CUSTOMER RETENTION PROJECT DATASET ANALYSIS

FLIP ROBO

Prepared by -SHREYASI SAHA

Data Science Intern at Flip Robo
Technologies

SME: **Khushboo Garg**

Problem Statement

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.
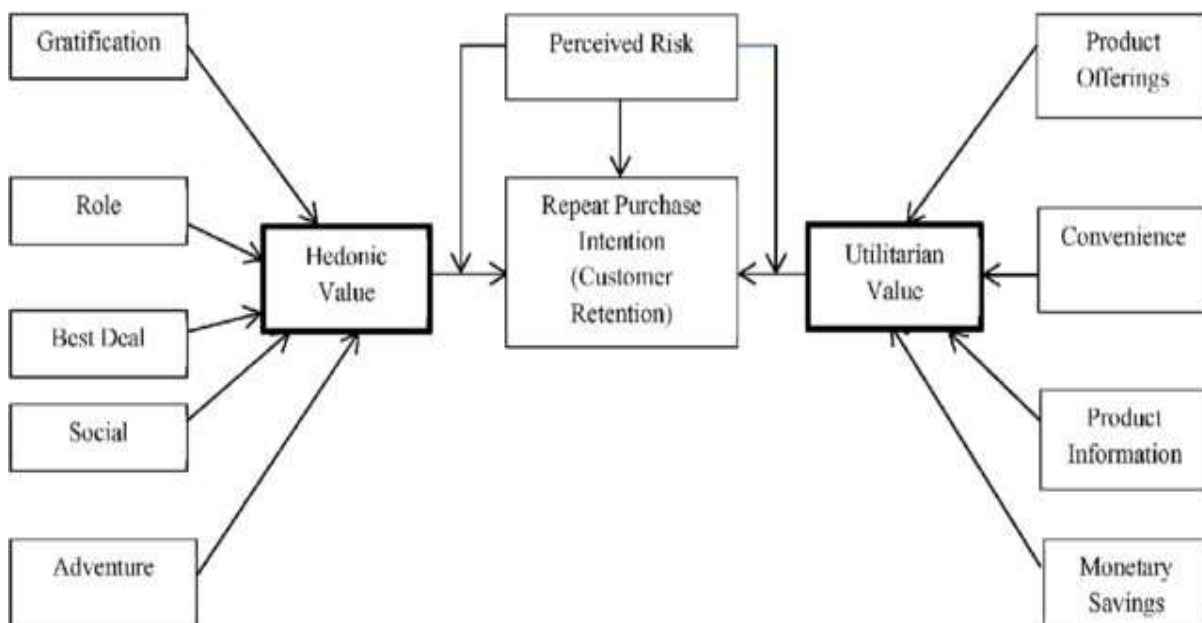
In the above use case diagram, we can see that the Repeat Purchase Intention basically our Customer Retention strategy relies on Hedonic value and Utilitarian value. Also, we see that there are perceived risks affecting the purchase and repurchase intentions of our customers. The Hedonic value has 5 major parts such as gratification, role, best deal, social aspect and adventure feeling criterions. Whereas in Utilitarian value we have product offerings, convenience, product information and monetary savings.

Motivation for the Problem Undertaken

Our main objective of doing this project is to analyze whether the users are shopping products from e-commerce websites. How did they give feedback to these websites on the basis of several positive and negative factors and also the details of the users on the basis of factors like age, gender, city etc.

Benefits of Customer Retention:

1. Retention is cheaper than acquisition
   - While the old adage about "it costs five times as much to acquire a new customer" may not be accurate in every case, the basic principle is spot on: it's more cost-effective to keep someone in the fold than to bring in new customers.
   - Even still, if it's data you want, there has been plenty of research into acquisition vs retention, and every one of them has come back with the economics favoring retention as the more economically viable focus.
   - One caveat though: retention is cheaper than acquisition, but it isn't necessarily easier.
2. Loyal customers are more profitable
   - Not only is loyalty cheaper, it has better returns. According to research, engaged consumers buy 90% more frequently, spend 60% more per transaction and are five times more likely to indicate it is the only brand they would purchase in the future.

- On average, they're delivering 23% more revenue and profitability over the average customer.
- While loyal customers are more profitable, don't take their loyalty for granted.
- They'll be more open to price increases, but be cautious not to raise prices simply to see how long they'll stick around.
- Consider the flipside: "Actively disengaged" customers (people who oppose the brand and may be actively spreading that opinion) can cost a brand 13% of its revenue.

3. Your brand will stand out from the crowd
- Put your consumer hat on, and consider how many brands you interact with that actually seem to value your patronage.
- You can probably only think of one or two.
- Most brands focus on acquisition, which makes the retention-centric among us stand out even more.
- People see around 10,000 marketing exposures a day, but only engage with a few of them.
- The ones that earn continual engagement are those with whom they feel an emotional connection with on some level.
- Forget a unique selling proposition; the best brands have a unique retention proposition.

4. You'll earn more word-of-mouth referrals
- Your loyal customers will be your best source of new business.
- Despite all the efforts into online and mobile marketing and social media, people are still most strongly influenced by referrals from friends and family.
- Millennials in particular will spread the word of a brand's exploits: 90% share their brand preferences online.

5. Engaged Customers Provide More Feedback
- Feedback is critical to the success of any business.
- Customers who provide feedback are often willing to give brands the benefit of the doubt.
- They're telling you how to earn their business repeatedly. As research has shown, people who have complained and seen their issue resolved are 84% less likely to decrease their spend.
- Need help dealing with the customers who are providing nasty feedback?

6. Customers will explore your brand
- That's a nice way of saying you'll be able to sell them more stuff.
- Once a brand has proven itself with one product or service, customers are six times more likely to say they would try a new product or service from the brand as soon as it becomes available.
- That's not just valuable for sales, but these folks can be utilized to help with #5 above as beta testers - a critical element in product development.

7. Loyal Customers are more forgiving
   - An Accenture study states over $1.6 trillion is lost each year due to customers bailing after a poor service experience.
   - We've gone so far as to claim that it's the top reason people will ditch a brand.
   - But customers who consider themselves loyal will let some misdeeds slide - just don't let it happen too often.
8. Customers will welcome your marketing
   - No one likes being marketed to.
   - Except for loyal customers!
   - Those folks are four times more likely to say they "appreciate when this brand reaches out to me" and seven times more likely to "always respond to this brand's promotional offers."

9. You earn wiggle room to try new things
   - Loyalty is fickle, so too many changes could chase people away.
   - But once you've established a core base of proven customers, your brand can expand its boundaries.
   - Maybe it's new messaging or a new product line, or even a new logo. The bottom line is as long as you maintain the basic premises that keep people in your corner; they'll stick with you through thin and thin.
   - In fact, some of them will be excited to see what you can do.
   - Existing customers are 50% more likely to try new products, according to a study.

Client Lifecycle Stages
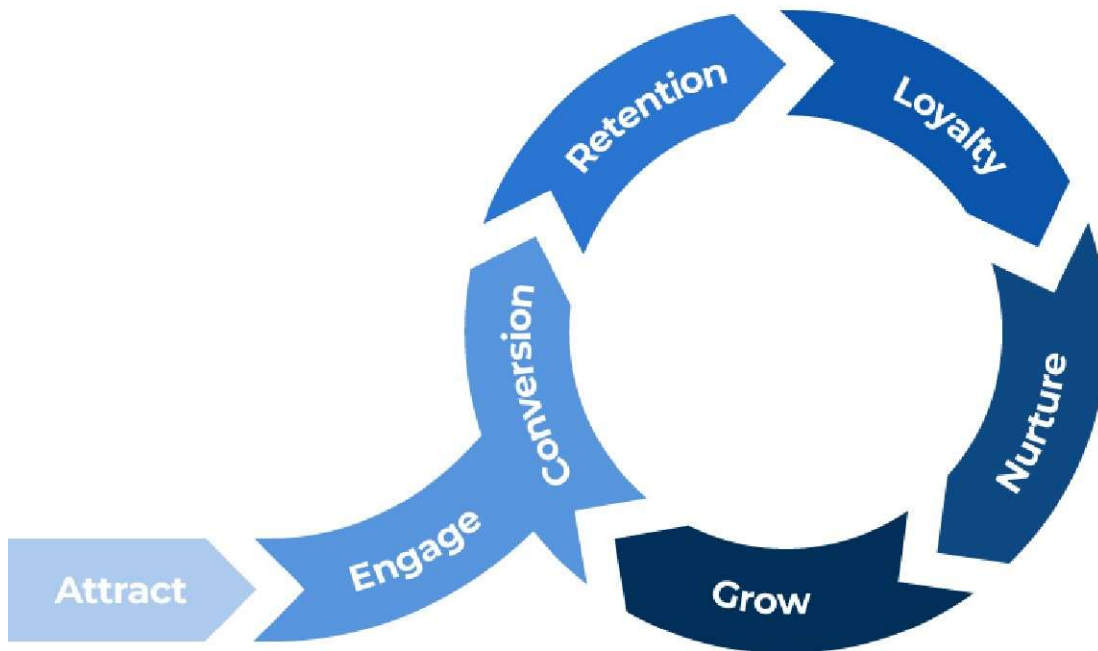
Lifetime revenue is the end goal, not just today's revenue.
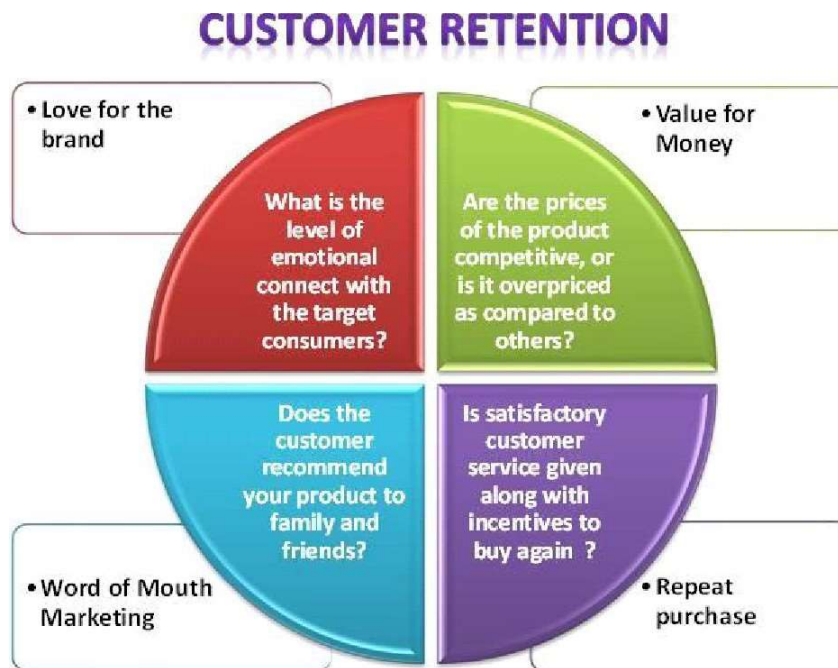
Need for Customer Retention:

Keeping current customers happy is generally more cost-effective than acquiring first-time customers. According to the Harvard Business Review, acquiring a new customer can be five to 25 times more expensive than holding on to an existing one.

Companies don't need to spend big on marketing, advertising, or sales outreach. It is easier to turn existing customers into repeating ones, since they already trust your brand from previous purchases. New customers, however, often require more convincing when it comes to that initial sale.

Customer loyalty won't just give companies repeat business. Loyal customers are more likely to give free recommendations to their colleagues, friends, and family. Creating that cycle of retained customers and buzz marketing is one way a company can cultivate customer loyalty for long-term success.

Improving customer retention means improving the customer experience. In fact, 77 percent of customers surveyed in a 2021 Customer Experience Trend Report being more loyal to a company that offers a good customer experience if they have an issue. 72 percent are willing to spend more from a

company the offers good customer experiences. And 50 percent say that customer experience is more important to them now compared to a year ago.

CUSTOMER RETENTION

- Love for the brand

What is the level of emotional connect with the target consumers?

- Value for Money

Are the prices of the product competitive, or is it overpriced as compared to others?

Does the customer recommend your product to family and friends?

Is satisfactory customer service given along with incentives to buy again ?

- Word of Mouth Marketing

- Repeat purchase

Since the cost of getting a new customer is an estimated five to ten times more than keeping an old one, nurturing loyal customers is a powerful strategy that helps businesses grow.

Dataset Details:

First, I imported all the necessary libraries and dependencies to create a detailed data analysis in Python.

```
In [1]:   1  #Import libraries required for data analysis
          2  import pandas as pd
          3  import numpy as np
          4  import seaborn as sns
          5  import matplotlib.pyplot as plt
          6  %matplotlib inline
          7  import warnings
          8  warnings.filterwarnings("ignore")
```

Then I loaded the dataset from the excel file which was provided by Flip Robo Technologies.

```
In [2]:   1  df = pd.read_excel('customer_retention_dataset.xlsx',sheet_name='datasheet')
          2
          3  df
```

Out[2]:

	1Gender of respondent	2 How old are you?	3 Which city do you shop online from?	4 What is the Pin Code of where you shop online from?	5 Since How Long You are Shopping Online ?	6 How many times you have made an online purchase in the past 1 year?	7 How do you access the internet while shopping on-line?	8 Which device do you use to access the online shopping?	9 What is the screen size of your mobile device?	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
266	Female	41-50 yaers	Bangalore	560010	2-3 years	Less than 10 times	Mobile internet	Laptop	Others	Window/windows Mobile	...	Amazon.in	Snapdeal.com	Amazon.in
267	Female	Less than 20 years	Solan	173229	2-3 years	Less than 10 times	Wi-Fi	Smartphone	5.5 inches	Android	...	Amazon.in	Amazon.in, Myntra.com, Snapdeal.com	Amazon.in
268	Female	41-50	Ghaziabad	201009	2-3 years	31-40 times	Mobile Internet	Smartphone	5.5 inches	Android	...	Amazon.in	Amazon.in	Amazon.in

Exploratory Data Analysis (EDA):

First I made sure to see the shape of the dataset i.e how many number of rows and columns are there in the dataset.Next I will see the column names of the dataset and after that I will see the datatype of the dataset.The next cell which I have shown here is information about the dataset.

```
In [3]:    1  df.shape
Out[3]: (269, 71)
```

Out[4]: Index(['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\t
       '10 What is the operating system (OS) of your device?\t\t\t\t
       '11 What browser do you run on your device to access the website?\t\t\t
       ',
       '12 Which channel did you follow to arrive at your favorite online store for th
       ',
       '13 After first visit, how do you reach the online retail store?\t\t\t\t
       ',
       '14 How much time do you explore the e- retail store before making a purchase c
       ',
       '15 What is your preferred payment Option?\t\t\t\t\t\t
       ',
       '16 How frequently do you abandon (selecting an items and leaving without makir
\t\t   ',
       '17 Why did you abandon the "Bag", "Shopping Cart"?\t\t\t\t\t
       ',
       '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 prc
       '20 Complete information on listed seller and product being offered is importar
       '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 transactior
       '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, onli
       '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 dec
       '34 Gaining access to loyalty programs is a benefit of shopping online',
       '35 Displaying quality Information on the website improves satisfaction of cust
       '36 User derive satisfaction while shopping on a good quality website or applic
       '37 Net Benefit derived from shopping online can lead to users satisfaction',

' 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?'],
  dtype='object')
',

In [5]:   1 df.dtypes

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

```
In [6]:    1  #Getting Information about my dataset
           2
           3  df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 269 entries, 0 to 268
Data columns (total 71 columns):
 #   Column
Non-Null Count  Dtype
---  ------
--------------  -----
 0   1Gender of respondent
269 non-null    object
 1   2 How old are you?
269 non-null    object
 2   3 Which city do you shop online from?
269 non-null    object
 3   4 What is the Pin Code of where you shop online from?
269 non-null    int64
 4   5 Since How Long You are Shopping Online ?
269 non-null    object
 5   6 How many times you have made an online purchase in the past 1 year?
269 non-null    object
```

After checking the information about the dataset, I will proceed to see a statistical summary of the data using the describe function and I will also be checking for any missing values in the dataset.

