FLIP ROBO

NAME OF THE PROJECT

“Customer retention case study”

Submitted by:

Leena Chatterjee

ACKNOWLEDGMENT:

- I have taken efforts in this project. However, it would not have been possible without the kind support and help of each individual of DATA TRAINED organizations. I would like to extend my sincere thanks to all of them.
- I am highly indebted to all team of Data trained for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project.

- I would like to express my special gratitude and thanks to my mentor for guiding for this project

Bibliography:

- https://www.searchstartnow.com/web?qo=semQuery&ad=semA&q=git hub%20for%20beginners&o=1468511&ag=fw4&an=msn_s&adid=79096 273683706&agid=1265538600064869&campaignid=416218294&clickid =57fa256a7fcb1776d83445cf499fe6e2&clid=aj-shopnet-it&kwid=kwd79096564506817%3Aloc-90&rch=intl835&utm_medium=bcpc&utm_source=b
- https://www.kaggle.com/learn

INTRODUCTION

Problem Statement:

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

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.

.

We are working on this dataset where we need to analyse the dataset to find the factor which all are factor which is making customer to repurchase product from same website.

The sample data is provided to us for academic purpose. In order to improve the factors, we need to analysis the dataset which is playing vital role to hold the customer so here we will be analysis the data based on customer feedback. In this dataset target variable **is' How many times you have made an online purchase in the past 1 year'** which will represent value as 1 to 5 based on numbers of time customer made purchase throughout the year . **Label '1' indicates that least headcount of people according to number of purchase Label '5' indicates Less than 10 times (maximum people purchased)**

Conceptual Background of the Domain Problem

Data science is the field where we can predict the probability. Here basically we need to analyse   the factor which will be helping all ecommerce website to grow their business

Basic summery

E-commerce refers to the process of buying or selling products or services over the Internet. Online shopping is becoming increasingly popular because of speed and ease of use for customers. E-commerce activities such as selling online can be directed at consumers or other businesses.

All e-commerce platforms should consider implementing a churn model to add value to their businesses as it is a bare essential component for customer retention. Customer retention is the ability of a company to retain its customers over a specified period of time.

So here we have consider 1 year time , in this dataset we do have some specific factor which play vital role to make customer repurchase , basically we need to analyse we positive factor for holding customer and need to find out all negative factor which caused to lose the customer, so that E-commerce can minimize the error for which their business may run in loss

Review of Literature

AS mentioned here output will be 0 to 5 based on that we will find reliable customer who had made shopping n number of times from the same ecommerce platform after the person made his/her 1st purchase. Output will depend on some variable for customer to choose the shopping website.

In this dataset those variables are ====□

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?

10 What is the operating system (OS) of your device?

11 What browser do you run on your device to access the website?

12 Which channel did you follow to arrive at your favorite online store for the first time?

13 After first visit, how do you reach the online retail store?

14 How much time do you explore the e- retail store before making a purchase decision?

15 What is your preferred payment Option?

16 How frequently do you abandon (selecting an items and leaving without making payment) your shopping cart?

17 Why did you abandon the “Bag”, “Shopping Cart”?

18 The content on the website must be easy to read and understand

19 Information on similar product to the one highlighted is important for product comparison

20 Complete information on listed seller and product being offered is important for purchase decision

21 All relevant information on listed products must be stated clearly

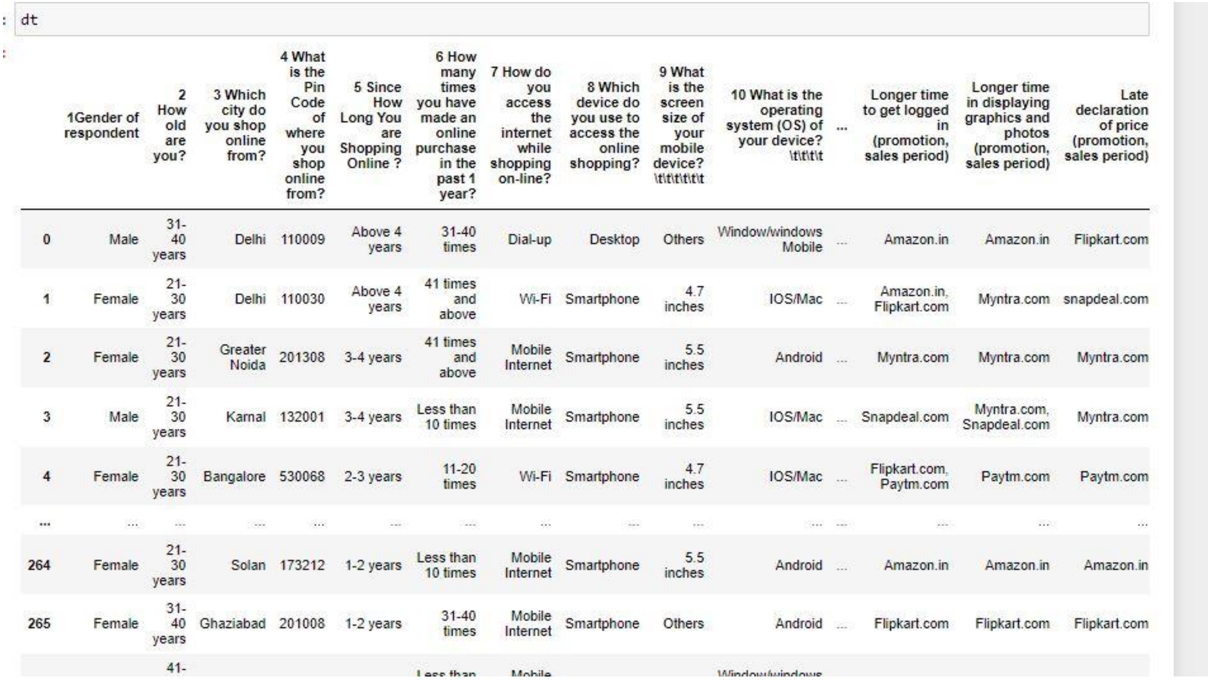
22 Ease of navigation in website
23 Loading and processing speed
24 User friendly Interface of the website
25 Convenient Payment methods
26 Trust that the online retail store will fulfill its part of the transaction at the stipulated time
27 Empathy (readiness to assist with queries) towards the customers
28 Being able to guarantee the privacy of the customer
29 Responsiveness, availability of several communication channels (email, online rep, twitter, phone etc.)
30 Online shopping gives monetary benefit and discounts
31 Enjoyment is derived from shopping online
32 Shopping online is convenient and flexible

33 Return and replacement policy of the e-tailer is important for purchase decision
34 Gaining access to loyalty programs is a benefit of shopping online
35 Displaying quality Information on the website improves satisfaction of customers
36 User derive satisfaction while shopping on a good quality website or application
37 Net Benefit derived from shopping online can lead to users satisfaction

38 User satisfaction cannot exist without trust
39 Offering a wide variety of listed product in several category
40 Provision of complete and relevant product information
41 Monetary savings
42 The Convenience of patronizing the online retailer
43 Shopping on the website gives you the sense of adventure
44 Shopping on your preferred e-tailer enhances your social status
45 You feel gratification shopping on your favorite e-tailer
46 Shopping on the website helps you fulfill certain roles
47 Getting value for money spent

- From the following, tick any (or all) of the online retailers you have shopped from;
- Easy to use website or application
- **Visual appealing web-page layout**
- Wild variety of product on offer
- Complete, relevant description information of products
- Fast loading website speed of website and application
- Reliability of the website or application
- Quickness to complete purchase
- Availability of several payment options
- Speedy order delivery
- Privacy of customers' information
- Security of customer financial information
- Perceived Trustworthiness
- Presence of online assistance through multi-channel
- 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)
- Limited mode of payment on most products (promotion, sales period)
- Longer delivery period
- Change in website/Application design
- Frequent disruption when moving from one page to another
- Website is as efficient as before
- Which of the Indian online retailer would you recommend to a friend?

Data Sources and their formats



```
: dt
```

	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\t	10 What is the operating system (OS) of your device? \t\t\t\t	...	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)
0	Male	31-40 years	Delhi	110009	Above 4 years	31-40 times	Dial-up	Desktop	Others	Window/windows Mobile	...	Amazon.in	Amazon.in	Flipkart.com
1	Female	21-30 years	Delhi	110030	Above 4 years	41 times and above	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	...	Amazon.in, Flipkart.com	Myntra.com	snapdeal.com
2	Female	21-30 years	Greater Noida	201308	3-4 years	41 times and above	Mobile Internet	Smartphone	5.5 inches	Android	...	Myntra.com	Myntra.com	Myntra.com
3	Male	21-30 years	Karnal	132001	3-4 years	Less than 10 times	Mobile Internet	Smartphone	5.5 inches	IOS/Mac	...	Snapdeal.com	Myntra.com, Snapdeal.com	Myntra.com
4	Female	21-30 years	Bangalore	530068	2-3 years	11-20 times	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	...	Flipkart.com, Paytm.com	Paytm.com	Paytm.com
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
264	Female	21-30 years	Solan	173212	1-2 years	Less than 10 times	Mobile Internet	Smartphone	5.5 inches	Android	...	Amazon.in	Amazon.in	Amazon.in
265	Female	31-40 years	Ghaziabad	201008	1-2 years	31-40 times	Mobile Internet	Smartphone	Others	Android	...	Flipkart.com	Flipkart.com	Flipkart.com
		41-				Less than	Mobile			Window/windows				

Dataset what we have received that is csv file. We saved the file in current working directory of our local system as csv file After that using panda.read_csv we uploaded to jupyter note book in df variable [Panda is in built library in jupyter Notebook

df.info()--- it provided object type of each columns .our dataset content of (269 rows , 71 columns)

2.df.dypes= its provided info that what the data type belongs to (float , in t)

3 df.isnull.sum()--- we found there is no null value

4 df.head()--- it shows first five columns in the dataset

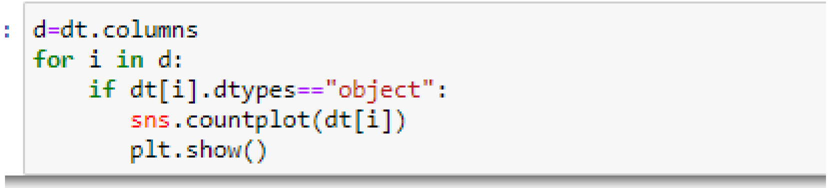
5 df.columns—it shows total columns of the dataset

Data visualization:

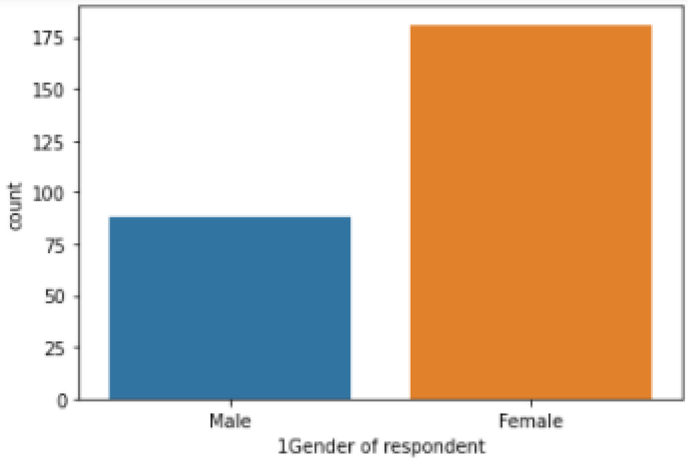
Now we will be plotting and seeing how features variables are related to Target

Univariate analysis :

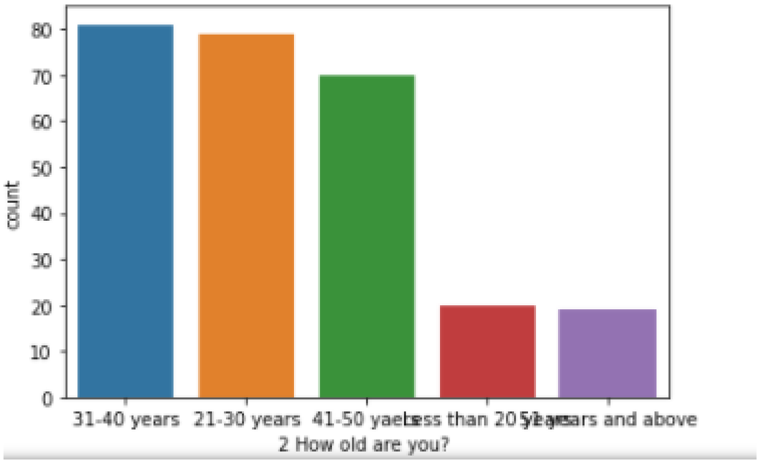
We checked dataset is object type dataset , through count plot we will be checking which category of data is giving us highest output :



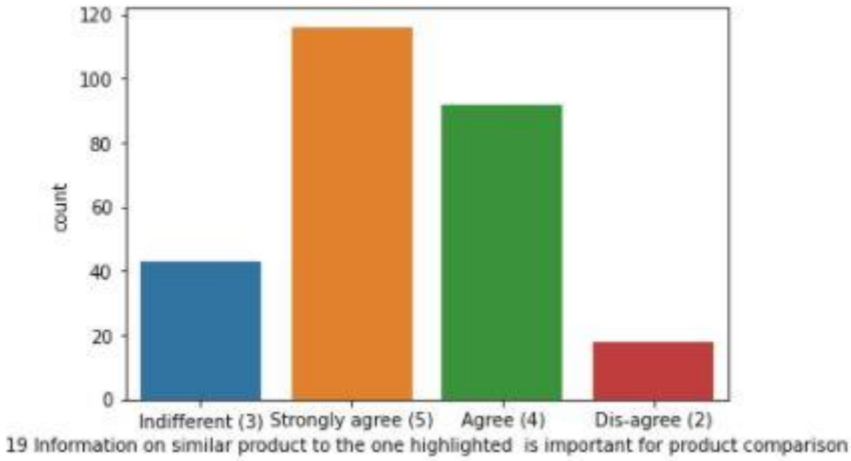
```
:  d=dt.columns
   for i in d:
       if dt[i].dtypes=="object":
           sns.countplot(dt[i])
           plt.show()
```



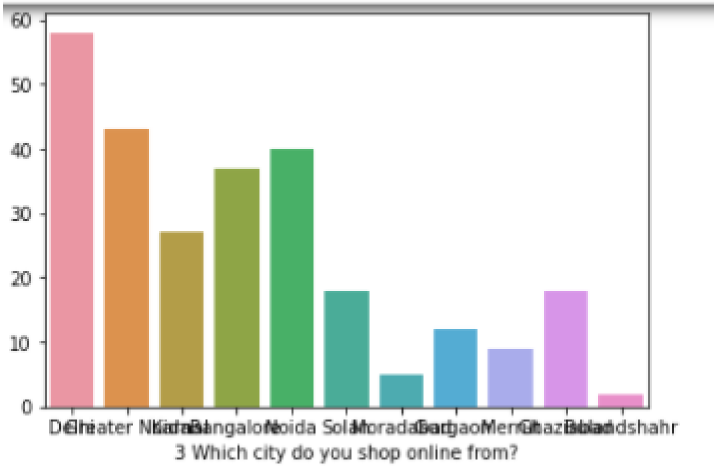
Observation – Here we found it is object column and found that it is having class imbalance issue , so female category giving highest value for shopping



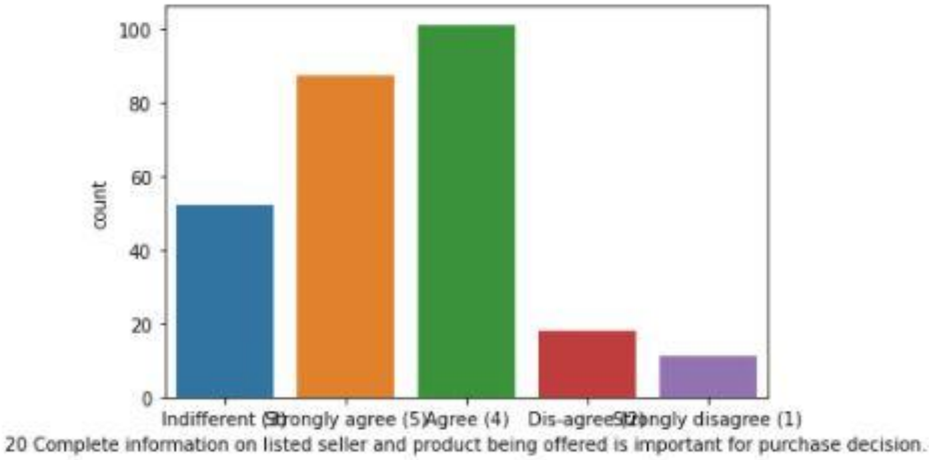
Observation – here 31-40 yrs category did highest shopping



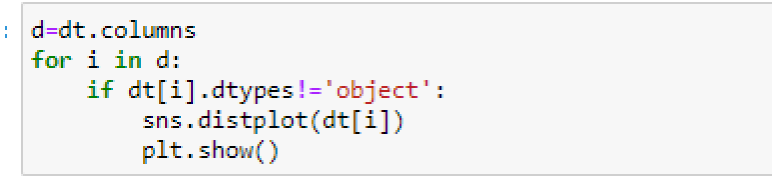
Observation— here maximum customers did strongly agree



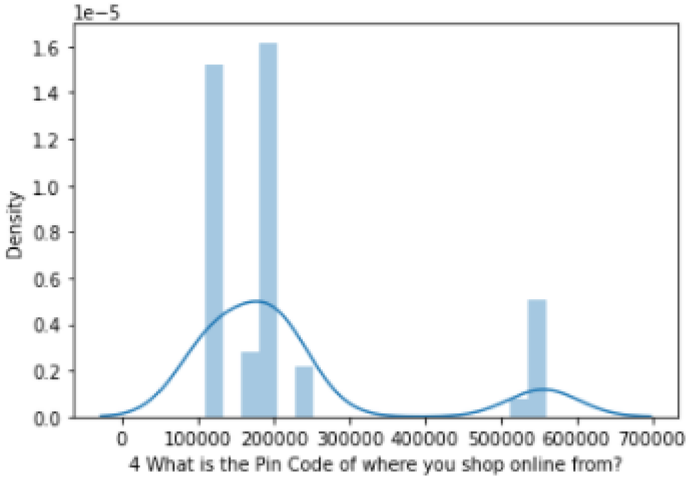
Observation – From Delhi people did maximum shopping and this column is also having class imbalance issue



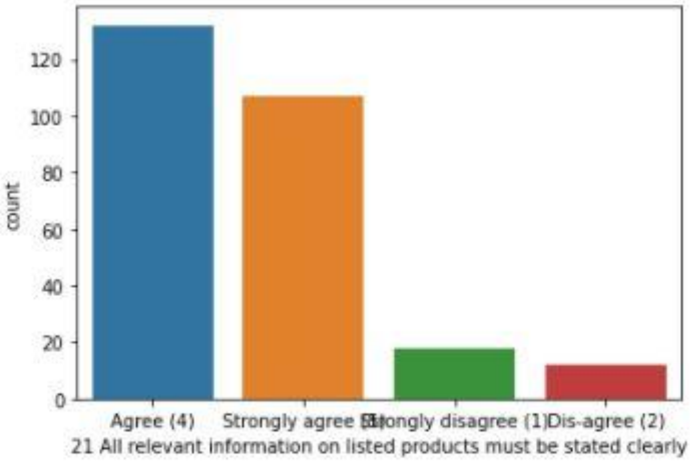
Observation—here maximum customers did agree



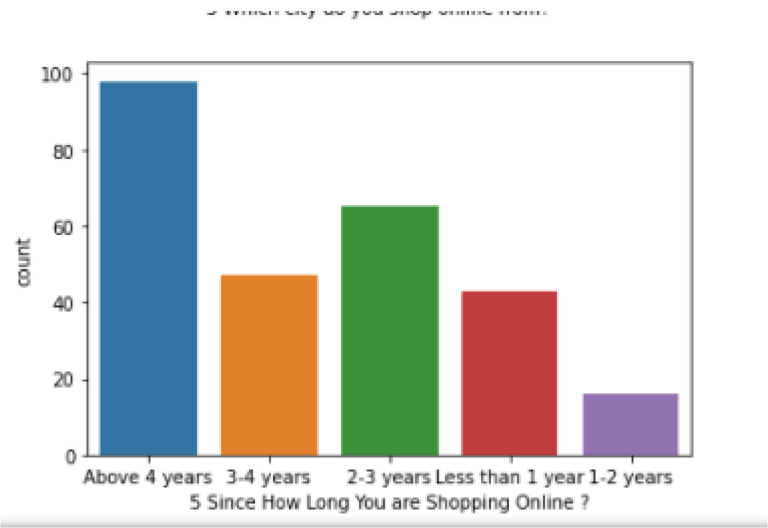
```
d=dt.columns
for i in d:
    if dt[i].dtypes!='object':
        sns.distplot(dt[i])
        plt.show()
```



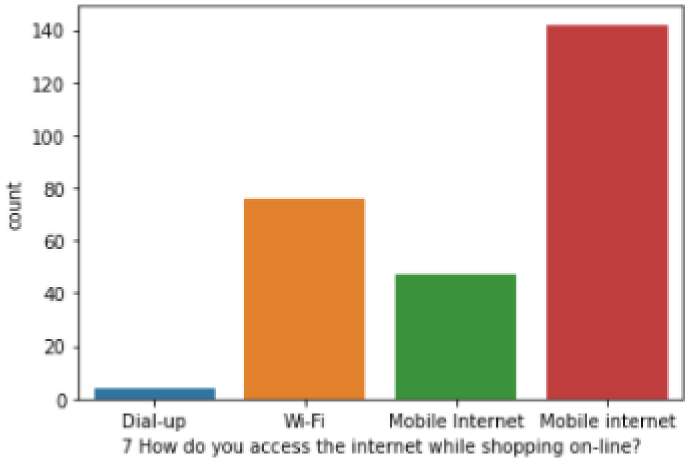
Observation-- **It is not a object type of column , it is numeric column , we used didplot and data is not normally distributed**



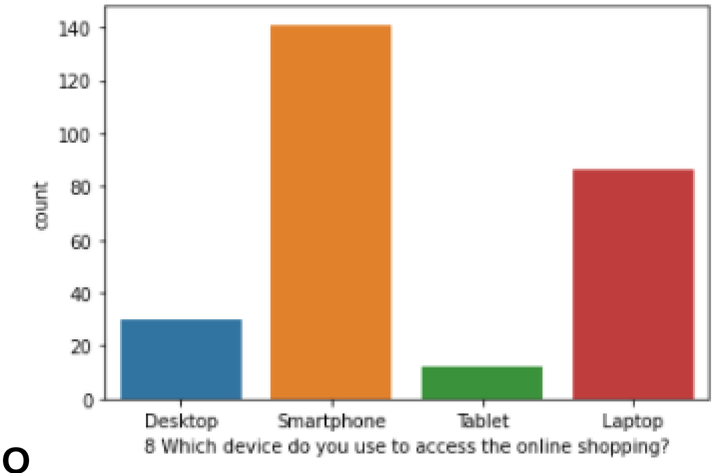
Observation--here maximum customers did agree



