

MGMT586 Final Project

Members: Meghan Ue Yan Lam, Chi Wu

1. Introduction to the dataset

We utilized the "[Customer Purchase Behavior - Electronic Sales Data](#)" dataset from Kaggle, which contains detailed sales transaction information from an electronics vendor, spanning from 2023 to 2024. The dataset provides a comprehensive overview of customer demographics, product categories, and various purchase behaviors. Our focus is on analyzing how demographic factors and purchase behaviors—such as add-on purchases, loyalty membership, shipping type, and purchase date—influence customer ratings and total spending within the electronic sales data. This report includes exploratory data analysis, hypothesis testing, and correlation analysis to uncover these relationships, aiming to better inform customer relationship management (CRM), marketing, and sales strategies.

Here are the key variables in this dataset that are used in this report:

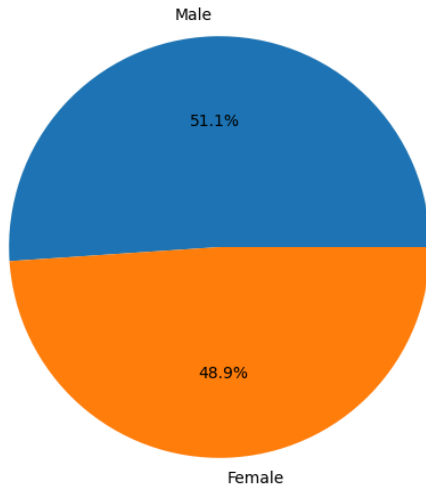
1. Customer ID: An unique identifier for each customer.
2. Age: Age of the customer
3. Gender: Gender of the customer (Male or Female)
4. Loyalty Member: (Yes/No) (Values change by time, indicating membership cancellation and sign up)
5. Product Type: Type of electronic product sold (e.g., Smartphone, Laptop, Tablet)
6. Rating: Customer rating of the product (1-5 stars) (Should have no Null Ratings)
7. Order Status: Status of the order (Completed, Cancelled)
8. Payment Method: Method used for payment (e.g., Cash, Credit Card, Paypal)
9. Total Price: Total price of the transaction, which is total sales
10. Unit Price: Price per unit of the product
11. Quantity: Number of units purchased
12. Purchase Date: Date of the purchase
13. Shipping Type: Type of shipping chosen (e.g., Standard, Overnight, Express)

2. Exploratory Data Analysis

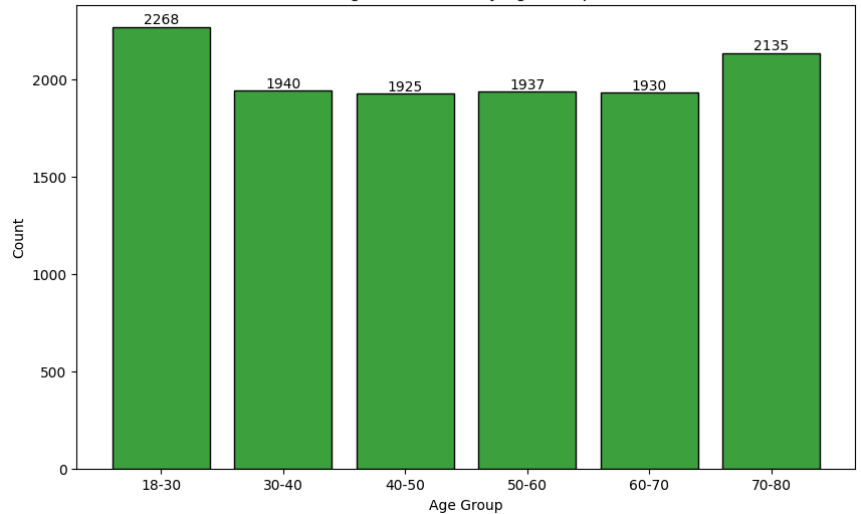
a. Demographic distribution

This dataset shows an equal distribution across gender and age groups. The graphs indicate that the number of data points is very similar for each gender and age group.

