



CUSTOMER RETENTION PROJECT

Submitted by:

Akash chaudhary

Intenship 17

ACKNOWLEDGMENT

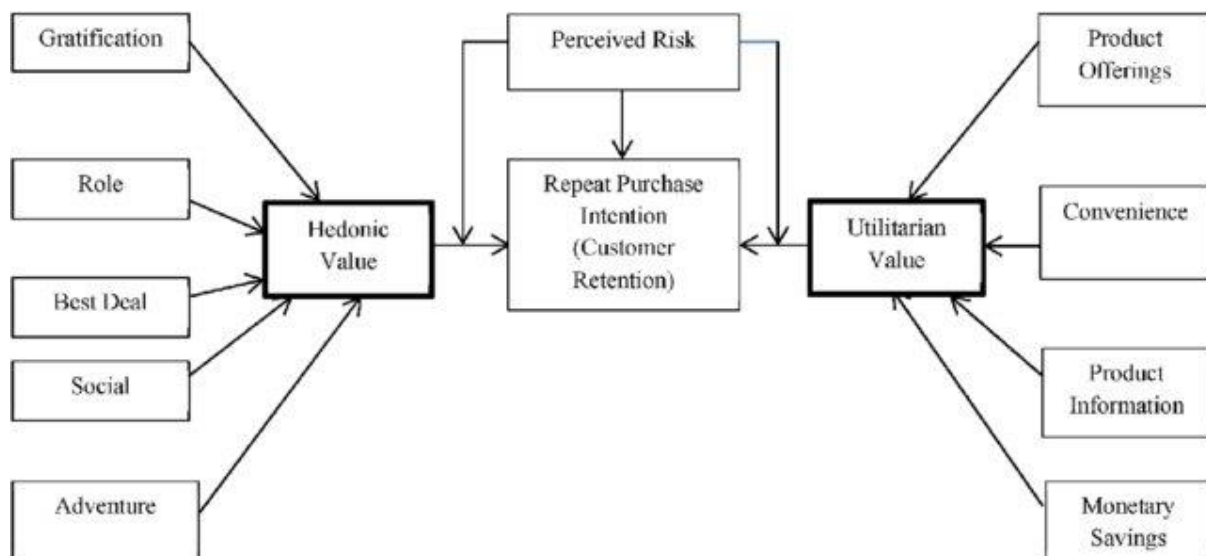
The project customer retention on e-commerce is a problem with different types of consumers who do or do use use online shopping sites in india on the behalf of the different types of clients. The data is collected from the Indian online shoppers. My mentor Mr. sajid chaudhary have provide me these data and help me a lot to address the problems and solutions.

INTRODUCTION

- **Business Problem Framing**

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.



- **Conceptual Background of the Domain Problem**

E-commerce is a crucial part of our lives in this age. The boundaries of online retail are expanding and providing us with complete flexibility while shopping.

The idea that originated out of convenience has now become a symbol of flagship and authentic products. The number of online retail businesses are increasing due to the internet era and huge popularity, this brings in a whole field of competition between these corporations. Data analysis is the key for keeping the top spot.

Companies must be prepared to face the problem of customer retention. While attracting customers is made easy by marketing the monetary benefits, keeping the consumers loyal is one hard task. Retention directly translates to satisfaction in our case. Higher the number of satisfied consumers, more will be the number of retentions. We now need to understand the factors affecting the happiness of the consumers.

This is where data analysis comes in. The data collected by surveys and other methods are analysed methodically keeping in mind the target is to achieve user recommendation. More recommendations give the general attitude of the users which can be used to understand user contentment.

The main pillar of any business is customer satisfaction, this is also true in case of Indian e-commerce retailers. There are multiple factors that concern the survival of an online retailer by affecting customer satisfaction. The aim of this project is to determine these features by drawing out detailed conclusions through analysis. The factors may be divided based on utilitarian and hedonic values. Utilitarian value is drawn out based on necessity while hedonic concerns personal enjoyment. Successfully solving this problem and pointing out the variables affecting the satisfaction of users requires some insight in customer service.

- **Review of Literature**

Exploratory data analysis entails multiple steps – identification of the type of variable, recognizing data types and dealing with their wrong assignment, checking basic nature such as size of the dataset. It involves both graphical and non-graphical analysis which helps in answering important questions concerning the problem. It is also necessary to treat missing values and outliers to gain insights on the data during data pre-processing. Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends and patterns in data. Visualization is one of the best ways to analyse data.

- **Motivation for the Problem Undertaken**

The sample data frame was provided to us from fliprobo for client database. In order to predict the future growth and scope of e-commerce by customer retention for the online shopping or knowing is it safe to for personal finance freedom and safety of the their data to a particular person, the company wants some predictions that could help them in further investment and improvement in selection of customers. So these could be predict by the machine learning algorithms.

Analytical Problem Framing

- **Mathematical/ Analytical Modeling of the Problem**

The mathematical problems are included are this was the huge data so getting null values is not a big deal. So there is some null values we need to tackle with them because of huge data there are many rows i.e. 269 rows and 71 columns first of all we have imported the data using pandas for collecting data we need to understand which information we will need and understand which will work for us to make analysis. First of all we need to check the data is balanced or not. Checking for outliers adjust them and some other works are accured at the time of making the model.

- **Data Sources and their formats**

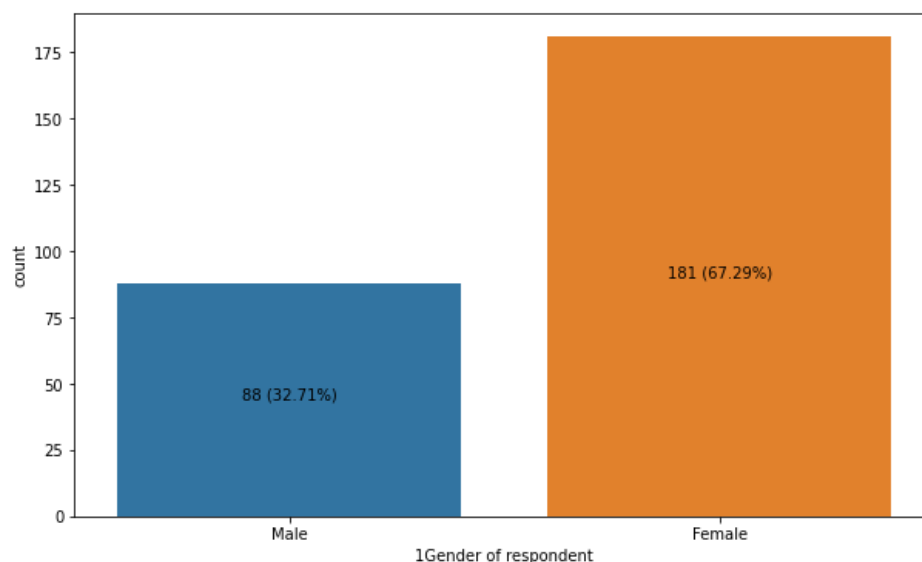
The data is gathered from the customers on their opinion on online shopping. The project was assigned to me by mr.sajid chaudhary data scientist and sme.

The data used for analysis purpose the chances of growing the e-commerce in india and predict is it safe or not. The data contains 269 rows and 71 columns. The data includes both in integer and string forms.

