

# Import Libraries

```
In [67]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

import sklearn
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, mean_absolute_error
from sklearn.model_selection import train_test_split

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

# Loading The Dataset

```
In [2]: df= pd.read_excel(r'C:\python\Cus_retention.xlsx')
```

```
In [3]: data= df.iloc[:,0:47]
```

```
In [4]: data
```

Out[4]:

	1Gender of respondent	2 How old are you?	3 Which city do you shop online from?	4 What is the Pin Code of where you shop online from?	5 Since How Long You are Shopping Online ?	6 How many times you have made an online purchase in the past 1 year?	7 How do you access the internet while shopping on-line?	8 Which device do you use to access the online shopping?	9 What is the screen size of your mobile device? ltltltltlt	10 What is the operating system (OS) of your device? ltltltlt	...	3 satisf
0	0	3	0	110009	5	4	4	3	5	1	...	
1	1	2	0	110030	5	5	2	1	2	3	...	
2	1	2	4	201308	4	5	3	1	4	2	...	
3	0	2	6	132001	4	1	3	1	4	3	...	
4	1	2	1	530068	3	2	2	1	2	3	...	
...	...	...	...	...	...	...	...	...	...	...	...	
264	1	2	10	173212	2	1	3	1	4	2	...	
265	1	3	3	201008	2	4	3	1	5	2	...	
266	1	4	1	560010	3	1	3	2	5	1	...	
267	1	1	10	173229	3	1	2	1	4	2	...	
268	1	4	3	201009	3	4	3	1	4	2	...	

269 rows × 47 columns

```
In [6]: data.columns
```

```
Out[6]: Index(['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?', '11 What browser do you run on your device to access the website?', '12 Which channel did you follow to arrive at your favorite online store for the first time?', '13 After first visit, how do you reach the online retail store?', '14 How much time do you explore the e-retail store before making a purchase decision?', '15 What is your preferred payment Option?', '16 How 4 do you abandon (selecting an items and leaving without making payment) your shopping cart?', '17 Why did you abandon the "Bag", "Shopping Cart"?', '18 The content on the website must be easy to read and understand', '19 Information on similar product to the one highlighted is important for product comparison', '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 customer', '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'],
dtype='object')
```



1Gender of respondent	2 How old are you?	3 Which city do you shop online from?	4 What is the Pin Code of where you shop online from?	5 Since How Long You are Shopping Online ?	6 How many times you have made an online purchase in the past 1 year?	14 How much time do you explore the e-retail store before making a purchase decision?	15 What is your preferred payment Option?	16 How 4 do you abandon (selecting an items and leaving without making payment) your shopping cart?	17 Why did you abandon the “Bag”, “Shopping Cart”?	30 sh
3	0	2	6	132001	4	1	3	1	1	2 ...
4	1	2	1	530068	3	2	5	1	4	2 ...
...	...	...	...	...	...	...	...	...	...	...
264	1	2	10	173212	2	1	5	2	1	2 ...
265	1	3	3	201008	2	4	1	2	3	1 ...
266	1	4	1	560010	3	1	5	1	3	2 ...
267	1	1	10	173229	3	1	2	2	3	4 ...
268	1	4	3	201009	3	4	4	2	1	5 ...

```
In [13]: dataset.columns
```

# Removing Duplicates

```
In [14]: dataset.duplicated().sum()
dataset.drop_duplicates(inplace=True)
```

```
In [15]: dataset
```

```
Out[15]:
```

	1Gender of respondent	2 How old are you?	3 Which city do you shop online from?	4 What is the Pin Code of where you shop online from?	5 Since How Long You are Shopping Online ?	6 How many times you have made an online purchase in the past 1 year?	14 How much time do you explore the e-retail store before making a purchase decision?	15 What is your preferred payment Option? \t\t\t\t\t	16 How 4 do you abandon (selecting an items and leaving without making payment) your shopping cart? \t\t\t\t\t\t\t	17 Why did you abandon the “Bag”, “Shopping Cart”? \t\t\t\t\t	30 sh
0	0	3	0	110009	5	4	3	4	3	3	...
1	1	2	0	110030	5	5	5	1	5	5	...
2	1	2	4	201308	4	5	4	4	3	5	...
3	0	2	6	132001	4	1	3	1	1	2	...
4	1	2	1	530068	3	2	5	1	4	2	...
...	...	...	...	...	...	...	...	...	...	...	...
106	1	3	3	201008	2	4	1	2	3	1	...
107	1	1	3	201001	1	1	1	2	1	2	...
112	0	3	2	203001	2	1	3	2	3	2	...
116	1	2	1	560037	3	2	5	1	4	2	...
117	0	3	9	201308	5	5	5	4	4	2	...

104 rows × 28 columns

## Renaming of the columns

```
In [16]: dataset= dataset.rename(columns={'1Gender of respondent': 'Gender',
                                         '2 How old are you? ': 'Age',
                                         '3 Which city do you shop online from?': 'Shopping_city',
                                         '4 What is the Pin Code of where you shop online from?':
                                         '5 Since How Long You are Shopping Online ?': 'From_when'
```

```
In [17]: dataset= dataset.rename(columns={'6 How many times you have made an online purchase in th
                                         '14 How much time do you explore the e- retail store befo
                                         '15 What is your preferred payment Option?\t\t\t\t\t
                                         '16 How 4 do you abandon (selecting an items and leaving
                                         '17 Why did you abandon the “Bag”, “Shopping Cart”?\t\t\t\t\t
```

```
In [18]: dataset= dataset.rename(columns={'18 The content on the website must be easy to read and u
'19 Information on similar product to the one highlighted
'20 Complete information on listed seller and product bei
'22 Ease of navigation in website': 'Website_navigation',
```

```
In [19]: dataset= dataset.rename(columns={'26 Trust that the online retail store will fulfill its
'27 Empathy (readiness to assist with queries) towards th
'28 Being able to guarantee the privacy of the customer':
'30 Online shopping gives monetary benefit and discounts'
'32 Shopping online is convenient and flexible':'Ease_of_
'33 Return and replacement policy of the e-tailer is impo
```

```
In [20]: dataset= dataset.rename(columns={'37 Net Benefit derived from shopping online can lead to
'39 Offering a wide variety of listed product in several
'42 The Convenience of patronizing the online retailer':
'43 Shopping on the website gives you the sense of advent
'45 You feel gratification shopping on your favorite e-ta
'47 Getting value for money spent': 'Money_value',
'41 Monetary savings': 'Monetary_savings'})

dataset
```

Out[20]:

	Gender	Age	Shopping_city	Pincode	From_when	Time_spent/year	Exploring_time	Payment_options	Cart
0	0	3	0	110009	5	4	3	4	
1	1	2	0	110030	5	5	5	1	
2	1	2	4	201308	4	5	4	4	
3	0	2	6	132001	4	1	3	1	
4	1	2	1	530068	3	2	5	1	
...	...	...	...	...	...	...	...	...	...
106	1	3	3	201008	2	4	1	2	
107	1	1	3	201001	1	1	1	2	
112	0	3	2	203001	2	1	3	2	
116	1	2	1	560037	3	2	5	1	
117	0	3	9	201308	5	5	5	4	

104 rows × 28 columns

```
In [21]: #
```

```
In [22]: #dataset['Shopping_city'] = dataset['Shopping_city'].tostring()
```

# Remove NaN values from dataset

```
In [23]: dataset.isnull()
```

Out[23]:

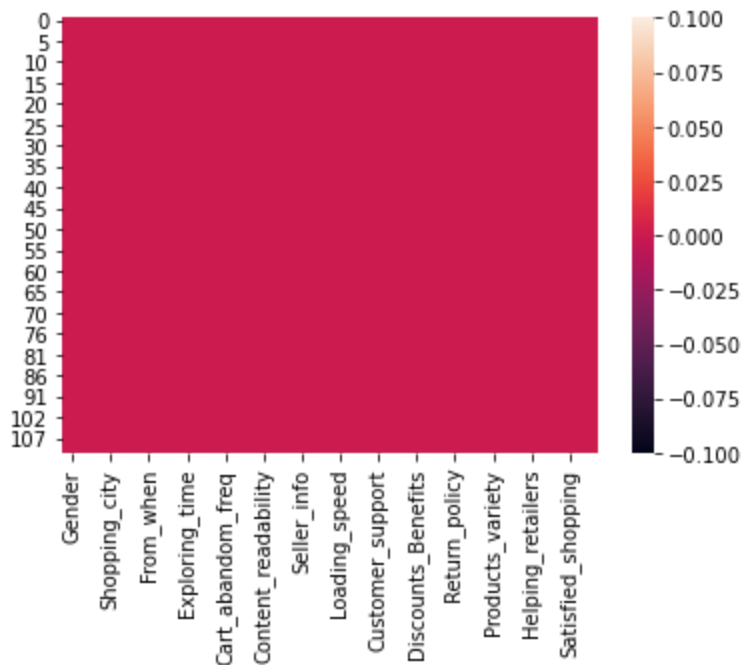
	Gender	Age	Shopping_city	Pincode	From_when	Time_spent/year	Exploring_time	Payment_options	Car
0	False	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	
...	...	...	...	...	...	...	...	...	...
106	False	False	False	False	False	False	False	False	
107	False	False	False	False	False	False	False	False	
112	False	False	False	False	False	False	False	False	
116	False	False	False	False	False	False	False	False	
117	False	False	False	False	False	False	False	False	

	Gender	Age	Shopping_city	Pincode	From_when	Time_spent/year	Exploring_time	Payment_options	Car
2	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...
106	False	False	False	False	False	False	False	False	False
107	False	False	False	False	False	False	False	False	False
112	False	False	False	False	False	False	False	False	False
116	False	False	False	False	False	False	False	False	False
117	False	False	False	False	False	False	False	False	False

104 rows × 28 columns

```
In [24]: sns.heatmap(dataset.isnull())
```

Out[24]: <AxesSubplot:>



```
In [25]: dataset.dtypes
```

```
Out[25]: Gender                int64
Age                int64
Shopping_city      int64
Pincode            int64
From_when          int64
Time_spent/year    int64
Exploring_time     int64
Payment_options    int64
Cart_abandon_freq  int64
Abandon_reason     int64
Content_readability int64
Comparision_info   int64
Seller_info        int64
Website_navigation int64
Loading_speed      int64
action            int64
```

```
Customer_support      int64
Customers_privacy      int64
Discounts_Benefits     int64
Ease_of_shopping      int64
Return_policy         int64
Used_satisfaction      int64
Products_variety       int64
Monetary_savings       int64
Helping_retailers      int64
Adventurous_shopping  int64
Satisfied_shopping     int64
Money_value           int64
dtype: object
```

```
In [26]: dataset=dataset.drop(['Pincode'],axis=1)
```

```
In [27]: dataset.describe()
```

```
Out[27]:
```

	Gender	Age	Shopping_city	From_when	Time_spent/year	Exploring_time	Payment_options	Cat
count	104.000000	104.000000	104.000000	104.000000	104.000000	104.000000	104.000000	
mean	0.634615	2.971154	4.115385	3.509615	2.846154	3.961538	1.817308	
std	0.483870	1.028306	3.434004	1.487783	1.676620	1.148431	1.147089	
min	0.000000	1.000000	0.000000	1.000000	1.000000	1.000000	1.000000	
25%	0.000000	2.000000	1.000000	3.000000	1.000000	3.000000	1.000000	
50%	1.000000	3.000000	4.000000	4.000000	3.000000	4.000000	1.000000	
75%	1.000000	4.000000	7.000000	5.000000	4.000000	5.000000	2.000000	
max	1.000000	5.000000	10.000000	5.000000	5.000000	5.000000	4.000000	

8 rows × 27 columns

