A Project on E-retail Factors for Customer Activation and Retention in E-commerce Companies

Importance of Customer Retention & Engagement for E-Commerce Giants

ABSTRACT

Using Data Analytics and visualization techniques to find out the factors which could play a vital role in increasing the customer engagement and retention of an e-commerce company like Amazon, Flipkart, Myntra and etcetera.

Presented By-

ASHUTOSH MISHRA

(DATA SCIENCE INTERN)

FLIP ROBO TECHNOLOGIES

ABSTRACT

Here the **Problem Statement** is to find out those factors which could play a vital role in increasing the customer engagement and retention of an e-commerce company like Amazon, Flipkart, Myntra and etcetera. As we all know that Customer engagement and customer retention are the two most important aspects in any type of businesses whether it's B-2-B, B-2-C or C-2-C and when it comes to e-commerce platforms it's somehow get more challenging for the customers to maintain their loyalty respect to these virtually presence businesses as the competition is really tough due to the n-numbers of e-retail presence. Also, If we as a customer compare buying something in a physical store with that of e-store then definitely the customers satisfaction are going to be on higher side w.r.t. physical store as we can touch, taste or opt for a trial depending on the products but the same is not true with online companies as customer gets delivery after a specific time. Hence these companies wants to go ahead w.r.t it's competition and in-order to do so they're giving a lots of inputs for feedbacks while customers are doing shopping so that they can increase the customer trust, loyalty, satisfaction and this will lead to customer retention.

I've got a dataset having 269 rows and 71 attributes and will do some Exploratory Data Analysis and Data Visualizations on the same and will try to find out the attributes that are really important for these e-commerce company in order to increase its sales, customer satisfaction and retention.

EXPLORATORY DATA ANALYSIS

EDA is a critical first step in analyzing the data from an experiment. Here are the main reasons we use EDA.

- Detection of mistakes or finding null values, repetition etcetera
- Determining Relationships among the explanatory variables
- Assessing the direction and rough size of relationships between explanatory and outcome variables.
- Making Data Clean and Normalize for the further model-building purpose.

Data-Frame Description

After installing all the important libraries for data manipulation, data visualization and statistics I've successfully imported the dataset with the help of pandas Excel and CSV data-frame. There are total 269 rows and 71 columns present and since the attributes are really higher hence it's going to be a complex situation to find out the best attributes for the decision making towards customer satisfaction, customer retention and customer engagement. Our data frame has only 1 integer column and rest 70 features are in categorical object format having two or more than two categories which indicates that we have to use encoder technique to convert it all in Numeric format while going further, so that manipulation and data cleansing can be perform easily.

We can refer below attached screenshot to check out our data-frame's dimension, shape, size, index and layout.

```
print(df.shape)
print(df.shape[0])
print(df.shape[1])
```

```
(269, 71)
269
71
```

This shows that this datframe contains 269 rows and 71 columns

```
#it's a 2 dimensional dataframe
df.ndim
```

5]: 2

```
df.size
```

6]: 19099

There are total Nineteen thousand ninty hundred data's are avialble under all columns alltogether

In [8]:
```
#indicates indexation which is from 0-269.
df.index
```

Out[8]: RangeIndex(start=0, stop=269, step=1)

In [9]:
```
#it shows top 5 rows of the dataframe
df.head()
```

Out[9]:

	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	...	Longer time to get logged in (promotion, sales period)	Longer time in displaying graphics and photos (promotion, sales period)	La declarati of pri (promotic sales perio
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.co
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.co
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.co
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.co
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.co

5 rows × 71 columns

Null Value Presence

Also the dataset is not having any null value presence as we can see in the below attached figure.

```
In [13]: #This will check-out the null value presence in each columns
         df.isnull().sum()

Out[13]: 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
```

As we can see in each columns there are no null value present,i.e. no rows are having missing values

```
In [14]: df.isnull().sum().sum()

Out[14]: 0
```

The sum of all the missing value in each columns alltogether is zero

Factors Uniqueness

There are almost more than two categories present for each attributes. Attribute four (column no. 4) of our data set contains almost 39 category of pin code number and going forward in this thesis I don't think it as an important attribute so we can drop this from the data-frame. Also, we can also refer to the below image for the same.

Checking Uniqueness of each Factor

```
In [55]: df.nunique()

Out[55]: 1Gender of respondent                                          2
         2 How old are you?                                            5
         3 Which city do you shop online from?                        11
         4 What is the Pin Code of where you shop online from?        39
         5 Since How Long You are Shopping Online ?                    5
                                                                      ..
         Longer delivery period                                        6
         Change in website/Application design                          7
         Frequent disruption when moving from one page to another      8
         Website is as efficient as before                             8
         Which of the Indian online retailer would you recommend to a friend?    8
         Length: 71, dtype: int64
```

As we can see in above output there are almost more than two categories present in each attributes which shows that e-commerce companies are giving a lots of inputs to their customers to get geniune feedbacks from the customers so that they can prepare the best out of it.

While doing EDA I've found that most of the attributes are having unique values more than two, but in some cases categories meanings are same but it's written in different way so I've rename it and made it as a single category itself as any mistake of information can lead towards false decision-making by the companies and it will affect their customer retention and engagement decision. As you can see in in the below image Mobile Internet occurs twice which have same meaning but its showing as a different category, hence I've made it a single category by replacing it as Mobile Internet.

