

Customer Retention Case Study Report



Prepared by Sonali Daga

Data Science Intern at Flip Robo Technologies



SME Name:

Sristi Maan

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 **Sristi Maan** 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 e-commerce store have been identified as: **Service quality, System quality, Information quality, Trust and Net benefit.**

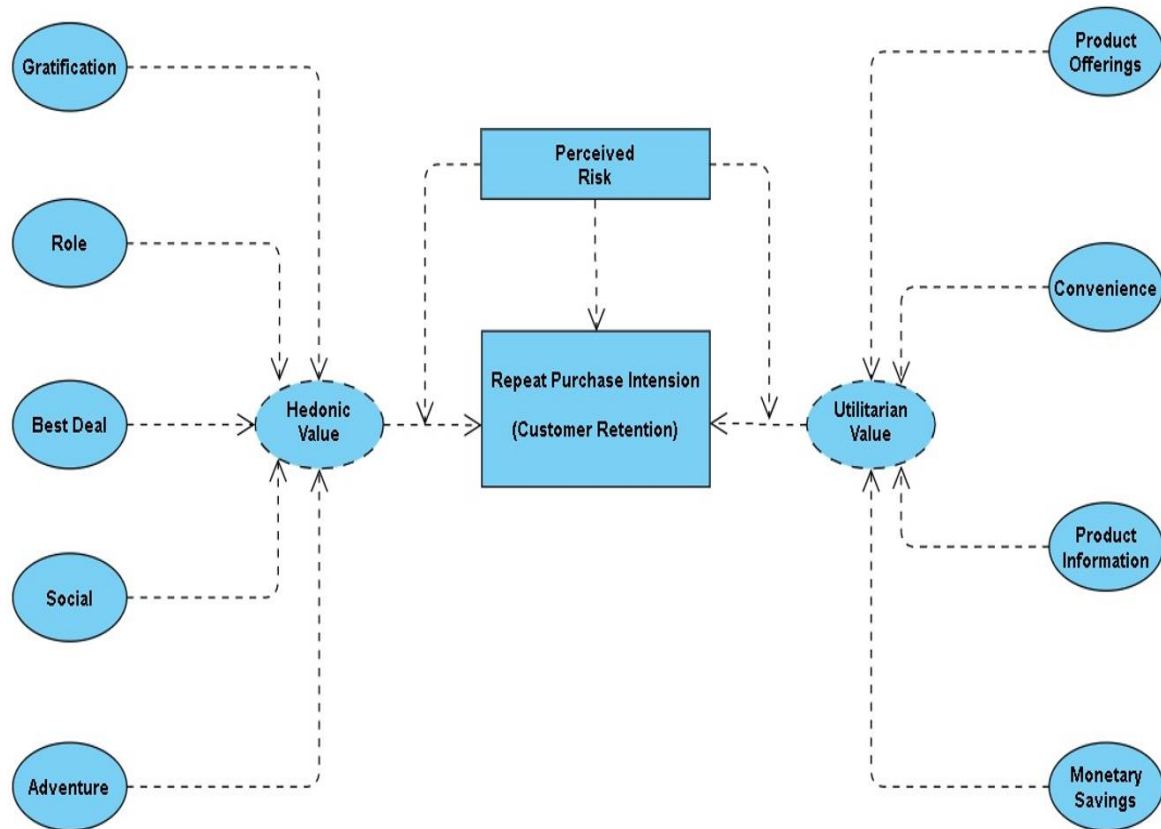
The research furthermore investigated the factors that influence the online customers repeat purchase intention.

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 use case diagram above, you can see that the intent of repeated purchases is basically based on a customer retention strategy based on pleasure value and utilitarian value. We also find that the perceived risks affect your purchase and repurchase intent. Pleasure value is made up of five key components: satisfaction, role, best offer, social aspects, and adventurous standards. In terms of utility value, we have product offerings, convenience, product information, and money savings.

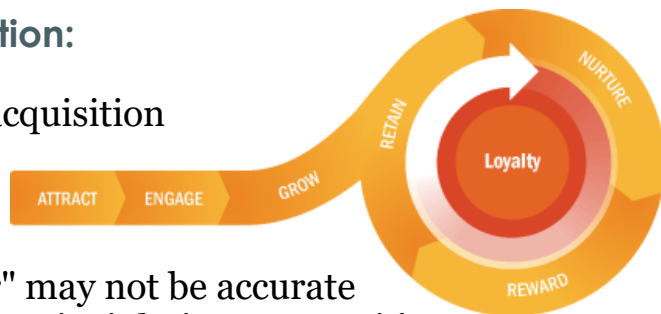
Motivation for the Problem Undertaken

The main goal of this project is to analyze whether users are purchasing products from e-commerce websites. How did you provide feedback to these websites based on multiple positive and negative factors, and based on user details based on factors such as 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 long-term 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.

Importing dataset:

```
: #importing required libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import datetime as dt

import warnings
warnings.filterwarnings('ignore')
```

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

```
#Importing dataset
df = pd.read_excel(r'customer_retention_dataset.xlsx')
df.head()
```

Exploratory Data Analysis (EDA):



Exploratory Data Analysis, or EDA, is an important step in any Data Analysis or Data Science project. EDA is the process of investigating the dataset to discover patterns, and anomalies (outliers), and form hypotheses based on our understanding of the dataset.

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.

```
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.

```
#Checking null values in the dataset
df.isnull().sum()
```

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

```
#Describing the object datatypes columns
df.describe(include="all").T
```

```
#Checking unique values of each column
df.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 df.columns:
    print(df[i].value_counts())
    print('*****')
```

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 used two for loops to generate a count chart for all columns, showing the percentage of data coverage and the observations after each chart as user input.

```

#Count plot for all categorical columns

for col in df[categorical_columns]:
    print(df[col].value_counts())
    plt.figure(figsize=(15,10))
    col_name = col
    values = df[col_name].value_counts()
    index = 0
    ax = sns.countplot(df[col_name], palette="viridis")

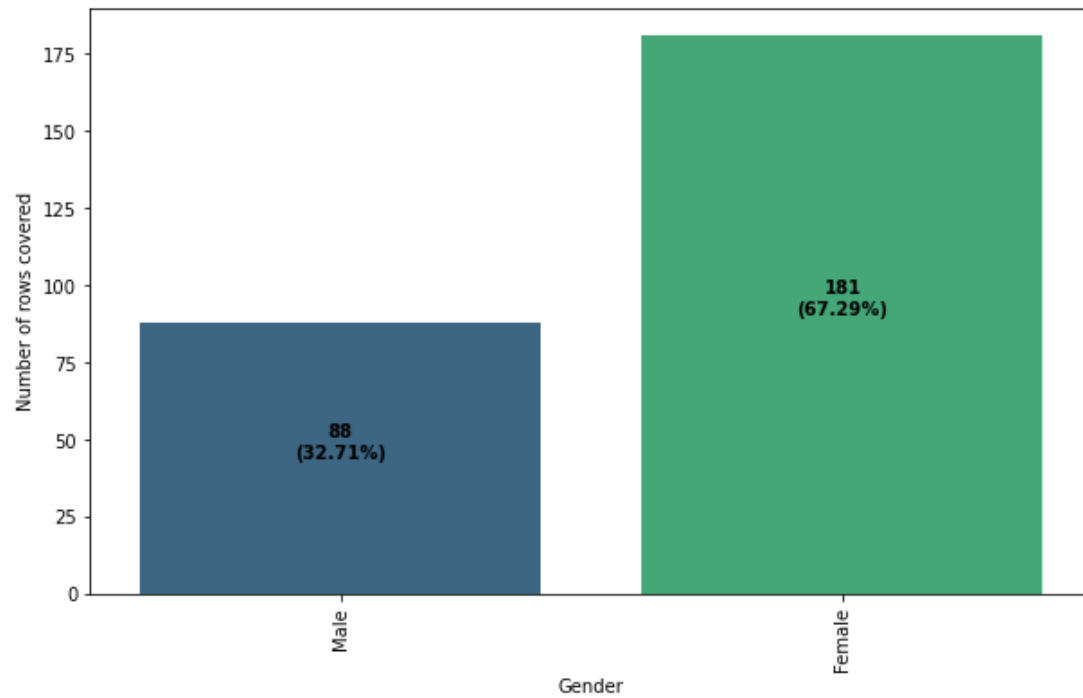
    for i in ax.patches:
        h = i.get_height() # getting the count of each value
        t = len(df[col_name]) # getting the total number of records using length
        s = f"{h}\n({round(h*100/t,2)}%)" # making the string for displaying in count bar
        plt.text(index, h/2, s, ha="center", fontweight="bold")
        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()
    print('Observation: \n')
    input('- ')
    print('-----'*20)
    print('\n')