Observation—people who are shopping above 4 years they are having highest vote

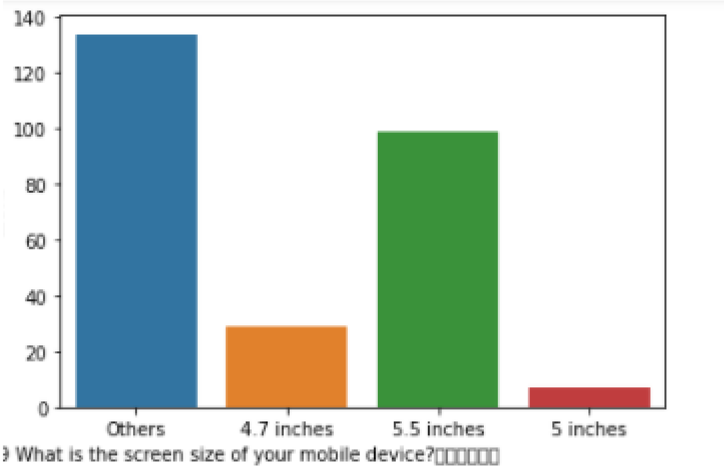


Observation – Using Mobile internet people shopped highest

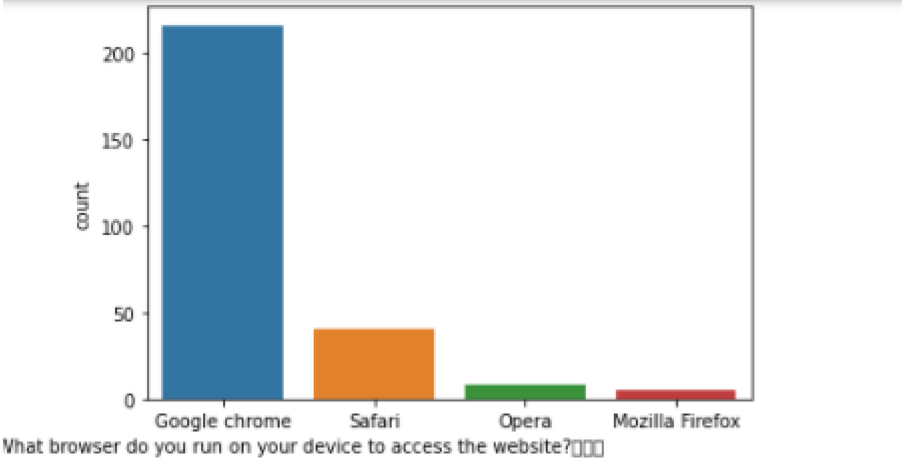


O

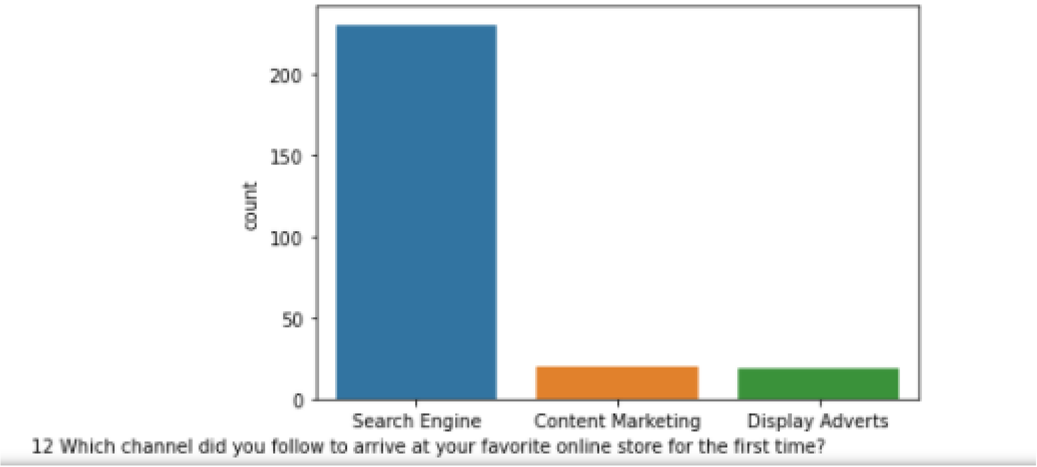
Observation – Using smartphone people bought more



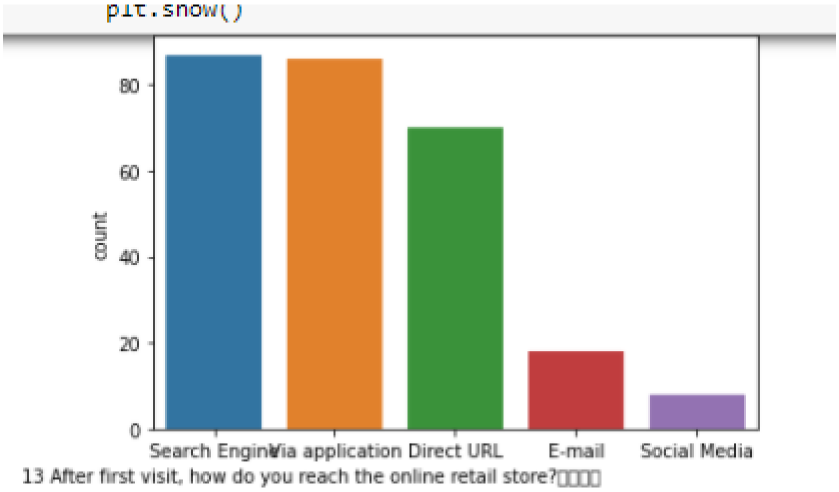
Observation—Others screen size of mobiles which screen measurement not mentioned or we did not get complete data due to some privacy reason contributing more for online shopping



Observation – People used google chrome more for online shopping

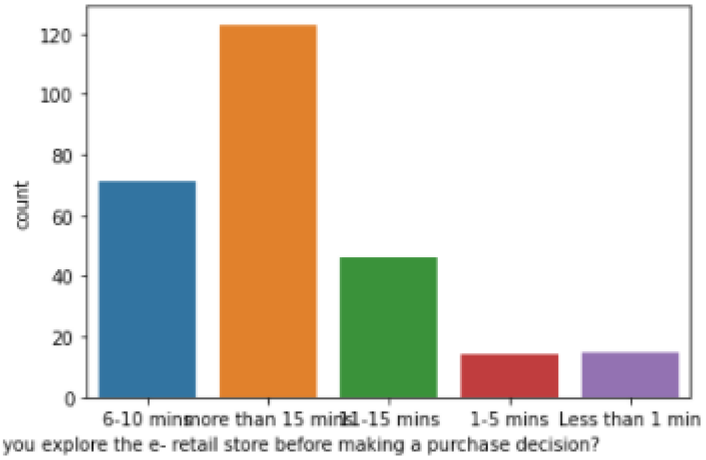


Observation—Search -engine is the favourite choice made by people for online shopping

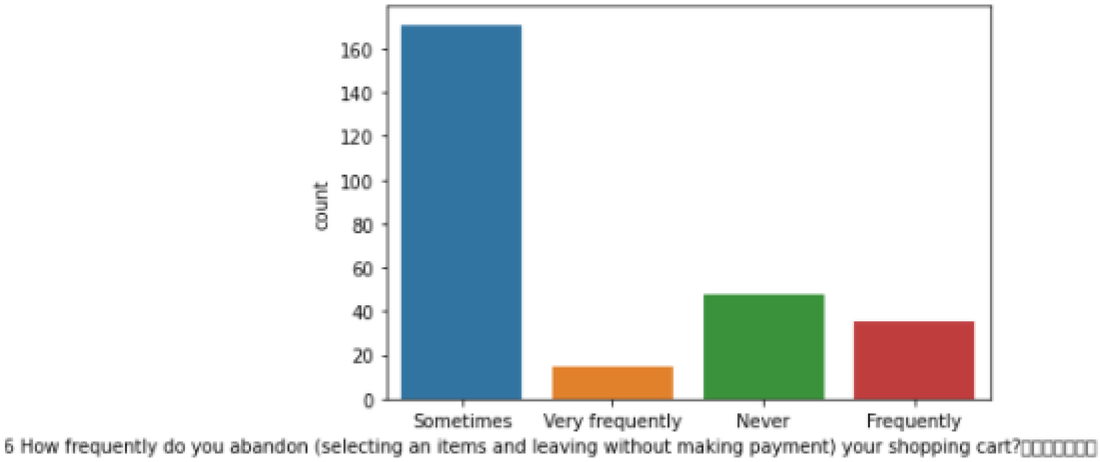


```
plt.show()
```

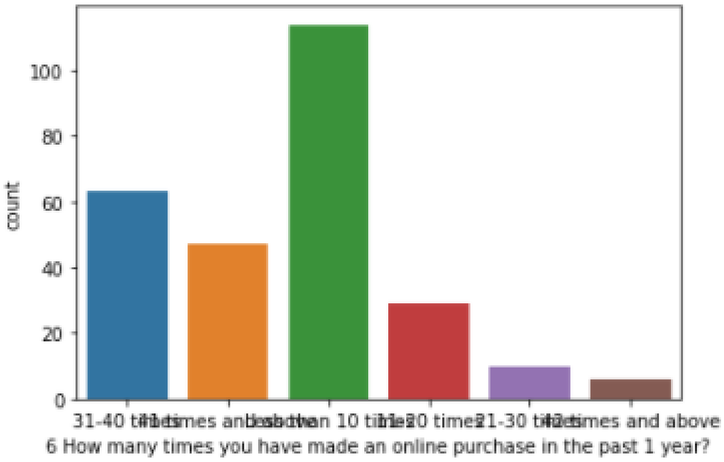
Observation-SO here , search engine , application providing highest equal value for reaching people to website after 1st visit



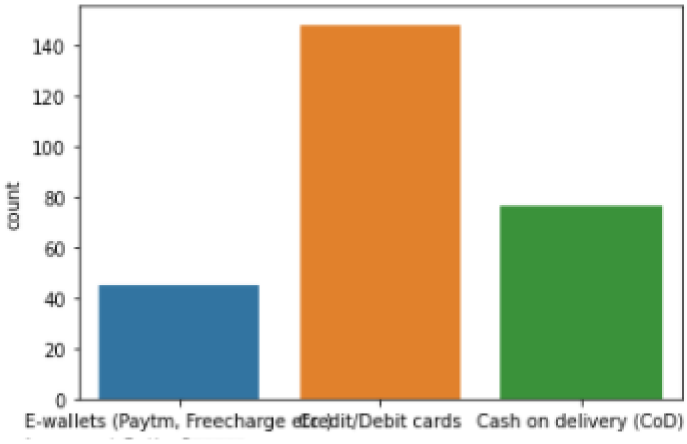
Observation—Maximum People spent more than 15 min before made final decision of shopping



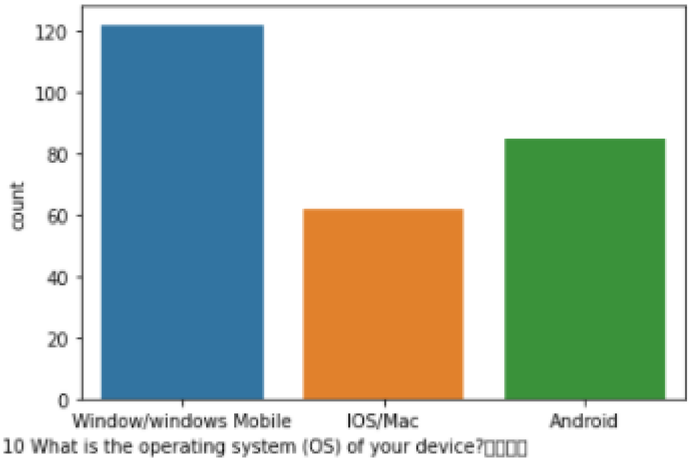
Observation- People rarely sometimes abandon shopping cart



Observation – so this is the target variable where Less than 10 times giving highest value

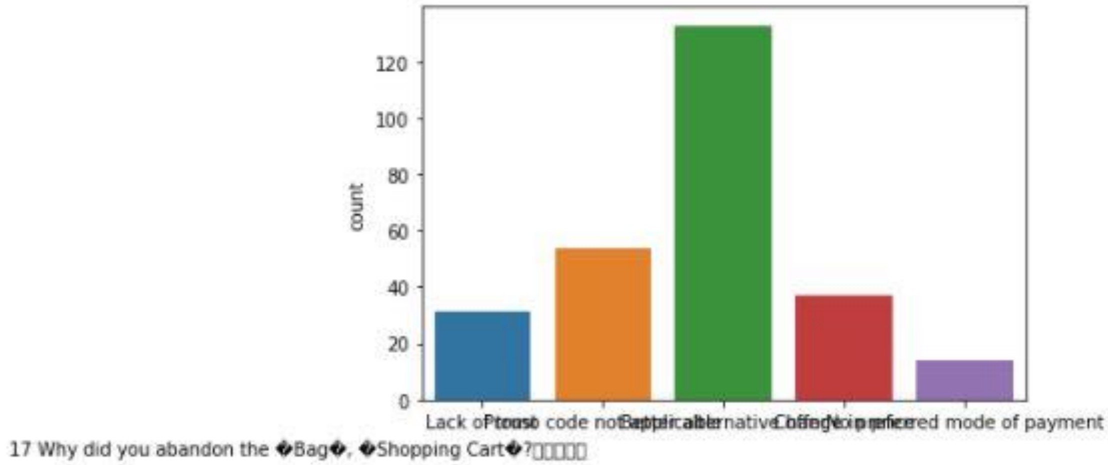


Observation—People prefer Debit/credit options more while doing purchase

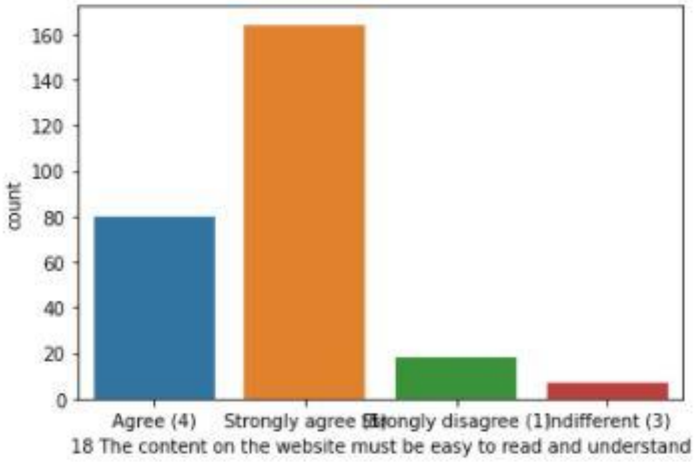


Observation-windows mobile people used more for shopping

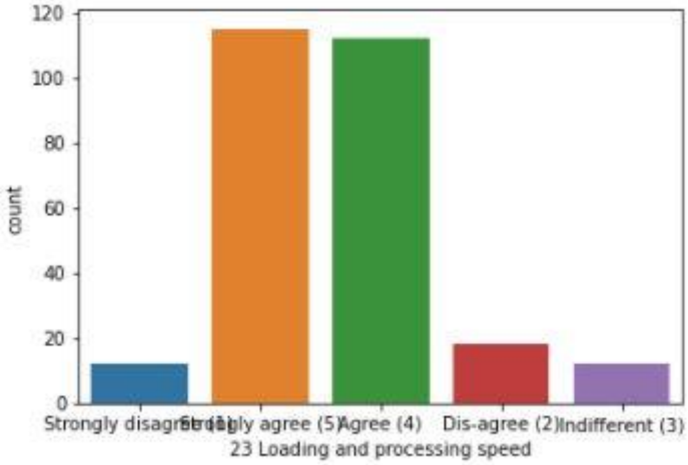
16 How frequently do you abandon (selecting an items and leaving without making payment) your shopping cart?□□□□□□□□



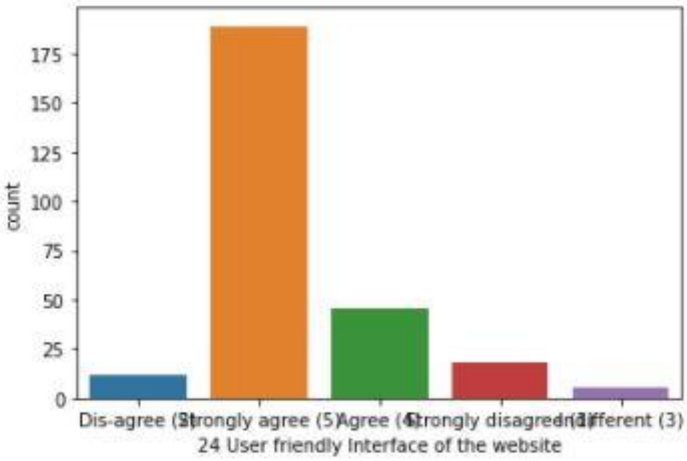
Observation- Reason of alternatives options people abandon shopping cart



Observation-the content available easy to understand is stated by most of the people we purchased

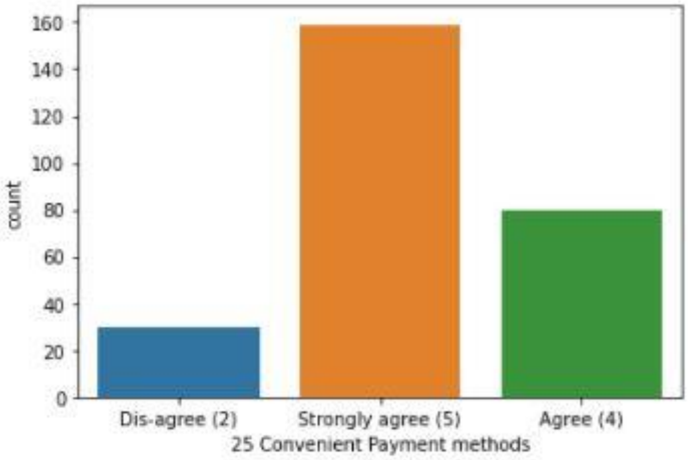


O→Highest vote for agree and strongly agree

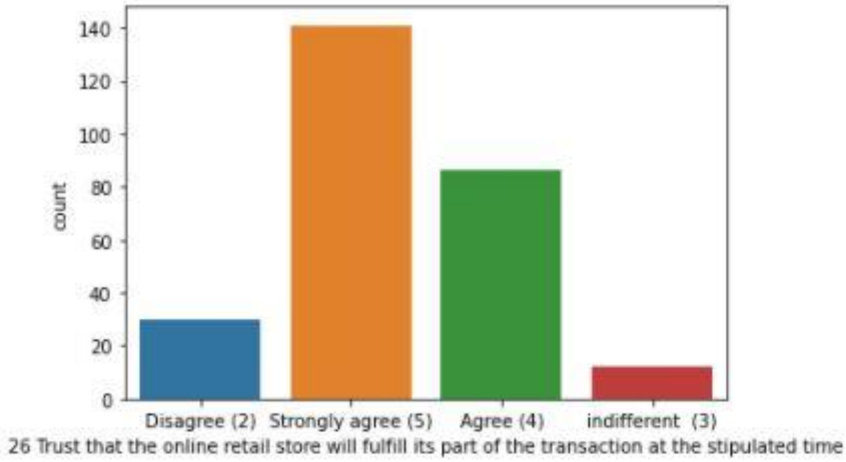


P

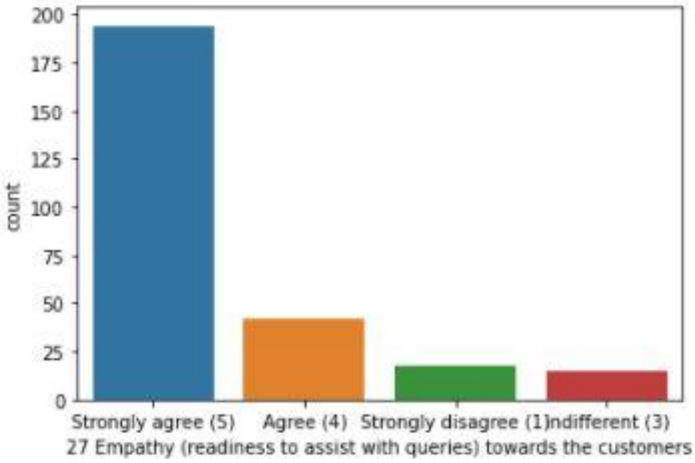
0->people did highest shopping from user friendly platform



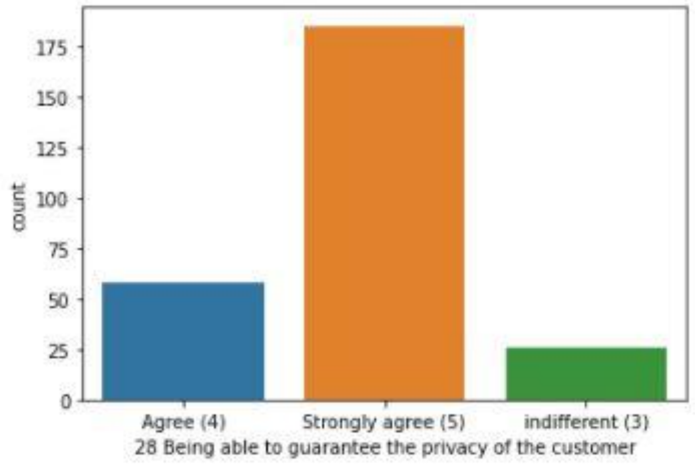
O->Convenient payment method is the first choice of customer



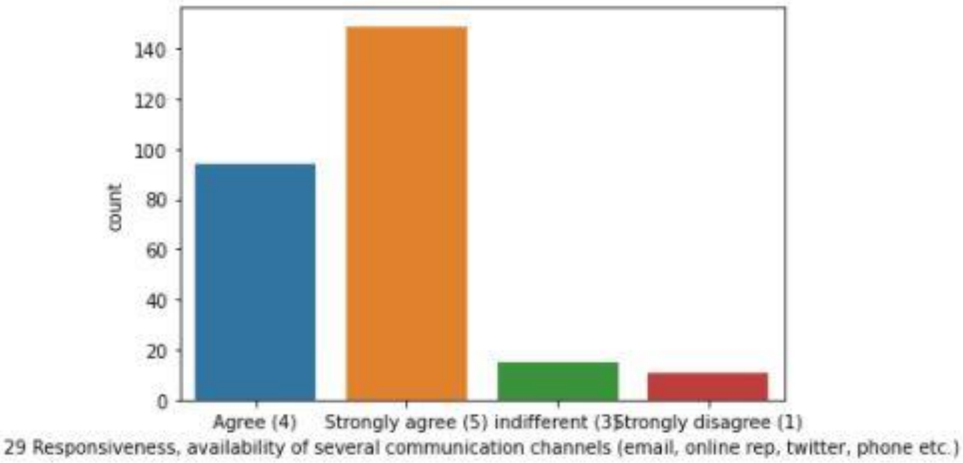
O->people choose trustworthy website for shopping



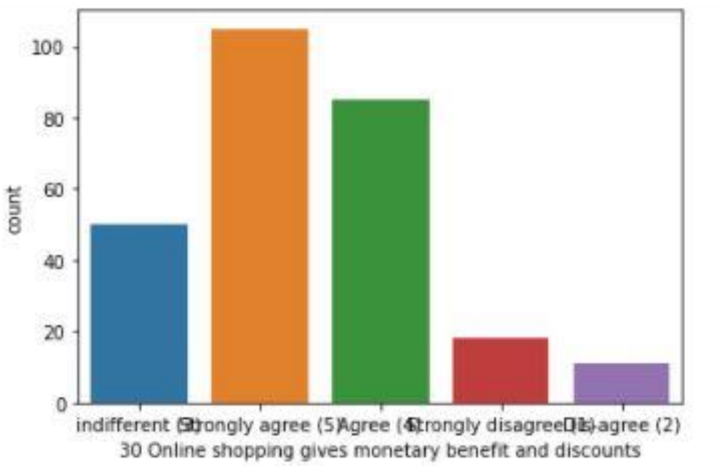
O→Here people strongly agree with this



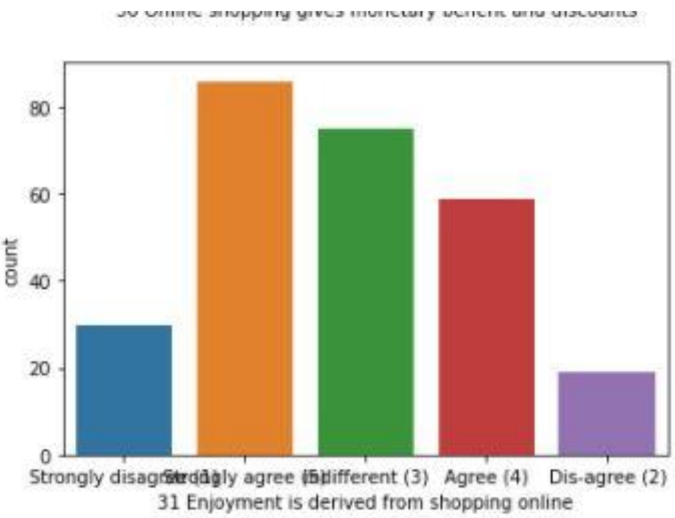
O→here people strongly agree with this



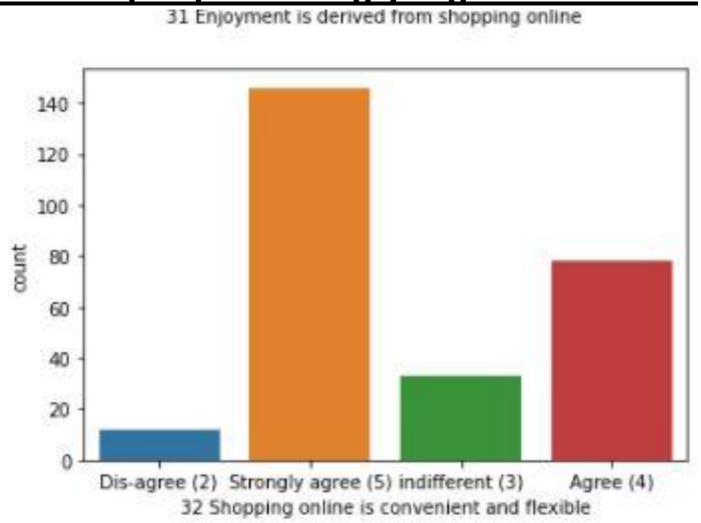
O->People strongly support this factor



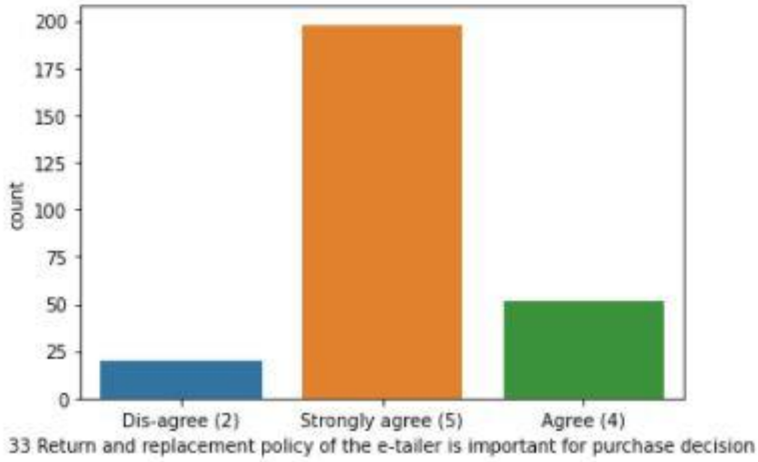
O→Here people strongly agree with this



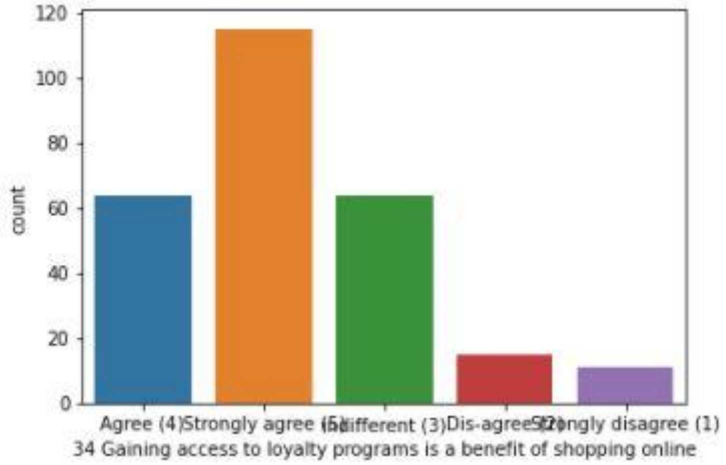
- O→Here people strongly agree with this