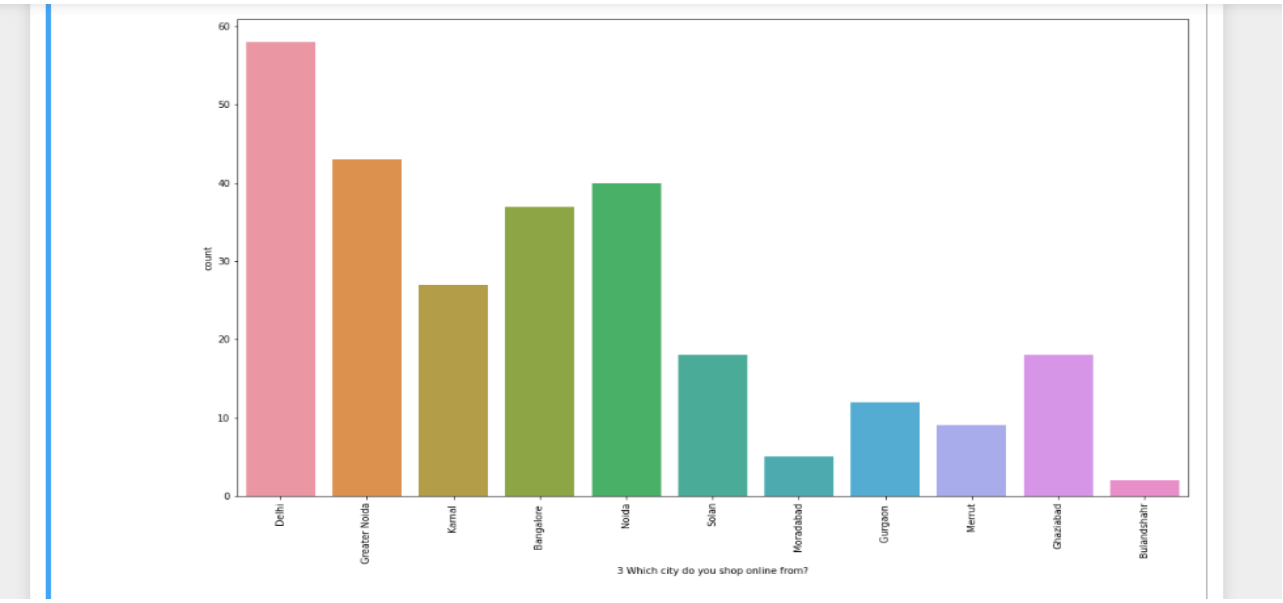
```
In [61]: df['7 How do you access the internet while shopping on-line?'].value_counts()

Out[61]: Mobile internet    142
         Wi-Fi               76
         Mobile Internet     47
         Dial-up              4
         Name: 7 How do you access the internet while shopping on-line?, dtype: int64
```

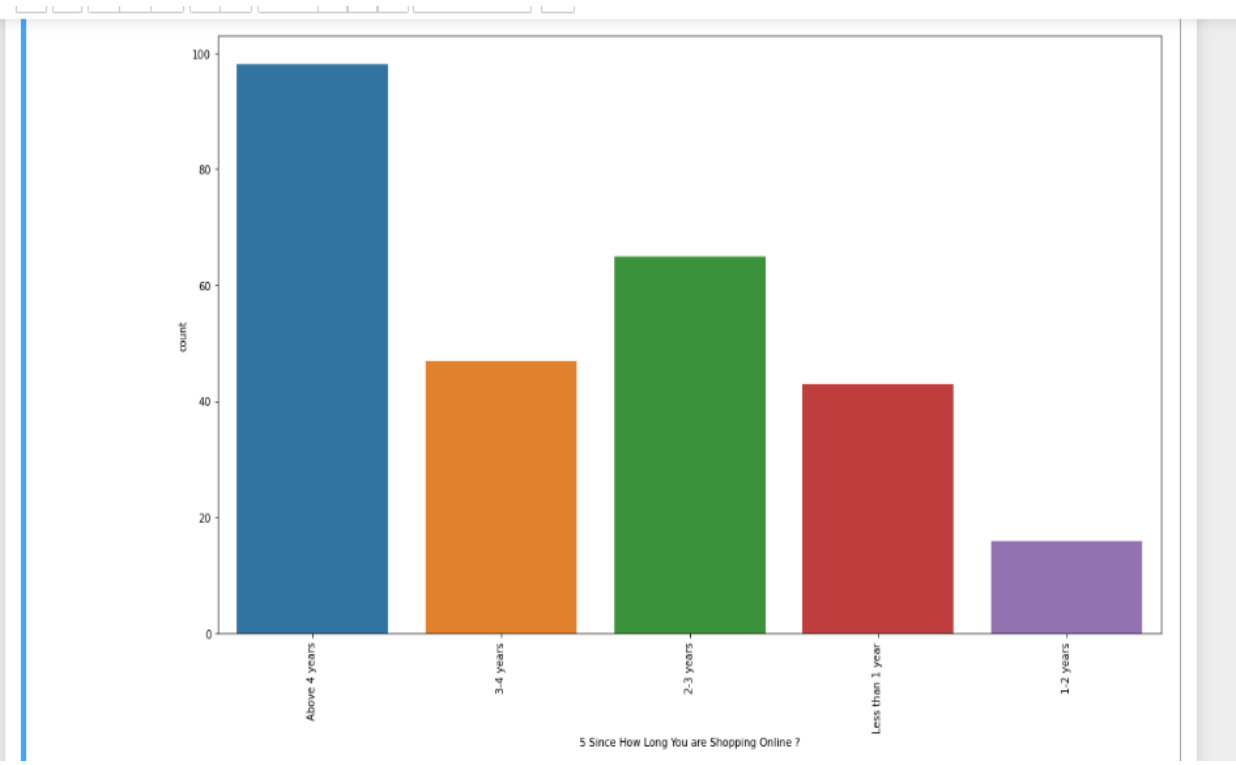
As we can see that There are four type of categorisation is given but out of them two are same in meaning,i.e. Mobile internet and Mobile Internet both are same in meaning hence will try to replace and convert it into single category

