Project Report on

Customer Retention Case Study

(E-retail factors for customer activation and retention: A case study from Indian e-commerce customers)

Submitted By-

Sankalp Mahapatra

Internship-29

Submitted to-

Shwetank Mishra

SME of Internship Batch-29

ACKNOWLEDGMENT

I would like to express my sincere gratitude to my mentors from Data Trained academy and FlipRobo Technologies for giving me the opportunity to work on this project. This project helped me a lot for using my theoretical knowledge about Exploratory data analysis(EDA) and many other data science concepts.

TABLE OF CONTENTS

INTRODUCTION

Customer retention is a business's ability to keep existing customers and continue to generate revenue from them. Companies use different tactics to convert first-time buyers into repeat shoppers. In other words, customer retention allows a business to increase the profitability of an existing customer and maximize their lifetime value (LTV).

Think of customer retention as a process where a business aims to convince existing customers to keep purchasing their products or services. Since a customer has already made a purchase, it's different from lead generation, which is the effort involved in capturing contact information of businesses or individuals who are likely to buy a product or service.

Instead, customer retention is focused on existing customers. The goal is to increase repeat purchases by building customer loyalty through excellent customer service, product value and a distinct advantage over similar products or services.

Why is Customer retention important ?

Customers tend to move on for a myriad of reasons, which can include poor customer service, too much friction in the buying process and a lack of perceived value. This is why it's a good idea to map out the customer journey to know where the leaks are. It's also a best practice to solicit customer feedback and incorporate it into the company's larger plans.

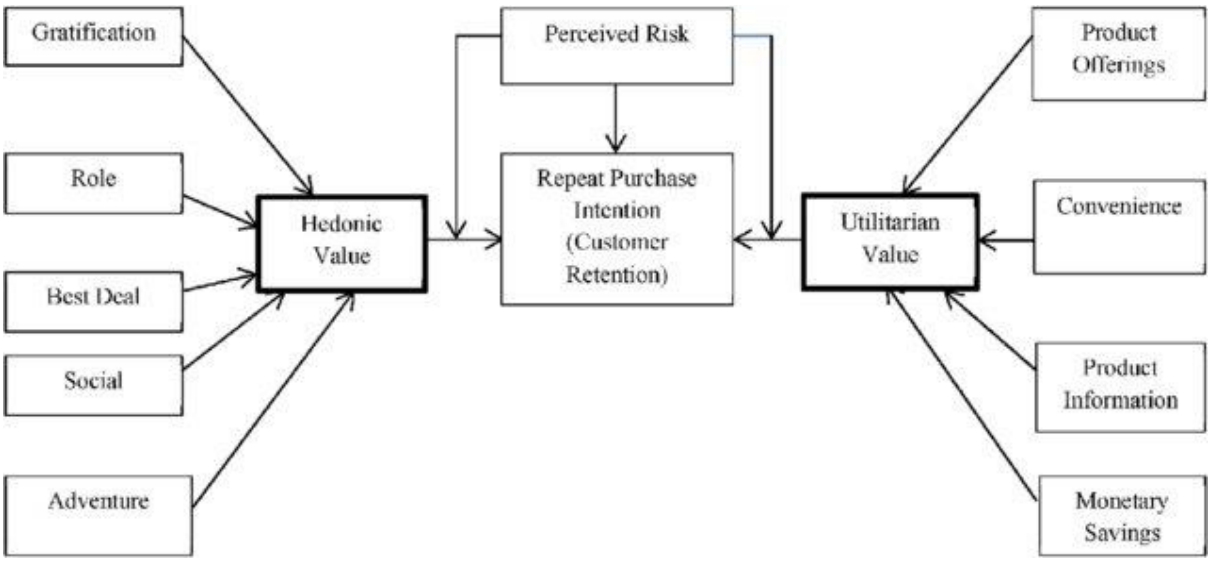
Customer retention is vital in driving repeat purchases and on-going value from your customer base. One oft-cited rule of thumb is that it costs five times as much to acquire a new customer as it does to retain an existing customer. Two of the most important factors in improving customer retention is understanding your customers' satisfaction and loyalty. Businesses also need to understand any operations that may turn off potential and existing customers, such as slow or poor customer service or a faulty product.

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.

Utilitarian value: Utilitarian value is an objective value which provides some functional benefits to the consumers and helps consumers to accomplish practical tasks.

Hedonistic value: Hedonistic value is subjective (Psychological) value which provides an experiential satisfaction. In other words, the immediate psychological gratification that comes from experiencing some activity or from consumption of a product.



In the above diagram, we can see that the Repeat Purchase Intention (customer Retention) 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 motivation for undertaking this project was to find out how many people are shopping from e commerce websites online. There are also local stores available in the market which people can directly visit and get the things of their need, but here we will find out how are the e-commerce giants like flipcart, amazon, snapdeal etc are retaining the online traffic so that customers would buy things of their need from them . Here we are provided with a dataset which contains responses from different customers of e commerce websites and their age , sex and some other information are also present.

Benefits of Customer Retention

The main benefit of customer retention is the ability to maximize the amount of money you can extract from each customer. There are also other benefits including the following:

- **Increased profits:** Many companies generate the majority of their revenue from existing customers—61% of SMBs said this was the case, per a BIA/Kelsey report—so focusing on this part of your business should be the priority. It will not only increase your revenue, but also your business's profitability.
- **Lower costs:** Retaining an existing customer is anywhere from *5-25 times cheaper* than acquiring a new one, according to Bain & Company, so it's a much more cost-effective strategy in the long run.
- **Increased average order value (AOV):** Repeat customers tend to spend more over time while increasing their average order value. That's why just a 5% increase in retention rate can lead to profits growing 25-95%, per Bain & Company. And loyal customers are 23% more likely to buy again than others, according to a Gallup study.
- **Acquire brand ambassadors:** Word of mouth is one of the best ways to grow your business organically. The more loyal your customers, the more likely that they'll share positive experiences and recommend your company to others.

Customer Acquisition vs. Customer Retention

Customer acquisition refers to the actions or processes designed to help a business gain new customer. This includes any efforts focused on finding new leads or turning prospects into paying customers.

Customer retention, on the other hand, happens after you acquire the customer. Once they make a purchase, you're trying to build loyalty and drive repeat business.

Why Do Customers Leave a Company?

Customers tend to move on for a myriad of reasons, which can include poor customer service, too much friction in the buying process and a lack of perceived value. This is why it's a good idea to map out the customer journey to know where the leaks are. It's also a best practice to solicit customer feedback and incorporate it into the company's larger plans.

When to Focus on Retention

The maturity of your business will determine whether you need to focus on customer retention. For instance, if your company just launched, you should focus on customer acquisition, as there are no existing clients to retain. The focal point at that stage should be developing strategies that will cultivate your initial customer base. That's because you're not getting any sales or customers, making it a moot point in trying to retain them. Tactics can include creating co-branded content, content produced by your business and a non-competitor, or creating a paid ad campaign.

But it doesn't mean you can plant the seeds. Actions such as engaging with consumers and making the purchase process as frictionless as possible helps a business to gain traction. As this happens, there will be more data to look at what's working—perhaps start tactics such as retention email campaigns or surveys. That way your business is working towards encouraging existing and past customers to make additional purchases.

Once you're more established, you can incorporate more customer retention tactics—making it more of a priority than customer acquisition—as you start to

generate more consistent sales. As your sales grow at a steady rate and you have a decent-sized customer base, you can shift more time and attention to your customer retention efforts. At this stage, things like loyalty or referral programs make sense, as you'll have a steady (and hopefully loyal) customer base to draw from.

Customer Retention Statistics

Here are some more statistics that demonstrate why focusing on customer retention is vital to your business:

- Poor customer service would convince 39% of people never to use a company again and 37% to change suppliers, according to research from New Voice Media.
- Temkin Group found that 77% of customers would recommend a business to a friend after having just one positive experience.
- It takes 12 positive customer experiences to make up for one negative experience, according to Ruby Newell-Legner's "Understanding Customers."

Measuring Customer Retention and Key Metrics

Your customer retention strategies should be guided by data beyond sales numbers that can help quantify your efforts. Let's take a look at some of the key metrics you can use to determine your customer retention rate.

Attrition Rate Formula
The attrition rate is the number of customers a company lost in a specific time frame relative to its existing customer base. To calculate the attrition rate, take the number of customers your business lost by the end of a specified period and divide it by the total number of customers at the beginning of the period.

Customer Retention Rate Formula
To calculate customer retention rate, determine the number of customers you acquired over a specific period. Subtract that figure from the total number of customers at the end of that period. Then you take this number and divide it by the number of customers you had at the beginning of the period.

Repeat Customer Rate
The repeat customer rate measures the chances an existing customer will make more than one purchase.

To calculate your repeat customer rate, take the number of customers who made more than one purchase and divide it by the total number of unique customers.

Purchase Frequency
The purchase frequency formula is related to the repeat customer rate and represents the average number of orders placed by each customer. Take the same period used for your repeat customer rate (such as a month or quarter) and divide the total number of orders by the total number of unique customers.

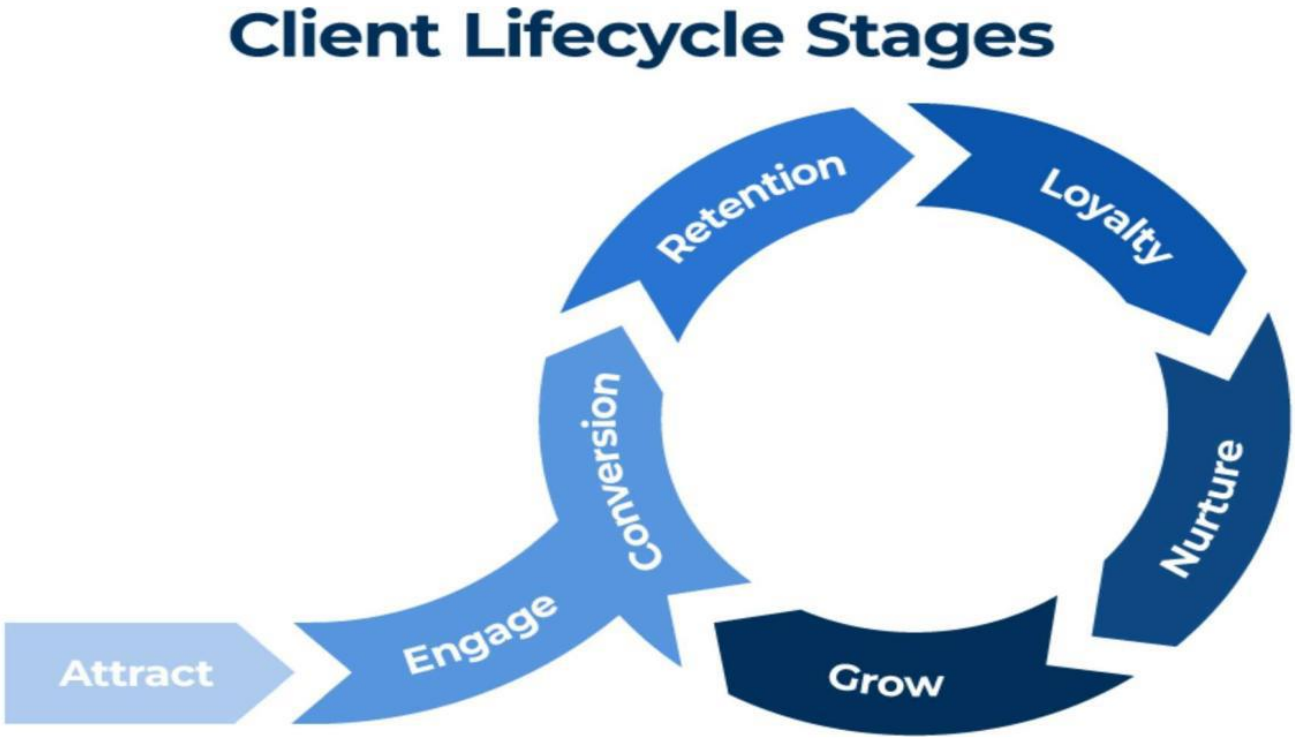
Average Order Value (AOV)
AOV shows the average amount spent per purchase. Using the same period for the repeat purchase rate, divide your total annual revenue by the number of orders processed.

5 Strategies to Improve Customer Retention

Here are some practical ideas for improving customer retention:

1. **Engage with customers:** Look at your marketing channels and identify the best ways to engage with your clients. Do they respond best to social media, email marketing, online events or something else? Let customers weigh in on upcoming products and services, so they feel like they're part of the brand.
2. **Reduce friction in the purchase process:** The fewer obstacles or challenges customers face when purchasing your product or service, the better. When it comes to ecommerce, fast page load times and a fast, simple checkout experience is critical. In a store, eliminate friction by making sure a staff member is always available to help a customer when they're ready to check out.
3. **Improve customer support:** Offer multiple ways for customers to reach you. This can include live chat on your website, a dedicated telephone number, email, social media and a comprehensive FAQ page online. Additionally, you want to ensure fast response rates. Training your staff well and measuring their performance with benchmarks will help you meet customer expectations for communication.

4. **Create a community:** Having exclusive membership programs or forums where your company representatives and customers can interact with each other will help generate more brand loyalty and buzz. Other ideas include giving discount codes to loyal customers and creating referral programs that offer current clients an incentive.
5. **Start a loyalty program:** Loyalty programs can be a great way to motivate existing customers to make additional purchases and at a higher frequency. Ensure that your loyalty program has rewards that existing customers would find valuable, like free products or significant discounts.



The above figure shows the lifecycle of a consumer.

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

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.

Exploratory Data Analysis (EDA)

First I have imported all the libraries in jupyter notebook and then I have imported the data from the excel file.

```
In [177]:
import pandas as pd
import numpy as np
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
import matplotlib.pyplot as plt
import seaborn as sns
import pickle

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

```
In [178]:
data=pd.read_excel(r"C:\Users\sanka\Downloads\Compressed\Customer_retention_dataset-_2\Customer_retention_dataset\customer_r
```

Then as the number of columns are more than 70 so all the columns are not visible after using head() method. So I have used set_options() method to display all the columns and rows.

```
#displaying all the columns in the dataset
pd. set_option("display.max_columns", None)

#displaying all the rows in the dataset
pd.set_option("display.max_rows",None)
```

```
data.head()
```

Initial observations from the dataset

1.The dataset is consist of columns having only object data of some e commerce websites.

2.the column names are very big here, we need to rename the columns with small names for further analysis.

3.The dataset is consist of both numerical and catagorical data columns. We need to encoe the columns having catagorical data.

4.The dataset contains both utilitarian value and hedonistic values. These values effects the repeat purchase intention of the customers.

Then I have printed the number of rows and number of columns using data.shape. and after that I have checked for the null values.

```
#finding number of rows and columns
data.shape
```

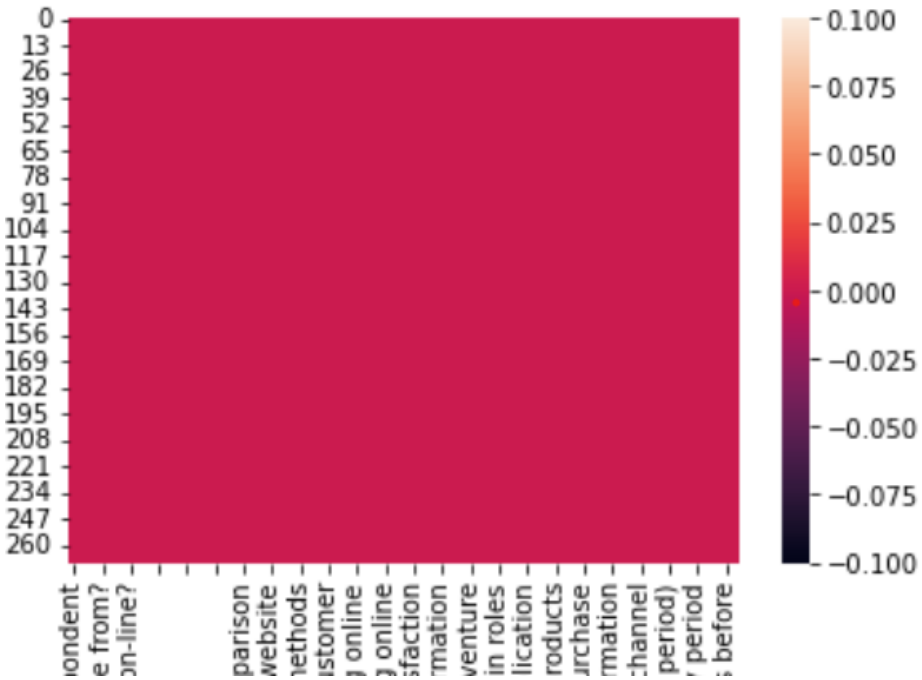
```
(269, 71)
```

The dataset has 269 rows and 71 columns.

