



## **Customer Retention**

Submitted by:

R. Rajesh kannan

**ACKNOWLEDGEMENT**

Reference: [pandas.pydata.org](https://pandas.pydata.org), [seaborn.pydata.org](https://seaborn.pydata.org),  
[matplotlib.org](https://matplotlib.org)

Resource: [stackoverflow.com](https://stackoverflow.com), [geeksforgeeks.org](https://www.geeksforgeeks.org)

Data Sources are from Indian online shoppers

Other Resources are Project Use case.

## **Introduction**

Business Problem Framing:

Data Sources are from Indian online shoppers

For the success of online product seller customer satisfaction is most important. It has been posited as a key stimulant of purchase, repurchase intentions and customer loyalty.

Problem: Data is based on customer review based on that review of literature, theories and models have been carried out to propose the models for customer activation and customer retention

Conceptual Background of the Domain Problem:

Every retailer business improvement is based on five major factor they are,

1. service quality
2. system quality
3. information quality
4. trust
5. net benefit

Motivation for the Problem Undertaken:

Aim of this project is to analysis the data which get from e-commerce in India to get findings and conclusion, it is based on review which provided by the customers. These reviews are

based on service quality, system quality, information quality, trust and net benefit.

By analysis of data retailers can predict that customer activation and retention.

### **Analytical Problem Framing**

Mathematical/ Analytical Modeling of the Problem:

Here all the data are categorical data and some ordinal data, so there are no require of mathematical and statistical modeling.

For analytical purpose I used visualization to analysis the data. Here I used matplotlib.pyplot, seaborn and simple plot function in pandas.

# Data Sources and their formats:

## Data Sources: From Indian E-commerce

### Formats:

```
In [6]: df.dtypes
Out[6]: gender of respondent      object
        how old are you          object
        which city do you shop online from  object
        what is the pin code of where you shop online from  int64
        since how long you are shopping online  object
        how many times you have made an online purchase in the past year  object
        how do you access the internet while shopping on-line  object
        which device do you use to access the online shopping  object
        what is the screen size of your mobile device  object
        what is the operating system (os) of your device  object
        what browser do you run on your device to access the website  object
        which channel did you follow to arrive at your favorite online store for the first time  object
        after first visit, how do you reach the online retail store  object
        how much time do you explore the e- retail store before making a purchase decision  object
        what is your preferred payment option  object
        how frequently do you abandon (selecting an item and leaving without making payment) your shopping cart  object
        why did you abandon the "bag", "shopping cart"  object
        the content on the website must be easy to read and understand  object
        information on similar product to the one highlighted is important for product comparison  object
        complete information on listed seller and product being offered is important for purchase decision.  object
        all relevant information on listed products must be stated clearly  object
        ease of navigation in website  object
        loading and processing speed  object
        user friendly interface of the website  object
        convenient payment methods  object
        trust that the online retail store will fulfill its part of the transaction at the stipulated time  object
        empathy (readiness to assist with queries) towards the customers  object
        being able to guarantee the privacy of the customer  object
        responsiveness, availability of several communication channels (email, online rep, twitter, phone etc.)  object
        online shopping gives monetary benefit and discounts  object
        enjoyment is derived from shopping online  object
        shopping online is convenient and flexible  object
        return and replacement policy of the e-tailer is important for purchase decision  object
        gaining access to loyalty programs is a benefit of shopping online  object
        displaying quality information on the website improves satisfaction of customers  object
        user derive satisfaction while shopping on a good quality website or application  object
        net benefit derived from shopping online can lead to users satisfaction  object
        user satisfaction cannot exist without trust  object
        offering a wide variety of listed product in several category  object
        provision of complete and relevant product information  object
        monetary savings  object
        the convenience of patronizing the online retailer  object
        shopping on the website gives you the sense of adventure  object
        shopping on your preferred e-tailer enhances your social status  object
        you feel gratification shopping on your favorite e-tailer  object
        shopping on the website helps you fulfill certain roles  object
        getting value for money spent  object
        from the following, tick any (or all) of the online retailers you have shopped from;  object
        easy to use website or application  object
        visual appealing web-page layout  object
        wide variety of product on offer  object
        complete, relevant description information of products  object
        fast loading website speed of website and application  object
        reliability of the website or application  object
        quickness to complete purchase  object
        availability of several payment options  object
        speedy order delivery  object
        privacy of customers' information  object
        security of customer financial information  object
        perceived trustworthiness  object
        presence of online assistance through multi-channel  object
        longer time to get logged in (promotion, sales period)  object
        longer time in displaying graphics and photos (promotion, sales period)  object
        late declaration of price (promotion, sales period)  object
        longer page loading time (promotion, sales period)  object
        limited mode of payment on most products (promotion, sales period)  object
        longer delivery period  object
        change in website/application design  object
        frequent disruption when moving from one page to another  object
        website is as efficient as before  object
        which of the indian online retailer would you recommend to a friend  object
        dtype: object
```

Every data type is object data type except the Pincode feature.

Necessary:

Here all the categorical and ordinal data are necessary, we can treat that necessary variable as our requirement. If the data have more than 60% is null then we have to drop that, if there replacement value for null value in description then treat by replacement, but here there is no presence of null value.

Data Preprocessing:

There is no presence of lot of null values in the given data set.

```
In [43]: df.isnull().sum()
```

```
Out[43]: gender of respondent
how old are you
which city do you shop online from
what is the pin code of where you shop online from
since how long you are shopping online
how many times you have made an online purchase in the past year
how do you access the internet while shopping on-line
which device do you use to access the online shopping
what is the screen size of your mobile device
what is the operating system (os) of your 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
after first visit, how do you reach the online retail store
how much time do you explore the e- retail store before making a purchase decision
what is your preferred payment option
how frequently do you abandon (selecting an items and leaving without making payment) your shopping cart
why did you abandon the "bag", "shopping cart"
the content on the website must be easy to read and understand
information on similar product to the one highlighted is important for product comparison
complete information on listed seller and product being offered is important for purchase decision.
all relevant information on listed products must be stated clearly
ease of navigation in website
loading and processing speed
user friendly interface of the website
convenient payment methods
trust that the online retail store will fulfill its part of the transaction at the stipulated time
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.)
online shopping gives monetary benefit and discounts
enjoyment is derived from shopping online
shopping online is convenient and flexible
return and replacement policy of the e-tailer is important for purchase decision
gaining access to loyalty programs is a benefit of shopping online
displaying quality information on the website improves satisfaction of customers
user derive satisfaction while shopping on a good quality website or application
net benefit derived from shopping online can lead to users satisfaction
user satisfaction cannot exist without trust
offering a wide variety of listed product in several category
provision of complete and relevant product information
monetary savings
the convenience of patronizing the online retailer
shopping on the website gives you the sense of adventure
shopping on your preferred e-tailer enhances your social status
you feel gratification shopping on your favorite e-tailer
shopping on the website helps you fulfill certain roles
getting value for money spent
from the following, tick any (or all) of the online retailers you have shopped from:
easy to use website or application
visual appealing web-page layout
wild variety of product on offer
complete, relevant description information of products
fast loading website speed of website and application
reliability of the website or application
quickness to complete purchase
availability of several payment options
speedy order delivery
privacy of customers' information
security of customer financial information
perceived trustworthiness
presence of online assistance through multi-channel
longer time to get logged in (promotion, sales period)
longer time in displaying graphics and photos (promotion, sales period)
late declaration of price (promotion, sales period)
longer page loading time (promotion, sales period)
limited mode of payment on most products (promotion, sales period)
longer delivery period
change in website/application design
frequent disruption when moving from one page to another
website is as efficient as before
which of the indian online retailer would you recommend to a friend
dtype: int64
```

Categorical data will not contain outliers and also there will be no presence of skewed in data here all values are get by click the give check box.

Data Inputs and Output Logic Relationships:

There is no presence of output features in given dataset.

But we can predict it by certain models. Output data will be binary classification i.e here we have to predict customer activation and retention.

Visualization techniques like bi-variant visualization can also perform to see relationship between input features and also to visualize its % of occupation in total data.

To find:

Apply analytical skills to get findings and conclusions in detailed from data.

Hardware and Software Requirements and Tools Used:

Hardware: i5 processor, 8 GB RAM.

Software: OS(windows),

Tools: Jupiter Notebook or Py charm

Libraries: numpy, pandas, matplotlib, seaborn, sklearn

Packages: Pyplot, metrics, model\_selection, and respective model packages.

Analysis by Visualizations:

Here I divide data into three different category (user details, Rating, website prefer).

```
In [10]: #seperate dataframe into three different dataframe for analysis
df.columns.get_loc(" the content on the website must be easy to read and understand")

Out[10]: 17

In [11]: #details of user
df1=df.iloc[:,17:47]

In [12]: df1.shape

Out[12]: (269, 30)

In [13]: df.columns.get_loc("from the following, tick any (or all) of the online retailers you have shopped from; ")

Out[13]: 47

In [14]: #ranking
df2=df.iloc[:,47:]

In [15]: df2.shape

Out[15]: (269, 24)

In [16]: #website prefer
df0=df.iloc[:,0:17]
```

Visualization on user details:

Uni variant visualization:

Count of city:



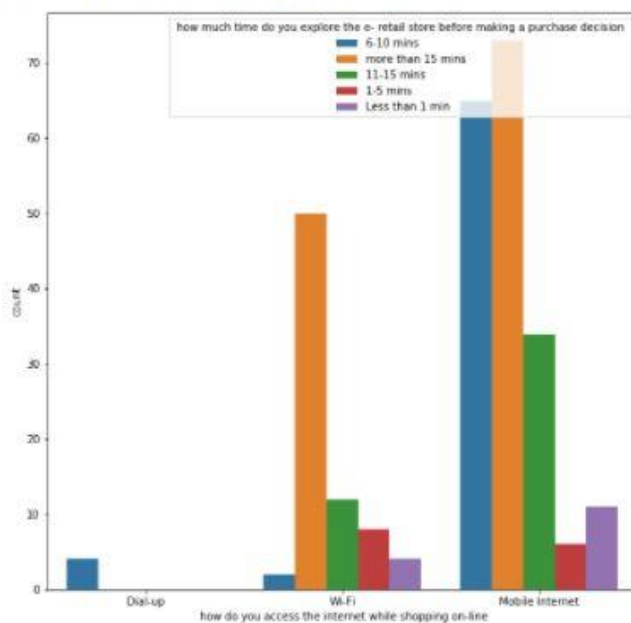


Most of the customers are shopping from (Delhi, Greater Noida, Noida, Bangalore, Karnal).

Bi Variant analysis:

Relation between spent time on shopping and internet source.

```
In [20]: plt.figure(figsize=(10,10))
sns.countplot(x=df[" how do you access the internet while shopping on-line"],hue=df[" how much time do you explore the e- retail
Out[20]: <matplotlib.axes._subplots.AxesSubplot at 0x19d36552448>
```

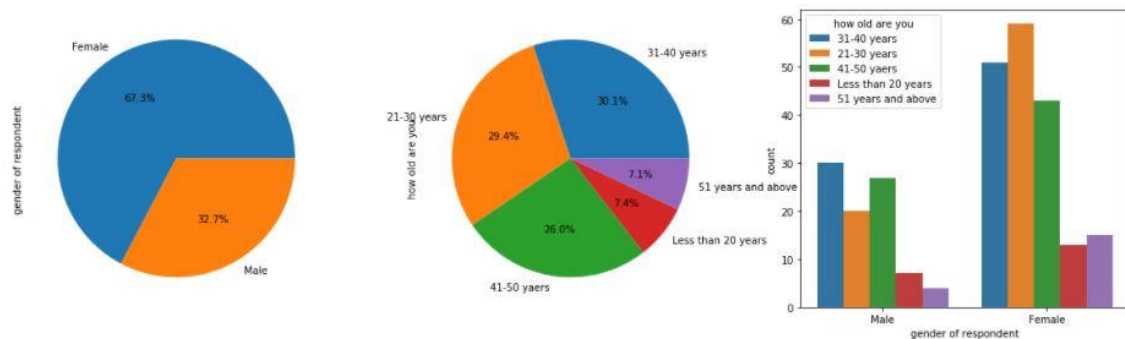


Customers use mobile internet will spend more time on e-retail.

