

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 Libraries and Load the data of Customer Retention

```
In [547]: import pandas as pd
import seaborn as sns
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
```

```
In [548]: Customer_Retention_df = pd.read_csv('customer_retention.csv')
```

```
In [549]: Customer_Retention_df
```

Out[549]:

Gender of respondent	How old are you?	Which city do you shop online from?	What is the Pin Code of where you shop online from?	Since How Long You are Shopping Online ?	How many times you have made an online purchase in the past 1 year?	How do you access the internet while shopping on-line?	Which device do you use to access the online shopping?	What is the screen size of your mobile device?	What is the operating system (OS) of your device?
24-40	Above 40	24-40	24-40	24-40	24-40	24-40	24-40	24-40	24-40

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

```
Customer_Retention_df.columns
```

show the columns present in data set


```

Genderofrespondent 0
Howoldareyou? 0
Whichcitydoyoushoponlinefrom? 0
WhatisthePinCodeofwhereyoushoponlinefrom? 0
SinceHowLongYouareShoppingOnline? 0
Howmanytimesyouhavemadeanonlinepurchaseinthepast1year? 0
Howdoyouaccesstheinternetwhileshoppingon-line? 0
Whichdevicedoyouusetooaccesstheonlineshopping? 0
Whatisthescreensizeofyourmobiledevice? 0
Whatistheoperatingsystem(OS)ofyourdevice? 0
Whatbrowserdoyourunonyourdevicetoaccessthewebsite? 0
Whichchanneldidyoufollowtoarriveatyourfavoriteonlinestoreforthefirsttime? 0
Afterfirstvisit,howdoyoureachtheonlinetailstore? 0
Howmuchtimedoyouexplorethee-retailstorebeforemakingapurchasedecision? 0
WhatisyourpreferredpaymentOption? 0
Howfrequentlydoyouabandon(selectinganitemsandleavingwithoutmakingpayment)yourshoppingcart? 0
Whydidyouabandonthe"Bag","ShoppingCart"? 0
Thecontentonthewebsitemustbeeasytoreadandunderstand 0
Informationonsimilarproducttotheonehighlightedisimportantforproductcomparison 0
Completeinformationonlistedsellerandproductbeingofferedisimportantforpurchasedecision. 0
Allrelevantinformationonlistedproductsmustbestatedclearly 0

```

Statistical description show only numeric

```
value Customer_Retention_df.describe()
```

WhatisthePinCodeofwhereyoushoponlinefrom?	
count	269.000000
mean	220465.747212
std	140524.341051
min	110008.000000
25%	122018.000000
50%	201303.000000
75%	201310.000000
max	560037.000000

statistical data type description

```
Customer_Retention_df.dtypes
```

Genderofrespondent	object
Howoldareyou?	object
Whichcitydoyoushoponlinefrom?	object
WhatisthePinCodeofwhereyoushoponlinefrom?	int64
SinceHowLongYouareShoppingOnline?	object
Howmanytimesyouhavemadeanonlinepurchaseinthepast1year?	object
Howdoyouaccesstheinternetwhileshoppingon-line?	object
Whichdevicedoyouusetooaccesstheonlineshopping?	object
Whatisthescreensizeofyourmobiledevice?	object
Whatistheoperatingsystem(OS)ofyourdevice?	object
Whatbrowserdoyourunonyourdevicetoaccessthewebsite?	object
Whichchanneldidyoufollowtoarriveatyourfavoriteonlinestoreforthefirsttime?	object
Afterfirstvisit,howdoyoureachtheonlinetailstore?	object
Howmuchtimedoyouexplorethee-retailstorebeforemakingapurchasedecision?	object
WhatisyourpreferredpaymentOption?	object
Howfrequentlydoyouabandon(selectinganitemsandleavingwithoutmakingpayment)yourshoppingcart?	object
Whydoyouabandonthe"Bag","ShoppingCart"?	object
Thecontentonthewebsitemustbeeasytoreadandunderstand	object
Informationonsimilarproducttotheonehighlightedisimportantforproductcomparison	object
Completeinformationonlistedsellerandproductbeingofferedisimportantforpurchasedecision.	object
Allrelevantinformationonlistedproductsmustbestatedclearly	object
Easeofnavigationinwebsite	object
Loadingandprocessingspeed	object
UserfriendlyInterfaceofthewebsite	object
ConvenientPaymentmethods	object
Trustthattheonlinetailstorewillfulfillitspartofthetransactionatthestipulatedtime	object
Empathy(readinesstoassistwithqueries)towardsthecustomers	object
Beingabletoguaranteetheprivacyofthecustomer	object
Responsiveness,availabilityofseveralcommunicationchannels(email,onlinerep,twitter,phoneetc.)	object
Onlineshoppinggivesmonetarybenefitanddiscounts	object
Recommendationfromotheronline	object

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 Customer_Retention_df.shape

```
In [558]: ##### check the shape of datatypes
          Customer_Retention_df.shape
```