Numerical variable:

```
In [7]:    1  df.describe()
```

Out[7]:

	4 What is the Pin Code of where you shop online from?
count	269.000000
mean	220465.747212
std	140524.341051
min	110008.000000
25%	122018.000000
50%	201303.000000
75%	201310.000000
max	560037.000000

Categorical variables:

```
In [8]:    1  df.describe(include=['O'])
```

Out[8]:

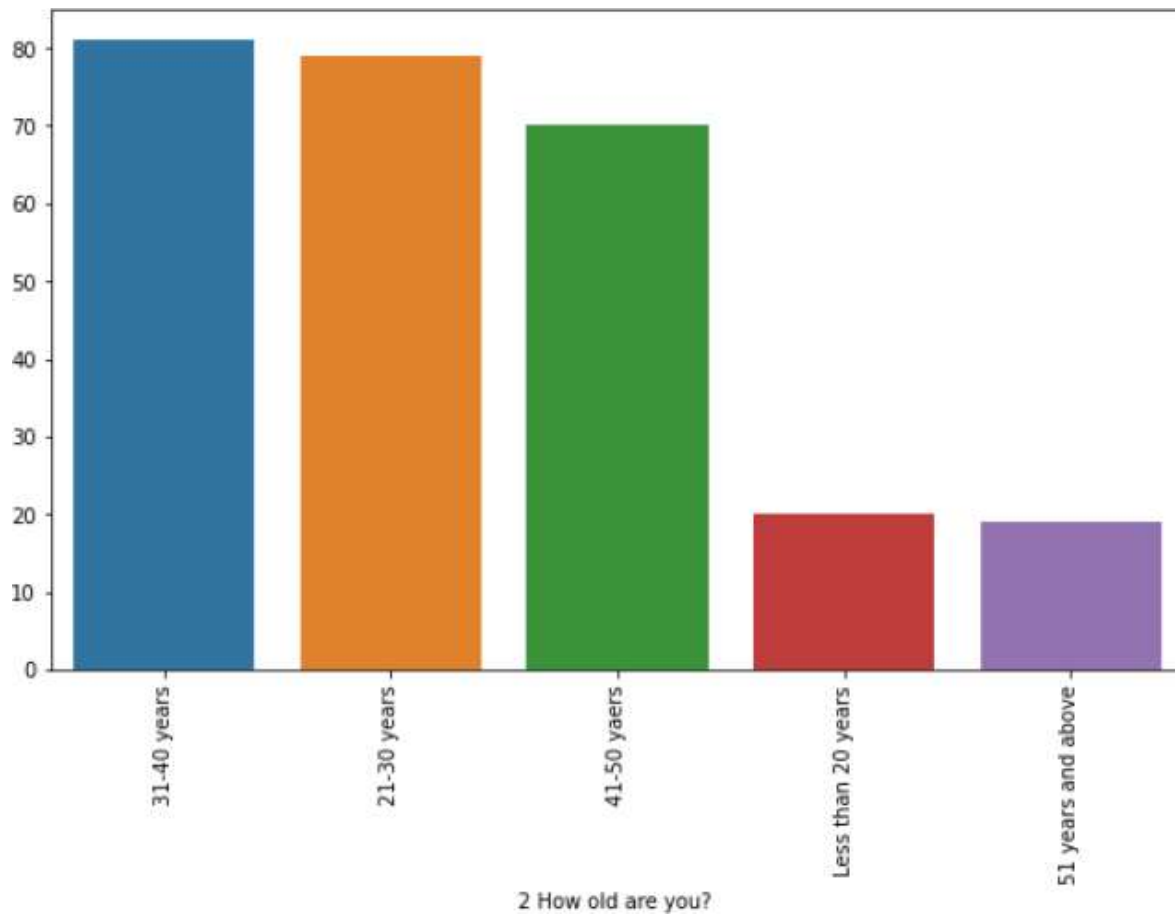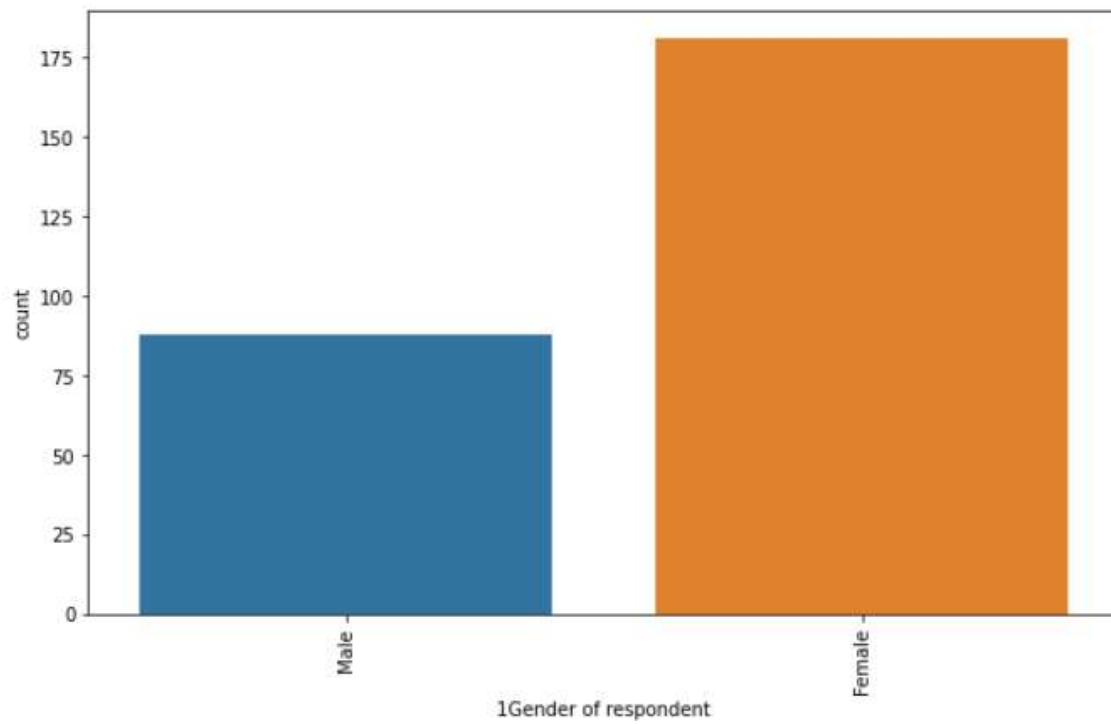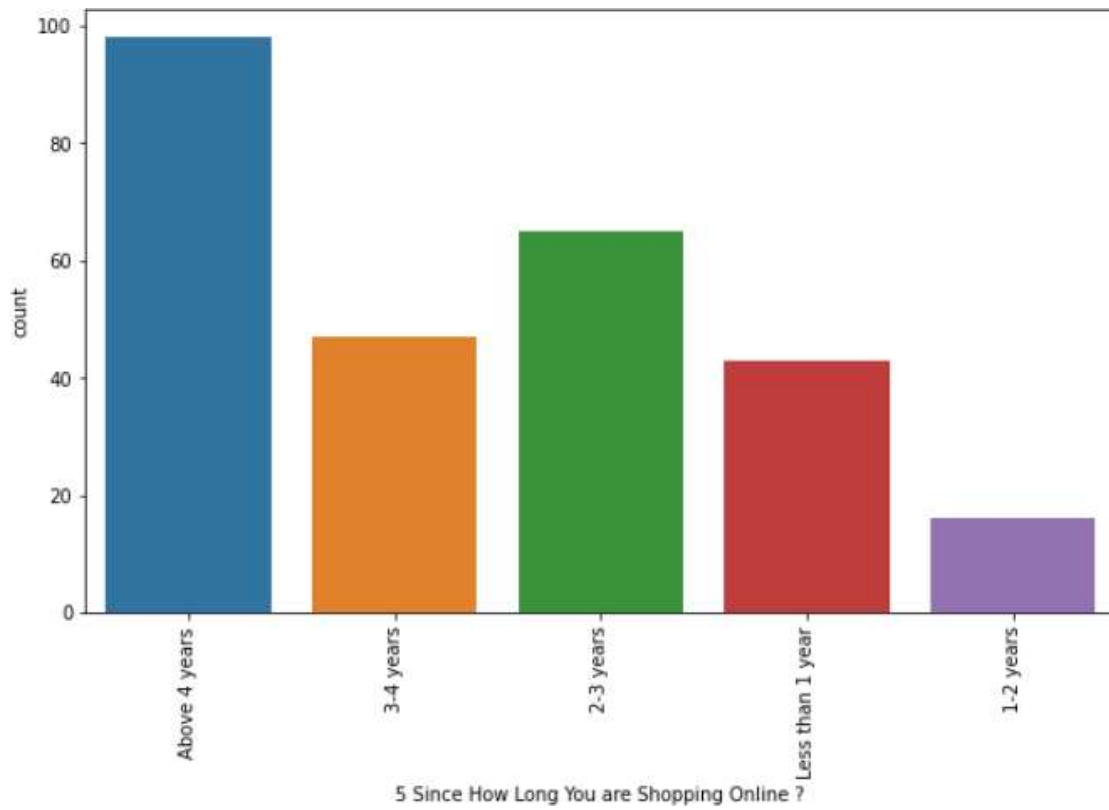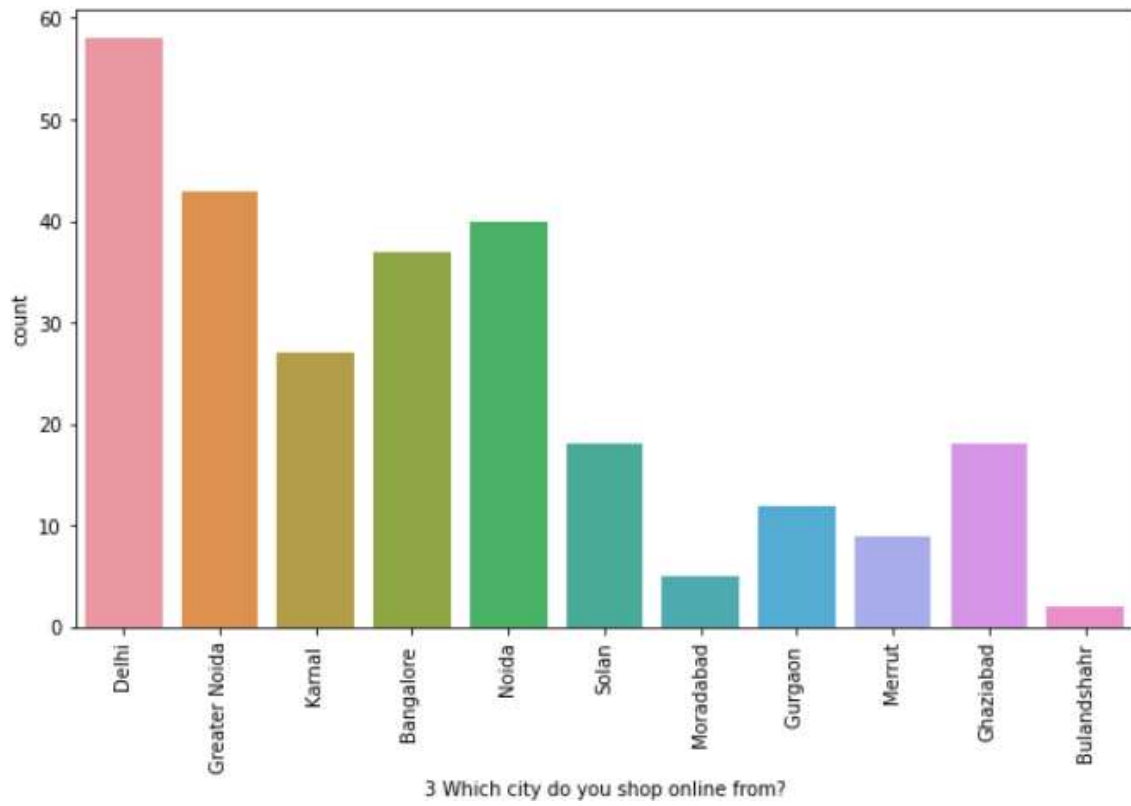|	1Gender of respondent	2 How old are you?	3 Which city do you shop online from?	5 Since How Long You are Shopping Online ?	6 How many times you have made an online purchase in the past 1 year?	7 How do you access the internet while shopping on-line?	8 Which device do you use to access the online shopping?	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	11 What browser do you run on your device to access the website? \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)	(r
count	269	269	269	269	269	269	269	269	269	269	...	269	269	269	
unique	2	5	11	5	6	4	4	4	3	4	...	10	10	8	
top	Female	31-40 years	Delhi	Above 4 years	Less than 10 times	Mobile internet	Smartphone	Others	Window/windows Mobile	Google chrome	...	Amazon.in	Amazon.in, Flipkart.com	Myntra.com	M
freq	181	81	58	98	114	142	141	134	122	216	...	57	60	75	

4 rows × 70 columns

```
In [9]:    1  df.isnull().sum().sum()

Out[9]:  0
```

In the dataset there are no missing values present.

In the dataset,there are no missing values present.

Now I will check for distinct elements ......

```
In [10]:    1  df.nunique()

Out[10]:  1Gender of respondent                                                2
          2 How old are you?                                                  5
          3 Which city do you shop online from?                              11
          4 What is the Pin Code of where you shop online from?              39
          5 Since How Long You are Shopping Online ?                          5
                                                                             ..
          Longer delivery period                                             6
          Change in website/Application design                               7
          Frequent disruption when moving from one page to another           8
          Website is as efficient as before                                  8
          Which of the Indian online retailer would you recommend to a friend?  8
          Length: 71, dtype: int64
```

Here I have completed the structural part or text analysis part of the dataset.I will proceed to data visualization part.

