

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

```
In [2]: !pip install openpyxl
```

Requirement already satisfied: openpyxl in c:\users\rakesh lodem\anaconda3\lib\site-packages (3.0.7)
Requirement already satisfied: et-xmlfile in c:\users\rakesh lodem\anaconda3\lib\site-packages (from openpyxl) (1.0.1)

```
In [3]: df=pd.read_excel(r'C:\Users\RAKESH~1\AppData\Local\Temp\Rar$Di10956.44429\customer_retention_dataset.xlsx')
df.head()
```

Out[3]:

	1 Gender 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)
0	Male	31-40 years	Delhi	110009	Above 4 years	31-40 times	Dial-up	Desktop	Others	Window/windows Mobile	...	Amazon.in	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	Myntra.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
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
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

5 rows × 71 columns

```
In [4]: #Setting option to show max rows and max columns
pd.set_option("display.max_columns",None)
pd.set_option("display.max_rows", None)
```

```
In [5]: ## preprocessing the column names
```

```
In [6]: from string import digits

#Removing tab spaces
df.columns = df.columns.str.replace('\t','')

#Removing digits
remove_digits = str.maketrans('', '', digits)
df.columns = df.columns.str.translate(remove_digits)

#Removing leading and trailing spaces
df.columns = df.columns.str.strip()
```

```
In [7]: df.head()
```

Out[7]:

	Gender of respondent	How old are you?	Which city do you shop online from?	What is the Pin Code of where you shop	Since How Long You are Shopping Online ?	How many times you have made an online purchase in the	How do you access the internet while shopping	Which device do you use to access the online shopping?	What is the screen size of your mobile	What is the operating system (OS) of your device?	What browser do you run on your device to access	Which channel did you follow to arrive at your favorite online	After first visit, how do you reach the online retail
--	----------------------	------------------	-------------------------------------	--	--	--	---	--	--	---	--	--	---

				online from?		past year?	on-line?		device?		the website?	store for the first time?	store?
0	Male	31-40 years	Delhi	110009	Above 4 years	31-40 times	Dial-up	Desktop	Others	Window/windows Mobile	Google chrome	Search Engine	Search Engine
1	Female	21-30 years	Delhi	110030	Above 4 years	41 times and above	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	Google chrome	Search Engine	Via application
2	Female	21-30 years	Greater Noida	201308	3-4 years	41 times and above	Mobile Internet	Smartphone	5.5 inches	Android	Google chrome	Search Engine	Via application
3	Male	21-30 years	Karnal	132001	3-4 years	Less than 10 times	Mobile Internet	Smartphone	5.5 inches	IOS/Mac	Safari	Search Engine	Search Engine
4	Female	21-30 years	Bangalore	530068	2-3 years	11-20 times	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	Safari	Content Marketing	Via application

In [8]: `df.shape`

Out[8]: (269, 71)

In [9]: `df.dtypes`

Out[9]: Gender of respondent obje
 ct
 How old are you? obje
 ct
 Which city do you shop online from? obje
 ct
 What is the Pin Code of where you shop online from? int
 64
 Since How Long You are Shopping Online ? obje
 ct
 How many times you have made an online purchase in the past year? obje
 ct
 How do you access the internet while shopping on-line? obje
 ct
 Which device do you use to access the online shopping? obje
 ct
 What is the screen size of your mobile device? obje
 ct
 What is the operating system (OS) of your device? obje
 ct
 What browser do you run on your device to access the website? obje
 ct
 Which channel did you follow to arrive at your favorite online store for the first time? obje
 ct
 After first visit, how do you reach the online retail store? obje
 ct
 How much time do you explore the e- retail store before making a purchase decision? obje
 ct
 What is your preferred payment Option? obje
 ct
 How frequently do you abandon (selecting an items and leaving without making payment) your shopping cart? obje
 ct
 Why did you abandon the “Bag”, “Shopping Cart”? obje
 ct
 The content on the website must be easy to read and understand obje
 ct
 Information on similar product to the one highlighted is important for product comparison obje
 ct
 Complete information on listed seller and product being offered is important for purchase decision. obje
 ct
 All relevant information on listed products must be stated clearly obje
 ct
 Ease of navigation in website obje
 ct
 Loading and processing speed obje
 ct

User friendly Interface of the website	obje
ct	
Convenient Payment methods	obje
ct	
Trust that the online retail store will fulfill its part of the transaction at the stipulated time	obje
ct	
Empathy (readiness to assist with queries) towards the customers	obje
ct	
Being able to guarantee the privacy of the customer	obje
ct	
Responsiveness, availability of several communication channels (email, online rep, twitter, phone etc.)	obje
ct	
Online shopping gives monetary benefit and discounts	obje
ct	
Enjoyment is derived from shopping online	obje
ct	
Shopping online is convenient and flexible	obje
ct	
Return and replacement policy of the e-tailer is important for purchase decision	obje
ct	
Gaining access to loyalty programs is a benefit of shopping online	obje
ct	
Displaying quality Information on the website improves satisfaction of customers	obje
ct	
User derive satisfaction while shopping on a good quality website or application	obje
ct	
Net Benefit derived from shopping online can lead to users satisfaction	obje
ct	
User satisfaction cannot exist without trust	obje
ct	
Offering a wide variety of listed product in several category	obje
ct	
Provision of complete and relevant product information	obje
ct	
Monetary savings	obje
ct	
The Convenience of patronizing the online retailer	obje
ct	
Shopping on the website gives you the sense of adventure	obje
ct	
Shopping on your preferred e-tailer enhances your social status	obje
ct	
You feel gratification shopping on your favorite e-tailer	obje
ct	
Shopping on the website helps you fulfill certain roles	obje
ct	
Getting value for money spent	obje
ct	
From the following, tick any (or all) of the online retailers you have shopped from;	obje
ct	
Easy to use website or application	obje
ct	
Visual appealing web-page layout	obje
ct	
Wild variety of product on offer	obje
ct	
Complete, relevant description information of products	obje
ct	
Fast loading website speed of website and application	obje
ct	
Reliability of the website or application	obje
ct	
Quickness to complete purchase	obje
ct	
Availability of several payment options	obje
ct	
Speedy order delivery	obje
ct	
Privacy of customers' information	obje
ct	
Security of customer financial information	obje
ct	
Perceived Trustworthiness	obje
ct	
Presence of online assistance through multi-channel	obje
ct	
Longer time to get logged in (promotion, sales period)	obje
ct	
Longer time in displaying graphics and photos (promotion, sales period)	obje
ct	
Late declaration of price (promotion, sales period)	obje
ct	
Longer page loading time (promotion, sales period)	obje

ct	
Limited mode of payment on most products (promotion, sales period)	obje
ct	
Longer delivery period	obje
ct	
Change in website/Application design	obje
ct	
Frequent disruption when moving from one page to another	obje
ct	
Website is as efficient as before	obje
ct	
Which of the Indian online retailer would you recommend to a friend?	obje
ct	
dtype: object	

```
In [10]: df.isnull().sum()
```

```
Out[10]: Gender of respondent                                0
How old are you?                                           0
Which city do you shop online from?                       0
What is the Pin Code of where you shop online from?       0
Since How Long You are Shopping Online ?                  0
How many times you have made an online purchase in the past year? 0
How do you access the internet while shopping on-line?    0
Which device do you use to access the online shopping?    0
What is the screen size of your mobile device?            0
What is the operating system (OS) of your device?         0
What browser do you run on your device to access the website? 0
Which channel did you follow to arrive at your favorite online store for the first time? 0
After first visit, how do you reach the online retail store? 0
How much time do you explore the e- retail store before making a purchase decision? 0
What is your preferred payment Option?                    0
How frequently do you abandon (selecting an items and leaving without making payment) your shopping cart? 0
Why did you abandon the "Bag", "Shopping Cart"?           0
The content on the website must be easy to read and understand 0
Information on similar product to the one highlighted is important for product comparison 0
Complete information on listed seller and product being offered is important for purchase decision. 0
All relevant information on listed products must be stated clearly 0
Ease of navigation in website                             0
Loading and processing speed                               0
User friendly Interface of the website                    0
Convenient Payment methods                                0
Trust that the online retail store will fulfill its part of the transaction at the stipulated time 0
Empathy (readiness to assist with queries) towards the customers 0
Being able to guarantee the privacy of the customer       0
Responsiveness, availability of several communication channels (email, online rep, twitter, phone etc.) 0
Online shopping gives monetary benefit and discounts      0
Enjoyment is derived from shopping online                 0
Shopping online is convenient and flexible                0
Return and replacement policy of the e-tailer is important for purchase decision 0
Gaining access to loyalty programs is a benefit of shopping online 0
Displaying quality Information on the website improves satisfaction of customers 0
User derive satisfaction while shopping on a good quality website or application 0
Net Benefit derived from shopping online can lead to users satisfaction 0
User satisfaction cannot exist without trust              0
Offering a wide variety of listed product in several category 0
Provision of complete and relevant product information    0
Monetary savings                                          0
The Convenience of patronizing the online retailer        0
Shopping on the website gives you the sense of adventure 0
Shopping on your preferred e-tailer enhances your social status 0
You feel gratification shopping on your favorite e-tailer 0
Shopping on the website helps you fulfill certain roles   0
Getting value for money spent                             0
From the following, tick any (or all) of the online retailers you have shopped from; 0
Easy to use website or application                       0
Visual appealing web-page layout                          0
Wild variety of product on offer                          0
Complete, relevant description information of products    0
Fast loading website speed of website and application    0
Reliability of the website or application                 0
Quickness to complete purchase                           0
Availability of several payment options                   0
Speedy order delivery                                    0
Privacy of customers' information                         0
Security of customer financial information                 0
Perceived Trustworthiness                                0
Presence of online assistance through multi-channel      0
Longer time to get logged in (promotion, sales period)   0
Longer time in displaying graphics and photos (promotion, sales period) 0
```

Late declaration of price (promotion, sales period)	0
Longer page loading time (promotion, sales period)	0
Limited mode of payment on most products (promotion, sales period)	0
Longer delivery period	0
Change in website/Application design	0
Frequent disruption when moving from one page to another	0
Website is as efficient as before	0
Which of the Indian online retailer would you recommend to a friend?	0
dtype: int64	

```
In [11]: df.nunique()
```

Gender of respondent	2
How old are you?	5
Which city do you shop online from?	11
What is the Pin Code of where you shop online from?	39
Since How Long You are Shopping Online ?	5
How many times you have made an online purchase in the past year?	6
How do you access the internet while shopping on-line?	4
Which device do you use to access the online shopping?	4
What is the screen size of your mobile device?	4
What is the operating system (OS) of your device?	3
What browser do you run on your device to access the website?	4
Which channel did you follow to arrive at your favorite online store for the first time?	3
After first visit, how do you reach the online retail store?	5
How much time do you explore the e- retail store before making a purchase decision?	5
What is your preferred payment Option?	3
How frequently do you abandon (selecting an items and leaving without making payment) your shopping cart?	4
Why did you abandon the "Bag", "Shopping Cart"?	5
The content on the website must be easy to read and understand	4
Information on similar product to the one highlighted is important for product comparison	4
Complete information on listed seller and product being offered is important for purchase decision.	5
All relevant information on listed products must be stated clearly	4
Ease of navigation in website	4
Loading and processing speed	5
User friendly Interface of the website	5
Convenient Payment methods	3
Trust that the online retail store will fulfill its part of the transaction at the stipulated time	4
Empathy (readiness to assist with queries) towards the customers	4
Being able to guarantee the privacy of the customer	3
Responsiveness, availability of several communication channels (email, online rep, twitter, phone etc.)	4
Online shopping gives monetary benefit and discounts	5
Enjoyment is derived from shopping online	5
Shopping online is convenient and flexible	4
Return and replacement policy of the e-tailer is important for purchase decision	3
Gaining access to loyalty programs is a benefit of shopping online	5
Displaying quality Information on the website improves satisfaction of customers	3
User derive satisfaction while shopping on a good quality website or application	3
Net Benefit derived from shopping online can lead to users satisfaction	4
User satisfaction cannot exist without trust	5
Offering a wide variety of listed product in several category	4
Provision of complete and relevant product information	4
Monetary savings	4
The Convenience of patronizing the online retailer	3
Shopping on the website gives you the sense of adventure	5
Shopping on your preferred e-tailer enhances your social status	5
You feel gratification shopping on your favorite e-tailer	5
Shopping on the website helps you fulfill certain roles	5
Getting value for money spent	3
From the following, tick any (or all) of the online retailers you have shopped from;	9
Easy to use website or application	10
Visual appealing web-page layout	10
Wild variety of product on offer	9
Complete, relevant description information of products	11
Fast loading website speed of website and application	10
Reliability of the website or application	10
Quickness to complete purchase	9
Availability of several payment options	11
Speedy order delivery	6
Privacy of customers' information	11
Security of customer financial information	11
Perceived Trustworthiness	9
Presence of online assistance through multi-channel	10
Longer time to get logged in (promotion, sales period)	10
Longer time in displaying graphics and photos (promotion, sales period)	10
Late declaration of price (promotion, sales period)	8
Longer page loading time (promotion, sales period)	11
Limited mode of payment on most products (promotion, sales period)	8
Longer delivery period	6
Change in website/Application design	7

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: int64

8
8
8

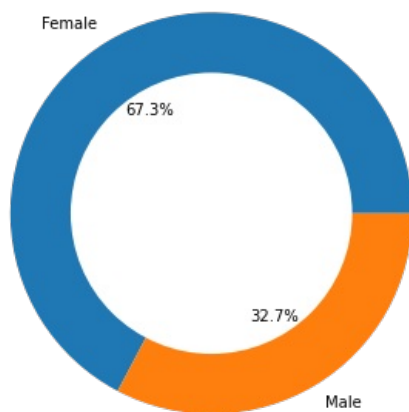
```
In [12]: ## all columns are categorical datatypes
```

```
In [14]: ## univariate analysis
```

```
In [15]: personal_info=['Gender of respondent','How old are you?','Which city do you shop online from?',  
                        'What is the Pin Code of where you shop online from?','Since How Long You are Shopping Online ?',  
                        'How many times you have made an online purchase in the past year?']
```