```
In [28]: dataset.skew()
```

```
Out[28]: Gender      -0.567322
Age                0.058517
Shopping_city      0.269727
From_when         -0.533029
Time_spent/year    0.071437
Exploring_time    -0.747139
Payment_options    1.153734
Cart_abandon_freq -0.504534
Abandon_reason     0.742680
Content_readability -2.158898
Comparision_info   -0.653158
Seller_info        -0.707972
Website_navigation -1.956092
Loading_speed      -1.444516
Trust_on_transaction -1.221767
Customer_support   -2.335985
Customers_privacy  -1.303716
Discounts_Benefits -1.107883
Ease_of_shopping   -1.137267
Return_policy      -2.340356
Used_satisfaction  -1.174983
Products_variety   -0.571452
Monetary_savings   -1.336802
Helping_retailers  0.086649
Adventurous_shopping -0.454868
Loading_speed      -0.383867
```



Money\_value -0.197862  
dtype: float64

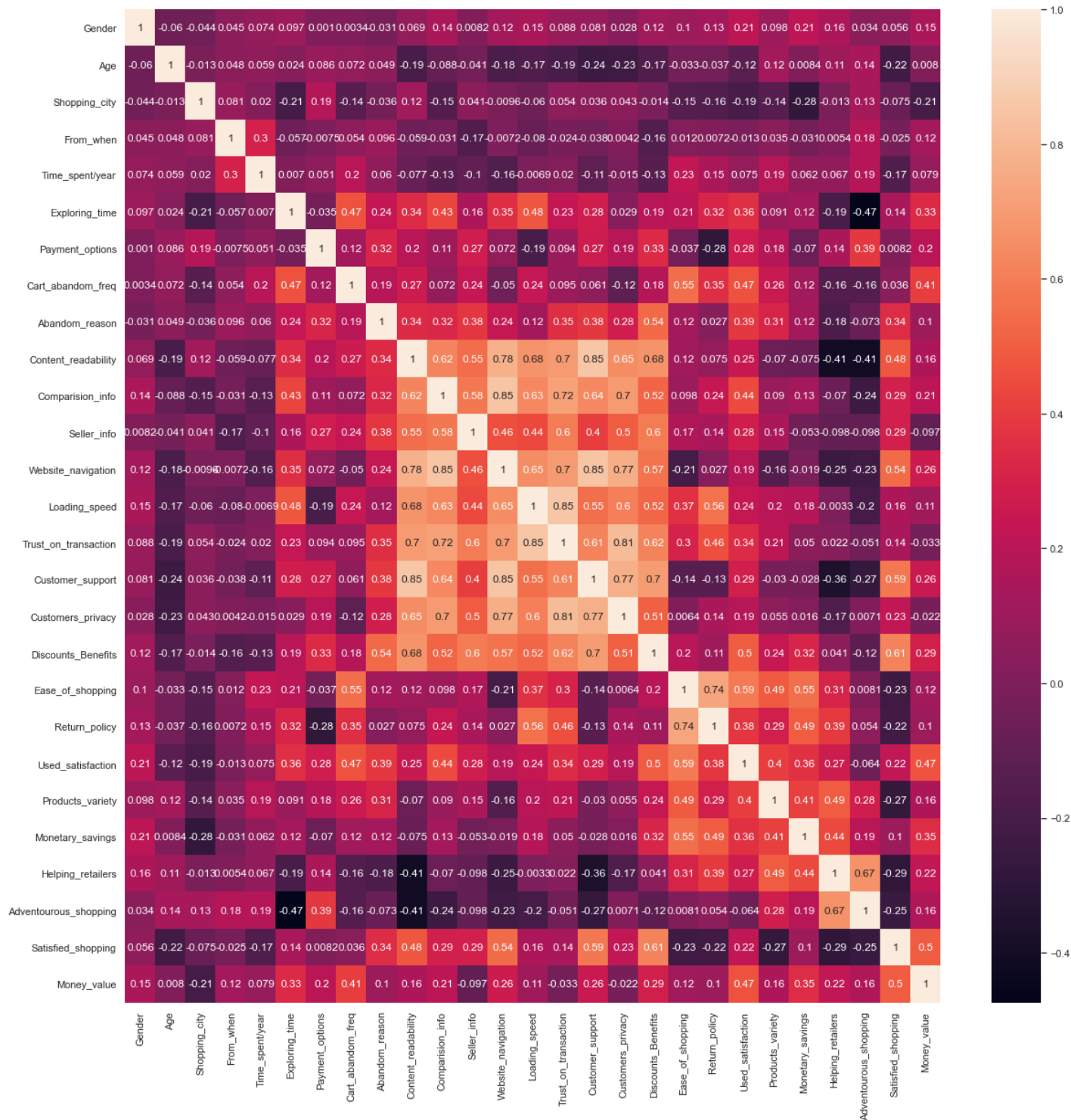
In [71]: #

In [73]: #

In [74]: #

## To Find Correlation between columns

```
In [58]: corr_hmap= dataset.corr()  
plt.figure(figsize=(20,20))  
sns.heatmap(corr_hmap,annot=True)  
plt.show()
```



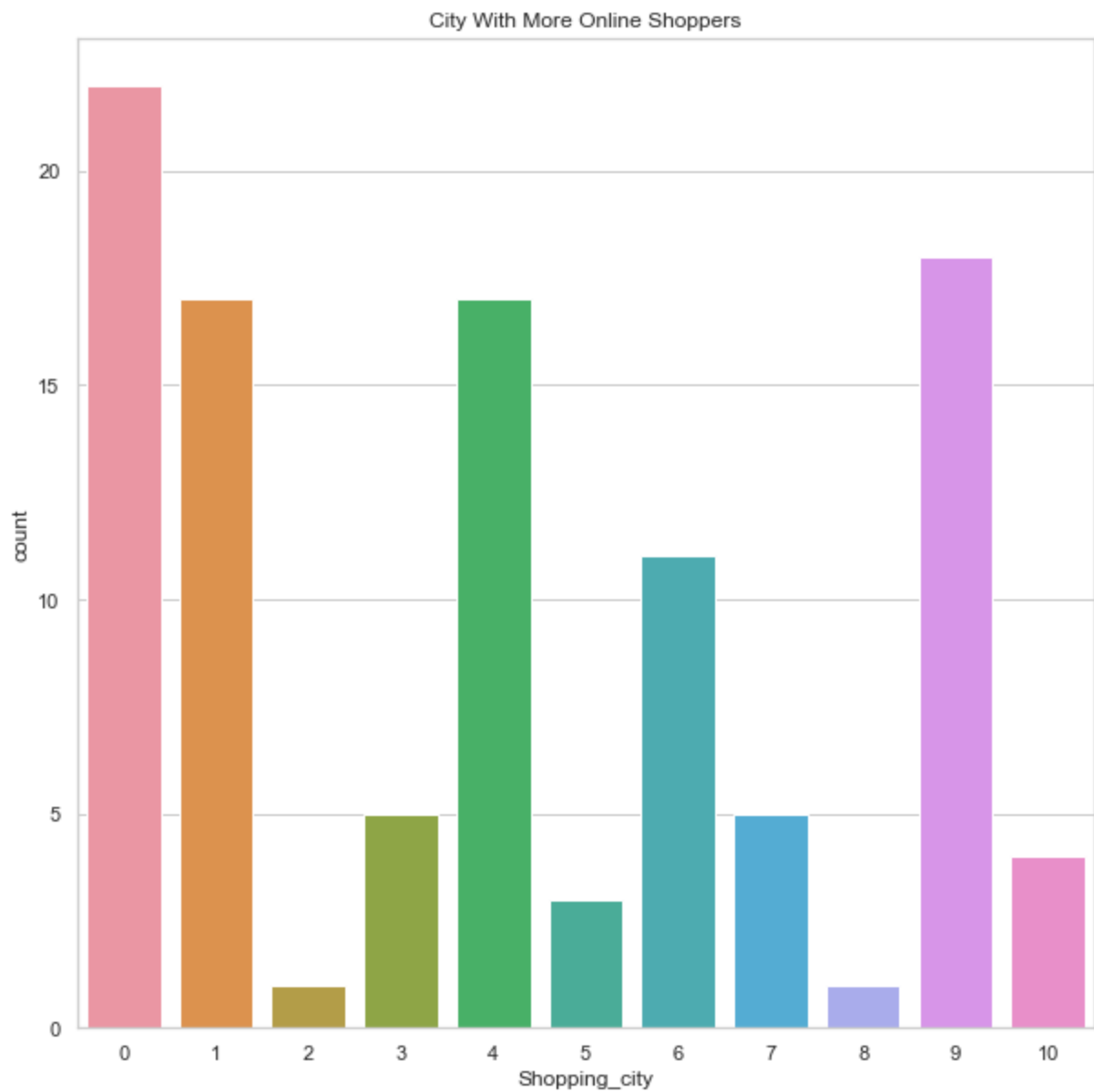
## Visualization

Count Plot

## City With More Online Customers

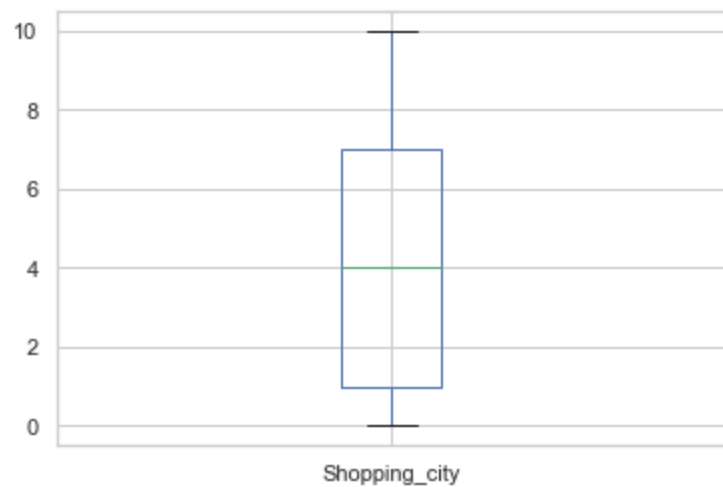
```
In [59]: sns.countplot(dataset.Shopping_city)
fig=plt.gcf()
fig.set_size_inches(10,10)
plt.title('City With More Online Shoppers') # delhi records more online shoppers
```

Out[59]: Text(0.5, 1.0, 'City With More Online Shoppers')



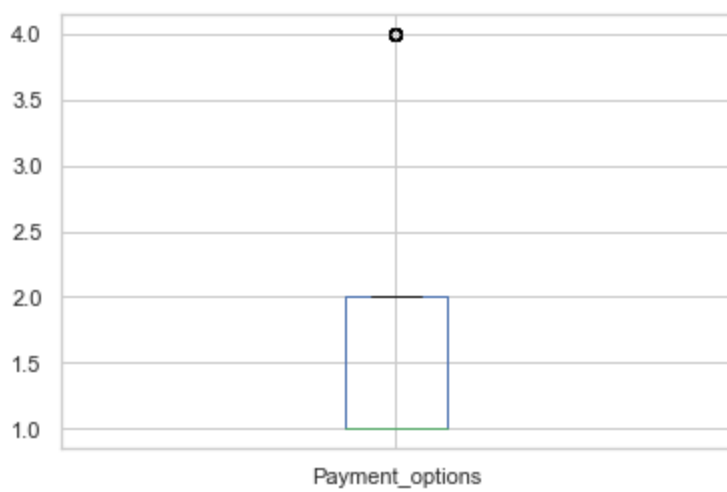
```
In [47]: dataset['Shopping_city'].plot.box()
```

