Customer Preference Analytics

Anant Patel - 0866771

Over the past decade, there has been a transformation in global retail with online sales contributing 20% of total global retail sales. According to Statista (2023) the online sales for the fashion market were valued at $759.5 billion in 2022 and are expected to increase to $1.2 trillion by 2027 showing a significant shift of consumers towards digital platforms and changes in the purchasing behavior of consumers. On one hand, new clothing trends are constantly emerging, largely influenced by online influencers and popular film stars. These trends are quickly spread to people through various online social media platforms, movies, and advertisements Giri, Thomassey, and Zeng (2019). Social Media, in particular, has made fashion more accessible by providing users with instant exposure to the latest clothing styles, while targeted advertisements further drive consumer interest.@cleofasbrand has also quoted that fashion giants like H&M and Zara have very high influence over the customers as they can launch a new fashion trend that will be widely accepted by consumers around the globe. This evolution has not only diversified the fashion industry but also shown new ways for customers to engage with the purchase of clothing and accessories.@glinska2017customer analyzed the Polish data and also highlighted that there has been a surge of 11.6% in online shoppers from 2012 to 2016. Not only this, the variety of clothing, and convenience of purchasing have broadly influenced customers to shift from offline retail market to online platforms.

On the other hand, this shift has also occurred due to recent pandemic situations around the globe. After COVID-19, many behavioral changes occurred in individuals that impacted how they purchase goods. During the COVID period, the fashion industry was one of the industries that were severely affected by the lockdown which made consumers shift towards e-commerce platforms Cleofas et al. (n.d.). Additionally, Wibowo (2024) has also mentioned how a famous fashion industry, Santoon, in Indonesia, faced problems with e-commerce despite being a market leader in the country.

These shifts need to be understood to help new online businesses survive in this competitive market. Customer preferences can be identified by analyzing past data, which helps develop effective strategies for new companies. Analyzing past data is a crucial step for understanding customer needs, purchase patterns, and much more. The customer needs a more personalized service that can be carried out by analyzing past data and finding meaningful insights from the data @ giri2019customer. According to @ giri2019customer, the main strategy of the

fashion industry should be expanding its customer base and finding hidden insights that allow businesses to retain their customers.

Factors Affecting Customer Preference

The purchasing patterns of the customers are influenced by many factors. These factors vary from offline to online retail. First of all, detailed product information has more influence on customer purchase.@wibowo2024exploring identified that quality and size are the primary factors influencing customers' preferences for fashion products. Also, highlighted that product information is crucial for e-commerce platforms, which was missing on Santoon's social media platform. The author identified that customers view products at offline stores but purchase them online this suggests that product information needs to be detailed.

Secondly, past customer reviews are also important as they can leveraged to improve products and services. Not only this,Wibowo (2024) described that customer reviews play a crucial role in enhancing the online product. Moreover, the writer found that online exclusive offer such as free shipping and limited-time discounts affects the customer preference for purchasing online. In addition, secure payment options, return policies, and efficient delivery improve the recency of customers.

Furthermore,Cleofas et al. (n.d.) found in the study that brand was the most important factor contributing 43% among all the factors followed by clothing function, and place of purchase. Kod (n.d.) also quoted that across all the respondents brand name, lower price, and shopping from home were the motivators for the online shoppers.

On the one hand,Kod (n.d.)'s research on the Polish market found that age does not diversify customers' decisions despite having significant results of the Mann-Whitney U test for different age groups. On the other hand,Glińska and Tomaszewska (2017) also researched online retail in Poland and found that there are significant differences between age groups' purchasing patterns. The author found that individuals aged less than 25 saw online shopping as taking part in loyalty programs and spending time whereas the motives of people who had an age above 25 were making purchases as around-the-clock availability, more information about products, and possibilities to find special products. Additionally, Delimarta and Rahadi (2021) found that brand image, price, quality, and design are positively related to the customer's willingness to purchase the product.

The Role of Data Analytics in Understanding and Responding to Consumer Preferences:

The role of a data analyst in general is to uncover hidden patterns, and trends, and generate insights that help organizations make informed decisions. This can be carried out with the help of various tools to visualize and analyze the data. The duty of data analysts is also crucial for identifying customer preferences. According to Giri, Thomassey, and Zeng (2019), analytics

carried out to find the hidden patterns in customer purchases is called Customer Analytics. The Company's business goal defines how to incorporate it into business intelligence. The writer also identified the sources of the data extraction. Firstly, website traffic contains data such as page views, click-through rates, and time spent on the website. Moreover, purchase history shows records of purchases which can be used to identify the average order value, preferred brands, and categories. Additionally, from the customer feedback insights like product quality, customer service, and overall satisfaction can be derived. These customers can be clustered into separate groups (Sulianta, Ulfah, and Amalia 2024) using the clustering technique of k-mean clustering. The author clustered customers into 4 clusters based on price, quality, and brand. Moreover, the cluster also has customers who buy costly or premium products. Okofu et al. (2024) carried out a sentiment-based analysis for identifying customer behavior and based on the Bayesian Network further derived parameters of interest such as Usefulness, Usage Ease, Purchase Intension, User-trust.

