CUSTOMER RETENTION DATASET

Submitted by:

Ruzaina Khan

CONTENTS:

INTRODUCTION

Customer Retention refers to the action and strategies a business uses to try and keep existing customers. To enable these actions, customer retention analytics provide predictive metrics of which customer might churn-which enable them to get ahead of it. 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 hedonic 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
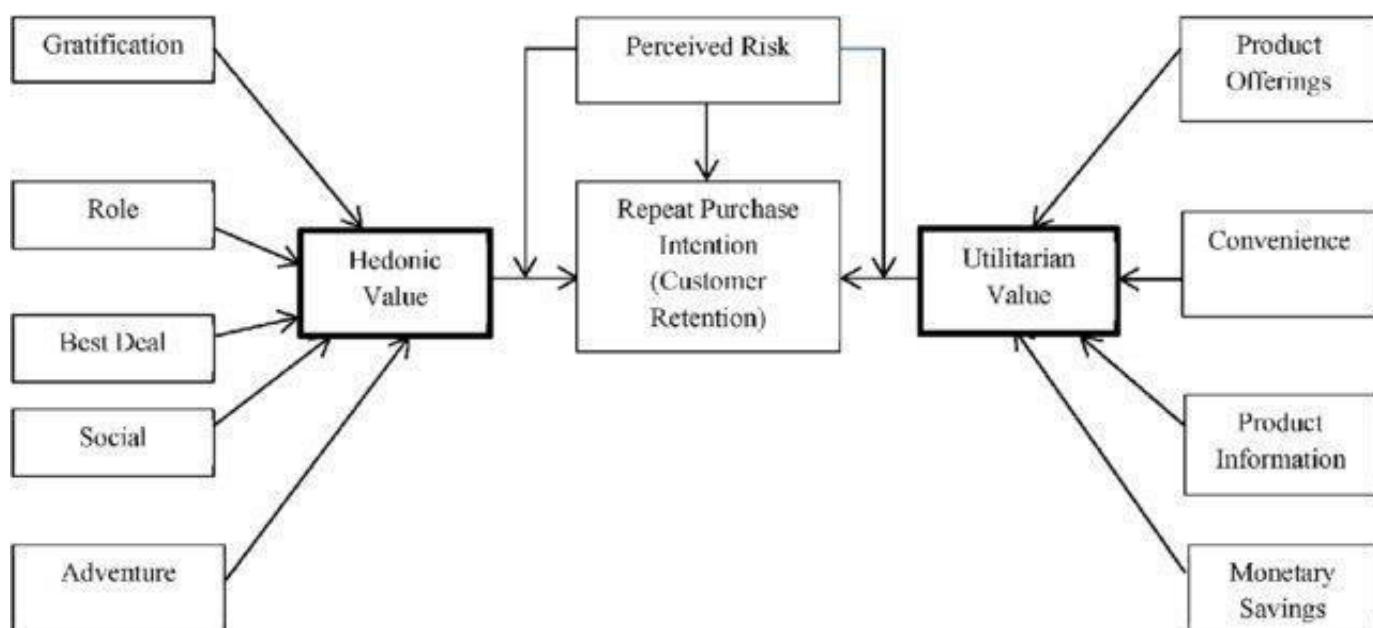
1. What is Customer Retention?

Customer Retention refers to a company's ability to turn customers into repeat buyers and prevent them from switching to a competitor. It indicates whether your product and the quality of the service please your existing customers. 3 Customer Retention strategies are the processes and initiatives businesses put in place to build customer loyalty and improve customer lifetime value.

2. Why is customer retention important?

Customer retention increases your customers' lifetime value and boosts your revenue. It also helps you build amazing relationship with your customers. You aren't just another website or store. They trust you with their money because you give them value in exchange. According to the Harvard Business Review, acquiring a new customer can be 5 to 25 times more expensive than holding on to an existing one. You don't need to spend big on marketing, advertising or sales outreach. It is easier to turn existing customers into repeating ones, since they already trust your brand from previous purchases. New customers, however, often require more convincing when it comes to the initial sale.

3. Customer Retention benefits:

• **Retention is Cheaper than Acquisition**

• **Loyal Customers are more profitable.**

• **Your Brand will stand out from the crowd.**

• **Engage customers provide more feedback.**

• **Loyal customers are more forgiving.**

• **Customers will explore your brand**

Analytical Problem Framing

The studies were identified using different keywords that are related mainly to customer retention and analysing other factors such as satisfaction, trust, and commitment in several searching processes until reaching a suitable number of related and targeted articles and studies to select and analyse. In the dataset there was 1 numerical columns and all the categorical columns has been analysed using all categorical plots to get better insight on the dataset. After visualizing the features it bought us a good insight that what actually customers are expecting from the e-commerce sellers.

Out[2]:

	1Gender 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? \t\t\t\t\t	10 What is the operating system (OS) of your device? \t\t\t	...	Longer tim to get logge i (promotio sales perioc
0	Male	31-40 years	Delhi	110009	Above 4 years	31-40 times	Dial-up	Desktop	Others	Window/windows Mobile	...	Amazon.I
1	Female	21-30 years	Delhi	110030	Above 4 years	41 times and above	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	...	Amazon.ir Flipkart.cor
2	Female	21-30 years	Greater Noida	201308	3-4 years	41 times and above	Mobile Internet	Smartphone	5.5 inches	Android	...	Myntra.co
3	Male	21-30 years	Karnal	132001	3-4 years	Less than 10 times	Mobile Internet	Smartphone	5.5 inches	IOS/Mac	...	Snapdeal.cor
4	Female	21-30 years	Bangalore	530068	2-3 years	11-20 times	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	...	Flipkart.con Paytm.cor

5 rows × 71 columns

Data:

In this project we are given excel file containing dataset of customer retention. There are 71 columns by which we can analyse how e-commerce sector can look after its customers to improve their service All the variables in the dataset are of object type except pincode, however it is categorical in nature. The dataset contains 269 rows and 71 columns. The data contains no null value however there is a lot of unwanted characters in the columns.

Hardware and Software Requirements and Tools Used

• **A mid level computer that runs on Intel i3- i5 8th generation, 4gb ram or A10/A11or any other equivalent chipset and a suitable processor.**

• **Juypter Notebook/Google chrome**

• **Libraries and packeges used:**

import pandas as pd

import numpy as np

import matplotlib. pyplot as plt

import seaborn as sn

 import warnings warnings.filterwarnings("ignore")

We will use Python through Jupyter notebook for data processing. Also we will use liabraries here is numpy, matplotlib, pandas and seaborn. The matpotplotlib and seaborn library has been used to make charts to visualize and understand the problem, correlation, outliers and many other things, the pandas and numpy library issued to handle dataset and perform various tasks.

Shape of the Dataset – 269 Rows and 71 Columns.

Data Type of the Dataset - 1 Integer Column and 70 are Categorical Columns.

Null Values in Dataset – No Null Values .

Exploratory Data Analysis:

- 1st Columns 1Gender of respondent

```
In [10]:   1  #Checking 1st Columns 1Gender of respondent (Using Countpplot)
           2  plt.figure(figsize=(5,4),facecolor='blue')
           3  sns.countplot(x='1Gender of respondent',data=data)
           4  plt.savefig("1Gender of respondent.png")
           5  data['1Gender of respondent'].value_counts()
```

```
Out[10]:   Female    181
           Male       88
           Name: 1Gender of respondent, dtype: int64
```
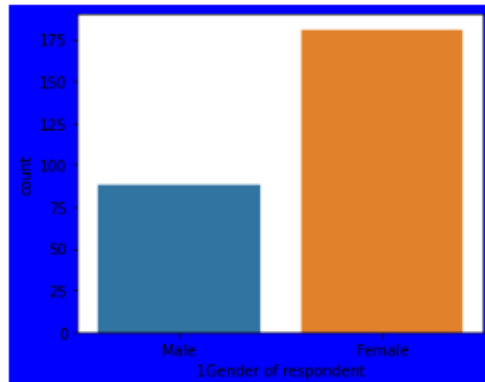


Here we can see that Gender of respondent is ,

Male - 88

Female - 181

Majority of People are Female.

- 2nd Column 2 How old are you?

```
In [11]:   1  #Checking 2nc Column 2 How old are you? (Using Countplot)
           2
           3  sns.countplot(x="2 How old are you? ",data=data)
           4  plt.savefig('2 How old are you.png')
           5  data['2 How old are you? '].value_counts()

Out[11]:  31-40 years        81
          21-30 years        79
          41-50 yaers        70
          Less than 20 years 20
          51 years and above 19
          Name: 2 How old are you? , dtype: int64
```
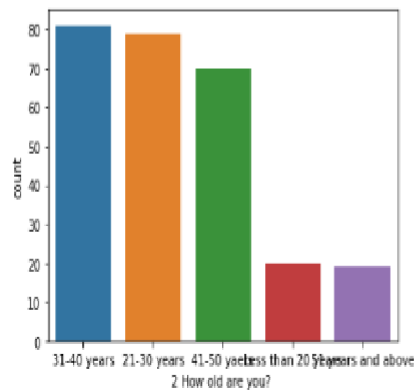

Here we can se that ,

81 of the customers are from age group 31-40 years.

**79 of the customers are form age froup 21-30 years.**

70 of the customer are from age from 41-50 years

20 of the customer are less then 20 year old

19 of the customer are above 51 years old

- **3rd Columns Which city do you shop online from?**

```
12]:    1  #Checking the 3rd Columns Which city do you shop online from?
        2  plt.figure(figsize=(20,15))
        3  sns.countplot(x='3 Which city do you shop online from?',data=data)
        4  plt.savefig('3 Which city do you shop online from.png')
        5  data['3 Which city do you shop online from?'].value_counts()
```
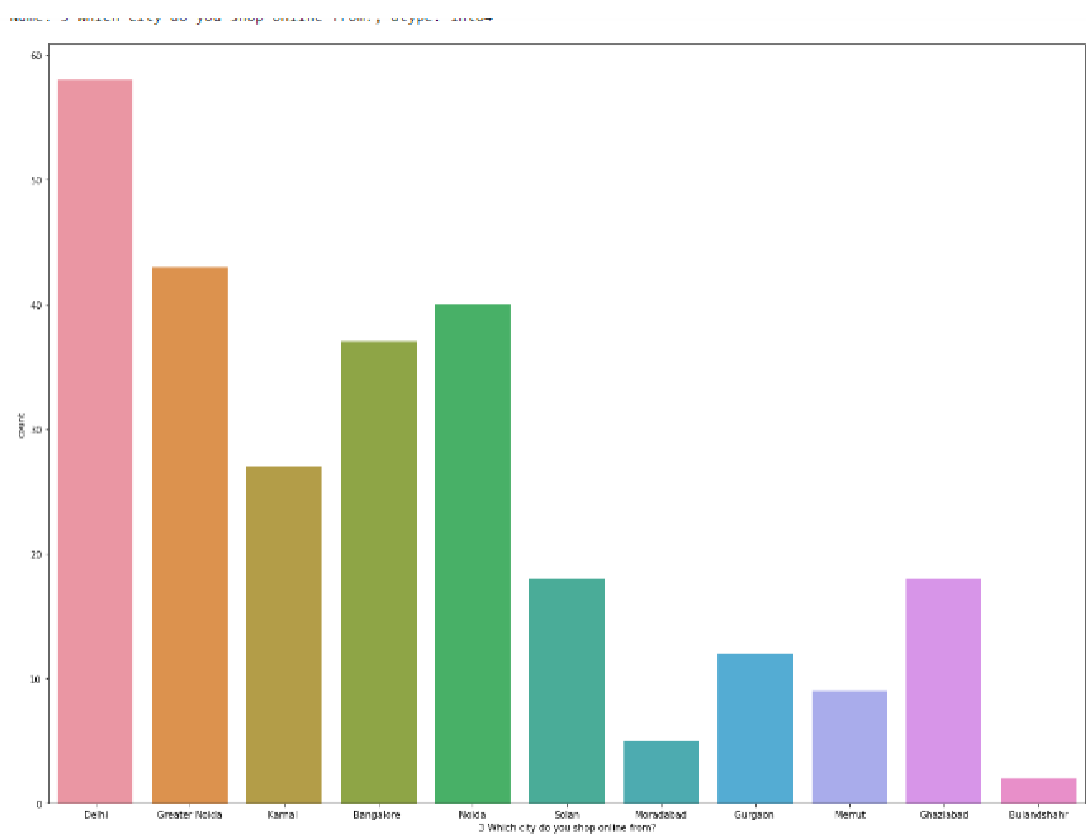

Majority of the People From Delhi.

