Data Visualization Digital Assignment – 2 & 3

Name : Mahesh Jagtap
Reg No.24MCS1017

Title: Retail Data Analysis: Enhancing Customer Insights and Buiness processes

Executive Summary:

In the competitive world of retail, understanding customer behavior is crucial for success. This project highlights the importance of analyzing retail data to optimize business processes and enhance customer satisfaction. By examining customer transactions and preferences, retailers can make informed decisions that not only meet customer expectations but also drive profitability.

The significance of this project lies in its ability to convert raw data into meaningful insights. As the retail market continues to grow and evolve, leveraging data analytics becomes essential for tracking trends, forecasting sales, and improving marketing strategies. Retailers that effectively analyze customer behavior are better positioned to offer tailored experiences and improve overall operational efficiency.

For this analysis, we utilized the following tools:

- **Tableau**: A powerful data visualization tool that helped us create interactive dashboards and visual representations of customer data, making it easier to identify trends and patterns.
- **Excel**: Used for data cleaning and preparation, ensuring that the datasets were accurate and ready for analysis.

Introduction:

In an increasingly competitive retail market, understanding customer behavior has become essential for businesses striving to achieve long-term success. Retailers must not only meet the needs of their customers but also anticipate their preferences and trends. This project aims to analyze retail data to uncover insights into customer transactions and purchasing behavior, which can ultimately help improve service delivery and drive profitability.The analysis focuses on three main datasets: customer demographics, product categories, and transaction records. By examining these datasets, we can identify patterns in consumer behavior, such as which products are most popular among different customer segments and how demographic factors influence purchasing decisions.

Through this project, we seek to provide actionable insights that will allow retailers to optimize their strategies, enhance customer satisfaction, and increase sales. The findings will be presented using data visualization tools, making it easier for stakeholders to understand and apply the results effectively. This analysis is crucial for retailers looking to maintain a competitive edge and foster a customer-centric approach in their operations.

Problem Statement:

To analyze day-to-day transactions and track customers across various locations, focusing on their purchases and returns across different product categories. This analysis aims to uncover insights into customer behavior, identify popular product categories, and evaluate the performance of different store types to enhance service delivery and optimize business strategies.

Methodology :

1. **Data Source**

The analysis is based on three primary datasets:

Customer Dataset: Contains demographic information, including customer ID, date of birth, gender, and city code.
Product Dataset: Details the product hierarchy, including product category codes and descriptions.
Transaction Dataset: Records individual transactions, including transaction ID, customer ID, transaction date, product details, quantity, rate, tax, total amount, and store type.

2. **Data Preprocessing**

- **Data Cleaning**: Remove any duplicates, handle missing values, and ensure data consistency across the three datasets. For example, ensuring that customer IDs match in the transactions dataset and customer dataset.
- **Data Transformation**: Convert date formats and numerical values as necessary for analysis. This includes changing date strings to date objects and ensuring monetary values are in the correct format.

3. **Data Processing**

- **Data Merging**: Combine the three datasets into a unified dataset, Customer_Final, using appropriate joins (e.g., inner join) to ensure all relevant information is retained for customers with transaction history.
- **Feature Engineering**: Create new variables where needed, such as calculating customer age from the date of birth and categorizing transactions into positive (purchases) and negative (returns).

Results and Discussion

This section presents a comprehensive analysis of data, addressing key questions related to customer behavior, product preferences, and transaction trends. Various visualizations and statistical measures have been employed to provide clear insights. The following charts and analyses answer the specific questions posed in this study:

1. Preparing a summary report for the merged data set.

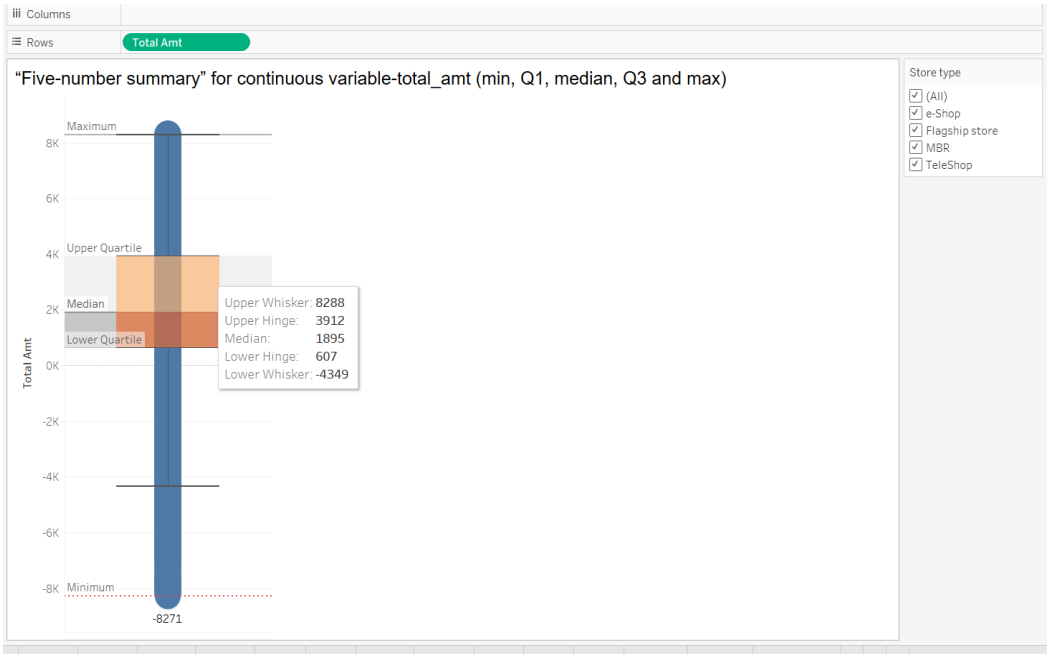

Fig 1. **box plot** represent a "Five-number summary" for a continuous variable, total amount. The plot provides a quick visual summary of the data's distribution, highlighting its central tendency and variability.