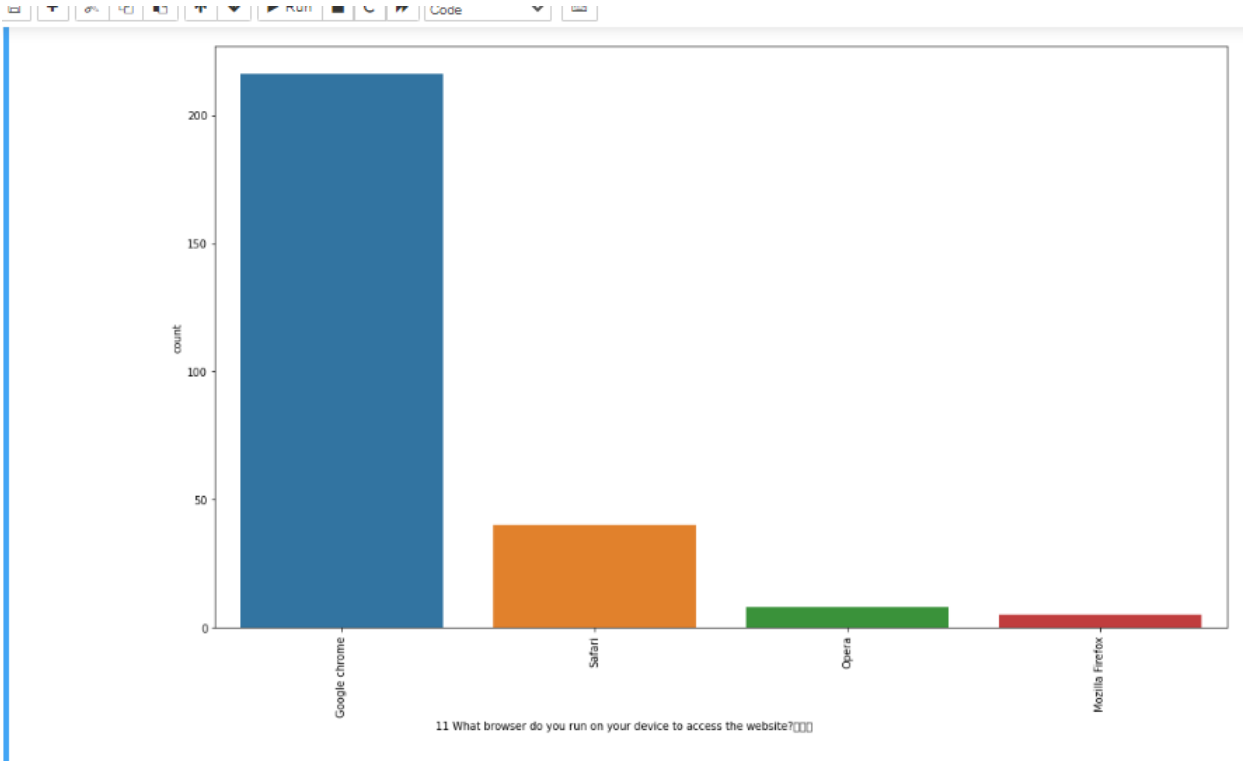
```
In [63]: #Replacing 'Mobile internet' with 'Mobile Internet' and making it single category
         df['7 How do you access the internet while shopping on-line?']=df['7 How do you access the internet while shopping on-line?']

In [64]: #Now will get the required category
         df['7 How do you access the internet while shopping on-line?'].value_counts()

Out[64]: Mobile Internet    189
         Wi-Fi               76
         Dial-up              4
         Name: 7 How do you access the internet while shopping on-line?, dtype: int64
```

