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

SUBMITTED BY : ARUNPRASAD.K

GITHUB LINK :
https://github.com/arunprasad-k

CONTENT

Abstract

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 on the basis of the Means End Chain theory (MEC) and Prospect theory. By hypothesis that a combination of both utilitarian value and hedonistic values are needed to affect the repeat purchase intention (loyalty) positively, Structural equation model has been presented on the primary data collected from the Indian online shoppers. Results indicate the e-retail success factors, which are very much critical for customer satisfaction. By increasing the utilitarian value and hedonistic values derived by the customers, customer satisfaction and hence the customers repeat purchase intention can be increased significantly.

General Terms

Exploratory Data Analytic, Uni-variate Analysis, Bi-variate Analysis, Outliers.

Keywords

- ✓ Data mining
- ✓ Feature Engineering
- ✓ EDA

1. Introduction

Large number of customers are getting attracted towards online retailing; this is because e-stores usually offer them a variety of services and products according to their preferences, Convenience, round the clock availability, flexible pricing, discounts as well as free door step delivery are some of the major benefits of shopping online. Presently, more number of online retailers are beginning to experience increase in demand for products and services . Indian online retail industry has been experiencing good times since the last six years; as a result of the constantly growing internet penetration, deployment of modern infrastructures, and a robust ecosystem for e-retail start-ups . Several e-commerce start-ups have commenced operation with innovative strategies, which differs from what was pioneered by first generation e-commerce companies.

“More than 1200 start-ups came up in 2018, including eight unicorns, taking the total number to 7200 in India”. India’s B2C e-commerce revenue grew from €20 billion in 2017 to reach €25 billion in 2018 at a growth rate of 20% . The number of internet users in India has been growing at a CAGR of 35% since the year 2007 according to a report by IAMAI-IMRB (2017). Having grown from just over a 100 million internet users in 2010, India has since touched 500 million internet users in 2018.

Online retail businesses are positioning themselves to be able to take advantage of the massive internet user base by turning them into online shoppers. Indian e-commerce market is poised to surpass that of the United States, making it the second largest in the world in less than twenty years from now. According to global payments firm Worldpay , this rapid growth is driven by a host of e-commerce players comprising of small, medium and large firms in terms of revenue and assets. An amalgamation of factors contributed to India’s rapid increase in number of online retail shoppers, which includes; digital friendly policies of the Govt. Large investments made by the online retailers, growing internet and smartphone penetration, demographics distributions, emerging middle class and young population.

With western markets getting saturated (matured) and China becoming more restrictive, India is becoming the main battleground for the e-retailers. The focus is attributed to the increased internet penetration, per capita income, rising middle class, urbanization and changes in consumer spending trends in India. Some foreign e-retailers like Amazon. com and Walmart (through Flipkart.com, Myntra.com, Jabong.com) are already playing with a good market share, while others like Alibaba.com, Rakuten.com are subtly trying to gain access to market through investment in small or big companies.

Therefore, knowing how to maximise the repurchase intention of Indians online consumers’ is vital for an online retailer in India to achieve its business goals. This may further lead to develop a general reference model for successful online retail business. Success of an online retail website depends on its system quality and how much consumer motivator values are derived through shopping from it. “System quality is the level of user satisfaction with the technical and functional aspects and is dependent on factors like: response time, usability, availability, reliability, and adaptability” . On the other hand, motivational values are referred to as either utilitarian value or hedonic value . The present work has focused upon these values and attempted to see their role in the customer satisfaction and purchase decisions. The primary objective of this study is to propose and empirically test the online retail success model with emphasis on proving values to the customers.

2. **Theoretical Background**

Studies have considered “customer repurchase intention” as a reflection of loyalty and are being frequently used as a construct of e retail success estimation instead of “intention to use”. As a result of the increase in popularity and prevalence of online shopping, DeLone & McLean’s e-commerce extended success model was unable to completely measure an e-Retail’s websites operation level of success. Studies have suggested that past online shopping experience, perceived usefulness, and customer satisfaction are factors capable of influencing a customer’s repurchase intention .

Website quality (e-store) and the usability of the e-store have also been proposed as being very vital for e-retail success by studies . In order to recommend a comprehensive model, the work adopted the Quality-Value-Satisfaction-Loyalty model; as the mechanism for which an online retailer may successful. The customer retention model illustrates that for a customer to become loyal to an online retail brand, there must be satisfaction, which arises when the e-tailer possesses a quality system backed-up by the proper mix of values. Shopping values refers to the evaluations of experience of shopping with an e-store; such as the extra convenience of shopping online. It focuses on the experience of purchasing the preferred product, instead of the product itself . online customers is divided into two most critical predictors a) shopping (hedonic and utilitarian) value (in formativeness, efficiency and entertainment) and website features. In comparison to this, previous studies have used variables such as websites, internet performance, perceived profit, perceived risk, variables for website selection, and differences in lifestyle.

This category of e-retail customer would be motivated to shop on an e-vendor with widely selection of products, informative, convenient, and fast websites . Both categories of shopping values have been presented in Fig. 1.

