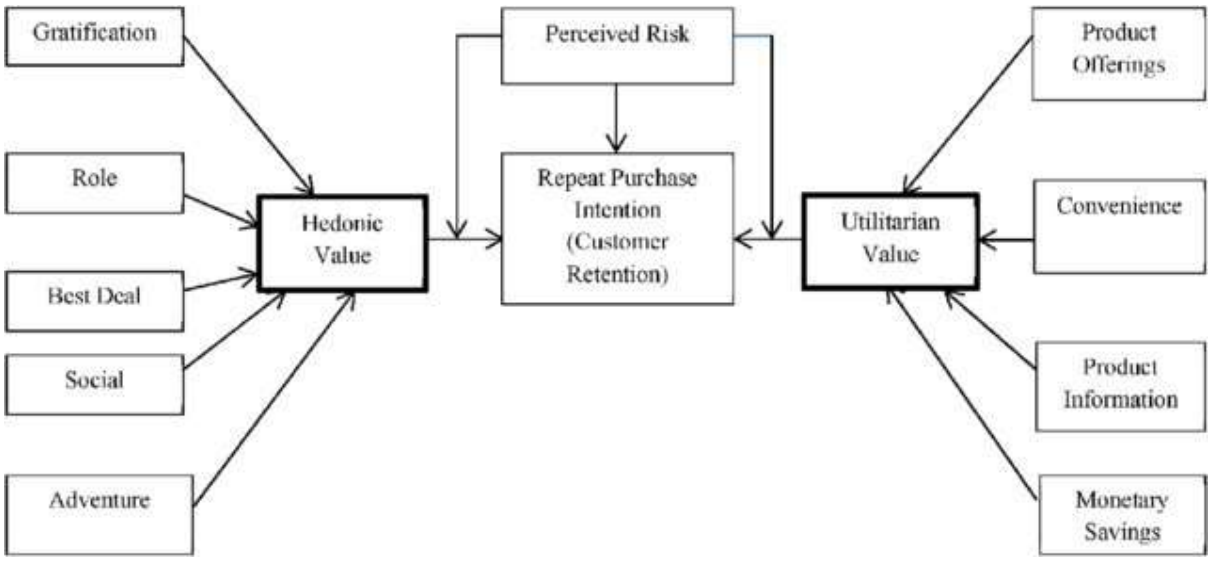
PROJECT REPORT

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

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



Solution

1) Importing required libraries

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

2) Importing raw data and encoded data from the excel file given

```
crr = pd.read_excel('C:/Users/Shashanka S/Desktop/Flip robo internship/3 - Customer_retention_dataset-/raw_data.xlsx')
cr = pd.read_excel('C:/Users/Shashanka S/Desktop/Flip robo internship/3 - Customer_retention_dataset-/customer_retention_dataset.xlsx')
```

3) Saving it in Dataframe format

4) Checking out dimensions of data

```
crr.shape
(269, 71)
```

5) Checking out the null or missing values & handling

```
crr.isnull().sum().sum()
```

No null values or missing values found

6) Checking out types of data present in different columns

Only Pincode column has int type, all others have object type

7) Encoding the data

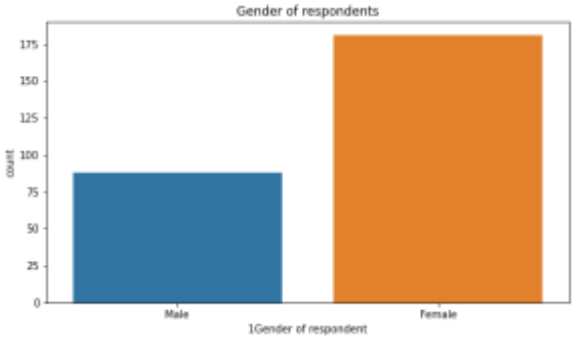
(We already have encoded data (cr), no need to encode again)

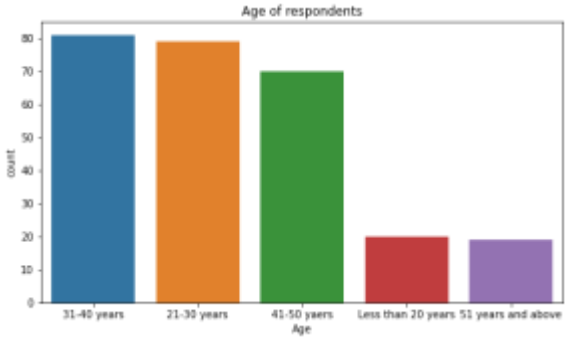
```
from sklearn.preprocessing import OrdinalEncoder
enc = OrdinalEncoder()

cr_encoded = pd.DataFrame()
for i in crr.columns:
        cr_encoded[i] = enc.fit_transform(crr[i].values.reshape(-1,1))
```

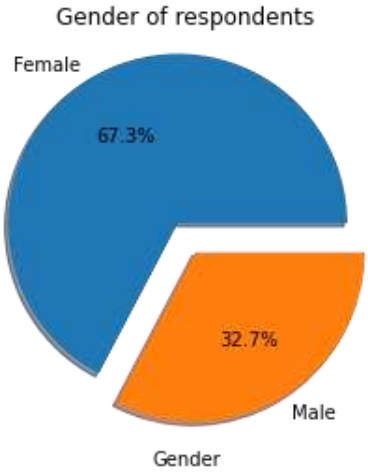
8) Checking out the column names

```
crr.columns
```

9) Exploring the data, visualization of data regarding gender and age of respondents

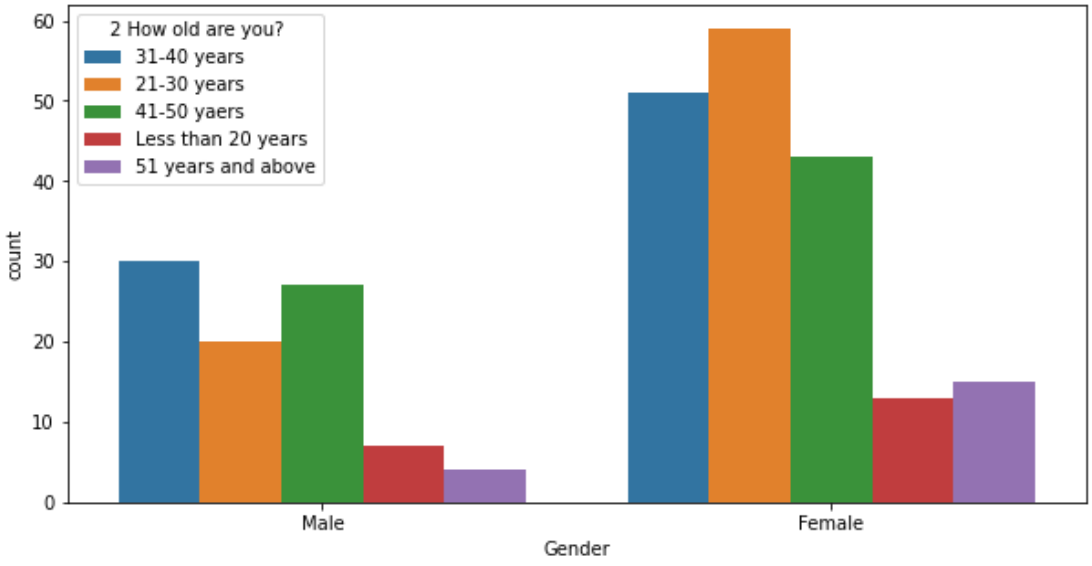






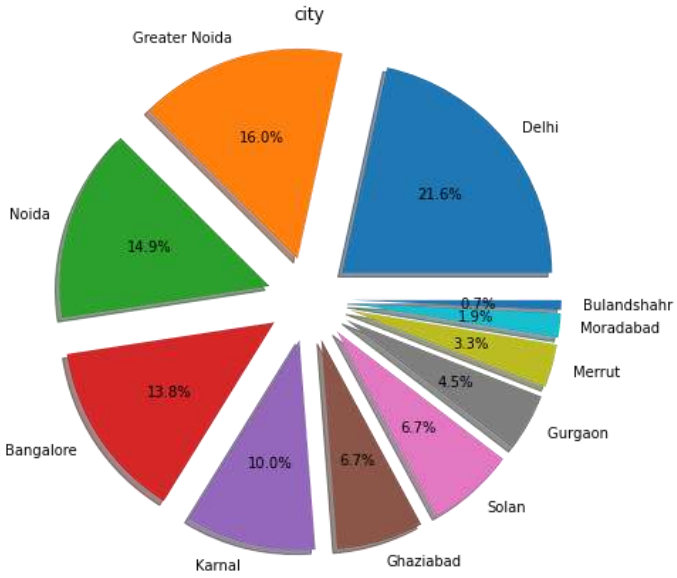
We can infer that most of the respondents are female from the above plots

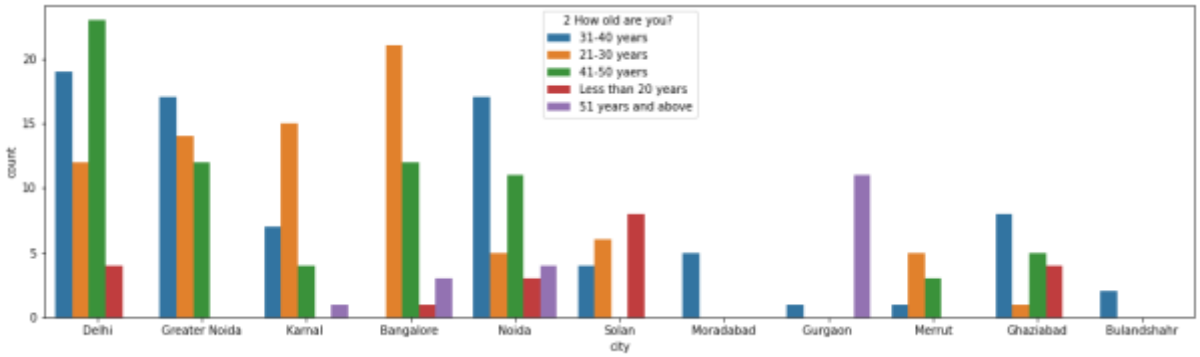
Exploring the data, visualization of data regarding gender and age of respondents together

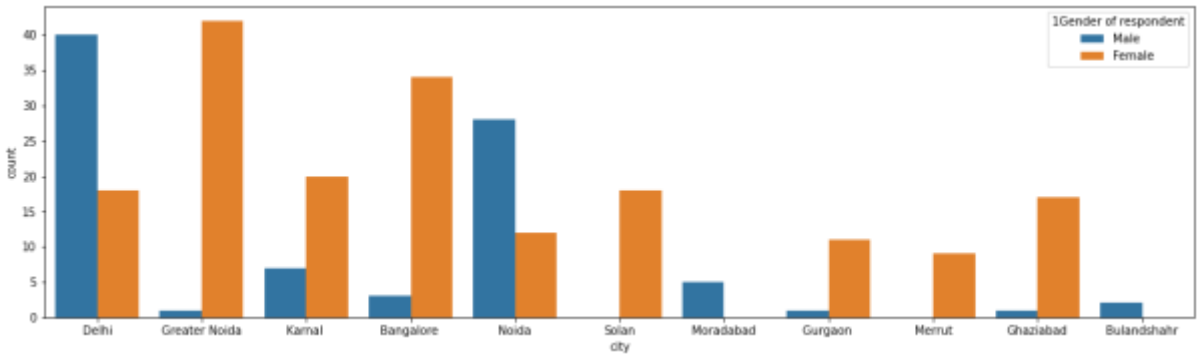


We can infer that most of the respondents are female with the age group of 21-30 years

Exploring the data, visualization of data regarding city of the customers

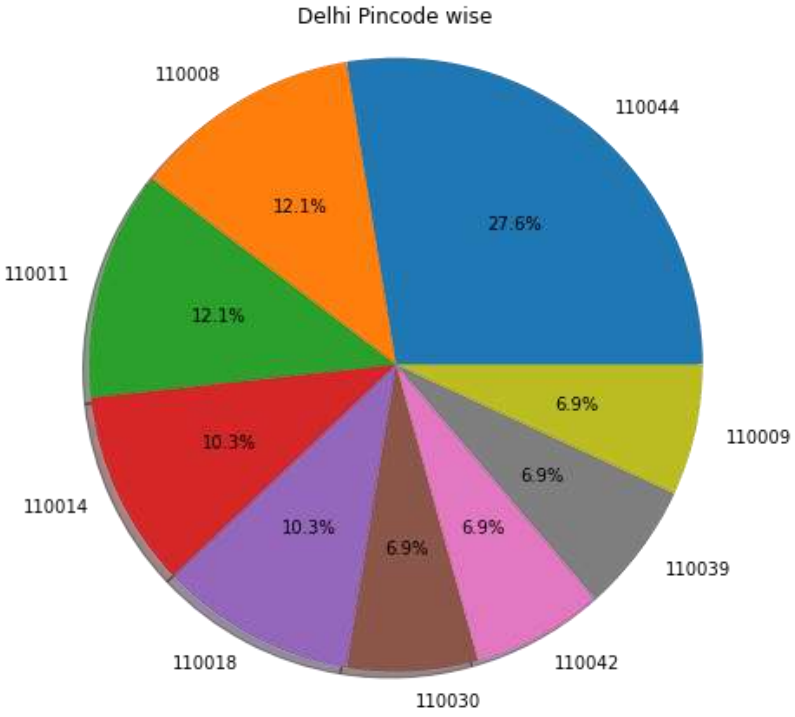


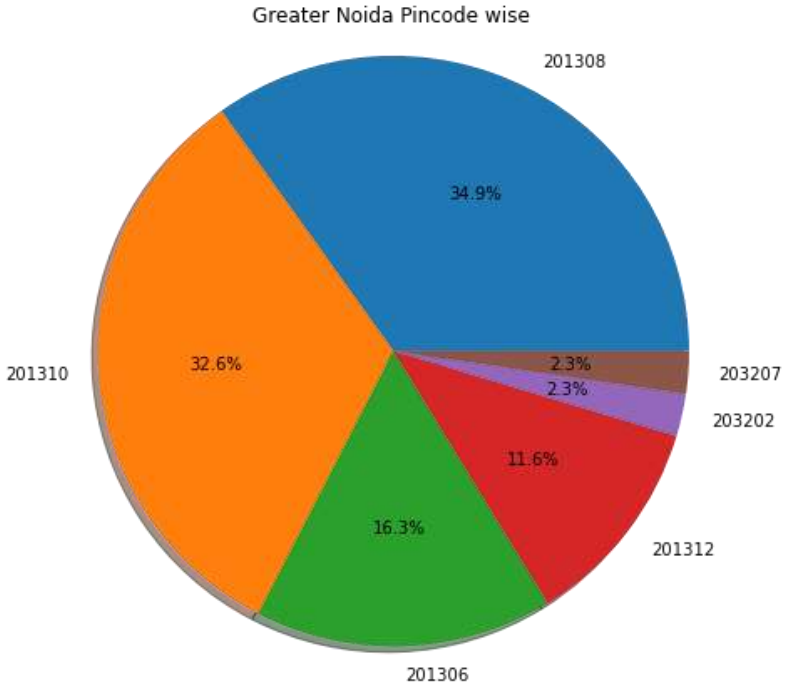


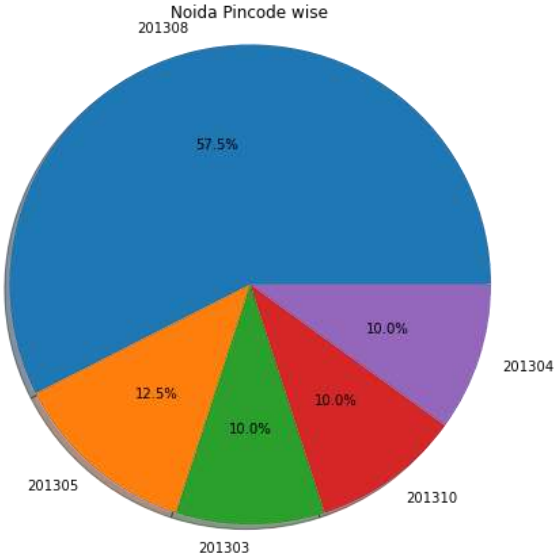


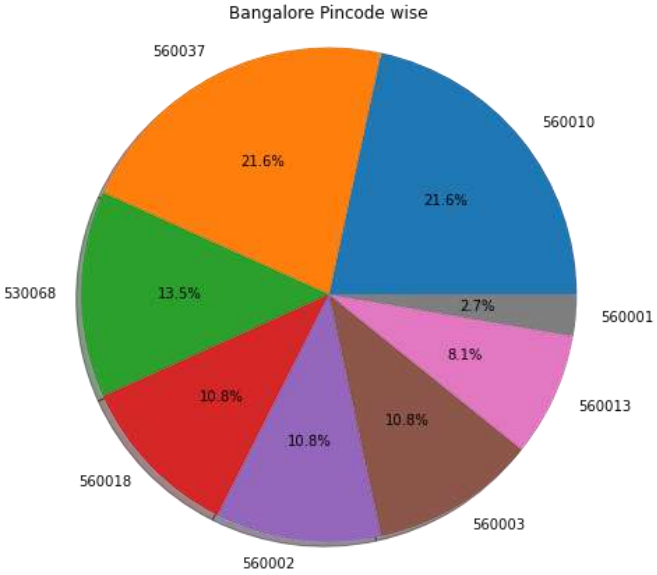
We can infer that in
1) 21.6% of the respondents are from Delhi
2) Most of the respondents from Delhi belongs to 41-50 years age group
3) Majority of the respondents from Delhi are male
Inference varies from city to city