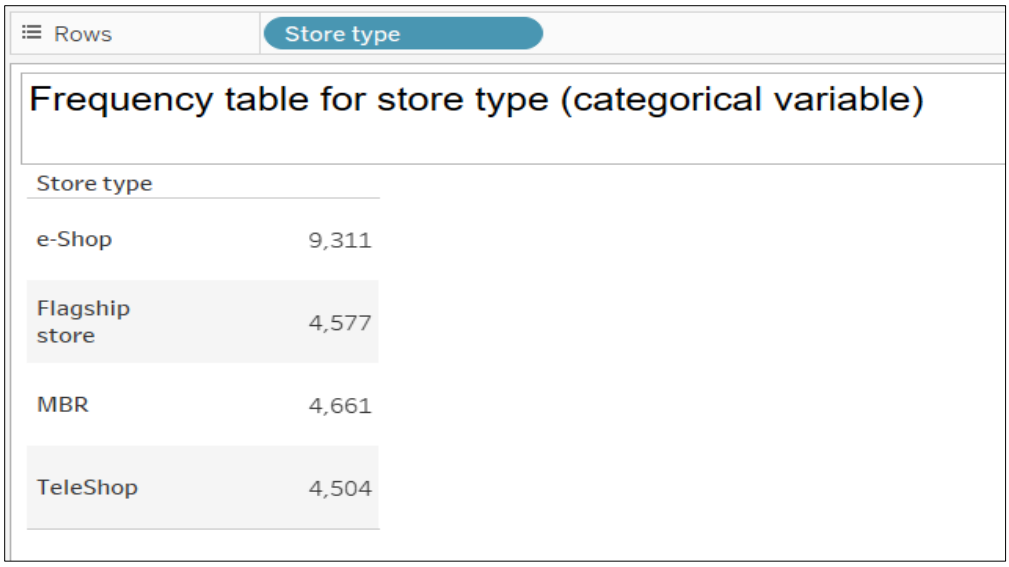

Fig 2. **frequency table** for different store types. It lists the number of occurrences for each type of store. This helps to compare the popularity or usage of different store types within the dataset.