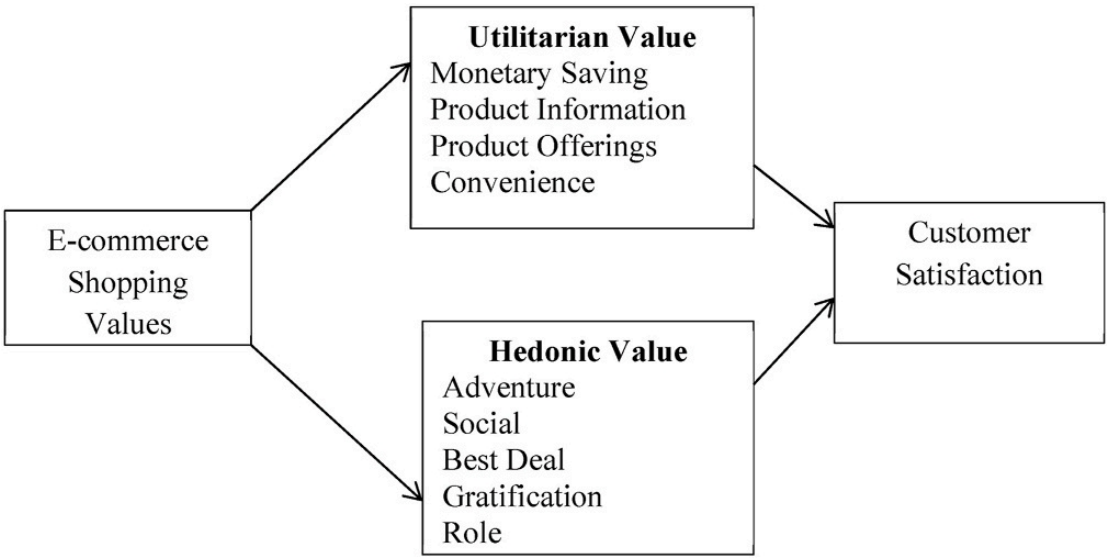


Fig. 1. E-commerce shopping values

3.Analytical Problem Framing

DATA ANALYTICS AND ITS CATEGORIES

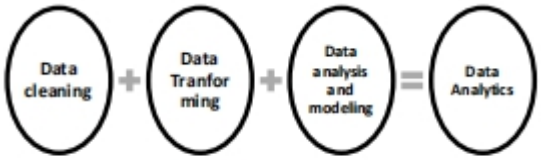


Fig 1: Data Analytics

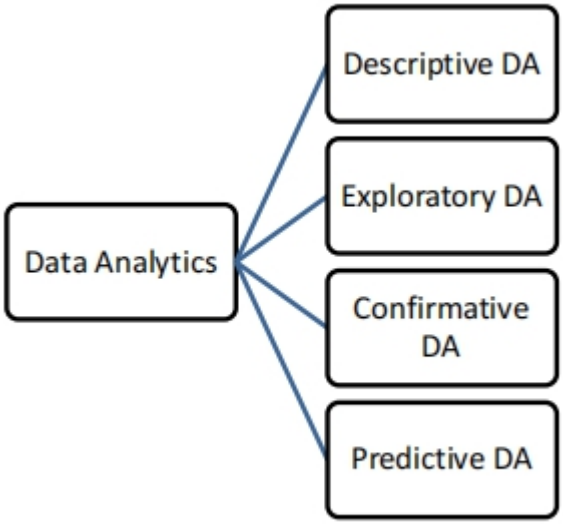


Fig 2: Categories of Data Analytics

4) BASIC STATISTICS ABOUT OUR DATASET

4.1 Data formats

#	Column	Non-Null Count	Dtype
0	gender	269 non-null	object
1	age	269 non-null	object
2	city	269 non-null	object
3	pincode	269 non-null	int64
4	years of shopping online	269 non-null	object
5	no.of.times purchased online	269 non-null	object
6	accessing internet	269 non-null	object
7	device used	269 non-null	object
8	screen size	269 non-null	object
9	os	269 non-null	object
10	browser used	269 non-null	object

11	channel follow	269 non-null	object
12	reaching store after 1st visit	269 non-null	object
13	time to explore	269 non-null	object
14	preferred payment option	269 non-null	object
15	abandoning cart frequently	269 non-null	object
16	why abandon the bag	269 non-null	object
17	easy to read	269 non-null	object
18	info on product comparision	269 non-null	object
19	info on purchase decision	269 non-null	object
20	info on product stated clearly	269 non-null	object
21	ease of navigation	269 non-null	object
22	processing speed	269 non-null	object
23	user friendly interface	269 non-null	object
24	easy Payment methods	269 non-null	object
25	trust on online retail	269 non-null	object
26	empathy towards customers	269 non-null	object
27	gaurantee on privacy	269 non-null	object
28	availability on communication channel	269 non-null	object
29	benifits and discounts	269 non-null	object
30	enjoyment is derived	269 non-null	object
31	convenient/flexible	269 non-null	object
32	return/replacement policy	269 non-null	object
33	gaining access to loyalty	269 non-null	object
34	display quality info satisty customers	269 non-null	object
35	user derive satisfaction	269 non-null	object
36	Net Benefit from shopping online can lead to users satisfaction	269 non-null	object
37	User satisfaction cannot exist without trust	269 non-null	object
38	Offering a variety of product in several category	269 non-null	object
39	Provision of complete and relevant product information	269 non-null	object
40	monetary savings	269 non-null	object
41	The Convenience of patronizing the online retailer	269 non-null	object
42	Shopping on the website gives sense of adventure	269 non-null	object
43	Shopping on e-tailer enhances social status	269 non-null	object
44	feel gratification shopping on favorite e-tailer	269 non-null	object
45	Shopping on the website helps fulfill roles	269 non-null	object
46	Getting value for money spent	269 non-null	object
47	Tick any (or all) of the online retailers you have shopped from	269 non-null	object
48	Easy to use website/app	269 non-null	object
49	Visual appealing web-page layout	269 non-null	object
50	Wild variety of product on offer	269 non-null	object
51	Complete, relevant description information of products	269 non-null	object
52	speed of website/app	269 non-null	object
53	Reliability of the website/app	269 non-null	object
54	Quickness to complete purchase	269 non-null	object
55	Availability of several payment options	269 non-null	object
56	Speedy order delivery	269 non-null	object
57	Privacy of customer info	269 non-null	object
58	Security of customer financial info	269 non-null	object
59	Perceived Trustworthiness	269 non-null	object
60	Presence of online assistance through multi-channel	269 non-null	object
61	Longer time to get log in	269 non-null	object