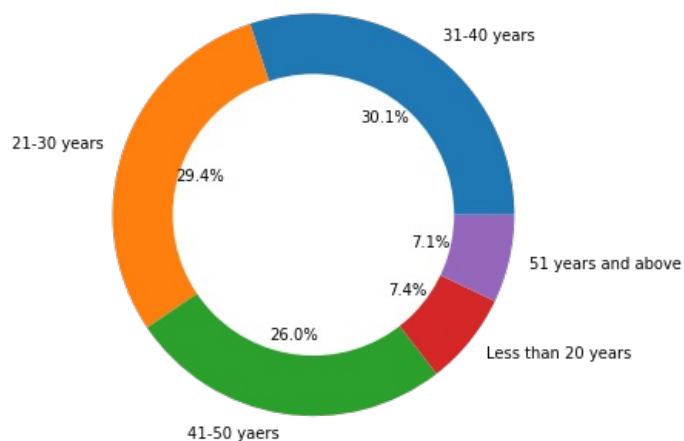
```
In [16]: ## personal info
```

```
In [17]: for i in personal_info:  
    if i!='What is the Pin Code of where you shop online from?':  
        plt.figure(figsize=(8,6))  
        df[i].value_counts().plot.pie(autopct='%1.1f%%')  
        centre=plt.Circle((0,0),0.7,fc='white')  
        fig=plt.gcf()  
        fig.gca().add_artist(centre)  
        plt.xlabel(i)  
        plt.ylabel('')  
        plt.figure()
```



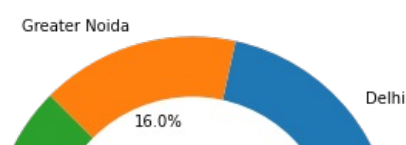
Gender of respondent

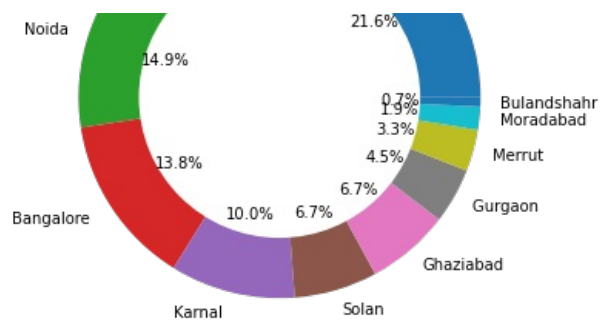
<Figure size 432x288 with 0 Axes>



How old are you?

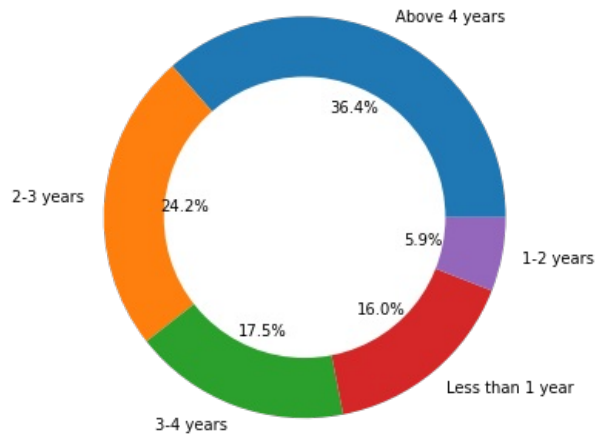
<Figure size 432x288 with 0 Axes>





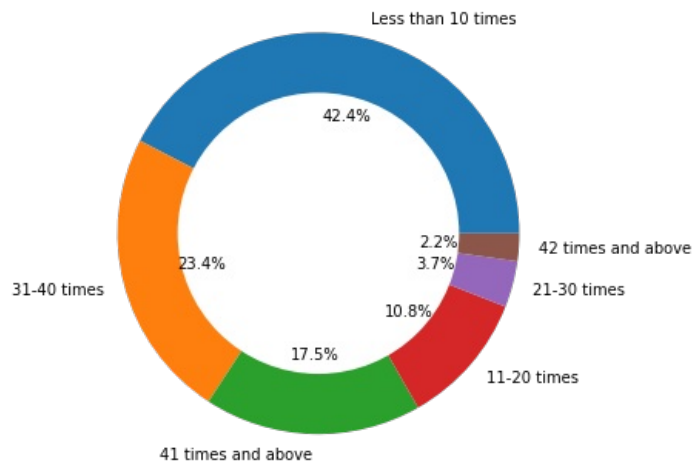
Which city do you shop online from?

<Figure size 432x288 with 0 Axes>



Since How Long You are Shopping Online ?

<Figure size 432x288 with 0 Axes>



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

<Figure size 432x288 with 0 Axes>

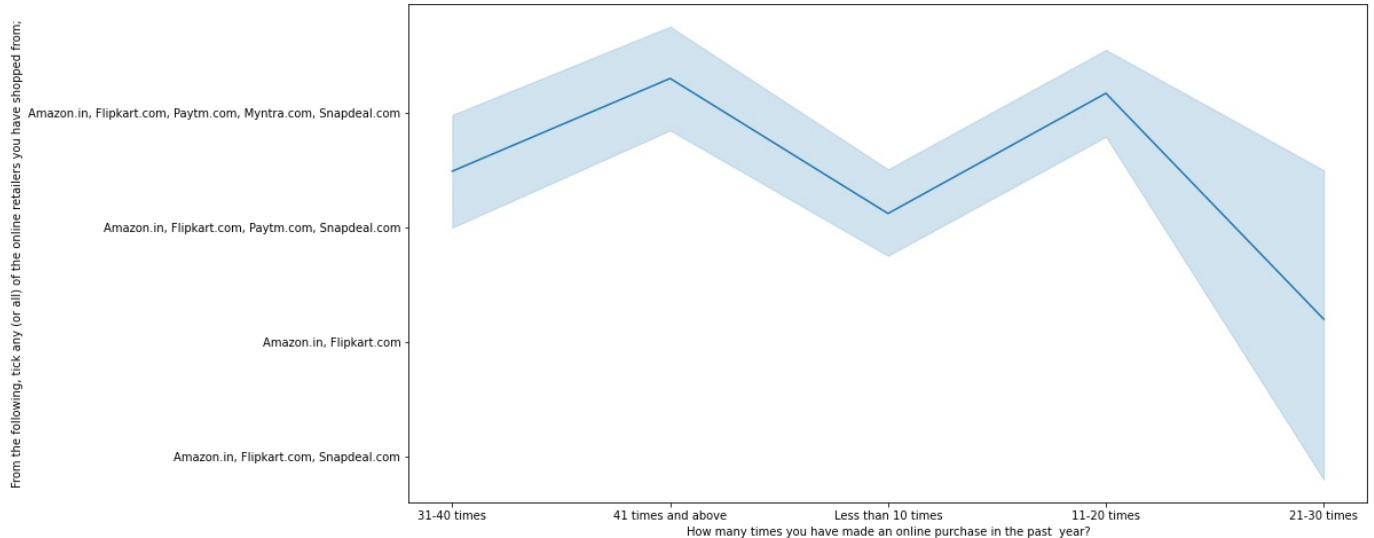
```
In [18]: ## intention of repeat purchase
```

```
In [19]: #Resolving ambiguity of column
#Changing 42 times and above to 41 times and above
df['How many times you have made an online purchase in the past year?'].replace('42 times and above','41 times and above',inplace=True)
```

```
In [20]: plt.figure(figsize=(15,8))
```

```
sns.lineplot(df['How many times you have made an online purchase in the past year?'],
            df['From the following, tick any (or all) of the online retailers you have shopped from;'])
```

Out[20]: <AxesSubplot:xlabel='How many times you have made an online purchase in the past year?', ylabel='From the following, tick any (or all) of the online retailers you have shopped from; '>



In [21]: *## converting years to numbers for better understanding and analysis*

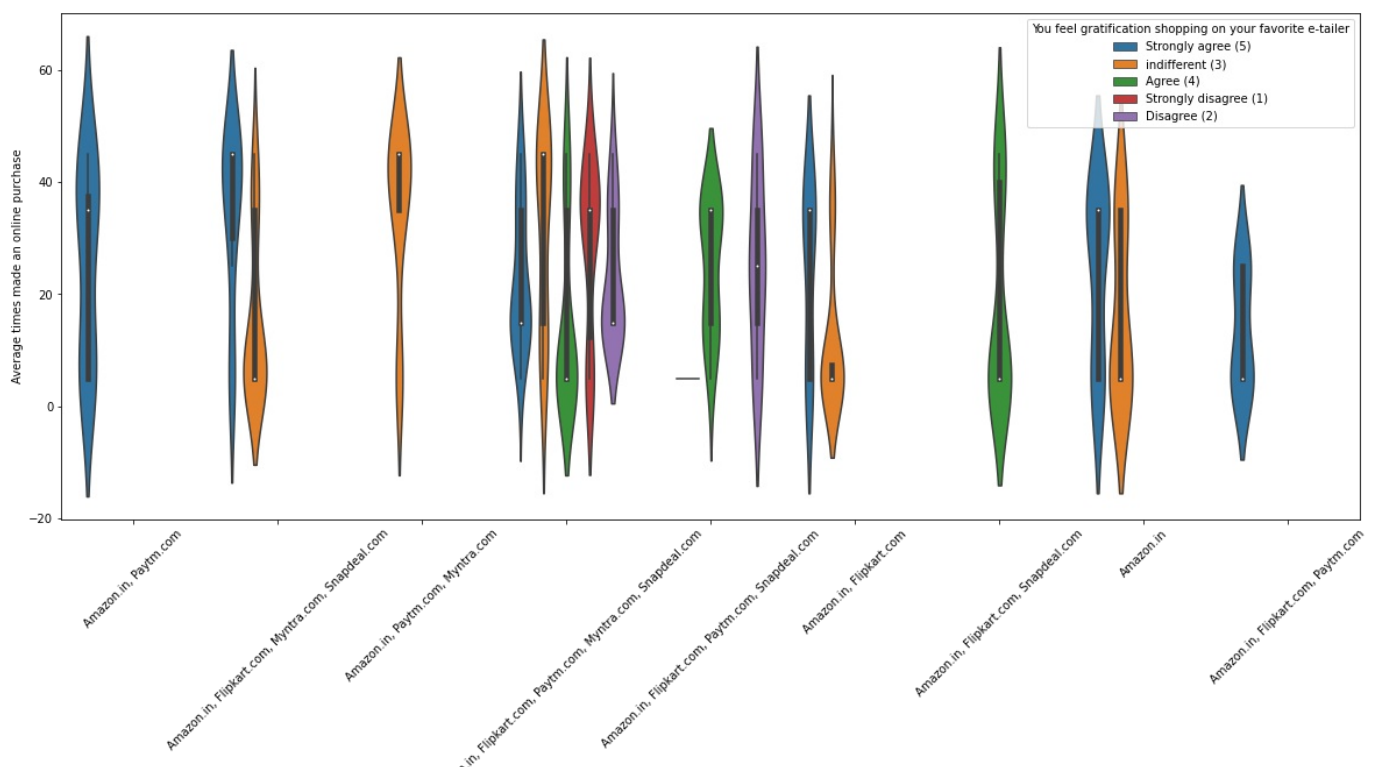
In [22]:

```
dict={'31-40 times':35,'41 times and above':45,'Less than 10 times':5,'11-20 times':15,'21-30 times':25}
df['Average times made an online purchase']=df['How many times you have made an online purchase in the past year']
```

In [23]:

```
plt.figure(figsize=(20,8))
sns.violinplot(df['From the following, tick any (or all) of the online retailers you have shopped from;'],
              df['Average times made an online purchase'],hue=df['You feel gratification shopping on your favorite e-tailer'],
              plt.xticks(rotation=45))
```

Out[23]: (array([0, 1, 2, 3, 4, 5, 6, 7, 8]),
[Text(0, 0, 'Amazon.in, Paytm.com'),
Text(1, 0, 'Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com'),
Text(2, 0, 'Amazon.in, Paytm.com, Myntra.com'),
Text(3, 0, 'Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com'),
Text(4, 0, 'Amazon.in, Flipkart.com, Paytm.com, Snapdeal.com'),
Text(5, 0, 'Amazon.in, Flipkart.com'),
Text(6, 0, 'Amazon.in, Flipkart.com, Snapdeal.com'),
Text(7, 0, 'Amazon.in'),
Text(8, 0, 'Amazon.in, Flipkart.com, Paytm.com')])



From the following, tick any (or all) of the online retailers you have shopped from;