	1 Gender of respondent	2 How old are you?	3 Which city do you shop online from?	4 What is the Pin Code of where you shop online from?	5 Since How Long You are Shopping Online ?	6 How many times you have made an online purchase in the past 1 year?	7 How do you access the internet while shopping on-line?	8 Which device do you use to access the online shopping?	9 What is the screen size of your mobile device?	10 What is the operating system (OS) of your device?	...	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)
0	Male	31-40 years	Delhi	110009	Above 4 years	31-40 times	Dial-up	Desktop	Others	Window/windows Mobile	...	Amazon.in	Amazon.in	Flipkart.com
1	Female	21-30 years	Delhi	110030	Above 4 years	41 times and above	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	...	Amazon.in, Flipkart.com	Myntra.com	snapdeal.com
2	Female	21-30 years	Greater Noida	201308	3-4 years	41 times and above	Mobile Internet	Smartphone	5.5 inches	Android	...	Myntra.com	Myntra.com	Myntra.com
3	Male	21-30 years	Karnal	132001	3-4 years	Less than 10 times	Mobile Internet	Smartphone	5.5 inches	IOS/Mac	...	Snapdeal.com	Myntra.com, Snapdeal.com	Myntra.com
4	Female	21-30 years	Bangalore	530068	2-3 years	11-20 times	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	...	Flipkart.com, Paytm.com	Paytm.com	Paytm.com

- **Data Preprocessing Done**

First of all we got data the data is in excel file so we import this into our model. we need to cleaning up the data although it is a very messed data because we need to doing the proper cleaning so we just added only those columns which we need the most so first we need to clean the data in the data there are 269 rows and 71 columns. Than we correlate the targeted variable with some important independent variables. we have dropped some of unnecessary columns which will not help us to making the model. The data is very inbalanced because there are 67.27% female respondents and 32.63% male respondents. So we need to take steps accordingly.



so we can see there are more number of women respondents as compare to. data is imbalanced because only 32% respondents are men rest of all others are women.

- **Data Inputs- Logic- Output Relationships**

The data is collected from the customers on the behalf of their reviews on online shopping thw questioner includes the questions like how often you used

e-commerce for shopping and how old are you gender and how much you reliable which site is easy to access and which is trustworthy etc. The data points to remember are:

- There are no null values in the dataset.
- There may be some respondents who never used e-commerce yet.
- The dataset is imbalanced. Label '1' has approximately 67.5% records, while, label '0' has approximately 32.5% records.
- For some features, there may be values which might not be realistic. We need to observe them and treat them with a suitable explanation.

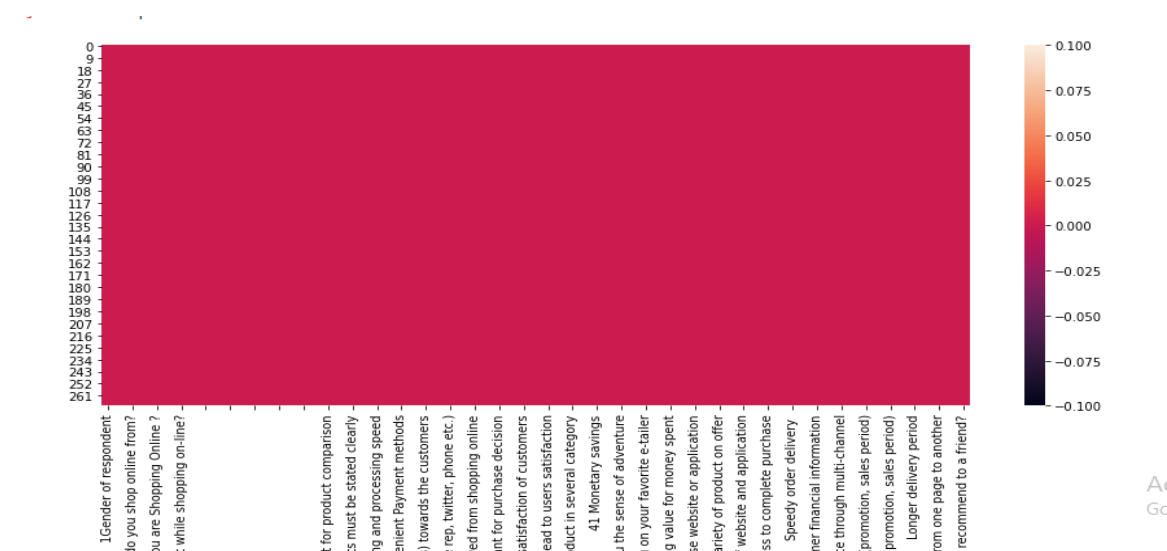
• Hardware and Software Requirements and Tools Used

We have used jupyter notebook to make the model and imported some python libraries these are includes numpy, pandas, maths, stats, seaborn, matplotlib, sklearn etc.

• Visualizations and Interpretation of the Results

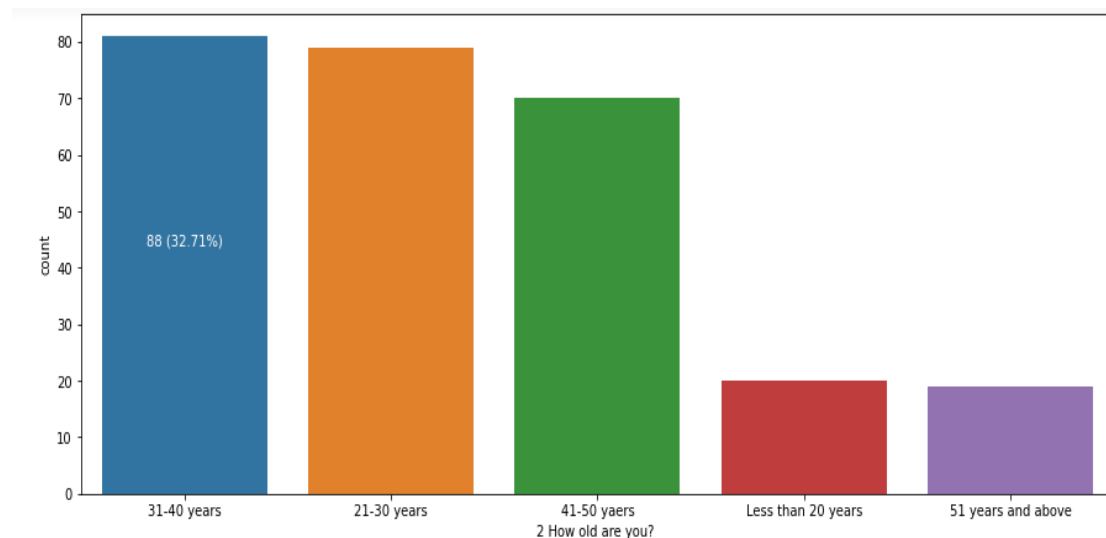
The visualizations are includes:

1. Heat map for null values



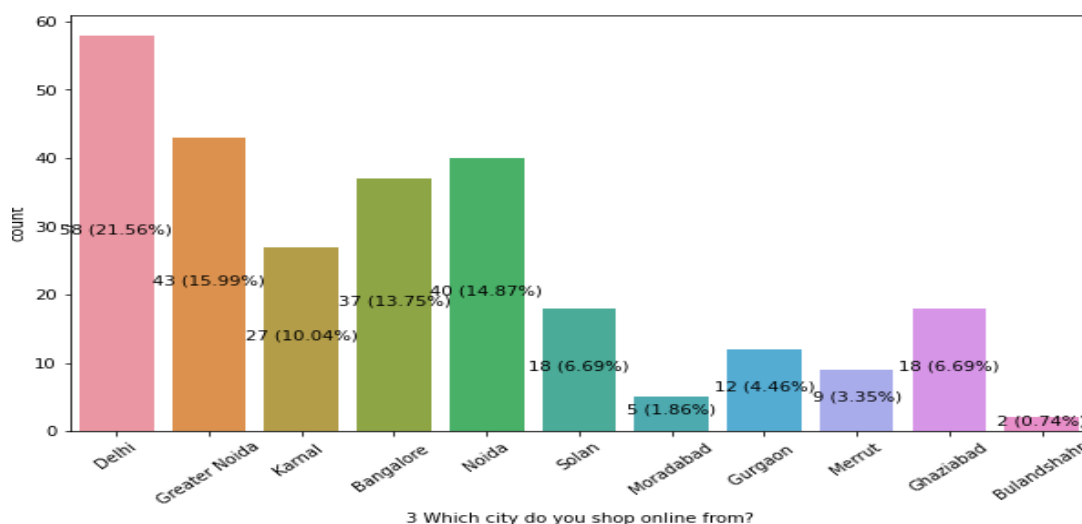
the dataset is free of null values.

2. checking the data distribution using count plot for age groups



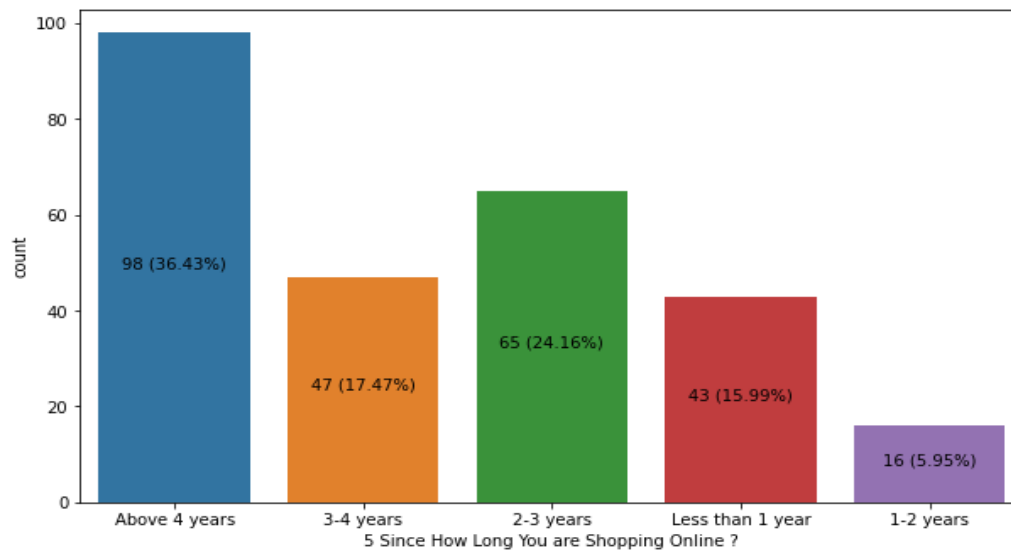
we can see macimum number of people are from 31-40 age group. less than 20 years group people and 51 years age group people are not use online shopping sites.

3. checking the data distribution using count plot for city.



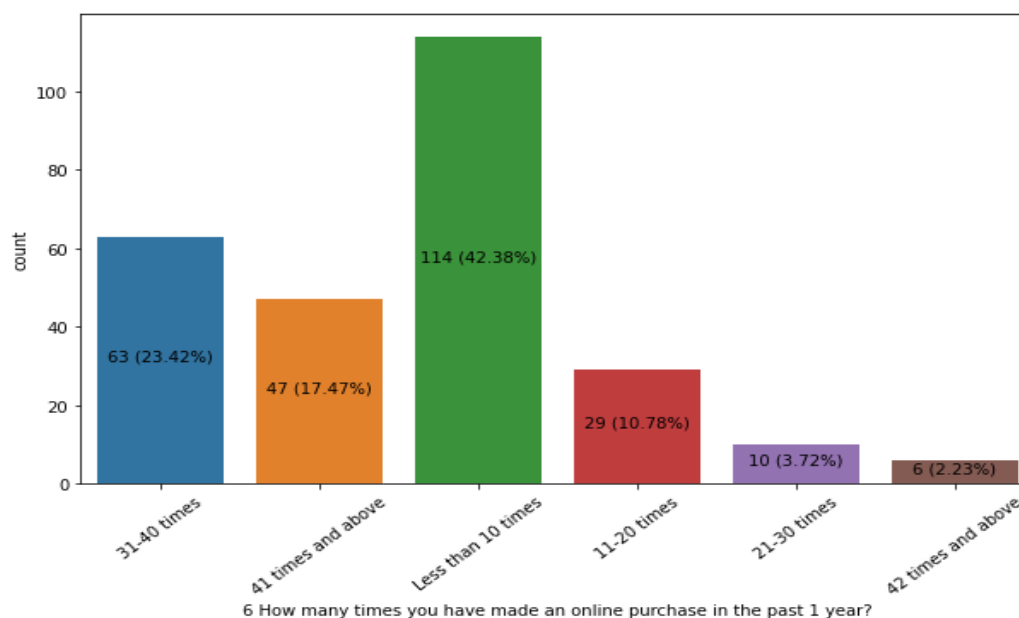
delhi has maximum number of people using e-commerce sites.than nida banglore and greater noida. only fer people are using online sites in bulandshehar and moradabad.

4. checking the data distribution using count plot for time to use ecommerce.



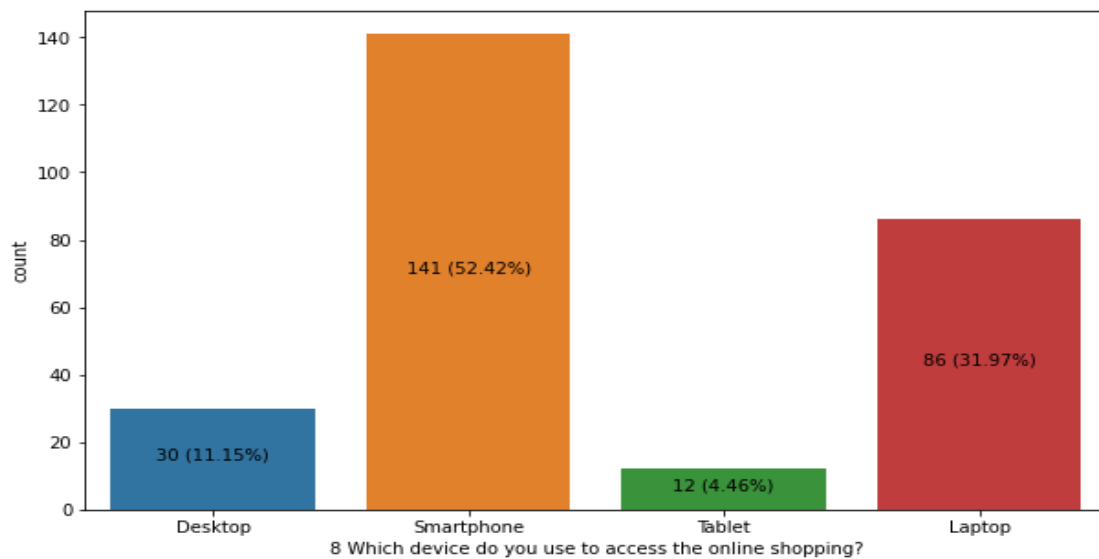
maximum people are using the ecommerce sites more than 4 years and numbers are increasing for the people using shopping sites.