62	Longer time in displaying graphics and photos	269 non-null	object
63	Late declaration of price	269 non-null	object
64	Longer page loading time	269 non-null	object
65	Limited mode of payment on most products	269 non-null	object
66	Longer delivery period	269 non-null	object
67	Change in website/Application design	269 non-null	object
68	disruption moving from one page to another	269 non-null	object
69	Website is as efficient as before	269 non-null	object
70	Which of the online retailer would you recommend to a friend?	269 non-null	object

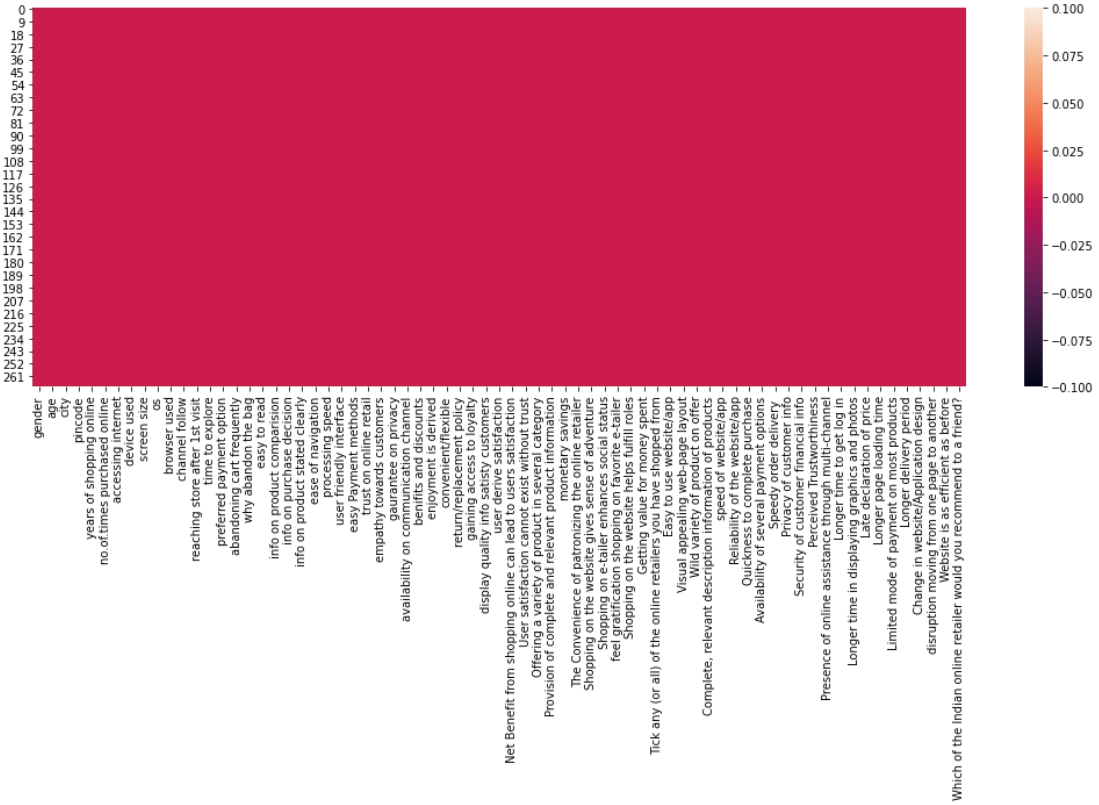
4.2 Data Dimensions

The data provided consist of **269 rows and 71 columns**

Now let us explore our data set by knowing the influence of each attribute on the column label. We will create histograms, Bar plots to achieve this.

5.Data Cleaning

Before applying any type of data analytics on the data set, the data should be first cleaned.Here we find there are no missing values in the data set which needed to be handled in case if any present**.**



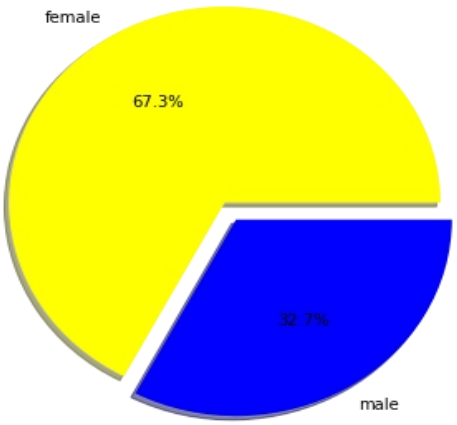
6. EXPLORATORY DATA ANALYSIS

We are going to perform exploratory data analysis for our problem in the first stage. In exploratory data analysis data set is explored to figure out the features which would influence the customer retention. The data is deeply analyzed by finding a relationship between each attribute.

6.1 Uni variate Analysis

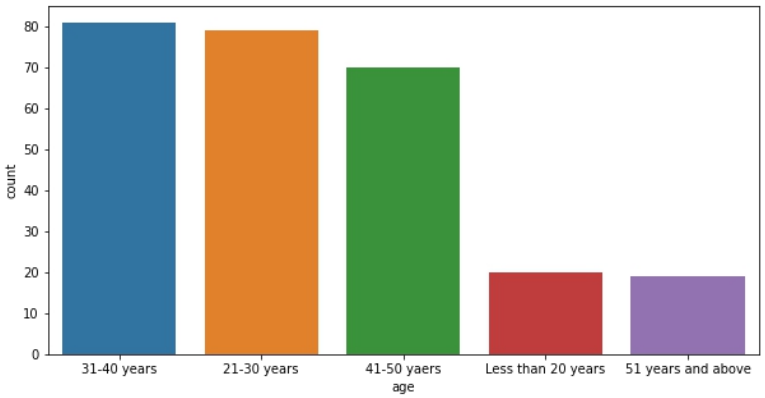
6.1.1 Gender

The demographics of the respondents showed that the number of females was higher with a share of 68%, while 32% were males.