Gender distribution among unique customers



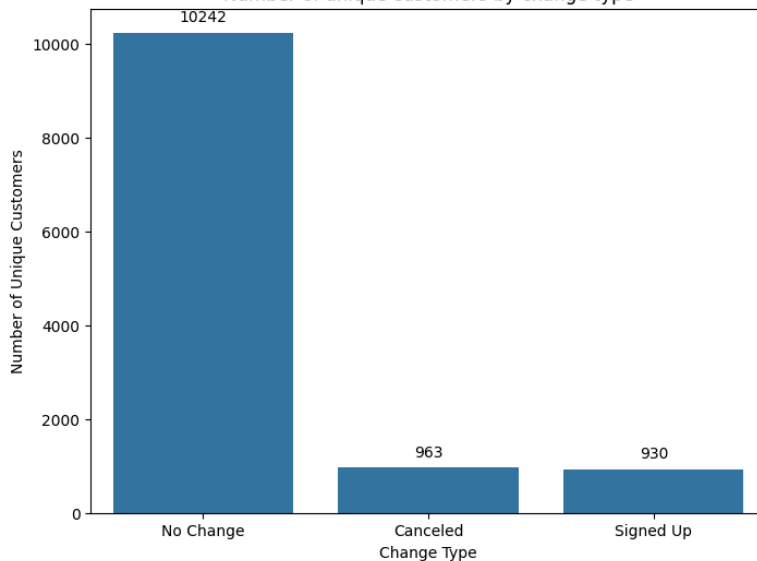
Age Distribution by Age Group



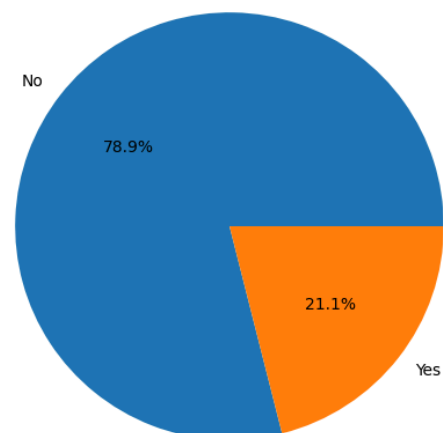
b. Loyalty Membership changes over the period

The analysis of membership type changes reveals that only 21% of customers now are current loyalty members, as shown in the pie chart, suggesting that the loyalty program is not effectively converting or retaining customers. The bar plot indicates that 10,242 out of 12,135 unique customers (about 84%) experienced "No Change" in their membership status, which means most customers either stayed non-members or remained members without change. The number of cancellations (963) slightly exceeds sign-ups (930), indicating a net negative shift and low overall membership growth.