- **4 What is the Pin Code of where you shop online from?**

```
In [13]:  1  #Lets plot 4th columns(''4 What is the Pin Code of where you shop online from?'')
          2  plt.figure(figsize=(12,10))
          3  sns.countplot(x='4 What is the Pin Code of where you shop online from?',data=data)
          4  plt.xticks(rotation='vertical')
          5  plt.savefig('4 What is the Pin Code of where you shop online from.png')
          6  data['4 What is the Pin Code of where you shop online from?'].value_counts()
```
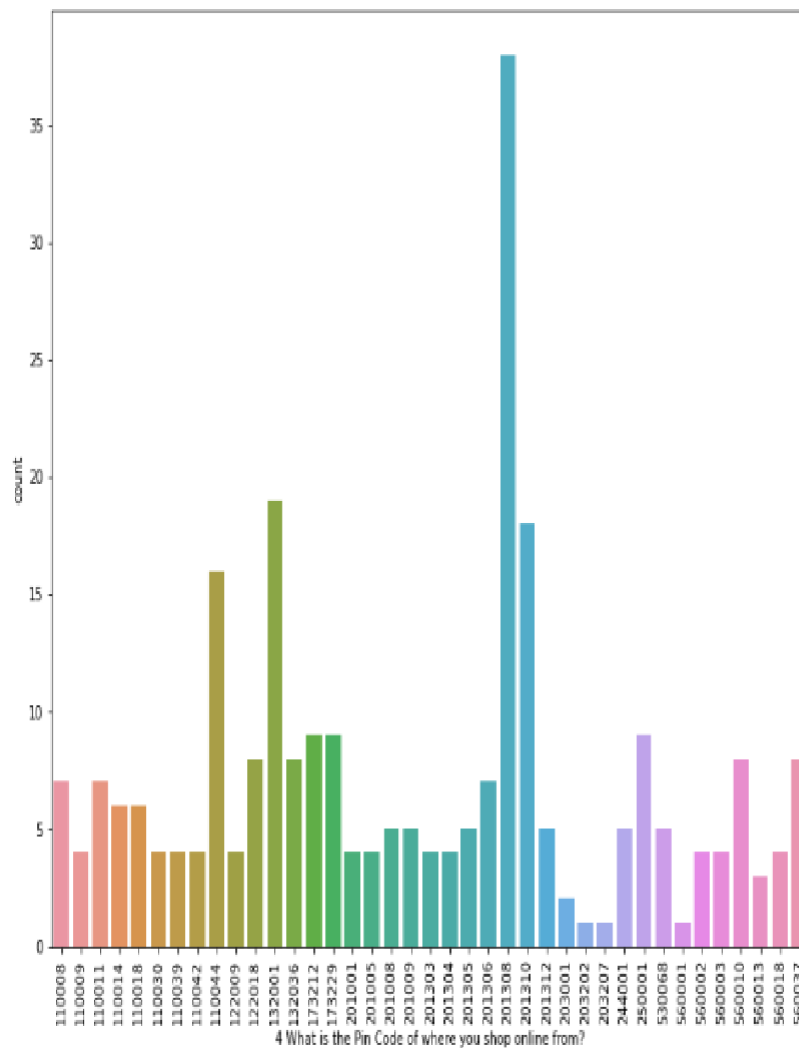

Majority of the People from 201308

- **5 Since How Long You are Shopping Online ?**

```
In [14]:   1  #Lets plot 5th Columns '5 Since How Long You are Shopping Online ?'
           2  plt.figure(figsize=(20,15))
           3  sns.countplot(x='5 Since How Long You are Shopping Online ?',data=data)
           4  plt.savefig('5 Since How Long You are Shopping Online.png')
           5  data['5 Since How Long You are Shopping Online ?'].value_counts()
```
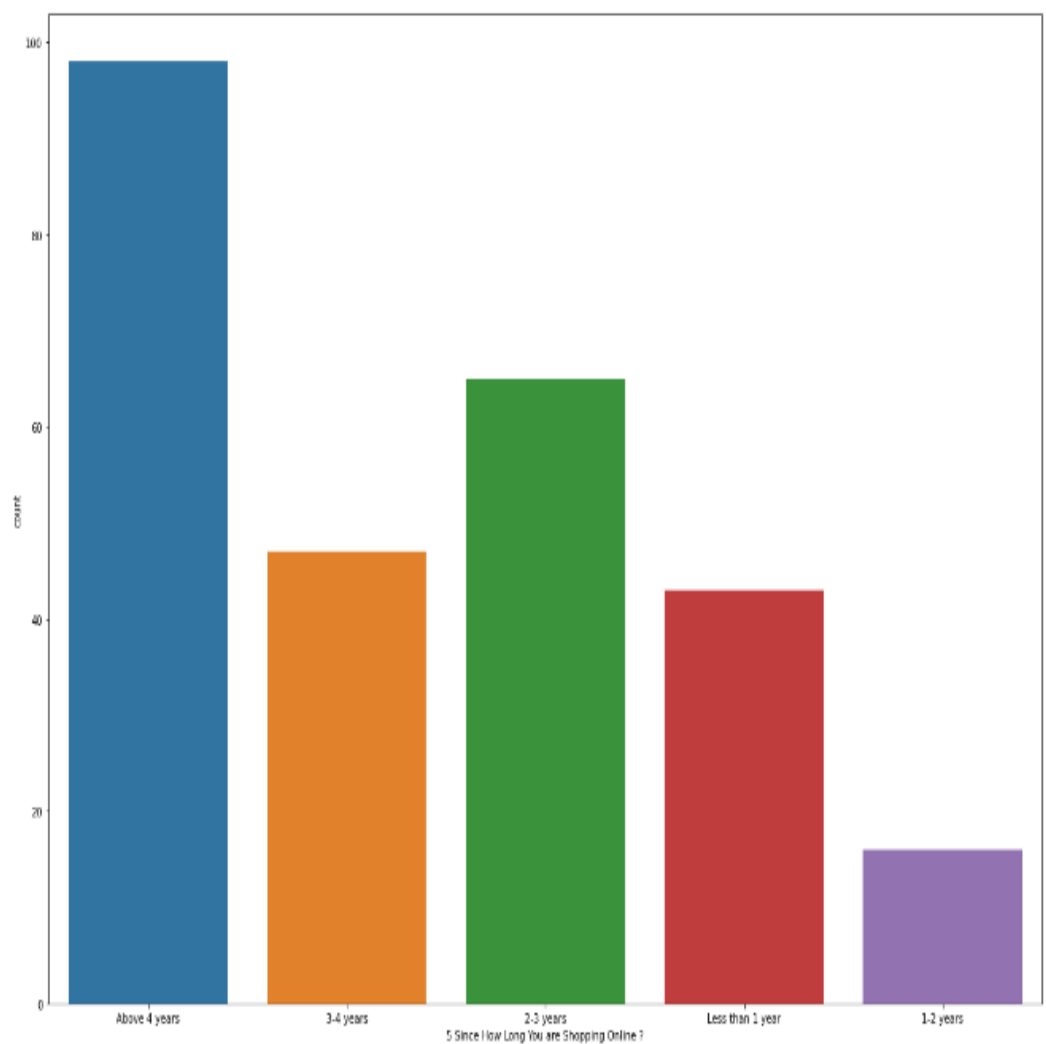

Majority of the People from Above 4 Year.

6 How many times you have made an online purchase in the past 1 year?

```
In [15]:  1  #Lets plot 6th columns (''6 How many times you have made an online purchase in the past 1 year?'')
          2  plt.figure(figsize=(20,15))
          3  sns.countplot(x='6 How many times you have made an online purchase in the past 1 year?',data=data)
          4  plt.savefig('6 How many times you have made an online purchase in the past 1 year.png')
          5  data['6 How many times you have made an online purchase in the past 1 year?'].value_counts()
```
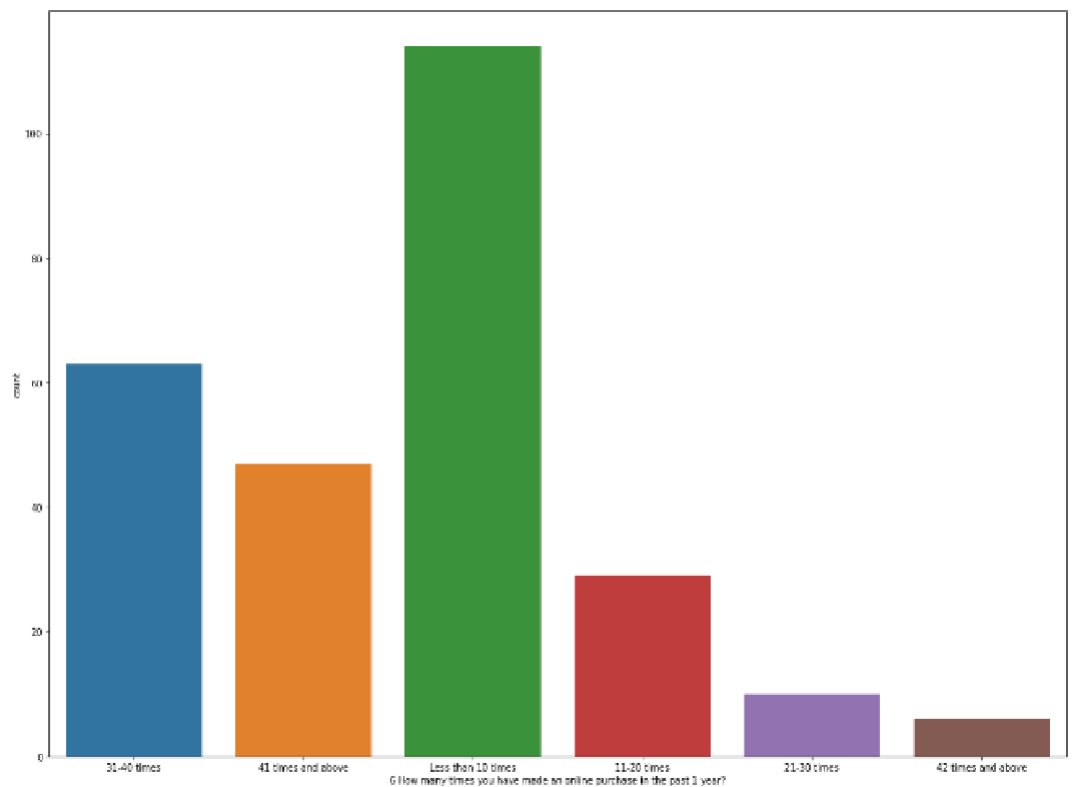

Here we can see theat ('' How many times you have made an online purchase in the past 1 year?)