```

#finding the null values in each column
data.isna().sum()
```

To get a more clear view about the null values I have used heatmap plot.

```
# Let's visualize the null values clearly
sns.heatmap(data.isna())
plt.show()
```



I have found out that there was no null values present in the dataset.

Then I have printed the number of columns and I have observed that the column names were very big so I have renamed the columns.

```
#no. of columns in the dataset
data.columns
```

The column names are very big, lets shorten the titles of the columns with some proper names.

```
# Renaming the column names for better understanding
df1 = ['Gender','Age','Shopping_City','Pin_code','Shopping_Since','Shopping_Frequency','Internet_Accessibility','Device_Used
       'Screen_Size','Operating_System', 'Browser_Used','Channel_First_Used','Login_Mode','Time_Explored','Payment_Mode'
      'Abandon_Reason','Content_Readability','Similar_Product_Info','Seller_Product_Info','Product_Info_Clarity','Naviga
      'Loading_Processing_Speed','User_Friendly_Interface','Convenient_Payment_Mode','Timely_Fulfilment_Trust','Customer
      'Customer_Privacy_Guarantee','Various_Channel_Responses','Benefits','Enjoy','Convenience','Return_Replacement_Poli
      'Info_Satisfaction','Site_Quality_Satisfaction','Net_Benefit_Satisfaction','Trust','Product_Several_Category','Rel
      'Patronizing_Convenience','Adventure_Sense','Social_Status','Gratification','Role_Fulfilment','Money_Worthy','Shop
       'Visually_Appealing_WebApp','Product_Variety','Complete_Product_Info','Fast_WebApp','Reliable_WebApp','Quick_Purc
       'Fast_Delivery','Customer_Privacy_Info','Financial_Security_Info','Perceived_Trustworthiness','Multichannel_Assis
       'Late_Price_Declare','Long_Loading_Time','Limited_Payment_Mode','Late_Delivery','WebApp_Design_Change','Page_Disr
       'Recommendation']

data.columns = df1
```

After that I have checked for different types of data types present in the dataset and I have found out that there are 70 columns having object(string) datatype values. We need to change them into numerical values.

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 269 entries, 0 to 268
Data columns (total 71 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   Gender                      269 non-null    object
 1   Age                         269 non-null    object
 2   Shopping_City               269 non-null    object
 3   Pin_code                    269 non-null    int64
 4   Shopping_Since              269 non-null    object
 5   Shopping_Frequency          269 non-null    object
 6   Internet_Accessibility      269 non-null    object
 7   Device_Used                 269 non-null    object
 8   Screen_Size                 269 non-null    object
 9   Operating_System            269 non-null    object
 10  Browser_Used                269 non-null    object
 11  Channel_First_Used          269 non-null    object
 12  Login_Mode                  269 non-null    object
 13  Time_Explored               269 non-null    object
 14  Payment_Mode                269 non-null    object
```

Then I have used describe() method to know about the statistics of the dataset. In this dataset only one column with continuous values is present.

In [190]:
```
#understanding the data at high level i.e. checking the statistics of the data
data.describe()
```

Out[190]:

	Pin_code
count	269.000000
mean	220465.747212
std	140524.341051
min	110008.000000
25%	122018.000000
50%	201303.000000
75%	201310.000000
max	560037.000000

Observations from the statistics of the data

1. Pin_code is the only numerical column present in the dataset.

2. Its min value is 110008.0 and max value is 560037.0. even if pincode column has numerical data, but those are only pincodes. so, we can not get anything from the mean and the standard deviation of the column.

Then I have found out the number of unique values in each column.

```
# lets check the number of unique values in each column
data.nunique().to_frame("Unique values")
```

	Unique values
Screen_Size	4
Operating_System	3
Browser_Used	4
Channel_First_Used	3
Login_Mode	5
Time_Explored	5
Payment_Mode	3
Abandon_Frequency	4
Abandon_Reason	5

We have found out the number of unique values in each column , now lets find out the type of catagorical values in each column for selecting the proper encoding technique.

```
# lets check the value counts of each type in each column
for d in data.columns:
        print(data[d].value_counts())
        print('-'*200)
```

```
Female    181
Male       88
Name: Gender, dtype: int64
----------------------------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------
31-40 years           81
21-30 years           79
41-50 yaers           70
Less than 20 years    20
51 years and above    19
Name: Age, dtype: int64
----------------------------------------------------------------------------------------------------------------------------
```

Observation after finding out the value counts.

1. In Shopping_Frequency column we have, 41 times and above having 47 values and 42 times and above having 6 values.They can be consider as one unit.

2. In Internet_Accessibility column we have, Mobile internet having 142 values and Mobile Internet having 47 values.They can be consider as one unit.

Then I have removed the duplicate values in the above columns.

```
#replacing duplicate values

# Replacing duplicate values in  Shopping_Frequency column
data["Shopping_Frequency"]=data["Shopping_Frequency"].replace('42 times and above','41 times and above')

# Replacing duplicate values in  Internet_Accessibility column
data["Internet_Accessibility"]=data["Internet_Accessibility"].replace('Mobile internet','Mobile Internet')
```

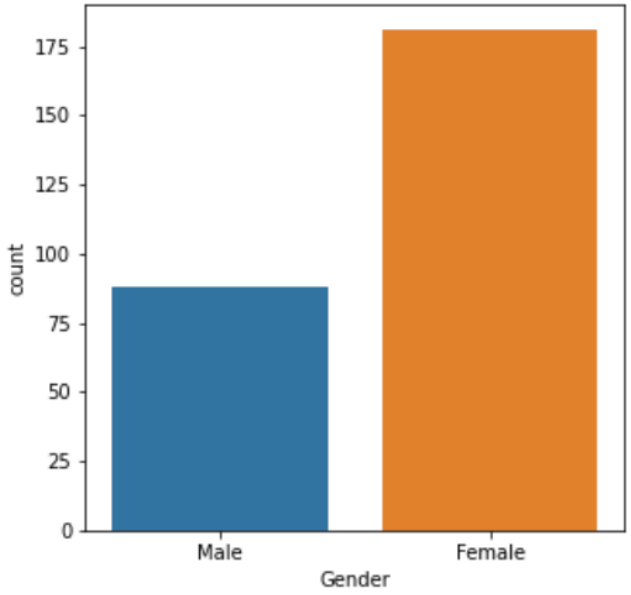
After that I have plot the value counts of all the categorical columns using countplot.

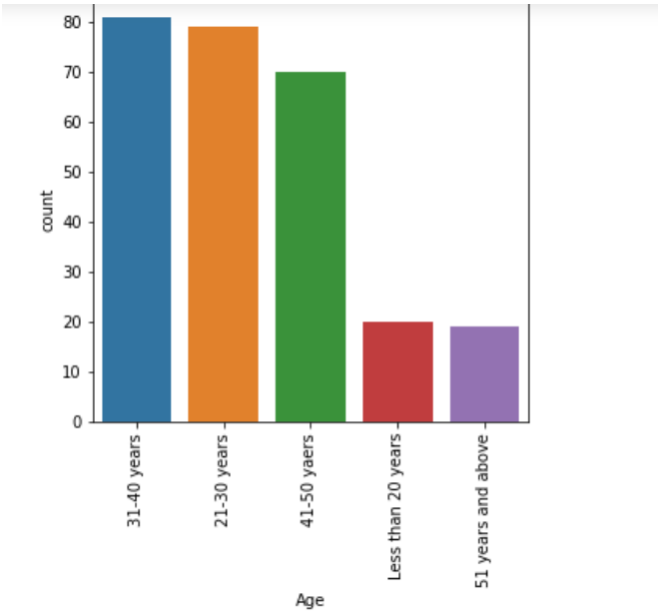
```
def value_counts(column):
    counts=len(data[column].value_counts())
    if counts<5:
        plt.figure(figsize=(5,5))
    elif counts<10:
        plt.figure(figsize=(5,5))
        plt.xticks(rotation=90)
    elif counts<20:
        plt.figure(figsize=(5,5))
        plt.xticks(rotation=90)
    else:
        plt.figure(figsize=(5,5))
        plt.xticks(rotation=90)
    sns.countplot(x=column,data=data)
    plt.show()
    print("*"*200)
```

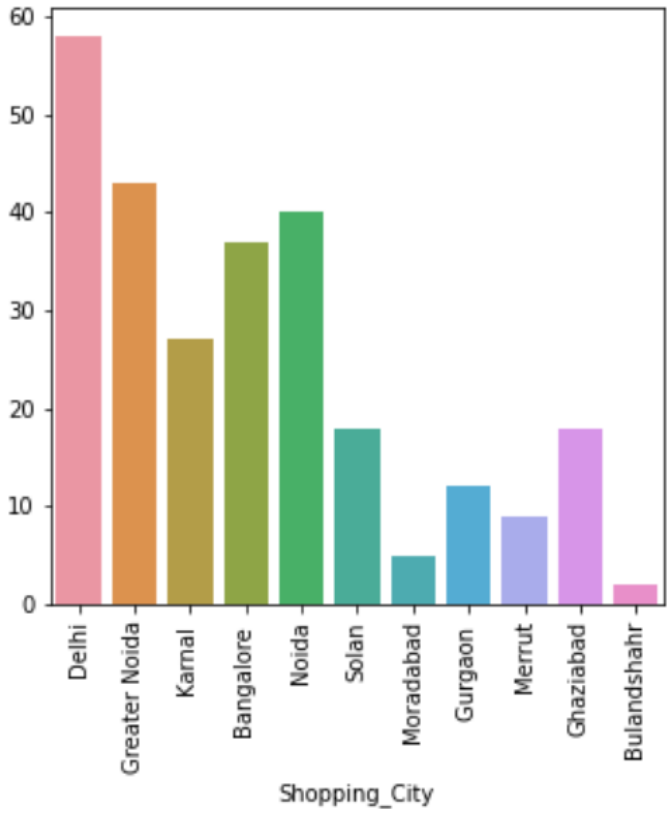
```
CR1=data.iloc[:,[0,1,2,4,5,6,7,8,10,12,13,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,4
for i in CR1:
    print(CR1[i].value_counts())
    value_counts(i)
```

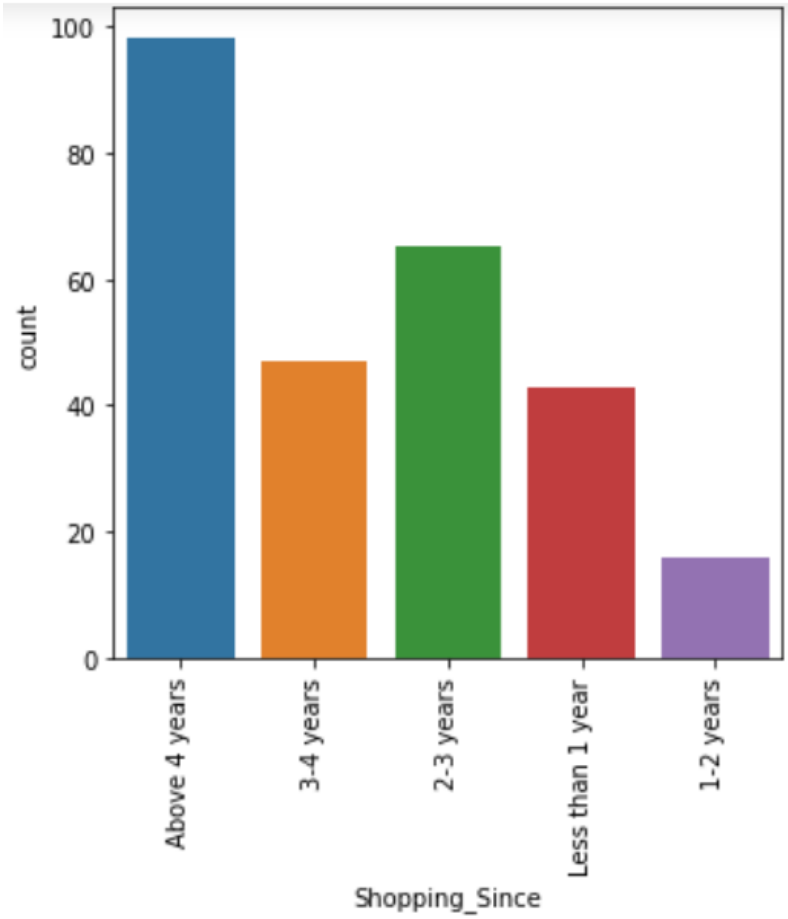
Plots for all the unique values in each column.

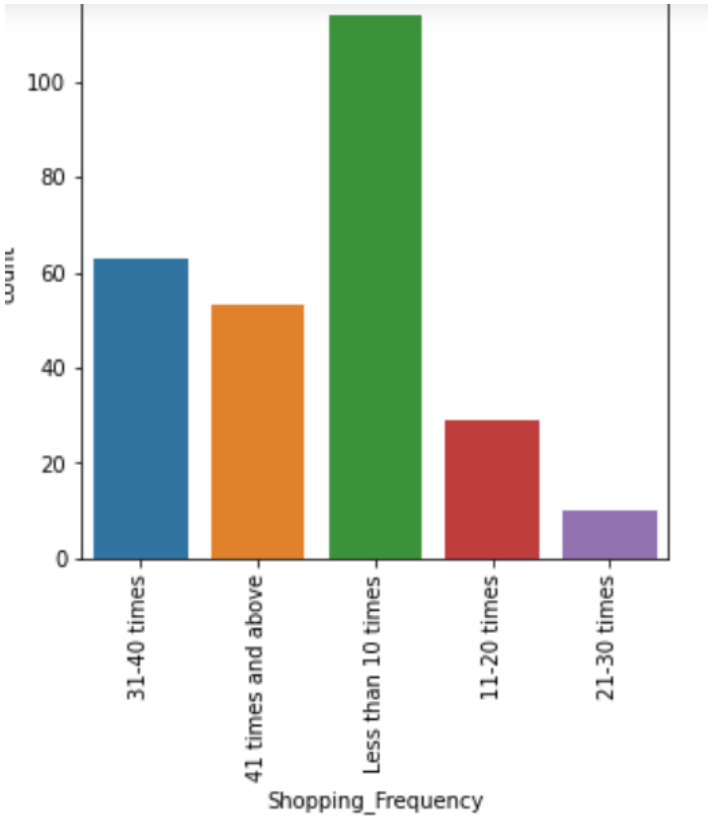
```
Female    181
Male       88
Name: Gender, dtype: int64
```

