**A staircase in a building

Description automatically generated with low confidenceLogo, company name

Description automatically generated**

**Author:** Ibrahim A. Shukoor

**Retention**

**Customer**

**Problem Statement**

Word of mouth is regarded as one of the most organic techniques of sales generation, requiring no additional cash from companies (E-tailers) for promotion. On the contrary, greater expenditures in the company's services are required to improve the website's efficiency and performance in order to gain more trust and speed up transactions. But which is more significant, hedonistic or utilitarian values? A case study of Indian e-commerce customers brought to light a list of E retail factors that may possibly play a role in customer activation and retention. The dataset contained a total of 71 columns, the first 10-11 columns contain questions aimed to extract individual user details such as age, gender, and city *etcetera.* The next 36 questions include questions which are designed to understand user’s shopping habits and to assess the importance of E-retail factors for individual users. For example:

1. How frequently do you abandon (selecting an item and leaving without making payment) your shopping cart? **{user shopping habit}**
2. Displaying quality Information on the website improves satisfaction of customers. **{importance of e-retail factor for the user}**

The remaining 24 questions were open-ended and carefully selected to extract the name of the online retailer that the user believed would be the greatest match for each question. It included questions like:

1. Easy to use website or application
2. Reliability of the website or application
3. Limited mode of payment on most products (promotion, sales period)

The provided dataset was subjected to data analysis, and numerous questions were addressed.

**Scope of the study**

The main objective was to evaluate how important hedonistic and utilitarian values were to the questioned user population. It also looked to see whether there were any gender-related differences in user opinions. Additionally, it revealed which E-tailer was believed to be more well-liked and which were considered unsatisfactory according to the users.

**Exploratory Data Analysis and Data Visualization**

The dataset was examined for null values before being divided into two and having the column names modified to something more succinct and clearer. The initial 47 questions were included in the first dataset, while the remaining subjective questions were included in the second dataset.

