



# CUSTOMER RETENTION CASE STUDY REPORT

Submitted by:

KAYALVIZHI . P

Data Science Intern at Flip Robo Technologies

## **ACKNOWLEDGMENT**

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 project.

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 *Shwetank Mishra* 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, 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.

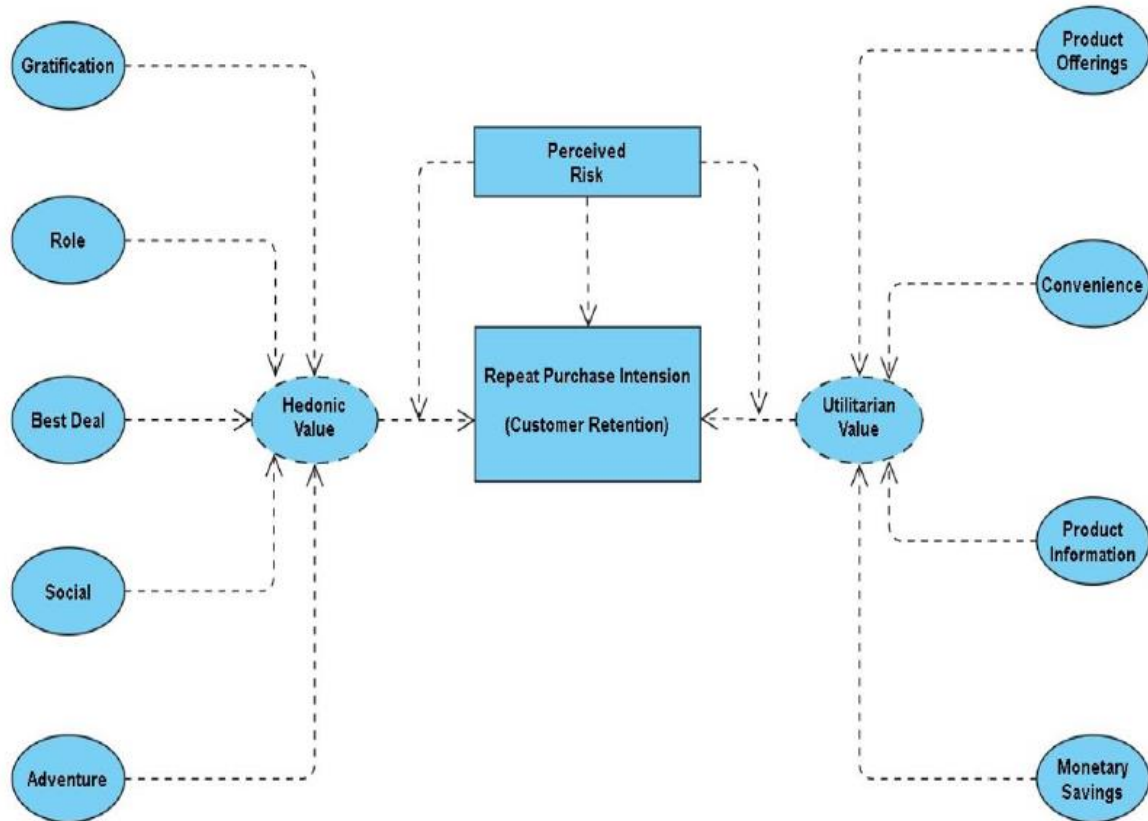
Five major factors that contributed to the success of an e-commerce store have been identified as: service quality, system quality, information quality, trust and net benefit.

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 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.

# 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.





## Dataset Details:

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

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

Then, I imported the datasets.

```
xls = pd.ExcelFile('customer_retention_dataset.xlsx')
```

Then, I separated the imported datasheets present in Excel spreadsheet and make them as a two different dataframe variables df1 and df2.

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

## Exploratory Data Analysis (EDA):

After importing the dataset, next process is EDA. Before that I notice that large numbers of rows and columns are present in dataset, so that information was truncated. To overcome these issues 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 analyse the data. With the help of rename function I was able to change the names of columns that were too lengthy and could have been accommodated in shorter formats.

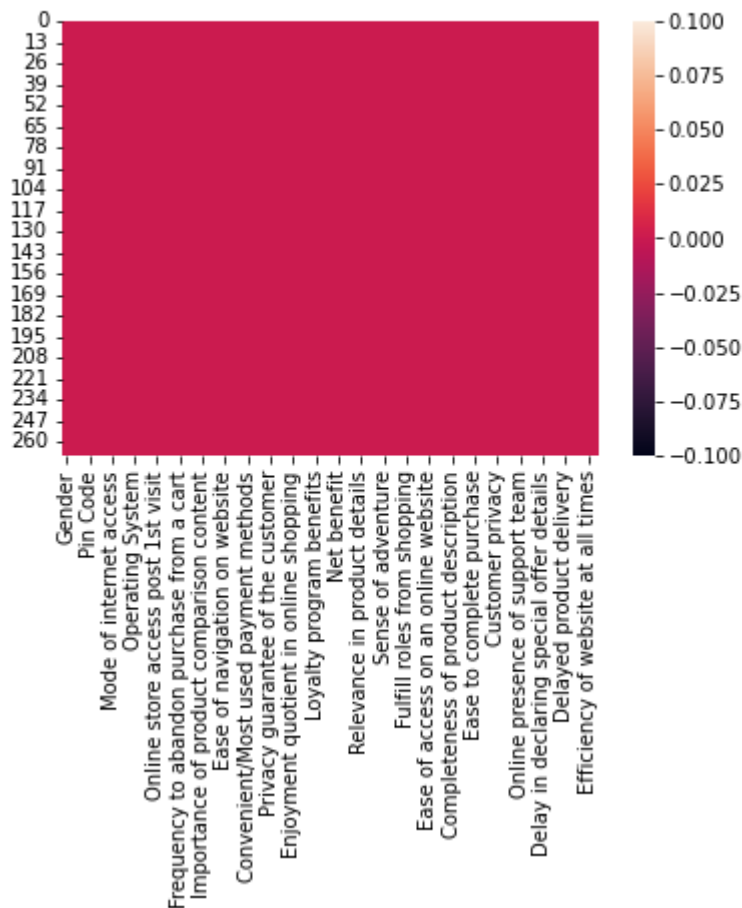
Now, it was the time to take a look at any kind of missing values or null value that might have been present in our dataset by using below given python code.

```
df1.isnull().sum()
```

Luckily, I was able to see that there were no missing values in our entire dataset that is clearly visible in heat map below.

```
sns.heatmap(df1.isnull())
```

<AxesSubplot:>



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

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

```
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 col in object_datatype:
    print(col)
    print(df1[col].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:

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

## Univariate Analysis:

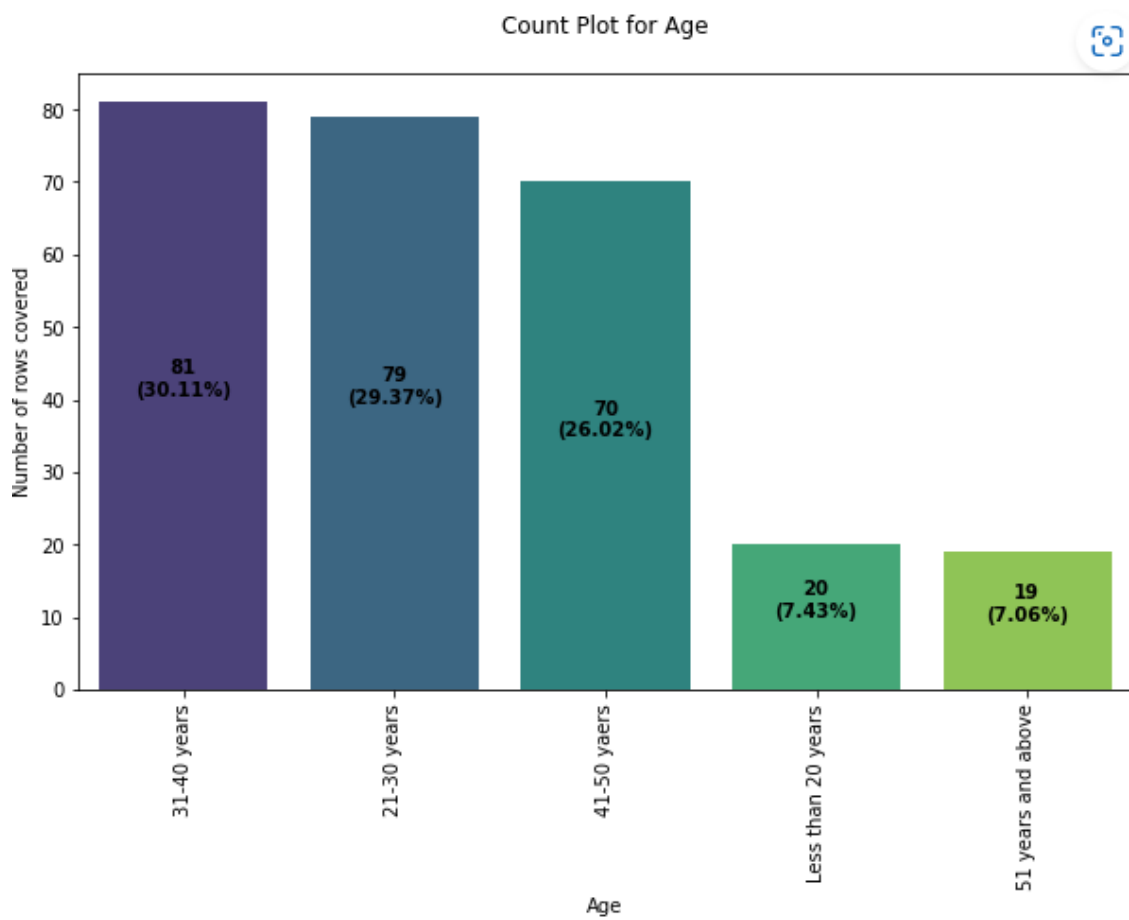
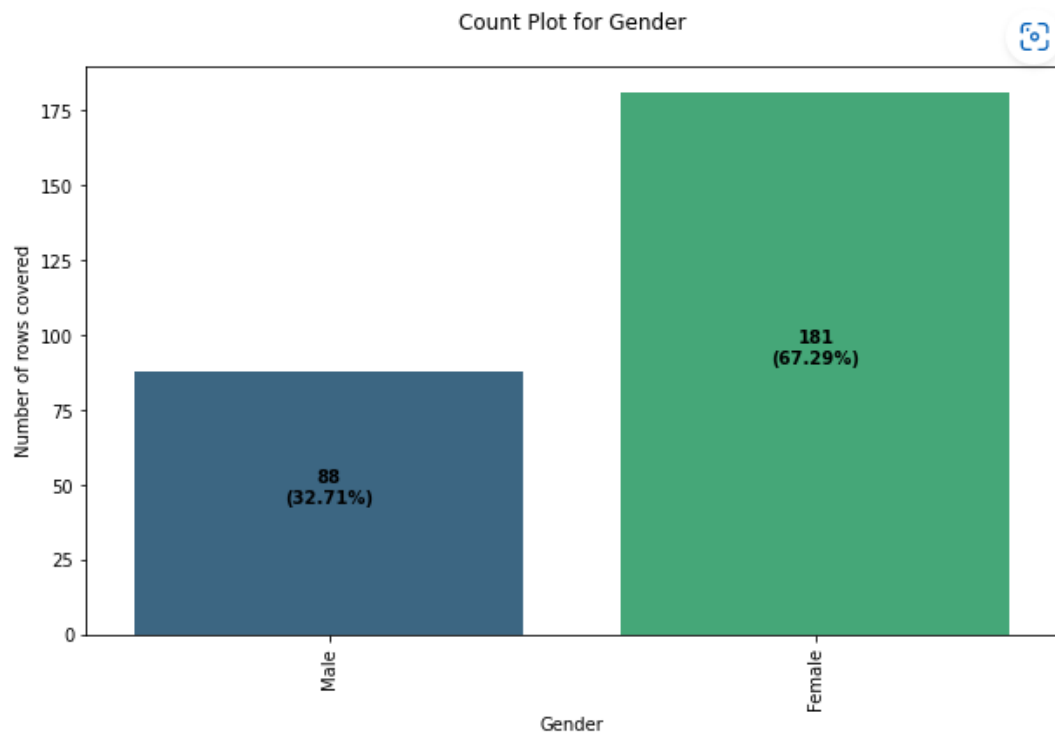
I made use of 2 for loops to generate count plots for all our columns showing the percentage of data coverage.

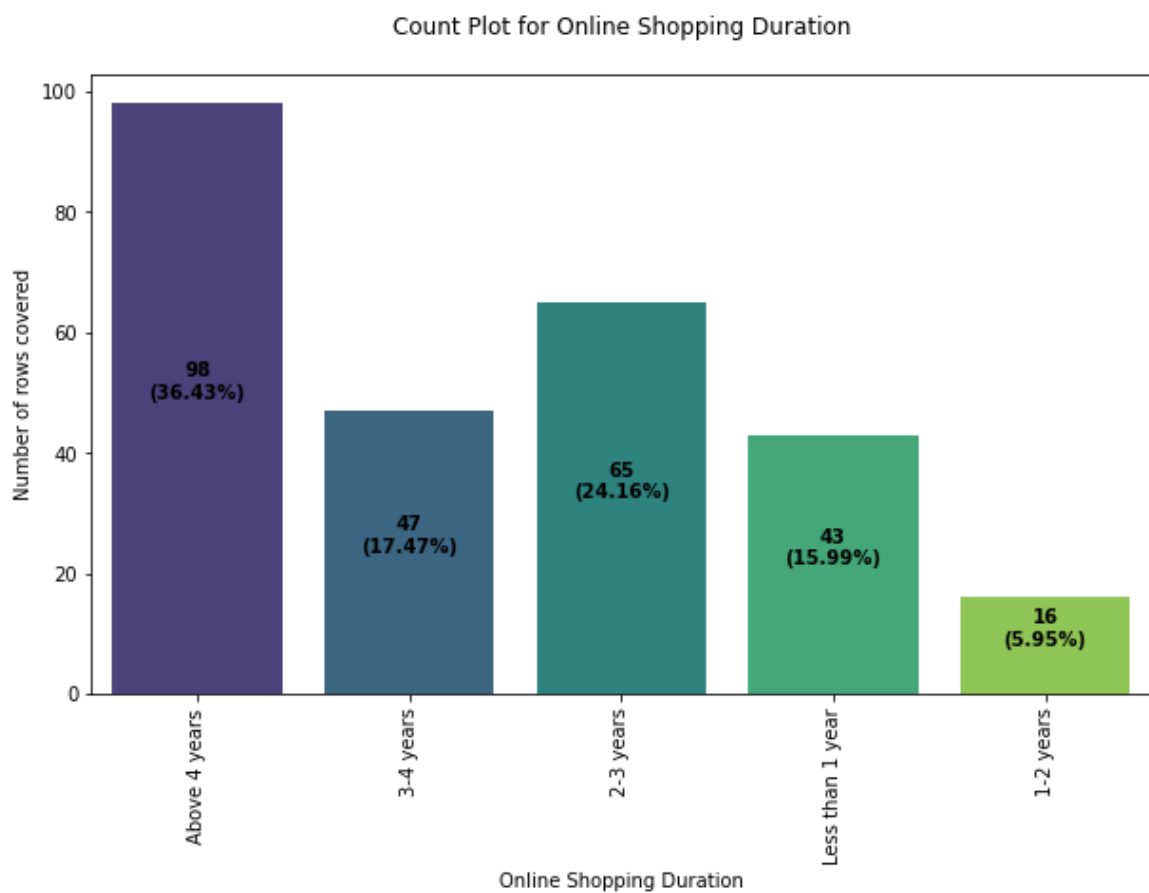
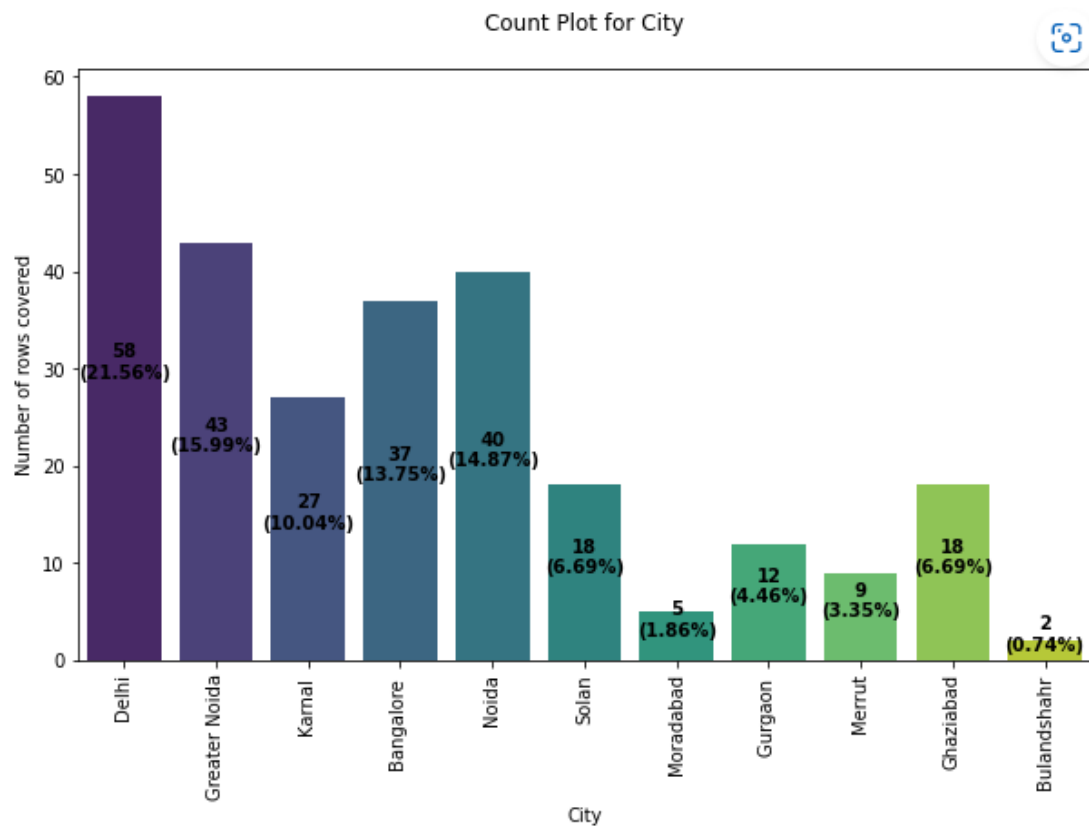
```
for col in df1[object_datatype]:
    plt.figure(figsize=(10,6))
    col_name = col
    values = df1[col_name].value_counts()
    index = 0
    ax = sns.countplot(df1[col_name], palette="viridis")

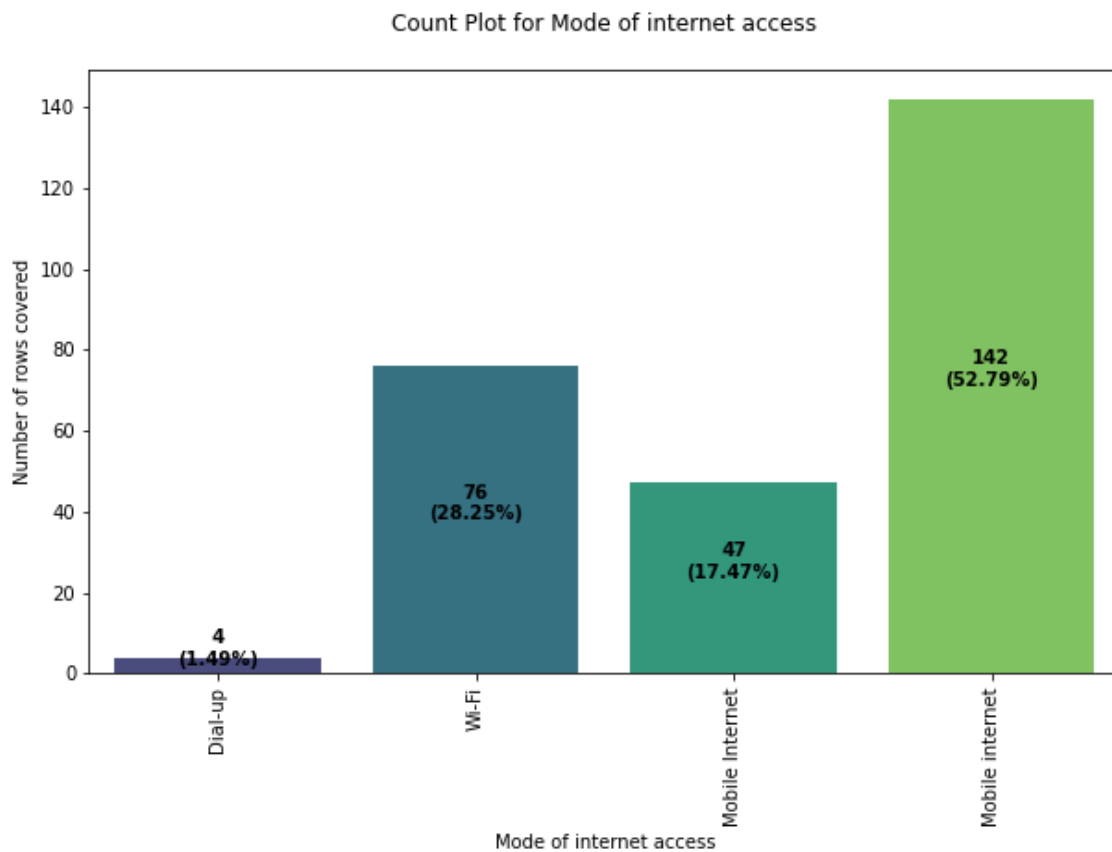
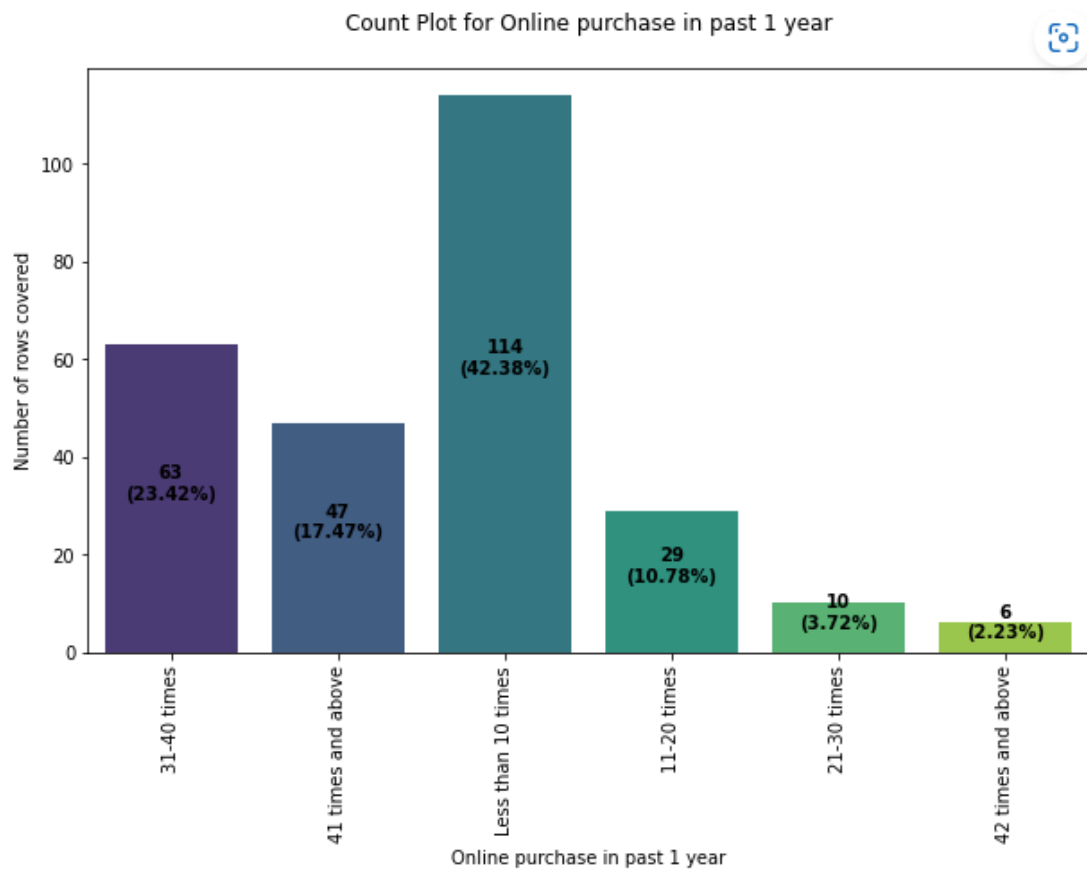
    for i in ax.patches:
        h = i.get_height() # getting the count of each value
        t = len(df1[col_name]) # getting the total number of records using length
        s = f"{h}\n({round(h*100/t,2)}%)" # making the string for displaying in count bar
        plt.text(index, h/2, s, ha="center", fontweight="bold")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

