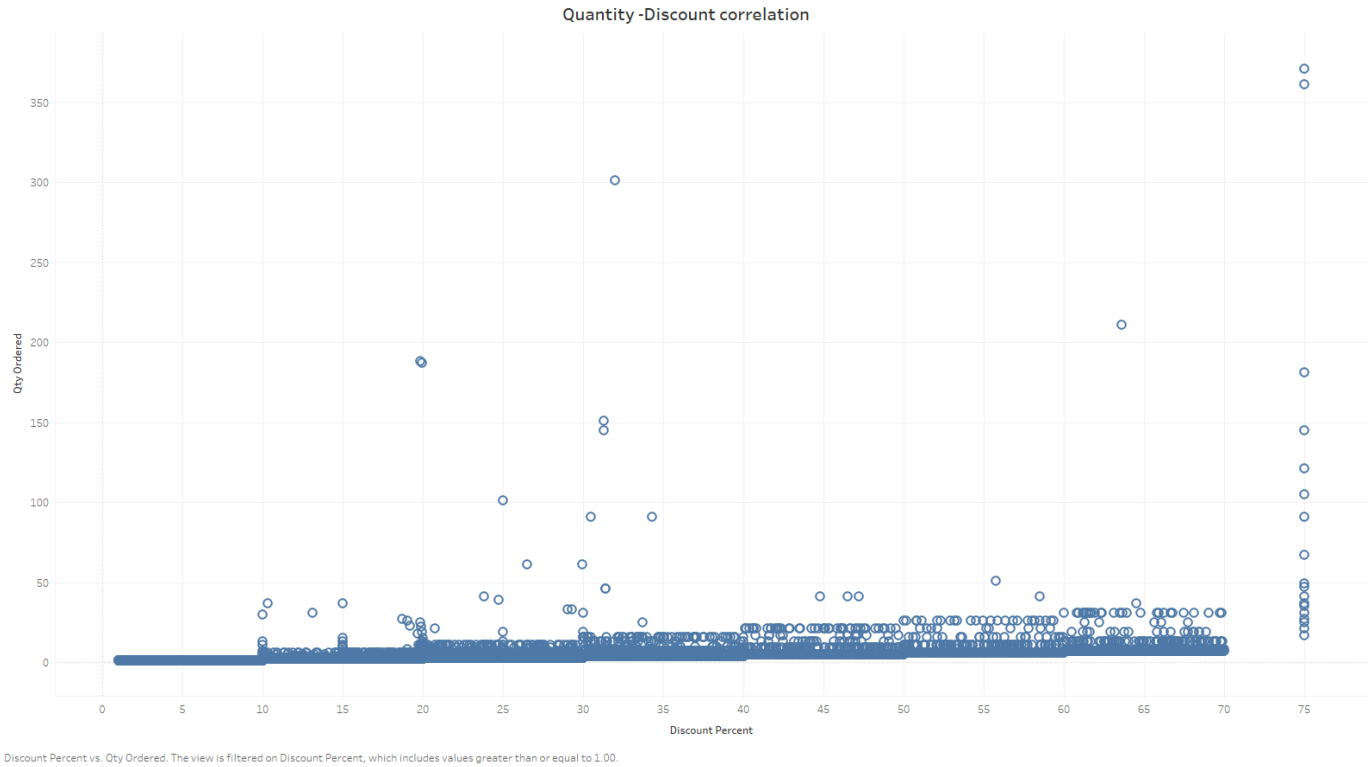
As December to January has the highest number of orders canceled, there is a significant decrease in revenue.

Revenue based on Age



I utilized bar graphs to compare items in different groups. This bar graph compares ages and makes it much easier to identify patterns. The 30-40 age group spent around 45 million, followed by the 60-70 age group, which spent 41 million. The 0-20 age group spent $7 million.