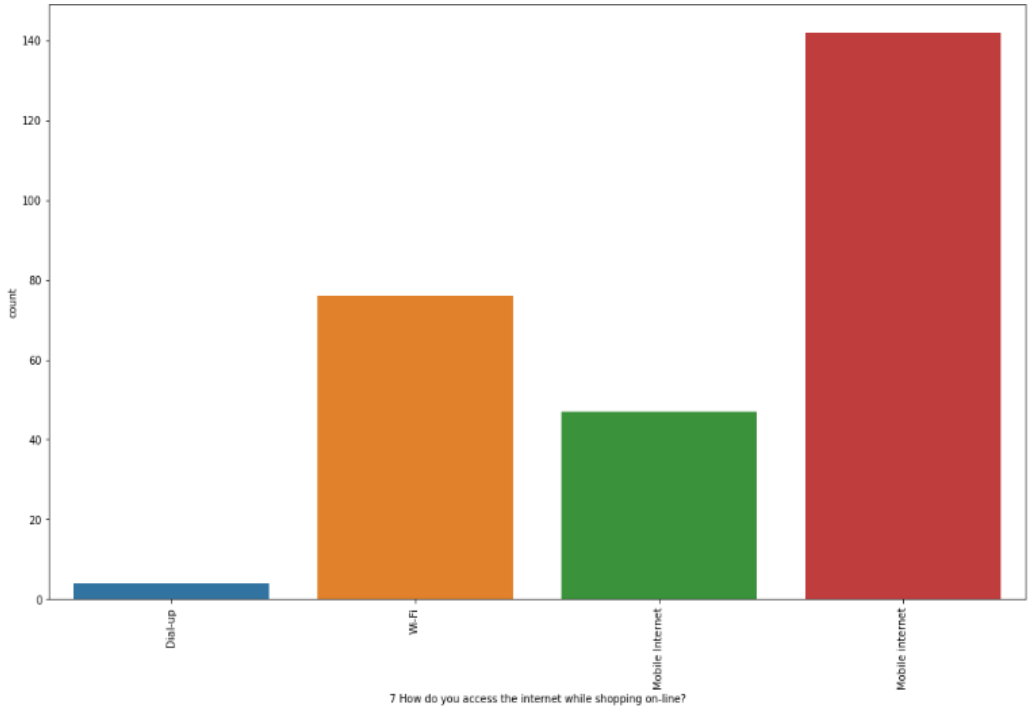
Data Visualizations

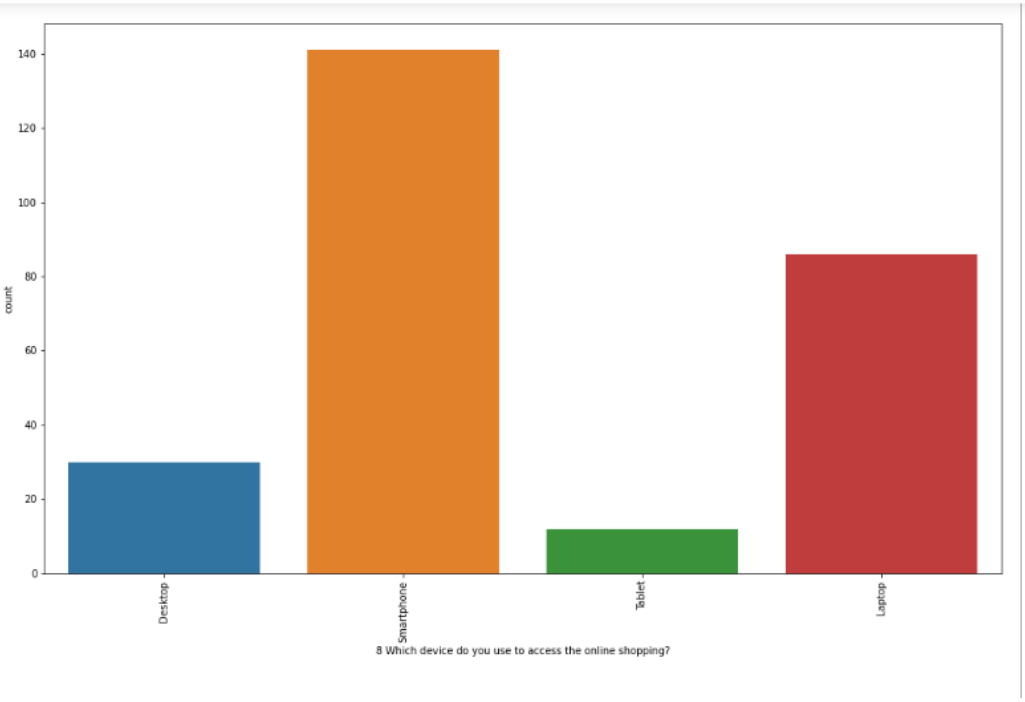
1. COUNTPLOT




100
80
60
40
20
0
count
Above 4 years
3-4 years
2-3 years
Less than 1 year
1-2 years
5 Since How Long You are Shopping Online ?

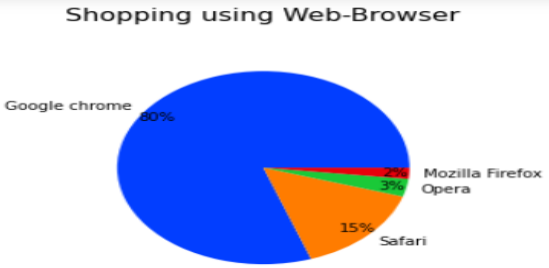

Code
200
150
100
50
0
count
Google chrome
Safari
Opera
Mozilla Firefox
11 What browser do you run on your device to access the website?





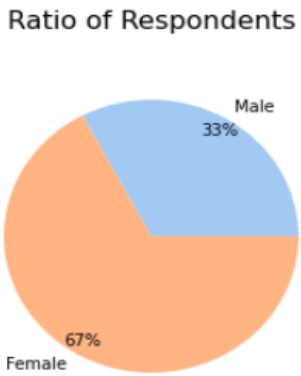
As, by seeing all the above countplot diagrams we can say that, Most of the respondents are females of Delhi having age more than 4 Years and they are using mobile internet through smartphone and accessing Google Chrome Browser for shopping their favorite products and when it comes to referring the best e-commerce platform for shopping to any friends or family member then the name they are suggesting is Amazon.