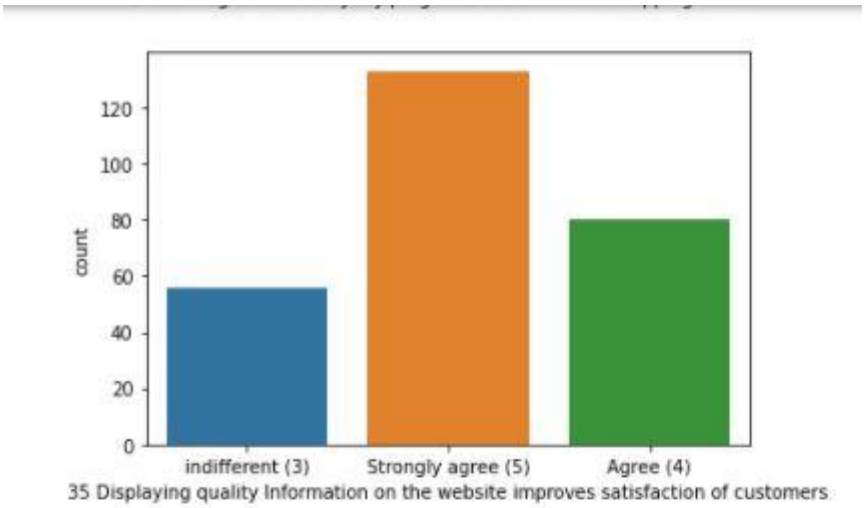
O→Here people strongly agree with this



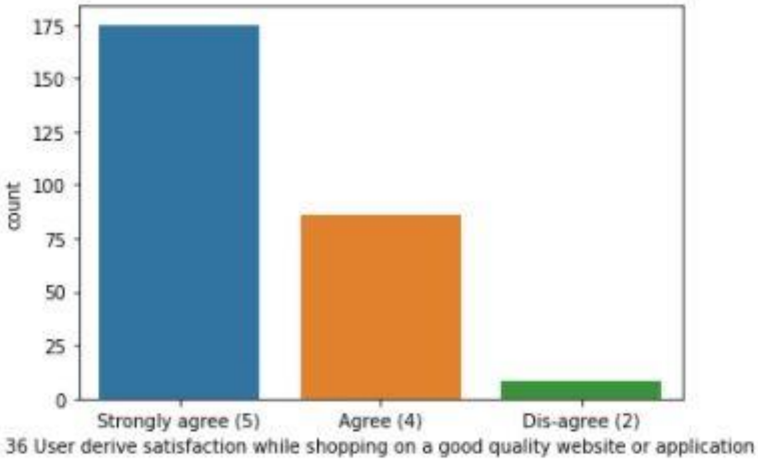
O→Here people strongly agree with this



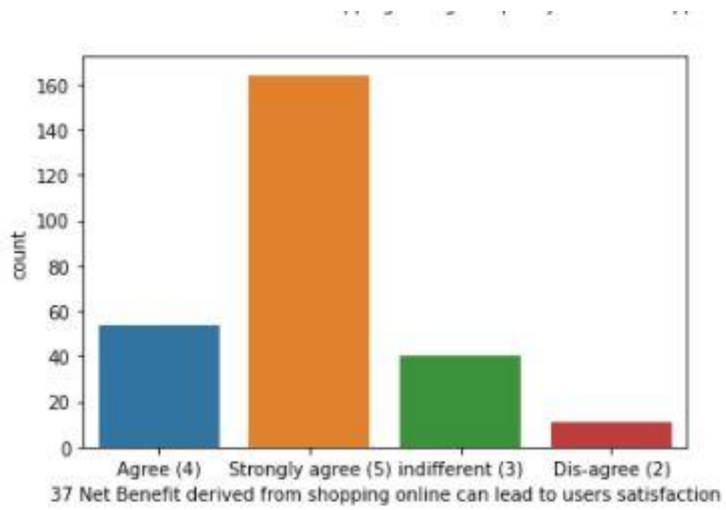
O→Here people strongly agree with this



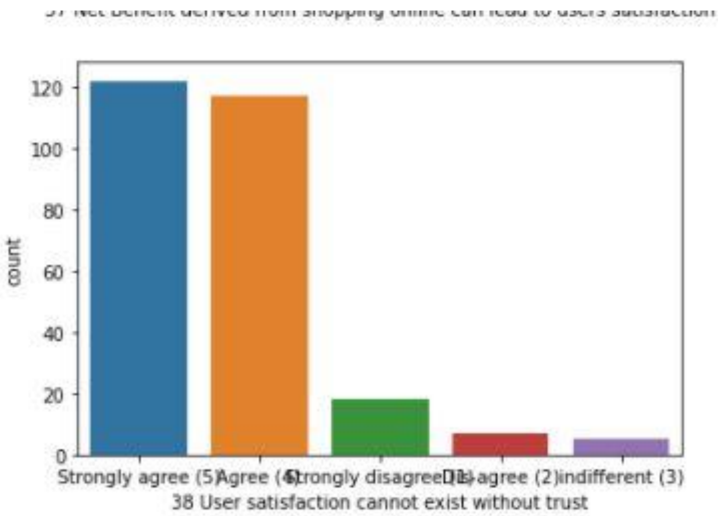
O→Here people strongly agree with this



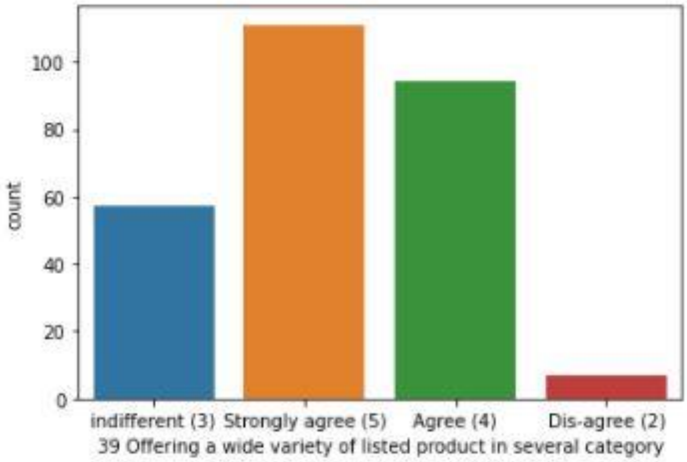
O→Here people strongly agree with this



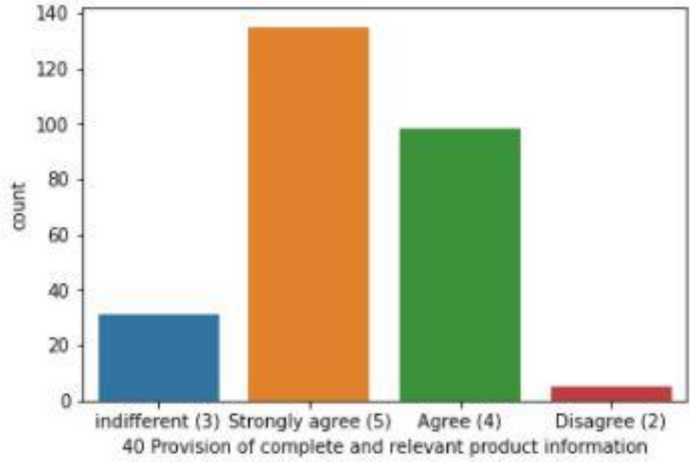
O→Here people strongly agree with this



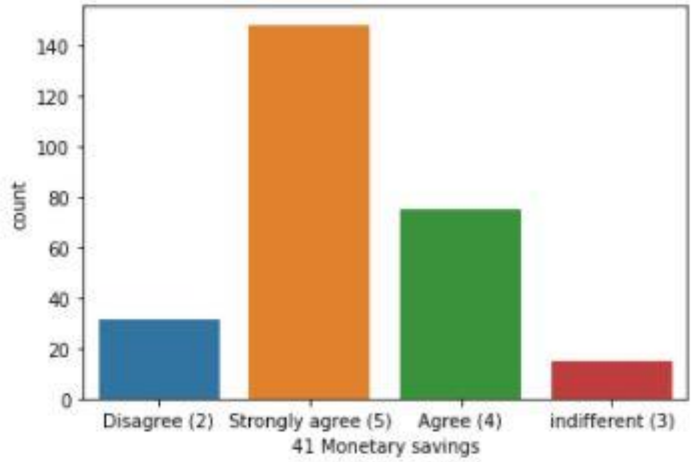
O→Here people strongly agree with this



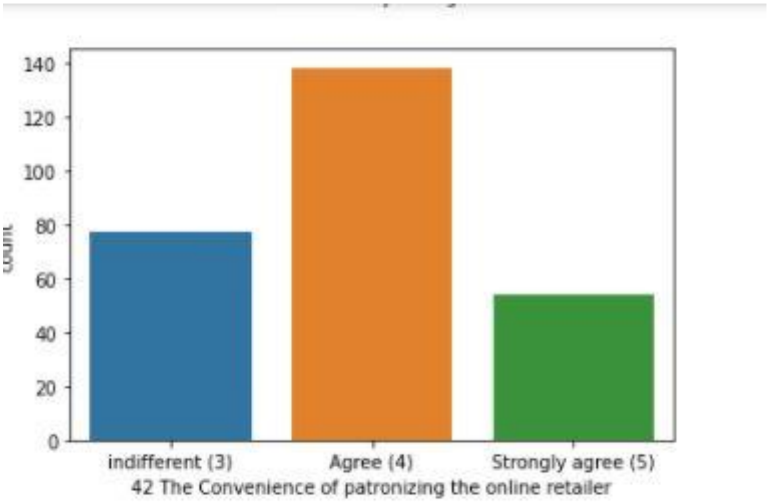
O→Here people strongly agree with this



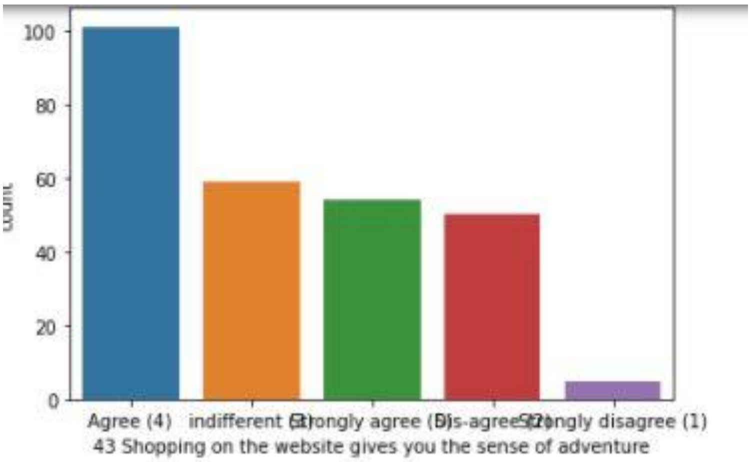
O→Here people strongly agree with this



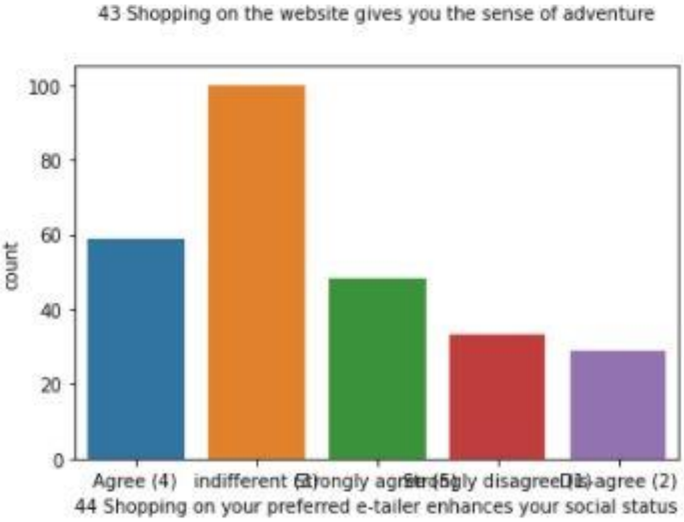
O→Here people strongly agree with this



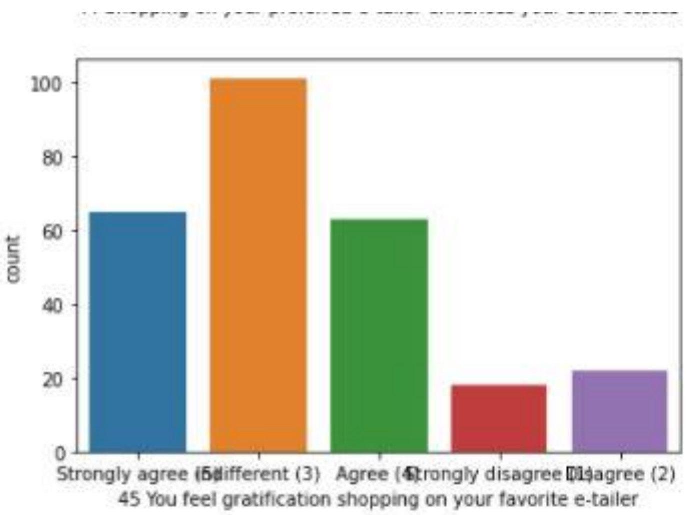
O→Here people strongly agree with this

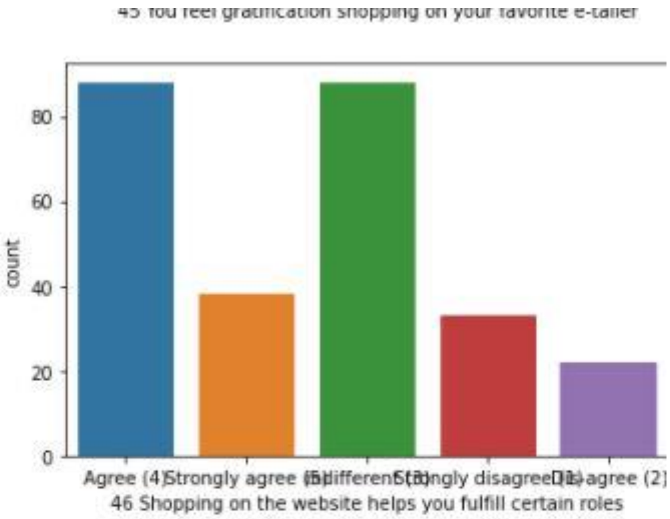


O→Here people strongly agree with this

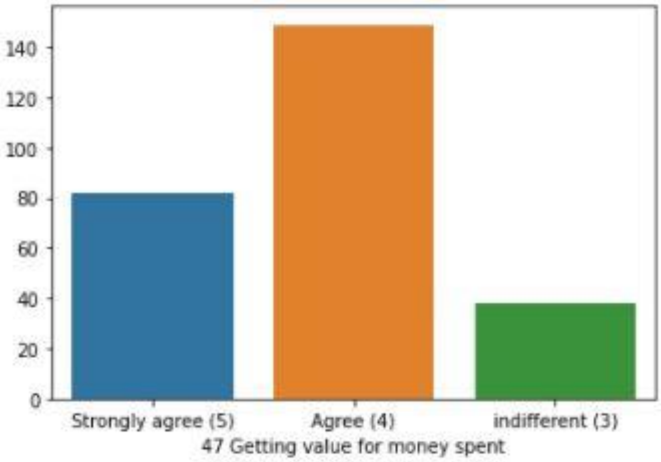


O→Here people strongly agree with this





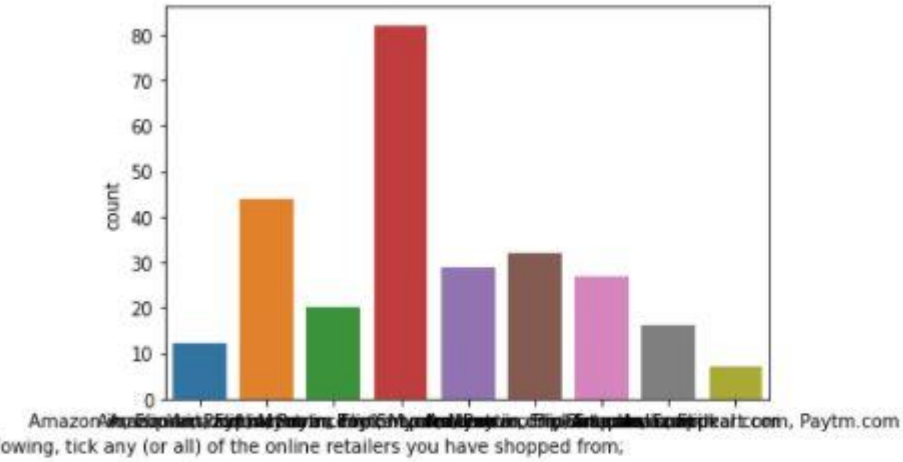
This column has class imbalance issue , frequency of data is not equally distributed

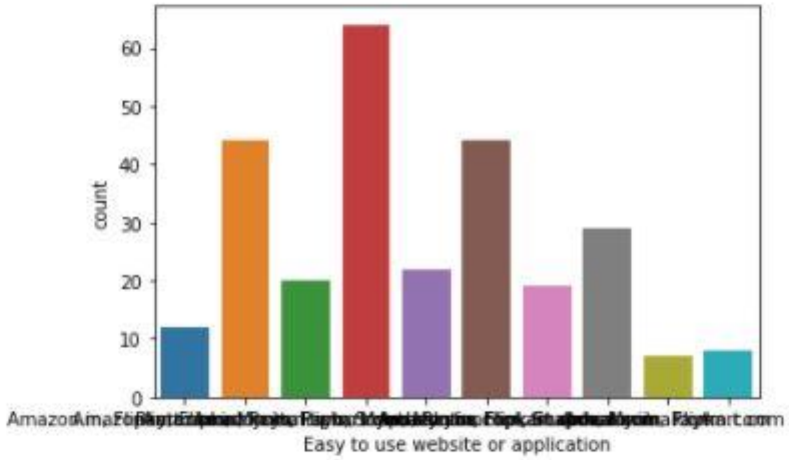


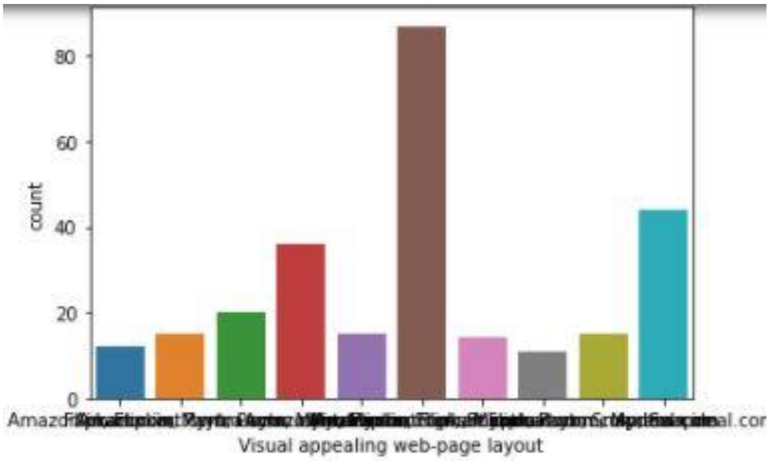
T

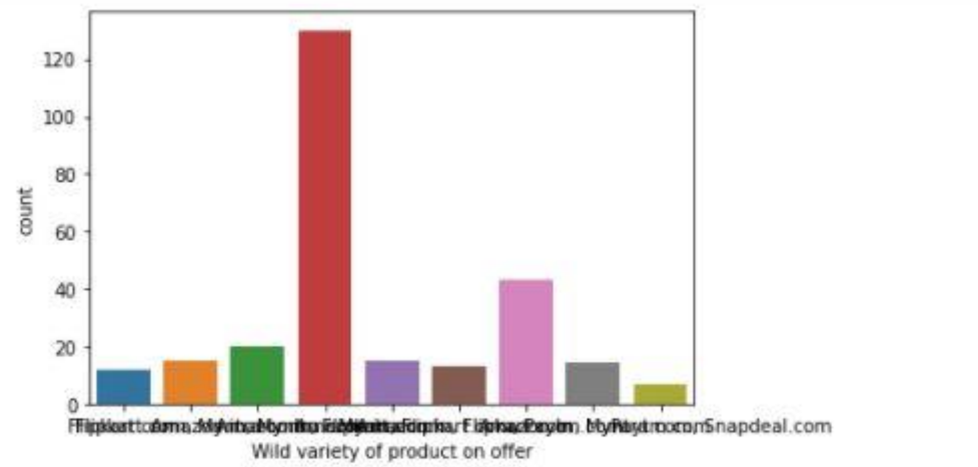
This column has class imbalance issue , frequency of data is not equally distributed

