E-RETAIL FACTORS FOR CUSTOMER ACTIVATION AND RETENTION

A case study from Indian e-commerce customers

Submitted by:

SWATHI R

ACKNOWLEDGMENT

The internship opportunity I had with Flip Robo was a great chance for learning and professional development. Therefore, I consider myself as a very lucky individual as I was provided with an opportunity to be a part of it.

I would like to thank our SME for suggesting this project and for his whole hearted cooperation and constant encouragement throughout the project. And I also like to thank the data trained mentors and Technical team members for helping me with the technical queries.

And these are the following website which I referred for the reference

1. https://www.kaggle.com/
2. https://scikit-learn.org/
3. www.stackoverflow.com
4. www.google.com
5. www.geeksforgeeks.org

INTRODUCTION

Online shopping took over a significant segment of the retail market during the first decade of the 21 st century. As number of personal computers increased and established, retailers began to offer their products over the Internet.

Some of the various ways in which online marketing is done in India are company websites, shopping portals, online auction sites. Interestingly, India is expected to be the third largest Internet market in the world in the next five years.

A survey by Associated Chambers of Commerce and Industry of India (ASSO CHAM) revealed that shopping trends of Indian consumers have witnessed a significant change as online shopping show more than 55% rise in 2013 as compared to previous year's 85%. Generally, higher levels of education, metropolitan cities and personal income correspond to more favorable perceptions of shopping online.

But despite the high growth of online shopping in recent years, India's e-commerce industry is still in its nascent stage. Online shopping accounts for less than one percent of the total shopping in the country, but has a lot of potential to shoot up.

PROBLEM STATEMENT

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.

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.

Data set

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

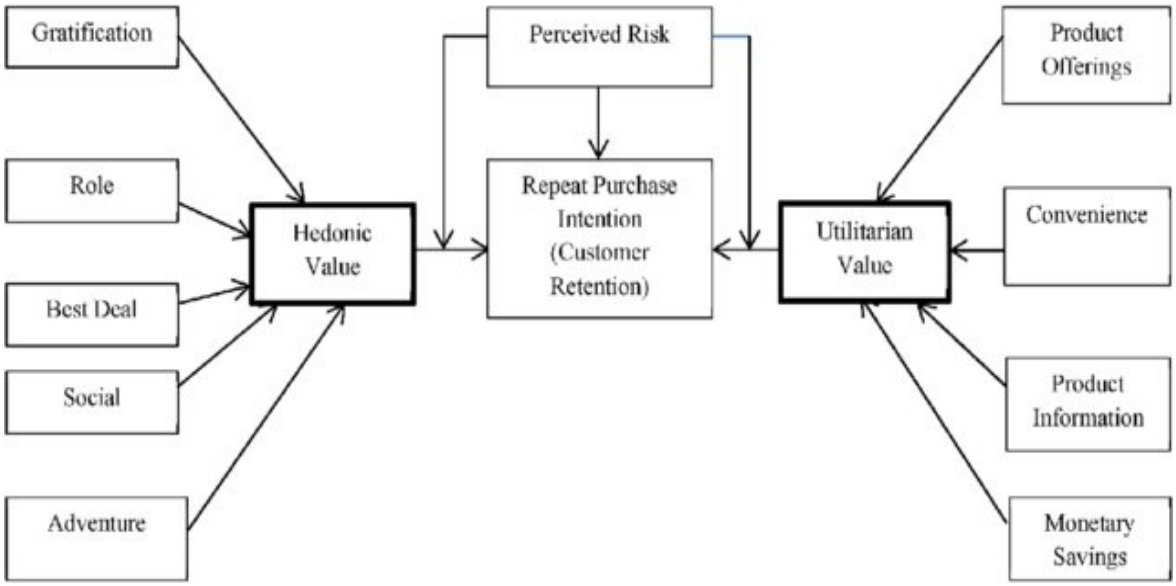
The data is been given by a highly-confidential company and they gave it to us in an excel file.

They also had provided the problem statement by explaining what they need from us and also the required criteria to be satisfied.

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.

DIAGRAMMATIC REPRESENTATION OF CUSTOMER RETENTION



UPLOADING DATA SET

After importing Libraries we uploading the data using the excel file provided in two different Data Frames

- df → Contain the Detail Description of Dataset. Enabling us to analyse the data in encoded form
- dfencoded → Contain the Encoded form of the Dataset. Enabling us to analyse the data in encoded form. Screen shot of the data set is attached

+ Code + Text

Reconnect | Editing

```
df = pd.read_excel('customer_retention_dataset.xlsx','datasheet')
dfencoded = pd.read_excel('customer_retention_dataset.xlsx','codedsheet')
```

```
df.head()
```

	1Gender of respondent	2 How old are you?	3 Which city do you shop online from?	4 What is the Pin Code of where you shop online from?	5 Since How Long You are Shopping Online ?	6 How many times you have made an online purchase in the past 1 year?	7 How do you access the internet while shopping on-line?	8 Which device do you use to access the online shopping?	9 What is the screen size of your mobile device? \t\t\t\t\t\t	10 What is the operating system (OS) of your device? \t\t\t\t	11 What browser do you run on your device to access the website? \t\t\t	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? \t\t\t\t	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 frequently 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	18 The content on the website must be easy to read and understand	Inf on pr high i for co
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	6-10 mins	E-wallets (Paytm, Freecharge etc.)	Sometimes	Lack of trust	Agree (4)	Indi
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	more than 15 mins	Credit/Debit cards	Very frequently	Promo code not applicable	Strongly agree (5)	
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	11-15 mins	E-wallets (Paytm, Freecharge etc.)	Sometimes	Promo code not applicable	Strongly agree (5)	
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	6-10 mins	Credit/Debit cards	Never	Better alternative offer	Agree (4)	