```
Out[47]: <AxesSubplot:>
```



```
In [60]: dataset['Payment_options'].plot.box()
```

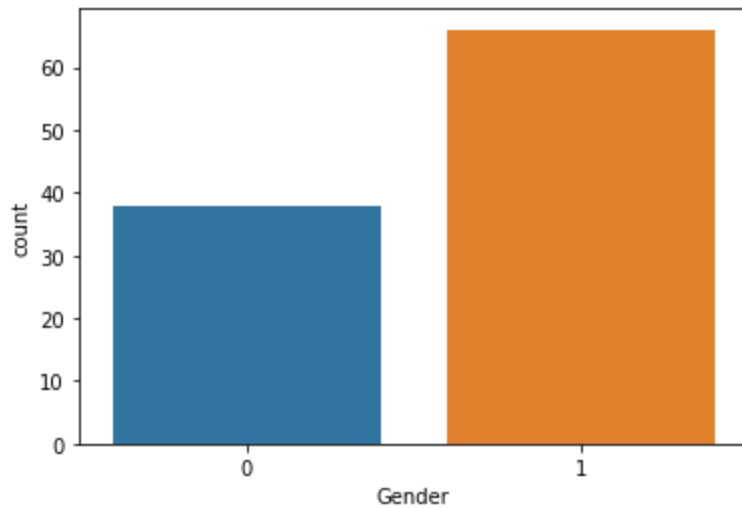
```
Out[60]: <AxesSubplot:>
```



## Female shoppers are more than male

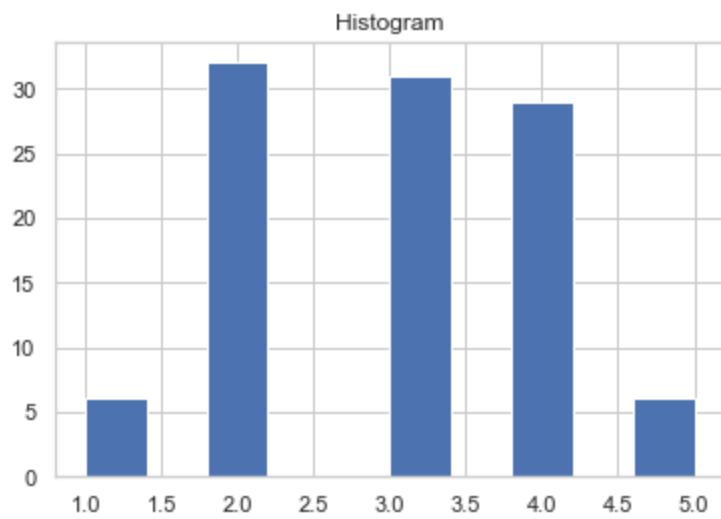
```
In [99]: sns.countplot(dataset.Gender)
```

```
Out[99]: <AxesSubplot:xlabel='Gender', ylabel='count'>
```



## Which age group of people makes frequent purchases

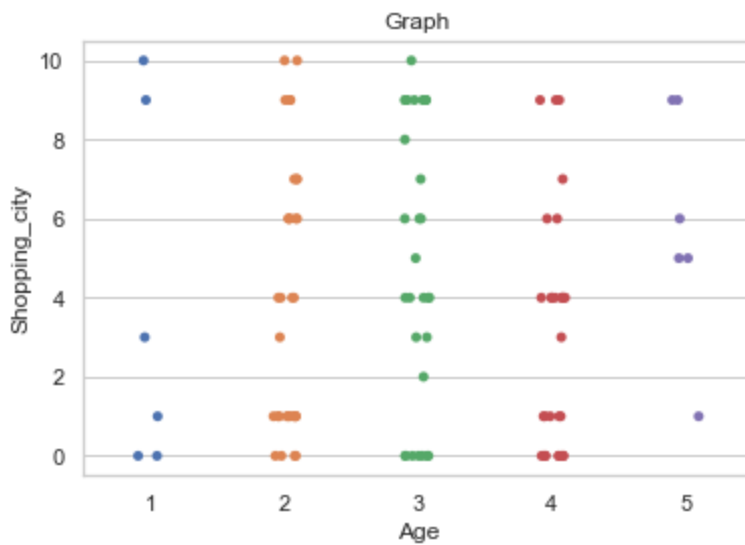
```
In [115... x= dataset['Age']  
plt.hist(x)  
plt.title('Histogram')  
plt.show()
```



## Age and gender based shopping in different cities

In [33]:

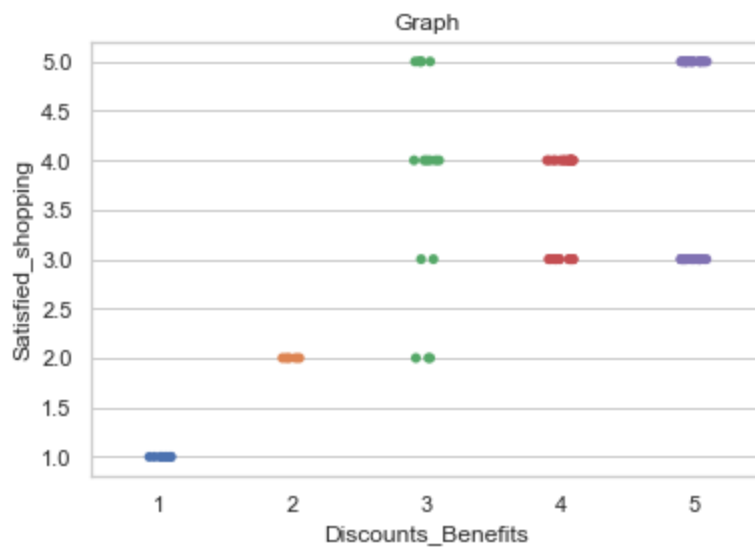
```
sns.set(style='whitegrid')
a=sns.stripplot(x='Age', y='Shopping_city', data=dataset)
plt.title('Graph')
plt.show() #female customers within age 21-30 shops more in online stores
```



In [31]:

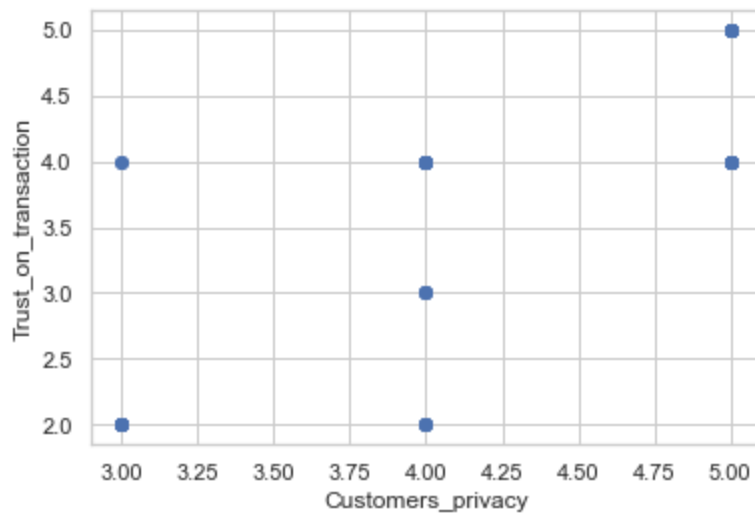
```
sns.set(style='whitegrid')
a=sns.stripplot(x='Gender', y='Shopping_city', data=dataset)
plt.title('Graph')
plt.show()
```





Customers Privacy and trust on transaction

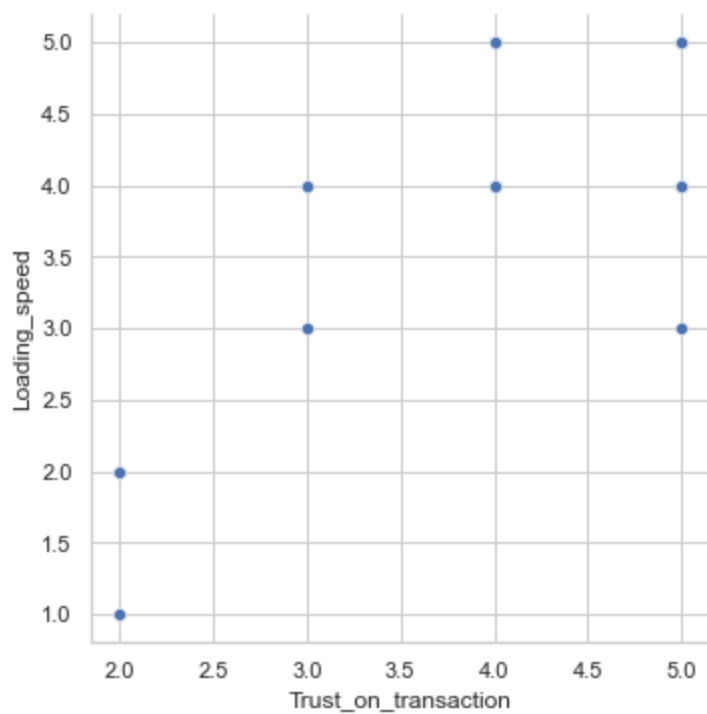
```
In [41]: x= dataset['Customers_privacy']
y= dataset['Trust_on_transaction']
plt.scatter(x,y)
plt.xlabel('Customers_privacy')
plt.ylabel('Trust_on_transaction')
plt.show()
```



Loading Speed and trust on transaction

```
In [120... sns.relplot(x='Trust_on_transaction', y='Loading_speed', data= dataset)
```

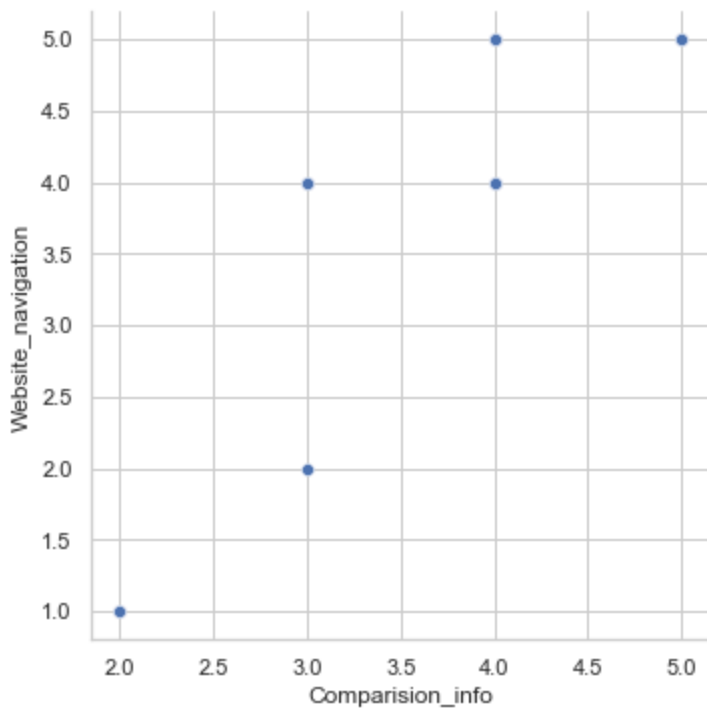
```
Out[120... <seaborn.axisgrid.FacetGrid at 0xeafe1c0>
```



Information on product comparision and website navigation

```
In [121... sns.relplot(x='Comparision_info', y='Website_navigation', data= dataset)
```