## Relation between gender and age (also uni-variant):

```
In [22]: ["gender of respondent", "how old are you"]
plt.figure(figsize=(20,12))
plt.subplot(2,3,1)
df0["gender of respondent"].value_counts().plot.pie(autopct="%1f%%")
plt.subplot(2,3,2)
df0["how old are you"].value_counts().plot.pie(autopct="%1f%%")
plt.subplot(2,3,3)
sns.countplot(df0["gender of respondent"],hue=df0["how old are you"])
```

Out[22]: <matplotlib.axes.\_subplots.AxesSubplot at 0x19d36515148>

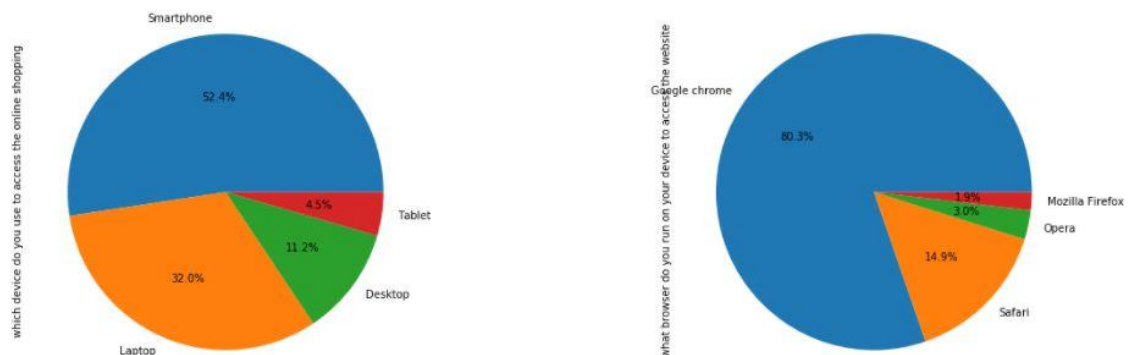


1. Most of the customers are females nearly about 67%.
2. Majority of Females are between 31-40, 21-30 and 41-50.

## Uni-variant visualization on browser use and device use:

```
In [23]: plt.figure(figsize=(20,15))
plt.subplot(2,2,1)
df0["which device do you use to access the online shopping"].value_counts().plot.pie(autopct="%1f%%")
plt.subplot(2,2,2)
df0["what browser do you run on your device to access the website"].value_counts().plot.pie(autopct="%1f%%")
```

Out[23]: <matplotlib.axes.\_subplots.AxesSubplot at 0x19d3762a3c8>



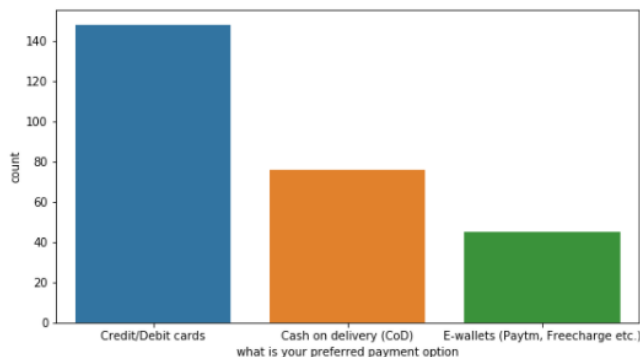
1. Lot of users are access shopping on smartphone and Laptop.

2. Mainly users are using Google chrome as browser to access website

Uni-varient visualization on Payment option:

```
In [24]: plt.figure(figsize=(20,5))
plt.subplot(1,2,1)
sns.countplot(df0[" what is your preferred payment option "],order=df0[" what is your preferred payment option "].value_counts())

Out[24]: <matplotlib.axes._subplots.AxesSubplot at 0x19d36474888>
```

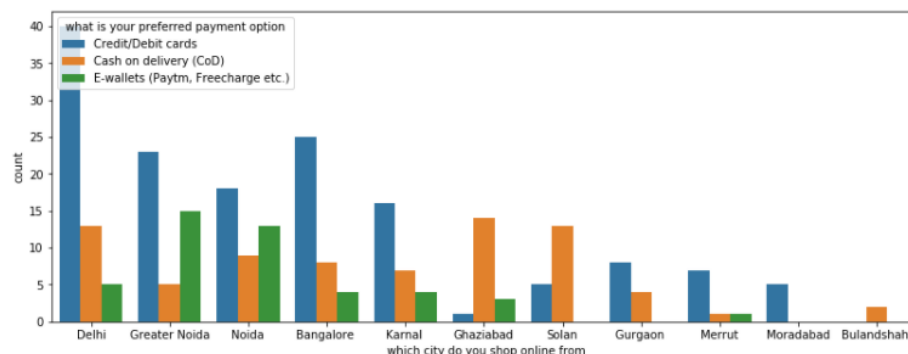


Card way transaction are high in payment method than other mode of transaction.

Combine visual of cities and payment method:

```
In [25]: plt.figure(figsize=(30,5))
plt.subplot(1,2,1)
sns.countplot(df0[" which city do you shop online from"],order=df0[" which city do you shop online from"].value_counts().index, hue="what is your preferred payment option")

Out[25]: <matplotlib.axes._subplots.AxesSubplot at 0x19d363ee048>
```



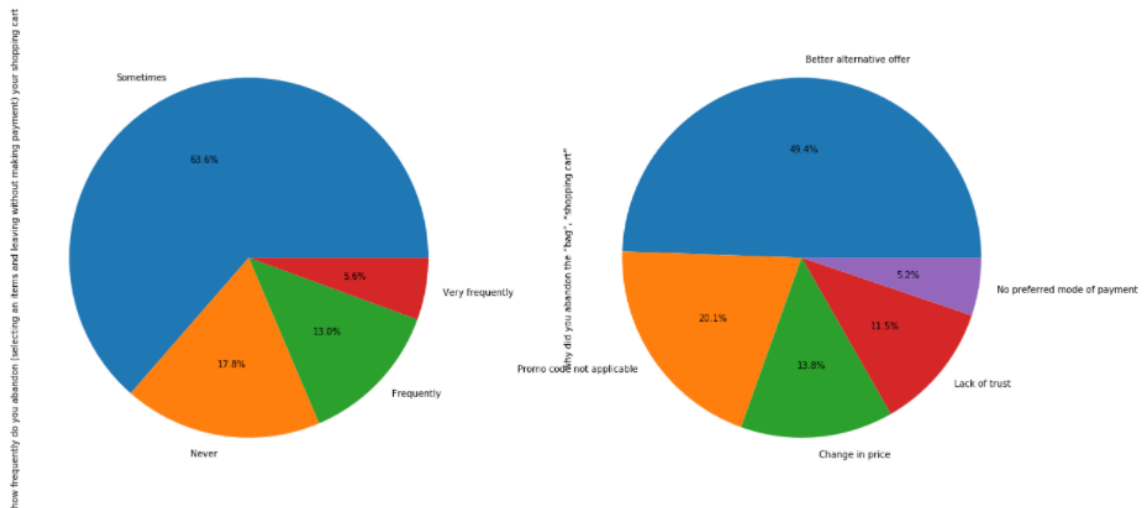
1. Cities like Ghziabad, Solan, cash on delivery payment method plays major roll.