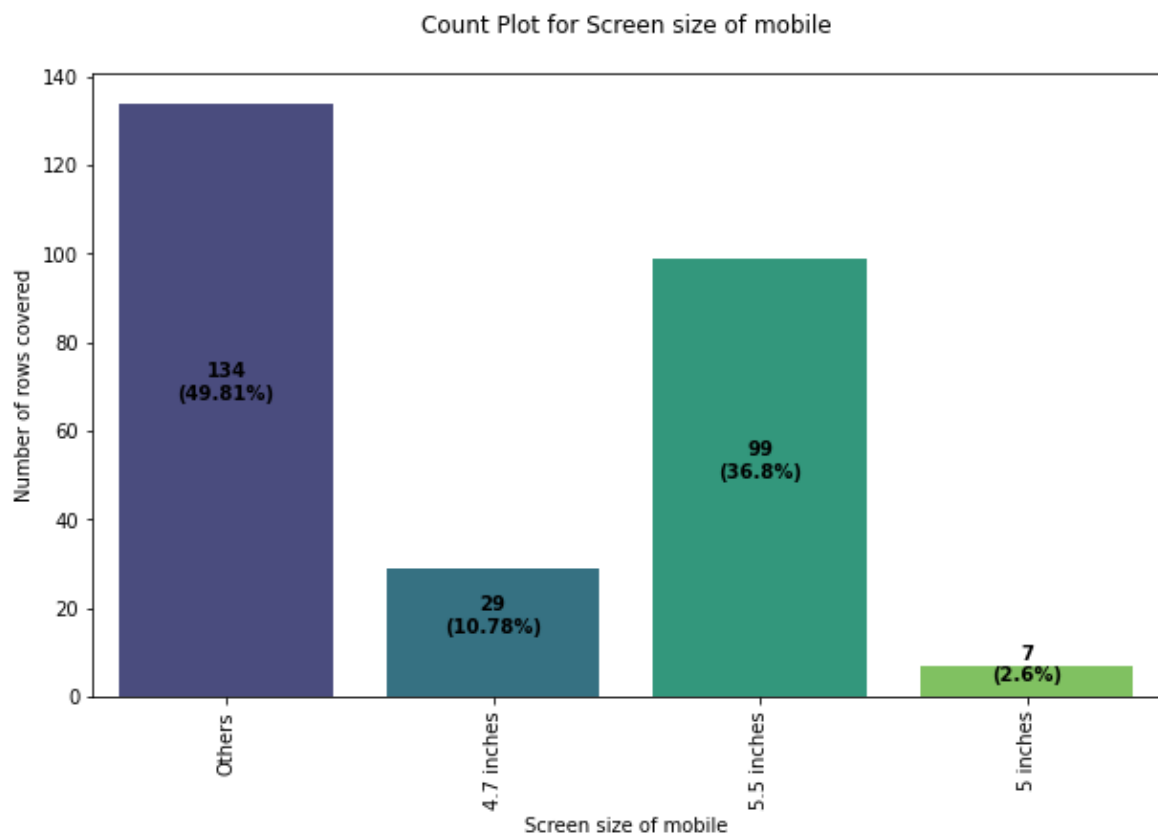
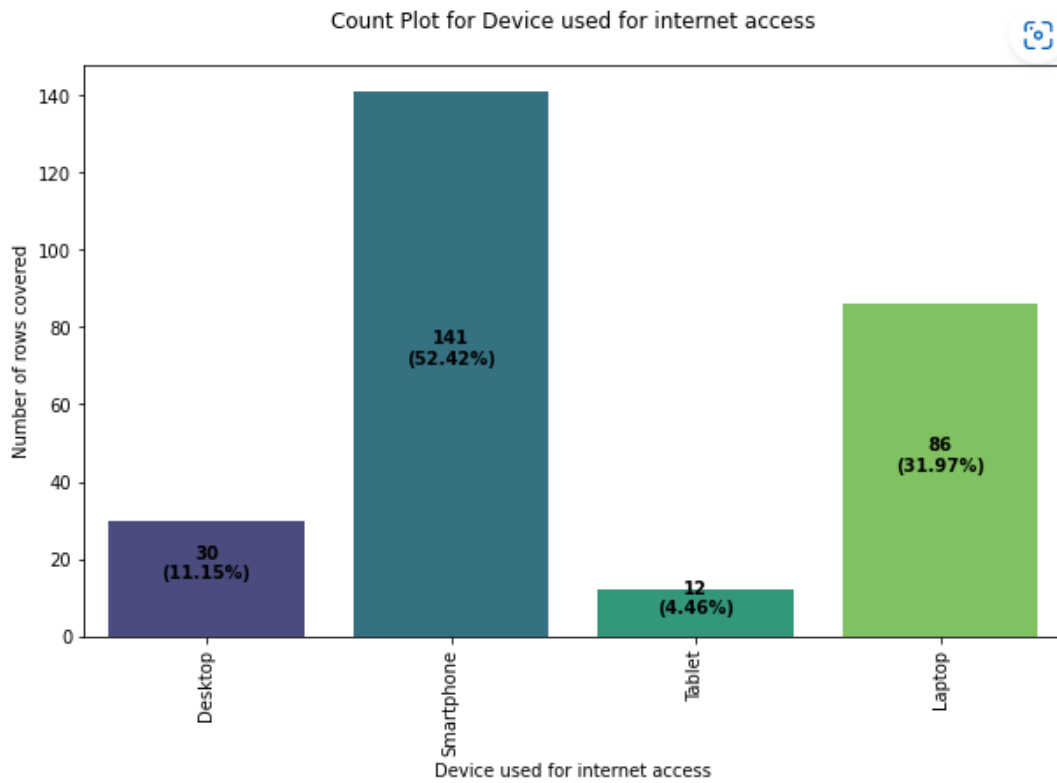
    plt.title(f"Count Plot for {col_name}\n")
    plt.xlabel(col_name)
    plt.ylabel(f"Number of rows covered")
    plt.xticks(rotation=90)
    plt.show()
```

The above code generated multiple count plots as displayed below.

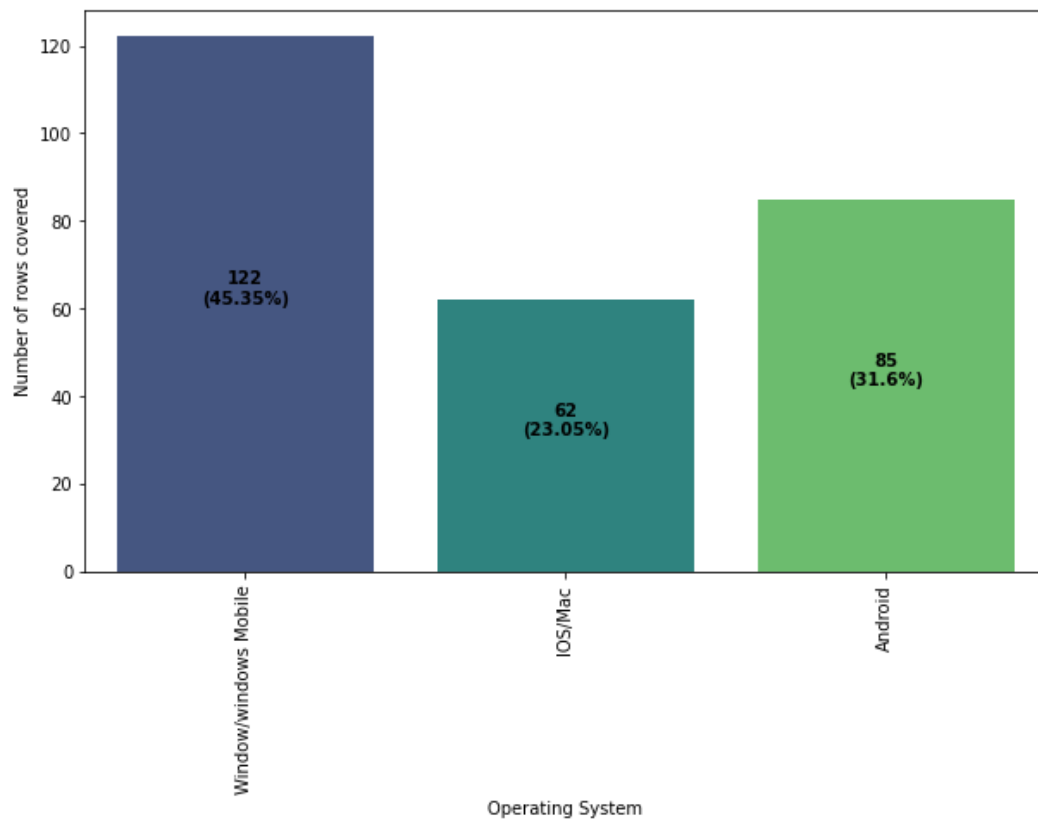




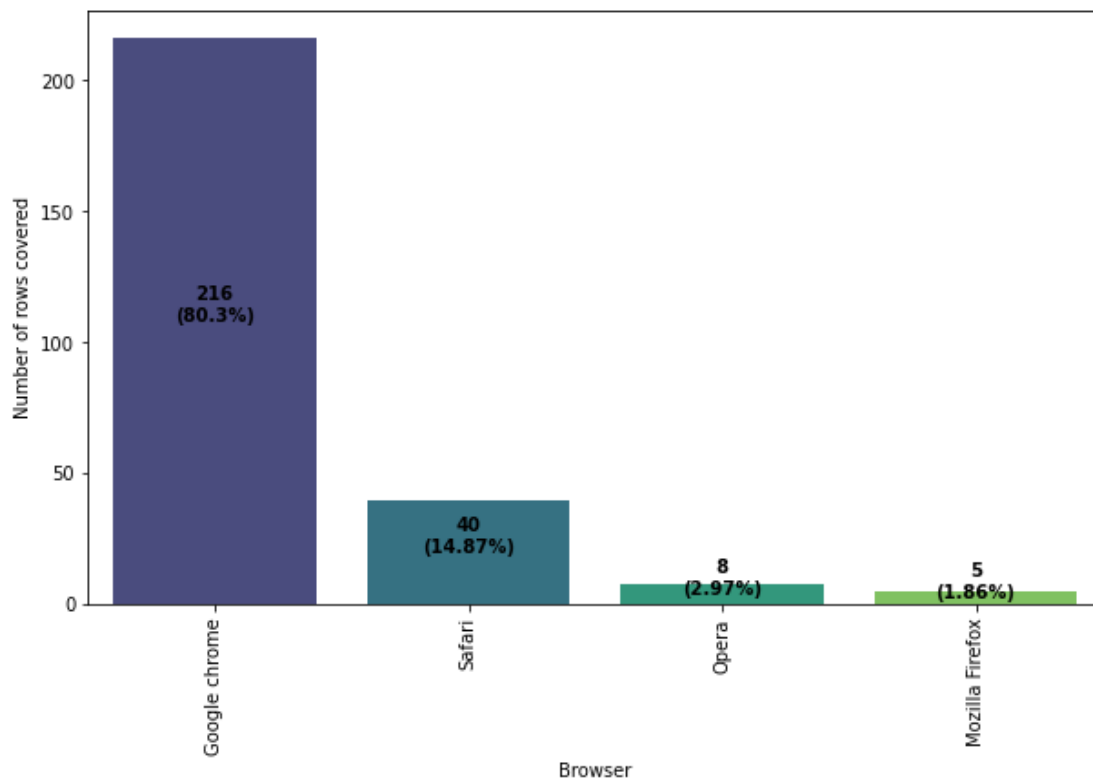




Count Plot for Operating System

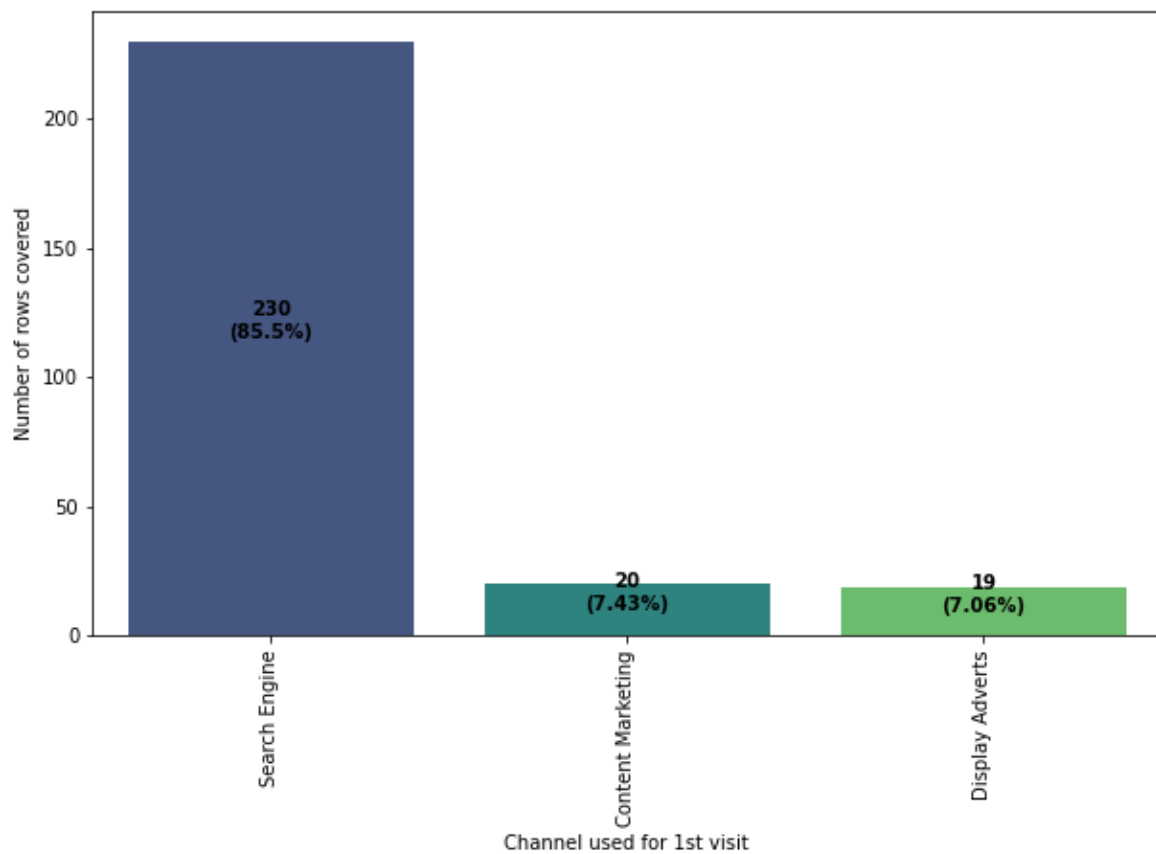


Count Plot for Browser

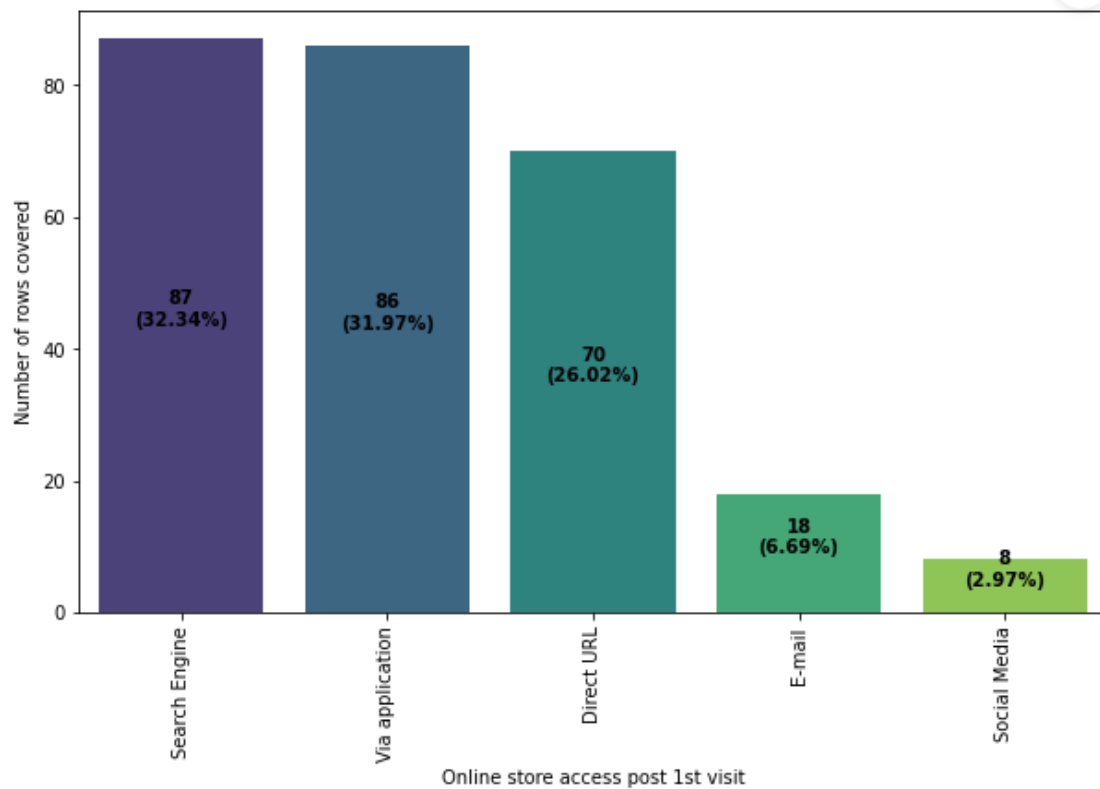




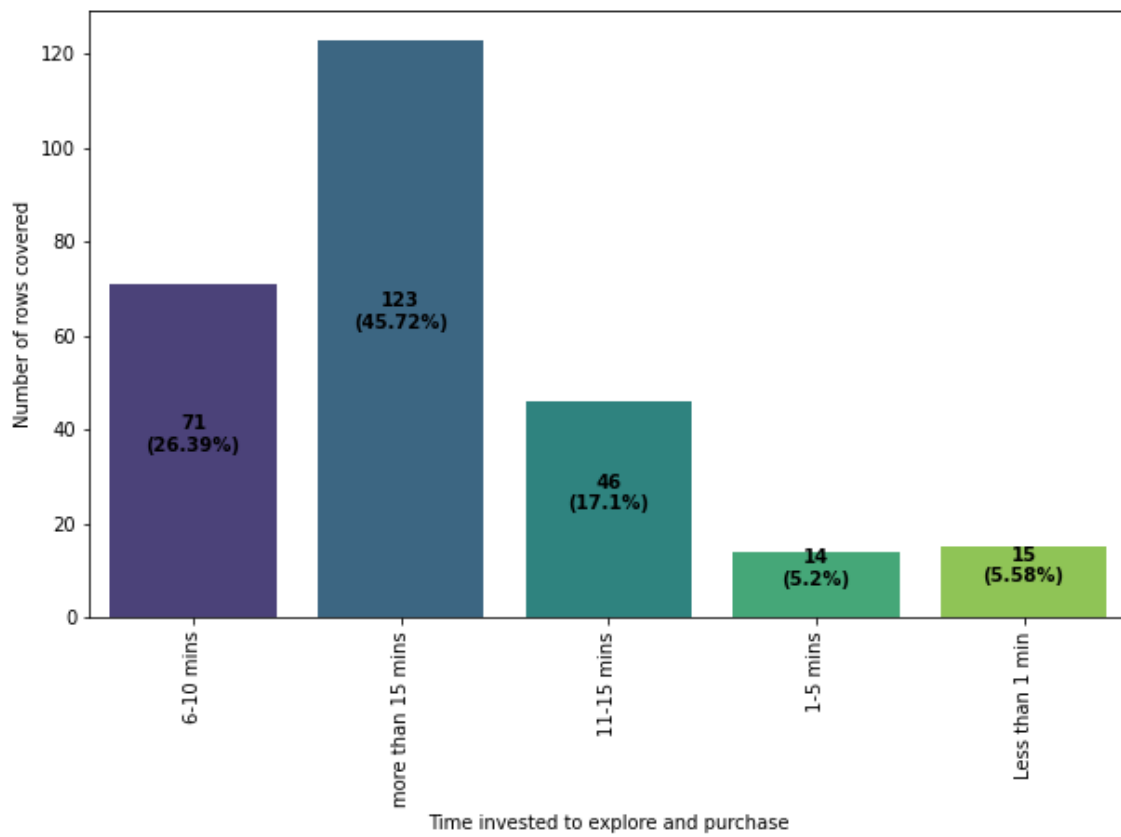
Count Plot for Channel used for 1st visit



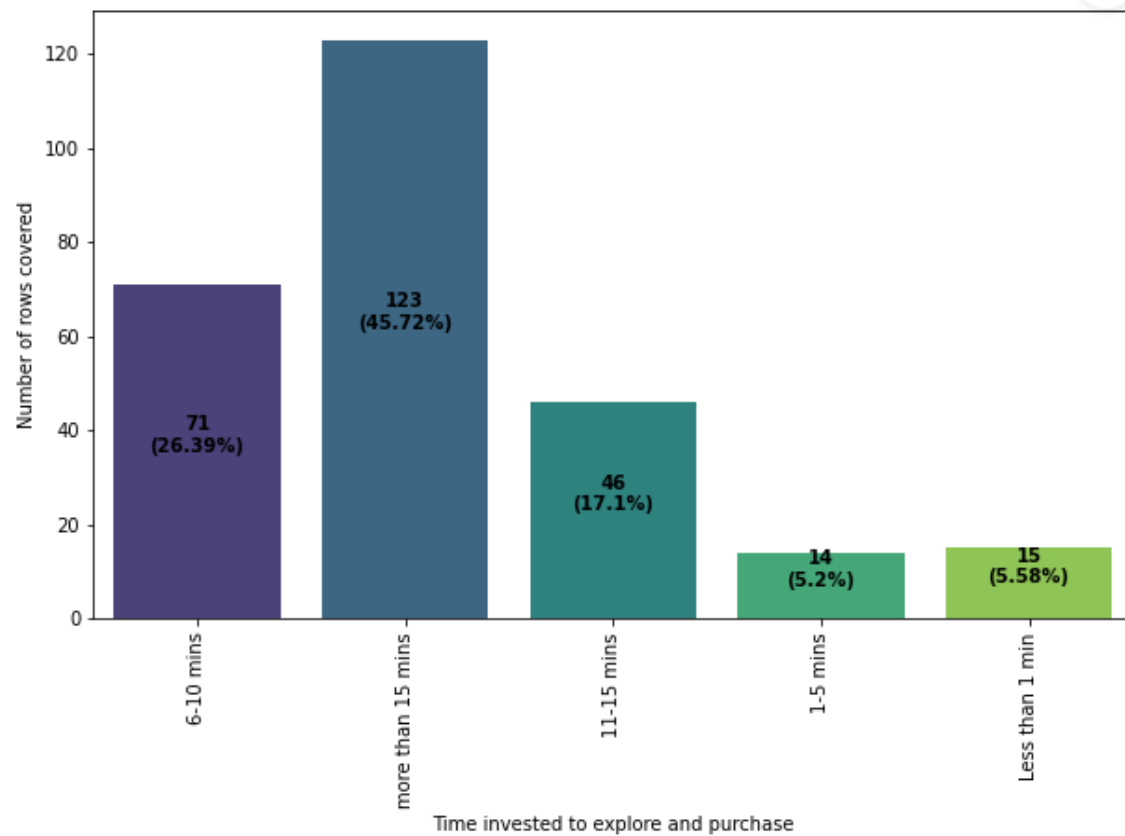
Count Plot for Online store access post 1st visit

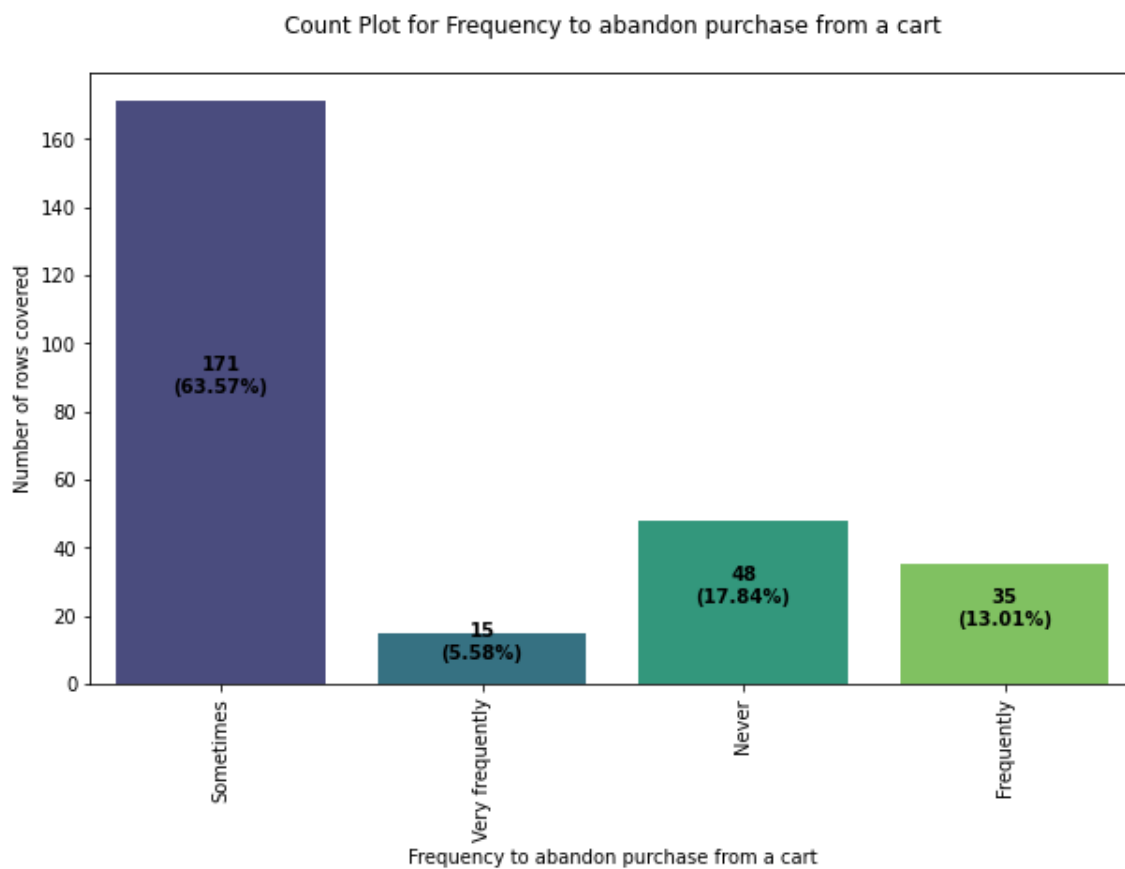
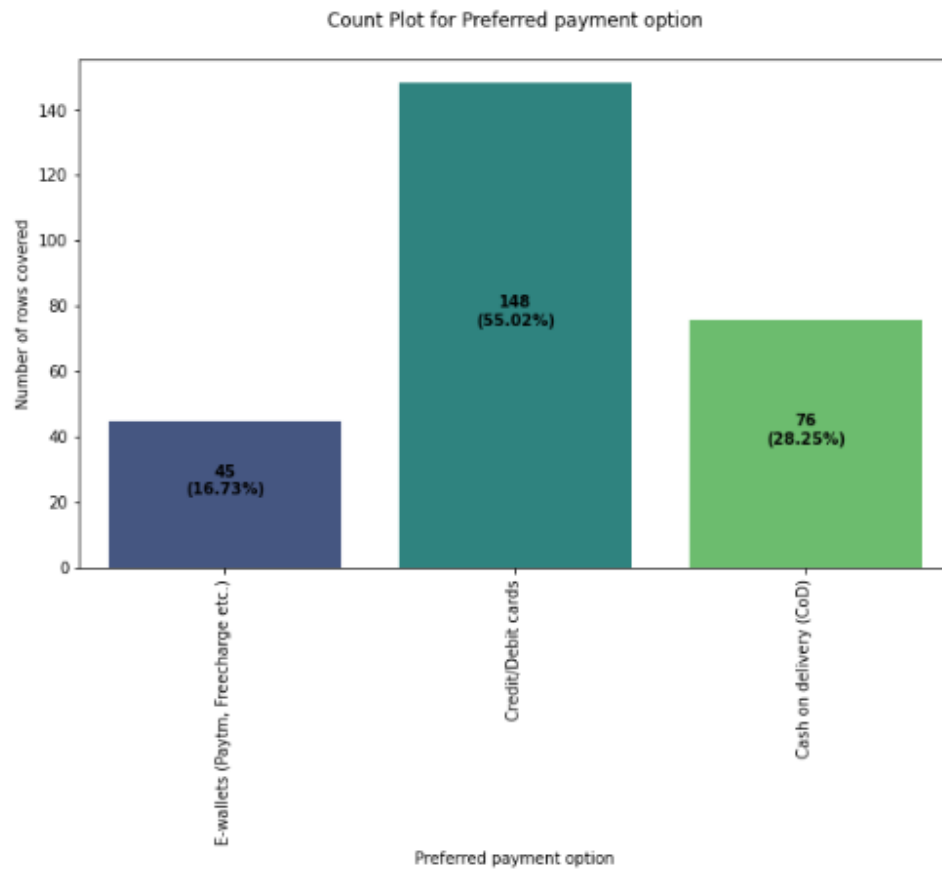


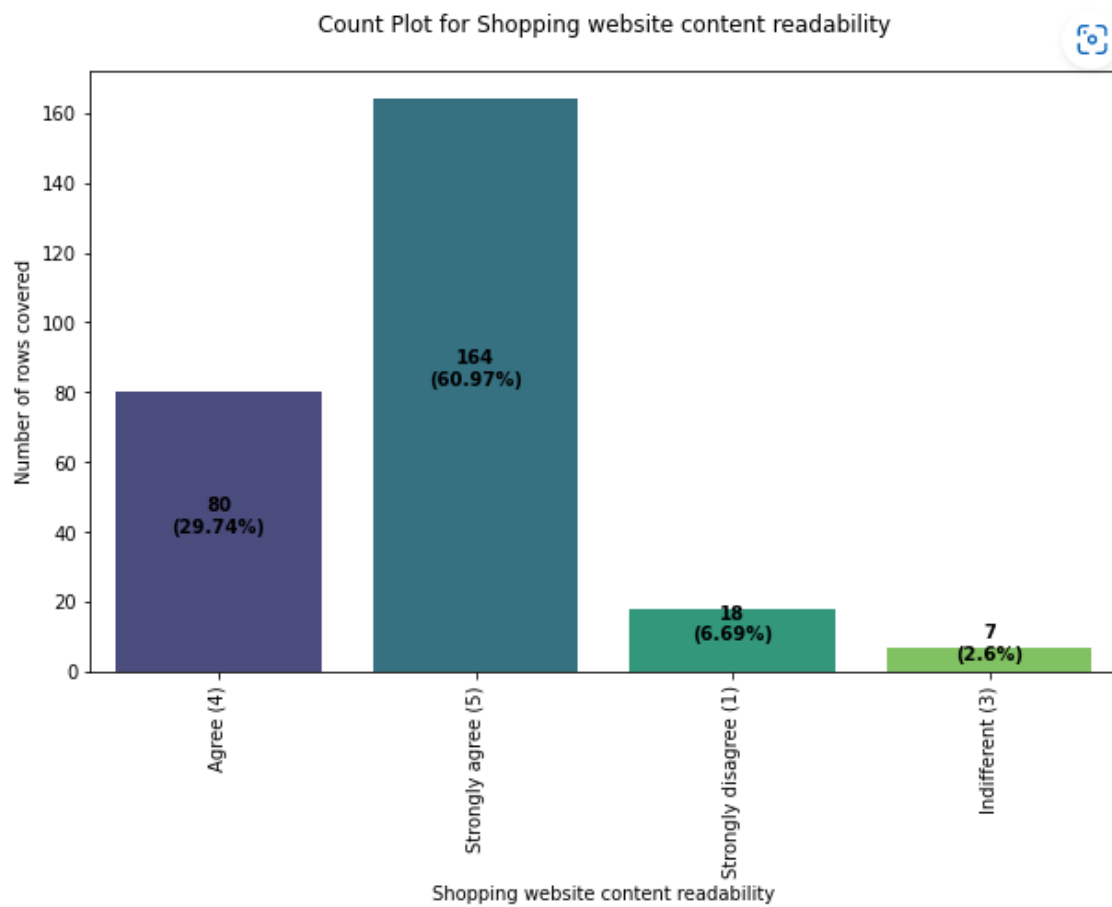
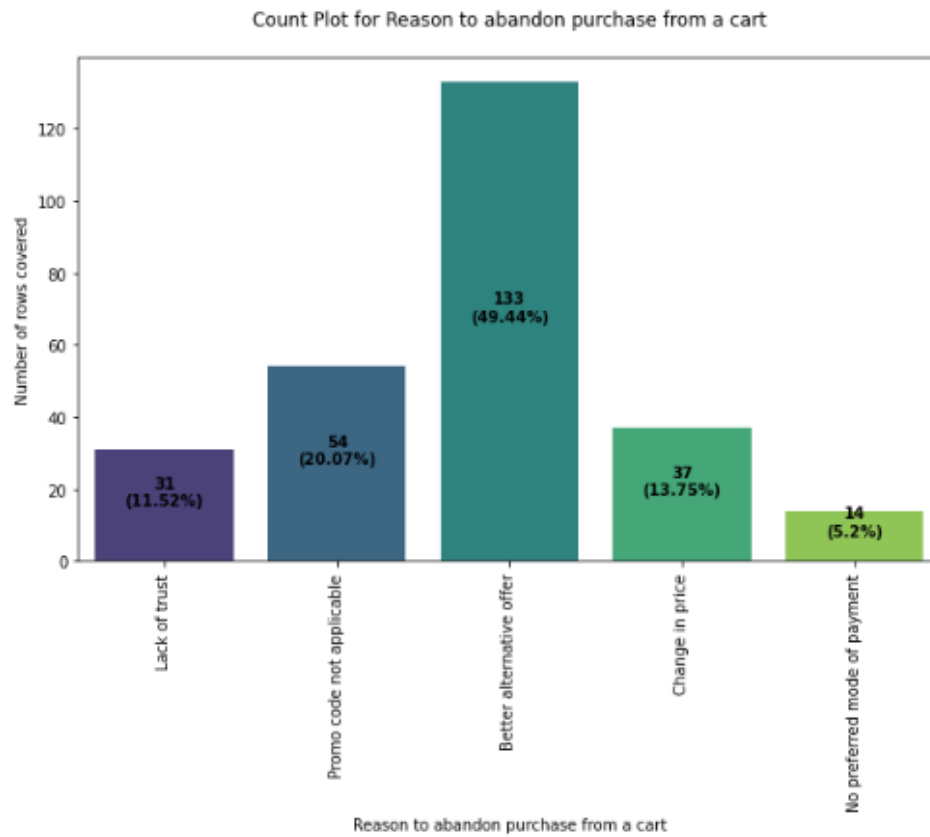
Count Plot for Time invested to explore and purchase

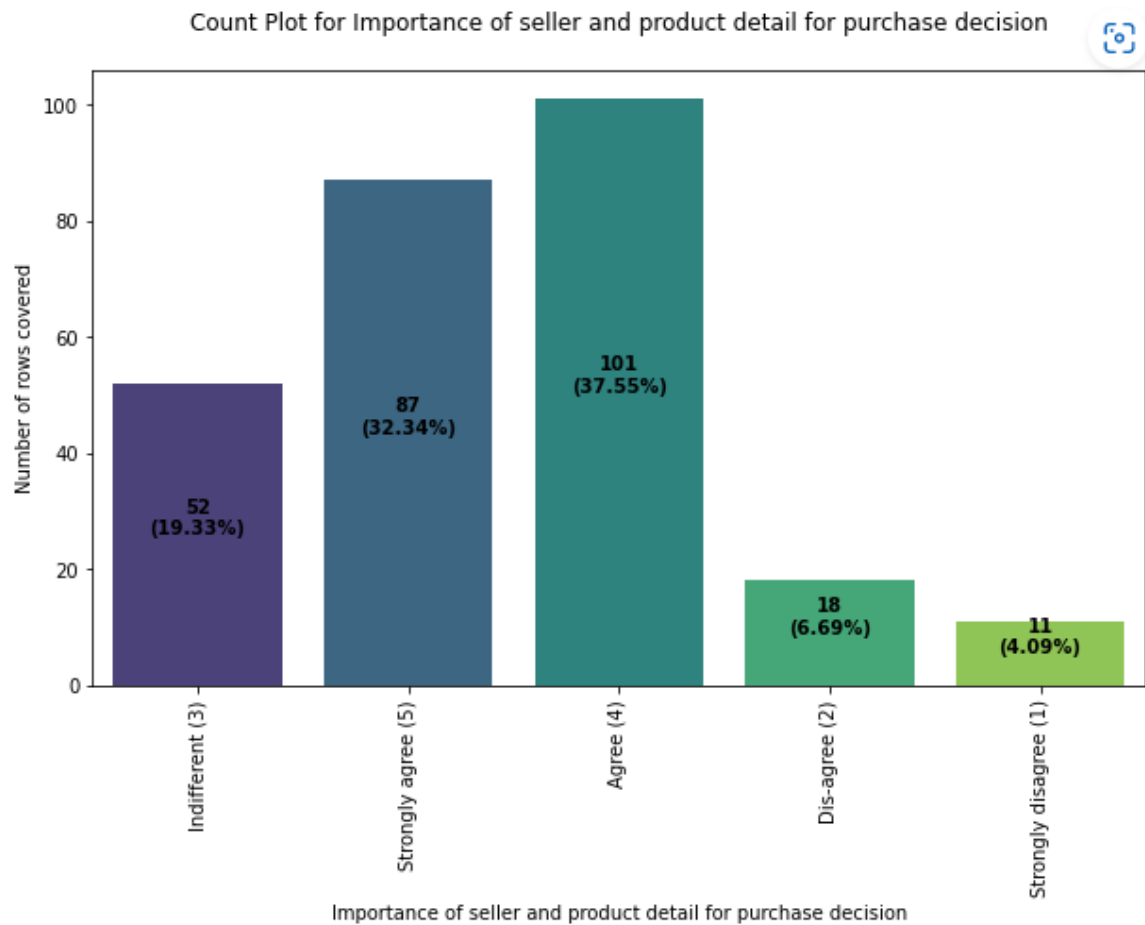
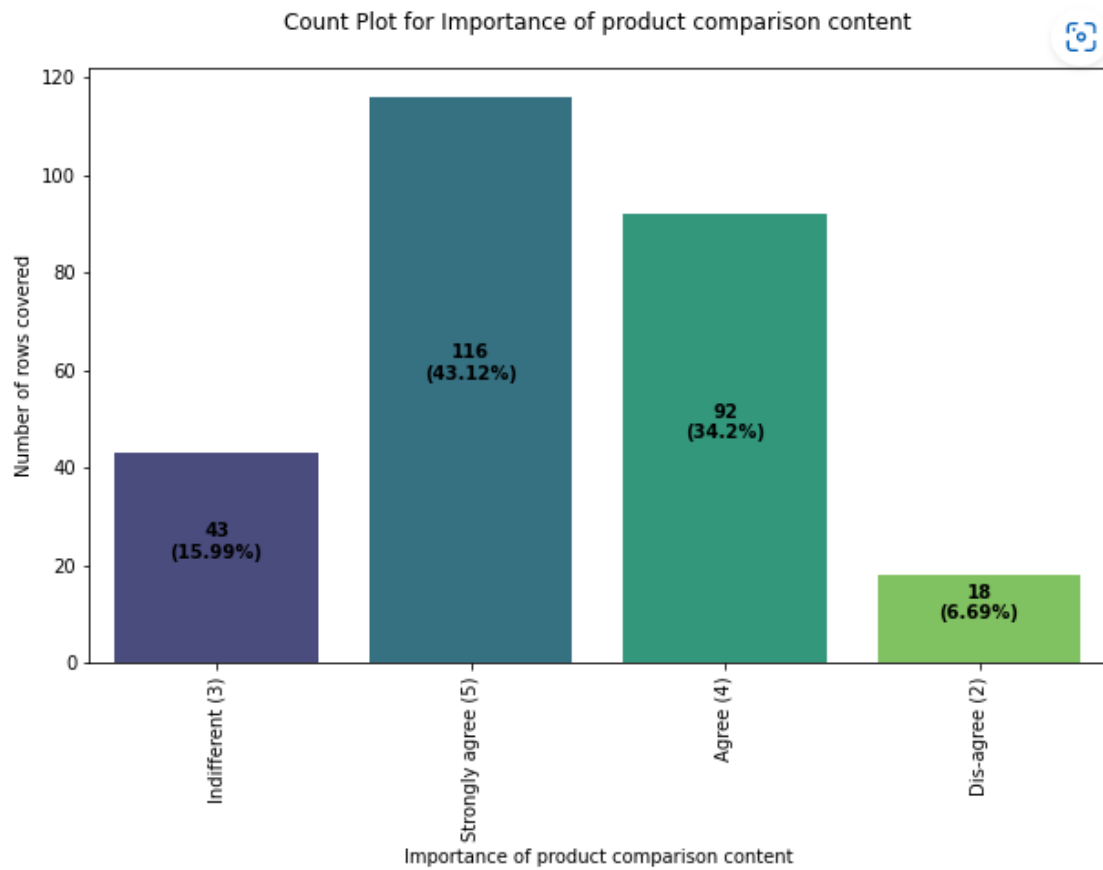


Count Plot for Time invested to explore and purchase

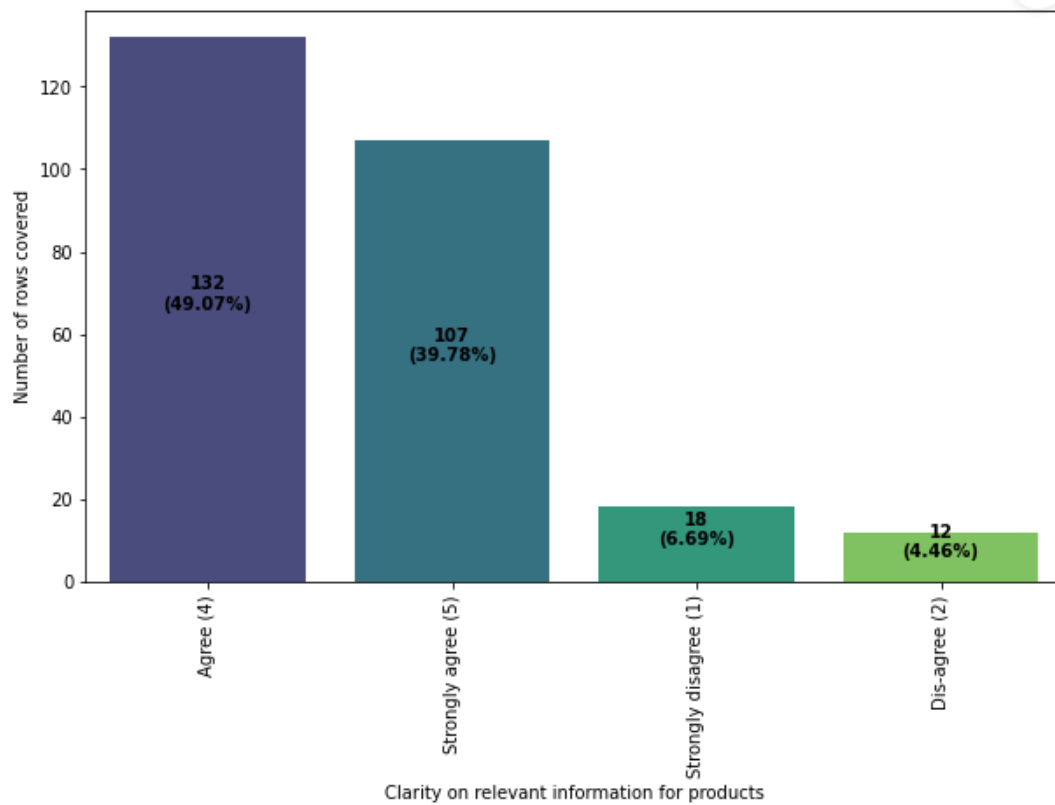




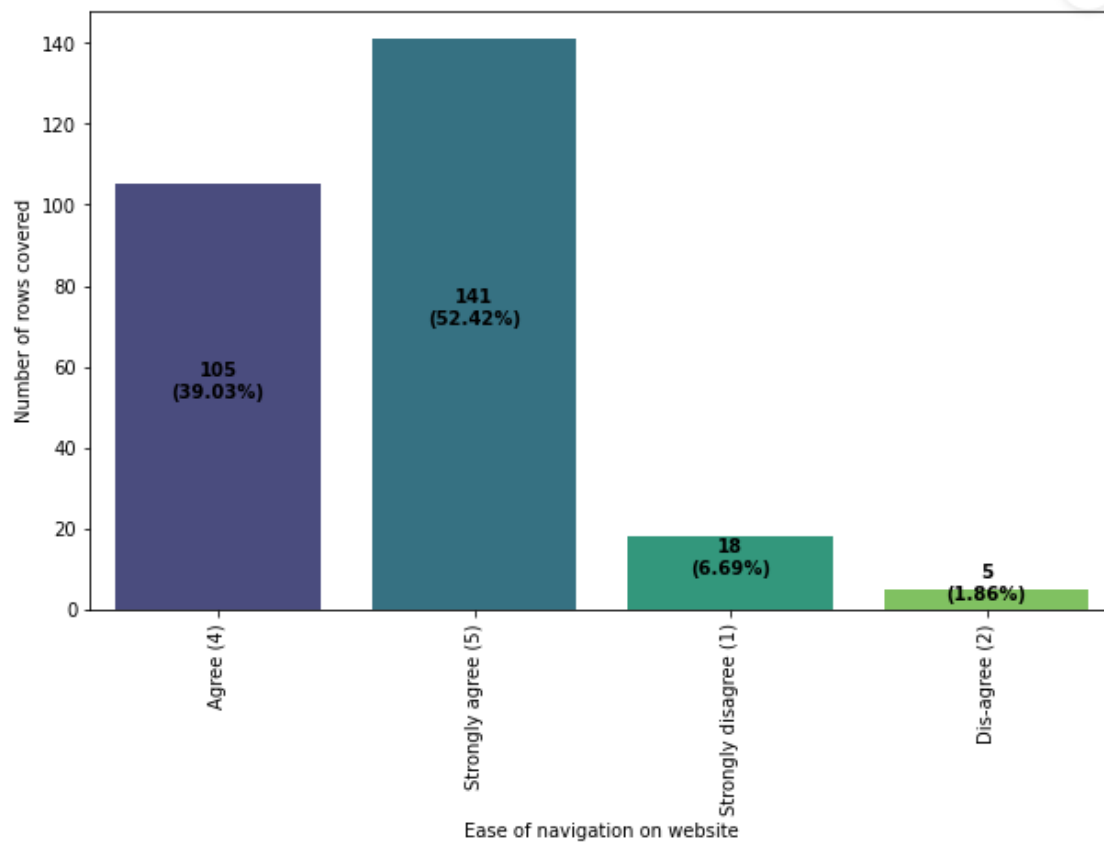




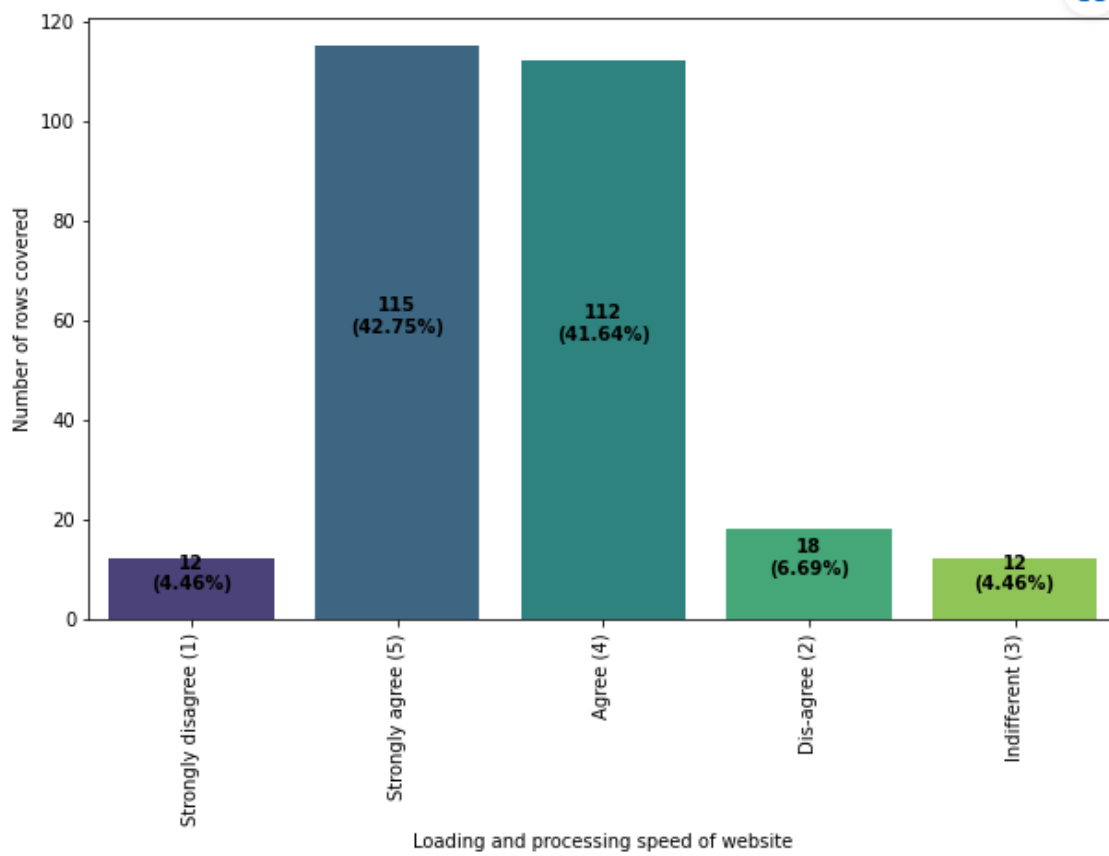