```
Out[121... <seaborn.axisgrid.FacetGrid at 0xc2fd3d0>
```

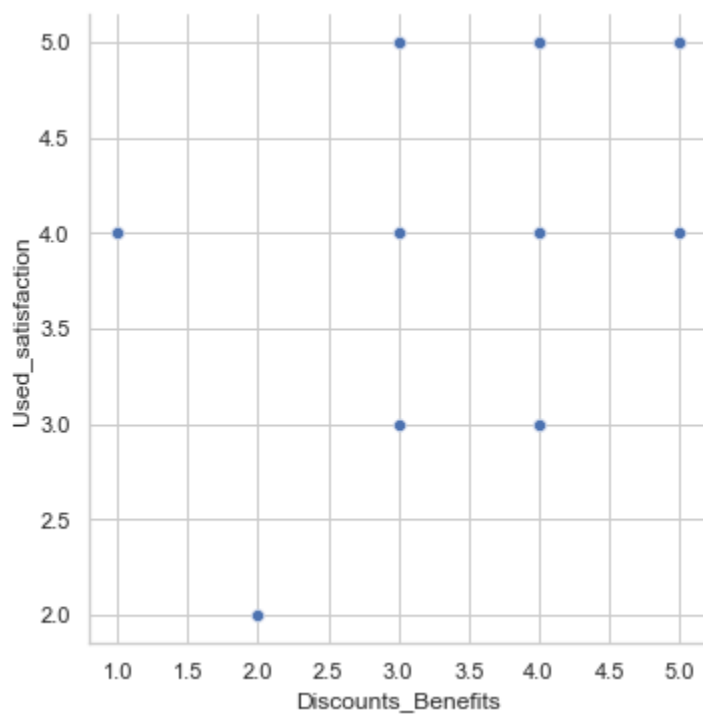


Discounts and user satisfaction

```
In [136... sns.relplot(x='Discounts_Benefits', y='Used_satisfaction', data= dataset)
```

```
Out[136... <seaborn.axisgrid.FacetGrid at 0xfd10e20>
```

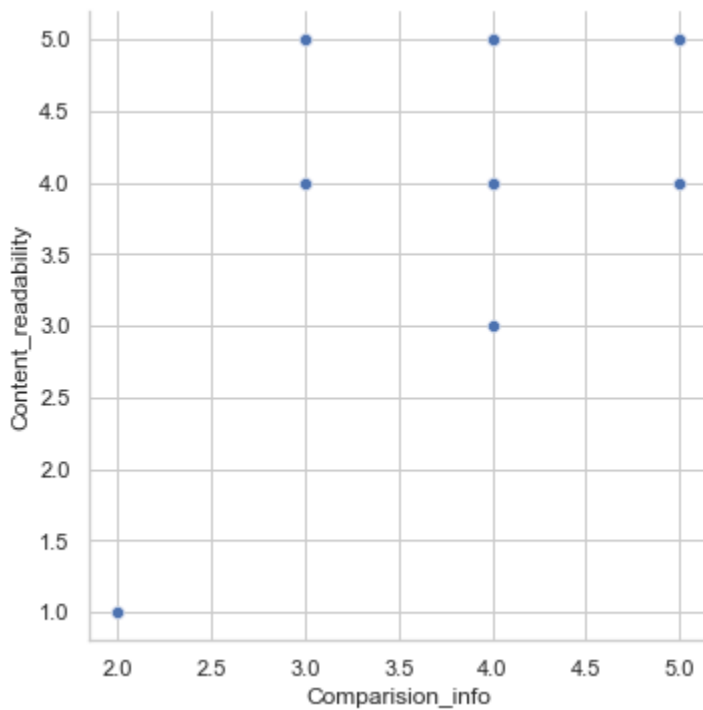




Comparison Information and ease of content reading

```
In [42]: sns.relplot(x='Comparison_info', y='Content_readability', data= dataset)
```

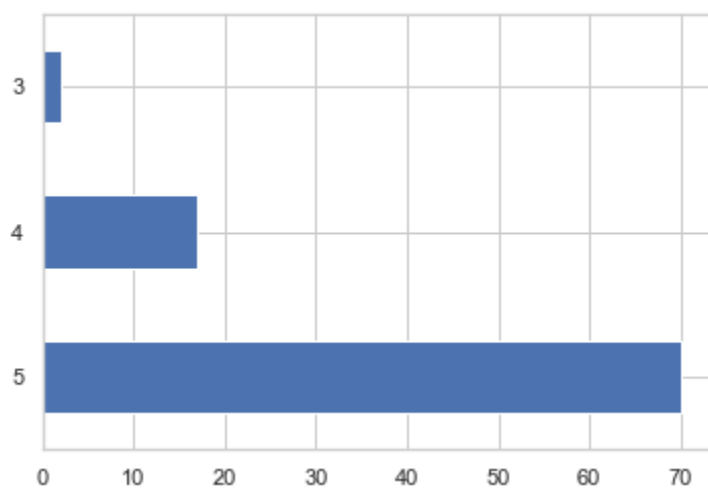
```
Out[42]: <seaborn.axisgrid.FacetGrid at 0xcc2f4f0>
```



Customers Privacy

```
In [98]: #sns.histplot(dataset['Customers_privacy'], bins=5)
dataset['Customers_privacy'].value_counts().plot(kind='barh')
```

```
Out[98]: <AxesSubplot:>
```

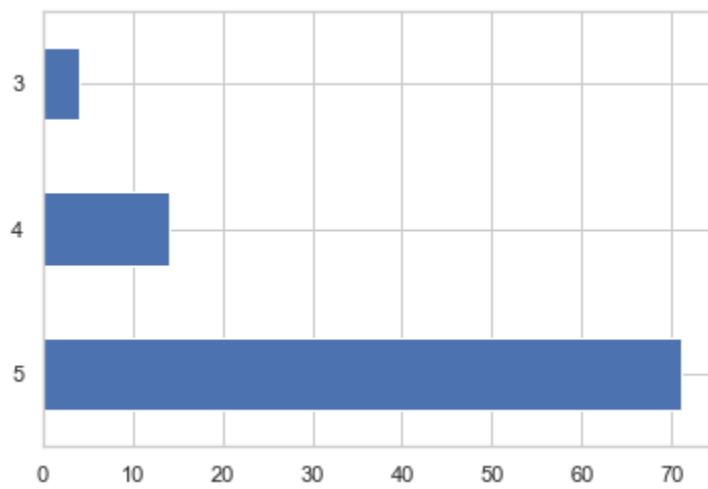


Customer\_Support

In [100...

```
dataset['Customer_support'].value_counts().plot(kind='barh')
```

Out[100... <AxesSubplot:>

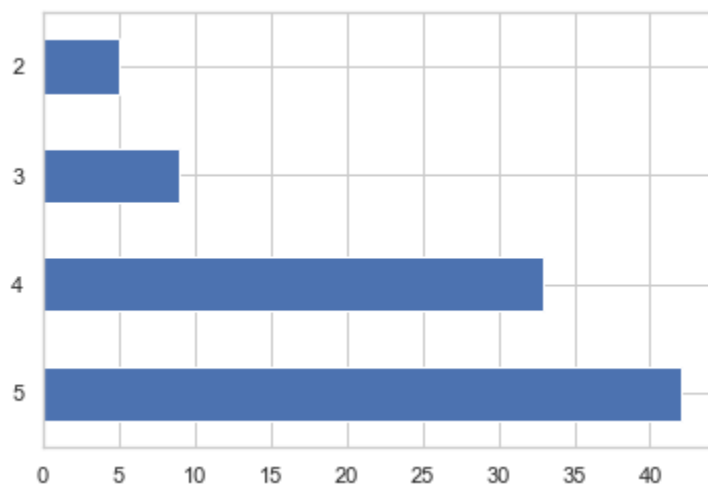


Discounts

In [99]:

```
dataset['Discounts_Benefits'].value_counts().plot(kind='barh')
```

Out[99]: <AxesSubplot:>



In [72]:

```
#removing Outliers
```

```
from scipy.stats import zscore
z=np.abs(zscore(dataset))
z
```

```
Out[72]: array([[1.31789306, 0.02818797, 1.2042252 , ..., 0.36735282, 1.30959215,
1.26332089],
[0.75878691, 0.948995 , 1.2042252 , ..., 0.61225471, 1.30959215,
1.26332089],
[0.75878691, 0.948995 , 0.03376332, ..., 0.36735282, 0.4144279 ,
0.28238938],
...,
[1.31789306, 0.02818797, 0.61899426, ..., 2.57146976, 0.4144279 ,
1.82809964],
[0.75878691, 0.948995 , 0.91160973, ..., 0.61225471, 1.30959215,
1.26332089],
[1.31789306, 0.02818797, 1.42931402, ..., 1.34696035, 0.4144279 ,
1.26332089]])
```

```
In [75]: # threshold for zscore is 3....., zscore greater than 3 is outliers
threshold =3
print(np.where(z>3))

(array([ 0,  7,  7, 21, 21, 31, 38, 38, 44, 52, 52, 53, 53, 59, 64, 64, 65,
        65, 66, 70, 70, 81, 96], dtype=int64), array([19,  9, 15,  9, 15, 19,  9, 15, 19,
        15, 19,  9, 15, 19, 19], dtype=int64))
```

```
In [76]: z[0][19]
```

```
Out[76]: 3.2504661418861547
```

```
In [68]: dataset_new= dataset[(z<3).all(axis=1)]
```

```
In [69]: dataset.shape
```

```
Out[69]: (104, 27)
```

```
In [70]: dataset_new.shape
```

```
Out[70]: (89, 27)
```

```
In [77]: dataset= dataset_new
dataset.shape
```

```
Out[77]: (89, 27)
```

```
In [78]: x=dataset.iloc[:,0:-1]
x.head()
```

```
Out[78]:
```

	Gender	Age	Shopping_city	From_when	Time_spent/year	Exploring_time	Payment_options	Cart_abandonm_fr
1	1	2	0	5	5	5	1	
2	1	2	4	4	5	4	4	
3	0	2	6	4	1	3	1	
4	1	2	1	3	2	5	1	

Gender	Age	Shopping_city	From_when	Time_spent/year	Exploring_time	Payment_options	Cart_abandon_fr
5	1	3	9	5	5	5	4

5 rows × 26 columns

```
In [81]: y= dataset.iloc[:, -1]
y.head()
```

```
Out[81]: 1    5
2    4
3    4
4    5
5    5
Name: Money_value, dtype: int64
```

```
In [82]: x.shape
```

```
Out[82]: (89, 26)
```

```
In [83]: y.shape
```

```
Out[83]: (89,)
```

```
In [84]: xtrain,xtest,ytrain,ytest= train_test_split(x,y, test_size=.20, random_state=42)
```

```
In [85]: xtrain.shape
```

```
Out[85]: (71, 26)
```

```
In [86]: xtest.shape
```

```
Out[86]: (18, 26)
```

```
In [87]: ytrain.shape
```

```
Out[87]: (71,)
```

```
In [88]: ytest.shape
```

```
Out[88]: (18,)
```

## Linear Regression

```
In [92]: lm=LinearRegression()
```

```
In [93]: lm.fit(xtrain,ytrain)
```

```
In [94]: lm.score(xtrain,ytrain)
```

Out[94]: 1.0

```
In [96]: #predict the value
pred= lm.predict(xtest)
print("predicted test result:", pred)
print("Actual :", ytest)
```

predicted test result: [5. 4. 4. 4. 4. 5.]  
4. 5. 4. 4. 5. 4.  
4. 4. 6.19275634 5. 4. 6.19275634]  
Actual : 50 5  
62 4  
34 4  
14 4  
57 4  
1 5  
78 4  
20 5  
12 4  
25 4  
5 5  
37 4  
48 4  
98 4  
102 4  
45 5  
92 4  
75 4  
Name: Money\_value, dtype: int64

```
In [97]: print('error:')

print('Mean_squared_error:', mean_squared_error(ytest,pred))
print('Mean_absolute_error:', mean_absolute_error(ytest,pred))

print('Root mean squared value:',np.sqrt(mean_squared_error(ytest,pred)))
```

error:  
Mean\_squared\_error: 0.5342422648556444  
Mean\_absolute\_error: 0.24363959376634176  
Root mean squared value: 0.7309187812990199

# Dataset2

