

Customer Retention Case Study Report

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



SME name:
Sapna Verma.

Prepared by:
Abhimanyu Varpe.
(Data Science Intern at Flip Robo Technologies)

ACKNOWLEDGMENT

I would like to express my special thanks of gratitude to **Flip Robo Technologies** (Bangalore) and as well as **Datatrained education platform** for giving me the golden opportunity to do this wonderful project on the topic (E-retail factors for customer activation and retention), which also helped me in doing a lot of Research and understanding of the E-retail business and customer retention also came to know about so many new things I am really thankful to them.

It is my deepest pleasure and gratification to present this report. Working on this project was an incredible experience that has given me a very informative knowledge regarding the data analysis process.

Secondly, I would also like to thank my **SME Sapna Varma** mam for helping me throughout the project and giving the required time frame for project completion.

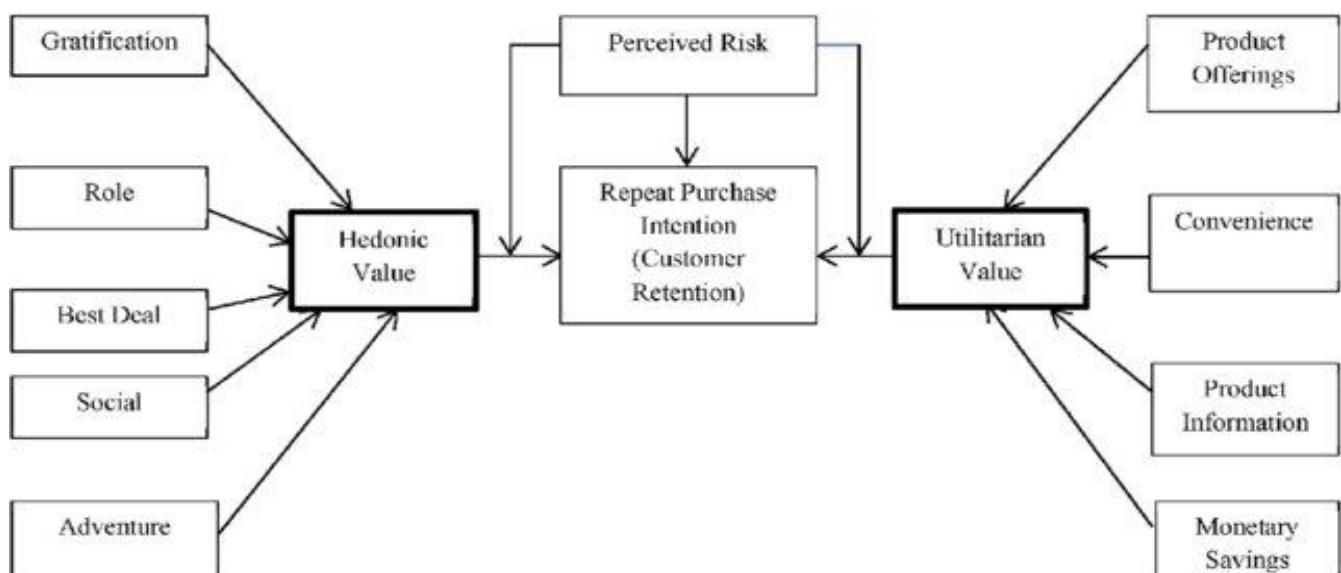
INTRODUCTION:

With the rapid global growth in electronic commerce (e-commerce), businesses are attempting to gain a competitive advantage by using e-commerce to interact with Nowadays, online shopping is a fast-growing phenomenon. Growing numbers of consumers shop online to purchase goods and services, gather product information or even browse for enjoyment. Online shopping environments are therefore playing an increasing role in the overall relationship between marketers and their consumers. That is, consumer-purchases are mainly based on the cyberspace appearance such as pictures, images, quality information, and video clips of the product, not on the actual experience. As the Internet has now become a truly global phenomenon, this growing and diverse Internet population means that people having diverse tastes and purposes are now going to the Web for information and to buy products and services. Thus, the impact of these online shopping environments on consumer response necessitates a critical understanding for marketing planning.

Problem Statement:

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



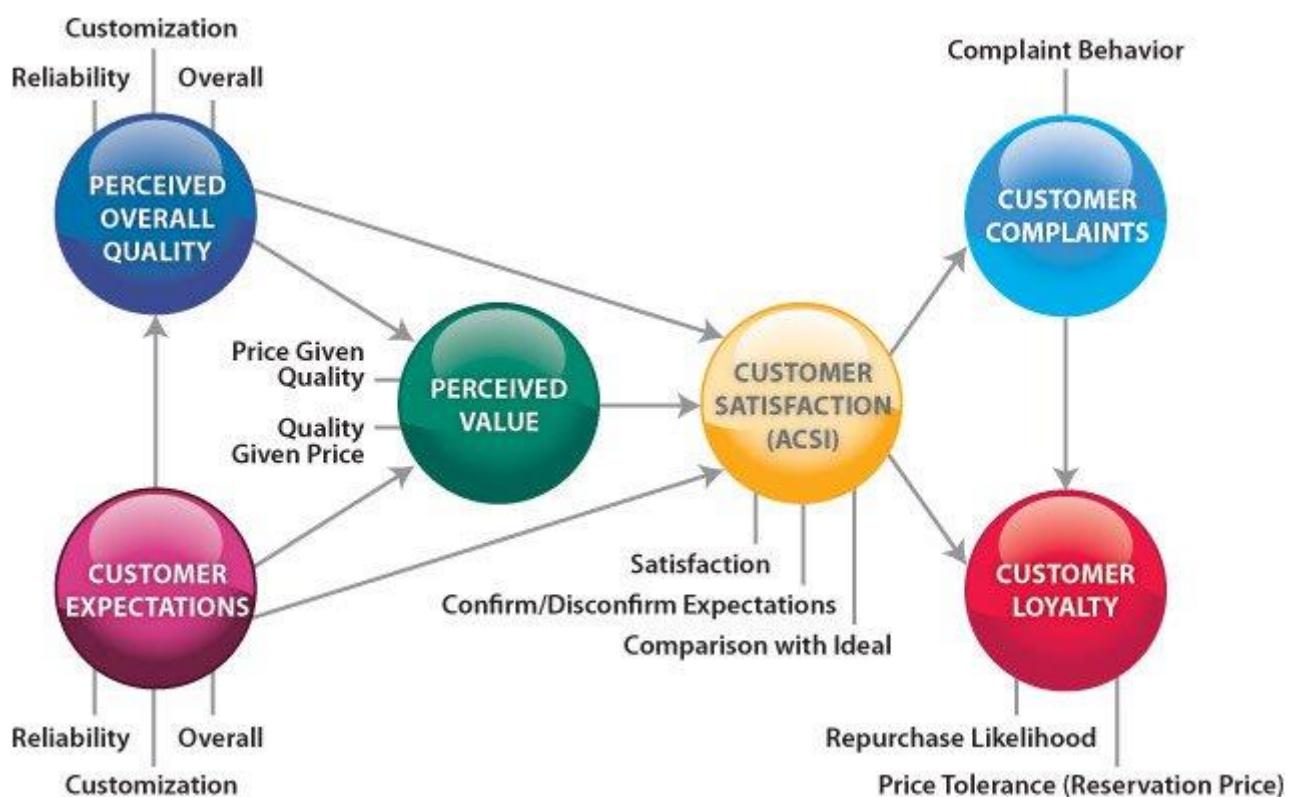
In the above use case diagram, we can see that the Repeat Purchase Intention basically our Customer Retention strategy relies on Hedonic value and Utilitarian value. Also, we see that there are perceived risks affecting the purchase and re purchase intentions of our customers. The Hedonic value has 5 major parts such as gratification, role, best deal, social aspect and adventure feeling criterions. Where as in Utilitarian value we have product offerings, convenience, product information and monetary savings.

Motivation for the Problem Undertaken:

Our main objective of doing this project is to analyse whether the users are shopping products from e-commerce websites, how did they give feedbacks to these websites on the basis of several positive? And then find out the way to customer retention through data analysis.

Customer satisfaction:

Customer satisfaction is defined as a measurement that determines how happy customers are with a company's products, services, and capabilities. Customer satisfaction information, including surveys and ratings, can help a company determine how to best improve or changes its products and services.



What is Customer Retention?

Customer retention refers to the activities and actions companies and organizations take to reduce the number of customer defections. The goal of customer retention programs is to help

companies retain as many customers as possible, often through customer loyalty and brand loyalty initiatives. It is important to remember that customer retention begins with the first contact a customer has with a company and continues throughout the entire lifetime of the relationship. While most companies traditionally spend more money on customer acquisition because they view it as a quick and effective way of increasing revenue, customer retention often is faster and, on average, costs up to seven times less than customer acquisition. Selling to customers with whom you already have a relationship is often a more effective way of growing revenue because companies don't need to attract, educate, and convert new ones.

Benefits of Customer Retention:

- Retention is cheaper than acquisition: Retaining customers is less expensive than attracting new customers, and to retain customers retailers must give them reasons to be loyal. A study from Harvard Business School showed that an increase in customer retention rates of 5% can increase profits by 25–95%.
- Loyal customers are more profitable: Not only is loyalty cheaper, it has better returns. According to research, engaged consumers buy 90% more frequently, spend 60% more per transaction and are five times more likely to indicate it is the only brand they would purchase in the future. On average, they're delivering 23% more revenue and profitability over the average customer. While loyal customers are more profitable, don't take their loyalty for granted.
- Fostering more repeat purchases from your customers should be the apex strategy for every digital retail brand. This strategy is based upon the following statistics.
- 91 % of the customers are likely to purchase from the brands that recognize them and send them personalized & contextual offers.
- On average, 65 % of the e-commerce revenue comes from repeat customers.
- New customer acquisitions can cost five times more than that of retaining the existing customer base.
- 86 % of the customers prefers to purchase from a brand that offers a great customer experience.

The above stats clearly show the power of the win-win relationship between a customer and a brand – as they both derive significant value from each other as long as the association lasts.



About Data:

dataset is the collection of the questionnaire which were asked in survey of Online retailers in Indian Market.

Detailed analysis was carried out on dataset containing information gathered from 269 customers in anaconda Jupiter notebook with Python, various libraries used for analysis.

Steps by step procedure:

- First, I imported all the necessary libraries and dependencies to create a detailed data analysis in Python.

```
1 #Importing the important packages and libraries of the python.  
2  
3 import pandas as pd  
4 import numpy as np  
5 import seaborn as sns  
6 import warnings  
7 warnings.filterwarnings('ignore')  
8 import matplotlib.pyplot as plt  
9 import plotly.express as px
```

- Then I have imported the dataset:

```
1 #Importing dataset  
2 df = pd.read_excel(r"D:\data scientist\Internship\Customer_retention_dataset\customer_retention_dataset.xlsx")  
3 df
```

- I made sure to rename the column names that were ill formatted and quite long which made no sense to me. With the help of rename I was able to change the names of columns that were too lengthy and could have been accommodated in shorter formats.

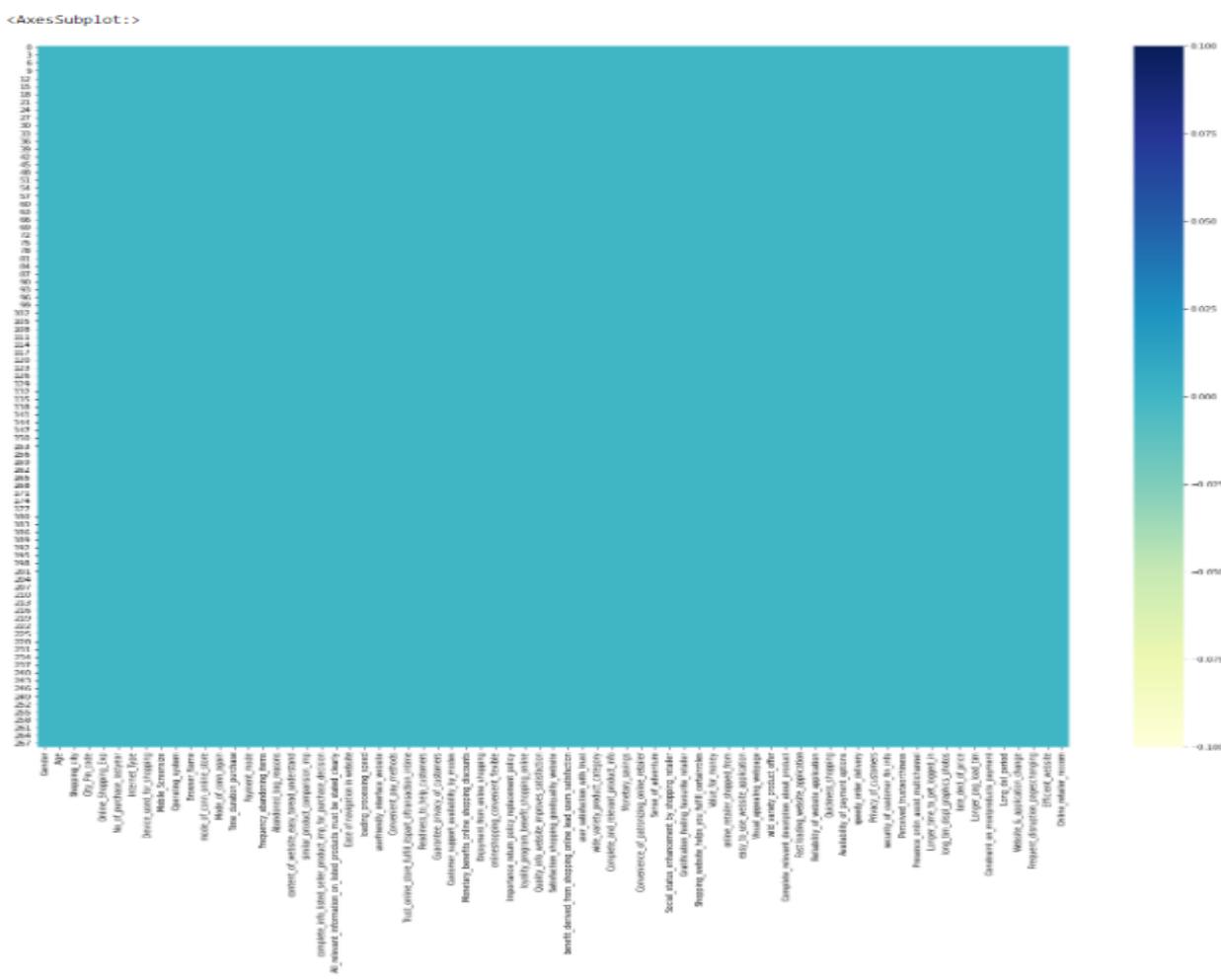
Exploratory Data Analysis (EDA):

- Now was the time to take a look at any kind of missing values or null value that might have been present in our dataset.

```
1 # checking the null values in the dataframe  
2  
3  
4 for i in df:  
5     percentage = np.round((df[i].isnull().sum()/614)*100,2)  
6     print(i,':', percentage, '%')
```

Luckily, I was able to see that there were no missing values in our entire dataset that is prominently visible in the heatmap visuals below.

```
1 # Let's visualize the null values clearly
2 plt.figure(figsize=(25,20))
3 sns.heatmap(df.isnull(),cmap="YlGnBu")
```



- I went ahead to take a look into each record information by making use of df.info () , and nunique methods:

I found there, the data set have 70 object type of columns and only 1 column is integer type (i.e., City_Pin_code).

Since We have mostly categorical data present in the dataset, we are not going to worry about removing outliers or skewness.

- Then I Checked the value counts of each of the columns using For loop:

```
1 #Checking the value counts of each of the columns
2 for i in df.columns:
3     print(f"""
4 Column: '{i}' :
5 {df[i].value_counts()}
6 """)
```

- In column 'Internet_Type', I found there are 2 values for same parameter 'Mobile Internet', so I merged them while processing.

```
1 df['Internet_Type']= df['Internet_Type'].replace('Mobile internet','Mobile Internet')
```

Now, the data got ready for visualization.

Data Visualization:

Data visualization is defined as a graphical representation that contains the information and the data.

Benefits of Good Data Visualization:

Data visualization is another technique of visual art that grabs our interest and keeps our main focus on the message captured with the help of eyes.

Different Types of Analysis for Data Visualization are:

1. **Univariate Analysis:** In the univariate analysis, we will be using a single feature to analyse almost all of its properties.
2. **Bivariate Analysis:** When we compare the data between exactly 2 features then it is known as bivariate analysis.
3. **Multivariate Analysis:** In the multivariate analysis, we will be comparing more than 2 variables.

I. Univariate Analysis:

I made use of 'for loop' to generate count plots and pie charts for all our columns showing the percentage of data coverage.

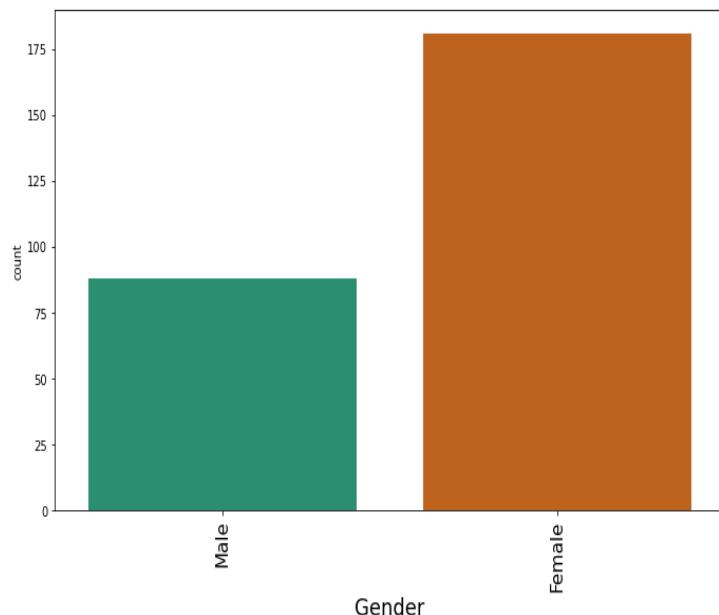
```
1 for col in categorical_col:
2     plt.figure(figsize=(25,7))
3
4     plt.subplot(1,2,1)
5
6     sns.countplot(df[col], palette ="Dark2" )
7
8     plt.title(f'count plot for "{col}"\n',fontsize = 20)
9     plt.xlabel(col,fontsize = 18)
10    plt.xticks(fontsize = 15,rotation=90)
11
12
13
14    plt.subplot(1,2,2)
15
16    df[col].value_counts().plot.pie(title=f'count plot for "{col}"\n', autopct='%.1f%%', fontsize=15)
17    plt.title(f' count plot for "{col}"\n',fontsize = 20)
18    plt.show()
```

This piece of code generated multiple count plots and pie charts of each column; images as displayed below.

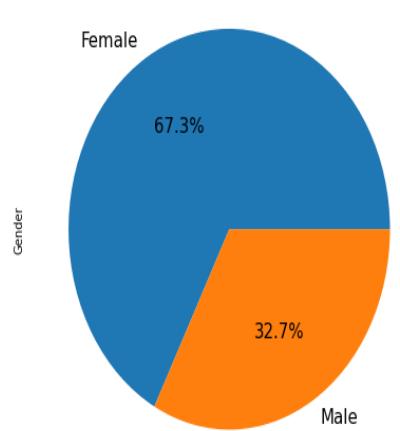
❖ Observations and Data Analysis of Univariate data visualization:

1. Analysis of Customer Gender:

count plot for "Gender"



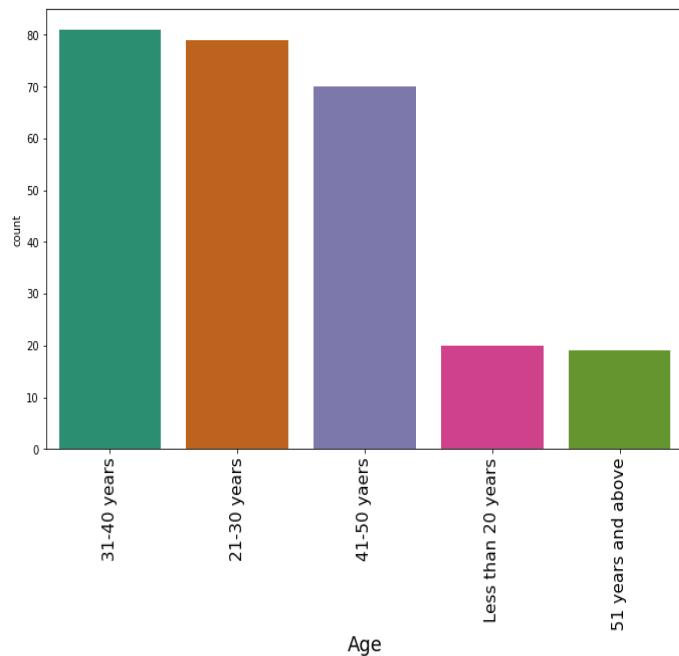
count plot for "Gender"



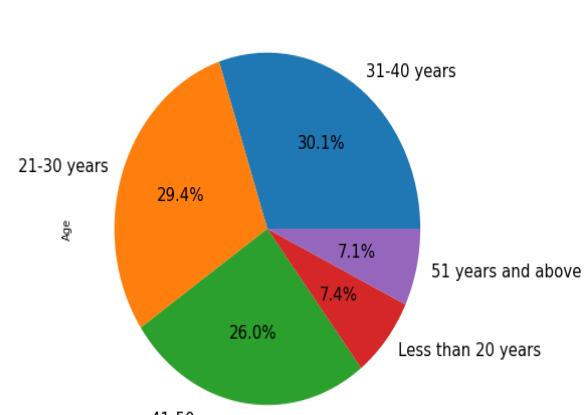
- plots shows that 67.3% of the customers visiting are females where as males are only 32.7%.
- Which means Females are more likely to shopping online than Males.
- If companies want to pull the Male customers also for online shopping, they can provide offers on products for men.

2. Analysis of Customer Age:

count plot for "Age"

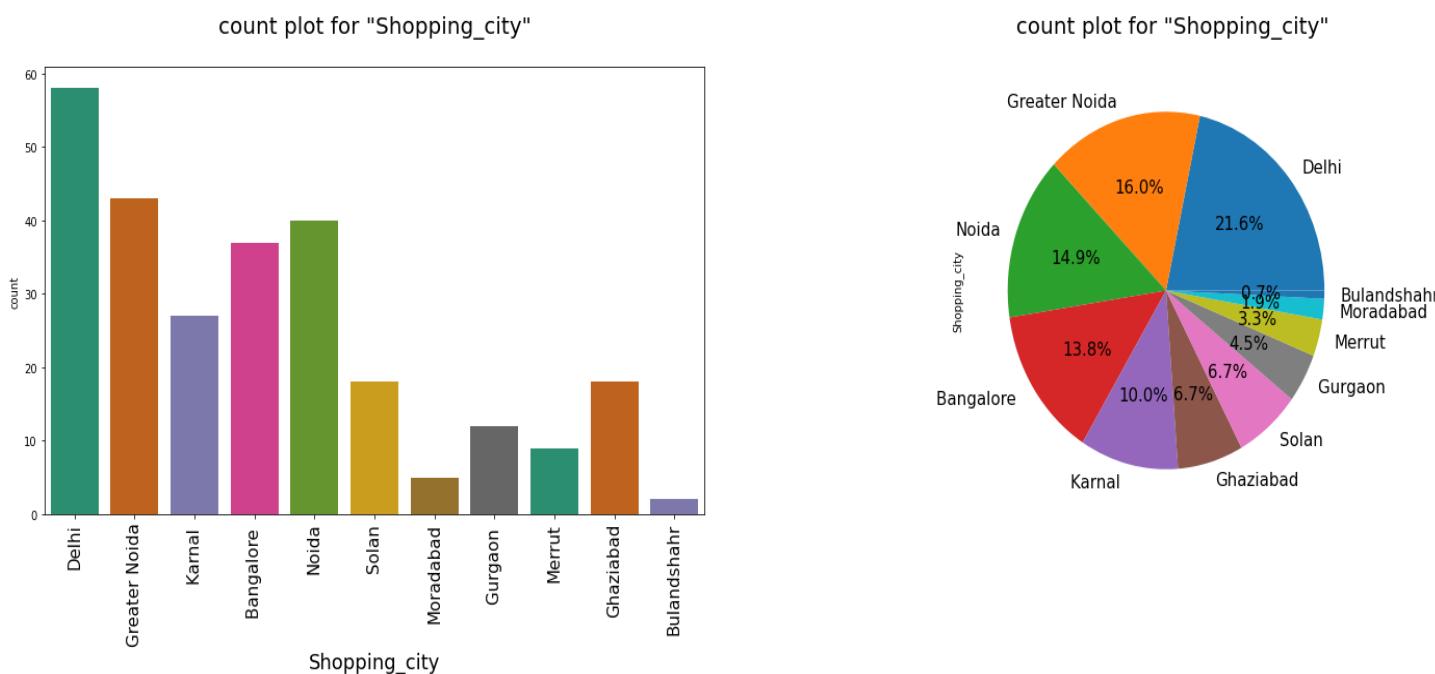


count plot for "Age"



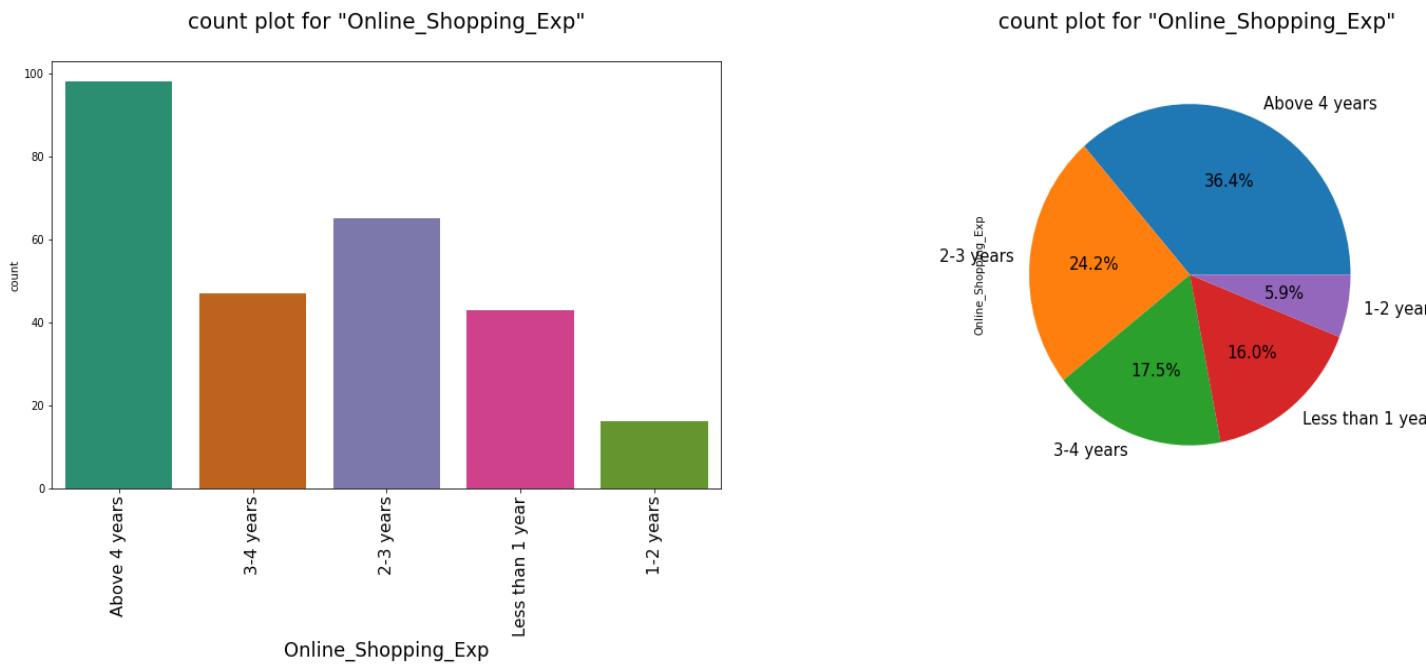
- From the plots we can clearly see that most of the respondents fall into the following three categories: -
 - 31-40 years (30.1%),
 - 21-30 years (29.4%),
 - 41-50 years (26%)
- Very few respondents fall into the categories: - (may be because of lack of awareness of e. shopping)
 - Less than 20 years (7.4%)
 - 51 years and above. (7.1%)
- Which means, the population who does the online shopping belongs in 20 to 50 age-group.