Getting value for money spent

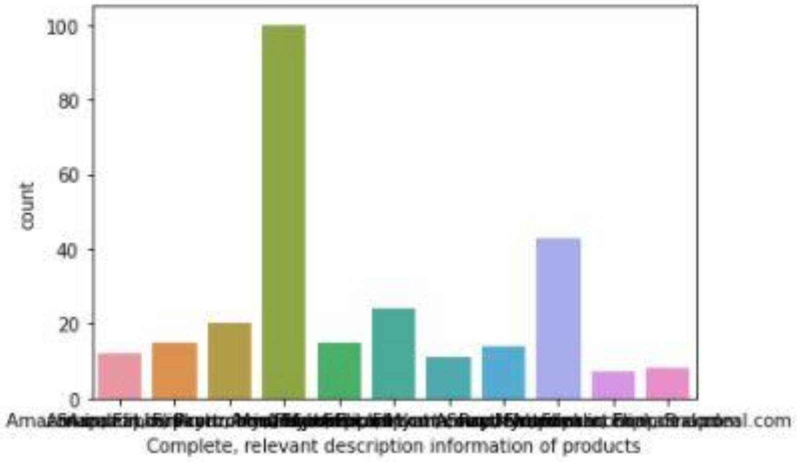




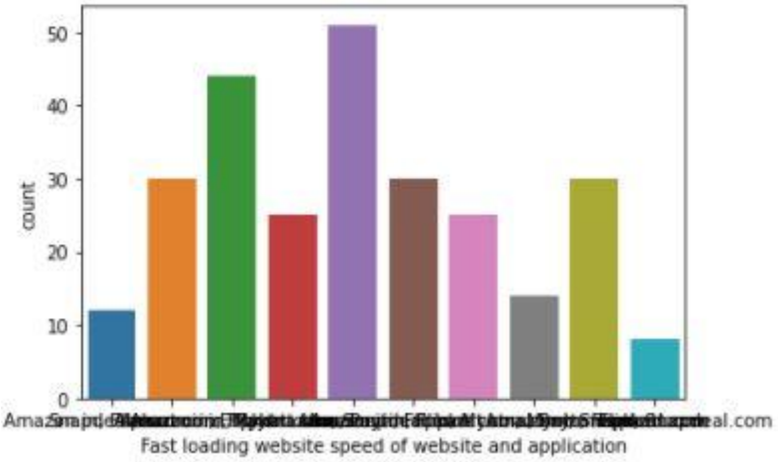


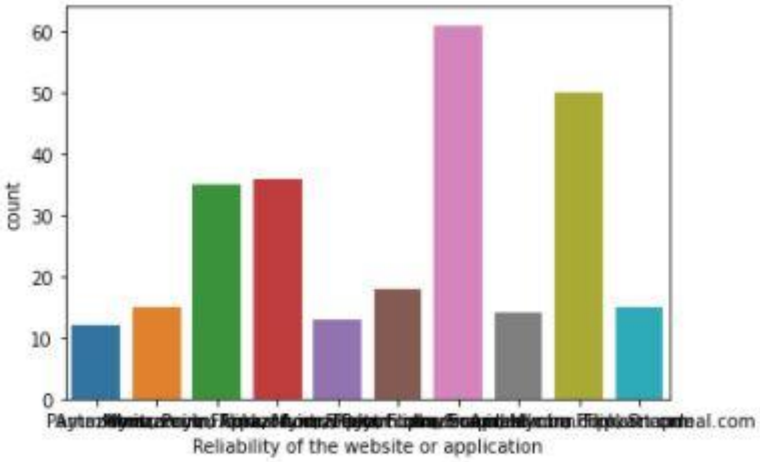


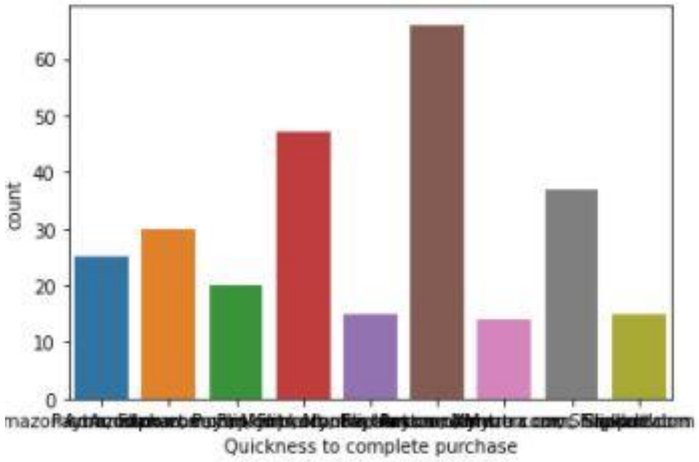
Wild variety of product on offer

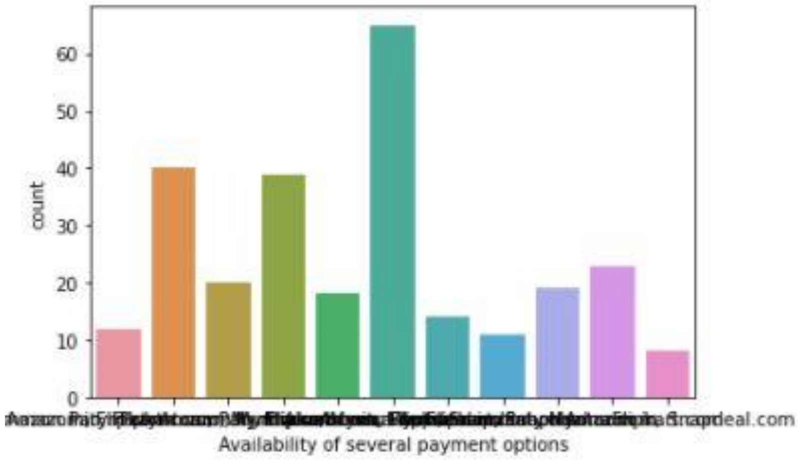


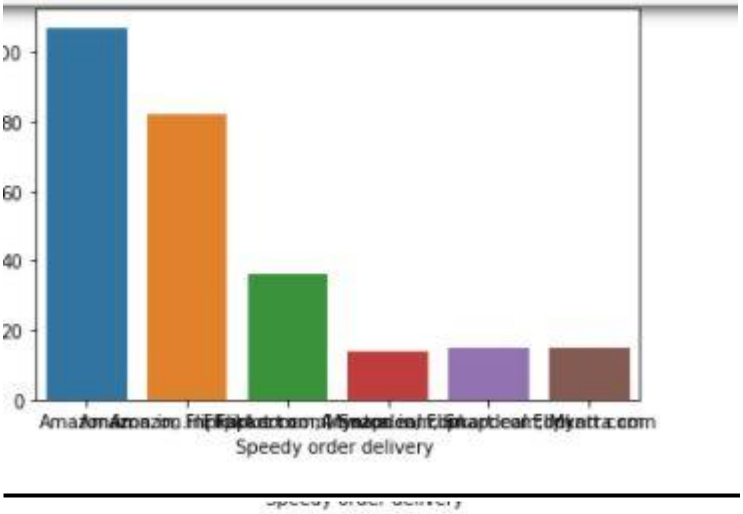
Complete, relevant description information of products

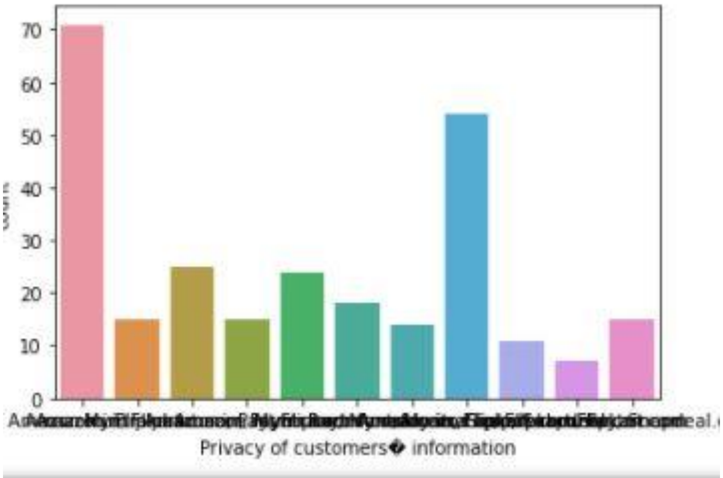


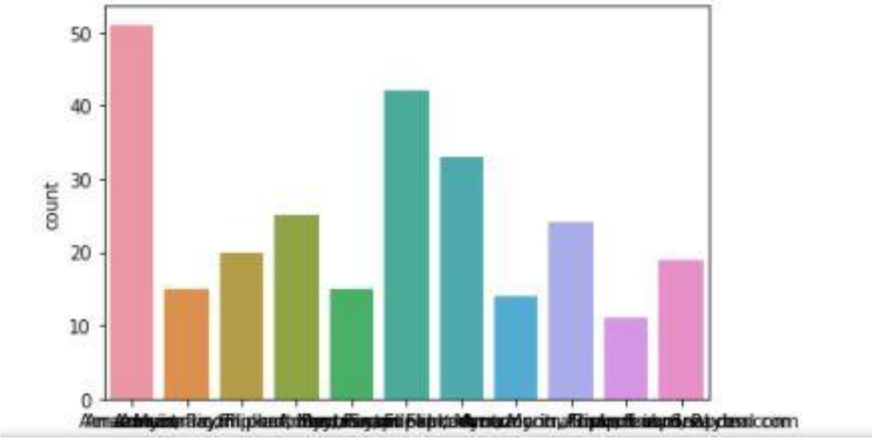

count
0
10
20
30
40
50
60
Reliability of the website or application

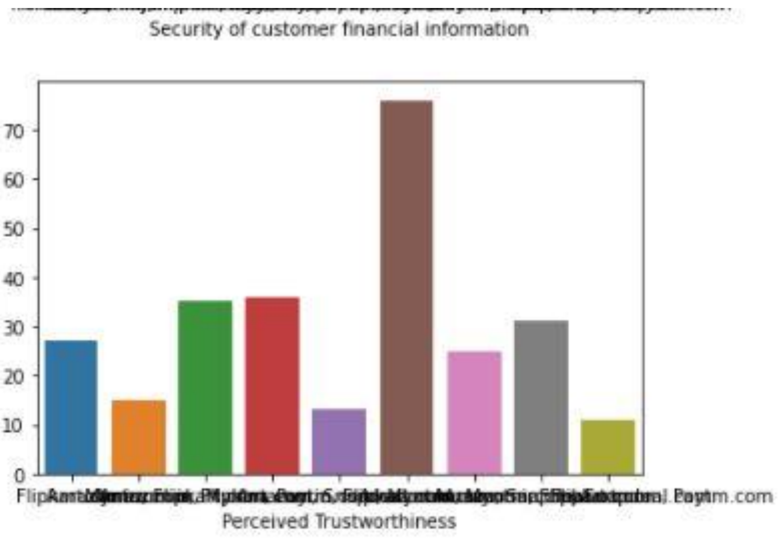

count
0
10
20
30
40
50
60
Quickness to complete purchase

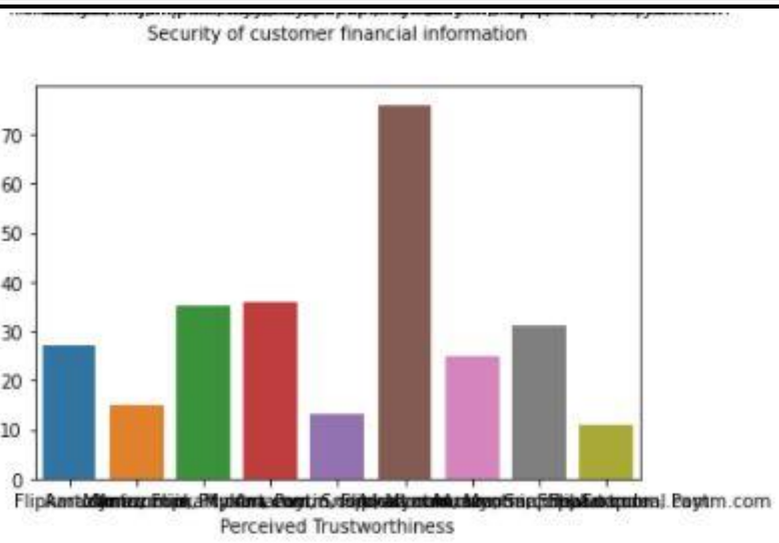

count
0
10
20
30
40
50
60
Availability of several payment options

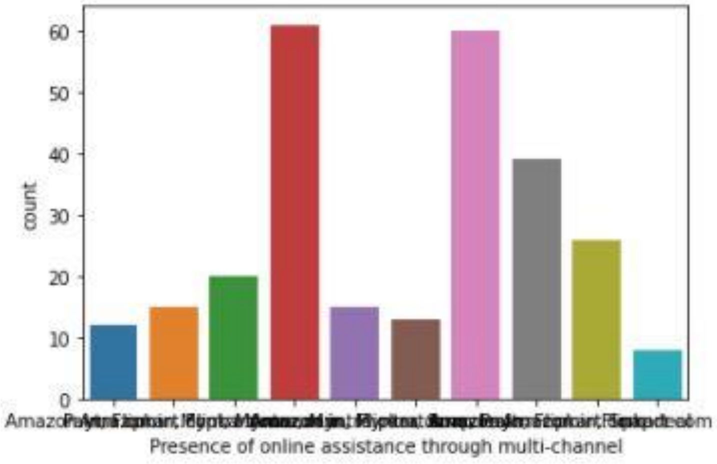

00
80
60
40
20
0
Speedy order delivery

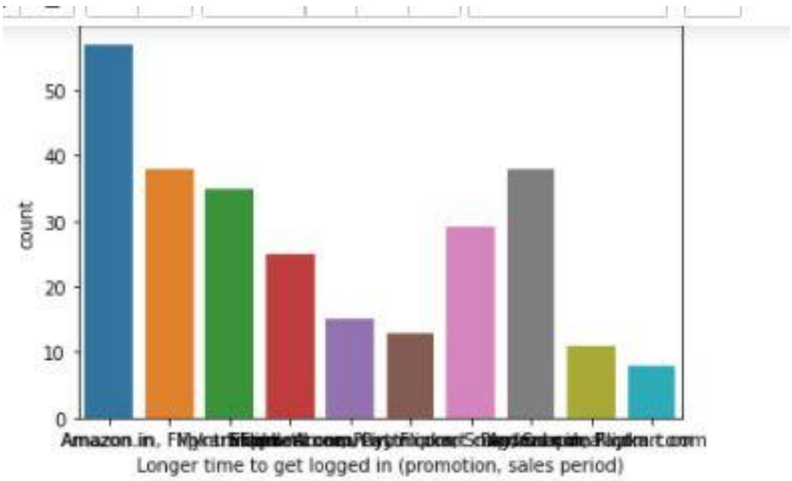

70
60
50
40
30
20
10
0
Privacy of customers� information

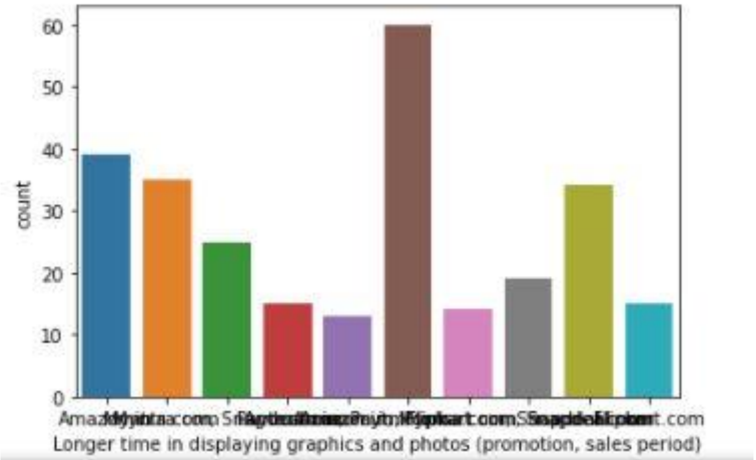

50
40
30
count
20
10
0

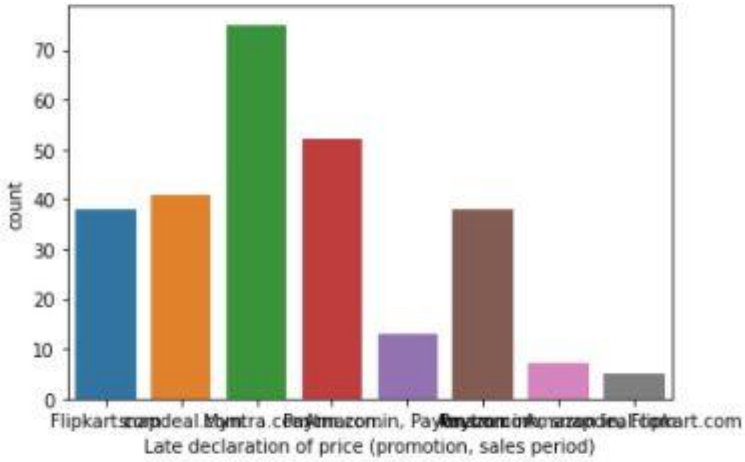

Security of customer financial information
count
0
10
20
30
40
50
60
70
Perceived Trustworthiness

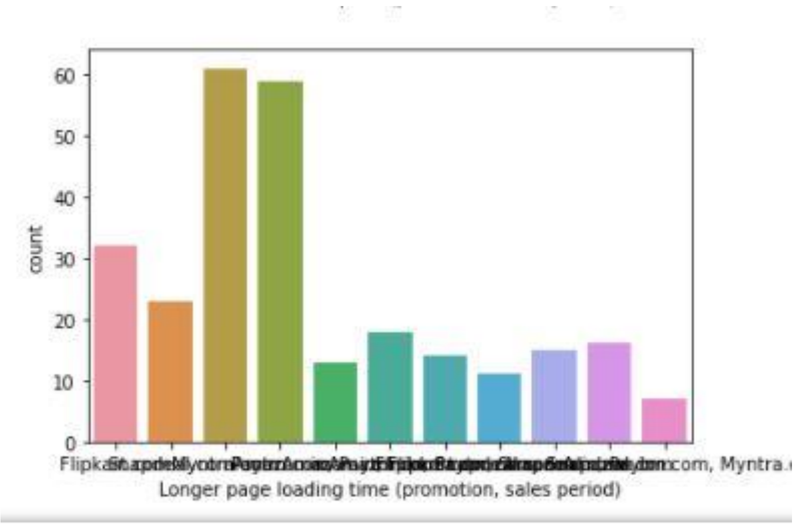

Security of customer financial information
count
0
10
20
30
40
50
60
70
Perceived Trustworthiness

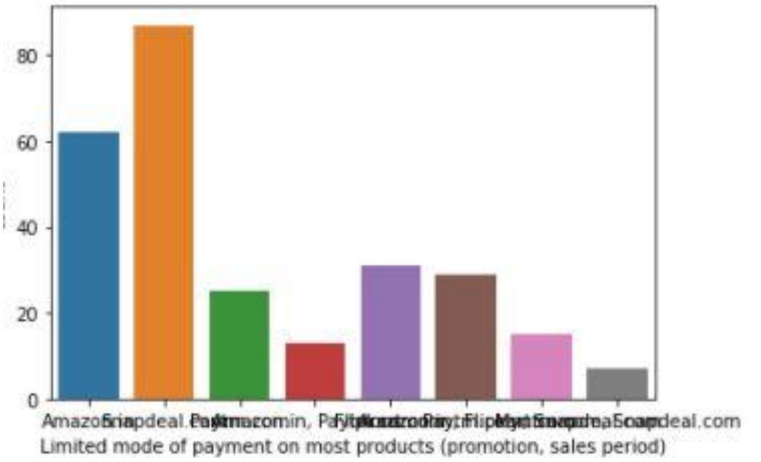

count
0
10
20
30
40
50
60
Presence of online assistance through multi-channel

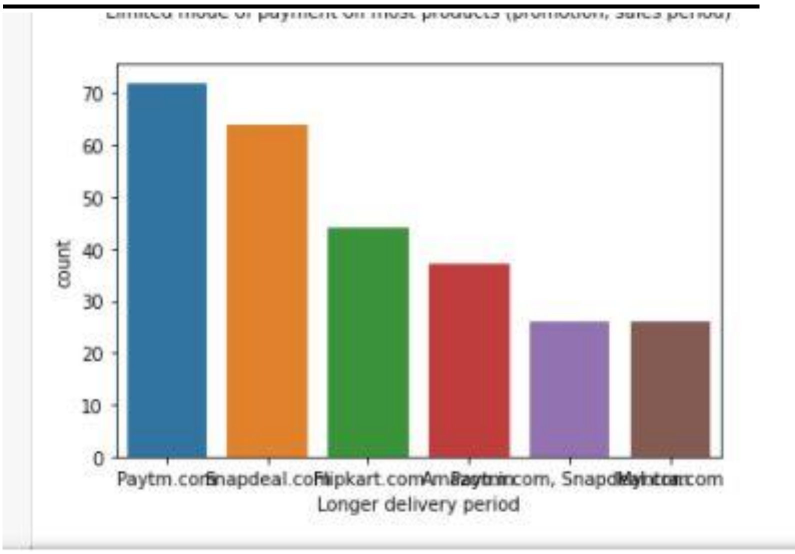

count
0
10
20
30
40
50
Amazon.in, Flipkart.com
Longer time to get logged in (promotion, sales period)

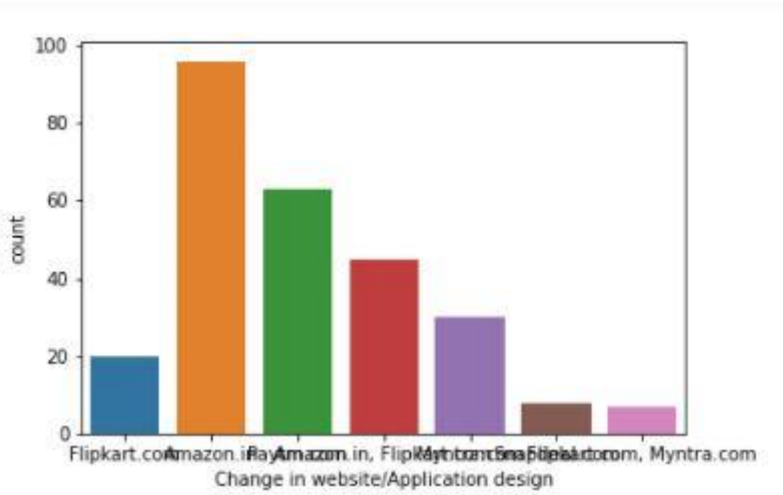

count
0
10
20
30
40
50
60
Flipkart.com
Longer time in displaying graphics and photos (promotion, sales period)

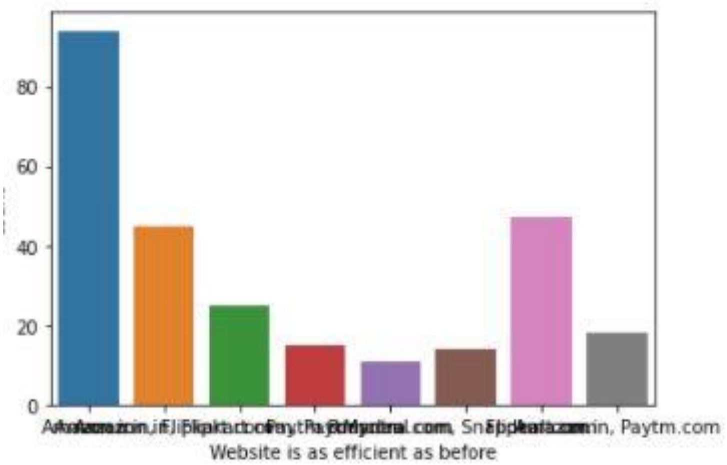

count
0
10
20
30
40
50
60
70
Flipkart.com
Late declaration of price (promotion, sales period)

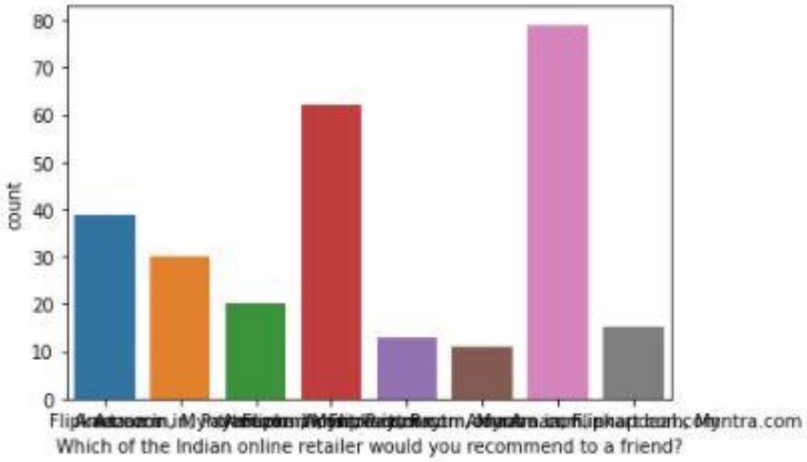

count
0
10
20
30
40
50
60
Longer page loading time (promotion, sales period)


0
20
40
60
80
Amazo
deal.com
Limited mode of payment on most products (promotion, sales period)


Limited mode of payment on most products (promotion, sales period)
count
0
10
20
30
40
50
60
70
Paytm.com
Snapdeal.com
Flipkart.com
Amazon.in
Paytm.com, Snapdeal.com
Myntra.com
Longer delivery period


100
80
60
40
20
0
count
Flipkart.com
Flipkart.com, Myntra.com
Change in website/Application design


80
60
40
20
0
Amazon.in, Paytm.com
Website is as efficient as before


80
70
60
50
40
30
20
10
0
count
Flipkart.com, Myntra.com
Which of the Indian online retailer would you recommend to a friend?