2. Generating histograms for all continuous variables and frequency bars for categorical variables.

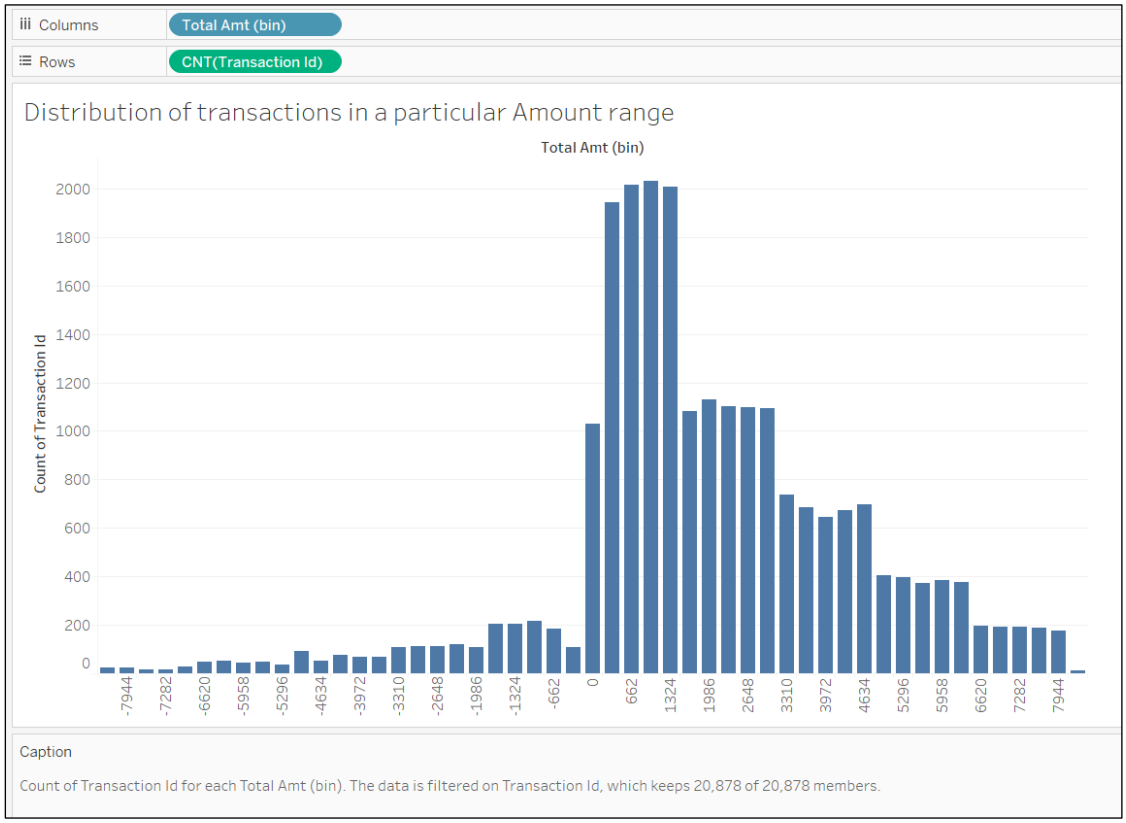

Fig 3. Histogram shows how transactions are distributed across various amount ranges, which can be useful for analyzing financial data or consumer behavior.