3. Analysis of Customer city:



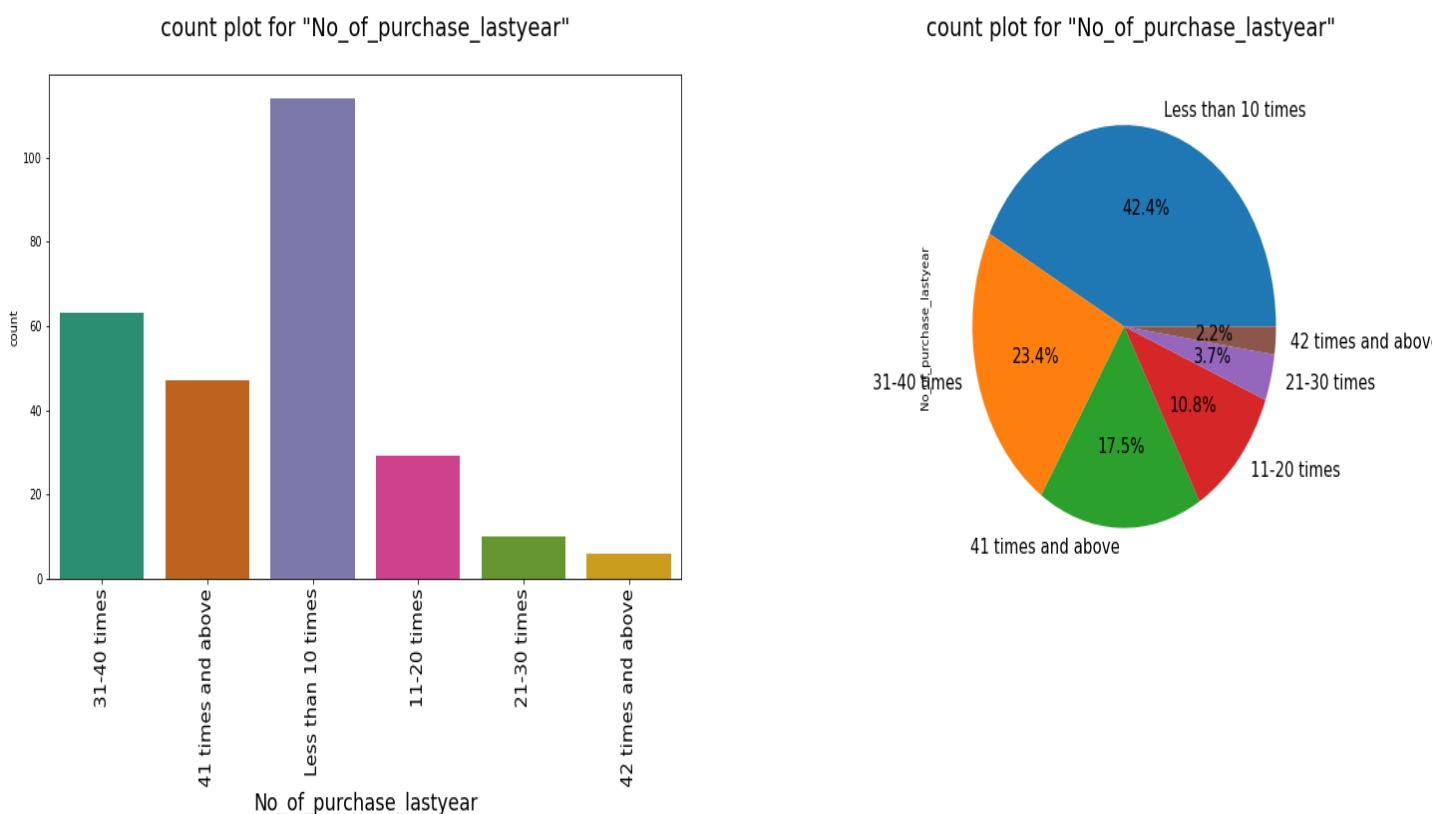
- We can see a greater number of customers are from Delhi followed by Noida, Greater Noida and Bangalore and then Karnal.
- It would be beneficial, if companies consider to develop a business strategies for small cities also.

4. Analysis of Customers Online Shopping Experience:



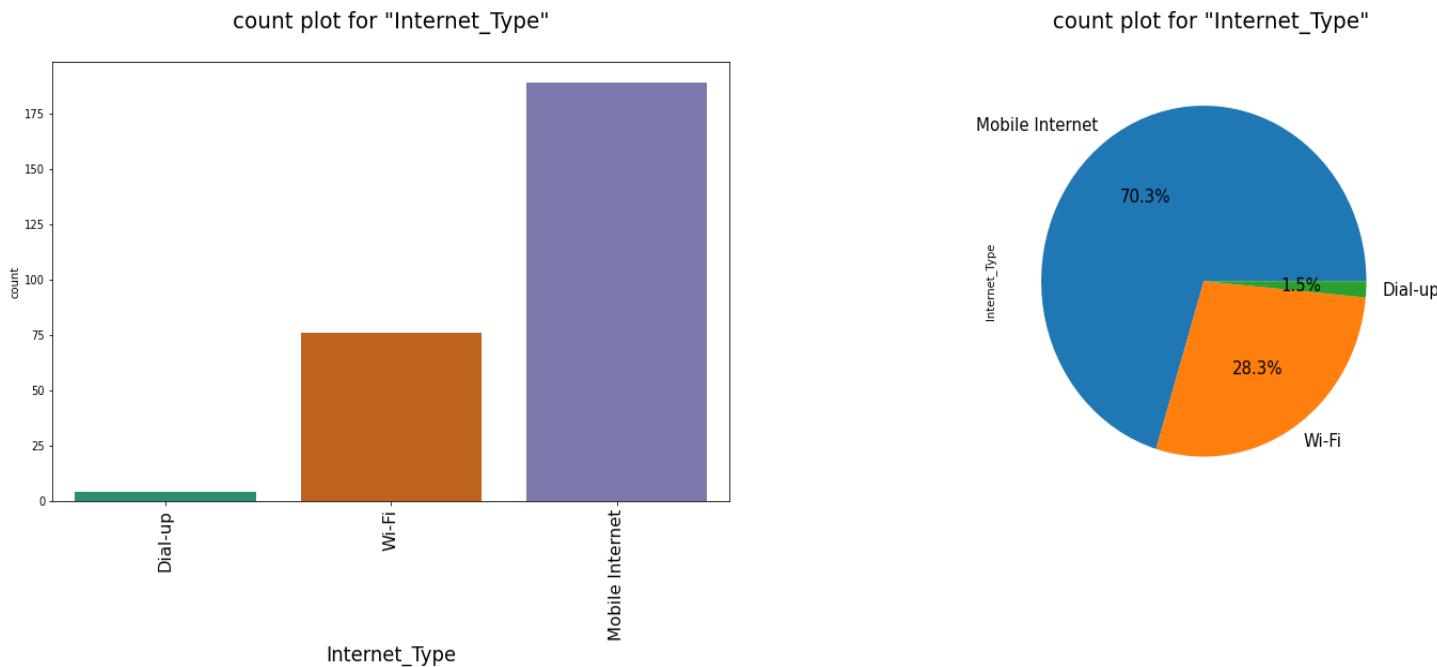
- We can see that most of the online shoppers of our dataset have more than 4 years' experience (36.4%), followed by 2-3 years (24.2%), 3-4 years (17.5%).
- Very few of the shoppers have experience of 1-2 years.

5. Number of purchase last year:



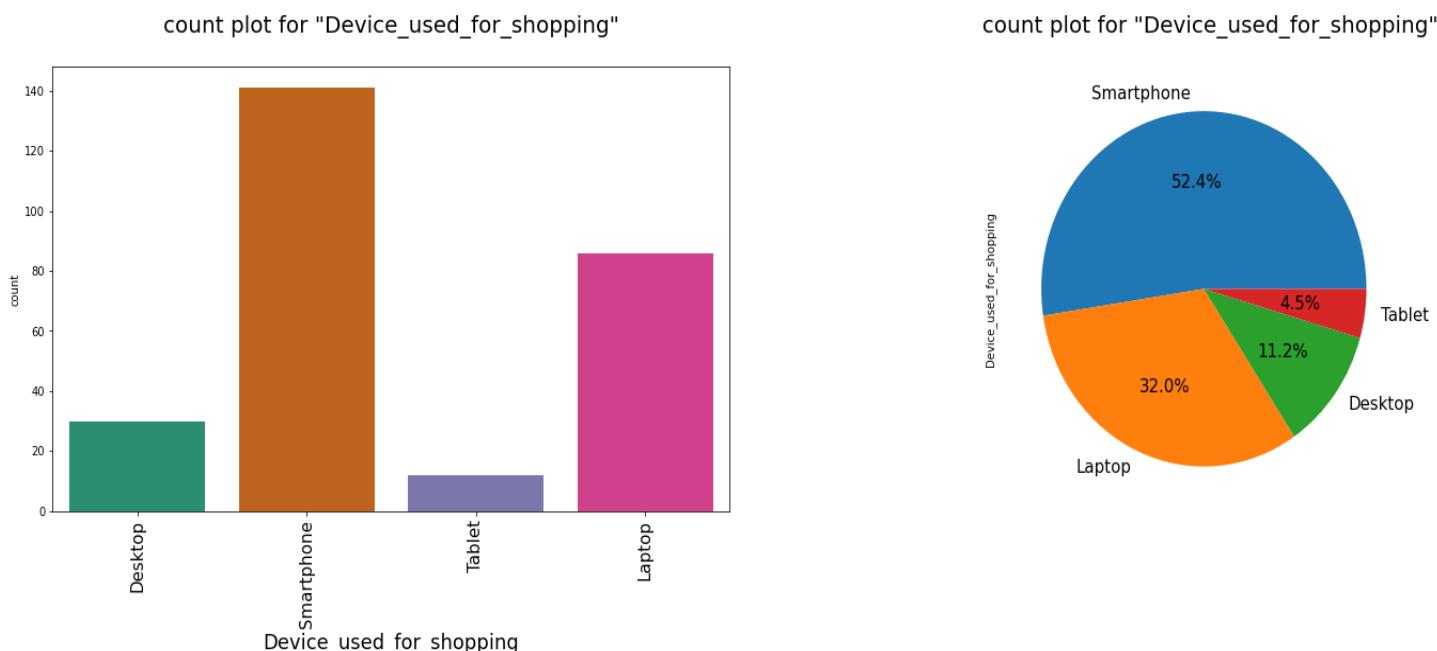
- We can see, there are more than 42% of peoples who have purchased less than 10 times in online shopping.
- Also 23.4% Customers Purchased product 31-40 times in past 1 year.

6. Analysis of Internet type:



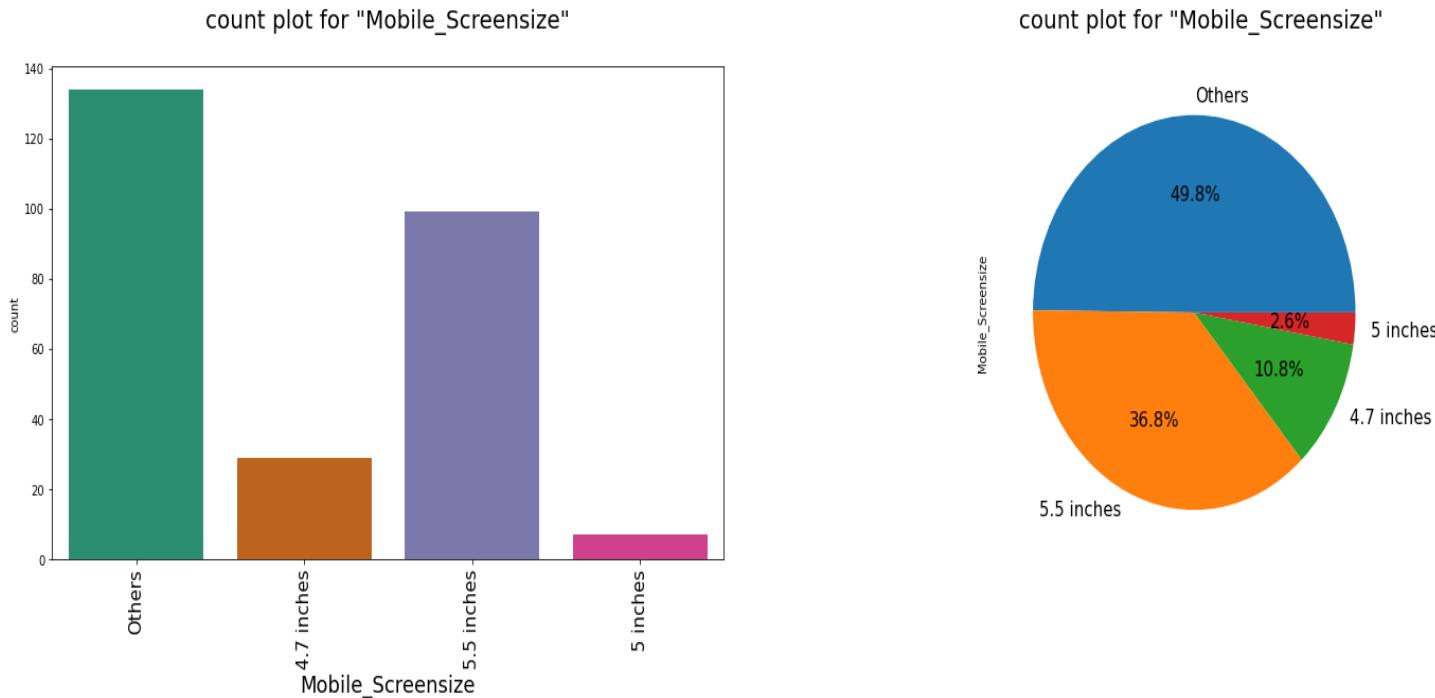
- Most of people depend on "Mobile Internet" (more than 70%) & followed by "Wi-Fi" (28.3%) for the online shopping.

7. Analysis of Device_used_for_shopping:



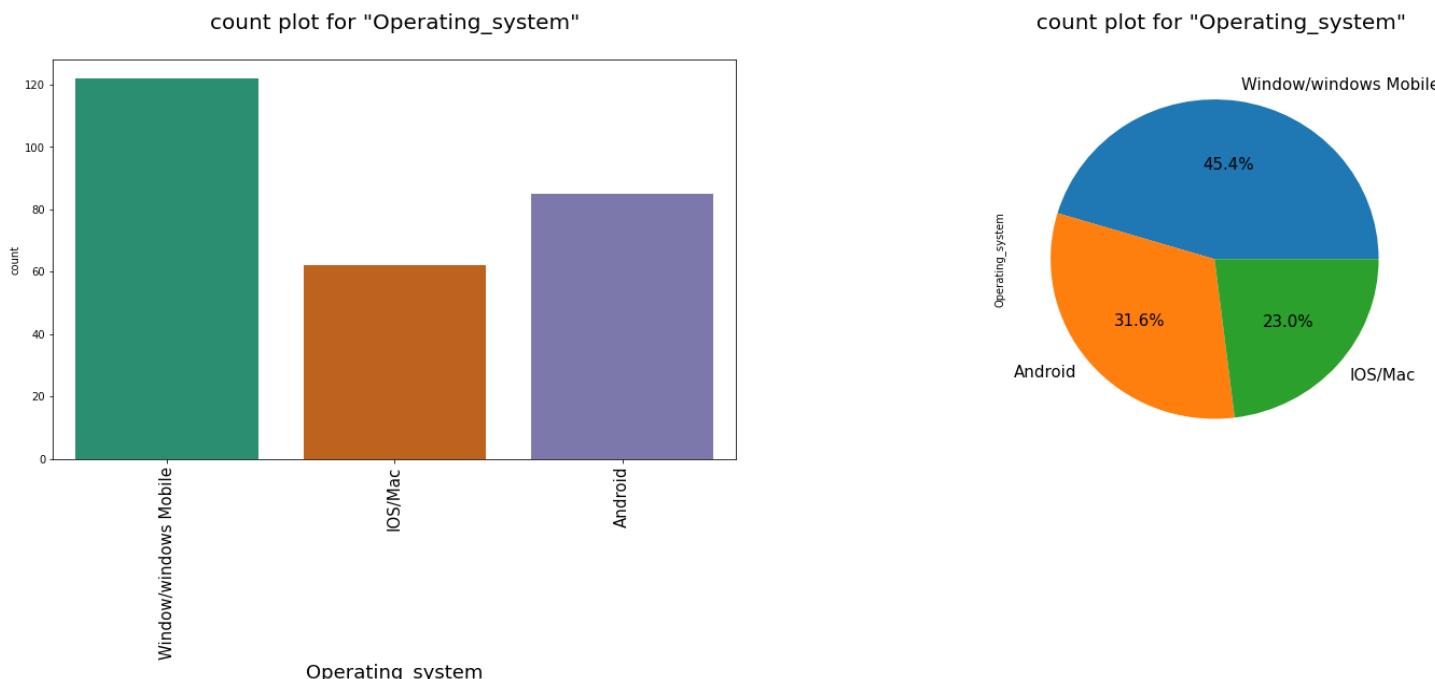
- Smartphone (52.4%) is the most popular device followed by the Laptop (32%).

8. Analysis of Mobile Screen size:



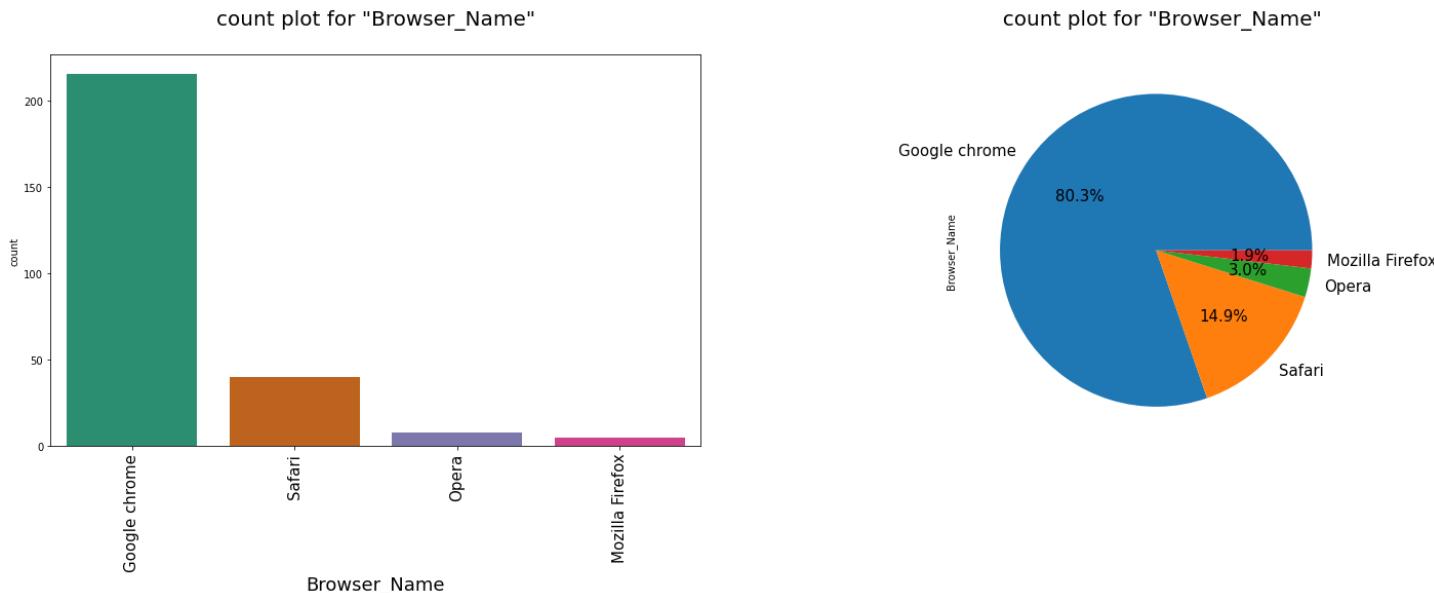
- The data which we have it is saying most people have screen size apart from 4.7", 5.5" & 5 inches.
- Next to that most people have 5.5 inches of screen size of the mobile.

9. Analysis of Operating system:



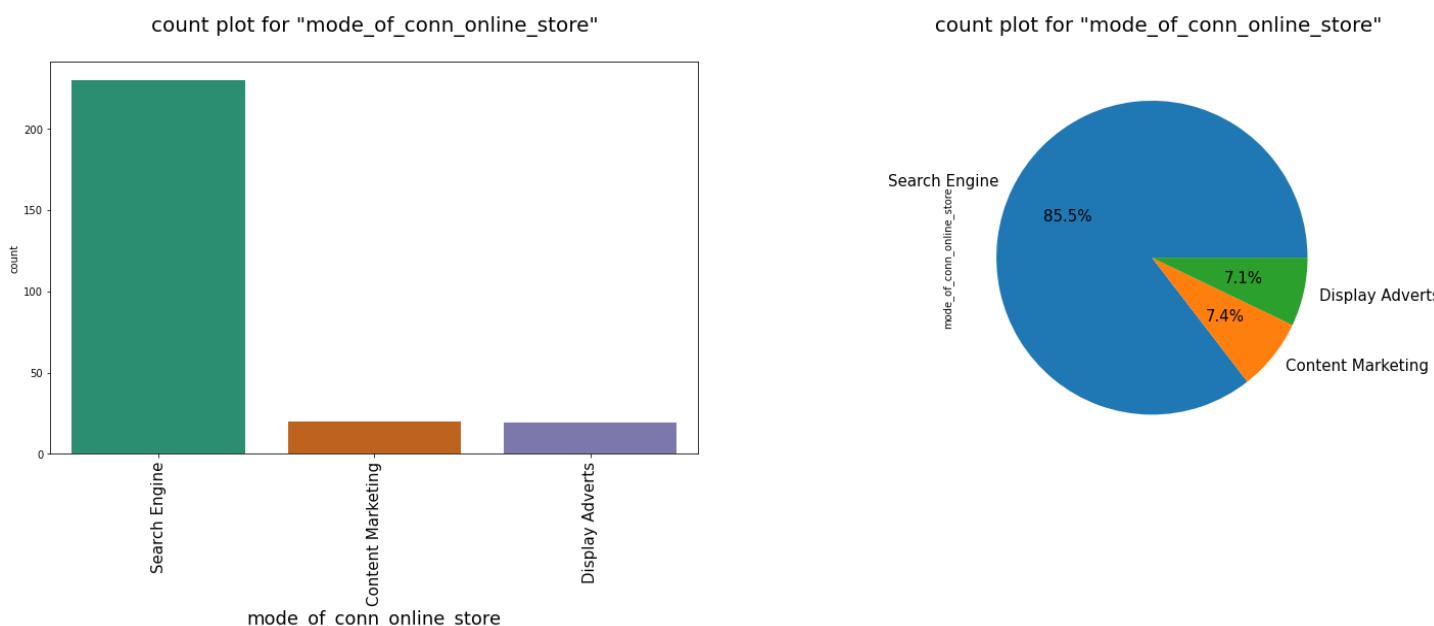
- We can see a greater number of Windows users, followed by Android and IOS/MAC. It would be beneficial if companies consider to develop a windows app.

