Submitted By: Prerna Jain

TABLE OF CONTENTS

INTRODUCTION

BUSINESS PROBLEM FRAMING

The objective was to perform extensive data analysis on a given dataset and produce valuable insights that will help in customer retention

CONCEPTUAL BACKGROUND OF THE DOMAIN PROBLEM

Customer satisfaction has emerged as one of the most important factors that guarantee the success of the 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.

REVIEW OF LITERATURE

In today's challenging economy and competitive business world, retaining their customer base is critical to organizational success. If the company doesn't give their customer some good reason to stay, the organization's competitors will give the customer a reason to leave. Customer retention and customer satisfaction drive profits. It's far less expensive to cultivate an organization existing customer base and sell more service to the customer than it's to seek new, single-transaction customers. Most surveys across industries show that keeping one existing customer is five to seven times more profitable than attracting one new customer. A customer-focused approach among its employees is still not present. In this era of intense competition. It is very important for any service company to understand that merely acquiring customer is not sufficient because there is a direct link between customer retention over time and profitability & growth. Customer retention to a great extent depends on service quality and customer satisfaction. Complaints are a natural part of any service activity as mistakes are an unavoidable feature of all human endeavoured thus also of service recovery. Service recovery is the process of putting things right after something goes wrong in the service delivery. Customer retention is the maintenance of continuous trading relationships with customers over the long term.

Customer retention is the mirror image of customer defection. High retention is equivalent to low defection. In an industry where there are multiple purchases over the years, the organization's entire team should be very focused on retaining those customers:

i. Delivering service that's consistent with your value proposition and brand.
ii. Cross-selling, up-selling and asking for referrals from existing customers.
iii. Developing programs to increase customer loyalty and decrease turnover.
iv. Prioritizing retention as a major focus in your annual marketing plan.
v. Knowing the lifetime value for different segments and using that data to improve the marketing.

Studies say it costs ten times more to generate a new customer than to maintain an existing one. If organization has a small number of customers, losing a few could cripple company. Even if there are a large number of customers, a small increase in the rate should dramatically increase profits. The maintenance of the patronage of people who have purchased a company's goods or services once and the gaining of repeat purchases. Customer retention occurs when a customer is loyal to a company, brand, or to a specific product or service, expressing long-term commitment and refusing to purchase from competitors. Of critical importance to such strategies are the wider concepts of customer service, customer relations, and relationship marketing. Companies can build loyalty and retention through the use of a number of techniques, including database marketing, the issue of loyalty cards, redeemable against a variety of goods or service, preferential discounts, free gifts, special promotions, newsletters or magazines, members' clubs or customized products in limited editions. It has been argued that customer retention is linked to employee loyalty, since loyal employees build up long-term relationships with customers. Customer retention has always been an important topic for the marketing. For sure, the advantages of loyal clients are obvious. Often CRM is only implementing new systems for data mining and client segmentation or operational system like a complaint management. But the thing is: data mining system or client clubs are not the basis. They are the cherry of the cake called client retention. A key principle of relationship marketing is the retention of customers through varying means and practices to ensure repeated trade from pre-existing customers by satisfying requirements above those of competing companies through a mutually beneficial relationship. This technique is now used as a means of counter balancing new customer and opportunities with Current and existing customers as a means of maximizing profit and counteracting the “leaky bucket theory of business” in which new customer gained in order direct marketing-oriented businesses were at the expense of or coincided with the loss of older customers. This process of "churning" is less economically viable than retaining all or the majority of customers using both direct and relationship management as lead generation via new customers requires more investment. Many companies in competing markets will redirect or allocate large amounts of resources or attention towards customer retention as in markets with increasing competition it may cost 5 times more to attract new customers than it would to retain current customers, as direct or "offensive" marketing requires much more extensive resources to cause defection from competitors. However, it is suggested that because of the extensive classic marketing theories center on means of attracting customer and creating transactions rather than maintaining them, the majority usage of direct marketing used in the past is now gradually being used more alongside relationship marketing as its importance becomes more recognizable. According to Buchanan and Gilles the increased profitability associated with customer retention efforts occurs because of several factors that occur once a relationship has been established with a customer.

i. The cost of acquisition occurs only at the beginning of the relationship, so the longer the relationship, the lower the amortized cost.
ii. Account maintenance costs decline as a percentage of total costs or as a percentage of revenue.
iii. Long-term customers tend to be less inclined to switch, and also tend to be fewer prices sensitive. This can result in stable unit sales volume and increase in dollar-sales volume.
iv. Long-term customer may initiate free word of mouth promotions and referrals.
v. Long-term customers are more likely to purchase ancillary products and high margin supplemental products.
vi. Customer that stays with company tend to be satisfied with the relationship andare less likely to switch to competitors, making it difficult for competitors to enter the market or gain market share.

IMPORTANCE OF CUSTOMER RETENTION:

There are a number of reasons for this. To begin with, to acquire a customer a company incurs promotional costs like advertising, sales promotion etc. It is said that it costs five times more to attract a new customer than retaining one. The operating cost decrease when a customer stay. Service being rich in experience and credence qualities, it takes some time for customers to get accustomed to it and once they are used to the service and are satisfied with the service provider, they tend to purchase more over a period of time. As they remain satisfied with a service provider, they spread a positive word of mouth, which is very effective in case of service for attracting new customers. Longer the customer stays with an organization, more the organization knows about him, which enables it to offer a customized service which makes it difficult for the customer to defect. This may even provide opportunities to the organization to charge price premium by offering individualized service which may be difficult for the competitors to offer. Considering the importance of retaining customers in service business,

Reichheld & Sasser coined a term ‘Zero Defection’. They highlighted that companies can boost profits by almost 100% by retaining just 5% more of their customers. Further, it is also very important to understand the life time value of a customer. Further, if by a positive word of mouth, he brings just one more customer to the organization, his value to the organization doubles. Therefore, it is important for all the employees in the organization to understand the life time value of their customers.

ADVANTAGES OF CUSTOMER RETENTION POSSIBILITY OF REPEAT BUSINESS:

This is probably the most obvious advantage of customer retention. Effective services that lead to customer satisfaction will make customer coming back to again, thus giving repeat business. Repeat business is a win-win proposition for the business or service and the customer. The business reduces the cost of customer acquisition, while the customer reduces the cost of finding a reliable vendor and thus also saves on costs associated with switching vendors.

REDUCED COSTS FOR CUSTOMER ACQUISITION

Acquiring a customer has certain associated costs. These include the costs associated with advertising, following up, sales demos, travel and meeting cost etc. having a repeat customer means that the customer means that the customer is already aware of your processes and can predict certain quality of output, thus minimizing the cost involved in new customer acquisition. Having a repeat customer also has the potential to open up another channel to advertise your business word of mouth. Word of mouth advertising recommendations are perhaps the most important outcome of having a satisfied customer.

FOSTERING GREATER INTERACTION BETWEEN BUSINESS AND CUSTOMER

Today's markets are increasingly moving away from mass produced standard products and service, towards a more customized market, where products and service are tailored to meet customers' specific requirements. Having a repeat customer is an opportunity for you to build a more focused relationship based on your customers' specific needs and requirements. Being ensured of having a customer who comes back, you have more confidence to suggest improvements, provide an insight to better understand their needs and consequently design products and services that are relevant. Having a repeat business also provides an opportunity for the buyer and the seller to co-create products and services.

HAVING MORE DELIGHTED CUSTOMERS

Effective customer retention strategies allow you to move from the zone of customer satisfaction to customer delight. Studies have shown that customer delight is achieved only when there is a perfect synergy between the buyer needs and the buyer understands what the seller can deliver exactly what the customer need. If you are able to deliver your customers, you have better chance of them coming back to you, since they now know why you are different from the rest of competition.

CUSTOMER RETENTION: STATISTICS

- Acquiring a new customer can cost five times more than satisfying and retaining current customers.
- 2% increase in customer retention has the same effect on profits as cutting costs by 10%.
- The average company losses 10% of its customer each year.
- 5% reduction in the customer defection rate can increase profits by 25-125%, depending on the industries.
- The customer profitability rate tends to increase over the life of a retained customer.
- Companies can boost profits anywhere from 25 to 125% by retaining merely5% more existing customers.

MEASURING CUSTOMER RETENTION:

Retention rate is normally calculated as the number of customers who have been lost over a period of time, usually calculated over a quarterly or annual period. The key is to calculate the percentage versus existing customers, and not underestimate the loss rate by tallying new customer acquisitions into the mix. The customer retention rate refers to the number of customers lost over a period of time. It is normally calculated by the percentage of lost customer versus existing customers over a quarterly or annual period, without tallying new customer acquisitions. While there are obvious benefits to keeping customers loyal and maintaining retention rates, it can be extremely challenging for management to keep retention rates up. Some companies can measure retention rate using their CRM system, since any of the vendor swith solid sales modules should offer this capability. Customer service expert Lori Bock lender commends that companies look for this functionality when evaluating CRM solutions, even though it is unlikely to be the differentiating factor. Companies like witness, Performix, AIM, and Merced offer these types of tools. To measure this, some companies combine data from the CRM system and data from other systems, such as your systems, such as your quality monitoring system, ACD or CTI solution handling contact routing and reporting. There are no hard and fast rules on calculating customer defection and customer retention, according to Lowenstein. It can depend on the industries or the type of business, since companies have long-term arrangements with customers. However, several consulting and database management companies have succeeded in creating them. However, the appropriate interval over which retention rate should be measured is not always one year. Rather, it depends on the customer repurchase cycle. Car insurance and magazine subscriptions are bought on an annual basis. Carpet tiles and hi-fis are not. If the normal hi-fire placement cycle is four years, then retention rate is more meaningful if it is measured over four years instead of twelve months. Additional complexity is added when companies a sell a range of products and services, each with different repurchase cycles. Automobile dealers might sell cars, parts, fuel and service to a single customer. These products have different repurchase cycles which make it very difficult for the dealer to have a whole of customer perspective on retention. Sometimes companies are not clear about whether an individual customer has defected. This is because of the location of customer related data, which might be retained in product silos, channel silos or functional silos.

TYPE OF CUSTOMER RETENTION RATE

RAW CUSTOMER RETENTION RATE:

This is the number of customers doing business with a firm at the end of a trading period, expressed as percentage of those who were active customer at the beginning of the period.

SALES-ADJUSTED RETENTION RATE:

This is the value of sales achieved from the retained customers, expressed as a percentage of the sales achieved from all customers who were active at the beginning of the period.

PROFIT-ADJUSTED RETENTION RATE:

This is the profit earned from the retained customers, expressed as a percentage of the profit earned from all customers who were active at the beginning of the period.

MOTIVATION FOR THE PROBLEM UNDERTAKEN

The combination of both utilitarian value and hedonistic values are needed to affect the repeat purchase intention (loyalty) positively. The objective behind the problem is to help eCommerce websites to find E-retail factors for customer activation and retention.

```
In [3]: data.head()
```

Out[3]:

	1Gender of respondent	2 How old are you?	3 Which city do you shop online from?	4 What is the Pin Code of where you shop online from?	5 Since How Long You are Shopping Online ?	6 How many times you have made an online purchase in the past 1 year?	7 How do you access the internet while shopping on-line?	8 Which device do you use to access the online shopping?	9 What is the screen size of your mobile device?\t\t\t\t\t\t	10 What is the operating system (OS) of your device?\t\t\t\t	...	Longer time to get logged in (promotion, sales period)	Longer time in displaying graphics and photos (promotion, sales period)	Late declaration of price (promotion, sales period)
0	Male	31-40 years	Delhi	110009	Above 4 years	31-40 times	Dial-up	Desktop	Others	Window/windows Mobile	...	Amazon.in	Amazon.in	Flipkart.com
1	Female	21-30 years	Delhi	110030	Above 4 years	41 times and above	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	...	Amazon.in, Flipkart.com	Myntra.com	snapdeal.com
2	Female	21-30 years	Greater Noida	201308	3-4 years	41 times and above	Mobile internet	Smartphone	5.5 inches	Android	...	Myntra.com	Myntra.com	Myntra.com
3	Male	21-30 years	Karnal	132001	3-4 years	Less than 10 times	Mobile internet	Smartphone	5.5 inches	IOS/Mac	...	Snapdeal.com	Myntra.com, Snapdeal.com	Myntra.com
4	Female	21-30 years	Bangalore	530068	2-3 years	11-20 times	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	...	Flipkart.com, Paytm.com	Paytm.com	Paytm.com

5 rows × 71 columns

ANALYTICAL PROBLEM FRAMING

MATHEMATICAL/ ANALYTICAL MODELING OF THE PROBLEM

In this project we have performed various mathematical and statistical analysis such as we checked description or statistical summary of the data using describe, checked correlation using corr and also visualized it using heatmap.

- The dataset consists of 71 features and 269 rows.
- All of the attributes were of 'object' type except the pin code (int).
- Dataset did not contain any null values.

In [6]:
```
df.head()
```

Out[6]:

	1Gender of respondent	2 How old are you?	3 Which city do you shop online from?	4 What is the Pin Code of where you shop online from?	5 Since How Long You are Shopping Online ?	6 How many times you have made an online purchase in the past 1 year?	7 How do you access the internet while shopping on-line?	8 Which device do you use to access the online shopping?	9 What is the screen size of your mobile device? \t\t\t\t\t\t	10 What is the operating system (OS) of your device? \t\t\t\t	...	Longer time to get logged in (promotion, sales period)	Longer time in displaying graphics and photos (promotion, sales period)	Late declaration of price (promotion, sales period)
0	Male	31-40 years	Delhi	110009	Above 4 years	31-40 times	Dial-up	Desktop	Others	Window/windows Mobile	...	Amazon.in	Amazon.in	Flipkart.com
1	Female	21-30 years	Delhi	110030	Above 4 years	41 times and above	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	...	Amazon.in, Flipkart.com	Myntra.com	snapdeal.com
2	Female	21-30 years	Greater Noida	201308	3-4 years	41 times and above	Mobile Internet	Smartphone	5.5 inches	Android	...	Myntra.com	Myntra.com	Myntra.com
3	Male	21-30 years	Karnal	132001	3-4 years	Less than 10 times	Mobile Internet	Smartphone	5.5 inches	IOS/Mac	...	Snapdeal.com	Myntra.com, Snapdeal.com	Myntra.com
4	Female	21-30 years	Bangalore	530068	2-3 years	11-20 times	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	...	Flipkart.com, Paytm.com	Paytm.com	Paytm.com

5 rows × 71 columns

DATA SOURCES AND QUESTIONS

The data has been collected from Indian Online Retailers. Results indicate the e-retail success factors which are very much critical for customer saitisfaction.

Questions that we will try to answer by performing EDA are as follows:

- Online Shopping is preferred in which cities?
- Who are the potential customers?
- Which age group they belong to?
- Who shops the most? Males or Females?
- How long people have been shopping for?
- Are they satisfied?
- Do they trust the online store?
- How many times shopped in past 1 year? Did they purchase or cancelled often?
- Time spent before deciding what to purchase?
- Average time spent shopping before making a decision by males and females?
- Preferred payment option?
- group by age groups?
- gender
- How do people access online retail store?
- internet accessibility
- device used
- screen size of the device
- os of the device
- browser used
- channel used2
- how did u reach online retail store after first visit?
- Why do people cancel their orders?
- What factors about the app matters ?
- content readabilty should be simple?
- info on the product?
- navigation of the website

- User friendly Interface
- convenient payment methods
- trust on that retail store will be transacting on time
- guarantees privacy of the customer?
- the availability of retails stores on multiple channels?
- Return and replacement policy of the e-tailer is important for purchase decision
- offers a wide product variety?
- if you are getting value for the money spent?
- loading and processing speed
- What are the benefits of online shopping?
- gives monetary benefit and discounts
- access to loyalty programs
- enhances social status
- feel gratification
- helps fulfill certain roles
- gives a sense of adventure
- monetary savings
- convenient and flexible
- enjoyment derived
- User satisfaction depends on what factors?
    - cannot exist without trust (38)
    - net benefit derived lead to user satisfaction(37)
    - achieved while shopping on good quality app/website (36)
    - display info (35)
- Which online retail store performs better?
    - easy to use app?
    - visually appealing web page layout?
    - variety of products on?
    - fast loading websites?
    - reliability of website or appn?

- complete relevant info of products?
- easy to do a purchase on?
- several payments option?
- speedy order?

EDA STEPS AND VISUALIZATION

IMPORT NECESSARY PACKAGES:-

```
In [1]: # Importing Necessary Libraries
        import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import seaborn as sns
        import warnings
        import joblib
        warnings.filterwarnings('ignore')
        %matplotlib inline

In [2]: !pip install plotly

        Requirement already satisfied: plotly in c:\users\prerna jain\anaconda3\lib\site-packages (5.5.0)
        Requirement already satisfied: tenacity>=6.2.0 in c:\users\prerna jain\anaconda3\lib\site-packages (from plotly) (8.0.1)
        Requirement already satisfied: six in c:\users\prerna jain\anaconda3\lib\site-packages (from plotly) (1.16.0)

In [3]: import plotly.graph_objects as go
        from plotly.subplots import make_subplots
        from  sklearn.preprocessing import LabelEncoder
        import plotly.express as px

In [4]: #Importing the dataset
        df = pd.read_excel('customer_retention_dataset.xlsx','datasheet')
        encoded_df = pd.read_excel('customer_retention_dataset.xlsx','codedsheet')

In [5]: #Checking shape of dataset
        print('Original Data :', df.shape)
        print('Encoded Data :', encoded_df.shape)

        Original Data : (269, 71)
        Encoded Data : (269, 71)
```

<u>UNDERSTAND DATA STRUCTURE</u>:-

Firstly, imported all the required libraries, followed by loading the data. Then did a statistical analysis on the dataset. Then deeply analyzed the data by finding a relationship between each attribute and the Customer according to the given details and tried to find the factors that will help in Customer Retention.

> Below images shows that in dataset 71 columns and 269 rows in it.

```
In [6]: df.head()
Out[6]:
```

	1Gender of respondent	2 How old are you?	3 Which city do you shop online from?	4 What is the Pin Code of where you shop online from?	5 Since How Long You are Shopping Online ?	6 How many times you have made an online purchase in the past 1 year?	7 How do you access the internet while shopping on-line?	8 Which device do you use to access the online shopping?	9 What is the screen size of your mobile device? #######	10 What is the operating system (OS) of your device? ####	...	Longer time to get logged in (promotion, sales period)	Longer time in displaying graphics and photos (promotion, sales period)	Late declaration of price (promotion, sales period)
0	Male	31-40 years	Delhi	110009	Above 4 years	31-40 times	Dial-up	Desktop	Others	Window/windows Mobile	...	Amazon.in	Amazon.in	Flipkart.com
1	Female	21-30 years	Delhi	110030	Above 4 years	41 times and above	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	...	Amazon.in, Flipkart.com	Myntra.com	snapdeal.com
2	Female	21-30 years	Greater Noida	201308	3-4 years	41 times and above	Mobile Internet	Smartphone	5.5 inches	Android	...	Myntra.com	Myntra.com	Myntra.com
3	Male	21-30 years	Karnal	132001	3-4 years	Less than 10 times	Mobile Internet	Smartphone	5.5 inches	IOS/Mac	...	Snapdeal.com	Myntra.com, Snapdeal.com	Myntra.com
4	Female	21-30 years	Bangalore	530068	2-3 years	11-20 times	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	...	Flipkart.com, Paytm.com	Paytm.com	Paytm.com

5 rows × 71 columns

```
In [7]: encoded_df.head()
Out[7]:
```

	1Gender of respondent	2 How old are you?	3 Which city do you shop online from?	4 What is the Pin Code of where you shop online from?	5 Since How Long You are Shopping Online ?	6 How many times you have made an online purchase in the past 1 year?	7 How do you access the internet while shopping on-line?	8 Which device do you use to access the online shopping?	9 What is the screen size of your mobile device? #######	10 What is the operating system (OS) of your device? ####	...	Longer time to get logged in (promotion, sales period)	Longer time in displaying graphics and photos (promotion, sales period)	Late declaration of price (promotion, sales period)	Longer loading (prom sales pe
0	0	3	Delhi	110009	5	4	4	3	5	1	...	Amazon.in	Amazon.in	Flipkart.com	Flipkar
1	1	2	Delhi	110030	5	5	2	1	2	3	...	Amazon.in, Flipkart.com	Myntra.com	snapdeal.com	Snapdea
2	1	2	Greater Noida	201308	4	5	3	1	4	2	...	Myntra.com	Myntra.com	Myntra.com	Myntr
3	0	2	Karnal	132001	4	1	3	1	4	3	...	Snapdeal.com	Myntra.com, Snapdeal.com	Myntra.com	Paytn
4	1	2	Bangalore	530068	3	2	2	1	2	3	...	Flipkart.com, Paytm.com	Paytm.com	Paytm.com	Paytn

5 rows × 71 columns

We have renamed our columns label for easy access

```
In [7]: # Let's rename the new column names to the original datastes

columns = ['gender','age','city','pinCode','shoppingSince','shoppingFrequency','internetAccessibility','deviceUsed',
           'screenSize','OS', 'browserUsed','channelFirstUsed','loginMode','timeSpentDeciding','paymentMode','cancellingFrequenc
          'cancellationReason','contentReadability','similarProductInfo','sellerProductInfo','productInfoClarity','navigationEase
          'loadingProcessingSpeed','userFrienlyInterface','conveninetPaymentMode','timelyFulfilmentTrust','custSupportResponse',
          'custPrivacyGuarantee','variousChannelResponses','benefit','enjoy','convenience','returnReplacementPolicy','loyaltyProg
          'infoSatisfaction','siteQualitySatisfaction','netBenefitSatisfaction','trust','productSeveralCategory','relevantProduct
          'patronizingConvenience','adventureSense','socialStatus','gratification','roleFulfilment','moneyWorthy','shoppedFrom',
           'visuallyAppealingWebApp','productVariety','completeProductInfo','fastWebApp','reliableWebApp','quickPurchase','paymen
           'fastDelivery','custInfoPrivacy','finInfoSecurity','perceivedTrustworthiness','multiChannelAssistance','longLoginTime'
           'latePriceDeclare','longLoadingTime','limitedPaymentMode','lateDelivery','webAppDesignChange','pageDisruption','webApp
           'recommendation']

data.columns = columns
encoded_data.columns = columns
```

```
In [8]: data.head()
```

Out[8]:

	gender	age	city	pinCode	shoppingSince	shoppingFrequency	internetAccessibility	deviceUsed	screenSize	OS	...	longLoginTime
0	Male	31-40 years	Delhi	110009	Above 4 years	31-40 times	Dial-up	Desktop	Others	Window/windows Mobile	...	Amazon.in
1	Female	21-30 years	Delhi	110030	Above 4 years	41 times and above	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	...	Amazon.in, Flipkart.com
2	Female	21-30 years	Greater Noida	201308	3-4 years	41 times and above	Mobile Internet	Smartphone	5.5 inches	Android	...	Myntra.com
3	Male	21-30 years	Karnal	132001	3-4 years	Less than 10 times	Mobile Internet	Smartphone	5.5 inches	IOS/Mac	...	Snapdeal.com
4	Female	21-30 years	Bangalore	530068	2-3 years	11-20 times	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	...	Flipkart.com, Paytm.com

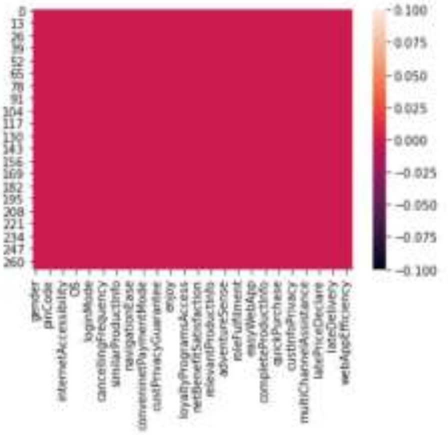
5 rows × 71 columns

```
In [12]: # let's check the datatype of each feature

data.columns.groupby(data.dtypes)

Out[12]: {int64: ['pinCode'], object: ['gender', 'age', 'city', 'shoppingSince', 'shoppingFrequency', 'internetAccessibility', 'deviceUs
ed', 'screenSize', 'OS', 'browserUsed', 'channelFirstUsed', 'loginMode', 'timeSpentDeciding', 'paymentMode', 'cancellingFrequen
cy', 'cancellationReason', 'contentReadability', 'similarProductInfo', 'sellerProductInfo', 'productInfoClarity', 'navigationEa
se', 'loadingProcessingSpeed', 'userFrienlyInterface', 'conveninetPaymentMode', 'timelyFulfilmentTrust', 'custSupportResponse',
'custPrivacyGuarantee', 'variousChannelResponses', 'benefit', 'enjoy', 'convenience', 'returnReplacementPolicy', 'loyaltyProgra
msAccess', 'infoSatisfaction', 'siteQualitySatisfaction', 'netBenefitSatisfaction', 'trust', 'productSeveralCategory', 'relevan
tProductInfo', 'monetarySavings', 'patronizingConvenience', 'adventureSense', 'socialStatus', 'gratification', 'roleFulfilmen
t', 'moneyWorthy', 'shoppedFrom', 'easyWebApp', 'visuallyAppealingWebApp', 'productVariety', 'completeProductInfo', 'fastWebAp
p', 'reliableWebApp', 'quickPurchase', 'paymentOptionsAvailability', 'fastDelivery', 'custInfoPrivacy', 'finInfoSecurity', 'per
ceivedTrustworthiness', 'multiChannelAssistance', 'longLoginTime', 'longDisplayTime', 'latePriceDeclare', 'longLoadingTime', 'l
imitedPaymentMode', 'lateDelivery', 'webAppDesignChange', 'pageDisruption', 'webAppEfficiency', 'recommendation']}
```

```
In [13]: # Let's check the null values in our dataset

data.isnull().sum()

Out[13]: gender                0
         age                   0
         city                  0
         pinCode               0
         shoppingSince         0
                              ..
         lateDelivery          0
         webAppDesignChange    0
         pageDisruption        0
         webAppEfficiency      0
         recommendation        0
         Length: 71, dtype: int64
```

We do not have any null values in our dataset

In [14]:
```python
# Let's plot the Heat map to visualise the null values

sns.heatmap(data.isnull());
```



In [15]:
```python
# Let's check the  categorical and continuos variables in our dataset

categorical=[x for x in data.columns if data[x].dtype==object]
print(len(categorical))
continous=[x for x in data.columns if x not in categorical]
print(len(continous))
```

```
70
1
```

We can observe that we have only one numeral datatype rest all are catagoriacal datatypes

In [16]:
```python
# Let's check the statistical summary

data.describe()
```

Out[16]:

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

In [17]:
```python
# Let's describe the object types

data.describe(include='object')
```

Out[17]:

	gender	age	city	shoppingSince	shoppingFrequency	internetAccessibility	deviceUsed	screenSize	OS	browserUsed	...	longLoginTi
count	269	269	269	269	269	269	269	269	269	269	...	2
unique	2	5	11	5	6	4	4	4	3	4	...	
top	Female	31-40 years	Delhi	Above 4 years	Less than 10 times	Mobile internet	Smartphone	Others	Window/windows Mobile	Google chrome	...	Amazon
freq	181	81	58	98	114	142	141	134	122	216	...	

4 rows × 70 columns

Data Visualization

Univariate Analysis

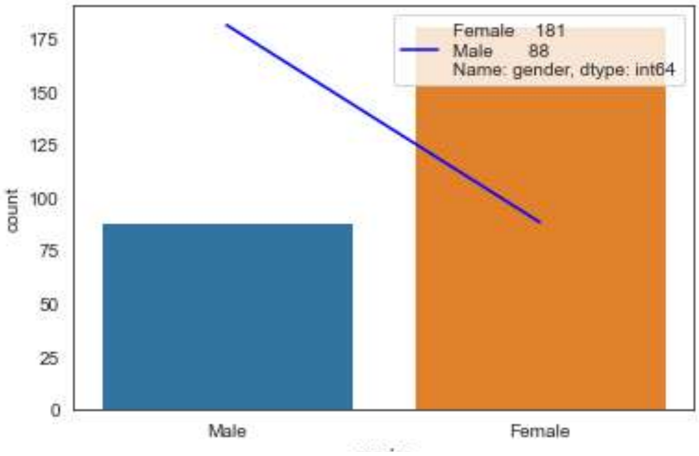
Using Set_Style and countplot:

In [20]:
```
def bia_vari(a):
    plt.figure(figsize=(8,6))
    sns.set_style('darkgrid')
    sns.lineplot(x='Which__of__the__Indian__online__retailer__would__you__recommend__to__a__friend',y=a,data=df,markers=False)
    plt.xticks(rotation=90)
    plt.show()
```

In [21]:
```
def value_counts(data):
    s=df.columns
    for i in s:
        if df[i].dtype=='object':
            sns.set_style('white')  #darkgrid
            a=df[i].value_counts()
            print(a)
            sns.countplot(x=df[i],data=df)
            plt.plot(a,color='b',label=a)
            plt.legend()
            plt.show()
            print("*******************************************************************************")
            print('\n')

value_counts(df)
```

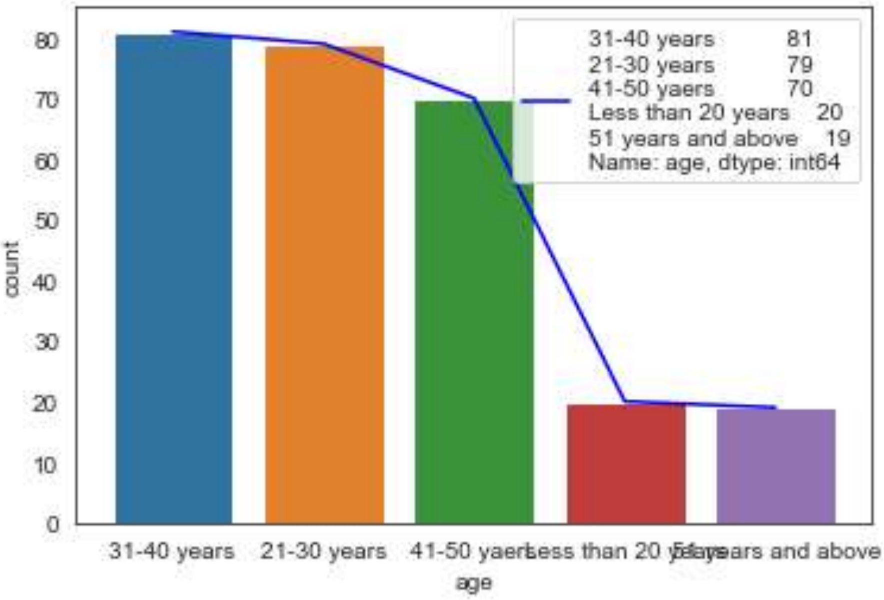
VISUALIZATIONS:-

```
Female    181
Male       88
Name: gender, type: int64
```



****

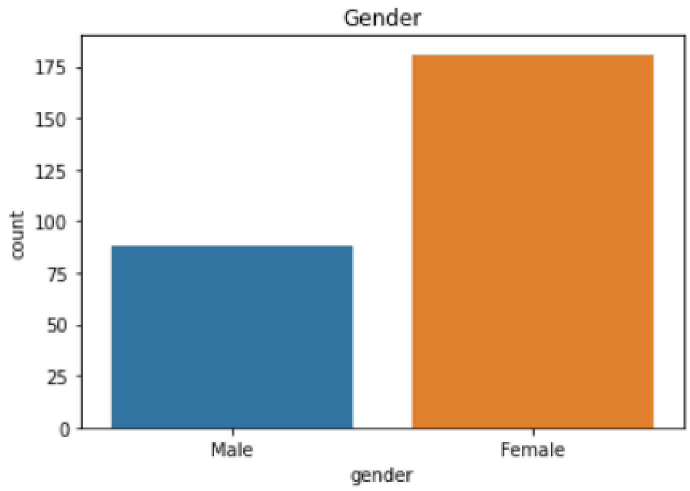
```
31-40 years           81
21-30 years           79
41-50 years           70
Less than 20 years    20
51 years and above    19
Name: age, dtype: int64
```



VISUALIZATIONS

In [18]:
```
sns.countplot(data['gender'])
plt.title('Gender')
print(round(data['gender'].value_counts()/269*100),2)
```

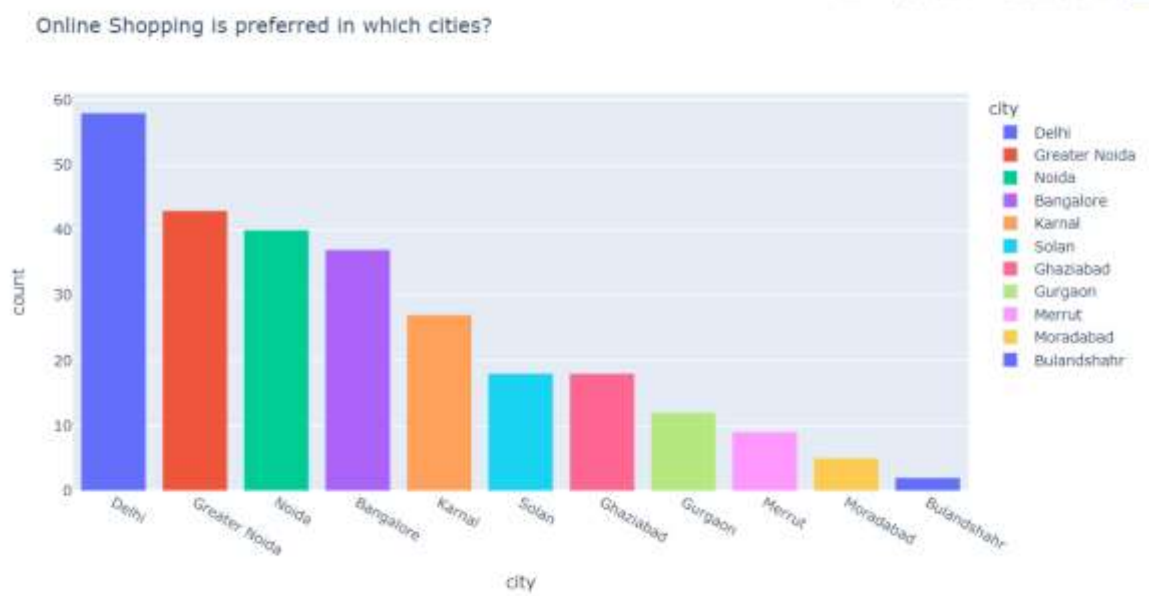
```
Female    67.0
Male      33.0
Name: gender, dtype: float64 2
```



Observation:

Out of the total, 67% customer are female

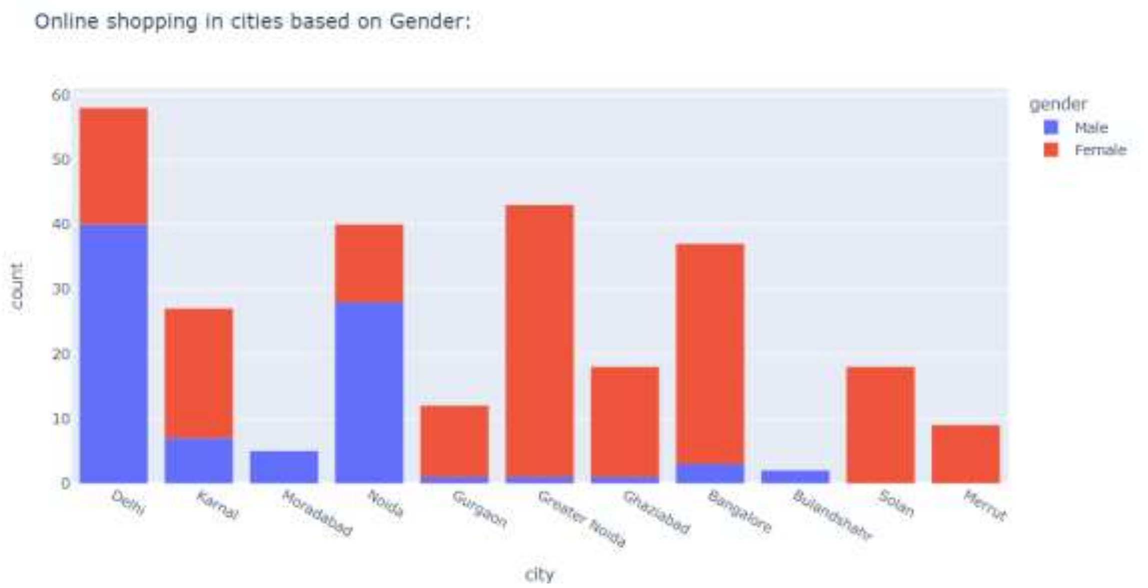
In [19]:
```
# Let's check in whihc city people shop the most

city_count = pd.DataFrame(data['city'].value_counts()).reset_index()
city_count.columns = ['city','count']


fig = px.bar(city_count, x='city',y='count',
             color='city',
             title = 'Online Shopping is preferred in which cities?')
fig.show(height=200,width=200)
```



Observation:

Delhi shops the most online, followed by Noida and Bangalore.

In [20]:
```
dataset = data[['age','city','gender']]

fig = px.histogram(dataset,x='city',color='gender',title='Online shopping in cities based on Gender:')
fig.show()
```

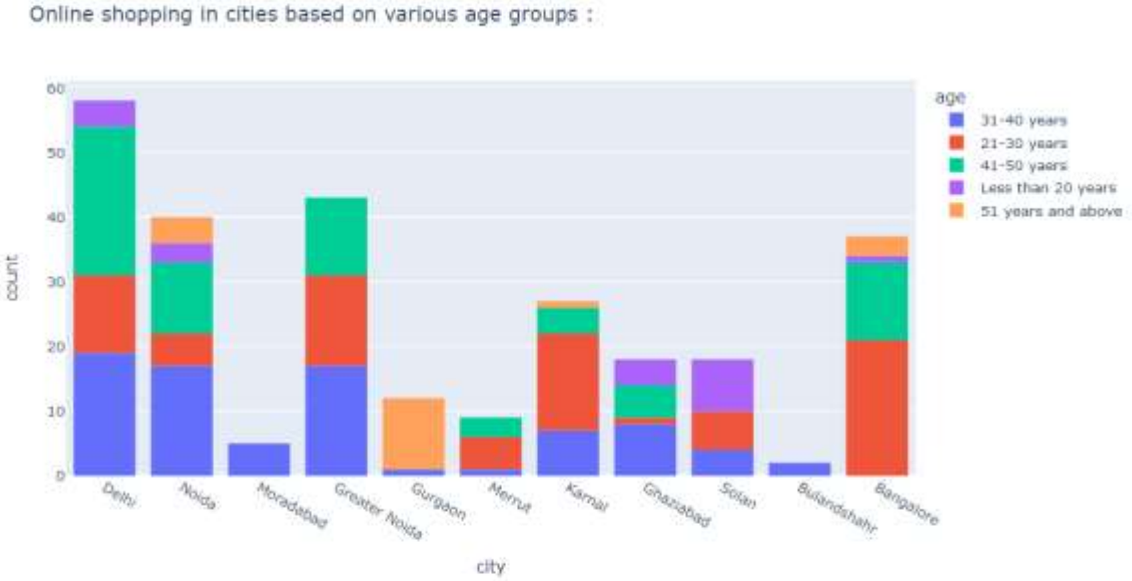


Observations:

1. Surprisingly, in Delhi and Noida, males prefer online shopping more than females. So we know our target audience in these cities,right?
2. Bulandshahr and Moradabad - both cities in UttarPradesh no female shoppers at all.

In [21]:
```
def value_count(column):
    diff_count=len(data[column].value_counts())
    if diff_count<5:
        plt.figure(figsize=(10,5))
    elif diff_count<10:
        plt.figure(figsize=(10,6))
        plt.xticks(rotation=90)
    elif diff_count<20:
        plt.figure(figsize=(25,6))
        plt.xticks(rotation=90)
    else:
        plt.figure(figsize=(20,6))
        plt.xticks(rotation=90)
    sns.countplot(x=column,data=data,orient='v')
    plt.show()
    #checking percentage of data classification in each string attribute
    print(round(data[i].value_counts()/269*100),2)
```

In [22]:
```
fig = px.histogram(dataset,x='city',color='age',title='Online shopping in cities based on various age groups : ')
fig.show()
```



Observations:

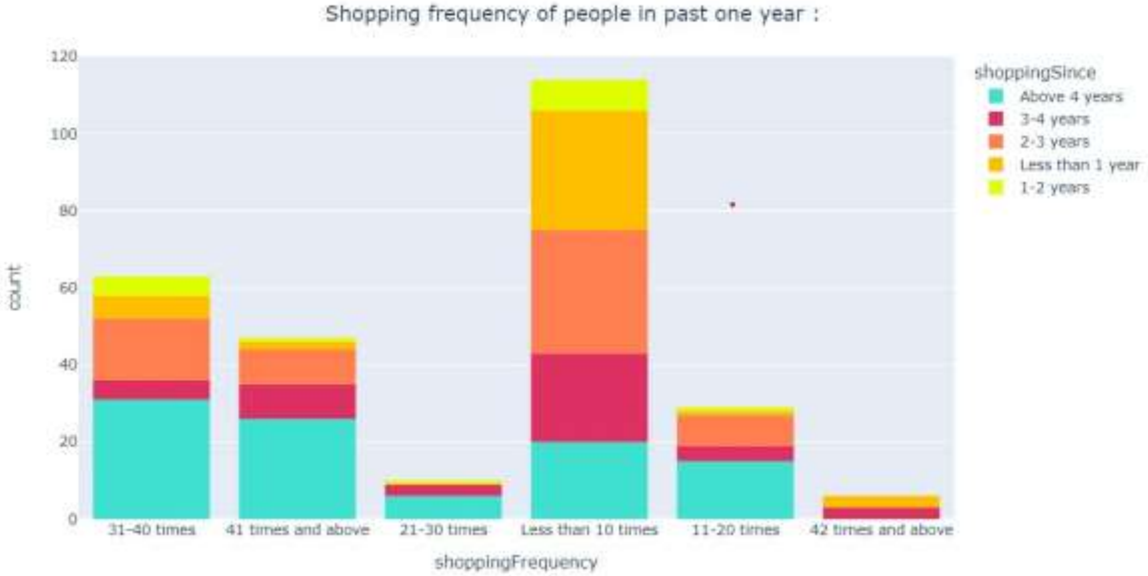
1. In Delhi, people aged between 41-50 years, prefer to shop on online more, when compared to other age groups.

2. In Bangalore, it is people aged between 21-30 years of age.

3. In the rest of the cities, on an average, people aged between 31-40 years prefer online shopping, except Gurgaon wher e senior citizens are more involved

In [23]:
```
# Let's check the Average Shopping frequency in last 1 year

fig = px.histogram(data,x='shoppingFrequency',color='shoppingSince',
                   color_discrete_map={'Above 4 years':'#40E0D0','3-4 years':'#DE3163','2-3 years':'#FF7F50',
                                       'Less than 1 year':'#FFBF00','1-2 years':'#DFFF00'})
fig.update_layout(title='Shopping frequency of people in past one year : ', title_x=0.5)

fig.show()
```

In [23]:
```
# Let's check the Average Shopping frequency in last 1 year

fig = px.histogram(data,x='shoppingFrequency',color='shoppingSince',
                   color_discrete_map={'Above 4 years':'#40E0D0','3-4 years':'#DE3163','2-3 years':'#FF7F50',
                                       'Less than 1 year':'#FFBF00','1-2 years':'#DFFF00'})
fig.update_layout(title='Shopping frequency of people in past one year : ', title_x=0.5)

fig.show()
```



Observations:

1. In the past 1 year, on an average people have shopped approximately 30-40 times.

2. From the data, we can see that, people who have been shopping for more then 3-4 years are the ones who frequently shops.

In [24]:
```
# Let's check the Cancellation reasons

reasons = pd.DataFrame(data['cancellationReason'].value_counts()).reset_index()
reasons.columns = ['Reason','Count']

fig = go.Figure(go.Funnelarea(
    text = reasons['Reason'],
    values = reasons['Count'],
    marker = {"colors": ["deepskyblue", "lightsalmon", "tan", "teal", "silver"],
              "line": {"color": ["wheat", "wheat", "blue", "wheat", "wheat"], "width": [0, 1, 5, 0, 4]}}
    ))

fig.update_layout(title = 'Reasons for not purchasing any product :',title_x=0.5)
fig.show()
```

Observations:

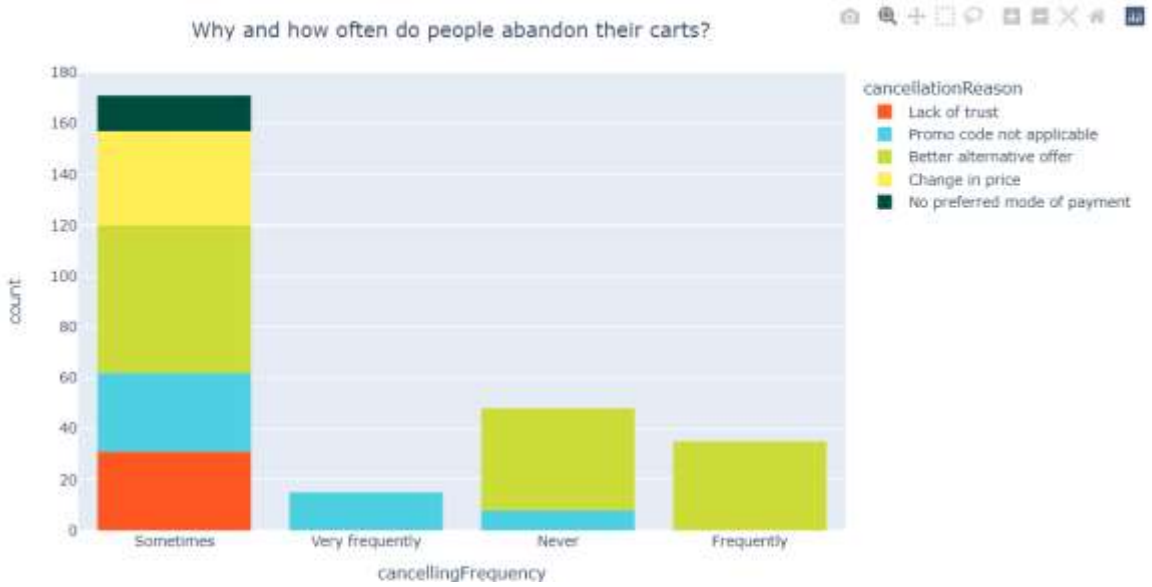
1. We can observe that most of the people, abandoned their cart as they were having better alternative offers.

2. People even abandon their carts, because they do not have their preferred mode of payment.

3. So when the preferred payment mode of people who cancel their cart was analysed, it appears that they preferred CoD, but was not available for that product, which can also imply that the customer may have some difficulty in trusting the retail store?

