Customer retention analysis

A Data Science Project by - Anshul Dubey

Introduction: A good method for companies to increase revenue is to invest in pre-existing customers. Retaining customers becomes essential for consumer dependent businesses. If customers are on the verge of leaving the service, it would be prudent for the company to identify what factors influence a customer's lack of interest in the product. As the cost of retaining a customer is far lesser than getting a new one, analyzing customer churn can reveal valuable insights. The loss of customers is known as customer churn or customer attrition. Companies analyze customer churn to uncover which factors lead to a customer voluntarily switching to a rival business.

Customer churn occurs when customers or subscribers stop doing business with a company or service. Also known as customer attrition, customer churn is a critical metric because it is less expensive to retain existing customers than it is to acquire new customers – earning business from new customers means working leads all the way through the sales funnel, utilizing your marketing and sales resources throughout the process. Customer retention, on the other hand, is generally more cost-effective, as you have already earned the trust and loyalty of existing customers. Using a machine learning approach, we can find patterns which cause customer churn and forecast it to obtain a prognosis on which factors impact customer retention. It can help make sense of relationships between data. The models can predict customers with high probability to churn based on analyzing customers personal, demographic and behavioural data to provide personalized and customer-oriented marketing campaigns to gain customer satisfaction.

Project Description: 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.

Projects Goals: Gaining an understanding of the customers' **churn reasons is a powerful component of designing a data-driven customer retention strategy. Identifying the customers that are likely to churn and preventing attrition is a challenging task. Simple business heuristics often fall short on.**

That is why, the goal of our project was to perform a deep analysis of data in terms of customer retention, building a mechanism for identifying customers at risk of churn, and supporting the prevention of churn.

Process Of Customer Retention Analysis: In the first stage of the project, it's very important to **analyze business processes and perform data analysis. We run statistical analysis of all available attributes, analyze existing data structure, as well as customer care department actions and all related business aspects.**

Tasks that have been performed from a business point of view:

1. **Gathering requirements from business departments**
2. **Customer care processes analysis**
3. **Analysis of existing IT infrastructure**
4. **Preparation of the project plan**

Tasks that have been performed from the data point of view:

1. **Analysis of available data types**
2. **Visual data analysis**
3. **Correlation analysis**
4. **Outlier detection (dbscan, isolation forest)**
5. **Missing values analysis**
6. Analysis and definition of the “target” variable

Based on the results and insights obtained regarding these steps, we have a better understanding of what variables we will be able to generate at the data preparation stage and what the system architecture will look like.

Why We Will Do Customer Retention Analysis:

Customer retention analysis helps to understand how customers are influenced by different business decisions. By analyzing customer churn, business users are able to understand and see trends in product or service satisfaction/dissatisfaction. Analysis based on cohorts and demographic data can be very helpful. It delivers insights on what is impacting particular customer decisions (price changes, new products or services, product upgrades or changes in customer communication, and other).

Dashboards provide a convenient interface where you can visualize and analyze data and focus on key performance indicators (KPIs) from across your organization, helping you gain valuable insight and make quick and accurate decisions.

Key customer retention benefits:

1. **Split customers into cohorts and custom lists to find out who is driving your business growth and answer complex questions about your next investments,**
2. **Ability to conduct in-depth analysis to gain insight and correlations between different subscriptions and business activities,**

Monitor all KPIs in one place to understand business performance.

Data Overview:

In this tutorial, we will use the customer retention data-set. Each row represents a customer.We have 47 type of data of each customer in 47 columns of the dataset. Below is a short description of each feature inside the data.

1. 1 Gender of respondent
   - 2 How old are you?
2. 3 Which city do you shop online from?
3. 4 What is the Pin Code of where you sh op online from?
   - 5 Since How Long You are Shopping Online ?