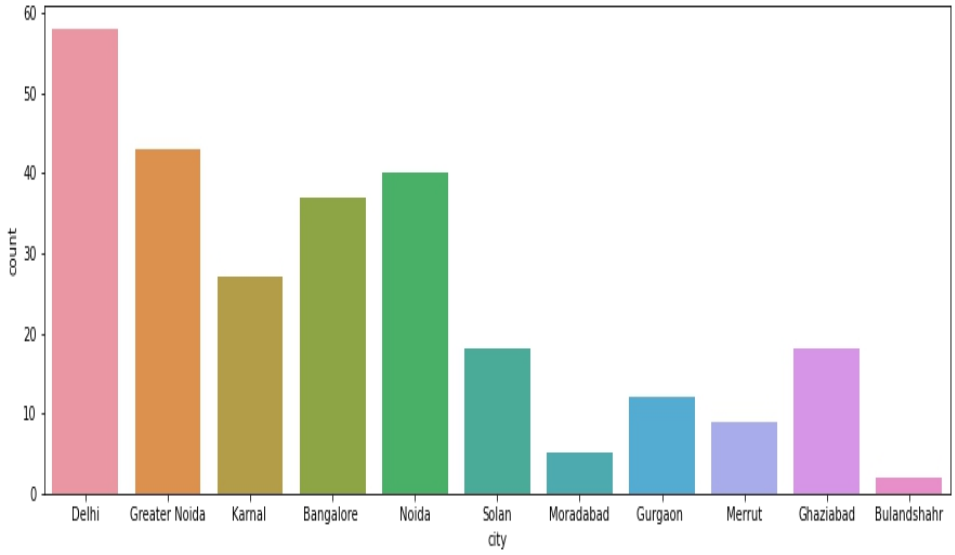
6.1.2 Age

Here we can see that people of age group between **31-40 years** seems to use the online retailing the most when compared to other age groups whereas the people of age group of less than **20 years and 51 years and above** were the least to use online retailing.



6.1.3 City

Regarding city wise the capital of India **Delhi** placed at the top among other cities with respect to people who shop online from , followed by **Bangalore**.The least among them being **Bulandshahr**



Customer activation model

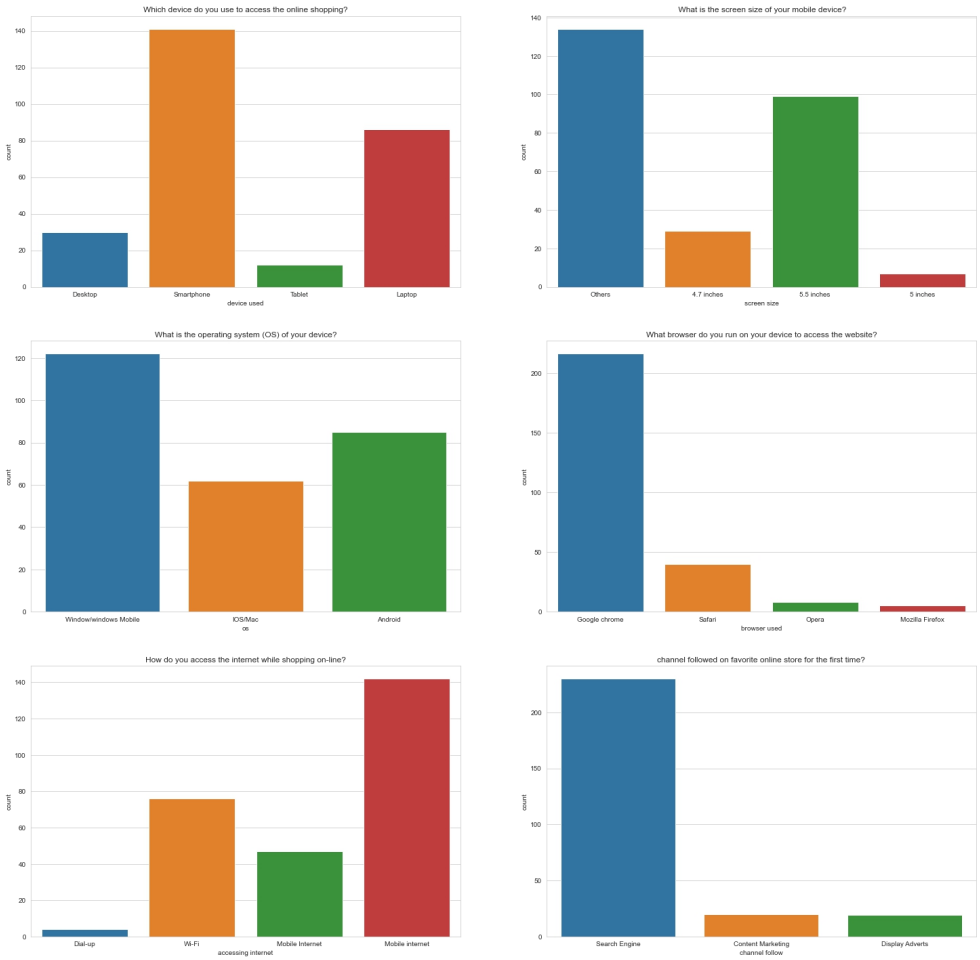
The research model in this study is based on the Online Customer Lifecycle used in e-commerce, it tracks both the customer's qualitative and quantitative journey in relation to a product . The model is based on statistical records of customer's progression through the phases of the process; focusing on how to optimize the performance of several touch points and channels of interaction. Goal of every e-commerce company is to acquire customers ensuring that there is a continuously acceptable conversion rate .

The model adopts the three qualities (System Quality, Information Quality, Use, and Service Quality) and its impacts on User Satisfaction, thus, intention of the customer to use the e-retail website or application is also affected. In this research model, "User Satisfaction" represents one of the principal measurements of success of an online retail website. Wang et al. also stated that the measurement of "User Satisfaction" is an objective based attitude.

6.1.4 system quality

This can be described as the users' perception of the online retail website's ability to efficiently provide requested information as well as its delivery method. System quality analyses the performance of the website; some of the metric includes; page load speed, ease of navigation between pages, website design, appearance, website availability, and website layout.