Benefits of Data Analytics

The insights can be utilized by fashion retailers to enhance their customer service and marketing. Glińska and Tomaszewska (2017) concluded the research and identified age-dependent differences in consumer behavior which can be used in marketing strategies to target age groups with personalized offers and discounts on products, recommendations, and content that has customer-specific interest.

Cleofas et al. (n.d.) also identified that purchases are driven by purchases for everyday use and recreation without immediate purchase intention. This understanding helps inform product design assortment planning, and inventory management, ensuring that businesses offer products that meet customer demands. With the help of data analysis, we can enhance the customer experience, improve website usability, and improve the purchasing process which leads to customer satisfaction (Okofu et al. 2024).

Methods used

1. **Thematic analysis**: It can find recurring patterns and generate a theme (Wibowo 2024). This method minimally organizes and describes the data set in detail. Wibowo (2024) used thematic analysis to find recurring patterns related to customer preference of purchasing fashion items online versus offline.

2. **Triangulation**: According to Wibowo (2024), triangulation is used to validate research findings by comparing data from different sources. Suppose one source triangulated data from customer interviews, company owner observations, and existing research on customer decision-making in the fashion industry.

3. **Conjoint Analysis**: This method is used to find the relative importance of different product attributes in customer purchase decisions. Cleofas et al. (n.d.) and Suzianti, Faradilla, and Anjani (2015) used the conjoint analysis for Customer preference analysis.

4. **CRISP-DM (Cross-Industry Standard Process for Data Mining)**: Sulianta, Ulfah, and Amalia (2024) used this methodology which has stages like data collection, cleaning, analysis, and interpretation. These phases produce specific outputs that lead to the next steps in the project.

5. **K-Means Clustering**: It is an unsupervised machine learning algorithm used to cluster or group the data points into clusters based on their similarity. This method can be used to segment customers based on shared characteristics such as price sensitivity or brand preference (Sulianta, Ulfah, and Amalia 2024).

6. **Elbow method**: The elbow method is used in conjunction with K-Means Clustering to identify the optimal number of clusters. Sulianta, Ulfah, and Amalia (2024) used the elbow method to identify the number of clusters before passing it into the K-Means Clustering algorithm.

These are the common methods used by the researchers to conduct Customer Preference Analysis.

> Research Objective This study aims to explore the ever-changing market of fashion retail while focusing specifically on factors influencing customer preferences for online e-commerce fashion platforms and how this analysis can be useful in developing business strategies.

ⓘ Research Objective

This study aims to explore the ever-changing market of fashion retail while focusing specifically on factors influencing customer preferences for online e-commerce fashion platforms and how this analysis can be useful in developing business strategies.

Data description

According to the aim of the study to identify customer preferences based on behavioral and demographic traits such as age, gender, recency, preferred delivery method, and transaction amount, the dataset must contain these features or variables. The dataset that this research is going to utilize is available on bhadramohit (2024). This dataset contains data collected from several **E-commerce** and **Retail** platforms that depict US consumer trends. This dataset can be used in several analyses such as Customer segmentation, demographic analysis, and much more. Not only this due to the diversity of data, a variety of visualizations can be inherited from this dataset. The dataset also utilizes various GitHub repositories that collected data related to demographic data *(Age, Gender, Purchase Amount, and Season)* . The dataset contains a total of 3900 rows and 22 columns. The description of the Columns is listed in the below given table:

Table 1: Table for Column Description

Feature Name	Description
Customer.ID	Unique identifier for each customer.
Age	Age of the customer.
Gender	Gender of the customer.
Item.Purchased	Item that was purchased by the customer.
Category	Category of the purchased item.
Purchase.Amount..USD.	Purchase amount in USD.
Location	Geographical location of the customer.
Size	Size of the purchased item.
Color	Color of the purchased item.
Season	Season during which the item was purchased.
Review.Rating	Customer's rating for the item.
Subscription.Status	Whether the customer has a subscription.
Payment.Method	Payment method used for the purchase.

`Shipping.Type`	Type of shipping used for the purchase.
`Discount.Applied`	Whether a discount was applied during the purchase.
`Promo.Code.Used`	Whether a promo code was used for the purchase.
`Previous.Purchases`	Number of previous purchases made by the customer.
`Preferred.Payment.Method`	Customer's preferred method of payment.
`Frequency.of.Purchases`	Frequency with which the customer makes purchases.

