



Customer Retention Project

Improving customer loyalty through targeted retention strategies



SUBMITTED BY:
Vaibhav Rajesh Godawale

Introduction

-Business Problem Framing

This is a case study on Indian E-commerce customers to know which are the major factor for activation and retention of the customers, 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.

-Conceptual Background of Domain Problem

In this case we have a dataset in which first 19 columns tells about the customers, in next 30 columns we will get know how users are being satisfied with online e-retail shops in different areas like payment, speed of the delivery etc, In next 22 columns we will get to know which is the users favorite online retail shop under different categories.

-Review of literatures

The data for this project is been provided from the client side, apart from that to accomplish this project in better way went through several websites like geeksforgeeks, towards datascience and many more for effective visualization and also took the help of data trained institute.

Read books like data visualization using python, and did some research over stack overflow subject on the same topic.

-Motivation for problem undertaken

The particular project is undertaken to advice our client reagrding the factorsthat affect most for the activation and retention of the customers, Adding to that we would always like to see our client at top of chart in this reckless competition.

Analytical Problem Framing

-Mathematical/Analytical Modeling of the problem

Since all most all columns are of type categorical we label encoded them to make the machine understand, As there was no response variable so we looked at the correlation between 2 independent variable if there is any significant correlation between those then we analyse that pair using suitable technique to derive maximum cooperative information.

-Data sources and their formats

The data what we got is completely of type categorical, only pincode was of integer type but we saw it as a category as it infers the location of a particular place.

- Hardware and Software Requirements and Tools Used

We carried out this project in jupyter notebook of Anaconda Navigator For better visualization and for data processing we used libraries like pandas, matplotlib, seaborn and sklearn.

-Project Flow

The project flow is comparatively small compared to other projects because only analysis part is there.

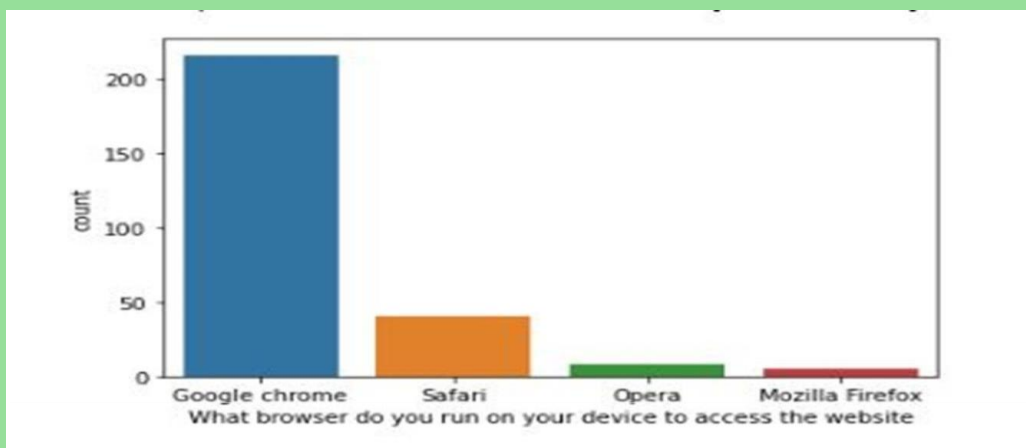
1. Importing the dataset
2. Data processing
3. Exploratory Data Analysis
 - A. Univariate Analysis
 - B. Bi variate analysis
 - C. Multivariate analysis

Informations Derived from Exploratory

Data Analysis

Univariate Analysis (Information regarding users)

1. Out of total users 67% were females.
2. Most participating age group in online retail shopping is from 30-50 years.
3. Most of the users in this dataset is from Delhi followed by Noida and bangalore.
4. In this dataset we have majority of users who are shopping through online from past 4 years and above followed by 2-3 years.
5. Around 44 % of users had ordered less than 10 times in past one year followed by 31- 40 times which consist of 22% user.
6. Around 67% of user use mobile internet for online shopping followed by WiFi which consist of 22%
7. Regarding the operating system of users around 45% of user has got windows/windows mobile followed by android and IOS/Mac of 31% and 24% respectively.
8. Around 80% of user use Google Chrome as their web browser followed by safari, opera and Mozilla firefox of 15%, 3% and 2% respectively.

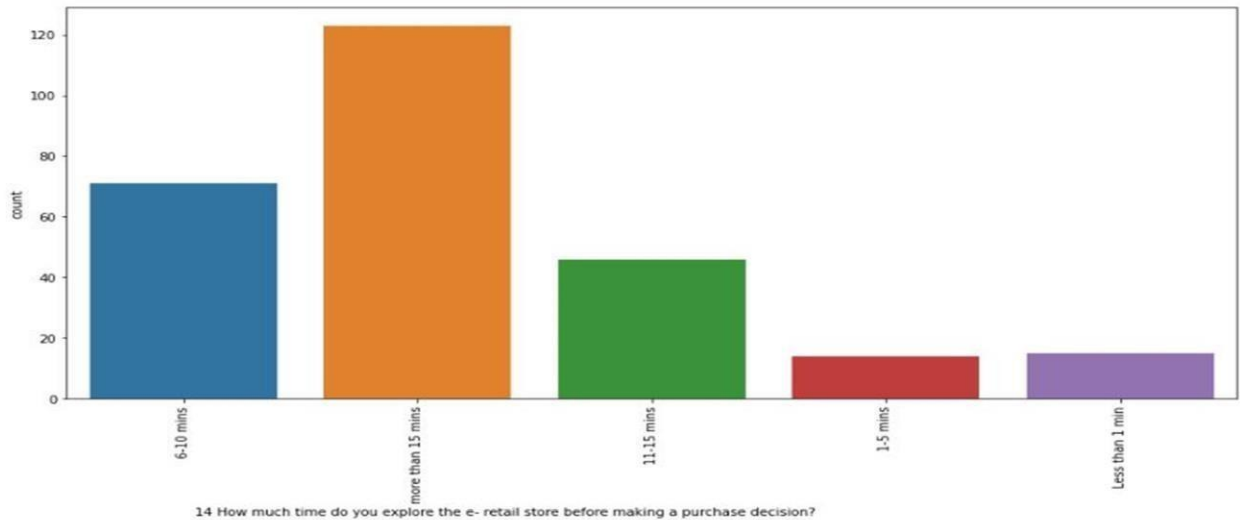


9. Out of 269 users around 85 users again visited by search engine, around 85 again visited through applications and around 70 have visited through direct url.

Column 14: How much time do you explore the e-retail store before making a decision

-Around 44% of user spend more than 15 min before they make up the decision of purchasing

-Around 24% of user spend 6-10 minutes before making a decision of purchasing

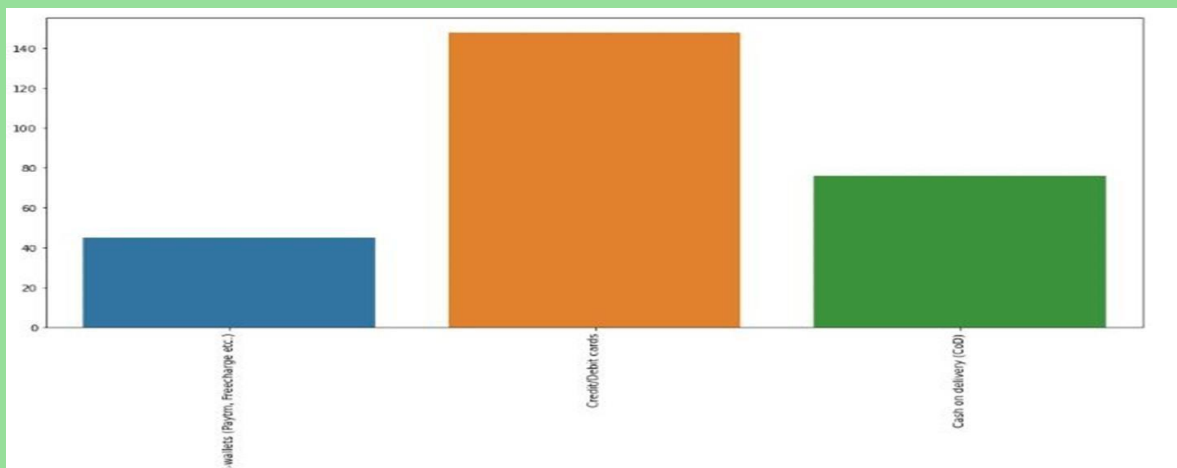


column 15: what is your preferred payment option?