**Text, letter

Description automatically generated**

Figure 1: Renaming columns

**Analysis**

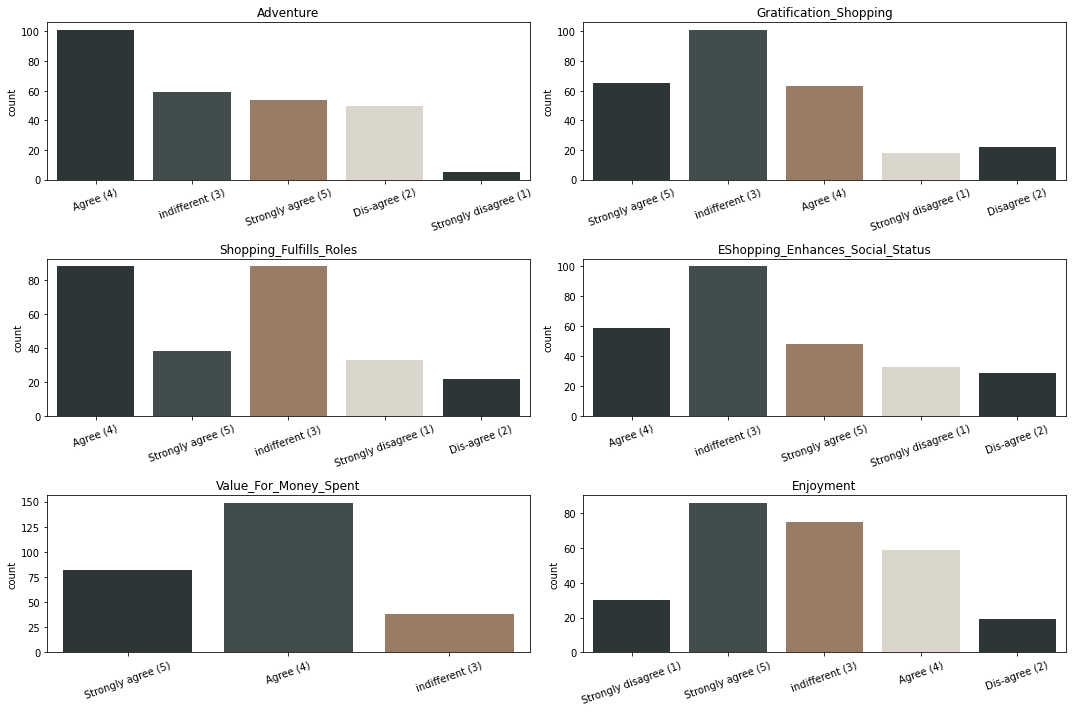
1. Hedonistic Values

Figure 2: Importance of Hedonistic values among users

Customers appear to agree wholeheartedly that elements like a sense of adventure, receiving value for the money invested, and enjoyment play a role in their purchasing decision. However, they find that factors including sense of gratification, fulfilling roles, and enhancing social status are irrelevant. We can examine this case in more detail to identify differences in opinion in terms of gender.

1. Hedonistic Values and Gender

The hedonistic values vary depending on the gender. Most often, this conflict/difference can be seen in:

* In the case of gratification, most females opt for ‘indifferent’ and most guys choose to ‘agree’.
* Shopping Fulfils Roles: The majority of women choose ‘indifferent’, and the majority of men ‘agree’ that shopping helps them fulfil certain roles.
* E Shopping Enhances Social Status: Both the genders opt for ‘indifferent’
* When it comes to enjoyment, most Females chose ‘strongly agree’, whereas the majority of males selected ‘indifferent’.

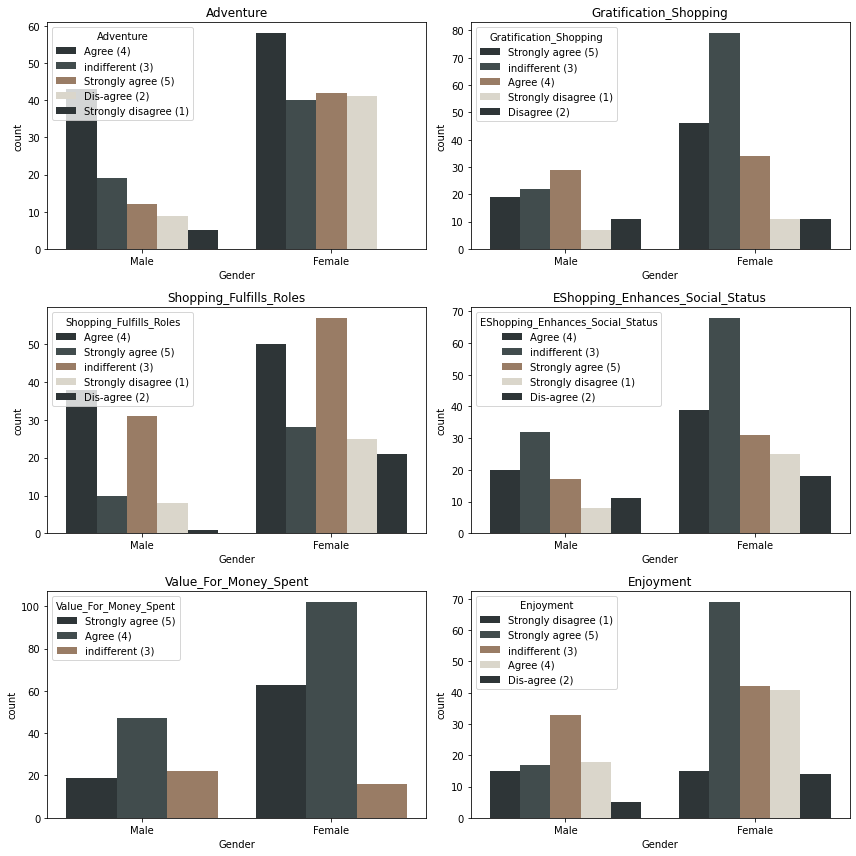
We can safely say that Hedonistic values are not constant in terms of gender, and they vary quite a lot.