4. Lack of trust is also a reason for cancelling their product

In [25]:
```
fig = px.histogram(data,x='cancellingFrequency',color='cancellationReason',
                   color_discrete_map={'Lack of trust':'#FF5722','Promo code not applicable':'#4DD0E1',
                                       'Better alternative offer':'#CDDC39','Change in price':'#FFEE58',
                                       'No preferred mode of payment':'#004D40'})
fig.update_layout(title='Why and how often do people abandon their carts? ', title_x=0.4)

fig.show()
```



```
In [26]: fig = px.histogram(data,x='cancellationReason',color='paymentMode')
fig.update_layout(title='Payment Mode vs Cancellation Reasons : ', title_x=0.5)

fig.show()
```

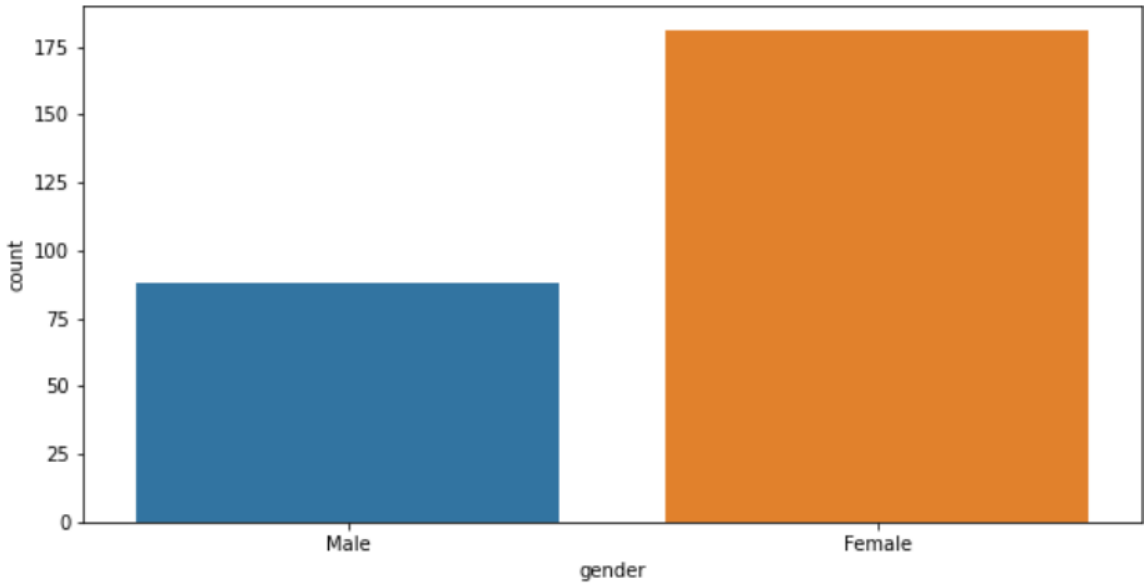
Observations:

1. So when the preferred payment mode of people who cancel their cart was analysed, it appears that they preferred CoD, but was not available for that product, which can also imply that the customer may have some difficulty in trusting the retail store?

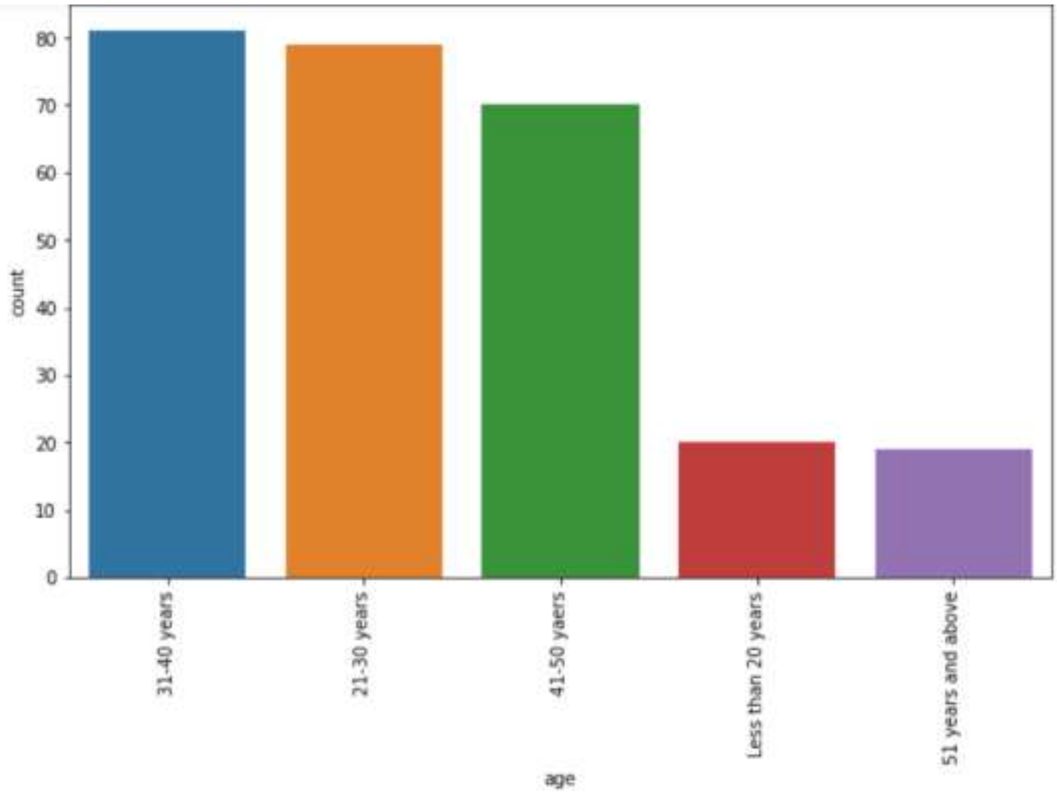
2. Lack of trust is also a reason for cancelling their product, and the payment mode is e-wallets, which they believe might not be reliable?

```
In [27]: def value_count(column):
    diff_count=len(data[column].value_counts())
    if diff_count<5:
        plt.figure(figsize=(10,5))
    elif diff_count<10:
        plt.figure(figsize=(10,6))
        plt.xticks(rotation=90)
    elif diff_count<20:
        plt.figure(figsize=(25,6))
        plt.xticks(rotation=90)
    else:
        plt.figure(figsize=(20,6))
        plt.xticks(rotation=90)
    sns.countplot(x=column,data=data,orient='v')
    plt.show()

# Let's check the percentage of data classification in each string attribute
    print(round(data[i].value_counts()/269*100),2)
```

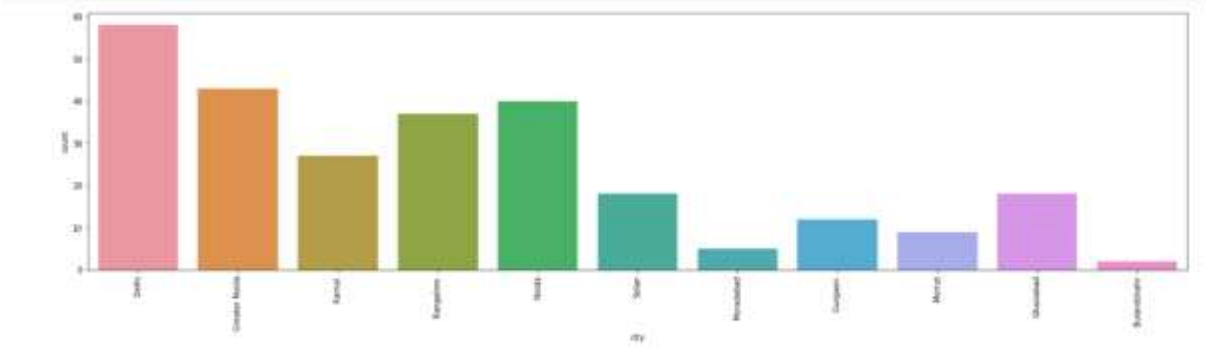
```
In [28]: for i in categorical:
    value_count(i)
```



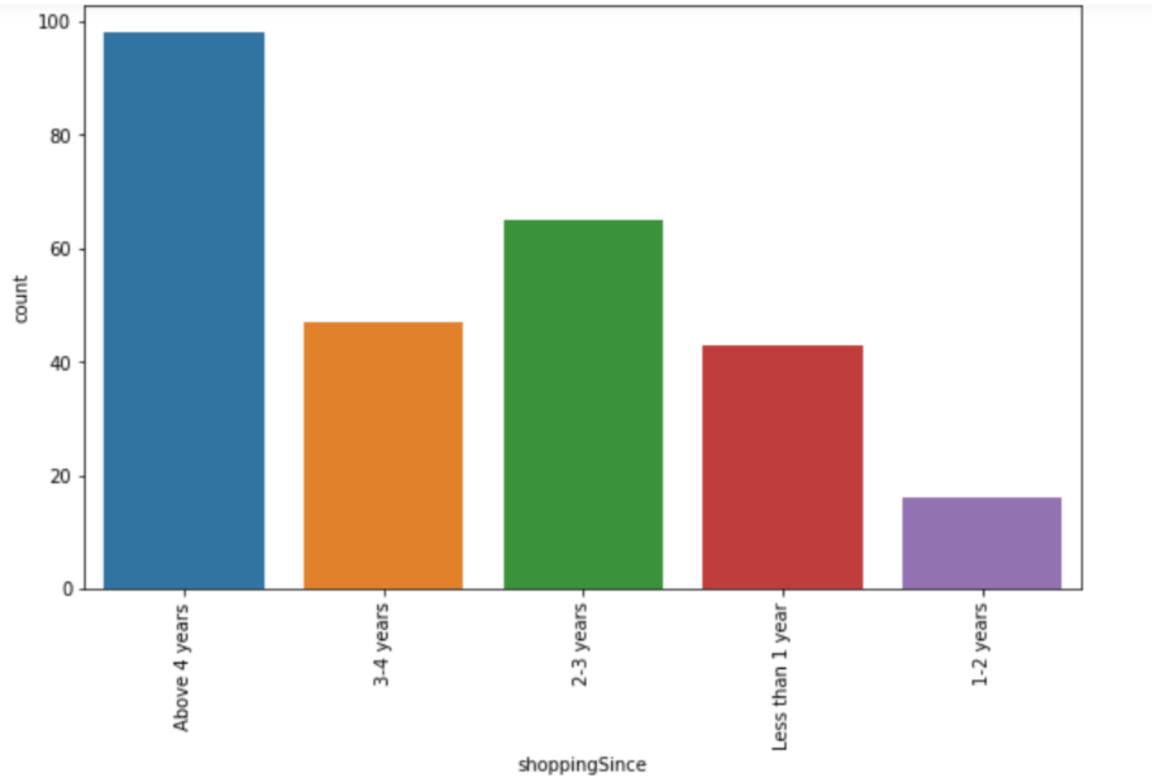

```
Female    67.0
Male      33.0
Name: gender, dtype: float64 2
```



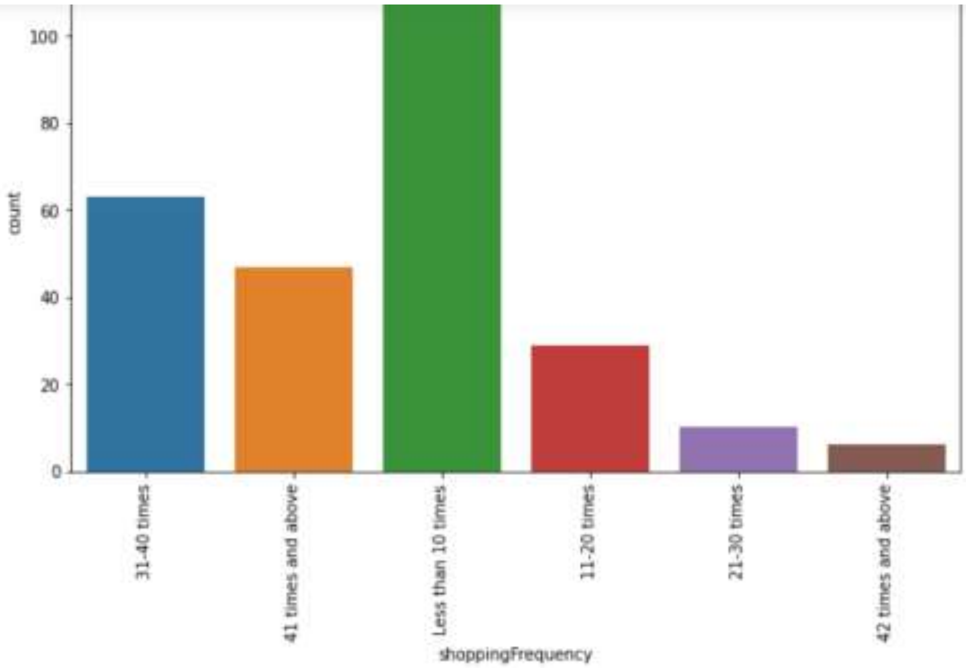

```
31-40 years          30.0
21-30 years          29.0
41-50 yaers          26.0
Less than 20 years    7.0
```



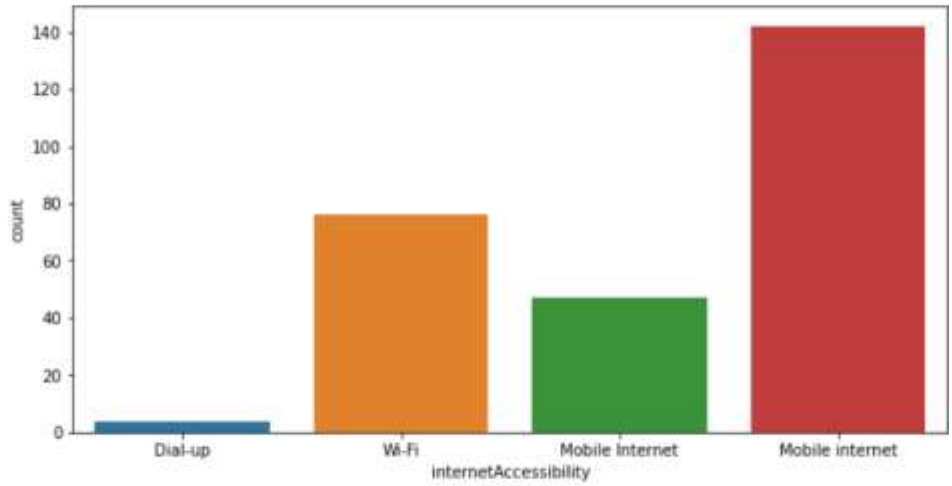
```
Delhi            22.0
Greater Noida    16.0
Noida            15.0
Bangalore        14.0
Karnal           10.0
Solan             7.0
Ghaziabad         7.0
Gurgaon           4.0
Merrut            3.0
Moradabad         2.0
Bulandshahr       1.0
Name: city, dtype: float64 2
```



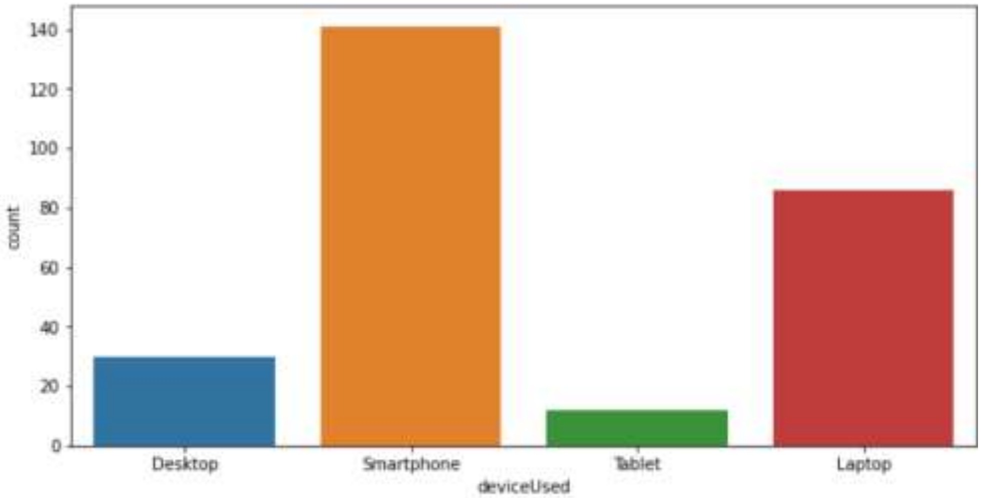

```
Above 4 years       36.0
2-3 years           24.0
3-4 years           17.0
Less than 1 year    16.0
1-2 years            6.0
```



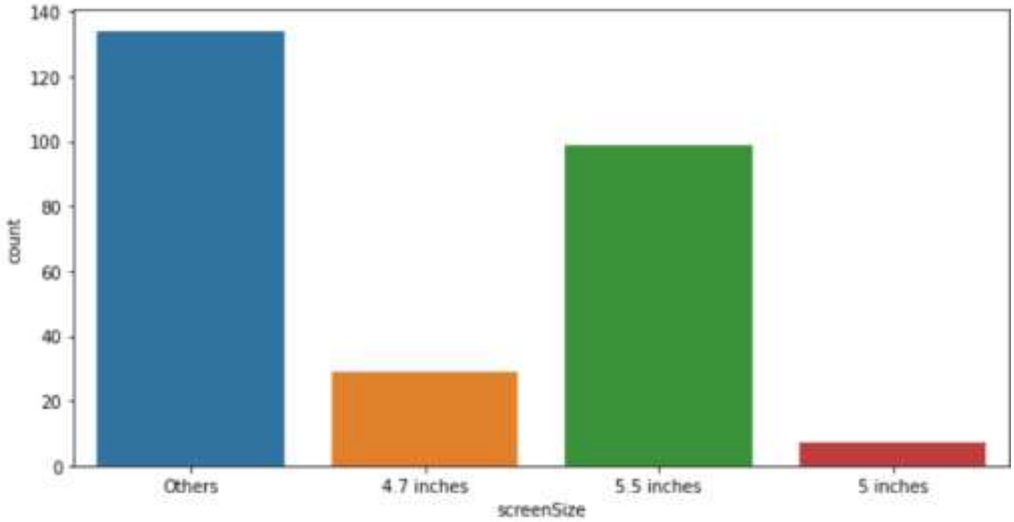

```
Less than 10 times    42.0
31-40 times           23.0
41 times and above    17.0
11-20 times           11.0
21-30 times            4.0
42 times and above     2.0
```



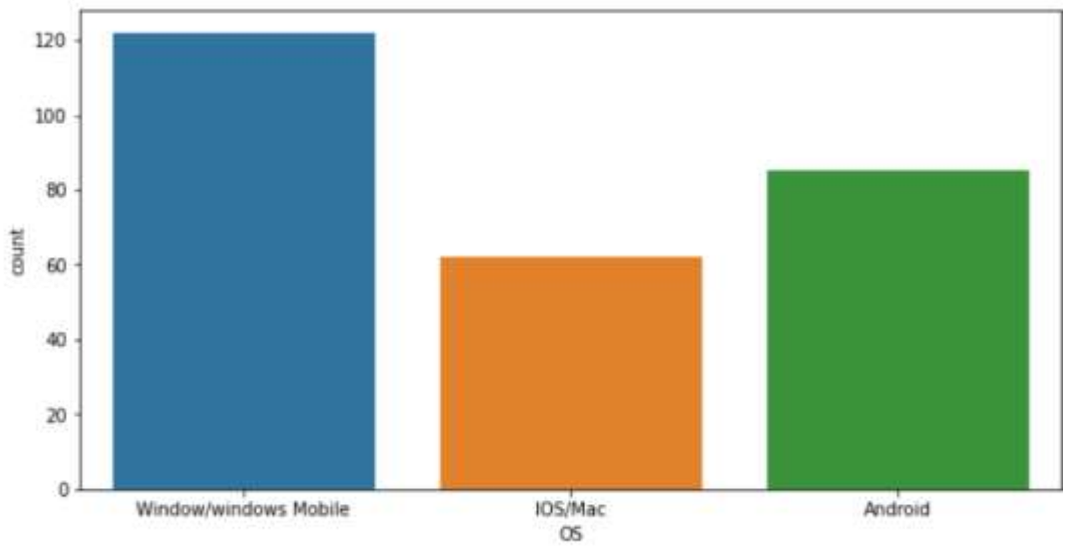

```
Mobile internet    53.0
Wi-Fi              28.0
Mobile Internet    17.0
Dial-up             1.0
Name: internetAccessibility, dtype: float64 2
```



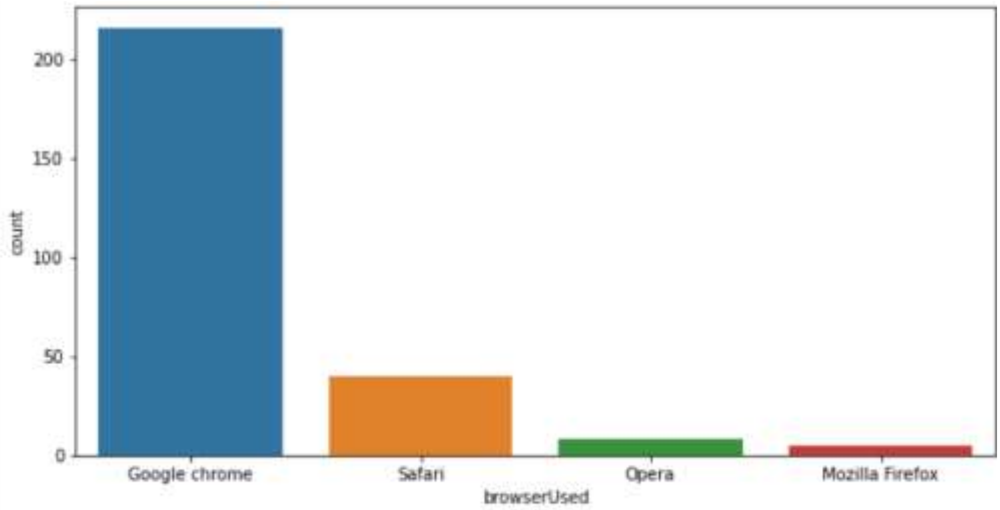

```
Smartphone    52.0
Laptop        32.0
Desktop       11.0
Tablet         4.0
Name: deviceUsed, dtype: float64 2
```



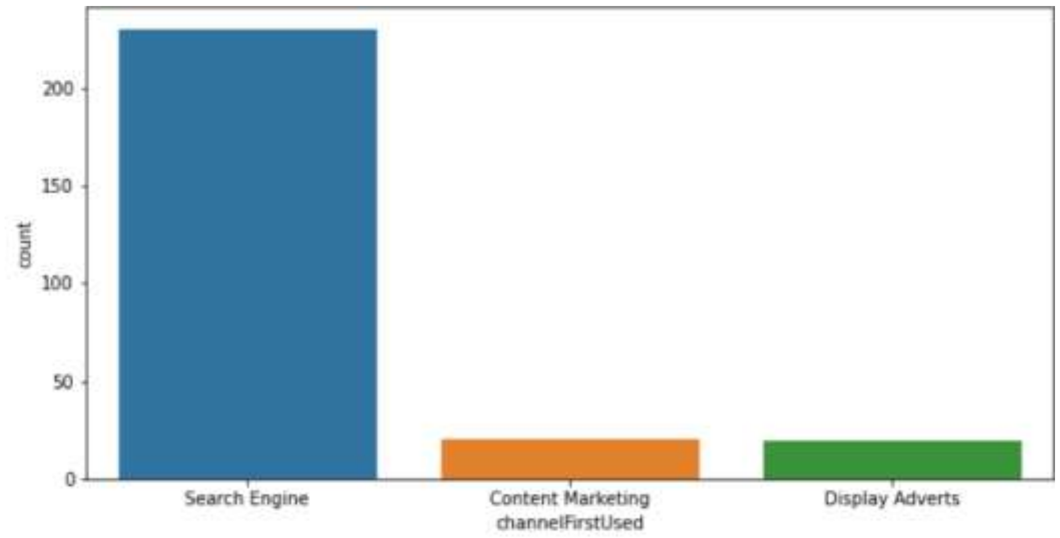

```
Others        50.0
5.5 inches    37.0
4.7 inches    11.0
5 inches       3.0
Name: screenSize, dtype: float64 2
```



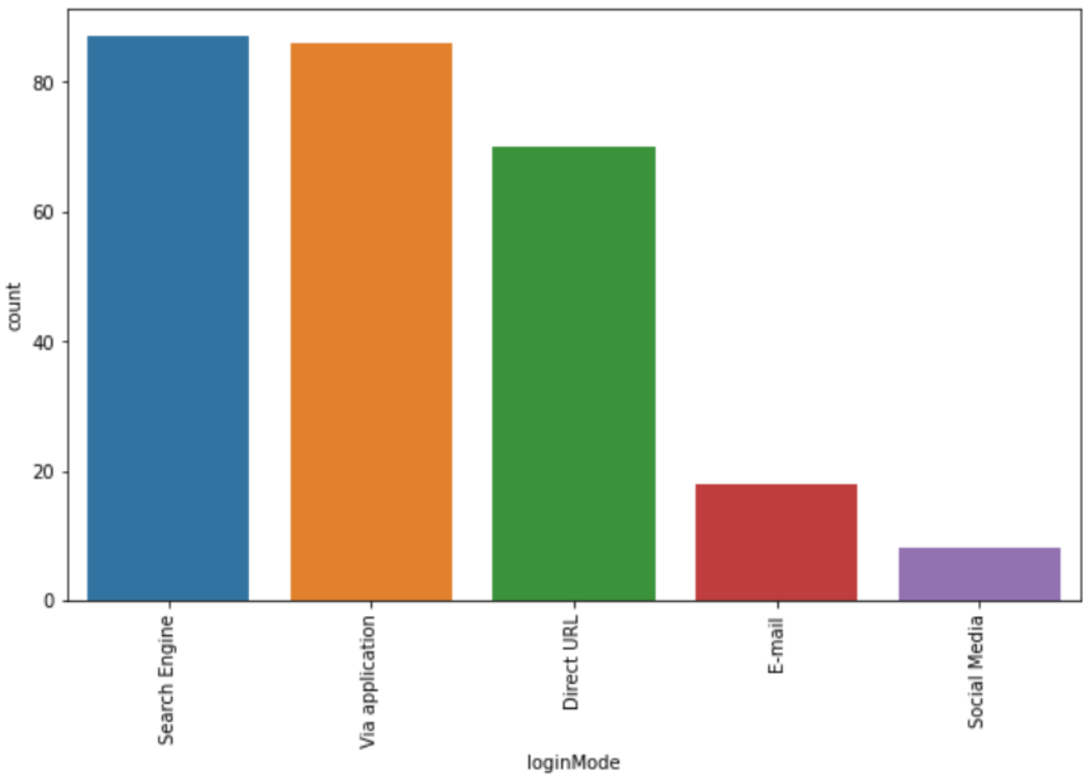

```
Window/windows Mobile    45.0
Android                  32.0
IOS/Mac                  23.0
Name: OS, dtype: float64 2
```



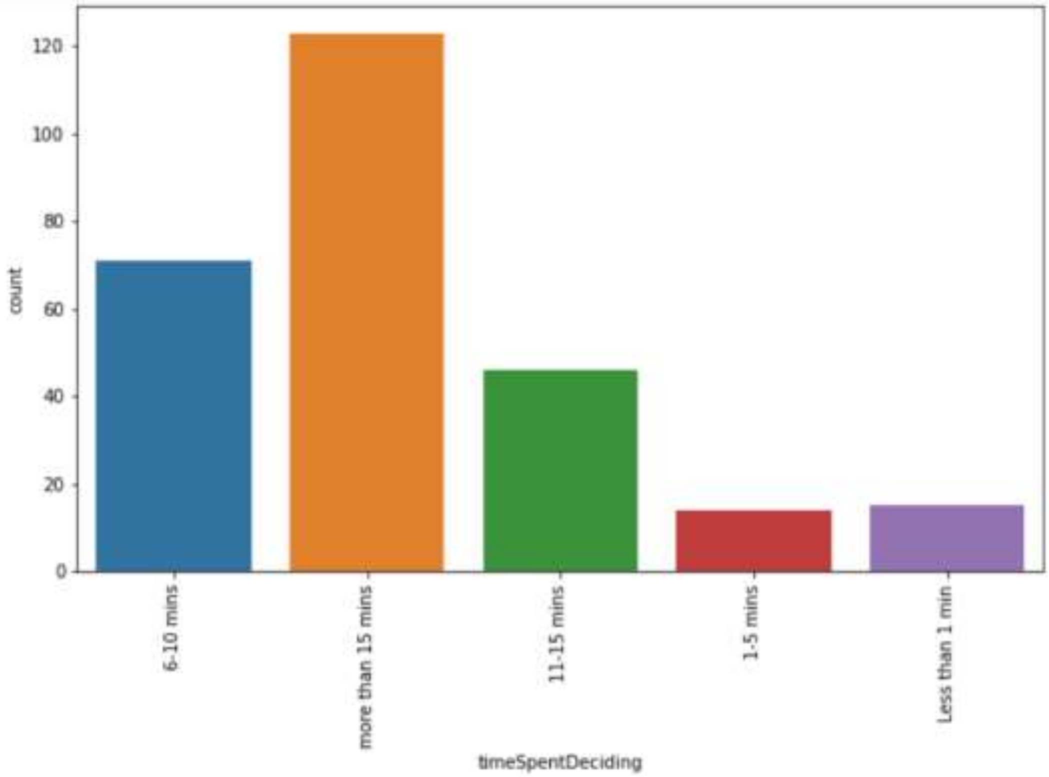

```
Google chrome      80.0
Safari             15.0
Opera               3.0
Mozilla Firefox     2.0
Name: browserUsed, dtype: float64 2
```



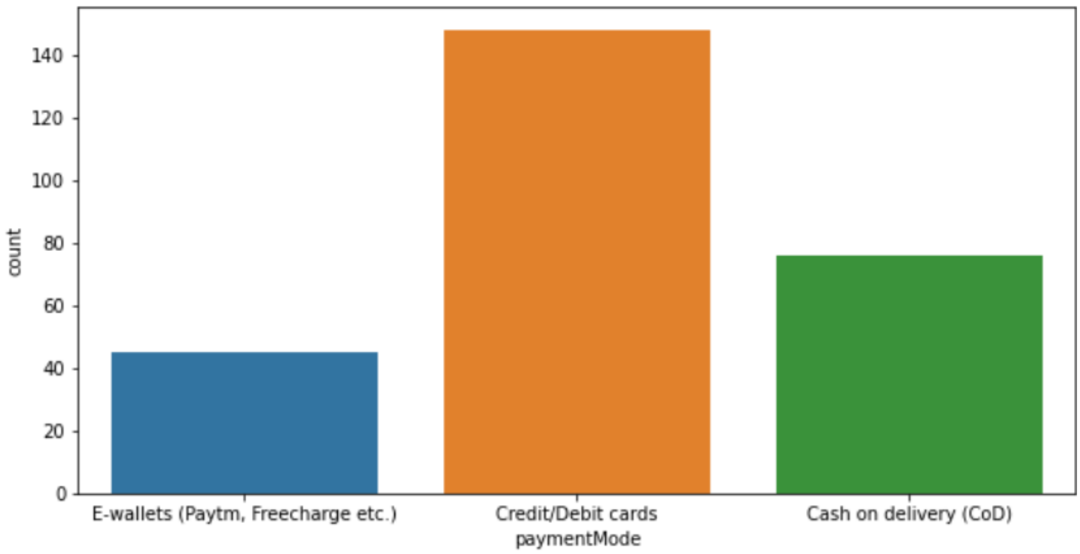

```
Search Engine        86.0
Content Marketing     7.0
Display Adverts       7.0
Name: channelFirstUsed, dtype: float64 ?
```



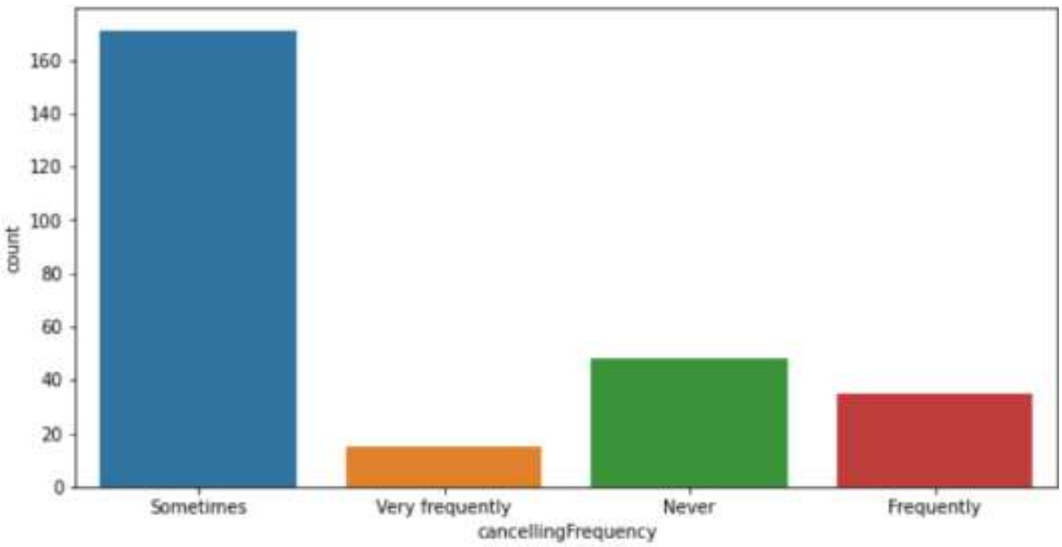

```
Search Engine      32.0
Via application    32.0
Direct URL         26.0
E-mail              7.0
Social Media        3.0
```



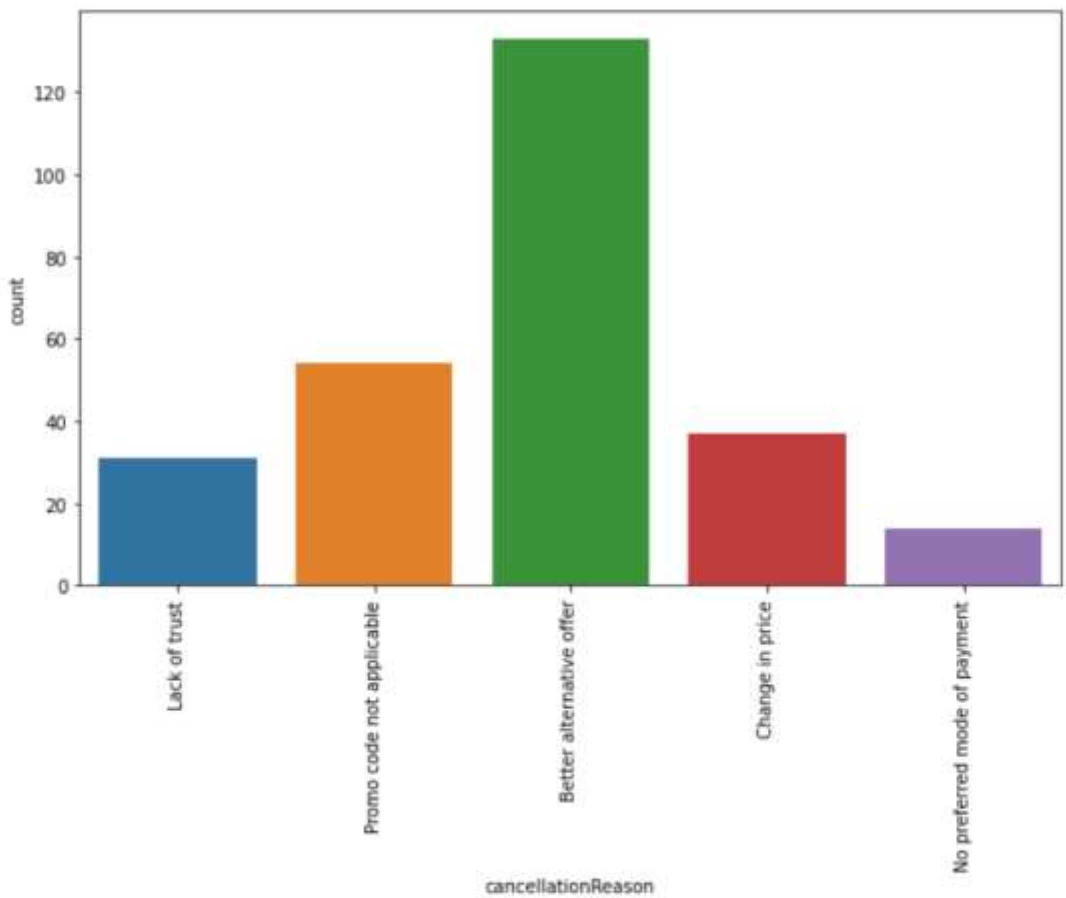

```
more than 15 mins     46.0
6-10 mins             26.0
11-15 mins            17.0
```




```
Credit/Debit cards                    55.0
Cash on delivery (CoD)                28.0
E-wallets (Paytm, Freecharge etc.)    17.0
Name: paymentMode, dtype: float64 2
```

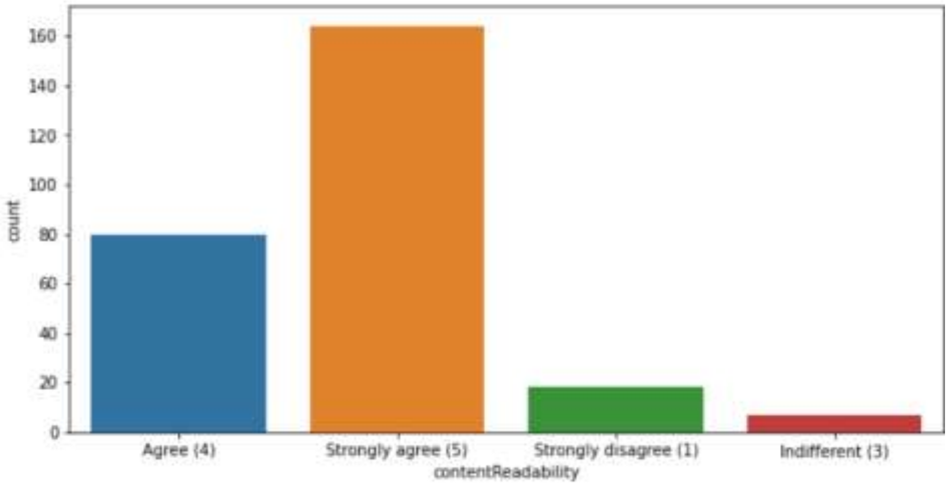



```
Sometimes          64.0
Never              18.0
Frequently         13.0
Very frequently     6.0
Name: cancellingFrequency, dtype: float64 2
```



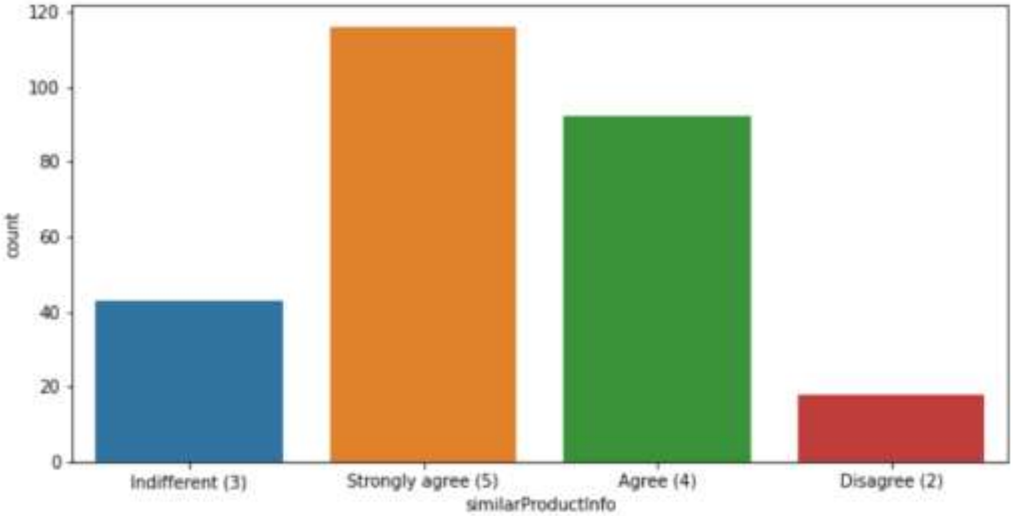
cancellationReason

```
Better alternative offer        49.0
Promo code not applicable       20.0
Change in price                 14.0
Lack of trust                   12.0
No preferred mode of payment     5.0
Name: cancellationReason, dtype: float64 2
```

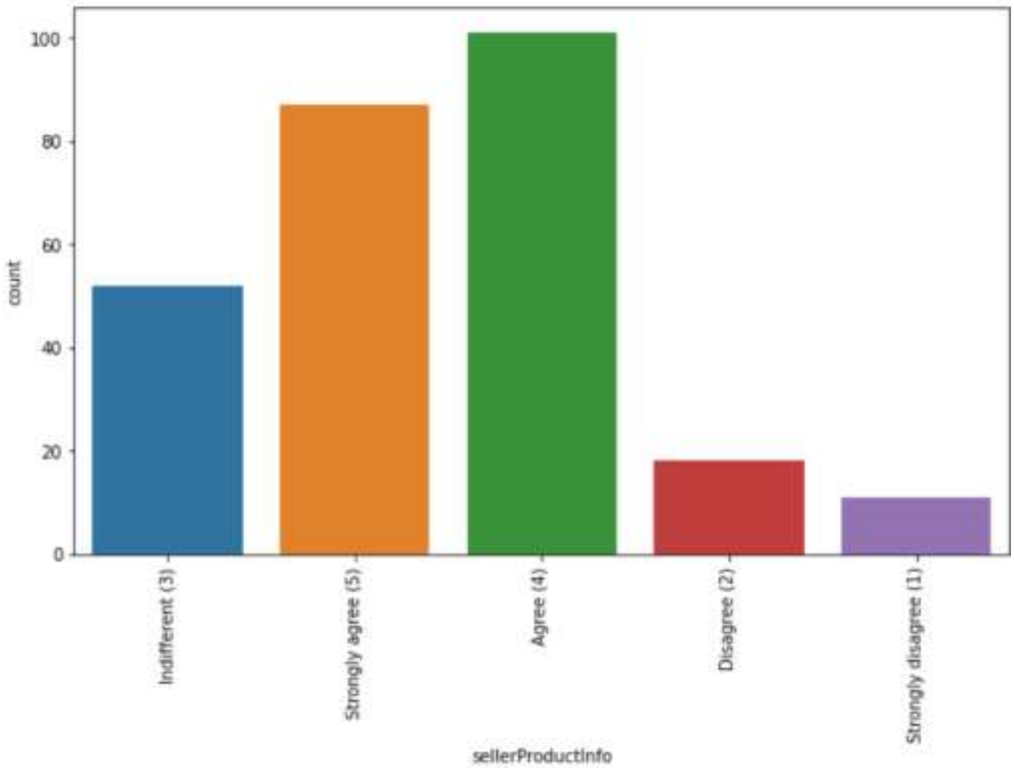



```
Strongly agree (5)       61.0
Agree (4)                30.0
Strongly disagree (1)     7.0
Indifferent (3)           3.0
Name: contentReadability, dtype: float64 2
```

```
Strongly agree (5)       61.0
Agree (4)                30.0
Strongly disagree (1)     7.0
Indifferent (3)           3.0
Name: contentReadability, dtype: float64 2
```

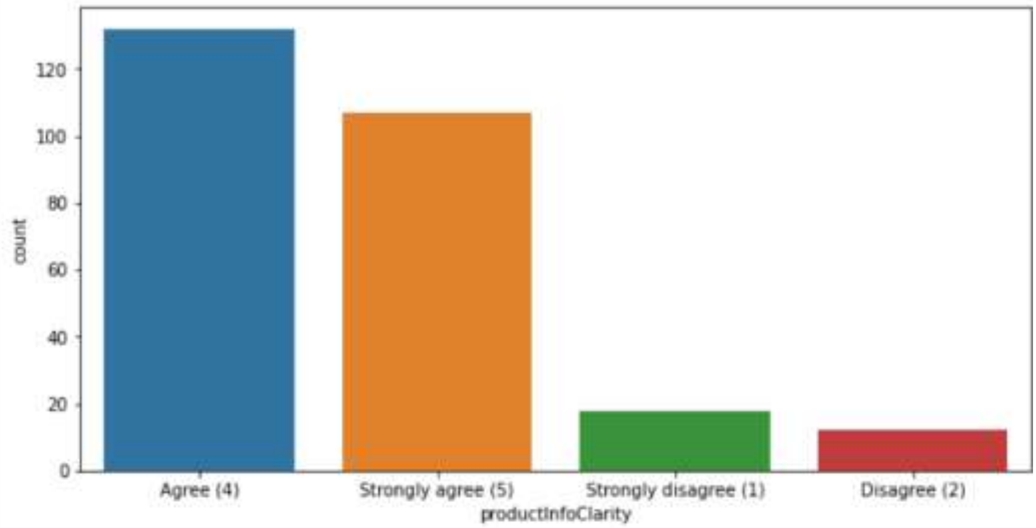



```
Strongly agree (5)    43.0
Agree (4)             34.0
Indifferent (3)       16.0
Disagree (2)           7.0
Name: similarProductInfo, dtype: float64 2
```

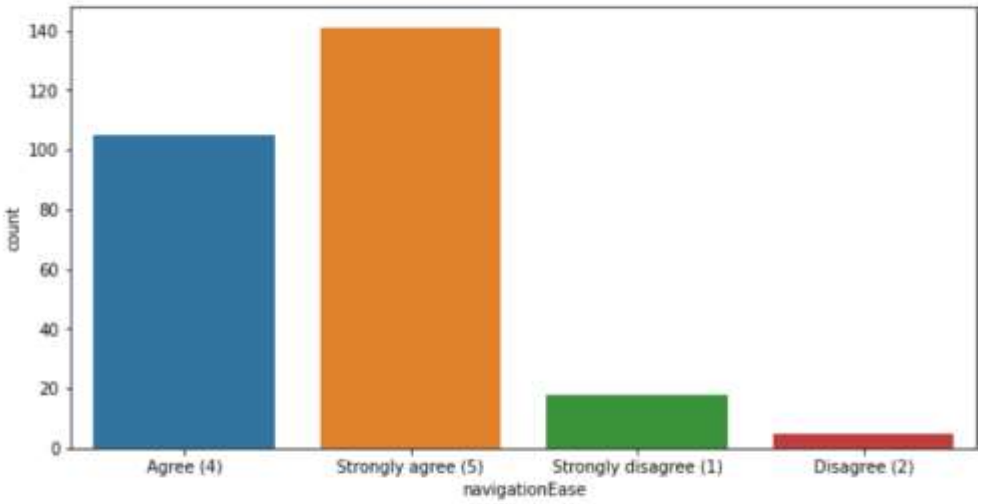


sellerProductInfo

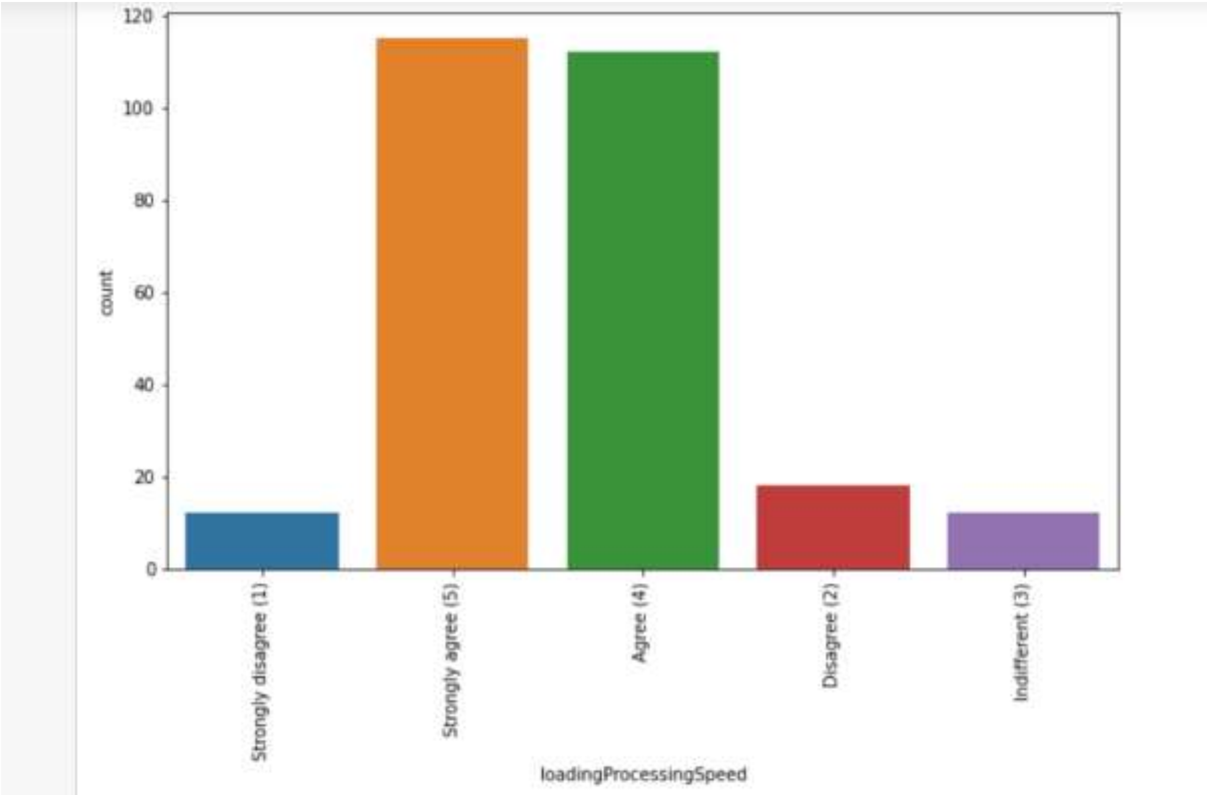
```
Agree (4)                38.0
Strongly agree (5)       32.0
Indifferent (3)          19.0
Disagree (2)              7.0
Strongly disagree (1)     4.0
Name: sellerProductInfo, dtype: float64 2
```




```
Agree (4)                49.0
Strongly agree (5)       40.0
Strongly disagree (1)     7.0
Disagree (2)              4.0
Name: productInfoClarity, dtype: float64 2
```