10. Analysis of Browser_Name:



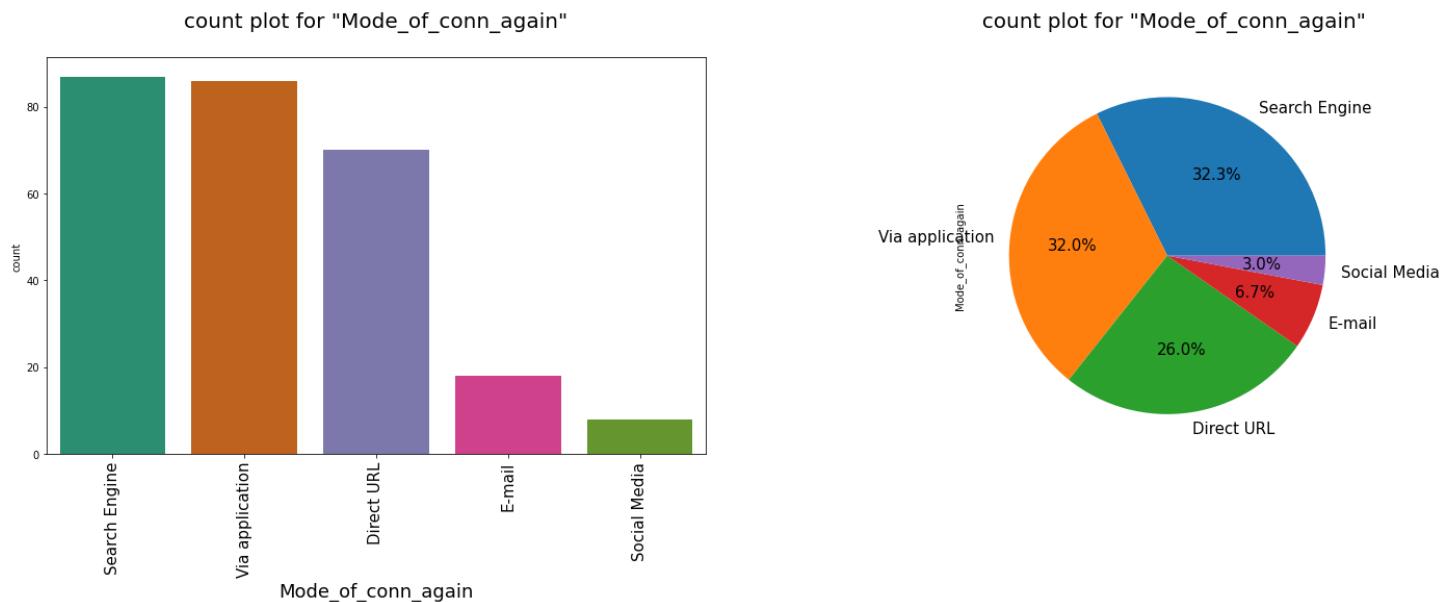
- We can see Google Chrome has the monopoly in browsing e-shopping Websites (80.3%).

11. Analysis of mode of connecting online store:



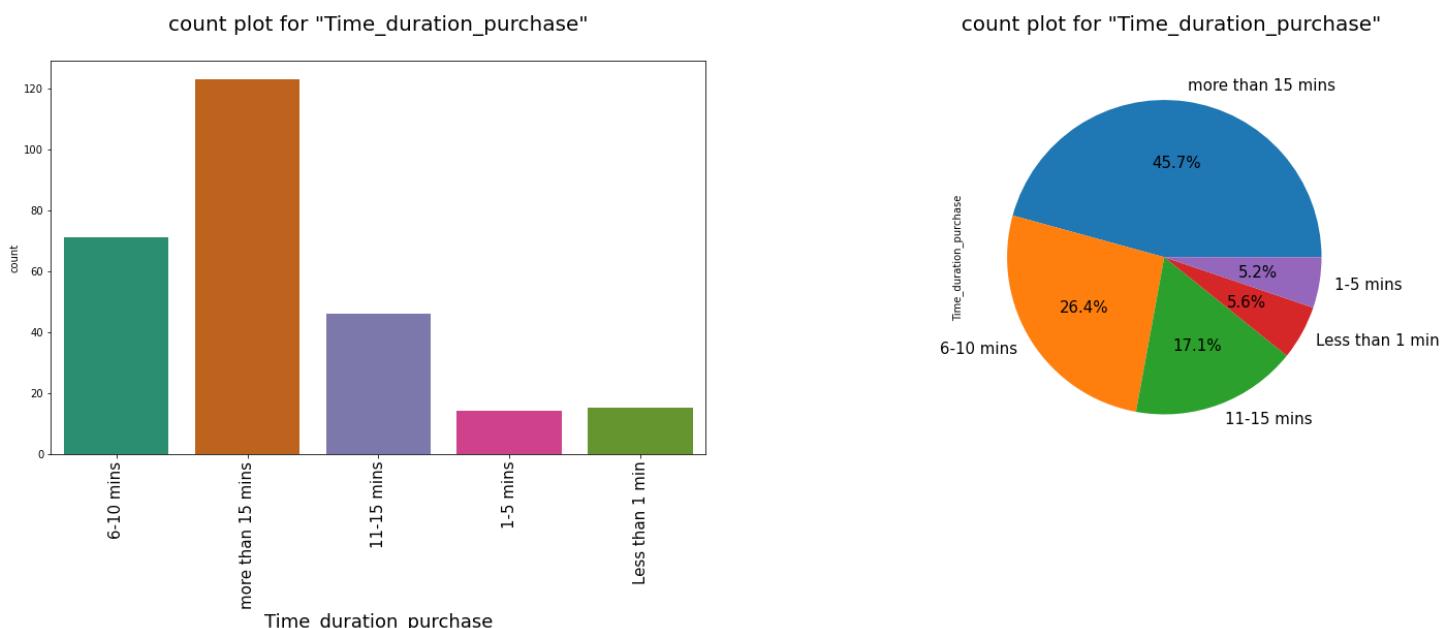
- We can see most of the Customers use Search engines(85.5%) to reach out to the website and most used search engine is Google Chrome.

12. Analysis of Mode of connecting to the store after first time:



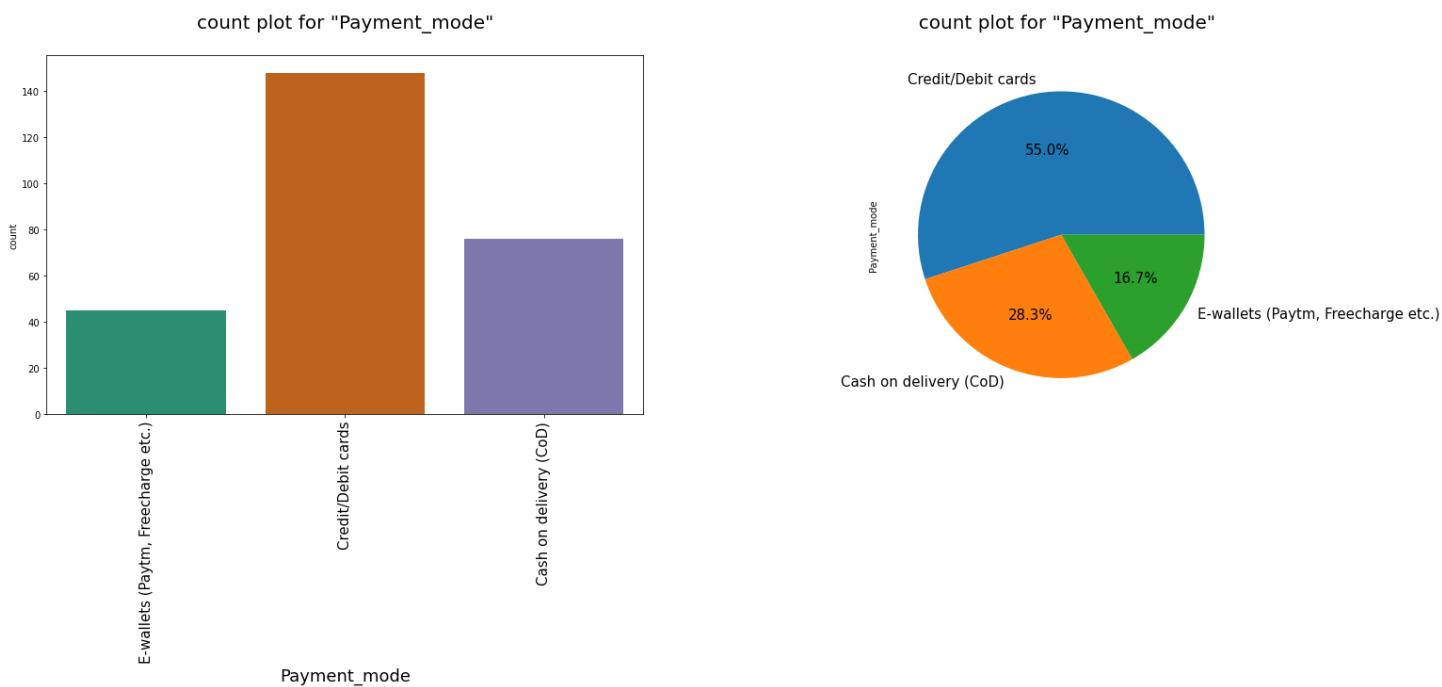
- We can see most of the customers use website (32.3%) and then App (32%) most of the time, some of the customers use direct URL to revisit the website.

13. Analysis of Time_duration_purchase:



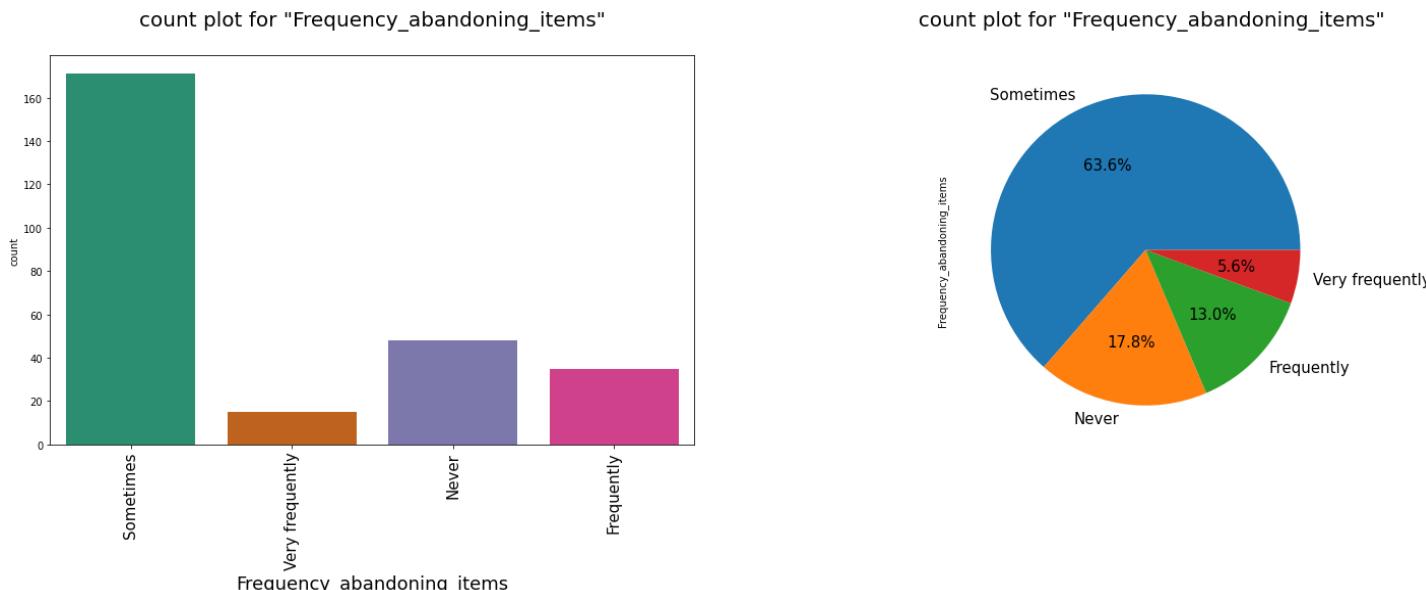
- We found that more than 45% of peoples spend more than 15 Min to make the decision of purchase when they visit an online retail store.

14. Analysis of Payment_mode:



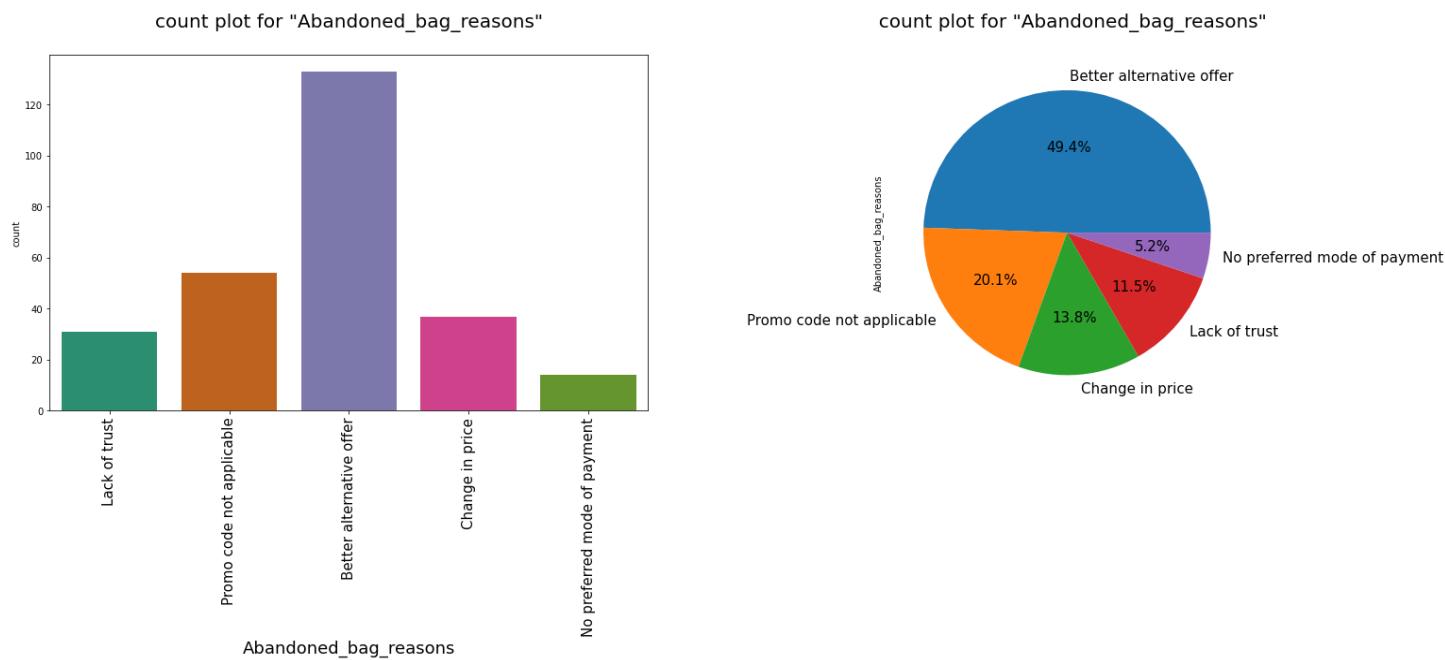
- We found that most of customers prefer the Credit/Debit card option for online payment.
- After Credit/Debit card comes the Cash on Delivery (COD).
- Least people go with E-wallets (Paytm, Freecharge).

15. Analysis of Frequency_abandoning_items:



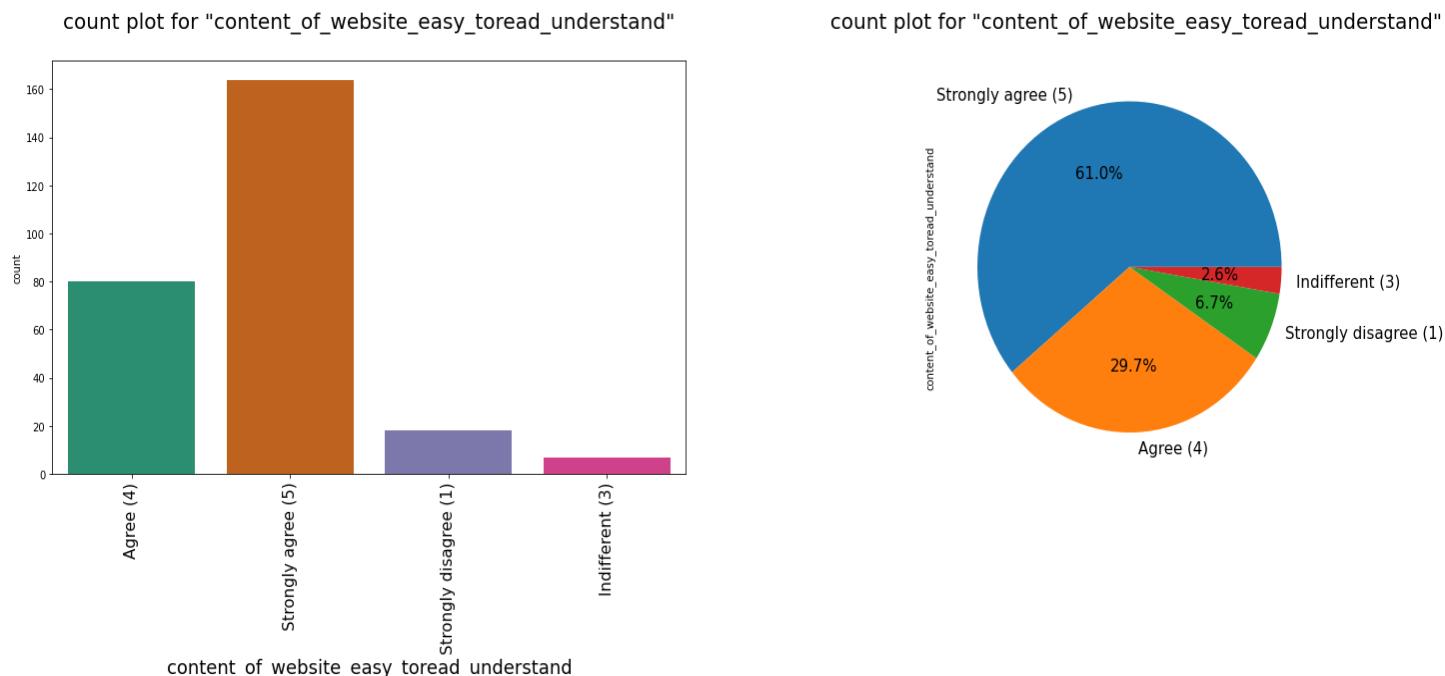
- Sometimes (63.6%) most of the people do abandon the items in list while shopping online.

16. Analysis of Abandoned_bag_reasons:



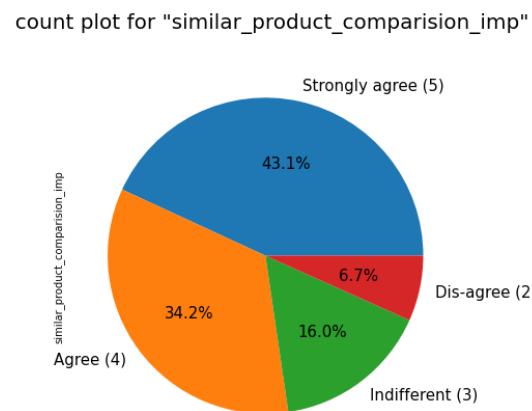
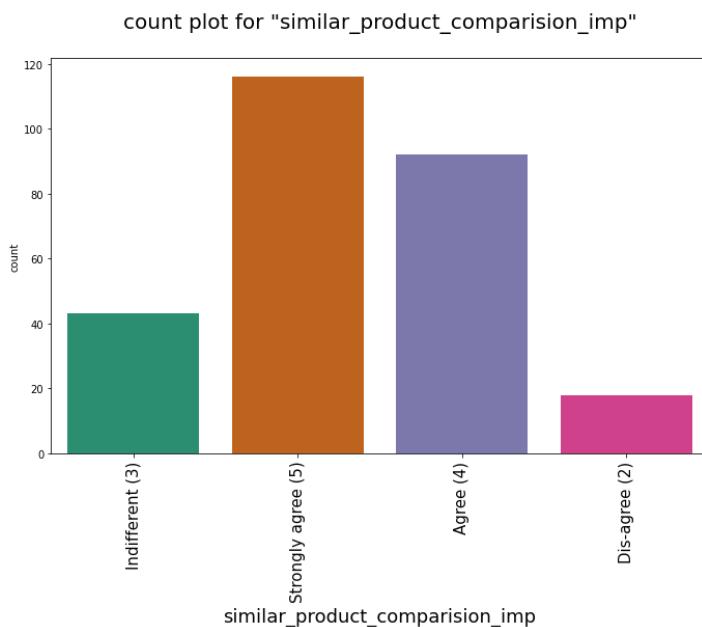
- We can see most of the customers abandon the cart because they get an better alternative offers, also some customers found their promocode is not applicable for the payment.

17. Analysis of content_of_website_easy_toread_understand:



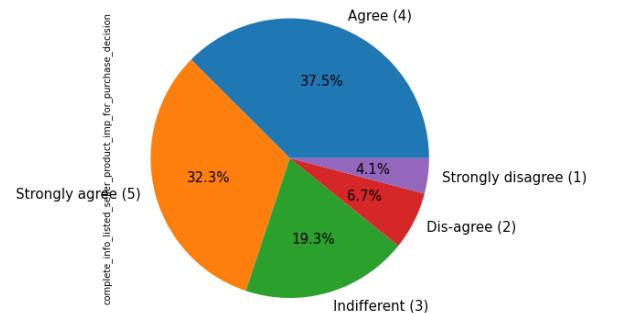
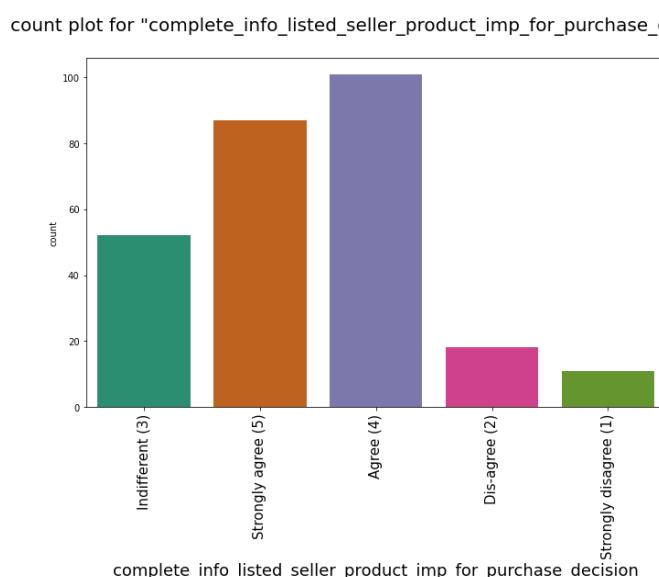
- 61% people strongly agree, and almost 30% are agree on, websites should be easy to understand.
- so, if companies want to increase their sales their websites should be user-friendly and understandable.

18. Analysis of similar_product_comparision_imp:



- Most of the Peoples say that similar Product comparison should be shown on website/App.

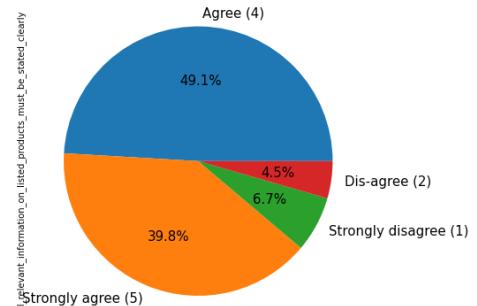
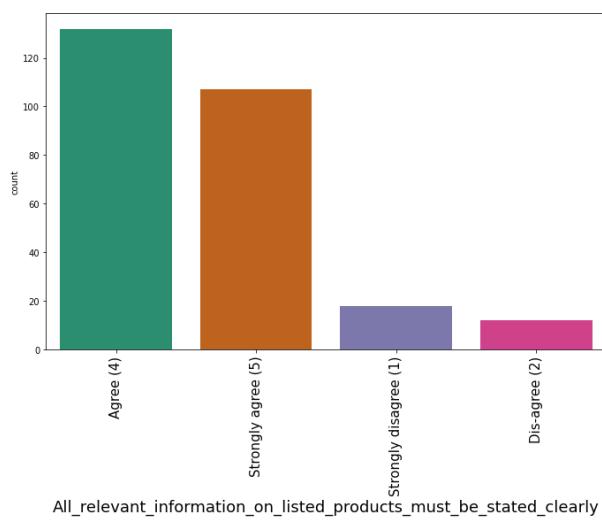
19. Analysis of complete info listed seller product imp for purchase decision:



- Most of the customers (70%) thinks that the complete information about the product helps to make a buying the product.