Figure 3: Hedonistic values with respect to Gender

1. Utilitarian Values

From the illustration below (Figure 4), it is clear that variables like complete product information, monetary savings, discounts, value for money spent, information on similar products, and convenience are do play a role in users’ purchase decisions. Thus, we can conclude that utilitarian principles are essential.

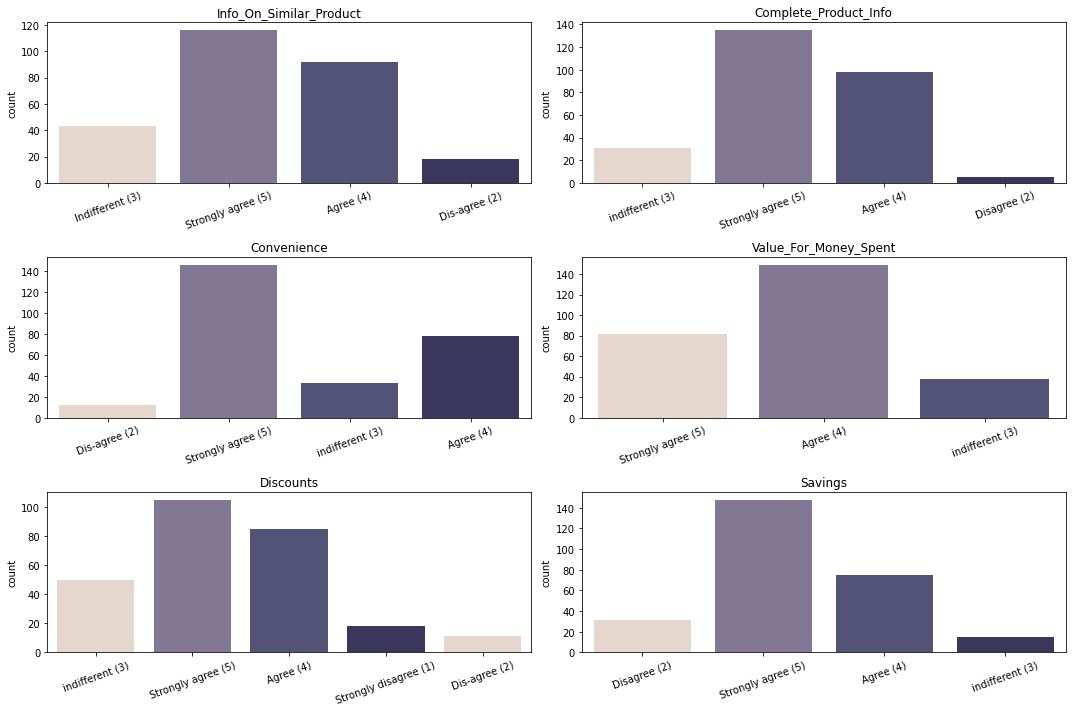


Figure 4: Importance of Utilitarian values among users

1. Utilitarian Values and Gender

Qr code

Description automatically generatedThere appears to be no distinction between the genders in terms of utilitarian values. The following elements—complete product information, financial savings, discounts, value for money spent, information on similar products, and convenience—appears to be deemed important by both the genders.

Figure 5: Utilitarian values and with respect to Gender

1. Service Quality

**Background pattern

Description automatically generated**Empathy, privacy, responsiveness, availability of returns and replacements, quality information, ease of navigation, responsiveness, friendliness of the user interface, order fulfilment trust, and satisfaction with good UI/UX are all service qualities that are critical to boosting sales and boosting customer loyalty, as can be seen in the figure below. (Figure 6A, 6B)