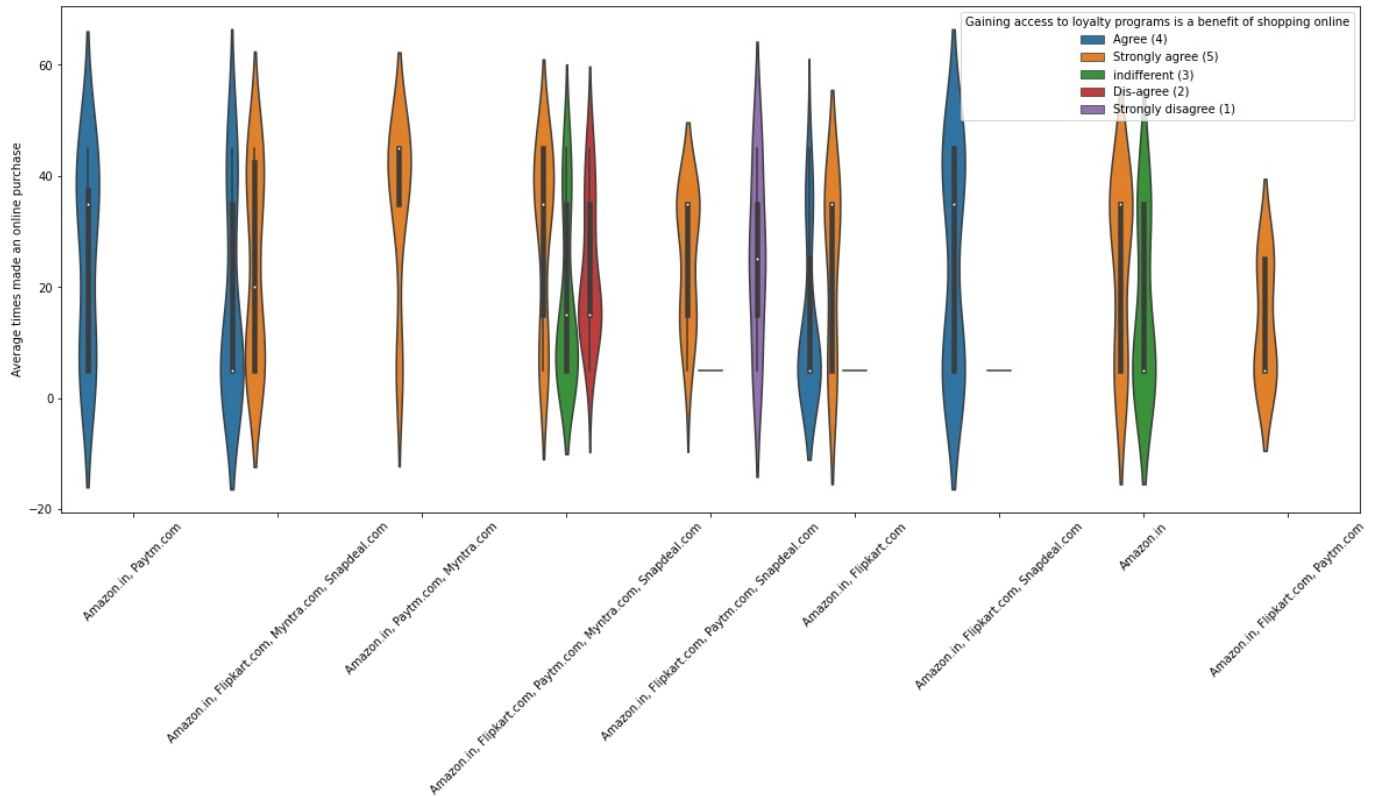
In []:

In [24]:

```
plt.figure(figsize=(20,8))
sns.violinplot(df['From the following, tick any (or all) of the online retailers you have shopped from;'],
              df['Average times made an online purchase'],hue=df['Gaining access to loyalty programs is a benefit'],
              plt.xticks(rotation=45))
```

Out[24]:

```
(array([0, 1, 2, 3, 4, 5, 6, 7, 8]),
 [Text(0, 0, 'Amazon.in, Paytm.com'),
  Text(1, 0, 'Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com'),
  Text(2, 0, 'Amazon.in, Paytm.com, Myntra.com'),
  Text(3, 0, 'Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com'),
  Text(4, 0, 'Amazon.in, Flipkart.com, Paytm.com, Snapdeal.com'),
  Text(5, 0, 'Amazon.in, Flipkart.com'),
  Text(6, 0, 'Amazon.in, Flipkart.com, Snapdeal.com'),
  Text(7, 0, 'Amazon.in'),
  Text(8, 0, 'Amazon.in, Flipkart.com, Paytm.com')])
```



From the following, tick any (or all) of the online retailers you have shopped from;

In [25]:

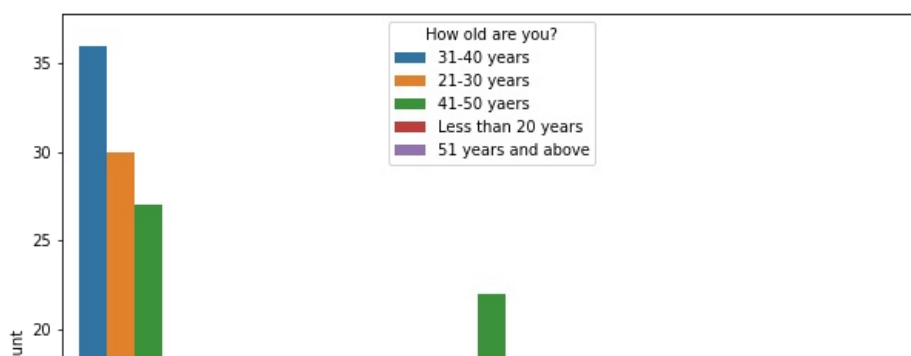
```
## online retailing
```

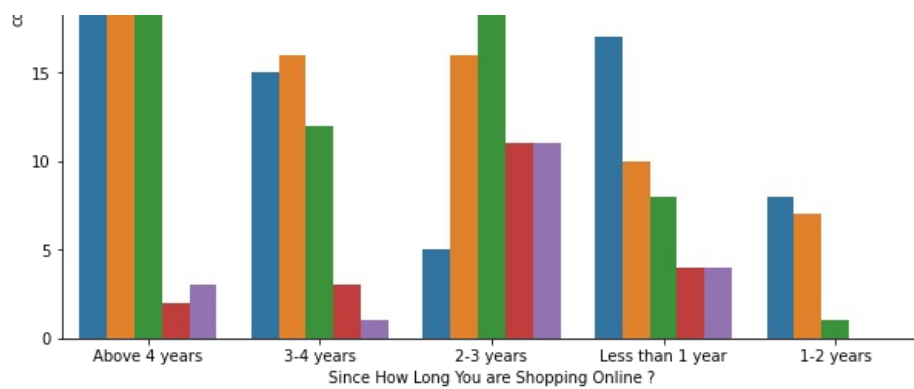
In [26]:

```
plt.figure(figsize=(10,8))
sns.countplot(df['Since How Long You are Shopping Online ?'],hue=df['How old are you?'])
```

Out[26]:

```
<AxesSubplot:xlabel='Since How Long You are Shopping Online ?', ylabel='count'>
```





In [27]: `##Highest number of people have been shopping online for above 4 years except for the age group below 20 years and
##Converting Years to numbers for better analysis`

In [28]: `df['Since How Long You are Shopping Online ?'].unique()`

Out[28]: `array(['Above 4 years', '3-4 years', '2-3 years', 'Less than 1 year',
'1-2 years'], dtype=object)`

In [29]: `dict={'Above 4 years':4.5,'3-4 years':3.5,'2-3 years':2.5,'1-2 years':1.5,'Less than 1 year':0.5}
df['Average years of shopping online']=df['Since How Long You are Shopping Online ?'].replace(dict)`

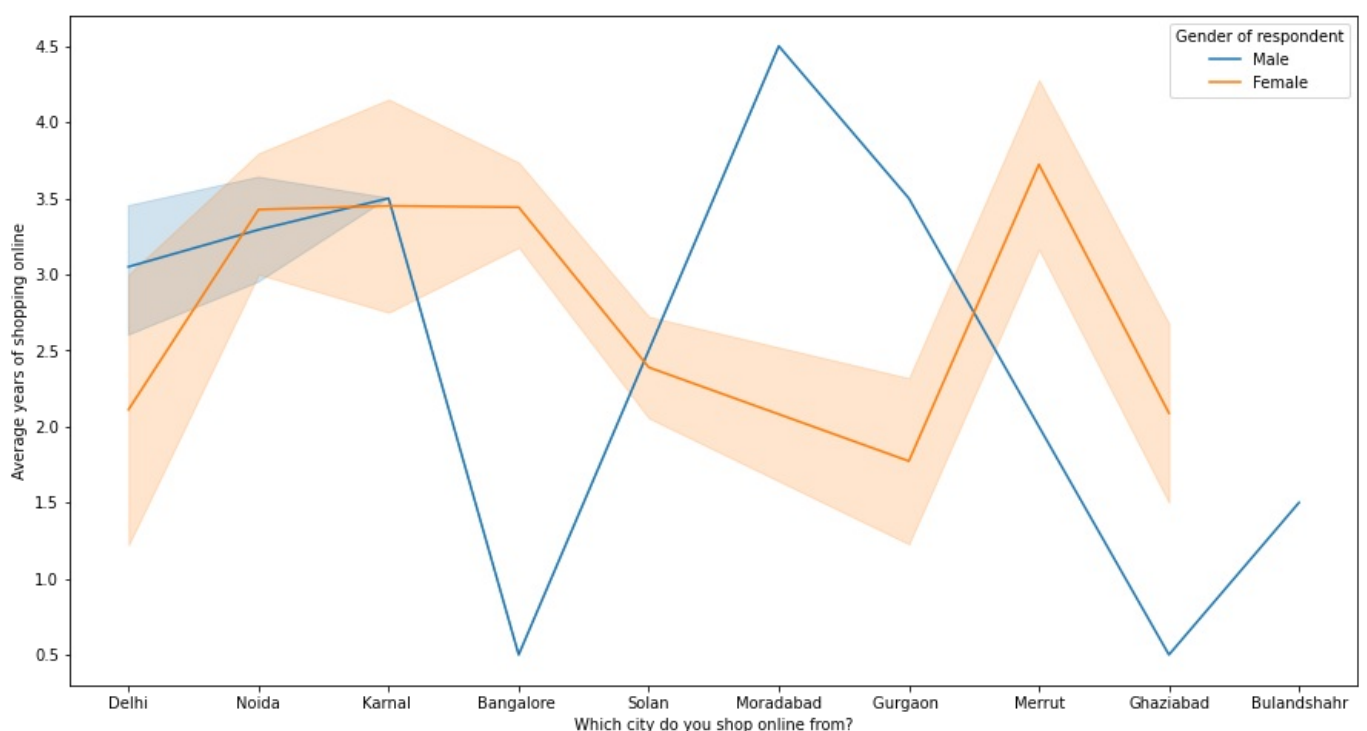
In [30]: `df['Which city do you shop online from?'].unique()`

Out[30]: `array(['Delhi', 'Greater Noida', 'Karnal ', 'Bangalore ', 'Noida',
'Solani', 'Moradabad', 'Gurgaon ', 'Merrut', 'Ghaziabad',
'Bulandshahr'], dtype=object)`

In [31]: `#Changing Greater noida to noida
df['Which city do you shop online from?'].replace({'Greater Noida':'Noida'},inplace=True)`

In [32]: `plt.figure(figsize=(15,8))
sns.lineplot(df['Which city do you shop online from?'],df['Average years of shopping online'],hue=df['Gender of respondent'])`

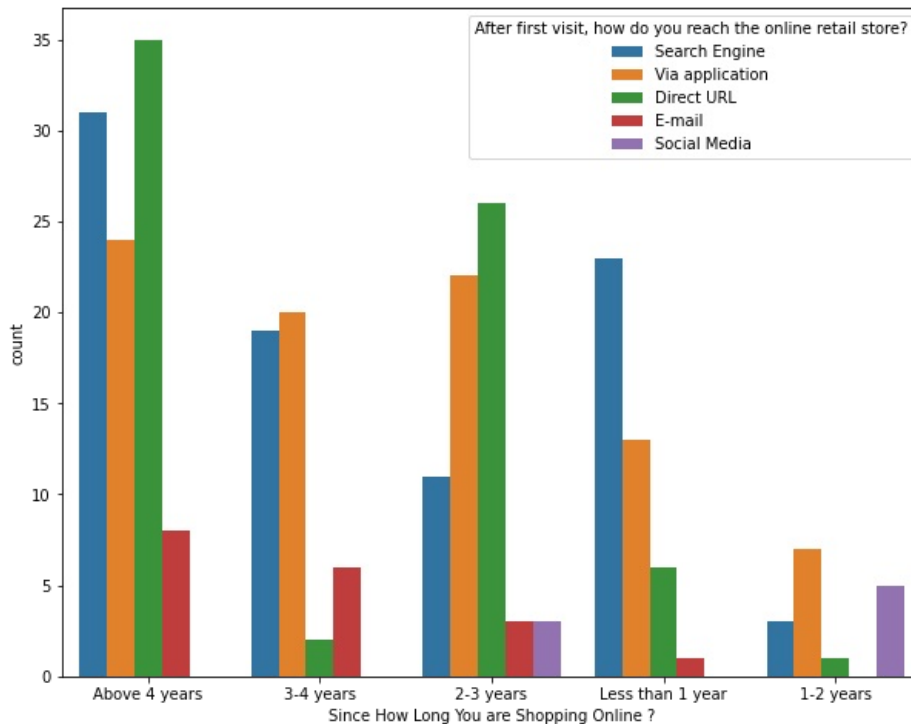
Out[32]: `<AxesSubplot:xlabel='Which city do you shop online from?', ylabel='Average years of shopping online'>`



In [33]: *##n lines, we can see that density of female customers is more than male. Men living in banglore and ghaziabad st*

In [34]: `plt.figure(figsize=(10,8))
sns.countplot(df['Since How Long You are Shopping Online ?'],
hue=df['After first visit, how do you reach the online retail store?'])`

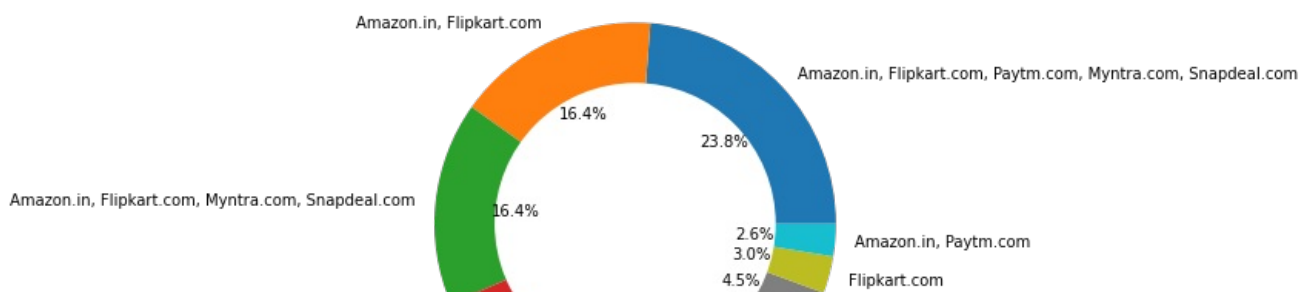
Out[34]: `<AxesSubplot:xlabel='Since How Long You are Shopping Online ?', ylabel='count'>`