Detail’s observation of above plotted graph

1 Gender of respondent- Observation – Here we found it is object column and found that it is having class imbalance issue , so female category giving highest value for shopping

2 How old are you? - Observation – here 31-40 yrs category did highest shopping

3 Which city do you shop online from? - Observation – From Delhi people did maximum shopping and this column is also having class imbalance issue

4 What is the Pin Code of where you shop online from? Observation-- It is not an object type of column , it is numeric column , we used did plot and data is not normally distributed

5 Since How Long You are Shopping Online? - Observation—people who are shopping above 4 years they are having highest vote

6 How many times you have made an online purchase in the past 1 year? - Observation – so this is the target variable where Less than 10 times giving highest value

7 How do you access the internet while shopping on-line? - Observation – Using Mobile internet people shopped highest

8 Which devices do you use to access the online shopping? - Observation – Using smartphone people bought more

9 What is the screen size of your mobile device? Observation—Other’s screen size of mobiles which screen measurement not mentioned or we did not get complete data due to some privacy reason contributing more for online shopping

10 What is the operating system (OS) of your device? Observation- windows mobile people used more for shopping

11 What browsers do you run on your device to access the website? Observation – People used google chrome more for online shopping

12 Which channels did you follow to arrive at your favourite online store for the first time? Observation—Search -engine is the favourite choice made by people for online shopping

13 After first visit, how do you reach the online retail store? Observation-SO here, search engine, application providing highest equal value for reaching people to website after 1st visit

14 How much time do you explore the e- retail store before making a purchase decision? Observation—Maximum People spent more than 15 min before made final decision of shopping

15 What is your preferred payment Option? Observation—People prefer Debit/credit options more while doing purchase

16 How frequently do you abandon (selecting an item and leaving without making payment) your shopping cart? Observation- People rarely sometimes abandon shopping cart

17 Why did you abandon the “Bag”, “Shopping Cart”? Observation-Reason of alternatives options people abandon shopping cart

18 The content on the website must be easy to read and understand-Observation-the content available easy to understand is stated by most of the people we purchased

19 Information on similar product to the one highlighted is important for product comparison---
Observation—here all customers did strongly agree

20 Complete information on listed seller and product being offered is important for purchase decision
Observation—here all customers did strongly agree

21 All relevant information on listed products must be stated clearly-Observation--here maximum customers did agree

23 Loading and processing speed: Highest vote for agree and strongly agree
24 User friendly Interface of the website-0->people did highest shopping from user friendly platform

25 Convenient Payment methods- O->Convenient payment method is the first choice of customer

26 Trust that the online retail store will fulfil its part of the transaction at the stipulated time- O->people choose trustworthy website for shopping

27 Empathy (readiness to assist with queries) towards the customers-O→Here people strongly agree with this

28 Being able to guarantee the privacy of the customer O→Here people strongly agree with this

29 Responsiveness, availability of several communication channels (email, online rep, twitter, phone etc.) O→Here people strongly agree with this

30 Online shopping gives monetary benefit and discounts- O→Here people strongly agree with this

31 Enjoyment is derived from shopping online- O→Here people strongly agree with this

32 Shopping online is convenient and flexible- O→Here people strongly agree with this

33 Return and replacement policy of the e-tailer is important for purchase decision- O→Here people strongly agree with this

34 Gaining access to loyalty programs is a benefit of shopping online- O→Here people strongly agree with this

35 Displaying quality Information on the website improves satisfaction of customers O→Here people strongly agree with this

36 User derive satisfaction while shopping on a good quality website or application O→Here people strongly agree with this

37 Net Benefit derived from shopping online can lead to users satisfaction O→Here people strongly agree with this

38 User satisfaction cannot exist without trust O→Here people strongly agree with this

39 Offering a wide variety of listed product in several category- O→Here people strongly agree with this

40 Provision of complete and relevant product information- O→Here people strongly agree with this

41 Monetary savings- O→Here people strongly agree with this

42 The Convenience of patronizing the online retailer- O→Here people strongly agree with this

43 Shopping on the website gives you the sense of adventure- O→Here people strongly agree with this

44 Shopping on your preferred e-tailer enhances your social status- O→Here people strongly agree with this

45 You feel gratification shopping on your favourite e-tailer- O→Here people strongly agree with this

46 Shopping on the website helps you fulfil certain roles- This column has class imbalance issue, frequency of data is not equally distributed

47 Getting value for money spent- This column has class imbalance issue, frequency of data is not equally distributed

From the following, tick any (or all) of the online retailers you have shopped from; - **This column has class imbalance issue , frequency of data is not equally distributed**

-

Easy to use website or application- **This column has class imbalance issue, frequency of data is not equally distributed**

-

Visual appealing web-page layout- This column has class imbalance issue, frequency of data is not equally distributed

-

Wild variety of product on offer- **This column has class imbalance issue , frequency of data is not equally distributed**

-

Complete, relevant description information of products- **This column has class imbalance issue, frequency of data is not equally distributed**

-

Fast loading website speed of website and application- **This column has class imbalance issue, frequency of data is not equally distributed**

-

Reliability of the website or application- **This column has class imbalance issue, frequency of data is not equally distributed**

-

Quickness to complete purchase- **This column has class imbalance issue, frequency of data is not equally distributed**

-

Availability of several payment options- **This column has class imbalance issue, frequency of data is not equally distributed**

-

Speedy order delivery- **This column has class imbalance issue, frequency of data is not equally distributed**

-

Privacy of customers’ information- **This column has class imbalance issue, frequency of data is not equally distributed**

- Security of customer financial information- **This column has class imbalance issue, frequency of data is not equally distributed**
- Perceived Trustworthiness- **This column has class imbalance issue, frequency of data is not equally distributed**
- Presence of online assistance through multi-channel **This column has class imbalance issue, frequency of data is not equally distributed**
- Longer time to get logged in (promotion, sales period) **This column has class imbalance issue, frequency of data is not equally distributed**
- Longer time in displaying graphics and photos (promotion, sales period) **This column has class imbalance issue, frequency of data is not equally distributed**
- Late declaration of price (promotion, sales period) **This column has class imbalance issue, frequency of data is not equally distributed**
- Longer page loading time (promotion, sales period) **This column has class imbalance issue, frequency of data is not equally distributed**
- Limited mode of payment on most products (promotion, sales period) **This column has class imbalance issue, frequency of data is not equally distributed**
- Longer delivery period **This column has class imbalance issue, frequency of data is not equally distributed**

Change in website/Application design **This column has class imbalance issue, frequency of data is not equally distributed**

-

Frequent disruption when moving from one page to another **This column has class imbalance issue, frequency of data is not equally distributed**

-

Website is as efficient as before **This column has class imbalance issue, frequency of data is not equally distributed**

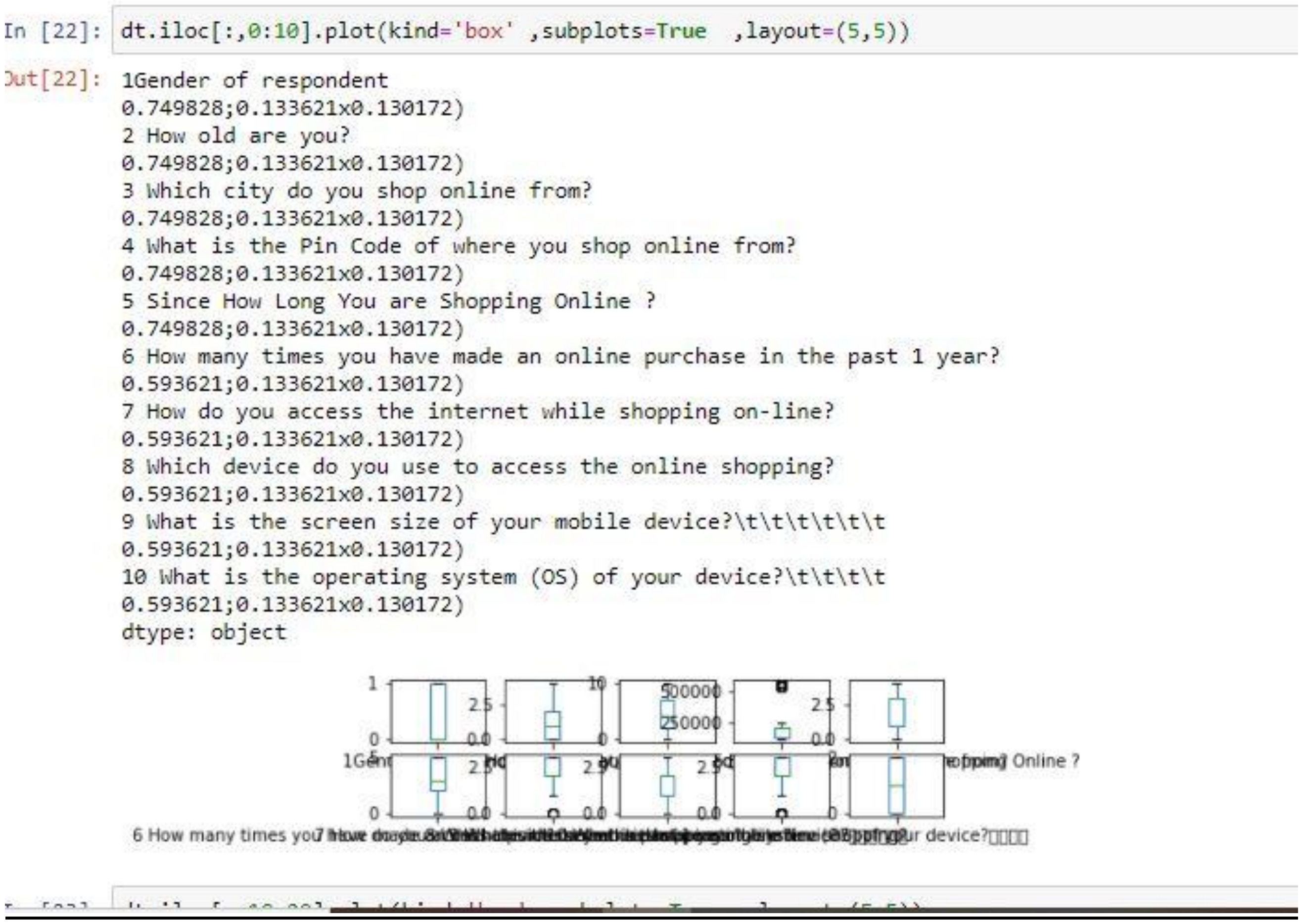
-

Which of the Indian online retailer would you recommend to a friend? **This column has class imbalance issue , frequency of data is not equally distributed**

-

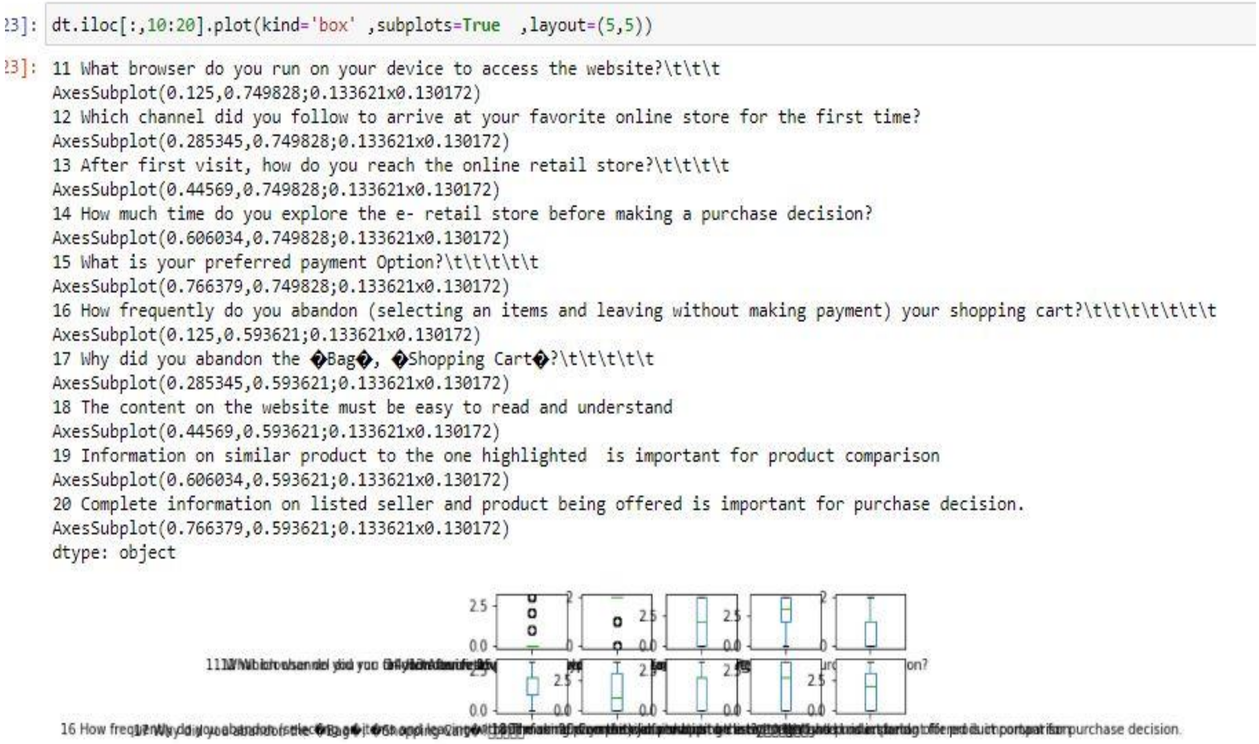
BOXPLOT: -

In statistics, an outlier is a data point that differs significantly from other observations. An outlier may be due to variability in the measurement or it may indicate experimental error; the latter are sometimes excluded from the data set. An outlier can cause serious problems in statistical
But from string data we can not remove outlier or else we will be losing information for getting graphical view we have plotted boxplot below



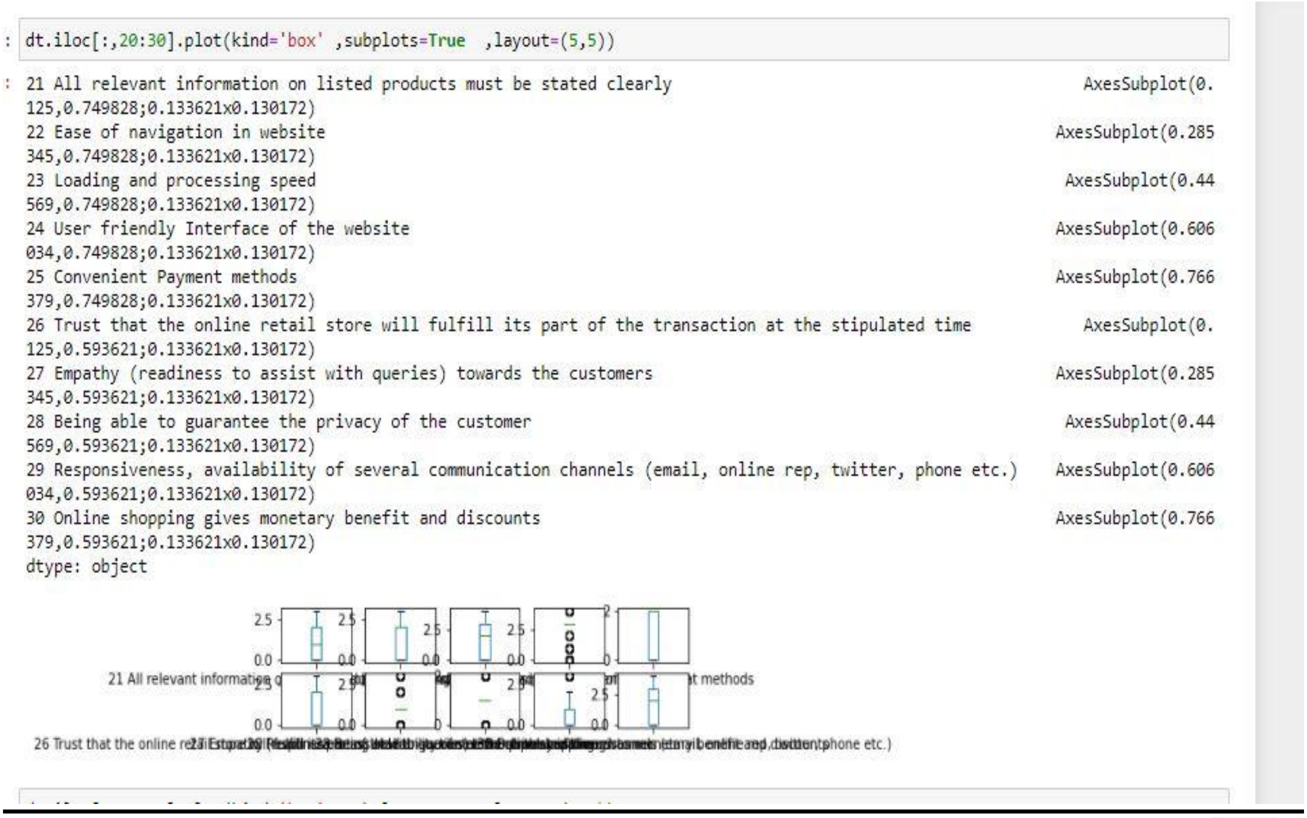
```
In [22]: dt.iloc[:,0:10].plot(kind='box' ,subplots=True  ,layout=(5,5))
```

```
Out[22]: 1Gender of respondent
         0.749828;0.133621x0.130172)
         2 How old are you?
         0.749828;0.133621x0.130172)
         3 Which city do you shop online from?
         0.749828;0.133621x0.130172)
         4 What is the Pin Code of where you shop online from?
         0.749828;0.133621x0.130172)
         5 Since How Long You are Shopping Online ?
         0.749828;0.133621x0.130172)
         6 How many times you have made an online purchase in the past 1 year?
         0.593621;0.133621x0.130172)
         7 How do you access the internet while shopping on-line?
         0.593621;0.133621x0.130172)
         8 Which device do you use to access the online shopping?
         0.593621;0.133621x0.130172)
         9 What is the screen size of your mobile device?\t\t\t\t\t\t\t
         0.593621;0.133621x0.130172)
         10 What is the operating system (OS) of your device?\t\t\t\t
         0.593621;0.133621x0.130172)
         dtype: object
```



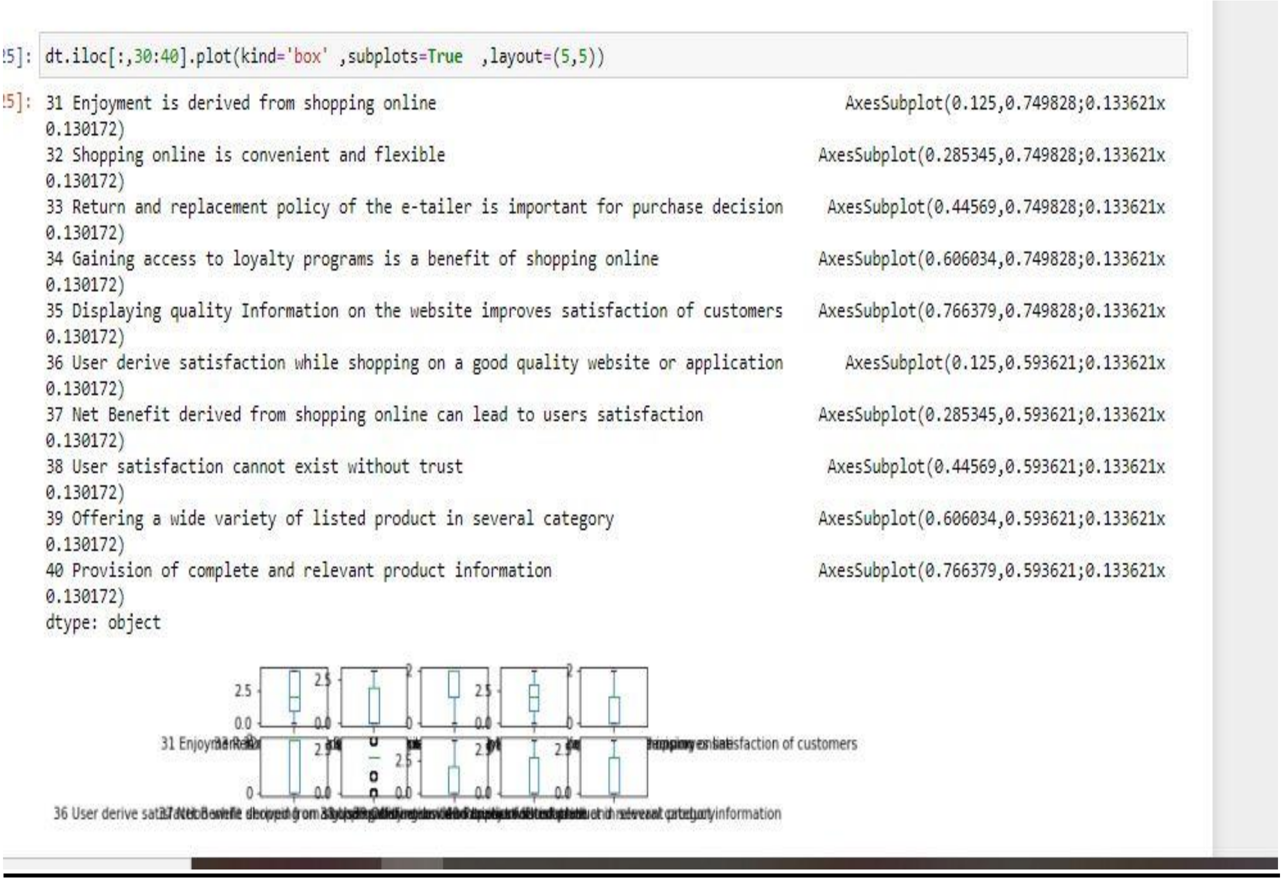
```
In [23]: dt.iloc[:,10:20].plot(kind='box' ,subplots=True  ,layout=(5,5))
```

```
Out[23]: 11 What browser do you run on your device to access the website?\t\t\t
      AxesSubplot(0.125,0.749828;0.133621x0.130172)
      12 Which channel did you follow to arrive at your favorite online store for the first time?
      AxesSubplot(0.285345,0.749828;0.133621x0.130172)
      13 After first visit, how do you reach the online retail store?\t\t\t\t
      AxesSubplot(0.44569,0.749828;0.133621x0.130172)
      14 How much time do you explore the e- retail store before making a purchase decision?
      AxesSubplot(0.606034,0.749828;0.133621x0.130172)
      15 What is your preferred payment Option?\t\t\t\t\t
      AxesSubplot(0.766379,0.749828;0.133621x0.130172)
      16 How frequently do you abandon (selecting an items and leaving without making payment) your shopping cart?\t\t\t\t\t\t\t\t
      AxesSubplot(0.125,0.593621;0.133621x0.130172)
      17 Why did you abandon the �Bag�, �Shopping Cart�?\t\t\t\t\t
      AxesSubplot(0.285345,0.593621;0.133621x0.130172)
      18 The content on the website must be easy to read and understand
      AxesSubplot(0.44569,0.593621;0.133621x0.130172)
      19 Information on similar product to the one highlighted  is important for product comparison
      AxesSubplot(0.606034,0.593621;0.133621x0.130172)
      20 Complete information on listed seller and product being offered is important for purchase decision.
      AxesSubplot(0.766379,0.593621;0.133621x0.130172)
      dtype: object
```