Less than 10 times - 114 People

31-40 times - 63 People

41 times and above - 47 People

11-20 times - 29 People

21-30 times - 10 People

42 times and above - 6 People

7 How do you access the internet while shopping on-line?

```
In [16]:  1  #Lets plot 7th Columns('7 How do you access the internet while shopping on-line?')
          2  plt.figure(figsize=(20,15))
          3  sns.countplot(x='7 How do you access the internet while shopping on-line?',data=data)
          4  plt.savefig('7 How do you access the internet while shopping on-line.png')
          5  data['7 How do you access the internet while shopping on-line?'].value_counts()
```
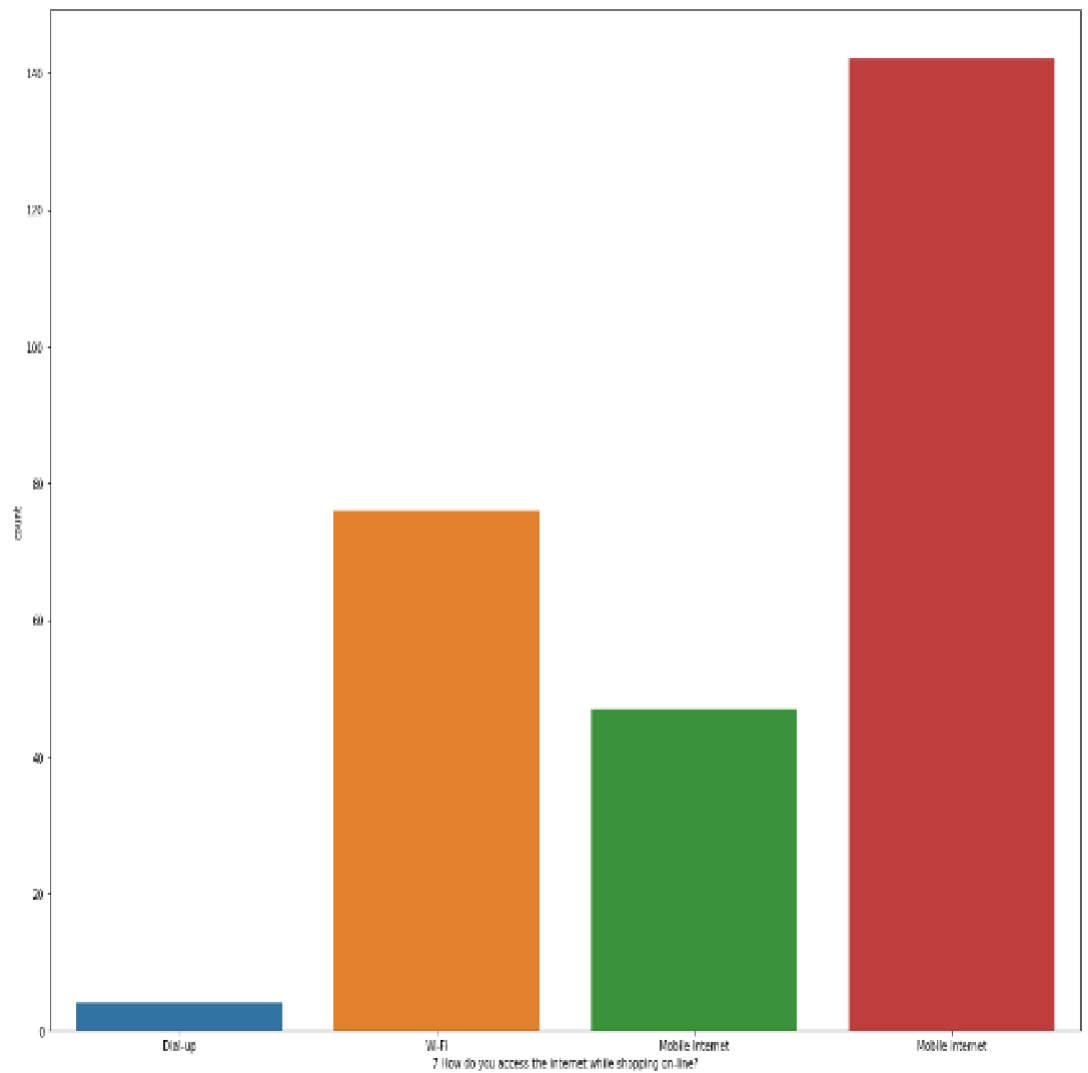

Majority of the people are using Mobile data - 142 People

Wifi Uses - 76 People

Mobile Internet - 74 People

Dail up - 4

8 Which device do you use to access the online shopping?

```
In [17]:  1  #Lets plot 8th columns ('8 Which device do you use to access the online shopping?')
          2  plt.figure(figsize=(20,15))
          3  sns.countplot(x='8 Which device do you use to access the online shopping?',data=data)
          4  plt.savefig('8 Which device do you use to access the online shopping.png')
          5  data['8 Which device do you use to access the online shopping?'].value_counts()
```
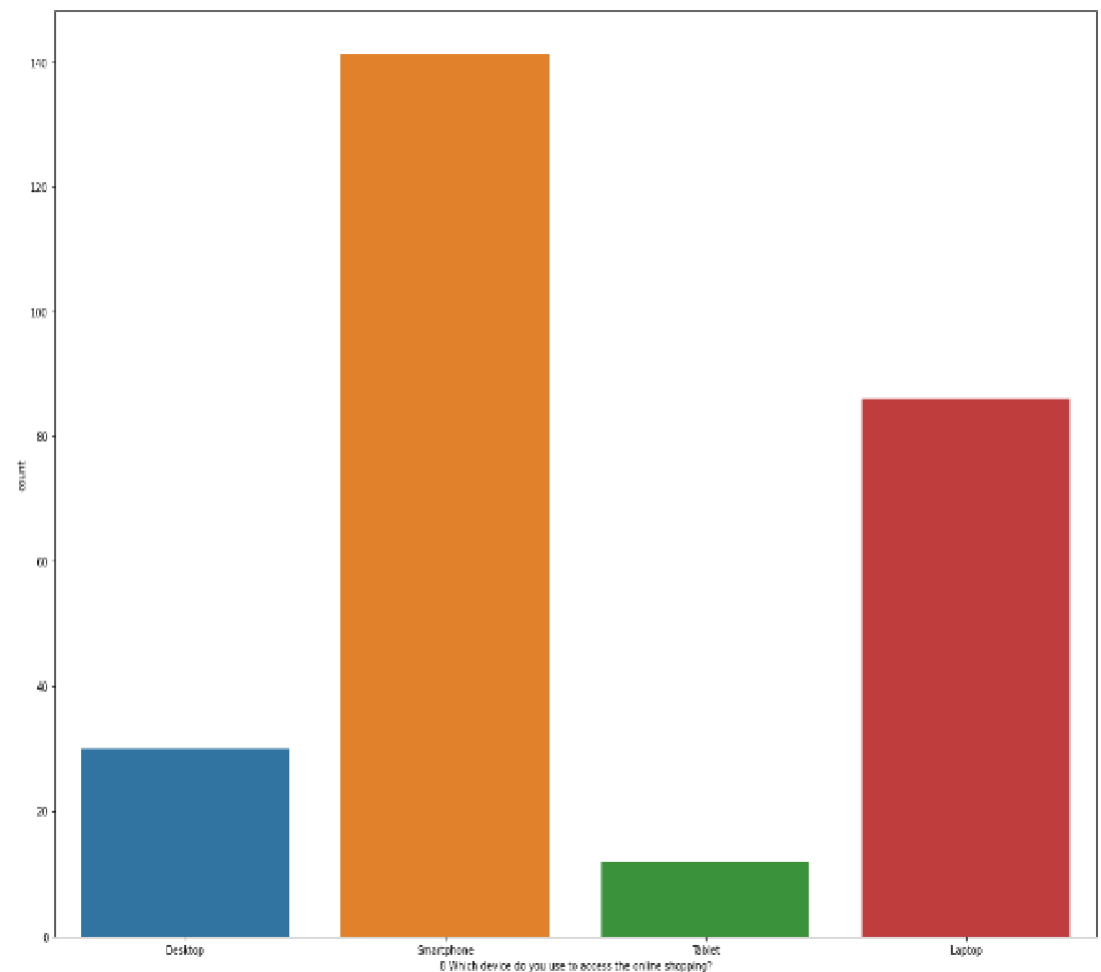

Smartphone 141 Peoples

Laptop 86 Peoples

Desktop 30 Peoples

Tablet 12 Peolpls

9 What is the screen size of your mobile device?

```
In [18]: 1 #Lets plot 9th Columns " What is the screen size of your mobile device?\t\t\t\t\t\t"
         2 plt.figure(figsize=(20,15))
         3 sns.countplot(x='9 What is the screen size of your mobile device?\t\t\t\t\t\t                    ",data=
         4 plt.savefig('9 What is the screen size of your mobile device.png')
         5 data["9 What is the screen size of your mobile device?\t\t\t\t\t\t                    '].value_counts()
         6
```
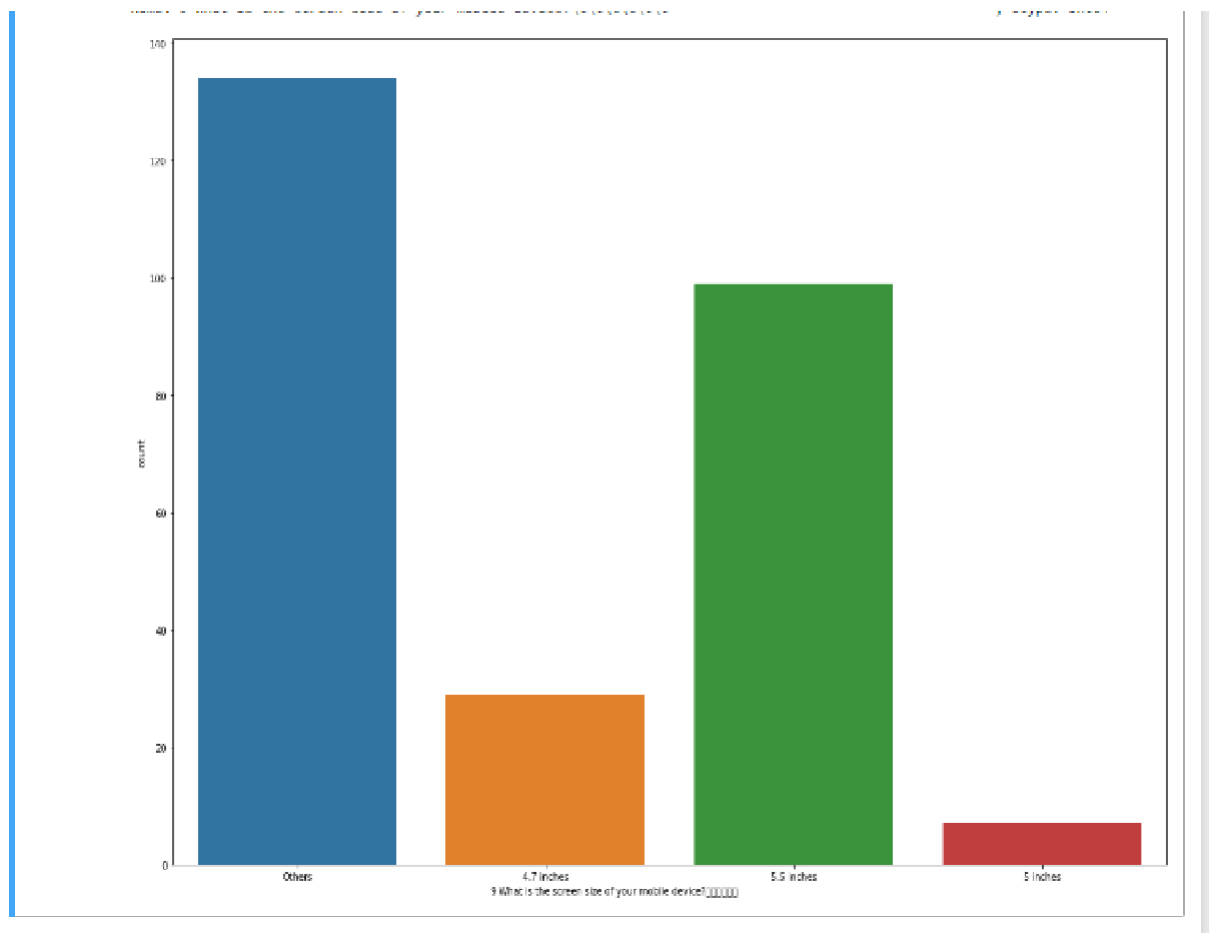