5. checking the data distribution using count plot for last time orderd.



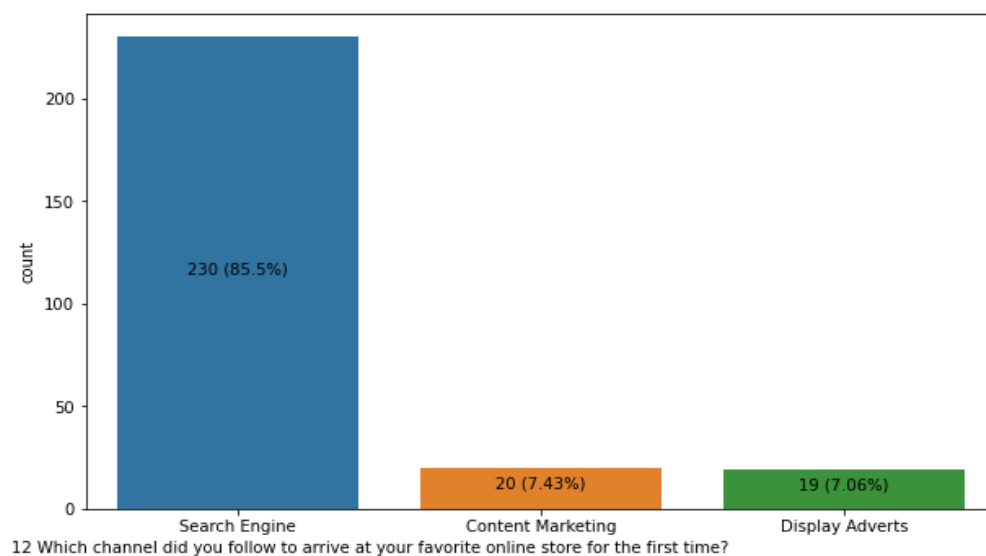
most of the people are place order less than 10 times and very fer people order more than 42 times in a year.

6. checking the data distribution using count plot for mode of shopping.



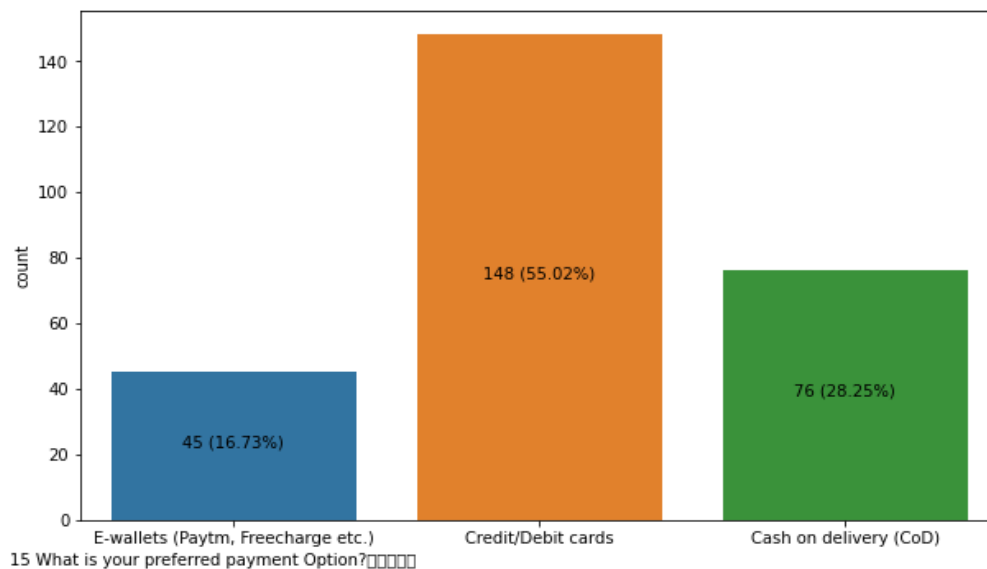
most people are using smart phones and laptop for online shopping and few people are using tablets for online shopping.

7. checking the data distribution using count plot for channel for online shopping.



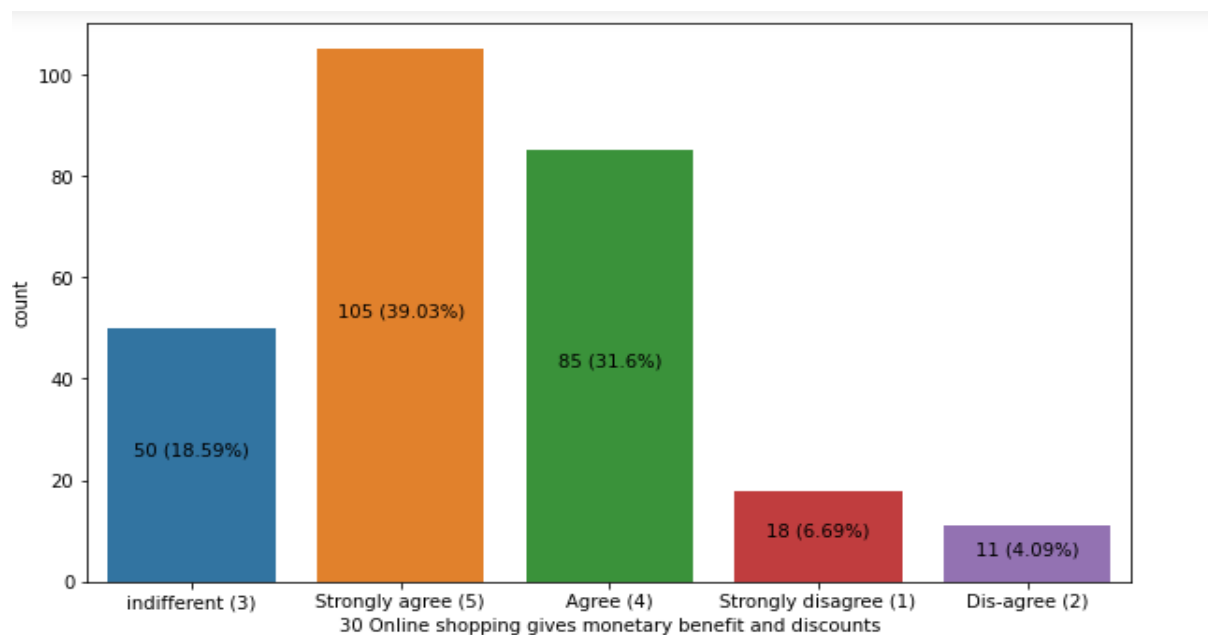
we can see most of the people are using search engine for online shopping. and very few people are using content marketing and display advt.

8. checking the data distribution using count plot for preferred payment method.