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 Google Colab 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

- Renaming the new column names to the original datasets

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

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

    df.columns = columns
    dfencoded.columns = columns
```

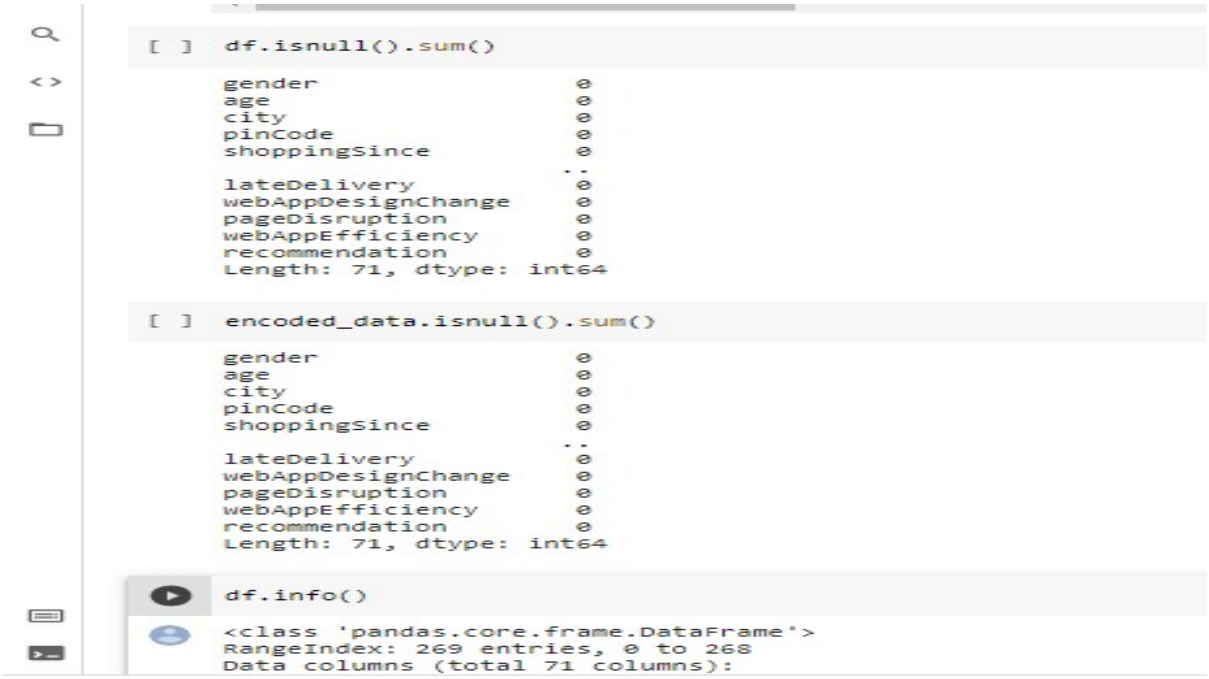
```
[ ] df.head()
```

	gender	age	city	pinCode	shoppingSince	shoppingFrequency	internetAccessibility	deviceUsed	screenSize	OS	browserUsed	channelFirstUsed	loginMode	timeSpentDeciding	paymentMode
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	6-10 mins	E-wallets (Paytm, Freecharge etc.)
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	more than 15 mins	Credit/Debit cards
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	11-15 mins	E-wallets (Paytm, Freecharge etc.)
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	6-10 mins	Credit/Debit cards
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	more than 15 mins	Credit/Debit cards

```
[ ] dfencoded.head()
```

	1Gender of respondent	2 How old are you?	3 Which city do you shop online from?	4 What is the Pin Code of where you shop online from?	5 Since How Long You are Shopping Online ?	6 How many times you have made an online purchase in the past 1 year?	7 How do you access the internet while shopping on-line?	8 Which device do you use to access the online shopping?	9 What is the screen size of your mobile device? \t\t\t\t\t\t	10 What is the operating system (OS) of your device? \t\t\t\t	11 What browser do you run on your device to access the website? \t\t\t	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? \t\t\t\t	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	18 The content on the website must be easy to read and understand	1 Information on similar product to the one highlighted is important for product comparison
0	0	3	Delhi	110009	5	4	4	3	5	1	1	1	1	3	4	3	3	4	
1	1	2	Delhi	110030	5	5	2	1	2	3	1	1	4	5	1	5	5	5	
2	1	2	Greater Noida	201308	4	5	3	1	4	2	1	1	4	4	4	3	5	5	
3	0	2	Karnal	132001	4	1	3	1	4	3	2	1	1	3	1	1	2	4	
4	1	2	Bangalore	530068	3	2	2	1	2	3	2	3	4	5	1	4	2	5	

- Checking for null values in the data set.
- There are no null values present in the dataset.