20. All_relevant_information_on_listed_products_must_be_stated_clearly:

count plot for "All_relevant_information_on_listed_products_must_be_stated_clearly"

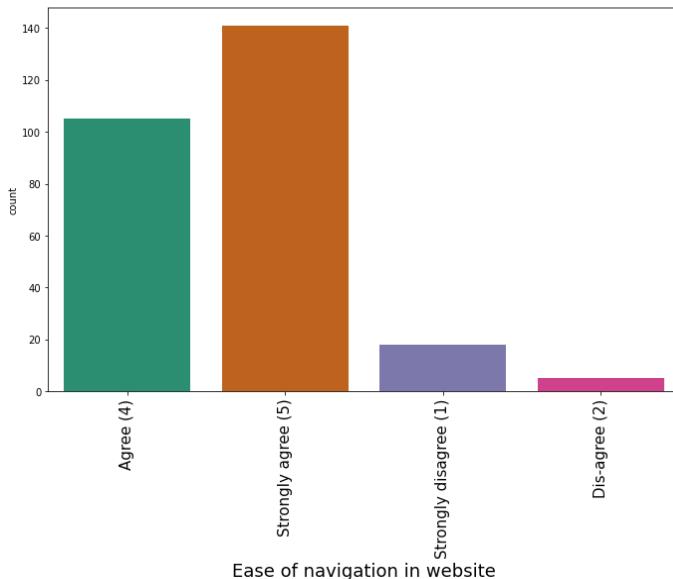


All_relevant_information_on_listed_products_must_be_stated_clearly

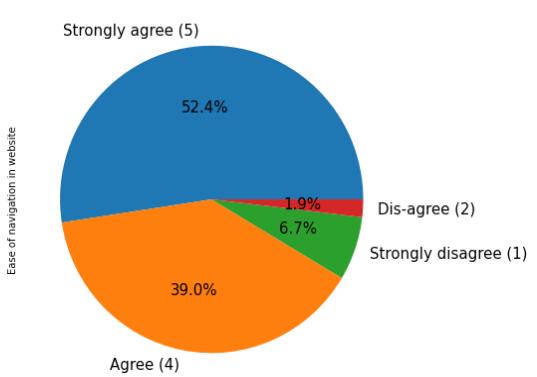
- almost 90% of population agree on the point: 'All relevant information on listed products must be stated clearly'.

21. Analysis of Ease of navigation in website:

count plot for "Ease of navigation in website"

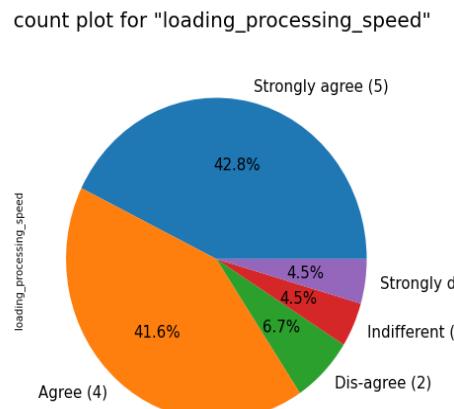
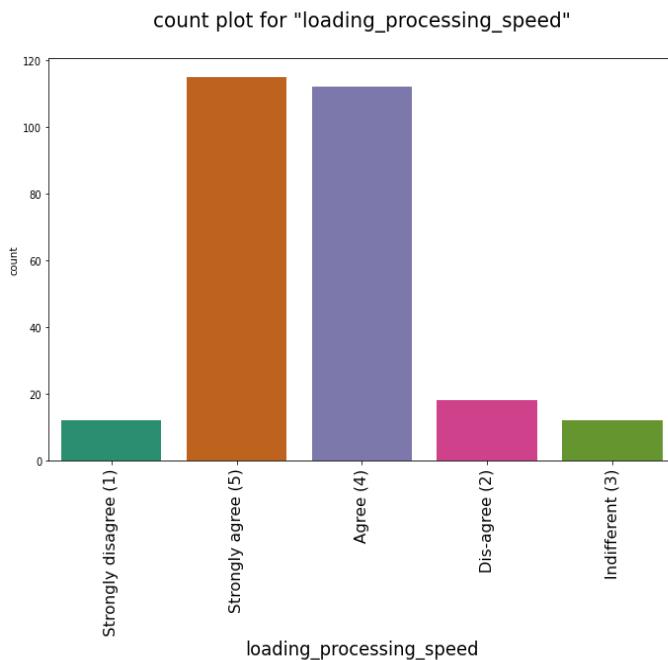


count plot for "Ease of navigation in website"



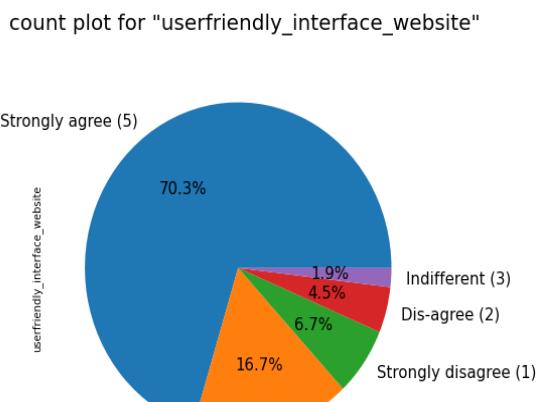
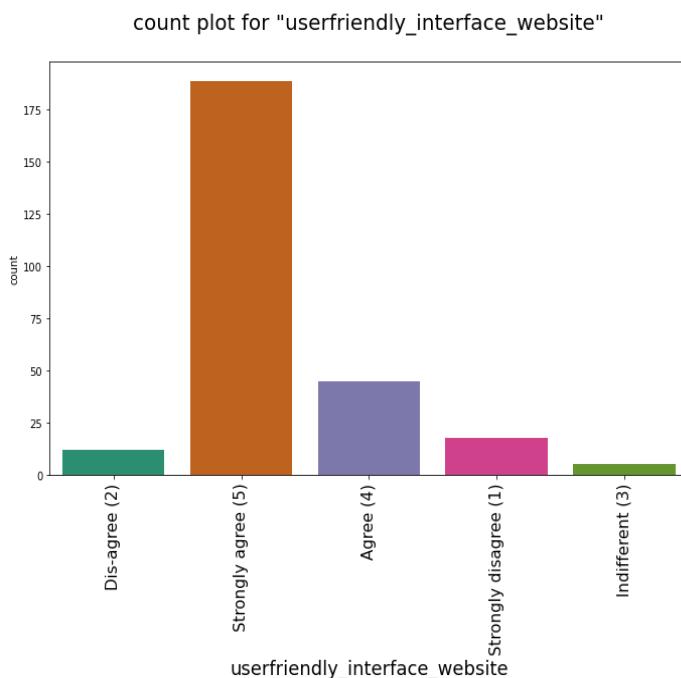
- Mostly people strongly agree that "ease in navigation of website" is very important.
- Very less people dis-agree with "Important of ease in navigation of website".

22. Analysis of loading_processing_speed:



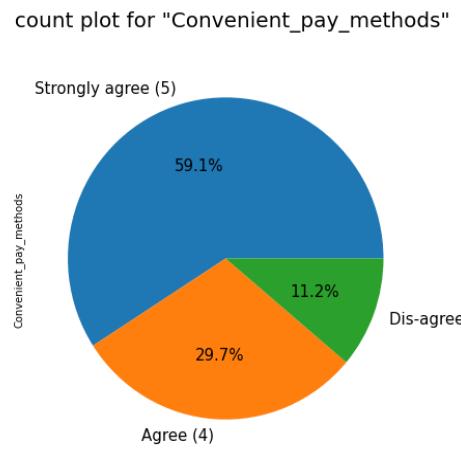
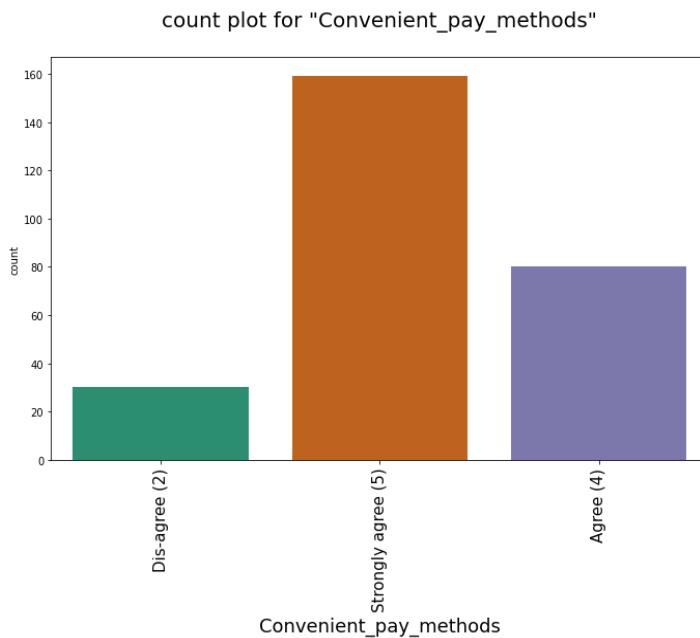
- Majority of Customers agrees on loading and Processing speed should be good.

23. Analysis of userfriendly_interface_website:



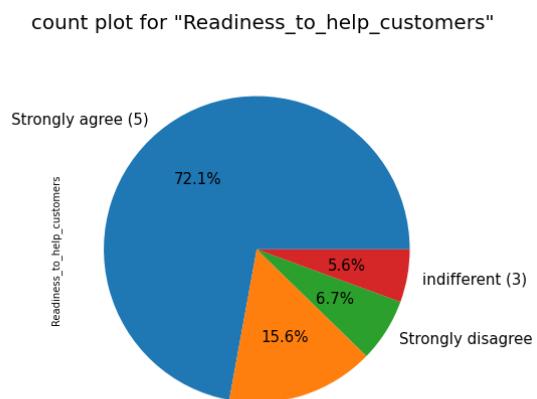
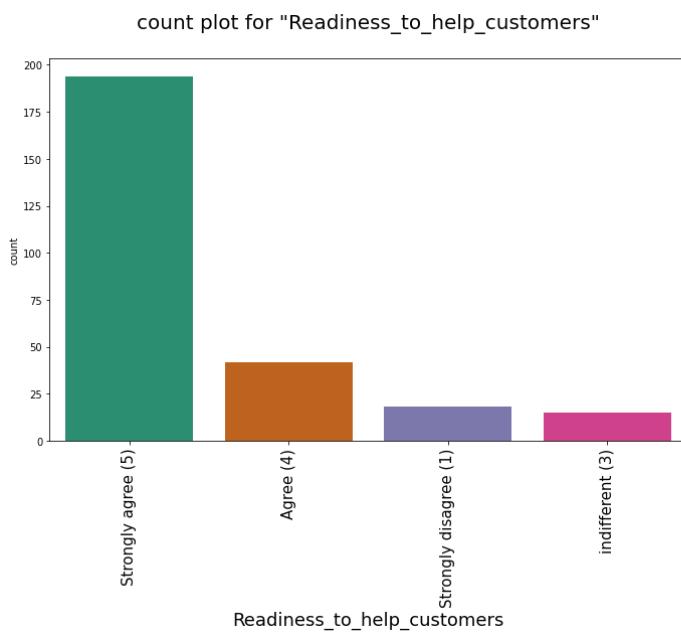
- We can see that most of our respondents strongly agree to the point that "user-friendly interface of the website" is very important.

24. Analysis of Convenient_pay_methods:



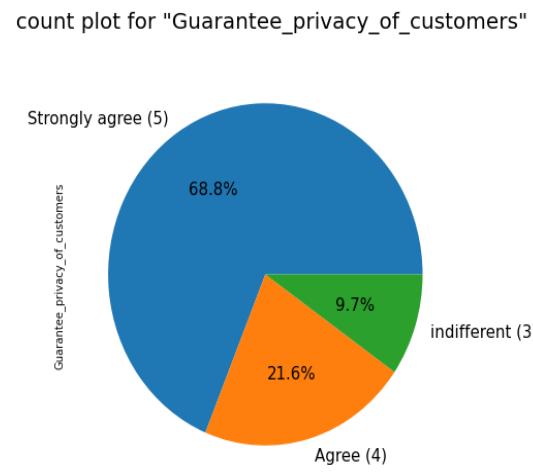
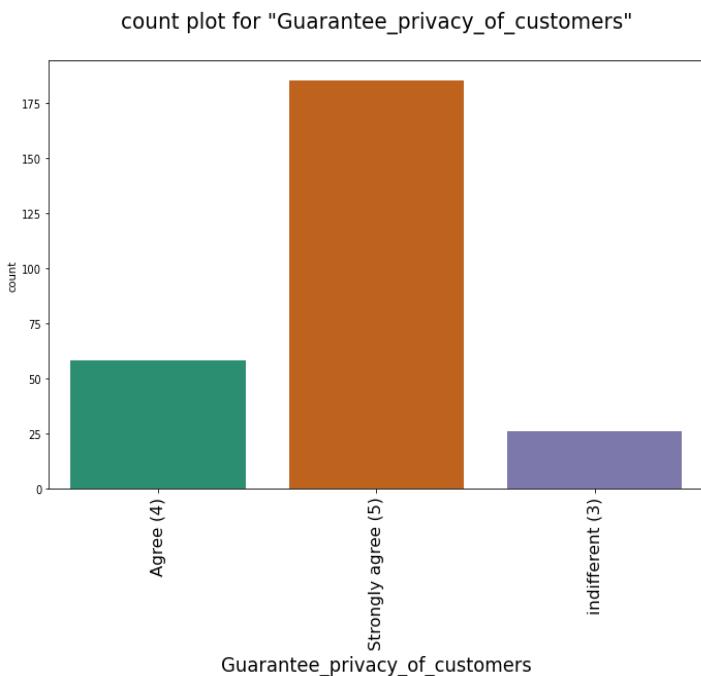
- Many of the respondents strongly believe that if payment methods are convenient, then it will be good for online shoppers.

25. Analysis of Readiness_to_help_customers:



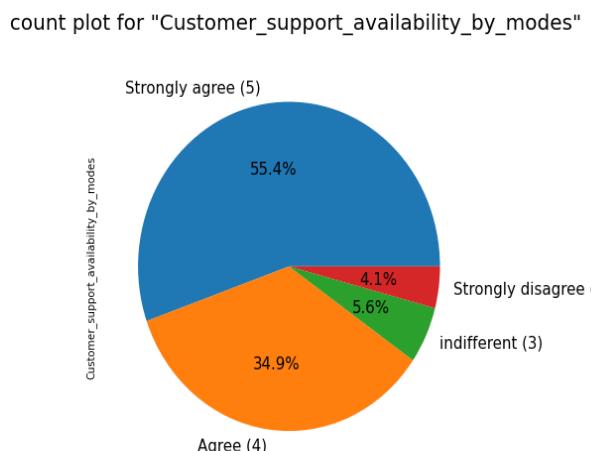
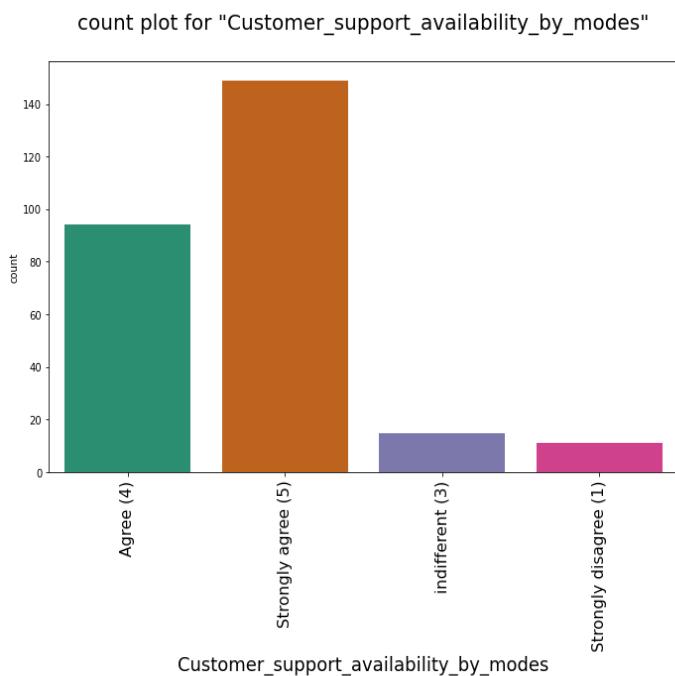
- Most of Customers firmly believe that approach of "Readiness to help the customers always" is the most important concern.

26. Analysis of Guarantee_privacy_of_customers & Trust_online_store_fulfill_itspart_of_transac tion_intime:



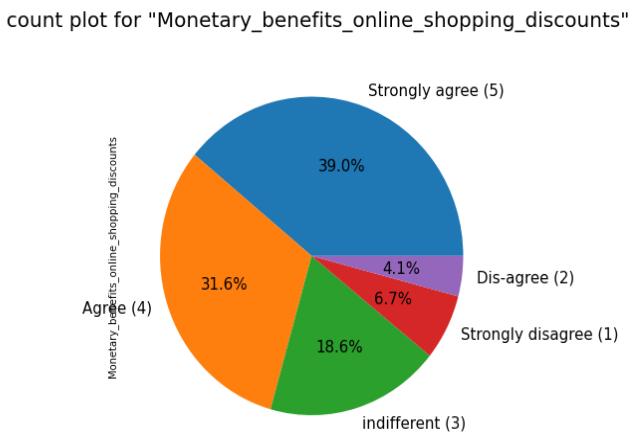
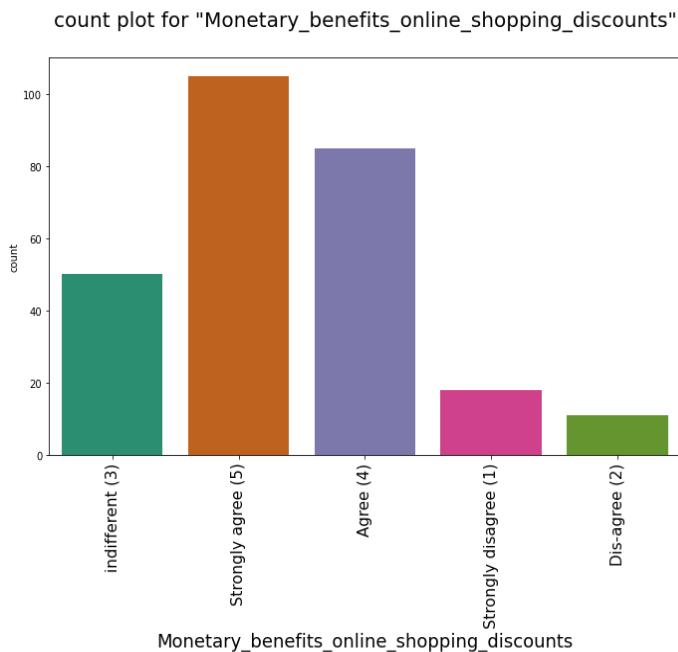
- As we all know that privacy is the most vital thing as all the customers data like credit-card number, debit-card number, are stored in the retailer's database, and customer believes on the company that it will not breach the private data policy.
- Our records also shows that most of customers strongly believe that online store must guarantee the privacy of the data of customers.

27. Analysis of Customer_support_availability_by_modes:



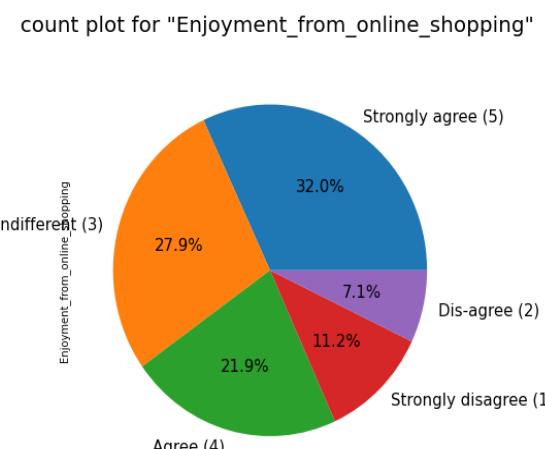
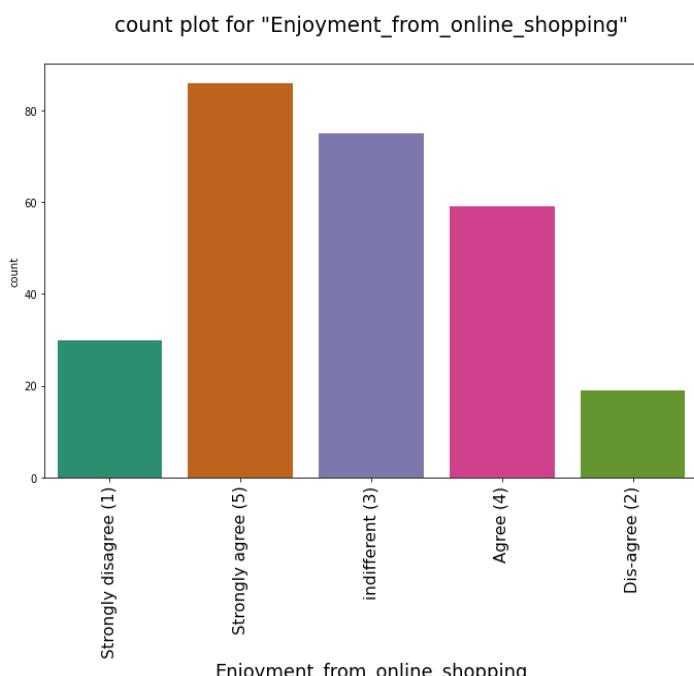
- Majority of the customers expect that their online store must provide the assistance of 'customer support' to them through various modes like ("Email", "Text messages", "Phone Call").

28. Analysis of Monetary_benefits_online_shopping_discounts:



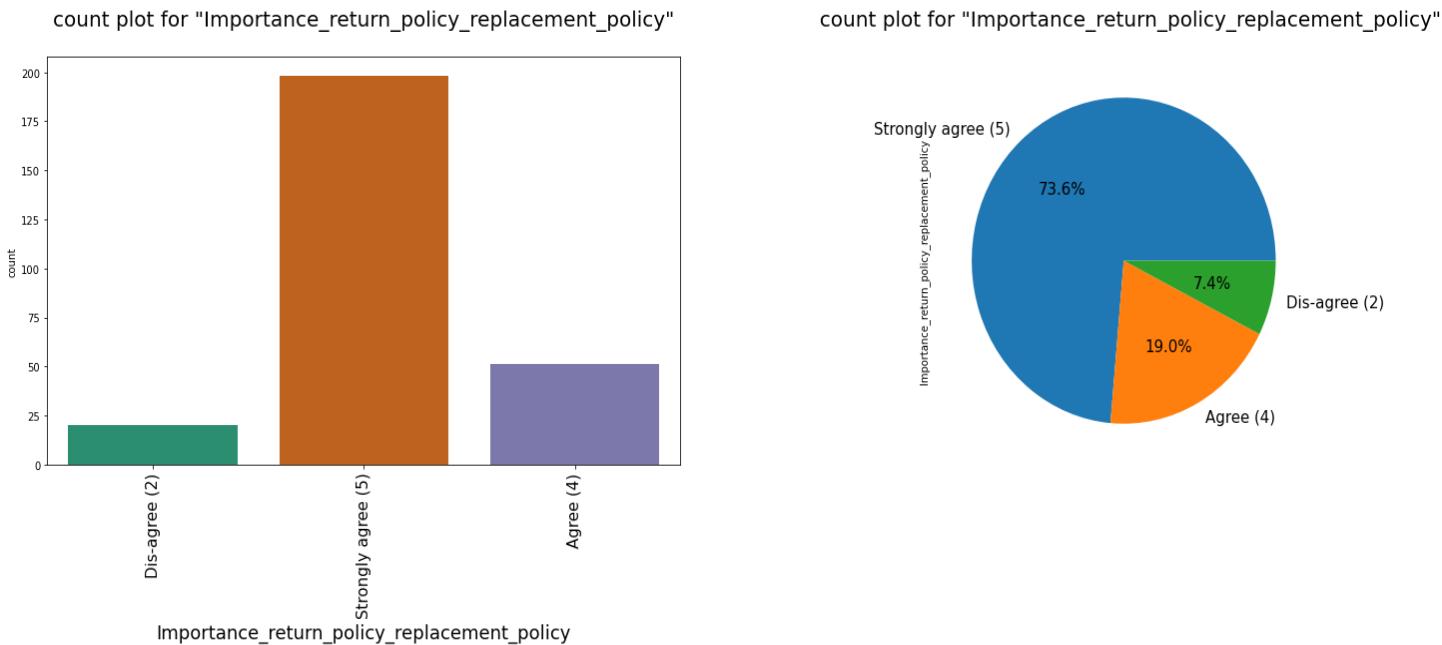
- Mostly of the customers are inclined towards the perks of monetary benefits while shopping in the forms of coupons, discount offers etc.