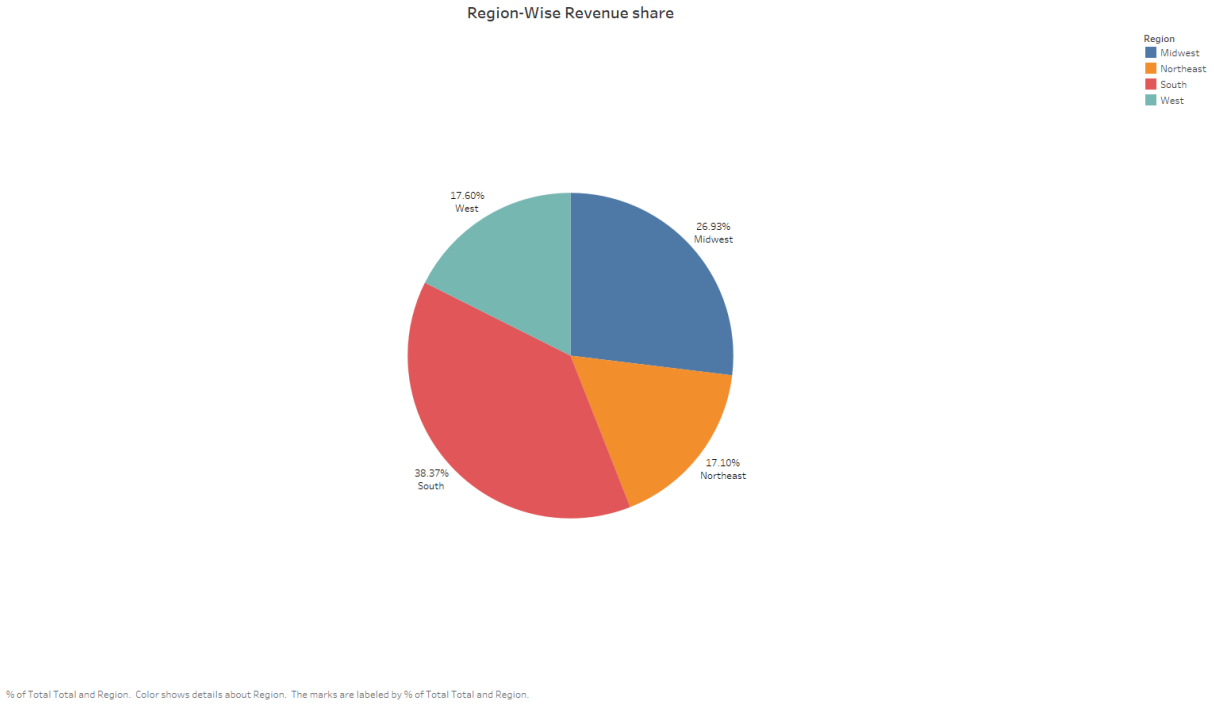
Quantity – Discount percentage correlation



As the discount percentage increases, so does the quantity order, and there is a positive correlation.

As a result, if the discount percentage is high, so is the quantity order.

Percentage of revenue per region:

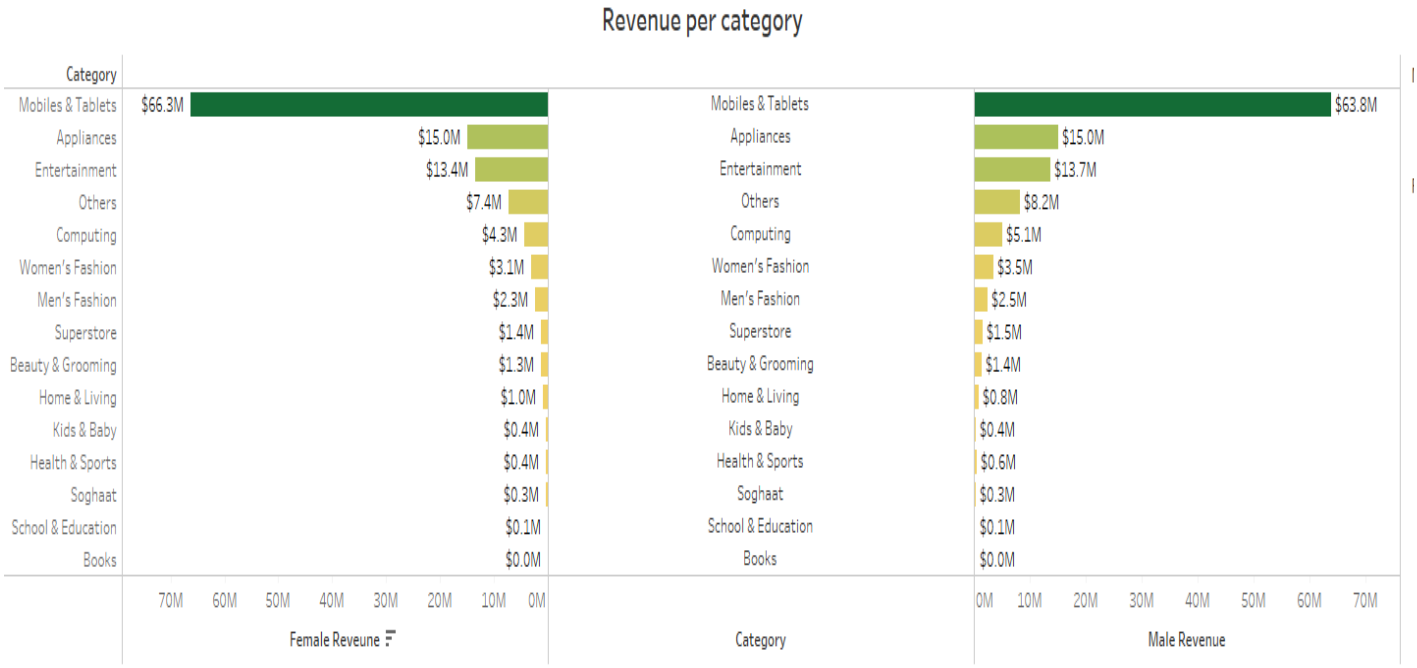


The pie chart shows the percentage of revenue generated from each region of the country.

The South region generated 38.37% revenue which is the highest of all the regions. The Midwest region generated 26.93% revenue which is next to the south region.

The west and the northeast regions generated 17.60% and 17.10% revenue which are the least of all the regions.

Revenue per category per gender:



Sum of Female Reveune, sum of zero axis and sum of Male Revenue for each Category. For pane Sum of Female Reveune: Color shows sum of Female Reveune. The marks are labeled by sum of Female Reveune. For pane Sum of Male Revenue: Color shows sum of Male Revenue. The marks are labeled by sum of Male Revenue.

The graph above depicts money generated by gender—male and female—in several categories.

Females can spend up to $63.8 million, while men can spend up to $66.3 million.

The following image shows that mobiles and tablets have a considerable impact on revenue, with ladies contributing $66.3 million more than males. Furthermore, the money earned by the Appliances is around $15 million for both genders.

The third-largest revenue-generating industry is entertainment, with male and female contributions reaching $13.4 million and $13.7 million respectively.

Except for mobile & tablets and home & living, men tend to have more revenue in each category than women.

