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

Submitted by:

HARESH KHARSEL

ACKNOWLEDGMENT

I have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere thanks to all of them.

I am highly indebted to Flip Robo Technologies Bangalore for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project.

I want to thank my SME SWATI MAHASETH for providing the Dataset and helping us to solve the problem and addressing out ourQuery in right time.

I would like to express my gratitude towards my parents & members of Flip Robo for their kind co-operation and encouragement which help me in completion of this project.

I would like to express my special gratitude and thanks to industry persons for giving me such attention and time.

INTRODUCTION

Customer satisfaction has emerged as one of the most important factors that guarantee the success of online store; it has been posited as a key stimulant of purchase, repurchase intentions and customer loyalty. A comprehensive review of the literature, theories and models have been carried out to propose the models for customer activation and customer retention.

Five major factors that contributed to the success of an e-commerce store have been identified as: service quality, system quality, information quality, trust and net benefit. The research furthermore investigated the factors that influence the online customers repeat purchase intention. The combination of both utilitarian value and hedonistic values are needed to affect the repeat purchase intention (loyalty) positively.

Conceptual Background of the Domain Problem

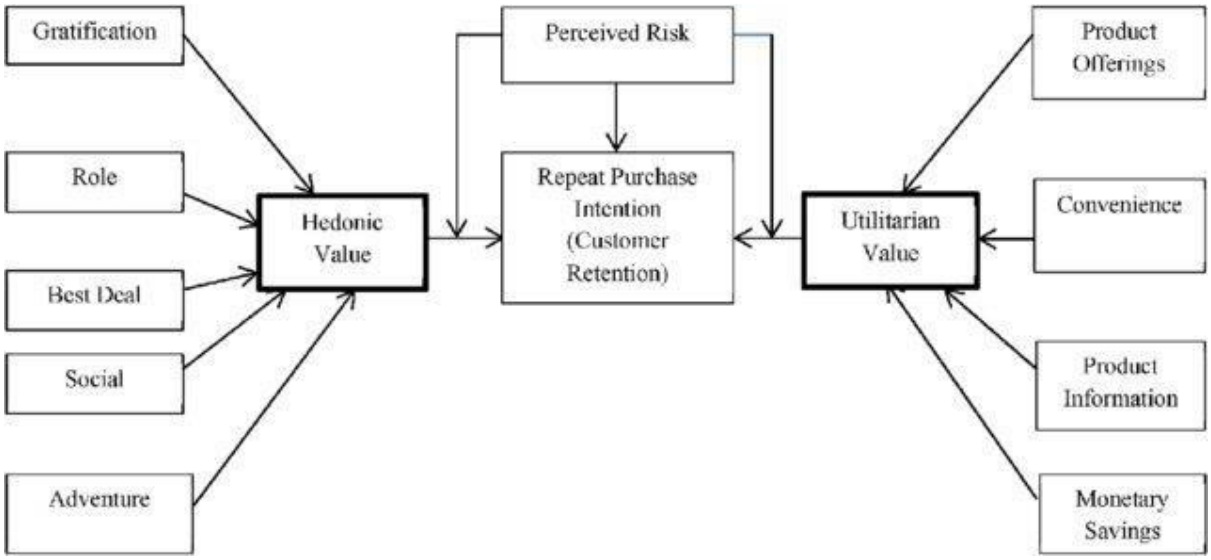
The data is collected from the Indian online shoppers. Results indicate the e-retail success factors, which are very much critical for customer satisfaction.

Five major factors that contributed to the success of an e-commerce store have been identified as: service quality, system quality, information quality, trust and net benefit. The research furthermore investigated the factors that influence the online customers repeat purchase intention.

Motivation for the Problem Undertaken

Our main objective of doing this project is to analyse whether the users are shopping products from e-commerce websites, how did they give feedbacks to these websites on the basis of several positive and negative factors and also the details of the users on basis of factors like age, gender, etc.

Diagrammatic Representation of Customer Retention



The Hedonic value consists of factors like Gratification, Role, Best Deal, Social and Adventure.

The Utilitarian value consists of factors like Product Offerings, Convenience, Product Information and Monetary Savings.

Customer Retention is based on 3 factors, according to the above diagram. They are:

Perceived Risk, Hedonic value and Utilitarian value

Data Sources and their formats

The data is been given by a highly-confidential company and they gave it to us in an excel file. They also had provided the problem statement by explaining what they need from us and also the required criteria to be satisfied.

Let's check the data now. Below I have attached the snapshot below to give an overview.

Loading the dataset

In [2]:
```
import pandas as pd
df=pd.read_excel('datasheet.xlsx')   #Dataset in excel format
df.head()  #Checking out the top 5 rows of the dataset
```

Out[2]:

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

5 rows × 71 columns

-> There are totally 269 rows and 71 columns in this dataset

-> Our objective is to find the insights of the data and to do thorough data analysis.

Hardware and Software Requirements and Tools Used

For doing this project, the hardware used is a laptop with high end specification and a stable internet connection. While coming to software part, I had used anaconda navigator and in that I have used **Jupyter notebook** to do my python programming and analysis.

For using an excel file, Microsoft excel is needed. In Jupyter notebook, I had used lots of python libraries to carry out this project and I have mentioned below with proper justification:

1. Pandas- a library which is used to read the data, visualisation and analysis of data.

2. NumPy- used for working with array and various mathematical techniques.

3. Seaborn- visualization tool for plotting different types of plot.

4. Matplotlib- It provides an object-oriented API for embedding plots into applications.

Data Analysis

There are 70 columns of object type and 1 column of int type

```
In [5]: df.isnull().sum()   #Checking for null values in the dataset
Out[5]: 1Gender of respondent                                          0
        2 How old are you?                                             0
        3 Which city do you shop online from?                          0
        4 What is the Pin Code of where you shop online from?          0
        5 Since How Long You are Shopping Online ?                     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
        Length: 71, dtype: int64
```

There are no null values in this dataset and 70 columns are of object datatype and only 1 column is of int data type.