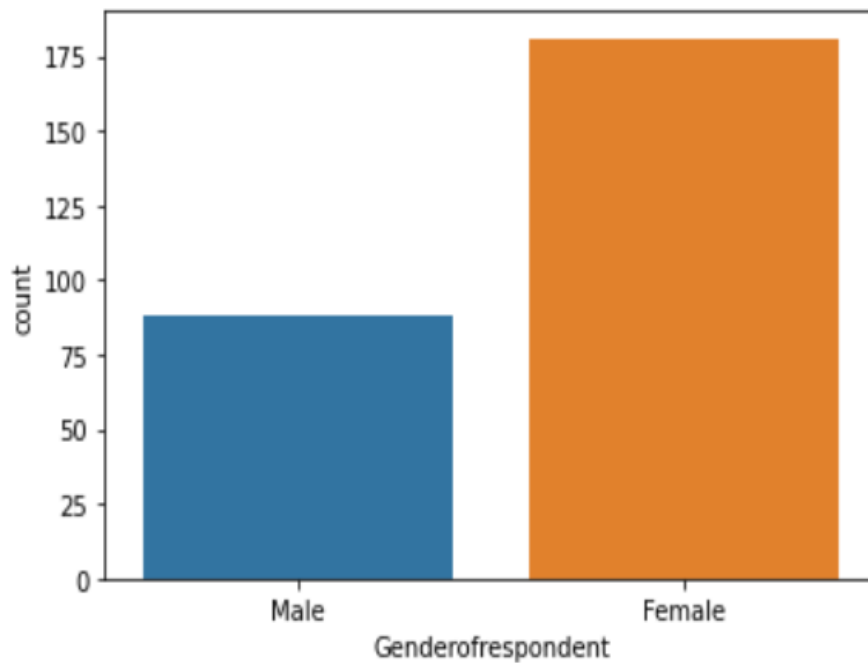
```
Out[558]: (269, 71)
```

There are 269 Rows and 71 Columns

some Data visualization to understand the data set

```
In [561]: sns.countplot(Customer_Retention_df['Genderofrespondent'])
```

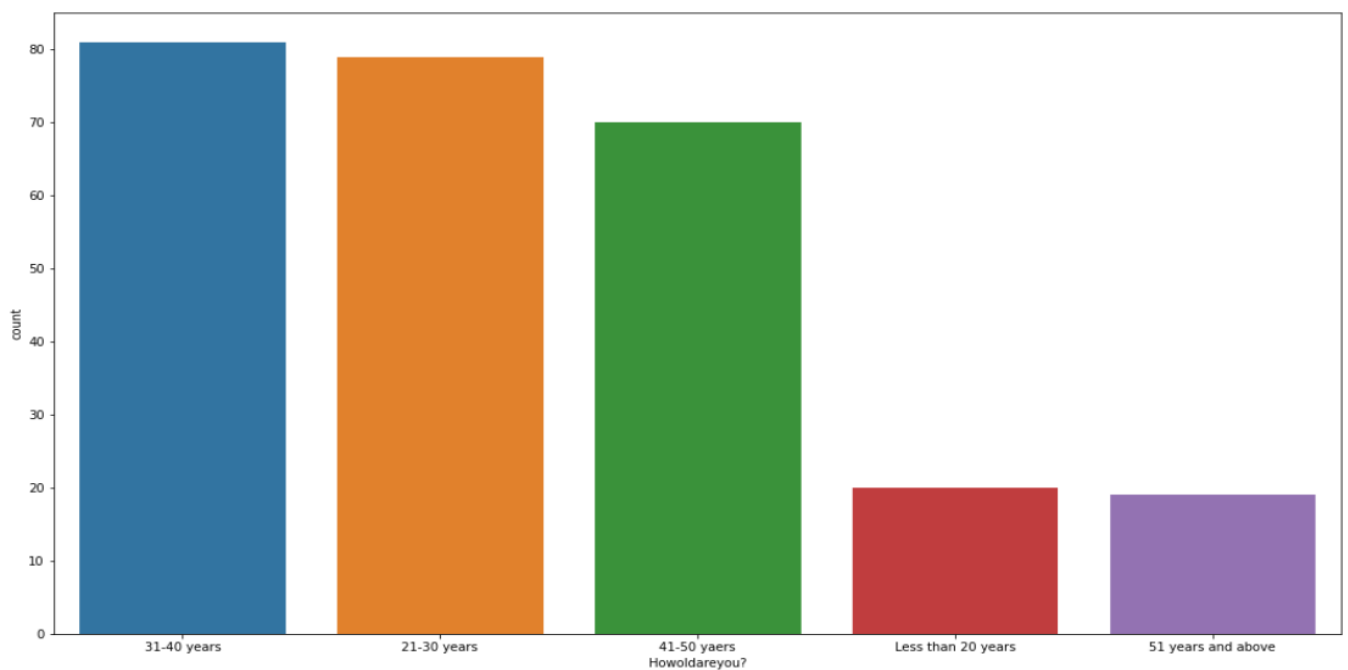
```
Out[561]: <AxesSubplot:xlabel='Genderofrespondent', ylabel='count'>
```



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

```
In [562]: plt.figure(figsize=(18,10))  
sns.countplot(Customer_Retention_df['Howoldareyou?'])
```

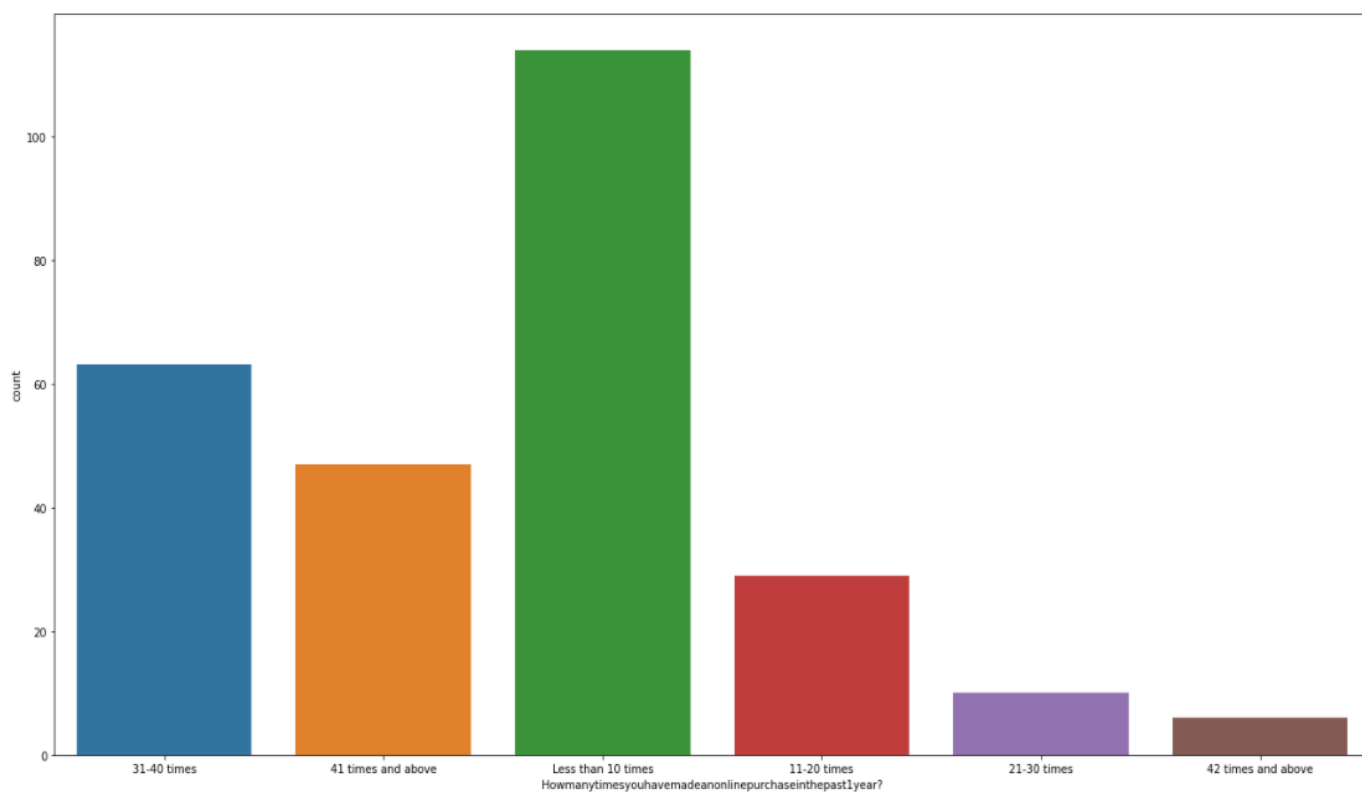
```
Out[562]: <AxesSubplot:xlabel='Howoldareyou?', ylabel='count'>
```



