

Project Report

Customer

Retention Analysis

Submitted by:

Vishal Kumar

What is Customer Retention?

The customer retention is the process of engaging existing customers to continue buying products or services from their business. The goal of customer retention is retaining as many as customer as possible in the company. One of the best ways to understand customer retention is to explore how to calculate customer retention rate for a given period.

How to Measure Customer Retention?

The simplest way to measure a customer retention rate is to look at the number of new customers acquired, number of previous customers, and total number of customers at the end of any chosen period. Retention rate =
$$\frac{((\text{No of Total Customers} - \text{No of New Customers}) / \text{No of Previous Customers})}{100}$$
 The calculation provides a percentage that

indicates how many customers have been retained over the period

Why is Customer Retention Important?

- Customer retention is an express route to a business' financial success. It allows the companies to build long term, meaningful relationships with customers.
- It empowers customers to share feedback with the company team.
- It helps you understand how loyal and satisfied your customers are, how strong your customer service is, and your products are really worth their money or not
- .➤ It enhances the brand reputation and understands future needs of the customers

Problem Undertaken:

The main objective behind this problem is to help the ecommerce websites to find the e-

retail factors for customers activation and retention. The online shopping was considered by almost everyone to be convenient. The store is always open and it is easy, less stress and time saving to shop from home. Participants also seemed to agree that internet shopping gives access to a large range of stores and products from the entire world and offers best price to their consumers. So, in this project we will be analyzing how these online stores attracts the customers and whether the customers satisfied by their products or not.

ANALYTICAL PROBLEM FRAMING

The main thing that I found in problem statement is, the data given for us is unsupervised data. The problem statement contains both utilitarian value and hedonic value. I have performed both univariate and bivariate analysis to analyses these values using

different plots like pie plot, count plot, distribution plot, factor plot etc. These plots give better pattern for analyzing the data. In this project I have done various mathematical and statistical analysis such as describing the statistical summary of the columns in which I found that the count is same for all the columns which means no null values present. Since the dataset contains object data type, I used label encoding method to convert the object data into numerical data. Checked for correlation between the features and visualized it using heat map

Data Sources and their formats:

- The data is collected from the Indian online shoppers. Results indicate the e-retail success factors, which are very much critical for customer satisfaction. The dataset is

provided by Flip Robo which is in the format xlsx

- There are two excel sheets one is detailed datasheet and other one is encoded datasheet. I have used detailed datasheet for the processing.
- The dataset contains 269 rows and 71 columns which is comprised of categorical columns. All the features contain object data type except the feature Pin code (integer data type)
- While describing the data I found skewness and outliers present in the columns. Since all the columns are categorical so I haven't removed outliers and skewness.

Data Pre-processing Done:

- Firstly, I have imported the necessary libraries and dataset

```
In [2]: #Importing necessary libraries
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
import os
import scipy as stats
from sklearn.preprocessing import LabelEncoder
%matplotlib inline
warnings.filterwarnings('ignore')
```

```
In [3]: # Importing Dataset
# Reading the excel file
df = pd.read_excel("customer_retention_dataset.xlsx")
pd.set_option("display.max_columns",None)
df
```

Out[3]:

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	5 Since How Long You are Shopping Online ?	6 How many times you have made an online purchase in the past 1	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	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	1
------------------------	--------------------	---------------------------------------	---	--	---	--	--	--	--	--	---	---	--	---

import numpy as np

It is defined as a Python package used for performing the various numerical computations and processing of the multidimensional and single dimensional array elements. The calculations using Numpy arrays are faster than the normal Python array.

import pandas as pd:

Pandas is a Python library that is used for faster data analysis, data cleaning and data pre-processing. The data-frame term is coming from Pandas only.

import matplotlib.pyplot as plt and import seaborn as sns:

Matplotlib and Seaborn acts as the backbone of data visualization through Python

From sklearn.preprocessing import LabelEncoder:

There are several encoding techniques like Label Encoder, One Hot Encoder, Ordinal Encoder. In this project I have used Label Encoder technique to convert categorical data or object type data into numerical data

- Checked the dimension of the dataset, data types of the features. Checked the information of the data frame using info() and checked the number of unique values present in the dataset
- Since the columns had no proper names, so I renamed them with new name as below.

```
In [5]: # Checking the dimension of dataset
df.shape

Out[5]: (269, 71)

In [7]: columns = ['Gender', 'Age', 'Shopping_City', 'Pincode', 'Shopping_Since', 'Shopping_Frequency', 'Internet_Accessibility', 'Device_Used',
'Screen_Size', 'OS', 'Browser_Used', 'Channel_First_Used', 'Login_Mode', 'Time_Explored', 'Payment_Mode', 'Abandon_Frequ',
'Abandon_Reason', 'Content_Readability', 'Similar_Product_Info', 'Seller_Product_Info', 'Product_Info_Clarity', 'Navigat',
>Loading_Processing_Speed', 'User_Friendly_Interface', 'Convenient_Payment_Mode', 'Timely_Fulfillment_Trust', 'Customer',
'Customer_Privacy_Guarantee', 'Various_Channel_Responses', 'Benefits', 'Enjoy', 'Convenience', 'Return_Replacement_Polic',
'Info_Satisfaction', 'Site_Quality_Satisfaction', 'Net_Benefit_Satisfaction', 'Trust', 'Product_Sevelar_Category', 'Rela',
'Patronizing_Convenience', 'Adventure_Sense', 'Social_Status', 'Gratification', 'Role_Fulfillment', 'Money_Worthy', 'Shop',
'Visually_Appealing_WebApp', 'Product_Variety', 'Complete_Product_Info', 'Fast_WebApp', 'Reliable_WebApp', 'Quick_Purch',
'Fast_Delivery', 'Customer_Privacy_Info', 'Financial_Security_Info', 'Perceived_Trustworthiness', 'Multichannel_Assist',
'Late_Price_Declare', 'Long>Loading_Time', 'Limited_Payment_Mode', 'Late_Delivery', 'WebApp_Design_Change', 'Page_Disrup',
'Recommendation']

df.columns = columns

In [8]: # Checking the new column names after renaming them
df.columns

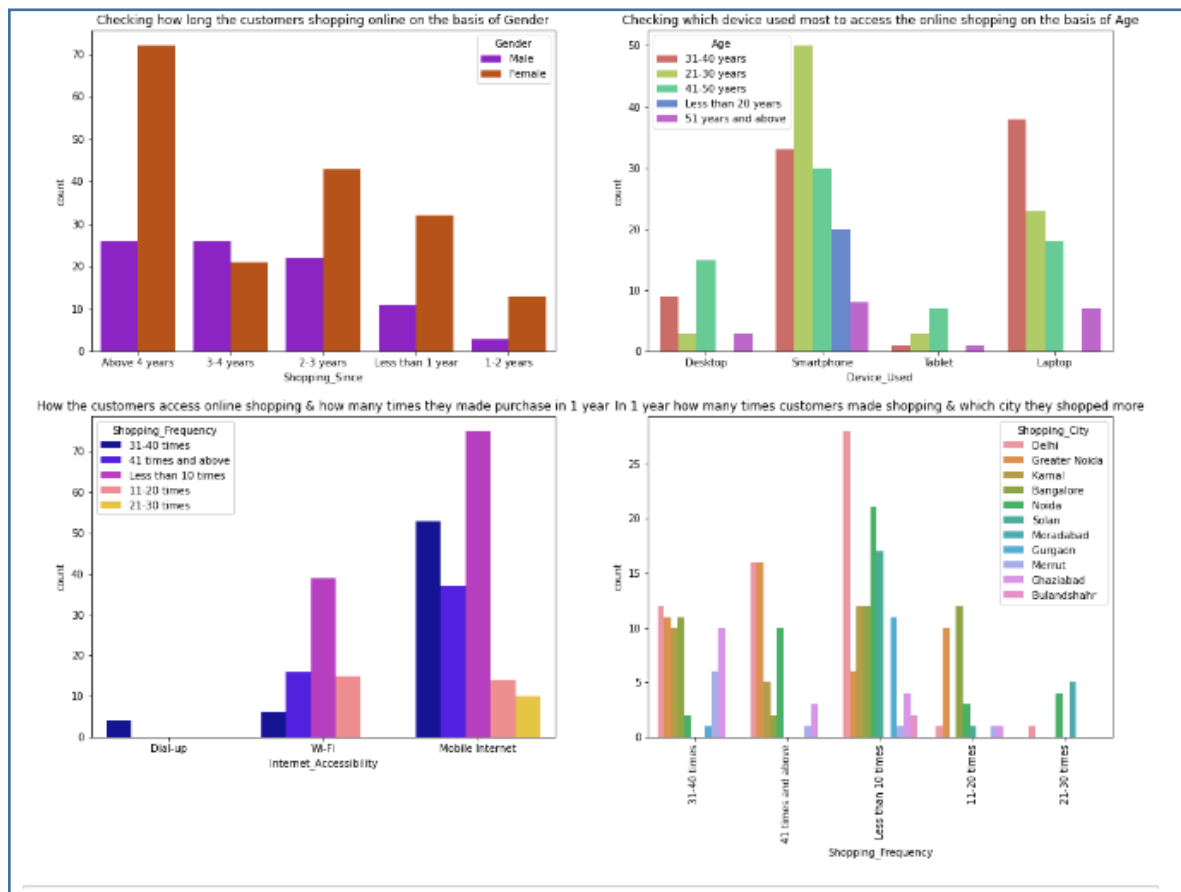
Out[8]: Index(['Gender', 'Age', 'Shopping_City', 'Pincode', 'Shopping_Since',
'Shopping_Frequency', 'Internet_Accessibility', 'Device_Used',
'Screen_Size', 'OS', 'Browser_Used', 'Channel_First_Used', 'Login_Mode',
'Time_Explored', 'Payment_Mode', 'Abandon_Frequency', 'Abandon_Reason',
'Content_Readability', 'Similar_Product_Info', 'Seller_Product_Info',
'Product_Info_Clarity', 'Navigation_Ease', 'Loading_Processing_Speed',
'User_Friendly_Interface', 'Convenient_Payment_Mode',
'Timely_Fulfillment_Trust', 'Customer_Support_Response',
'Customer_Privacy_Guarantee', 'Various_Channel_Responses', 'Benefits',
'Enjoy', 'Convenience', 'Return_Replacement_Policy',
'Loyalty_Programs_Access', 'Info_Satisfaction',
'Site_Quality_Satisfaction', 'Net_Benefit_Satisfaction', 'Trust',
```

- I have checked the null values and found no null values present in the dataset and visualized it using heat map.
- Checked the value count of each column. By checking value count function, I found some of the columns contains duplicate and unwanted entries, so I have replaced them using appropriate values.
- Visualized each feature using seaborn and matplotlib libraries by plotting count plot, pie plot, distribution plot, box plot and factor plot.
- Performed label encoding.