```
[ ]  df.isnull().sum()

gender                  0
age                     0
city                    0
pinCode                 0
shoppingSince           0
                       ..
lateDelivery            0
webAppDesignChange      0
pageDisruption          0
webAppEfficiency        0
recommendation          0
Length: 71, dtype: int64

[ ]  encoded_data.isnull().sum()

gender                  0
age                     0
city                    0
pinCode                 0
shoppingSince           0
                       ..
lateDelivery            0
webAppDesignChange      0
pageDisruption          0
webAppEfficiency        0
recommendation          0
Length: 71, dtype: int64

     df.info()

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

- 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.

```
DATA VISUALIZATION

[ ]  for i in df.columns:
       print(df[i].value_counts())
       print("--------------------------------")

203207     1
203202     1
Name: pinCode, dtype: int64
--------------------------------
Above 4 years       98
2-3 years           65
3-4 years           47
Less than 1 year    43
1-2 years           16
Name: shoppingSince, dtype: int64
--------------------------------
Less than 10 times    114
31-40 times            63
41 times and above     47
11-20 times            29
21-30 times            10
42 times and above      6
Name: shoppingFrequency, dtype: int64
--------------------------------
Mobile internet    142
Wi-Fi               76
Mobile Internet     47
Dial-up              4
Name: internetAccessibility, dtype: int64
--------------------------------
Smartphone    141
Laptop         86
Desktop        30
Tablet         12
Name: deviceUsed, dtype: int64
--------------------------------
```

Correlation between Variables

```
[ ] df.apply(lambda x : pd.factorize(x)[0]).corr(method='pearson', min_periods=1)
```

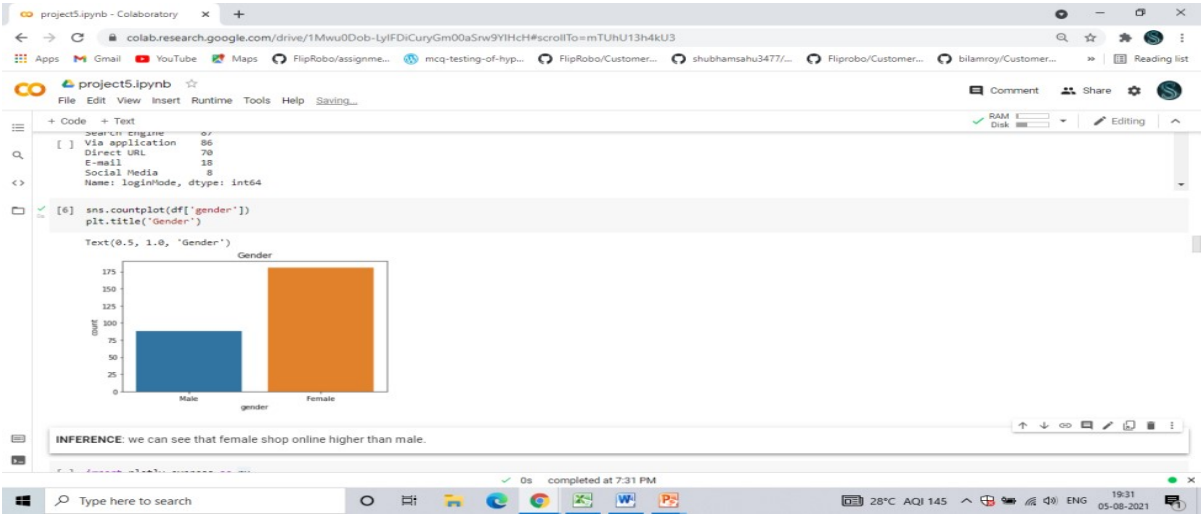
	gender	age	city	pinCode	shoppingSince	shoppingFrequency	internetAccessibility	deviceUsed	screenSize	OS	browserUsed	channelFirstUsed	loginMode	timeSpentDec
gender	1.000000	0.034475	0.196291	0.077270	0.037380	-0.171123	-0.201063	0.063097	-0.015089	-0.019243	-0.167459	-0.185259	-0.060113	0.2
age	0.034475	1.000000	0.163532	-0.019154	0.090323	-0.005596	-0.040077	-0.159394	0.127427	0.108475	-0.123541	0.013896	0.317130	0.1
city	0.196291	0.163532	1.000000	0.347878	0.208610	-0.035784	-0.156361	-0.094786	0.197274	0.155224	0.035707	-0.058391	0.086455	0.2
pinCode	0.077270	-0.019154	0.347878	1.000000	0.437144	0.031963	0.231474	-0.106708	0.156820	0.095841	0.230006	0.076020	-0.002740	0.2
shoppingSince	0.037380	0.090323	0.208610	0.437144	1.000000	0.087416	-0.040878	-0.054688	0.256882	0.187290	0.342688	0.178333	0.007927	0.2
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
lateDelivery	0.070612	0.279299	0.277034	0.159754	0.055680	0.011163	-0.100512	0.357144	-0.127673	-0.195657	-0.199827	-0.110715	0.507256	0.1
webAppDesignChange	0.067428	0.057116	-0.129196	-0.004688	0.127201	0.081318	-0.001550	0.453325	-0.025407	-0.120316	0.015245	0.064448	-0.375580	-0.1
pageDisruption	0.020797	-0.087987	0.237886	-0.087549	0.015835	0.171922	-0.058379	0.297947	0.094254	0.155314	0.082094	-0.137411	-0.073928	0.1
webAppEfficiency	-0.035515	0.048604	0.160467	0.029517	0.225815	0.119941	-0.036510	0.379451	-0.160743	-0.060454	0.024764	0.053000	-0.083362	0.0
recommendation	-0.085315	0.337155	0.314567	0.164432	0.029341	0.017889	0.010465	0.086864	0.145320	0.032682	0.072979	0.138541	0.569372	-0.0

