Project Name

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

Submitted by:

Rupali Rane

ACKNOWLEDGMENT

I would like to express my deep sense of gratitude to **Flip Robo Technologies** who gave me the golden opportunity to do this data analysis project on **Customer Retention Dataset Analysis**, which also helped me in doing lots of research and I came to know about so many new things.

I have put in my all efforts while doing this project.

Rupali Rane

INTRODUCTION

• Business Problem Framing:

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

Customer retention refers to the ability of a company or product to retain its customers over some specified period. High customer retention means customers of the product or business tend to return to, continue to buy or in some other way not defect to another product or business, or to non-use entirely. Selling organizations generally attempt to reduce customer defections. Customer retention starts with the first contact an organization has with a customer and continues throughout the entire lifetime of a relationship and successful retention efforts take this entire lifecycle into account. A company's ability to attract and retain new customers is related not only to its product or services, but also to the way it services its existing customers, the value the customers actually perceive as a result of utilizing the solutions, and the reputation it creates within and across the marketplace.

• Motivation for the Problem Undertaken

Successful customer retention involves more than giving the customer what they expect. Generating loyal advocates of the brand might mean exceeding customer expectations. Creating customer loyalty puts 'customer value rather than maximizing profits and shareholder value at the centre of business strategy'. The key differentiation in a competitive environment is often the delivery of a consistently high standard of customer service. Furthermore, in the emerging world of Customer Success, retention is a major objective.

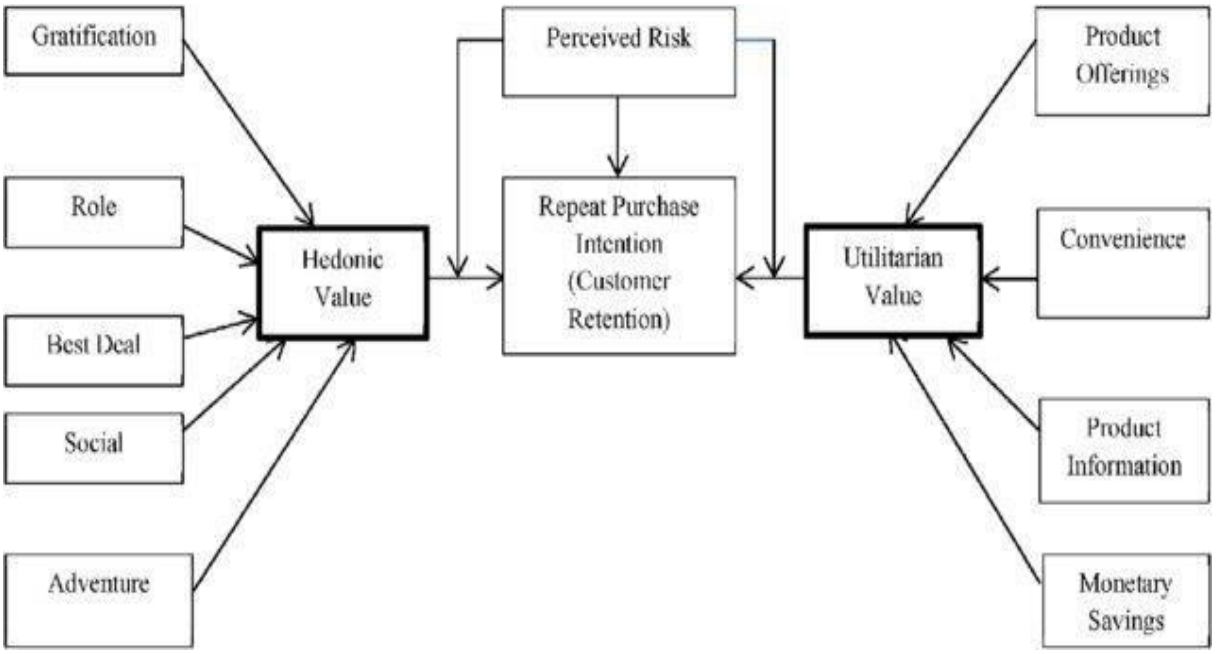
- **PROBLEM STATEMENT**

In this project, I have a dataset of different Online Shopping Site Like Amazon, Flipkart, Paytm, Snapdeal, Myntra. These All Sites Are Very Popular Everyone Use To But Something From these sites Some People Order More Things And Some People Order Fewer Things But While Use These App Some People Find Little Difficulty And They Also Don't Have Sometimes There Choice Payment Method Or Maybe They Are Getting That Particular Product At Higher Price As Compare To Other Sites Or Maybe Market And People Also Face Late Order Delivery.

Now To Solve All These Problems What People Do People Use To Give Feedback On Sites Now Why They Give There Feedback Because They Don't Want To Face The Same Problem That They Are Facing And Company Use To Take There Feedback Tries To Improve Them May Be They Give Discount To There Customers And Maybe They Reduce The Delivery Time Or Maybe They Can Add A Option As We Have On Amazon That Give 100rs Extra And Product Will Deliver By Tomorrow.

So, In This Data, I Have To Figure Out What Are The Problems People Are Facing Or What Company Can So To Increase There Sales And Can Earn More Profit And How They Can Improve There Site For Customers So They Don't Have To Suffer Anyone And Can Do Shopping More.

Diagram Representation of Customer Retention



The hedonic value has 5 values : gratification, role, best deal, social, and adventure.

The Utilitarian have four values: Product offering, Convenience, Product Information, and Monetary Saving

The customer retention is based on three main factors Hedonic value, Perceived Risk, Utilitarian value.

Hardware and Software Requirements and Tools Used:

- Hardware Used:
  - RAM: 8 GB
  - CPU: AMD A8 Quad Core 2.2 Ghz
  - GPU: AMD Redon R5 Graphics
- Software Tools used:
  - Programming language: Python 3.0
  - Distribution: Anaconda Navigator
  - Browser-based language shell: Jupyter Notebook
- Libraries/Packages Used:
  - Pandas
  - Numpy
  - Matplotlib
  - Seaborn

Analytical Problem Framing:

- Mathematical/ Analytical Modelling of the Problem

The provided dataset is in Microsoft Excel (xlsx) format. We will begin with loading the dataset and reading the dataset from the xlsx file using the read_excel() function from the Pandas Python package. Next, we will perform Non-Graphical Exploratory Data Analysis (EDA) such as checking the data types and missing values using pandas info() function, Next, we will move on to perform graphical EDA to get more insights from our dataset.

- Data Sources and their formats

The dataset is being provided by Flip Robo Technologies in .xlsx (Microsoft Excel) format and contains 269 records with 71 features as shown below:

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?**
11. **What browser do you run on your device to access the website?**
12. **Which channel did you follow to arrive at your favorite online store for the first time?**
13. **After first visit, how do you reach the online retail store?**
14. **How much time do you explore the e- retail store before making a purchase decision?**
15. **What is your preferred payment Option?**
16. **How 4 do you abandon (selecting an items and leaving without making payment) your shopping cart?**
17. **Why did you abandon the "Bag", "Shopping Cart"?**
18. **The content on the website must be easy to read and understand**
19. **Information on similar product to the one highlighted is important for product comparison**
20. **Complete information on listed seller and product being offered is important for purchase decision.**
21. **All relevant information on listed products must be stated clearly**
22. **Ease of navigation in website**
23. **Loading and processing speed**
24. **User friendly Interface of the website**
25. **Convenient Payment methods**
26. **Trust that the online retail store will fulfill its part of the transaction at the stipulated time**
27. **Empathy (readiness to assist with queries) towards the customers**
28. **Being able to guarantee the privacy of the customer**
29. **Responsiveness, availability of several communication channels (email, online rep, twitter, phone etc.)**
30. **Online shopping gives monetary benefit and discounts**
31. **Enjoyment is derived from shopping online**
32. **Shopping online is convenient and flexible**
33. **Return and replacement policy of the e-tailer is important for purchase decision**
34. **Gaining access to loyalty programs is a benefit of shopping online**
35. **Displaying quality Information on the website improves satisfaction of customers**
36. **User derive satisfaction while shopping on a good quality website or application**
37. **Net Benefit derived from shopping online can lead to users satisfaction**
38. **User satisfaction cannot exist without trust**
39. **Offering a wide variety of listed product in several category**
40. **Provision of complete and relevant product information**

41. **Monetary savings**
42. **The Convenience of patronizing the online retailer**
43. **Shopping on the website gives you the sense of adventure**
44. **Shopping on your preferred e-tailer enhances your social status**
45. **You feel gratification shopping on your favorite e-tailer**
46. **Shopping on the website helps you fulfill certain roles**
47. **Getting value for money spent**
48. **From the following, tick any (or all) of the online retailers you have shopped from;**
49. **Easy to use website or application**
50. **Visual appealing web-page layout**
51. **Wild variety of product on offer**
52. **Complete, relevant description information of products**
53. **Fast loading website speed of website and application**
54. **Reliability of the website or application**
55. **Quickness to complete purchase**
56. **Availability of several payment options**
57. **Speedy order delivery**
58. **Privacy of customers' information**
59. **Security of customer financial information**
60. **Perceived Trustworthiness**
61. **Presence of online assistance through multi-channel**
62. **Longer time to get logged in (promotion, sales period)**
63. **Longer time in displaying graphics and photos (promotion, sales period)**
64. **Late declaration of price (promotion, sales period)**
65. **Longer page loading time (promotion, sales period)**
66. **Limited mode of payment on most products (promotion, sales period)**
67. **Longer delivery period**
68. **Change in website/Application design**
69. **Frequent disruption when moving from one page to another**
70. **Website is as efficient as before**
71. **Which of the Indian online retailer would you recommend to a friend?**

Let's load it and start the analysis.