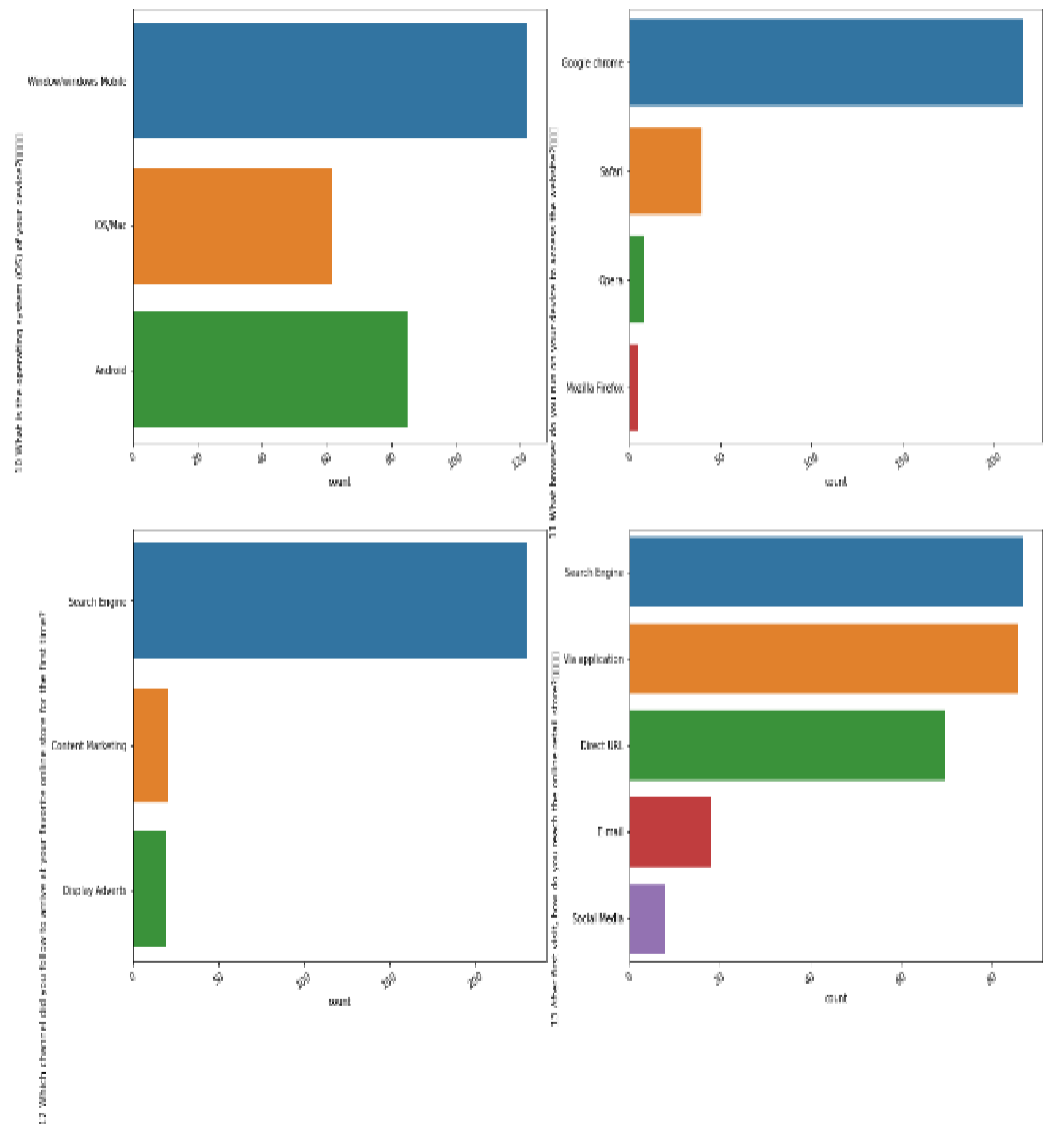Others 134 (Maybe its Computer)

5.5 inches 99

4.7 inches 29

5 inches 7
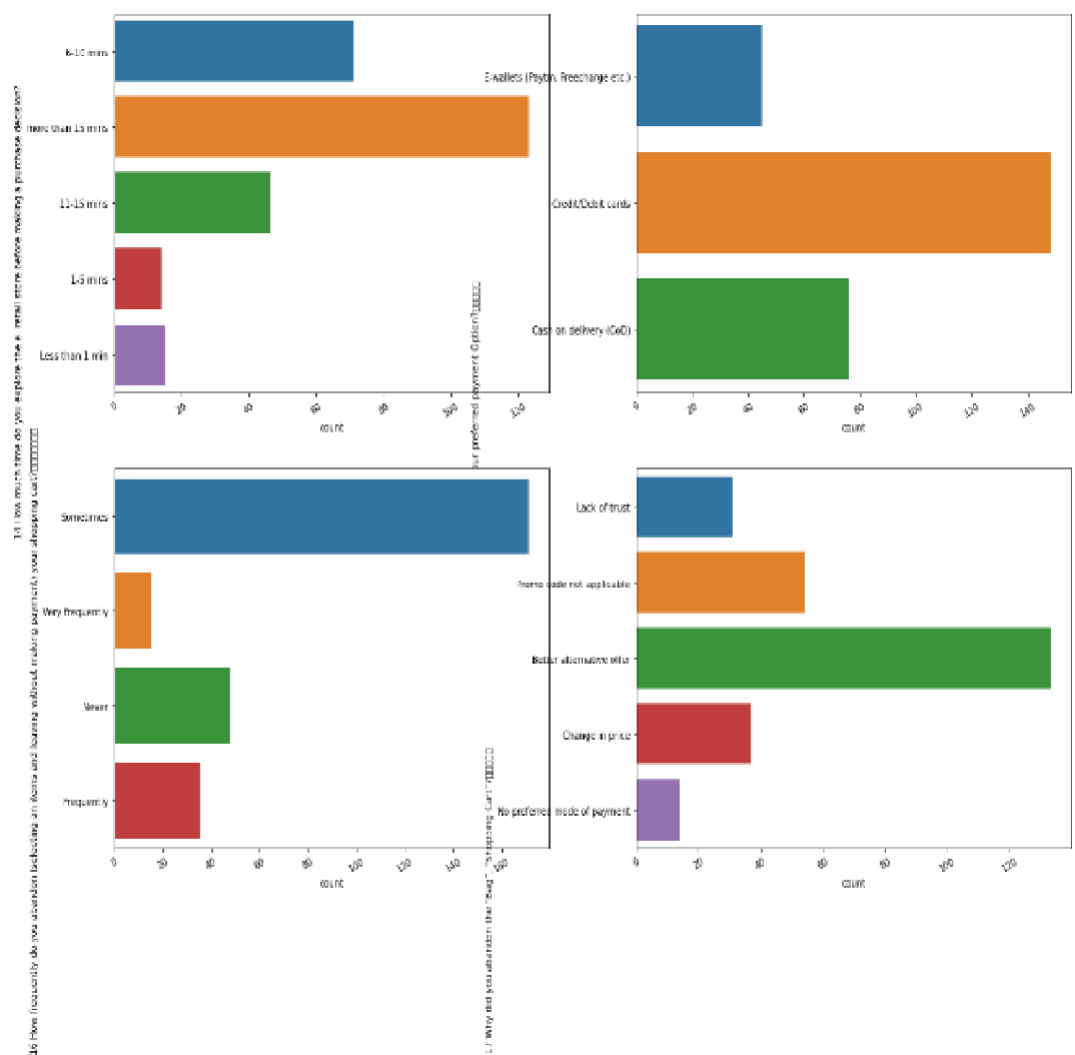
Plotting 4 Column in One Code..(9 to 13)

```
In [19]:   1  #Lets plot 9 to 13 columns in one code .
           2  plt.figure(figsize=(20,15))
           3  plotnumber=1
           4  for col in data.columns[9:13]:
           5      plt.subplot(2,2,plotnumber)
           6      plotnumber +=1
           7      ax=sns.countplot(y=data[col])
           8      plt.xticks(rotation=30)
           9  plt.savefig('9to13.png')
```
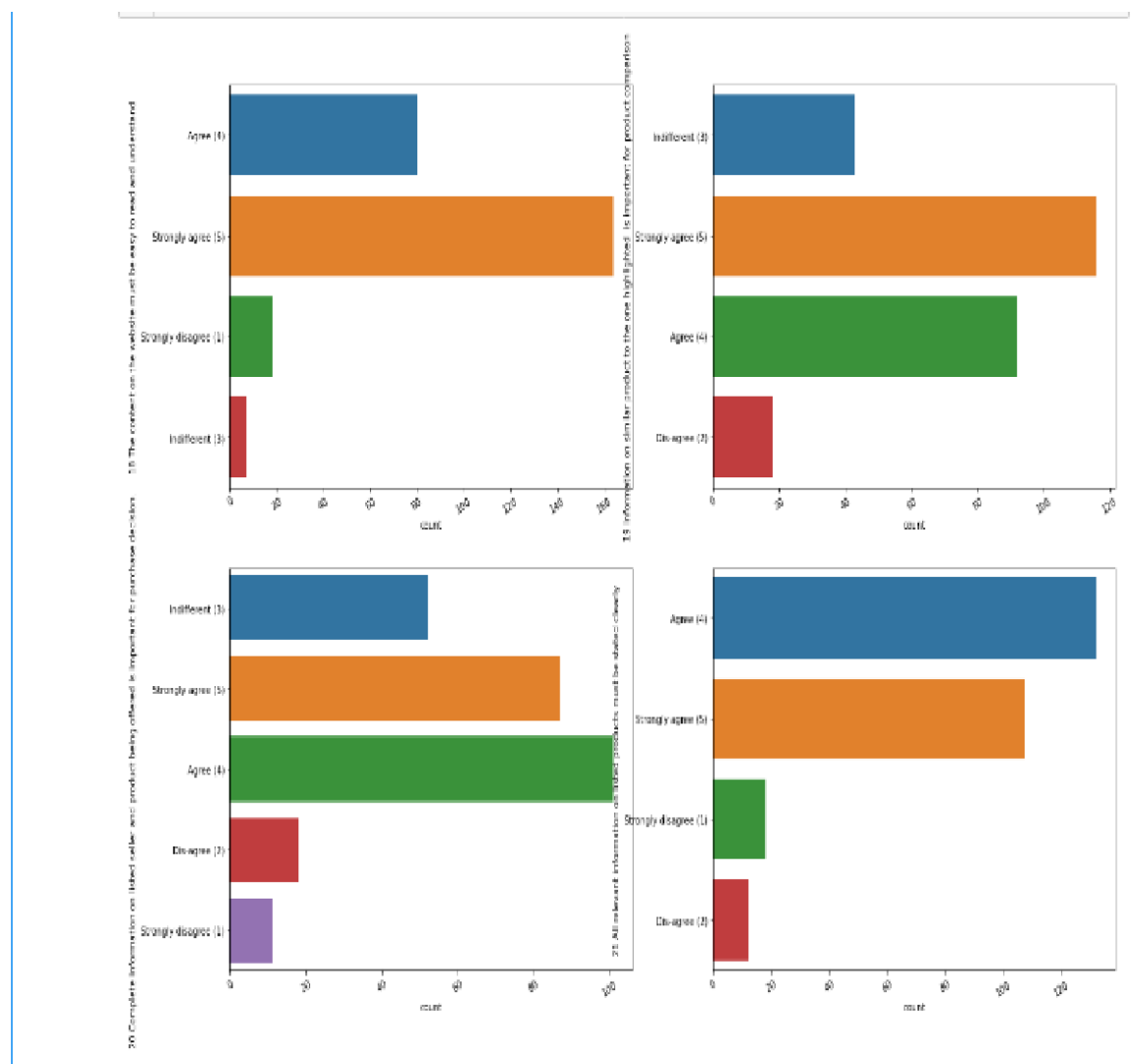
Lets plot 13 to 17 columns in one code .

```
In [20]:   1  #Lets plot 13 to 17 columns in one code .
           2  plt.figure(figsize=(20,15))
           3  plotnumber=1
           4  for col in data.columns[13:17]:
           5      plt.subplot(2,2,plotnumber)
           6      plotnumber +=1
           7      ax=sns.countplot(y=data[col])
           8      plt.xticks(rotation=30)
           9  plt.savefig('13to17.png')
          10
```
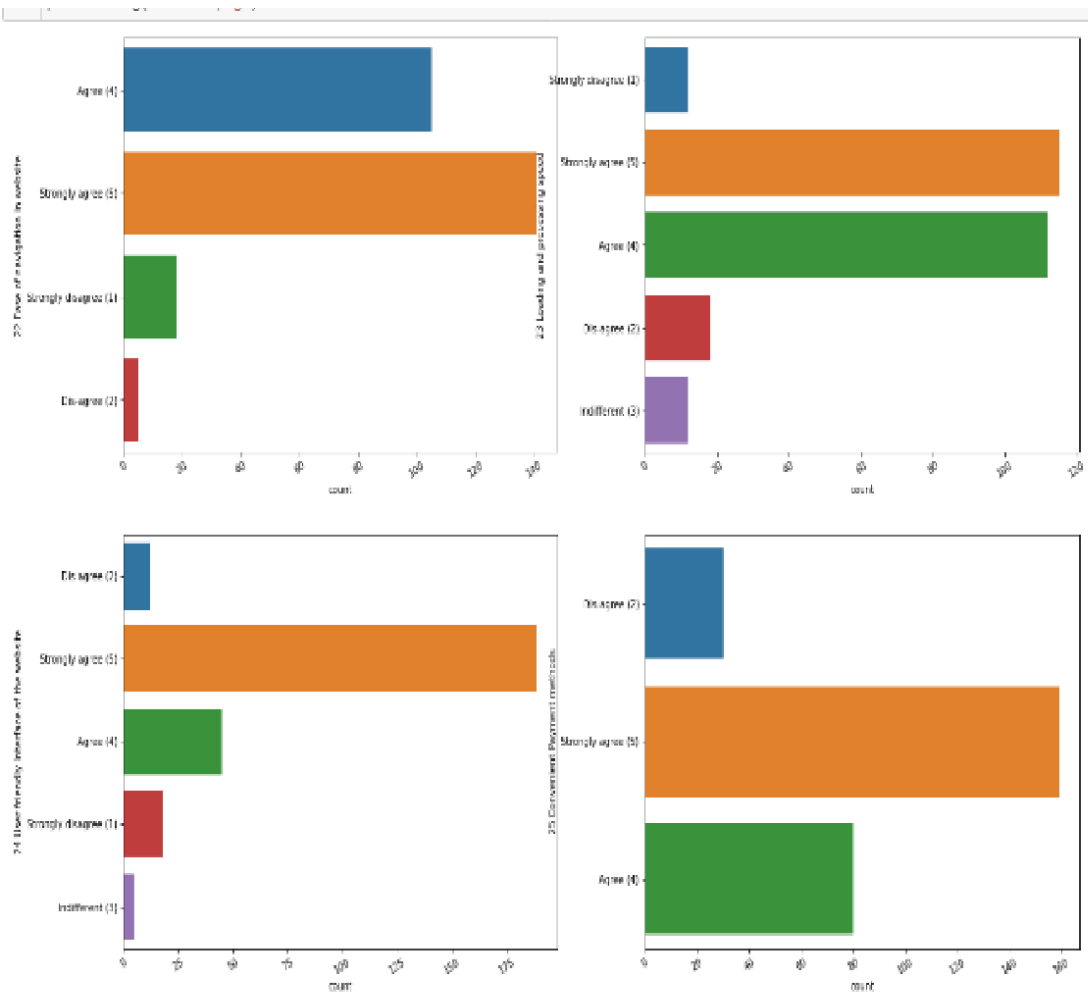
Lets plot 17 to 21 columns in one code

```
In [21]:   1  #Lets plot 17 to 21 columns in one code
           2  plt.figure(figsize=(20,15))
           3  plotnumber=1
           4  for col in data.columns[17:21]:
           5      plt.subplot(2,2,plotnumber)
           6      plotnumber +=1
           7      ax=sns.countplot(y=data[col])
           8
           9      plt.xticks(rotation=30)
          10  plt.savefig('17to21.png')
```