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

```
In [567]: plt.figure(figsize=(22,12))
sns.countplot(Customer_Retention_df['Howmanytimesyouhavemadeanonlinepurchaseinthepast1year?'])
```

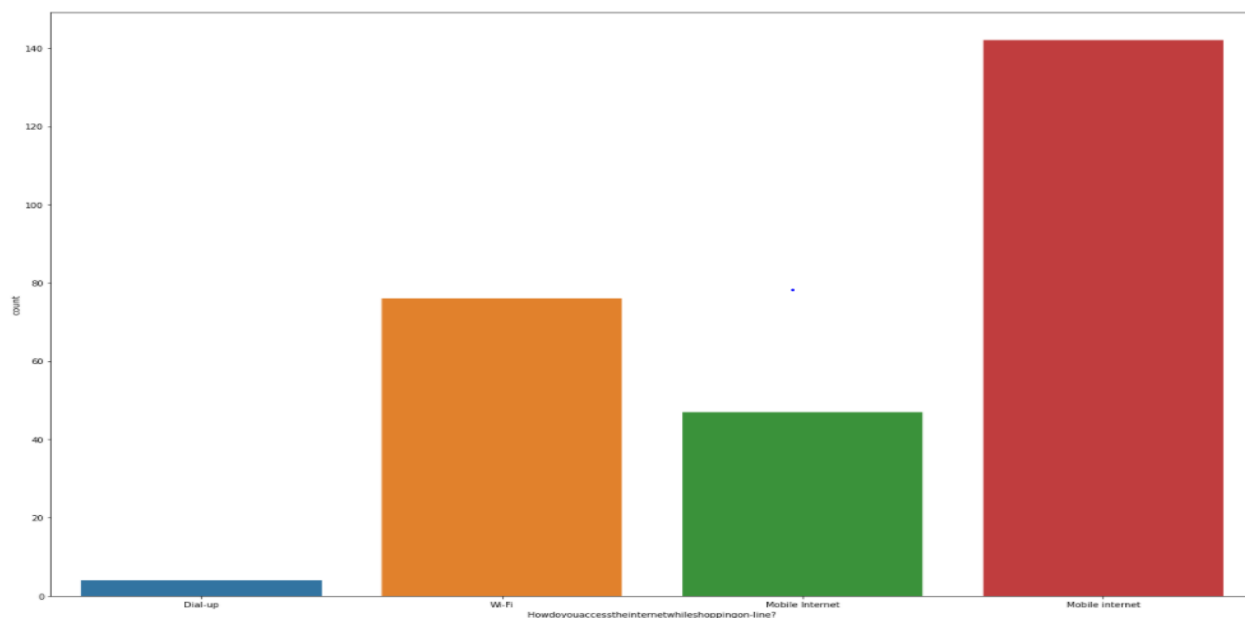
```
Out[567]: <AxesSubplot:xlabel='Howmanytimesyouhavemadeanonlinepurchaseinthepast1year?', ylabel='count'>
```