- The categorical variables are Gender (Male, Female), Item.Purchased (Dress, Backpack,…), Category (Clothing, Footwear, Outerwear, Accessories), Location (New Jersey, Ohio,…), Size (S, M, L, XL), Color (Yellow, White, …), Season (Winter, Spring, Summer, Fall), Payment.Method (Credit, Debit, Venmo, Bank Transfer, Cash, Paypal), Preferred.Payment.Method (Credit, Debit, Venmo, Bank Transfer, Cash, Paypal), Shipping.Type (Express, Free shipping, Next Day Air, Standard, 2-Day Shipping, Store Pickup)

- Subscription.Status, Discount Applied, and Promo.Code.Used are binary categorical variable with Yes or No as possible values.

- Age, Purchase.Amount..USD., Review.Rating, Previous.Purchases, Frequency.of.Purchases are a continuous variable.

As the requirement of the study, there are all the features such as Age, gender, Frequency of Purchase, Size of fashion product, Promo code utilized or not, Shipping type, and many more. Moreover, features such as Subscription, Category of product, and Payment method can be utilized for further identifying the unexplored relationship. These relationships can be used to highlight some new factors affecting customer preferences. Additionally, Business can utilize these findings to strategies their business plan for all-round development. This data set can be utilized for further segmentation based on Demographic and Behavioural traits. Demographic traits can be used to target customers based on age and gender such as separate deals for Men and Women or special deals to children based on analysis. Behavioral Traits are Purchase Amount, Frequency of Purchases, and Preferred Payment Methods which can utilized to create clusters for loyal customers and new customers and then give offers based on the behavioural traits. Not only this, based on the item purchased and category, a Market Basket Analysis can also be conducted to find the products that are sold together such as t-shirts and pants, or shoes and socks.

Methods

In this study, I am going to use CRISP-DM also known as *Cross-Industry Standard Process for Data Mining* which is a methodology used by Sulianta, Ulfah, and Amalia (2024) in their study on *Revealing Consumer Preferences in the Fashion Industry Using K-Means Clustering*. This method includes following 4 steps:

1. Data Collection: Here, in this study data is collected from retail and e-commerce platforms across the US, and several github repository which was then aggregated and made available by bhadramohit (2024) on *Kaggle*.

2. Data Cleaning: In this step the dataset must be cleaned by removing NA values and making it ready for analysis by doing basic feature engineering.

3. Data Analysis: In this step, the data that is preprocessed needs to be explored and analyzed for generating meaningful insights.

4. Interpretation: The insights generated in previous step will be utilized and interpreted for understanding the trends and patterns to answer the question of research.

1. Data Collection

The data is collected by bhadramohit (2024) which can be downloaded and loaded into Rstudio using following code.

```
df <- read.csv("shopping_trends.csv")
df_2 <- df
```

The original dataframe is copied into another variable, `df_2`. This variable will be used in further analysis.

2. Data Cleaning

Firstly, the categorical features have to be converted into `factors` for analysis and better structuring of data. Therefore, looping through each column and comparing their class with `character` which then is converted into factor column can be done in following way. Moreover, the Size column has a natural order of S < M < L < XL, so Size feature can be factorized using `ordered = TRUE`.

```
# Loop to convert character column in factor column
for (i in names(df_2)){
  if(class(df_2[[i]]) == "character"){
    df_2[[i]] <- factor(df_2[[i]])


  }
}

# Factorizing Size column for ordering the levels
df_2$Size <- factor(df_2$Size, levels=c("S","M","L","XL"), ordered = TRUE)

str(df_2)
```

Secondly, the dataset should be checked for any NA values. This can be done by finding the column sum from each variable after applying `is.na()`.

```
# Check null values
colSums(is.na(df_2))
```

The age is continuous variable, to check if age have any effect in customer preference it needs to be binned into two categories: `Under 25 ( Age <= 25 ), and Above 25 ( Age > 25 )` (Glińska and Tomaszewska (2017)).

```
# Binning the age in to two categories
df_2$Age.bin <- ifelse(df_2$Age <= 25,"Under 25","Above 25")

# Factorizing the column
df_2$Age.bin <- factor(df_2$Age.bin)

str(df_2$Age.bin)
```

3. Data Analysis

First of all, the new binned age (`Age.bin`) needs to be analyzed with other features to find if there is a significant difference in the Age groups.

Frequency of Purchase may vary for individuals under 25 and above 25. There may be a significance that Age has a relationship with Frequency of purchase. To visualize both the groups based on the count of different frequencies of purchase category a horizontal bar chart for each age group can be created.

```r
age_frequency <- data.frame(table(df_2$Age.bin, df_2$Frequency.of.Purchases))
colnames(age_frequency) <- c("Age","Frequency.of.Purchase","Count")

age_frequency %>%
  ggplot(aes(fill = Age, x = Frequency.of.Purchase, y = Count)) +
  geom_bar(stat = "identity", position = "stack") +
  coord_flip() +
  labs(x = "Frequency of Purchase")
```

A chi-squared test of independence is conducted to find if there is any significant effect of age on the frequency of purchase.

```r
# Chi square test of independence
chisq.test(table(df_2$Age.bin,df_2$Frequency.of.Purchases))
```