   6 How many times you have made an onli ne purchase in the past 1

   year?
4. 7 How do you access the internet while shopping on-line?
   - 8 Which device do you use to access th e online shopping?
   - 9 What is the screen size of your mobi le device?\t\t\t\t\t\t
   - 10 What is the operating system (OS) o f your device?\t\t\t\t
   - 11 What browser do you run on your dev ice to access the website?\t\t\t

5. 12 Which channel did you follow to arr ive at your favorite online store for t he first time?
   - 13 After first visit, how do you reach the online retail store?\t\t\t\t
6. 14 How much time do you explore the e-retail store before making a purchase d ecision?
   - 15 What is your preferred payment Opti on?\t\t\t\t\t
7.
1. 16 How frequently do you abandon (sele cting an items and leaving without mak ing payment) your shopping cart?\t\t\t \t\t\t\t
2. 17 Why did you abandon the “Bag”, “Sho pping Cart”?\t\t\t\t\t
   - 18 The content on the website must be e asy to read and understand
1. 19 Information on similar product to t he one highlighted is important for p roduct comparison
2. 20 Complete information on listed sell er and product being offered is import ant for purchase decision.
   - 21 All relevant information on listed p roducts must be stated clearly
3. 22 Ease of navigation in website
1. 23 Loading and processing speed

2. 24 User friendly Interface of the webs ite

25 Convenient Payment methods
26 Trust that the online retail store w ill fulfill its part of the transactio n at the stipulated time

- 27 Empathy (readiness to assist with q ueries) towards the customers

1. • 28 Being able to guarantee the privacy of the customer

29 Responsiveness, availability of sev eral communication channels (email, on line rep, twitter, phone etc.)

- 30 Online shopping gives monetary bene fit and discounts
- 31 Enjoyment is derived from shopping online

1. • 32 Shopping online is convenient and f lexible
2. 33 Return and replacement policy of th e e-tailer is important for purchase d ecision
   - 34 Gaining access to loyalty programs i s a benefit of shopping online
1. 35 Displaying quality Information on t he website improves satisfaction of cu stomers
2. 36 User derive satisfaction while shop ping on a good quality website or appl

ication

- 37 Net Benefit derived from shopping o nline can lead to users satisfaction
- 38 User satisfaction cannot exist with out trust

39 Offering a wide variety of listed product in several category

- 40 Provision of complete and relevant p roduct information

3. 41 Monetary savings
4. 42 The Convenience of patronizing the o nline retailer
   - 43 Shopping on the website gives you t he sense of adventure
   - 44 Shopping on your preferred e-tailer enhances your social status
   - 45 You feel gratification shopping on y our favorite e-tailer
   - 46 Shopping on the website helps you f ulfill certain roles
5. 47 Getting value for money spent

In [5]:
```
df
```

Out[5]:

	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)	Longer page loading time (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	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	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	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	Paytm.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	Paytm.com
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
		21-													

In [6]:
```
df.isnull().sum()
```

Out[6]:
```
1Gender of respondent                                                0
2 How old are you?                                                   0
3 Which city do you shop online from?                                0
4 What is the Pin Code of where you shop online from?                0
5 Since How Long You are Shopping Online ?                           0
                                                                    ..
Longer delivery period                                               0
Change in website/Application design                                 0
Frequent disruption when moving from one page to another             0
Website is as efficient as before                                    0
Which of the Indian online retailer would you recommend to a friend? 0
Length: 71, dtype: int64
```

so here in the datset we can see that there is no null values present to be handled

Data Exploration with consumer analysis

In [19]:
```
df['1Gender of respondent'].unique()
```

Out[19]:
```
array(['Male', 'Female'], dtype=object)
```

In [21]:
```
df['1Gender of respondent'].value_counts()
```

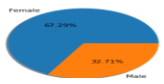
Out[21]:
```
Female    181
Male       88
Name: 1Gender of respondent, dtype: int64
```

using pic plot we will see how data looks like

In [24]:
```
import matplotlib.pyplot as plt
```

In [25]:
```
labels = 'Female','Male'
plt.figure(figsize=(15,6))
fig, ax = plt.subplots()
ax.pie(df['1Gender of respondent'].value_counts(),labels = labels,radius =1,autopct = '%1.2f%%', shadow=True,)
plt.title('Percentage of male and female')
plt.show()
```