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

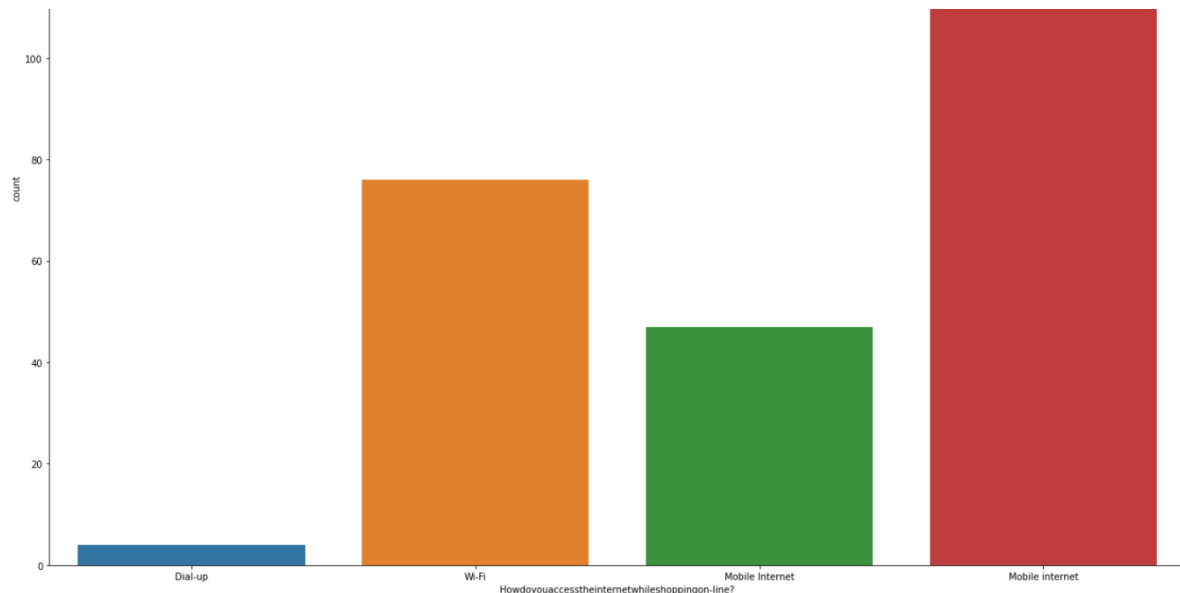
```
In [571]: plt.figure(figsize=(22,15))
sns.countplot(Customer_Retention_df['Howdoyouaccesstheinternetwhilesoppingon-line?'])
```

```
Out[571]: <AxesSubplot:xlabel='Howdoyouaccesstheinternetwhilesoppingon-line?', ylabel='count'>
```



Mostly People use smartphone for online shopping.

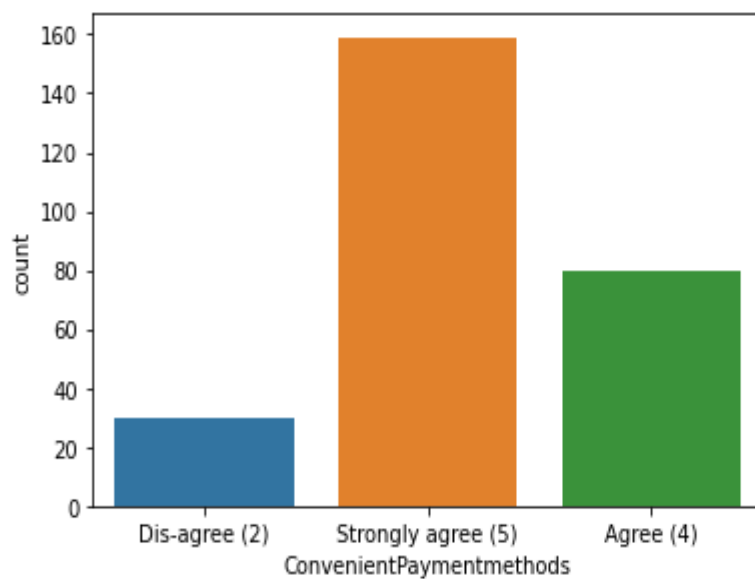
```
In [571]: plt.figure(figsize=(22,15))
sns.countplot(Customer_Retention_df['Howdoyouaccesstheinternetwhileshoppingon-line?'])
```



Mostly People use mobile internet for shopping.

```
In [576]: sns.countplot(Customer_Retention_df['ConvenientPaymentmethods'])
```

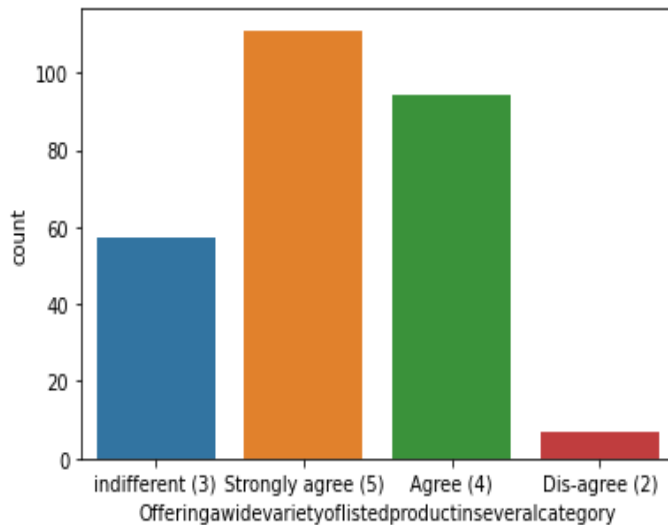
```
Out[576]: <AxesSubplot:xlabel='ConvenientPaymentmethods', ylabel='count'>
```



Continent Payment Method is Strongly agree for payment method.

```
In [578]: sns.countplot(Customer_Retention_df["Offeringawidevarietyoflistedproductinseveralcategory"])
```

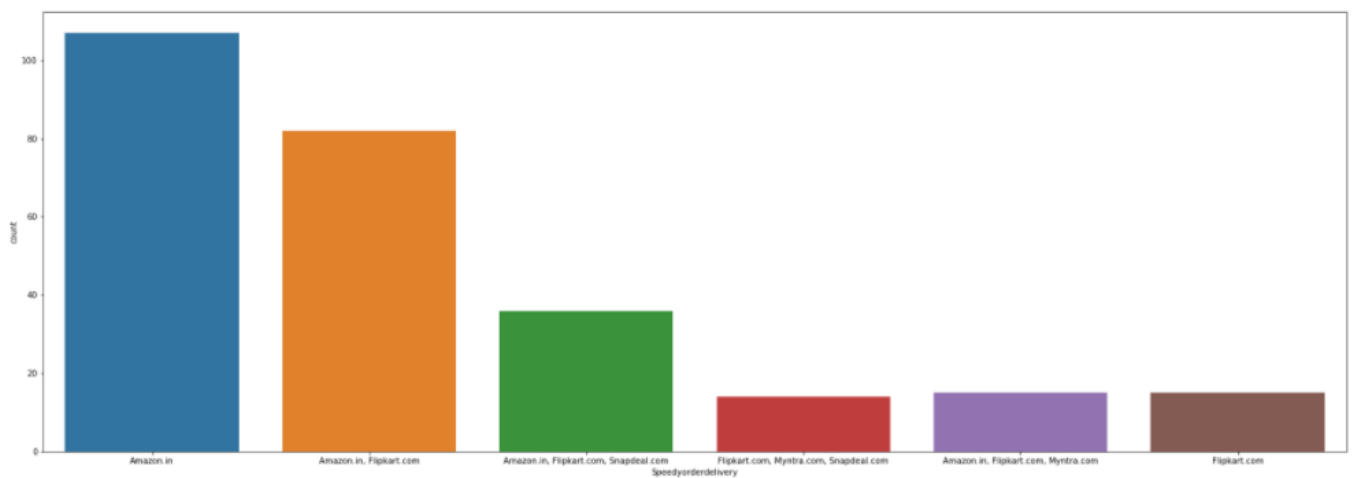
```
Out[578]: <AxesSubplot:xlabel='Offeringawidevarietyoflistedproductinseveralcategory', ylabel='count'>
```