Count Plot for Clarity on relevant information for products



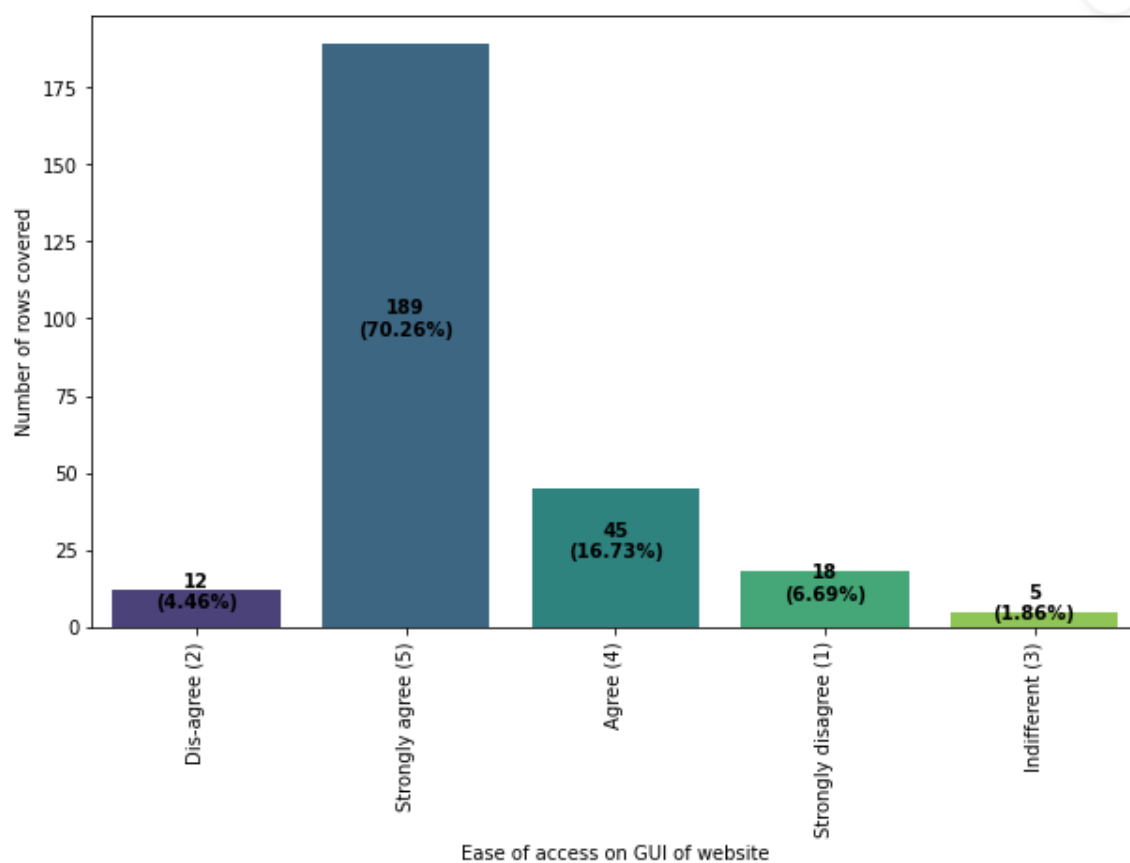
Count Plot for Ease of navigation on website



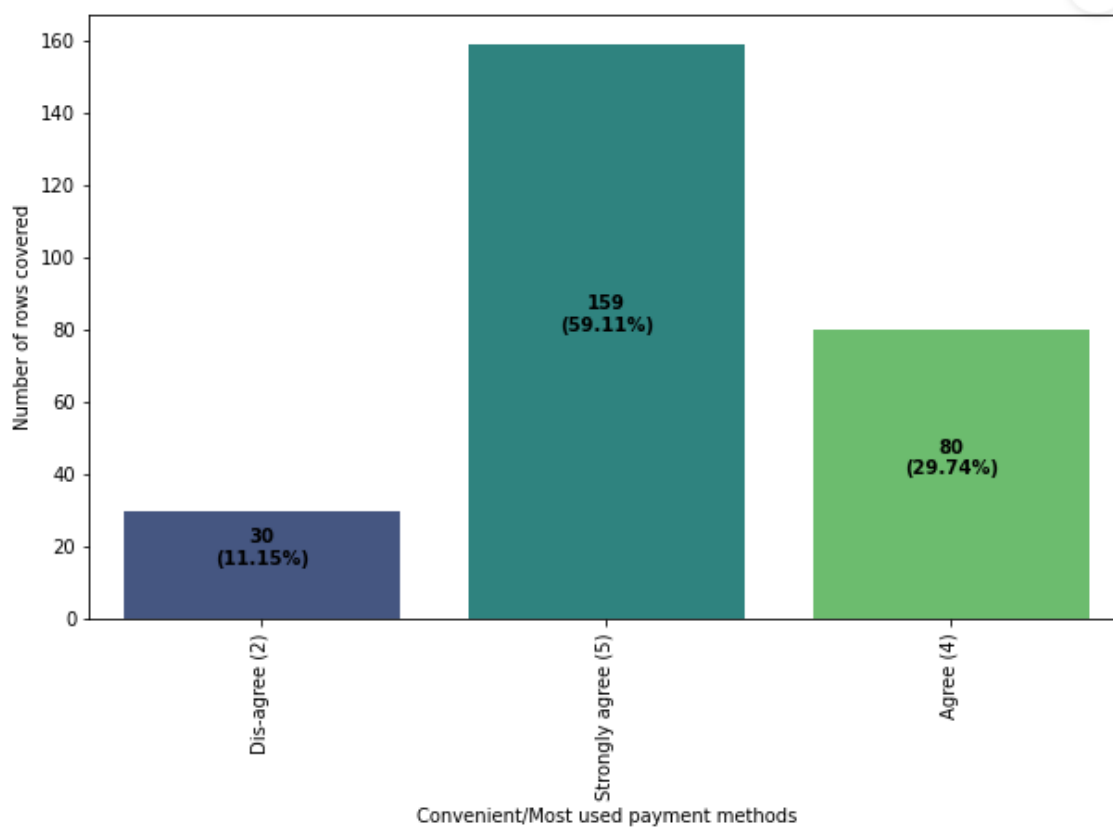
Count Plot for Loading and processing speed of website



Count Plot for Ease of access on GUI of website

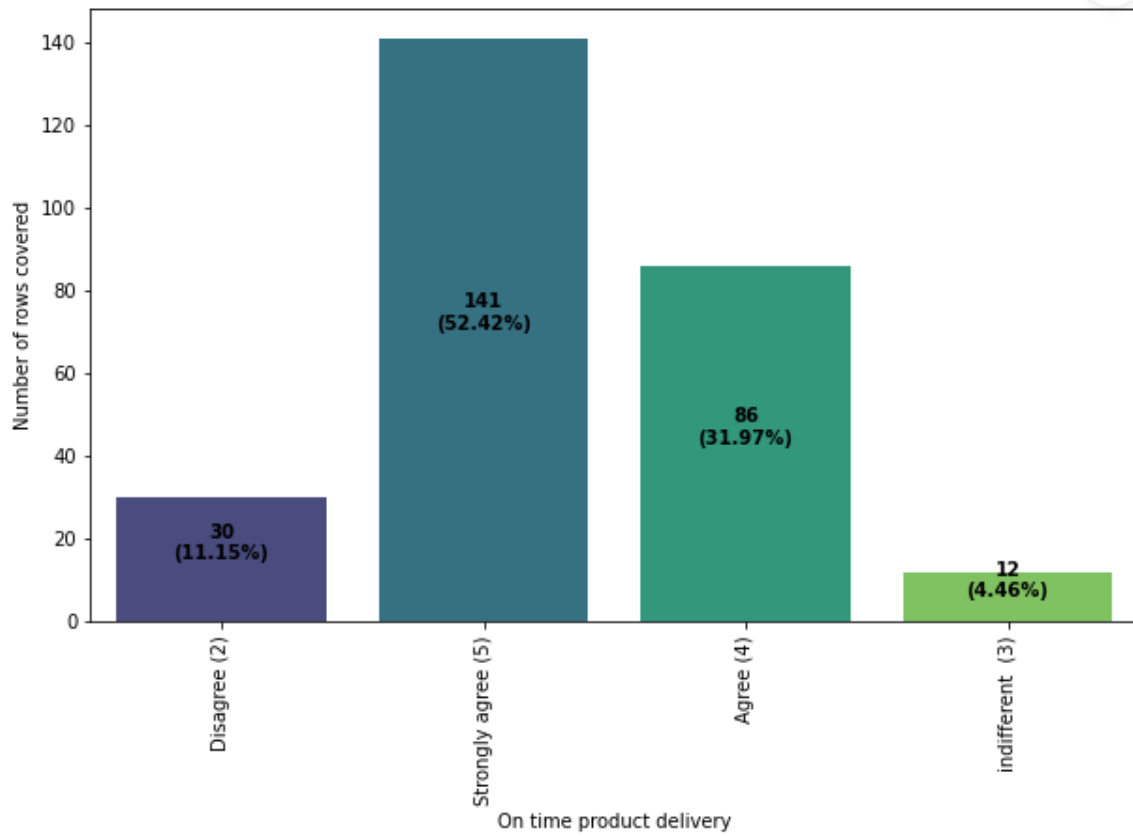


Count Plot for Convenient/Most used payment methods

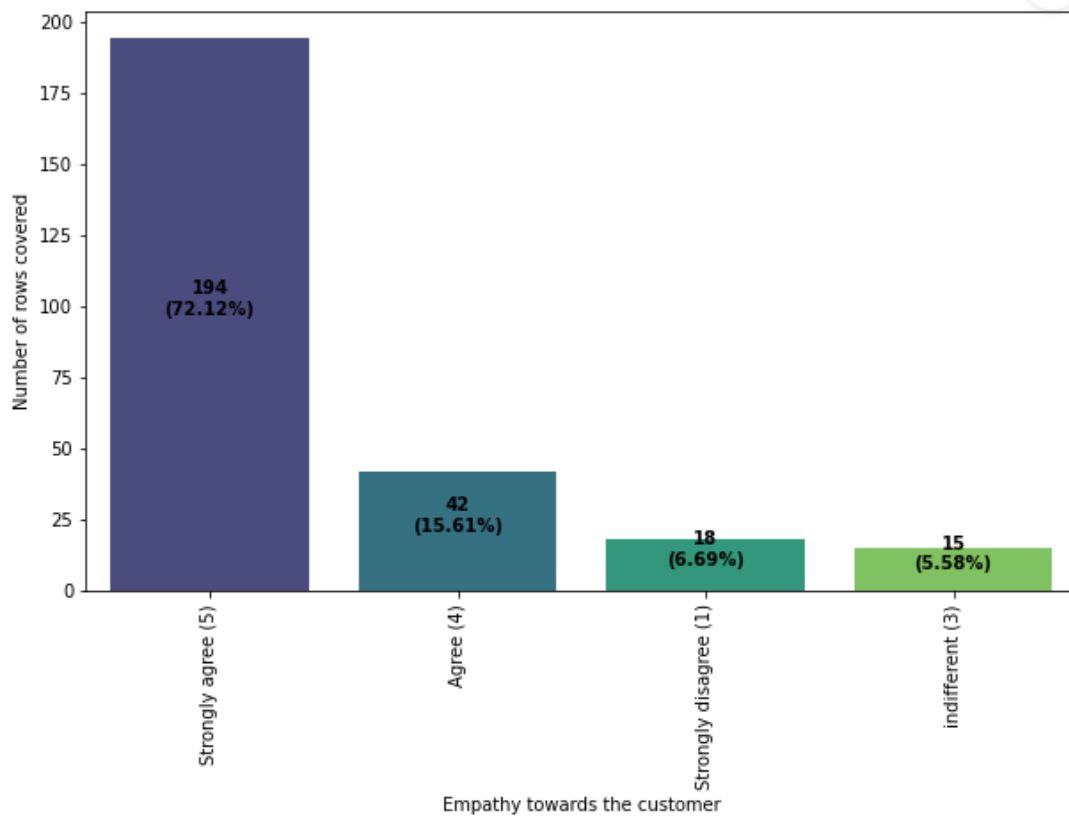




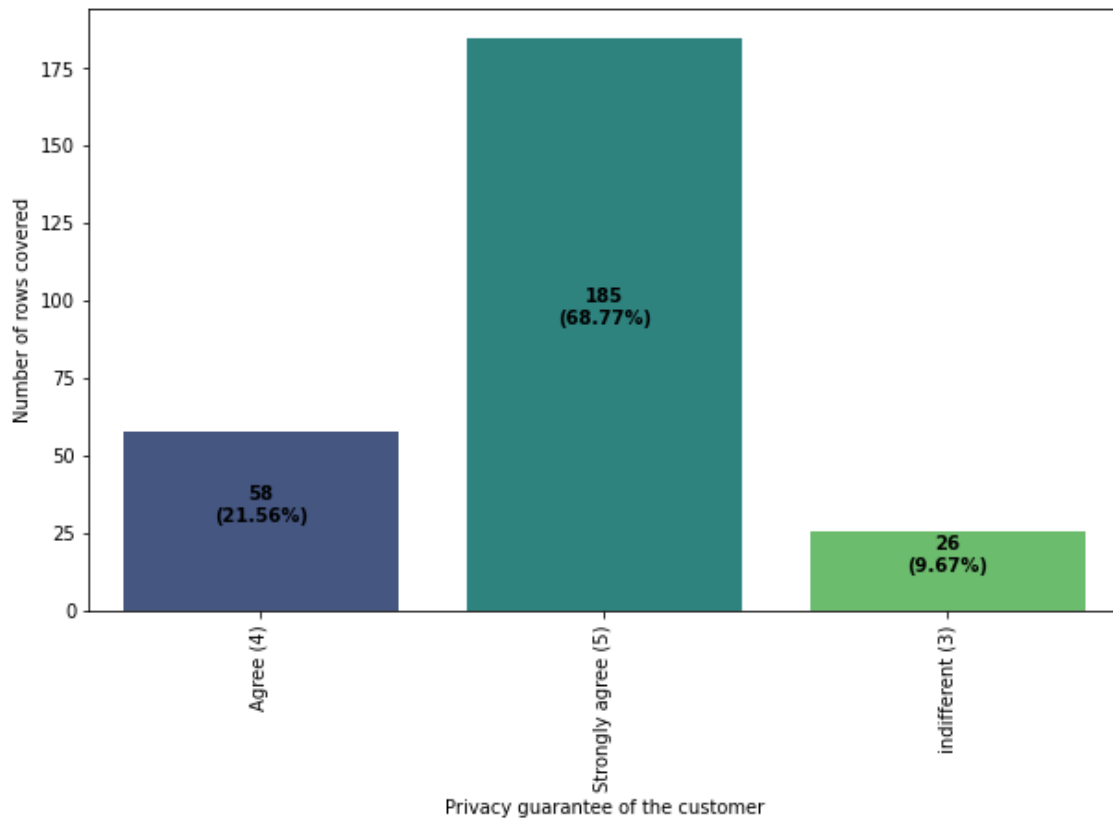
Count Plot for On time product delivery



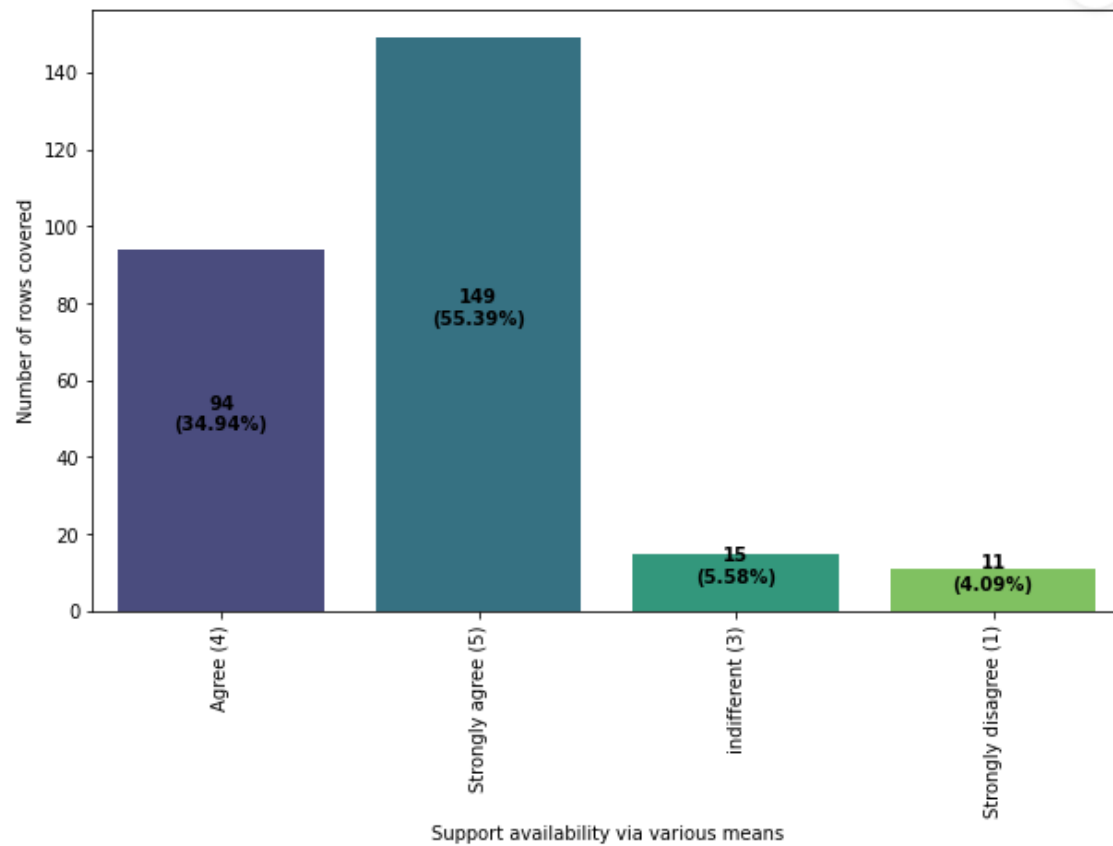
Count Plot for Empathy towards the customer



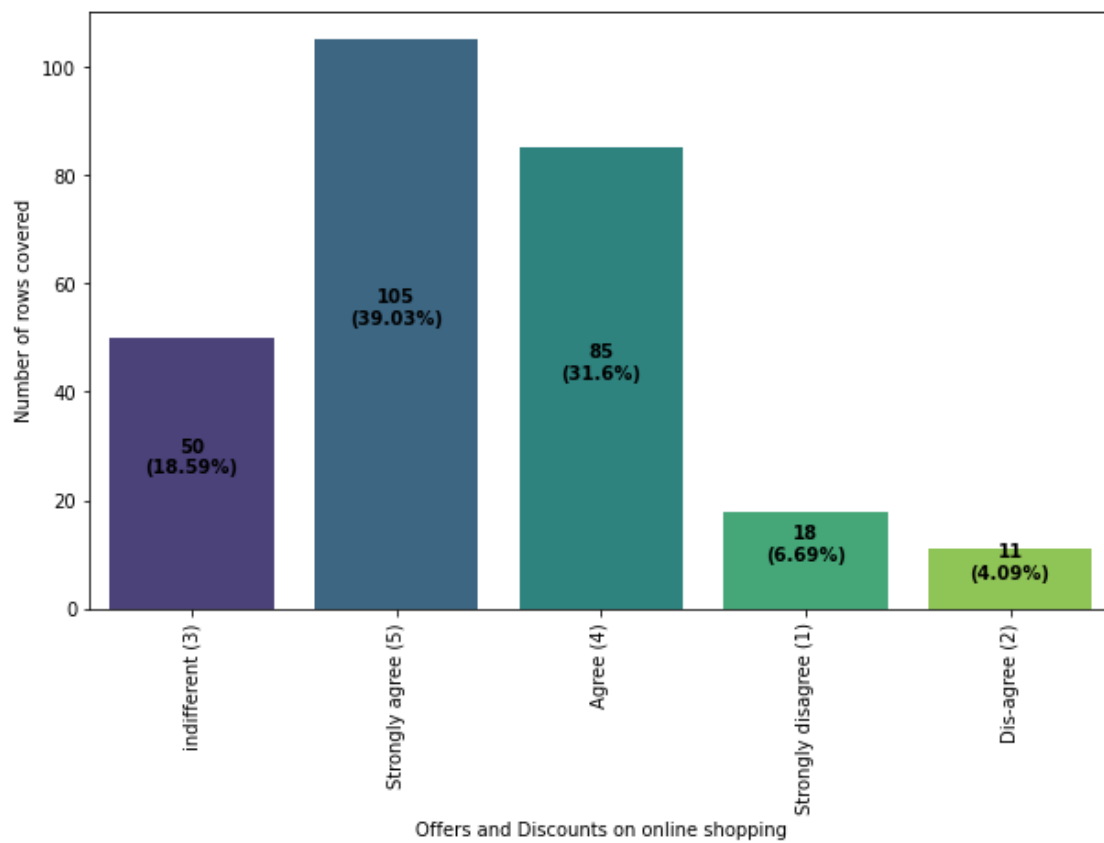
Count Plot for Privacy guarantee of the customer



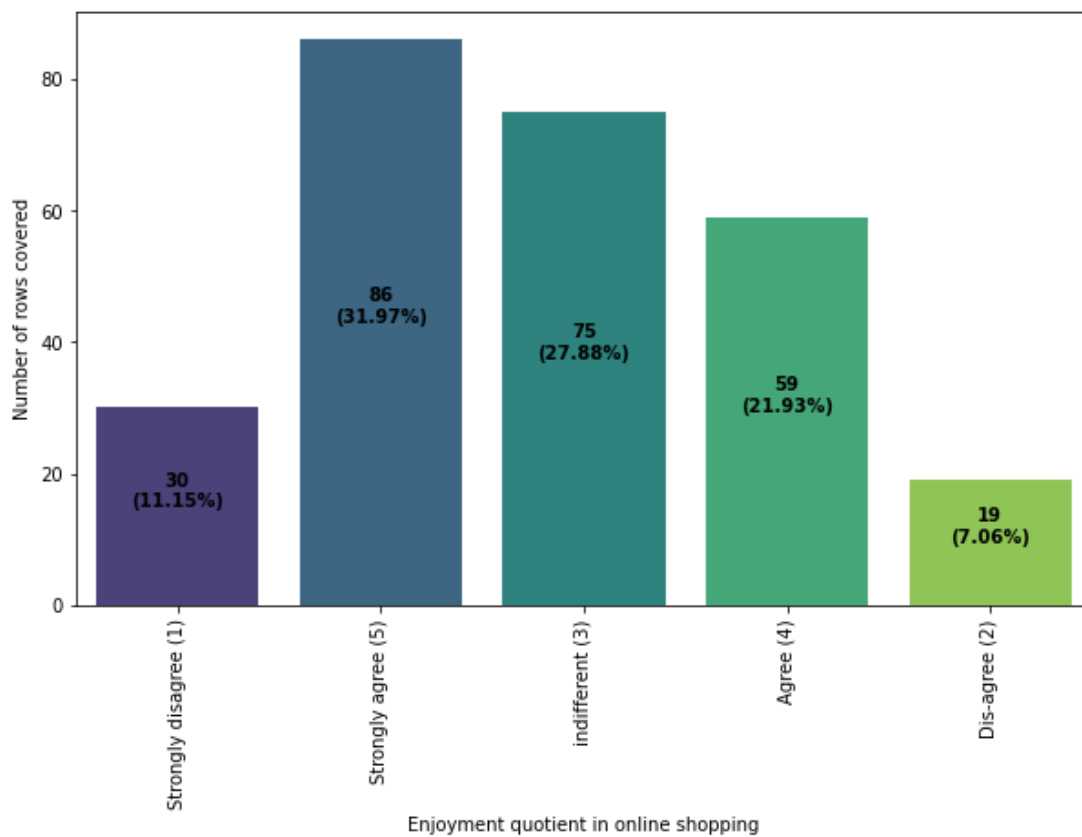
Count Plot for Support availability via various means



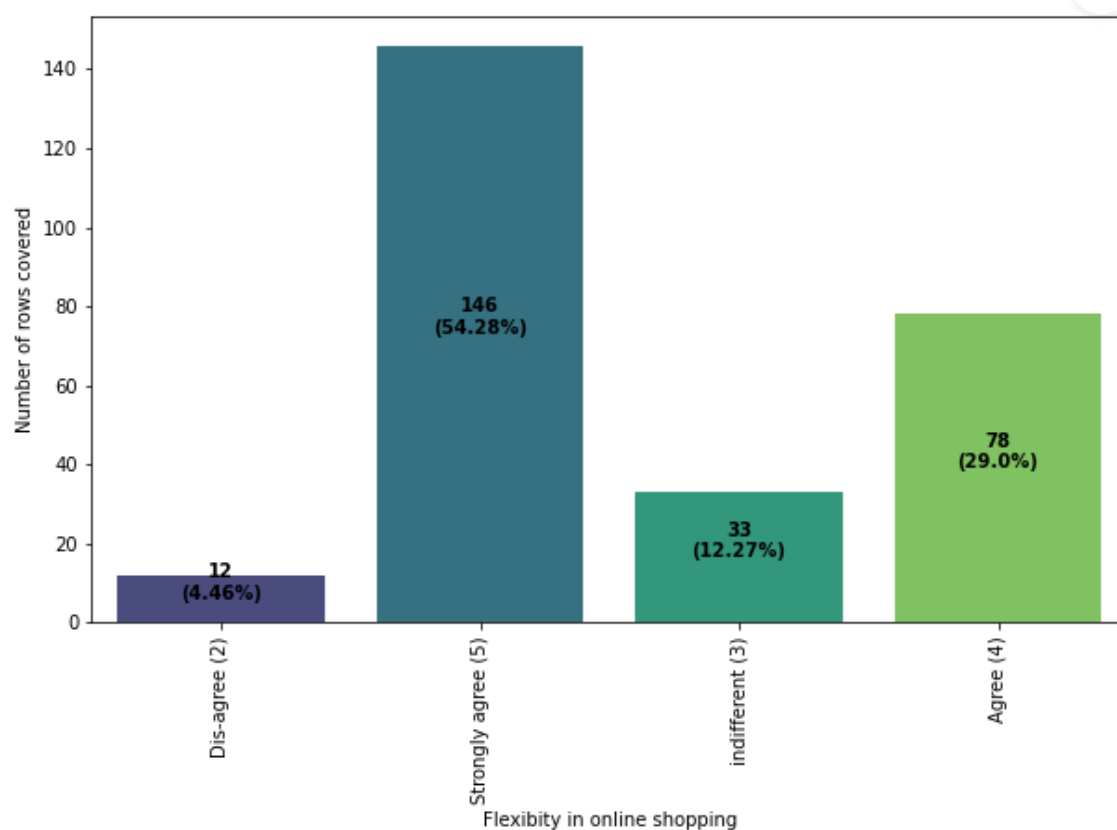
Count Plot for Offers and Discounts on online shopping



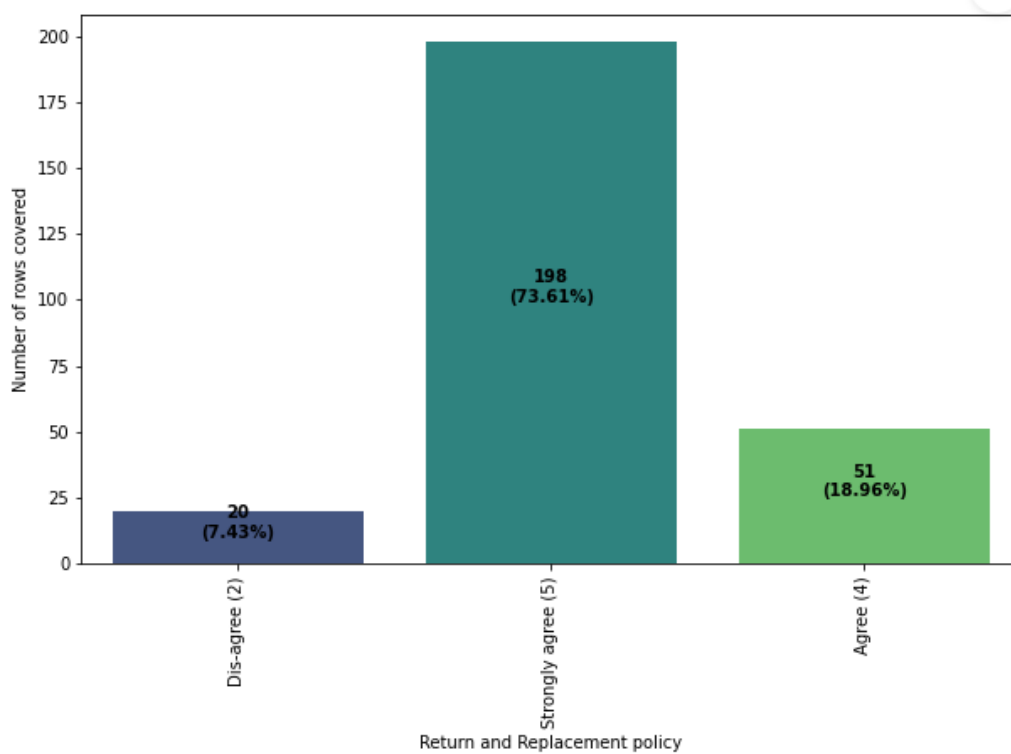
Count Plot for Enjoyment quotient in online shopping

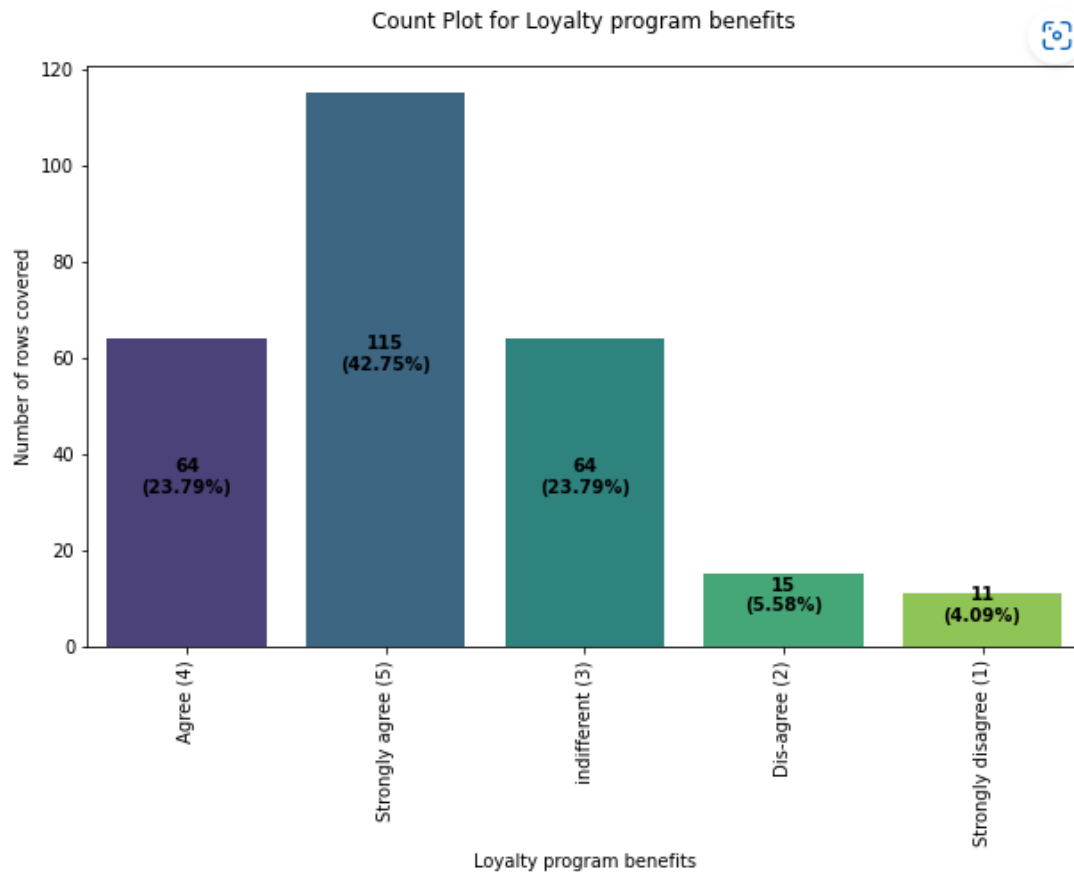


Count Plot for Flexibity in online shopping

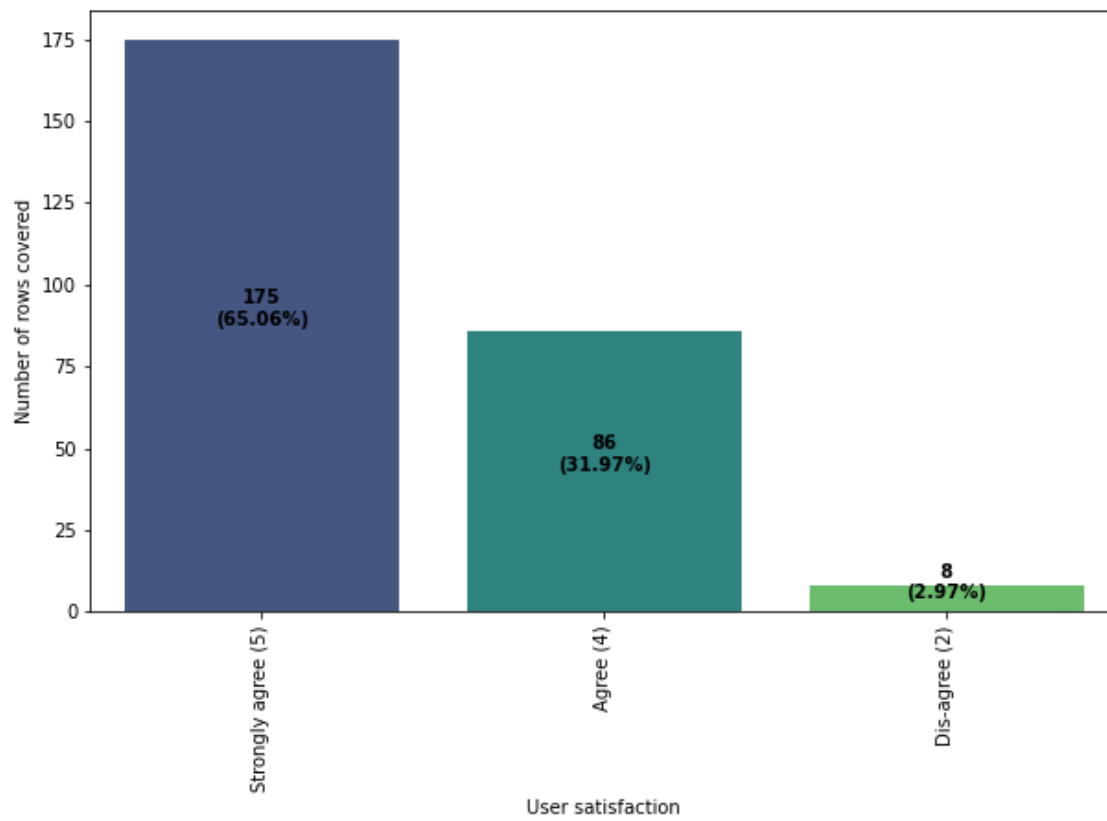


Count Plot for Return and Replacement policy

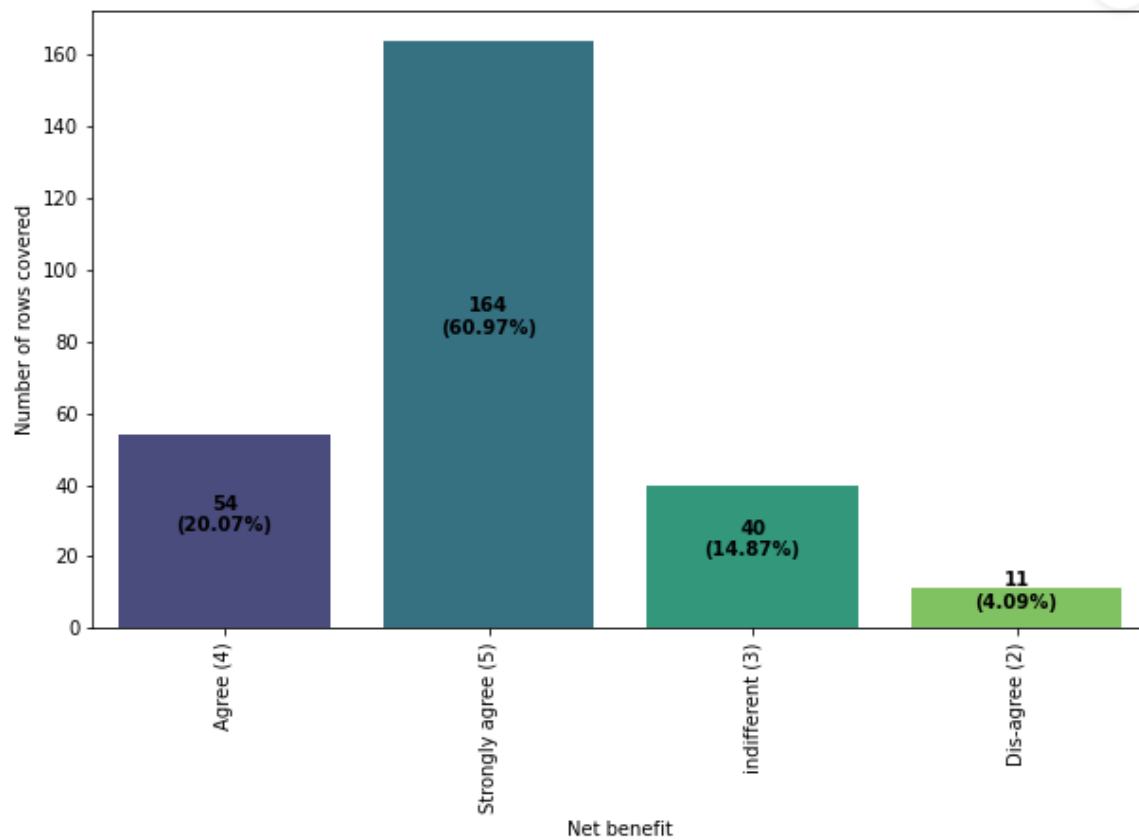




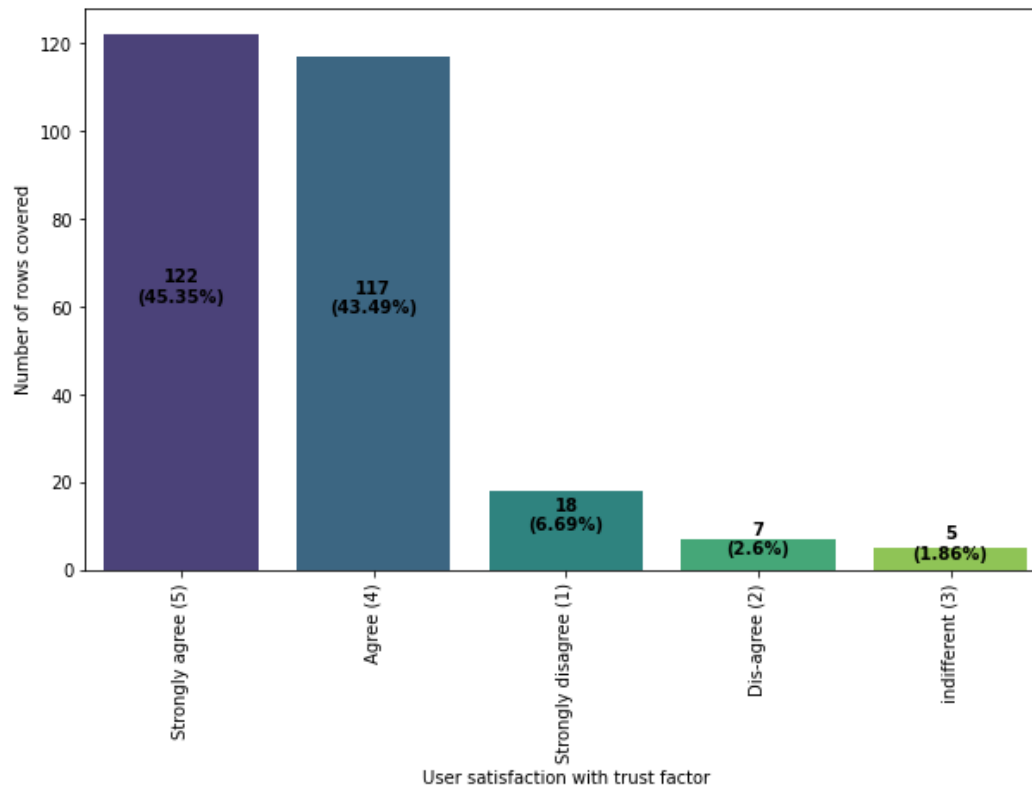