```
In [113... df1= pd.read_excel(r'C:\python\Cus_retention1.xlsx')
```

```
In [114... df1
```

Out[114...

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	Quick to complete purchase
--	------------------------------------	----------------------------------	----------------------------------	--	---	---	----------------------------

	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
0	Amazon.in, Paytm.com	Paytm.com	Flipkart.com	Flipkart.com	Snapdeal.com	Snapdeal.com	Paytm.com	Paytm.com
1	Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com	Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com	Amazon.in, Myntra.com	Flipkart.com, Myntra.com	Amazon.in, Flipkart.com, Myntra.com	Amazon.in, Flipkart.com, Myntra.com	Myntra.com	Amazon.in, Flipkart.com, Myntra.com
2	Amazon.in, Paytm.com, Myntra.com	Amazon.in, Paytm.com, Myntra.com	Amazon.in, Paytm.com, Myntra.com	Amazon.in, Myntra.com	Amazon.in, Paytm.com, Myntra.com	Amazon.in, Paytm.com	Amazon.in, Paytm.com, Myntra.com	Amazon.in, Paytm.com, Myntra.com
3	Amazon.in, Flipkart.com, Paytm.com, Myntra.com...	Amazon.in, Flipkart.com, Paytm.com, Myntra.com...	Amazon.in, Flipkart.com, Paytm.com, Myntra.com...	Amazon.in, Flipkart.com	Amazon.in, Flipkart.com	Amazon.in, Flipkart.com, Snapdeal.com	Amazon.in, Flipkart.com, Paytm.com	Amazon.in, Flipkart.com, Paytm.com
4	Amazon.in, Flipkart.com, Paytm.com, Myntra.com...	Amazon.in, Flipkart.com, Paytm.com, Myntra.com...	Myntra.com	Myntra.com	Amazon.in, Flipkart.com, Paytm.com, Myntra.com...	Amazon.in	Amazon.in, Paytm.com, Myntra.com	Amazon.in, Flipkart.com, Paytm.com, Myntra.com
...	...	...	...	...	...	...	...	...
264	Amazon.in	Amazon.in	Amazon.in	Amazon.in	Amazon.in	Amazon.in	Amazon.in	Amazon.in
265	Amazon.in, Flipkart.com	Flipkart.com	Amazon.in	Amazon.in	Flipkart.com	Flipkart.com	Flipkart.com	Flipkart.com
266	Amazon.in, Flipkart.com, Snapdeal.com	Amazon.in	Amazon.in	Amazon.in	Amazon.in	Amazon.in	Amazon.in	Amazon.in
267	Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com	Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com	Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com	Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com	Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com	Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com	Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com	Flipkart.com, Myntra.com, Snapdeal.com
268	Amazon.in	Amazon.in	Amazon.in	Amazon.in	Amazon.in	Amazon.in	Amazon.in	Amazon.in

269 rows × 24 columns

```
In [ ]: df1.info()
```

```
In [115... df1.columns
```

```
Out[115... Index(['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',
      '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 [116... sns.heatmap(df1.isnull())
```

```
Out[116... <AxesSubplot:>
```



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

Complete, relevant description information of products

Visual appealing web-page layout

Reliability of the website or application

Availability of several payment options

Privacy of customers' information

Perceived Trustworthiness

Longer time to get logged in (promotion, sales period)

Late declaration of price (promotion, sales period)

Limited mode of payment on most products (promotion, sales period)

Change in website/Application design

Website is as efficient as before

In [125... df1.dtypes

Out[125... From the following, tick any (or all) of the online retailers you have shopped from;

object

Easy to use website or application

object

Visual appealing web-page layout

object

Wild variety of product on offer



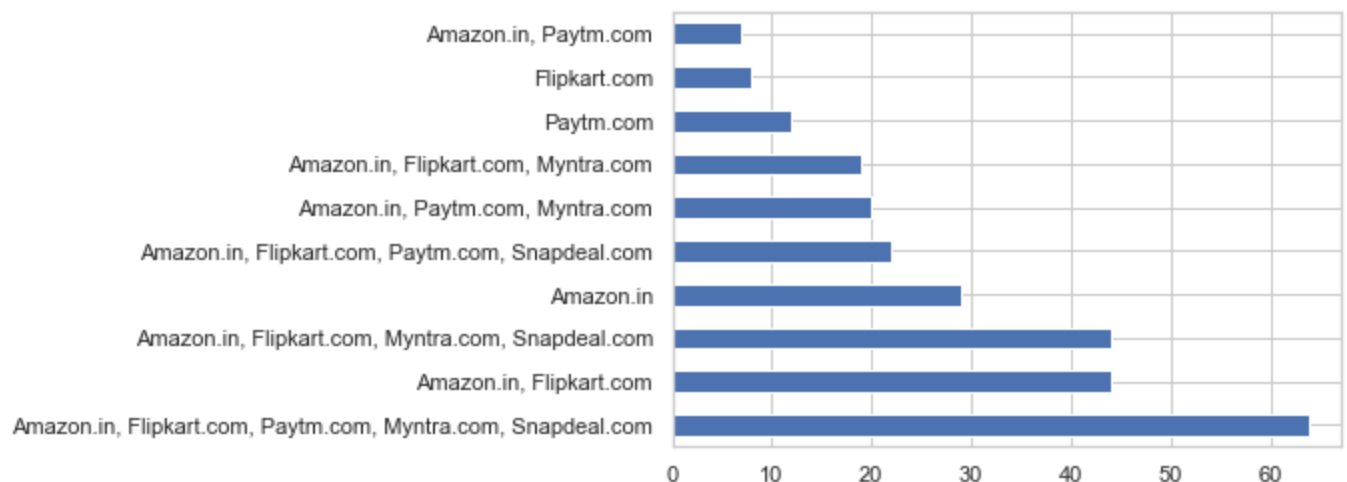
```

Complete, relevant description information of products
object
Fast loading website speed of website and application
object
Reliability of the website or application
object
Quickness to complete purchase
object
Availability of several payment options
object
Speedy order delivery
object
Privacy of customers' information
object
Security of customer financial information
object
Perceived Trustworthiness
object
Presence of online assistance through multi-channel
object
Longer time to get logged in (promotion, sales period)
object
Longer time in displaying graphics and photos (promotion, sales period)
object
Late declaration of price (promotion, sales period)
object
Longer page loading time (promotion, sales period)
object
Limited mode of payment on most products (promotion, sales period)
object
Longer delivery period
object
Change in website/Application design
object
Frequent disruption when moving from one page to another
object
Website is as efficient as before
object
Which of the Indian online retailer would you recommend to a friend?
object
dtype: object

```

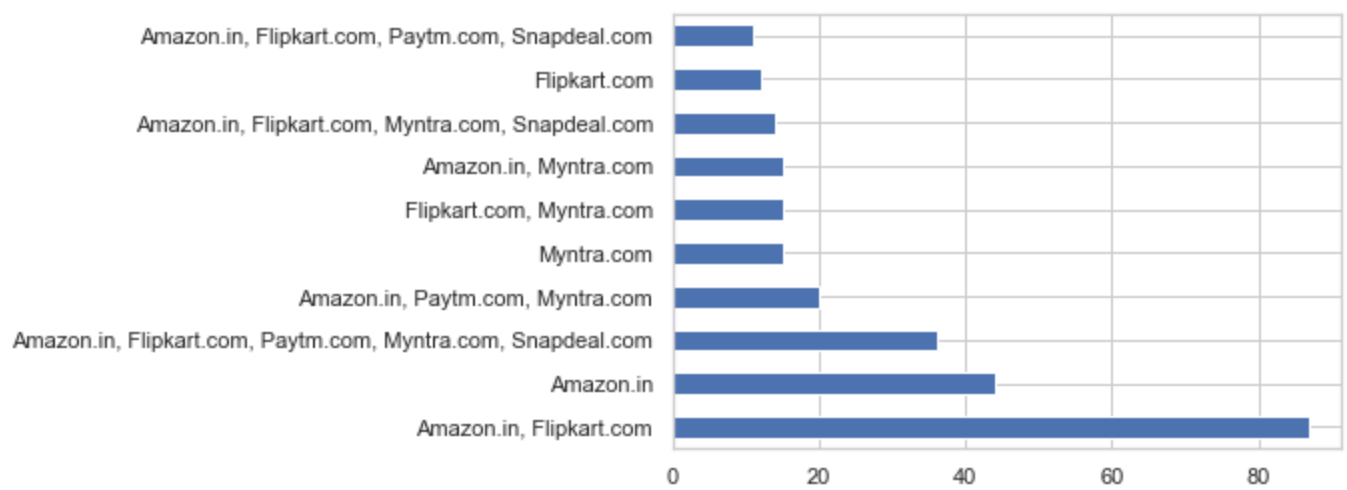
```
In [123... df1['Easy to use website or application'].value_counts().plot(kind='barh')
```

```
Out[123... <AxesSubplot:>
```



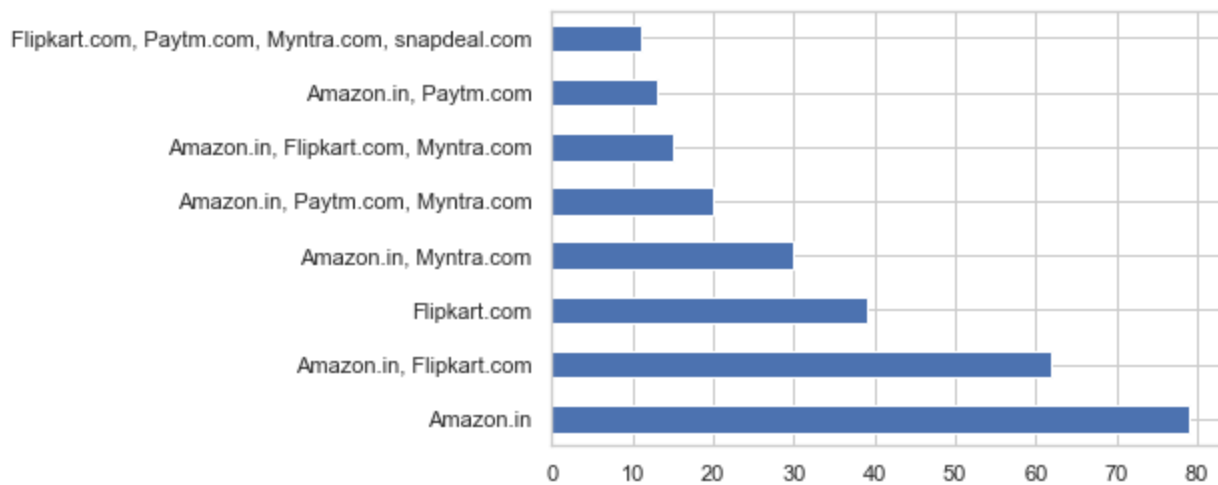
```
In [124... df1['Visual appealing web-page layout'].value_counts().plot(kind='barh')
```