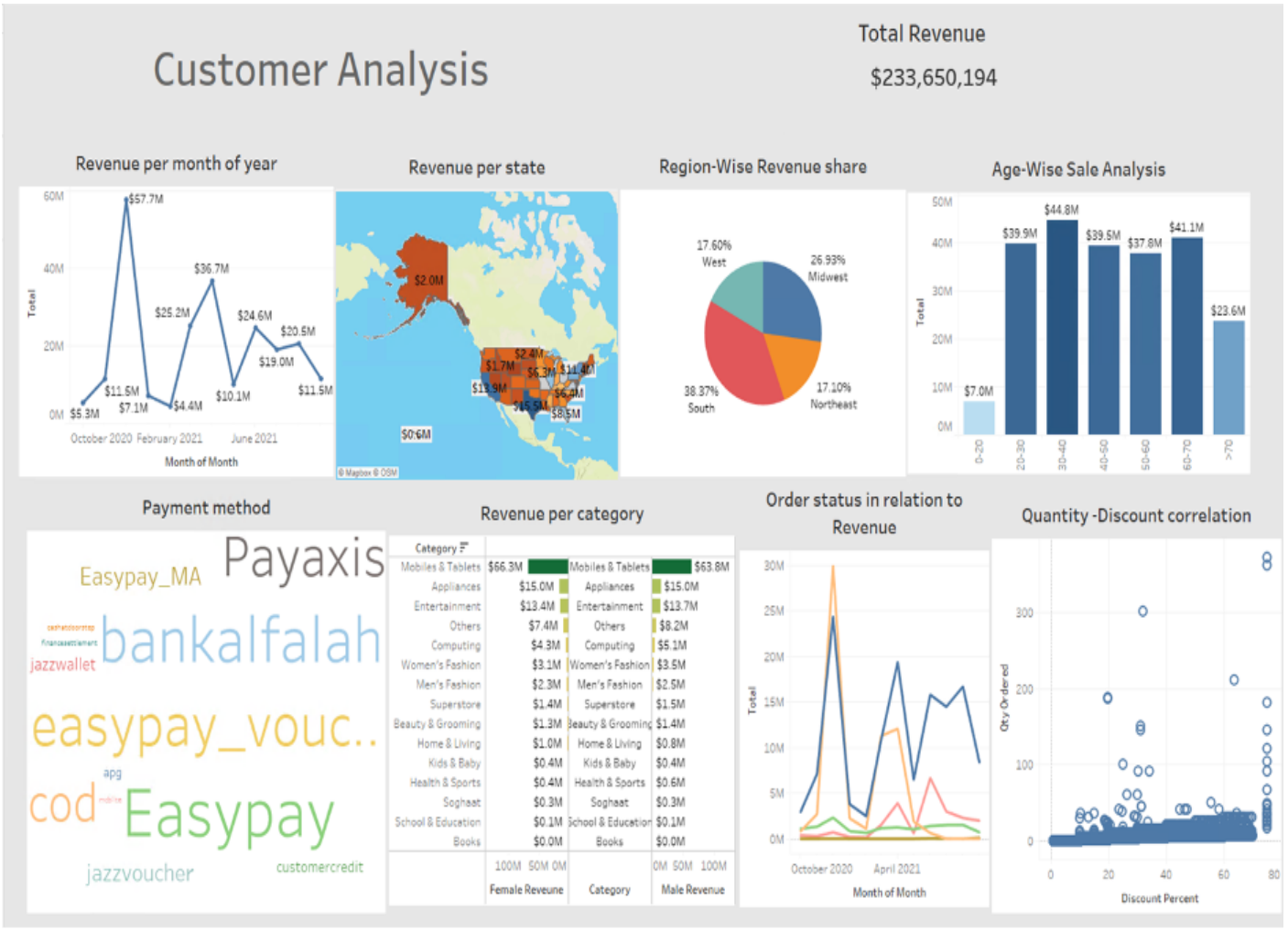
Payment Method:

The word cloud tells us that the terms that are huge and bold are the most regularly utilized. And the ones that are little are the ones that are least popular.

Easy pay is the most popular payment method among customers, followed by bank alfalah.

The least popular payment method is cash at doorstep.

Tableau Dashboard:



Conclusion:

- TX has the greatest revenue of $15.5 million
- The month of December generates the most revenue ($57.7 million), while February generates the least revenue ($4.4 million).
- Discount percentage increases, so does the quantity order.
- The age 30-40 group accounts for the majority of revenue (45 million dollars).
- The South region generated the most revenue (38.37% of all regions), while the Northeast region generated the least (17.10% of all regions).
- Mobile phones and tablets have a significant impact on revenue, with women contributing $66.3 million and men contributing 63 million.
- Customers' preferred payment method is easy pay while Cash at the door is the least popular payment method.