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

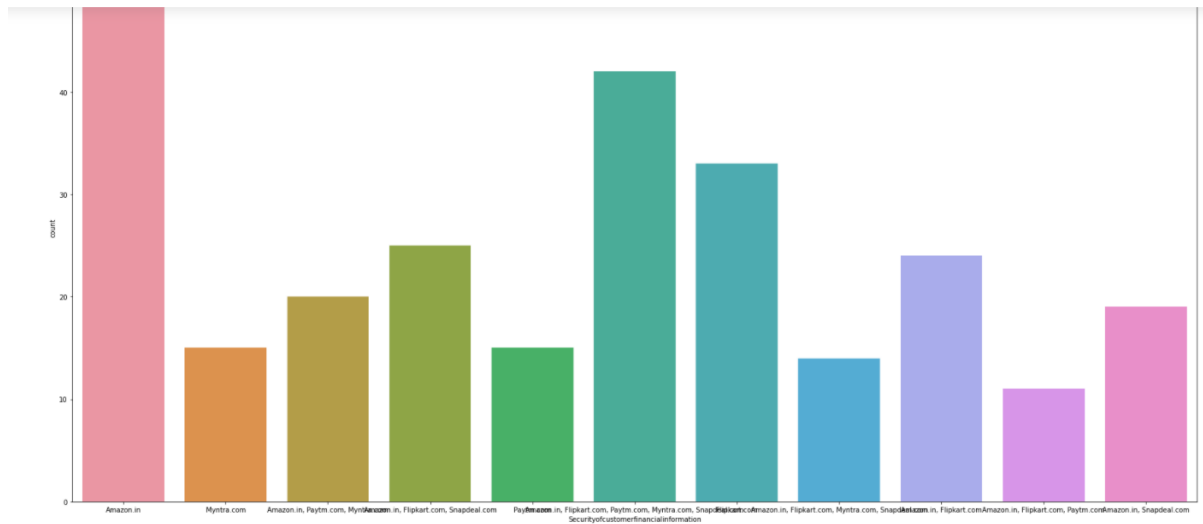
```
In [580]: plt.figure(figsize=(28,10))
sns.countplot(Customer_Retention_df['Speedyorderdelivery'])
```

```
Out[580]: <AxesSubplot:xlabel='Speedyorderdelivery', ylabel='count'>
```



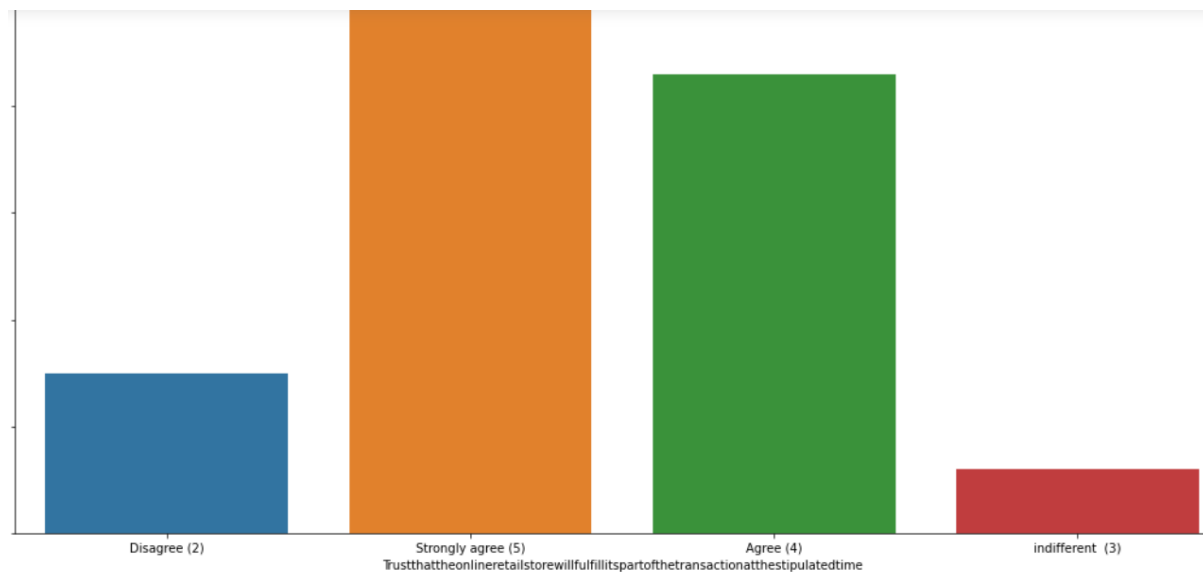
Mostly Amazon gives the speed of the delivery.

```
In [582]: plt.figure(figsize=(30,15))
sns.countplot(Customer_Retention_df['Securityofcustomerfinancialinformation'])
```