What is Data Visualization?
Data visualization is defined as a graphical representation that contains the information and the data.

Benefits of Good Data Visualization?
Data visualization is another technique of visual art that grabs our interest and keeps our main focus on the message captured with the help of eyes.

Different Types of Analysis for Data Visualization are:
1. Univariate Analysis: In the univariate analysis, we will be using a single feature to analyze almost all of its properties.
2. Bivariate Analysis: When we compare the data between exactly 2 features then it is known as bivariate analysis.
3. Multivariate Analysis: In the multivariate analysis, we will be comparing more than 2 variables.

Univariate Analysis:

The term univariate analysis refers to the analysis of one variable. You can remember this because the pre1x "uni" means "one."

There are three common ways to perform univariate analysis on one variable:

1. Summary statistics – Measures the center and spread of values.

2. Frequency table – Describes how often diIerent values occur.

3. Charts – Used to visualize the distribution of values.

For Univariate Analysis ,I will explore categorical variables that has 70 columns which I will be using count plot to analyze properly.

COUNT PLOT:

seaborn.countplot() method is used to Show the counts of observations in each categorical bin using bars.

```
In [11]:    1  categorical_variables = df.select_dtypes(include=['O'])
            2  categorical_variables
```

Out[11]:

	1Gender of respondent	2 How old are you?	3 Which city do you shop online from?	5 Since How Long You are Shopping Online ?	6 How many times you have made an online purchase in the past 1 year?	7 How do you access the internet while shopping on-line?	8 Which device do you use to access the online shopping?	9 What is the screen size of your mobile device? \t\t\t\t\t	10 What is the operating system (OS) of your device? \t\t\t\t	11 What browser do you run on your device to access the website? \t\t\t	...	Longer time to get logged in (promotion, sales period)	Longer time in displaying graphics and photos (promotion, sales period)	La declaratio of pri (promotio sales perio
0	Male	31-40 years	Delhi	Above 4 years	31-40 times	Dial-up	Desktop	Others	Window/windows Mobile	Google chrome	...	Amazon.in	Amazon.in	Flipkart.co
1	Female	21-30 years	Delhi	Above 4 years	41 times and above	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	Google chrome	...	Amazon.in, Flipkart.com	Myntra.com	snapdeal.co
2	Female	21-30 years	Greater Noida	3-4 years	41 times and above	Mobile Internet	Smartphone	5.5 inches	Android	Google chrome	...	Myntra.com	Myntra.com	Myntra.co
3	Male	21-30 years	Karnal	3-4 years	Less than 10 times	Mobile Internet	Smartphone	5.5 inches	IOS/Mac	Safari	...	Snapdeal.com	Myntra.com, Snapdeal.com	Myntra.co
4	Female	21-30 years	Bangalore	2-3 years	11-20 times	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	Safari	...	Flipkart.com, Paytm.com	Paytm.com	Paytm.co
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
264	Female	21-30 years	Solan	1-2 years	Less than 10 times	Mobile Internet	Smartphone	5.5 inches	Android	Opera	...	Amazon.in	Amazon.in	Amazon.
265	Female	31-40 years	Ghaziabad	1-2 years	31-40 times	Mobile Internet	Smartphone	Others	Android	Google chrome	...	Flipkart.com	Flipkart.com	Flipkart.co
266	Female	41-50 yaers	Bangalore	2-3 years	Less than 10 times	Mobile internet	Laptop	Others	Window/windows Mobile	Google chrome	...	Amazon.in	Snapdeal.com	Amazon.

I will explore the categorical variables by using count plot

```
In [12]:    1  print("Count Plot for categorical variables")
            2  def generate_count_plot():
            3      for i in categorical_variables:
            4          plt.figure(figsize=(10,6))
            5          c = sns.countplot(df[i])
            6          plt.xticks(rotation=90)
            7  generate_count_plot()
```

Count Plot for categorical variables
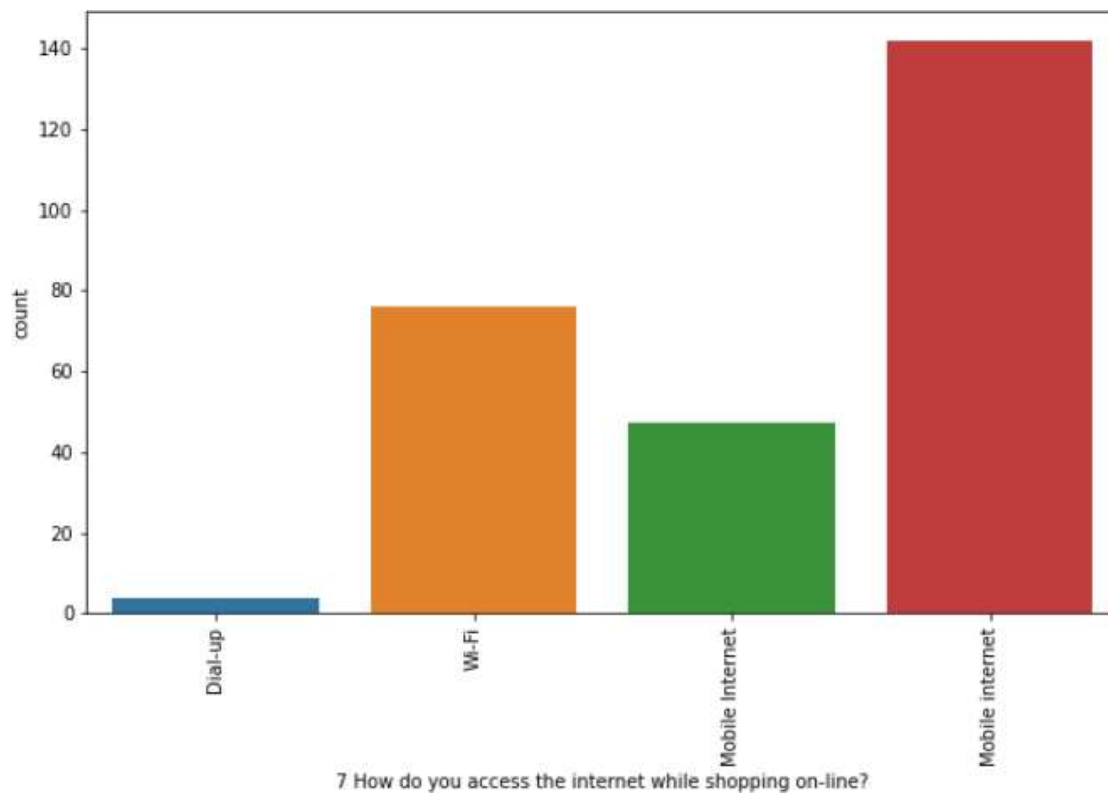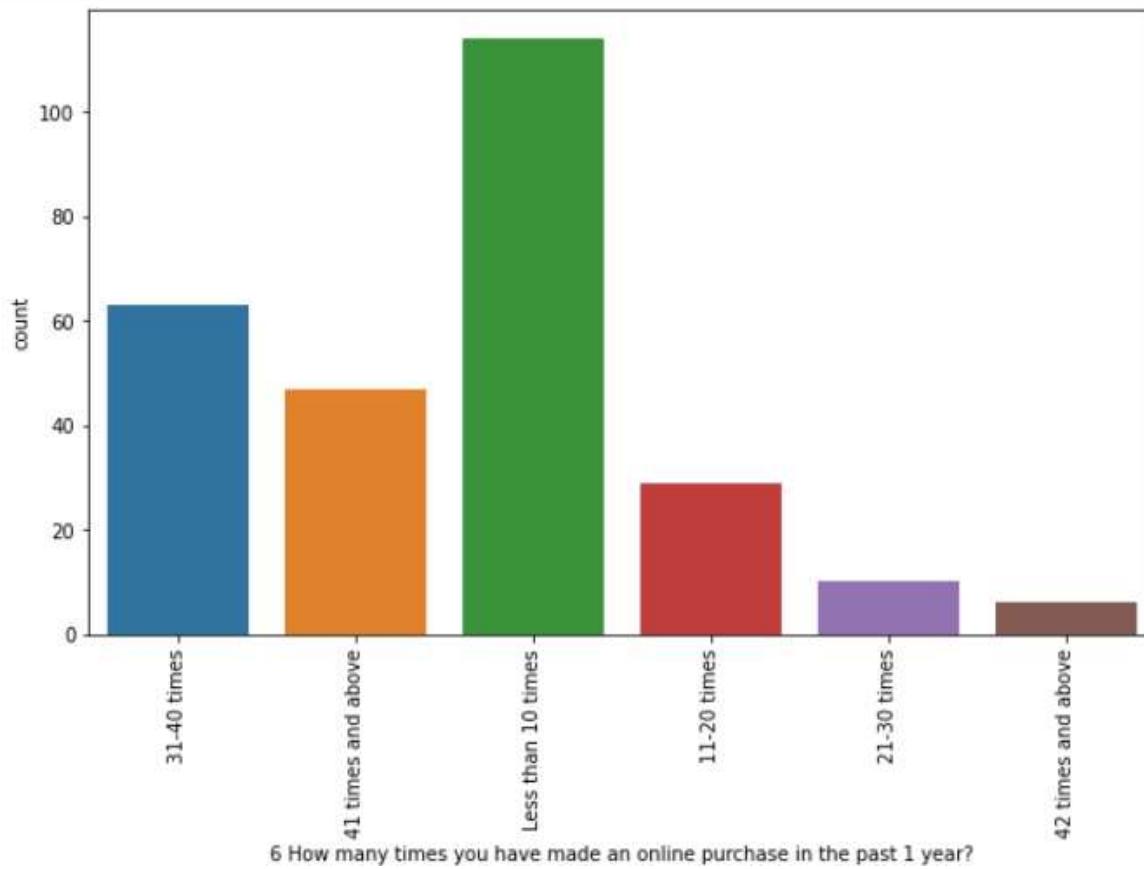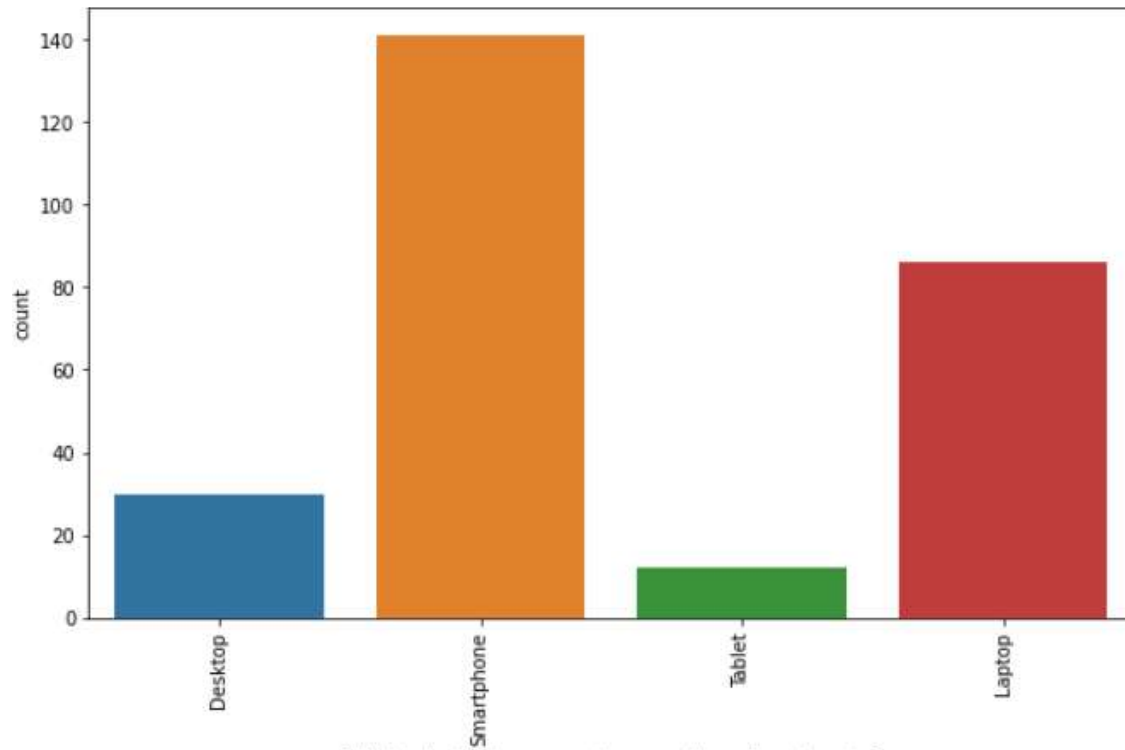
1Gender of respondent