71 rows × 71 columns

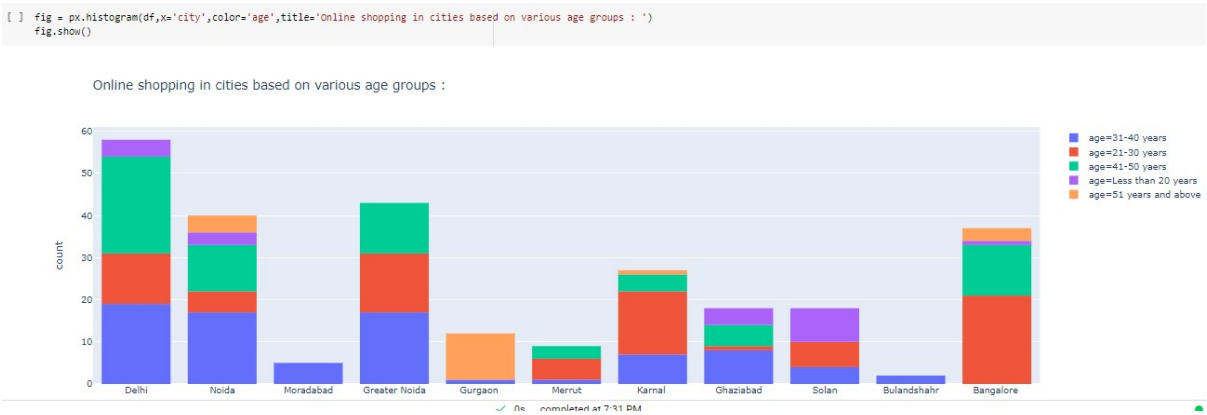
- People from Delhi, Greater Noida and Noida shop frequently when compared to Merrut, Moradabad and Bulandshahr.

```
[ ] import plotly.express as px

city_count = pd.DataFrame(df['city'].value_counts()).reset_index()
city_count.columns = ['city','count']
fig = px.bar(city_count, x='city',y='count',
             color='city',
             title = 'Online Shopping is preferred in which cities?')
fig.show()
```

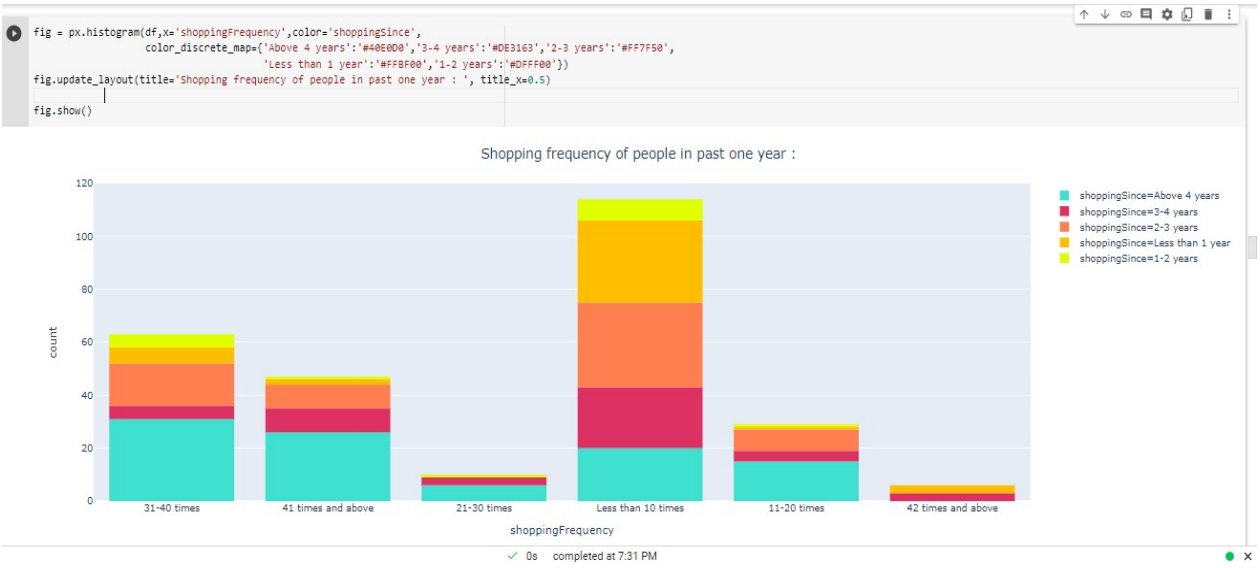
- We can see that female shop online higher than male.



