

Contents

1. INTRODUCTION	1
1.1 Importing the dataset	1
2. EXPLORATORY DATA ANALYSIS.....	2
2.1 Null Values.....	2
2.2 Summary Statistics	3
2.3 Count Plot individual column.....	6
2.4 Count Plot based on Gender.....	13
2.5 Group By function	15
2.6 Catplot	16
3. MODEL BUILDING	17
3.1 Label Encoding	17
3.2 Skewness	18
3.3 Scaling.....	18
3.4 SMOTE for imbalance.....	18
3.5 Model Training	19
CONCLUSION	20
REFERENCES.....	21

E-Retail: Customer Retention Case Study Data Visualization

1. Introduction

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:

1. service quality
2. system quality
3. information quality
4. trust and
5. net benefit.

The research has investigated the factors that influence the online customers repeat purchase intention. The combination of both useful value and pleasure-seeking values are needed to affect the repeat purchase intention (loyalty) positively. Results indicate the e-retail success factors, which are very much critical for customer satisfaction. Below figure 1 gives the roadmap in retaining the customer.



Figure 1.1 Customer retention roadmap [1]

1.1 Importing the dataset

The data was collected from the Indian online shoppers. There are two datasets which were provided one regular and the other one encoded. Dataset was provided in excel format. Before importing the data set necessary libraries required for further analysis were imported as shown below.

E-Retail: Customer Retention Case Study Data Visualization

```
In [1]: # Importing Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

Dataset without encoding was used for analyzing. After importing top 5 rows are plotted using dot head. Since the number of columns are 71, to see all the column names display.max_columns is used.

```
In [2]: cr_df = pd.read_excel(r'G:\AI Professional\Data Trained\Internship Flipprobo Technologies\Customer_retention_dataset\customer_retention_data.xlsx')
cr_df.head()
```

Out[2]:

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

5 rows x 71 columns

```
In [3]: # To display all the columns
pd.set_option('display.max_columns', None)
cr_df.head(2)
```

```
In [4]: # Dimensions of the dataset
cr_df.shape
```

Out[4]: (269, 71)

2. Exploratory Data Analysis

2.1 Null Values

Firstly, the basic information of the data is plotted as shown below.

```
In [5]: # Basic information of the dataset like column names, data type of column, count of Non null values in each column
cr_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
```

Checking for Null values using isnull() function. The data has no null values as shown below.

```
In [6]: # Checking for null values in the data
cr_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

In [7]: cr_df.isnull().sum().sum()

Out[7]: 0
```

Next, obtained the names of all columns to use the names for further analysis.

```
In [8]: # Getting all the column names
cr_df.columns

Out[8]: 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?'],
              dtype=object)
```

2.2 Summary Statistics

Since all the columns present in the dataset are of object datatype, we need to check the count of each value in the column along with the frequency of each value. From this we will come to know which value in each column is more dominant which will give us the how the respondents feel about using the online platforms for shopping. For loop is used for the same. Later for statistical data describe() function is used which will give the following stats of object datatype like count, number of unique values, top value and frequency of the top value as shown in Table 2.1 below. This table provides the overall picture of customer response through which we can come to know whether the customer will continue buying online or not.

E-Retail: Customer Retention Case Study Data Visualization

```
In [9]: # Count of each categorical feature through which we can identify the dominant values and
# also duplicated values with change in name of the value
for i in cr_df.columns:
    print(cr_df[i].value_counts())
```