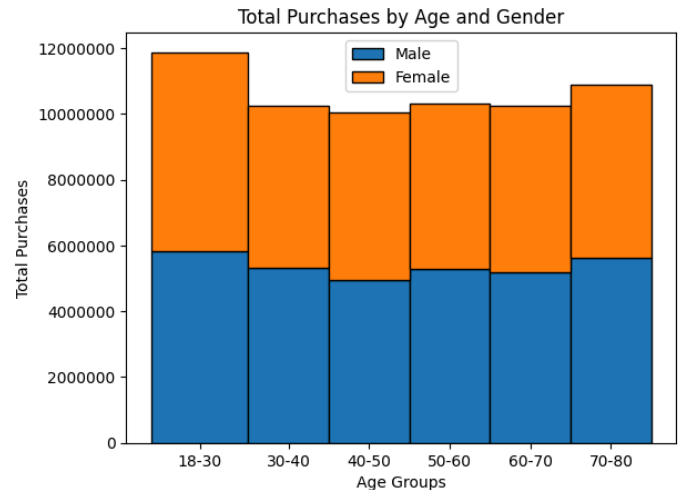
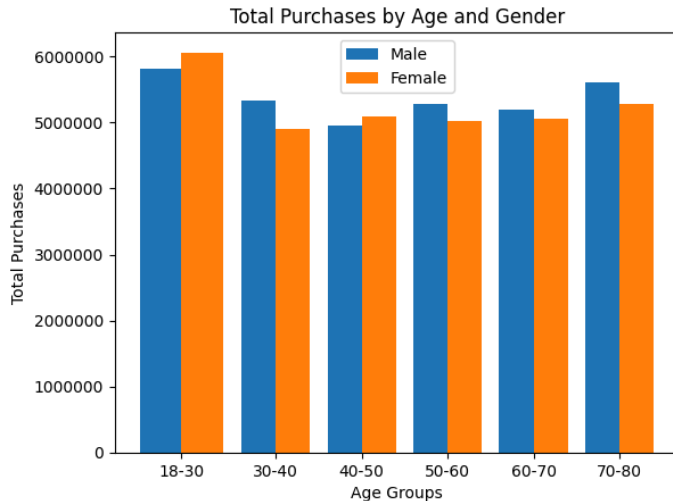
Number of unique customers by change type



Current Distribution of Loyalty Member Status



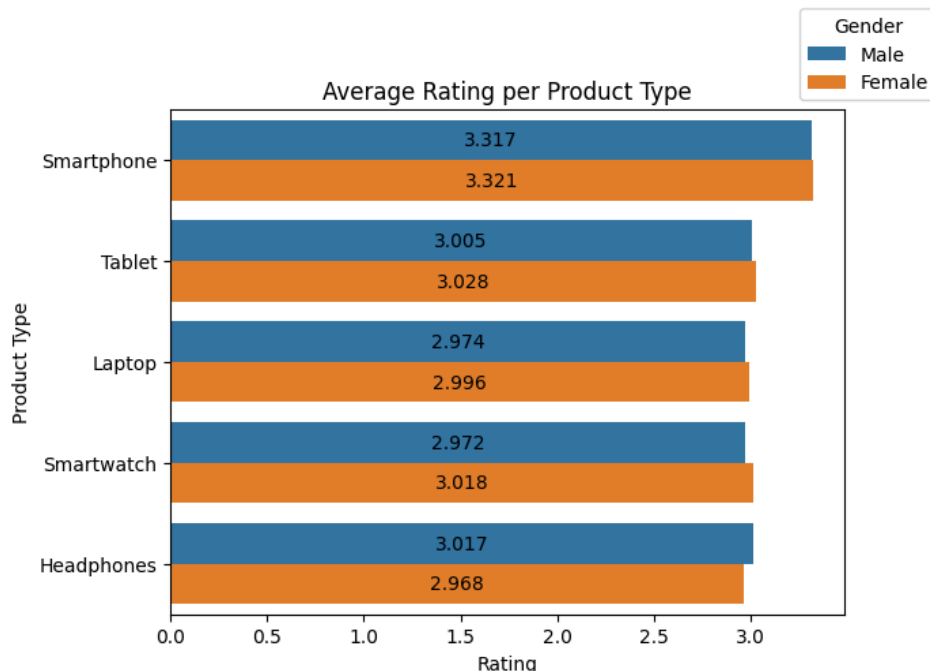
c. Total Purchase by Gender and Age



Based on the graphs, the 18-30 age group is the most represented, likely contributing to their substantial purchasing power. The 70-80 age group also shows a significant purchase volume, driven by a relatively large population size. Across most age groups, males generally have higher total purchases than females. However, in the 18-30 age group, females slightly outspend males, highlighting young females as a key consumer base.

