

```
In [1]: # Importing all the important libraries necessary
import pandas as pd
import numpy as np
```

```
In [2]: import matplotlib.pyplot as plt  
import seaborn as sns  
%matplotlib inline
```

```
In [3]: # loading the dataset into a dataframe  
# loading the 1st sheet names datasheet for analysis  
df = pd.read_excel(r'C:\Users\bhanu\DS0522\Internship\Project 1 - Customer retention\Cu
```

```
In [4]: # expanding the view to see all the columns in the given dataset  
pd.set_option('display.max_columns', 500)
```

In [88]: df.shape

```
Out[88]: (269, 71)
```

In [ ]:

```
In [6]: #Checking the name of the columns and replacing them with question numbers  
df.columns
```

'20 Complete information on listed seller and product being offered is important for purchase decision.',  
 '21 All relevant information on listed products must be stated clearly',  
 '22 Ease of navigation in website', '23 Loading and processing speed',  
 '24 User friendly Interface of the website',  
 '25 Convenient Payment methods',  
 '26 Trust that the online retail store will fulfill its part of the transaction at the stipulated time',  
 '27 Empathy (readiness to assist with queries) towards the customers',  
 '28 Being able to guarantee the privacy of the customer',  
 '29 Responsiveness, availability of several communication channels (email, online rep, twitter, phone etc.)',  
 '30 Online shopping gives monetary benefit and discounts',  
 '31 Enjoyment is derived from shopping online',  
 '32 Shopping online is convenient and flexible',  
 '33 Return and replacement policy of the e-tailer is important for purchase decision',  
 '34 Gaining access to loyalty programs is a benefit of shopping online',  
 '35 Displaying quality Information on the website improves satisfaction of customers',  
 '36 User derive satisfaction while shopping on a good quality website or application',  
 '37 Net Benefit derived from shopping online can lead to users satisfaction',  
 '38 User satisfaction cannot exist without trust',  
 '39 Offering a wide variety of listed product in several category',  
 '40 Provision of complete and relevant product information',  
 '41 Monetary savings',  
 '42 The Convenience of patronizing the online retailer',  
 '43 Shopping on the website gives you the sense of adventure',  
 '44 Shopping on your preferred e-tailer enhances your social status',  
 '45 You feel gratification shopping on your favorite e-tailer',  
 '46 Shopping on the website helps you fulfill certain roles',  
 '47 Getting value for money spent',  
 'From the following, tick any (or all) of the online retailers you have shopped from; ',  
 'Easy to use website or application',  
 'Visual appealing web-page layout', 'Wild variety of product on offer',  
 'Complete, relevant description information of products',  
 'Fast loading website speed of website and application',  
 'Reliability of the website or application',  
 'Quickness to complete purchase',  
 'Availability of several payment options', 'Speedy order delivery ',  
 'Privacy of customers' information',  
 'Security of customer financial information',  
 'Perceived Trustworthiness',  
 'Presence of online assistance through multi-channel',  
 '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 page loading time (promotion, sales period)',  
 'Limited mode of payment on most products (promotion, sales period)',  
 'Longer delivery period', 'Change in website/Application design',  
 'Frequent disruption when moving from one page to another',  
 'Website is as efficient as before',  
 'Which of the Indian online retailer would you recommend to a friend?'],  
 dtype='object')

In [7]: #Overview of the dataset  
df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 269 entries, 0 to 268
Data columns (total 71 columns):
 #   Column
```

		Non-Null Count	Dtype
0	1Gender of respondent	269 non-null	object
1	2 How old are you?	269 non-null	object
2	3 Which city do you shop online from?	269 non-null	object
3	4 What is the Pin Code of where you shop online from?	269 non-null	object
4	5 Since How Long You are Shopping Online ?	269 non-null	int64
5	6 How many times you have made an online purchase in the past 1 year?	269 non-null	object
6	7 How do you access the internet while shopping on-line?	269 non-null	object
7	8 Which device do you use to access the online shopping?	269 non-null	object
8	9 What is the screen size of your mobile device?	269 non-null	object
9	10 What is the operating system (OS) of your device?	269 non-null	object
10	11 What browser do you run on your device to access the website?	269 non-null	object
11	12 Which channel did you follow to arrive at your favorite online store for the first time?	269 non-null	object
12	13 After first visit, how do you reach the online retail store?	269 non-null	object
13	14 How much time do you explore the e- retail store before making a purchase decision?	269 non-null	object
14	15 What is your preferred payment Option?	269 non-null	object
15	16 How frequently do you abandon (selecting an items and leaving without making payment) your shopping cart?	269 non-null	object
16	17 Why did you abandon the "Bag", "Shopping Cart"?	269 non-null	object
17	18 The content on the website must be easy to read and understand	269 non-null	object
18	19 Information on similar product to the one highlighted is important for product comparison	269 non-null	object
19	20 Complete information on listed seller and product being offered is important for purchase decision.	269 non-null	object

20 21 All relevant information on listed products must be stated clearly  
21 22 Ease of navigation in website  
22 23 Loading and processing speed  
23 24 User friendly Interface of the website  
24 25 Convenient Payment methods  
25 26 Trust that the online retail store will fulfill its part of the transaction at the stipulated time  
26 27 Empathy (readiness to assist with queries) towards the customers  
27 28 Being able to guarantee the privacy of the customer  
28 29 Responsiveness, availability of several communication channels (email, online re p, twitter, phone etc.)  
29 30 Online shopping gives monetary benefit and discounts  
30 31 Enjoyment is derived from shopping online  
31 32 Shopping online is convenient and flexible  
32 33 Return and replacement policy of the e-tailer is important for purchase decision  
33 34 Gaining access to loyalty programs is a benefit of shopping online  
34 35 Displaying quality Information on the website improves satisfaction of customers  
35 36 User derive satisfaction while shopping on a good quality website or application  
36 37 Net Benefit derived from shopping online can lead to users satisfaction  
37 38 User satisfaction cannot exist without trust  
38 39 Offering a wide variety of listed product in several category  
39 40 Provision of complete and relevant product information  
40 41 Monetary savings  
41 42 The Convenience of patronizing the online retailer

42	43 Shopping on the website gives you the sense of adventure	269 non-null object
43	44 Shopping on your preferred e-tailer enhances your social status	269 non-null object
44	45 You feel gratification shopping on your favorite e-tailer	269 non-null object
45	46 Shopping on the website helps you fulfill certain roles	269 non-null object
46	47 Getting value for money spent	269 non-null object
47	From the following, tick any (or all) of the online retailers you have shopped from;	269 non-null object
48	Easy to use website or application	269 non-null object
49	Visual appealing web-page layout	269 non-null object
50	Wide variety of product on offer	269 non-null object
51	Complete, relevant description information of products	269 non-null object
52	Fast loading website speed of website and application	269 non-null object
53	Reliability of the website or application	269 non-null object
54	Quickness to complete purchase	269 non-null object
55	Availability of several payment options	269 non-null object
56	Speedy order delivery	269 non-null object
57	Privacy of customers' information	269 non-null object
58	Security of customer financial information	269 non-null object
59	Perceived Trustworthiness	269 non-null object
60	Presence of online assistance through multi-channel	269 non-null object
61	Longer time to get logged in (promotion, sales period)	269 non-null object
62	Longer time in displaying graphics and photos (promotion, sales period)	269 non-null object
63	Late declaration of price (promotion, sales period)	269 non-null object

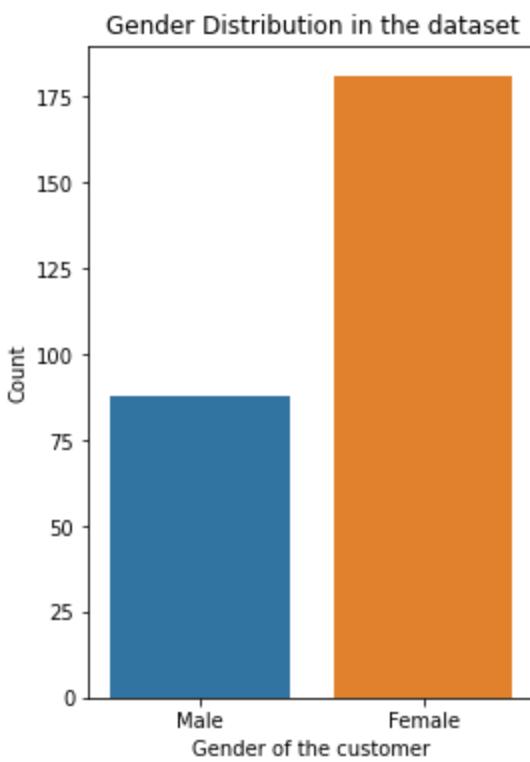
```
64 Longer page loading time (promotion, sales period)          269 non-null    object
65 Limited mode of payment on most products (promotion, sales period)
66 Longer delivery period                                     269 non-null    object
67 Change in website/Application design                      269 non-null    object
68 Frequent disruption when moving from one page to another
69 Website is as efficient as before                         269 non-null    object
70 Which of the Indian online retailer would you recommend to a friend?      269 non-null    object
dtypes: int64(1), object(70)
memory usage: 149.3+ KB
```

```
In [8]: df.columns=np.arange(0,71)
```

```
In [9]: # Checking the number of male and female respondents
gender_dist = df[0]
gender_dist.value_counts()
```

```
Out[9]: Female    181
Male      88
Name: 0, dtype: int64
```

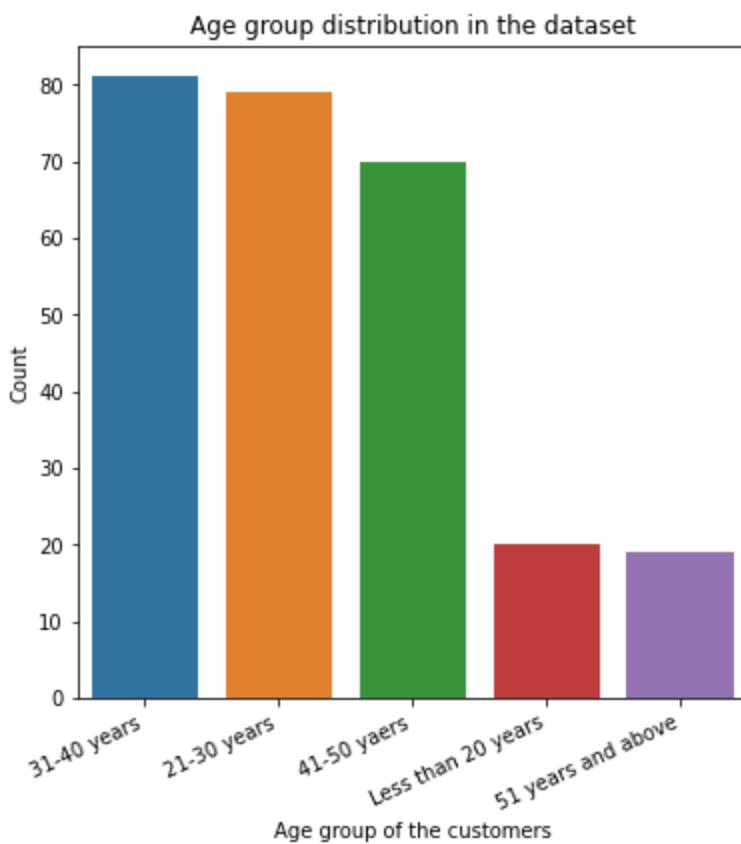
```
In [10]: plt.figure(figsize=(4,6))
plt.title('Gender Distribution in the dataset')
sns.countplot(x=df[0])
plt.xlabel('Gender of the customer')
plt.ylabel('Count')
plt.show()
```



```
In [11]: # Checking the age groups of the customers  
age_dist = df[1]  
age_dist.value_counts()
```

```
Out[11]: 31-40 years      81  
21-30 years      79  
41-50 years      70  
Less than 20 years    20  
51 years and above   19  
Name: 1, dtype: int64
```

```
In [12]: # Visualizing the distribution of various age groups in the dataset  
plt.figure(figsize=(6, 6))  
plt.title('Age group distribution in the dataset')  
sns.countplot(x=df[1])  
plt.xlabel('Age group of the customers')  
plt.ylabel('Count')  
plt.xticks(rotation=25, ha='right')  
plt.show()
```



In [ ]:

In [13]:

```
#Checking the various location from where the purchases were made
loc = df[2]
loc.value_counts()
```

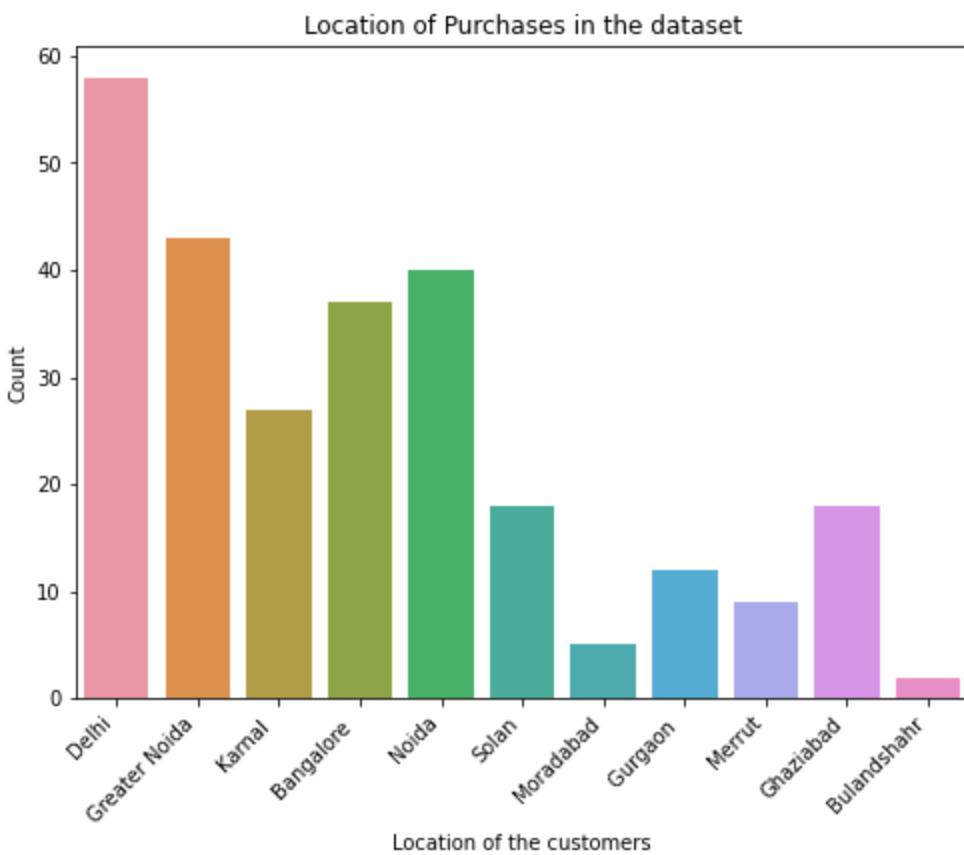
Out[13]:

Delhi	58
Greater Noida	43
Noida	40
Bangalore	37
Karnal	27
Solan	18
Ghaziabad	18
Gurgaon	12
Merrut	9
Moradabad	5
Bulandshahr	2

Name: 2, dtype: int64

In [14]:

```
# Visualizing the distribution of location of Purchases in the dataset
plt.figure(figsize=(8,6))
plt.title('Location of Purchases in the dataset')
sns.countplot(x=df[2])
plt.xlabel('Location of the customers')
plt.ylabel('Count')
plt.xticks(rotation=45, ha='right')
plt.show()
```



In [ ]:

In [15]:

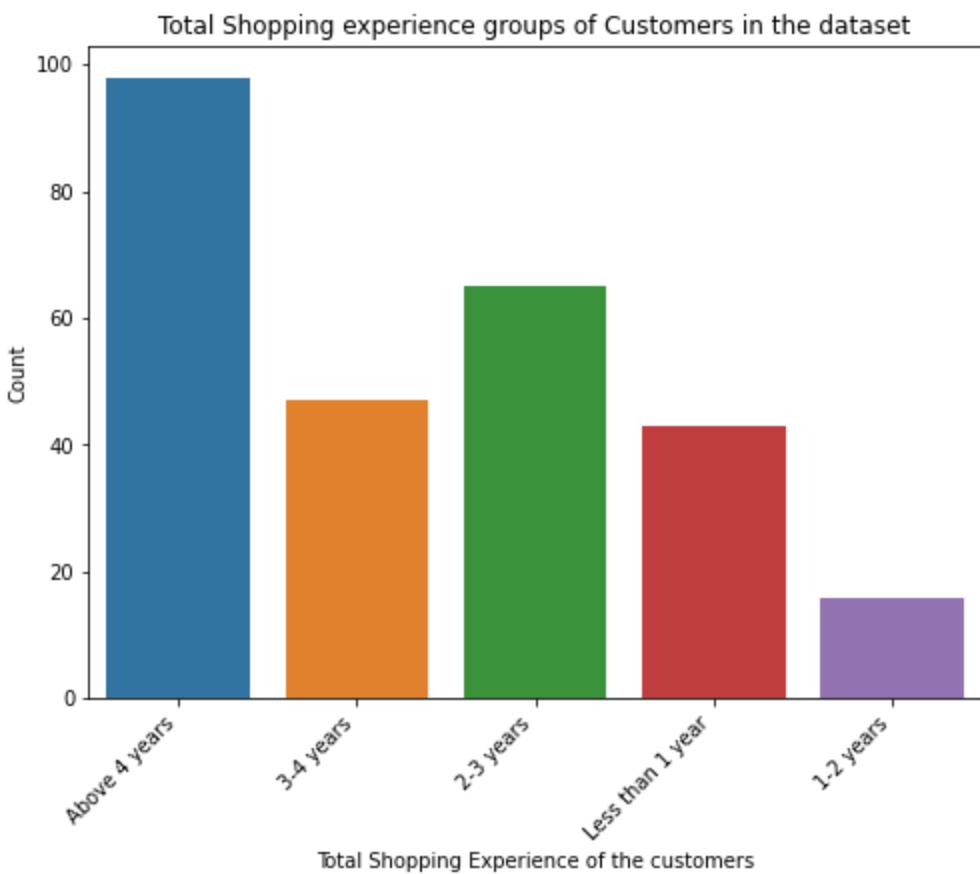
```
# Checking teh total shopping experience of the customers
tot_shop_exp = df[4]
tot_shop_exp.value_counts()
```

Out[15]:

Above 4 years	98
2-3 years	65
3-4 years	47
Less than 1 year	43
1-2 years	16
Name: 4, dtype: int64	

In [16]:

```
# Visualizing the distribution of Total Shopping experience of the customers in the dataset
plt.figure(figsize=(8,6))
plt.title('Total Shopping experience groups of Customers in the dataset')
sns.countplot(x=df[4])
plt.xlabel('Total Shopping Experience of the customers')
plt.ylabel('Count')
plt.xticks(rotation=45, ha='right')
plt.show()
```



In [ ]:

In [17]:

```
# Checking the number of purchases made by the customers in the past 1 year
prev_1yr_shopping = df[5]
prev_1yr_shopping.value_counts()
```

Out[17]:

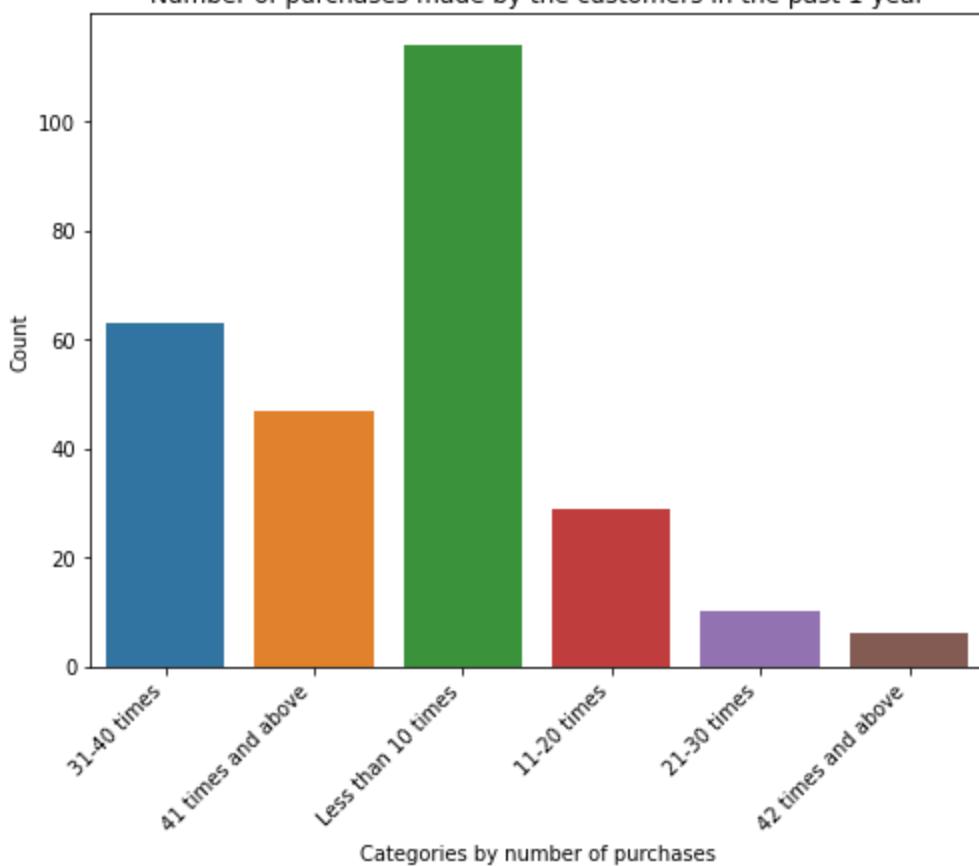
Less than 10 times	114
31-40 times	63
41 times and above	47
11-20 times	29
21-30 times	10
42 times and above	6

Name: 5, dtype: int64

In [18]:

```
# Visualizing the number of purchases made by the customers in the past 1 year
plt.figure(figsize=(8,6))
plt.title('Number of purchases made by the customers in the past 1 year')
sns.countplot(x=df[5])
plt.xlabel('Categories by number of purchases')
plt.ylabel('Count')
plt.xticks(rotation=45, ha='right')
plt.show()
```

Number of purchases made by the customers in the past 1 year



In [19]:

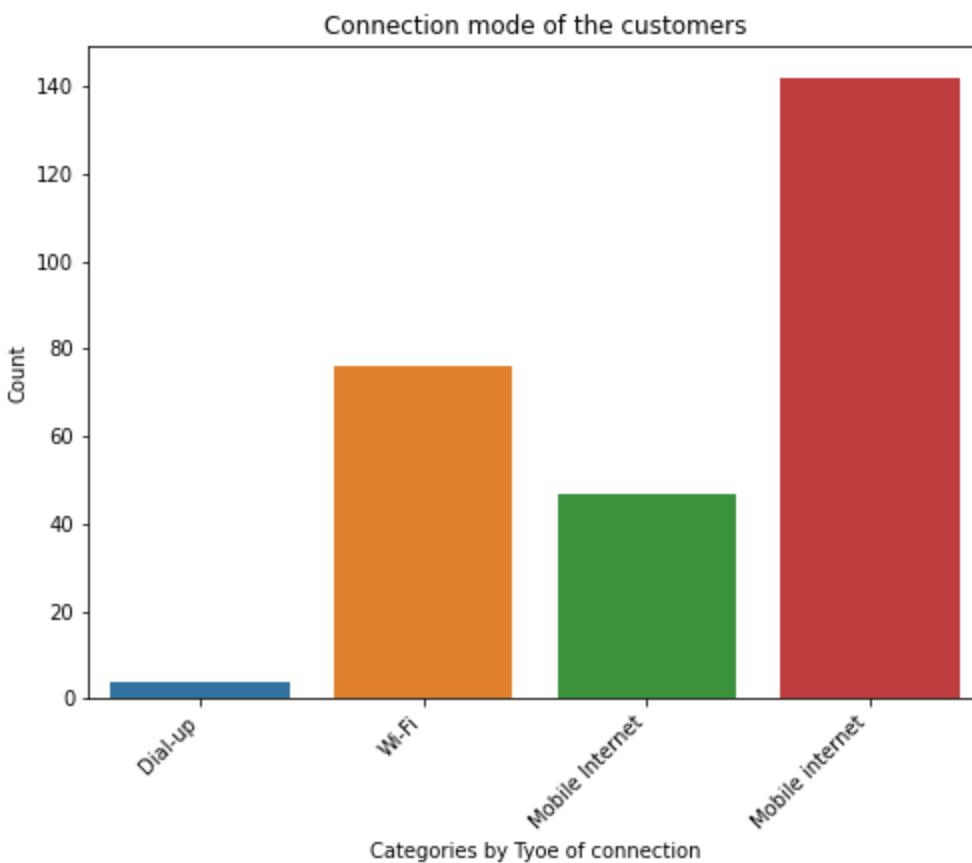
```
# Type of connection
connection = df[6]
connection.value_counts()
```

Out[19]:

```
Mobile internet    142
Wi-Fi             76
Mobile Internet   47
Dial-up            4
Name: 6, dtype: int64
```

In [20]:

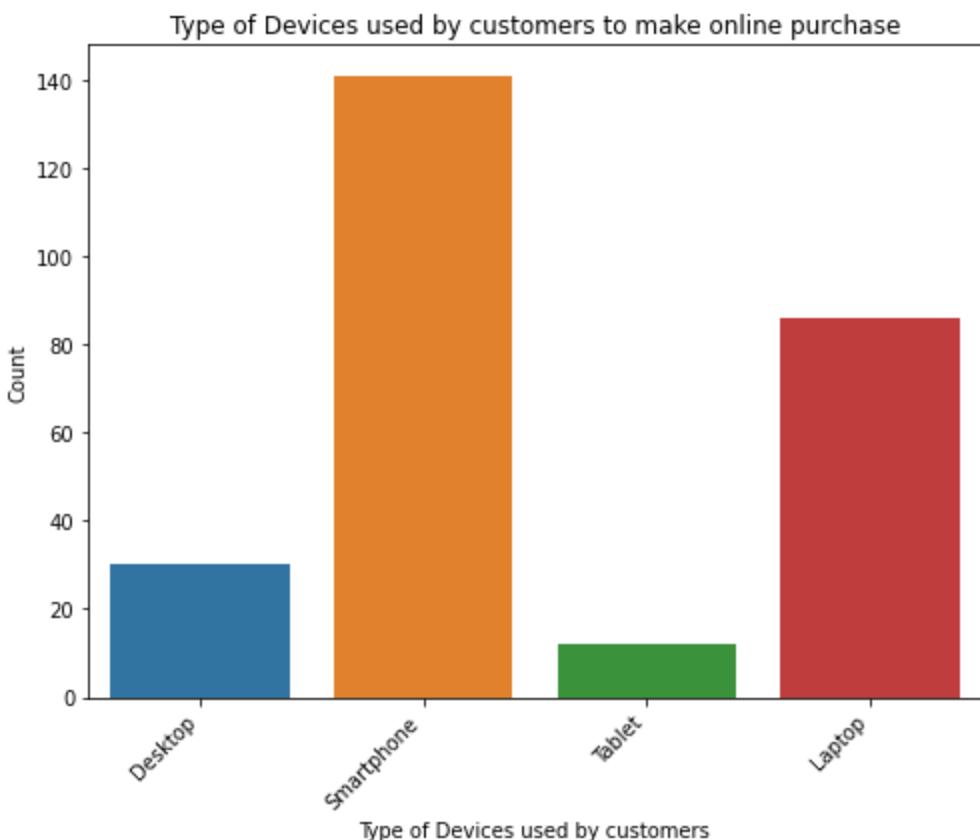
```
# Visualizing the Connection mode of the customers
plt.figure(figsize=(8,6))
plt.title('Connection mode of the customers')
sns.countplot(x=df[6])
plt.xlabel('Categories by Tyoe of connection')
plt.ylabel('Count')
plt.xticks(rotation=45, ha='right')
plt.show()
```



```
In [21]: # Checking the Type of Devices used by customers to make online purchase  
device = df[7]  
device.value_counts()
```

```
Out[21]: Smartphone    141  
Laptop        86  
Desktop       30  
Tablet        12  
Name: 7, dtype: int64
```

```
In [22]: # Visualizing the Type of Devices used by customers to make online purchase  
plt.figure(figsize=(8,6))  
plt.title('Type of Devices used by customers to make online purchase')  
sns.countplot(x=df[7])  
plt.xlabel('Type of Devices used by customers')  
plt.ylabel('Count')  
plt.xticks(rotation=45, ha='right')  
plt.show()
```



In [23]:

```
# Checking the Screen Size of Devices used by customers to make online purchase
device_size = df[8]
device_size.value_counts()
```

Out[23]:

Others	134
5.5 inches	99
4.7 inches	29
5 inches	7

Name: 8, dtype: int64

In [24]:

```
# Checking the OS of Devices used by customers to make online purchase
device_os = df[9]
device_os.value_counts()
```

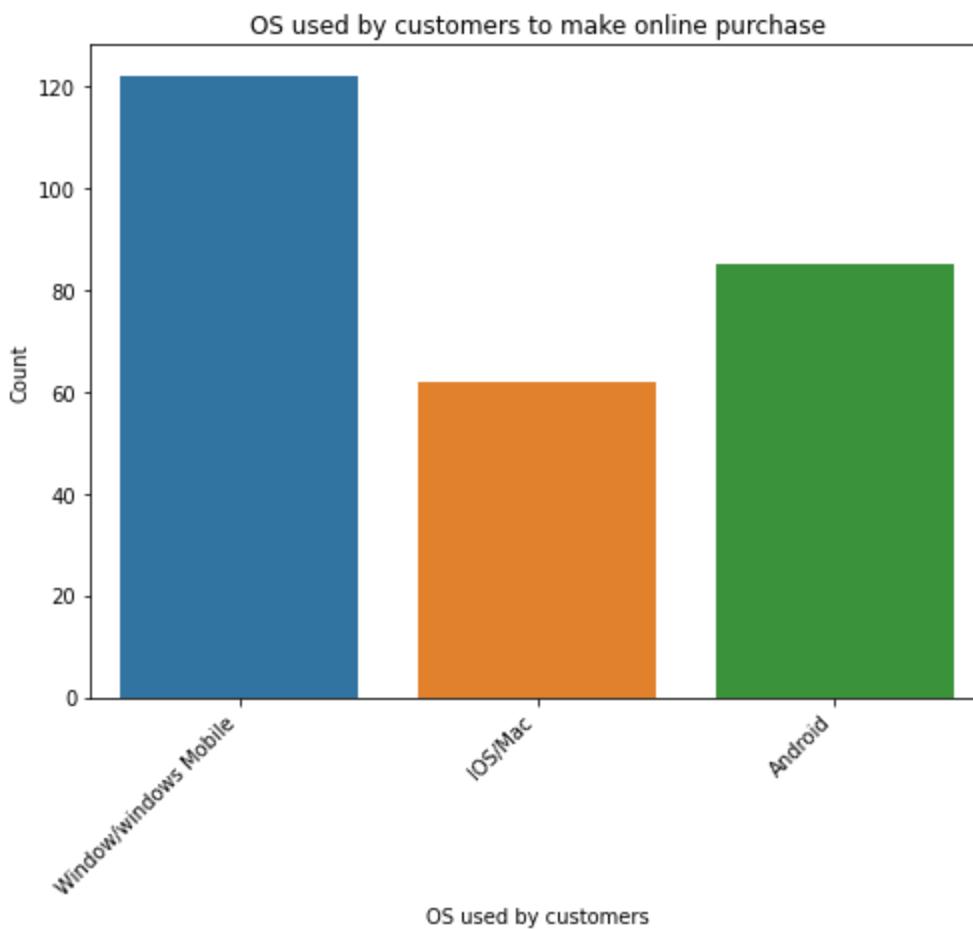
Out[24]:

Window/windows Mobile	122
Android	85
IOS/Mac	62

Name: 9, dtype: int64

In [25]:

```
plt.figure(figsize=(8,6))
plt.title('OS used by customers to make online purchase')
sns.countplot(x=df[9])
plt.xlabel('OS used by customers')
plt.ylabel('Count')
plt.xticks(rotation=45, ha='right')
plt.show()
```

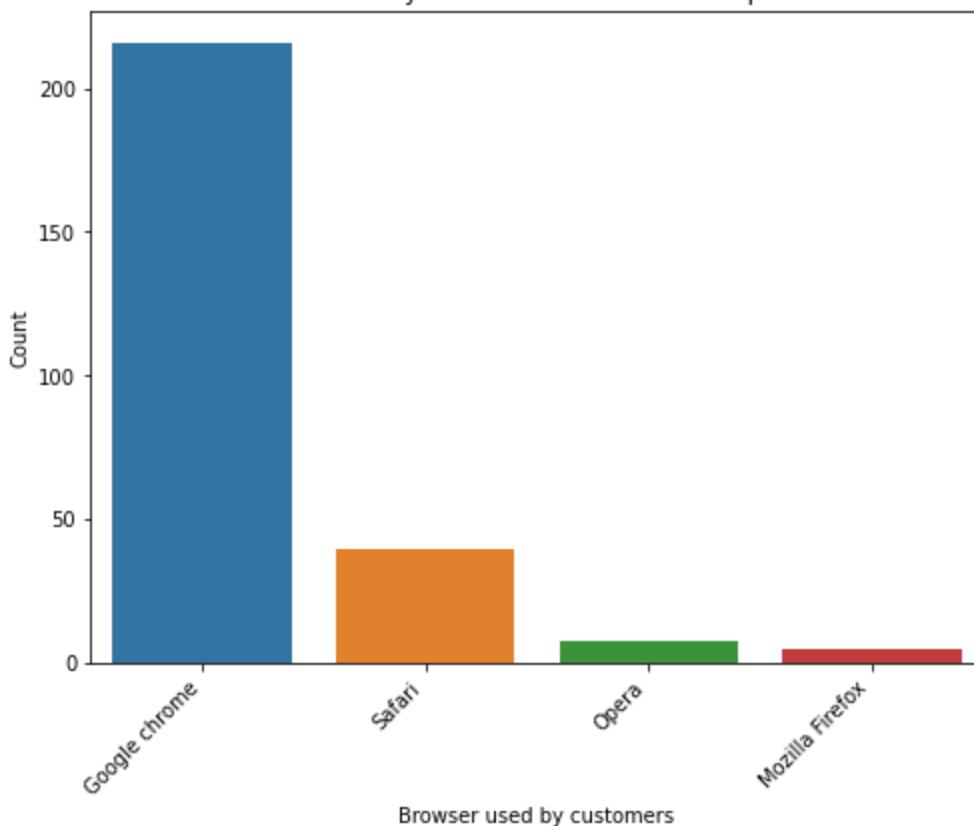


```
In [26]: # Checking the Browser used by customers to make online purchase
browser = df[10]
browser.value_counts()
```

```
Out[26]: Google chrome      216
Safari                  40
Opera                   8
Mozilla Firefox        5
Name: 10, dtype: int64
```

```
In [27]: # Visualizing the Browser used by customers to make online purchase
plt.figure(figsize=(8,6))
plt.title('Browser used by customers to make online purchase')
sns.countplot(x=df[10])
plt.xlabel('Browser used by customers')
plt.ylabel('Count')
plt.xticks(rotation=45, ha='right')
plt.show()
```

Browser used by customers to make online purchase



```
In [28]: # Checking the time spent by the customers before making the purchase  
time_bef_pur = df[13]  
time_bef_pur.value_counts()
```

```
Out[28]: more than 15 mins    123  
6-10 mins          71  
11-15 mins          46  
Less than 1 min      15  
1-5 mins            14  
Name: 13, dtype: int64
```

```
In [29]: # Preferred payment method of teh customers  
pref_payment = df[14]  
pref_payment.value_counts()
```

```
Out[29]: Credit/Debit cards        148  
Cash on delivery (CoD)           76  
E-wallets (Paytm, Freecharge etc.) 45  
Name: 14, dtype: int64
```

```
In [30]: # Checking effect of easy UI while purchasing  
easy_ui = df[17]  
easy_ui.value_counts()
```

```
Out[30]: Strongly agree (5)       164  
Agree (4)                      80  
Strongly disagree (1)           18  
Indifferent (3)                 7  
Name: 17, dtype: int64
```

```
In [31]: #Checking the importance of recommended similiar product info  
comp_prod_info = df[18]  
comp_prod_info.value_counts()
```

```
Out[31]: Strongly agree (5)      116  
          Agree (4)             92  
          Indifferent (3)       43  
          Dis-agree (2)          18  
          Name: 18, dtype: int64
```

```
In [32]: #Checking the importance of seller and complete product info  
comp_seller_prod_info = df[19]  
comp_seller_prod_info.value_counts()
```

```
Out[32]: Agree (4)              101  
          Strongly agree (5)    87  
          Indifferent (3)       52  
          Dis-agree (2)          18  
          Strongly disagree (1) 11  
          Name: 19, dtype: int64
```

```
In [33]: #Checking the speed of the website/application response  
sys_res = df[22]  
sys_res.value_counts()
```

```
Out[33]: Strongly agree (5)      115  
          Agree (4)            112  
          Dis-agree (2)          18  
          Strongly disagree (1) 12  
          Indifferent (3)        12  
          Name: 22, dtype: int64
```

```
In [34]: #Easy Payments  
conv_pay = df[24]  
conv_pay.value_counts()
```

```
Out[34]: Strongly agree (5)      159  
          Agree (4)            80  
          Dis-agree (2)          30  
          Name: 24, dtype: int64
```

```
In [35]: #Checking the privacy  
privacy = df[27]  
privacy.value_counts()
```

```
Out[35]: Strongly agree (5)      185  
          Agree (4)            58  
          Indifferent (3)        26  
          Name: 27, dtype: int64
```

```
In [36]: #Checking the service option of the seller  
service = df[28]  
service.value_counts()
```

```
Out[36]: Strongly agree (5)      149  
          Agree (4)            94  
          Indifferent (3)        15  
          Strongly disagree (1) 11  
          Name: 28, dtype: int64
```

```
In [37]: #Shopping Online is convenient and flexible  
online_shop_flex = df[31]  
online_shop_flex.value_counts()
```

```
Out[37]: Strongly agree (5)      146
```

```
Agree (4)          78  
indifferent (3)   33  
Dis-agree (2)     12  
Name: 31, dtype: int64
```

```
In [38]: # Quality of the website  
website_quatity = df[35]  
website_quatity.value_counts()
```

```
Out[38]: Strongly agree (5)    175  
Agree (4)           86  
Dis-agree (2)       8  
Name: 35, dtype: int64
```

```
In [39]: # Checking the effect of wide variety of products in each category  
variety = df[38]  
variety.value_counts()
```

```
Out[39]: Strongly agree (5)    111  
Agree (4)           94  
indifferent (3)    57  
Dis-agree (2)       7  
Name: 38, dtype: int64
```

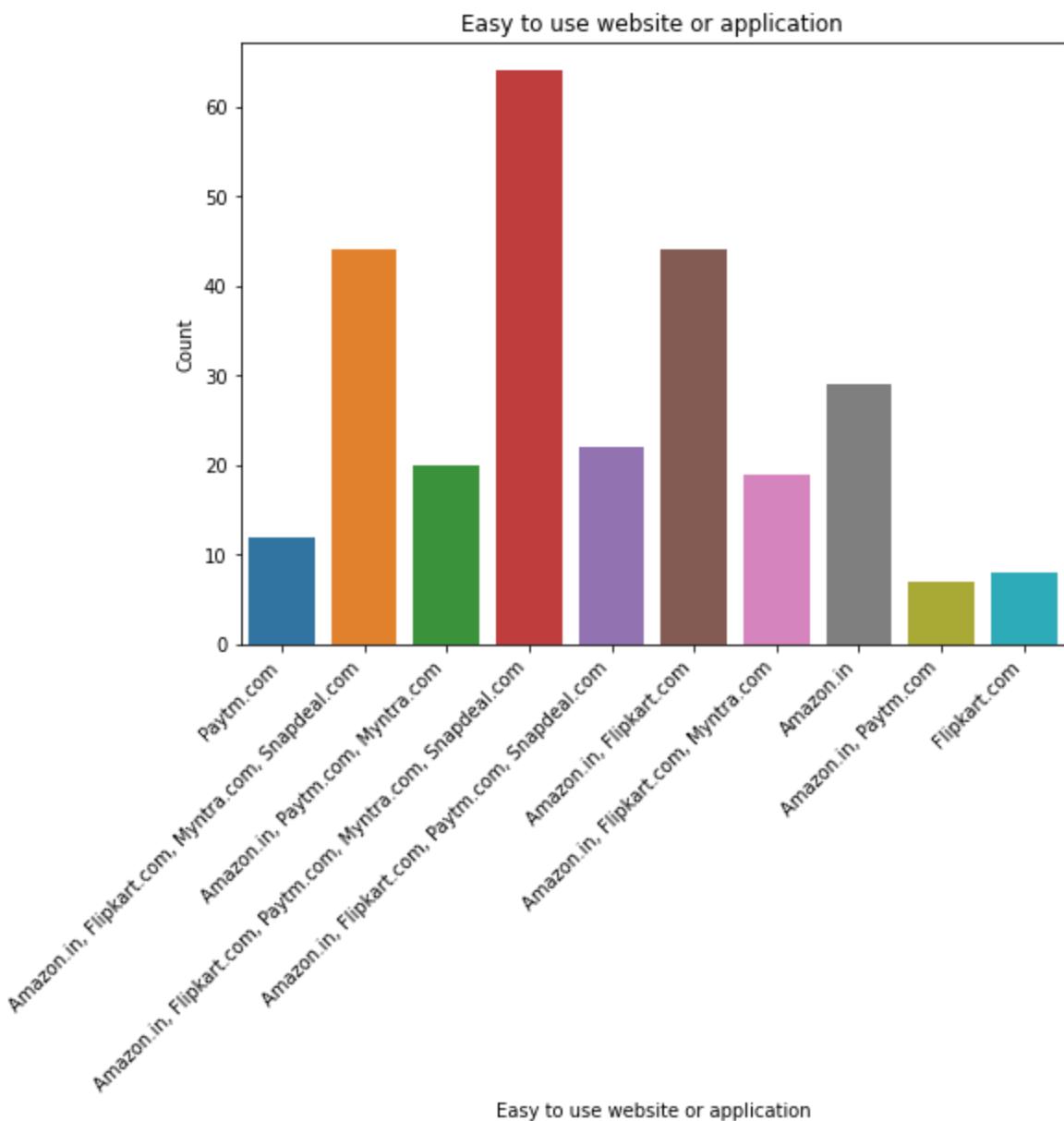
```
In [40]: value_for_money = df[46]  
value_for_money.value_counts()
```

```
Out[40]: Agree (4)           149  
Strongly agree (5)        82  
indifferent (3)          38  
Name: 46, dtype: int64
```

```
In [41]: #checking the column Easy to use website or application  
  
easy_website = df[48]  
easy_website.value_counts()
```

```
Out[41]: Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com      64  
Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com                      44  
Amazon.in, Flipkart.com                                         44  
Amazon.in                                                 29  
Amazon.in, Flipkart.com, Paytm.com, Snapdeal.com                  22  
Amazon.in, Paytm.com, Myntra.com                                    20  
Amazon.in, Flipkart.com, Myntra.com                                19  
Paytm.com                                              12  
Flipkart.com                                            8  
Amazon.in, Paytm.com                                         7  
Name: 48, dtype: int64
```

```
In [42]: plt.figure(figsize=(8,6))  
plt.title('Easy to use website or application')  
sns.countplot(x=df[48])  
plt.xlabel('Easy to use website or application')  
plt.ylabel('Count')  
plt.xticks(rotation=45, ha='right')  
plt.show()
```

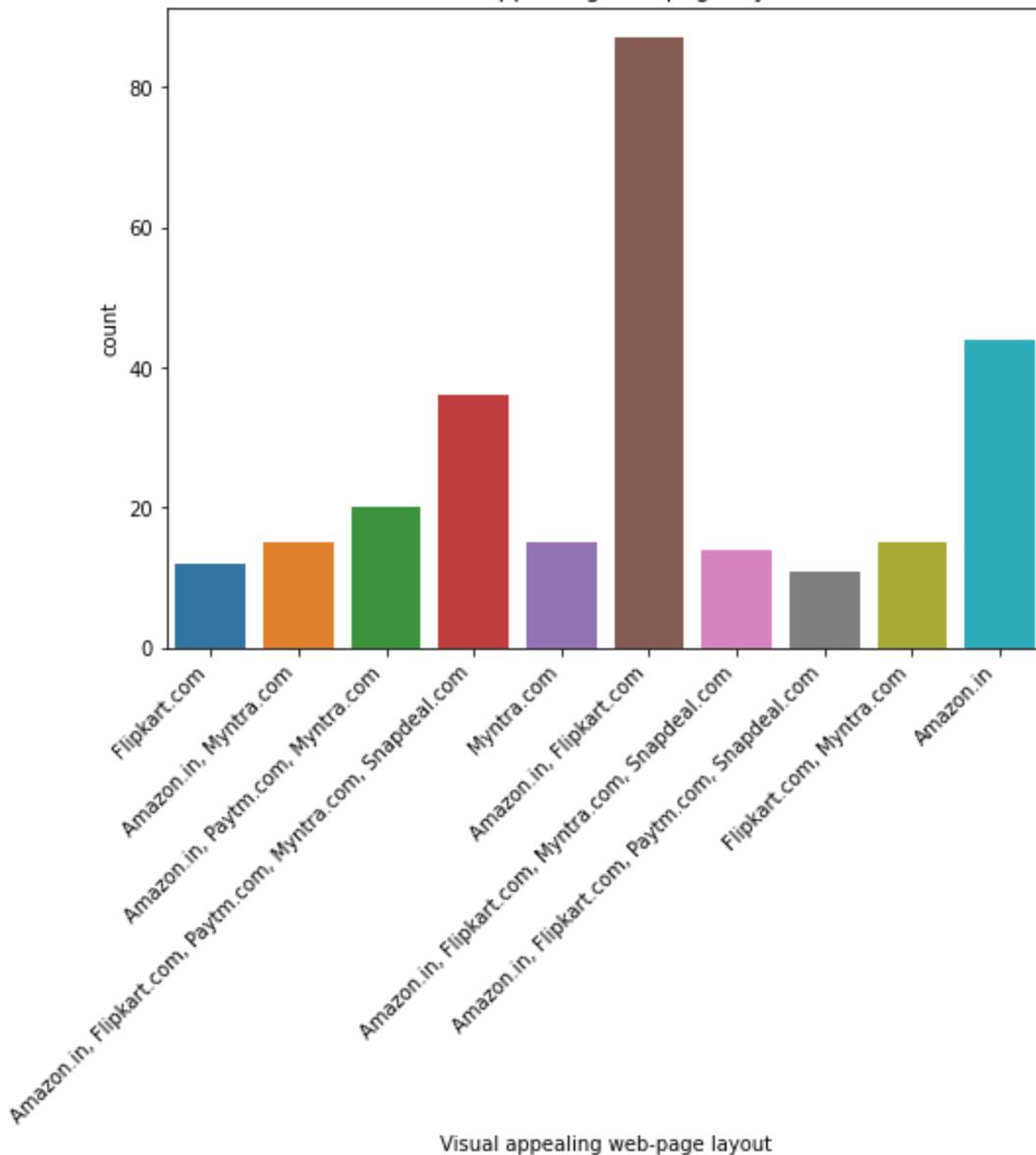


```
In [43]: #checking the column Visual appealing web-page layout
web_layout = df[49]
web_layout.value_counts()
```

```
Out[43]: Amazon.in, Flipkart.com          87
          Amazon.in                      44
          Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com 36
          Amazon.in, Paytm.com, Myntra.com           20
          Amazon.in, Myntra.com                  15
          Myntra.com                         15
          Flipkart.com, Myntra.com             15
          Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com        14
          Flipkart.com                        12
          Amazon.in, Flipkart.com, Paytm.com, Snapdeal.com         11
          Name: 49, dtype: int64
```

```
In [44]: plt.subplots(figsize=(8, 6))
sns.countplot(x=df[49])
plt.title("Visual appealing web-page layout")
plt.xticks(rotation=45, ha='right')
plt.xlabel('Visual appealing web-page layout')
plt.ylabel("count")
plt.show()
```

### Visual appealing web-page layout



In [45]:

```
# Checking the reliability of the website / application
reliability = df[53]
reliability.value_counts()
```

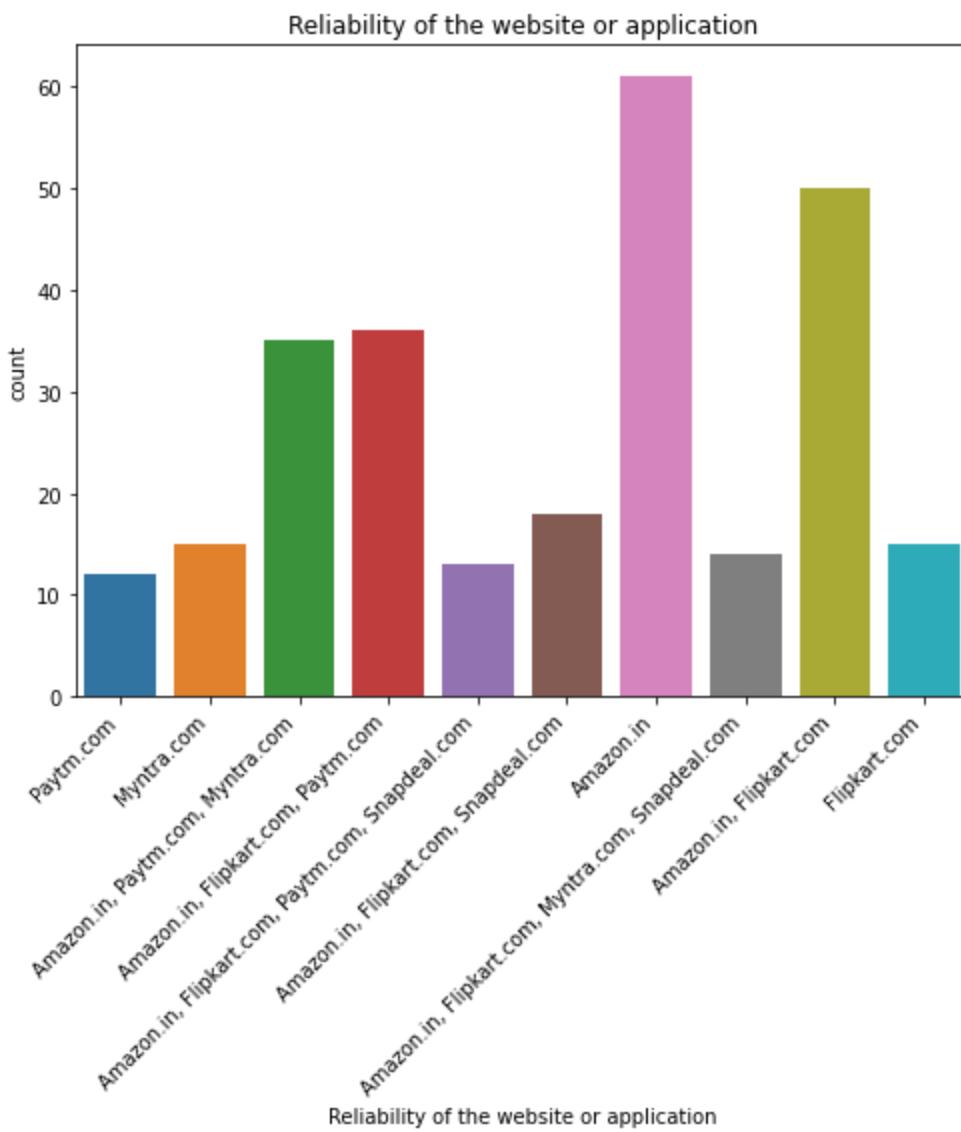
Out[45]:

Amazon.in	61
Amazon.in, Flipkart.com	50
Amazon.in, Flipkart.com, Paytm.com	36
Amazon.in, Paytm.com, Myntra.com	35
Amazon.in, Flipkart.com, Snapdeal.com	18
Myntra.com	15
Flipkart.com	15
Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com	14
Amazon.in, Flipkart.com, Paytm.com, Snapdeal.com	13
Paytm.com	12

Name: 53, dtype: int64

In [46]:

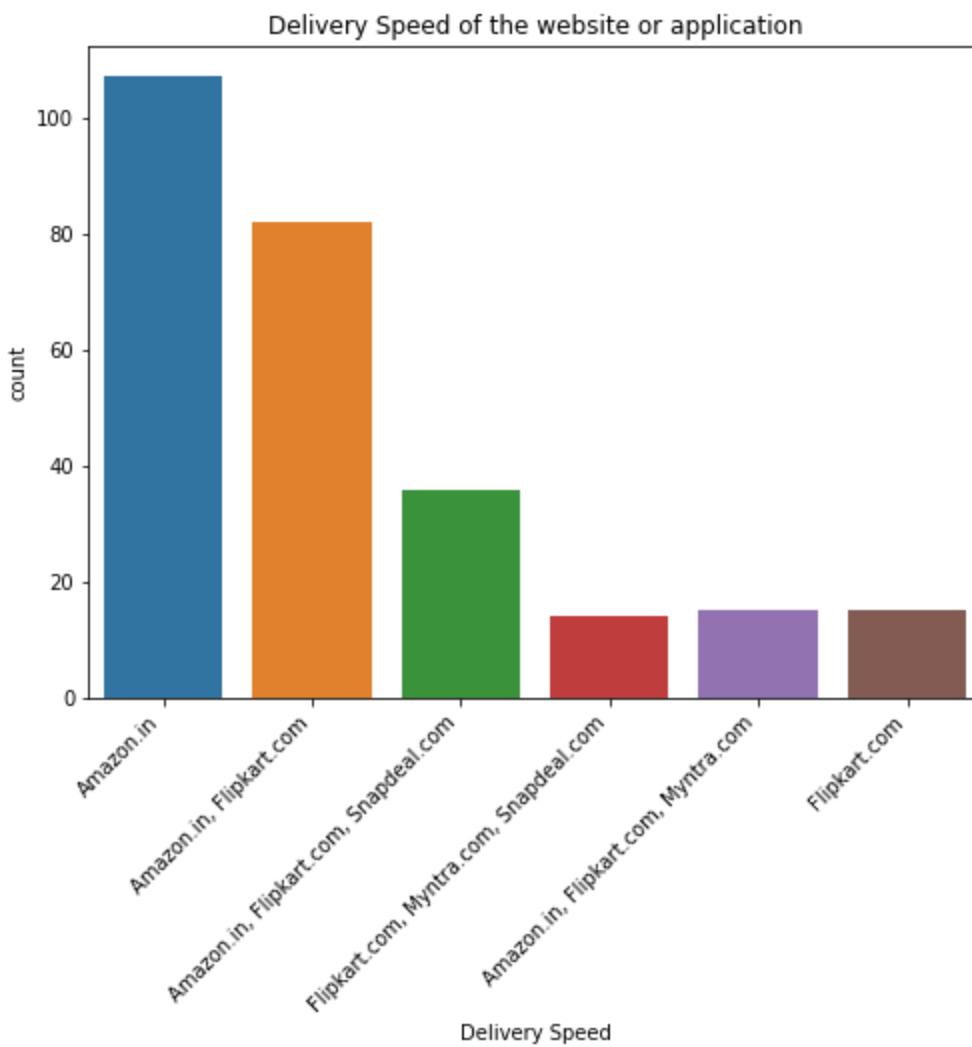
```
plt.subplots(figsize=(8, 6))
sns.countplot(x=df[53])
plt.title("Reliability of the website or application")
plt.xlabel('Reliability of the website or application')
plt.xticks(rotation=45, ha='right')
plt.ylabel("count")
plt.show()
```



```
In [47]: # Checking the speed of delivery of the website
delivery_speed = df[56]
delivery_speed.value_counts()
```

```
Out[47]: Amazon.in                               107
Amazon.in, Flipkart.com                         82
Amazon.in, Flipkart.com, Snapdeal.com           36
Amazon.in, Flipkart.com, Myntra.com              15
Flipkart.com                                    15
Flipkart.com, Myntra.com, Snapdeal.com          14
Name: 56, dtype: int64
```

```
In [48]: plt.subplots(figsize=(8, 6))
sns.countplot(x=df[56])
plt.title("Delivery Speed of the website or application")
plt.xlabel('Delivery Speed')
plt.xticks(rotation=45, ha='right')
plt.ylabel("count")
plt.show()
```



```
In [49]: df.head(1)
```

```
Out[49]: 0 1 2 3 4 5 6 7 8 9 10 11 12
          Male 31-40 years Delhi 110009 Above 4 years Dial-up Desktop Others Window/windows Mobile Google chrome Search Engine Search Engine 6-mi
```

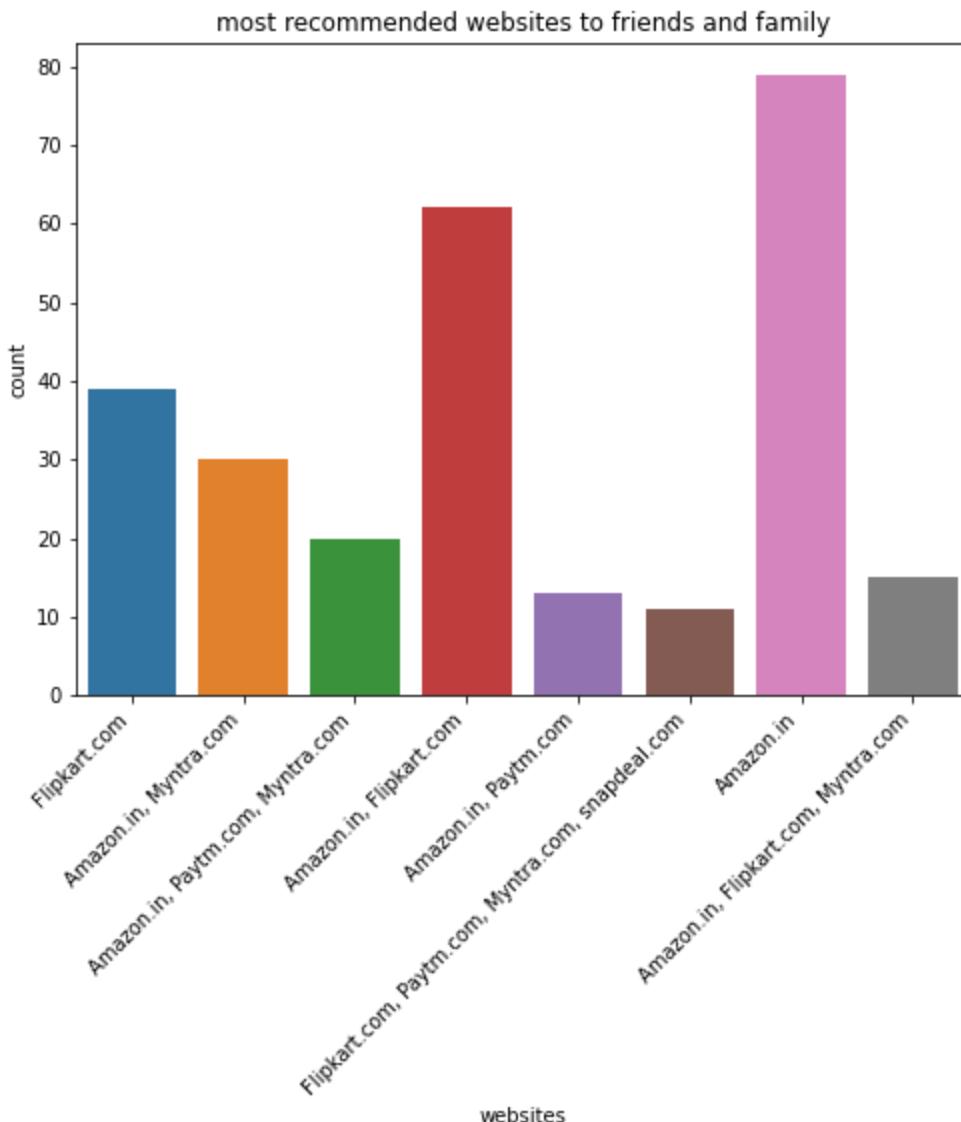
0	Male	31-40 years	Delhi	110009	Above 4 years	Dial-up	Desktop	Others	Window/windows	Mobile	Google chrome	Search Engine	Search Engine	6-mi
---	------	-------------	-------	--------	---------------	---------	---------	--------	----------------	--------	---------------	---------------	---------------	------

```
In [50]: # Checking the most recommended website
recommendation = df[70]
recommendation.value_counts()
```

```
Out[50]: Amazon.in 79
         Amazon.in, Flipkart.com 62
         Flipkart.com 39
         Amazon.in, Myntra.com 30
         Amazon.in, Paytm.com, Myntra.com 20
         Amazon.in, Flipkart.com, Myntra.com 15
         Amazon.in, Paytm.com 13
         Flipkart.com, Paytm.com, Myntra.com, snapdeal.com 11
Name: 70, dtype: int64
```

```
In [51]: plt.subplots(figsize=(8,6))
sns.countplot(x=df[70])
plt.title("most recommended websites to friends and family")
plt.xlabel('websites')
plt.xticks(rotation=45, ha='right')
```

```
plt.ylabel("count")
plt.show()
```



## Creating Categories of Satisfaction and Loyalty

In [52]:

```
df[44]
```

Out[52]:

```
0      Strongly agree (5)
1      Strongly agree (5)
2          indifferent (3)
3            Agree (4)
4      Strongly agree (5)

        ...
264      indifferent (3)
265      Strongly agree (5)
266            Agree (4)
267      indifferent (3)
268      Strongly agree (5)
Name: 44, Length: 269, dtype: object
```

In [53]:

```
df[44] = [i.replace('Strongly agree (5)', 'loyal') for i in df[44]]
df[44] = [i.replace('Agree (4)', 'loyal') for i in df[44]]

df[44] = [i.replace('Strongly disagree (1)', 'disloyal') for i in df[44]]
df[44] = [i.replace('Disagree (2)', 'disloyal') for i in df[44]]
```

```
df[44] = [i.replace('indifferent (3)', 'possible churners') for i in df[44]]
```

```
In [54]: df[44].value_counts()
```

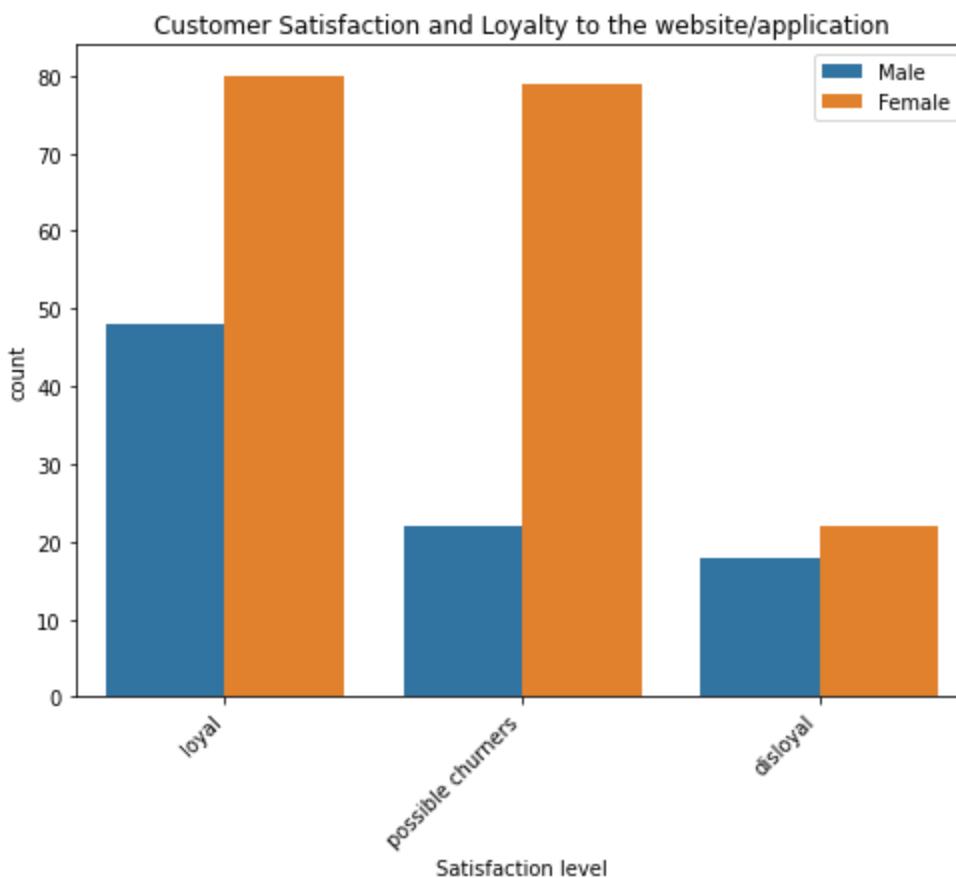
```
Out[54]: loyal           128  
possible churners    101  
disloyal            40  
Name: 44, dtype: int64
```

```
In [55]: plt.subplots(figsize=(8, 6))  
sns.countplot(x=df[44])  
plt.title("Customer Satisfaction and Loyalty to the website/application")  
plt.xlabel('Satisfaction level')  
plt.xticks(rotation=45, ha='right')  
plt.ylabel("count")  
plt.show()
```



```
In [56]: plt.subplots(figsize=(8, 6))
```

```
sns.countplot(x=df[44], hue=df[0])  
plt.title("Customer Satisfaction and Loyalty to the website/application")  
plt.xlabel('Satisfaction level')  
plt.xticks(rotation=45, ha='right')  
plt.ylabel("count")  
plt.show()
```

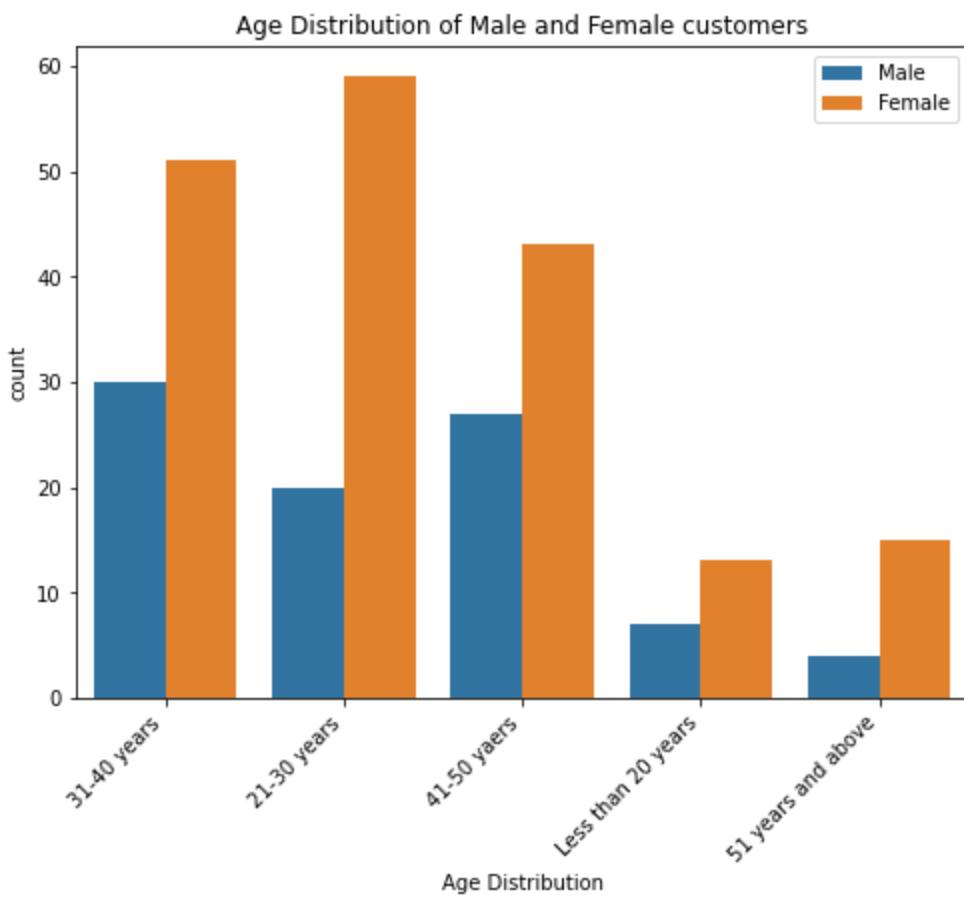


```
In [57]: df[2]
```

```
Out[57]: 0           Delhi
1           Delhi
2    Greater Noida
3       Karnal
4      Bangalore
...
264        Solan
265    Ghaziabad
266    Bangalore
267        Solan
268    Ghaziabad
Name: 2, Length: 269, dtype: object
```

```
In [58]: #Age Distribution of Male and Female customers
```

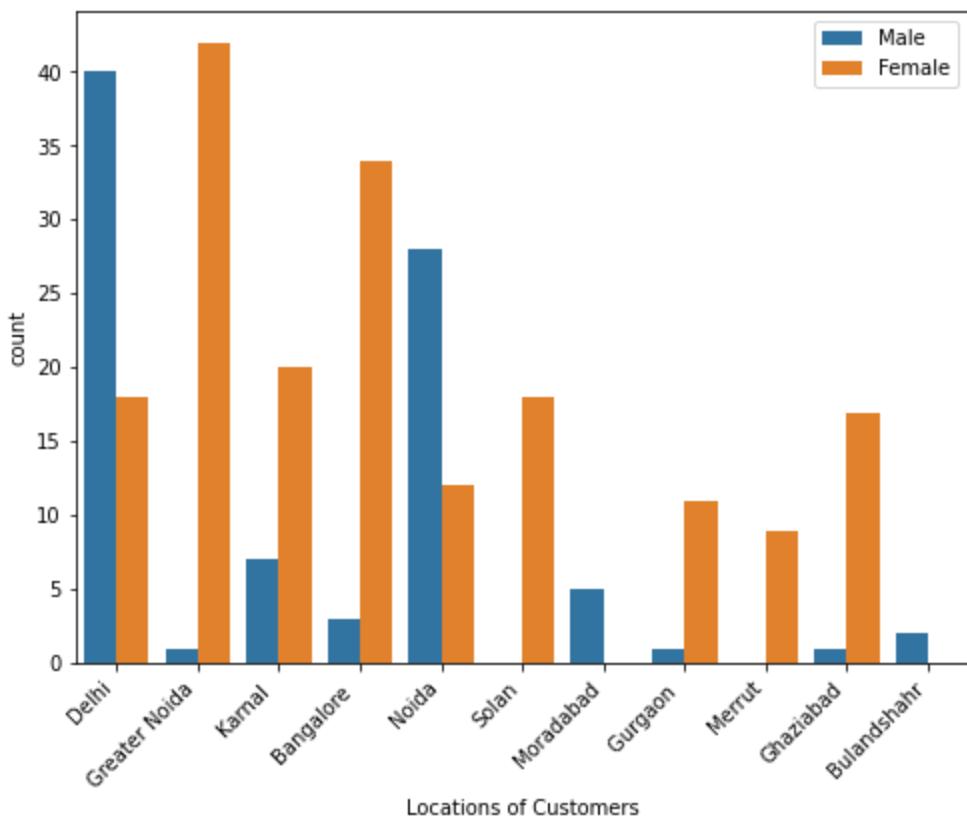
```
plt.subplots(figsize=(8,6))
sns.countplot(x=df[1],hue=df[0])
plt.title("Age Distribution of Male and Female customers")
plt.xlabel('Age Distribution')
plt.xticks(rotation=45,ha='right')
plt.ylabel("count")
plt.show()
```



In [59]:

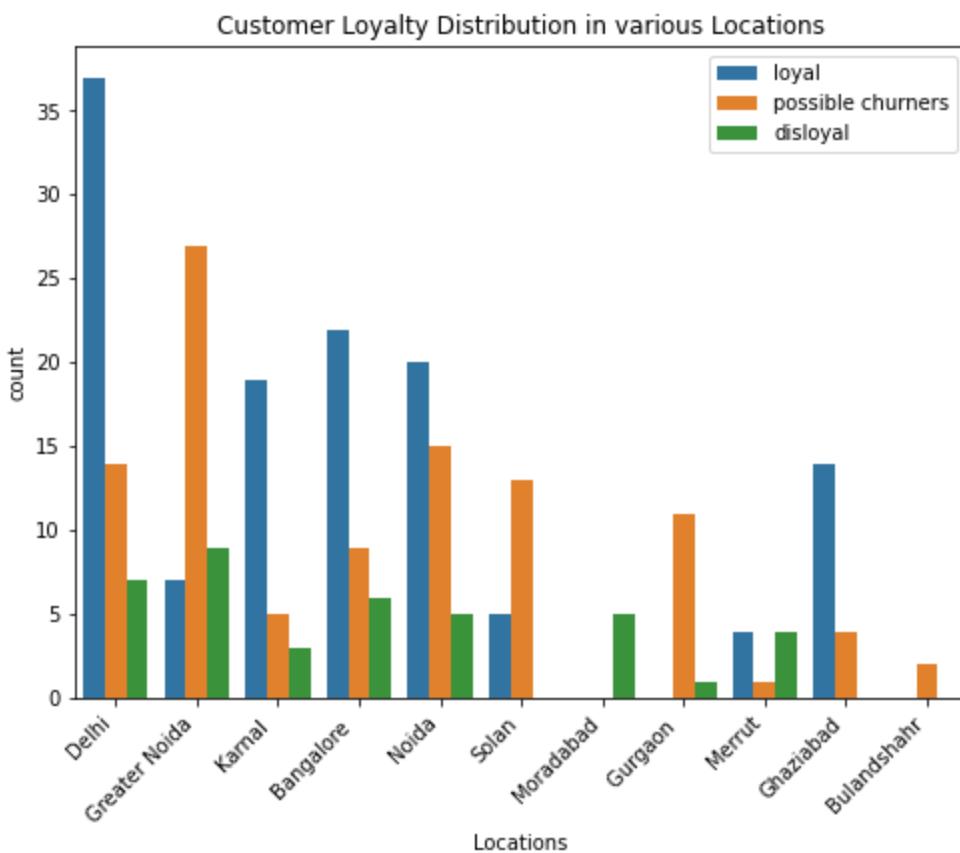
```
#Male and female Distribution in various Locations
plt.subplots(figsize=(8, 6))
sns.countplot(x=df[2],hue=df[0])
plt.title("Male and female Distribution in various Locations")
plt.xlabel('Locations of Customers')
plt.xticks(rotation=45,ha='right')
plt.ylabel("count")
plt.show()
```

Male and female Distribution in various Locations



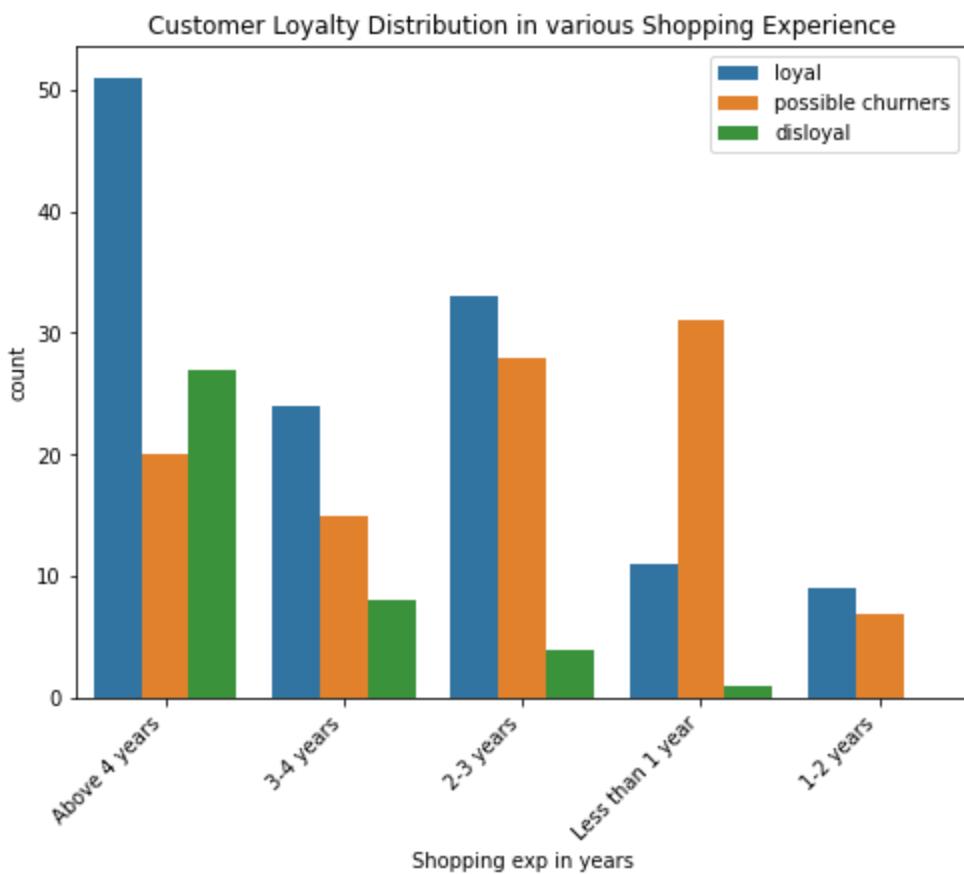
In [60]:

```
#Customer Loyalty Distribution in various Locations
plt.subplots(figsize=(8,6))
sns.countplot(x=df[2],hue=df[44])
plt.title("Customer Loyalty Distribution in various Locations")
plt.xlabel('Locations')
plt.xticks(rotation=45,ha='right')
plt.legend(loc='upper right')
plt.ylabel("count")
plt.show()
```



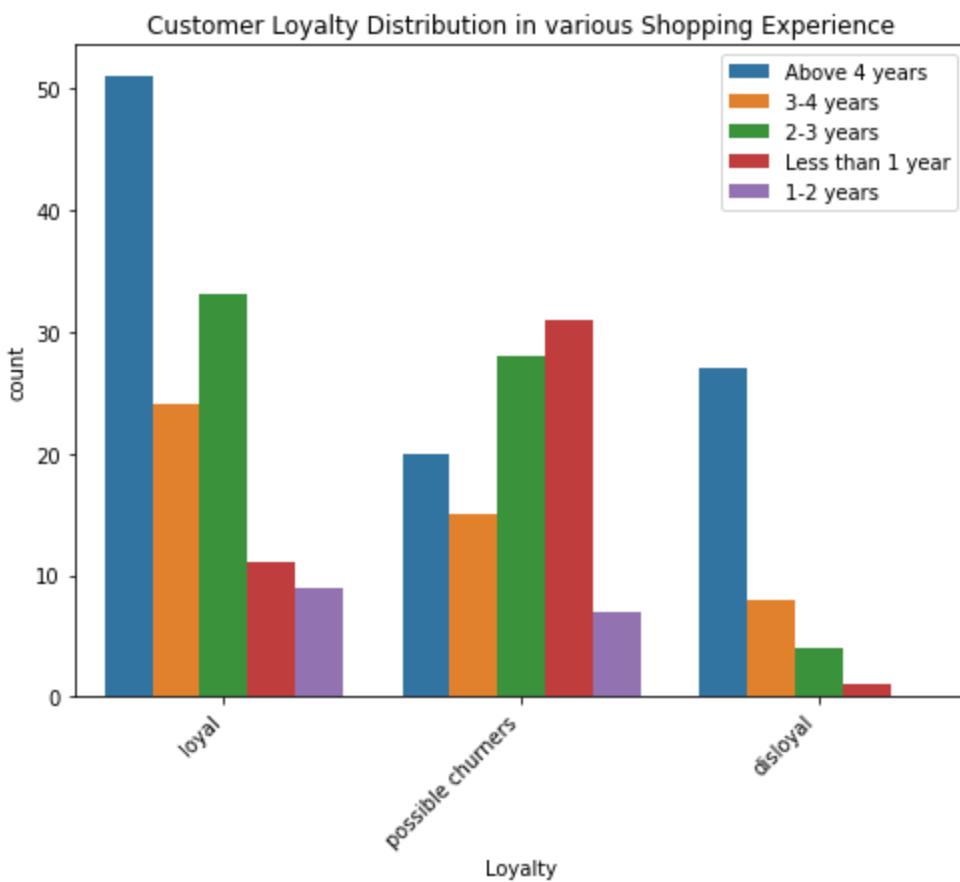
In [61]:

```
#Customer Loyalty Distribution in various Shopping Experience
plt.subplots(figsize=(8, 6))
sns.countplot(x=df[4], hue=df[44])
plt.title("Customer Loyalty Distribution in various Shopping Experience")
plt.xlabel('Shopping exp in years')
plt.xticks(rotation=45, ha='right')
plt.legend(loc='upper right')
plt.ylabel("count")
plt.show()
```



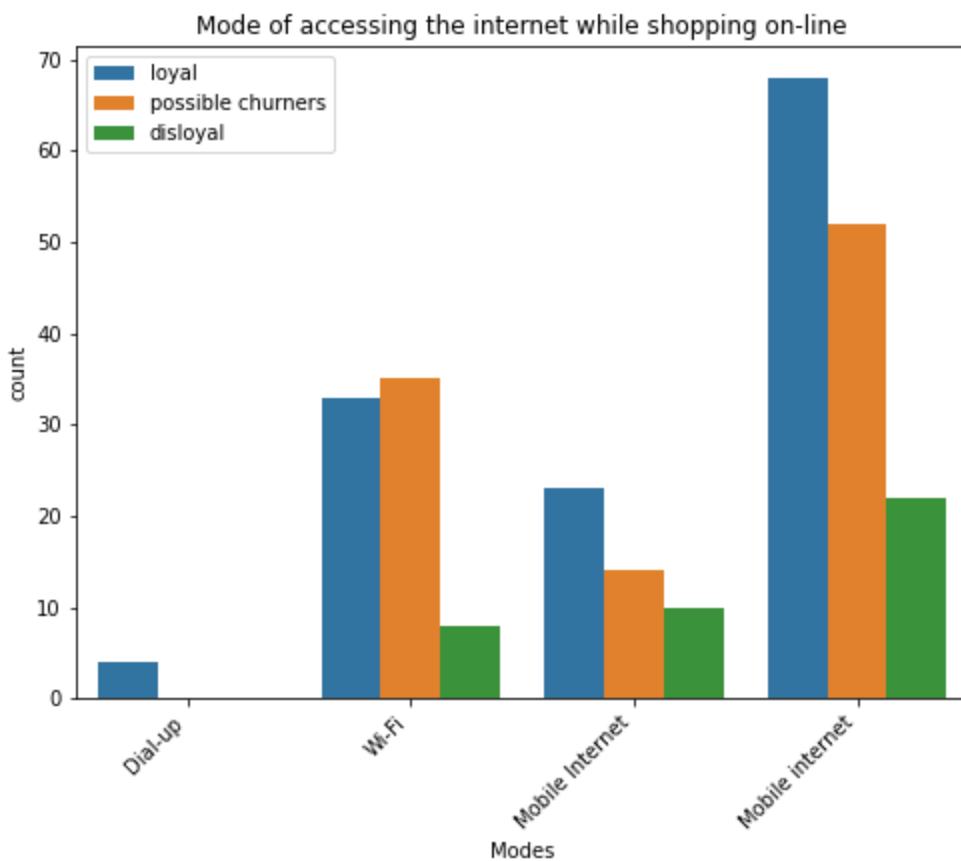
In [62]:

```
#Customer Loyalty Distribution in various Shopping Experience
plt.subplots(figsize=(8, 6))
sns.countplot(x=df[44], hue=df[4])
plt.title("Customer Loyalty Distribution in various Shopping Experience")
plt.xlabel('Loyalty')
plt.xticks(rotation=45, ha='right')
plt.legend(loc='upper right')
plt.ylabel("count")
plt.show()
```



In [63]:

```
# Mode of accessing the internet while shopping on-line
plt.subplots(figsize=(8, 6))
sns.countplot(x=df[6], hue=df[44])
plt.title("Mode of accessing the internet while shopping on-line")
plt.xlabel('Modes')
plt.xticks(rotation=45, ha='right')
plt.legend(loc='upper left')
plt.ylabel("count")
plt.show()
```



In [64]:

```
df[6]
```

Out[64]:

```
0          Dial-up
1          Wi-Fi
2      Mobile Internet
3      Mobile Internet
4          Wi-Fi
...
264     Mobile Internet
265     Mobile Internet
266     Mobile internet
267          Wi-Fi
268     Mobile Internet
Name: 6, Length: 269, dtype: object
```

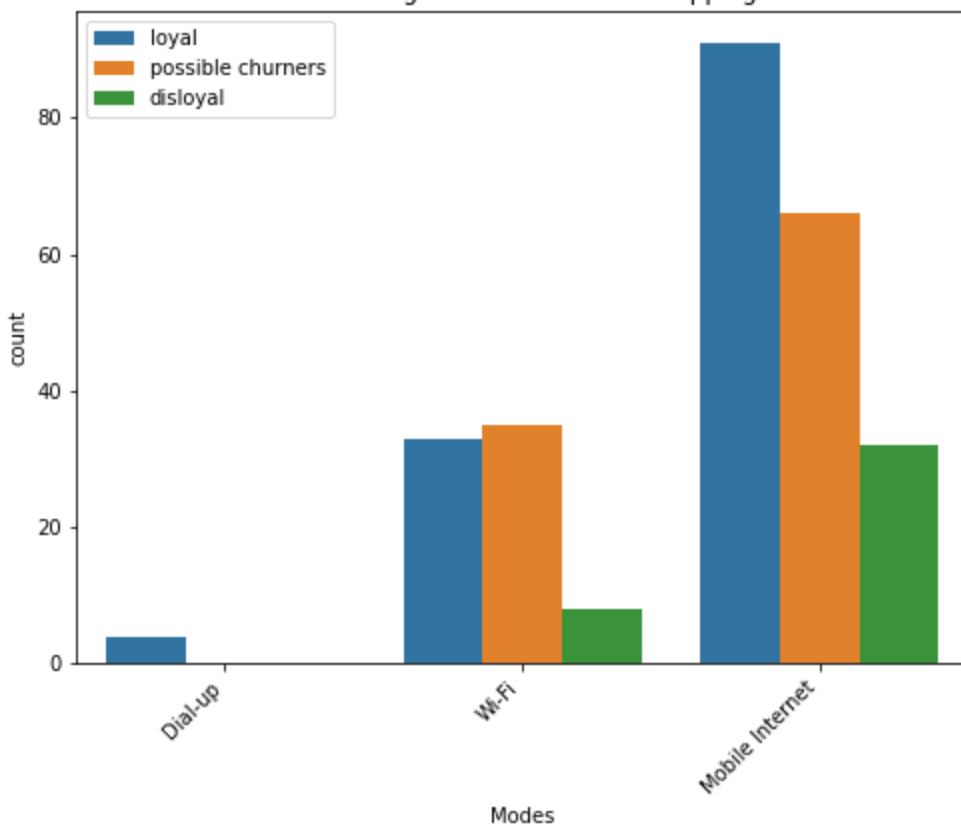
In [65]:

```
df[6] = [i.replace('Mobile internet','Mobile Internet') for i in df[6]]
```

In [66]:

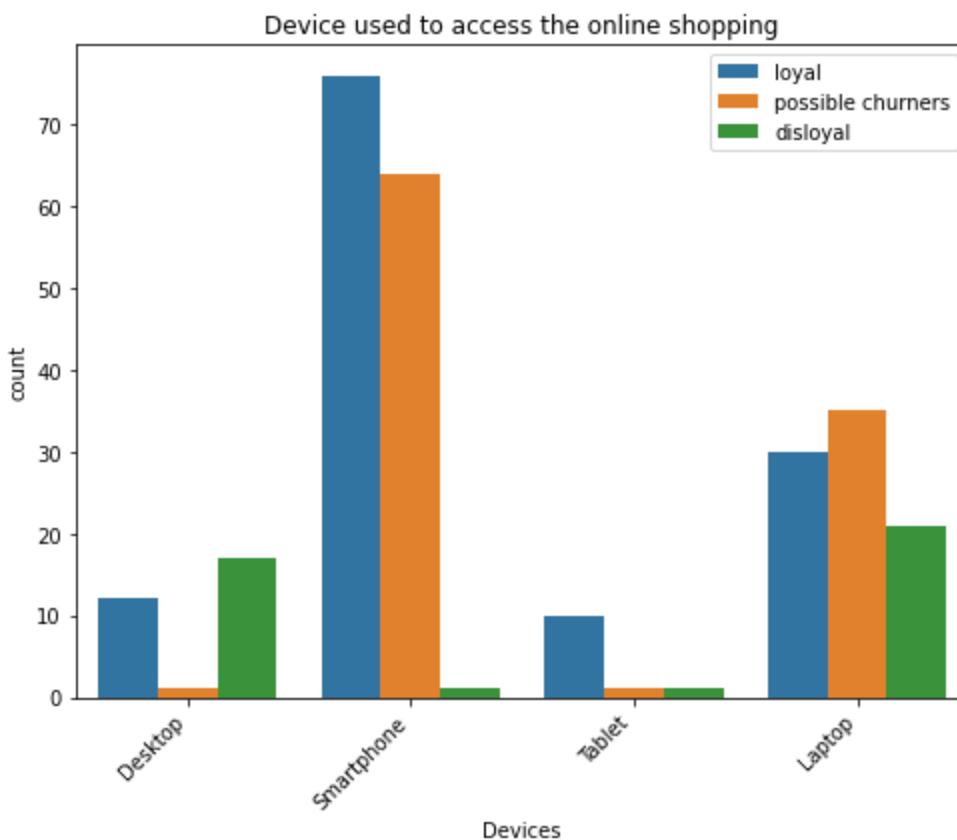
```
# Mode of accessing the internet while shopping on-line
plt.subplots(figsize=(8,6))
sns.countplot(x=df[6],hue=df[44])
plt.title("Mode of accessing the internet while shopping on-line")
plt.xlabel('Modes')
plt.xticks(rotation=45,ha='right')
plt.legend(loc='upper left')
plt.ylabel("count")
plt.show()
```

Mode of accessing the internet while shopping on-line



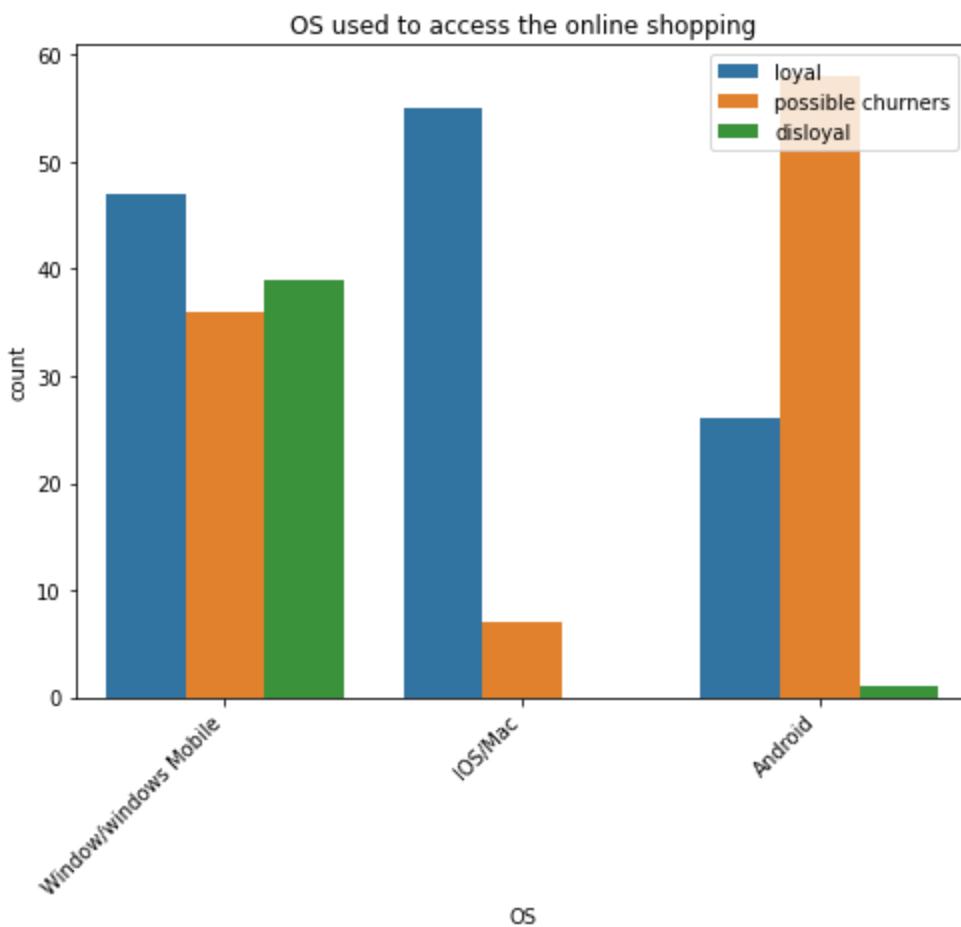
In [67]:

```
# Device used to access the online shopping
plt.subplots(figsize=(8, 6))
sns.countplot(x=df[7], hue=df[44])
plt.title("Device used to access the online shopping")
plt.xlabel('Devices')
plt.xticks(rotation=45, ha='right')
plt.legend(loc='upper right')
plt.ylabel("count")
plt.show()
```



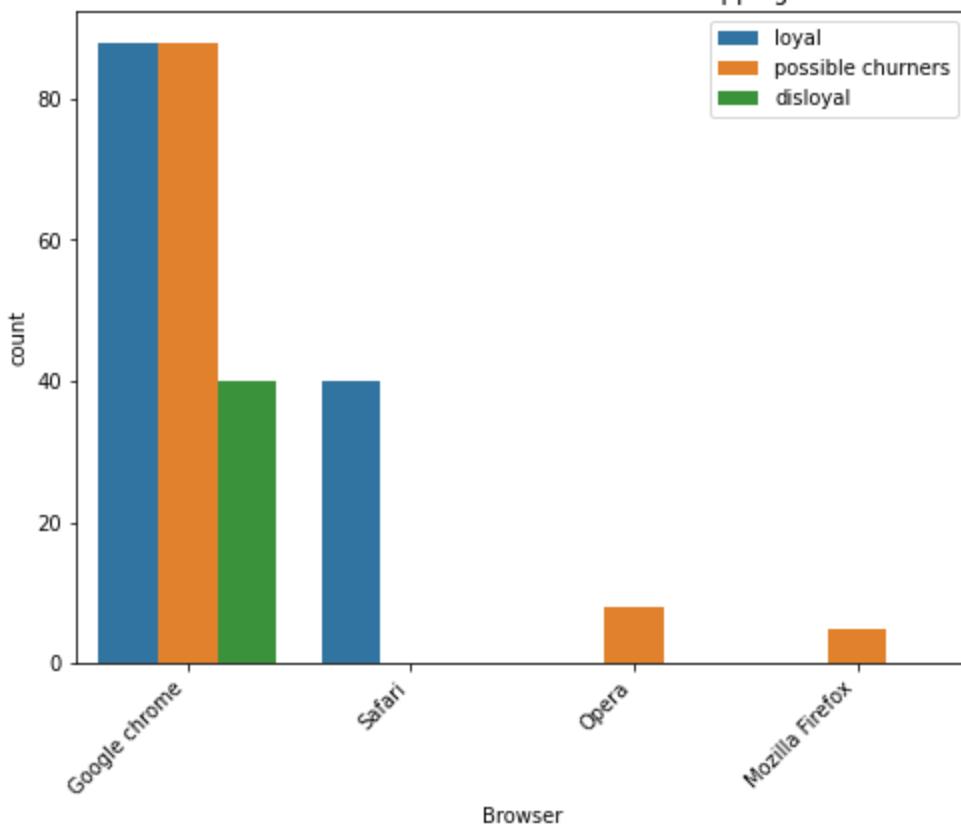
In [68]:

```
# OS used to access the online shopping
plt.subplots(figsize=(8, 6))
sns.countplot(x=df[9],hue=df[44])
plt.title("OS used to access the online shopping")
plt.xlabel('OS')
plt.xticks(rotation=45,ha='right')
plt.legend(loc='upper right')
plt.ylabel("count")
plt.show()
```



```
In [69]: # Browser used to access the online shopping
plt.subplots(figsize=(8, 6))
sns.countplot(x=df[10], hue=df[44])
plt.title("Browser used to access the online shopping")
plt.xlabel('Browser')
plt.xticks(rotation=45, ha='right')
plt.legend(loc='upper right')
plt.ylabel("count")
plt.show()
```

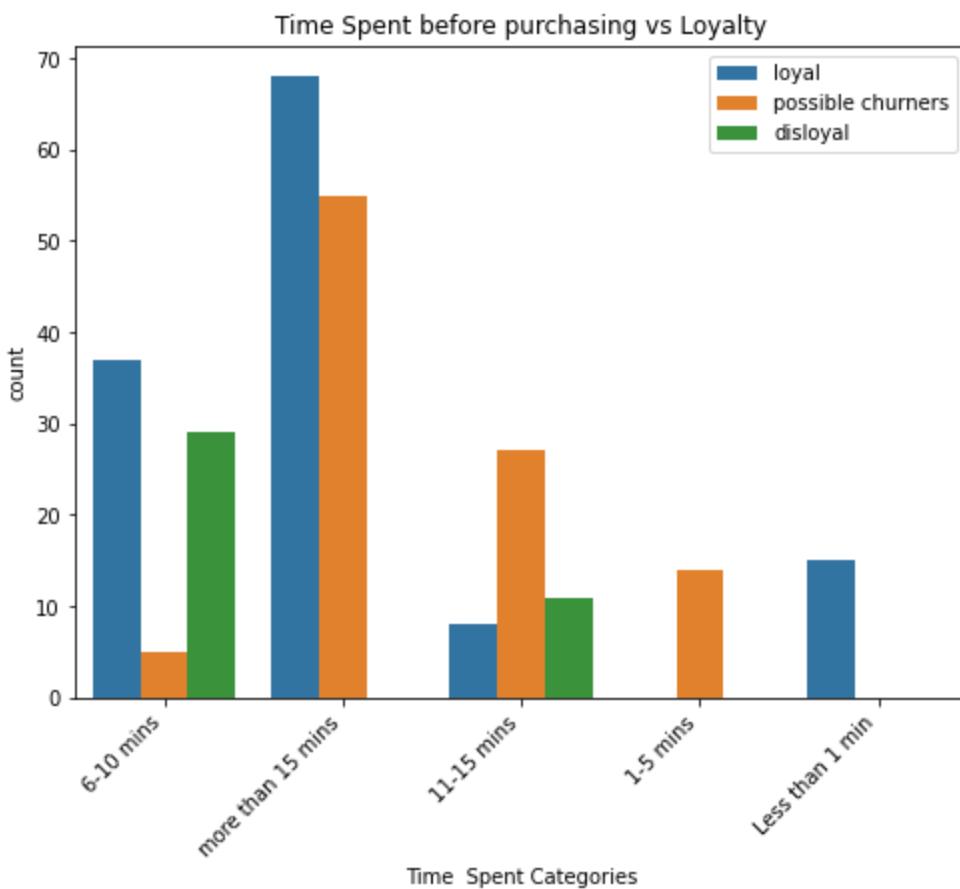
Browser used to access the online shopping