Lets plot 21 to 25 columns in one code
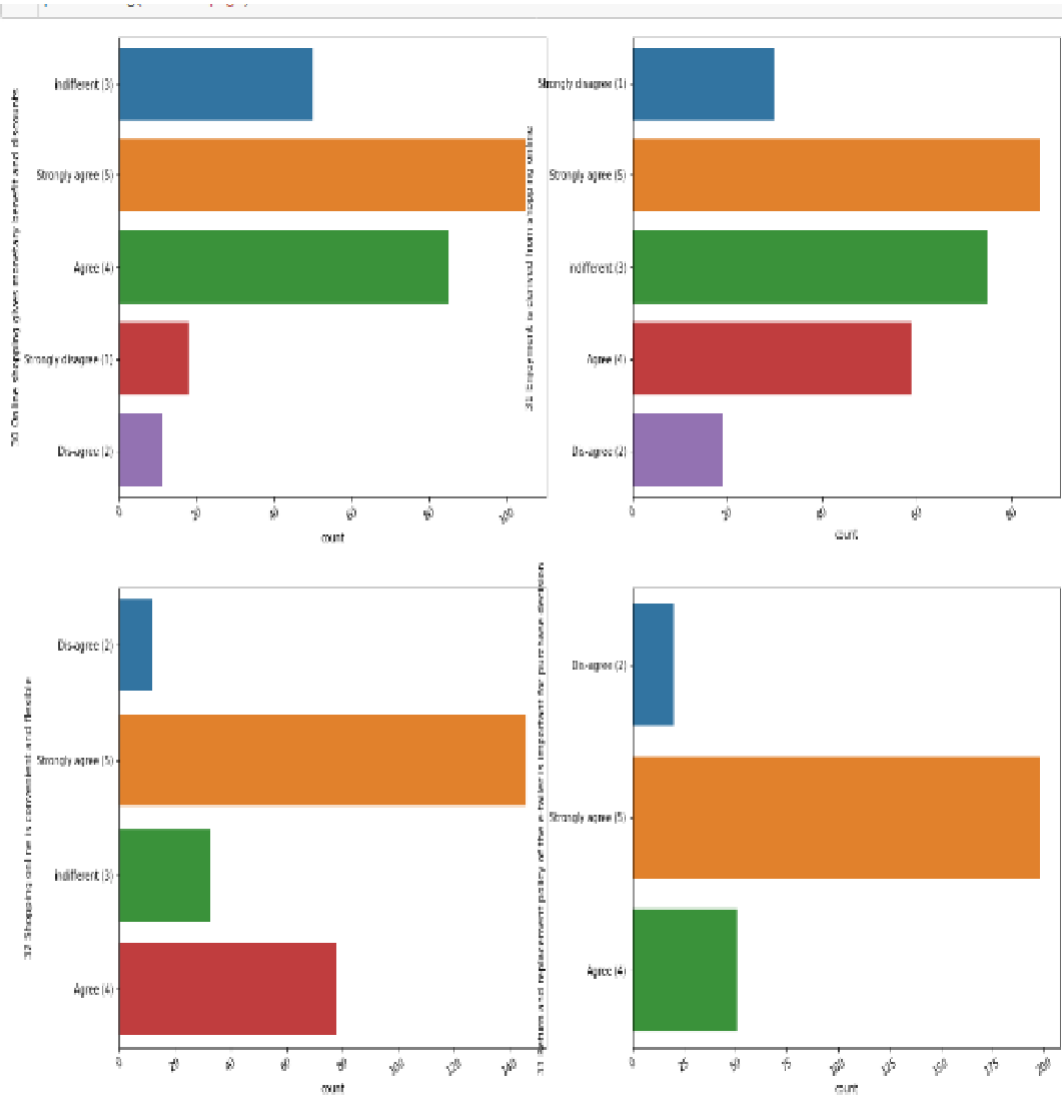
```
In [22]:   1  #Lets plot 21 to 25 columns in one code .
           2  plt.figure(figsize=(20,15))
           3  plotnumber=1
           4  for col in data.columns[21:25]:
           5      plt.subplot(2,2,plotnumber)
           6      plotnumber +=1
           7      ax=sns.countplot(y=data[col])
           8      plt.xticks(rotation=30)
           9  plt.savefig('21to26.png')
```