- Checked the skewness and outliers. Since all the columns are categorical, so I haven't removed skewness and outliers.
- Checked the correlation between the features and visualized it using heat map.
- Described the data using `describe()`.

Visualizations:

I have performed both univariate and bivariate analysis to visualize the data. In univariate analysis I have used pie plots, count plots and distribution plot and in bivariate analysis I have used count plot, factor plot and boxplots. Here I will be showing only bivariate analysis plots to analyses the data

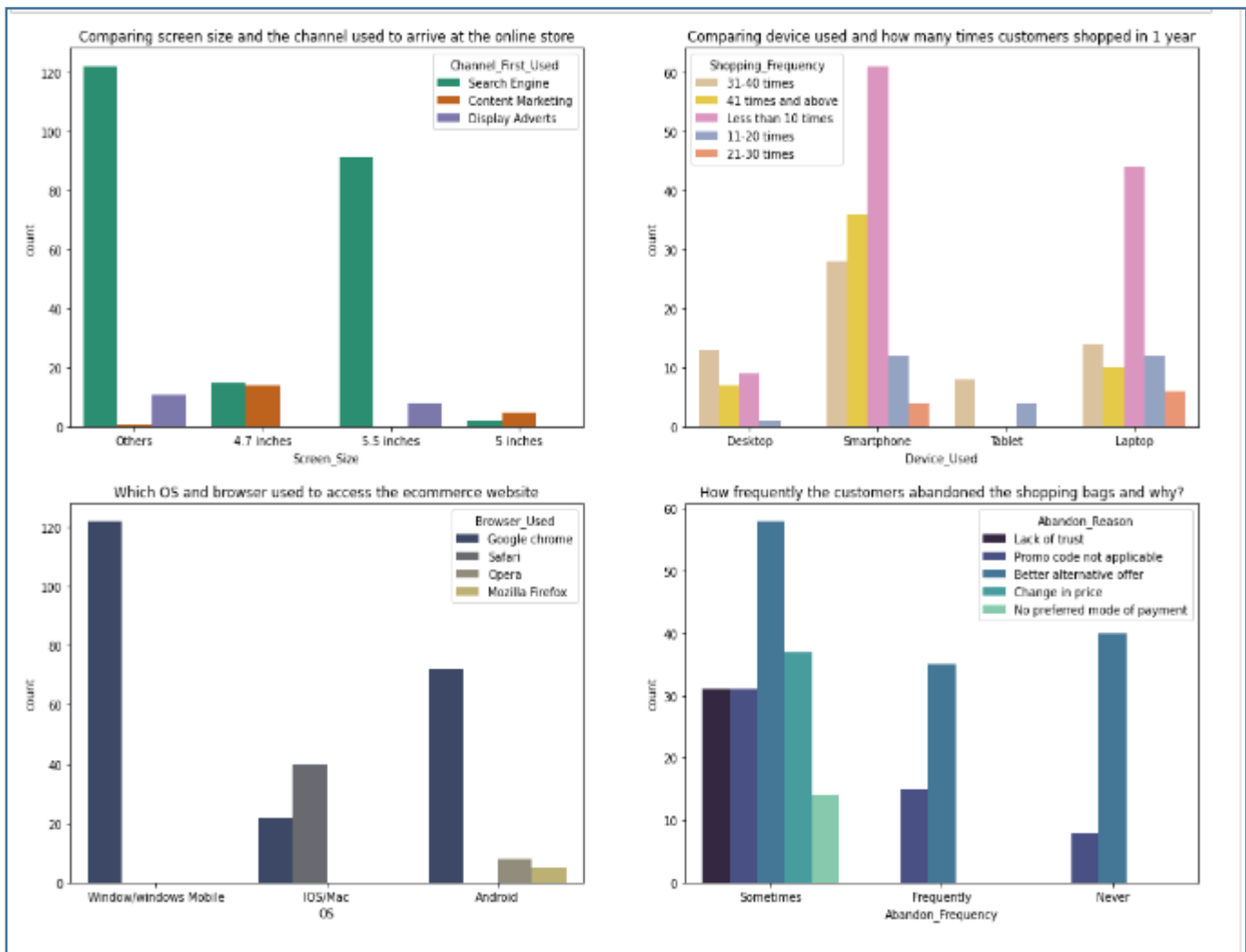


Observations from the above graphs:

Most of the female customers shopped online from more than 4 years and the count is also high for the females who shopped from 2-3 years. And only few male customers shop online more than 4 years. Which means the female customers are more enthusiastic to buy products from the

online shopping websites. Many customers whose age between 31-40 years and 21-30 years used Smartphones followed by Laptops to access the online shopping websites. Most of the customers access the shopping websites more than 31-40 times in 1 year through Mobile Internet to shop the products also most of the customers who used mobile internet to access the online shopping website made online purchase less than 10 times in a year. And only few of the customers used WiFi network to access the shopping store. Most of the customers used ecommerce websites less than 10 times in a year from the city Delhi to shop the products.

Comparison between two variables



Observations from the above plots:

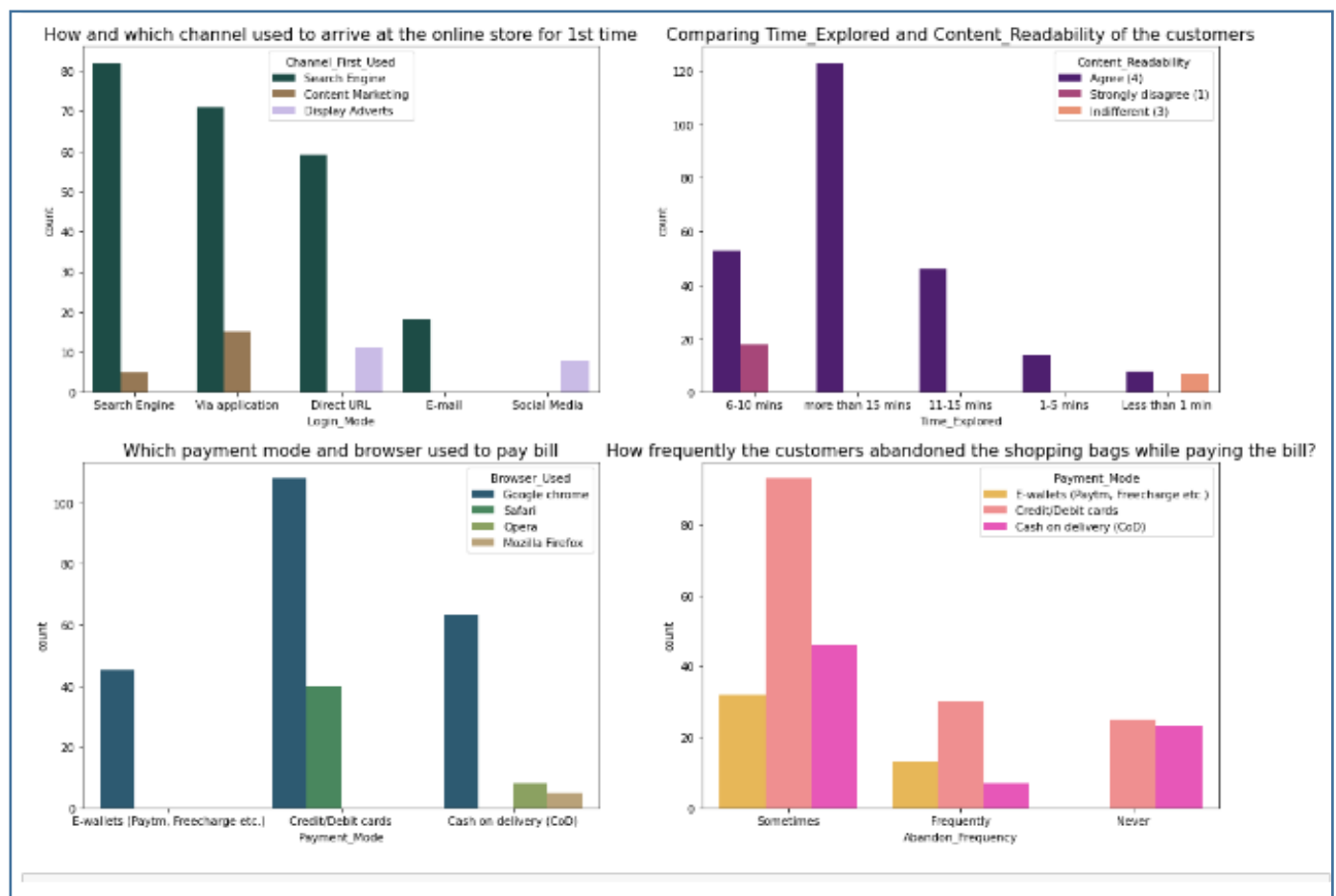
- The customers having their mobile screen size say 6 inches (others) have followed search engine channel to arrive

at their favorite online store for the first time. Also the customers who have their screen size 5.5 inches also used search engine channel to access the online shopping store.

- . Most of the customers used Smartphones 31-40 times in a year to access the ecommerce websites to shop the products.
- . Many customers having windows operating system in their device ran Google chrome to access the ecommerce shopping websites and some of the customers having IOS/Mac operating system used Google chrome as well as Safari to reach the online shopping store.
- . Due to Lack of trust on the ecommerce websites, sometimes most of the customers abandoned the websites and

some of the customers abandoned the shopping website due to the promo code not applicable. Which means, if the product is having the special price or some catalogue price rule is applicable on it. Then coupon code should not be applicable on the products.

So it is important for the ecommerce companies to create discount price, offers, coupon codes to retain the customers.