Exploratory Data Analysis (EDA)

In [7]:
```
#Importing Matplotlib and Seaborn
import seaborn as sns
import matplotlib.pyplot as plt
```

In [8]:
```
#Checking the value counts for all features in the dataset
for i in df.columns:
    print(i)
    print(df[i].value_counts())
    print("\n")
```

```
1Gender of respondent
Female    181
Male       88
Name: 1Gender of respondent, dtype: int64


2 How old are you?
31-40 years           81
21-30 years           79
41-50 yaers           70
Less than 20 years    20
51 years and above    19
Name: 2 How old are you? , dtype: int64


3 Which city do you shop online from?
Delhi            58
Greater Noida    43
Noida            40
Bangalore        37
Karnal           27
Solan            18
Ghaziabad        18
```

We checked the value counts of all 71 columns above and we iterated using a for loop. We can see some value counts of the columns like gender, age, city, etc. Below I had attached the value counts of other columns.

```
8 Which device do you use to access the online shopping?
Smartphone    141
Laptop         86
Desktop        30
Tablet         12
Name: 8 Which device do you use to access the online shopping?, dtype: int64


9 What is the screen size of your mobile device?
Others        134
5.5 inches     99
4.7 inches     29
5 inches        7
Name: 9 What is the screen size of your mobile device?\t\t\t\t\t\t                                    , dtype: int64


10 What is the operating system (OS) of your device?
Window/windows Mobile    122
Android                   85
IOS/Mac                   62
Name: 10 What is the operating system (OS) of your device?\t\t\t\t                                      , dtype: int64


11 What browser do you run on your device to access the website?
Google chrome      216
Safari              40
Opera                8
Mozilla Firefox      5
Name: 11 What browser do you run on your device to access the website?\t\t\t
, dtype: int64


12 Which channel did you follow to arrive at your favorite online store for the first time?
Search Engine        230
Content Marketing     20
Display Adverts       19
Name: 12 Which channel did you follow to arrive at your favorite online store for the first time?
, dtype: int64
```

Analysis of website feedbacks obtained

We can see that after column 47, there are both positive and negative feedbacks of the websites, which are given by the correspondents. We will analyse those data by using data analysis process.

```
In [11]: #Extracting dataframe from where the websites feedback start
         df_feedback=df.iloc[:,47:]
         df_feedback
```

Out[11]:

	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	...	Longe to get l (prom sales p
0	Amazon.in, Paytm.com	Paytm.com	Flipkart.com	Flipkart.com	Snapdeal.com	Snapdeal.com	Paytm.com	Paytm.com	Patym.com	Amazon.in	...	Ama
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.com, Flipkart.com, Myntra.com	Amazon.in, Flipkart.com, Myntra.com	Amazon.in, Flipkart.com	...	Ama Flipka
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.com, Paytm.com, Myntra.com	Patym.com, Myntra.com	Amazon.in	...	Myntr
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.com, Flipkart.com, Paytm.com	Amazon.in, Flipkart.com, Myntra.com	Amazon.in, Flipkart.com, Snapdeal.com	...	Snapde
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.com, Flipkart.com, Paytm.com, Myntra.co...	Amazon.in, Flipkart.com, Patym.com, Myntra.com...	Amazon.in	...	Flipkar Paytr
...	...	...	...	...	...	...	...	...	...	...	...	
264	Amazon.in	Amazon.in	Amazon.in	Amazon.in	Amazon.in	Amazon.in	Amazon.in	Amazon.com	Amazon.in	Amazon.in	...	Ama
265	Amazon.in, Flipkart.com	Flipkart.com	Amazon.in	Amazon.in	Flipkart.com	Flipkart.com	Flipkart.com	Flipkart.com	Flipkart.com	Flipkart.com	...	Flipka

First, we will extract only the feedbacks data and then save it in a new data frame, which will be used for further process.

#A separate dataframe for displaying the positive feedback

dfnew1=df_feedback.drop(["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", "Frequent disruption when moving from one page to another"], axis=1)

A separate dataframe for displaying the negative feedback

dfnew2=df_feedback[["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", "Frequent disruption when moving from one page to another"]]

Now, we will analyse the negative feedbacks first by checking the count of websites and the type of feedbacks given to each website. Then, we will save the obtained data in a new data frame and rename the column names.

In [14]:
```
#Analysing the negative feedbacks separately
website_list=['Amazon.in','Flipkart.com','Paytm.com','Myntra.com','Snapdeal.com']  #Website list
col_names=[]   #Empty list for column names
websites=[]    #Empty list for websites
count=[]       #Empty list for checking the count of no of times the websites are mentioned

for i in dfnew2.columns:
    for j in website_list:
        present=len(dfnew2[dfnew2[i].str.contains(j)])  #Checking if websites in dataframe are available in the website list
        col_names.append(i)   #Appending the column names
        websites.append(j)    #Appending the website names
        count.append(present) #Appending the count of website present in feedback
```

In [15]:
```
#Creating negative feedback dataframe
negative_df=pd.DataFrame([col_names,websites,count])
negative_df
```

Out[15]:

	0	1	2	3	4	5	6	7	8	9	...	25	26	
0	Longer time to get logged in (promotion, sales...	Longer time to get logged in (promotion, sales...	Longer time to get logged in (promotion, sales...	Longer time to get logged in (promotion, sales...	Longer time to get logged in (promotion, sales...	Longer time in displaying graphics and photos ...	Longer time in displaying graphics and photos ...	Longer time in displaying graphics and photos ...	Longer time in displaying graphics and photos ...	Longer time in displaying graphics and photos ...	...	Longer delivery period	Longer delivery period	
1	Amazon.in	Flipkart.com	Paytm.com	Myntra.com	Snapdeal.com	Amazon.in	Flipkart.com	Paytm.com	Myntra.com	Snapdeal.com	...	Amazon.in	Flipkart.com	Payt
2	135	103	77	35	67	126	94	28	74	92	...	37	44	

3 rows × 35 columns

In [16]:
```
#Transposing the dataframe to see clearly
negative_df=negative_df.transpose()
negative_df
```

In [17]:
```
#Replacing the index value of columns with feature names
negative_df.rename(columns={0:'Particulars',1:'Website',2:'People_Count'},inplace=True)
negative_df   #Checking the dataframe after replacing
```

Out[17]:

	Particulars	Website	People_Count
0	Longer time to get logged in (promotion, sales...	Amazon.in	135
1	Longer time to get logged in (promotion, sales...	Flipkart.com	103
2	Longer time to get logged in (promotion, sales...	Paytm.com	77
3	Longer time to get logged in (promotion, sales...	Myntra.com	35
4	Longer time to get logged in (promotion, sales...	Snapdeal.com	67
5	Longer time in displaying graphics and photos ...	Amazon.in	126
6	Longer time in displaying graphics and photos ...	Flipkart.com	94
7	Longer time in displaying graphics and photos ...	Paytm.com	28
8	Longer time in displaying graphics and photos ...	Myntra.com	74
9	Longer time in displaying graphics and photos ...	Snapdeal.com	92
10	Late declaration of price (promotion, sales pe...	Amazon.in	56
11	Late declaration of price (promotion, sales pe...	Flipkart.com	43
12	Late declaration of price (promotion, sales pe...	Paytm.com	72
13	Late declaration of price (promotion, sales pe...	Myntra.com	75
14	Late declaration of price (promotion, sales pe...	Snapdeal.com	0
15	Longer page loading time (promotion, sales per...	Amazon.in	68
16	Longer page loading time (promotion, sales per...	Flipkart.com	61
17	Longer page loading time (promotion, sales per...	Paytm.com	94
18	Longer page loading time (promotion, sales per...	Myntra.com	68
19	Longer page loading time (promotion, sales per...	Snapdeal.com	63

Now, we will analyse the positive feedbacks by checking the count of websites and the type of feedbacks given to each website. Then, we will save the obtained data in a new data frame and rename the column names.

```
website_list=['Amazon.in','Flipkart.com','Paytm.com','Myntra.com','Snapdeal.com']  #Website list
col_names=[]   #Empty list for column names
websites=[]    #Empty list for websites
count=[]       #Empty list for checking the count of no of times the websites are mentioned

for i in dfnew1.columns:
    for j in website_list:
        present=len(dfnew1[dfnew1[i].str.contains(j)])  #Checking if websites in dataframe are available in the website list
        col_names.append(i)   #Appending the column names
        websites.append(j)    #Appending the website names
        count.append(present) #Appending the count of website present in feedback
```

In [22]:
```
#Creating positive feedback dataframe
positive_df=pd.DataFrame([col_names,websites,count])
positive_df
```

Out[22]:

	0	1	2	3	4	5	6	7	8	9	...	75	76	
0	From the following, tick any (or all) of the o...	From the following, tick any (or all) of the o...	From the following, tick any (or all) of the o...	From the following, tick any (or all) of the o...	From the following, tick any (or all) of the o...	Easy to use website or application	Easy to use website or application	Easy to use website or application	Easy to use website or application	Easy to use website or application	...	Website is as efficient as before	Website is as efficient as before	Web as el as I
1	Amazon.in	Flipkart.com	Paytm.com	Myntra.com	Snapdeal.com	Amazon.in	Flipkart.com	Paytm.com	Myntra.com	Snapdeal.com	...	Amazon.in	Flipkart.com	Paytr
2	269	221	150	146	182	249	201	125	147	130	...	182	117	

3 rows × 85 columns

In [23]:
```
#Transposing the dataframe and renaming the index with column names
positive_df=positive_df.transpose()
positive_df.rename(columns={0:'Particulars',1:'Website',2:'People_Count'},inplace=True)
positive_df
```

Calculating the percentage wise feedback analysis

Now, we will calculate the percentage of people giving the feedbacks to the website for both positive and negative data

In [29]:
```
#Adding a new column to the dataframe
positive_df['% user feedbacks']=0
positive_df.sort_values(by='Website',inplace=True)
positive_df.reset_index()
```

Out[29]:

	index	Particulars	Website	People_Count	% user feedbacks
0	0	From the following, tick any (or all) of the o...	Amazon.in	269	0
1	70	Change in website/Application design	Amazon.in	141	0
2	55	Security of customer financial information	Amazon.in	206	0
3	20	Complete, relevant description information of ...	Amazon.in	238	0
4	30	Reliability of the website or application	Amazon.in	227	0
...	...	...	...	...	...
80	44	Availability of several payment options	Snapdeal.com	90	0
81	49	Speedy order delivery	Snapdeal.com	50	0
82	59	Security of customer financial information	Snapdeal.com	100	0
83	19	Wild variety of product on offer	Snapdeal.com	14	0
84	84	Which of the Indian online retailer would you ...	Snapdeal.com	0	0