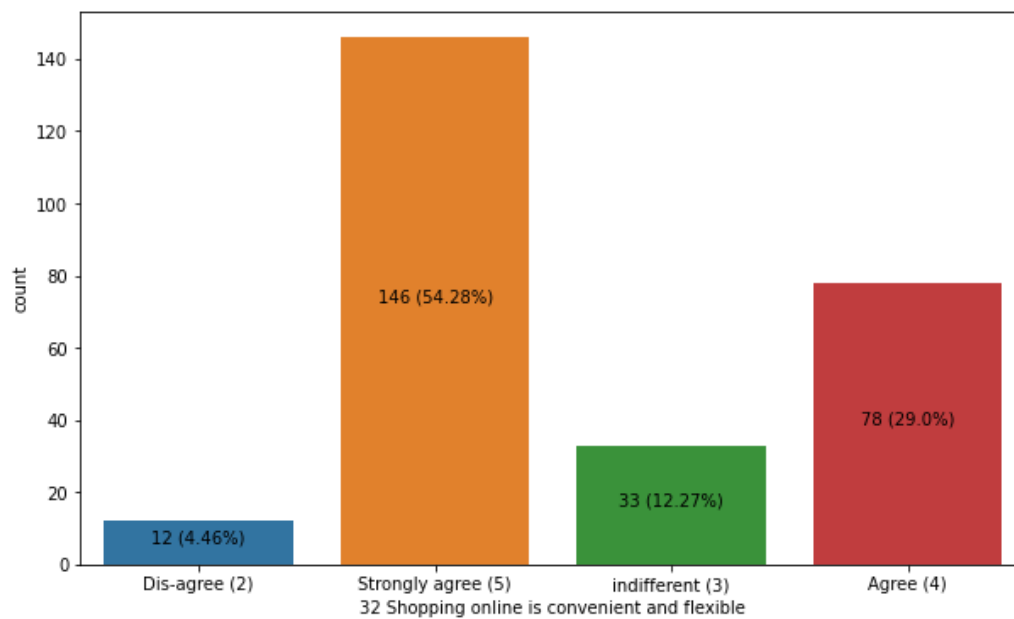
most of the customers are using credit/debit card for payments. and few people are using e wallets.

9. checking the data distribution using count plot for does online shopping gives benefit.



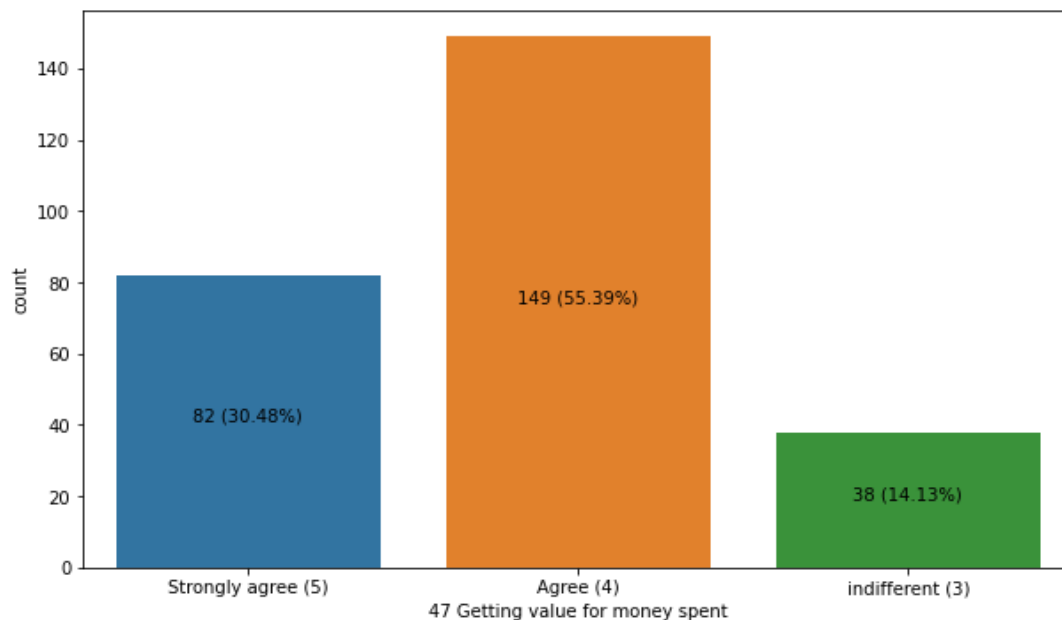
most of the people believe that online shopping can gives tham moonetary benefits and online fer people suppose its not.

10. checking the data distribution using count plot for convenience or not.



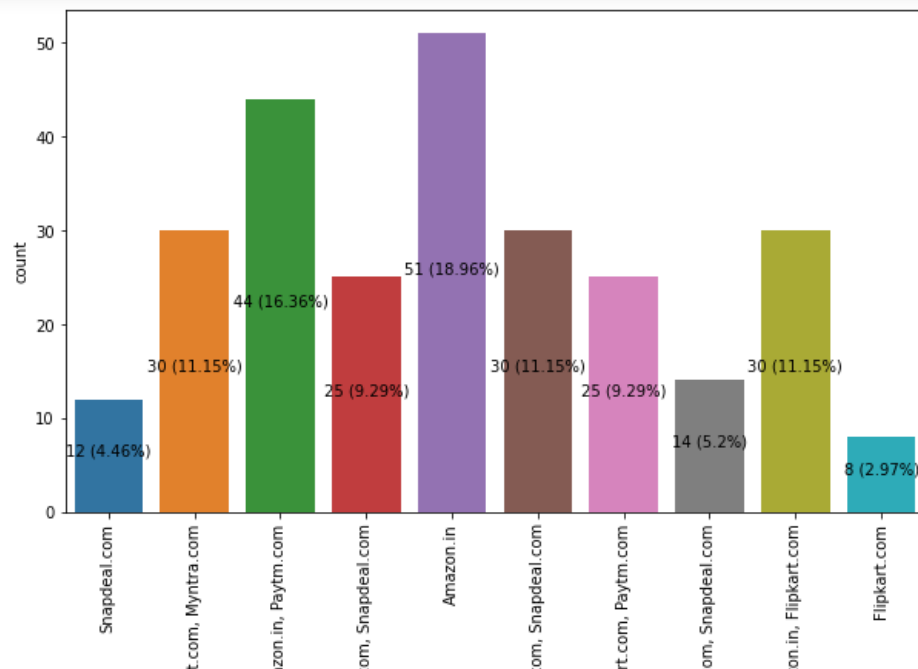
most of the people are strongly agree that online shopping is easy and convinint and very fer people believe its not.

11.checking the data distribution using count plot for Getting value for money spent.