2. *PIE-CHART*

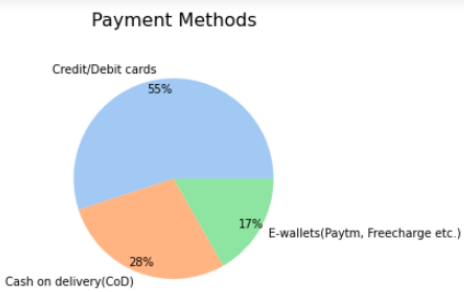


80% of the customers prefer using Google chrome while making online shoppings

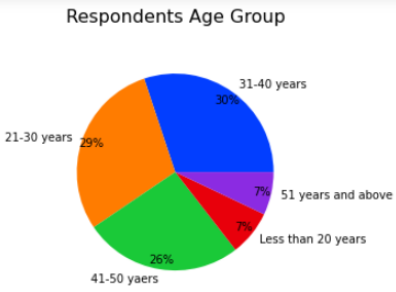
Delhi has highest online shoppings



Females are shooping more than males

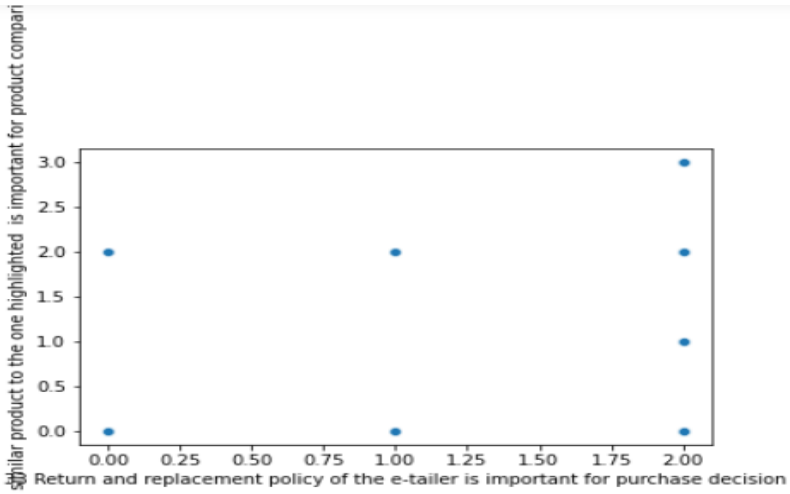


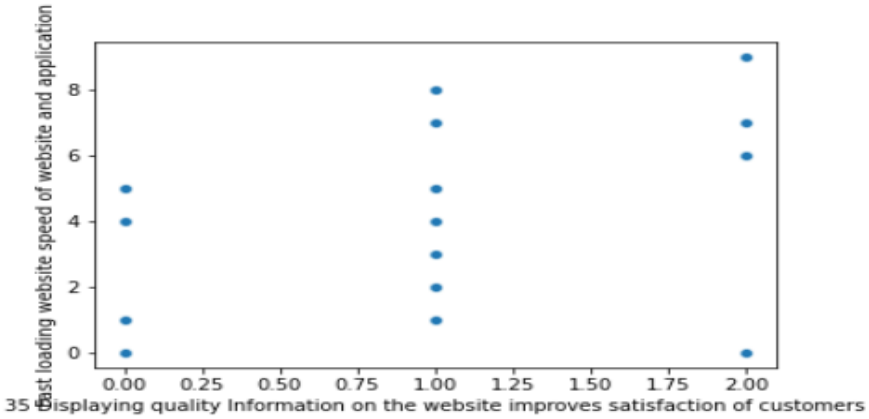
55% of the customers are using credit/debit card for the payments



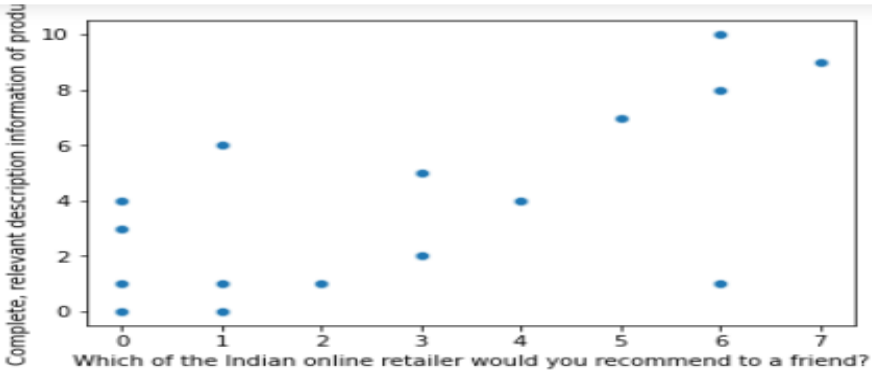
21-30 is the age group for the most of the respondets,who are shopping online

3. *SCATTERPLOT*





Positive correlation exist



This is showing a strong positive correlation having less outliers

From the above scatterplot diagram and my findings in jupyter notebook we can say that there are strong positive correlation b/w ('Which of the Indian online retailer would you recommend to a friend?' and 'Complete, relevant description information of products') & ('Speedy order delivery ' and 'Website is as efficient as before') & ('Displaying quality Information on the website improves satisfaction of customers' and 'Fast loading website speed of website and application') & ('Return and replacement policy of the e-tailer is important for purchase decision' and 'Information on similar product to the one highlighted is important for product comparison'). There are many more but we can't plot that much correlation as the inputs are 75 that's why I've plotted only the important relations.

DESCRIPTIVE STATISTICS

```
23]: df.describe(include='all').T
```

Out[23]:

	count	mean	std	min	25%	50%	75%	max
1Gender of respondent	269.0	0.327138	0.470042	0.0	0.0	0.0	1.0	1.0
2 How old are you?	269.0	1.330855	1.183774	0.0	0.0	1.0	2.0	4.0
3 Which city do you shop online from?	269.0	4.494424	3.187687	0.0	2.0	4.0	7.0	10.0
4 What is the Pin Code of where you shop online from?	269.0	18.163569	10.343865	0.0	10.0	19.0	24.0	38.0
5 Since How Long You are Shopping Online ?	269.0	2.323420	1.176357	0.0	1.0	3.0	3.0	4.0
...	...	...	...	...	...	...	...	...
Longer delivery period	269.0	2.736059	1.725857	0.0	1.0	3.0	4.0	5.0
Change in website/Application design	269.0	2.189591	2.152169	0.0	0.0	1.0	5.0	6.0
Frequent disruption when moving from one page to another	269.0	3.587361	2.557690	0.0	1.0	4.0	6.0	7.0
Website is as efficient as before	269.0	2.133829	2.171029	0.0	0.0	1.0	4.0	7.0
Which of the Indian online retailer would you recommend to a friend?	269.0	2.397770	2.362991	0.0	0.0	1.0	5.0	7.0

71 rows × 8 columns

Mean is basically the average of respective columns

Std. is standard deviation from the mean

25% percentile is representing the value below which there are 25% data are there similarly for 50 percentile and 75 percentile too.