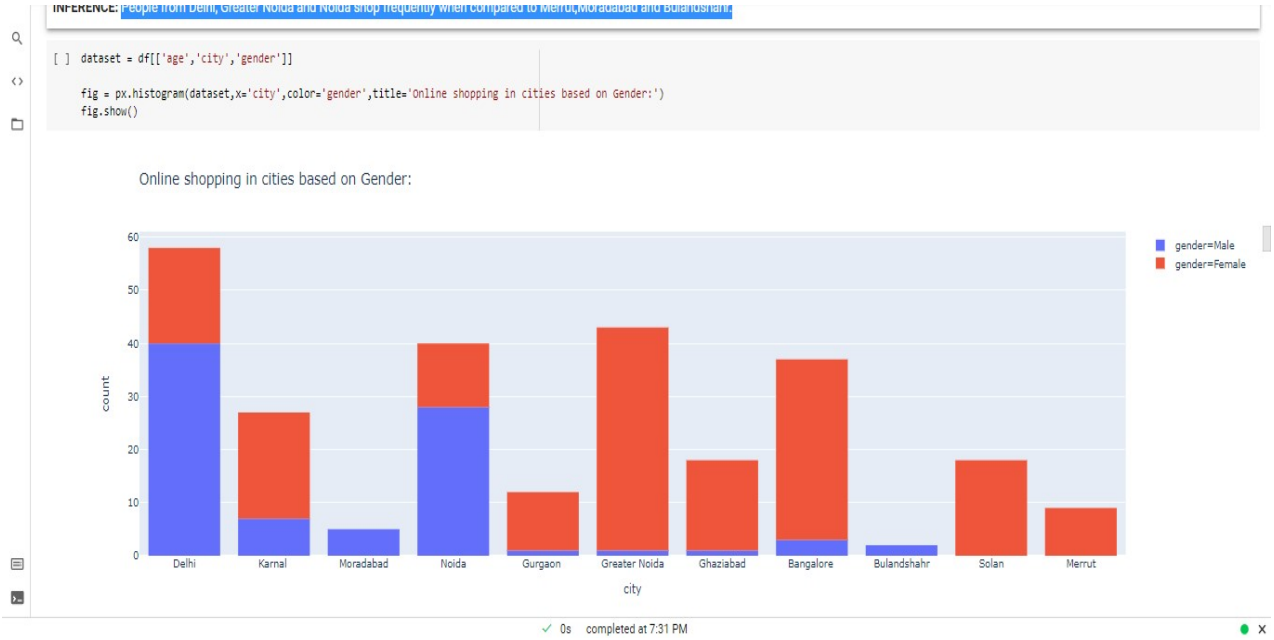
- In city like Bangalore and Karnal Young people(21-30) shop higher.
- In city like Delhi, Noida, Greater Noida Mid age(31-40) prople shop higher.
- And In city like Solan we can see that kids and teen age(under 20) people shop higher.



- Shopping frequency of people in past one year

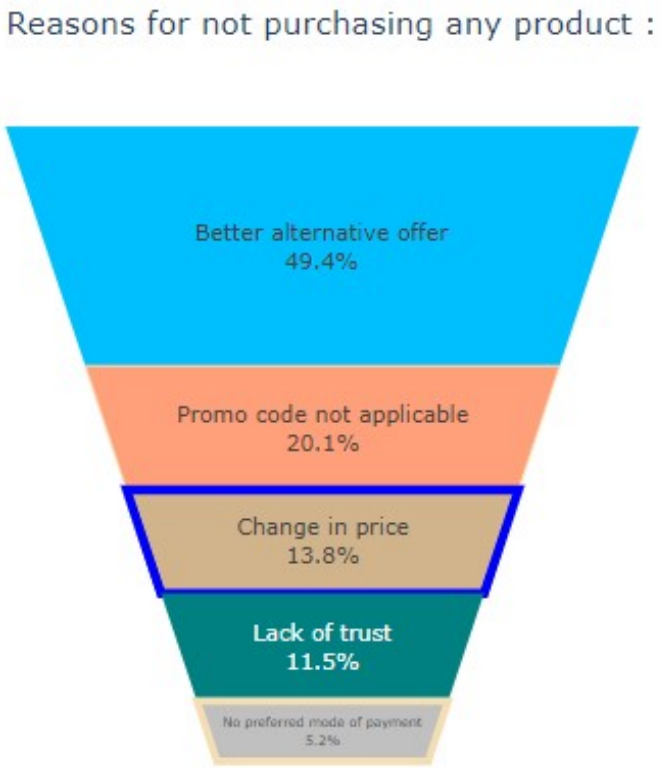


- Only In Delhi and Noida Male shoppers are higher when compared to other city.