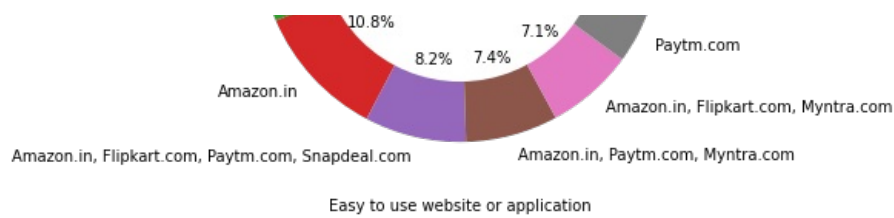


In [35]: *##Even though people who are shopping online for more than 3 years donot use the application rather use search en*
##Brand image

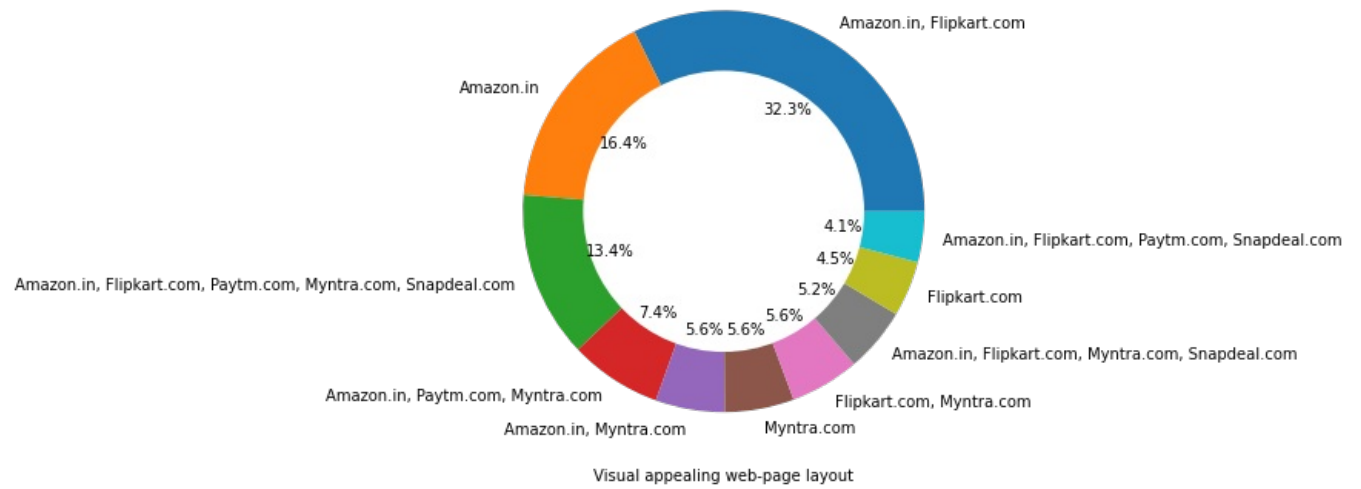
In [36]: `performance=['Easy to use website or application',
'Visual appealing web-page layout', 'Wild variety of product on offer',
'Complete, relevant description information of products',
'Fast loading website speed of website and application',
'Reliability of the website or application',
'Quickness 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']`

In [37]: `for i in performance:
plt.figure(figsize=(8,6))
df[i].value_counts().plot.pie(autopct='%1.1f%%')
centre=plt.Circle((0,0),0.7,fc='white')
fig=plt.gcf()
fig.gca().add_artist(centre)
plt.xlabel(i)
plt.ylabel('')
plt.figure()`

