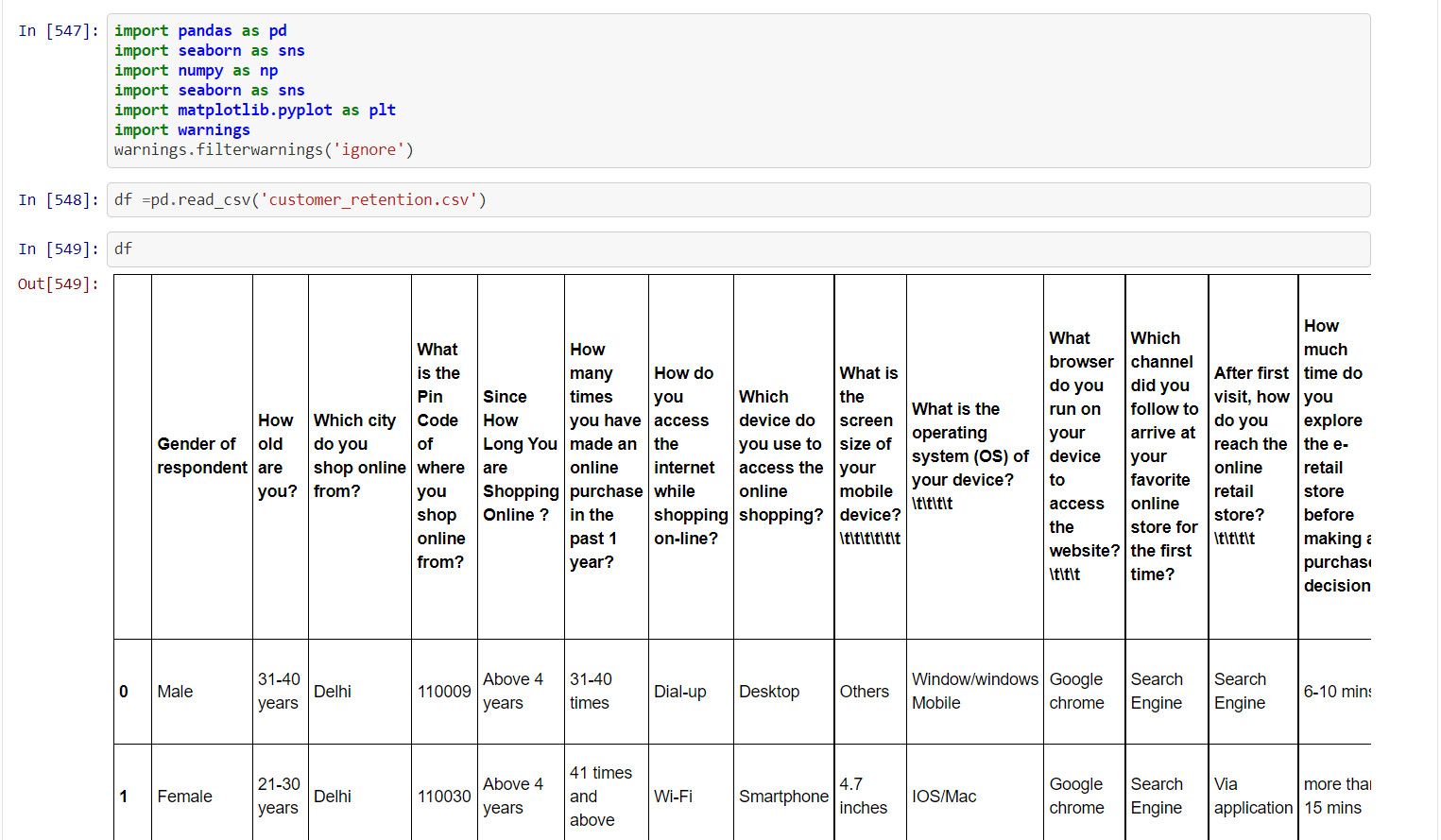
**Customer Retention**

Customer retention refers to the activities and actions companies and organizations take to reduce the number of customer defections. The goal of customer retention programs is to help companies retain as many customers as possible, often through customer loyalty and brand loyalty initiatives. It is important to remember that customer retention begins with the first contact a customer has with a company and continues throughout the entire lifetime of the relationship.

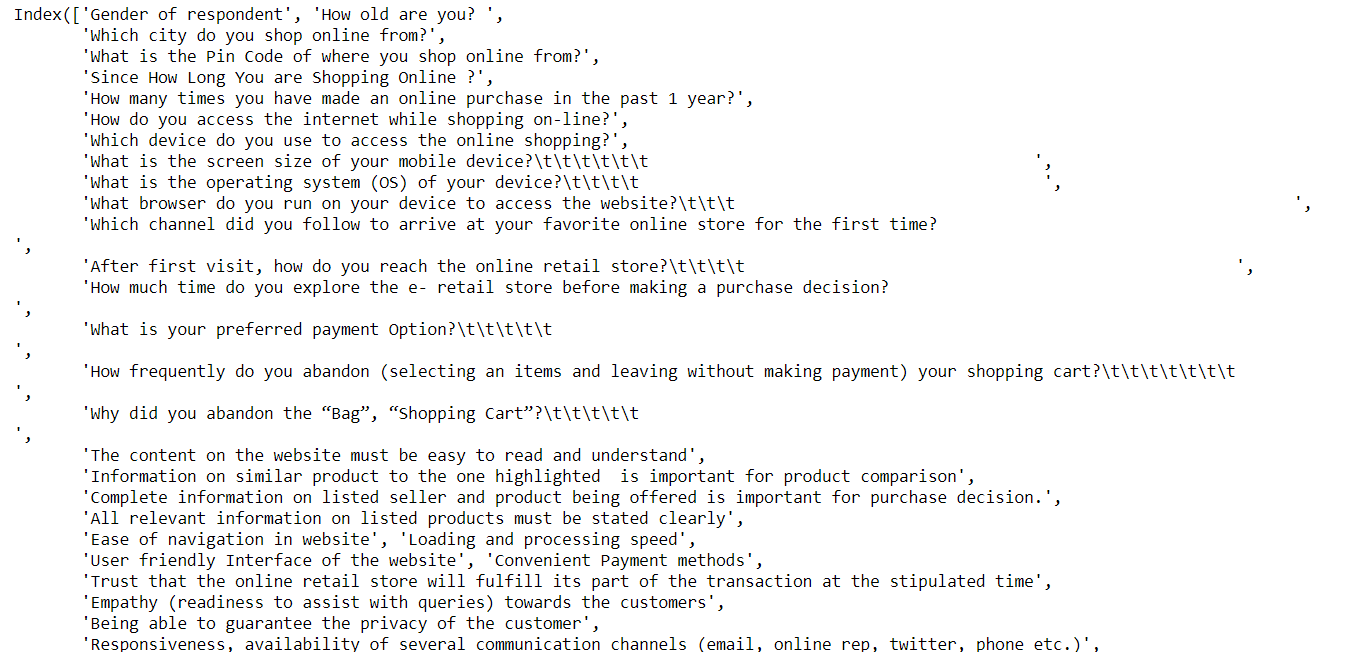
Load the basic libariries and Load the data of customer Retention



There many columns that show how customer action towards company in this case e-retailer e-commerce platform

df.columns

show the columns present in data set



#### display max columns and rows

pd.set\_option("display.max\_columns", None)

pd. set\_option("display.max\_rows", None)

### Replace the space and unwanted space from columns

df.columns = df.columns.str.replace(' ', '')

df.columns = df.columns.str.replace('\t\t\t\t\t\t', '')

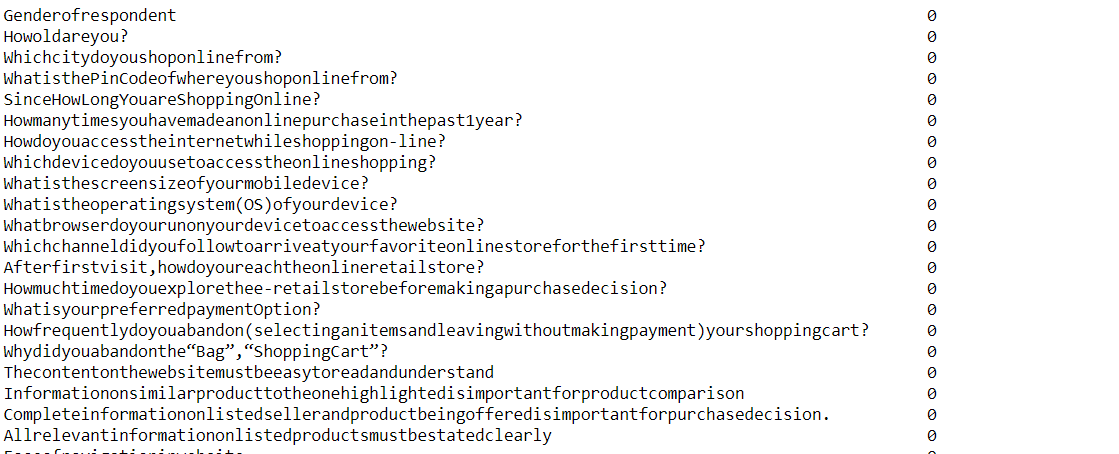
df.columns = df.columns.str.replace('\t\t\t\t', '')

df.columns = df.columns.str.replace('\t\t\t', '')

df.columns = df.columns.str.replace('\t', '')

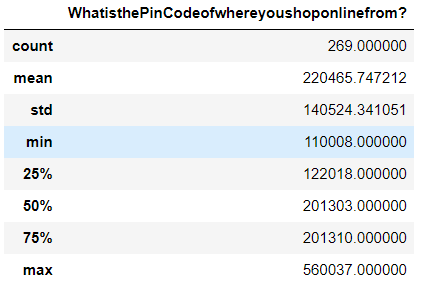
### check count of null value

df.isnull().sum()



#### statistical description show only numeric value

df.describe()



#### statistical data type description

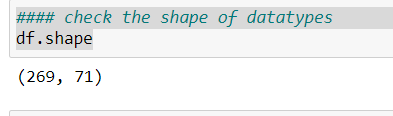
df.dtypes



Most of the columns are in categorical form so it is clear to apply encoding technique like label encoder/one hot encoding