85 rows × 5 columns

In [30]:
```
#Index range values of all website with Maximum users value
#positive_df.iloc[0:17]   #Amazon data range and maximum users are 269
#positive_df.iloc[17:34]  #Flipkart data range and maximum users are 221
#positive_df.iloc[34:51]  #Myntra data range and maximum users are 147
#positive_df.iloc[51:68]  #Paytm data range and maximum users are 150
#positive_df.iloc[68:85]  #Snapdeal data range and maximum users are 182
```

In [31]:
```
#Calculating the percentage for all data
positive_df['% user feedbacks'].iloc[0:17]=(positive_df['People_Count']*100)/269 # Amazon users
positive_df['% user feedbacks'].iloc[17:34]=(positive_df['People_Count']*100)/221 # Flipkart users
positive_df['% user feedbacks'].iloc[34:51]=(positive_df['People_Count']*100)/147 # Myntra users
positive_df['% user feedbacks'].iloc[51:68]=(positive_df['People_Count']*100)/150 # Paytm users
positive_df['% user feedbacks'].iloc[68:85]=(positive_df['People_Count']*100)/182 # Snapdeal users
```

In [32]:
```
positive_df.reset_index()
```

Out[32]:

	index	Particulars	Website	People_Count	% user feedbacks
0	0	From the following, tick any (or all) of the o...	Amazon.in	269	100
1	70	Change in website/Application design	Amazon.in	141	52.4164
2	55	Security of customer financial information	Amazon.in	206	76.5799
3	20	Complete, relevant description information of ...	Amazon.in	238	88.4758
4	30	Reliability of the website or application	Amazon.in	227	84.3866
...	...	...	...	...	...
80	44	Availability of several payment options	Snapdeal.com	90	49.4505
81	49	Speedy order delivery	Snapdeal.com	50	27.4725
82	59	Security of customer financial information	Snapdeal.com	100	54.9451
83	19	Wild variety of product on offer	Snapdeal.com	14	7.69231
84	84	Which of the Indian online retailer would you ...	Snapdeal.com	0	0

85 rows × 5 columns

In [33]:
```
#Sorting values in Descending order
new_positive=positive_df.sort_values(by=['% user feedbacks'],ascending=False)
```

Observations:

1. Amazon and Flipkart rank about 90% in satisfying customers, followed by Myntra.
2. The maximum percentage Paytm and Snapdeal could score here is 83 and 71 respectively.
3. No one is willing to refer Snapdeal to their contacts as it has the less percentage among all websites.
4. On an average, Snapdeal and Paytm scores are less when compared to amazon, flipkart and Myntra.

For negative:

In [37]:
```
#As there are less number of negative feedbacks, we will calculate the people count percentage first
for i in range(0,35,5):
    negative_df['Percentage']=(negative_df['People_Count']*100)/(negative_df.iloc[i:i+5,2].sum())

negative_df
```

Out[37]:

	Particulars	Website	People_Count	Percentage
0	Longer time to get logged in (promotion, sales...	Amazon.in	135	42.3197
1	Longer time to get logged in (promotion, sales...	Flipkart.com	103	32.2884
2	Longer time to get logged in (promotion, sales...	Paytm.com	77	24.1379
3	Longer time to get logged in (promotion, sales...	Myntra.com	35	10.9718
4	Longer time to get logged in (promotion, sales...	Snapdeal.com	67	21.0031
5	Longer time in displaying graphics and photos ...	Amazon.in	126	39.4984
6	Longer time in displaying graphics and photos ...	Flipkart.com	94	29.4671
7	Longer time in displaying graphics and photos ...	Paytm.com	28	8.77743
8	Longer time in displaying graphics and photos ...	Myntra.com	74	23.1975
9	Longer time in displaying graphics and photos ...	Snapdeal.com	92	28.8401
10	Late declaration of price (promotion, sales pe...	Amazon.in	56	17.5549
11	Late declaration of price (promotion, sales pe...	Flipkart.com	43	13.4796
12	Late declaration of price (promotion, sales pe...	Paytm.com	72	22.5705
13	Late declaration of price (promotion, sales pe...	Myntra.com	75	23.511
14	Late declaration of price (promotion, sales pe...	Snapdeal.com	0	0
15	Longer page loading time (promotion, sales per...	Amazon.in	68	21.3166
16	Longer page loading time (promotion, sales per...	Flipkart.com	61	19.1223

In [38]:
```
#We will now calculate the percentage of feedbacks in terms of websites
negative_df['% user feedbacks']=0
negative_df.sort_values(by='Website',inplace=True)
negative_df.reset_index()
```

Out[38]:

	index	Particulars	Website	People_Count	Percentage	% user feedbacks
0	0	Longer time to get logged in (promotion, sales...	Amazon.in	135	42.3197	0
1	20	Limited mode of payment on most products (prom...	Amazon.in	104	32.6019	0
2	25	Longer delivery period	Amazon.in	37	11.5987	0
3	15	Longer page loading time (promotion, sales per...	Amazon.in	68	21.3166	0
4	5	Longer time in displaying graphics and photos ...	Amazon.in	126	39.4984	0
5	30	Frequent disruption when moving from one page ...	Amazon.in	78	24.4514	0
6	10	Late declaration of price (promotion, sales pe...	Amazon.in	56	17.5549	0
7	16	Longer page loading time (promotion, sales per...	Flipkart.com	61	19.1223	0
8	26	Longer delivery period	Flipkart.com	44	13.7931	0
9	21	Limited mode of payment on most products (prom...	Flipkart.com	60	18.8088	0
10	11	Late declaration of price (promotion, sales pe...	Flipkart.com	43	13.4796	0
11	6	Longer time in displaying graphics and photos ...	Flipkart.com	94	29.4671	0
12	31	Frequent disruption when moving from one page ...	Flipkart.com	62	19.4357	0
13	1	Longer time to get logged in (promotion, sales...	Flipkart.com	103	32.2884	0
14	8	Longer time in displaying graphics and photos ...	Myntra.com	74	23.1975	0
15	13	Late declaration of price (promotion, sales pe...	Myntra.com	75	23.511	0
16	3	Longer time to get logged in (promotion, sales	Myntra.com	35	10.9718	0