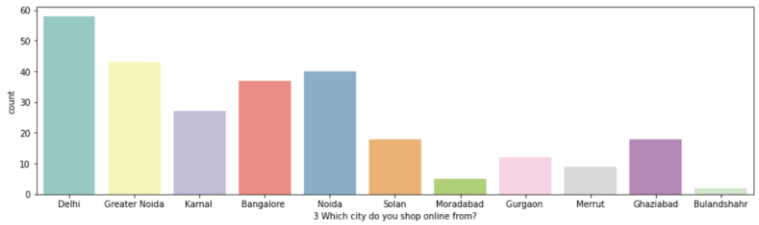
```
<Figure size 1080x432 with 0 Axes>
```



Highest Online shopping is done in these cities of India

In [30]:
```
plt.figure(figsize=(15,4),facecolor='white')
sns.countplot(df['3 Which city do you shop online from?'], palette="Set3")
```

Out[30]:
```
<AxesSubplot:xlabel='3 Which city do you shop online from?', ylabel='count'>
```



Here we can see that Delhi tops in this column metrics and then Greater Noida and Karnal is being followed by it

shopping According to the Gender

Customer online Shopping and what they prefer more

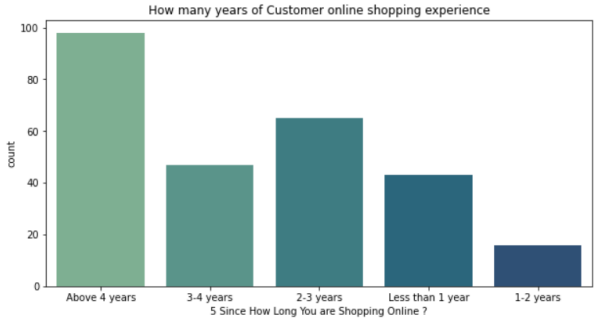
In [42]:
```
df['5 Since How Long You are Shopping Online ?'].value_counts()
```

Out[42]:
```
Above 4 years       98
2-3 years           65
3-4 years           47
Less than 1 year    43
1-2 years           16
Name: 5 Since How Long You are Shopping Online ?, dtype: int64
```

As we can see that there are max person who have been shopping for 4 yrs followed by 2-3 yrs

In [44]:
```
plt.figure(figsize=(10,5))
sns.countplot(df['5 Since How Long You are Shopping Online ?'], palette="crest")
plt.title('How many years of Customer online shopping experience')
```

Out[44]:
```
Text(0.5, 1.0, 'How many years of Customer online shopping experience')
```



Customer online Shopping and what they prefer more

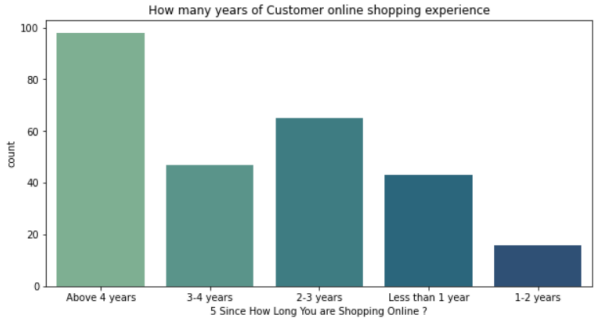
In [42]:
```
df['5 Since How Long You are Shopping Online ?'].value_counts()
```

Out[42]:
```
Above 4 years       98
2-3 years           65
3-4 years           47
Less than 1 year    43
1-2 years           16
Name: 5 Since How Long You are Shopping Online ?, dtype: int64
```

As we can see that there are max person who have been shopping for 4 yrs followed by 2-3 yrs

In [44]:
```
plt.figure(figsize=(10,5))
sns.countplot(df['5 Since How Long You are Shopping Online ?'], palette="crest")
plt.title('How many years of Customer online shopping experience')
```

Out[44]:
```
Text(0.5, 1.0, 'How many years of Customer online shopping experience')
```