2 How old are you?

3 Which city do 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?
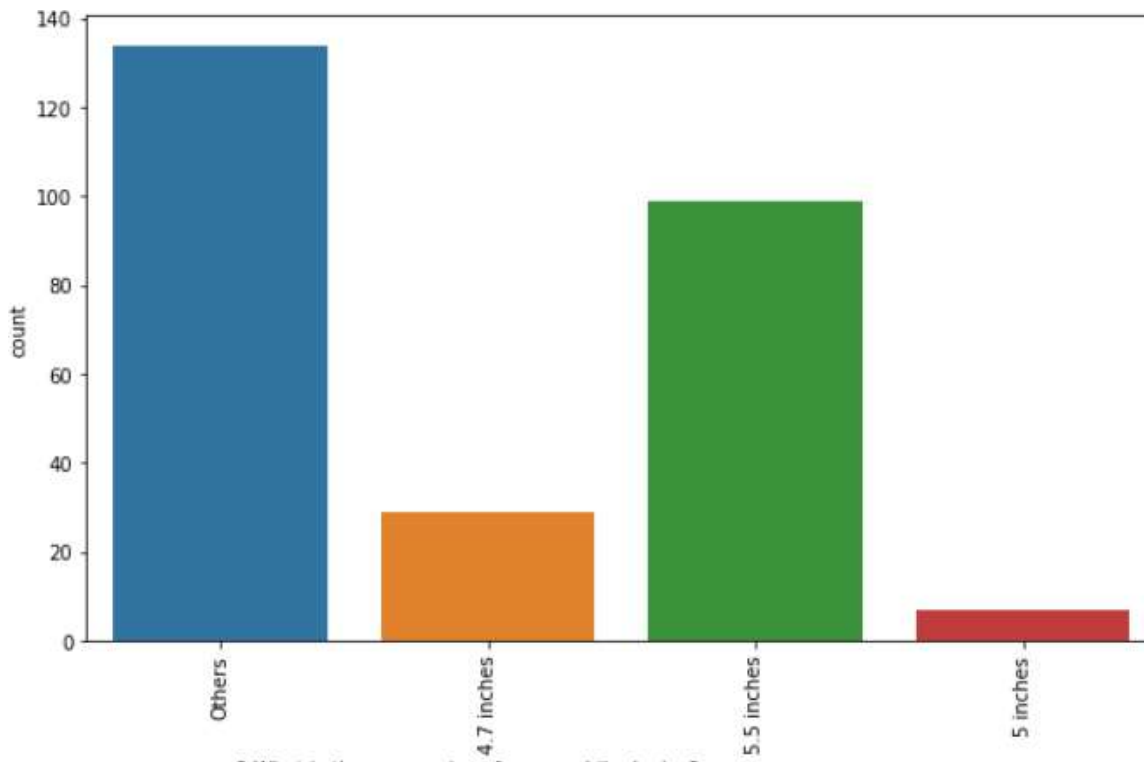
9 What is the screen size of your mobile device?
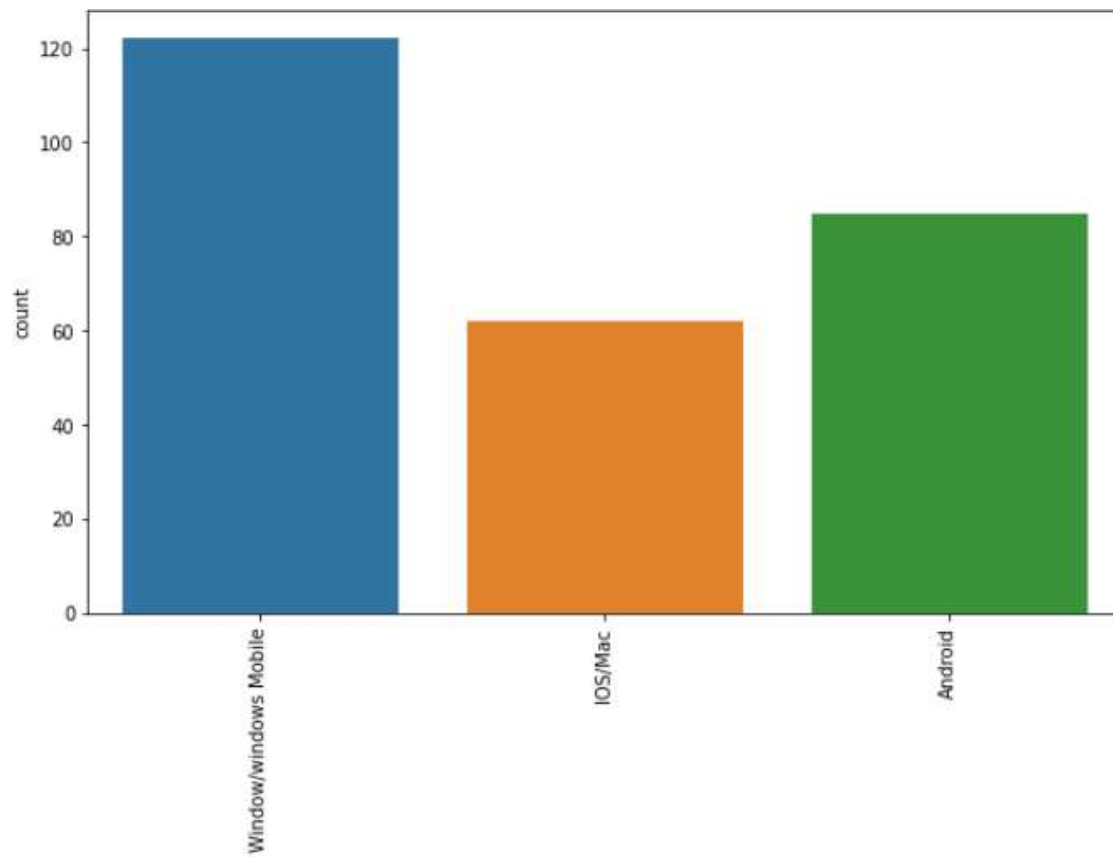
10 What is the operating system (OS) of your device?
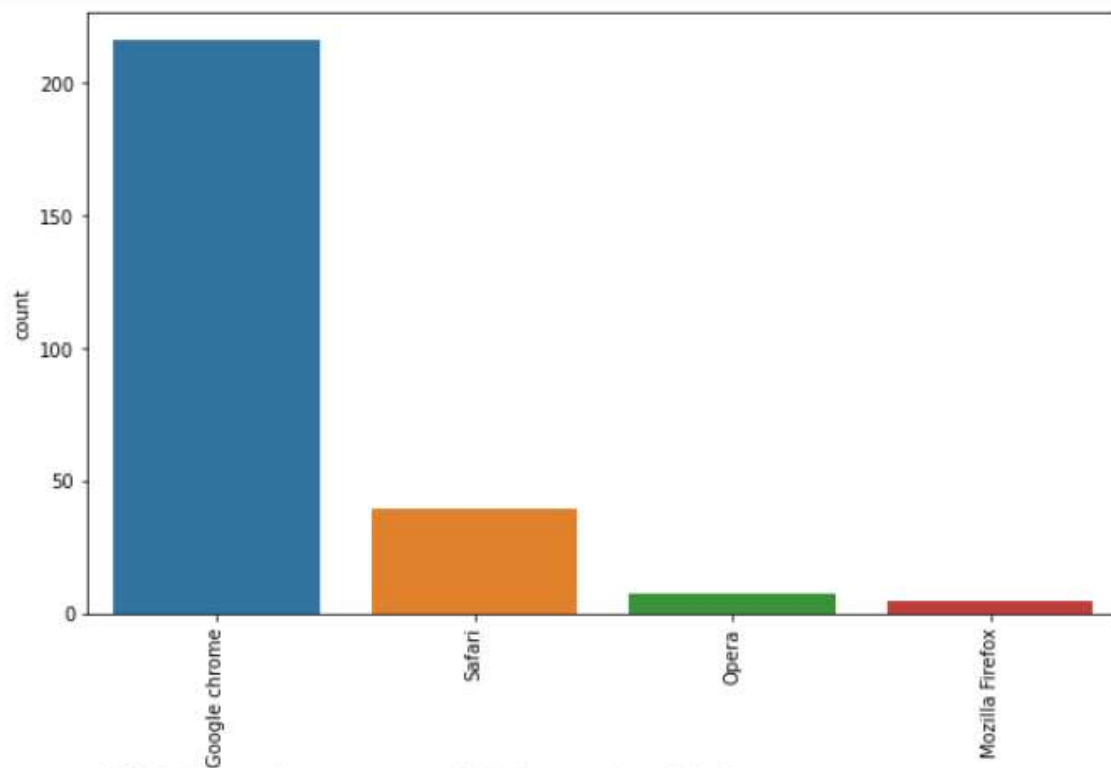
11 What browser do you run on your device to access the website?
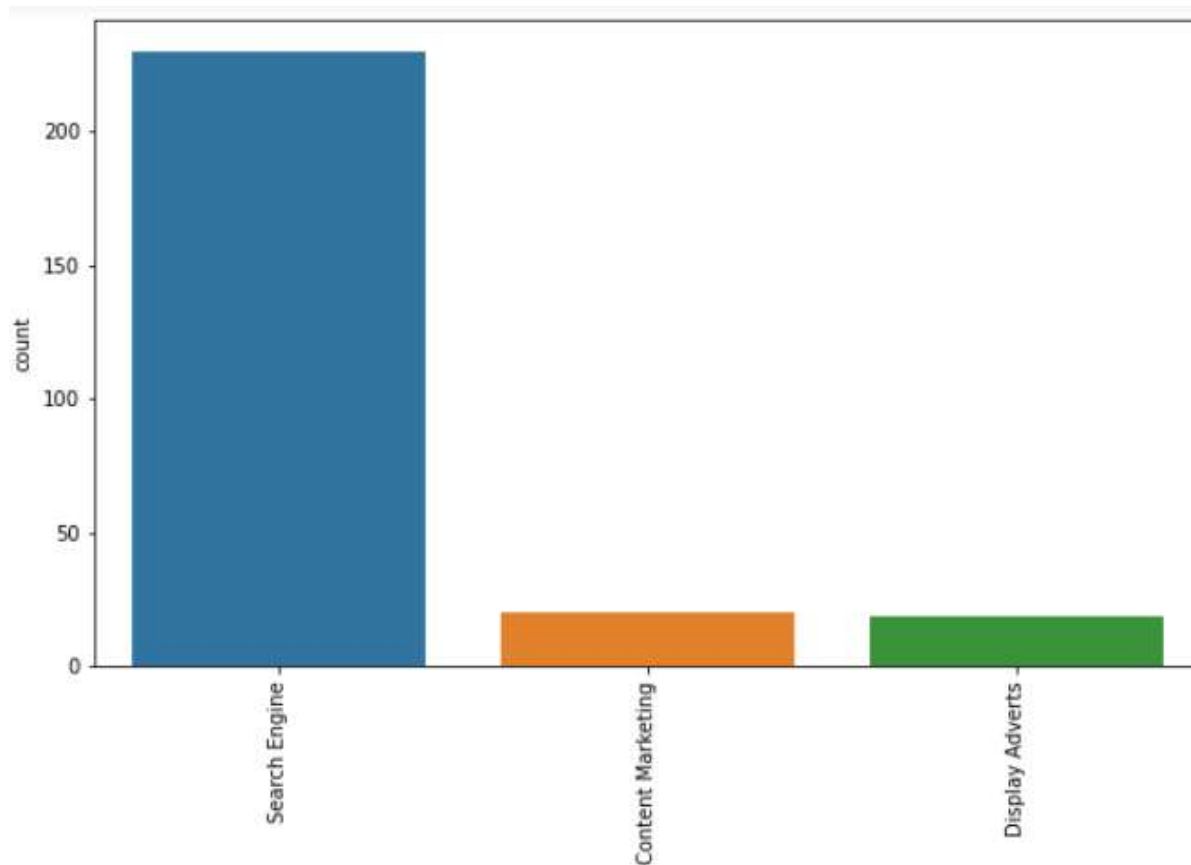
L2 Which channel did you follow to arrive at your favorite online store for the first time?
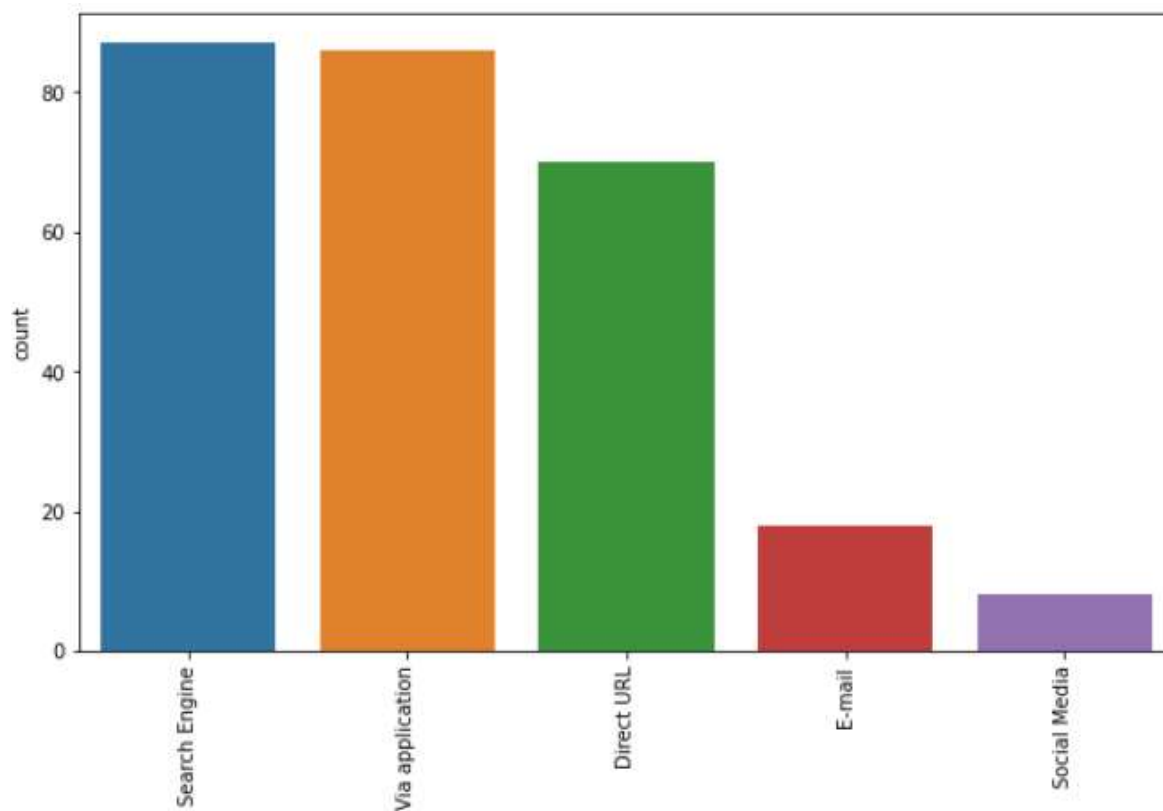
13 After first visit, how do you reach the online retail store?
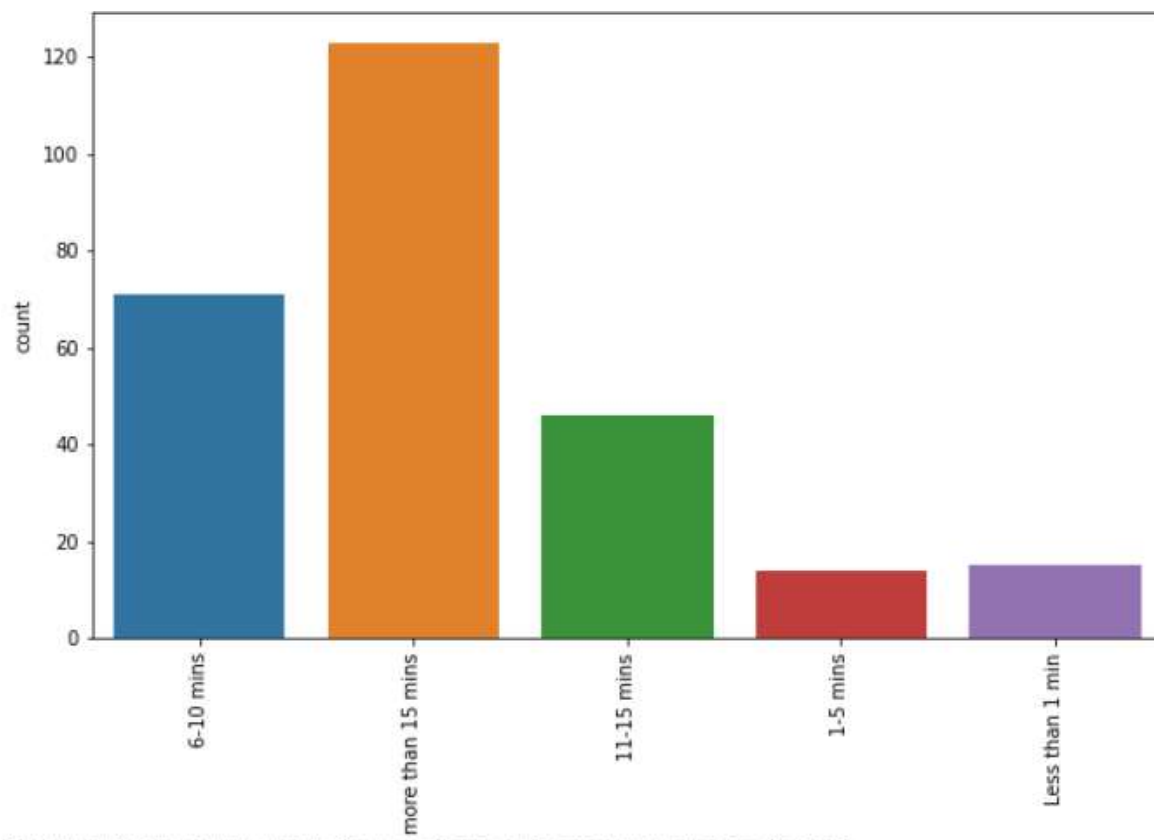
14 How much time do you explore the e- retail store before making a purchase decision?
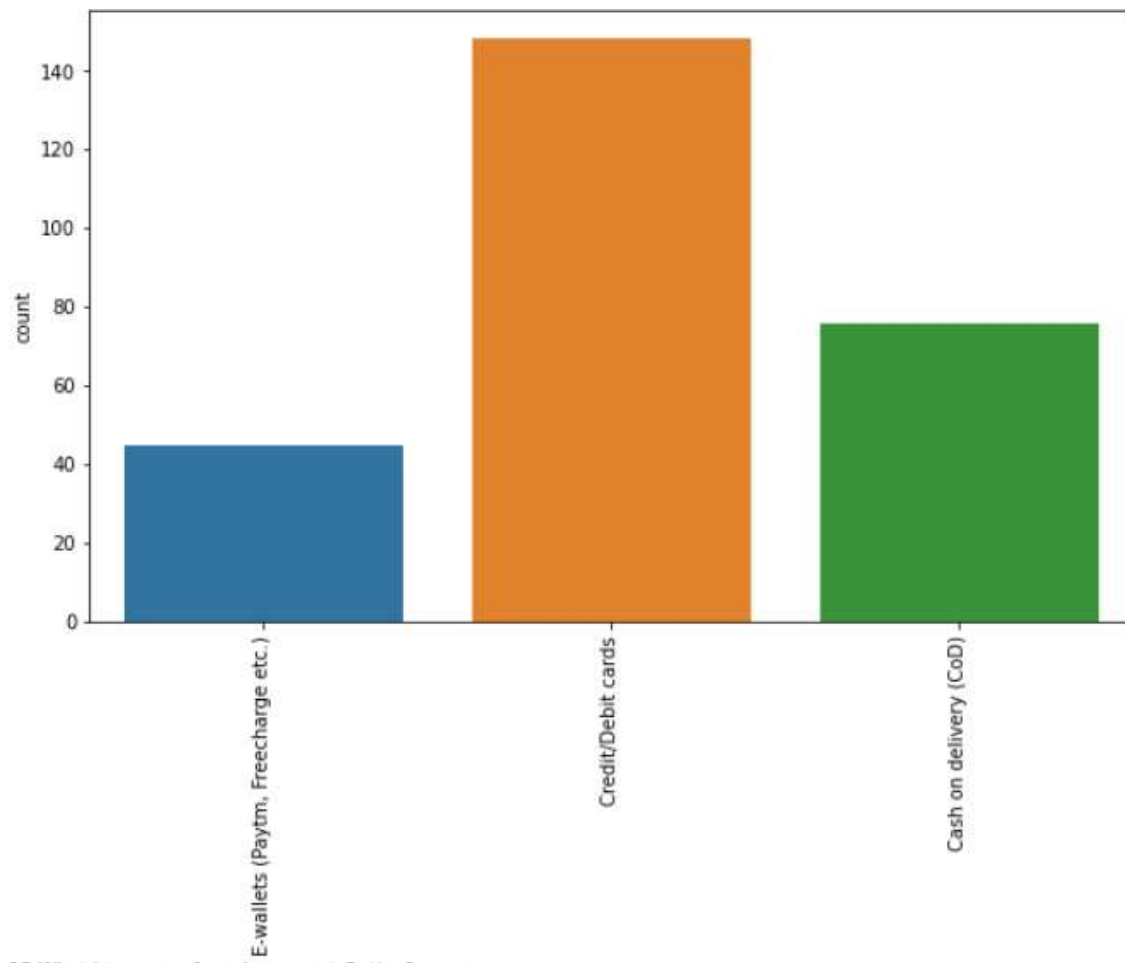
count