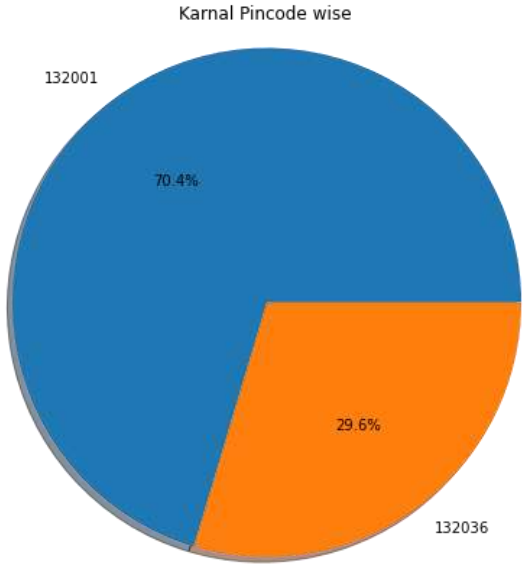
Visualization of data regarding city of the customers along with various pincode inside that city



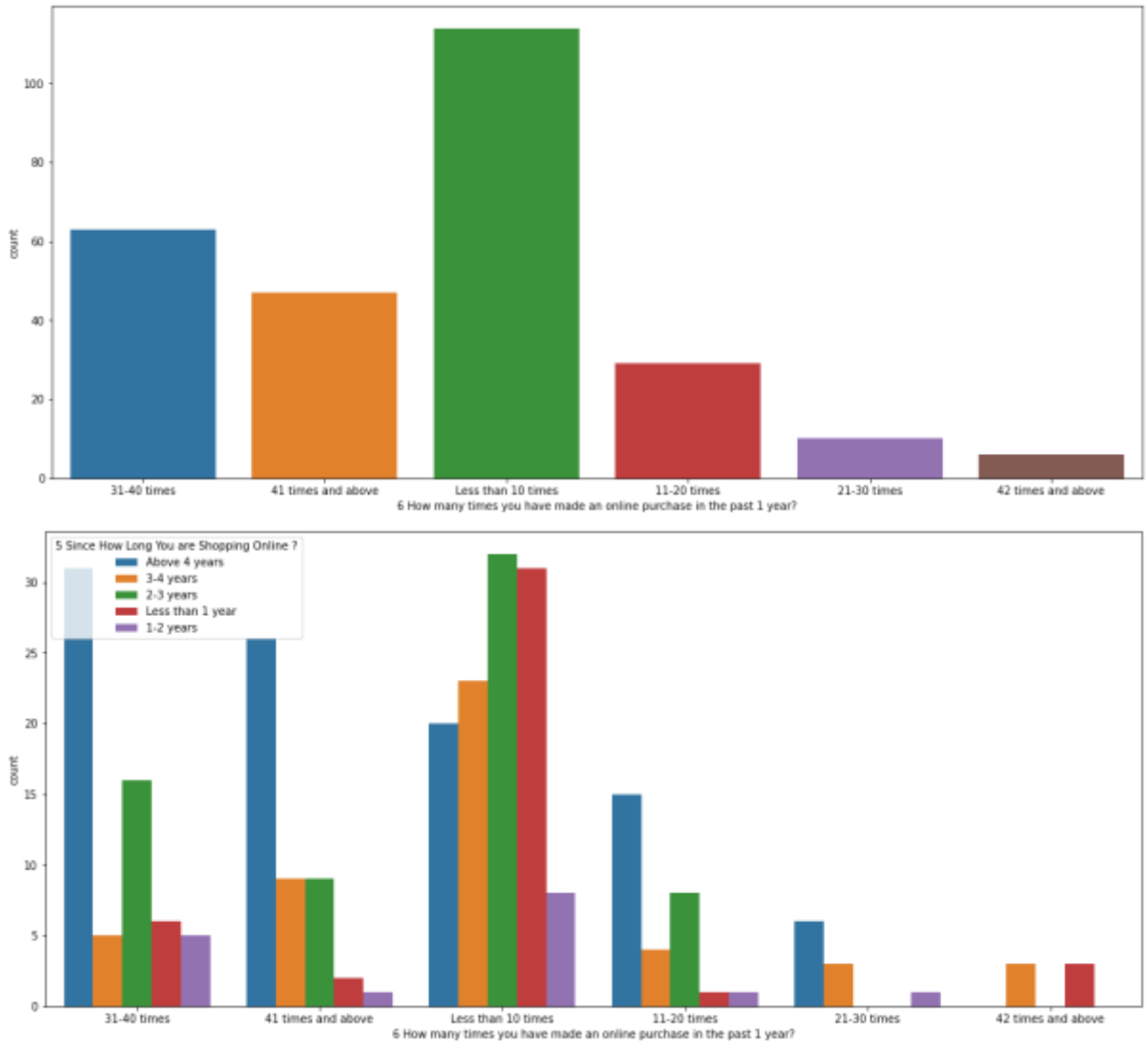



Noida Pincode wise
201308
57.5%
10.0%
201304
12.5%
10.0%
10.0%
201305
201310
201303


Bangalore Pincode wise
560037
560010
21.6%
21.6%
530068
13.5%
2.7%
560001
8.1%
560013
10.8%
10.8%
10.8%
560018
560003
560002


Karnal Pincode wise
132001
70.4%
29.6%
132036

Visualization of data regarding experience and frequency of online shopping



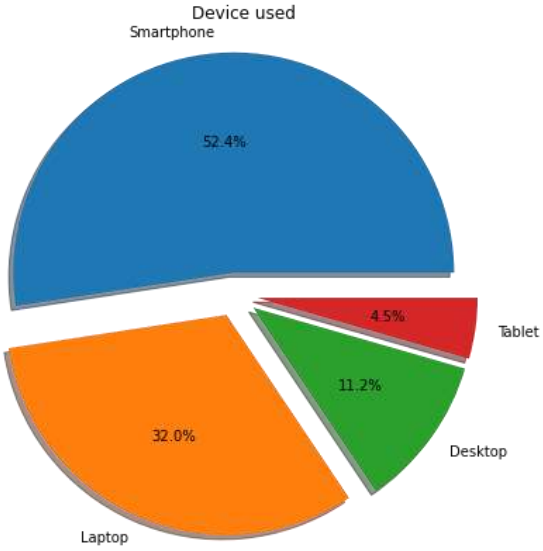
We can infer that:

1) Frequency: Most of the online shopping customers have shopped less than 10 times online in past 1 year

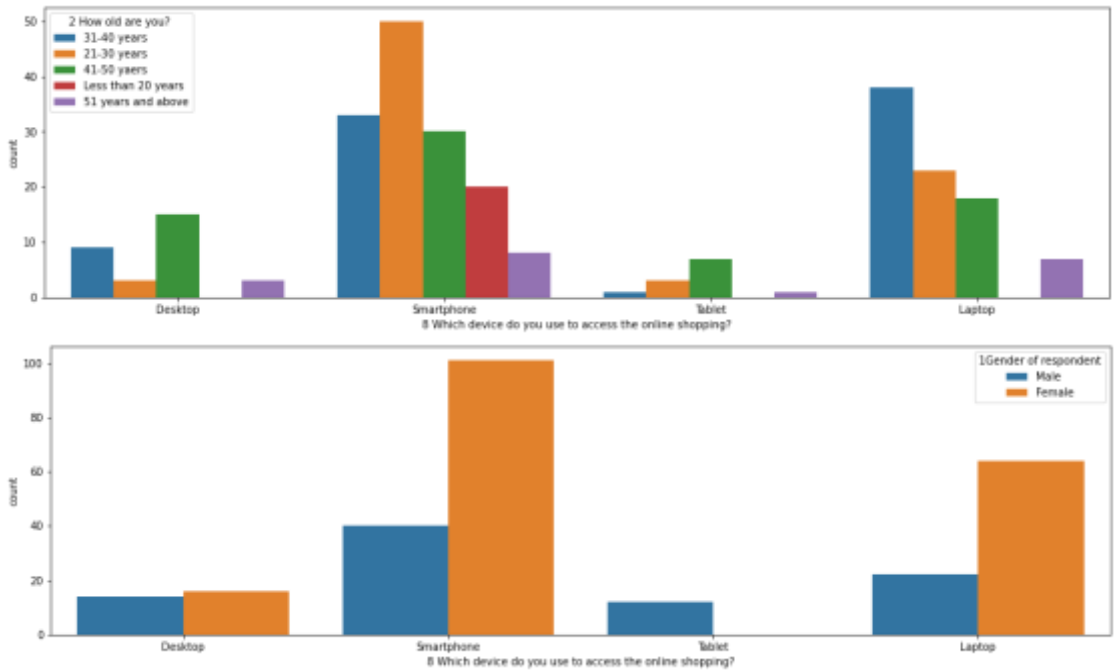
2) Experience: Most of the online shopping customers have shopped has more than 4 years of online shopping experience

3) But for customers who have shopped less than 10 times online in past 1 year, most of them have 2-3 years of online shopping experience

Visualization of data regarding device used for online shopping



We can infer that most of the customers prefer to do online shopping through smartphones

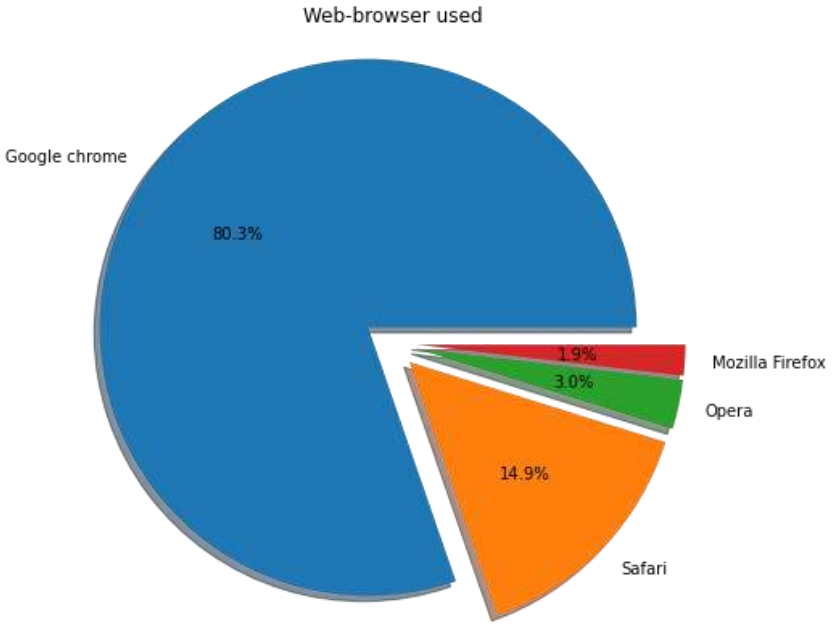


We can infer that

1) Most of the customers using smartphone for online shopping belong to age group of 21-30 years

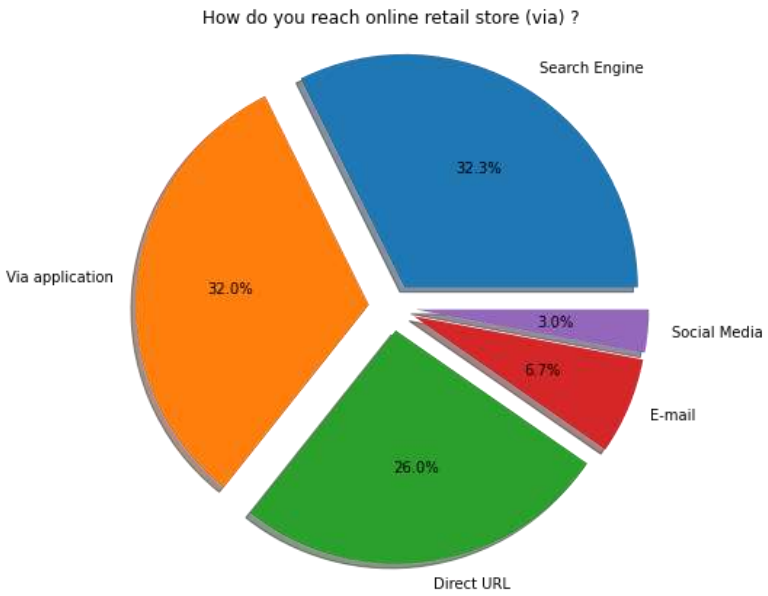
2) Most of the customers using smartphone for online shopping are female