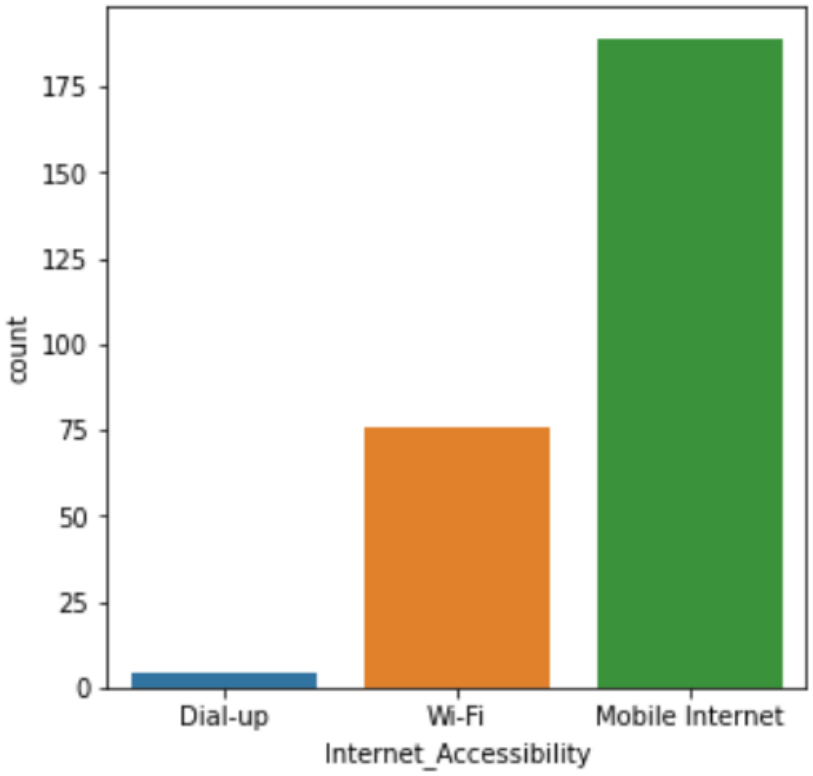


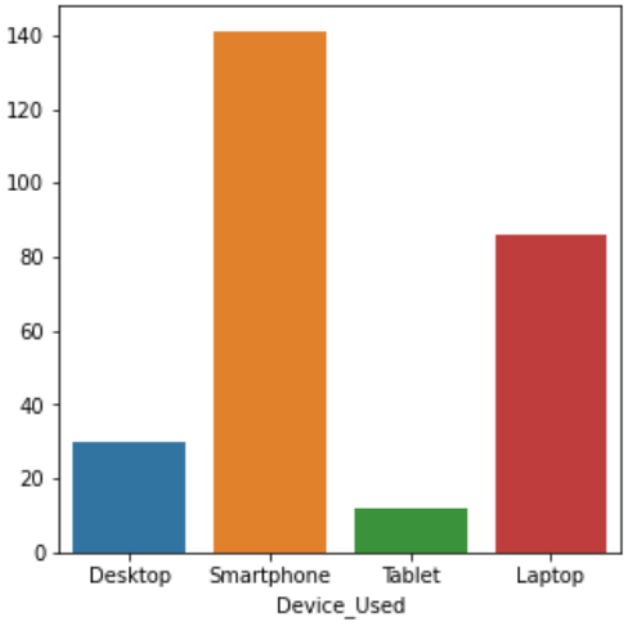


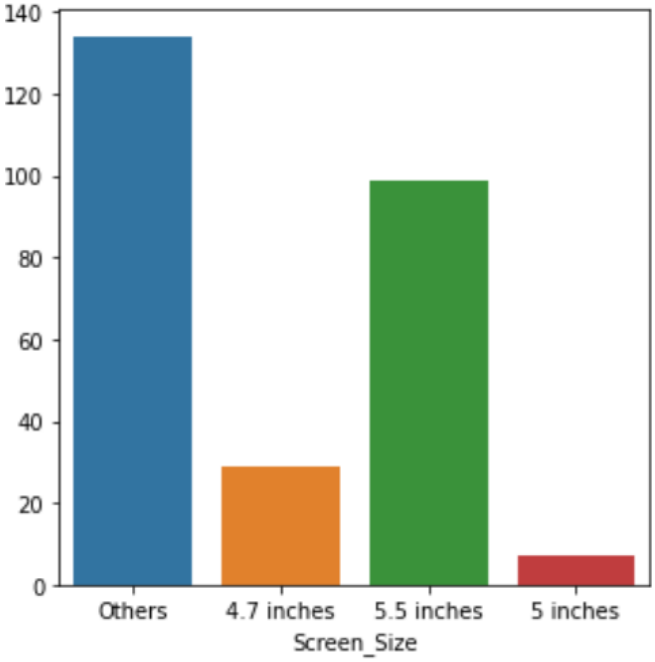

60
50
40
30
20
10
0
Delhi
Greater Noida
Karnal
Bangalore
Noida
Solan
Moradabad
Gurgaon
Merrut
Ghaziabad
Bulandshahr
Shopping_City

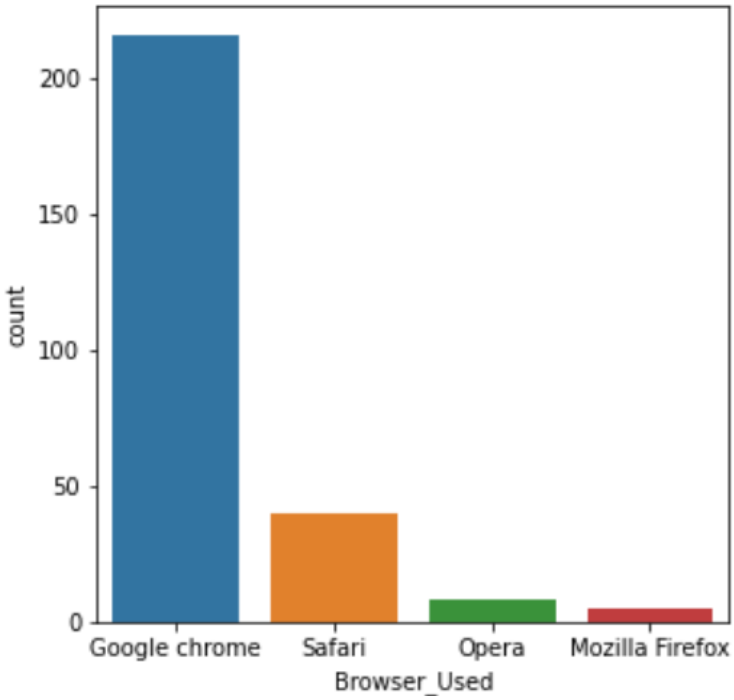

100
80
60
40
20
0
count
Above 4 years
3-4 years
2-3 years
Less than 1 year
1-2 years
Shopping_Since

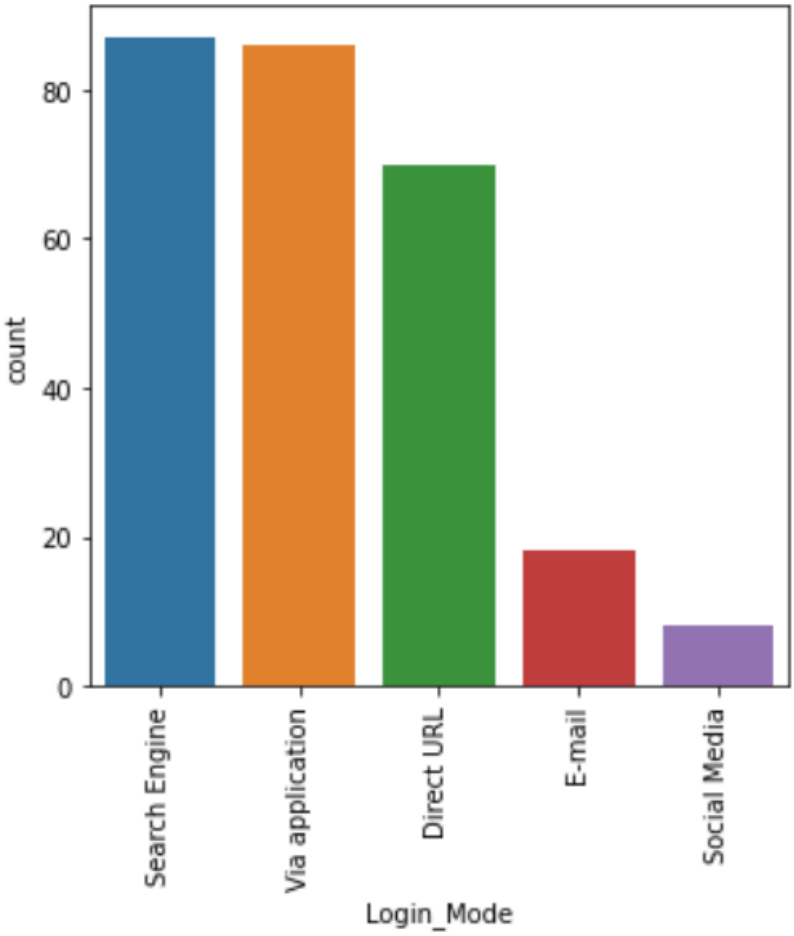

100
80
60
40
20
0
31-40 times
41 times and above
Less than 10 times
11-20 times
21-30 times
Shopping_Frequency

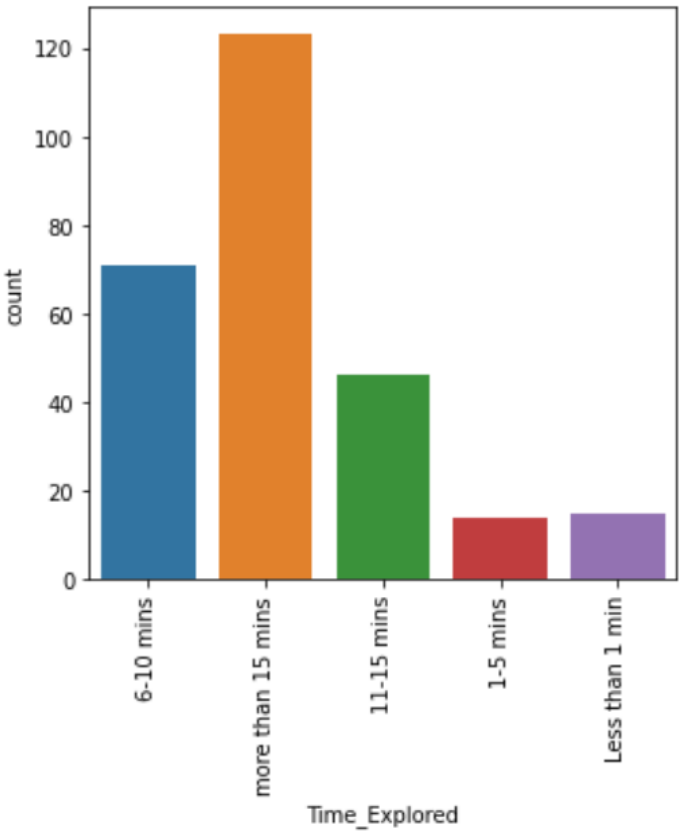

175
150
125
100
75
50
25
0
count
Dial-up
Wi-Fi
Mobile Internet
Internet_Accessibility

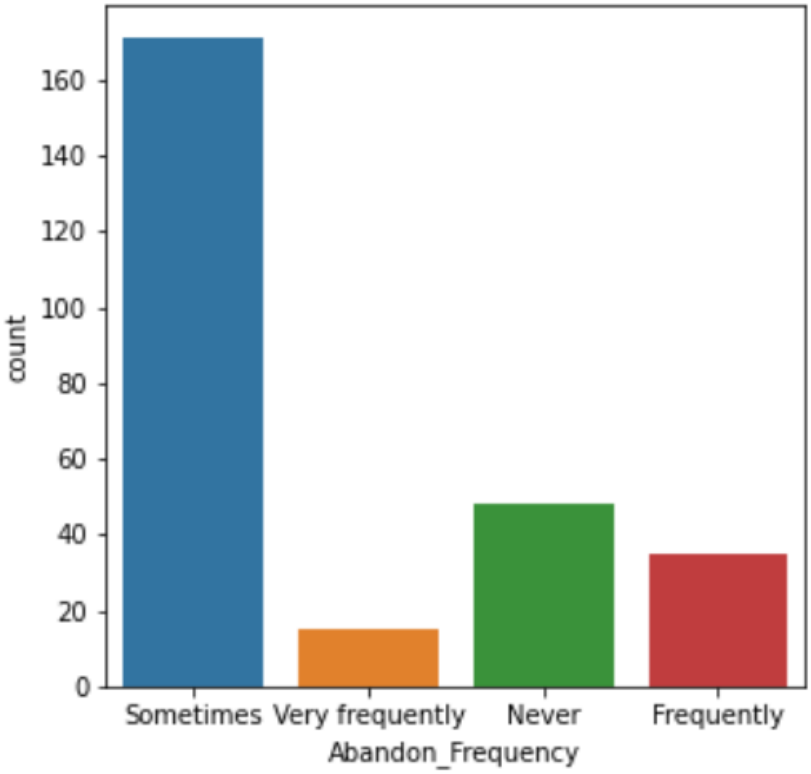

140
120
100
80
60
40
20
0
Desktop
Smartphone
Tablet
Laptop
Device_Used

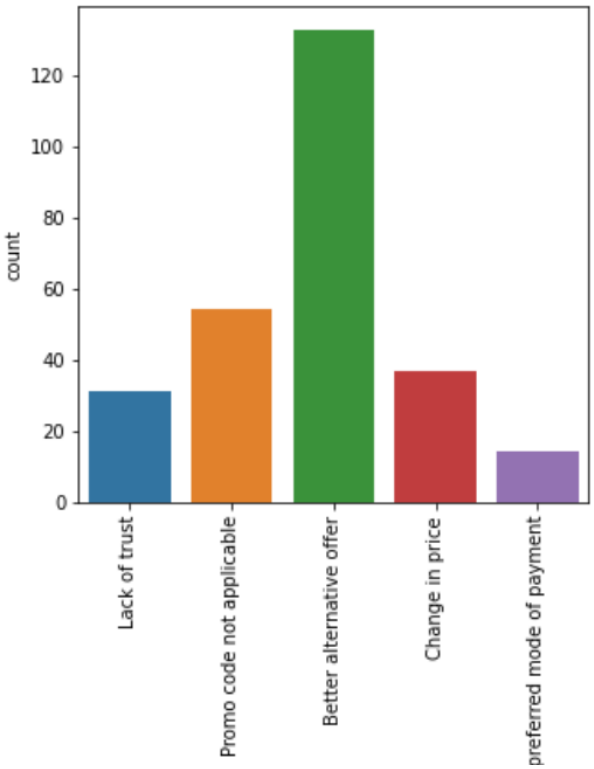

140
120
100
80
60
40
20
0
Others
4.7 inches
5.5 inches
5 inches
Screen_Size

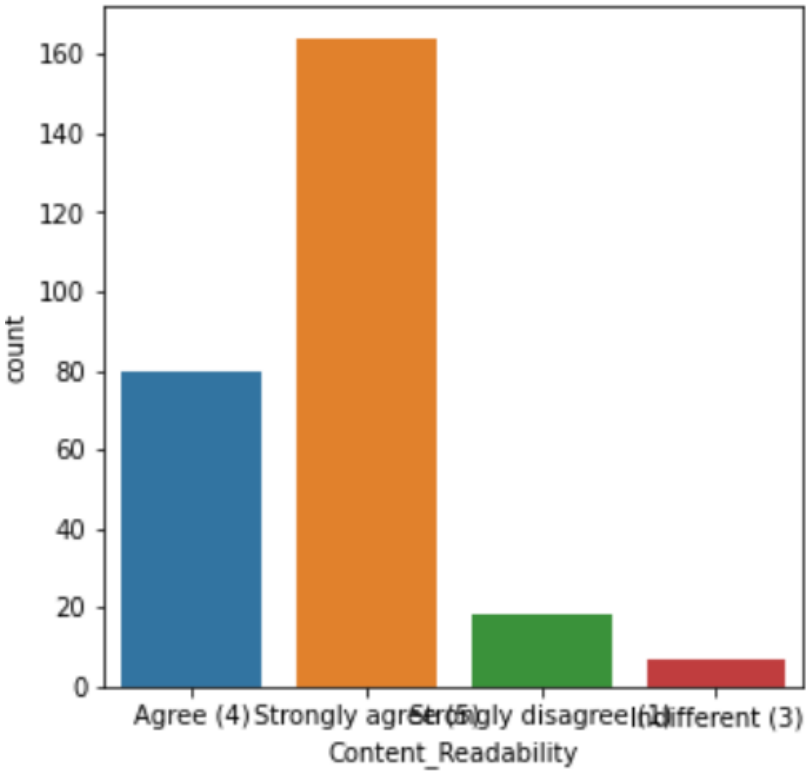

200
150
100
50
0
count
Google chrome
Safari
Opera
Mozilla Firefox
Browser_Used

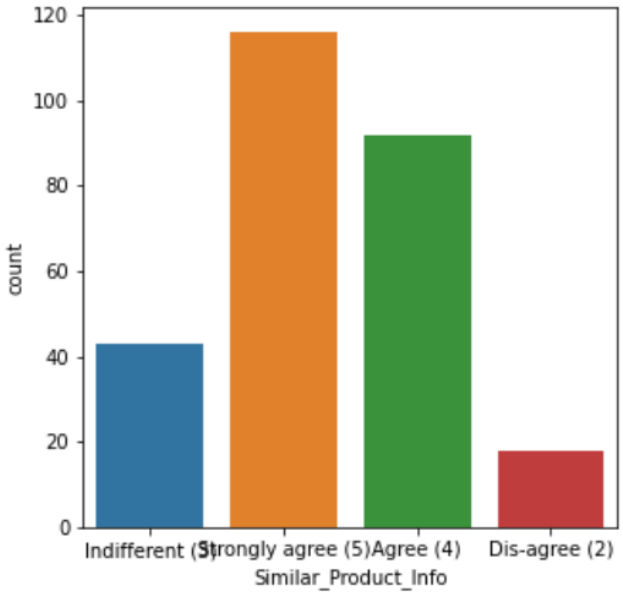

80
60
40
20
0
count
Search Engine
Via application
Direct URL
E-mail
Social Media
Login_Mode