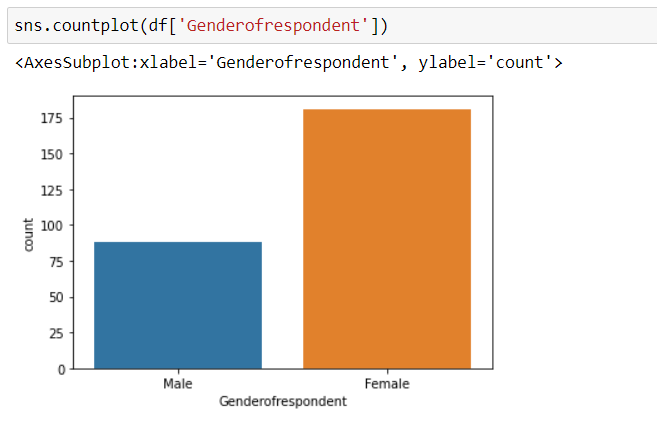
#### check the shape of datatypes

df.shape

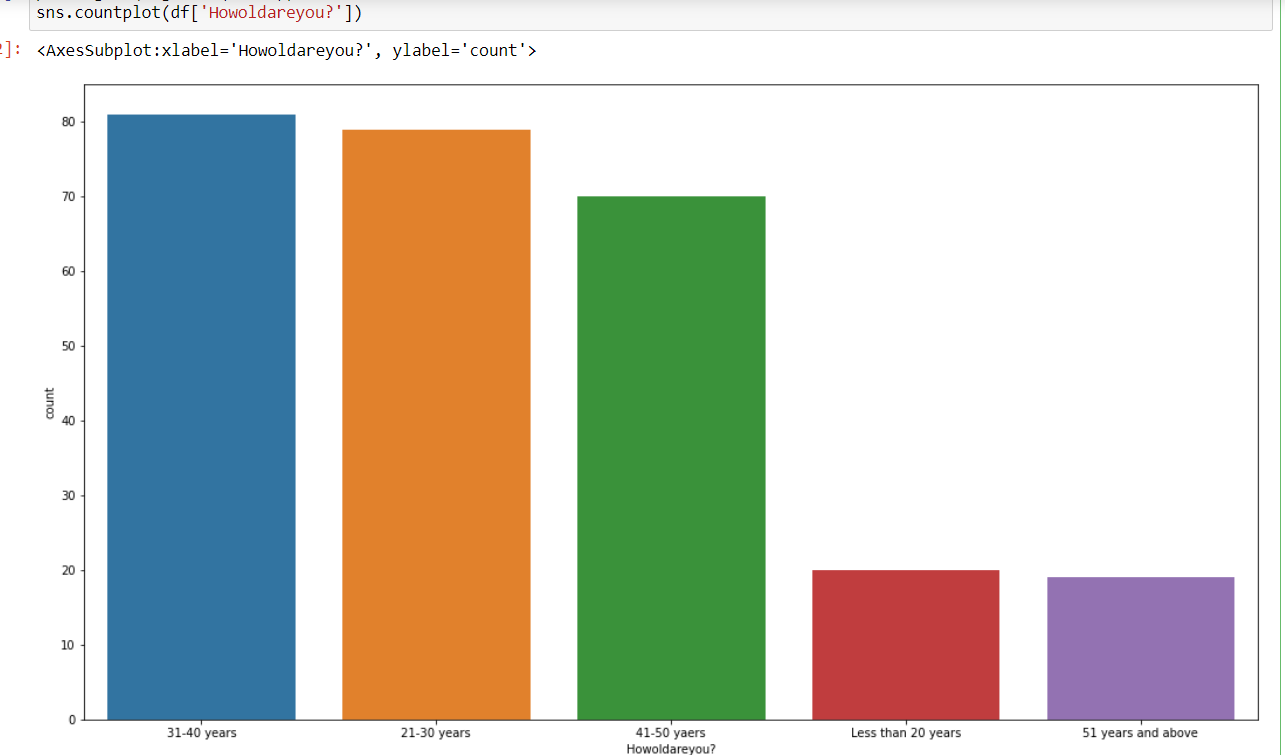


There are 269 Rows and 71 Columns

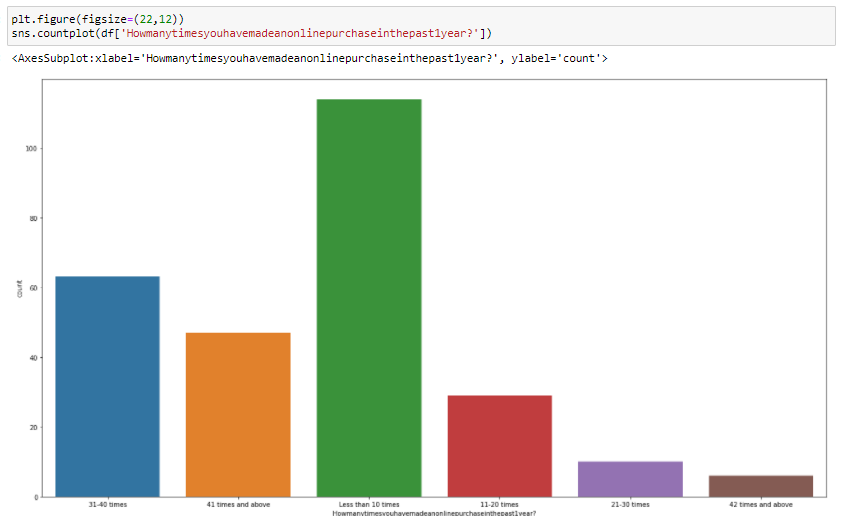
#### some Data visualization to understand the data set



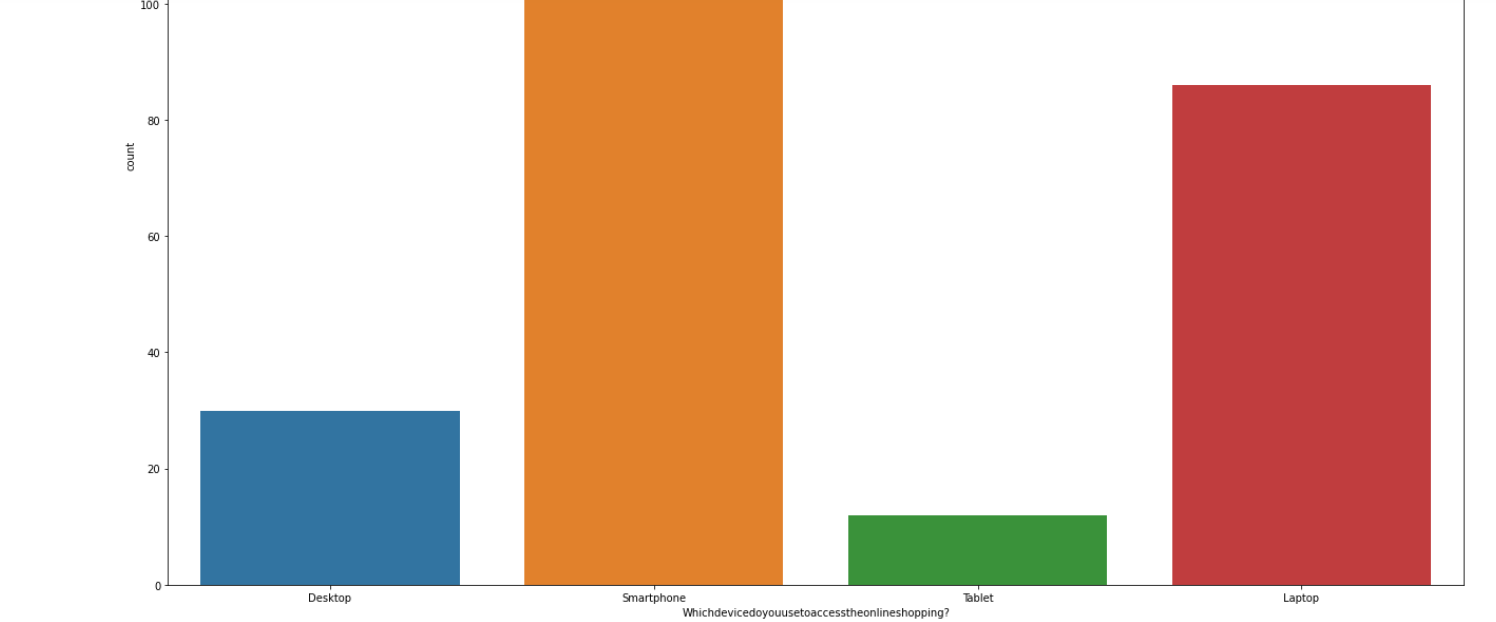
Female count is more than male in term of gender of respondent in e-commerce retailer



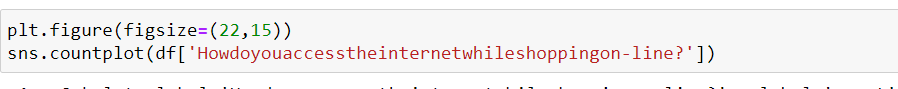
Mostly 31-40 years and 21-30 years use e-retailer shop more than others

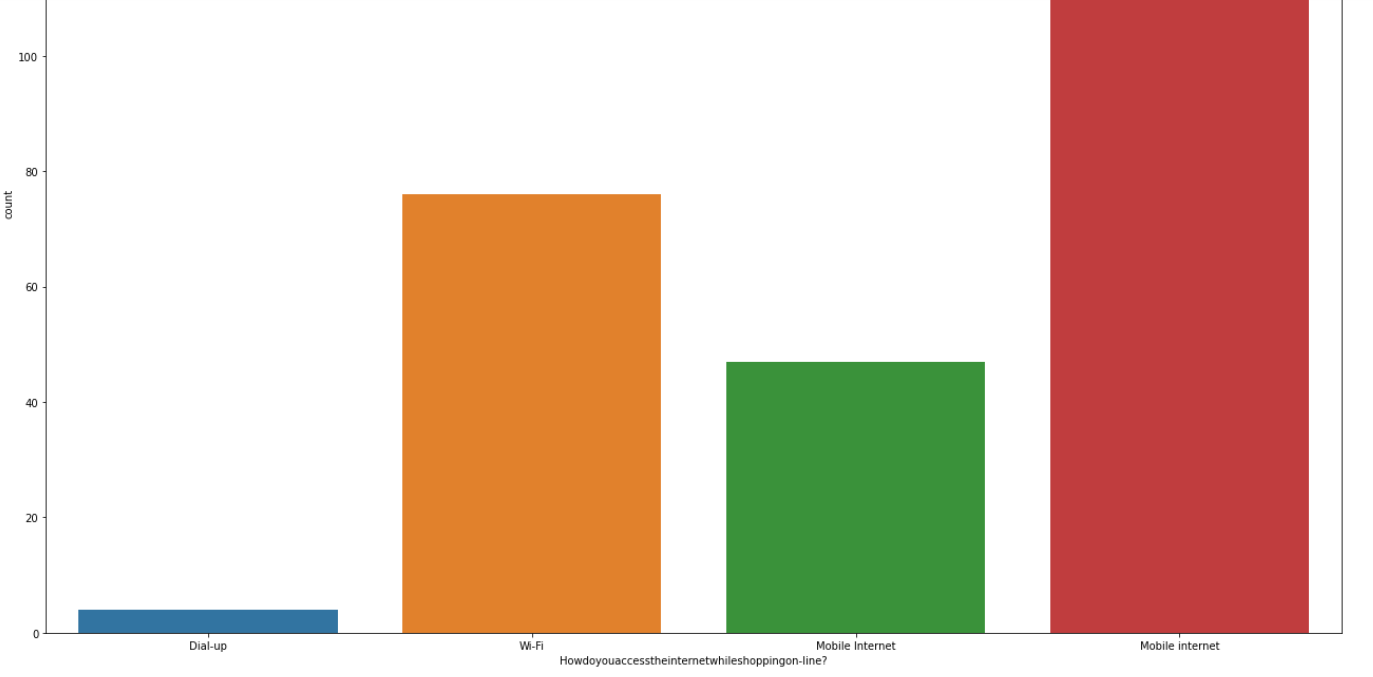


Less than 10 times people did purchase in past 1 years.it means it is not good for the e-commerce platform.