In [39]:
```
#Index range values of all website with Maximum users value
#negative_df.iloc[0:7]   #Amazon data range and maximum users are 269
#negative_df.iloc[7:14]  #Flipkart data range and maximum users are 221
#negative_df.iloc[14:21]  #Myntra data range and maximum users are 147
#negative_df.iloc[21:28]  #Paytm data range and maximum users are 150
#negative_df.iloc[28:35]  #Snapdeal data range and maximum users are 182
```

In [40]:
```
#Calculating the percentage for all data
negative_df['% user feedbacks'].iloc[0:7]=(negative_df['People_Count']*100)/269 # Amazon users
negative_df['% user feedbacks'].iloc[7:14]=(negative_df['People_Count']*100)/221 # Flipkart users
negative_df['% user feedbacks'].iloc[14:21]=(negative_df['People_Count']*100)/147 # Myntra users
negative_df['% user feedbacks'].iloc[21:28]=(negative_df['People_Count']*100)/150 # Paytm users
negative_df['% user feedbacks'].iloc[28:35]=(negative_df['People_Count']*100)/182 # Snapdeal users
```

In [41]:
```
negative_df.sort_values(by=['% user feedbacks','Website'],ascending=False)
```

Observations:

1. Around 65% of Paytm customers are not happy with their delivery period and longer term in loading pages.

2. Approx. 60% of Snapdeal customers are not happy about their limited mode of payment and nearly 50% of people are not satisfied in longer time of displaying graphics.

3. We can observe that even though with count wise, Amazon and Flipkart showed more negative reviews. When we take percentage, in top 10, Amazon has appeared only once and flipkart has not even appeared even one time.

4. The highest percentage Myntra got is 51, whereas flipkart's highest percentage is 46. However, other websites like Paytm, snapdeal.com have got highest percentage for negative reviews around 60-67%.

5. In terms of less dissatisfaction, myntra.com and flipkart are better, followed by amazon.

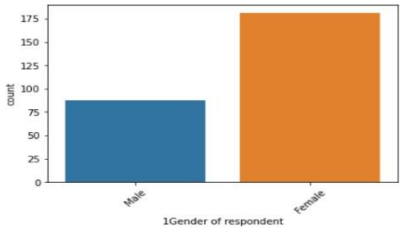
Visualizations

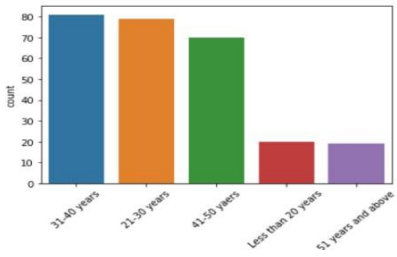
Now, we will see the different plots done with this dataset in order to know the insight of the data present. Below are the codes given for the plots and the output obtained:

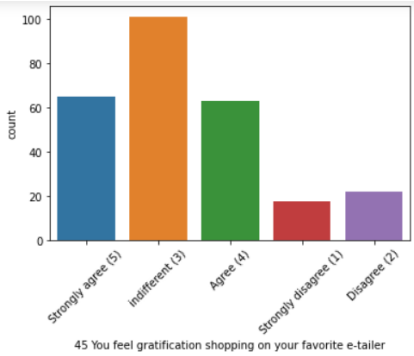
```
In [9]: #Plotting countplot for all the columns present
        for i in df.columns:
            plt.figure(i)
            print(sns.countplot(df[i]))
            plt.xticks(rotation=45)
            print("\n")
```

```
AxesSubplot(0.125,0.125;0.775x0.755)


AxesSubplot(0.125,0.125;0.775x0.755)


AxesSubplot(0.125,0.125;0.775x0.755)


AxesSubplot(0.125,0.125;0.775x0.755)


AxesSubplot(0.125,0.125;0.775x0.755)
```

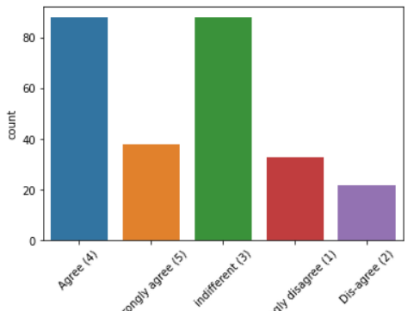
Below are some of the outputs obtained after running the above code:

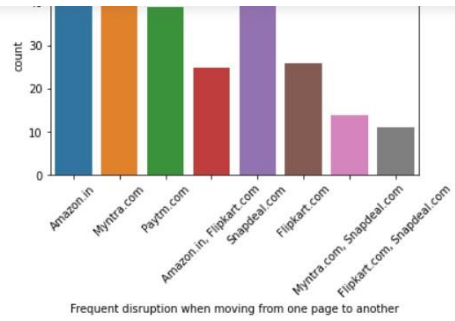




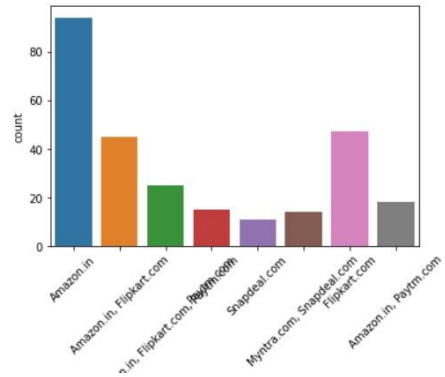


45 You feel gratification shopping on your favorite e-tailer





Frequent disruption when moving from one page to another



Observations from the count plot:

- There are more women respondents than men. It could be that data collection is mainly focused on women.

