Customer Retention Case Study Report

Prepared by Megha Singh

Data Science Intern at Flip Robo Technologies

FLIP ROBO

SME Name: Khushboo Garg

Acknowledgement

It is my deepest pleasure and gratification to present this report. Working on this project was an incredible experience that has given me a very informative knowledge regarding the data analysis process.

All the required information and dataset are provided by Flip Robo Technologies (Bangalore) that helped me to complete the project. I want to thank my SME Khushboo Garg for giving the dataset and instructions to perform the complete case study process.

INTRODUCTION

Problem Statement

Customer satisfaction has emerged as one of the most important factors that guarantee the success of online store; it has been posited as a key stimulant of purchase or repurchase intentions and customer loyalty.

A comprehensive review of the literature, theories and models have been carried out to propose the models for customer activation and customer retention.

Five major factors that contributed to the success of an ecommerce store have been identified as: service quality, system quality, information quality, trust and net benefit.

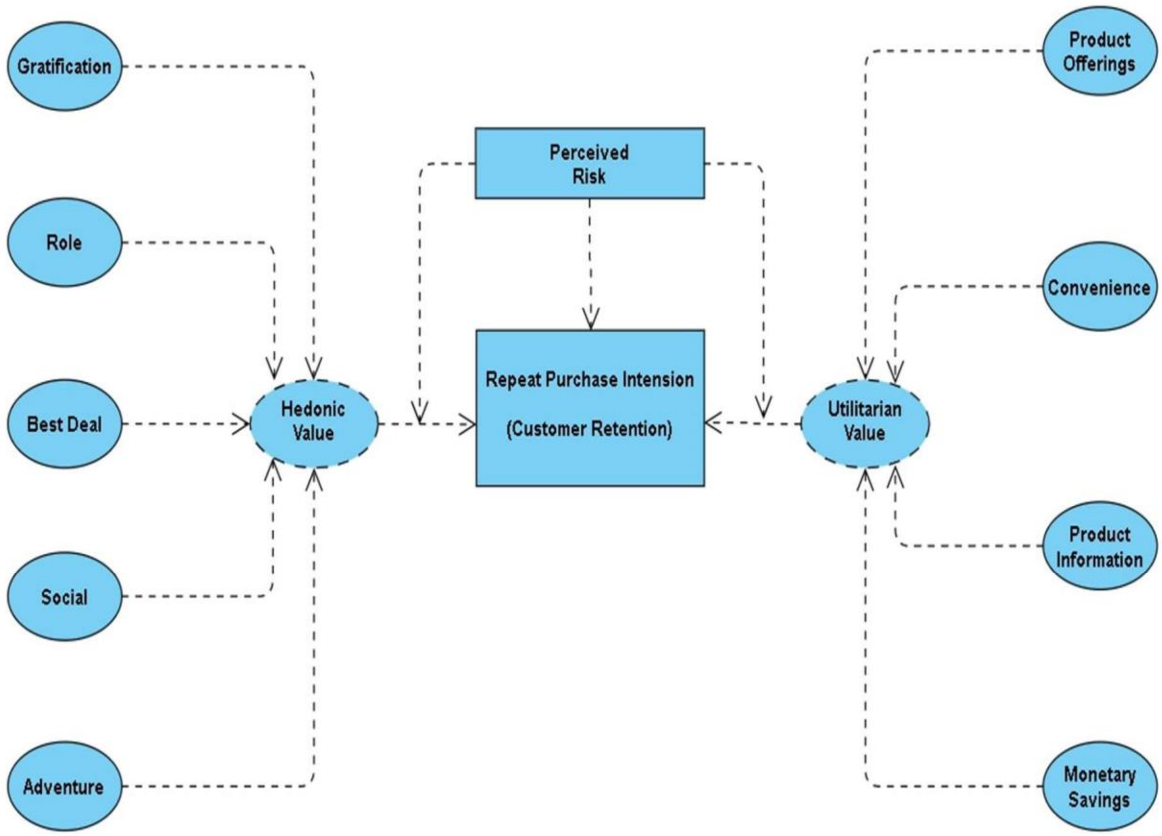
The research furthermore investigated the factors that influence the online customers repeat purchase intention.

The combination of both utilitarian value and hedonistic values are needed to affect the repeat purchase intention (loyalty) positively.

The data is collected from the Indian online shoppers.

Results indicate the e-retail success factors, which are very much critical for customer satisfaction.

Use Case Diagram



In the above use case diagram, we can see that the Repeat Purchase Intention basically our Customer Retention strategy relies on Hedonic value and Utilitarian value. Also, we see that there are perceived risks affecting the purchase and re purchase intentions of our customers. The Hedonic value has 5 major parts such as gratification, role, best deal, social aspect and adventure feeling criterions. Where as in Utilitarian value we have product offerings, convenience, product information and monetary savings.

Motivation for the Problem Undertaken

Our main objective of doing this project is to analyze whether the users are shopping products from e-commerce websites. How did they give feedbacks to these websites on the basis of several positive and negative

factors and also the details of the users on basis of factors like age, gender, city etc.

Benefits of Customer Retention:

1. Retention is cheaper than acquisition
    - While the old adage about "it costs five times as much to acquire a new customer" may not be accurate in every case, the basic principle is spot on: it’s more cost-effective to keep someone in the fold than to bring in new customers.
    - Even still, if it’s data you want, there has been plenty of research into acquisition vs retention, and every one of them has come back with the economics favoring retention as the more economically viable focus.
    - One caveat though: retention is cheaper than acquisition, but it isn’t necessarily easier.
2. Loyal customers are more profitable
    - Not only is loyalty cheaper, it has better returns. According to research, engaged consumers buy 90% more frequently, spend 60% more per transaction and are five times more likely to indicate it is the only brand they would purchase in the future.
    - On average, they’re delivering 23% more revenue and profitability over the average customer.
    - While loyal customers are more profitable, don’t take their loyalty for granted.
    - They’ll be more open to price increases, but be cautious not to raise prices simply to see how long they’ll stick around.
    - Consider the flipside: “Actively disengaged” customers (people who oppose the brand and may be actively spreading that opinion) can cost a brand 13% of its revenue.
3. Your brand will stand out from the crowd
    - Put your consumer hat on, and consider how many brands you interact with that actually seem to value your patronage.
    - You can probably only think of one or two.
    - Most brands focus on acquisition, which makes the retention-centric among us stand out even more.
    - People see around 10,000 marketing exposures a day, but only engage with a few of them.
    - The ones that earn continual engagement are those with whom they feel an emotional connection with on some level.
    - Forget a unique selling proposition; the best brands have a unique retention proposition.

4. You’ll earn more word-of-mouth referrals
    - Your loyal customers will be your best source of new business.
    - Despite all the efforts into online and mobile marketing and social media, people are still most strongly influenced by referrals from friends and family.
    - Millennials in particular will spread the word of a brand’s exploits: 90% share their brand preferences online.
5. Engaged Customers Provide More Feedback
    - Feedback is critical to the success of any business.
    - Customers who provide feedbacks are often willing to give brands the benefit of the doubt.
    - They’re telling you how to earn their business repeatedly. As research has shown, people who have complained and seen their issue resolved are 84% less likely to decrease their spend.
    - Need help dealing with the customers who are providing nasty feedback?
6. Customers will explore your brand
    - That’s a nice way of saying you’ll be able to sell them more stuff.
    - Once a brand has proven itself with one product or service, customers are six times more likely to say they would try a new product or service from the brand as soon as it becomes available.
    - That’s not just valuable for sales, but these folks can be utilized to help with #5 above as beta testers - a critical element in product development.
7. Loyal Customers are more forgiving
    - An Accenture study states over $1.6 trillion is lost each year due to customers bailing after a poor service experience.
    - We've gone so far as to claim that it's the top reason people will ditch a brand.
    - But customers who consider themselves loyal will let some misdeeds slide - just don't let it happen too often.
8. Customers will welcome your marketing No one likes being marketed to.
    - Except for loyal customers!
    - Those folks are four times more likely to say they “appreciate when this brand reaches out to me” and seven times more likely to “always respond to this brand’s promotional offers.”

9. You earn wiggle room to try new things
    - Loyalty is fickle, so too many changes could chase people away.
    - But once you’ve established a core base of proven customers, your brand can expand its boundaries.

- Maybe it’s new messaging or a new product line, or even a new logo. The bottom line is as long as you maintain the basic premises that keep people in your corner; they’ll stick with you through thin and thin.
- In fact, some of them will be excited to see what you can do.
- Existing customers are 50% more likely to try new products, according to a study.

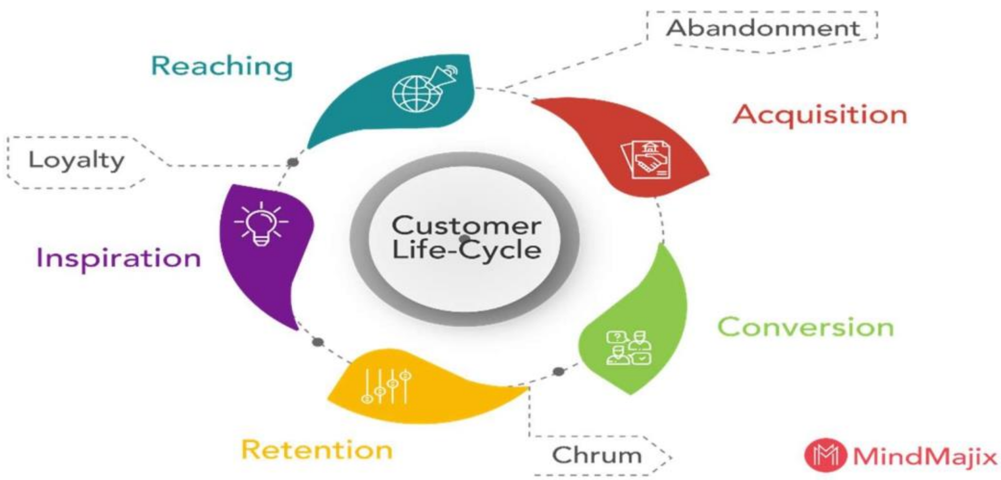
Lifetime revenue is the end goal, not just today’s revenue.

Need for Customer Retention:

Keeping current customers happy is generally more cost-effective than acquiring first-time customers. According to the Harvard Business Review, acquiring a new customer can be five to 25 times more expensive than holding on to an existing one.
Companies don’t need to spend big on marketing, advertising, or sales outreach. It is easier to turn existing customers into repeating ones, since they already trust your brand from previous purchases. New customers, however, often require more convincing when it comes to that initial sale.

Customer loyalty won’t just give companies repeat business. Loyal customers are more likely to give free recommendations to their colleagues, friends, and family. Creating that cycle of retained customers and buzz marketing is one way a company can cultivate customer loyalty for longterm success.

Improving customer retention means improving the customer experience. In fact, 77 percent of customers surveyed in a 2021 Customer Experience Trend Report being more loyal to a company that offers a good customer experience if they have an issue. 72 percent are willing to spend more from a company the offers good customer experiences. And 50 percent say that customer experience is more important to them now compared to a year ago.



Since the cost of getting a new customer is an estimated five to ten times more than keeping an old one, nurturing loyal customers is a powerful strategy that helps businesses grow.

Dataset Details:

First, I imported all the necessary libraries and dependencies to create a detailed data analysis in Python.

```python
import warnings
warnings.simplefilter("ignore")
warnings.filterwarnings("ignore")
import pandas_profiling
import missingno
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
from sklearn.preprocessing import OrdinalEncoder
```

Then I separated the sheets present in our Excel spreadsheet and stored them in 2 different dataframe variables.

```python
xls = pd.ExcelFile('customer_retention_dataset.xlsx')
df1 = pd.read_excel(xls, 'datasheet') # sheet 1
df2 = pd.read_excel(xls, 'codedsheet') # sheet 2
```

Exploratory Data Analysis (EDA):

After I got the dataset in our Jupyter Notebook I was able to notice that due to large number of rows and columns the information was truncated. Therefore, to overcome this challenge I used the pandas code as shown below.

```python
pd.set_option('display.max_columns', None) # show all columns in a dataframe
pd.set_option('display.max_rows', None) # show all rows in a dataframe
```

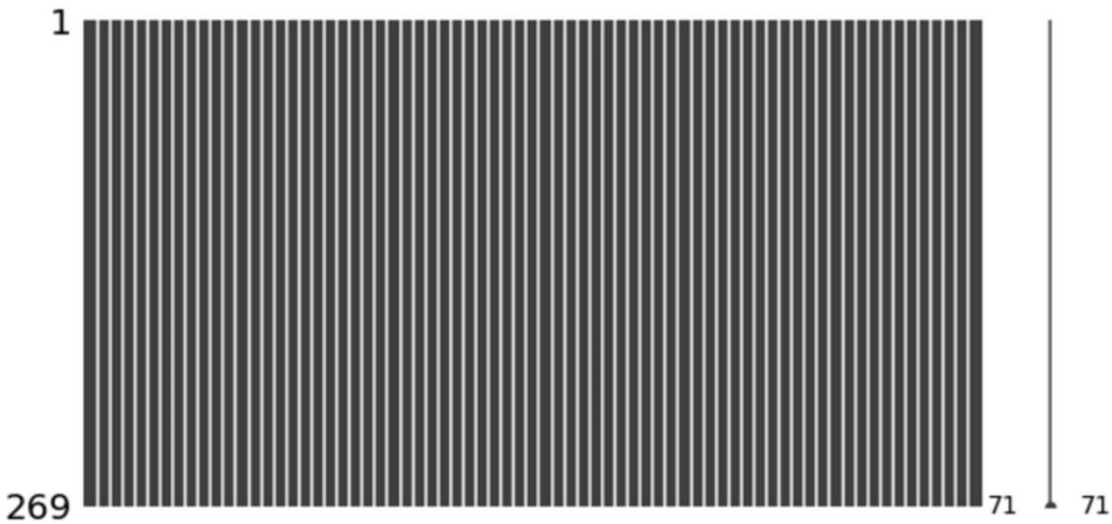
I made sure to rename the column names that were ill formatted and quite long which made no sense to me. With the help of rename I was able to change the names of columns that were too lengthy and could have been accommodated in shorter formats.

Now was the time to take a look at any kind of missing values or null value that might have been present in our dataset.

```
df1.isna().sum() # checking for missing values
```

```
missingno.matrix(df1, figsize = (10,5))
```

Luckily, I was able to see that there were no missing values in our entire dataset that is prominently visible in the matrix visual below.



I went ahead to take a look into each record information by making use of describe, info and nunique methods.

```
df1.describe(include="all").T
```

```
df1.info() # checking the datatype information on columns
```

```
df1.nunique().to_frame("Unique Values")
```

I used a for loop to take a look at all the unique values present in the categorical columns covering the number of rows in the dataset.

```
for i in df1.columns:
    print(i)
    print(df1[i].value_counts().to_frame("Values Counts"))
    print("*"*120)
```

After checking for unique values, I checked if dataset contains any duplicates values or not and I found some.

```
df1.duplicated().sum()
```

```
166
```

But I kept these duplicates because it is possible that respondents may have same choices and deleting this much would lead to too much loss of data

Visualization:

What is Data Visualization?
Data visualization is defined as a graphical representation that contains the information and the data.

Benefits of Good Data Visualization?
Data visualization is another technique of visual art that grabs our interest and keeps our main focus on the message captured with the help of eyes.

Different Types of Analysis for Data Visualization are:

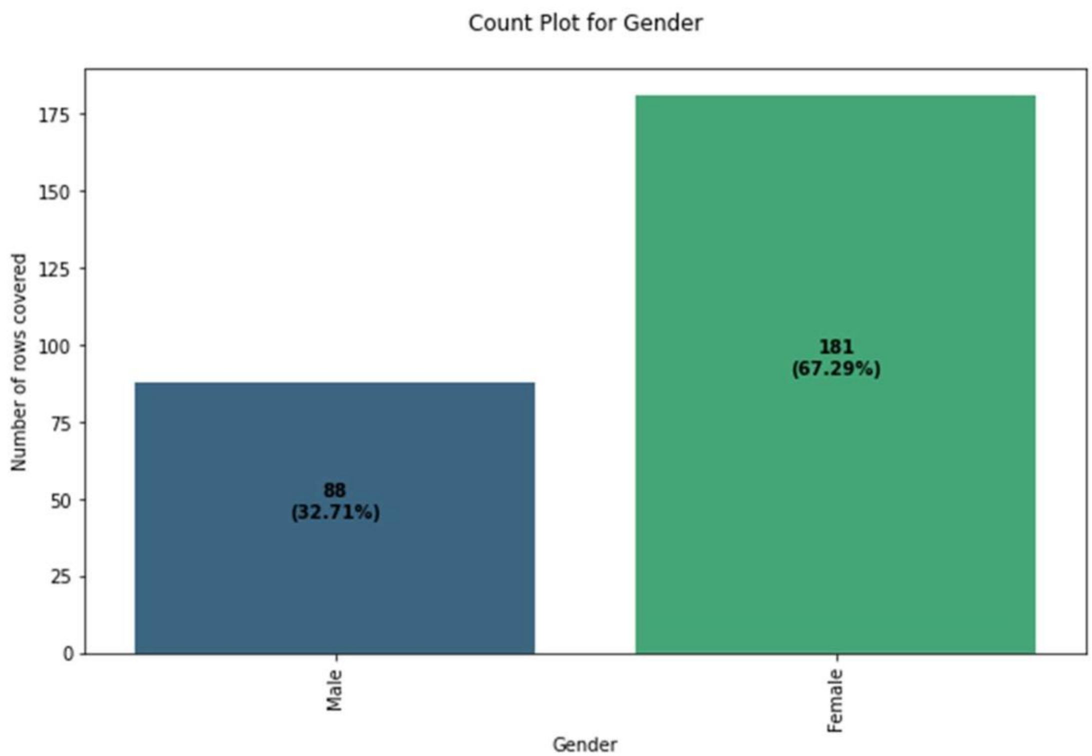
1. Univariate Analysis: In the univariate analysis, we will be using a single feature to analyze almost all of its properties.
2. Bivariate Analysis: When we compare the data between exactly 2 features then it is known as bivariate analysis.
3. Multivariate Analysis: In the multivariate analysis, we will be comparing more than 2 variables.

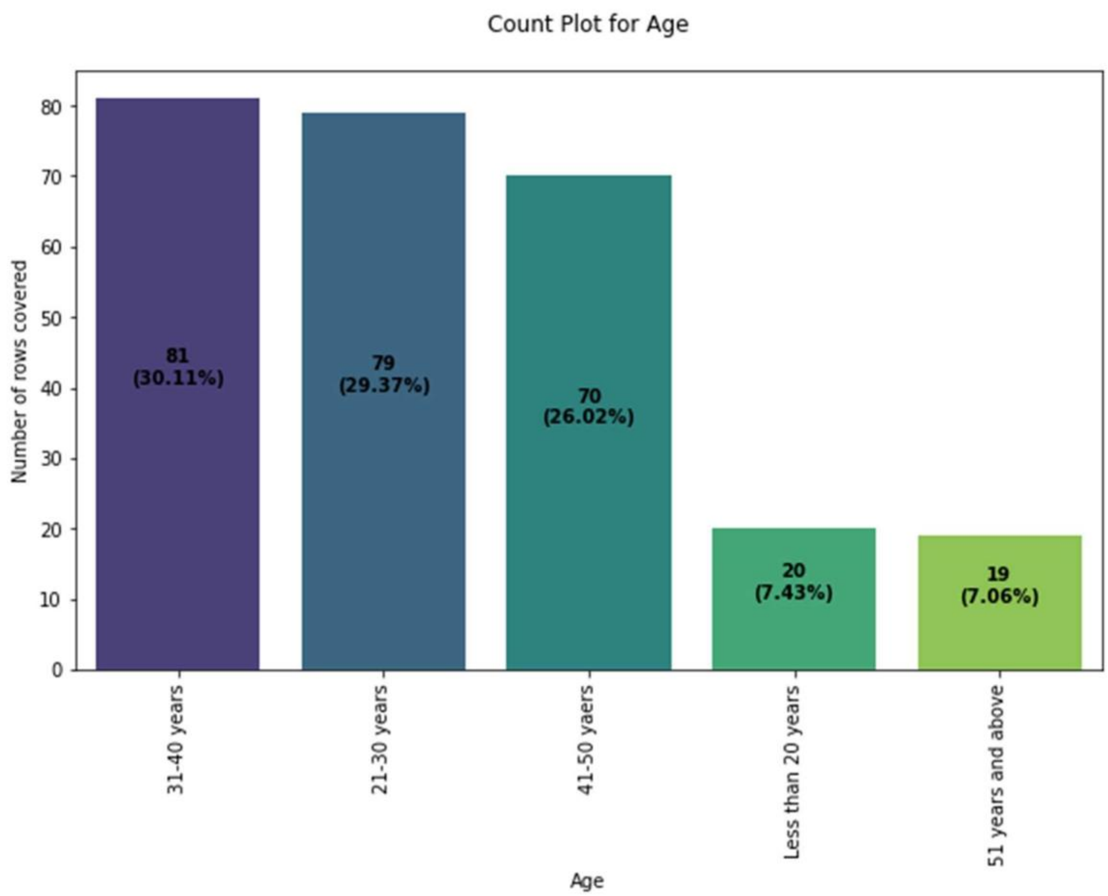
Univariate Analysis:

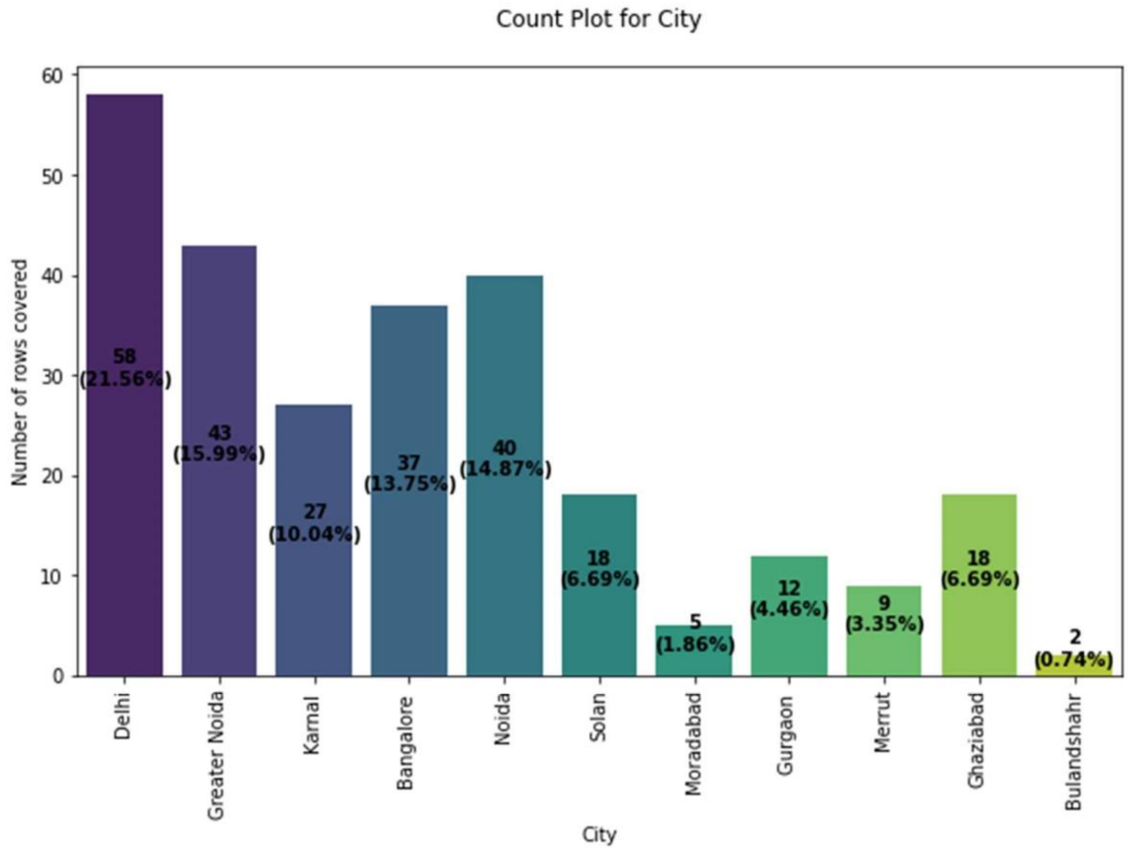
I made use of 2 for loops to generate count plots for all our columns showing the percentage of data coverage.

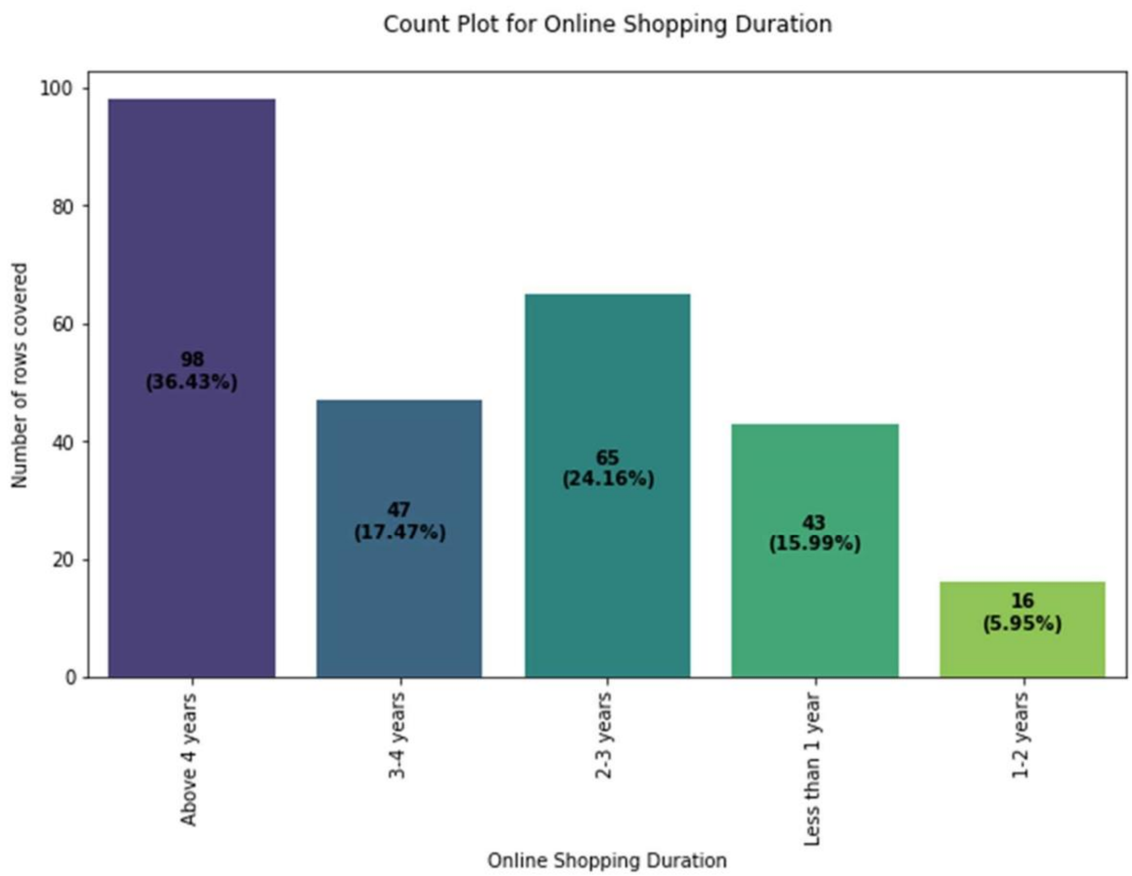
```
for col in df1[object_datatype]:
    plt.figure(figsize=(10,6))
    col_name = col
    values = df1[col_name].value_counts()
    index = 0
    ax = sns.countplot(df1[col_name], palette="viridis")

    for i in ax.patches:
        h = i.get_height() # getting the count of each value
        t = len(df1[col_name]) # getting the total number of records using length
        s = f"{h}\n({round(h*100/t,2)}%)" # making the string for displaying in count bar
        plt.text(index, h/2, s, ha="center", fontweight="bold")
        index += 1

    plt.title(f"Count Plot for {col_name}\n")
    plt.xlabel(col_name)
    plt.ylabel(f"Number of rows covered")
    plt.xticks(rotation=90)
    plt.show()
```

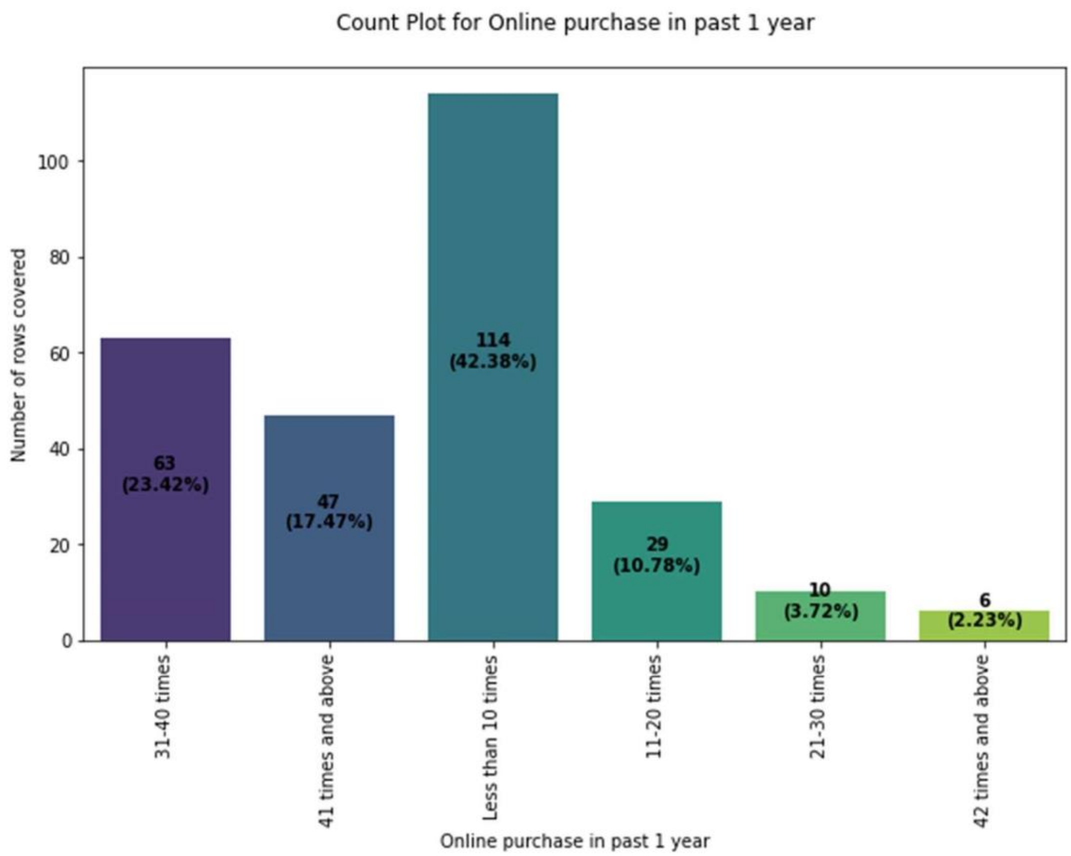
This piece of code generated multiple count plot images as displayed below.