-Around 55% of user prefer credit/debit cards

-Around 29% of user prefer cash on delivery

-Around 16% of user prefer e-wallets



column 16: How frequently do you abandon (selecting an items and leaving without making payment) your shopping cart?

-Around 63% of user abandoned sometimes without making an payment

-Around 18% of user never abandoned without making an payment

-Around 13% of user frequently abandoned without making an payment

column 17: Why did you abandon the “Bag”, “Shopping Cart”

-Around 48% of user abandoned the cart because of better alternative offer

-Around 20% of user abandoned the cart because of promo code not applicable

-Around 14% of user abandoned the cart because of change in price

-Around 11% of user abandoned the cart because of lack of trust

-Around 7% of user abandoned the cart because of no preferred mode of payment

column 18: The content on the website must be easy to read and understand

-Around 59% of the user strongly agree that the content on the website is easy to read and understand

-Around 29% of the user agree that the content on the website is easy to read and understand

-Around 7% of the user strongly disagree with the content on the website

Users opinion on online retail shops

-Around 5% says that the contents are indifferent

column 19: Information on similar product to the one highlighted is important for product comparison

-Around 42% strongly agree that information on similar products to the one highlighted is important for product comparison.

-Around 35% agree that information on similar products to the one highlighted is important for product comparison.

-Around 15% have opinion that information on similar products to the one highlighted is important for product comparison is indifferent.

-Around 8% disagree that information on similar products to the one highlighted is important for product comparison.

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

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

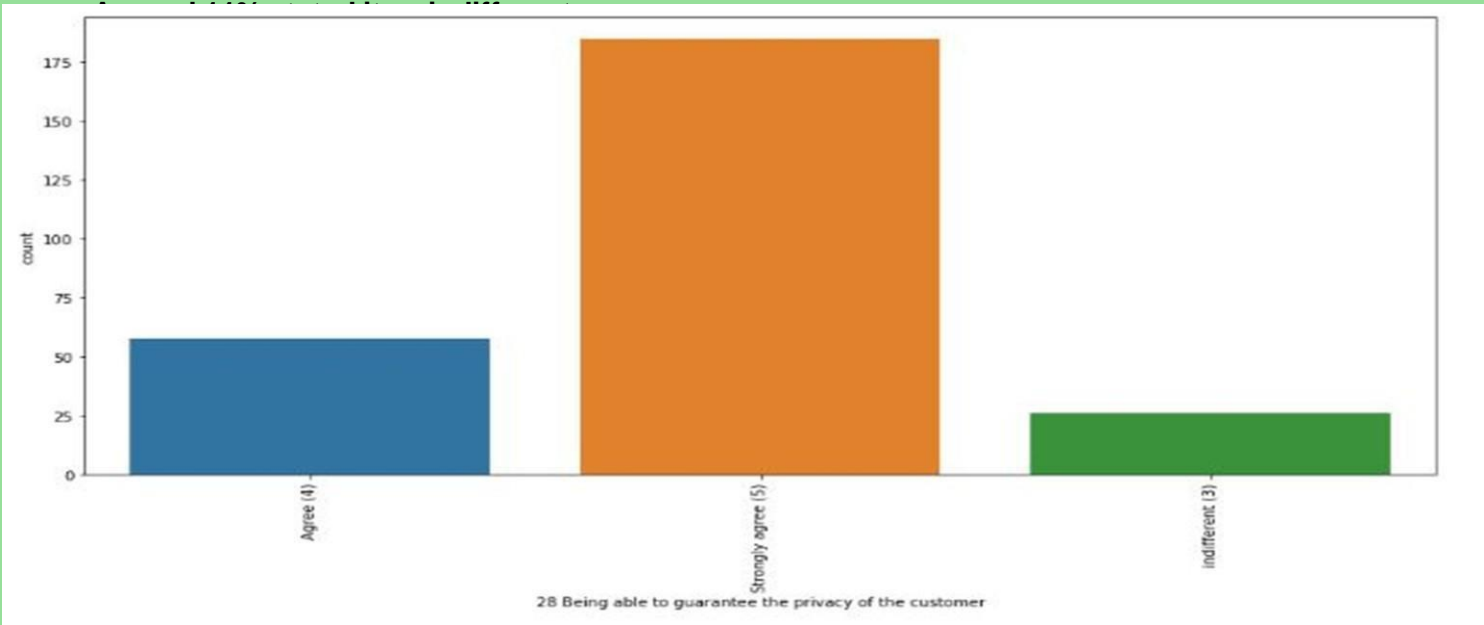
- Around 52% of user strongly agree that they have trust in the online retail store will fulfill its part of the transaction at the stipulated time
- Around 31% of user agree that they have trust in the online retail store will fulfill its part of the transaction at the stipulated time
- Around 11% of the user disagree with the fact that online retail store will fulfill its part of the transaction at the stipulated time.
- Around 6% of user stated that its indifferent.

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

- Around 71% of user strongly agree with the fact that the online retail stores are always ready to assist with queries of customers.
- Around 15% of user agree with the fact that the online retail stores are always ready to assist with queries of customers.
- Around 8% of user disagree with the fact that the online retail stores are always ready to assist with queries of customers.
- Around 6% of user stated that this is indifferent.

column 28: Being able to guarantee the privacy of the customer

- Around 68% of user strongly agree with the fact that online retail store will guarantee their privacy
- Around 21% of user agree that online retail store will guarantee their privacy



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

- Around 55% of user strongly agree that there are several communication channels
- Around 35% of user agree that there are several communication channels
- Around 6% of user agree that its indifferent
- Around 4% of user strongly disagree that there are several communicationchannels

column 30: Online shopping gives monetary benefit and discounts

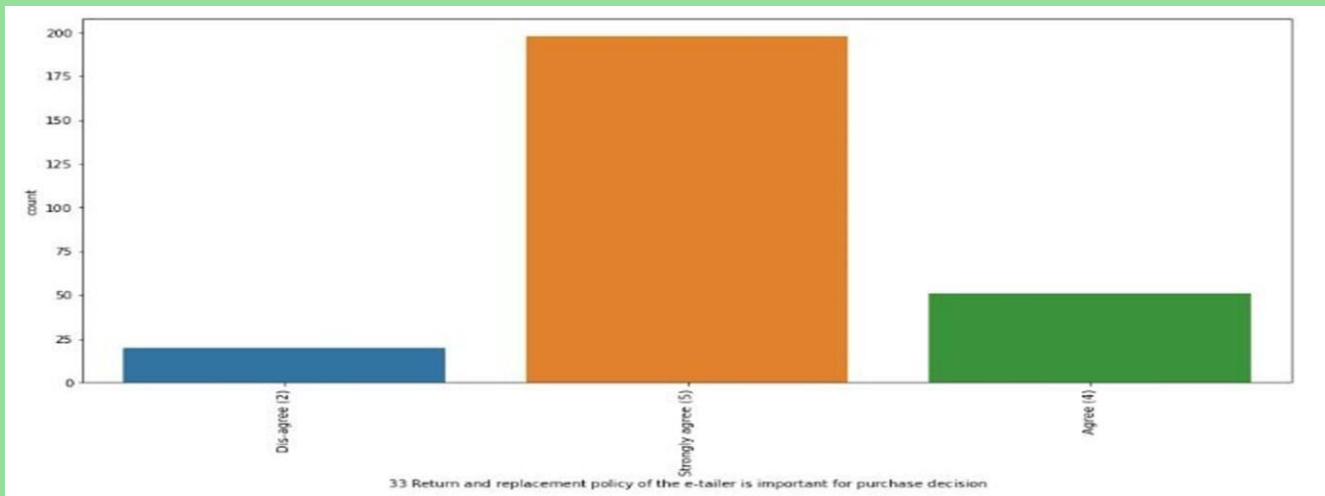
- Around 39% of user strongly agree that online shopping gives monetary benefits
- Around 31% of user agree that online shopping gives monetary benefits
- Around 18% stated that its indifferent
- Around 7% strongly disagree with fact that online shopping gives monetary benefits
- Around 5% of user disagree with the fact that online shopping gives moentorybenefits column