```
Female    181
Male       88
Name: 1Gender of respondent, dtype: int64
31-40 years    81
21-30 years    79
41-50 yaers    70
Less than 20 years    20
51 years and above    19
Name: 2 How old are you? , dtype: int64
Delhi        58
```

Table 1 Summary Statistics of Customer Retention

Column Names	count	unique	top	freq
1Gender of respondent	269	2	Female	181
2 How old are you?	269	5	31-40 years	81
3 Which city do you shop online from?	269	11	Delhi	58
5 Since How Long You are Shopping Online ?	269	5	Above 4 years	98
6 How many times you have made an online purchase in the past 1 year?	269	6	Less than 10 times	114
7 How do you access the internet while shopping on-line?	269	3	Mobile Internet	189
8 Which device do you use to access the online shopping?	269	4	Smartphone	141
9 What is the screen size of your mobile device?	269	4	Others	134
10 What is the operating system (OS) of your device?	269	3	Window/windows Mobile	122
11 What browser do you run on your device to access the website?	269	4	Google chrome	216
12 Which channel did you follow to arrive at your favorite online store for the first time?	269	3	Search Engine	230
13 After first visit, how do you reach the online retail store?	269	5	Search Engine	87
14 How much time do you explore the e- retail store before making a purchase decision?	269	5	more than 15 mins	123
15 What is your preferred payment Option?	269	3	Credit/Debit cards	148
16 How frequently do you abandon (selecting an items and leaving without making payment) your shopping cart?	269	4	Sometimes	171
17 Why did you abandon the “Bag”, “Shopping Cart”?	269	5	Better alternative offer	133
18 The content on the website must be easy to read and understand	269	4	Strongly agree (5)	164
19 Information on similar product to the one highlighted is important for product comparison	269	4	Strongly agree (5)	116
20 Complete information on listed seller and product being offered is important for purchase decision.	269	5	Agree (4)	101
21 All relevant information on listed products must be stated clearly	269	4	Agree (4)	132
22 Ease of navigation in website	269	4	Strongly agree (5)	141
23 Loading and processing speed	269	5	Strongly agree (5)	115

E-Retail: Customer Retention Case Study Data Visualization

24 User friendly Interface of the website	269	5	Strongly agree (5)	189
25 Convenient Payment methods	269	3	Strongly agree (5)	159
26 Trust that the online retail store will fulfill its part of the transaction at the stipulated time	269	4	Strongly agree (5)	141
27 Empathy (readiness to assist with queries) towards the customers	269	4	Strongly agree (5)	194
28 Being able to guarantee the privacy of the customer	269	3	Strongly agree (5)	185
29 Responsiveness, availability of several communication channels (email, online rep, twitter, phone etc.)	269	4	Strongly agree (5)	149
30 Online shopping gives monetary benefit and discounts	269	5	Strongly agree (5)	105
31 Enjoyment is derived from shopping online	269	5	Strongly agree (5)	86
32 Shopping online is convenient and flexible	269	4	Strongly agree (5)	146
33 Return and replacement policy of the e-tailer is important for purchase decision	269	3	Strongly agree (5)	198
34 Gaining access to loyalty programs is a benefit of shopping online	269	5	Strongly agree (5)	115