```
Out[124... <AxesSubplot:>
```



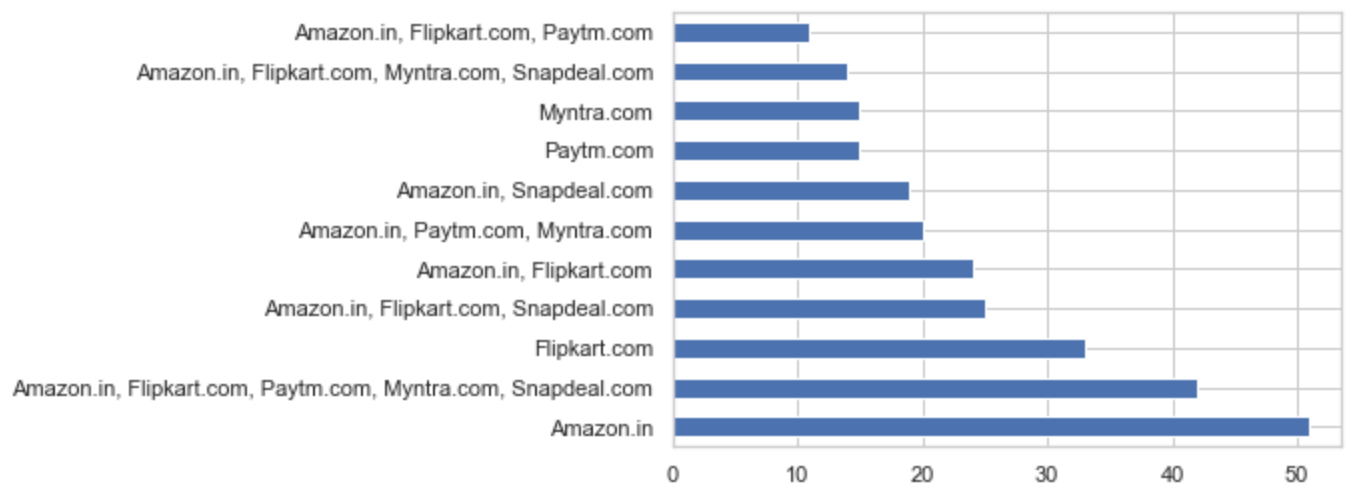
In [126... `df1['Which of the Indian online retailer would you recommend to a friend?'].value_counts()`

Out[126... `<AxesSubplot:>`



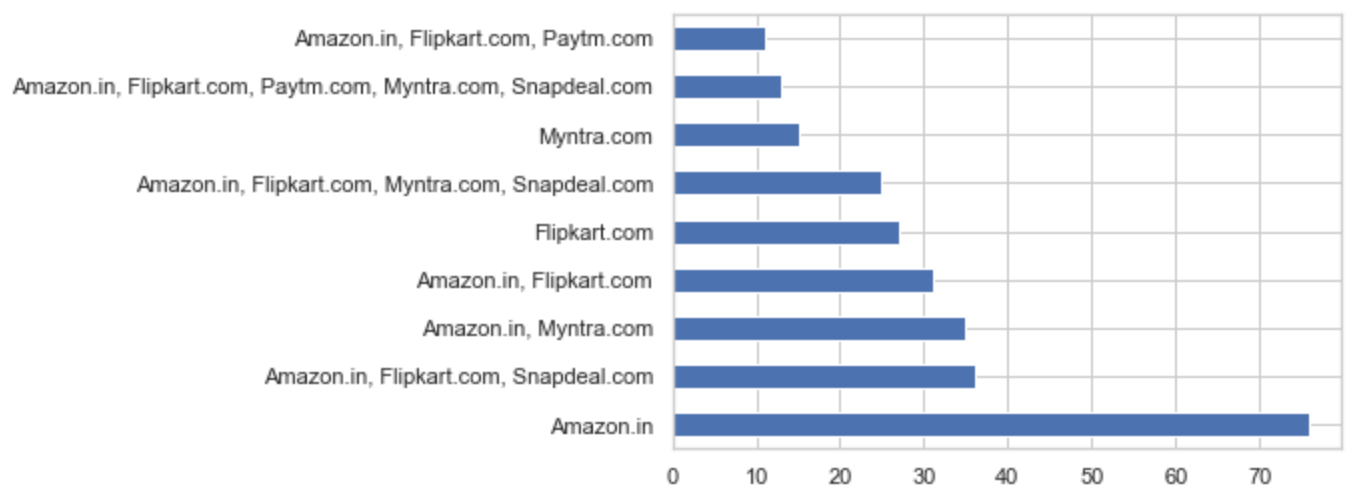
In [130... `df1['Security of customer financial information'].value_counts().plot(kind='barh')`

Out[130... `<AxesSubplot:>`



In [132... `df1['Perceived Trustworthiness'].value_counts().plot(kind='barh')`

Out[132... `<AxesSubplot:>`



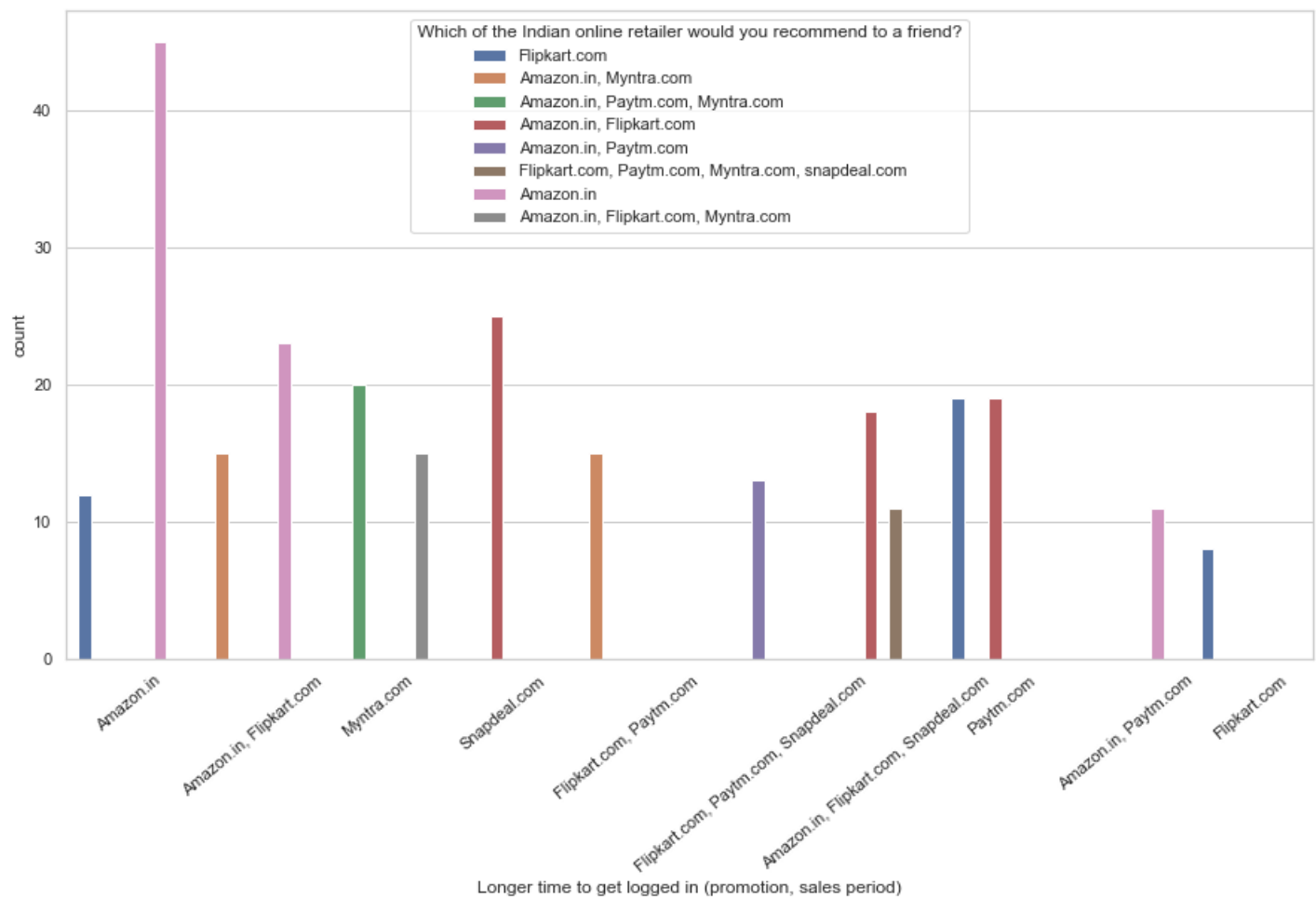
## Factors that negatively impact customer retention

In [127...

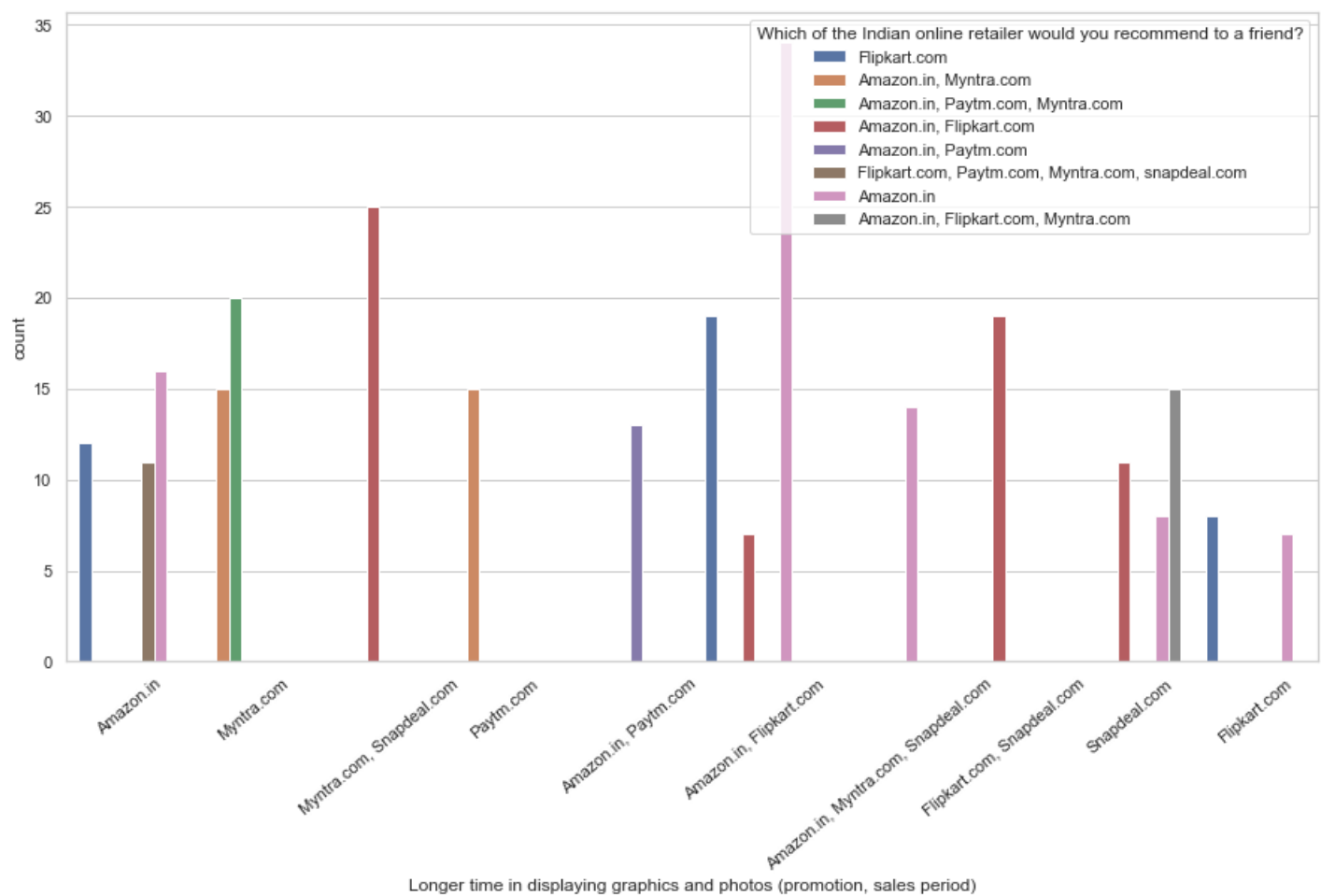
```
nv=['Longer time to get logged in (promotion, sales period)',
    'Longer time in displaying graphics and photos (promotion, sales period)',
    'Late declaration of price (promotion, sales period)',
    'Longer page loading time (promotion, sales period)',
    'Limited mode of payment on most products (promotion, sales period)',
    'Longer delivery period', 'Change in website/Application design',
    'Frequent disruption when moving from one page to another']
```

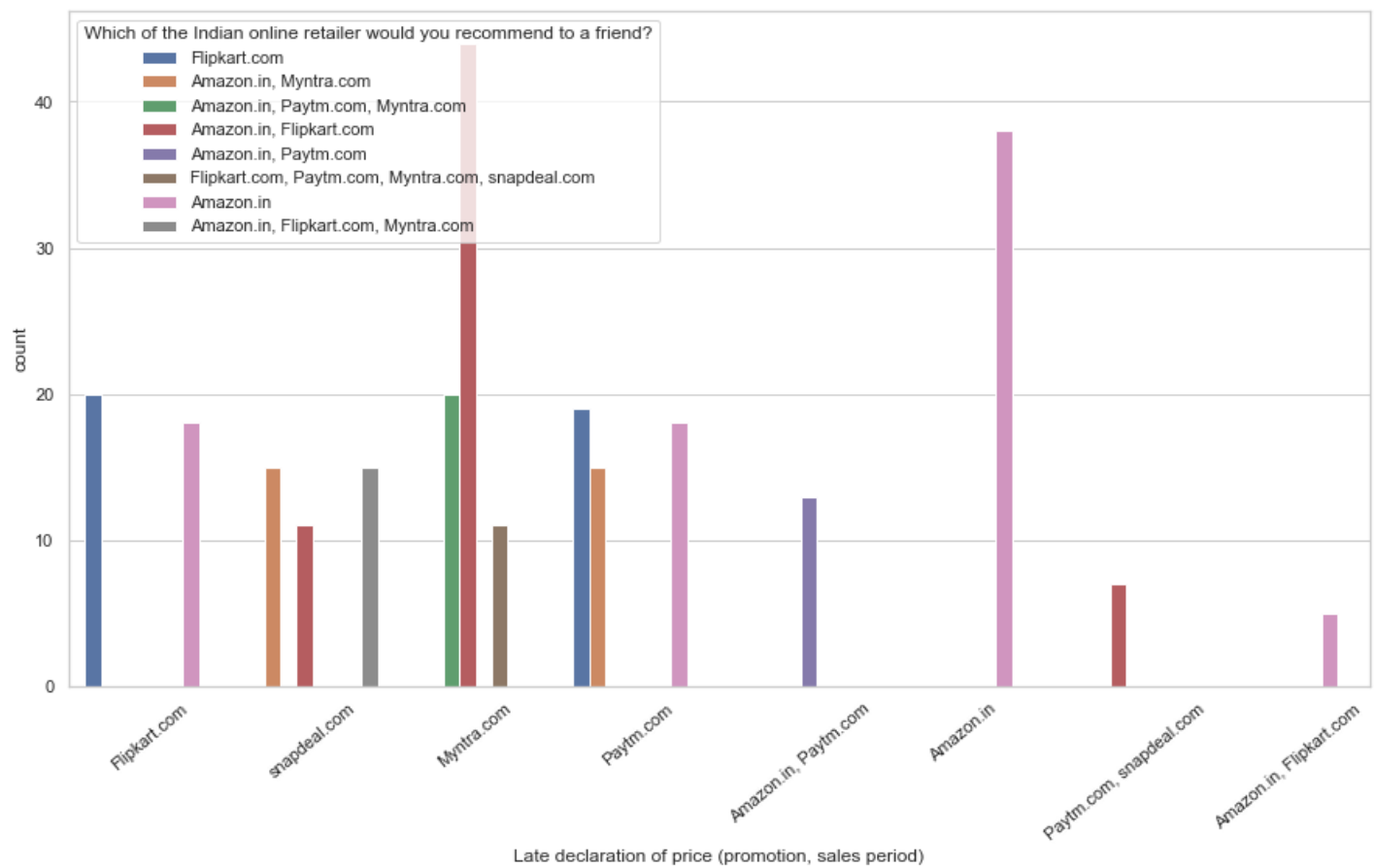
In [129...