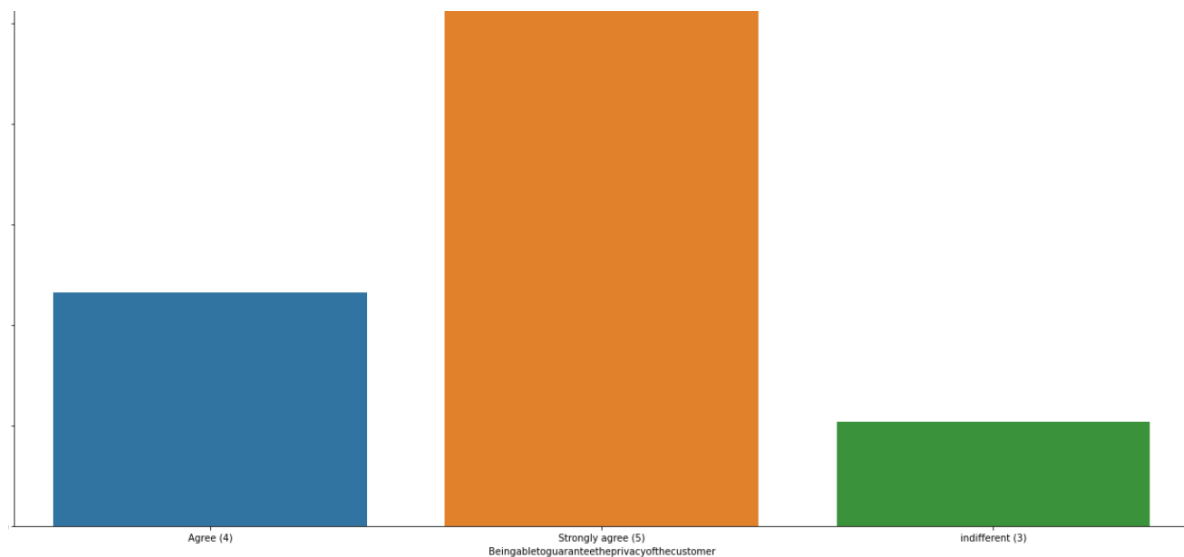
Amazon keeps the the security of customer financial information mostly shown in graph and its seems to good for the customer.

```
In [584]: plt.figure(figsize=(18,12))
sns.countplot(Customer_Retention_df['Trustthattheonlinetailstorewillfulfillitspartofthetransactionatthetipulatedtime'])
```



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

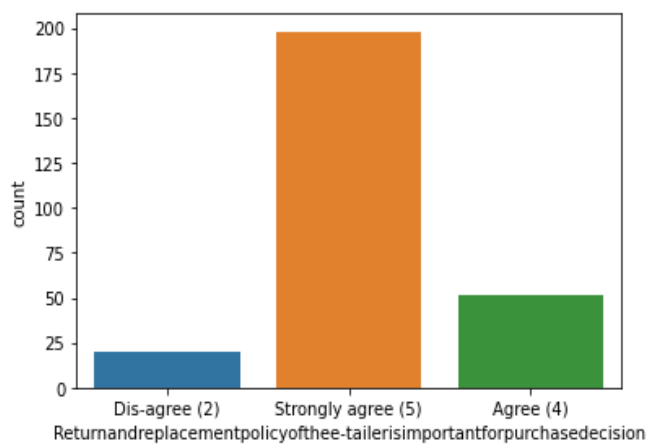
```
In [586]: plt.figure(figsize=(22,15))
sns.countplot(Customer_Retention_df['Beingabletoguaranteetheprivacyofthecustomer'])
```



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

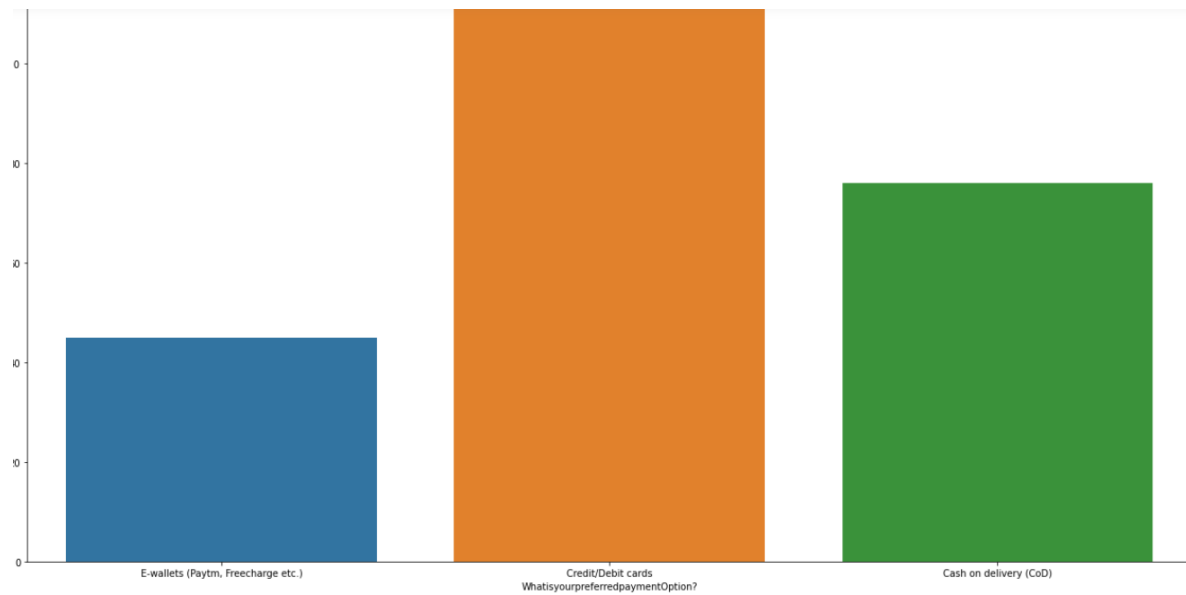
```
In [588]: sns.countplot(Customer_Retention_df['Returnandreplacementpolicyofthee-tailerisimportantforpurchasedecision'])
```

```
Out[588]: <AxesSubplot:xlabel='Returnandreplacementpolicyofthee-tailerisimportantforpurchasedecision', ylabel='count'>
```