Importing Libraries:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
```

Loading the dataset:

```
df = pd.read_excel (r'C:/Users/Rupalipatil14/Desktop/customer_retention_dataset.xlsx', sheet_name='datasheet')
print (df)
```

```
     1Gender of respondent 2 How old are you?   \
0                     Male          31-40 years
1                   Female          21-30 years
2                   Female          21-30 years
3                     Male          21-30 years
4                   Female          21-30 years
..                     ...                 ...
264                 Female          21-30 years
265                 Female          31-40 years
266                 Female          41-50 yaers
267                 Female  Less than 20 years
268                 Female          41-50 yaers

     3 Which city do you shop online from?  \
0                                    Delhi
```

The data type information:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 269 entries, 0 to 268
Data columns (total 71 columns):
 #   Column
Non-Null Count  Dtype
---  ------
--------------  -----
 0   1Gender of respondent
269 non-null    object
 1   2 How old are you?
269 non-null    object
 2   3 Which city do you shop online from?
269 non-null    object
 3   4 What is the Pin Code of where you shop online from?
269 non-null    int64
 4   5 Since How Long You are Shopping Online ?
269 non-null    object
 5   6 How many times you have made an online purchase in the past 1 year?
269 non-null    object
```

Observation: We can see that there are 70 categorical data type columns and 1 numerical data type column present in the dataset.

Checking missing/null values present in the dataset:

```
#checking null count for each column
df.isnull().sum()
```

```
1Gender of respondent                                                0
2 How old are you?                                                   0
3 Which city do you shop online from?                                0
4 What is the Pin Code of where you shop online from?                0
5 Since How Long You are Shopping Online ?                           0
                                                                    ..
Longer delivery period                                               0
Change in website/Application design                                 0
Frequent disruption when moving from one page to another             0
Website is as efficient as before                                    0
Which of the Indian online retailer would you recommend to a friend?    0
Length: 71, dtype: int64
```

```
df.isnull().sum().sum()
```

```
0
```

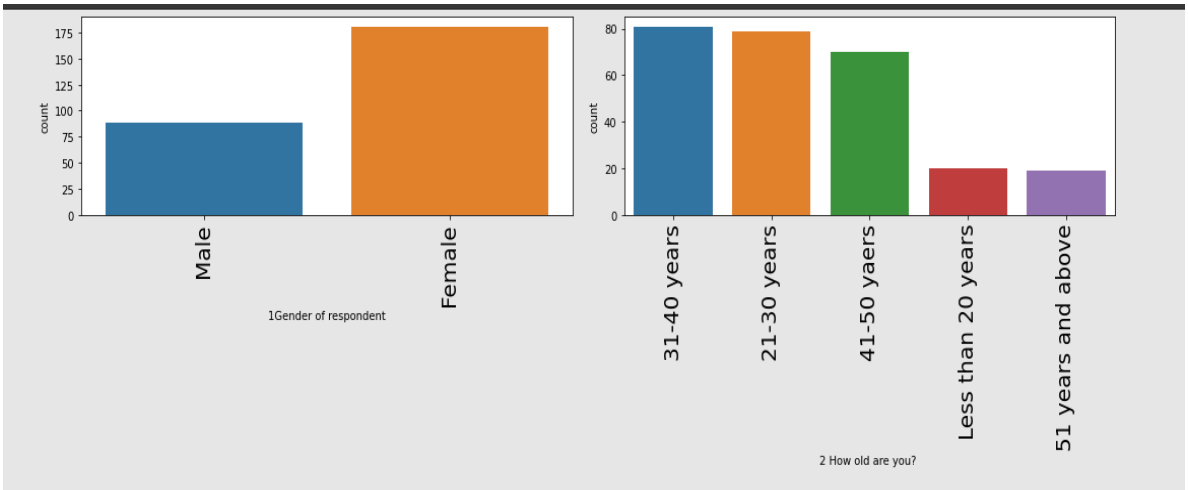
There are no missing values present in the dataset.

Visualizations:

To better understand the data, the following types of visualizations have been used: 1. Univariate, 2. Bivariate.

1. Univariate Analysis: Univariate analysis is the simplest form of data analysis where the data being analyzed contains only one variable. In this project analysis, I have used count-plot because it helped to understand each feature of data like people of which age used which site more, which type of device more to order any product, and people of which age or gender researched about the product before buying it, etc.

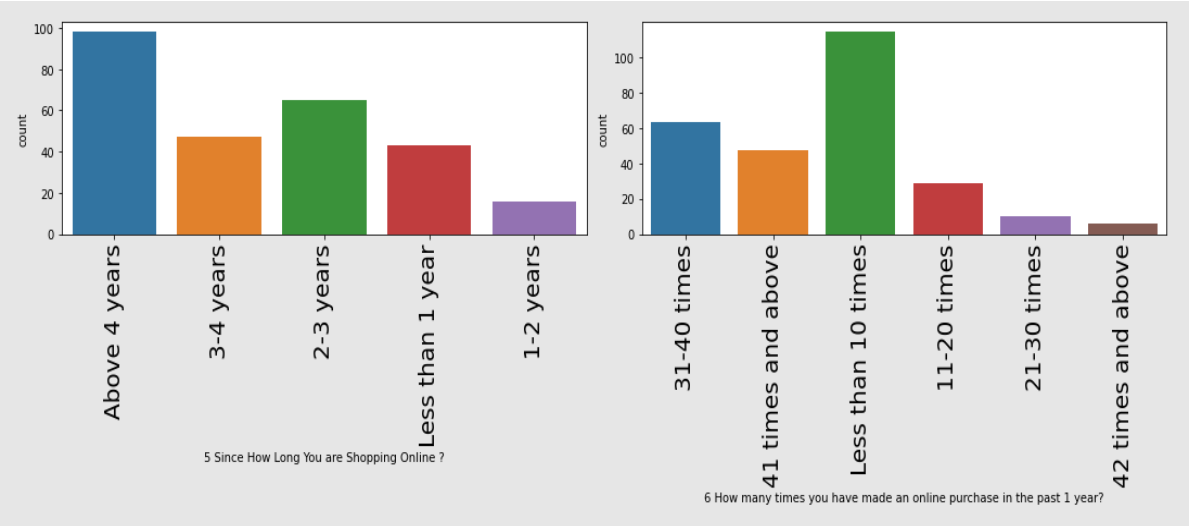
   Some of the count-plots and the observations made from them are shown below.



Observation:

1. From the count plot of gender, we can say male customers are fewer while female customers are more which means females use to do more shopping

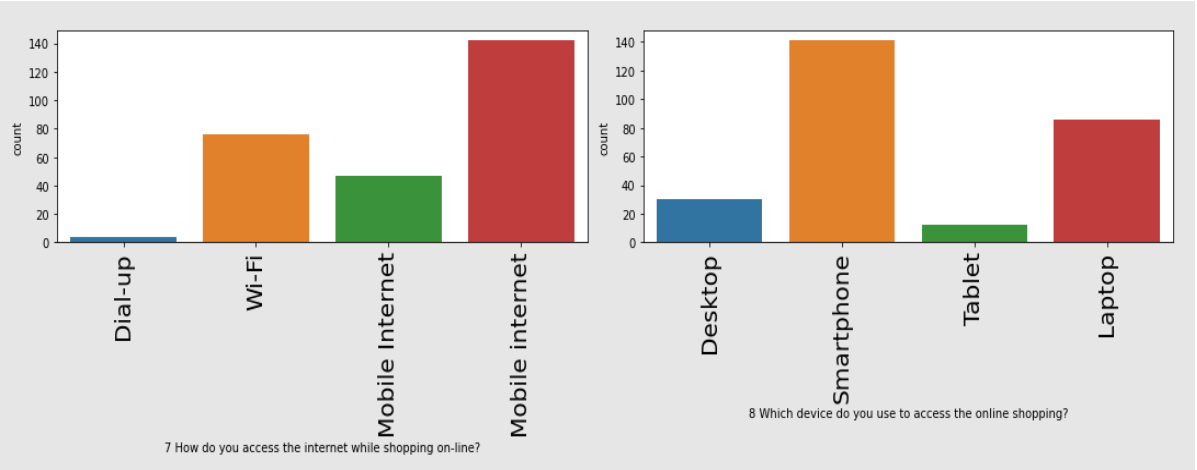
2. From count plot of age, we see that 21-30 years and 31-40 years of people use to do more shopping and people who are 51 years and above use do very less shopping