```
for i in nv:
    plt.figure(figsize=(15,8))
    sns.countplot(df1[i],hue=df1['Which of the Indian online retailer would you recommend'])
    plt.xticks(rotation=40)
    plt.figure()
```

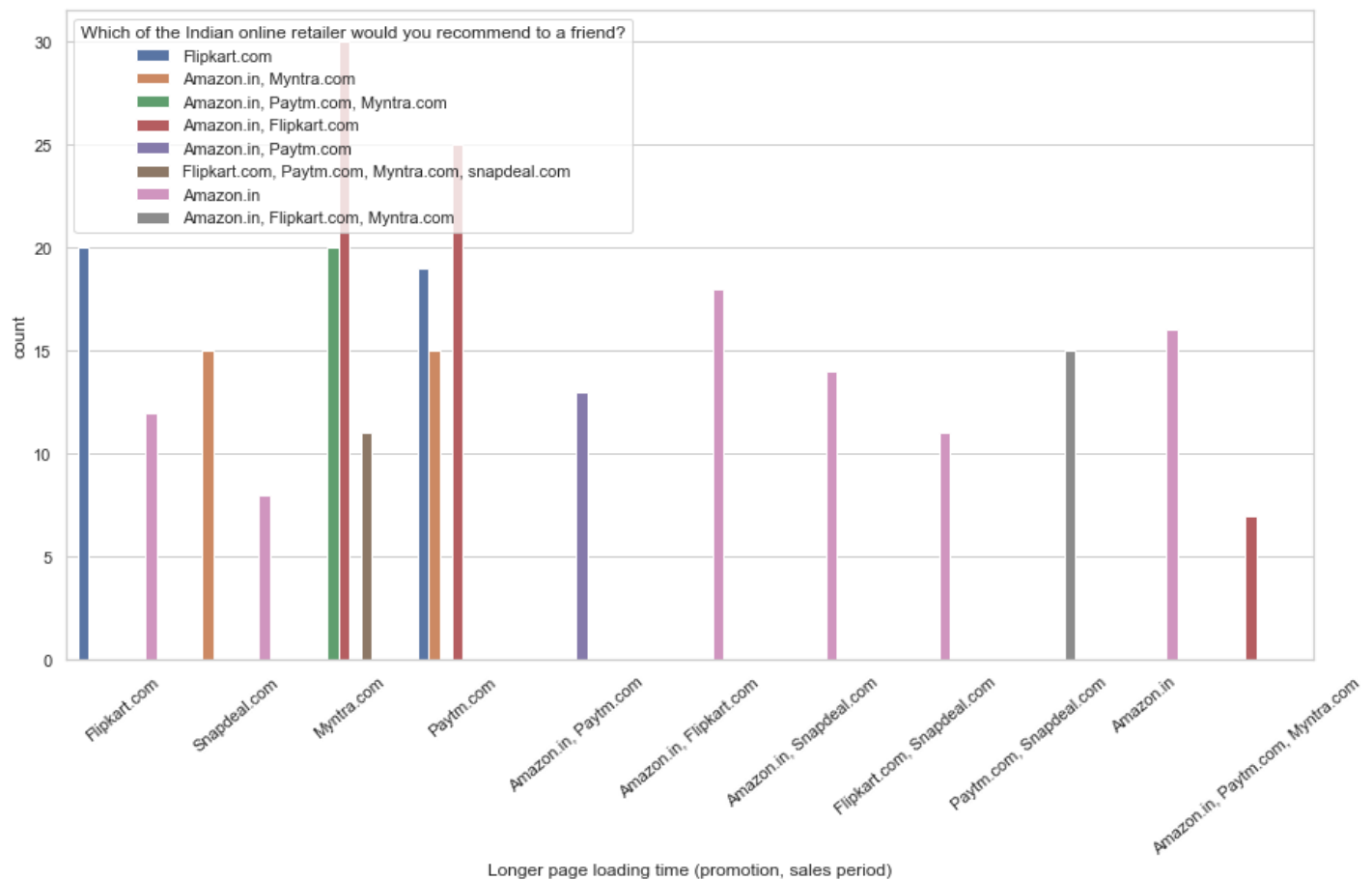


<Figure size 432x288 with 0 Axes>

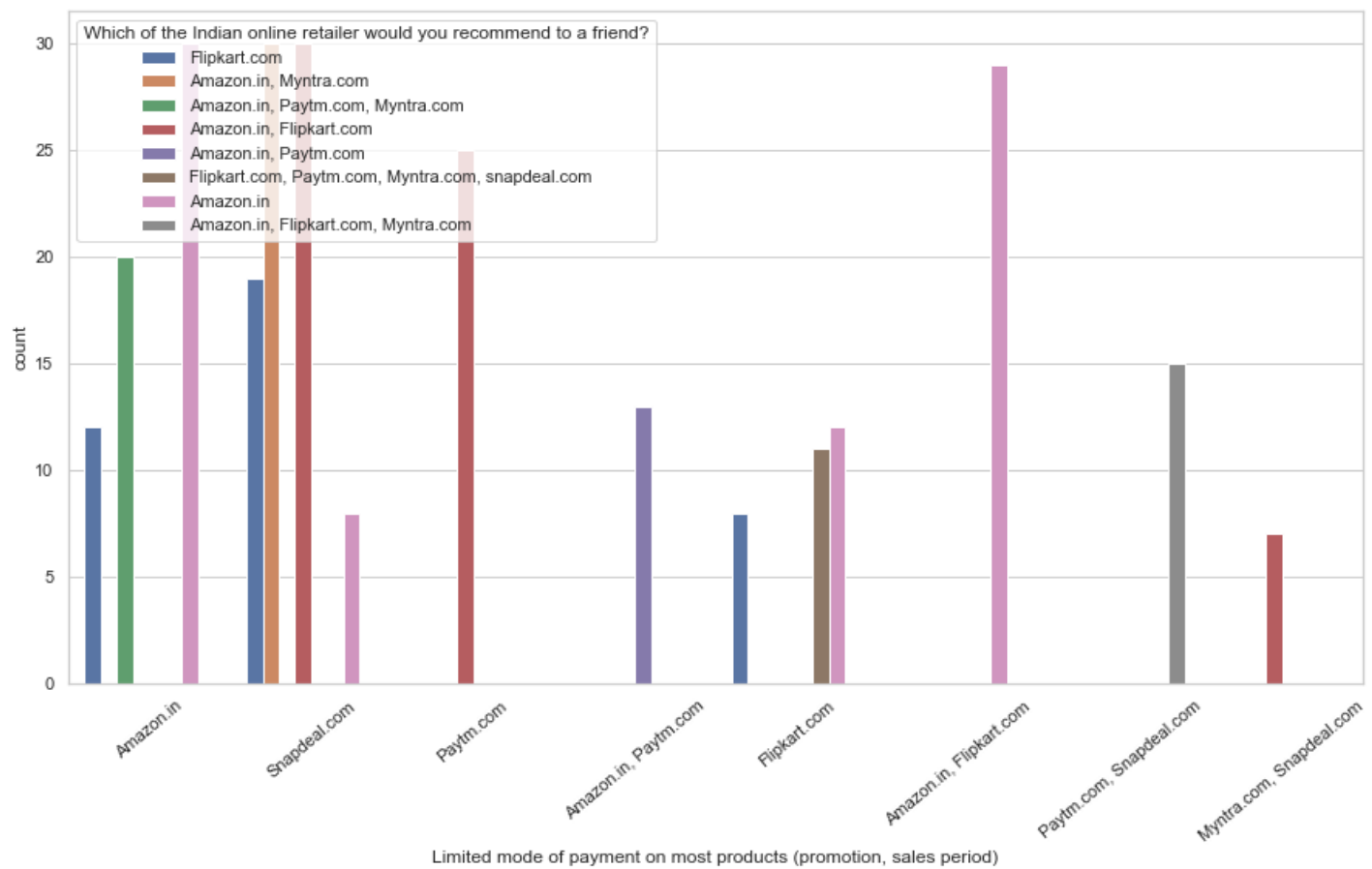