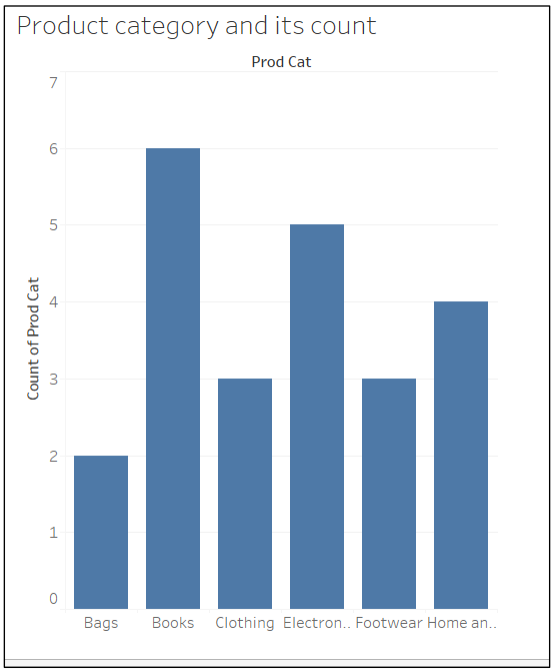

Fig 4. Product category and its count

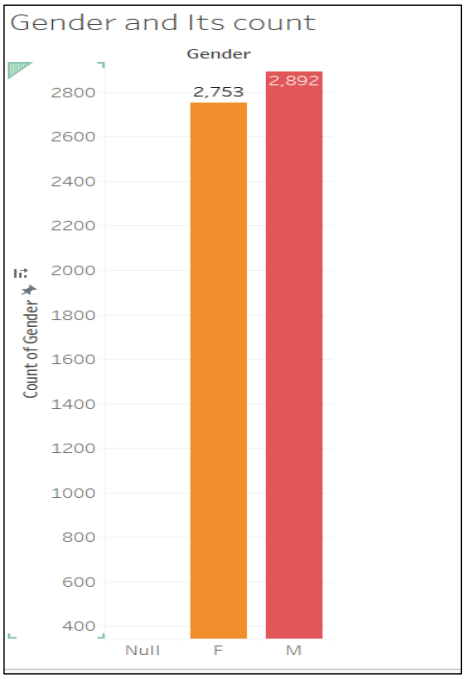

fig 5. Gender vs count

3.1 Time period of the available transaction data

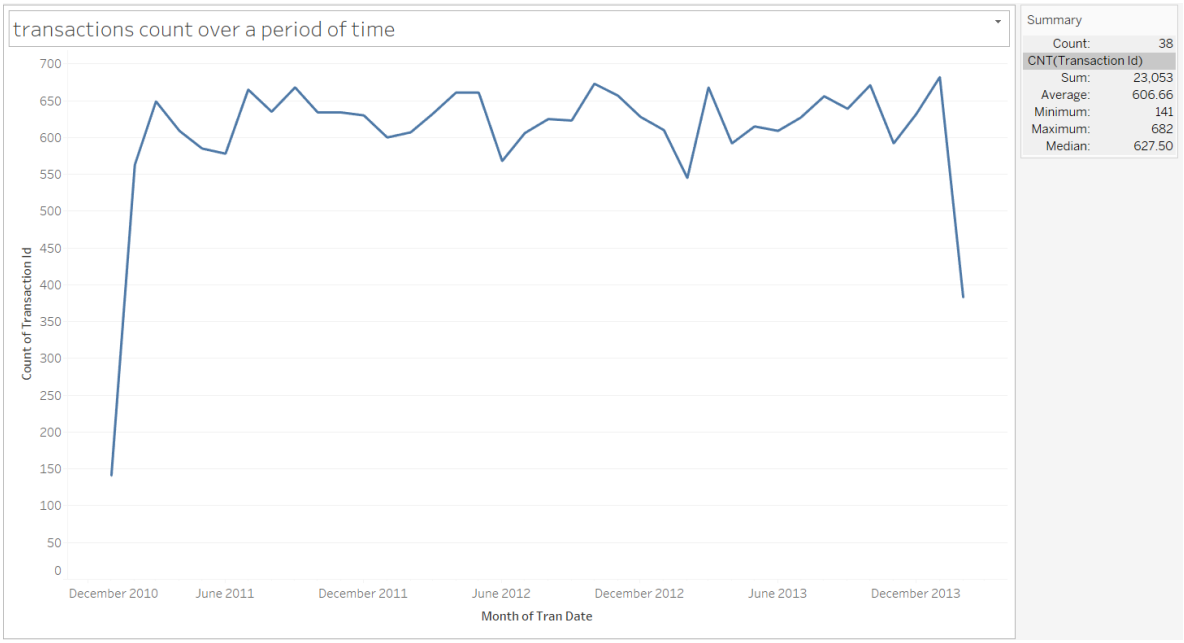

Fig 6. Month vs No. of transactions

3.2 Count of transactions where the total amount of transaction was negative

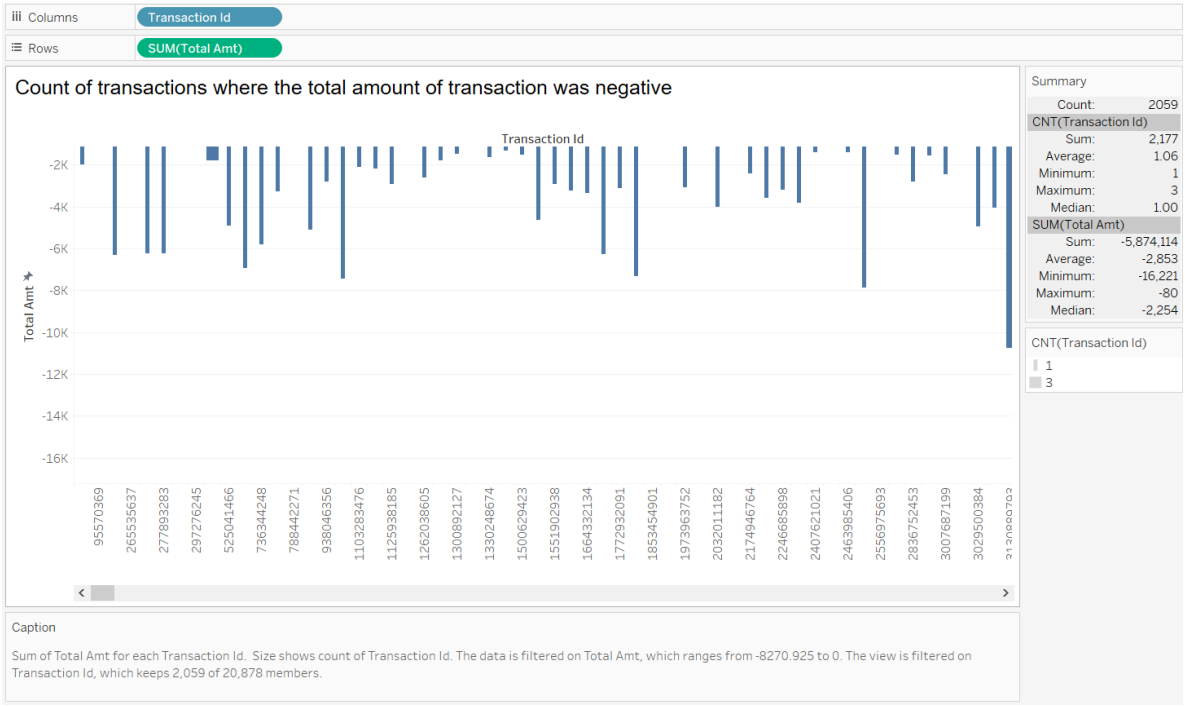

Fig 7. Transaction id vs total amt(Negative)