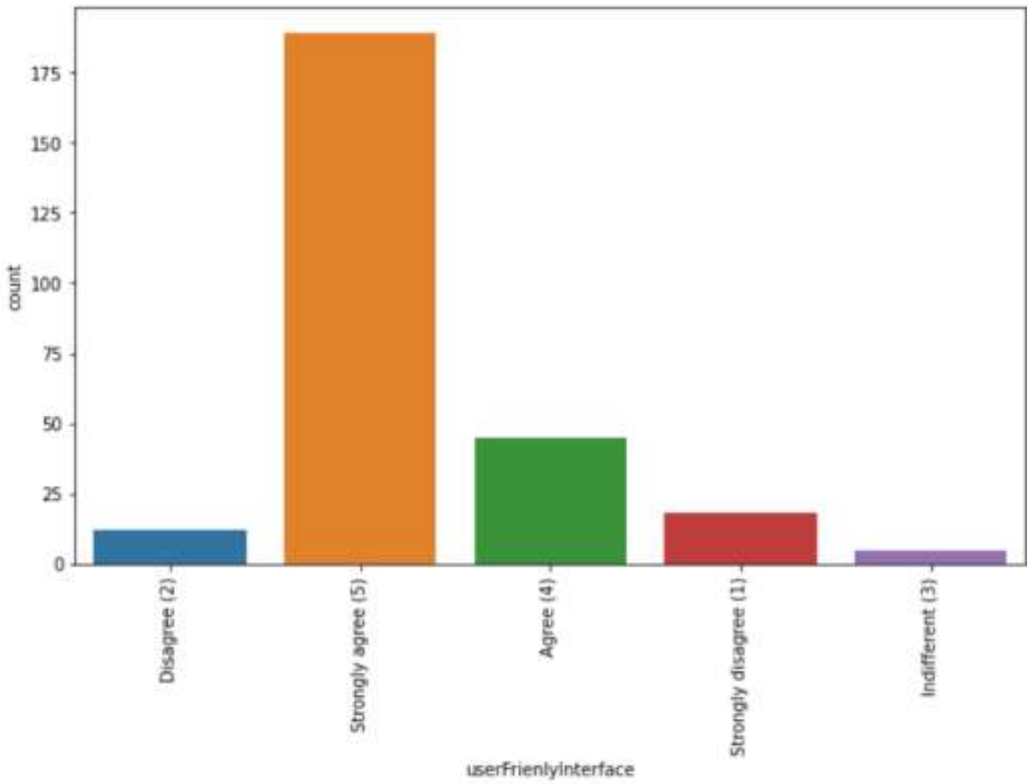



```
Strongly agree (5)       52.0
Agree (4)                39.0
Strongly disagree (1)     7.0
Disagree (2)              2.0
Name: navigationEase, dtype: float64 2
```

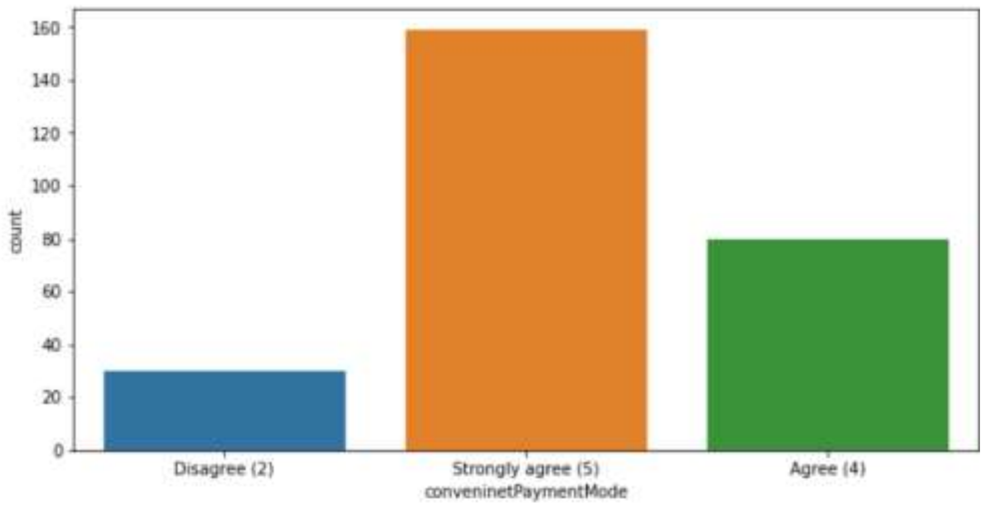


loadingProcessingSpeed

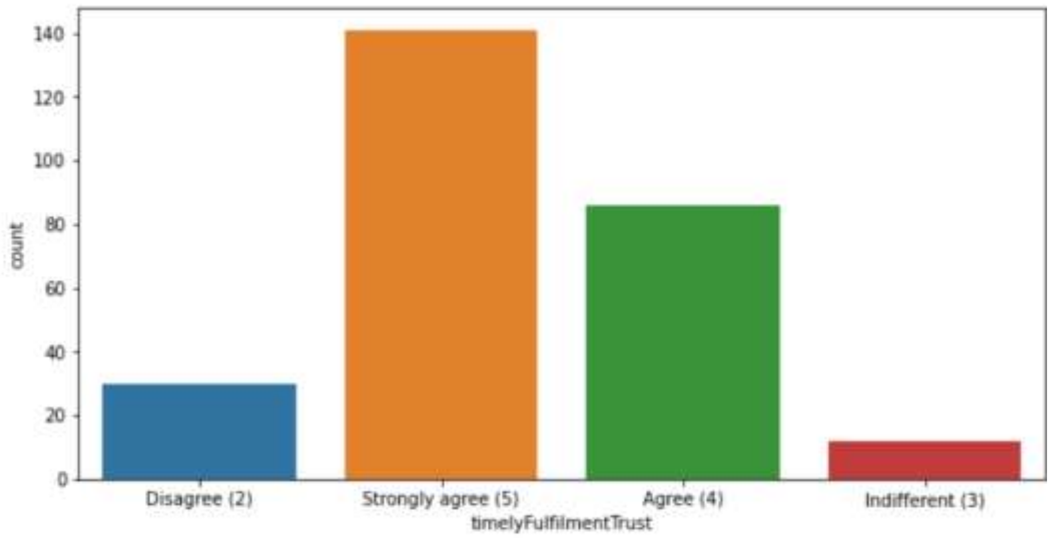
```
Strongly agree (5)       43.0
Agree (4)                42.0
Disagree (2)              7.0
Strongly disagree (1)     4.0
Indifferent (3)           4.0
Name: loadingProcessingSpeed, dtype: float64 2
```



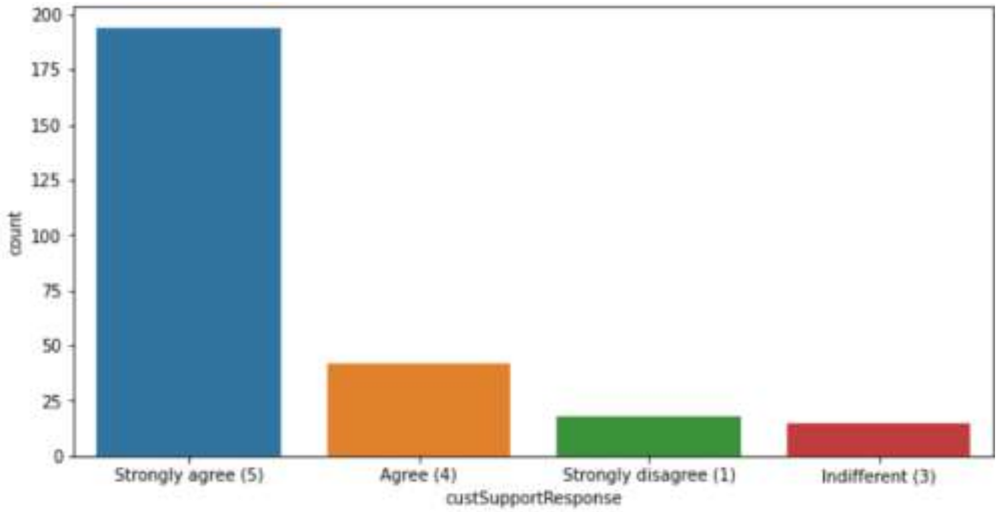

```
Strongly agree (5)       70.0
Agree (4)                17.0
Strongly disagree (1)     7.0
Disagree (2)              4.0
Indifferent (3)           2.0
Name: userFrienlyInterface, dtype: float64 2
```



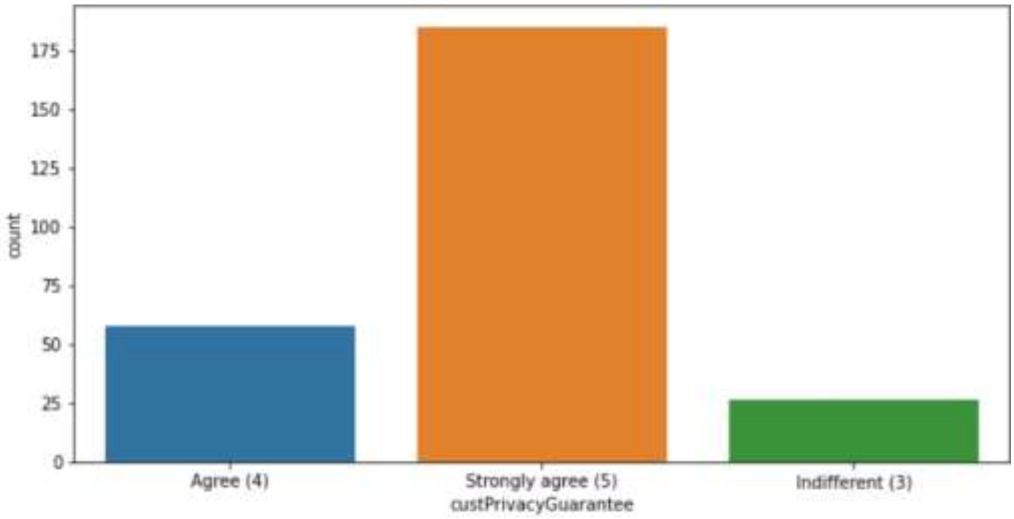

```
Strongly agree (5)     59.0
Agree (4)              30.0
Disagree (2)           11.0
Name: conveninetPaymentMode, dtype: float64 2
```



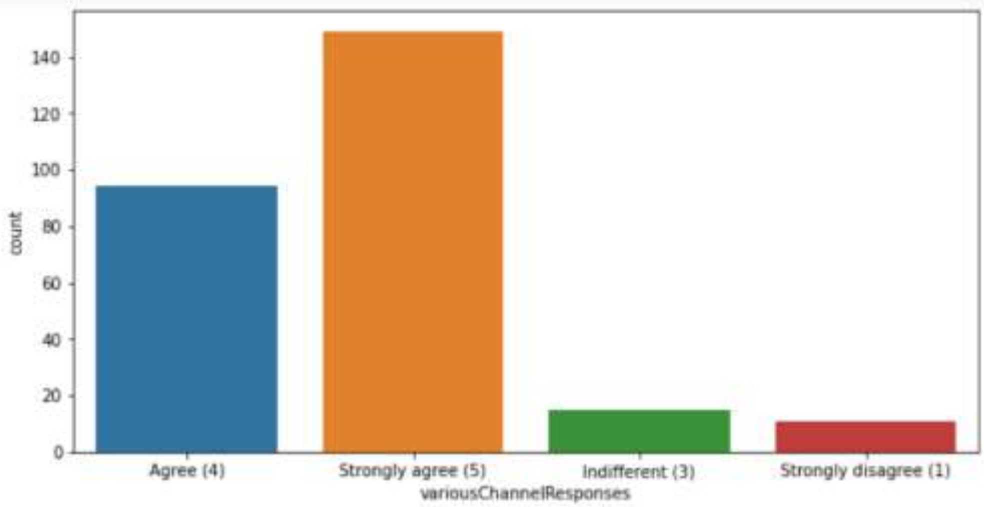

```
Strongly agree (5)     52.0
Agree (4)              32.0
Disagree (2)           11.0
Indifferent (3)         4.0
Name: timelyFulfilmentTrust, dtype: float64 2
```



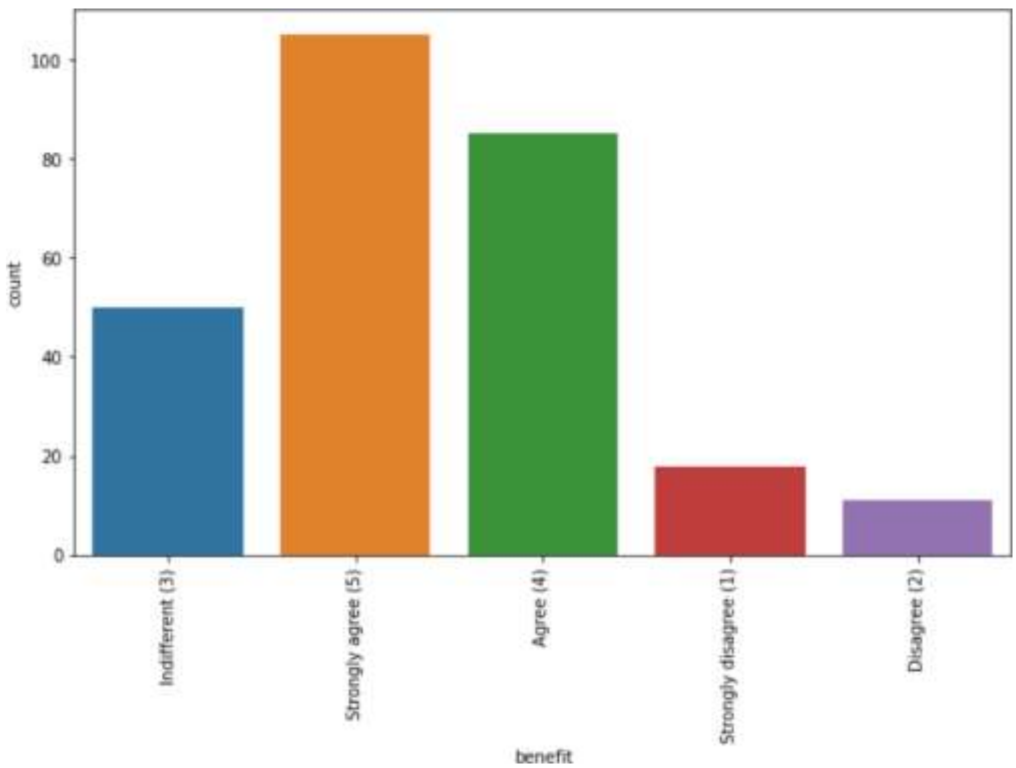

```
Strongly agree (5)       72.0
Agree (4)                16.0
Strongly disagree (1)     7.0
Indifferent (3)           6.0
Name: custSupportResponse, dtype: float64 2
```



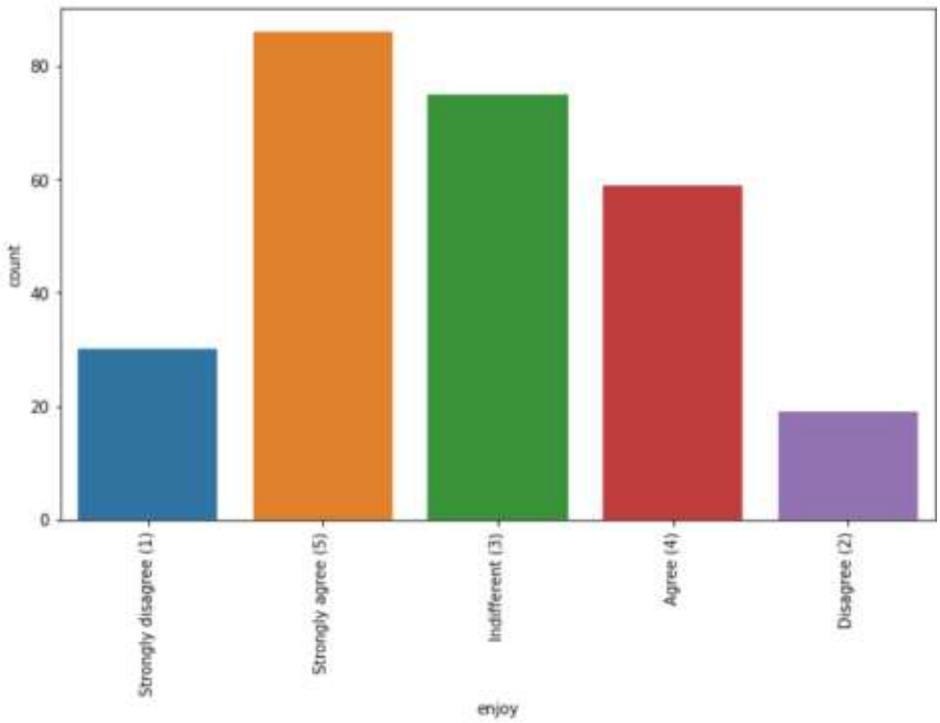

```
Strongly agree (5)    69.0
Agree (4)             22.0
Indifferent (3)       10.0
Name: custPrivacyGuarantee, dtype: float64 2
```



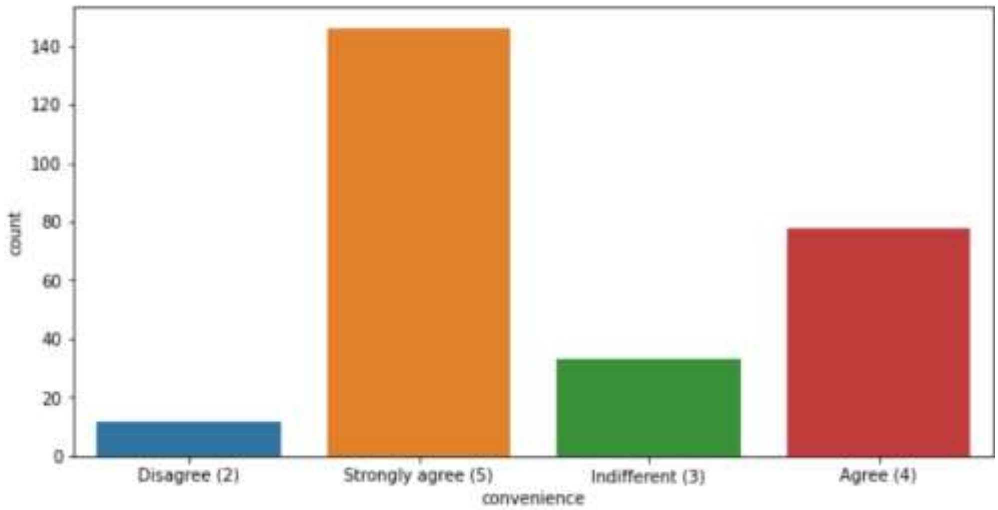

```
Strongly agree (5)       55.0
Agree (4)                35.0
Indifferent (3)           6.0
Strongly disagree (1)     4.0
Name: variousChannelResponses, dtype: float64 2
```



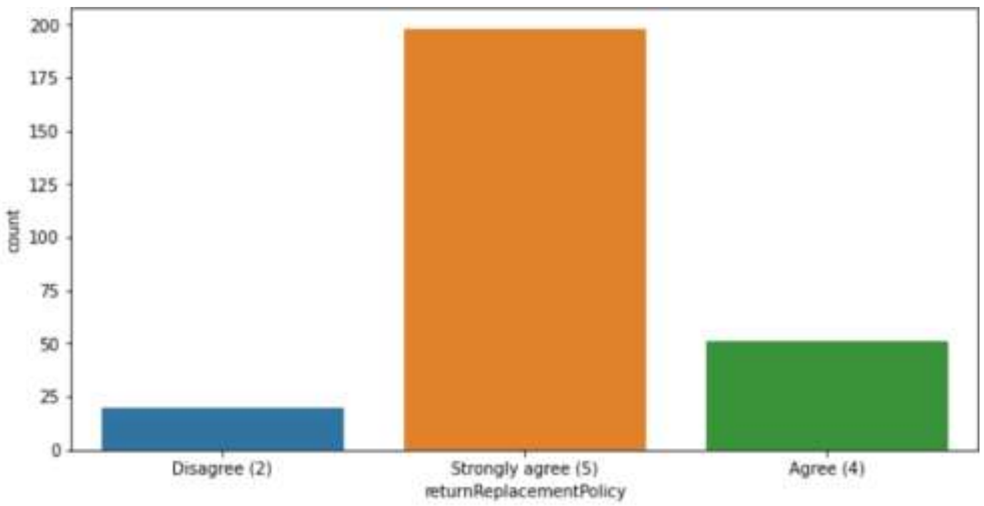

```
Strongly agree (5)       39.0
Agree (4)                32.0
Indifferent (3)          19.0
Strongly disagree (1)     7.0
Disagree (2)              4.0
Name: benefit, dtype: float64 2
```



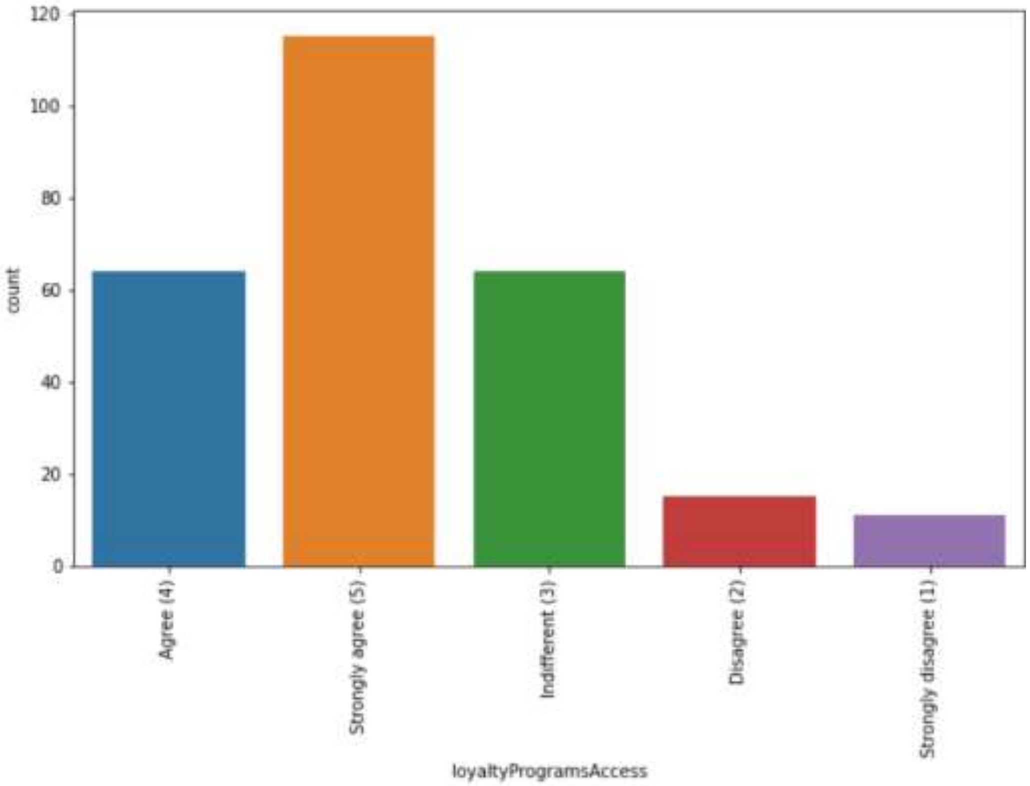

```
Strongly agree (5)       32.0
Indifferent (3)          28.0
Agree (4)                22.0
Strongly disagree (1)    11.0
Disagree (2)              7.0
Name: enjoy, dtype: float64 2
```



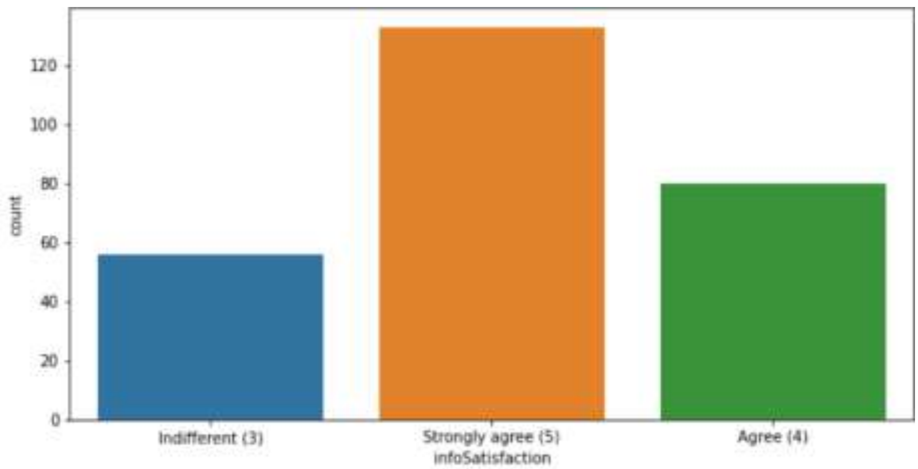

```
Strongly agree (5)    54.0
Agree (4)             29.0
Indifferent (3)       12.0
Disagree (2)           4.0
Name: convenience, dtype: float64 2
```



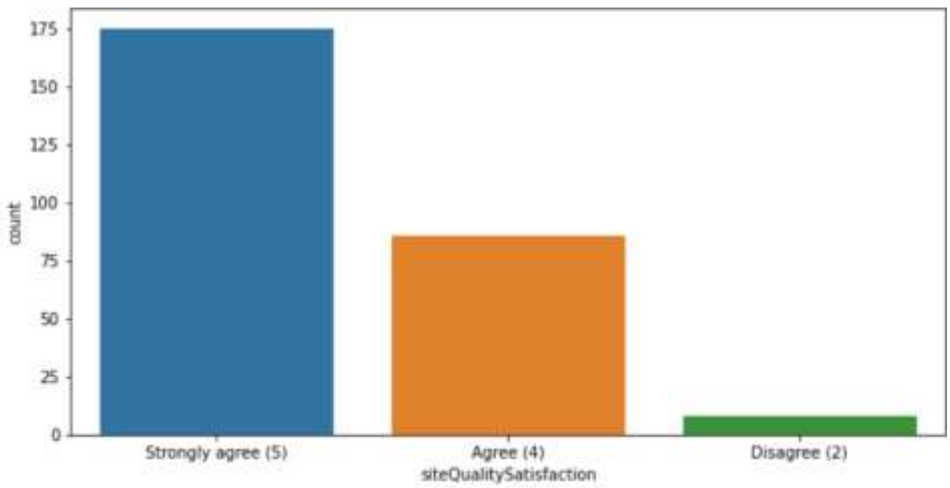

```
Strongly agree (5)     74.0
Agree (4)              19.0
Disagree (2)            7.0
Name: returnReplacementPolicy, dtype: float64 2
```



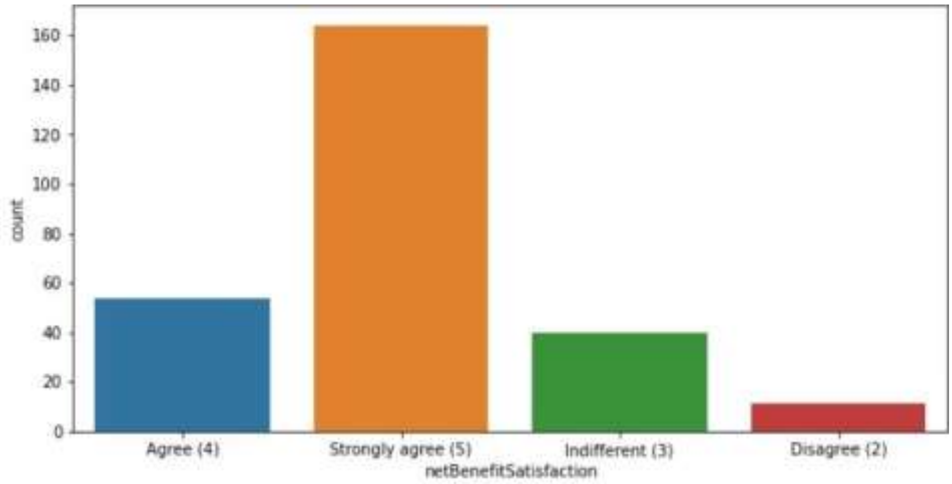

```
Strongly agree (5)       43.0
Agree (4)                24.0
Indifferent (3)          24.0
Disagree (2)              6.0
Strongly disagree (1)     4.0
Name: loyaltyProgramsAccess, dtype: float64 2
```



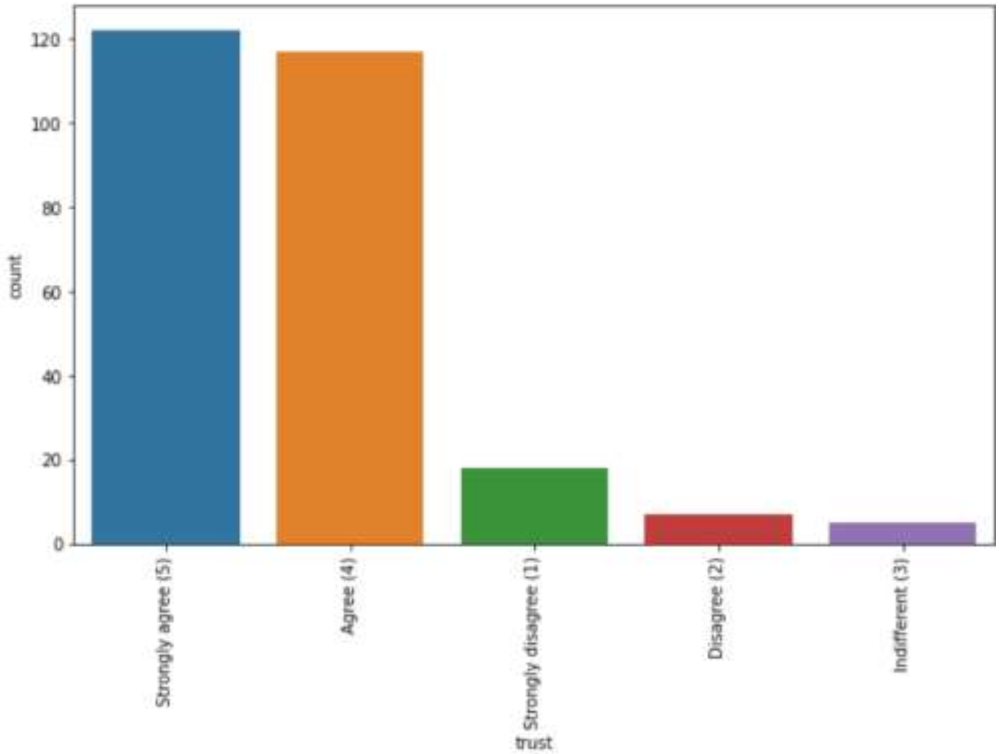

```
Strongly agree (5)     49.0
Agree (4)              30.0
Indifferent (3)        21.0
Name: infoSatisfaction, dtype: float64 2
```



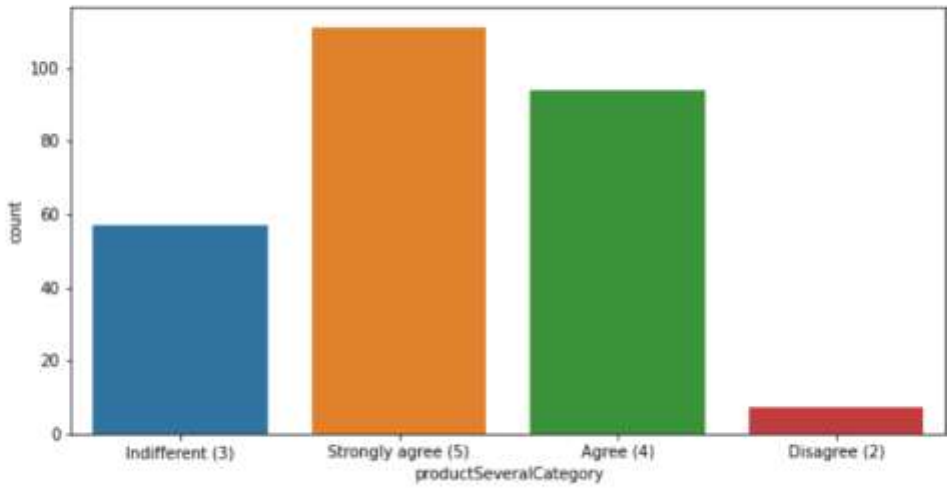

```
Strongly agree (5)     65.0
Agree (4)              32.0
Disagree (2)            3.0
Name: siteQualitySatisfaction, dtype: float64 2
```



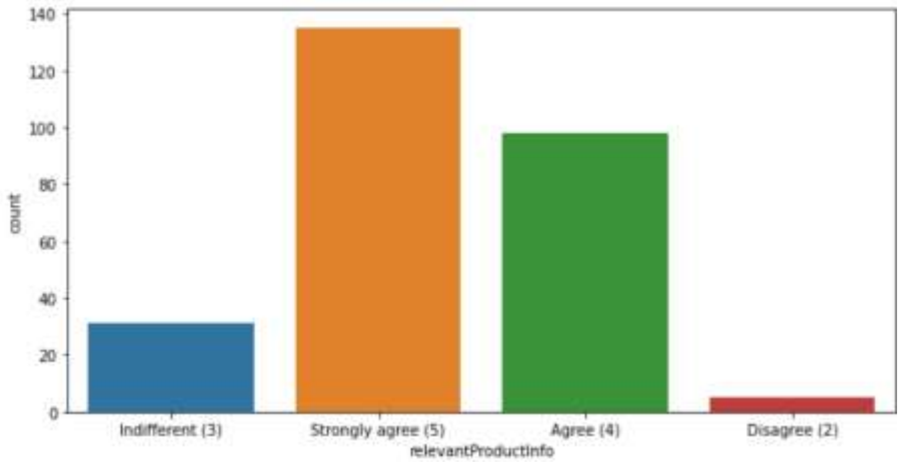

```
Strongly agree (5)     61.0
Agree (4)              20.0
Indifferent (3)        15.0
Disagree (2)            4.0
Name: netBenefitSatisfaction, dtype: float64 2
```



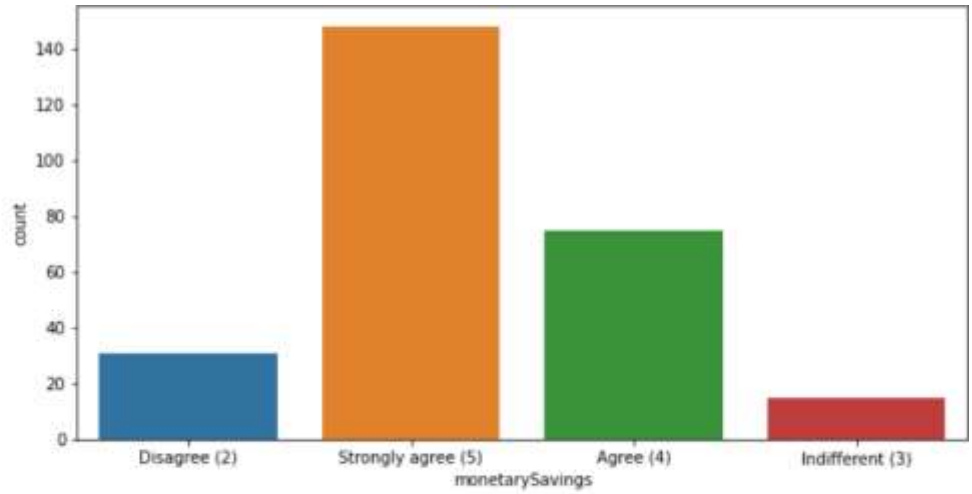

```
Strongly agree (5)       45.0
Agree (4)                43.0
Strongly disagree (1)     7.0
Disagree (2)              3.0
Indifferent (3)           2.0
Name: trust, dtype: float64 2
```



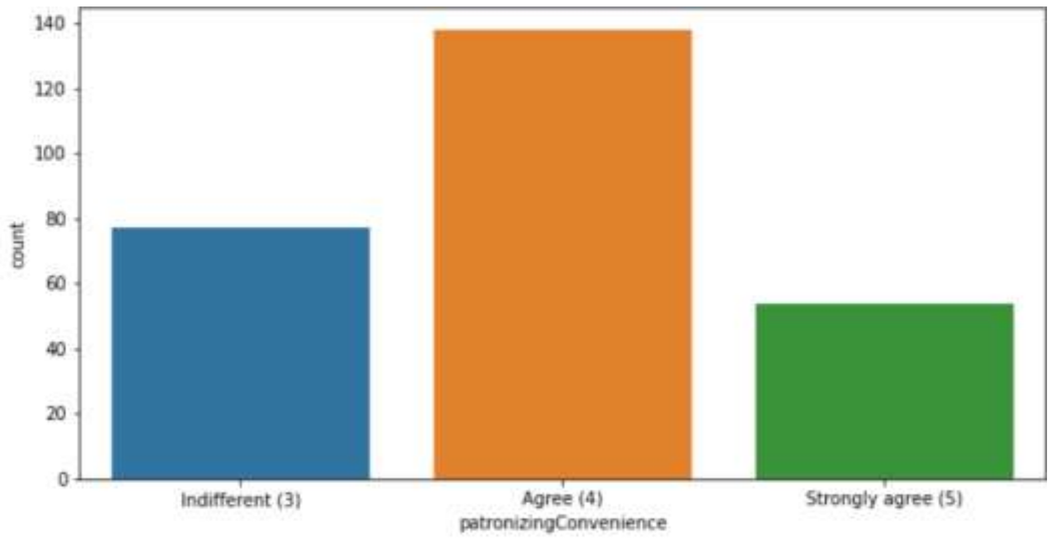

```
Strongly agree (5)    41.0
Agree (4)             35.0
Indifferent (3)       21.0
Disagree (2)           3.0
Name: productSeveralCategory, dtype: float64 2
```



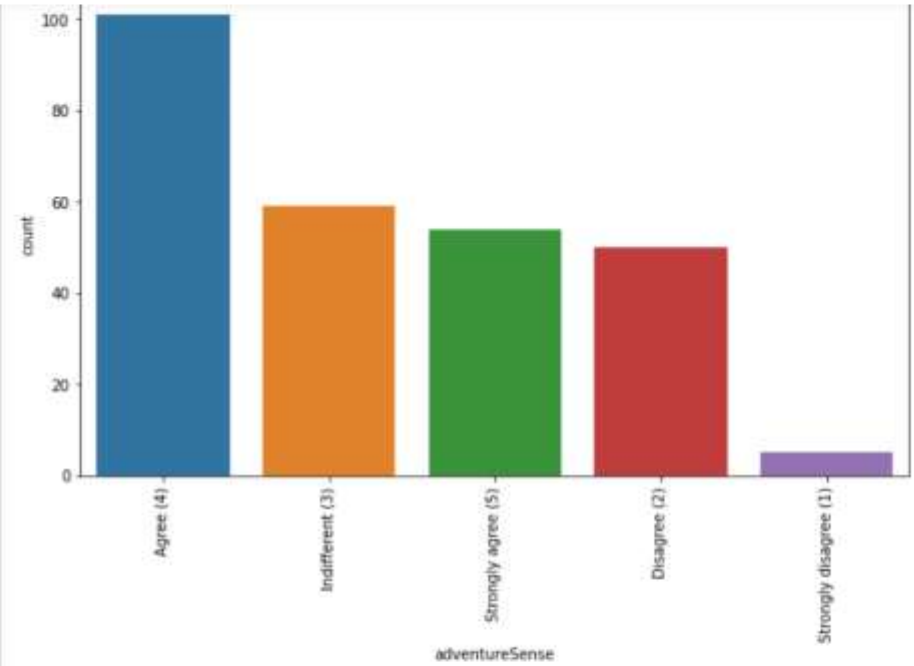

```
Strongly agree (5)    50.0
Agree (4)             36.0
Indifferent (3)       12.0
Disagree (2)           2.0
Name: relevantProductInfo, dtype: float64 2
```



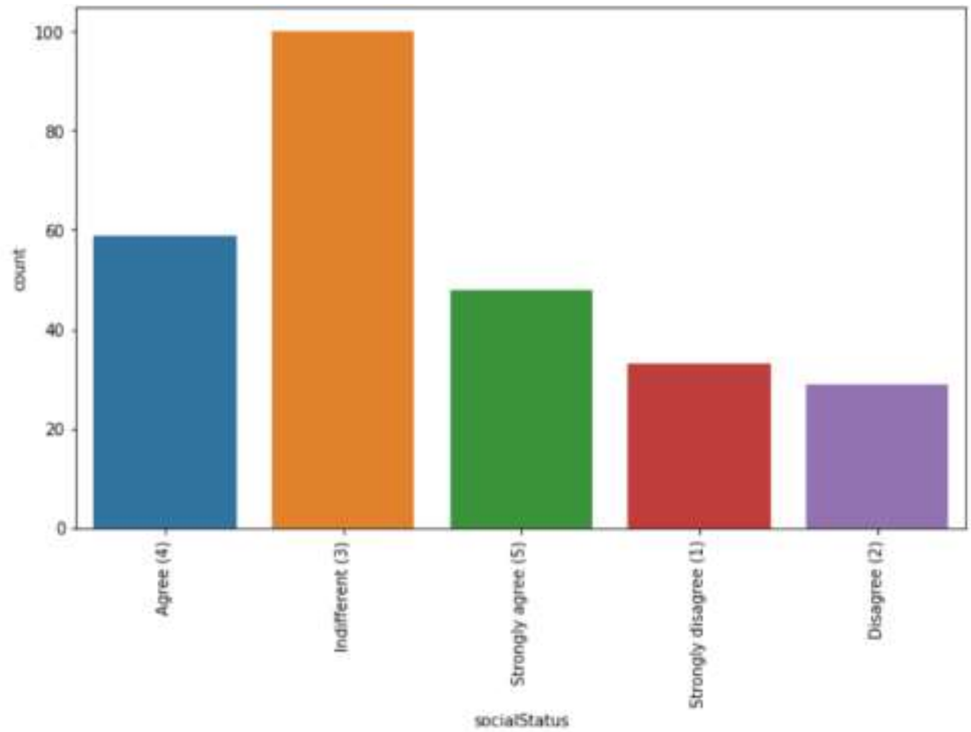

```
Strongly agree (5)    55.0
Agree (4)             28.0
Disagree (2)          12.0
Indifferent (3)        6.0
Name: monetarySavings, dtype: float64 2
```



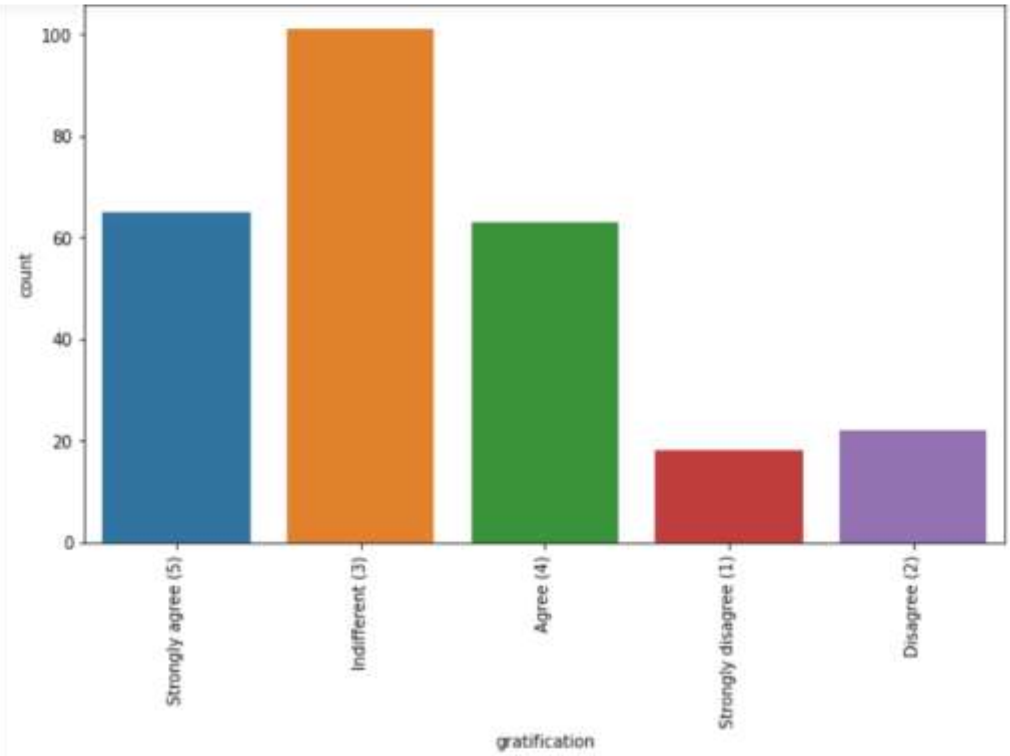

```
Agree (4)             51.0
Indifferent (3)       29.0
Strongly agree (5)    20.0
Name: patronizingConvenience, dtype: float64 2
```