Observation:

1. From Since How Long You are Shopping Online plot, we can say since the age of 4+ people has started to shop online.

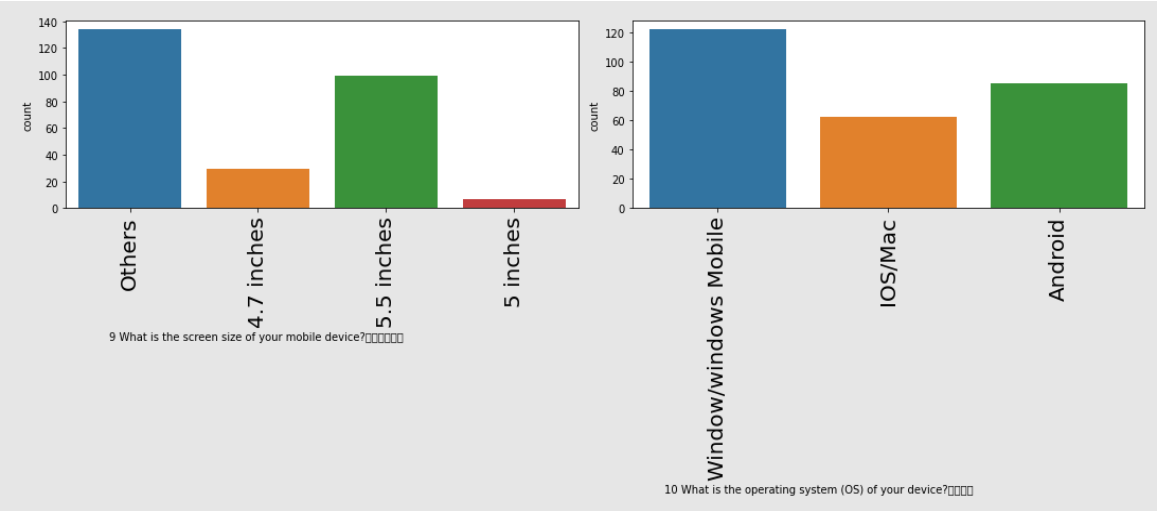
2. From How many times you have made an online purchase in the past 1 year? plot, we can say less than 10 times has high count, and 42 times and above have least count. It means that most people do online purchase less than 10 times a year.



Observation:

1. From How do you access the internet while shopping online? plot, we can see people use to shop using mobile internet more and very few use dial-up.

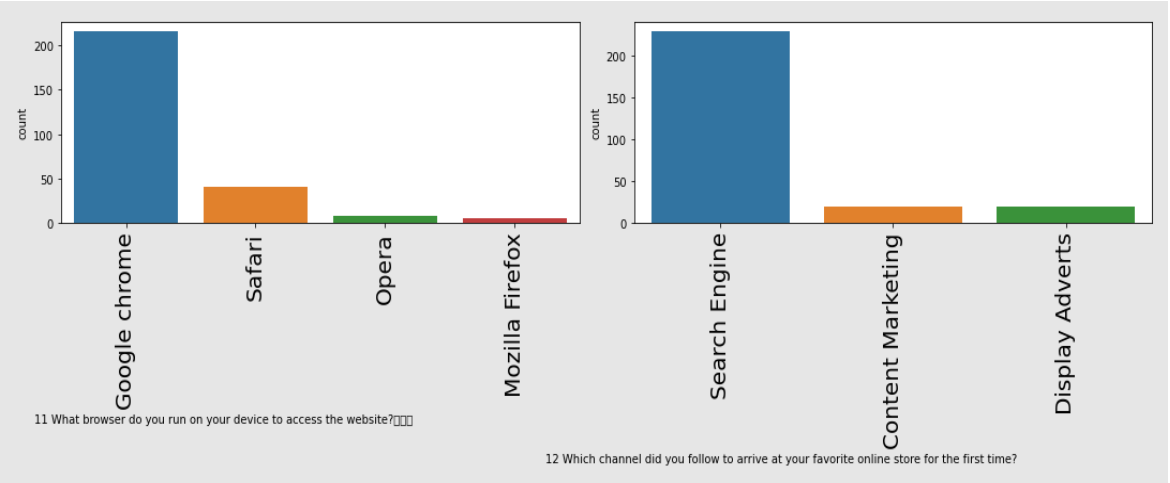
2. From Which device do you use to access online shopping? plot, we can say people use their smartphones more than laptops, and very few uses tablet.



Observation:

1. From What is the screen size of your mobile device? plot, we can say others have the highest count, followed by 5.5 inches. 5 inches has the least count.

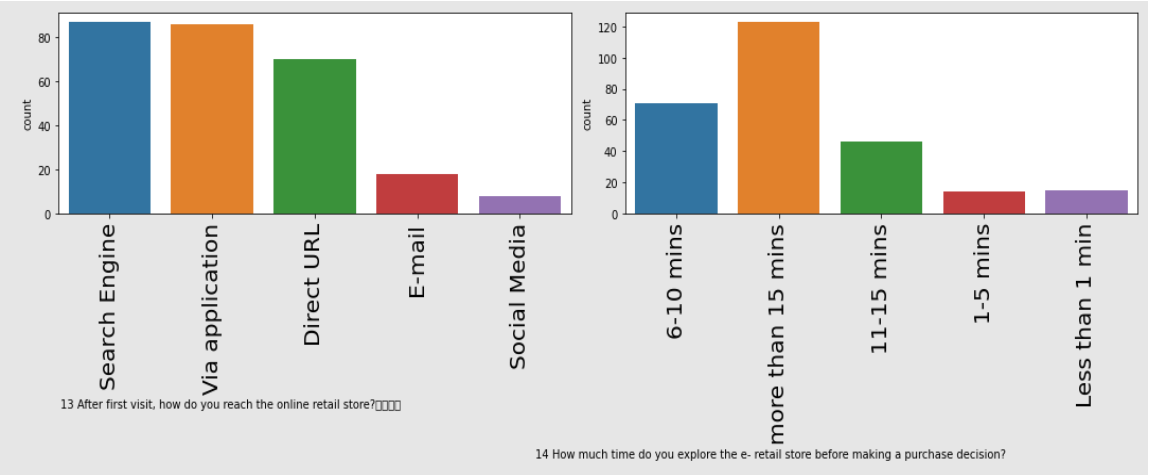
2. From What is the operating system (OS) of your device? plot, we can say most people use Windows phones then android and lastly few people use IOS/Mac for online shopping.



Observation:

1. From What browser do you run on your device to access the website? plot, we can say the majority of people use Google chrome and very few people use firefox.

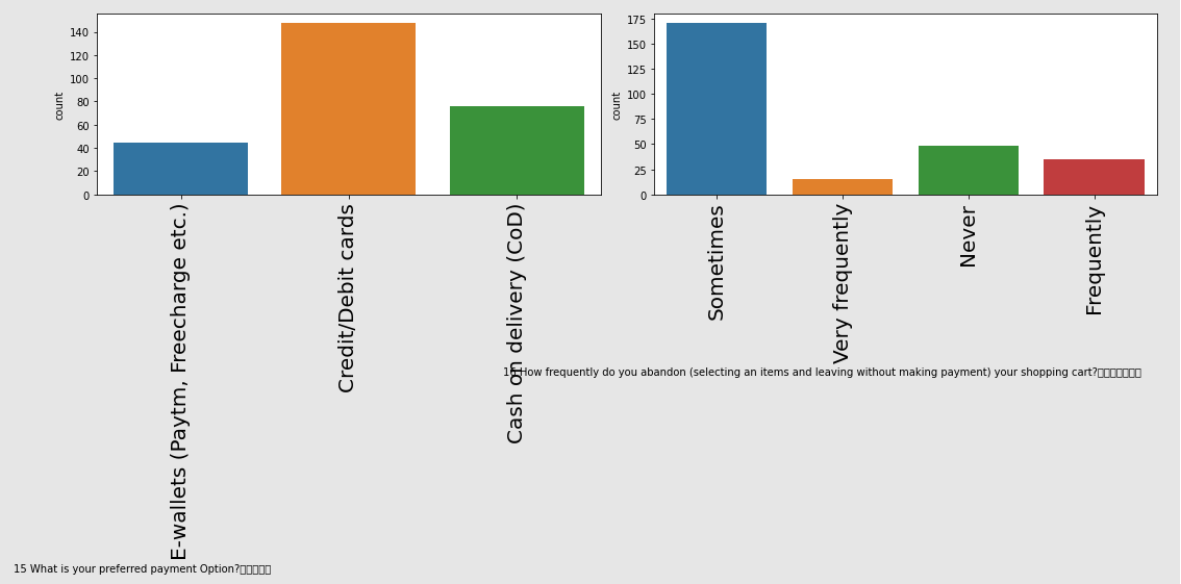
2. From Which channel did you follow to arrive at your favorite online store for the first time? plot, we see that People mostly use the search engine to buy something.