Lets plot 29 to 33 columns in one code
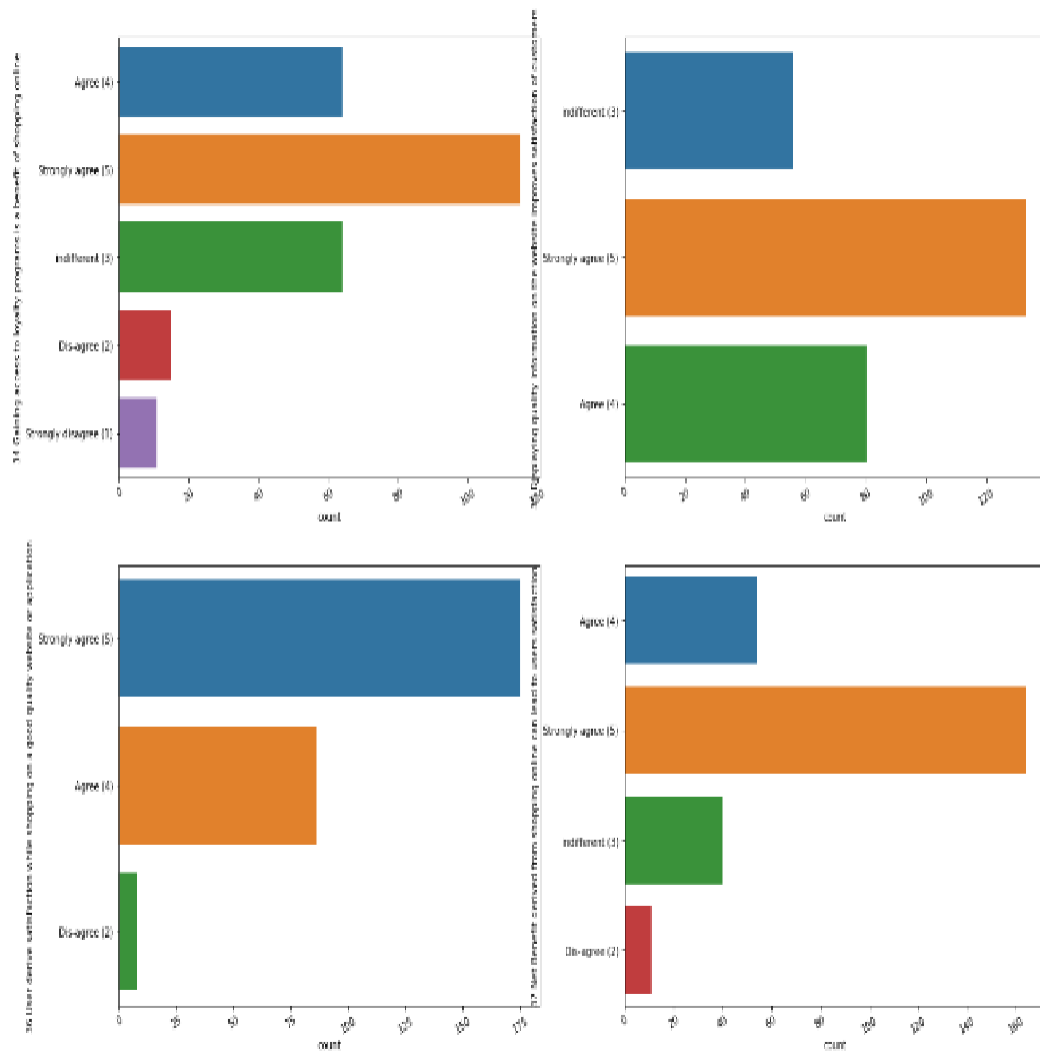
```
In [24]:   1  #Lets plot 29 to 33 columns in one code .
           2  plt.figure(figsize=(20,15))
           3  plotnumber=1
           4  for col in data.columns[29:33]:
           5      plt.subplot(2,2,plotnumber)
           6      plotnumber +=1
           7      ax=sns.countplot(y=data[col])
           8      plt.xticks(rotation=30)
           9  plt.savefig('29to33.png')
```
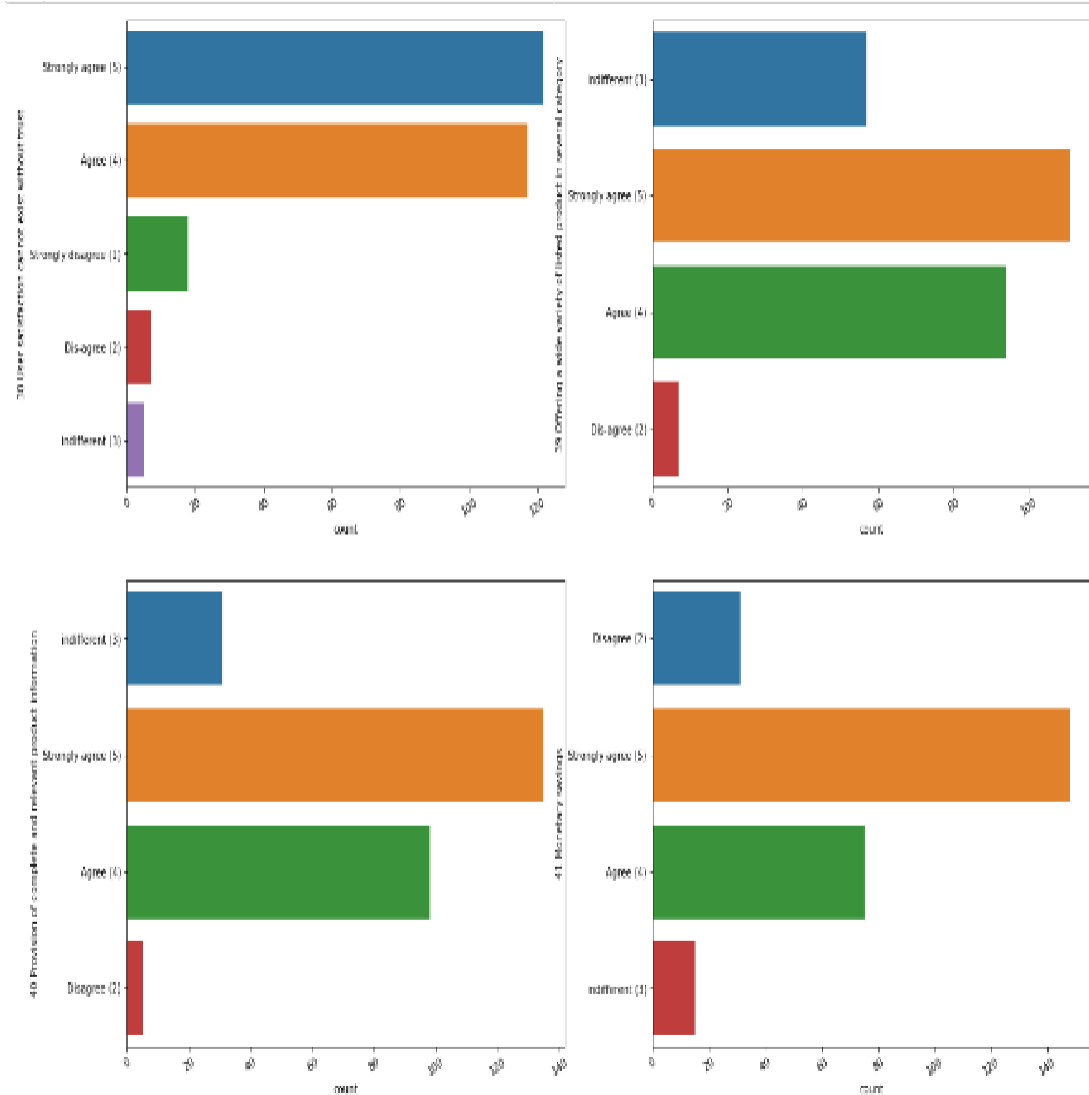
Lets plot 33 to 37 columns in one code .

```
In [25]:   1  #Lets plot 33 to 37 columns in one code .
           2  plt.figure(figsize=(20,15))
           3  plotnumber=1
           4  for col in data.columns[33:37]:
           5      plt.subplot(2,2,plotnumber)
           6      plotnumber +=1
           7      ax=sns.countplot(y=data[col])
           8      plt.xticks(rotation=30)
           9  plt.savefig('33to37.png')
```
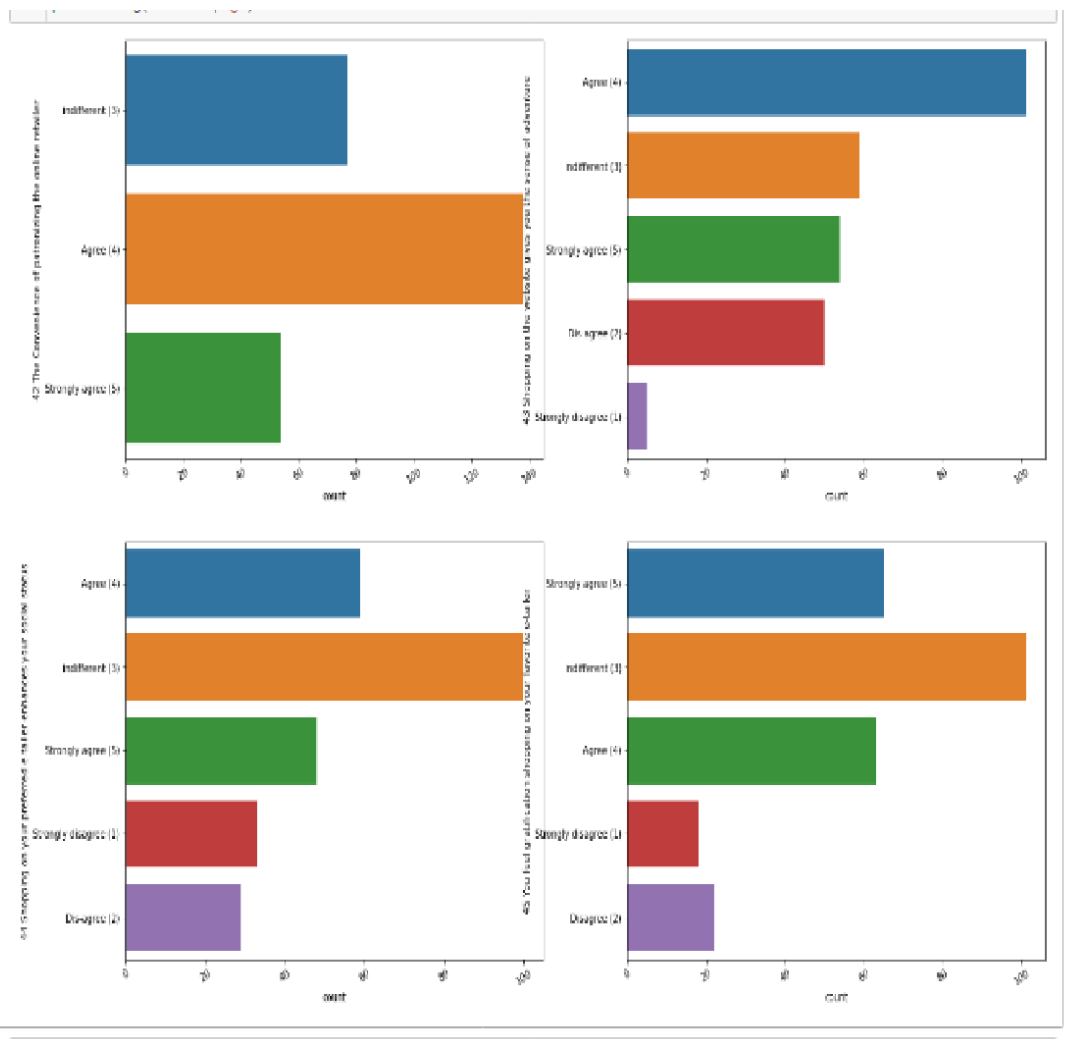
Lets plot 37 to 41 columns in one code .

```
In [26]:   1  #Lets plot 37 to 41 columns in one code .
           2  plt.figure(figsize=(20,15))
           3  plotnumber=1
           4  for col in data.columns[37:41]:
           5      plt.subplot(2,2,plotnumber)
           6      plotnumber +=1
           7      ax=sns.countplot(y=data[col])
           8      plt.xticks(rotation=30)
           9  plt.savefig('37to41.png')
```
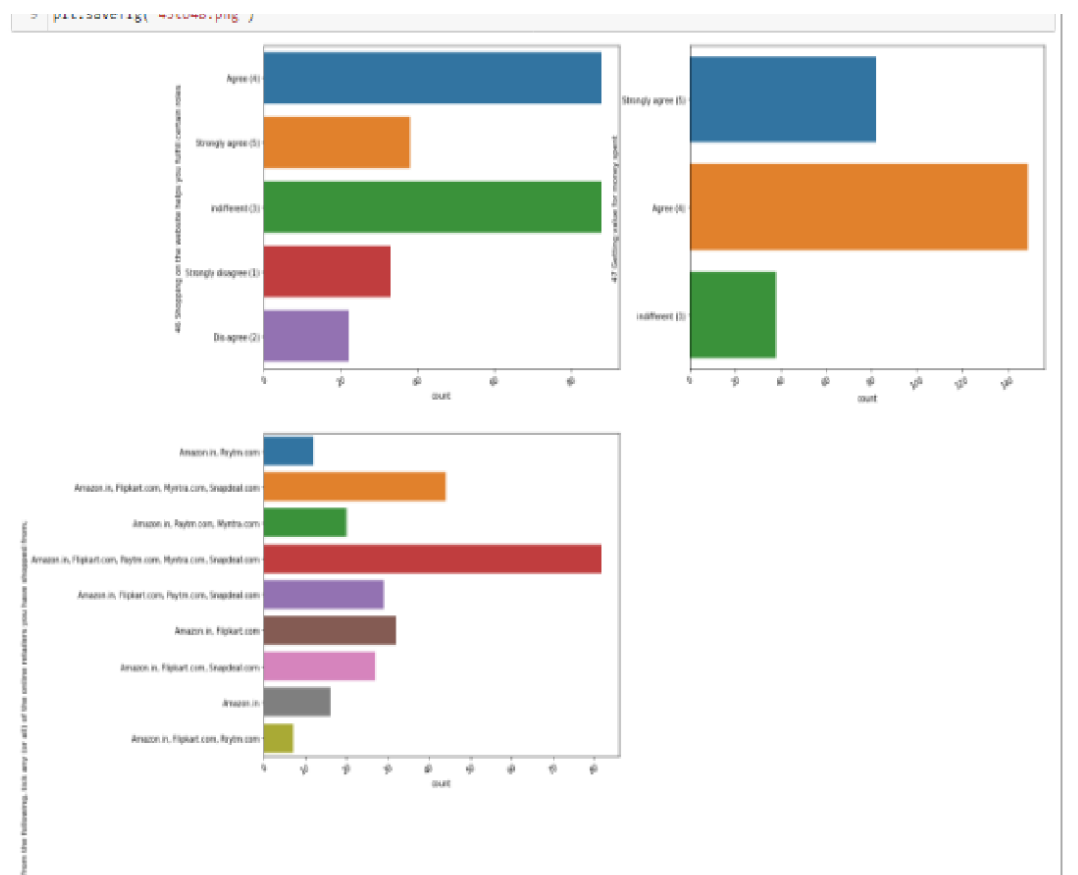
Lets plot 41 to 45 columns in one code .

```
In [27]:   1  #Lets plot 41 to 45 columns in one code .
           2  plt.figure(figsize=(20,15))
           3  plotnumber=1
           4  for col in data.columns[41:45]:
           5      plt.subplot(2,2,plotnumber)
           6      plotnumber +=1
           7      ax=sns.countplot(y=data[col])
           8      plt.xticks(rotation=30)
           9  plt.savefig('41to45.png')
```

Lets plot 45 to 48 columns in one code

```
28]:   1  #Lets plot 45 to 48 columns in one code .
       2  plt.figure(figsize=(20,15))
       3  plotnumber=1
       4  for col in data.columns[45:48]:
       5      plt.subplot(2,2,plotnumber)
       6      plotnumber +=1
       7      ax=sns.countplot(y=data[col])
       8      plt.xticks(rotation=30)
       9  plt.savefig('45to48.png')
```

Lets plot 48 to 52 columns in one code .