4. Analysis of product categories that are more popular among females vs male customers.

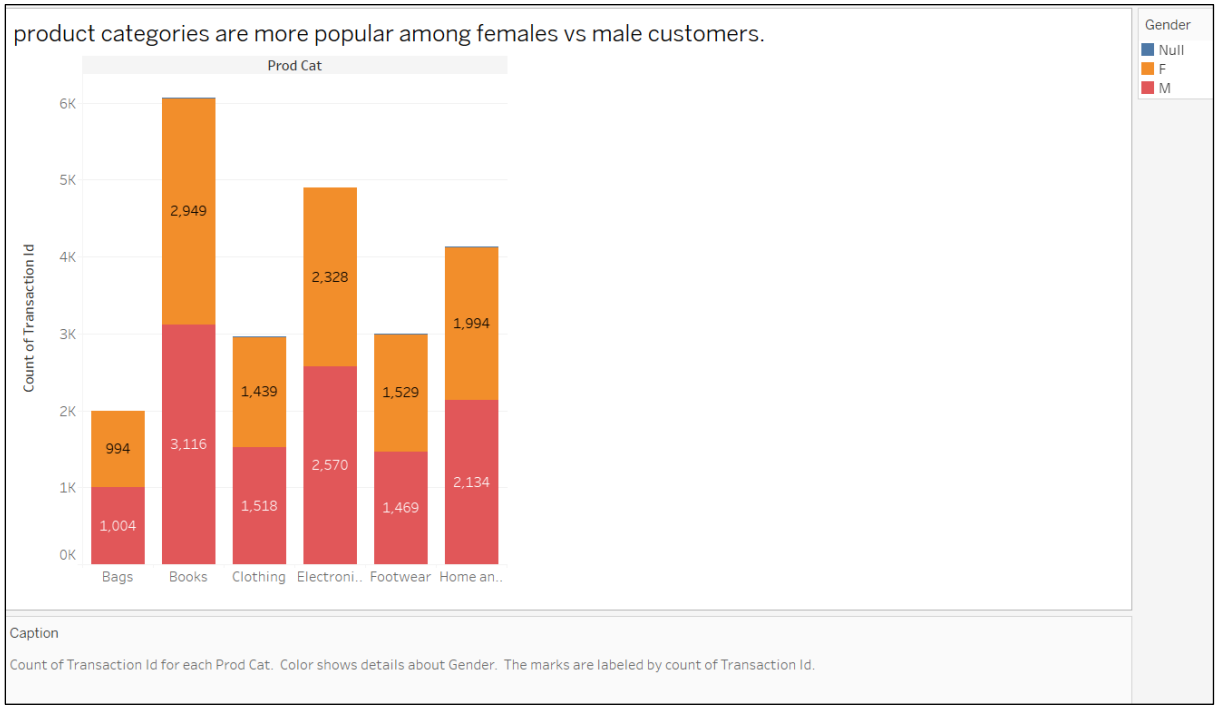

Fig 8. Product Categories vs transactions

5. **Which City code has the maximum customers and what was the percentage of customers from that city?**

➔ City code 3 has the the maximum customers (595) constituting 10.54% of total customers.

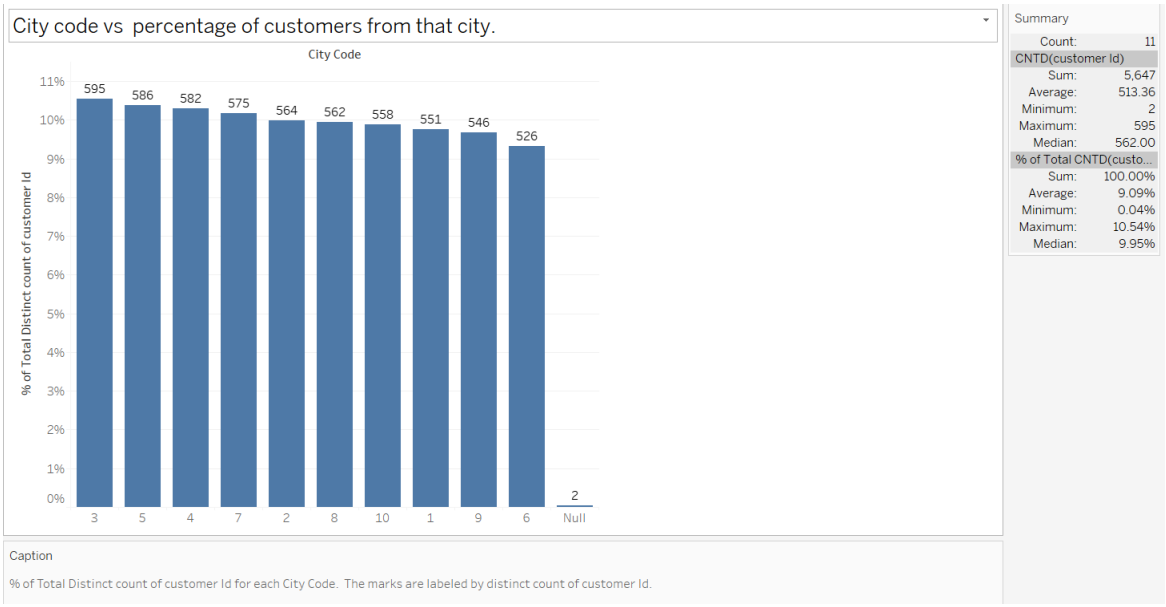

Fig 9. City code vs percentage of customers from that city.

6. **Which store type sells the maximum products by value and by quantity?**
   - ➔ E shop sells the maximum products (22763) with total amount of around 19.8 Million.