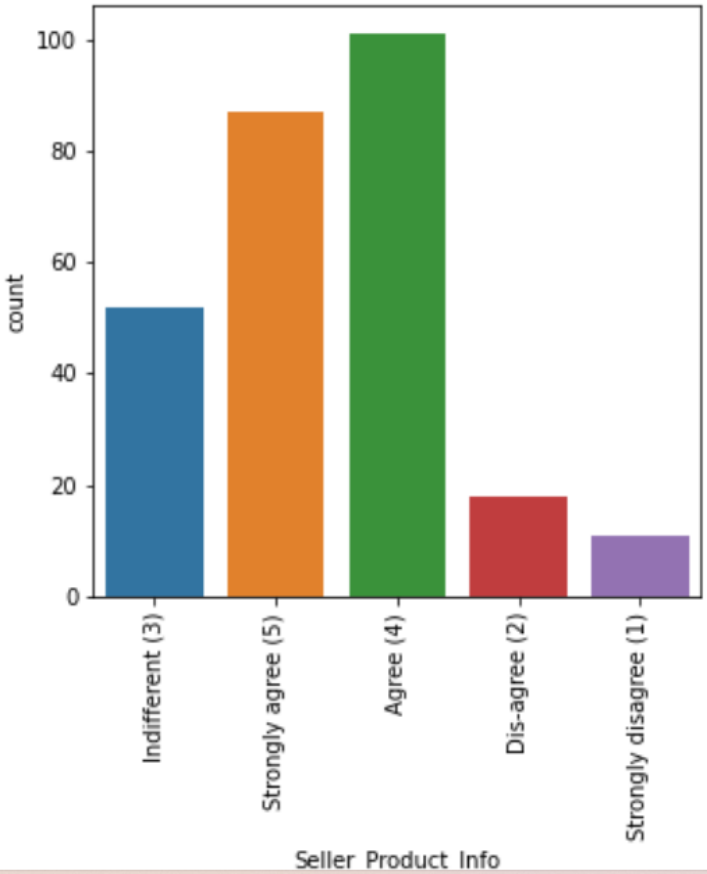

120
100
80
60
40
20
0
count
6-10 mins
more than 15 mins
11-15 mins
1-5 mins
Less than 1 min
Time_Explored

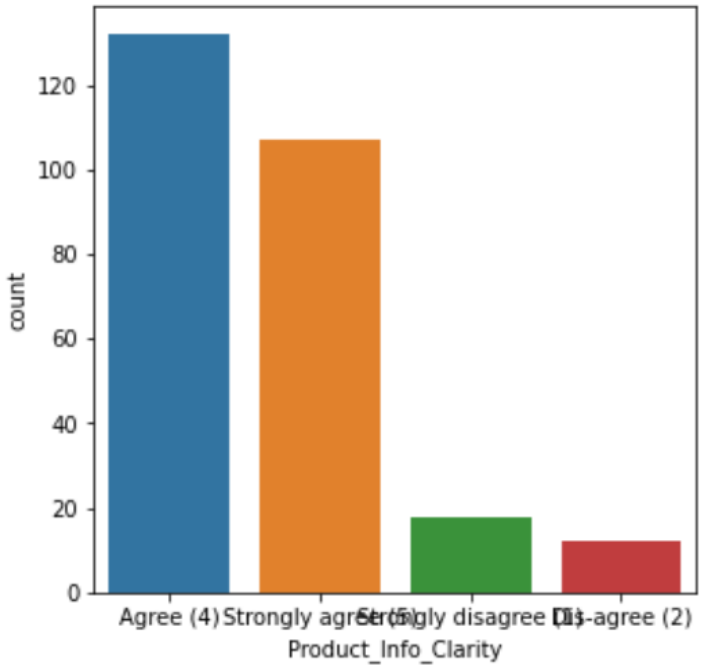

160
140
120
100
80
60
40
20
0
count
Sometimes
Very frequently
Never
Frequently
Abandon_Frequency

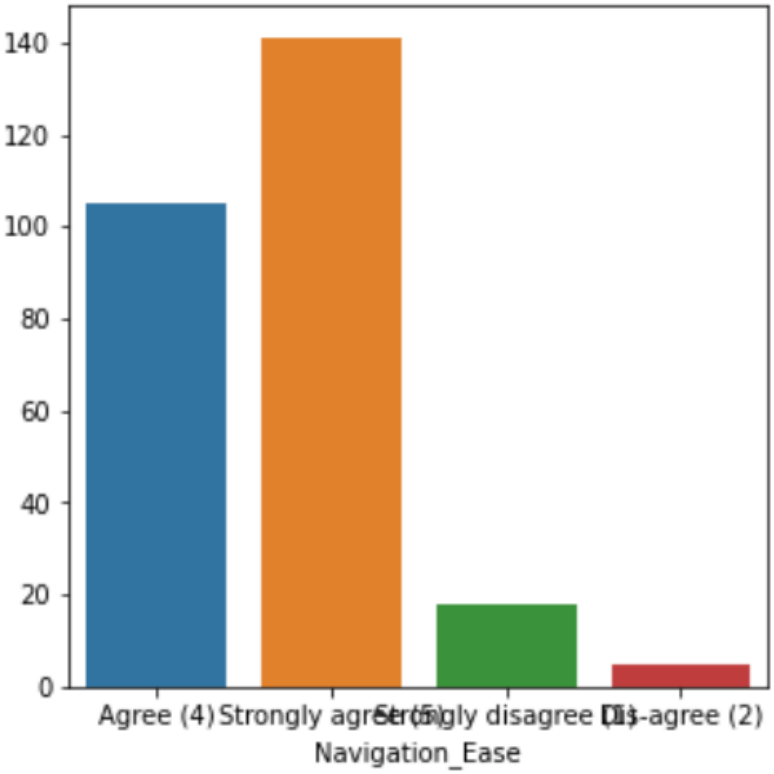

count
120
100
80
60
40
20
0
Lack of trust
Promo code not applicable
Better alternative offer
Change in price
preferred mode of payment

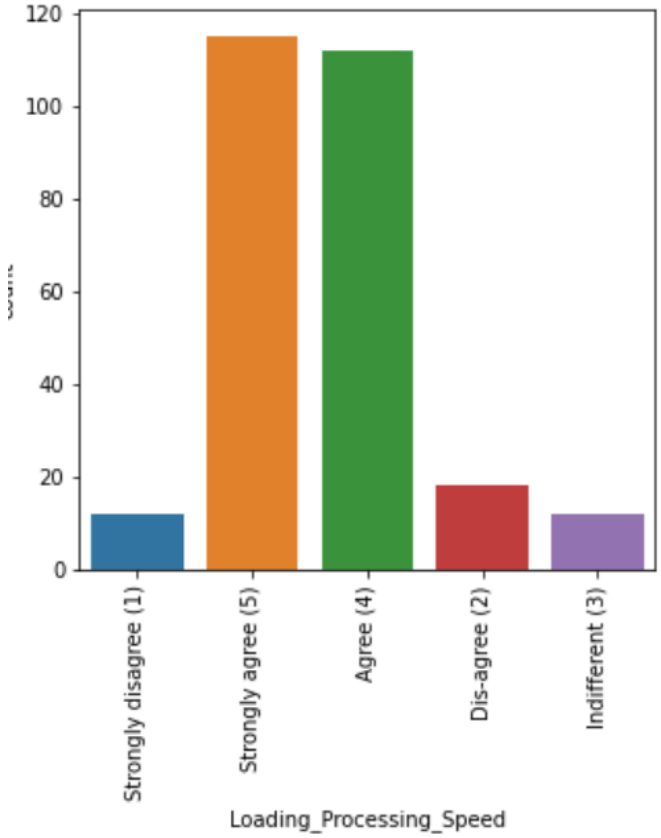

count
160
140
120
100
80
60
40
20
0
Agree (4)
Strongly agree (5)
Strongly disagree (1)
Indifferent (3)
Content_Readability

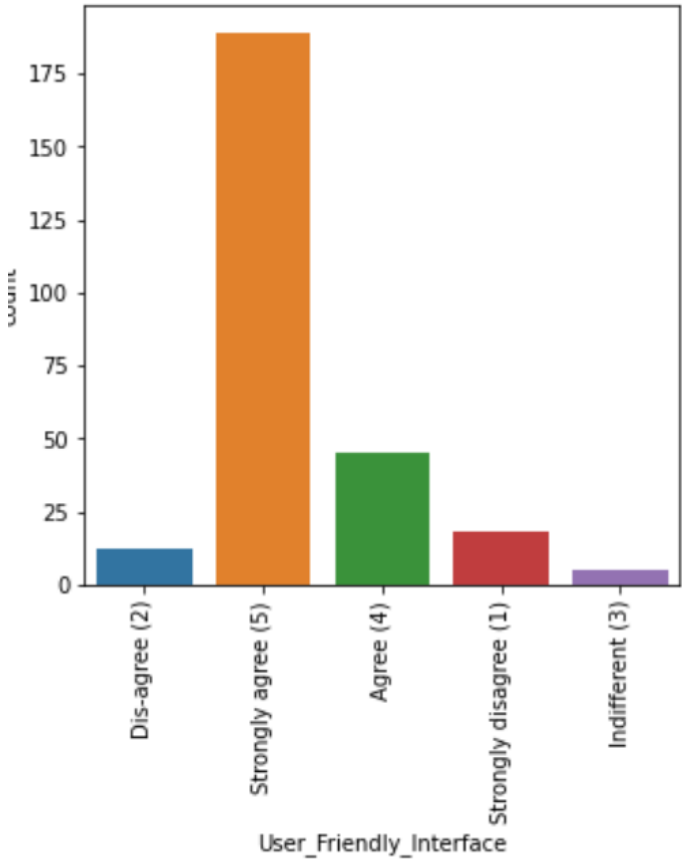

120
100
80
60
40
20
0
count
Indifferent (3)
Strongly agree (5)
Agree (4)
Dis-agree (2)
Similar_Product_Info

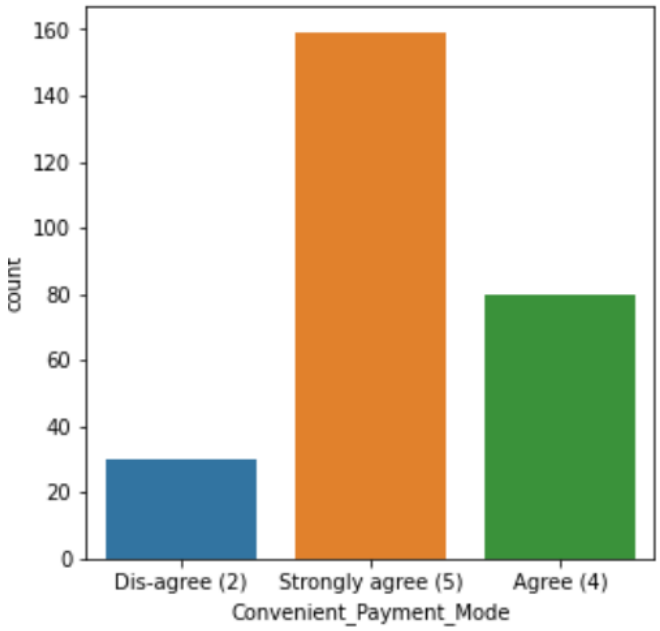

100
80
60
40
20
0
count
Indifferent (3)
Strongly agree (5)
Agree (4)
Dis-agree (2)
Strongly disagree (1)
Seller_Product_Info

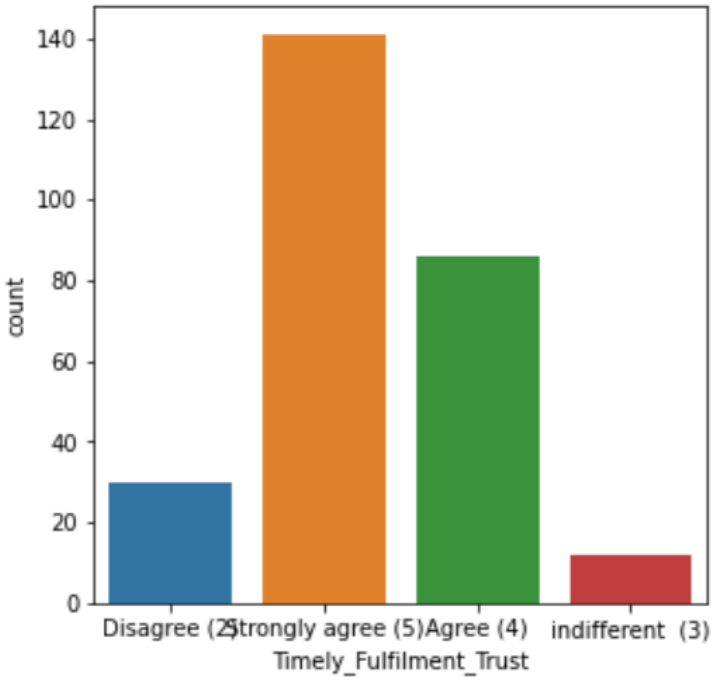

120
100
80
60
40
20
0
count
Agree (4)
Strongly agree (5)
Strongly disagree (1)
Dis-agree (2)
Product_Info_Clarity