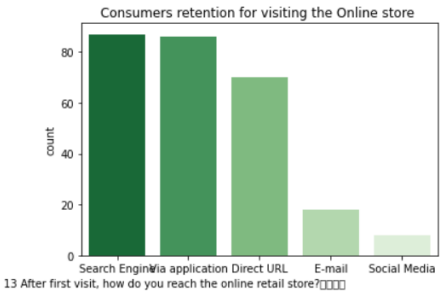


In [82]:
```
df[       '13 After first visit, how do you reach the online retail store?\t\t\t\t
].value_counts()
```

Out[82]:
```
Search Engine      87
Via application    86
Direct URL         70
E-mail             18
Social Media        8
Name: 13 After first visit, how do you reach the online retail store?\t\t\t\t
```

In [83]:
```
sns.countplot(df[       '13 After first visit, how do you reach the online retail store?\t\t\t\t
], palette="Greens_r")
plt.title('Consumers retention for visiting the Online store ')
```

Out[83]:
```
Text(0.5, 1.0, 'Consumers retention for visiting the Online store ')
```

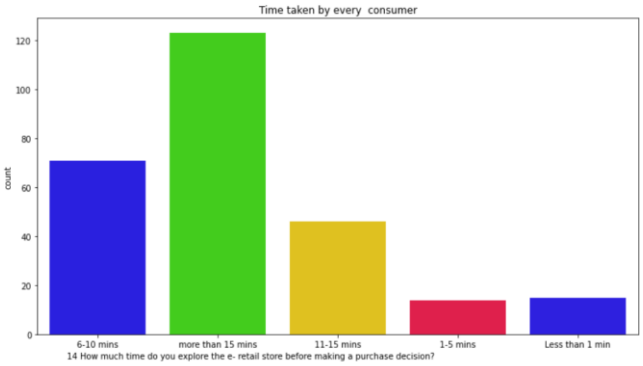


Here we can see that the Search engine and the Application has the higest number of consumers retention part

How much a consumer take time to purchase something in e-retail store

In [86]:
```
plt.figure(figsize=(13,7))
sns.countplot(df[       '14 How much time do you explore the e- retail store before making a purchase decision?
], palette ='prism')
plt.title('Time taken by every  consumer')
```

Out[86]:
```
Text(0.5, 1.0, 'Time taken by every  consumer')
```



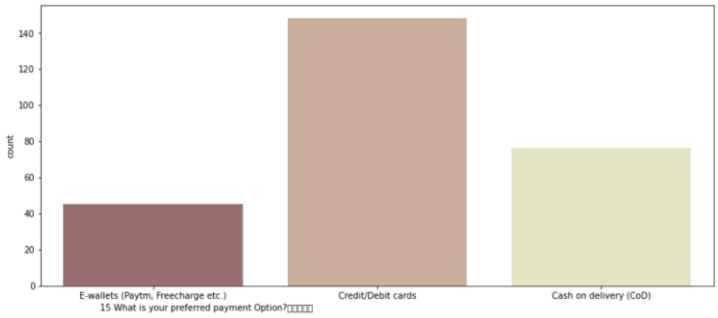
So as we can see that the max people before buying something in any retail store takes around 15 mins and next followed by 6-10 mins and very few people are there who takes less then 1 min of time.

In [88]:
```
df[       '15 What is your preferred payment Option?\t\t\t\t\t
].value_counts()
```

Out[88]:
```
Credit/Debit cards                  148
Cash on delivery (CoD)               76
E-wallets (Paytm, Freecharge etc.)   45
Name: 15 What is your preferred payment Option?\t\t\t\t\t
, dtype: int64
```

In [89]:
```
plt.figure(figsize =(14,6))
sns.countplot(df[       '15 What is your preferred payment Option?\t\t\t\t\t
], palette ='pink')
```

Out[89]:
```
<AxesSubplot:xlabel='15 What is your preferred payment Option?\t\t\t\t\t
', ylabel='count'>
```