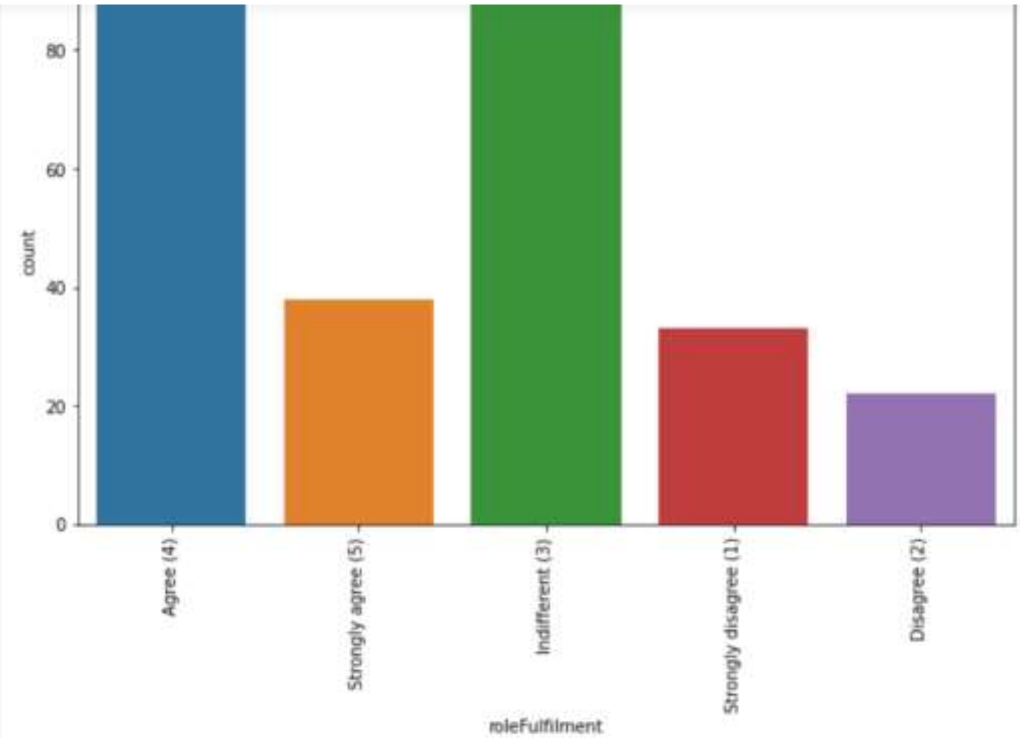

```
Agree (4)              38.0
Indifferent (3)        22.0
Strongly agree (5)     20.0
Disagree (2)           19.0
```



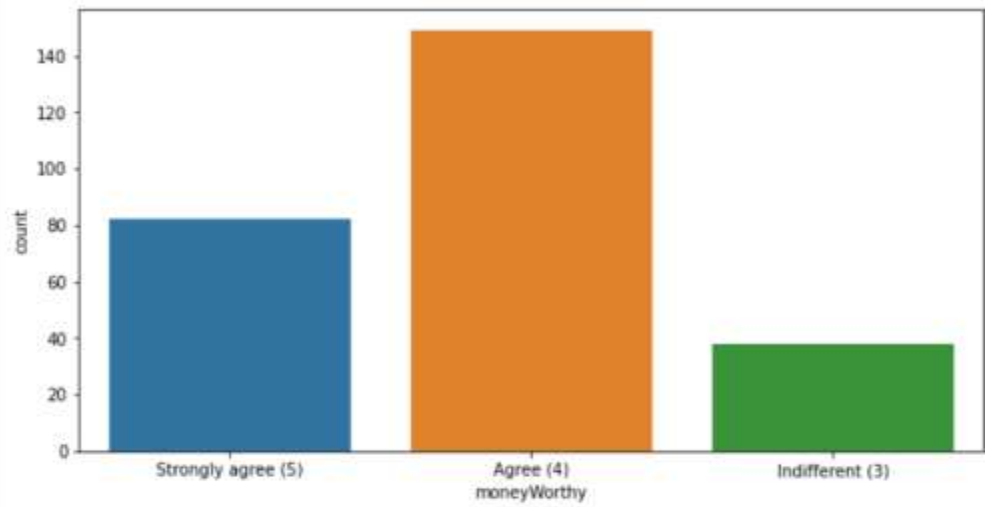

```
Indifferent (3)         37.0
Agree (4)               22.0
Strongly agree (5)      18.0
```



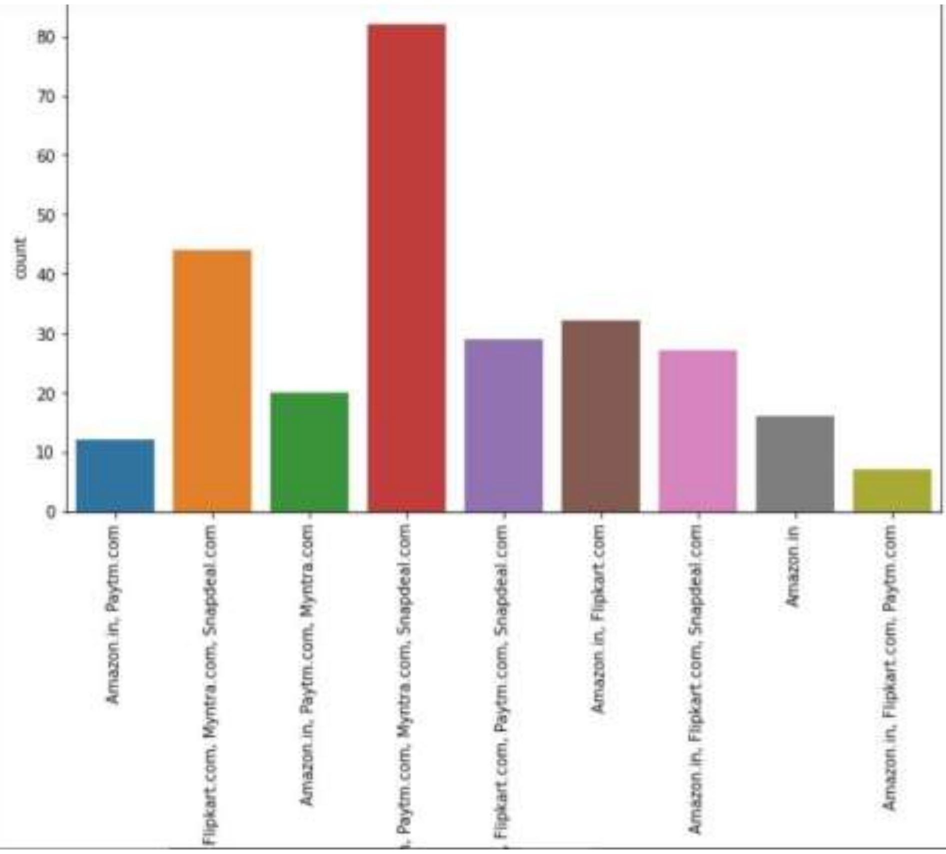

```
Indifferent (3)          38.0
Strongly agree (5)       24.0
Agree (4)                23.0
Disagree (2)              8.0
```



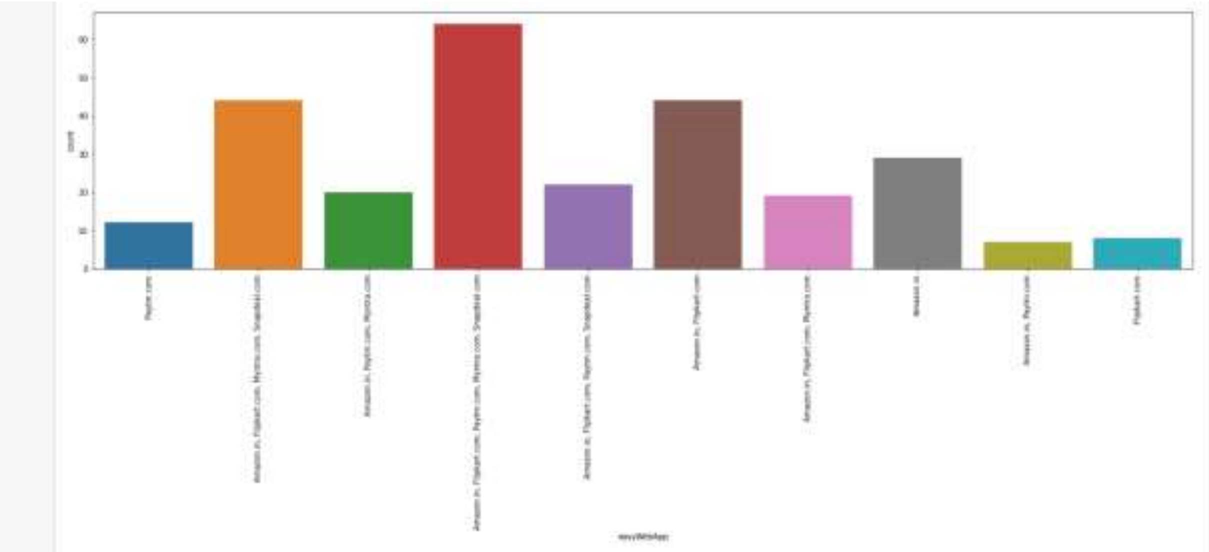

```
Agree (4)                33.0
Indifferent (3)          33.0
Strongly agree (5)       14.0
Strongly disagree (1)    12.0
Disagree (2)              8.0
```




```
Agree (4)             55.0
Strongly agree (5)    30.0
Indifferent (3)       14.0
Name: moneyWorthy, dtype: float64 2
```

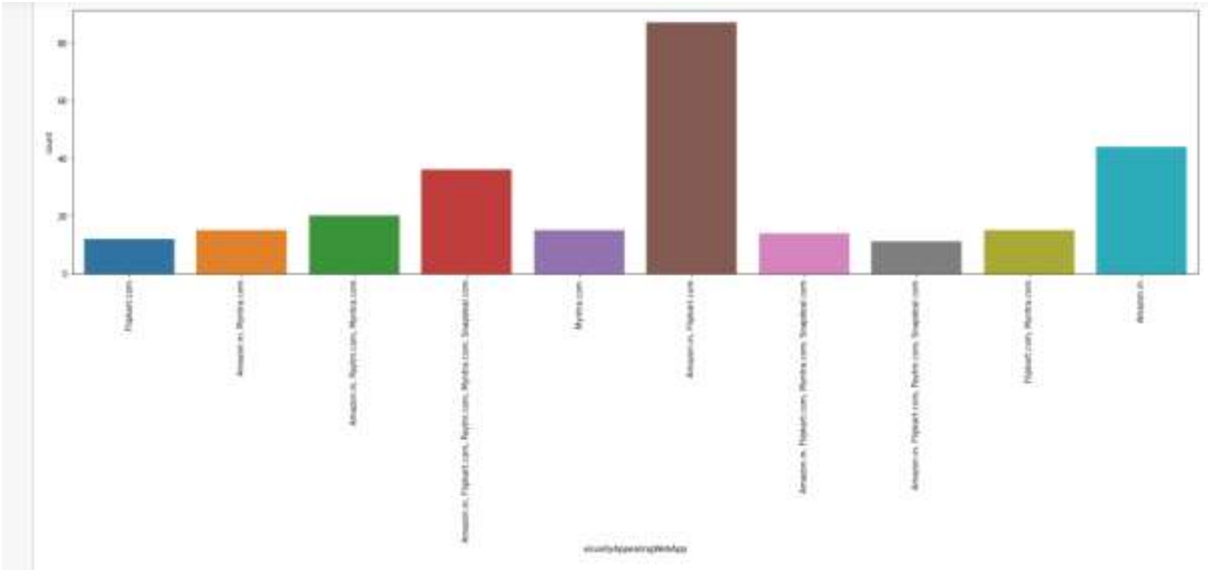


```
Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com    30.0
Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com               16.0
Amazon.in, Flipkart.com                                         12.0
Amazon.in, Flipkart.com, Paytm.com, Snapdeal.com                11.0
Amazon.in, Flipkart.com, Snapdeal.com                           10.0
Amazon.in, Paytm.com, Myntra.com                                 7.0
Amazon.in                                                        6.0
Amazon.in, Paytm.com                                             4.0
Amazon.in, Flipkart.com, Paytm.com                               3.0
Name: shoppedFrom, dtype: float64 2
```

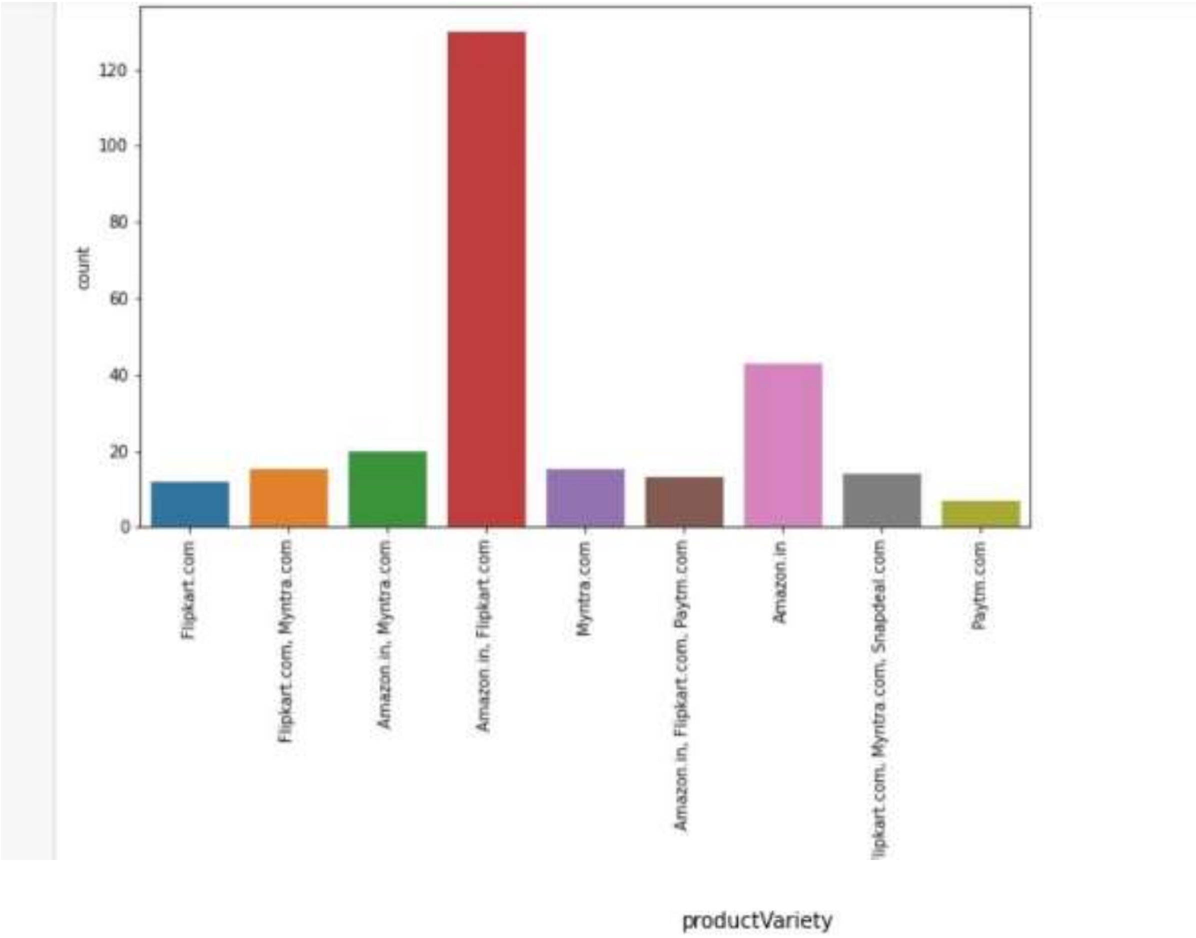


easyWebApp

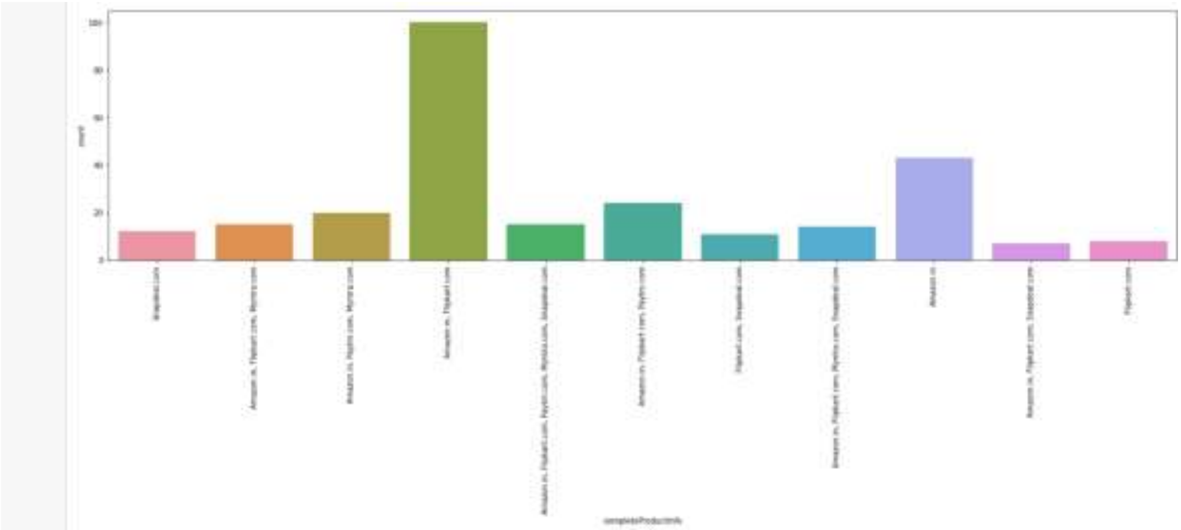
```
Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com    24.0
Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com               16.0
Amazon.in, Flipkart.com                                         16.0
Amazon.in                                                       11.0
Amazon.in, Flipkart.com, Paytm.com, Snapdeal.com                 8.0
Amazon.in, Paytm.com, Myntra.com                                 7.0
Amazon.in, Flipkart.com, Myntra.com                              7.0
Paytm.com                                                        4.0
Flipkart.com                                                     3.0
Amazon.in, Paytm.com                                             3.0
Name: easyWebApp, dtype: float64 2
```



```
                                                         visuallyAppealingWebApp
Amazon.in, Flipkart.com                                          32.0
Amazon.in                                                        16.0
Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com     13.0
Amazon.in, Paytm.com, Myntra.com                                  7.0
Amazon.in, Myntra.com                                             6.0
Myntra.com                                                        6.0
Flipkart.com, Myntra.com                                          6.0
Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com                 5.0
Flipkart.com                                                      4.0
Amazon.in, Flipkart.com, Paytm.com, Snapdeal.com                  4.0
Name: visuallyAppealingWebApp, dtype: float64 2
```

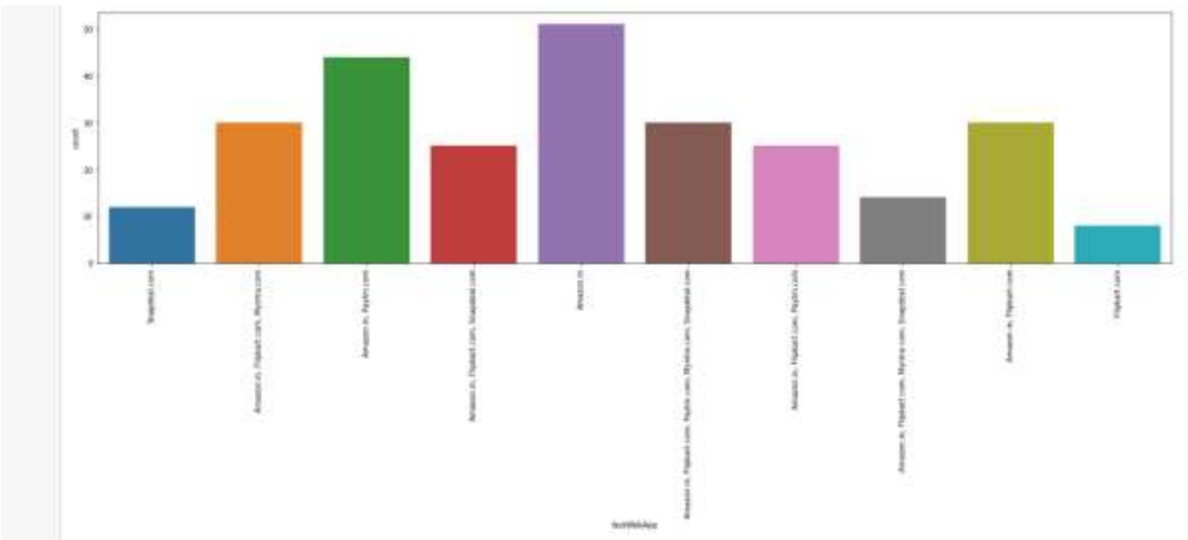



```
Amazon.in, Flipkart.com                              48.0
Amazon.in                                            16.0
Amazon.in, Myntra.com                                 7.0
Myntra.com                                            6.0
Flipkart.com, Myntra.com                              6.0
Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com     5.0
Amazon.in, Flipkart.com, Paytm.com                    5.0
Flipkart.com                                          4.0
Paytm.com                                             3.0
Name: productVariety, dtype: float64 2
```

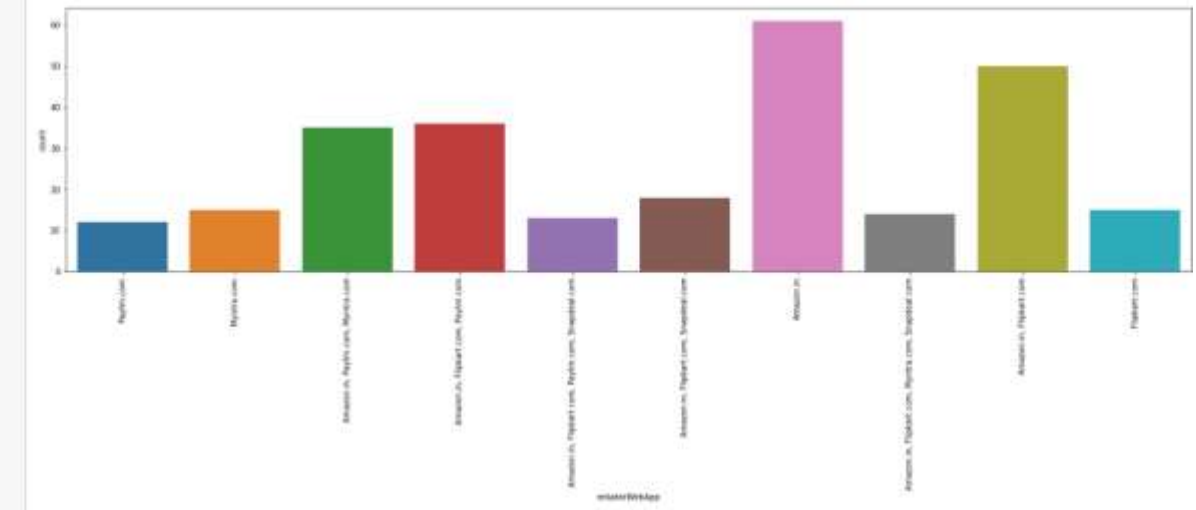


completeProductInfo

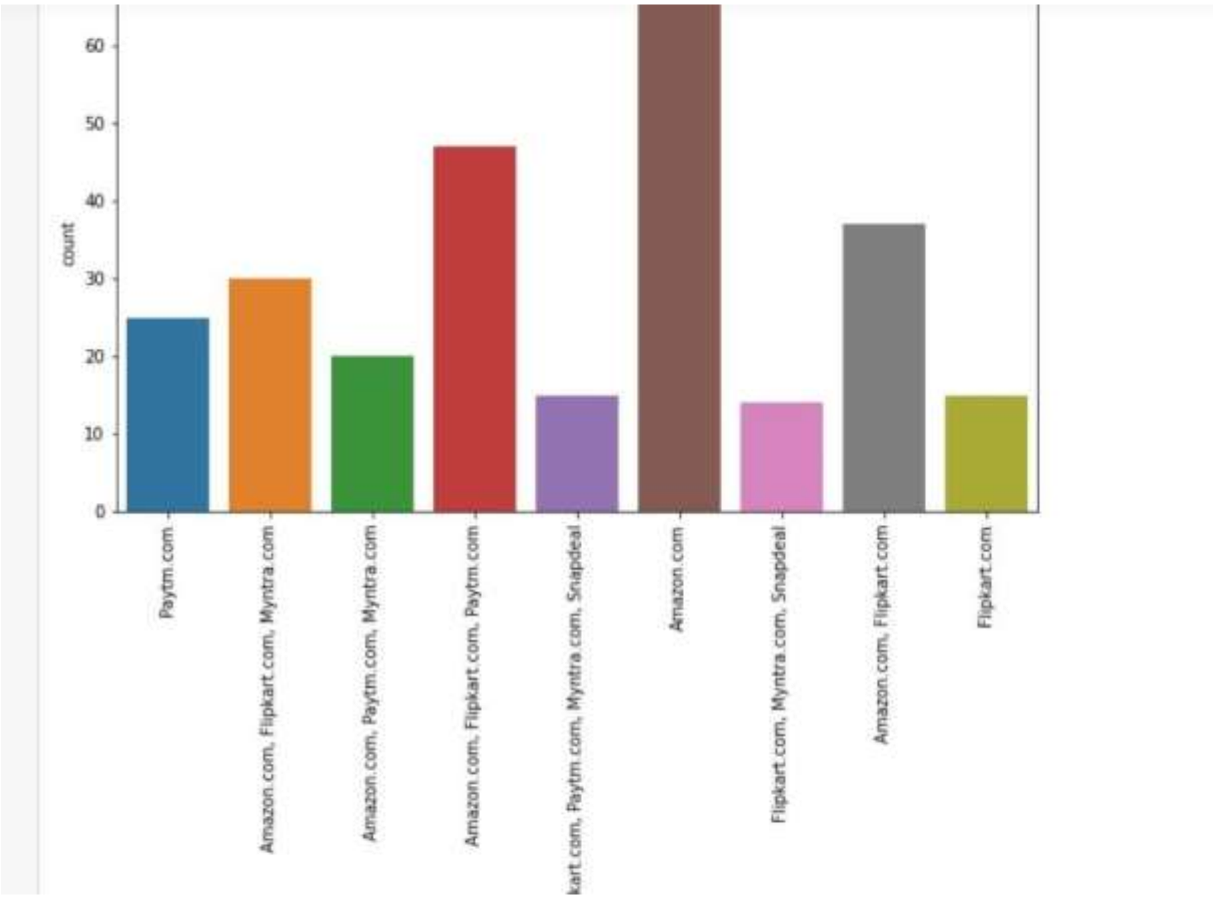
```
Amazon.in, Flipkart.com                                          37.0
Amazon.in                                                        16.0
Amazon.in, Flipkart.com, Paytm.com                                9.0
Amazon.in, Paytm.com, Myntra.com                                  7.0
Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com      6.0
Amazon.in, Flipkart.com, Myntra.com                               6.0
Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com                 5.0
Snapdeal.com                                                      4.0
Flipkart.com, Snapdeal.com                                        4.0
Flipkart.com                                                      3.0
Amazon.in, Flipkart.com, Snapdeal.com                             3.0
Name: completeProductInfo, dtype: float64 2
```



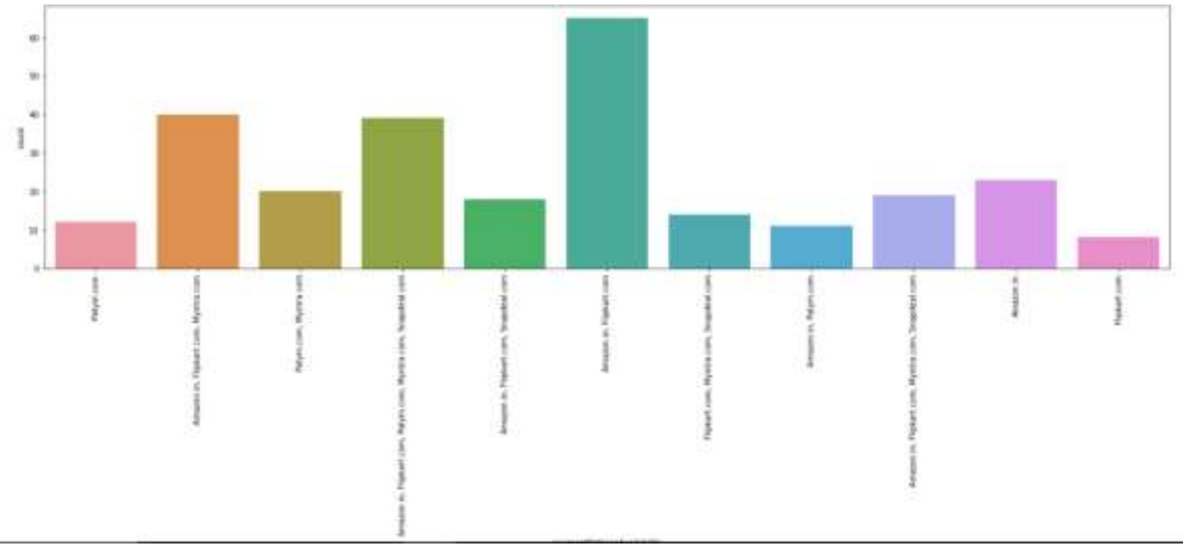
```
                                                         fastWebApp
Amazon.in                                                19.0
Amazon.in, Paytm.com                                     16.0
Amazon.in, Flipkart.com                                  11.0
Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com   11.0
Amazon.in, Flipkart.com, Myntra.com                      11.0
Amazon.in, Flipkart.com, Snapdeal.com                     9.0
Amazon.in, Flipkart.com, Paytm.com                        9.0
Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com         5.0
Snapdeal.com                                              4.0
Flipkart.com                                              3.0
Name: fastWebApp, dtype: float64 2
```



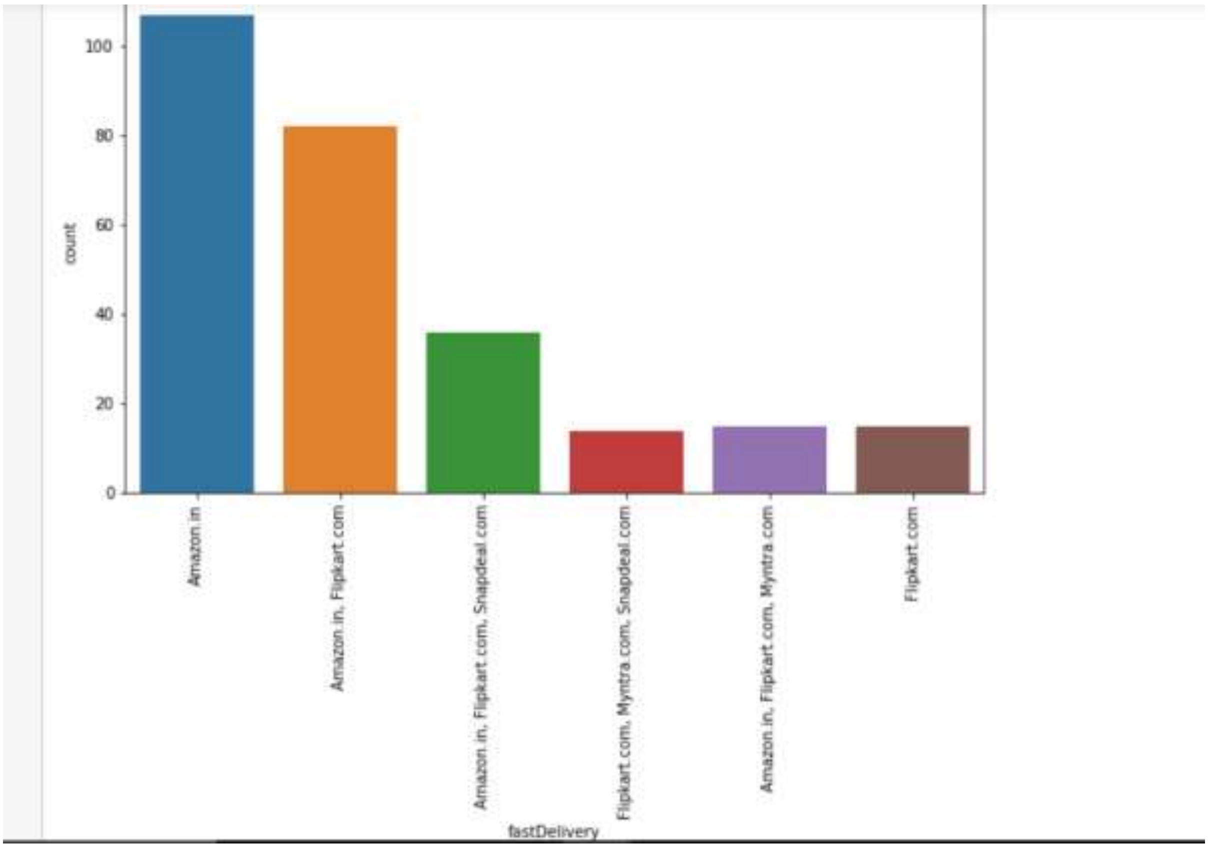
```
Amazon.in                                                23.0
Amazon.in, Flipkart.com                                  19.0
Amazon.in, Flipkart.com, Paytm.com                       13.0
Amazon.in, Paytm.com, Myntra.com                         13.0
Amazon.in, Flipkart.com, Snapdeal.com                     7.0
Myntra.com                                                6.0
Flipkart.com                                              6.0
Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com         5.0
Amazon.in, Flipkart.com, Paytm.com, Snapdeal.com          5.0
Paytm.com                                                 4.0
Name: reliableWebApp, dtype: float64 2
```




```
Amazon.com                                                25.0
Amazon.com, Flipkart.com, Paytm.com                       17.0
Amazon.com, Flipkart.com                                  14.0
Amazon.com, Flipkart.com, Myntra.com                      11.0
Paytm.com                                                  9.0
Amazon.com, Paytm.com, Myntra.com                          7.0
Amazon.com, Flipkart.com, Paytm.com, Myntra.com, Snapdeal  6.0
Flipkart.com                                               6.0
Flipkart.com, Myntra.com, Snapdeal                         5.0
Name: quickPurchase, dtype: float64 2
```

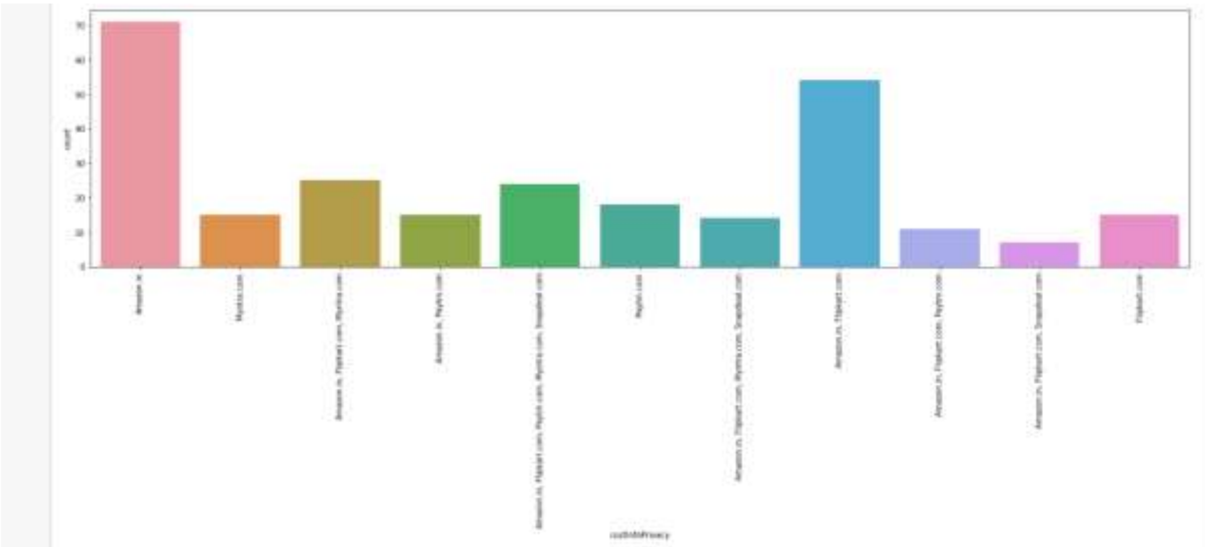


```
paymentOptionsAvailability

Amazon.in, Flipkart.com                                         24.0
Amazon.in, Flipkart.com, Myntra.com                             15.0
Amazon.in, Flipkart.com, Patym.com, Myntra.com, Snapdeal.com    14.0
Amazon.in                                                        9.0
Patym.com, Myntra.com                                            7.0
Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com                7.0
Amazon.in, Flipkart.com, Snapdeal.com                            7.0
Flipkart.com, Myntra.com, Snapdeal.com                           5.0
Patym.com                                                        4.0
Amazon.in, Patym.com                                             4.0
Flipkart.com                                                     3.0
Name: paymentOptionsAvailability, dtype: float64 2
```

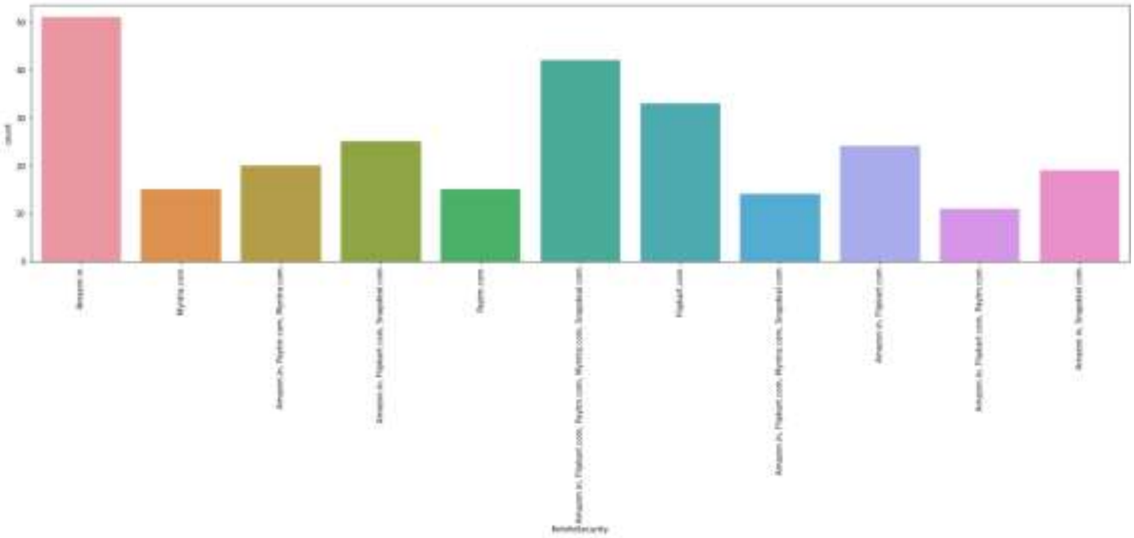


fastDelivery

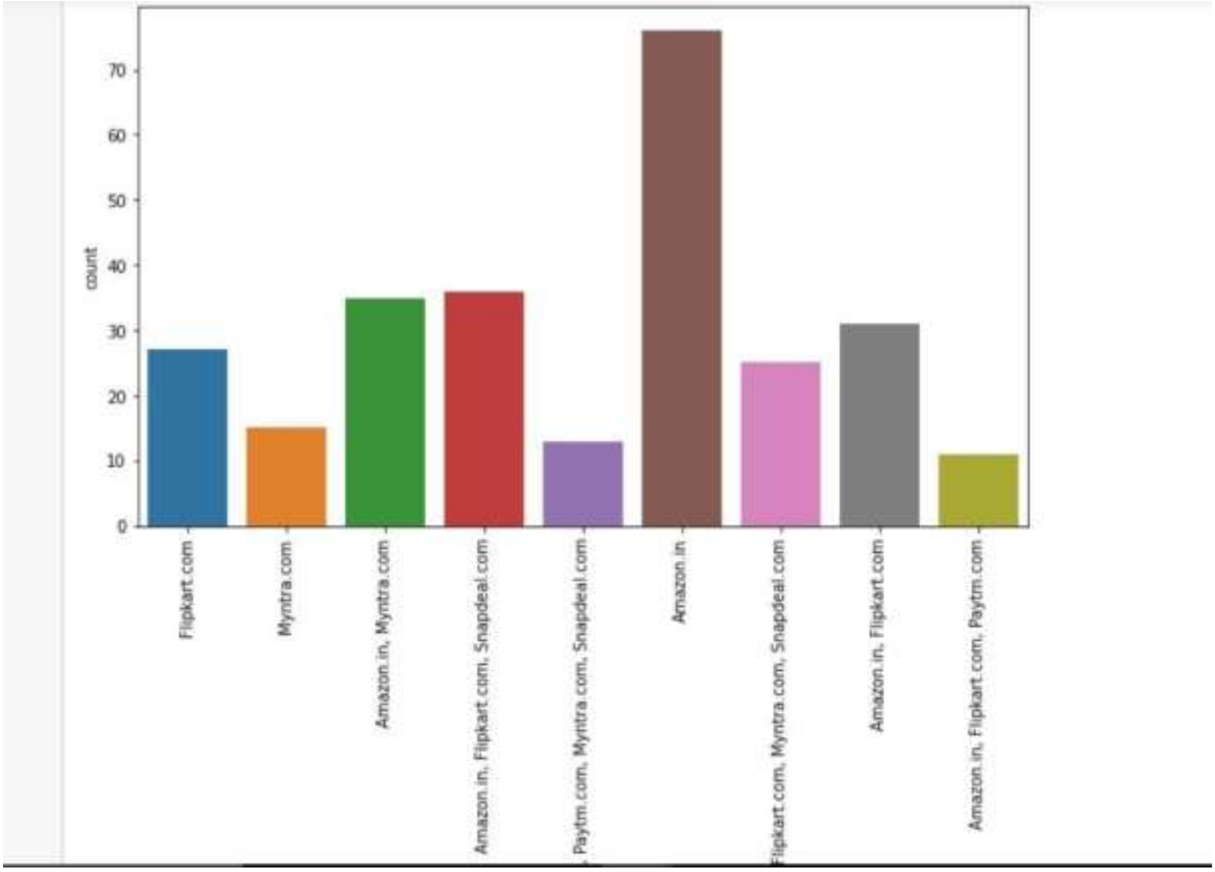
```
Amazon.in                                  40.0
Amazon.in, Flipkart.com                    30.0
Amazon.in, Flipkart.com, Snapdeal.com      13.0
Flipkart.com                                6.0
Amazon.in, Flipkart.com, Myntra.com         6.0
Flipkart.com, Myntra.com, Snapdeal.com      5.0
Name: fastDelivery, dtype: float64 2
```



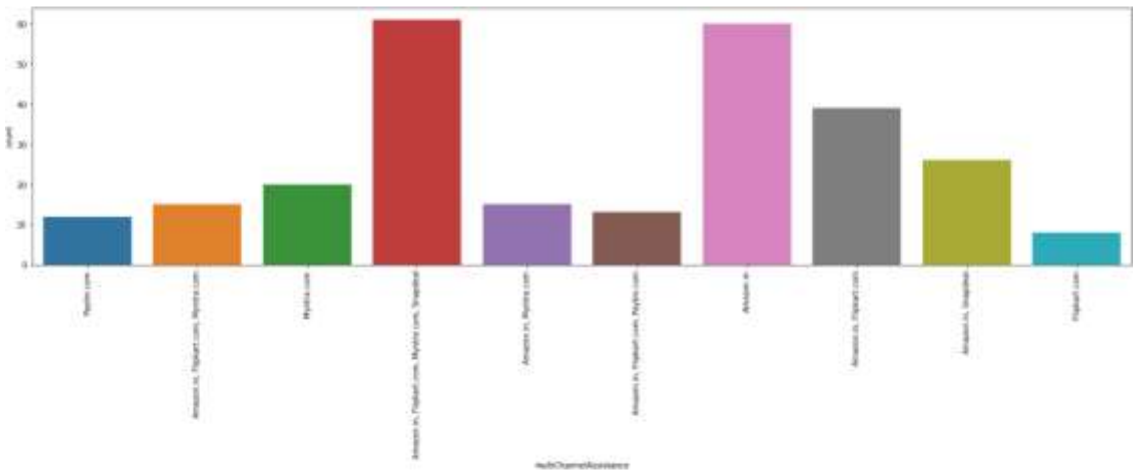
```
Amazon.in                                                   26.0
Amazon.in, Flipkart.com                                     20.0
Amazon.in, Flipkart.com, Myntra.com                          9.0
Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com 9.0
Paytm.com                                                    7.0
Myntra.com                                                   6.0
Flipkart.com                                                 6.0
Amazon.in, Paytm.com                                         6.0
Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com            5.0
Amazon.in, Flipkart.com, Paytm.com                           4.0
Amazon.in, Flipkart.com, Snapdeal.com                        3.0
Name: custInfoPrivacy, dtype: float64 2
```



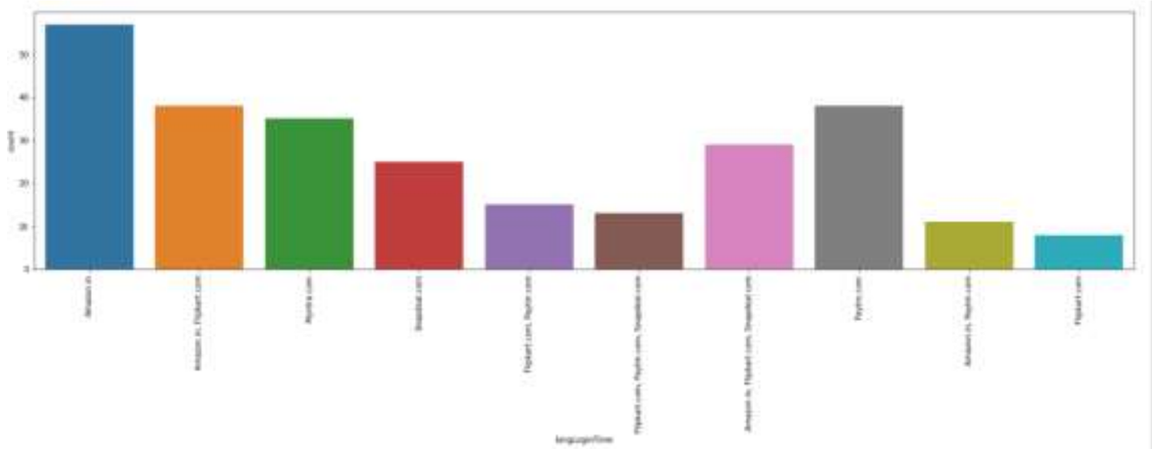
```
Amazon.in                                                   19.0
Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com 16.0
Flipkart.com                                                12.0
Amazon.in, Flipkart.com, Snapdeal.com                        9.0
Amazon.in, Flipkart.com                                      9.0
Amazon.in, Paytm.com, Myntra.com                             7.0
Amazon.in, Snapdeal.com                                      7.0
Myntra.com                                                   6.0
Paytm.com                                                    6.0
Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com            5.0
Amazon.in, Flipkart.com, Paytm.com                           4.0
Name: finInfoSecurity, dtype: float64 2
```



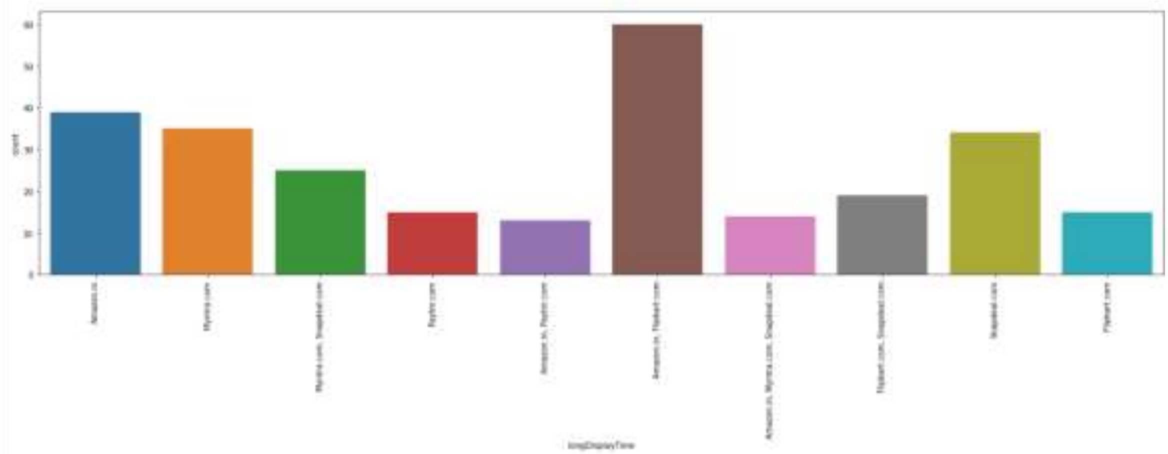

```
Amazon.in                                                    28.0
Amazon.in, Flipkart.com, Snapdeal.com                        13.0
Amazon.in, Myntra.com                                        13.0
Amazon.in, Flipkart.com                                      12.0
Flipkart.com                                                 10.0
Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com             9.0
Myntra.com                                                    6.0
Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com  5.0
Amazon.in, Flipkart.com, Paytm.com                            4.0
Name: perceivedTrustworthiness, dtype: float64 2
```