As mentioned by Glińska and Tomaszewska (2017) younger generation saw online shopping as part of loyalty programs. Promo codes are one of the aspects of loyalty programs where customer can get promo codes based on their frequency of purchase. Therefore, a relationship between age and promo code used can be analyzed using a heatmap with the count in each category and a chi-squared test to prove the effect.

```r
# 1.2 Promo codes
age_promo_table <- data.frame(table(df_2$Age.bin,df_2$Promo.Code.Used))

colnames(age_promo_table) <- c("Age.bin", "Promo.Code.Used", "Count")

age_promo_table %>%
  ggplot(aes(x = Age.bin, y = Promo.Code.Used, fill = Count)) +
    geom_tile() +
    labs(title = "Heatmap for Age group and Promo code used",
         x = "Age group", y = " Promo Code Used", fill = "Count") +
    theme_minimal() +
  geom_text(aes(label = Count), color = "white", size = 6)
```

```r
age_promo_code <- table(df_2$Age.bin, df_2$Promo.Code.Used)
chisq.test(age_promo_code)
```

Furthermore, Wibowo (2024) concludes that customer reviews play a crucial role. So, to find which age group gives a higher rating a boxplot of the distribution of ratings and a t-test to confirm the effect of age and reviews rating needs to be computed.

```
# 1.3 Review Rating
df_2 %>%
  ggplot(aes(x=Age.bin, y = Review.Rating, fill = Age.bin)) +
  geom_boxplot()
```

Before performing the t-test the assumptions need to be checked to verify if a parametric test can be conducted or not. After conducting a t-test and getting a significant p-value < 0.05, a Cohen's d test will be conducted to find the effect size of the relationship.

```
# T test
Age.grouping <- group_by(df_2,Age.bin)
```

```
# Assumptions checking
# Outlier checking
identify_outliers(Age.grouping, Review.Rating)
```

```
# Normality checking
shapiro_test(Age.grouping, Review.Rating)
```

```
# Homogeniety of variance
levene_test(df_2, Review.Rating ~ Age.bin)
```

```
t_test(Review.Rating ~ Age.bin, data = df_2)
```

```
cohens_d(df_2,Review.Rating ~ Age.bin, var.equal = FALSE)
```

The promo codes are very crucial during online purchases therefore, gender may have an influence on promo code usage. This can be conducted using a chi-squared test and post-hoc test to find which gender uses promo-code most.

```
# Dataframe containing Contingency table for gender vs promo-code use
gender_promo_table <- data.frame(table(df_2$Gender, df_2$Promo.Code.Used))

# renaming column names of the gender_promo_table dataframe
colnames(gender_promo_table) <- c("Gender", "Promo.Code.Used", "Count")

# Heat map to visualize the contingency table
gender_promo_table %>%
  ggplot(aes(x = Gender, y = Promo.Code.Used, fill = Count)) +
    geom_tile() +
```

```
    labs(title = "Heatmap for Gender and Promo Code Used",
         x = "Gender", y = "Promo Code Used", fill = "Count") +
    theme_minimal() +
    geom_text(aes(label = Count), color = "white", size = 6)
```

```
# Continegency table
gender_promo_code <- table(df_2$Gender, df_2$Promo.Code.Used)

# Chi-square test
chisq.test(gender_promo_code)

# Post-hoc test
chisq.posthoc.test(gender_promo_code)
```