35 Displaying quality Information on the website improves satisfaction of customers	269	3	Strongly agree (5)	133
36 User derive satisfaction while shopping on a good quality website or application	269	3	Strongly agree (5)	175
37 Net Benefit derived from shopping online can lead to users satisfaction	269	4	Strongly agree (5)	164
38 User satisfaction cannot exist without trust	269	5	Strongly agree (5)	122
39 Offering a wide variety of listed product in several category	269	4	Strongly agree (5)	111
40 Provision of complete and relevant product information	269	4	Strongly agree (5)	135
41 Monetary savings	269	4	Strongly agree (5)	148
42 The Convenience of patronizing the online retailer	269	3	Agree (4)	138
43 Shopping on the website gives you the sense of adventure	269	5	Agree (4)	101
44 Shopping on your preferred e-tailer enhances your social status	269	5	indifferent (3)	100
45 You feel gratification shopping on your favorite e-tailer	269	5	indifferent (3)	101
46 Shopping on the website helps you fulfill certain roles	269	5	Agree (4)	88
47 Getting value for money spent	269	3	Agree (4)	149
From the following, tick any (or all) of the online retailers you have shopped from;	269	9	Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com	82
Easy to use website or application	269	10	Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com	64
Visual appealing web-page layout	269	10	Amazon.in, Flipkart.com	87

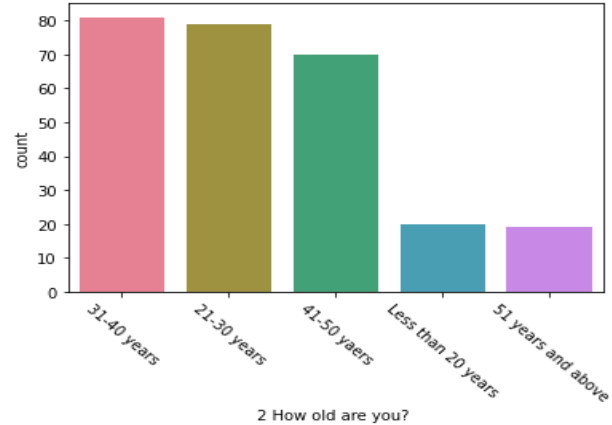
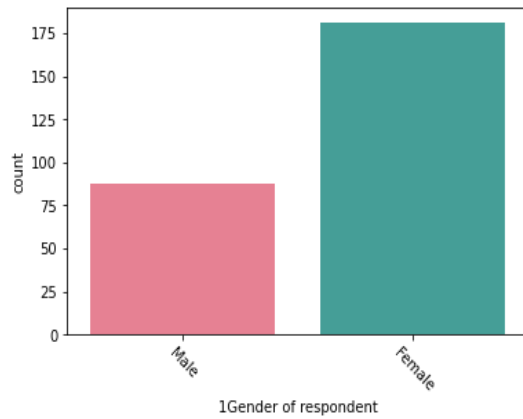
Wild variety of product on offer	269	9	Amazon.in, Flipkart.com	130
Complete, relevant description information of products	269	11	Amazon.in, Flipkart.com	100
Fast loading website speed of website and application	269	10	Amazon.in	51
Reliability of the website or application	269	10	Amazon.in	61
Quickness to complete purchase	269	9	Amazon.com	66
Availability of several payment options	269	11	Amazon.in, Flipkart.com	65
Speedy order delivery	269	6	Amazon.in	107
Privacy of customers' information	269	11	Amazon.in	71
Security of customer financial information	269	11	Amazon.in	51
Perceived Trustworthiness	269	9	Amazon.in	76
Presence of online assistance through multi-channel	269	10	Amazon.in, Flipkart.com, Myntra.com, Snapdeal	61
Longer time to get logged in (promotion, sales period)	269	10	Amazon.in	57
Longer time in displaying graphics and photos (promotion, sales period)	269	10	Amazon.in, Flipkart.com	60
Late declaration of price (promotion, sales period)	269	8	Myntra.com	75
Longer page loading time (promotion, sales period)	269	11	Myntra.com	61
Limited mode of payment on most products (promotion, sales period)	269	8	Snapdeal.com	87
Longer delivery period	269	6	Paytm.com	72
Change in website/Application design	269	7	Amazon.in	96
Frequent disruption when moving from one page to another	269	8	Amazon.in	53
Website is as efficient as before	269	8	Amazon.in	94
Which of the Indian online retailer would you recommend to a friend?	269	8	Amazon.in	79

2.3 Count Plot individual column

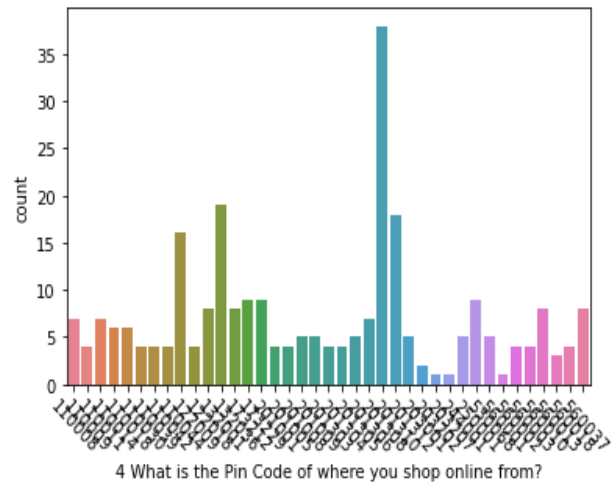
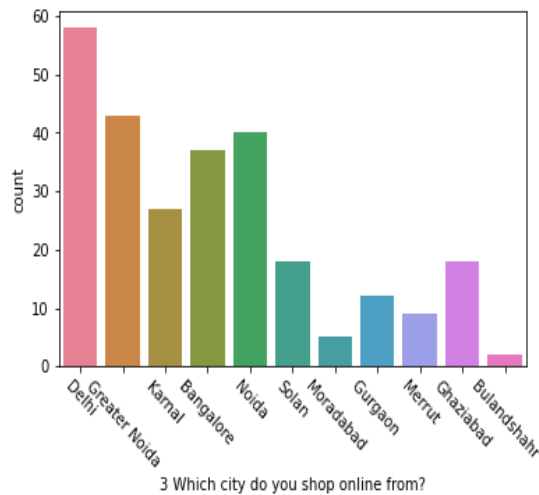
Now, let's plot some graphs like count plot of all the columns. Count plot will give us the clear picture of customer, their interest and opinion about different online shopping platforms and what medium they use such as internet facilities, mobile screen size.