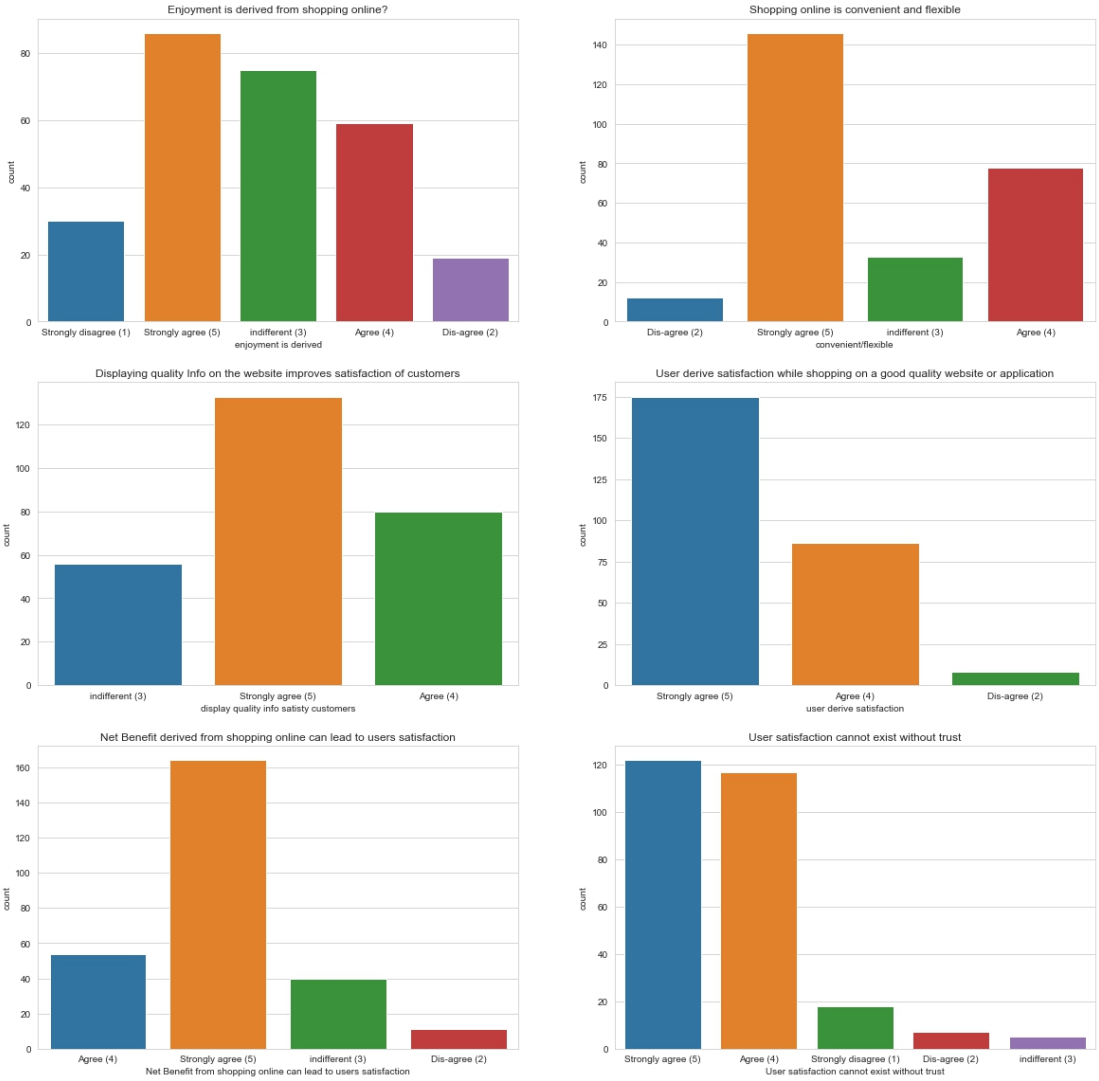
Technology acceptance model (TAM) states that, an e-retail website which is seen as being easy to use is most likely to bring about a feeling of positivity among the consumers. The performance of a website performs a significant role in deciding satisfaction of the customers while shopping on the website. When customers purchase a product from a website or web application, functionality snags (for example, Website crash, interruption) may result in a poor customer user experience . System quality positively impacts online customer satisfaction .



- From the count plot we can infer that the user who use **smartphone** device were more in number to access the online shopping
- People who uses smartphone prefer a big screen. from the graph we can see that most of the people (around 100) uses the phone with screen size of **5.5 inches**.
- Around 120 users use their device with **windows/windows** mobile as their operating system
- Most people prefer **google chrome** as the browser while accessing internet
- while accessing internet most people uses **mobile internet** to access the online shopping websites
- people prefer **search engine** as the channel followed on favorite online store for the first time. very least prefer the other two

6.1.5 User Satisfaction

Satisfaction has been defined as a personal feeling of contentment or pleasure, as a result of comparing the actual outcome of shopping online as against their expectation . Satisfaction of online customers will positively impact their future intention (repeat purchase) directly or indirectly.

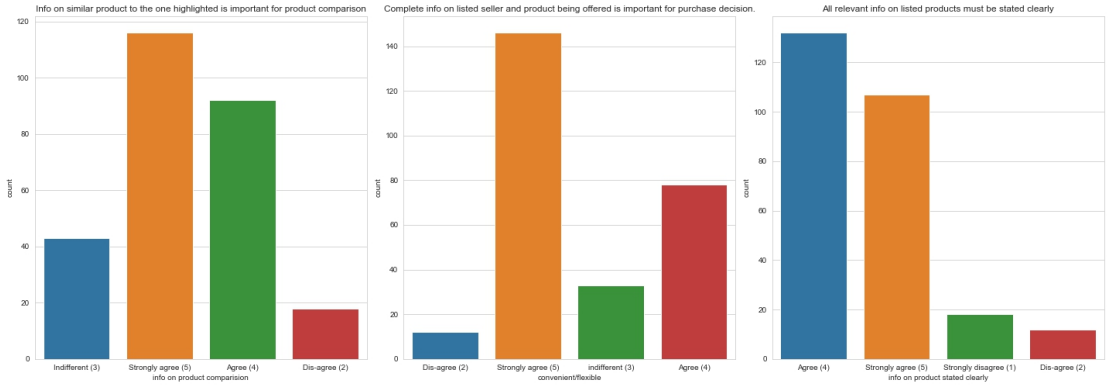


- **Enjoyment is derived from shopping online?**- For this question people around 85 in count strongly agree that shopping online gives them enjoyment whereas only 18 person disagree to this opinion
- **Shopping online is convenient and flexible?**- From the graph we infer that people strongly agree that shopping online is convenient and flexible rather than offline shopping.