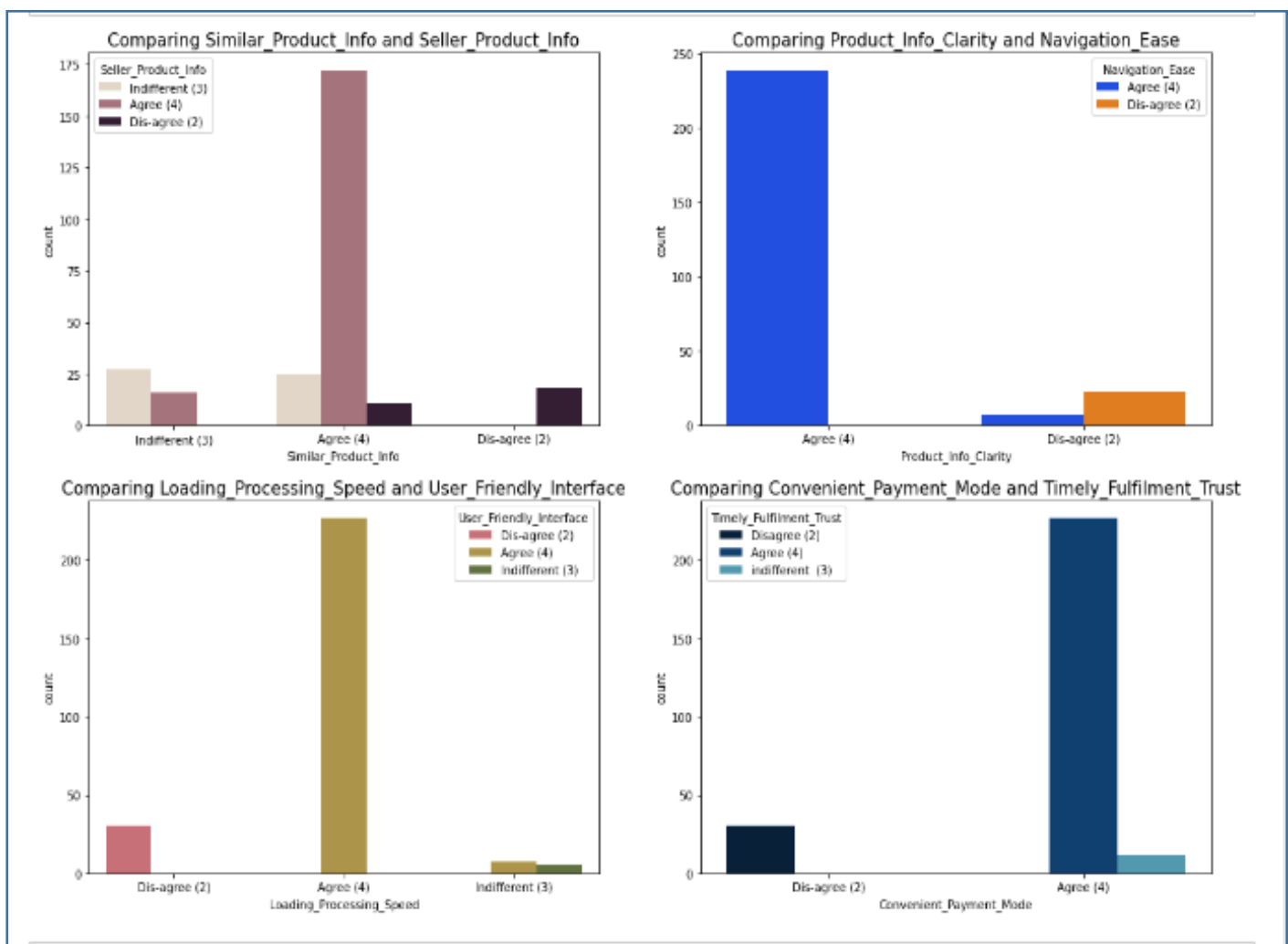


Observations:

- Search engine is the most used channel by the customers to arrive their favorite store for the first time and after visit the website for the first time, most of them used the same channel to reach the online retail store to re shopping the products.
- Most of the customers agreed that the content on the website is easy to read and understand also they explored more than 15 mins before making the purchase decision and some of the customers strongly disagreed that the content is not good and they explored 6-10 mins before making the purchase decision. So ecommerce websites should enable some images and it should contain clear structure, so that the

customers can easily read and understand the content of the product.

- Most of the customers used google chrome to reach the websites and they preferred to pay their product price using Credit/Debit cards and only few of the customers used Safari browser to reach the e-retail websites.



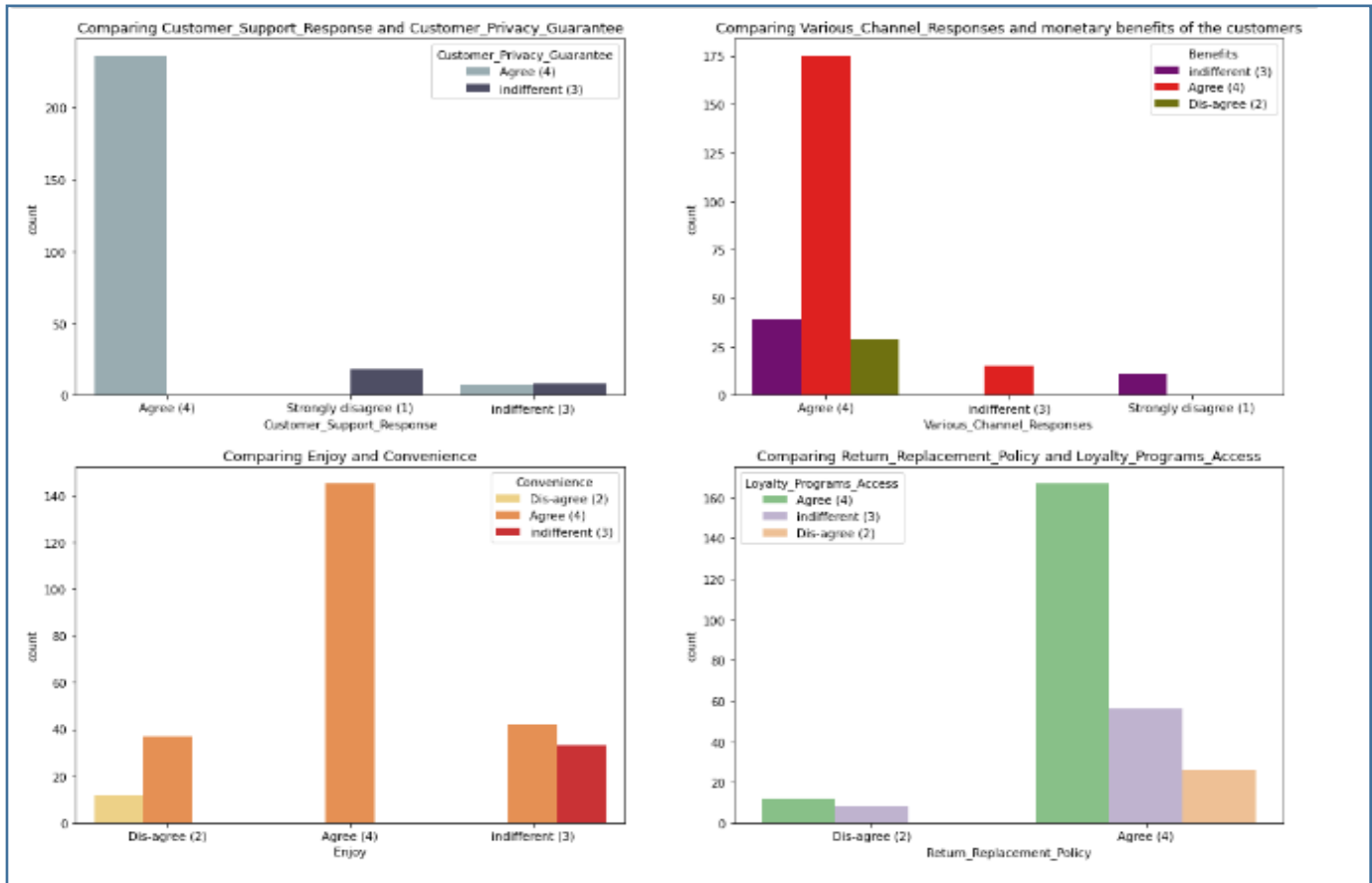
Observations:

- . Most of the customers agreed that the information on similar product to the one highlighted is important for product comparison and also Complete information on listed seller and product being offered is important for purchase decision. In order to buy a product, the ecommerce website must give the complete information about the product and seller information then only the customers can compare the product costs and its details in different websites and they tend to buy that particular product in a particular website.
- . Around 90% of the customers agreed that they should be able to navigate the website easily and the products information in the website must be

clearly stated their uses, lifetime, and benefits etc. Then only more customers tend to buy those products and can shop easily.

- . Most of the customers agreed with the user friendly interface of the websites which can be easily loaded and processed also these websites' loading and processing capacity is very fast so that the customers like to shop in ecommerce websites. If these websites do not have this much of loading and processing speed then customers don't want to buy the products in this website and they tend to other websites or other options rather than this.
- . Most of the customers agree to the trust that the online retail stores will fulfil its part of the transaction at the stipulated time also most of them very happy with

the convenient payment modes given by the websites. In other words, the websites must provide all the possible ways of payment methods then only the customers shop frequently all the time the mode of the payment for customers may not possible sometimes they may choose cash on delivery. So if the retailers provides all type of payment methods then the customers can easily make the payment also it enhances the sales of the ecommerce sites. And the transaction must also be given with some stipulated time otherwise the payments may be failed so they've to provide minimum amount of time which need to be fixed for all.



Observations:

- Almost all the customers agreed that ecommerce websites have empathy towards them and these sites being able to guarantee the privacy of the customers. That is the online retailers

must be able to resolve all the queries of the customers and they have to assure the customers keeping all their credential secured and should not share with others. If the websites give guarantee about the privacy, then the customers make shopping regularly which will enhance the company's sales.

- . Most of the customers agreed that the online shopping gives monetary benefits and responsiveness, availability of several communication channels will help them more while shopping online which means if one channel is not available then customers can easily reach out to other channel to fulfil their benefits. So, it is important for the online e-tailer companies to provide various channels to communicate with the customers. The ecommerce websites