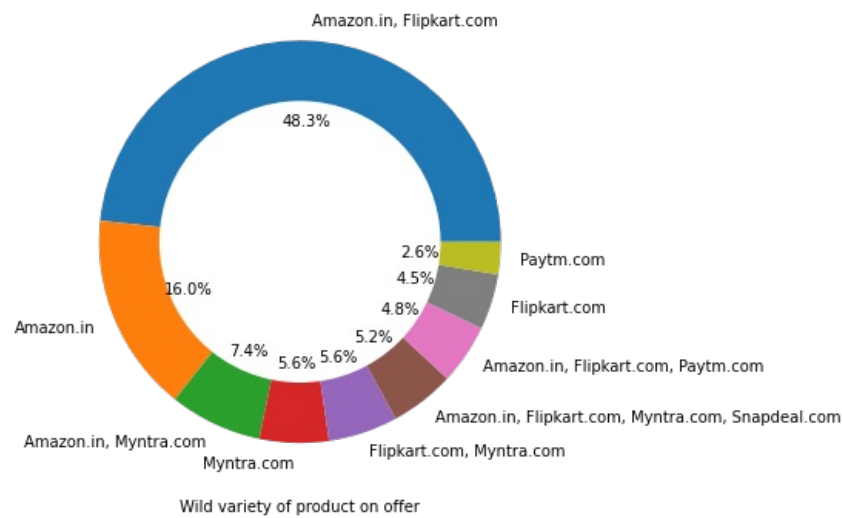




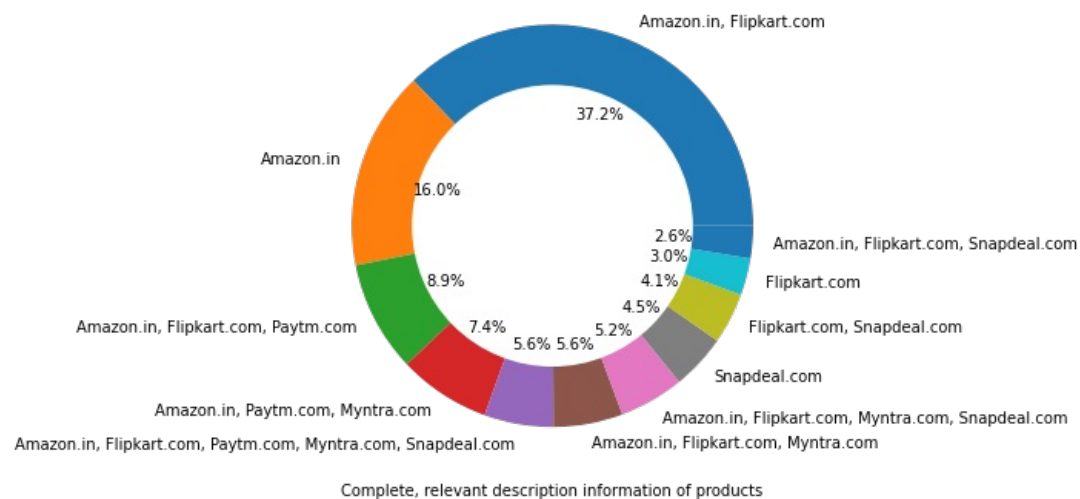
<Figure size 432x288 with 0 Axes>



<Figure size 432x288 with 0 Axes>



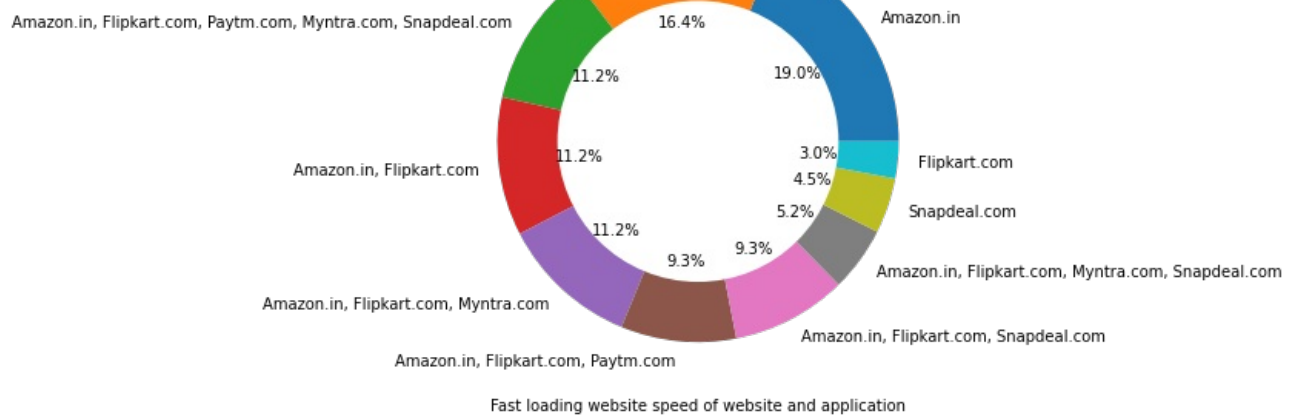
<Figure size 432x288 with 0 Axes>



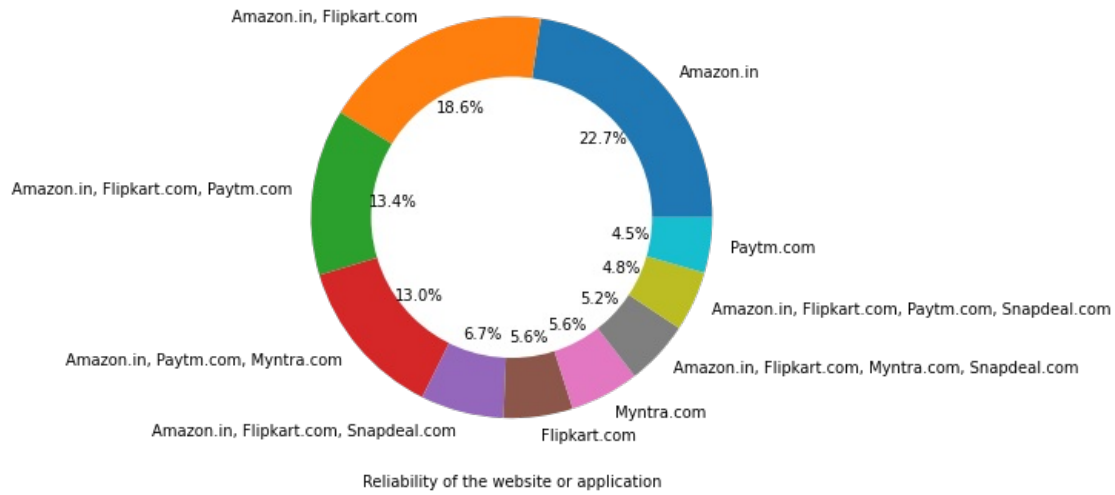
<Figure size 432x288 with 0 Axes>

Amazon.in, Paytm.com

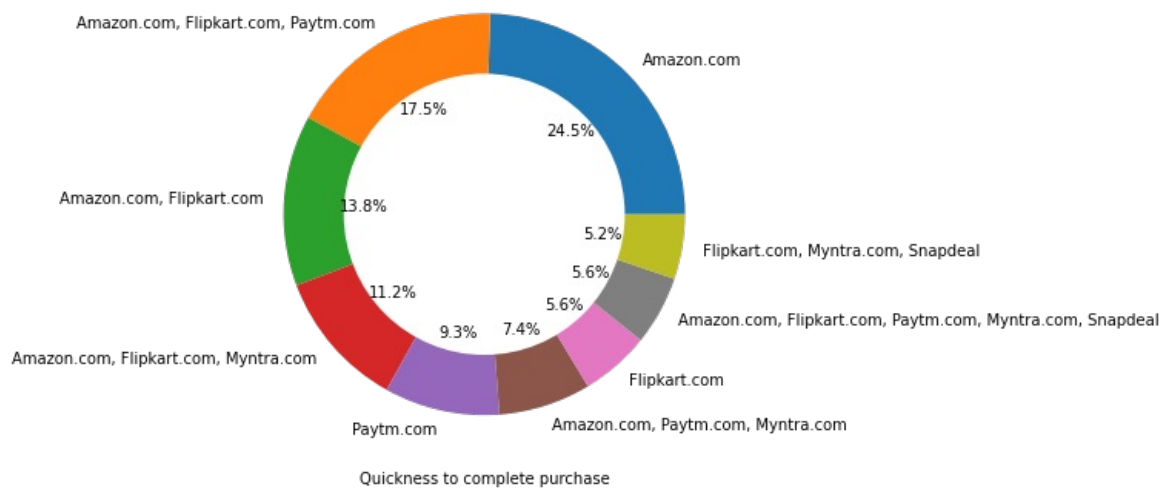




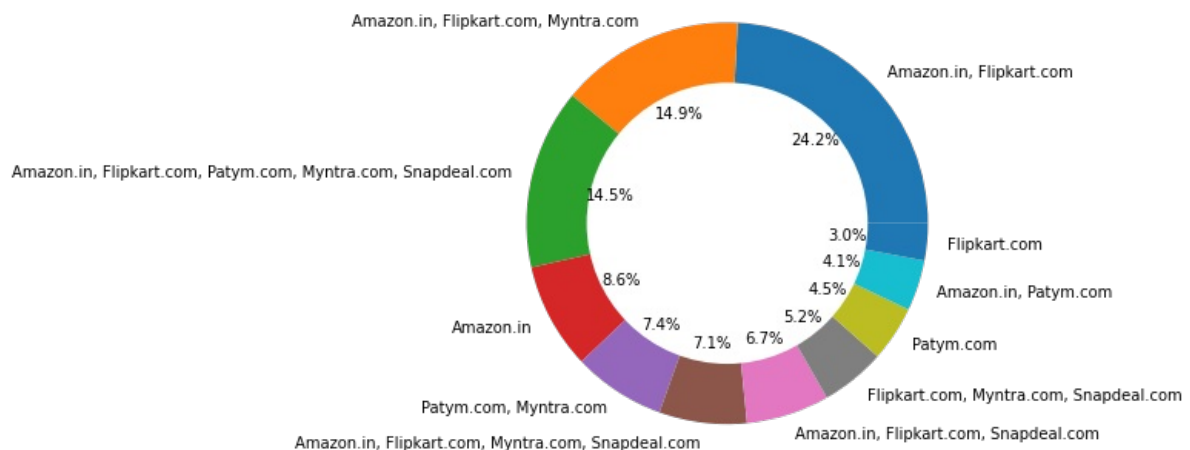
<Figure size 432x288 with 0 Axes>



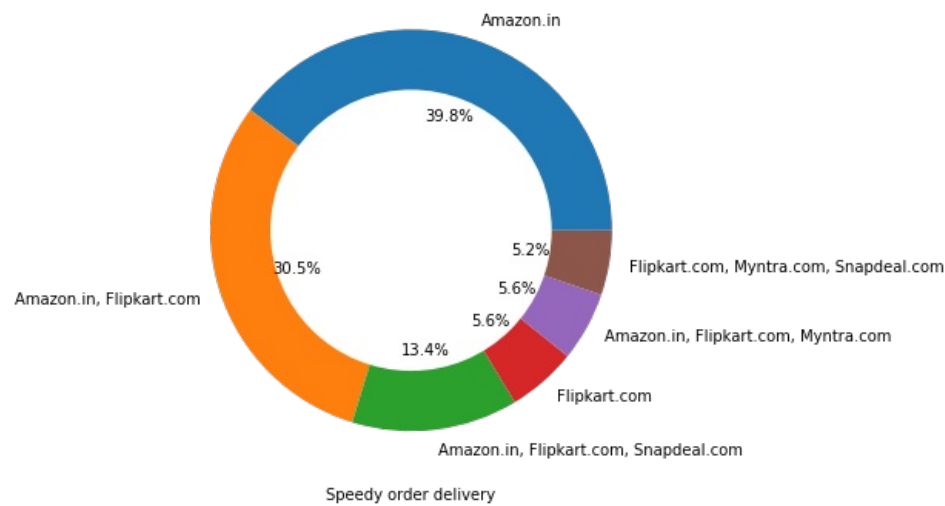
<Figure size 432x288 with 0 Axes>



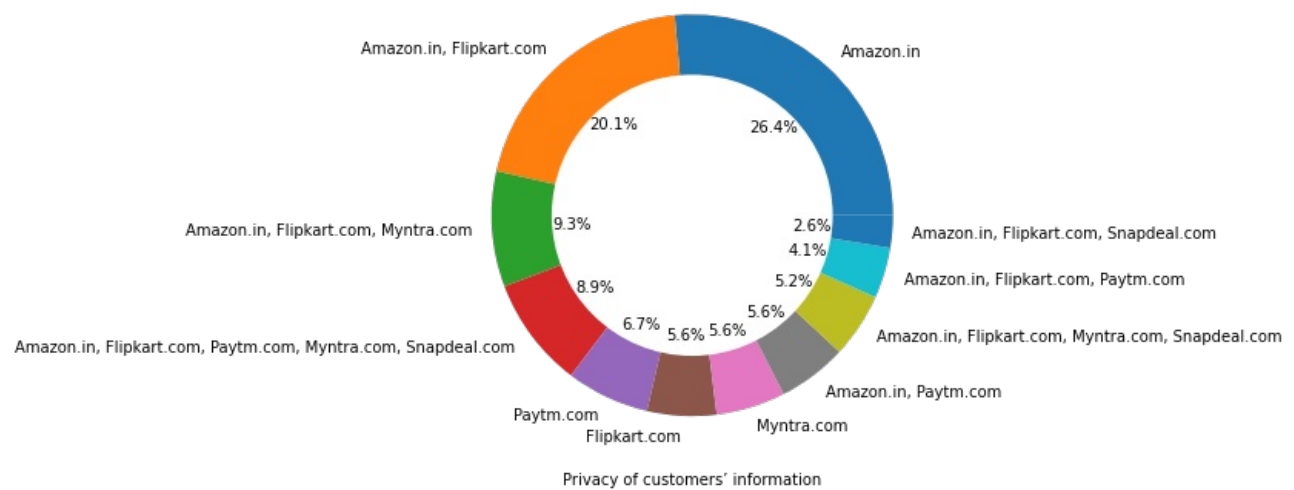
<Figure size 432x288 with 0 Axes>



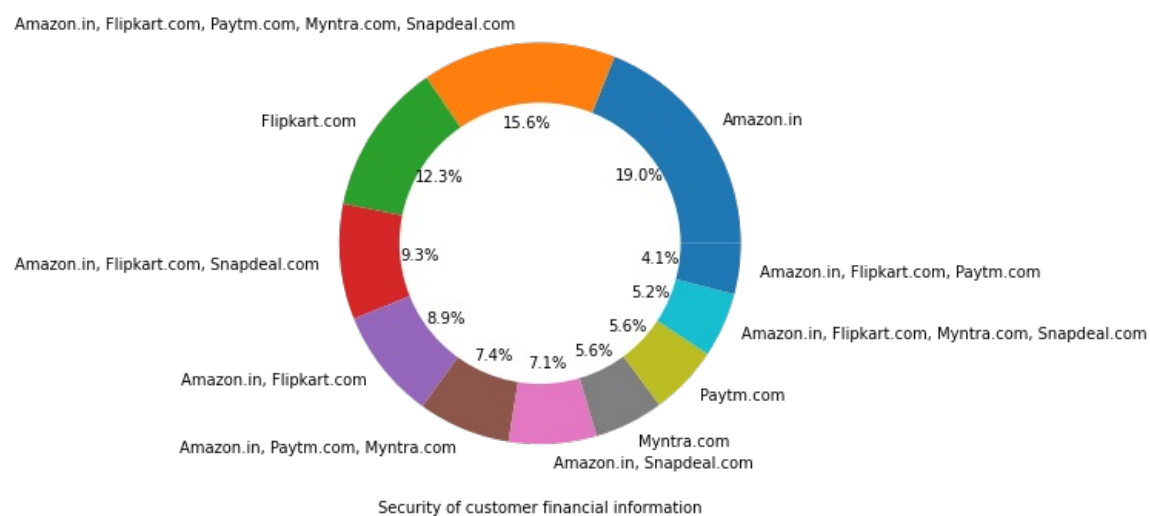
<Figure size 432x288 with 0 Axes>



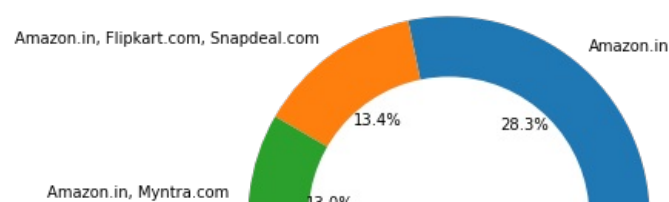
<Figure size 432x288 with 0 Axes>

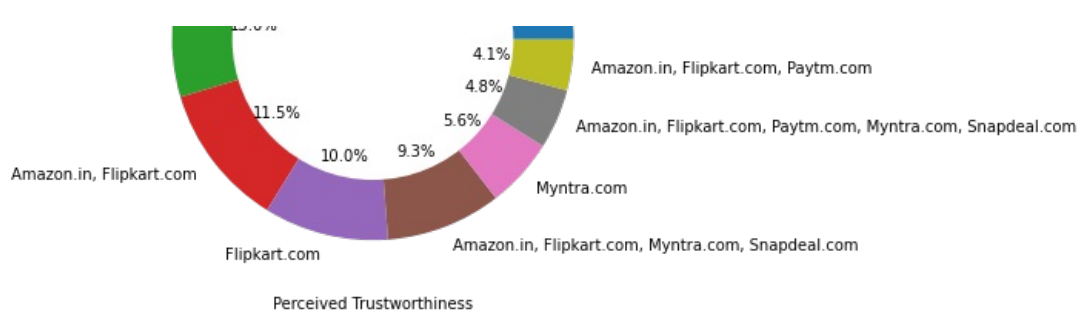


<Figure size 432x288 with 0 Axes>

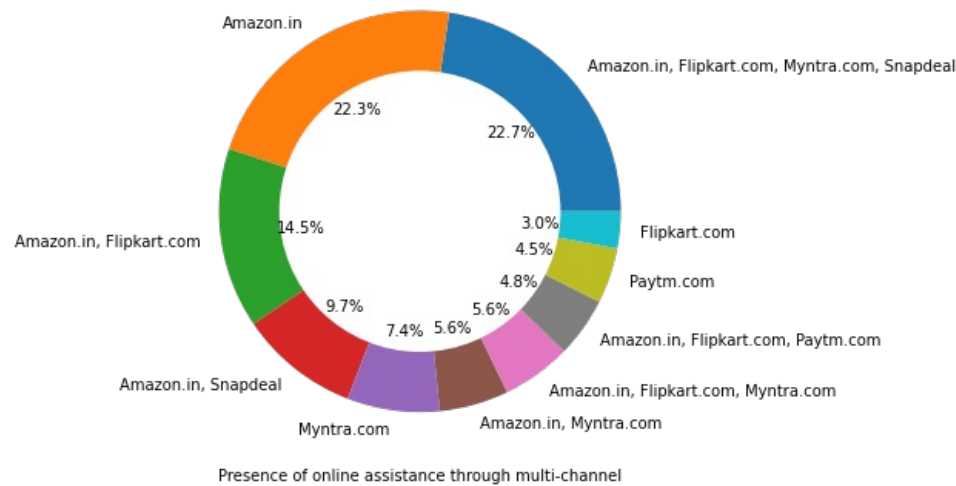


<Figure size 432x288 with 0 Axes>





<Figure size 432x288 with 0 Axes>

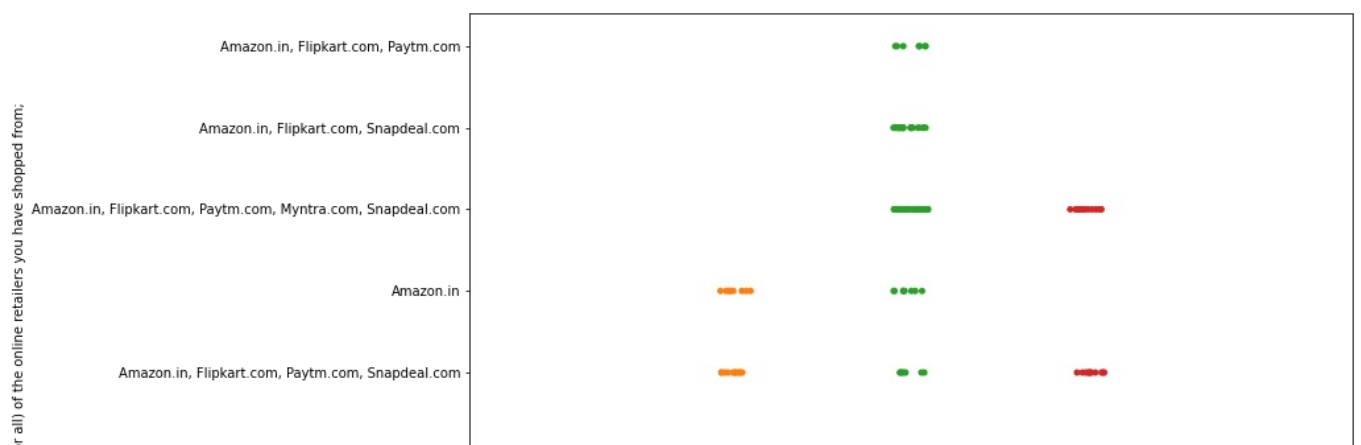


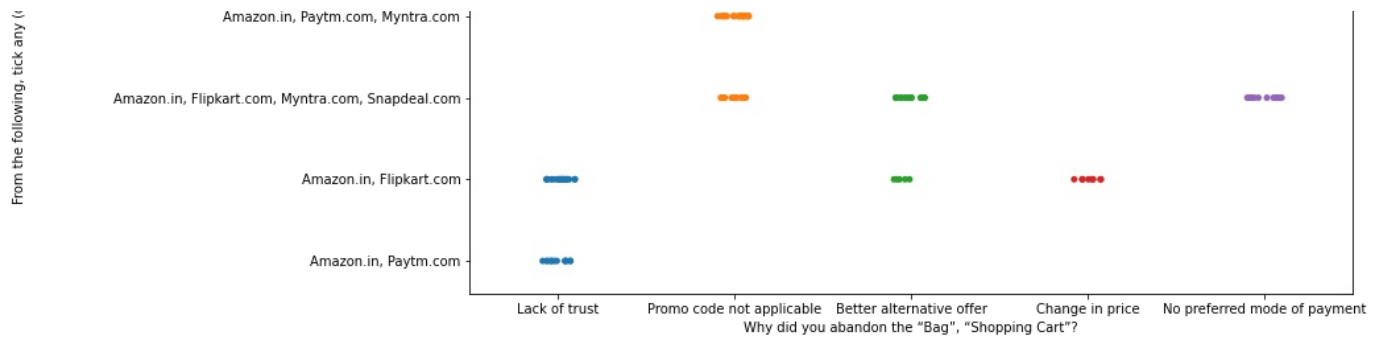
<Figure size 432x288 with 0 Axes>

In [38]: *##Amazon, Flipkart have been had the highest votes for having all the positive points and have maintained a very*

```
In [39]: plt.figure(figsize=(12,10))
sns.stripplot(df['Why did you abandon the "Bag", "Shopping Cart"?'],
              df['From the following, tick any (or all) of the online retailers you have shopped from;'])
```

Out[39]: <AxesSubplot:xlabel='Why did you abandon the "Bag", "Shopping Cart"?', ylabel='From the following, tick any (or a ll) of the online retailers you have shopped from;*>