50 percentile is also known as median of respective columns

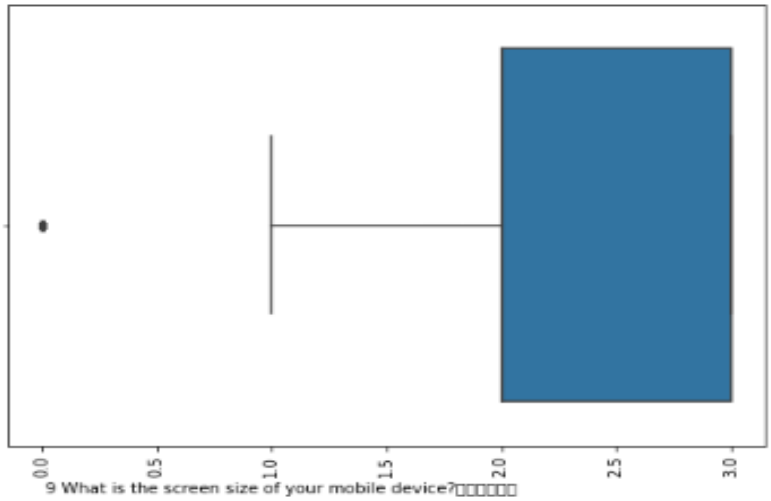
Min. and Max. Is the lowest and highest value of the respective columns?

As, we can see in above output most of the columns has median more than its mean and in the 4th column the interquartile difference is also varying hence there is possibility that outliers can be present in the dataset.

OUTLIERS

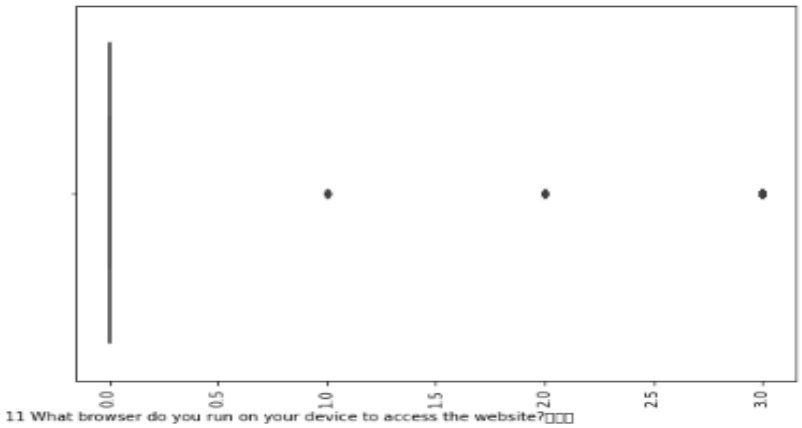
The Boxplot for attribute- "9 What is the screen size of your mobile device? " is-

AxesSubplot(0.125,0.125;0.775x0.755)



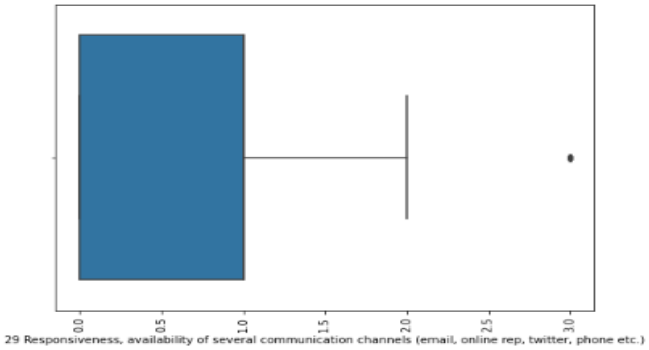
The Boxplot for attribute- "11 What browser do you run on your device to access the website? " is-

AxesSubplot(0.125,0.125;0.775x0.755)



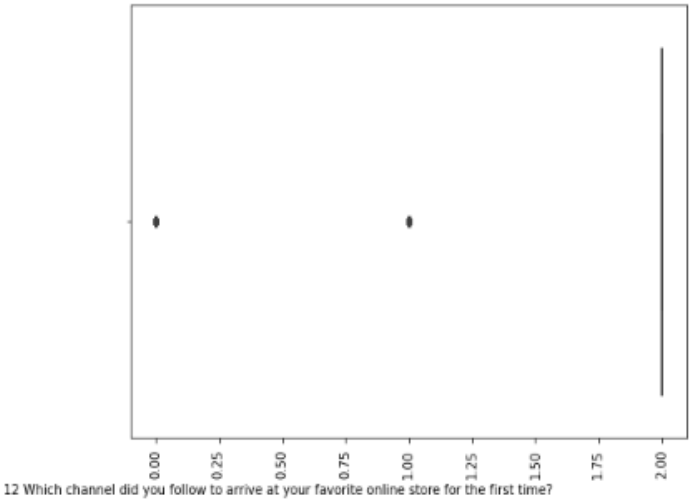
The Boxplot for attribute- "29 Responsiveness, availability of several communication channels (email, online rep, twitter, phone etc.)" is-