2. For other cities card payment plays major roll.

Uni-varient visual on abandon frequency on cart and why:

```
In [26]: plt.figure(figsize=(20,20))
plt.subplot(2,2,1)
df0[" how frequently do you abandon (selecting an items and leaving without making payment) your shopping cart "].value_counts()
plt.subplot(2,2,2)
df0[" why did you abandon the "bag", "shopping cart" "].value_counts().plot.pie(autopct="%1f%%")
```

Out[26]: <matplotlib.axes.\_subplots.AxesSubplot at 0x19d3649f9c8>

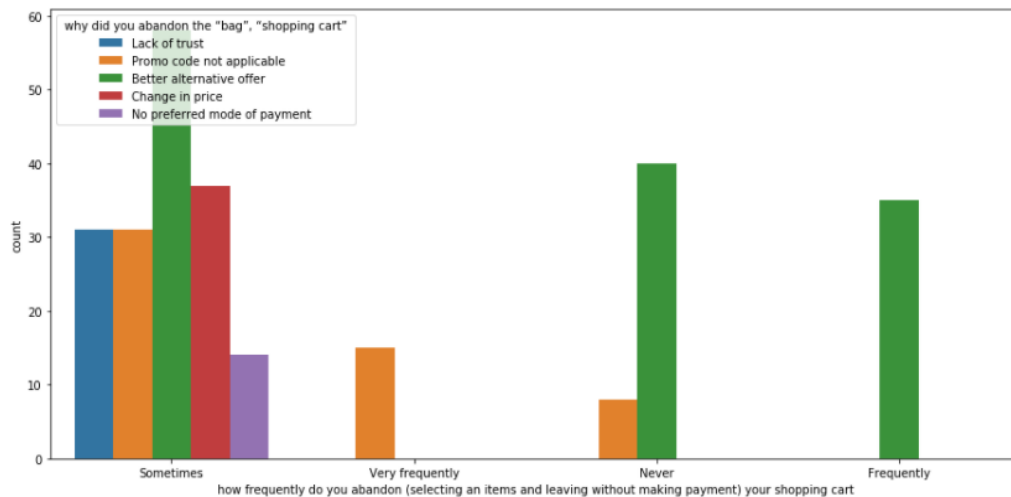


1. It shows mostly peoples are abandon cart or bag for better alternative offer.
2. 63% of peoples are abandon shopping cart for sometimes.

## Combine visual of abandon frequency on cart and why:

```
In [27]: plt.figure(figsize=(15,7))
sns.countplot(df0[" how frequently do you abandon (selecting an items and leaving without making payment) your shopping cart "],)
```

```
Out[27]: <matplotlib.axes._subplots.AxesSubplot at 0x19d36369388>
```



## Visualization on Rating data frame:

Divide rating category features as several category for better findings.

```
In [28]: #ranking dataframe are seperated by its use
website=["website","speed","Respons","shopping online"]
money=["Payment","money","discounts","saving"]
sat=["customer","user"]
ret=["e-tailer","retail"]
```

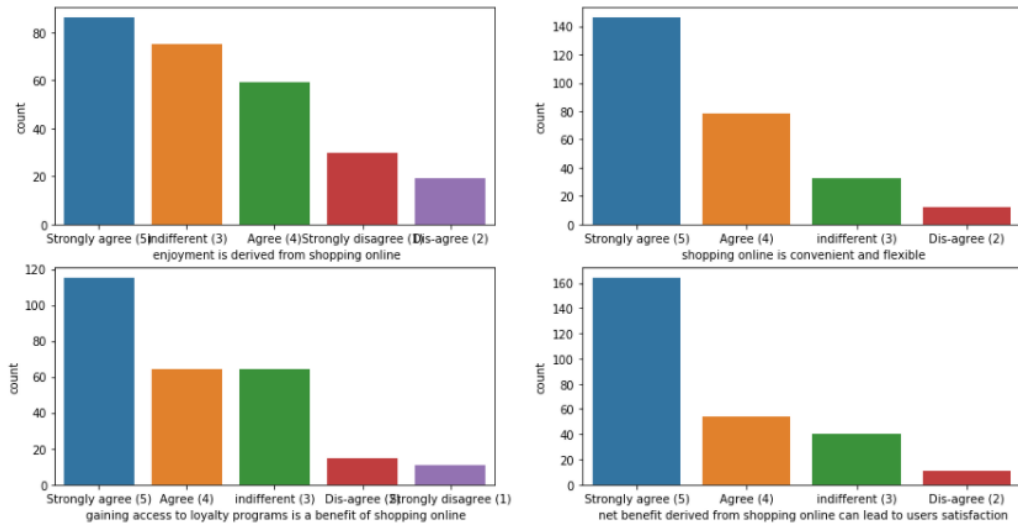
```
In [30]: dfw=pd.concat([df1.filter(like="website"),df1.filter(like="speed"),df1.filter(like="respons")],axis=1)
dfm=pd.concat([df1.filter(like="payment"),df1.filter(like="money"),df1.filter(like="discounts"),df1.filter(like="saving")],axis=1)
dfc=pd.concat([df1.filter(like="customer"),df1.filter(like="user")],axis=1)
dfp=df1.filter(like="product")
dfr=pd.concat([df1.filter(like="e-tailer"),df1.filter(like="retail")],axis=1)
dfs=df1.filter(like="shopping online")
```

```
In [31]: print(dfs.shape,dfm.shape,dfw.shape,dfp.shape,dfr.shape,dfc.shape)

(269, 4) (269, 4) (269, 9) (269, 5) (269, 5) (269, 7)
```