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

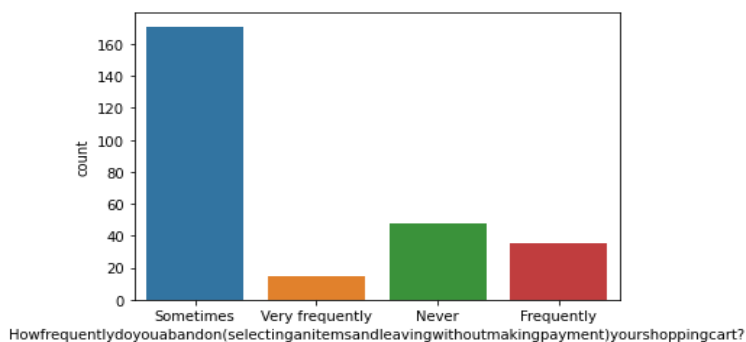
```
In [590]: plt.figure(figsize=(22,15))
sns.countplot(Customer_Retention_df['WhatisyourpreferredpaymentOption?'])
```



Credit/debit card is preferred payment option chosen highly by the customers for online shopping.

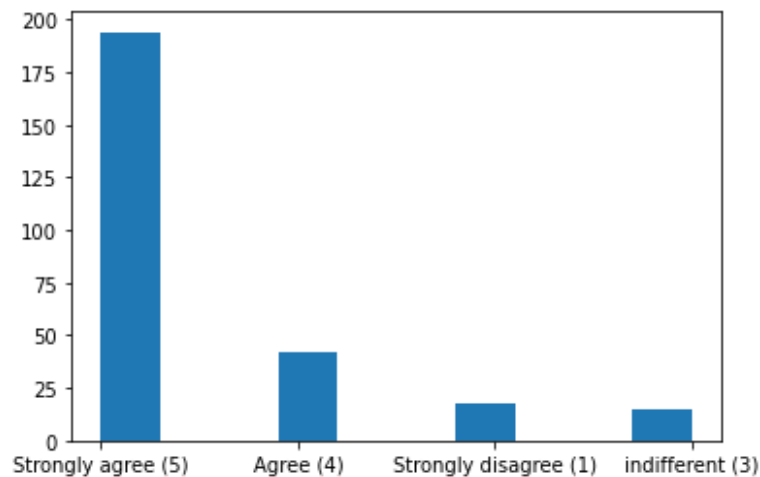
```
In [592]: sns.countplot(Customer_Retention_df['How frequently do you abandon (selecting an item and leaving without making payment) your shopping cart?'])
```

```
Out[592]: <AxesSubplot: xlabel='How frequently do you abandon (selecting an item and leaving without making payment) your shopping cart?', ylabel='count'>
```



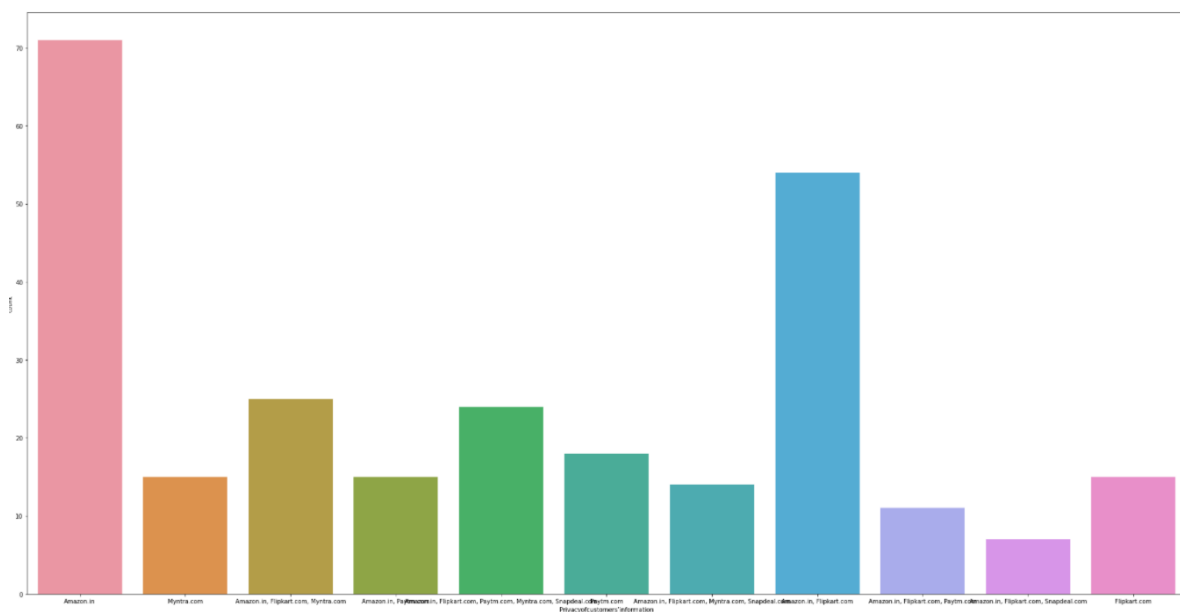
Sometimes is highly countable for selecting and leaving without making payment

```
In [594]: # Creating plot
plt.hist(Customer_Retention_df['Empathy(readinesstoassistwithqueries)towardsthecustomers'])
plt.show()
```



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

```
In [598]: plt.figure(figsize=(35,18))
sns.countplot(Customer_Retention_df['Privacyofcustomers'information'])
```



Amazon keep the customer policy information of customers

Pre-processing

```
In [604]: Customer_Retention_df['Genderofrespondent'].unique()
```

```
Out[604]: array(['Male', 'Female'], dtype=object)
```

```
In [605]: ##### import label encoder to convert catagorical value
import sklearn
from sklearn import preprocessing
le = preprocessing.LabelEncoder()
```

```
In [606]: Customer_Retention_df['Genderofrespondent'] = le.fit_transform(Customer_Retention_df['Genderofrespondent'])
```

```
In [607]: Customer_Retention_df['Genderofrespondent'].unique() # 1 for male and 0 for female
```

```
Out[607]: array([1, 0])
```