31: Enjoyment is derived from shopping online

- Around 32% of user strongly agree with the fact that they enjoy online shopping
- Around 29% of user stated it as indifferent
- Around 22% of user agree with the fact that they enjoy online shopping
- Around 11% of user strongly disagree with the fact that they enjoy online shopping
- Around 6% of user disagree with the fact that they enjoy online shopping column 32:

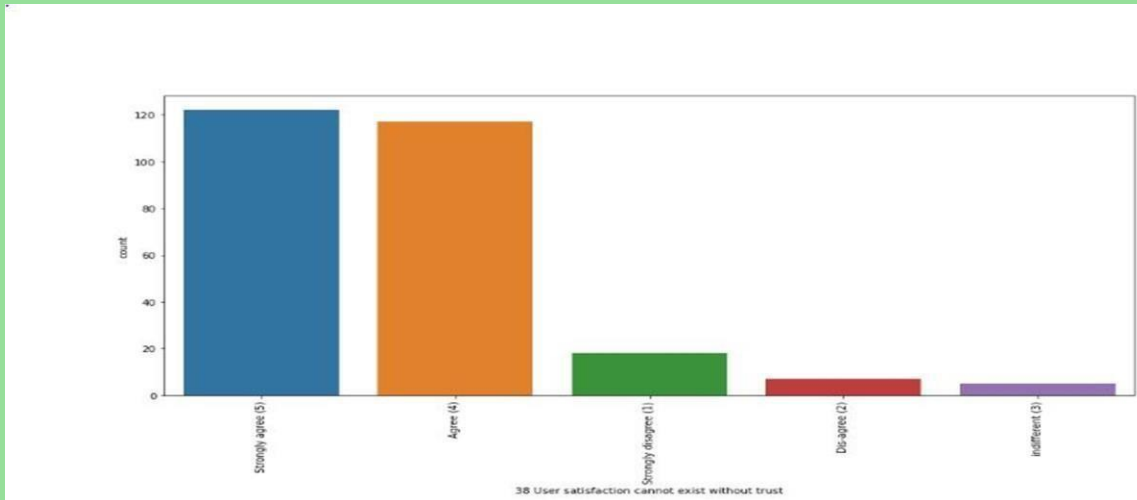
Return and replacement policy of the e-tailer is important for purchase decision

- Around 74% of user strongly agree that replacement policy is important for purchase
- Around 19% of user agree that replacement policy is important for purchase



column 38: User satisfaction cannot exist without trust

- Around 45% of user strongly agree that user satisfaction cannot exist without trust
- Around 43% of user agree that user satisfaction cannot exist without trust
- Around 6% of user strongly disagree that user satisfaction cannot exist without trust
- Around 3% of user disagree that user satisfaction cannot exist without trust



column 39: Offering a wide variety of listed product in several category

- Around 42% of user strongly agree with the fact that online retail store offers a wide variety of listed products in several category
- Around 35% of user agree with the fact that online retail store offers a wide variety of listed products in several category
- Around 2% of user disagree with the fact that online retail store offers a wide variety of listed products in several category
- Around 21% of user stated it as indifferent

column 40: Provision of complete and relevant product information

- Around 50% of user strongly agree that complete and relevant product information is available.
- Around 36% of user agree that complete and relevant product information is available.
- Around 12% of user stated it as indifferent
- Around 2% of user disagree that complete and relevant product information is available.

column 41: From the following, tick any (or all) of the online retailers you have

shopped from

- Everyone are familiar with amazon and flipkart.

Column 56: In which website the purchasing will be quick?

According to user amazon is so quick in terms of completing the purchase followed by flipkart paytm and myntra.

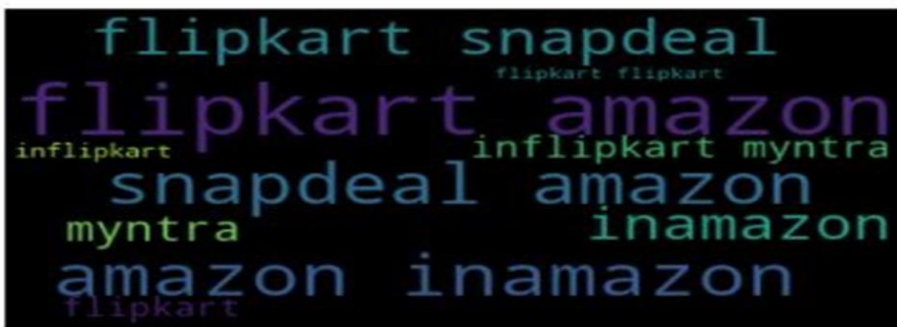
Column 57: Availability of several payment options?

According the user amazon has got several payment options followed by flipkart and myntra.

Column 58: Which website has got speedy delivery of products?

Amazon tops the chart followed by flipkart and snapdeal.

```
#column 57:Speedy order delivery
list1=[]
for i in df['Speedy order delivery ']:
    list1.append(i)
k=''.join(list1)
k1 = k.replace('com', '')
k1=k1.lower()
k1=word_tokenize(k1)
k1=''.join(k1)
wordcloud = WordCloud(max_font_size=50, max_words=100, background_color="black").generate(k1)
plt.figure(figsize=(9,5))
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis("off")
plt.show()
```



```
df['Speedy order delivery '].value_counts()

Amazon.in                                107
Amazon.in, Flipkart.com                   82
Amazon.in, Flipkart.com, Snapdeal.com     36
Flipkart.com                             15
Amazon.in, Flipkart.com, Myntra.com       15
Flipkart.com, Myntra.com, Snapdeal.com    14
Name: Speedy order delivery , dtype: int64
```

Column 59: Which web application takes most care regarding the privacy of customer's information?

In case of customers privacy amazon tops the chart followed by flipkart and paytm

Column 60: Which web application takes most care regarding the customer's financial

information?

In case of users financial privacy amazon tops the chart followed by flipkart and snapdeal

Column 61: Presence of online assistance through multiple channels?

In case of Presence of online assistance through multi-channel amazon tops the chart followed by flipkart and snapdeal

Column 62: Which website takes more time to get logged in?

Amazon tops the chart followed by flipkart in case of time taking to logged in.

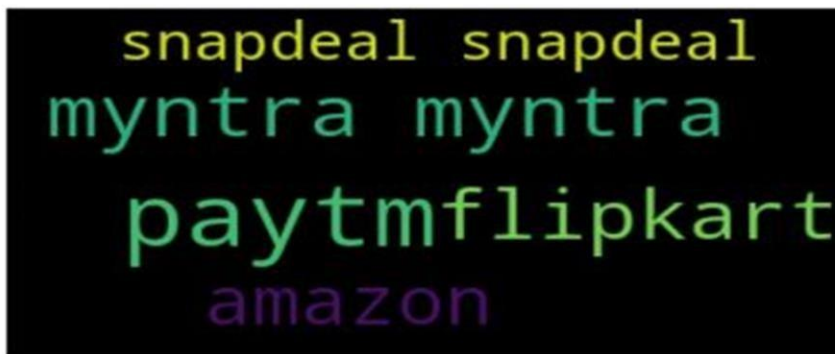
Column 63: Which website takes longer time in displaying the graphics and photos?

In case of time taking for displaying thee pictures Amazon takes a lot of time followed by flipkart and snapdeal

Column 64: Which website declares price of products in late?

In case of late declaration of price mynthra tops the chart followed by paytm and flipkart

```
#column 64:Late declaration of price (promotion, sales period)
list1=[]
for i in df['Late declaration of price (promotion, sales period)']:
    list1.append(i)
k=''.join(list1)
k1 = k.replace('com', '')
k1=k1.lower()
k1=word_tokenize(k1)
k1=''.join(k1)
wordcloud = WordCloud(max_font_size=50, max_words=5, background_color="black").generate(k1)
plt.figure(figsize=(9,5))
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis("off")
plt.show()
```



```
df['Late declaration of price (promotion, sales period)'].value_counts()

Myntra.com          75
Paytm.com           52
snapdeal.com        41
Flipkart.com        38
Amazon.in           38
Amazon.in, Paytm.com 13
Paytm.com, snapdeal.com 7
Amazon.in, Flipkart.com 5
Name: Late declaration of price (promotion, sales period), dtype: int64
```

Column 65: Change in website/application design

Amazon tops the chart followed by flipkart, paytm in case on Change in website/Application design

Column 66: Website is as efficient as before

Amazon tops the chart followed by flipkart and paytm in retaining the efficiency of website or company as before

Column 67: Which of the Indian online retailer would you recommend to a friend?