Figure 6A: Importance of Service Quality among the users

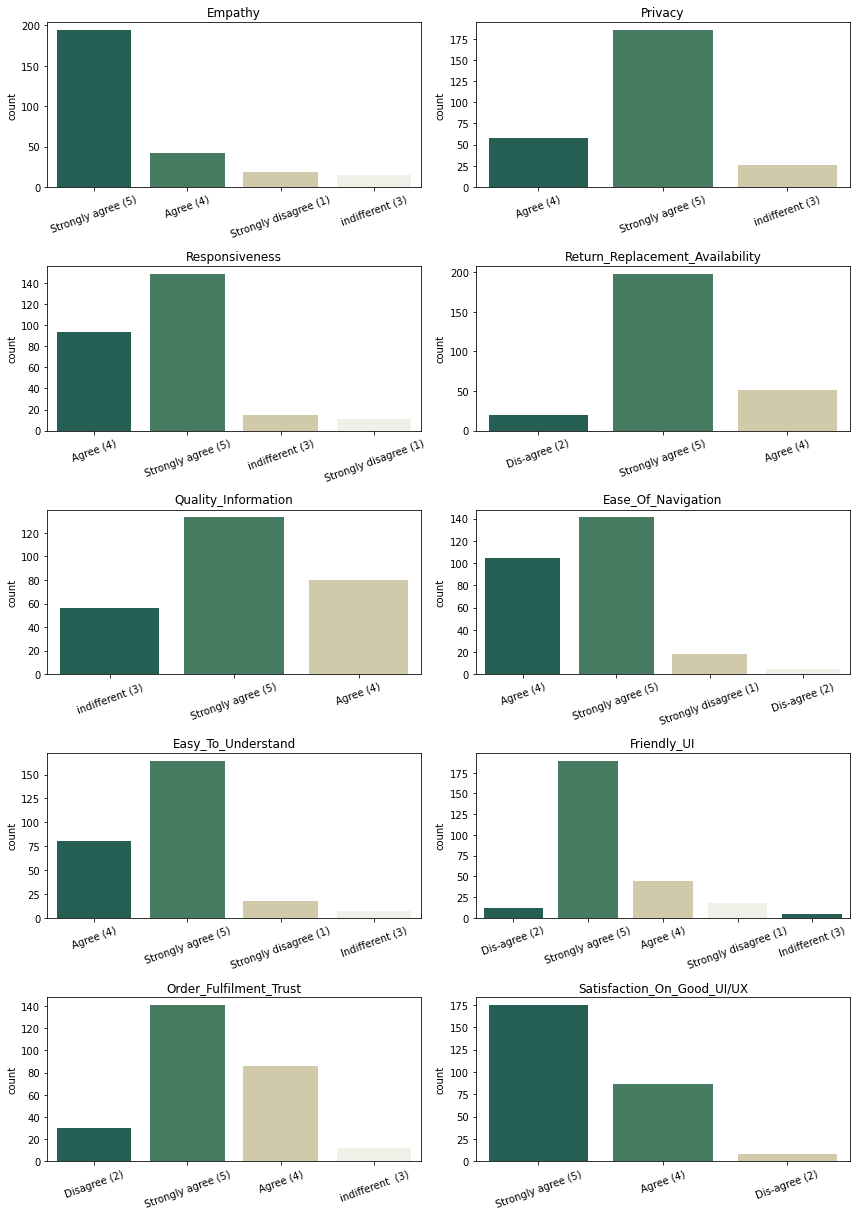
1. Population Characteristics

Figure 6B: Importance of Service Quality among the users

A picture containing outdoor, monitor, tower, light

Description automatically generatedAnalysis was done to understand the population characteristics (Figure 7)