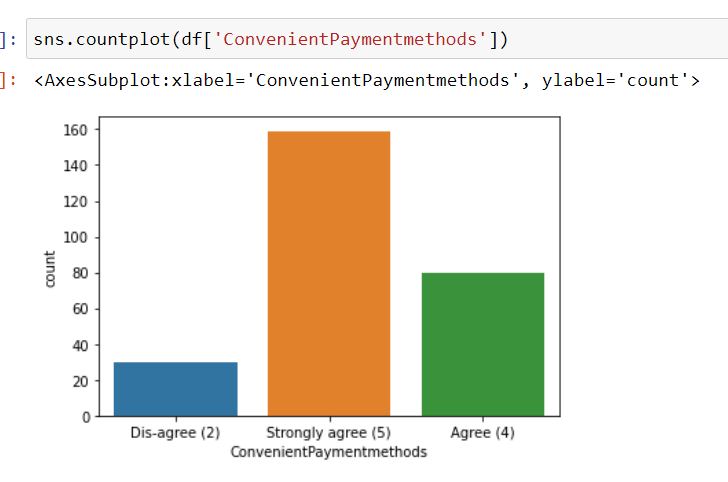


Mostly People use smartphone for oneline shopping.

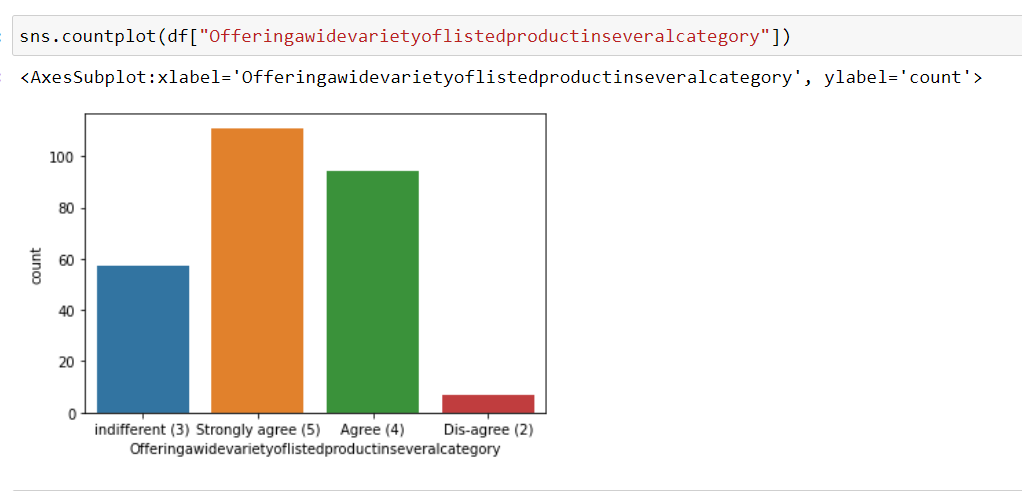




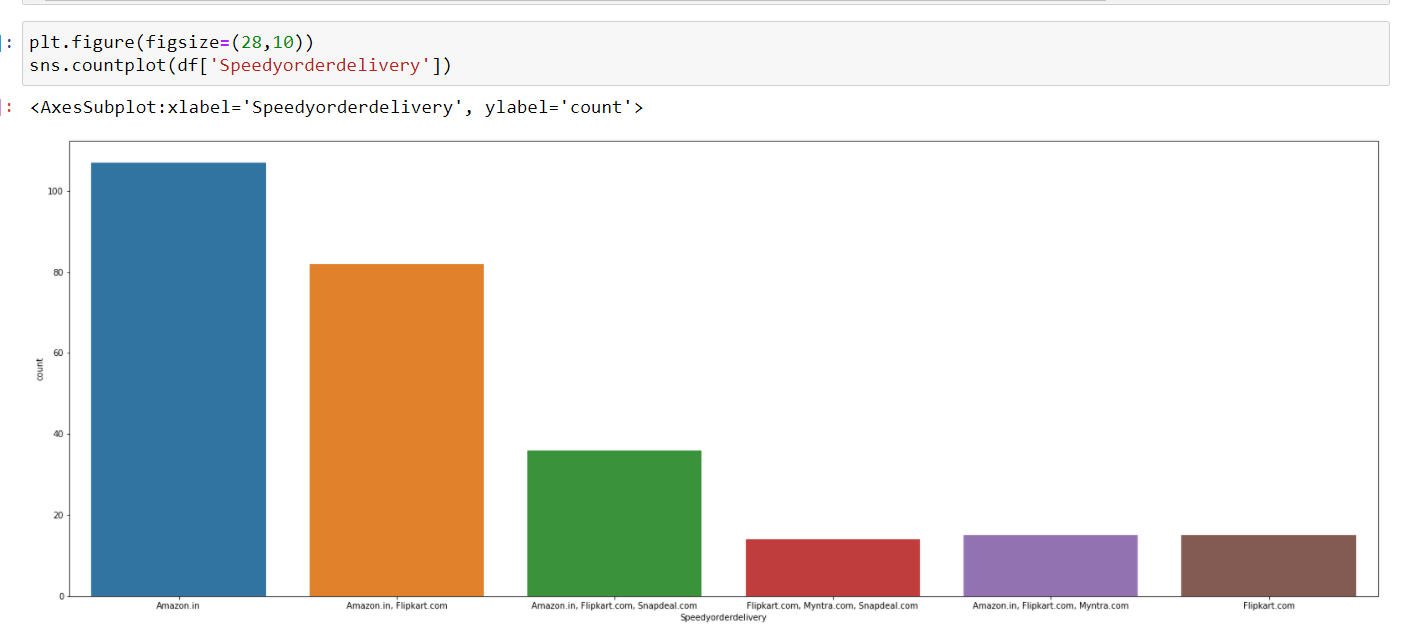
Mostly People use mobile internet for shopping.



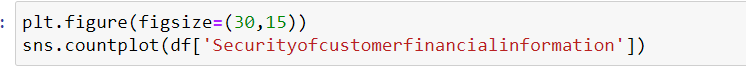
Continent Payment Method is Strongly agree for payment method.

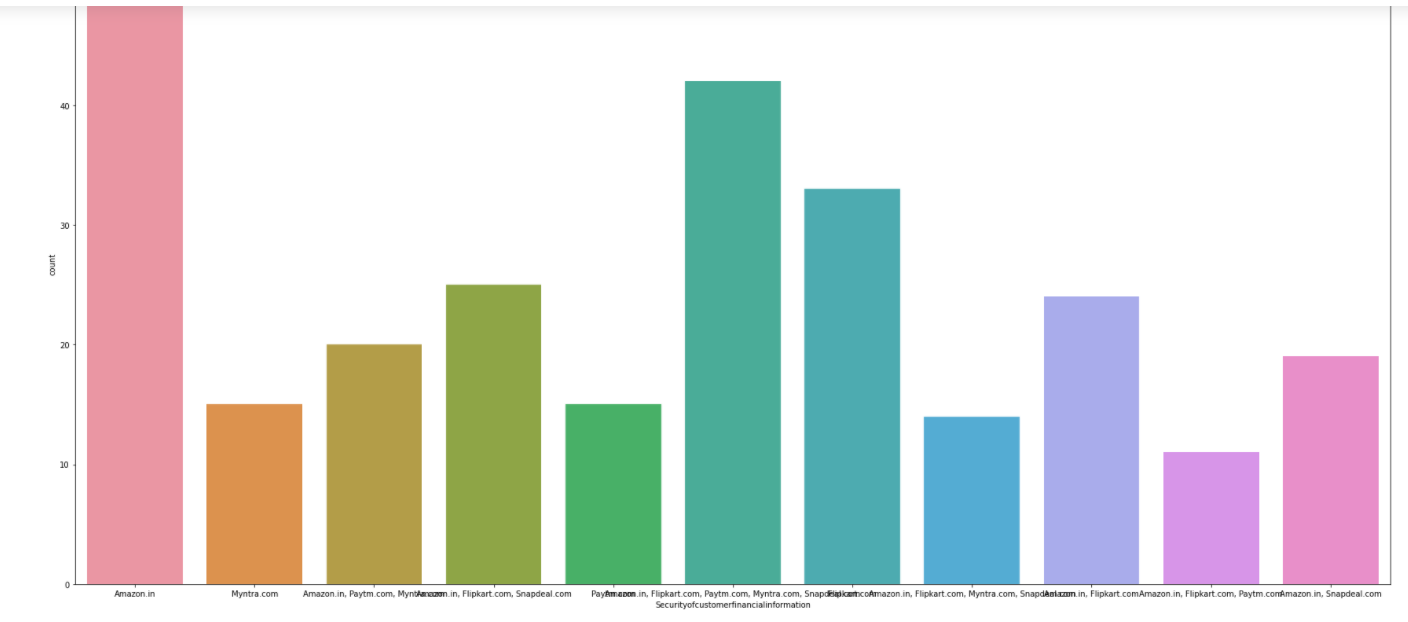


This graph shows that e-store provide wide variety of list of product in several category (Strongly agree)