29. Analysis of Enjoyment_from_online_shopping:



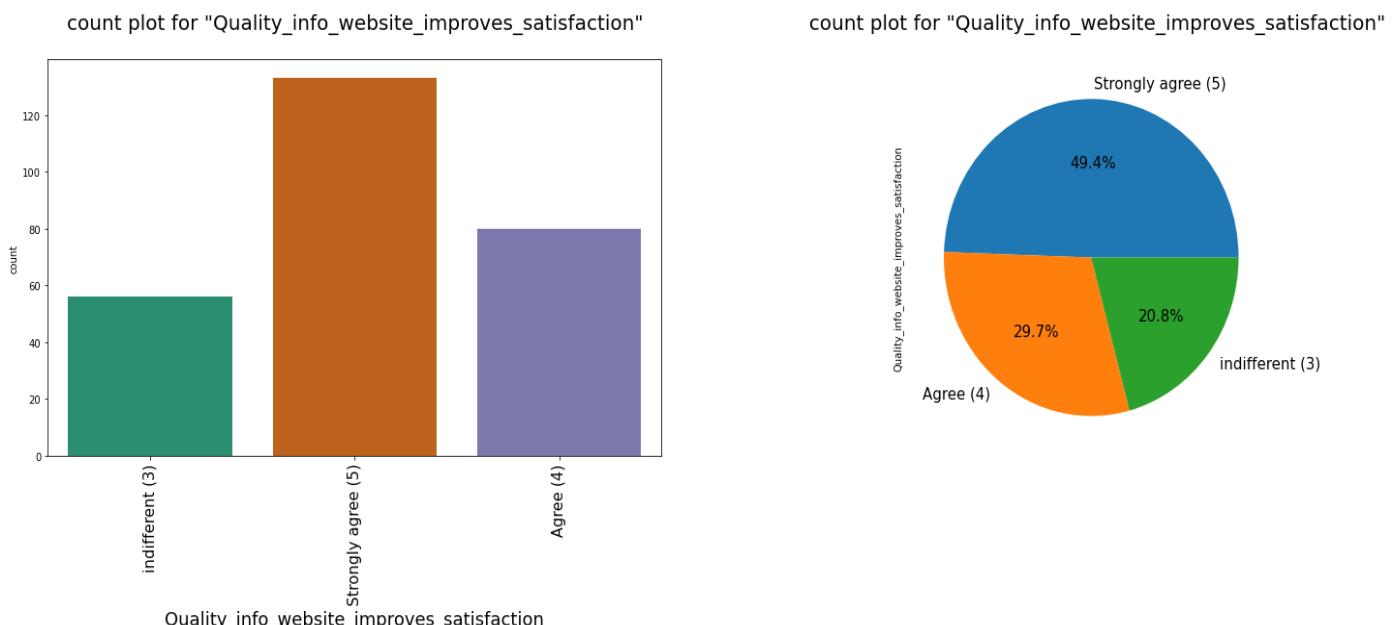
- Most people agree that they enjoy the online shopping.
- Many of respondents are also unconcerned about the enjoyment.
- Very few of them disagree with the fact of enjoyment while online shopping.

30. Analysis of Importance_return_policy_replacement_policy:



- As we all know that replacement and return policy is very much important for a customer since if the product which he/she buy didn't performing well then, they can return it to the seller without any problem by giving proper reason of return of product.
- Most of the customers are agreeing with the fact that there must be customer friendly return or replacement policy must be present in the company.

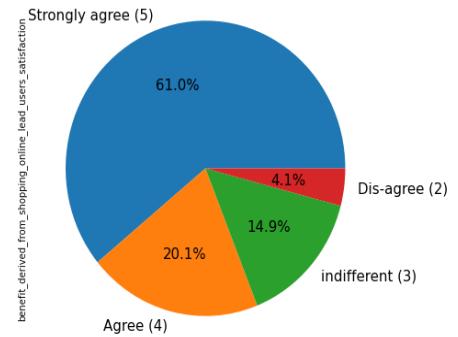
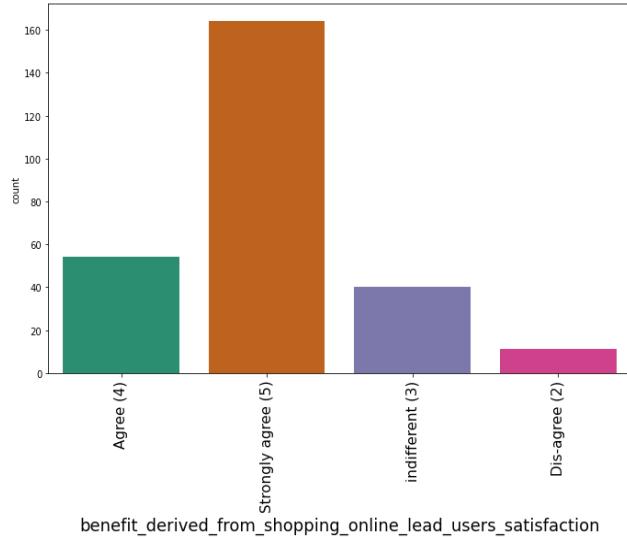
31. Analysis of Quality_info_website_improves_satisfaction:



- We all know that quality information about the product helps us to understand the product well and we can decide upon reading the details whether to go for the shopping or not.
- Customers are strongly agreeing with importance of the presence of the quality information present in the dashboard improves the satisfaction of the customers.

32. Analysis of benefit_derived_from_shopping_online_lead_users_satisfaction:

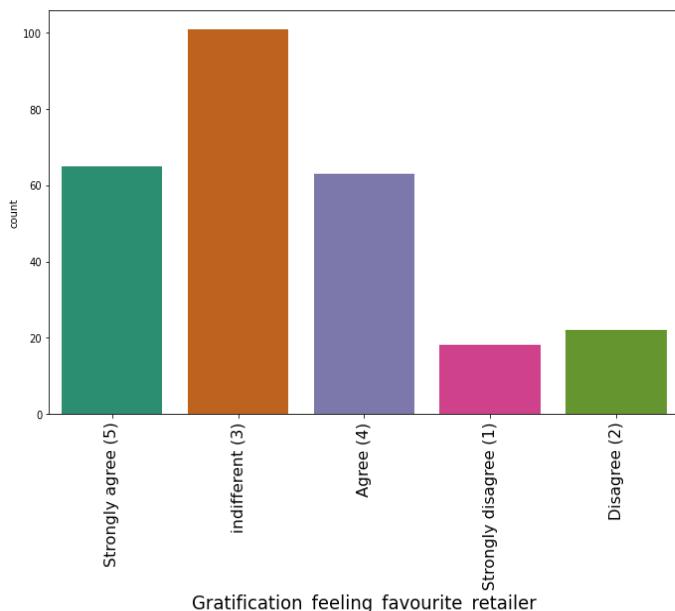
count plot for "benefit_derived_from_shopping_online_lead_users_satisfaction" count plot for "benefit_derived_from_shopping_online_lead_users_satisfaction"



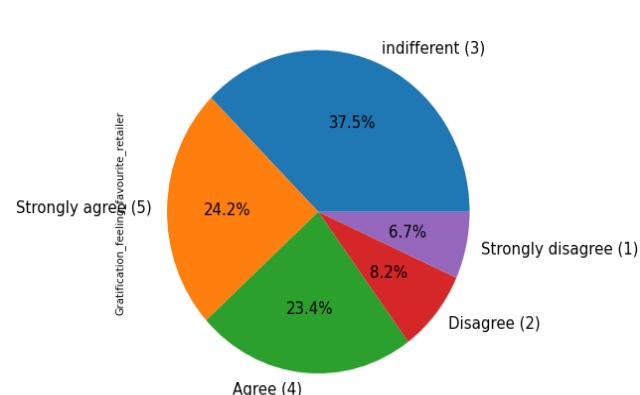
- We can see that from the plots that many customers believe that benefit derived from the online shopping leads to customer satisfaction.

33. Analysis of Gratification_feeling_favourite_retailer:

count plot for "Gratification_feeling_favourite_retailer"

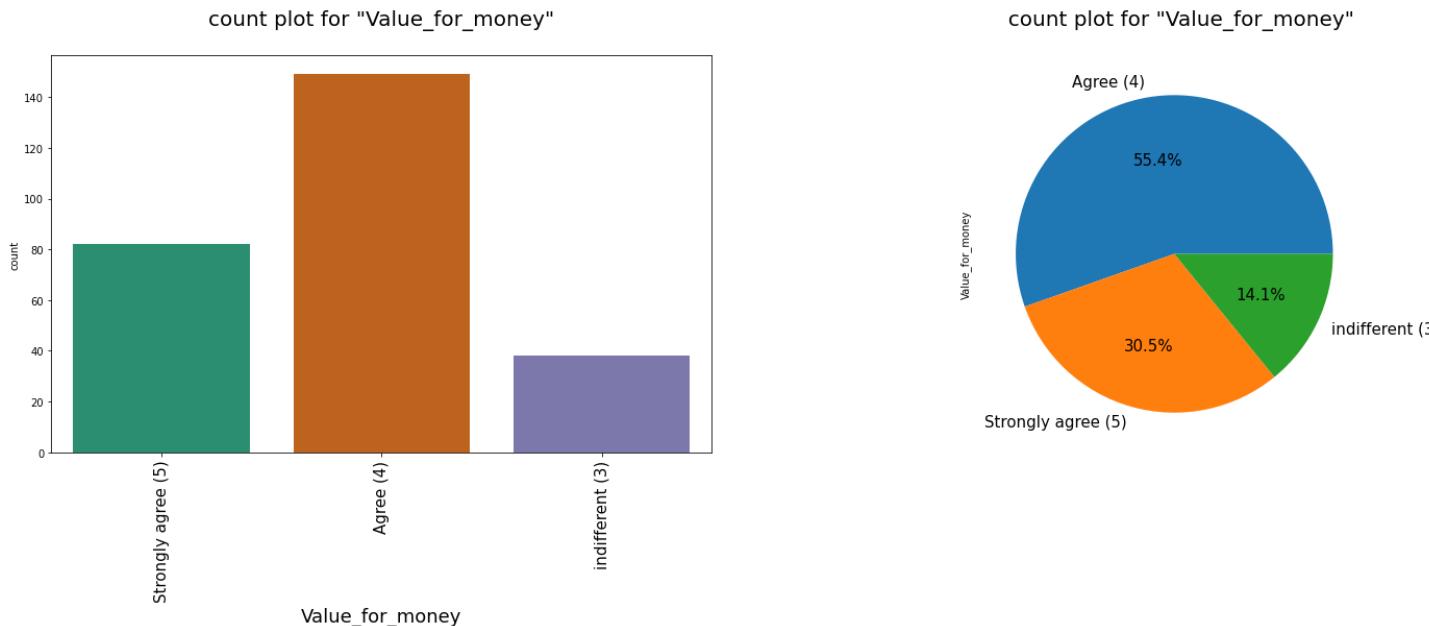


count plot for "Gratification_feeling_favourite_retailer"



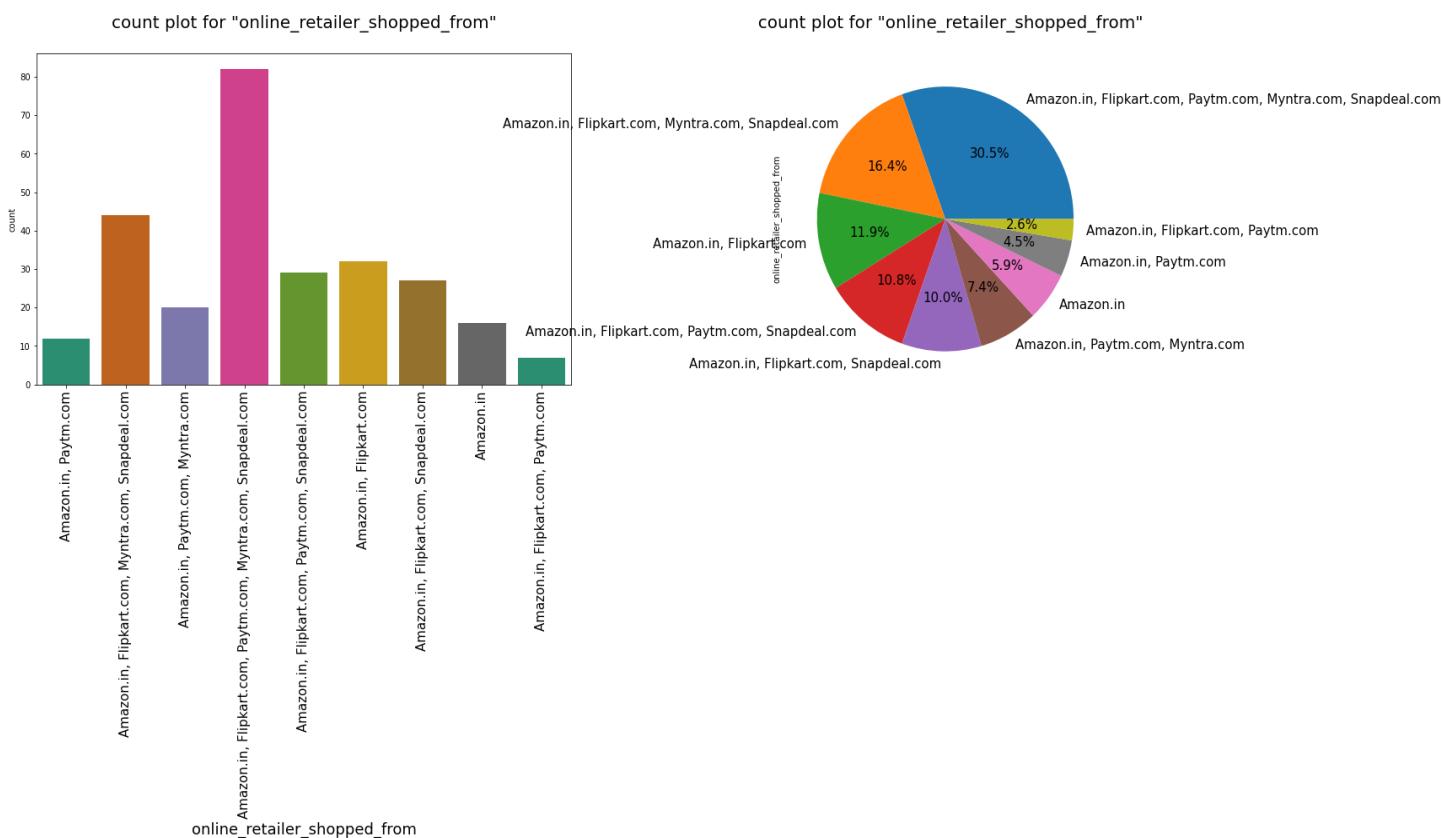
- We can see from the above plots that there is some mixed kind of response on "Gratification feeling while shopping from the favourite retailer", most of the respondent are not concerned about it, some of very strongly concerned about it.

34. Analysis of Value_for_money:



- All of us expect value for the money whenever we shop anything from anywhere.
- Here also the respondents are also stick to the value for money statement.

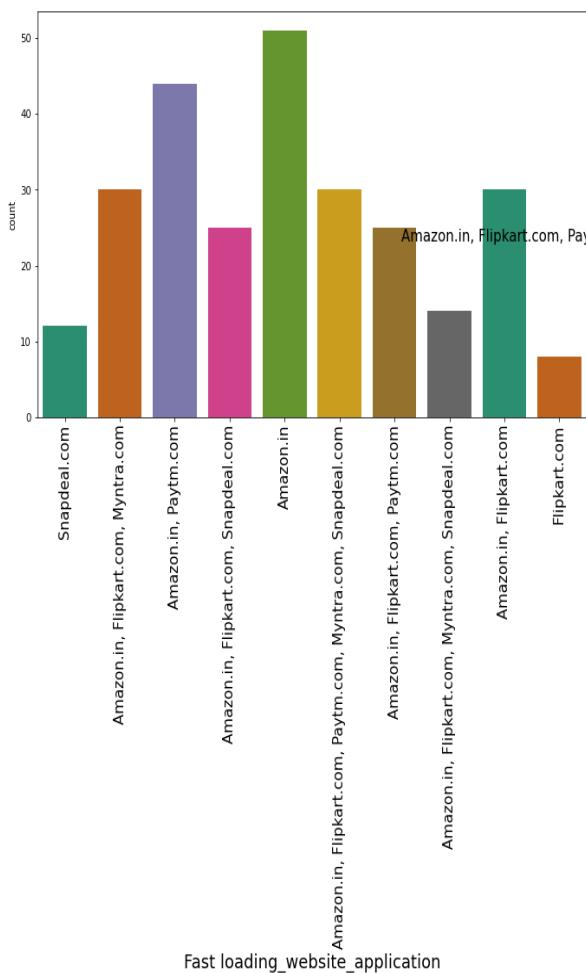
35. Analysis of online_retailer_shopped_from:



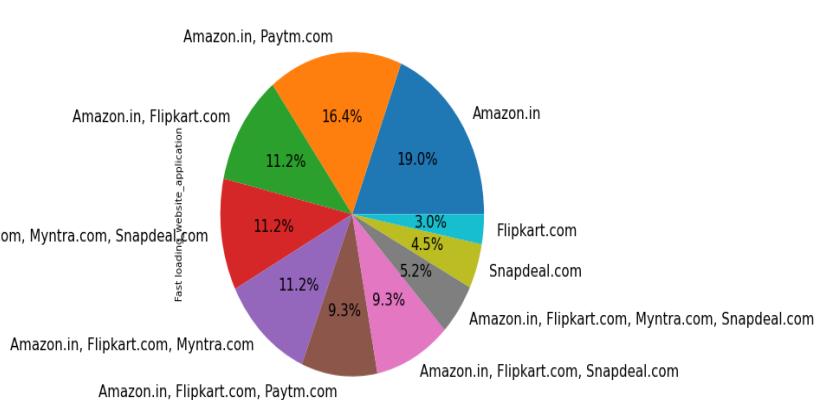
- Most of the Customers have shopped from all ("Amazon", "Flipkart", "Paytm", "Mynta", "Snapdeal").

36. Analysis of Fast loading_website_application:

count plot for "Fast loading_website_application"

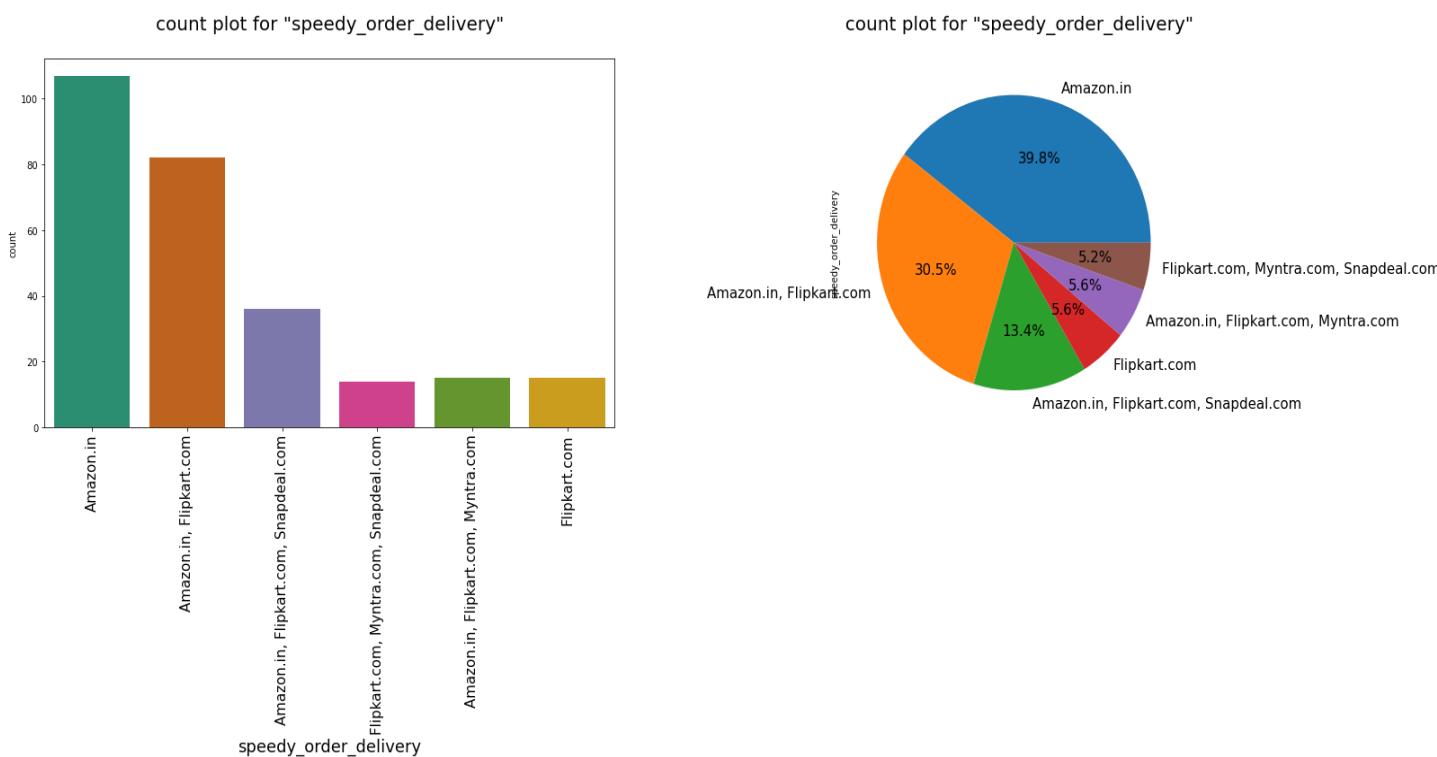


count plot for "Fast loading_website_application"



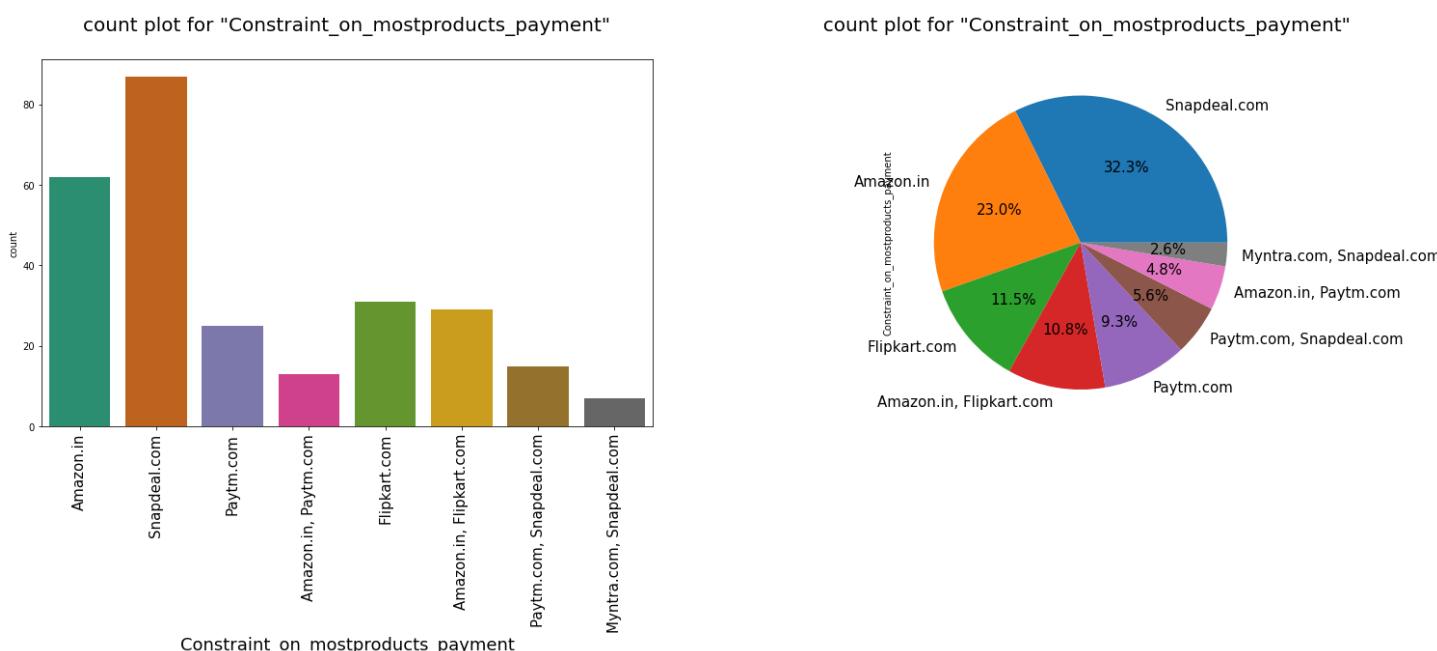
- We can see that most responsive and fast website is "Amazon.com", "Paytm" followed by "Flipkart.com"
- Slowest website is the "Snapdeal.com".