Visualization of data regarding browser used for online shopping

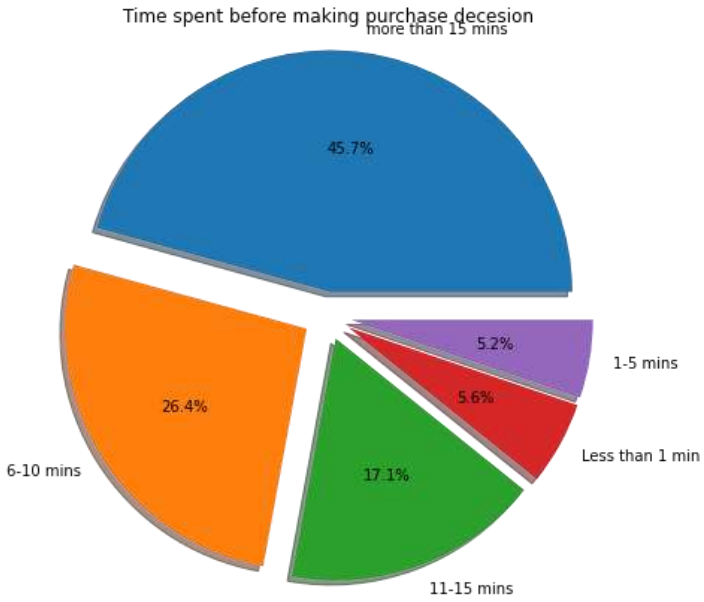


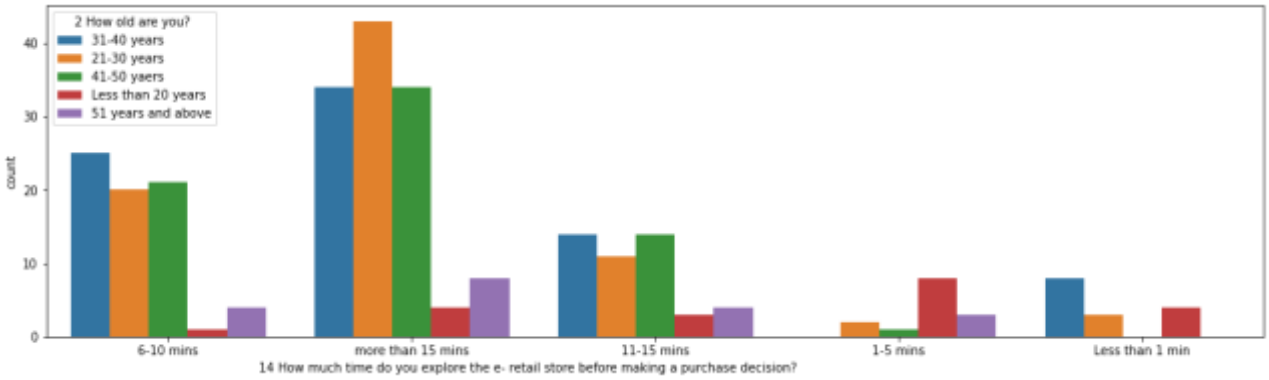
We can infer that most of the online shopping customers prefer google chrome as web-browser while shopping online

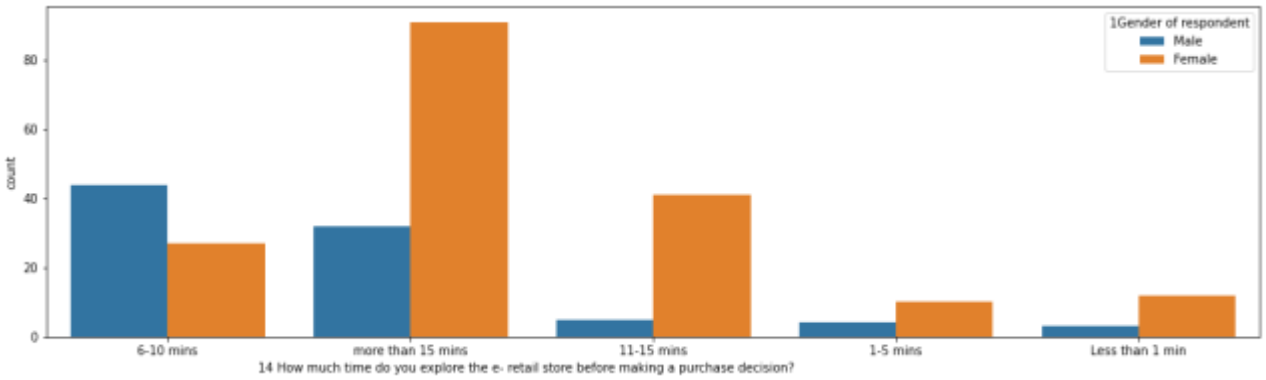
Visualization of data regarding how the online shopping customers reach online retailer store



Visualization of data regarding the spent to make purchase decision







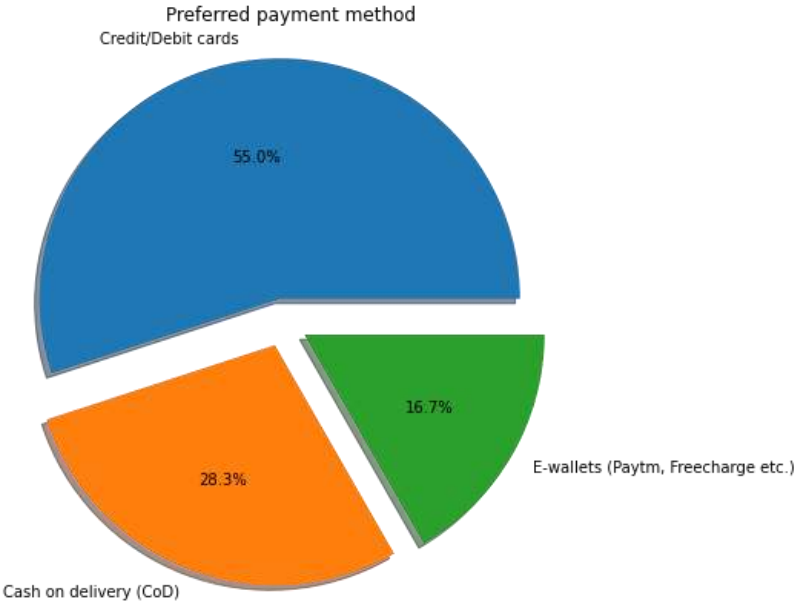
We can infer that

1) Majority of customers spend more than 15 minutes to make purchase decesion

2) Majority of customers who spend more than 15 minutes to make purchase decesion belongs to age group of 21-30 years

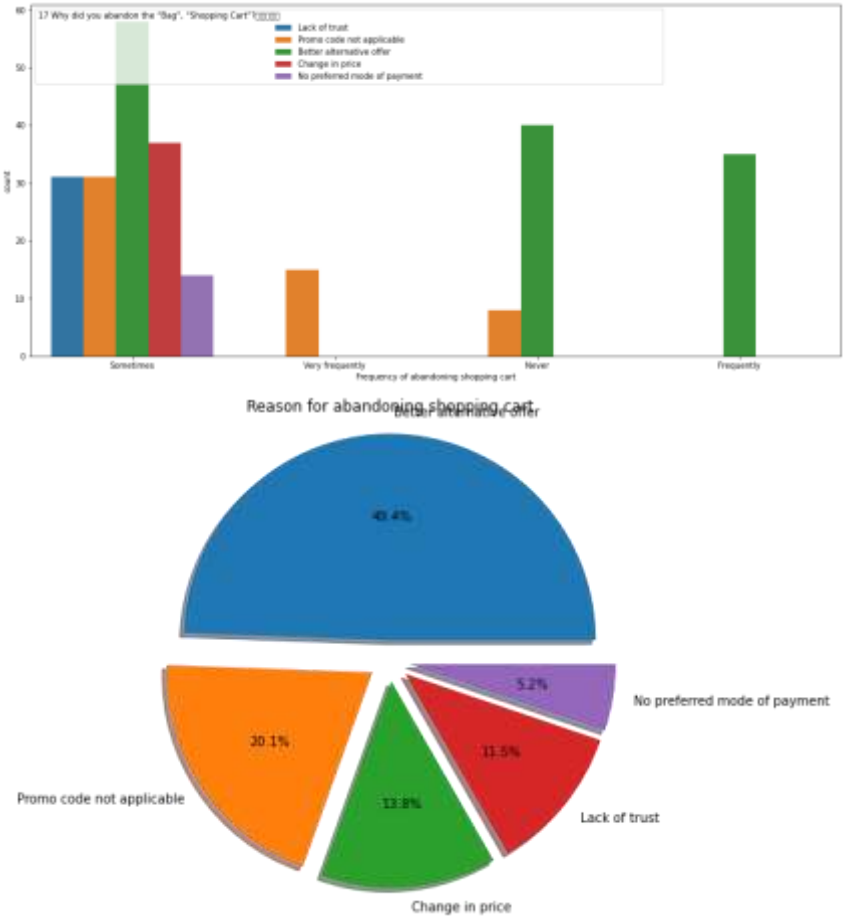
3) Majority of customers who spend more than 15 minutes to make purchase decesion are female

Visualization of data regarding preferred payment method



We can infer that most of the online shopping customers prefer payment through cards

Visualization of data regarding frequency & reason of abandoning shopping cart

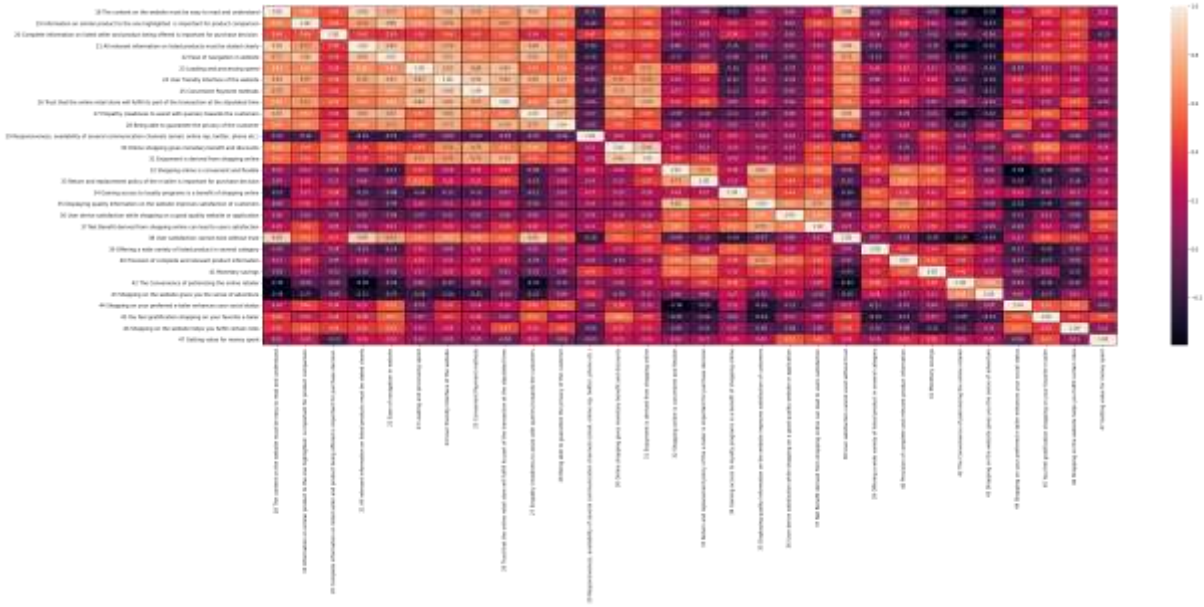


We can infer that sometimes online shopping customers abandon cart, due to better alternate offer available

10) Seperating factors affecting the customer retention or satisfaction as per feedback data

cr_coded = cr.iloc[:,17:47]

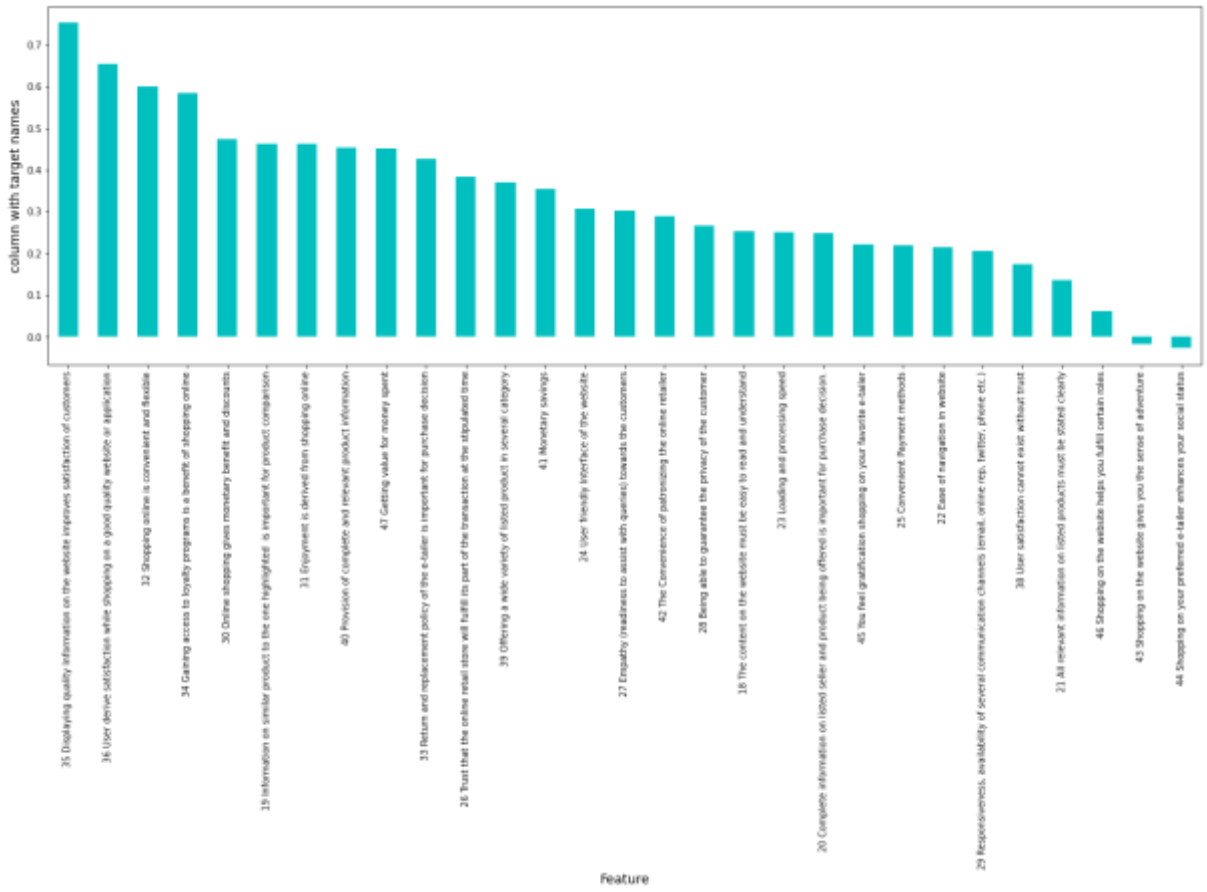
11) Finding coefficients of correlation of all against each other and plotting heatmap for vizualization



Since there are too many columns, lets prioritize and select factors based on highest value of coefficient of correlation with respect to target

12) Considering 'Net Benefit derived from shopping online can lead to users satisfaction' as target feature

```
35 Displaying quality Information on the website improves satisfaction of
customers                                  0.753178
36 User derive satisfaction while shopping on a good quality website or a
pplication                                 0.654770
32      Shopping      online      is      convenient      and      flexible
0.601184
34 Gaining access to loyalty programs is a benefit of shopping online
0.585258
30    Online    shopping    gives    monetary    benefit    and    discounts
0.473854
19 Information on similar product to the one highlighted is important for
product comparison                         0.464319
31      Enjoyment      is      derived      from      shopping      online
0.463977
40    Provision    of    complete    and    relevant    product    information
0.454384
47          Getting          value          for          money          spent
0.451091
33 Return and replacement policy of the e-tailer is important for purchas
e decision                                 0.427085
```