Count Plot for User satisfaction



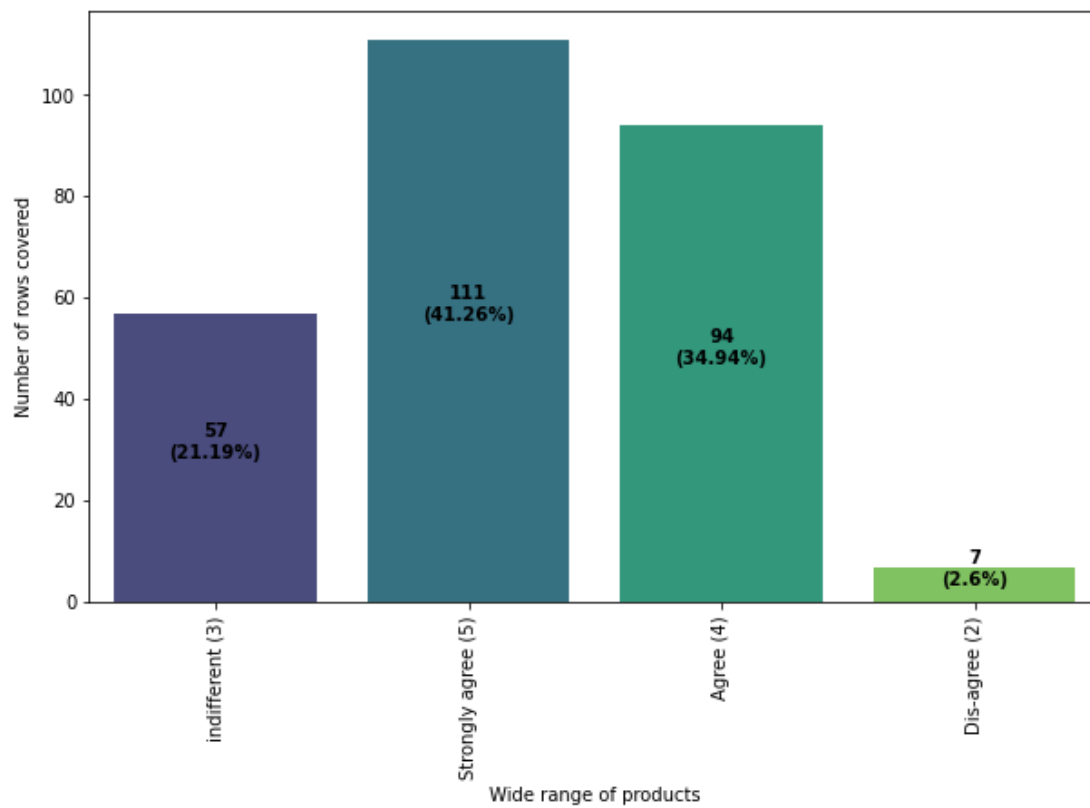
Count Plot for Net benefit



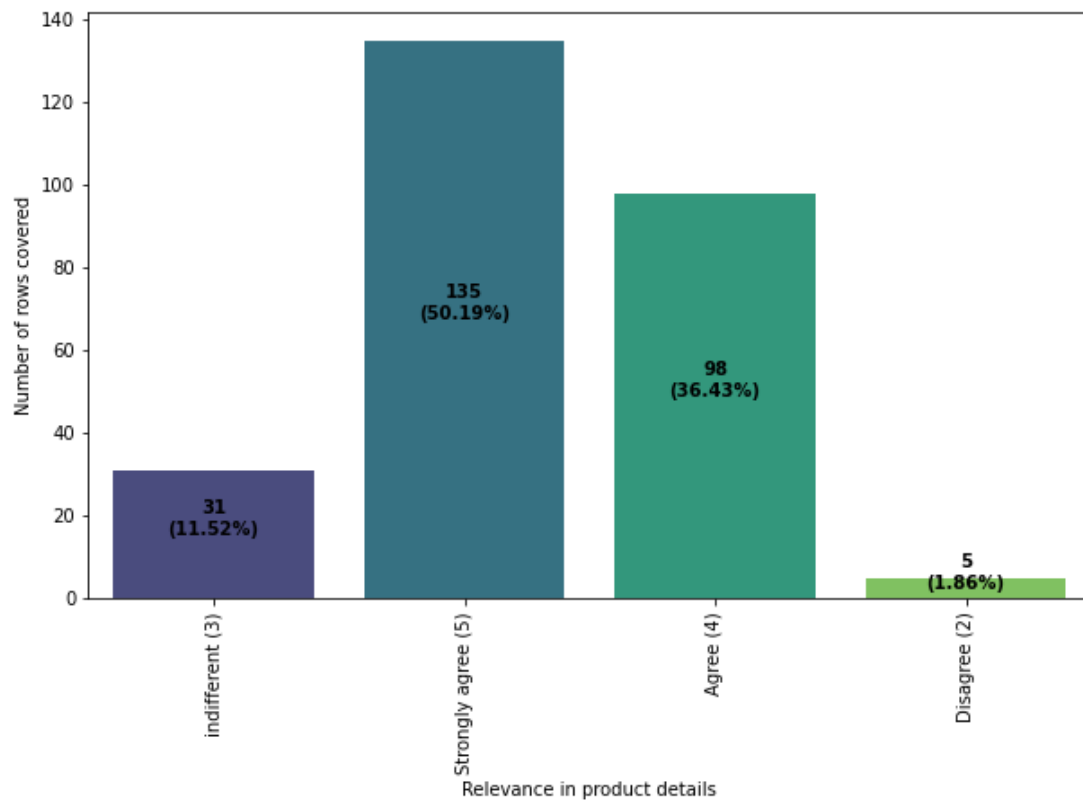
Count Plot for User satisfaction with trust factor



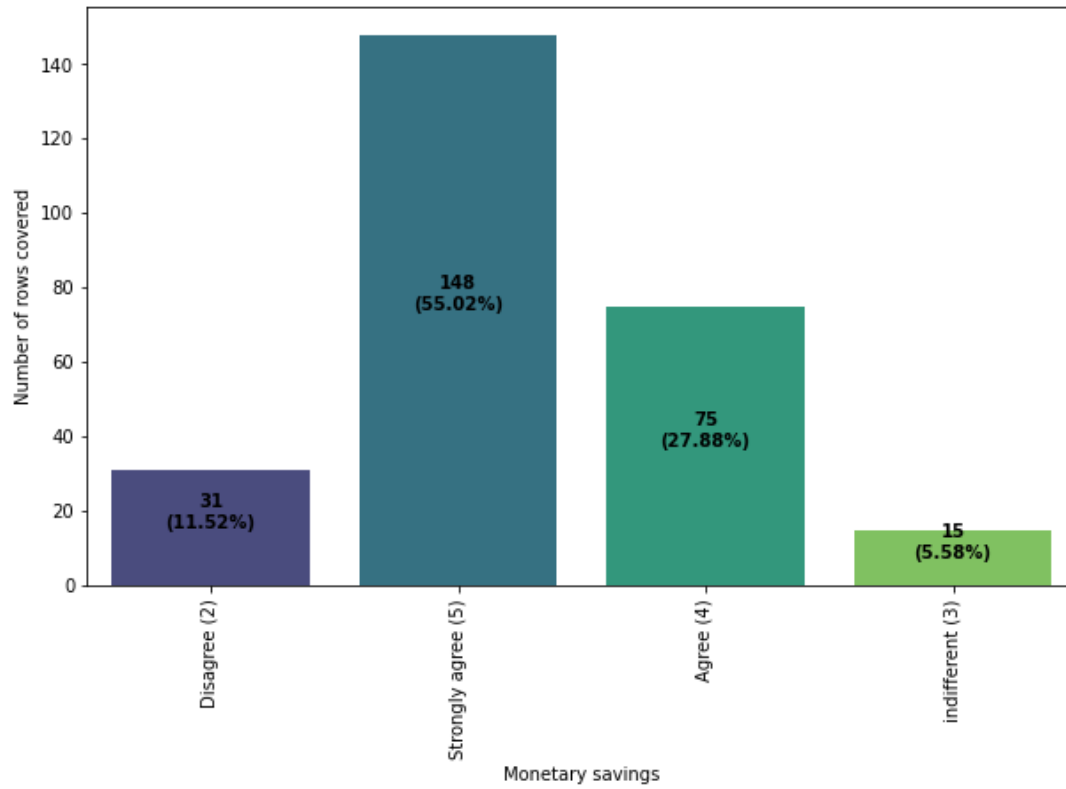
Count Plot for Wide range of products



Count Plot for Relevance in product details

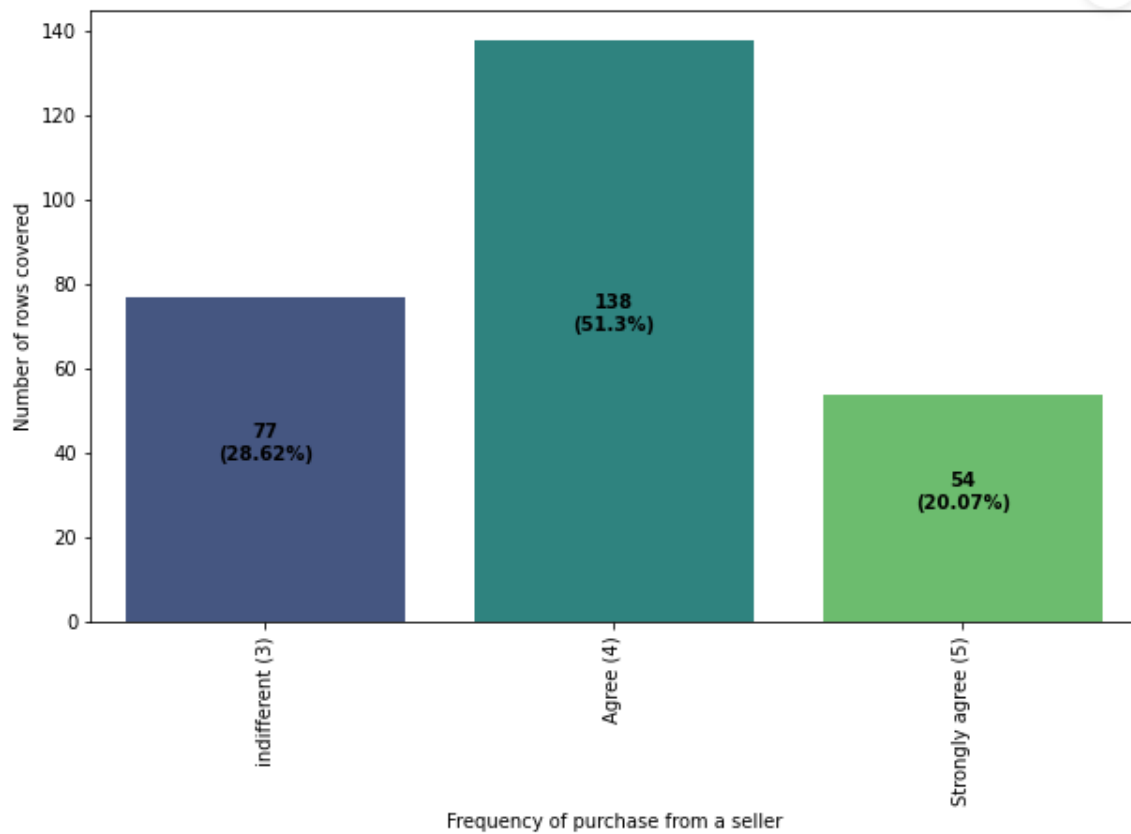


Count Plot for Monetary savings

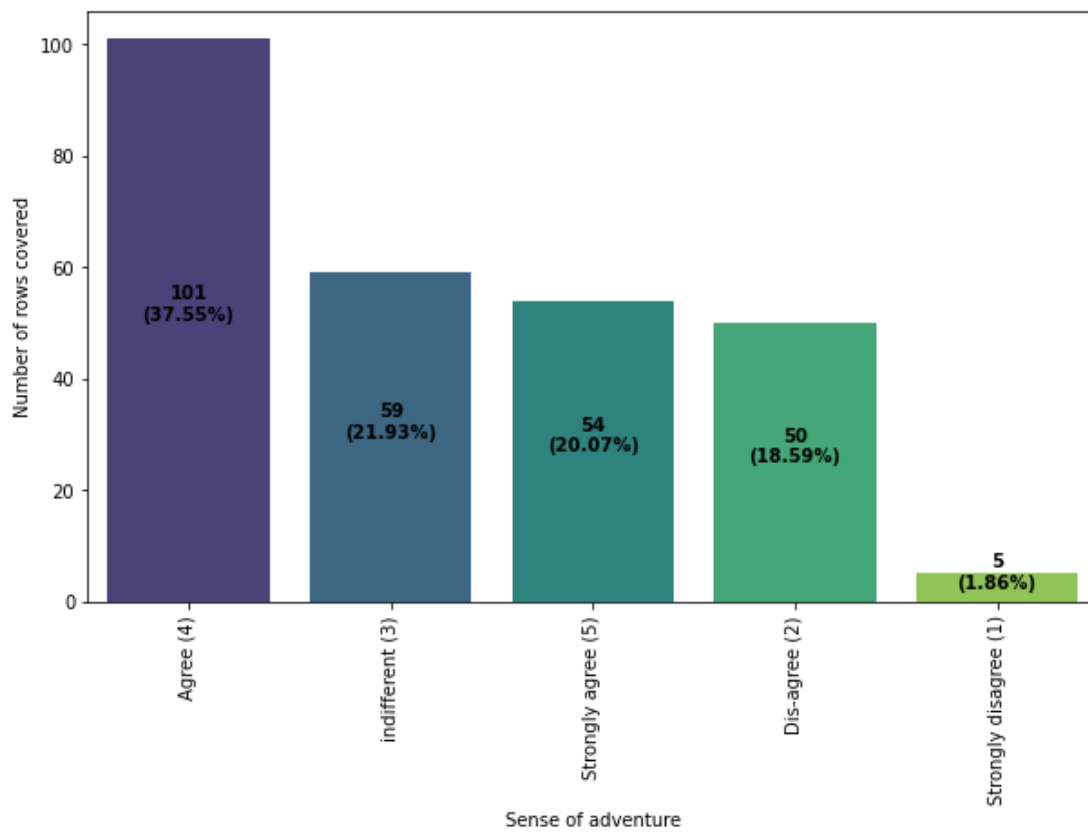


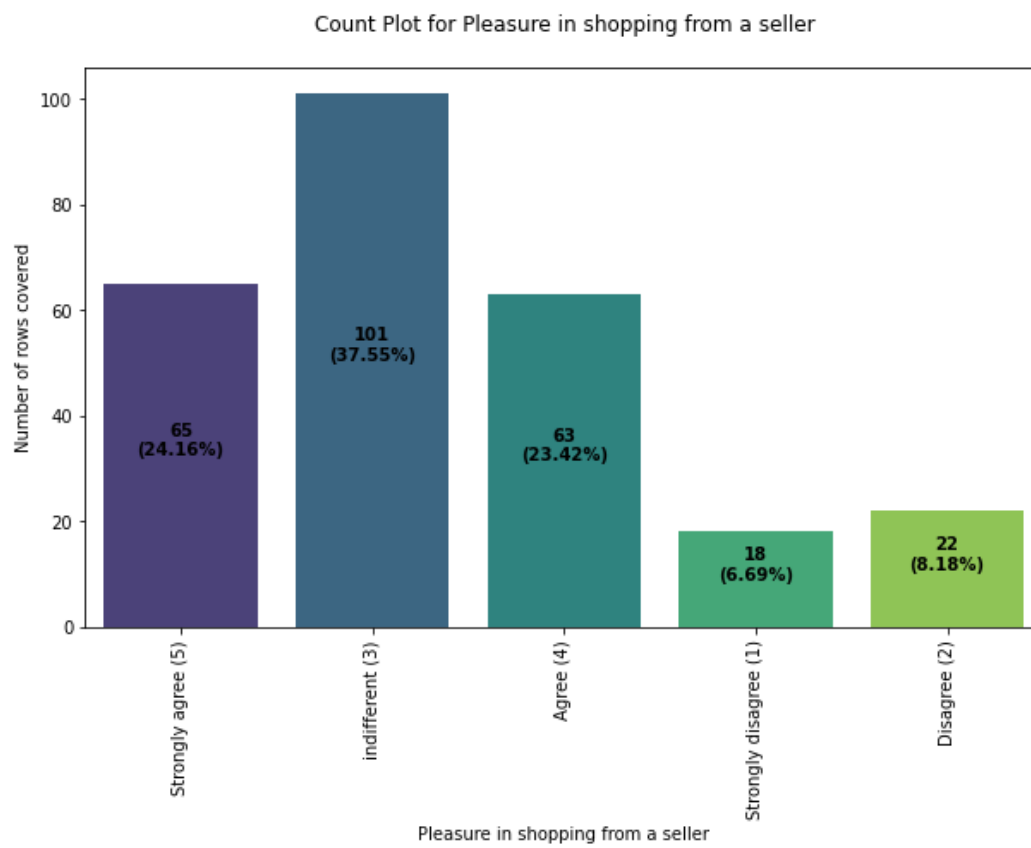
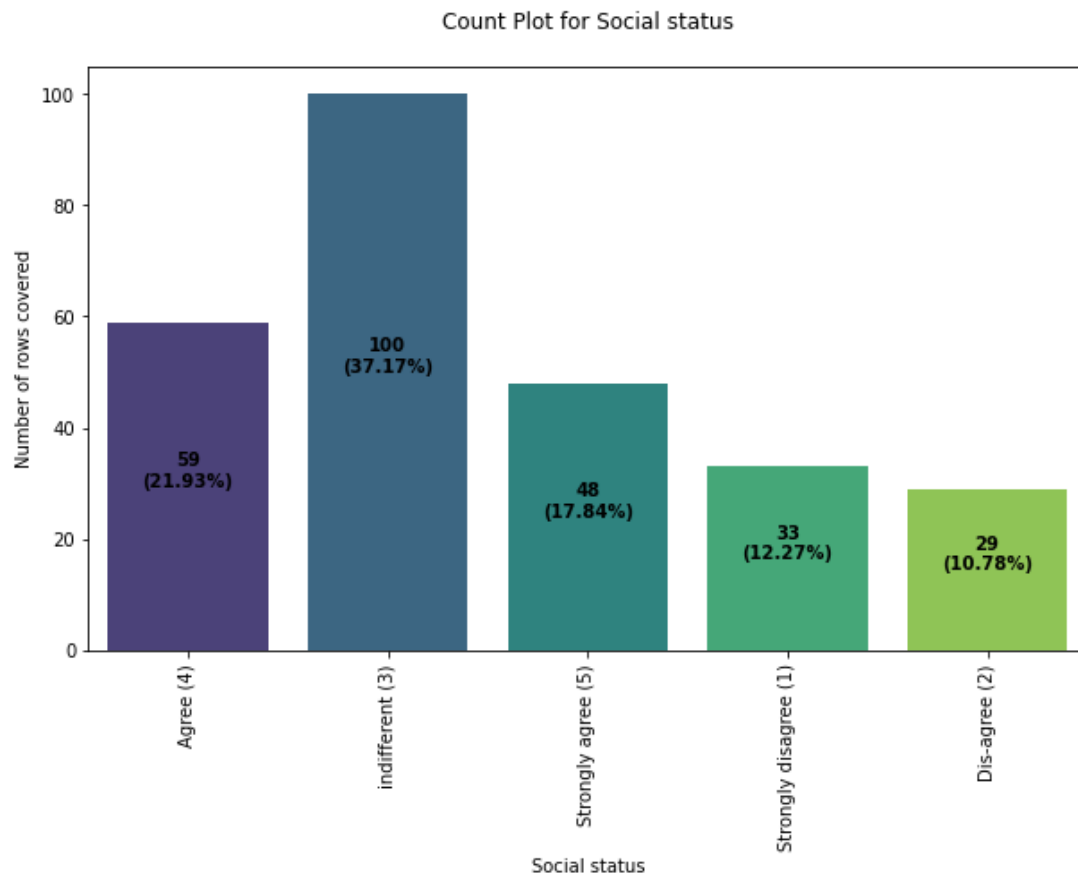


Count Plot for Frequency of purchase from a seller

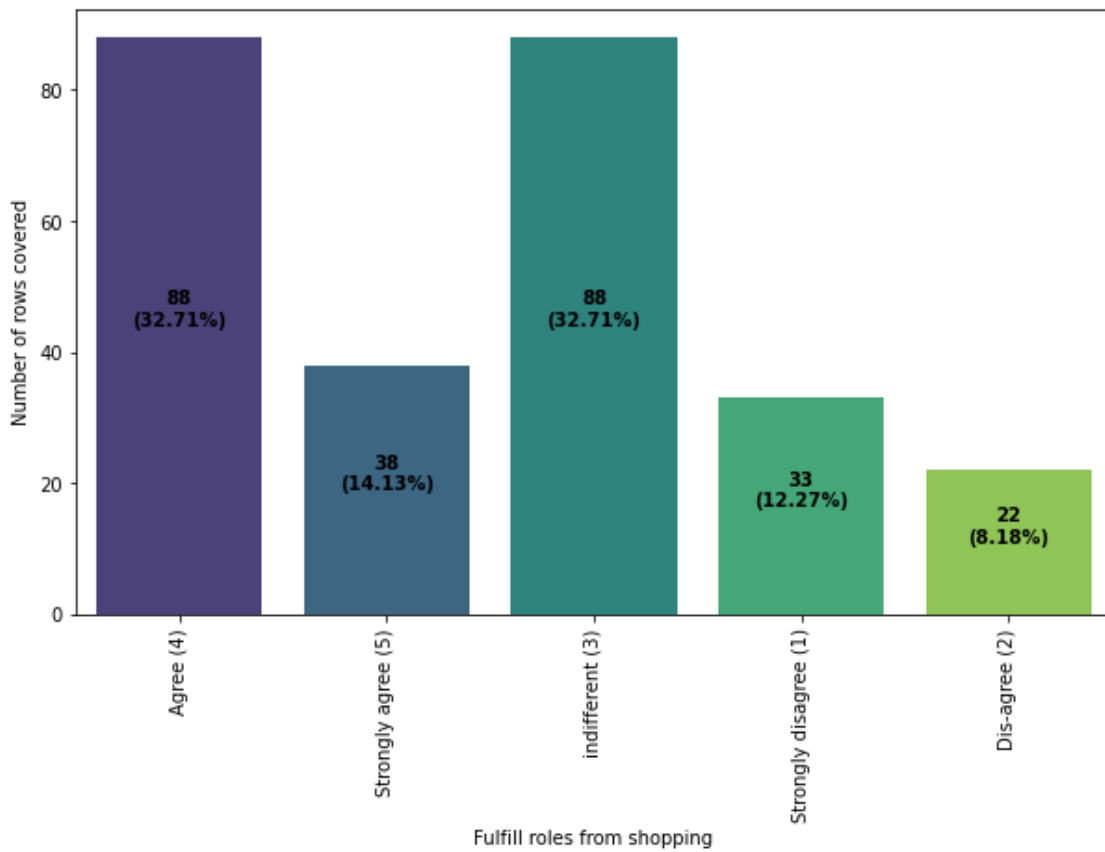


Count Plot for Sense of adventure

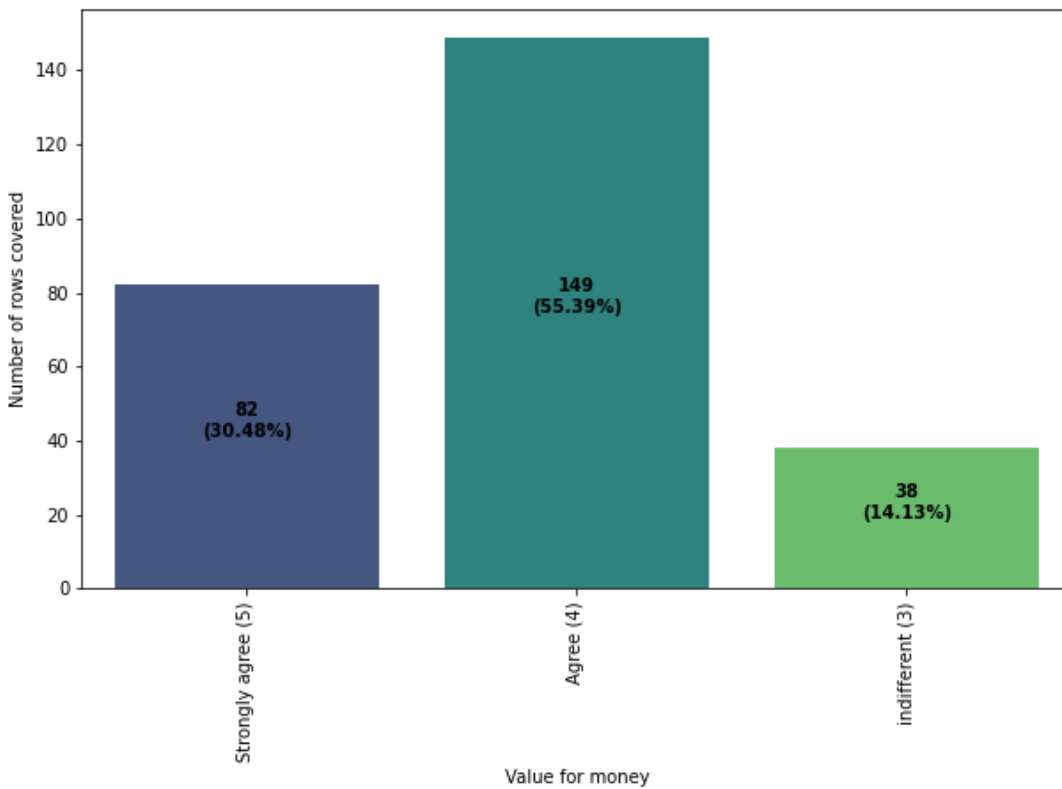


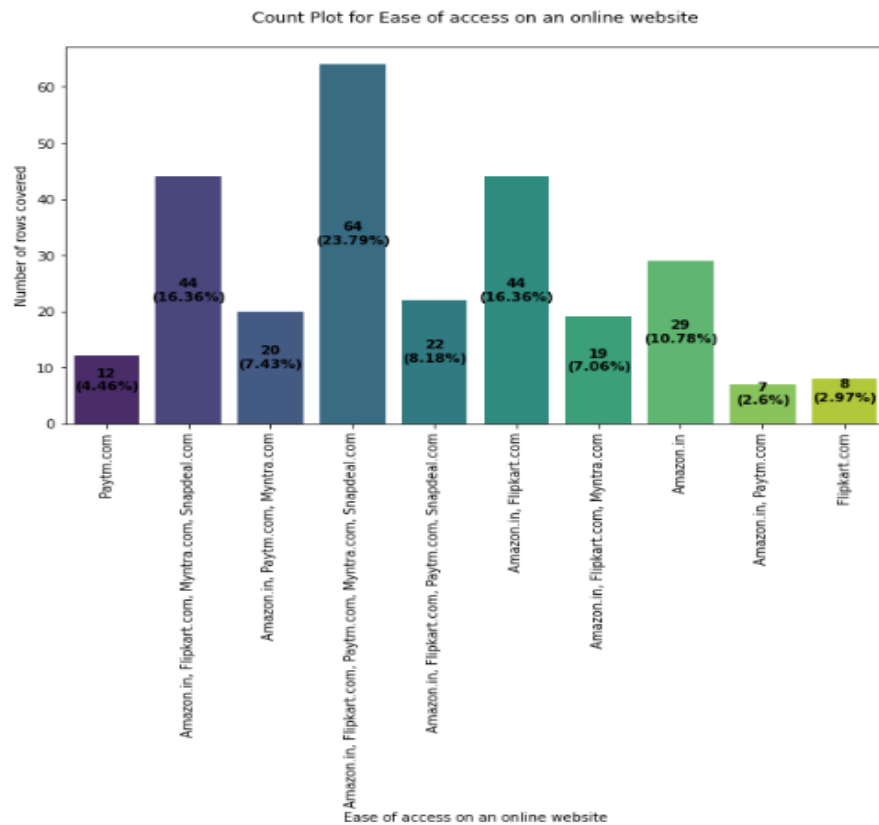
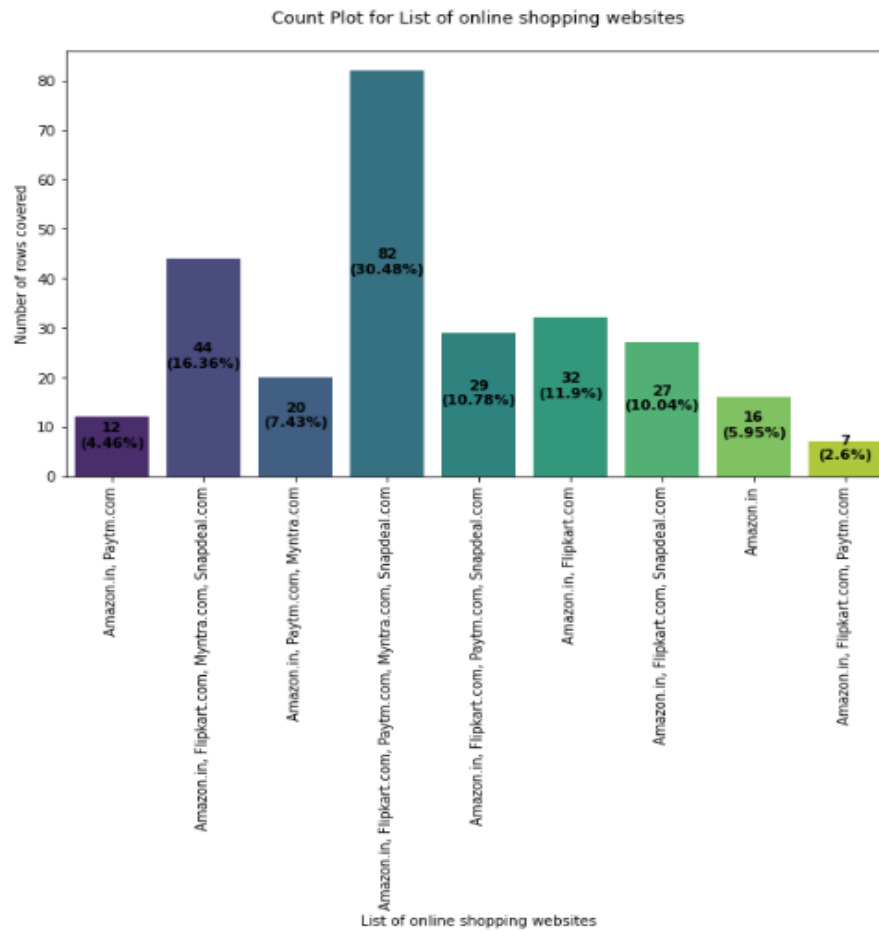


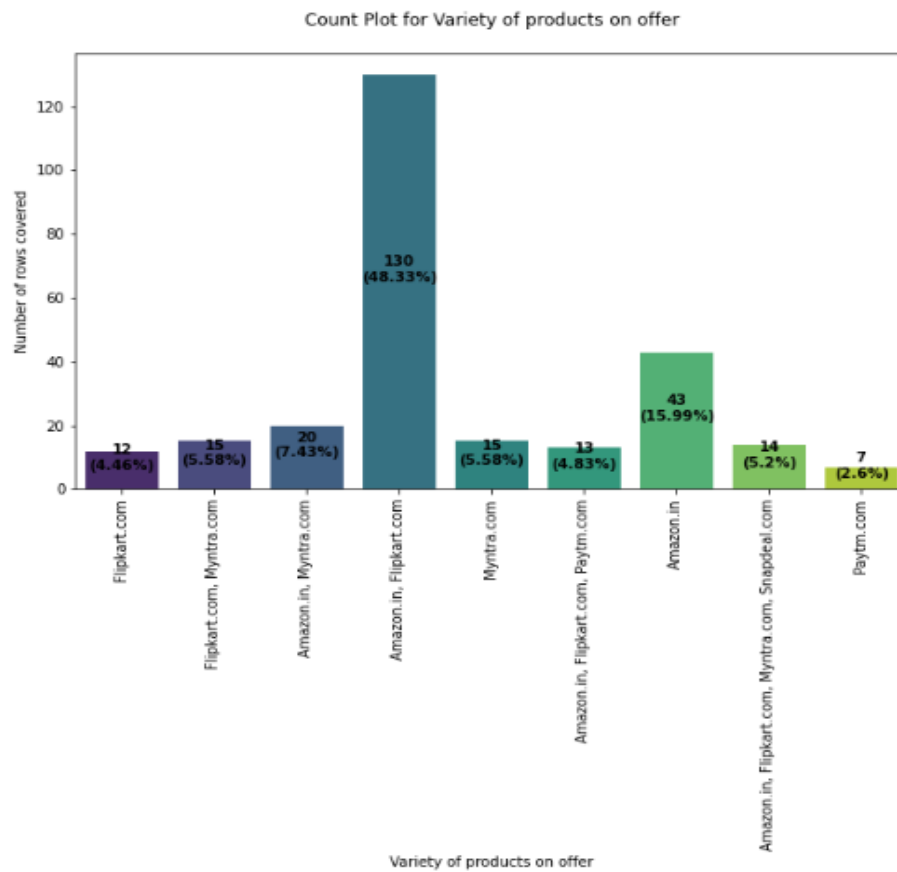
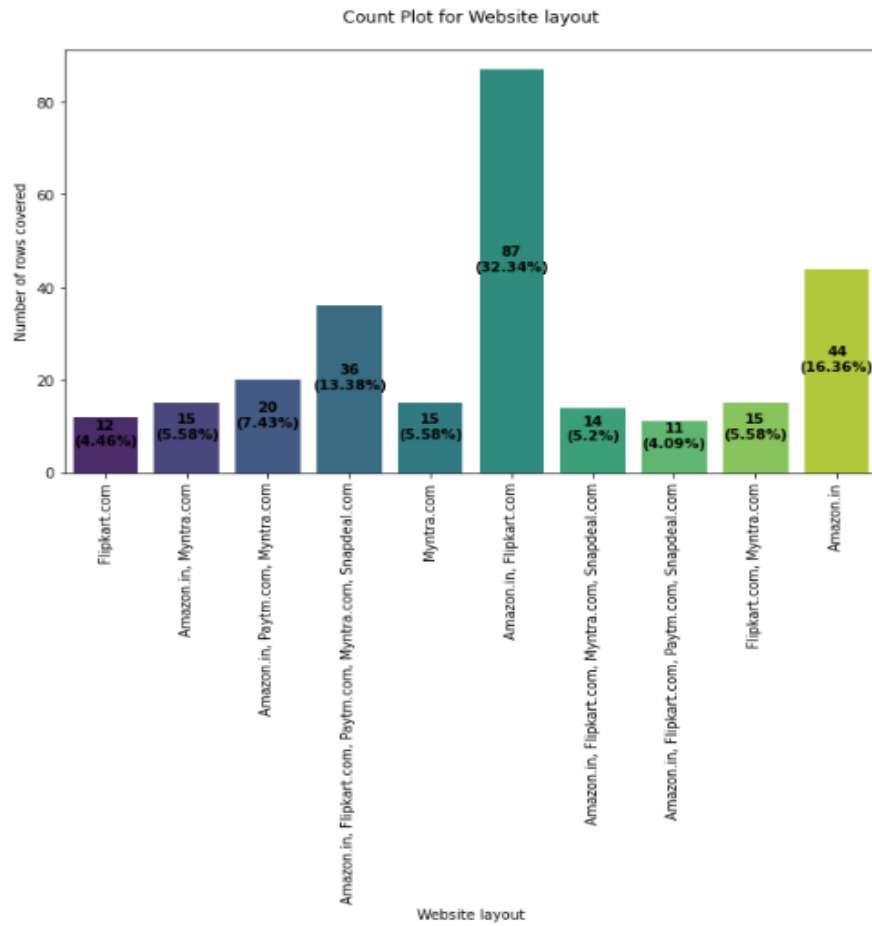
Count Plot for Fulfill roles from shopping

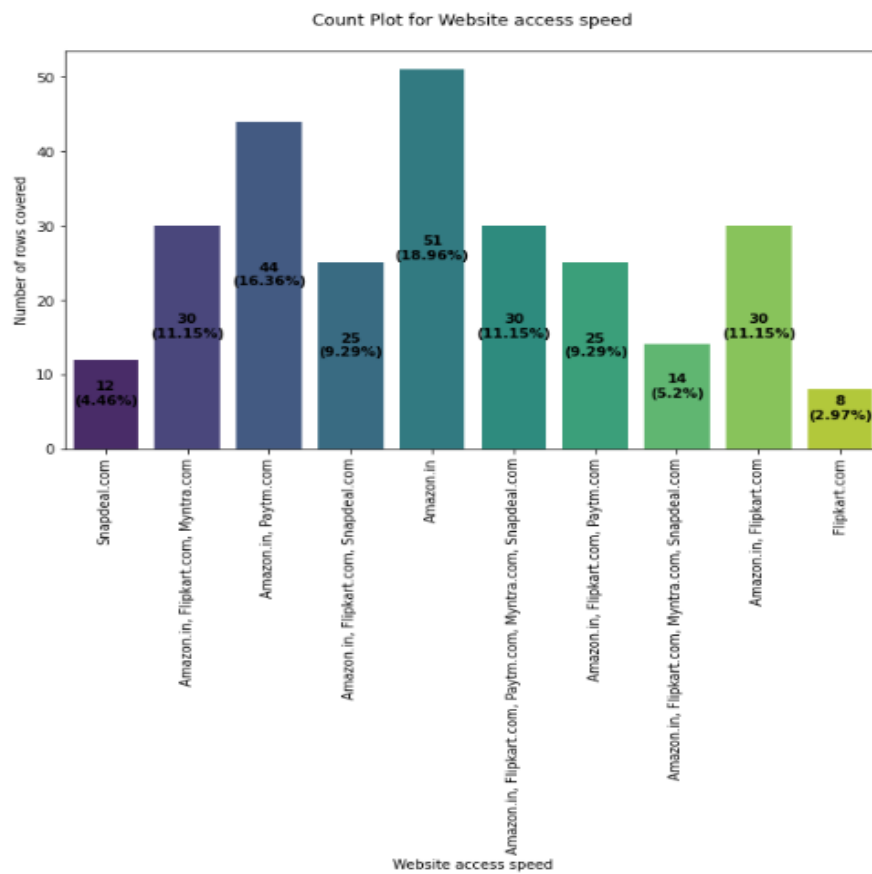
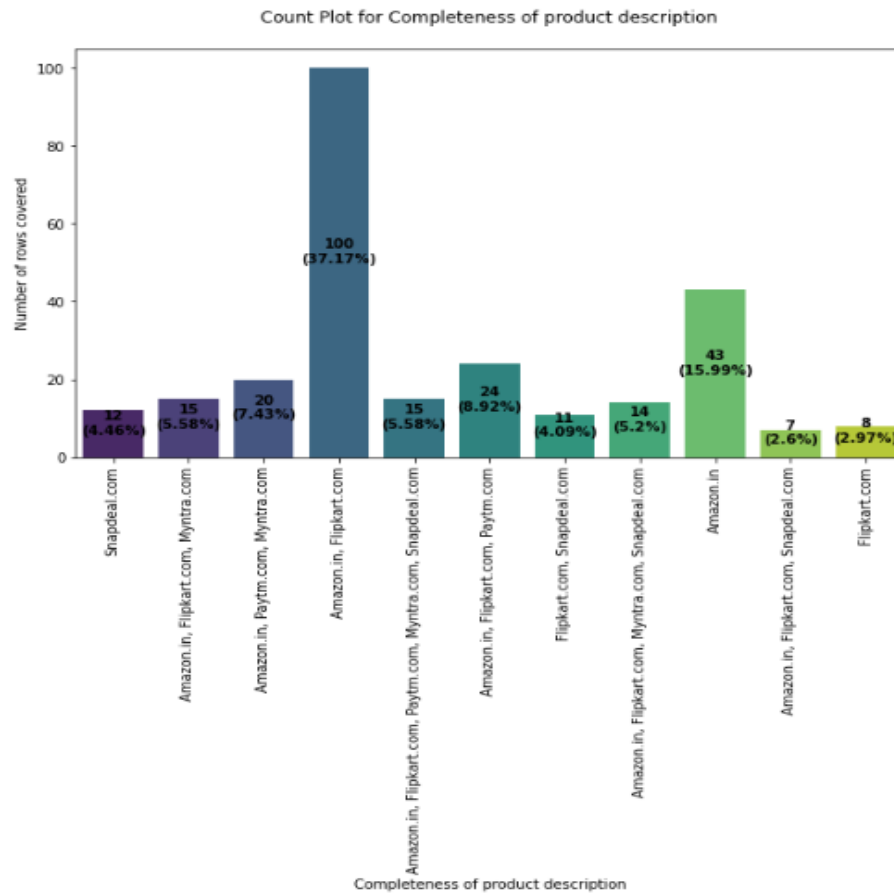


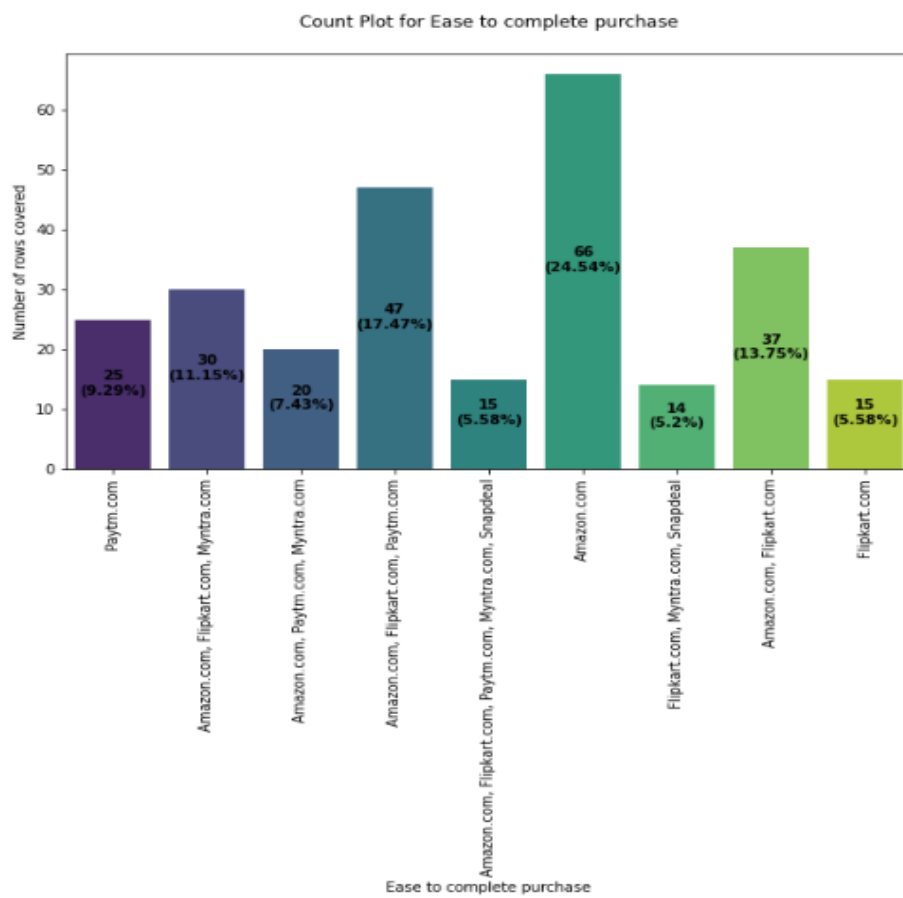
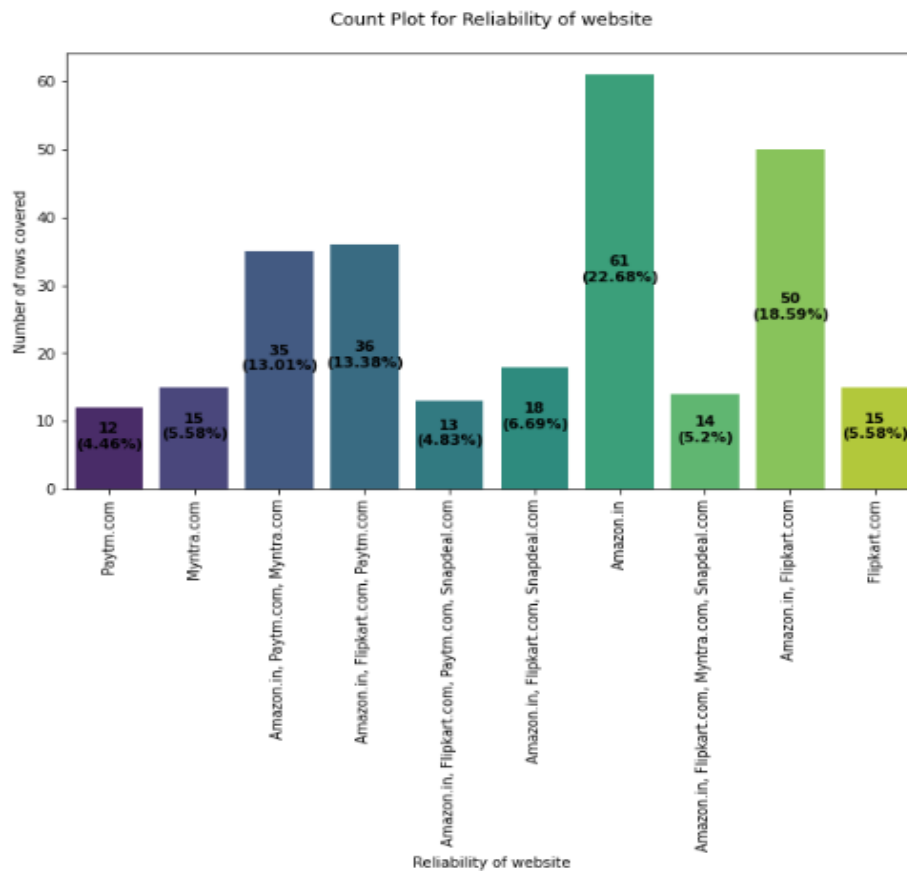
Count Plot for Value for money



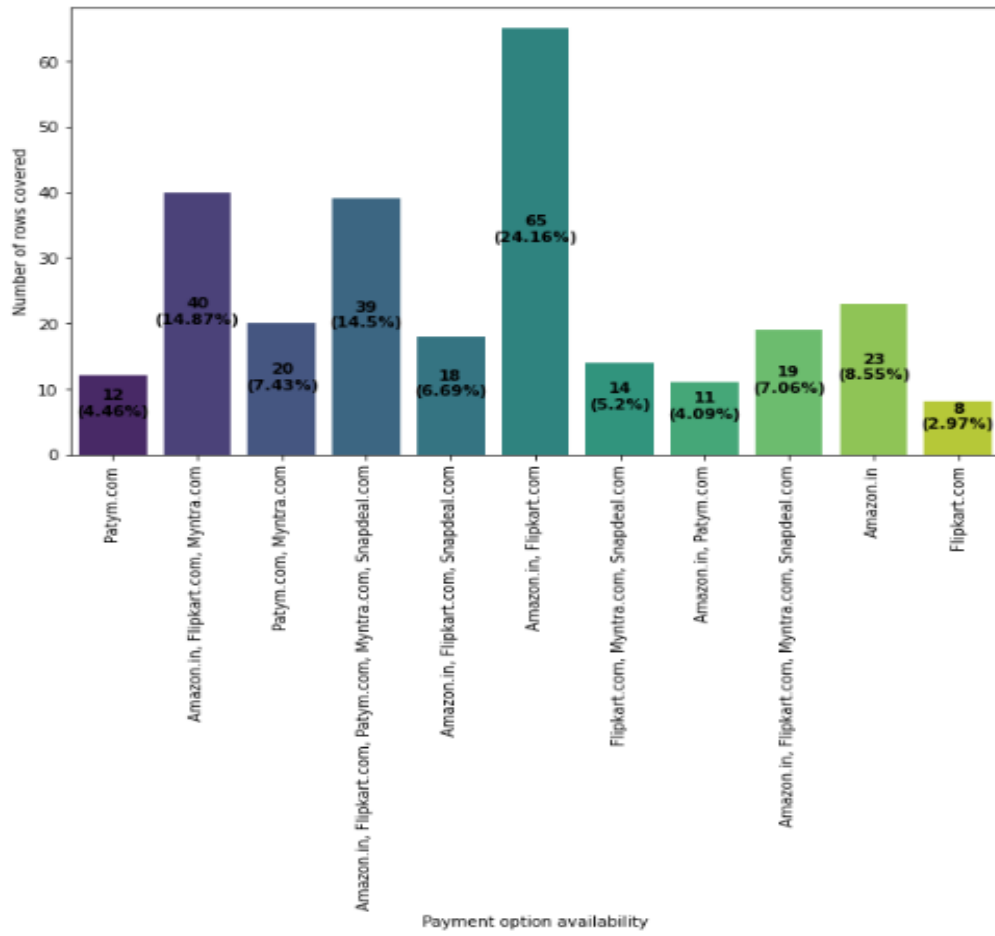




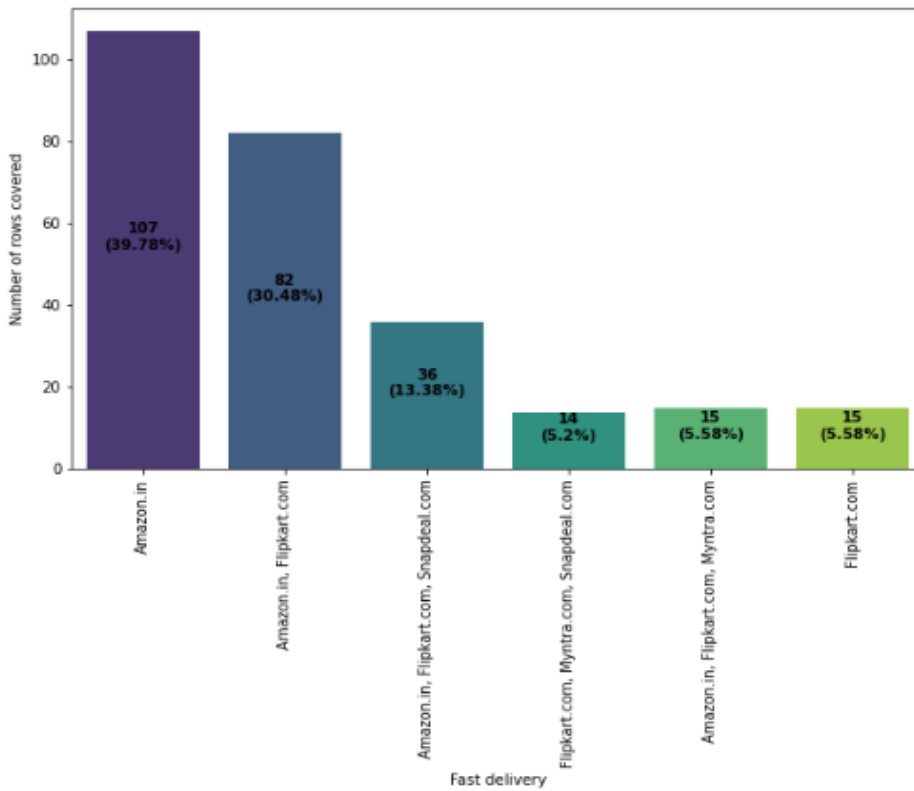




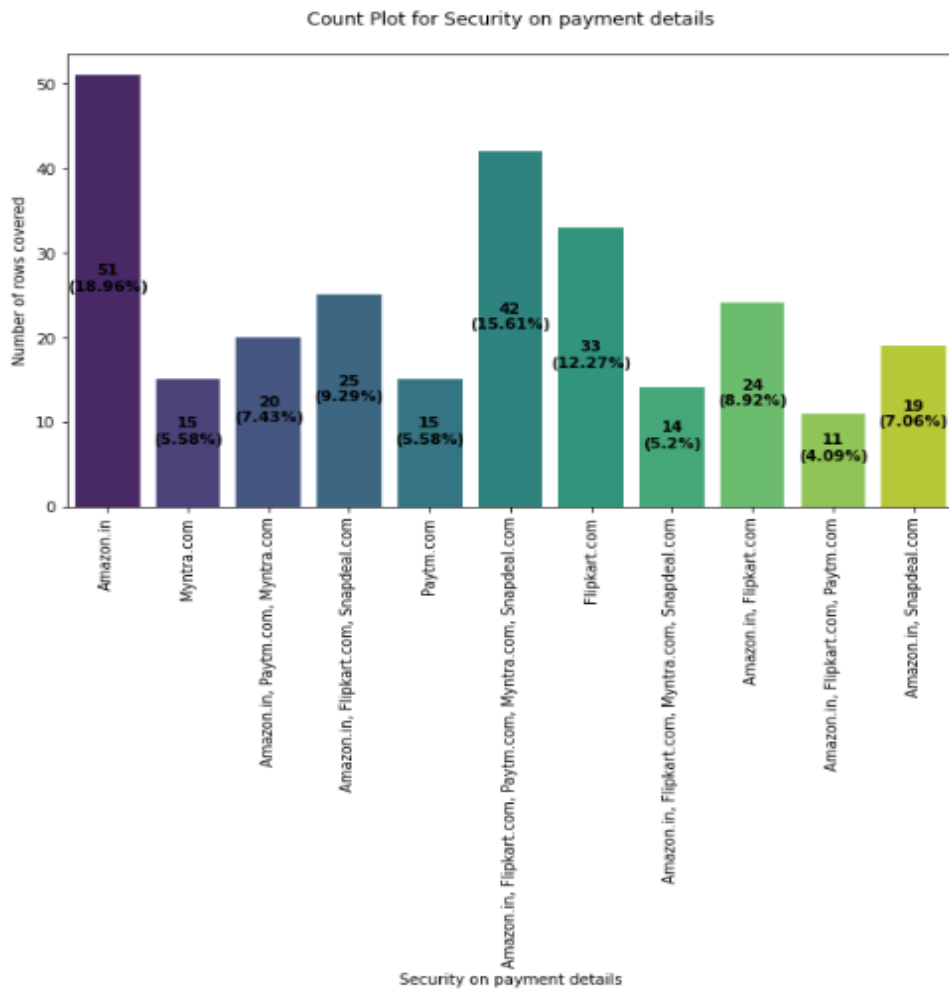