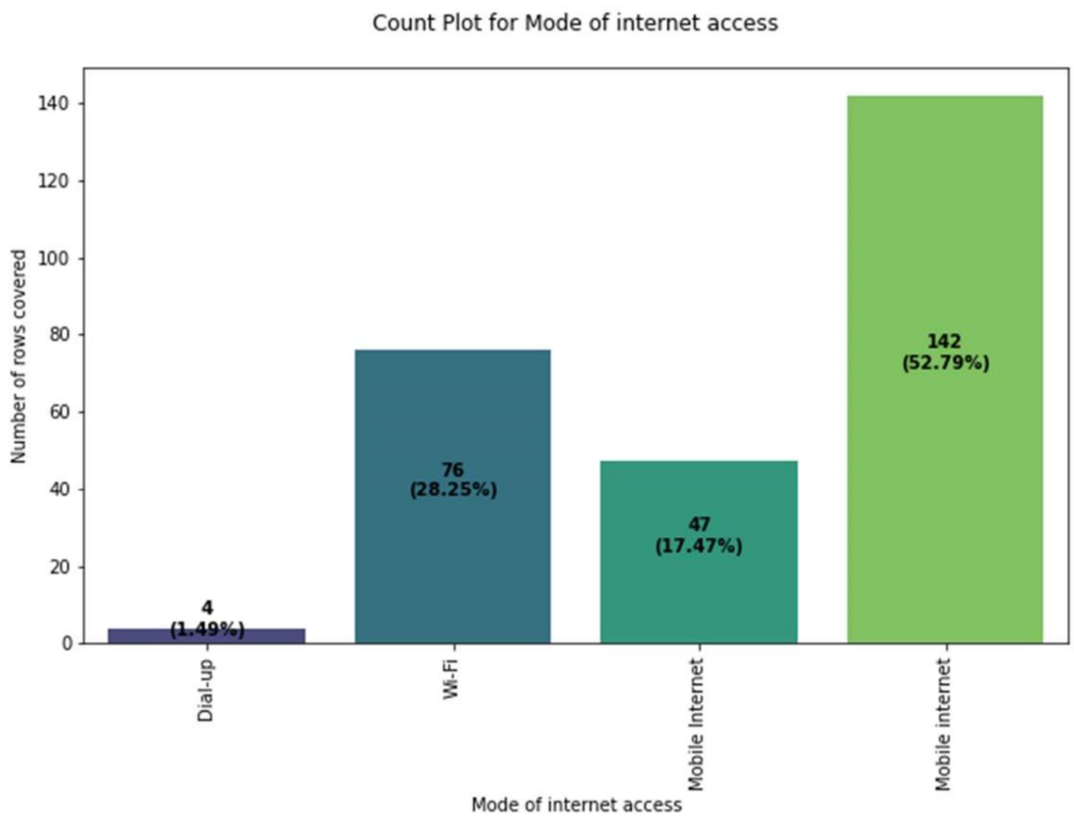

Count Plot for Gender
Number of rows covered
175
150
125
100
75
50
25
0
88
(32.71%)
181
(67.29%)
Male
Female
Gender

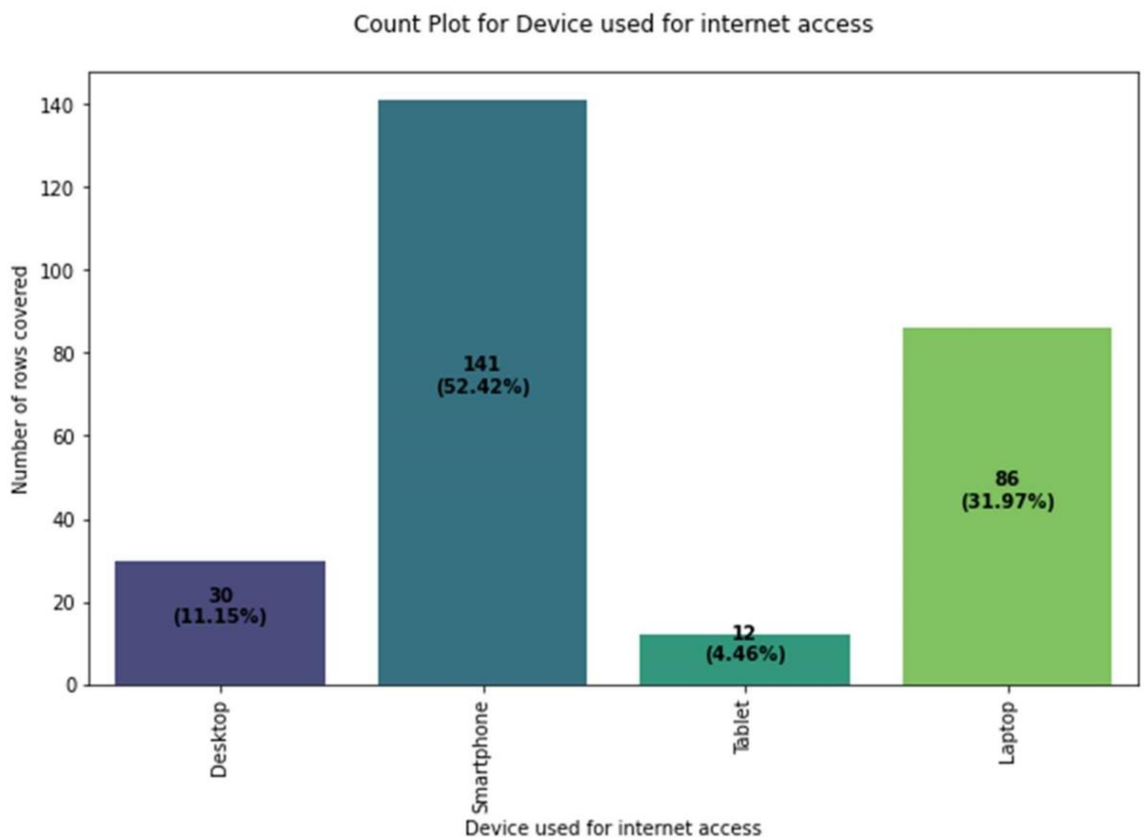

Count Plot for Age
Number of rows covered
80
70
60
50
40
30
20
10
0
81
(30.11%)
79
(29.37%)
70
(26.02%)
20
(7.43%)
19
(7.06%)
31-40 years
21-30 years
41-50 yaers
Less than 20 years
51 years and above
Age

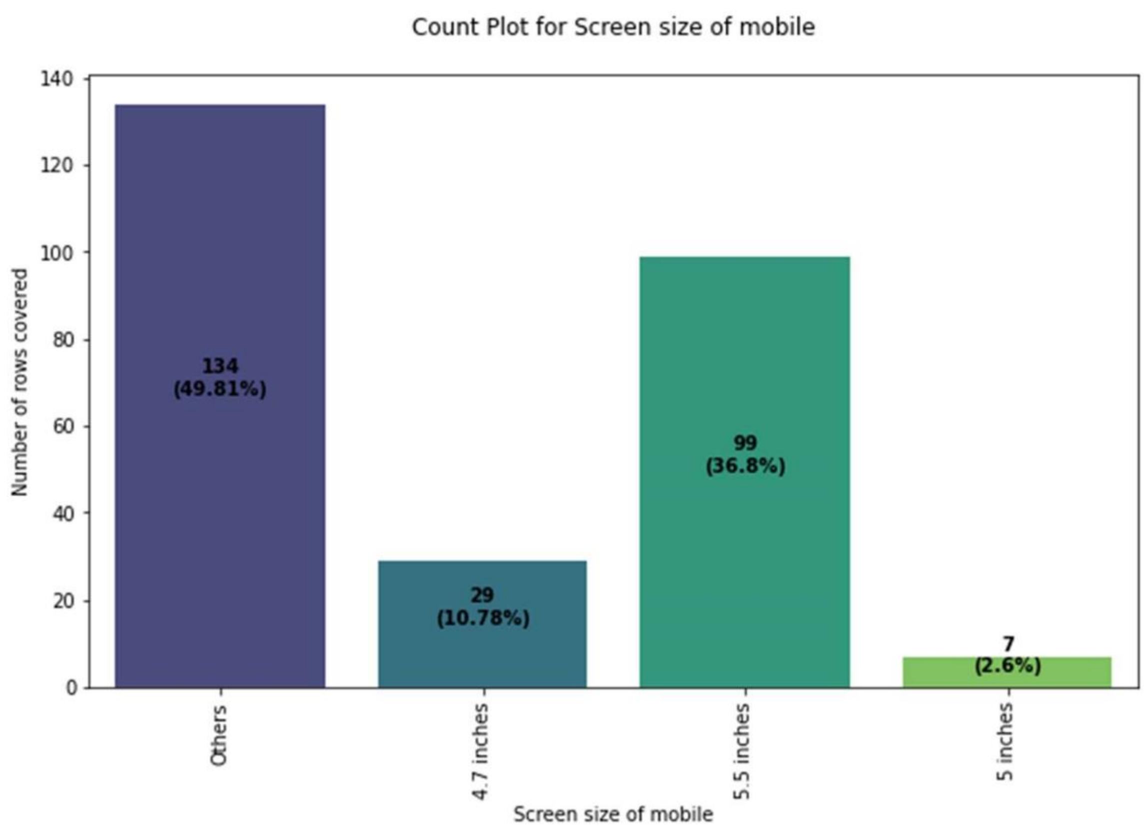

Count Plot for City
Number of rows covered
City
Delhi
Greater Noida
Karnal
Bangalore
Noida
Solan
Moradabad
Gurgaon
Merrut
Ghaziabad
Bulandshahr
58
(21.56%)
43
(15.99%)
27
(10.04%)
37
(13.75%)
40
(14.87%)
18
(6.69%)
5
(1.86%)
12
(4.46%)
9
(3.35%)
18
(6.69%)
2
(0.74%)

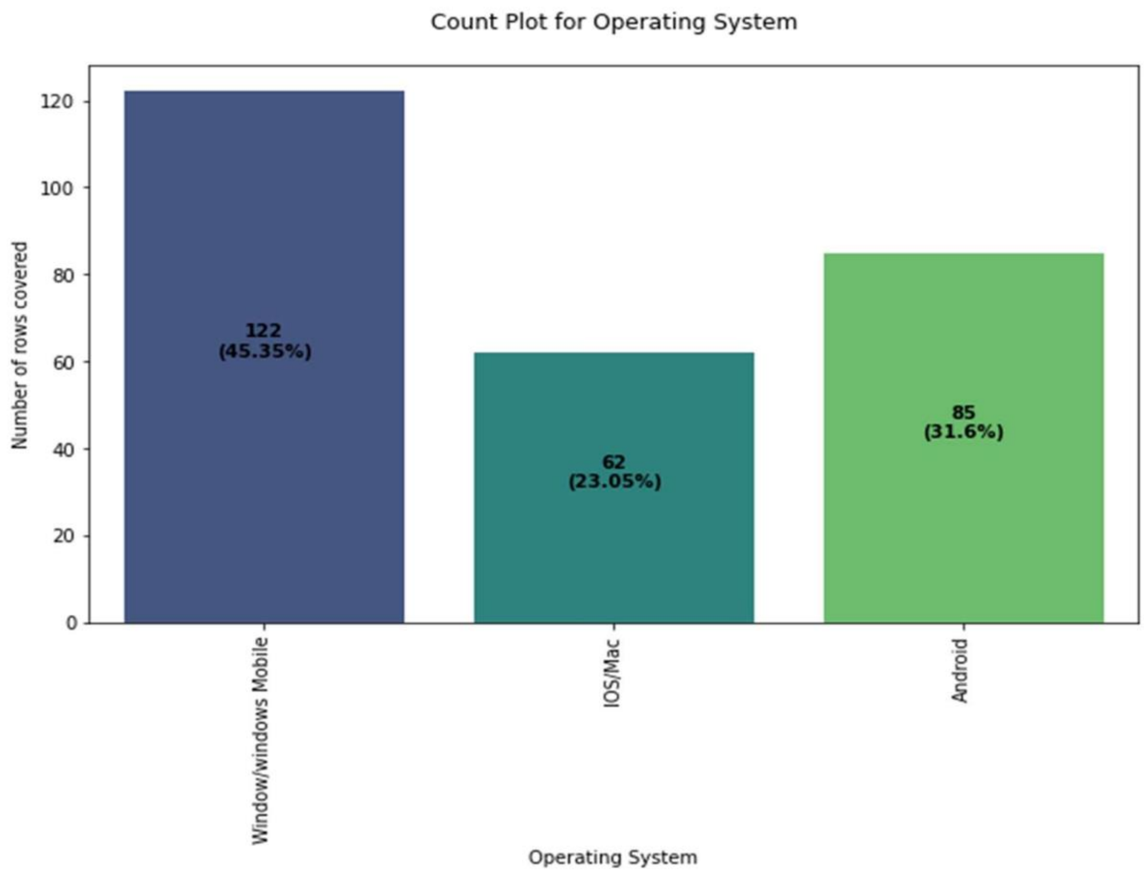

Count Plot for Online Shopping Duration
Number of rows covered
Online Shopping Duration
Above 4 years
3-4 years
2-3 years
Less than 1 year
1-2 years
98
(36.43%)
47
(17.47%)
65
(24.16%)
43
(15.99%)
16
(5.95%)


Count Plot for Online purchase in past 1 year
Number of rows covered
0
20
40
60
80
100
63
(23.42%)
47
(17.47%)
114
(42.38%)
29
(10.78%)
10
(3.72%)
6
(2.23%)
31-40 times
41 times and above
Less than 10 times
11-20 times
21-30 times
42 times and above
Online purchase in past 1 year

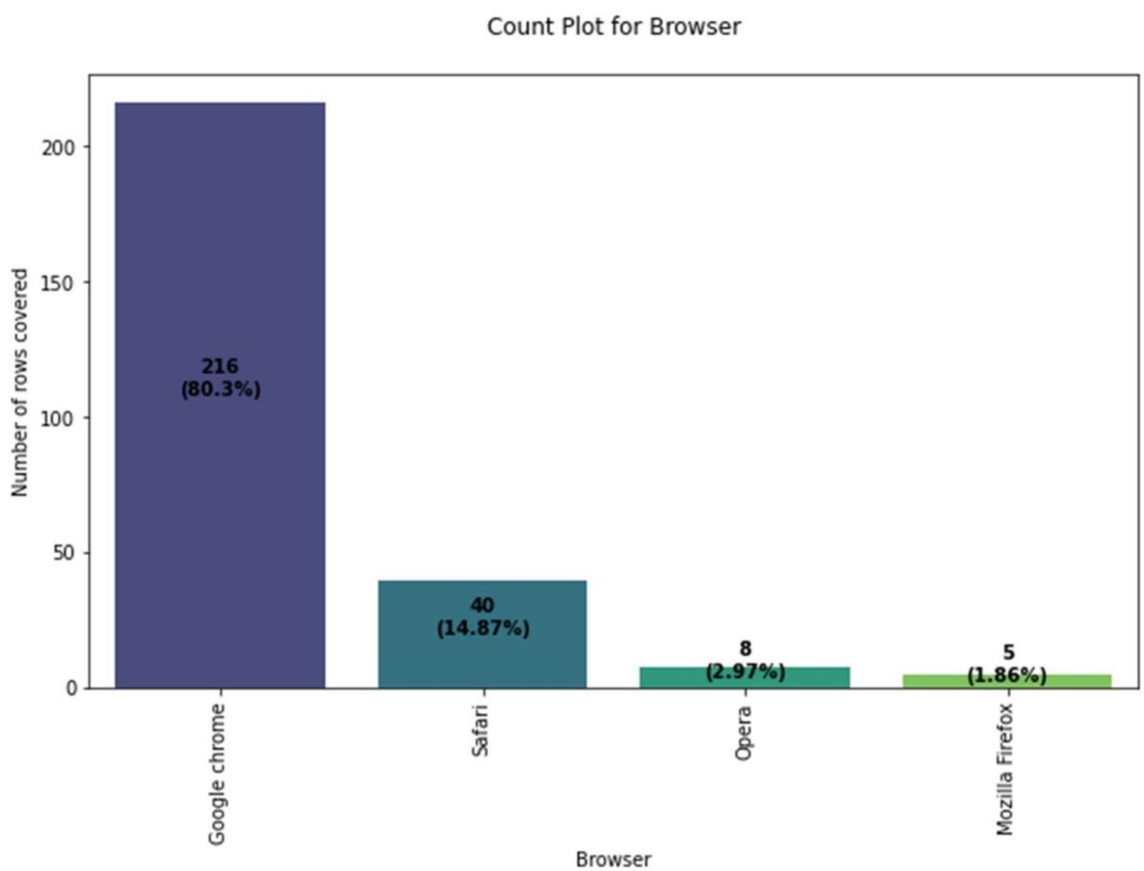

Count Plot for Mode of internet access
Number of rows covered
0
20
40
60
80
100
120
140
4
(1.49%)
76
(28.25%)
47
(17.47%)
142
(52.79%)
Dial-up
Wi-Fi
Mobile Internet
Mobile internet
Mode of internet access


Count Plot for Device used for internet access
Number of rows covered
0
20
40
60
80
100
120
140
30
(11.15%)
141
(52.42%)
12
(4.46%)
86
(31.97%)
Desktop
Smartphone
Tablet
Laptop
Device used for internet access


Count Plot for Screen size of mobile
Number of rows covered
0
20
40
60
80
100
120
140
134
(49.81%)
29
(10.78%)
99
(36.8%)
7
(2.6%)
Others
4.7 inches
5.5 inches
5 inches
Screen size of mobile

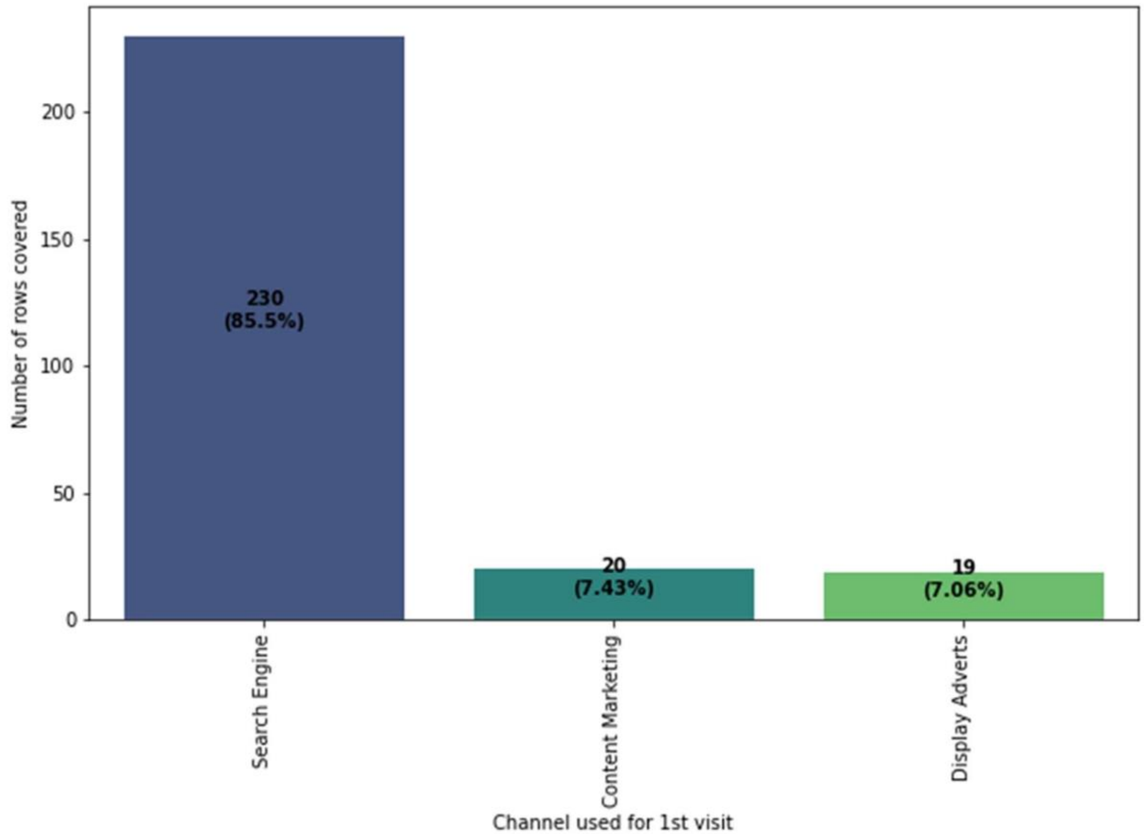


Count Plot for Operating System
Number of rows covered
122
(45.35%)
62
(23.05%)
85
(31.6%)
Window/windows Mobile
IOS/Mac
Android
Operating System

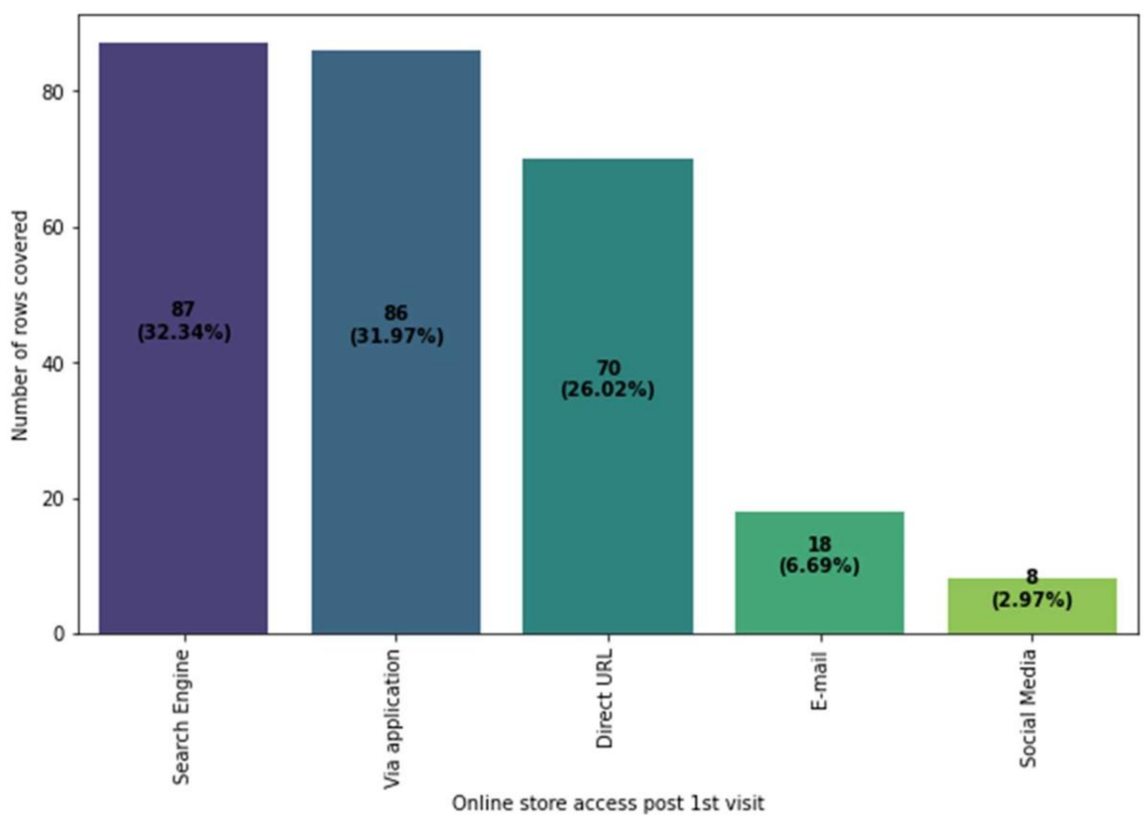


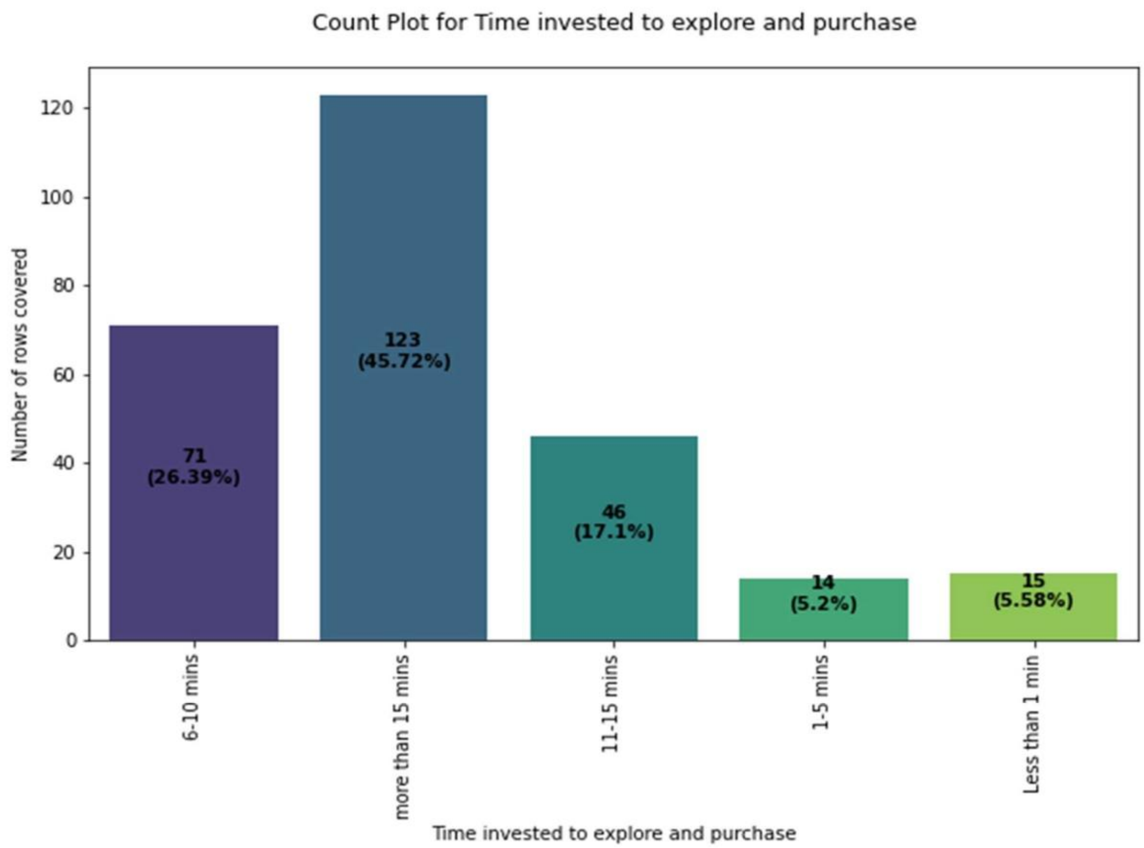
Count Plot for Browser
Number of rows covered
216
(80.3%)
40
(14.87%)
8
(2.97%)
5
(1.86%)
Google chrome
Safari
Opera
Mozilla Firefox
Browser

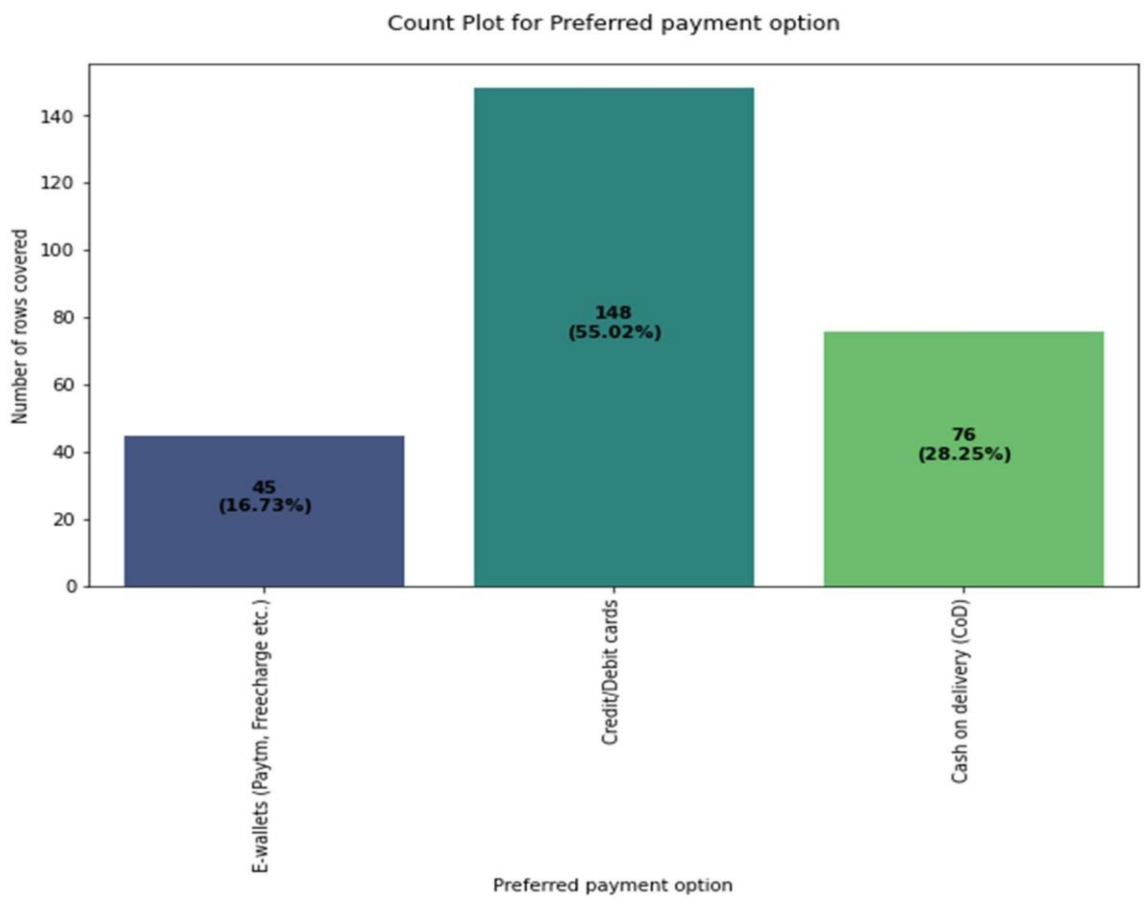
Count Plot for Channel used for 1st visit