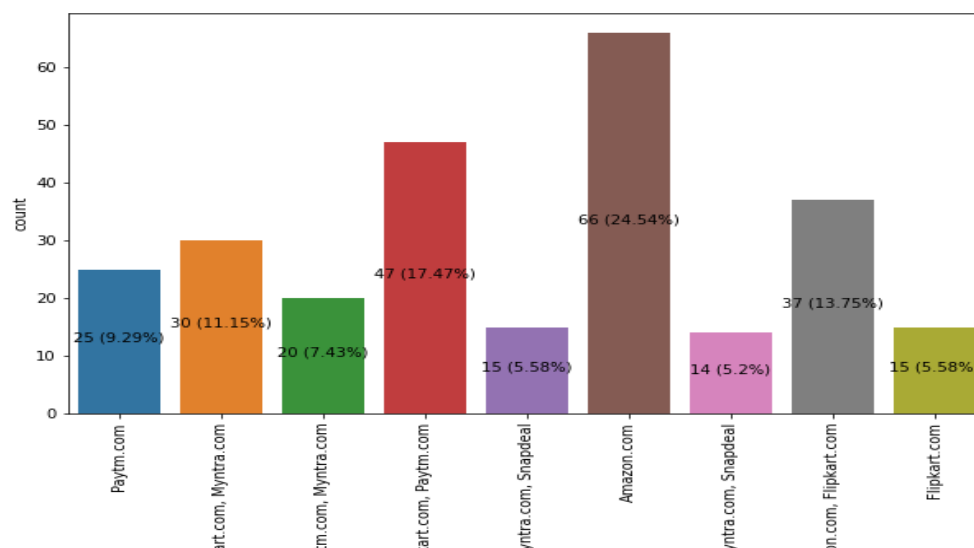
we can see maximum number of customers believe the they getting the value for money and very fer are not satisfy or they can't considered yet.

12. checking the data distribution using count plot for Fast loading website speed of website and application.



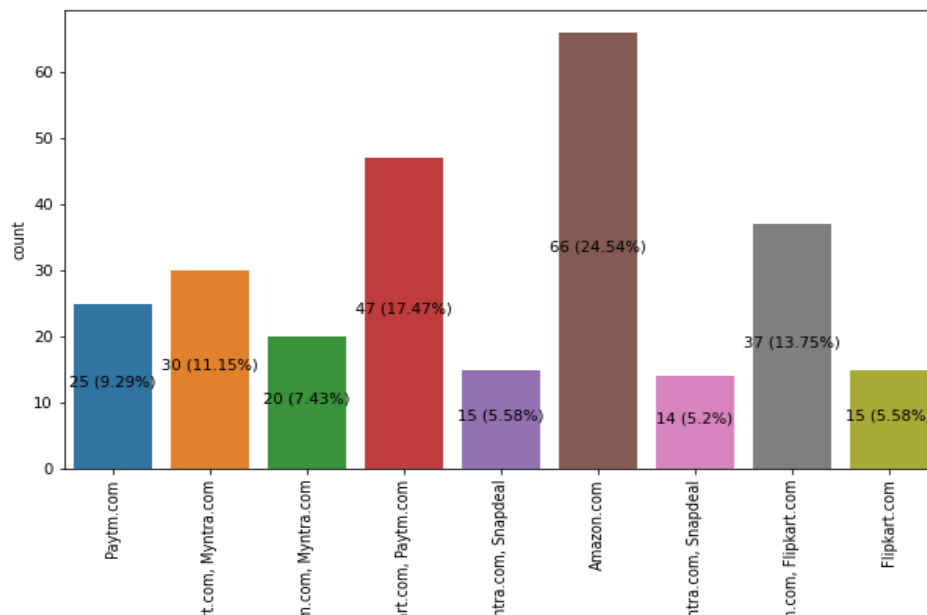
amazon is the leading website for e-commerce and its used by most of the customers and flipkart is back in the list.

13. checking the data distribution using count plot for Quickness to complete purchase.



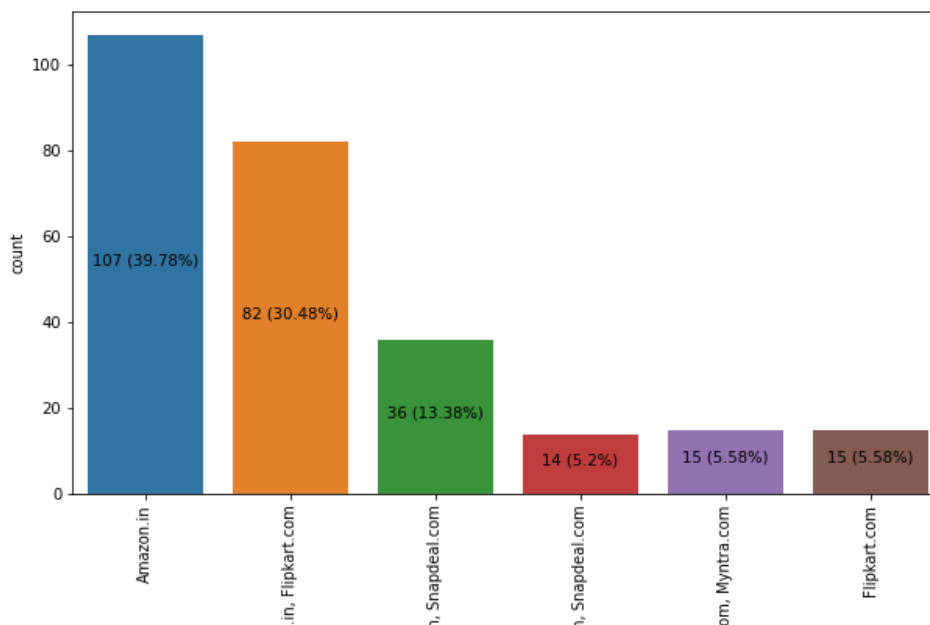
amazon is the quickest site to complete the purchase in other words customers can complete the purchase fastest as compared to other websites and flipkart is slower in the list.

14. checking the data distribution using count plot for Speedy order delivery.



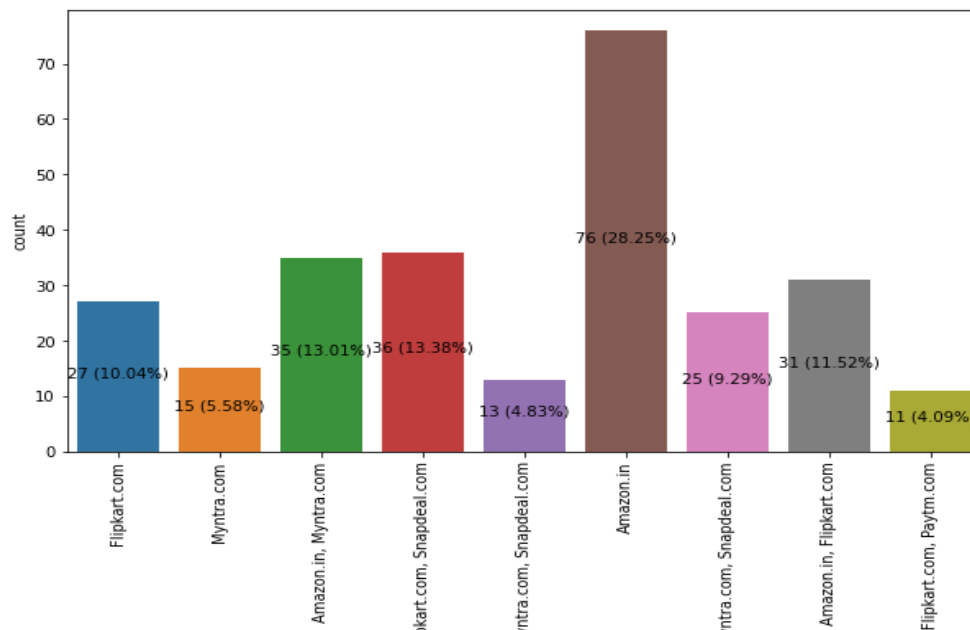
amazon is the fastest site to deliver the products to the customers and flipkart is slowest.