- **Displaying quality Info on the website improves satisfaction of customers?**- The data clearly shows 3 kind of people opt for this statement.In that option around 170 people strongly agree that Displaying quality Info on the website improves their satisfaction
- **User derive satisfaction while shopping on a good quality website or application?**- From the data we infer that most of the people strongly agree that a good website satisfy them while shopping online.
- **Net Benefit derived from shopping online can lead to users satisfaction?** - Most of the people strongly agree that they get satisfied by doing online shopping with net benefits
- **User satisfaction cannot exist without trust** - The count plot clearly shows people mostly believe that User satisfaction cannot exist without trust. Nearly 120 people strongly agree to it where as 115 people chooses agree.

6.1.6 Information Quality

This can be defined as the customers' perceptions towards the presentation and characteristics of information presented or displayed in an e-commerce website, web application or portal. It takes the characteristics of information into consideration, for example, timeliness, relevance, understandable, accuracy, and completeness. One of the basic responsibilities of an e-tailor is to ensure the availability of complete information on products being offered, the transitional process involved, and services to be rendered. Displaying high quality information on the website can stimulate an improved conversion rate as well as better customers' satisfaction. However, outdated or inaccurate information can lead to customers' dissatisfaction with the online retailer . An online retail website having the right information content quality is important for online retail customer satisfaction.



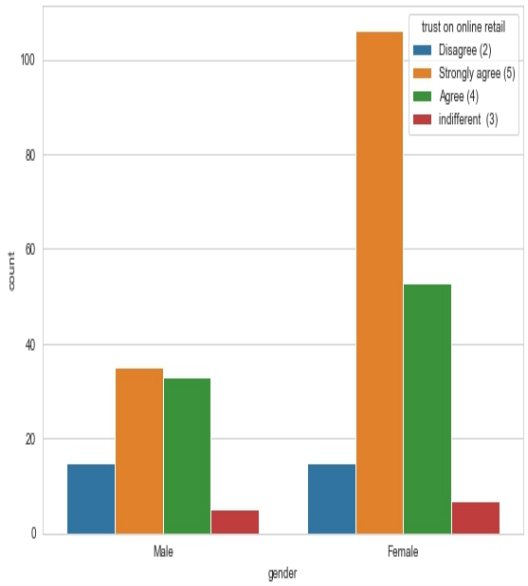
- Most of the respondent **strongly agrees** that Info on similar product to the one highlighted is important for product comparison.
- very few **disagree** to the fact that Complete info on listed seller and product being offered is important for purchase decision. Around 150 respondent **strongly agree** to it.

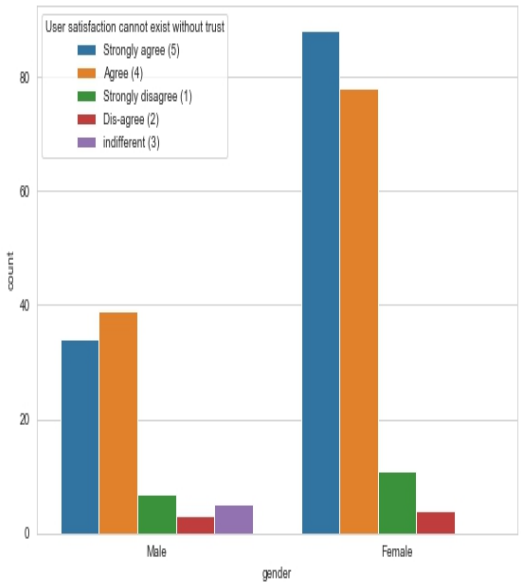
- only people around 110 choose **strongly agree** whereas around 130 choose **agree** on relevant info on listed products must be stated clearly. This clearly shows the e retailer must improve their information quality on the respective field.

6.2 Bi-variate Analysis

6.2.1 Trust

According to a study, trust is the perception of a consumer that the merchant will transact with all ethically, capability, and efficient standard. This is also supported by the Theory of Planned Behaviour. Which, further states that trust produces a positive feelings toward the online retailers; this increases the chances of customer's intention to buy a products, or use a service. Whenever there is no trust in a process, it stops online consumers from continuing with the transaction, as a result of this, they may never transact with the online retailer again if it does not improve on its trustworthiness . Generally, when an online retailer is not trusted, customer would not buy from the online retailers, this may be as a result of the perception that the e-retailer will not behave ethical, socially accepted manner. Trust begins as a result of an online retailer being able to fulfil its objectives.