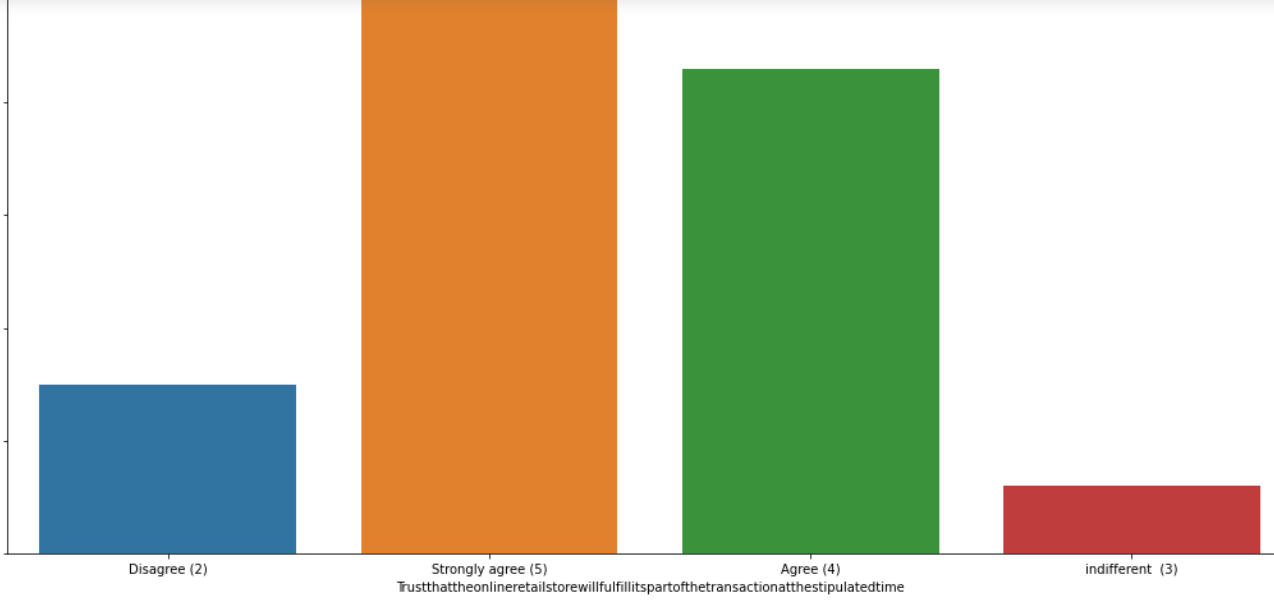
Mostly Amazon gives the speed of the delivery.



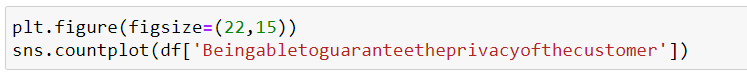


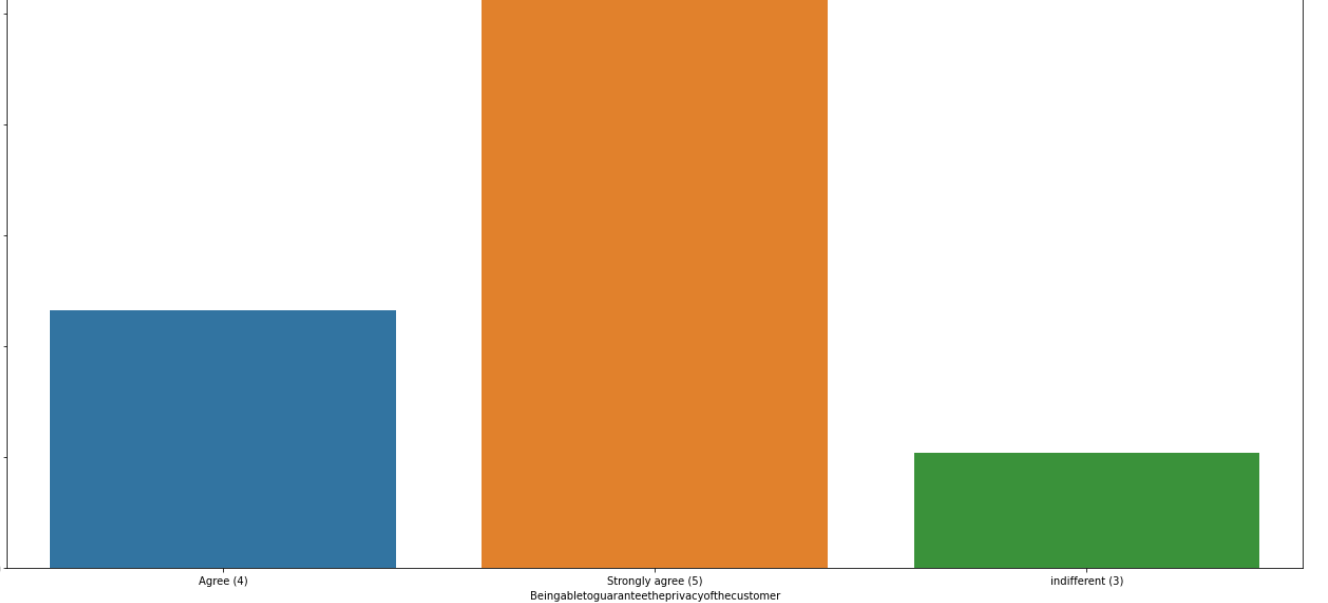
Amazon keeps the the security of customer financial information mostly shown in graph and its seems to good for the customer.



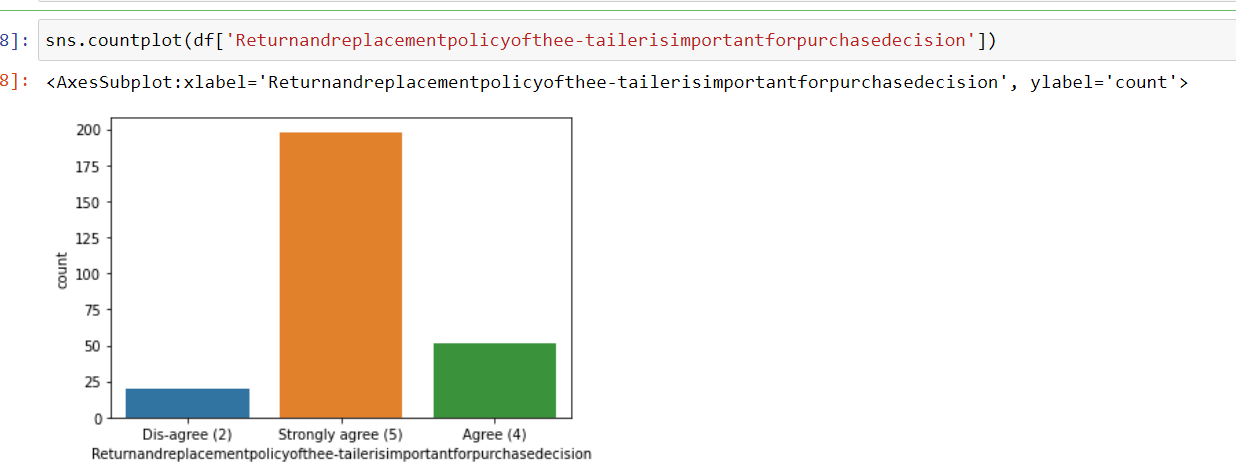


people stronly agree for the Trust that the online retail store will fulfill its part of the transaction at the stipulated time.

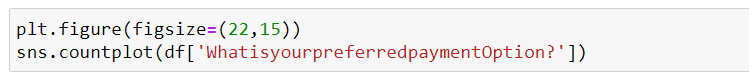


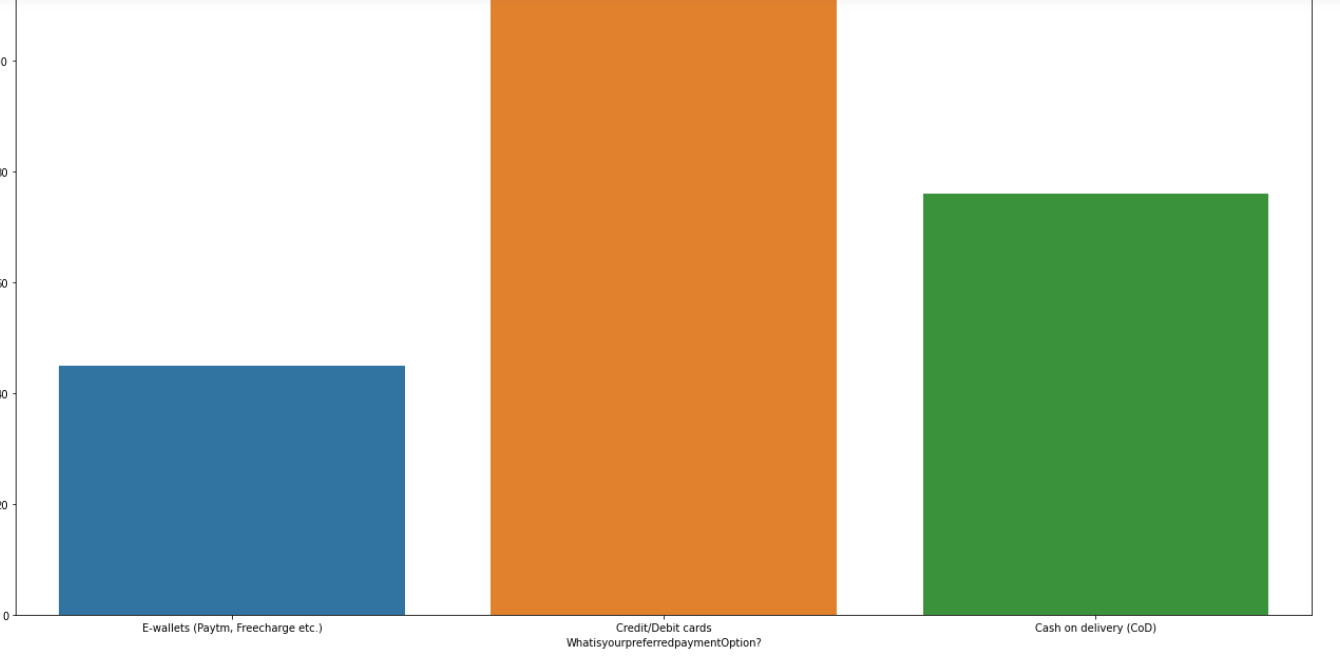


It Strongly agree for Being able to guarantee the privacy of the customer.