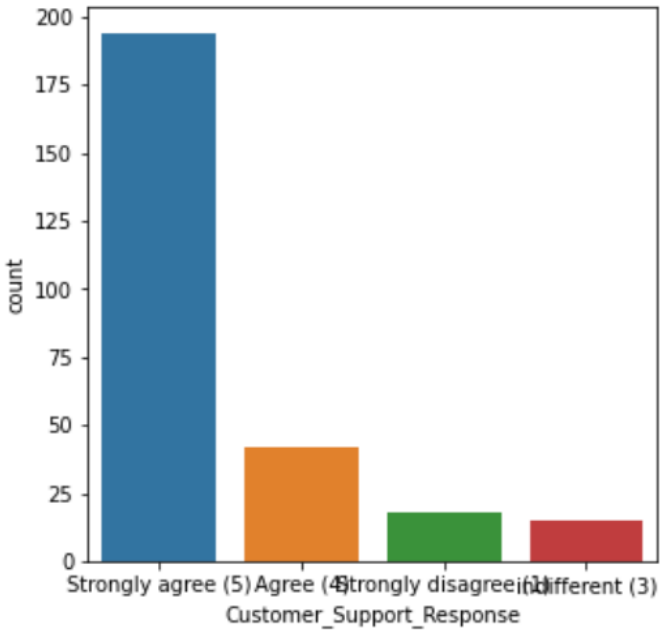

140
120
100
80
60
40
20
0
Agree (4)
Strongly agree (5)
Strongly disagree (1)
Dis-agree (2)
Navigation_Ease

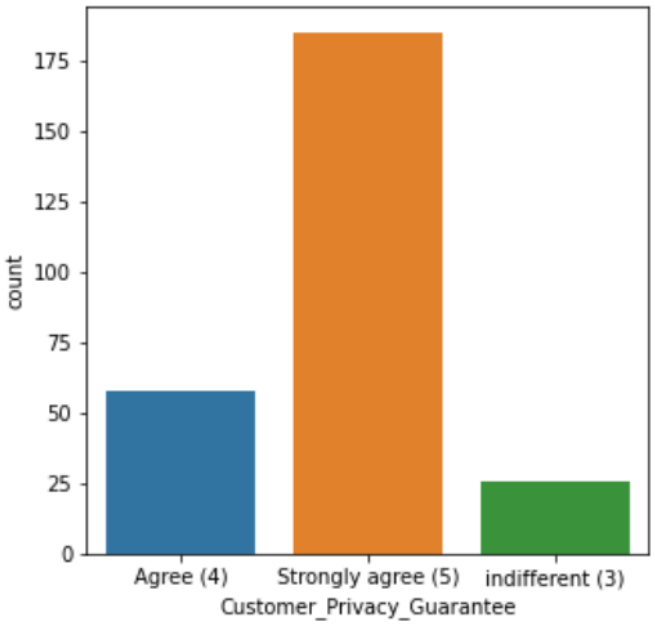

120
100
80
60
40
20
0
Strongly disagree (1)
Strongly agree (5)
Agree (4)
Dis-agree (2)
Indifferent (3)
Loading_Processing_Speed

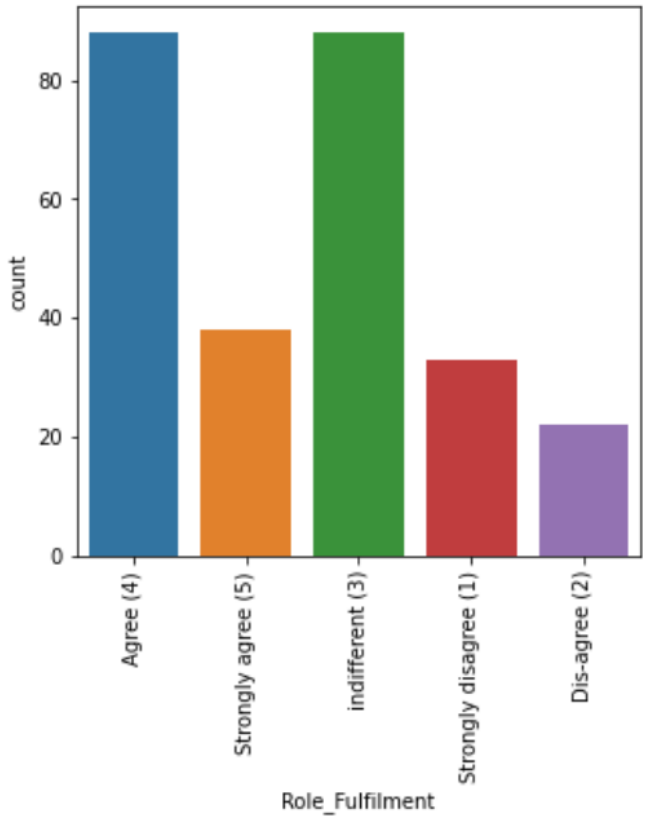

175
150
125
100
75
50
25
0
Dis-agree (2)
Strongly agree (5)
Agree (4)
Strongly disagree (1)
Indifferent (3)
User_Friendly_Interface

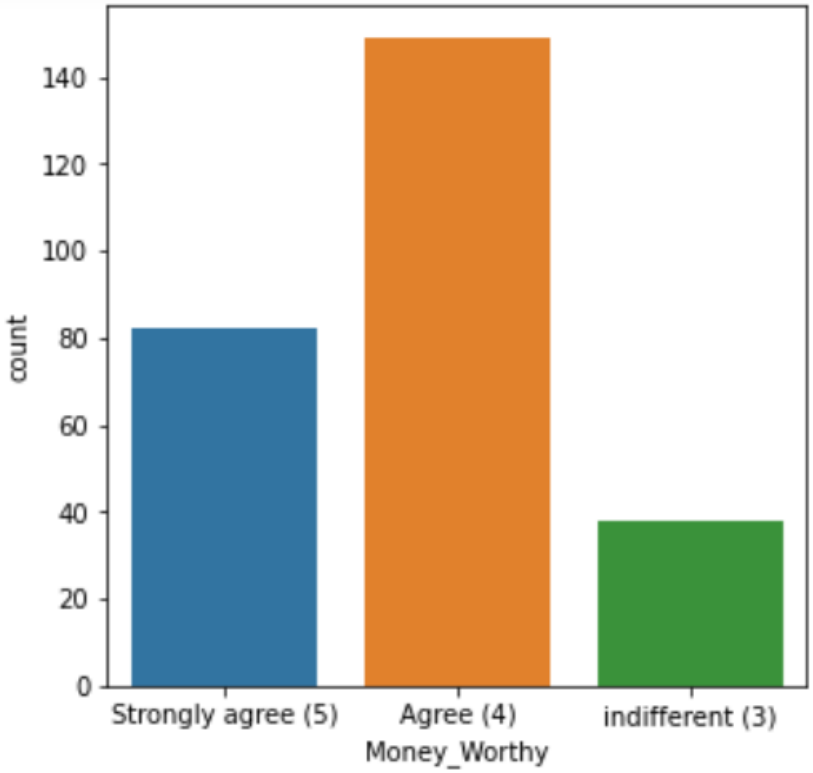

160
140
120
100
80
60
40
20
0
count
Dis-agree (2)
Strongly agree (5)
Agree (4)
Convenient_Payment_Mode

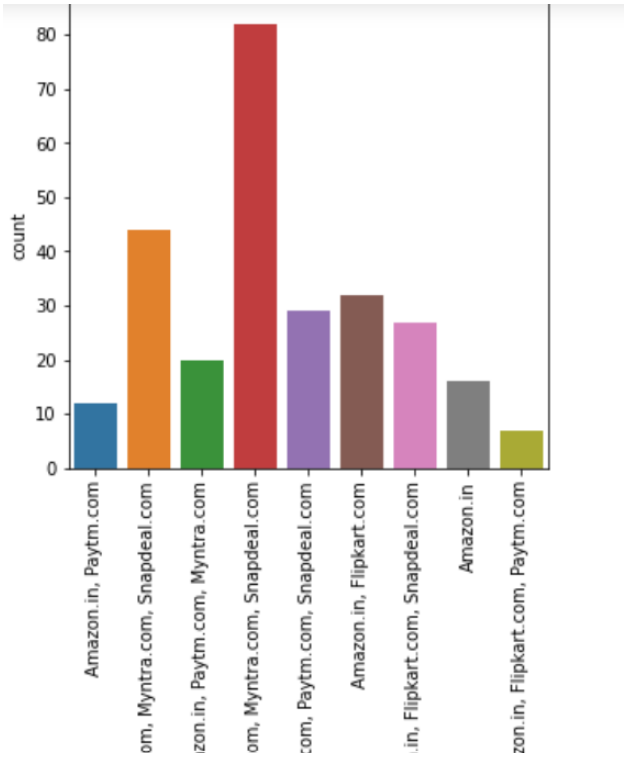

140
120
100
80
60
40
20
0
count
Disagree (2)
Strongly agree (5)
Agree (4)
indifferent (3)
Timely_Fulfilment_Trust

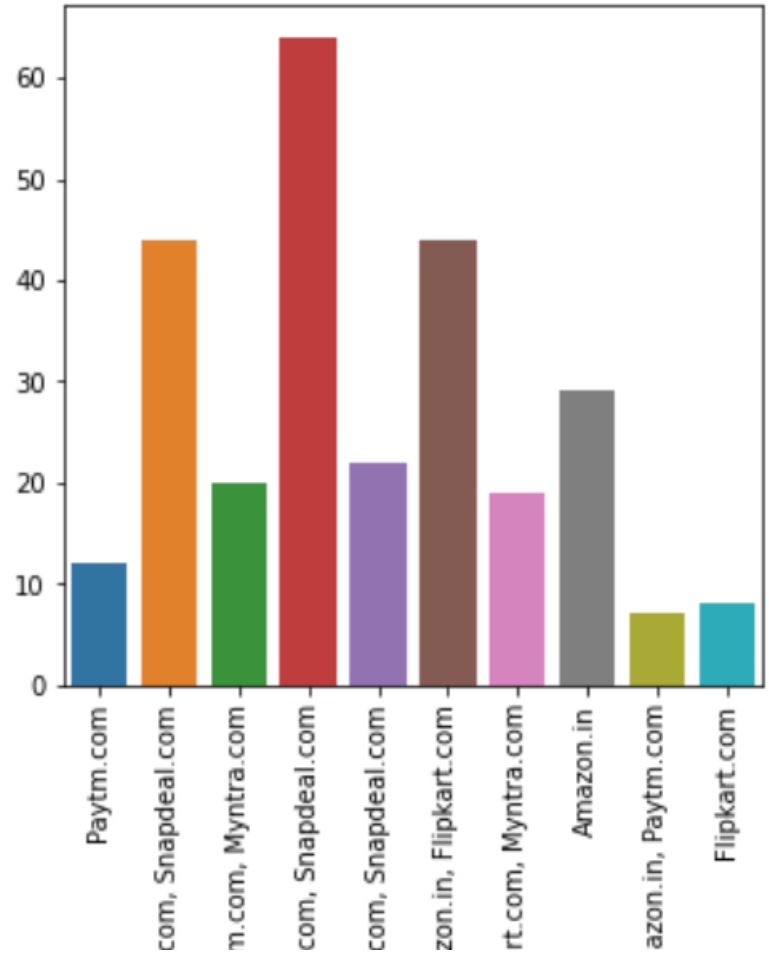

200
175
150
125
100
75
50
25
0
count
Strongly agree (5)
Agree (4)
Strongly disagree (1)
indifferent (3)
Customer_Support_Response

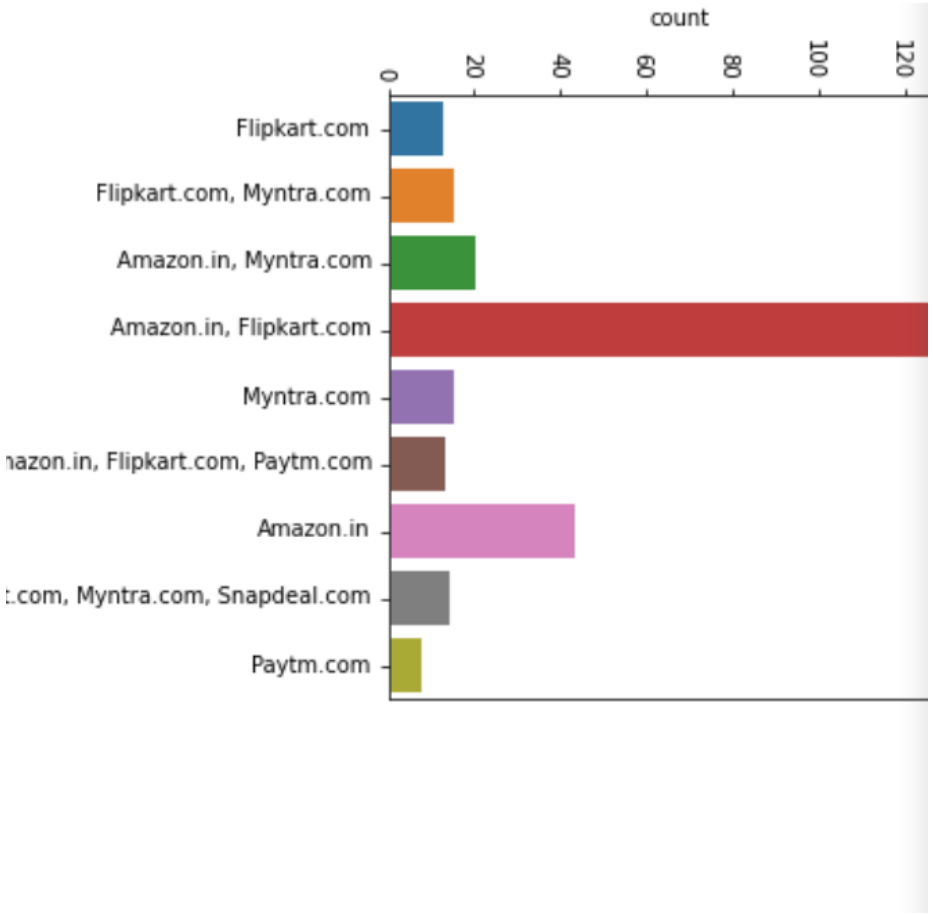

175
150
125
100
75
50
25
0
count
Agree (4)
Strongly agree (5)
indifferent (3)
Customer_Privacy_Guarantee

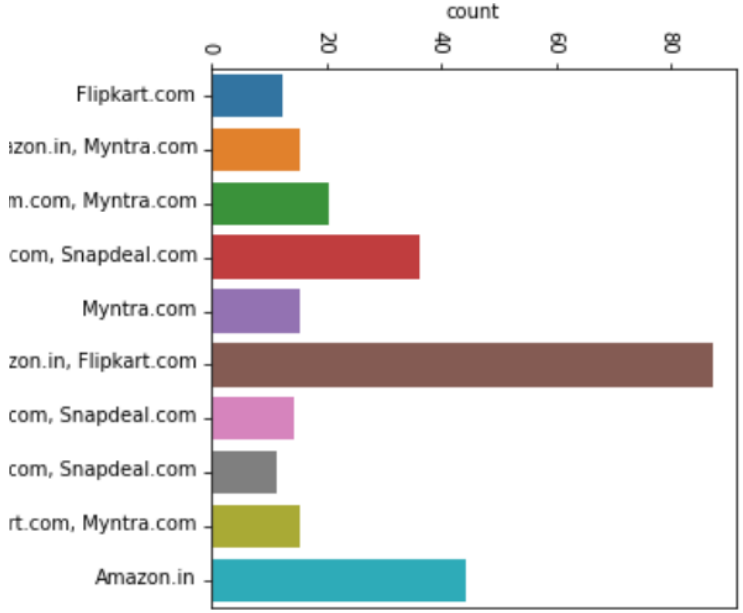

count
0
20
40
60
80
Agree (4)
Strongly agree (5)
indifferent (3)
Strongly disagree (1)
Dis-agree (2)
Role_Fulfilment