140

120

100

80

60

40

20

0

E-wallets (Paytm, Freecharge etc.)

Credit/Debit cards

Cash on delivery (CoD)

15 What is your preferred payment Option?
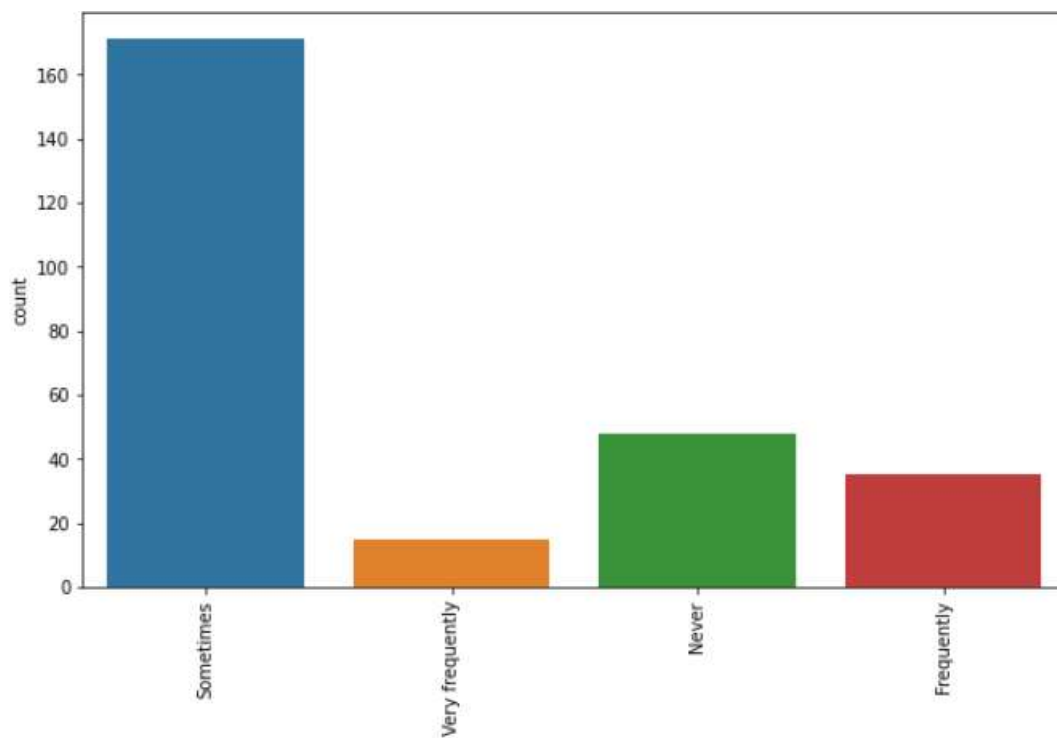
16 How frequently do you abandon (selecting an items and leaving without making payment) your shopping cart?　□□□□□□
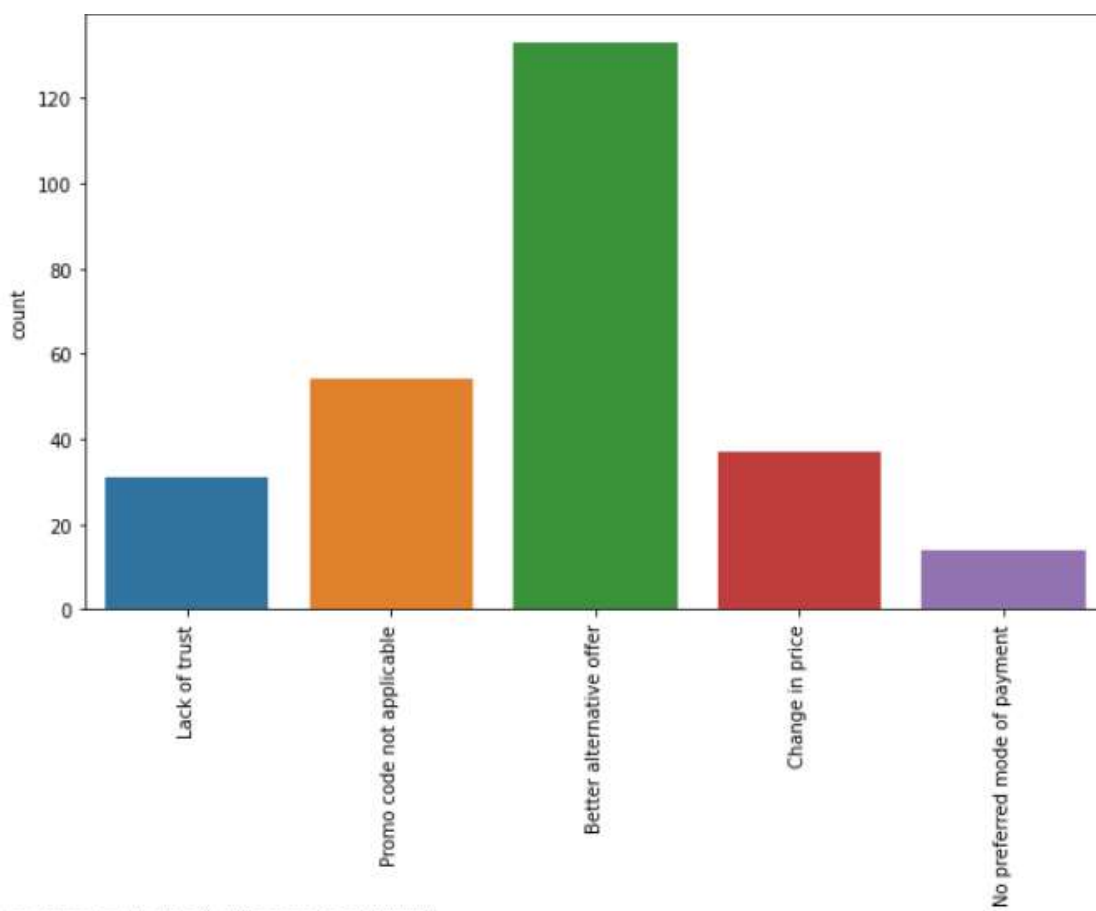
Why did you abandon the "Bag", "Shopping Cart"?　□□□□□
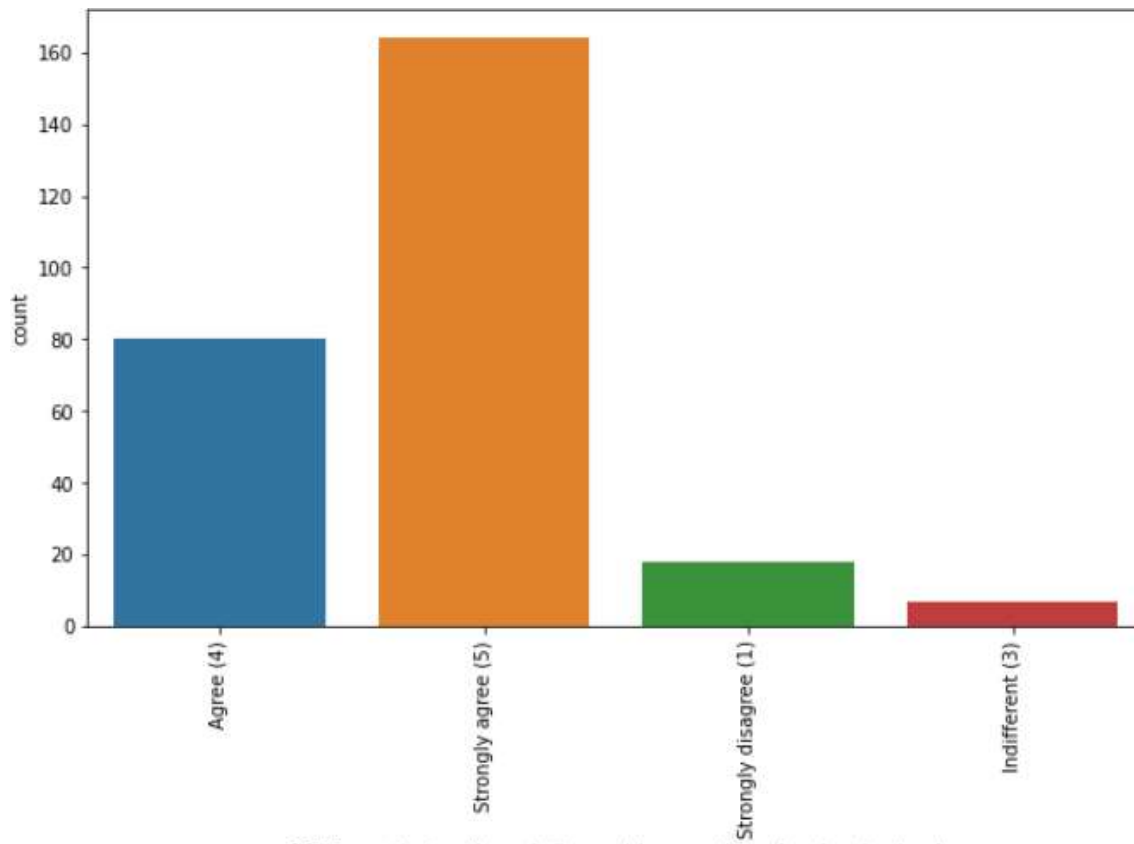
18 The content on the website must be easy to read and understand
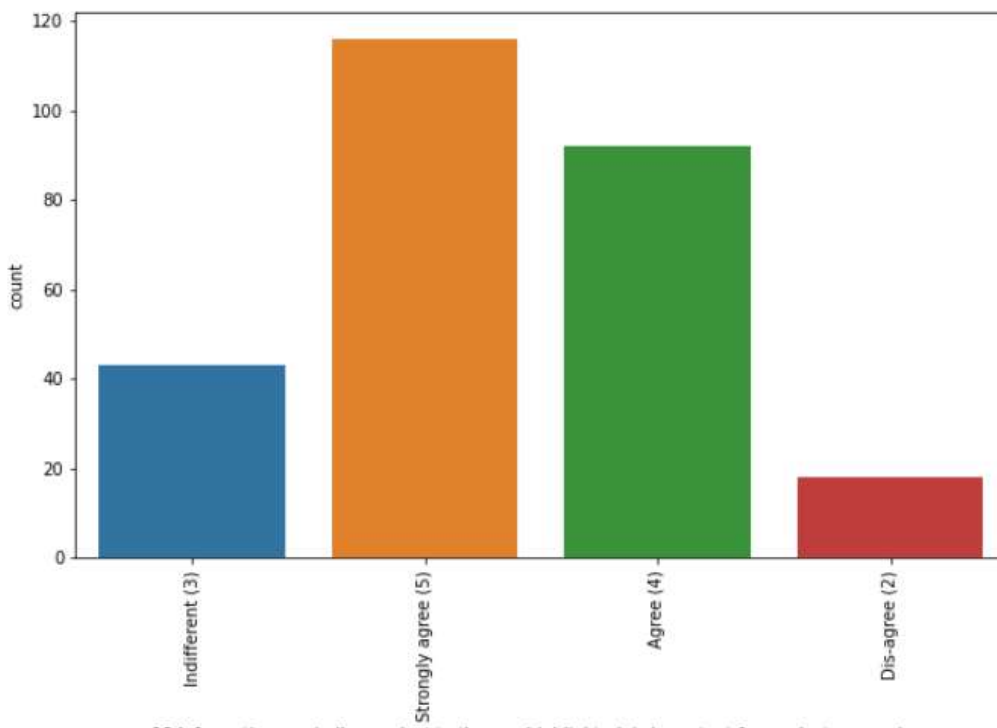
19 Information on similar product to the one highlighted is important for product comparison