37. Analysis of speedy_order_delivery:



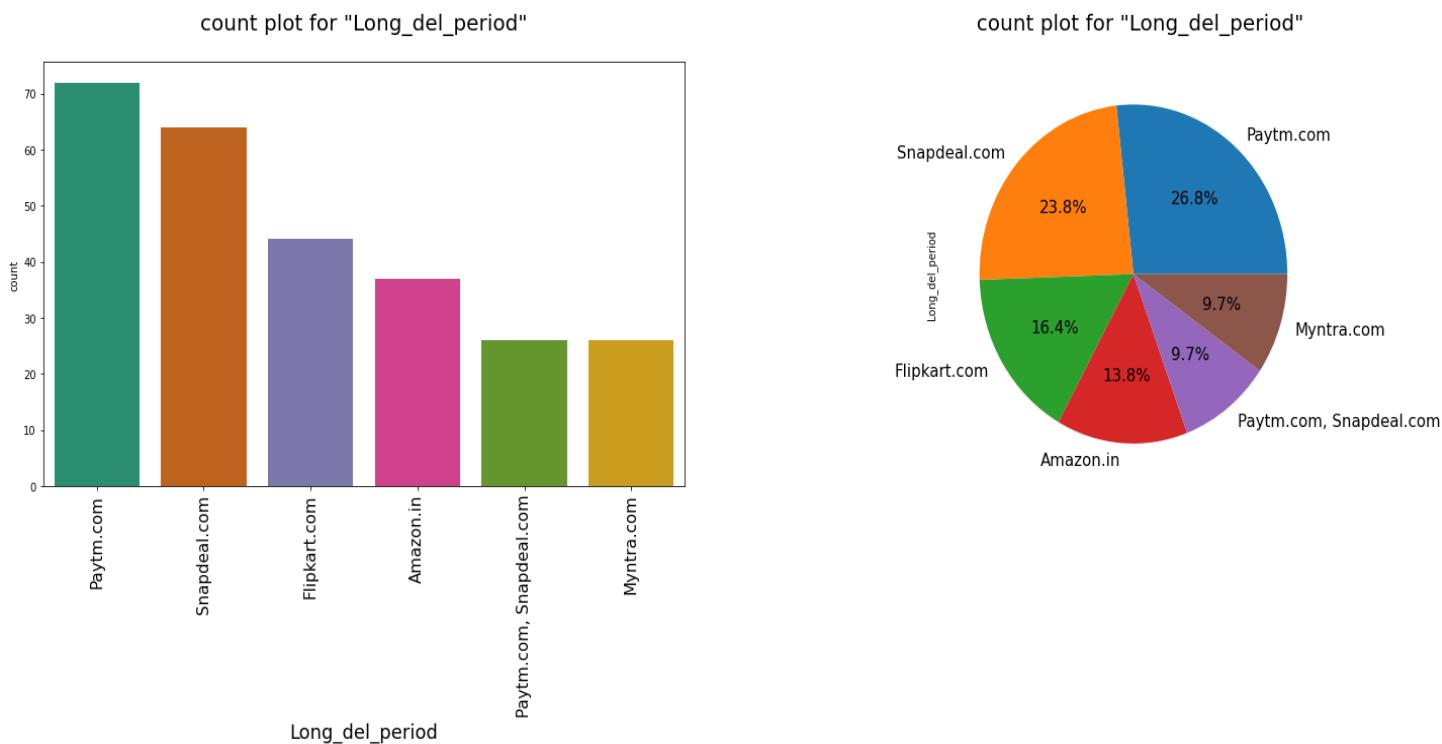
- We can see that "Amazon.com" delivers the product very fast followed by the "Flipkart.com".

38. Analysis of Constraint_on_mostproducts_payment:



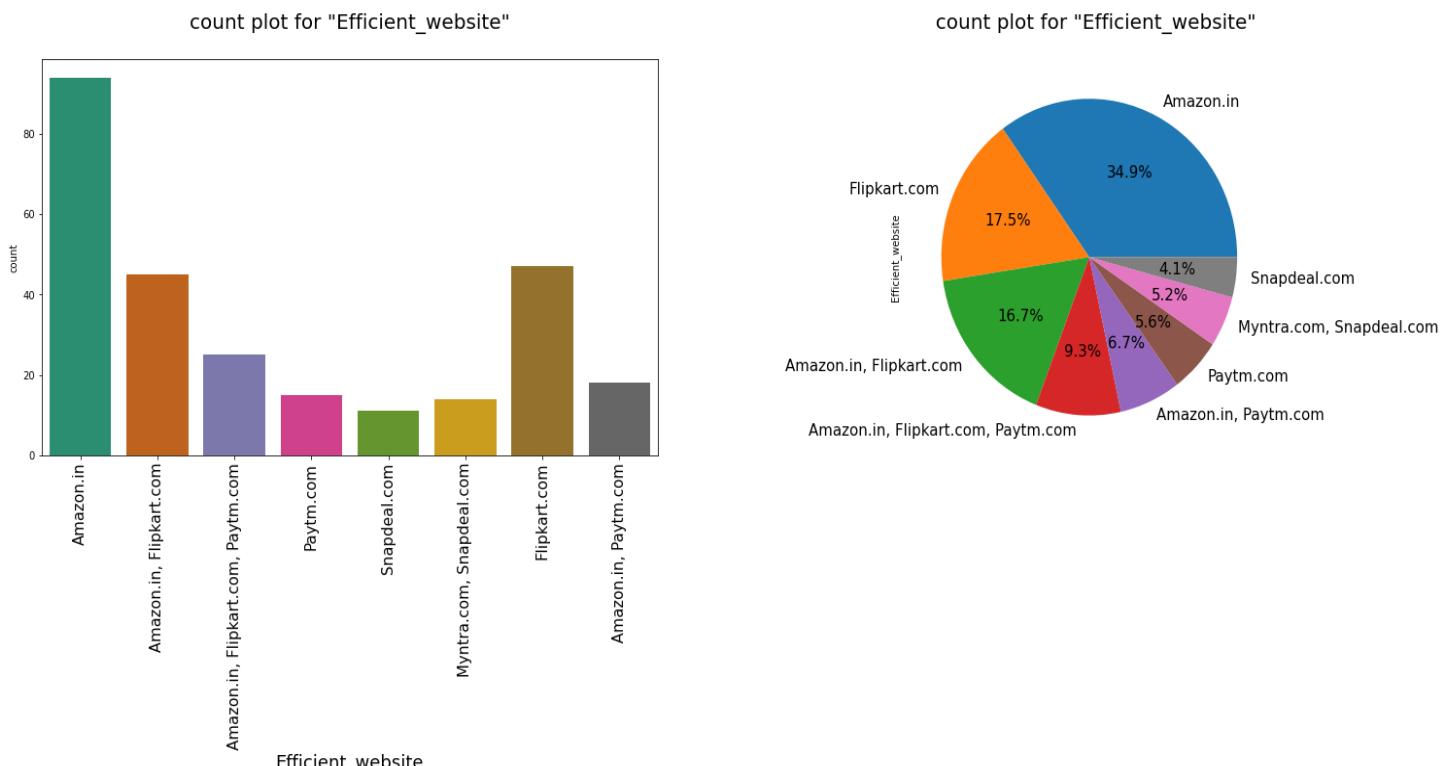
- We can clearly see from the plot that "Snapdeal" has constraints on payment methods for many products followed by "Amazon".
- Mynta and Snapdeal are open to different payment methods

39. Analysis of Long_del_period:



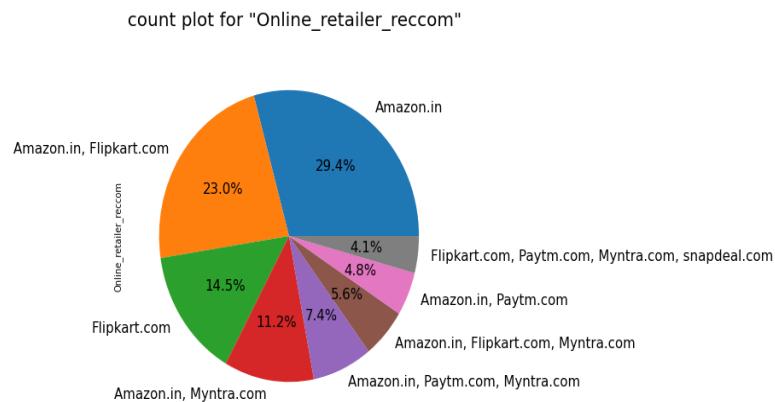
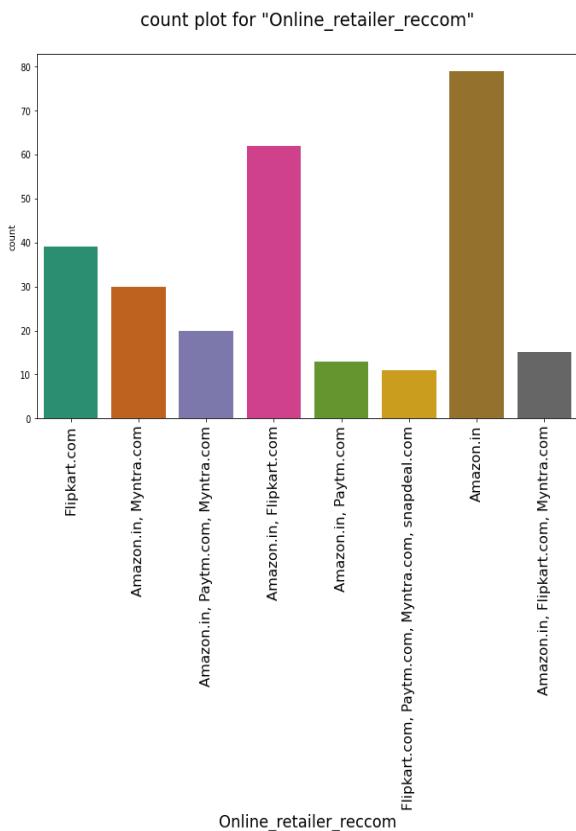
- From the above graph we can clearly say that most quickly delivering website is 'Myntra'
- From the data, most lately delivering website is "Paytm".
- Flipkart and Amazon both deliver the product on average time.

40. Analysis of Efficient_website:



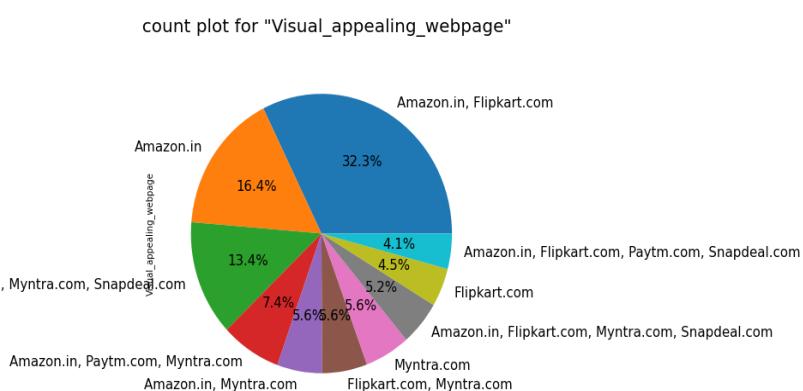
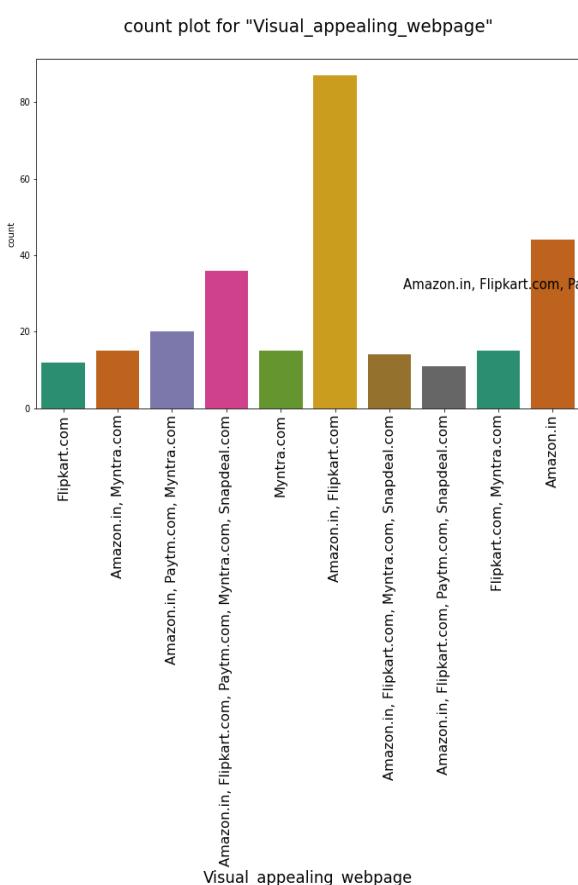
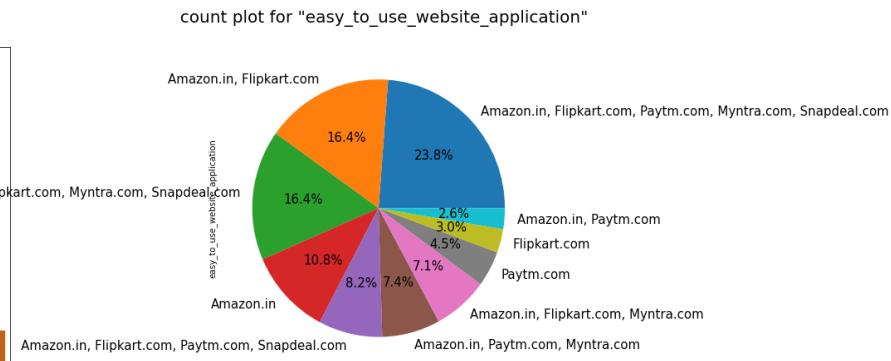
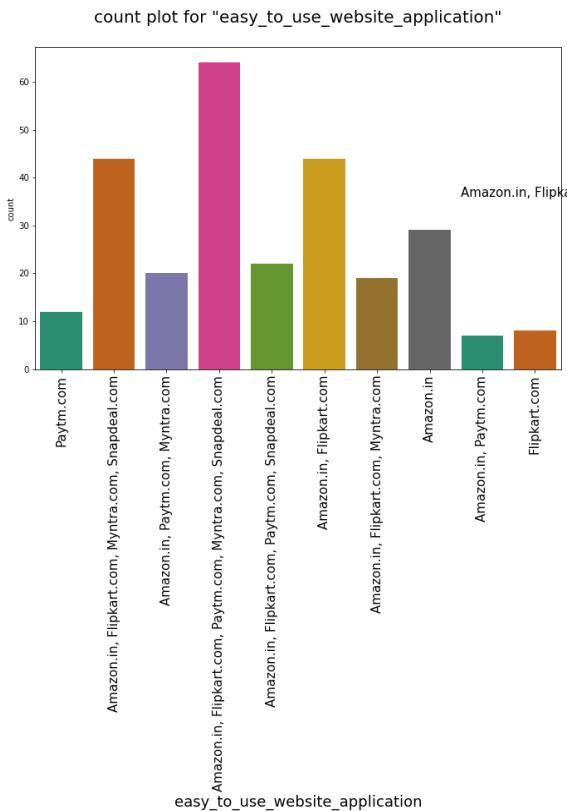
- From the polls we can say that most efficient website is "Amazon" followed by "Flipkart".
- Most inefficient website is "Snapdeal".

41. Analysis of Online_retailer_reccom:

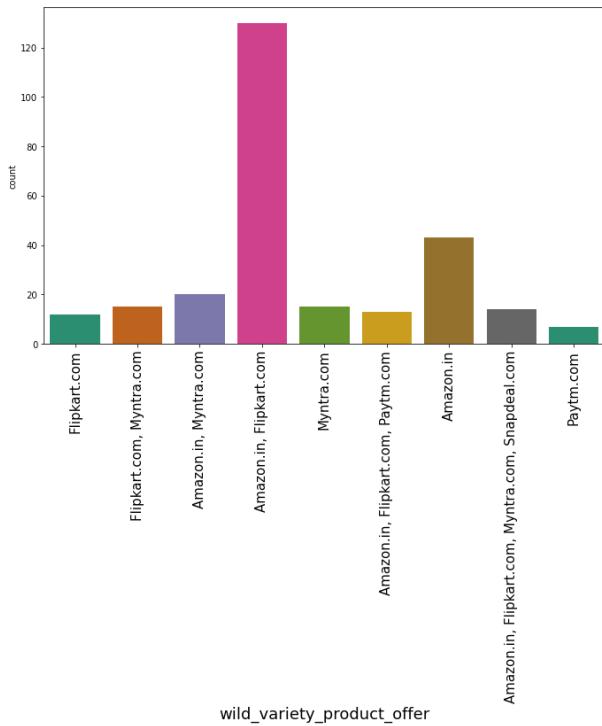


- We can see that mostly people are recommending "Amazon.com" & "Flipkart.com" to others.
- Least recommended website is "Paytm" & "Snapdeal".

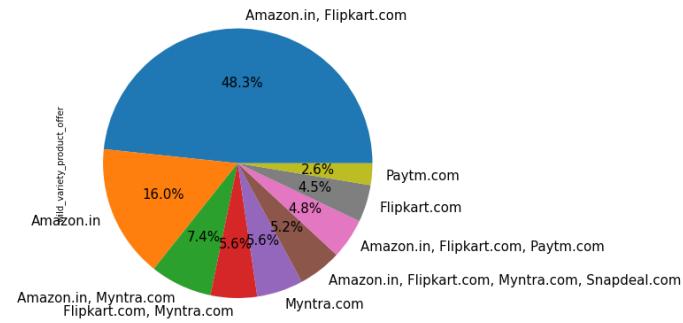
Remaining charts:



count plot for "wild_variety_product_offer"

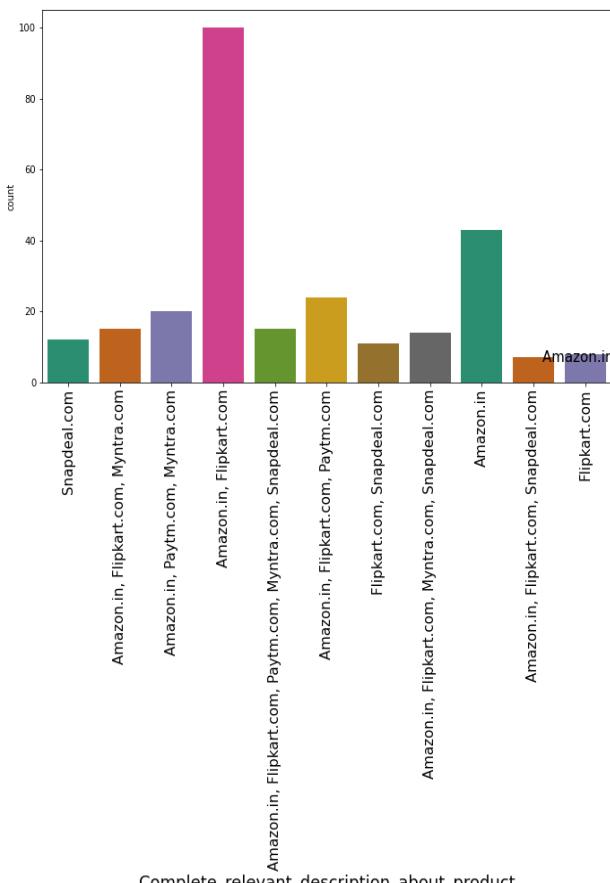


count plot for "wild_variety_product_offer"

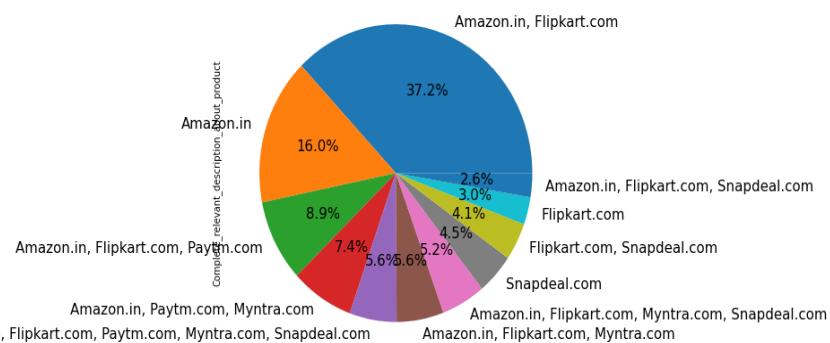


wild_variety_product_offer

count plot for "Complete_relevant_description_about_product"

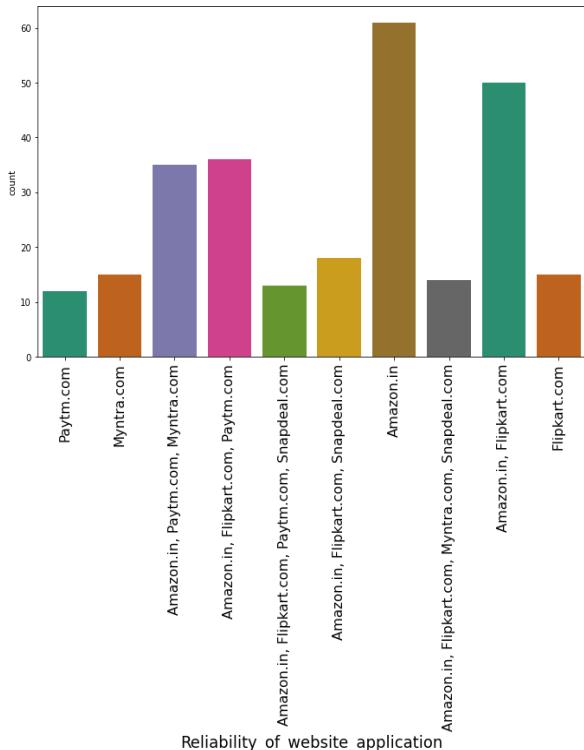


count plot for "Complete_relevant_description_about_product"

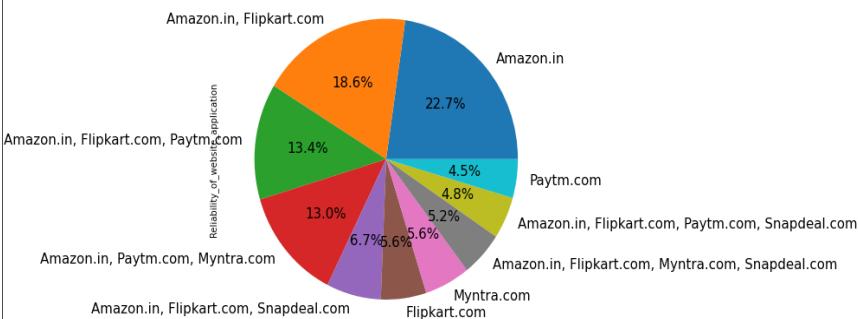


Complete_relevant_description_about_product

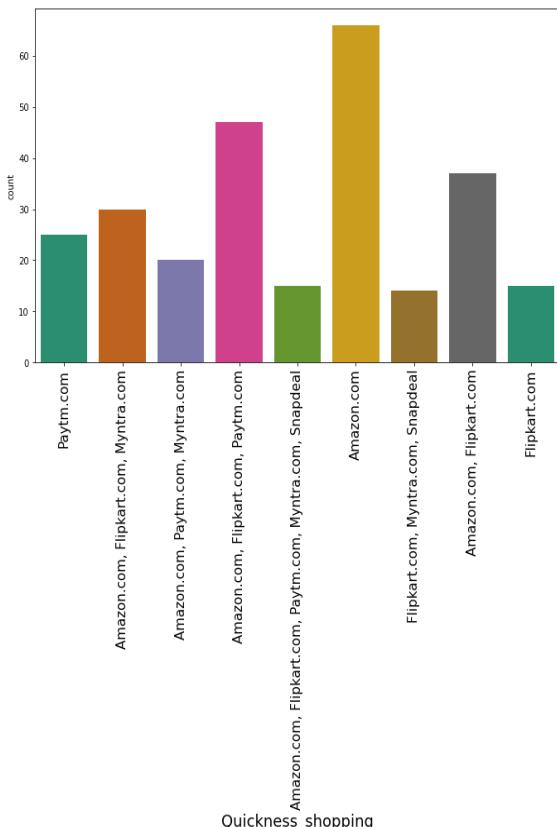
count plot for "Reliability_of_website_application"



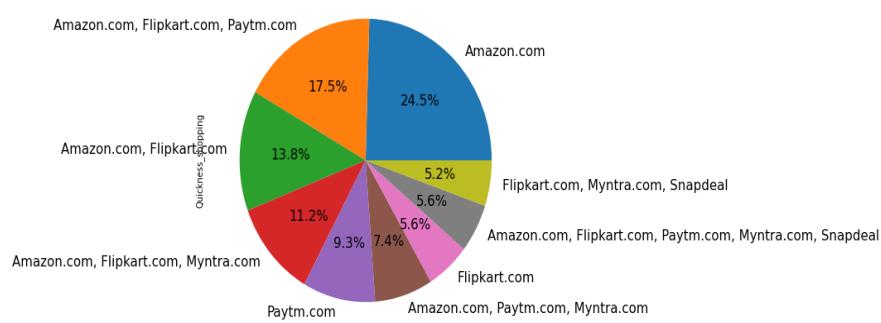
count plot for "Reliability_of_website_application"