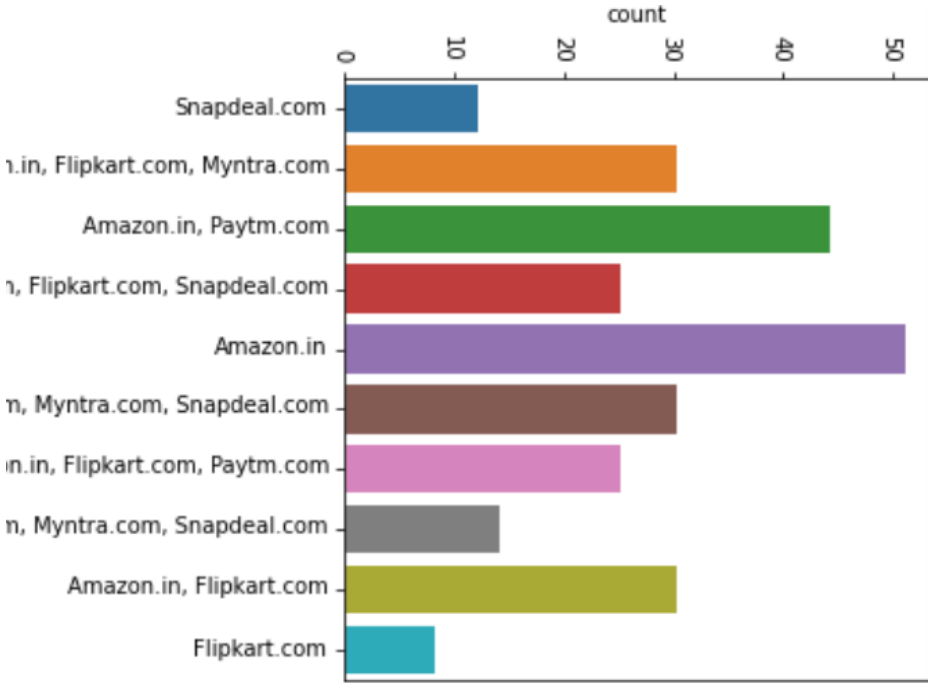

count
0
20
40
60
80
100
120
140
Strongly agree (5)
Agree (4)
indifferent (3)
Money_Worthy

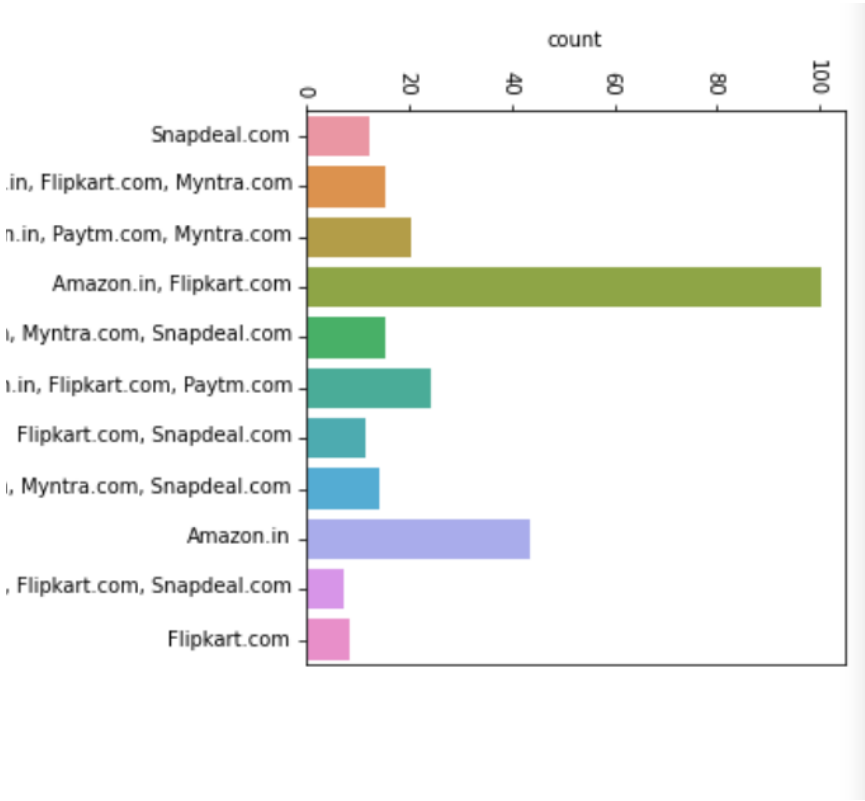

count
0
10
20
30
40
50
60
70
80
Amazon.in, Paytm.com
om, Myntra.com, Snapdeal.com
zon.in, Paytm.com, Myntra.com
om, Myntra.com, Snapdeal.com
com, Paytm.com, Snapdeal.com
Amazon.in, Flipkart.com
in, Flipkart.com, Snapdeal.com
Amazon.in
zon.in, Flipkart.com, Paytm.com

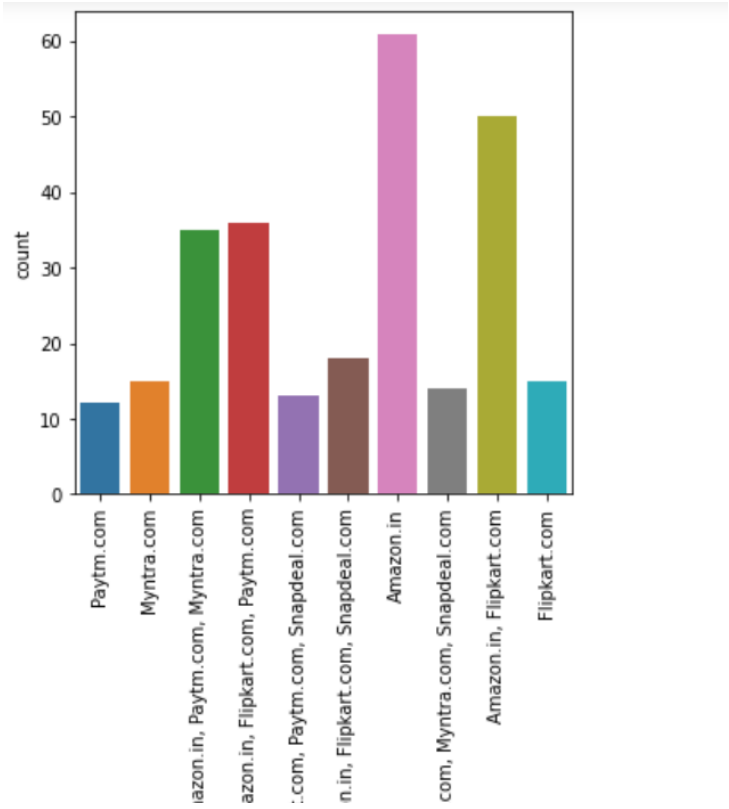

0
10
20
30
40
50
60
Paytm.com
com, Snapdeal.com
m.com, Myntra.com
com, Snapdeal.com
com, Snapdeal.com
zon.in, Flipkart.com
rt.com, Myntra.com
Amazon.in
azon.in, Paytm.com
Flipkart.com

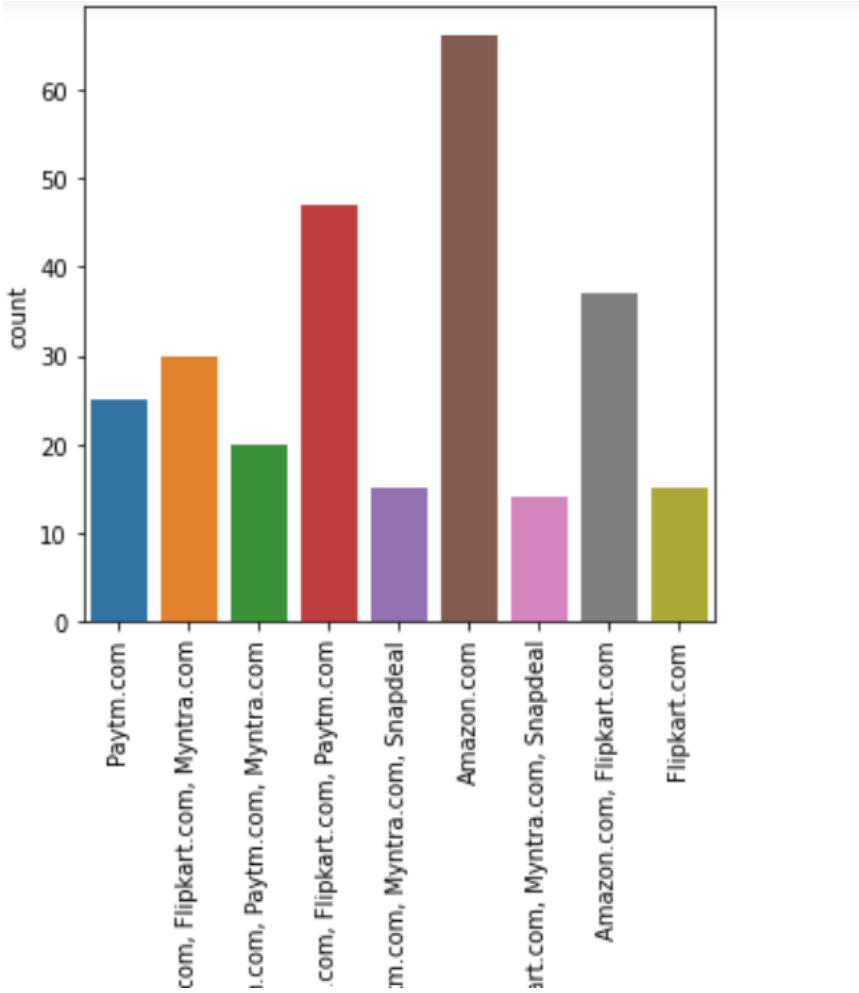

count
0
20
40
60
80
100
120
Flipkart.com
Flipkart.com, Myntra.com
Amazon.in, Myntra.com
Amazon.in, Flipkart.com
Myntra.com
hazon.in, Flipkart.com, Paytm.com
Amazon.in
t.com, Myntra.com, Snapdeal.com
Paytm.com

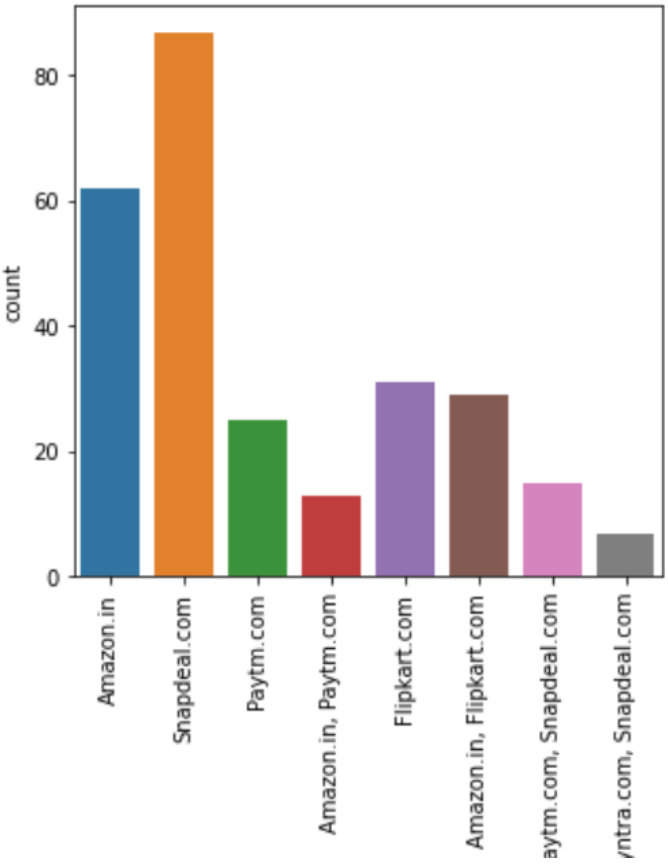

count
0
20
40
60
80
Flipkart.com
zon.in, Myntra.com
m.com, Myntra.com
com, Snapdeal.com
Myntra.com
zon.in, Flipkart.com
com, Snapdeal.com
com, Snapdeal.com
rt.com, Myntra.com
Amazon.in

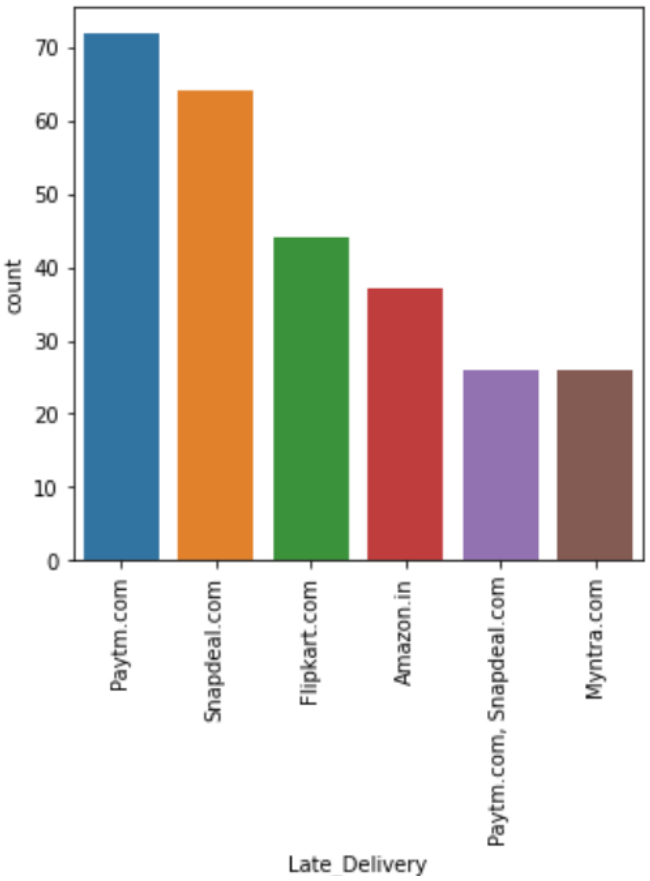

count
0
10
20
30
40
50
Snapdeal.com
ı.in, Flipkart.com, Myntra.com
Amazon.in, Paytm.com
ı, Flipkart.com, Snapdeal.com
Amazon.in
n, Myntra.com, Snapdeal.com
ın.in, Flipkart.com, Paytm.com
n, Myntra.com, Snapdeal.com
Amazon.in, Flipkart.com
Flipkart.com


count
0
20
40
60
80
100
Snapdeal.com
.in, Flipkart.com, Myntra.com
ı.in, Paytm.com, Myntra.com
Amazon.in, Flipkart.com
ı, Myntra.com, Snapdeal.com
ı.in, Flipkart.com, Paytm.com
Flipkart.com, Snapdeal.com
ı, Myntra.com, Snapdeal.com
Amazon.in
, Flipkart.com, Snapdeal.com
Flipkart.com


count
0
10
20
30
40
50
60
Paytm.com
Myntra.com
iazon.in, Paytm.com, Myntra.com
azon.in, Flipkart.com, Paytm.com
:.com, Paytm.com, Snapdeal.com
n.in, Flipkart.com, Snapdeal.com
Amazon.in
com, Myntra.com, Snapdeal.com
Amazon.in, Flipkart.com
Flipkart.com


count
0
10
20
30
40
50
60
Paytm.com
com, Flipkart.com, Myntra.com
.com, Paytm.com, Myntra.com
.com, Flipkart.com, Paytm.com
m.com, Myntra.com, Snapdeal
Amazon.com
art.com, Myntra.com, Snapdeal
Amazon.com, Flipkart.com
Flipkart.com


count
0
20
40
60
80
Amazon.in
Snapdeal.com
Paytm.com
Amazon.in, Paytm.com
Flipkart.com
Amazon.in, Flipkart.com
aytm.com, Snapdeal.com
yntra.com, Snapdeal.com


count
0
10
20
30
40
50
60
70
Paytm.com
Snapdeal.com
Flipkart.com
Amazon.in
Paytm.com, Snapdeal.com
Myntra.com
Late_Delivery

Observations from the countplots

- There were more female customers than male customers.
- People in age group od 21 to 30 years are more active on e commerce sites.
- Most of the customers from the city Delhi, Greater Noida, Noida and Bangalore are used to shop onine and the shopping count is high in these cities.
- Most of the customers found shopping online for more than 4 years and the count is high for the same.
- In last 1 year, most of the customers started purchasing online.
- Most of the customers used Smartphone device to access the online shopping.
- The count is high for others mobile screen size.
- Most of the customers used Google chrome to access the website.
- Most of the customers used Search Engine and Via application to reach the online retail store.
- Many customers took more than 15 mins before making the purchase decision.
- Around 133 customers abandoned their bag due to some better alternative offer.
- Around 77% of the customers agreed that the information on similar product to the one highlighted is important for product comparison.
- About 70% of the customers agreed that complete information on listed seller and product being offered is important for purchase decision.
- 88.84% of the customers agreed that all relevant information on listed products must be stated clearly and only 11% of the customers disagreed with it.
- 91% of the customers agreed that ease of navigation in website helps them more.
- Most of the customers agreed that they have no issues with the loading and processing speed.
- 87% of the customers agreed with user friendly website interface.
- 84% of the customers trusted that the online retail store will fulfill its part of the transaction at the stipulated time.
- 83% of the customers agreed that shopping online is convenient and flexible and 12% of the customers are indifferent which means either they are agreed to this or disagreed and only 5% of the customers completely disagreed with it.