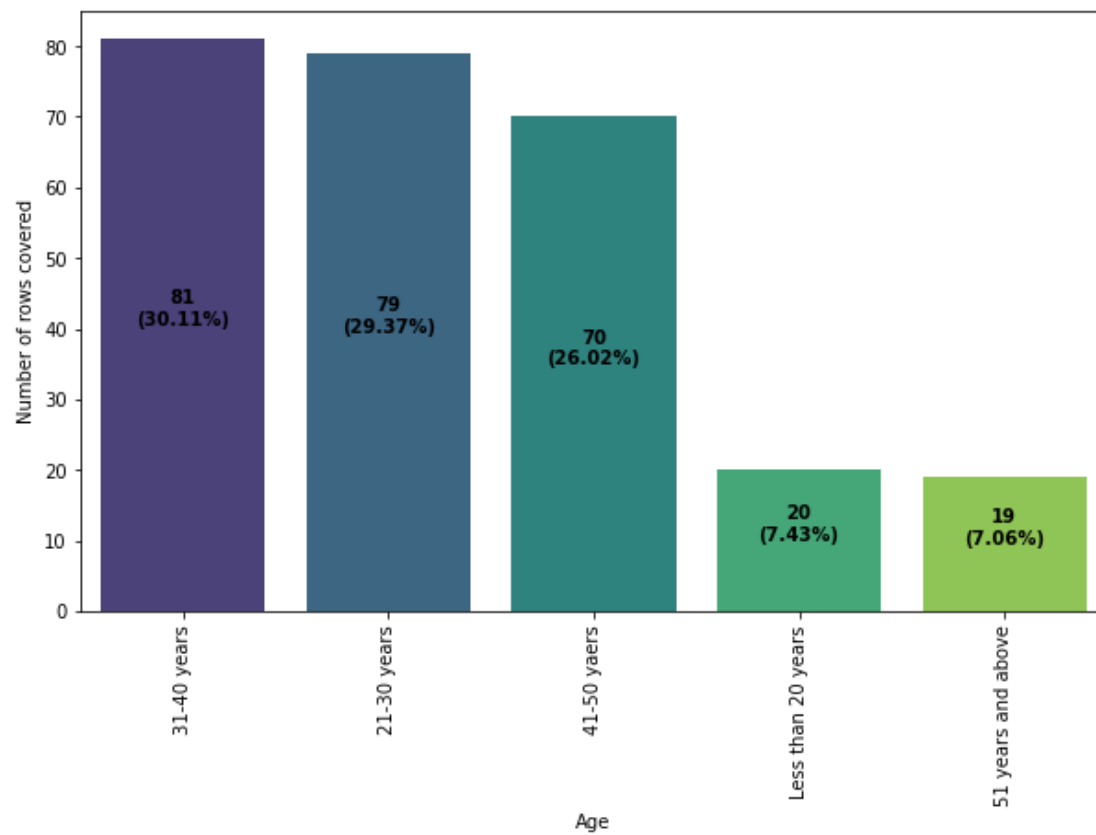
```

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

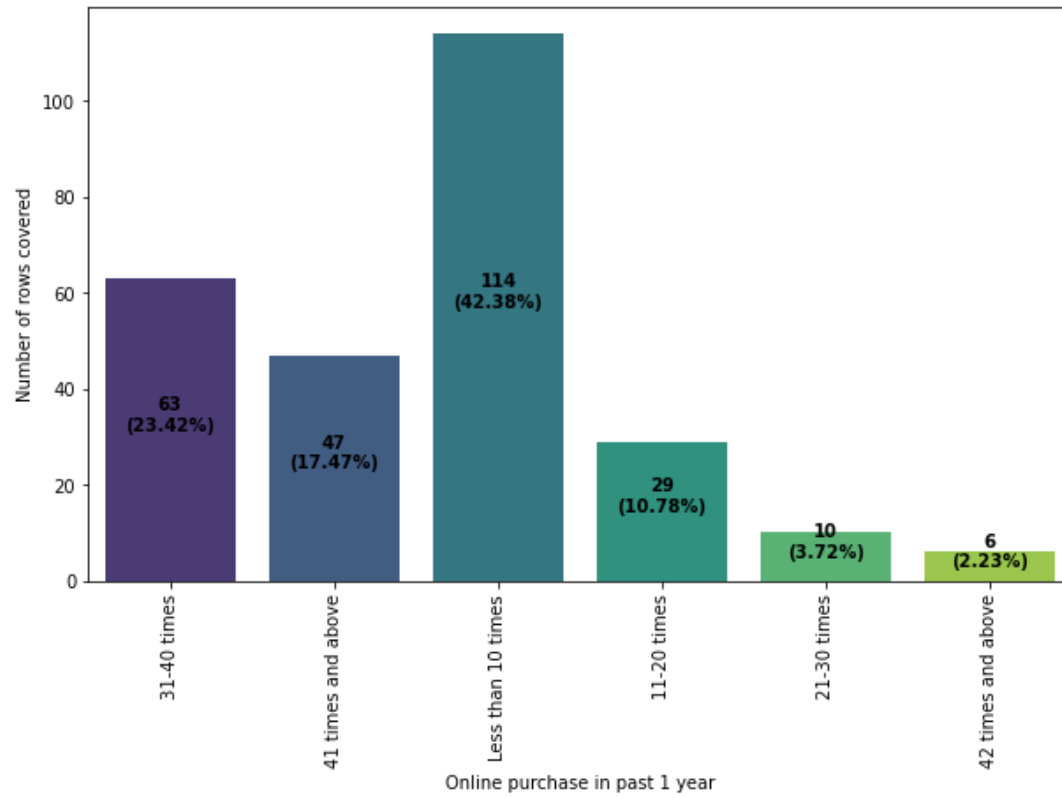
Count Plot for Gender



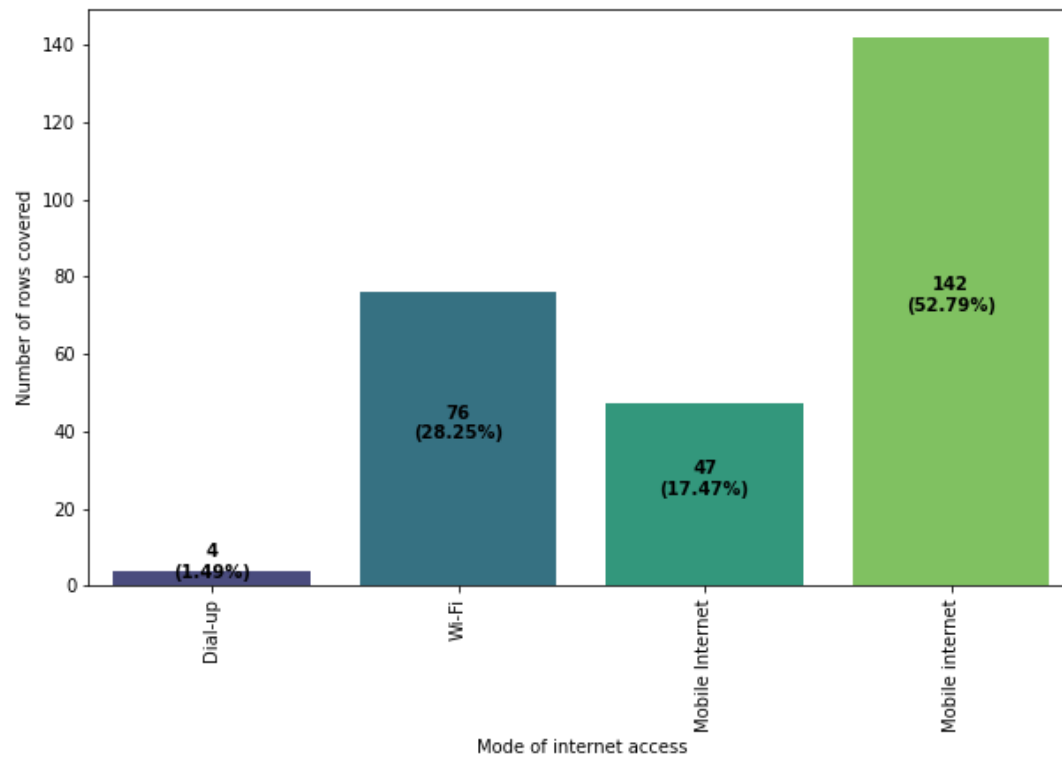
Count Plot for Age



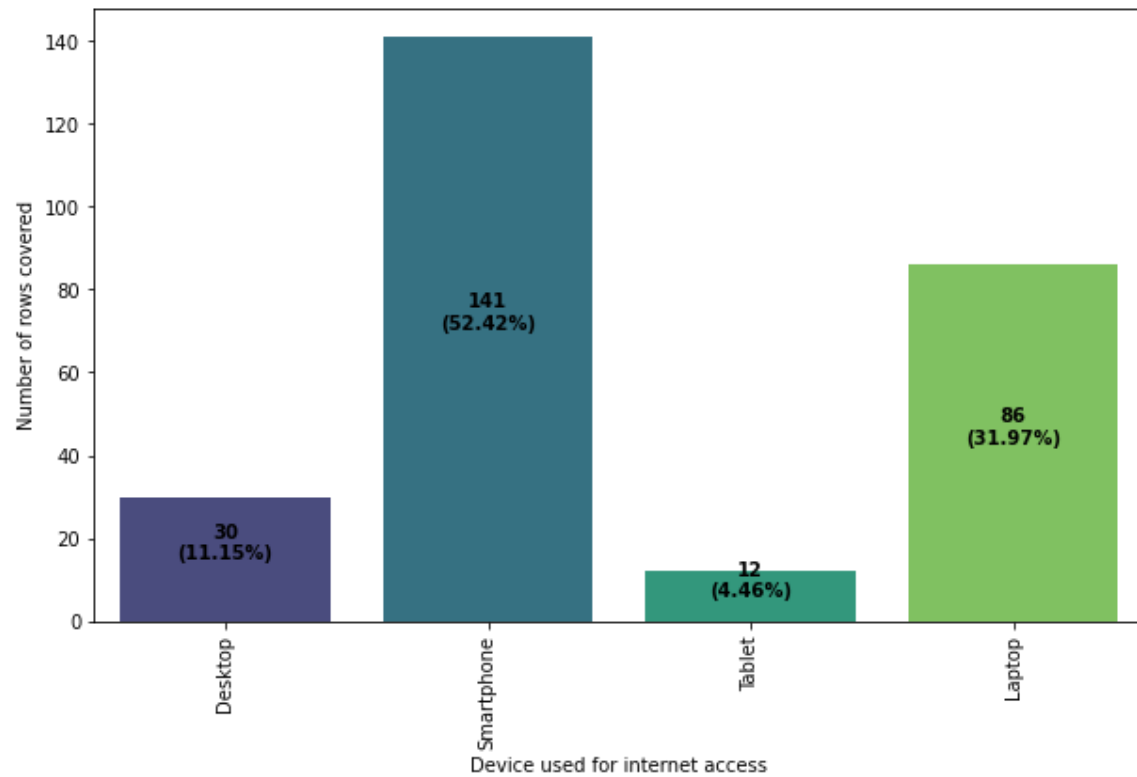
Count Plot for Online purchase in past 1 year



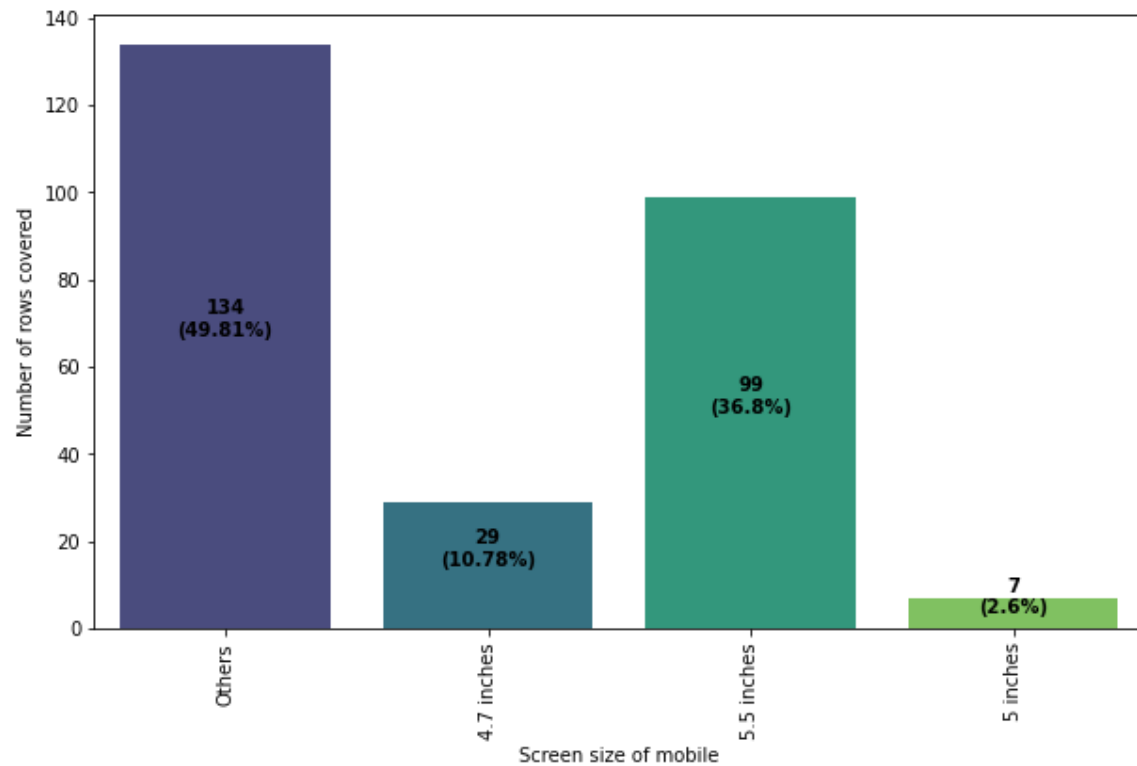
Count Plot for Mode of internet access

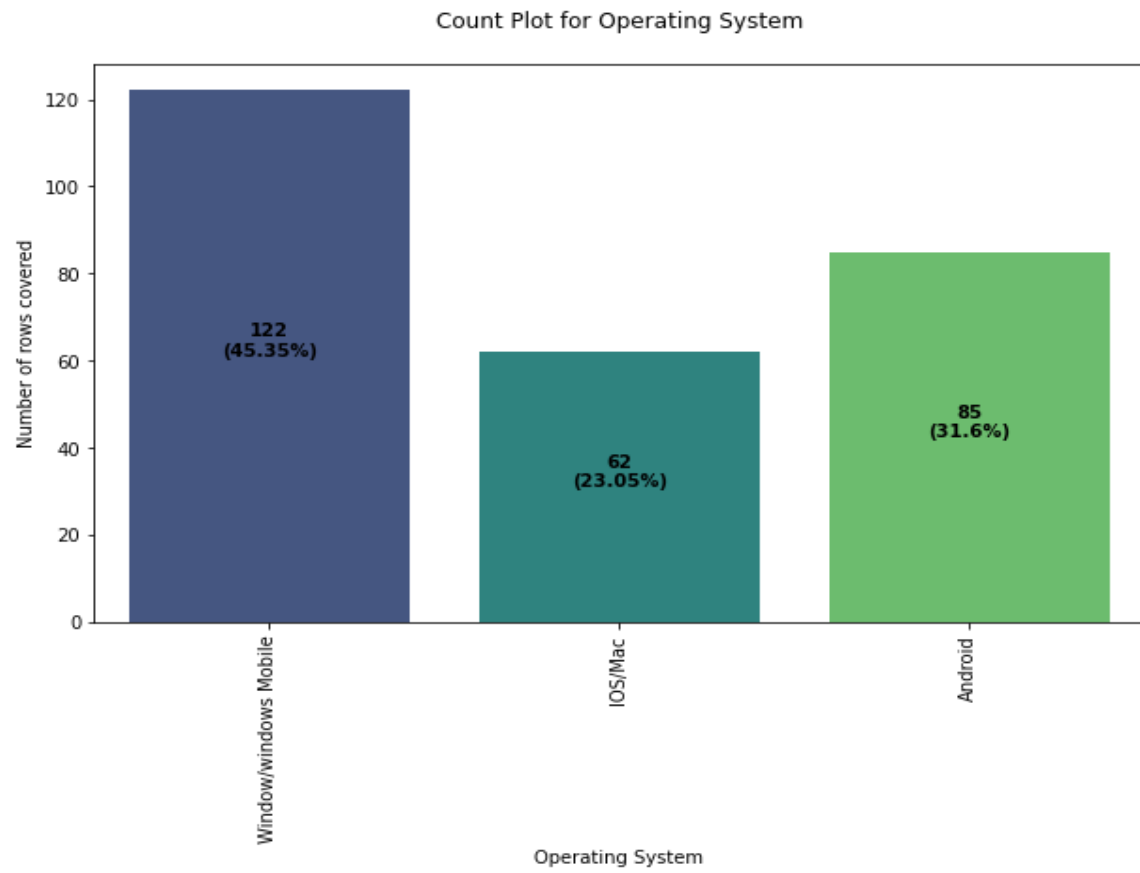


Count Plot for Device used for internet access

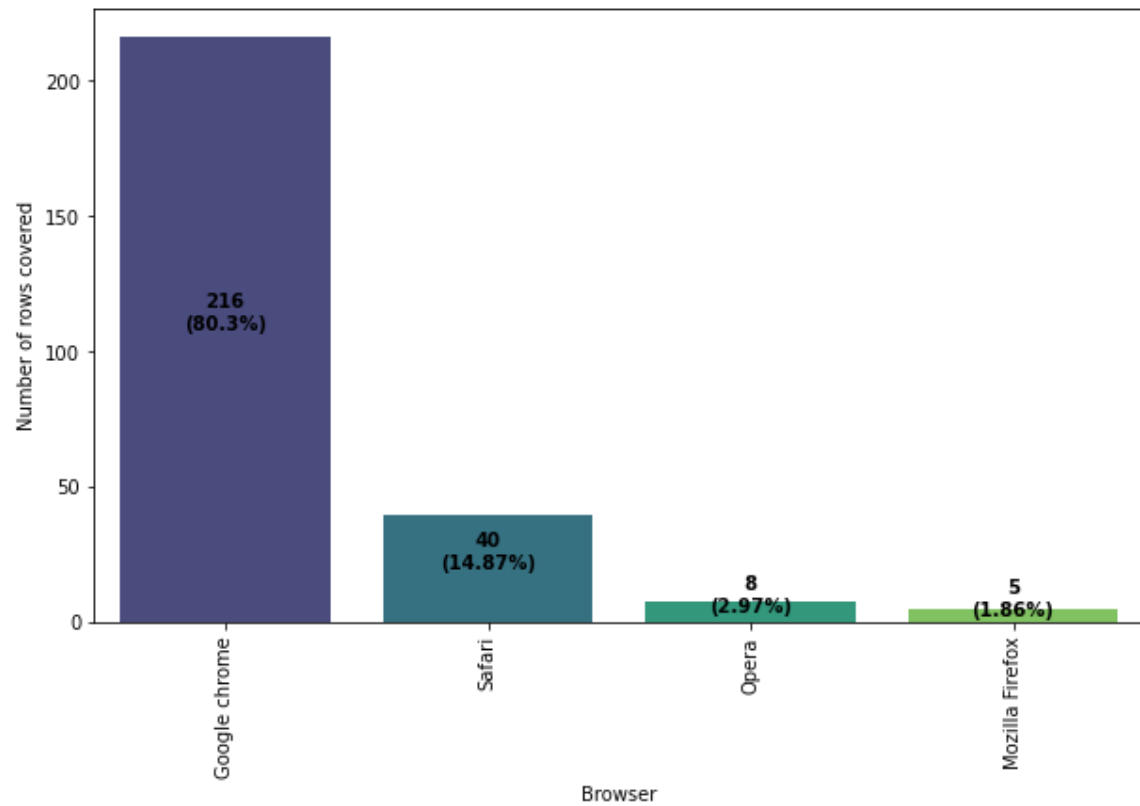


Count Plot for Screen size of mobile

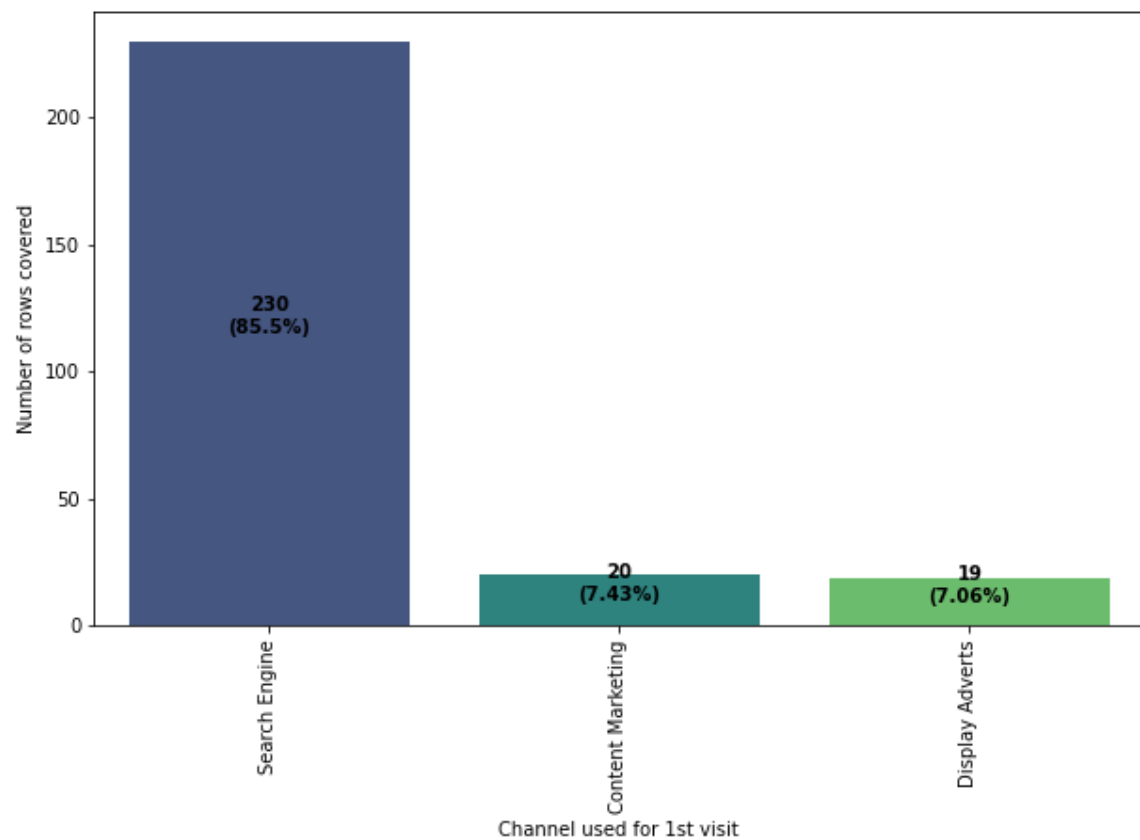




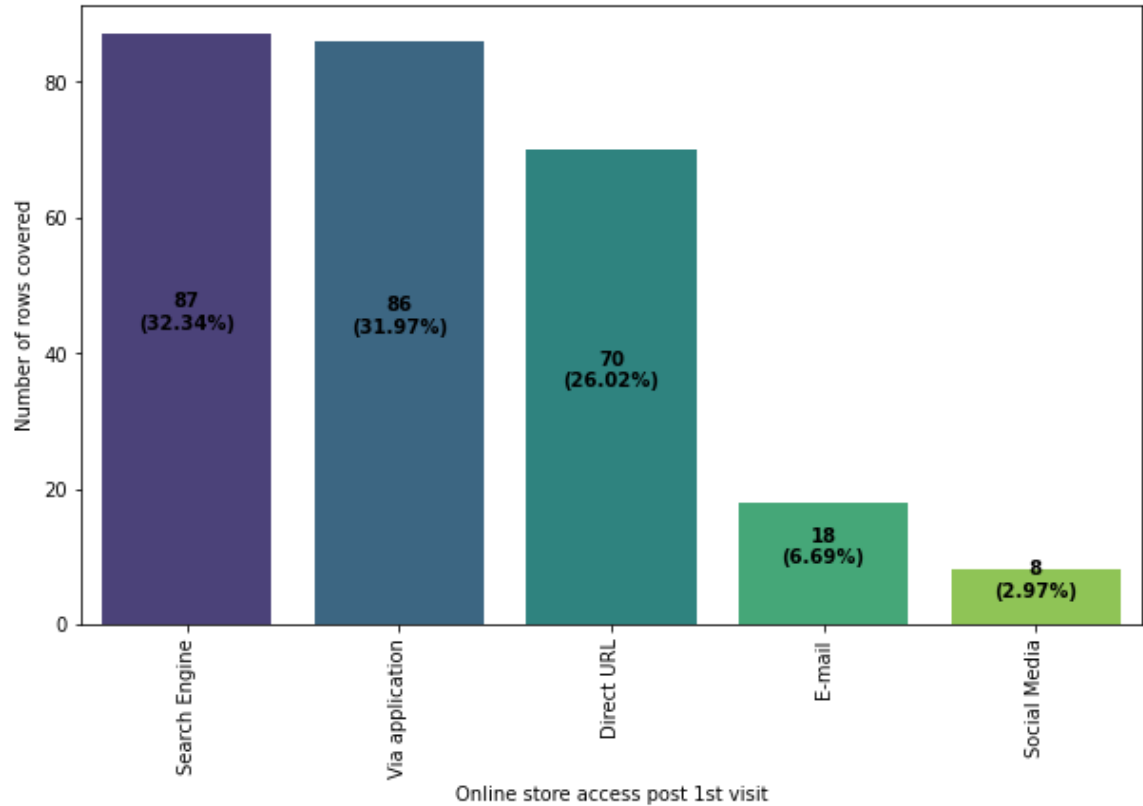
Count Plot for Browser



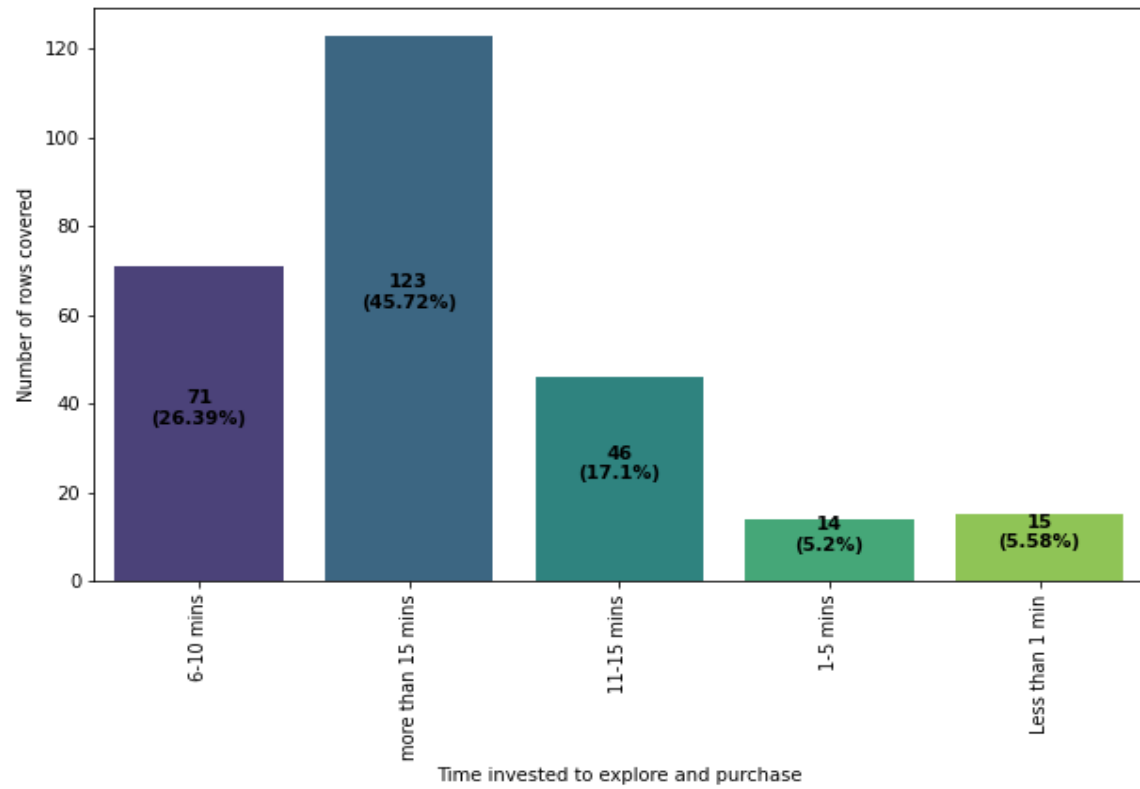
Count Plot for Channel used for 1st visit

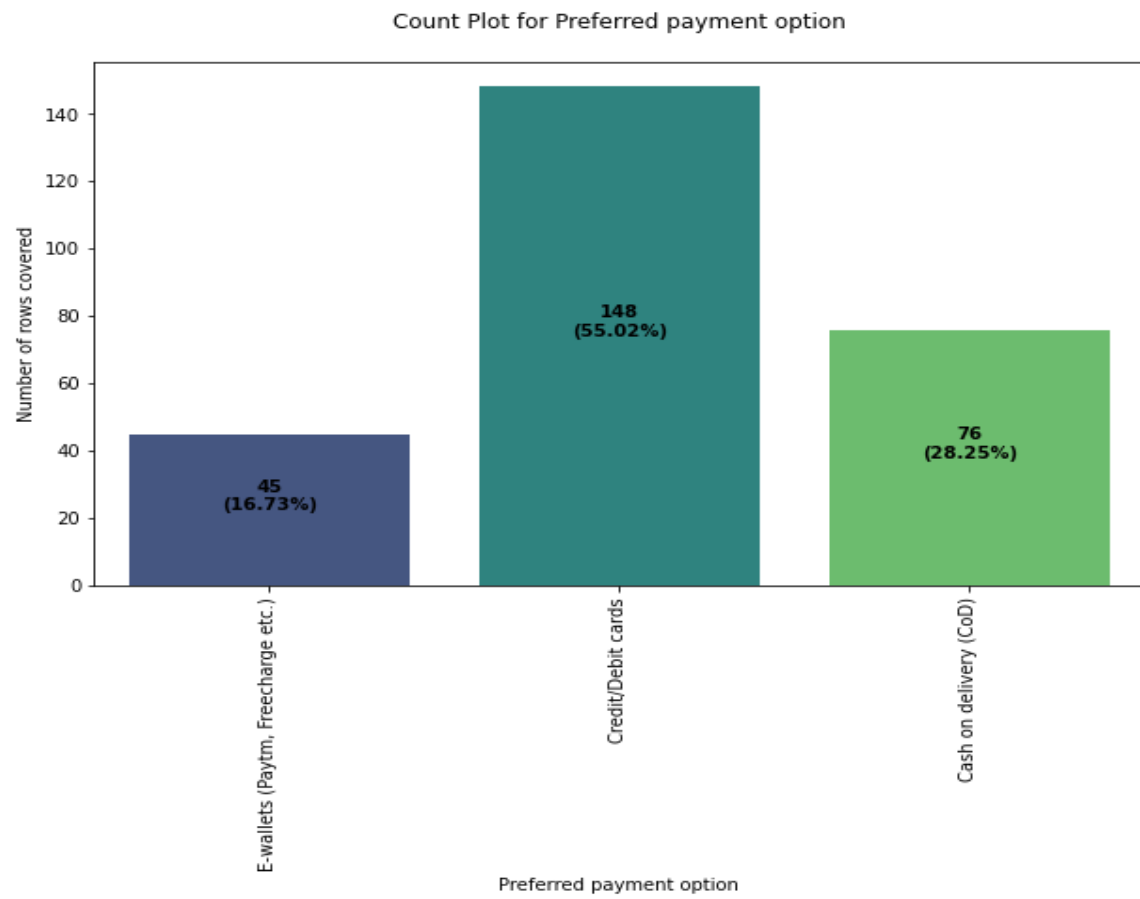


Count Plot for Online store access post 1st visit

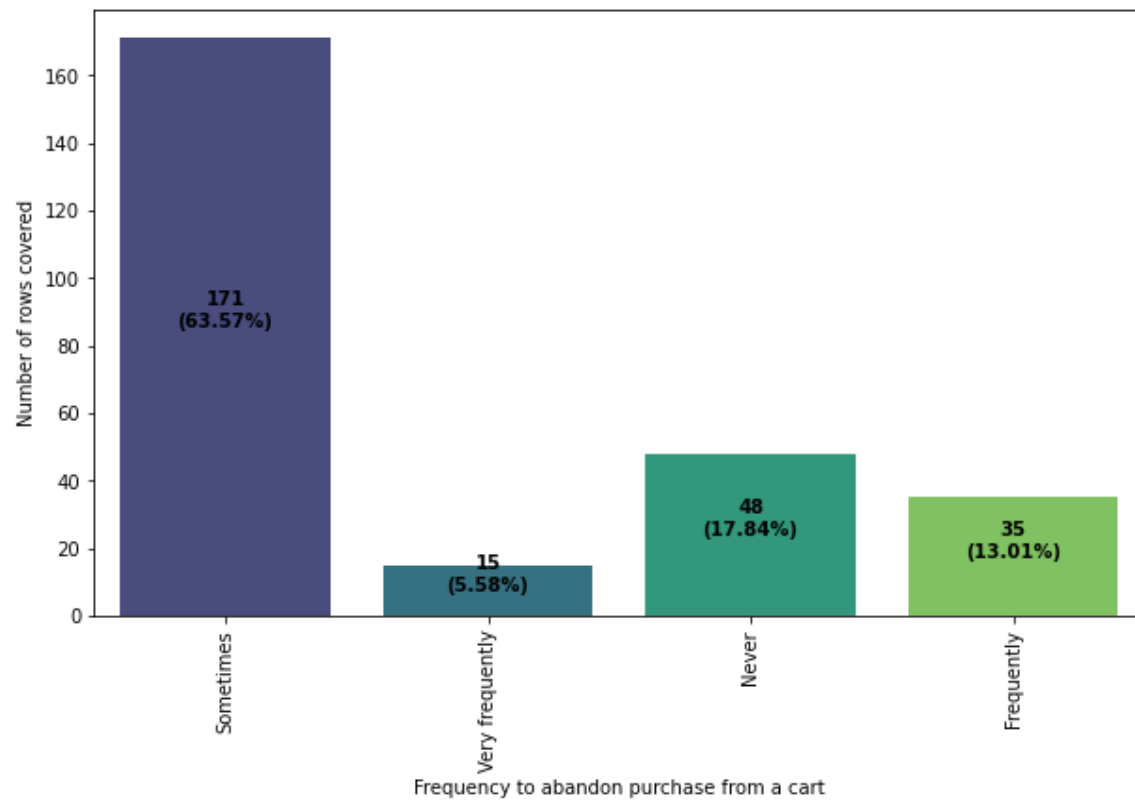


Count Plot for Time invested to explore and purchase

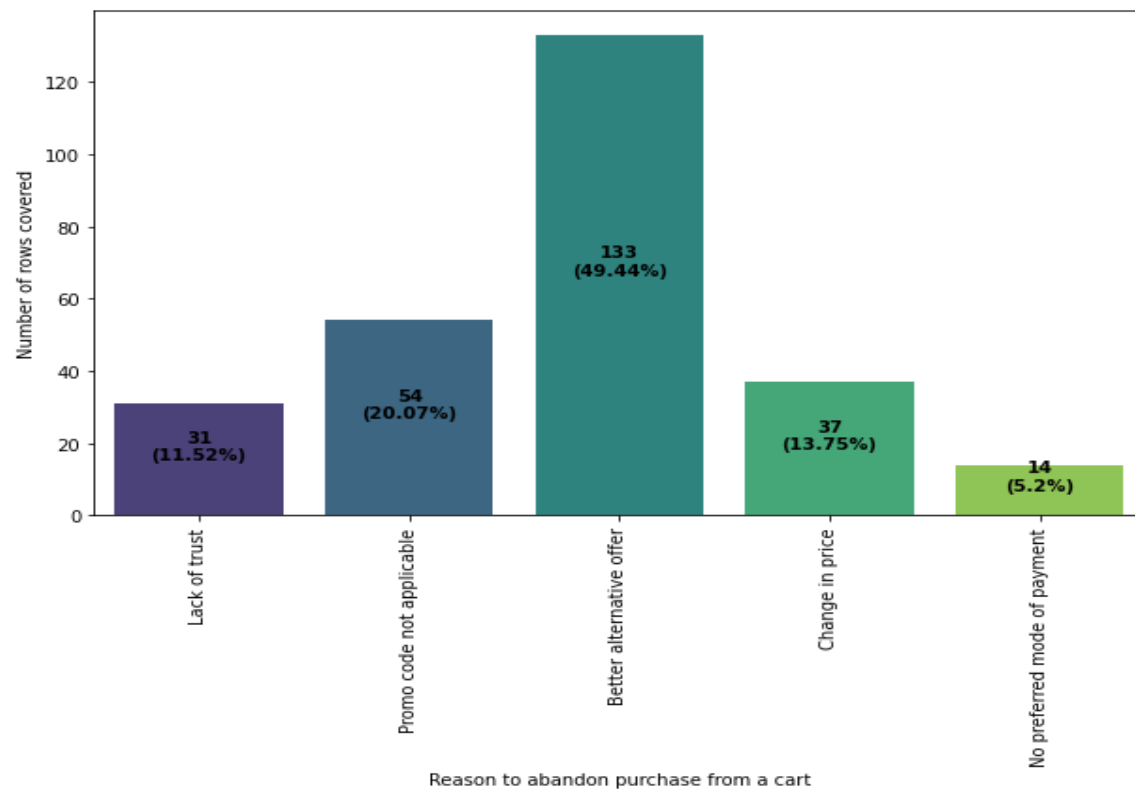




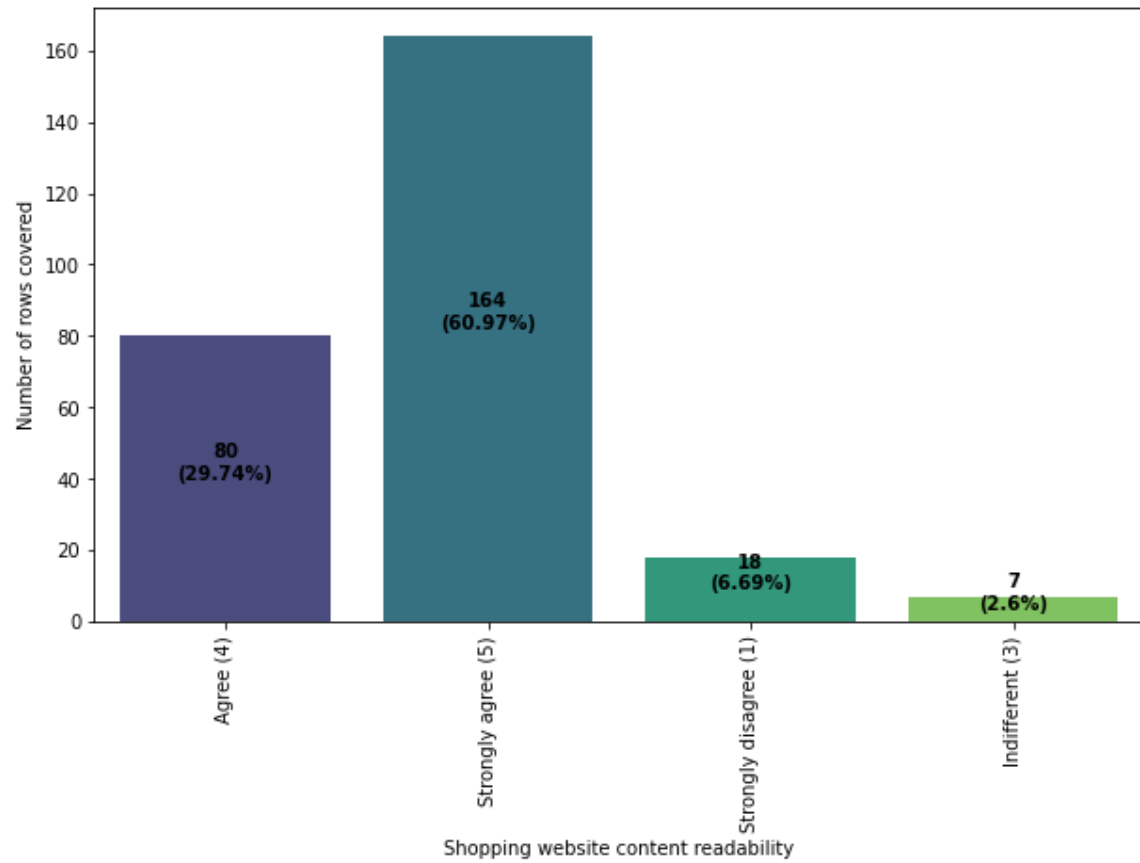
Count Plot for Frequency to abandon purchase from a cart

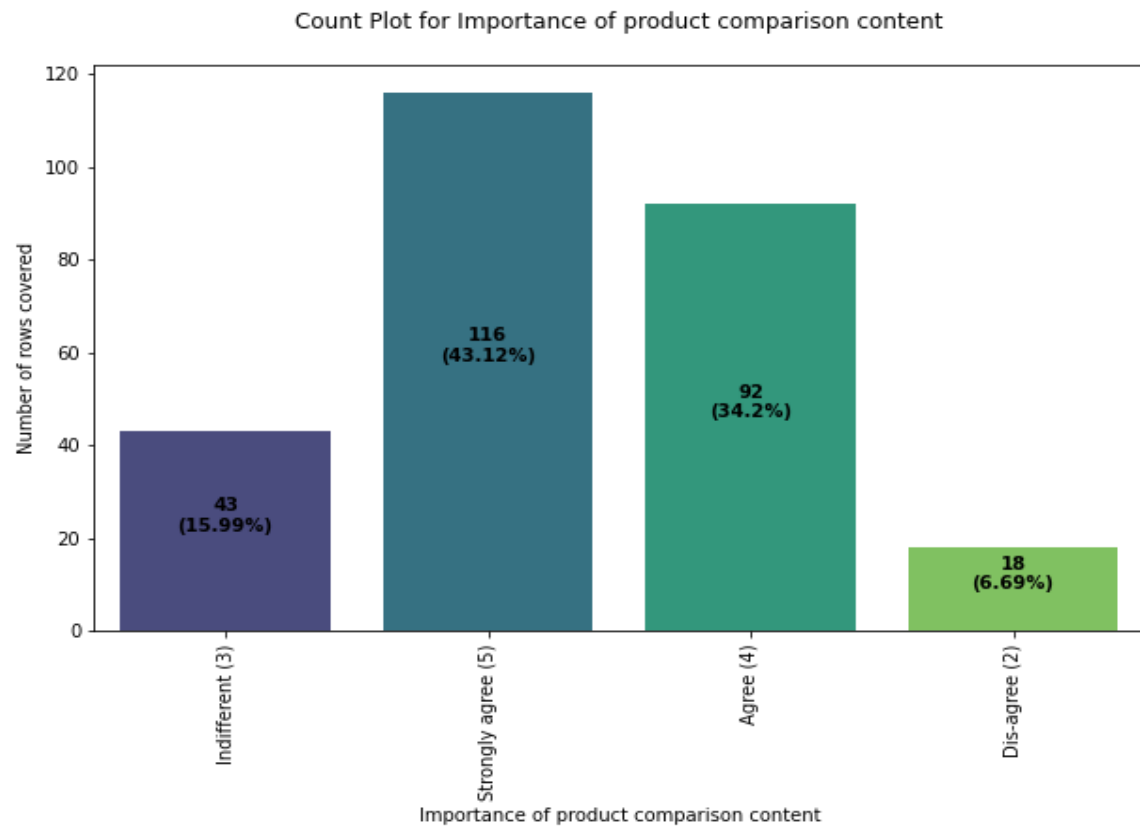


Count Plot for Reason to abandon purchase from a cart

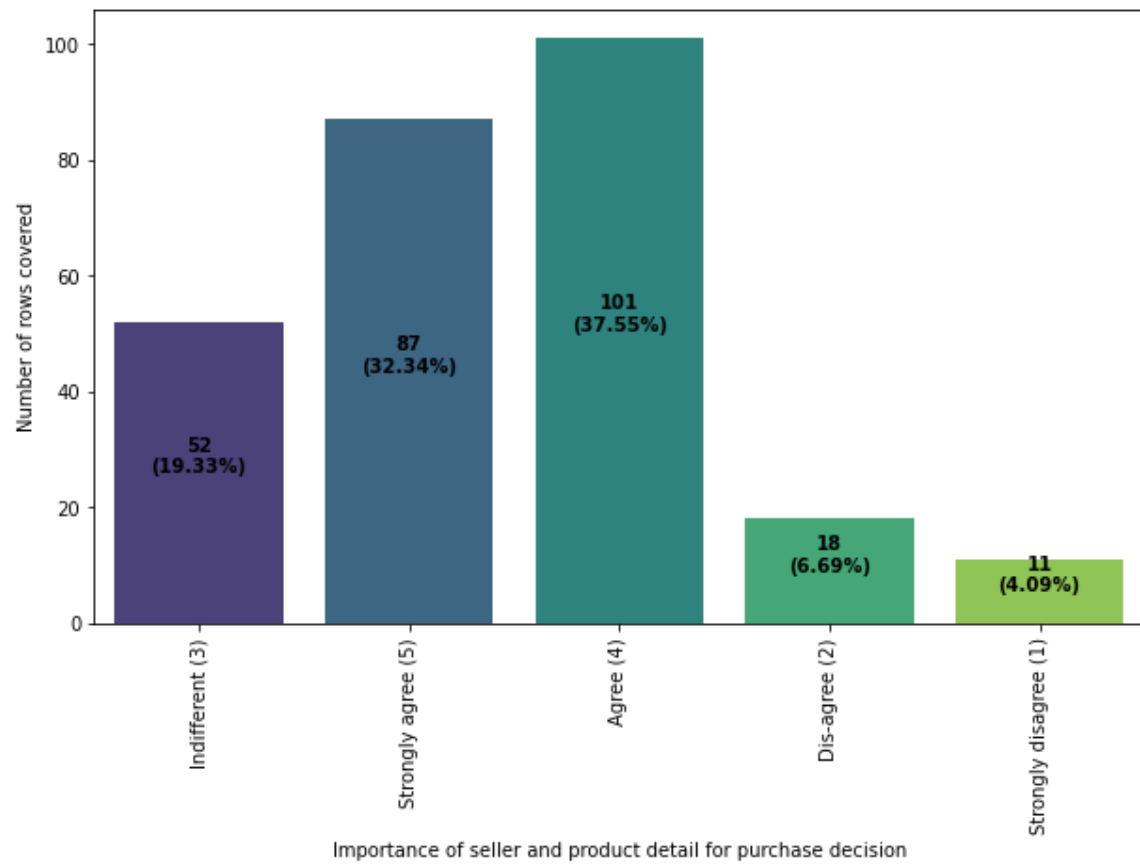