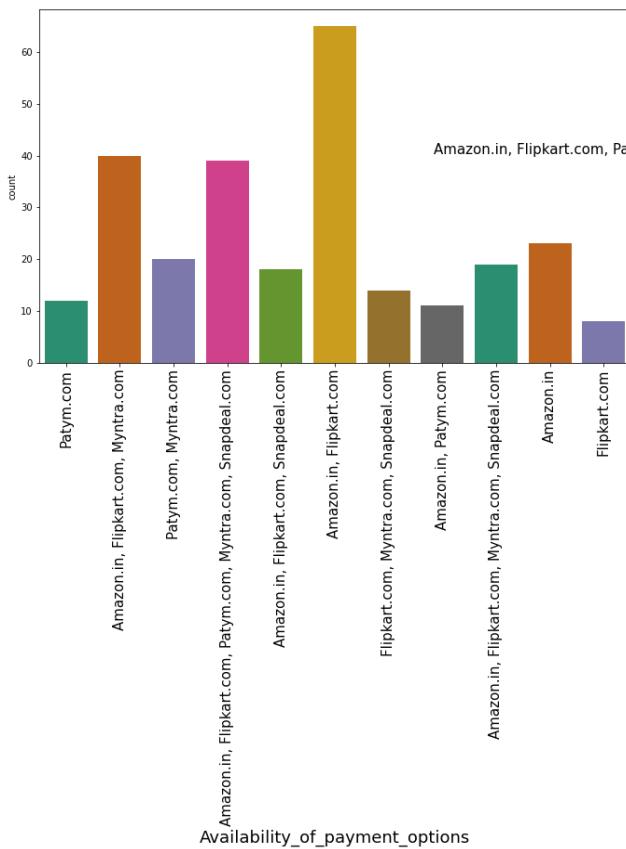
count plot for "Quickness_shopping"



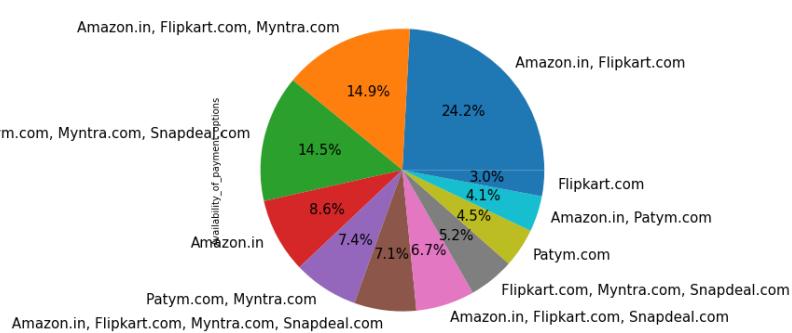
count plot for "Quickness_shopping"



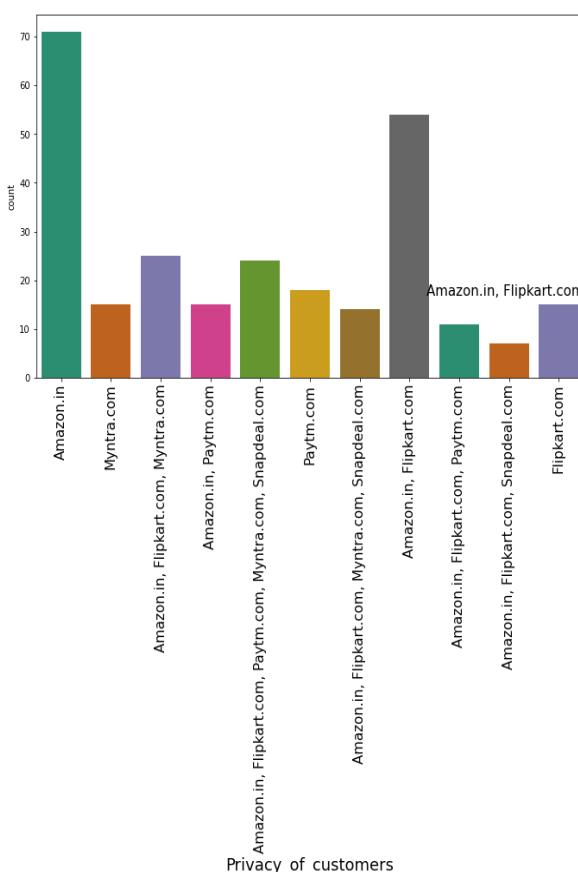
count plot for "Availability_of_payment_options"



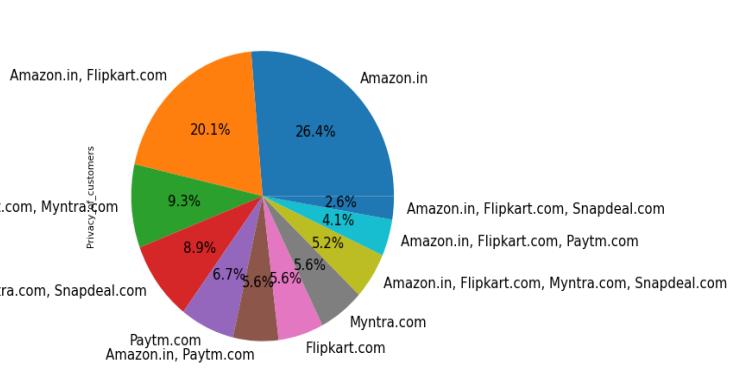
count plot for "Availability_of_payment_options"



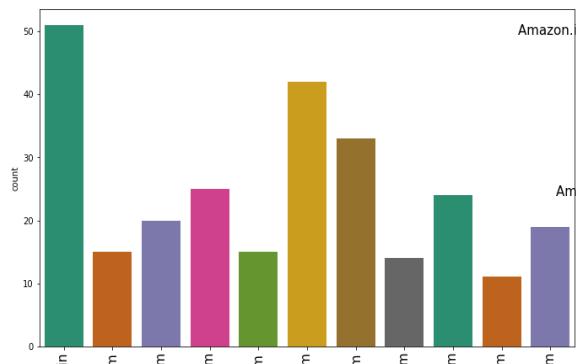
count plot for "Privacy_of_customers"



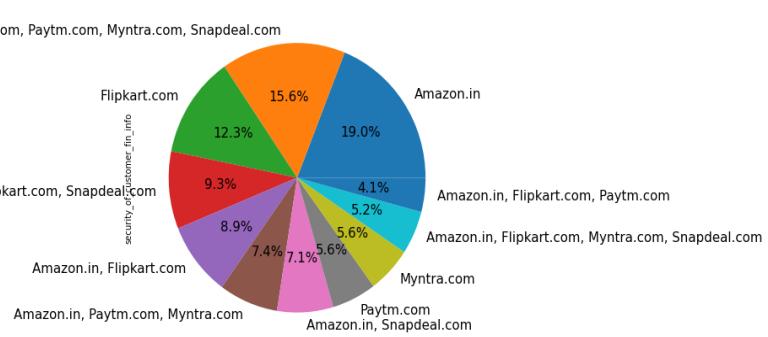
count plot for "Privacy_of_customers"



count plot for "security_of_customer_fin_info"

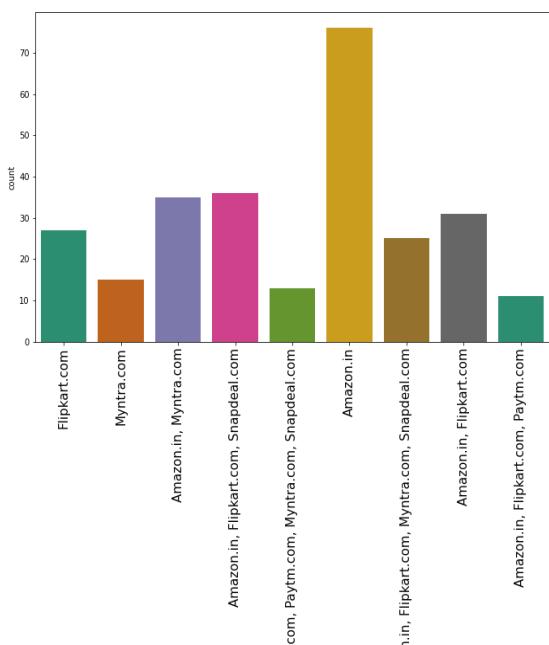


count plot for "security_of_customer_fin_info"

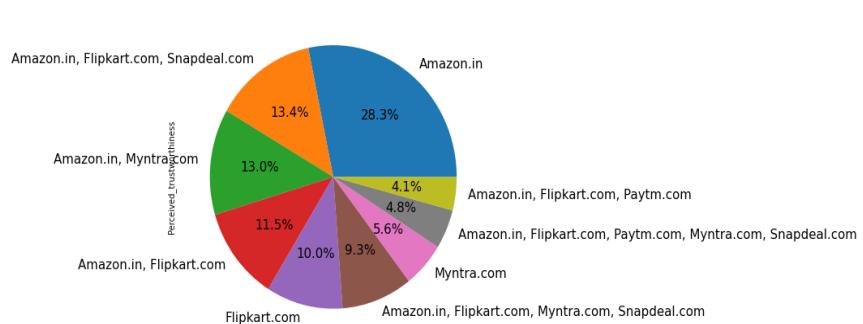


security_of_customer_fin_info

count plot for "Perceived_trustworthiness"

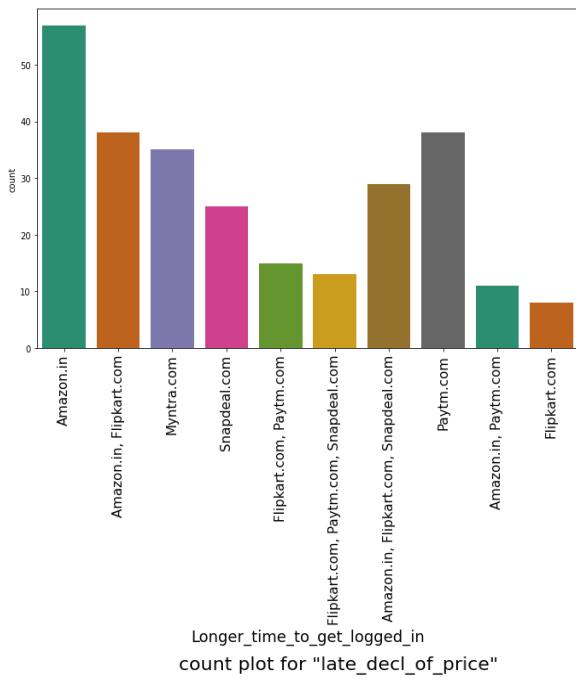


count plot for "Perceived_trustworthiness"

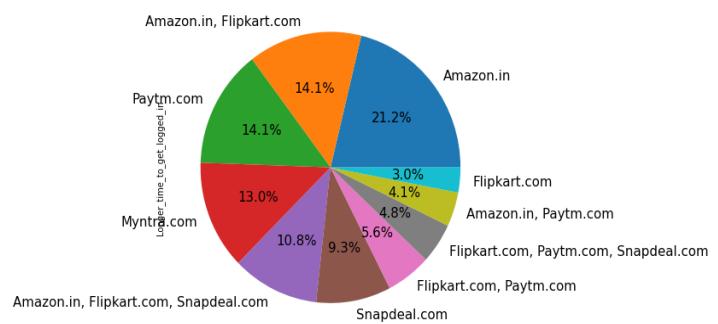


Perceived_trustworthiness

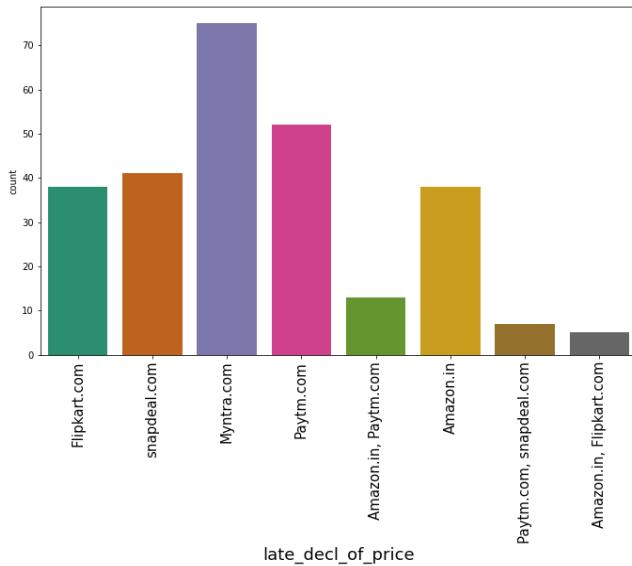
count plot for "Longer_time_to_get_logged_in"



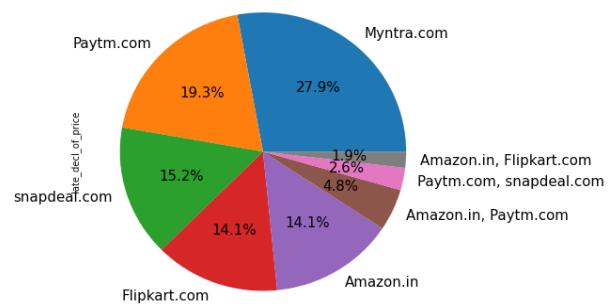
count plot for "Longer_time_to_get_logged_in"



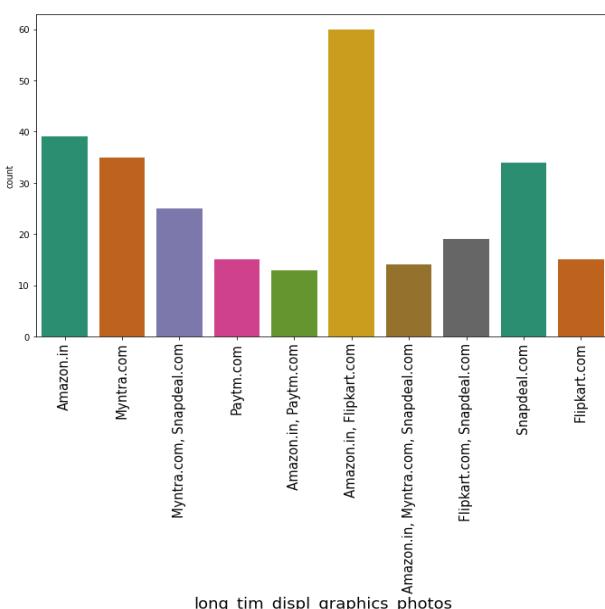
count plot for "late_decl_of_price"



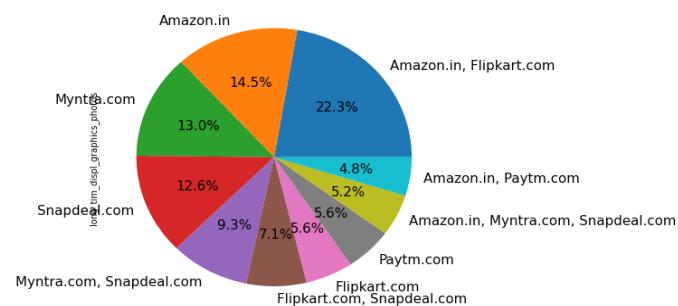
count plot for "late_decl_of_price"



count plot for "long_tim_displ_graphics_photos"



count plot for "long_tim_displ_graphics_photos"



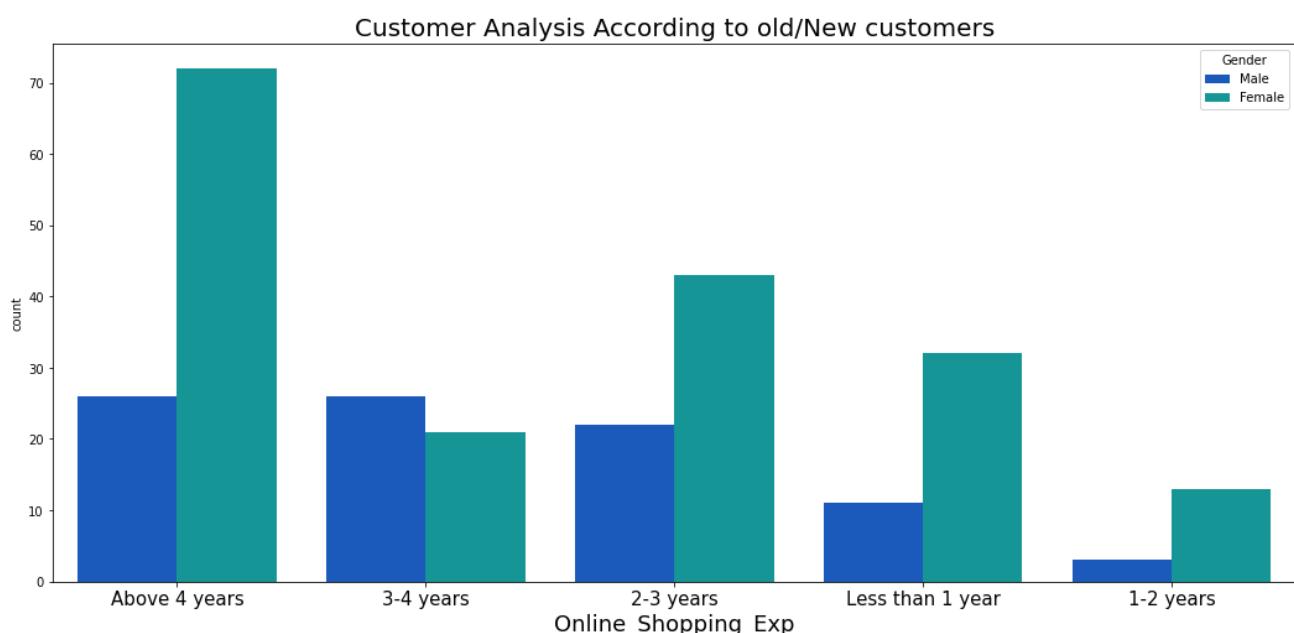
II. Bivariate Analysis:

Bivariate analysis is one of the simplest forms of quantitative (statistical) analysis. It involves the analysis of two variables, for the purpose of determining the empirical relationship between them.

I performed bivariate analysis using count plots again and changing the hue format. Please refer the codes and the outputs in below.

1. Analysis of relation between 'Gender' and 'Online_Shopping_Exp':

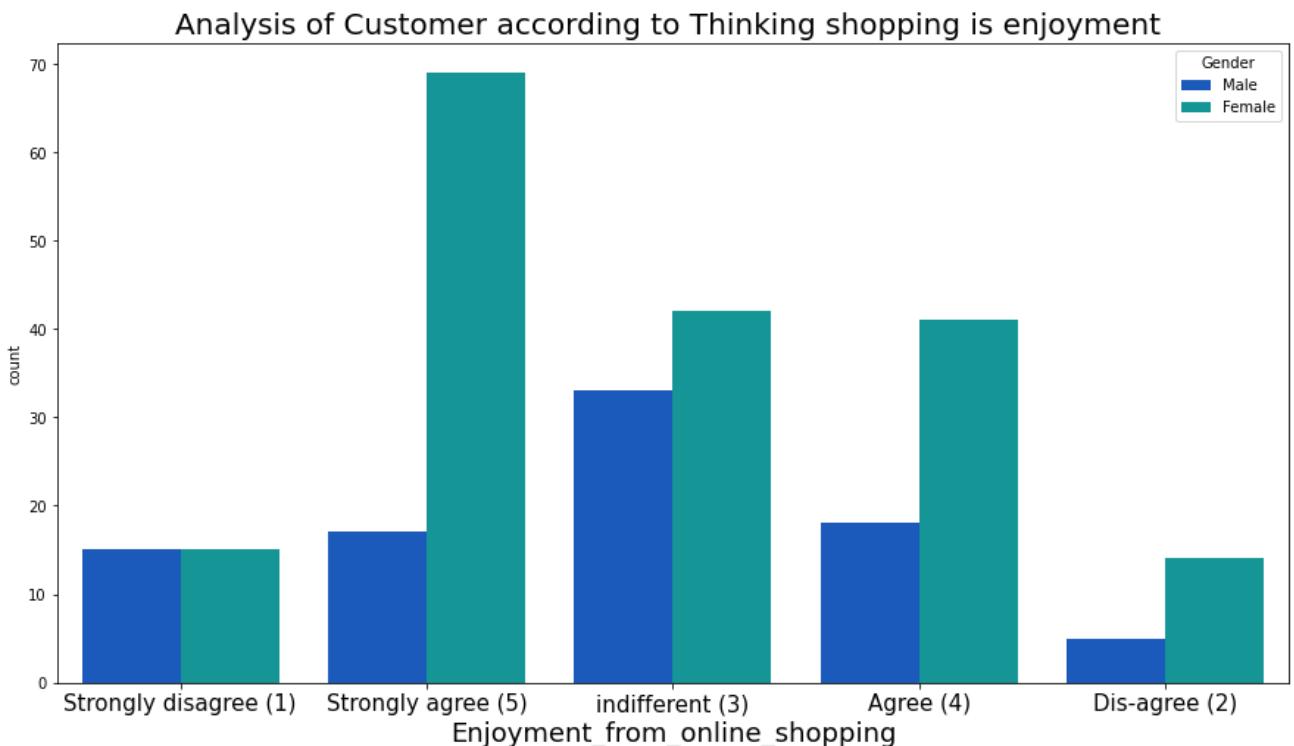
```
1 #relation between 'Gender' and 'OnLine_Shopping_Exp'
2 plt.figure(figsize = (18,8))
3 sns.countplot(x = 'Online_Shopping_Exp',hue = 'Gender' ,data = df, palette = "winter")
4
5 plt.title('Customer Analysis According to old/New customers', fontsize = 20)
6 plt.xlabel('Online_Shopping_Exp',fontsize = 18)
7 plt.xticks(fontsize = 15 )
8 plt.show()
```



- As We have seen earlier, most of the online shoppers of our dataset have more than 4 years' experience (37%) And most of them are Females.
- So, we can conclude that Females have more experience in online shopping than Males.

2. Analysis of relation between 'Gender' and 'Enjoyment_from_online_shopping':

```
1 #relation between 'Gender' and 'Enjoyment_from_online_shopping'
2 plt.figure(figsize = (15,8))
3 sns.countplot(x ='Enjoyment_from_online_shopping',hue = 'Gender',palette="winter" , data = df)
4 plt.title('Analysis of Customer according to Thinking shopping is enjoyment ',fontsize = 20)
5 plt.xlabel('Enjoyment_from_online_shopping',fontsize = 18)
6 plt.xticks(fontsize = 15 )
7
8 plt.show()
```



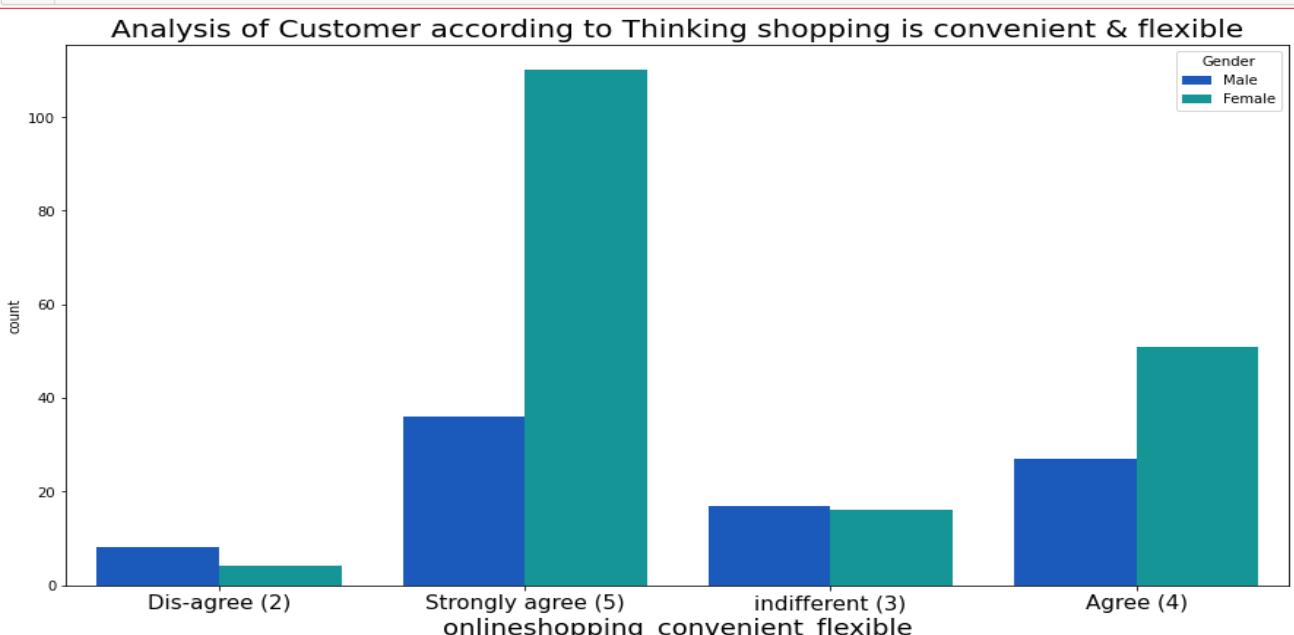
- Above plot shows that females find the online shopping experience enjoyable than the males.