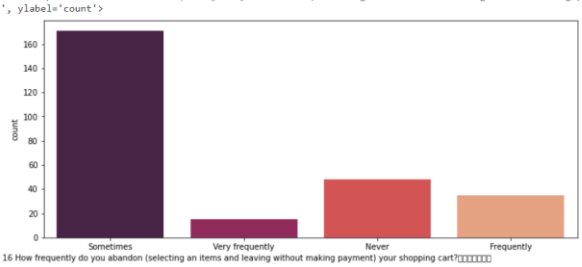


In [92]:
```
df[       '16 How frequently do you abandon (selecting an items and leaving without making payment) your shopping cart?\t\t\t\t\t\t\t
].value_counts()
```

Out[92]:
```
Sometimes           171
Never                48
Frequently           35
Very frequently      15
Name: 16 How frequently do you abandon (selecting an items and leaving without making payment) your shopping cart?\t\t\t\t\t\t\t
, dtype: int64
```

In [94]:
```
plt.figure(figsize=(12,5))
sns.countplot(df[       '16 How frequently do you abandon (selecting an items and leaving without making payment) your shopping cart?\t\t\t\t\t\t\t
],palette ='rocket')
```

Out[94]:
```
<AxesSubplot:xlabel='16 How frequently do you abandon (selecting an items and leaving without making payment) your shopping cart?\t\t\t\t\t\t\t
', ylabel='count'>
```



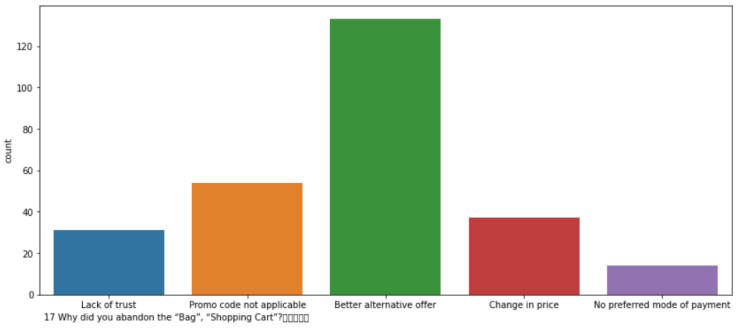
Here we can see most of the peope do online shopping sometimes and very less people to it frequently

In [96]:
```
df[       '17 Why did you abandon the “Bag”, “Shopping Cart”?\t\t\t\t\t
].value_counts()
```

Out[96]:
```
Better alternative offer        133
Promo code not applicable        54
Change in price                  37
Lack of trust                    31
No preferred mode of payment     14
Name: 17 Why did you abandon the “Bag”, “Shopping Cart”?\t\t\t\t\t
, dtype: int64
```

In [97]:
```
plt.figure(figsize=(14,6))
sns.countplot(df[       '17 Why did you abandon the “Bag”, “Shopping Cart”?\t\t\t\t\t
])
```

Out[97]:
```
<AxesSubplot:xlabel='17 Why did you abandon the “Bag”, “Shopping Cart”?\t\t\t\t\t
', ylabel='count'>
```



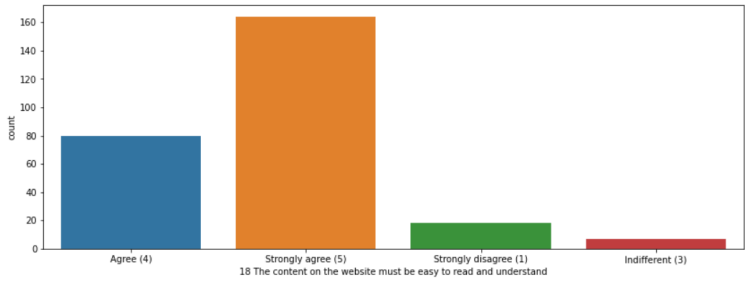
Diffrent Opinions by customers on website features

In [98]:
```
df[       '18 The content on the website must be easy to read and understand'
].value_counts()
```