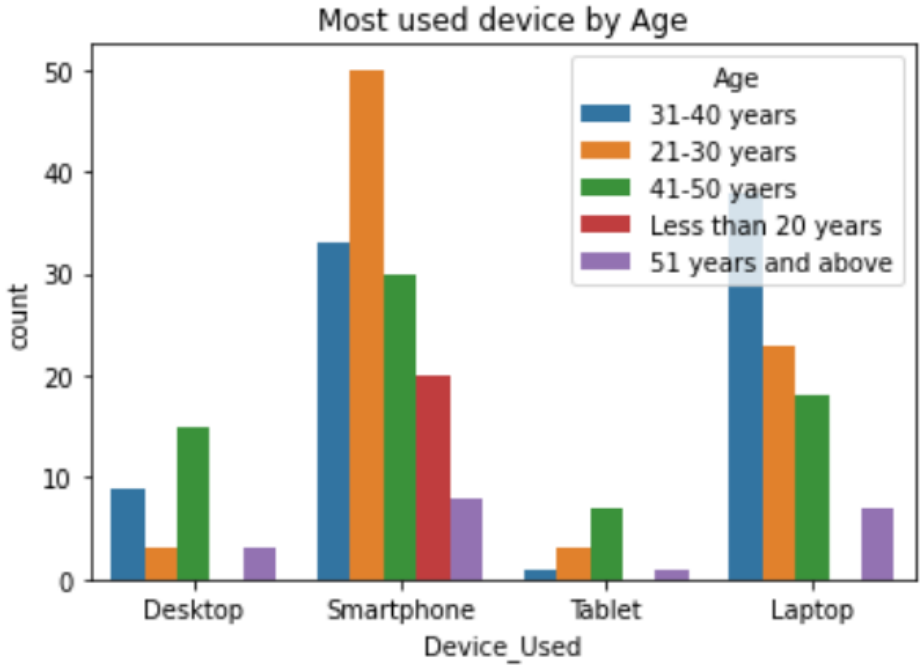
- Most of the customers agreed to offering a wide variety of listed product in several category and the count is high for the same.
- Around 86% of the customers would like to have provision of complete and relevant product information in the online shopping website.
- 47% of the customers agreed that shopping on the website helps them fulfilling certain roles and 33% of the customers are in confusion whether to agree or disagree and only 20% of the customers disagrees with it.
- Most of the people shopped from Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com companies and they think that it is easy to use website or applicatiion in these companies.
- Amazon.in and Flipkart.com have high visual appealing web-page layout compared to others.
- 48% of the customers says that amazon and flipkart shows wide variety of products in their shopping websited compared to other websites.
- 37% of the customers liked amazon and flipkart in displaying complete and relevant information of the products.
- Around 51 customers says that Amazon.in is the fast loading website and application and they liked it.
- The count is high for amazon followed by amazon and flipkart which means most of the customers liked the reliability of website or application in amazon and flipkart.
- Most of the customers likes Amazon's quickness to complete the purchase followed by Flipart's and only few of the customers likes Myntra website.
- In Amazon and flipkart websites there are several payment options available compared to the other shopping websites.
- Most of the customers liked Amazon's delivery speed.
- Most of the customers trusts amazon followed by flipkart in terms of keeping the privacy of their data information
- The count is high for the customers who belives that amazon website keeps their finanacial information secret.
- Most of the customers believed that Amazon has perceived trustworthiness comapared to others.
- Most of the customers like Amazon inerms of presence of online assistance through multi-channel followed by flipkart,Myntra and snapdeal.

- Most of the customers agreed that Amazon takes longer time to get logged them in.
- Customers believes that Amazon and flipkart takes longer time in display the graphics and photos in sales period.
- Customers says that Myntra and paytm have late declaration of price in promotion/sales period compared to others.
- Also Myntra and paytm takes longer page loading time.
- Snapdeal.com has limited mode of payment on most products followed by Amazon.in.
- In terms of time taken in product delivery Paytm has highest count followed by Snapdeal.com.
- Most of the customers disliked change in website/Application design on amazon followed by paytm.
- Most of the customers disliked frequent disruption when moving from one page to another on amazon, Myntra and snapdeal.
- Most of the customers believes that Amazon and flipkart website are as efficient as before.
- Most of the customers would like to recommend amazon to a friend.

I have also plotted some more plots to k now more about the dataset.

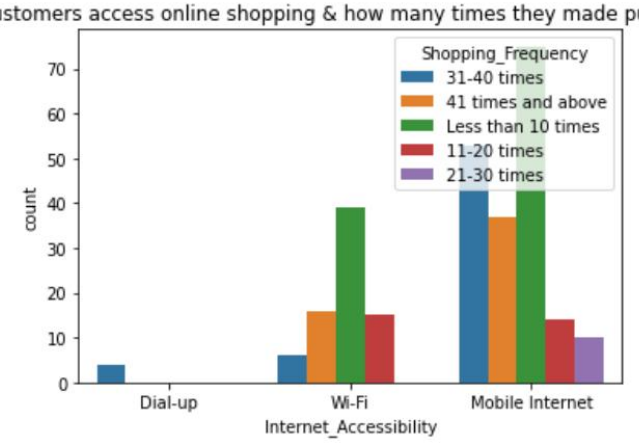
```
1
2 plt.title(' how long the customers shopping online on the basis of Gender')
3 sns.countplot(data['Shopping_Since'],hue=data['Gender']);
4
5
6
```

```
plt.title('Most used device by Age')
sns.countplot(data['Device_Used'],hue=data['Age']);

```

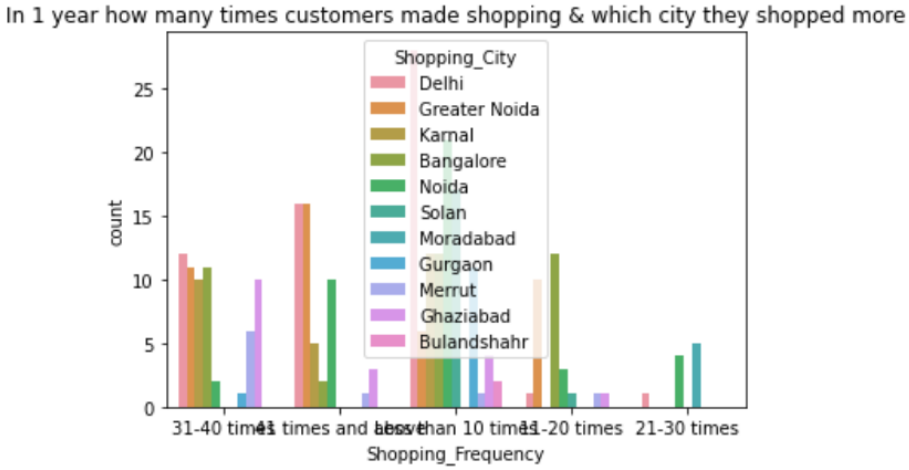



```
plt.title('How the customers access online shopping & how many times they made purchase in 1 year')
sns.countplot(data['Internet_Accessibility'],hue=data['Shopping_Frequency']);

```

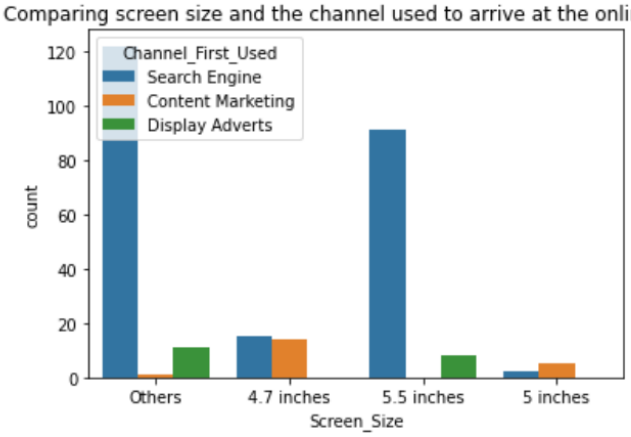
How the customers access online shopping & how many times they made purchase in 1 year



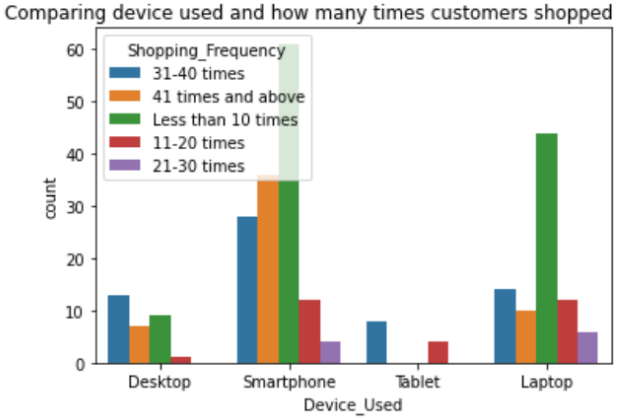
```
|:  1 plt.title('In 1 year how many times customers made shopping & which city they shopped more')
    2 sns.countplot(data['Shopping_Frequency'],hue=data['Shopping_City']);
```



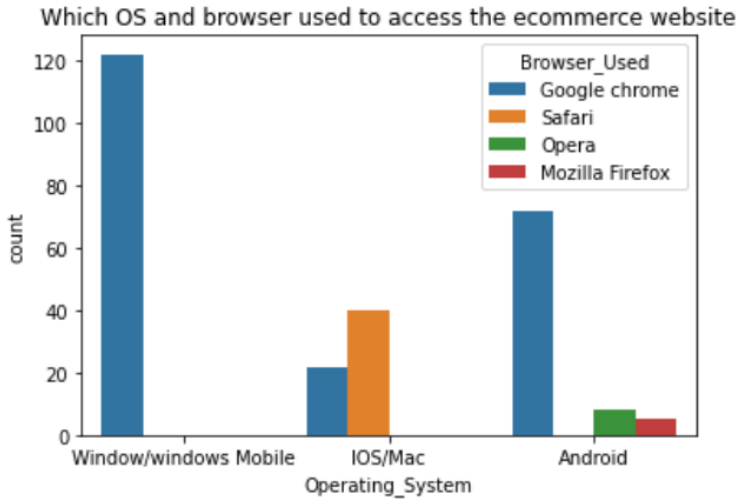

```
1 plt.title('Comparing screen size and the channel used to arrive at the online store',fontsize=12)
2 sns.countplot(data['Screen_Size'],hue=data['Channel_First_Used']);
3
4
```



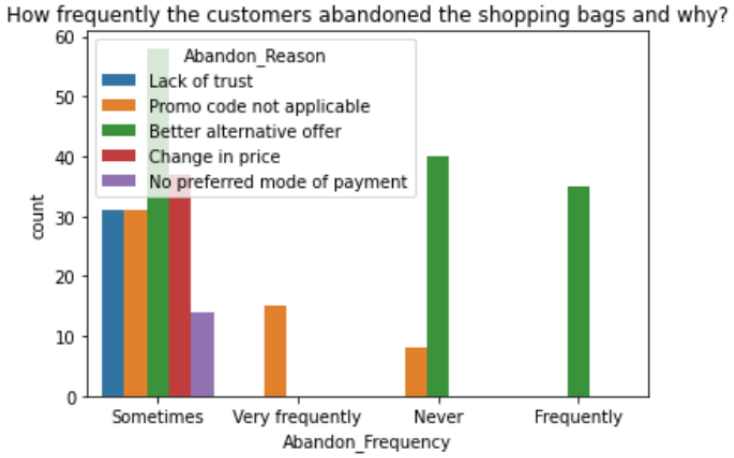

```
1 plt.title('Comparing device used and how many times customers shopped in 1 year',fontsize=12)
2 sns.countplot(data['Device_Used'],hue=data['Shopping_Frequency']);
3
4
```



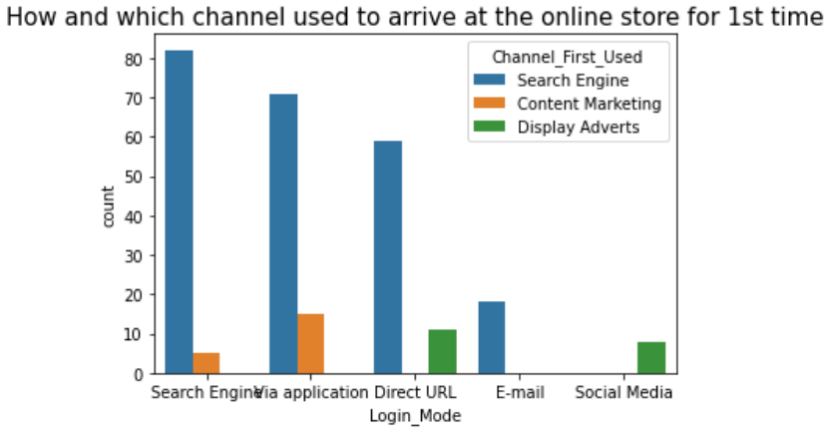
```
plt.title('Which OS and browser used to access the ecommerce website',fontsize=12)
sns.countplot(data['Operating_System'],hue=data['Browser_Used']);


```



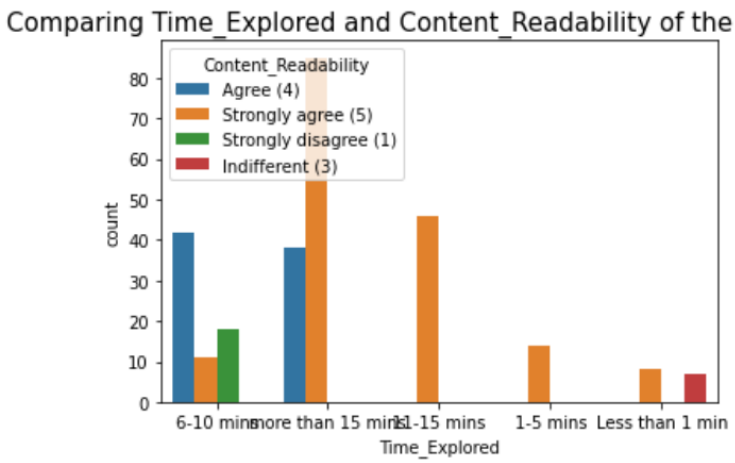

```
plt.title('How frequently the customers abandoned the shopping bags and why?',fontsize=12)
sns.countplot(data['Abandon_Frequency'],hue=data['Abandon_Reason']);
```