## Visualize features under shopping satisfaction:

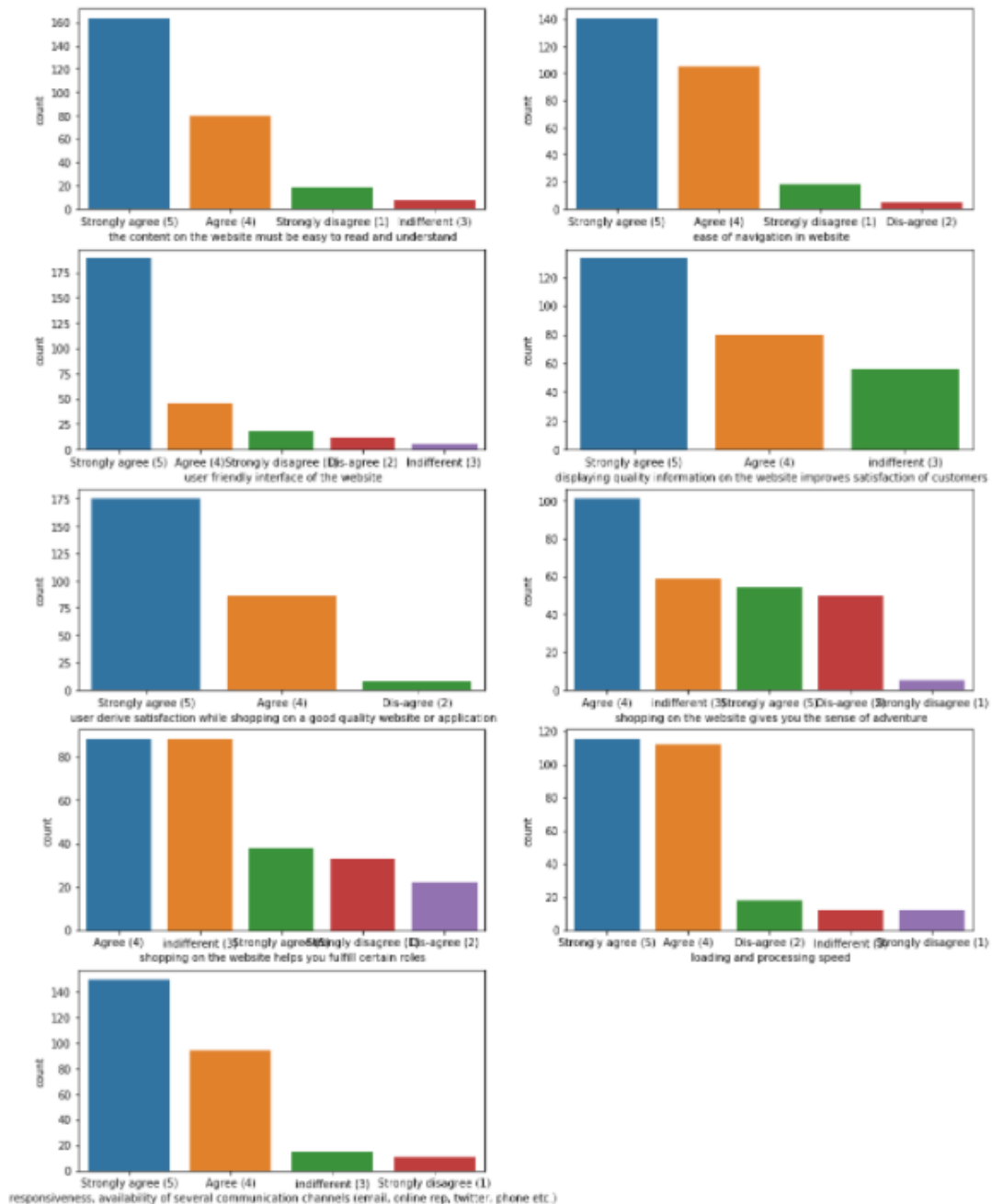
```
In [32]: #shopping satisfaction
plt.figure(figsize=(15,20))
for i in dfs.columns:
    plt.subplot(5,2,(dfs.columns.get_loc(i)+1))
    sns.countplot(df1[i],order=df[i].value_counts().index)
```



Mostly peoples are strongly agree with shopping benefit, satisfaction, flexible, enjoyment of shopping.

## Visualize features under Website Ratings:

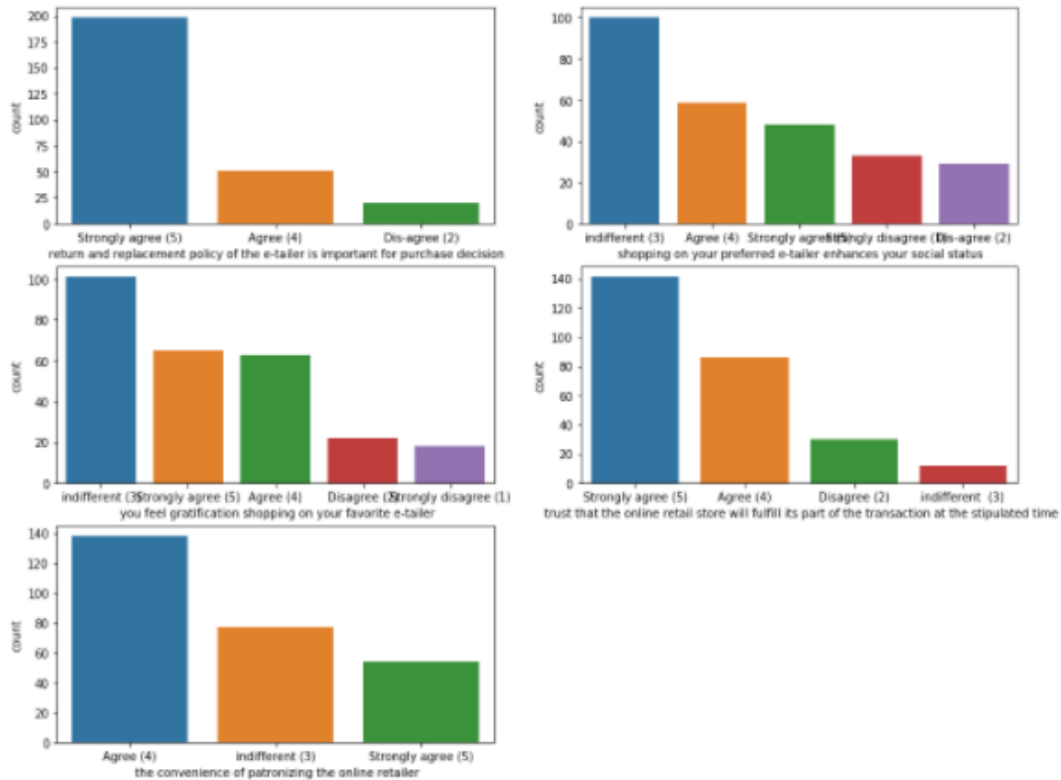
```
In [33]: #website ratings
plt.figure(figsize=(15,20))
for i in dfw.columns:
    plt.subplot(5,2,(dfw.columns.get_loc(i)+1))
    sns.countplot(df1[i],order=df1[i].value_counts().index)
```