Observation:

1. From After first visit, how do you reach the online retail store? plot, we see that people used a search engine or app for this.

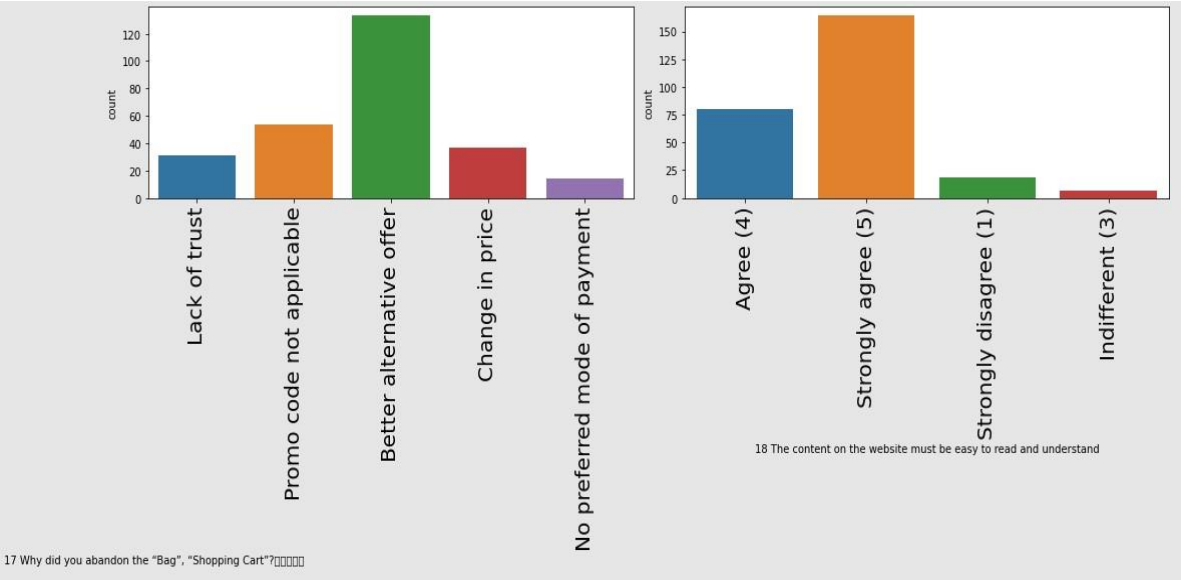
2. From How much time do you explore the e- retail store before making a purchase decision? plot, we can say most people take 15 minutes and a few people take 1-5 minutes only.



Observation:

1. From What is your preferred payment Option? plot, we can say most people use credit/debit cards. The 2nd preference of people is cash on delivery and least people use E-wallet.

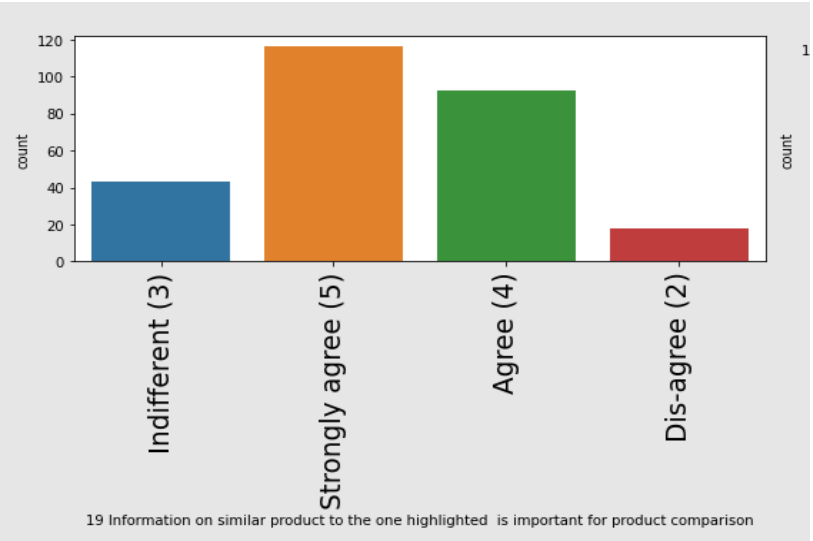
2. From How frequently do you abandon (selecting an items and leaving without making payment) your shopping cart? plot, we can say sometimes has the highest count and frequently has the least count.



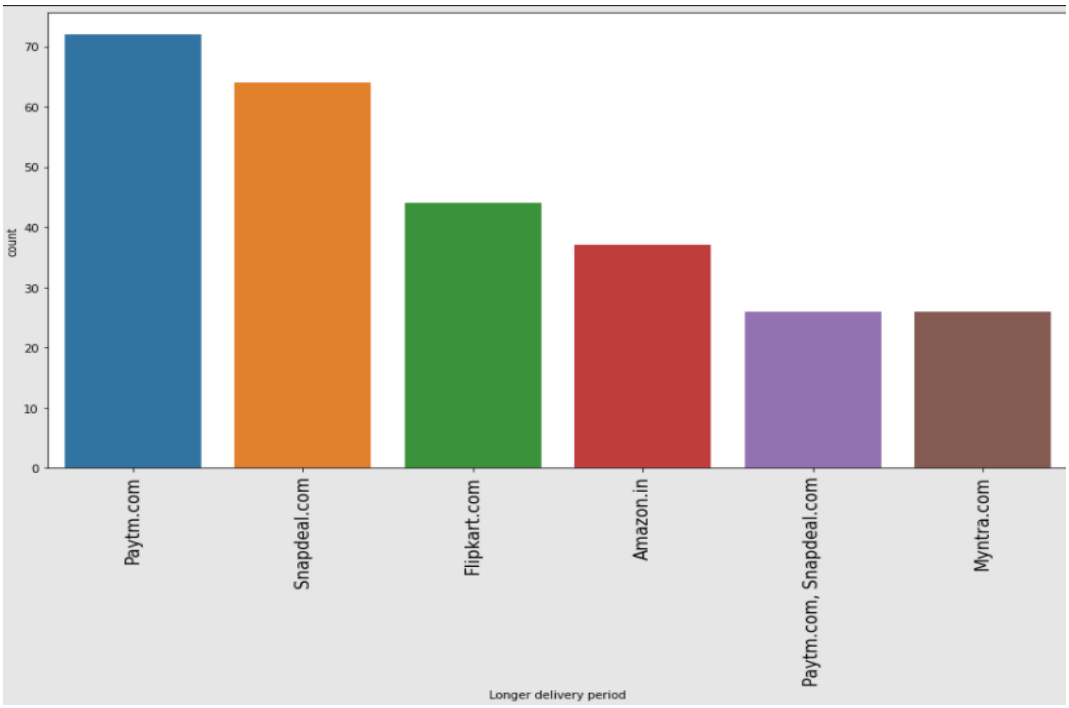
Observation:

1. From Why did you abandon the "Bag", "Shopping Cart plot, we can say a better alternative offer has the high count which means people get some other best offers while lack of trust and no preferred payment has the lowest counts.

2. From The content on the website must be easy to read and understand? plot, we see that people have voted more for strongly agree so that they can understand the content of the site.

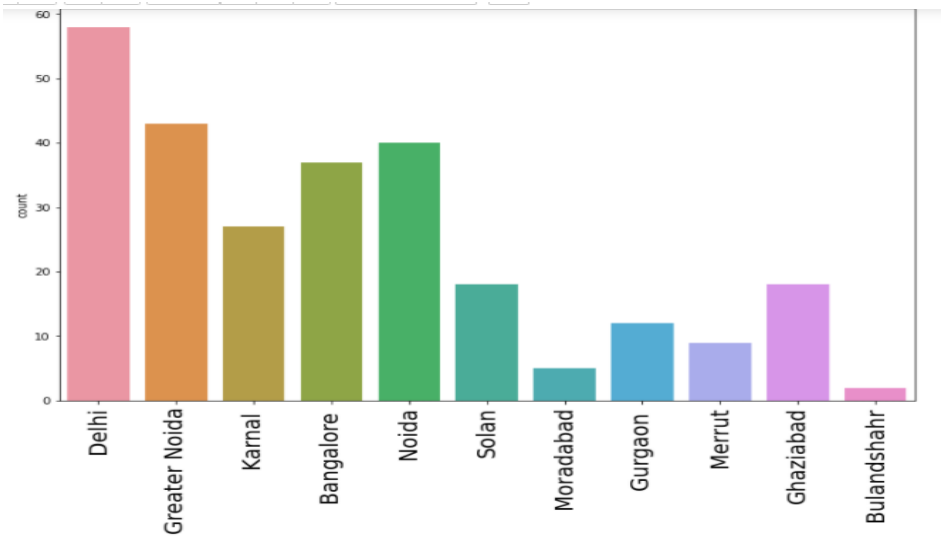


Observation: From Information on similar product to the one highlighted is important for product comparison plot, we can say people have voted more for strongly agree and agree.