Out[98]:
```
Strongly agree (5)       164
Agree (4)                 80
Strongly disagree (1)     18
Indifferent (3)            7
Name: 18 The content on the website must be easy to read and understand, dtype: int64
```

In [100…
```
plt.figure(figsize=(14,5))
sns.countplot(df[       '18 The content on the website must be easy to read and understand'
])
```

Out[100…
```
<AxesSubplot:xlabel='18 The content on the website must be easy to read and understand', ylabel='count'>
```



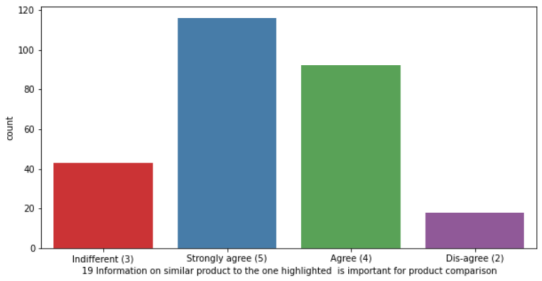
So we can see that the most of consumers sentiment is that all the online portals should be simple to understand and not very complex

In [101…
```
df[       '19 Information on similar product to the one highlighted  is important for product comparison'
].value_counts()
```

Out[101…
```
Strongly agree (5)    116
Agree (4)              92
Indifferent (3)        43
Dis-agree (2)          18
Name: 19 Information on similar product to the one highlighted  is important for product comparison, dtype: int64
```

In [102…
```
plt.figure(figsize=(10,5))
sns.countplot(df[       '19 Information on similar product to the one highlighted  is important for product comparison'
], palette ='Set1')
```

Out[102…
```
<AxesSubplot:xlabel='19 Information on similar product to the one highlighted  is important for product comparison', ylabel='count'>
```



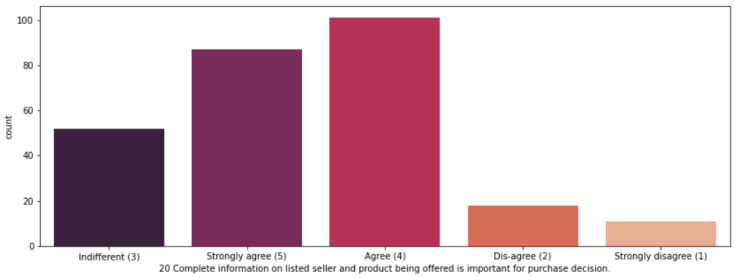
As here we can see that most people strongly agree with the consent

In [103…
```
df[       '20 Complete information on listed seller and product being offered is important for purchase decision.'
].value_counts()
```

Out[103…
```
Agree (4)                101
Strongly agree (5)        87
Indifferent (3)           52
Dis-agree (2)             18
Strongly disagree (1)     11
Name: 20 Complete information on listed seller and product being offered is important for purchase decision., dtype: int64
```

In [104…
```
plt.figure(figsize=(14,5))
sns.countplot(df[       '20 Complete information on listed seller and product being offered is important for purchase decision.'
],palette ='rocket')
```

Out[104…
```
<AxesSubplot:xlabel='20 Complete information on listed seller and product being offered is important for purchase decision.', ylabel='count'>
```



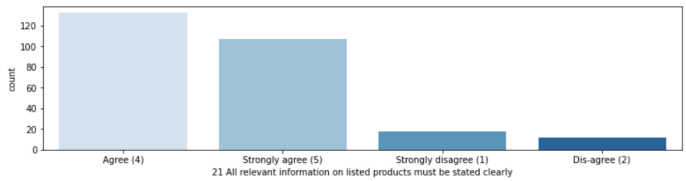
Here we can see the max of people ageree with that the complete info of seller should be given --to make a fast buying decisions

In [105…
```
df[       '21 All relevant information on listed products must be stated clearly'
].value_counts()
```

Out[105…
```
Agree (4)                132
Strongly agree (5)       107
Strongly disagree (1)     18
Dis-agree (2)             12
Name: 21 All relevant information on listed products must be stated clearly, dtype: int64
```