AxesSubplot(0.125,0.125;0.775x0.755)



The Boxplot for attribute- "12 Which channel did you follow to arrive at your favorite online store for the first time? " is-

AxesSubplot(0.125,0.125;0.775x0.755)



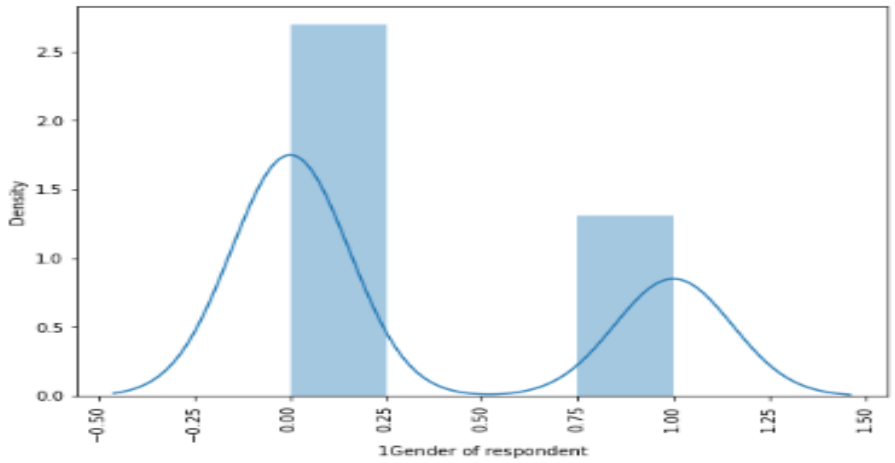
ALL THE BELOW ATTRIBUTES HAVE HIGHER OUTLIERS PRESENCE POSSIBILITY AS COMPARE TO OTHER ATTRIBUTES.

- 'WHAT IS THE SCREEN SIZE OF YOUR MOBILE DEVICE',
- 'WHAT BROWSER DO YOU RUN ON YOUR DEVICE TO ACCESS THE WEBSITE',
- 'WHICH CHANNEL DID YOU FOLLOW TO ARRIVE AT YOUR FAVORITE ONLINE STORE FOR THE FIRST TIME?'
- 'USER FRIENDLY INTERFACE OF THE WEBSITE','
- EMPATHY (READINESS TO ASSIST WITH QUERIES) TOWARDS THE CUSTOMERS',

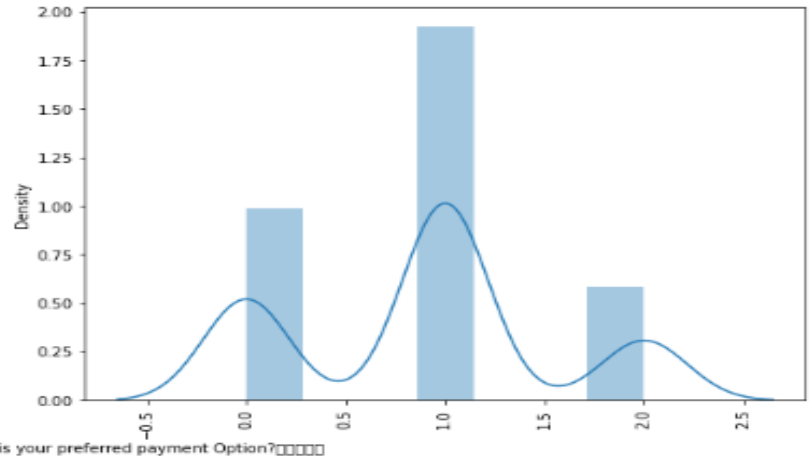
- 'BEING ABLE TO GUARANTEE THE PRIVACY OF THE CUSTOMER',
- 'RESPONSIVENESS, AVAILABILITY OF SEVERAL COMMUNICATION CHANNELS (EMAIL, ONLINE REP, TWITTER, PHONE ETC.)',
- 'NET BENEFIT DERIVED FROM SHOPPING ONLINE CAN LEAD TO USERS SATISFACTION'

DISTRIBUTIONPLOT

```
The Distplot for attribute- "1Gender of respondent" is-
 AxesSubplot(0.125,0.125;0.775x0.755)
```

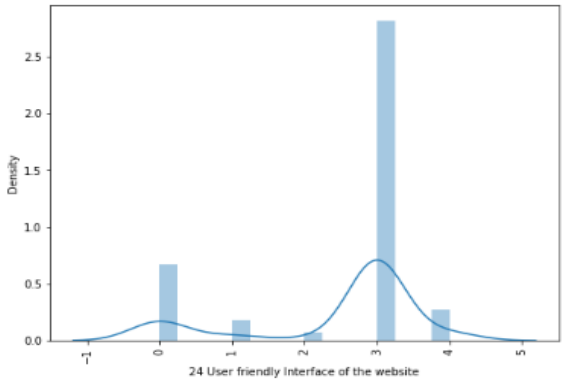



```
The Distplot for attribute- "15 What is your preferred payment Option?
" is-
 AxesSubplot(0.125,0.125;0.775x0.755)
```