```
In [70]: df[13].value_counts()
```

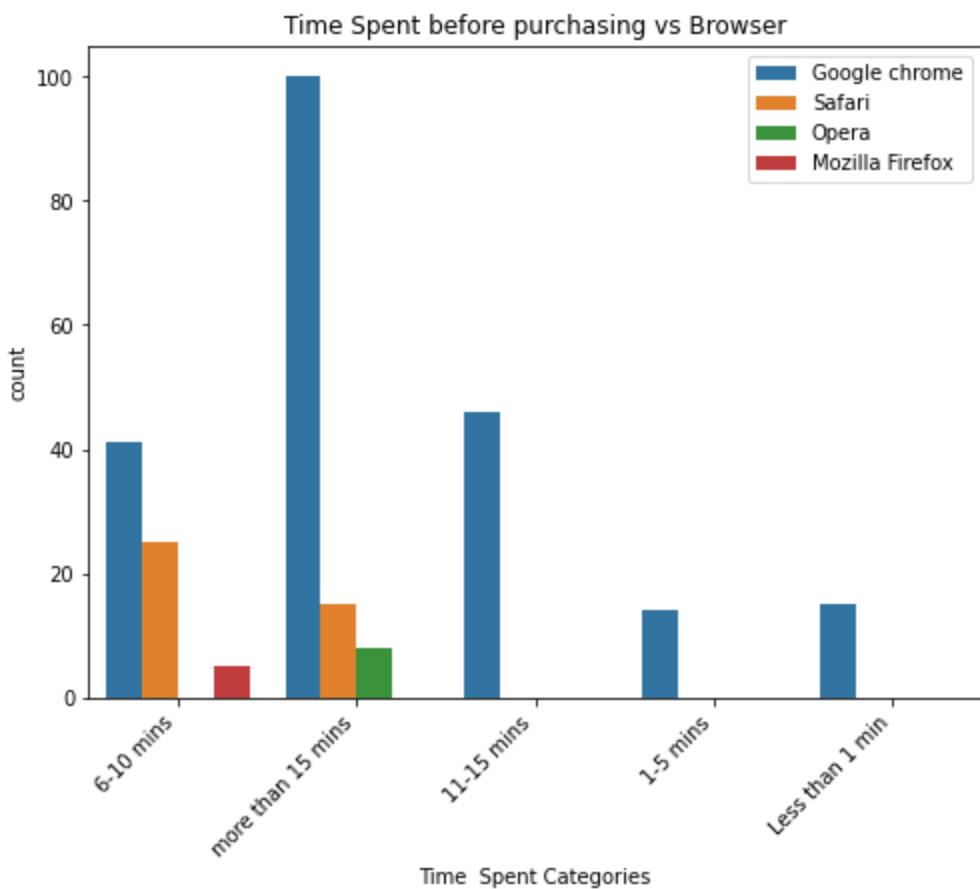
```
Out[70]: more than 15 mins    123
          6-10 mins        71
          11-15 mins        46
          Less than 1 min    15
          1-5 mins          14
          Name: 13, dtype: int64
```

```
In [71]: # Time Spent before purchasing vs Loyalty
plt.subplots(figsize=(8,6))
sns.countplot(x=df[13],hue=df[44])
plt.title("Time Spent before purchasing vs Loyalty")
plt.xlabel('Time Spent Categories')
plt.xticks(rotation=45,ha='right')
plt.legend(loc='upper right')
plt.ylabel("count")
plt.show()
```



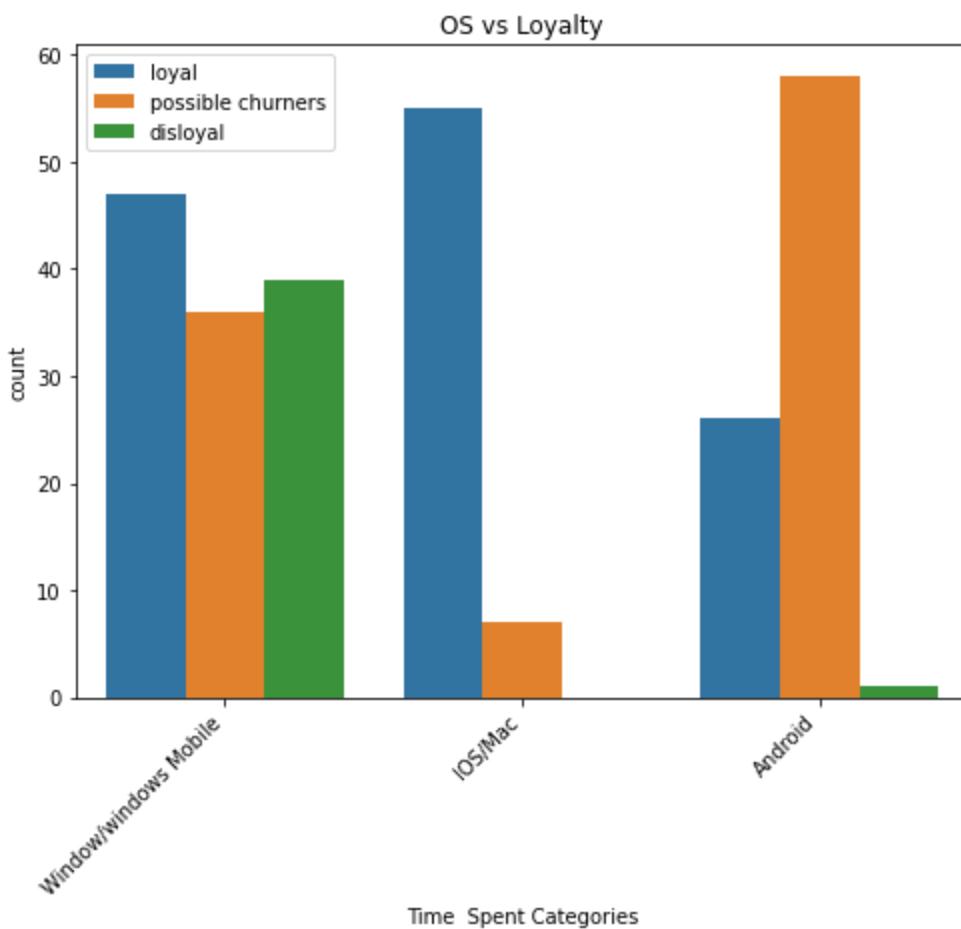
In [72]:

```
# Time Spent before purchasing vs Browser
plt.subplots(figsize=(8, 6))
sns.countplot(x=df[13], hue=df[10])
plt.title("Time Spent before purchasing vs Browser")
plt.xlabel('Time Spent Categories')
plt.xticks(rotation=45, ha='right')
plt.legend(loc='upper right')
plt.ylabel("count")
plt.show()
```

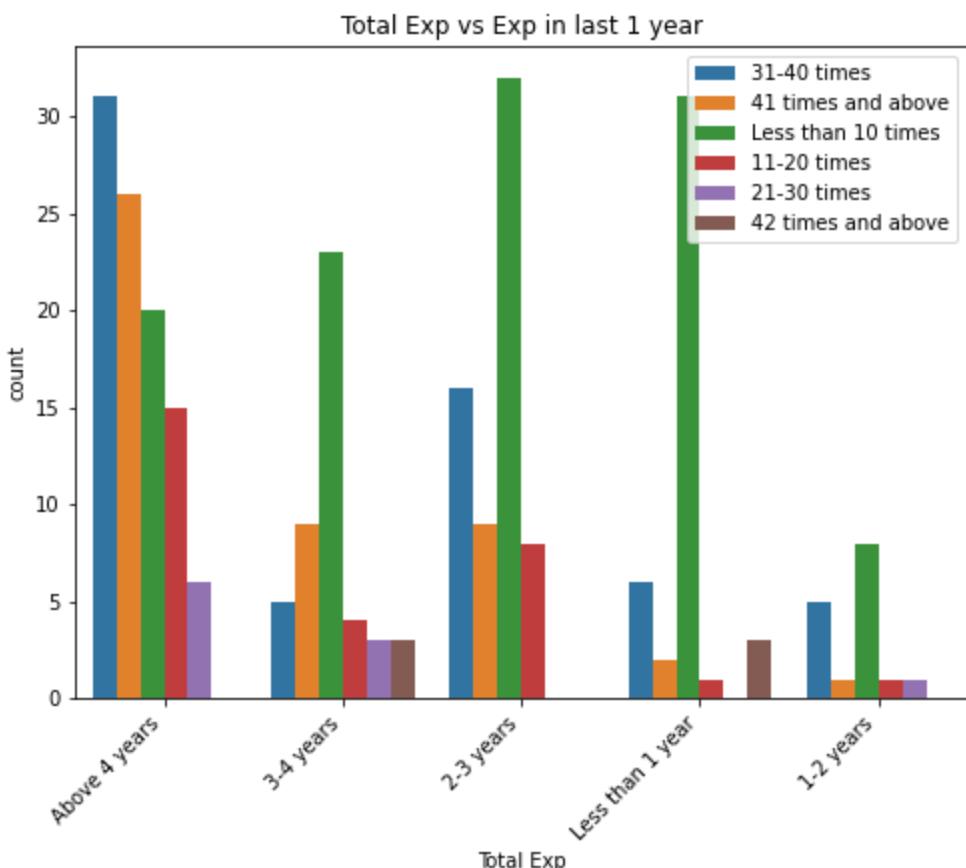


In [73]:

```
# OS vs Loyalty
plt.subplots(figsize=(8, 6))
sns.countplot(x=df[9], hue=df[44])
plt.title("OS vs Loyalty")
plt.xlabel('Time Spent Categories')
plt.xticks(rotation=45, ha='right')
plt.legend(loc='upper left')
plt.ylabel("count")
plt.show()
```

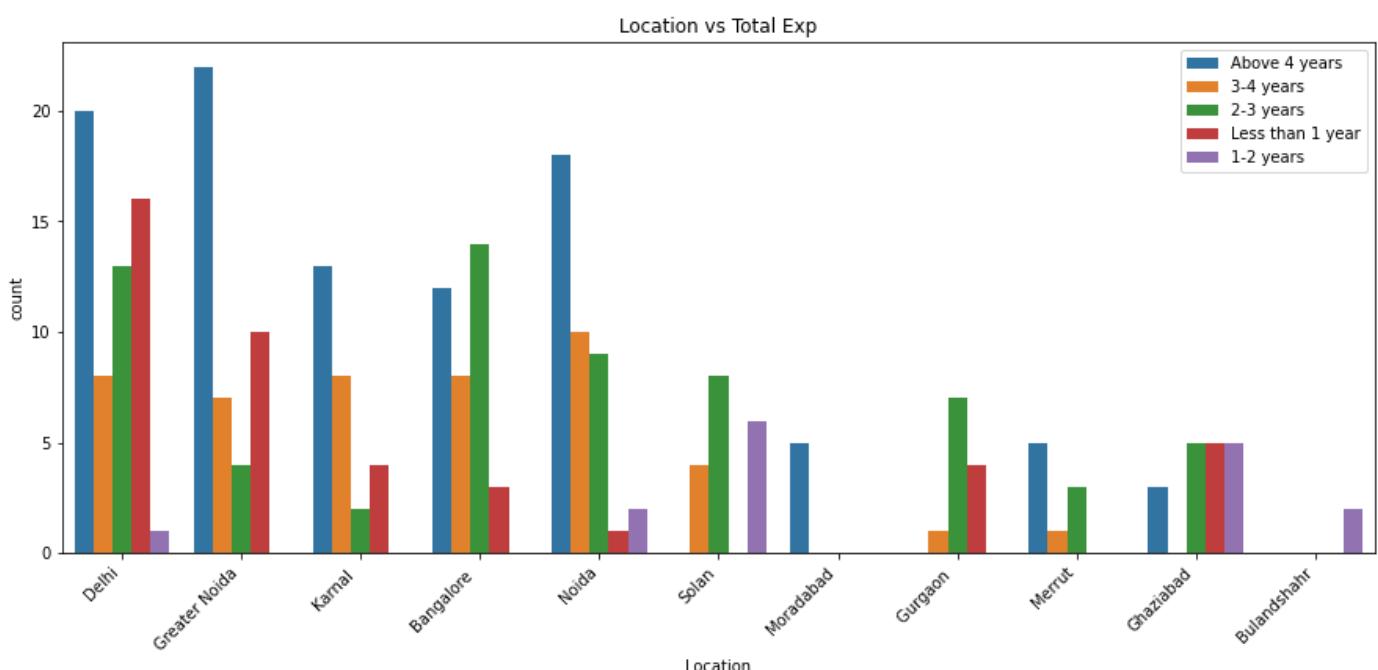


```
In [74]: # Total Exp vs Exp in last 1 year
plt.subplots(figsize=(8, 6))
sns.countplot(x=df[4], hue=df[5])
plt.title("Total Exp vs Exp in last 1 year")
plt.xlabel('Total Exp')
plt.xticks(rotation=45, ha='right')
plt.legend(loc='upper right')
plt.ylabel("count")
plt.show()
```



In [75]:

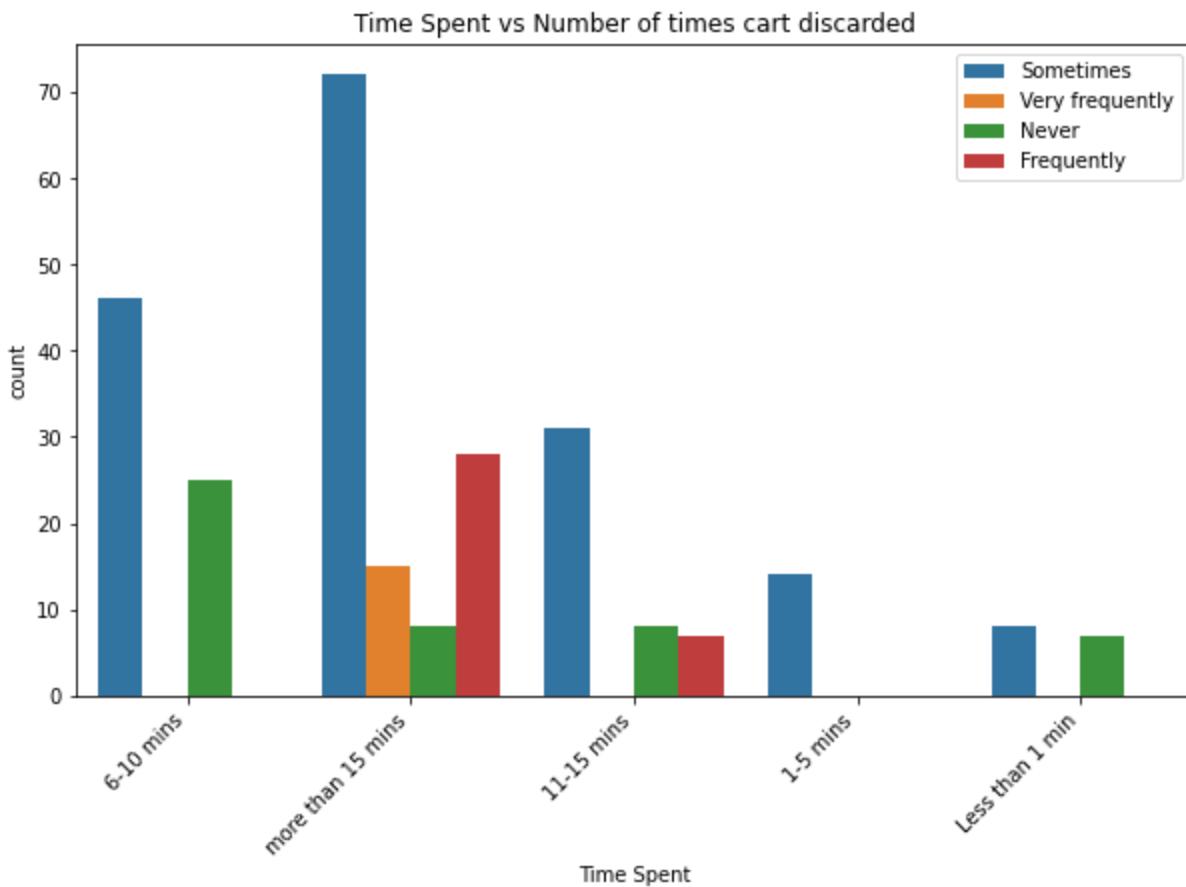
```
# Location vs Total Exp
plt.subplots(figsize=(15, 6))
sns.countplot(x=df[2], hue=df[4])
plt.title("Location vs Total Exp")
plt.xlabel('Location')
plt.xticks(rotation=45, ha='right')
plt.legend(loc='upper right')
plt.ylabel("count")
plt.show()
```



In [76]:

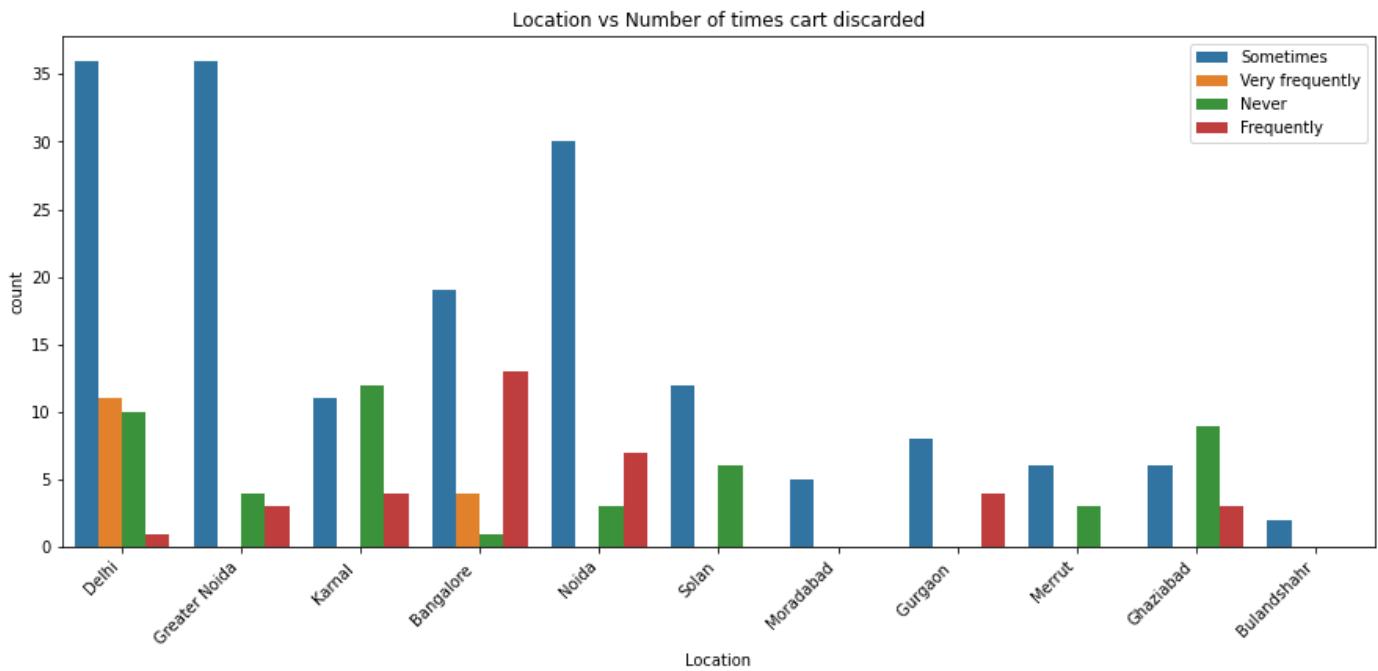
```
# Time Spent vs Number of times cart discarded
plt.subplots(figsize=(10, 6))
sns.countplot(x=df[13], hue=df[15])
```

```
plt.title("Time Spent vs Number of times cart discarded")
plt.xlabel('Time Spent')
plt.xticks(rotation=45,ha='right')
plt.legend(loc='upper right')
plt.ylabel("count")
plt.show()
```



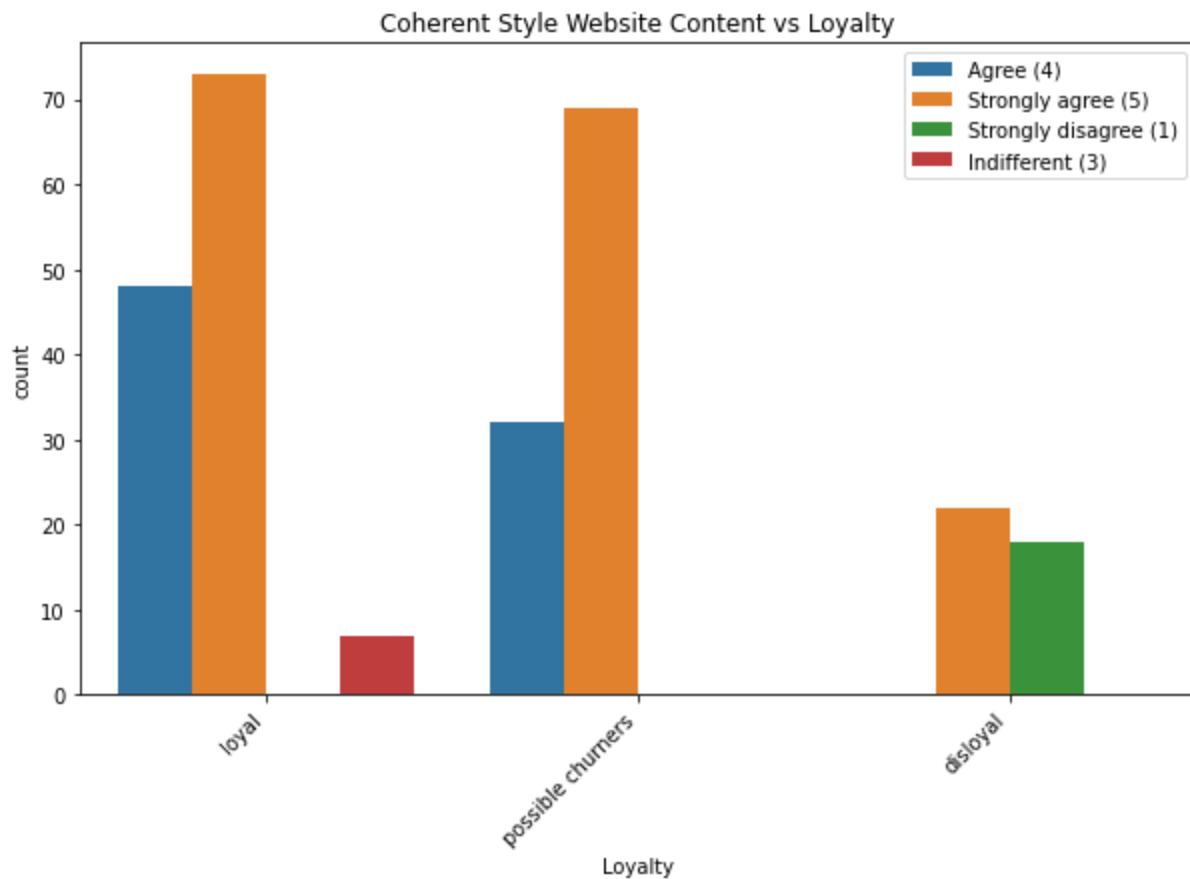
In [77]:

```
# Location vs Number of times cart discarded
plt.subplots(figsize=(15,6))
sns.countplot(x=df[2],hue=df[15])
plt.title("Location vs Number of times cart discarded")
plt.xlabel('Location')
plt.xticks(rotation=45,ha='right')
plt.legend(loc='upper right')
plt.ylabel("count")
plt.show()
```



In [78]:

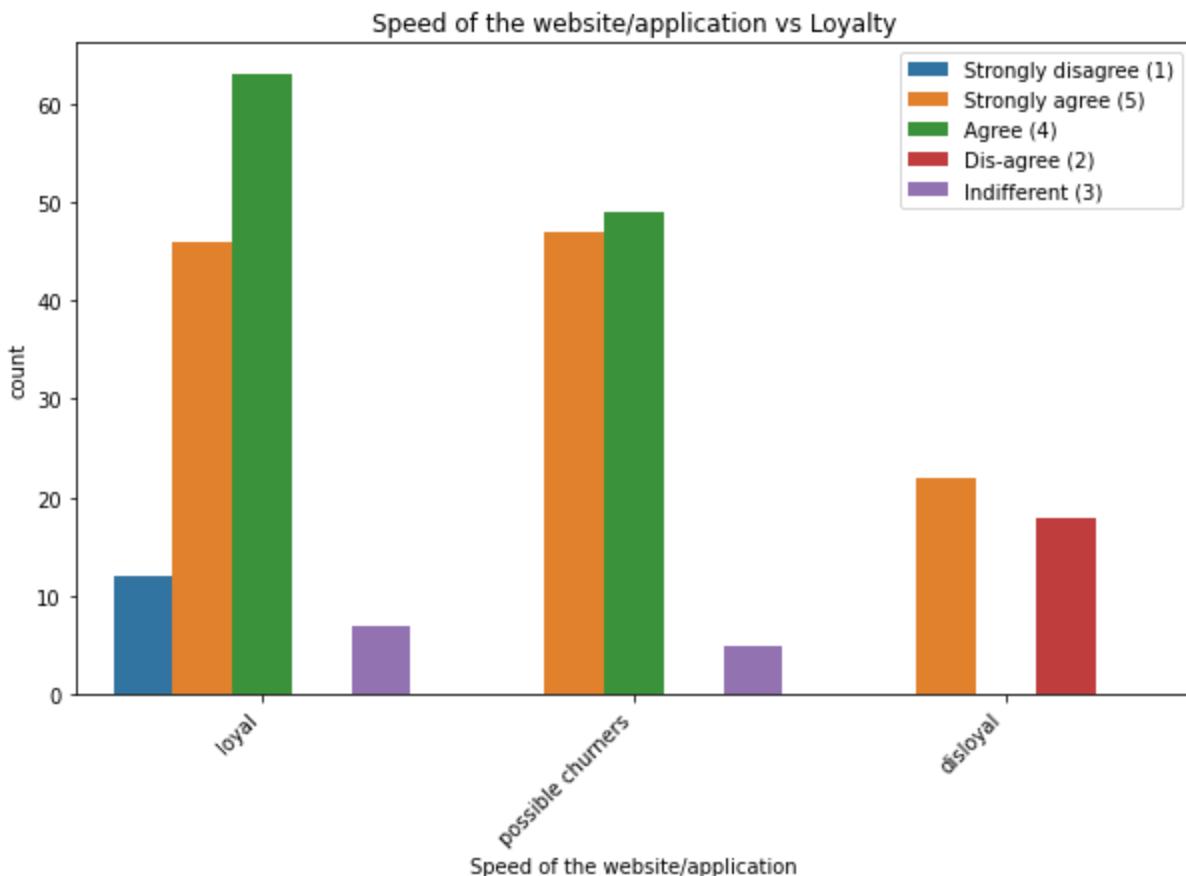
```
# Coherent Style Website Content vs Loyalty
plt.subplots(figsize=(10,6))
sns.countplot(x=df[44],hue=df[17])
plt.title("Coherent Style Website Content vs Loyalty")
plt.xlabel('Loyalty')
plt.xticks(rotation=45,ha='right')
plt.legend(loc='upper right')
plt.ylabel("count")
plt.show()
```



In [79]:

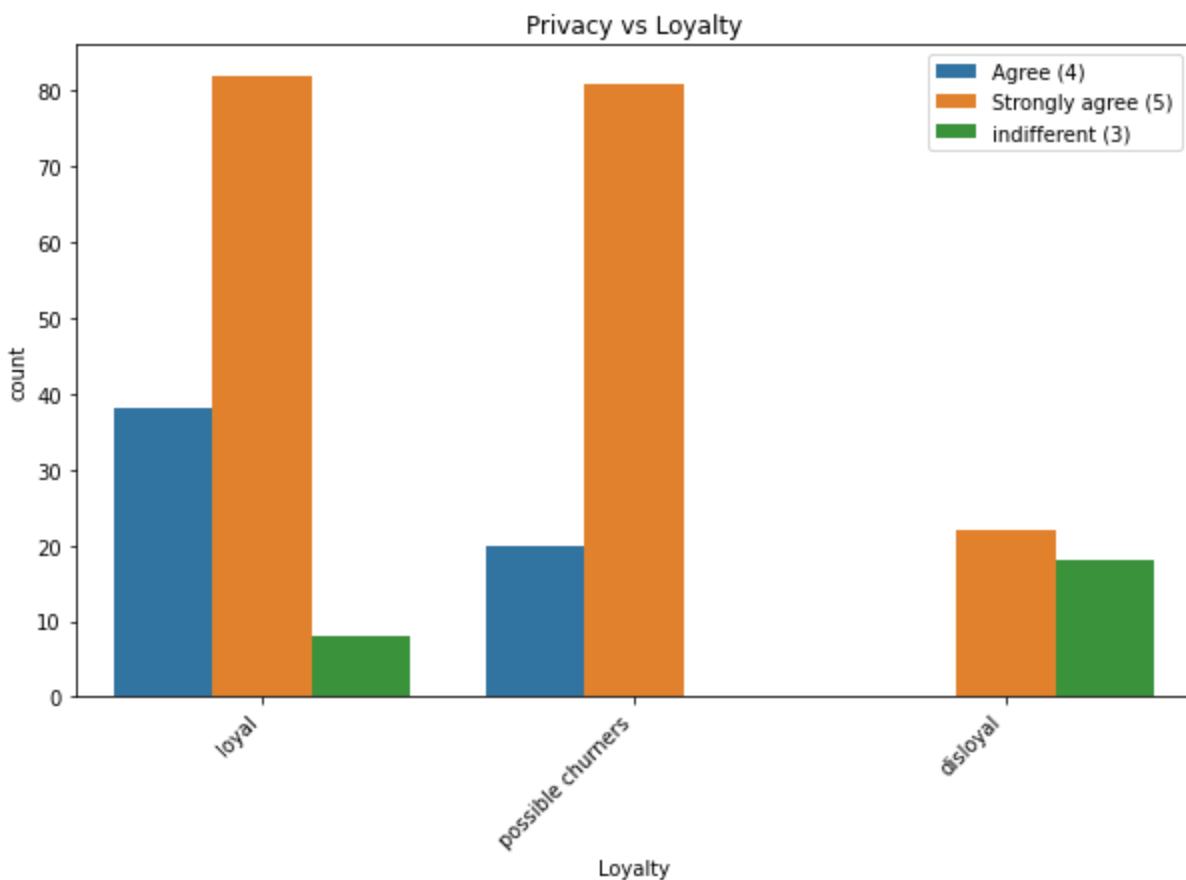
```
# Speed of the website/application vs Loyalty
plt.subplots(figsize=(10,6))
sns.countplot(x=df[44],hue=df[22])
```

```
plt.title("Speed of the website/application vs Loyalty")
plt.xlabel('Speed of the website/application')
plt.xticks(rotation=45,ha='right')
plt.legend(loc='upper right')
plt.ylabel("count")
plt.show()
```



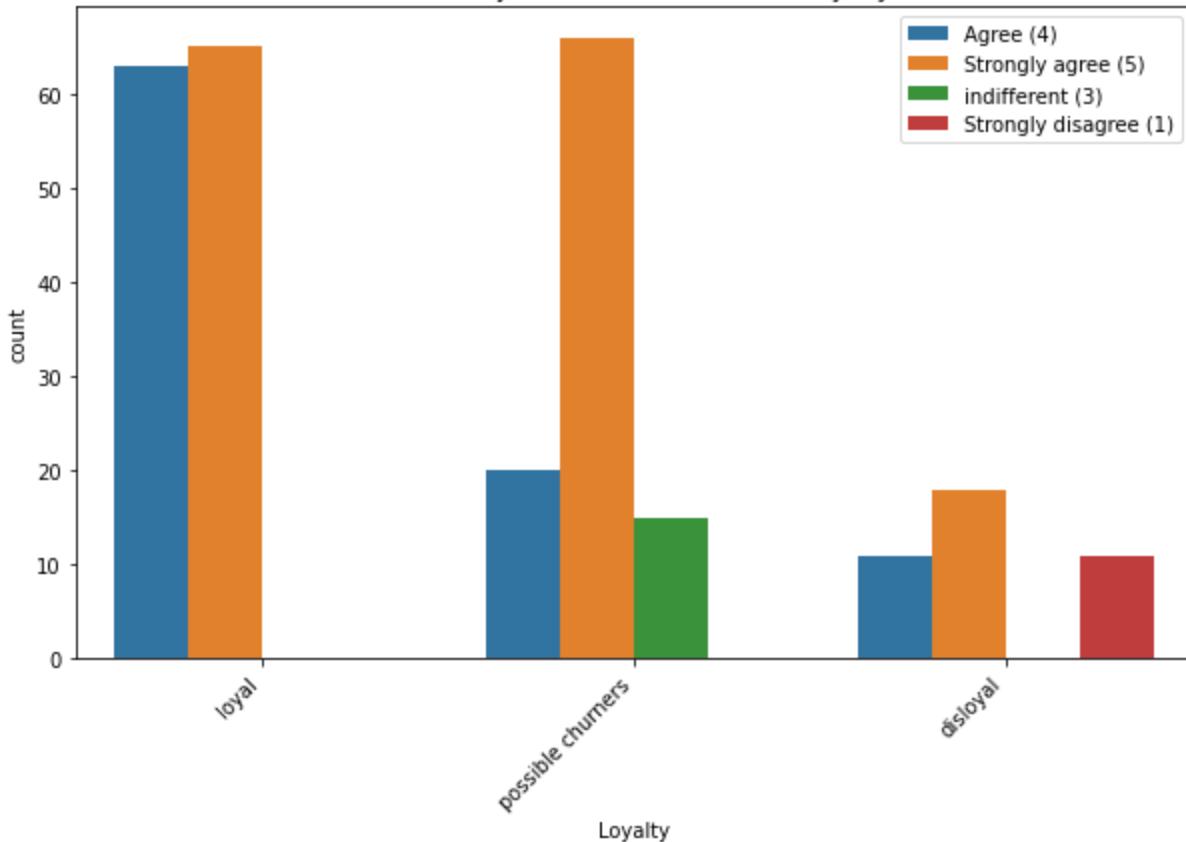
In [80]:

```
# Privacy vs Loyalty
plt.subplots(figsize=(10, 6))
sns.countplot(x=df[44], hue=df[27])
plt.title("Privacy vs Loyalty")
plt.xlabel('Loyalty')
plt.xticks(rotation=45,ha='right')
plt.legend(loc='upper right')
plt.ylabel("count")
plt.show()
```



```
In [81]: # Availability of Customer Service vs Loyalty
plt.subplots(figsize=(10,6))
sns.countplot(x=df[44], hue=df[28])
plt.title("Availability of Customer Service vs Loyalty")
plt.xlabel('Loyalty')
plt.xticks(rotation=45, ha='right')
plt.legend(loc='upper right')
plt.ylabel("count")
plt.show()
```

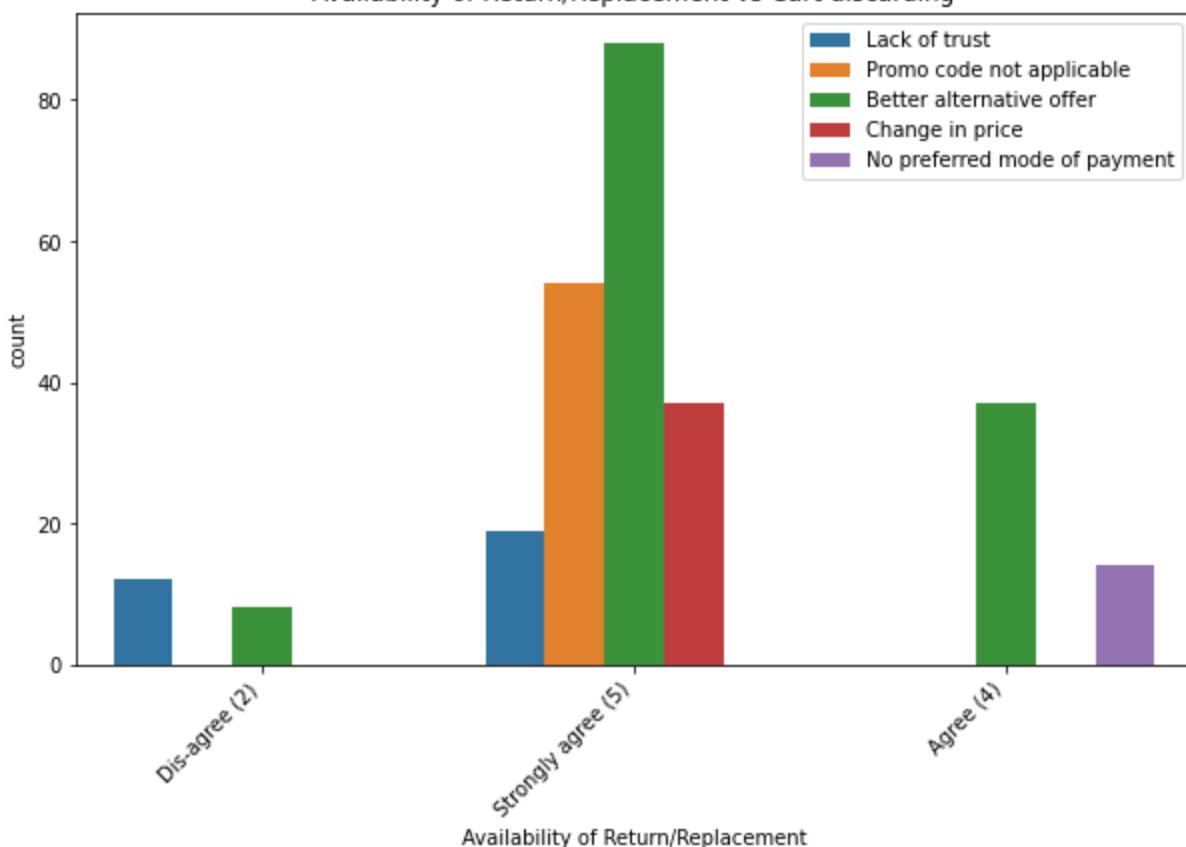
### Availability of Customer Service vs Loyalty



In [82]:

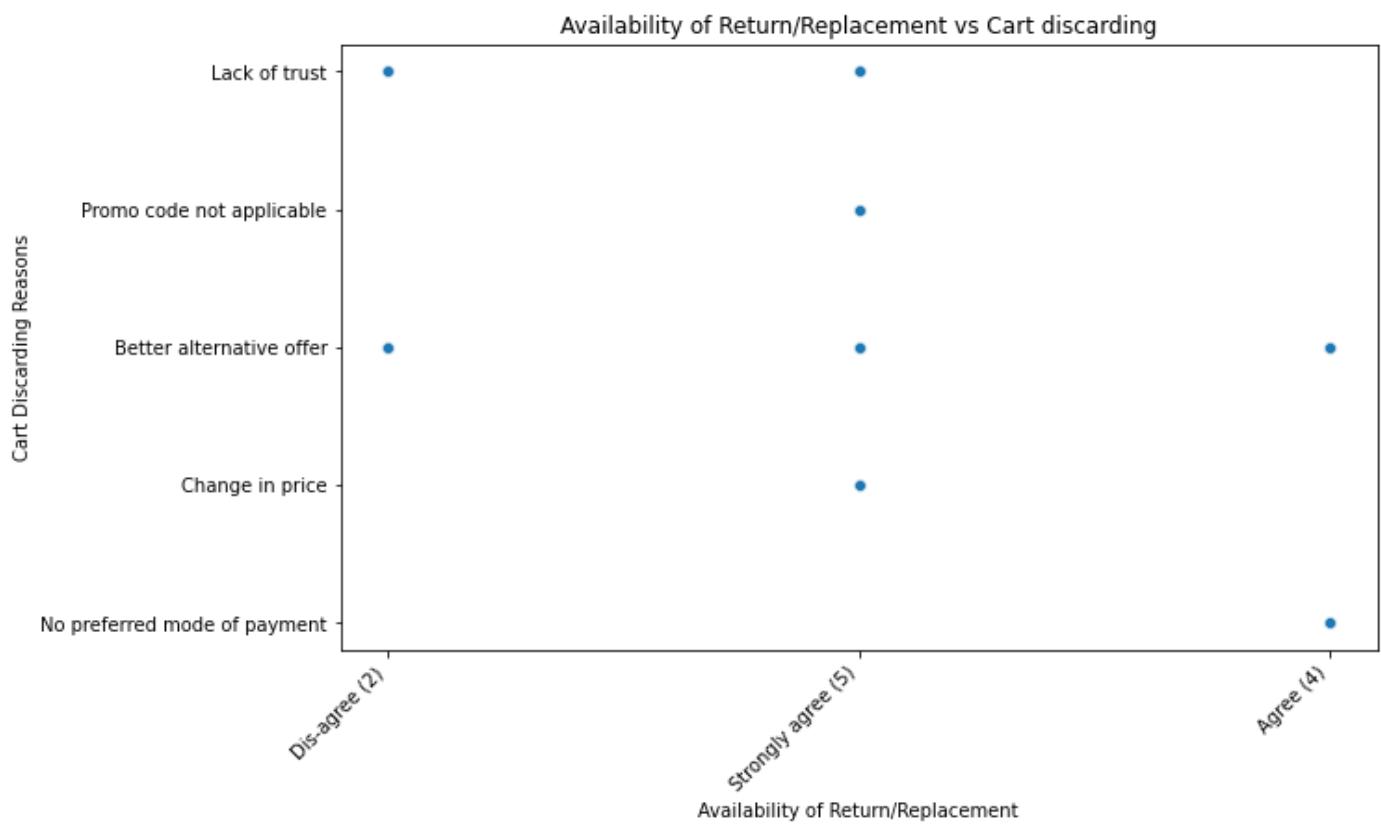
```
# Availability of Return/Replacement vs Cart discarding
plt.subplots(figsize=(10,6))
sns.countplot(x=df[32],hue=df[16])
plt.title("Availability of Return/Replacement vs Cart discarding")
plt.xlabel('Availability of Return/Replacement')
plt.xticks(rotation=45,ha='right')
plt.legend(loc='upper right')
plt.ylabel("count")
plt.show()
```

Availability of Return/Replacement vs Cart discarding

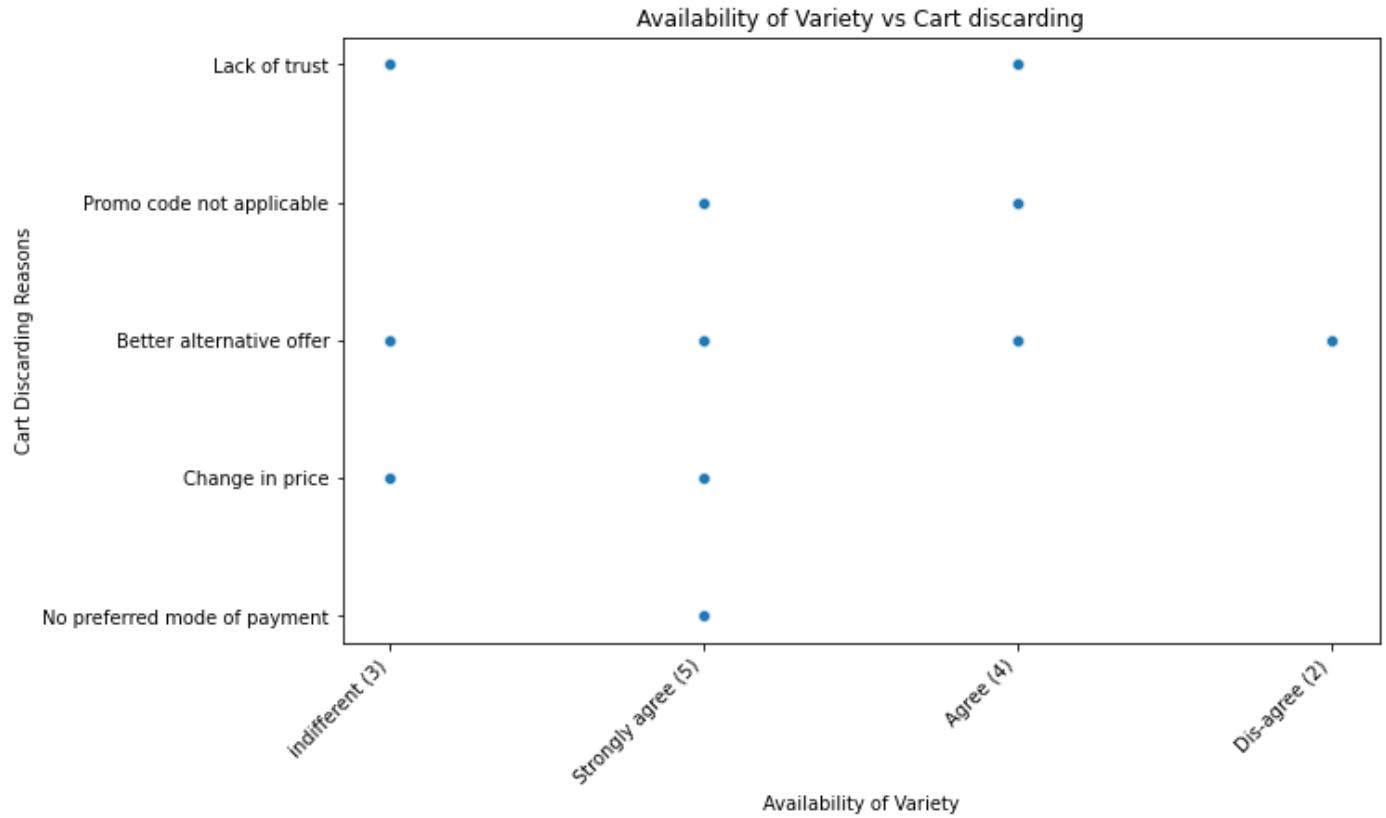


In [83]:

```
plt.subplots(figsize=(10, 6))
sns.scatterplot(x=df[32], y=df[16])
plt.title("Availability of Return/Replacement vs Cart discarding")
plt.xlabel('Availability of Return/Replacement')
plt.xticks(rotation=45, ha='right')
plt.ylabel("Cart Discarding Reasons")
plt.show()
```



```
In [84]: # Availability of Variety vs Cart discarding
plt.subplots(figsize=(10, 6))
sns.scatterplot(x=df[38],y=df[16])
plt.title("Availability of Variety vs Cart discarding")
plt.xlabel('Availability of Variety')
plt.xticks(rotation=45,ha='right')
plt.ylabel("Cart Discarding Reasons")
plt.show()
```



```
In [85]: df[48]
```

```
Out[85]: 0 Paytm.com
1 Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com
2 Amazon.in, Paytm.com, Myntra.com
3 Amazon.in, Flipkart.com, Paytm.com, Myntra.com...
4 Amazon.in, Flipkart.com, Paytm.com, Myntra.com...
...
264 Amazon.in
265 Flipkart.com
266 Amazon.in
267 Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com
268 Amazon.in
Name: 48, Length: 269, dtype: object
```

```
In [86]: df[70]
```

```
Out[86]: 0 Flipkart.com
1 Amazon.in, Myntra.com
2 Amazon.in, Paytm.com, Myntra.com
3 Amazon.in, Flipkart.com
4 Amazon.in, Myntra.com
...
264 Amazon.in
265 Flipkart.com
266 Amazon.in
267 Amazon.in
```

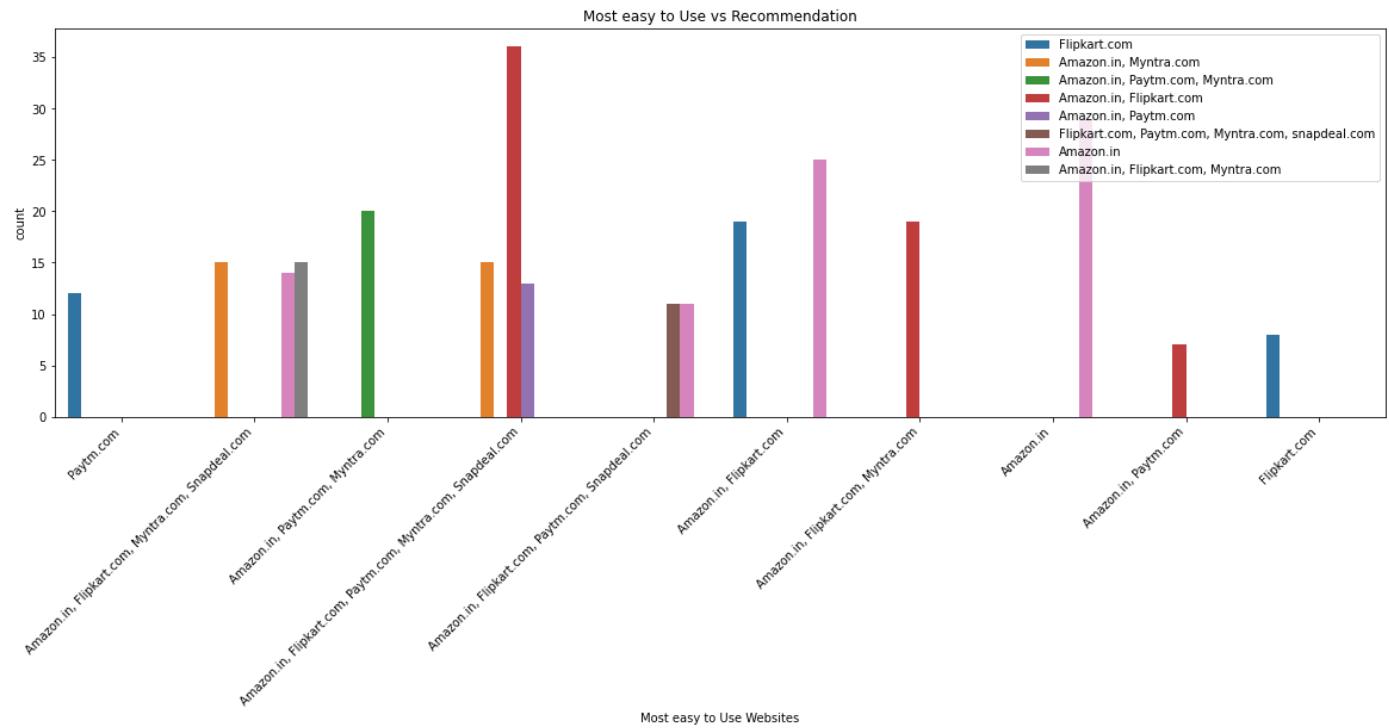
268

Amazon.in

Name: 70, Length: 269, dtype: object

In [87]:

```
# Most easy to Use vs Recommendation
plt.subplots(figsize=(20,6))
sns.countplot(x=df[48], hue=df[70])
plt.title("Most easy to Use vs Recommendation")
plt.xlabel('Most easy to Use Websites')
plt.xticks(rotation=45, ha='right')
plt.legend(loc='upper right')
plt.ylabel("count")
plt.show()
```



In [ ]:

# **Observations:**

## **Univariate Analysis**

- There are more female customers compared to men customers who made purchases
- 60% of purchases are made by customers aged between 20 – 40yrs
- The highest purchases were made from Delhi followed by Greater Noida, Noida and Bangalore and Bulandshar with the least number of purchases
- 50% of the customers have experience in online purchases more than 3 years
- 57% of people purchased less than 30 times over the last year
- 70% of customers used mobile internet during their purchases 28% used WIFI and 2% used dialup connections to make purchases
- 57% of purchases were through mobiles (52%) and tablets (5%) and rest 43% was made through laptops and desktops
- 45% of customers used Windows/windows devices to make purchases 32% through Android and 23%through IOS/MAC
- 81% of purchases were done through Google Chrome, 15% through Safari Browser and 4% through Opera and Firefox
- 63% of Customers spend a min of 10 mins on the website/application browsing before making the actual purchase while 11% spend less than 5 mins.
- 55% of the customers prefer to use either credit or debit cards to make purchase while 28% chose Cod as an option and remaining 17% use wallets.
- 91% of the customers feel that the User Interface i.e., the contents of the web page must be easy to read and understand
- 77% of the customers feel that the products for comparison should have complete info of that product
- 70% of the customers believe that they would prefer the complete details f the seller and the product to be listed before they decide to make any purchase. Out of this 46% of them strongly believe that these details are very much needed for making a purchase
- 84% of the customers feel that the loading and the processing time of the website or the application has an impact on the purchase decision. A slow website/application brings down the interest of the customer.
- 89% of the customers feel that easy and convenient payment methods have a high impact on purchases. Easier the payment methods higher the excitement of the customer towards purchases
- 90% of the customers are of the opinion that they tend to purchase products from sellers who are obliged to maintain the privacy of the customer
- 90% of the customers are of the opinion that they tend to purchase products from sellers who are able to provide service through various modes of communication
- 83% of the customers fee that they prefer to shop online because it is more convenient and flexible
- 97% of the customers feel that the quality of the website has more effect on the purchases. A good quality website attracts more customers and thereby encourage them to make purchases
- 76% of customers need more varieties in each category of products
- 86% of customers feel that they shop online because online shopping gives ore value for money mean cheaper prices for good quality products and service
- Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com are the easiest to use websites
- Amazon.in, Flipkart.com are the website which are more visually appealing to the customers

- Most of the customers feel that Amazon is the most reliable website to shop with followed by flipkart and paytm
- Most of them feel that Amazon has the highest delivery speed among all the websites
- Amazon is the website that the customers feel that they would recommend it to their friends and family.



# CUSTOMER RETENTION ANALYSIS

By

C. B. KARTHIK

# TABLE OF CONTENTS :-

1. INTRODUCTION
  - 1.1 WHAT IS CUSTOMER RETENTION?
  - 1.2 BACKGROUND OF THE PROBLEM
2. TOP 5 BENEFITS OF CUSTOMER RETENTION
3. DATA ANALYSIS
4. CONCLUSION
5. LIMITATION OF THE DATASET AND SCOPE FOR FUTURE WORK

# INTRODUCTION

## What is customer retention?

- Retention analysis (or survival analysis) is the process of analyzing user metrics to understand how and why customers churn.
- Retention analysis is key to gain insights on how to maintain a profitable customer base by improving retention and new user acquisition rates.

By running consistent retention analysis, what will we achieve:

- Reason why customers are churning.
- At what point of time the customers are more likely to leave.
- How does churning affect our bottom line.
- Strategies to follow to improve retention strategies.

# BACKGROUND OF THE PROBLEM

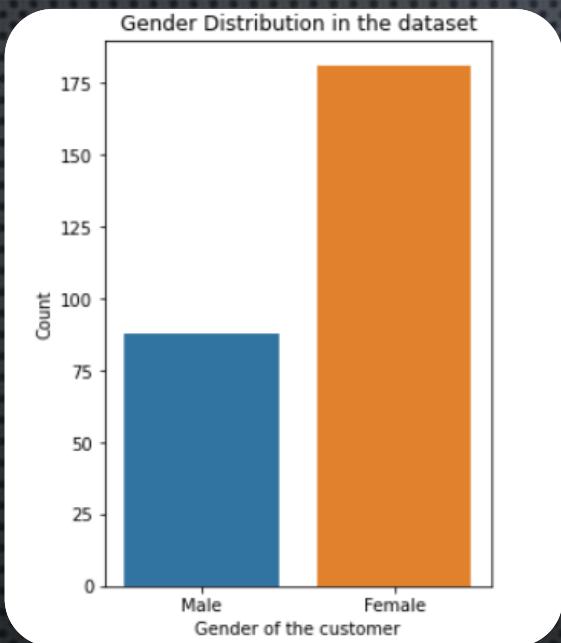
- KEEPING YOUR CURRENT CUSTOMERS HAPPY IS GENERALLY MORE COST-EFFECTIVE THAN ACQUIRING FIRST-TIME CUSTOMERS.
- ACCORDING TO THE HARVARD BUSINESS REVIEW, ACQUIRING A NEW CUSTOMER CAN BE 5 TO 25 TIMES MORE EXPENSIVE THAN HOLDING ON TO AN EXISTING ONE.
- THERE IS NOT NEED TO SPEND BIG ON MARKETING, ADVERTISING OR SALES OUTREACH.
- IT IS EASIER TO TURN EXISTING CUSTOMERS INTO REPEATING ONES, SINCE THEY ALREADY TRUST YOUR BRAND FROM PREVIOUS PURCHASES.
- NEW CUSTOMERS, HOWEVER, OFTEN REQUIRE MORE CONVINCING WHEN IT COMES TO THAT INITIAL SALE. CUSTOMER LOYALTY WON'T JUST GIVE YOU REPEAT BUSINESS.
- LOYAL CUSTOMERS ARE MORE LIKELY TO GIVE FREE RECOMMENDATIONS TO THEIR COLLEAGUES, FRIENDS AND FAMILY.
- CREATING THAT CYCLE OF RETAINED CUSTOMERS AND BUZZ MARKETING IS ONE WAY THE COMPANY CAN CULTIVATE CUSTOMER LOYALTY FOR LONG-TERM SUCCESS

## Top 5 benefits of customer retention

- REDUCING THE COST OF CUSTOMER ACQUISITION AS THE COST TO ACQUIRE NEW CUSTOMERS IS VERY HIGH COMPARED TO RETAINING OLD CUSTOMERS
- INCREASING THE VALUE OF EACH SALE
- GAINING INSIGHT INTO YOUR CUSTOMERS
- BUILDING REFERRALS AND LOYALTY
- REDUCING CUSTOMER CHURN

# Data analysis

## Univariate analysis: Gender of the Customer



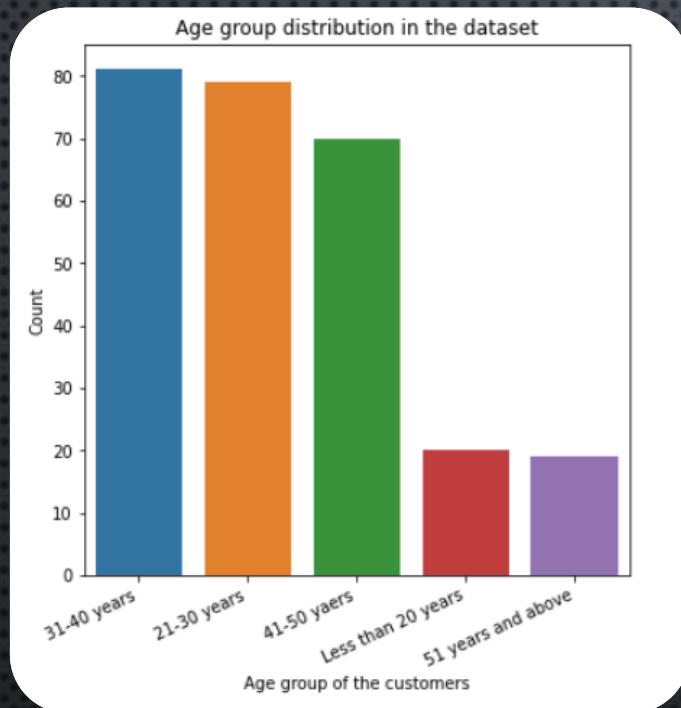
Female	181
Male	88

Observation:

Majority, 181 of the customers are Female whereas Male are 88.

There are more female customers compared to men customers who made purchases

## Univariate analysis: Gender of the Customer



31-40 years	81
21-30 years	79
41-50 yaers	70
Less than 20 years	20
51 years and above	19

Observation:

Majority of the customers are from age group 31-40 years.

60% of purchases are made by customers aged between 20 – 40yrs

## Univariate analysis: Location of the Customer



Delhi	58
Greater Noida	43
Noida	40
Bangalore	37
Karnal	27
Solan	18
Ghaziabad	18
Gurgaon	12
Merrut	9
Moradabad	5
Bulandshahr	2

Observation:

The highest purchases were made from Delhi followed by Greater Noida, Noida and Bangalore and Bulandshahr with the least number of purchases

## Univariate analysis: Experience of the Customer

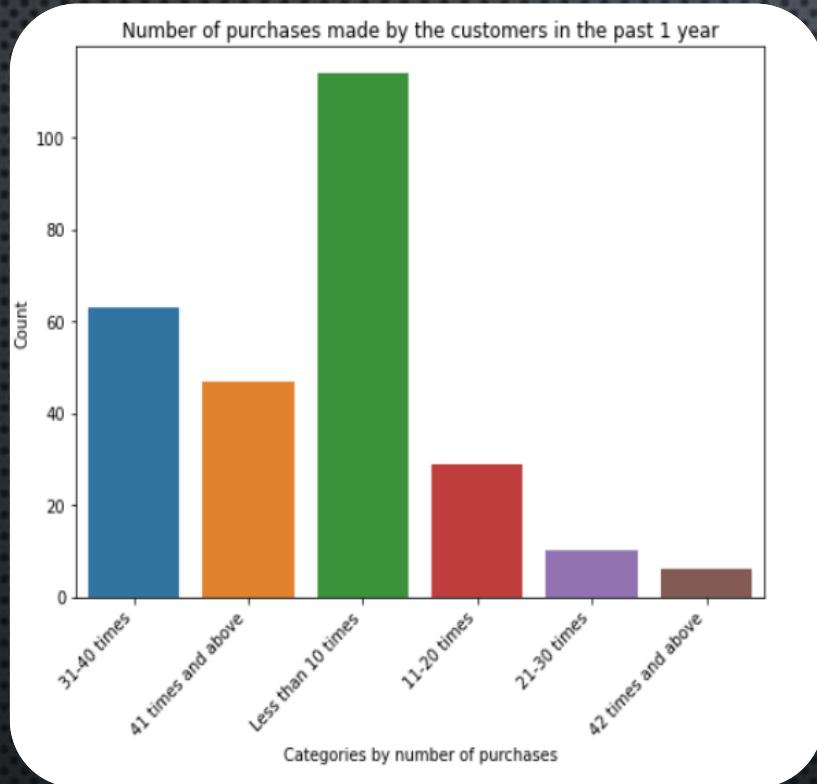


Above 4 years	98
2-3 years	65
3-4 years	47
Less than 1 year	43
1-2 years	16

Observation:

50% of the customers have experience in online purchases more than 3 years

## Univariate analysis: Experience of the Customer in last 1 year

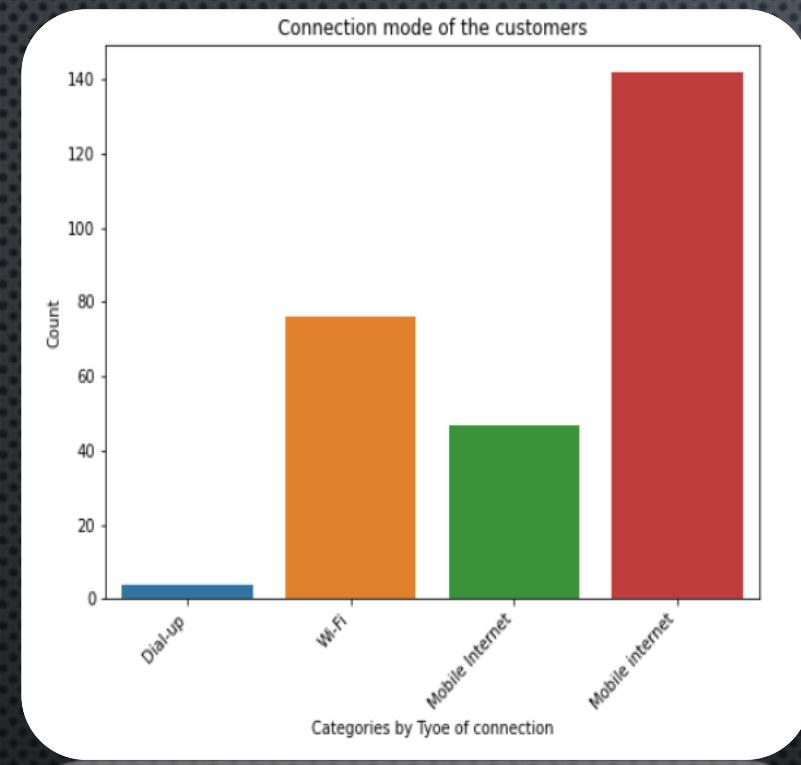


Less than 10 times	114
31-40 times	63
41 times and above	47
11-20 times	29
21-30 times	10
42 times and above	6

Observation:

57% of people purchased less than 30 times over the last year

## Univariate analysis: Connection mode used by the Customer

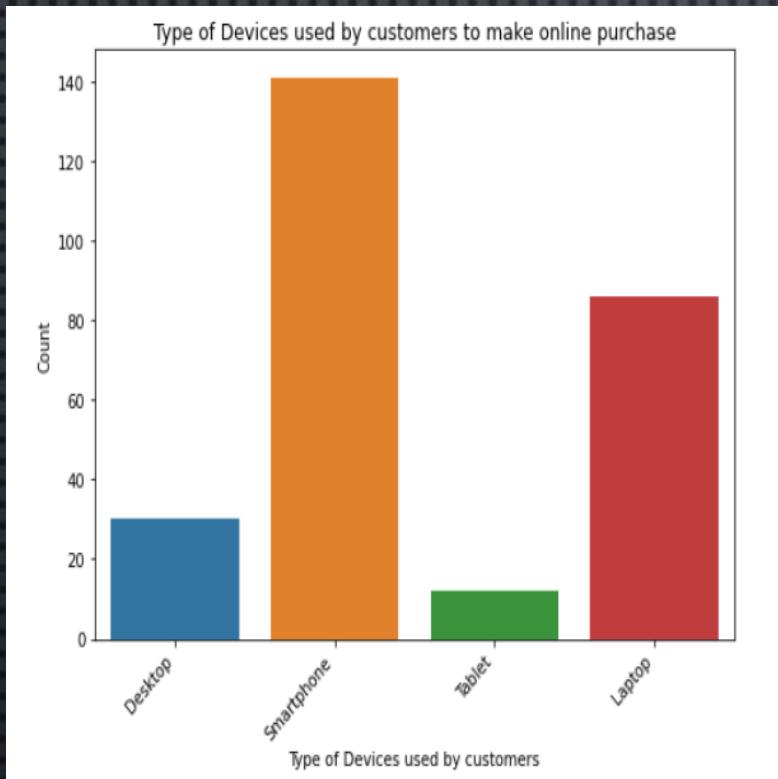


Mobile internet	142
Wi-Fi	76
Mobile Internet	47
Dial-up	4

Observation:

70% of customers used mobile internet during their purchases 28% used WIFI and 2% used dialup connections to make purchases

## Univariate analysis: Devices used by the Customer

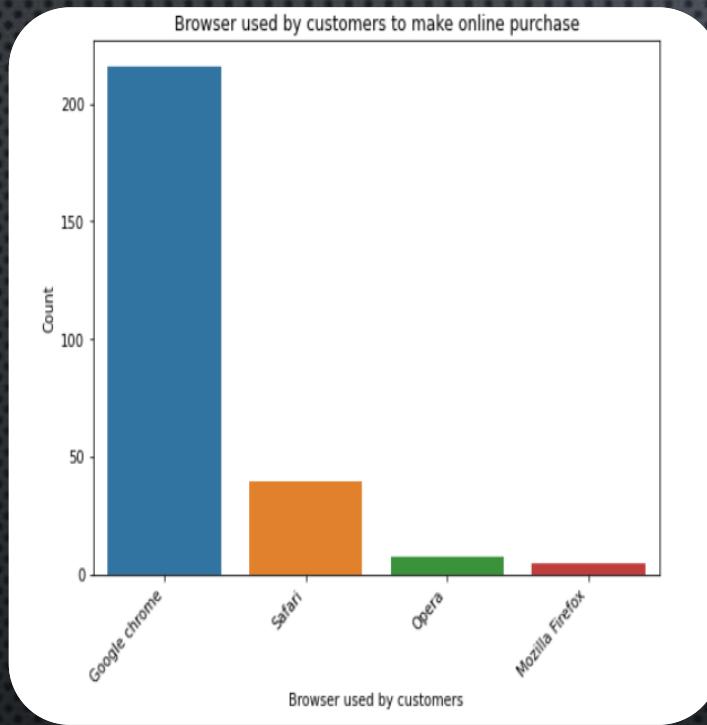


Smartphone	141
Laptop	86
Desktop	30
Tablet	12

Observation:

57% of purchases were through mobiles (52%) and tablets (5%) and rest 43% was made through laptops and desktops

## Univariate analysis: Browser used by the Customer

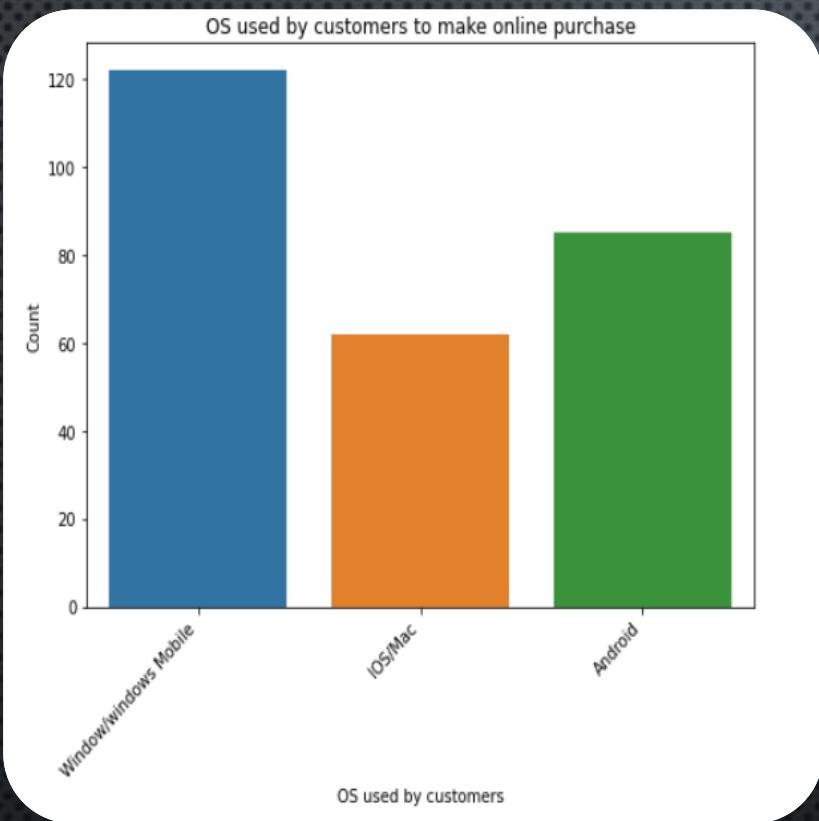


Google chrome	216
Safari	40
Opera	8
Mozilla Firefox	5

Observation:

81% of purchases were done through Google Chrome, 15% through Safari Browser and 4% through Opera and Firefox