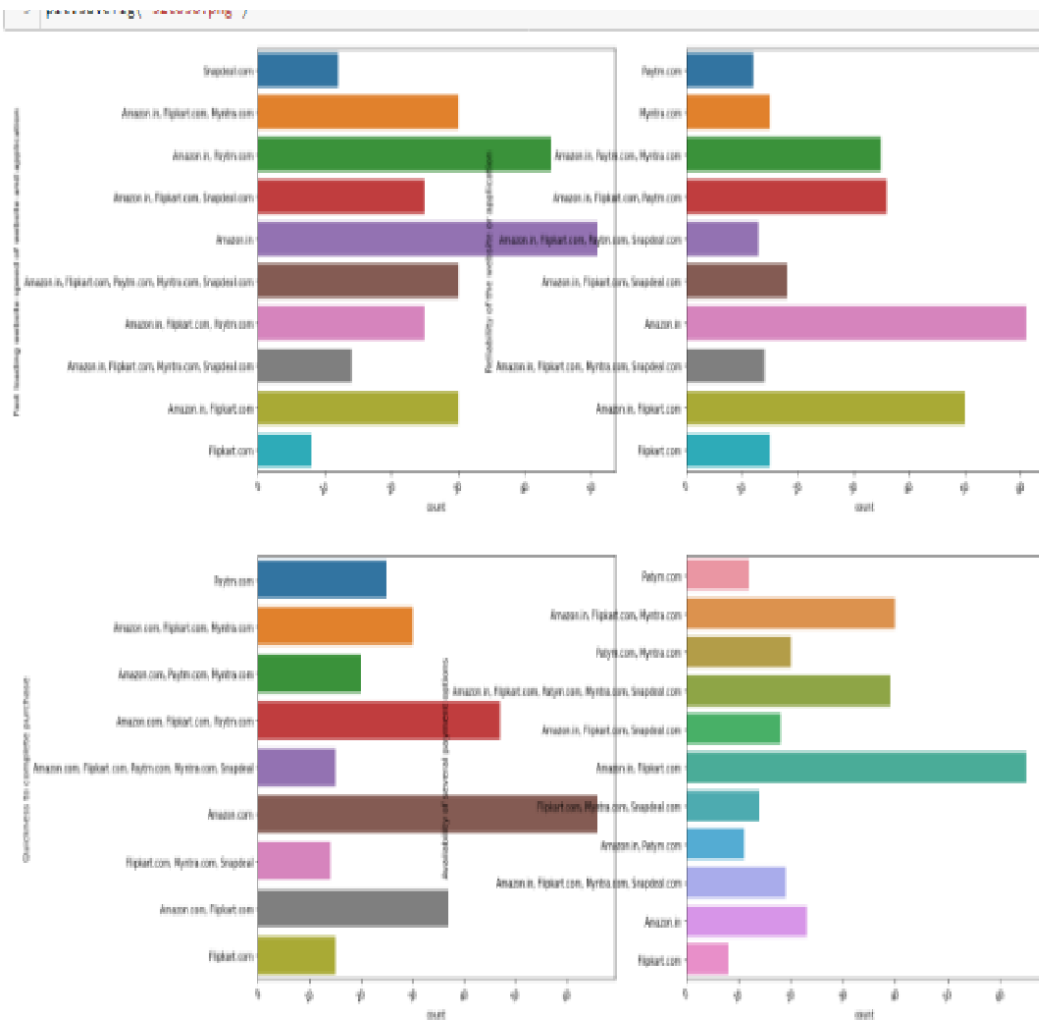```
In [29]:   1 #Lets plot 48 to 52 columns in one code .
           2 plt.figure(figsize=(20,15))
           3 plotnumber=1
           4 for col in data.columns[48:52]:
           5   plt.subplot(2,2,plotnumber)
           6   plotnumber +=1
           7   ax=sns.countplot(y=data[col])
           8   plt.xticks(rotation=30)
           9 plt.savefig('48to52.png')
```

## Lets plot 52 to 56 columns in one code
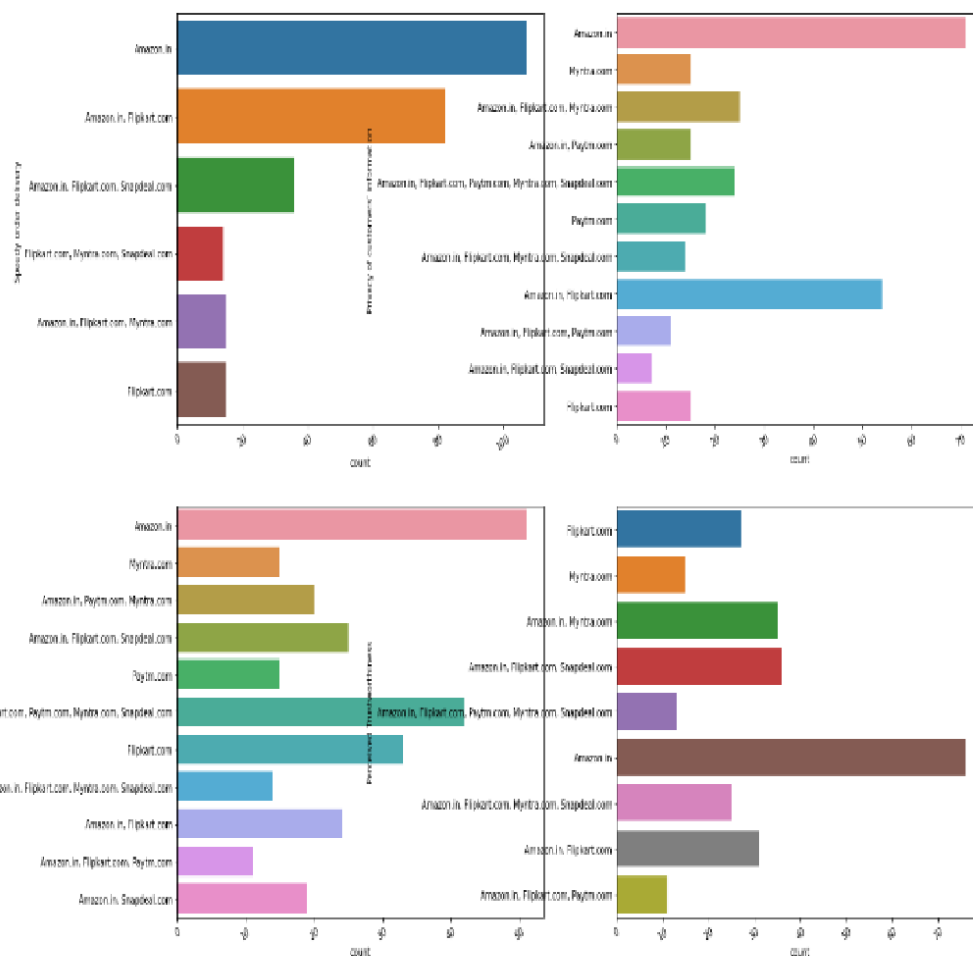


```
In [30]: 1 #Lets plot 52 to 56 columns in one code .
          2 plt.figure(figsize=(20,15))
          3 plotnumber=1
          4 for col in data.columns[52:56]:
          5   plt.subplot(2,2,plotnumber)
          6   plotnumber +=1
          7   ax=sns.countplot(y=data[col])
          8   plt.xticks(rotation=30)
          9 plt.savefig('52to56.png')
```
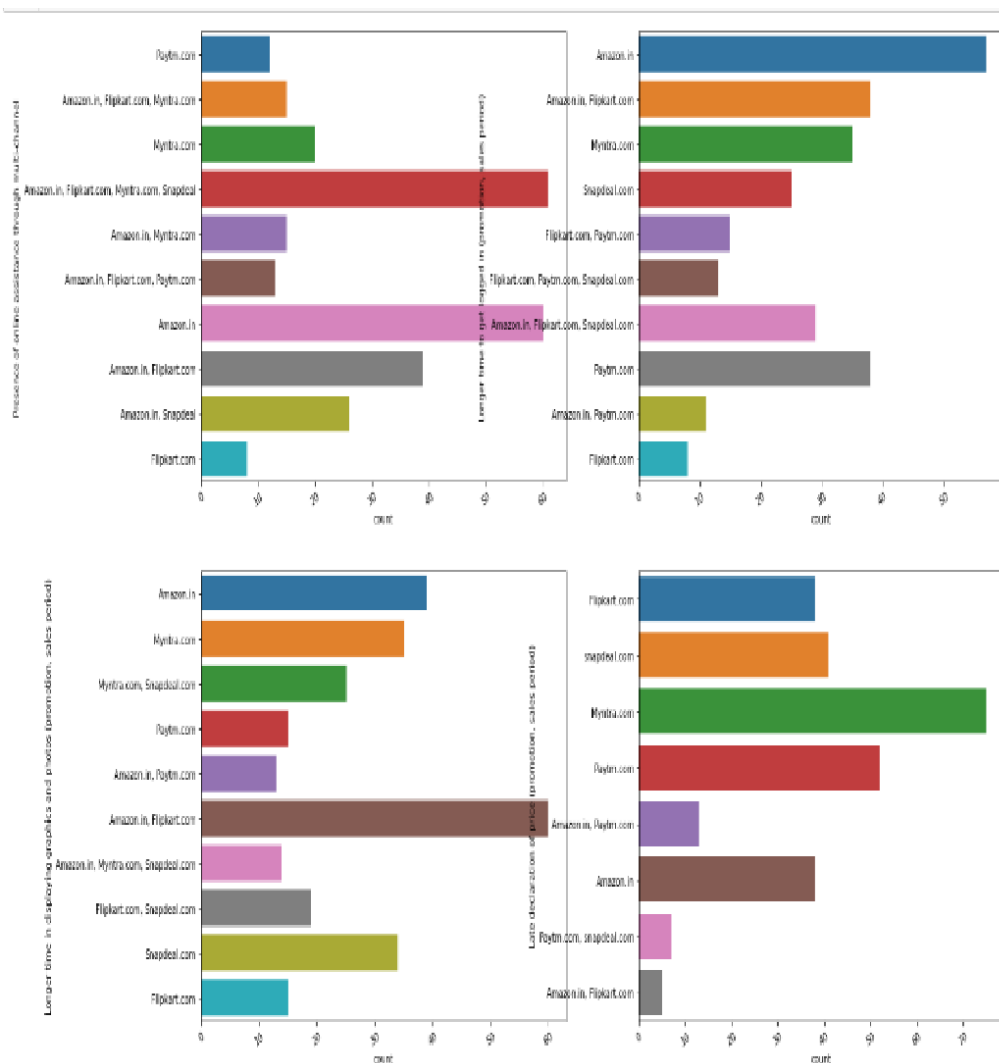
## Lets plot 56 to 60 columns in one code