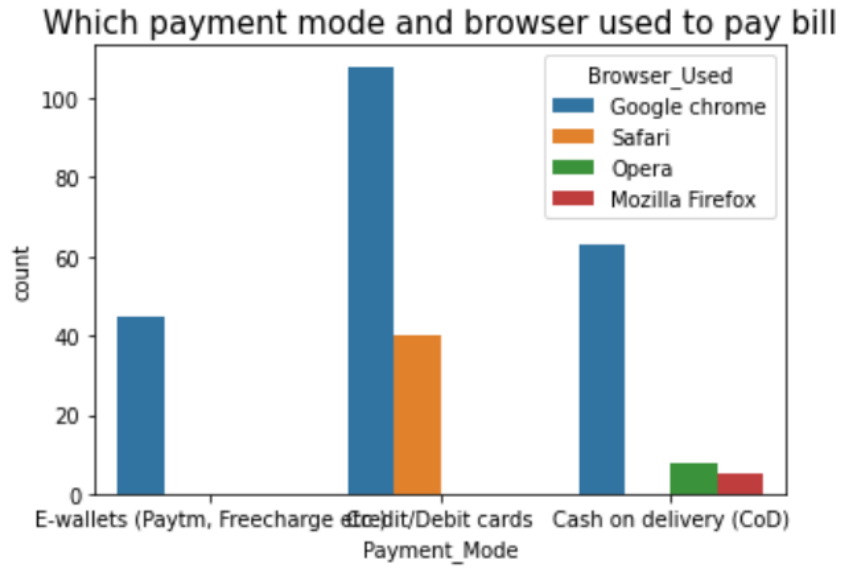
```
plt.title('How and which channel used to arrive at the online store for 1st time',fontsize=15)
sns.countplot('Login_Mode',hue='Channel_First_Used',data=data);

```




```
plt.title('Comparing Time_Explored and Content_Readability of the customers',fontsize=15)
sns.countplot('Time_Explored',hue='Content_Readability',data=data);

```



```
plt.title('Which payment mode and browser used to pay bill',fontsize=15)
sns.countplot('Payment_Mode',hue='Browser_Used',data=data);

```




```

plt.title('How frequently the customers abandoned the shopping bags while paying the bill?',fontsize=15)
sns.countplot('Abandon_Frequency',hue='Payment_Mode',data=data);
```

I have converted the dataset into numerical and categorical columns. And then I have used LabelEncoder to encode the categorical columns.

```
from sklearn.preprocessing import LabelEncoder
lab_enc=LabelEncoder()
```

```
# now lets check for numerical columns
numerical_col=[]
for i in data.dtypes.index:
    if data.dtypes[i]!='object':
        numerical_col.append(i)
print(numerical_col)
```

```
['Pin_code']
```

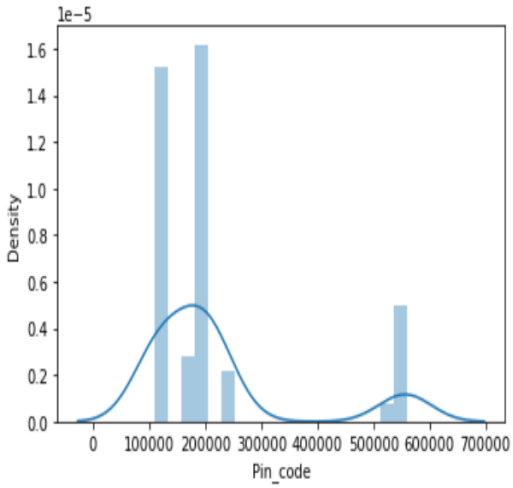
```
# lets check for categorical columns
categorical_col=[]
for i in data.dtypes.index:
    if data.dtypes[i]=='object':
        categorical_col.append(i)
print(categorical_col)
```

```
#encoding the catagorical columns.
data[categorical_col]=data[categorical_col].apply(lab_enc.fit_transform)
data[categorical_col]
```

	Gender	Age	Shopping_City	Shopping_Since	Shopping_Frequency	Internet_Accessibility	Device_Used	Screen_Size	Operating_System	Browser_Used
0	1	1	2	3	2	0	0	3	2	0
1	0	0	2	3	3	2	2	0	1	0
2	0	0	4	2	3	1	2	2	0	0
3	1	0	6	2	4	1	2	2	1	3
4	0	0	0	1	0	2	2	0	1	3
5	0	1	9	3	3	2	2	2	0	0
6	1	2	2	3	2	2	3	3	0	0
7	1	2	2	2	4	1	0	3	2	0
8	0	4	10	1	4	2	2	2	0	0
9	0	1	2	4	4	2	1	3	2	0
10	1	1	8	3	1	1	1	3	2	0

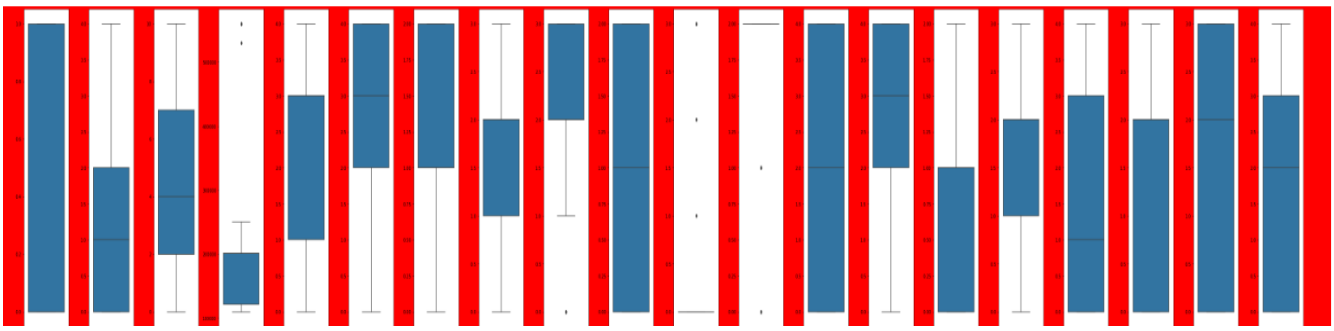
Then I have checked the distribution of the columns having continuous data.

```
# now the data looks good and there is no missing values so we can start visualizing the type of distribution for each featu
# we will only evaluate the type of distribution for features having continious data here

sns.distplot(data["Pin_code"])
plt.show()
```



The data was found to be normally distributed int he Pin_code column and we can observe a little skewness in the right side. Now lets find out the outliers by using box plots.

```
# checking for the outliers by ploting box plots

plt.figure(figsize=(50,40), facecolor='red')
plotnumber=1

for column in data:
    if plotnumber<=71:
        plt.subplot(4,20,plotnumber)
        ax=sns.boxplot(data=data[column])
        plt.xlabel(column, fontsize=20)

    plotnumber+=1
plt.tight_layout()
```



Pincode has some outliers we have removed the outliers, But for rest of the catagorical columns we have kept the outliers as it is.

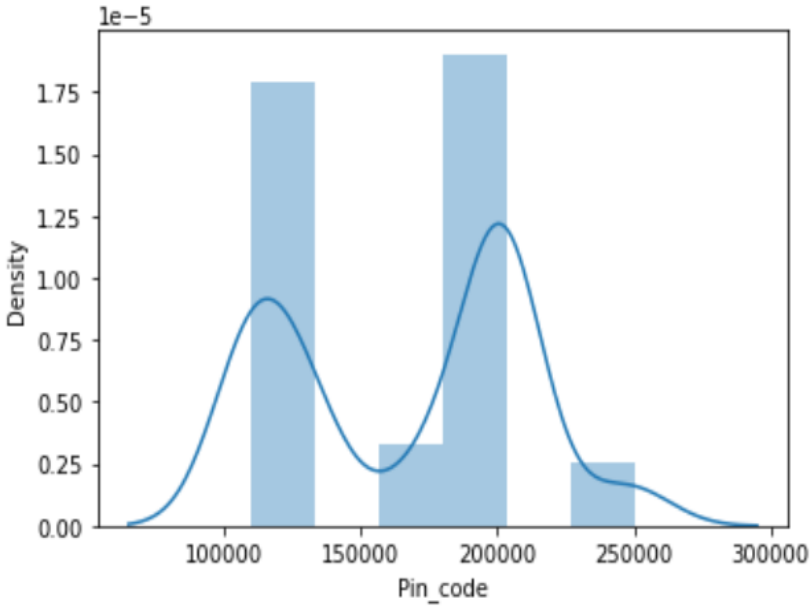
```
#assigning our datset to a new variable for testing the feasibility after removing the outliers
df=data
```

```
#Finding the IQR(Inter Quantile range) to finding the outliers

#1st quantile
q1=df.quantile(0.25)
#3rd quantile
q3=df.quantile(0.75)

#IQR
iqr=q3 -q1

iqr
```

```
#removing the outliers for Pin_code
pc_high=q3.Pin_code + (1.5 * iqr.Pin_code)


index=np.where(df['Pin_code'] > pc_high)

df=df.drop(df.index[index])
print(df.shape)

df.reset_index()
```

Then I have checked whether the distribution is somewhat better after removing the outliers.

```
#checking the distribution of Pin_code after removing the outliers
sns.distplot(data["Pin_code"])
plt.show()
```



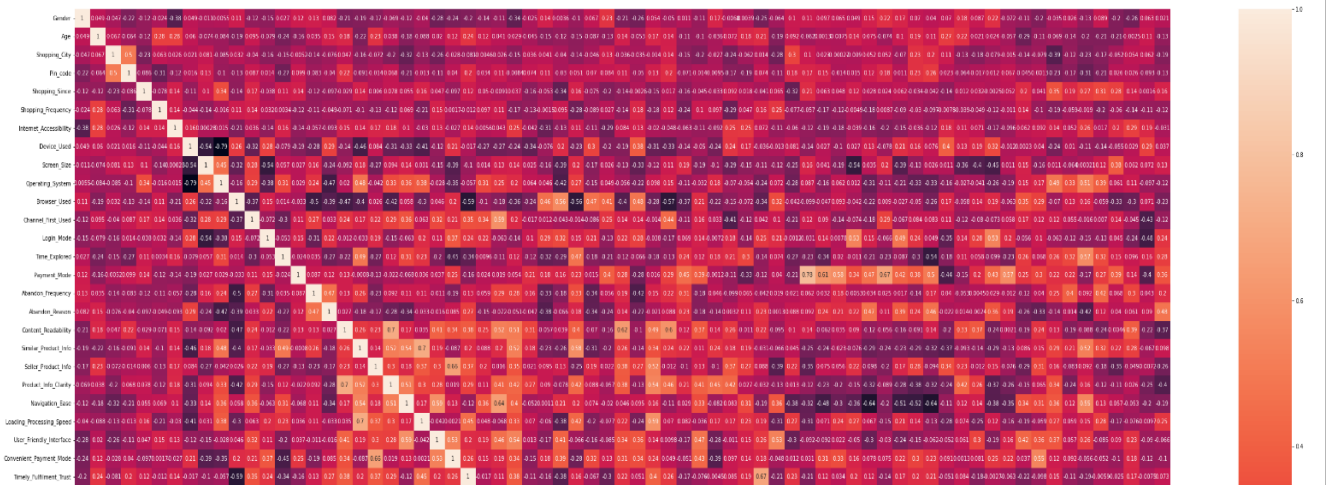
Then I have checked for multi colinearity among the features. And then I have plotted a heatmap for finding out all the correlations among the features.

```
#finding the corelation between the features in the data set
data.corr()
```

	Gender	Age	Shopping_City	Pin_code	Shopping_Since	Shopping_Frequency	Internet_Accessibility	Device_Used	Scre
Gender	1.000000	0.048932	-0.047285	-0.217051	-0.122845	-0.024250	-0.383177	0.049402	-0
Age	0.048932	1.000000	0.066757	-0.063737	-0.119878	0.284435	0.281308	0.059700	-0
Shopping_City	-0.047285	0.066757	1.000000	0.502607	-0.230181	0.063008	0.025934	0.020943	0
Pin_code	-0.217051	-0.063737	0.502607	1.000000	-0.085614	-0.306467	-0.118551	0.016476	0
Shopping_Since	-0.122845	-0.119878	-0.230181	-0.085614	1.000000	-0.078333	0.139136	-0.114961	0
Shopping_Frequency	-0.024250	0.284435	0.063008	-0.306467	-0.078333	1.000000	0.137331	-0.044425	-0
Internet_Accessibility	-0.383177	0.281308	0.025934	-0.118551	0.139136	0.137331	1.000000	0.156645	0
Device_Used	0.049402	0.059700	0.020943	0.016476	-0.114961	-0.044425	0.156645	1.000000	-0
Screen_Size	-0.011343	-0.073781	0.080901	0.126744	0.103752	-0.139950	0.000279	-0.541429	1
Operating_System	0.005531	-0.084156	-0.085307	-0.102719	0.336034	-0.015781	0.014885	-0.787589	0
Browser_Used	0.114896	-0.191970	0.032330	-0.127408	-0.143613	0.114144	-0.205750	0.261736	-0

```
#now lets find out the co relation using heatmap plot for a better understanding.
plt.figure(figsize=(50,40))
sns.heatmap(data.corr(),annot=True)
```

```
<AxesSubplot:>
```



We can see there are some features having very good colinearity among each other but this is not not good for the model. lets Find out the VIF(variance inflation factor) scores for each factor to find more about multicolinearity.

```
#using standard scalar for scaling the model
from sklearn.preprocessing import StandardScaler
scalar= StandardScaler()
x_scaled=scalar.fit_transform(data)
```

```
x_scaled.shape[1]
```

```
71
```

```python
#computing vif for all the features
from statsmodels.stats.outliers_influence import variance_inflation_factor
vif=pd.DataFrame()
vif['VIF']=[variance_inflation_factor(x_scaled,i) for i in range (x_scaled.shape[1])]
vif['Features']=data.columns

vif
```

2	2.753899	Shopping_City
3	3.063231	Pin_code
4	2.240240	Shopping_Since
5	2.413503	Shopping_Frequency
6	2.688888	Internet_Accessibility
7	35.808590	Device_Used
8	21.703059	Screen_Size
9	34.707684	Operating_System
10	inf	Browser_Used
11	inf	Channel_First_Used

Observation from the VIF scores:

- Device_used, Screen Size and Operating system have high vif values whicha re greater than 10 so they are highly multicolinear.
- Also some features like gender, age, shoping city Pincode etc are moderately multicolinear.
- Maximum Features have Vif= inf which means the features are perfect corelation with other features.

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.

Conclusions from the data Analysis

- This study was performed to find out about the influencing factors towards online shopping from different e-commerce websites.
- In this customer retention project during the Exploratory Data Analysis(EDA), in data cleaning part I have replaced the duplicate values in different columns. I have found no null values. i have also renamed the columns as their original named were way too big. I have also encoded the catagorical values in all the columns by using label encoder. I have Visualized the data using count plot, box plot and dist plot.I have also Checked the statistics of the data for columnshaving continious values and also checked for skewness, outliers and correlation between the features. I have also removed the outliers by using IQR.
- From the analysis it was observed that consumers purchasing decisions are dependent on various factors including both on the combination of both utilitarian value and hedonistic values. All these factors influence consumers to purchase products online from e-commerce websites. According to consumers' opinions, "time saving" is the most important influencing factors for shopping online from e-commerce websites.
- Again "information availability", "open 24/7", "huge range of products/ brands", "reasonable prices", "various offers for online products", "easy ordering system", and "shopping fun" are other infuencing factors for shopping online. Also, "online payment system", "personal privacy or security issues",

"delaying of delivery" and "lacks of personal customer service" are the main inhibitions of online shopping to the customers.

- After the Visualization of the data with proper visualization techniques I found out that Amazon is the best online store where the customers trust on buying products and it has positive impact on the customers.
- It was also concluded that online shopping is not trustworthy and reliable to some customers due to only online payment system and personal privacy. In addition, online security is a major concern for the consumer particularly in terms of fraud, privacy and hacking.