15. checking the data distribution using count plot for Perceived Trustworthiness.



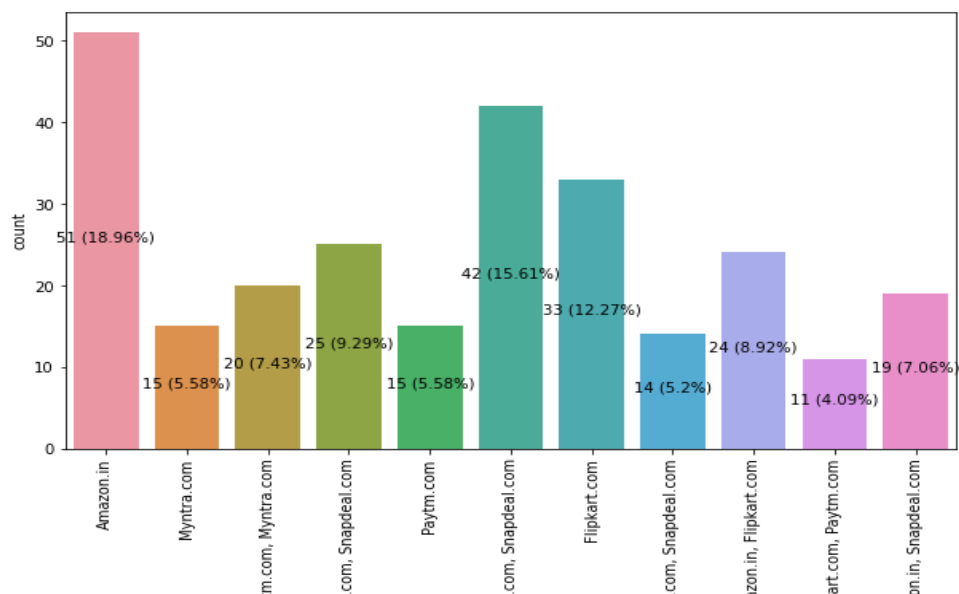
amazon is the fastest site to deliver the products to the customers and flipkart is slowest.

16. checking the data distribution using count plot for Perceived Trustworthiness.



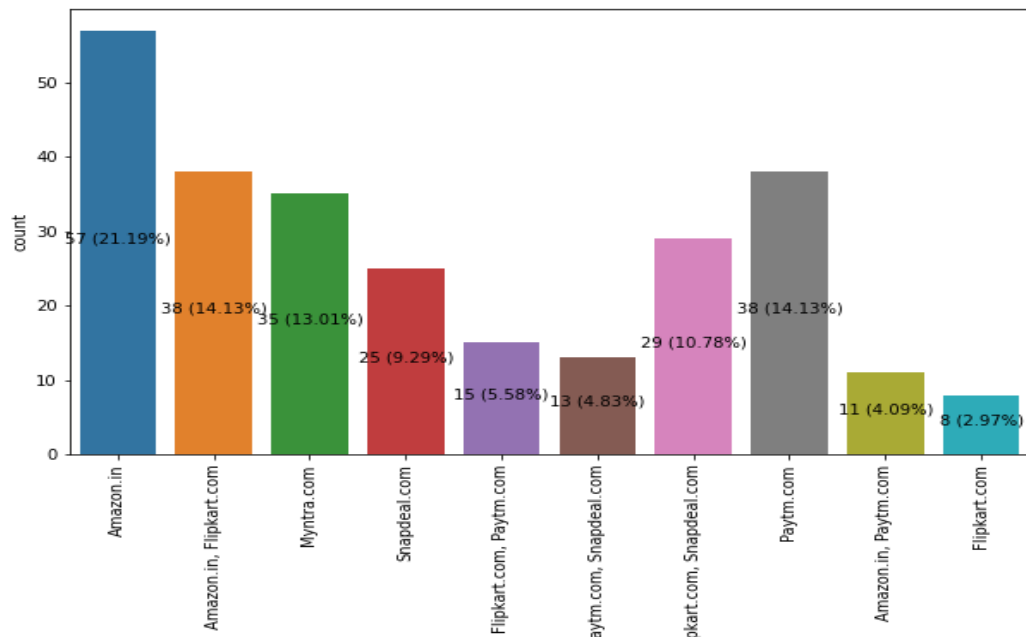
amazon is top in trustworthiness for customers and myntra is not upto trustworthiness.

17. checking the data distribution using count plot for Security of customer financial information.



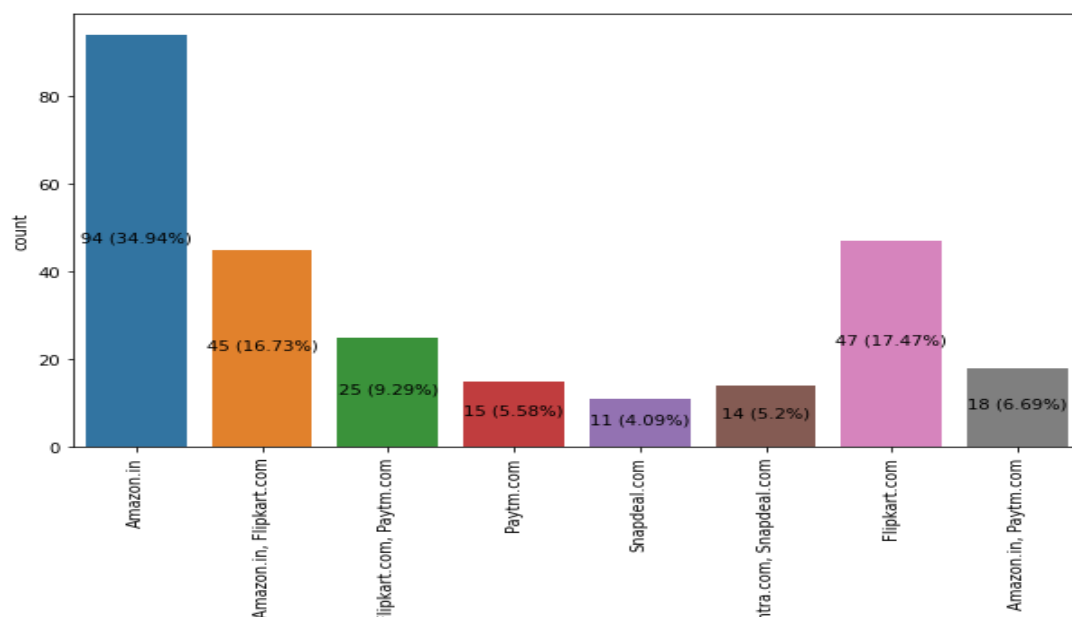
amazon is the best website for security of customer financial information and myntra does not suggested by costomers.

18. checking the data distribution using count plot for Longer time to get logged in (promotion, sales period).