**A picture containing clock, outdoor, tower, light

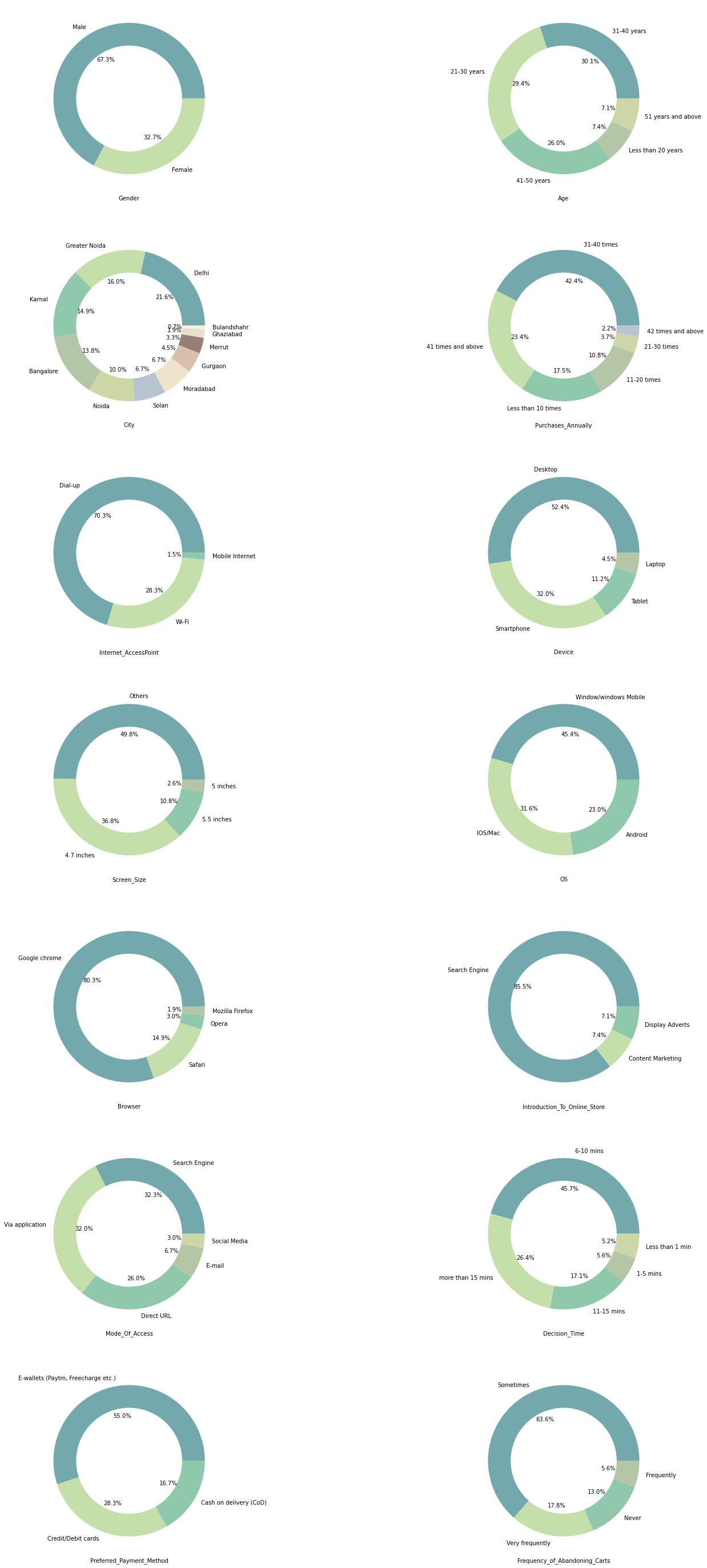
Description automatically generated**

Figure 7: Population characteristics

1. Additional Analysis

In order to obtain more information, further analysis was done. It was discovered throughout this analysis that

1. The purchase decision-making process varied by gender, with the majority of males requiring less than one minute and the majority of females taking more than 15 minutes to make a purchase decision. (Figure 8)

Chart, histogram

Description automatically generated

Figure 8: Decision time with regards to Gender

1. Chart, line chart

   Description automatically generatedThe best decision-making window was between 1–5 and 6–10 minutes. Users who made purchases during this window had greater annual purchases. Users who took longer than 15 minutes to complete a transaction often made less than 10 purchases per year.

Figure 9: Decision time vs Purchases Annually

1. Chart, line chart

   Description automatically generatedAdditionally, we also saw that Credit/Debit cards and Window/windows Mobile are popular amongst Females and E-wallets (Paytm, Freecharge etc.) and IOS/Mac are more popular amongst Males.