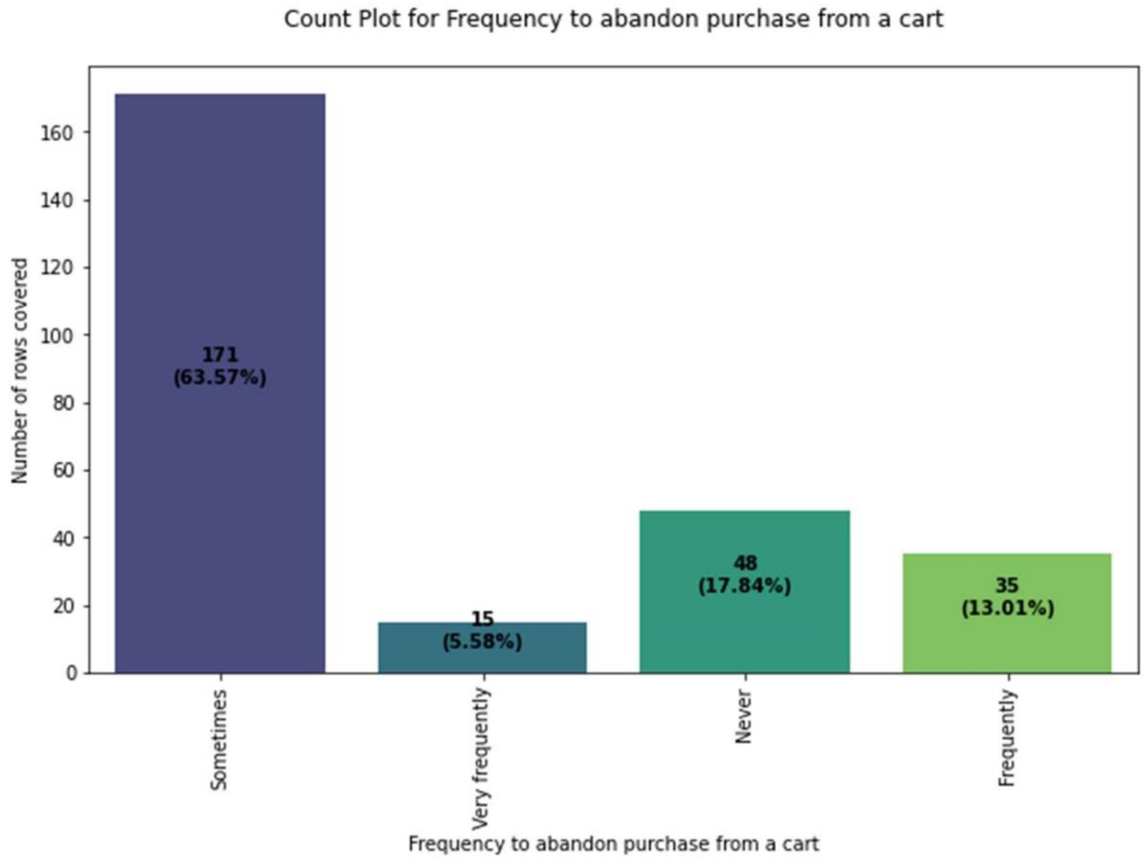


Count Plot for Online store access post 1st visit

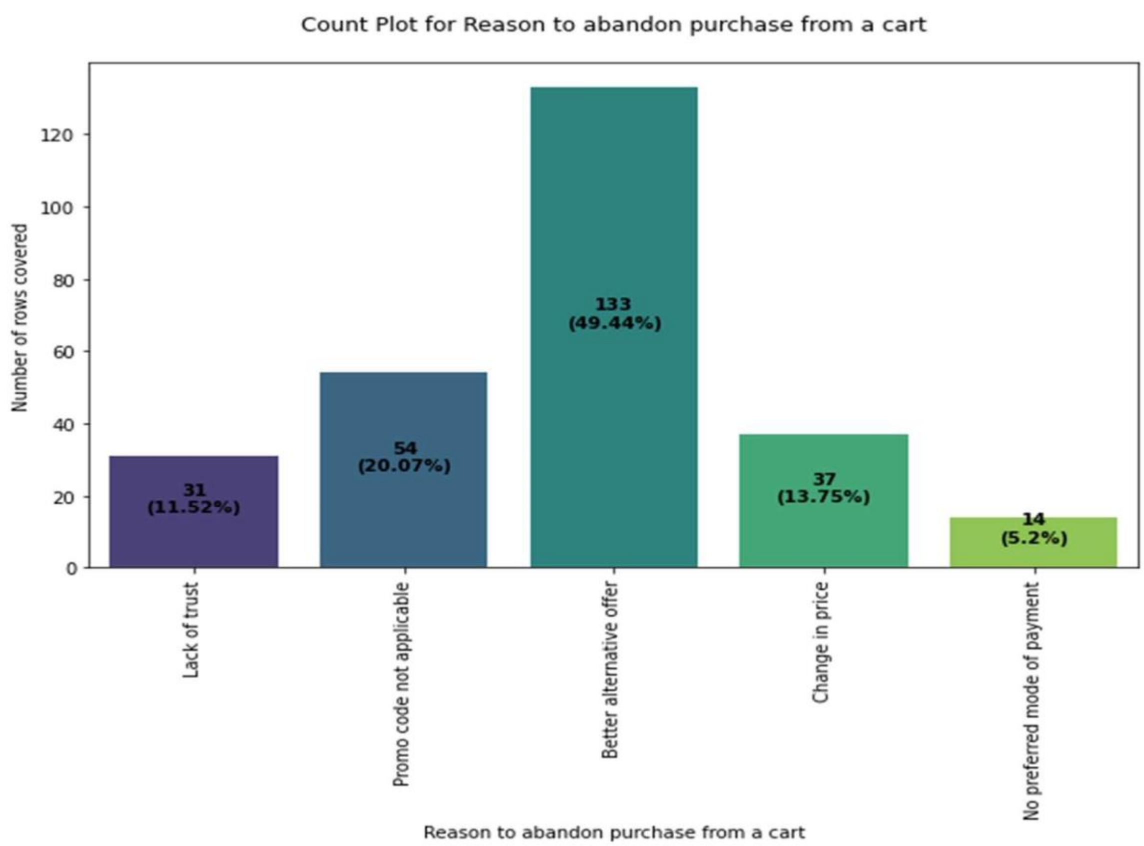



Count Plot for Time invested to explore and purchase
Number of rows covered
71
(26.39%)
123
(45.72%)
46
(17.1%)
14
(5.2%)
15
(5.58%)
6-10 mins
more than 15 mins
11-15 mins
1-5 mins
Less than 1 min
Time invested to explore and purchase


Count Plot for Preferred payment option
Number of rows covered
45
(16.73%)
148
(55.02%)
76
(28.25%)
E-wallets (Paytm, Freecharge etc.)
Credit/Debit cards
Cash on delivery (CoD)
Preferred payment option

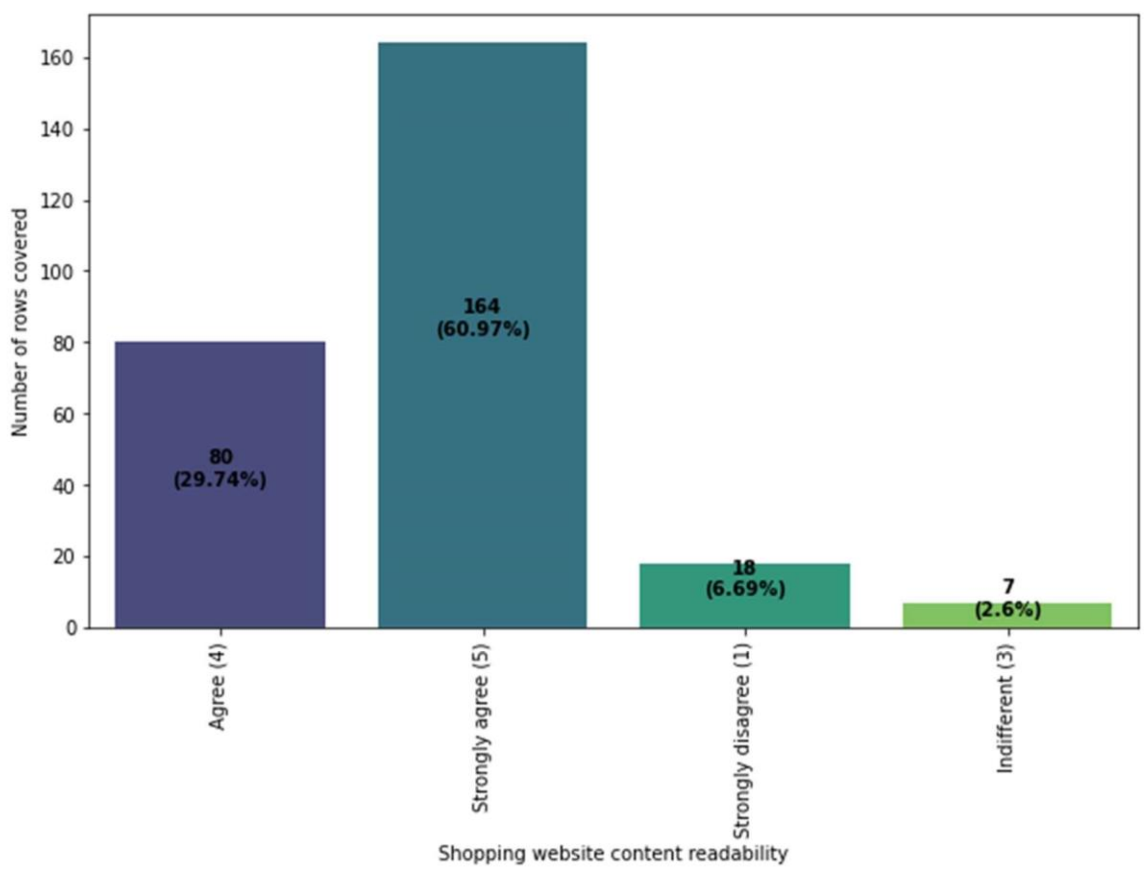


Count Plot for Frequency to abandon purchase from a cart
Number of rows covered
171
(63.57%)
15
(5.58%)
48
(17.84%)
35
(13.01%)
Sometimes
Very frequently
Never
Frequently
Frequency to abandon purchase from a cart

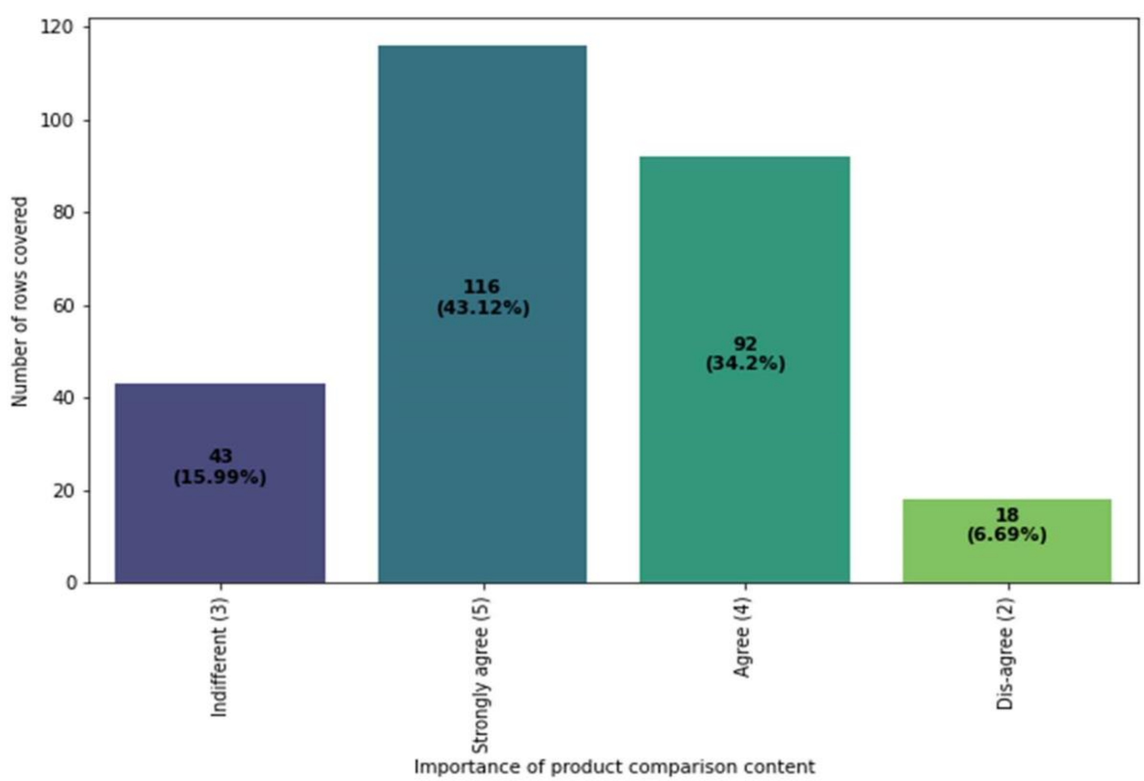


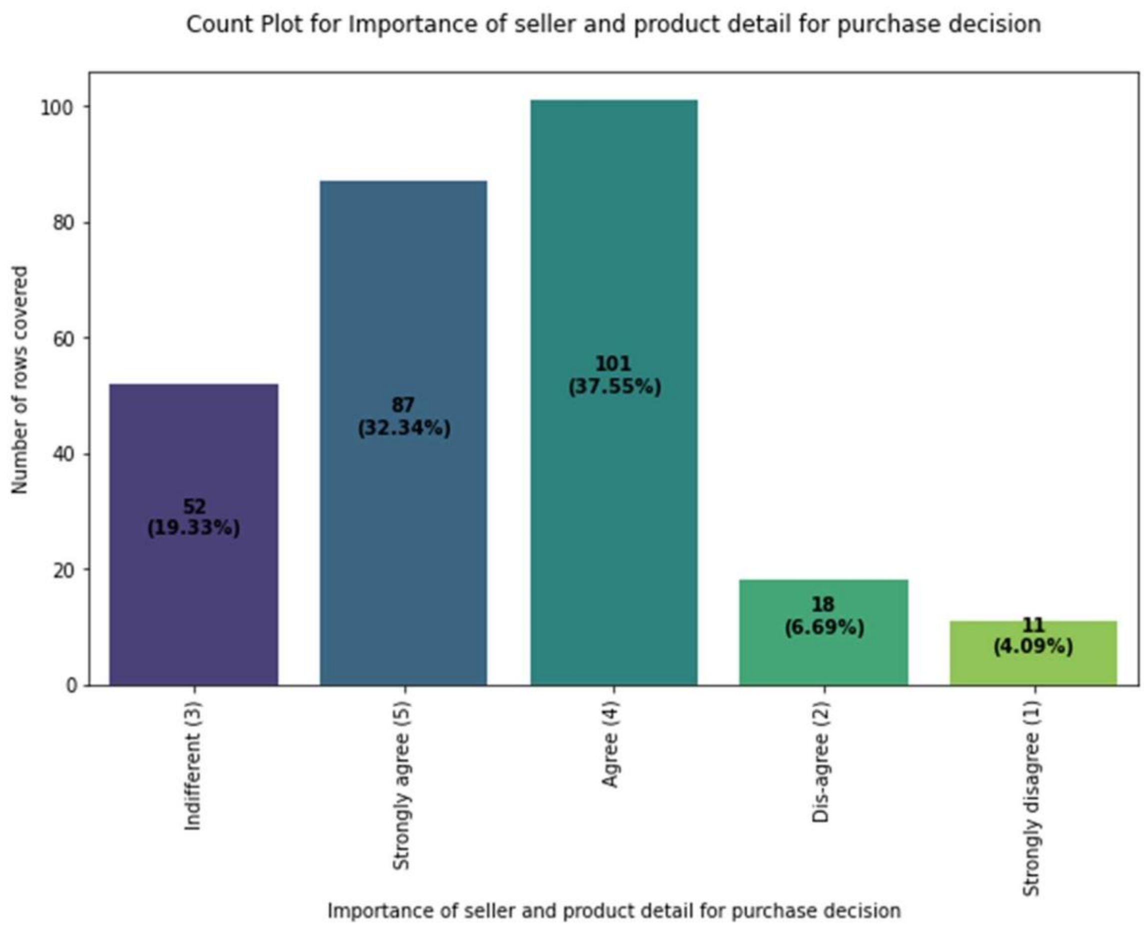
Count Plot for Reason to abandon purchase from a cart
Number of rows covered
31
(11.52%)
54
(20.07%)
133
(49.44%)
37
(13.75%)
14
(5.2%)
Lack of trust
Promo code not applicable
Better alternative offer
Change in price
No preferred mode of payment
Reason to abandon purchase from a cart

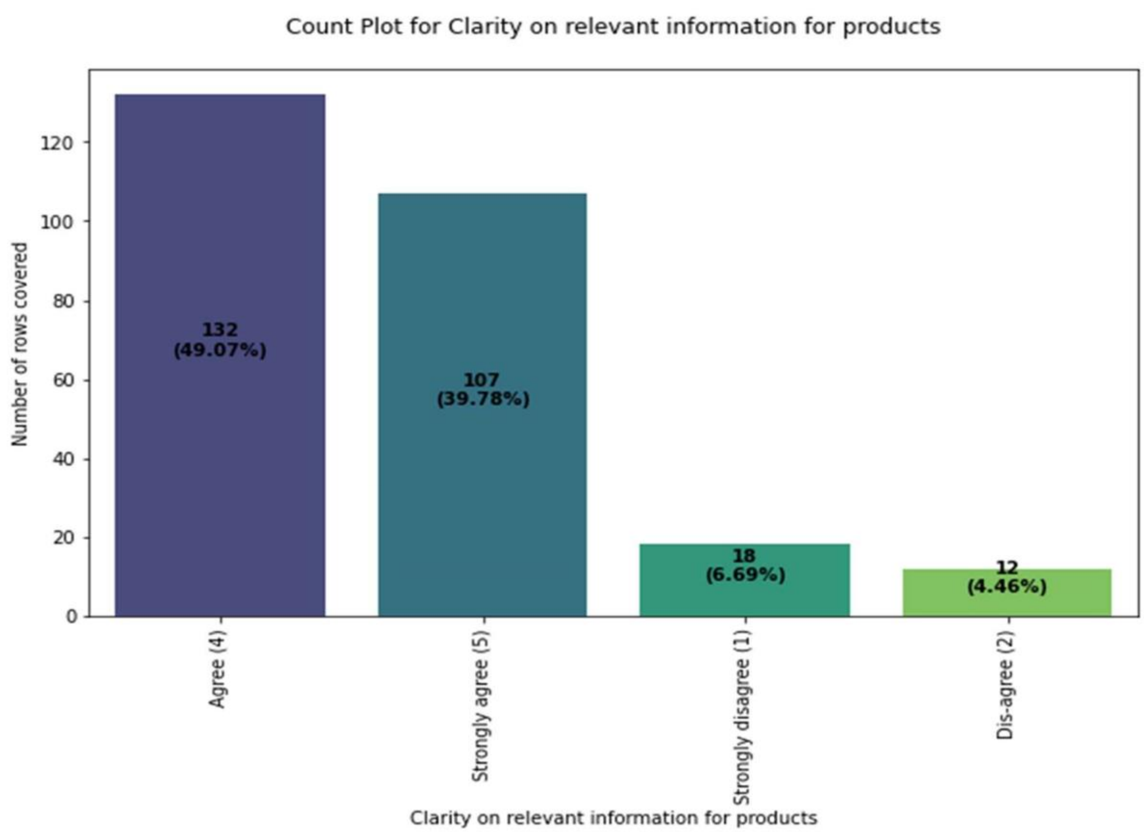
Count Plot for Shopping website content readability

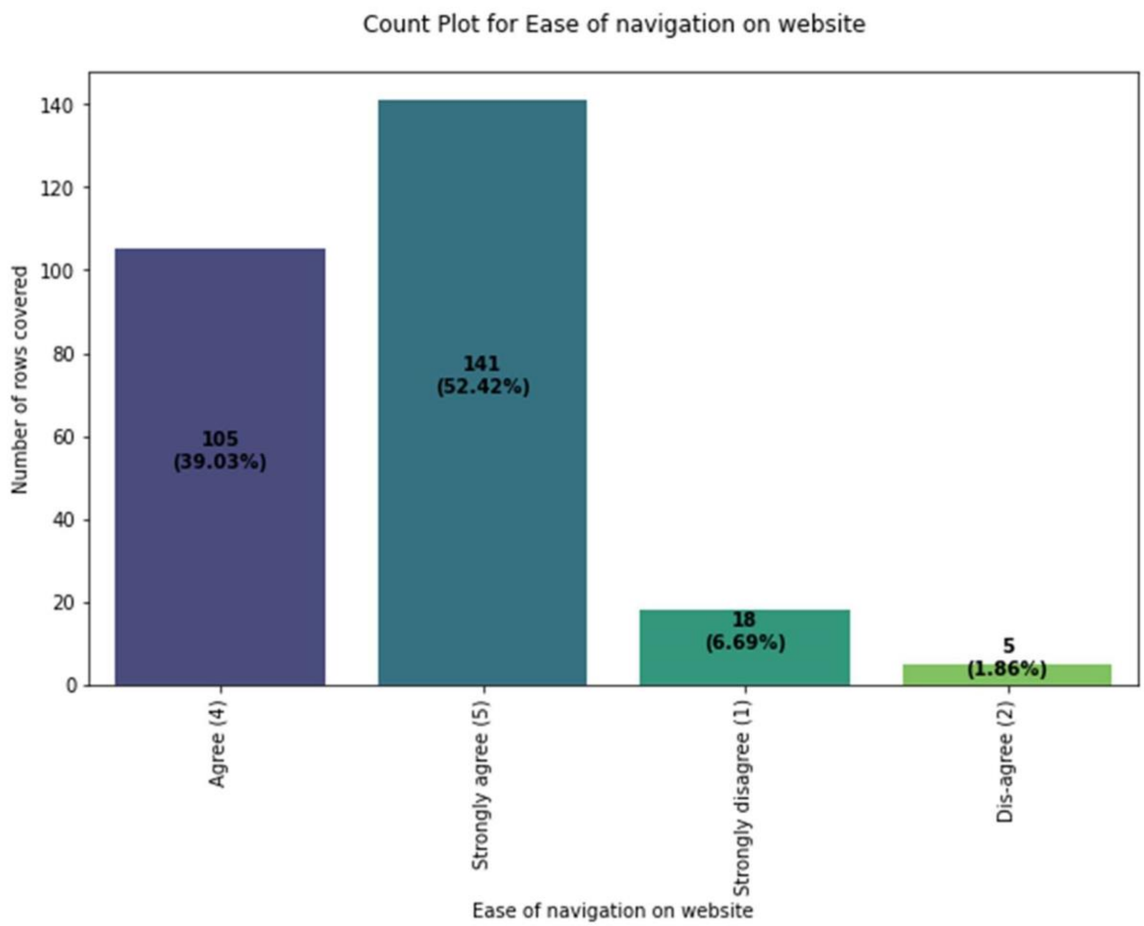


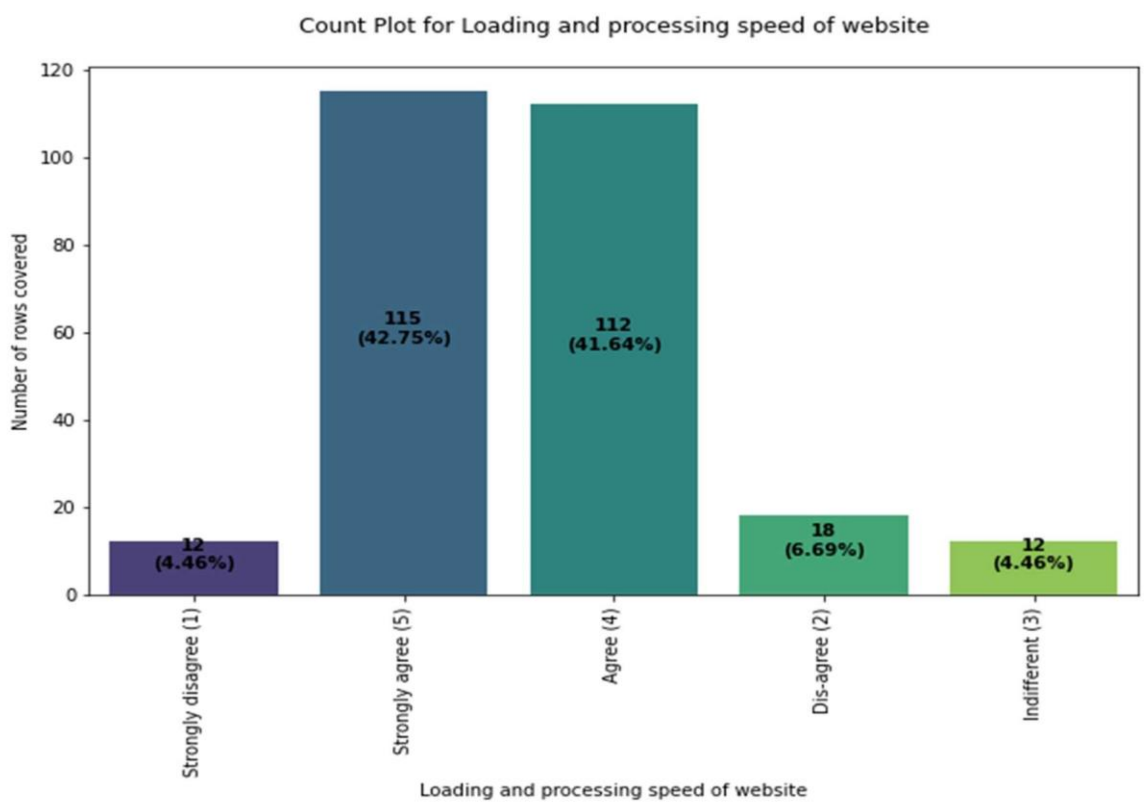
Count Plot for Importance of product comparison content

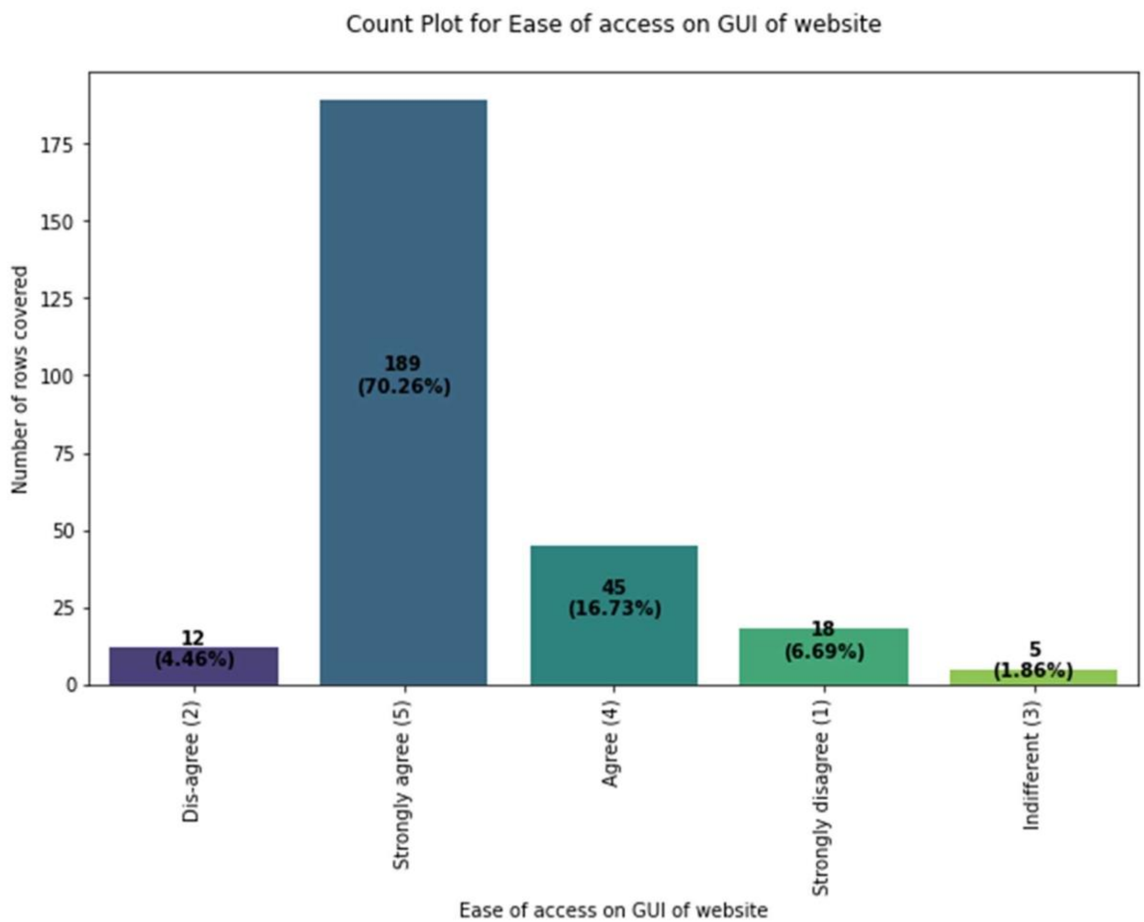


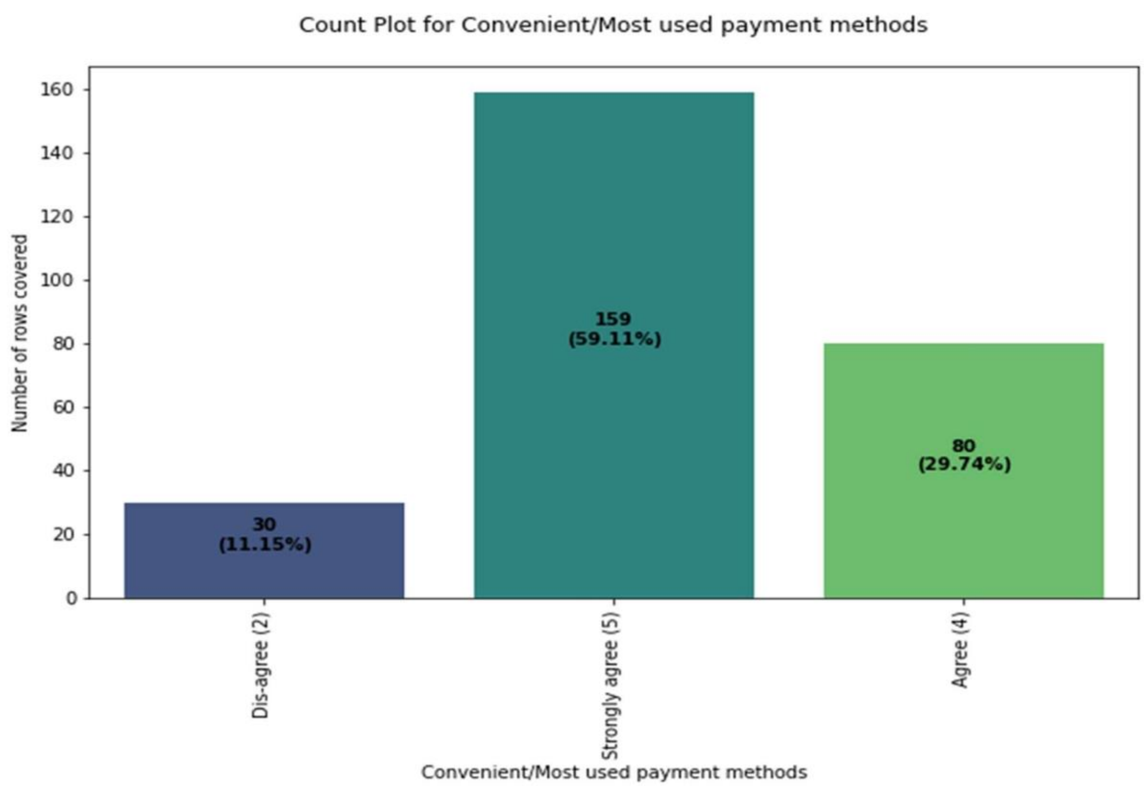

Count Plot for Importance of seller and product detail for purchase decision
Number of rows covered
0
20
40
60
80
100
52
(19.33%)
87
(32.34%)
101
(37.55%)
18
(6.69%)
11
(4.09%)
Indifferent (3)
Strongly agree (5)
Agree (4)
Dis-agree (2)
Strongly disagree (1)
Importance of seller and product detail for purchase decision

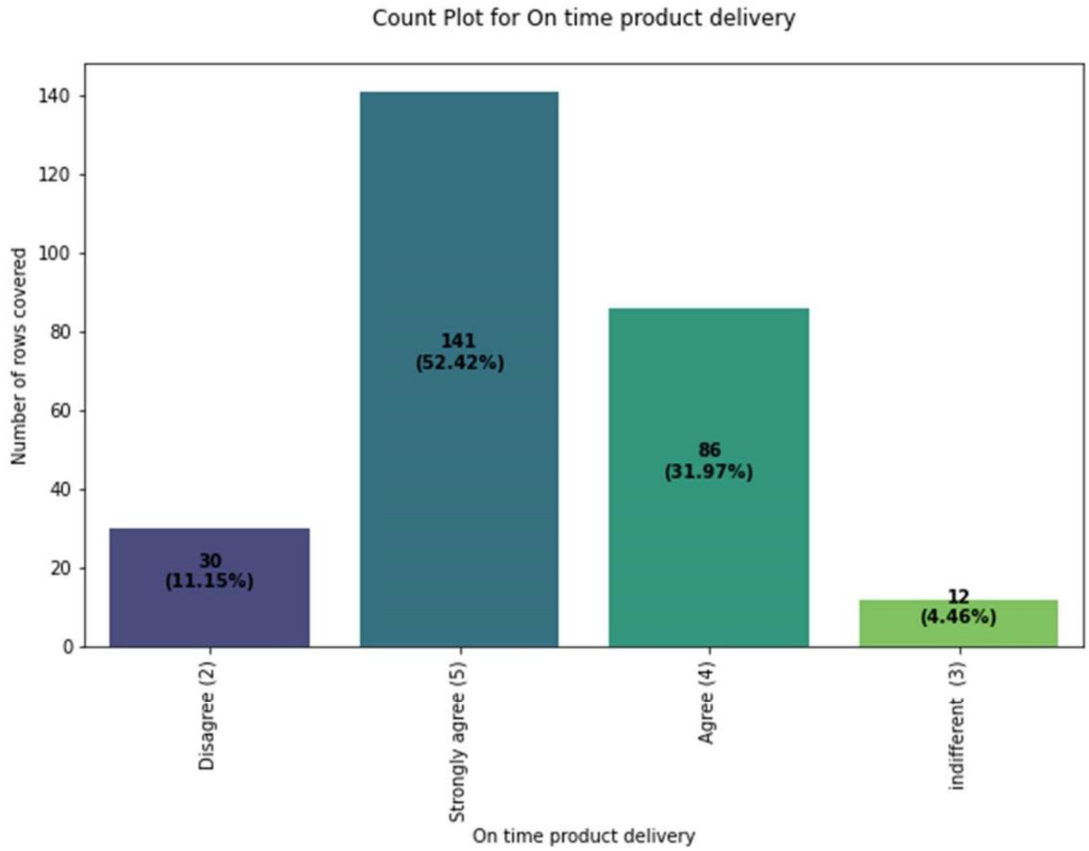

Count Plot for Clarity on relevant information for products
Number of rows covered
0
20
40
60
80
100
120
132
(49.07%)
107
(39.78%)
18
(6.69%)
12
(4.46%)
Agree (4)
Strongly agree (5)
Strongly disagree (1)
Dis-agree (2)
Clarity on relevant information for products