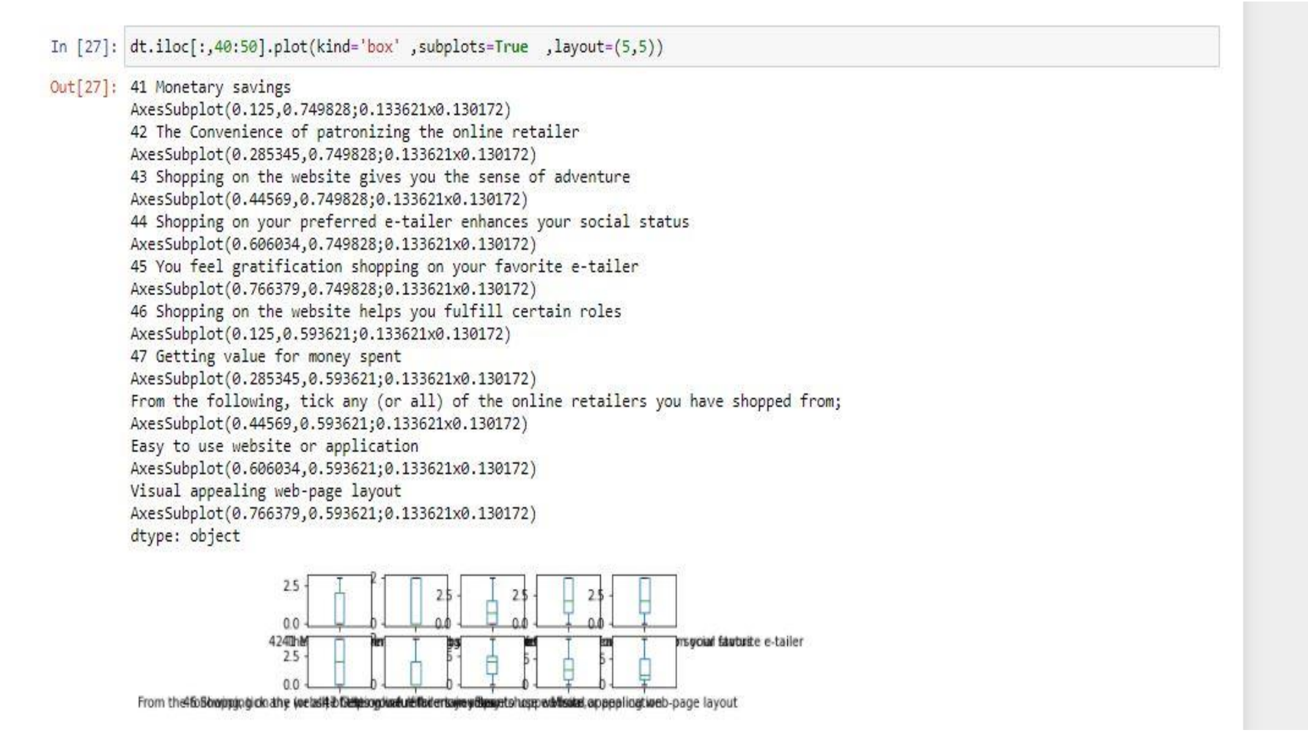
```
: dt.iloc[:,20:30].plot(kind='box' ,subplots=True  ,layout=(5,5))
```

```
: 21 All relevant information on listed products must be stated clearly                                AxesSubplot(0.
  125,0.749828;0.133621x0.130172)
  22 Ease of navigation in website                                                                   AxesSubplot(0.285
  345,0.749828;0.133621x0.130172)
  23 Loading and processing speed                                                                     AxesSubplot(0.44
  569,0.749828;0.133621x0.130172)
  24 User friendly Interface of the website                                                          AxesSubplot(0.606
  034,0.749828;0.133621x0.130172)
  25 Convenient Payment methods                                                                      AxesSubplot(0.766
  379,0.749828;0.133621x0.130172)
  26 Trust that the online retail store will fulfill its part of the transaction at the stipulated time    AxesSubplot(0.
  125,0.593621;0.133621x0.130172)
  27 Empathy (readiness to assist with queries) towards the customers                                AxesSubplot(0.285
  345,0.593621;0.133621x0.130172)
  28 Being able to guarantee the privacy of the customer                                              AxesSubplot(0.44
  569,0.593621;0.133621x0.130172)
  29 Responsiveness, availability of several communication channels (email, online rep, twitter, phone etc.)   AxesSubplot(0.606
  034,0.593621;0.133621x0.130172)
  30 Online shopping gives monetary benefit and discounts                                            AxesSubplot(0.766
  379,0.593621;0.133621x0.130172)
  dtype: object
```



```
[25]: dt.iloc[:,30:40].plot(kind='box' ,subplots=True  ,layout=(5,5))
```

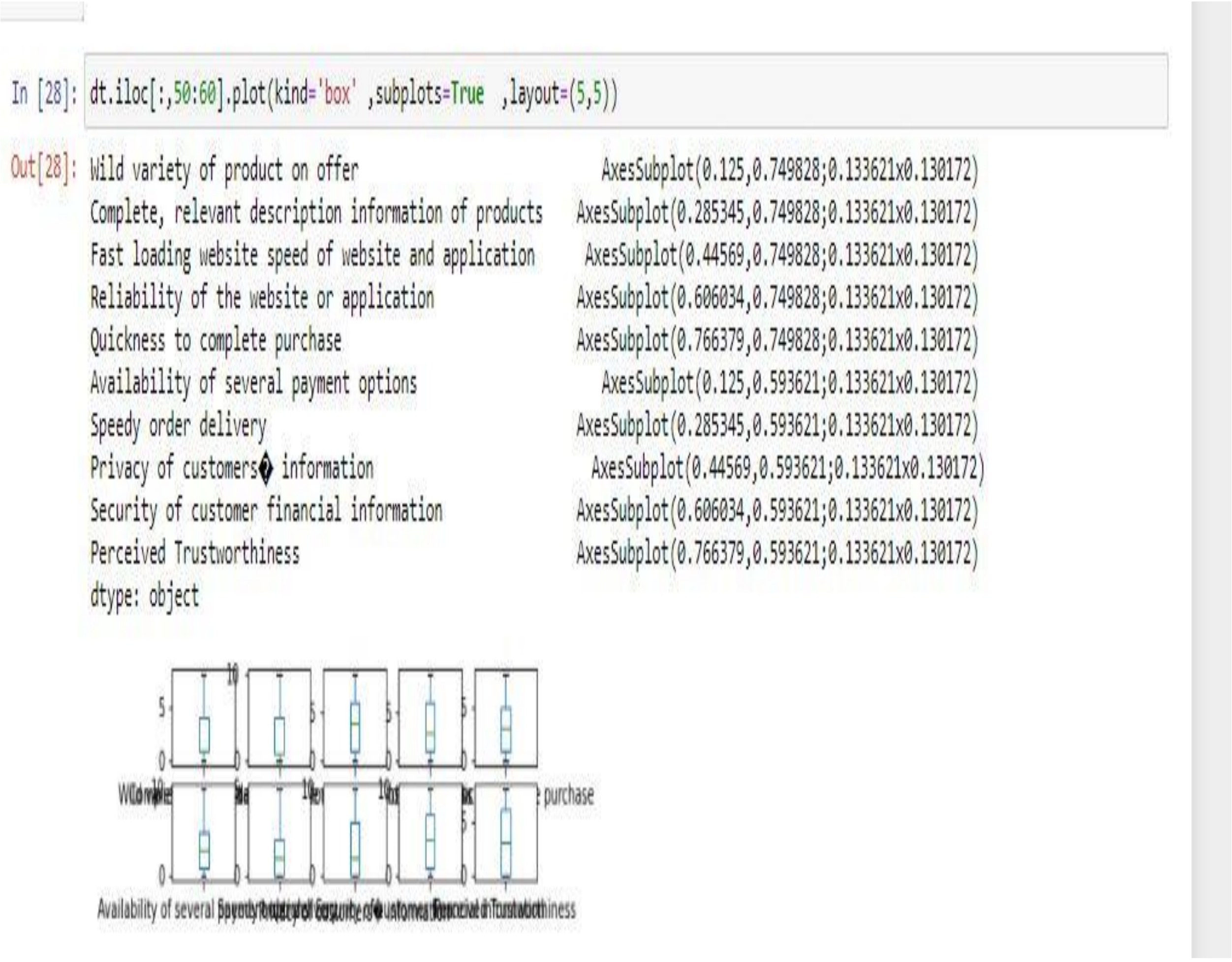
```
[25]: 31 Enjoyment is derived from shopping online                                    AxesSubplot(0.125,0.749828;0.133621x
      0.130172)
      32 Shopping online is convenient and flexible                                   AxesSubplot(0.285345,0.749828;0.133621x
      0.130172)
      33 Return and replacement policy of the e-tailer is important for purchase decision   AxesSubplot(0.44569,0.749828;0.133621x
      0.130172)
      34 Gaining access to loyalty programs is a benefit of shopping online           AxesSubplot(0.606034,0.749828;0.133621x
      0.130172)
      35 Displaying quality Information on the website improves satisfaction of customers   AxesSubplot(0.766379,0.749828;0.133621x
      0.130172)
      36 User derive satisfaction while shopping on a good quality website or application   AxesSubplot(0.125,0.593621;0.133621x
      0.130172)
      37 Net Benefit derived from shopping online can lead to users satisfaction      AxesSubplot(0.285345,0.593621;0.133621x
      0.130172)
      38 User satisfaction cannot exist without trust                                 AxesSubplot(0.44569,0.593621;0.133621x
      0.130172)
      39 Offering a wide variety of listed product in several category               AxesSubplot(0.606034,0.593621;0.133621x
      0.130172)
      40 Provision of complete and relevant product information                      AxesSubplot(0.766379,0.593621;0.133621x
      0.130172)
      dtype: object
```



```
In [27]: dt.iloc[:,40:50].plot(kind='box' ,subplots=True  ,layout=(5,5))
```

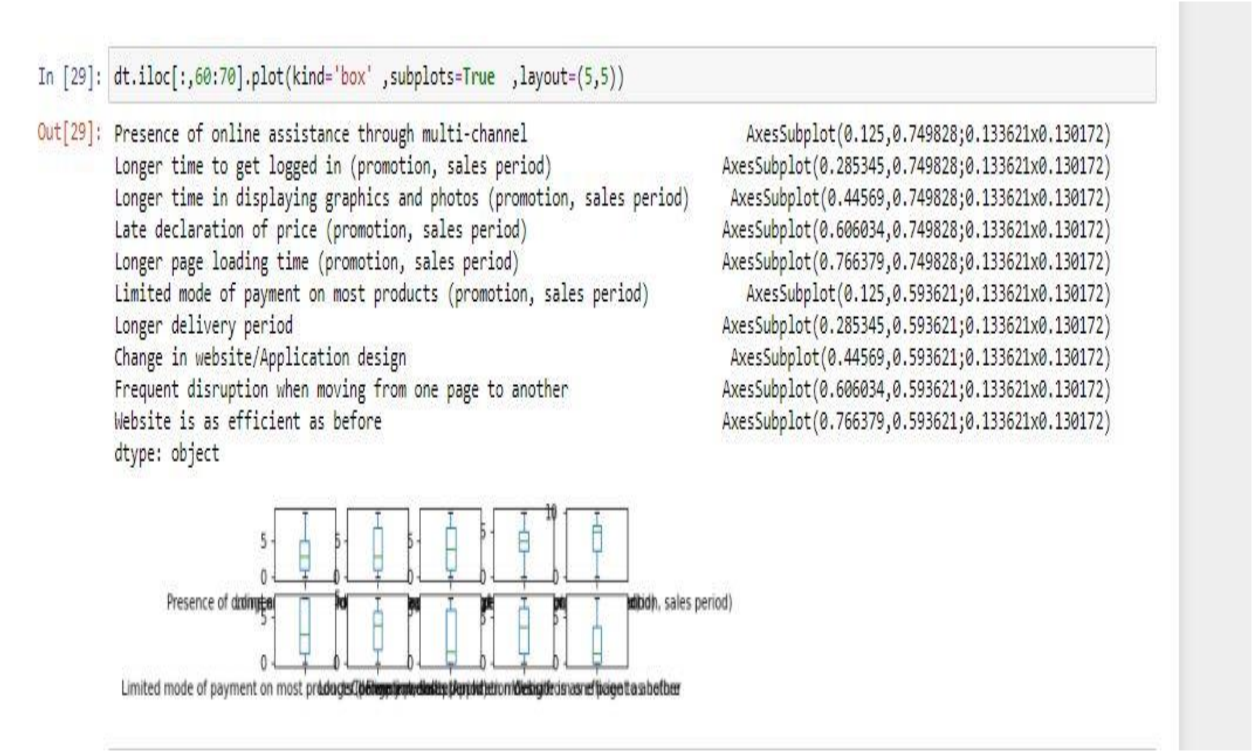
```
Out[27]: 41 Monetary savings
         AxesSubplot(0.125,0.749828;0.133621x0.130172)
         42 The Convenience of patronizing the online retailer
         AxesSubplot(0.285345,0.749828;0.133621x0.130172)
         43 Shopping on the website gives you the sense of adventure
         AxesSubplot(0.44569,0.749828;0.133621x0.130172)
         44 Shopping on your preferred e-tailer enhances your social status
         AxesSubplot(0.606034,0.749828;0.133621x0.130172)
         45 You feel gratification shopping on your favorite e-tailer
         AxesSubplot(0.766379,0.749828;0.133621x0.130172)
         46 Shopping on the website helps you fulfill certain roles
         AxesSubplot(0.125,0.593621;0.133621x0.130172)
         47 Getting value for money spent
         AxesSubplot(0.285345,0.593621;0.133621x0.130172)
         From the following, tick any (or all) of the online retailers you have shopped from;
         AxesSubplot(0.44569,0.593621;0.133621x0.130172)
         Easy to use website or application
         AxesSubplot(0.606034,0.593621;0.133621x0.130172)
         Visual appealing web-page layout
         AxesSubplot(0.766379,0.593621;0.133621x0.130172)
         dtype: object
```

2.5
0.0
2
0
2.5
0.0
2.5
0.0
2.5
0.0
social status
favorite e-tailer
2.5
0.0
From the following, tick any (or all) of the online retailers
Visual appealing web-page layout

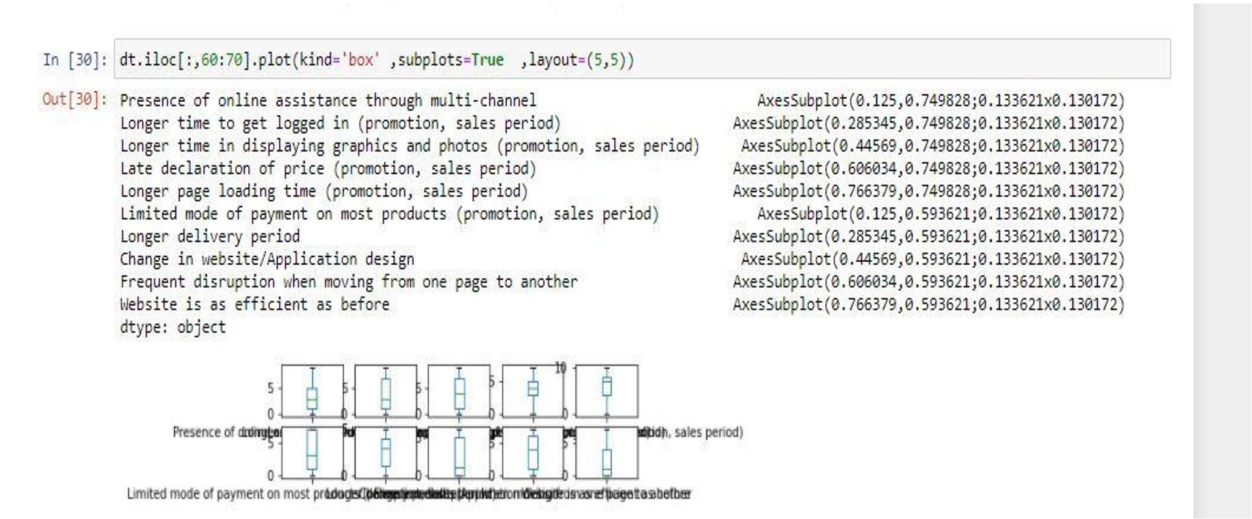


```
In [28]: dt.iloc[:,50:60].plot(kind='box' ,subplots=True  ,layout=(5,5))
```

```
Out[28]: Wild variety of product on offer                        AxesSubplot(0.125,0.749828;0.133621x0.130172)
         Complete, relevant description information of products  AxesSubplot(0.285345,0.749828;0.133621x0.130172)
         Fast loading website speed of website and application    AxesSubplot(0.44569,0.749828;0.133621x0.130172)
         Reliability of the website or application               AxesSubplot(0.606034,0.749828;0.133621x0.130172)
         Quickness to complete purchase                          AxesSubplot(0.766379,0.749828;0.133621x0.130172)
         Availability of several payment options                   AxesSubplot(0.125,0.593621;0.133621x0.130172)
         Speedy order delivery                                   AxesSubplot(0.285345,0.593621;0.133621x0.130172)
         Privacy of customers� information                        AxesSubplot(0.44569,0.593621;0.133621x0.130172)
         Security of customer financial information              AxesSubplot(0.606034,0.593621;0.133621x0.130172)
         Perceived Trustworthiness                               AxesSubplot(0.766379,0.593621;0.133621x0.130172)
         dtype: object
```

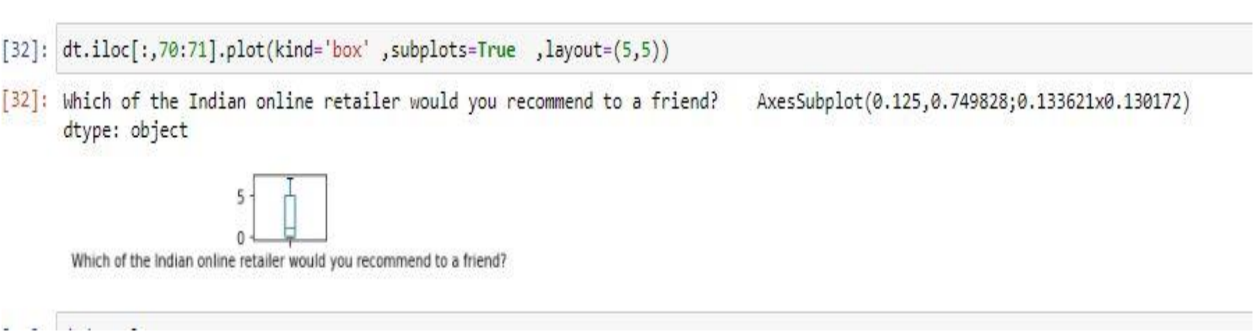


```
In [29]: dt.iloc[:,60:70].plot(kind='box' ,subplots=True  ,layout=(5,5))

Out[29]: Presence of online assistance through multi-channel                    AxesSubplot(0.125,0.749828;0.133621x0.130172)
         Longer time to get logged in (promotion, sales period)              AxesSubplot(0.285345,0.749828;0.133621x0.130172)
         Longer time in displaying graphics and photos (promotion, sales period)  AxesSubplot(0.44569,0.749828;0.133621x0.130172)
         Late declaration of price (promotion, sales period)                 AxesSubplot(0.606034,0.749828;0.133621x0.130172)
         Longer page loading time (promotion, sales period)                  AxesSubplot(0.766379,0.749828;0.133621x0.130172)
         Limited mode of payment on most products (promotion, sales period)     AxesSubplot(0.125,0.593621;0.133621x0.130172)
         Longer delivery period                                              AxesSubplot(0.285345,0.593621;0.133621x0.130172)
         Change in website/Application design                                 AxesSubplot(0.44569,0.593621;0.133621x0.130172)
         Frequent disruption when moving from one page to another            AxesSubplot(0.606034,0.593621;0.133621x0.130172)
         Website is as efficient as before                                   AxesSubplot(0.766379,0.593621;0.133621x0.130172)
         dtype: object
```



```
In [30]: dt.iloc[:,60:70].plot(kind='box' ,subplots=True  ,layout=(5,5))

Out[30]: Presence of online assistance through multi-channel                    AxesSubplot(0.125,0.749828;0.133621x0.130172)
         Longer time to get logged in (promotion, sales period)              AxesSubplot(0.285345,0.749828;0.133621x0.130172)
         Longer time in displaying graphics and photos (promotion, sales period)  AxesSubplot(0.44569,0.749828;0.133621x0.130172)
         Late declaration of price (promotion, sales period)                 AxesSubplot(0.606034,0.749828;0.133621x0.130172)
         Longer page loading time (promotion, sales period)                  AxesSubplot(0.766379,0.749828;0.133621x0.130172)
         Limited mode of payment on most products (promotion, sales period)     AxesSubplot(0.125,0.593621;0.133621x0.130172)
         Longer delivery period                                              AxesSubplot(0.285345,0.593621;0.133621x0.130172)
         Change in website/Application design                                 AxesSubplot(0.44569,0.593621;0.133621x0.130172)
         Frequent disruption when moving from one page to another            AxesSubplot(0.606034,0.593621;0.133621x0.130172)
         Website is as efficient as before                                   AxesSubplot(0.766379,0.593621;0.133621x0.130172)
         dtype: object
```

5
0
10
Presence of online
sales period)
Limited mode of payment on most pr



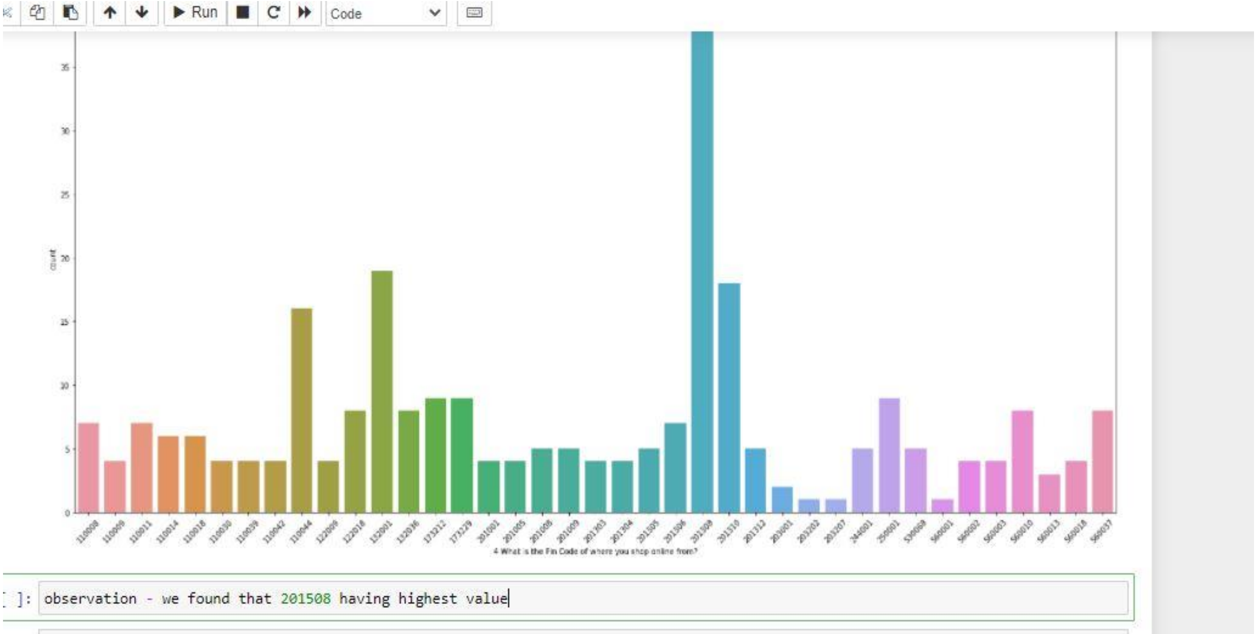
```
[32]: dt.iloc[:,70:71].plot(kind='box' ,subplots=True  ,layout=(5,5))

[32]: Which of the Indian online retailer would you recommend to a friend?    AxesSubplot(0.125,0.749828;0.133621x0.130172)
      dtype: object
```

5
0
Which of the Indian online retailer would you recommend to a friend?