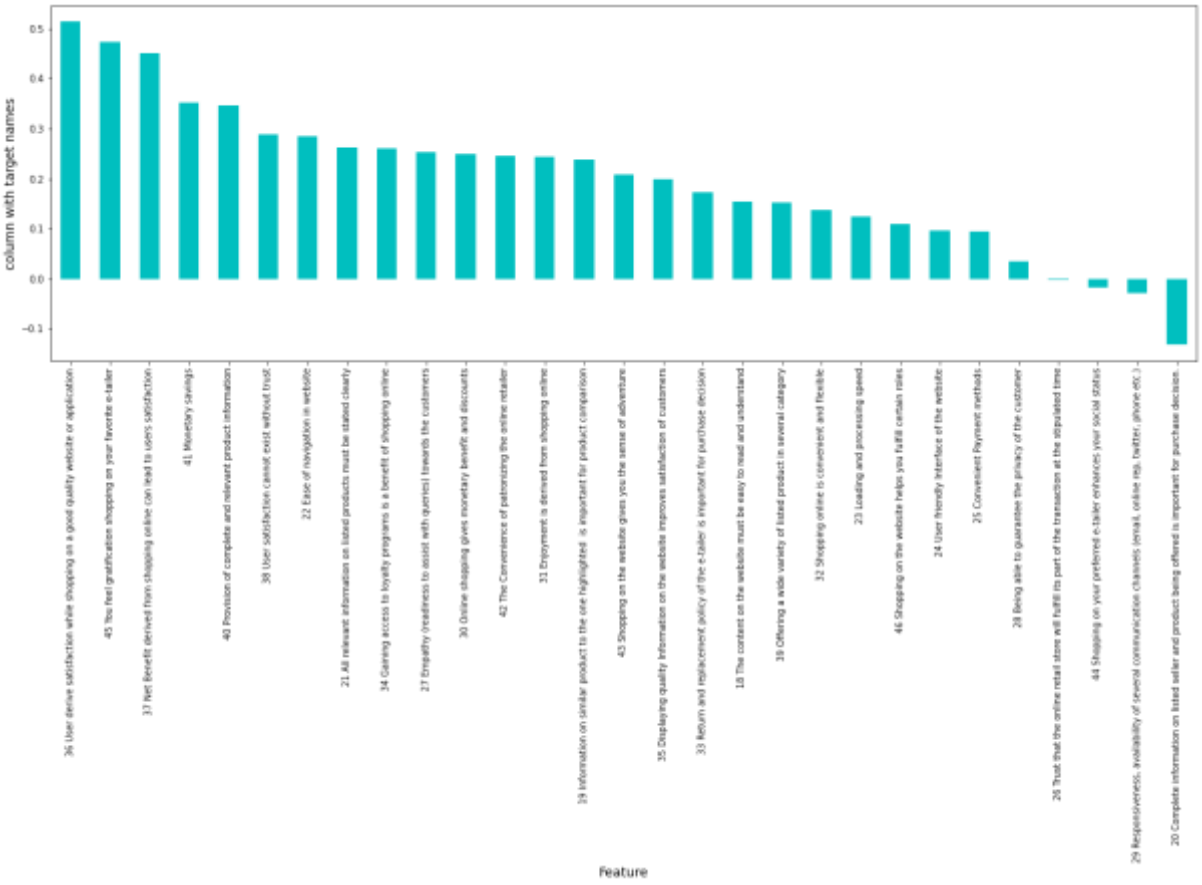


Factors affecting are displayed in descending order of coefficient of correlation

This means: 'Displaying quality Information on the website improves satisfaction of customers' influences the more to online shopping customers to buy products online

13) Considering 'Getting value for money spent' as target feature

```
36 User derive satisfaction while shopping on a good quality website or a
pplication                                  0.513614
45   You   feel   gratification   shopping   on   your   favorite   e-tailer
0.473175
37 Net Benefit derived from shopping online can lead to users satisfactio
n                                           0.451091
41                              Monetary                                savings
0.352887
40   Provision   of   complete   and   relevant   product   information
0.347117
38     User     satisfaction     cannot     exist     without     trust
0.289592
22          Ease          of          navigation          in          website
0.284420
21 All relevant information on listed products must be stated clearly
0.261774
34 Gaining access to loyalty programs is a benefit of shopping online
0.261204
27 Empathy (readiness to assist with queries) towards the customers
0.253112
```



Factors affecting are displayed in descending order of coefficient of correlation

This means: 'User derive satisfaction while shopping on a good quality website or application' influences online shopping customers to buy products online

14) Lets check how various categorized factors correlated to the target

Let target be = '37 Net Benefit derived from shopping online can lead to users satisfaction'

Hedonic Value factors

1) Gratification:'45 You feel gratification shopping on your favorite e-tailer'

2) Role: '46 Shopping on the website helps you fulfill certain roles'

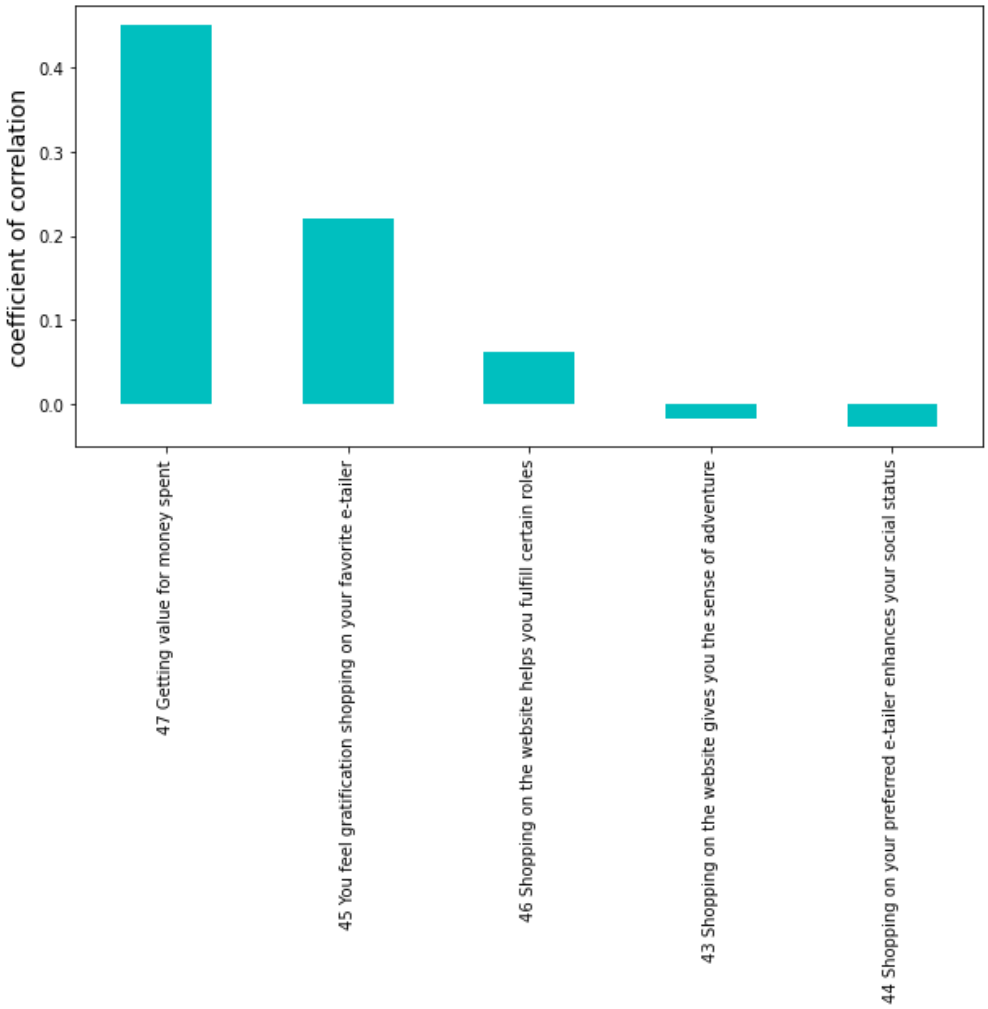
3) Best deal:'47 Getting value for money spent'

4) Social:'44 Shopping on your preferred e-tailer enhances your social status'

5) Adventure:'43 Shopping on the website gives you the sense of adventure'

```
47 Getting value for money spent                                  0.451091
45 You feel gratification shopping on your favorite e-tailer      0.221208
46 Shopping on the website helps you fulfill certain roles        0.061530
43 Shopping on the website gives you the sense of adventure      -0.017818
```

```
44 Shopping on your preferred e-tailer enhances your social status   -0.026833
```



Factors affecting are displayed in descending order of coefficient of correlation

This means: 'Getting value for money spent' influences the more online shopping customers to buy products online

Risk presceived factors:

'26 Trust that the online retail store will fulfill its part of the transaction at the stipulated time'

'27 Empathy (readiness to assist with queries) towards the customers'

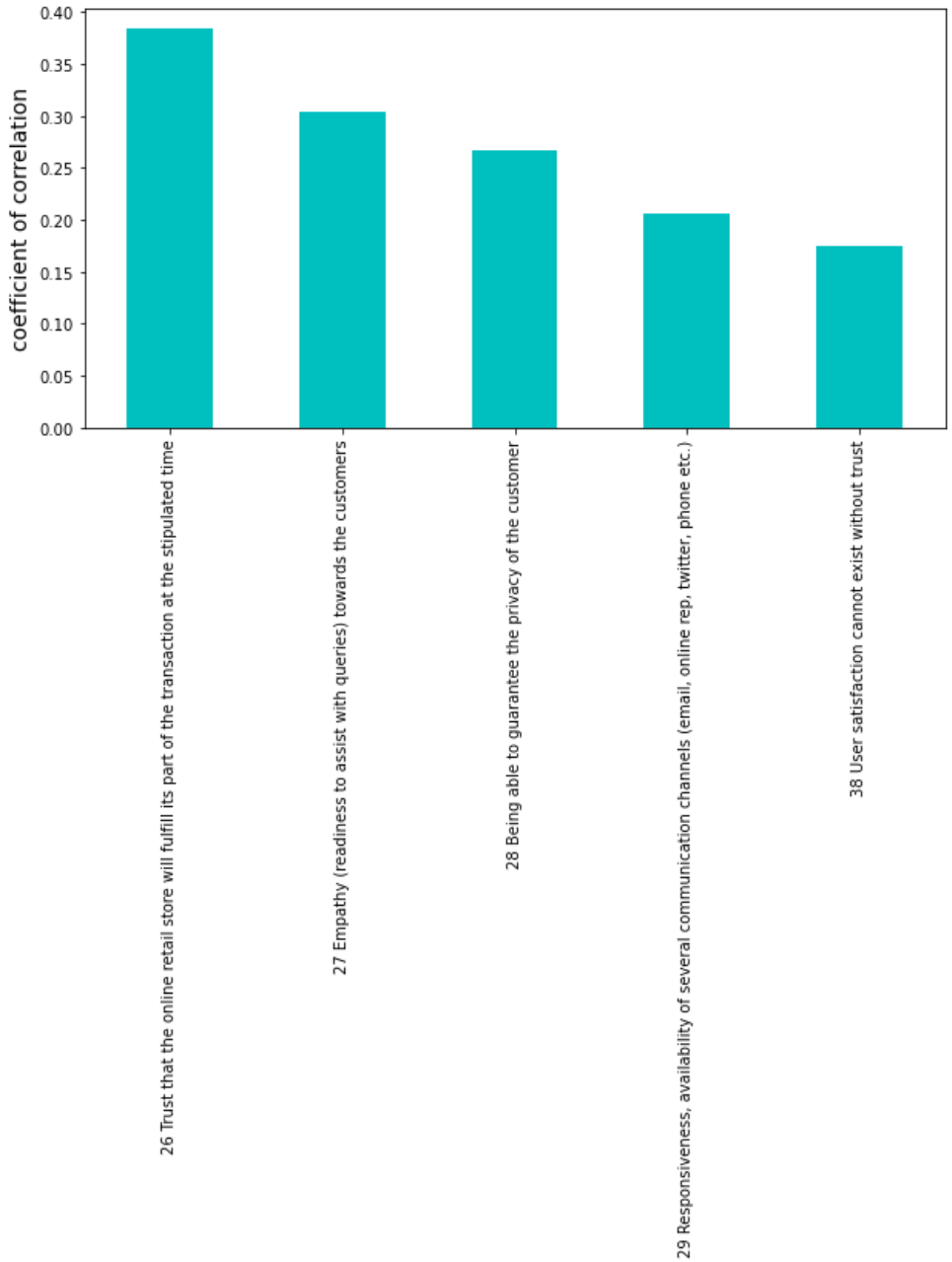
'28 Being able to guarantee the privacy of the customer'

'29 Responsiveness, availability of several communication channels (email, online rep, twitter, phone etc.)'

'38 User satisfaction cannot exist without trust'

```
26 Trust that the online retail store will fulfill its part of the tran
saction at the stipulated time         0.384287
```

```
27 Empathy (readiness to assist with queries) towards the customers
0.303216
28 Being able to guarantee the privacy of the customer
0.266504
29 Responsiveness, availability of several communication channels (emai
l, online rep, twitter, phone etc.)    0.205948
38 User satisfaction cannot exist without trust
0.174921
```



Factors affecting are displayed in descending order of coefficient of correlation

This means: 'Trust that the online retail store will fulfill its part of the transaction at the stipulated time' influences more online shopping customers to buy products online

Utilitarian value factors:

1) Product offerings:

'39 Offering a wide variety of listed product in several category'

2) Convenience:

'25 Convenient Payment methods'

'32 Shopping online is convenient and flexible'

'42 The Convenience of patronizing the online retailer'

3) Product information:

'20 Complete information on listed seller and product being offered is important for purchase decision.'

'21 All relevant information on listed products must be stated clearly'

'35 Displaying quality Information on the website improves satisfaction of customers'

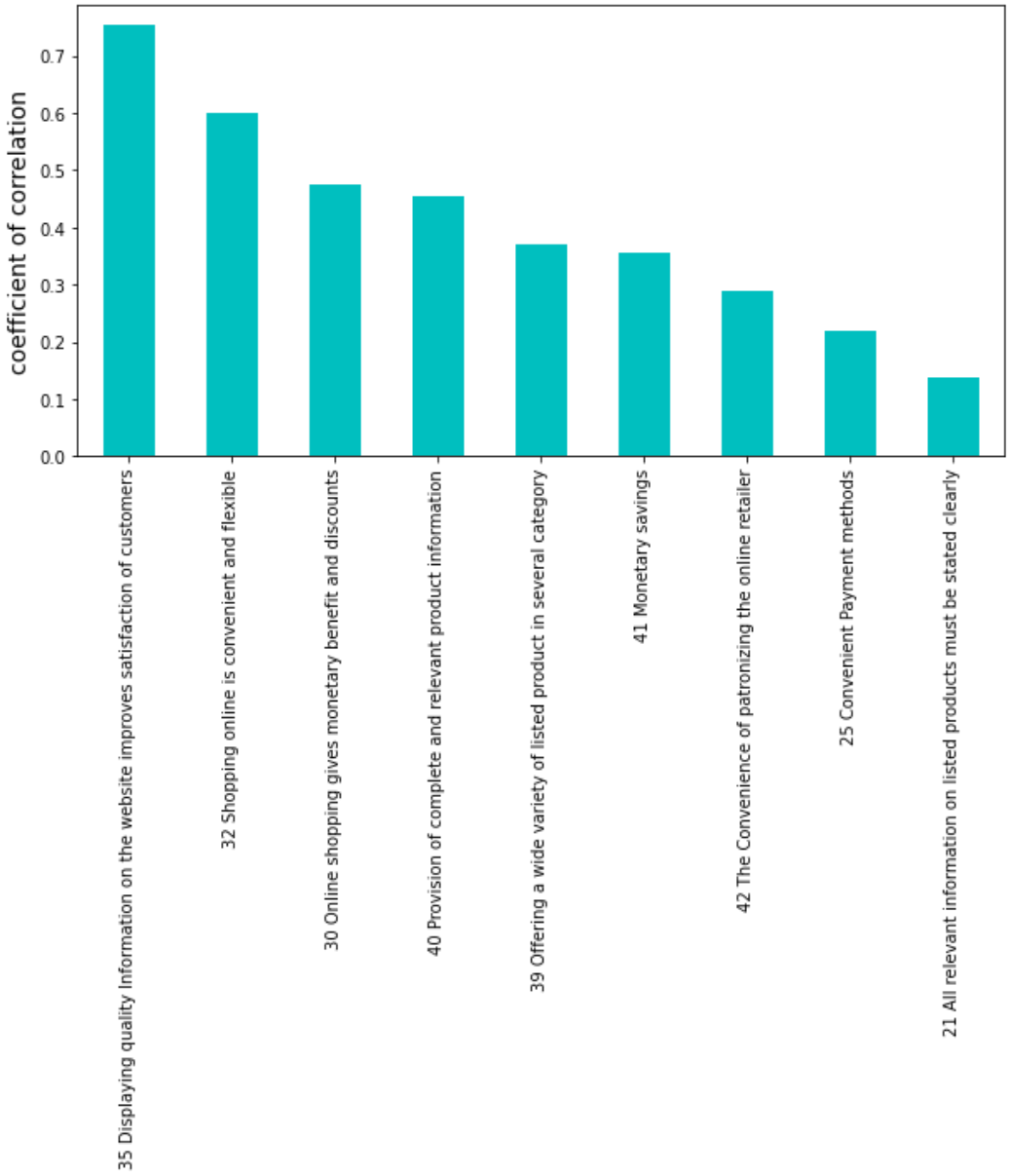
'40 Provision of complete and relevant product information'