Figure 10: Preferred Payment Method vs OS

1. Chart, histogram

   Description automatically generatedThe most popular devices for making purchases are smartphones, followed by laptops and then desktops.

Figure 11: Device with regards to Purchases Annually

1. Best and Least Satisfactory E-tailers (according to the users)

*The following were the factors considered while assessing the* ***best online store*** *according to the users:*

*Online Retailers Used, Visually Appealing, Easy to Use, Variety of Products, Complete Product Info, Loading Speed, Reliability, Transaction Speed, Availability of Payment Options, Fast Delivery, Privacy, Security, Trust, Online Assistance, Efficient and Recommendation*

|  |  |  |
| --- | --- | --- |
| Rank | E-tailer Name | # of times mentioned |
| 1 | Amazon | 3860 |
| 2 | Flipkart | 2774 |
| 3 | Myntra | 1436 |
| 4 | Paytm | 1208 |
| 5 | Snapdeal | 1128 |

Table 1: Best online store assessment

*The following were the factors considered while assessing the* ***least satisfactory*** *according to the users:*

*Longer Log In, Longer time to display photos, Late price declaration, longer loading time, limited payment options, longer delivery, changes in user interface and frequent page disruptions.*

|  |  |  |
| --- | --- | --- |
| Rank | E-tailer Name | # of times mentioned |
| 1 | Amazon | 745 |
| 2 | Snapdeal | 551 |
| 3 | Flipkart | 539 |
| 4 | Paytm | 524 |
| 5 | Myntra | 388 |

Table 2: Least satisfactory online store assessment

As seen above (Table 1 and Table 2) customers appreciated Amazon in terms of its system, trust, reliability, and utilitarian values. Because Amazon excels in these areas, users appear to downplay the unpleasant aspects, such as slower loading times, fewer payment options, etc.

**EDA Conclusion and Remarks:**

We investigated the hedonistic values and discovered that the user base appears to be divided between competing ideals. It is more **difficult for an E tailer to satisfy the hedonistic desires** of the user community because **males and females tend to have different hedonistic values**.

The **utilitarian values analysis revealed that both genders reach a general understanding** (i.e., similar values), making it simpler for an E tailer to concentrate on enhancing its utilities, system qualities, and service qualities. In this way, consumer loyalty can be maximized.

Given that **80% of people use smartphones and laptops**, online retailers can adapt their advertisements to be focused on these devices to attract more customers. Additionally, **since 55% of the user base uses e-wallets** to finish their purchases, **they can offer more deals on e-wallets**.

Moreover, the analysis revealed that **shorter decision times—typically under 10 minutes**—**significantly boost the number of transactions a user makes**, **online retailers can therefore work to reduce the time customers take to make decisions**. This can be **accomplished through speeding up loading times**, **log-in times, price disclosures, and UI improvements.**