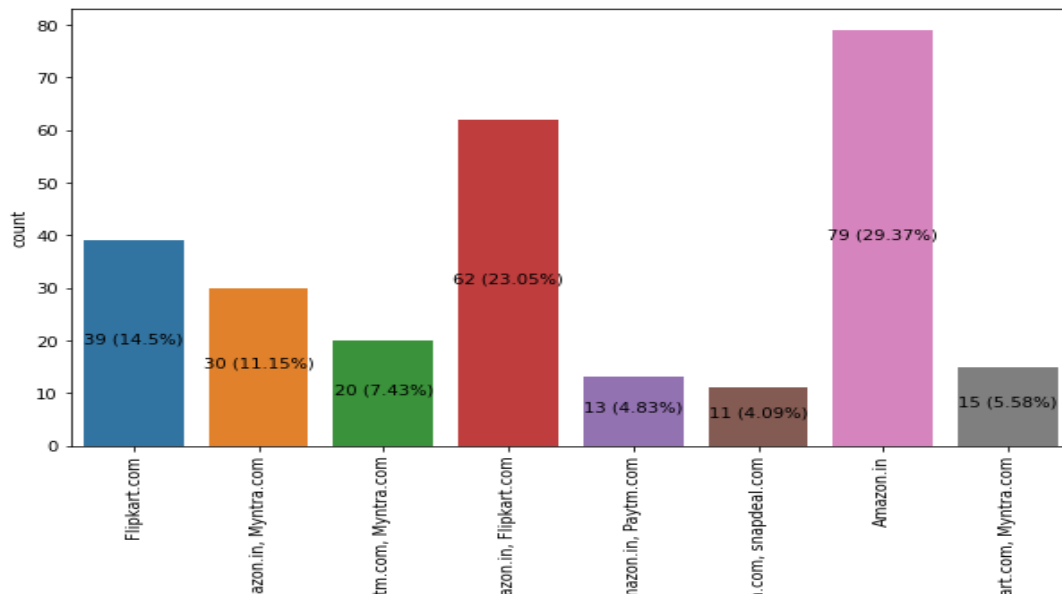
amazon have all the good things but amazon takes too long to get logged in for users. and flipkart is the quickest site for logging in.

19. checking the data distribution using count plot for Website is as efficient as before.



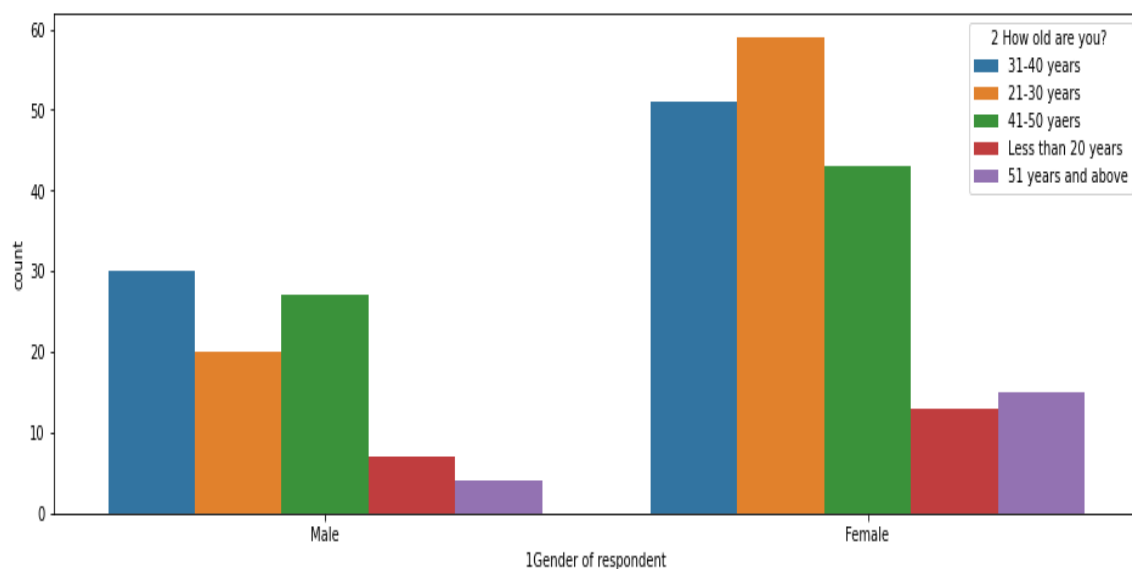
according to our data amazon.in is as efficient as it was before but snapdeal, paytm and flipkart are changing.

20. checking the data distribution using count plot for Which of the Indian online retailer would you recommend to a friend?



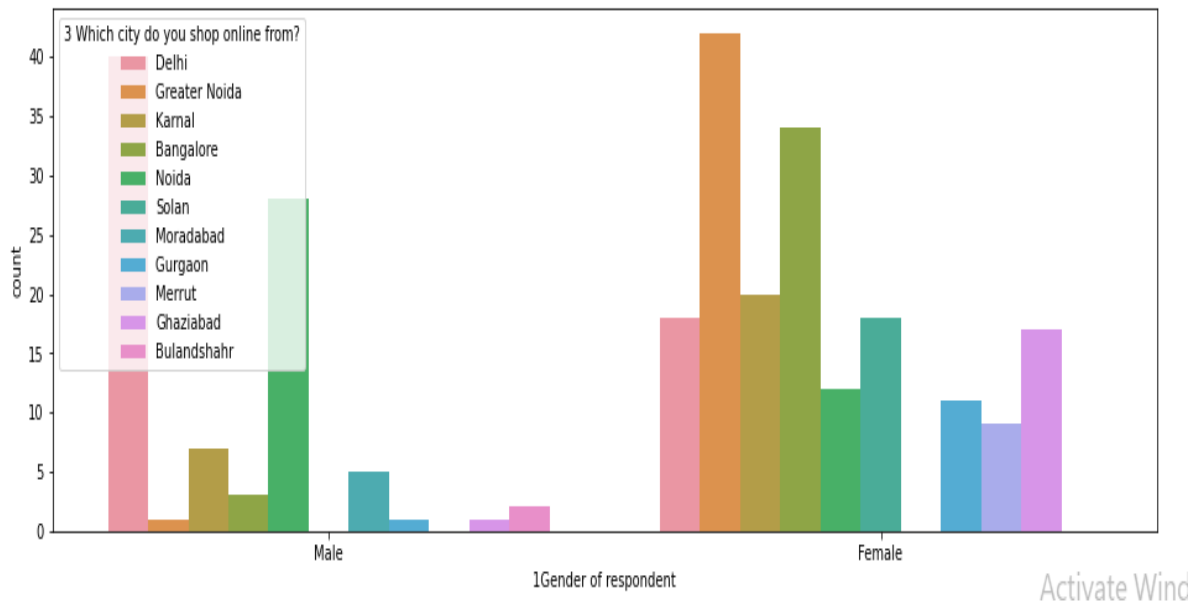
most of the people are recommending amazon.in to their family and friends while very few people are recommending paytm to their family and friends.

21. Checking relationship between gender of respondent and age.



we can see most of male and female customers are of ages 21 to 50 years. minimum number of male and female customers are of age less than 20 years.

22. Checking relationship between gender of respondent and city.



we can see most of female respondents are greater noida and most of male customers are from delhi. males and females both customers are from moradabad have less customers.

CONCLUSION

• Key Findings and Conclusions of the Study

We can see in the above analysis the e-commerce is a booming sector. and the number of users are increasing day by day as most of the customer find it enjoyable, value for money, they believe it's reliable and there are some trustworthy sites like amazon which provide the security to the user and financial protection also. some sites are delaying the orders but they are changing with the time so people are using the e-commerce site for shopping as using it is fun. so we can conclude that the e-commerce sites will boom in the future as number of customers are increasing and it's convenient. and amazon is leading the e-commerce market with the maximum number of buyers and maximum number of varieties of product on their site.