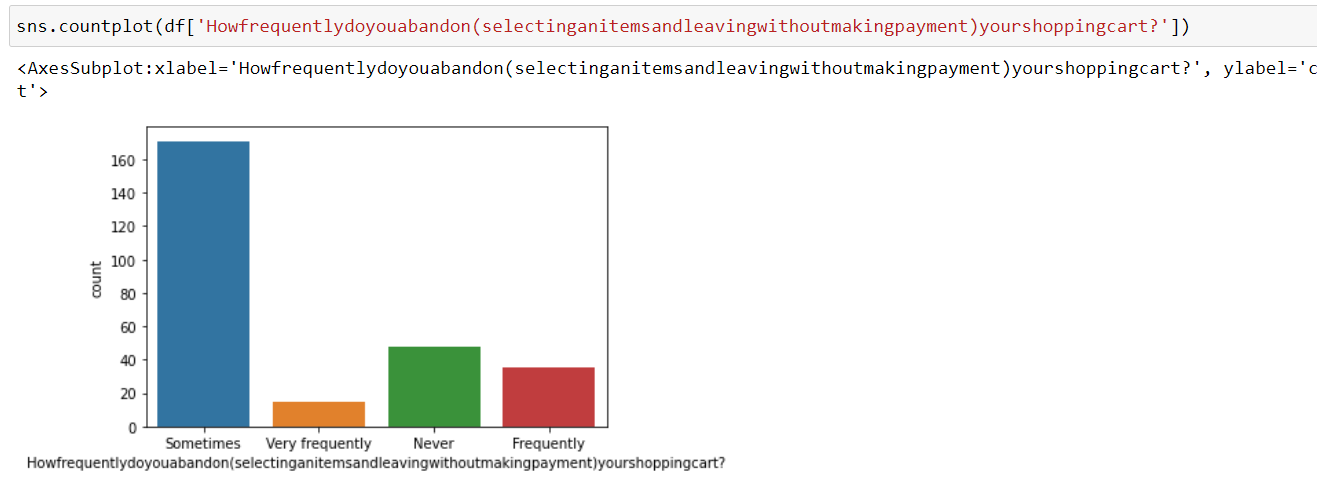


strongly agreed for the Return and Replacement policy is important for purchase decision is strongly

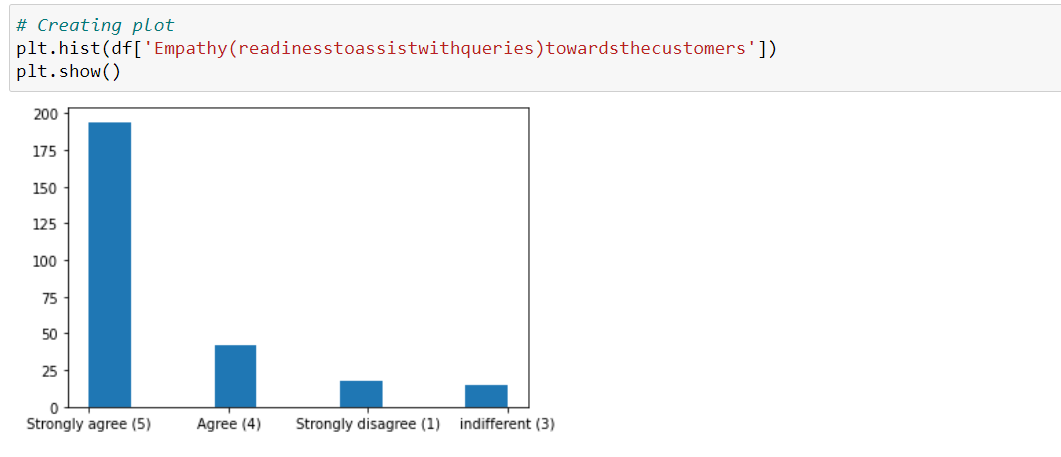




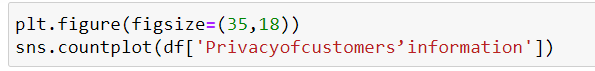
credit/debit card is perfered payment option choosen higly by the cutomers for online shopping.

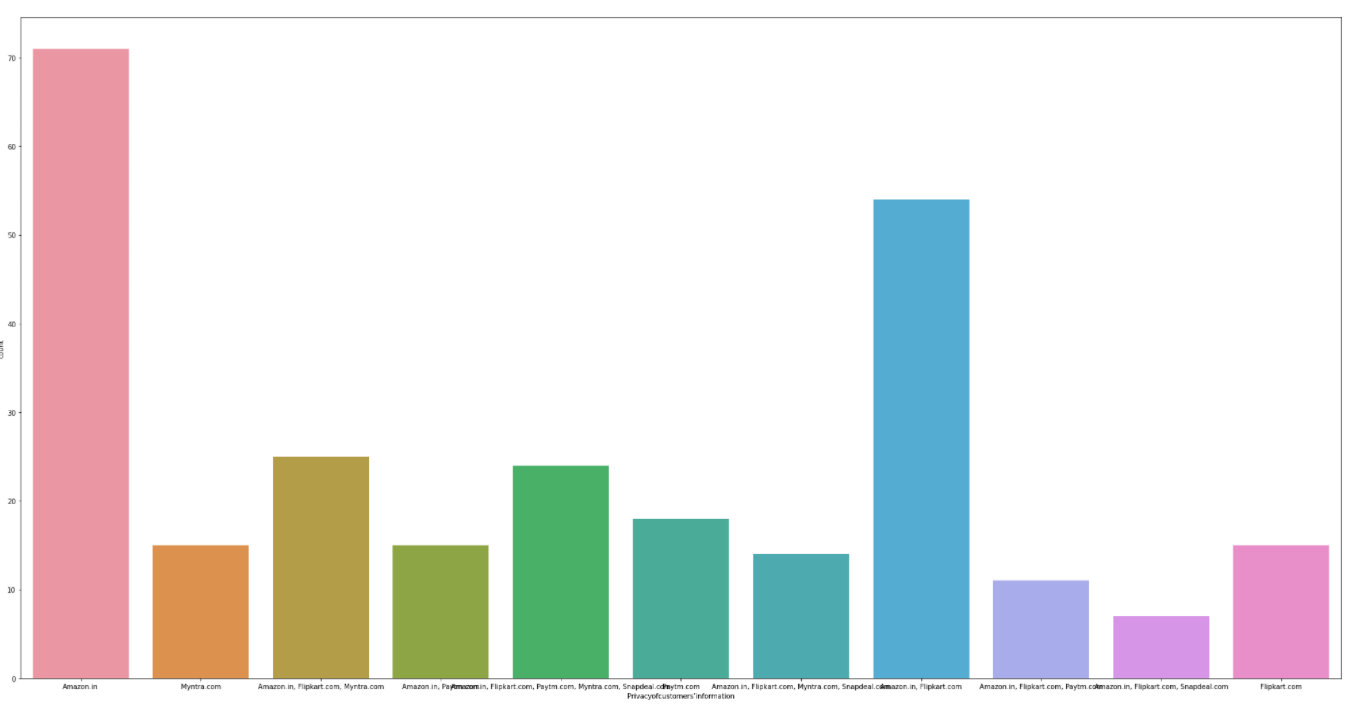


Sometimes is highly countable for selecting and leaving without making payment



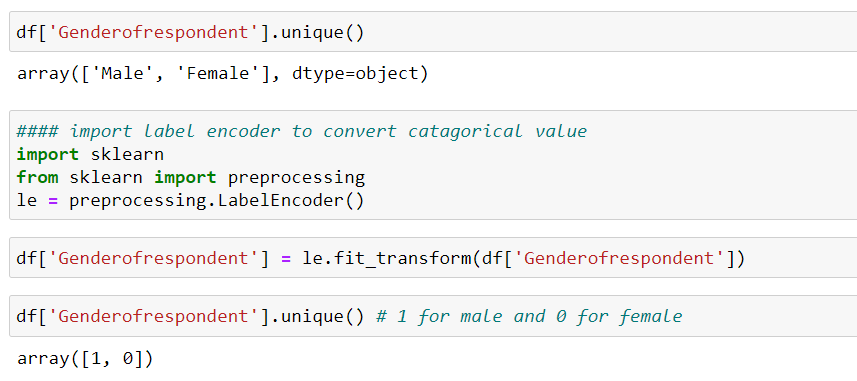
Online Strongly agreed towards Empathy readiness to assist with queries towards customer which helps more to bind the relationship.





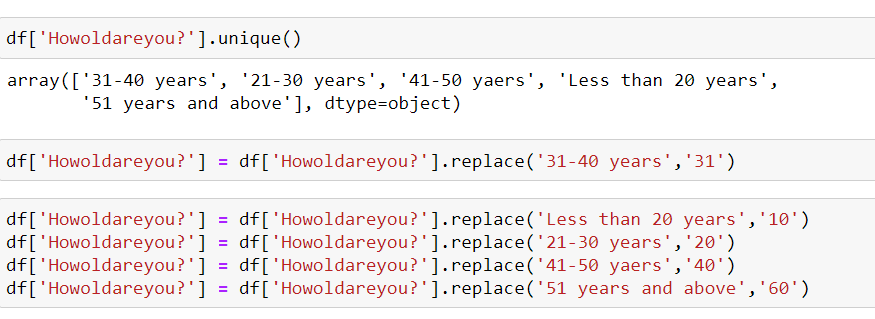
Amazon keep the customer policy information of customers