**Model Building:**

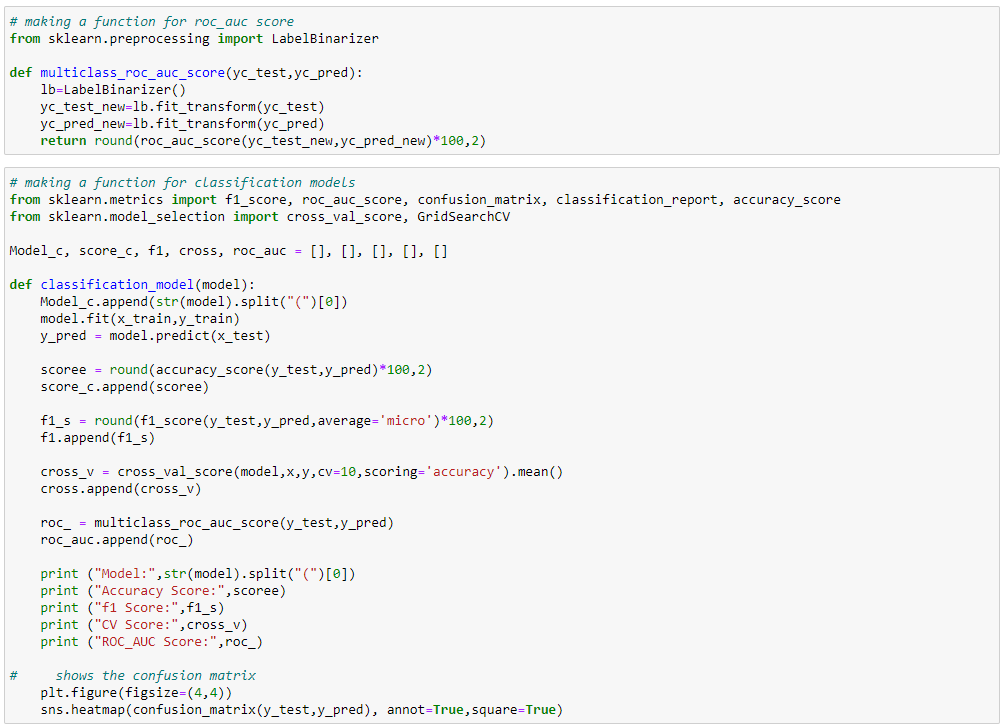
Text

Description automatically generatedThe number of unique values was reduced through feature engineering, and the column headers were made brief and concise. This helped in the understanding and visualizing the columns. Once the data was divided into **x** (independent variables) and **y** (target variable). The data was encoded using Ordinal Encoder and Label Encoder from the sklearn library.

Figure 12: Encoding **x** and **y**

The values were plotted using "**boxplots**" and "**kdeplots**" to visually identify outliers and column skewness.

1. Outliers: The outliers were treated using Z-score because IQR resulted in huge amounts of data loss.
2. Skewness: Outliers were removed, and the data was normalized using Min Max Scaler.

****A picture containing background pattern

Description automatically generatedSkewness and multicollinearity were handled after the outliers were removed. To balance the target variable, SMOTE over sampling was used. Following that, different classification models were applied to the dataset in order to compare and select the best model.