Majority of the user refer amazon followed by flipkart and mynthra.

Column 68: Which website takes longer time to load the page?

In case of longer page loading paytm tops the chart followed by myntraspnapdeal,flipkart

```
#column 65:Longer page loading time (promotion, sales period)
list1=[]
for i in df['Longer page loading time (promotion, sales period)']:
    list1.append(i)
k=''.join(list1)
k1 = k.replace('com', '')
k1=k1.lower()
k1=word_tokenize(k1)
k1=''.join(k1)
wordcloud = WordCloud(max_font_size=50, max_words=7, background_color="black").generate(k1)
plt.figure(figsize=(9,5))
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis("off")
plt.show()
```



```
df['Longer page loading time (promotion, sales period)'].value_counts()
```

Myntra.com	61
Paytm.com	59
Flipkart.com	32
Snapdeal.com	23
Amazon.in, Flipkart.com	18
Amazon.in	16
Paytm.com, Snapdeal.com	15
Amazon.in, Snapdeal.com	14
Amazon.in, Paytm.com	13
Flipkart.com, Snapdeal.com	11
Amazon.in, Paytm.com, Myntra.com	7

Column 69: Limited mode of payment on most products (promotion, sales period)

In case of limited mode of payment on most of the products snapdeal tops the chart followed by flipkart and amazon.

```
df['Limited mode of payment on most products (promotion, sales period)'].value_counts()
Snapdeal.com      87
Amazon.in         62
Flipkart.com      31
Amazon.in, Flipkart.com  29
Paytm.com         25
Paytm.com, Snapdeal.com  15
Amazon.in, Paytm.com  13
Myntra.com, Snapdeal.com  7
Name: Limited mode of payment on most products (promotion, sales period), dtype: int64
```



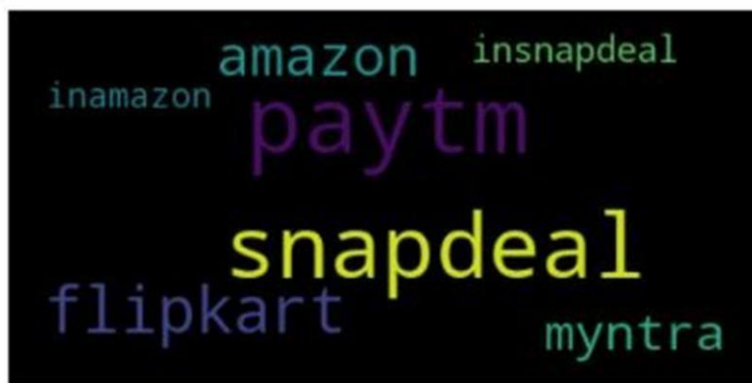
```
#column 66:Limited mode of payment on most products (promotion, sales period)
list1=[]
for i in df['Limited mode of payment on most products (promotion, sales period)']:
    list1.append(i)
k=''.join(list1)
k1 = k.replace('com', '')
k1=k1.lower()
k1=word_tokenize(k1)
k1=''.join(k1)
wordcloud = WordCloud(max_font_size=50, max_words=7, background_color="black").generate(k1)
plt.figure(figsize=(9,5))
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis("off")
plt.show()
```



Column 70: Which website has longer delivery period?

In case of longer delivery period snapdeal tops the chart followed by paytm and flipkart.

```
: #column 67:Longer delivery period
list1=[]
for i in df['Longer delivery period']:
    list1.append(i)
k=''.join(list1)
k1 = k.replace('com', '')
k1=k1.lower()
k1=word_tokenize(k1)
k1=''.join(k1)
wordcloud = WordCloud(max_font_size=50, max_words=7, background_color="black").generate(k1)
plt.figure(figsize=(9,5))
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis("off")
plt.show()
```

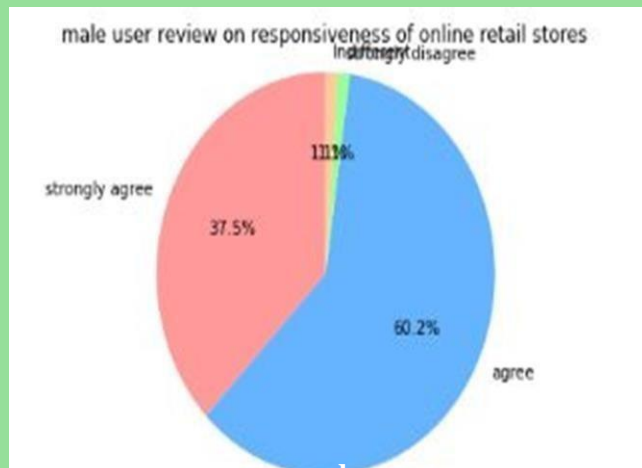
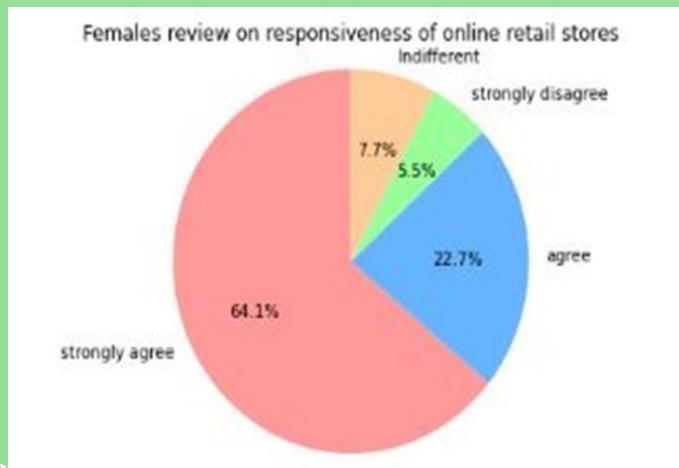


```
: df['Longer delivery period'].value_counts()
: Paytm.com 72
: Snapdeal.com 64
: Flipkart.com 44
: Amazon.in 37
: Paytm.com, Snapdeal.com 26
: Myntra.com 26
Name: Longer delivery period, dtype: int64
```

Results of Multivariate and Bi variate analysis

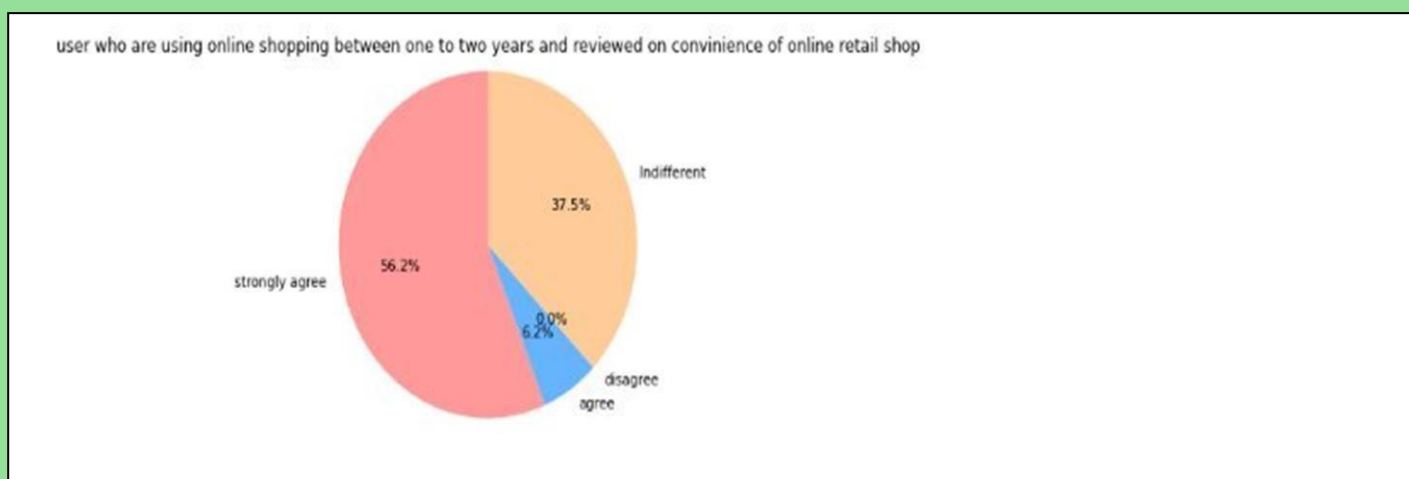
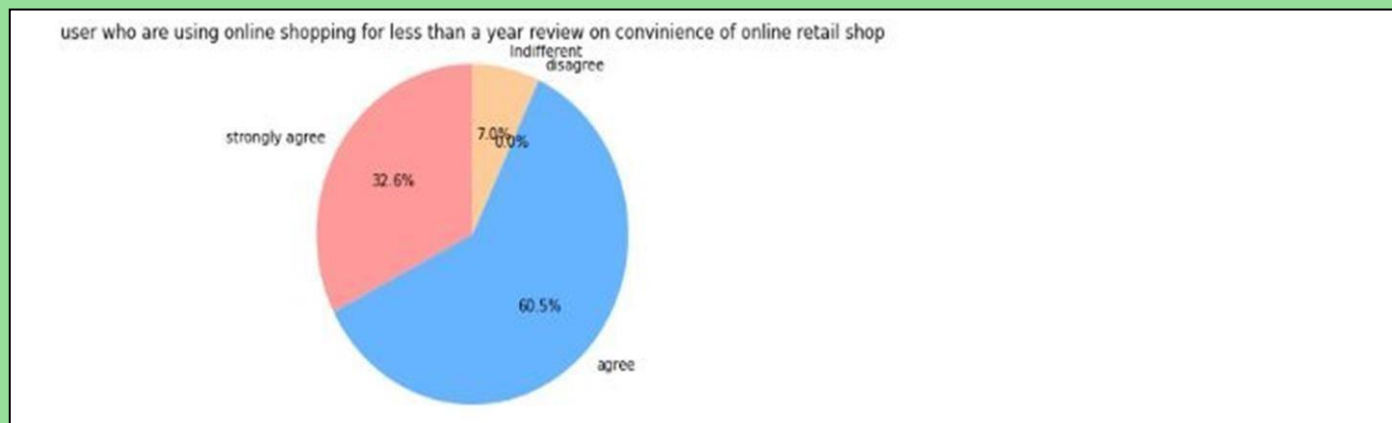
Gender has got significant correlation with the following:

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

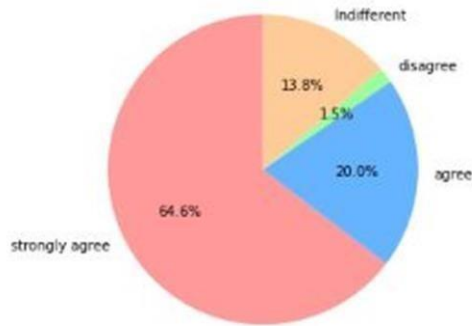


Both male and female users are equally agree that online retail stores are responsive and graph shows that if there is any disagreement the probability that it's from female user is more compared to male user.

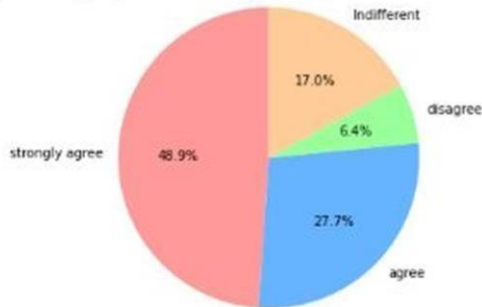
2. Users remark on the convenience of the online retail shop based on their experience in online shopping.



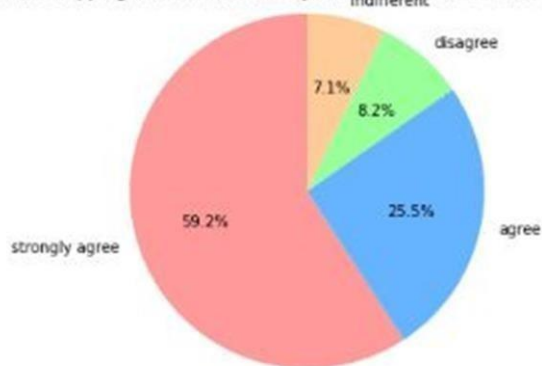
user who are using online shopping for two to three years and reviewed on convinience of online retail shop



user who are using online shopping for three to four years and reviewed on convinience of online retail shop



user who are using online shopping for four and above years and reviewed on convinience of online retail shop

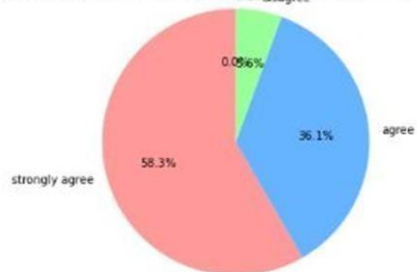


In the long run on online retail store as customers started to having more and more years of experience through online retail store the category of strongly agree will remain almost constant butdisagreeeness will slightly increase.

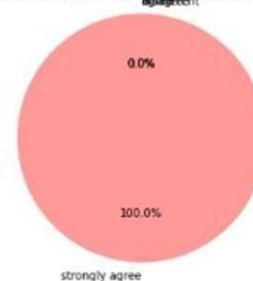
3. Customers enjoys the shopping when they derive the internet from wifi or mobile.

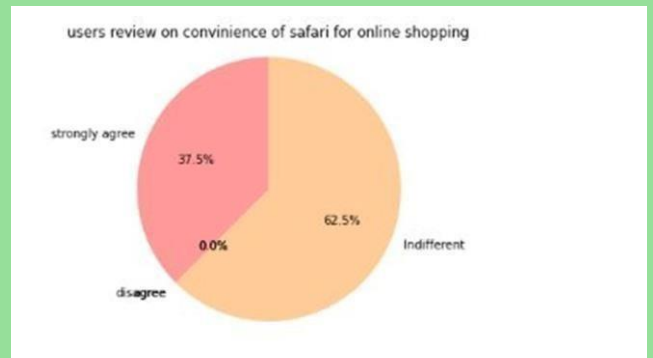
4. Browser we gonna use has significant affect over the flexibility and convinience of online shopping

users review on convinience of google chrome for online shopping



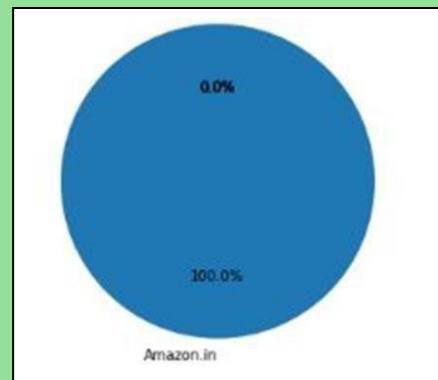
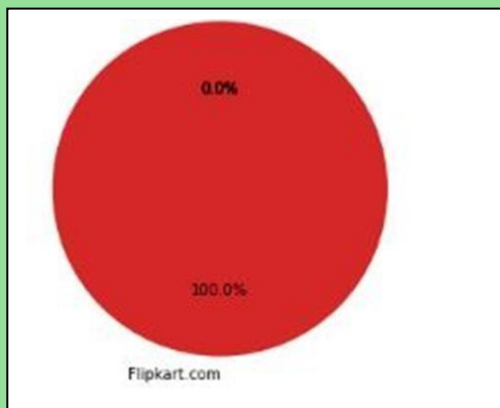
users review on convinience of mozilla firefox for online shopping



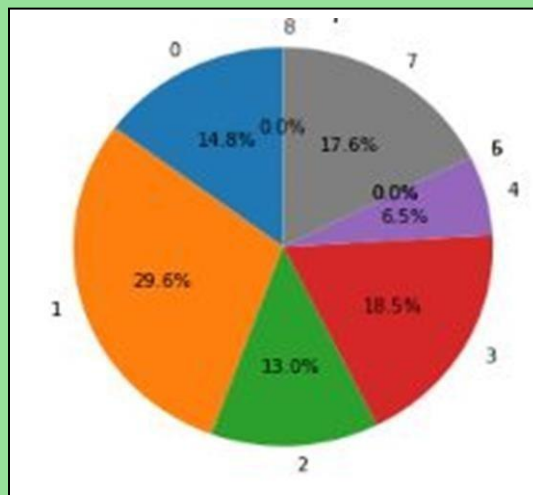


In terms of convinience mozilla firefox tops the chart followed by chrome , safari and opera

5. There is relationship between how quickly they purchase and in which website they are purchasing, Inthis dataset those who completes purchase within one minute belongs to flipkart and those who completes within one to five minutes belongs to amazon.