4) Monitory saving:

'41 Monetary savings'

'30 Online shopping gives monetary benefit and discounts'

```
35 Displaying quality Information on the website improves satisfaction o
f customers    0.753178
32     Shopping     online     is     convenient     and     flexible
0.601184
30   Online   shopping   gives   monetary   benefit   and   discounts
0.473854
40   Provision   of   complete   and   relevant   product   information
0.454384
39 Offering a wide variety of listed product in several category
0.370172
41 Monetary savings                                                  0
.354810
42   The   Convenience   of   patronizing   the   online   retailer
0.290370
25              Convenient              Payment              methods
0.219503
21 All relevant information on listed products must be stated clearly
0.136561
```

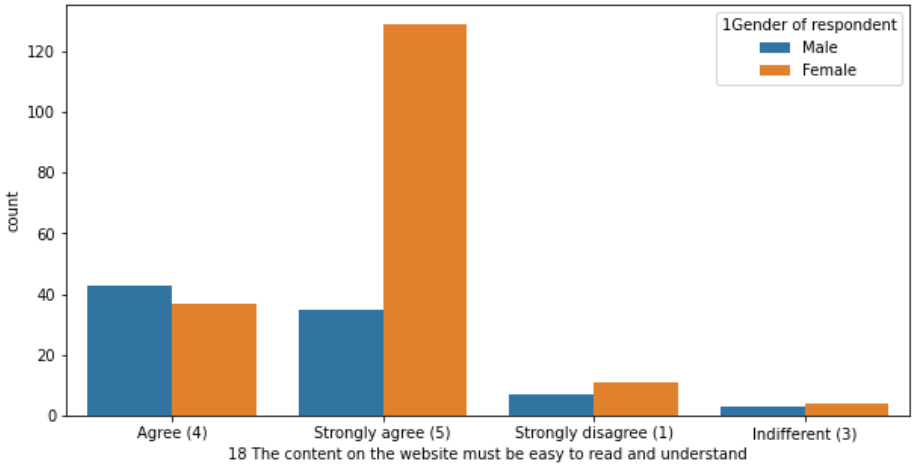


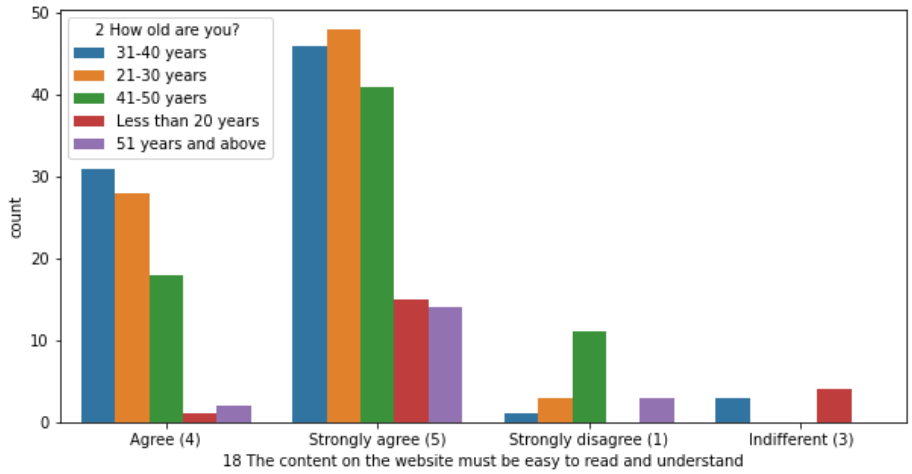
Factors affecting are displayed in descending order of coefficient of correlation

This means: 'Displaying quality Information on the website improves satisfaction of customers' influences more online shopping customers to buy products online

15) Infering from the above correlation analysis, lets consider following factors for further data analysis for customer satisfaction & retention

18 The content on the website must be easy to read and understand
21 All relevant information on listed products must be stated clearly
27 Empathy (readiness to assist with queries) towards the customers
22 Ease of navigation in website
24 User friendly Interface of the website
25 Convenient Payment methods
35 Displaying quality Information on the website improves satisfaction of customers
36 User derive satisfaction while shopping on a good quality website or application
32 Shopping online is convenient and flexible
34 Gaining access to loyalty programs is a benefit of shopping online
30 Online shopping gives monetary benefit and discounts

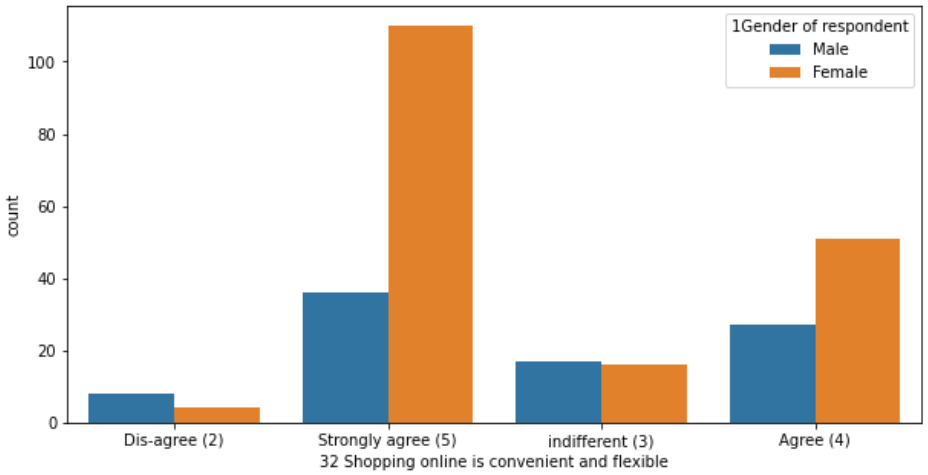


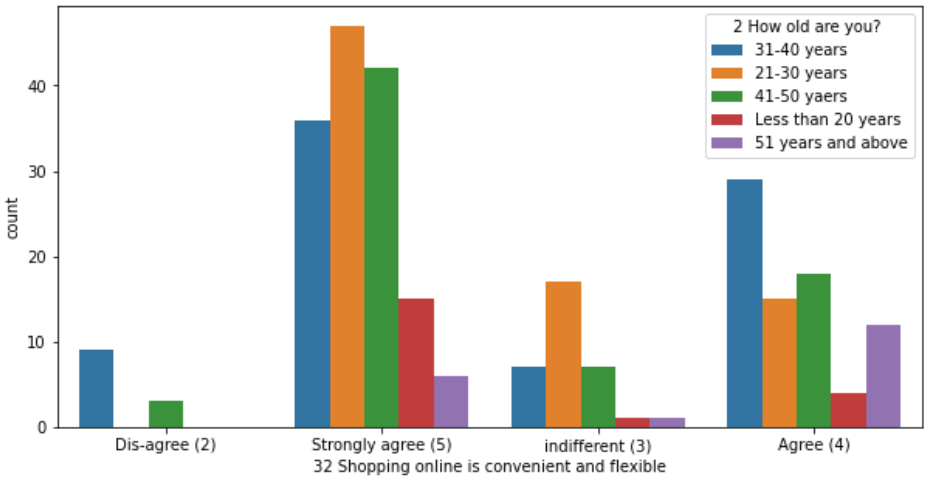


Inference:

1) Most of the online shopping customers who has opinion that 'The content on the website must be easy to read and understand' are female

2) Most of the online shopping customers who has opinion that 'The content on the website must be easy to read and understand' are of age group 21-30 years

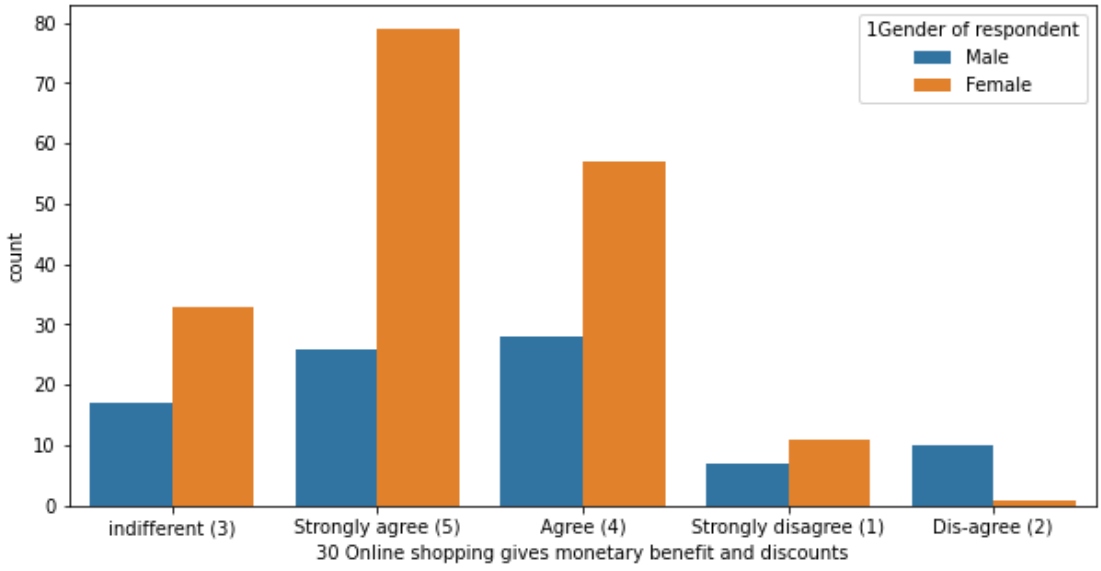


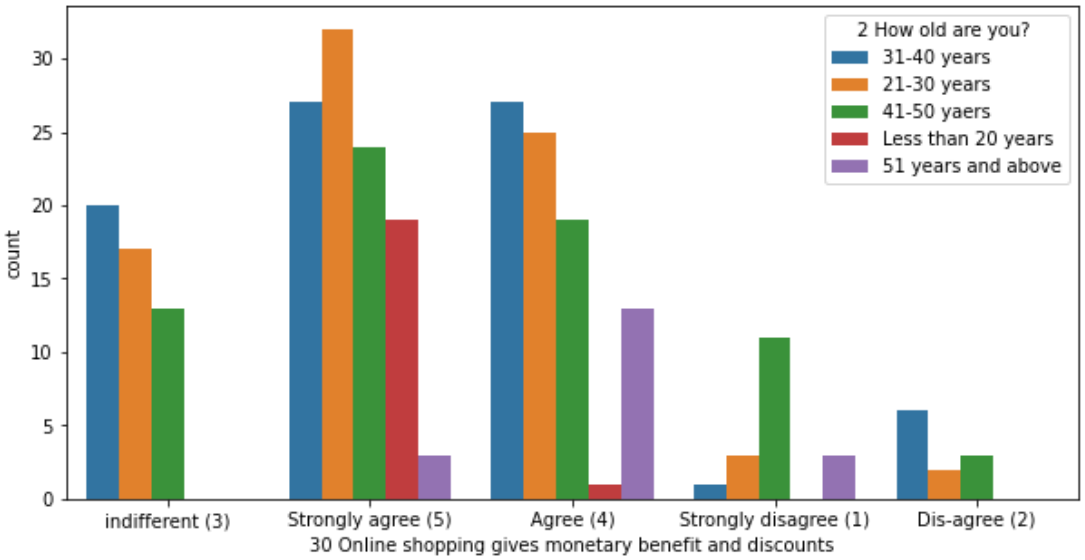


Inference:

1) Most of the online shopping customers who has opinion that 'Shopping online is convenient and flexible' are female

2) Most of the online shopping customers who has opinion that 'Shopping online is convenient and flexible' are of age group 21-30 years



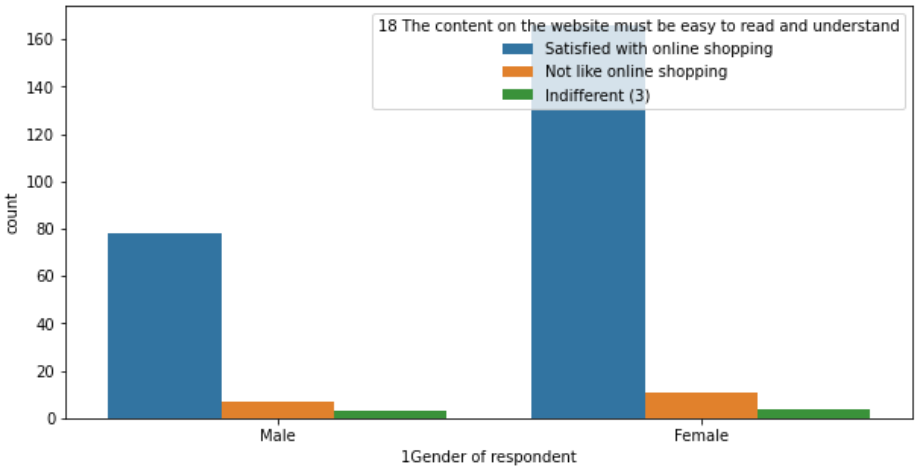


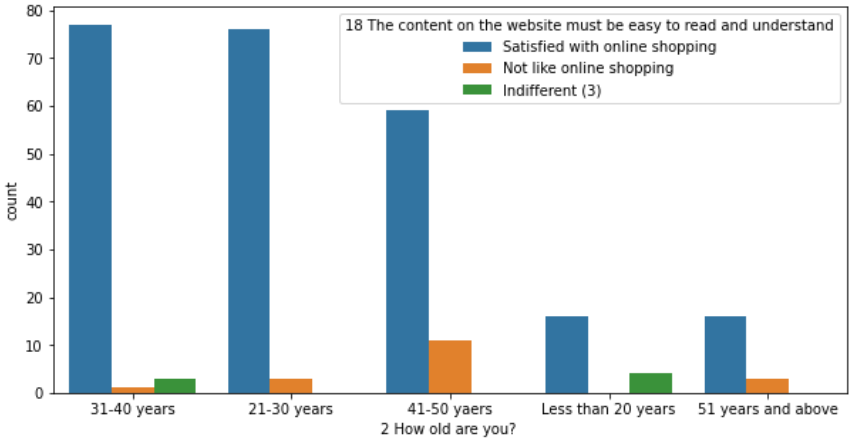
Inference:

1) Most of the online shopping customers who has opinion that 'Online shopping gives monetary benefit and discounts' are female

2) Most of the online shopping customers who has opinion that 'Online shopping gives monetary benefit and discounts' are of age group 21-30 years