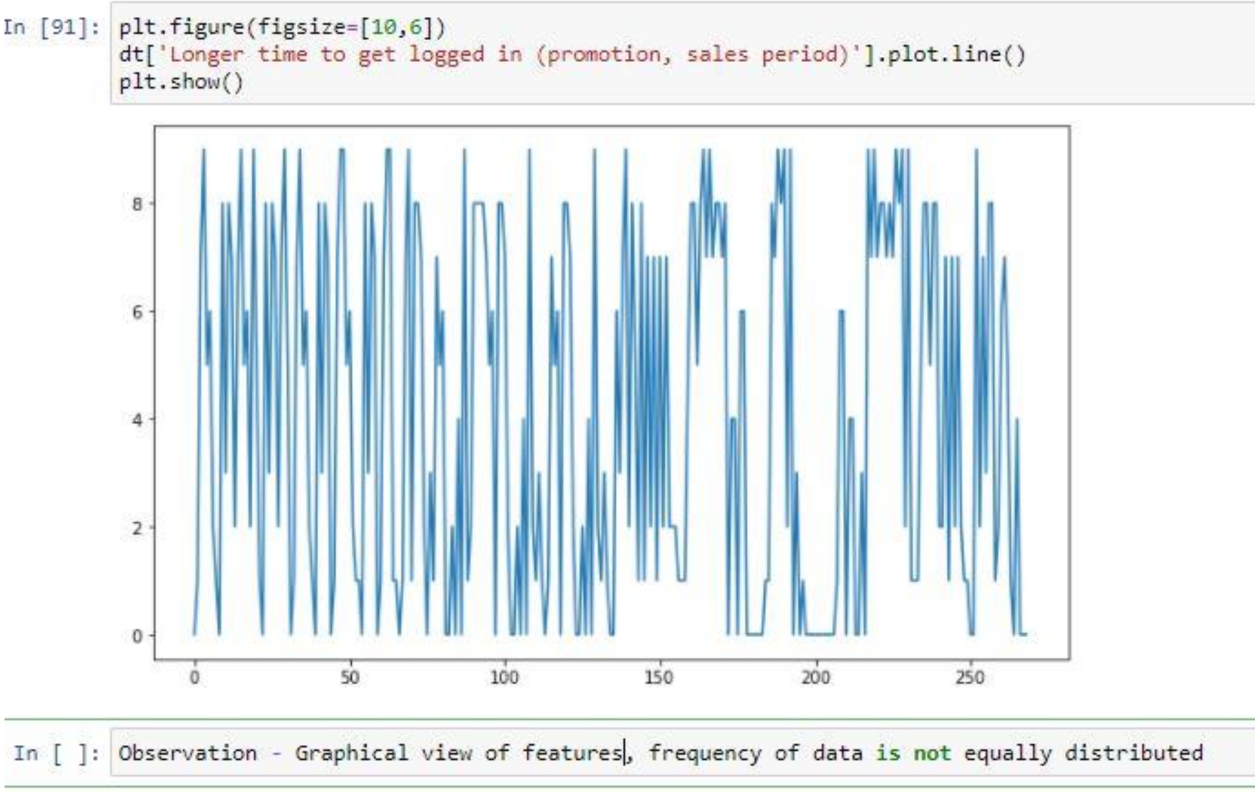
Observation – where dots present above or below the vertices it seems represent outlier basically data due to skew but we can not remove as all these columns type are object from object type of data we can not remove outlier

Some more EDA:

plt.figure(figsize=[25,12])
sns.countplot(x = '4 What is the Pin Code of where you shop online from?', data = dt)
plt.xticks(rotation = 45)

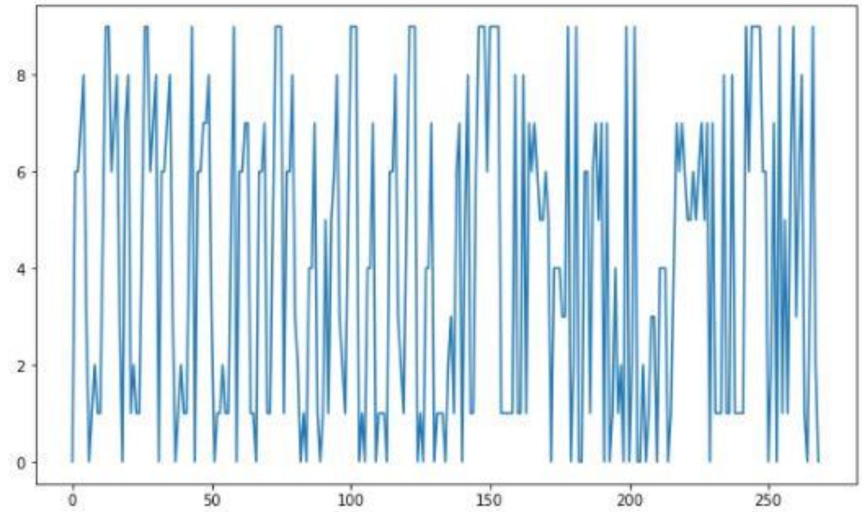


```
[ ]: observation - we found that 201508 having highest value
```



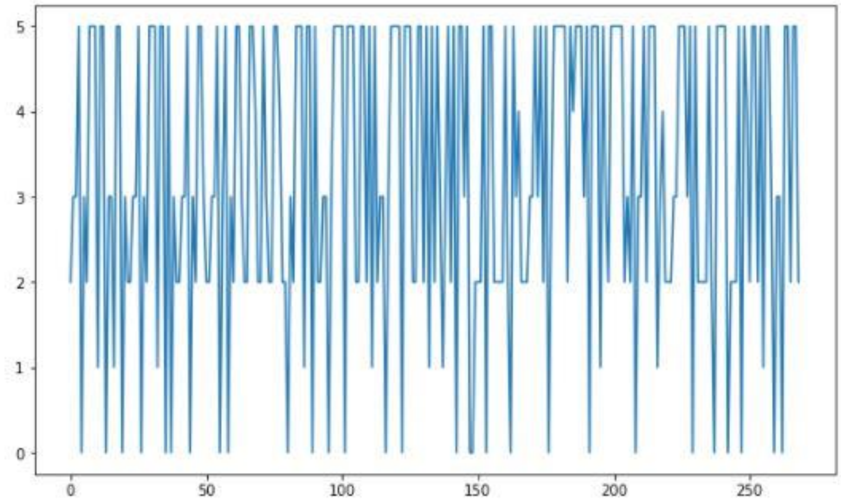
```
In [91]: plt.figure(figsize=[10,6])
         dt['Longer time to get logged in (promotion, sales period)'].plot.line()
         plt.show()
```

```
In [ ]: Observation - Graphical view of features, frequency of data is not equally distributed
In [92]: plt.figure(figsize=[10,6])
         dt['Longer time in displaying graphics and photos (promotion, sales period)'].plot.line()
         plt.show()
```



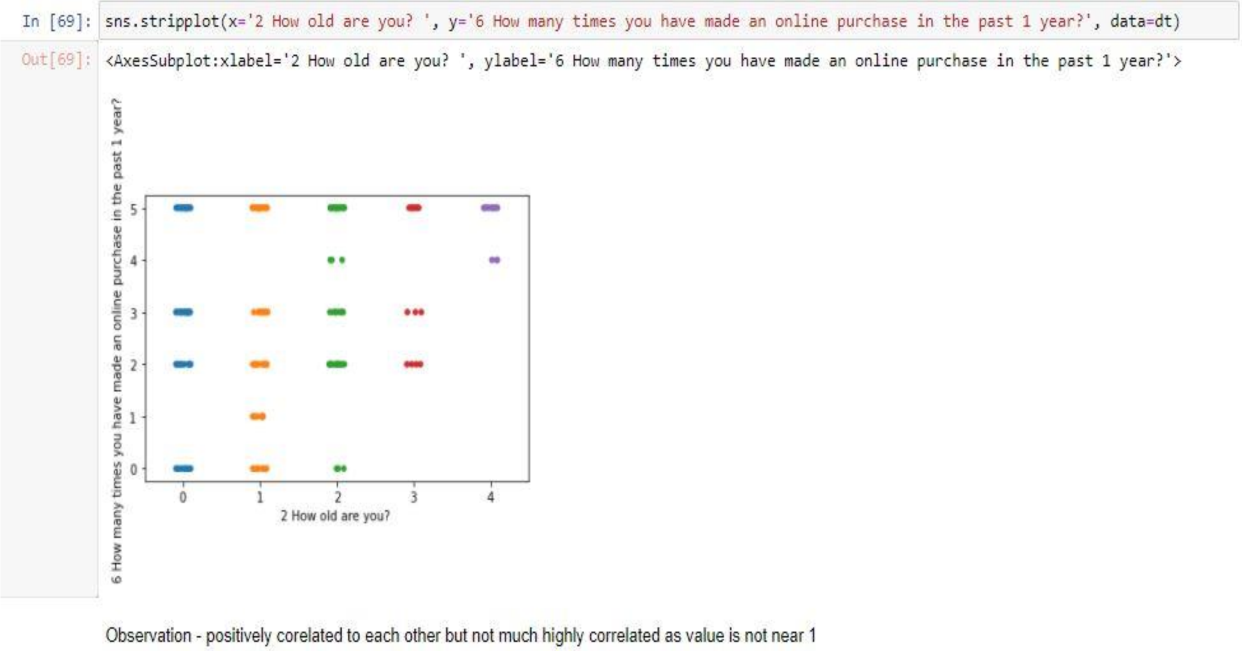
Observation - Graphical view of features, frequency of data is not equally distributed

```
In [20]: plt.figure(figsize=[10,6])
         dt['6 How many times you have made an online purchase in the past 1 year?'].plot.line()
         plt.show()
```



Observation - Graphical view of target variable , frequency of data is not equally distributed

Bivariate---- From df.corr () we get correlationship value, from there we found which variables are highly correlated with each other and which are negatively correlated. Here we plotted graphical representation using strip

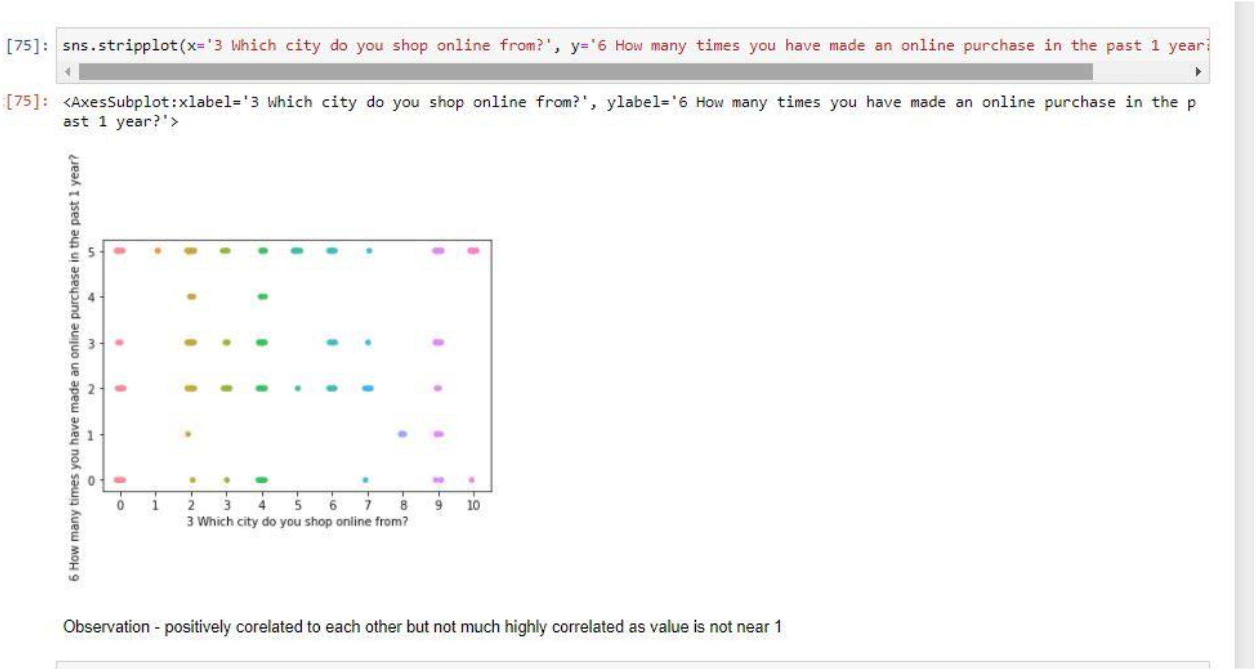


```
In [69]: sns.stripplot(x='2 How old are you? ', y='6 How many times you have made an online purchase in the past 1 year?', data=dt)
```

```
Out[69]: <AxesSubplot:xlabel='2 How old are you? ', ylabel='6 How many times you have made an online purchase in the past 1 year?'>
```

6 How many times you have made an online purchase in the past 1 year?
0
1
2
3
4
5
0
1
2
3
4
2 How old are you?

Observation - positively corelated to each other but not much highly correlated as value is not near 1

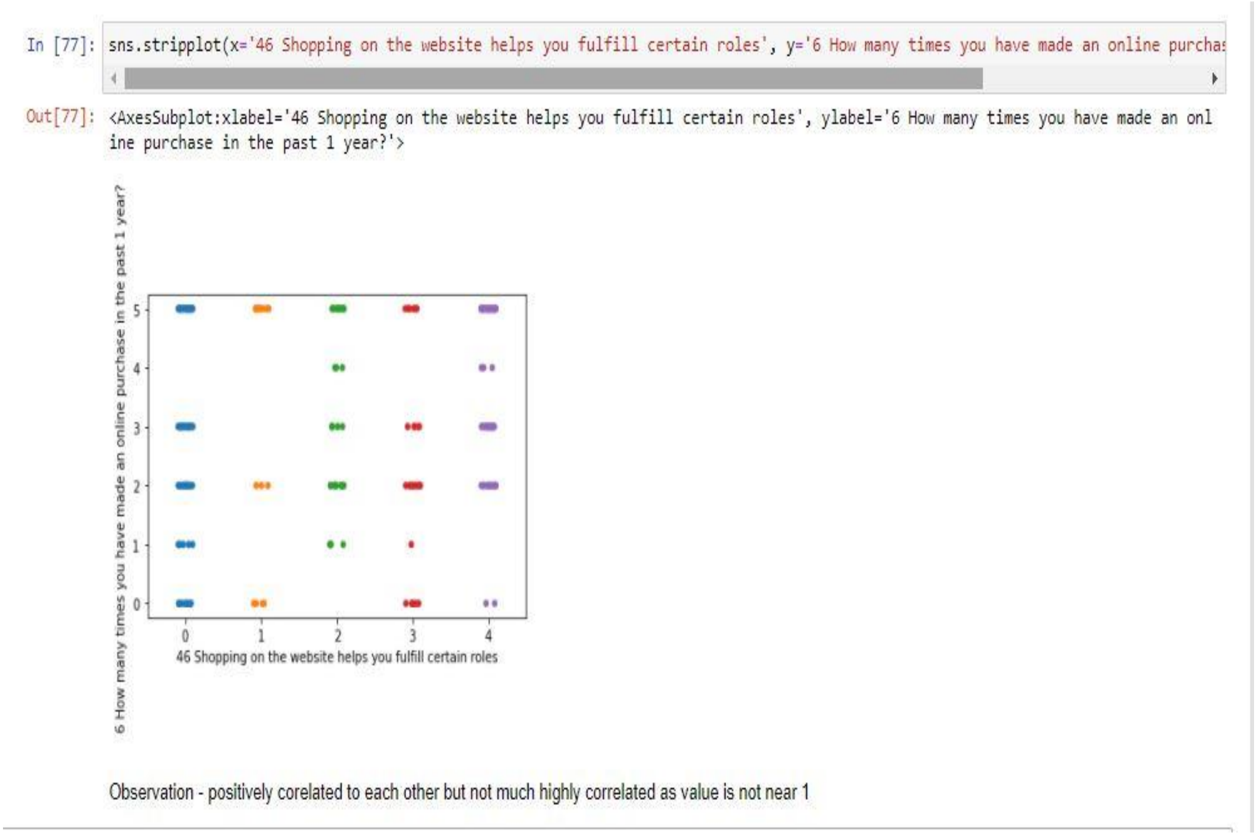


```
[75]: sns.stripplot(x='3 Which city do you shop online from?', y='6 How many times you have made an online purchase in the past 1 year
```

```
[75]: <AxesSubplot:xlabel='3 Which city do you shop online from?', ylabel='6 How many times you have made an online purchase in the p
ast 1 year?'>
```

6 How many times you have made an online purchase in the past 1 year?
0
1
2
3
4
5
0
1
2
3
4
5
6
7
8
9
10
3 Which city do you shop online from?

Observation - positively corelated to each other but not much highly correlated as value is not near 1

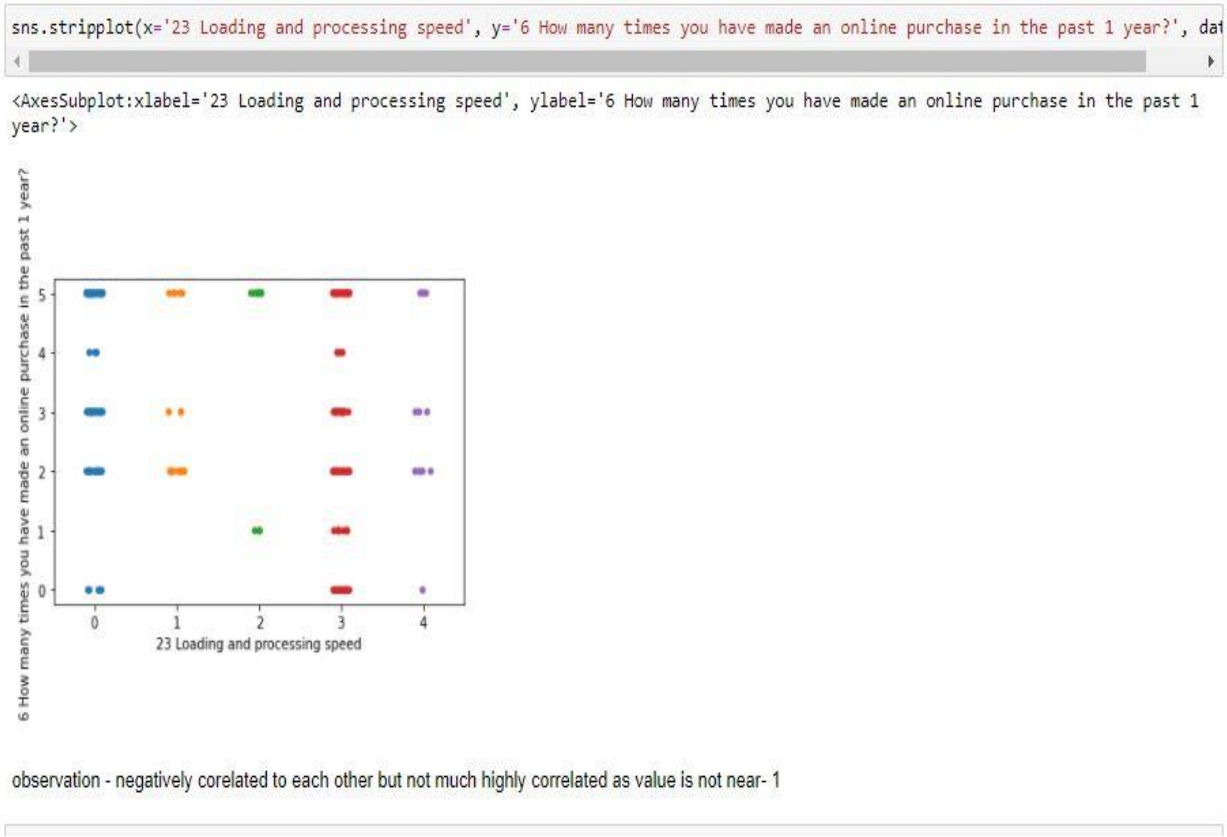


```
In [77]: sns.stripplot(x='46 Shopping on the website helps you fulfill certain roles', y='6 How many times you have made an online purchas
```

```
Out[77]: <AxesSubplot:xlabel='46 Shopping on the website helps you fulfill certain roles', ylabel='6 How many times you have made an onl
ine purchase in the past 1 year?'>
```

6 How many times you have made an online purchase in the past 1 year?
5
4
3
2
1
0
0
1
2
3
4
46 Shopping on the website helps you fulfill certain roles

Observation - positively corelated to each other but not much highly correlated as value is not near 1

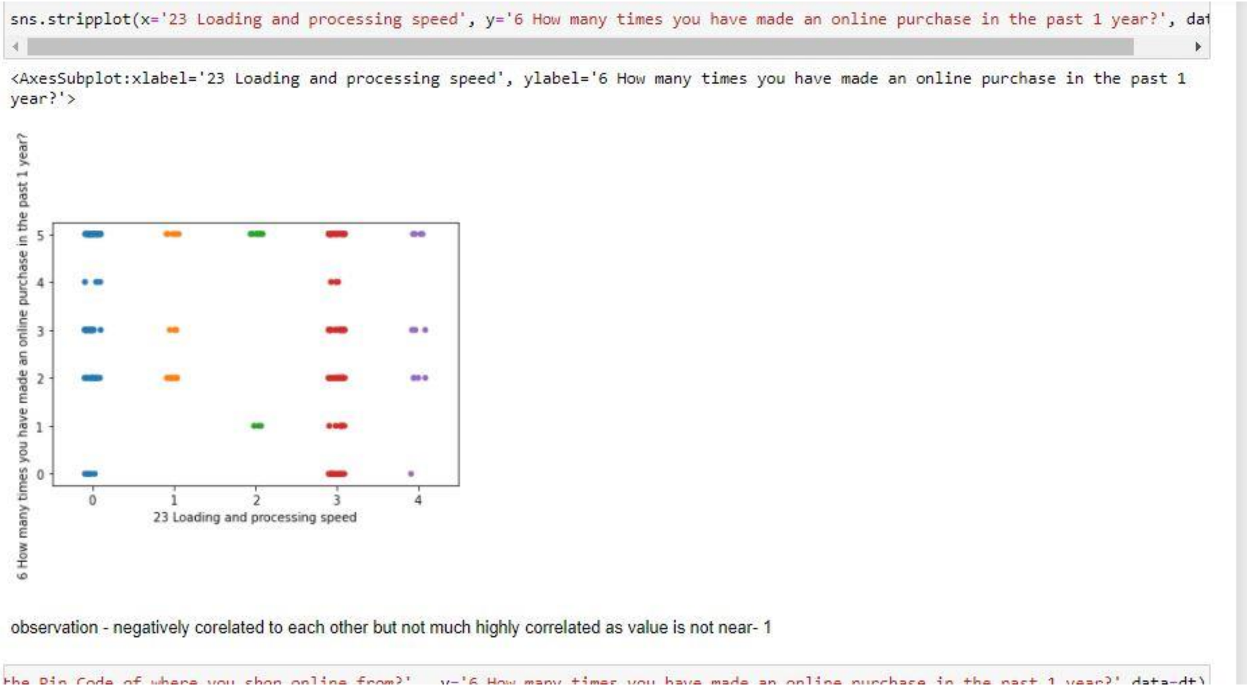


```
sns.stripplot(x='23 Loading and processing speed', y='6 How many times you have made an online purchase in the past 1 year?', dat
```

```
<AxesSubplot:xlabel='23 Loading and processing speed', ylabel='6 How many times you have made an online purchase in the past 1
year?'>
```

6 How many times you have made an online purchase in the past 1 year?
5
4
3
2
1
0
0
1
2
3
4
23 Loading and processing speed

observation - negatively corelated to each other but not much highly correlated as value is not near- 1



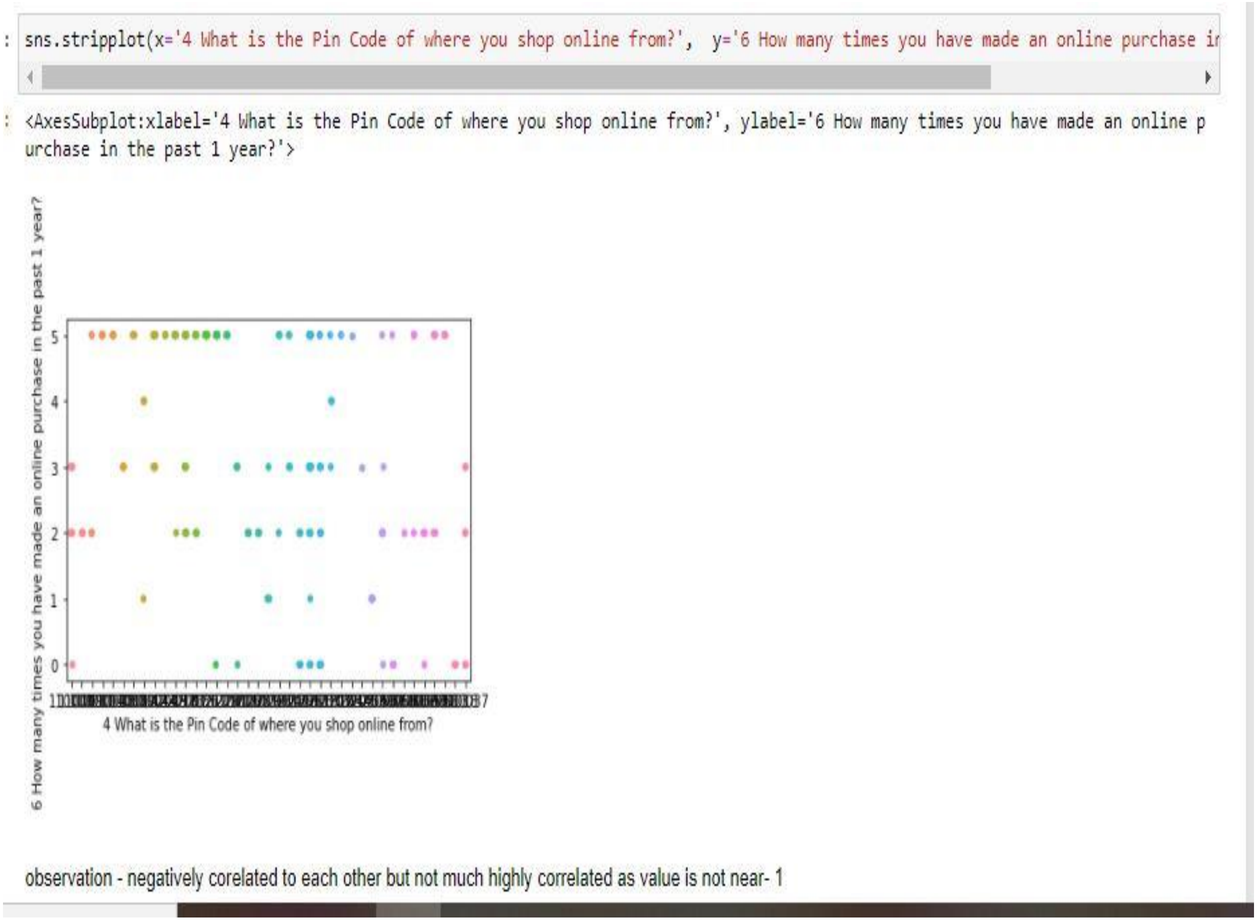
```
sns.stripplot(x='23 Loading and processing speed', y='6 How many times you have made an online purchase in the past 1 year?', dat
```

```
<AxesSubplot:xlabel='23 Loading and processing speed', ylabel='6 How many times you have made an online purchase in the past 1
year?'>
```