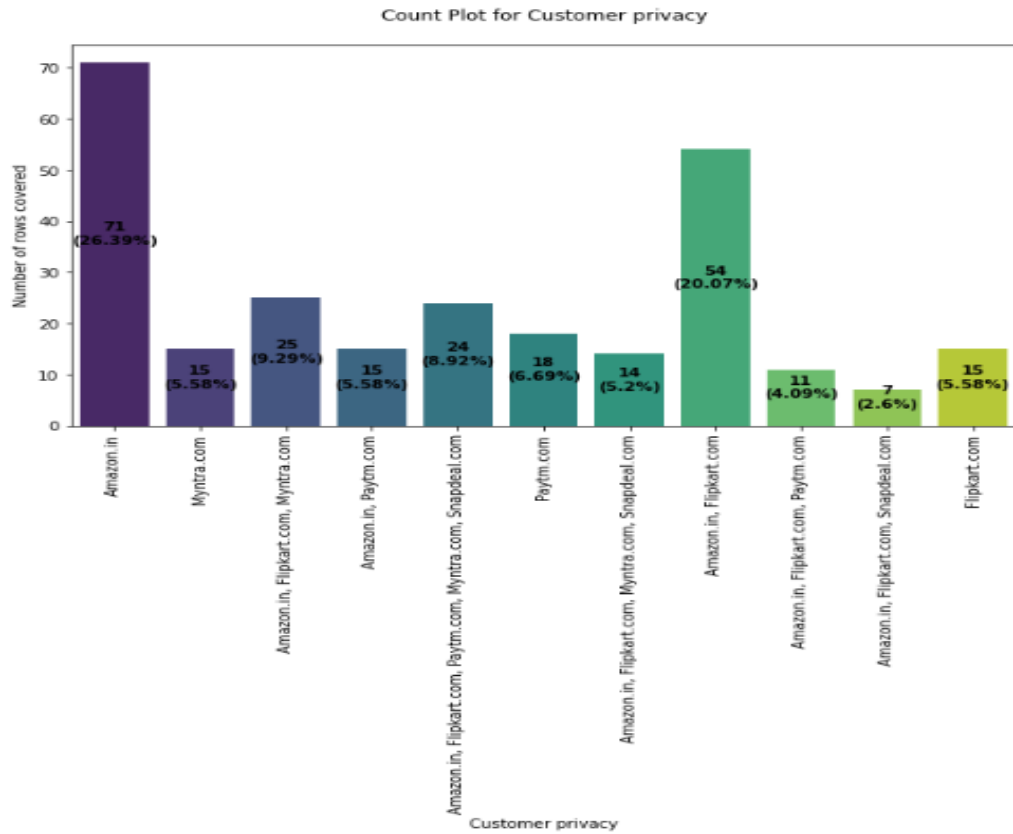
Count Plot for Payment option availability

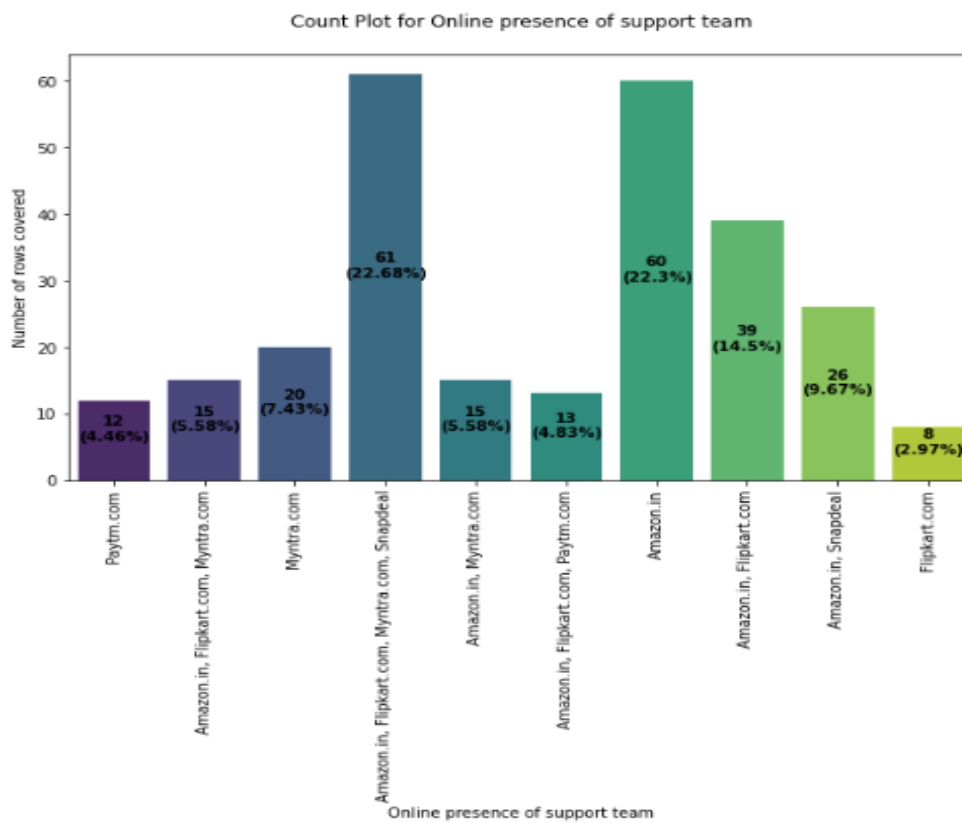
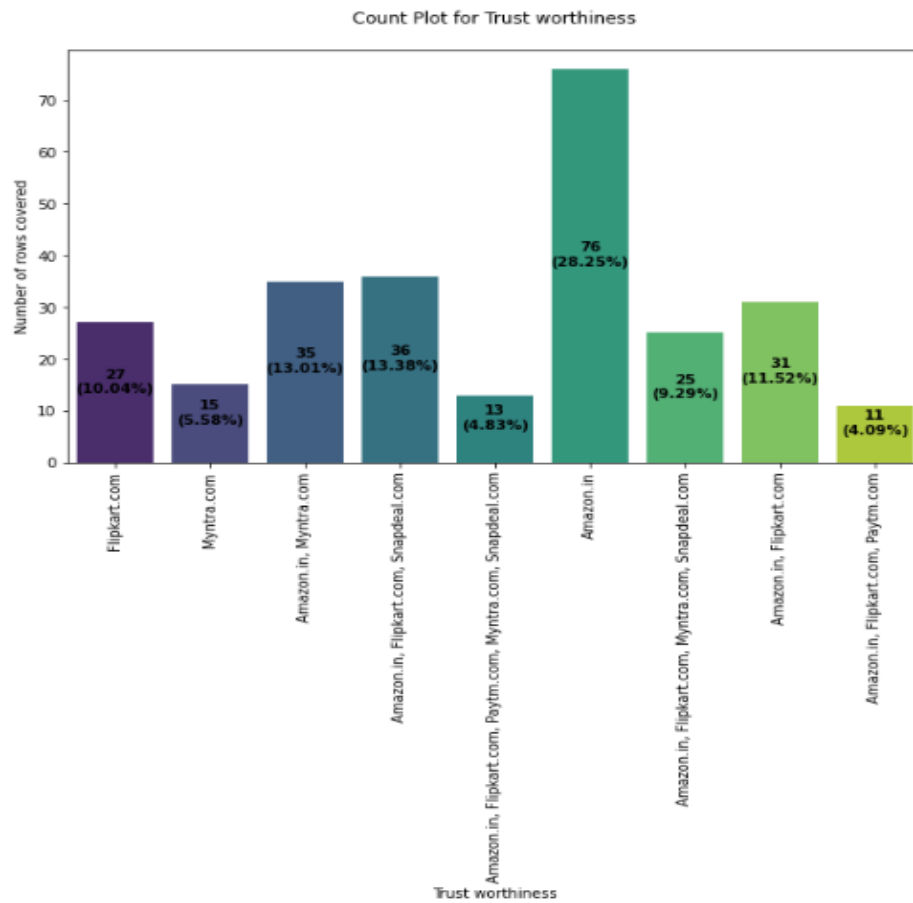


Count Plot for Fast delivery

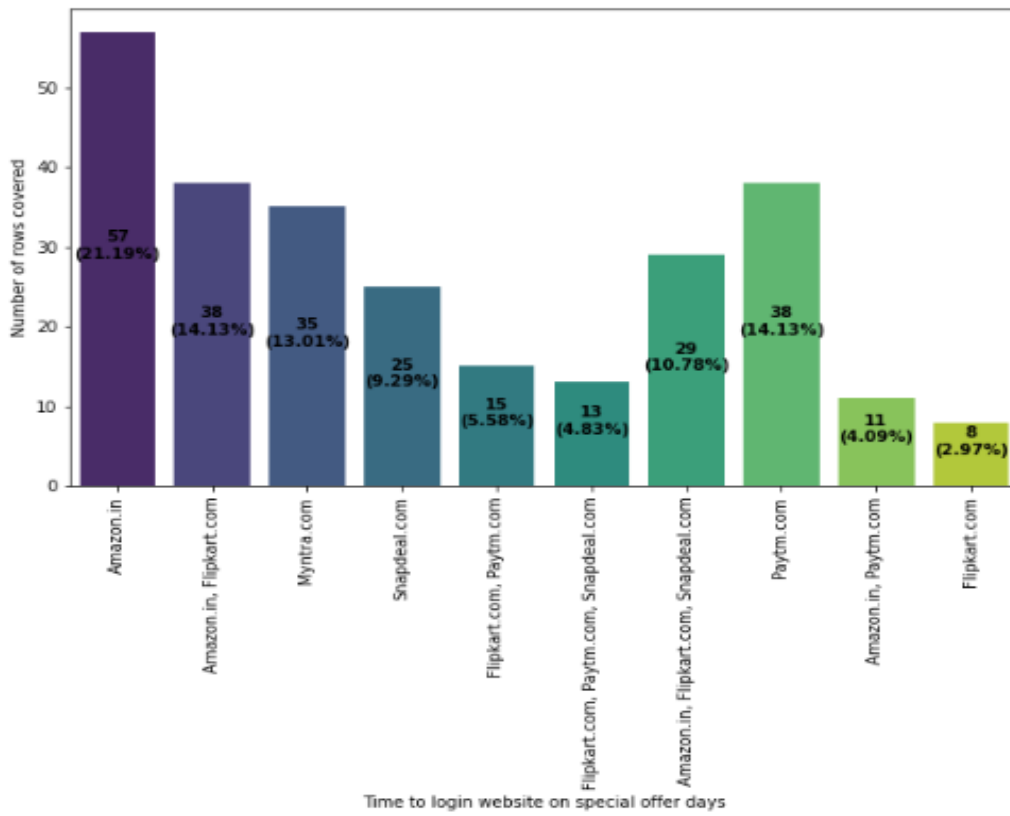




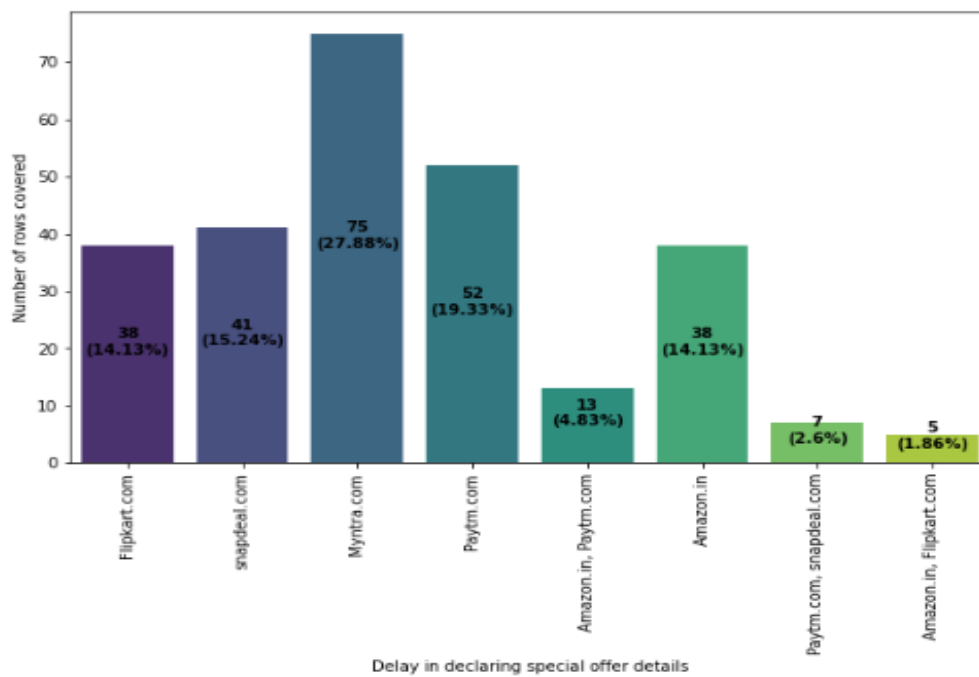


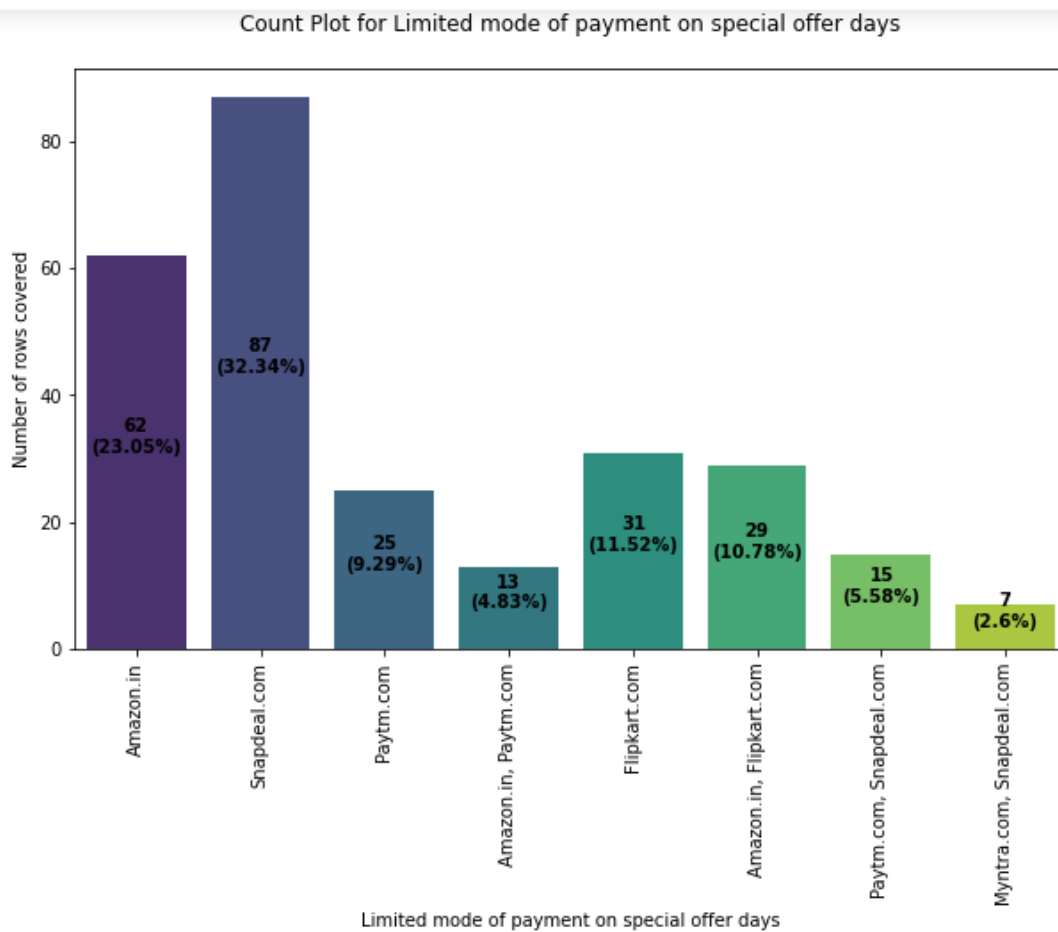
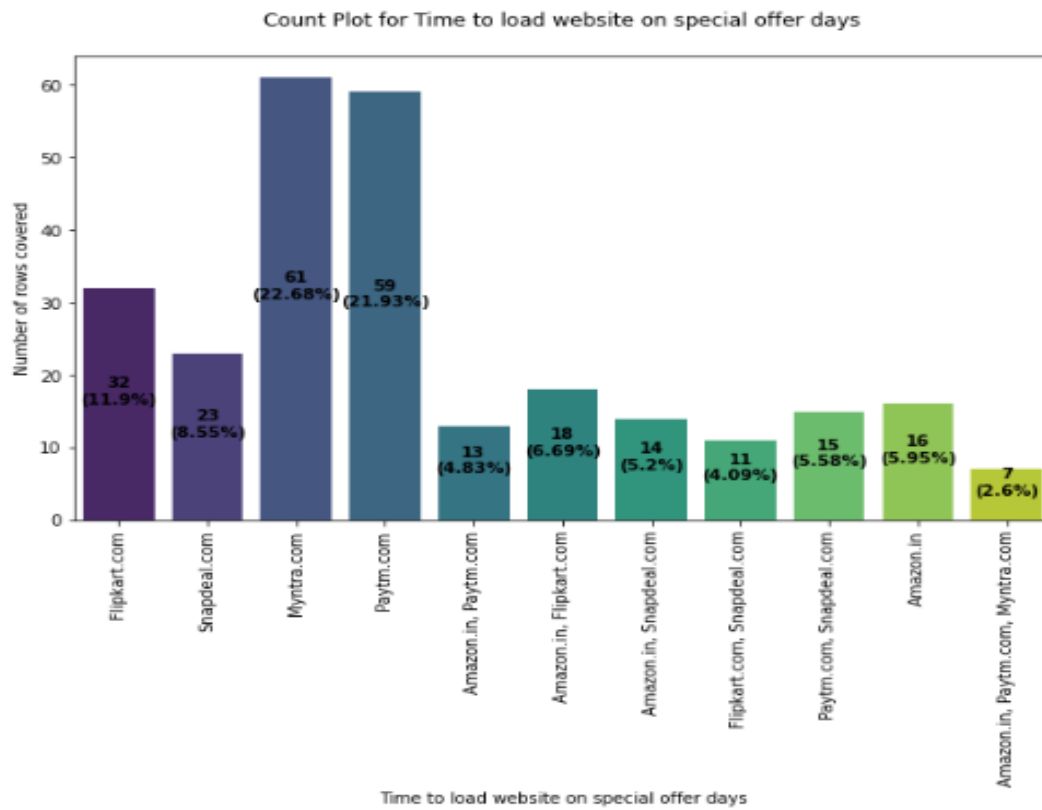


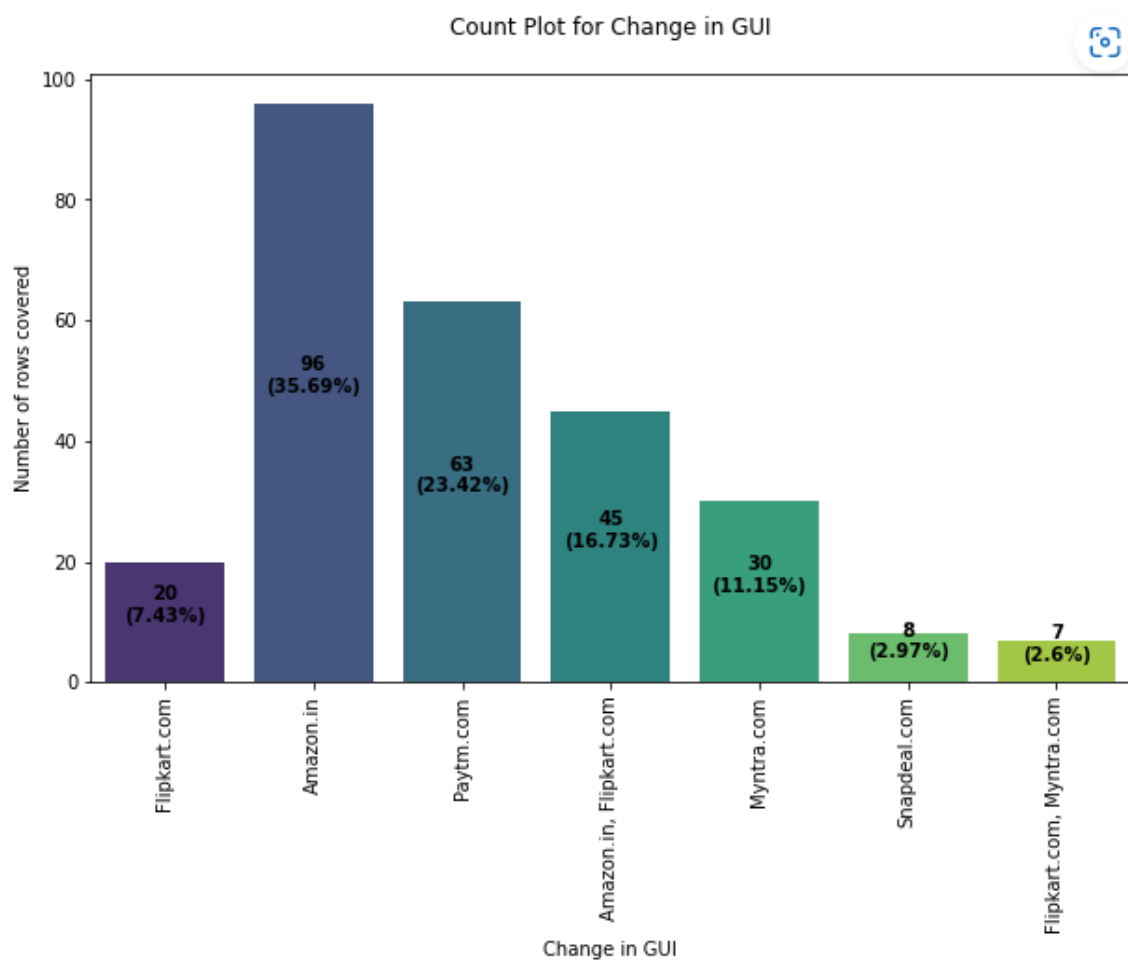
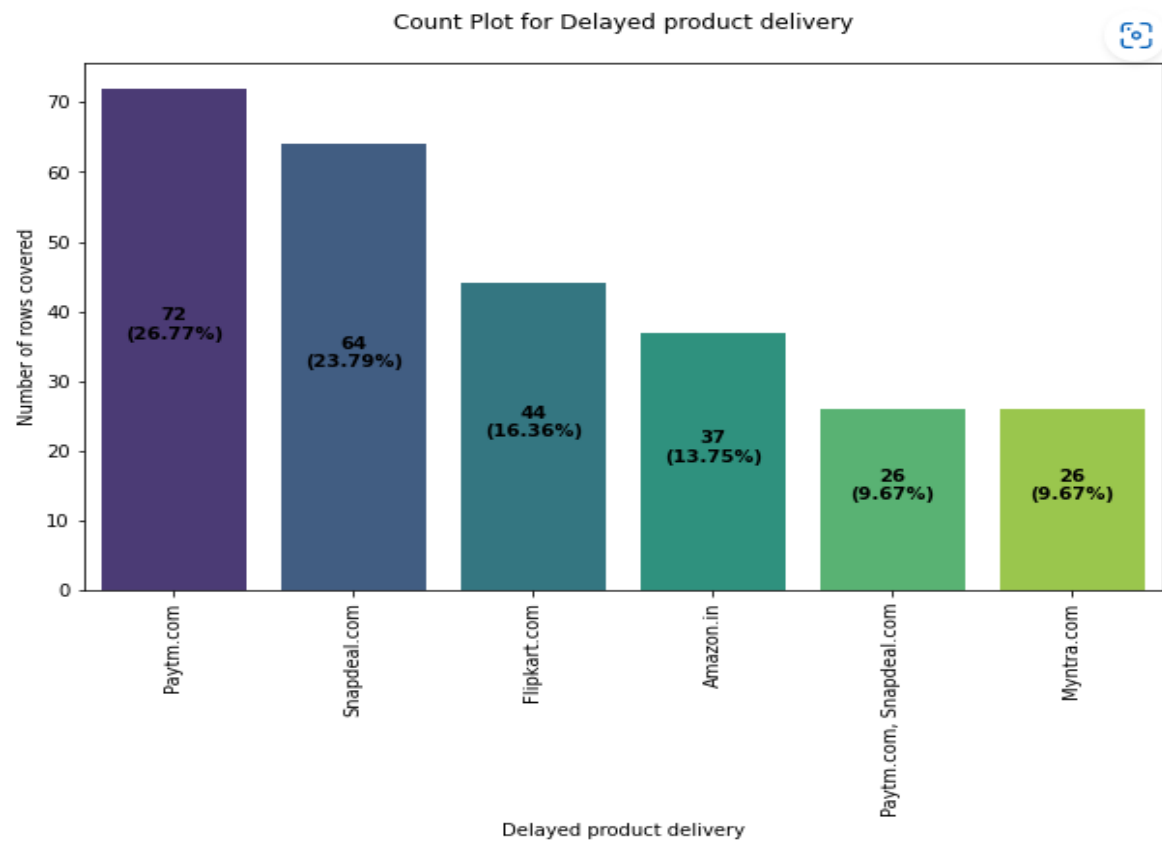
Count Plot for Time to login website on special offer days

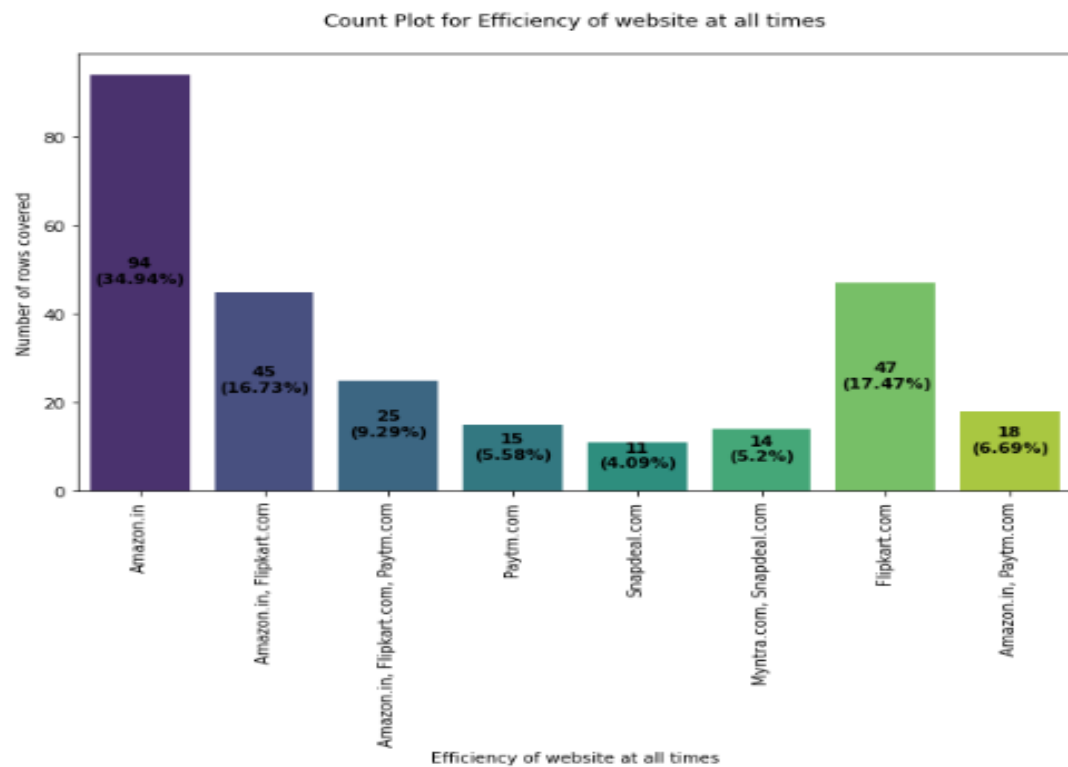
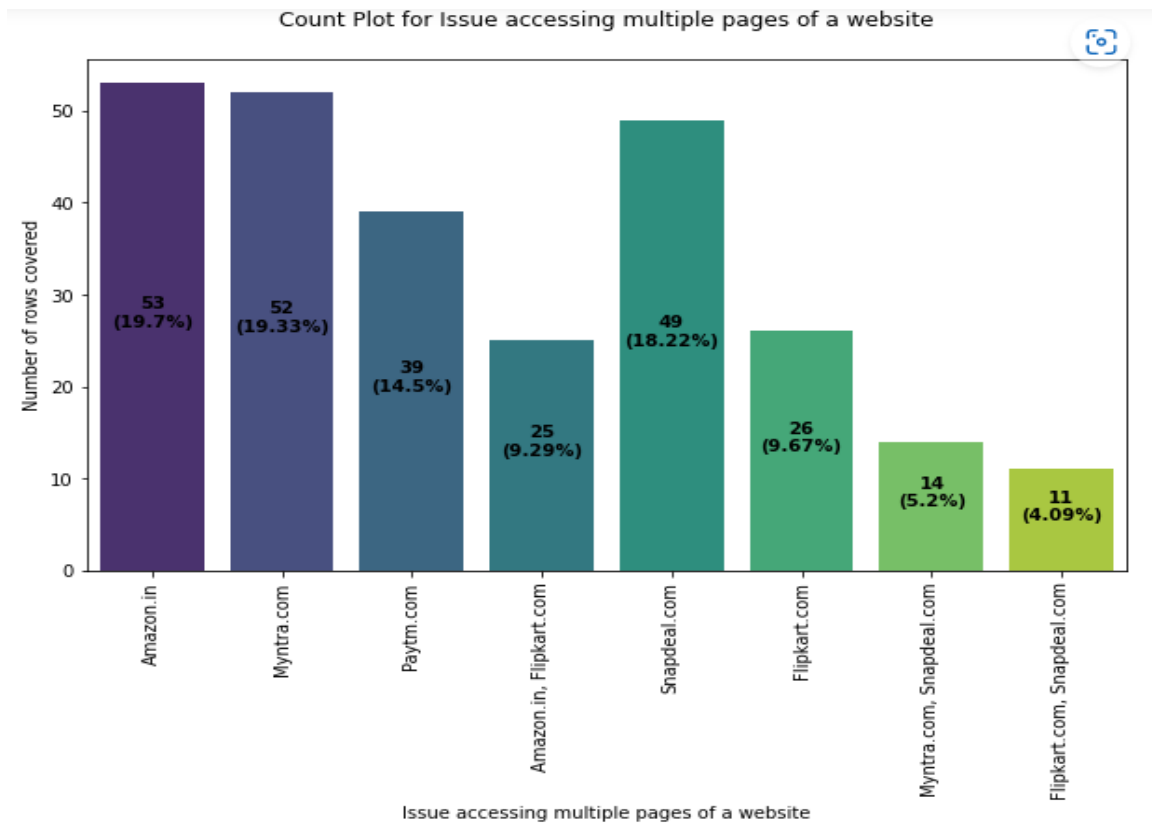


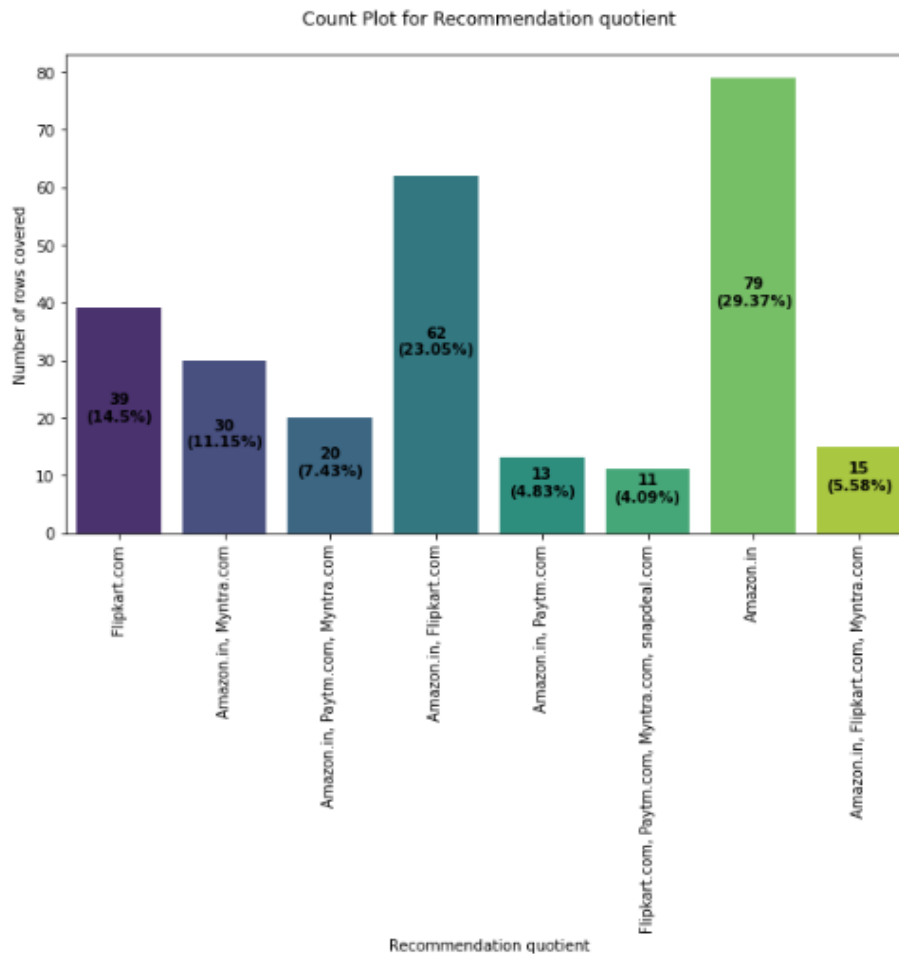
Count Plot for Delay in declaring special offer details









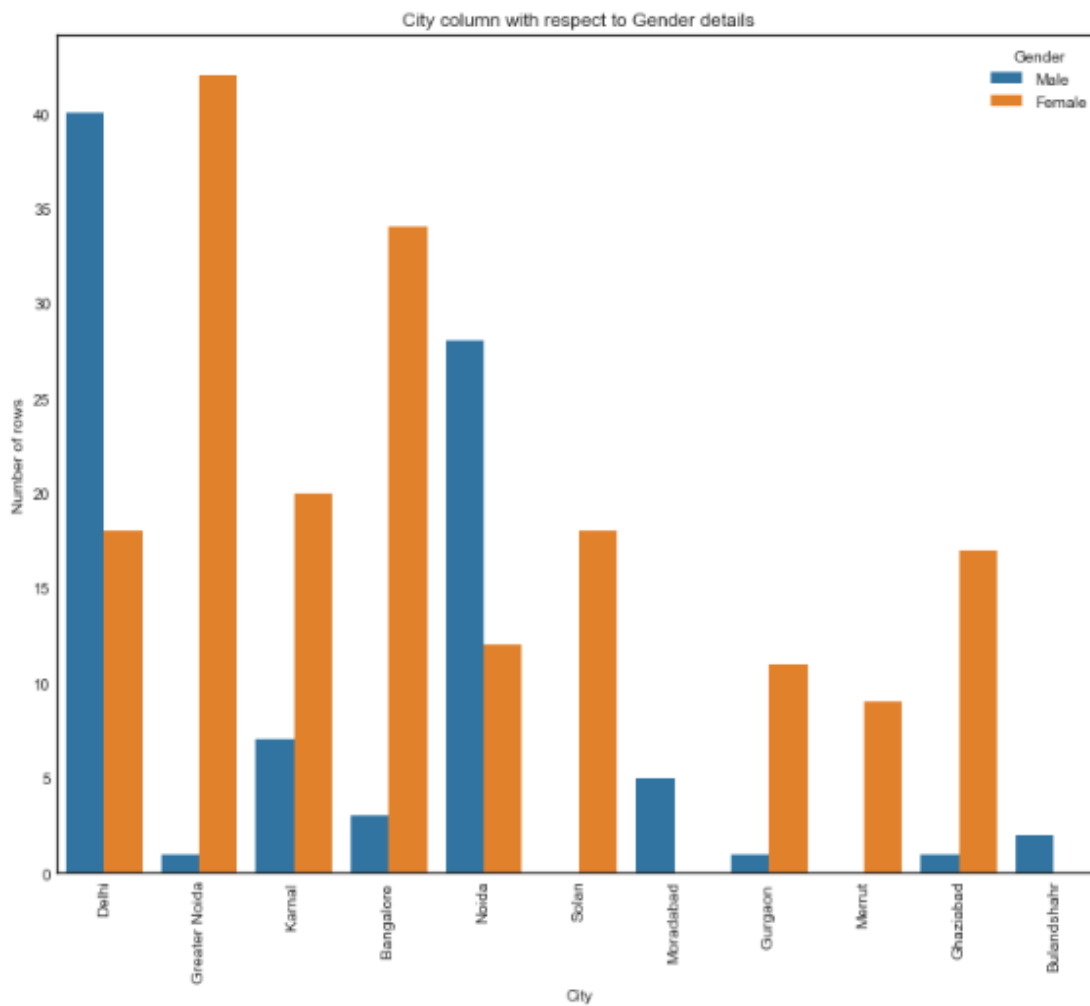


## Bivariate Analysis:

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:

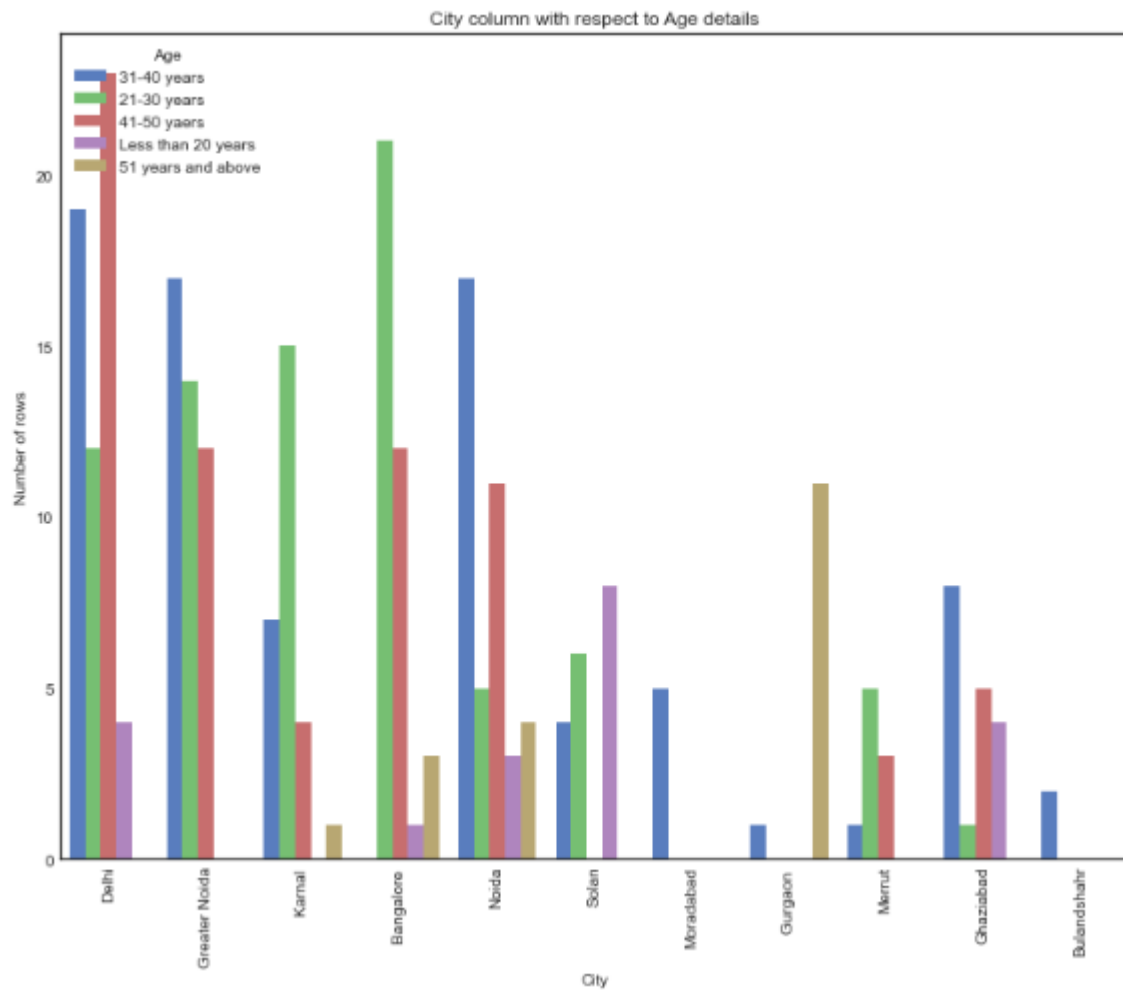
```
for col in df1:
    if col=='Gender':
        pass
    elif col=="Pin Code":
        pass
    else:
        plt.style.use('seaborn-white')
        plt.figure(figsize=(10,8))
        sns.countplot(x=col, data=df1, hue="Gender")
        plt.title("{} column with respect to Gender details.".format(col))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