Merging Agree & Strongly agree and Disagree & Strongly disagree

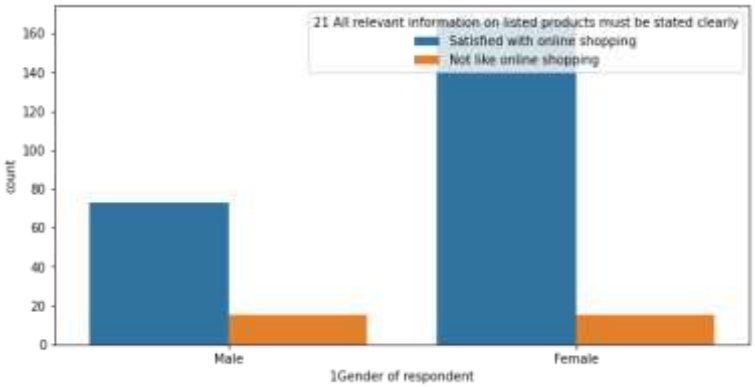


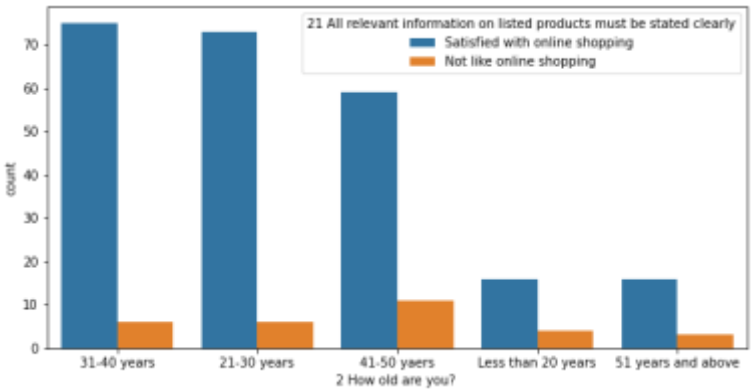


Inference:

1) Most of the customers influenced strongly by 'The content on the website must be easy to read and understand' are female

2) Most of the customers influenced strongly by 'The content on the website must be easy to read and understand' are of age group 21-50 years

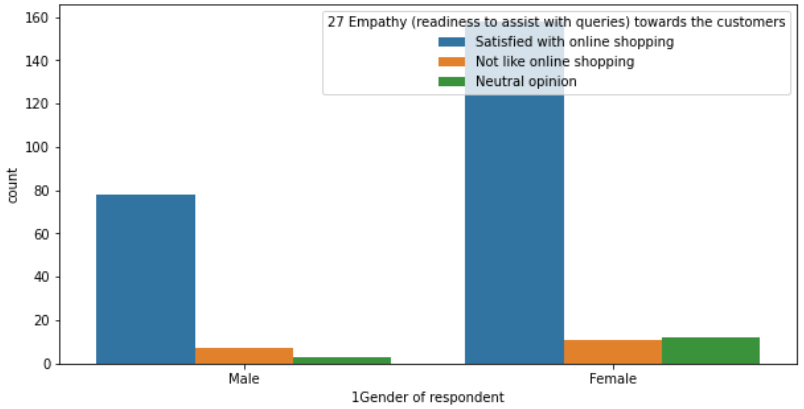


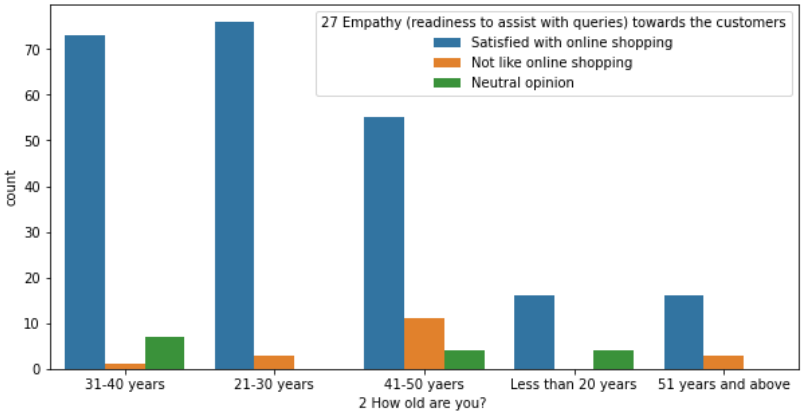


Inference:

1) Most of the customers influenced strongly by 'All relevant information on listed products must be stated clearly' are female

2) Most of the customers influenced strongly by 'All relevant information on listed products must be stated clearly' are of age group 21-50 years

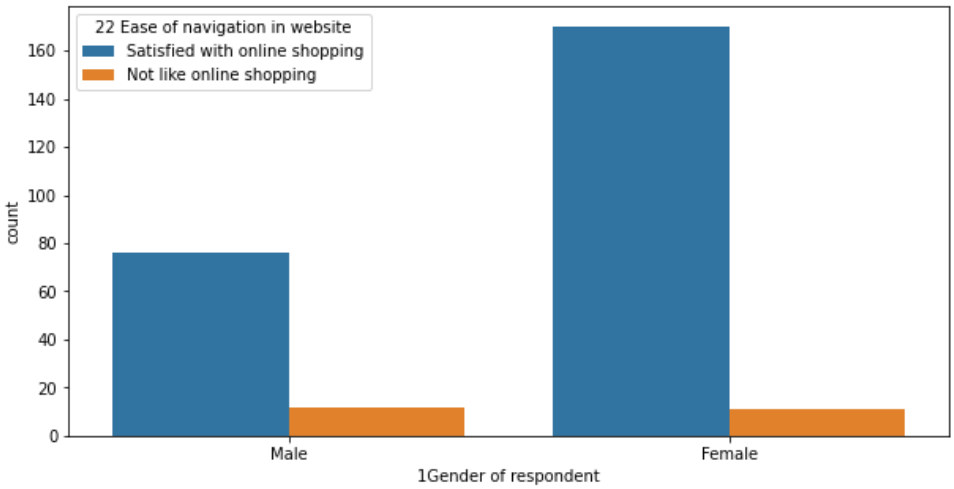


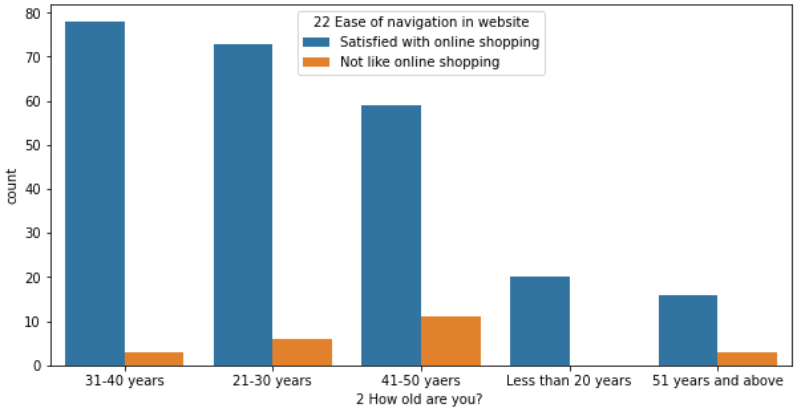


Inference:

1) Most of the customers influenced strongly by 'Empathy (readiness to assist with queries) towards the customers' are female

2) Most of the customers influenced strongly by 'Empathy (readiness to assist with queries) towards the customers' are age group 21-50 years

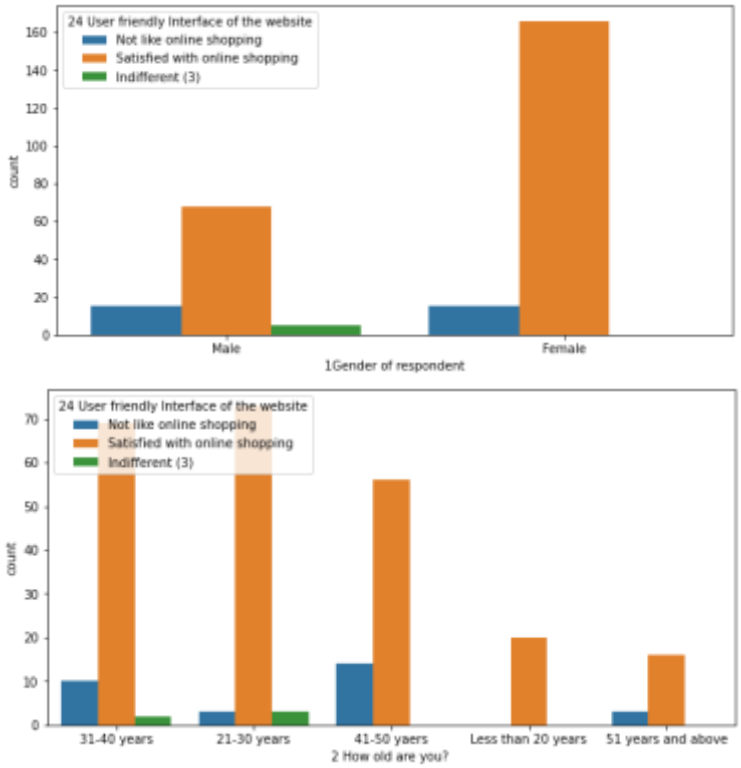




Inference:

1) Most of the customers influenced strongly by 'Ease of navigation in website' are female

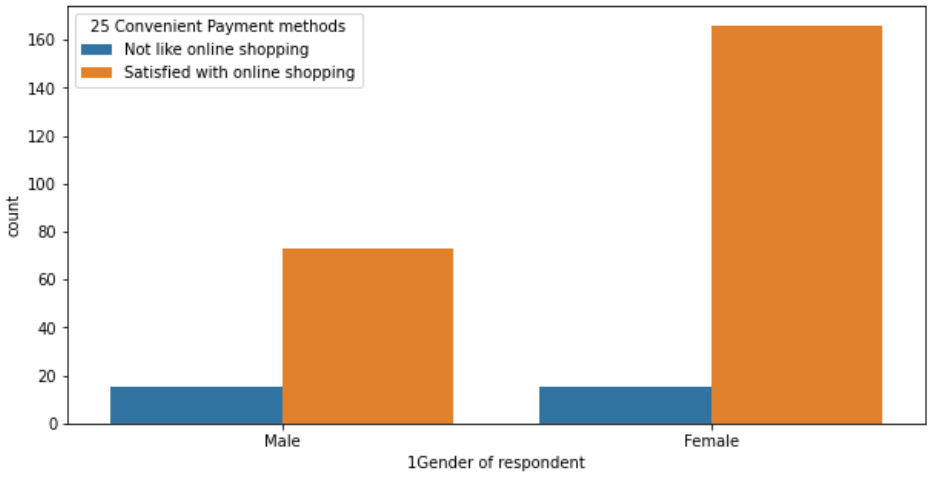
2) Most of the customers influenced strongly by 'Ease of navigation in website' are of age group 21-50 years

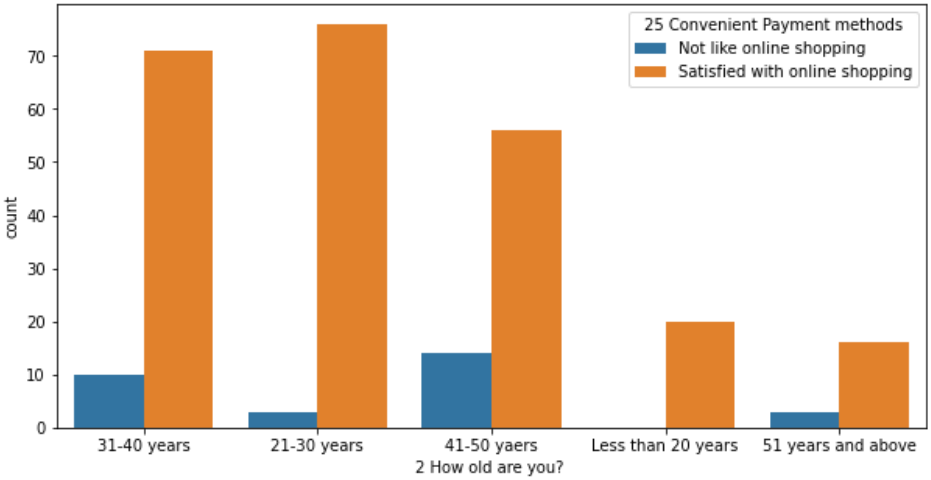


Inference:

1) Most of the customers influenced strongly by 'User friendly Interface of the website' are female

2) Most of the customers influenced strongly by 'User friendly Interface of the website' are age group 21-50 years

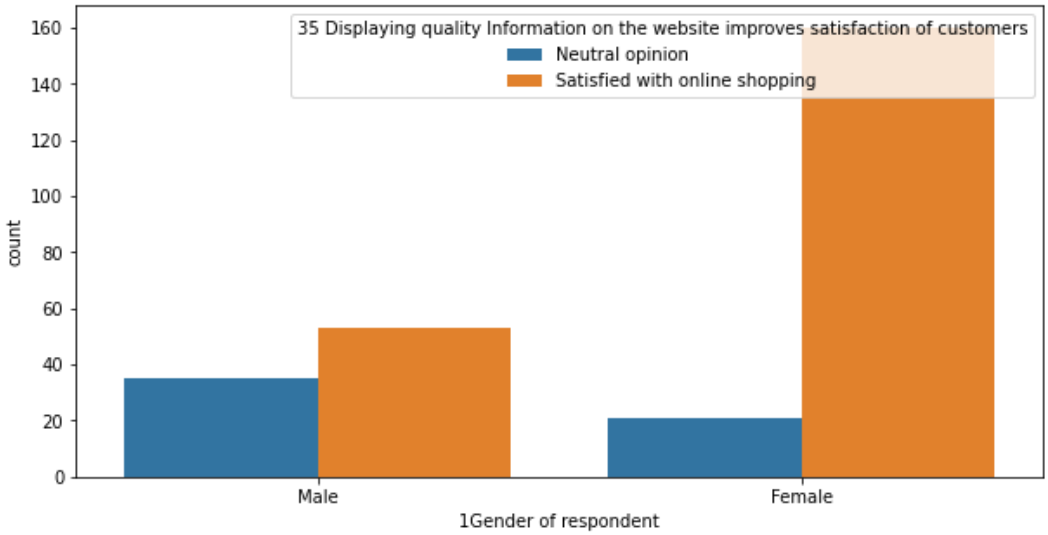


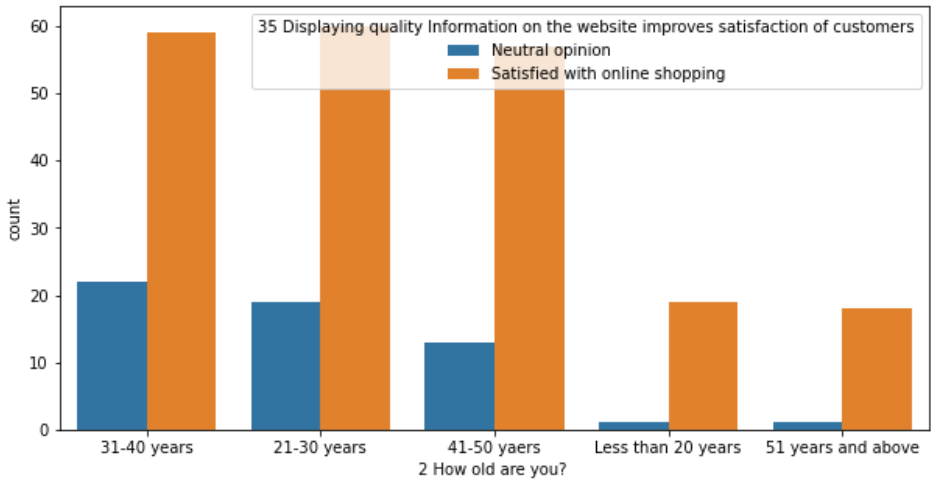


Inference:

1) Most of the customers influenced strongly by 'Convenient Payment methods' are female