20 Complete information on listed seller and product being offered is important for purchase decision.
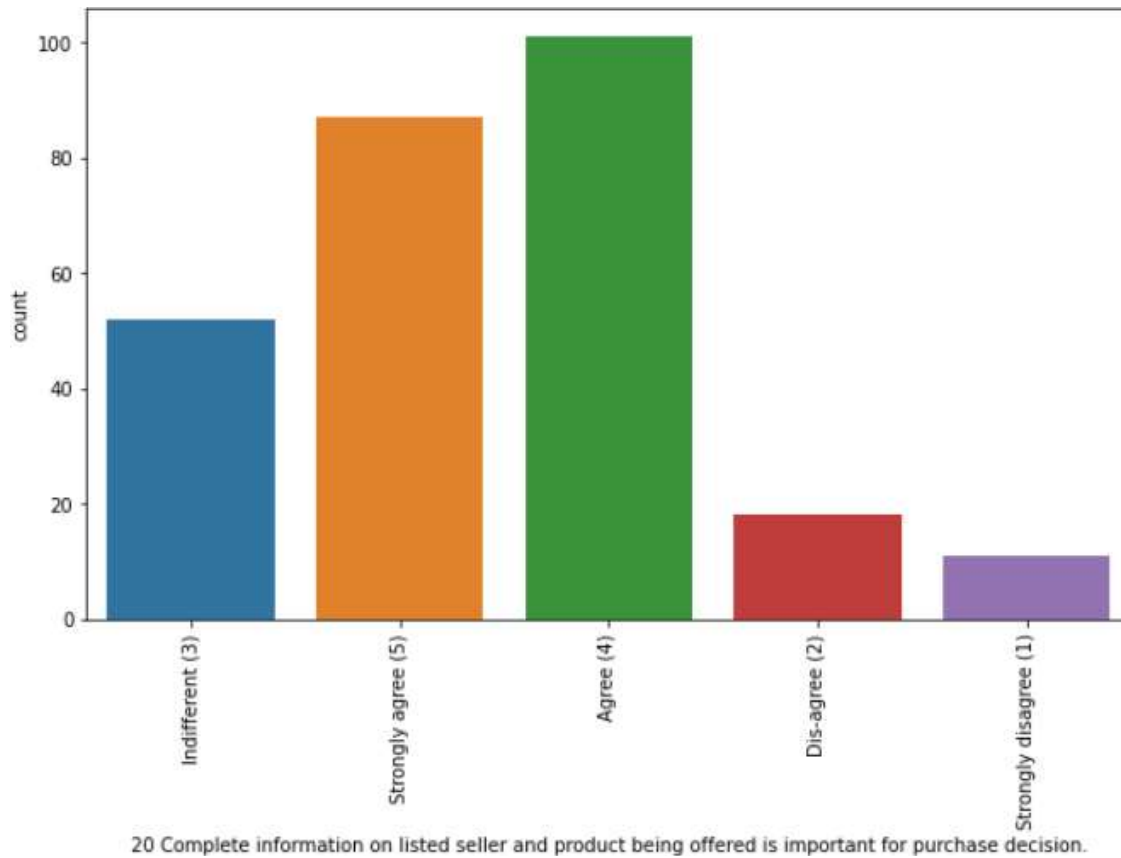
The remaining columns with count plot has been shown in the jupyter notebook.

Next part of the analysis, I will check the numerical variable i.e only 1 variable is there.

I will use the histogram to check the distribution for numerical variable.

```
In [13]:   1  #Noting numerical variable
           2  numerical_variables = df.select_dtypes(include=['number'])
           3  numerical_variables
```

Out[13]:

	4 What is the Pin Code of where you shop online from?
0	110009
1	110030
2	201308
3	132001
4	530068
...	...
264	173212
265	201008
266	560010
267	173229
268	201009

269 rows × 1 columns

Histogram:

A histogram is **a graph that shows the frequency of numerical data using rectangles**. The height of a rectangle (the vertical axis) represents the distribution frequency of a variable (the amount, or how often that variable appears).

```
In [14]:   1  #Using histogram for analysis
           2  plt.figure(figsize=(12,8))
           3  plt.hist(df['4 What is the Pin Code of where you shop online from?'],edgecolor='black',linewidth=3)
           4  plt.xlabel('Pincode')
           5  plt.show()
```
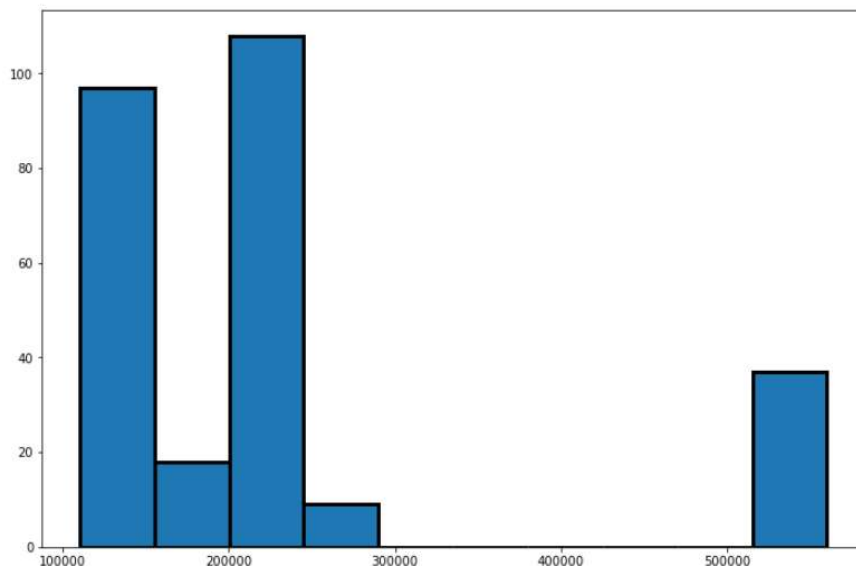
I also plotted the density plot.

```
In [15]:    1  #Density Plot
            2  plt.figure(figsize=(12,8))
            3  sns.distplot(df['4 What is the Pin Code of where you shop online from?'],color='blue')
            4  plt.xlabel('Pincode')
            5  plt.show()
```
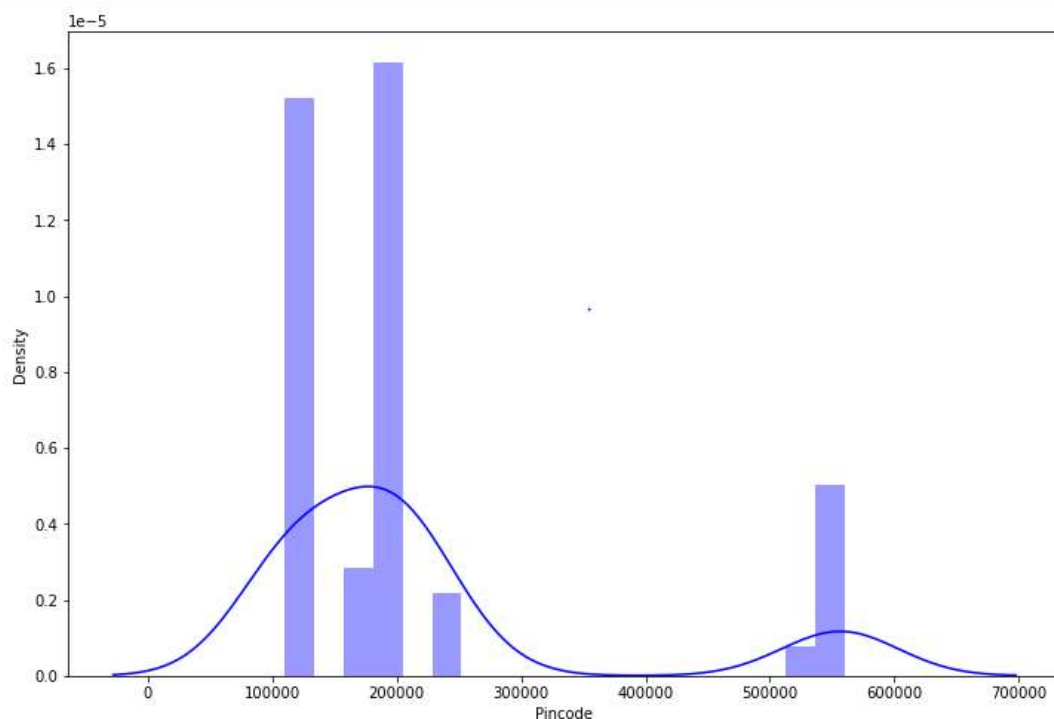

Bivariate Analysis:

I performed bivariate analysis using count plots again and changing the hue format.

Code:

```
In [16]:    1  feature = df.drop(['1Gender of respondent'],axis=1)
```

```
In [17]:    1  print("------------------------------------------Bivariate Analysis------------------------------------------"
            2  def generate_count_plot_for_bivariate_analysis():
            3      for i in feature:
            4          plt.figure(figsize=(10,6))
            5          c = sns.countplot(df[i],hue=df['1Gender of respondent'])
            6          plt.xticks(rotation=90)
            7  generate_count_plot_for_bivariate_analysis()
```

Output:

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?⬜⬜⬜⬜⬜
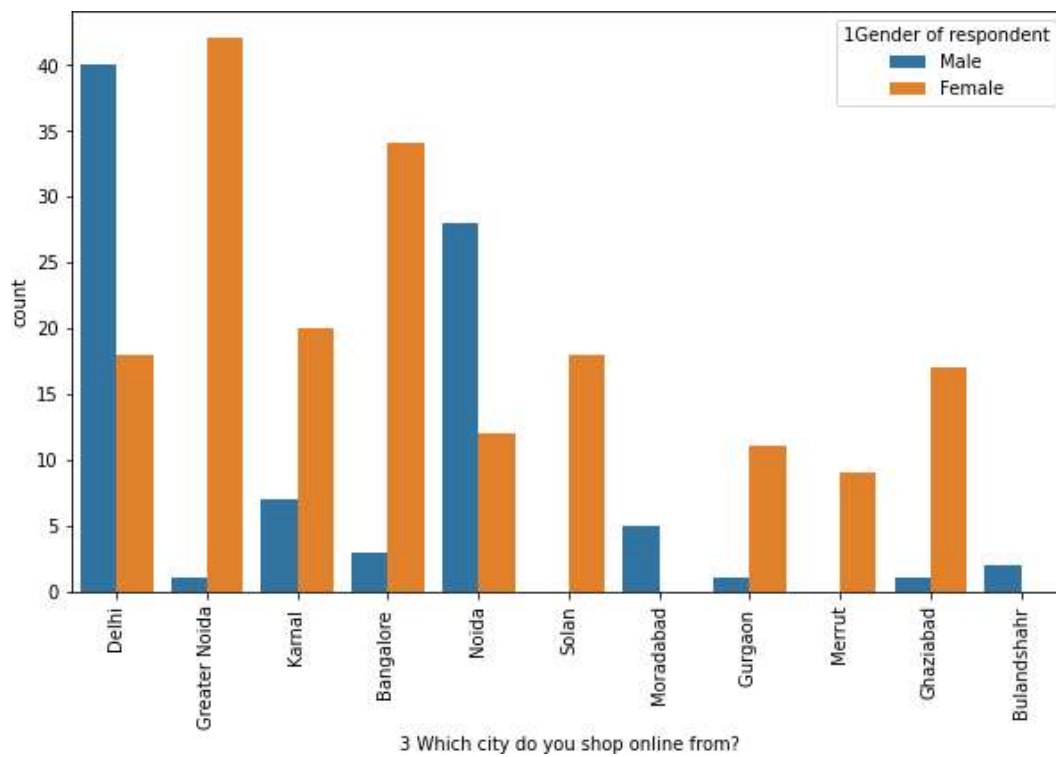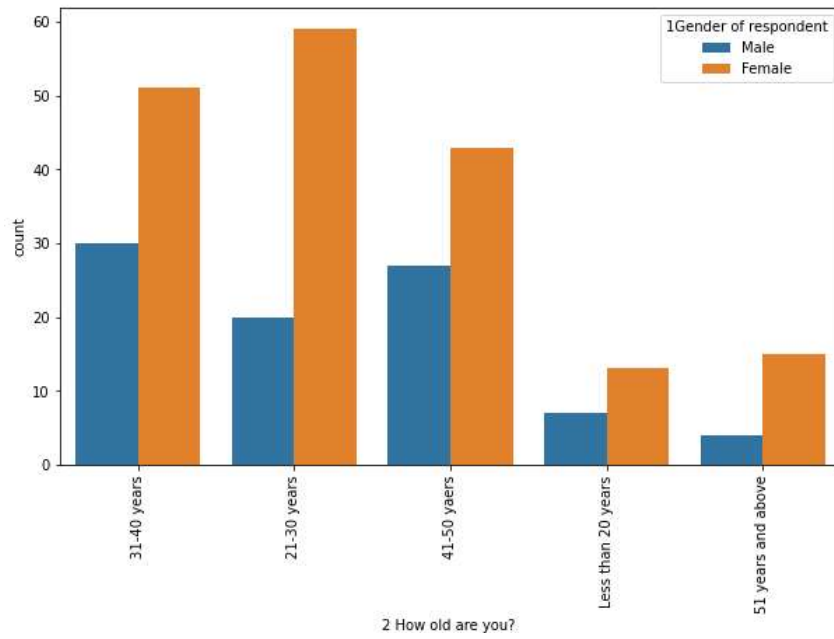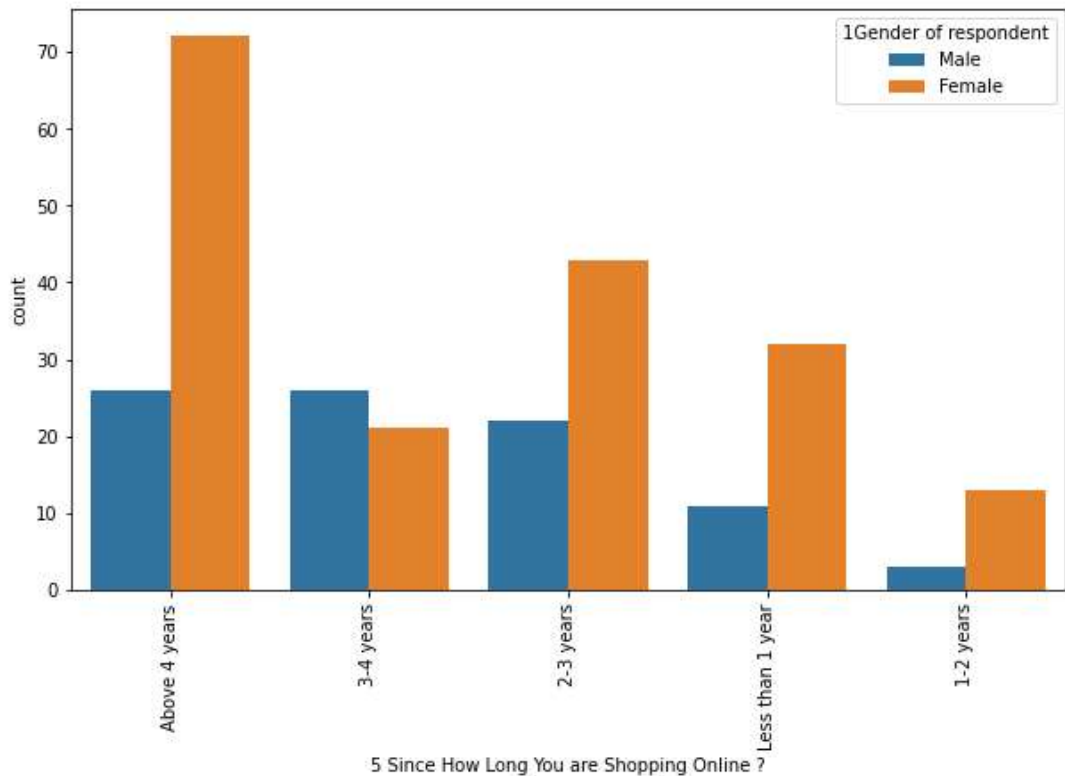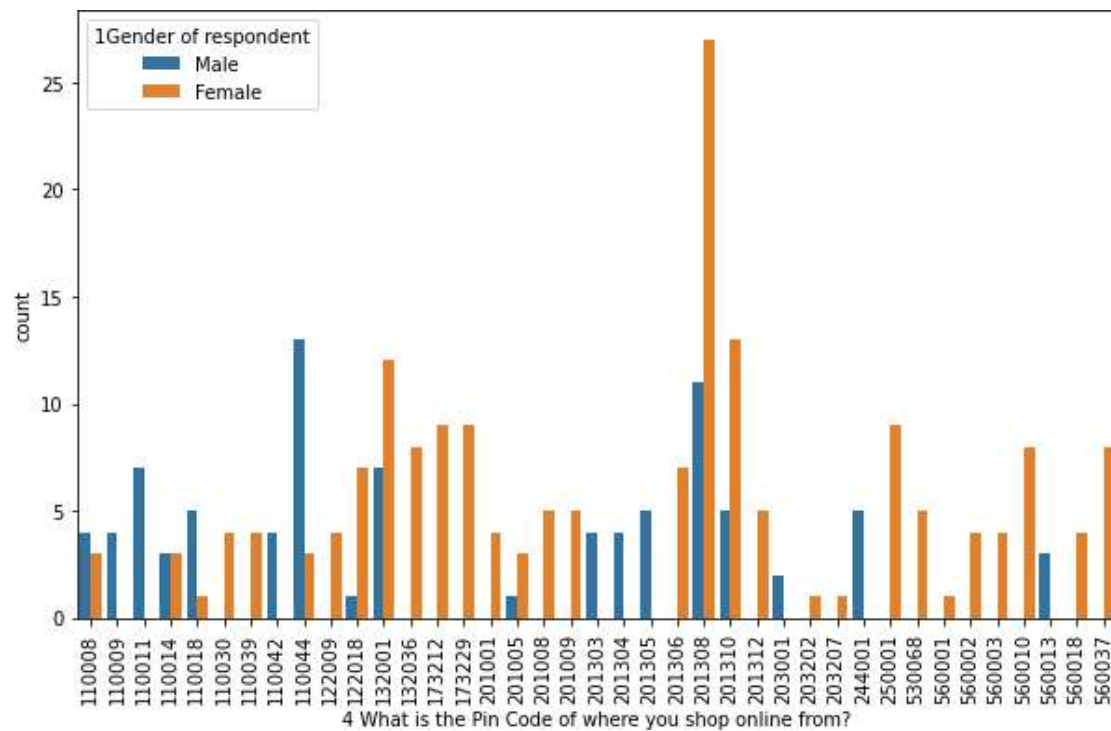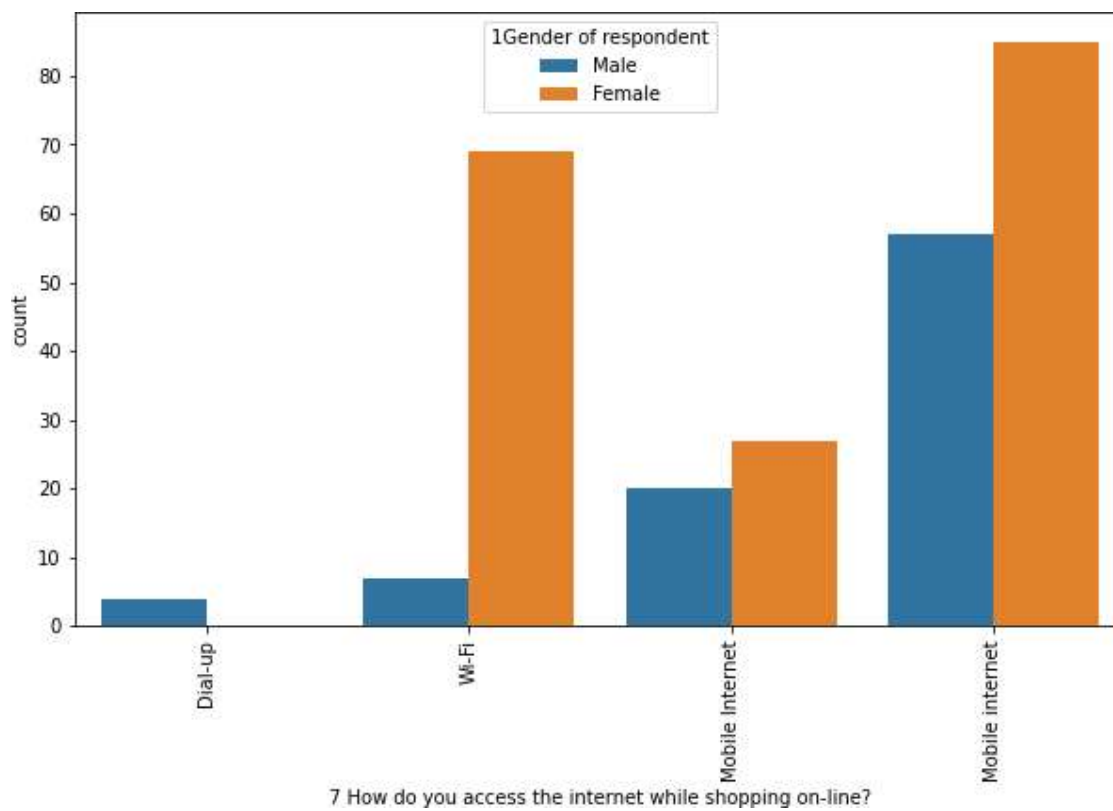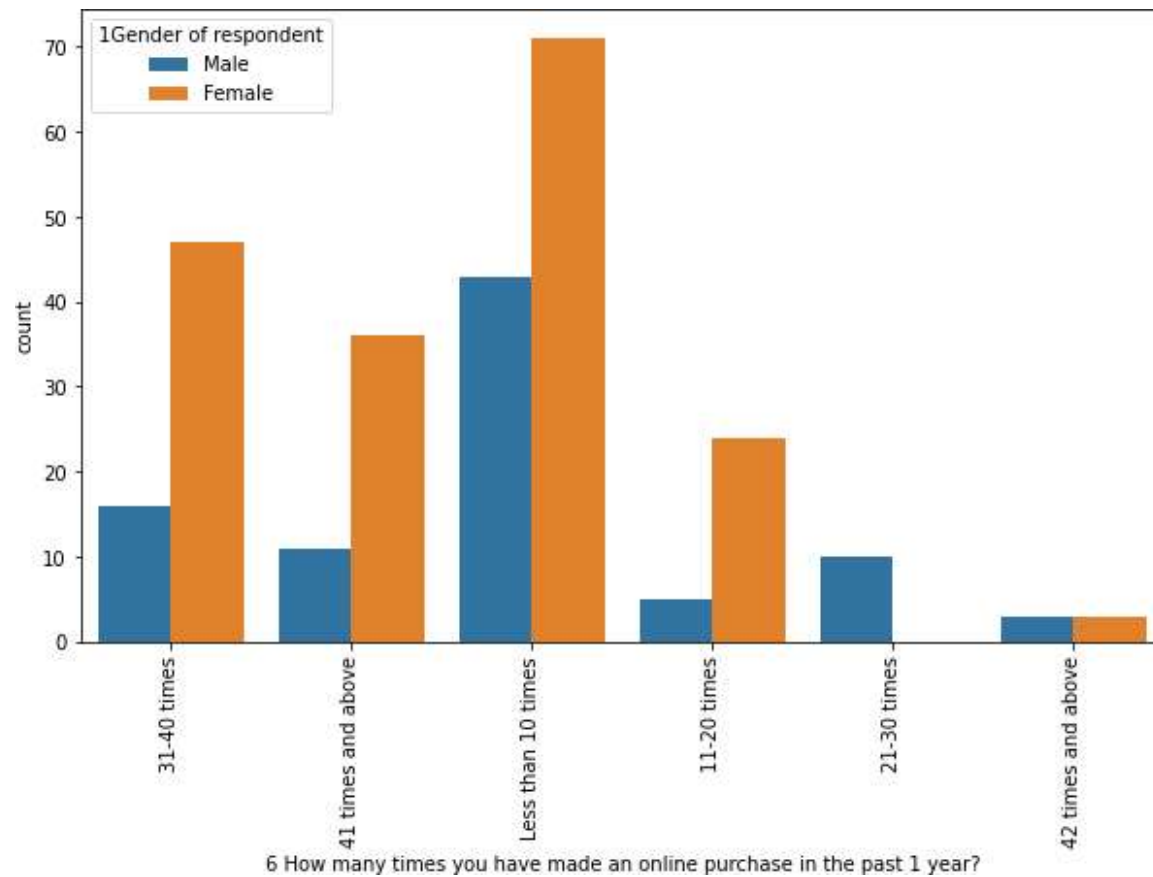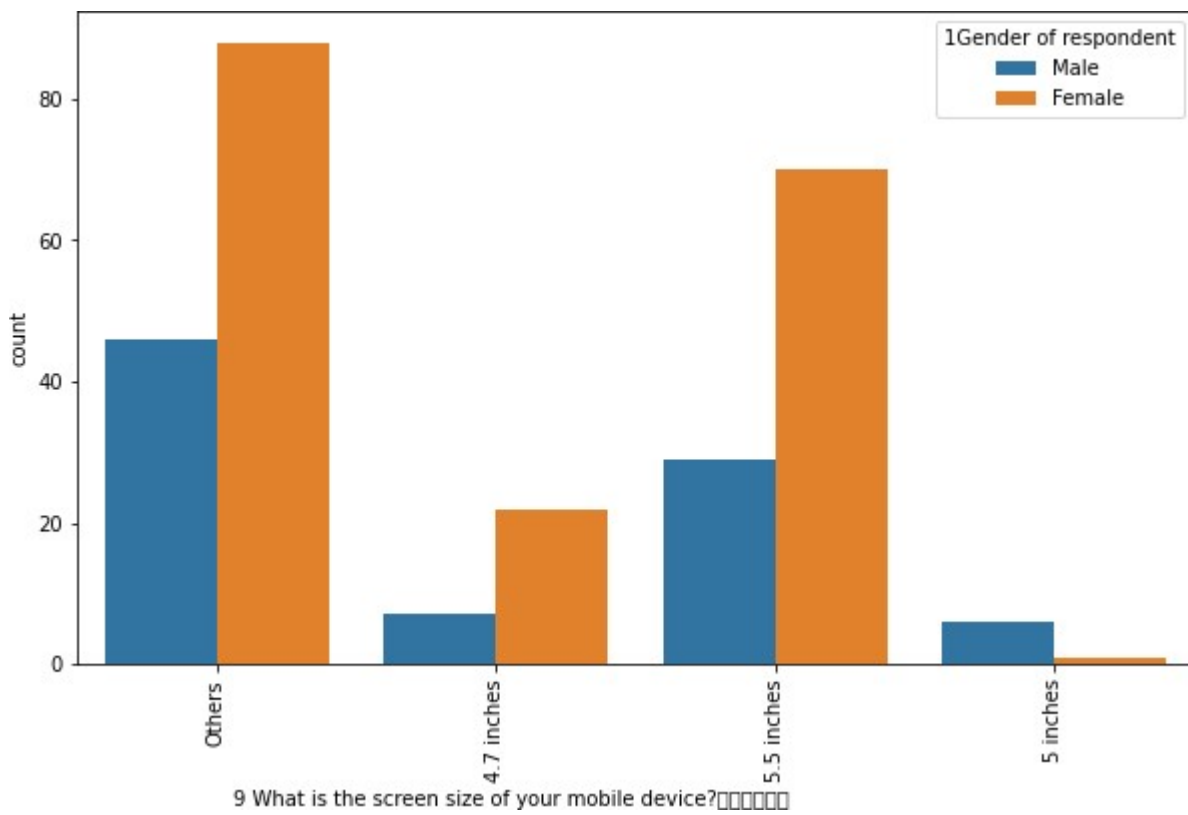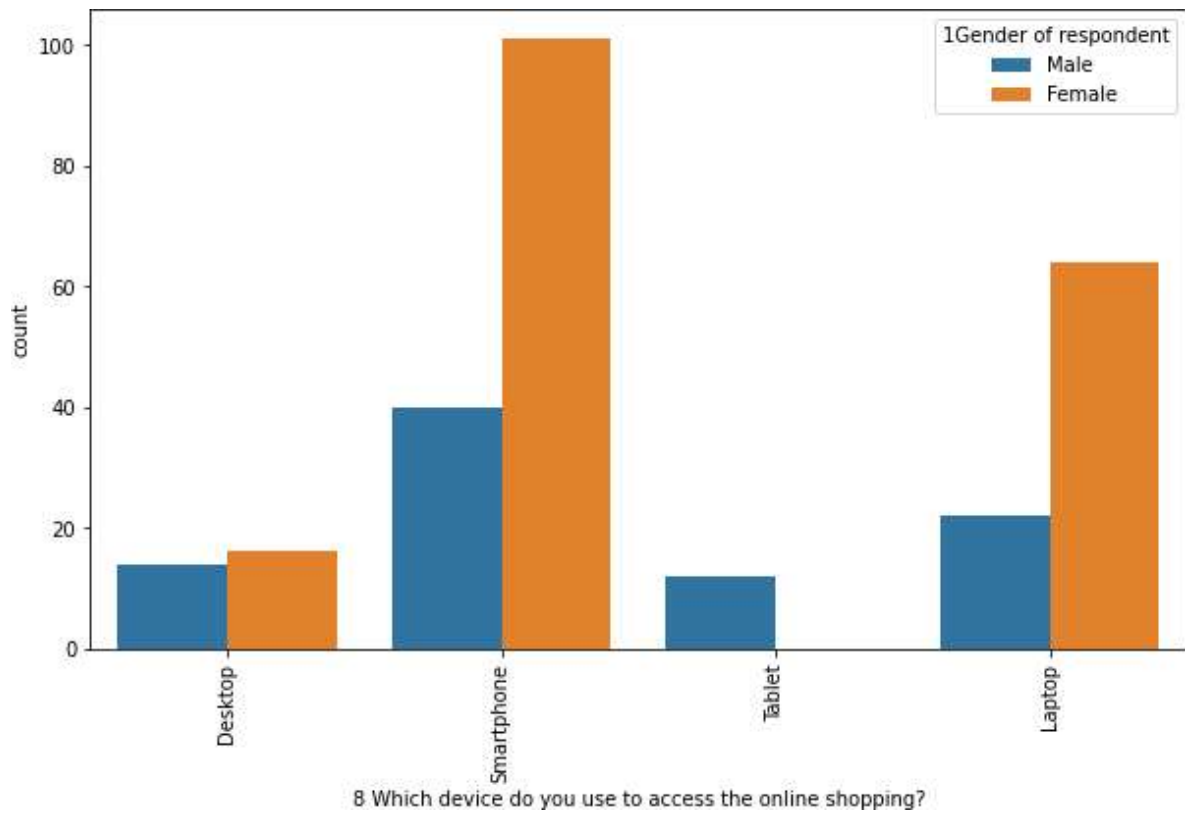
The remaining columns have been plotted in the jupyter notebook.