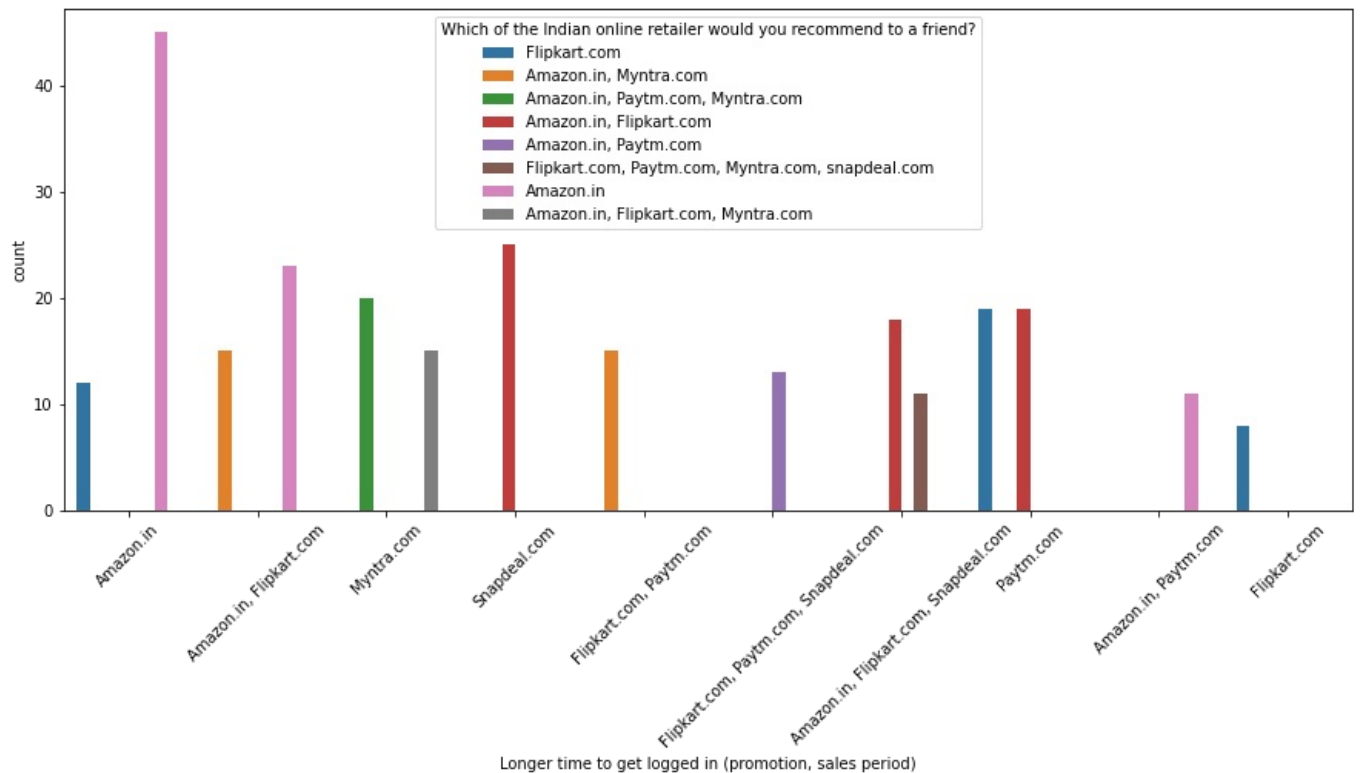


In [40]: *#We can clearly see that most of the time people abandon the bag is beacuse they get a better alternative offer c*

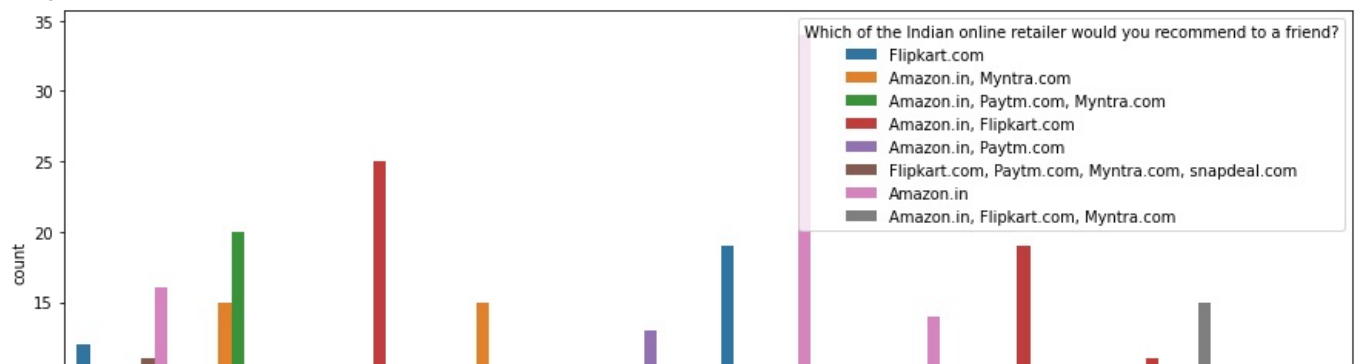
In [41]: *##LOYALTY*

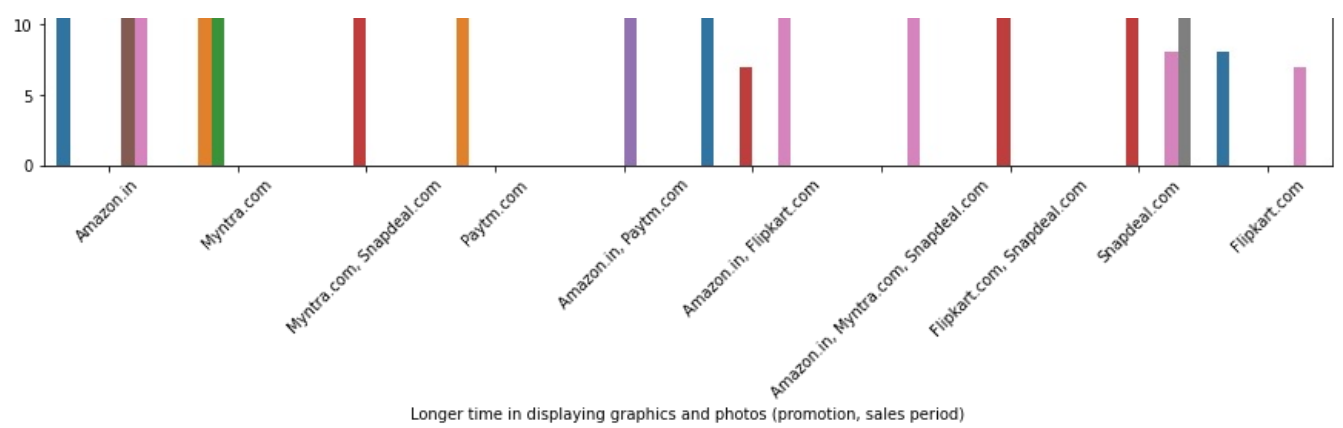
In [42]: *#Collecting all the negative remarks about a brand*
 bad=['Longer time to get logged in (promotion, sales period)',
 'Longer time in displaying graphics and photos (promotion, sales period)',
 'Late declaration of price (promotion, sales period)',
 'Longer page loading time (promotion, sales 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']

In [43]: *for i in bad:*
 plt.figure(figsize=(15,6))
 sns.countplot(df[i],hue=df['Which of the Indian online retailer would you recommend to a friend?'])
 plt.xticks(rotation=45)
 plt.figure()