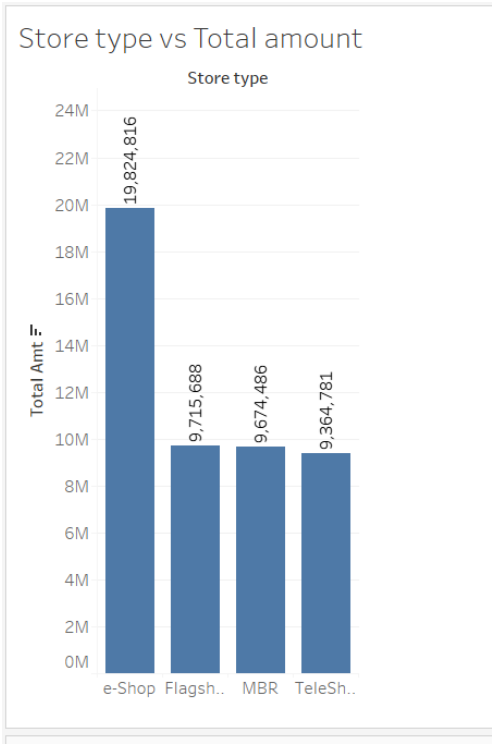

Fig 10

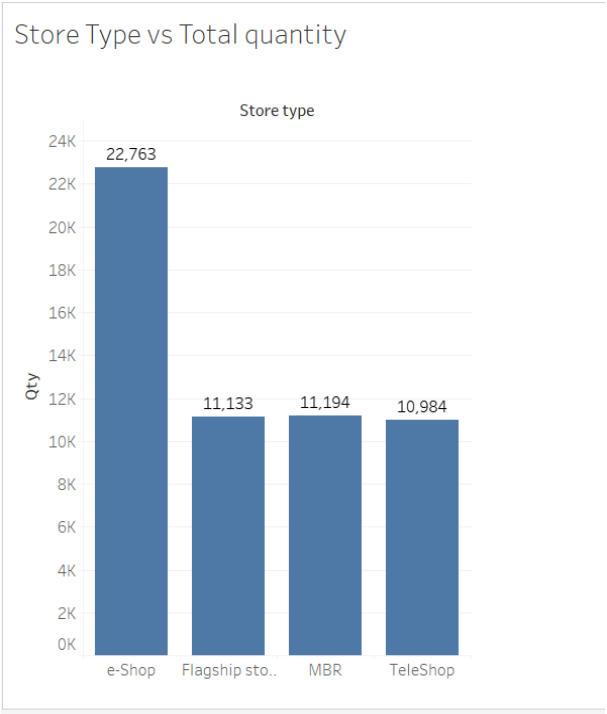

fig 11

7. **What was the total amount earned from the "Electronics" and "Clothing" categories from Flagship Stores?**
   - ➔ 3,409,559

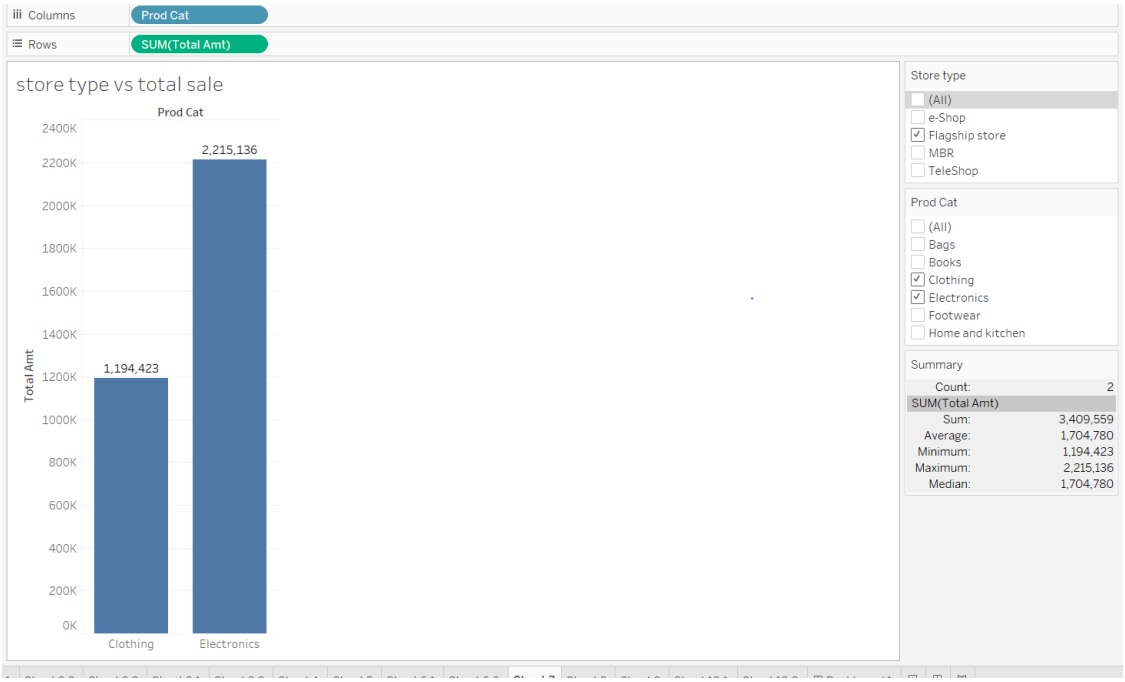

Fig 12. Store type vs Total sale

8. What was the total amount earned from "Male" customers under the "Electronics" category?

➔ 5,703,109

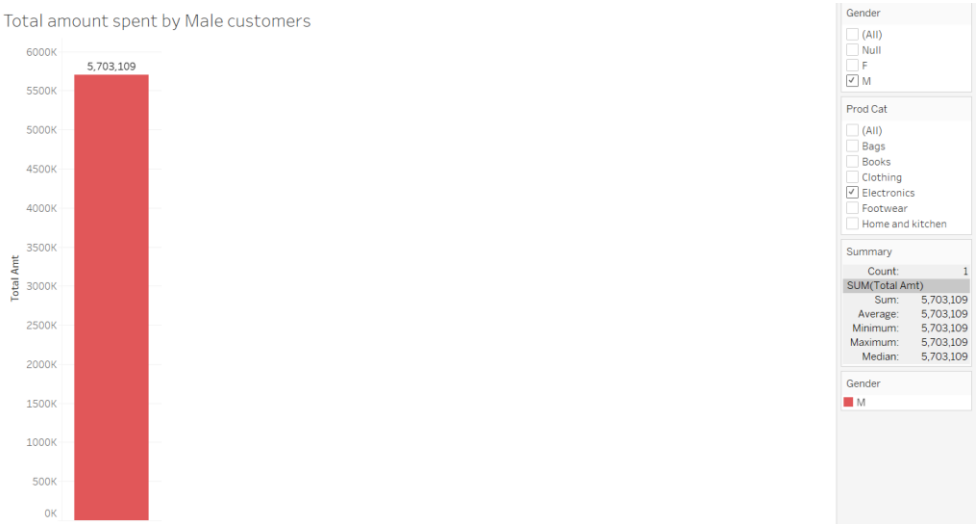

Fig 13. Total amount spent by Male customers

9. How many customers have more than 10 unique transactions, after removing all transactions which have any negative amounts?