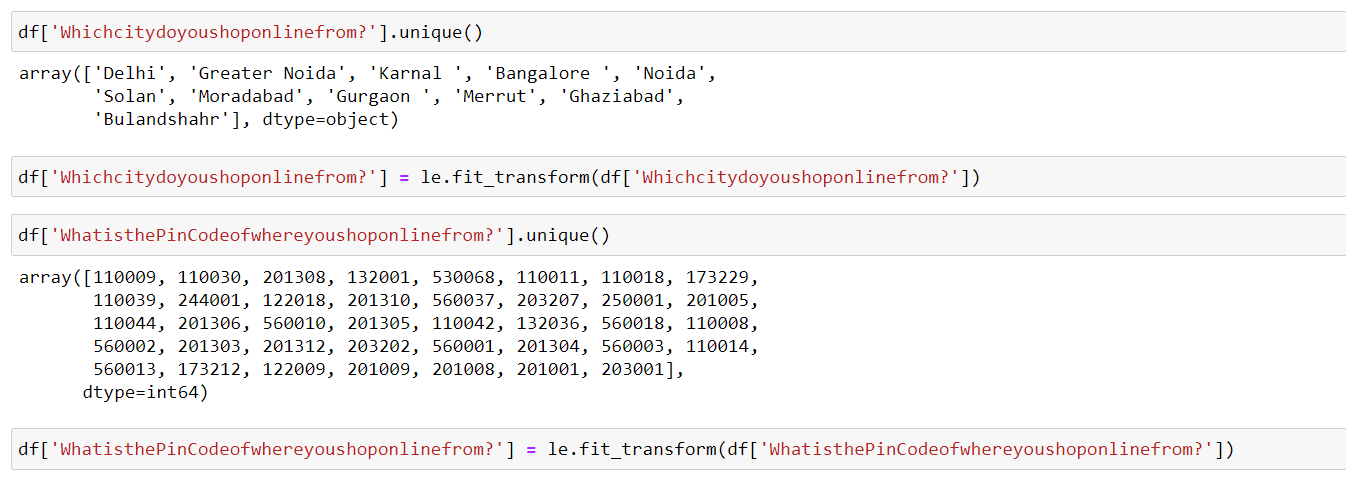
**Pre-processing**

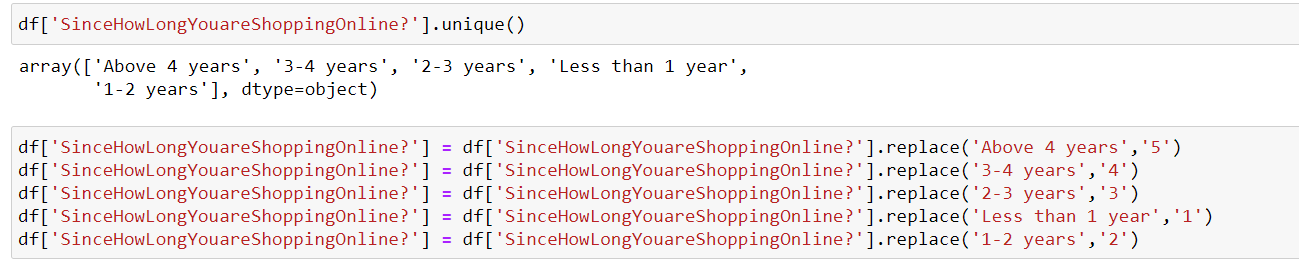


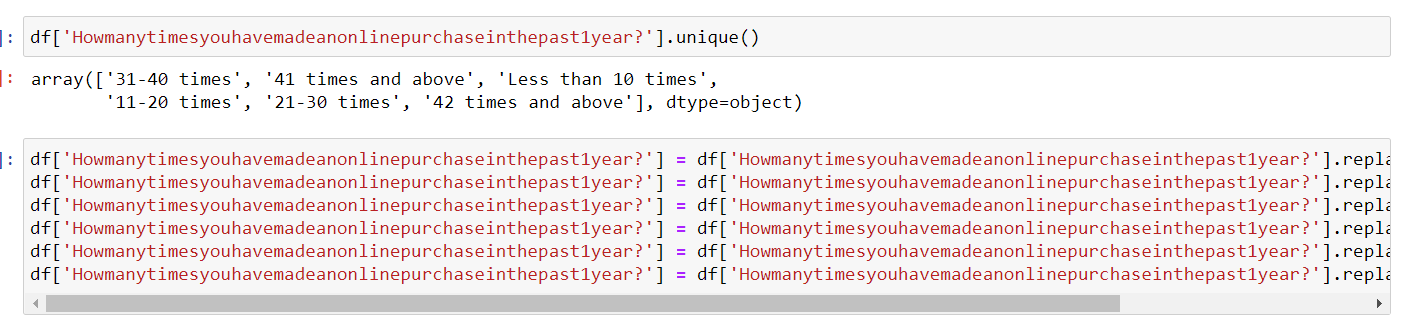
Gender of Respondent have two catagorical value which convert into 0 and 1.

## All columns are in categorical in behaviour so each column have been encoded with the help of encoder or use replace function to convert categorical value into numerical value.

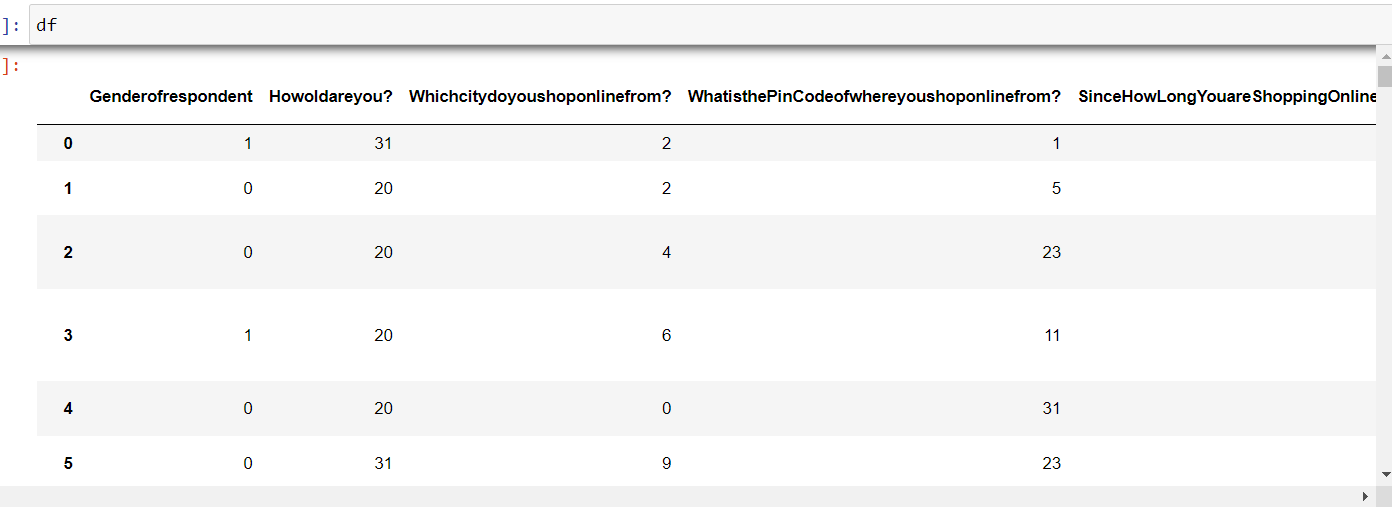








Check weather the columns are in categorical is converting into numerical or not



Its converting all columns in numerical so we will use same concept to convert the categorical fields.

df['Howdoyouaccesstheinternetwhileshoppingon-line?'] = le.fit\_transform(df['Howdoyouaccesstheinternetwhileshoppingon-line?'])

df['Whichdevicedoyouusetoaccesstheonlineshopping?'] = le.fit\_transform(df['Whichdevicedoyouusetoaccesstheonlineshopping?'])

df['Whatistheoperatingsystem(OS)ofyourdevice?'] =le.fit\_transform(df['Whatistheoperatingsystem(OS)ofyourdevice?'])

df['Whatbrowserdoyourunonyourdevicetoaccessthewebsite?'] = le.fit\_transform(df['Whatbrowserdoyourunonyourdevicetoaccessthewebsite?'])

df['Whichchanneldidyoufollowtoarriveatyourfavoriteonlinestoreforthefirsttime?'] = le.fit\_transform(df['Whichchanneldidyoufollowtoarriveatyourfavoriteonlinestoreforthefirsttime?'])

df['Afterfirstvisit,howdoyoureachtheonlineretailstore?'] = le.fit\_transform(df['Afterfirstvisit,howdoyoureachtheonlineretailstore?'])

df['Howfrequentlydoyouabandon(selectinganitemsandleavingwithoutmakingpayment)yourshoppingcart?'] = le.fit\_transform(df['Howfrequentlydoyouabandon(selectinganitemsandleavingwithoutmakingpayment)yourshoppingcart?'])

df['WhatisyourpreferredpaymentOption?'] = le.fit\_transform(df['WhatisyourpreferredpaymentOption?'])

df['Whydidyouabandonthe“Bag”,“ShoppingCart”?'] = le.fit\_transform(df['Whydidyouabandonthe“Bag”,“ShoppingCart”?'])

df['Thecontentonthewebsitemustbeeasytoreadandunderstand'] = le.fit\_transform(df['Thecontentonthewebsitemustbeeasytoreadandunderstand'])

df['Informationonsimilarproducttotheonehighlightedisimportantforproductcomparison'] = le.fit\_transform(df['Informationonsimilarproducttotheonehighlightedisimportantforproductcomparison'])

df['Completeinformationonlistedsellerandproductbeingofferedisimportantforpurchasedecision.'] = le.fit\_transform(df['Completeinformationonlistedsellerandproductbeingofferedisimportantforpurchasedecision.'])

Same code in image to show its work to convert



df['Whydidyouabandonthe“Bag”,“ShoppingCart”?'] = le.fit\_transform(df['Whydidyouabandonthe“Bag”,“ShoppingCart”?'])

df['Thecontentonthewebsitemustbeeasytoreadandunderstand'] = le.fit\_transform(df['Thecontentonthewebsitemustbeeasytoreadandunderstand'])

df['Informationonsimilarproducttotheonehighlightedisimportantforproductcomparison'] = le.fit\_transform(df['Informationonsimilarproducttotheonehighlightedisimportantforproductcomparison'])

df['Completeinformationonlistedsellerandproductbeingofferedisimportantforpurchasedecision.'] = le.fit\_transform(df['Completeinformationonlistedsellerandproductbeingofferedisimportantforpurchasedecision.'])

df['Allrelevantinformationonlistedproductsmustbestatedclearly'] = le.fit\_transform(df['Allrelevantinformationonlistedproductsmustbestatedclearly'])

df['Easeofnavigationinwebsite'] = le.fit\_transform(df['Easeofnavigationinwebsite'])

df['Loadingandprocessingspeed'] = le.fit\_transform(df['Loadingandprocessingspeed'])

df['UserfriendlyInterfaceofthewebsite'] = le.fit\_transform(df['UserfriendlyInterfaceofthewebsite'])

df['ConvenientPaymentmethods'] = le.fit\_transform(df['ConvenientPaymentmethods'])

df['Beingabletoguaranteetheprivacyofthecustomer'] = le.fit\_transform(df['Beingabletoguaranteetheprivacyofthecustomer'])

df['Trustthattheonlineretailstorewillfulfillitspartofthetransactionatthestipulatedtime'] = le.fit\_transform(df['Trustthattheonlineretailstorewillfulfillitspartofthetransactionatthestipulatedtime'])

df['Empathy(readinesstoassistwithqueries)towardsthecustomers'] = le.fit\_transform(df['Empathy(readinesstoassistwithqueries)towardsthecustomers'])

df['Responsiveness,availabilityofseveralcommunicationchannels(email,onlinerep,twitter,phoneetc.)'] = le.fit\_transform(df['Responsiveness,availabilityofseveralcommunicationchannels(email,onlinerep,twitter,phoneetc.)'])