6. The significant factor for higher sales is cash on delivery, why amazon is leading because it provides more cash on delivery options compared to other web applications.



7. There are situations or instances where user abandon the shopping without purchasing, one of the factor which is closely related to this according to this dataset is which web browser they use.

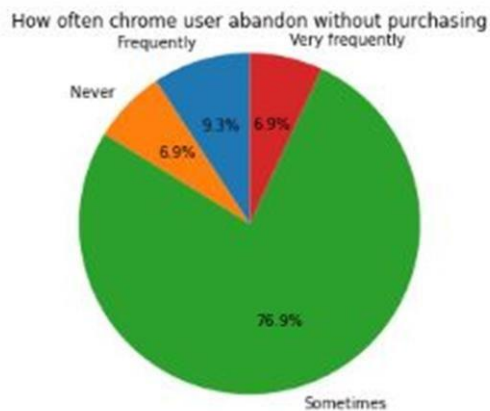
```

In [87]: #Lets see how google chrome user abandon the shopping without purchasing
k2=df1[(df1[df1.columns[15]]==0) & (df1[df1.columns[10]]==0)].count()
k3=df1[(df1[df1.columns[15]]==1) & (df1[df1.columns[10]]==0)].count()
k4=df1[(df1[df1.columns[15]]==2) & (df1[df1.columns[10]]==0)].count()
k5=df1[(df1[df1.columns[15]]==3) & (df1[df1.columns[10]]==0)].count()

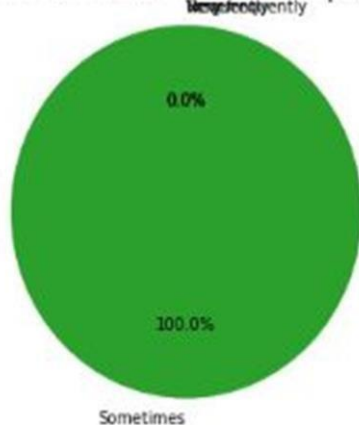
labels = ['Frequently','Never','Sometimes','Very frequently']
sizes = [k2[0],k3[0],k4[0],k5[0]]

fig1, ax1 = plt.subplots()
patches, texts, autotexts = ax1.pie(sizes,labels=labels, autopct='%1.1f%%',startangle=90)
for text in texts:
    text.set_color('black')
for autotext in autotexts:
    autotext.set_color('black')
# Equal aspect ratio ensures that pie is drawn as a circle
ax1.axis('equal')
plt.title('How often chrome user abandon without purchasing')
plt.tight_layout()
plt.show()

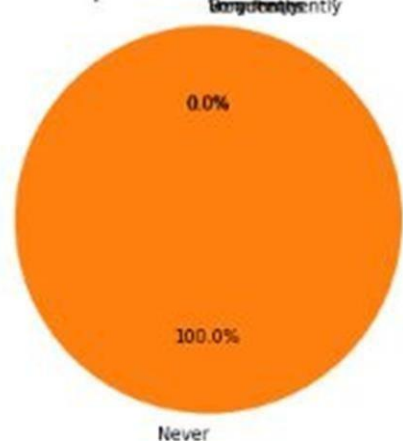
```



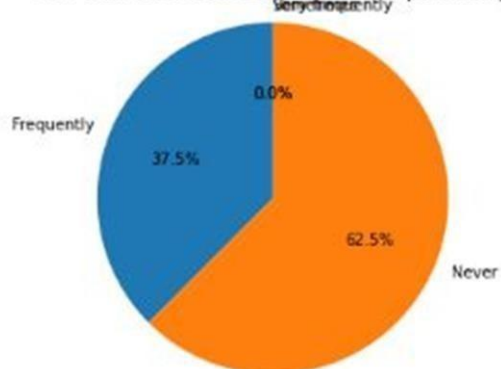
How often Mozilla user abandon without purchasing



How often opera abandon without purchasing

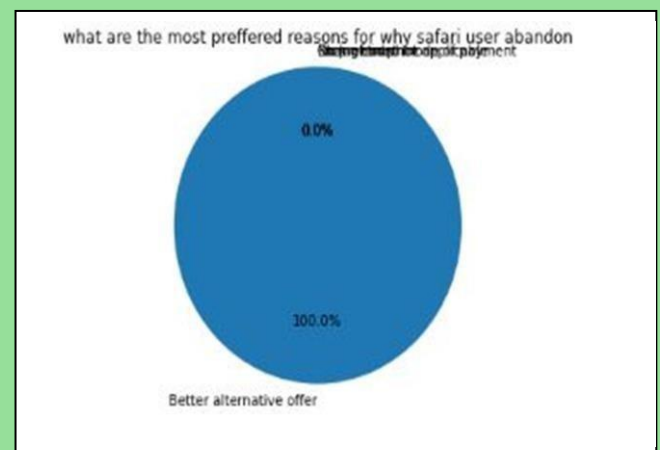
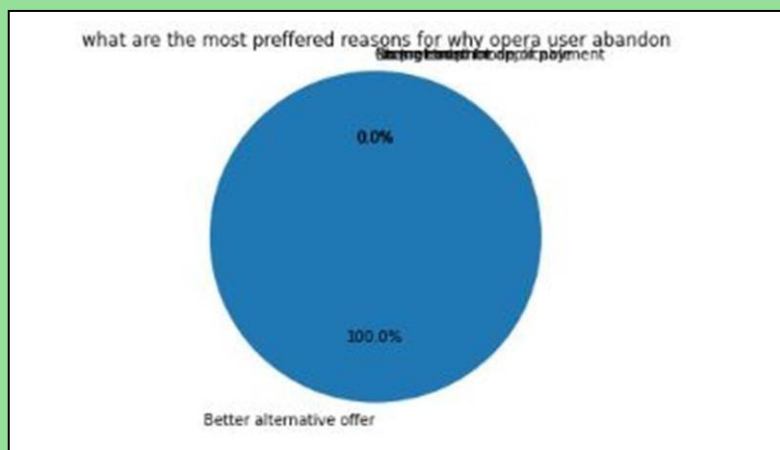
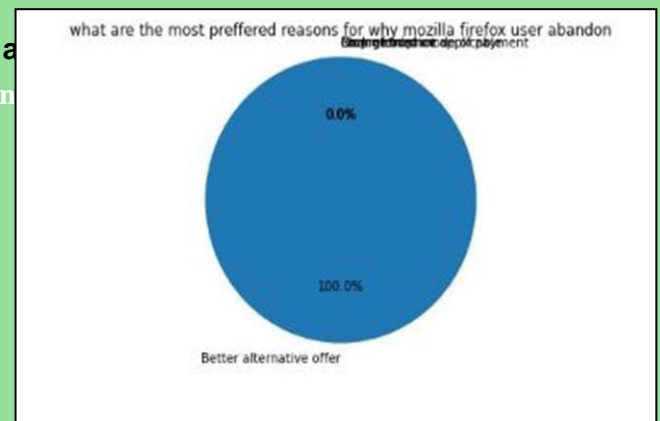
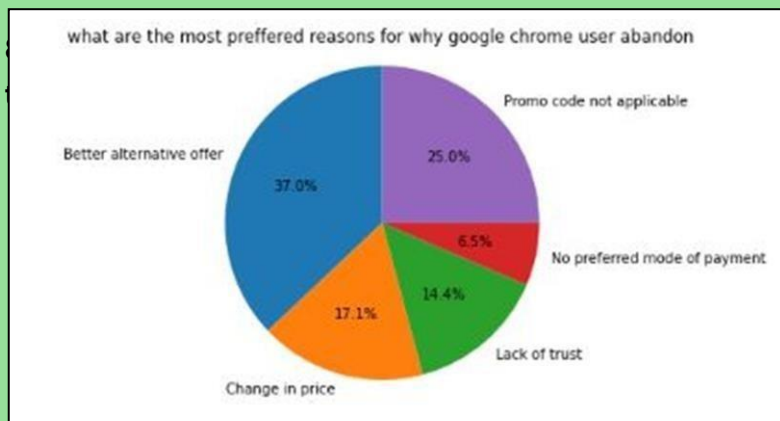