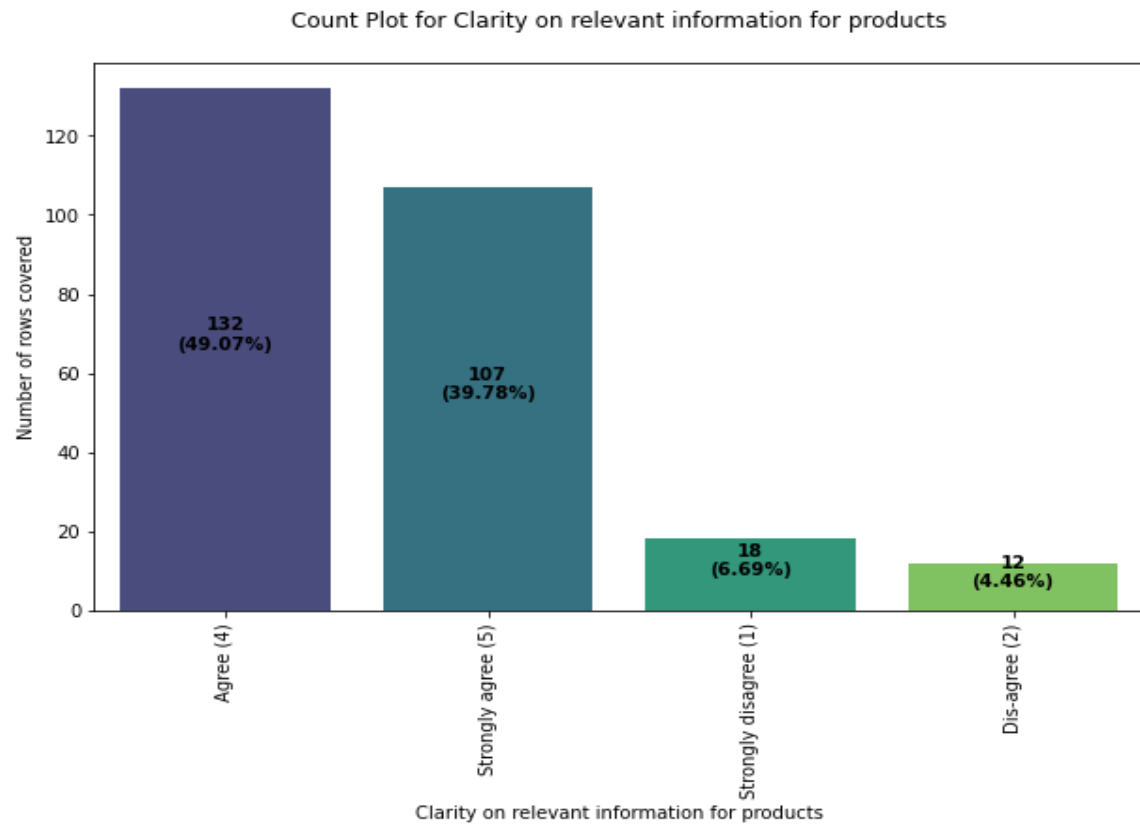
Count Plot for Shopping website content readability



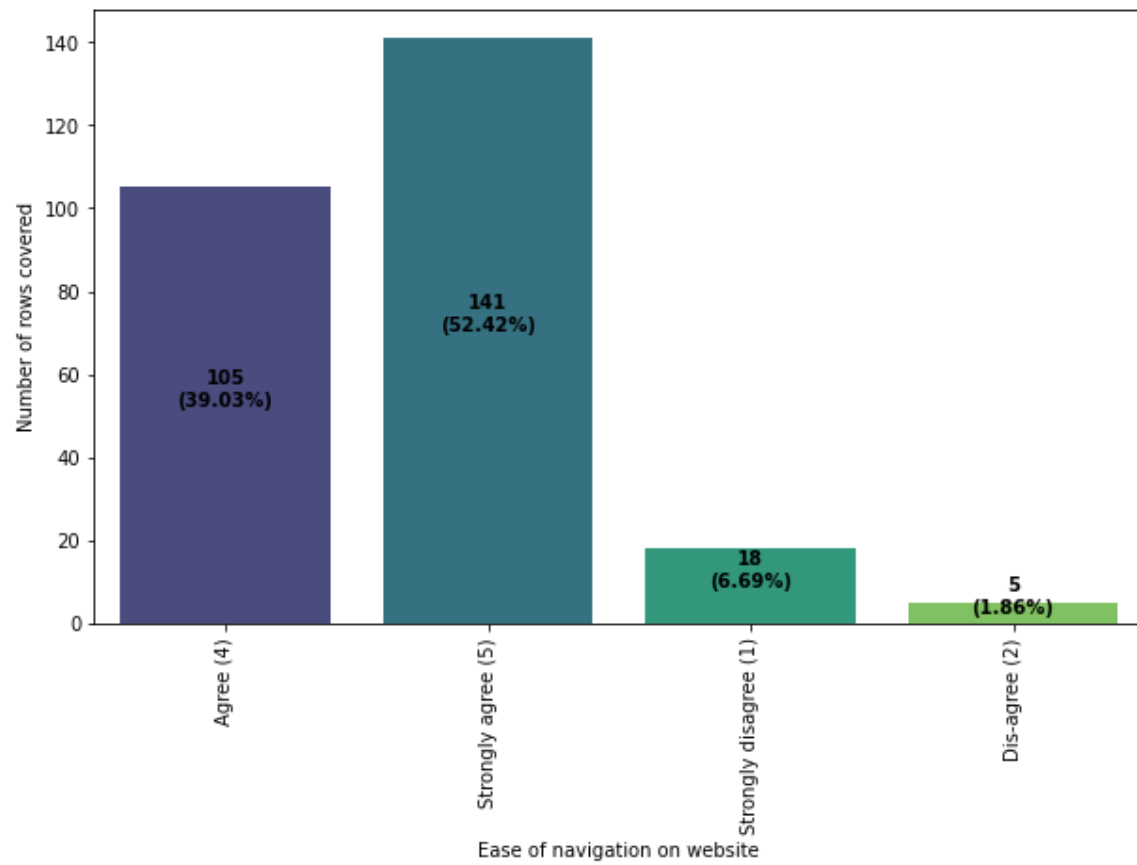


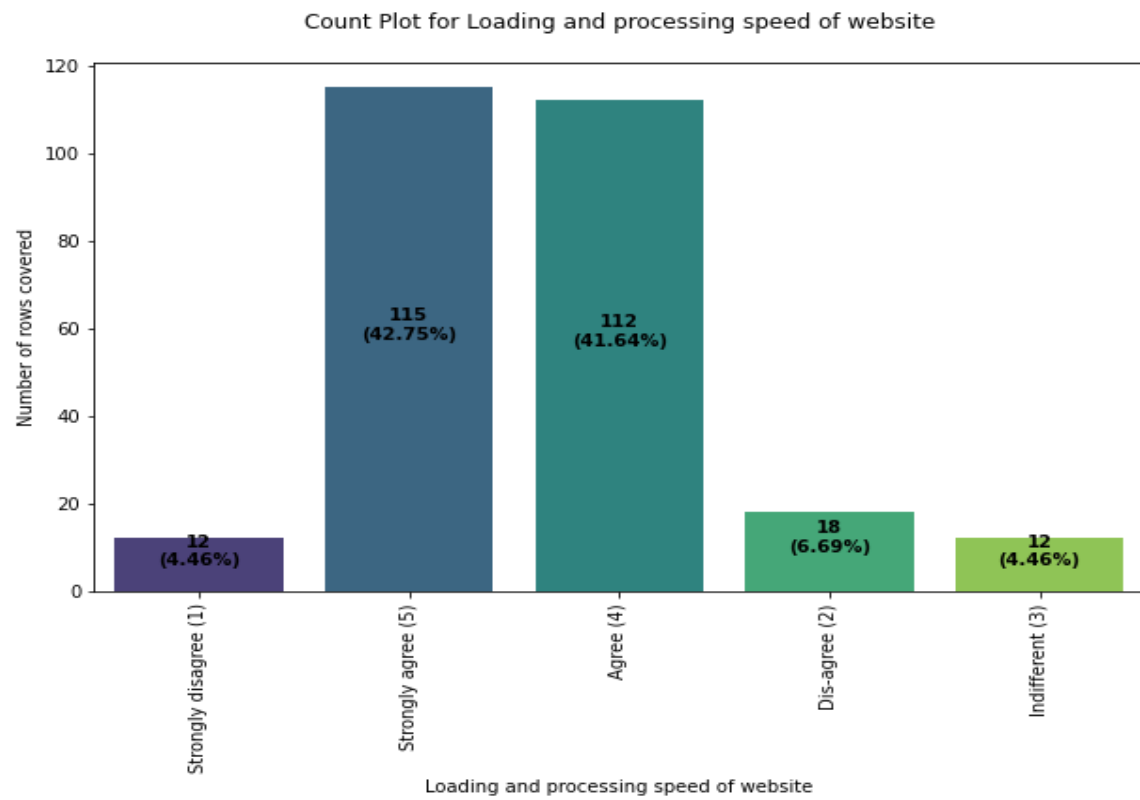
Count Plot for Importance of seller and product detail for purchase decision



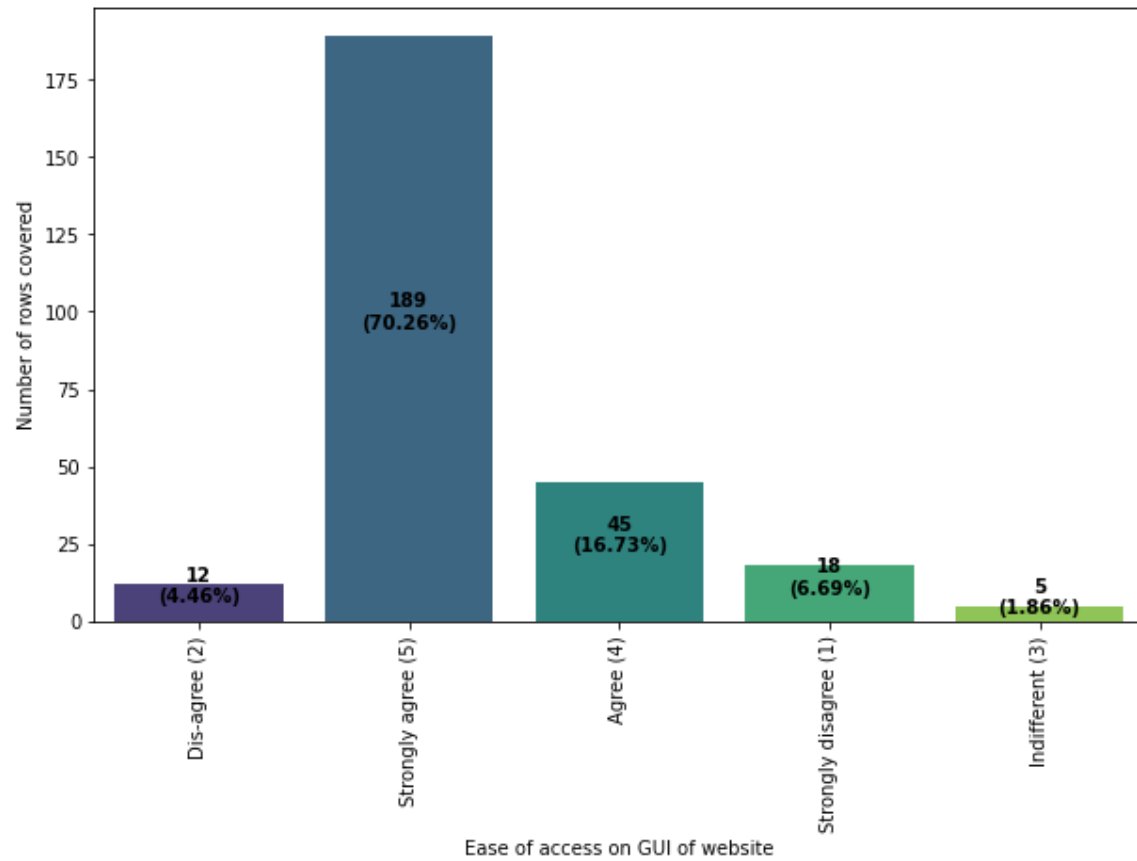


Count Plot for Ease of navigation on website

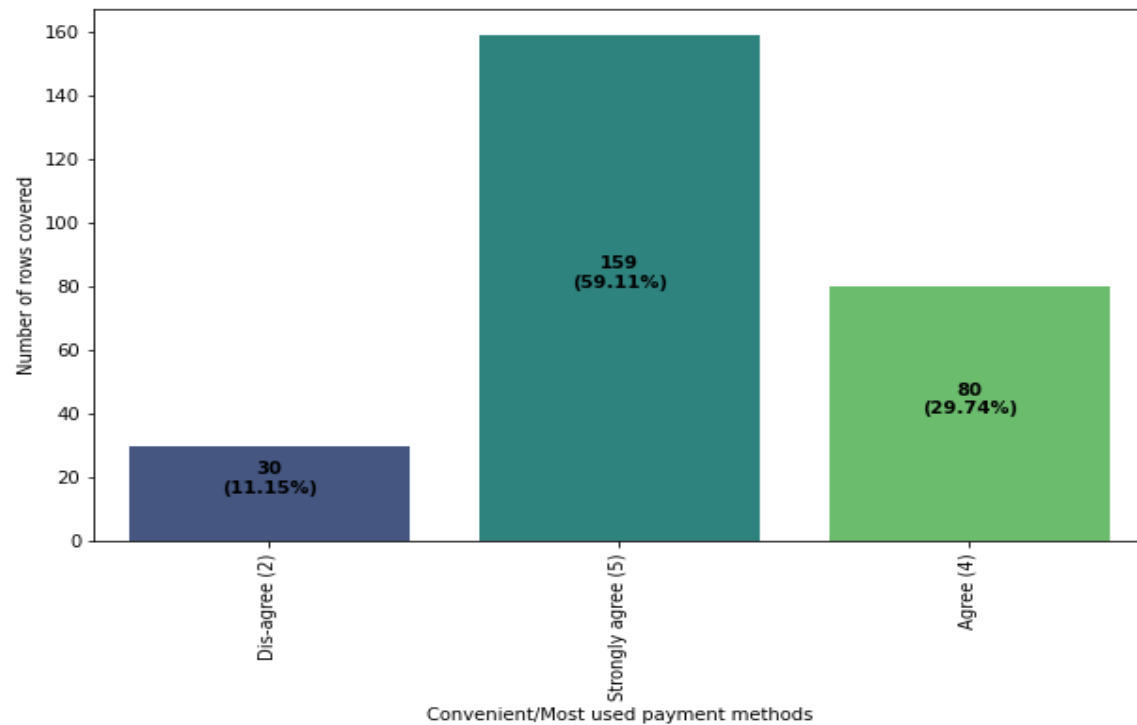




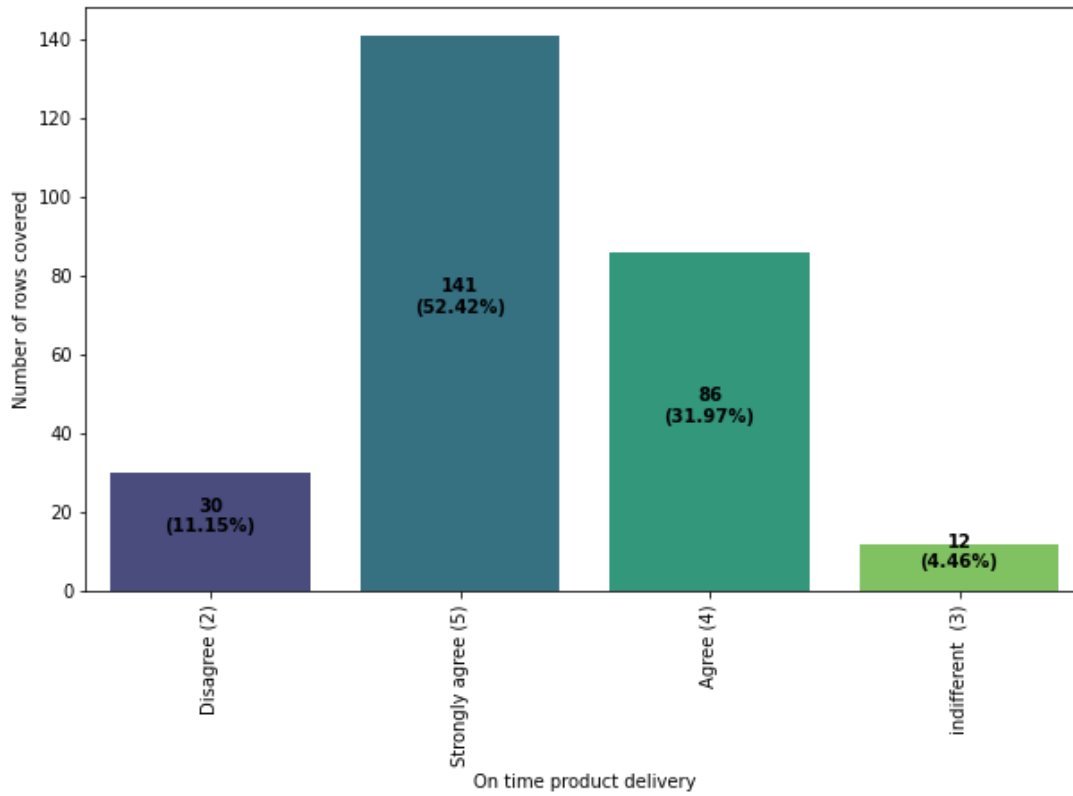
Count Plot for Ease of access on GUI of website



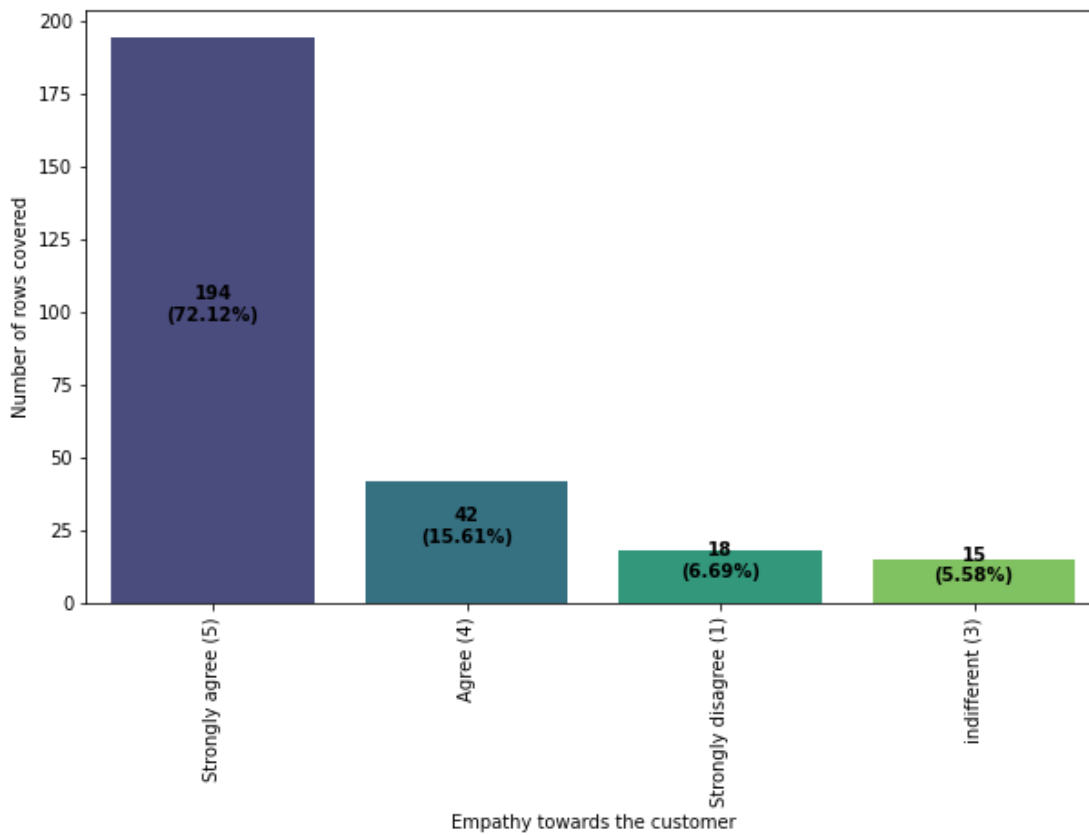
Count Plot for Convenient/Most used payment methods



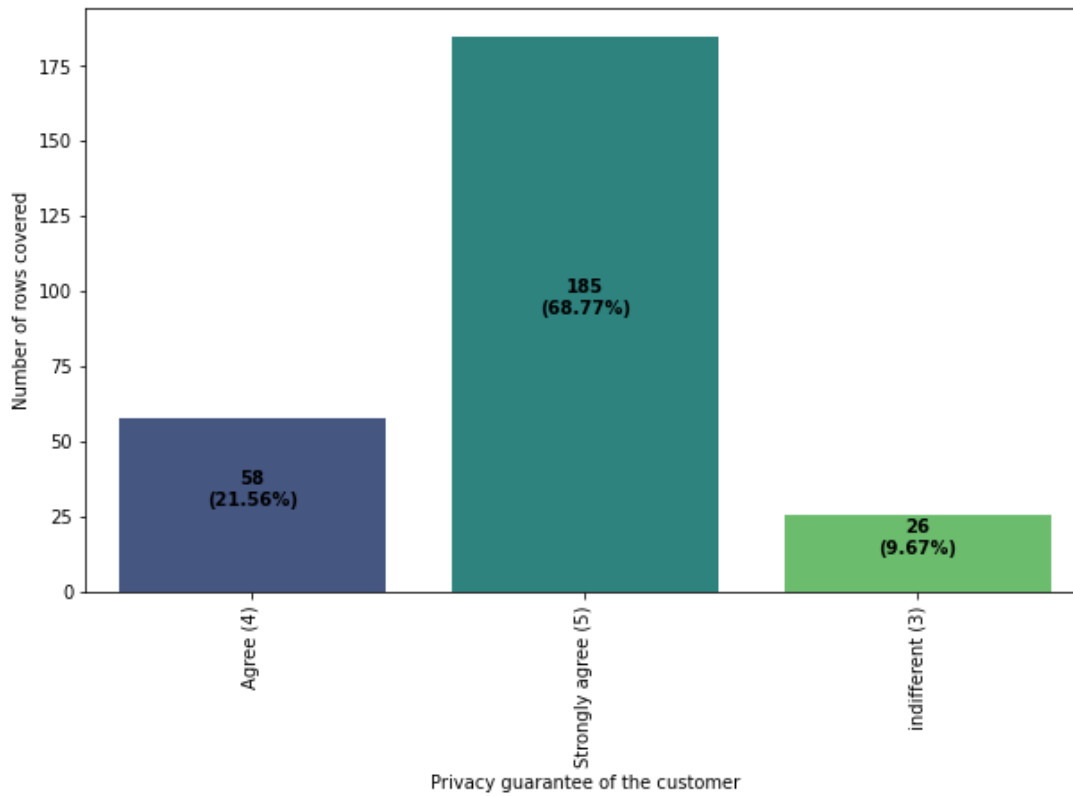
Count Plot for On time product delivery



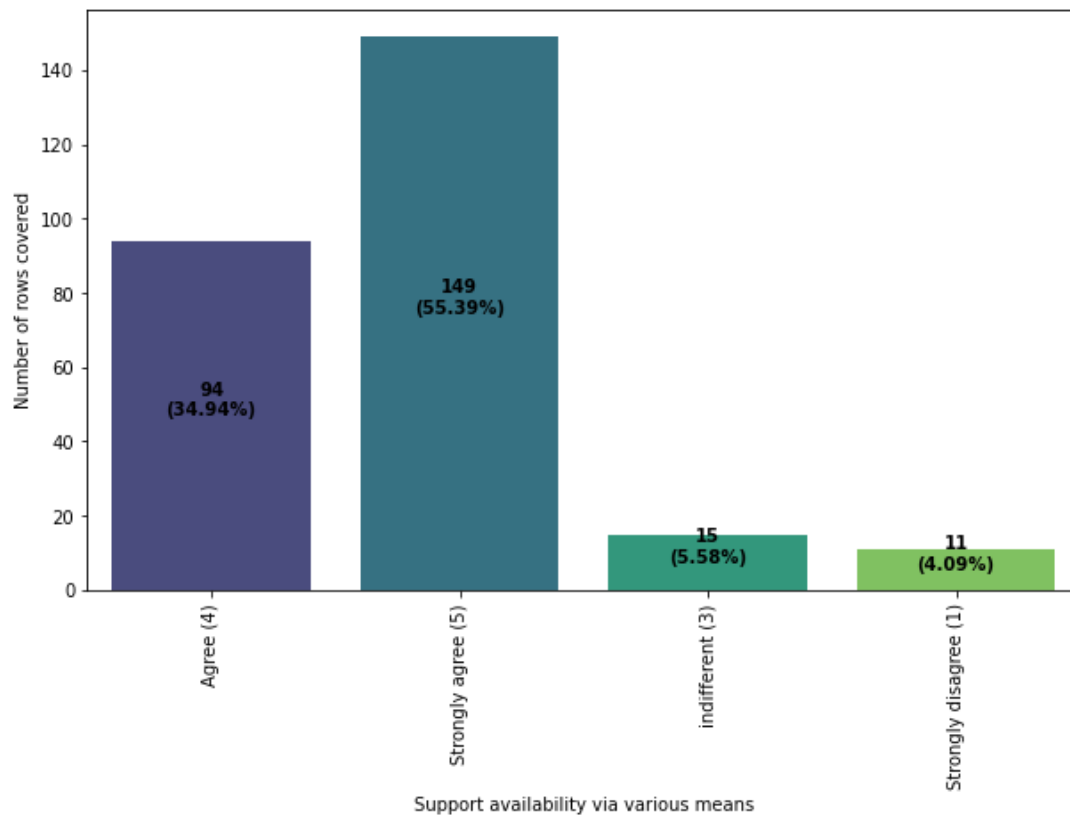
Count Plot for Empathy towards the customer



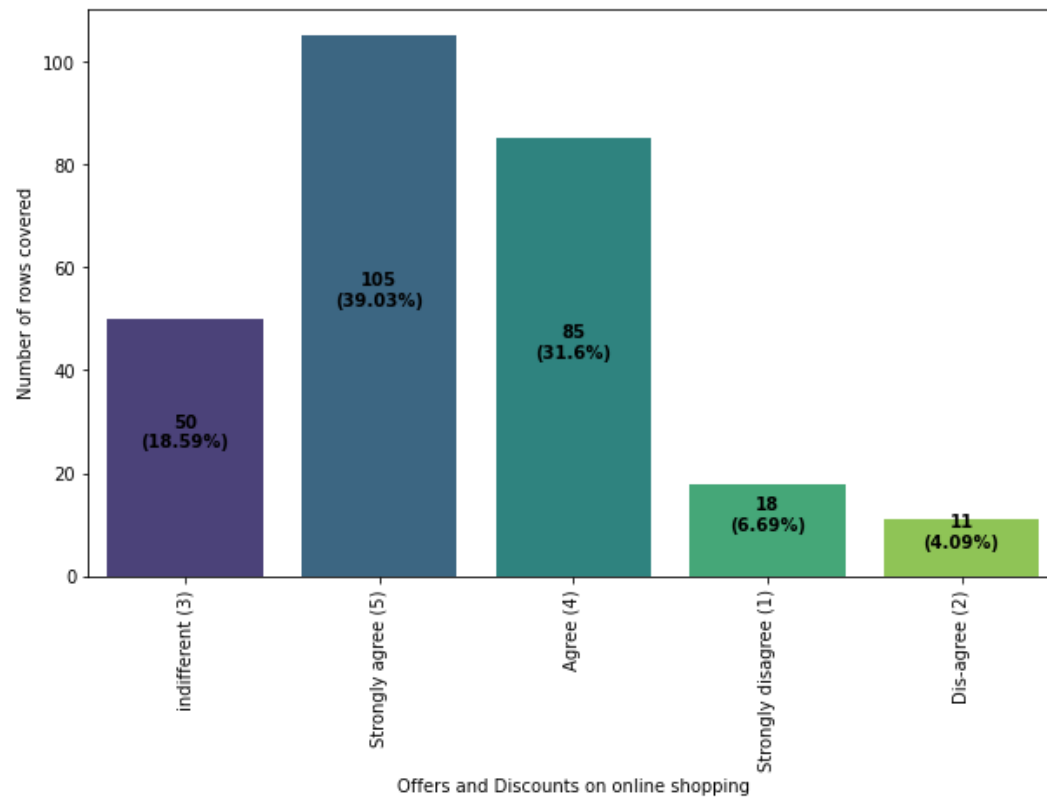
Count Plot for Privacy guarantee of the customer



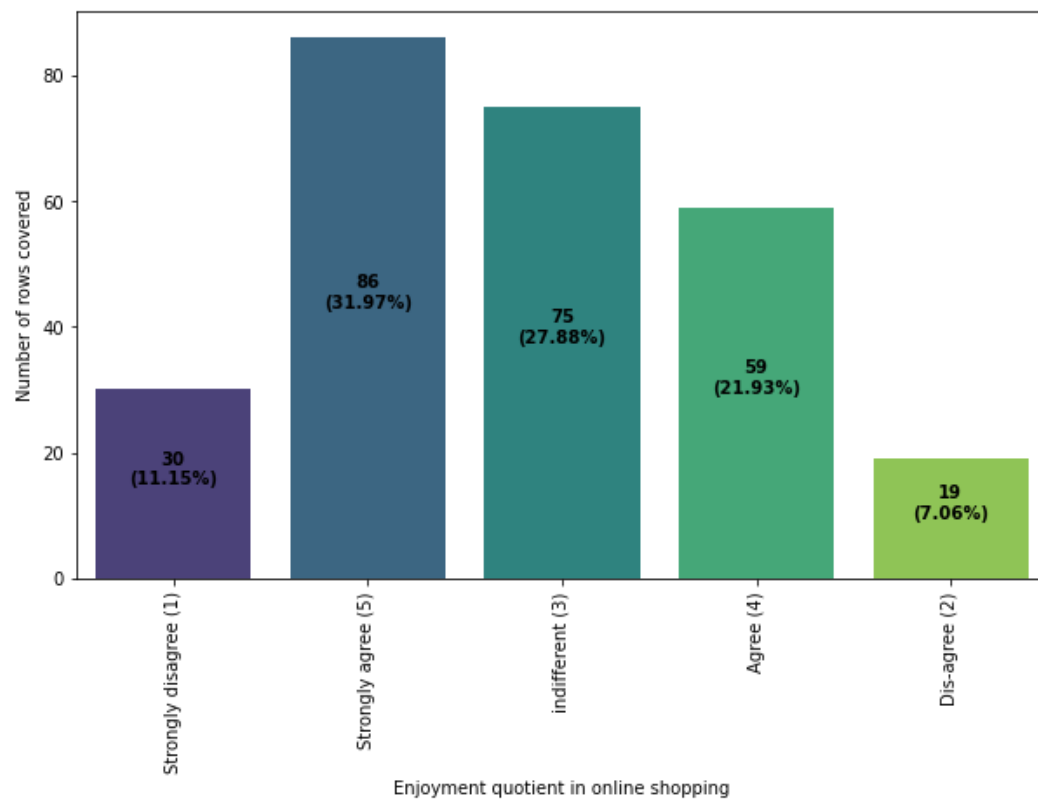
Count Plot for Support availability via various means



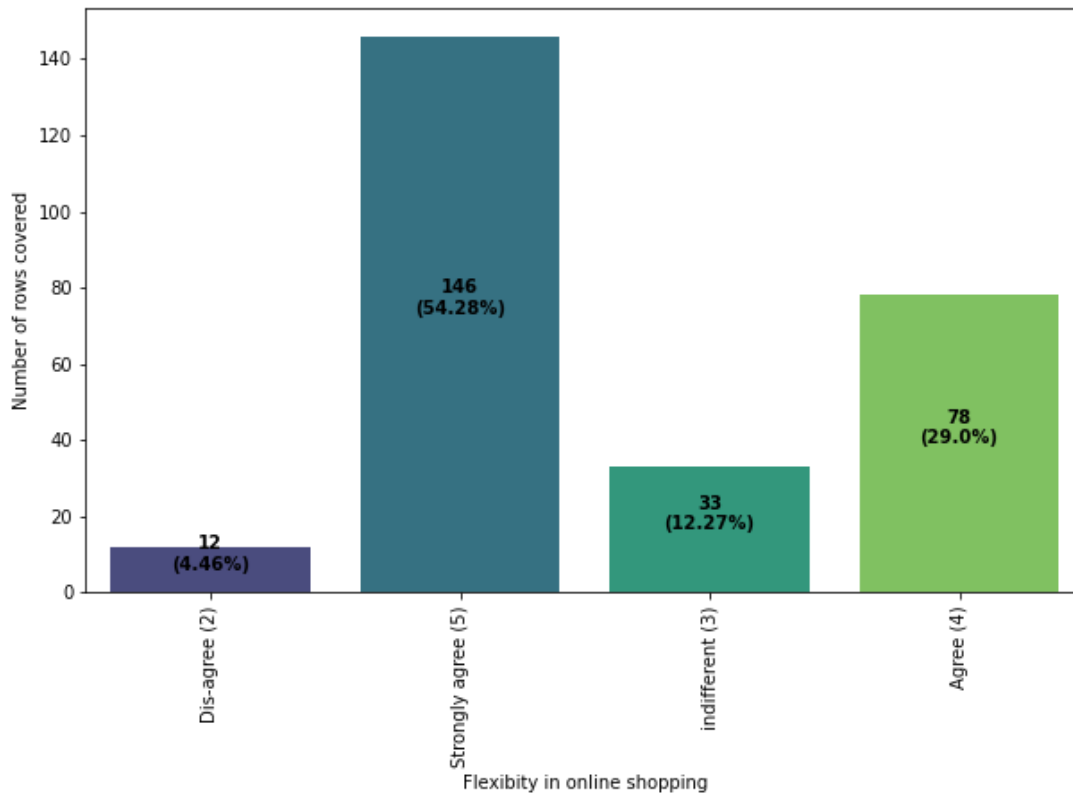
Count Plot for Offers and Discounts on online shopping



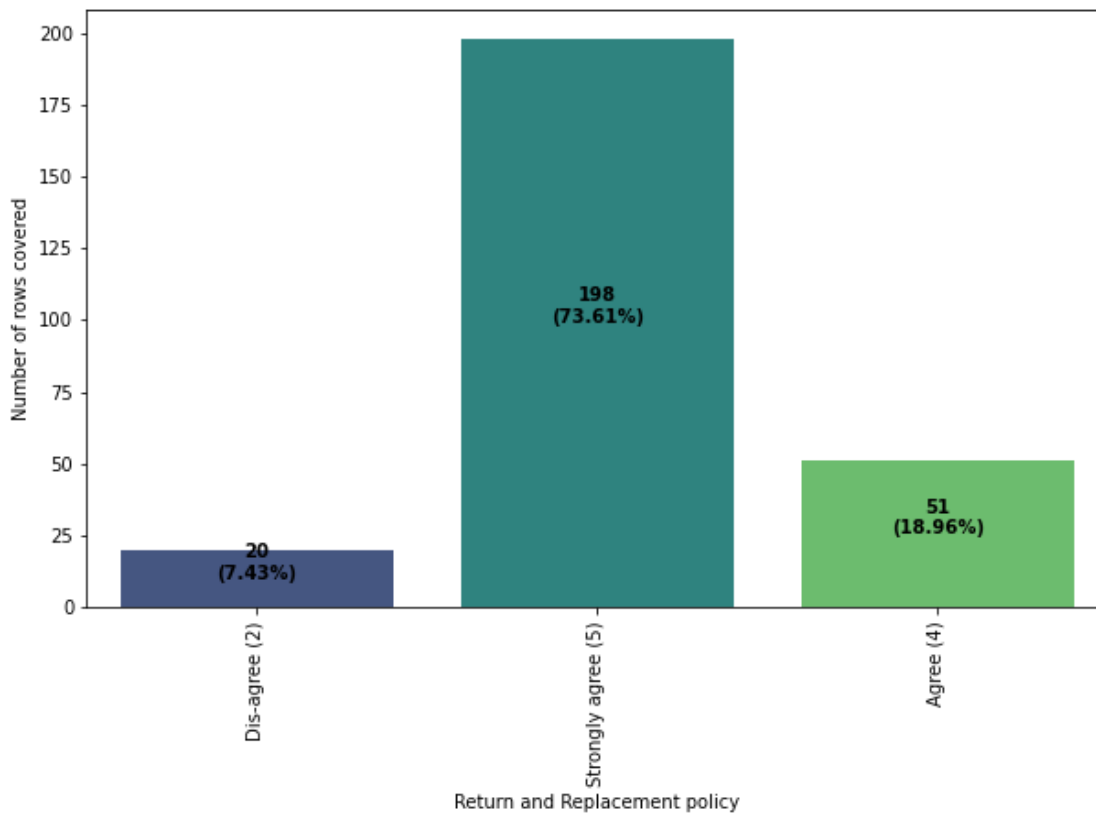
Count Plot for Enjoyment quotient in online shopping

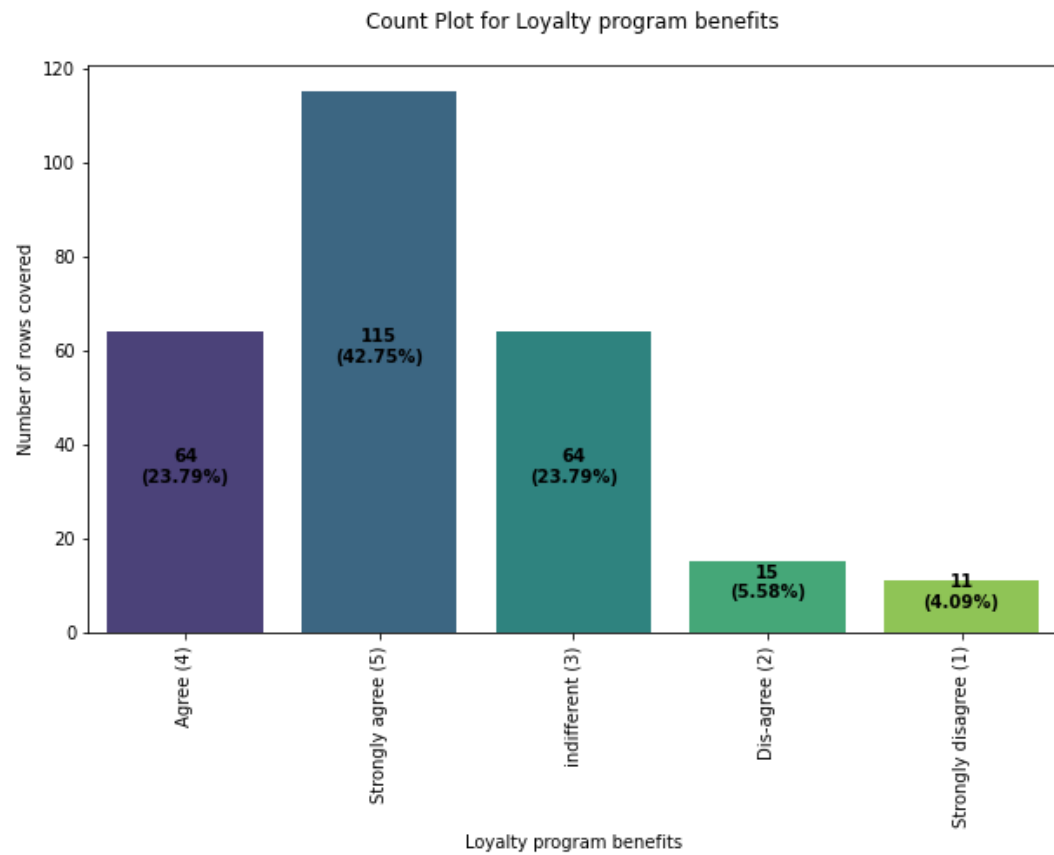


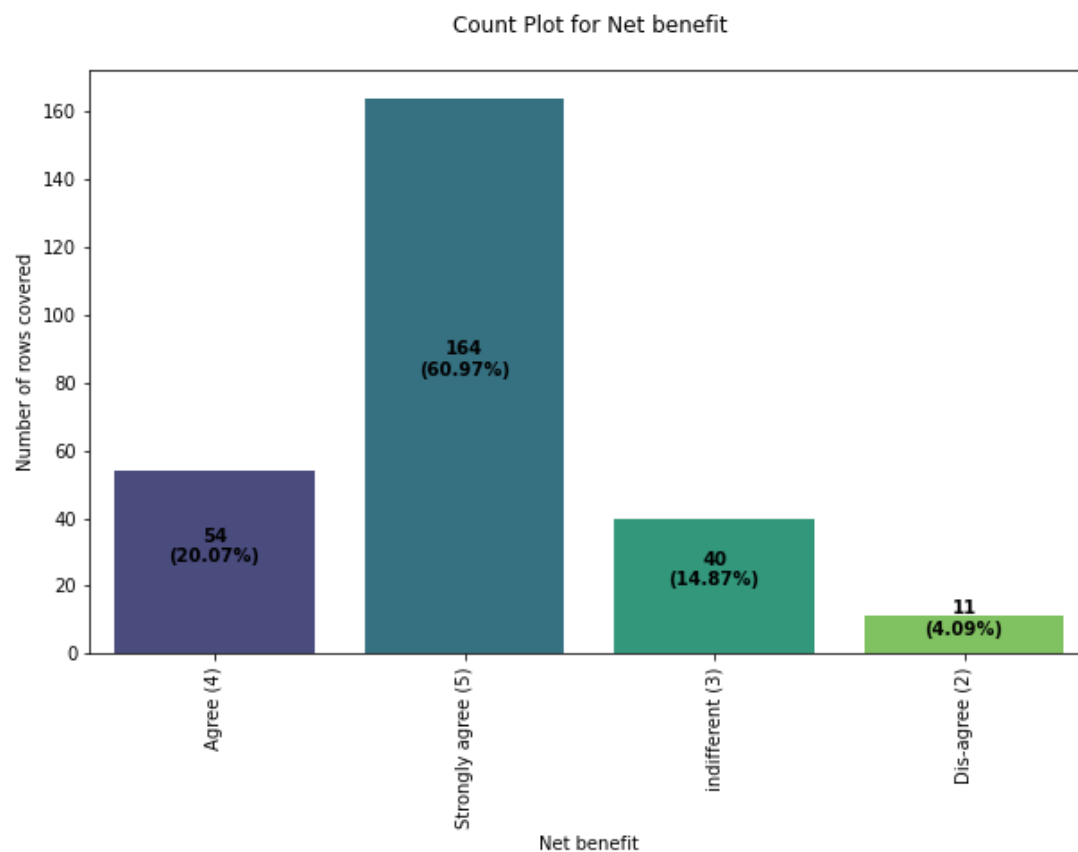