```
In [12]: # Since all the columns are of object data type
# Lets plot all at once
for i in cr_df.columns:
    print(cr_df[i].value_counts()) # Gives frequency of each value in the column
    # Count Plot frequency of values in the column for clearer picture
    sns.countplot(x = i, data = cr_df, palette = 'husl')
    plt.xticks(rotation = -45)
    plt.show()
```

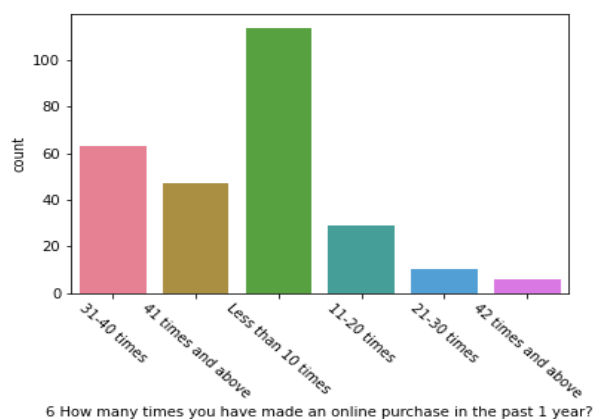
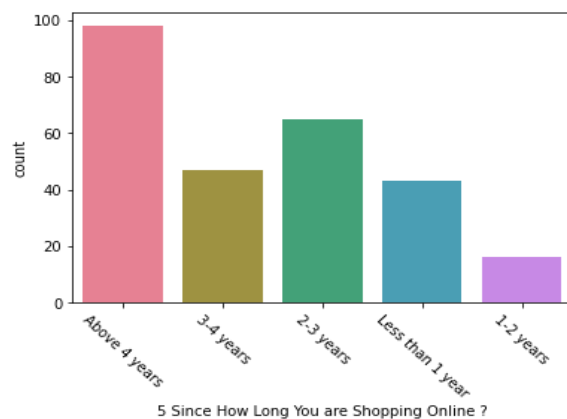
E-Retail: Customer Retention Case Study Data Visualization



More female customers than male. Most online buyers in range of 21 to 50 yrs later it decreases.

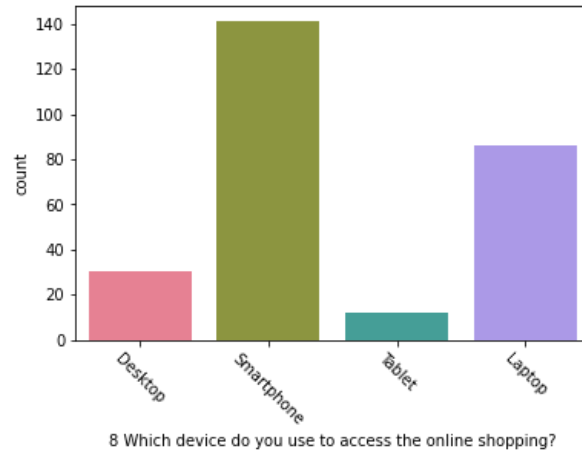
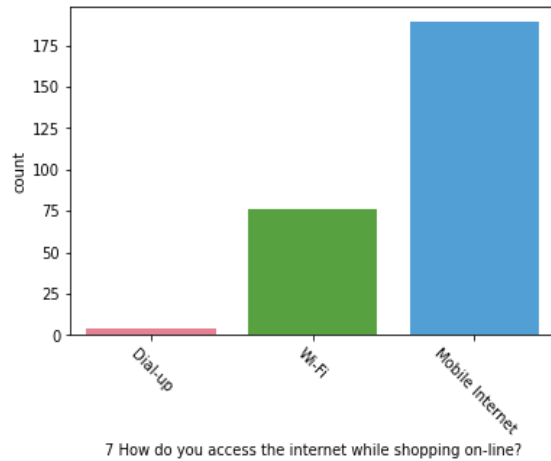


People shop more from Delhi and least from Bulandshahr.