I have just changed the hue to 3Which city do you shop online from?

```
In [18]:   1 feature1 = df.drop(['3 Which city do you shop online from?'],axis=1)
```

```
In [19]:   1 print("-----------------------------Bivariate Analysis-------------------------------")
           2 def generate_count_plot_for_bivariate_analysis_city():
           3     for i in feature1:
           4         plt.figure(figsize=(10,6))
           5         c = sns.countplot(df[i],hue=df['3 Which city do you shop online from?'],edgecolor='black',linewidth=2)
           6         plt.xticks(rotation=90)
           7 generate_count_plot_for_bivariate_analysis_city()
```

Output has been plotted in the jupyter notebook.

Next I have changed hue to 2How old are you?

```
In [20]:   1 feature2 = df.drop(['2 How old are you? '],axis=1)
```

```
In [21]:   1 print("-----------------------------Bivariate Analysis--------------------------------
           2 def generate_count_plot_for_bivariate_analysis_age():
           3     for i in feature2:
           4         plt.figure(figsize=(10,6))
           5         c = sns.countplot(df[i],hue=df['2 How old are you? '],edgecolor='black',linewidth=2)
           6         plt.xticks(rotation=90)
           7 generate_count_plot_for_bivariate_analysis_age()
```

Output has been plotted in a jupyter notebook.

I have just completed the bivariate analysis part.I will proceed to the Encoding part.

ENCODING

Ordinal Encoding : Encode categorical features as an integer array.

The input to this transformer should be an array-like of integers or strings, denoting the values taken on by categorical (discrete) features. The features are converted to ordinal integers. This results in a single column of integers (0 to n_categories - 1) per feature.

Then I performed Ordinal Encoding on all the object data type columns before I could proceed with any kind of multivariate analysis.

Code:

```
In [23]:   1  from sklearn.preprocessing import OrdinalEncoder
```

```
In [24]:   1  #Assigning object
           2  ord = OrdinalEncoder()
           3  ord
```

```
Out[24]: OrdinalEncoder()
```
In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

```
In [25]:   1  for i in categorical_variables:
           2      df[[i]] = ord.fit_transform(df[[i]])
           3  df.head()
```

Out[25]:

|	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)	(pr
0	1.0	1.0	2.0	110009	3.0	2.0	0.0	0.0	3.0	2.0	...	0.0	0.0	3.0	5.0	
1	0.0	0.0	2.0	110030	3.0	3.0	3.0	2.0	0.0	1.0	...	1.0	6.0	7.0	10.0	
2	0.0	0.0	4.0	201308	2.0	3.0	1.0	2.0	2.0	0.0	...	7.0	6.0	4.0	7.0	
3	1.0	0.0	6.0	132001	2.0	5.0	1.0	2.0	2.0	1.0	...	9.0	7.0	4.0	8.0	
4	0.0	0.0	0.0	530068	1.0	0.0	3.0	2.0	0.0	1.0	...	5.0	8.0	5.0	8.0	

5 rows × 71 columns

I was able to obtain object data type conversion to numeric datatype with the help of Ordinal Encoding method.

Multivariate Analysis: The statistical study of data where multiple measurements are made on each experimental unit and where the relationships among multivariate measurements and their structure are important.

code:

```
In [26]:   1  df.hist(figsize=(40,60))
           2  plt.show()
```

Output:
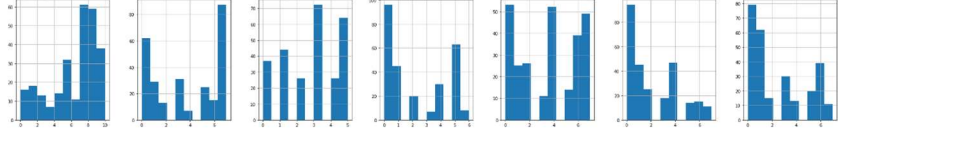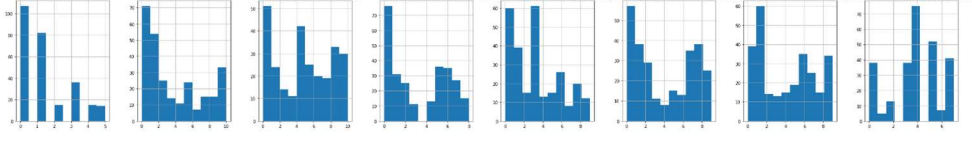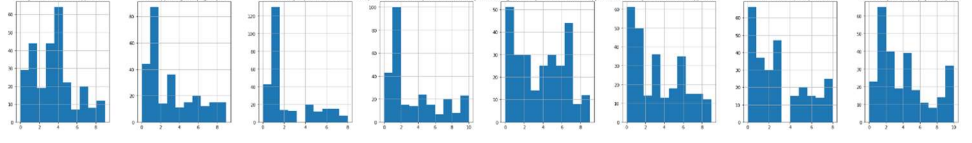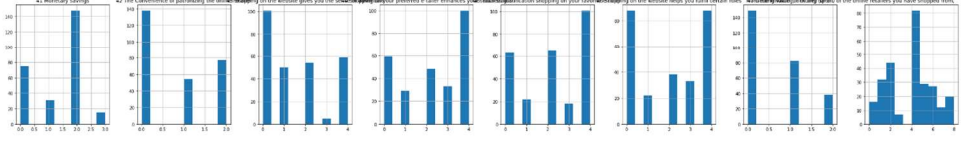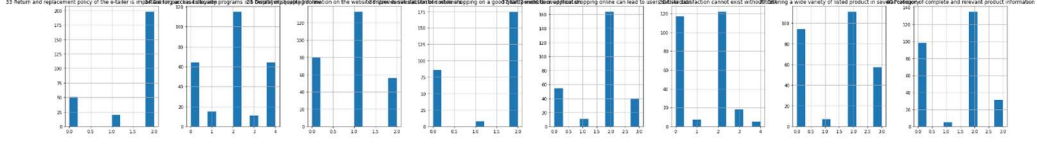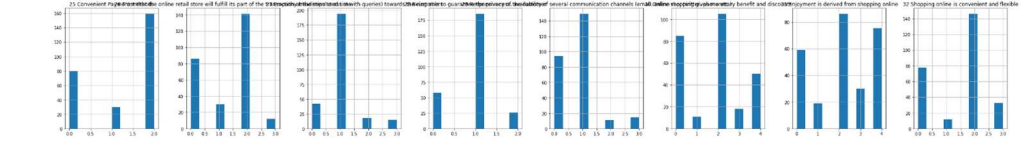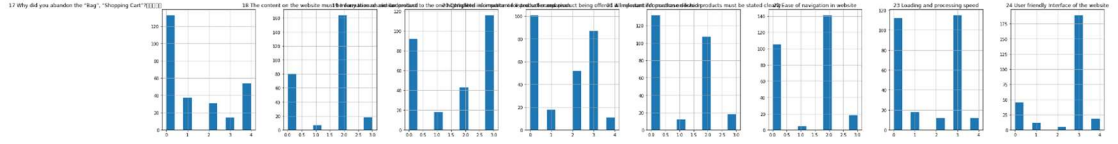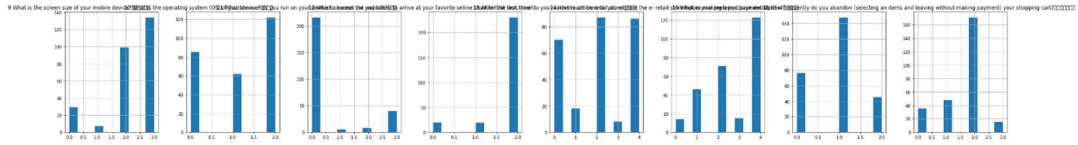
Correlation Matrix:

I generated a heatmap using the correlation values between the dataset columns. The correlation details are bifurcated majorly into positive and negative parts.
Positive correlation - A correlation of +1 indicates a perfect positive correlation, meaning that both variables move in the same direction together.
Negative correlation - A correlation of −1 indicates a perfect negative correlation, meaning that as one variable goes up, the other goes down.

Code:

```
In [28]:    1  plt.figure(figsize=(20,20))
            2  sns.heatmap(df_corr,annot=True)

Out[28]:  <AxesSubplot:>
```

Output:

In the above heatmap ,I cannot figure out which columns have multicollinearity problem.So I used Variance Inflation Factor(VIF)to check .

Code:

```
In [29]:  1  from statsmodels.stats.outliers_influence import variance_inflation_factor
          2
          3  vif = pd.DataFrame()
          4  vif['vif'] = [variance_inflation_factor(df.values,i) for i in range(df.shape[1])]
          5  vif['features'] = df.columns
```

Output:

The better picture has been plotted in the jupyter notebook.

The jupyter notebook also includes outlier detection and some more visualization techniques used like joint plots.It has been wonderfully plotted in the jupyter notebook.

Findings:

- In this project we have investigated ecommerce quality in online businesses and developed new knowledge to understand the most important dimensions of E-retail factor for customer activation and retention.
- This project aimed to enhance prior understanding of how e-commerce websites affected customer satisfaction,customer trust and customer behavior i.e repurchase intention,customer loyalty and site revisit.
- The dimensions like information about products,convenient payment mode,Trust,Fulfillment,website design change,security/privacy and many others had a positive impact on the e-commerce websites for customers.

Observations:

1. Flipkart
To be improved:
- During promotions, try to give a disturbance free shopping experience to customers.
- Flipkart has to bring some certification
of product
Positive Feedback:
- Convenient to use and also a good website for shopping.
- Presence of online assistance through multi-channels

2. Amazon.com
To be improved:
- Reduce the delivery time of the products.
- Give more payment options to customers
- Flipkart and Amazon almost share the same feedbacks with varying percentages as the only difference.
Positive feedback summary:
1. Availability of complete information of the products.
2. Trustworthy , Return Replacement facilities in Maximum Products.

3. Myntra.com
To be improved:
- Try to give the price early during promotions.
- Reduce the delivery time of the products during promotions.
Positive feedback summary:
1. Different Brand Option Wide range of Brands.
2. Wide variety of product to offer.

4. Paytm.com
To be improved:
- Try to give the price early during promotion.
- Frequent disturbance is occurring while moving from one page to another.

Positive feedback summary
- Cashback & Vouchers on Purchases
- Around 20% of the customers believe that Paytm has a wild variety of products on offer.

5. Snapdeal.com
To be improved:
- Reduce the delivery time of the products during promotions.
- Late declaration of price and discounts.
- No one has expressed to recommend Snapdeal to a contact as it has the most negative feedback among all other websites.

Positive feedback summary:
- Convenient to use.
- 54% of the customers are happy about the availability of financial information security.

Conclusion:

Based on overall observations, the first 47 features provide insights into how e-tailer is helpful & growing based on customer inputs. The data explained how the online platform has been used more often in which CITY, PIN CODE, AGE etc. It also showed that in some factors there is less importance given to contributing to the success of an e-commerce store, so based on that we could remove those factors & keep all the important factors, also we could improve on some factors that influence the online customers repeat purchase intention.

Apart from the first 47 features, the rest of the features showed which online platform has been used more based on the success factors. Based on the case study for customer activation & retention, Amazon is most reliable and has fulfilled the customer requirements. After Amazon, data showed Flipkart has been used more for online shopping.

The case study from Indian e-commerce customers showed Amazon and Flipkart have been used mostly for Online Shopping and most recommended by Friends. So, based on the research factors, Amazon & Flipkart are the e-commerce platforms, which are having the combination of both utilitarian and hedonistic values to keep the repeat purchase intention

(loyalty) positively.