## Univariate analysis: OS used by the customers



Window/windows Mobile	122
Android	85
IOS/Mac	62

Observation:

45% of customers used Windows/windows devices to make purchases 32% through Android and 23%through IOS/MAC

## Univariate analysis: Time spent by the customers before making the purchase

more than 15 mins	123
6-10 mins	71
11-15 mins	46
Less than 1 min	15
1-5 mins	14

Observation:

63% of Customers spend a minimum of 10 mins on the website/application browsing before making the actual purchase while 11% spend less than 5 mins.

## Univariate analysis: Preferred payment method of the customers

Credit/Debit cards	148
Cash on delivery (CoD)	76
E-wallets (Paytm, Freecharge etc.)	45

Observation:

55% of the customers prefer to use either credit or debit cards to make purchase while 28% chose Cod as an option and remaining 17% use wallets.

### **Univariate analysis: Importance of recommended similar product info**

Strongly agree (5)	116
Agree (4)	92
Indifferent (3)	43
Dis-agree (2)	18

Observation:

77% of the customers feel that the products for comparison should have complete info of that product

### **Univariate analysis: Preferred payment method of the customers**

Strongly agree (5)	164
Agree (4)	80
Strongly disagree (1)	18
Indifferent (3)	7

Observation:

91% of the customers feel that the User Interface i.e., the contents of the web page must be easy to read and understand

## Univariate analysis: Importance of seller and complete product info

Agree (4)	101
Strongly agree (5)	87
Indifferent (3)	52
Dis-agree (2)	18
Strongly disagree (1)	11

Observation:

70% of the customers believe that they would prefer the complete details f the seller and the product to be listed before they decide to make any purchase. Out of this 46% of them strongly believe that these details are very much needed for making a purchase.

## Univariate analysis: Speed of the website/application response

Strongly agree (5)	115
Agree (4)	112
Dis-agree (2)	18
Strongly disagree (1)	12
Indifferent (3)	12

Observation:

84% of the customers feel that the loading and the processing time of the website or the application has an impact on the purchase decision. A slow website/application brings down the interest of the customer.

### **Univariate analysis: Easy Payments**

Strongly agree (5)	159
Agree (4)	80
Dis-agree (2)	30

Observation:

89% of the customers feel that easy and convenient payment methods have a high impact on purchases. Easier the payment methods higher the excitement of the customer towards purchases

### **Univariate analysis: Privacy of the Customer Information**

Strongly agree (5)	185
Agree (4)	58
indifferent (3)	26

Observation:

90% of the customers are of the opinion that they tend to purchase products from sellers who are obliged to maintain the privacy of the customer

### **Univariate analysis: Service availability of the Seller**

Strongly agree (5)	149
Agree (4)	94
indifferent (3)	15
Strongly disagree (1)	11

Observation:

90% of the customers are of the opinion that they tend to purchase products from sellers who are able to provide service through various modes of communication.

### **Univariate analysis: Convenience and flexibility of Online Shopping**

Strongly agree (5)	146
Agree (4)	78
indifferent (3)	33
Dis-agree (2)	12

Observation:

83% of the customers fee that they prefer to shop online because it is more convenient and flexible

### **Univariate analysis: Quality of the website**

Strongly agree (5)	175
Agree (4)	86
Dis-agree (2)	8

Observation:

97% of the customers feel that the quality of the website has more effect on the purchases. A good quality website attracts more customers and thereby encourage them to make purchases

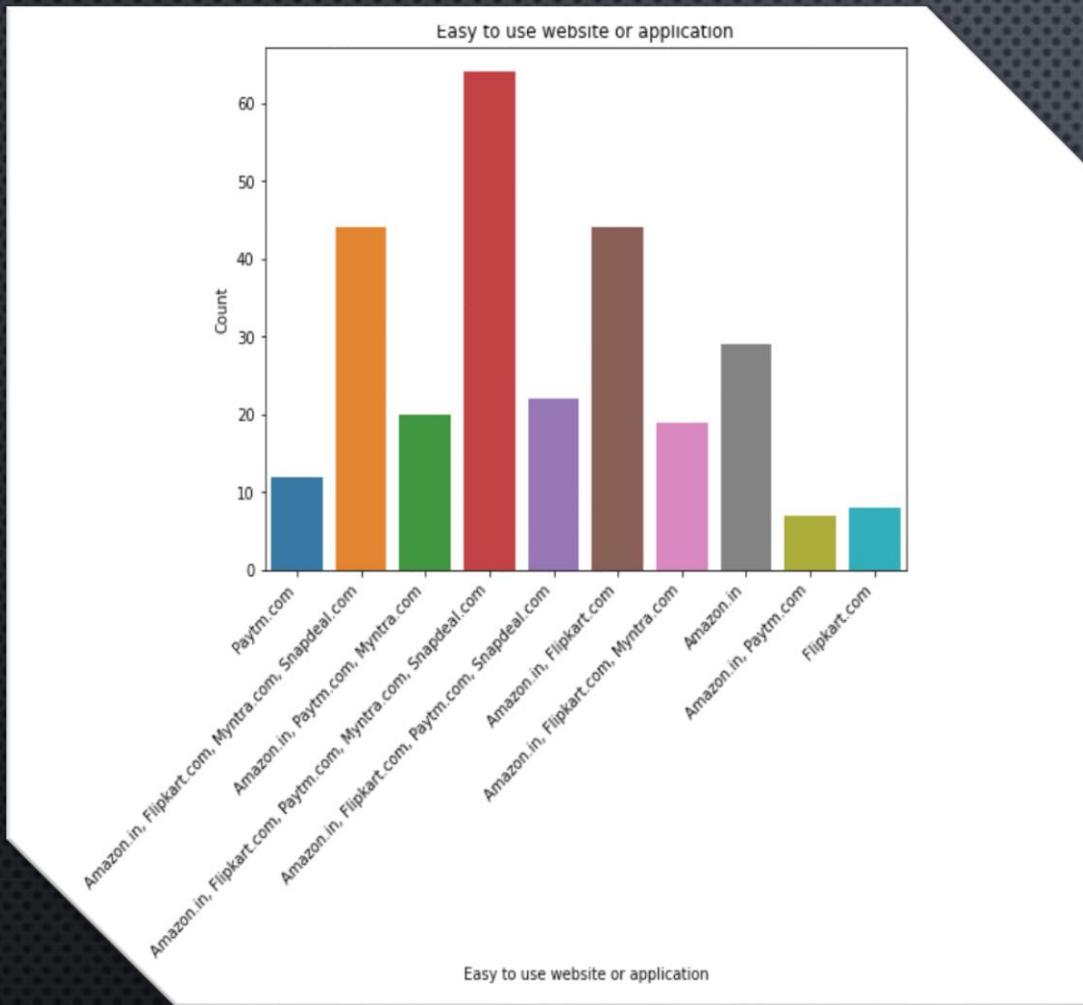
### **Univariate analysis: Effect of wide variety of products in each category**

Strongly agree (5)	111
Agree (4)	94
indifferent (3)	57
Dis-agree (2)	7

Observation:

76% of customers need more varieties in each category of products

## Univariate analysis: Service availability of the Seller

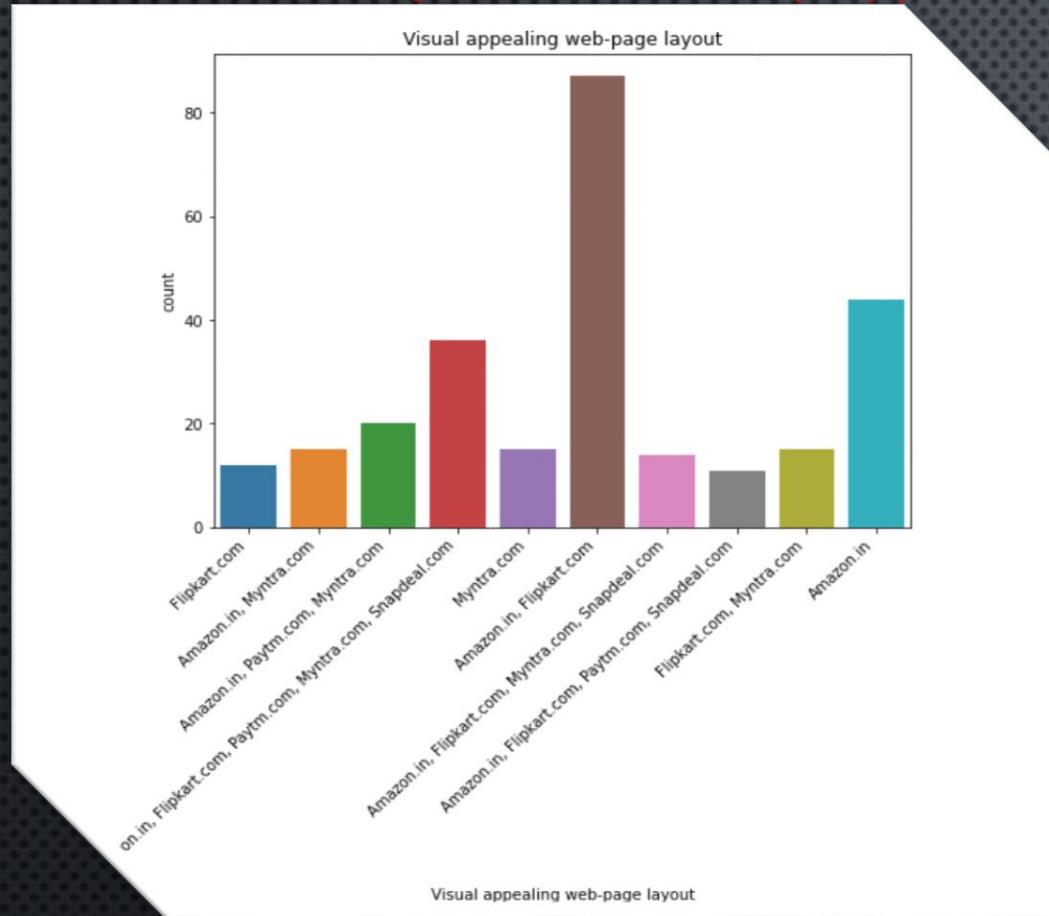


amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com	64
Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com	44
Amazon.in, Flipkart.com	44
Amazon.in	29
Amazon.in, Flipkart.com, Paytm.com, Snapdeal.com	22
Amazon.in, Paytm.com, Myntra.com	20
Amazon.in, Flipkart.com, Myntra.com	19
Paytm.com	12
Flipkart.com	8
Amazon.in, Paytm.com	7

Observation:

Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com are the easiest to use websites

## Univariate analysis: Most Visually appealing web-page layout

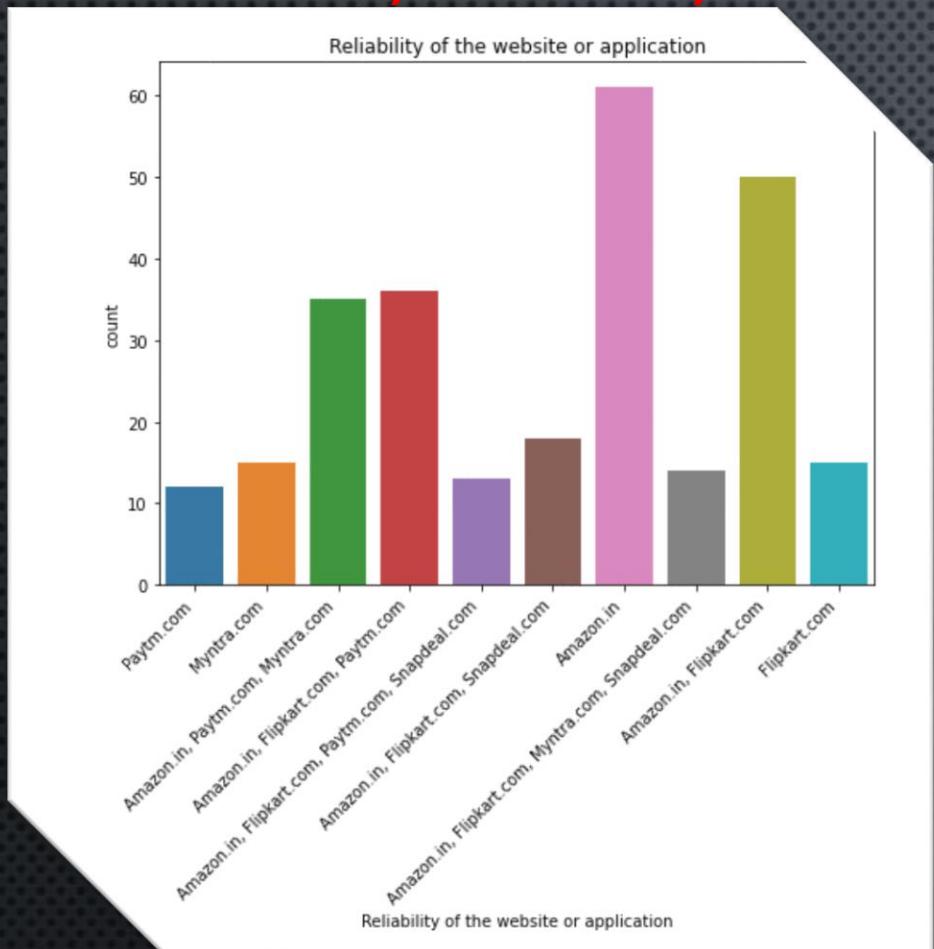


Amazon.in, Flipkart.com	87
Amazon.in	44
Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com	36
Amazon.in, Paytm.com, Myntra.com	20
Amazon.in, Myntra.com	15
Myntra.com	15
Flipkart.com, Myntra.com	15
Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com	14
Flipkart.com	12
Amazon.in, Flipkart.com, Paytm.com, Snapdeal.com	11

Observation:

Amazon.in, Flipkart.com are the website which are more visually appealing to the customers

## Univariate analysis: Reliability of the website / application

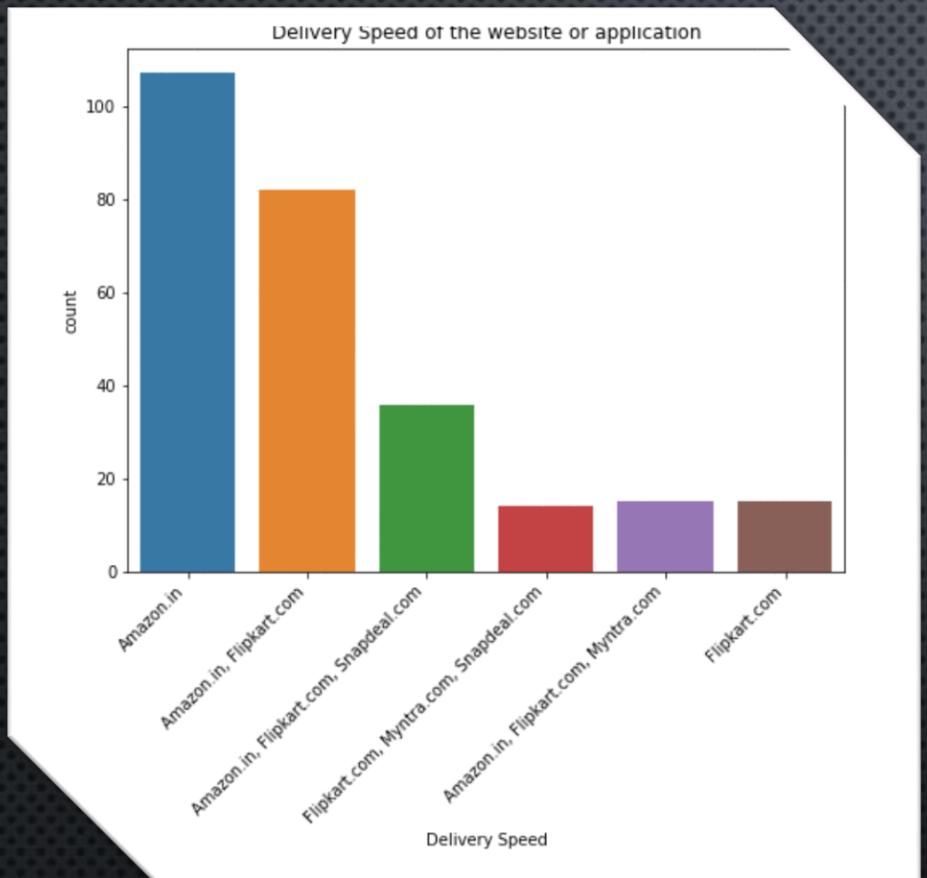


Amazon.in	61
Amazon.in, Flipkart.com	50
Amazon.in, Flipkart.com, Paytm.com	36
Amazon.in, Paytm.com, Myntra.com	35
Amazon.in, Flipkart.com, Snapdeal.com	18
Myntra.com	15
Flipkart.com	15
Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com	14
Amazon.in, Flipkart.com, Paytm.com, Snapdeal.com	13
Paytm.com	12

Observation:

Most of the customers feel that Amazon is the most reliable website to shop with followed by Flipkart and Paytm

## Univariate analysis: Delivery Speed of the Seller

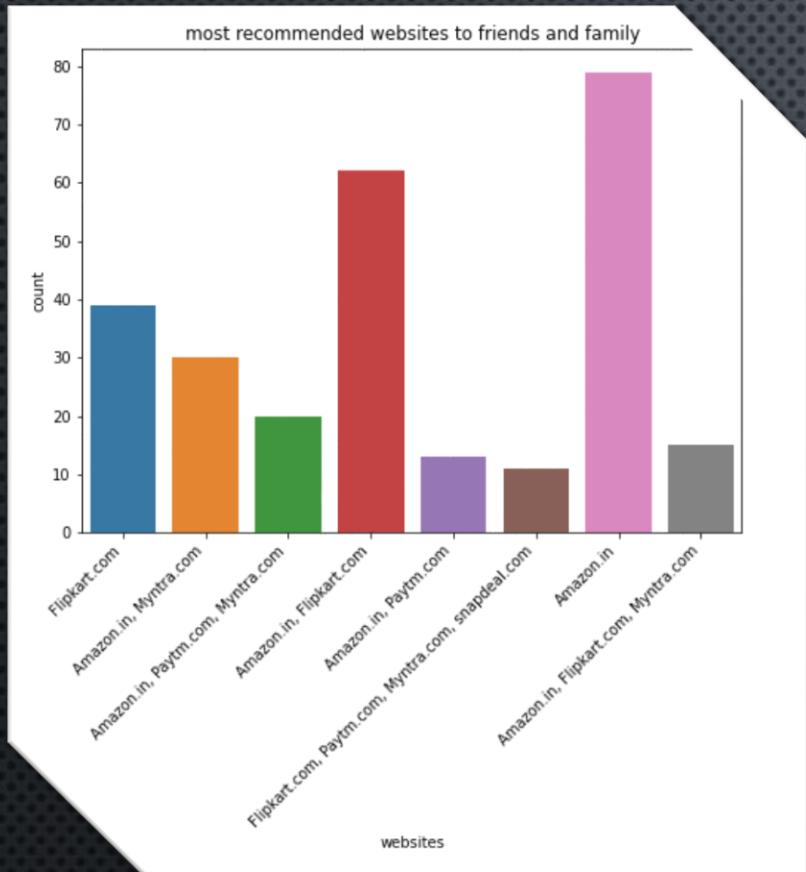


Amazon.in	107
Amazon.in, Flipkart.com	82
Amazon.in, Flipkart.com, Snapdeal.com	36
Amazon.in, Flipkart.com, Myntra.com	15
Flipkart.com	15
Flipkart.com, Myntra.com, Snapdeal.com	14

Observation:

Most of them are of the opinion that Amazon has the highest delivery speed among all the websites

## Univariate analysis: Recommendation to Friends/Family

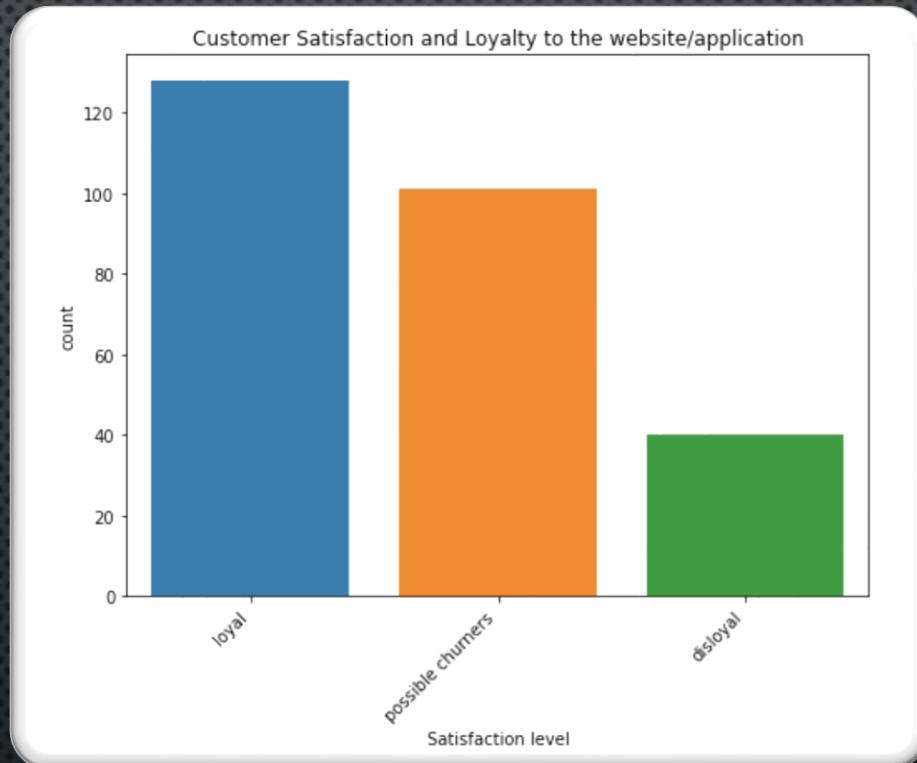


Amazon.in	79
Amazon.in, Flipkart.com	62
Flipkart.com	39
Amazon.in, Myntra.com	30
Amazon.in, Paytm.com, Myntra.com	20
Amazon.in, Flipkart.com, Myntra.com	15
Amazon.in, Paytm.com	13
Flipkart.com, Paytm.com, Myntra.com, snapdeal.com	11

Observation:

Amazon is the website that the customers feel that they would recommend it to their friends and family.

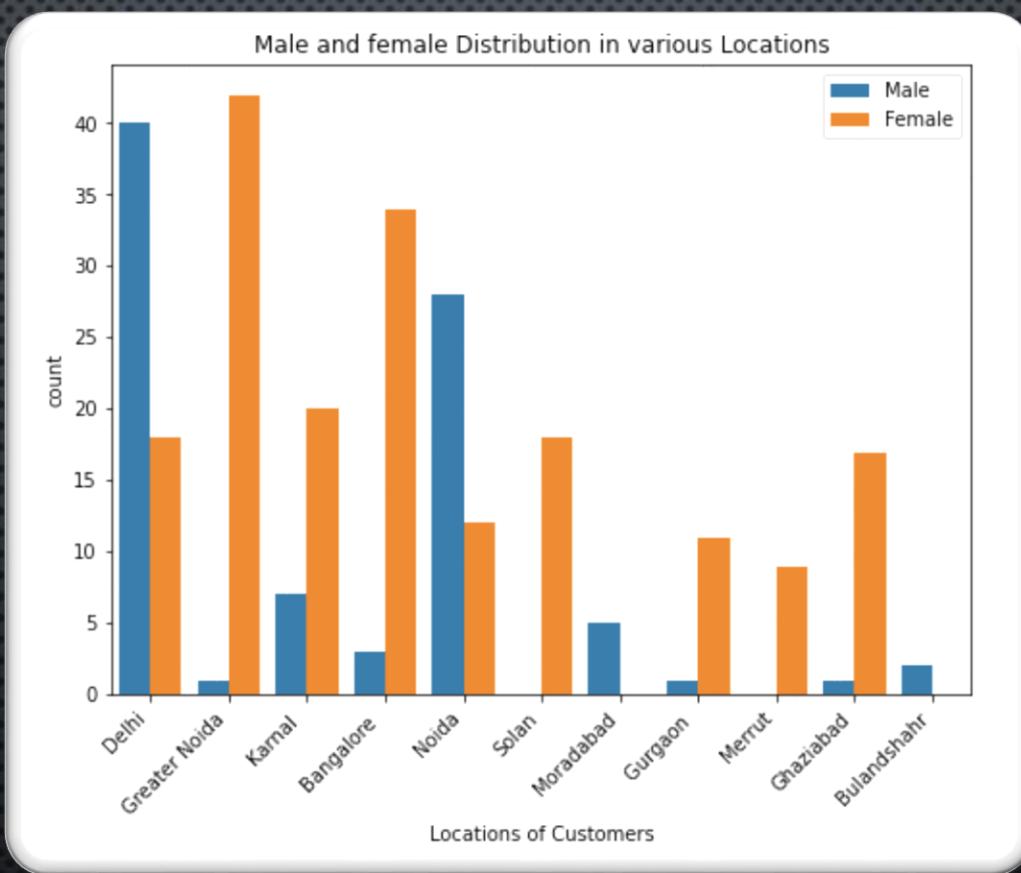
## Segmentation of Customers based on their Loyalty to the App/ Website/ Seller



Observation:

I have segmented the customers based on their loyalty. Anyone with satisfaction level 4 and above are loyal, anyone with satisfaction level 2 and below are disloyal and anyone with satisfaction level 3 as possible churners

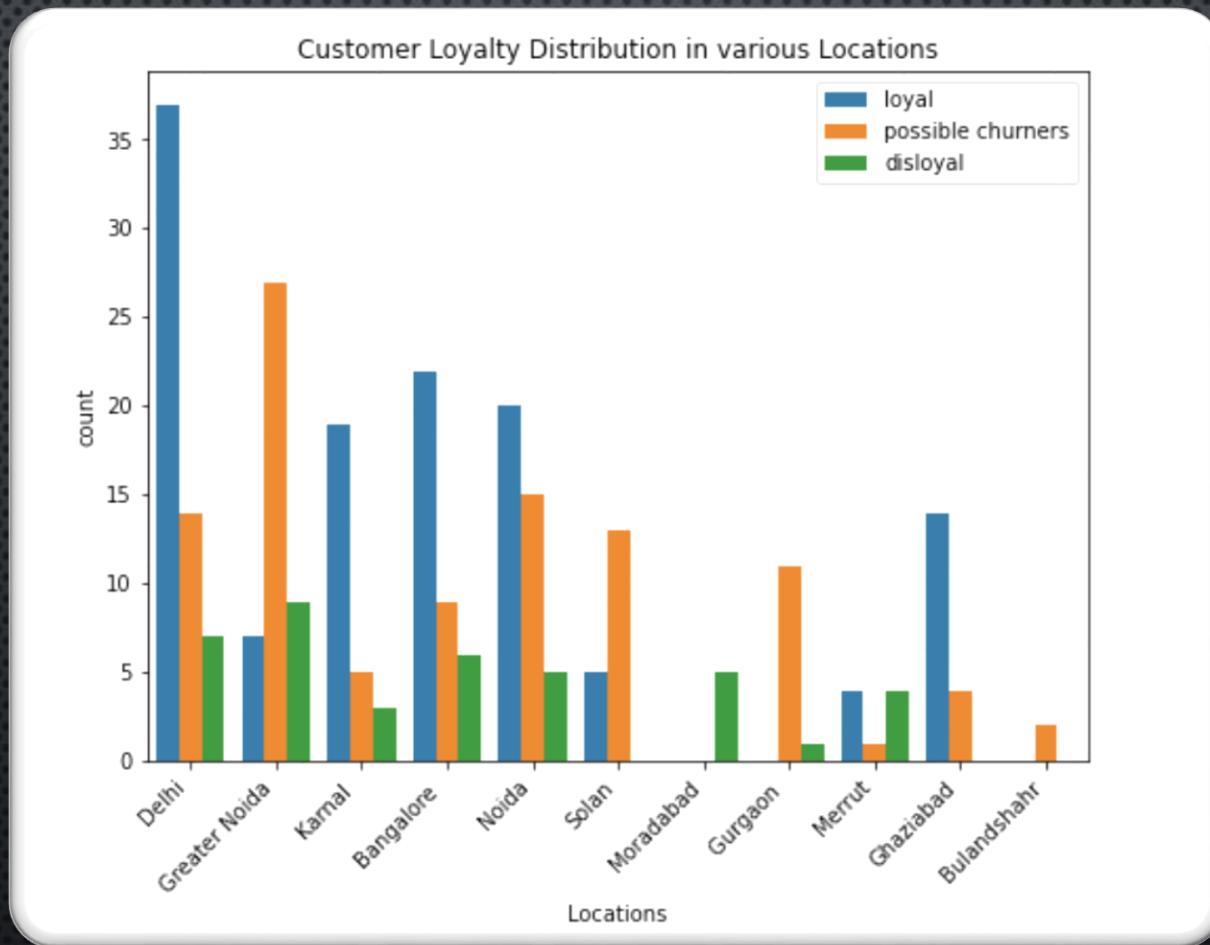
## Distribution of Customers gender over different Locations



Observation:

The number of female customers is on a higher note in all locations except Delhi, Noida, Moradabad and Bulandshahr

## Distribution of Customers Loyalty over different Locations



# CONCLUSION

- CUSTOMER RETENTION IS THE MOST EFFECTIVE WAY TO GROW A BUSINESS AND IT IS ONE OF THE KEYS TO SUCCESS.
- ACQUIRING NEW CUSTOMERS ARE ALWAYS MORE EXPENSIVE AND TAKES MORE EFFORT THAN RETAINING THEM.
- THEREFORE IT IS CRUCIAL FOR BUSINESS, TO TRY TO KEEP THE RETENTION RATE UP, WITH THE TIPS MENTIONED HERE OR MORE, IN ORDER TO HAVE BETTER CONVERSION RATES, LOWER MARKETING COSTS, AND HIGHER PROFITS.

# Limitations of this dataset and Scope for Future Work

- THE SIZE OF THE DATASET BEING VERY SMALL, MAY LEAD TO BIAS OF THE INSIGHTS DRAWN.
- THE DATASET HAS ONLY 269 OBSERVATIONS COLLECTED OVER A LARGE GEOGRAPHICAL AREA (ENTIRE COUNTRY).
- IF WE TRY TO MAKE DATA DRIVEN DECISIONS FROM THE GIVEN DATASET, WE MIGHT NOT BE ABLE TO COME TO A PROPER CONCLUSION WHICH EXACTLY REFLECTS THE ACTUAL REALITY.

## Scope for future work:

- COLLECT MORE SAMPLES EVENLY COVERING THE ENTIRE GEOGRAPHICAL AREA TO MAKE SURE THAT THE DATASET IS NOT IMBALANCED