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

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

```
In [608]: Customer_Retention_df['Howoldareyou?'].unique()
```

```
Out[608]: array(['31-40 years', '21-30 years', '41-50 yaers', 'Less than 20 years',
                '51 years and above'], dtype=object)
```

```
In [609]: Customer_Retention_df['Howoldareyou?'] = Customer_Retention_df['Howoldareyou?'].replace('31-40 years', '31')
```

```
In [610]: Customer_Retention_df['Howoldareyou?'] = Customer_Retention_df['Howoldareyou?'].replace('Less than 20 years', '10')
Customer_Retention_df['Howoldareyou?'] = Customer_Retention_df['Howoldareyou?'].replace('21-30 years', '20')
Customer_Retention_df['Howoldareyou?'] = Customer_Retention_df['Howoldareyou?'].replace('41-50 yaers', '40')
Customer_Retention_df['Howoldareyou?'] = Customer_Retention_df['Howoldareyou?'].replace('51 years and above', '60')
```

```
In [611]: Customer_Retention_df['Whichcitydoyoushoponlinefrom?'].unique()
```

```
Out[611]: array(['Delhi', 'Greater Noida', 'Karnal ', 'Bangalore ', 'Noida',
                'Solani', 'Moradabad', 'Gurgaon ', 'Merrut', 'Ghaziabad',
                'Bulandshahr'], dtype=object)
```

```
In [612]: Customer_Retention_df['Whichcitydoyoushoponlinefrom?'] = le.fit_transform(Customer_Retention_df['Whichcitydoyoushoponlinefrom?'])
```

```
In [613]: Customer_Retention_df['WhatisthePinCodeofwhereyoushoponlinefrom?'].unique()
```

```
Out[613]: array([110009, 110030, 201308, 132001, 530068, 110011, 110018, 173229,
                110039, 244001, 122018, 201310, 560037, 203207, 250001, 201005,
                110044, 201306, 560010, 201305, 110042, 132036, 560018, 110008,
                560002, 201303, 201312, 203202, 560001, 201304, 560003, 110014,
                560013, 173212, 122009, 201009, 201008, 201001, 203001],
                dtype=int64)
```

```
In [614]: Customer_Retention_df['WhatisthePinCodeofwhereyoushoponlinefrom?'] = le.fit_transform(Customer_Retention_df['WhatisthePinCodeofwh
```

```
Out[619]:
```

	Genderofrespondent	Howoldareyou?	Whichcitydoyoushoponlinefrom?	WhatisthePinCodeofwherelyoushoponlinefrom?	SinceHowLongYouareShoppingOnline
0	1	31	2	1	
1	0	20	2	5	
2	0	20	4	23	
3	1	20	6	11	
4	0	20	0	31	
5	0	31	9	23	

```
In [631]: Customer_Retention_df['DisplayingqualityInformationonthewebsiteimprovesatisfactionofcustomers'] = le.fit_transform(Customer_Rete
Customer_Retention_df['Userderivesatisfactionwhileshoppingonagoodqualitywebsiteorapplication'] = le.fit_transform(Customer_Retent
Customer_Retention_df['Usersatisfactioncannotexistwithoutrust'] = le.fit_transform(Customer_Retention_df['Usersatisfactioncannot
Customer_Retention_df['Offeringawidevarietyoflistedproductinseveralcategory'] = le.fit_transform(Customer_Retention_df['Offeringa
Customer_Retention_df['Provisionofcompleteandrelevantproductinformation'] = le.fit_transform(Customer_Retention_df['Provisionofcc

Customer_Retention_df['Monetarysavings'] = le.fit_transform(Customer_Retention_df['Monetarysavings'])
Customer_Retention_df['TheConvenienceofpatronizingtheonlineretailer'] = le.fit_transform(Customer_Retention_df['TheConvenienceofp
Customer_Retention_df['Shoppingonthewebsitegivesyouthesenseofadventure'] = le.fit_transform(Customer_Retention_df['Shoppingonthev
Customer_Retention_df['Shoppingonyourpreferredetailerenhancesyoursocialstatus'] = le.fit_transform(Customer_Retention_df['Shoppi
Customer_Retention_df['Youfeelgratificationshoppingonyourfavoriteetailer'] = le.fit_transform(Customer_Retention_df['Youfeelgrat

Customer_Retention_df['Shoppingonthewebsitehelpsyoufulfillcertainroles'] = le.fit_transform(Customer_Retention_df['Shoppingonthev
Customer_Retention_df['Gettingvalueformoneyspent'] = le.fit_transform(Customer_Retention_df['Gettingvalueformoneyspent'])
Customer_Retention_df['NetBenefitderiveCustomer_Retention_dfromshoppingonlinecanleadtouserssatisfaction'] =le.fit_transform(Custo
```

```
In [632]: Customer_Retention_df.head()
```

head function show top 5 rows from the dataset

```
In [633]: Customer_Retention_df['Fromthefollowing,tickany(orall)oftheonlineretailersyouhaveshoppedfrom;'].unique()
```

```
Out[633]: array(['Amazon.in, Paytm.com',
                'Amazon.in, Flipkart.com, Myntra.com, Snapdeal.com',
                'Amazon.in, Paytm.com, Myntra.com',
                'Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com',
                'Amazon.in, Flipkart.com, Paytm.com, Snapdeal.com',
                'Amazon.in, Flipkart.com', 'Amazon.in, Flipkart.com, Snapdeal.com',
                'Amazon.in', 'Amazon.in, Flipkart.com, Paytm.com'], dtype=object)
```

In [731]: Customer_Retention_df

Out[731]:

	Genderofrespondent	Howoldareyou?	Whichcitydoyoushoponlinefrom?	WhatisthePinCodeofwhereyoushoponlinefrom?	SinceHowLongYouareShoppingOnline
0	1	2	2	1	
1	0	1	2	5	
2	0	1	4	23	
3	1	1	6	11	
4	0	1	0	31	
5	0	2	9	23	
6	1	3	2	2	
7	1	3	2	4	
8	0	0	10	14	
9	0	2	2	6	

Conclusion:

Customer aspect various things from the online store like privacy, wide 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.