IE 6600- Computation and Visualization for Analytics

Final Report

Group 5
Analyzing Retail Sales Data: Uncovering Purchasing Trends and Patterns

Heet Manesh Khatri

Sai Varun Kumar Namburi

Harsh Yogeshbhai Panchal

Madupu Srilekha Reddy

Tejaswi Yaramada

Introduction

Customer sales analysis is the process of looking at sales data from a company to learn more about customer patterns, preferences, and behavior. Businesses can use this study to pinpoint their most valued clients, comprehend their purchasing behaviors, and fine-tune their sales and marketing tactics to boost sales and profitability.

Data on customer demographics, past purchases, product preferences, and purchasing behavior are routinely collected and analyzed as part of customer sales analysis. Businesses can segment their consumer base and create customized marketing campaigns by evaluating this data to better meet the demands and preferences of each category.

Overall, customer sales analysis is a valuable tool for businesses looking to improve their sales and marketing strategies and build stronger relationships with their customers. By gaining a deeper understanding of their customers, businesses can make data-driven decisions that drive growth and success.

Our analysis will focus on several key aspects of revenue. Firstly, we will examine the revenue generated per state, allowing us to identify any geographical trends in our earnings. Secondly, we will analyze revenue on a month-by-month basis, providing insight into any seasonal fluctuations in our income. Thirdly, we will investigate how revenue correlates with the age of our customers, helping us to better understand our target demographic. Additionally, we will explore the correlation between quantity and discount percentage to gain a deeper understanding of customer purchasing behavior. Furthermore, we will examine the percentage of revenue generated per region, providing insight into which areas are most lucrative for our business. Lastly, we will analyze revenue per category with a focus on gender, allowing us to identify any gender-based purchasing patterns. By conducting this comprehensive analysis, we aim to gain a holistic understanding of our revenue and make informed decisions about future business strategies.

Data description

The dataset is taken from Git hub. Link to the dataset: https://github.com/stanley-george-joseph/Customer-Analysis-Tableau

The provided data is a transactional dataset capturing online customer purchases. It contains information on order ID, order date, item ID, SKU, quantity ordered, price, payment method, and customer information like name, gender, age, email, etc. Additionally, the data includes information on the order's status, such as received, complete, canceled, and order refunded, which can help to identify potential issues in the purchasing process.

Moreover, the data includes details on the item categories purchased, allowing for the identification of popular items and trends among customers. The dataset can be used for various purposes, such as analyzing customer purchase behavior, identifying the most popular products, and understanding payment preferences.

By analyzing this data, businesses can gain insights into the shopping preferences of their customers, optimize their sales strategy, and improve customer satisfaction. For example, businesses can identify which products sell better during certain times of the year, which payment methods are preferred, and which marketing strategies are most effective.

Methodology:

After checking the data for missing values, we confirmed that no null values were found. Subsequently, we proceeded to remove a few columns.

To remove any unnecessary columns that didn't provide meaningful information, we went through the data set and identified the columns that were irrelevant or redundant and removed them to streamline the data set. Columns such as "bi_st", "ref_num", and "Discount_Percent" seem to be unnecessary, so we dropped them from the dataset.

Reviewed column headers to ensure they were standardized and easy to understand. This step improved the readability of the data set and reduced the risk of confusion when analyzing it.

To standardize the format of data in each column to ensure consistency, we have reviewed each column to ensure that the data was formatted consistently. Converting date fields to a consistent format, the "order_date" field is in the format "dd-mm-yyyy", but the "month" field is in the format "MMM-yy". We have converted all date fields to a consistent format, such as "yyyy-mm-dd".

Converting categorical variables to consistent format: The "category" and "payment_method" columns should be converted to a consistent format, such as lowercase.

To fix data types, some fields such as "qty_ordered" and "age" are numeric but are stored as strings. We converted them to the appropriate data type.

Checked for outliers and errors in the data and handled them accordingly, identified any data points that appeared to be outliers or errors and handled them accordingly. For example, removing or correcting these values could help prevent biases and inaccuracies in the data. By following these steps, we ensured that the data set was clean, consistent, and ready for analysis.

Heatmaps are often used to analyze correlations between variables. By examining the colors of the cells, we can quickly identify which variables are positively or negatively correlated with each other. This information is useful in identifying relationships between variables, as well as in identifying potential outliers or areas of interest in the dataset.