Figure 13(A): outlier visualization

Figure 14: user-defined function for obtaining the ROC-AUC score

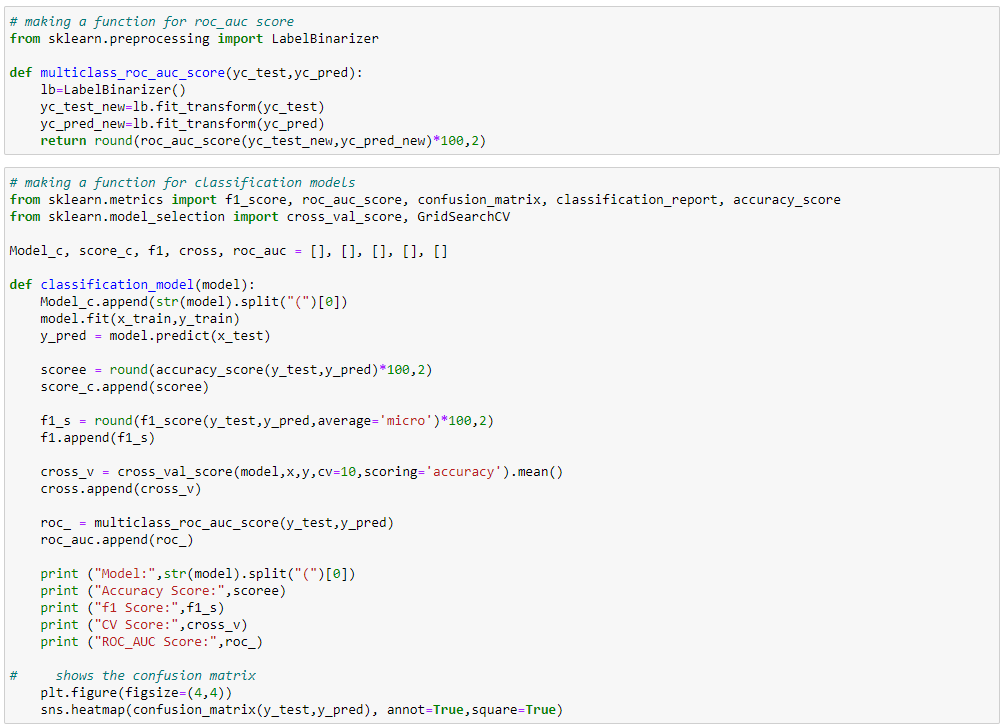
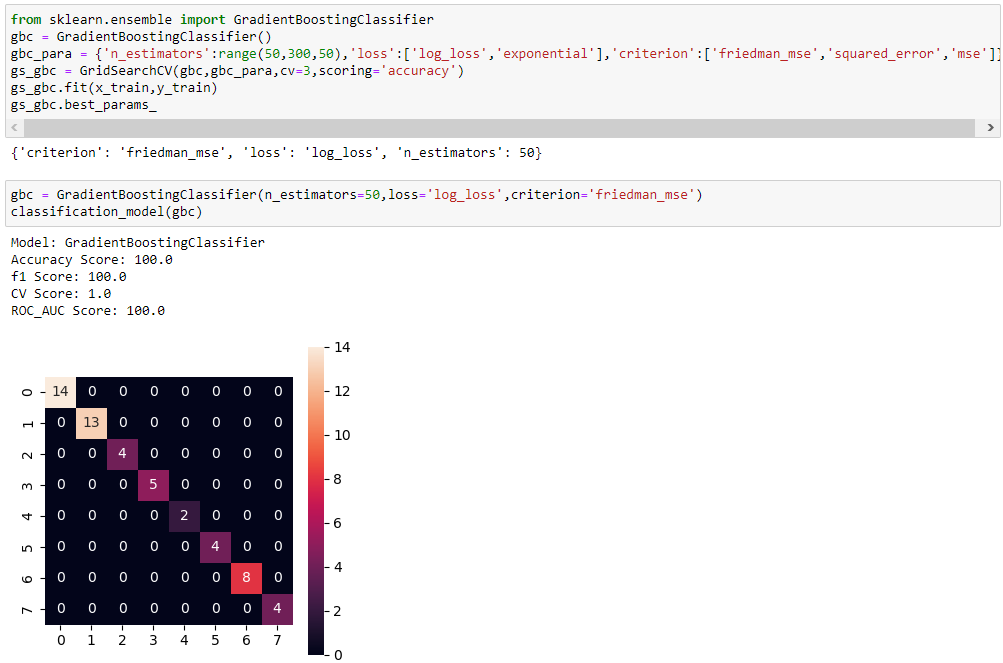
As a result Gradient Boosting Classifier had the best score of 100% accuracy, CV score and ROC\_AUC score.

Figure 15: user-defined function for obtaining the performance metrics of the models

**Feature Importance**

A feature importance plot is one way to interpret model results. It tells us what features are important for the machine learning model in deciding which variables are important for more recommendations.

Figure 16: Feature importance

Chart

Description automatically generated **A picture containing text

Description automatically generatedModel** **Conclusion:**

The top 10 essential features that can effectively predict the "recommended e-tailer," according to the model with the highest accuracy, are:

1. Efficiency of the website
2. Retailers used by the customer
3. Security of the E-tailer
4. Online assistance provided by the E-tailer
5. Transaction speed
6. Offers and Discounts (Savings)
7. Responsiveness of the website
8. Loyalty programs
9. Easy-to-use
10. Availability of payment options

**Jupyter Notebook**

The jupyter notebook with the visualizations performed and models built can be found here: <https://github.com/ibvhim/Blog-Notebooks/blob/main/Customer_Retention_ML%20version.ipynb>