3. Analysis of relation between 'Gender' and 'onlineshopping_convenientFlexible'

```

1 #relation between 'Gender' and 'onlineshopping_convenientFlexible'
2 plt.figure(figsize = (15,8))
3 sns.countplot(x ='onlineshopping_convenientFlexible',hue = 'Gender',palette="winter", data = df)
4 plt.title('Analysis of Customer according to Thinking shopping is convenient & flexible',fontsize = 20)
5 plt.xlabel('onlineshopping_convenientFlexible',fontsize = 18)
6 plt.xticks(fontsize = 15 )
7
8 plt.show()

```



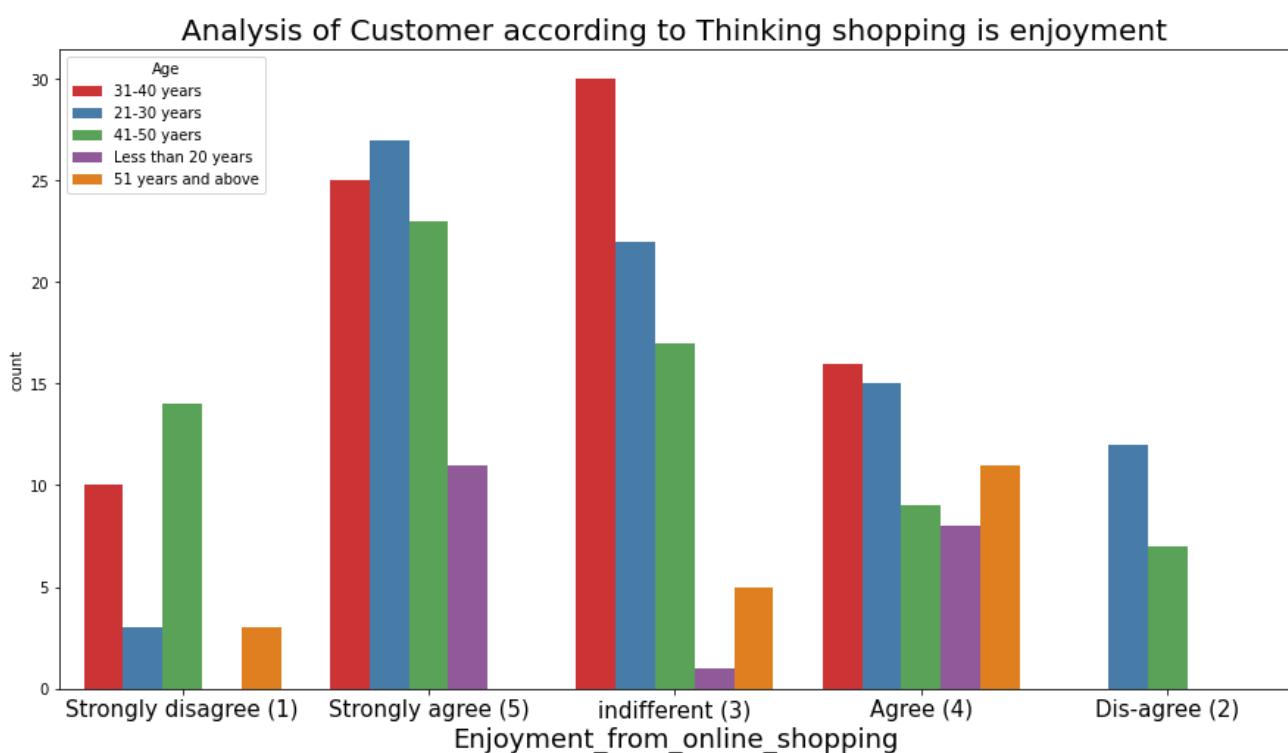
- Above plot shows that most of females find the online shopping convenient and flexible.

4. Analysis of relation between 'Age' and 'onlineshopping_convenient_flexible':

```

1 #relation between 'Age' and 'onlineshopping_convenient_flexible'
2 plt.figure(figsize = (15,8))
3 sns.countplot(x ='Enjoyment_from_online_shopping',hue = 'Age',palette="Set1", data = df)
4 plt.title('Analysis of Customer according to Thinking shopping is enjoyment',fontsize = 20)
5 plt.xlabel('Enjoyment_from_online_shopping',fontsize = 18)
6 plt.xticks(fontsize = 15 )
7
8 plt.show()

```



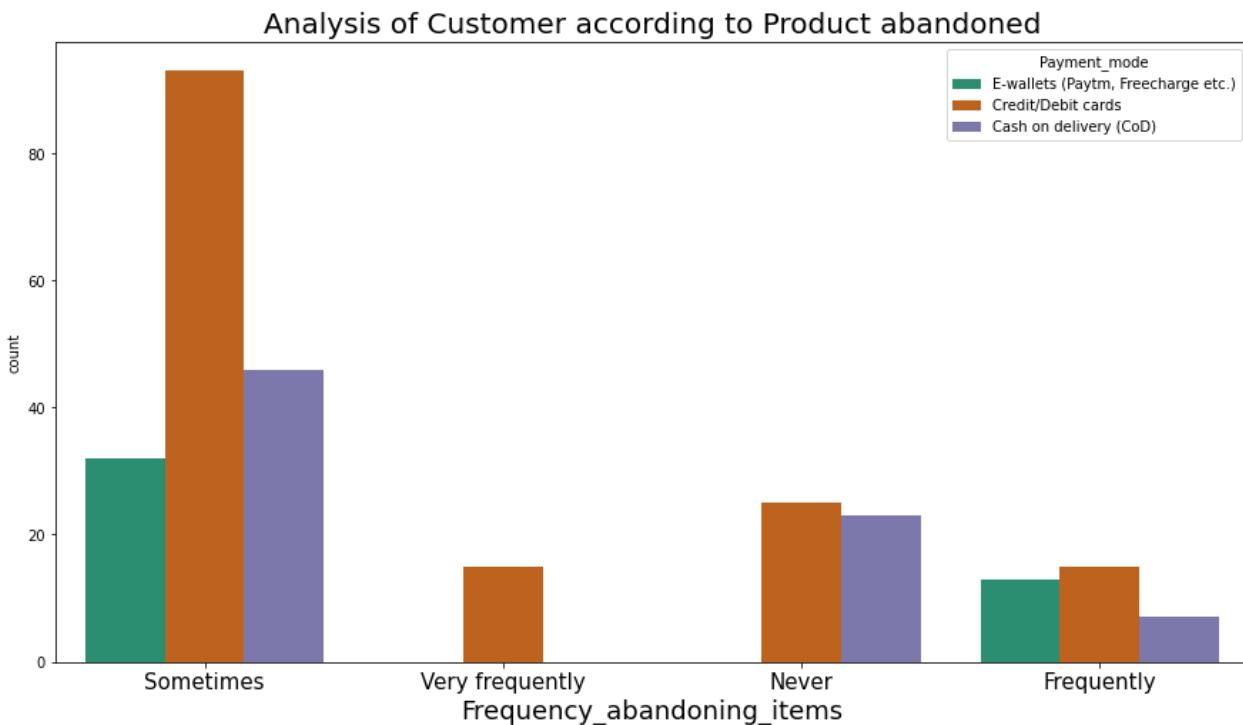
- We can see in above plots customers less than 30 years find the online shopping experience more enjoyable.

5. Analysis of Customer according to Product abandoned:

```

1 # Analysis of Customer according to Product abandoned
2 plt.figure(figsize = (15,8))
3 sns.countplot(x ='Frequency_abandoning_items',hue = 'Payment_mode',palette="Dark2", data = df)
4 plt.title('Analysis of Customer according to Product abandoned',fontsize = 20)
5 plt.xlabel('Frequency_abandoning_items',fontsize = 18)
6 plt.xticks(fontsize = 15 )
7
8 plt.show()

```



- We can see people using COD have abandoned products very rarely while customers using online payment methods are frequent

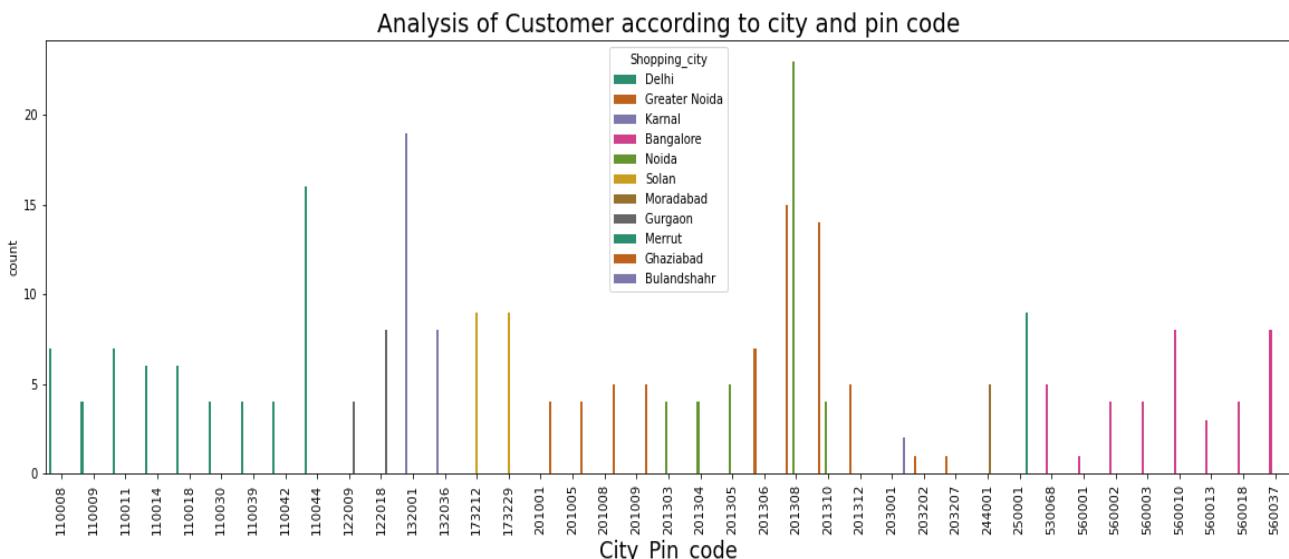
6. Analysis of Customer according to city and pin code:

```
1 df.groupby(['Shopping_city', 'City_Pin_code']).size()
Shopping_city  City_Pin_code
Bangalore      530068           5
                560001           1
                560002           4
                560003           4
                560010           8
                560013           3
                560018           4
                560037           8
Bulandshahr   203001           2
Delhi          110008           7
                110009           4
                110011           7
                110014           6
                110018           6
                110030           4
                110039           4
                110042           4
                110044          16
Ghaziabad     201001           4
                201005           4
                201008           5
                201009           5
Greater Noida 201306           7
                201308          15
                201310          14
                201312           5
                203202           1
                203207           1
Gurgaon        122009           4
                122018           8
Karnal         132001          19
                132036           8
Merrut          250001           9
Moradabad     244001           5
Noida          201303           4
                201304           4
                201305           5
                201308          23
                201310           4
Solan          173212           9
                173229           9
dtype: int64
```

```

1 # Analysis of Customer according to city and pin code
2 plt.figure(figsize = (20,6))
3 sns.countplot(x ='City_Pin_code',hue = 'Shopping_city',palette="Dark2", data = df)
4 plt.title('Analysis of Customer according to city and pin code',fontsize = 20)
5 plt.xlabel('City_Pin_code',fontsize = 18)
6 plt.xticks(rotation=90)
7 plt.show()

```



- We can see in Bangalore there is even distribution of pin codes used by customer except for two codes which are little more used than the others.
- Bulandshahr, Solan, Merrut, Moradabad, being smaller cities has one or two pin codes.
- Delhi also has good distribution of pin codes used by customers except one code (110044) which is used by more than 25 % of customers.
- In Greater Noida also more than 80% of times (201308 & 201310) these codes are used by the customers.
- In Kernal more orders are placed from 132001 pin code.
- In Noida also 70% of customers are using 201308 codes for order delivery.
- IN above analysis it can be predicted that in metro cities there are certain regions where a company can have more delivery focus
- Above plot shows that our data has very large number of Peoples who are doing inline shopping from past 2 to 1 years.

Correlation between columns

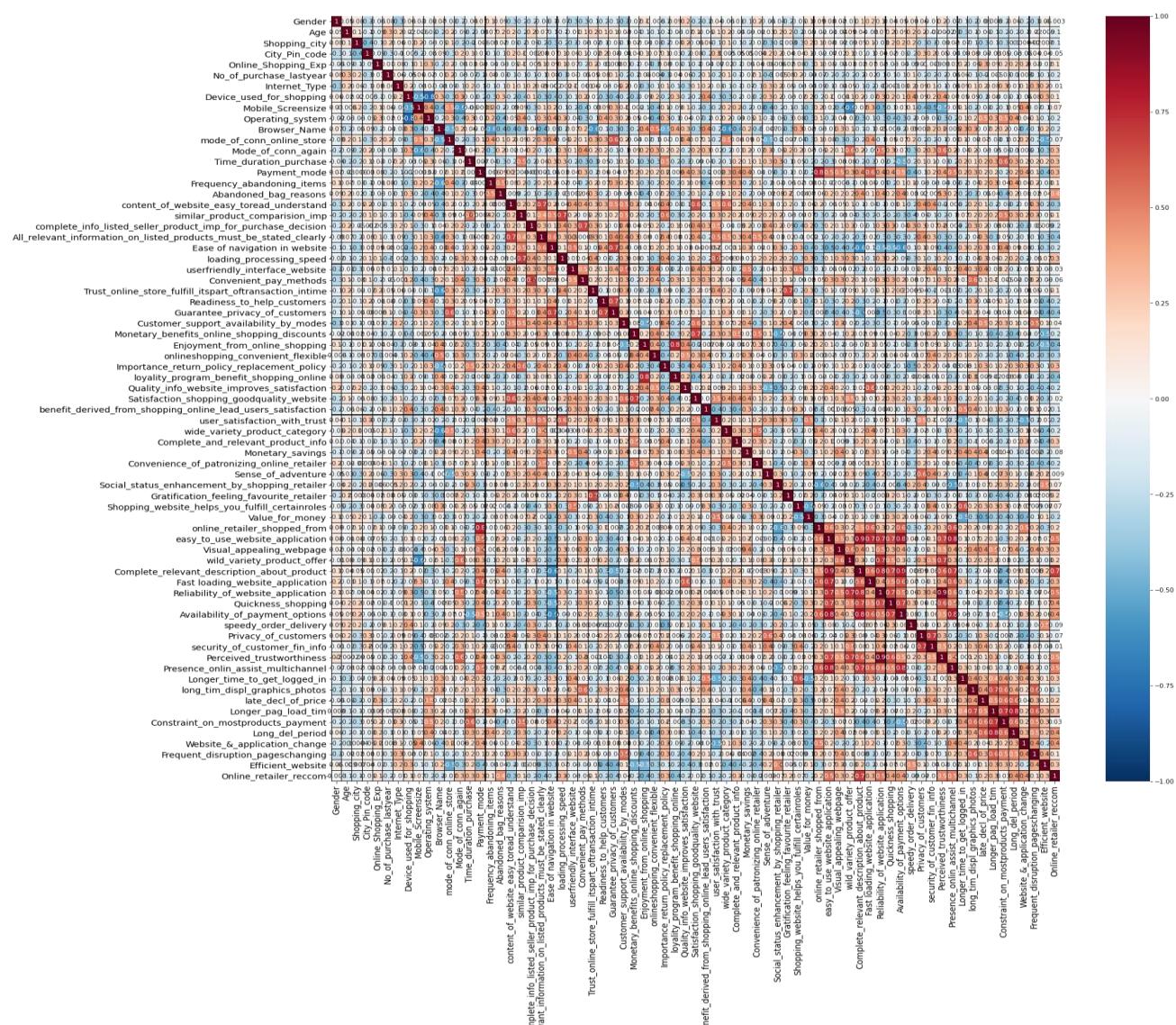
At this point anyone would be sure that there is relation between the columns so, I decided to find out that. And for that I Encode the categorical columns using Label Encoding technique:

```
1 from sklearn.preprocessing import LabelEncoder  
2 LE=LabelEncoder()  
3 df[categorical_col]= df[categorical_col].apply(LE.fit_transform)
```

Then, I generated a heatmap using the correlation values between the dataset columns. The correlation details are bifurcated majorly into positive and negative parts.

- Positive correlation - A correlation of +1 indicates a perfect positive correlation, meaning that both variables move in the same direction together.
 - Negative correlation - A correlation of -1 indicates a perfect negative correlation, meaning that as one variable goes up, the other goes down.

```
1 corr=df.corr()
2 plt.figure(figsize=(25,25))
3 sns.heatmap(df.corr(),linewidhts=.1,vmin=-1, vmax=1, fmt='.1g',
4             annot = True, linecolor="black",annot_kws={'size':10},cmap="RdBu_r")
5 plt.yticks(rotation=0,fontsize = 14)
6 plt.xticks(fontsize = 14)
```

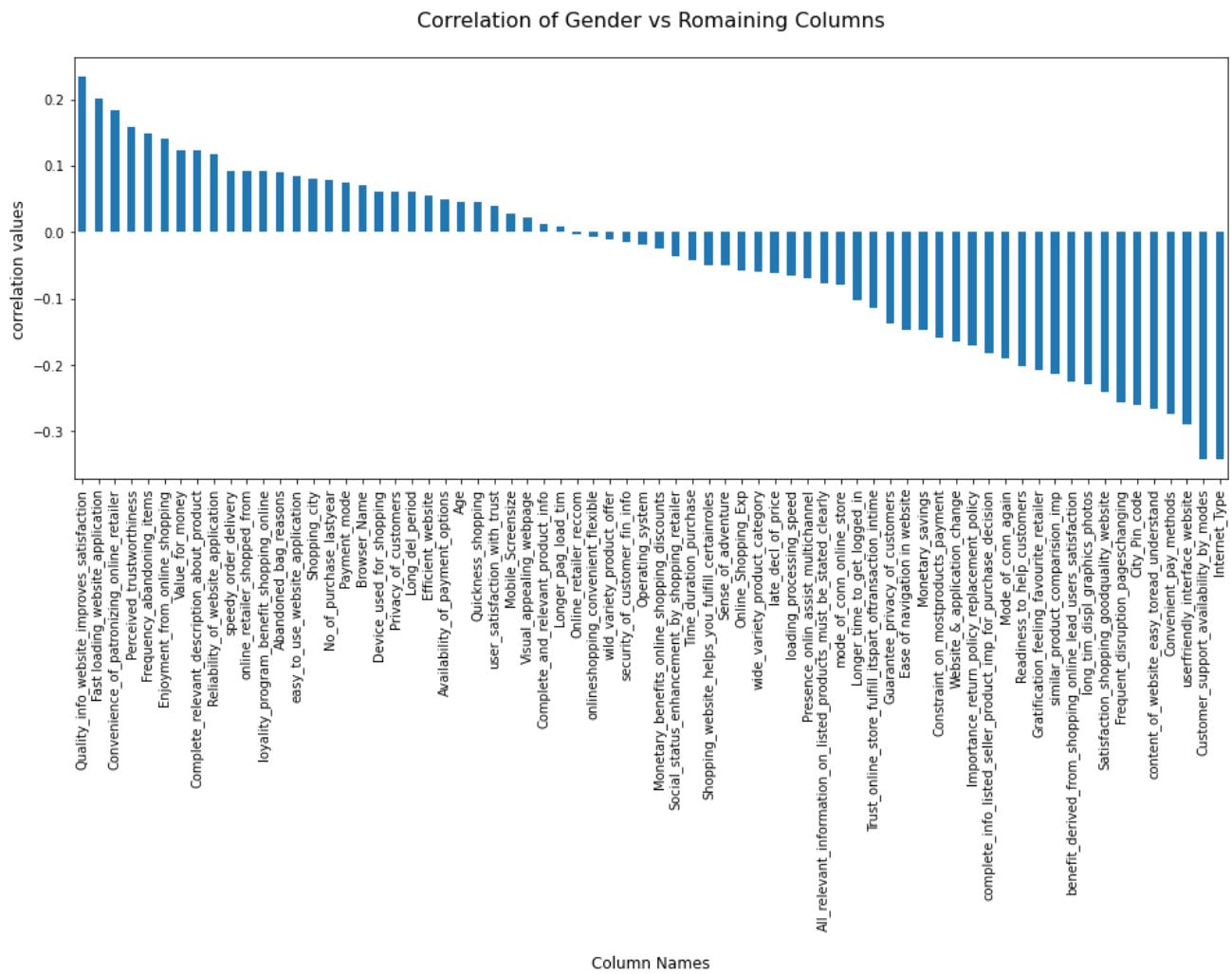


In the above heatmap due to lot of columns we are not able to see the correlation details however we can observe the color-coding details and get a hint that there is no multi collinearity concern between the column values.

```

1 column_names = df.columns
2
3 for col in df[column_names]:
4
5     df_corr= df.corr()
6     plt. figure(figsize=(16,6))
7     df_corr[col].sort_values(ascending=False).drop(col).plot.bar()
8
9     plt.title(f"Correlation of {col} vs Remaining Columns\n". format(col), fontsize=16)
10    plt.xlabel("\nColumn Names", fontsize=12)
11    plt.ylabel("correlation values", fontsize=12)
12
13    plt.show()

```



Since the heatmap was not clear in terms of its values I have generated this bar plot for each column vs remaining column showing the positive and negative correlation data.

Inference of companies:

1. Amazon.com:

Positive feedback summary:

- Convenient to use and also a good website for shopping.
- Fast delivery of products.
- Availability of complete information of the products.
- Presence of online assistance through multi-channels.
- Reliable website or app, perceived trustworthiness.

To be improved:

- During promotions, try to give a disturbance free shopping experience to customers.
- Give more payment options to customers.
- Try to show price early during promotion.
- Reduce the delivery time of the products.

2. Flipkart.com:

Positive feedback summary:

- Convenient to use and also a good website for shopping.
- Fast delivery of products.
- Availability of complete information of the products.
- Presence of online assistance through multi-channels.
- Reliable website or app, perceived trustworthiness.
- Wild variety of products to offer.

To be improved:

- During promotions, try to give a disturbance free shopping experience to customers.
- Give more payment options to customers.
- Try to give the price early during promotion.
- Reduce the delivery time of the products.
- Flipkart and Amazon almost share the same feedbacks with varying percentages as the only difference.

3. Myntra.com

Positive feedback summary:

- Convenient to use and also a good website.
- Availability of several payment options.
- Faster products delivery.
- Complete information of products available.
- Reliable website or app, perceived trustworthiness.
- Wild variety of product to offer.

To be improved:

- During promotions, try to give a disturbance free shopping experience to customers.
- Try to give the price early during promotions.
- Reduce the delivery time of the products during promotions.

4. Paytm.com

Positive feedback summary:

- Convenient to use and a good website.
- Quickness to complete a purchase.
- About 64% of the customers feel that either web or app is reliable.
- Around 20% of the customers believe that Paytm has a wide variety of products on offer.

To be improved:

- Reduce the delivery time of the products during promotions.
- Try to give the price early during promotion.
- During promotions, try to give a disturbance free shopping experience to customers.
- Late declaration of price and discounts.
- Frequent disturbance is occurring while moving from one page to another.

5. Snapdeal.com

Positive feedback summary:

- Convenient to use.
- 54% of the customers are happy about the availability of financial information security.

To be improved:

- Reduce the delivery time of the products during promotions.
- Try to give the price early during promotion.
- During promotions, try to give a disturbance free shopping experience to customers.
- Late declaration of price and discounts.
- No one has expressed to recommend Snapdeal to a contact as it has the most negative feedbacks among all other websites.

❖ Summary

- Study shows that females are more likely to shop online than the males.
- Online shopping is mostly used in big cities like Delhi and Noida, Bangalore more than the smaller cities. Also, there are certain pin codes from this metro cities which were more frequently used for order delivery than rest of the codes.
- It was concluded that 95 % of shoppers are likely to purchase again when they first use the online shopping.
- Most of the customers falls in age group 20 to 40 years. There are less responders in age 51 to 60.
- Peoples generally use mobile or tabs to shop online while some of them also prefer laptop, usually reach to their favourite retail store by searching on a browser and Google Chrome is most used browser followed by Safari.
- Study shows that most of the customers spends more than 10 Min when they visit the online retail store. Also, more than 75% of customer uses online payment methods and 25% of them still go with cash on delivery.
- 50% of customers abandon the products from their carts due to they get and better alternative offer on other retail store.
- Customers prefer that information of similar products should be displayed for comparison as it helps them making a buy decision.
- Users are more likely to shop from a faster loading website. It was found that online shopping is Hedonic for some customers 90% of customers says that shopping online gives them enjoyment, satisfaction and also increases their social status. Nearly 80% of customers believes that online shopping is adventures for them.
- Amazon.in is the people's favourite choice for the online shopping, since it is offering the sales of its products with various additional benefits like "Fastest delivering orders", "Fastest loading and responsive website", "More products are available for any section of item" etc.
- Flipkart is also performing good but it comes after the Amazon.
- Least rated online retailer companies are Snapdeal.com and Paytm.com, very less proportion of the population are following them due to the various reasons like "Very long delivery periods", "Slow loading website", "Less varieties of the products are listed on them" etc.

❖ Conclusions:

- Based on overall observations, the first 47 features provide insights into how e-tailer is helpful & growing based on customer inputs. The data explained how the online platform has been used more often in which CITY, PIN CODE, AGE etc. It also showed that in some factors there is less importance given to contribute to the success of an e-commerce store, so based on that we could remove those factors & keep all the important factors, also we could improve on some factors that influence the online customers repeat purchase intention.
- Apart from the first 47 features, the rest of the features showed which online platform has been used more based on the success factors. Based on the case study for customer activation & retention, Amazon is most reliable and has been fulfilled the customer requirements. After Amazon, data showed Flipkart has been used more for online shopping.
- The case study from Indian e-commerce customers showed Amazon and Flipkart has been used mostly for Online Shopping and most recommended by Friends. So, based on the research factors, Amazon & Flipkart are the e-commerce platform, which are having the combination of both utilitarian and hedonistic values to keep the repeat purchase intention (loyalty) positively.