How often safari user abandon without purchasing



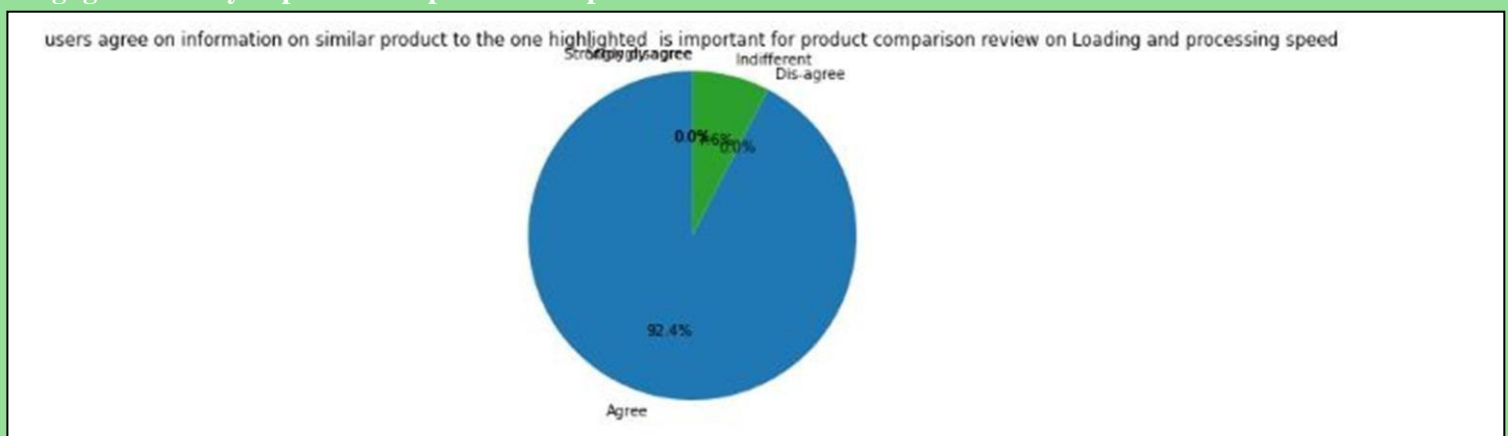
1. The study shows that around 77% of chrome user sometimes abandon shopping without purchasing
2. The study shows that the every mozilla user sometimes abandon shopping without purchasing
3. The study shows that the no opera user abandon the shopping without purchasing

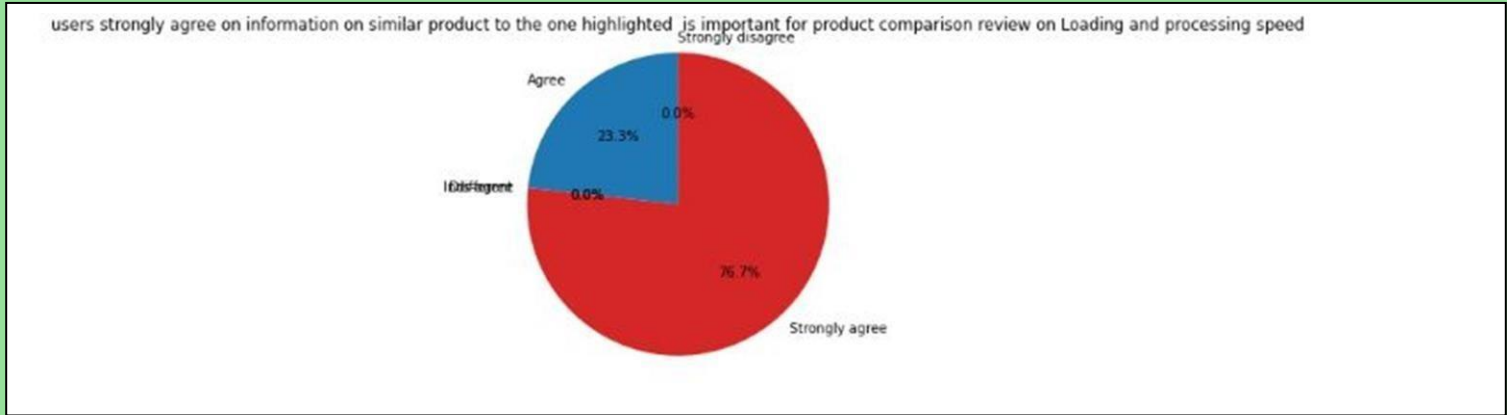
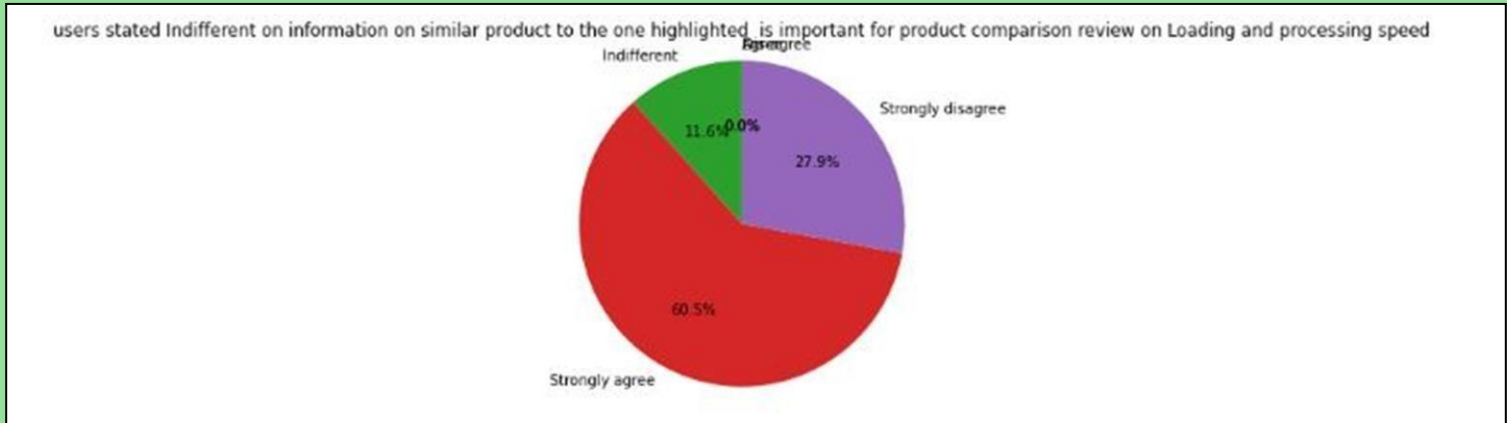
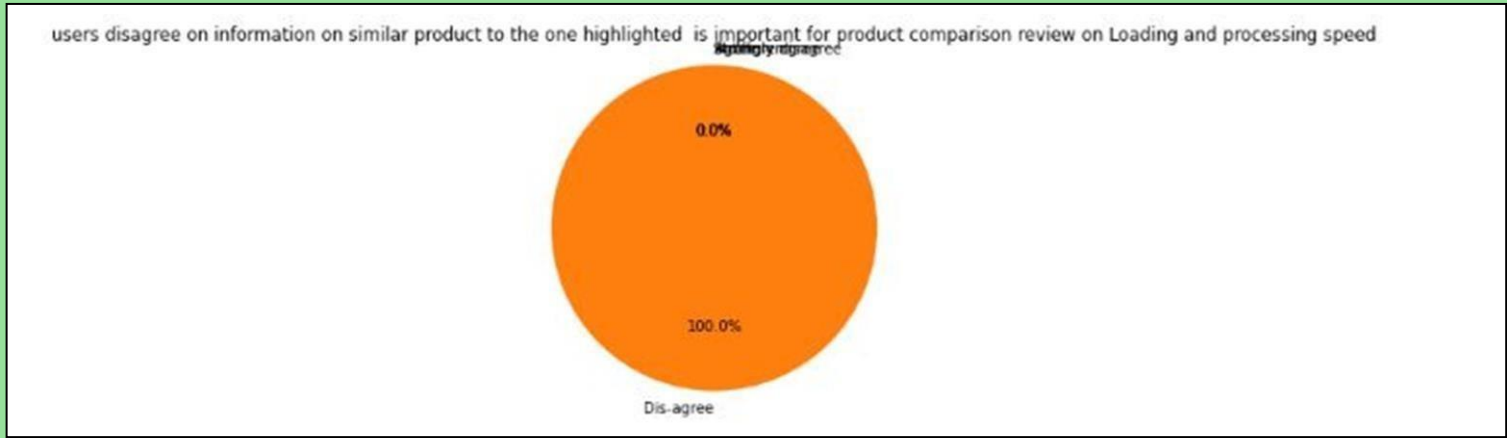
abandon



Better alternative offer is the reason why almost all type of user gives for abandoning the shoppingcart

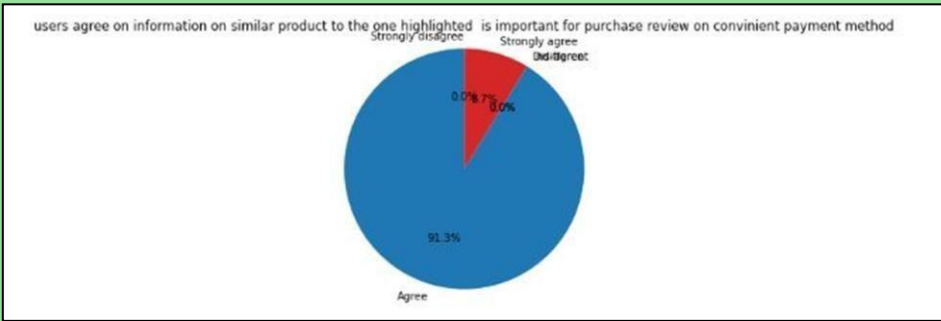
9. There is a significant correlation between the loading / processing speed and product description of one highlighted is very important for product comparison.