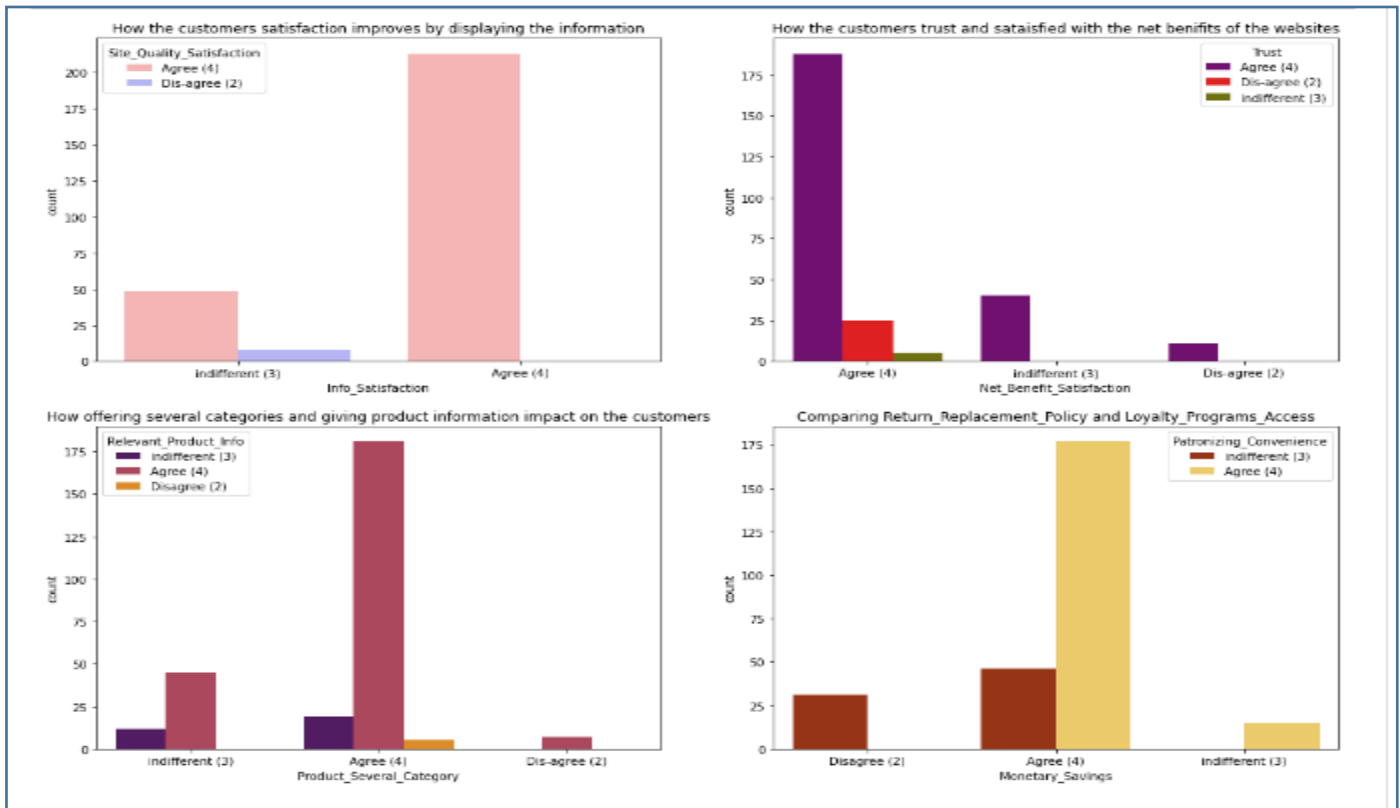
should ask the feedback regarding their services, ratings of the products, reviews etc and also they try to communicate with the customers in different social platform then only customers get satisfied by the e-tilers sites and make more shopping on the particular websites regularly which intends to increase the sales of the company. If one website gives less price and more discount for particular product then the customers tend to shop more in that particular website. So, the companies must try to give less price then customers like their offers and retention also increases.

- . Most of the customers believed that they enjoy online shopping also shopping online is convenient and flexible and some of the customers who

disagreed with the enjoyment of the shopping, they are not convenient with the online shopping. Some customer's shops online for their enjoyment purpose they are termed to be hedonistic, for them shopping online gives experiential satisfaction. They contribute much for the ecommerce companies by buying all the costly products randomly.

- . Most of the customers agreed that return and replacement policy of the e-tailer is important for purchase decision also gaining access to loyalty programs is a benefit of shopping online. Many return policies have conditional agreements, such as time limits, that must be clearly defined and expressed at the time of purchase or else the customers won't get the chance to

return their damaged or dissatisfied products due to this they may not access the same website if they want to shop again. It is evident from the fact that the customers actually not liking the products completely, they are just purchasing the products and returning them in case of any dissatisfaction. So it is important for the online shopping websites to make easy return and replacement policy if they want to retain their customers. Also, by gaining access to loyalty programs, the customers get more and more rewards, increasing their engagement rate and thus bringing more profit to both company and customer.



Observations:

- Many customers agreed that displaying quality information on the website improves satisfaction of customers since they believe that displaying quality information have significant association with customer satisfaction and they are satisfied and happy while shopping on good quality websites. In order to obtain

high levels of customer satisfaction, high service quality is needed, which often leads to favorable behavioral intentions also a website with good system quality, information quality, and electronic service quality is a key to success in e-commerce. So, the online e-tailers must display all the information about the product then only customers get an idea to buy the products regularly.

- . Most of the customers agreed that net Benefit derived from shopping online can lead to users' satisfaction also they believe that user satisfaction cannot exist without trust. The e-tailer should provide crediting points (net benefits) so that the customers tend to buy frequently in order to gain points. Trust is also a major factor for customers to decide whether to buy products from

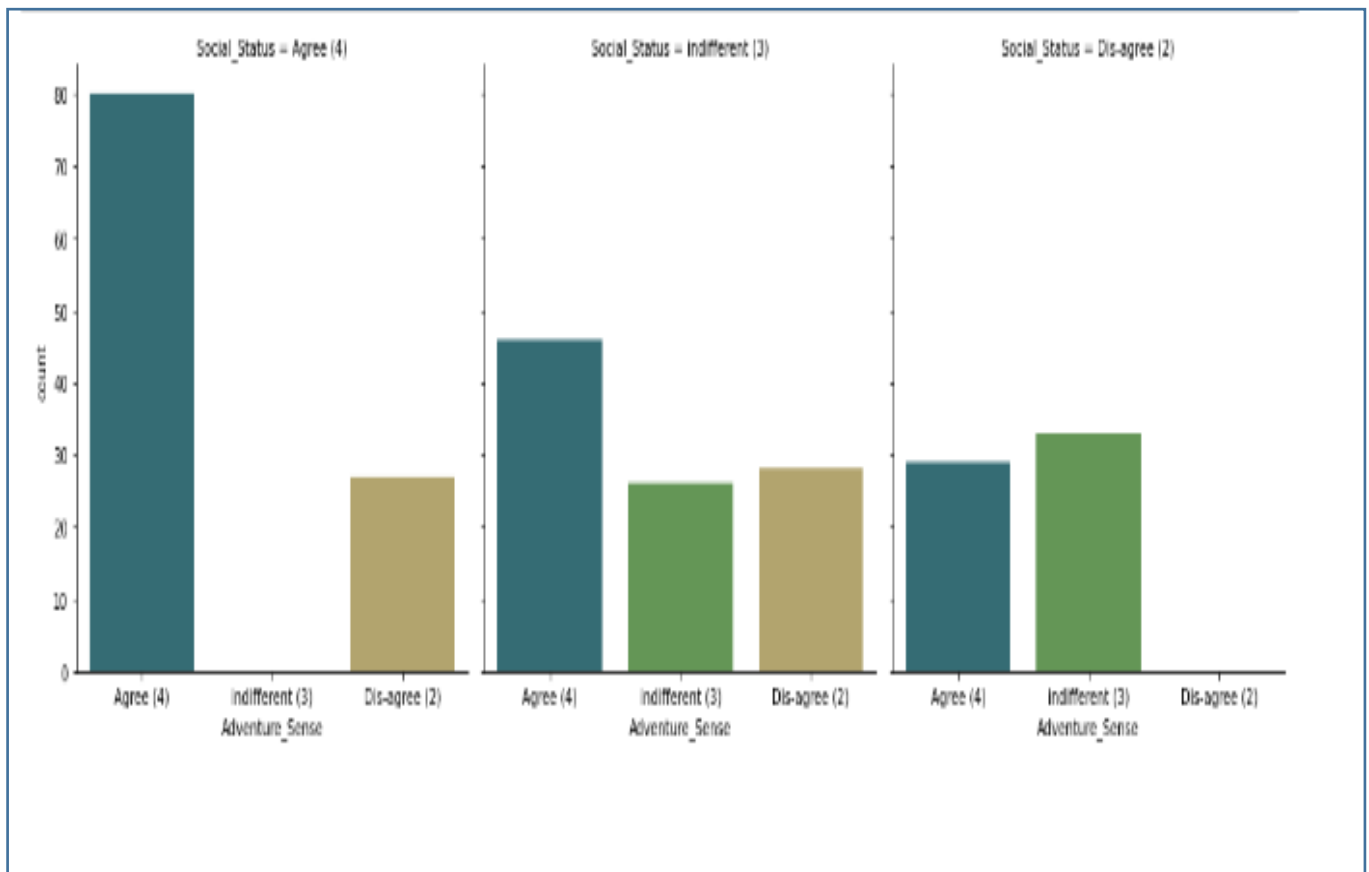
online stores or not also trust helps reduce uncertainty when the degree of familiarity between the customer and transaction security mechanism is insufficient. If customers have a high level of trust toward the website, it is more likely for them to have intention to purchase so it's important for the ecommerce website to make the customers get trust on them.

- The customers are more likely to purchase on the same websites if that website offers them a wide variety of products in several category and giving relevant information about the products. Having multiple product lines may allow to grow the ecommerce business and finding accurate and up-to-date information of the product must be stated clearly in the website so that the

customers can buy the products without any confusion.

- . In this digital and competitive world, everyone wants to save money, the ecommerce company need to know that the best way to sell online is to make the consumer feel that he is saving money doing so. And not just feel, online shopping should result in a lot of saving for the consumer. This saving would automatically get converted into trust and brand equity for the seller. To do this the online companies should offer the best deals and bargains to the consumer through social platforms. If the retailers give some discounted prices then the customers can make money savings and they tend to purchase in the same websites regularly. Convenience is

the important thing for ecommerce and most of the customers agreed with it.

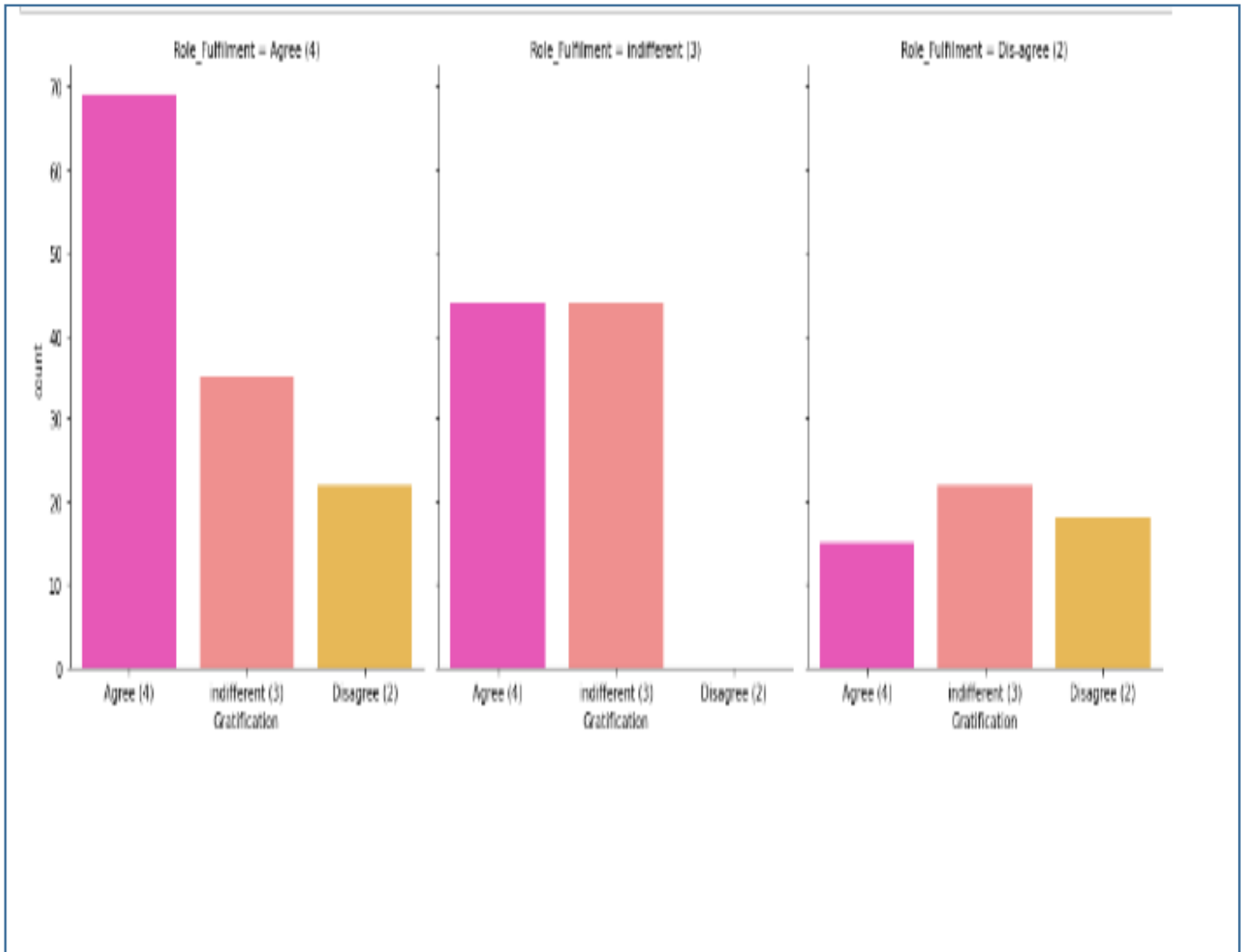


Observations from the above plots:

- Most of the customers agreed that shopping on the website gives the sense of adventure. The adventures in the shopping websites gives positive activity to experience an amplified enjoyment to

the customers while shopping on websites. They also believe that shopping on preferred e-tailer enhances the social status of the customers. Many customers think they are adventuring while shopping online as they search for low cost and high discount products to buy and prefer the same to the others. In this way they think that shopping in the website gives them the adventure.

- Shopping online won't affect anyone's status and the customers agreed that shopping on preferred e-tailer enhances their social status

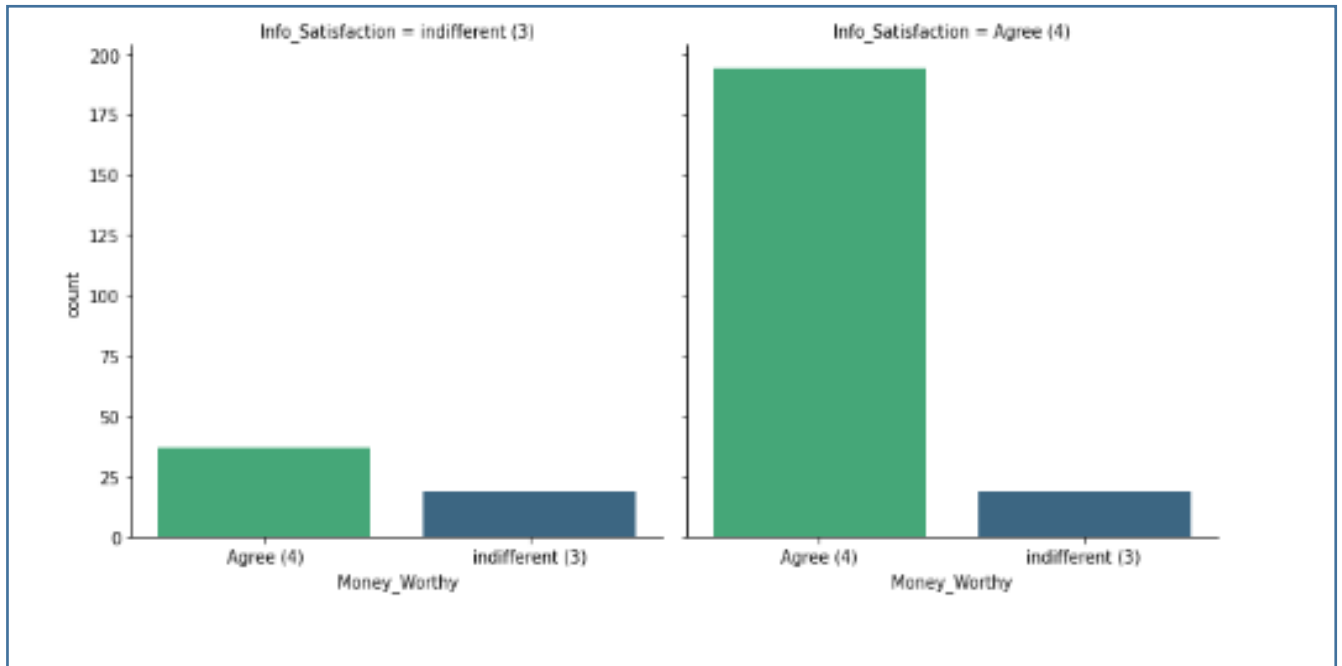


Observations from the above plots

- Most of the customers agreed that they felt gratified while shopping on their favorites e-tailer. This is because the e-tailer companies can successfully make up for a mistake or a dissatisfied customer is to be equally

expedient in addressing the customer's needs.

- . Also, most of the customers agreed that shopping on online website helps them fulfil their certain roles. Fulfilment refers to activities that ensure customers receive what they ordered, including the time of delivery, order accuracy, and delivery condition, also the customers cannot see the product directly before they purchase it. Companies must ensure delivery timeliness, order accuracy, and delivery conditions to provide superior service quality for customers. The companies must understand that the customer satisfaction is an indication of the customer's belief of the probability of a service leading to a positive feeling. If the companies give positive vibration to the customers by giving chance to fulfil their roles then they shop more on that particular website.

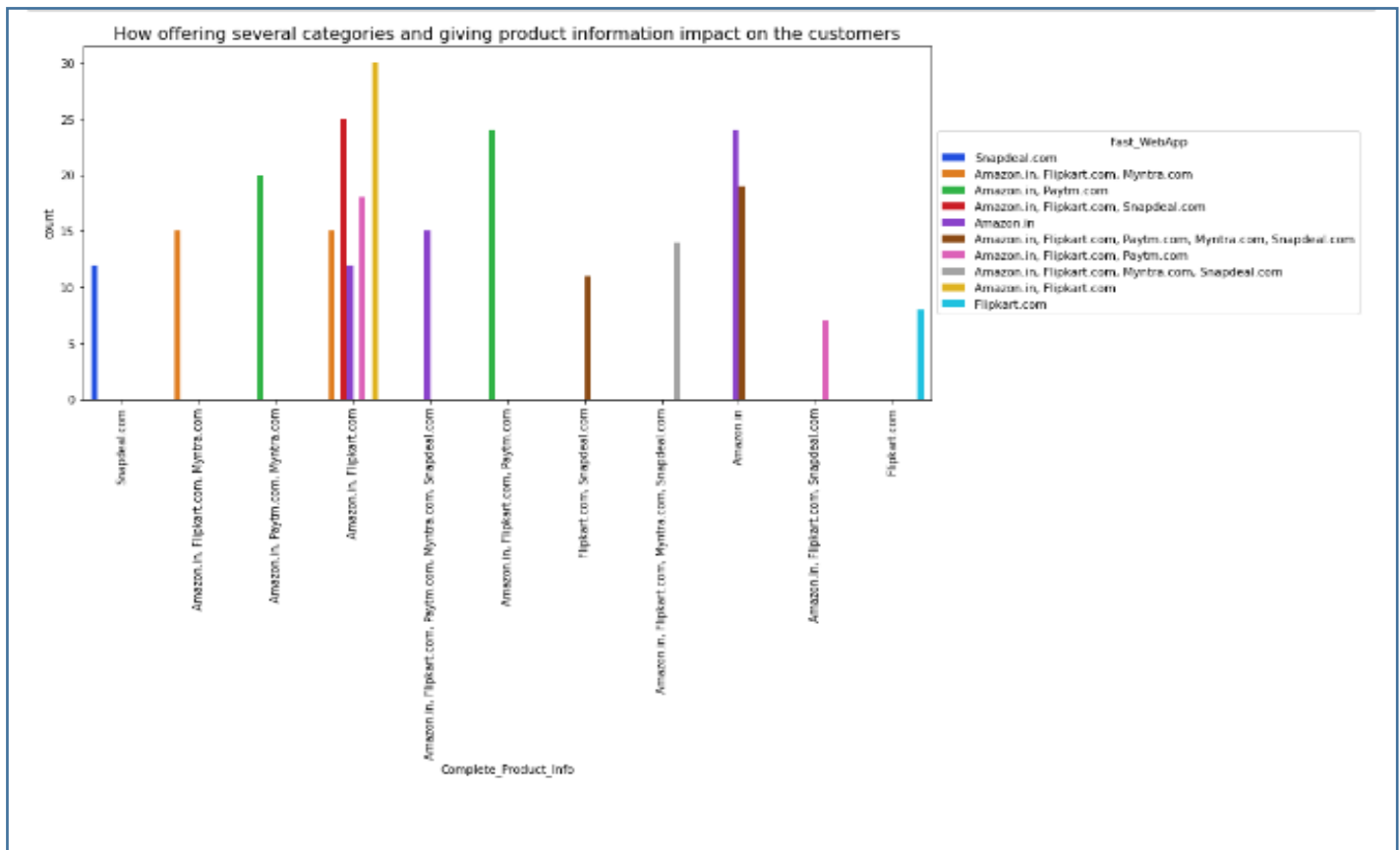


Observations from the above plots:

- The customers should satisfy with their product that they shopped from the online store then only they agreed that they got value for the money they spent. The companies should display the quality information about the products so that the customers being able to purchase their product and thinks that it worth for money and this comes under utilitarian value.

lots of varieties of similar products with different brand.