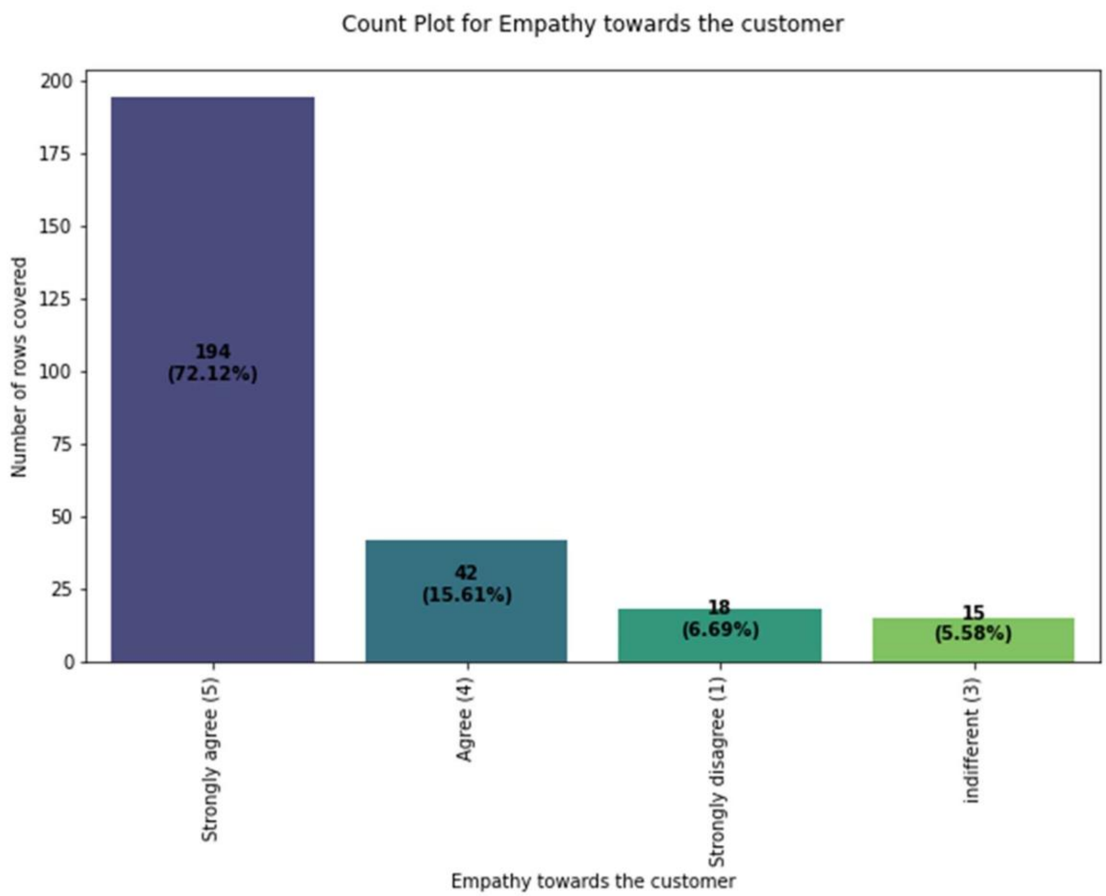

Count Plot for Ease of navigation on website
Number of rows covered
0
20
40
60
80
100
120
140
105
(39.03%)
141
(52.42%)
18
(6.69%)
5
(1.86%)
Agree (4)
Strongly agree (5)
Strongly disagree (1)
Dis-agree (2)
Ease of navigation on website

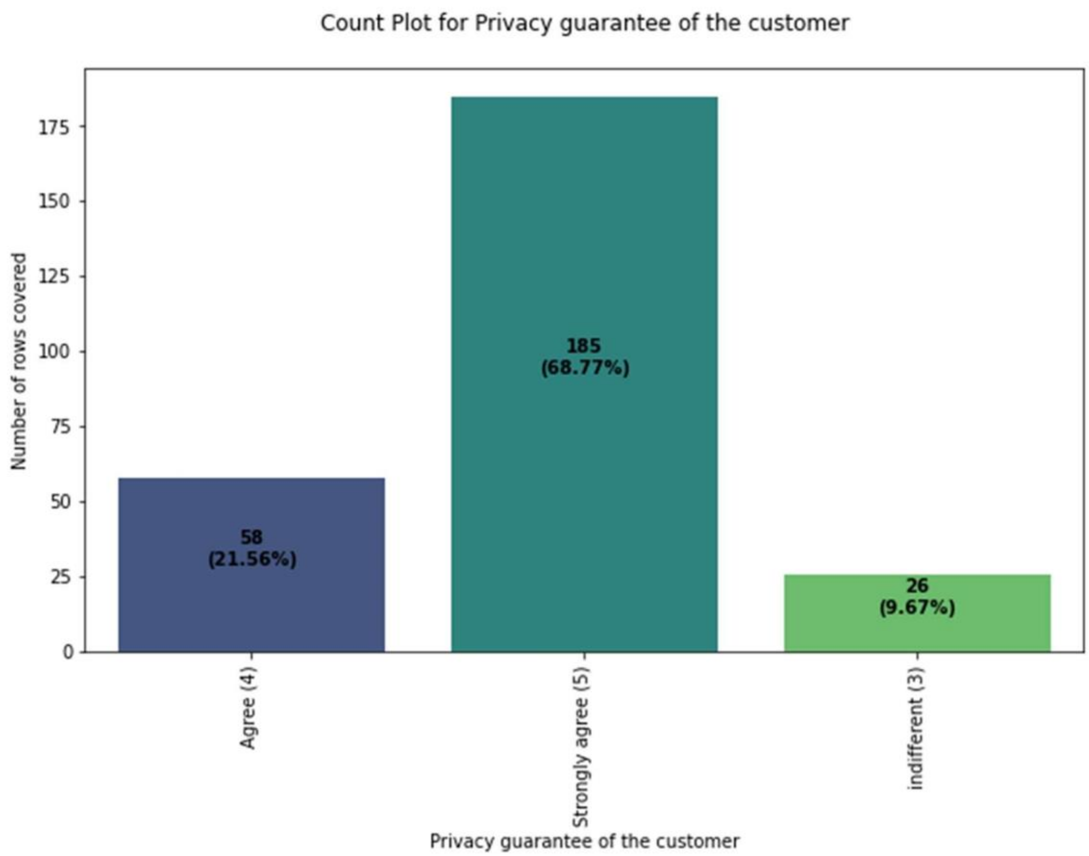

Count Plot for Loading and processing speed of website
Number of rows covered
0
20
40
60
80
100
120
12
(4.46%)
115
(42.75%)
112
(41.64%)
18
(6.69%)
12
(4.46%)
Strongly disagree (1)
Strongly agree (5)
Agree (4)
Dis-agree (2)
Indifferent (3)
Loading and processing speed of website

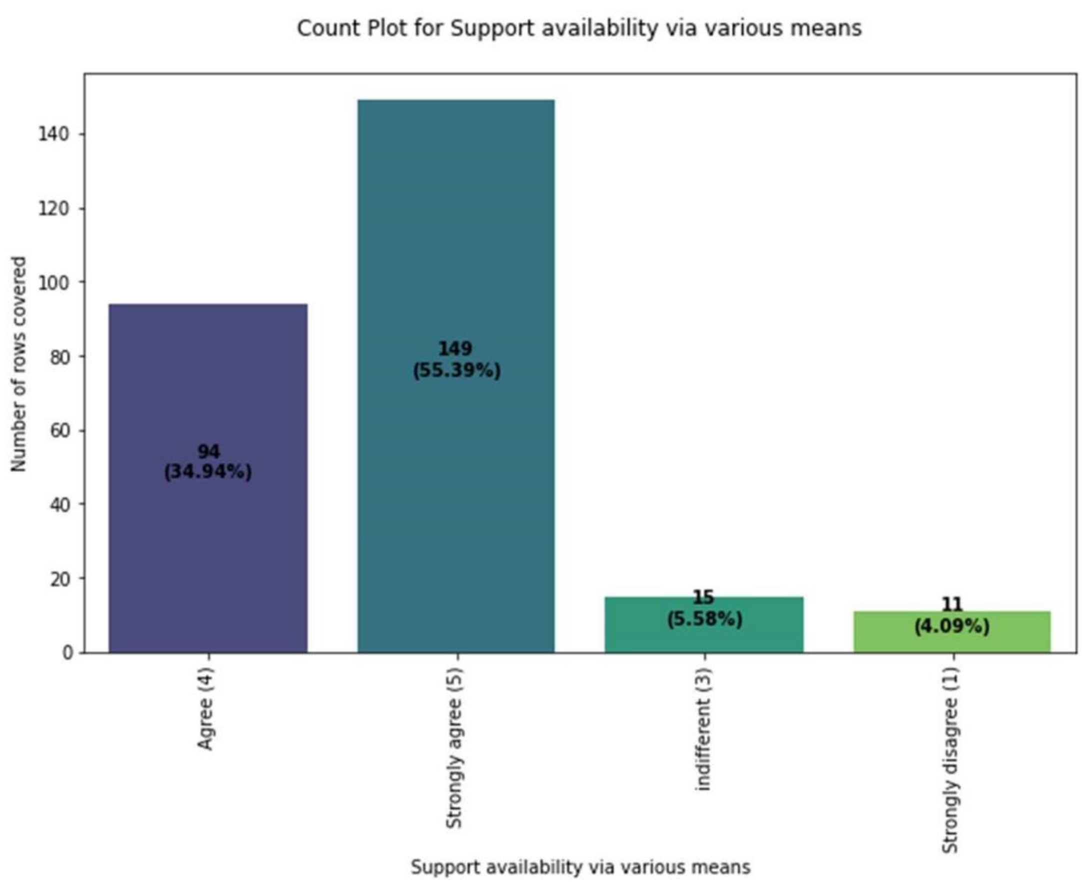

Count Plot for Ease of access on GUI of website
Number of rows covered
0
25
50
75
100
125
150
175
12
(4.46%)
189
(70.26%)
45
(16.73%)
18
(6.69%)
5
(1.86%)
Dis-agree (2)
Strongly agree (5)
Agree (4)
Strongly disagree (1)
Indifferent (3)
Ease of access on GUI of website

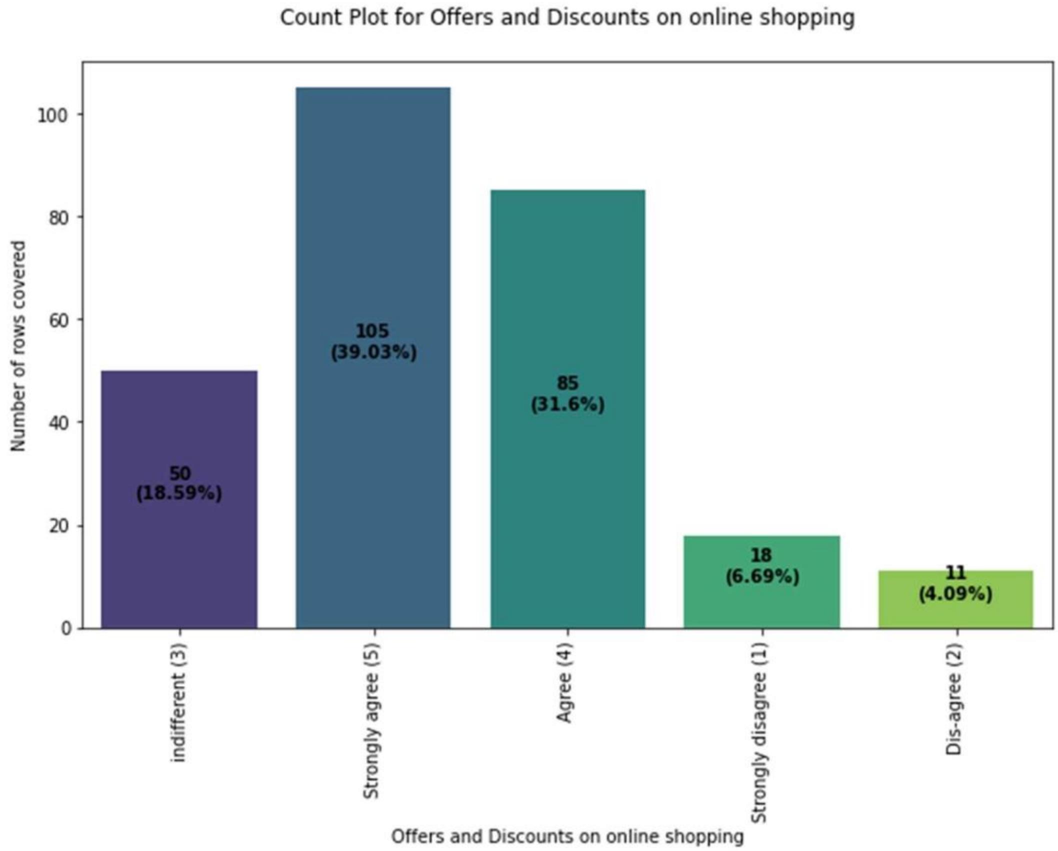

Count Plot for Convenient/Most used payment methods
Number of rows covered
0
20
40
60
80
100
120
140
160
30
(11.15%)
159
(59.11%)
80
(29.74%)
Dis-agree (2)
Strongly agree (5)
Agree (4)
Convenient/Most used payment methods

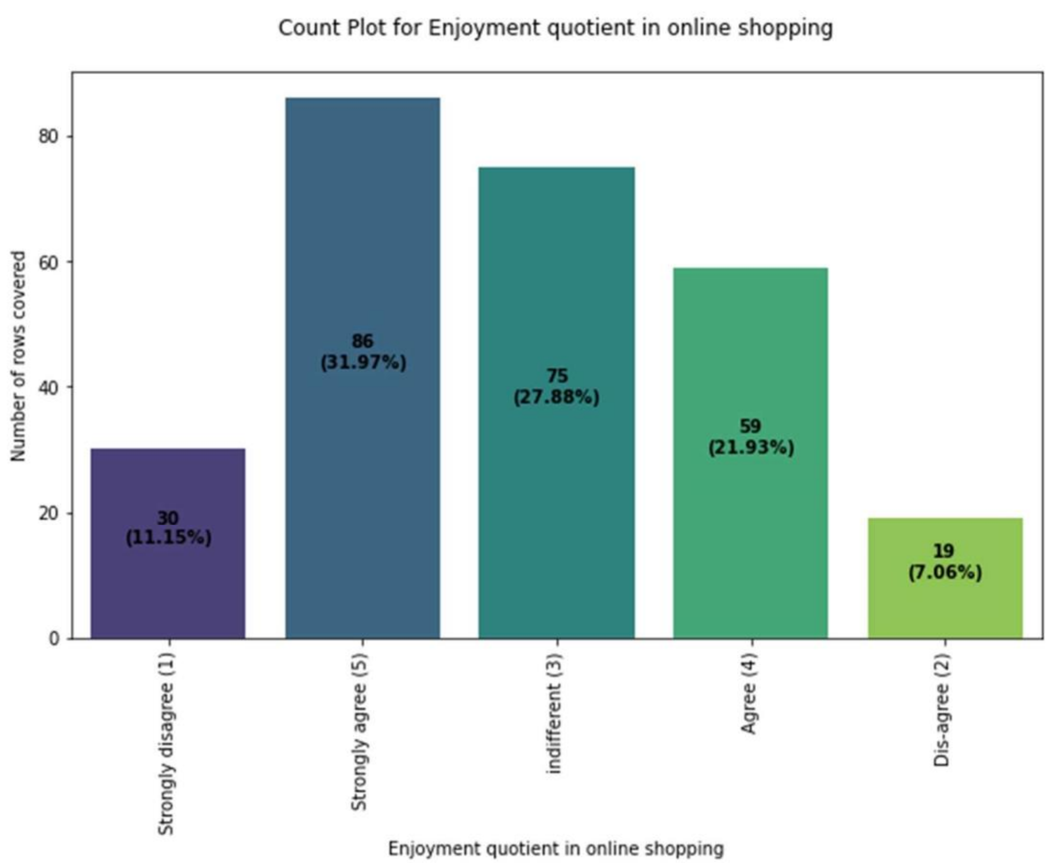

Count Plot for On time product delivery
Number of rows covered
0
20
40
60
80
100
120
140
30
(11.15%)
141
(52.42%)
86
(31.97%)
12
(4.46%)
Disagree (2)
Strongly agree (5)
Agree (4)
indifferent (3)
On time product delivery

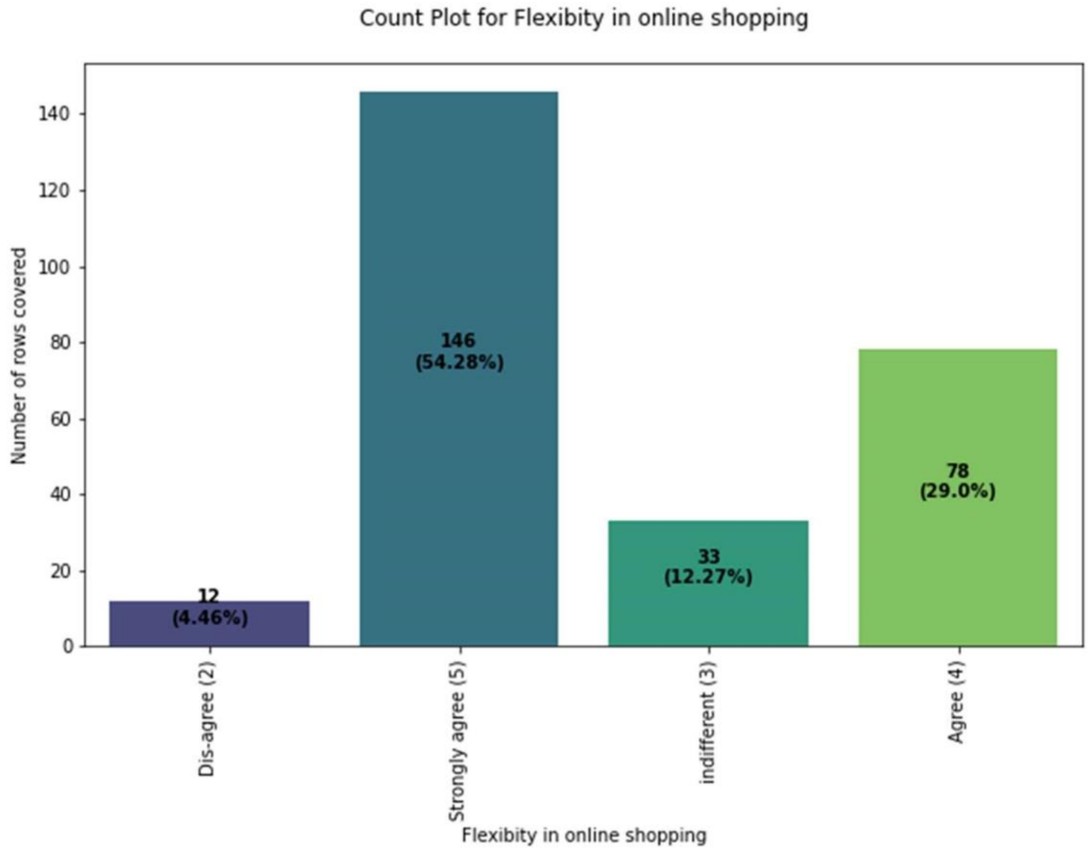

Count Plot for Empathy towards the customer
Number of rows covered
0
25
50
75
100
125
150
175
200
194
(72.12%)
42
(15.61%)
18
(6.69%)
15
(5.58%)
Strongly agree (5)
Agree (4)
Strongly disagree (1)
indifferent (3)
Empathy towards the customer


Count Plot for Privacy guarantee of the customer
Number of rows covered
0
25
50
75
100
125
150
175
58
(21.56%)
185
(68.77%)
26
(9.67%)
Agree (4)
Strongly agree (5)
indifferent (3)
Privacy guarantee of the customer

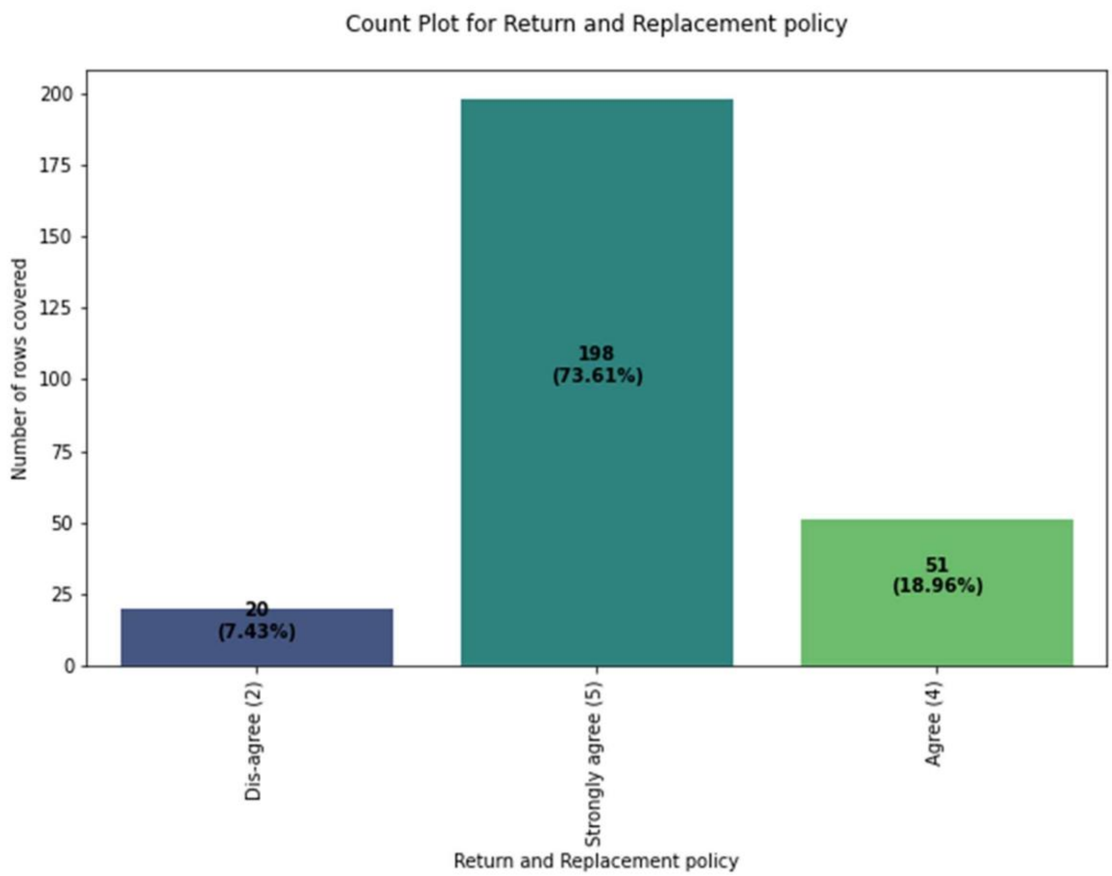

Count Plot for Support availability via various means
Number of rows covered
0
20
40
60
80
100
120
140
94
(34.94%)
149
(55.39%)
15
(5.58%)
11
(4.09%)
Agree (4)
Strongly agree (5)
indifferent (3)
Strongly disagree (1)
Support availability via various means


Count Plot for Offers and Discounts on online shopping
Number of rows covered
0
20
40
60
80
100
50
(18.59%)
105
(39.03%)
85
(31.6%)
18
(6.69%)
11
(4.09%)
indifferent (3)
Strongly agree (5)
Agree (4)
Strongly disagree (1)
Dis-agree (2)
Offers and Discounts on online shopping


Count Plot for Enjoyment quotient in online shopping
Number of rows covered
0
20
40
60
80
30
(11.15%)
86
(31.97%)
75
(27.88%)
59
(21.93%)
19
(7.06%)
Strongly disagree (1)
Strongly agree (5)
indifferent (3)
Agree (4)
Dis-agree (2)
Enjoyment quotient in online shopping

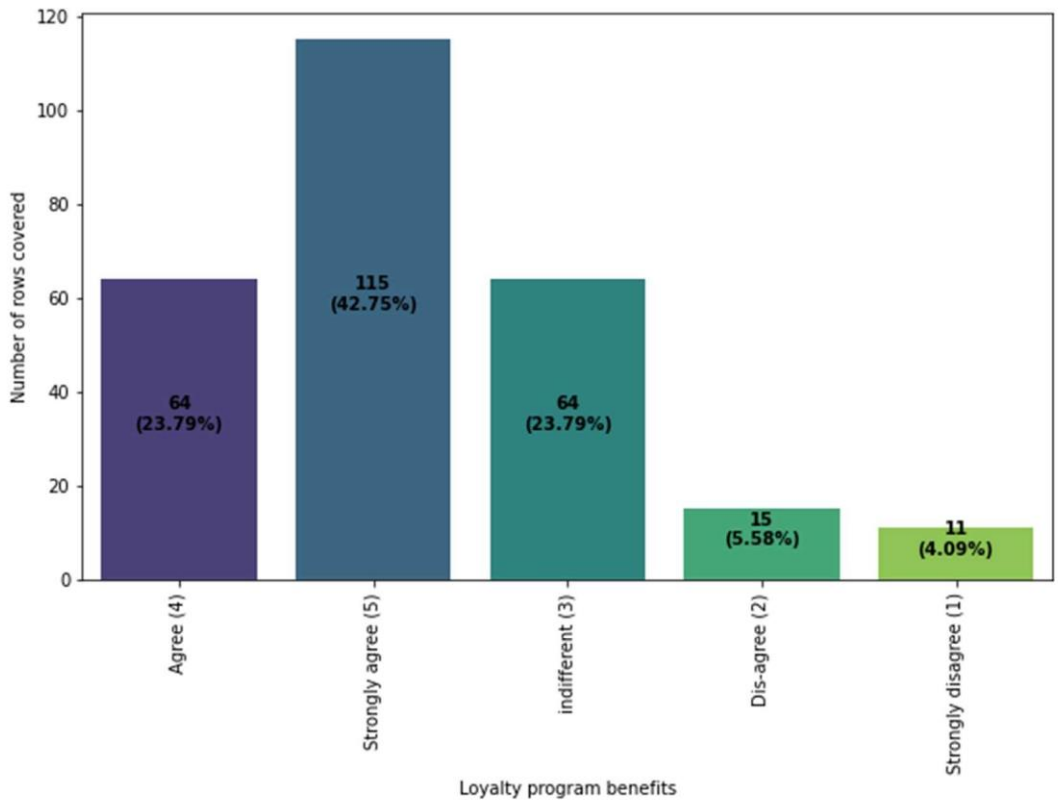


Count Plot for Flexibity in online shopping
Number of rows covered
0
20
40
60
80
100
120
140
12
(4.46%)
146
(54.28%)
33
(12.27%)
78
(29.0%)
Dis-agree (2)
Strongly agree (5)
indifferent (3)
Agree (4)
Flexibity in online shopping

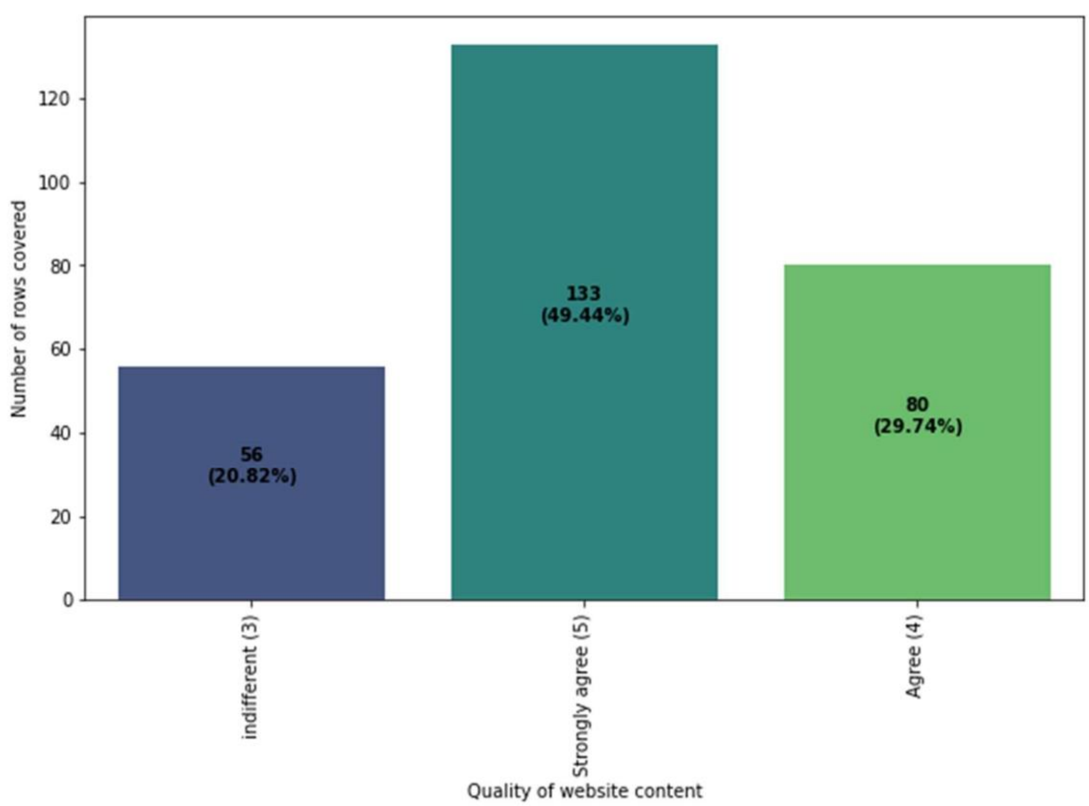


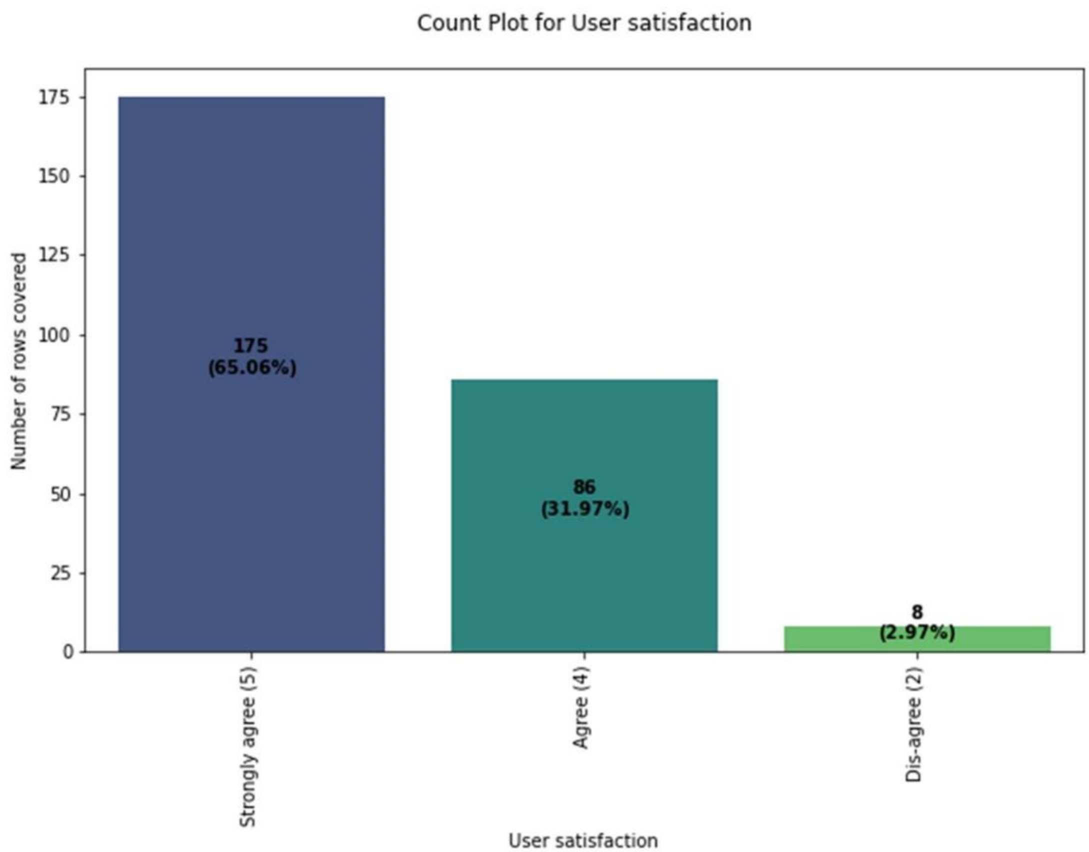
Count Plot for Return and Replacement policy
Number of rows covered
0
25
50
75
100
125
150
175
200
20
(7.43%)
198
(73.61%)
51
(18.96%)
Dis-agree (2)
Strongly agree (5)
Agree (4)
Return and Replacement policy