The analysis reveals that both younger and older demographics are key market segments, with significant purchasing power across age groups. While male customers generally drive higher total purchases, young female customers also demonstrate strong purchasing potential, particularly in the 18-30 age group.

d. Average Rating by Product Type and Gender

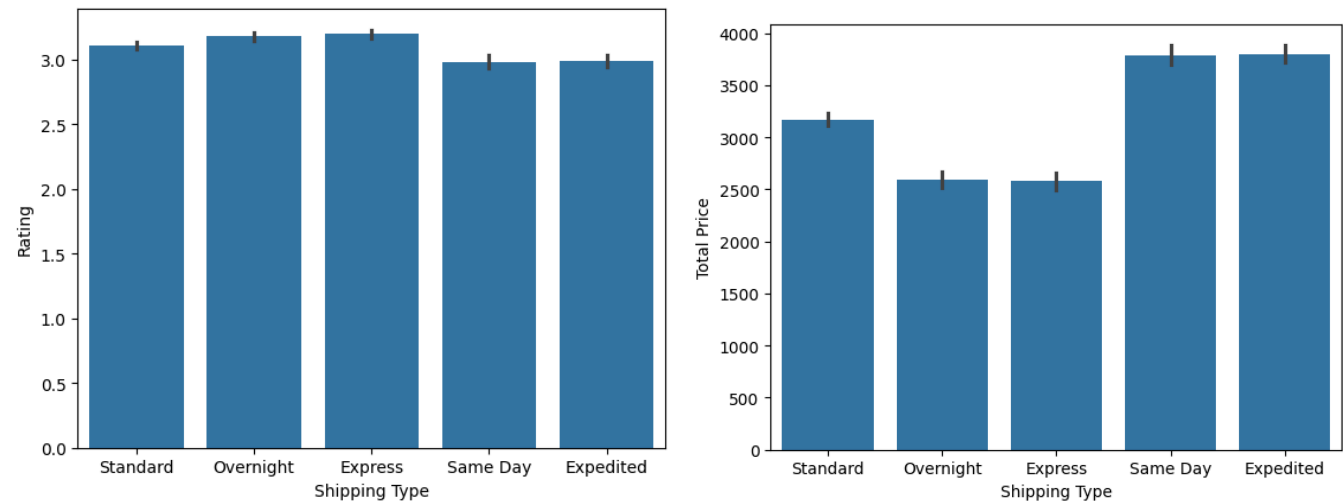


The graph results indicate that most product ratings fall between 2.9 and 3.4, which is close to the median rating of 3.0. While average ratings are similar across product types, smartphones stand out with the highest overall rating, whereas laptops receive the lowest ratings, with both genders rating them below 3.0.

The low ratings for laptops may point to specific issues causing general dissatisfaction among customers. Additionally, while the gender differences are subtle, females tend to rate most product types slightly higher than males, except for headphones. This suggests that females may be marginally more satisfied with the products, which could be an opportunity for targeted marketing and product improvement strategies.

e. Rating and Total Spendings Across Shipping Type

The fbar chart on the left shows that customers ratings are consistent across all shipping types, averaging around 3.0 on a scale of 1-5, indicating the choice of shipping has little impact of satisfaction. Whilst the chart on the right reveals that Same Day and Expedited shipping have the highest total prices, while Overnight and Express shipping have lower sales.