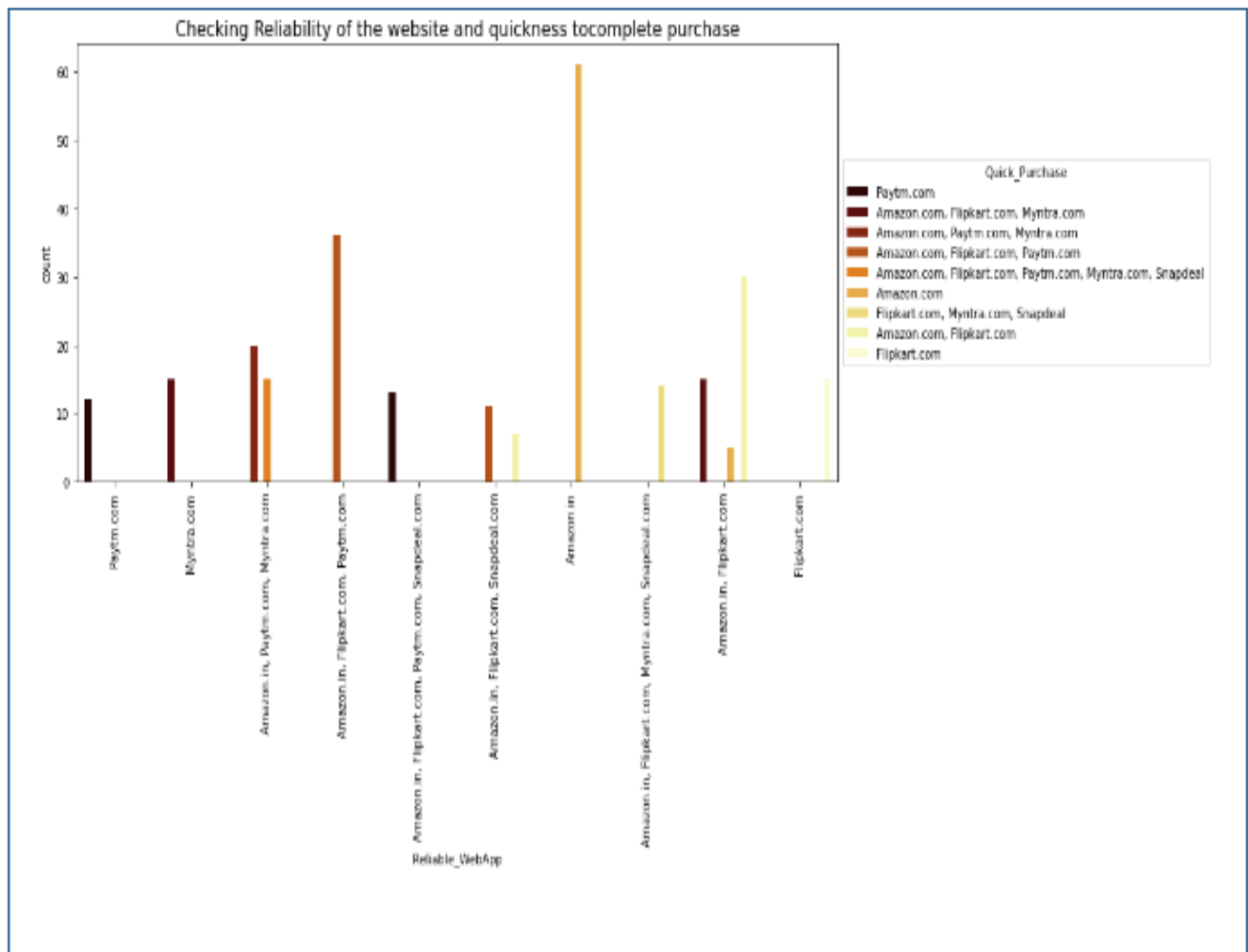
- . Amazon and Flipkart have high visual appealing web-page layout compared to others that means these websites provides some colorful graphics on the homepage. The more people find the website attractive, there are higher chances that they will stay a little longer in that website, and also these websites provide wild variety of products in an attractive manner which makes the customers to buy the product.



Observations from the above plot:

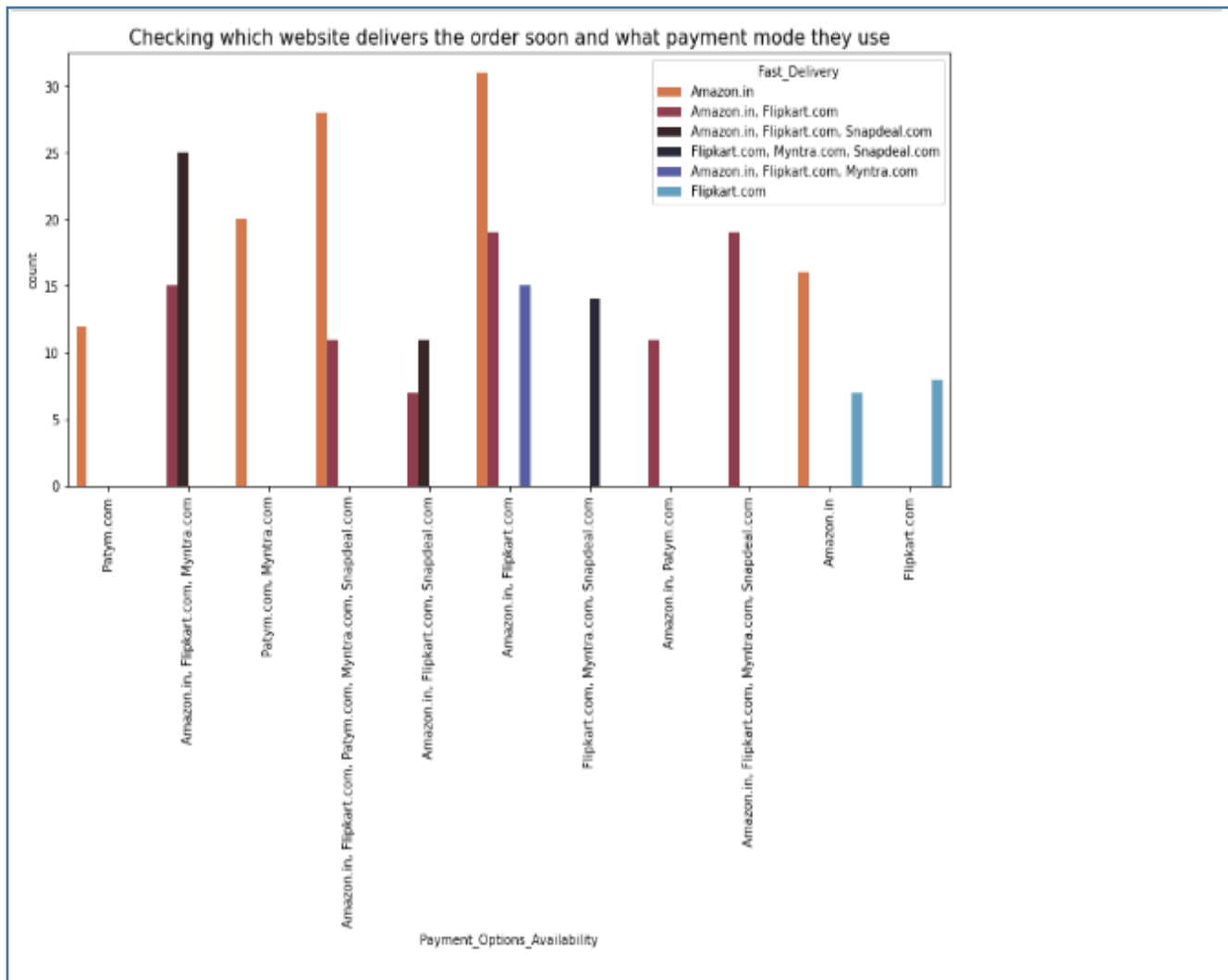
- The most common problem which is faced by the customers is server problem, in many websites these problems are common. Also, only few websites provide the relevant information about the products. These things may lead the customers to have bad impression on these websites.

- From the plot we can visualize that the amazon and flip kart websites gives complete and relevant information and these websites have no issue with the server and most of the customer liked the web speed of both amazon and flipkart.



Observations from the above plot:

- . The consumer determines the shop's reliability based on the information transmitted by the shop and certain sites offer customers the opportunity to purchase items that are used which means they are likely to be the most reliable. Some of the customers completes their purchase very quickly due to the discount, less price, free delivery charges etc provided by the ecommerce websites.
- . From the plot we can notice amazon site is more reliable and most of the customers complete their purchase on amazon very quickly.

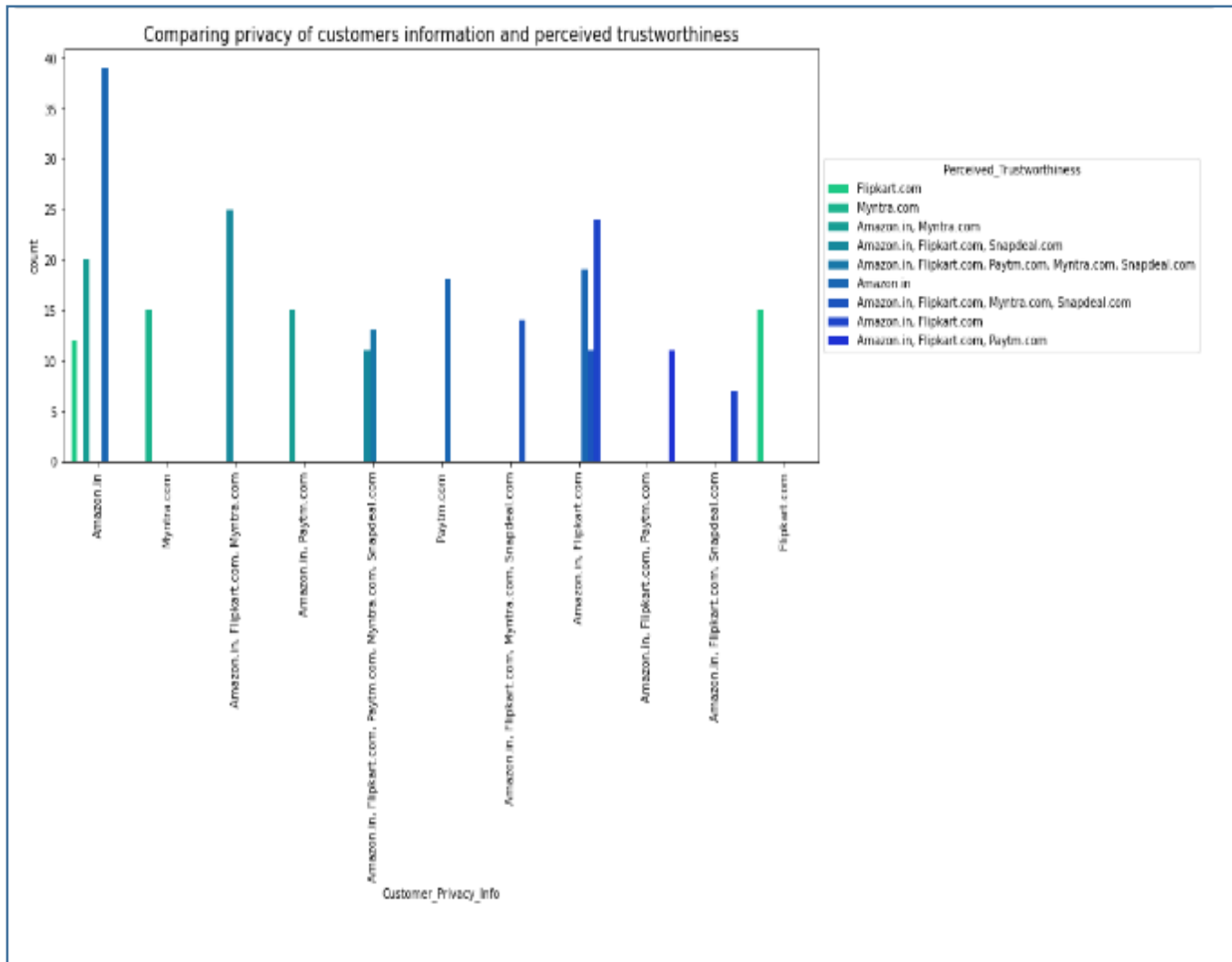


Observations from the above plot:

- Having different types of payment methods will help the customers to pay the invoice easily using their choice of payment and if the websites have

speedy delivery methods without delivery charge, then the customers like to buy the products in those websites.