Wibowo (2024) also highlighted that offers like free shipping affects customer preference. Therefore, this difference can be visualized using a horizontal grouped bar chart. In addition, a chi-square test can be performed to identify if there is any significance in relationship between gender and shipping type. Also, a post-hoc test will be performed to find which shipping method is prefered by each gender.

```
# 2.6 Shipping Type
gender_shipping <- data.frame(table(df_2$Gender, df_2$Shipping.Type))
colnames(gender_shipping) <- c("Gender","Shipping_Type","Count")

gender_shipping %>%
  ggplot(aes(fill = Gender, y = Count, x = Shipping_Type)) +
  geom_bar(position = "dodge", stat = "identity") +
  coord_flip()
```

```
shipping_table <- table(df_2$Gender,df_2$Shipping.Type)
chisq.test(shipping_table)
```

```
# Posthoc test for shipping vs gender
chisq.posthoc.test(shipping_table)
```
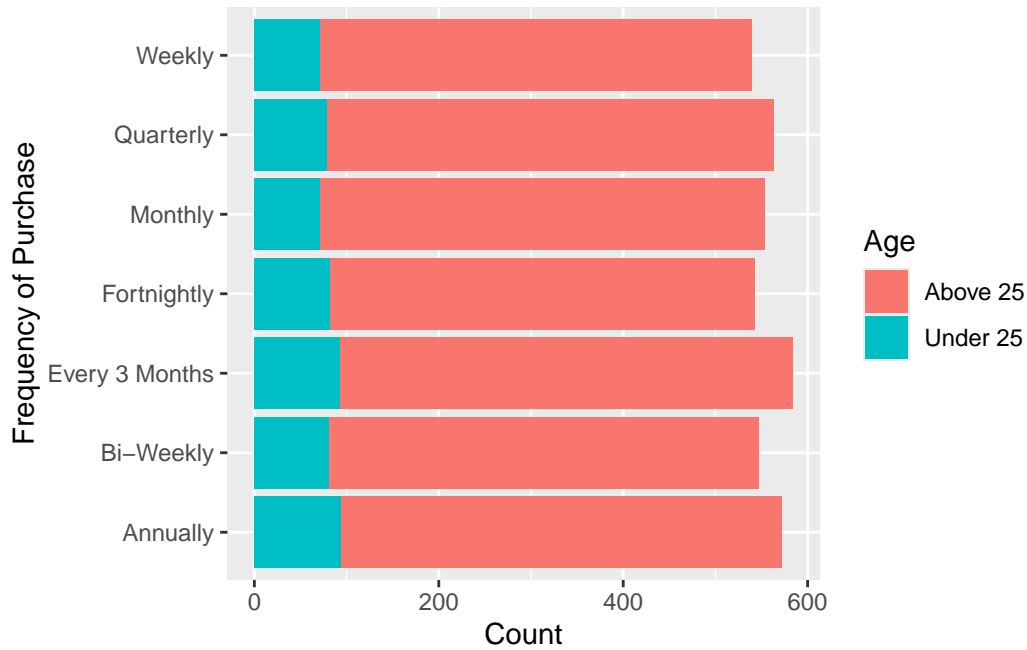
Results

Figure 1: Plots showing the frequency purchase in different age group

The plot shows that the proportions of people Above 25 is quite higher than Under 25. Moreover, both groups have high preference for the purchasing every 3 months. Additionally, People both under and above 25 have higher frequency of purchase annually. Whereas, rest of the Frequency of purchase categories also have some proportion of customers. From the plot it can be said that there are not much of a difference in frequency of purchase within groups.

```
    Pearson's Chi-squared test

data:  table(df_2$Age.bin, df_2$Frequency.of.Purchases)
X-squared = 4.8901, df = 6, p-value = 0.558
```

The Chi-square test of Independence were not significant, $X^2(6, 3900) = 4.8901, p > 0.05$, this suggests that the Frequency of Purchase and Age groups are independent of each other and there is not enough evidence to conclude that there is any relation between two variables.
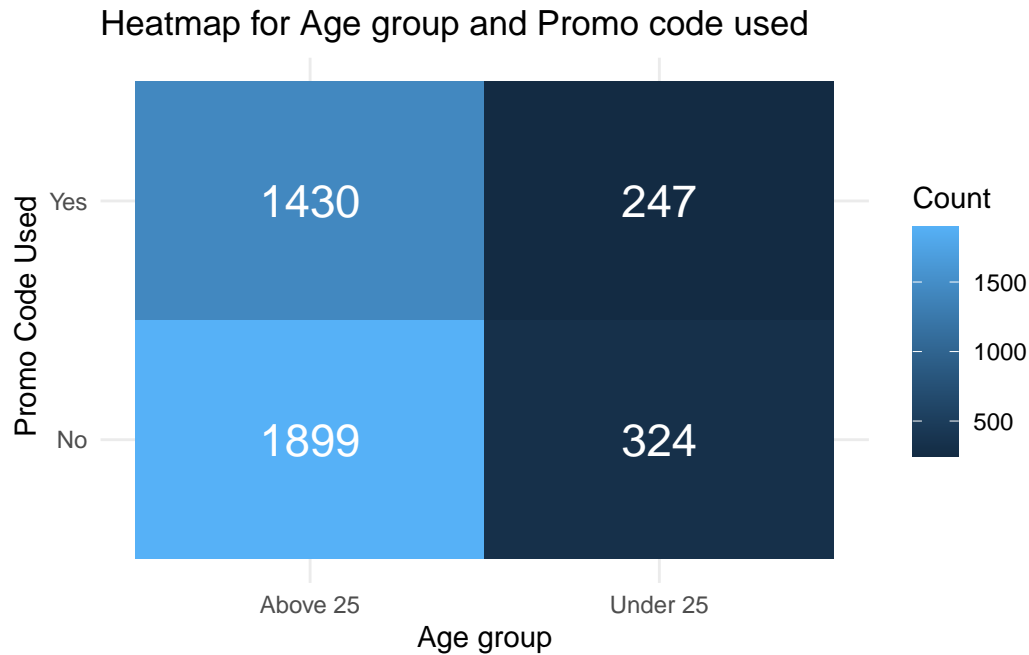
12