- Amongst the respondents, the major class targeted is between 21-40 years, followed by 41-50 and less than 20 years. We can

understand that the correspondents are mostly from working class.

- The respondents are majorly residing in cities like Delhi, Greater Noida, Noida and Bangalore.

- Majority are shopping online for more than 4 years. There are considerable people who are shopping online since less than one year and also it shows that many new customers are being added every year.

- We can observe that many have shopped less than 10 times in the past year.

- Many of them use mobile to shop online, followed by laptop, desktop and tablet.

- Windows constitute the major OS of the customer device, followed by Android and Mac.

- Google Chrome is majorly used to access the shopping website.

- People are becoming customers of their favourite stores by using the search engine. Content marketing or display advertisements are not that impactful when coming to online marketing. So, companies should spend more on advertising on search engines.

- For repeated visits, people use search engine first, followed by app and direct URL. We can see that difference between app and search engine is small.

- Majority of the people spend more than 15 minutes before making a purchase, followed by 6-10 minutes.

- The major payment method used by all is credit/debit cards, followed by COD and e-wallets.

- People have mentioned that sometimes they would leave the cart without purchasing and the major reason they have mentioned is that they are finding some better alternative offer. It means that people are comparing from many online websites before making any purchase.

- Customers strongly agree that content of website must be easy to read and understandable.

- Majority of customers want information of similar products to make purchase.

- Majority of the customers want complete information on listed sellers and their products being offered.

- Customers want all relevant information on the listed products and very less customers disagree to that.

- The customers wanted the websites to be easily navigated.

- Majority of the customers wanted high loading and processing speed, user friendly interface of website, convenient payment method, high trust on website, empathy towards customers, guarantee privacy of customers, responsiveness-availability of several communication channels, etc.

- People feel that online shopping provides monitory benefits and discounts.

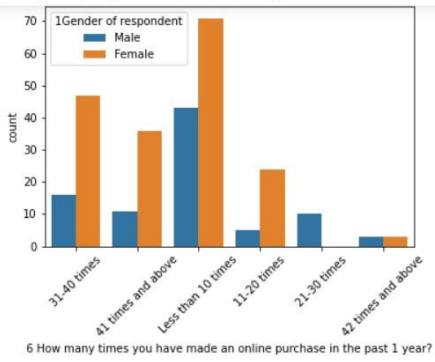
- Customers also feel that shopping online is convenient and flexible.

- Return policy is important for deciding the product purchase to many customers.
- Many customers find shopping through online helps them financially because of cost and discount factors.

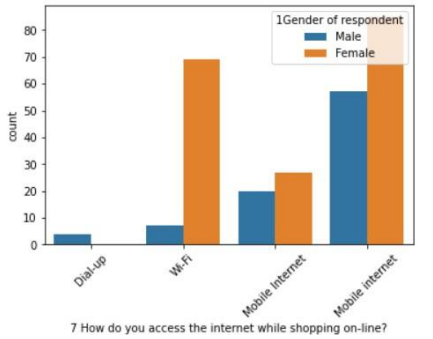
- When it comes to certain factors like gratification, social status enhancement because of shopping, or whether shopping online gives a thrill or adventure, customers are more indifferent to these. So, there is an ample scope in giving more enhanced experienced to customers in this regard.

Count plot for gender:

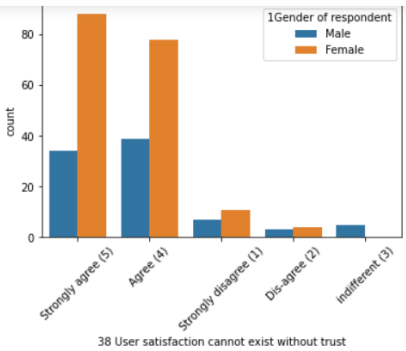
```
In [10]: #Plotting countplot according to gender
         for i in df.columns:
             plt.figure(i)
             sns.countplot(df[i],hue=df['1Gender of respondent'])
             plt.xticks(rotation=45)
             print("\n")
```



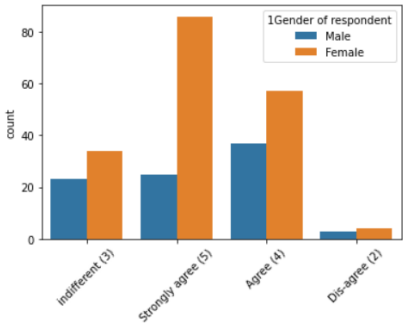
6 How many times you have made an online purchase in the past 1 year?



7 How do you access the internet while shopping on-line?



38 User satisfaction cannot exist without trust



Observations:

- Above 41-50 years and less than 20 years, female and male respondents count difference is not much.
- From Bangalore and Greater Noida, many respondents are female.
- From Noida and Delhi, many respondents are male.
- Both men and women shopping from desktop count are almost same. However, more women shop from either smartphone or laptop.
- Most of the women come back to shopping website by using search engine.
- Many women prefer to use search engine or app, rather than direct URL. However, men prefer to use search engine and URL

and app little less. So, we can understand that women use app more than men.

- Women spend more time than men during online shopping and the time is mostly more than 15 mins, followed by 11-15 mins.

- Women too compare the products with other websites and is one of the reasons to leave the cart without shopping.

- Women prefer more loyal points than men.