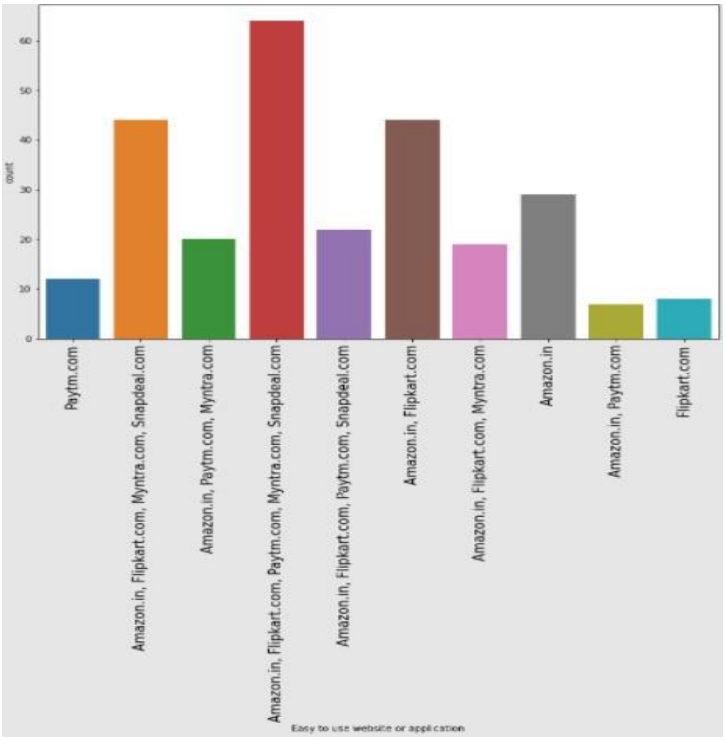
Observation:

From the above plot, we see that paytm.com takes the longest time to deliver items followed by snapdeal.com, flipkart.com then Amazon.com. Myntra.com takes the least time to deliver products.



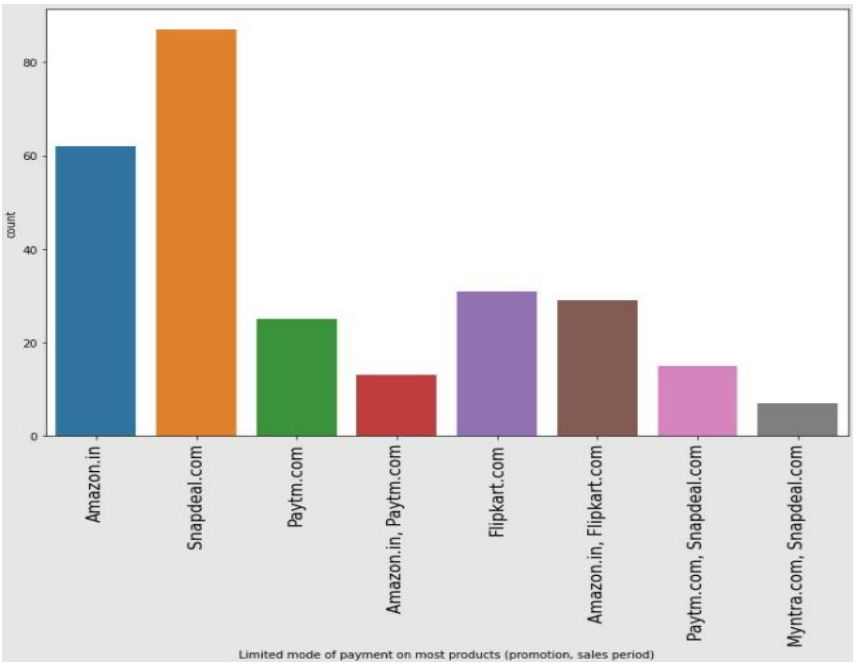
Observation:

From the above plot, we can say Delhi has the highest count which means that people in Delhi order more. After Delhi, Greater Noida has the 2nd highest count and Bulandshahr has the lowest count which means people who live in Bulandshahr usually buy fewer online items.



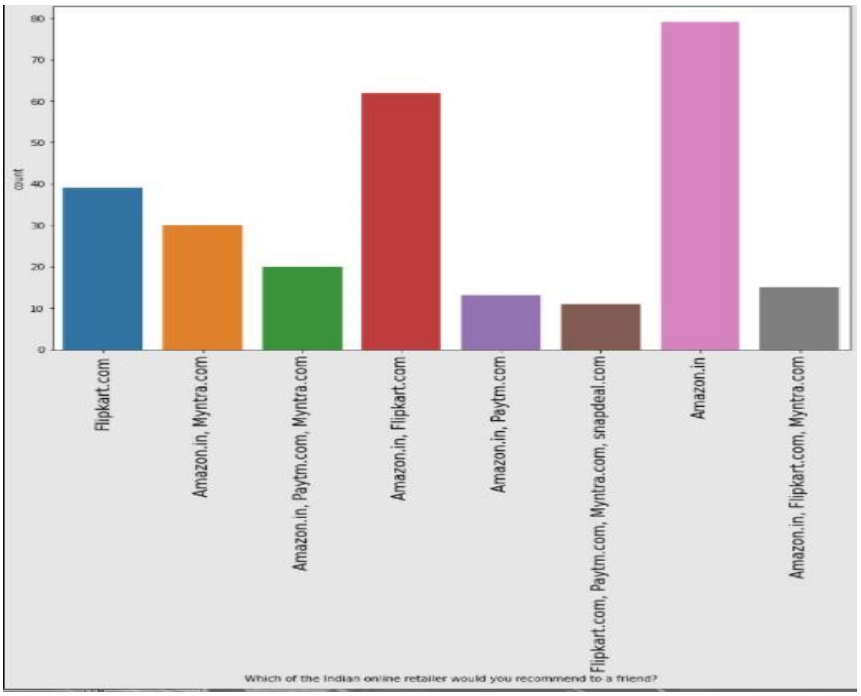
Observation:

From the above plot, we can say that values that have a single variable among these, only Amazon.com has the higher count as compared to others and paytm.com has the least count.



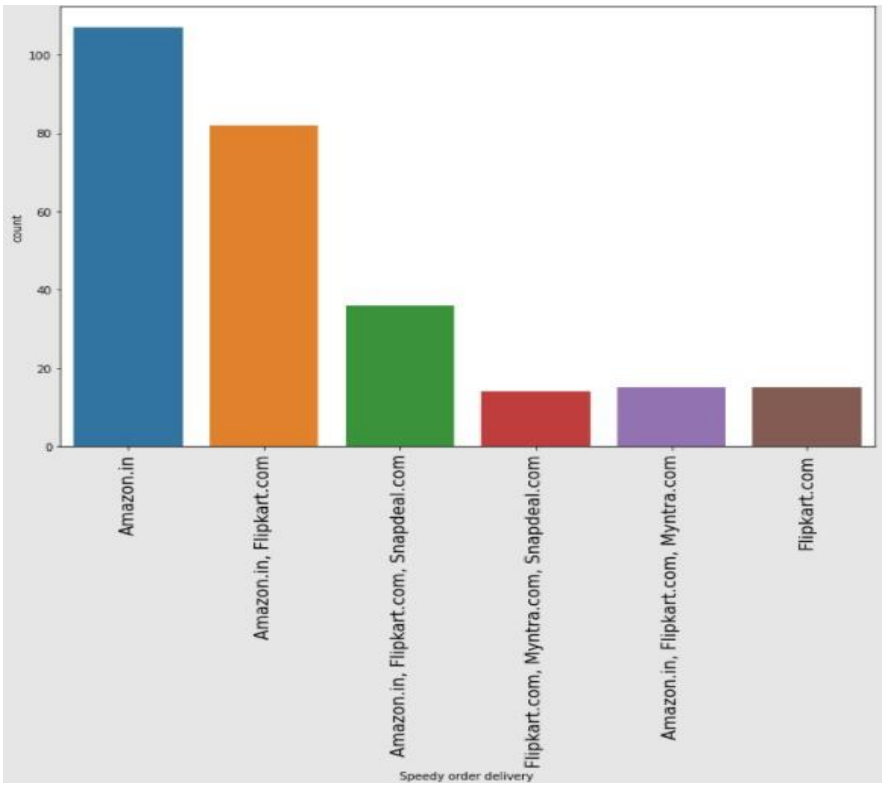
Observation:

From the plot above, we can say Snapdeal has the highest limited mode of payment, and Paytm has the least.



Observation:

From the above plot, we can say that most of the people recommended amazon to their f riends. Next, people also recommended Flipkart.