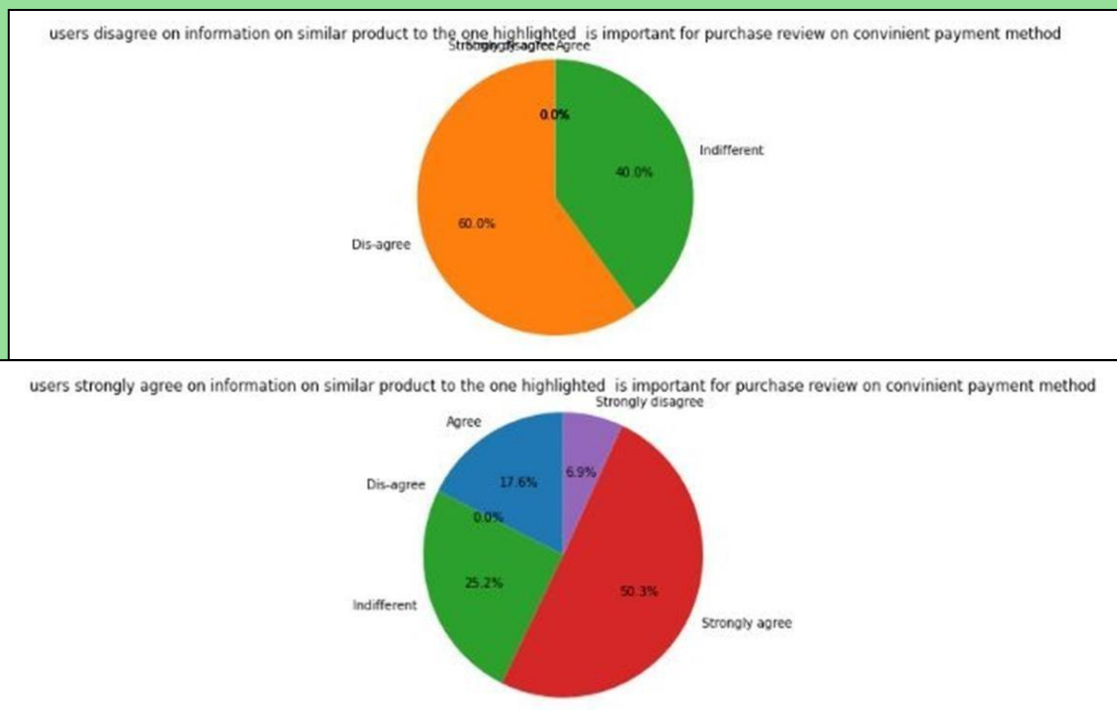




From the above graph we can say that users behave similarly on two topics 1.information on similar products to the one highlighted is very important for the product comparison and loading and processing speed.

10. The columns convenient payment methods and complete information on listed seller and product being offered is important for purchase decision has got significant relationship in following way.





Users react similarly on following two columns:

- a. Complete information on listed seller and product being offered is important for purchase decision
- b. Convenient Payment methods

11. If all the relevent information on listed products stated clearly then it would make the navigation soeasy for the users,

The ease of navigation over the website is depends on how much complete information is thereregarding the products.

The ease of navigation is also depends on how many payment options are there, if there are enoughamount of payment option especially cash on delivery then user will comment it as so easy for the navigation

12 .To be stated as the best user friendly interface of the website the retail shop should be veryresponsiveness and there should be several communication channels and the website should be easy for the navigation.

13. The study shows that if online retail store fulfill its part of the transaction in stipulated time then userwill feel gratified.

14. If Online retail shop want to get good rating for empathy towards the customer then they should assure the privacy of the customer at its best,the customer should get value for the money inferingthat there should be best service and quality of the product.

15. The above study shows that to make the customer satisfy the online retail shop should have enough responsiveness and should be available to the customer with enough communicationchannels.

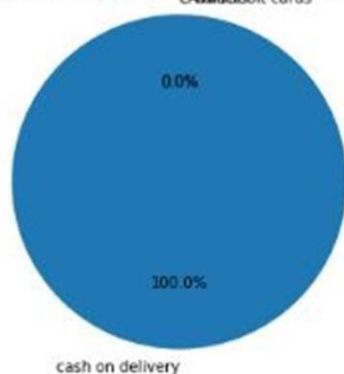
16. To retain the customer in this competitive world the online retail shop should give enough onlineassistance through multiple channels for retaining the customers, The website should be fast loading and there should be several payment option especially cash on delivery.

17. In case of payment option majority of the user likes amazon, the following pie chart shows which payment option most of the user uses in amazon.

```
: #amazon in case of payment option
k2=df1[(df1[df1.columns[14]]==0) & (df1[df1.columns[47]]==0)].count()
k3=df1[(df1[df1.columns[14]]==1) & (df1[df1.columns[47]]==0)].count()
k4=df1[(df1[df1.columns[14]]==2) & (df1[df1.columns[47]]==0)].count()
labels = [ 'cash on delivery','credit/debit cards','E-wallets']
sizes = [k2[0],k3[0],k4[0]]

fig1, ax1 = plt.subplots()
patches, texts, autotexts = ax1.pie(sizes,labels=labels, radius=3,autopct='%1.1f%%',startangle=90)
for text in texts:
    text.set_color('black')
for autotext in autotexts:
    autotext.set_color('black')
# Equal aspect ratio ensures that pie is drawn as a circle
ax1.axis('equal')
plt.title('which one of the payment option of amazon is most favorite for the users')
plt.tight_layout()
plt.show()
```

which one of the payment option of amazon is most favorite for the users



All user opted for the cash on delivery payment option as their most favorite payment option.

18. Amazon tops the chart in case of easy to use application why because it has got several payment option, It has got good online assistance, the page loads fast comparatively and it provides complete, relevant description information of products,
19. To be the most appealing webpage there should be wild variety of products on offer, there should be more payment options especially cash on delivery and the web page should be easy to use.
20. The fast loading of the website makes the user to easy to use the website
21. If any online retail website want to score top on the topic privacy of customers information then it should make sure that it will keep customers financial information as safe as possible.
22. Frequent disruptions when moving from one page to another is mostly caused by longer page loading.
23. To get your online retail shop offered to someone following are the required factor

There should be complete, relevant descriptive information of products, The website should be reliable. The website should be easy to use, The website should provide better online assistance through multi-channel

Conclusion

Key findings and the conclusions of the study:

To excel in this competitive world the retail shop should have very good assistance though many channels, It should provide several mode of payments especially cash on delivery adding to that it should have a appealing web page and clear and relevant product description, along with quick loading of web application.

Learning outcomes of the study in respect of Data Science

In this particular project learnt many things starting with we shouldn't directly jump into the project we should understand each and every variable and we should plan before entering like how to analyse etc, The preprocessing of data is very much important as otherwise it waste much time in later stage, We should have as much as grip over data visualization, We should write set of questions regarding what we are going to analyse, We should not deviate from our problem statement of the project.

There are certain limitations, Since all the users are not well versed with all type of web

Limitations of this work and scope of future work

applications in that case their opinion will be only for those applications which they have used, so the data was quite not fair.

As a part of scope for future work we can build the model.

!! Thank you