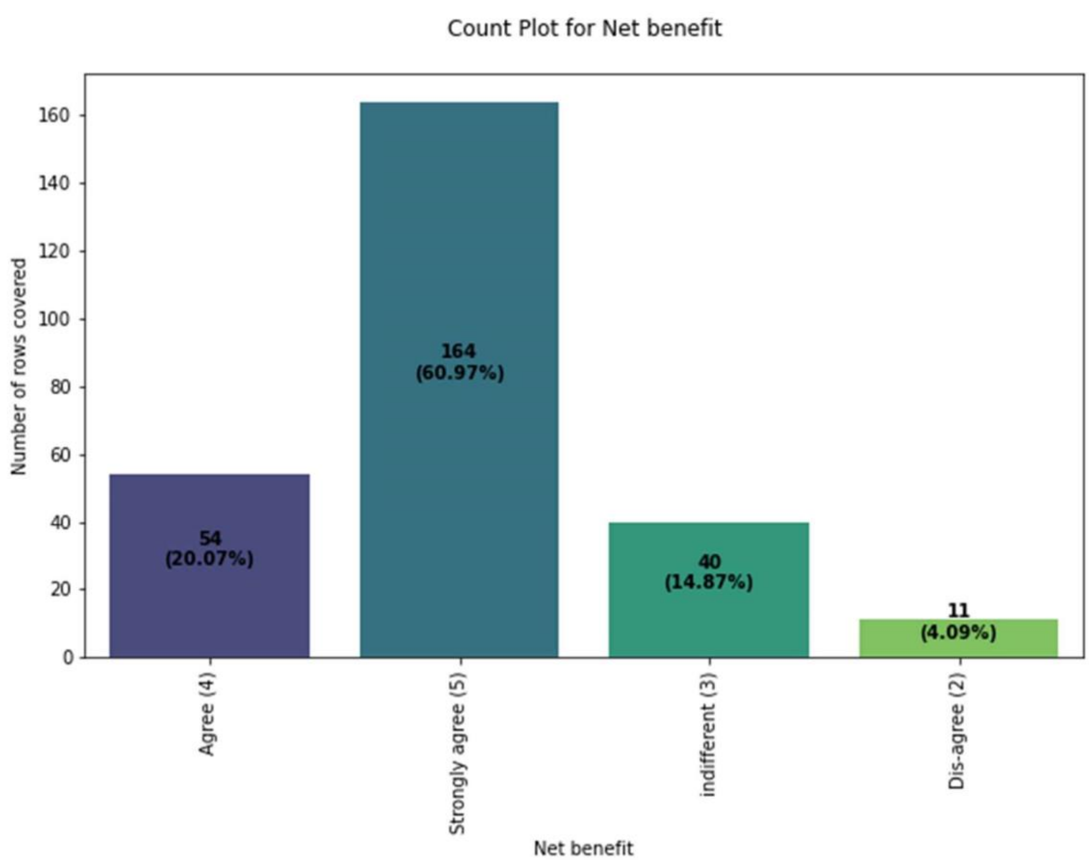
Count Plot for Loyalty program benefits

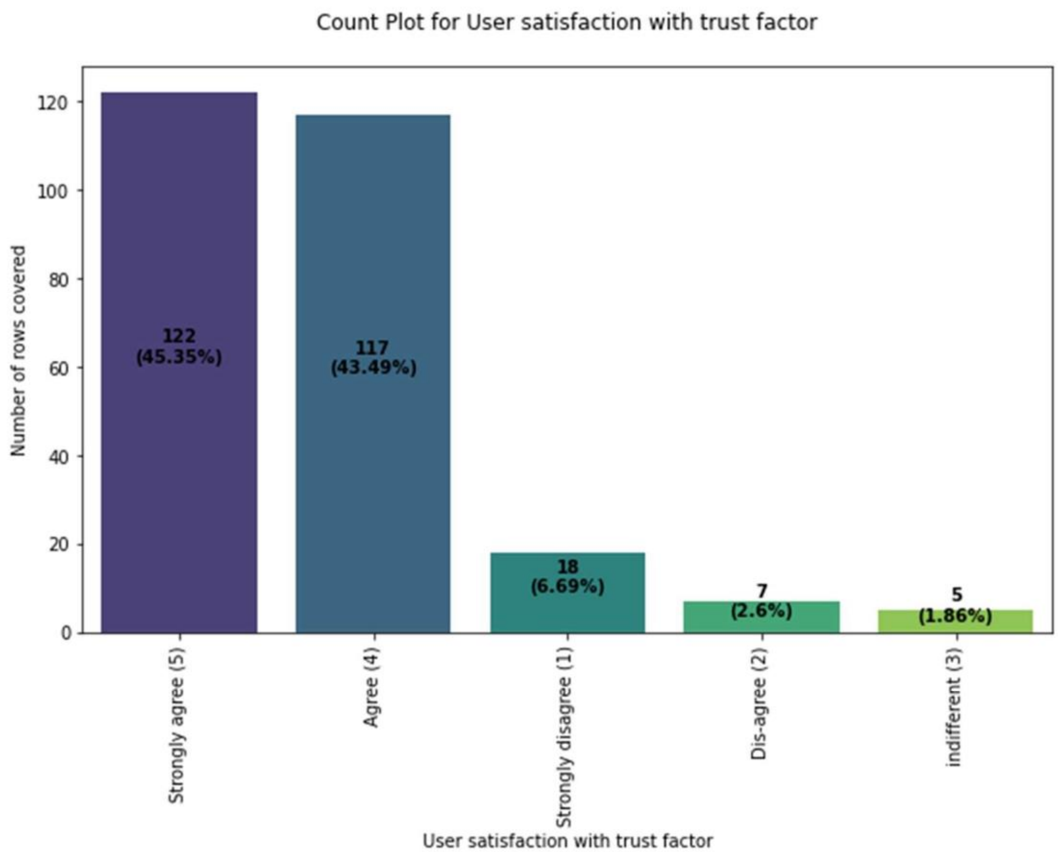


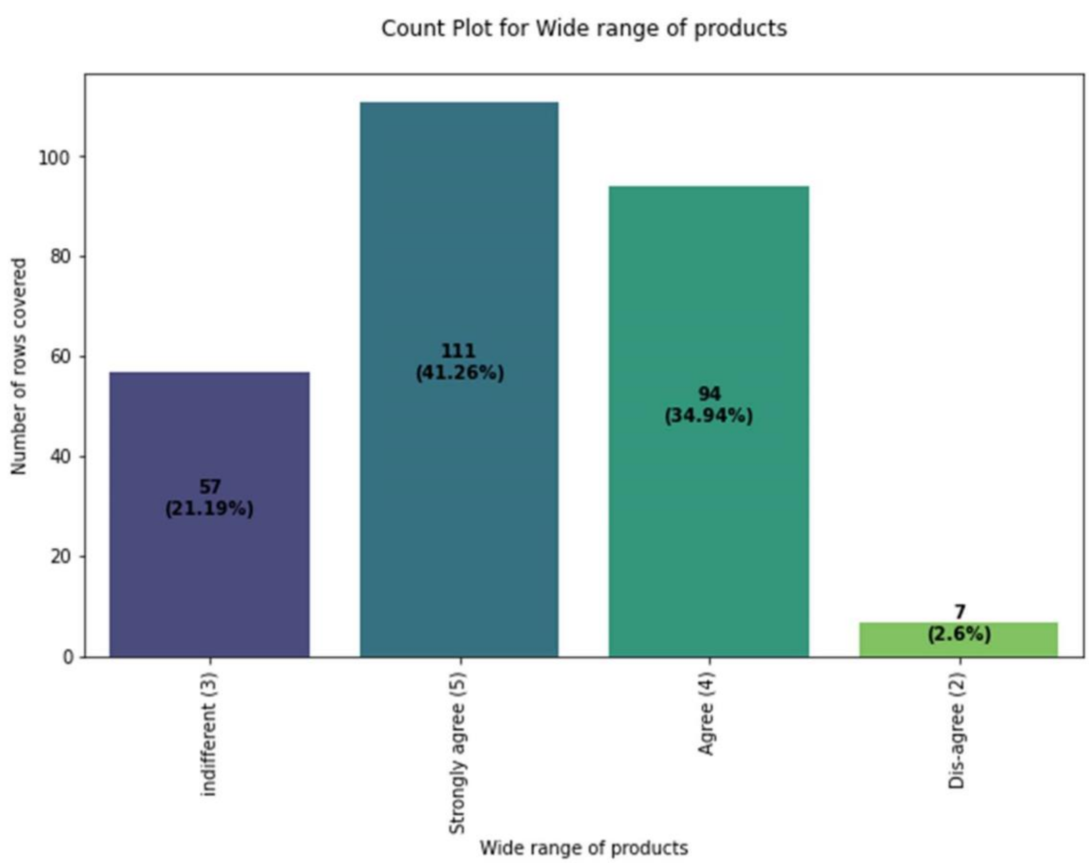
Count Plot for Quality of website content

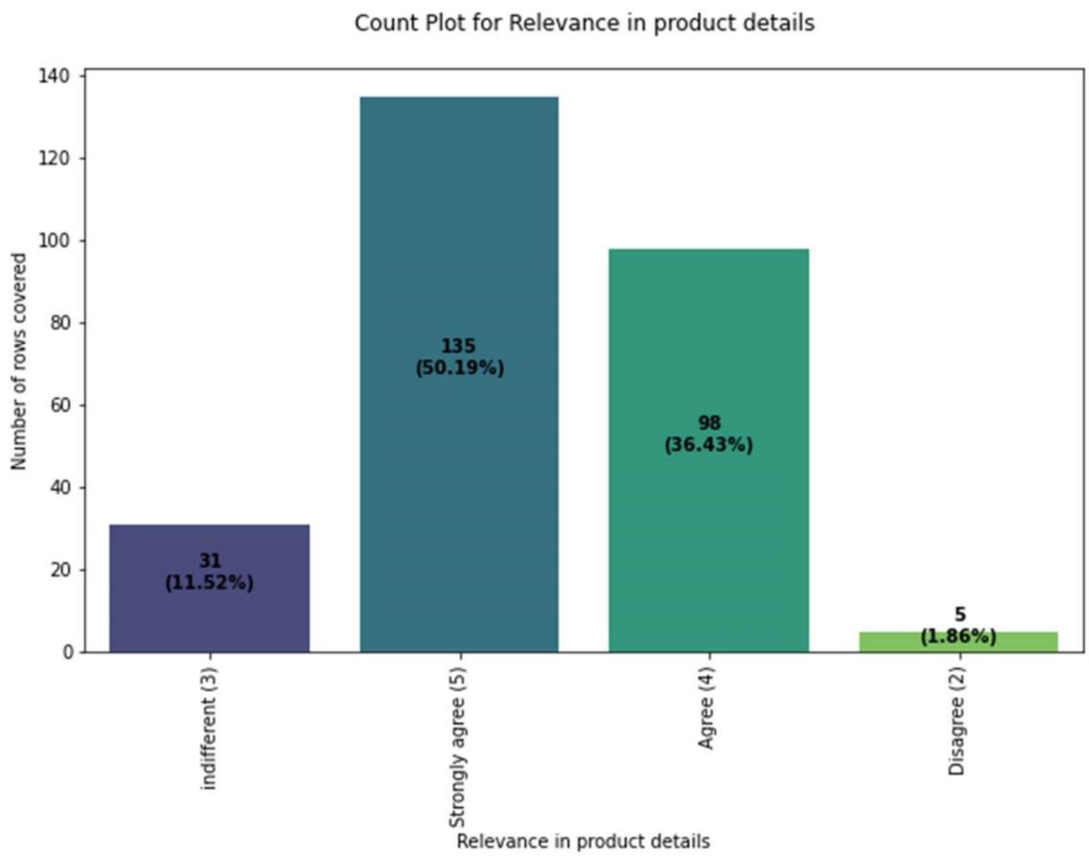


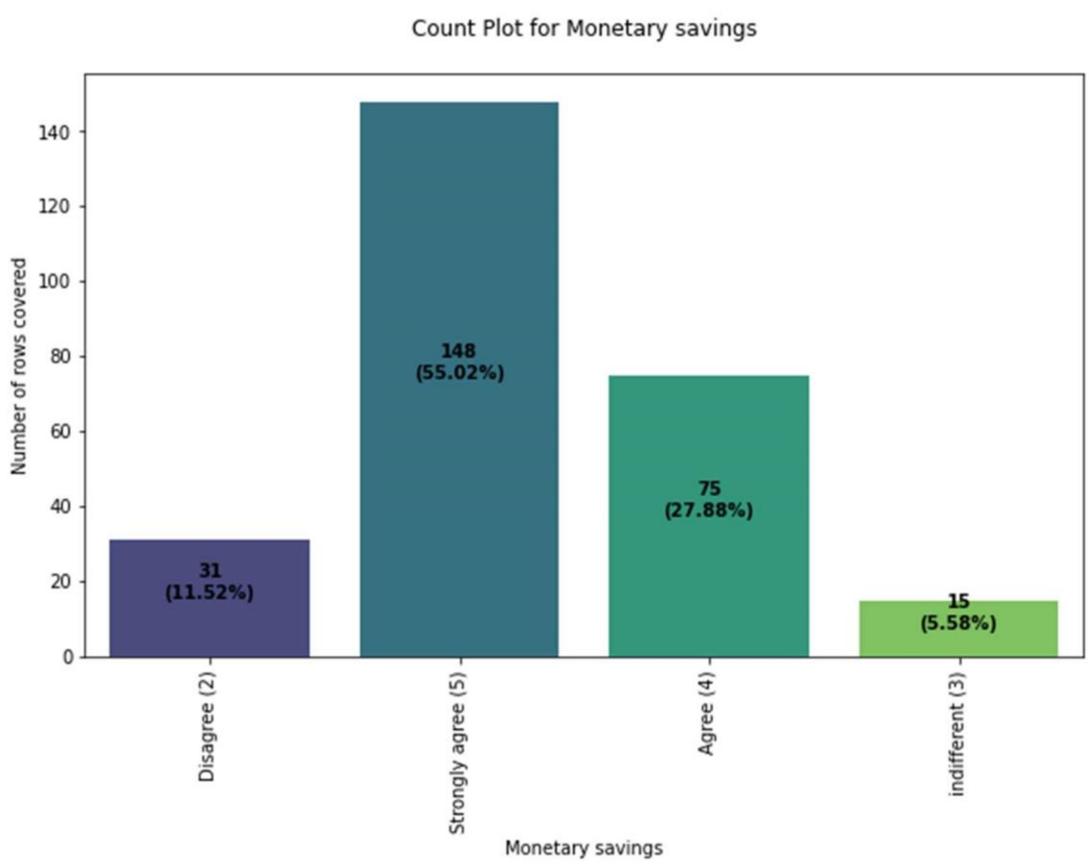

Count Plot for User satisfaction
Number of rows covered
175
150
125
100
75
50
25
0
175
(65.06%)
86
(31.97%)
8
(2.97%)
Strongly agree (5)
Agree (4)
Dis-agree (2)
User satisfaction

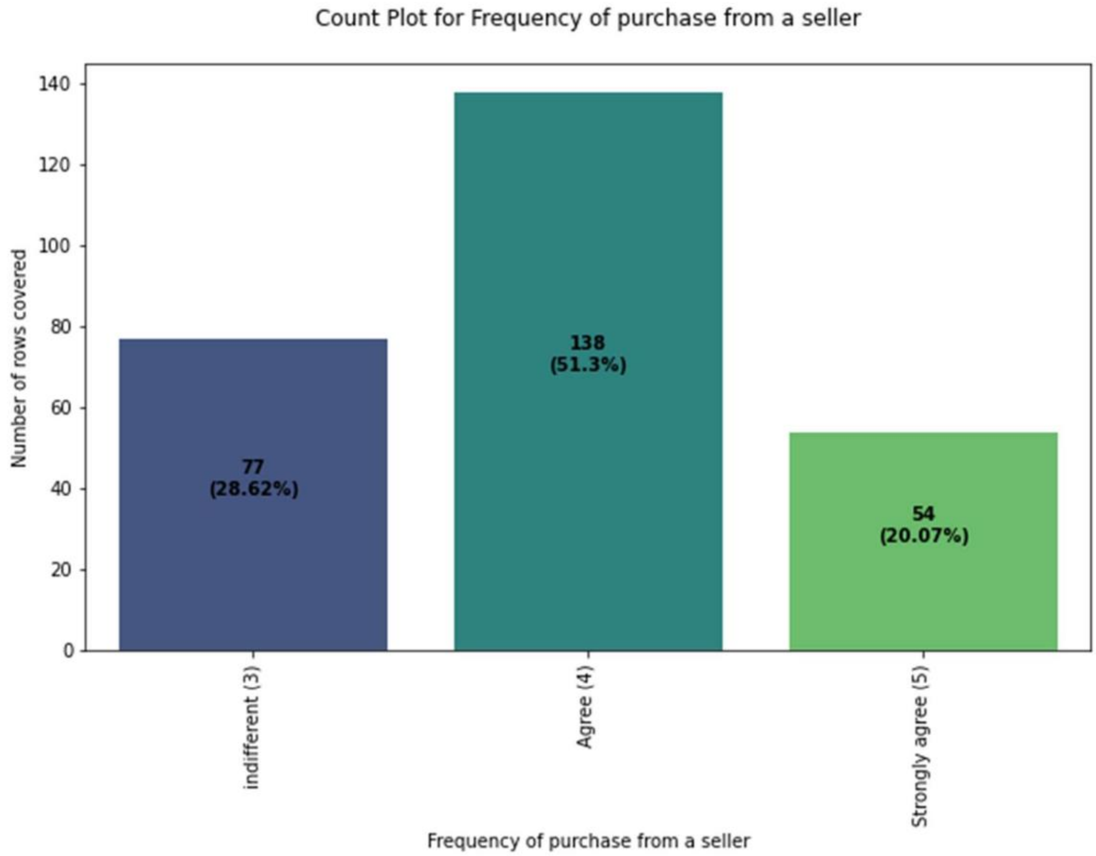

Count Plot for Net benefit
Number of rows covered
160
140
120
100
80
60
40
20
0
54
(20.07%)
164
(60.97%)
40
(14.87%)
11
(4.09%)
Agree (4)
Strongly agree (5)
indifferent (3)
Dis-agree (2)
Net benefit

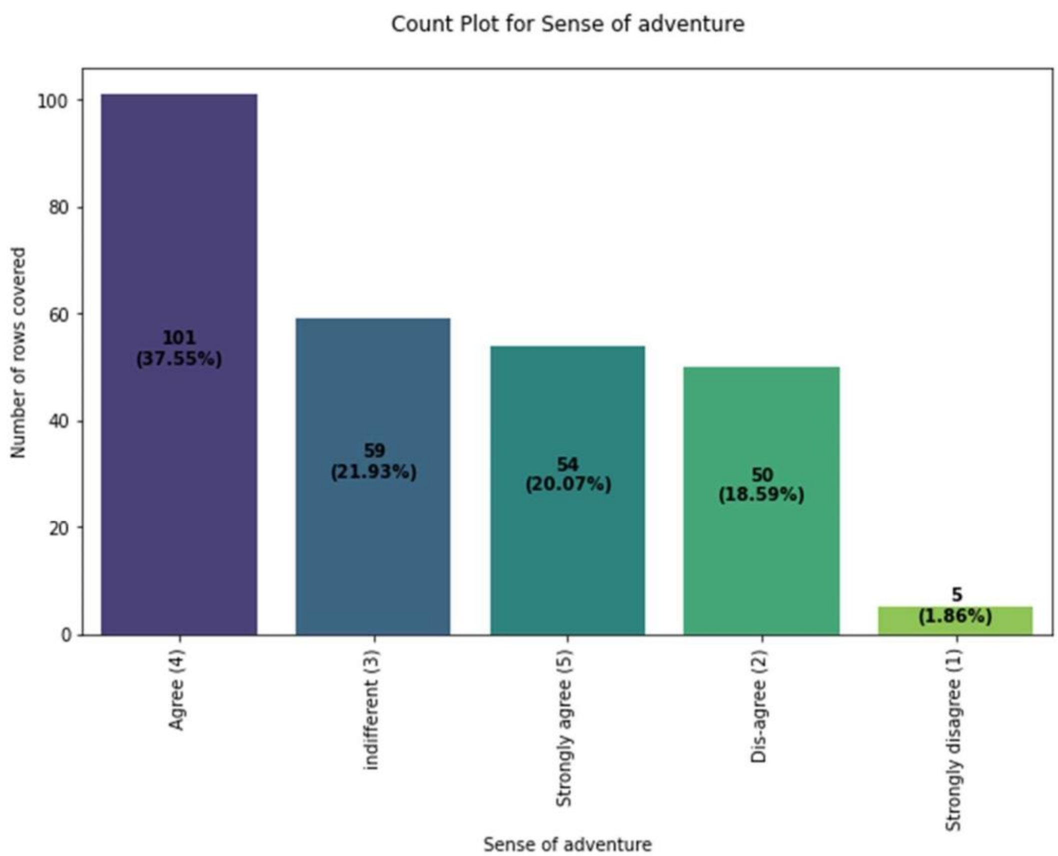

Count Plot for User satisfaction with trust factor
Number of rows covered
0
20
40
60
80
100
120
122
(45.35%)
117
(43.49%)
18
(6.69%)
7
(2.6%)
5
(1.86%)
Strongly agree (5)
Agree (4)
Strongly disagree (1)
Dis-agree (2)
indifferent (3)
User satisfaction with trust factor

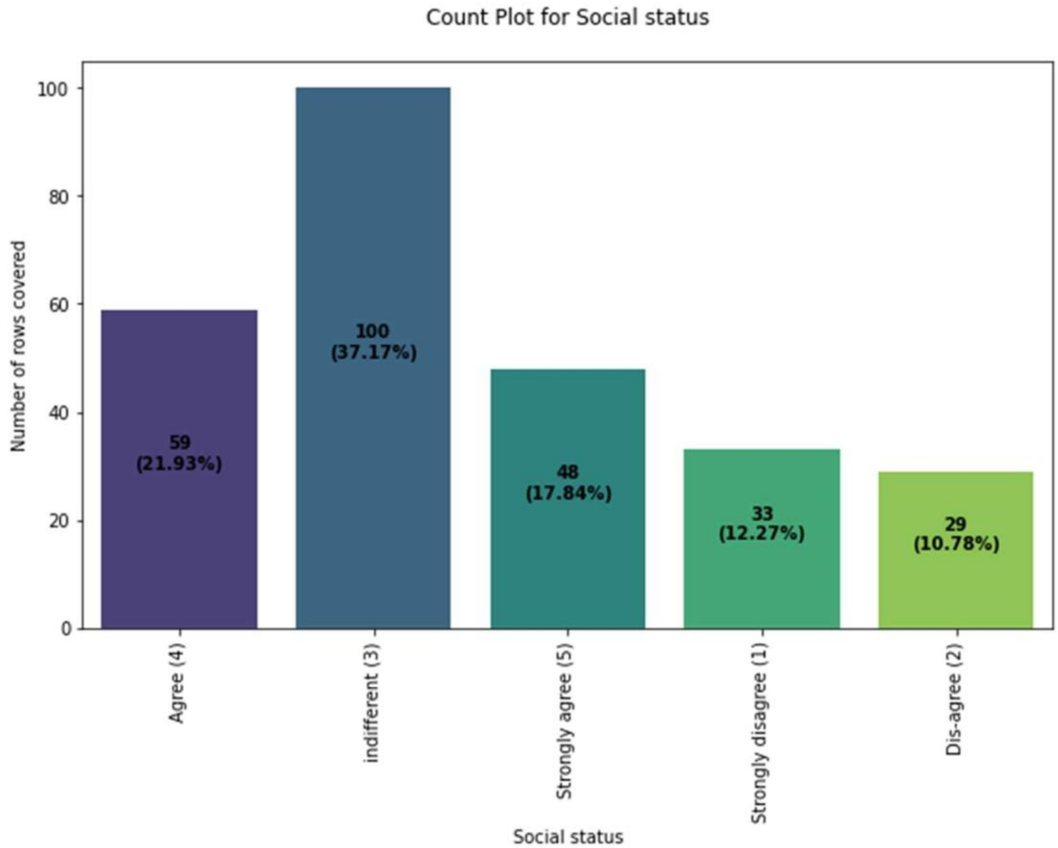

Count Plot for Wide range of products
Number of rows covered
0
20
40
60
80
100
57
(21.19%)
111
(41.26%)
94
(34.94%)
7
(2.6%)
indifferent (3)
Strongly agree (5)
Agree (4)
Dis-agree (2)
Wide range of products


Count Plot for Relevance in product details
Number of rows covered
0
20
40
60
80
100
120
140
31
(11.52%)
135
(50.19%)
98
(36.43%)
5
(1.86%)
indifferent (3)
Strongly agree (5)
Agree (4)
Disagree (2)
Relevance in product details

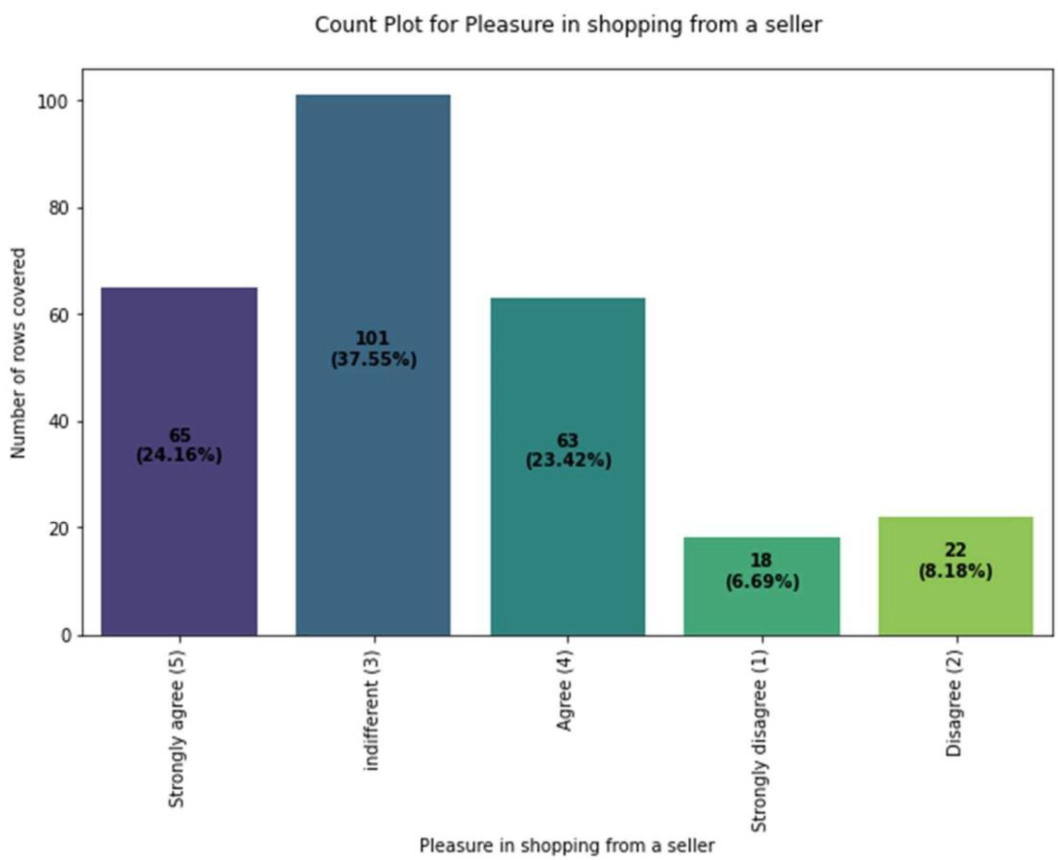

Count Plot for Monetary savings
Number of rows covered
0
20
40
60
80
100
120
140
31
(11.52%)
148
(55.02%)
75
(27.88%)
15
(5.58%)
Disagree (2)
Strongly agree (5)
Agree (4)
indifferent (3)
Monetary savings


Count Plot for Frequency of purchase from a seller
Number of rows covered
77
(28.62%)
138
(51.3%)
54
(20.07%)
indifferent (3)
Agree (4)
Strongly agree (5)
Frequency of purchase from a seller

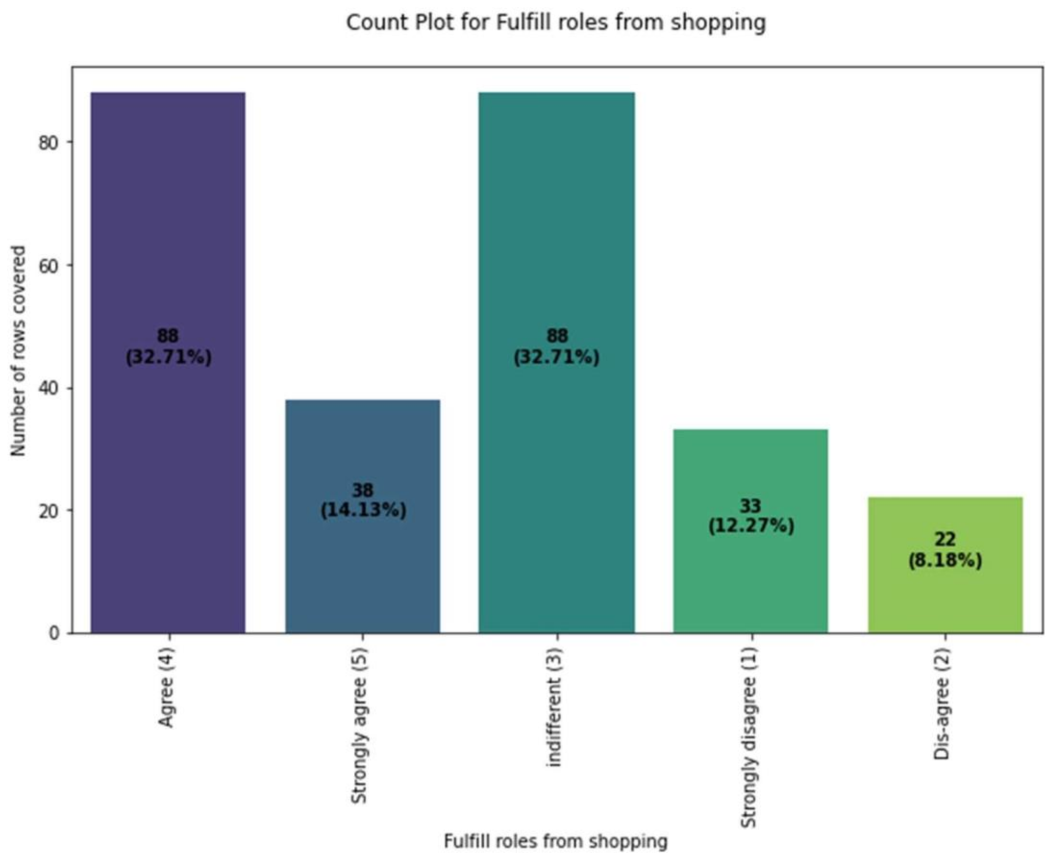

Count Plot for Sense of adventure
Number of rows covered
101
(37.55%)
59
(21.93%)
54
(20.07%)
50
(18.59%)
5
(1.86%)
Agree (4)
indifferent (3)
Strongly agree (5)
Dis-agree (2)
Strongly disagree (1)
Sense of adventure