Figure 2: Heatmap showing contingency table of Promo code used vs Age Group

The heatmap shows the contingency table between Age group and Promo Code Used or not. Although there are less individual under 25 than above 25. In addition, both the groups have higher number of individuals **not using promo code** than those using promo code.

```
        Pearson's Chi-squared test with Yates' continuity correction

data:  age_promo_code
X-squared = 0.0078762, df = 1, p-value = 0.9293
```

The pearson's chi-squared test shows that there is **no relationship** between age groups and promo code use.
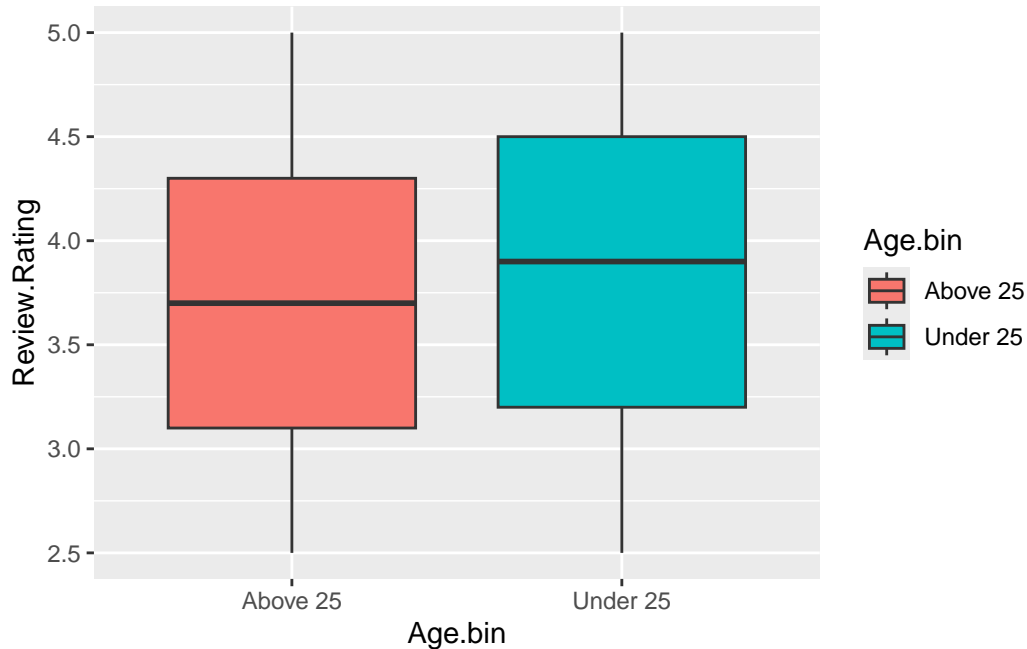
Figure 3: Boxplot showing distributions of Review ratings in different Age Groups

The boxplots depicts that overall both the groups rate `between 2.5 and 5`. However, shoppers that are Above 25 having median rating below 3.75 while Under 25 usually rate higher with median rating above 3.75. This hypothesis can be checked using a t-test of independence.

Before conducting the test, assumption of t test needs to be checked

i. **Extreme Outliers**: Here, the continuous variable is checked to identify if there are any extreme outliers in the dataset.

```
 [1] Age.bin                 Customer.ID            Age
 [4] Gender                  Item.Purchased        Category
 [7] Purchase.Amount..USD.   Location              Size
[10] Color                   Season                Review.Rating
[13] Subscription.Status     Payment.Method        Shipping.Type
[16] Discount.Applied        Promo.Code.Used       Previous.Purchases
[19] Preferred.Payment.Method Frequency.of.Purchases  is.outlier
[22] is.extreme
<0 rows> (or 0-length row.names)
```

The output shows 0 rows which means there are no outliers or extreme outliers in the Review rating variable.

ii. Normality Test: In this test the Shapiro-Wilks test is conducted to check if there is normality in the distribution between groups. Here if p-value is > than 0.05 than the data is normally distributed.

```
# A tibble: 2 x 4
  Age.bin  variable      statistic       p
  <fct>    <chr>             <dbl>   <dbl>
1 Above 25 Review.Rating    0.957 2.02e-30
2 Under 25 Review.Rating    0.948 3.02e-13
```

Unfortunately, the Shapiro-Wilks test failed to prove normality as p-values are less than the significance level (0.05). A non-parametric needs to be conducted.

iii. Homogeneity Test: In this assumption test, with the help of the Levene's test, variance of different groups are checked if they are homogeneous or heterogeneous.

```
# A tibble: 1 x 4
    df1   df2 statistic       p
  <int> <int>     <dbl> <dbl>
1     1  3898      1.15 0.283
```

Levene's test proved the homogeneity in the variance of the two groups as p-value(0.283) > 0.05.

```
# A tibble: 1 x 8
  .y.           group1   group2       n1    n2 statistic    df       p
* <chr>         <chr>    <chr>     <int> <int>     <dbl> <dbl>   <dbl>
1 Review.Rating Above 25 Under 25   3329   571     -2.37  773. 0.0178
```