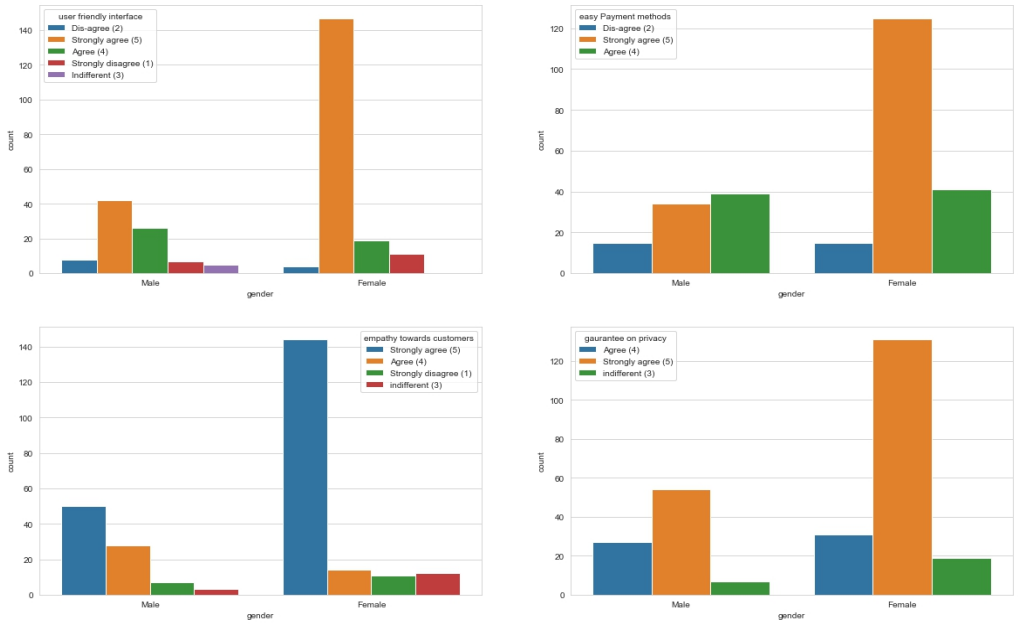




- Trust and commitment are the central tenets in building successful long-term relationships in the online retailing context. Here female respondents **strongly agree that trust is an important factor** for successful customer retention
- most of the Male Respondent doesn't agree that user satisfaction cannot exist without trust.But seeing female respondent view it is complete reciprocal to the male respondent view.over 85 female users **agree that trust is the main factor for user satisfaction.**

6.2.2 Service Quality

Service quality refers to how well the services delivered by an online retail store are able to match the expectations of the customer. Examples of such services may include; contact, responsiveness, and privacy. For online retailers, making available several channel for customers to be able to communicate across to the online vendor when assistance is required. This is crucial, if the quality of an e-retailer service operation has to improve. Responsiveness, which is concerned with the competence to handle customer concerns and returns can be improved upon by making available support through email, text, presence of online representatives and telephone. Online consumers will not shop on an online portal, which does not guarantee that credit or debit card information are secured against third party misuse.



- Among the respondents,female customers **strongly agrees** to the fact that the E-retail Shopping must be **user friendly** interface. But the male Respondents does not agree much to this fact.
- Female Respondents believe that **easier payment method** to shop online will be best for better E-retail shopping.Around 130 female users **strongly agrees** this whereas only half of the male respondents w.r.t female users agrees to this.
- showing empathy helps customers to feel that their concerns matter. Female users **strongly agrees** that **empathy towards customers** will increase the trust on service quality whereas many male respondents does not agree to this point.
- Handling and protection of the sensitive personal information of the users is very important. Female respondents around 140 **strongly agrees** on **Guarantee on privacy** increase the customer retention and male respondents around 50 **strongly agrees** to it.

7. Categorical encoding using Label-Encoder

Label Encoding refers to converting the **labels** into numeric form so as to convert it into the machine-readable form. Machine learning algorithms can then decide in a better way on how those **labels** must be operated. It is an important pre-processing step for the structured dataset in supervised learning.

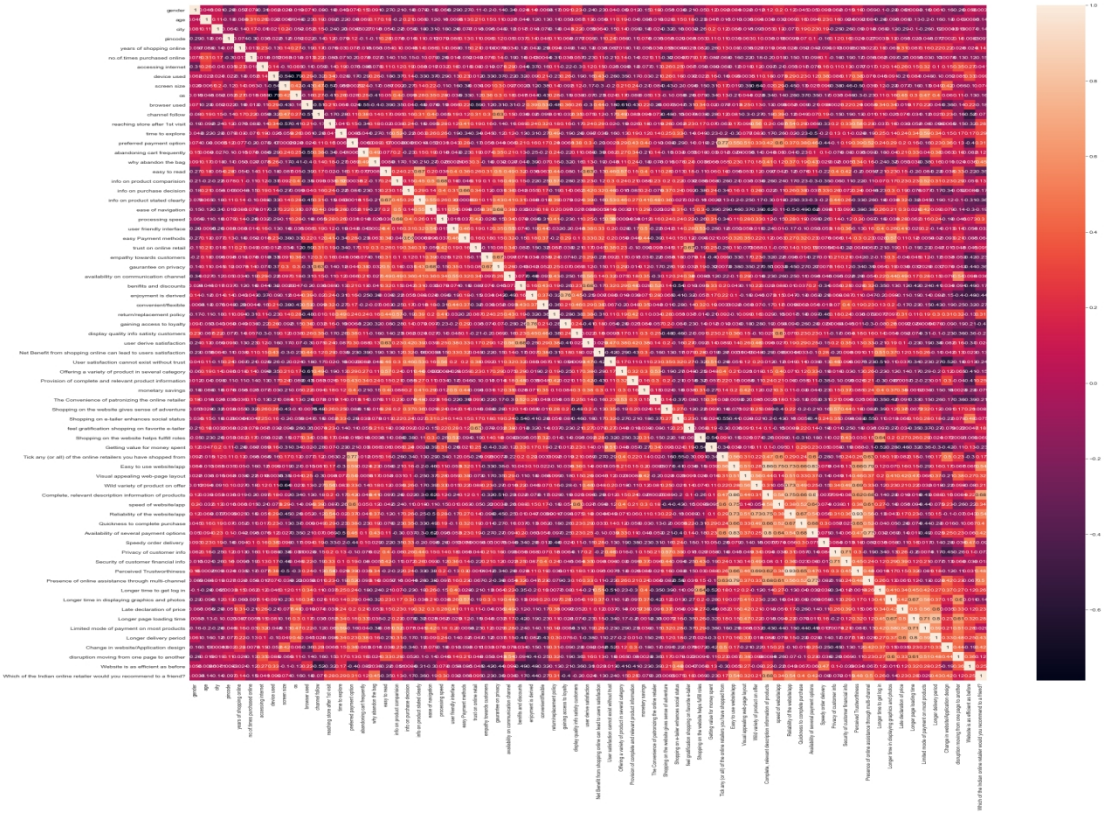
```
#From the given data-set we can infer that all the data types are categorical
#we convert them to integer type by using the label encoder method
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
for columns in df.columns:
    df[columns]=le.fit_transform(df[columns])
```

Once the Encoding is done we proceed with Feature Engineering.