Appendix

Sidhant thakur:

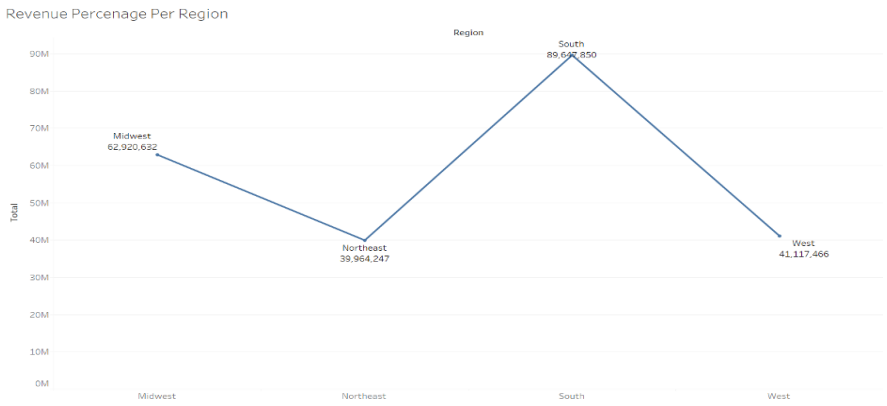
There was a lot to learn in this course as well as in the group project. In the group project, I made some visualizations and discussed with the team whether the plots I made were according to its data, audience, and message.

I ran a quick data analysis, and proposed adding a calculated field called Total, which is ((Price*quantity ordered)-discount).

I had made various charts such as treemap,bar graphs, mosaic plots and found out that line charts are best to analyze the trend. I made a word chart to show various payment methods.
This course exposed me to the world of Data Visualization, before taking the course I didn't have much knowledge about Data Visualization and its implementation in real-world applications. This did not only teach me to use Tableau but also R and its useful libraries like ggplot2. I also gained knowledge of data manipulation and appending two or more datasets in Tableau for creating better visualization. Overall, it was a great experience to take this class and work on this project.

Preetham Sai Kandimalla:

In this project I wanted to visualize the revenue generated from each region of the USA from the sales. The pie chart is a good visualization for this representation as the size of each pie explains the distribution of revenue from each region. I also tried different visualizations like line graph, bar chart and tree map. The line graph shows the trend of how much revenue was generated but it was not clear enough.



The bar graph was easy to compare the different regions by checking the height of the bars, but it was hard to notice the exact difference of revenue.

In this course I learned about different types of data visualizations and different ways data can be visualized based on the type of data and the audience. I also learned the tools like Tableau and RStudio which help in creating the visualizations. The properties of visualizations like shape, size and color are very important to create better visualizations. It was very interesting to learn about the human perspective of understanding the data from the visualizations and how to create better visualizations so that it is easy for the audience to understand the data.

Saransh Thakur :

This project explores a wide range of exploratory analyses to understand and analyze the sales dataset.The data consist of 286369 observations and 36 features. Overall we looked at the revenue to understand the distribution.I have created a geographical plot to better understand the revenue per state and using a divergent color. I built a scatterplot between discount percentage and quantity ordered, which shows a positive link between both variables. Finally, I have revenue based on age, which shows that the 30-40 age group has greater revenue than others.

I used tableau to build the majority of my visualization, the data mining and cleaning process was done in R.

Sumera Fatima Khatoon:

The Final report has a variety of visuals to analyze the sale data. By looking at the data it required some cleaning which was done by using R. I also visualized the revenue per category per gender where it explains how male, and females spend the money in 15 different categories. By taking this class and participating in this group project I was not only introduced but also gained knowledge about Data Visualization and the class also taught me to use Tableau and R. I had a good time by taking this class and working in the group.

Parth Babubhai Patel:

I learned a lot about the value of data visualization in this course.Making sure that the audience can grasp what we're trying to say is the whole objective of visualization. To ensure that the data is comprehended, it is crucial that we select the appropriate type of graph or plot and level of color saturation.I had never used tableau before, and it was a great learning experience. Not just with tableau, but also with R, using professor Eli Brown's notes from my earlier "Fundamentals of Data Science" course made things much simpler.I was able to assist with presentation preparation by updating and adding some ideas.