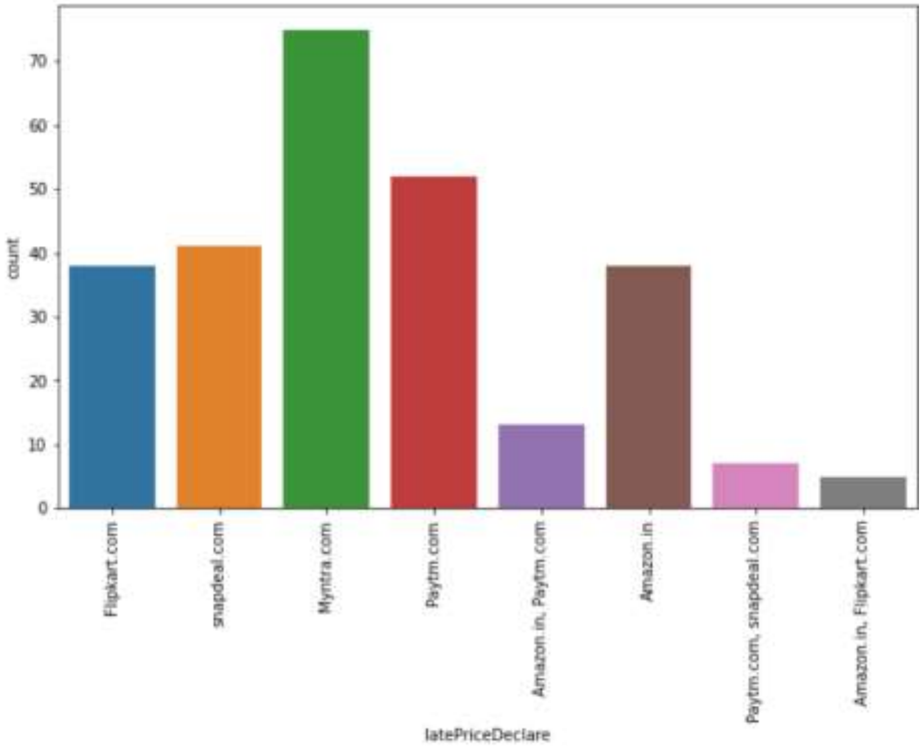
```
Amazon.in, Flipkart.com, Myntra.com, Snapdeal    23.0
Amazon.in                                        22.0
Amazon.in, Flipkart.com                          14.0
Amazon.in, Snapdeal                              10.0
Myntra.com                                        7.0
Amazon.in, Myntra.com                             6.0
Amazon.in, Flipkart.com, Myntra.com               6.0
Amazon.in, Flipkart.com, Paytm.com                5.0
Paytm.com                                         4.0
Flipkart.com                                      3.0
Name: multiChannelAssistance, dtype: float64 2
```



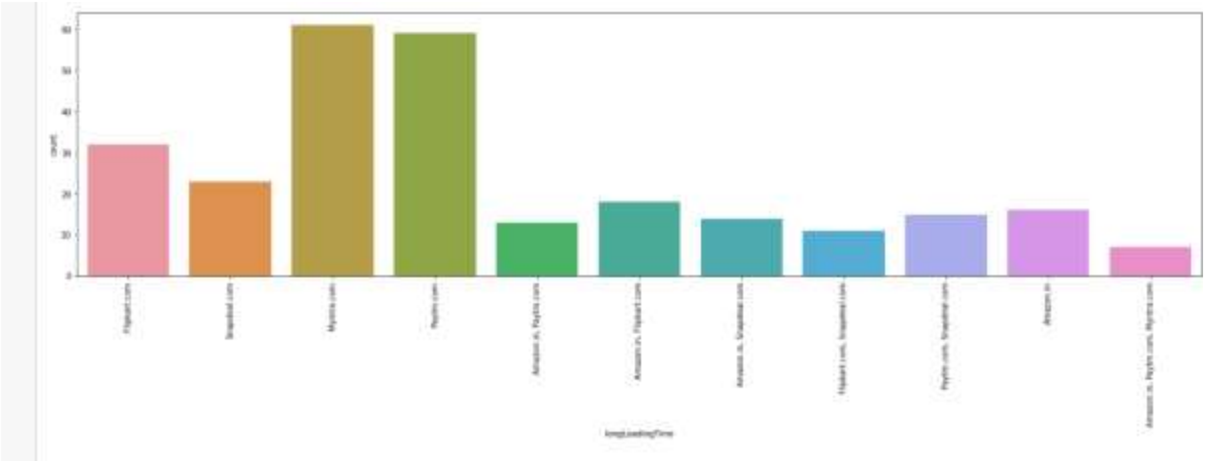
```
Amazon.in                                 21.0
Amazon.in, Flipkart.com                   14.0
Paytm.com                                 14.0
Myntra.com                                13.0
Amazon.in, Flipkart.com, Snapdeal.com     11.0
Snapdeal.com                               9.0
Flipkart.com, Paytm.com                    6.0
Flipkart.com, Paytm.com, Snapdeal.com      5.0
Amazon.in, Paytm.com                       4.0
Flipkart.com                               3.0
Name: longLoginTime, dtype: float64 2
```



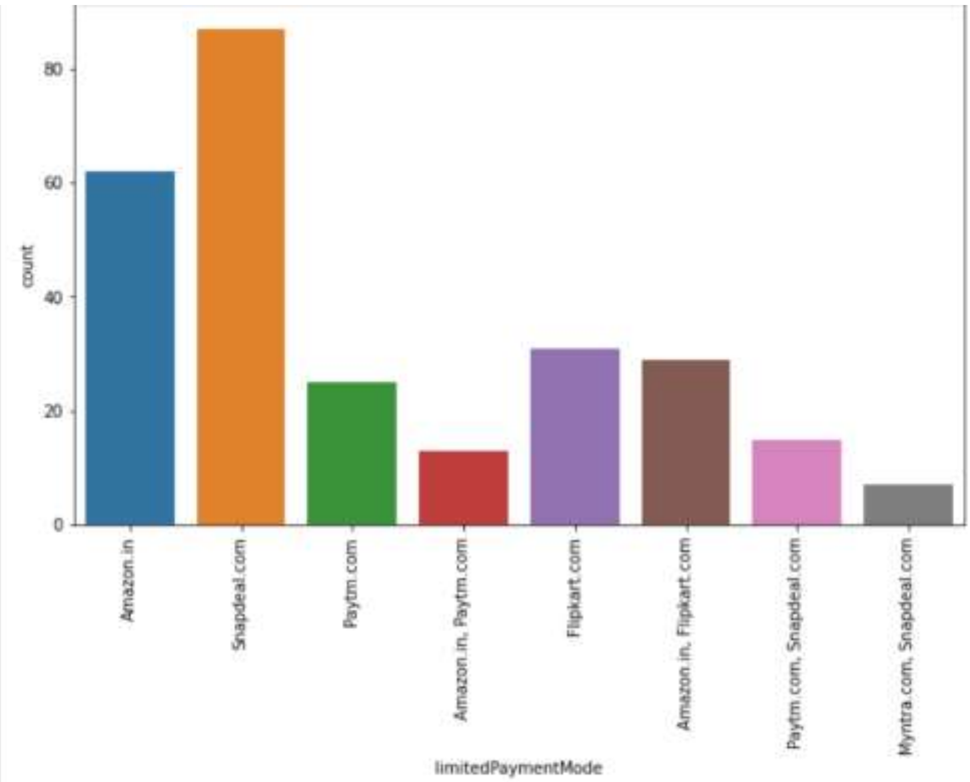
```
Amazon.in, Flipkart.com                   22.0
Amazon.in                                 14.0
Myntra.com                                13.0
Snapdeal.com                              13.0
Myntra.com, Snapdeal.com                   9.0
Flipkart.com, Snapdeal.com                 7.0
Paytm.com                                  6.0
Flipkart.com                               6.0
Amazon.in, Myntra.com, Snapdeal.com        5.0
Amazon.in, Paytm.com                       5.0
Name: longDisplayTime, dtype: float64 2
```



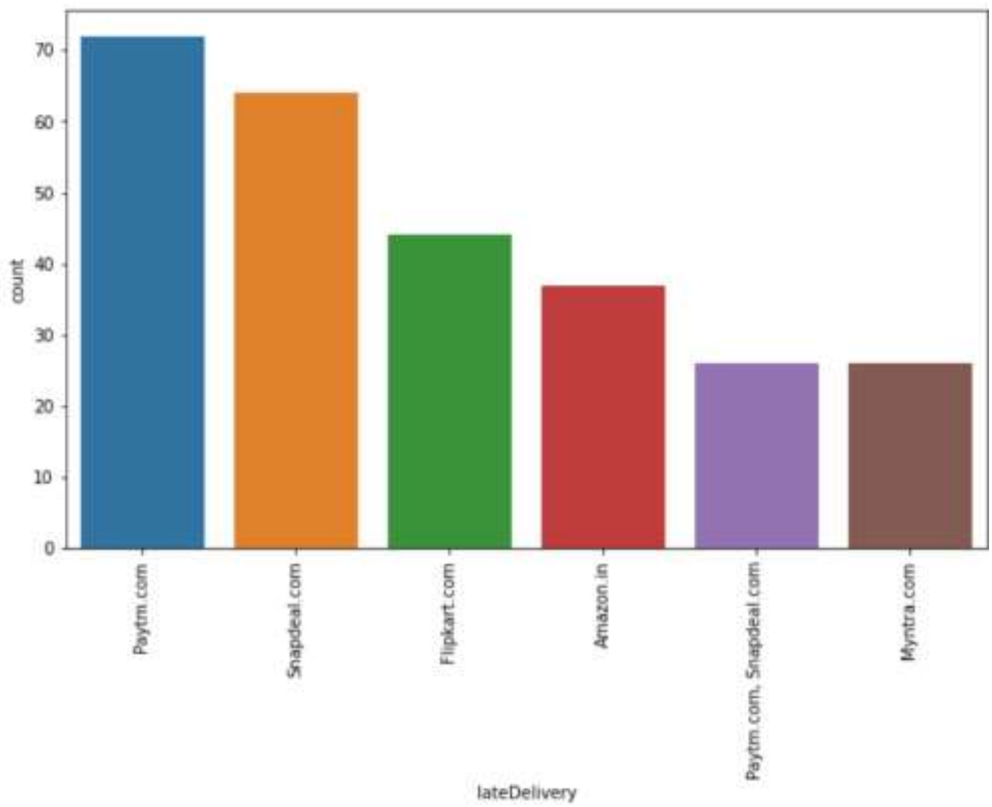

```
Myntra.com                   28.0
Paytm.com                    19.0
snapdeal.com                 15.0
Amazon.in                    14.0
Flipkart.com                 14.0
Amazon.in, Paytm.com          5.0
Paytm.com, snapdeal.com       3.0
Amazon.in, Flipkart.com       2.0
Name: latePriceDeclare, dtype: float64 2
```



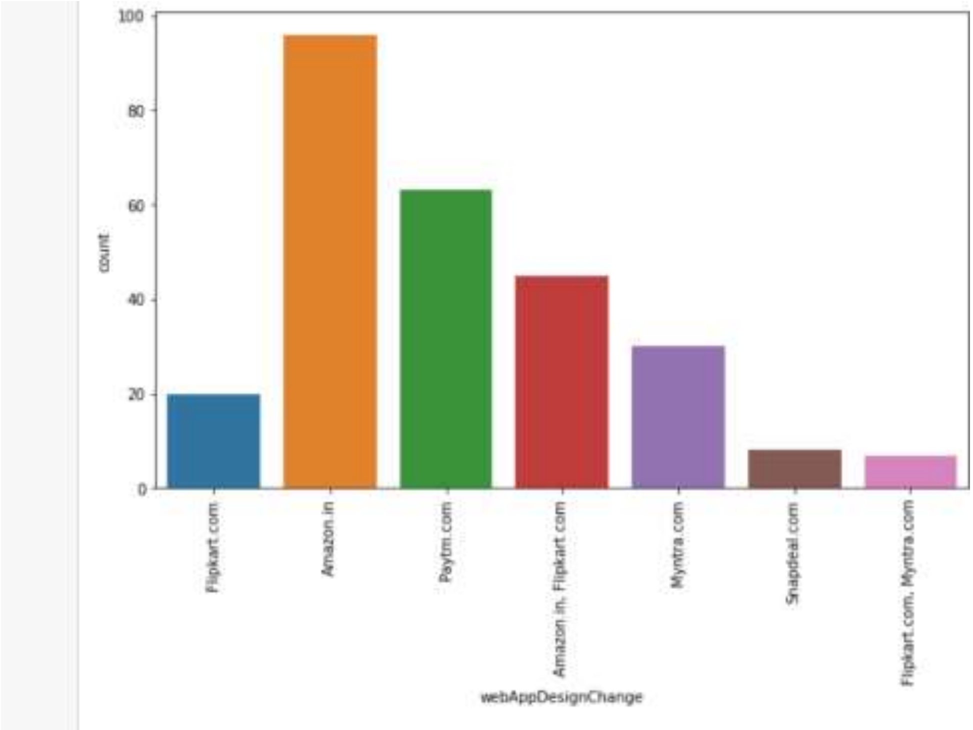
```
Myntra.com                           23.0
Paytm.com                            22.0
Flipkart.com                         12.0
Snapdeal.com                          9.0
Amazon.in, Flipkart.com               7.0
Amazon.in                             6.0
Paytm.com, Snapdeal.com               6.0
Amazon.in, Snapdeal.com               5.0
Amazon.in, Paytm.com                  5.0
Flipkart.com, Snapdeal.com            4.0
Amazon.in, Paytm.com, Myntra.com      3.0
Name: longLoadingTime, dtype: float64 2
```



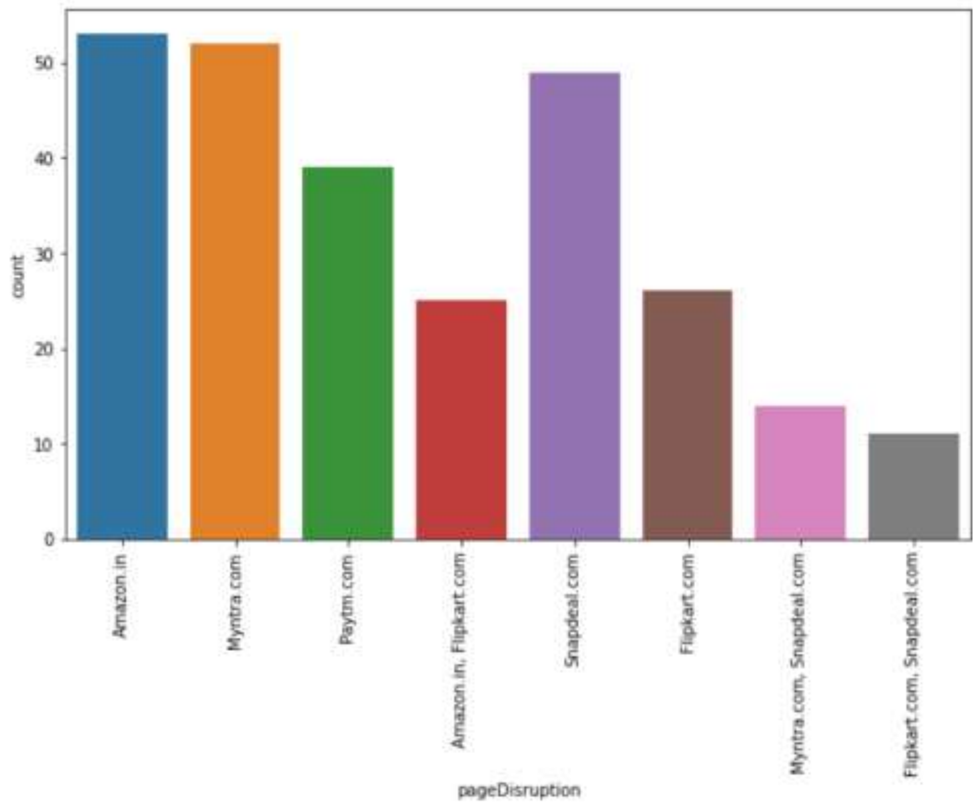
```
Snapdeal.com                 32.0
Amazon.in                    23.0
Flipkart.com                 12.0
Amazon.in, Flipkart.com      11.0
Paytm.com                     9.0
Paytm.com, Snapdeal.com       6.0
Amazon.in, Paytm.com          5.0
Myntra.com, Snapdeal.com      3.0
Name: limitedPaymentMode, dtype: float64 2
```



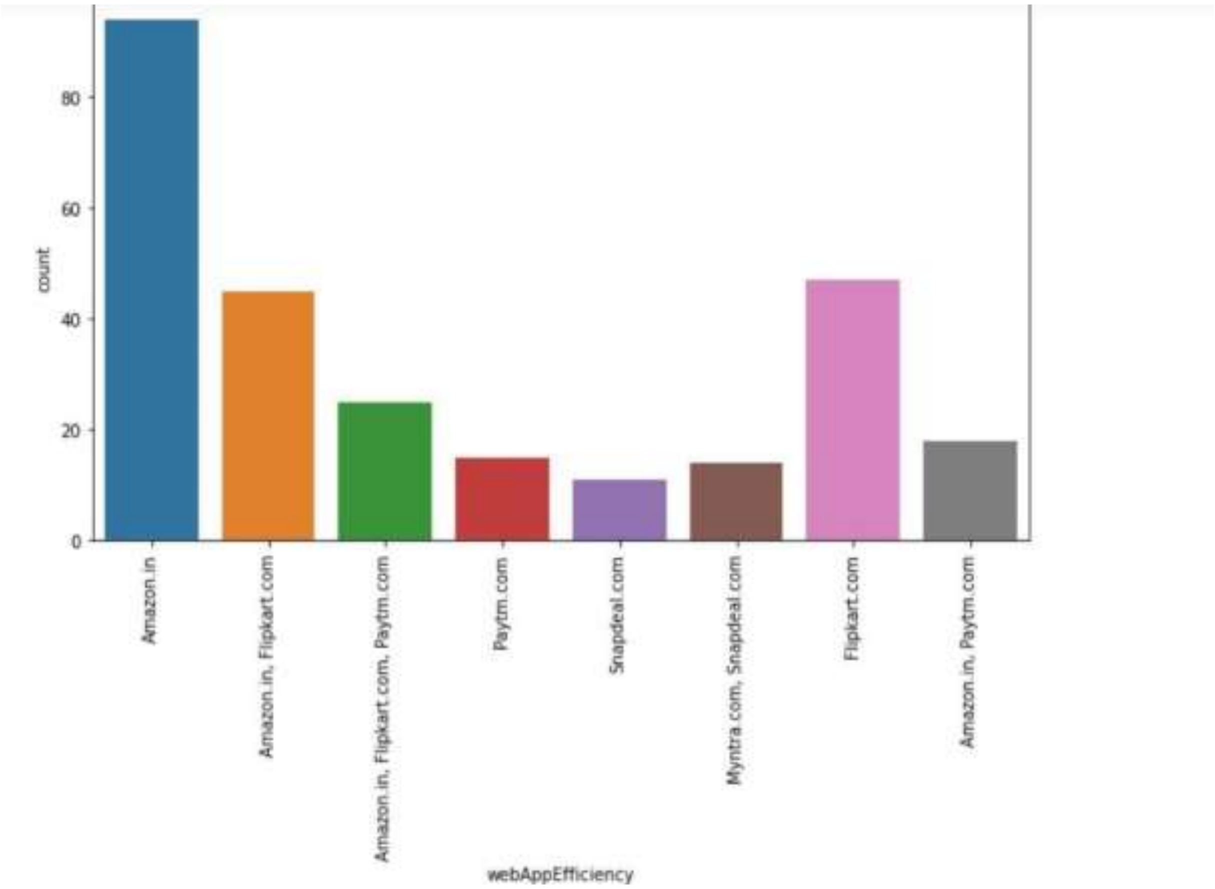

```
Paytm.com                  27.0
Snapdeal.com               24.0
Flipkart.com               16.0
Amazon.in                  14.0
Paytm.com, Snapdeal.com    10.0
Myntra.com                 10.0
Name: lateDelivery, dtype: float64 2
```



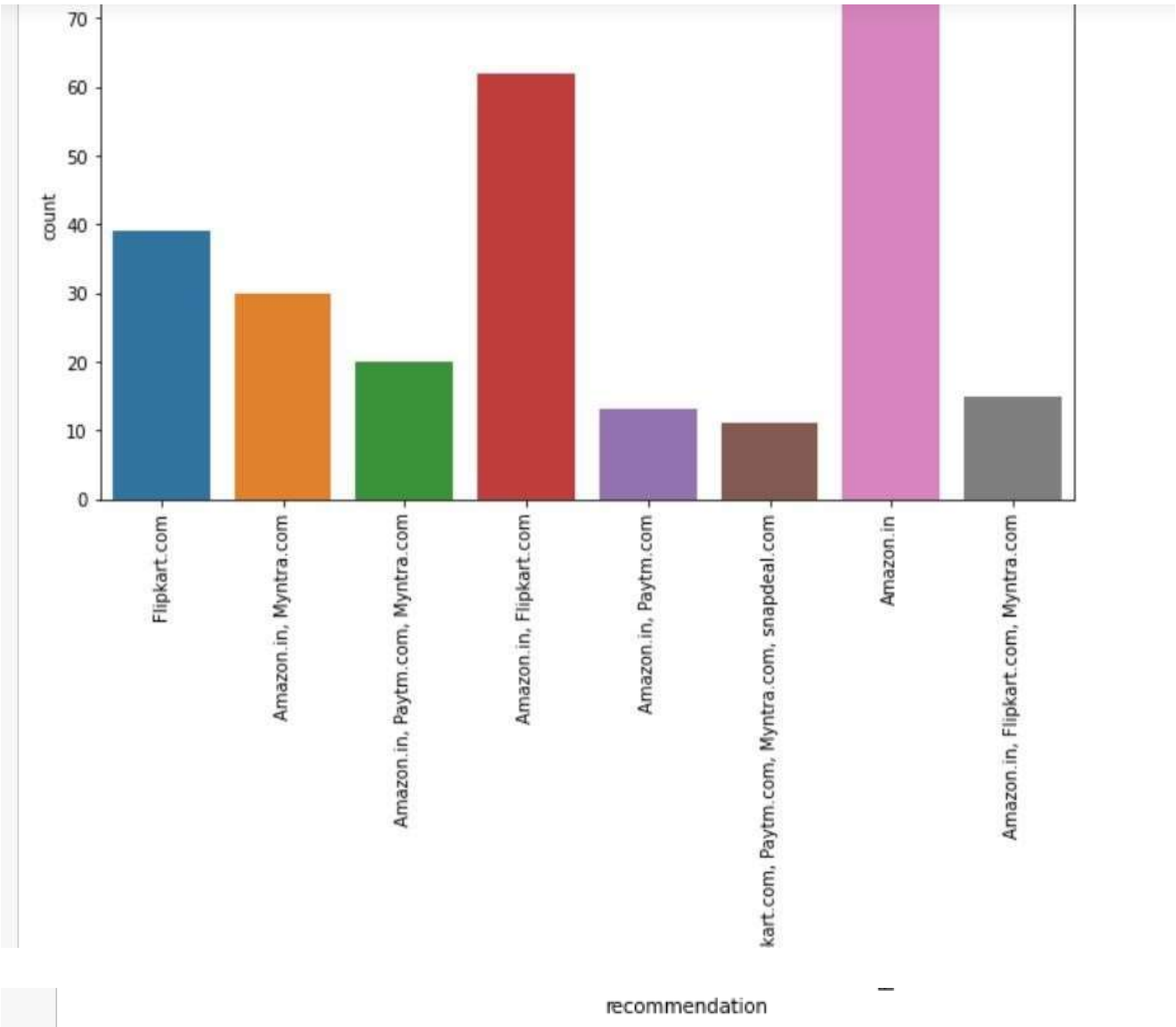

```
Amazon.in                  36.0
Paytm.com                  23.0
Amazon.in, Flipkart.com    17.0
Myntra.com                 11.0
Flipkart.com                7.0
Snapdeal.com                3.0
Flipkart.com, Myntra.com    3.0
Name: webAppDesignChange, dtype: float64 2
```




```
Amazon.in                    20.0
Myntra.com                   19.0
Snapdeal.com                 18.0
Paytm.com                    14.0
Flipkart.com                 10.0
Amazon.in, Flipkart.com       9.0
Myntra.com, Snapdeal.com      5.0
Flipkart.com, Snapdeal.com    4.0
Name: pageDisruption, dtype: float64 2
```




```
Amazon.in                              35.0
Flipkart.com                           17.0
Amazon.in, Flipkart.com                17.0
Amazon.in, Flipkart.com, Paytm.com      9.0
Amazon.in, Paytm.com                    7.0
Paytm.com                               6.0
Myntra.com, Snapdeal.com                5.0
Snapdeal.com                            4.0
Name: webAppEfficiency, dtype: float64 2
```




```
Amazon.in                                          29.0
Amazon.in, Flipkart.com                            23.0
Flipkart.com                                       14.0
Amazon.in, Myntra.com                              11.0
Amazon.in, Paytm.com, Myntra.com                    7.0
Amazon.in, Flipkart.com, Myntra.com                 6.0
Amazon.in, Paytm.com                                5.0
Flipkart.com, Paytm.com, Myntra.com, snapdeal.com   4.0
Name: recommendation, dtype: float64 2
```

OBSERVATIONS:

- Female customers (67%) are more than male customers
- Customers between 20 to 50 years shopped more from the online store. 75 % of customers are between 20 to 50
- Delhi, Bangalore, Noida are the cities with high number of buyers

- 36% customers were found shopping online for more than 4 years , 24% were shopping for 2-3 years
- IN last one year 42% customers have purchased online less than 10 times,23% customers purchased 31-40 times only 2 percent customers purchased more than 42 times
- 70% customers used mobile internet for online purchase
- 52% customers used Smartphone for online purchase, 32% used laptop least 4% used tablet
- 45% customers use Windows phone
- 80% customers used Google chrome for online purchase
- 86% customers uses Search engine to reach their favorite online store
- After first visit 32 % customers used search engine to reach online store and 32% uses application. We can assume that these customers have been retained by the E commerce website as they have downloaded the application on their phone and have a fair chance of shopping again. Also, there are a few customers who are using Desktop/Laptop, for which, there are no applications for online shopping and you have to stick with the web browser.
- 46% customers take at least 15 minutes before making a purchase, 26% take 6-10 minutes only
- 55% customers prefer Credit/Debit cards to make payments ,28% preferred COD and 17 % preferred Wallets
- 64% customers abandon their shopping cart sometimes, 19% abandon their cart frequently
- 49% customers abandon their bag due to better alternative offer, 20% abandon due to promo code not applicable
- 61% customers have strongly agreed to have easy website content which is easy to understand
- 43% customers agree that Information on similar product to the one highlighted is important for product comparison.
- 70% customers agree that Complete information on listed seller and product being offered is important for purchase decision
- 90% customers agree all relevant information on listed products must be stated clearly.
- 90% customers agreed that the website should be easily navigable

- 85% customers had no issues with the loading and processing speed
- 87% customers agree with user friendly website interface. The online stores should invest heavily in creating user friendly apps and websites, so that the customers do not have to work around much and the overall shopping experience is smooth.
- 59% customers strongly agree with the convenient paying methods. The reluctance to make online payments is gradually reducing. The acceptance has been built by addressing the trust issue over time by giving customers an option to pay online or to pay Cash On Delivery (COD)
- 52% customers trust that online store will fulfill its part of transaction at stipulated time
- 85% customers like the organization's readiness to assist with queries
- Being able to guarantee the privacy of the customer: This also got 69% strongly agree. Costumers are concerned about the unauthorized access to their data. Protecting user privacy will enable stores to drive more revenue and gain more customers
- Responsiveness, availability of several communication channels (email, online rep, twitter, phone etc.): 90% customers agreed to it. In case one channel is not available, customers can reach out to multiple channels which again is an important factor. Being able to communicate easily can make the difference in both their shopping experience as well as fulfill business goals of online store.
- 50% agree that online shopping gives monetary benefit and discounts to the customer: Most online shopping sites offer amazing round-the-year discounts. Banks and Digital Wallets have happily jumped on the online shopping bandwagon by providing Reward Points, and instant Cashback offers in addition to brand discounts, which will help to gain new customers
- 60% customers enjoys online shopping
- 84% customers agree that online shopping is convenient and flexible:
- 90% customers agree that return and replacement policy helps them making purchase decision. It is evident from the fact that people cannot actually try & touch the products, they are purchasing before it reaches home and they would want to return or replace in case of dissatisfaction. Online shopping

websites should make strategies around easy return and replacement policy if they want to retain their customers.

- Around 50% customers agree that gaining access to loyalty program is a benefit of shopping online
- Displaying quality Information on the website improves satisfaction of customers: 80% customer agreed to it. It means displaying quality information have a significant association with customer satisfaction, and repurchase intentions
- 95% customers are satisfied while shopping on a good quality website
- Net Benefit derived from shopping online can lead to users satisfaction:90% customers agreed
- User satisfaction cannot exist without trust: 85% customers agree that customer satisfaction cannot be built without trust. Trust is important factor to attract e-commerce buyers. So, it is important for companies to learn how to manage consumers' trust
- Offering a wide variety of listed product in several categories: 75% use agreed.
- 86% customers like to have complete and relevant information
- 80% customers agree to receive monetary savings while shopping online
- 50% agree with the convenience of patronizing the online retailer
- 50% customers agree that online shopping give the sense of adventure
- only 30% customers agree that online shopping enhances their social status
- 45% customers feel gratified while shopping with their favorite retailer
- 45% customers feel that shopping online helps them fulfill certain roles
- 85% customers agree they get value for their money while shopping online

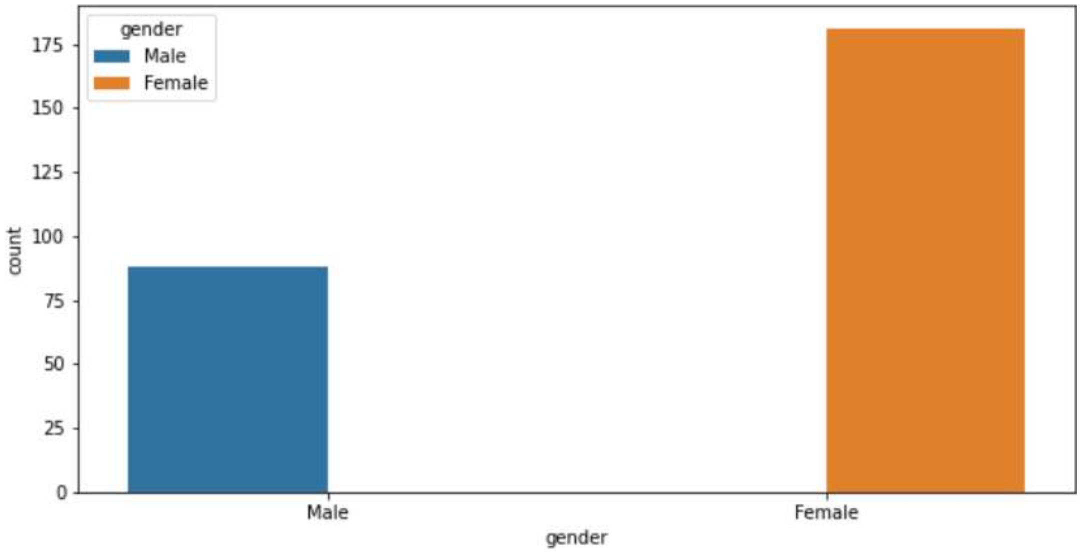
- Observations from Multiple Options based Questions i.e. related to company name specific parameters, after 47th rows
  - Maximum people have shopped from these 5 companies - Amazon.in, Flipkart.com, Paytm.com, Myntra.com, and Snapdeal.com.
  - 48% customers says flipkart, amazon shows wide variety of products
  - 37% customers like flipkart and amazon in terms of displaying complete and relevant information of the products

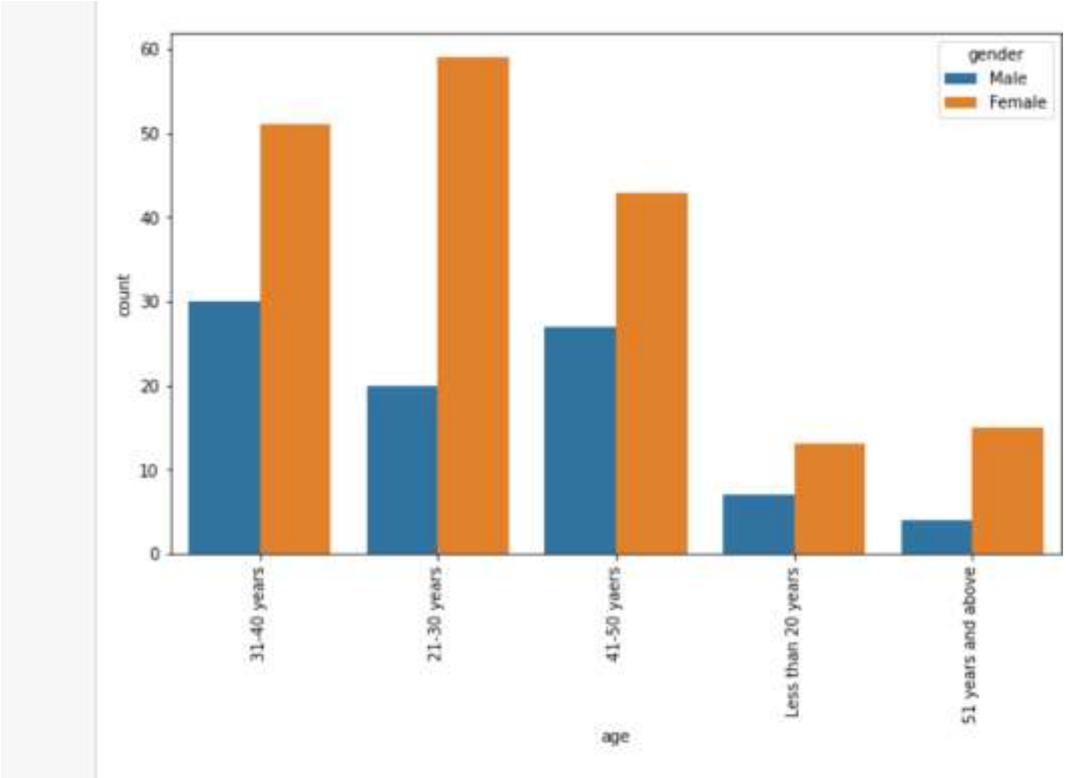
- In terms of speed Amazon.in is liked by 90% customers , 60 % like Flipkart , 27% Myntra
- 78% customers trust Amazon to be reliable , 55% likes flipkart , 25% Myntra
- 80% customers likes Amazon's quickness to complete the purchase , 60% likes Flipart's, 30% likes Myntra's
- 86% likes Amazon's delivery speed , 56% likes Flipkart's , 5% likes Myntra's
- 82% customers trust amazon in terms of keeping the privacy of their data, 56% trust flipkart 20% trust Myntra , 12 % trust Snapdeal
- 70% trust amazon in terms of keeping their financial information secured, 56% Flipakrt, 34% Myntra , 21% Snapdeal , 22@ Paytm
- 84% customers believe Amazon perceived trustworthiness, 43% - Flipkart , 33% - Myntra ,Snapdeal 27% Snapdeal
- 86% customers like Amazon in terms of online assistance through multi-channel , 51% -Flipkart ,42% Myntra
- 50% customers agree that Amazon takes longer time to log them in while in sales period/promotions , 39% goes with flipkart, 13% goes with Myntra , 16% Snapdeal
- 46% customers says that amazon takes longer time in displaying photos in sales/promotion , 35%- flipkart, 21% Snapdeal, 14% Myntra, 11% Paytm
- In terms of late declaration of price in promotion/sales 28% goes with Myntra, 21% amazon , 16% flipkart , 18% Snapdeal, 23% Paytm
- In terms of longer page loading in promotion/sales 26% customer goes with Myntra, 23% flipkart, 23% amazon, 24% Snapdeal
- In terms of limited mode of payment on most products during sales period/promotions 41% goes with Snapdeal, 40% -amazon, 23% flipkart, 20% Paytm and only 3% Myntra. Hence Myntra gives most payment options during sales/promotion.
- In terms of time taken in product delivery Paytm has highest votes of 37%, Snapdeal 24%, flipkart 16%, amazon 14% and Snapdeal and Myntra 10%. Hence Snapdeal and Myntra take minimum time for delivery among all.
- 20% customers dislikes disruptions while moving to another page on amazon, 24% on Myntra, 27% Snapdeal, 14% Paytm, 23% Snapdeal
- 61% customers says that amazon website is as efficient as before, 43% for flipkart, 22% for Paytm, 9% for Snapdeal

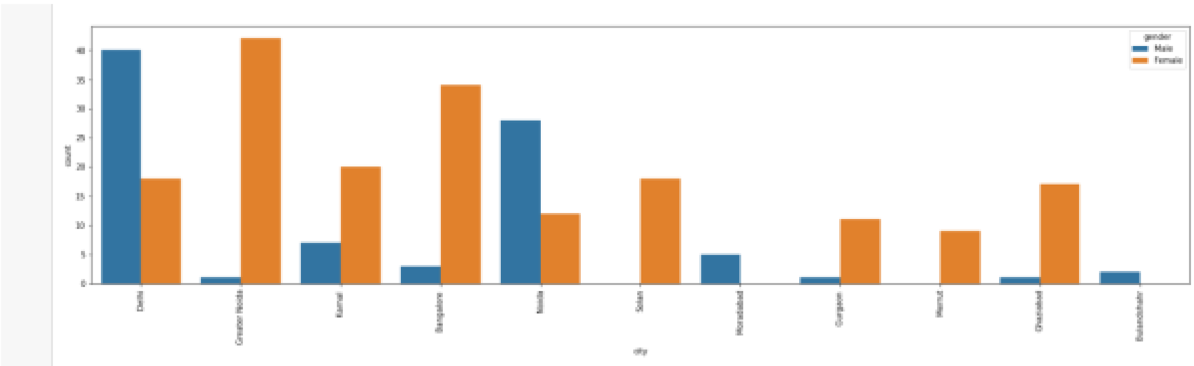
○ 81% customers would like to recommend amazon to a friend, 43% would like flipkart , 22% would like for Myntra , 9% for Paytm and least 4 % would like to recommend Snapdeal

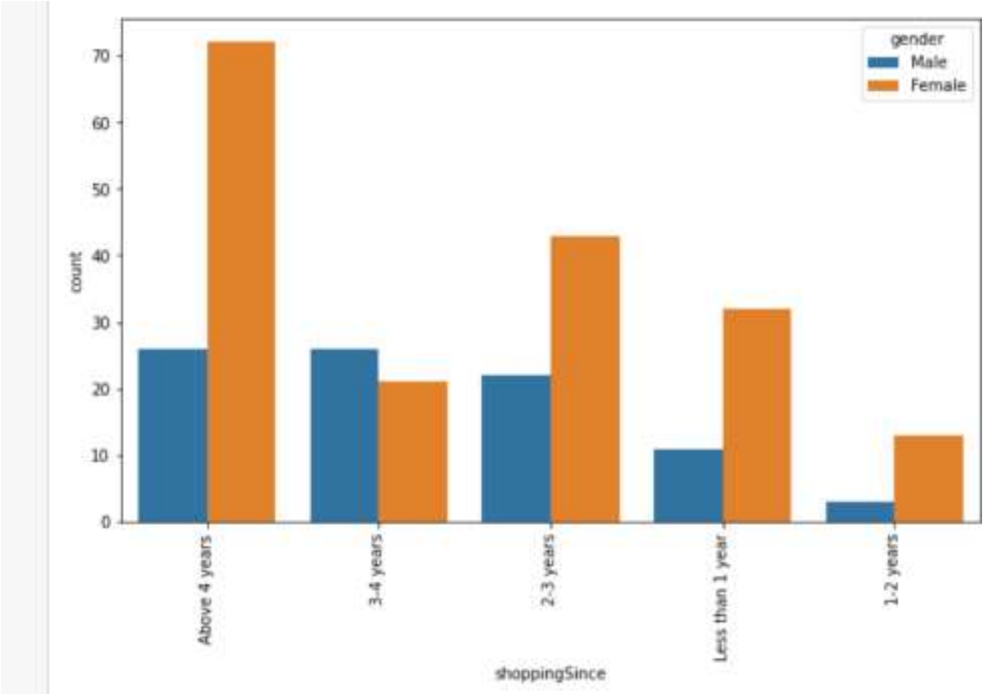
```
In [29]: def value_count(column):
             diff_count=len(data[column].value_counts())
             if diff_count<5:
                 plt.figure(figsize=(10,5))
             elif diff_count<10:
                 plt.figure(figsize=(10,6))
                 plt.xticks(rotation=90)
             elif diff_count<20:
                 plt.figure(figsize=(25,6))
                 plt.xticks(rotation=90)
             else:
                 plt.figure(figsize=(20,6))
                 plt.xticks(rotation=90)
             sns.countplot(x=column,hue='gender',data=data)
             plt.show()
```

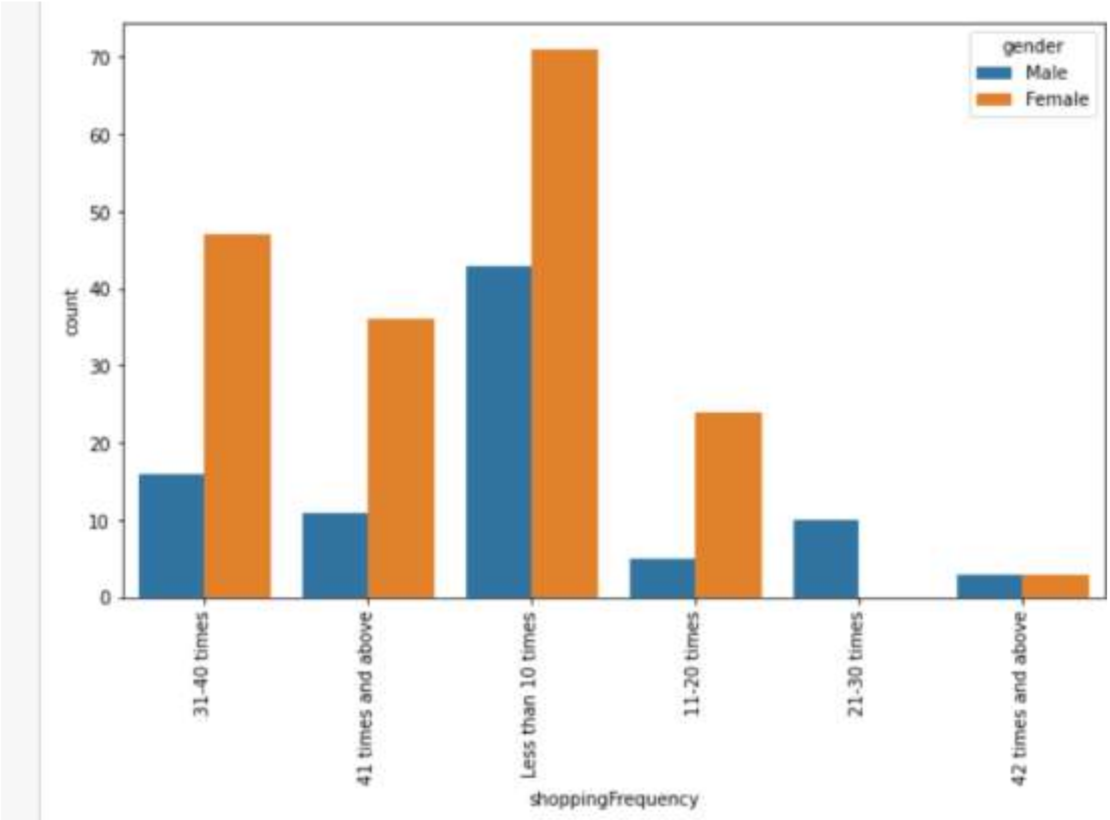
```
In [30]: data2=data.iloc[:,[0,1,2,4,5,13,16,18,26,31,32,34,37,38,46,47,70]]
         for i in data2:
             value_count(i)
```