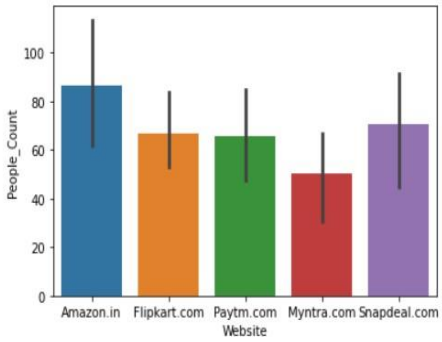
- More women disagree that online shopping is a kind of adventure. So, websites need to work towards giving real time experience as this can be a big marketing strategy.

- Women don't feel that online shopping fulfils certain roles.

- Rest of all other observations are similar as observed in the before count plots.

Website vs People count:

```
In [18]: #Plotting barplot for people_count vs website relationship
         sns.barplot(x='Website',y='People_Count',data=negative_df)

Out[18]: <matplotlib.axes._subplots.AxesSubplot at 0x1e03ad19b80>
```



-> During the promotion time, Amazon has received more negative feedback from customers followed by Snapdeal, Flipkart and Paytm.

-> We can note that difference between negative feedbacks of the websites is not very huge and it needs to be improved in order to handle such situations.

#Taking a list of numbers

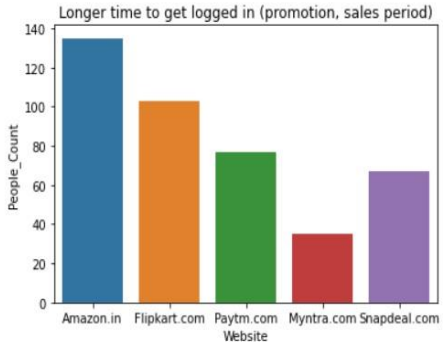
import numpy as np

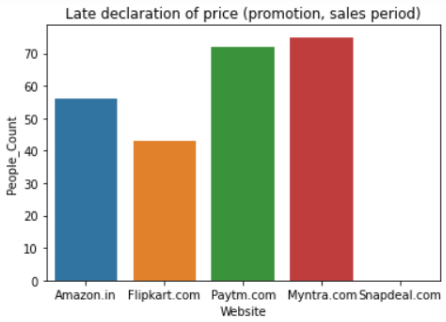
b=np.array(range(17))

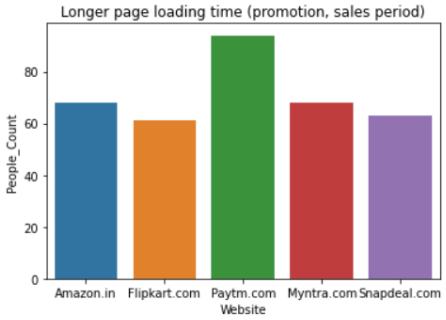
b=b*5

b=list(b)

In [20]:
```
#Extracting feature wise comparison from negative feedback dataframe
try:
    for i in b:
        a=negative_df.iloc[i:i+5,:]
        sns.barplot(x='Website',y='People_Count',data=a)
        plt.title(a['Particulars'][i])
        plt.show()
        print("\n")
except ValueError:   #As the value has some error while iterating, we are passing it by using except
    pass
```





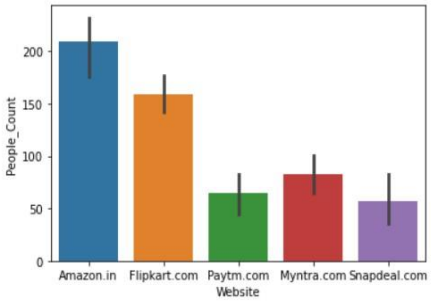


Observations:

1. Amazon takes longer time to get logged in during promotion, followed by flipkart, Paytm and Snapdeal.
2. Amazon takes longer time in displaying graphics and photos followed by flipkart and snapdeal.com.
3. Myntra and Paytm makes late declaration of price during promotion.
4. Paytm takes longer time to load the page during promotion.
5. Snapdeal and Amazon have limited mode of payment on most of products during promotion.
6. Paytm and Snapdeal take a longer delivery period, whereas Myntra and Amazon takes lesser delivery period.
7. Amazon, Snapdeal and Myntra have frequent discrepancies, when moving from one page to another.

In [24]:
```
#Plotting barplot for people_count vs website relationship
sns.barplot(x='Website',y='People_Count',data=positive_df)
```

Out[24]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1e035d09310>
```



We can observe that Amazon has received overall good feedback from the customers, followed by Flipkart, and snapdeal has received less number of feedbacks.

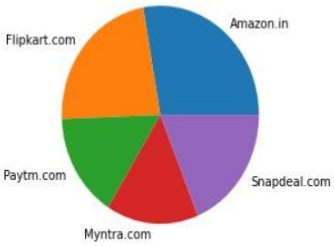
In [25]:
```
b=np.array(range(17))
b=b*5
b=list(b)
b
```

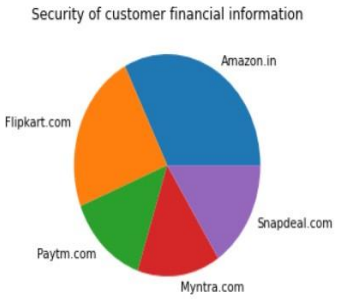
Out[25]:
```
[0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80]
```

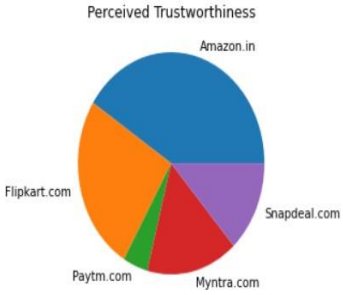
Plotting pie-chart for website vs people count:

In [26]:
```
#Extracting feature wise comparison from positive feedback dataframe and plotting the data using pie chart
for i in b:
    a=positive_df.iloc[i:i+5,:]
    plt.pie('People_Count', labels='Website',data=a)
    plt.title(a['Particulars'][i])
    plt.show()
    print("\n")
```

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

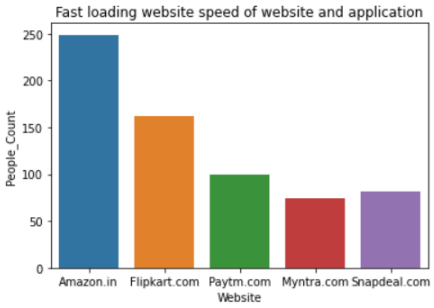


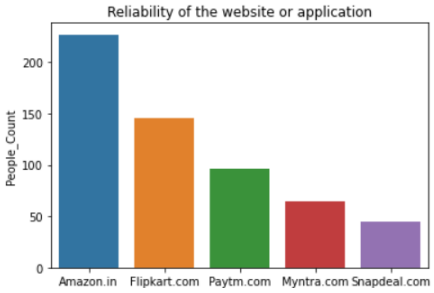




Feature wise comparison from positive feedback data frame and plotting bar plot:

In [27]:
```
#Extracting feature wise comparison from positive feedback dataframe and plotting bar plot
for i in b:
    a=positive_df.iloc[i:i+5,:]
    sns.barplot(x='Website',y='People_Count',data=a)
    plt.title(a['Particulars'][i])
    plt.show()
    print("\n")
```





Observations:

1. Many customers have shopped from Amazon and Flipkart.
2. Amazon and Flipkart have been named most as easy to use website
3. Amazon and Flipkart have been named as the most visually appealing web page layout and also having wild variety of products.
4. Paytm and Snapdeal had not been given more marks on availability of wild variety of products.
5. Amazon and Flipkart have got more positive feedbacks than other websites with relevant to Complete, relevant description information of products, Fast loading of websites, Reliability of website, quickness to complete purchase, availability of several payment options, speedy order delivery, privacy of customers information, security of customer financial information, etc.

6. Paytm has got less feedbacks in perceived trustworthiness, presence of online assistance through multi-channel, speed order delivery.
7. Snapdeal.com has got a smaller number of feedbacks in change of website/application design.
8. Myntra has got the least feedbacks in website as efficient as before, followed by Snapdeal.
9. Only one person has recommended Snapdeal.com overall.

CONCLUSION

Key Findings and Conclusions of the Study

➔**Individual recommendations and feedbacks to the websites**

1. Amazon.com

To be improved:

1. During promotions, try to give a disturbance free shopping experience to customers.
2. Give more payment options to customers.
3. Try to give price early during promotion.
4. Reduce the delivery time of the products.

Positive feedback summary:

1. Convenient to use and also a good website for shopping.
2. Fast delivery of products.

3. Availability of complete information of the products.
4. Presence of online assistance through multi-channels.
5. Reliable website or app, perceived trustworthiness.

2. Flipkart.com

To be improved:

1. During promotions, try to give a disturbance free shopping experience to customers.
2. Give more payment options to customers.
3. Try to give the price early during promotion.
4. Reduce the delivery time of the products.
5. Flipkart and Amazon almost share the same feedbacks with varying percentages as the only difference.

Positive feedback summary:

1. Convenient to use and also a good website for shopping.
2. Fast delivery of products.
3. Availability of complete information of the products.
4. Presence of online assistance through multi-channels.
5. Reliable website or app, perceived trustworthiness.
6. Wild variety of products to offer.

3. Myntra.com

To be improved:

1. During promotions, try to give a disturbance free shopping experience to customers.
2. Try to give the price early during promotions.
3. Reduce the delivery time of the products during promotions.

Positive feedback summary:

1. Convenient to use and also a good website.
2. Availability of several payment options.
3. Faster products delivery.
4. Complete information of products available.
5. Reliable website or app, perceived trustworthiness.
6. Wild variety of product to offer

4. Paytm.com

To be improved:

1. Reduce the delivery time of the products during promotions.
2. Try to give the price early during promotion.
3. During promotions, try to give a disturbance free shopping experience to customers.
4. Late declaration of price and discounts.

5. Frequent disturbance is occurring while moving from one page to another.

Positive feedback summary

1. Convenient to use and a good website.
2. Quickness to complete a purchase.
3. About 64% of the customers feel that either web or app is reliable.
4. Around 20% of the customers believe that Paytm has a wild variety of products on offer.

5. Snapdeal.com

To be improved:

1. Reduce the delivery time of the products during promotions.
2. Try to give the price early during promotion.
3. During promotions, try to give a disturbance free shopping experience to customers.
4. Late declaration of price and discounts.
5. No one has expressed to recommend Snapdeal to a contact as it has the most negative feedbacks among all other websites.

Positive feedback summary:

1. Convenient to use.
2. 54% of the customers are happy about the availability of financial information security.

General suggestions and recommendations to all the e-commerce websites

1. Improve the experience of shopping for customers, as there is a lot of scope in enhancing the shopping experience to the customers using AI.
2. Continue giving more financial benefits like coupons, cashbacks, etc. as customers are very much attracted to it.
3. Trustworthiness and approachability through various channels are still highly rated by customers.
4. Majority of the customers are working class women and their age is between 20-40. Always bring variety of products targeting them.
5. Provide more customer friendly approach like fast delivery, complaint resolution, etc.

Therefore, we had analysed the given dataset by using various data analysis process and also, we had concluded the analysis by observing the positive and negative feedbacks obtained. We recommended some suggestions for the websites to improve further in the future.