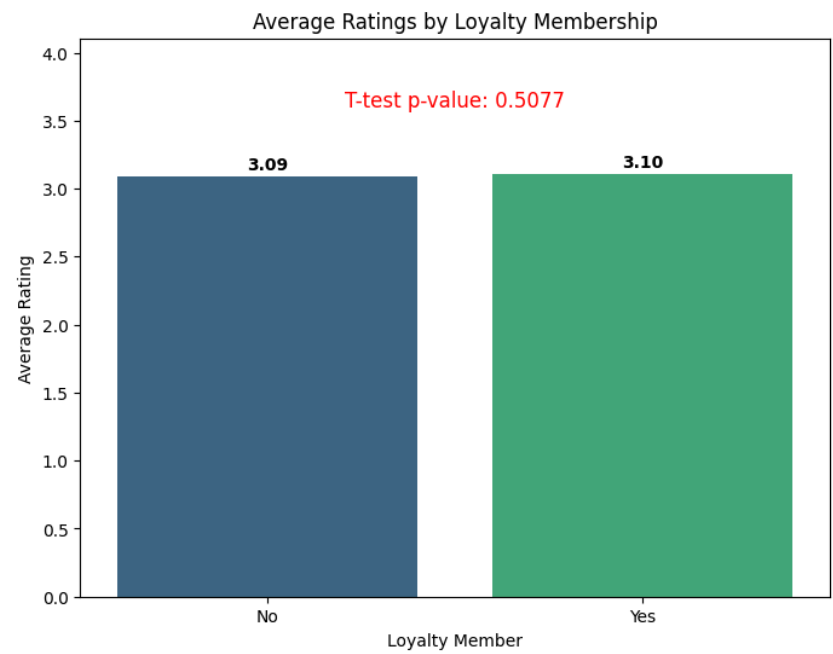


3. Hypothesis Testing

Our previous EDA indicated no significant difference in product ratings between genders. Given this, we decided to investigate whether loyalty membership might have a notable effect on ratings, as customers in the program are expected to have higher satisfaction with the products, making them more likely to join the program.

To test this, we conducted a hypothesis test. The null hypothesis (H0) states that loyalty membership does not affect the average rating, while the alternative hypothesis (H1) suggests that it does. The results, visualized below, show that we cannot reject the null hypothesis. This indicates that being part of the loyalty program does not significantly influence customers' product ratings.

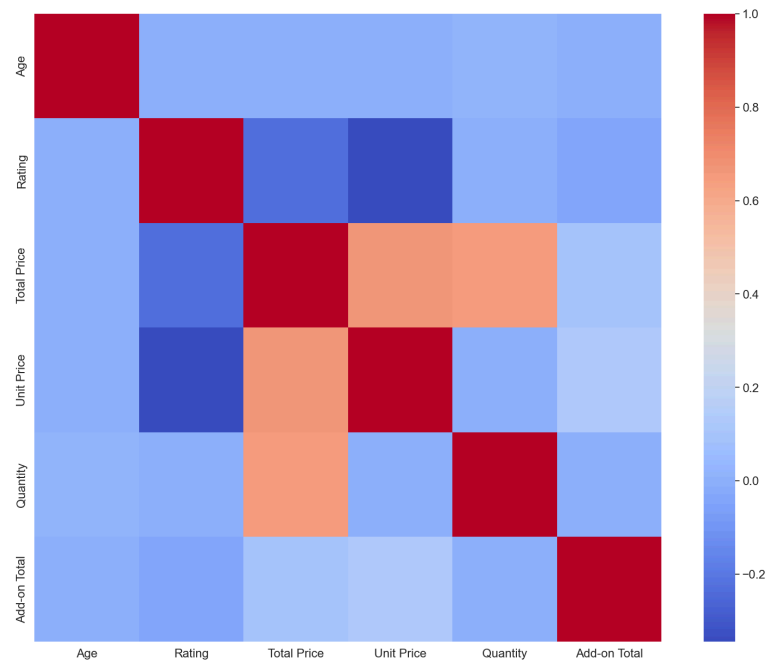
This finding suggests that customers may not join the loyalty program due to brand loyalty or preference toward the vendor.