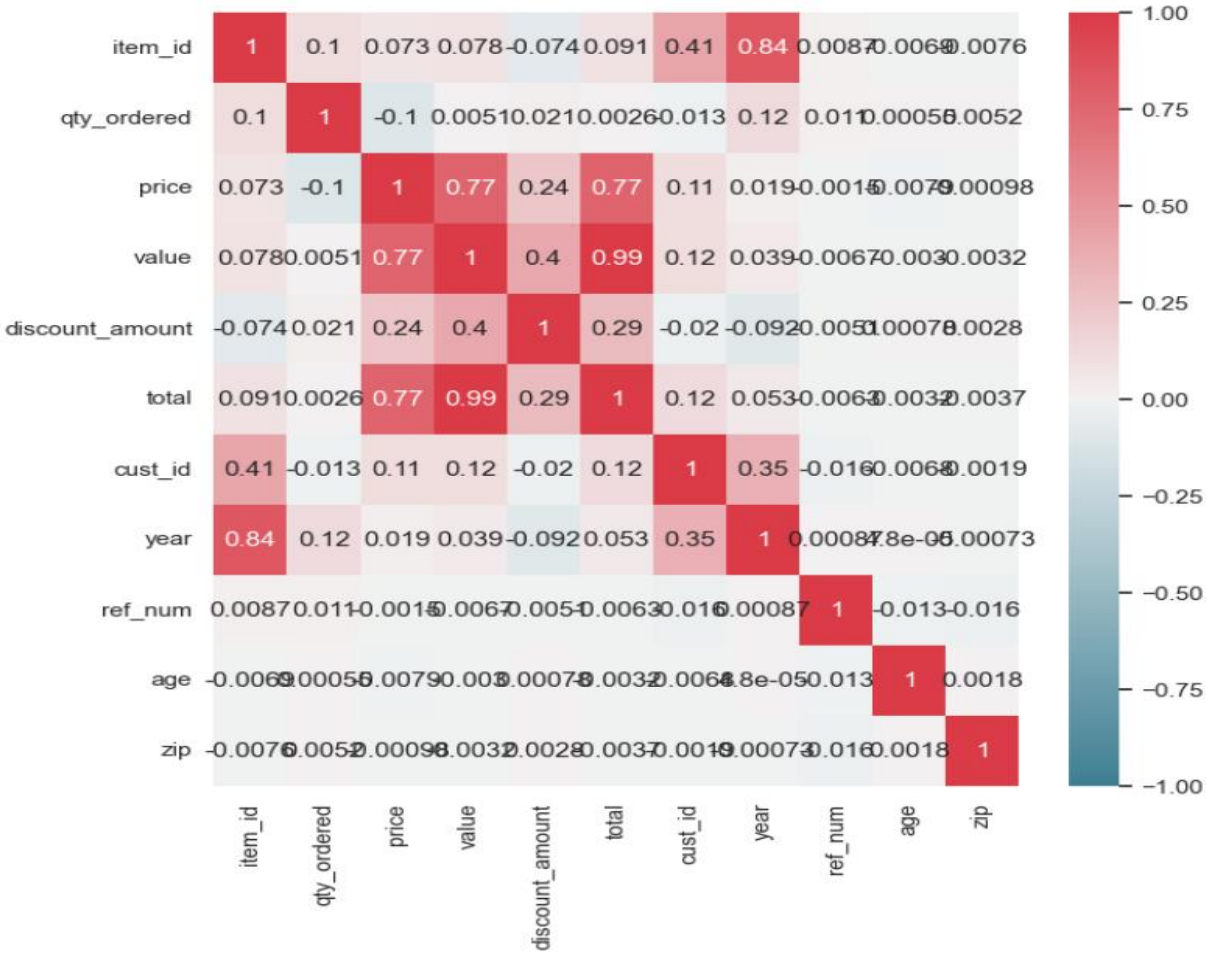

Fig 1: Heat Map

Interactive Dashboard:

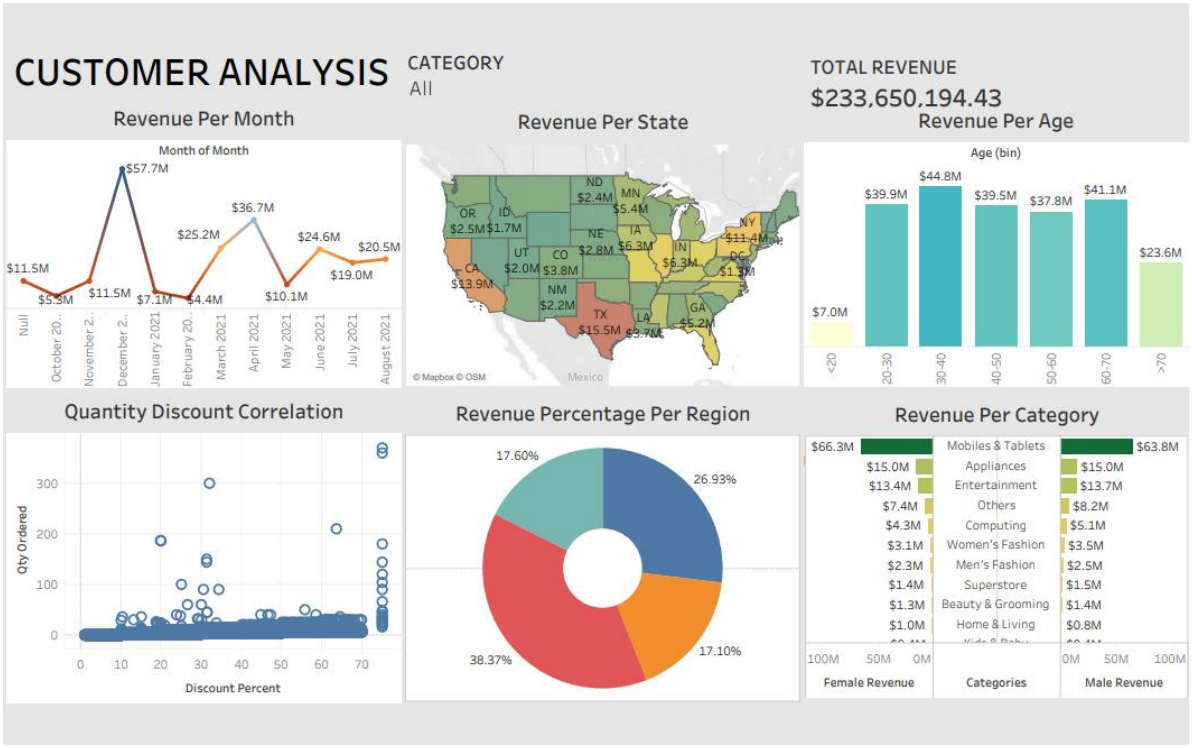

Fig 2: Total Customer Analysis

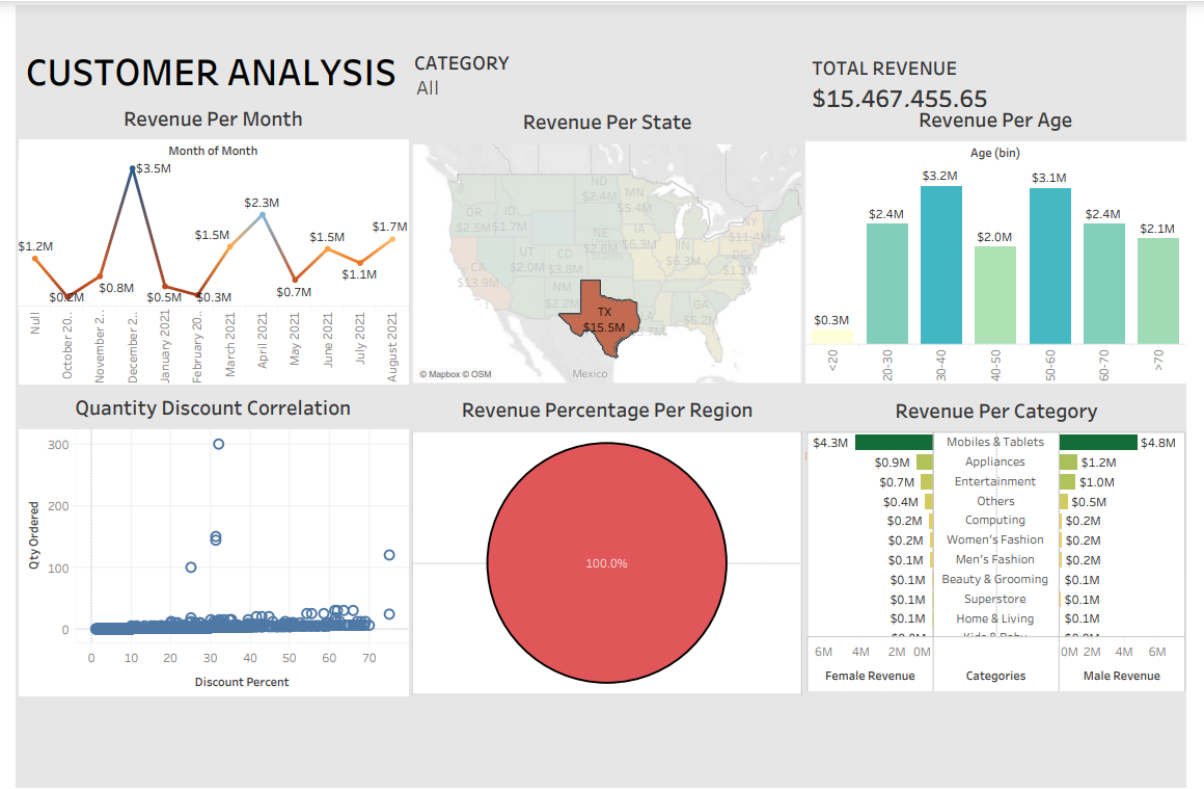

Fig 3: Customer Analysis of Texas

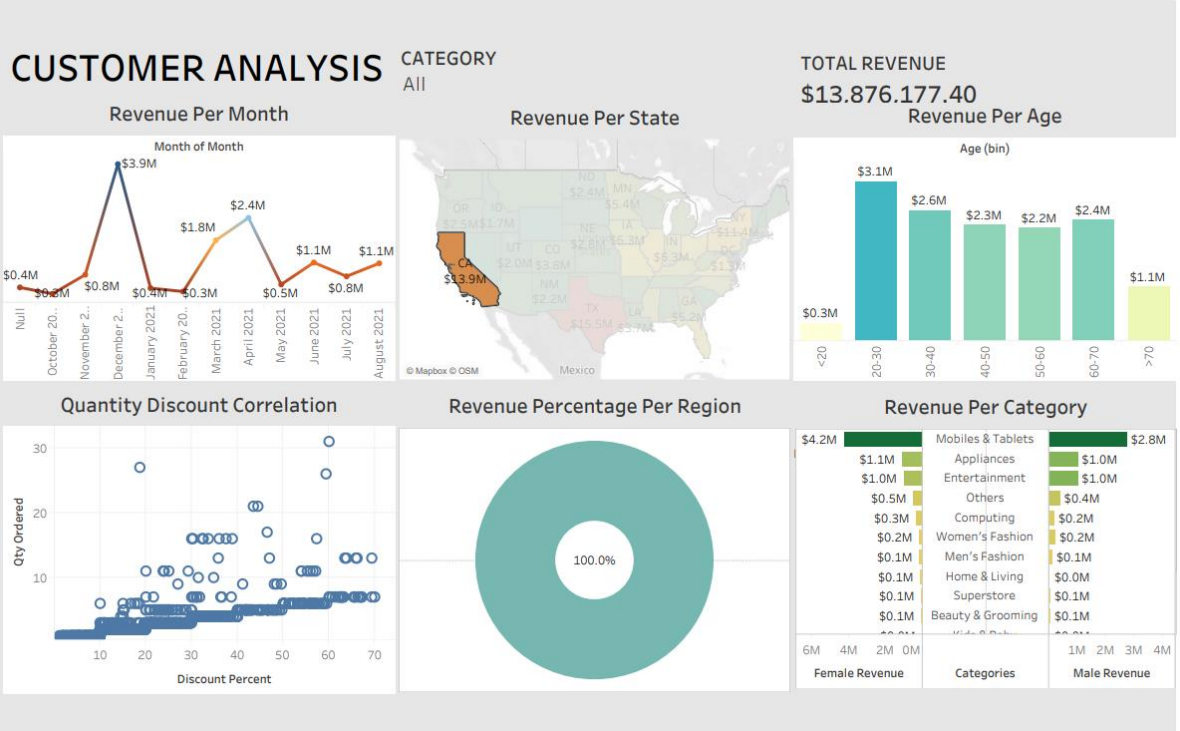

Fig 4: Customer Analysis of California

Insights:

To remove unnecessary columns and produce illuminating visuals, a Python analysis was first conducted. Next, the data was transferred to Tableau to design an interactive dashboard that includes seven visual representations, which offer a comprehensive and valuable perspective on the data.

Our analysis highlights that Texas ranks as the highest revenue-generating state in the US, with California closely following suit. Additionally, our findings indicate that the month of December experiences the highest sales figures, primarily attributed to Christmas. Moreover, the age demographic of 30-40 showcases the highest revenue numbers. We have also generated revenue breakdowns per category, revealing that females generate higher revenue than males when purchasing mobiles and tablets.

Additionally, our analysis indicates that the southern region boasts the highest revenue figures, accounting for 38.37% of total revenue. These are the trends that we have analyzed through our data exploration and visualization efforts.

Future scope

There are several potential future scopes for customer sales analysis. Here are a few of them:

- Predictive analysis: One potential future scope for customer sales analysis is to use predictive modeling techniques to forecast future sales and identify potential opportunities for growth. This can involve using machine learning algorithms to analyze historical data and identify patterns and trends that can be used to make predictions about future sales.

- Customer segmentation: Customer segmentation involves dividing customers into groups based on shared characteristics such as demographics, behavior, and purchasing patterns. Future customer sales analysis could involve using advanced segmentation techniques to identify specific customer groups that are most profitable or that have the highest potential for growth.