2) Most of the customers influenced strongly by 'Convenient Payment methods' are age group 21-50 years

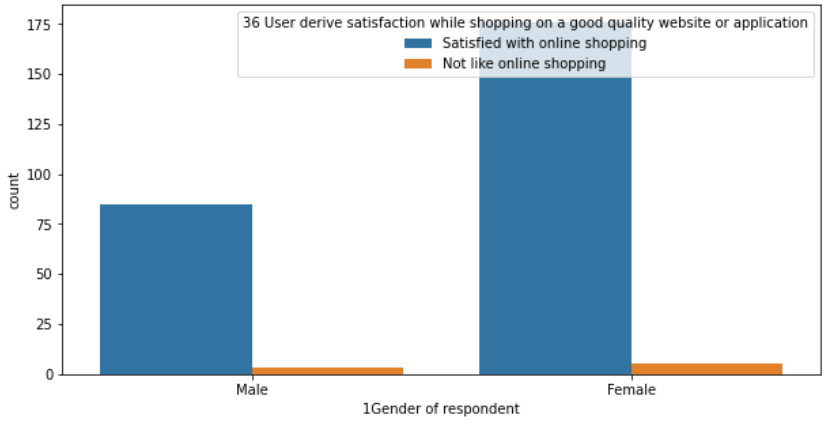


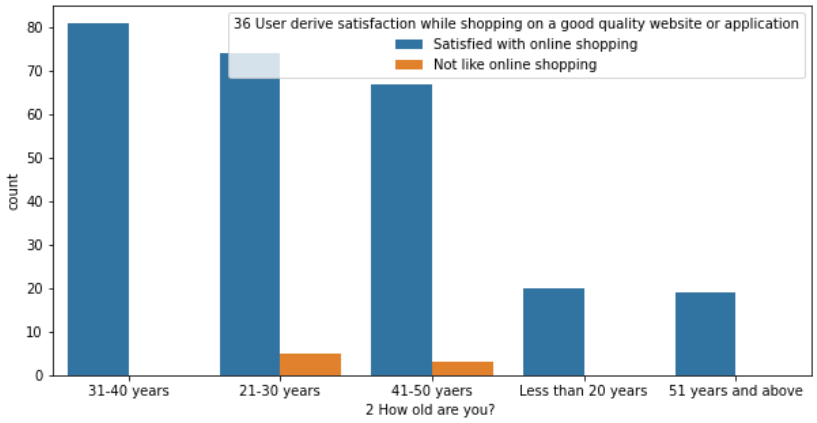


Inference:

1) Most of the customers influenced strongly by 'Displaying quality Information on the website improves satisfaction of customers' are female

2) Most of the customers influenced strongly by 'Displaying quality Information on the website improves satisfaction of customers' are age group 21-50 years

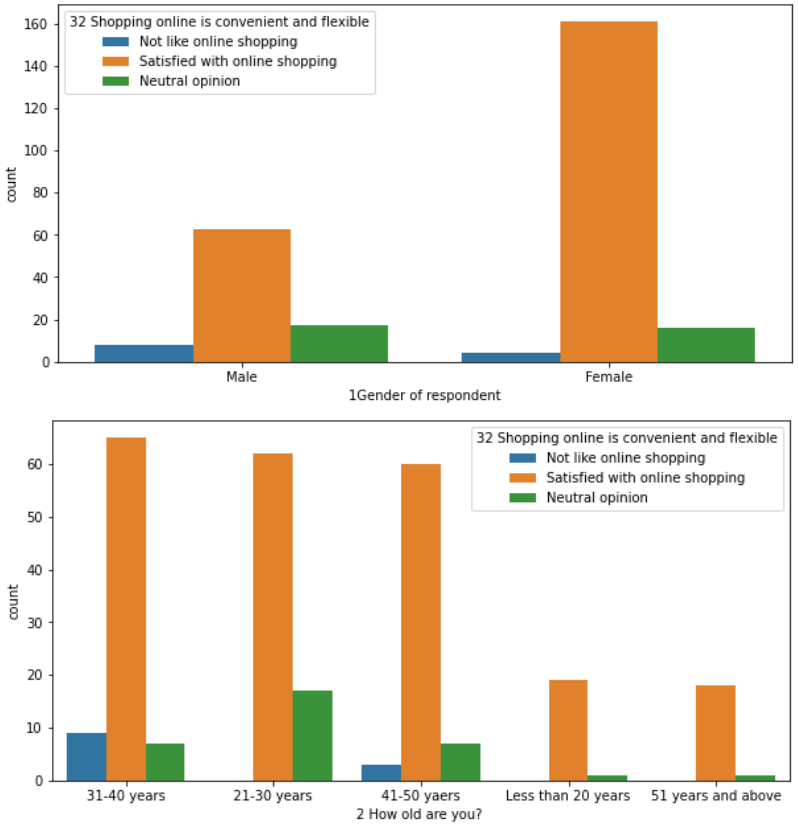




Inference:

1) Most of the customers influenced strongly by 'User derive satisfaction while shopping on a good quality website or application' are female

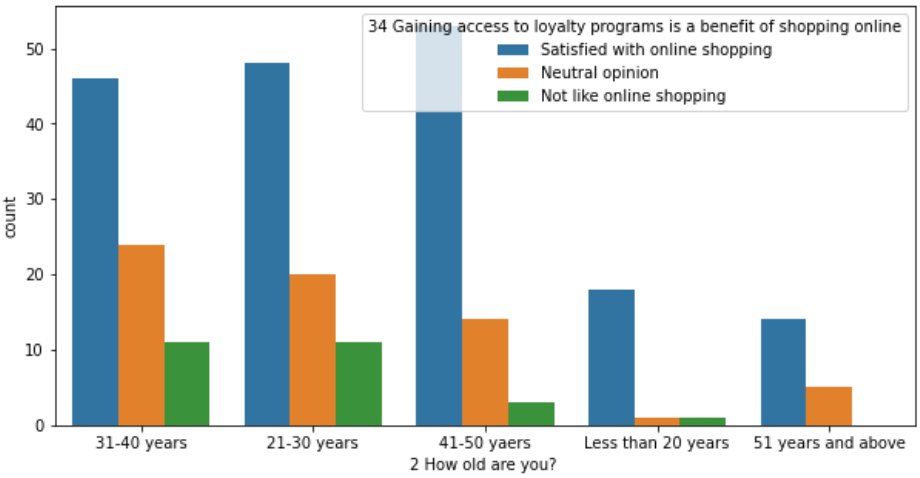
2) Most of the customers influenced strongly by 'User derive satisfaction while shopping on a good quality website or application' are age group 21-50 years



Inference:

1) Most of the customers influenced strongly by 'Shopping online is convenient and flexible' are female

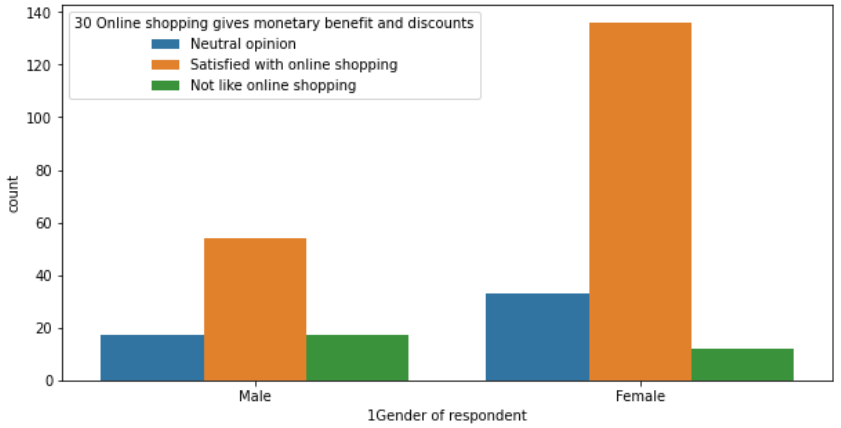
2) Most of the customers influenced strongly by 'Shopping online is convenient and flexible' are of age group 21-50 years

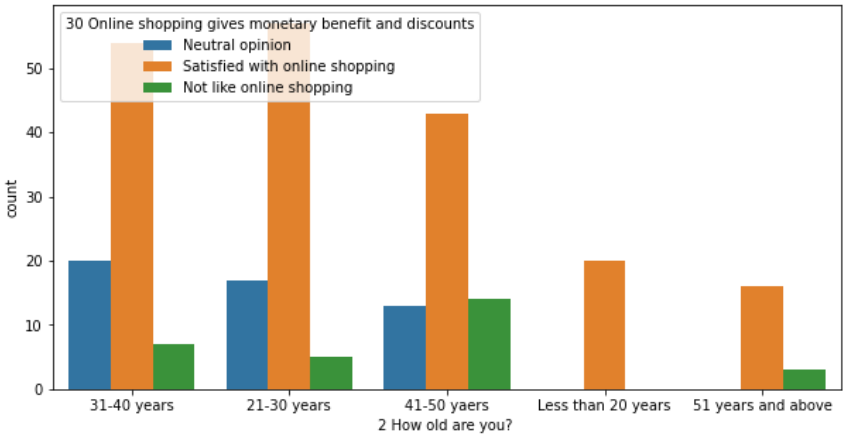


Inference:

1) Most of the customers influenced strongly by 'Gaining access to loyalty programs is a benefit of shopping online' are female

2) Most of the customers influenced strongly by 'Gaining access to loyalty programs is a benefit of shopping online' are of age group 21-50 years





Inference:

1) Most of the customers influenced strongly by 'Online shopping gives monetary benefit and discounts' are female

2) Most of the customers influenced strongly by 'Online shopping gives monetary benefit and discounts' are of age group 21-50 years

Overall:

We can infer that female under the age group of 21-50 years are agreeing/strongly agreeing for the considered factors influening the decesion for online shopping, hence it can be considered that they are satisfied for online shopping based on that particular factor is satisfying

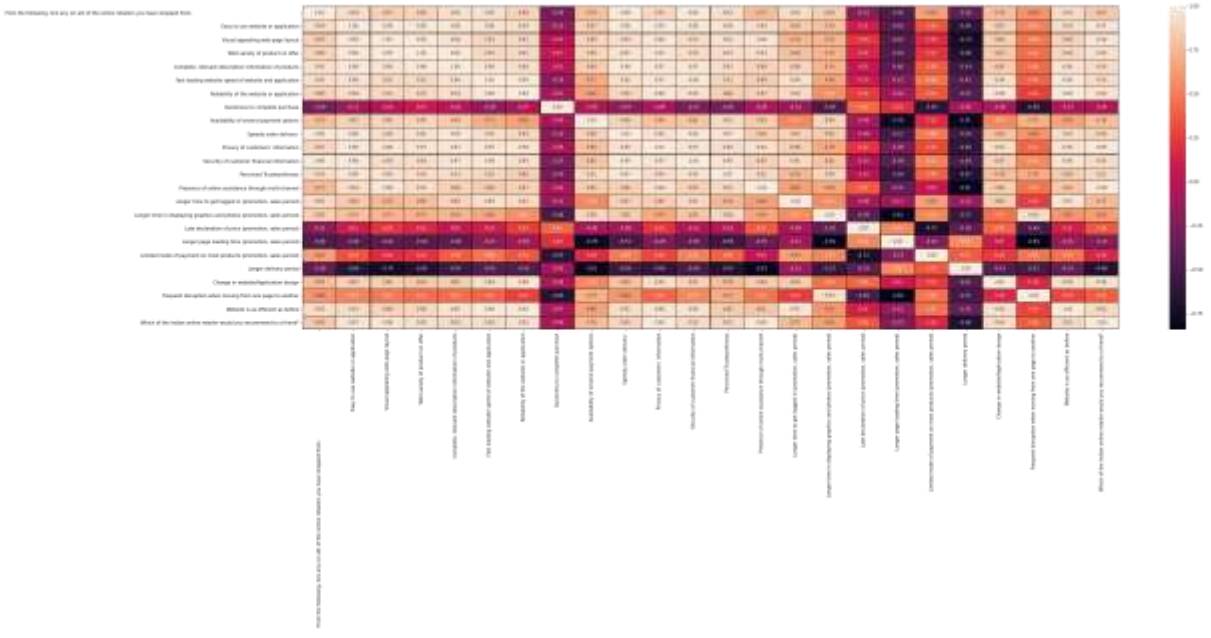
16) For remaining data from column no.48 onwards, regarding website or multiple website option chosen:

First we will unwrap the multidimensional list and form list containing single elements of chosen website as feedback data

```
data = pd.DataFrame()

for i in range(47,71):
    a=[]
    b=[]
    c=[]
    p=0
    column = list(crr.iloc[:,i])
    for ele in column:
        a.append(ele.split(','))
    for j in range(len(a)):
        for k in range(len(a[j])):
            b.append(a[j][k])
    for l in b:
        c.append(l.strip())
    m = pd.Series(c)
    m = m.value_counts(normalize=True)*100
    data[crr.iloc[:,i].name] = m

data = data.replace(np.nan,0)
#Snapdeal has nan values at some last column entries and some intermediate colum
```

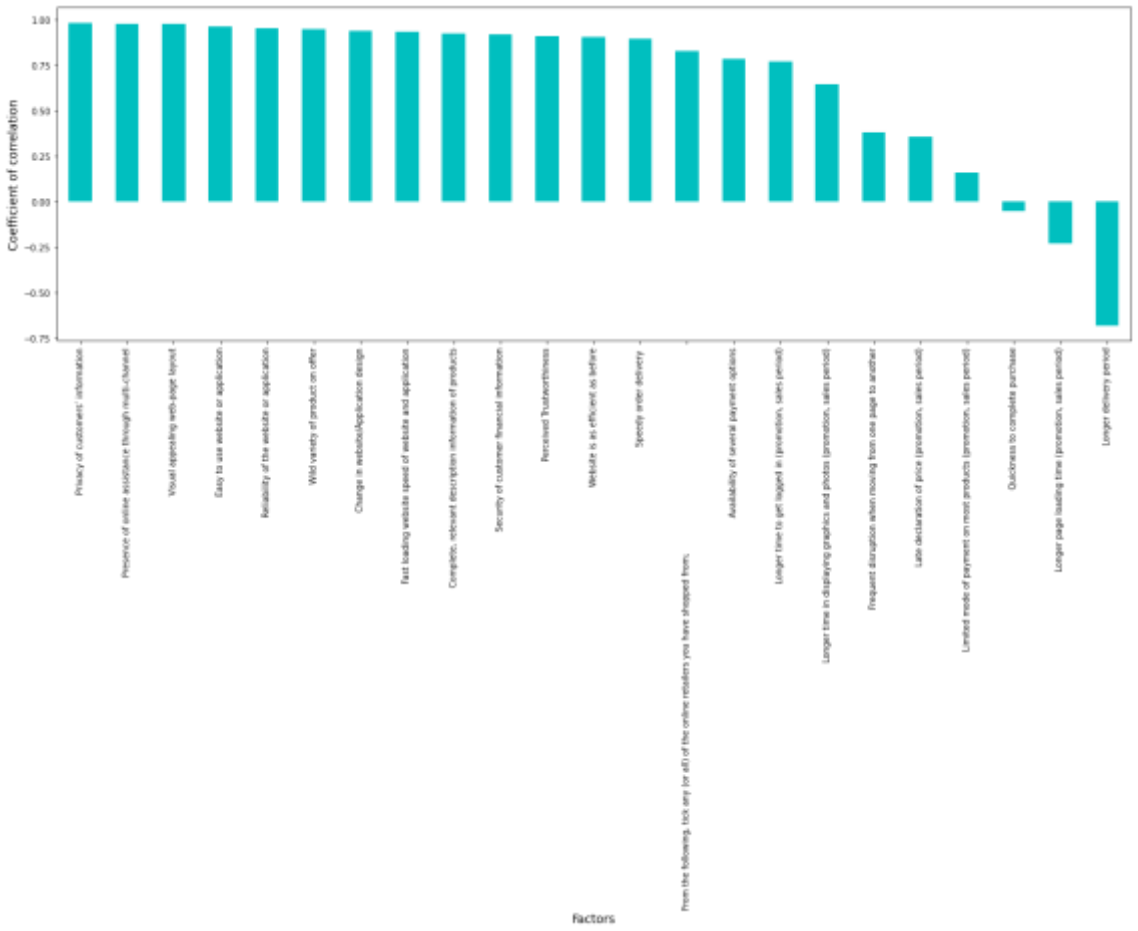
Since entries have multiple values or options selected, unwrapping and finding feedback chosen as whole will give fruitful results

Then doing correlation analysis and plotting heat-map



Since there are too many columns, lets prioritize and select factors based on highest value of coefficient of correlation with respect to target "Which of the Indian online retailer would you recommend to a friend?"