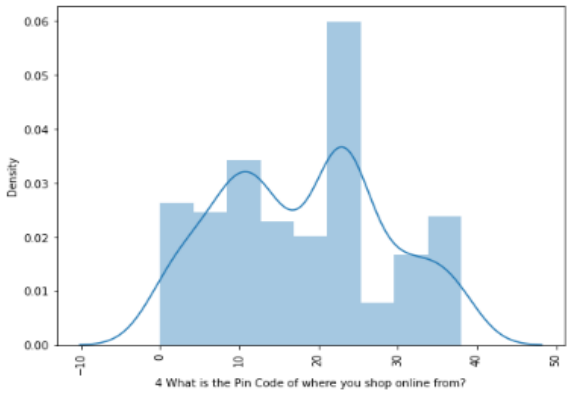
The Distplot for attribute- "24 User friendly Interface of the website" is-

AxesSubplot(0.125,0.125;0.775x0.755)



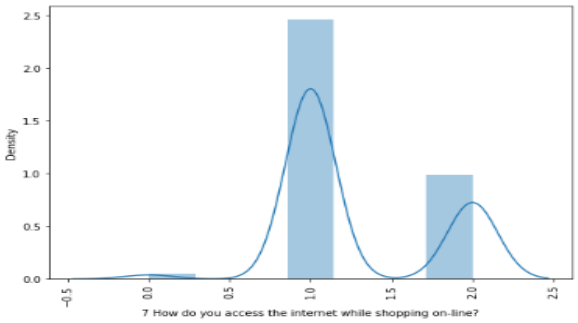
The Distplot for attribute- "4 What is the Pin Code of where you shop online from?" is-

AxesSubplot(0.125,0.125;0.775x0.755)



The Distplot for attribute- "7 How do you access the internet while shopping on-line?" is-

AxesSubplot(0.125,0.125;0.775x0.755)



All the above distribution plots are showing that almost no attributes are normally distributed and since all the attributes were in categorical format earlier, hence if skewness is present, we can't remove the skewness as it doesn't support its property on categorical attributes. Hence will skip this part and also as the integer column (which is Pin-code) is not meaningful as City is already given hence will ignore the presence of pin-code attribute in the dataset.

CONCLUSION

1. Out of 269 respondents Female are 181 which is more than twice as that of Male respondents.

2. Most of the respondents are having age group b/w 21-50.

3. Delhi, Greater Noida and Noida are top three cities when it comes to shopping online while Bulandshahr, Moradabad and Meerut are in bottom three for the same.

4. 98 respondents are shopping from last 4 years while the number is less and which 16, b/w the years 1-2 is.

5. 114 respondents have made less than 10 times online purchase in the past 1 year.

6. 189 Respondents are using mobile internet while shopping on-line unlike Wi-Fi and others.

7. 141 Respondents are using Smartphones while accessing to online shopping.

8. 230 Respondents inspired to do shopping because of Search-Engine advertisements.

9. 123 respondents takes more than 15 minutes while making their first purchase.

10. 148 Respondents uses Credit/Debit cards as the preferred payment Option while making payment.

11. 133 Respondents thinks they are getting better alternative offer that's why they abandon the Bags and Shopping Carts option.

12. The content on the website must be easy to read and understand that are said by 164 respondents.

13. 116 Respondents says Information on similar product to the one highlighted is must for product comparison.

14. 101 Respondents thinks complete information on listed seller and product being offered is important for purchase decision.

15. 132 Respondents thinks all relevant information on listed products should be stated clearly.

16. 141 Respondents thinks Ease of navigation in website is really must.

17. 189 Respondents thinks User friendly Interface of the website is must.

18. 194 Respondents thinks Empathy (readiness to assist with queries) towards the customers is must.

19. 185 Respondents thinks being able to guarantee the privacy of the customer is must.

20. 149 Respondents think Responsiveness, availability of several communication channels (email, online rep, twitter, phone etc.) is must.

21. 105 Respondents thinks online shopping monetary benefit and discounts is must.

22. 146 Respondents think shopping online must be convenient and flexible.

23. 198 Respondents think Return and replacement policy of the e-tailer is must for purchase decision

24. 122 Respondents think User satisfaction cannot exist without trust

25. 111 Respondents thinks Offering a wide variety of listed product in several category is must.

26. 135 Respondents thinks Provision of complete and relevant product information is must.

27. 149 Respondents think getting value for money spent is must.

According to me above 27 points are important to retain/add new set of customers, customer satisfaction.

1. 82 Respondents have shopped from e-commerce websites like- Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com

2. 64 Respondents thinks Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com is Easy to use website or application

3. 87 Respondents thinks Amazon.in and Flipkart.com have Visual appealing web-page layout

4. 130 Respondents thinks Amazon.in and Flipkart.com are providing Wild variety of product on offer unlike others.

5. 100 Respondents thinks Amazon.in and Flipkart.com have Complete, relevant description information of products

6. 66 Respondents thinks Amazon.in is quickest when it comes to purchase anything.

7. 65 Respondents thinks Amazon.in and Flipkart.com have Availability of several payment options unlike others.

8. 107 Respondents thinks Amazon.in have Speedy order delivery unlike others.

9. 107 Respondents thinks Amazon.in is best when it comes to the privacy of the customers.

10. 79 Respondents would like to refer Amazon.in to others

Above 10 points are just indicating that which e-commerce companies has great Positioning, Branding, Advertisement or Presence unlike others. Hence I would like conclude that- Amazon.in and Flipkart are among two companies that customers are loving it as they are doing lots of shopping because customers have great faith on their user friendly applications, Multiple Payments options, privacy & Security of customer financial information, speedy delivery, trust, offers, cashbacks and etcetera. Also, when it comes to referring the website, for shopping, to friends, family or others Amazon.in comes at first choice on their minds.