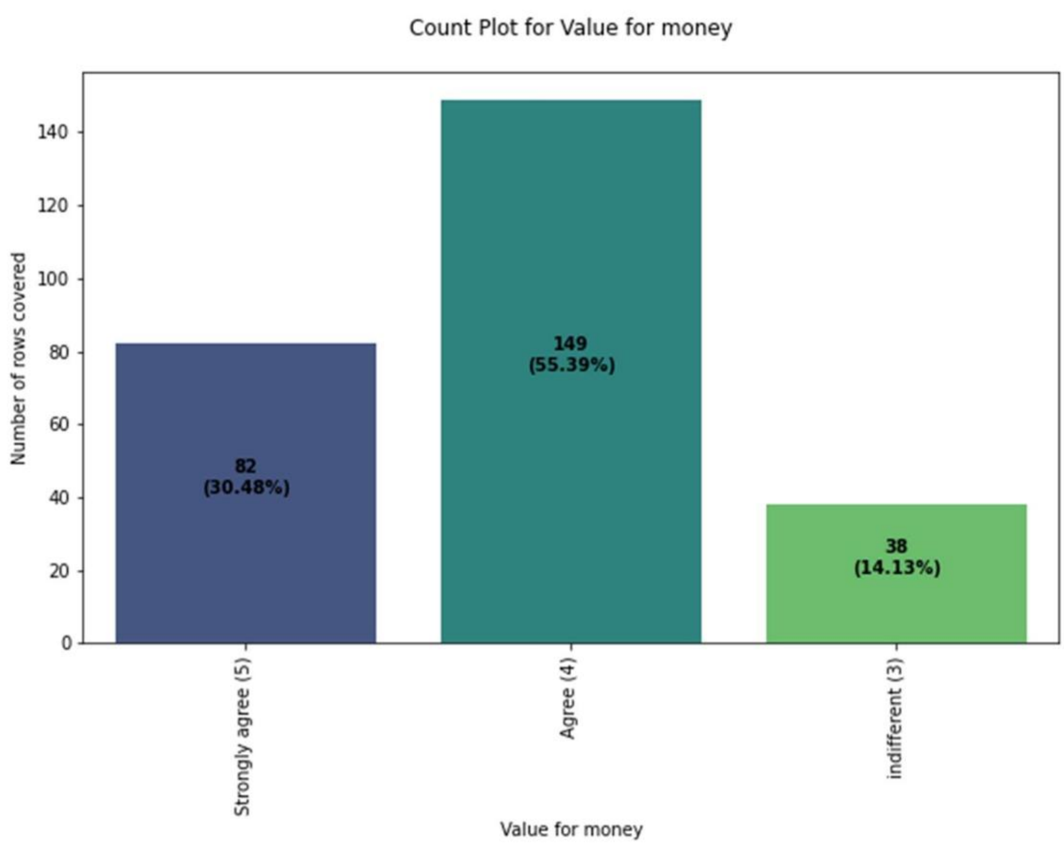


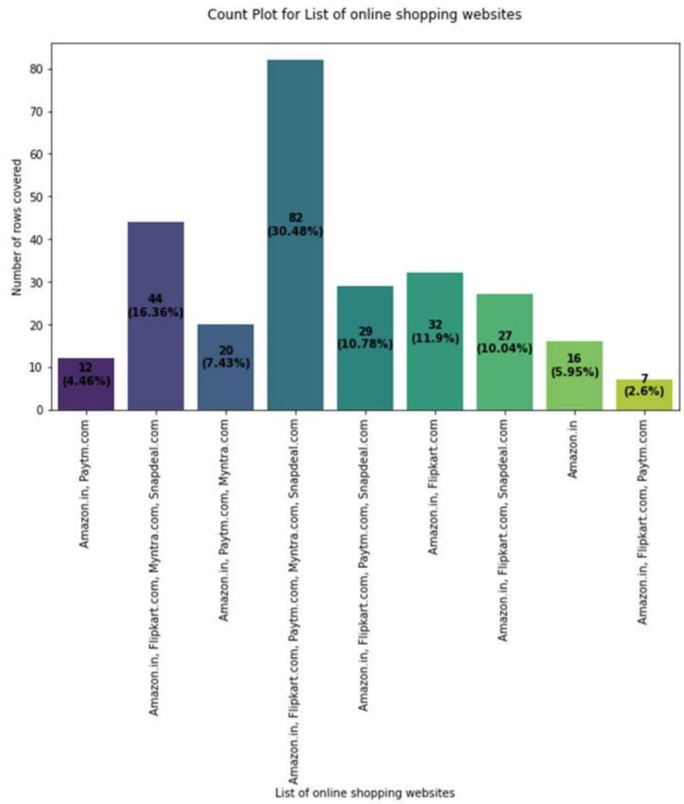
Count Plot for Social status
Number of rows covered
100
80
60
40
20
0
59
(21.93%)
100
(37.17%)
48
(17.84%)
33
(12.27%)
29
(10.78%)
Agree (4)
indifferent (3)
Strongly agree (5)
Strongly disagree (1)
Dis-agree (2)
Social status

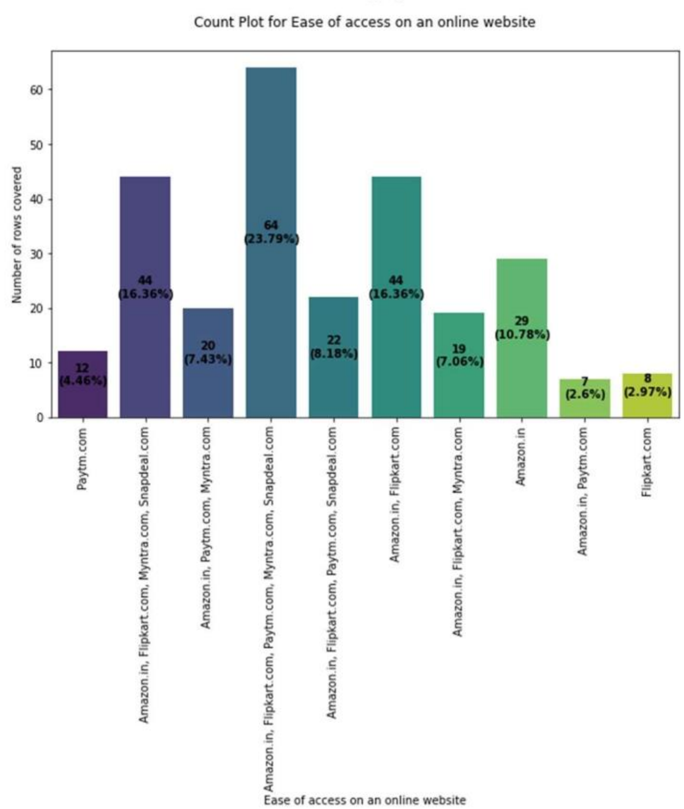


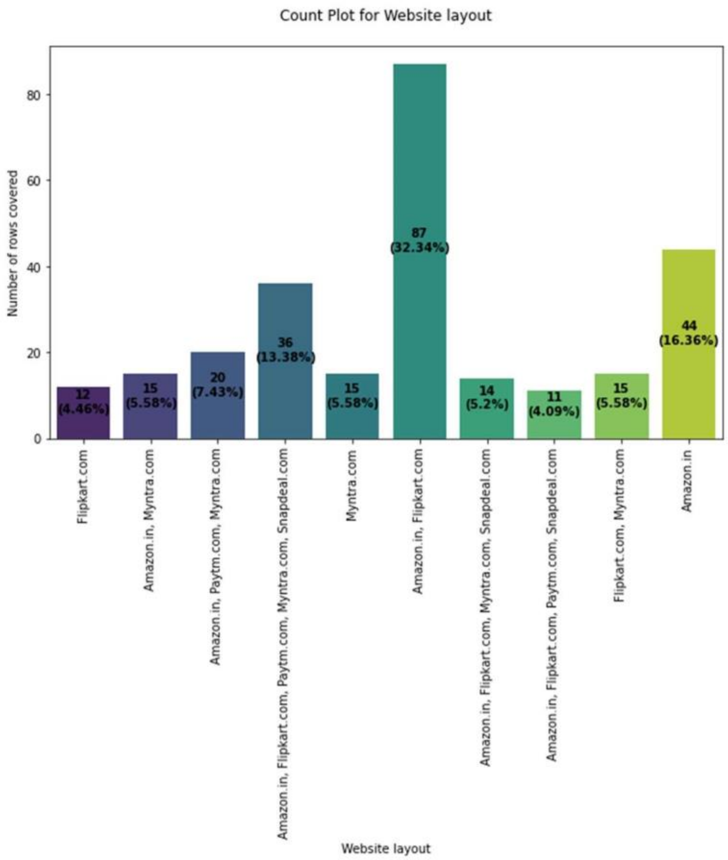
Count Plot for Pleasure in shopping from a seller
Number of rows covered
100
80
60
40
20
0
65
(24.16%)
101
(37.55%)
63
(23.42%)
18
(6.69%)
22
(8.18%)
Strongly agree (5)
indifferent (3)
Agree (4)
Strongly disagree (1)
Disagree (2)
Pleasure in shopping from a seller

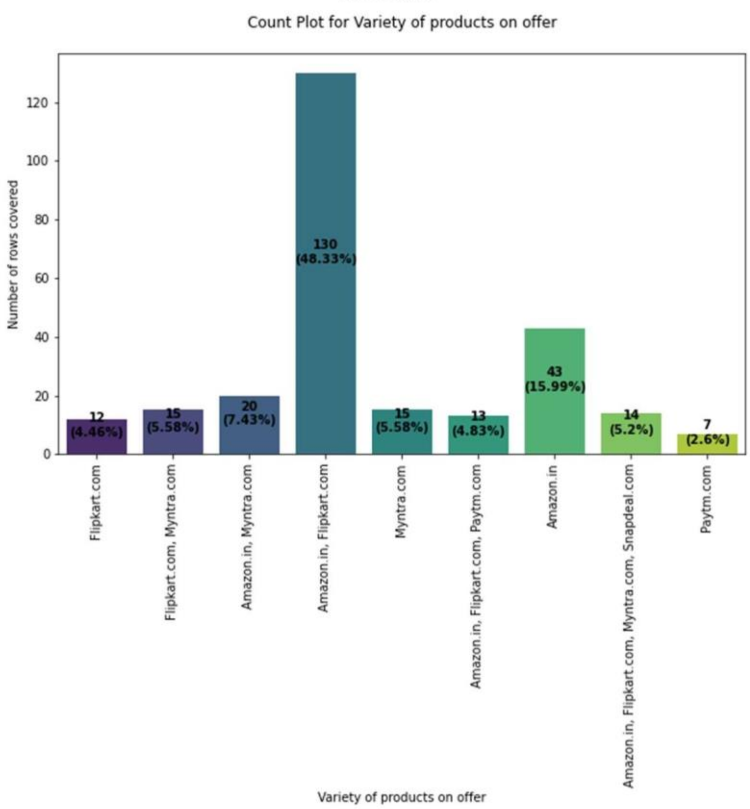

Count Plot for Fulfill roles from shopping
Number of rows covered
0
20
40
60
80
88
(32.71%)
38
(14.13%)
88
(32.71%)
33
(12.27%)
22
(8.18%)
Agree (4)
Strongly agree (5)
indifferent (3)
Strongly disagree (1)
Dis-agree (2)
Fulfill roles from shopping

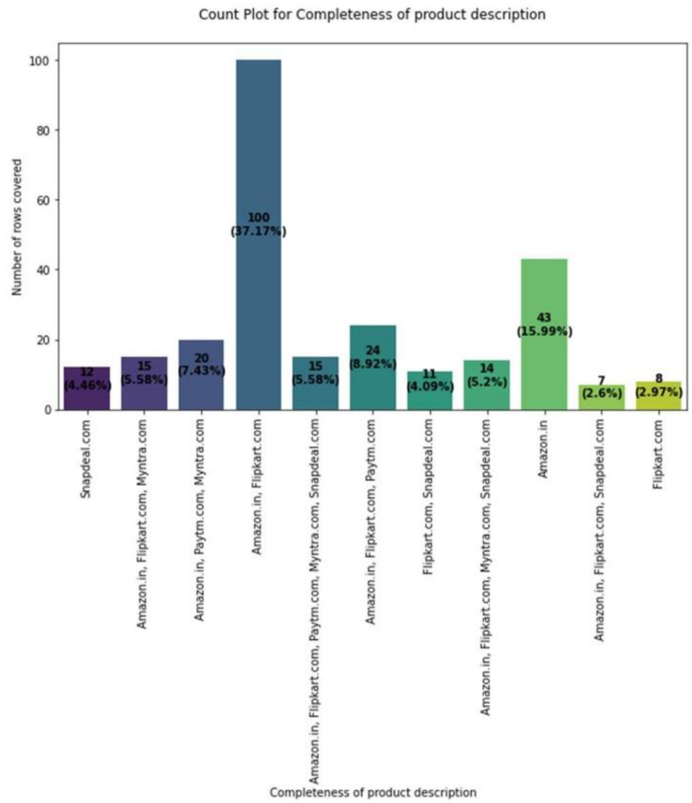
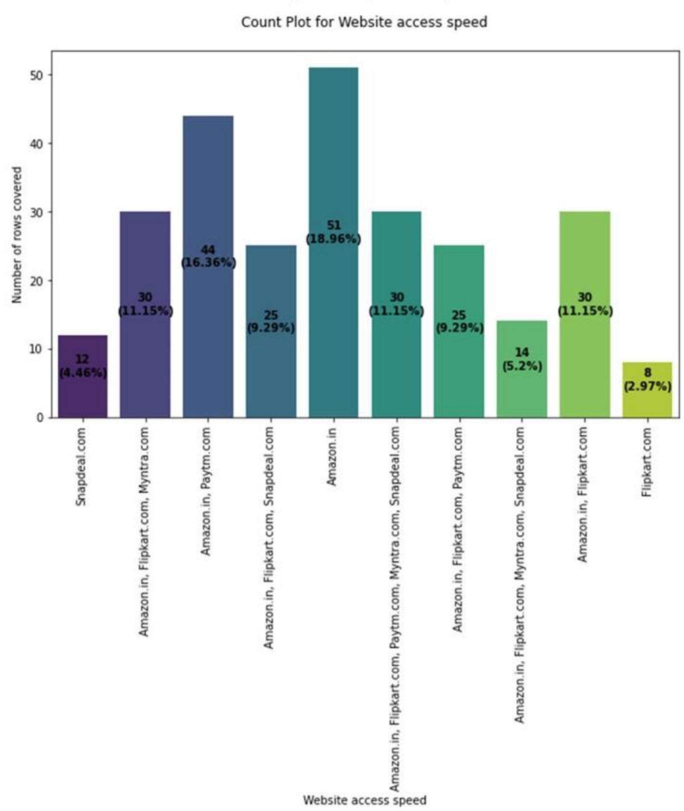

Count Plot for Value for money
Number of rows covered
0
20
40
60
80
100
120
140
82
(30.48%)
149
(55.39%)
38
(14.13%)
Strongly agree (5)
Agree (4)
indifferent (3)
Value for money

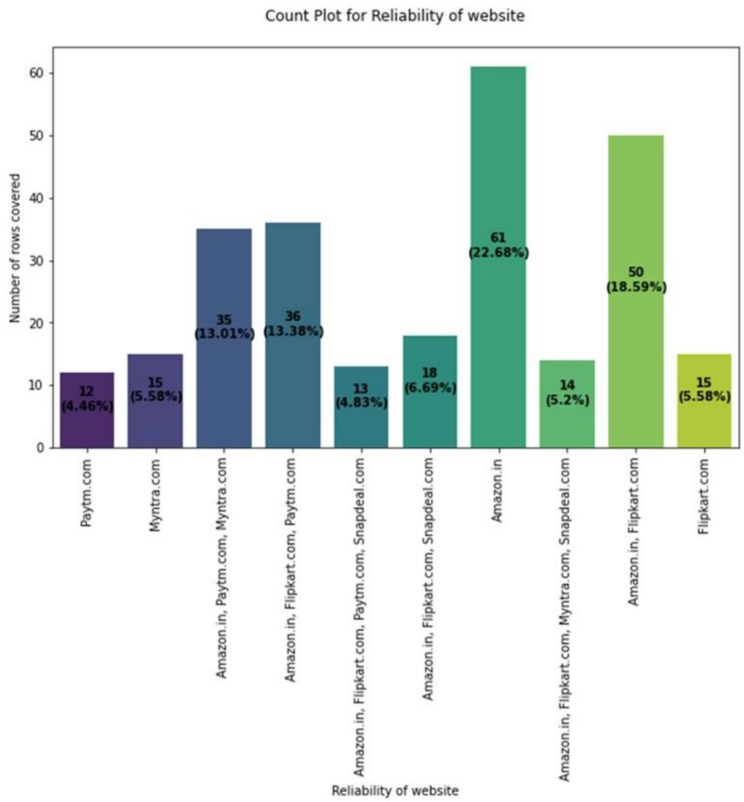

Count Plot for List of online shopping websites
Number of rows covered
12 (4.46%)
44 (16.36%)
20 (7.43%)
82 (30.48%)
29 (10.78%)
32 (11.9%)
27 (10.04%)
16 (5.95%)
7 (2.6%)
Amazon.in, Paytm.com
Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com
Amazon.in, Paytm.com, Myntra.com
Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com
Amazon.in, Flipkart.com, Paytm.com, Snapdeal.com
Amazon.in, Flipkart.com
Amazon.in, Flipkart.com, Snapdeal.com
Amazon.in
Amazon.in, Flipkart.com, Paytm.com
List of online shopping websites

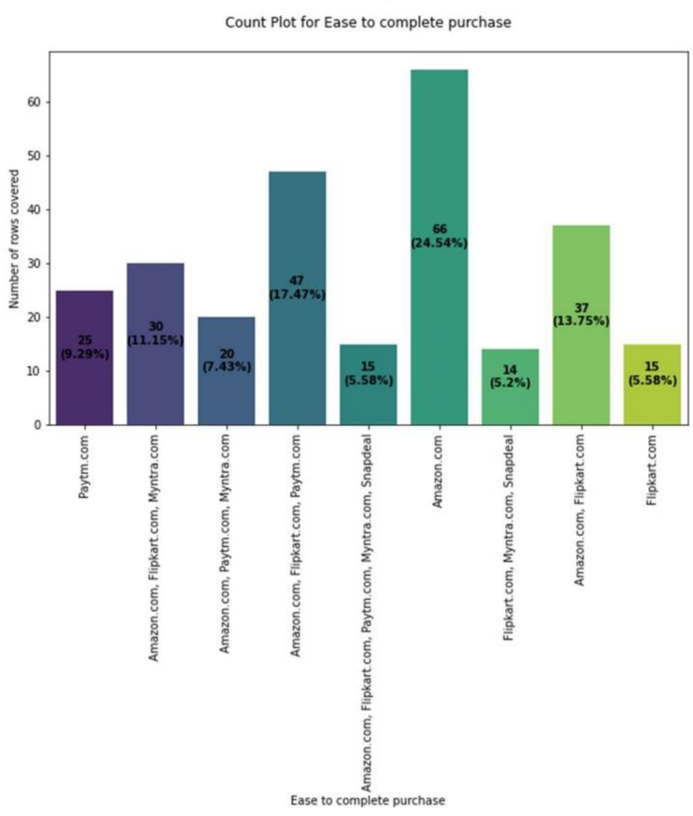

Count Plot for Ease of access on an online website
Number of rows covered
12 (4.46%)
44 (16.36%)
20 (7.43%)
64 (23.79%)
22 (8.18%)
44 (16.36%)
19 (7.06%)
29 (10.78%)
7 (2.6%)
8 (2.97%)
Paytm.com
Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com
Amazon.in, Paytm.com, Myntra.com
Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com
Amazon.in, Flipkart.com, Paytm.com, Snapdeal.com
Amazon.in, Flipkart.com
Amazon.in, Flipkart.com, Myntra.com
Amazon.in
Amazon.in, Paytm.com
Flipkart.com
Ease of access on an online website

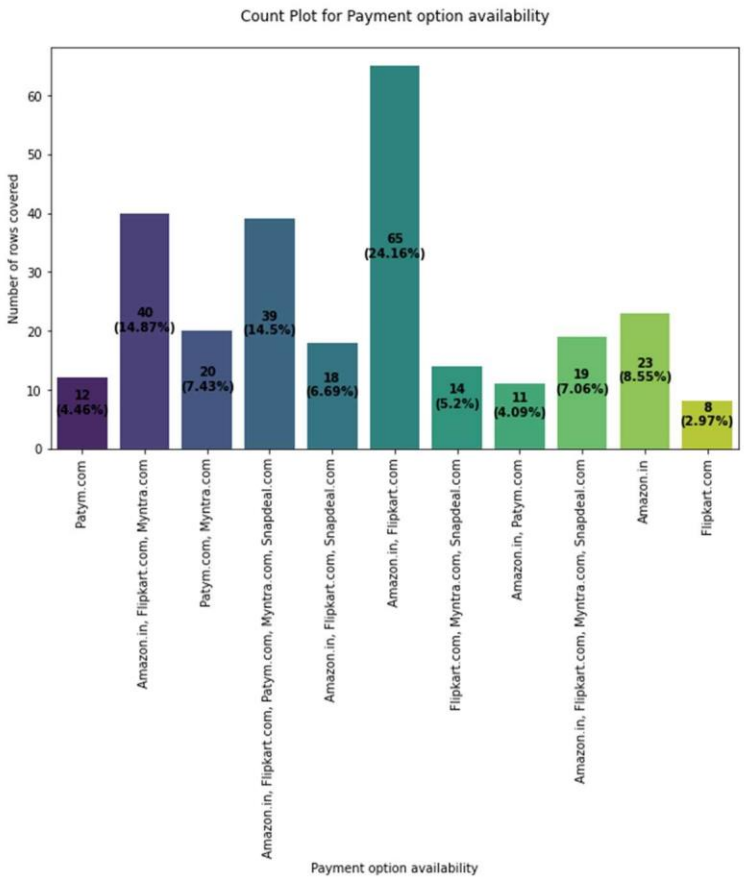

Count Plot for Website layout
Number of rows covered
Website layout
Flipkart.com
Amazon.in, Myntra.com
Amazon.in, Paytm.com, Myntra.com
Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com
Myntra.com
Amazon.in, Flipkart.com
Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com
Amazon.in, Flipkart.com, Paytm.com, Snapdeal.com
Flipkart.com, Myntra.com
Amazon.in
12 (4.46%)
15 (5.58%)
20 (7.43%)
36 (13.38%)
15 (5.58%)
87 (32.34%)
14 (5.2%)
11 (4.09%)
15 (5.58%)
44 (16.36%)

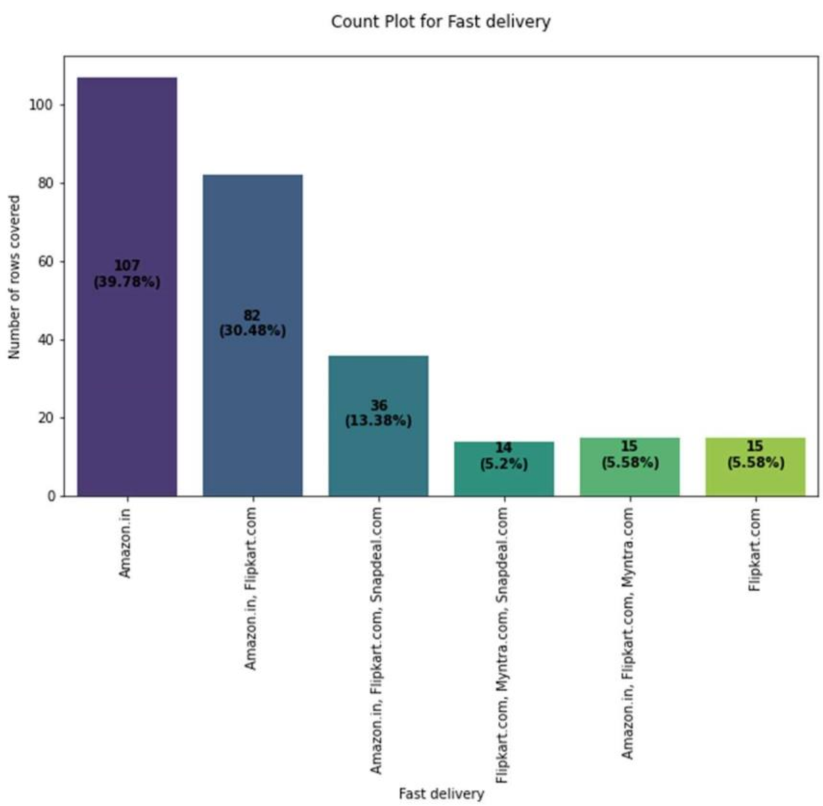

Count Plot for Variety of products on offer
Number of rows covered
Variety of products on offer
Flipkart.com
Flipkart.com, Myntra.com
Amazon.in, Myntra.com
Amazon.in, Flipkart.com
Myntra.com
Amazon.in, Flipkart.com, Paytm.com
Amazon.in
Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com
Paytm.com
12 (4.46%)
15 (5.58%)
20 (7.43%)
130 (48.33%)
15 (5.58%)
13 (4.83%)
43 (15.99%)
14 (5.2%)
7 (2.6%)

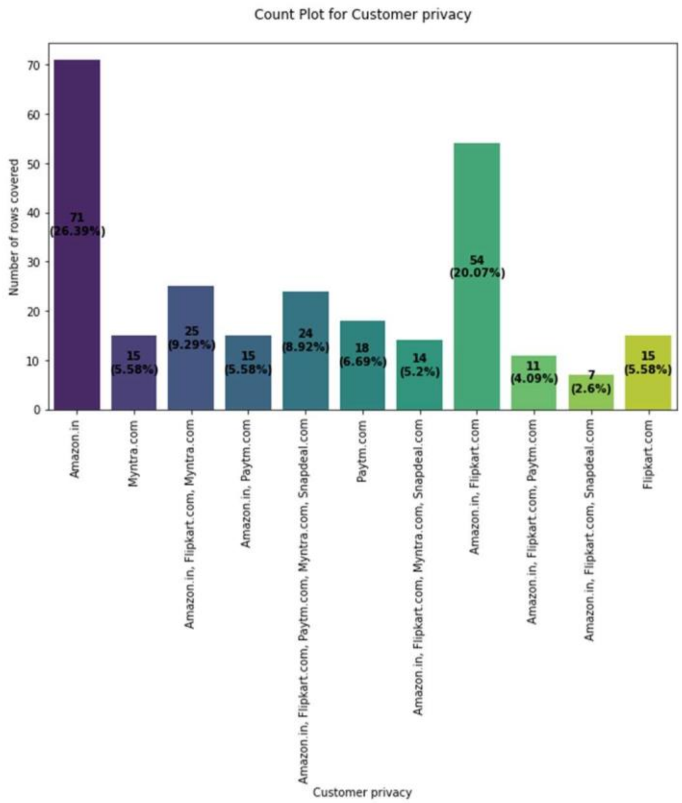

Count Plot for Completeness of product description
Number of rows covered
100
(37.17%)
43
(15.99%)
24
(8.92%)
20
(7.43%)
15
(5.58%)
12
(4.46%)
11
(4.09%)
14
(5.2%)
7
(2.6%)
8
(2.97%)
Snapdeal.com
Amazon.in, Flipkart.com, Myntra.com
Amazon.in, Paytm.com, Myntra.com
Amazon.in, Flipkart.com
Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com
Amazon.in, Flipkart.com, Paytm.com
Flipkart.com, Snapdeal.com
Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com
Amazon.in
Amazon.in, Flipkart.com, Snapdeal.com
Flipkart.com
Completeness of product description

Count Plot for Website access speed
Number of rows covered
12
(4.46%)
30
(11.15%)
44
(16.36%)
25
(9.29%)
51
(18.96%)
30
(11.15%)
25
(9.29%)
14
(5.2%)
30
(11.15%)
8
(2.97%)
Snapdeal.com
Amazon.in, Flipkart.com, Myntra.com
Amazon.in, Paytm.com
Amazon.in, Flipkart.com, Snapdeal.com
Amazon.in
Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com
Amazon.in, Flipkart.com, Paytm.com
Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com
Amazon.in, Flipkart.com
Flipkart.com
Website access speed

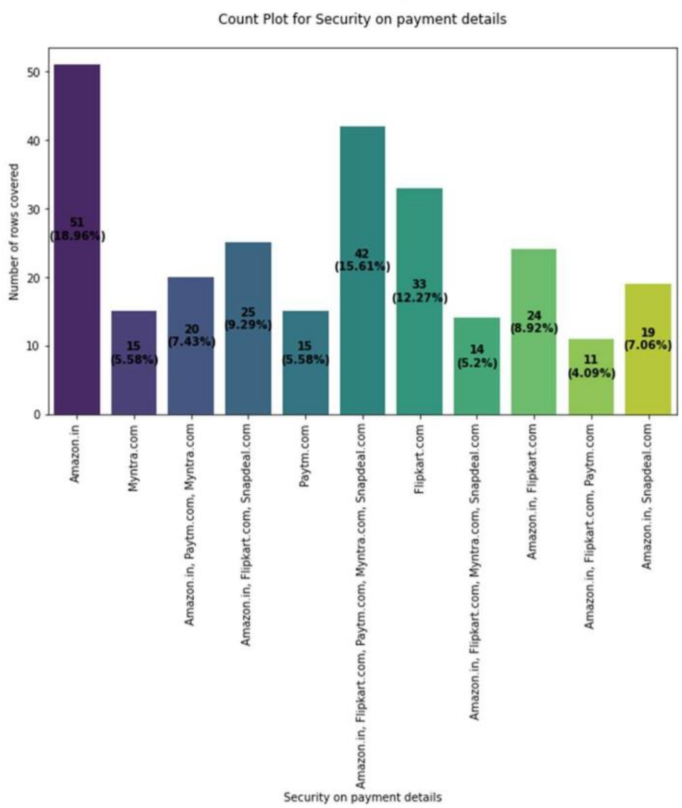

Count Plot for Reliability of website
Number of rows covered
Reliability of website
Paytm.com
Myntra.com
Amazon.in, Paytm.com, Myntra.com
Amazon.in, Flipkart.com, Paytm.com
Amazon.in, Flipkart.com, Paytm.com, Snapdeal.com
Amazon.in, Flipkart.com, Snapdeal.com
Amazon.in
Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com
Amazon.in, Flipkart.com
Flipkart.com
12 (4.46%)
15 (5.58%)
35 (13.01%)
36 (13.38%)
13 (4.83%)
18 (6.69%)
61 (22.68%)
14 (5.2%)
50 (18.59%)
15 (5.58%)

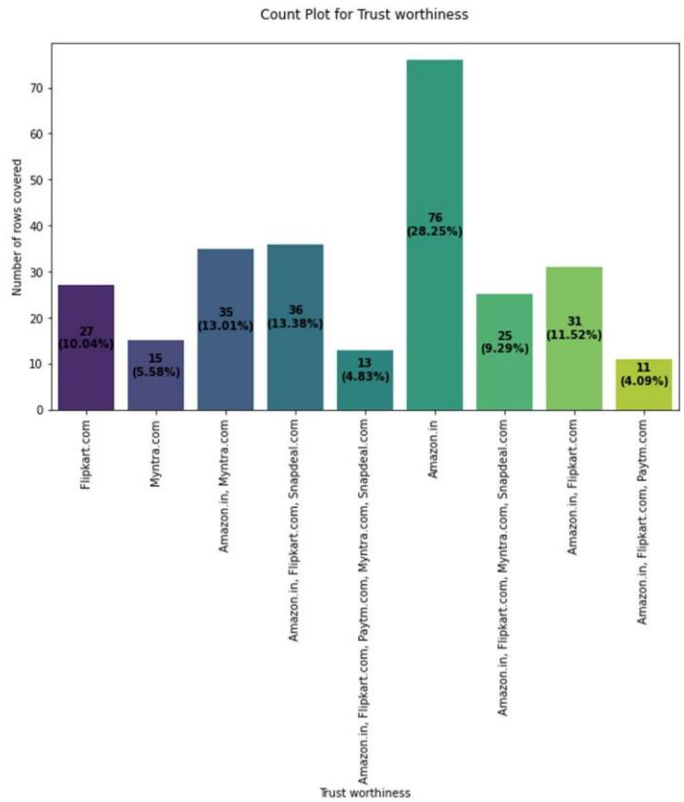

Count Plot for Ease to complete purchase
Number of rows covered
Ease to complete purchase
Paytm.com
Amazon.com, Flipkart.com, Myntra.com
Amazon.com, Paytm.com, Myntra.com
Amazon.com, Flipkart.com, Paytm.com
Amazon.com, Flipkart.com, Paytm.com, Myntra.com, Snapdeal
Amazon.com
Flipkart.com, Myntra.com, Snapdeal
Amazon.com, Flipkart.com
Flipkart.com
25 (9.29%)
30 (11.15%)
20 (7.43%)
47 (17.47%)
15 (5.58%)
66 (24.54%)
14 (5.2%)
37 (13.75%)
15 (5.58%)