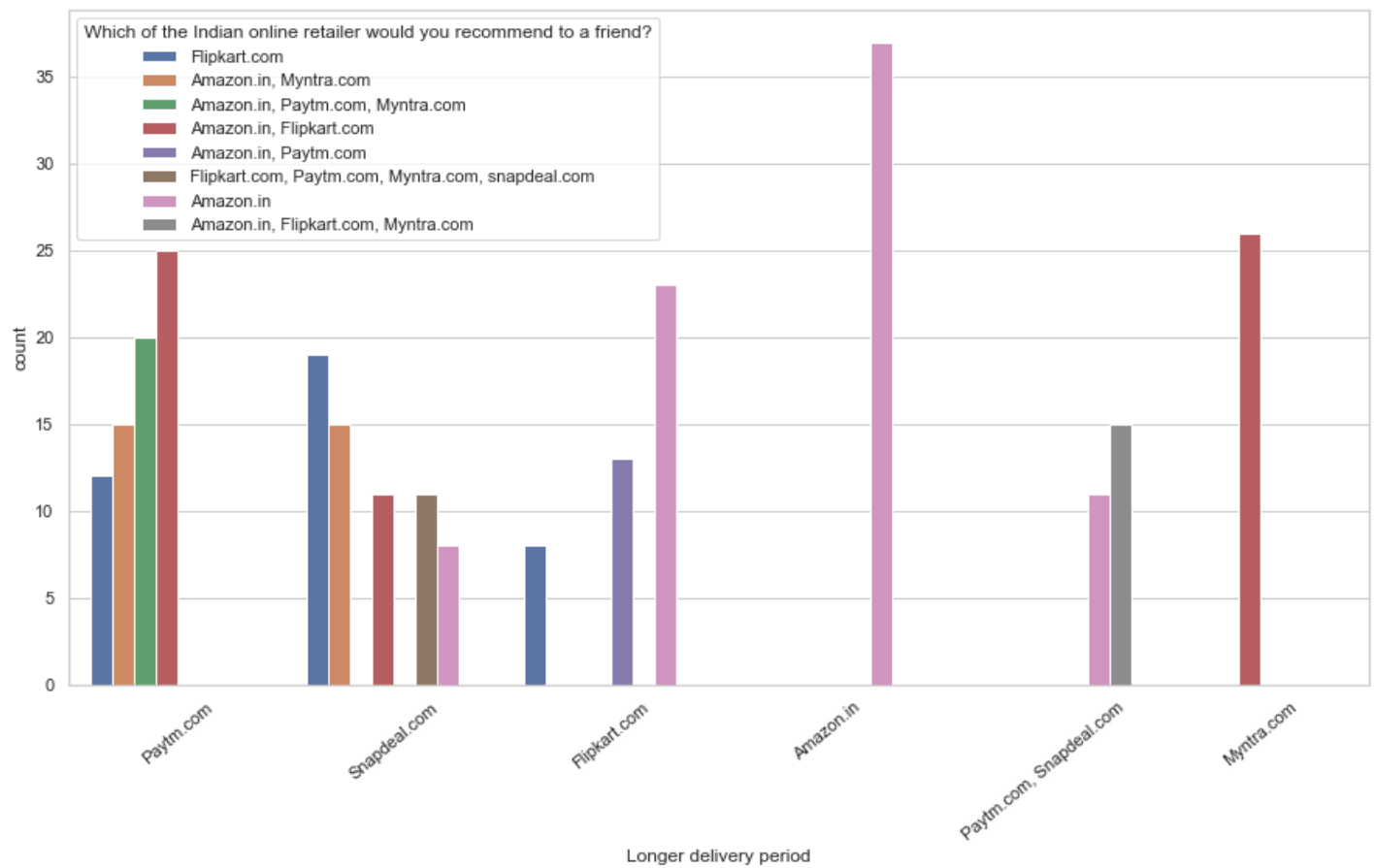
<Figure size 432x288 with 0 Axes>



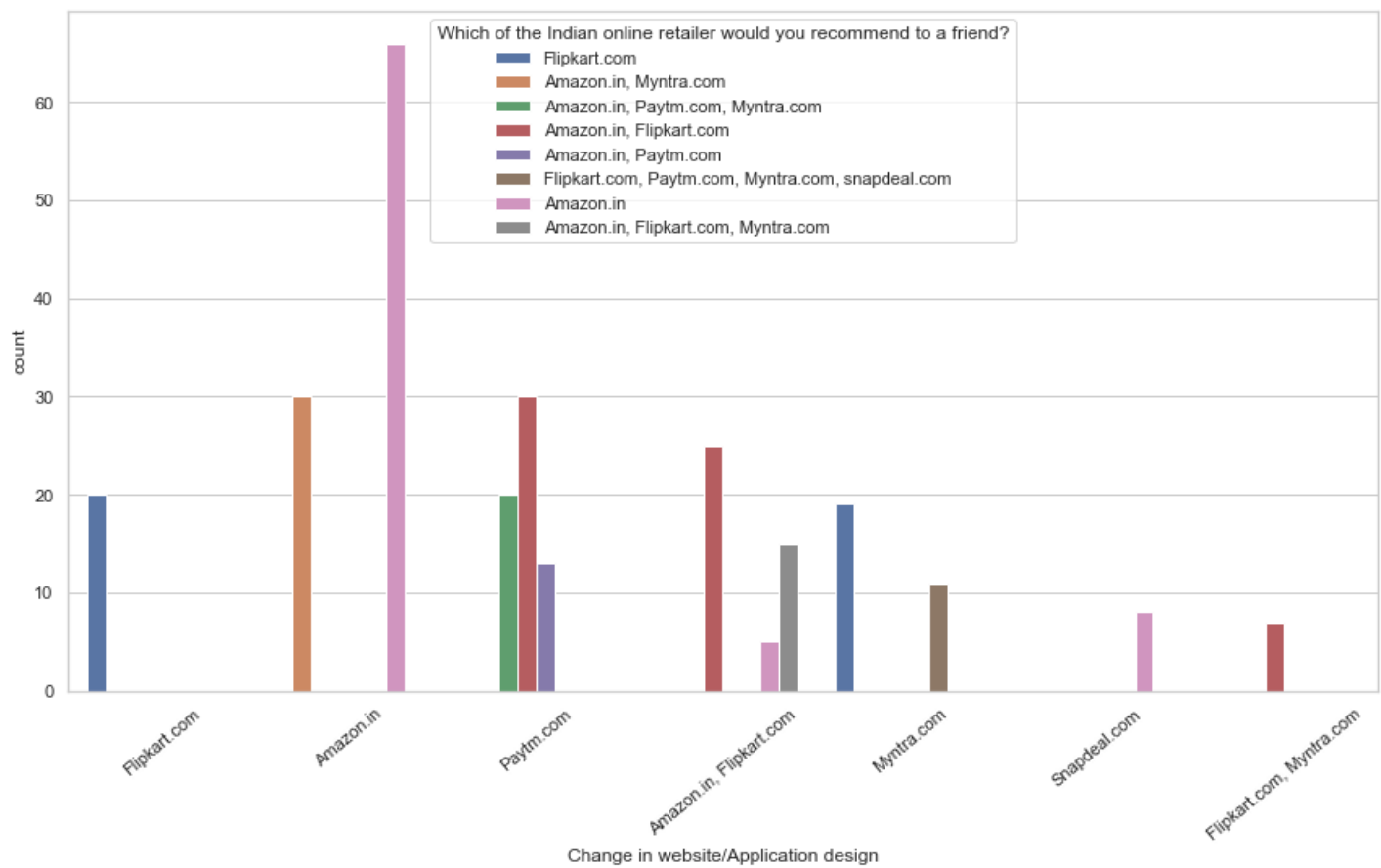
<Figure size 432x288 with 0 Axes>



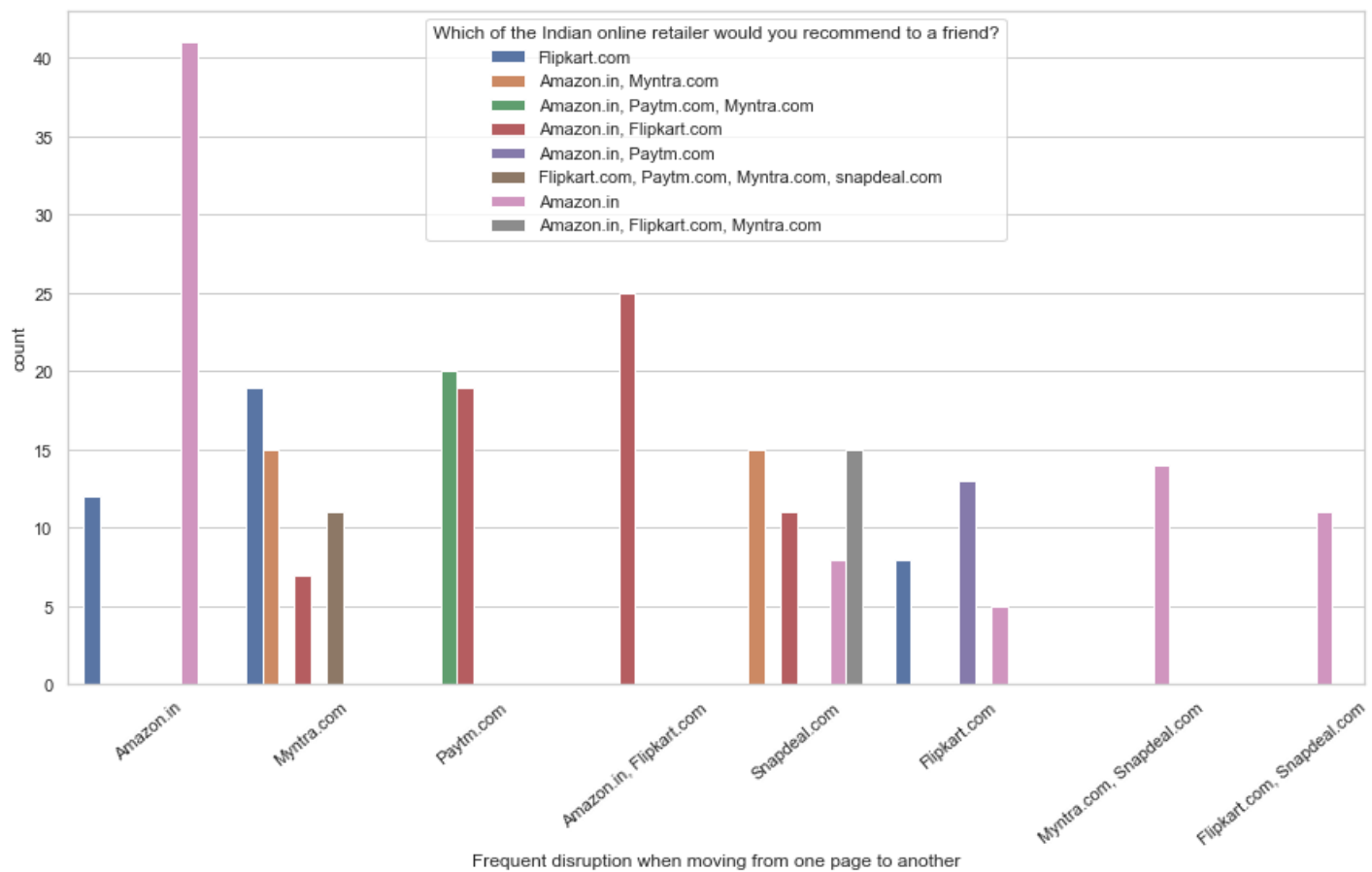
<Figure size 432x288 with 0 Axes>



<Figure size 432x288 with 0 Axes>



<Figure size 432x288 with 0 Axes>



<Figure size 432x288 with 0 Axes>

In [ ]: