

Customer Retention Dataset Project

Report

Problem Statement: We are given a E-retail factors for customer activation and retention: A case study from Indian e-commerce customers and we have to analyse the data and find keys points in retaining customers.

Exploratory Data Analysis:

1) Shape of data:

```
In [74]: 1 df.shape
```

Out[74]: (269, 71)

We have a small data of 269 rows and 71 columns.

2) Columns Present:

```
In [7]: 1 df.columns
```

[illegible]

3) Nulls Present:

Checking nulls

```
In [75]: 1 pd.set_option('display.max_rows', 500)
          2 df.isna().sum()
```

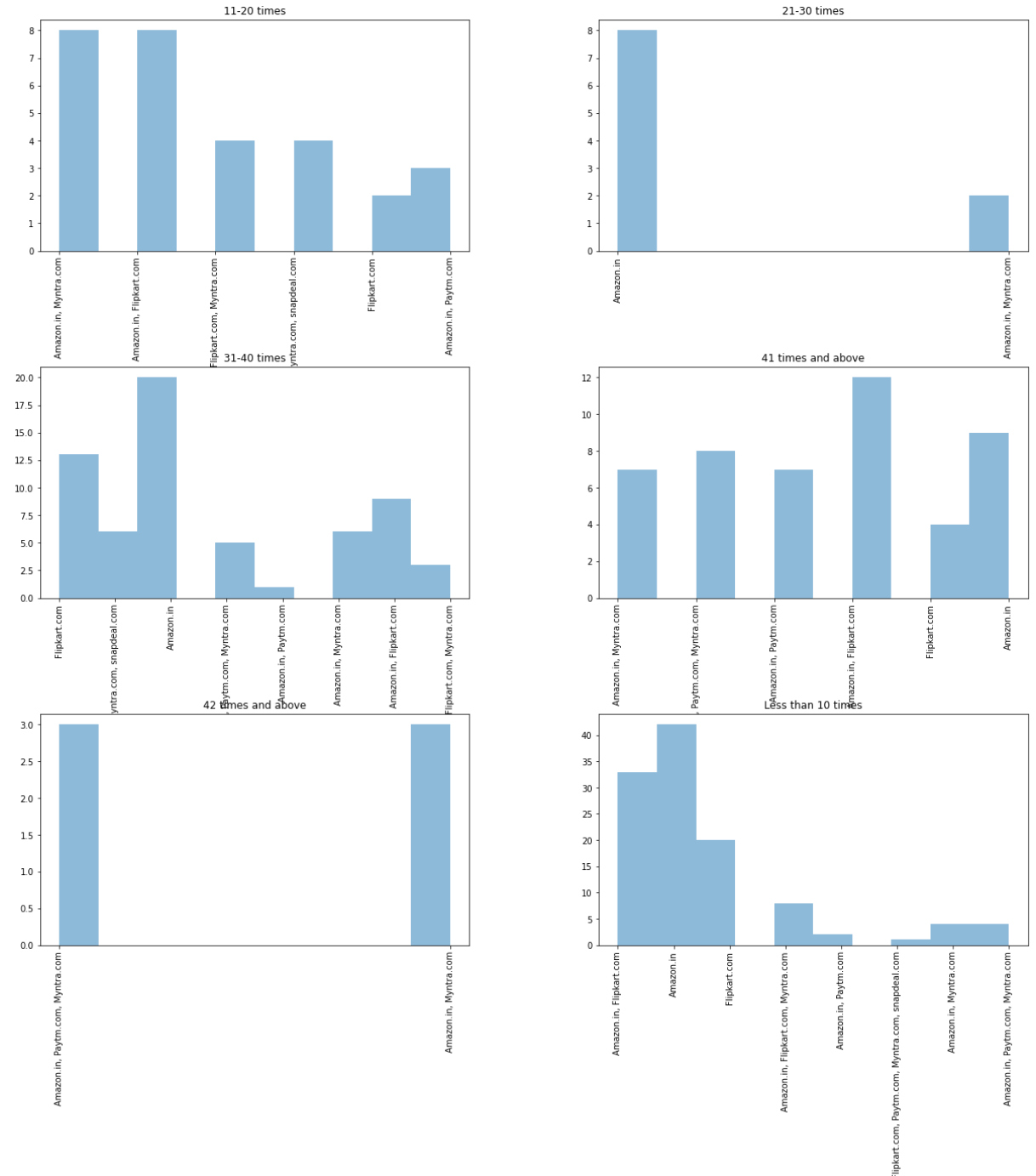
[illegible]

No nulls present.

Dataset has no null values.

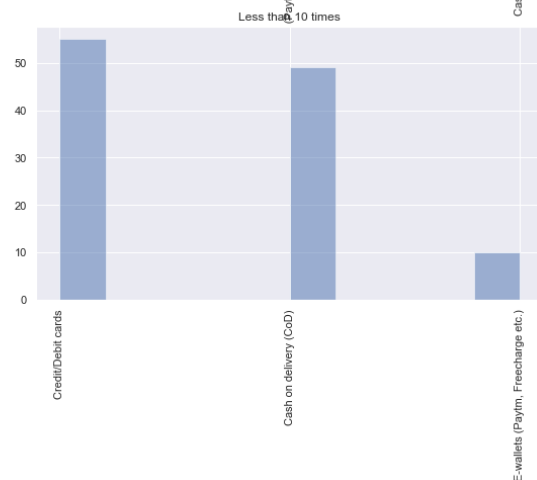
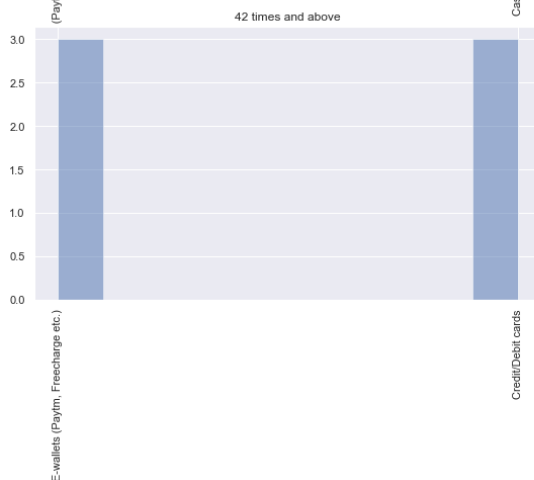
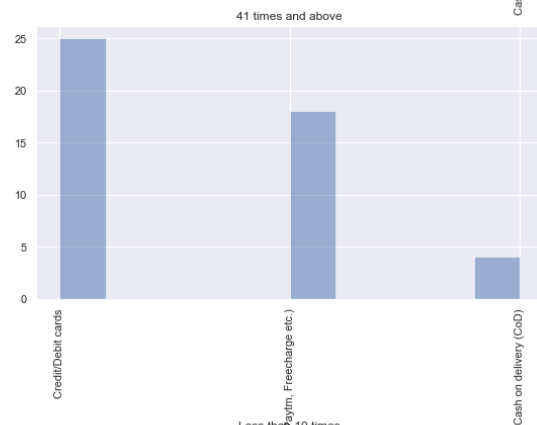
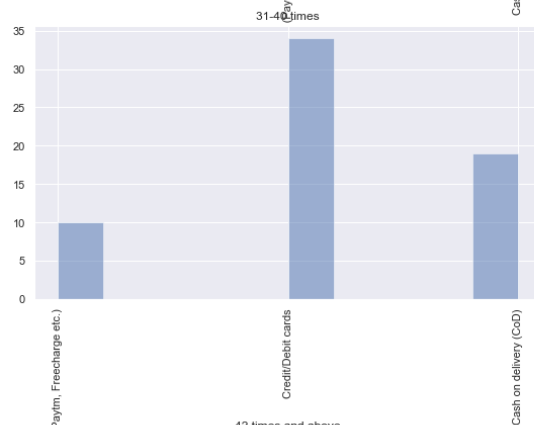
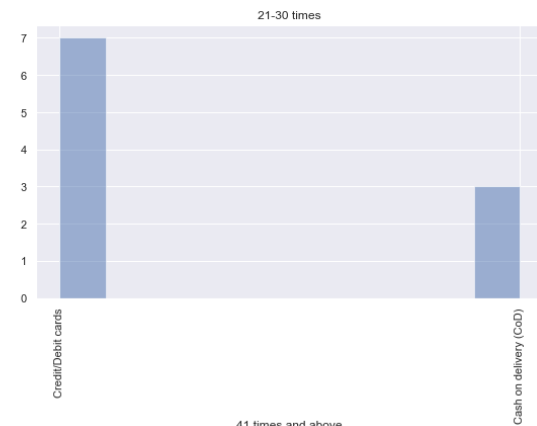
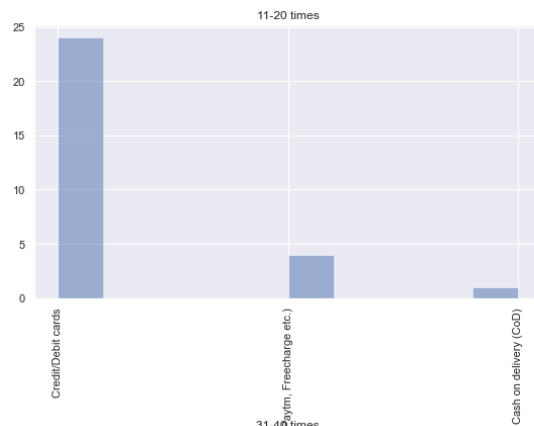
4) Analysing relationship between features:

- i) Between Which of the Indian online retailer would you recommend to a friend? And 6 How many times you have made an online purchase in the past 1 year?



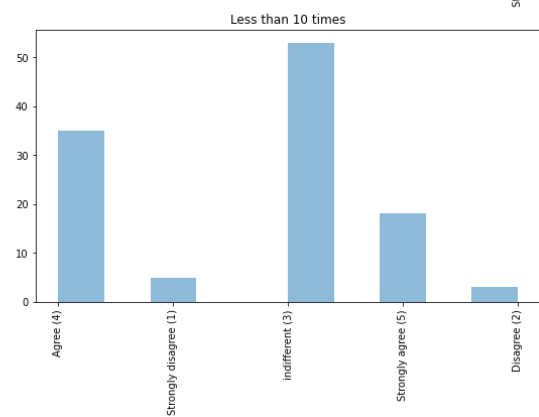
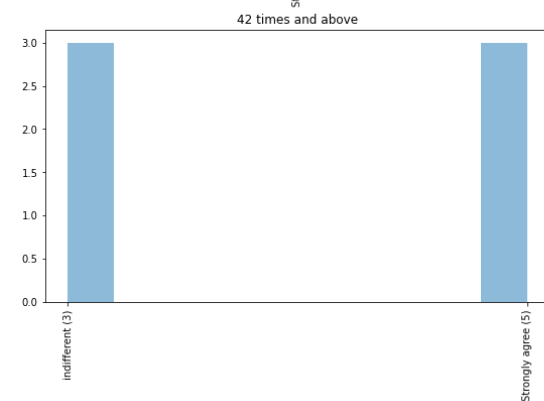
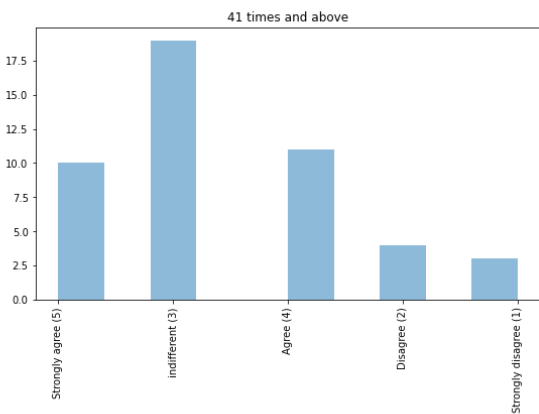
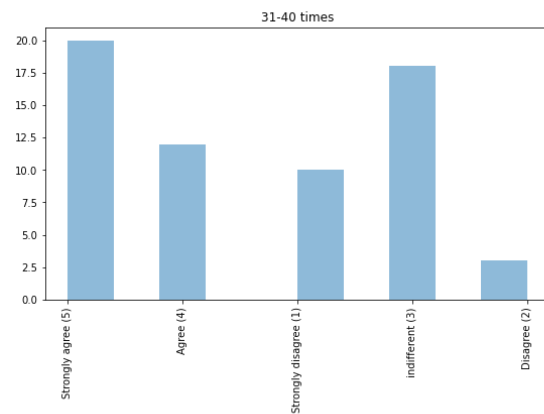
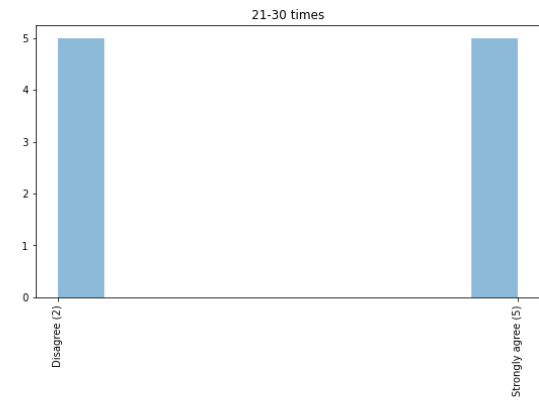
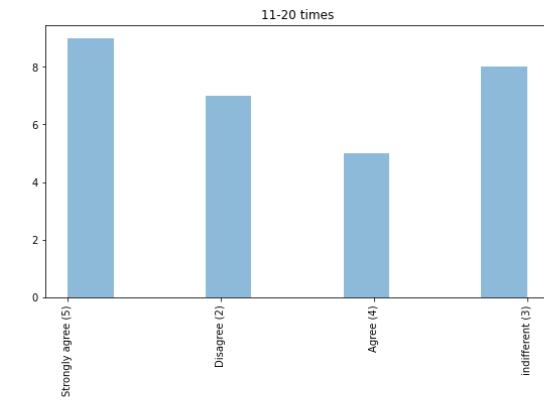
Observation: Irrespective of number of times purchase made online the most recommended website being flipkart and amazon.

ii) Between 15 What is your preferred payment Option? And How many times you have made an online purchase in the past 1 year?



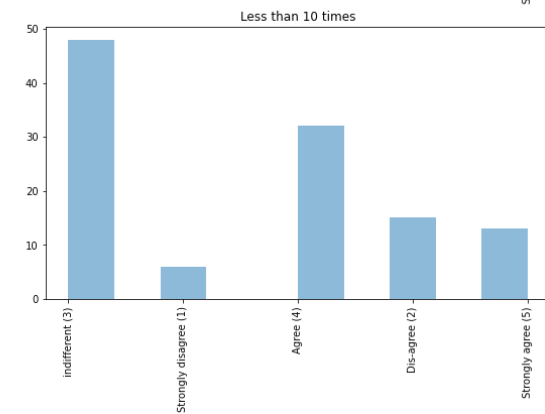
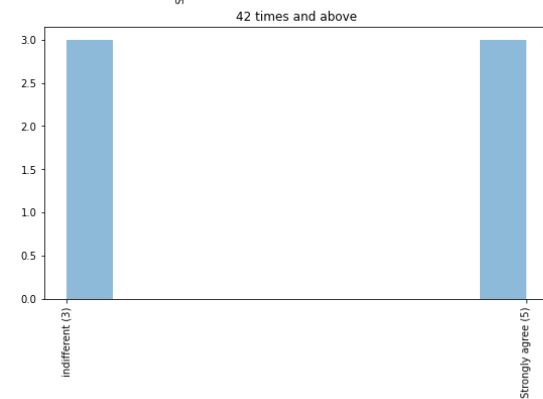
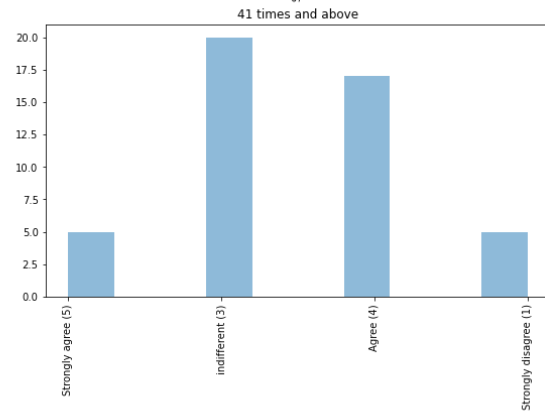
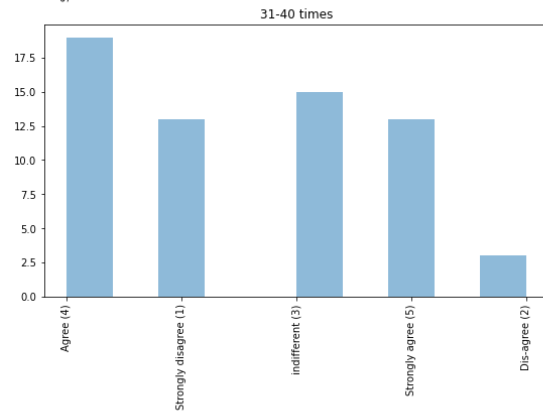
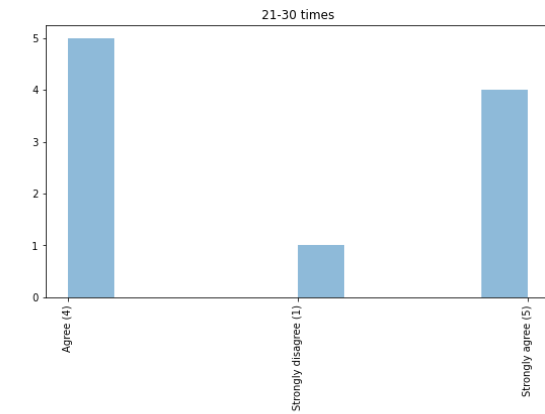
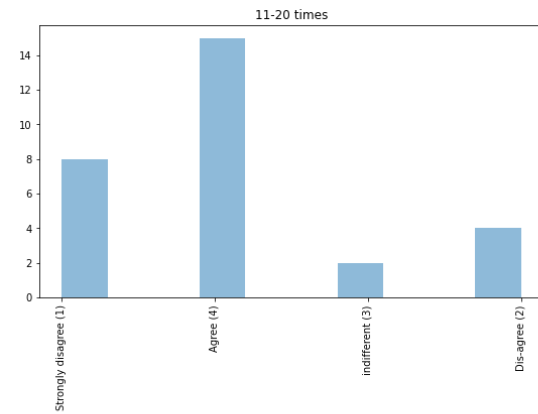
Observation: We could observe that as the number of times customer do online shopping increases, they pay using credit/debit card or E-wallets whereas if the number of times shopping less then COD option is used more. That means once customers gain confidence in that website, they use prepaid options.

- iii) Between You feel gratification shopping on your favorite e-tailer and How many times you have made an online purchase in the past 1 year?



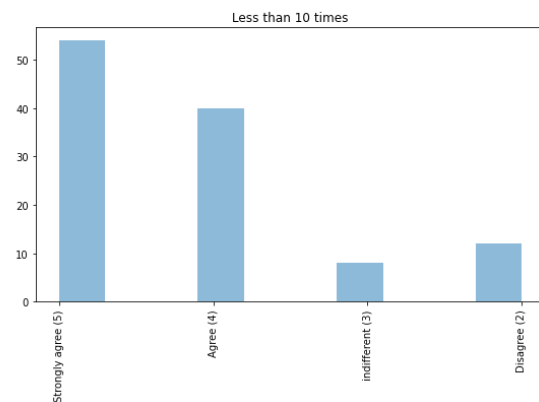
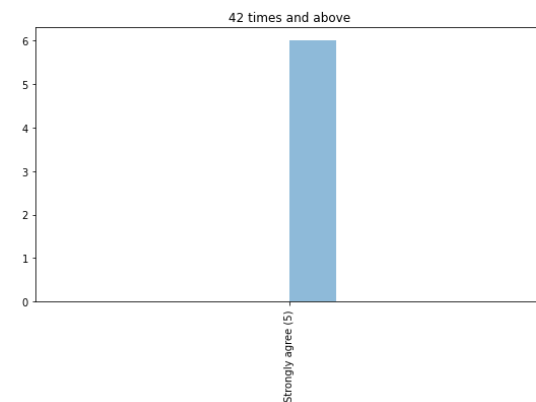
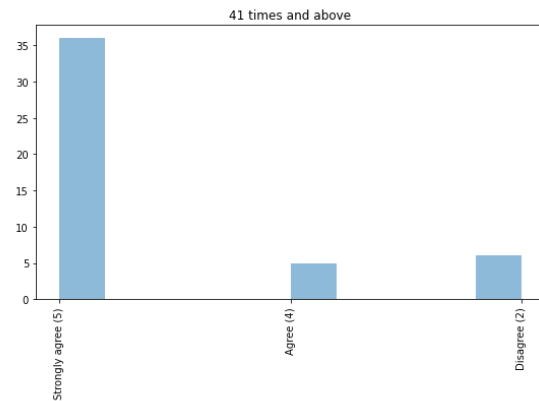
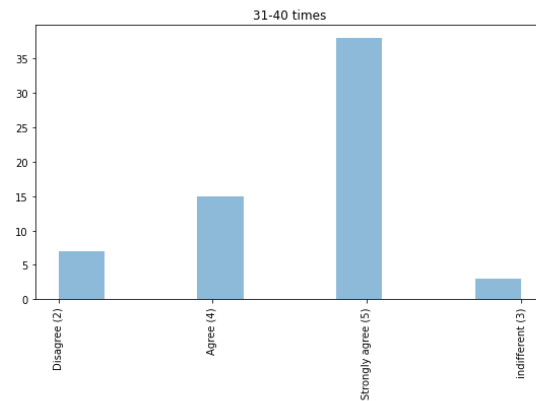
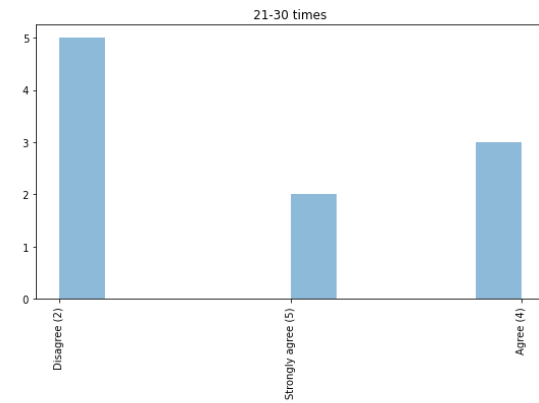
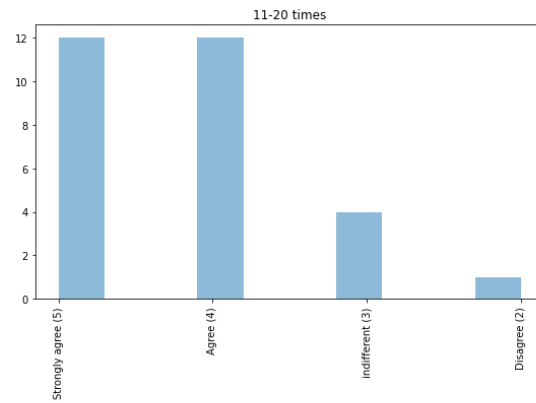
Observation: Most of the time customer is indifferent to feeling gratified.

- iv) Shopping on the website helps you fulfill certain roles and How many times you have made an online purchase in the past 1 year?



Observation: Most of the time customer is either indifferent or strongly agree that Shopping on the website helps fulfilling certain roles.

- v) Monetary savings and How many times you have made an online purchase in the past 1 year?



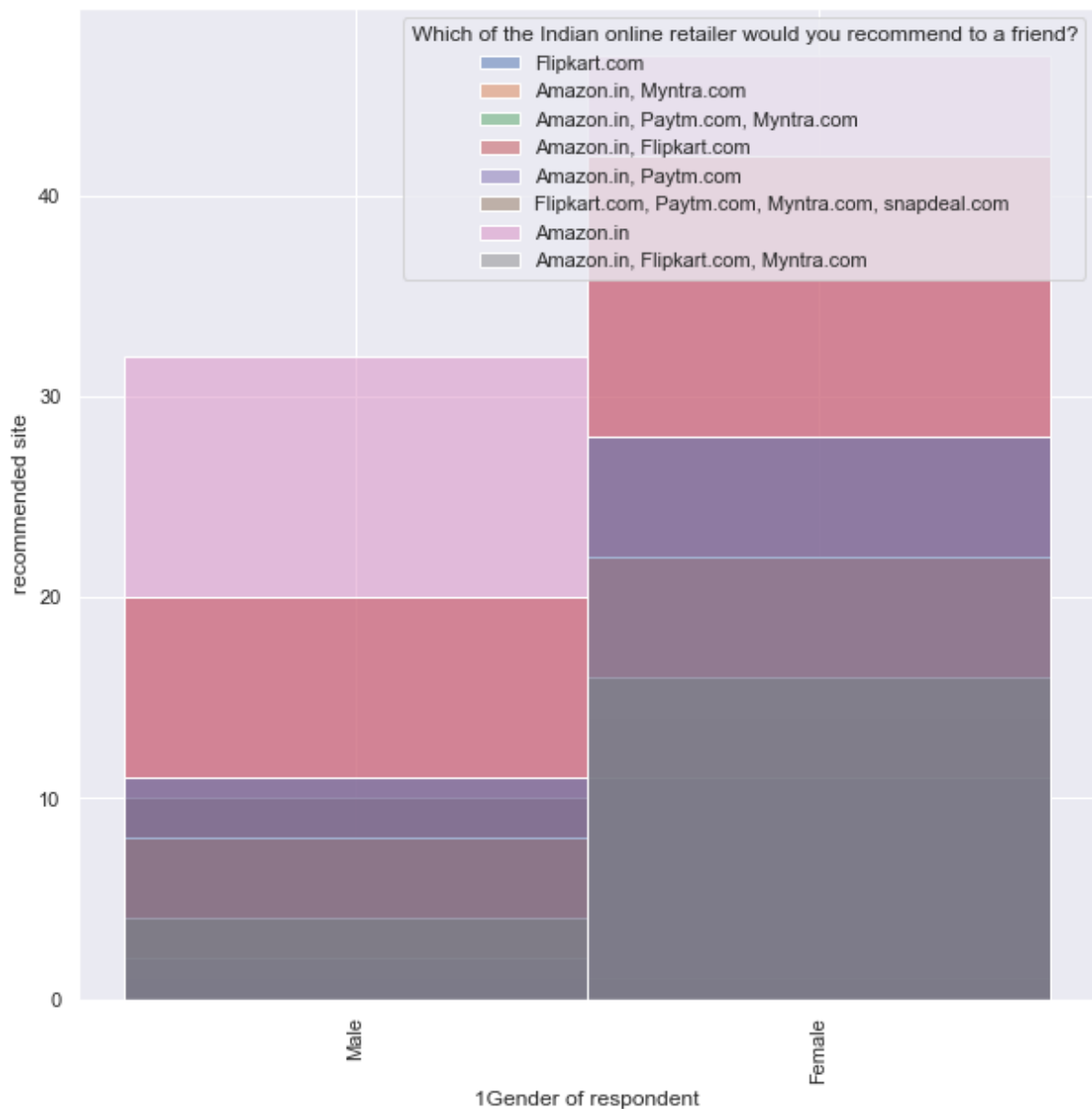
Observation: If the number of purchases is between 20-30 times, maximum customers disagree of monetary savings and in rest customers strongly agree of monetary savings.

5) Analysing Relationship between features and labels. Considering label as “Which of the Indian online retailer would you recommend to a friend?”

```
In [78]: 1 col=df.columns
          2 for i in col:
          3     plt.figure(figsize=(10,3))
          4     sns.histplot(x=i,hue='Which of the Indian online retailer would you recommend to a friend?',data=df)
          5     plt.xlabel(i)
          6     plt.xticks(rotation=90)
          7     plt.ylabel("recommended site")
          8     plt.show()
```

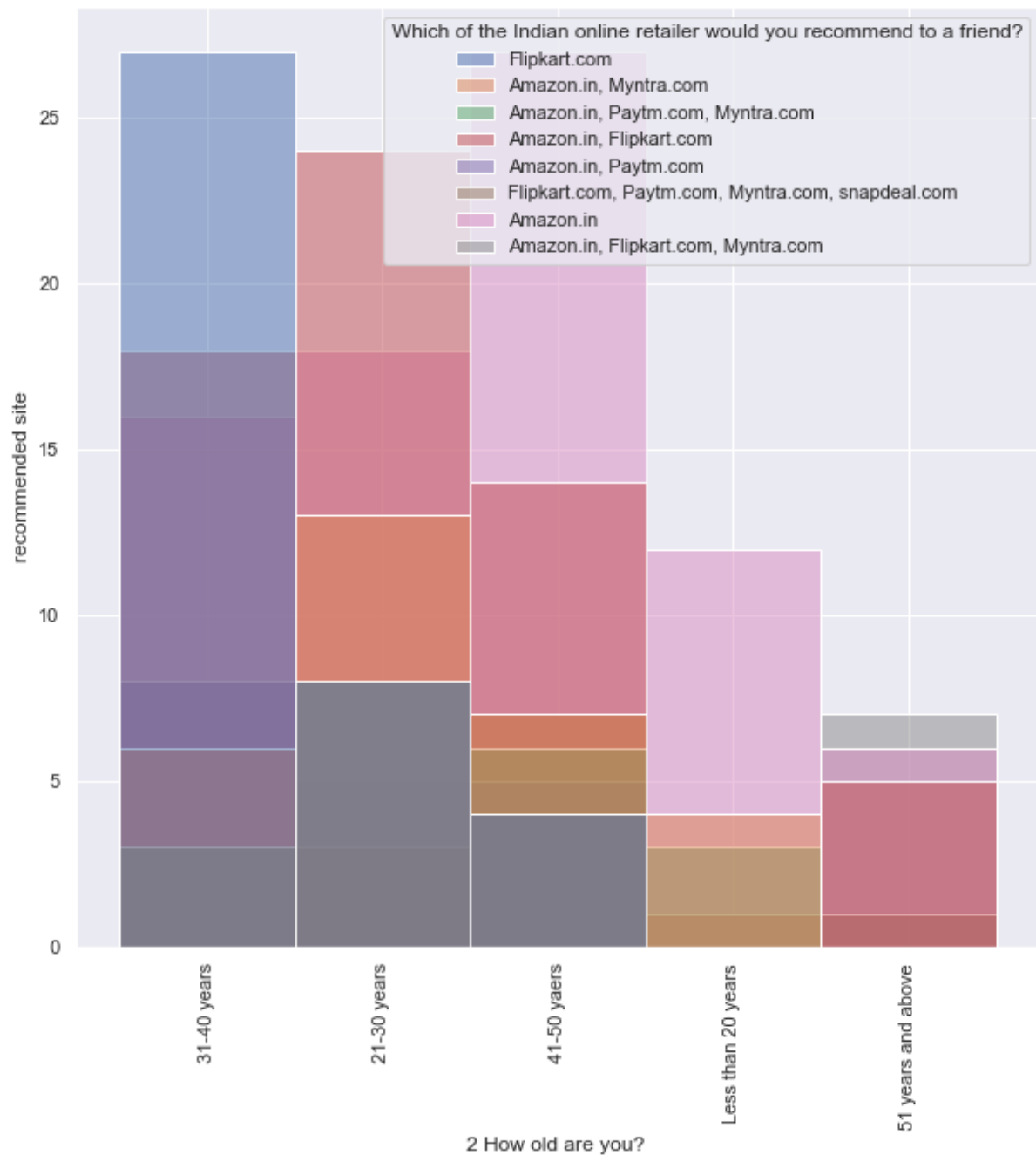
Output:

1) Gender:



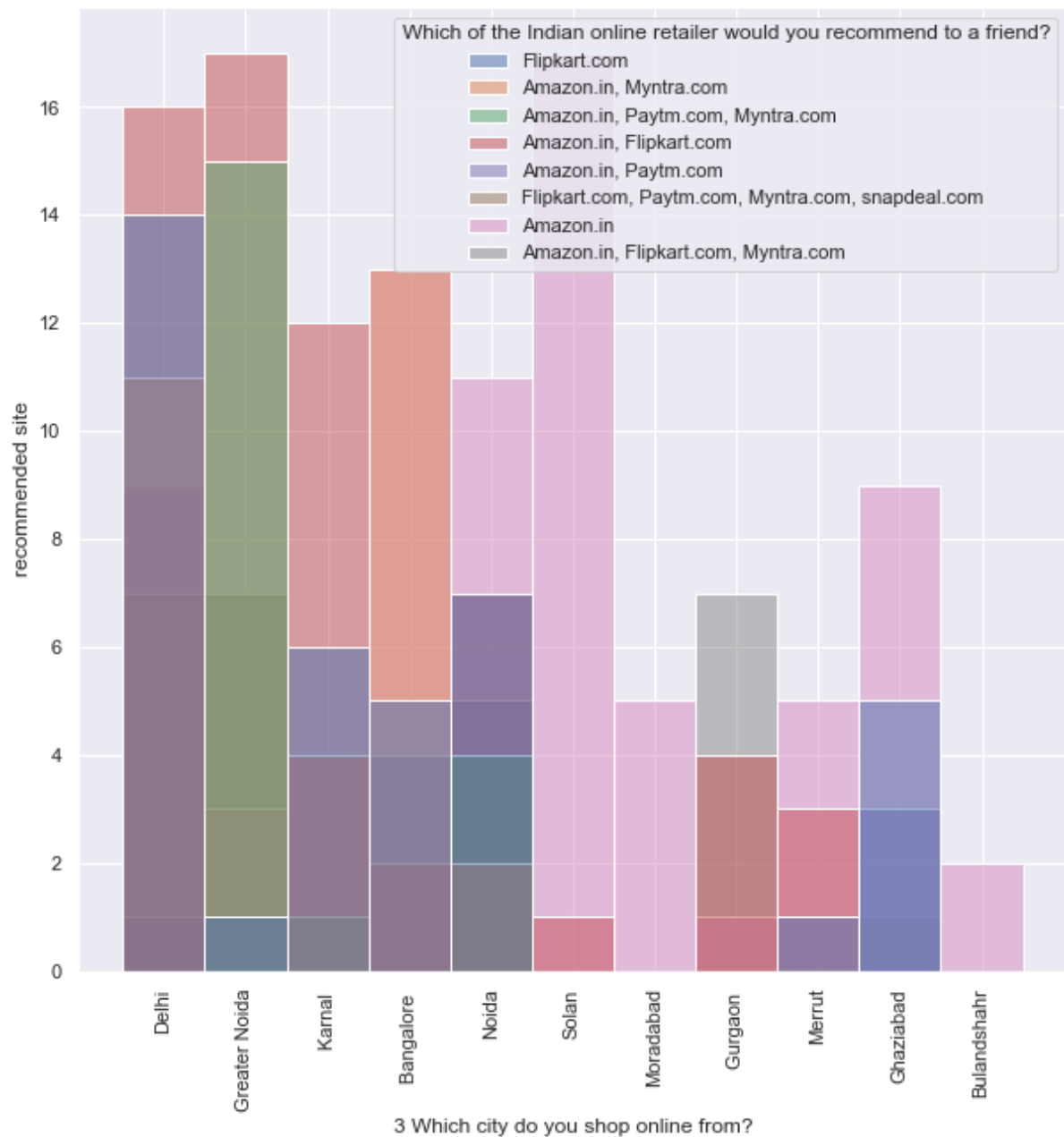
Observation: More number of female buyers as compared to males and maximum number of customers recommend either Amazon or Flipkart.

2) Age:



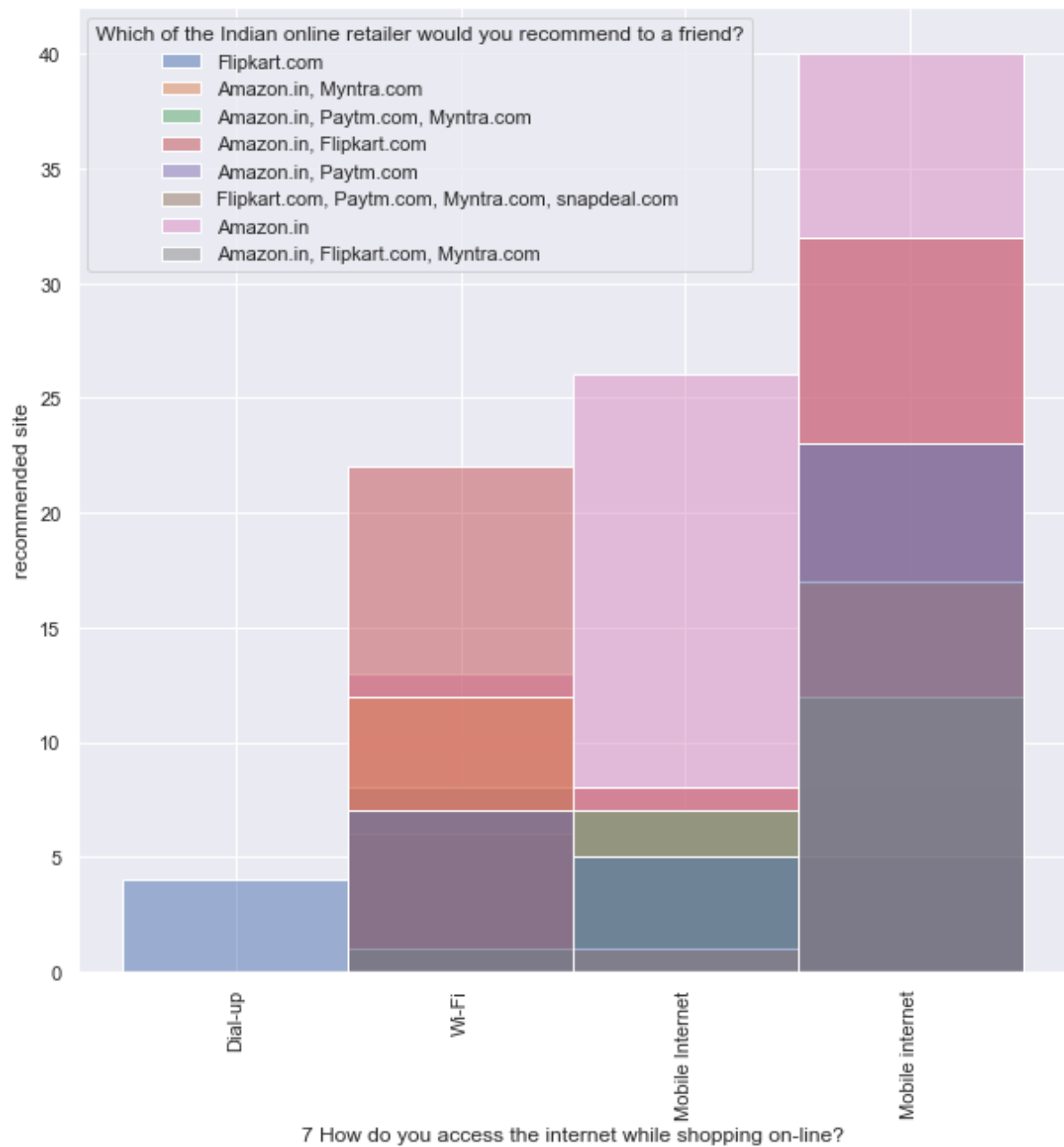
Observation: Maximum number of buyers are in the age between 31 to 40 or 41 to 50. The most recommended website from both the age groups are Flipkart and Amazon respectively.

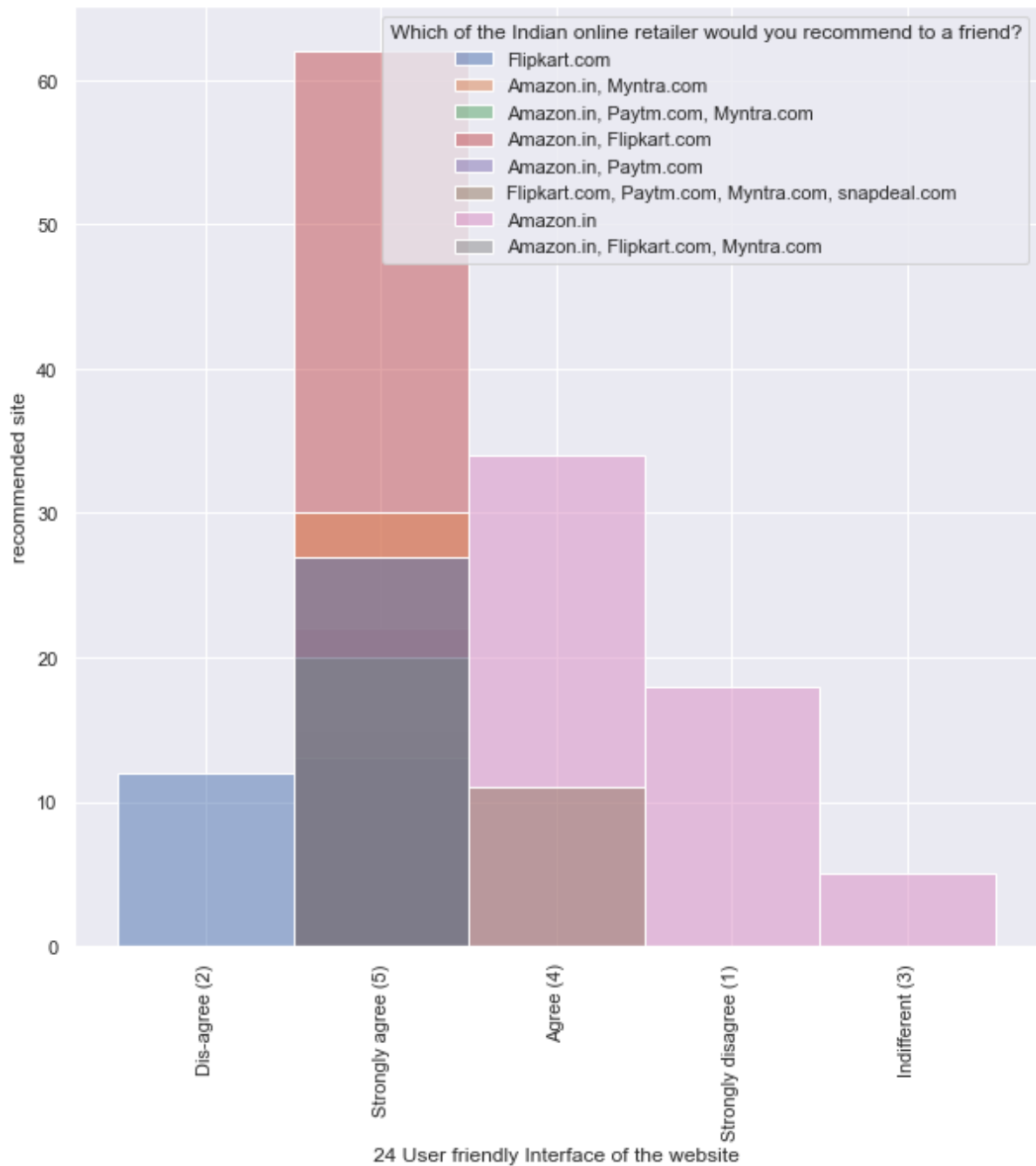
3) Location:



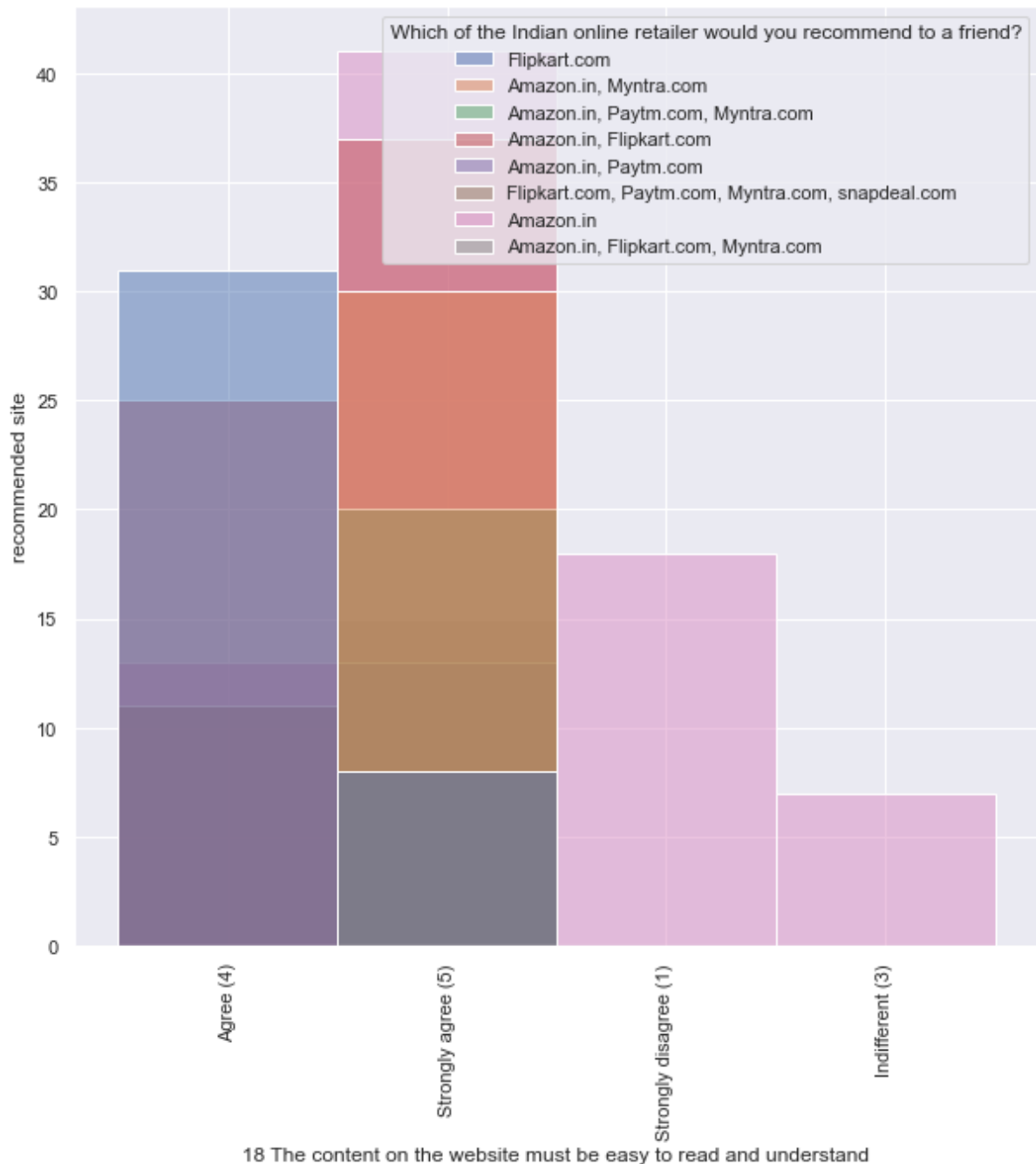
Observation: Most shopped city is Greater Noida, Solan and Delhi and buyers there recommend Amazon and Flipkart the most.

4) User friendliness of website/application:

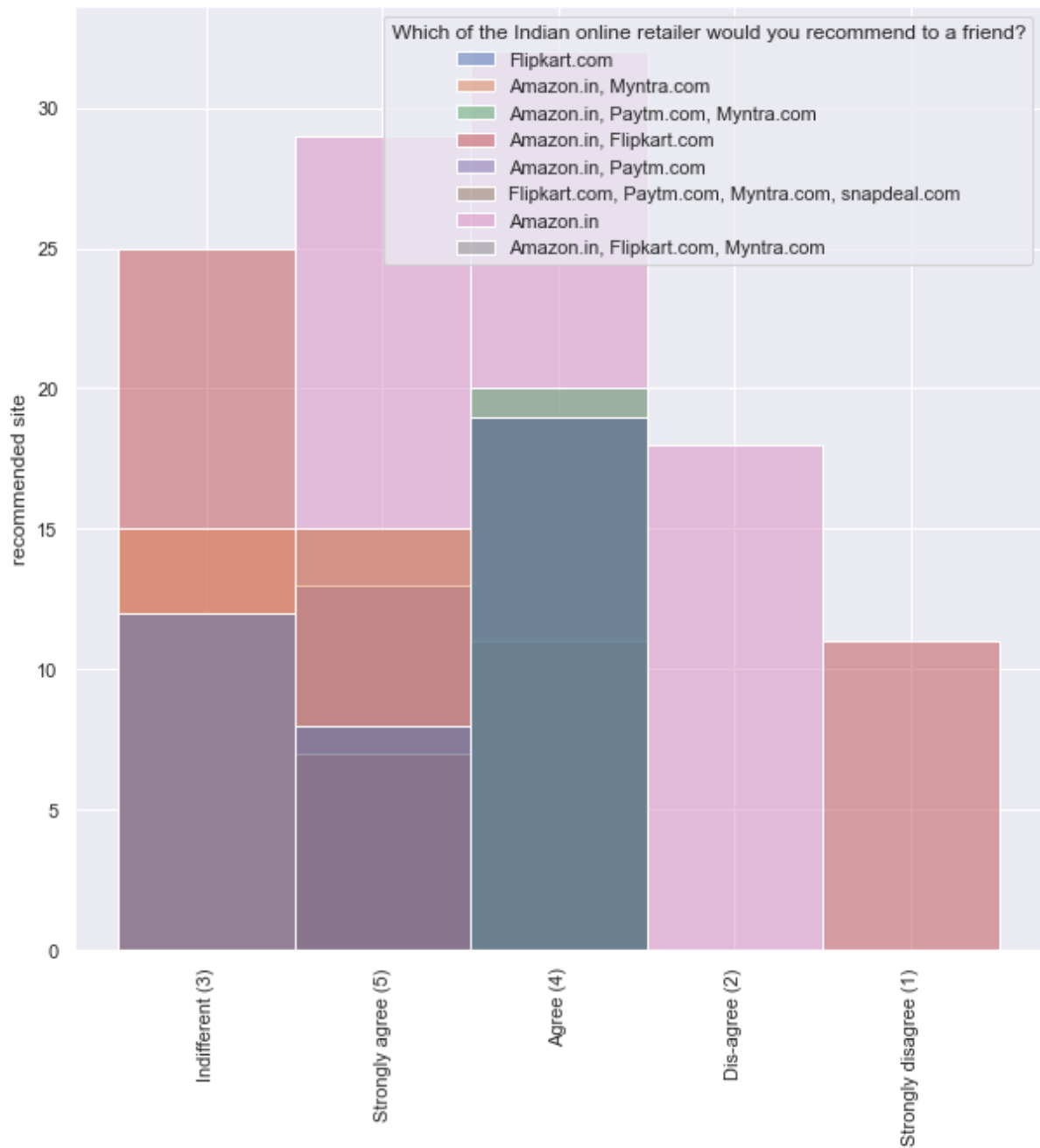




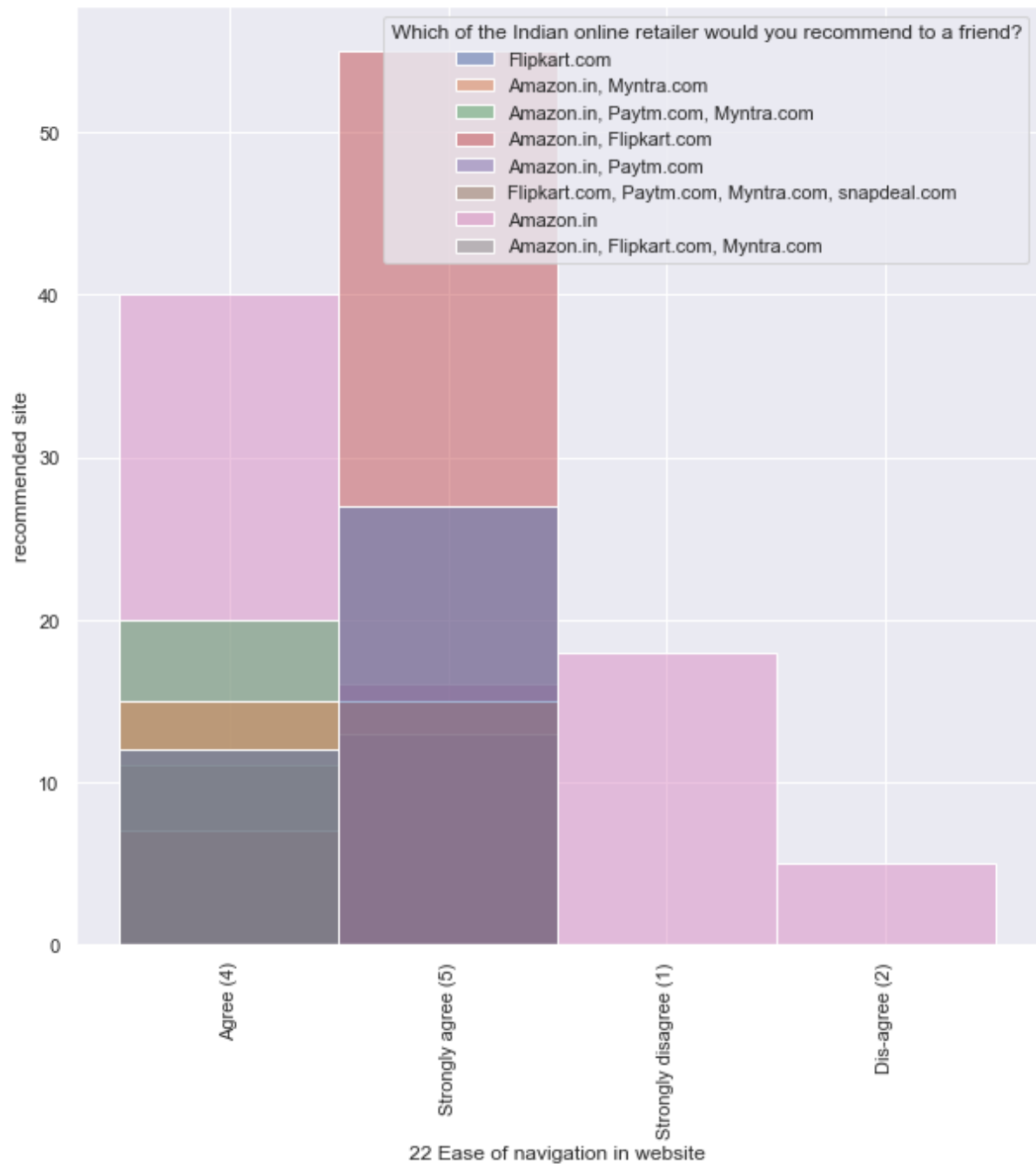
Observation: Most of the times the orders are placed using mobile phone thus websites has to be mobile friendly. Here also the most recommended websites are Amazon and Flipkart that means these provide good mobile screen and user friendly website and app.

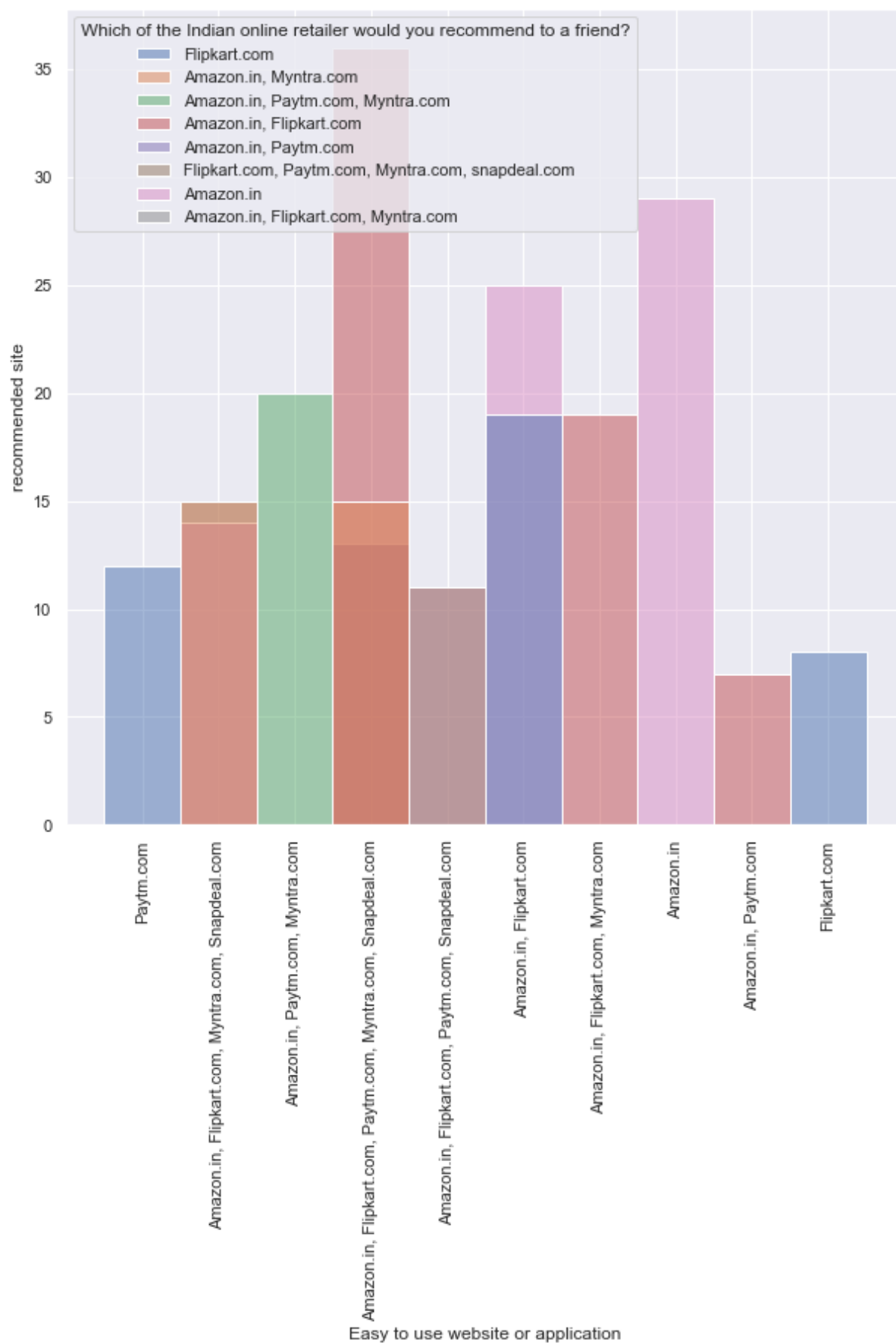


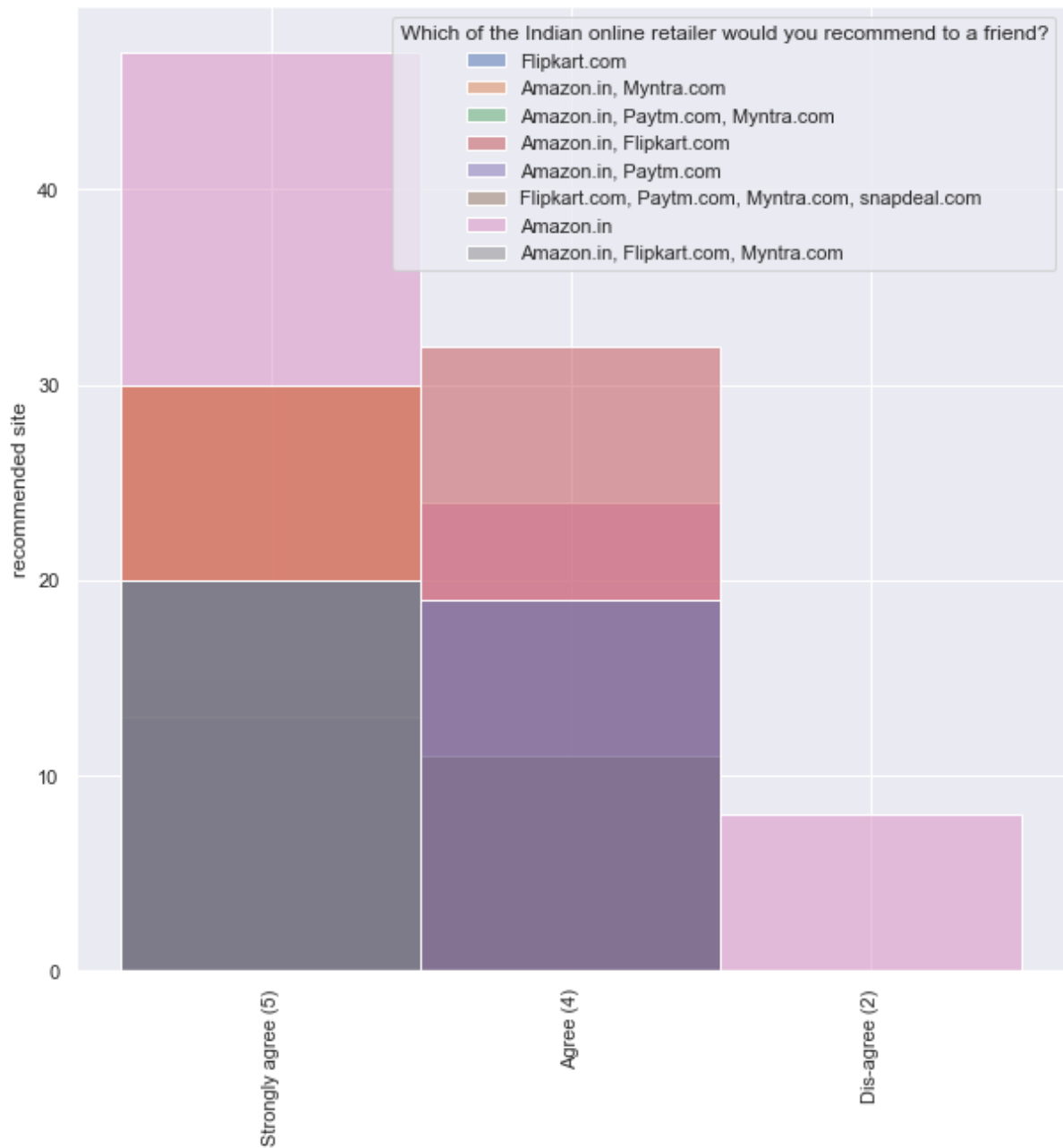
Observation: Whatever we observed in the above graph can be cross checked from this one. People strongly agree that the Amazon and Flipkart website content is easy to understand and read thus they prefer shopping from here.



Observation: Maximum people agree and strongly agree that the complete seller information display is important for purchase decision and thus they strongly recommend amazon and flipkart because of same feature.

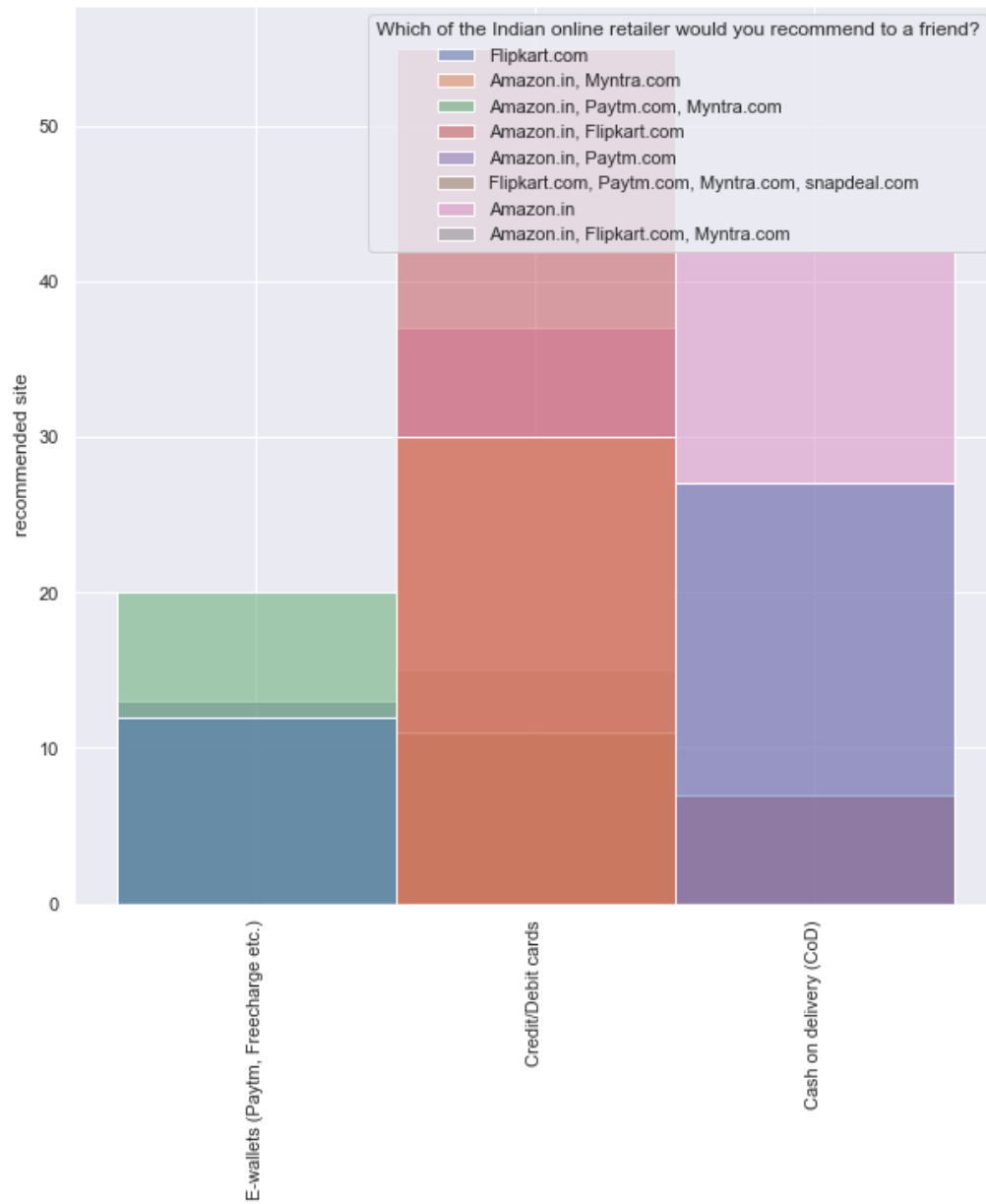


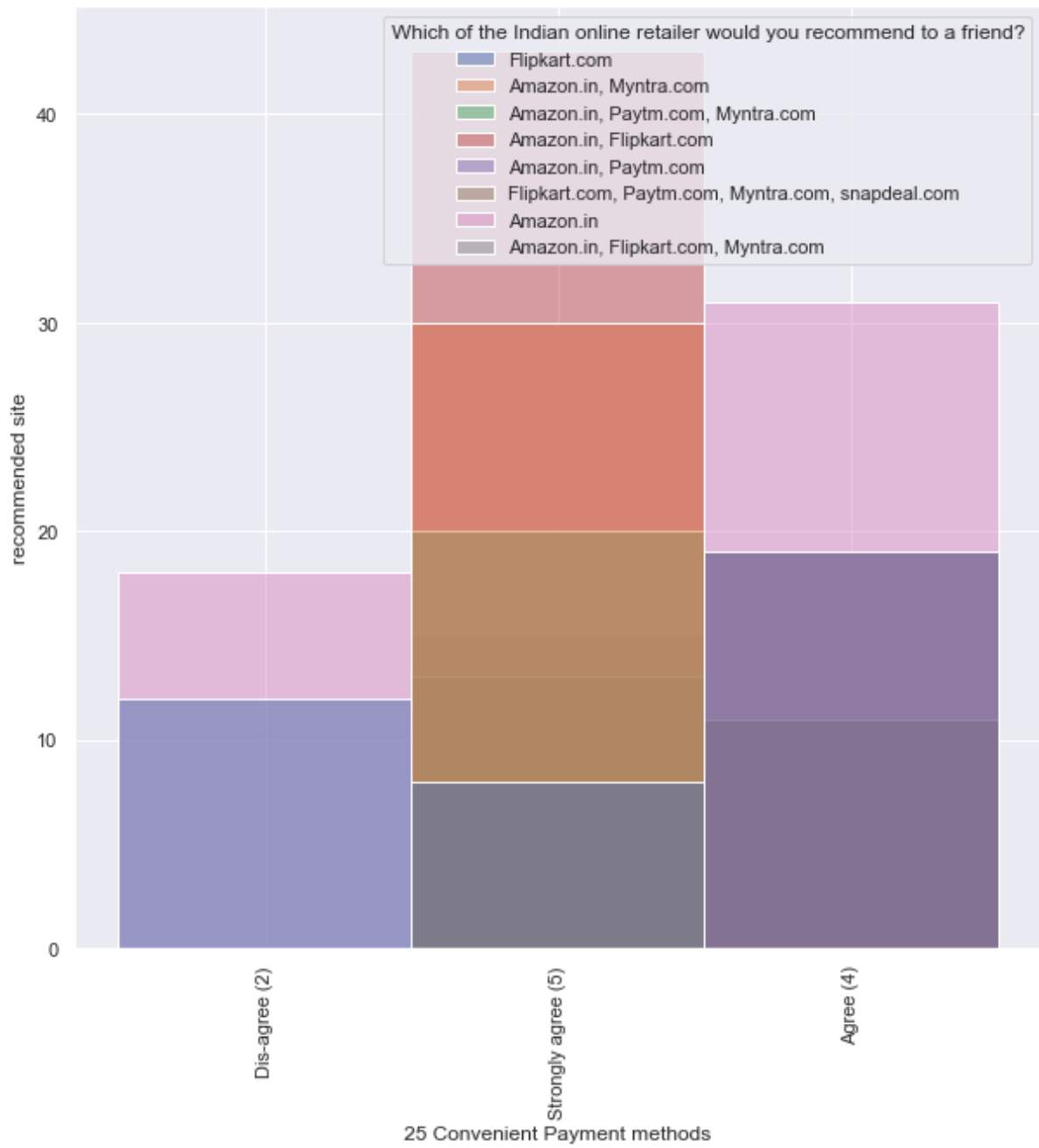


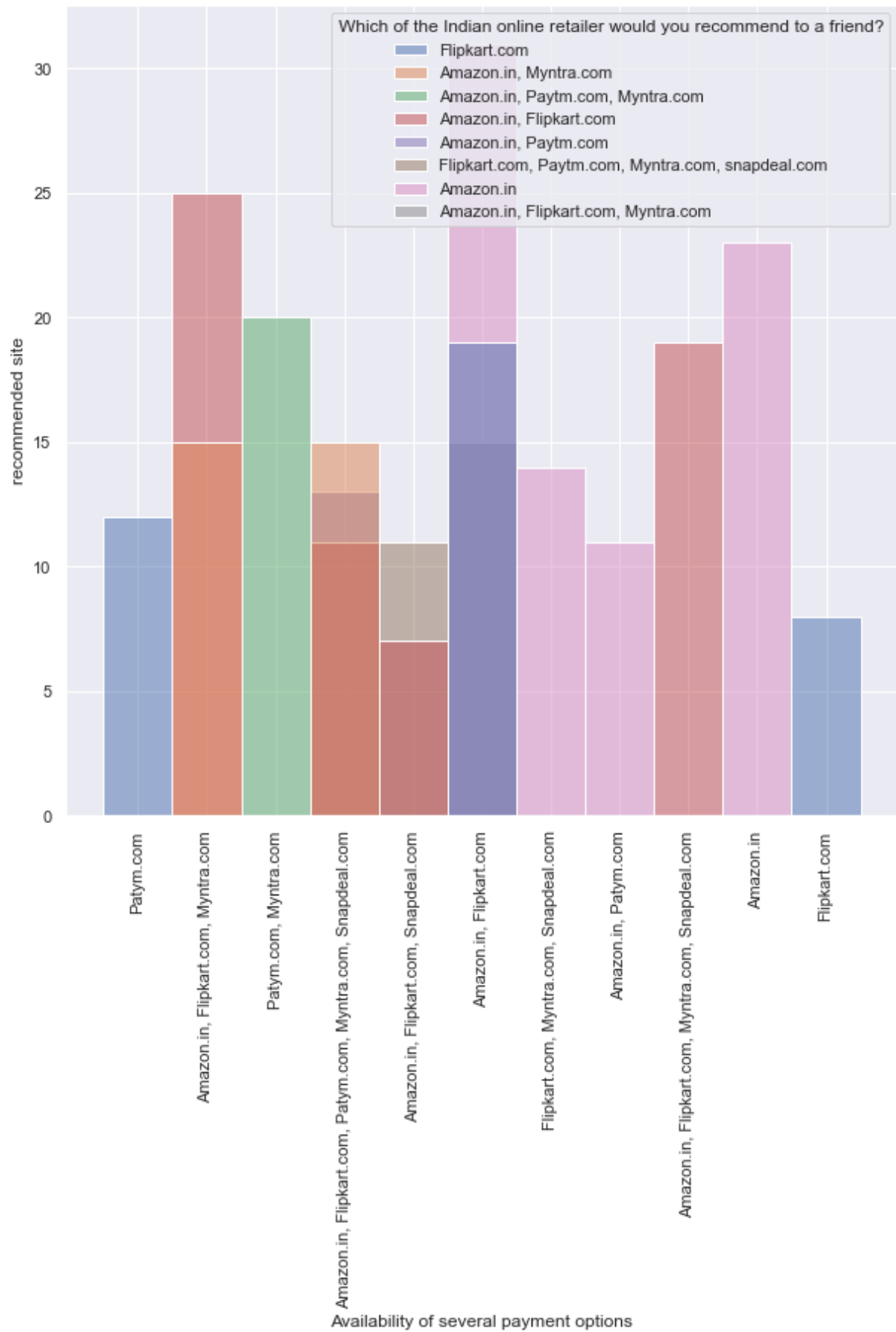


Observation: From the above graphs we can conclude that those websites are strongly recommended which provide ease of navigation in website and are of good quality.

5) Payment :

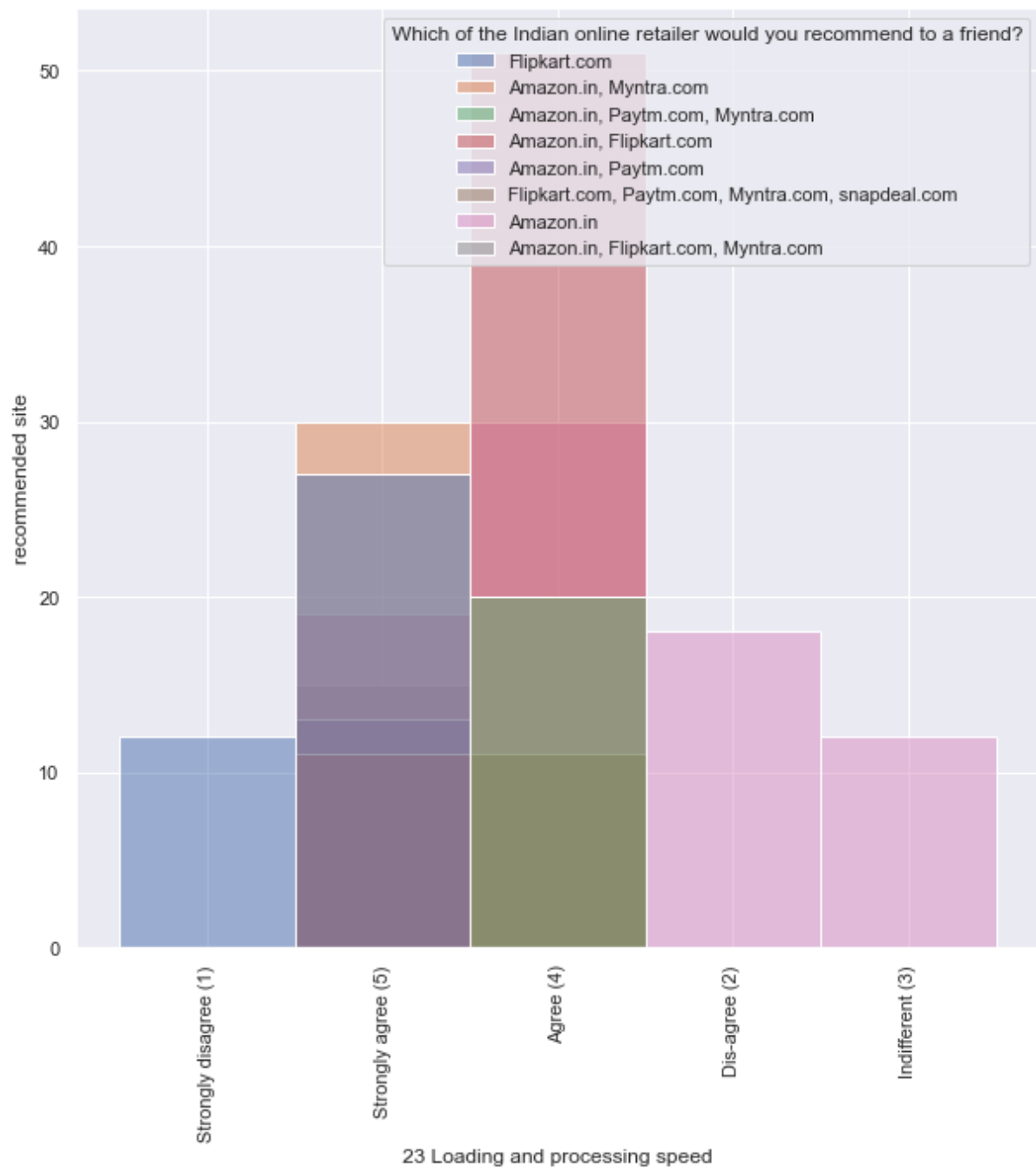


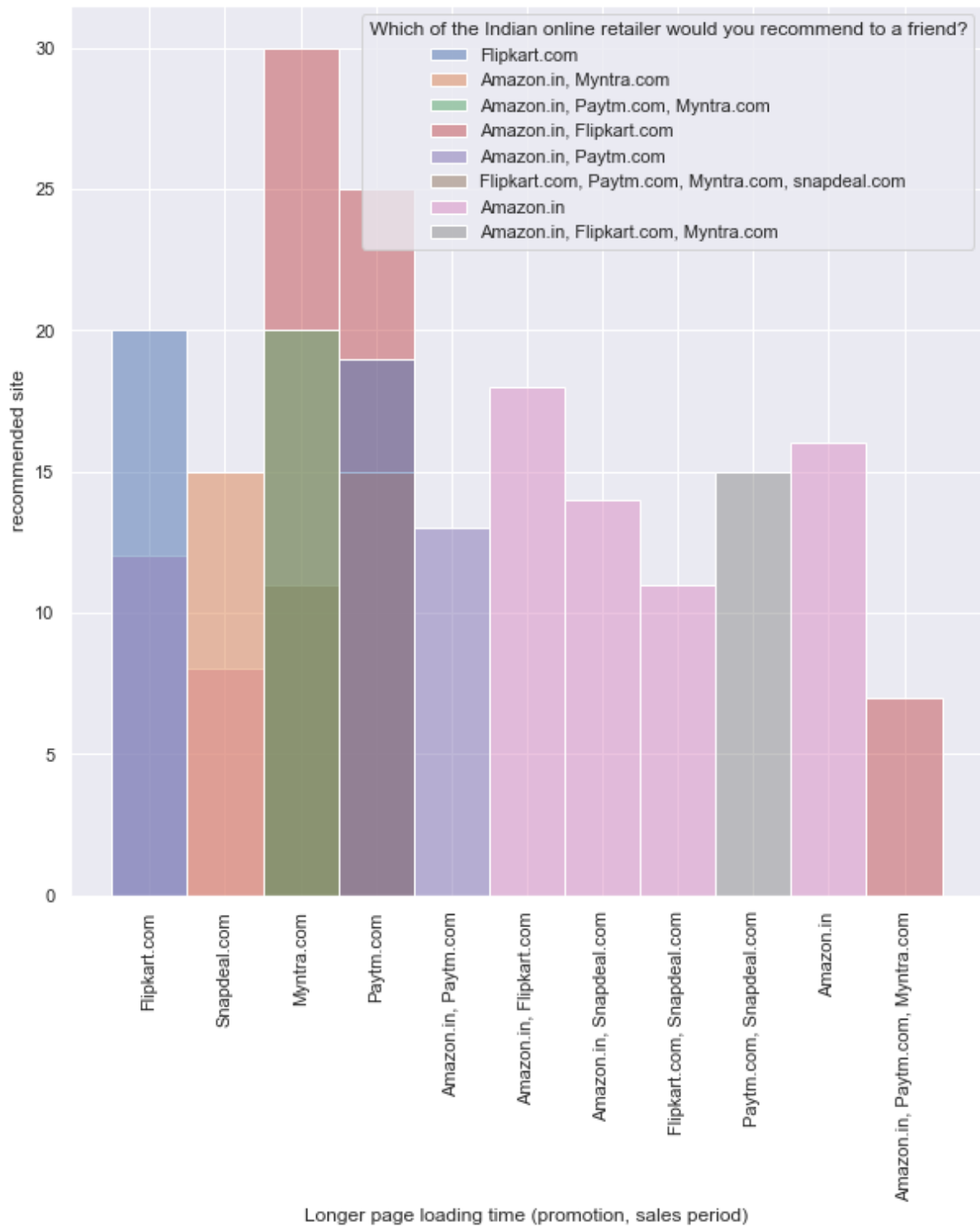




Observation: From the above graphs we can observe that mode of payment that is used most number of times is through credit/debit card and E-wallets are used least. Most of the convenient payment methods are provided by amazon and flipkart and so one of the most recommended websites. These websites also has availability of lot many payment options and so its easy for all section of customers to shop from here.

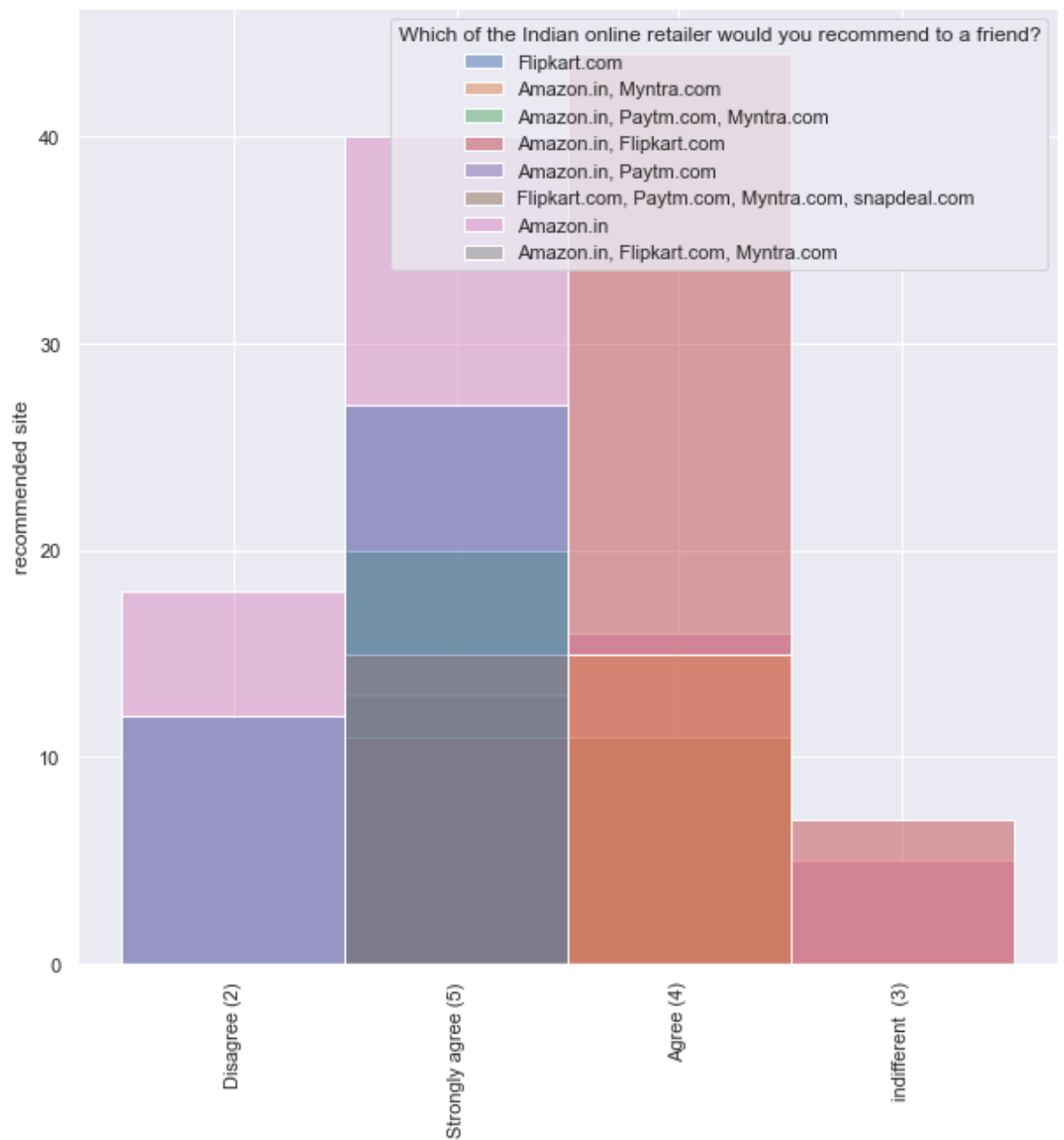
6) Loading and Processing speed:



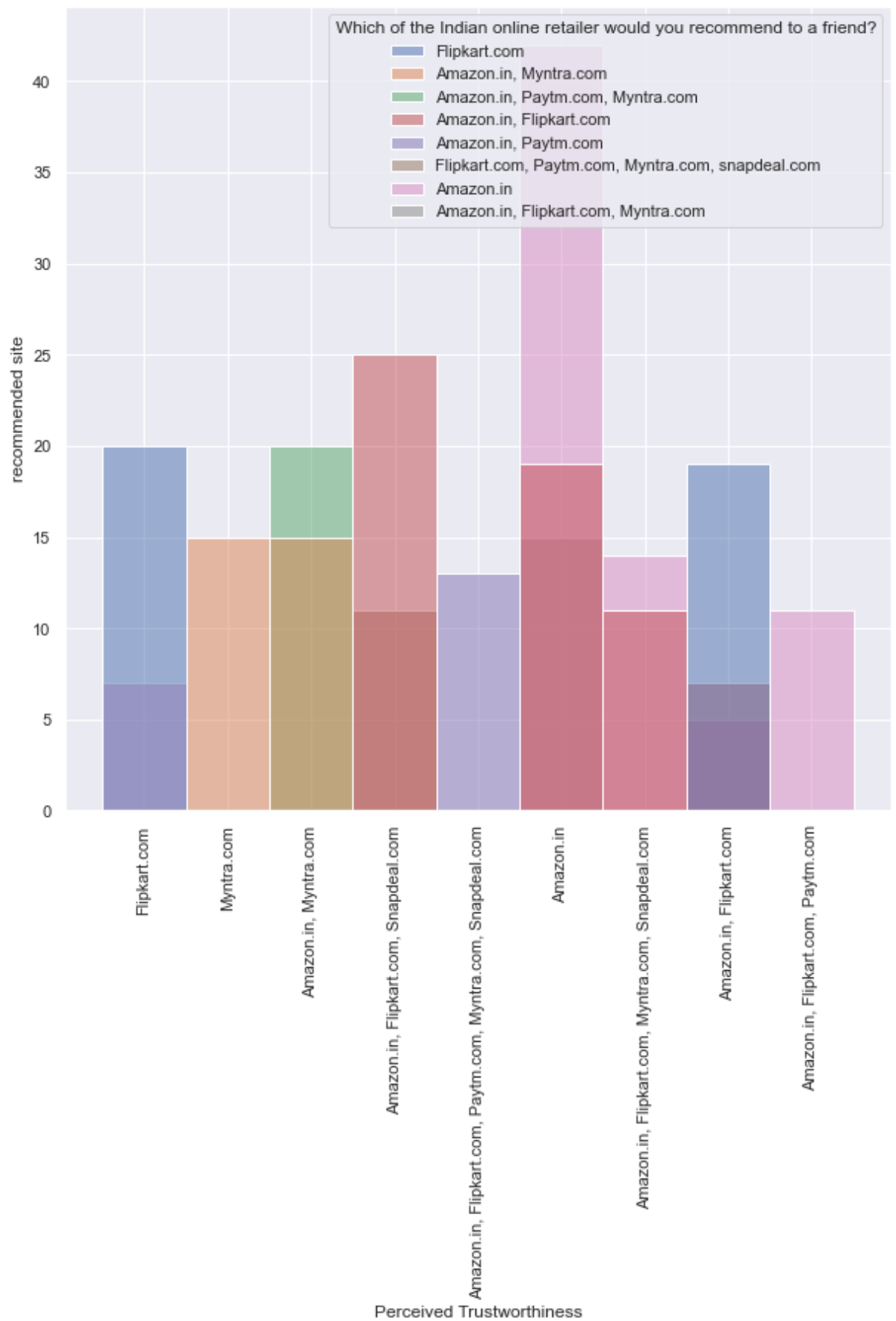


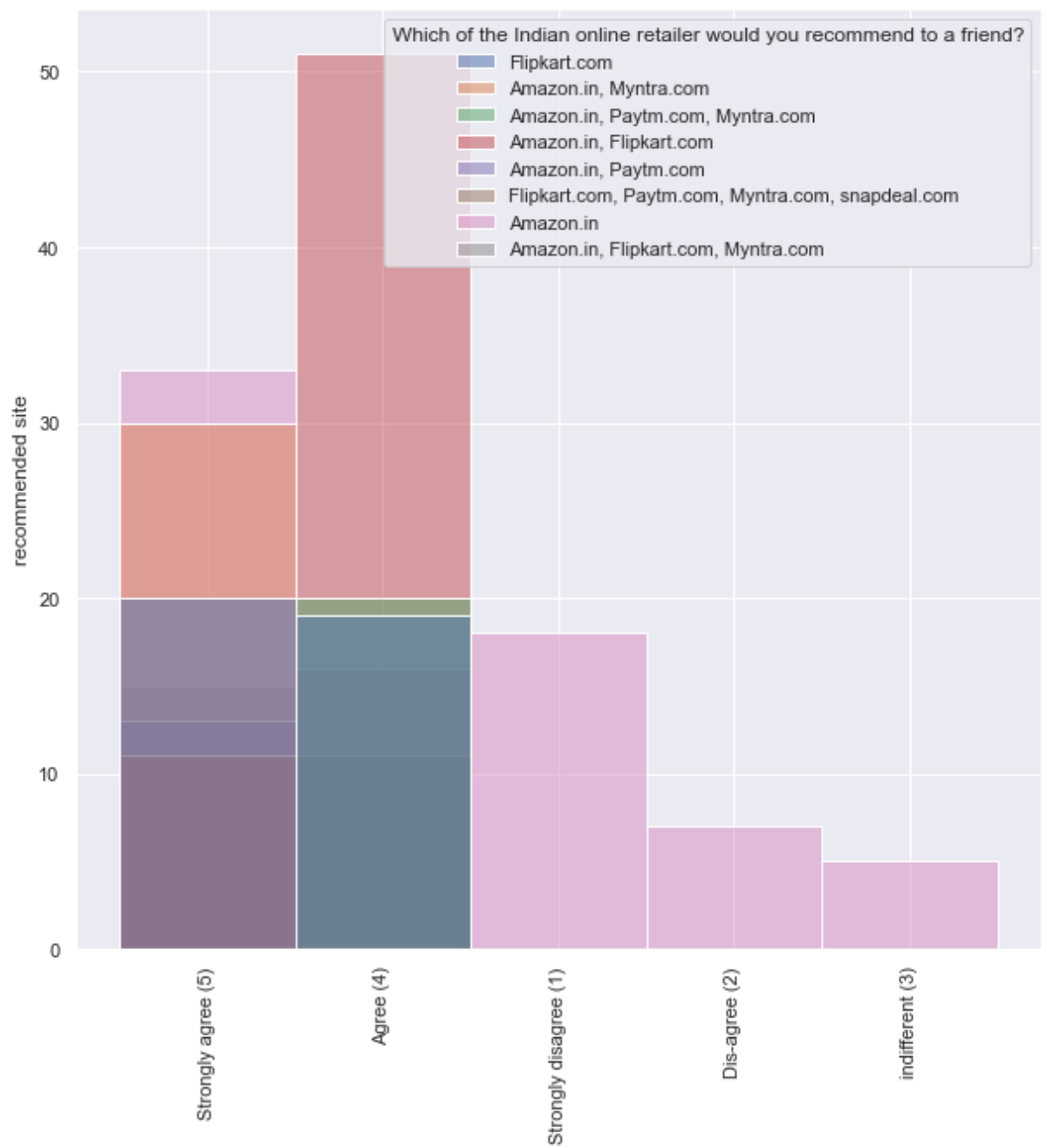
Observation: Loading and processing speed is best for amazon and flipkart thus most of the people recommend this and its an important aspect in doing online shopping with ease. Longer page loading time is of myntra and paytm whereas it is least for amazon.

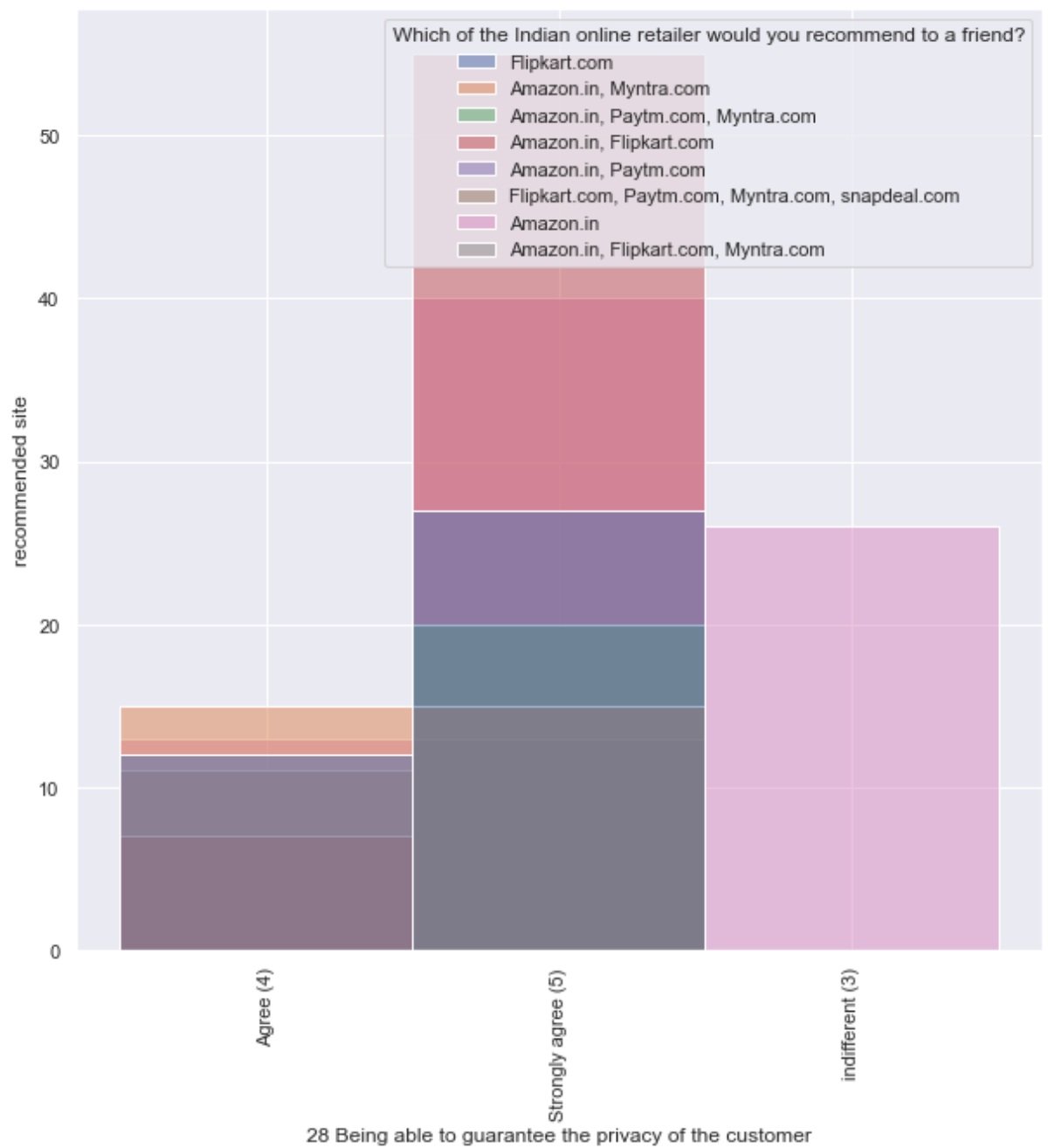
7) Trust and Privacy:

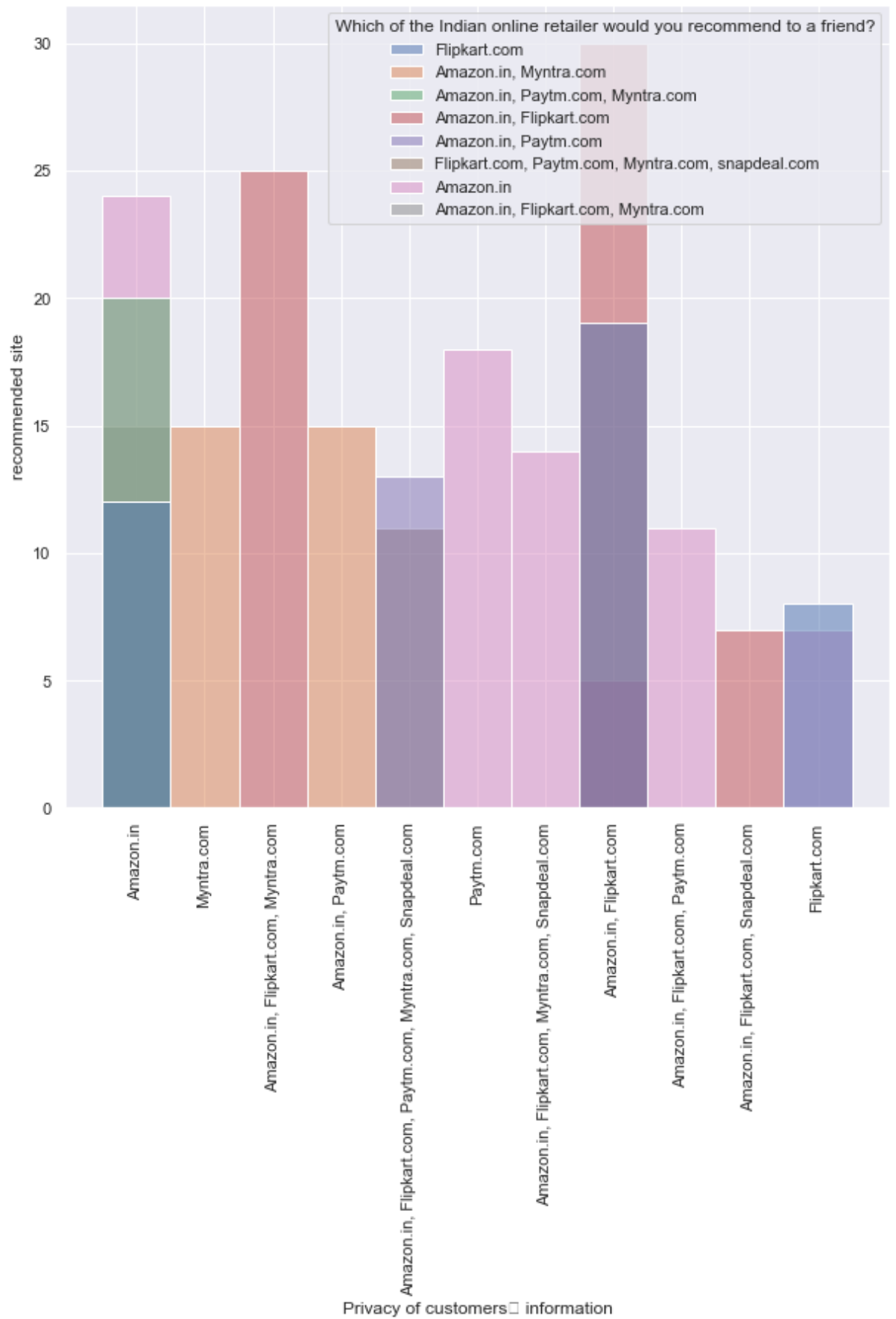


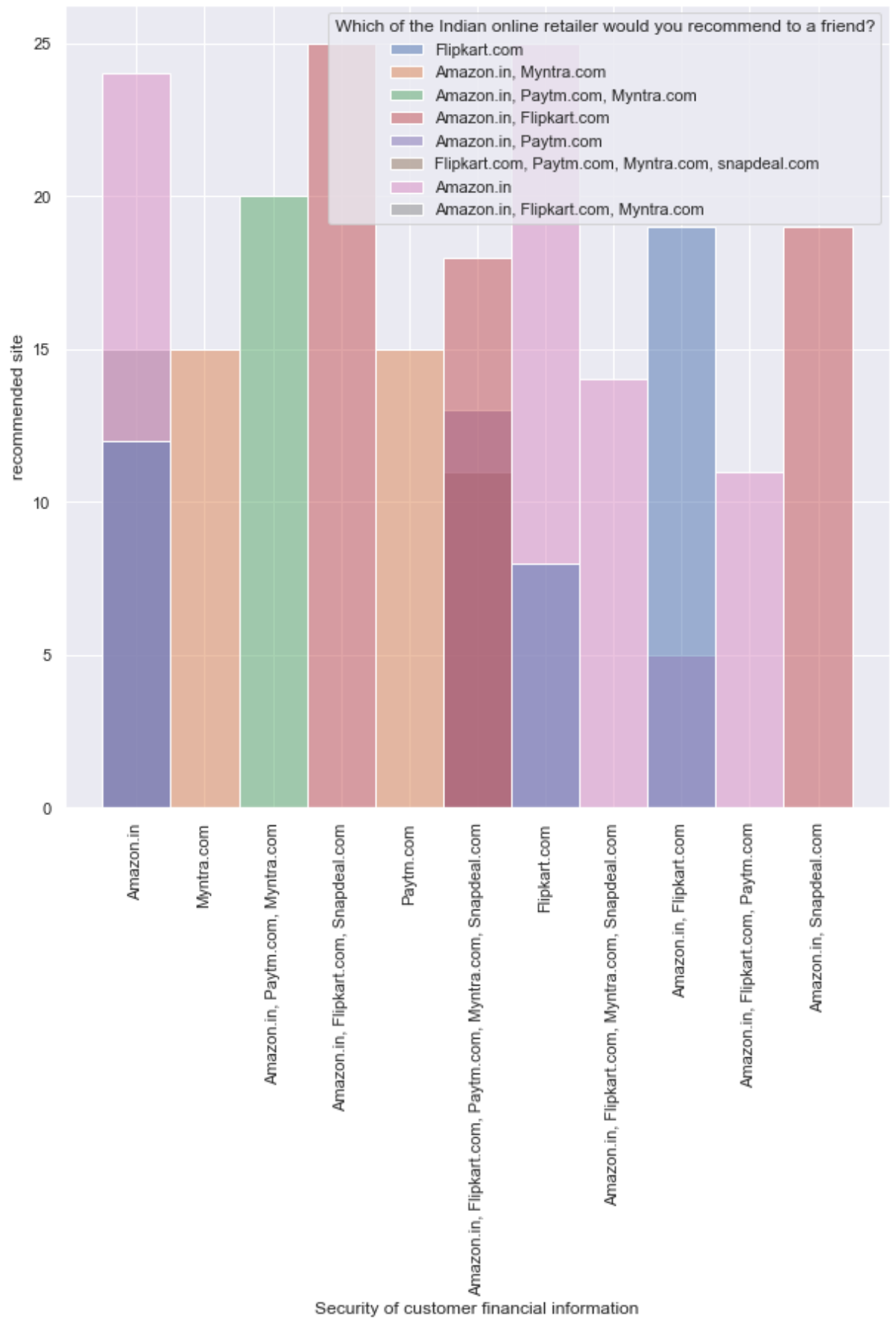
26 Trust that the online retail store will fulfill its part of the transaction at the stipulated time

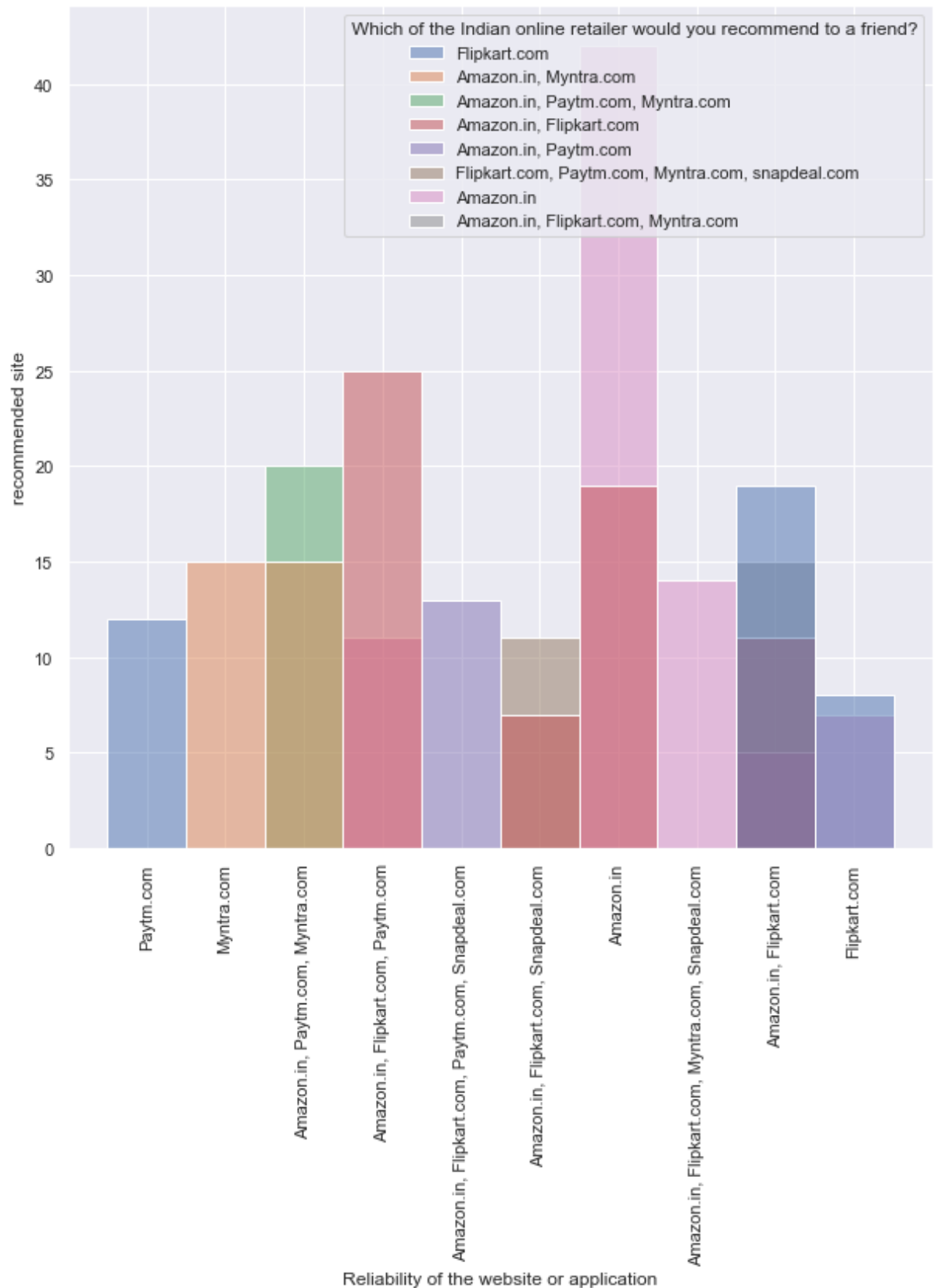








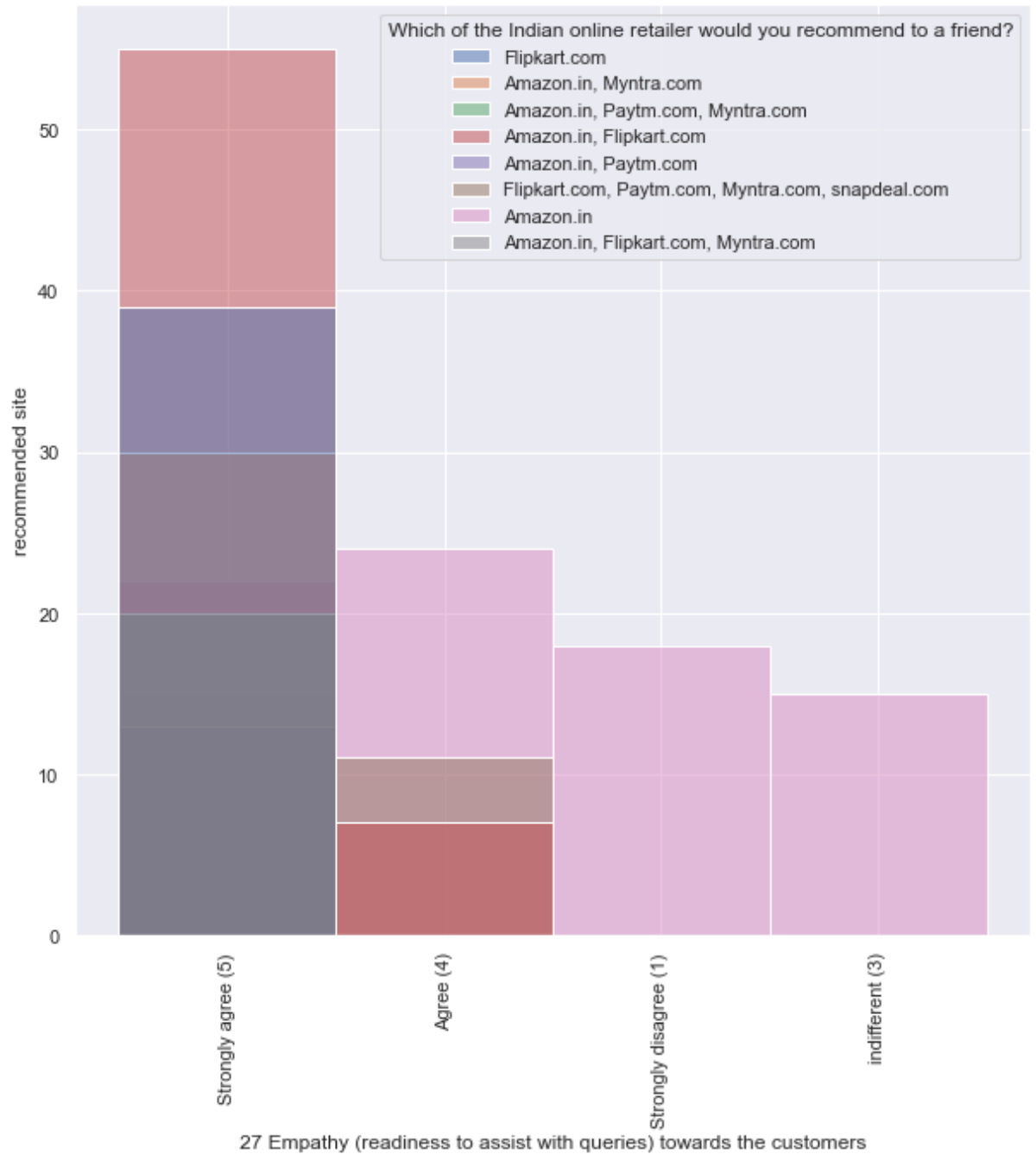


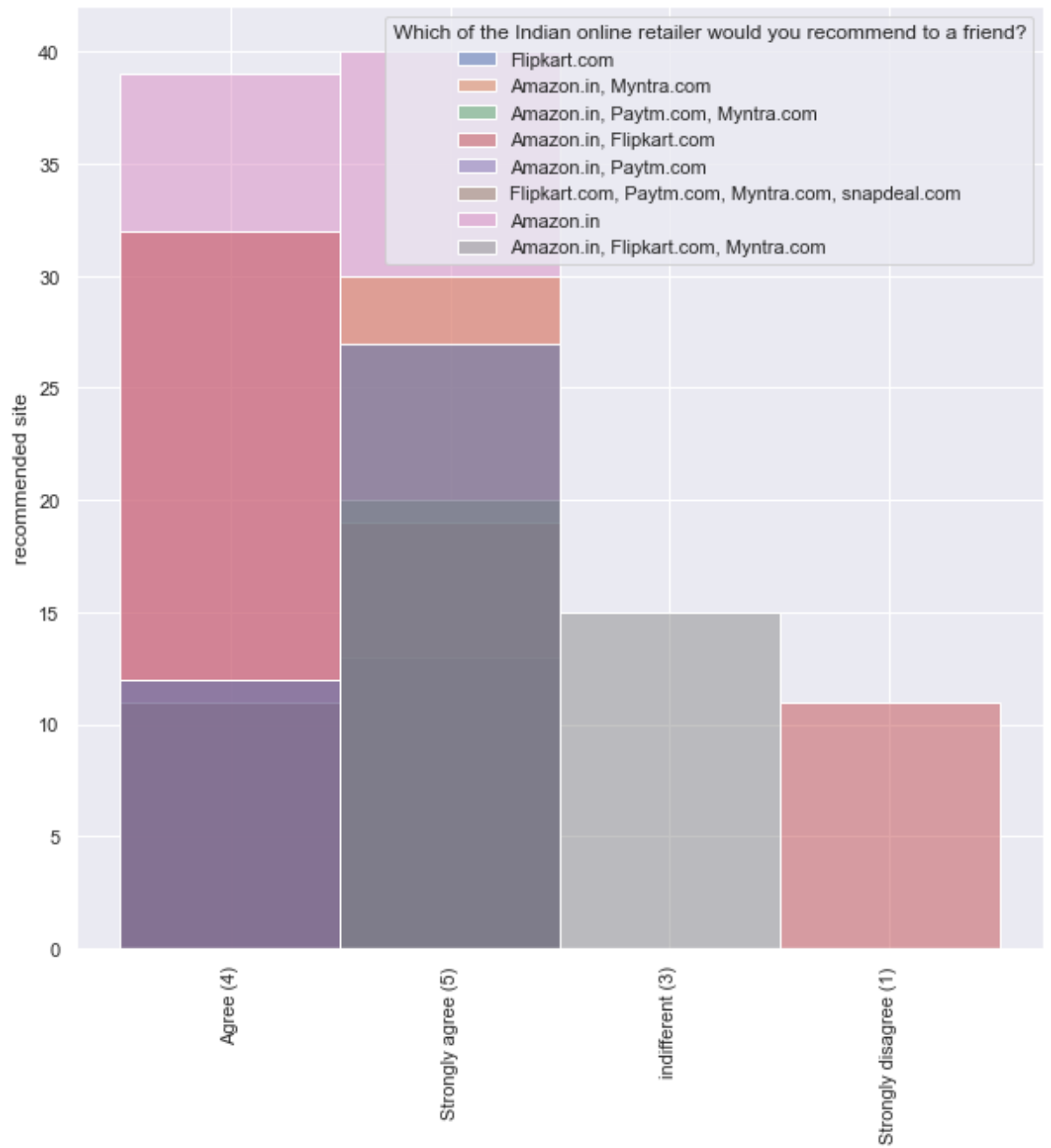


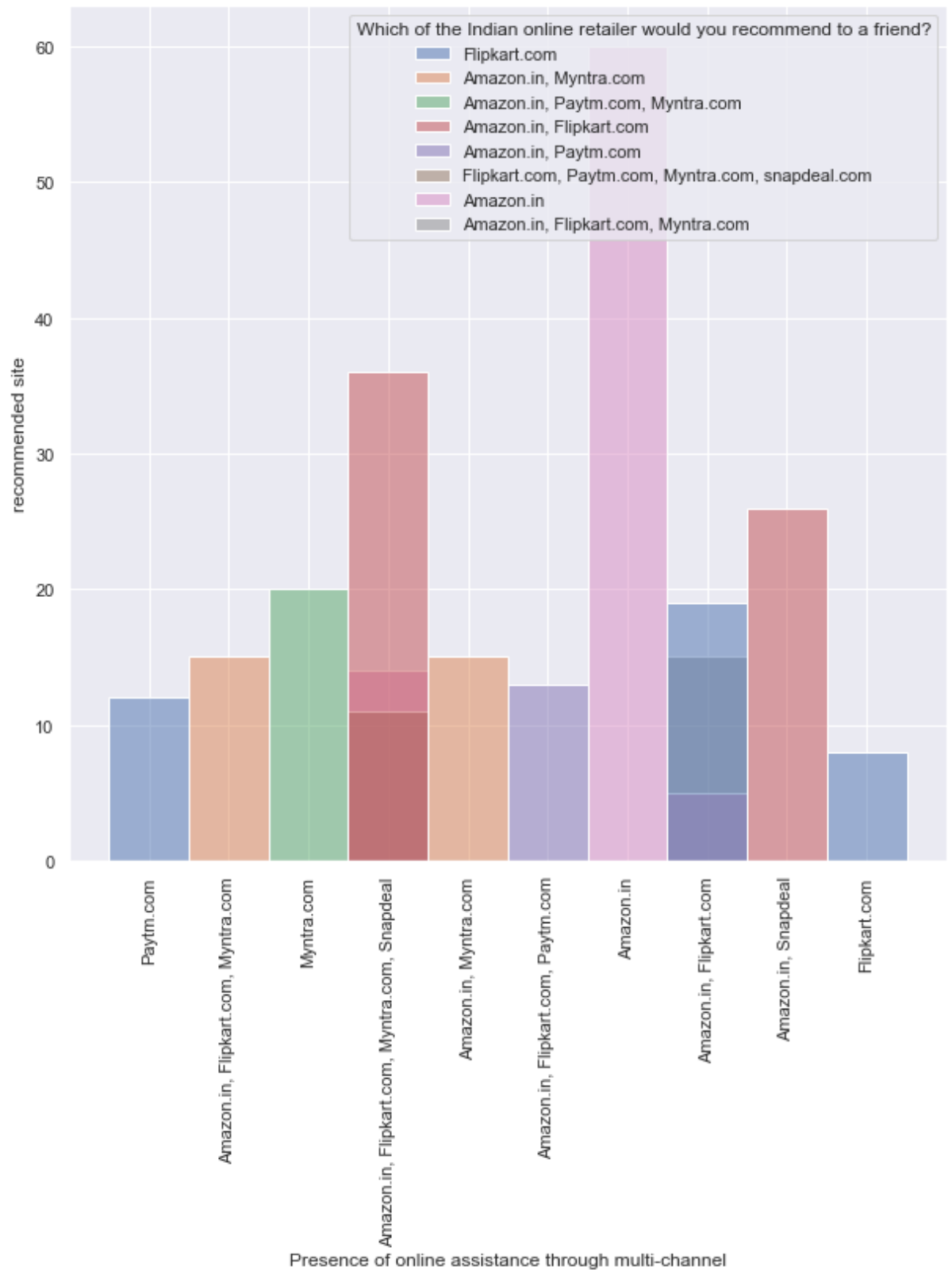
Observation: Customers strongly agree that the online shopping site should provide trust to customers and should give financial data security and privacy. From the above graphs we can conclude that amazon and

flipkart are the sites that give reliability, data privacy and customers trust them.

8) Empathy towards customers:

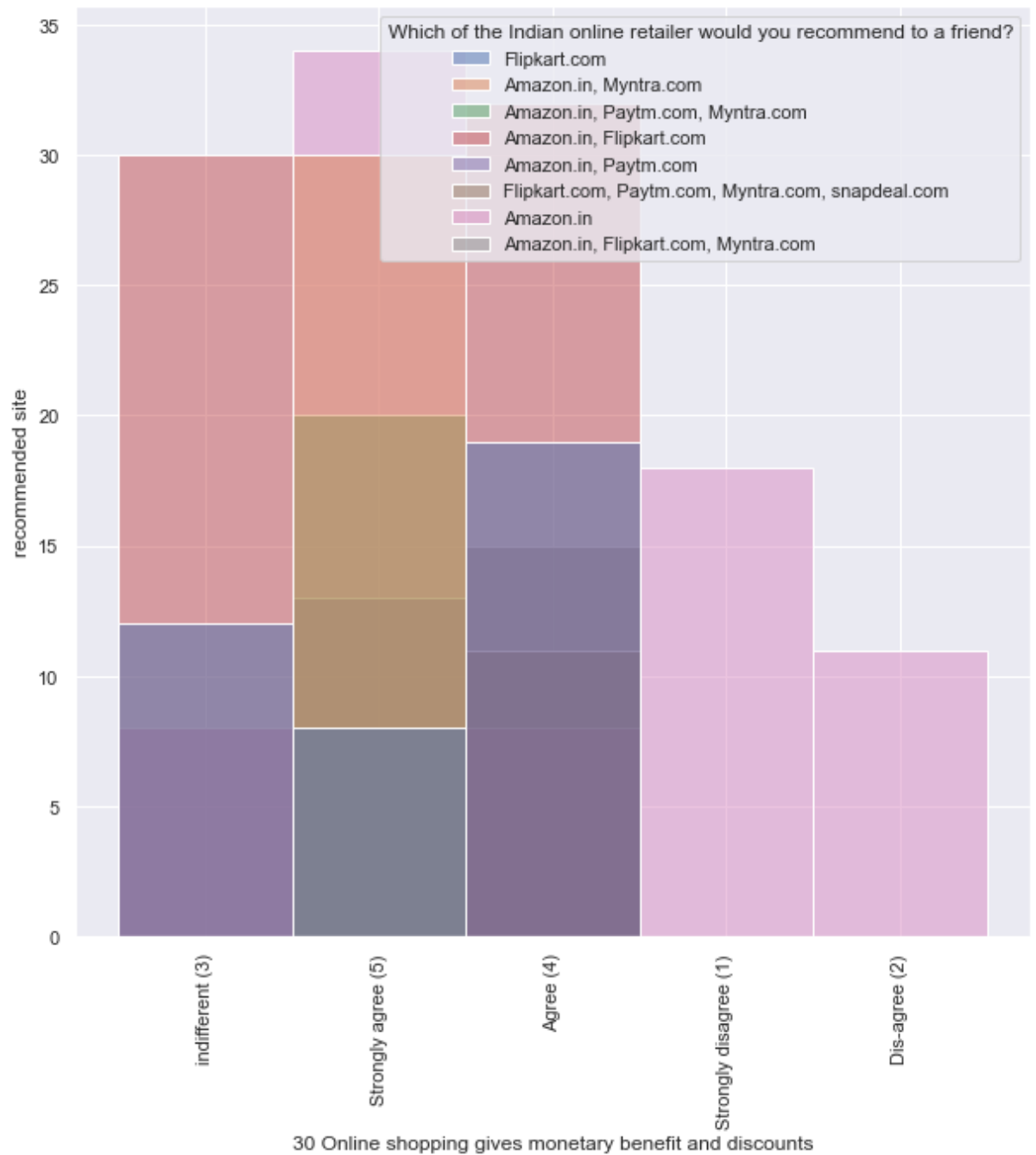


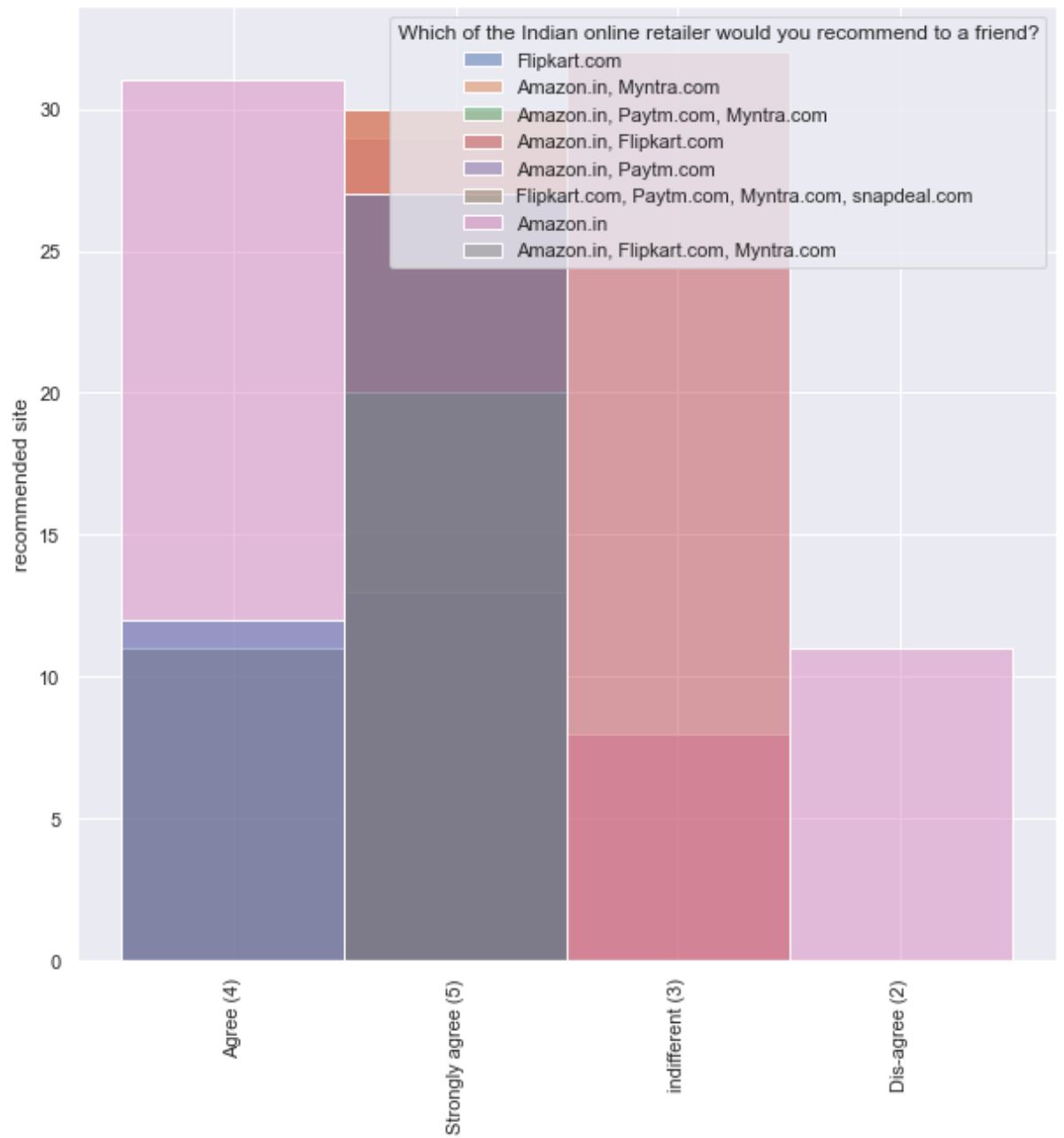


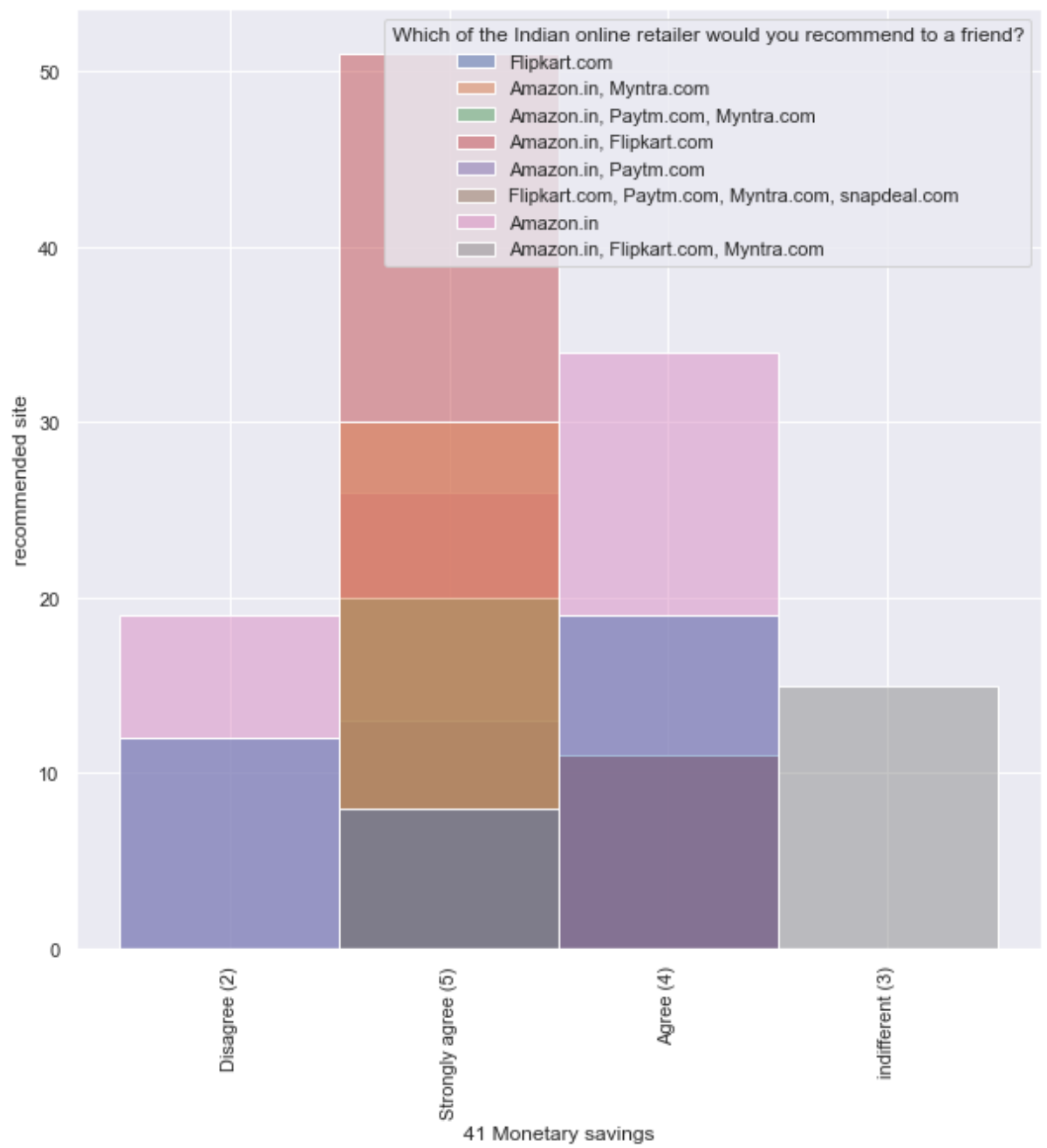


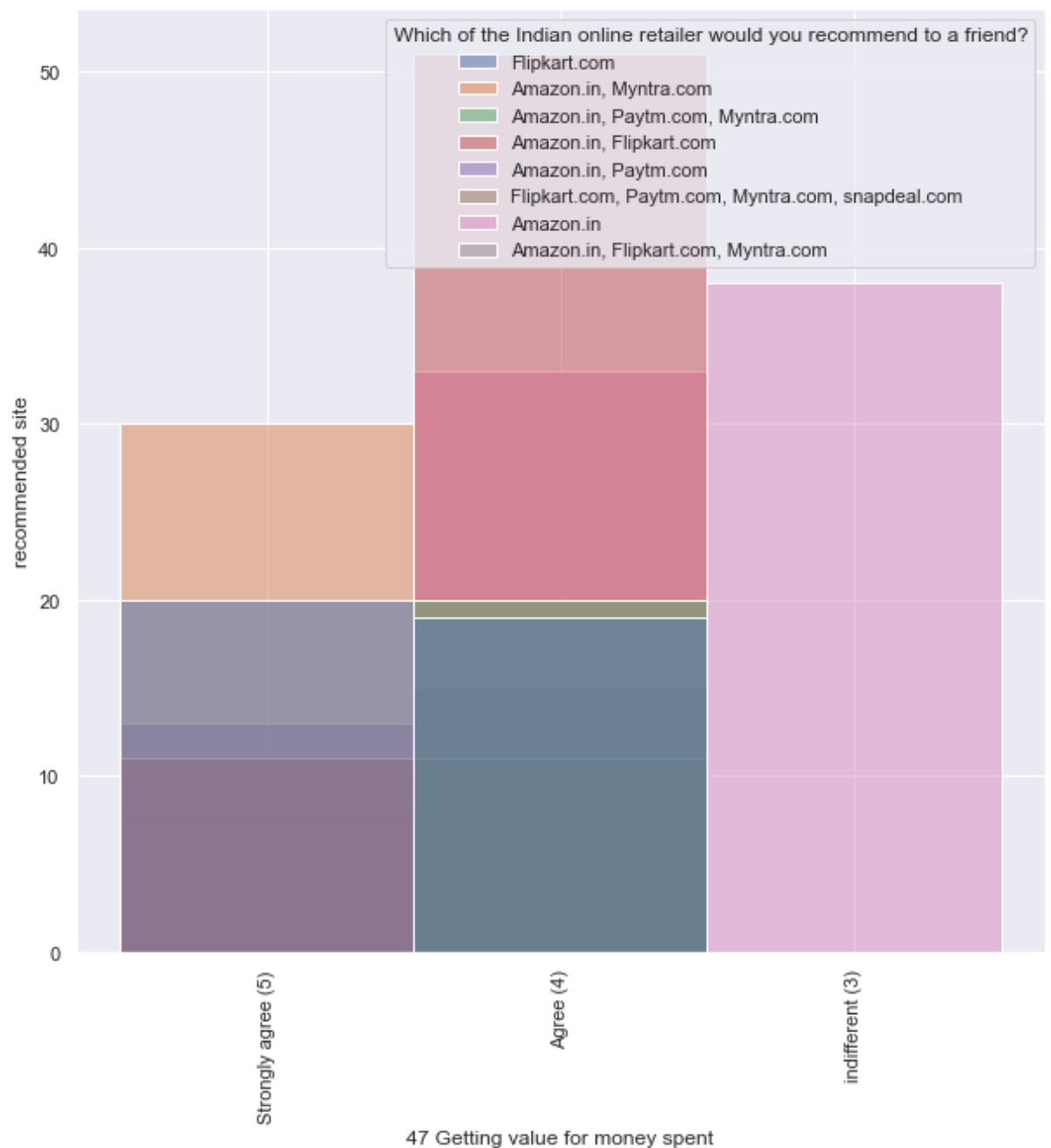
Observation: From the above graphs we could observe that customers prefer those websites that provide empathy, responsiveness and multiple channels to handle their query. Most recommended websites in this context is again amazon and flipkart.

9) Monetary Benefits:



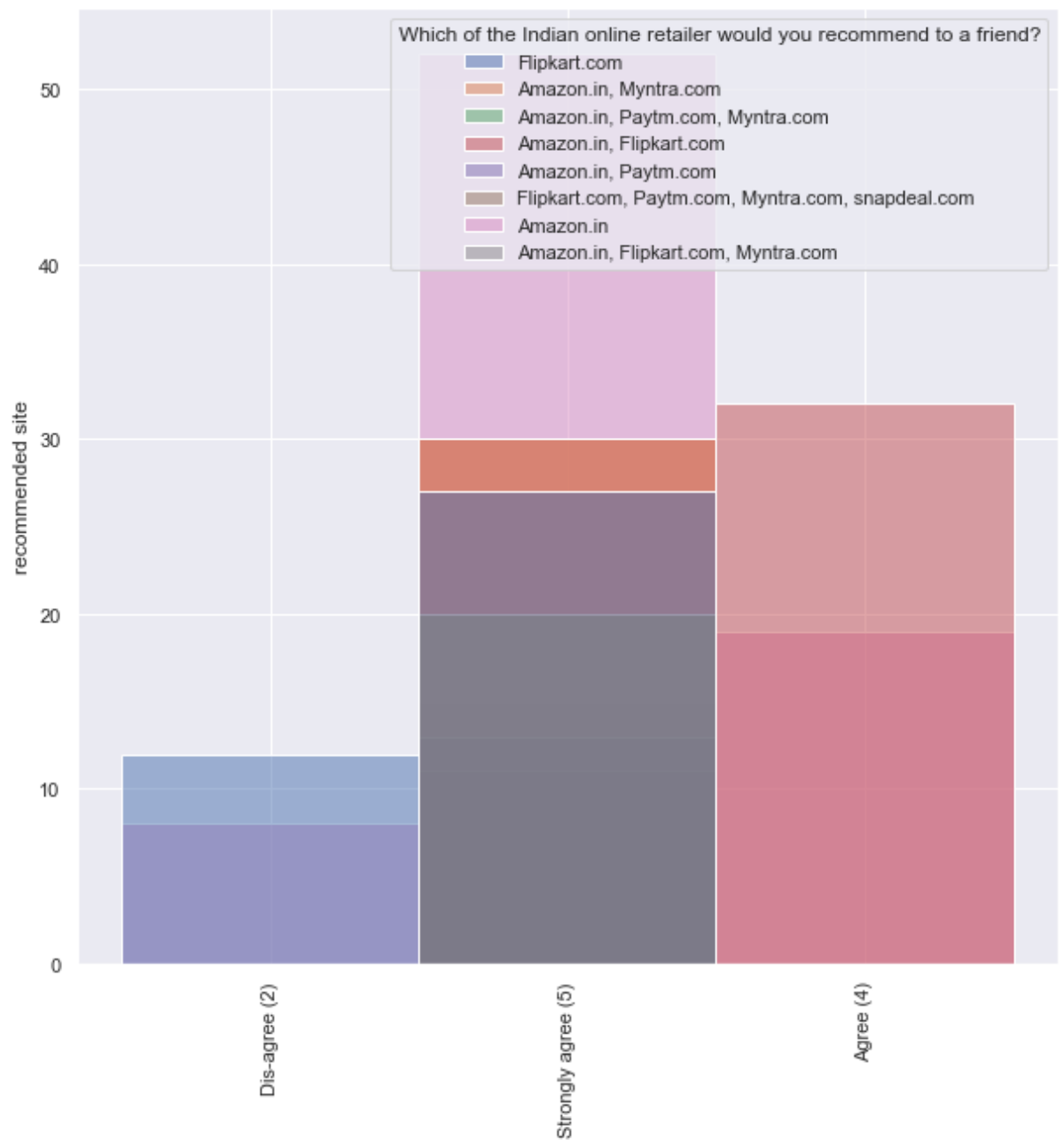






Observation: From the above graphs we can conclude that customers strongly agree and prefer those online shopping websites that provide discounts, monetary benefits and they feel that they are getting value for money.

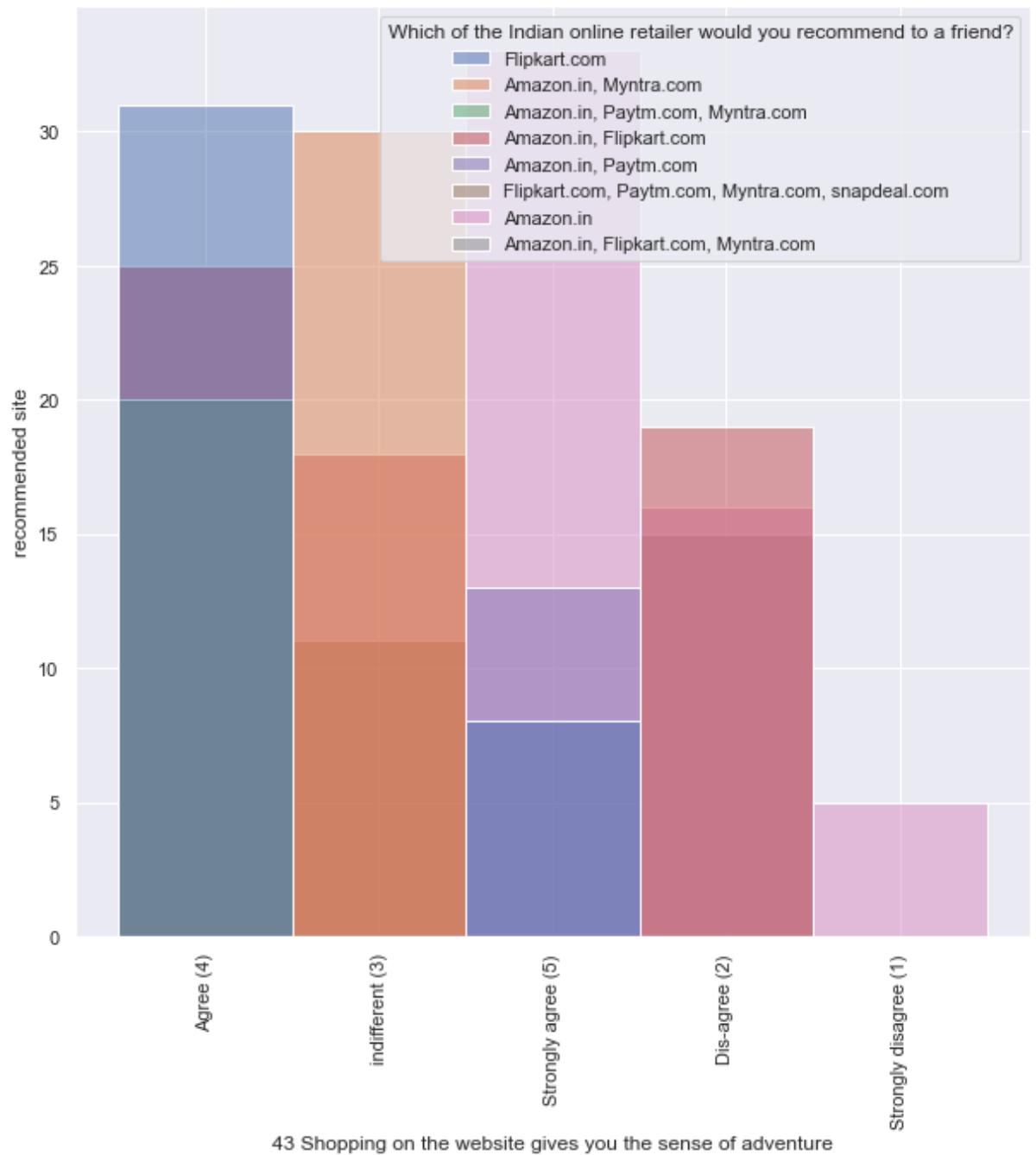
10)Return and Replacement:

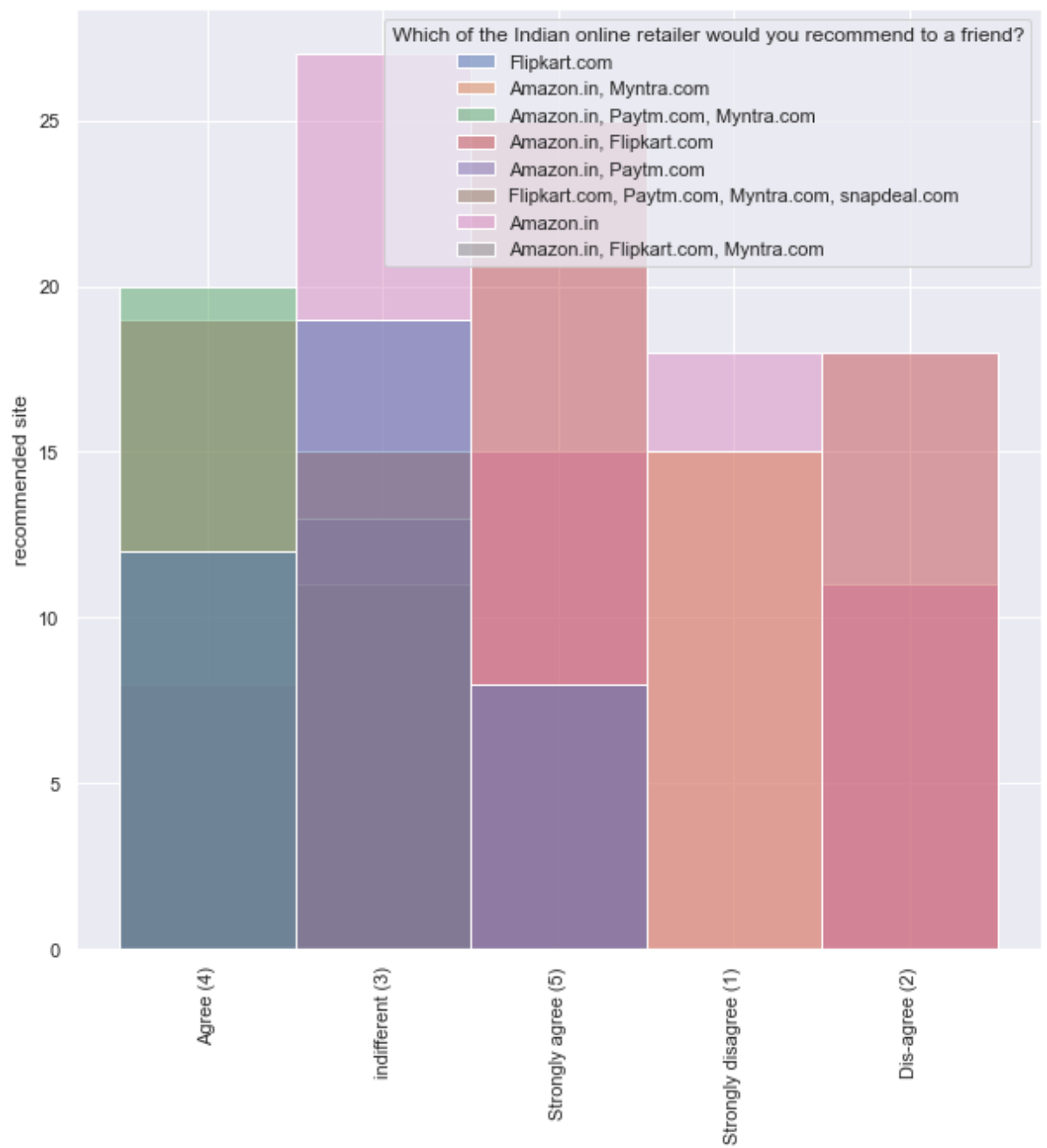


33 Return and replacement policy of the e-tailer is important for purchase decision

Observation: Return and replacement policy is very important factor for customers when doing online shopping.

11)Social Status and Adventure:





44 Shopping on your preferred e-tailer enhances your social status

Observation: Customers strongly agree that doing online shopping give them a sense of adventure but is indifferent that it provides social status.

6) Encoding: All are categorical feature so in order to know correlation between them we need to encode them.

Encoding

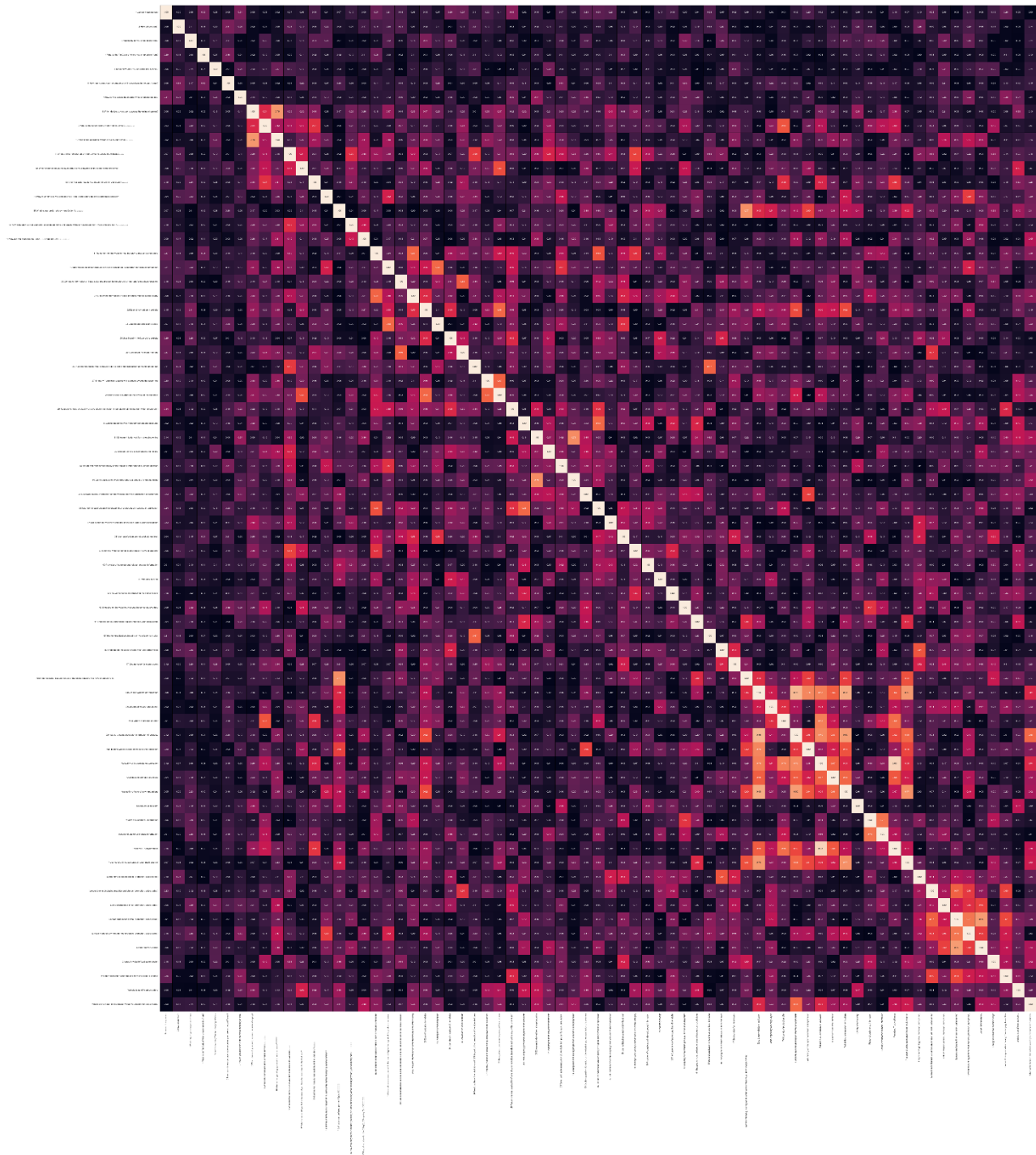
```
In [49]: 1 oe=OrdinalEncoder()
2 col=df.columns
3 for i in range(len(col)):
4     dfi=oe.fit_transform(df[[col[i]]])
5     df[col[i]]=dfi
6 df.head()
```

Out[49]:

| | 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? tititititit | 10 What is the operating system (OS) of your device? titititit | ... | Longer time to get logged in (promotion, sales period) | Longer time in displaying graphics and photos (promotion, sales period) | Late declaration of price (promotion, sales period) | Longer page loading time (promotion, sales period) | I |
|---|------------------------|--------------------|---------------------------------------|---|--|---|--|--|---|---|-----|--|---|---|--|---|
| 0 | 1.0 | 1.0 | 2.0 | 1.0 | 3.0 | 2.0 | 0.0 | 0.0 | 3.0 | 2.0 | ... | 0.0 | 0.0 | 3.0 | 5.0 | |
| 1 | 0.0 | 0.0 | 2.0 | 5.0 | 3.0 | 3.0 | 3.0 | 2.0 | 0.0 | 1.0 | ... | 1.0 | 6.0 | 7.0 | 10.0 | |
| 2 | 0.0 | 0.0 | 4.0 | 23.0 | 2.0 | 3.0 | 1.0 | 2.0 | 2.0 | 0.0 | ... | 7.0 | 6.0 | 4.0 | 7.0 | |
| 3 | 1.0 | 0.0 | 6.0 | 11.0 | 2.0 | 5.0 | 1.0 | 2.0 | 2.0 | 1.0 | ... | 9.0 | 7.0 | 4.0 | 8.0 | |
| 4 | 0.0 | 0.0 | 0.0 | 31.0 | 1.0 | 0.0 | 3.0 | 2.0 | 0.0 | 1.0 | ... | 5.0 | 8.0 | 5.0 | 8.0 | |

7) Correlation between features:

```
In [54]: 1 #Plotting heat map
2 #plt.figure(figsize=(20,10))
3 sns.set(rc = {'figure.figsize':(100,100)})
4 sns.heatmap(df.corr().abs(),annot=True,linewidth=0.5,linewidth='black',fmt='.2f')
```

Observation: Lot many multicollinearity exists. The multicollinear pairs are:

- 1) Which device do you use to access online shopping? and what is the operating system of your device?
- 2) What is your preferred payment option? and From the following, tick any (or all) of the online retailers you have shopped from;
- 3) Enjoyment is derived from shopping online and 34 Gaining access to loyalty programs is a benefit of shopping online

4)Easy to use website or application and Complete, relevant description information of products

5)Easy to use website or application and Fast loading website speed of website and application

6)Easy to use website or application and Reliability of the website or application

7)Easy to use website or application and Availability of several payment options

8)Wild variety of product on offer and Reliability of the website or application

9)Complete, relevant description information of products and Reliability of the website or application

10)Complete, relevant description information of products and Availability of several payment options

11)Availability of several payment options and Presence of online assistance through multi-channel

12)Presence of online assistance through multi-channel and easy to use website or application

13)Privacy of customers' information and Security of customer financial information

14)Reliability of the website or application and Percieved Trustworthyness

15)Longer page loading time (promotion, sales period) and Limited mode of payment on most products (promotion, sales period)

16)Longer page loading time (promotion, sales period) and Longer Delivery Period

We can get sure of this by using VIF

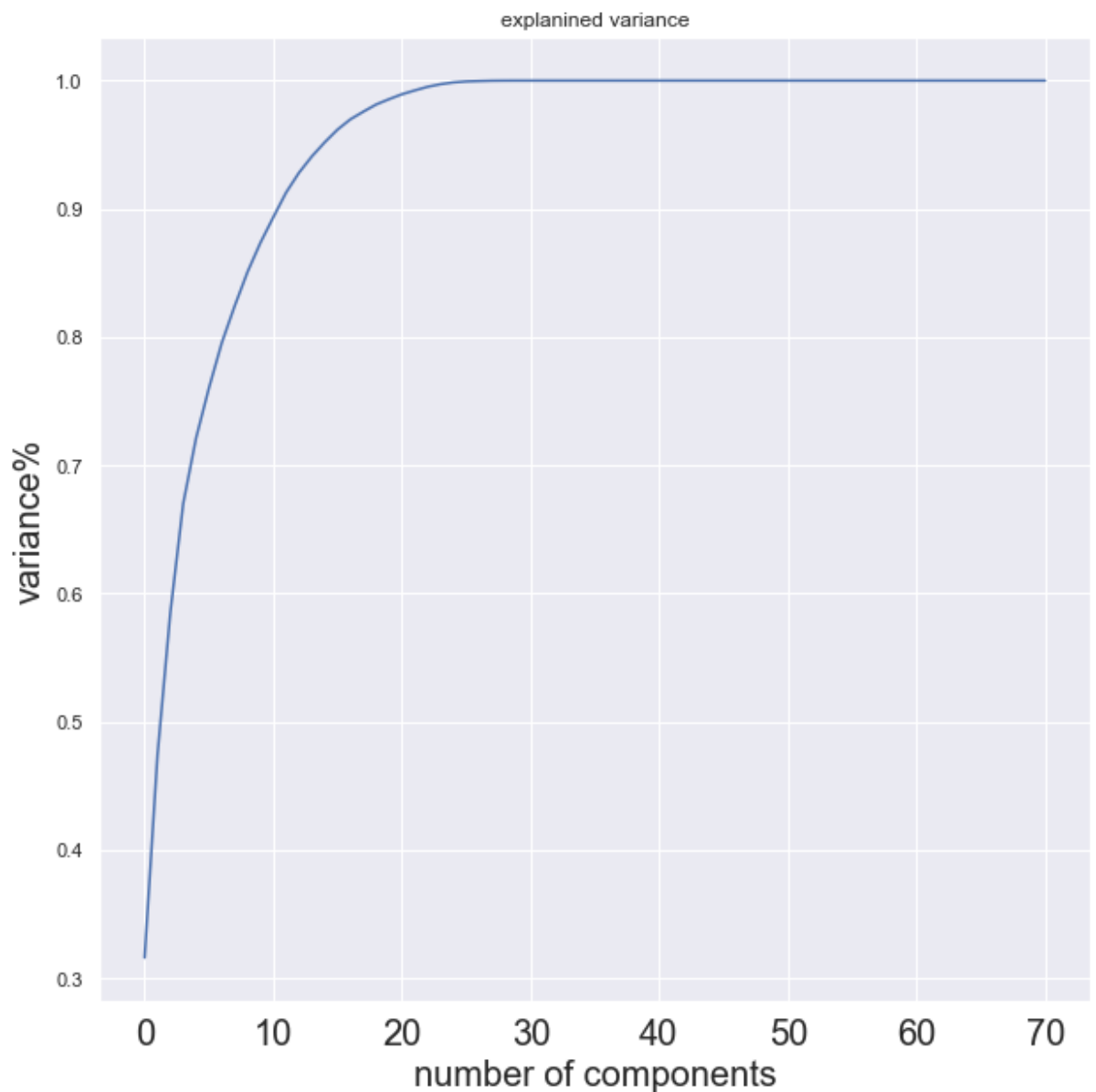
8) VIF (Variance Inflation Factor):

Out[56]:

| | vif | features |
|----|-----------|---|
| 0 | 2.287690 | 1 Gender of respondent |
| 1 | 1.881739 | 2 How old are you? |
| 2 | 1.694354 | 3 Which city do you shop online from? |
| 3 | 1.778355 | 4 What is the Pin Code of where you shop online? |
| 4 | 1.713943 | 5 Since How Long You are Shopping Online ? |
| 5 | 2.280625 | 6 How many times you have made an online purchase? |
| 6 | 2.406648 | 7 How do you access the internet while shopping? |
| 7 | 39.295620 | 8 Which device do you use to access the online store? |
| 8 | 30.503088 | 9 What is the screen size of your mobile device? |
| 9 | 38.680963 | 10 What is the operating system (OS) of your device? |
| 10 | inf | 11 What browser do you run on your device to access the online store? |
| 11 | inf | 12 Which channel did you follow to arrive at the online store? |
| 12 | inf | 13 After first visit, how do you reach the online store? |
| 13 | inf | 14 How much time do you explore the e-retailer's website? |
| 14 | inf | 15 What is your preferred payment Option? |
| 15 | inf | 16 How frequently do you abandon (selecting an item and not purchasing it)? |
| 16 | inf | 17 Why did you abandon the <input type="checkbox"/> Bag <input type="checkbox"/> Shopping Cart? |
| 17 | inf | 18 The content on the website must be easy to understand. |
| 18 | inf | 19 Information on similar product to the one highlighted. |
| 19 | inf | 20 Complete information on listed seller and product. |

Observation: Here we can see there is perfect correlation between variables thus vif score is inf that is infinity. So multicollinearity is confirmed and lot of features has infinity vif score so we will use PCA to combine correlated components and remove multicollinearity.

9) PCA(Principle Component Analysis):



Observation: We can see that approx. 98% of the variance is explained by 30 components so instead of passing all the 71 components in model lets use only 30.

Since all the variables are categorical thus we dont need to check outliers and skewness.

CONCLUSION

Following are the key factors that any online shopping organisation should consider to retain their customers:

- 1) User friendliness and easy to use website or application
- 2) Availability of almost all the payment options for all category of products mostly credit/debit card and cash on delivery.
- 3) Websites and applications should be created using such a technology so that the loading and processing speed is as fast as possible.
- 4) Customers query should be resolved as soon as possible and companies should show empathy towards them.
- 5) Online shopping companies should provide security and privacy to customers personal as well as financial data.
- 6) There should be frequent offers and discounts given by the site so that customers feel that every purchase is worth the cost.
- 7) There should be very flexible return and replacement policy.

Keeping all these points we found that customers mostly recommended 2 online shopping websites: Amazon and Flipkart.