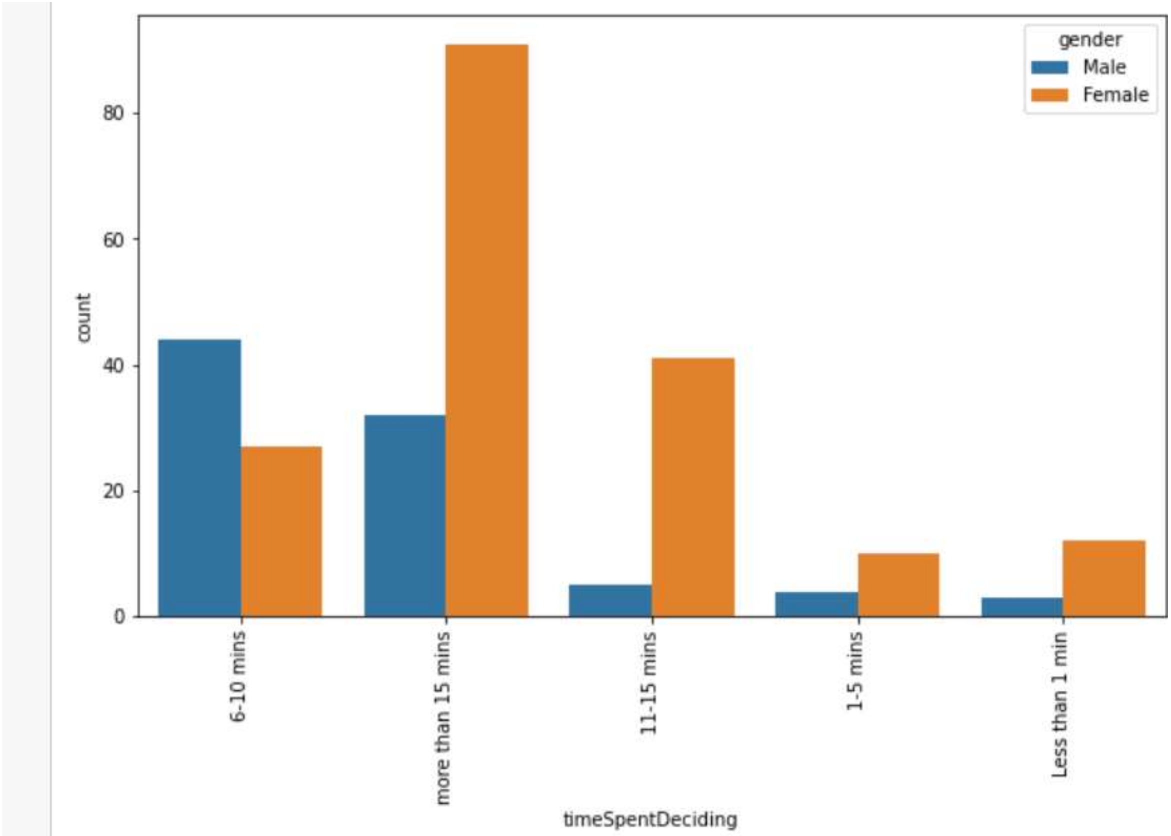


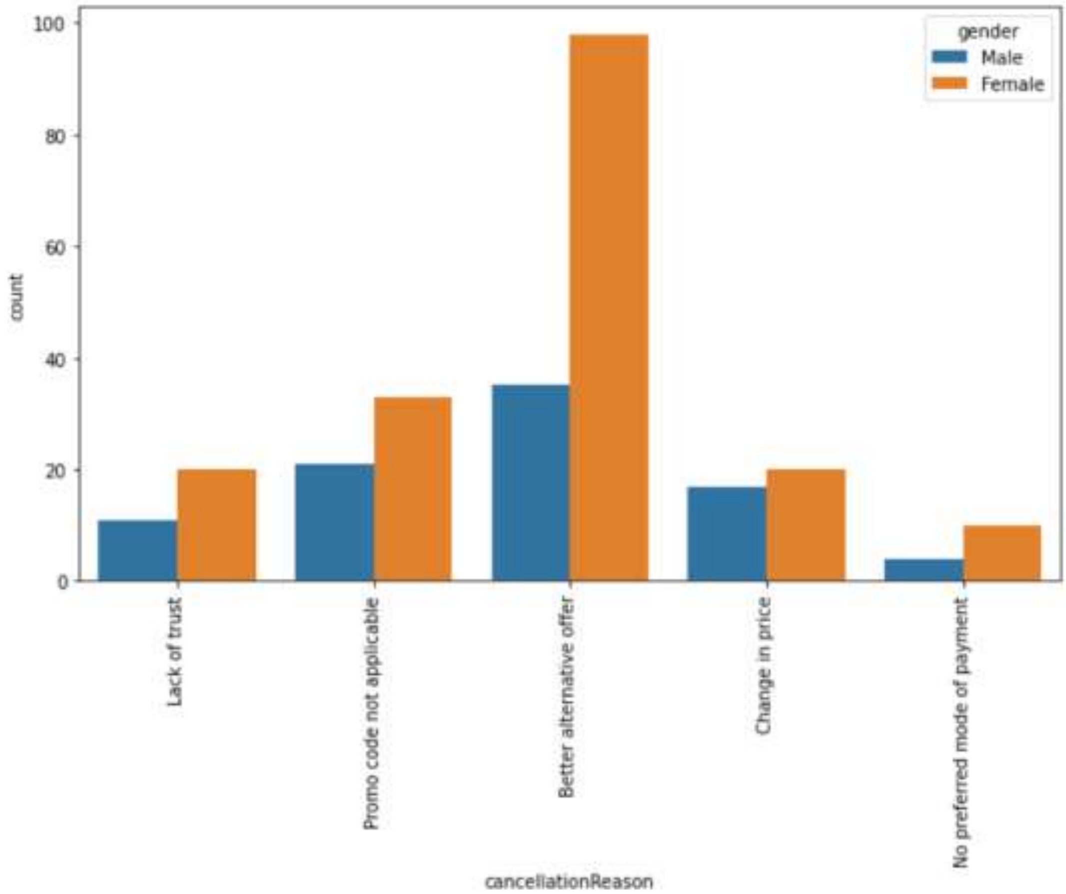

gender
Male
Female
count
60
50
40
30
20
10
0
31-40 years
21-30 years
41-50 yaers
Less than 20 years
51 years and above
age

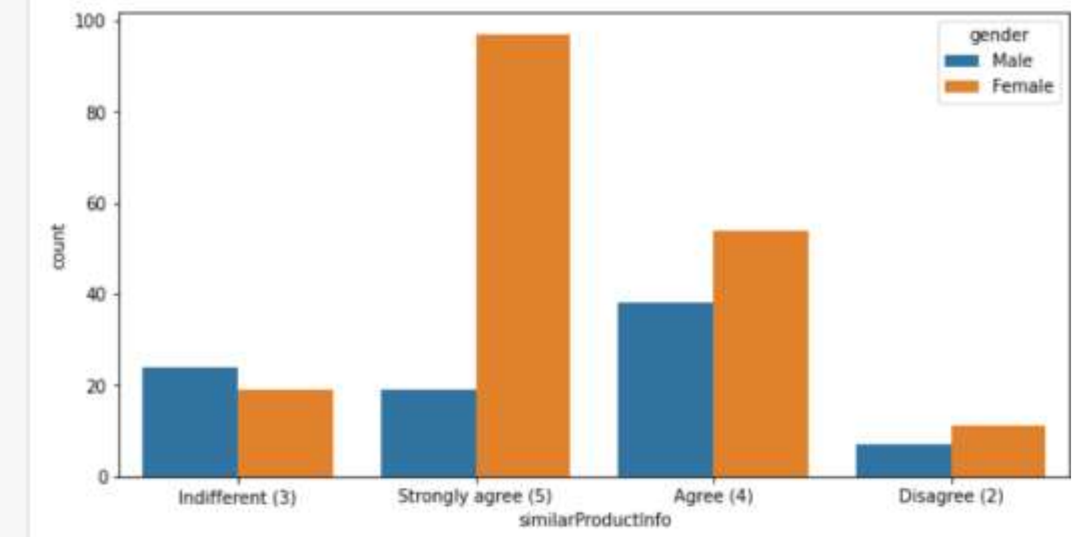


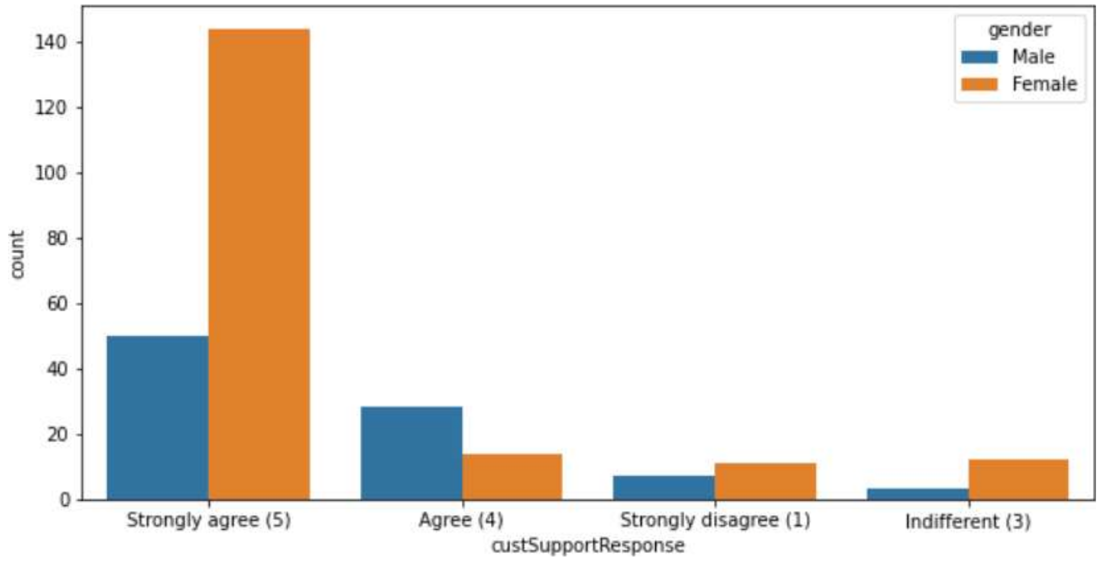

gender
Male
Female
count
Above 4 years
3-4 years
2-3 years
Less than 1 year
1-2 years
shoppingSince

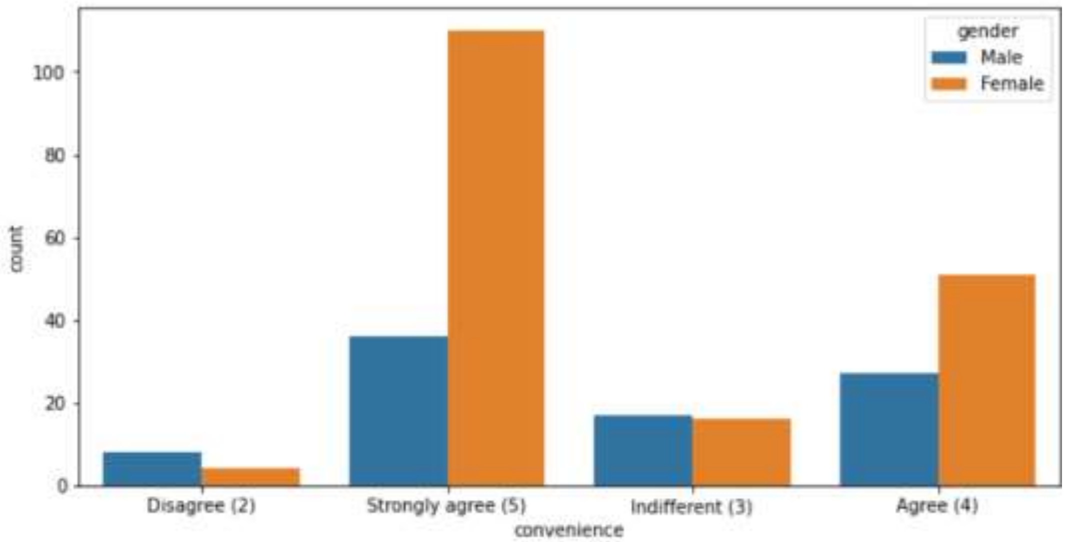

gender
Male
Female
count
31-40 times
41 times and above
Less than 10 times
11-20 times
21-30 times
42 times and above
shoppingFrequency

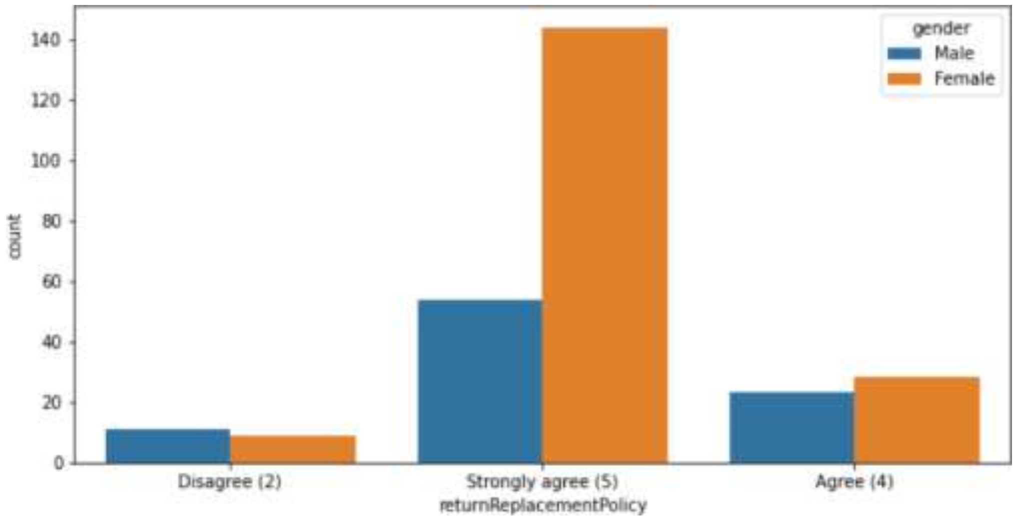

gender
Male
Female
count
80
60
40
20
0
6-10 mins
more than 15 mins
11-15 mins
1-5 mins
Less than 1 min
timeSpentDeciding

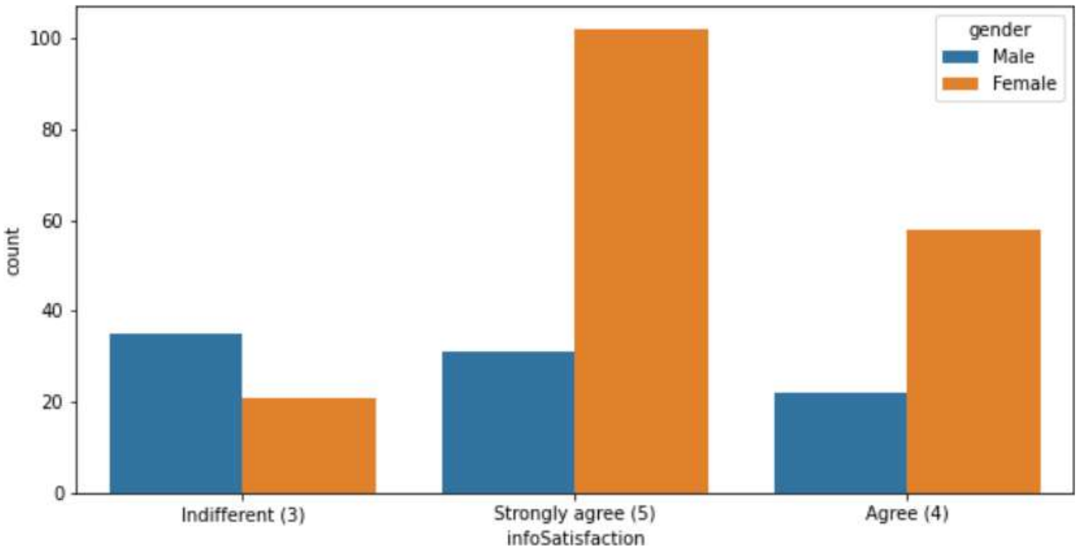

gender
Male
Female
count
100
80
60
40
20
0
Lack of trust
Promo code not applicable
Better alternative offer
Change in price
No preferred mode of payment
cancellationReason

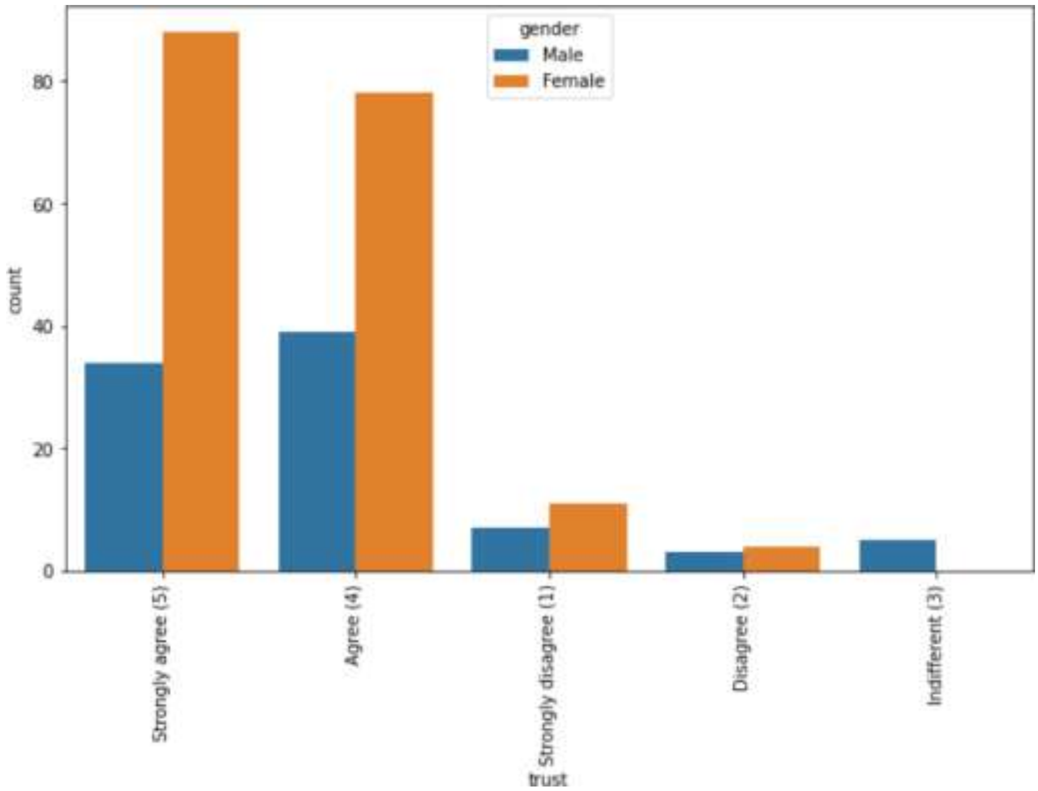

gender
Male
Female
count
100
80
60
40
20
0
Indifferent (3)
Strongly agree (5)
Agree (4)
Disagree (2)
similarProductInfo

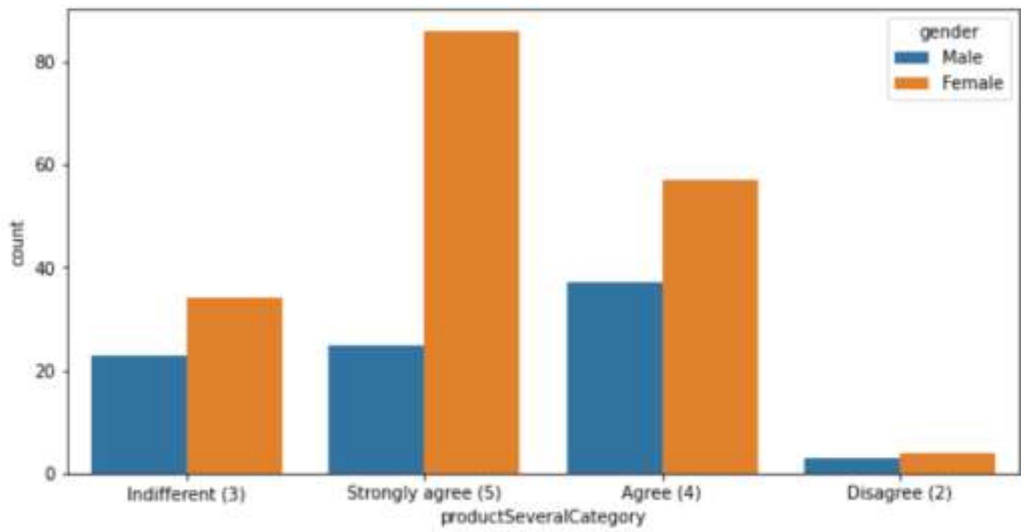

gender
Male
Female
count
140
120
100
80
60
40
20
0
Strongly agree (5)
Agree (4)
Strongly disagree (1)
Indifferent (3)
custSupportResponse

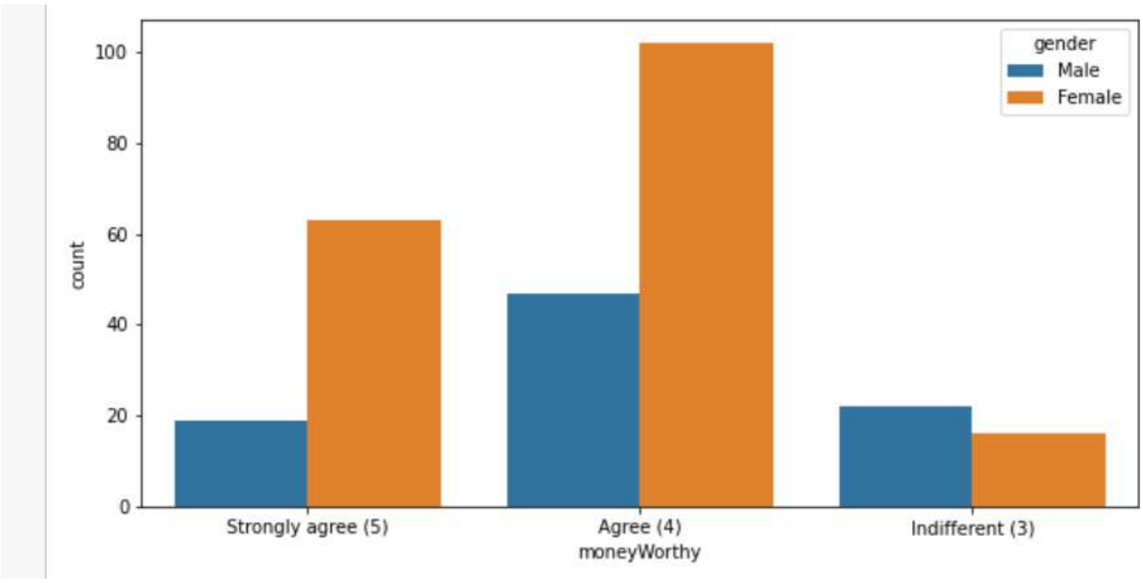

gender
Male
Female
count
100
80
60
40
20
0
Disagree (2)
Strongly agree (5)
Indifferent (3)
Agree (4)
convenience

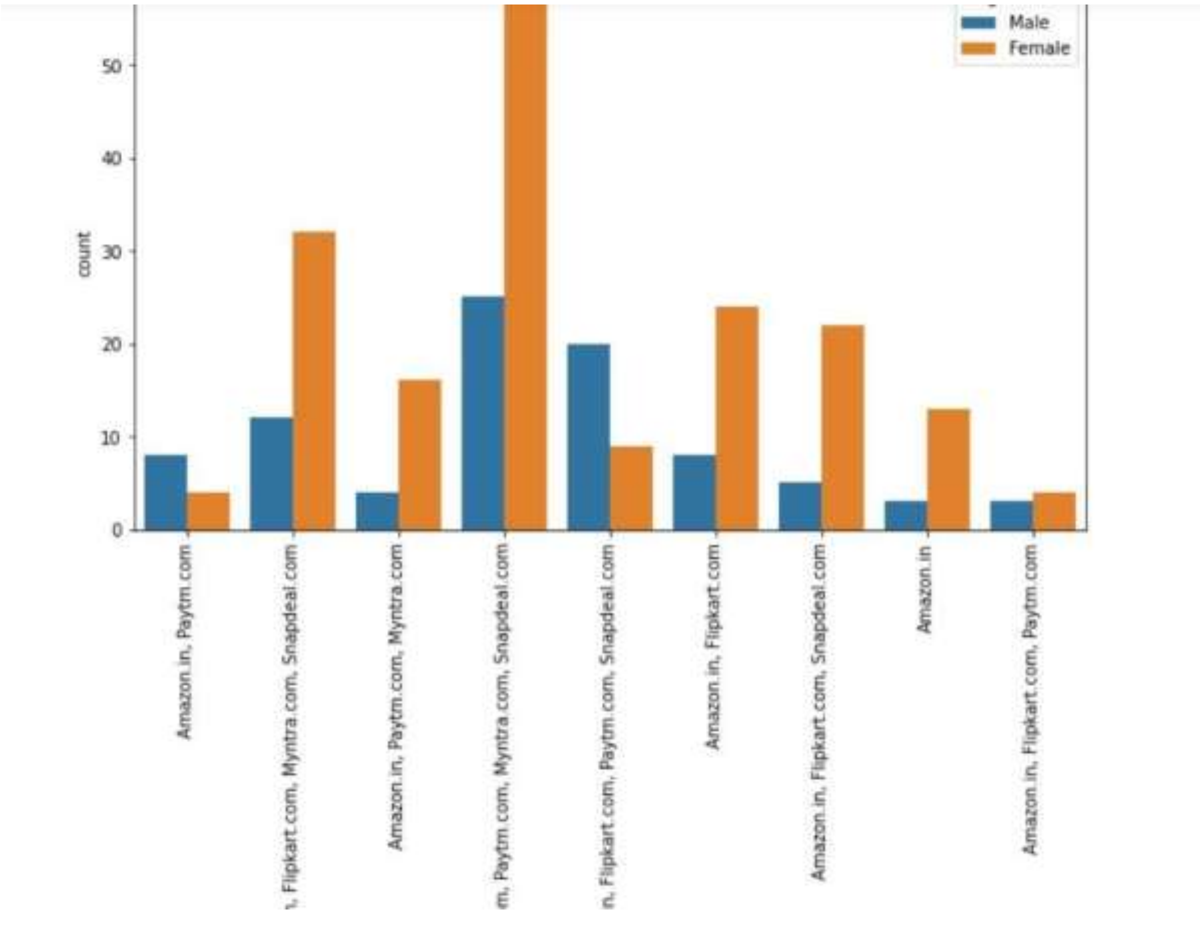

gender
Male
Female
count
Disagree (2)
Strongly agree (5)
Agree (4)
returnReplacementPolicy

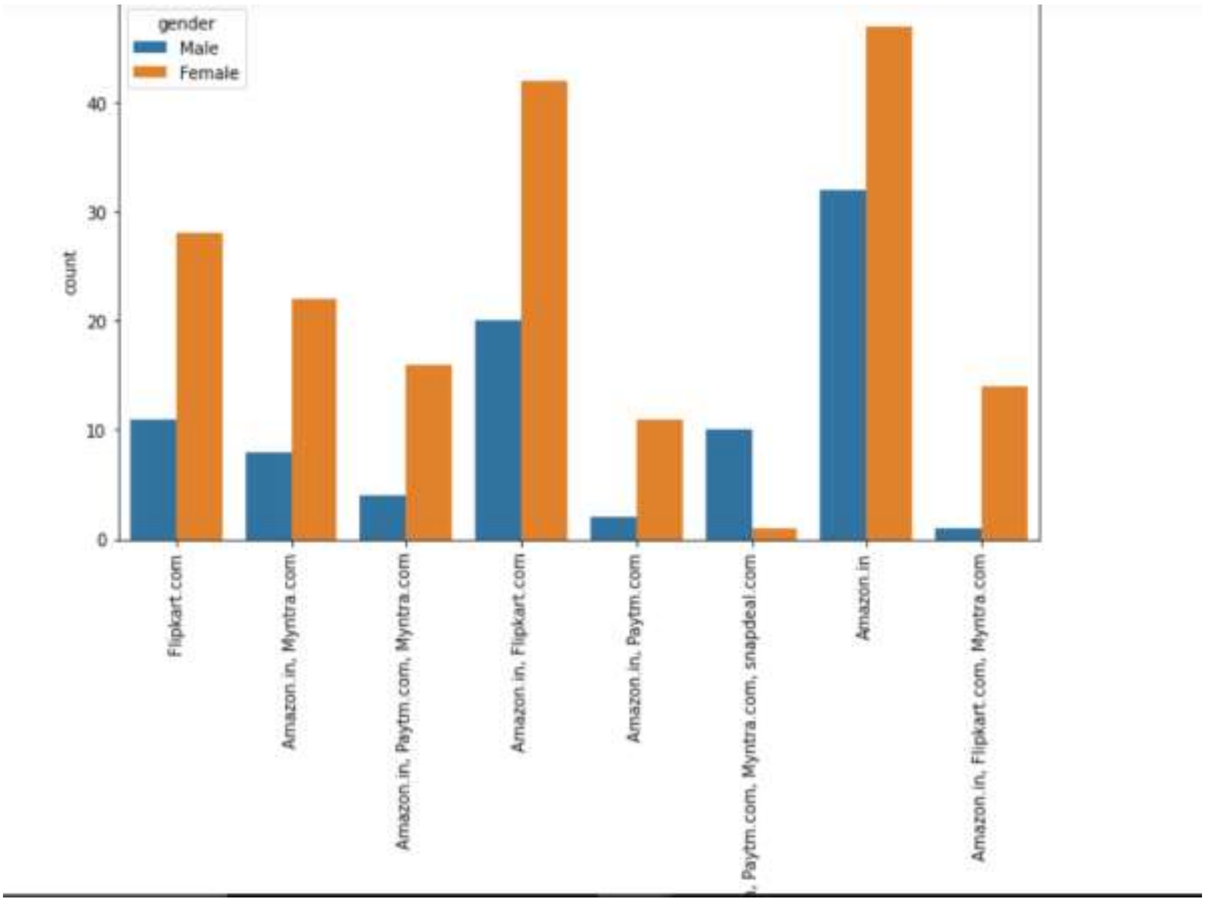

gender
Male
Female
count
Indifferent (3)
Strongly agree (5)
Agree (4)
infoSatisfaction


gender
Male
Female
count
80
60
40
20
0
Strongly agree (5)
Agree (4)
Strongly disagree (1)
Disagree (2)
Indifferent (3)
trust


gender
Male
Female
count
80
60
40
20
0
Indifferent (3)
Strongly agree (5)
Agree (4)
Disagree (2)
productSeveralCategory


gender
Male
Female
count
Strongly agree (5)
Agree (4)
Indifferent (3)
moneyWorthy


Male
Female
count
Amazon.in, Paytm.com
Amazon.in, Paytm.com, Myntra.com
Amazon.in, Flipkart.com
Amazon.in, Flipkart.com, Snapdeal.com
Amazon.in
Amazon.in, Flipkart.com, Paytm.com



OBSERVATIONS:

•Even though females are dominant in the dataset, female population using ecommerce is not dominant everywhere In Delhi and Noida more numbers of males are using ecommerce than female.

•Bulandshahr and Moradabad - both cities in Uttar Pradesh has no female shoppers at all.

• Numbers of females who are shopping since long are more than number of males.

• Frequency of females shopping online is more than males

•Most of the females usually take more than 15 meanwhile most of the male spend 6-10 min before making a purchase decision.

•More females strongly agree with the fact that there must be similar products to compare. Males do agree with the fact but the data suggests that it is okay if there exists a product but also okay if there does not exist one.

```
In [33]: # Let's check the Purchase decisions are based on the following factors :

purchase_factors = data[['contentReadability', 'productInfoClarity','relevantProductInfo', 'navigationEase',
                          'loadingProcessingSpeed','userFrienlyInterface','sellerProductInfo','similarProductInfo']]

x_data=[]

label = ['Strongly agree (5)', 'Agree (4)', 'Indifferent (3)', 'Disagree (2)','Strongly disagree (1)']
for each in purchase_factors.columns:
    data = pd.DataFrame(purchase_factors[each].value_counts()).reset_index()
    data.columns = ['factor','count']
    #data = data.sort_values(by='factor')

    category = [l for l in data['factor']]
    missing_label = list(set(label).difference(category))

    for miss in missing_label:
        i = len(data)+1
        data.loc[i,:] = miss, 0

    data = data.sort_values(by='factor')
    data['percentages'] = round((data['count']/data['count'].sum())*100,2)
    x = [p for p in data['percentages']]
    x_data.append(x)

top_labels = ['Agree<br>(4)', 'Disagree<br>(2)','Indifferent<br>(3)','Strongly<br>agree<br>(5)','Strongly<br>disagree<br>(1)']

y_data = ['Content on the website must be easy to read and understand.',
         'All relevant information on listed<br>products must be stated clearly',
         'Provision of complete<br>and relevant product Information',
```

```
y_data = ['Content on the website must be easy to read and understand.',
         'All relevant information on listed<br>products must be stated clearly',
         'Provision of complete<br>and relevant product information',
         'Ease of navigation in website',
         'Loading and processing speed',
         'Complete information on listed seller and<br>product being offered is important',
         'Information on similar product to the<br>one highlighted is important for comparison']

colors = ['#004D40','#00796B','#009688','#4DB6AC','#B2DFDB']

fig = go.Figure()

for i in range(0, len(x_data[0])):
    for xd, yd in zip(x_data, y_data):
        fig.add_trace(go.Bar(
            x=[xd[i]], y=[yd],
            orientation='h',
            marker=dict(
                color=colors[i],
                line=dict(color='rgb(248, 248, 249)', width=1)
            )
        ))
```

```
fig.update_layout(
    xaxis=dict(
        showgrid=False,
        showline=False,
        showticklabels=False,
        zeroline=False,
        domain=[0.15, 1]
    ),
    yaxis=dict(
        showgrid=False,
        showline=False,
        showticklabels=False,
        zeroline=False,
    ),
    barmode='stack',
    paper_bgcolor='rgb(248, 248, 255)',
    plot_bgcolor='rgb(248, 248, 255)',
    margin=dict(l=150, r=5, t=150, b=80),
    showlegend=False,
    title='Purchase decisions on online Retail Store are based on following factors : '
)

annotations = []
```

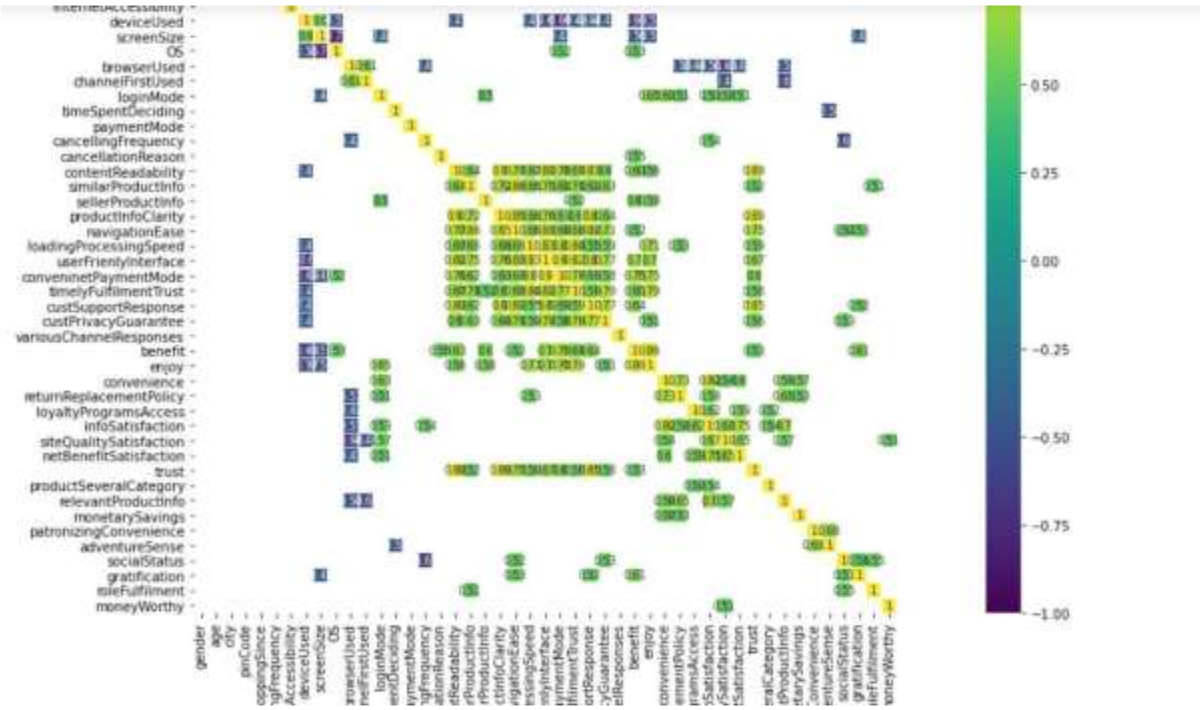
```
for yd, xd in zip(y_data, x_data):
    # labeling the y-axis
    annotations.append(dict(xref='paper', yref='y',
                            x=0.14, y=yd,
                            xanchor='right',
                            text=str(yd),
                            font=dict(family='Arial', size=12,
                                      color='rgb(67, 67, 67)'),
                            showarrow=False, align='right'))
    # labeling the first percentage of each bar (x_axis)
    annotations.append(dict(xref='x', yref='y',
                            x=xd[0] / 2, y=yd,
                            text=str(xd[0]) + '%',
                            font=dict(family='Arial', size=12,
                                      color='rgb(248, 248, 255)'),
                            showarrow=False))
    # labeling the first Likert scale (on the top)
    if yd == y_data[-1]:
        annotations.append(dict(xref='x', yref='paper',
                                x=xd[0] / 2, y=1.1,
                                text=top_labels[0],
                                font=dict(family='Arial', size=12,
                                          color='rgb(67, 67, 67)'),
                                showarrow=False))
```

```
    space = xd[0]
    for i in range(1, len(xd)):
            # labeling the rest of percentages for each bar (x_axis)
            annotations.append(dict(xref='x', yref='y',
                                    x=space + (xd[i]/2), y=yd,
                                    text=str(xd[i]) + '%',
                                    font=dict(family='Arial', size=11,
                                              color='rgb(248, 248, 255)'),
                                    showarrow=False))
            # labeling the Likert scale
            if yd == y_data[-1]:
                annotations.append(dict(xref='x', yref='paper',
                                        x=space + (xd[i]/2), y=1.1,
                                        text=top_labels[i],
                                        font=dict(family='Arial', size=11,
                                                  color='rgb(67, 67, 67)'),
                                        showarrow=False))
            space += xd[i]

fig.update_layout(annotations=annotations)

fig.show()
```

Observations:

- We can observe that, mostly agree with the fact that the product they are purchasing from the app or website, they hope to have complete information regarding the product.
- Most of them agrees that the functioning of app efficiently is also a major factor which helps in enhance user experience while doing online shopping.

In [34]:
```
# let's find the columns which has more correlation - for people to purchase online and the app that they use :

data = encoded_data.iloc[:,:47]
le = LabelEncoder()
data['city'] = le.fit_transform(data['city'])

corr =data.corr()
plt.figure(figsize=(25, 10))

sns.heatmap(corr[(corr >= 0.5) | (corr <= -0.4)],
            cmap='viridis', vmax=1.0, vmin=-1.0, linewidths=0.1,
            annot=True, annot_kws={"size": 8}, square=True);
plt.title("Correlation between variables: ")
```



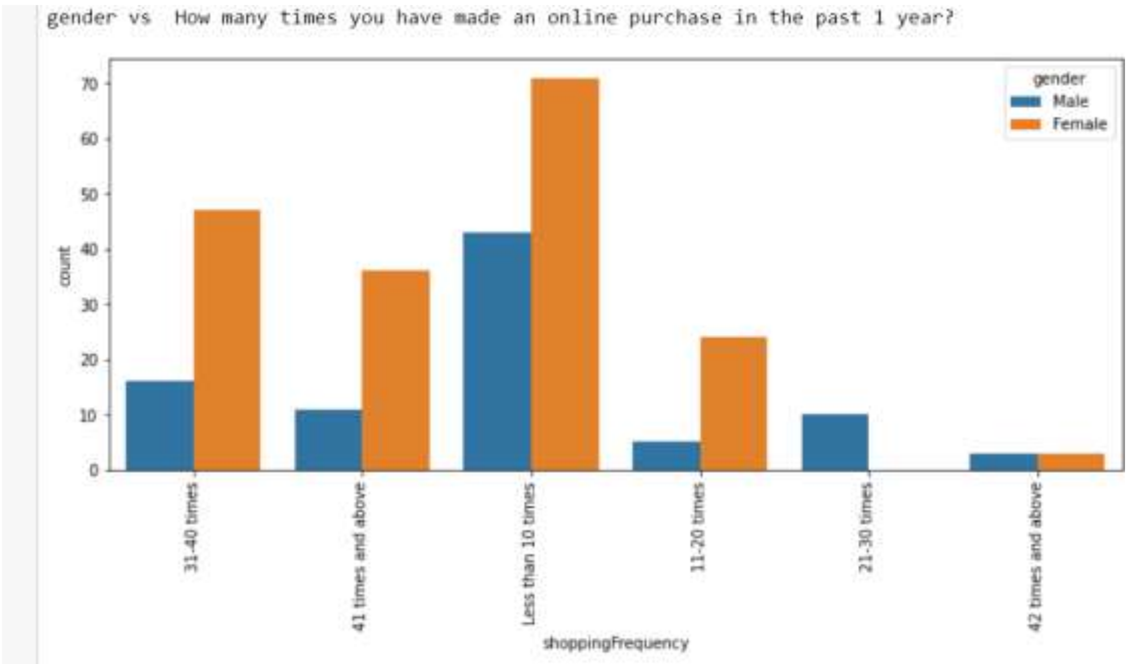
Observation:

- From the data we can see that, data related to how the person is accessing the app or website does not matter as it has negative correlation.
- Customer retention can be done, majorly with customer reviews and by finding out if they are satisfied with the quality of product and experience delivered to them.

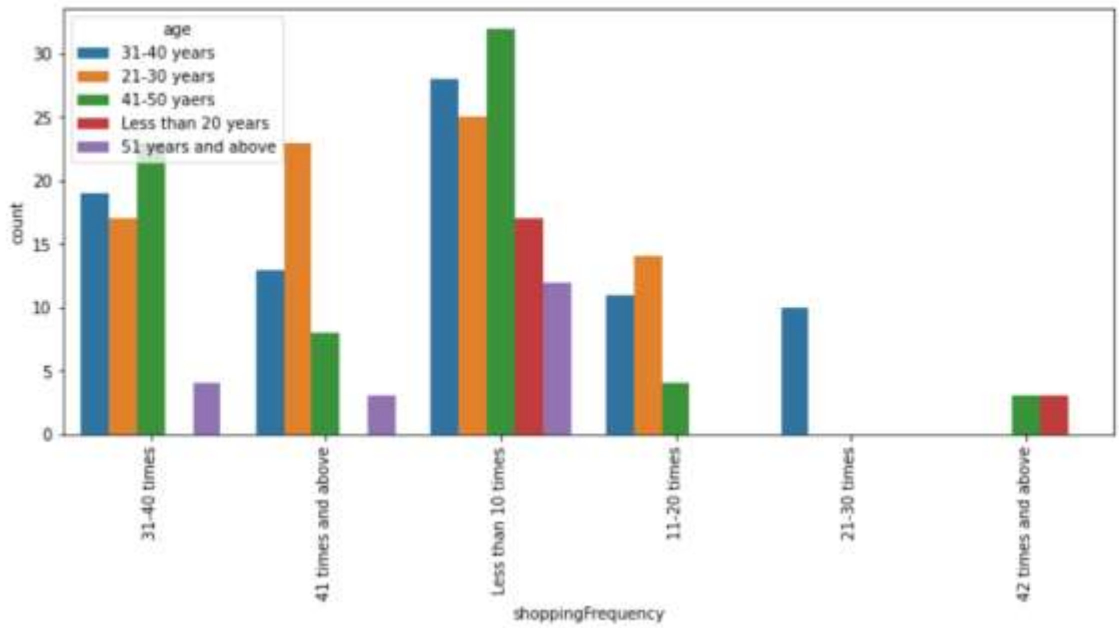
BIVARIATE ANALYSIS

```
In [31]: data['shoppingFrequency'].value_counts()

Out[31]: Less than 10 times      114
         31-40 times              63
         41 times and above       47
         11-20 times              29
         21-30 times              10
         42 times and above        6
         Name: shoppingFrequency, dtype: int64
```

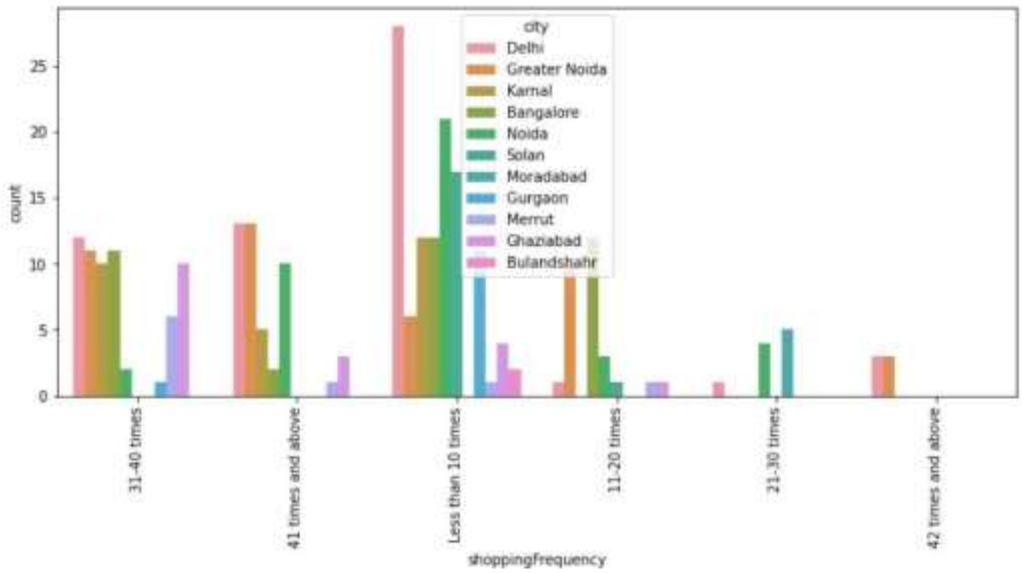
```
In [32]: data2=data.iloc[:,[0,1,2,4,5,13,14,16,26,31,32,34,37,38,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,
         for i in data2:
             print(i,'vs  How many times you have made an online purchase in the past 1 year?')
             plt.subplots(figsize=(12,5))
             sns.countplot(x='shoppingFrequency',hue=i,data=data2)
             plt.xticks(rotation=90)
             plt.show()

         gender vs  How many times you have made an online purchase in the past 1 year?
```



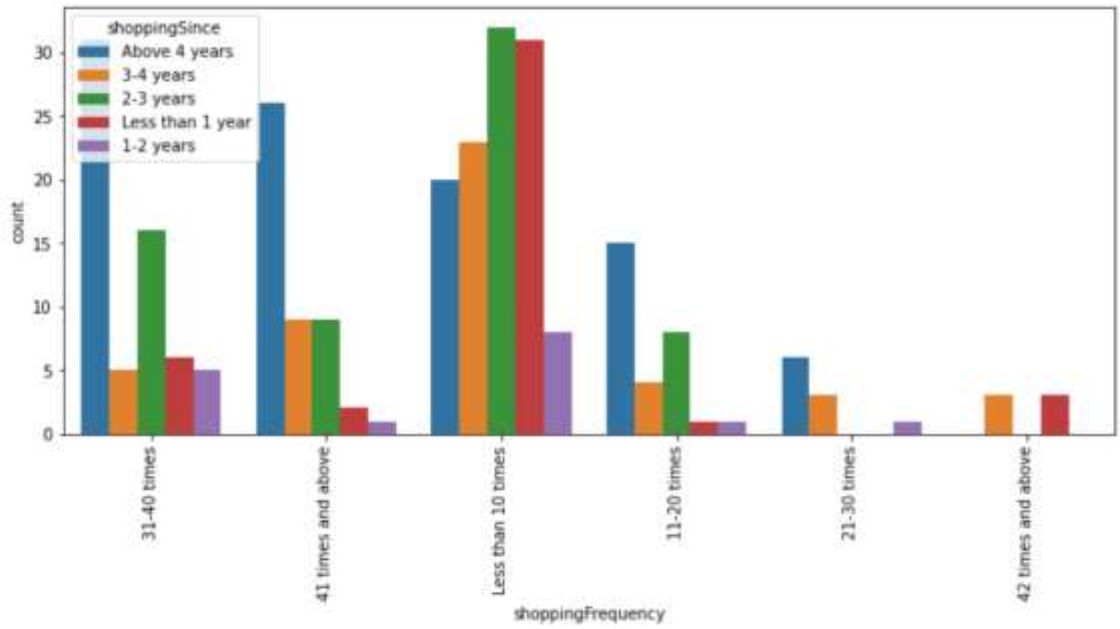
age vs How many times you have made an online purchase in the past 1 year?