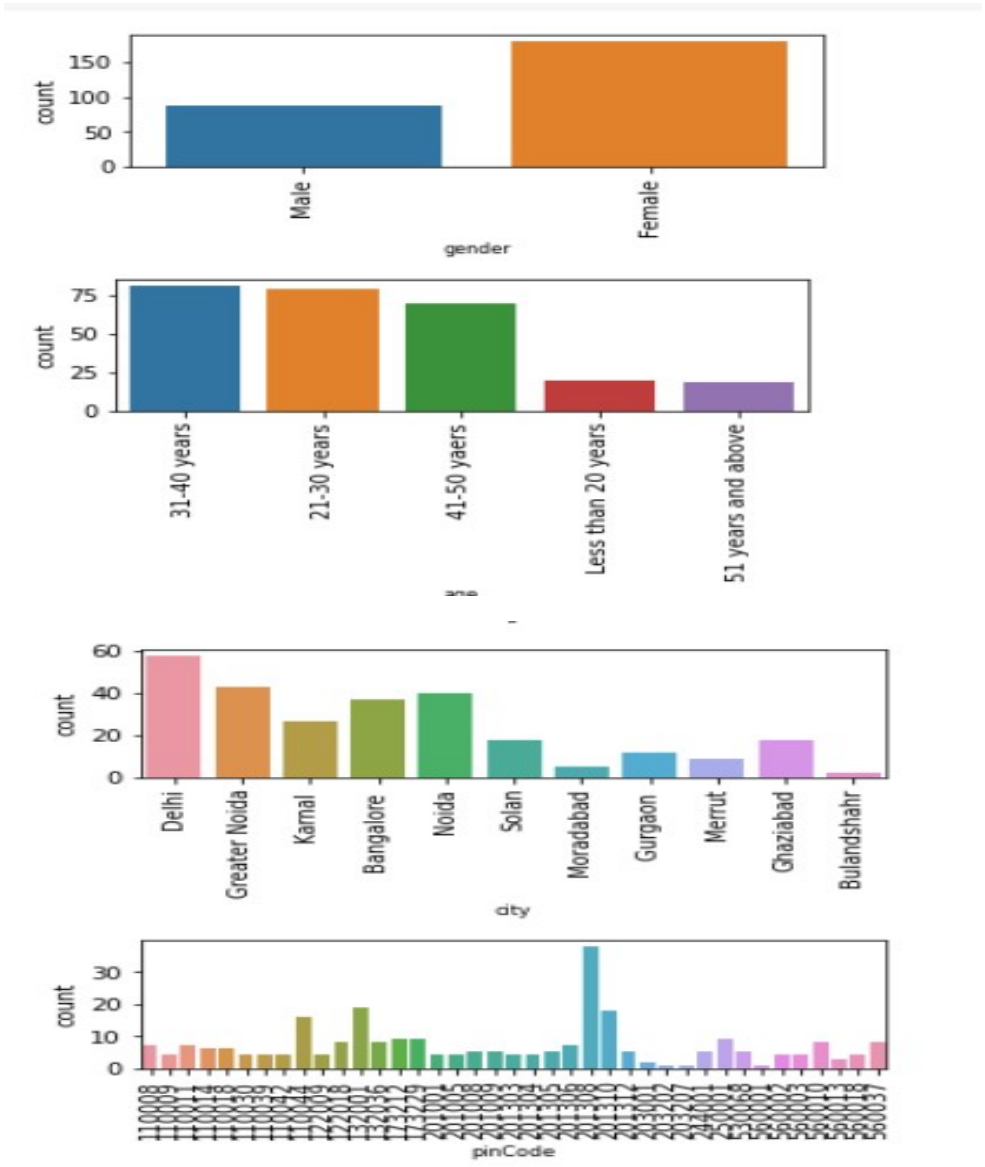
- Many people’s Reasons for not purchasing Better alternative options.

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

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

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



- We checked the Visualization of all 71 columns above and we iterated using a for loop. We can see some Visualization of the columns like gender, age, city, etc.

```
[ ]  plotno=1
     for column in df:
         if plotno<=70:
             plt.figure(figsize=(30,25))
             ax=plt.subplot(14,5,plotno)
             plt.xticks(rotation = 90)
             sns.countplot(df[column])
             plt.xlabel(column,fontsize=8)
         plotno+=1
     plt.show()
```



count
100
50
0
Above 4 years
3-4 years
2-3 years
Less than 1 year
1-2 years
shoppingSince

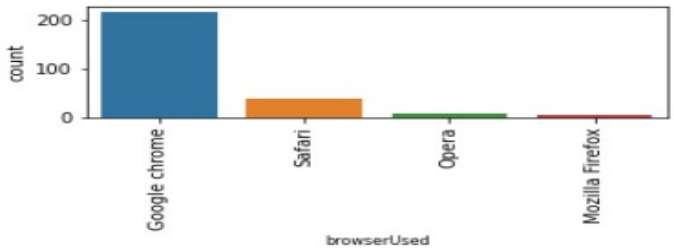
count
100
50
0
31-40 times
41 times and above
Less than 10 times
11-20 times
21-30 times
42 times and above

count
100
50
0
Dial-up
Wi-Fi
Mobile Internet
Mobile internet
internetAccessibility

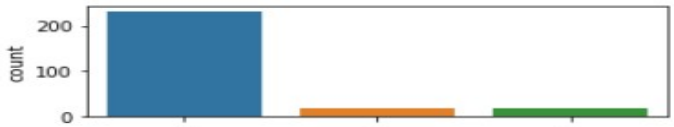
count
100
50
0
Desktop
Smartphone
Tablet
Laptop
deviceUsed

count
100
50
0
Others
4.7 inches
5.5 inches
5 inches
screenSize

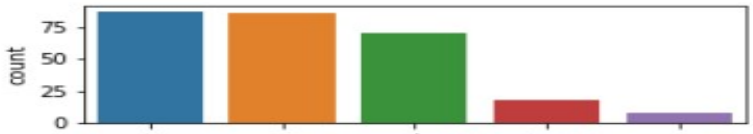
count
100
50
0
dow/windows Mobile
IOS/Mac
Android