df['Onlineshoppinggivesmonetarybenefitanddiscounts'] = le.fit\_transform(df['Onlineshoppinggivesmonetarybenefitanddiscounts'])

df['Enjoymentisderivedfromshoppingonline'] = le.fit\_transform(df['Enjoymentisderivedfromshoppingonline'])

df['Shoppingonlineisconvenientandflexible'] = le.fit\_transform(df['Shoppingonlineisconvenientandflexible'])

df['Returnandreplacementpolicyofthee-tailerisimportantforpurchasedecision'] = le.fit\_transform(df['Returnandreplacementpolicyofthee-tailerisimportantforpurchasedecision'])

df['Gainingaccesstoloyaltyprogramsisabenefitofshoppingonline'] = le.fit\_transform(df['Gainingaccesstoloyaltyprogramsisabenefitofshoppingonline'])



df['DisplayingqualityInformationonthewebsiteimprovessatisfactionofcustomers'] = le.fit\_transform(df['DisplayingqualityInformationonthewebsiteimprovessatisfactionofcustomers'])

df['Userderivesatisfactionwhileshoppingonagoodqualitywebsiteorapplication'] = le.fit\_transform(df['Userderivesatisfactionwhileshoppingonagoodqualitywebsiteorapplication'])

df['Usersatisfactioncannotexistwithouttrust'] = le.fit\_transform(df['Usersatisfactioncannotexistwithouttrust'])

df['Offeringawidevarietyoflistedproductinseveralcategory'] = le.fit\_transform(df['Offeringawidevarietyoflistedproductinseveralcategory'])

df['Provisionofcompleteandrelevantproductinformation'] = le.fit\_transform(df['Provisionofcompleteandrelevantproductinformation'])

df['Monetarysavings'] = le.fit\_transform(df['Monetarysavings'])

df['TheConvenienceofpatronizingtheonlineretailer'] = le.fit\_transform(df['TheConvenienceofpatronizingtheonlineretailer'])

df['Shoppingonthewebsitegivesyouthesenseofadventure'] = le.fit\_transform(df['Shoppingonthewebsitegivesyouthesenseofadventure'])

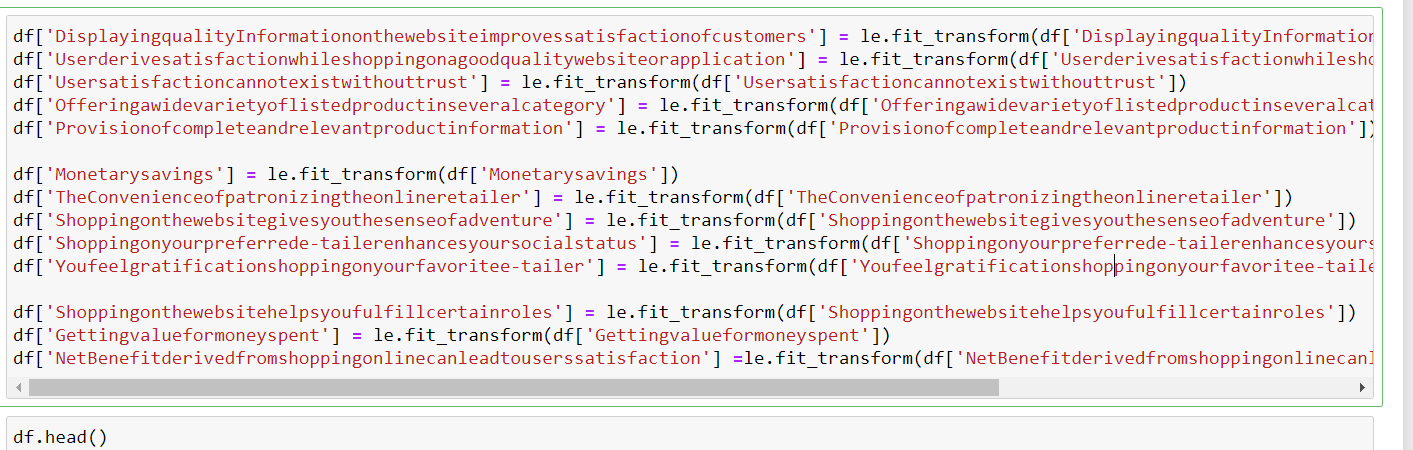
df['Shoppingonyourpreferrede-tailerenhancesyoursocialstatus'] = le.fit\_transform(df['Shoppingonyourpreferrede-tailerenhancesyoursocialstatus'])

df['Youfeelgratificationshoppingonyourfavoritee-tailer'] = le.fit\_transform(df['Youfeelgratificationshoppingonyourfavoritee-tailer'])

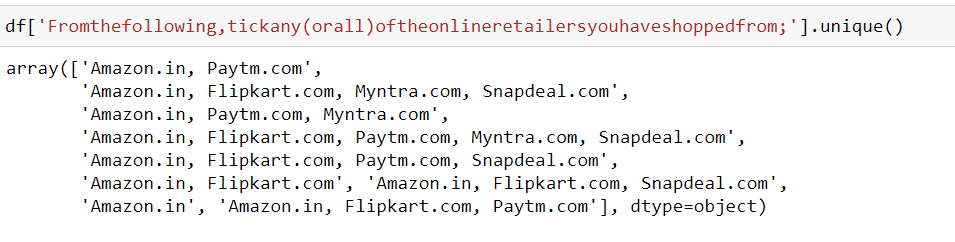
df['Shoppingonthewebsitehelpsyoufulfillcertainroles'] = le.fit\_transform(df['Shoppingonthewebsitehelpsyoufulfillcertainroles'])

df['Gettingvalueformoneyspent'] = le.fit\_transform(df['Gettingvalueformoneyspent'])

df['NetBenefitderivedfromshoppingonlinecanleadtouserssatisfaction'] =le.fit\_transform(df['NetBenefitderivedfromshoppingonlinecanleadtouserssatisfaction'])



### head function show top 5 rows from the dataset



Now split the comma and store all values in separate columns , it will be good to predict.

So lets split and separate in new one.

df['Fromthefollowing,tickany(orall)oftheonlineretailersyouhaveshoppedfrom1'] = df['Fromthefollowing,tickany(orall)oftheonlineretailersyouhaveshoppedfrom;'].str.split(',').str[1]

df['Fromthefollowing,tickany(orall)oftheonlineretailersyouhaveshoppedfrom2'] = df['Fromthefollowing,tickany(orall)oftheonlineretailersyouhaveshoppedfrom;'].str.split(',').str[2]

df['Fromthefollowing,tickany(orall)oftheonlineretailersyouhaveshoppedfrom3'] = df['Fromthefollowing,tickany(orall)oftheonlineretailersyouhaveshoppedfrom;'].str.split(',').str[3]

df['Fromthefollowing,tickany(orall)oftheonlineretailersyouhaveshoppedfrom4'] = df['Fromthefollowing,tickany(orall)oftheonlineretailersyouhaveshoppedfrom;'].str.split(',').str[4]

df['Fromthefollowing,tickany(orall)oftheonlineretailersyouhaveshoppedfrom5'] = df['Fromthefollowing,tickany(orall)oftheonlineretailersyouhaveshoppedfrom;'].str.split(',').str[5]

df.drop('Fromthefollowing,tickany(orall)oftheonlineretailersyouhaveshoppedfrom;',inplace=True,axis=1)

df['Easytousewebsiteorapplication1']= df['Easytousewebsiteorapplication'].str.split(',').str[1]

df['Easytousewebsiteorapplication2']= df['Easytousewebsiteorapplication'].str.split(',').str[2]

df['Easytousewebsiteorapplication3']= df['Easytousewebsiteorapplication'].str.split(',').str[3]

df['Easytousewebsiteorapplication4']= df['Easytousewebsiteorapplication'].str.split(',').str[4]

df['Easytousewebsiteorapplication5']= df['Easytousewebsiteorapplication'].str.split(',').str[5]

df.drop('Easytousewebsiteorapplication',inplace=True,axis=1)

#### separate with comma and store in new columns separately

df['Complete,relevantdescriptioninformationofproducts1']= df['Complete,relevantdescriptioninformationofproducts'].str.split(',').str[1]

df['Complete,relevantdescriptioninformationofproducts2']= df['Complete,relevantdescriptioninformationofproducts'].str.split(',').str[2]

df['Complete,relevantdescriptioninformationofproducts3']= df['Complete,relevantdescriptioninformationofproducts'].str.split(',').str[3]

df['Complete,relevantdescriptioninformationofproducts4']= df['Complete,relevantdescriptioninformationofproducts'].str.split(',').str[4]

df['Complete,relevantdescriptioninformationofproducts5']= df['Complete,relevantdescriptioninformationofproducts'].str.split(',').str[5]

#### separate with comma and store in new columns separately

df['Fastloadingwebsitespeedofwebsiteandapplication1']= df['Fastloadingwebsitespeedofwebsiteandapplication'].str.split(',').str[1]

df['Fastloadingwebsitespeedofwebsiteandapplication2']= df['Fastloadingwebsitespeedofwebsiteandapplication'].str.split(',').str[2]

df['Fastloadingwebsitespeedofwebsiteandapplication3']= df['Fastloadingwebsitespeedofwebsiteandapplication'].str.split(',').str[3]

df['Fastloadingwebsitespeedofwebsiteandapplication4']= df['Fastloadingwebsitespeedofwebsiteandapplication'].str.split(',').str[4]

df['Fastloadingwebsitespeedofwebsiteandapplication5']= df['Fastloadingwebsitespeedofwebsiteandapplication'].str.split(',').str[5]

#### separet ecah value from comma and store ine variable

#### separate with comma and store in new columns separately

df['Reliabilityofthewebsiteorapplication1']= df['Reliabilityofthewebsiteorapplication'].str.split(',').str[1]

df['Reliabilityofthewebsiteorapplication2']= df['Reliabilityofthewebsiteorapplication'].str.split(',').str[2]

df['Reliabilityofthewebsiteorapplication3']= df['Reliabilityofthewebsiteorapplication'].str.split(',').str[3]

df['Reliabilityofthewebsiteorapplication4']= df['Reliabilityofthewebsiteorapplication'].str.split(',').str[4]

df['Reliabilityofthewebsiteorapplication5']= df['Reliabilityofthewebsiteorapplication'].str.split(',').str[5]