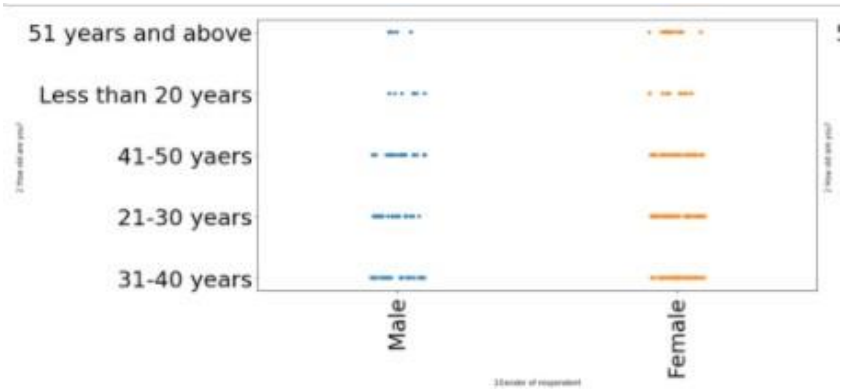


Observation: From the above plot, we can say Amazon.com provides the fastest delivery .

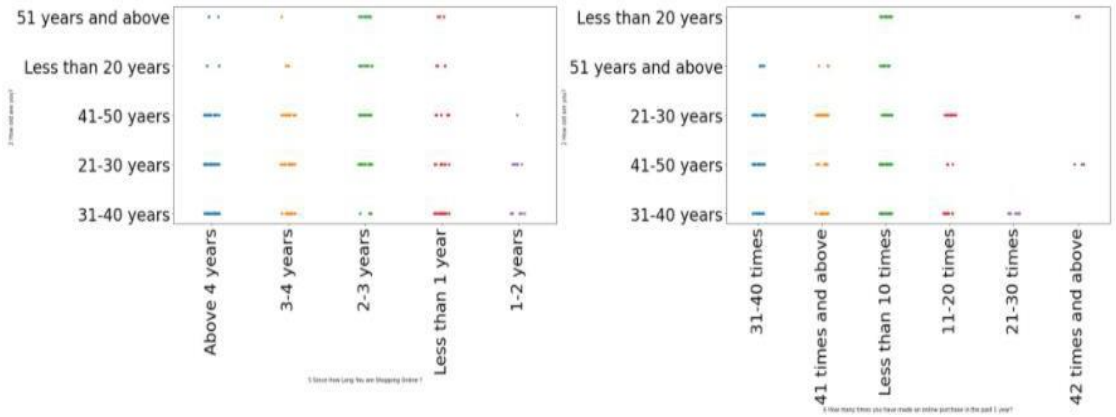
2. Bivariate Analysis: Bivariate analysis is one of the simplest forms of quantitative ana lysis. It involves the analysis of two variables to determine the empirical relationship betwee n them. We have analyzed the data and its relationship with features using a strip plot as sho wn below.

Strip plot helped to find out how different ages of people order their product, which age of people uses which site more, etc. I have done EDA on age vs all the features to understand how can sites improve their service for different ages people. After analysis, I understood that most 21-40 age people order more and spend their time on different sites to check products.

Some of the plots showing the relationship of Age with all the other features and the observation made from them are:



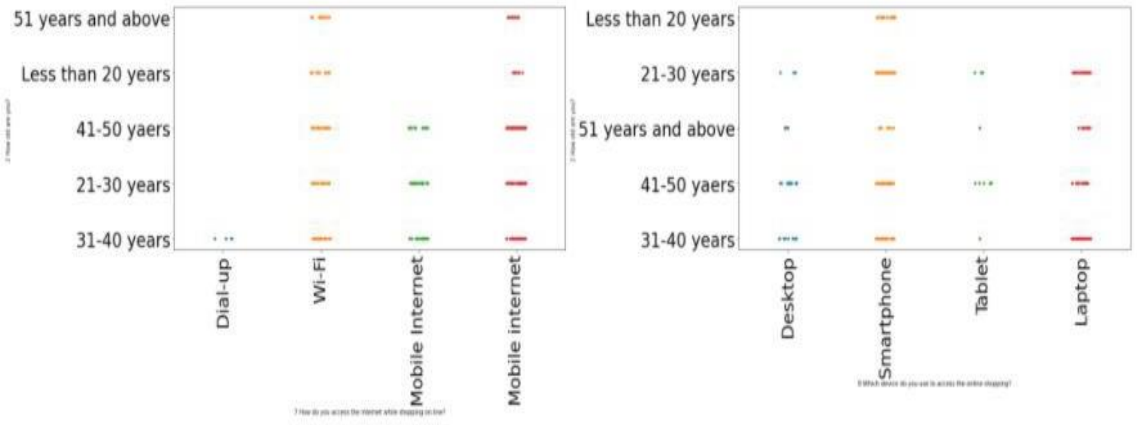
Observation: From the age vs gender plot, we can say for every age from young to old, females are more as compared to males



Observation:

1. From age vs Since How Long You are Shopping Online? plot, we can say the people who are doing shopping for above 4 years are mostly in age between 31-50 years, and the people who are in 3-4 years values are 21-30 years of age; with 2-3 years are mostly belonging to the age group 41-50 years and the people who are in less than 1 year are 31-40 years of age

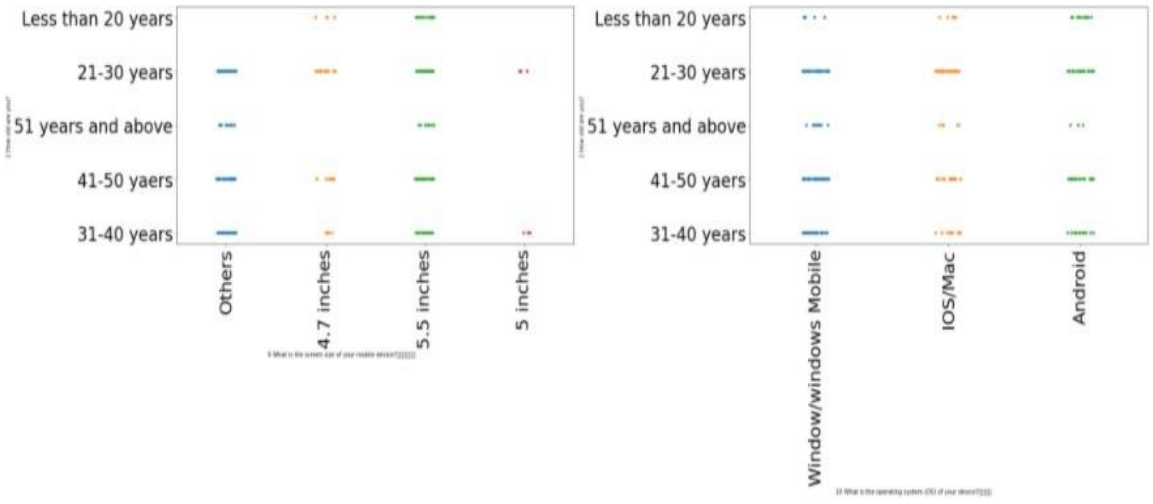
2. From age vs How many times you have made an online purchase in the past 1 year? plot, we can say most people who have ordered 31-40 times in a year falls into 41-50 age group; with 41 times and above have the age of 21-30 years; less than 10 times are in 21-50 age group, and 11-12 times are in the age between 21-30 years, so we can summarise that mostly 21-30 years of people are shopping more.



Observation:

1. From age vs How do you access the internet while shopping online? plot, we can say that some people use dial-up having age 31-40 years; and from wifi use, we can say all of the people use wifi but 21-40 age of people use it more. From mobile internet, we can say mostly 21-50 age of people use it more as compared to others.

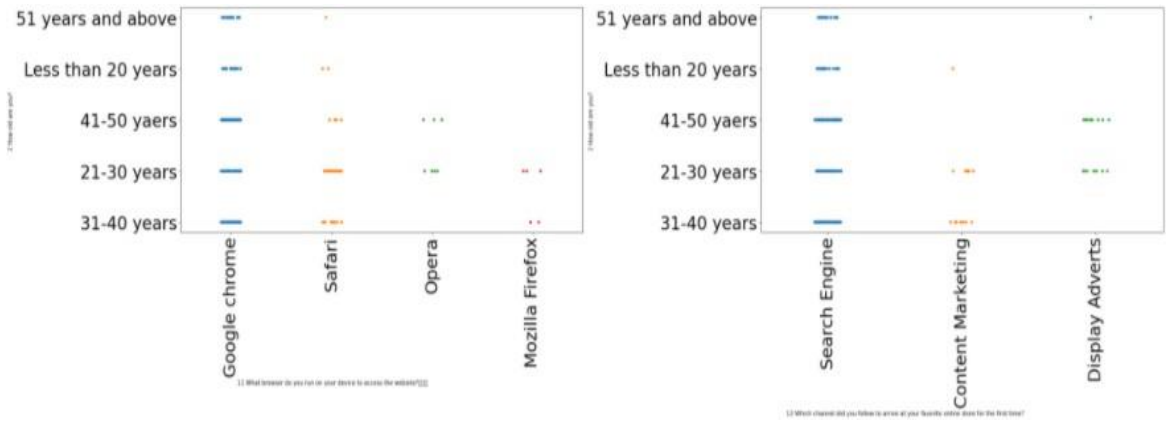
2. From age vs Which device do you use to access online shopping? plot, we can say people having age between 41-50 mostly use desktop and people whose age is between 21-40 years use mobile; some people use tablet having age of 41-50 years; some people use laptop have their age between 21-50 years, so we can say mostly above 41 years people use less mobile to order online, and some young people use mobile-only falls into age between 21-40years.