8. Feature Engineering

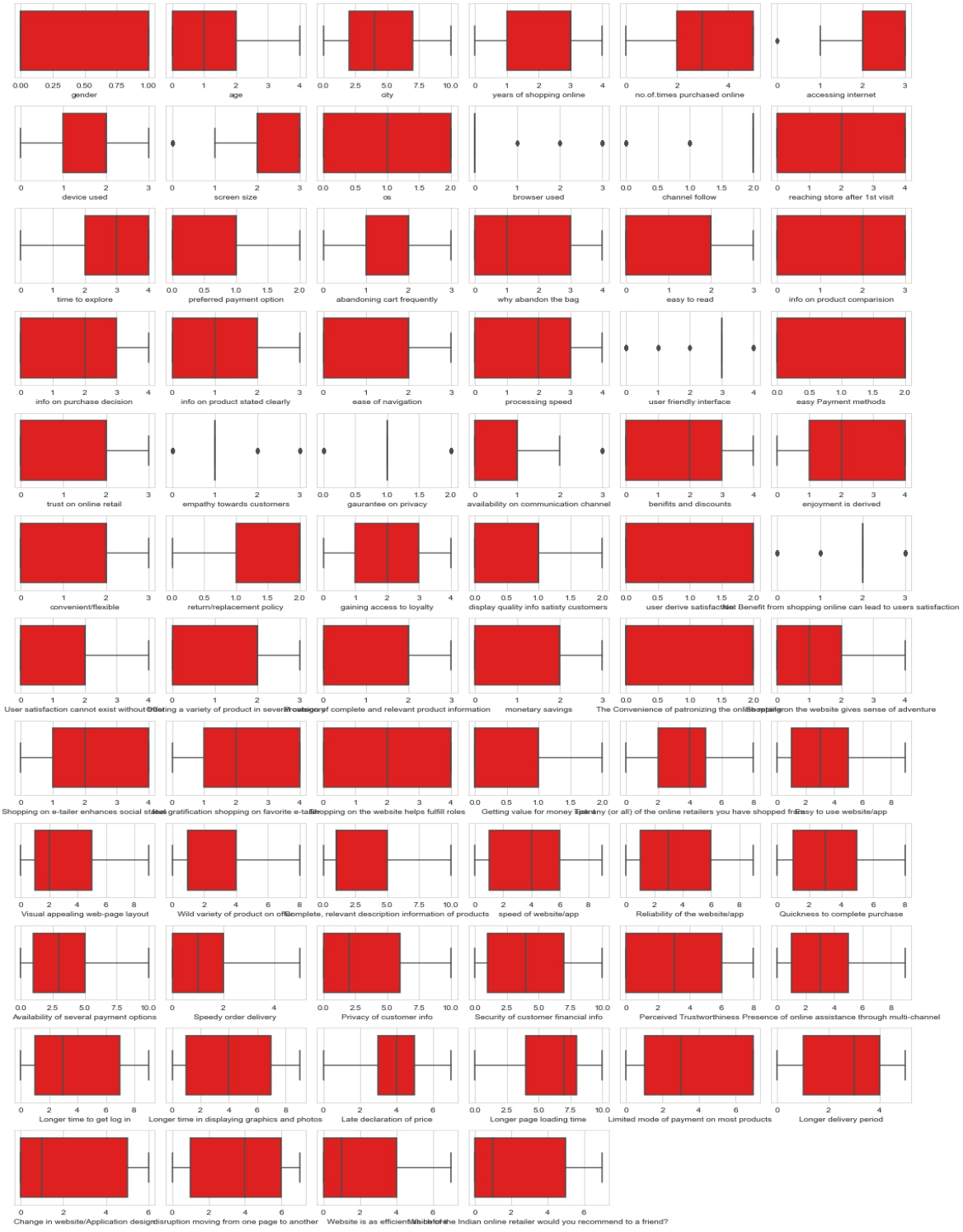
Feature engineering is the most important part of data analytic process. It deals with, selecting the features that are used in training and making predictions. In feature engineering the domain knowledge is used to find features in the dataset which are helpful in building machine learning model. It helps in understanding the dataset in terms of modeling. A bad feature selection may lead to less accurate or poor predictive model. The accuracy and the predictive power depend on the choice of correct features. It filters out all the unused or redundant features.

8.1 Correlation



8.2 Looking for Outliers

The difference between a good and an average machine learning model is often its ability to clean data. One of the biggest challenges in data cleaning is the identification and treatment of outliers. In simple terms, outliers are observations that are significantly different from other data points. Even the best machine learning algorithms will underperform if outliers are not cleaned from the data because outliers can adversely affect the training process of a machine learning algorithm, resulting in a loss of accuracy.



Percentage of Data Loss

loss_of_data=(269-249)/269*100

loss_of_data=7.43%

- ✓ when we try to remove outliers it results in loss of data around 7%.
- ✓ since the loss of data is not much as expected,so proceeding with outlier removal.

Conclusion

- ✓ Exploratory data analysis is one of the key competencies of a data scientist at a startup. We should be able to dig into a new data set and determine how to improve our product based on the results. EDA is a way of understanding the shape of a data set, exploring correlations within the data, and determining if there's a signal for modeling an outcome based on the different features.
- ✓ Here we have done the EDA for the given data set.
- ✓ From here we can proceed with train test split and predict using Machine learning algorithm