➔ **6**

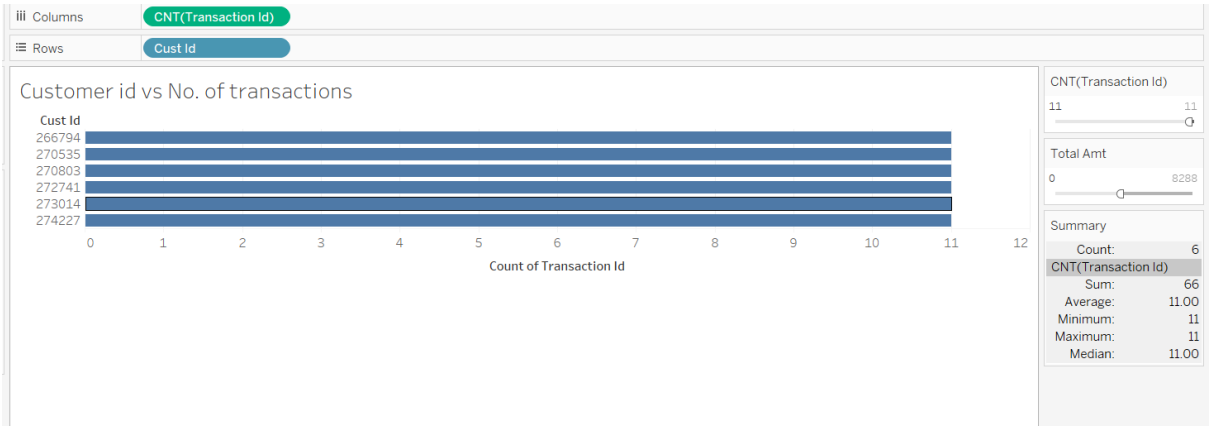

Fig 14.Customer id vs No of transactions

10. For all customers aged between 25 - 35, find out:

a. What was the total amount spent for "Electronics" and "Books" product categories ?

➔ 11,358,841

Fig 15. Product cat vs Total amt

b. What was the total amount spent by these customers between 1st Jan, 2014 to 1st Mar, 2014?

➔ **2,235,643**

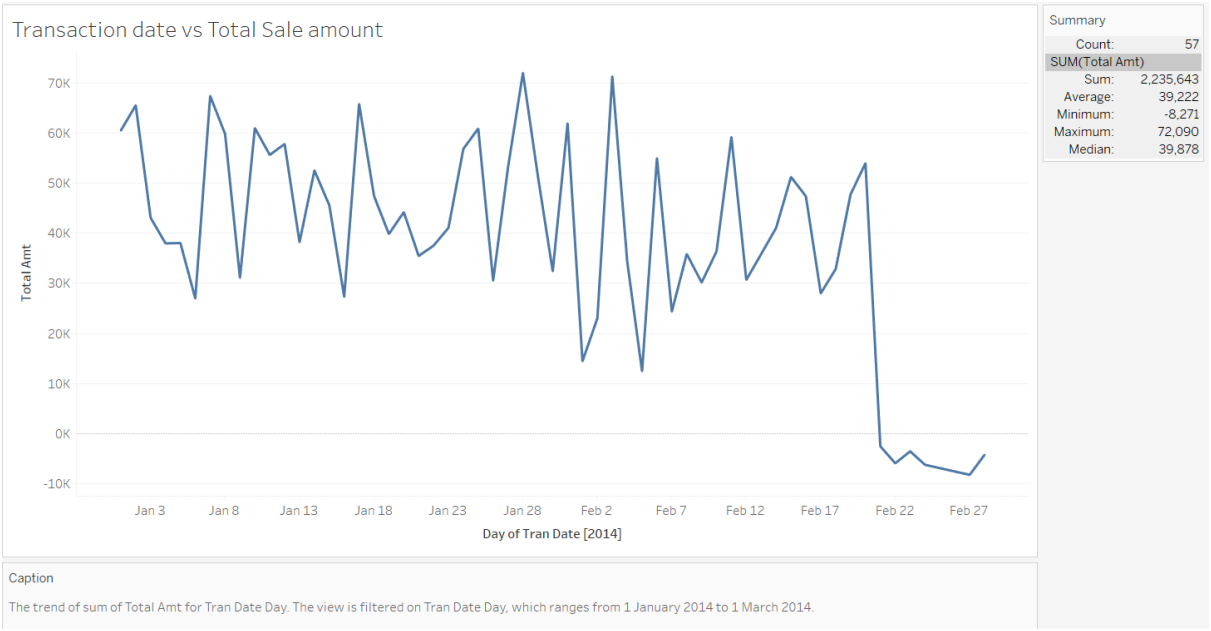

Fig 16. Transaction date vs Total Amt

Final Dashboard:

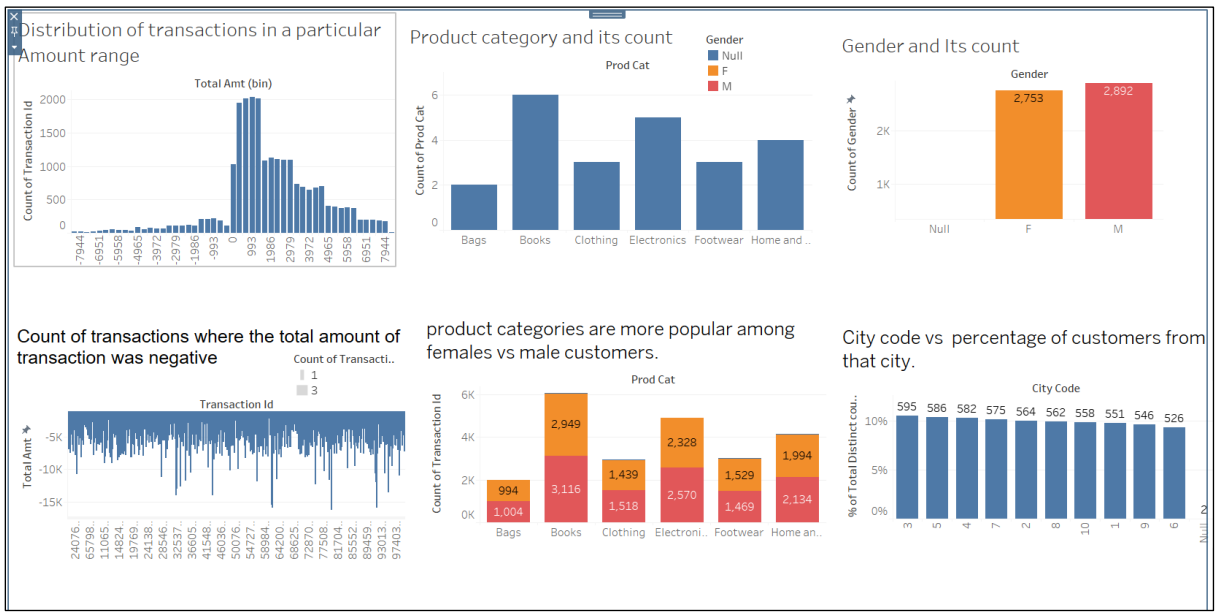

Fig 17

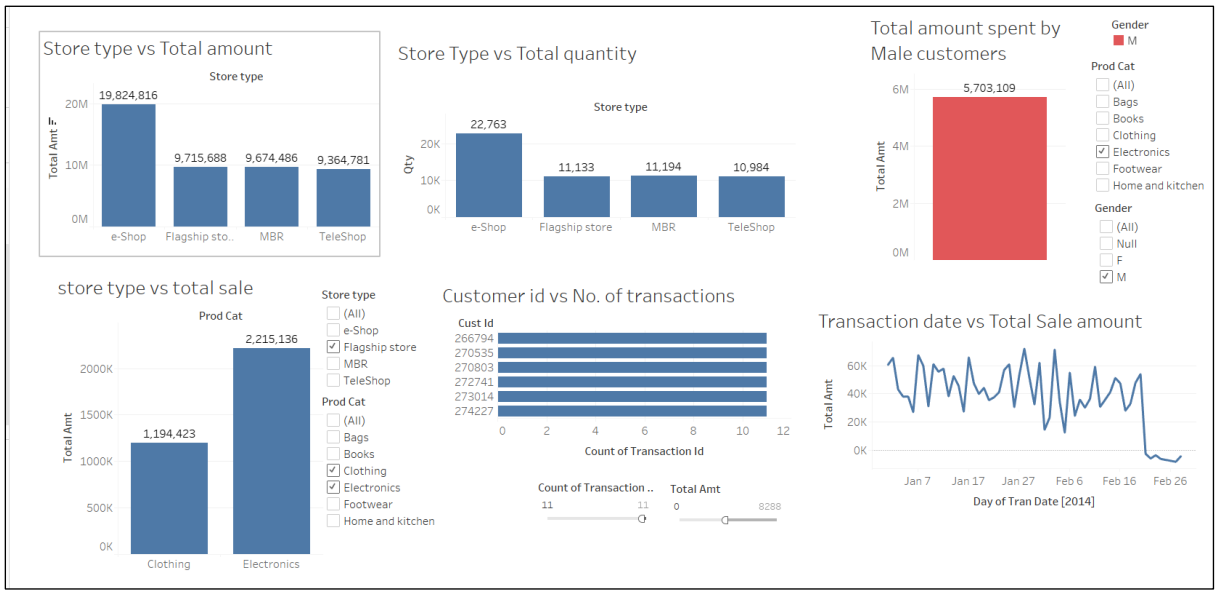

Fig 18

CONCLUSION:

This analysis offers key insights into customer behavior, product preferences, and store performance, enabling the retailer to make informed, data-driven decisions. By identifying high-value regions, popular products by gender, and effective sales channels, the retailer can optimize marketing, inventory, and resource allocation. Overall, these insights empower the business to enhance customer satisfaction and drive profitability in a competitive market.