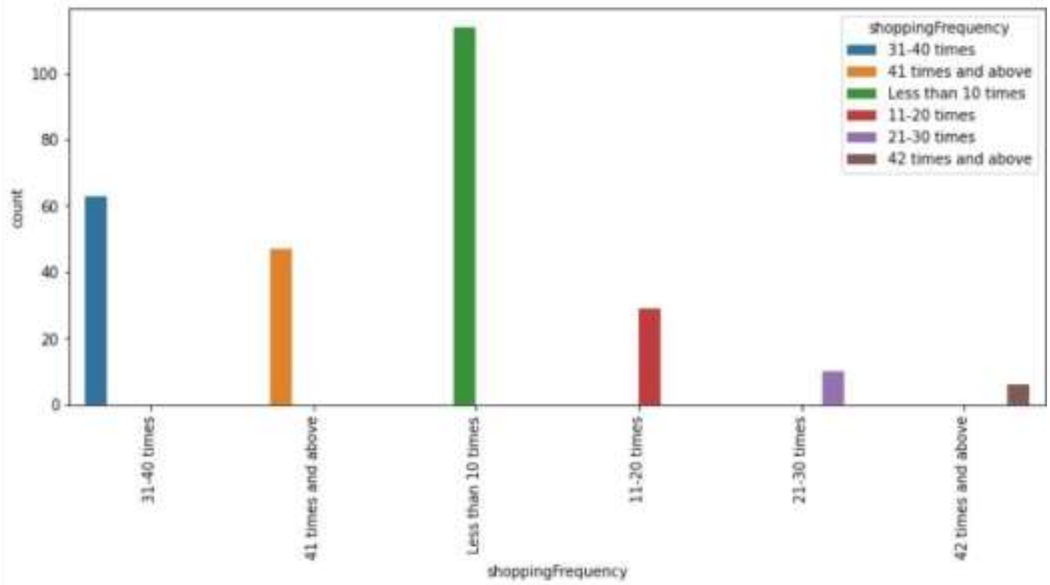
city vs How many times you have made an online purchase in the past 1 year?



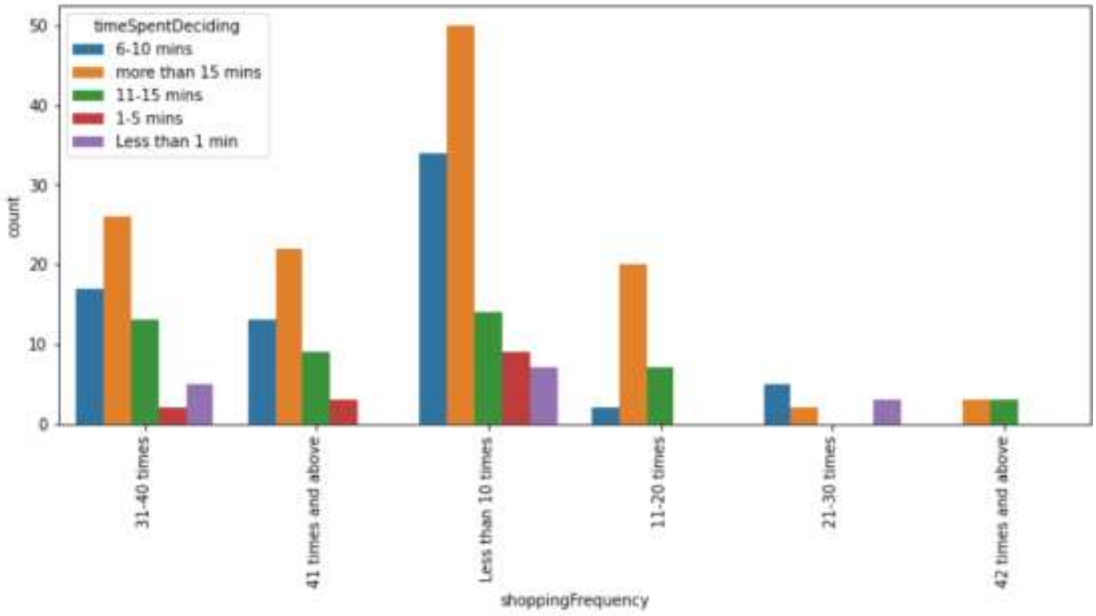
shoppingSince vs How many times you have made an online purchase in the past 1 year?



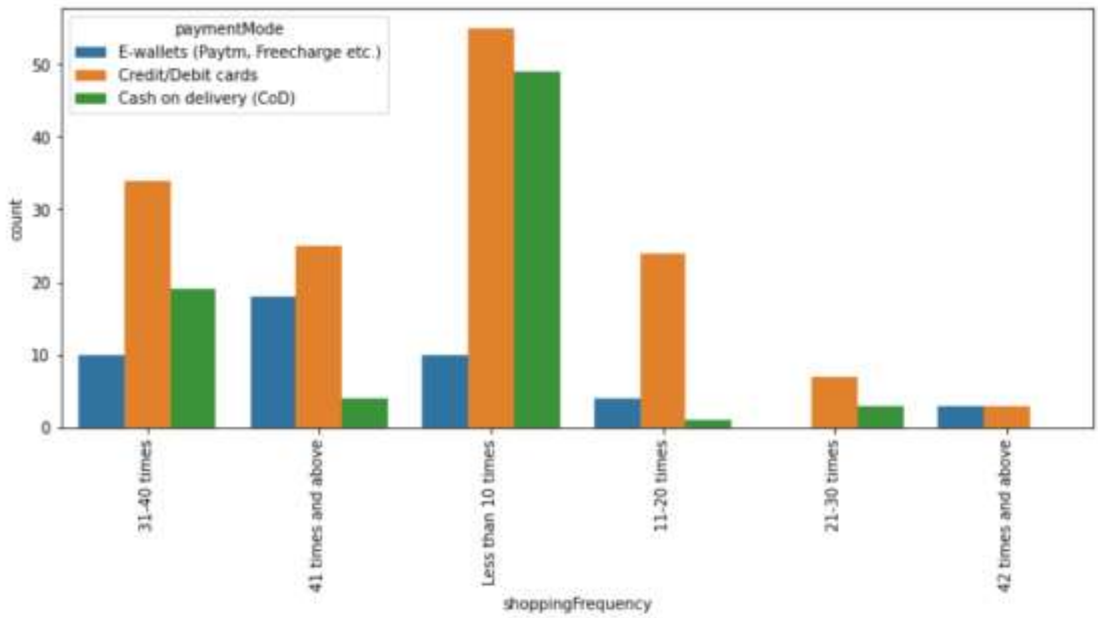
shoppingFrequency vs How many times you have made an online purchase in the past 1 year?



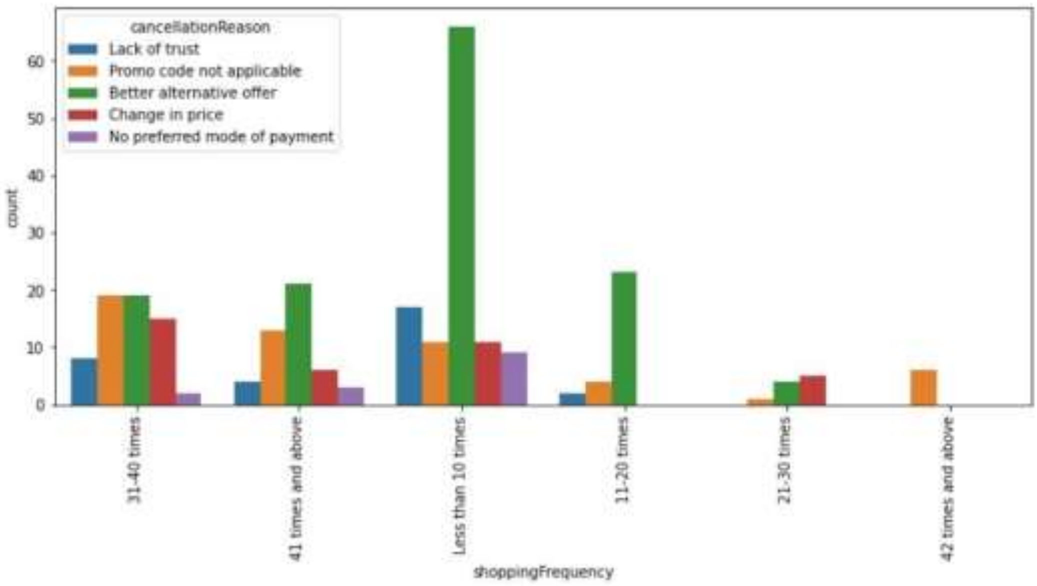
timeSpentDeciding vs How many times you have made an online purchase in the past 1 year?



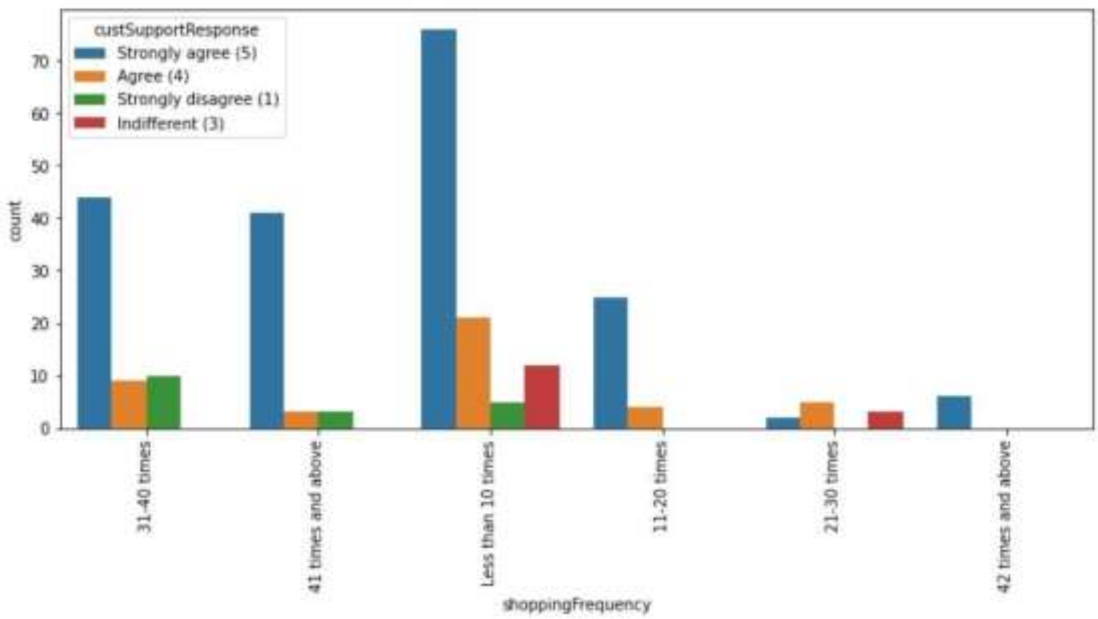
paymentMode vs How many times you have made an online purchase in the past 1 year?



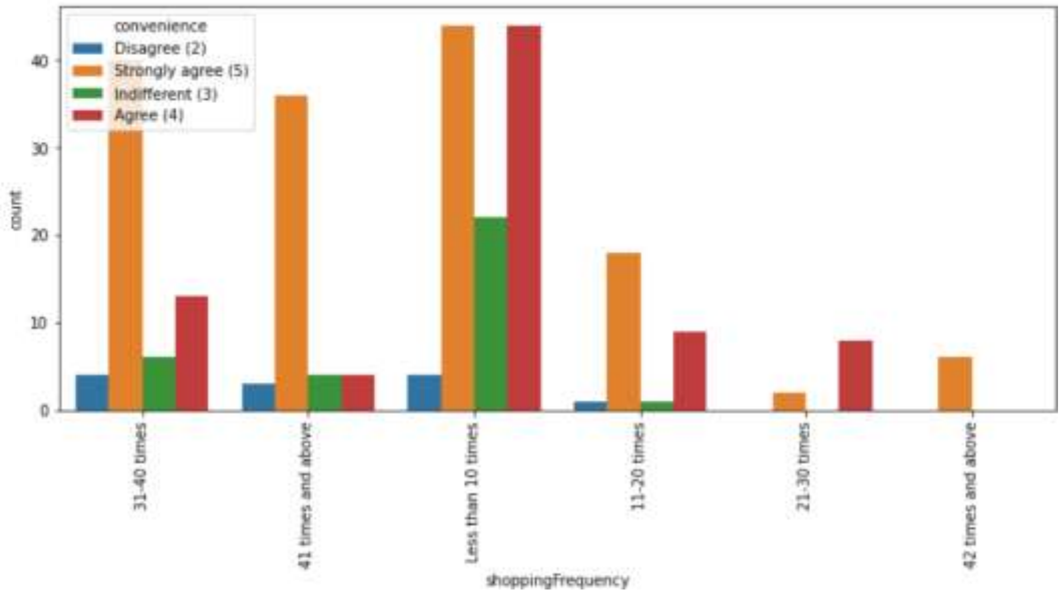
cancellationReason vs How many times you have made an online purchase in the past 1 year?



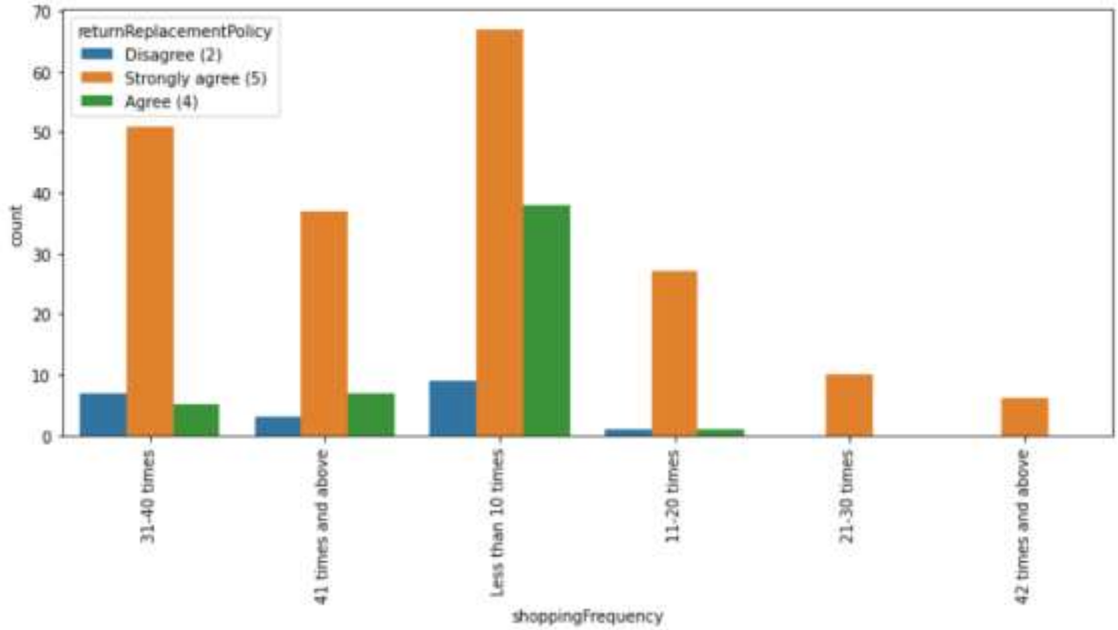
custSupportResponse vs How many times you have made an online purchase in the past 1 year?



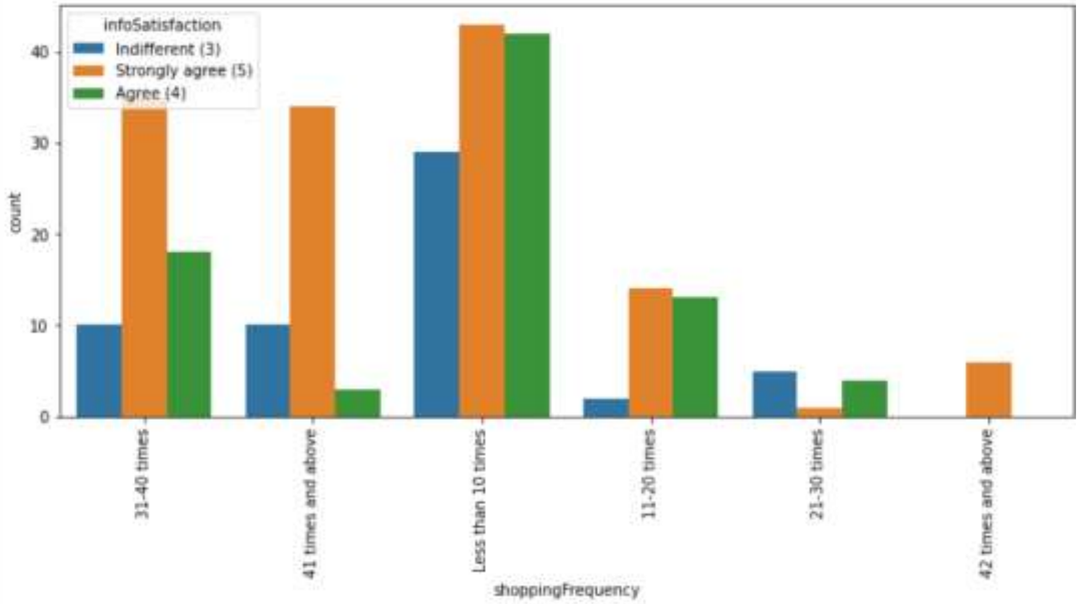
convenience vs How many times you have made an online purchase in the past 1 year?



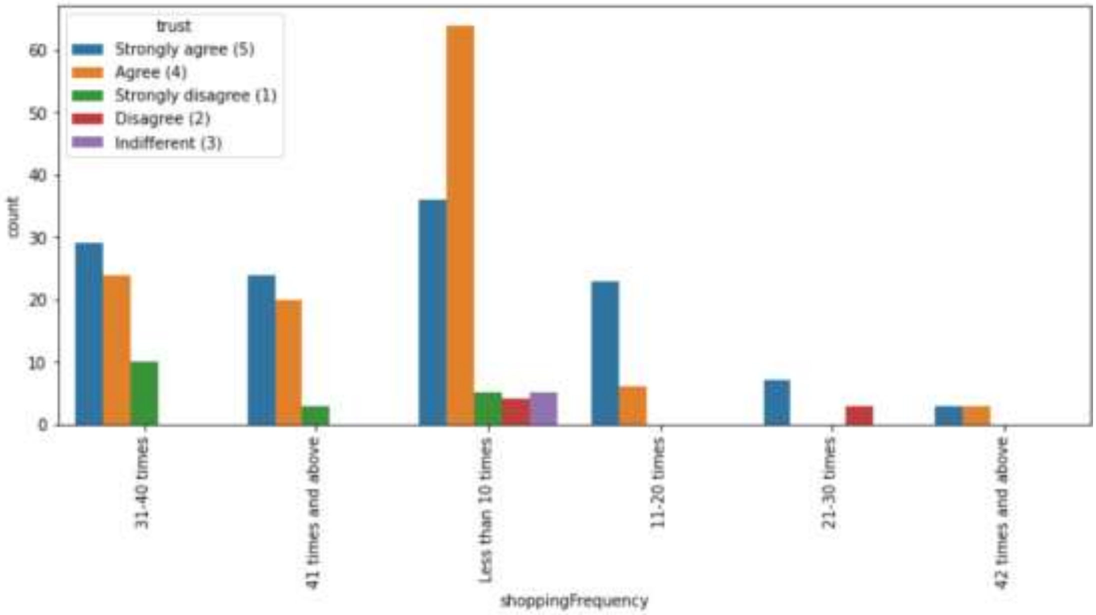
returnReplacementPolicy vs How many times you have made an online purchase in the past 1 year?



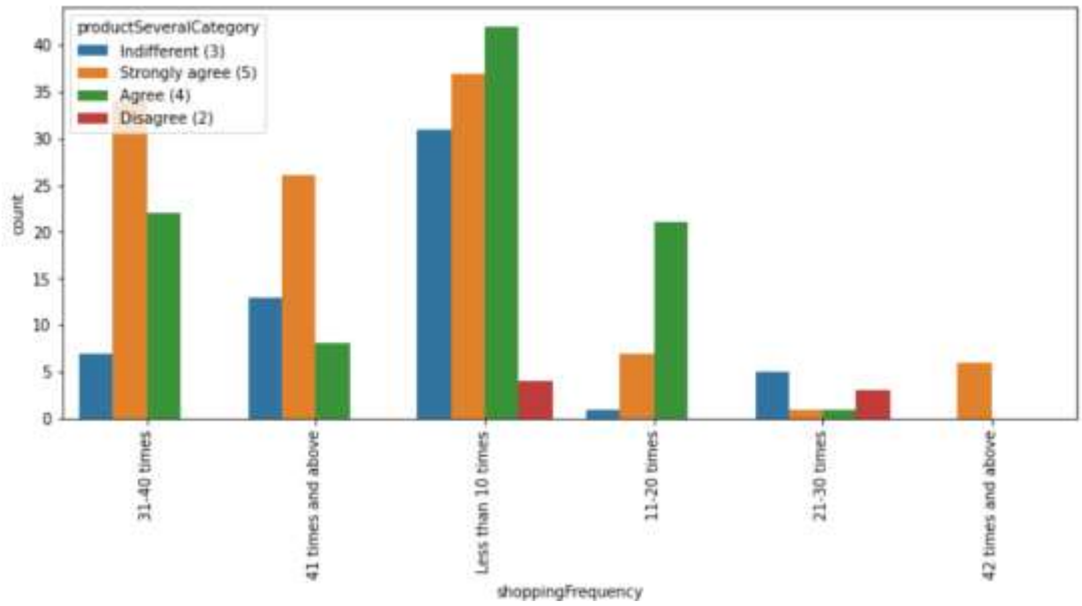
infoSatisfaction vs How many times you have made an online purchase in the past 1 year?



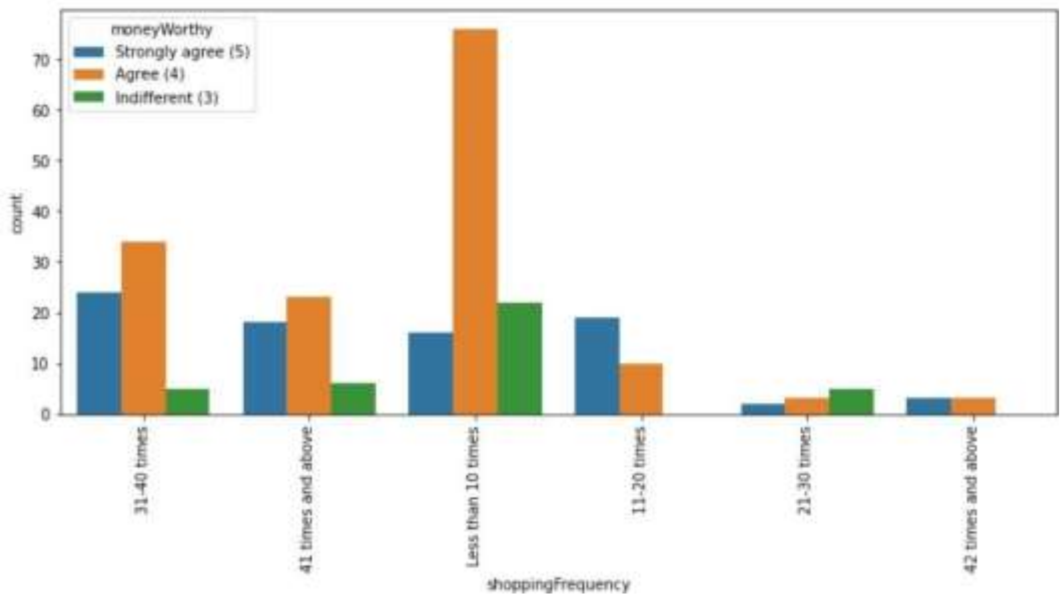
trust vs How many times you have made an online purchase in the past 1 year?



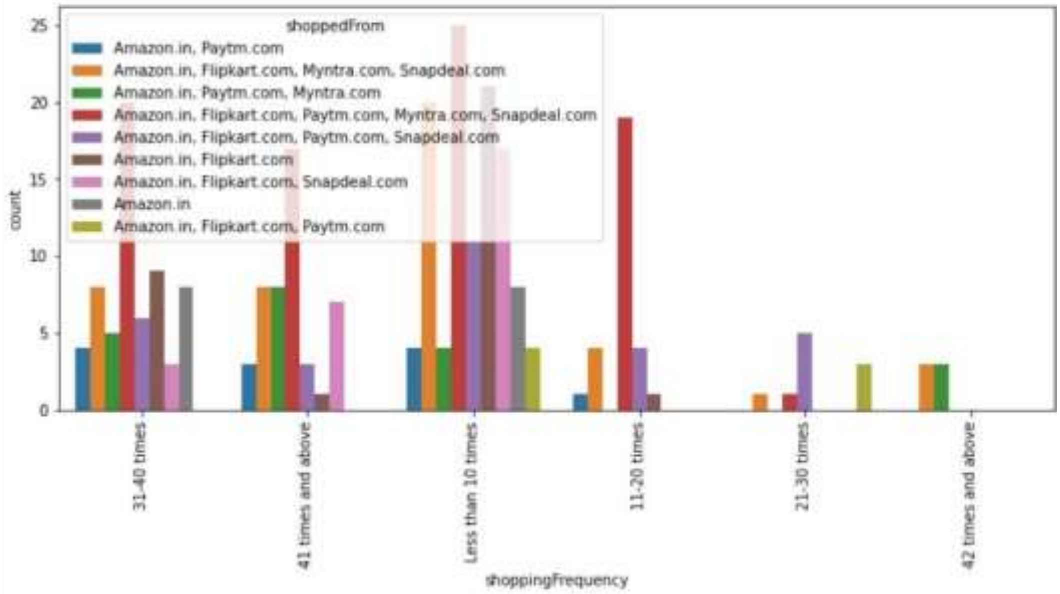
productSeveralCategory vs How many times you have made an online purchase in the past 1 year?



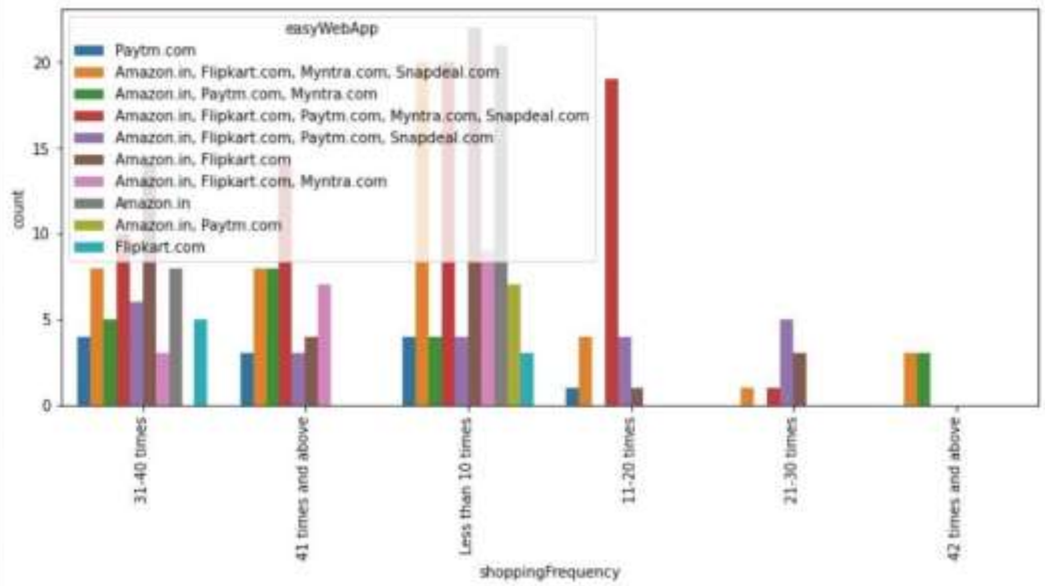
moneyWorthy vs How many times you have made an online purchase in the past 1 year?



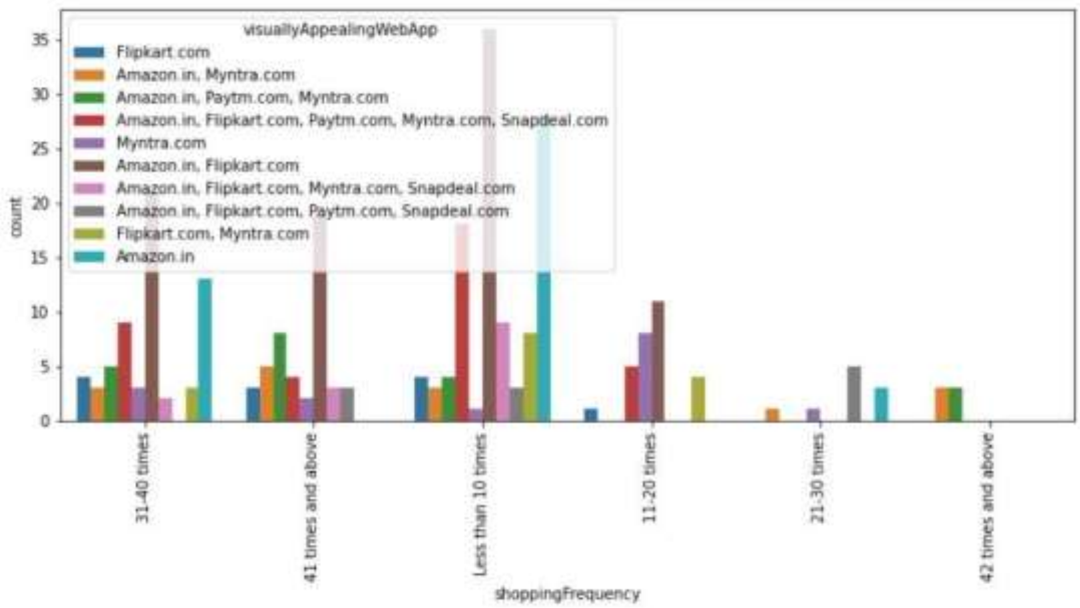
shoppedFrom vs How many times you have made an online purchase in the past 1 year?



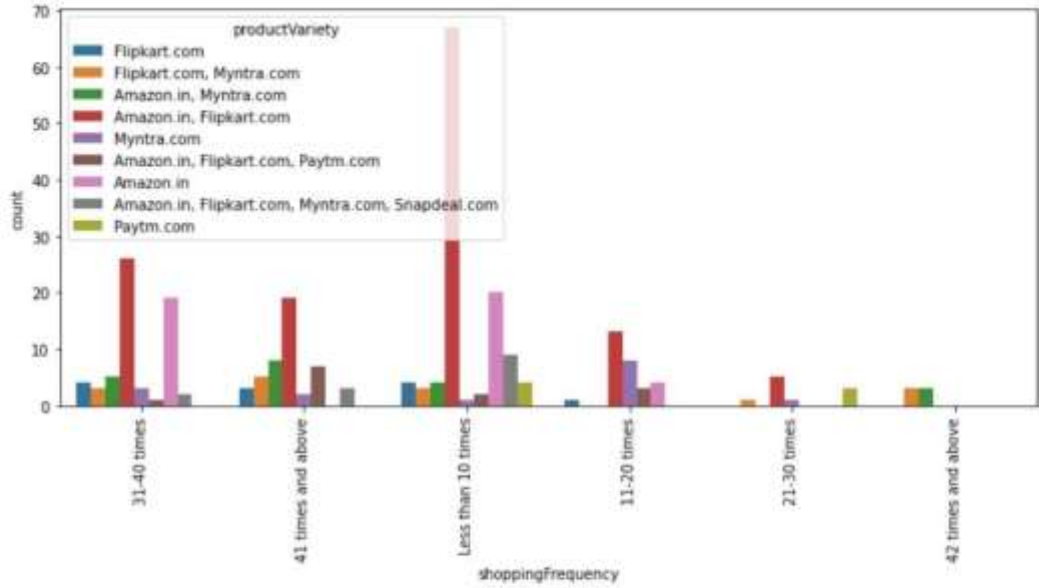
easyWebApp vs How many times you have made an online purchase in the past 1 year?



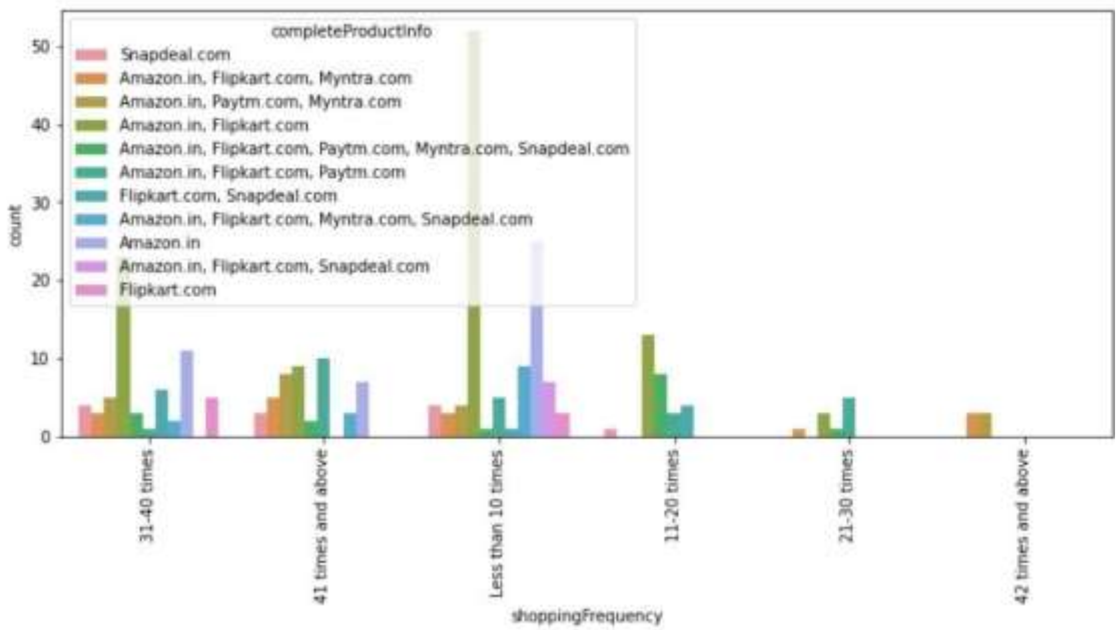
visuallyAppealingWebApp vs How many times you have made an online purchase in the past 1 year?



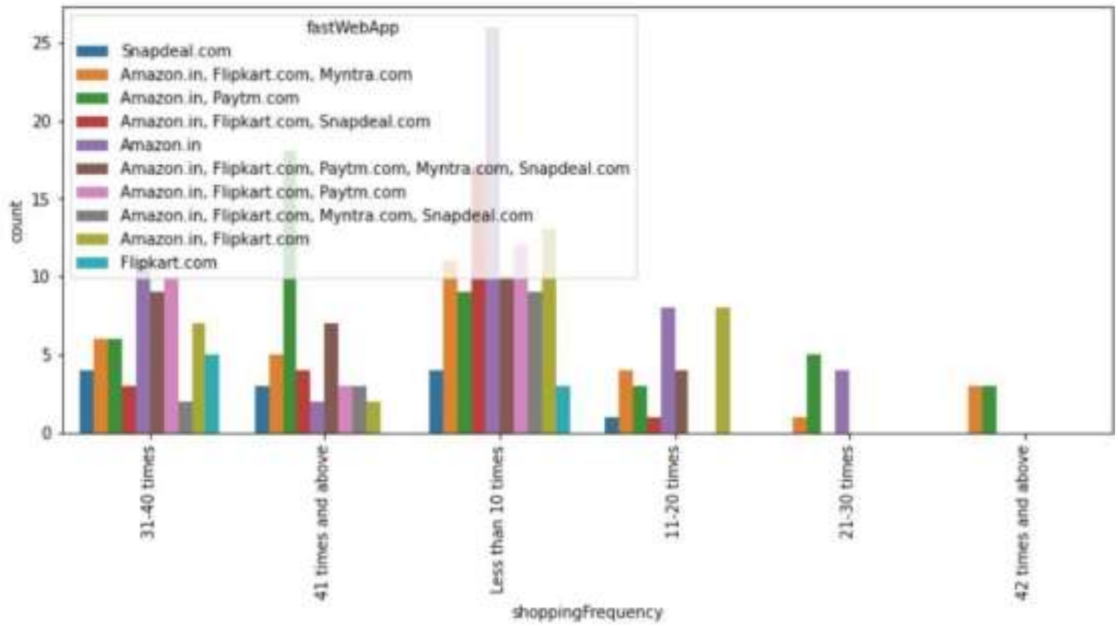
productVariety vs How many times you have made an online purchase in the past 1 year?



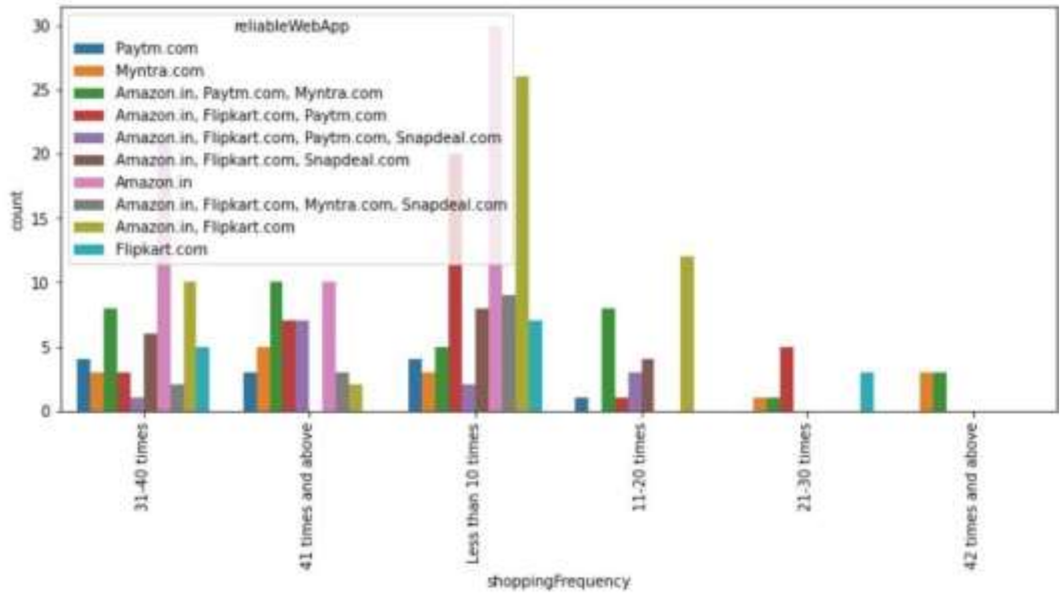
completeProductInfo vs How many times you have made an online purchase in the past 1 year?



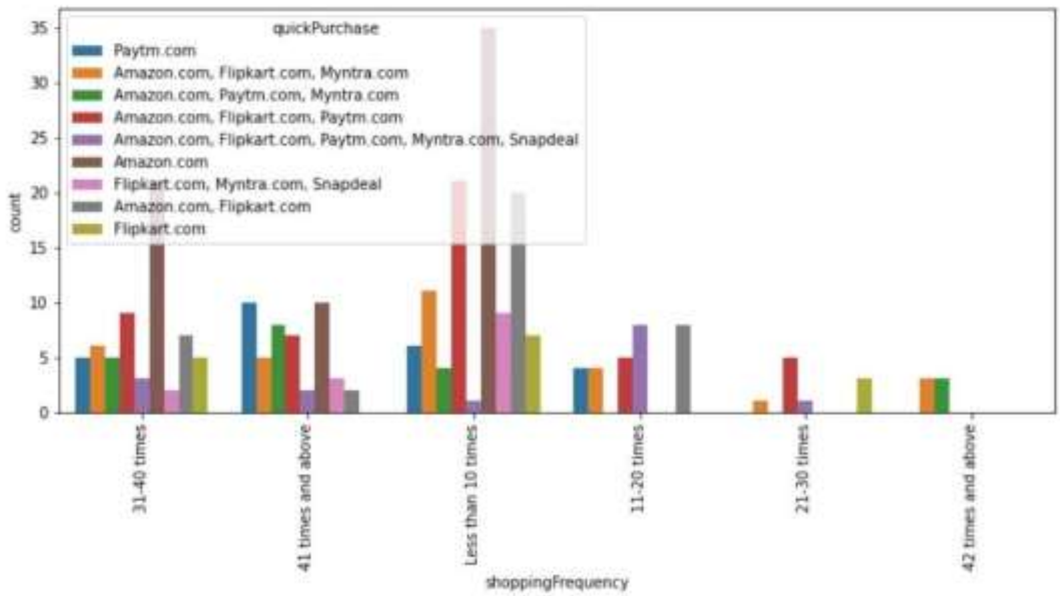
fastWebApp vs How many times you have made an online purchase in the past 1 year?



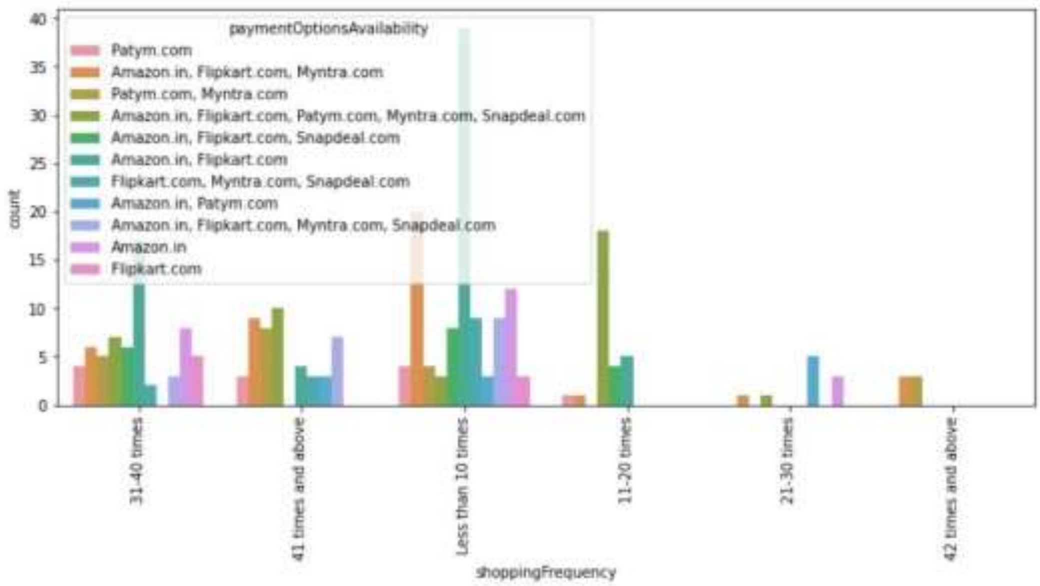
reliableWebApp vs How many times you have made an online purchase in the past 1 year?



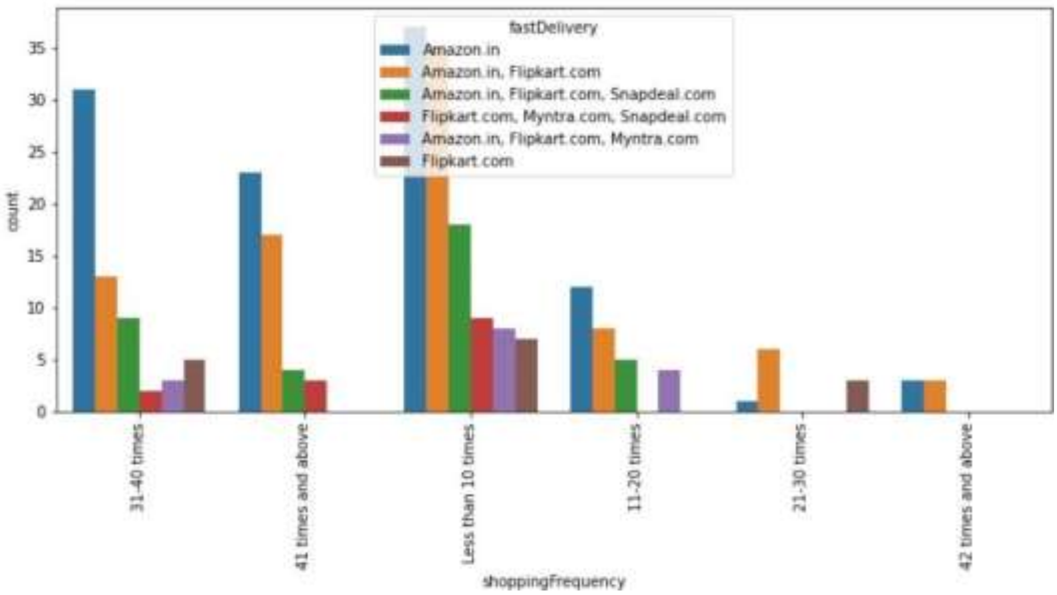
quickPurchase vs How many times you have made an online purchase in the past 1 year?



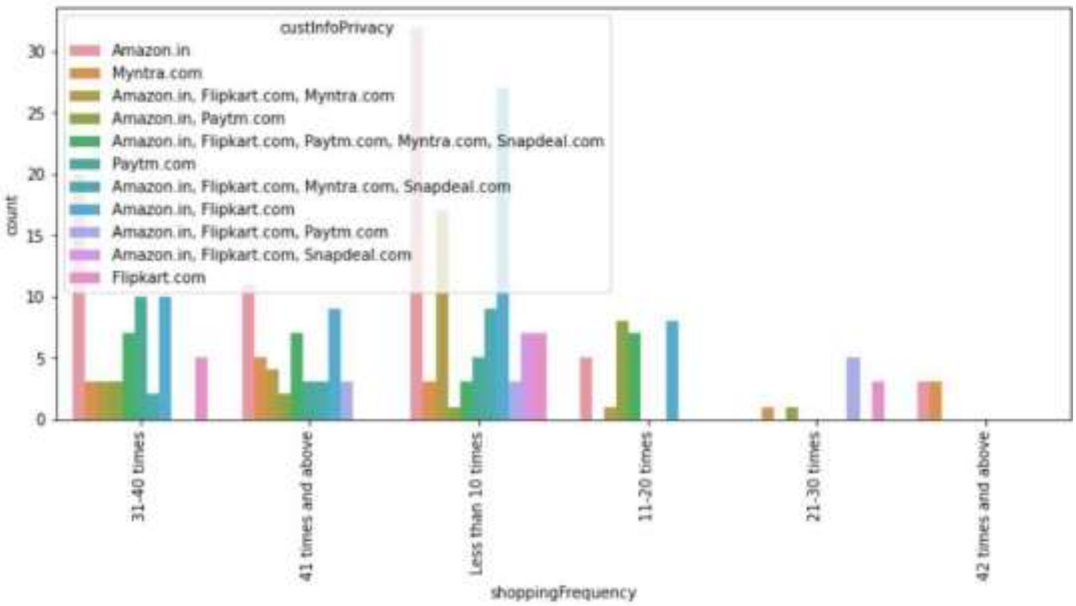
paymentOptionsAvailability vs How many times you have made an online purchase in the past 1 year?



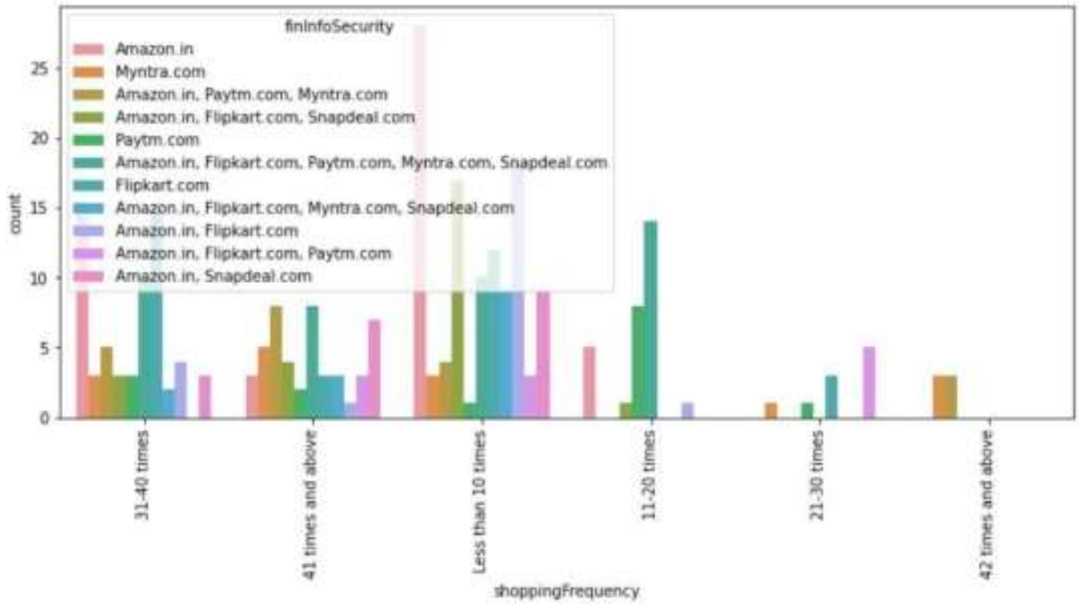
fastDelivery vs How many times you have made an online purchase in the past 1 year?