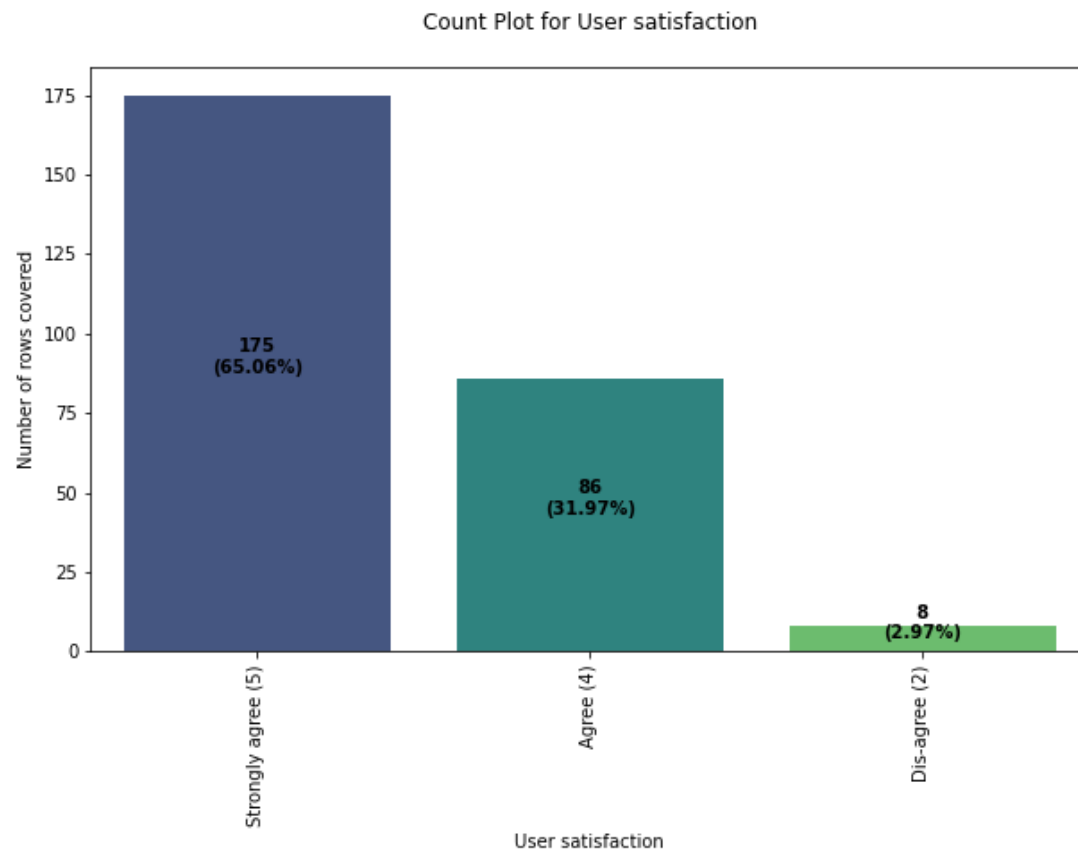
Count Plot for Flexibility in online shopping



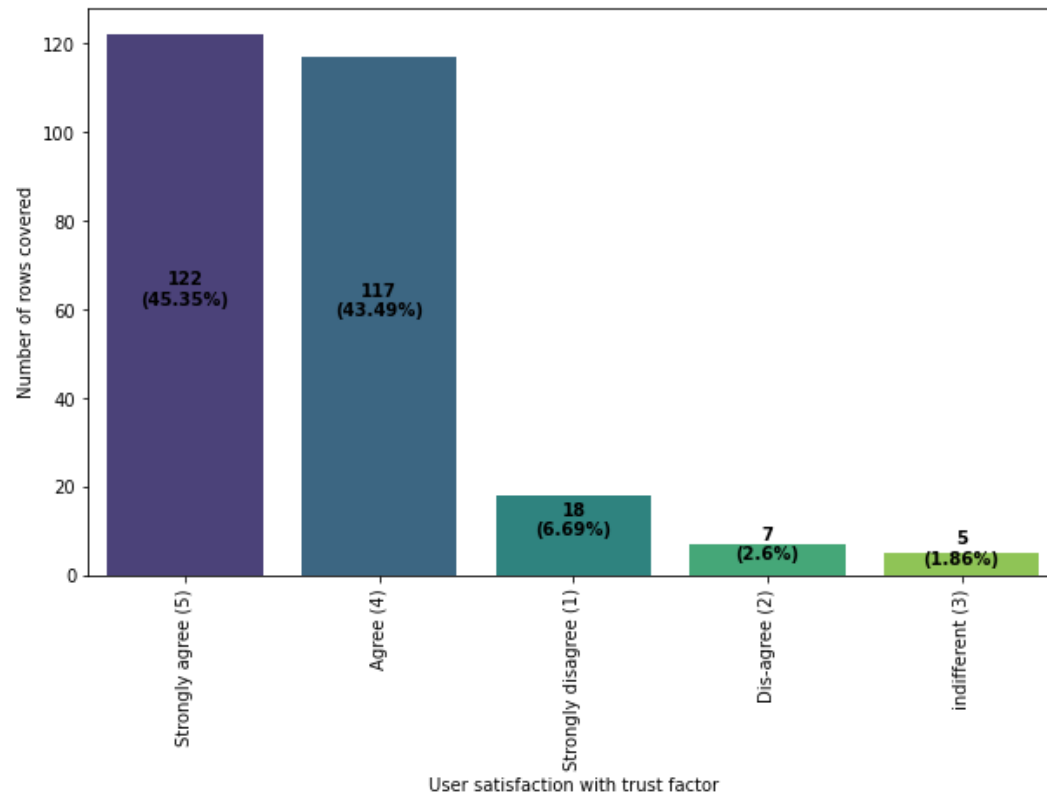
Count Plot for Return and Replacement policy



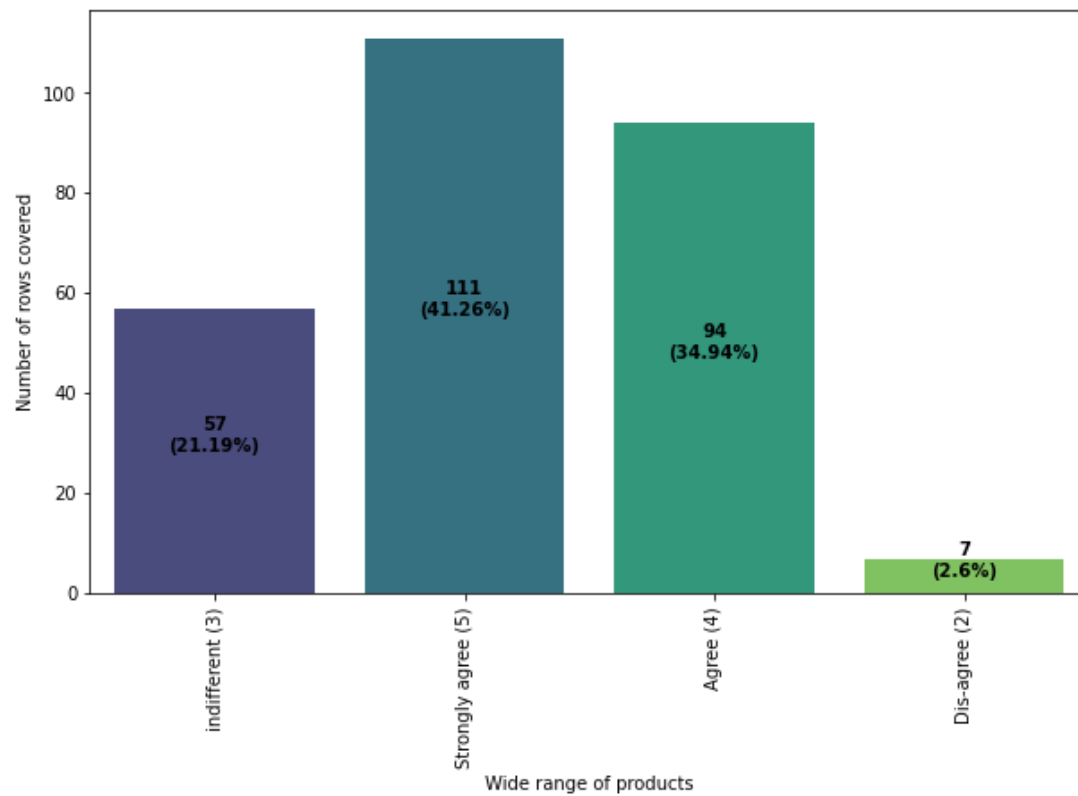


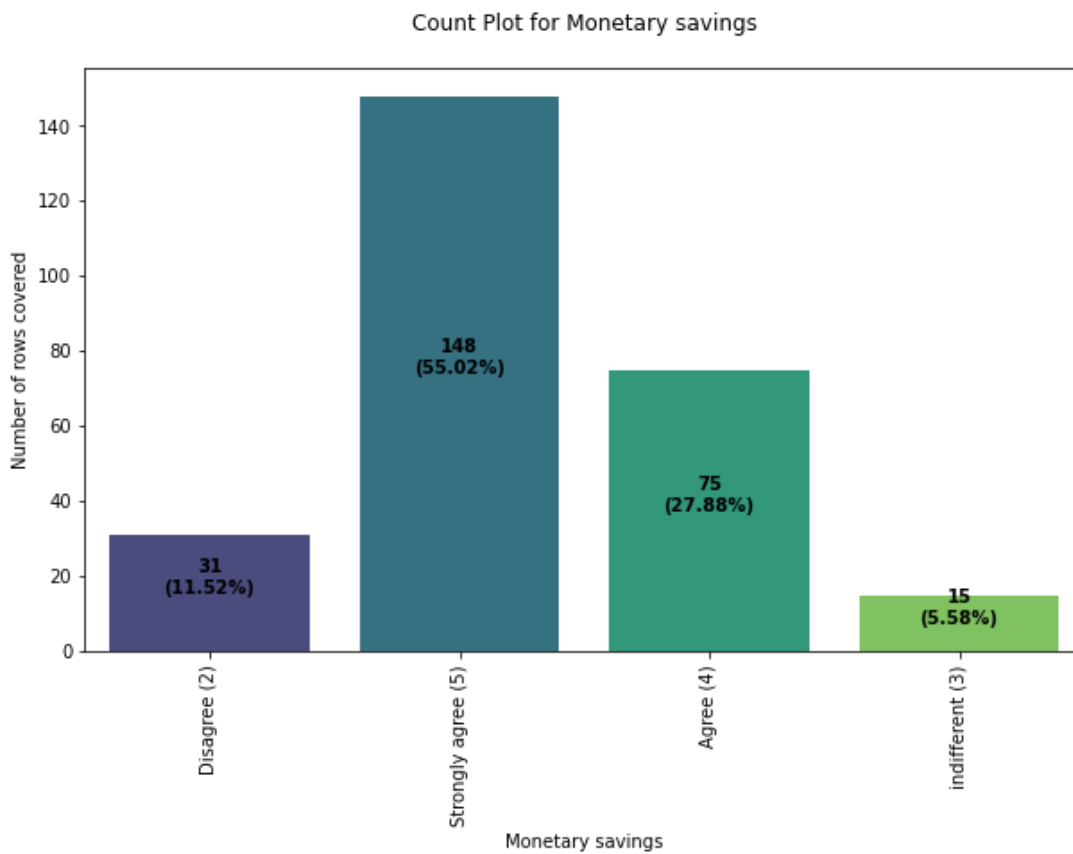
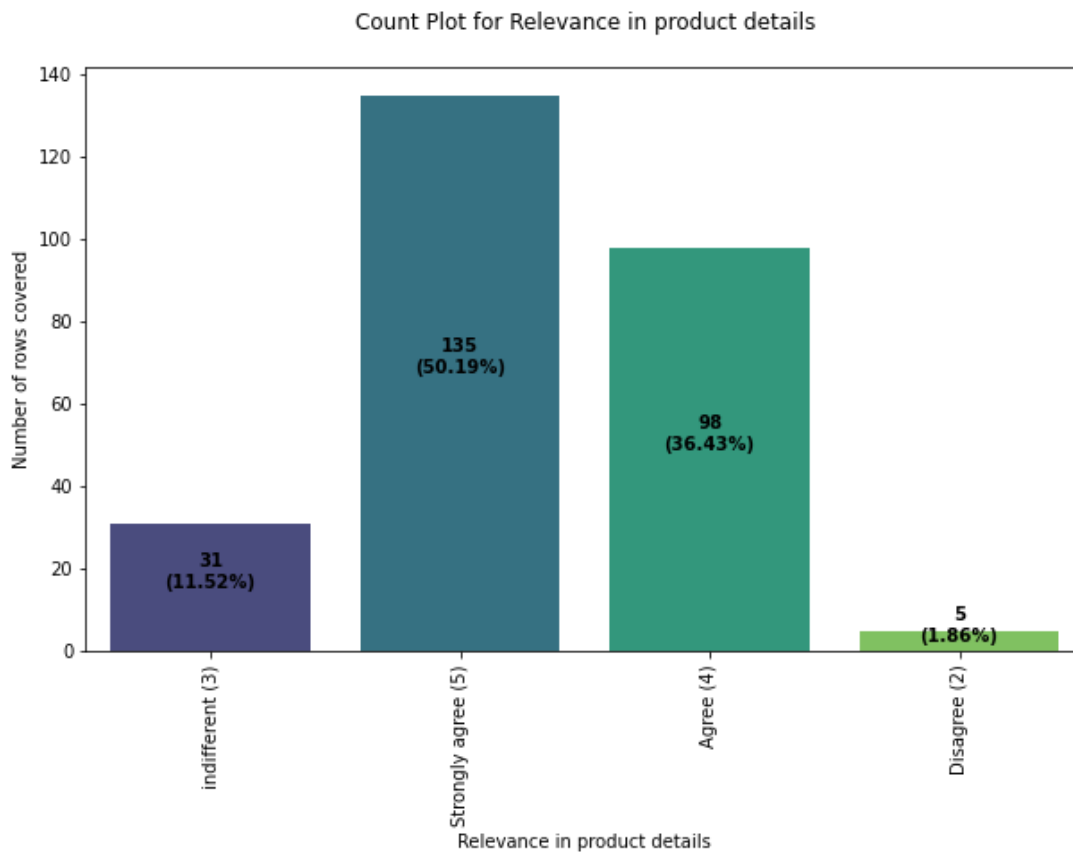


Count Plot for User satisfaction with trust factor

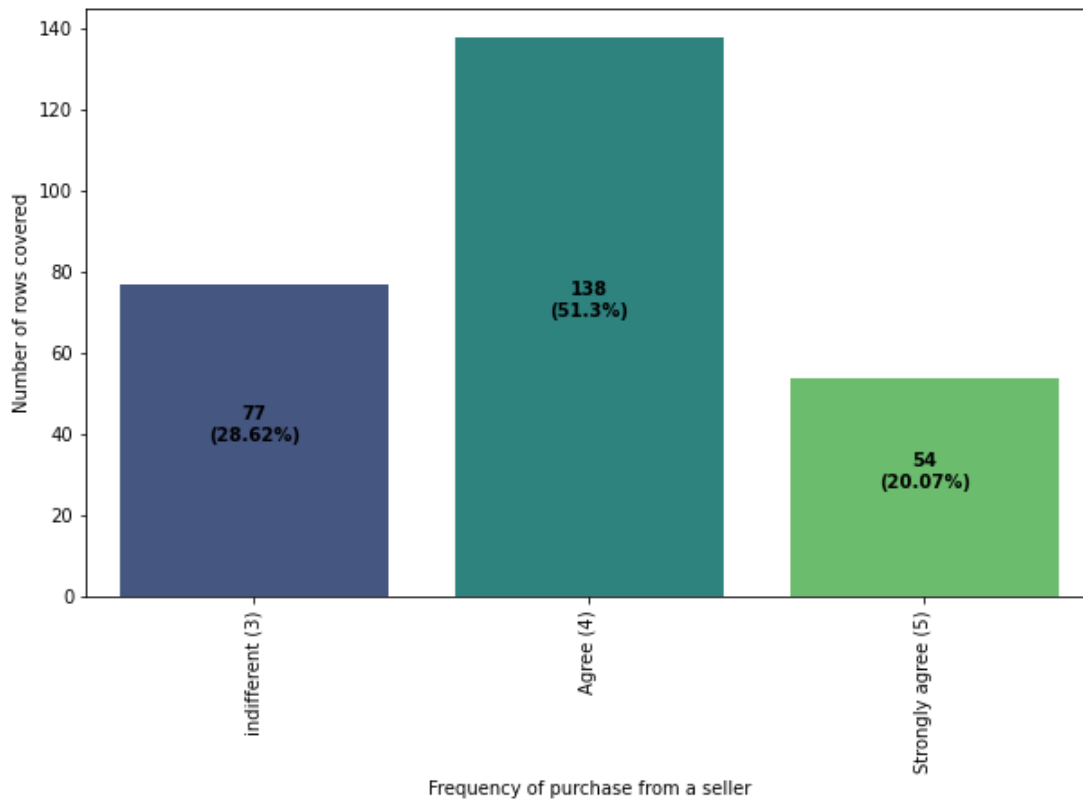


Count Plot for Wide range of products

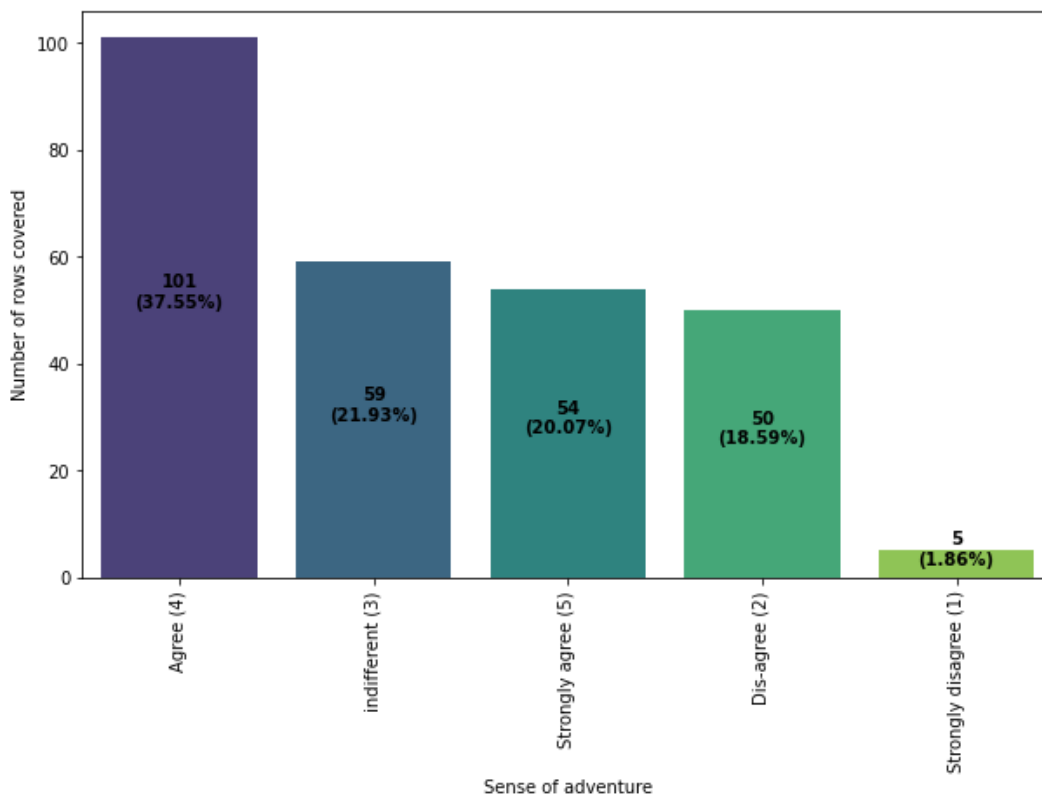


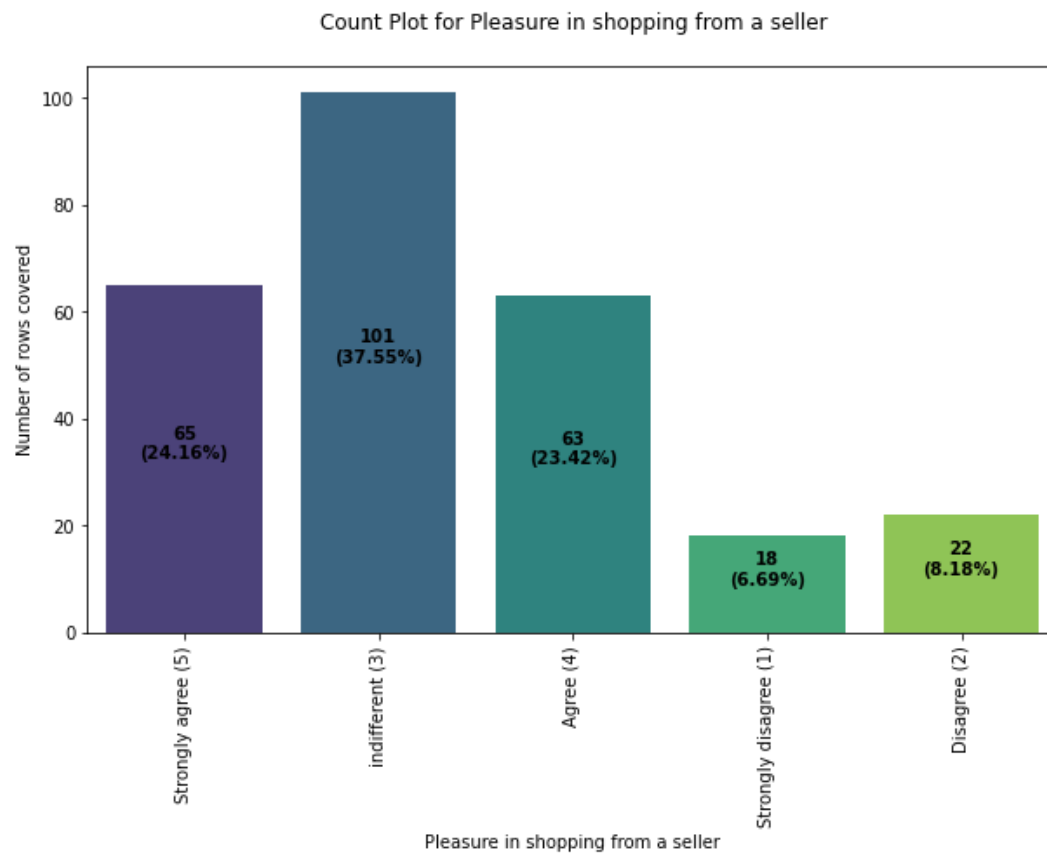
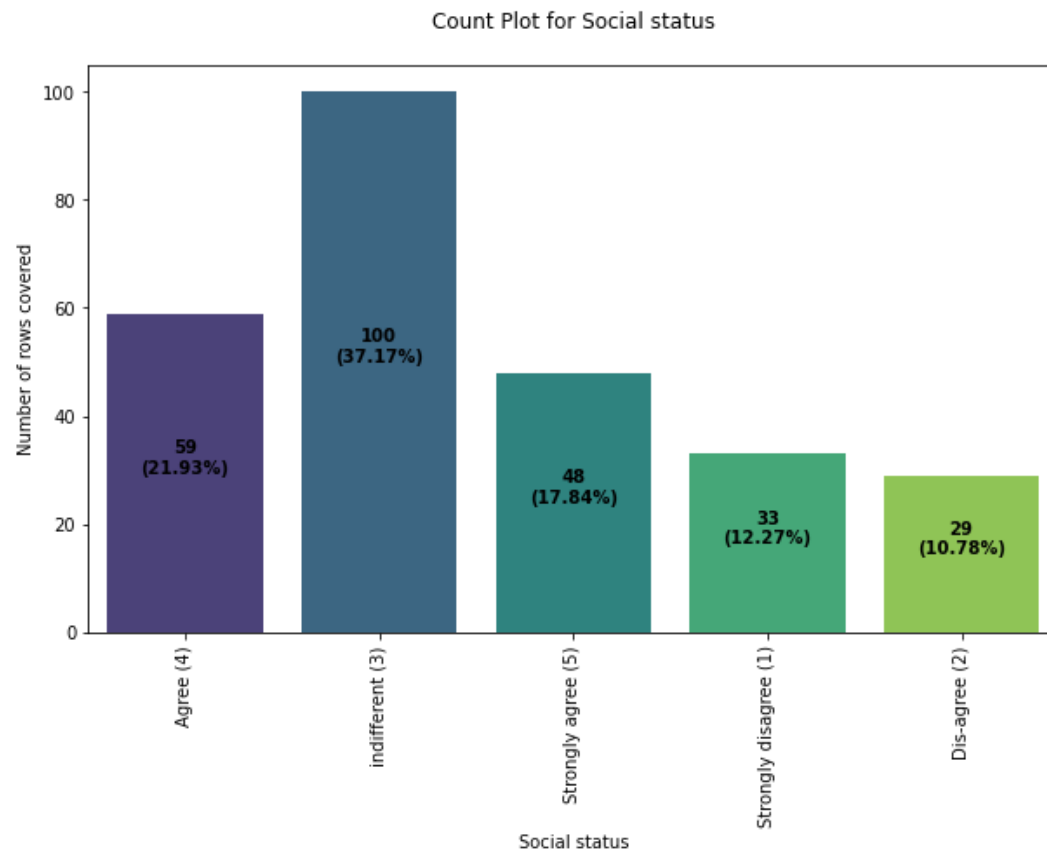


Count Plot for Frequency of purchase from a seller

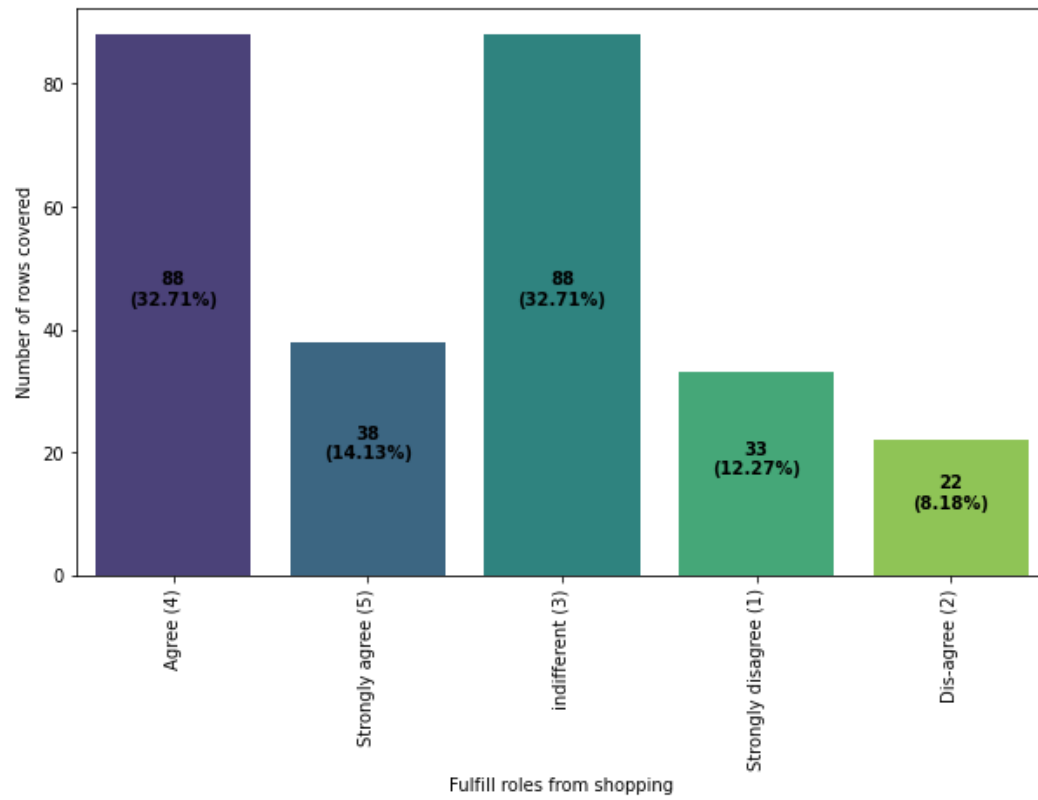


Count Plot for Sense of adventure

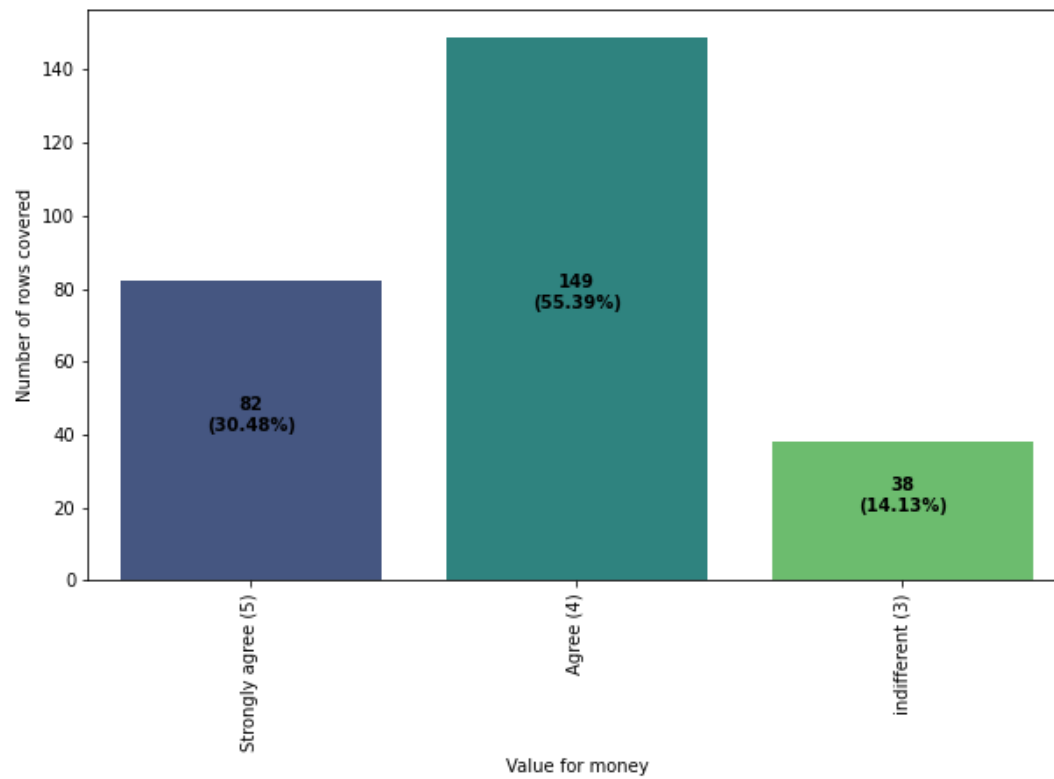


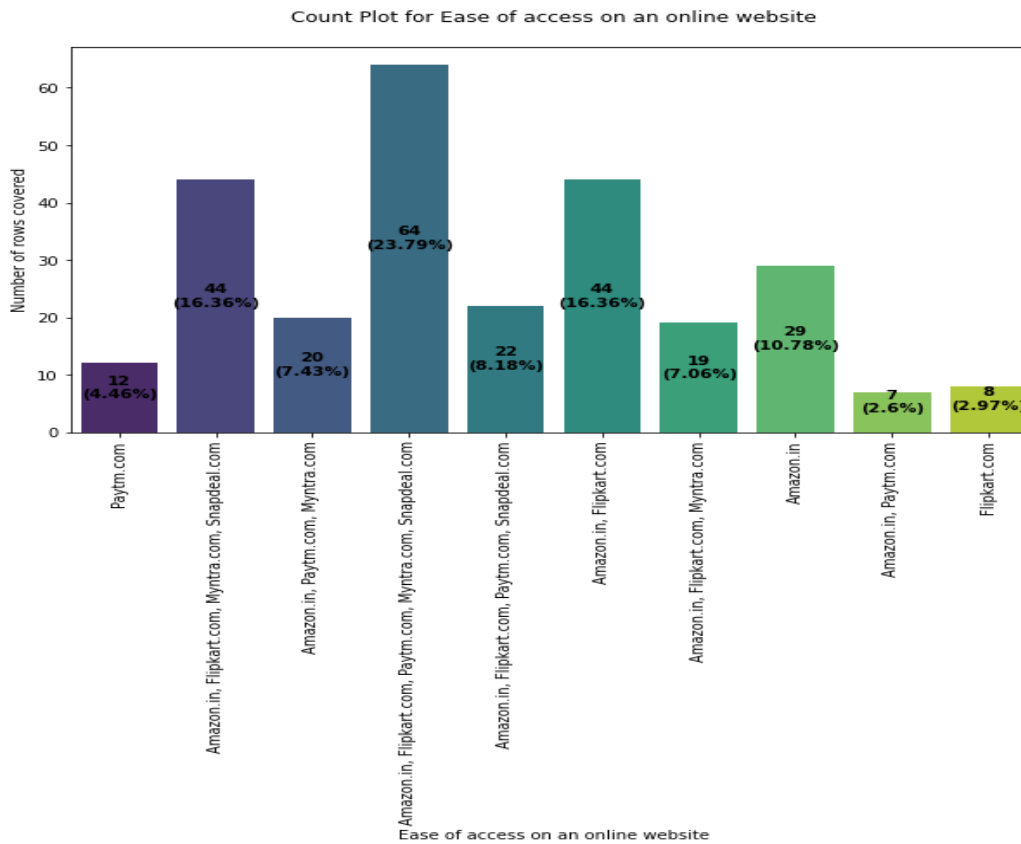
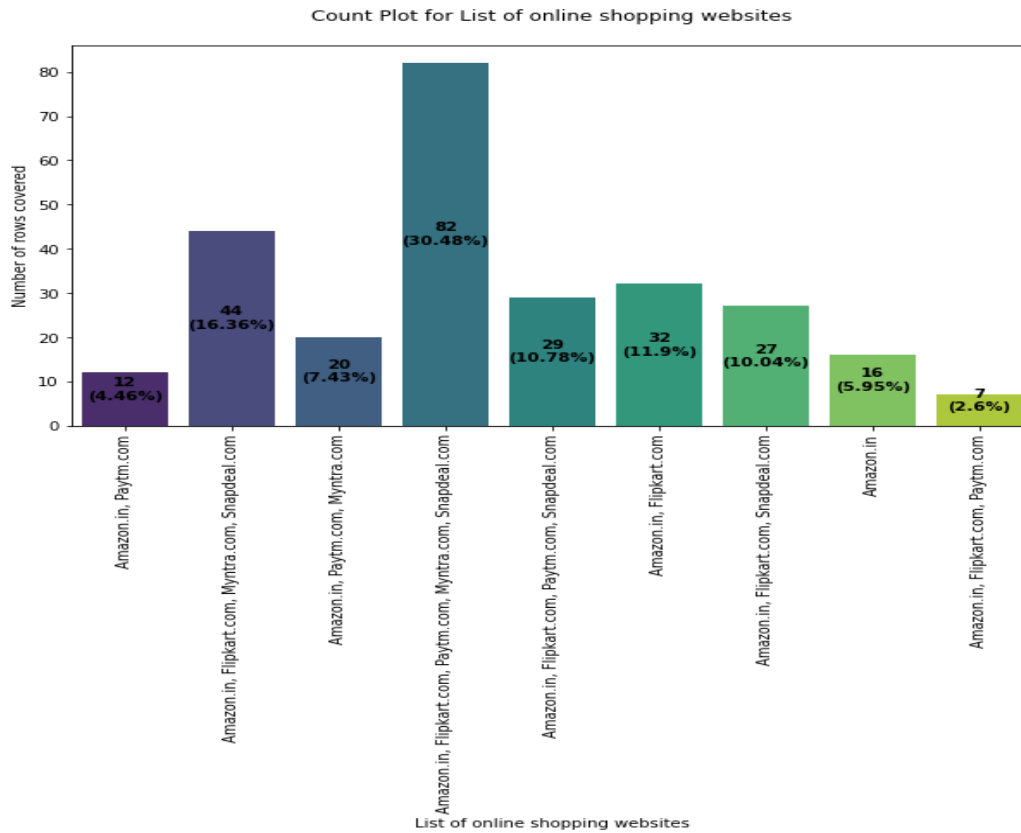


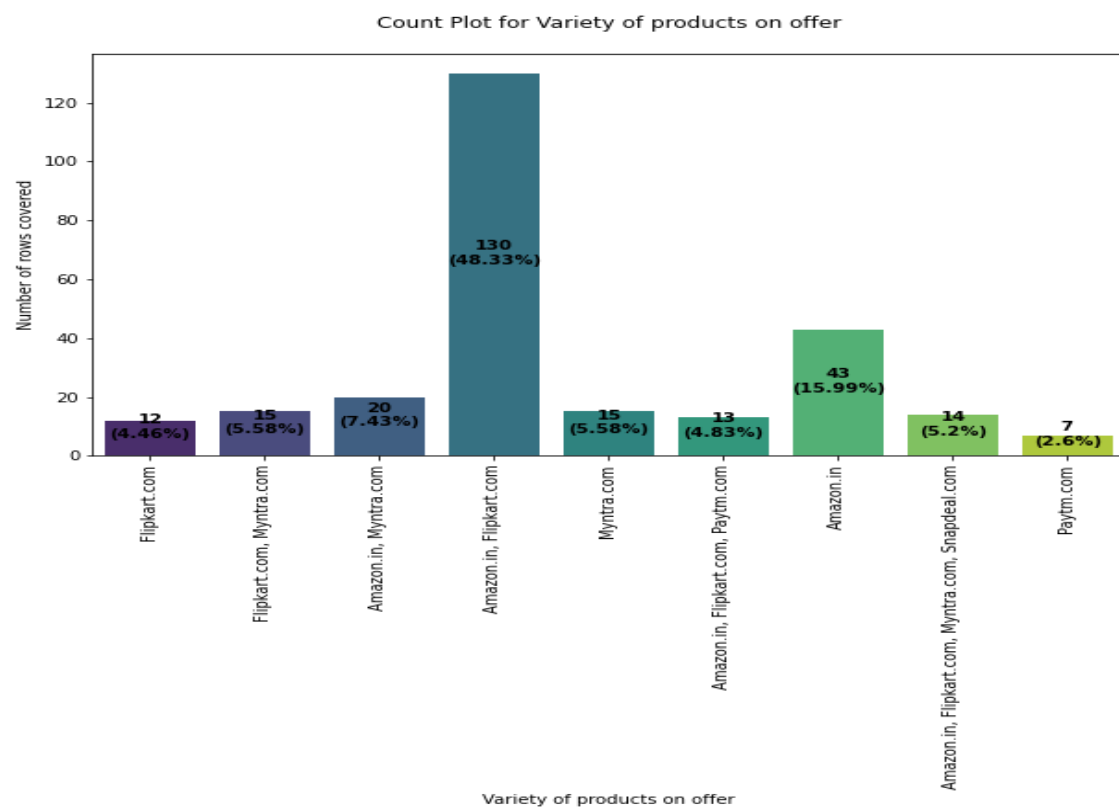
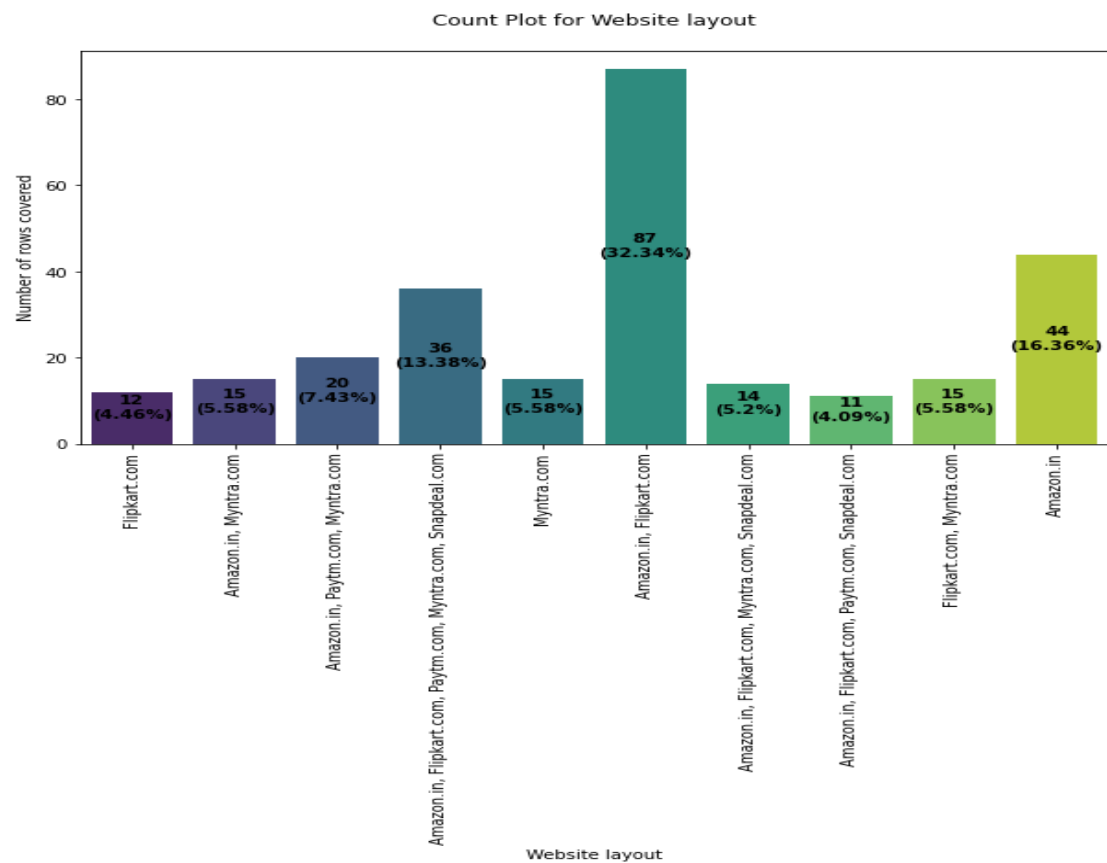
Count Plot for Fulfill roles from shopping

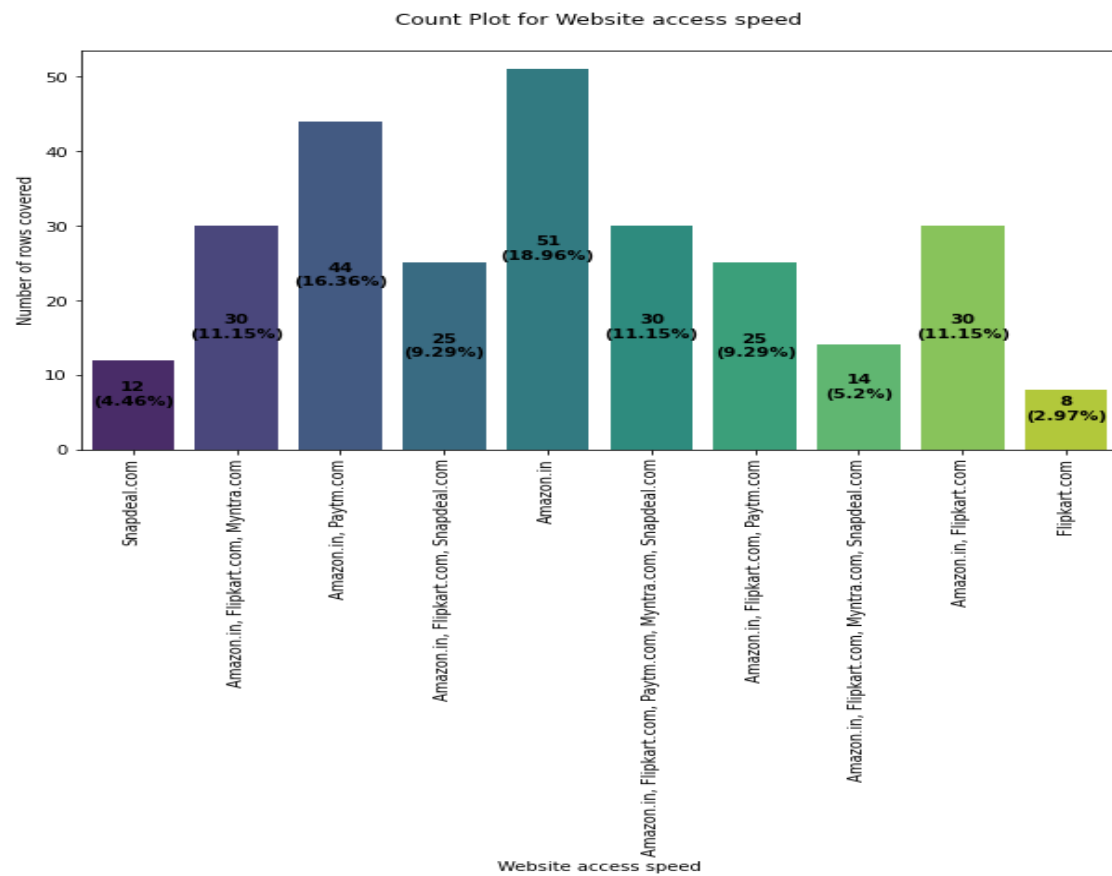
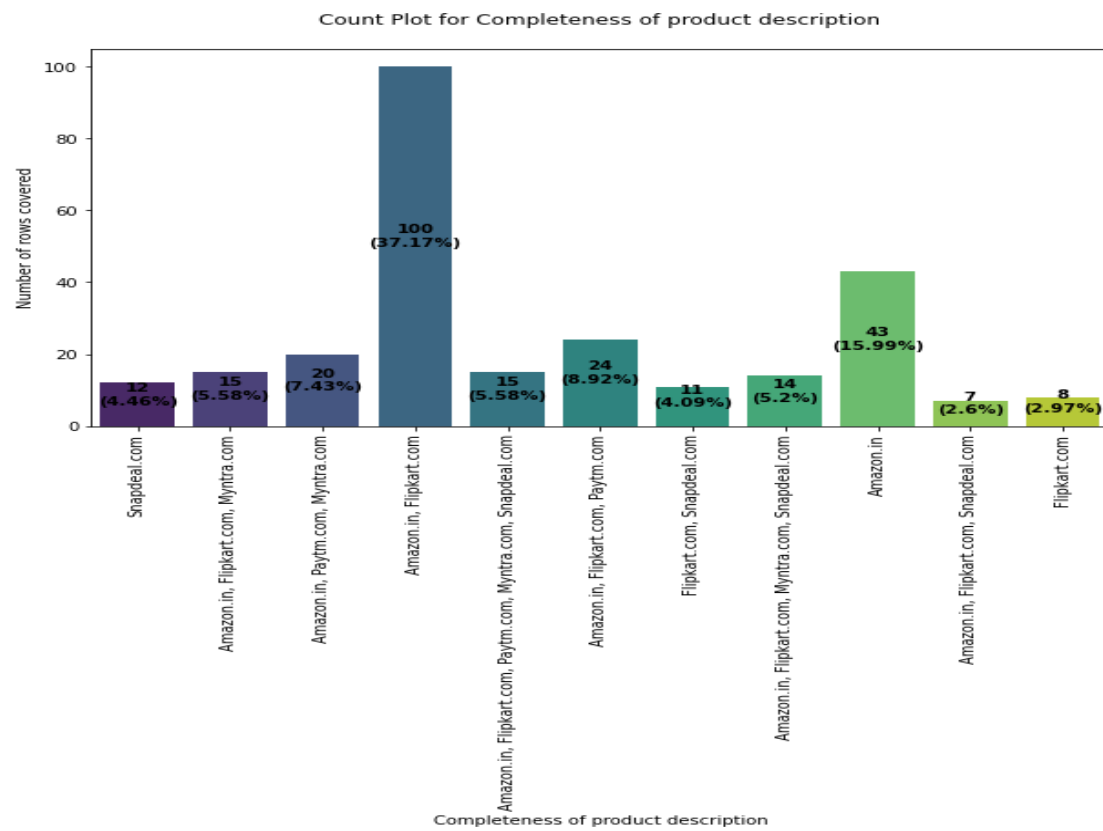


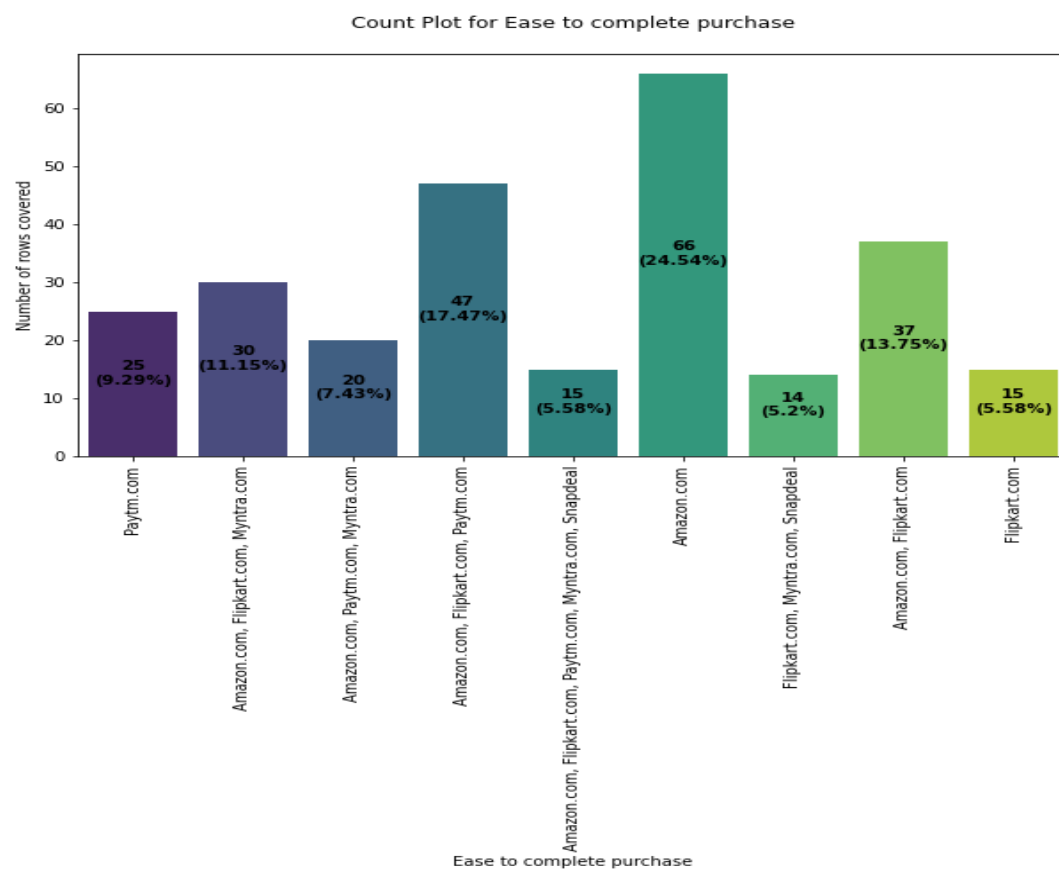
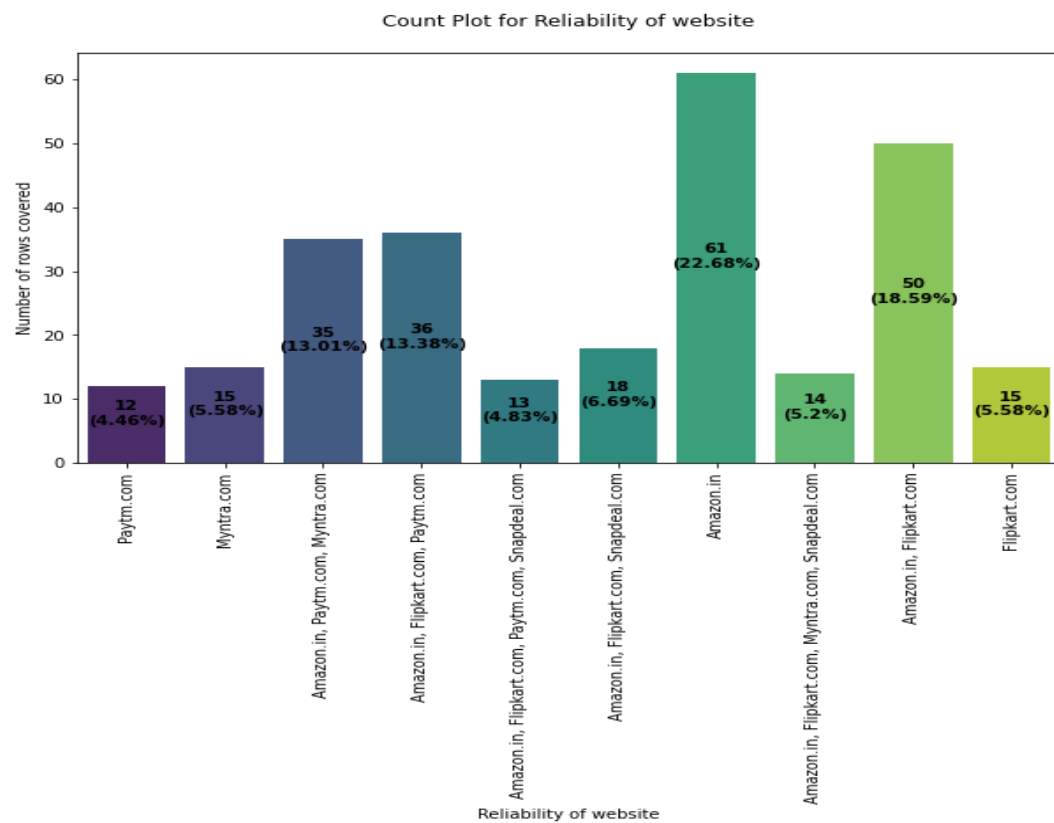
Count Plot for Value for money

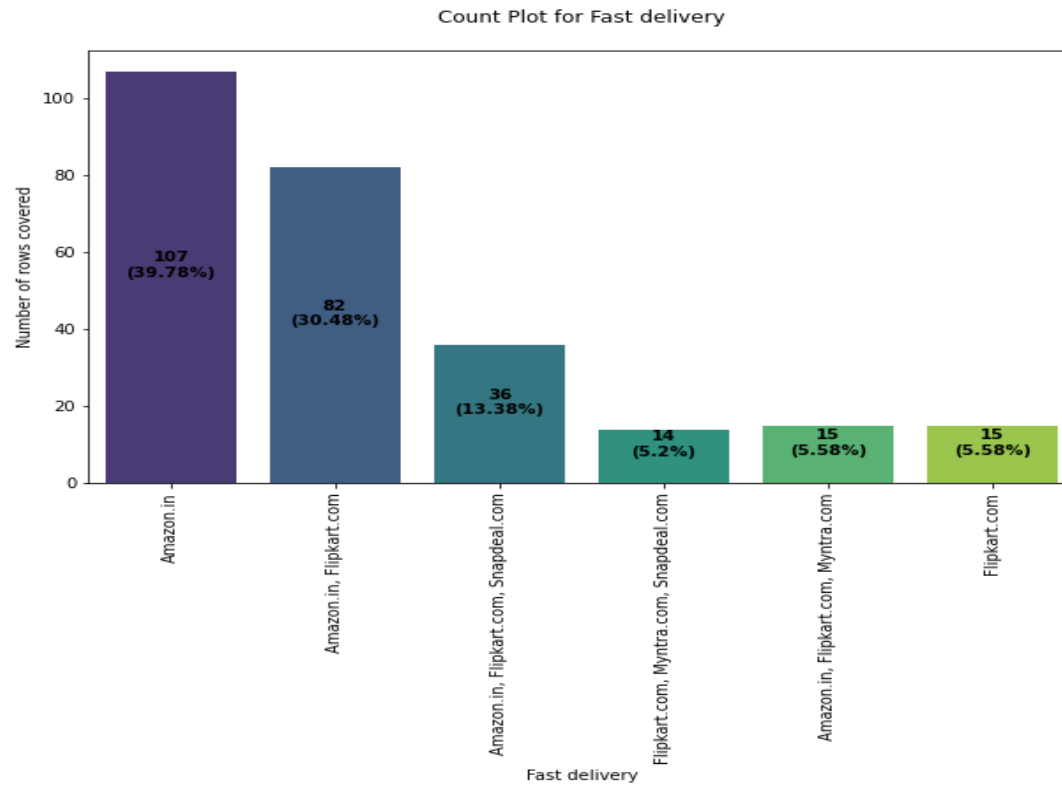
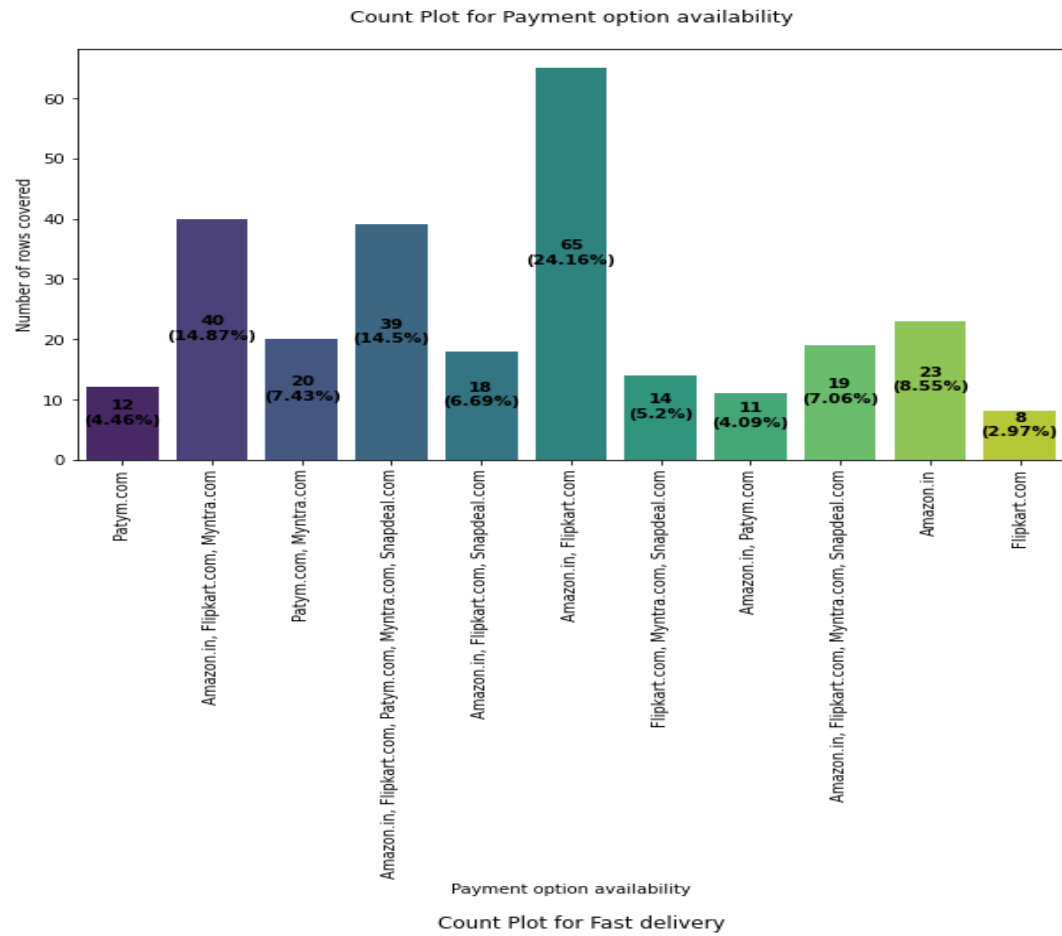


