Problem statement

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.

Solution

We are analysing the data set to see which factors influence customers to prefer one website over another, and how these factors can be improved upon to increase customer retention.

Steps taken for initial Exploratory Data Analysis of the entire data set:

1. Importing necessary libraries and the customer retention data set to Jupyter Notebook.

I imported both the data sheet and the encoded sheet as df and df2 to be able to analyse categorical data more clearly.

2. Basic exploratory data analysis by checking:

- Data frame's dimensions: 269 rows x 71 columns
- Data information
- Data frame’s columns
- Data frame’s columns’ data types: All but one column is categorical type. The pin code column is the only column with integer type data.
- Whether there are any null values: No null values
- Number of unique values in each column

3. Dropped unnecessary columns.

These were columns containing ordinal data like “strongly agree”, “agree”, “disagree”, etc, that did not give us any information about the website the customers were rating. So, they were removed from the data set.

4. Renaming columns

Certain columns containing direct questions like “How old are you?” and similar questions were renamed to “Age” and other shorter, more accurate names.

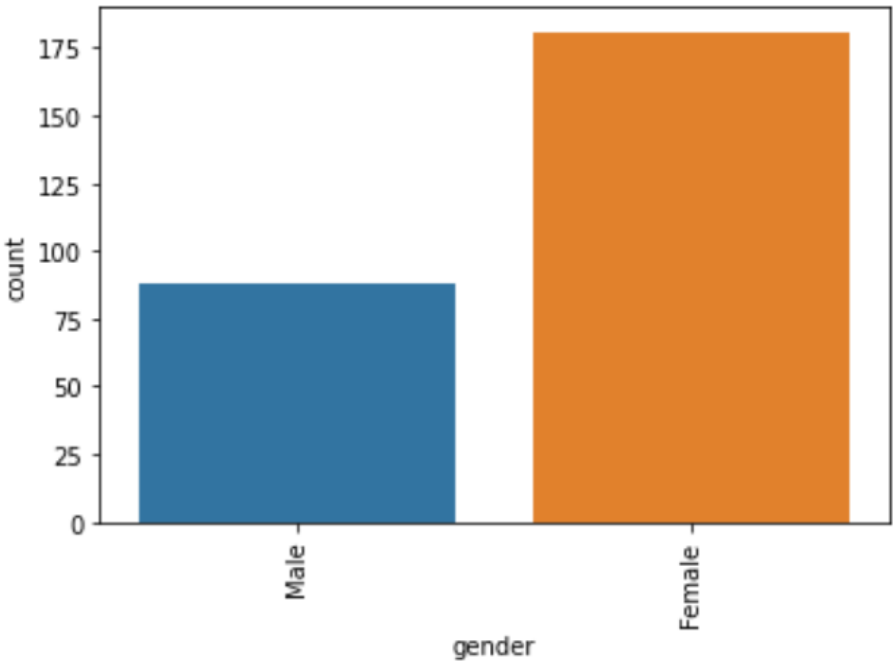
Steps taken for in-depth Exploratory Data Analysis:

1. **Gender**

- Checking the gender of the customers.

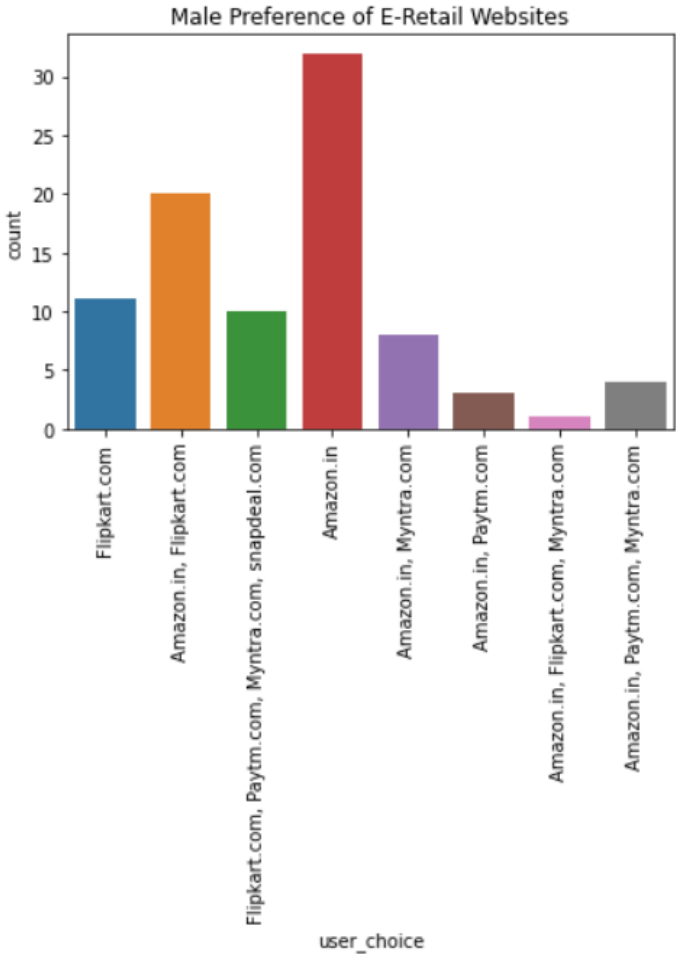
- I plotted a count plot to see how many of the customers are male and how many are female, and then found the exact value counts.

- Observation: Out of the 279 customers, there are 88 males and 181 females.

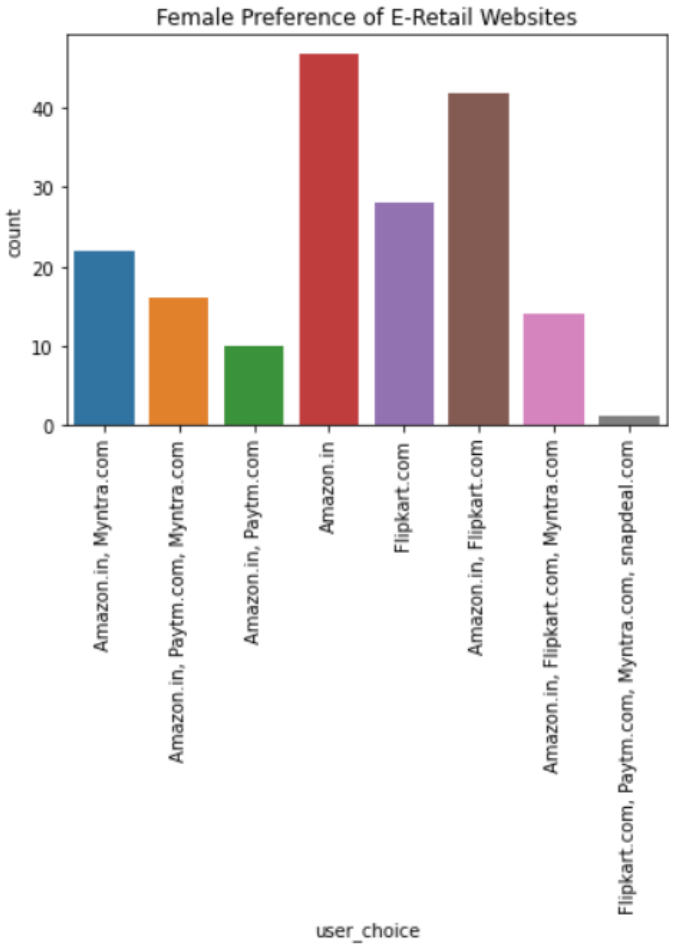


- Then, I plotted graphs to see the difference in male preference and female preference in websites, using the encoded dataset to get male and female values and their final website recommendation/preference.

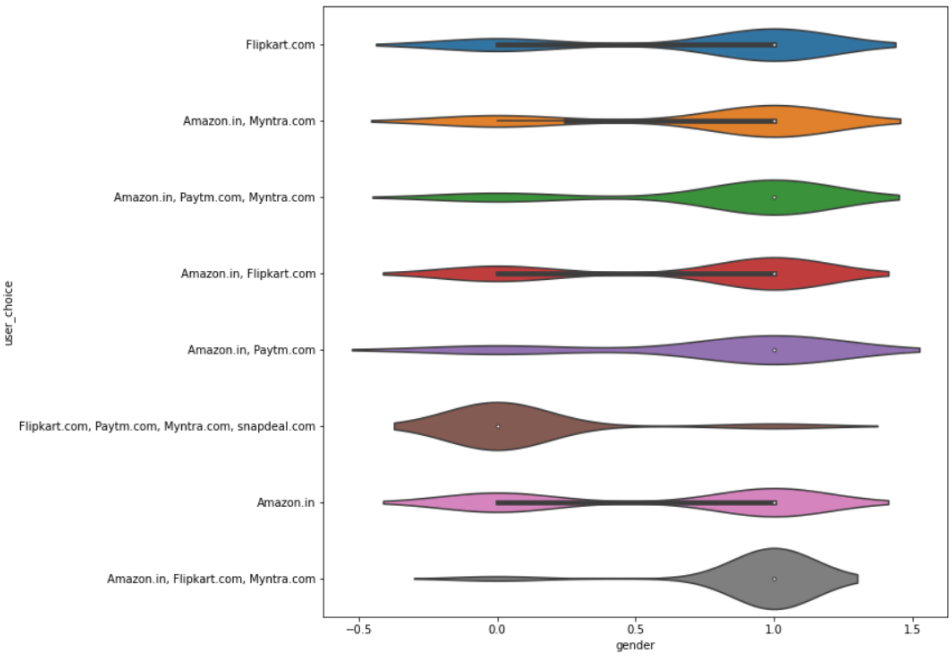
  Male preference:



Female preference:



- I plotted a violin plot to view the male and female preferences side by side. Male is encoded with value 0, while female is 1.



- Observations:

  1. Most males preferred just Amazon.in, and Amazon.in and Flipkart.com.

  2. Least males preferred Amazon.in, Flipkart.com, Myntra.com

  3. Most females preferred Amazon.in and Amazon, Flipkart.

  4. Least females preferred Flipkart, Paytm, Myntra and Snapdeal.

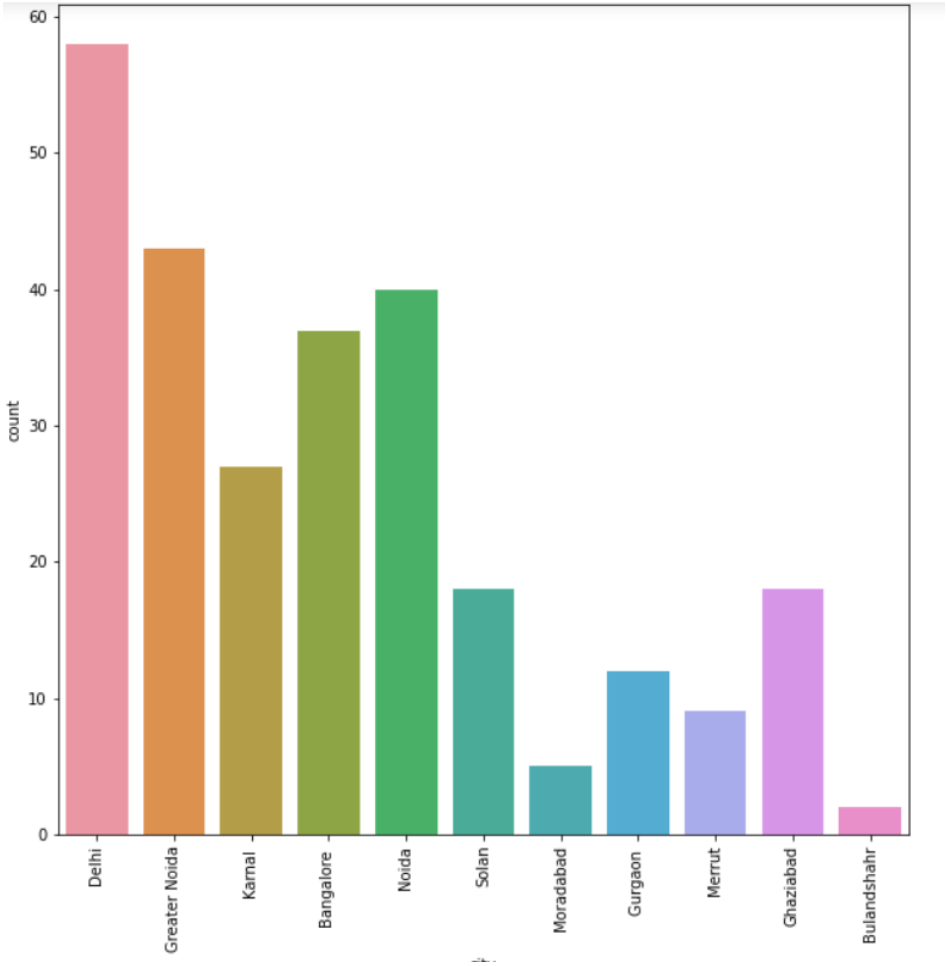
  5. Females preferred Amazon, Flipkart, Myntra significantly more than males, whereas males preferred Flipkart, Paytm, Myntra, Snapdeal significantly more than females.

  6. All the other distributions are more or less equivalent.

2. **City**

- Checking which cities the customers in this survey belong to.

- I plotted the count of customers from each city:



- Observation:

  1. Maximum number of customers are from the Delhi-NCR region.

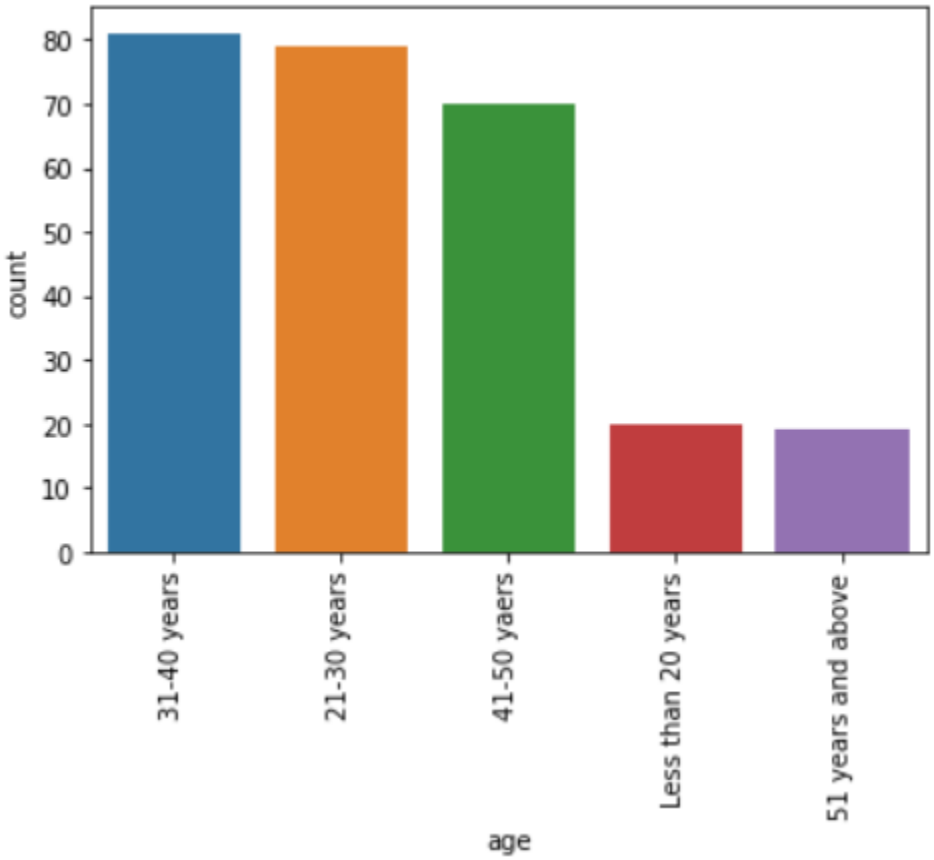
  2. Most cities are in North India, with the exception of Bangalore.

  3. Least number of customers are from Bulandshahr.

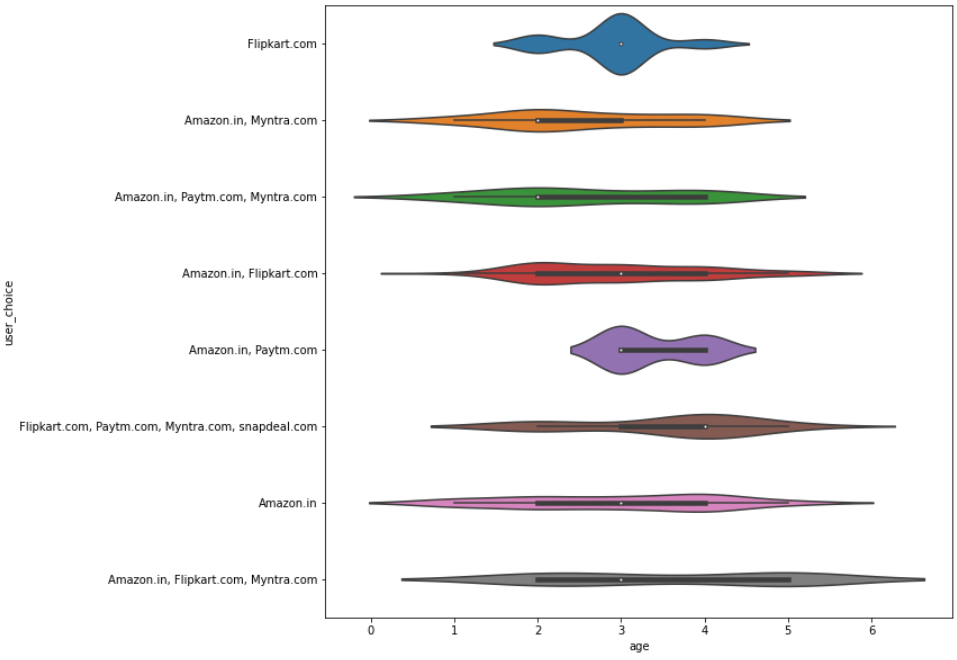
3. **Age**

- Checking if age influences website preference.

- I plotted the count of the different age groups:



- I plotted a violin plot to see the effect of age on website preference:



- Observation:

1. Most customers are in the age groups 31-40 and 21-30. Least customers are in age groups less than 20 years and 51 years and above.

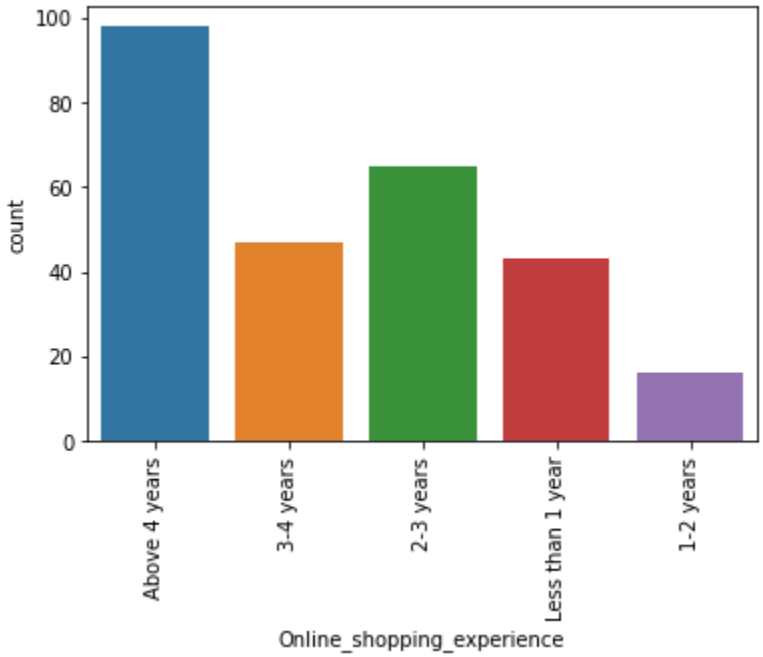
2. People in younger age groups preferred Myntra more, and people in older age groups preferred Paytm and Snapdeal more.

3. Amazon and Flipkart is preferred by all customers regardless of age.

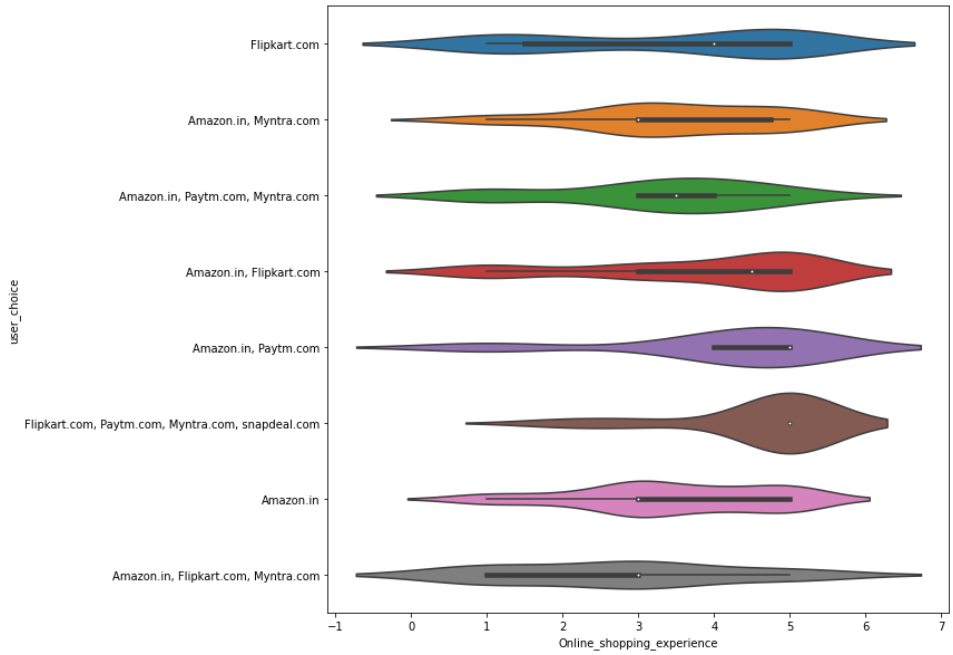
4. **Online Shopping Experience**

- How long have the customers in this survey been shopping online?

- I plotted the count of the categories:



- I plotted a violin plot to see the effect of age on website preference:



- Observation:

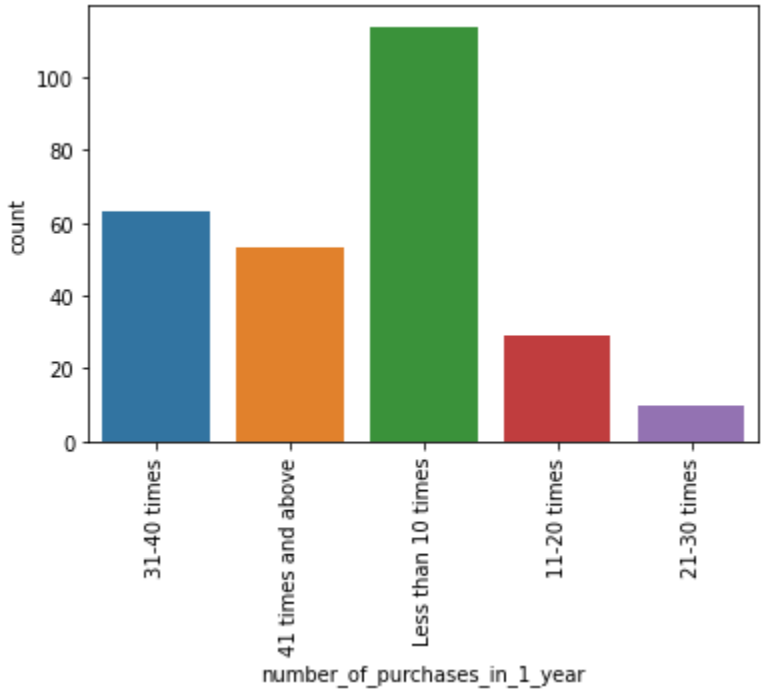
  1. Most customers in this survey have online shopping experience of 4+ years and 2-3 years.

  2. Least of them have been shopping for 1-2 years.

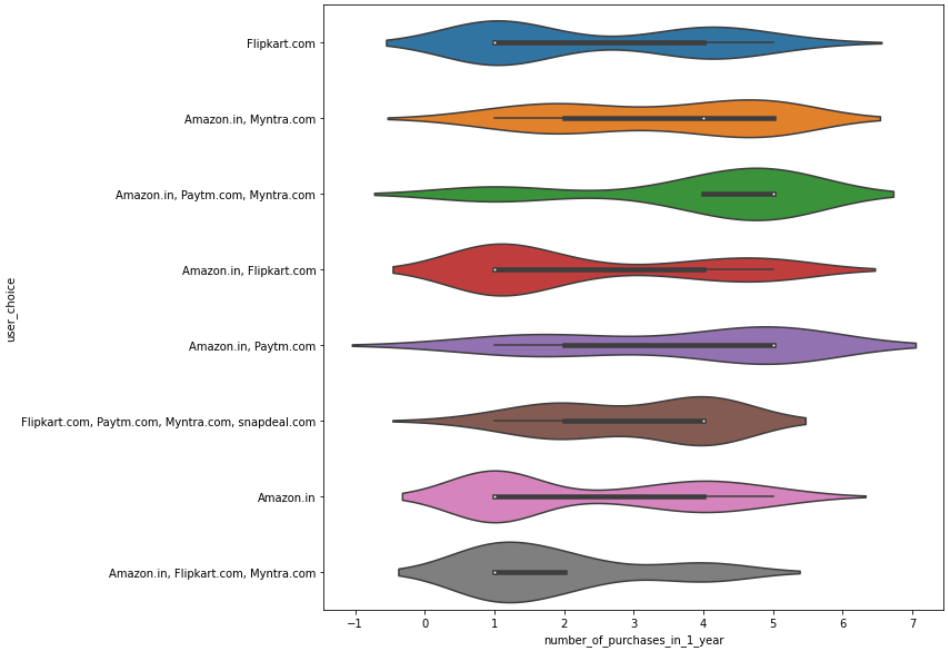
  3. Customers who have been shopping online longer preferred Amazon, Flipkart, Paytm, Myntra and Snapdeal. Customers who recently started shopping prefer Amazon and Flipkart.

5. **Number of online purchases in 1 year by customers**

- I plotted the count of the column:



- I plotted a violin plot to see the effect of number of online purchases in 1 year on website preference:



- Observation:

  1. Most customers ordered less than 1o times, or in contrast, 31-40 times.

  2. Least number of customers ordered 42+ times.

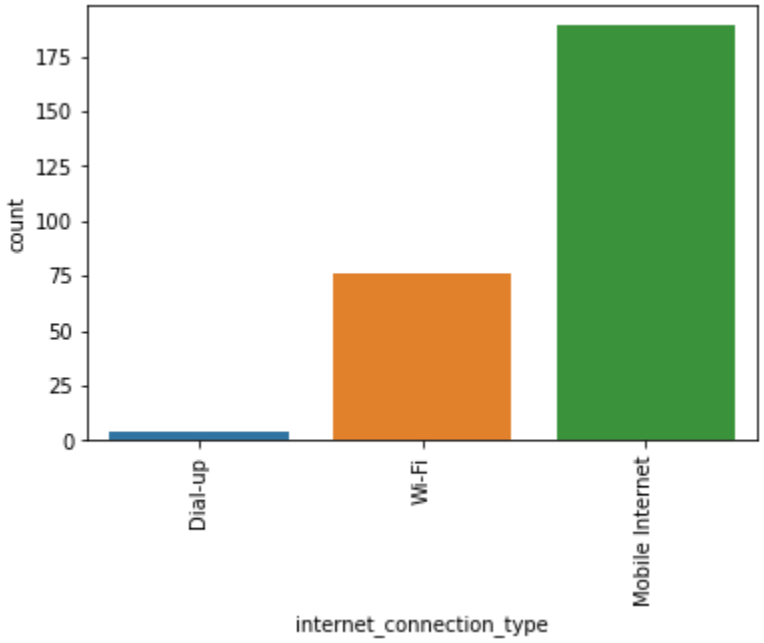
3. The customers who ordered infrequently preferred Amazon and Flipkart. The customers who ordered more frequently also preferred Myntra, Paytm and Snapdeal.

4. Most customers, regardless of shopping frequency preferred Amazon and Flipkart.

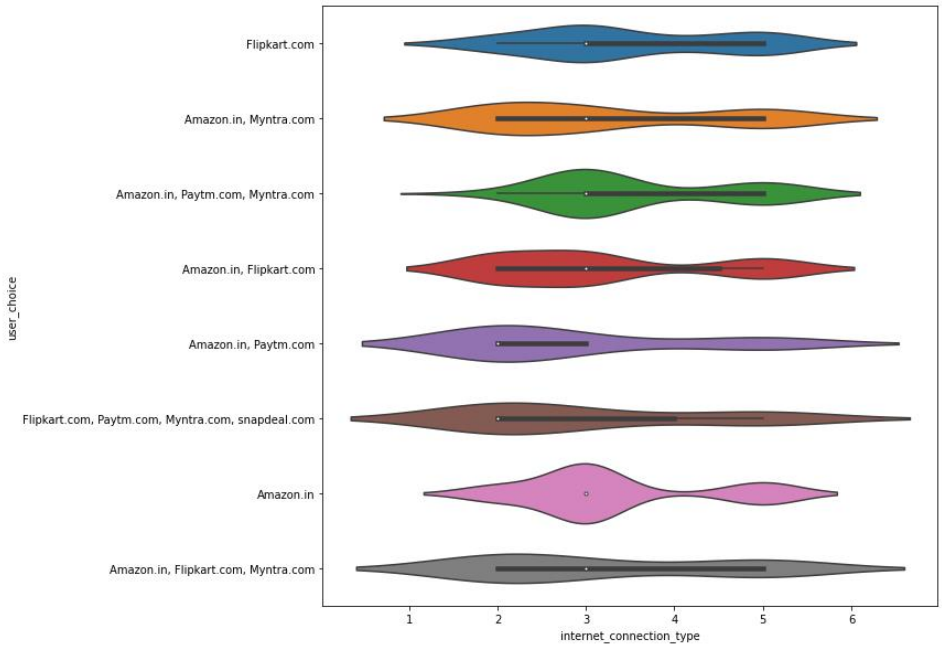
6. **Type of Internet Connection used by customers**

- I observed that Mobile internet type within the Internet connection category was being displayed twice due to a difference in upper and lower case. I combined the mobile internet data into one data type.

- I plotted the count of the column:



- I plotted the violin plot to see the effect of internet connection on website preference:



- Observation:

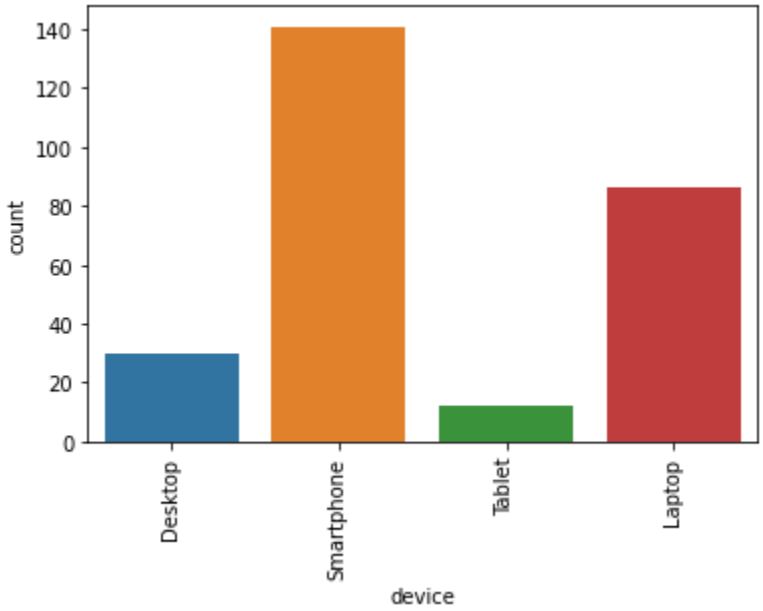
  1. Most customers used mobile internet to access the e-retail websites.

  2. A moderate amount used Wi-Fi, and least number of customers used dial-up.

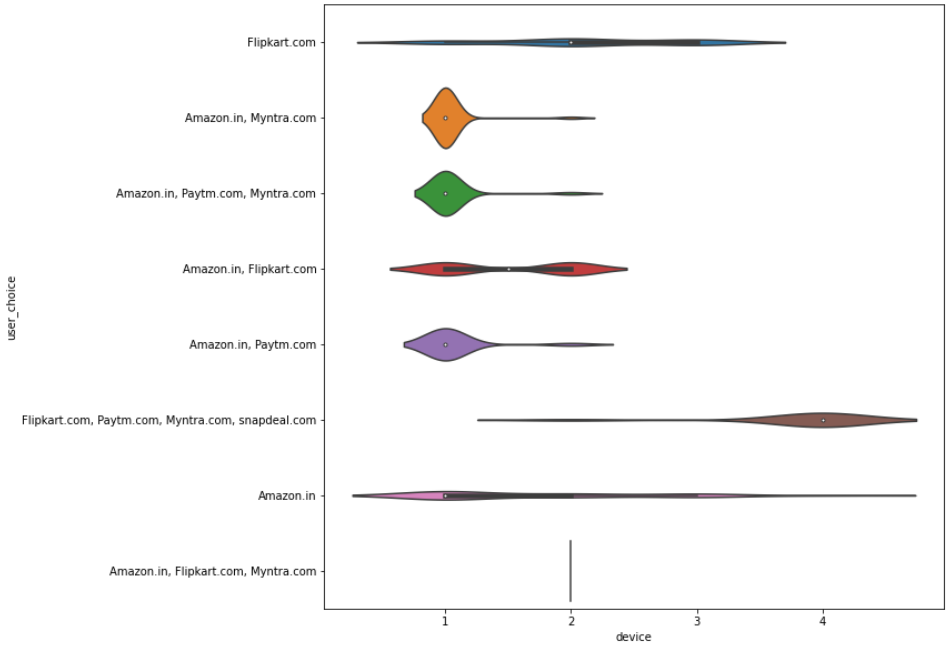
  3. Customers with dial up preferred Amazon, and customers with mobile internet preferred Amazon the most. Customers with Wi-Fi also preferred Amazon and Flipkart.

7. **Type of Device used by Customers**

- I plotted the count of the column:



- I plotted the violin plot to see the effect of type of device on website preference:



- Observation:

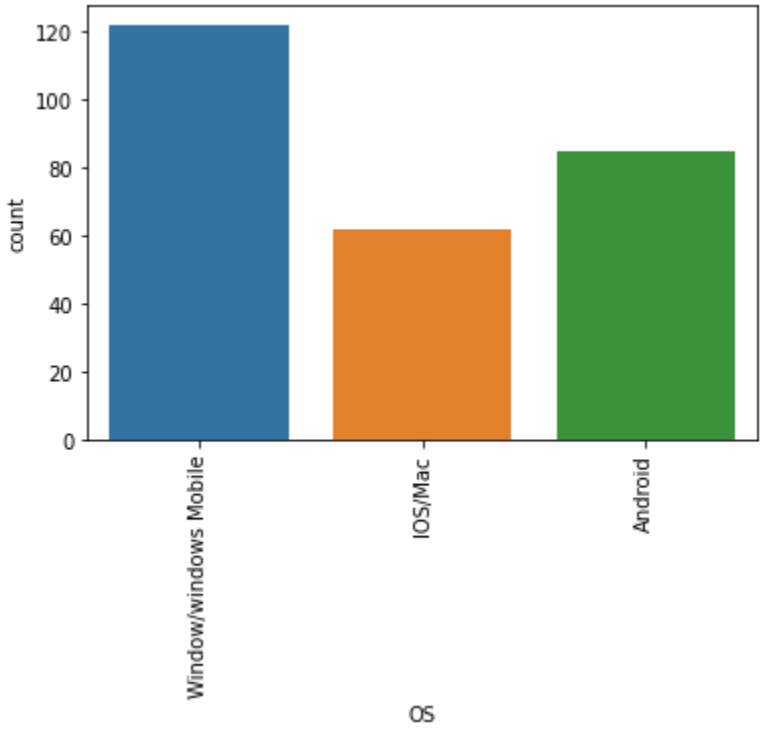
  1. Most customers used smartphone. Second majority of customers used laptop and 3rd majority used desktop.

  2. Least customers used tablet.

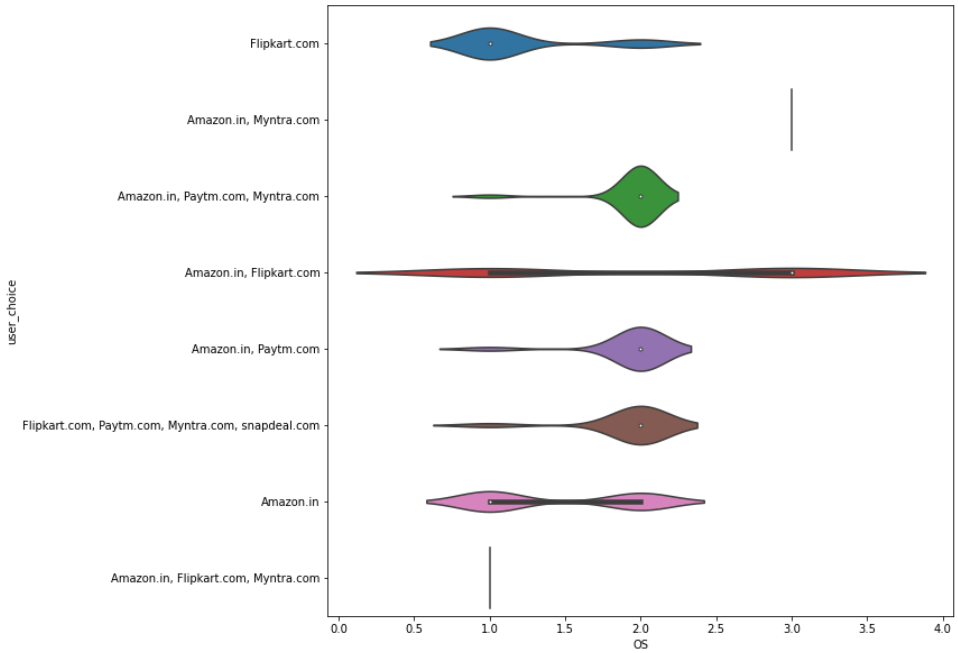
3. Customers who used tablet preferred Flipkart, Paytm, Myntra and Snapdeal the most. Customers who used smartphone preferred Amazon, Paytm, Myntra. Customers who used laptop preferred Amazon and Flipkart.

8. **Operating System used by customers**

- I plotted the count of the column:



- I plotted the violin plot to see the effect of type of device on website preference:



- Observation:

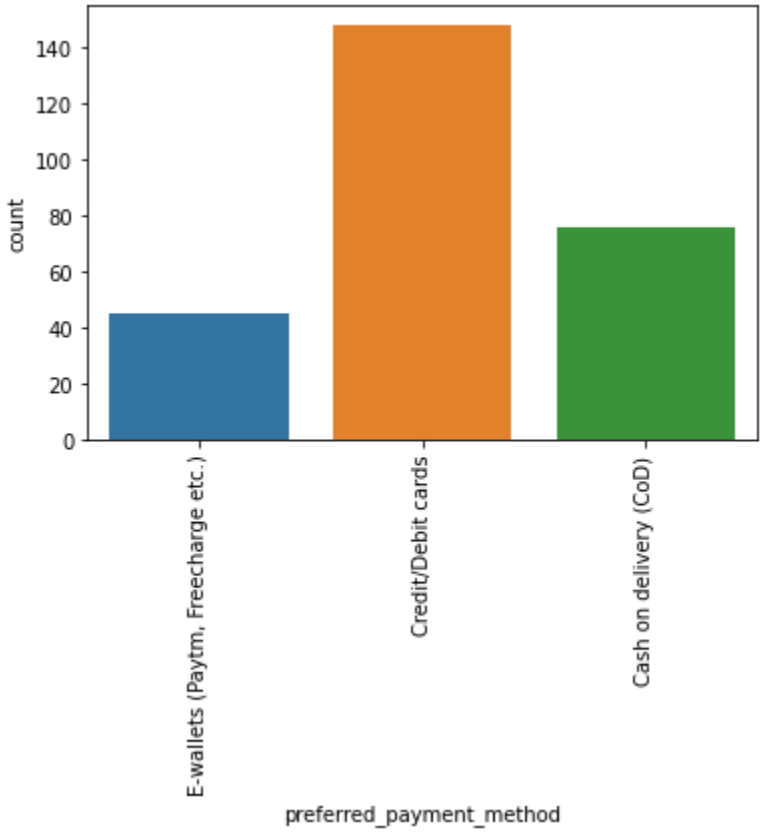
  1. Most customers used Windows or Windows phone.

  2. Second majority used Android, and least customers used IOS/Mac.

  3. Windows users preferred Amazon and Flipkart. Android users preferred Amazon, Paytm, Myntra and Snapdeal. IOS users preferred Amazon and Flipkart.

9. **Preferred Payment Mode of customers**

- I plotted the count of the column:



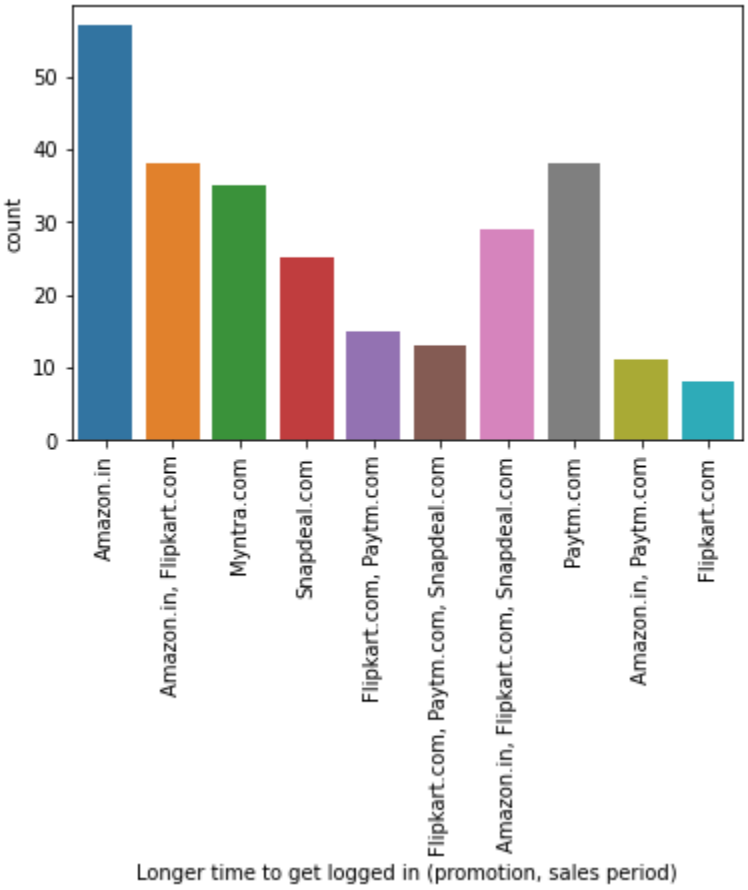
- Observation:

  1. Most customers preferred to pay using Credit/Debit cards.

  2. Second majority prefer cash on delivery (COD), and least customers use E-Wallets.

10. **Longer time to get logged in (promotion, sales period)**

- I plotted the count of the column:



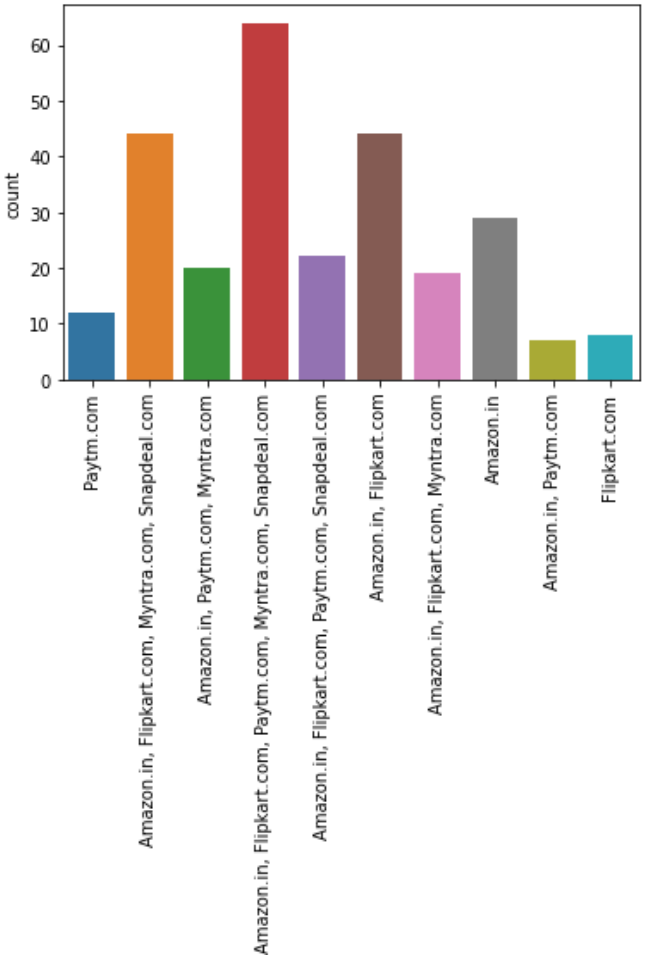
- Observation:

  1. Logging into Amazon takes the longest time during promotion or sales period.

  2. Logging into Flipkart takes the least amount of time during promotion or sales period.

11. **Easy to use website or application.**

- I plotted the count of the column:



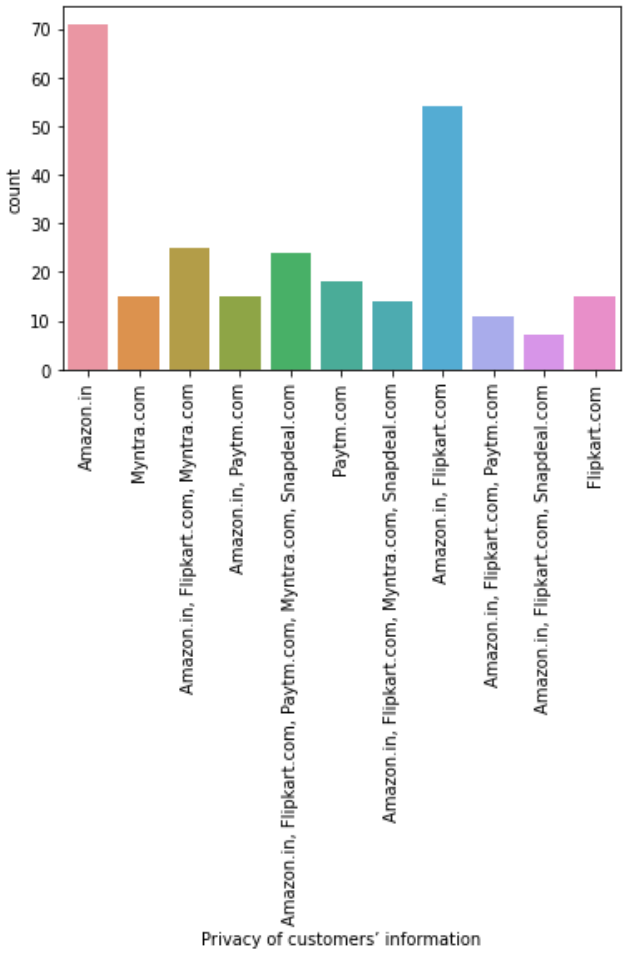
- Observation:

  1. Amazon and Flipkart are the easiest to use.

  2. Paytm is the hardest to use.

12. **Privacy of customer's information**

- I plotted the count of the column:



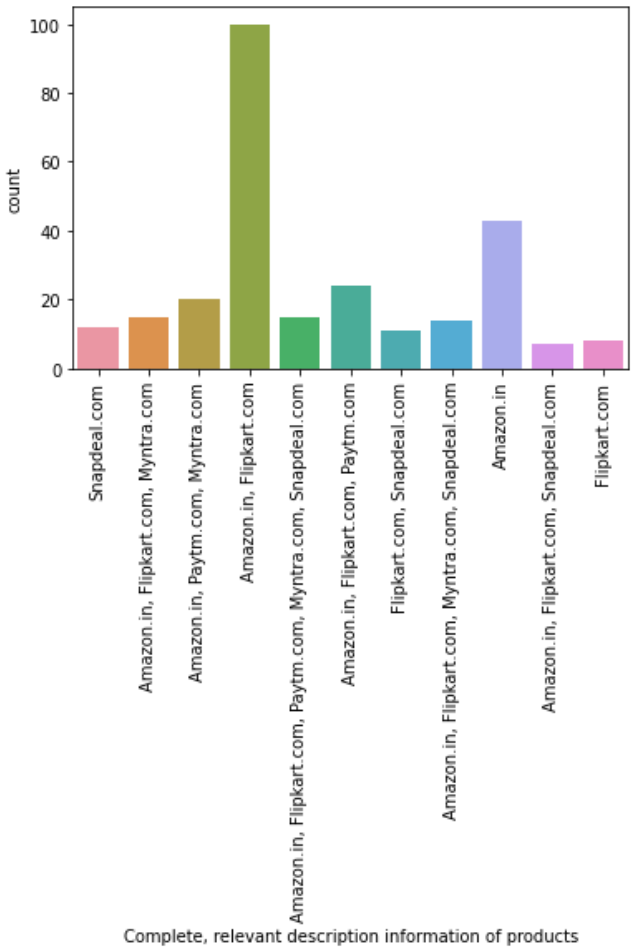
- Observation:

  1. Customers believed that Amazon respected the privacy of their information the most.

  2. Myntra, Snapdeal and Paytm ranked the lowest in privacy.

13. **Complete, relevant description information of products**

- I plotted the count of the column:



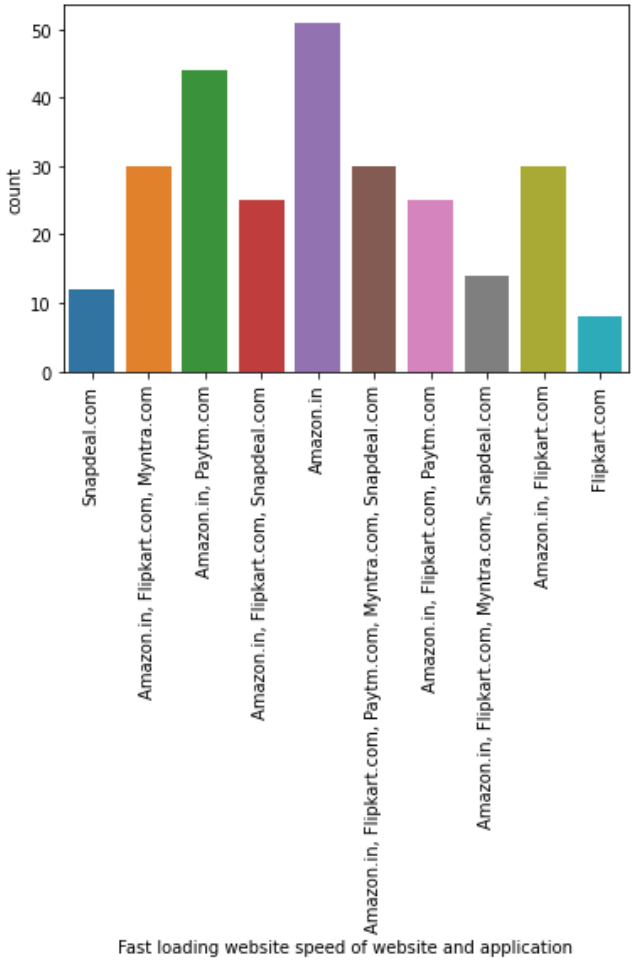
- Observation:

  1. Amazon and Flipkart have most description of products.

  2. Snapdeal and Paytm have the least information.

14. **Fast loading website speed of website and application**

- I plotted the count of the column:



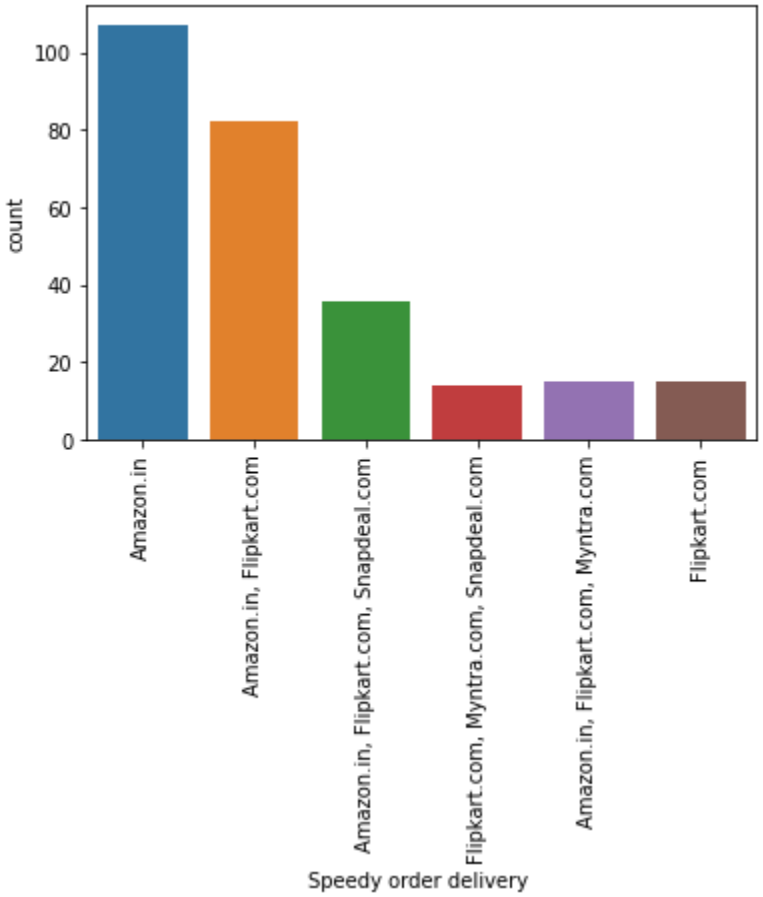
- Observation:

  1. Amazon, Paytm and Flipkart have fastest websites.

  2. Snapdeal has the slowest website.

15. **Speedy Order Delivery**

- I plotted the count of the column:



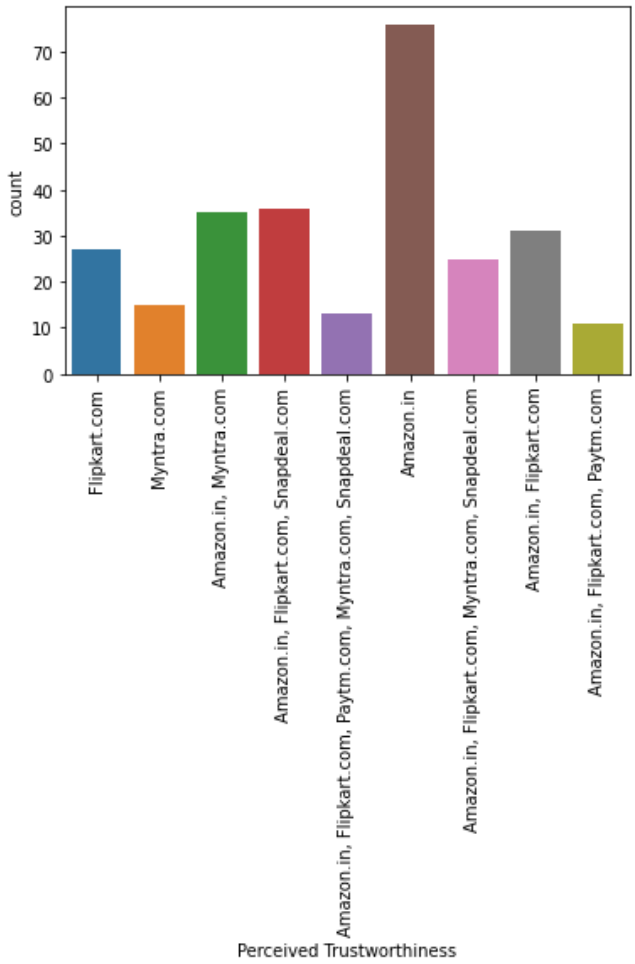
- Observation:

  1. Amazon and Flipkart have fastest delivery.

  2. Snapdeal and Myntra have the slowest delivery.

16. **Perceived Trustworthiness**

- I plotted the count of the column:



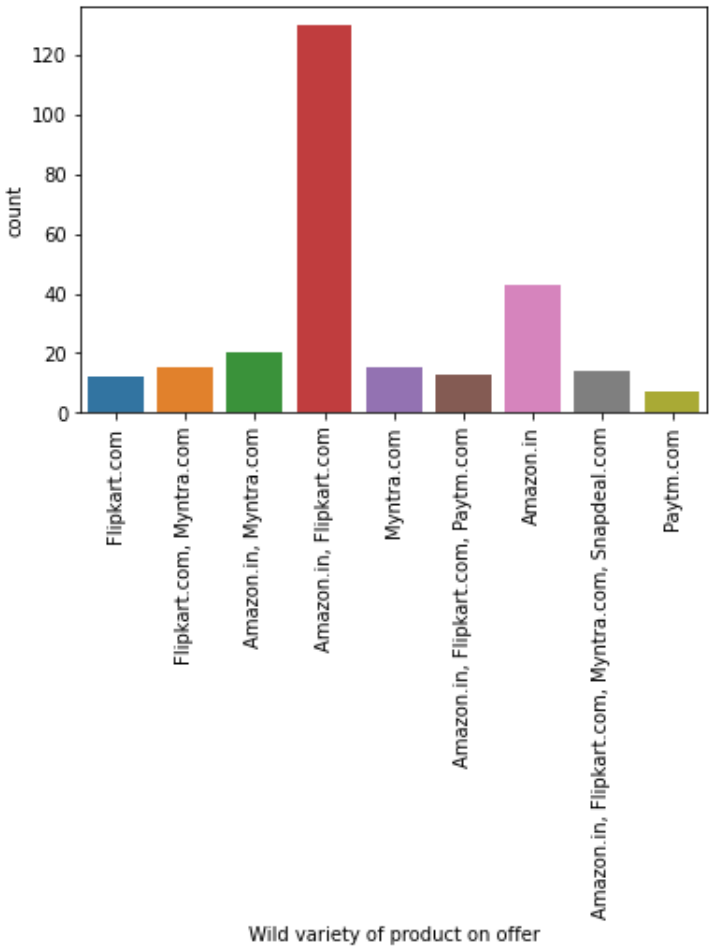
- Observation:

  1. Amazon is the most trustworthy website, followed by Flipkart and Snapdeal.

  2. Myntra and Paytm are least trustworthy.

17. **Wide Variety of Product**

- I plotted the count of the column:



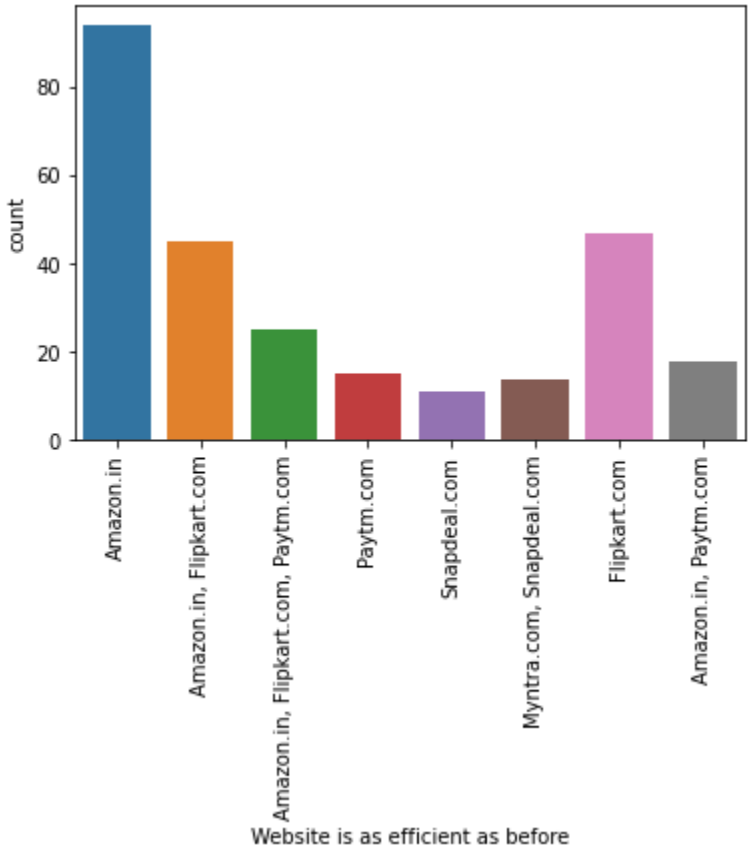
- Observation:

  1. Amazon and Flipkart have the widest variety of products.

  2. Paytm has the lowest variety.

18. **Website as Efficient as before**

- I plotted the count of the column:



- Observation:

  1. Amazon, along with Flipkart is as efficient as it was before.

  2. Customers believe that Snapdeal and Paytm are not as efficient as they used to be.

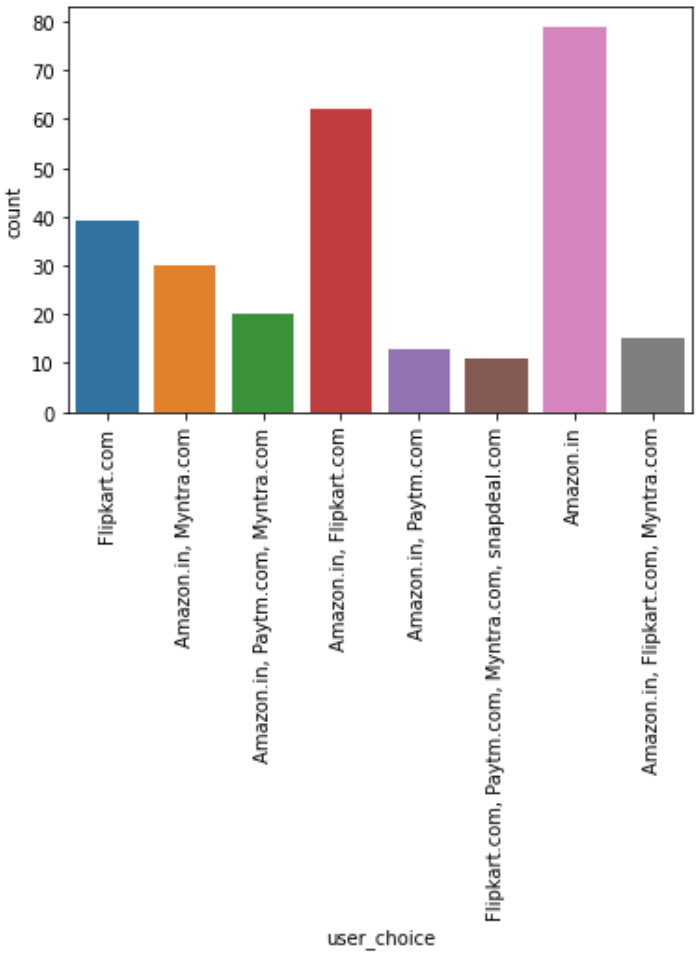
Conclusion

Based on all this information, we can conclude that Amazon is unanimously the most preferred website due to the following factors:

1. Wide range of products
2. Privacy of customer protected
3. Speedy Delivery

4. Trustworthy
5. Efficient as before

This can be proved by looking at this count plot that shows which websites are most preferred by customers:



Flipkart is second most preferred by customers but Amazon is much more ahead of it in terms of recommendations.