- . Here amazon and flip kart have several payment options and amazon indeed has speedy order delivery compared to other websites.



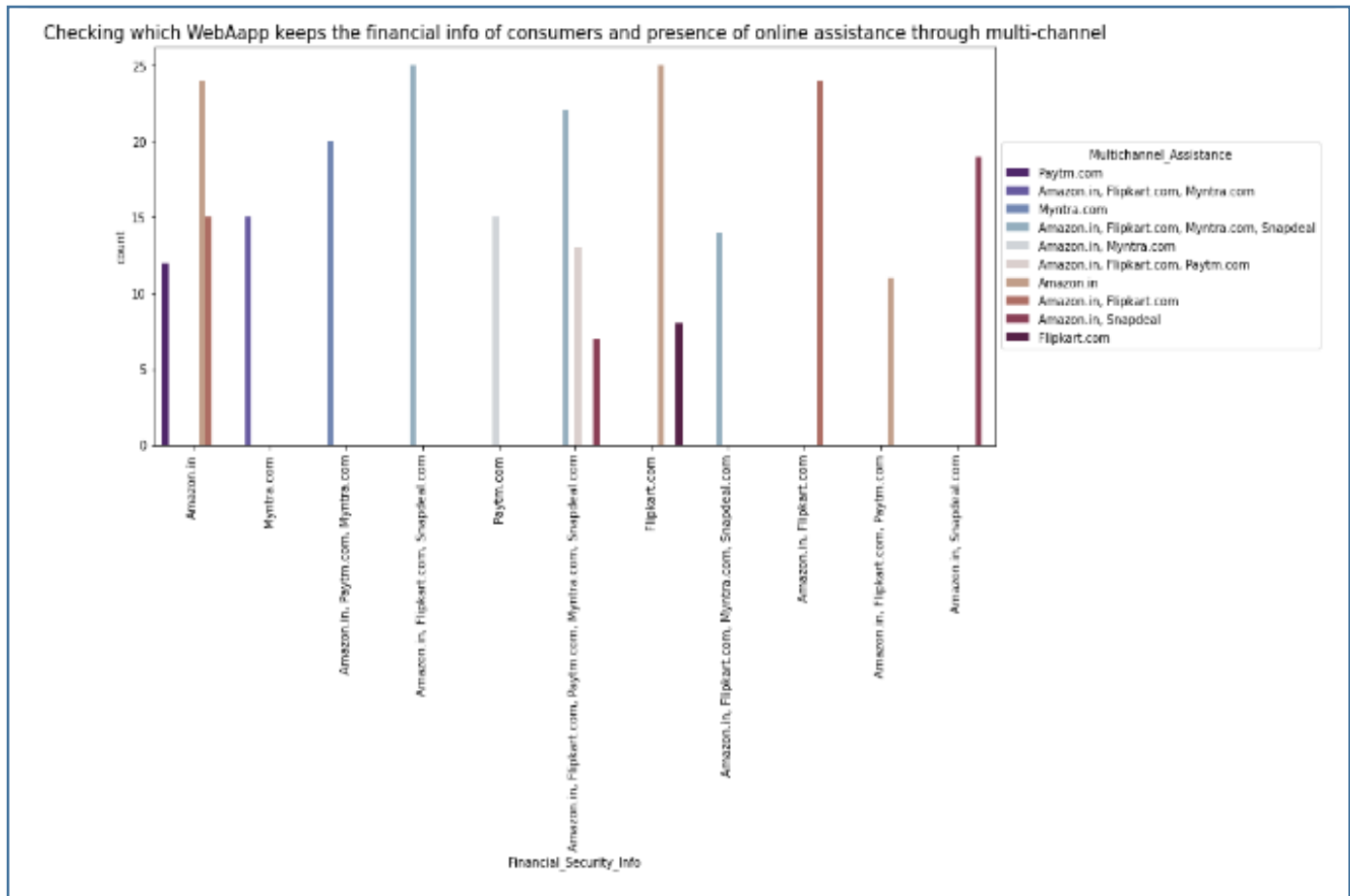
Observations from the above plot:

- Security/privacy refers to the security of credit card payments and privacy of shared information like name of the customer, address and phone number. Customers are always concerned whether the website would protect them

against fraud after a transaction. So, website security and privacy are important to assess the service quality of online stores. The customers think that buying online means taking risk, in this case trust is more important thing for both merchant and customer.

- . Most of the customers trusts amazon followed by flip kart in terms of keeping their privacy of data information secured and the customers who believes that amazon website keeps their financial information as secrete also trusts flip kart, Myntra, Snapdeal and Paytm in terms of keeping their financial information secured. Most of the customers believed that Amazon has perceived trustworthiness compared to others. Apart from this, customers

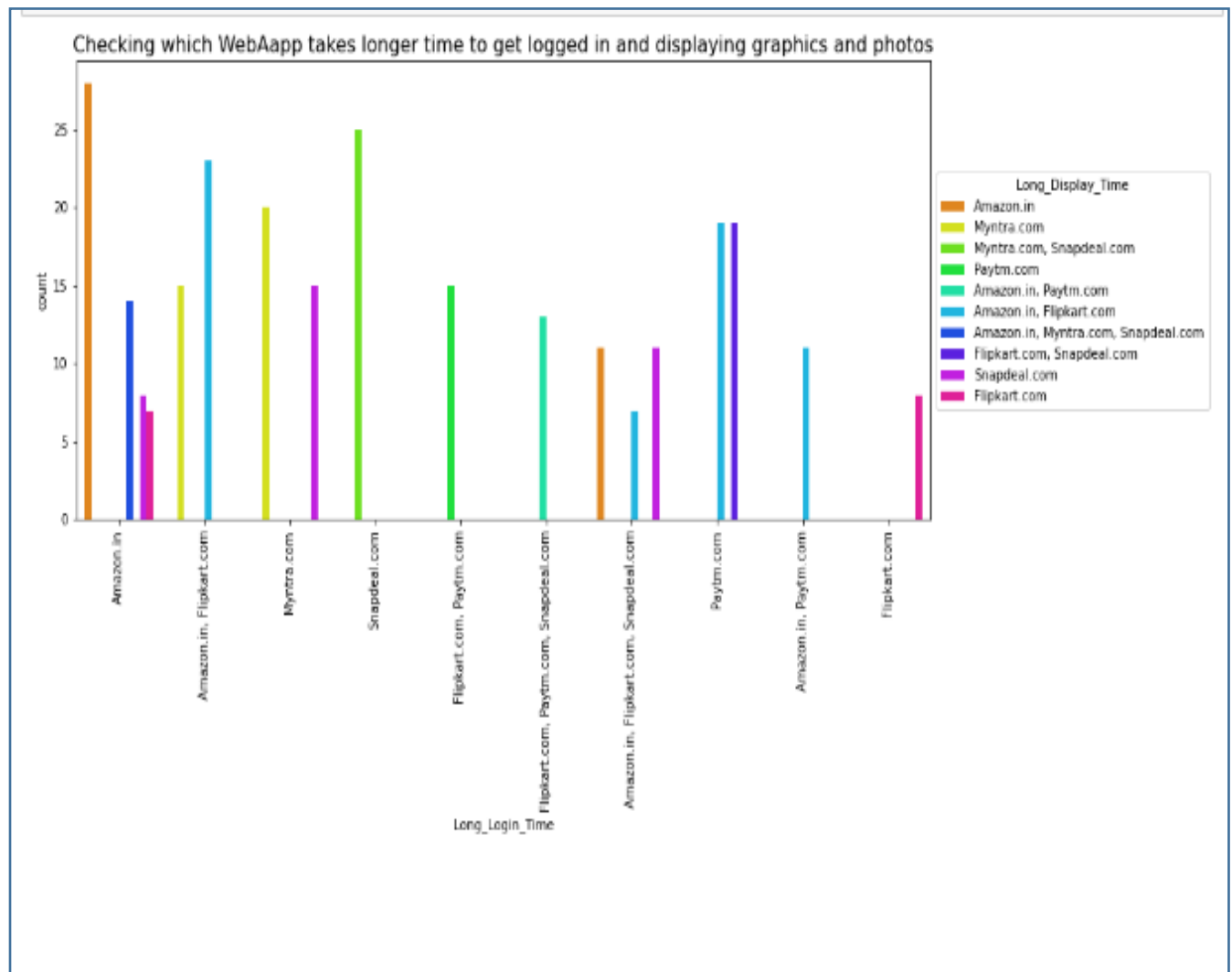
believed that flip kart and Myntra also have perceived trustworthiness.