The results of the t-test suggest there is some relationship between Review rating and Age groups, $t(773) = -2.37$, $p < 0.05$.

```
# A tibble: 1 x 7
  .y.           group1   group2   effsize    n1    n2 magnitude
* <chr>         <chr>    <chr>       <dbl> <int> <int> <ord>
1 Review.Rating Above 25 Under 25  -0.108  3329   571 negligible
```

As per Cohen's D test, the relationship has a negligible effect between Age and Review rating.
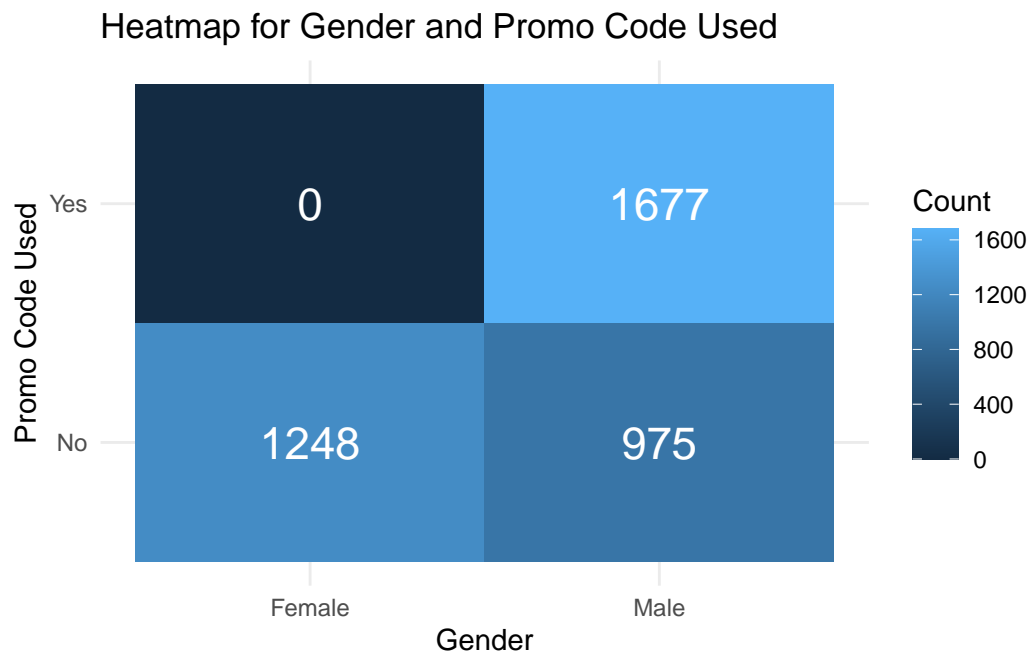

Figure 4: Heatmap showing the contingency table between Gender and Promo Code Use

It is clear from the heat map that Females do not like to use Promo codes whereas, on the other hand, Men highly use Promo codes.

```
    Pearson's Chi-squared test with Yates' continuity correction

data:  gender_promo_code
X-squared = 1381.9, df = 1, p-value < 2.2e-16


  Dimension      Value         No        Yes
1    Female Residuals   37.20914 -37.20914
2    Female  p values    0.00000   0.00000
3      Male Residuals  -37.20914  37.20914
4      Male  p values    0.00000   0.00000
```

The results of the chi-square confirm that the p-value for Promo-code use as `Yes` is less than 0.05 with residuals negative and males having residuals having residuals in positive.
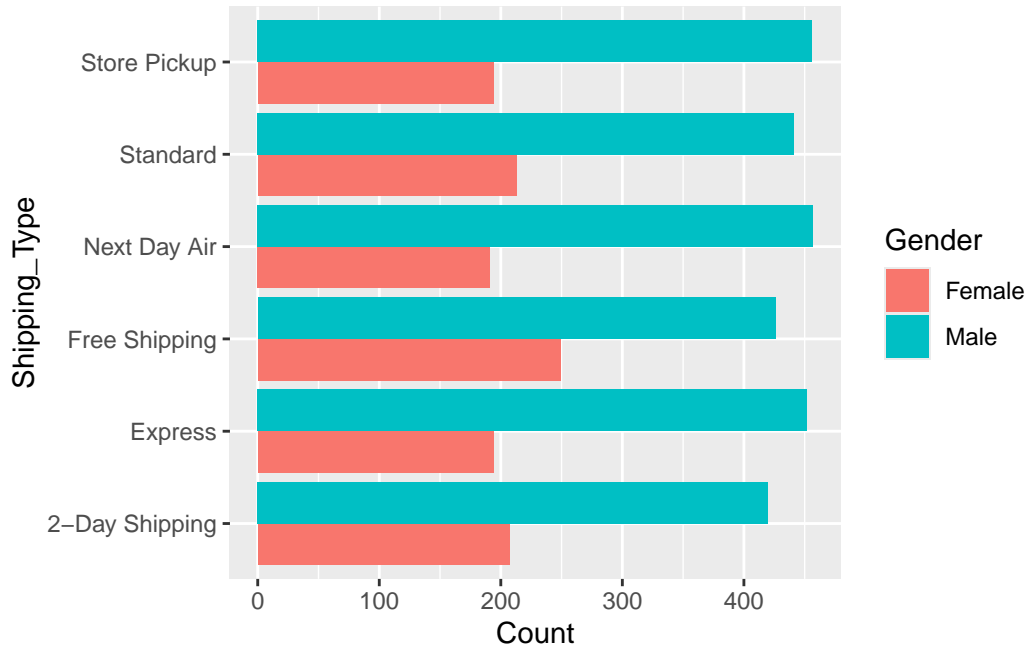
Figure 5: Horizontal barchart showing