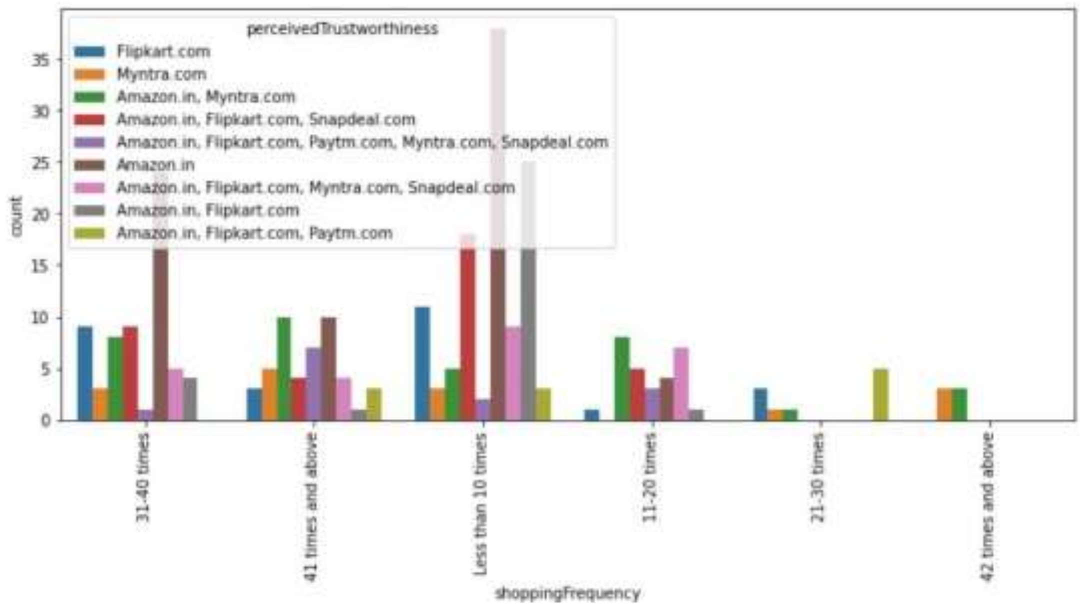
custInfoPrivacy vs How many times you have made an online purchase in the past 1 year?



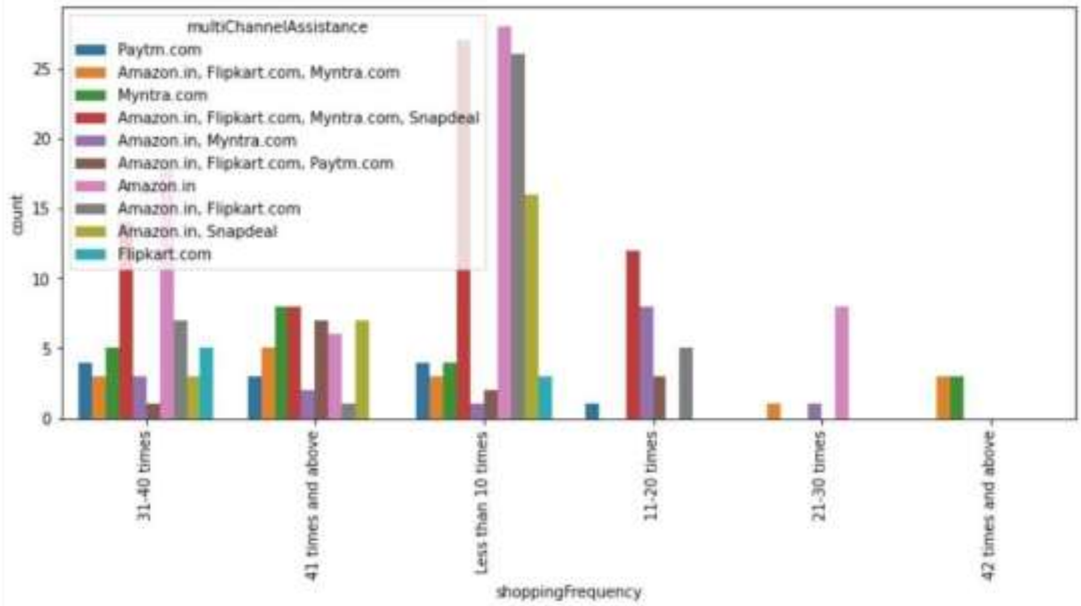
finInfoSecurity vs How many times you have made an online purchase in the past 1 year?



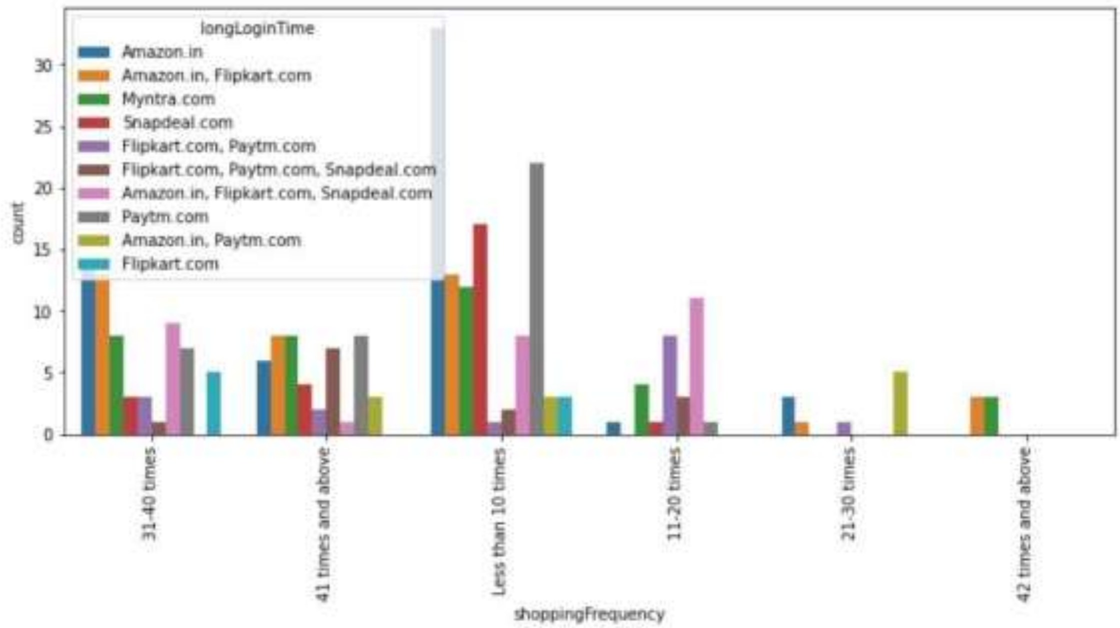
perceivedTrustworthiness vs How many times you have made an online purchase in the past 1 year?



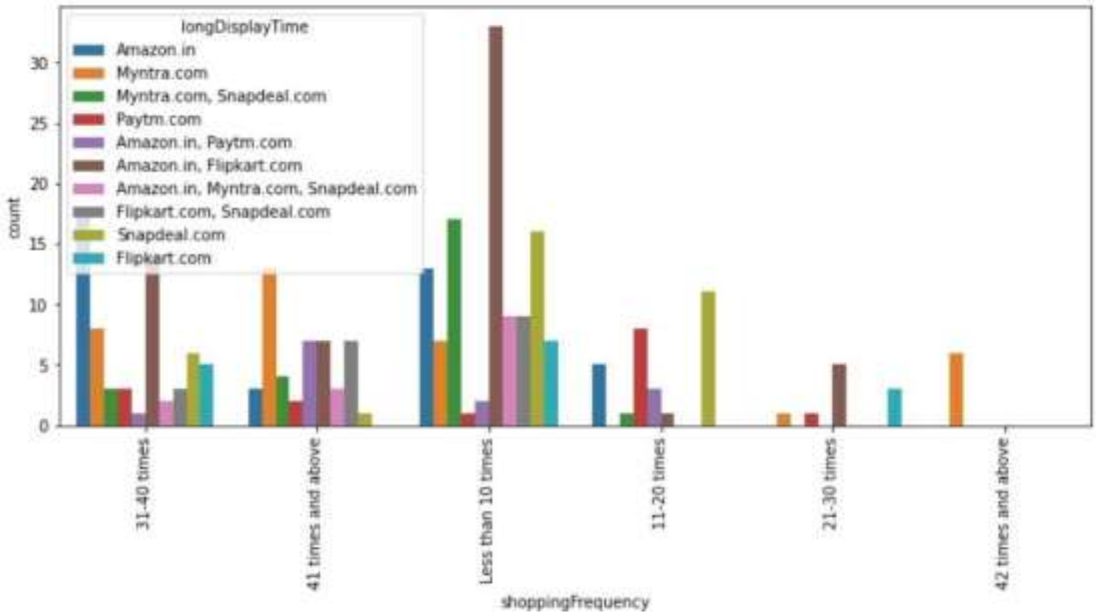
multiChannelAssistance vs How many times you have made an online purchase in the past 1 year?



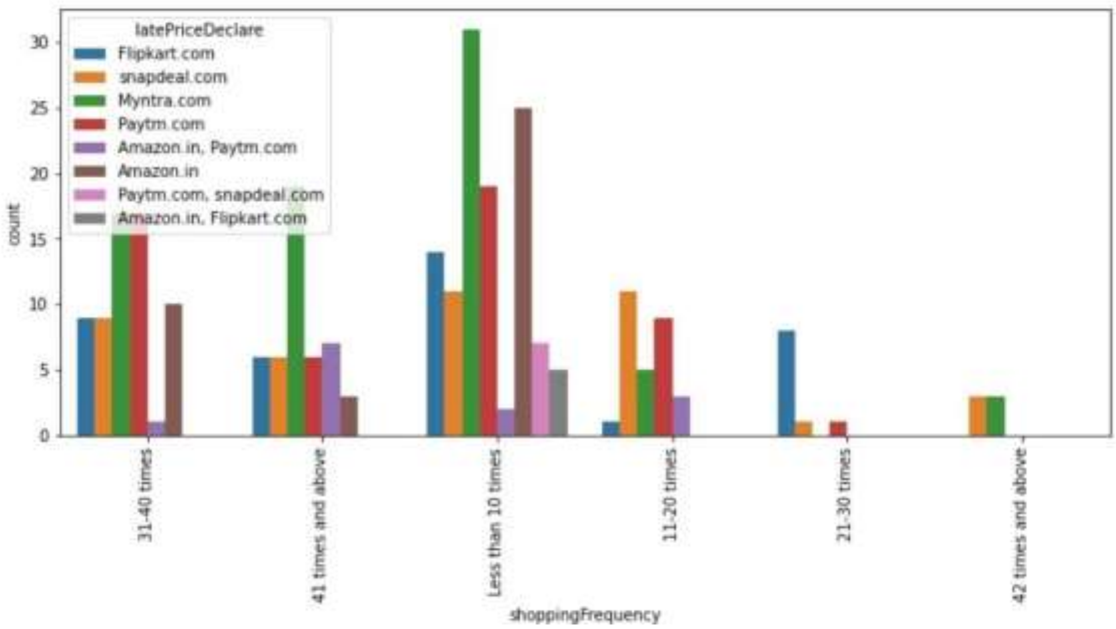
longLoginTime vs How many times you have made an online purchase in the past 1 year?



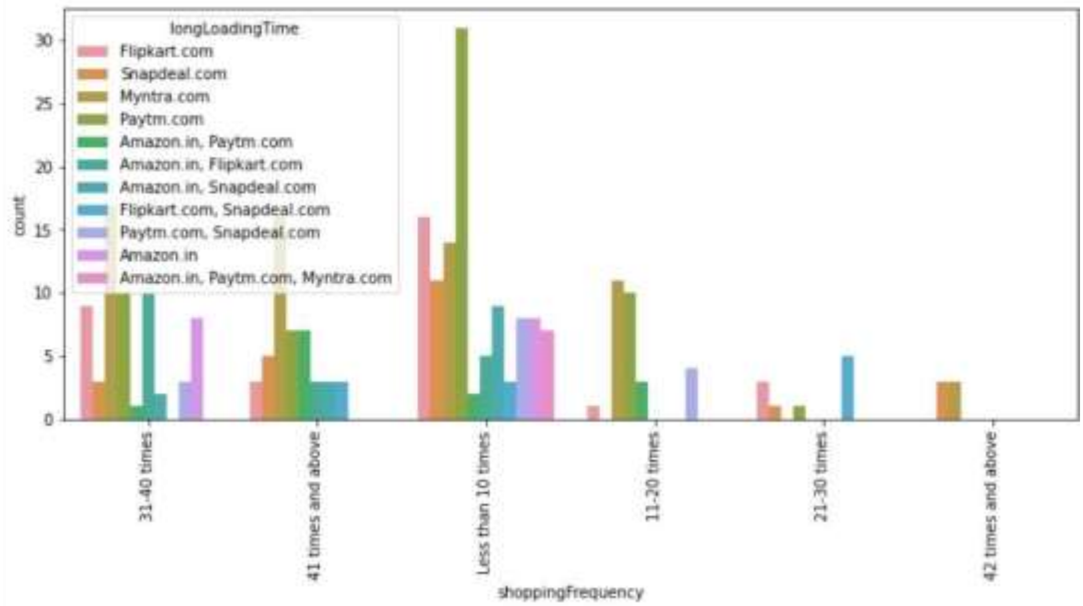
longDisplayTime vs How many times you have made an online purchase in the past 1 year?



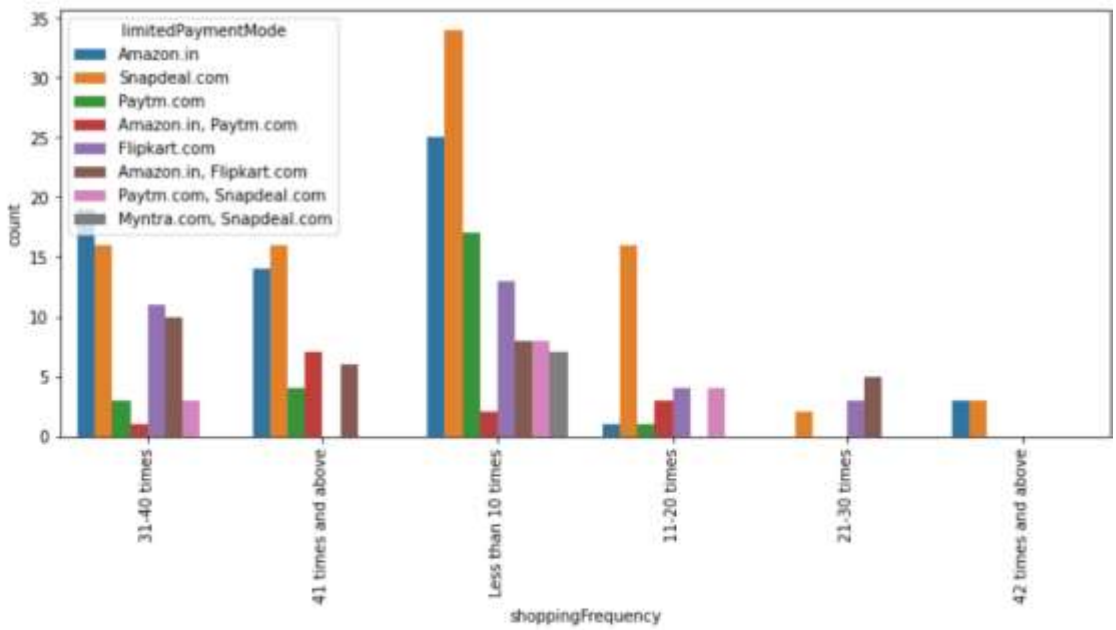
latePriceDeclare vs How many times you have made an online purchase in the past 1 year?



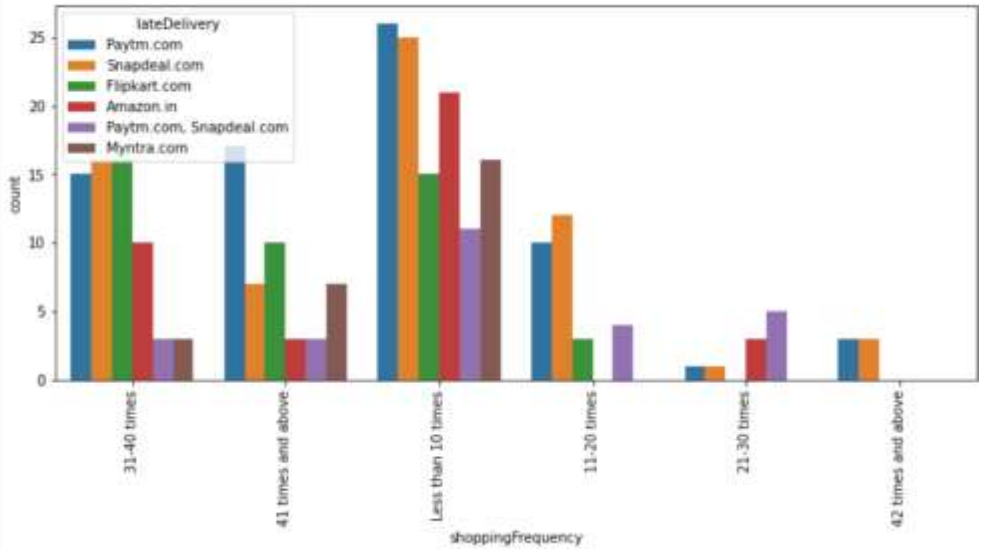
longLoadingTime vs How many times you have made an online purchase in the past 1 year?



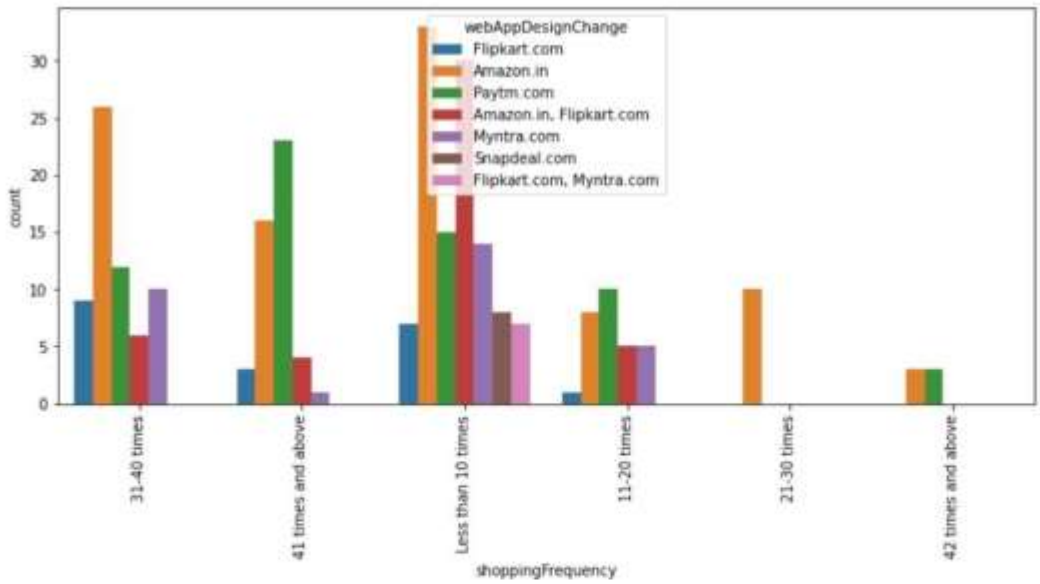
limitedPaymentMode vs How many times you have made an online purchase in the past 1 year?



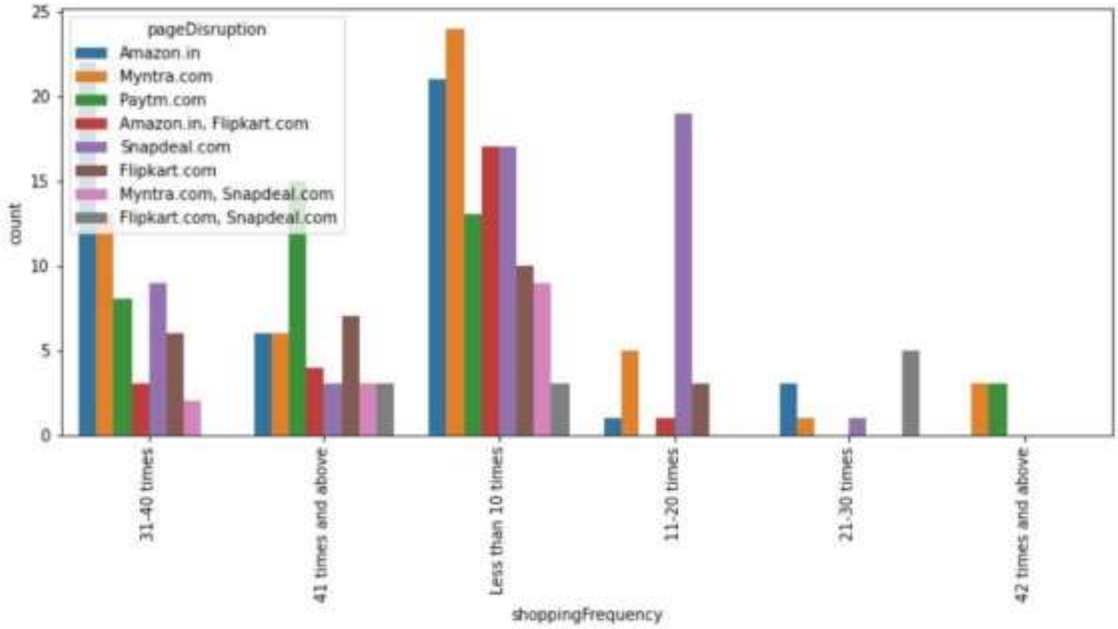
lateDelivery vs How many times you have made an online purchase in the past 1 year?



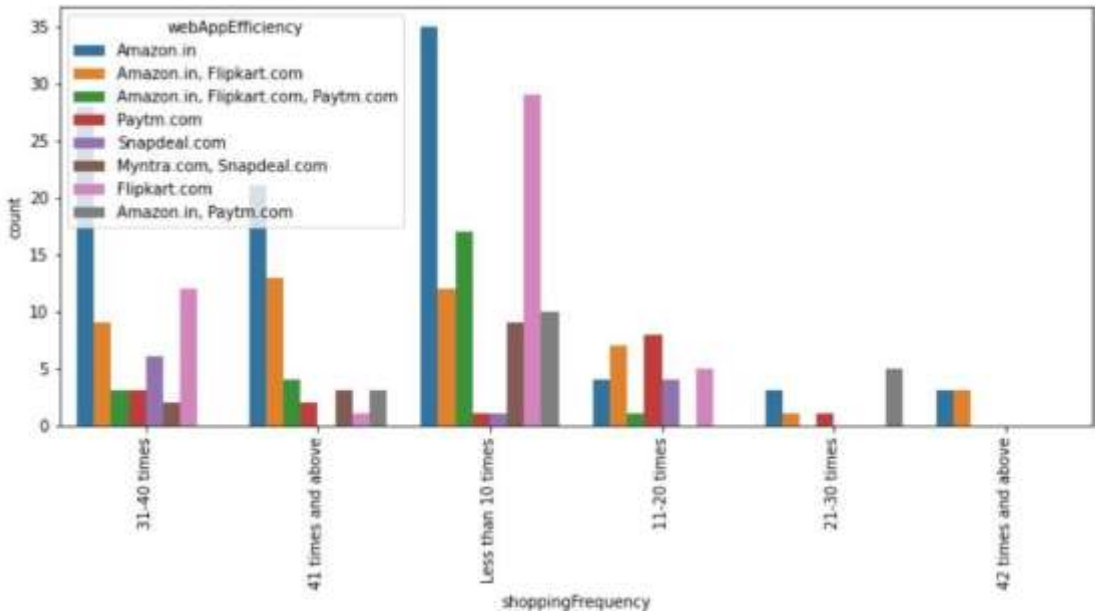
webAppDesignChange vs How many times you have made an online purchase in the past 1 year?



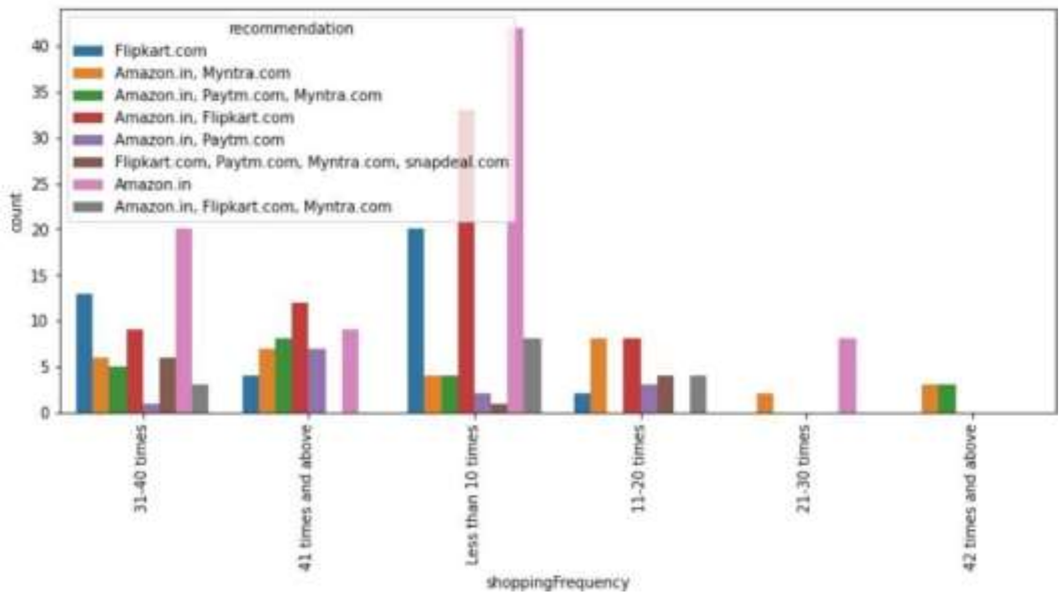
pageDisruption vs How many times you have made an online purchase in the past 1 year?



webAppEfficiency vs How many times you have made an online purchase in the past 1 year?



recommendation vs How many times you have made an online purchase in the past 1 year?



OBSERVATION

• Females have shopped a greater number of times in last 1 year

•Most of the individuals that do online shopping are from the age group 21-30 yrs. and, 31-40 yrs. People below 20 years, &people over 50 years of age do not do much online shopping.

•Customers from Delhi, Noida & Bangalore region have shopped the greatest number of times

•Most customers who are shopping online from 1 to 4 years have shopped less than 10 times and customers shopping online for above 4 years have highest purchase count in all more than 10 times group

•Most customers spend more than 15 minutes before making decision irrespective of the purchase count

•Most of the customers generally preferred credit/Debit cards for payment followed by cash on delivery. Very few customers opt for E-wallet payment.

•Most customers who purchased less than 10 times abandoned the cart due to better alternative offer. Customers purchasing more than 41 times abandoned due to promo code not applied or better alternative offer

•It can be seen clearly that all customers irrespective of the purchase count strongly agree with the empathy which means empathy is very important for customer satisfaction and hence retention. The reason is insecurity amongst the customers. If they have a provision of a helpline number associated with the website, that sense of insecurity vanishes and the customer can trust the website more. A 24X7 support is a must for these online retail stores.

•Maximum customers agree with online shopping being flexible, irrespective of purchase count as it provides freedom to shop without the influence or pressure from sales staff often encountered in offline stores.

•We can clearly see that return and replacement policy plays an important role in customer satisfaction.

•Displaying quality Information on the website improves satisfaction of customers: Most of the customers strongly agreed to this. Shoppers expect online retailers to provide all relevant and accurate information about the product. Providing appropriate information can help online retailers to dispel concerns and fears of consumers towards a particular product or online shopping

•User satisfaction cannot exist without trust: We see that, user with lack of trust have shopped less number of time. It means, trust is an area where online shopping still lags as several instances have been reported of customers being duped on receiving products that did not match the description claimed or turned out to be knock-offs. So, it is important for companies to learn how to manage consumers' trust

•Customers who have shopped more than 30 times strongly agree that net benefit from online purchase leads to user satisfaction Hence net benefit derived is an important factor for customer retention

•Offering a wide variety of listed product in several categories: Customers having purchase frequency of more than 30 times strongly agree to it. It means product variety increases consumers' likelihood of finding a good match with their preferences

• Customers with high purchase (>30) agree to get monetary savings

•Almost all customers with high purchase agree that they get value for money spent.it indicates that quality of product sold plays an important role in customer retention

• Highest number of customers have shopped from amazon, Paytm and Myntra

• Highest number of customers like amazon, Paytm and Myntra web page layout

•Highest number of customers like amazon and flipkart in terms of variety of product offered

•Highest number of customers like amazon and flipkart in terms of relevant and complete description of products

•High proportion of high purchase frequency customers like amazon and Paytm in terms of website speed

• Highest reliability is obtained by amazon and flipkart

•Highest number of customers like Amazon in terms of quickness to complete purchase

•In terms of payment option Amazon, Flipkart and Myntra are liked most by the high purchase frequency customers

•Amazon stands apart in terms of speed of order delivery and privacy of customers information

• Most of customers like amazon in terms of online assistance through multichannel

• Most user claims flipkart to take maximum login time during sales/promotion

•Highest customers claim Myntra take maximum time in displaying graphics and photos

INTERPRETATION OF THE RESULTS

Amazon & flipkart are liked by most number of people, in terms of online shopping.

Model Building :

```
In [65]: from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix,classification_report,auc
import warnings
warnings.filterwarnings('ignore')
#importing all the required libraries to find best Algorithm
from sklearn.naive_bayes import BernoulliNB
from sklearn.svm import SVC
from sklearn.tree import DecisionTreeClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import cross_val_score
model=[LogisticRegression(),KNeighborsClassifier(),BernoulliNB(),SVC(),DecisionTreeClassifier(),RandomForestClassifier(),KNeighb
for m in model:
    maxAccu=0
    maxRS=0
    for i in range(1,200):
        x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=.30,random_state=i)
        m.fit(x_train,y_train)
        predrf = m.predict(x_test)
        acc = accuracy_score(y_test, predrf)
        if acc>maxAccu:
            maxAccu=acc
            maxRS=i

    print("Best accuracy is",maxAccu," on Random_state ",maxRS, "for model ", m)
    x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=.30,random_state=maxRS)
```

```
    m.fit(x_train,y_train)
    pred=m.predict(x_test)
    print("*********************************************************************************")
    print('accuracy score of ->', m)
    print(accuracy_score(y_test,pred))
    print(confusion_matrix(y_test,pred))
    print(classification_report(y_test,pred))
    score=cross_val_score(m,x,y,cv=5)
    print(score)
    print(score.mean())
    print("Difference between Accuracy score and cross validatio score is - ",accuracy_score(y_test,pred)-score.mean())
    print("*********************************************************************************")
```

```
Best accuracy is 0.9382716049382716  on Random_state  47 for model  LogisticRegression()
*********************************************************************************
accuracy score of -> LogisticRegression()
0.9382716049382716
[[57  1]
 [ 4 19]]
              precision    recall  f1-score   support

           0       0.93      0.98      0.96        58
           1       0.95      0.83      0.88        23

    accuracy                           0.94        81
   macro avg       0.94      0.90      0.92        81
weighted avg       0.94      0.94      0.94        81

[0.74074074 0.83333333 0.98148148 0.85185185 0.90566038]
0.8626135569531795
Difference between Accuracy score and cross validatio score is -  0.07565804798509201
*********************************************************************************
Best accuracy is 0.9135802469135802  on Random_state  38 for model  KNeighborsClassifier()
*********************************************************************************
accuracy score of -> KNeighborsClassifier()
0.9135802469135802
[[52  5]
 [ 2 22]]
              precision    recall  f1-score   support

           0       0.96      0.91      0.94        57
           1       0.81      0.92      0.86        24

    accuracy                           0.91        81
   macro avg       0.89      0.91      0.90        81
weighted avg       0.92      0.91      0.91        81

[0.7037037  0.88888889 0.96296296 0.83333333 0.9245283 ]
0.8626834381551362
```

```
Difference between Accuracy score and cross validatio score is -  0.05089680875844045
*********************************************************************************
Best accuracy is 0.8641975308641975  on Random_state  122 for model  BernoulliNB()
*********************************************************************************
accuracy score of -> BernoulliNB()
0.8641975308641975
[[55  4]
 [ 7 15]]
              precision    recall  f1-score   support

           0       0.89      0.93      0.91        59
           1       0.79      0.68      0.73        22

    accuracy                           0.86        81
   macro avg       0.84      0.81      0.82        81
weighted avg       0.86      0.86      0.86        81

[0.64814815 0.68518519 0.85185185 0.64814815 0.83018868]
0.7327044025157232
Difference between Accuracy score and cross validatio score is -  0.13149312834847426
*********************************************************************************
Best accuracy is 0.9135802469135802  on Random_state  58 for model  SVC()
*********************************************************************************
accuracy score of -> SVC()
0.9135802469135802
[[57  2]
 [ 5 17]]
              precision    recall  f1-score   support

           0       0.92      0.97      0.94        59
           1       0.89      0.77      0.83        22

    accuracy                           0.91        81
   macro avg       0.91      0.87      0.89        81
weighted avg       0.91      0.91      0.91        81

[0.68518519 0.81481481 0.94444444 0.75925926 0.90566038]
```

```
[0.68518519 0.81481481 0.94444444 0.75925926 0.90566038]
0.8218728162124389
Difference between Accuracy score and cross validatio score is -  0.09170743070114129
*********************************************************************************
Best accuracy is 1.0  on Random_state  63 for model  DecisionTreeClassifier()
*********************************************************************************
accuracy score of -> DecisionTreeClassifier()
1.0
[[48  0]
 [ 0 33]]
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        48
           1       1.00      1.00      1.00        33

    accuracy                           1.00        81
   macro avg       1.00      1.00      1.00        81
weighted avg       1.00      1.00      1.00        81

[0.81481481 1.         1.         1.         1.        ]
0.962962962962963
Difference between Accuracy score and cross validatio score is -  0.03703703703703698
*********************************************************************************
Best accuracy is 0.9876543209876543  on Random_state  53 for model  RandomForestClassifier()
*********************************************************************************
accuracy score of -> RandomForestClassifier()
0.9753086419753086
[[58  0]
 [ 2 21]]
              precision    recall  f1-score   support

           0       0.97      1.00      0.98        58
           1       1.00      0.91      0.95        23

    accuracy                           0.98        81
   macro avg       0.98      0.96      0.97        81
weighted avg       0.98      0.98      0.97        81
```

```
 [ 2 21]]
              precision    recall  f1-score   support

           0       0.97      1.00      0.98        58
           1       1.00      0.91      0.95        23

    accuracy                           0.98        81
   macro avg       0.98      0.96      0.97        81
weighted avg       0.98      0.98      0.97        81

[0.66666667 0.98148148 1.         1.         1.        ]
0.9296296296296296
Difference between Accuracy score and cross validatio score is -  0.04567901234567906
*******************************************************************************
Best accuracy is 0.9629629629629629  on Random_state  28 for model  KNeighborsClassifier(n_neighbors=3)
*******************************************************************************
accuracy score of -> KNeighborsClassifier(n_neighbors=3)
0.9629629629629629
[[60  2]
 [ 1 18]]
              precision    recall  f1-score   support

           0       0.98      0.97      0.98        62
           1       0.90      0.95      0.92        19

    accuracy                           0.96        81
   macro avg       0.94      0.96      0.95        81
weighted avg       0.96      0.96      0.96        81

[0.7037037  0.96296296 1.         0.88888889 1.        ]
0.9111111111111111
Difference between Accuracy score and cross validatio score is -  0.05185185185185186
*******************************************************************************
```

Model Building conclusion:- Decision Tree Classifier is the best model with the accuracy of 100%.

Hyper Parameter Tuning

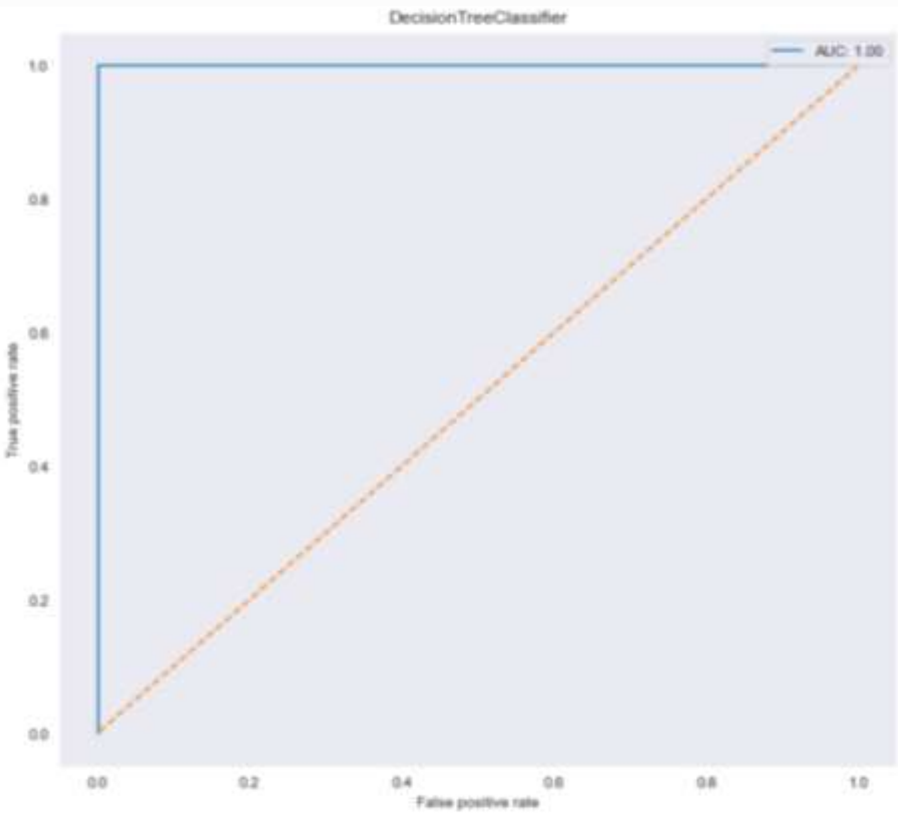
In [66]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=.30,random_state=63)
dtc=DecisionTreeClassifier()
dtc.fit(x_train,y_train)
dtc.score(x_train,y_train)
preddtc=dtc.predict(x_test)
print(accuracy_score(y_test,preddtc))
print(confusion_matrix(y_test,preddtc))
print(classification_report(y_test,preddtc))
```

```
1.0
[[48  0]
 [ 0 33]]
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        48
           1       1.00      1.00      1.00        33

    accuracy                           1.00        81
   macro avg       1.00      1.00      1.00        81
weighted avg       1.00      1.00      1.00        81
```

Hyper parameter tuning gives a 100% accuracy.

AUC ROC curve

In [67]:
```
from sklearn.metrics import roc_curve,auc
import matplotlib.pyplot as plt
fpr,tpr,thresholds=roc_curve(y_test,preddtc) # calculating fpr, tpr
rf_auc = auc(fpr, tpr) #Model Accuracy
plt.figure(figsize=(10,9)) #plotting the figure, size of 10*9
plt.plot(fpr, tpr, label = 'AUC: %0.2f' % rf_auc)
plt.plot([1,0],[1,0], linestyle = '--')
plt.legend(loc=0) #adding accuracy score at bottom right
plt.xlabel('False positive rate')
plt.ylabel('True positive rate')
plt.title('DecisionTreeClassifier')
plt.grid() #adding the grid
```



For this dataset, the Best model is DecisionTreeClassifier gives a 100% accuracy.

CONCLUSION:-

KEY FINDINGS AND CONCLUSIONS OF THE STUDY

Customer retention doesn't improve overnight. Customer retention is fickle when customer service is lacking. Few ways to improve customer retention can be as follows:

•**Easily navigable :** 90% customers agreed that the website should be easily navigable. Website navigation is a key to the success of any e-commerce website. It should be clean, clear and user-friendly. Online store should create easy-to-use navigation on website to make sure that, customers don't get confused while surfing the site.

•**User-friendly websites:** 87% customers agree with user friendly website interface. The online stores should invest heavily in creating user friendly apps and websites, so that the customers do not have to work around much and the overall shopping experience is smooth,& shoppers get what they want faster and without running into unnecessary complexity that can clog up the path to purchase. It should focus on the user experience by providing shopping categories, filters, and comparison capabilities. User-friendliness requires that your website works on all browsers and devices

•**Improve return & refund policy:** Return policies are an essential feature of any ecommerce website .90% customers agree that return and replacement policy help them making purchase decision. It is evident from the fact that people cannot actually try & touch the products, they are purchasing before it reaches home and they would want to return or replace in case of dissatisfaction. Online shopping websites should make strategies around easy return and replacement policy if they want to retain their customers. This is another trust-building feature of online selling. Which reassures buyers that if they are unhappy or just need a different size, the brand is there for them.

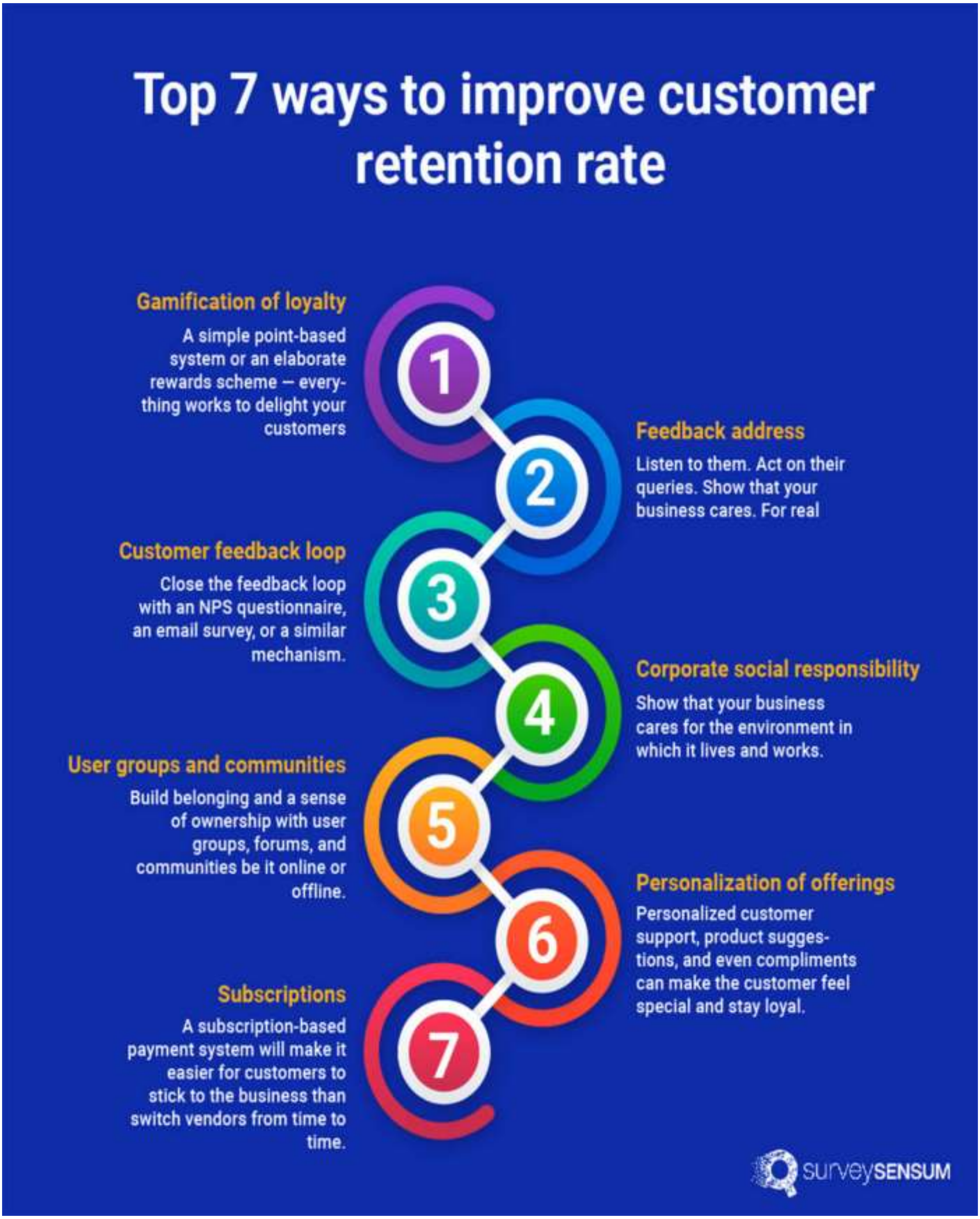
•**Privacy Policy:** Being able to guarantee the privacy of the customer:92% Customer agreed to this. Costumers are concerned about the unauthorized access to their data. Building trust with the customers is crucial for any e-commerce website. An e-commerce privacy policy statement makes business more transparent regarding how you collect, manage, and use data from site visitors.

•**Displaying quality Information on the website:** 90% customers agree all relevant information on listed products must be stated clearly. Content is one of the crucial challenges for any e-commerce website. it's simply not enough to just list a product name and image on a product page and expect the orders to roll in. Compiling a compelling array of product data, whether that's dimensions, MPNs or spec sheets can all help to convert customers better.

•**Responsiveness, availability of several communication channels (email, online rep, twitter, phone etc.):** 90% customers agreed to it. In case one channel is not available, customers can reach out to multiple channels which again is an important factor. Being able to communicate easily can make the difference in both their shopping experience as well as fulfill business goals of online store.

When customers are satisfied with a company or service, there is a high possibility that they will share their experience with other people Therefore it is crucial for E-commerce to take into account their customer satisfaction because this will retain customer loyalty as well as attract potential customers.

When customers are satisfied with a company or service, there is a high possibility that they will share their experience with other people Therefore it is crucial for E-commerce to take into account their customer satisfaction because this will retain customer loyalty as well as attract potential customers.


Top 7 ways to improve customer retention rate
Gamification of loyalty
A simple point-based system or an elaborate rewards scheme — everything works to delight your customers
1
2
Feedback address
Listen to them. Act on their queries. Show that your business cares. For real
Customer feedback loop
Close the feedback loop with an NPS questionnaire, an email survey, or a similar mechanism.
3
4
Corporate social responsibility
Show that your business cares for the environment in which it lives and works.
User groups and communities
Build belonging and a sense of ownership with user groups, forums, and communities be it online or offline.
5
6
Personalization of offerings
Personalized customer support, product suggestions, and even compliments can make the customer feel special and stay loyal.
Subscriptions
A subscription-based payment system will make it easier for customers to stick to the business than switch vendors from time to time.
7
surveySENSUM

LEARNING OUTCOMES OF THE STUDY IN RESPECT OF DATA SCIENCE

This project has demonstrated the importance of sampling effectively, modelling and predicting data.

Through different powerful tools of visualization, we were able to analyze and interpret different hidden insights about the data.

LIMITATIONS OF THIS WORK AND SCOPE FOR FUTURE WORK

Since the given dataset is very small. The analysis will be limited in scope. More data of such customers will definitely help increasing the probability of findings drawn above.