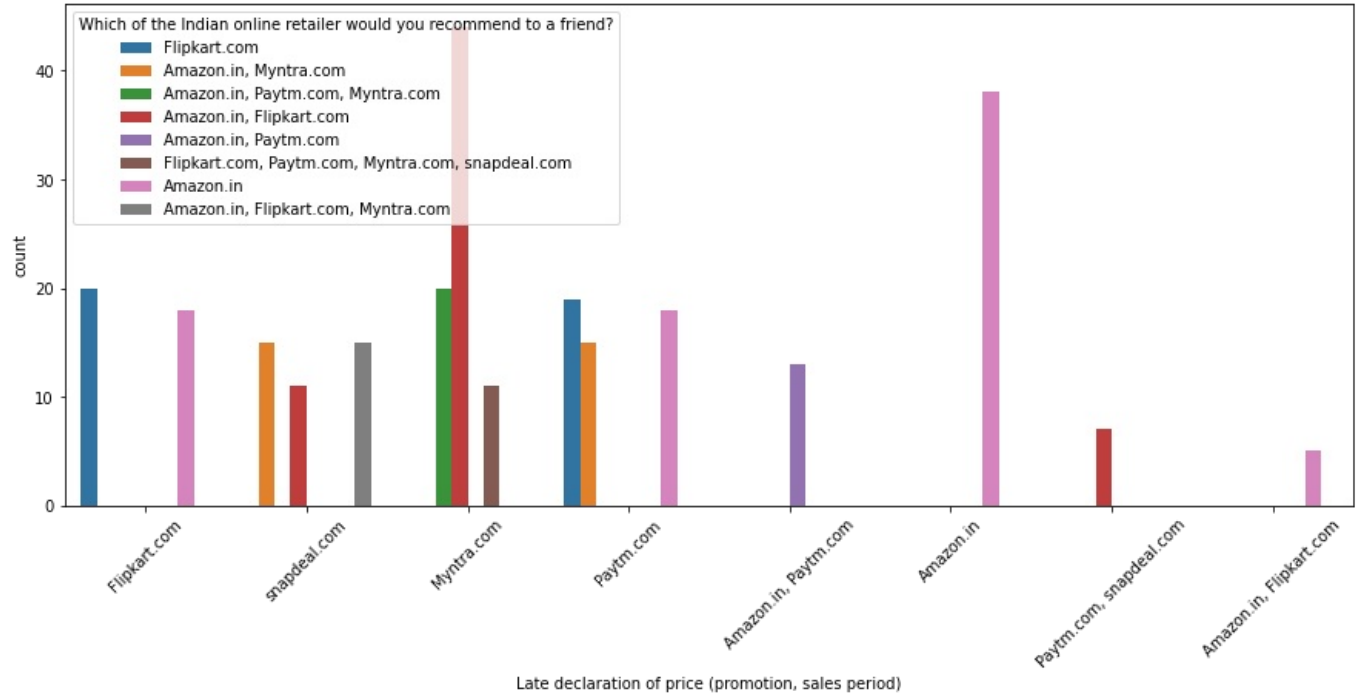


<Figure size 432x288 with 0 Axes>

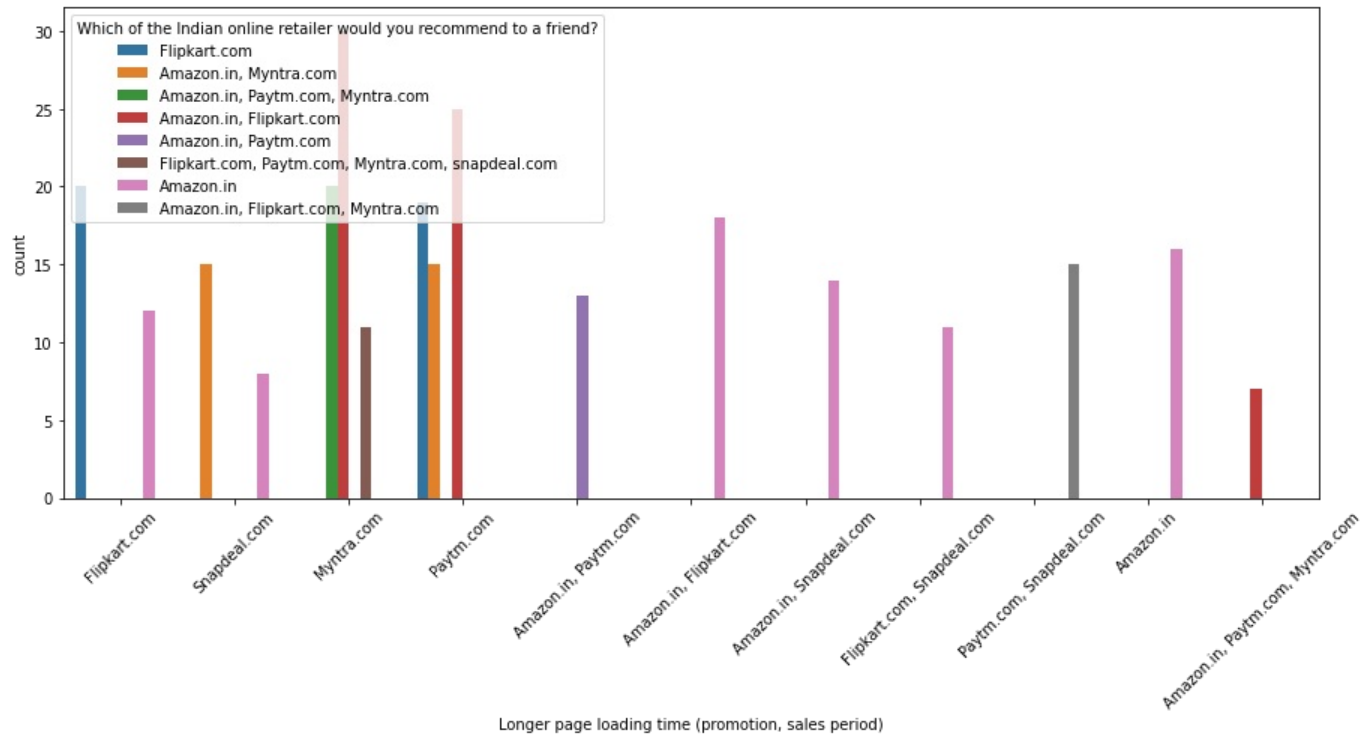




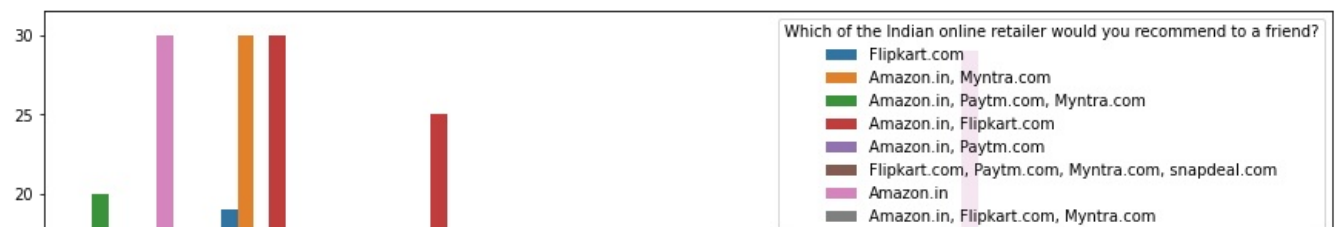
<Figure size 432x288 with 0 Axes>

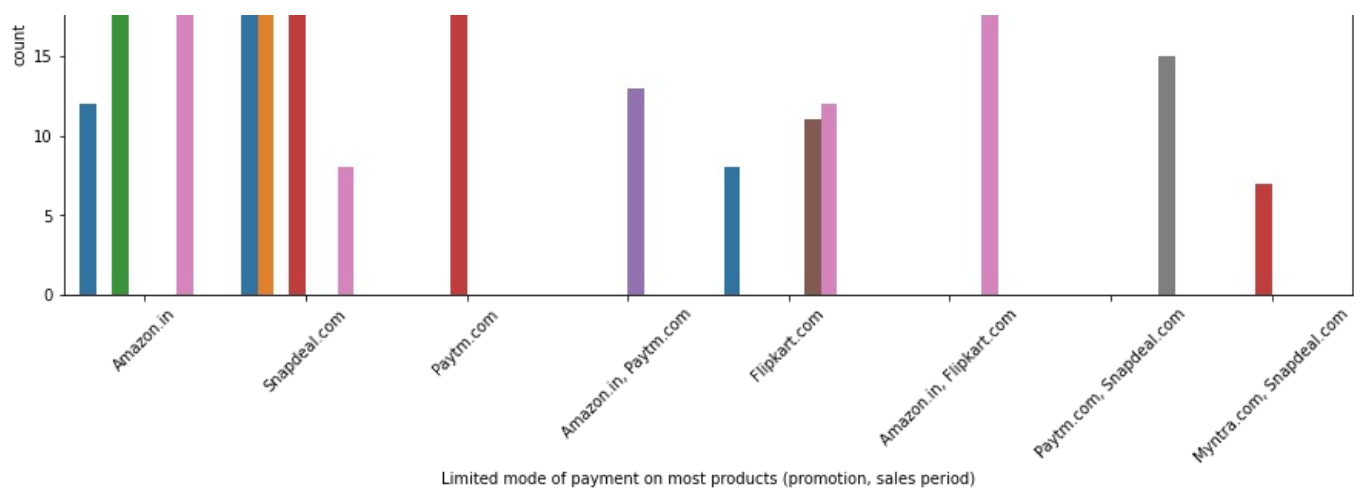


<Figure size 432x288 with 0 Axes>

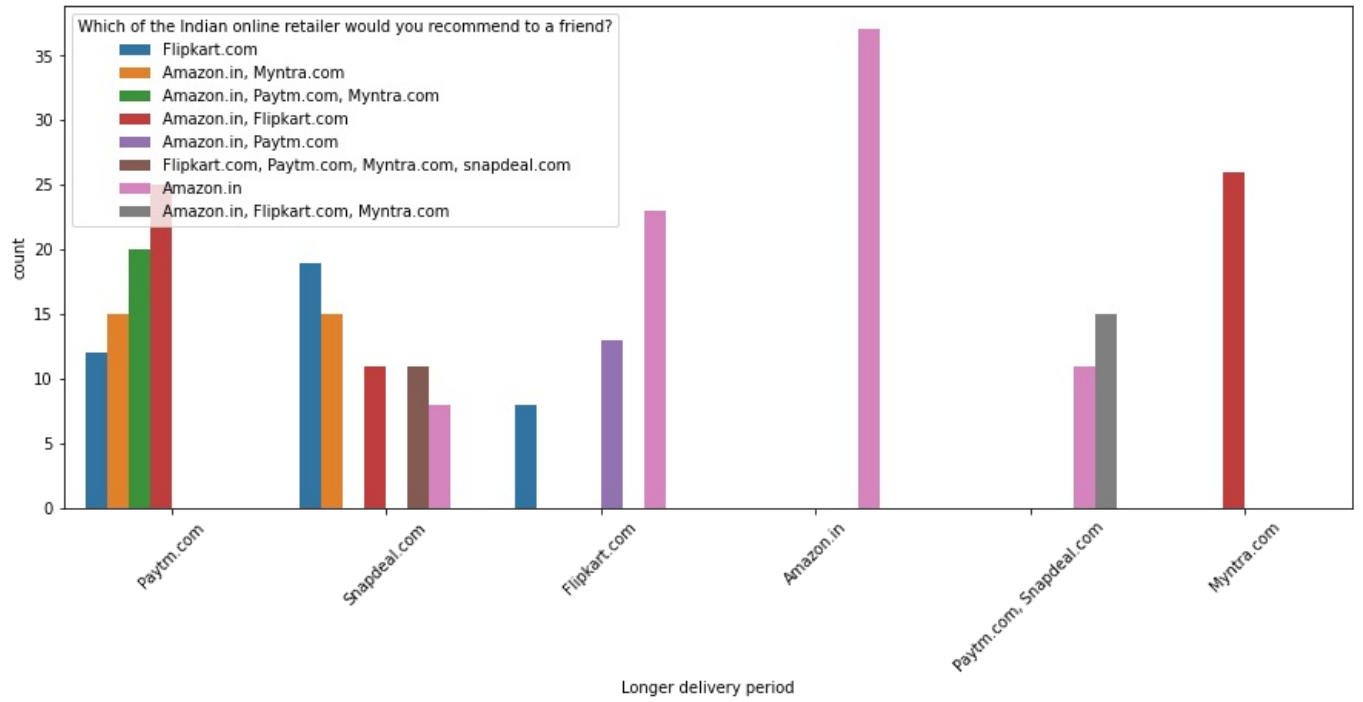


<Figure size 432x288 with 0 Axes>

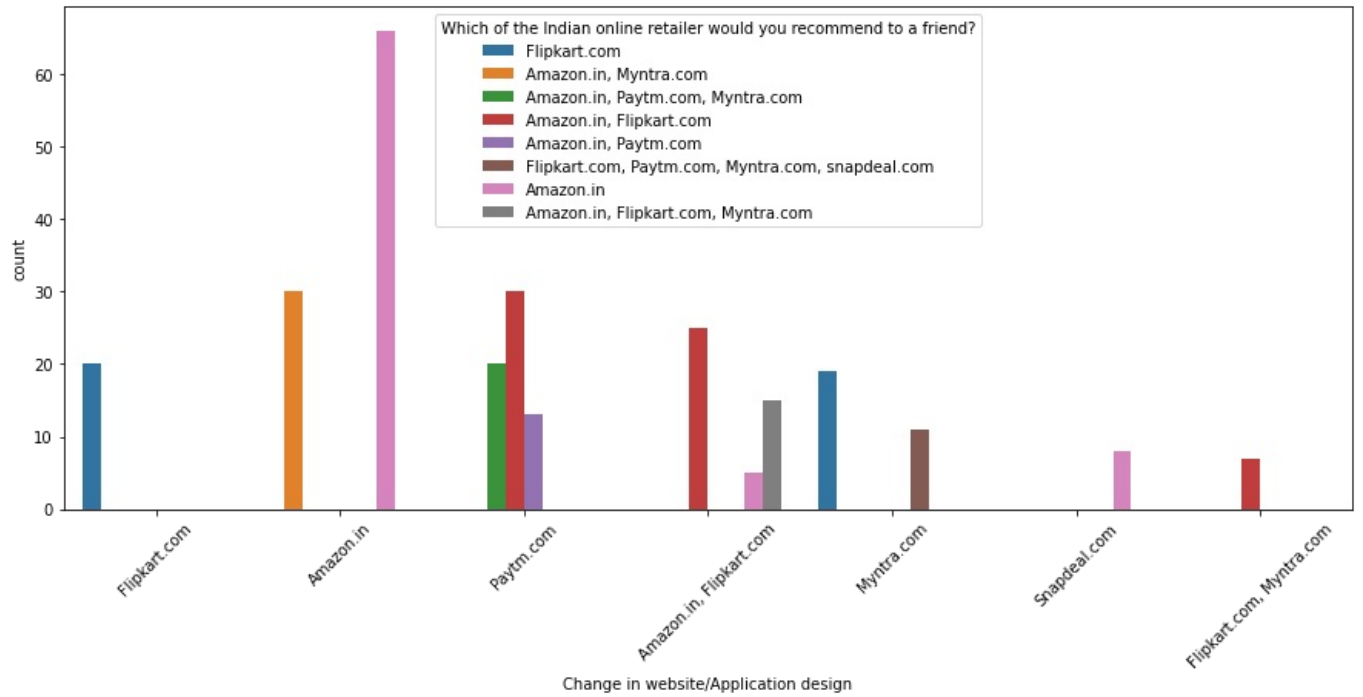




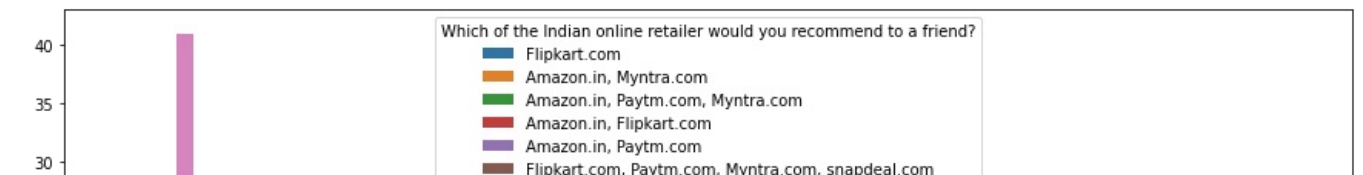
<Figure size 432x288 with 0 Axes>

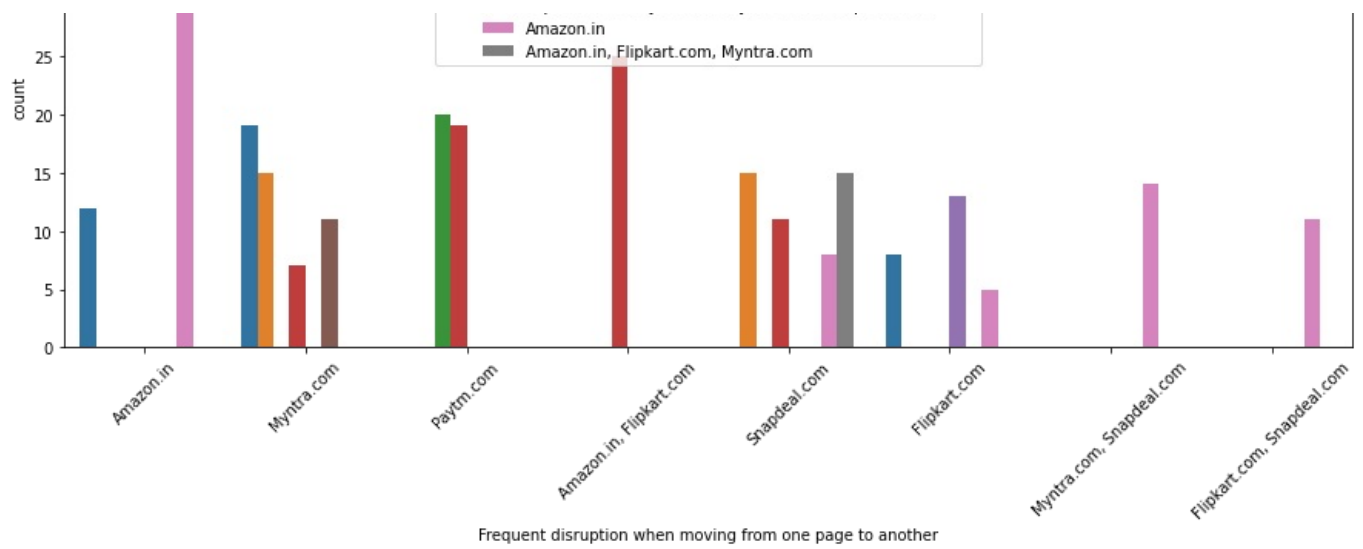


<Figure size 432x288 with 0 Axes>



<Figure size 432x288 with 0 Axes>





<Figure size 432x288 with 0 Axes>

```
In [44]: #Customers seem to be more loyal to amazon, flipkart and paytm as even though many of them have given negative re
```

```
In [45]: x=df.copy()
x.drop('Which of the Indian online retailer would you recommend to a friend?',axis=1,inplace=True)
y=df['Which of the Indian online retailer would you recommend to a friend?']
```

```
In [46]: cat=[i for i in x.columns if x[i].dtypes=='O']
```

```
In [47]: from sklearn.preprocessing import OrdinalEncoder,LabelEncoder
encode=OrdinalEncoder()
labe=LabelEncoder()
```

```
In [48]: #using ordinal encoder for independent features
for i in cat:
    x[i]=encode.fit_transform(x[i].values.reshape(-1,1))

#Using label encoder for Label Column
y=labe.fit_transform(y)
```

```
In [49]: ## scaling
```

```
In [51]: from sklearn.preprocessing import MinMaxScaler
scaler=MinMaxScaler()
```

```
In [52]: xd=scaler.fit_transform(x)
x=pd.DataFrame(xd,columns=x.columns)
```