```
In [31]:   1  #Lets plot 56 to 60 columns in one code .
           2  plt.figure(figsize=(20,15))
           3  plotnumber=1
           4  for col in data.columns[56:60]:
           5      plt.subplot(2,2,plotnumber)
           6      plotnumber +=1
           7      ax=sns.countplot(y=data[col])
           8      plt.xticks(rotation=30)
           9  plt.savefig('56to60.png')
```

Lets plot 60 to 64 columns in one code .
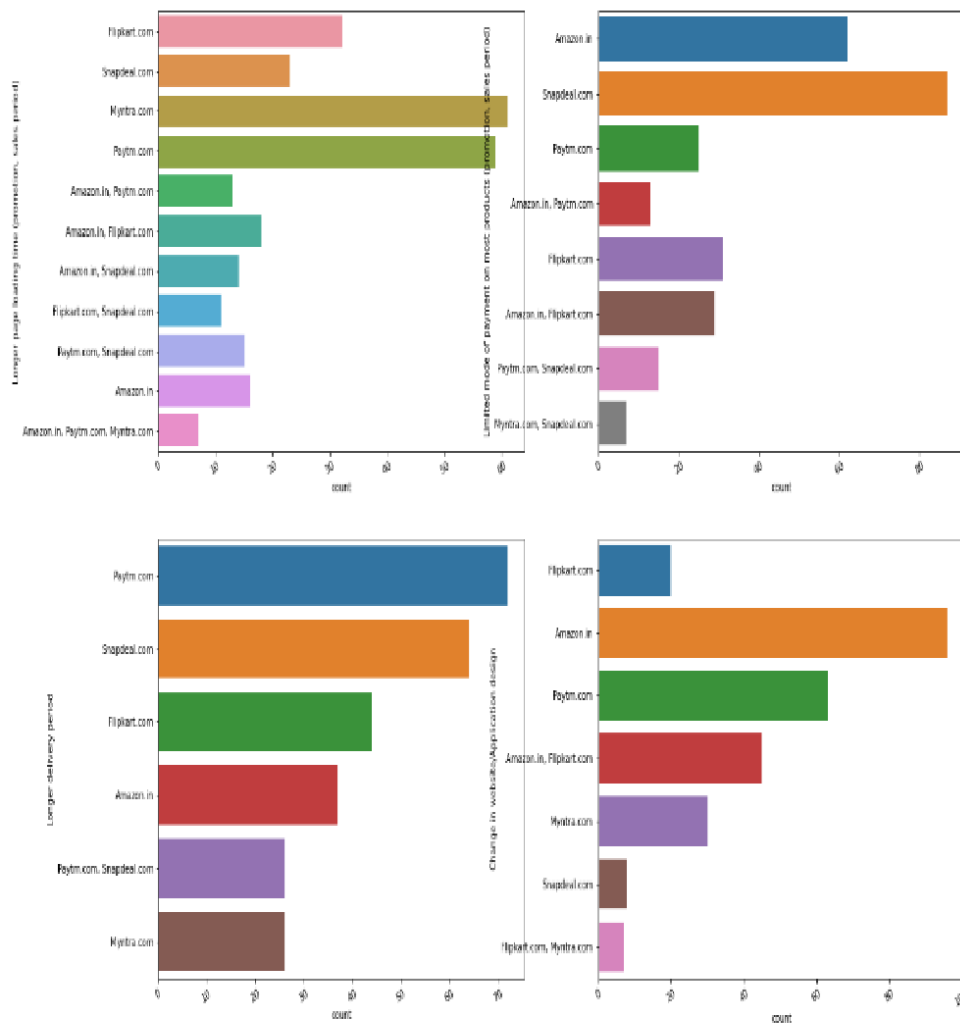
```
In [32]:   1  #Lets plot 60 to 64 columns in one code .
           2  plt.figure(figsize=(20,15))
           3  plotnumber=1
           4  for col in data.columns[60:64]:
           5      plt.subplot(2,2,plotnumber)
           6      plotnumber +=1
           7      ax=sns.countplot(y=data[col])
           8      plt.xticks(rotation=30)
           9  plt.savefig('60to64.png')
```

Lets plot 64 to 68 columns in one code.

count

```
In [33]:  1  #Lets plot 64 to 68 columns in one code .
          2  plt.figure(figsize=(20,15))
          3  plotnumber=1
          4  for col in data.columns[64:68]:
          5      plt.subplot(2,2,plotnumber)
          6      plotnumber +=1
          7      ax=sns.countplot(y=data[col])
          8      plt.xticks(rotation=30)
          9  plt.savefig('64to68.png')
```

Lets plot 68 to 72 columns in one code .

```
In [34]:   1  #Lets plot 68 to 72 columns in one code .
           2  plt.figure(figsize=(20,15))
           3  plotnumber=1
           4  for col in data.columns[68:72]:
           5      plt.subplot(2,2,plotnumber)
           6      plotnumber +=1
           7      ax=sns.countplot(y=data[col])
           8      plt.xticks(rotation=30)
           9  plt.savefig('68to72.png')
```
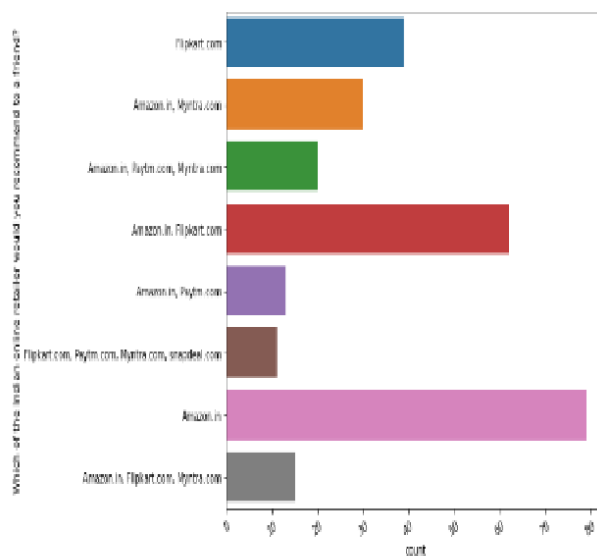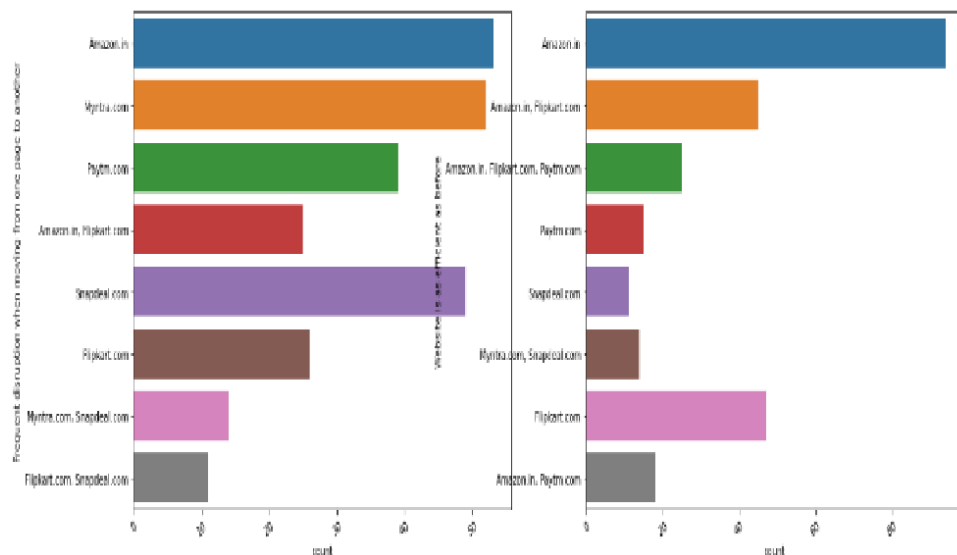
Lets plot 'Which of the Indian online retailer would you recommend to a friend?'

```python
#Lets plot 'Which of the Indian online retailer would you recommend to a friend?'
plt.figure(figsize=(20,15))
import matplotlib.pyplot as plt
import seaborn as sns
sns.countplot(y=data['Which of the Indian online retailer would you recommend to a friend?'])
plt.show()
```
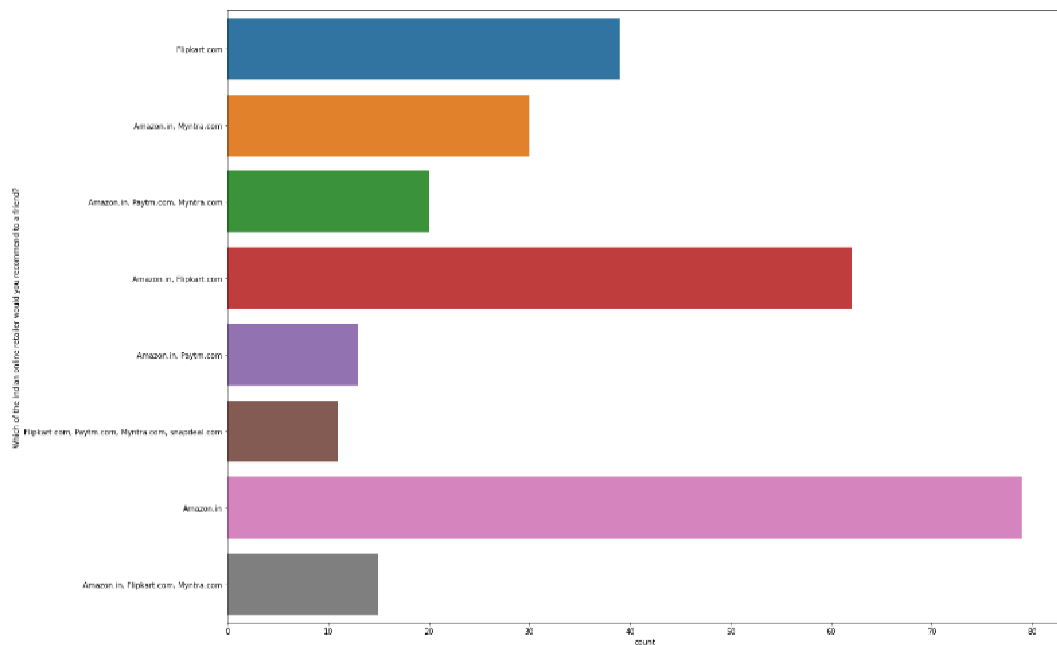

CONCLUSION

- **Amazon is recommended by 81.4% of the customers.**

- **Snapdeal is recommended by 4.1% of the customers.**

- **Flipkart is recommended by 47.2 of the customers.**

- **Myntra is recommended by 28% of the customers.**

- **Paytm is recommended by 16% of the customers**

As according to the dataset best online retailer they would recommend to a friend . Most of the respondents says Amazon because . It is providing all the feature that user want . Website is efficient and it is fast loading . it give complete relevant description and information of products. it is reliable and quick to complete the purchase . Amazom give speedy delivey to some places . customer and there is several payment option available . on the website . it provide online assistance through multichannels. providing good details on product . Amazon user friendly interface and has visual appealing webpage layout . Amazon also offers wide variety of product and its application is easy to use . Lastly the main thing why user recommend it is because of its trustworthiness and also its robust security in protecting customer financial information and their Privacy Information . Frequent distribution when moving from one page to another this company should see and improve it to give overall best experiance to the users . This paper investigated the factor that influence the online customer repeat purchase intenstion . During the process various data processing method has been used to clean the data . The project conatains EDA considering every aspect . The major finding is complete relevant description if products and reliability of the website increase chances of customer retetion. Howevee , if the content on the website is not easy to read and understand or can't guarantee the privacy of the customer will lead to chances of the customer retention . This project has increased my understanding of the concept . During the research i cam across various challenge and while solving them i learned a lot of new thing . How to plot different charts . . For example i learned how to plot subplot . How to How to group data and visualize that.