Visualization of descending order of coefficient of correlation



Privacy of customers' information 0.985095
Presence of online assistance through multi-channel 0.979067
Visual appealing web-page layout 0.977708
Easy to use website or application 0.965173
Reliability of the website or application 0.953425
Wild variety of product on offer 0.947364
Change in website/Application design 0.938144
Fast loading website speed of website and application 0.932608
Complete, relevant description information of products 0.926261
Security of customer financial information 0.921854

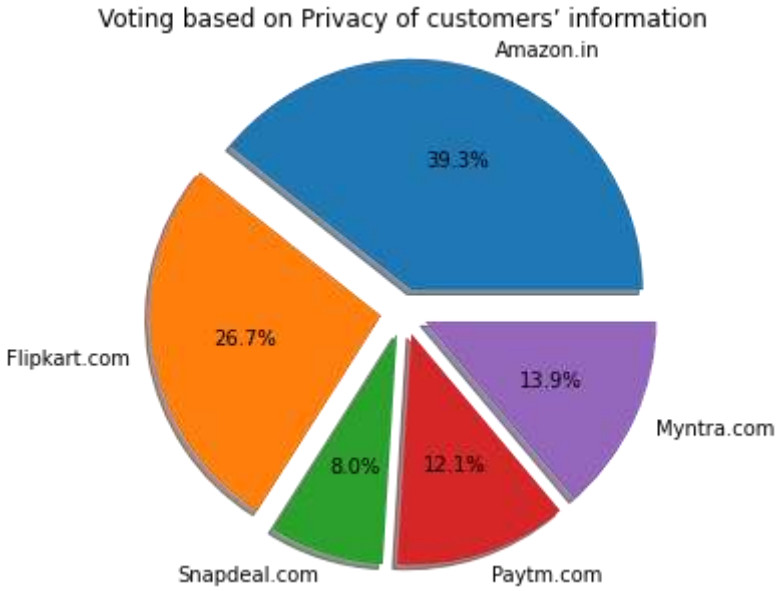
```
Perceived                                                  Trustworthiness
0.913052
Website          is          as          efficient          as          before
0.907795
Speedy                              order                              delivery
0.895488
From the following, tick any (or all) of the online retailers you have s
hopped                                                                  from;
0.828283
Availability          of          several          payment          options
0.783502
Longer    time    to    get    logged    in    (promotion,    sales    period)
0.773024
Longer time in displaying graphics and photos (promotion, sales period)
0.645129
Frequent    disruption    when    moving    from    one    page    to    another
0.381275
Late     declaration     of     price     (promotion,     sales     period)
0.358239
Limited mode of payment on most products (promotion, sales period)
0.158102
Quickness                to                complete                purchase
-0.055271
Longer     page     loading     time     (promotion,     sales     period)
-0.233774
Longer                              delivery                               period
-0.681979
```

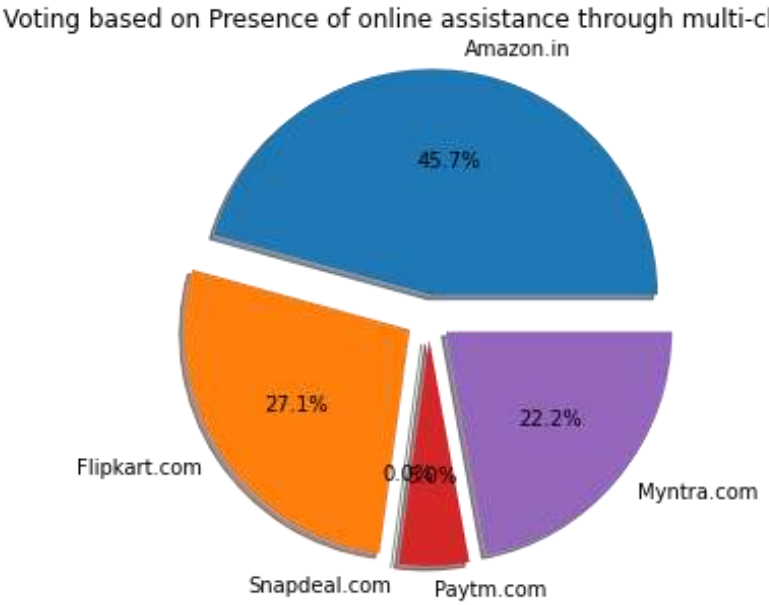
Inference:

1) High +ve correlation with target: 'Privacy of customers' information'

2) High -ve correlation with target: 'Longer delivery period'

Visualization of voting of customers for different online retailers

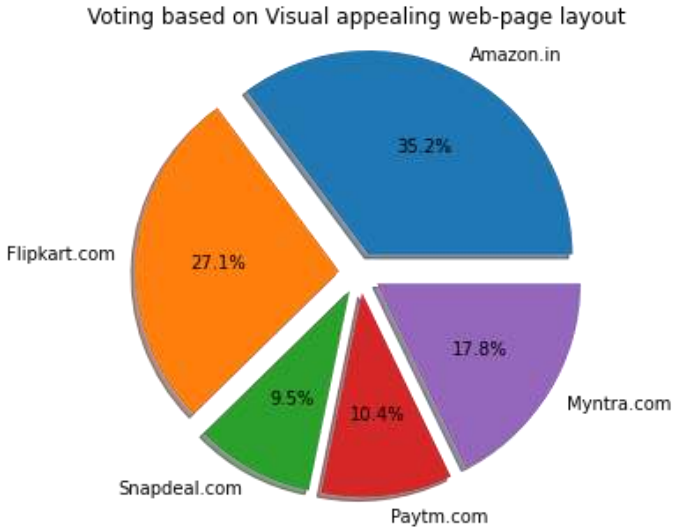


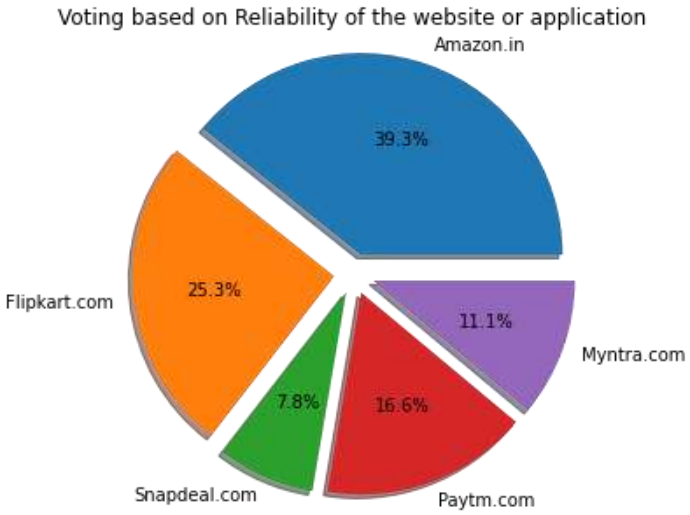


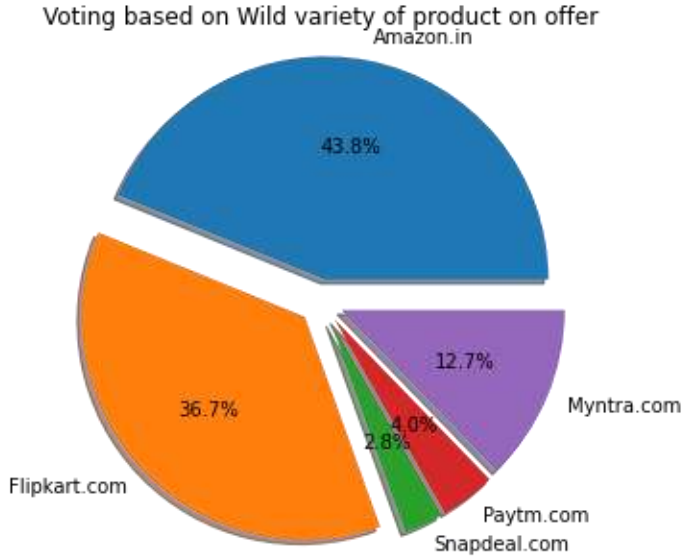
Inference:

1) Based on 'Privacy of customers' information' Amazon.in is preferred most for online shopping

2) Based on 'Presence of online assistance through multi-channel' Amazon.in is preferred most for online shopping





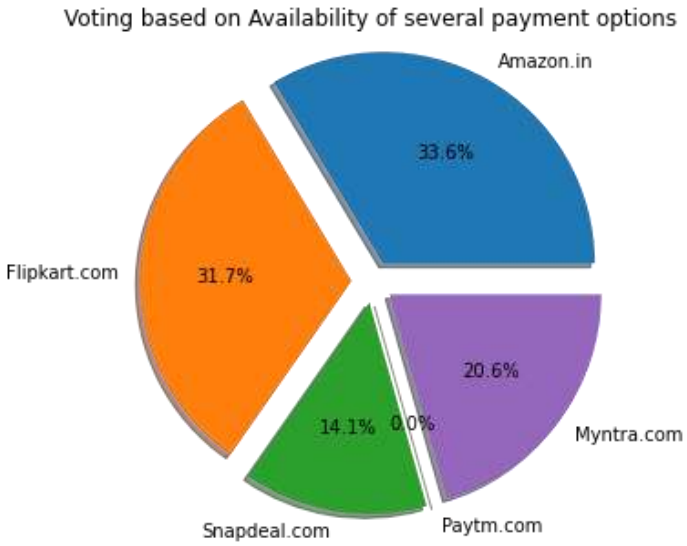


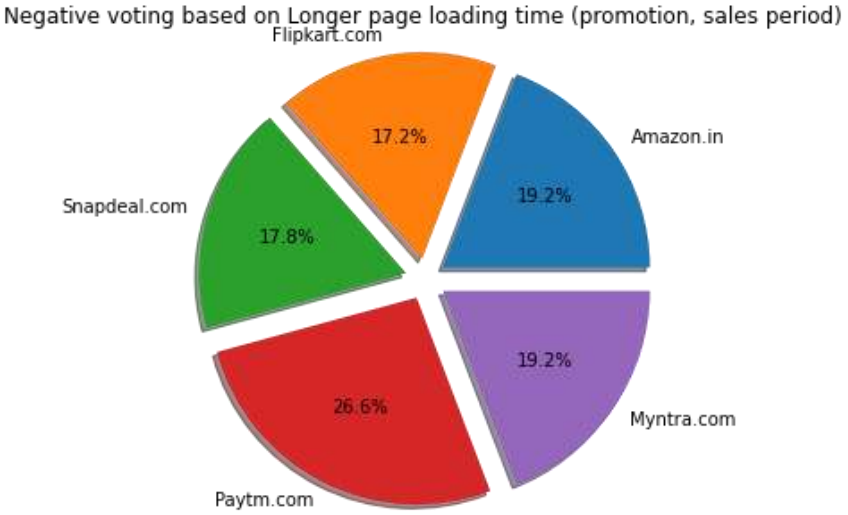
Inference:

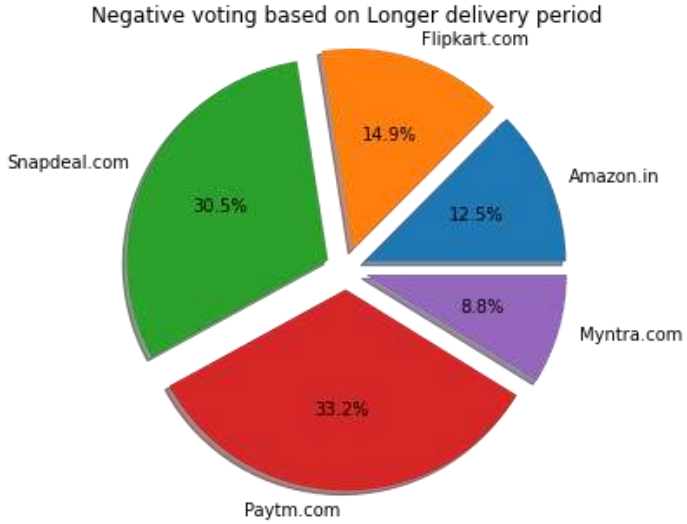
1) Based on 'Visual appealing web-page layout' Amazon.in is preferred most for online shopping

2) Based on 'Reliability of the website or application' Amazon.in is preferred most for online shopping

3) Based on 'Wild variety of product on offer' Amazon.in is preferred most for online shopping







Inference:

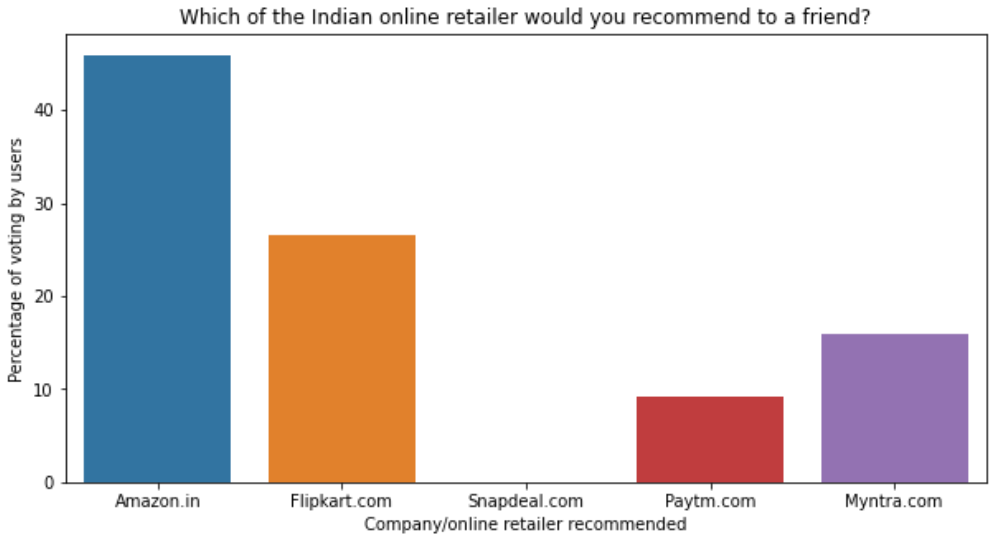
1) Based on 'Availability of several payment options' Amazon.in is preferred most for online shopping

2) Based on 'Longer page loading time (promotion, sales period)' Paytm.com is not preferred most for online shopping

3) Based on 'Longer delivery period' Paytm.com and Snapdeal.com are not preferred most for online shopping

Visualization of voting of customers for different online retailers:

Based on Company/online retailer recommendable to a friend



17) Conclusion:

This infers that majority of customers have voted for Amazon followed by Flipkart and Myntra as the retailer recommendable to a friend

This means they are also satsfied with their online shopping experience with that online retailer and have been retained as customers for future also