The horizontally grouped bar chart depicts that Males have a higher preference for Store Pickup, Next Day Air, and Express. Whereas Females prefer Free Shipping highly.

```
    Pearson's Chi-squared test

data:  shipping_table
X-squared = 12.243, df = 5, p-value = 0.03161
```

After conducting the chi-square test of independence between Gender and Shipping type, it shows there is significance, $X^2(5, 3900) = 12.243, p < 0.05$.

```
  Dimension      Value 2-Day Shipping   Express Free Shipping Next Day Air
1   Female Residuals      0.594365 -1.174532      2.994333    -1.508771
2   Female  p values      1.000000  1.000000      0.033006     1.000000
3     Male Residuals     -0.594365  1.174532     -2.994333     1.508771
4     Male  p values      1.000000  1.000000      0.033006     1.000000
    Standard Store Pickup
1  0.3418083    -1.289533
2  1.0000000     1.000000
```

```
3 -0.3418083     1.289533
4  1.0000000     1.000000
```

From the post-hoc test, it is clear that only free shipping has a p-value less than 0.05. Based on residuals Females have a higher preference for free shipping than Males. Lastly, other types of shipping methods do not have much significance with gender.

Bibiliography

bhadramohit. 2024. "Customer Shopping Latest Trends Dataset." https://www.kaggle.com/datasets/bhadramohit/customer-shopping-latest-trends-dataset.

Cleofas, Maria Arielle, Yogi Tri Prasetyo, Ardvin Kester S Ong, and Satria Fadil Persada. n.d. "Brand or Clothing Function? Consumer Preference Analysis on Clothing Apparel Attributes and Design: A Conjoint Analysis Approach."

Delimarta, Florencia Devina, and Raden Aswin Rahadi. 2021. "Customer Preferences on Sustainable Fashion Purchases: A Conceptual Model." *International Journal of Entrepreneurship and Management Practices* 4 (13): 78–88.

Giri, Chandadevi, Sebastien Thomassey, and Xianyi Zeng. 2019. "Customer Analytics in Fashion Retail Industry." In *Functional Textiles and Clothing*, 349–61. Springer.

Glińska, Ewa, and Ewelina Tomaszewska. 2017. "Customer Preferences Related to Shopping Online." *Annales Universitatis Mariae Curie-Skłodowska, Sectio H–Oeconomia* 51 (2): 87.

Kod, JEL. n.d. "Customer Preferences Related to Shopping Online." *Chemistry* 1: 56.

Okofu, Sebastina Nkechi, Kizito Eluemunor Anazia, Maureen Ifeanyi Akazue, Margaret Dumebi Okpor, Amanda Enadona Oweimieto, Clive Ebomagune Asuai, Geoffrey Augustine Nwokolo, Arnold Adimabua Ojugo, and Emmanuel Obiajulu Ojei. 2024. "Pilot Study on Consumer Preference, Intentions and Trust on Purchasing-Pattern for Online Virtual Shops." *Int. J. Adv. Comput. Sci. Appl* 15 (7): 804–11.

Statista. 2023. "Global Online Fashion Market Value from 2017 to 2027." https://www.statista.com/statistics/228706/global-online-fashion-market-value/.

Sulianta, Feri, Khaerani Ulfah, and Endang Amalia. 2024. "Revealing Consumer Preferences in the Fashion Industry Using k-Means Clustering." *International Journal of Engineering Continuity* 3 (2): 34–53.

Suzianti, Amalia, Nurza Dwi Prisca Faradilla, and Shabila Anjani. 2015. "Customer Preference Analysis on Fashion Online Shops Using the Kano Model and Conjoint Analysis." *International Journal of Technology* 6 (5): 881–85.

Wibowo, Amira Luthfiyya. 2024. "Exploring Santoon's Customers' Preferences That Affect Their Purchase Decision for Buying Fashion Products." *Journal of Consumer Studies and Applied Marketing* 2 (2): 139–45.