4. Correlation (Bivariate Analysis)

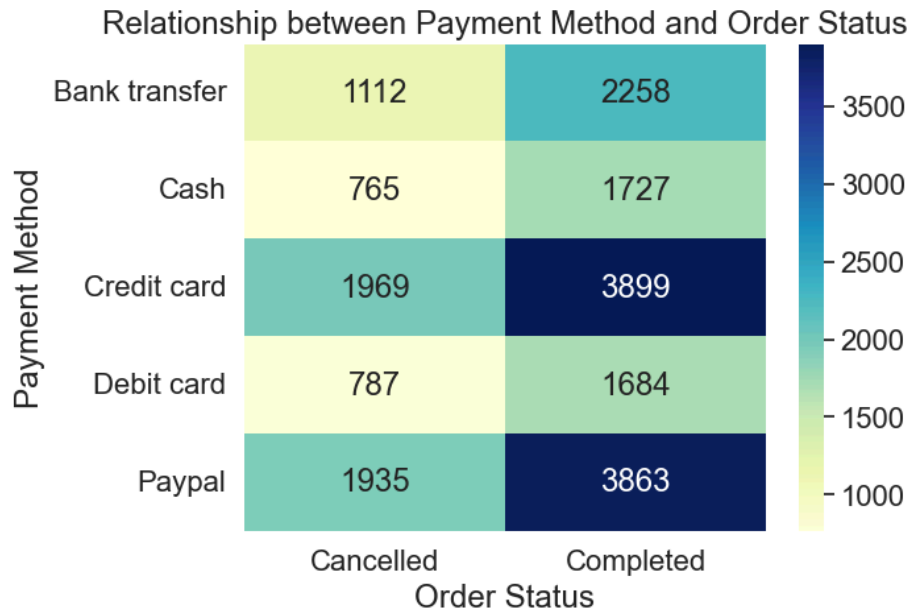
a. Correlations between Rating and other numeric variables (Pearson’s Correlation)

The correlation analysis reveals that higher-priced items and greater total spending are associated with slightly lower customer ratings. This trend suggests that customers may have higher expectations or be more critical of expensive products. Additionally, demographic factors such as customer age and behaviors like add-on purchases show weak correlations with customer ratings and total spending, indicating that these elements have limited influence on overall sales performance and customer satisfaction.



b. Relationship between Payment Method and Order Status (Contingency Table)

The heatmap shows that 'Credit card' and 'Paypal' are the most popular payment methods for both completed and canceled orders. However, they also have the highest cancellation counts (1969 and 1935), suggesting potential issues with these payment options. In contrast, 'Bank transfer', 'Cash', and 'Debit card' have fewer cancellations relative to their completed transactions, indicating potentially more secure options.



5. Business Insights and Recommendations

a. Revamp the loyalty program for greater engagement

The low acquisition rate and retention in the loyalty program suggest the need for improvements in both benefits and messaging. To enhance program effectiveness, introduce targeted campaigns for sign-ups and better incentives for existing members. Conduct quick surveys or research to identify reasons for cancellations and design more value-driven benefits to align with customer needs and boost membership growth.

b. Tailor marketing strategies by demographics and product preferences

Demographic analysis indicates key market segments in younger (18-30) and older (70-80) age groups, with young females showing significant purchasing potential. Vendors should create age- and gender-specific marketing strategies to effectively target these groups. Additionally, the low ratings for laptops and nuanced differences in product ratings by gender present an opportunity for product-specific improvements and targeted marketing to enhance satisfaction and cater to customer preferences.

c. Optimize shipping and payment strategies

Given that faster shipping options ('Same Day', 'Expedited') do not enhance satisfaction significantly and options like 'Overnight' and 'Express' are cost-intensive with low returns, the focus should shift to cost-effective

options that meet customer expectations. Moreover, high cancellation rates associated with 'Credit card' and 'Paypal' highlight the need to resolve payment inefficiencies. Encouraging more stable payment methods like 'Bank transfer' and 'Cash' through targeted incentives could reduce cancellations and improve the transaction completion rate.

d. Improve perceived value for premium-priced products

The correlation analysis shows that higher-priced products are associated with slightly lower ratings, indicating higher customer expectations. To mitigate this, manage expectations by improving product information, after-sales support, and offering value-added services to align with the perceived value of premium items. Since factors like age and add-on purchases do not significantly affect satisfaction, optimizing pricing and the overall purchasing experience is more effective than personalized marketing or add-on strategies.