Most customers are strongly agree with website speed, response, and website use.

## Visualize features under retailer ratings:

```
In [34]: #retailer ratings
plt.figure(figsize=(15,20))
for i in dfr.columns:
    plt.subplot(5,2,(dfr.columns.get_loc(i)+1))
    sns.countplot(df1[i],order=df1[i].value_counts().index)
```

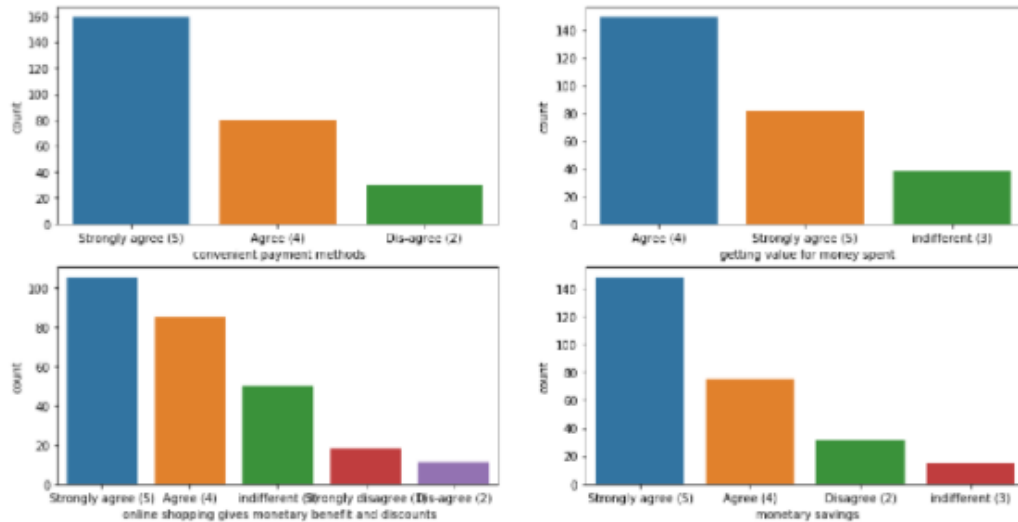


1. Most customers are agree with trust on retailer and replacement policy.
2. Most customers are rate indifferent for retailer enhances of social status and gratification shopping.



Visualize features under money satisfaction and safety:

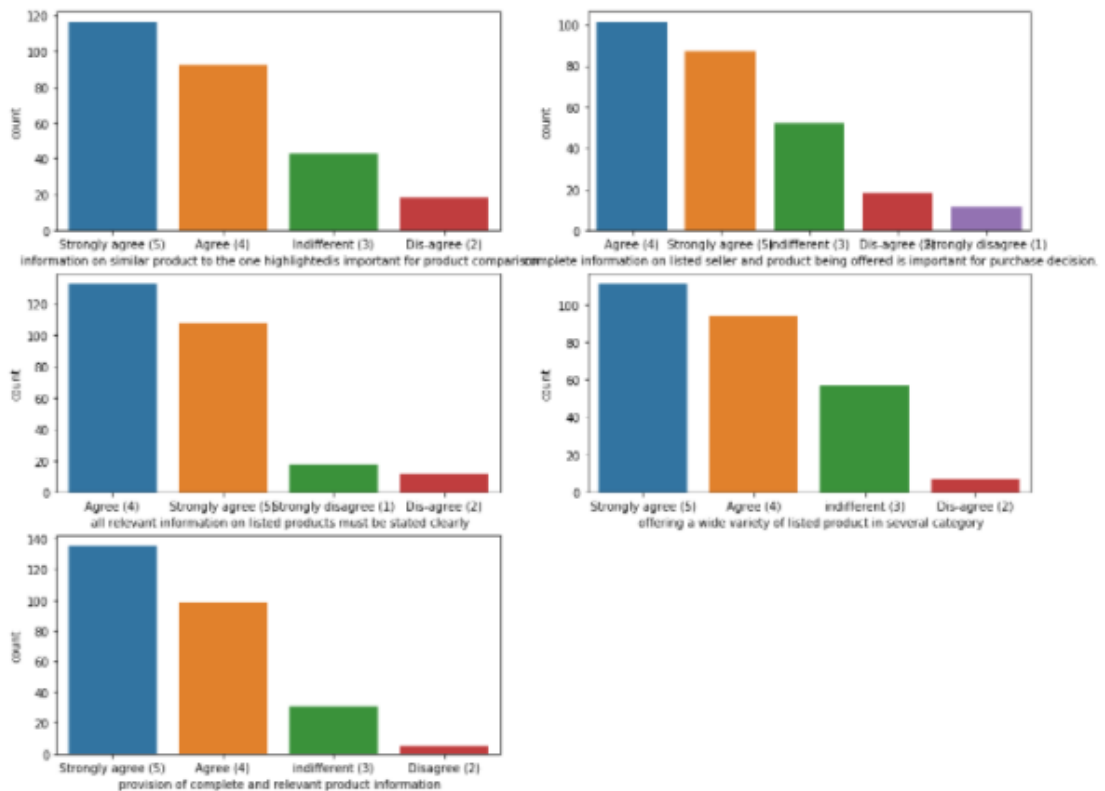
```
In [35]: #money satisfaction
plt.figure(figsize=(15,20))
for i in dfm.columns:
    plt.subplot(5,2,(dfm.columns.get_loc(i)+1))
    sns.countplot(df1[i],order=df1[i].value_counts().index)
```



Most customers are strongly agree with money satisfaction and trust.