Observation:

1. From age vs device size plot, we can say that there are people who have big device size or less size as compared to 5.5 inches and there are people who use 4.7 inches mobile who falls into the age group of 21-30 years and most people use 5.5-inch mobile whose age is 21-30 years and very fewer people use 5.5-inch mobile who have an age of above 51 years.

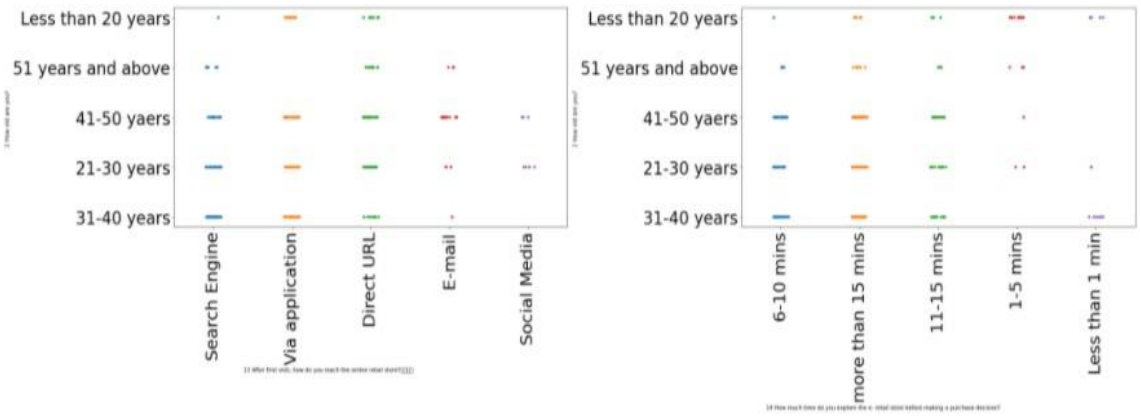
2. From Age vs device type plot, we can say people having age between 31-50 years use windows mobile, mostly 31-40 age of people use it more; people use IOS/Mac having age between 21-30 years indicating young people use it more often; and people use android more having age between 21-40 years, fewer people use android who have age between 41-50 and least number of people use android who have an age of 51+ years.



Observation:

1. From age vs What browser do you run on your device to access the website? plot, We can say people who have the age of 21-50 years everyone use Google chrome and fewer people use it who have age below 20 and 51+ years; people use safari more who have the age of 21-30 years, people with age 31-40 years also use it but have less count as compared to 21-30; there are only a few people who use opera having age of 21-30 years and very few people who use it have age of41-50 years; people use firefox who have age of 21-30 years more and very few uses it having age between 31-40 years.

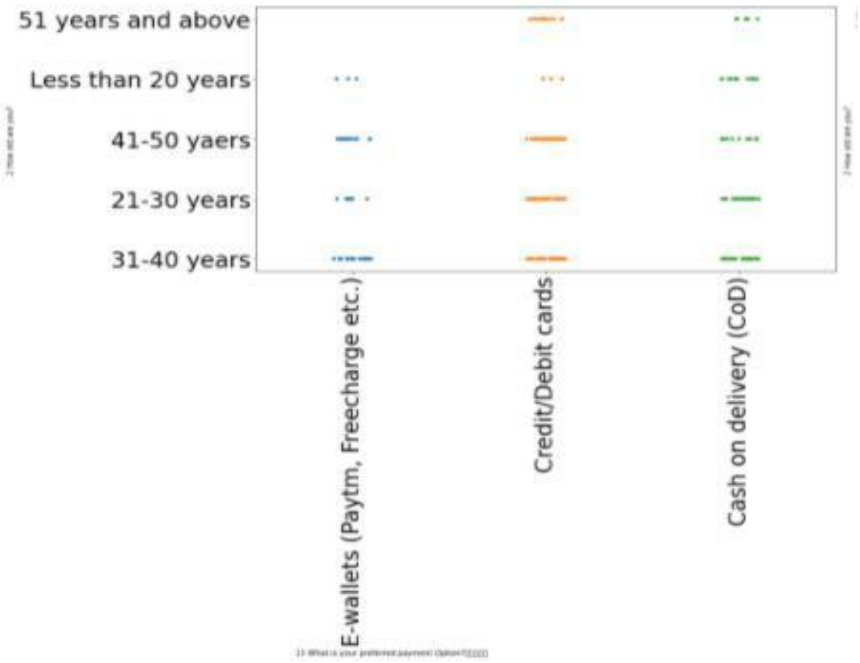
2. From age vs Which channel did you follow to arrive at your favorite online store for the first time? plot, we can say people of every age use search engine but fewer people use it whose age is below 20 or above 51 years as compared to people of age between 21-50 years; and a few people use content marketing as compared to others who have age between 21-30 years; and some people use display advertise who have the age of 41-50 years



Observation:

1. From age vs After first visit, how do you reach the online retail store? plot, we can say that people who have the age of 31-40 years reach the store by the search engine as compared to others; people who have the age of 21-50 years reach the store by application, and people who have the age of 31-40 years and below 21 have very less count as compared to others who reach the store by using direct URL, and people having ages between 41-50 and 51+ years mostly use email to reach the store.

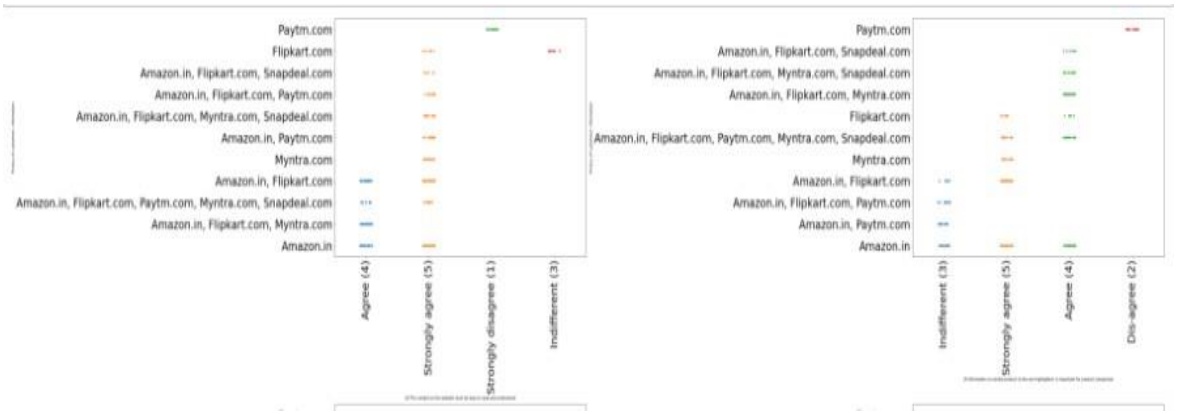
2. From age vs How much time do you explore the e- retail store before making a purchase decision? the plot, we can say people who are having age of 21-50 years have explored the store between 6-10 minutes; those having age between 21-51+ years have explored the store for more than 15 minutes; those having age below 20 years gave explored the store for 1-5 minutes; there are some people having age between 31-40 years have exploded it for less than a minute.



Observation:

From age vs What is your preferred payment Option? plot, we can say people whose age is between 31-50 years used e-wallet more as compared to others; people who have age 21-30 and 41-50 years used credit cards more, and people who have the age of 31-40 years also used credit card but have less count. From the cash on delivery option, we can say all the people use to take this option but people of 21-40 years of age have high count as compared to others.