Observations from the above plot

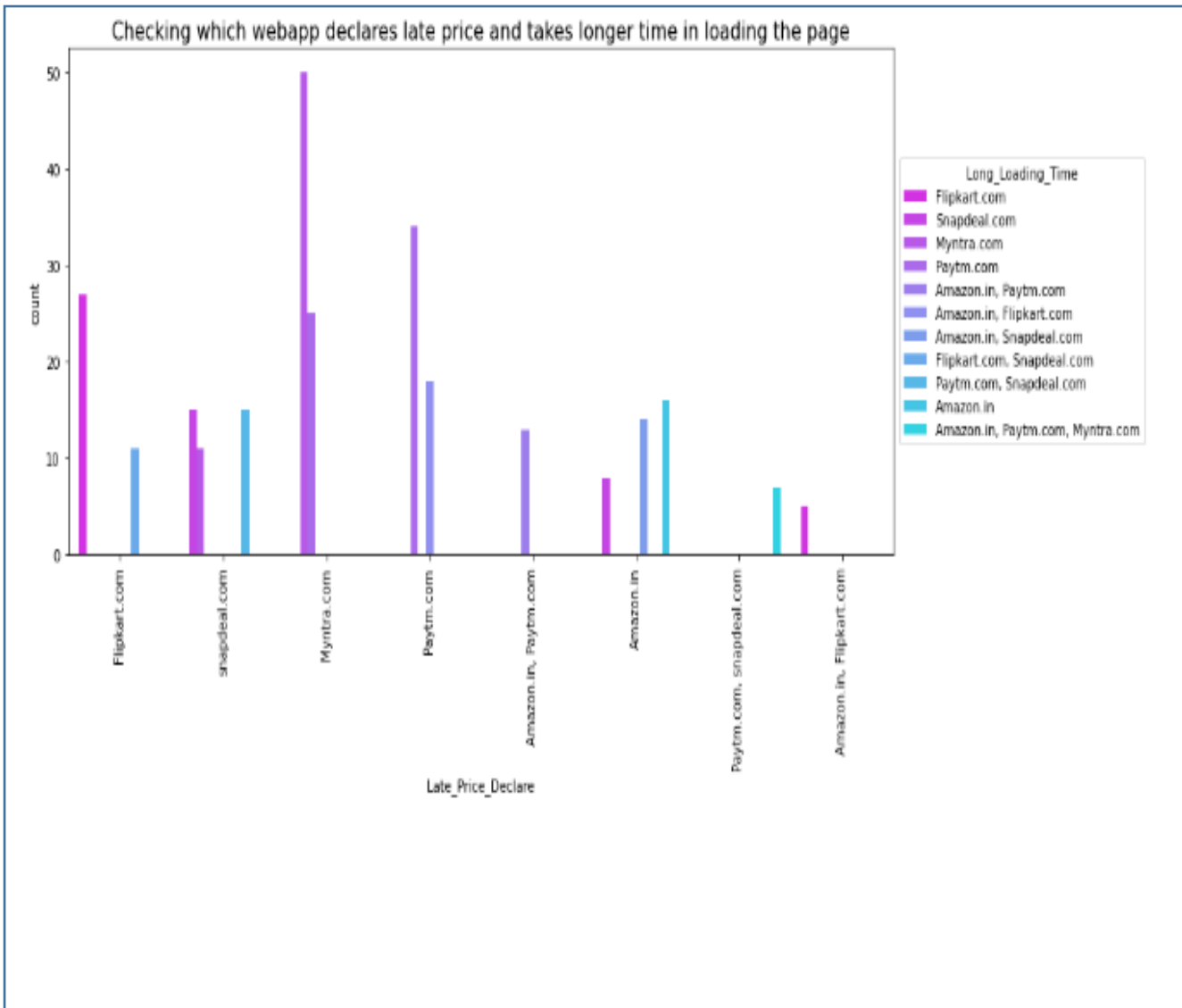
- The customer's trusts that amazon and flip kart keeps their financial information private and they never share any type of information to others.
- Multi-channel retailing provides several benefits which includes several shoppers like the convenience that is provided

through online channels in comparison to physical stores. Most of the customers like Amazon in terms of presence of online assistance through multi-channel.



Observations from the above plot:

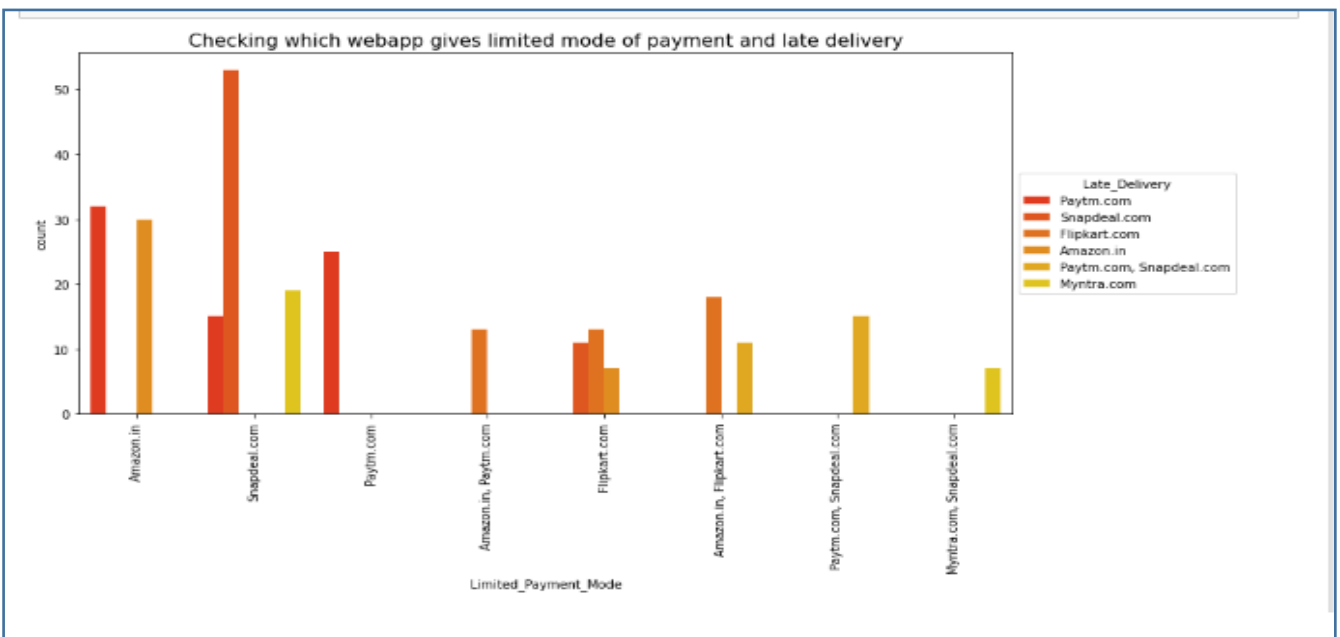
- . The customers mostly choose amazon website for buying products as it gives promotions and sales periods in some days, on these days most of the customers attracted by the offers provided by the websites, wants to buy the products. So, amazon will take more time to allow the customers to get login into the site.
- . When there is promotion or sales period, amazon and Myntra takes longer time to display the graphics and photos



Observations from the above plot:

- When there is promotion and sales, Myntra takes time to load the page and it has late declaration of price in these days.

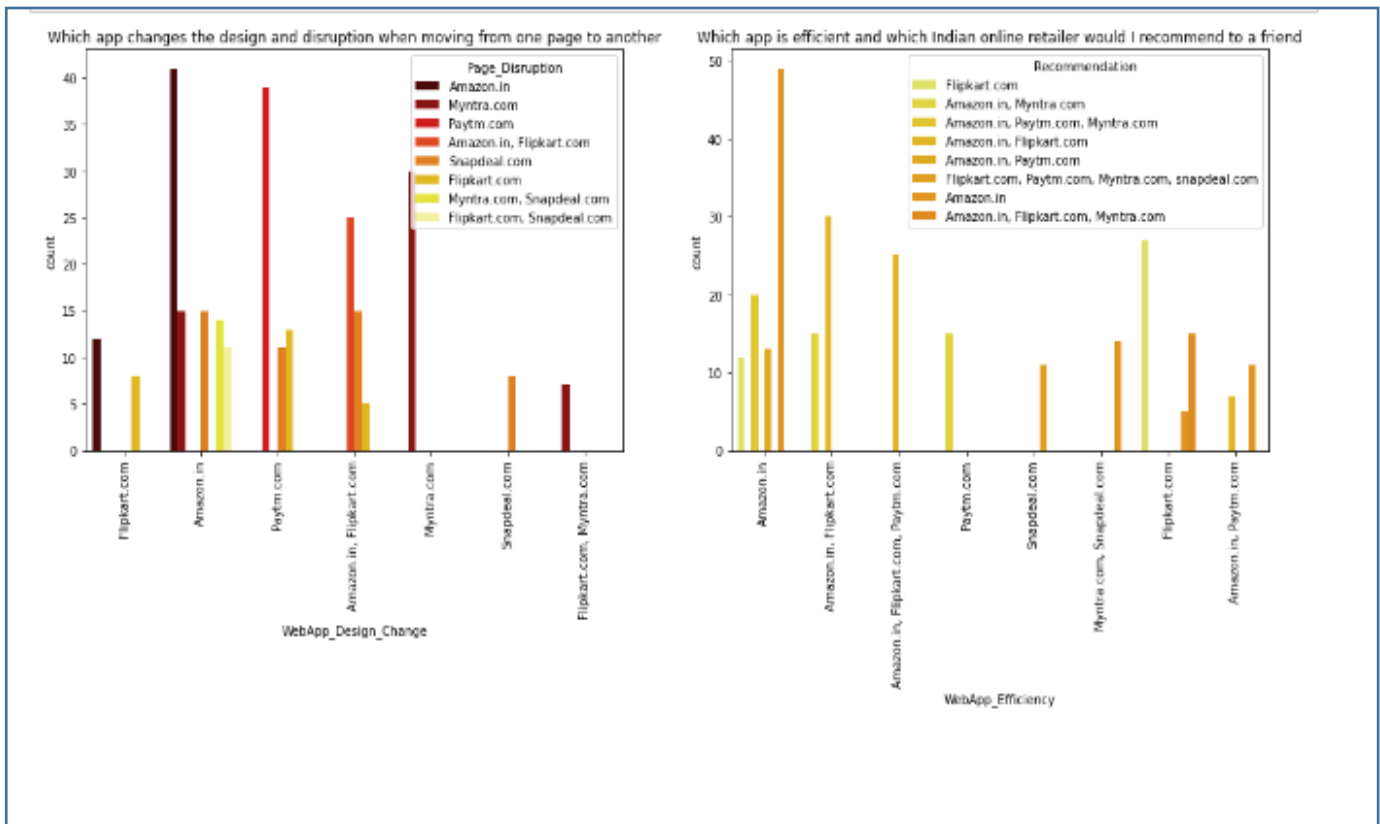
- Myntra declare the late price in order to clear the sales and they fix the price by comparing with other websites and they end up sales by providing benefits to the customers. In this time most of the customers tries to shop in this website so it takes long loading time.



Observations from the above plot:

- Snapdeal has limited mode of payment on most of the products followed by

Amazon. And paytm takes more time to deliver the product. So this website may not satisfy the customers due to late delivery.



Observations from the above plot:

- Amazon is the website where they frequently change their application designs in order to attract the customers and satisfies the customers' needs and they tend to make customers by updating everyday as per the

trend. But the disadvantages of this website are when moving from one page to other it slows down and sometimes it may shutdown.

- Amazon is the website which is more efficient as before and I suggest Amazon.com and Flipkart as a best Indian online retailer store for purchasing all types of products, as they provide enormous amounts of benefits.

Since we have observed all the columns contains object type data, so we need to convert them into numerical by using appropriate encoding techniques. Here I am using label encoding method to convert the data.