In [106…
```
plt.figure(figsize=(13,3))
sns.countplot(df[       '21 All relevant information on listed products must be stated clearly'
], palette ='Blues')
```

Out[106…
```
<AxesSubplot:xlabel='21 All relevant information on listed products must be stated clearly', ylabel='count'>
```



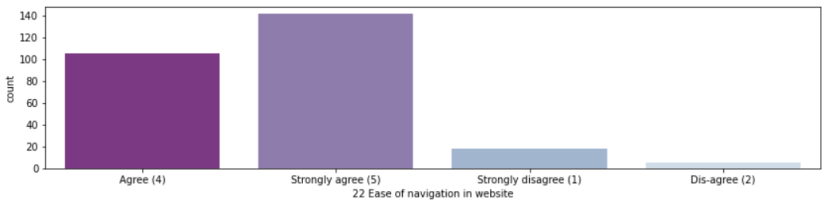
As we can see that most of the consumers want that allthe relevent info should be listed in the products clearly as it will help them to buy that easily and faster

In [107…
```
df['22 Ease of navigation in website'].value_counts()
```

Out[107…
```
Strongly agree (5)       141
Agree (4)                105
Strongly disagree (1)     18
Dis-agree (2)              5
Name: 22 Ease of navigation in website, dtype: int64
```

In [108…
```
plt.figure(figsize=(14,3))
sns.countplot(df['22 Ease of navigation in website'], palette='BuPu_r')
```

Out[108…
```
<AxesSubplot:xlabel='22 Ease of navigation in website', ylabel='count'>
```



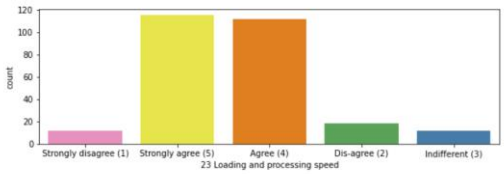
So we can see that the -most of the people want that they should get the Ease of navigation of website

In [109…
```
df[ '23 Loading and processing speed'].value_counts()
```

Out[109…
```
Strongly agree (5)       115
Agree (4)                112
Dis-agree (2)             18
Strongly disagree (1)     12
Indifferent (3)           12
Name: 23 Loading and processing speed, dtype: int64
```

In [110…
```
plt.figure(figsize=(10,3))
sns.countplot(df[ '23 Loading and processing speed'], palette ='Set1_r')
```

Out[110…
```
<AxesSubplot:xlabel='23 Loading and processing speed', ylabel='count'>
```



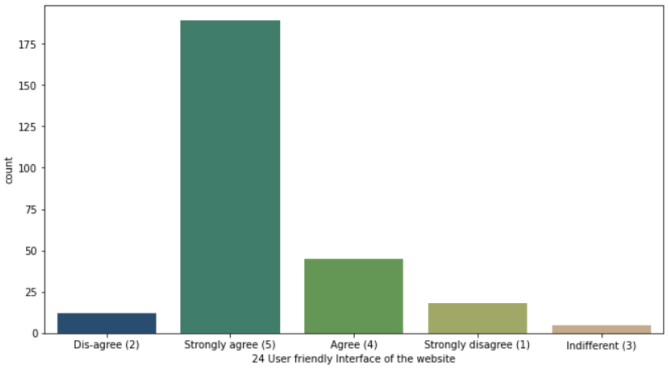
Everyone agreee that the loading and the processing speed of the website should be good

In [111…
```
df[       '24 User friendly Interface of the website'
].value_counts()
```

Out[111…
```
Strongly agree (5)       189
Agree (4)                 45
Strongly disagree (1)     18
Dis-agree (2)             12
Indifferent (3)            5
Name: 24 User friendly Interface of the website, dtype: int64
```

In [112…
```
plt.figure(figsize=(11,6))
sns.countplot(df[       '24 User friendly Interface of the website'
], palette ='gist_earth')
```