## Visualize features under product rating:

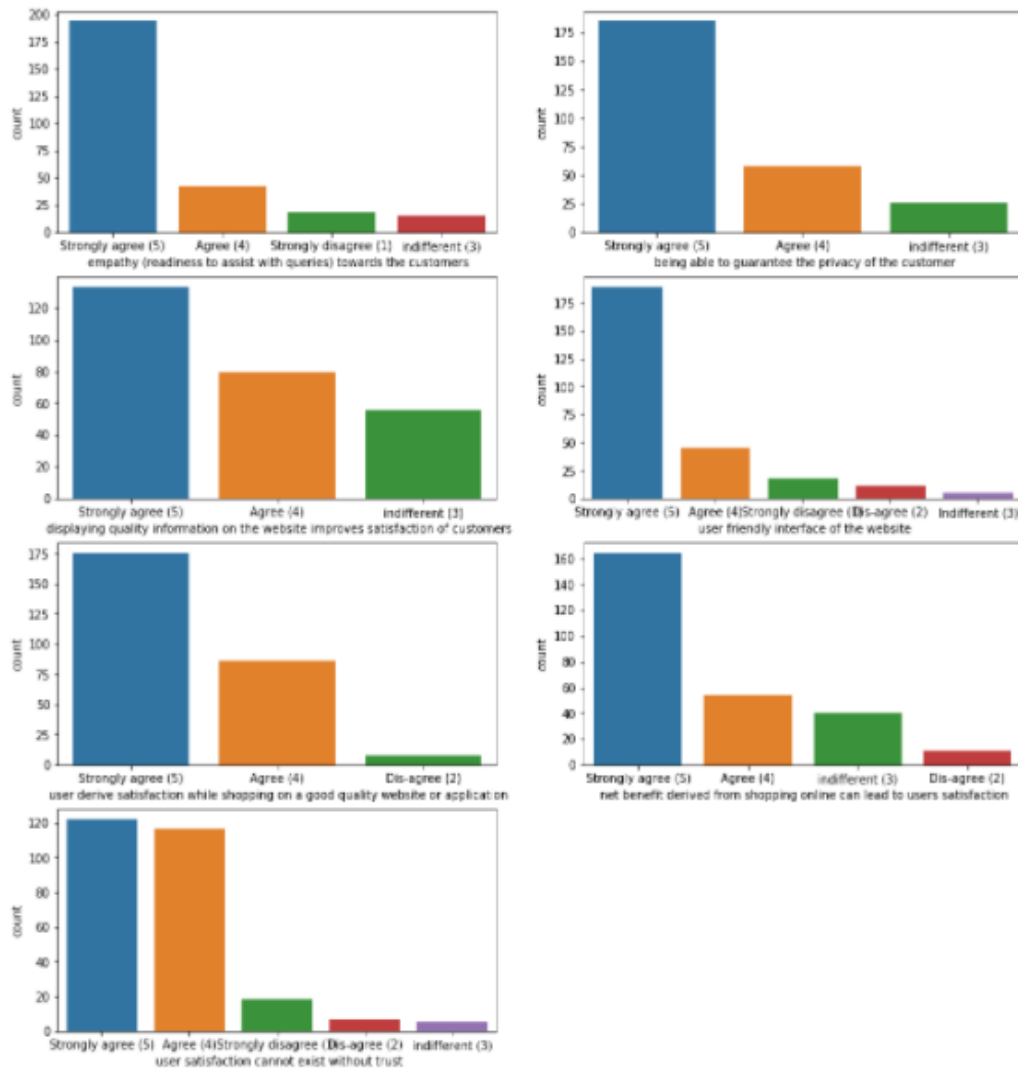
```
In [36]: #product ratings
plt.figure(figsize=(15,20))
for i in dfp.columns:
    plt.subplot(5,2,(dfp.columns.get_loc(i)+1))
    sns.countplot(dfp[i],order=dfp[i].value_counts().index)
```



Most customers are rated strongly agree with product related features.

## Visualize features under customer satisfaction ratings:

```
In [37]: #customers satisfaction ratings
plt.figure(figsize=(15,20))
for i in dfc.columns:
    plt.subplot(5,2,(dfc.columns.get_loc(i)+1))
    sns.countplot(df1[i],order=df[i].value_counts().index)
```



Most customers are strongly agree with their satisfaction and trust.

## Visualize on web ratings:

There are some positive ratings and negative ratings columns, so we have to separate it for better visualization.

```
In [44]: df2.columns.get_loc("longer time to get logged in (promotion, sales period)")
```

```
Out[44]: 14
```

```
In [45]: df2.columns.get_loc("frequent disruption when moving from one page to another")
```

```
Out[45]: 21
```

```
In [46]: dn=df2.iloc[:,14:22]          #filter negative rating columns
          dn.shape
```

```
Out[46]: (269, 8)
```

```
In [47]: dp=df2.drop(dn.columns,axis=1)  #filter positive rating columns
          dp.shape
```

```
Out[47]: (269, 16)
```

Encode the web ratings columns by develop dummies for every website ratings columns by each sites.

```
In [48]: list1=[]
          list2=[]
          for i in dp.columns:          #develop dummies for every website ratings columns by each sites
              list1.append(df2[i].str.get_dummies(sep=" ", ).add_prefix(" ").add_prefix(i))
          for i in dn.columns:
              list2.append(df2[i].str.get_dummies(sep=" ", ).add_prefix(" ").add_prefix(i))
```

```
In [49]: good=pd.concat(list1,axis=1)
          bad=pd.concat(list2,axis=1)
```

```
In [50]: good
```

	presence of online assistance through multi-channel Myntra.com	presence of online assistance through multi-channel Paytm.com	presence of online assistance through multi-channel Snapdeal	website is as efficient as before Amazon.in	website is as efficient as before Flipkart.com	website is as efficient as before Myntra.com	website is as efficient as before Paytm.com	website is as efficient as before Snapdeal.com	which of the indian online retailer would you recommend to a friend Amazon.in	which of the indian online retailer would you recommend to a friend Flipkart.com	which of the indian online retailer would you recommend to a friend Myntra.com	which of the indian online retailer would you recommend to a friend Paytm.com	which of the indian online retailer would you recommend to a friend snapdeal.com
0	0	1	0	1	0	0	0	0	0	1	0	0	0
1	1	0	0	1	1	0	0	0	1	0	1	0	0
0	1	0	0	1	0	0	0	0	1	0	1	1	0
1	1	0	1	1	1	0	1	0	1	1	0	0	0
0	1	0	0	0	0	0	1	0	1	0	1	0	0
1	0	1	0	1	0	0	0	0	1	0	0	1	0
1	1	0	1	0	0	0	0	1	0	1	1	1	1

```
In [51]: bad
```

```
Out[51]:
```