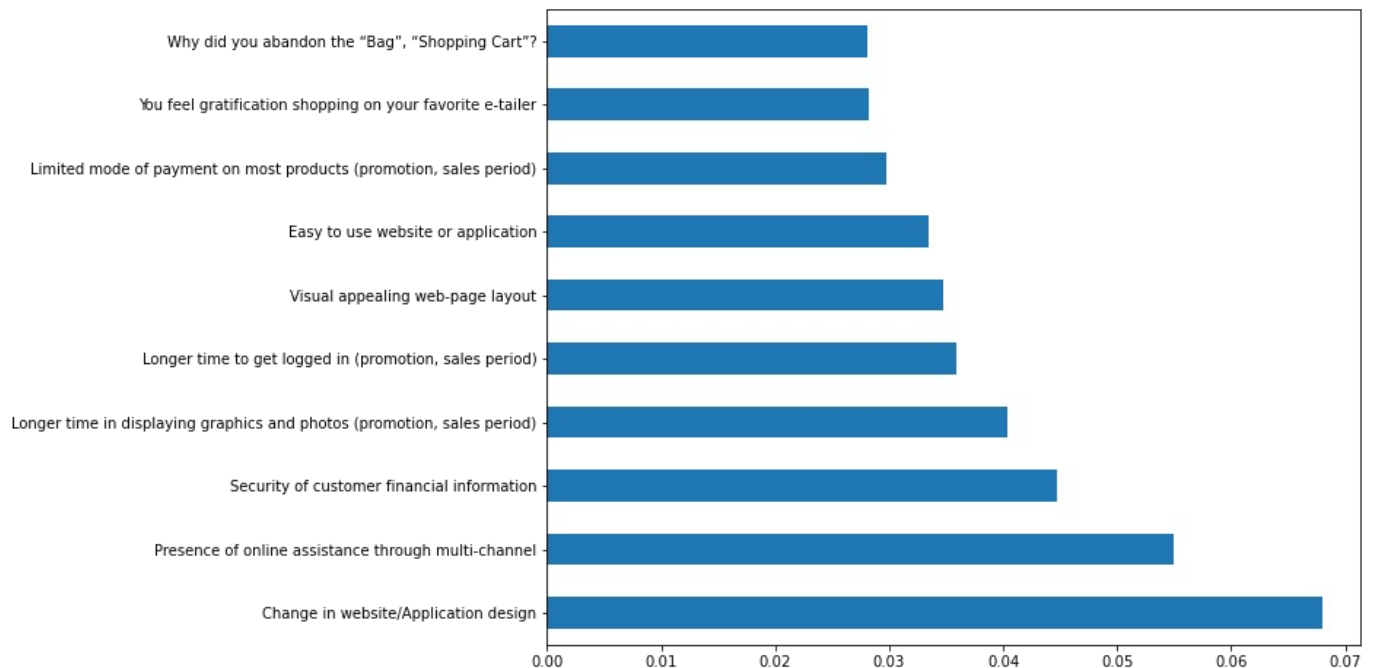
```
In [53]: ## feature selection
```

```
In [54]: from sklearn.ensemble import RandomForestClassifier
m=RandomForestClassifier()
```

```
m.fit(x,y)
```

```
Out[54]: RandomForestClassifier()
```

```
In [55]: #plot graph of feature importances for better visualization  
feat_importances = pd.Series(m.feature_importances_, index=x.columns)  
plt.figure(figsize=(10,8))  
feat_importances.nlargest(10).plot(kind='barh')  
plt.show()
```



```
In [56]: ##In the above chart we can see that above features are of most importance in determining which platform will a
```

```
In [57]: from sklearn.feature_selection import SelectKBest  
from sklearn.feature_selection import chi2
```

```
In [58]: selection = SelectKBest(score_func=chi2)  
fit = selection.fit(x,y)
```

```
In [59]: dfscores = pd.DataFrame(fit.scores_)  
dfcolumns = pd.DataFrame(x.columns)  
featureScores = pd.concat([dfcolumns,dfscores],axis=1)  
featureScores.columns = ['Features','Score'] #naming the dataframe columns
```

```
In [60]: print(featureScores.nlargest(10,'Score')) #print10 best features  
feat=list(featureScores.nlargest(10,'Score')['Features'])
```

	Features	Score
16	Why did you abandon the "Bag", "Shopping Cart"?	75.754028
22	Loading and processing speed	59.810983
42	Shopping on the website gives you the sense of...	59.253569
10	What browser do you run on your device to acce...	57.171099
67	Change in website/Application design	55.301526
49	Visual appealing web-page layout	54.245760
65	Limited mode of payment on most products (prom...	53.269266
61	Longer time to get logged in (promotion, sales...	48.222655
62	Longer time in displaying graphics and photos ...	48.130643
50	Wild variety of product on offer	47.605973

```
In [61]: ##PCA
```

```
In [62]: from sklearn.decomposition import PCA  
pca = PCA().fit(x)
```

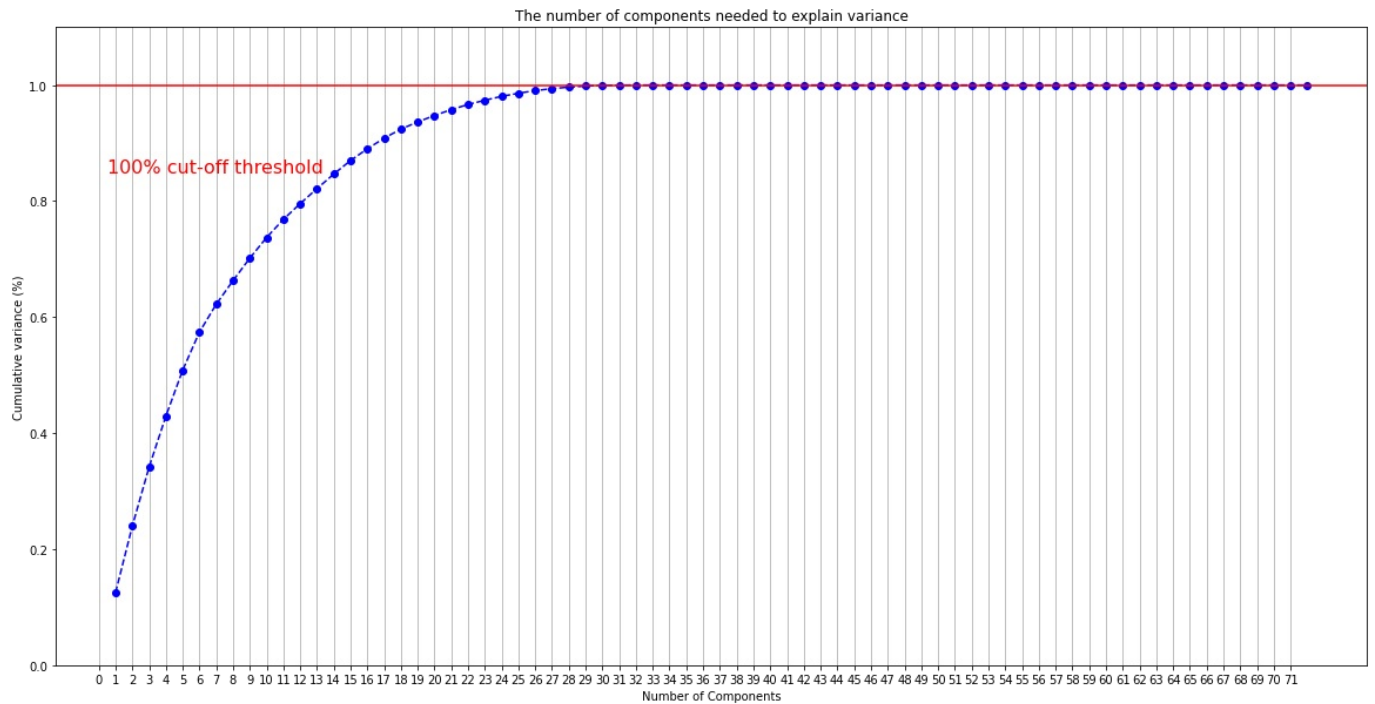
```
In [63]: fig, ax = plt.subplots(figsize=(20,10))
xi = np.arange(1, 73, step=1)
yi = np.cumsum(pca.explained_variance_ratio_)

plt.ylim(0.0,1.1)
plt.plot(xi, yi, marker='o', linestyle='--', color='b')

plt.xlabel('Number of Components')
plt.xticks(np.arange(0, 72, step=1)) #change from 0-based array index to 1-based human-readable label
plt.ylabel('Cumulative variance (%)')
plt.title('The number of components needed to explain variance')

plt.axhline(y=1, color='r', linestyle='-')
plt.text(0.5, 0.85, '100% cut-off threshold', color = 'red', fontsize=16)

ax.grid(axis='x')
plt.show()
```



```
In [64]: pca=PCA(n_components=29)
x=pca.fit_transform(x)
x=pd.DataFrame(x)
x.head()
```

```
Out[64]:
```

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	2.065419	-0.577759	-1.030081	-1.109784	0.652387	-1.137025	0.699876	-0.023177	-0.960103	-0.238855	-0.436650	-0.539191	0.130180
1	0.048667	-1.490547	1.081348	0.641617	0.066388	-0.820495	0.072214	-0.644870	0.087754	-0.296247	-0.157354	0.881935	0.648067
2	1.671684	-0.120022	0.775570	-1.481374	0.128287	0.836151	-0.793600	0.102789	0.448813	-0.515949	-0.033307	-0.086125	0.368685
3	-0.009522	2.146296	0.753236	-0.363176	-1.348954	-0.176575	0.567430	-0.548924	-0.142604	-0.084665	-0.341339	0.095133	0.089171
4	0.051352	-0.187387	2.386865	0.914150	0.273219	-0.992250	-0.511792	0.701105	-0.225943	0.735107	0.138216	-0.814487	-0.204856

```
In [65]: ## modeling phase
```

```
In [67]: from sklearn.model_selection import train_test_split,cross_val_score
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score,confusion_matrix,classification_report,roc_auc_score,roc_curve
```

```
In [68]: xtrain,xtest,ytrain,ytest=train_test_split(x,y,test_size=0.3,random_state=7)
```

```
In [69]: ##RANDOMFORESTCLASSIFIER
```

```
In [70]: model=RandomForestClassifier()
```

```
model.fit(xtrain,ytrain)
p=model.predict(xtest)
s=cross_val_score(model,x,y,cv=10)
```

```
In [71]: print('Accuracy',np.round(accuracy_score(p,ytest),4))
print('-----')
print('Mean of Cross Validation Score',np.round(s.mean(),4))
print('-----')
print('Confusion Matrix')
print(confusion_matrix(p,ytest))
print('-----')
print('Classification Report')
print(classification_report(p,ytest))
```

```
Accuracy 1.0
-----
Mean of Cross Validation Score 0.9926
-----
Confusion Matrix
[[26  0  0  0  0  0  0  0]
 [ 0 22  0  0  0  0  0  0]
 [ 0  0  4  0  0  0  0  0]
 [ 0  0  0  4  0  0  0  0]
 [ 0  0  0  0  5  0  0  0]
 [ 0  0  0  0  0  7  0  0]
 [ 0  0  0  0  0  0 11  0]
 [ 0  0  0  0  0  0  0  2]]
-----
Classification Report
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	26
1	1.00	1.00	1.00	22
2	1.00	1.00	1.00	4
3	1.00	1.00	1.00	4
4	1.00	1.00	1.00	5
5	1.00	1.00	1.00	7
6	1.00	1.00	1.00	11
7	1.00	1.00	1.00	2
accuracy			1.00	81
macro avg	1.00	1.00	1.00	81
weighted avg	1.00	1.00	1.00	81

```
In [72]: ## HYPER PARAMETER TUNING
```

```
In [73]: params={'n_estimators':[100, 300, 500, 700],
               'min_samples_split':[1,2,3,4],
               'min_samples_leaf':[1,2,3,4],
               'max_depth':[None,1,2,3,4,5,6,7,8,9,10,15,20,25,30,35,40]}
```

```
In [75]: from sklearn.model_selection import RandomizedSearchCV
```

```
In [76]: g=RandomizedSearchCV(RandomForestClassifier(),params,cv=10)
```

```
In [77]: g.fit(xtrain,ytrain)
```

```
Out[77]: RandomizedSearchCV(cv=10, estimator=RandomForestClassifier(),
                           param_distributions={'max_depth': [None, 1, 2, 3, 4, 5, 6, 7,
                                                                8, 9, 10, 15, 20, 25, 30,
                                                                35, 40],
                           'min_samples_leaf': [1, 2, 3, 4],
                           'min_samples_split': [1, 2, 3, 4],
                           'n_estimators': [100, 300, 500, 700]})
```

```
In [78]: print(g.best_estimator_)
print(g.best_params_)
print(g.best_score_)
```

```
RandomForestClassifier(max_depth=30, min_samples_split=3, n_estimators=700)
{'n_estimators': 700, 'min_samples_split': 3, 'min_samples_leaf': 1, 'max_depth': 30}
```

0.9947368421052631

```
In [79]: m=RandomForestClassifier(max_depth=20, min_samples_leaf=4, min_samples_split=4,n_estimators=700)
m.fit(xtrain,ytrain)
p=m.predict(xtest)
score=cross_val_score(m,x,y,cv=10)
```

```
In [80]: print('Accuracy',np.round(accuracy_score(p,ytest),4))
print('-----')
print('Mean of Cross Validation Score',np.round(s.mean(),4))
print('-----')
print('Confusion Matrix')
print(confusion_matrix(p,ytest))
print('-----')
print('Classification Report')
print(classification_report(p,ytest))
```

Accuracy 1.0

Mean of Cross Validation Score 0.9926

Confusion Matrix

```
[[26  0  0  0  0  0  0  0]
 [ 0 22  0  0  0  0  0  0]
 [ 0  0  4  0  0  0  0  0]
 [ 0  0  0  4  0  0  0  0]
 [ 0  0  0  0  5  0  0  0]
 [ 0  0  0  0  0  7  0  0]
 [ 0  0  0  0  0  0 11  0]
 [ 0  0  0  0  0  0  0 2]]
```

Classification Report

	precision	recall	f1-score	support
0	1.00	1.00	1.00	26
1	1.00	1.00	1.00	22
2	1.00	1.00	1.00	4
3	1.00	1.00	1.00	4
4	1.00	1.00	1.00	5
5	1.00	1.00	1.00	7
6	1.00	1.00	1.00	11
7	1.00	1.00	1.00	2
accuracy			1.00	81
macro avg	1.00	1.00	1.00	81
weighted avg	1.00	1.00	1.00	81

```
In [87]: model=RandomForestClassifier( min_samples_split=3, n_estimators=700, min_samples_leaf=1, max_depth= 30)
model.fit(xtrain,ytrain)
p=model.predict(xtest)
score=cross_val_score(model,x,y,cv=10)
```

In [88]:

In []:

Loading [MathJax]/extensions/Safe.js