Out[112…
```
<AxesSubplot:xlabel='24 User friendly Interface of the website', ylabel='count'>
```



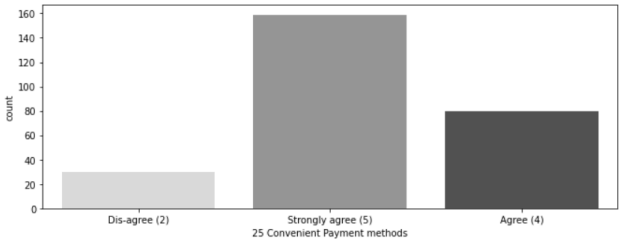
So we can see that the website should be strongly agree

In [113…
```
df[       '25 Convenient Payment methods'
].value_counts()
```

Out[113…
```
Strongly agree (5)    159
Agree (4)              80
Dis-agree (2)          30
Name: 25 Convenient Payment methods, dtype: int64
```

In [114…
```
plt.figure(figsize=(11,4))
sns.countplot(df[       '25 Convenient Payment methods'
], palette ='Greys')
```

Out[114…
```
<AxesSubplot:xlabel='25 Convenient Payment methods', ylabel='count'>
```

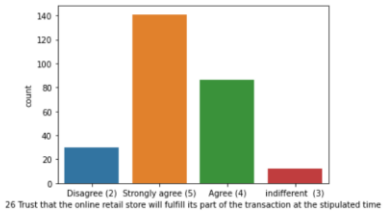


In [115…
```
df[       '26 Trust that the online retail store will fulfill its part of the transaction at the stipulated time'
].value_counts()
```

Out[115…
```
Strongly agree (5)    141
Agree (4)              86
Disagree (2)           30
indifferent  (3)       12
Name: 26 Trust that the online retail store will fulfill its part of the transaction at the stipulated time, dtype: int64
```

In [116…
```
sns.countplot(df[       '26 Trust that the online retail store will fulfill its part of the transaction at the stipulated time'
])
```

Out[116…
```
<AxesSubplot:xlabel='26 Trust that the online retail store will fulfill its part of the transaction at the stipulated time', ylabel='count'>
```

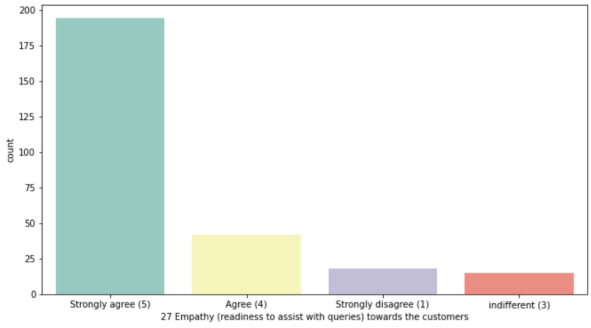


In [117…
```
df[       '27 Empathy (readiness to assist with queries) towards the customers'
].value_counts()
```

Out[117…
```
Strongly agree (5)       194
Agree (4)                 42
Strongly disagree (1)     18
indifferent (3)           15
Name: 27 Empathy (readiness to assist with queries) towards the customers, dtype: int64
```

In [118…
```
plt.figure(figsize=(11,6))
sns.countplot(df[       '27 Empathy (readiness to assist with queries) towards the customers'
], palette='Set3')
```

Out[118…
```
<AxesSubplot:xlabel='27 Empathy (readiness to assist with queries) towards the customers', ylabel='count'>
```



Coclusion:-from the above EDA we can decide that

1. We need to look at how to grow children(age below 21) and older(age above 51) customers.
2. The number of new customers is not increasing.

3.The customer abandon his/her “bag” or “shoping cart “ for better alternative offer.

4. We can see from the above data visualization that the number of male customers is half of female customers so we need to look at how to increase the number of male customers.
5. We can see that many customers are not able to decide whether their original role has been fulfilled.So we have to improve it.
6. We can see that many customers are not able to decide whether their social statuse has been enhancing.So we have to improve it.