6 How many times you have made an online purchase in the past 1 year?
0
1
2
3
4
5
0
1
2
3
4
23 Loading and processing speed

observation - negatively corelated to each other but not much highly correlated as value is not near- 1

```
the Pin Code of where you shop online from?',  y='6 How many times you have made an online purchase in the past 1 year?', data=dt)
```

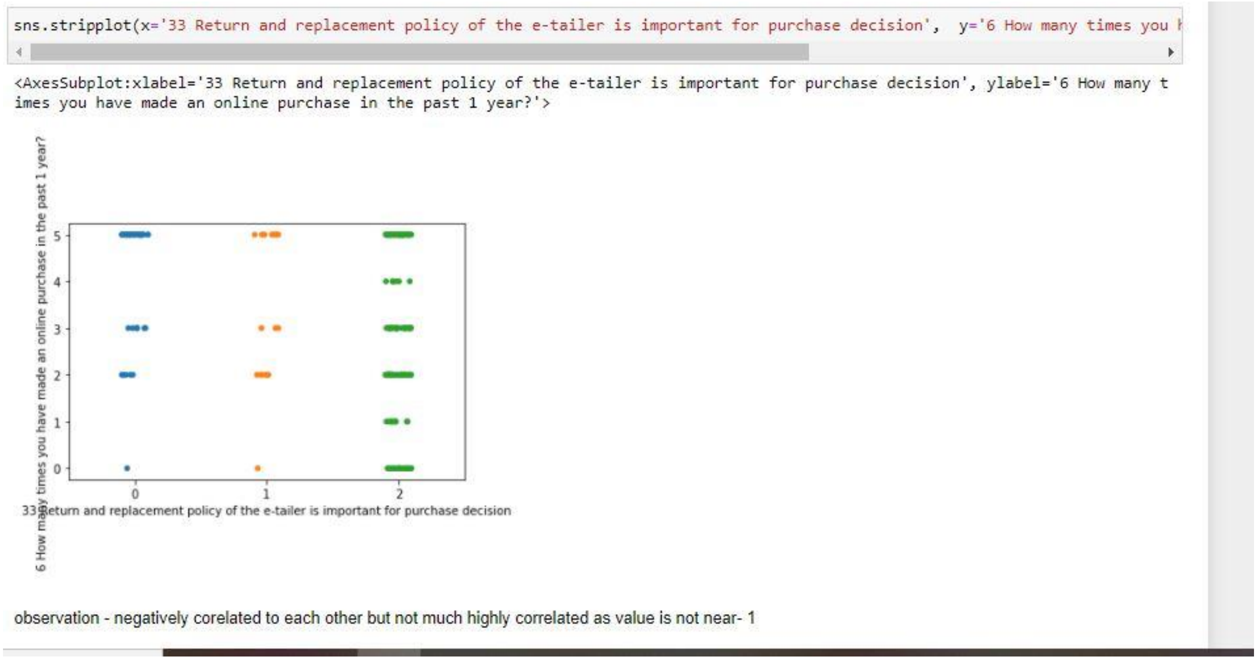


```
sns.stripplot(x='4 What is the Pin Code of where you shop online from?',  y='6 How many times you have made an online purchase ir
```

```
<AxesSubplot:xlabel='4 What is the Pin Code of where you shop online from?', ylabel='6 How many times you have made an online p
urchase in the past 1 year?'>
```

6 How many times you have made an online purchase in the past 1 year?
0
1
2
3
4
5
4 What is the Pin Code of where you shop online from?

observation - negatively corelated to each other but not much highly correlated as value is not near- 1

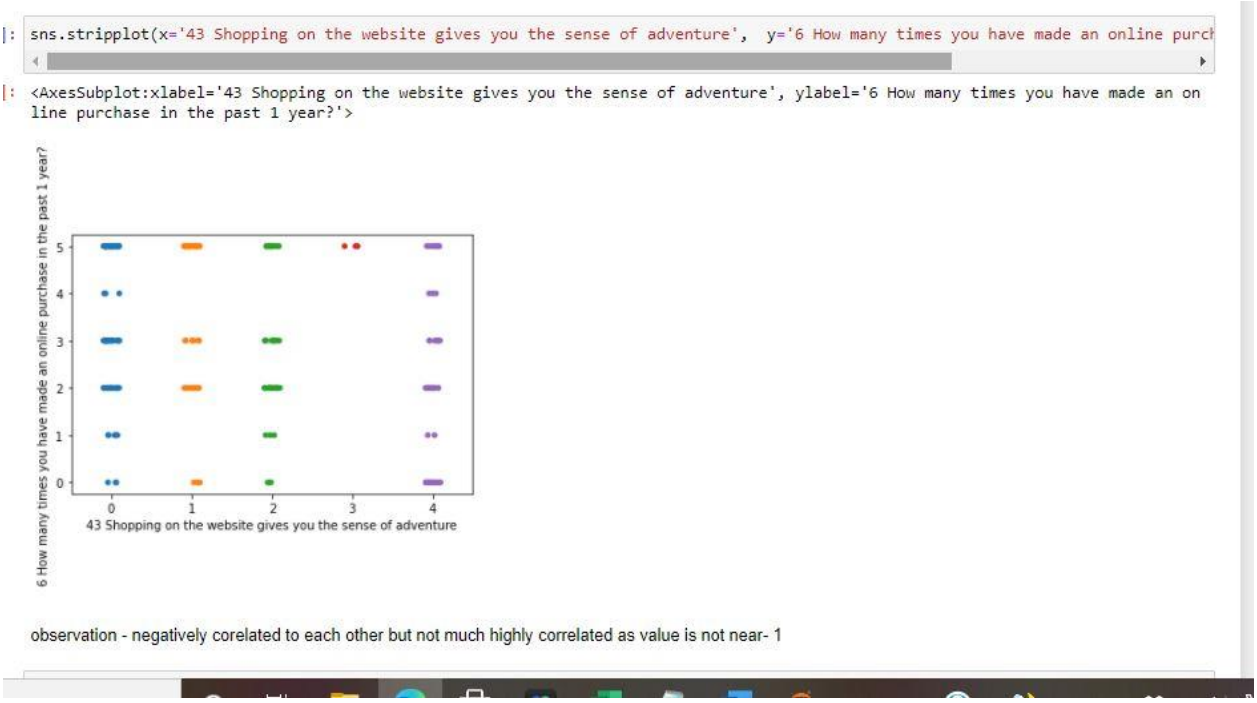


```
sns.stripplot(x='33 Return and replacement policy of the e-tailer is important for purchase decision',  y='6 How many times you h
```

```
<AxesSubplot:xlabel='33 Return and replacement policy of the e-tailer is important for purchase decision', ylabel='6 How many t
imes you have made an online purchase in the past 1 year?'>
```

6 How many times you have made an online purchase in the past 1 year?
5
4
3
2
1
0
0
1
2
33 Return and replacement policy of the e-tailer is important for purchase decision

observation - negatively corelated to each other but not much highly correlated as value is not near- 1



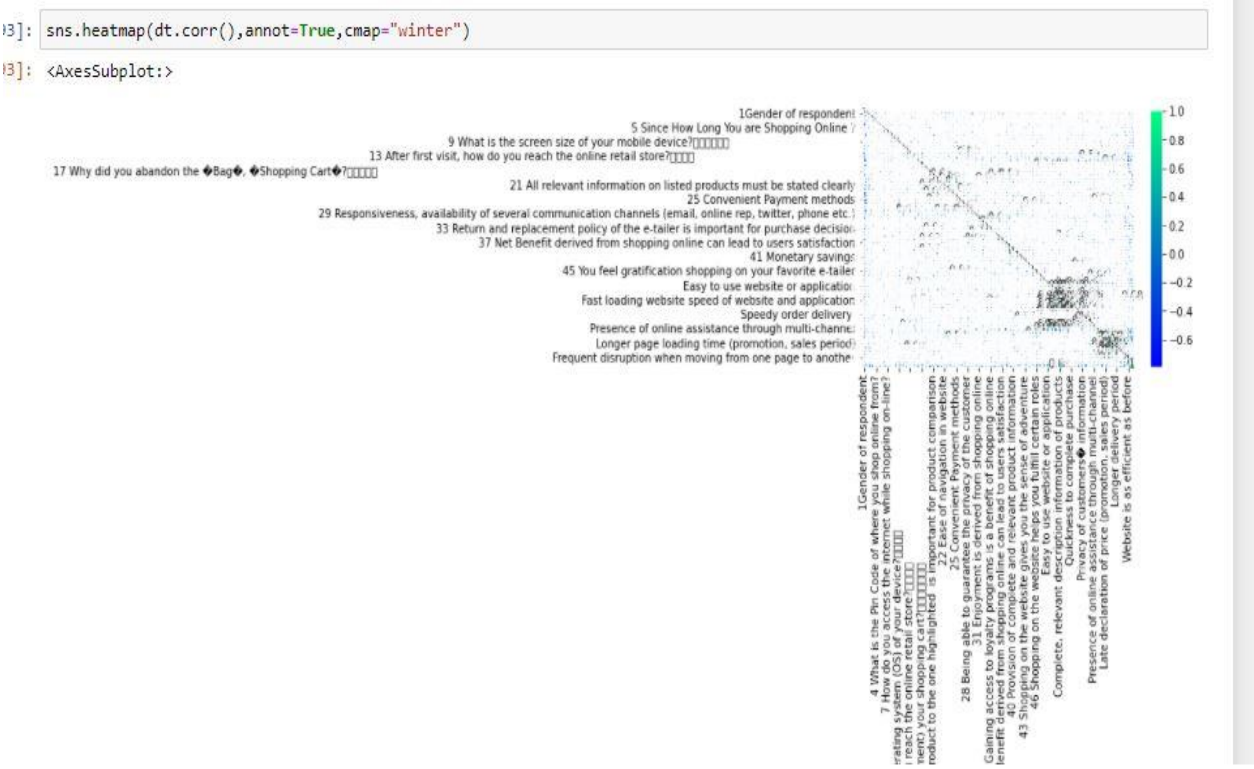
```
]: sns.stripplot(x='43 Shopping on the website gives you the sense of adventure',  y='6 How many times you have made an online purch
```

```
]: <AxesSubplot:xlabel='43 Shopping on the website gives you the sense of adventure', ylabel='6 How many times you have made an on
   line purchase in the past 1 year?'>
```

6 How many times you have made an online purchase in the past 1 year?
5
4
3
2
1
0
0
1
2
3
4
43 Shopping on the website gives you the sense of adventure

observation - negatively corelated to each other but not much highly correlated as value is not near- 1

Multivariate analysis—
We plot heatmap and pair plot to get multiplot idea



```
]3]: sns.heatmap(dt.corr(),annot=True,cmap="winter")
```

```
]3]: <AxesSubplot:>
```

1Gender of respondent
5 Since How Long You are Shopping Online ?
9 What is the screen size of your mobile device?□□□□□□
13 After first visit, how do you reach the online retail store?□□□□
17 Why did you abandon the �Bag�, �Shopping Cart�?□□□□□
21 All relevant information on listed products must be stated clearly
25 Convenient Payment methods
29 Responsiveness, availability of several communication channels (email, online rep, twitter, phone etc.)
33 Return and replacement policy of the e-tailer is important for purchase decisio
37 Net Benefit derived from shopping online can lead to users satisfaction
41 Monetary savings
45 You feel gratification shopping on your favorite e-tailer
Easy to use website or applicatio
Fast loading website speed of website and application
Speedy order delivery
Presence of online assistance through multi-channe
Longer page loading time (promotion, sales period)
Frequent disruption when moving from one page to anothe
1Gender of respondent
4 What is the Pin Code of where you shop online from?
7 How do you access the internet while shopping on-line?
erating system (OS) of your device?□□□□
reach the online retail store?□□□□
nent) your shopping cart?□□□□□□
roduct to the one highlighted is important for product comparison
22 Ease of navigation in website
25 Convenient Payment methods
28 Being able to guarantee the privacy of the customer
31 Enjoyment is derived from shopping online
Gaining access to loyalty programs is a benefit of shopping online
enefit derived from shopping online can lead to users satisfaction
40 Provision of complete and relevant product information
43 Shopping on the website gives you the sense of adventure
46 Shopping on the website helps you fulfill certain roles
Easy to use website or application
Complete, relevant description information of products
Quickness to complete purchase
Privacy of customers� information
Presence of online assistance through multi-channel
Late declaration of price (promotion, sales period)
Longer delivery period
Website is as efficient as before
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

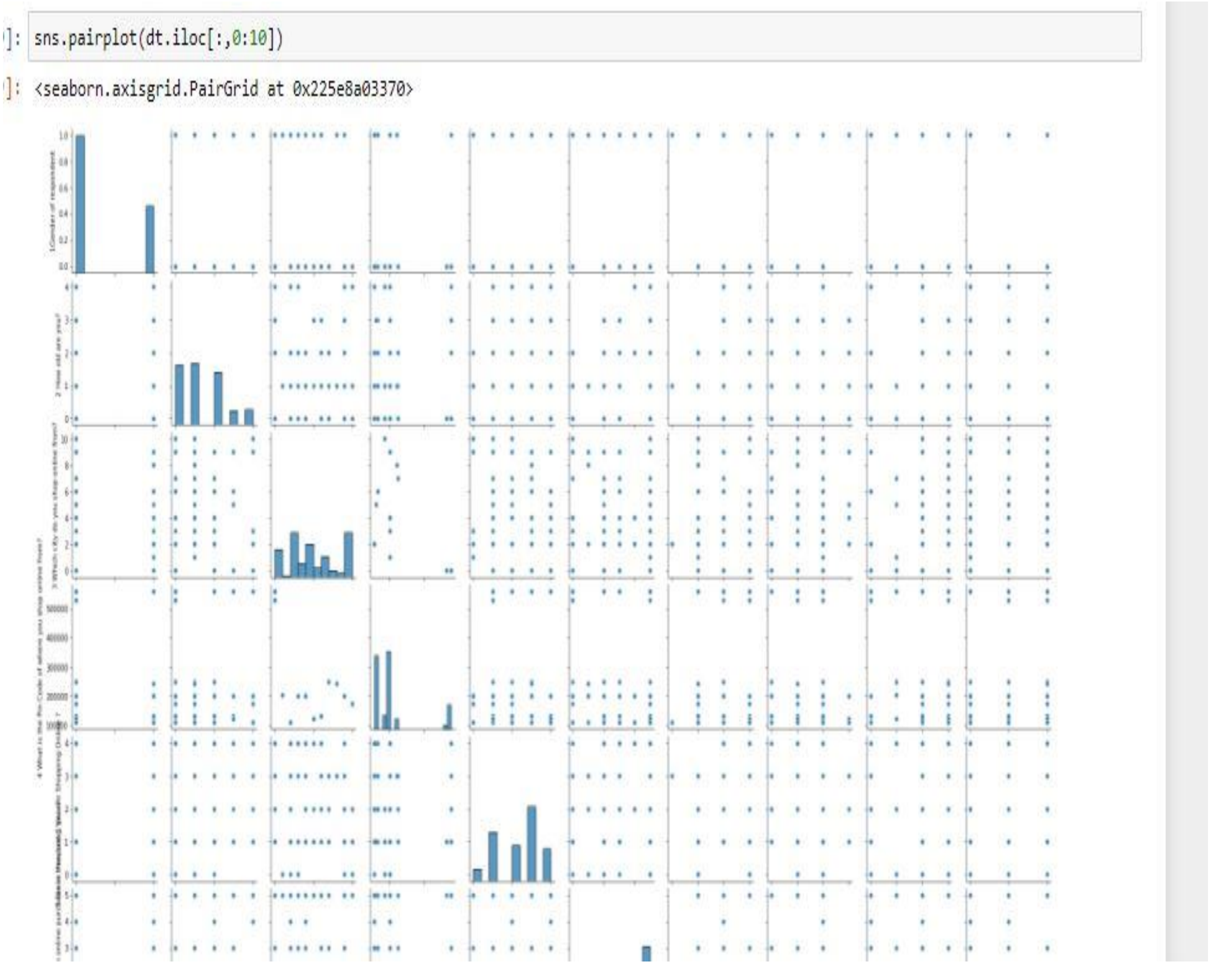
.

Pair plot:

Plot pairwise relationships in a dataset.

By default, this function will create a grid of Axes such that each numeric variable in data will by shared across the y-axes across a single row and the x-axes across a single column. The diagonal plots are treated differently: a univariate distribution plot is drawn to show the marginal distribution of the data in each column.

It is also possible to show a subset of variables or plot different variables on the rows and columns.



```
]: sns.pairplot(dt.iloc[:,0:10])

]: <seaborn.axisgrid.PairGrid at 0x225e8a03370>
```

Motivation for the Problem Undertaken

we study this model so this will help us to analyse. In order to improve the factors, we need to analysis the dataset which is playing vital role to hold the customer. so here we will be analysis the data based on customer feedback. In this dataset target variable is' How many times you have made an online purchase in the past 1 year' which will represent value as 1 to 5 based on numbers of time customer made purchase throughout the year. **Label '1' indicates that least headcount of people according to number of purchase Label '5' indicates Less than 10 times (maximum people purchased**

As we worked with real time data, we have gained knowledge that what are challenges has to face while working with real domain data (heavy data set), sometimes some information is uncertain so using this experience I believe we can work better on next project and that is being the best motivation behind this project work

Mathematical/ Analytical Modeling of the Problem

supervised learning uses labelled input and output data Supervised learning (SL) is the machine learning task of learning a function that maps an input to an output based on example input-output pairs.[It infers a function from *labelled training data* consisting of a set of *training examples*.[In supervised learning, each example is a *pair* consisting of an input object (typically a vector) and a desired output value (also called the *supervisory signal*). A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples. An optimal scenario will allow for the algorithm to correctly determine the class labels for unseen instances

Here our   dataset consist of Categorical data which is part of supervised learning so we will analyse with classification (Logistic classification)

Classification is a process in which an algorithm is used to analyze an existing data   set of known points. The understanding achieved through that analysis is then leveraged as a means of appropriately classifying the data. Classification is a form of machine learning that can be particularly helpful in analyzing very large, complex sets of data to help make more accurate predictions.

Data Sources and their formats

Data provided by Fliprobo which they have been provided by client

Using below command we got some basic information of data which is mentioned below

df.info()--- it provided object type of each columns .our dataset content of 209593 rows × 36 columns

2.df.dypes= its provided info that what the data type belongs to (float , int)

3 df.isnull.sum()--- we found there is no null value

4 df.head()--- it shows first five columns in the dataset

5 df.columns—it shows total columns of the dataset

6> df1[column name].value_counts()—provide unique value of this particular column

Data Pre-processing

Using label-encoder we converted categorical data to numeric as saved at df1 file We calculated correlation using df.corr () and plot as heat map for checking correlationship

As this is categorical data we cannot find mean so unable to calculate standard deviation , for categorical data that’s being the reason we cannot remove outlier or cannot define skewness and same informed by DATA trained mentor too.

Hardware and Software Requirements and Tools Used

Hardware

• Good performance PC [Minimum – 8gb RAM +SSD]

• Enough space in hard disk drive Software requirements

• jupyter note book • Sometimes you may need Google colab to cross check the output Package

• Numpy ---import numpy as np (For calculation)

• Panda-import pandas as pd (read data frame)

• Imblearn----- For class sampling Here the list of some other function

• For plotting- 1>import seaborn as sns

2> import matplotlib.pyplot as plt

• For ignore new version warning--- import warnings warnings.filterwarnings('ignore')'

• For class balancing----from imblearn.over_sampling import SMOTE • from sklearn.linear_model import LogisticRegression

• from sklearn.model_selection import train_test_split

• from sklearn.naive_bayes import MultinomialNB

• from sklearn.svm import SVC

• from sklearn.tree import DecisionTreeClassifier

• from sklearn.neighbors import KNeighborsClassifier

• from sklearn.ensemble import AdaBoostClassifier

• from sklearn.ensemble import RandomForestClassifier

• from sklearn.metrics import confusion_matrix, classification_report ,accuracy score

Model/s Development and Evaluation

Testing of Identified Approaches (Algorithms)

We have performed train test where we have send data to model (some data for training and some for testing). We have used 5 model to

- Decisions Classifier Model
- Random Forest Model
- Ada-boost Model
- SVC Model

ALGORITHIM

DecissionTree Classifier Model:

```
dtc=DecisionTreeClassifier()
dtc.fit(x_train,y_train)
preddtc=dtc.predict(x_test)
print ("acccuracy score" , accuracy_score(y_test,preddtc))
print("confusion matrix", confusion matrix(y_test,preddtc))
print("clasification report",classification_report(y_test,preddtc))
```

OutputRandom Forest Model- acccuracy score 0.9337748344370861

RandomForestClassifier

```
from sklearn.ensemble  import  RandomForestClassifier


rf=RandomForestClassifier( n_estimators=100,
random_state=42)
rf.fit(x_train, y_train) predrf=rf.predict(x_test)
print(accuracy_score (y_test,predrf))
print(confusion_matrix(y_test, predrf))
print(classification_report(y_test,predrf))
```

Output of accuracy score =
0.9337748344370861

Ada-boost Model

```
ad=AdaBoostClassifier(n_estimators=50
) ad.fit(x_train, y_train)
adprd=ad.predict(x_test)
print(accuracy_score(y_test,adprd))
print(confusion_matrix(y_test, adprd))
print(classification_report(y_test,adprd))
```

Output of accuracy score   -0.27

SVC model

```
from sklearn.svm import LinearSVC

clf = LinearSVC(random_state=0, tol=1e-5)

clf.fit(x_train, y_train.ravel())
predsvc=sv.predict(x_test)
print(sv.score(x_train,y_train.ravel()))
print("acccuracy score" ,
accuracy_score(y_test,predsvc))
print("confusion matrix",
confusion_matrix(y_test,predsvc))
print("clasification
report",classification_report(y_test,predsvc))
```

```
acccuracy score 0.31
```

Best model selection

We have calculated cross validation score of each model. Cross validation is a statistical method used to estimate the skill of machine learning models   and we found RandomForestClassifier has having less difference between accuracy and cross validation score .So as per logic RFC is our best model

Conclusion

- we have transformed categorical data to numeric using Label Encoder
- We have plotted graphical view of each column to understand data distribution using count plot as well as for finding outlier concept we plotted boxplot
- As this is categorical data we cannot remove outlier as mean concept not there in categorical data, same confirmed by Data Trained mentor

- we divided data x and y as a data and target
- we analysis all the model and found only RFC is having less difference between accuracy and cross_val_score
- We optimize model using hyper tuning parameter (hyper parameter optimization or tuning is the problem of choosing a set of optimal hyperparameters for a learning algorithm. These measures are called hyperparameters, and have to be tuned so that the model can optimally solve the machine learning problem)
- We got our final model
- We saved out final model in as .pkl file as per client requirement . It is basically Binary format of output

Limitation: -

The data could be incomplete. even the lack of a section or a substantial part of the data, could limit its usability.

We don’t get always accurate information as data might be not completed .

As it is real time data , it is complex data, took long time to execute

□

□

□