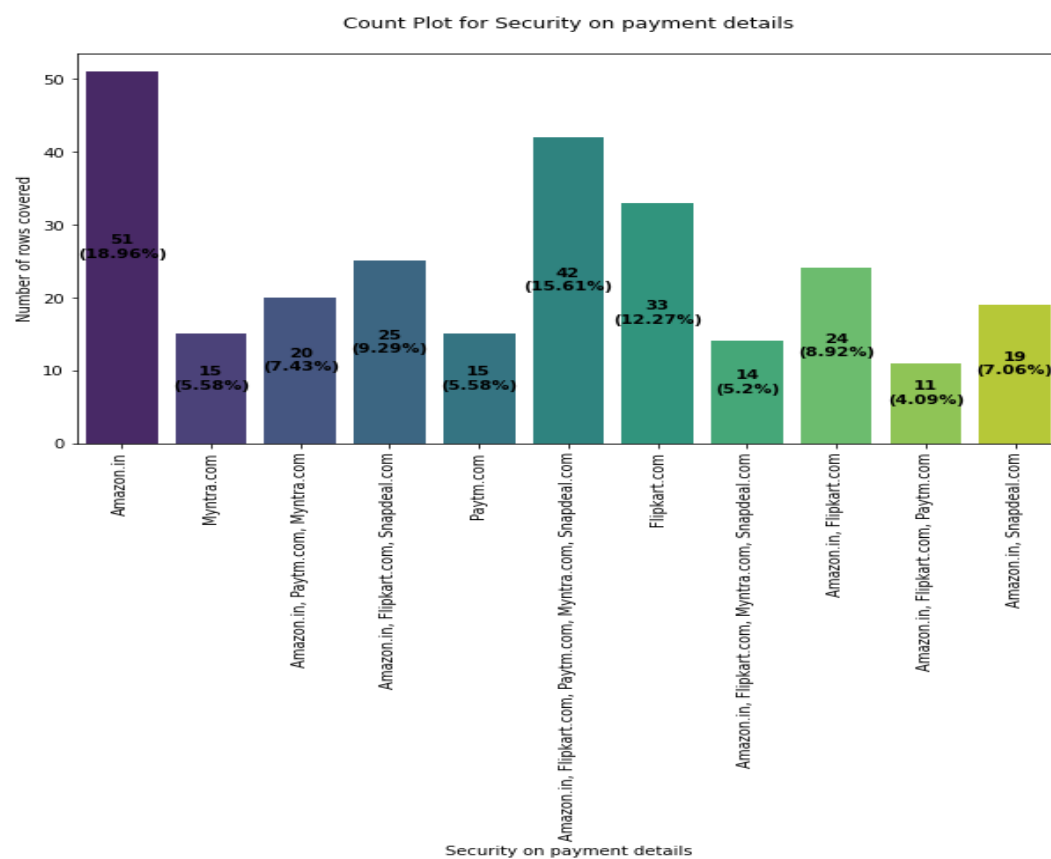
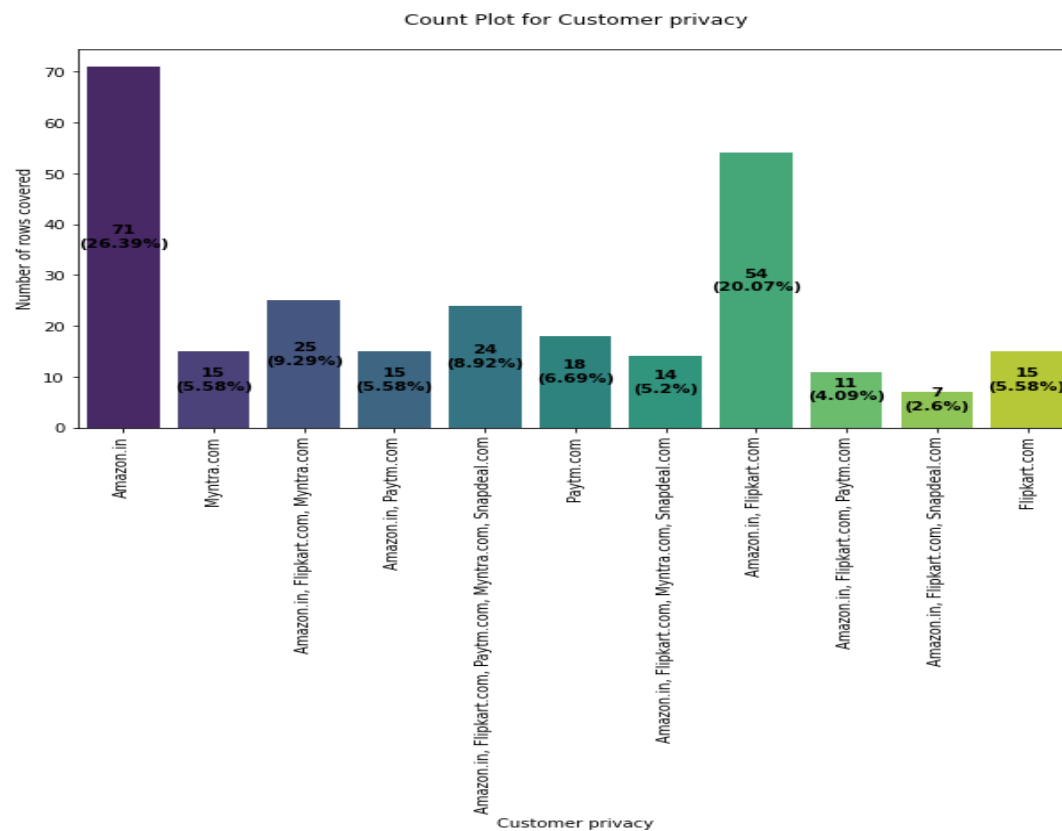


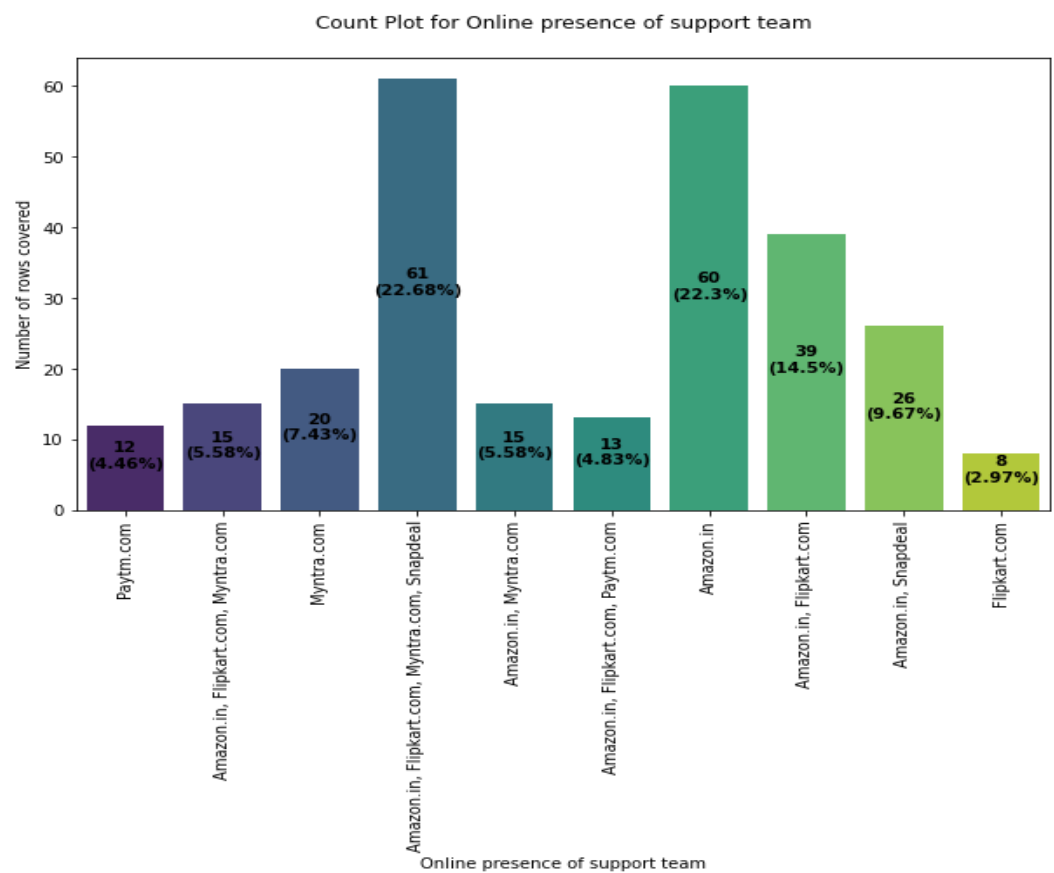
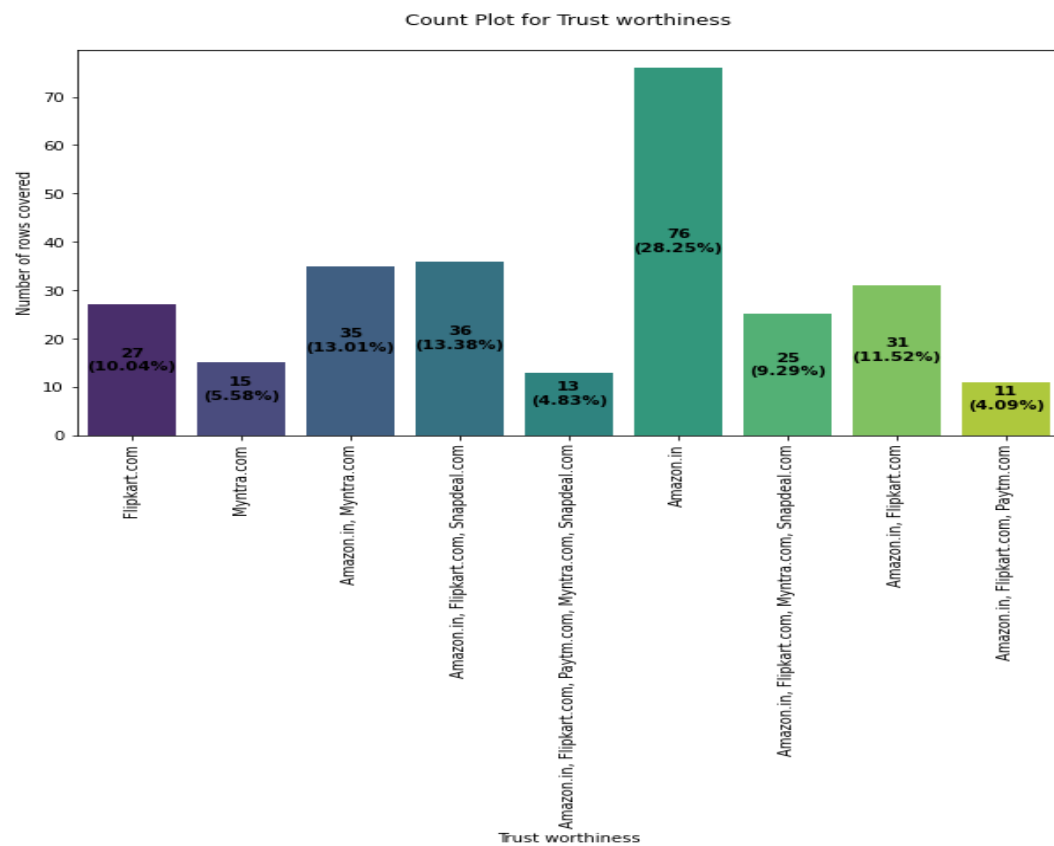




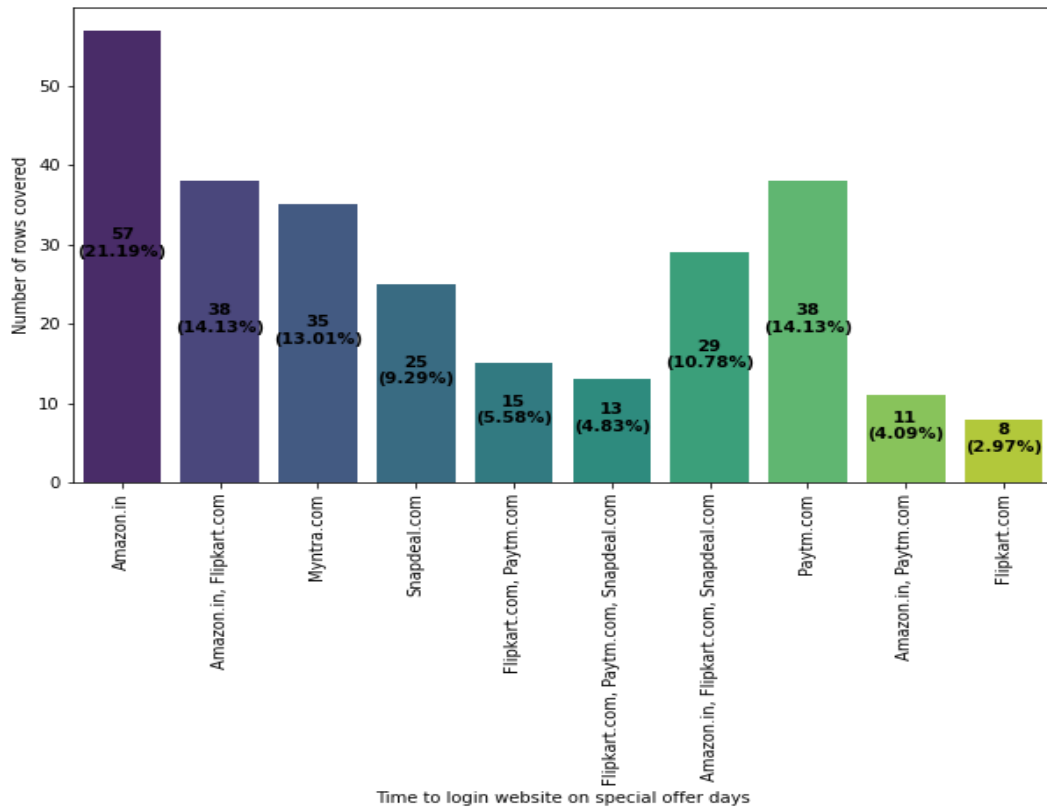




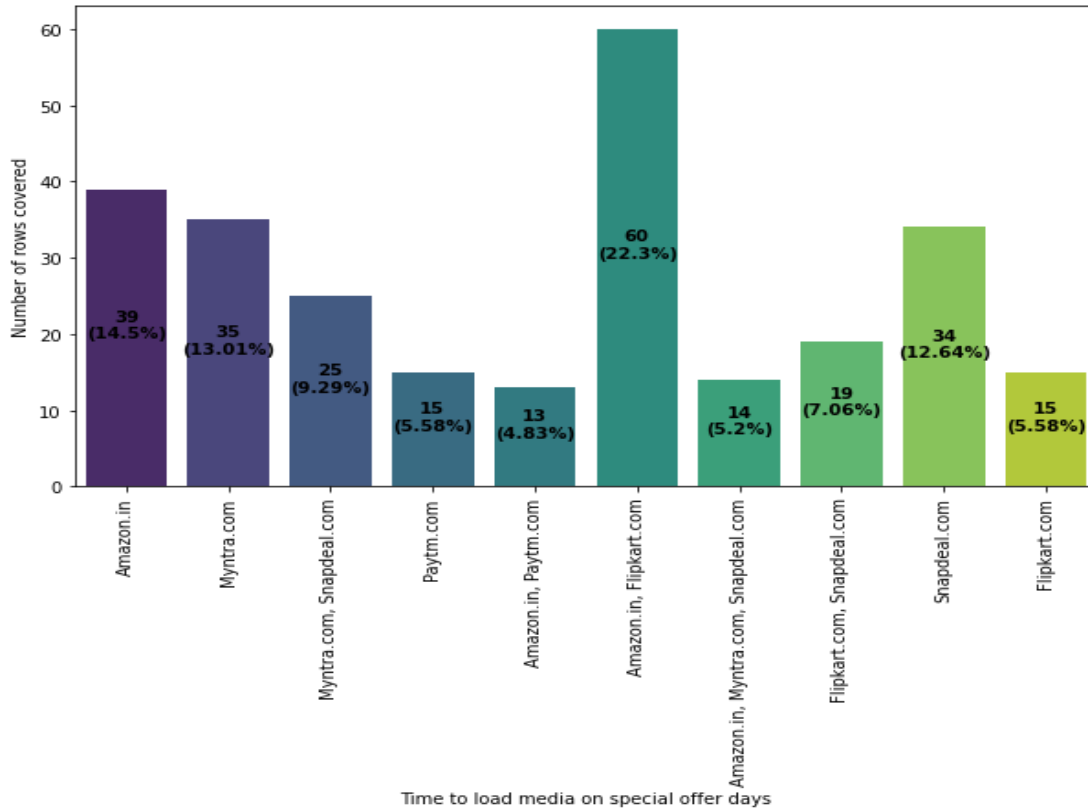




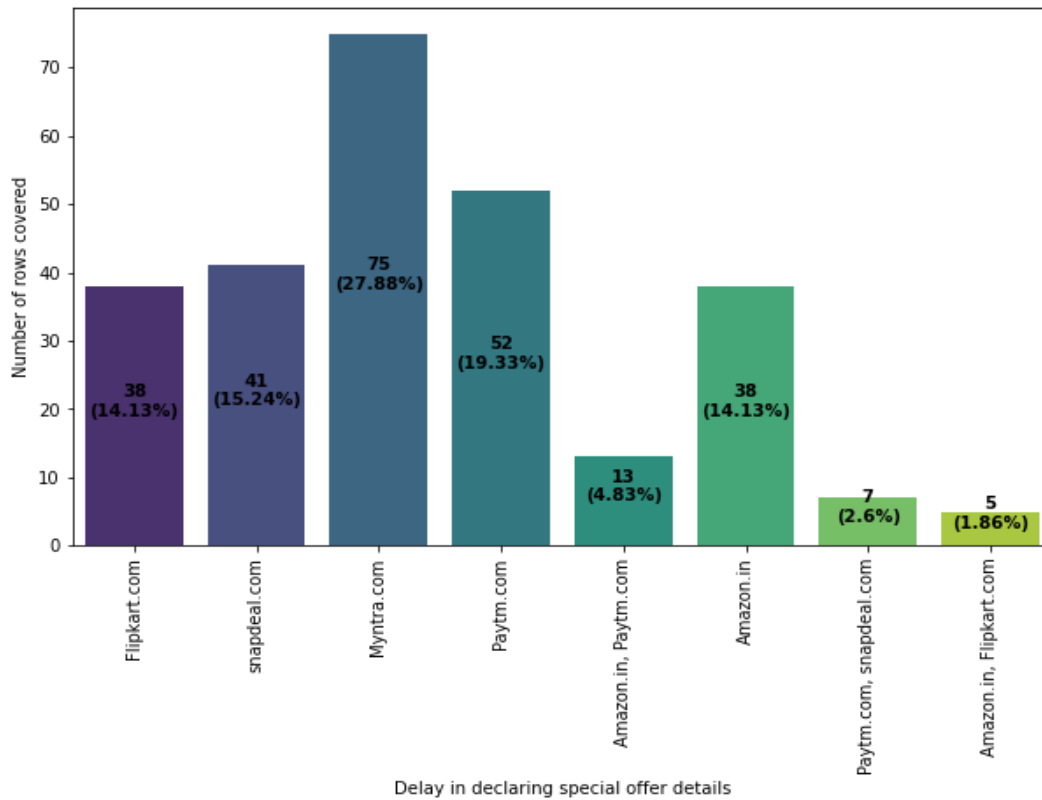
Count Plot for Time to login website on special offer days



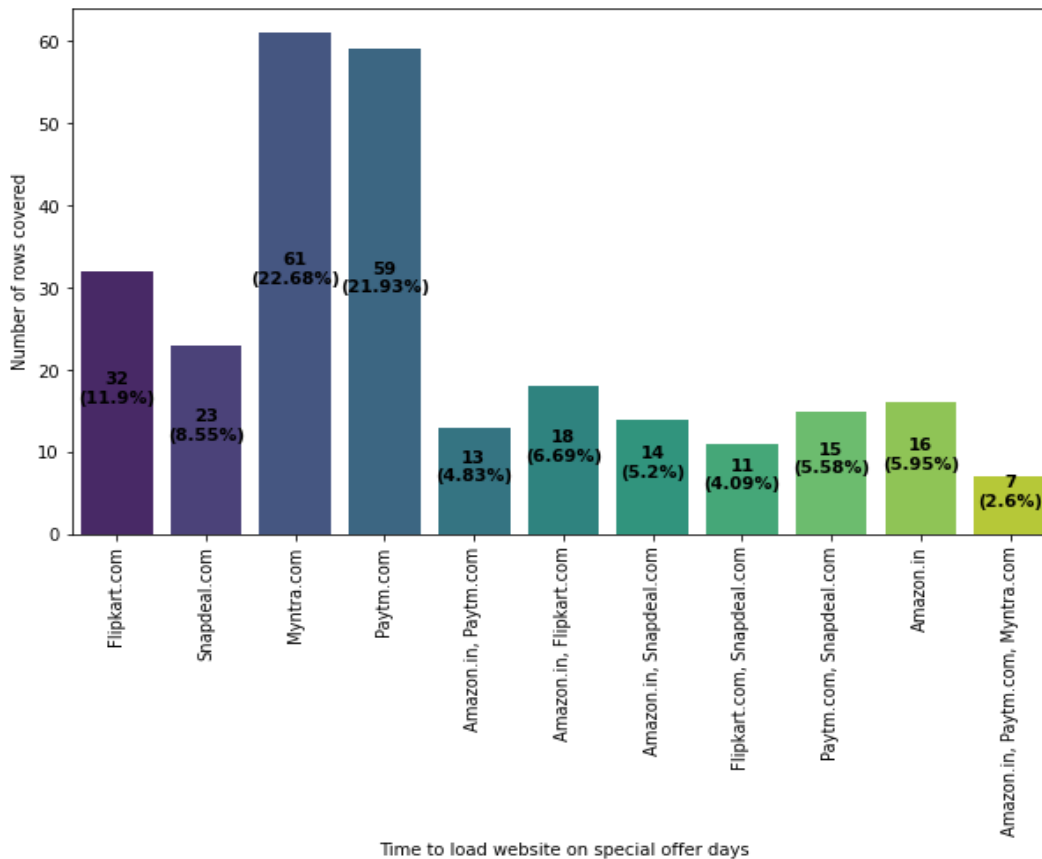
Count Plot for Time to load media on special offer days



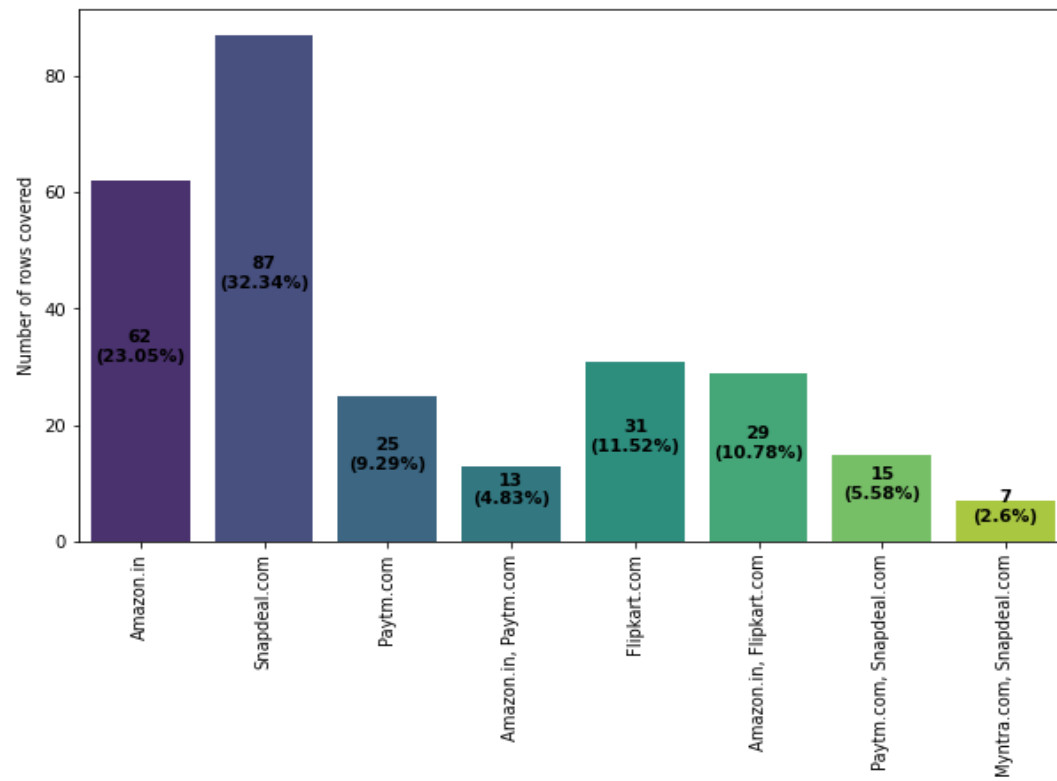
Count Plot for Delay in declaring special offer details



Count Plot for Time to load website on special offer days

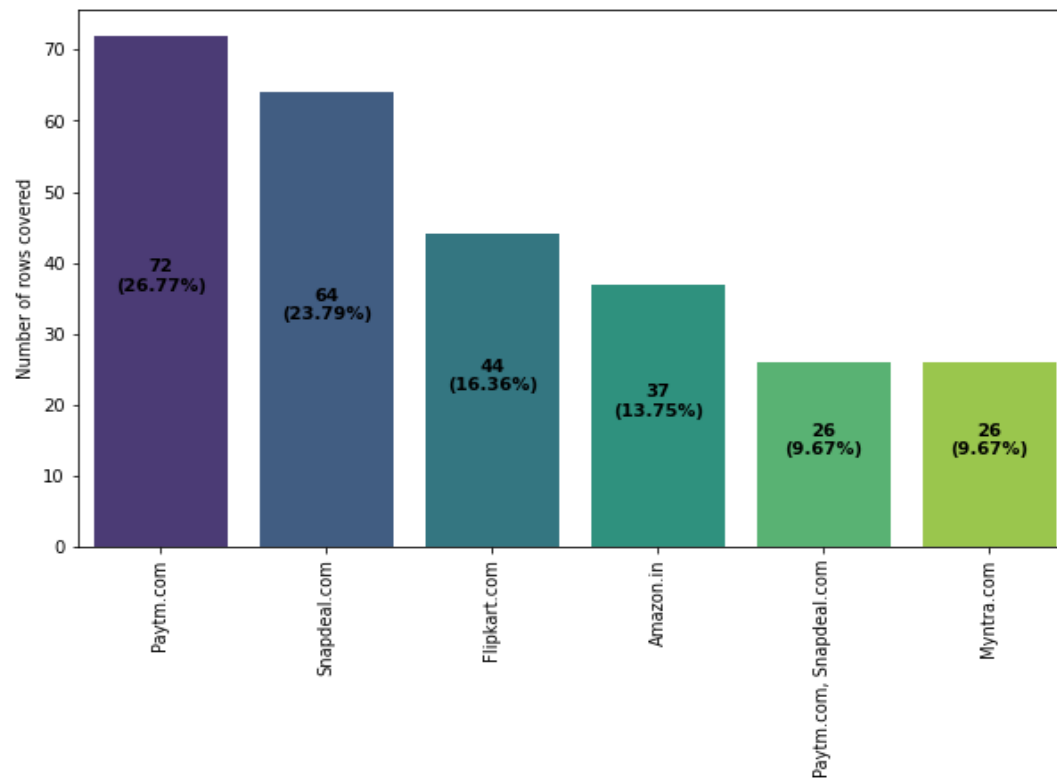


Count Plot for Limited mode of payment on special offer days

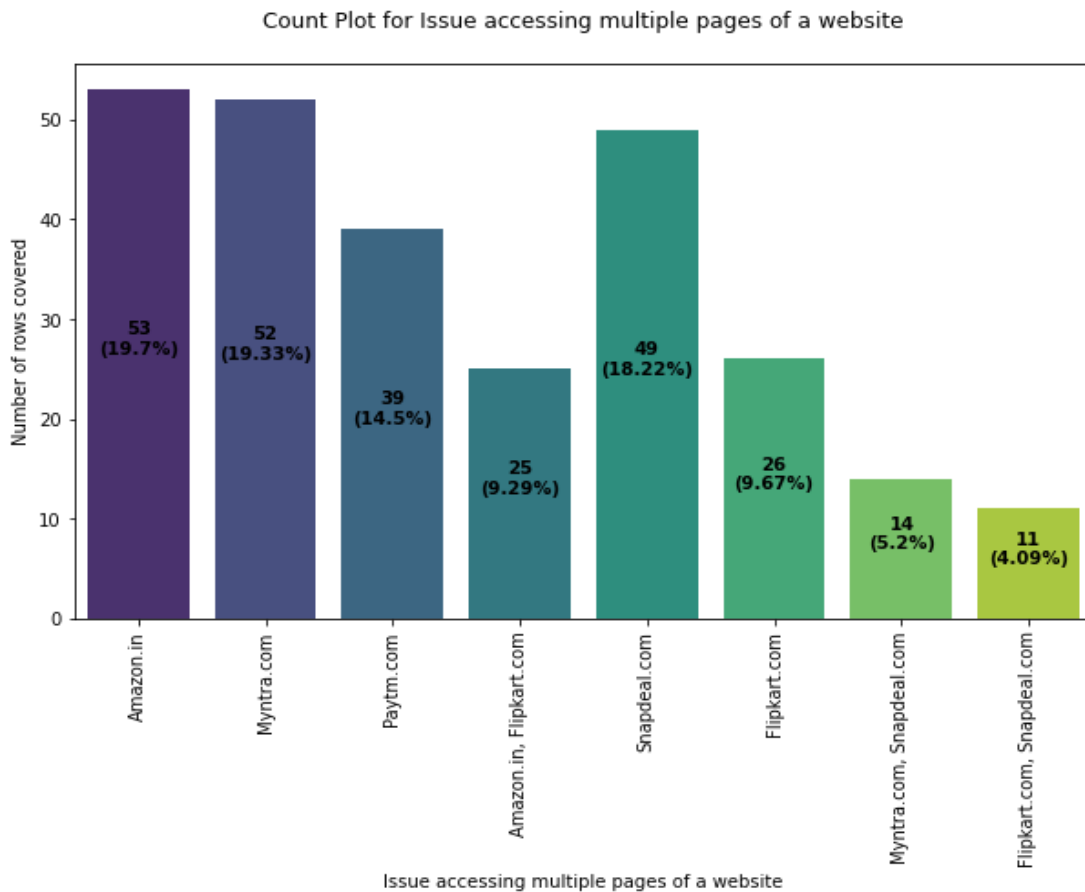
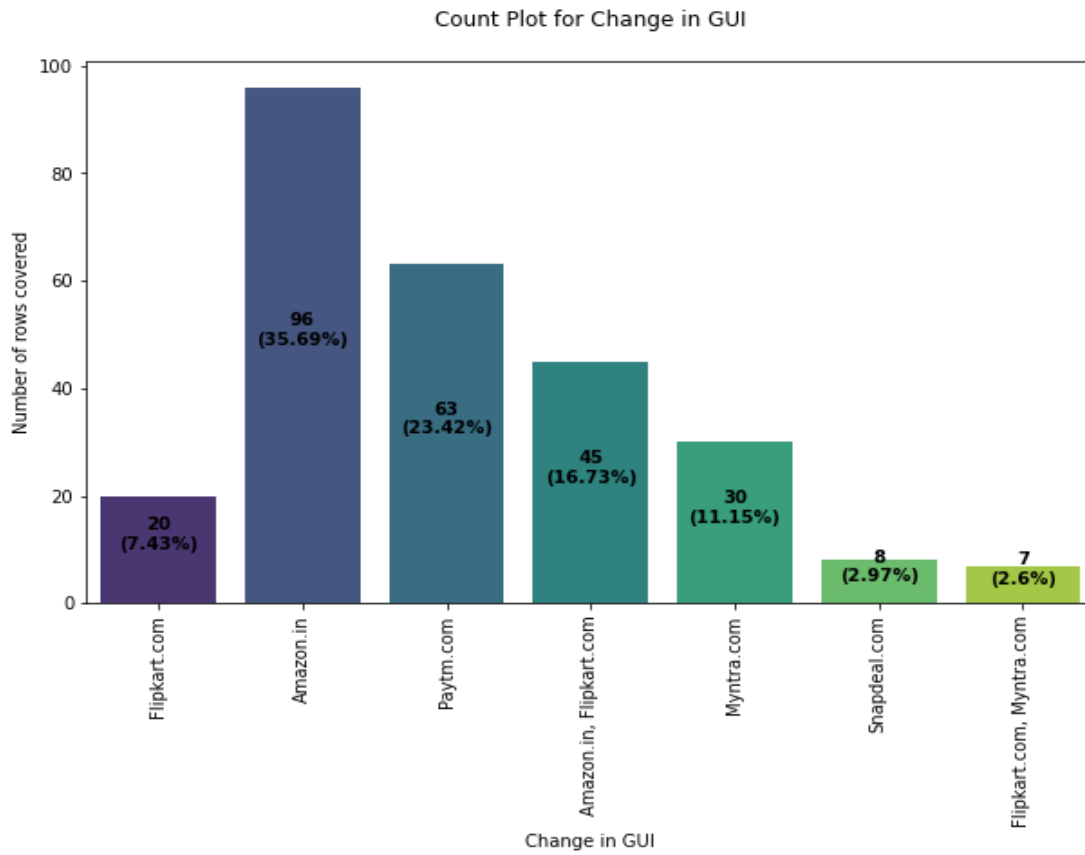


Limited mode of payment on special offer days

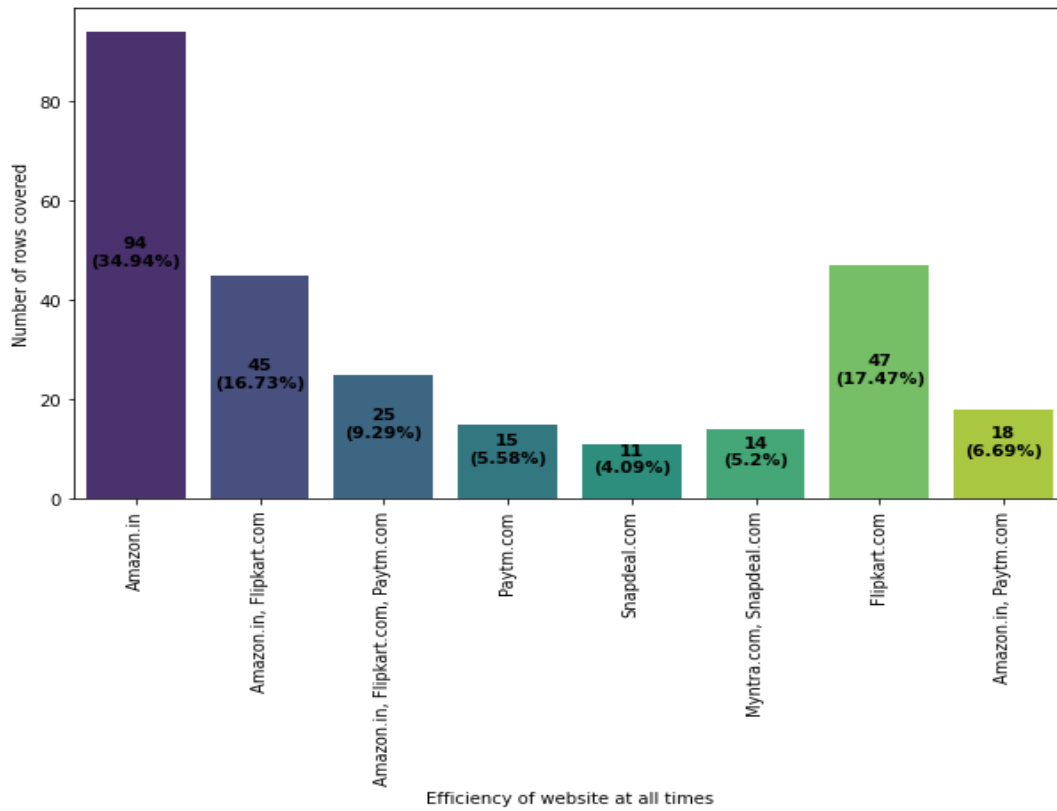
Count Plot for Delayed product delivery



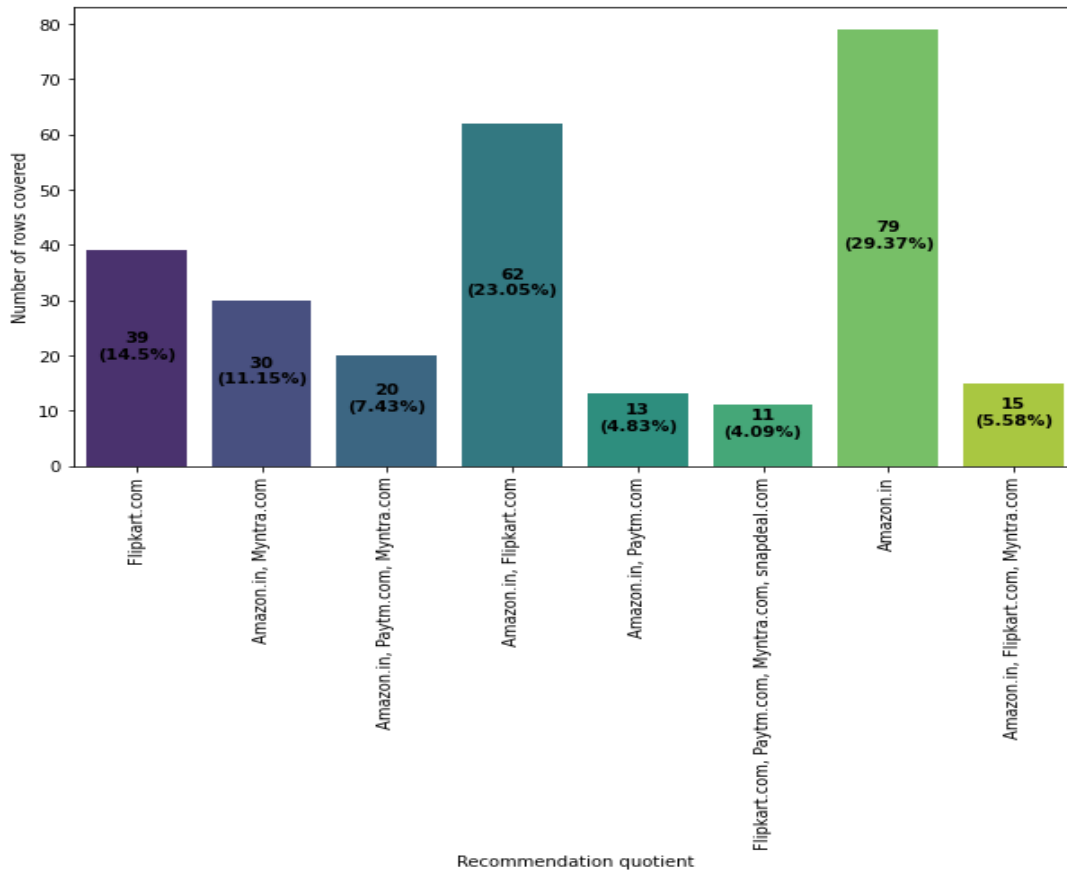
Delayed product delivery



Count Plot for Efficiency of website at all times



Count Plot for Recommendation quotient



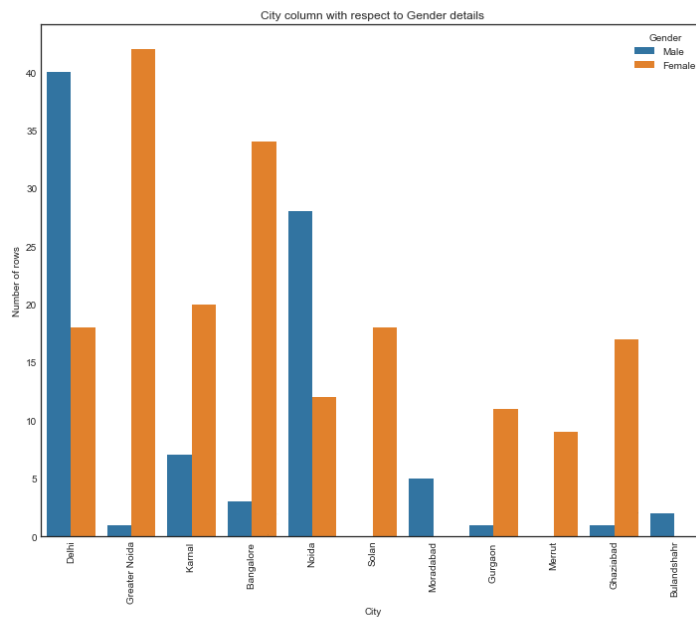
Bivariate Analysis:

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()
```