	longer time to get logged in (promotion, sales period) Amazon.in	longer time to get logged in (promotion, sales period) Flipkart.com	longer time to get logged in (promotion, sales period) Myntra.com	longer time to get logged in (promotion, sales period) Paytm.com	longer time to get logged in (promotion, sales period) Snapdeal.com	longer time in displaying graphics and photos (promotion, sales period) Amazon.in	longer time in displaying graphics and photos (promotion, sales period) Flipkart.com	longer time in displaying graphics and photos (promotion, sales period) Myntra.com	longer time in displaying graphics and photos (promotion, sales period) Paytm.com	longer time in displaying graphics and photos (promotion, sales period) Snapdeal.com	late declaration of price (promotion, sales period) Amazon.in	late declaration of price (promotion, sales period) Flipkart.com
0	1	0	0	0	0	1	0	0	0	0	0	0
1	1	1	0	0	0	0	0	1	0	0	0	0
2	0	0	1	0	0	0	0	1	0	0	0	0
3	0	0	0	0	1	0	0	1	0	1	0	0
4	0	1	0	1	0	0	0	0	1	0	0	0

```
In [76]: Total_positive_ratings=[]
Total_negative_ratings=[]
for i in ["Amazon", "Flipkart", "Mynta", "Paytm", "Snapdeal"]:
    Total_positive_ratings.append(good.sum().filter(like=1).sum())
    Total_negative_ratings.append(bad.sum().filter(like=1).sum())
print(Total_positive_ratings)
print(Total_negative_ratings)
```

[3633, 2628, 1372, 1030, 1072]  
[745, 539, 388, 524, 503]

```
In [82]: web_div=pd.DataFrame([Total_positive_ratings,Total_negative_ratings],columns=["Amazon", "Flipkart", "Mynta", "Paytm", "Snapdeal"],i
```

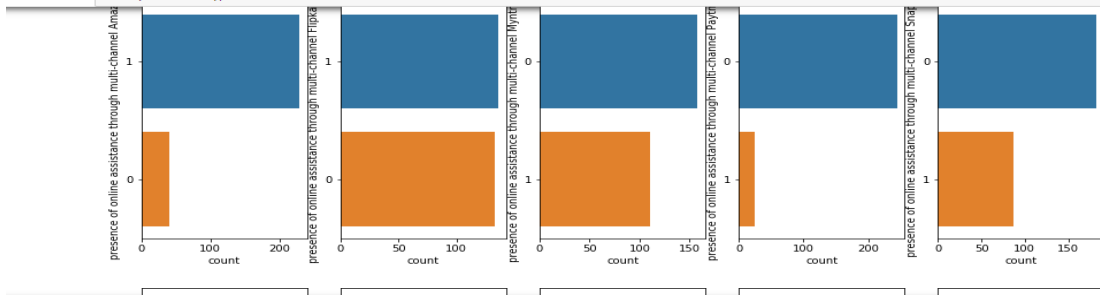
```
In [84]: web_div
```

```
Out[84]:
```

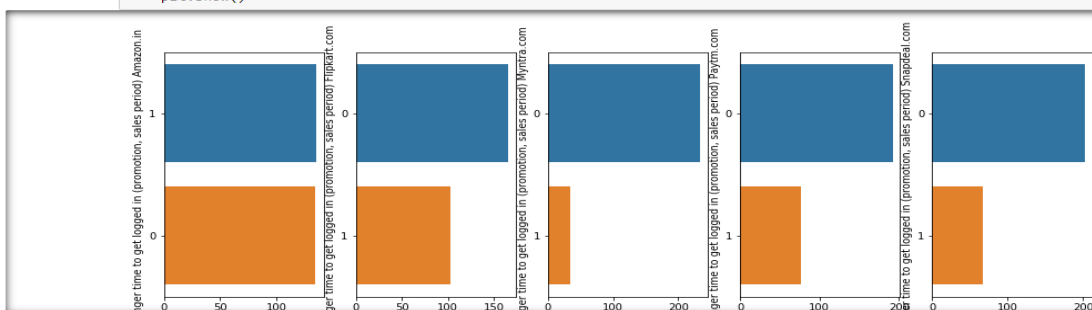
	Amazon	Flipkart	Mynta	Paytm	Snapdeal
positive	3633	2628	1372	1030	1072
negative	745	539	388	524	503

Amazon.in and flipkart.com are liked by most of the persons than different websites.

```
In [58]: for o in dp.columns:
cc=good.filter(like=o)
plt.figure(figsize=(15,5))
for i in cc.columns:
    plt.subplot(1,5,cc.columns.get_loc(i)+1)
    sns.countplot(y=cc[i],order=cc[i].value_counts().index)
plt.show()
```



```
In [59]: for o in dn.columns:
cc=bad.filter(like=o)
plt.figure(figsize=(15,5))
for i in cc.columns:
    plt.subplot(1,5,cc.columns.get_loc(i)+1)
    sns.countplot(y=cc[i],order=cc[i].value_counts().index)
plt.show()
```



## CONCLUSION

Key finding: Analysis in customer review and feedback.

Inferences: From the report it concluded that there are no wrong data. By analyze it and prediction was lead to get good model.

Observations:

1. For rating features mostly customer select strongly agree.
2. For customer details features(city-Delhi, browser-Google chrome, device-smartphone, net-Mobile Internet ) plays major role in online shopping.
3. For website rating features Amazon.in and flipkart.com are liked by most of the customers.

Learning Outcomes of the study in respect of Data Science

- I learned by visualize also can get important variables and also find how to extract information.
- Learned to analyse in categorical features.

Limitations and Future work:

Limitations: only presence of categorical data leads to less analysis.

Steps to follow further: Here I analyze all the data by visualization technique but I did not apply any model for prediction of customer activation and retention.