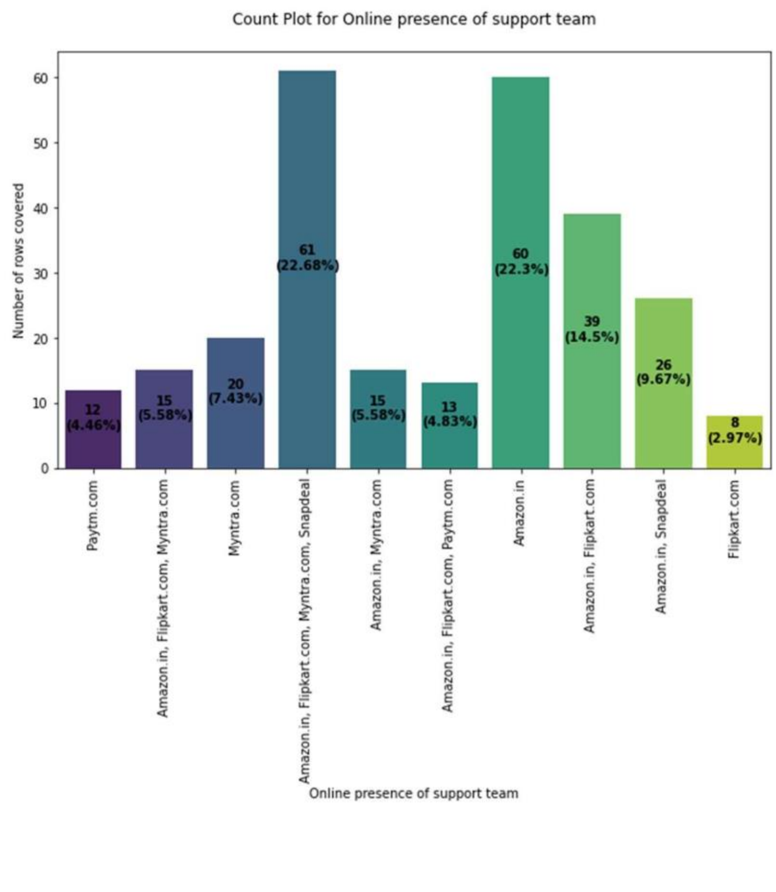

Count Plot for Payment option availability
Number of rows covered
Payment option availability
Patym.com
Amazon.in, Flipkart.com, Myntra.com
Patym.com, Myntra.com
Amazon.in, Flipkart.com, Patym.com, Myntra.com, Snapdeal.com
Amazon.in, Flipkart.com, Snapdeal.com
Amazon.in, Flipkart.com
Flipkart.com, Myntra.com, Snapdeal.com
Amazon.in, Patym.com
Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com
Amazon.in
Flipkart.com
12 (4.46%)
40 (14.87%)
20 (7.43%)
39 (14.5%)
18 (6.69%)
65 (24.16%)
14 (5.2%)
11 (4.09%)
19 (7.06%)
23 (8.55%)
8 (2.97%)

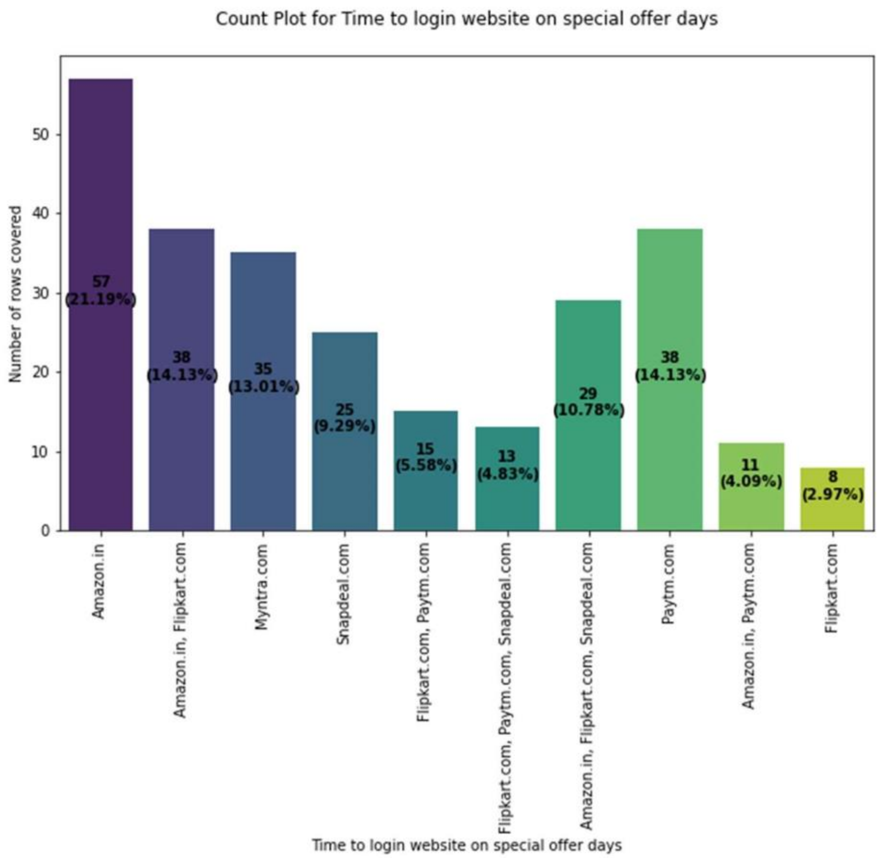

Count Plot for Fast delivery
Number of rows covered
Fast delivery
Amazon.in
Amazon.in, Flipkart.com
Amazon.in, Flipkart.com, Snapdeal.com
Flipkart.com, Myntra.com, Snapdeal.com
Amazon.in, Flipkart.com, Myntra.com
Flipkart.com
107 (39.78%)
82 (30.48%)
36 (13.38%)
14 (5.2%)
15 (5.58%)
15 (5.58%)

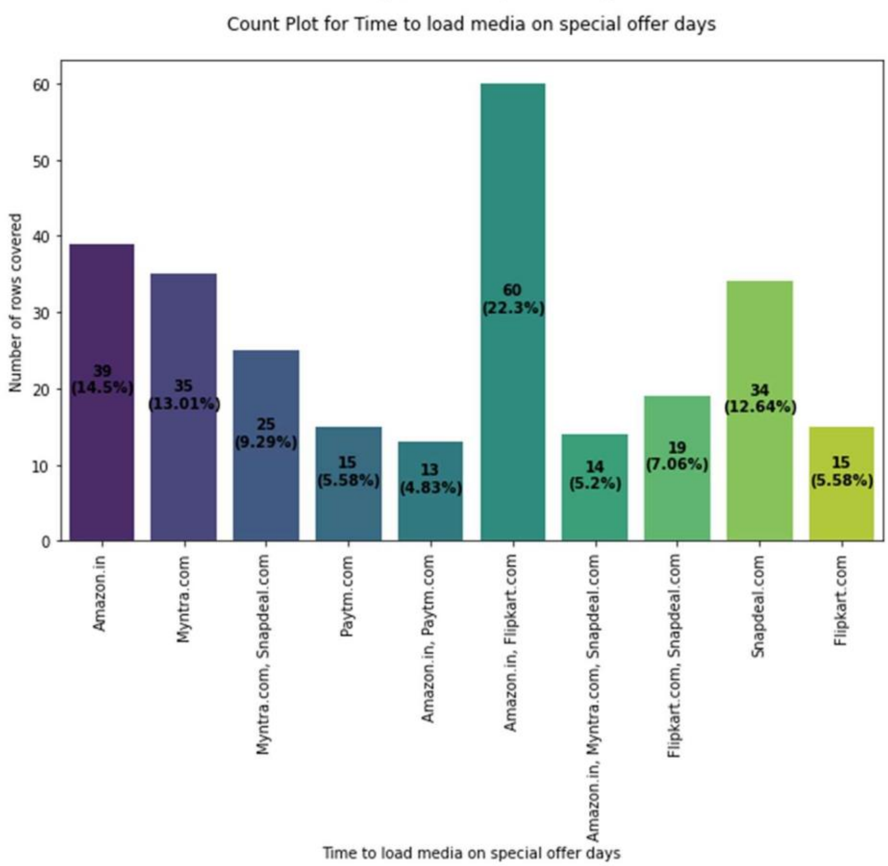

Count Plot for Customer privacy
Number of rows covered
71 (26.39%)
15 (5.58%)
25 (9.29%)
15 (5.58%)
24 (8.92%)
18 (6.69%)
14 (5.2%)
54 (20.07%)
11 (4.09%)
7 (2.6%)
15 (5.58%)
Amazon.in
Myntra.com
Amazon.in, Flipkart.com, Myntra.com
Amazon.in, Paytm.com
Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com
Paytm.com
Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com
Amazon.in, Flipkart.com
Amazon.in, Flipkart.com, Paytm.com
Amazon.in, Flipkart.com, Snapdeal.com
Flipkart.com
Customer privacy

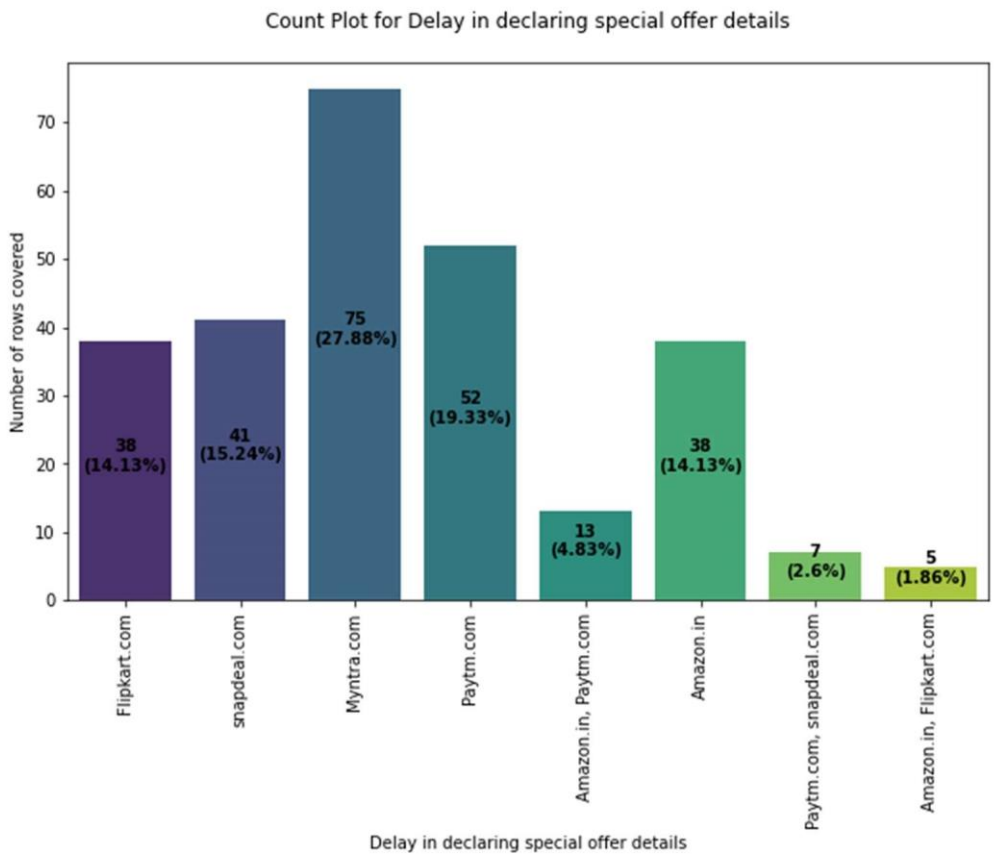

Count Plot for Security on payment details
Number of rows covered
51 (18.96%)
15 (5.58%)
20 (7.43%)
25 (9.29%)
15 (5.58%)
42 (15.61%)
33 (12.27%)
14 (5.2%)
24 (8.92%)
11 (4.09%)
19 (7.06%)
Amazon.in
Myntra.com
Amazon.in, Paytm.com, Myntra.com
Amazon.in, Flipkart.com, Snapdeal.com
Paytm.com
Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com
Flipkart.com
Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com
Amazon.in, Flipkart.com
Amazon.in, Flipkart.com, Paytm.com
Amazon.in, Snapdeal.com
Security on payment details

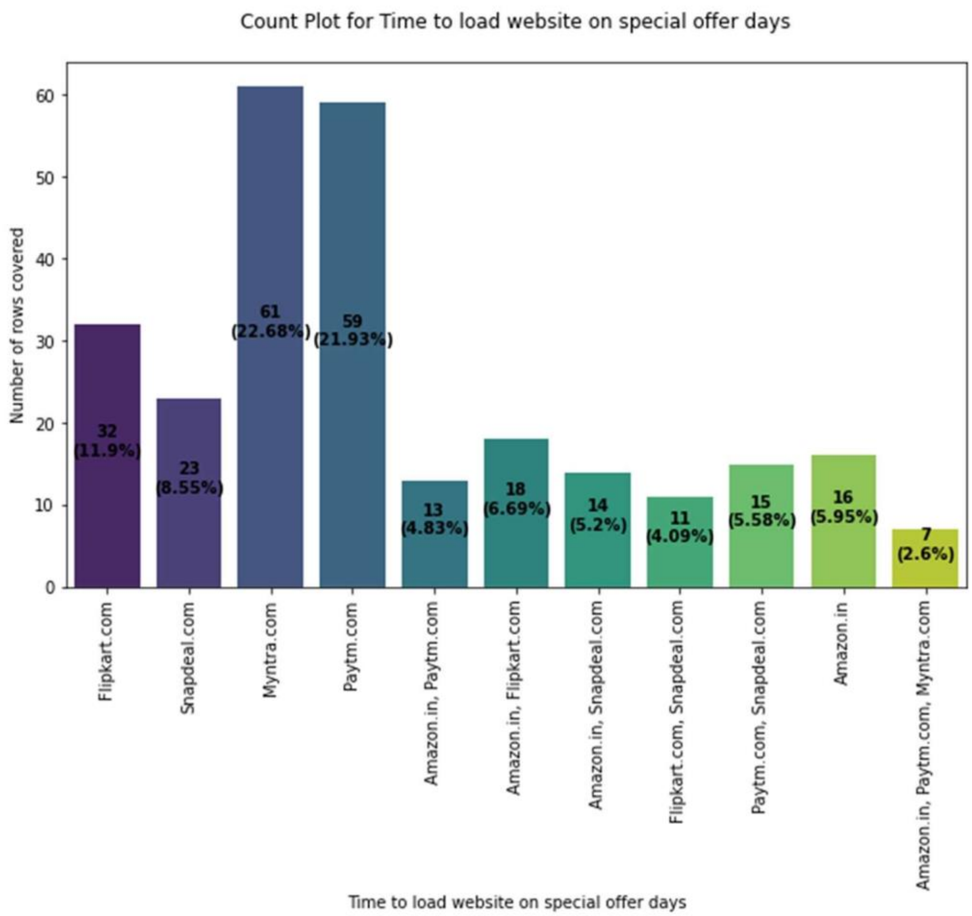

Count Plot for Trust worthiness
Number of rows covered
27 (10.04%)
15 (5.58%)
35 (13.01%)
36 (13.38%)
13 (4.83%)
76 (28.25%)
25 (9.29%)
31 (11.52%)
11 (4.09%)
Flipkart.com
Myntra.com
Amazon.in, Myntra.com
Amazon.in, Flipkart.com, Snapdeal.com
Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com
Amazon.in
Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com
Amazon.in, Flipkart.com
Amazon.in, Flipkart.com, Paytm.com
Trust worthiness

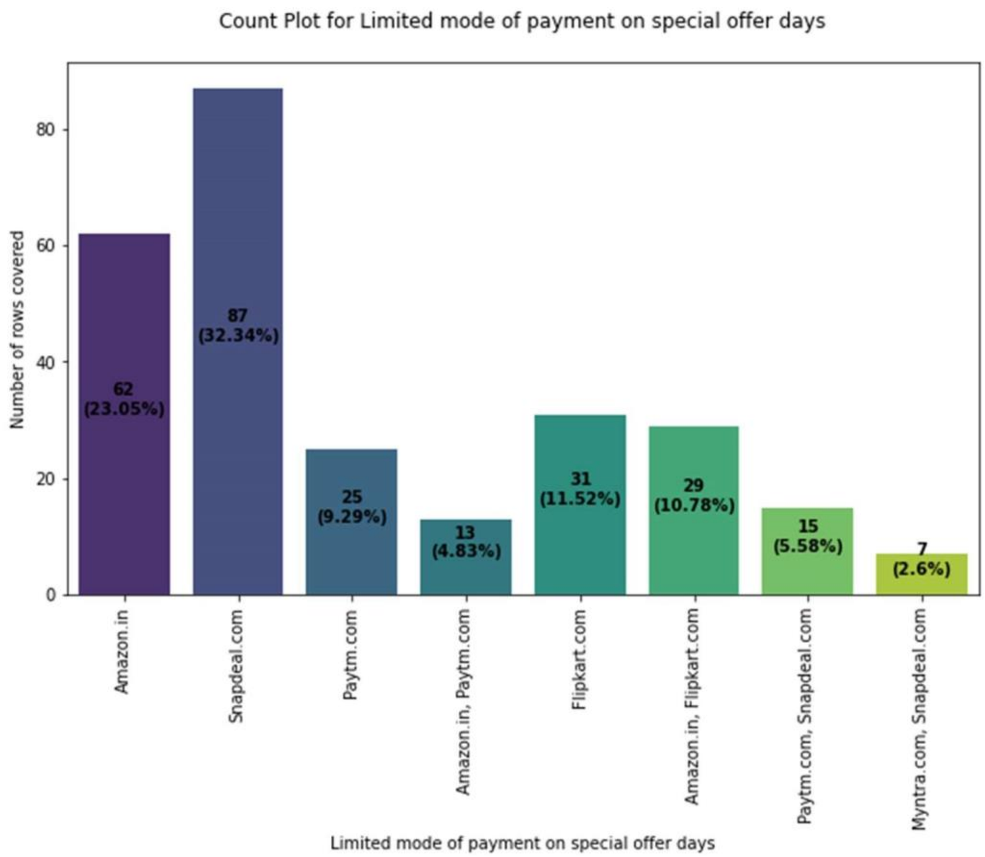

Count Plot for Online presence of support team
Number of rows covered
12 (4.46%)
15 (5.58%)
20 (7.43%)
61 (22.68%)
15 (5.58%)
13 (4.83%)
60 (22.3%)
39 (14.5%)
26 (9.67%)
8 (2.97%)
Paytm.com
Amazon.in, Flipkart.com, Myntra.com
Myntra.com
Amazon.in, Flipkart.com, Myntra.com, Snapdeal
Amazon.in, Myntra.com
Amazon.in, Flipkart.com, Paytm.com
Amazon.in
Amazon.in, Flipkart.com
Amazon.in, Snapdeal
Flipkart.com
Online presence of support team


Count Plot for Time to login website on special offer days
Number of rows covered
0
10
20
30
40
50
57
(21.19%)
38
(14.13%)
35
(13.01%)
25
(9.29%)
15
(5.58%)
13
(4.83%)
29
(10.78%)
38
(14.13%)
11
(4.09%)
8
(2.97%)
Amazon.in
Amazon.in, Flipkart.com
Myntra.com
Snapdeal.com
Flipkart.com, Paytm.com
Flipkart.com, Paytm.com, Snapdeal.com
Amazon.in, Flipkart.com, Snapdeal.com
Paytm.com
Amazon.in, Paytm.com
Flipkart.com
Time to login website on special offer days

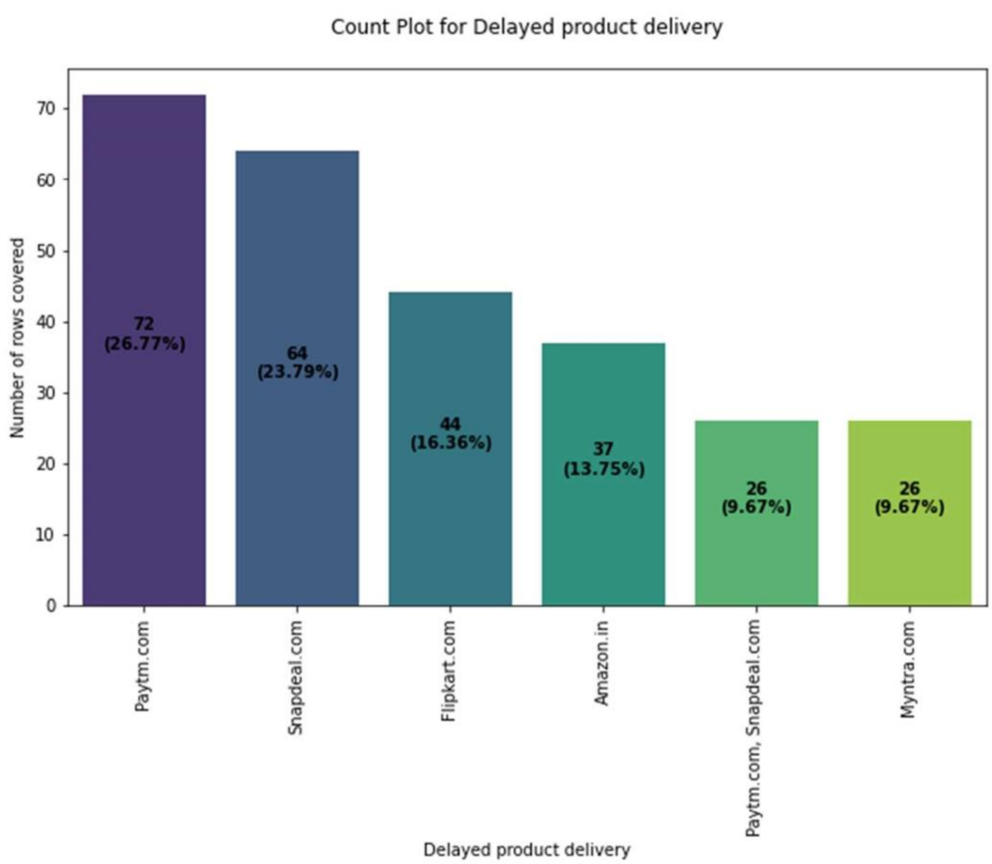

Count Plot for Time to load media on special offer days
Number of rows covered
0
10
20
30
40
50
60
39
(14.5%)
35
(13.01%)
25
(9.29%)
15
(5.58%)
13
(4.83%)
60
(22.3%)
14
(5.2%)
19
(7.06%)
34
(12.64%)
15
(5.58%)
Amazon.in
Myntra.com
Myntra.com, Snapdeal.com
Paytm.com
Amazon.in, Paytm.com
Amazon.in, Flipkart.com
Amazon.in, Myntra.com, Snapdeal.com
Flipkart.com, Snapdeal.com
Snapdeal.com
Flipkart.com
Time to load media on special offer days


Count Plot for Delay in declaring special offer details
Number of rows covered
0
10
20
30
40
50
60
70
38 (14.13%)
41 (15.24%)
75 (27.88%)
52 (19.33%)
13 (4.83%)
38 (14.13%)
7 (2.6%)
5 (1.86%)
Flipkart.com
snapdeal.com
Myntra.com
Paytm.com
Amazon.in, Paytm.com
Amazon.in
Paytm.com, snapdeal.com
Amazon.in, Flipkart.com
Delay in declaring special offer details

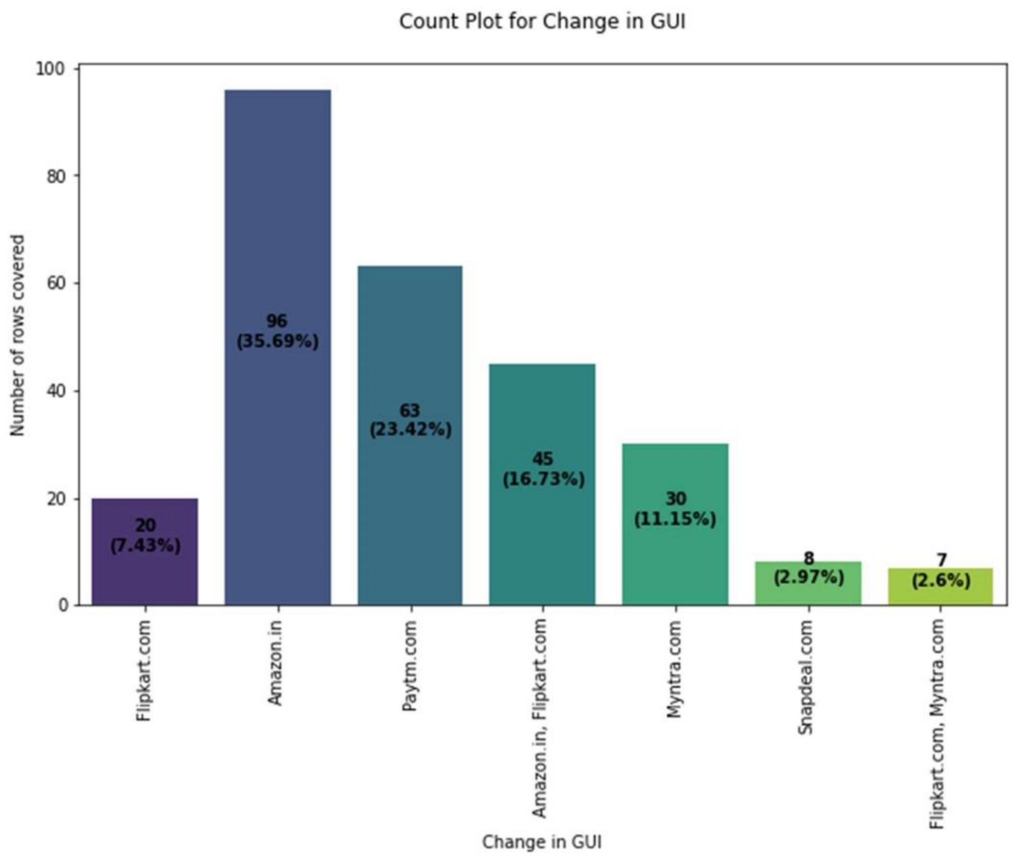

Count Plot for Time to load website on special offer days
Number of rows covered
0
10
20
30
40
50
60
32 (11.9%)
23 (8.55%)
61 (22.68%)
59 (21.93%)
13 (4.83%)
18 (6.69%)
14 (5.2%)
11 (4.09%)
15 (5.58%)
16 (5.95%)
7 (2.6%)
Flipkart.com
Snapdeal.com
Myntra.com
Paytm.com
Amazon.in, Paytm.com
Amazon.in, Flipkart.com
Amazon.in, Snapdeal.com
Flipkart.com, Snapdeal.com
Paytm.com, Snapdeal.com
Amazon.in
Amazon.in, Paytm.com, Myntra.com
Time to load website on special offer days

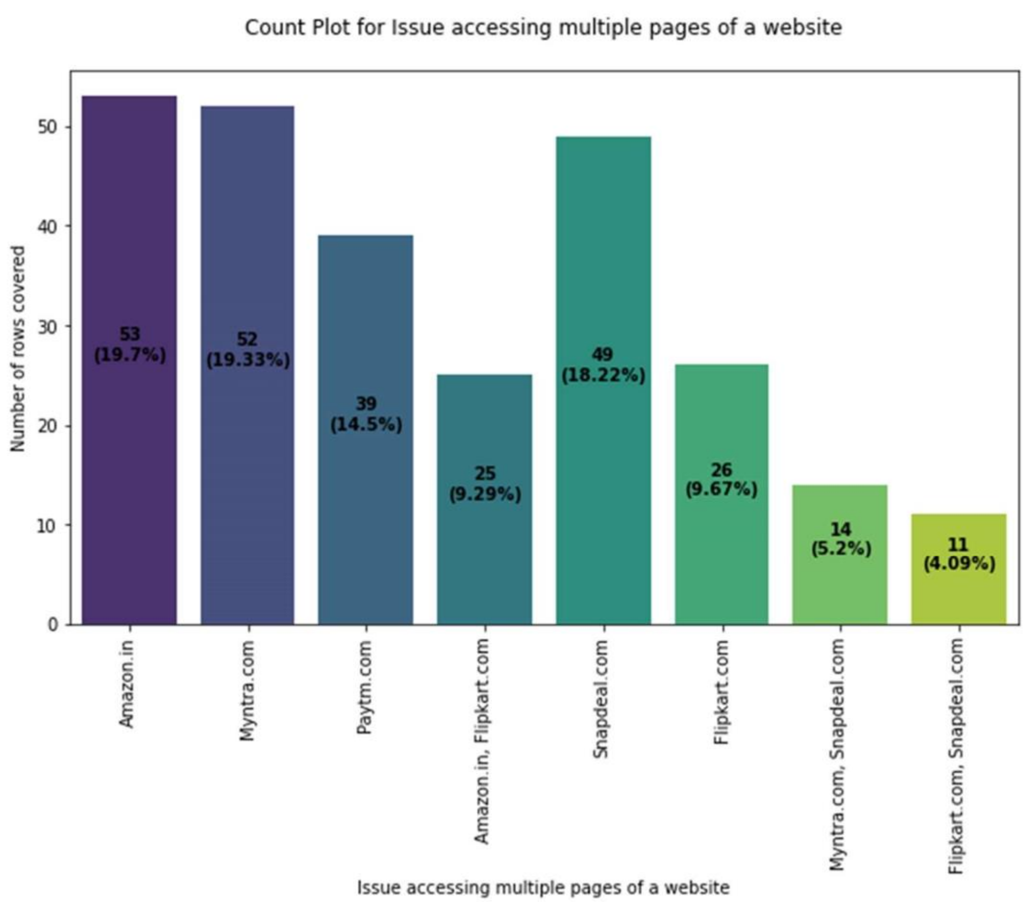


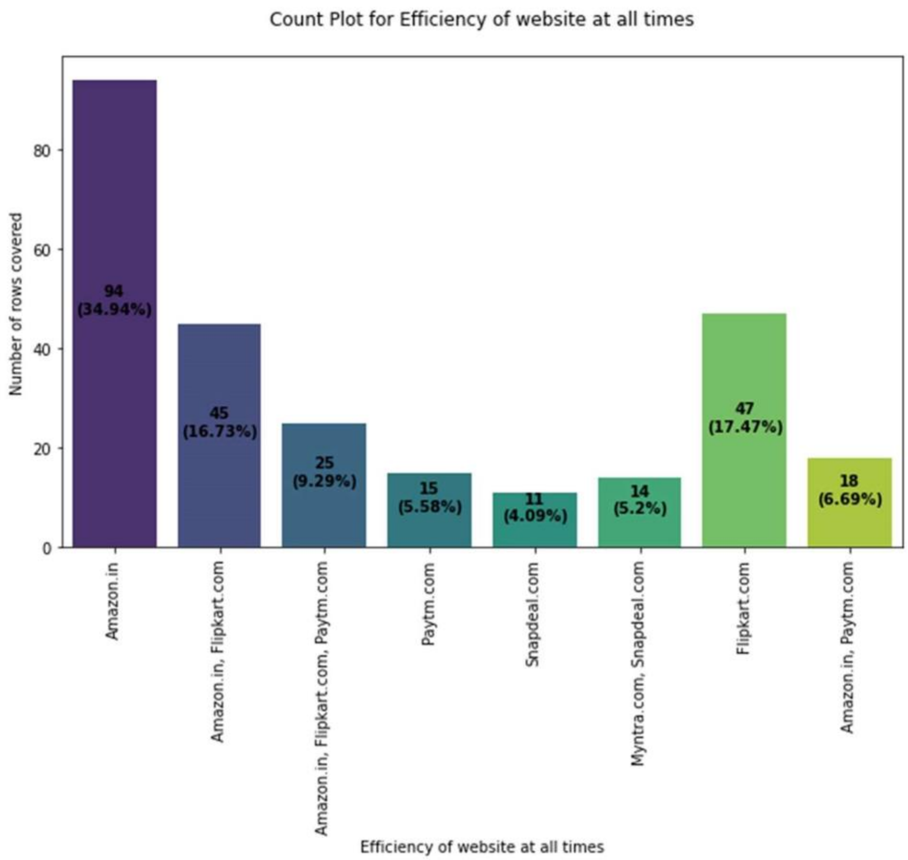
Count Plot for Limited mode of payment on special offer days
Number of rows covered
0
20
40
60
80
62 (23.05%)
87 (32.34%)
25 (9.29%)
13 (4.83%)
31 (11.52%)
29 (10.78%)
15 (5.58%)
7 (2.6%)
Amazon.in
Snapdeal.com
Paytm.com
Amazon.in, Paytm.com
Flipkart.com
Amazon.in, Flipkart.com
Paytm.com, Snapdeal.com
Myntra.com, Snapdeal.com
Limited mode of payment on special offer days