browserUsed



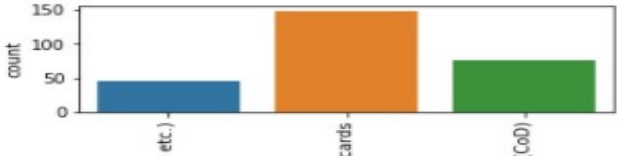
channelFirstUsed



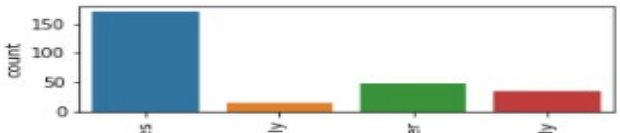
loginMode

n

timeSpentDeciding



paymentMode

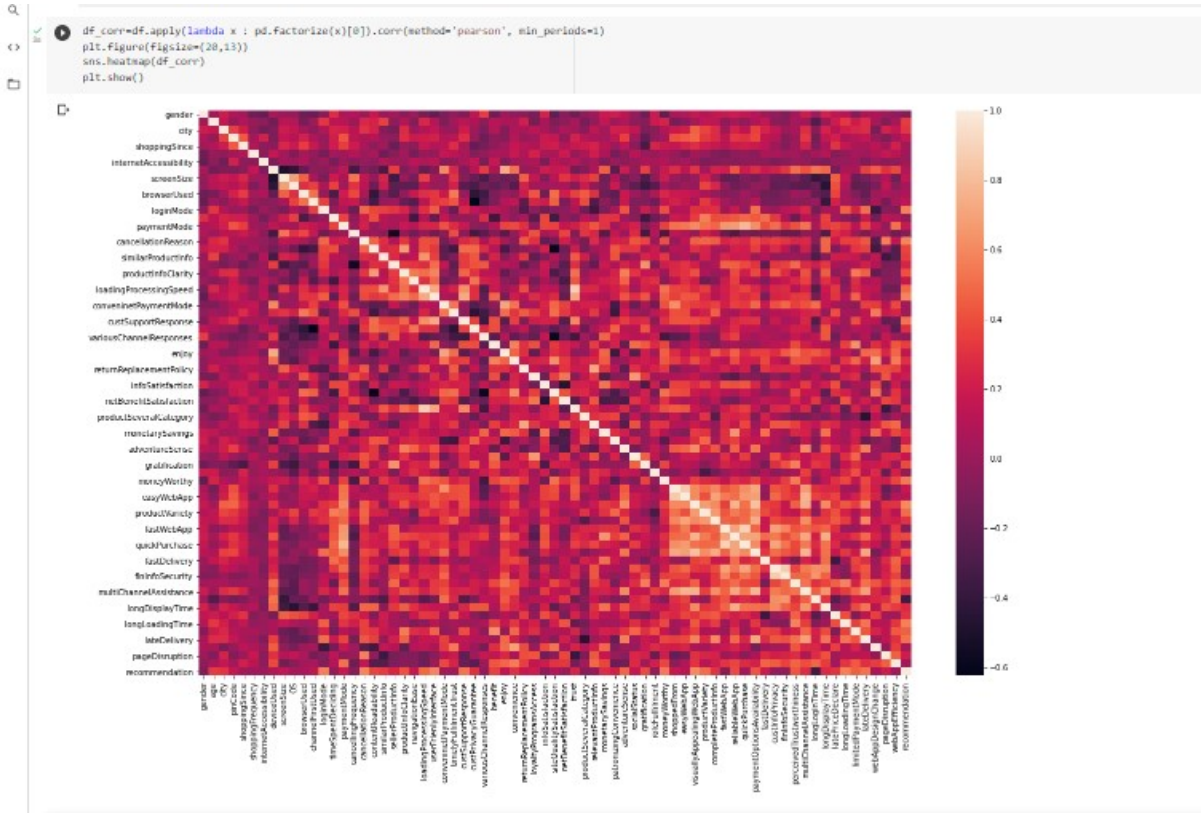


cancellingFrequency

INFERENCE FROM VISUALIZATIONS

- Age: 85% of the people are within age group 20-50 years.
- Very small percentage of people are less than 20 or more than 50 years old.
- Delhi, Greater Noida, Noida and Bangalore have the maximum participants.
- Maximum number of participants have been using Online Shopping platforms for more than 4 years.
- Purchases in last 1 year: Maximum People have made purchases less than 10 times. Next in line are 30-40 and 40+
- Most people use Mobile internet during shopping on-line.
- Very few people use desktops or tablets; while most are using Smartphones and laptops.
- A large percentage of people use Google Chrome.
- Search Engines are the most used channel which guide people to their favourite online store.
- Most people take purchase decision after spending more than 15 minutes.
- Credit/Debit cards are the most used mode of payment.
- Maximum people abondon the items in cart 'sometimes' as opposed to never or frequently.
- The biggest reason to abandon is 'Better alternative offer.

Correlation Heat map

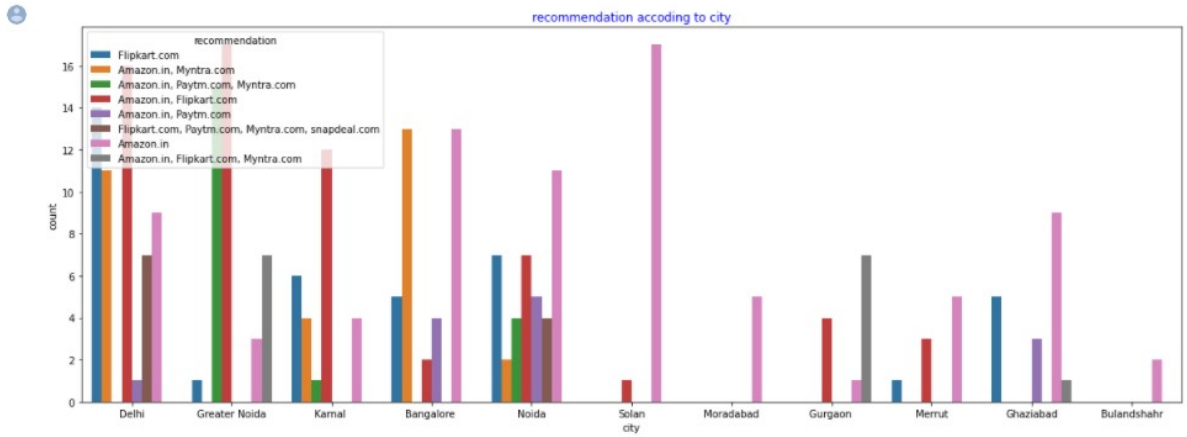


CORRELATION HEAT MAP

In the above Correlation Heat map, we have the following observations:

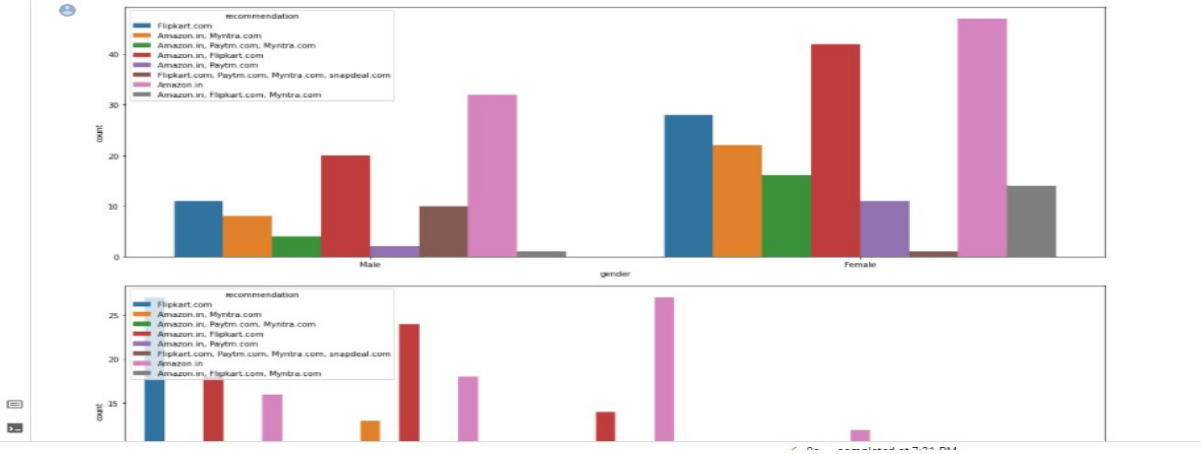
- Most of the high correlation is concentrated between “Ease to Read and Understand Website” and “Privacy of Customers”
- Screen Size and Operating System has the lowest Correlation.
- “Convenient Payment Method” and “User Friendly Interface” have high Correlation.

```
plt.figure(figsize=(20,7))
plt.title('recommendation accoding to city',color='blue')
sns.countplot(x='city',data=df,hue='recommendation')
plt.show()
```



- Recommendation of shopping website according to city

```
for i in df.columns:
  plt.figure(figsize=(20,7))
  sns.countplot(x=df[i],data=df,hue='recommendation')
  plt.show()
```



- We checked the Visualization and recommendation of all 71 columns with respect to suggested website above and we iterated using a for loop. We can see some Visualization of the columns like gender, age, city, etc.

INFERENCE from Visualizations

1.Maximum people have shopped from these 5 companies - Amazon.in Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com.

2. Amazon.com and Flipkart.com are the major choices of the customer.

3. In terms of Visual appealing webpage layout also, Amazon.com and Fli pkart.com seem to take the lead.

4. Talking about Wide variety of product on offer: Amazon.com and Flipk art.com are the major choices.

5. Maximum people have chosen to go with mazon.com and Flipkart.co m based on relevant description information of products.

6. Amazon have high score when it comes to Fast loading website, altho ugh paytm and Flipkart are not far behind.

7. Also Amazon have high score when it comes to Reliability of the websi te, although Flipkart and paytm are not far behind.

8. Amazon take the lead when it comes to Quickness in choosing produc t, Flipkart and paytm falls next to amazon.

9. Both Amazon and Flipkart are best in several payment options. Althou gh a lot of people also like to try Myntra.

10. Amazon seems to take the lead in Speedy order delivery, followed by Flipkart.

11. Amazon has a good reputation for maintaining privacy, followed by F lipkart

12. Amazon and Flipkart both having good Trustworthiness. Although M yntra.com and Snapdeal.com are also good in Trustworthiness.

13. Amazon.in, Flipkart.com, Myntra.com, Snapdeal provied best online assistance through multi-channel.

14. Amazon.in, Flipkart.com takes longer than ussula Longer time in displ aying graphics and photos.

15. Myntra, Paytm, Snapdeal. These companies should work on Late decl aration of price.

16. Myntra and Paytm have bad feedback in Longer page loading time. Fl ipkart should also have a look into it, as it is at the 3rd place.

17. Snapdeal is the most voted answer for mode of payment.

18. Paytm and Snapdeal takes more time to delivery their product.

19. Amazon.in needs to concentrate in Frequent disruption when movin g from one page to another.

20. Which of the Indian online retailer would you recommend to a friend ?: Amazon/Flipkart

OBSERVATIONS

- Many customers have shopped from Amazon and Flipkart.
- Amazon and Flipkart is suggested as most as easy to use website
- Amazon and Flipkart have been suggested as the most visually appealing web page layout and also having wild variety of products.
- Paytm and Snapdeal had not been given more marks on availability of wild variety of products.
- 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.
- Paytm has got less feedbacks in perceived trustworthiness, presence of online assistance through multi-channel, speed order delivery.
- Snapdeal.com has got a smaller number of feedbacks in change of website/application design.
- Myntra has got the least feedbacks in website as efficient as before, followed by Snapdeal.
- Only one person has recommended Snapdeal.com overall.

GENERAL SUGGESTIONS AND RECOMMENDATIONS TO ALL THE E-COMMERCE WEBSITES

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