Code:

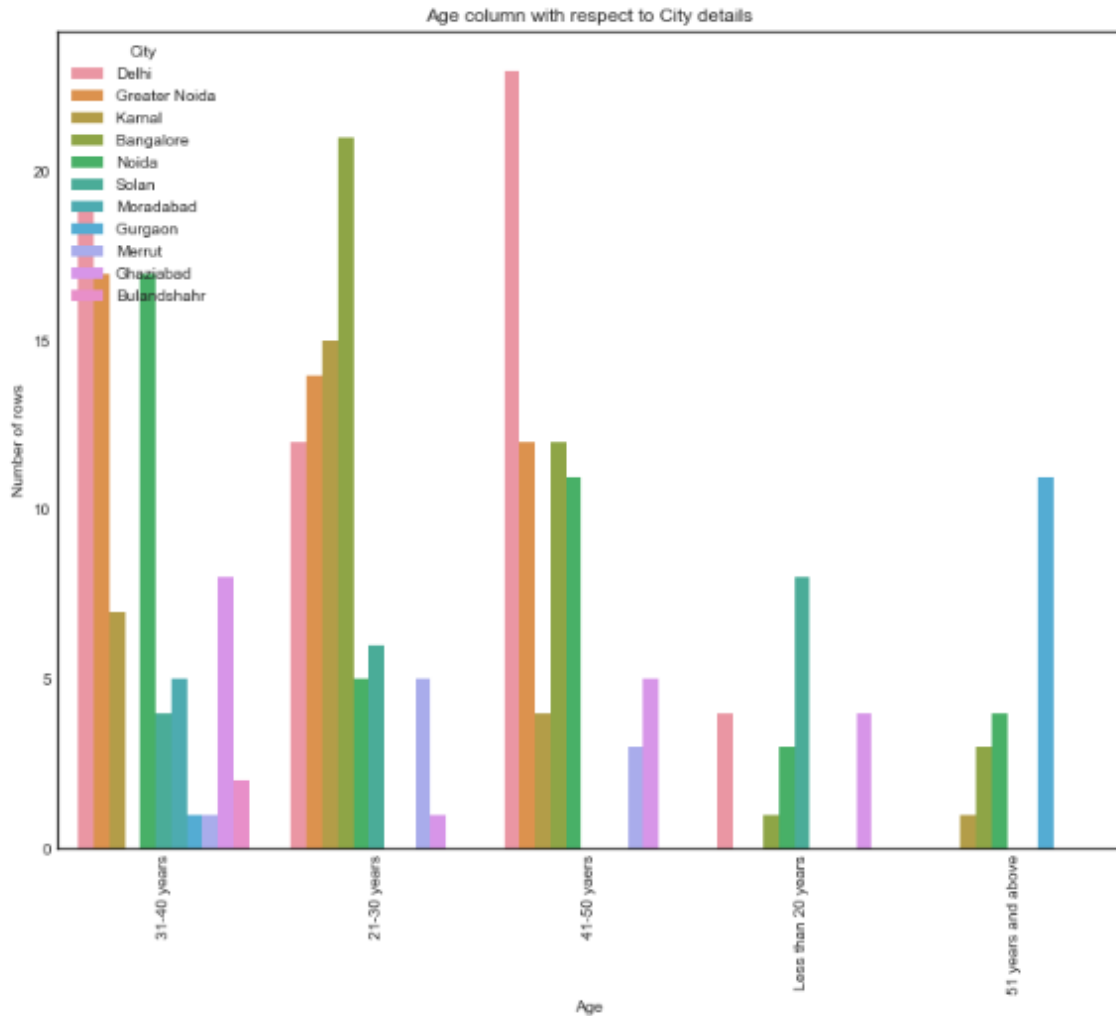
```
for col in df1:
    if col == "Age":
        pass
    elif col == "Pin Code":
        pass
    else:
        plt.style.use('seaborn-muted')
        plt.figure(figsize=(10,8))
        sns.countplot(x=col, data=df1, hue="Age")
        plt.title("{} column with respect to Age details".format(col))
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





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()
```



Then I performed Ordinal Encoding on all the object datatype columns before I move on to the multivariate analysis.

```
# Ordinal Encoding

oe = OrdinalEncoder()

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

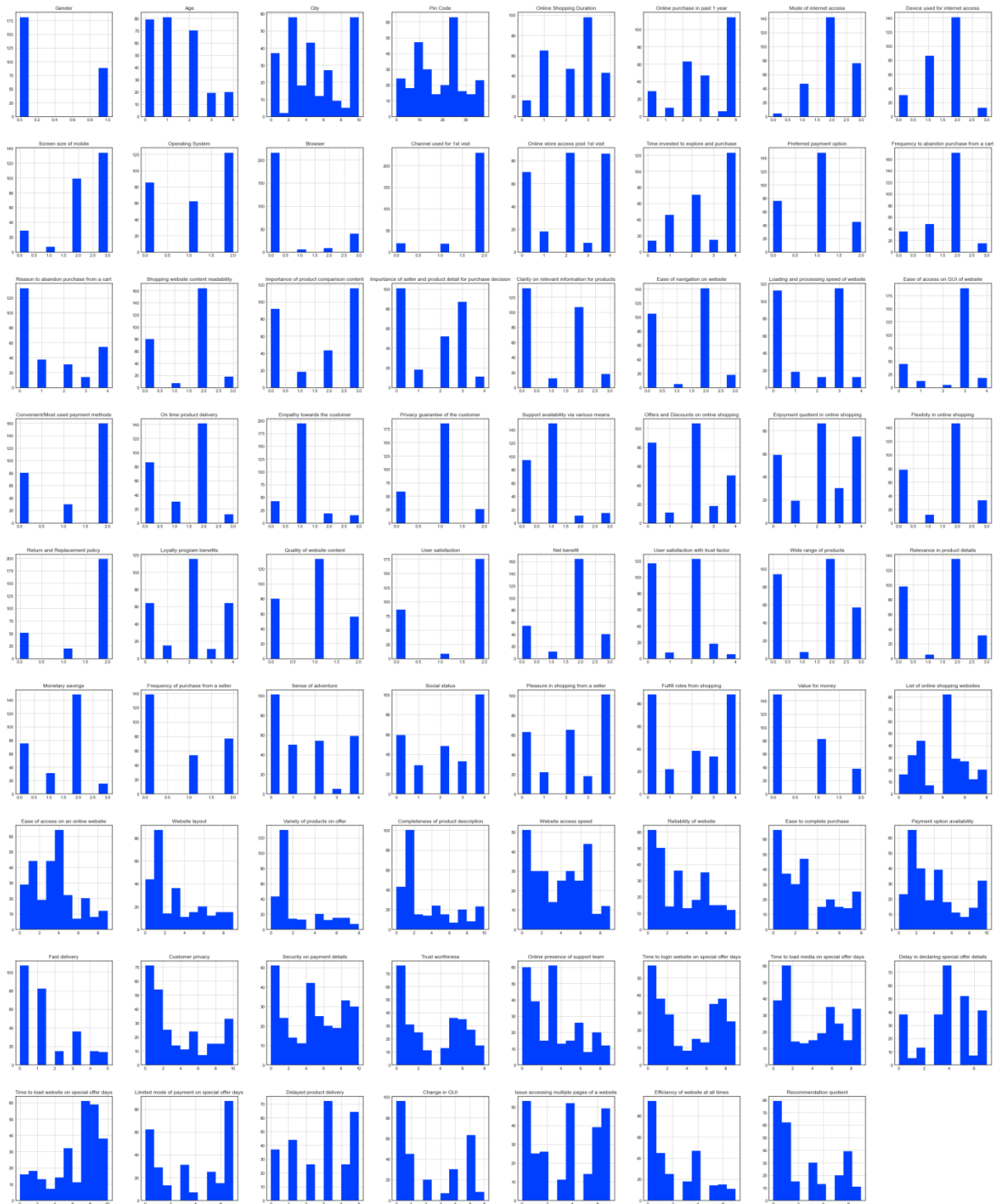
oe_col = df1.columns
df=ordinal_encode(df1, oe_col)
df
```

With the help of Ordinal Encoding method I converted object datatype to numeric datatype.

# Multivariate Analysis:

I generated a histogram plot for after encoding.

```
plt.style.use('seaborn-bright')  
  
df.hist(figsize=(40,50))  
plt.show()
```

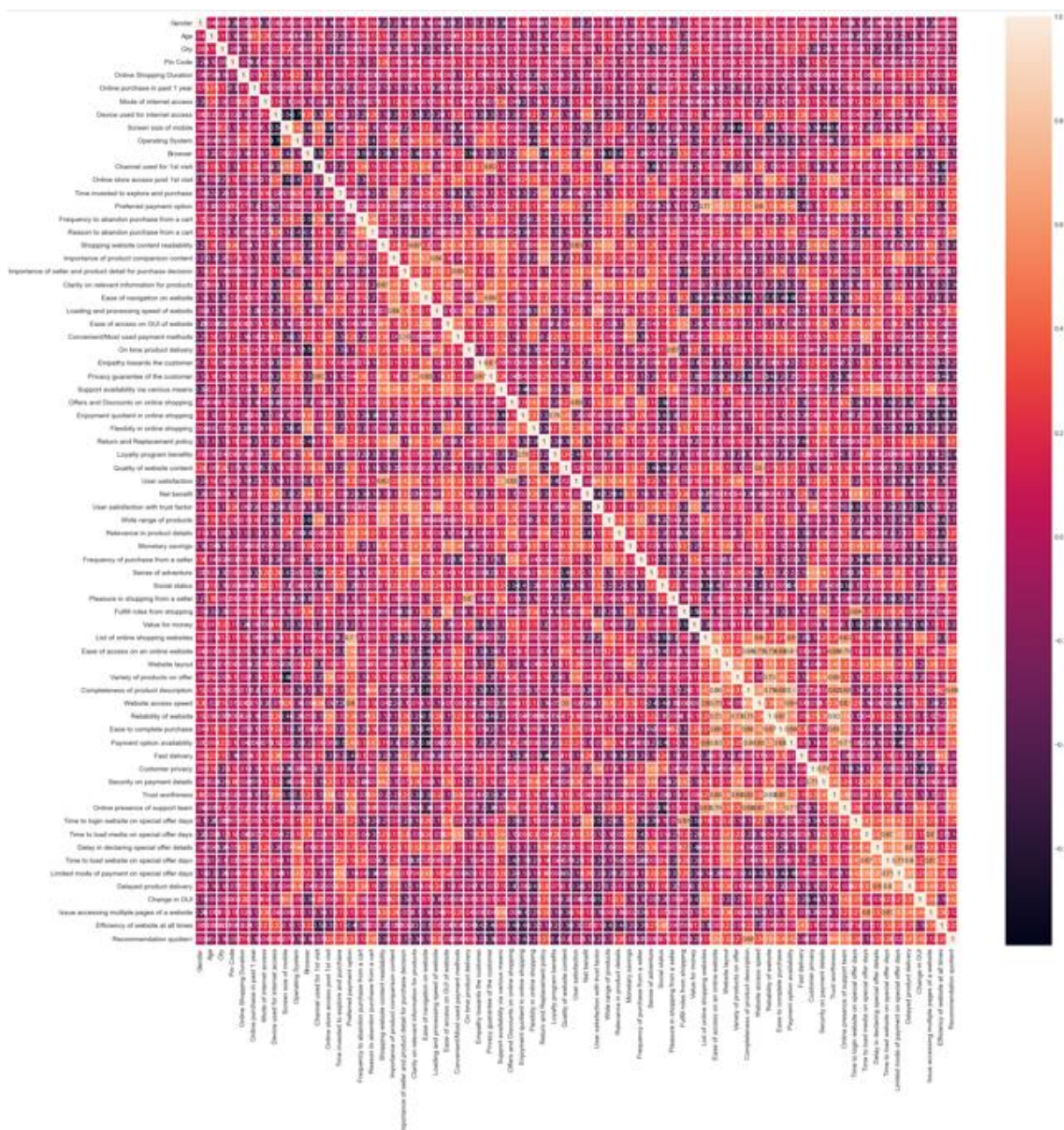


Then I use correlation method to find the correlation of the dataset, to check which columns are positively correlated and which columns are negatively correlated.

```
plt.figure(figsize=(25,25))
sns.heatmap(df1.corr(),linewidths=.1,annot=True)
```

Positive correlation - A correlation of +1 indicates a perfect positive correlation, meaning that both variables move in the same direction together.

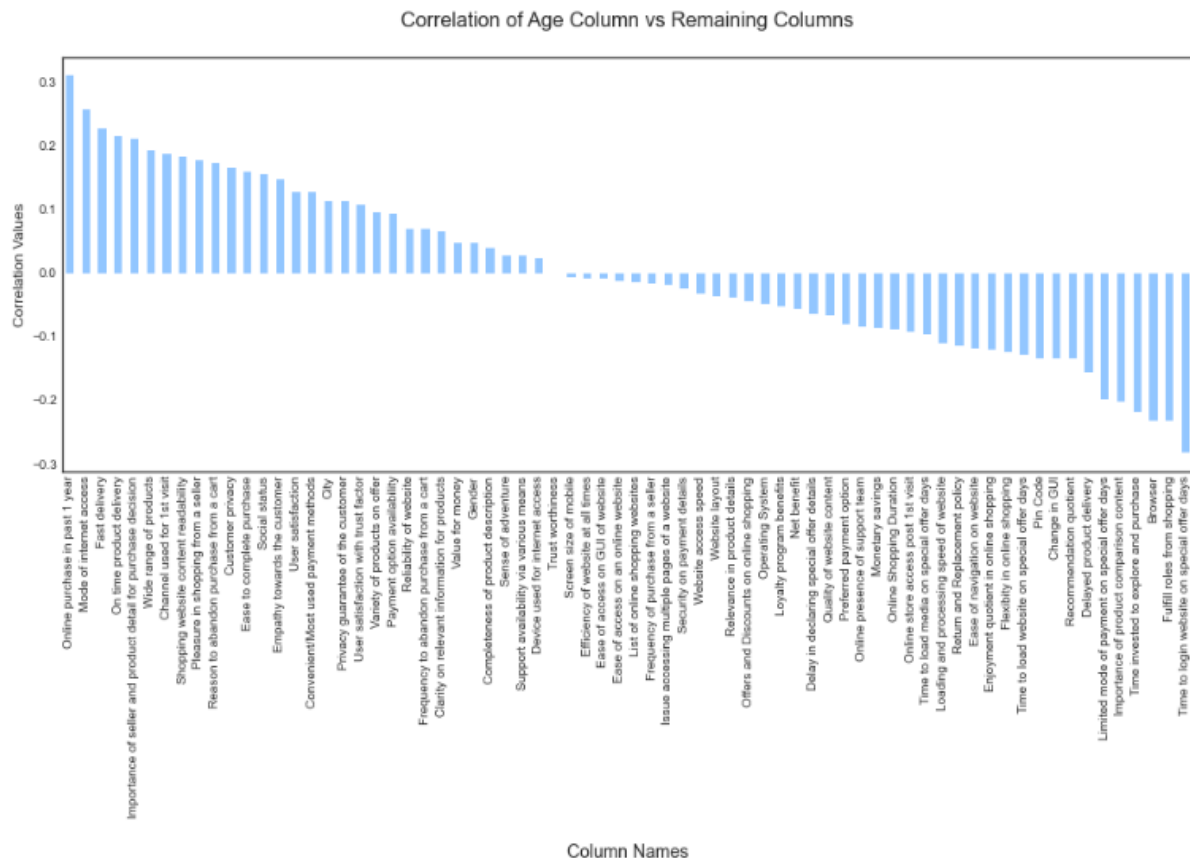
Negative correlation - A correlation of -1 indicates a perfect negative correlation, meaning that as one variable goes up, the other goes down.



In the above heatmap due to lot of columns, we are not able to see the correlation details clearly, so that I generate correlation bar plots. However we can observe the color-coding details and get a hint that there is no multicollinearity concern between the column values.

```
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()
```



The above bar plot for each column vs remaining column shows the positive and negative correlation of the 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 is 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 e-commerce 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, k-nearest neighbors for unsupervised machine learning, hierarchal clustering, apriori algorithm and neural networks.