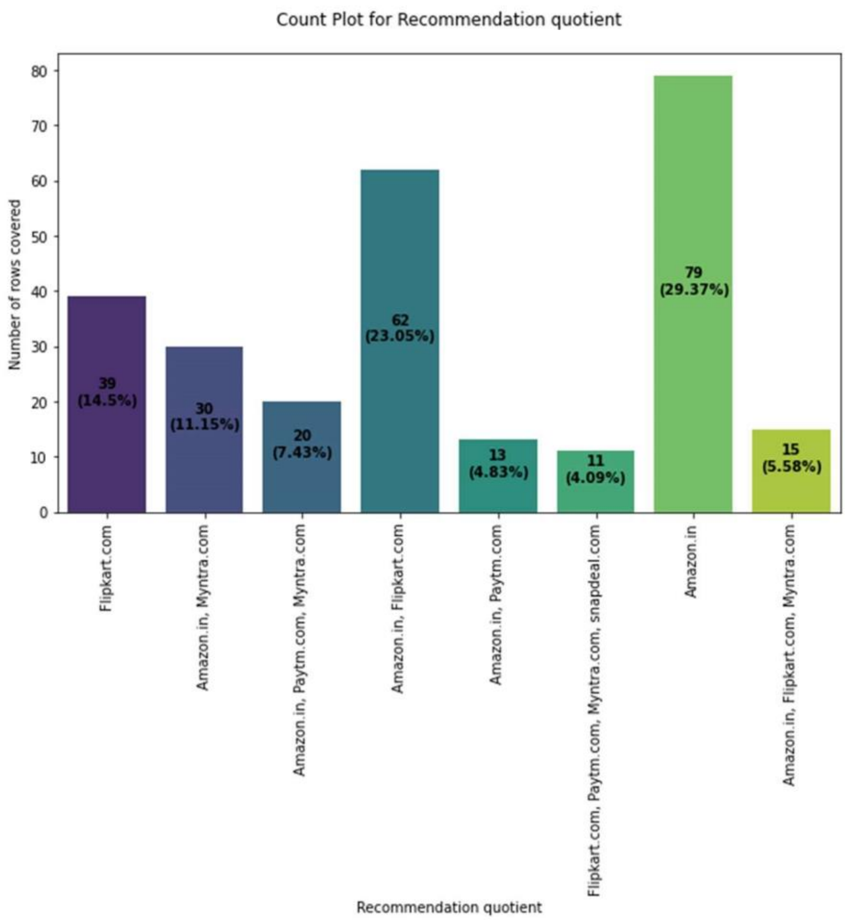


Count Plot for Delayed product delivery
Number of rows covered
0
10
20
30
40
50
60
70
72 (26.77%)
64 (23.79%)
44 (16.36%)
37 (13.75%)
26 (9.67%)
26 (9.67%)
Paytm.com
Snapdeal.com
Flipkart.com
Amazon.in
Paytm.com, Snapdeal.com
Myntra.com
Delayed product delivery


Count Plot for Change in GUI
Number of rows covered
0
20
40
60
80
100
20 (7.43%)
96 (35.69%)
63 (23.42%)
45 (16.73%)
30 (11.15%)
8 (2.97%)
7 (2.6%)
Flipkart.com
Amazon.in
Paytm.com
Amazon.in, Flipkart.com
Myntra.com
Snapdeal.com
Flipkart.com, Myntra.com
Change in GUI


Count Plot for Issue accessing multiple pages of a website
Number of rows covered
0
10
20
30
40
50
53 (19.7%)
52 (19.33%)
39 (14.5%)
25 (9.29%)
49 (18.22%)
26 (9.67%)
14 (5.2%)
11 (4.09%)
Amazon.in
Myntra.com
Paytm.com
Amazon.in, Flipkart.com
Snapdeal.com
Flipkart.com
Myntra.com, Snapdeal.com
Flipkart.com, Snapdeal.com
Issue accessing multiple pages of a website


Count Plot for Efficiency of website at all times
Number of rows covered
94
(34.94%)
45
(16.73%)
25
(9.29%)
15
(5.58%)
11
(4.09%)
14
(5.2%)
47
(17.47%)
18
(6.69%)
Amazon.in
Amazon.in, Flipkart.com
Amazon.in, Flipkart.com, Paytm.com
Paytm.com
Snapdeal.com
Myntra.com, Snapdeal.com
Flipkart.com
Amazon.in, Paytm.com
Efficiency of website at all times


Count Plot for Recommendation quotient
Number of rows covered
39
(14.5%)
30
(11.15%)
20
(7.43%)
62
(23.05%)
13
(4.83%)
11
(4.09%)
79
(29.37%)
15
(5.58%)
Flipkart.com
Amazon.in, Myntra.com
Amazon.in, Paytm.com, Myntra.com
Amazon.in, Flipkart.com
Amazon.in, Paytm.com
Flipkart.com, Paytm.com, Myntra.com, snapdeal.com
Amazon.in
Amazon.in, Flipkart.com, Myntra.com
Recommendation quotient

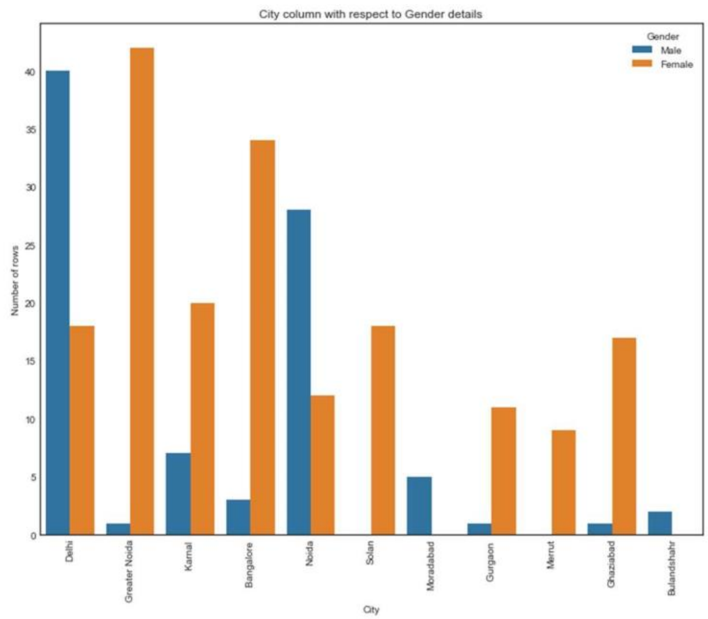
Bivariate Analysis:

I performed bivariate analysis using count plots again and changing the hue format. Please refer the code and the outputs in GIF formats below.

Code:

```
for col in df1:
    if col == "Gender":
        pass
    elif col == "Pin Code":
        pass
    else:
        plt.style.use('seaborn-white')
        plt.figure(figsize=(10,8))
        sns.countplot(x=col, data=df1, hue="Gender")
        plt.title("{} column with respect to Gender details".format(col))
        plt.tight_layout()
        plt.xticks(rotation=90)
        plt.ylabel("Number of rows")
        plt.show()
```

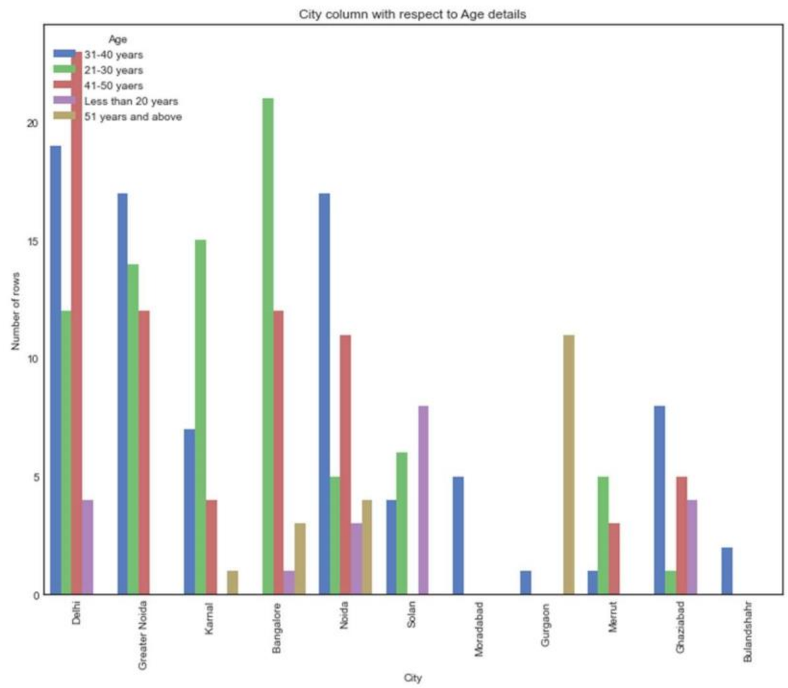
Output:



Code:

```
for col in df1:
    if col == "Age":
        pass
    elif col == "Pin Code":
        pass
    else:
        plt.style.use('seaborn-muted')
        plt.figure(figsize=(10,8))
        sns.countplot(x=col, data=df1, hue="Age")
        plt.title("{} column with respect to Age details".format(col))
        plt.tight_layout()
        plt.xticks(rotation=90)
        plt.ylabel("Number of rows")
        plt.show()
```

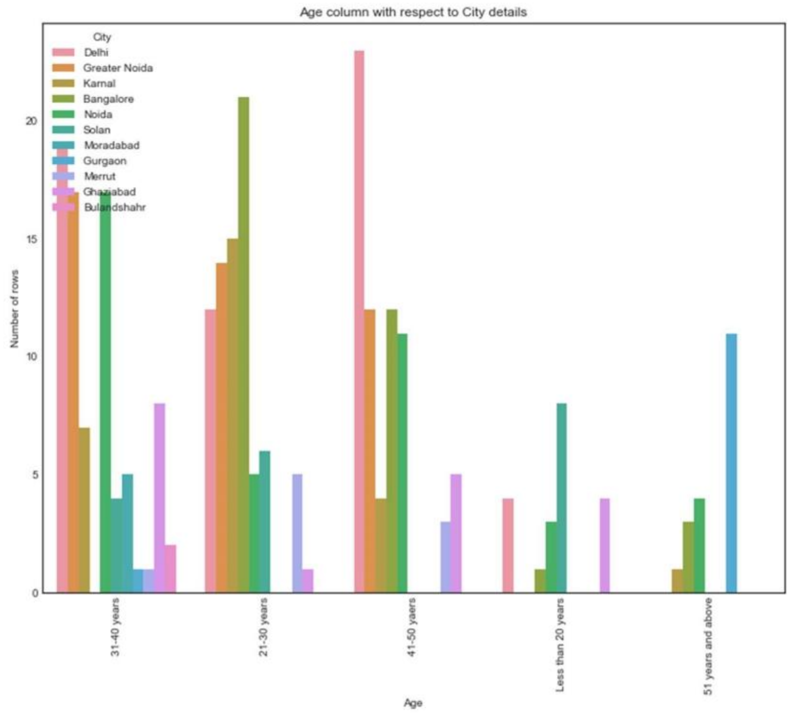
Output:



Code:

```
for col in df1:
    if col == "City":
        pass
    elif col == "Pin Code":
        pass
    else:
        plt.style.use('seaborn-colorblind')
        plt.figure(figsize=(10,8))
        sns.countplot(x=col, data=df1, hue="City")
        plt.title("{} column with respect to City details".format(col))
        plt.tight_layout()
        plt.xticks(rotation=90)
        plt.ylabel("Number of rows")
        plt.show()
```

Output:



Then I performed Ordinal Encoding on all the object datatype columns before I could proceed with any kind of multivariate analysis.

Code:

```
# Ordinal Encoding

oe = OrdinalEncoder()

def ordinal_encode(df, column):
    df[column] = oe.fit_transform(df[column])
    return df

oe_col = df1.columns
df=ordinal_encode(df1, oe_col)
df.head()
```

I was able to obtain object datatype conversion to numeric datatype with the help of Ordinal Encoding method.

Inferential Statistics:

First, I used Chi square test to check whether there is any relation between gender and online shopping companies and I got results that there is no relation between the two.

First Hypothesis

H0: There is no association between gender and e-retail company (Gender and e-retail company are independent)

H1: There is an association between gender and e-retail company

Chi Square Test

```
1]: import scipy.stats as stats
```

```
5]: # The 2 Columns needed are 1st and last
    dataset_table=pd.crosstab(df1['Gender'],
                              df1['Recommendation quotient'])
    dataset_table
```

```
: if chi_square_statistic>=critical_value:
      print("Reject H0,There is an association between gender and e-retail company recommended to a friend")
  else:
      print("Failed to Reject H0,There is no association between gender and e-retail company recommended to a friend")

  if p_value<=alpha:
      print("Reject H0,There is an association between gender and e-retail company recommended to a friend")
  else:
      print("Failed to Reject H0,There is no association between gender and e-retail company recommended to a friend")

  Failed to Reject H0,There is no association between gender and e-retail company recommended to a friend
  Failed to Reject H0,There is no association between gender and e-retail company recommended to a friend
```

Second Hypothesis

Then I checked if there is any relation between age and online shopping company and I got following results

H0: There is no association between age and e-retail company that person would recommend to a friend(age and e-retail company are independent

H1: There is an association between age and e-retail company that person would recommend to a friend

```
if chi_square_statistic>=critical_value:
    print("Reject H0,There is an association between age and e-retail company that person would recommend to a friend")
else:
    print("Failed to Reject H0, Therefore, There is no association between age and e-retail company that person would recommend t

if p_value<=alpha:
    print("Reject H0,There is an association between age and e-retail company that person would recommend to a friend")
else:
    print("Failed to Reject H0, Therefore, There is no association between age and e-retail company that person would recommend t
```

```
Failed to Reject H0, Therefore, There is no association between age and e-retail company that person would recommend to a frien
d
Failed to Reject H0, Therefore, There is no association between age and e-retail company that person would recommend to a frien
d
```

Multivariate Analysis:

For multivariate analysis I made use of Pandas Profiling in my Jupyter Notebook. pandas-profiling is an open-source Python module with which we can quickly do an exploratory data analysis with just a few lines of code. It generates interactive reports in web format that can be presented to any person, even if they don’t know programming.

It also offers report generation for the dataset with lots of features and customizations for the report generated. In short, what pandas-profiling does is save us all the work of visualizing and understanding the distribution of each variable. It generates a report with all the information easily available.

I took a screenshot of the initial output for pandas-profiling however we could scroll through for detailed analysis report on our dataset whilst browsing through different tabs as well. The single line code to get the embedded report is shown below:

```
pandas_profiling.ProfileReport(df1)
```

```
Summarize dataset: 100% 84/84 [00:42<00:00, 1.19it/s, Completed]
Generate report structure: 100% 1/1 [00:38<00:00, 38.55s/it]
Render HTML: 100% 1/1 [00:04<00:00, 4.32s/it]
```

Output:

Pandas Profiling Report — Overview | Variables | Interactions | Correlations | Missing values | Sample | Duplicate rows

Overview

Overview | Warnings 138 | Reproduction

Dataset statistics

Number of variables	71
Number of observations	269
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	60
Duplicate rows (%)	22.3%
Total size in memory	149.3 KiB
Average record size in memory	568.5 B

Variable types

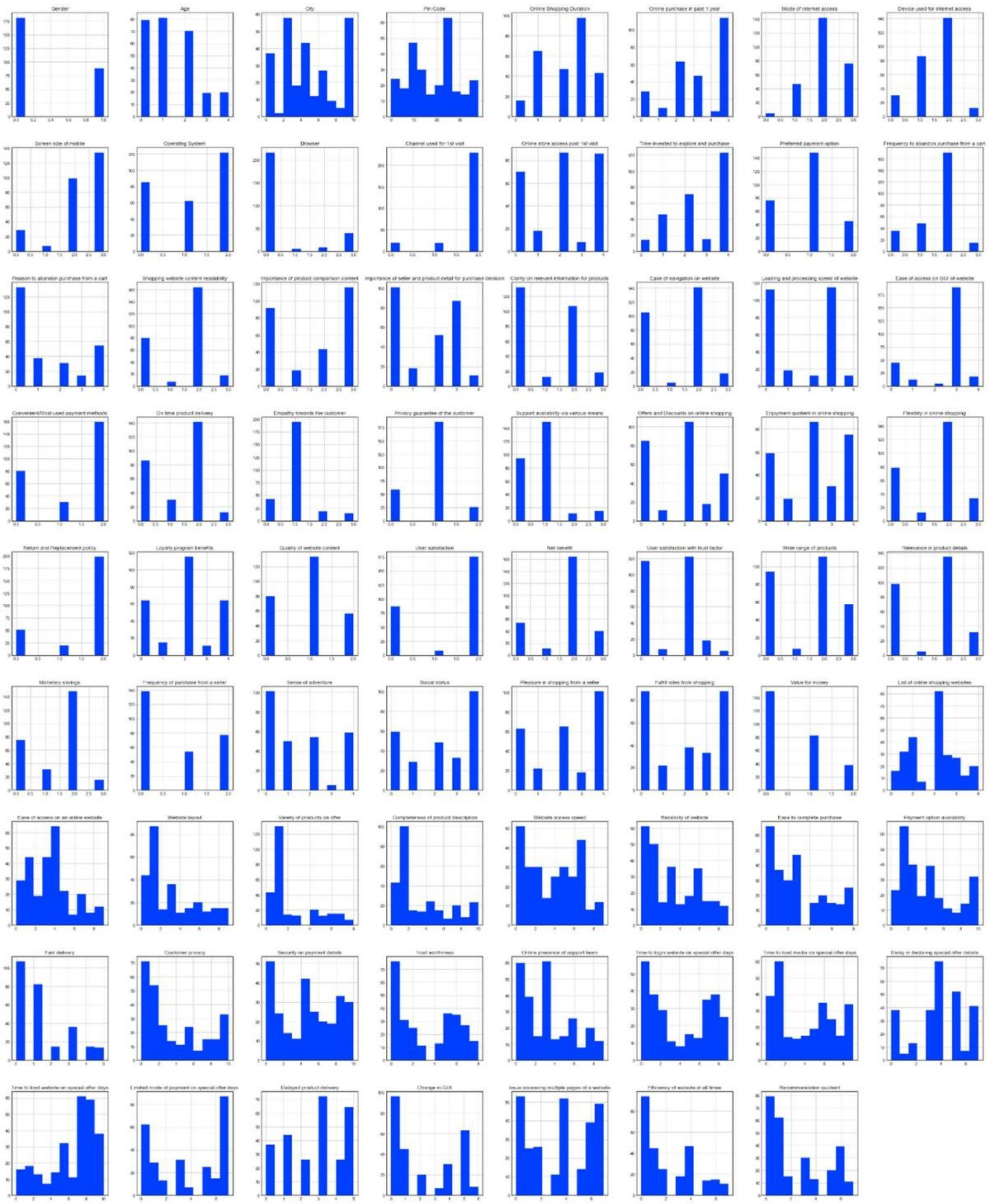
Categorical	70
Numeric	1

Along with the pandas-profiling method I generated a histogram post encoding all my column values.

Code:

```
plt.style.use('seaborn-bright')

df.hist(figsize=(40,50))
plt.show()
```

Output:



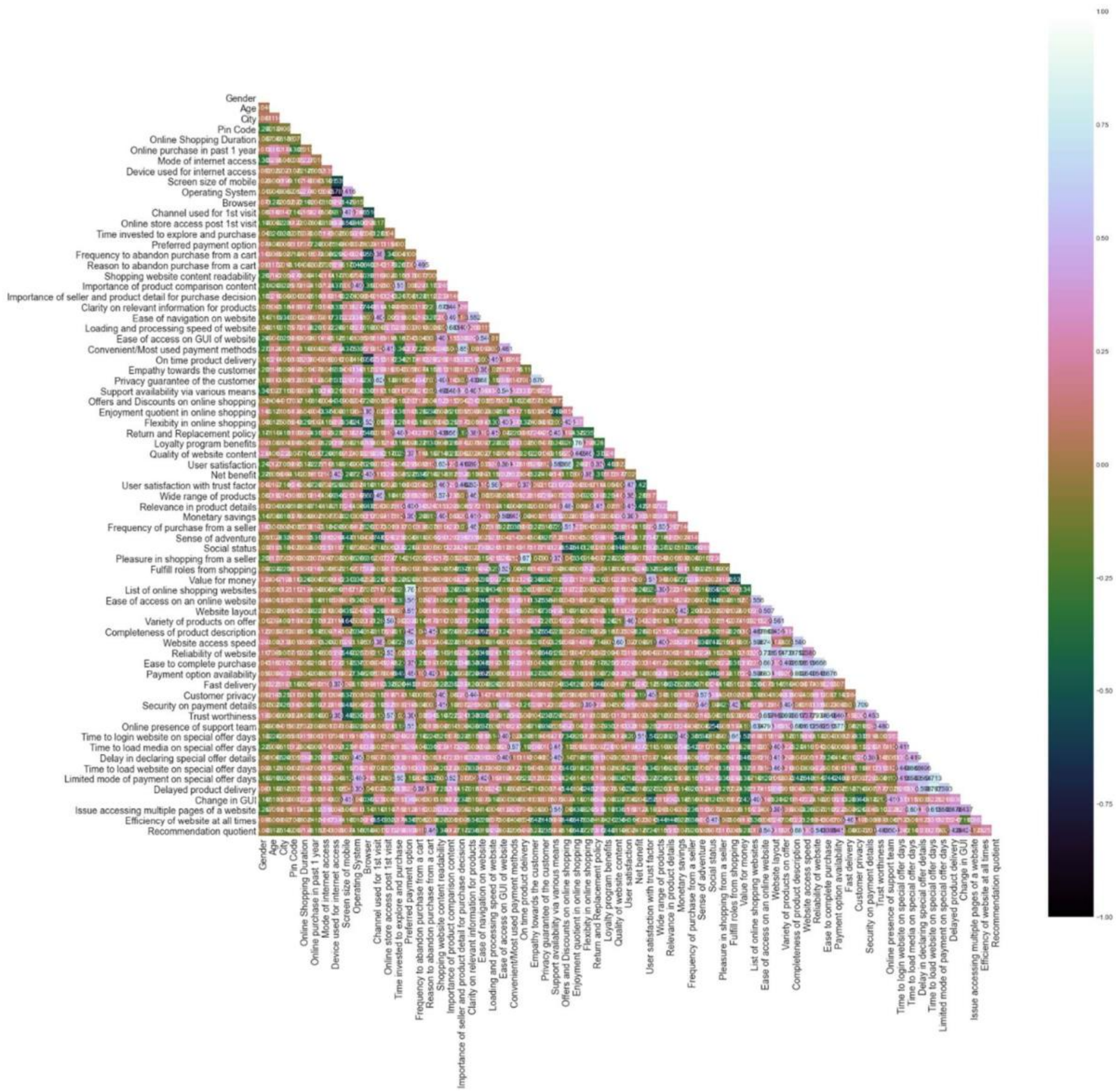
I generated a heatmap using the correlation values between the dataset columns. The correlation details are bifurcated majorly into positive and negative parts.

Positive correlation - A correlation of +1 indicates a perfect positive correlation, meaning that both variables move in the same direction together.
Negative correlation - A correlation of –1 indicates a perfect negative correlation, meaning that as one variable goes up, the other goes down.

Code:

```
upper_triangle = np.triu(df1.corr())
plt.figure(figsize=(25,25))
sns.heatmap(df1.corr(), vmin=-1, vmax=1, annot=True, square=True, fmt='0.3f',
            annot_kws={'size':10}, cmap="cubehelix", mask=upper_triangle)
plt.xticks(fontsize=14)
plt.yticks(fontsize=14)
plt.show()
```

Output:

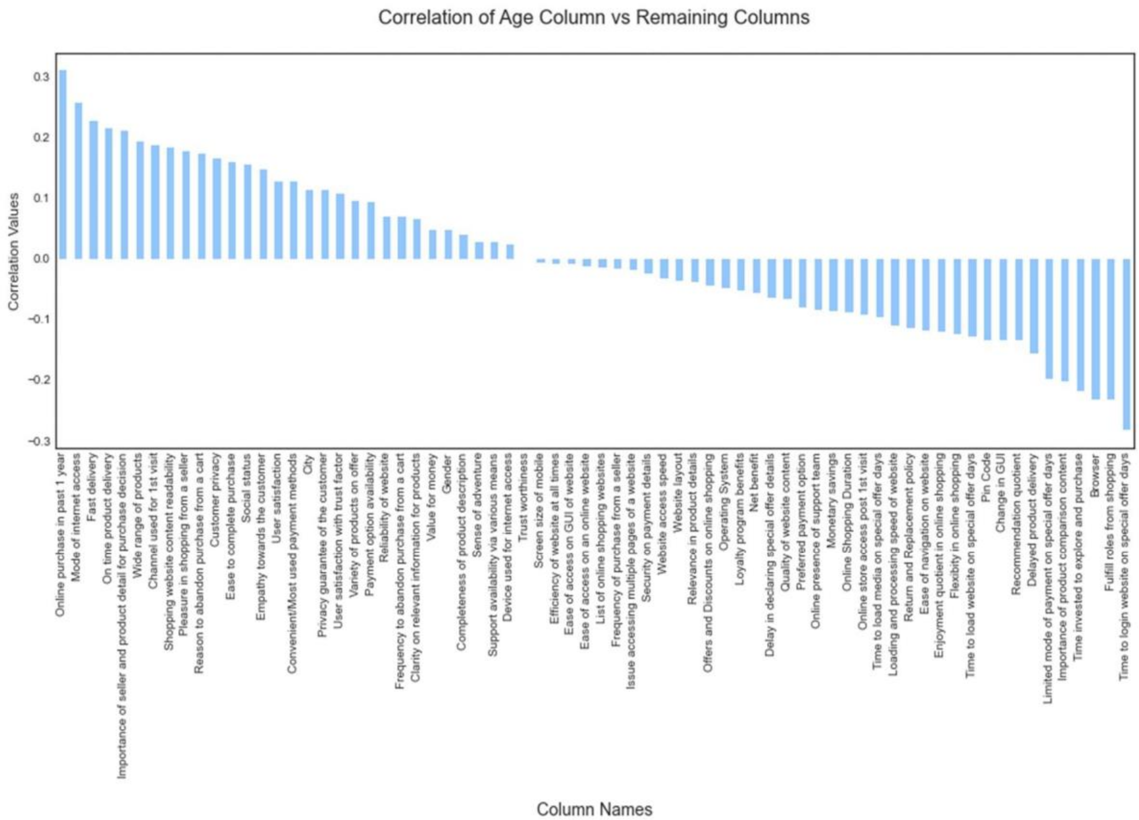


In the above heatmap due to lot of columns we are not able to see the correlation details however we can observe the color-coding details and get a hint that there is no multi collinearity concern between the column values.

Code:

```
plt.style.use('seaborn-pastel')

column_names = df1.columns
for col in df1[column_names]:
    df_corr = df1.corr()
    plt.figure(figsize=(16,6))
    df_corr[col].sort_values(ascending=False).drop(col).plot.bar()
    plt.title("Correlation of {} Column vs Remaining Columns\n".format(col), fontsize=16)
    plt.xlabel("\nColumn Names", fontsize=14)
    plt.ylabel("Correlation Values", fontsize=12)
    plt.show()
```

Output:



Since the heatmap was not clear in terms of its values I have generated this bar plot for each column vs remaining column showing the positive and negative correlation data.

Inference:

1. Amazon.com To be improved:
    - During promotions, try to give a disturbance free shopping experience to customers.
    - Give more payment options to customers.
    - Try to give price early during promotion.
    - Reduce the delivery time of the products.

Positive feedback summary:

- Convenient to use and also a good website for shopping.
- Fast delivery of products.
- Availability of complete information of the products.
- Presence of online assistance through multi-channels.
- Reliable website or app, perceived trustworthiness.

2. Flipkart.com To be
improved:

- During promotions, try to give a disturbance free shopping experience to customers.
- Give more payment options to customers.
- Try to give the price early during promotion.
- Reduce the delivery time of the products.
- Flipkart and Amazon almost share the same feedbacks with varying percentages as the only difference.

Positive feedback summary:

- Convenient to use and also a good website for shopping.
- Fast delivery of products.
- Availability of complete information of the products.
- Presence of online assistance through multi-channels.
- Reliable website or app, perceived trustworthiness.
- Wild variety of products to offer.

3. Myntra.com To
be improved:

- During promotions, try to give a disturbance free shopping experience to customers.
- Try to give the price early during promotions.
- Reduce the delivery time of the products during promotions.

Positive feedback summary:

- Convenient to use and also a good website.
- Availability of several payment options.
- Faster products delivery.
- Complete information of products available.
- Reliable website or app, perceived trustworthiness.
- Wild variety of product to offer

4. Paytm.com
To be improved:

- Reduce the delivery time of the products during promotions.
- Try to give the price early during promotion.
- During promotions, try to give a disturbance free shopping experience to customers.
- Late declaration of price and discounts.
- Frequent disturbance is occurring while moving from one page to another.

Positive feedback summary

- Convenient to use and a good website.
- Quickness to complete a purchase.
- About 64% of the customers feel that either web or app is reliable.
- Around 20% of the customers believe that Paytm has a wild variety of products on offer.

5. Snapdeal.com To be improved:

- Reduce the delivery time of the products during promotions.
- Try to give the price early during promotion.
- During promotions, try to give a disturbance free shopping experience to customers.
- Late declaration of price and discounts.
- No one has expressed to recommend Snapdeal to a contact as it has the most negative feedbacks among all other websites.

Positive feedback summary:

- Convenient to use.
- 54% of the customers are happy about the availability of financial information security.

Conclusion:

Based on overall observations, the first 47 features provide insights into how e-tailer is helpful & growing based on customer inputs. The data explained how the online platform has been used more often in which CITY, PIN CODE, AGE etc. It also showed that in some factors there is less importance given to contribute to the success of an e-commerce store, so based on that we could remove those factors & keep all the important factors, also we could improve on some factors that influence the online customers repeat purchase intention.

Apart from the first 47 features, the rest of the features showed which online platform has been used more based on the success factors. Based on the case study for customer activation & retention, Amazon is most reliable and has been fulfilled the customer requirements. After Amazon, data showed Flipkart has been used more for online shopping.

The case study from Indian e-commerce customers showed Amazon and Flipkart has been used mostly for Online Shopping and most recommended by Friends. So, based on the research factors, Amazon & Flipkart are the ecommerce platform, which are having the combination of both utilitarian and hedonistic values to keep the repeat purchase intention (loyalty) positively.

Future Work:

- I will need to perform some preprocessing on the data for example using the scaling techniques
- Since I have mostly categorical data present in the dataset, I am not going to worry about removing outliers or skewness
- Need to build some unsupervised machine learning models
- Will have to verify the clustering or association algorithm details that can be used on the dataset
- Some algorithms that I intend to work upon are k-means clustering, knearest neighbors for unsupervised machine learning, hierarchal clustering, apriori algorithm and neural networks.