#### separet ecah value from comma and store ine variable

df['Quicknesstocompletepurchase1']= df['Quicknesstocompletepurchase'].str.split(',').str[1]

df['Quicknesstocompletepurchase2']= df['Quicknesstocompletepurchase'].str.split(',').str[2]

df['Quicknesstocompletepurchase3']= df['Quicknesstocompletepurchase'].str.split(',').str[3]

df['Quicknesstocompletepurchase4']= df['Quicknesstocompletepurchase'].str.split(',').str[4]

df['Quicknesstocompletepurchase5']= df['Quicknesstocompletepurchase'].str.split(',').str[5]

#### separet ecah value from comma and store ine variable

df['Availabilityofseveralpaymentoptions1']= df['Availabilityofseveralpaymentoptions'].str.split(',').str[1]

df['Availabilityofseveralpaymentoptions2']= df['Availabilityofseveralpaymentoptions'].str.split(',').str[2]

df['Availabilityofseveralpaymentoptions3']= df['Availabilityofseveralpaymentoptions'].str.split(',').str[3]

df['Availabilityofseveralpaymentoptions4']= df['Availabilityofseveralpaymentoptions'].str.split(',').str[4]

df['Availabilityofseveralpaymentoptions5']= df['Availabilityofseveralpaymentoptions'].str.split(',').str[5]

#### separet ecah value from comma and store ine variable

df['Speedyorderdelivery1']= df['Speedyorderdelivery'].str.split(',').str[1]

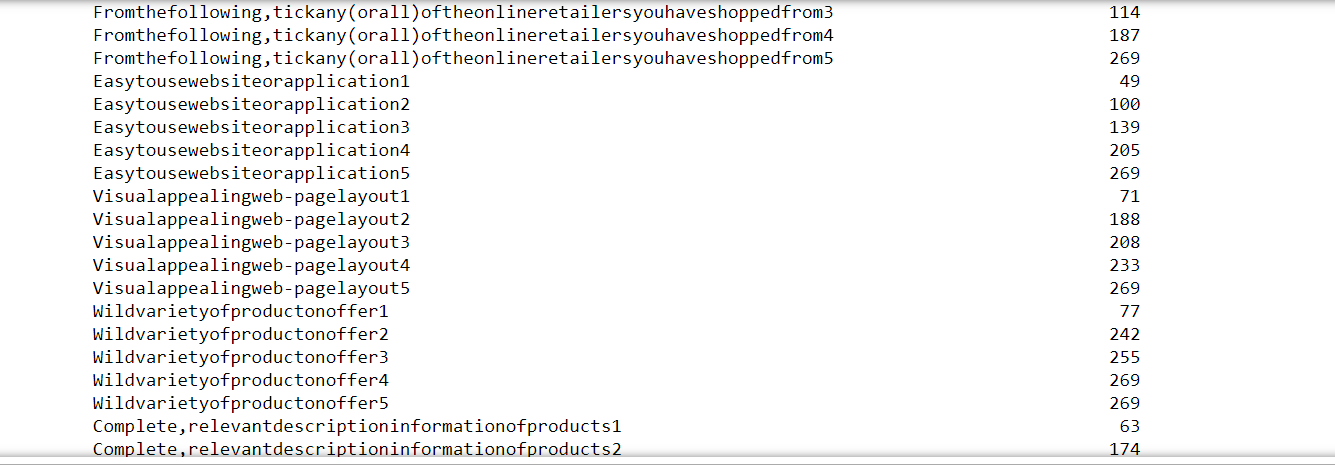
df['Speedyorderdelivery2']= df['Speedyorderdelivery'].str.split(',').str[2]

df['Speedyorderdelivery3']= df['Speedyorderdelivery'].str.split(',').str[3]

df['Speedyorderdelivery4']= df['Speedyorderdelivery'].str.split(',').str[4]

df['Speedyorderdelivery5']= df['Speedyorderdelivery'].str.split(',').str[5]

### now check for the null and handle the missing value.

df.isnull().sum()

### drop the mostly count missing value because its will may overfit the model



#### drop the unnecessary columns which no mean

df.drop(['Visualappealingweb-pagelayout5','Visualappealingweb-pagelayout4','Wildvarietyofproductonoffer3','Wildvarietyofproductonoffer4','Wildvarietyofproductonoffer5'],inplace=True,axis=1)

df.drop(['Fastloadingwebsitespeedofwebsiteandapplication5','Reliabilityofthewebsiteorapplication4','Reliabilityofthewebsiteorapplication5','Quicknesstocompletepurchase3','Quicknesstocompletepurchase4','Quicknesstocompletepurchase5','Availabilityofseveralpaymentoptions5'],inplace=True,axis=1)

df.drop(['Securityofcustomerfinancialinformation5','PerceivedTrustworthiness3','PerceivedTrustworthiness5','Presenceofonlineassistancethroughmulti-channel4','Presenceofonlineassistancethroughmulti-channel5','Longertimetogetloggedin(promotion,salesperiod)3','Longertimetogetloggedin(promotion,salesperiod)4','Longertimetogetloggedin(promotion,salesperiod)5'],inplace=True,axis=1)

df.drop(['Latedeclarationofprice(promotion,salesperiod)2','Longerpageloadingtime(promotion,salesperiod)2','Longerdeliveryperiod2','Frequentdisruptionwhenmovingfromonepagetoanother2'],inplace=True,axis=1)

df.drop(['WhichoftheIndianonlineretailerwouldyourecommendtoafriend?3','WhichoftheIndianonlineretailerwouldyourecommendtoafriend?4','WhichoftheIndianonlineretailerwouldyourecommendtoafriend?5'],inplace=True,axis=1)

df.drop(['Limitedmodeofpaymentonmostproducts(promotion,salesperiod)2','Longerpageloadingtime(promotion,salesperiod)5','Longerpageloadingtime(promotion,salesperiod)4','Longertimeindisplayinggraphicsandphotos(promotion,salesperiod)2'],axis=1,inplace=True)

df.drop(['Limitedmodeofpaymentonmostproducts(promotion,salesperiod)2','Longerpageloadingtime(promotion,salesperiod)5','Longerpageloadingtime(promotion,salesperiod)4','Longertimeindisplayinggraphicsandphotos(promotion,salesperiod)2'],axis=1,inplace=True)

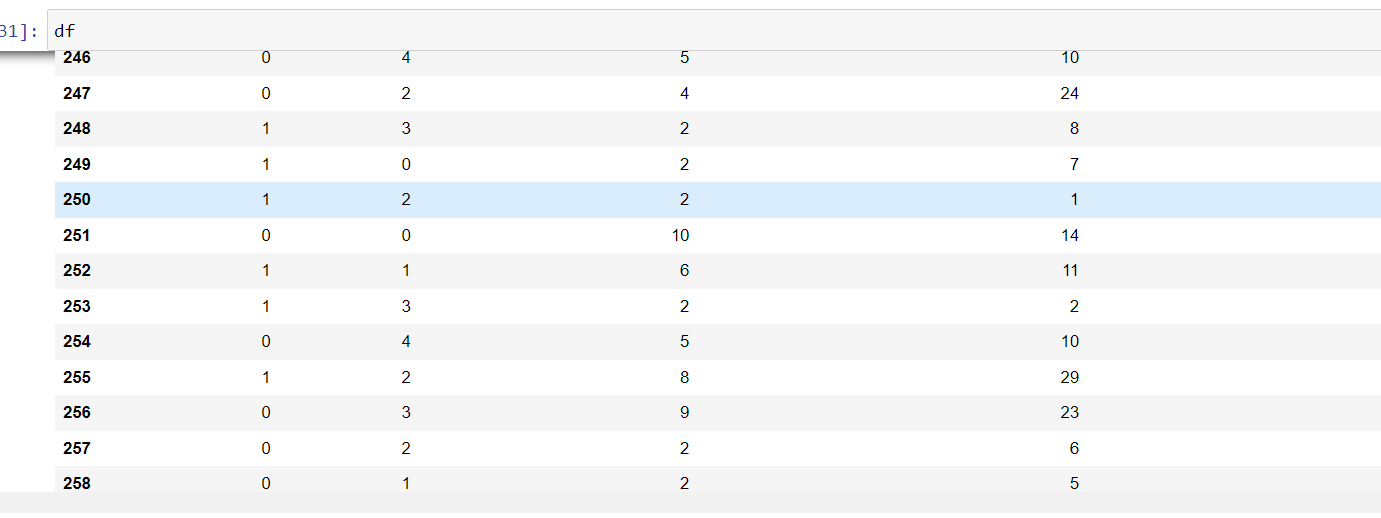
df.drop(['Longerpageloadingtime(promotion,salesperiod)3','Longertimeindisplayinggraphicsandphotos(promotion,salesperiod)5','Longertimeindisplayinggraphicsandphotos(promotion,salesperiod)3','Longertimeindisplayinggraphicsandphotos(promotion,salesperiod)4'],inplace=True,axis=1)

df.drop(['Websiteisasefficientasbefore3','Speedyorderdelivery3','Speedyorderdelivery4','Speedyorderdelivery5','Complete,relevantdescriptioninformationofproducts4','Fromthefollowing,tickany(orall)oftheonlineretailersyouhaveshoppedfrom5'],axis=1,inplace=True)

df.drop(['Easytousewebsiteorapplication5','Complete,relevantdescriptioninformationofproducts5','Privacyofcustomers’information5','PerceivedTrustworthiness4'],axis=1,inplace=True)

df.drop(['Securityofcustomerfinancialinformation1','WhichoftheIndianonlineretailerwouldyourecommendtoafriend?1','Changeinwebsite/Applicationdesign1','Reliabilityofthewebsiteorapplication3'],inplace=True,axis=1)

### in rest of missing value I have filled with None Value. Because most of the columns do not have the features to like fast delivery /security of customer information/easy to use website/long page reloading time most people not face this issue its means none so I have replaced all missing value with None vale and applied Label encoder to convert all categorical value or feature column into numeric now data is ready for modal.



Conclusion:

Customer aspect various things from the online store like privacy, vide variety of product, payment Efficiency, fast delivery orders, mostly used payment method many more. If customer gets the satisfaction, it automatically builds the relationship. From visualization customer agree with payment option , fast delivery/speedy delivery ,privacy of customer information and privacy of account/financial statement. I have visualized the scenario to understand what customer need during online shopping so any one feasible to keep relationship .After Visualization clean the data remove missing value and used scaling techniques for building the best model.