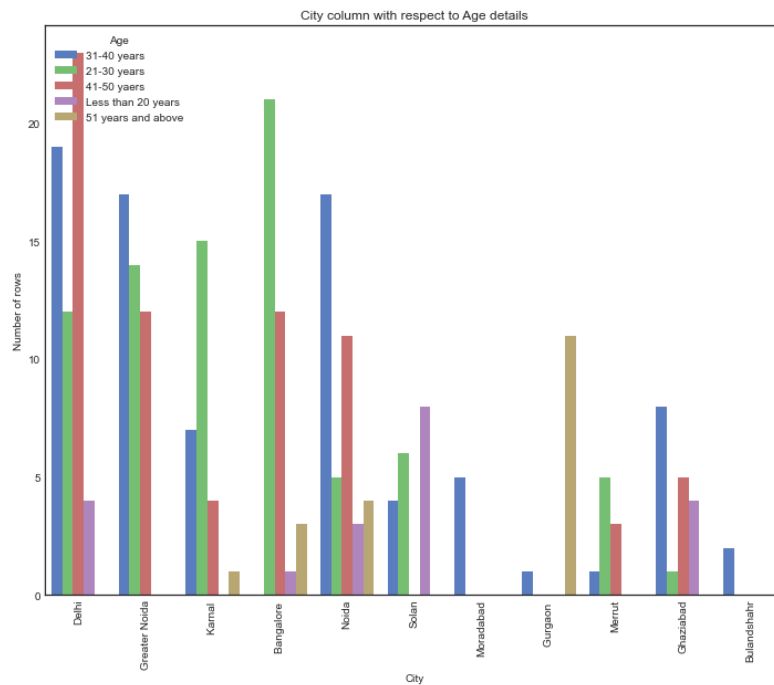
Output:



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()
```

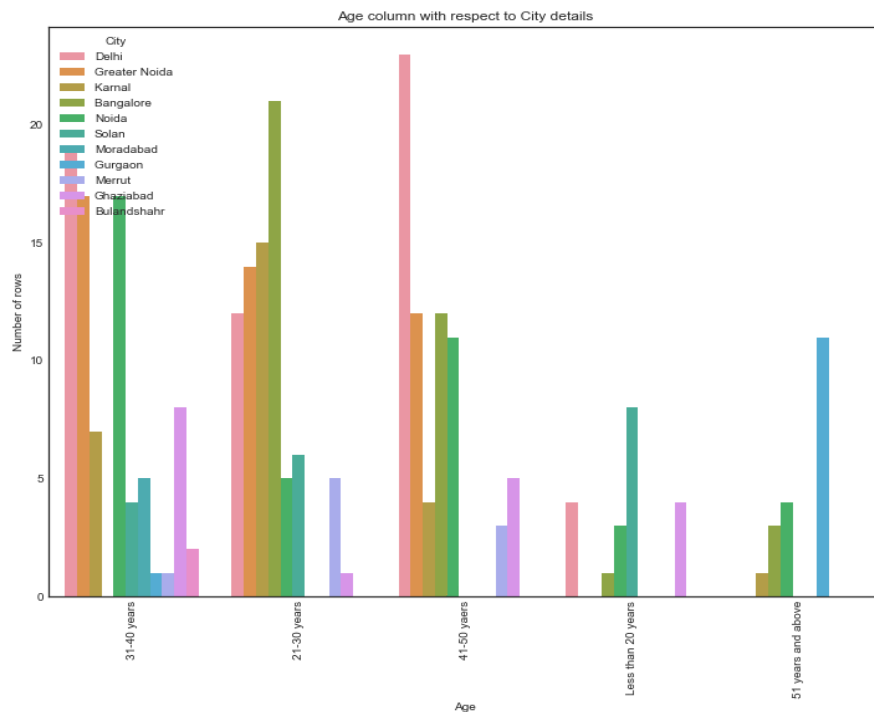
Output:



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:



Next, I imported Coded data and checked and compared all the ratings column.



Code:

```
#Let's also use the coded data for this part

coded_data = pd.read_excel('customer_retention_dataset.xlsx','codedsheet')
df = pd.read_excel('customer_retention_dataset.xlsx','datasheet')

rating_cols=df.iloc[:,17:47].columns
rating_cols=rating_cols.to_list()

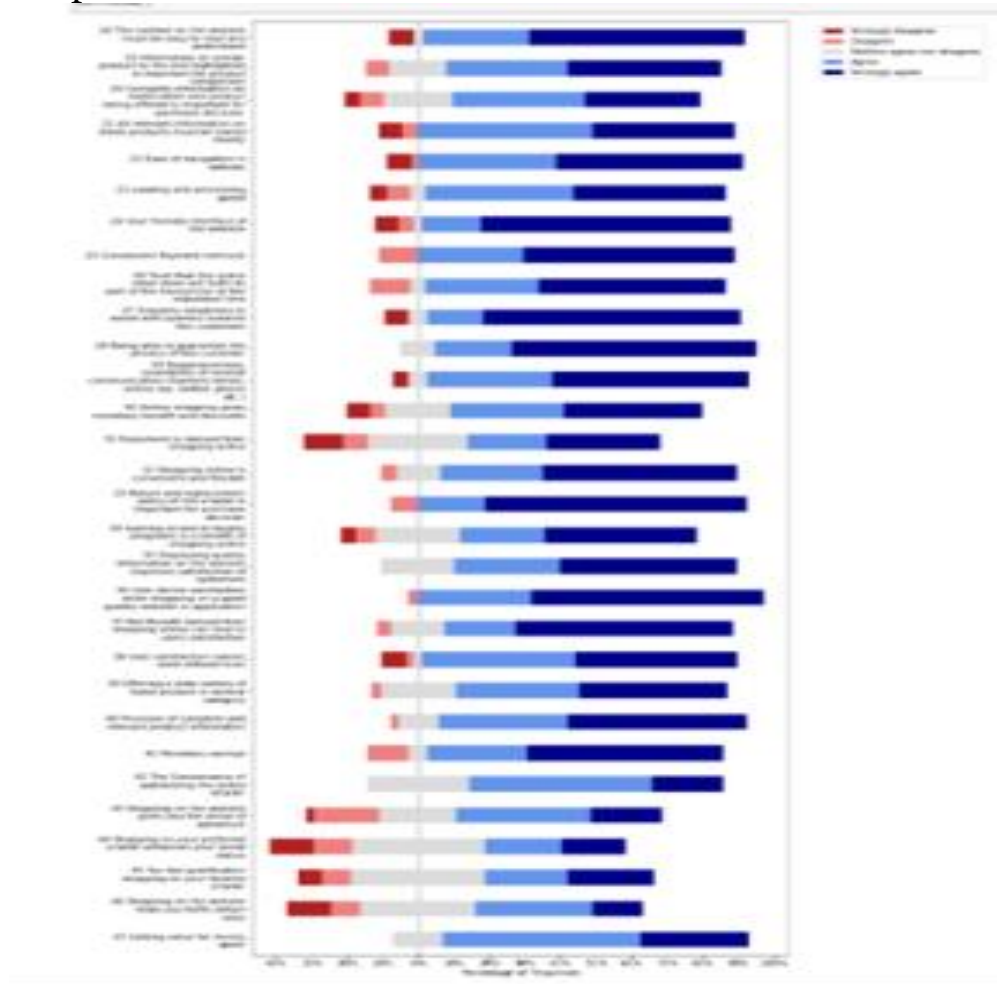
ratings = {1: 'Strongly disagree', 2: 'Disagree', 3: 'Neither agree nor disagree', 4: 'Agree', 5: 'Strongly agree'}

# Rename the values of the dataframe
for col in rating_cols:
    coded_data.replace({col: ratings},inplace=True)

import plot_likert

plot_likert.plot_likert(coded_data[rating_cols], plot_likert.scales.agree, plot_percentage=True,figsize=(10,25))
plt.show()
```

Output:



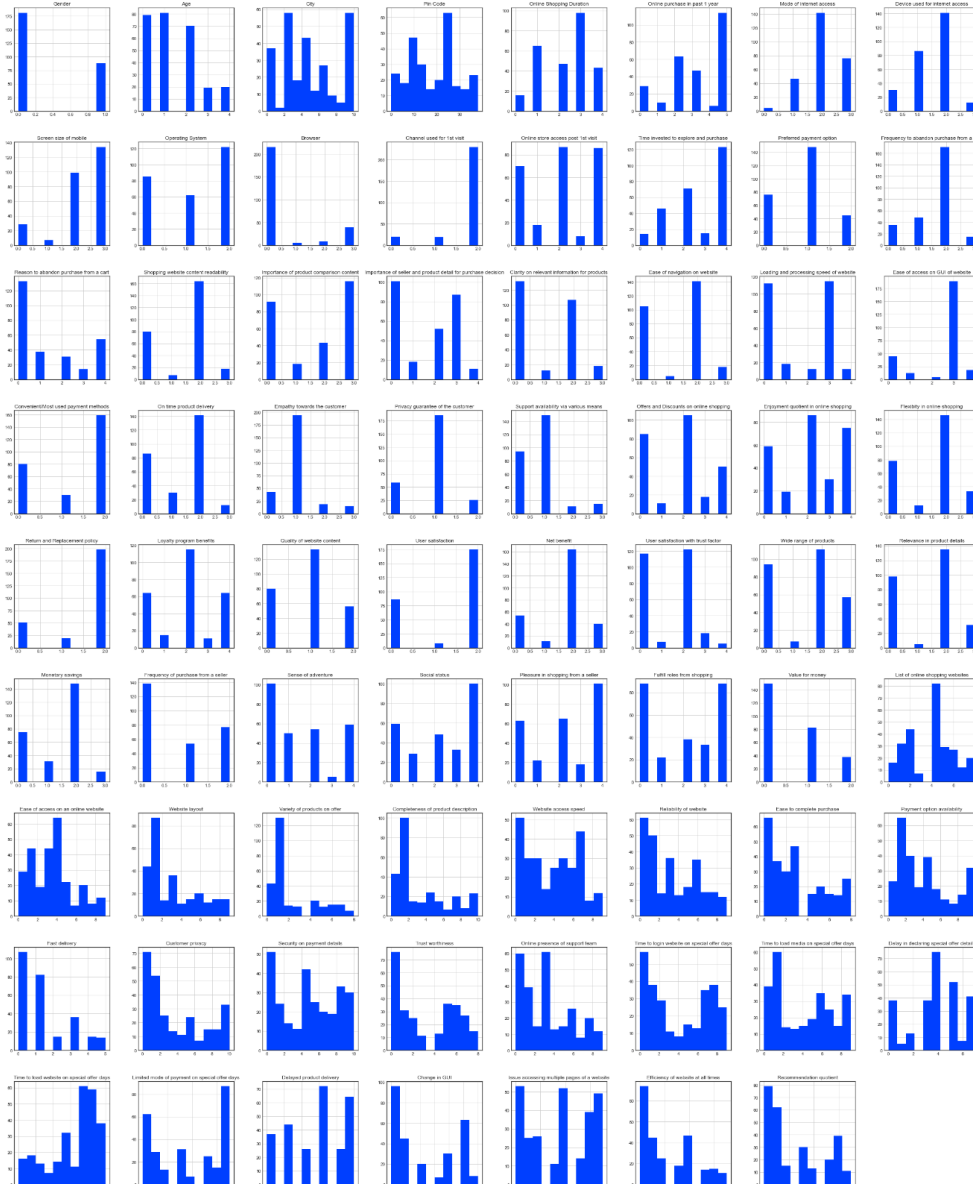
Multivariate Analysis:

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.

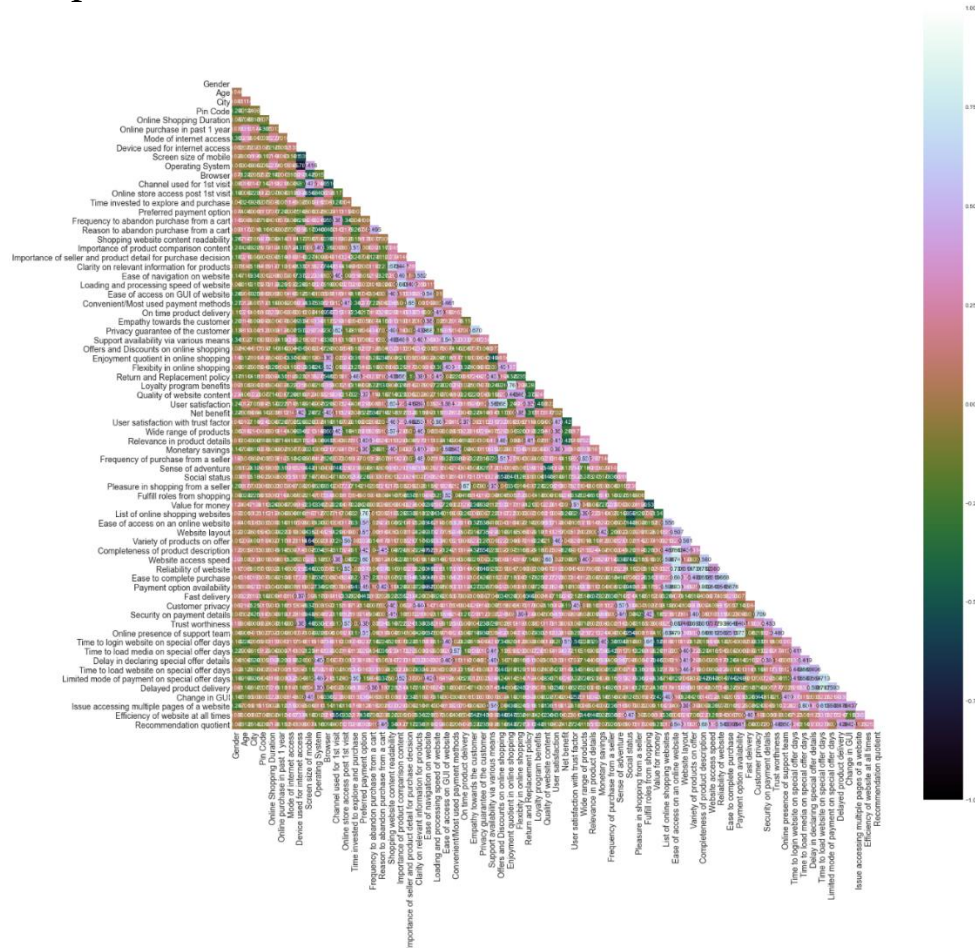
Negative correlation - A correlation of -1 indicates a perfect negative correlation, meaning that as one variable goes up, the other goes down.

Code:

```
#Correlation using a Heatmap

upper_triangle = np.triu(coded_data.corr())
plt.figure(figsize=(25,25))
sns.heatmap(coded_data.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:



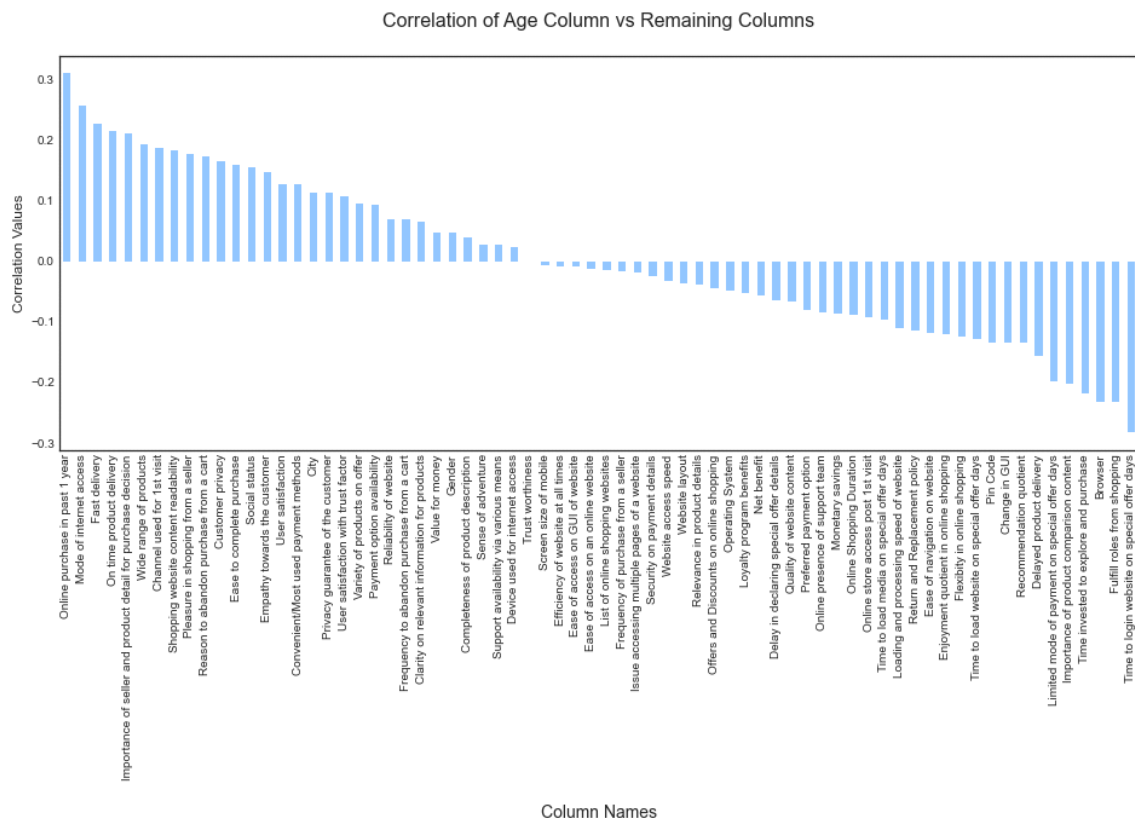
In the above heatmap due to lot of columns we are not able to see the correlation details however we can observe the color-coding details and get a hint that there is no multi collinearity concern between the column values.

Code:

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

column_names = df.columns
for col in df[column_names]:
    df_corr = df.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.

Building Machine Learning Model:

Code & Output:



Preprocessing Pipeline

Separating the label from rest of the features

```
In [77]: x= coded_data.drop('which of the Indian online retailer would you recommend to a friend?',axis=1,inplace=True)
y=df['which of the Indian online retailer would you recommend to a friend?']
```

Data Standardization

```
In [38]: from sklearn.preprocessing import MinMaxScaler
scaler=MinMaxScaler()
```

```
In [39]: xd=scaler.fit_transform(x)
x=pd.DataFrame(xd,columns=x.columns)
```

Using Feature importance of random forrest

```
In [40]: from sklearn.ensemble import RandomForestClassifier
m=RandomForestClassifier()
m.fit(x,y)
```

```
Out[40]: RandomForestClassifier()
```

```
In [41]: #plot graph of feature importances for better visualization
feat_importances = pd.Series(m.feature_importances_, index=x.columns)
plt.figure(figsize=(10,8))
```

Yay! we predicted the model with 100% accuracy!

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:

The results of this study suggest following outputs which might be useful for E-commerce websites to extend their business

1. The cost of the product, the reliability of the E-commerce company and the return policies all play an equally important role in deciding the buying behaviour of online customers. The cost is an important factor as it was the basic criteria used by online retailers to attract customers. The reliability of the E-commerce company is also important, as it is even required in offline retail. It is important because customers are paying online, so they need to be sure of security of the online transaction. The return policies are important because in online retail customer does not get to feel the product. Thus, he wants to be sure that it will be possible to return the product if he does not like it in real. Whereas, the logistics factor, which included Cash on delivery option, One day delivery and the quality of packaging plays a secondary role in this process though these are Must-be-quality. This is so because these all does not interfere with the real product and people believe that this is the basic value that E-commerce websites provide.

2. All the websites were not equally preferred by online customers. Amazon was the most preferred followed by Flipkart. This can be explained easily by previous result that we got. These two companies are most trusted in the industry and hence, have a huge reliability. Also, the sellers listed on these websites are generally from Tier 1 cities as compared to Snapdeal and PayTM which have more sellers from tier 2 and 3 cities. Also, these websites have the most lenient return policies as compared to others and also the time required to process a return is low for these.