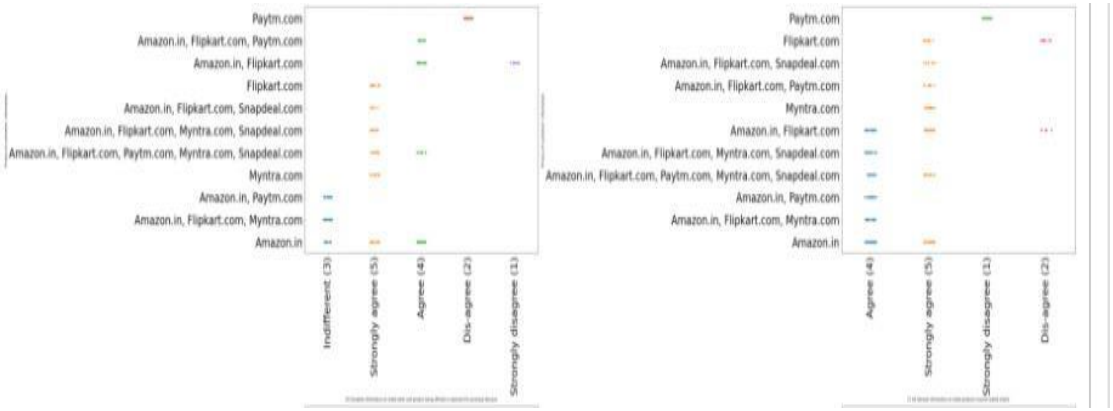
Some of the plots showing the relationship of All Site vs terms of different sites and the observations made from them are shown below:



Observation:

1. From Site vs The content on the website must be easy to read and understand plot, we can say every site strongly agree except for Paytm.

2. From Site vs Information on similar product to the one highlighted is important for product comparison, we see that Paytm disagrees while amazon, Flipkart, and Snapdeal strongly agree.

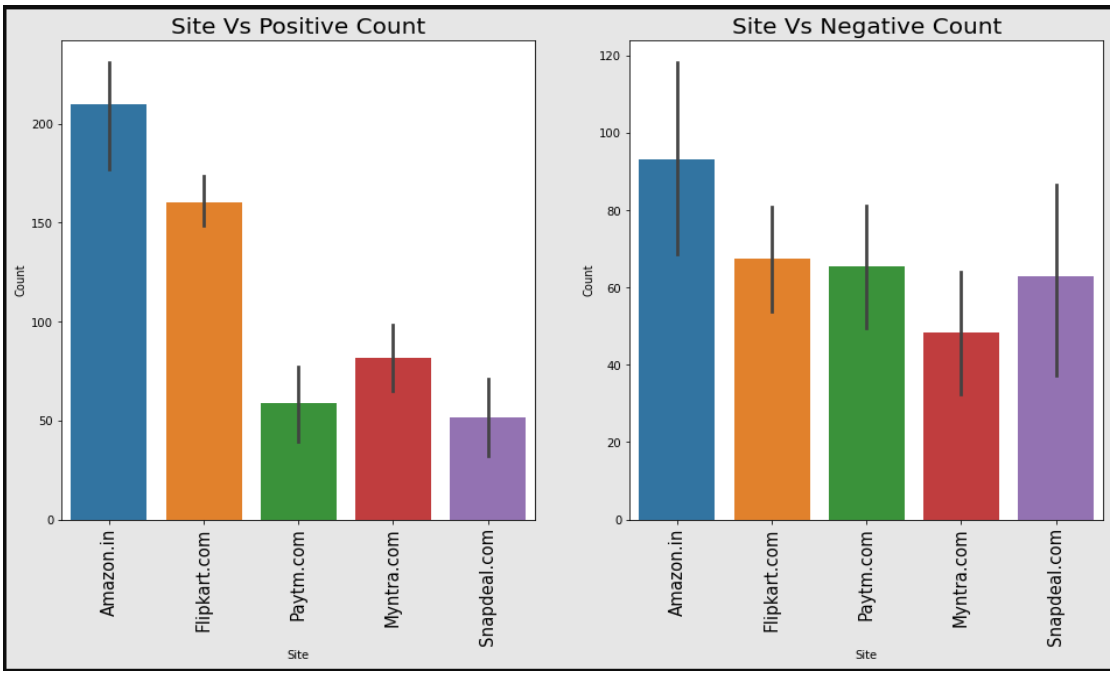


Observation:

1. From site vs Complete information on listed seller and product being offered is important for a purchase decision, Paytm disagree and other sites agree.

2. From site vs All relevant information on listed products must be stated clearly, Paytm disagree and others agree.

Next, I have defined a function to get the positive and negative review count of various online shopping sites and later used a count-plot to visualize the positive and negative review counts.



Observations:

From the above plot, We can see that Amazon.com has got the highest positive review as well as a highest negative review, but a positive review has a high count as compared to negative reviews, similar is the case with Flipkart.com and Myntra.com, while paytm.com and snapdeal.com have got lesser positive reviews and high negative reviews.

Interpretation of the Results / Conclusion:

Starting with univariate analysis, we found that the dataset is imbalanced by analyzing the count-plot of Feature 1: Gender of respondent which consists of a fewer number of Male and a higher number of Females. This needs to be handled during the train test split part of model training. Next, we moved on to Bivariate analysis and performed EDA on various features to have a better understanding of the data. Lastly, we checked what people are saying about their online shopping experience on various sites and the suggestions they are making for a smooth online shopping experience. The list of positive and negative reviews about various sites is shown below.

- **Amazon.com:**

Negative Reviews:

1. Takes a longer time to log in
2. Takes longer time in displaying graphics and photos
3. Late declaration of price
4. Takes Longer page loading time
5. More payment options required
6. Takes Longer delivery period

Positive Reviews:

1. Easy to use website or application
2. Visual appealing web-page layout
3. Wild variety of products on offer
4. Complete, relevant description information of a product
5. Fast loading website speed of the website
6. Reliability of the website or application
7. Quickness to complete purchase
8. Availability of several payment options
9. Speedy order delivery
10. Privacy of customers' information
11. Security of customer financial information
12. Perceived Trustworthiness
13. Presence of online assistance through multi-choice
14. Website is as efficient as before.

- **Flipkart.com:**

Positive reviews:

1. Easy to use website or application
2. Visual appealing web-page layout
3. Wild variety of products on offer

4. Complete, relevant description information of a product
5. Fast loading website speed of the website
6. Reliability of the website or application
7. Quickness to complete purchase
8. Availability of several payment options
9. Speedy order delivery
10. Privacy of customers' information
11. Security of customer financial information
12. Perceived Trustworthiness
13. Presence of online assistance through multi-choice
14. Website is as efficient as before
15. Which of the Indian online retailer would you choose

Negative Reviews:

1. Longer time to get logged in
2. Longer time in displaying graphics and photos
3. Late declaration of price
4. Longer page loading time
5. Limited mode of payment on most products
6. Longer delivery period
7. Change in website/Application design
8. Frequent disruption when moving from one page to another

- **Paytm.com:**

Positive Reviews with count:

1. Easy to use website or application - 125
2. Visual appealing web-page layout - 67
3. Wild variety of products on offer - 20

4. Complete, relevant description information of product - 59
5. Fast loading website speed of website - 99
6. Reliability of the website or application - 96
7. Privacy of customers' information - 68
8. Security of customer financial information - 88
9. Perceived Trustworthiness - 24
10. Presence of online assistance through multi-choice - 25
11. Website is as efficient as before - 58
12. Which of the Indian online retailer would you choose - 44

Negative Reviews with count:

1. Longer time to get logged-in - 77
2. Longer time in displaying graphics and photos - 28
3. Late declaration of price - 72
4. Longer page loading time - 94
5. Limited mode of payment on most products - 53
6. Longer delivery period - 98
7. Change in website/Application design - 63
8. Frequent disruption when moving from one page to another - 39

- **Myntra.com**

Positive Reviews with count:

1. Easy to use website or application - 147
2. Visual appealing web-page layout - 115
3. Wild variety of products on offer - 64
4. Complete, relevant description information of product - 64
5. Fast loading website speed of website - 74
6. Reliability of the website or application - 64

7 Quickness to complete purchase - 79

8. Availability of several payment options - 132

9. Speedy order delivery - 29

10. Privacy of customers' information - 78

11. Security of customer financial information - 91

12. Perceived Trustworthiness - 88

13. Presence of online assistance through multi-choice - 111

14. Website is as efficient as before - 14

15. Which of the Indian online retailer would you choose - 76

Negative Reviews with count:

1. Longer time to get logged-in - 35

2. Longer time in displaying graphics and photos - 74

3. Late declaration of price - 75

4. Longer page loading time - 68

5. Limited mode of payment on most products - 7

6. Longer delivery period - 26

7. Change in website/Application design - 37

8. Frequent disruption when moving from one page - 66

- **Snapdeal.com**

Positive Reviews with count:

1. Easy to use website or application - 130

2. Visual appealing web-page layout - 61

3. Wild variety of products on offer - 14

4. Complete, relevant description information of product - 59

5. Fast loading website speed of website - 81

6. Reliability of the website or application - 45

7. Availability of several payment options - 90

8. Speedy order delivery - 50

9. Privacy of customers' information - 45

10. Security of customer financial information - 100

11. Website is as efficient as before - 25

Negative reviews with count:

1. Longer time to get logged-in - 67

2. Longer time in displaying graphics and photos - 92

3. Late declaration of price - 0

4. Longer page loading time - 63

5. Limited mode of payment on most products - 109

6. Longer delivery period - 90

7. Change in website/Application design - 8

8. Frequent disruption when moving from one page to another – 74

Mentioned count for review means if more people are facing similar count then the site should improve it, but if it has a less count then no need to do anything..