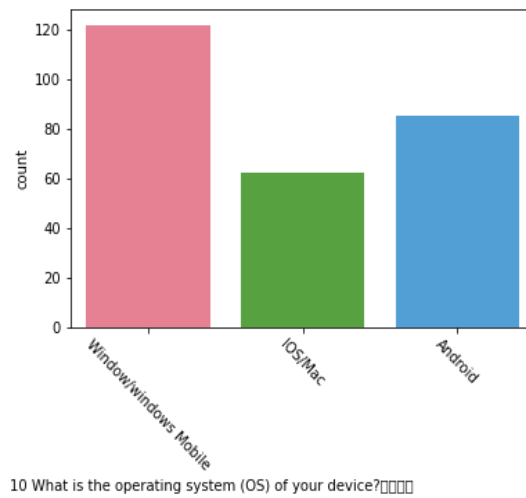
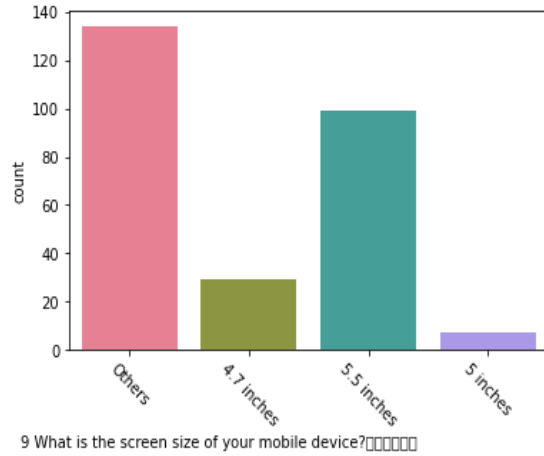


There are more people who are shopping online from past 4 years. More people who shopped less than 10 times online.

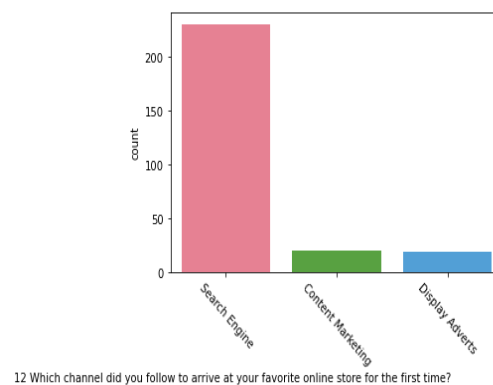
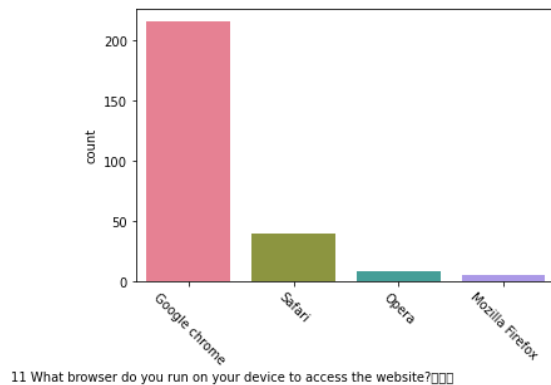
E-Retail: Customer Retention Case Study Data Visualization



Most use smartphone for online shopping, hence mobile internet.

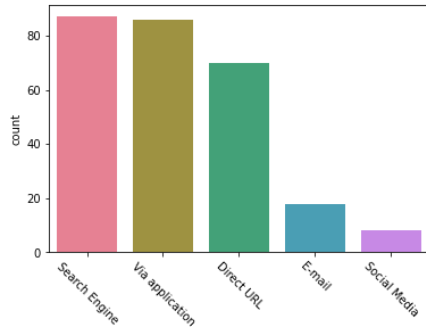


Mobile size comes of various sizes. People use all kinds of operating system.

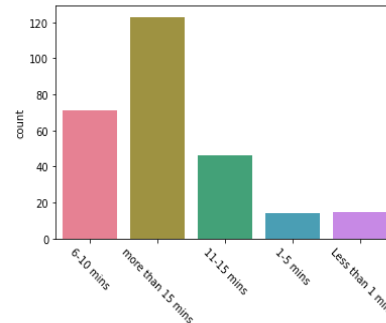


Most users use search engine for favorite online store mainly using google chrome.

E-Retail: Customer Retention Case Study Data Visualization

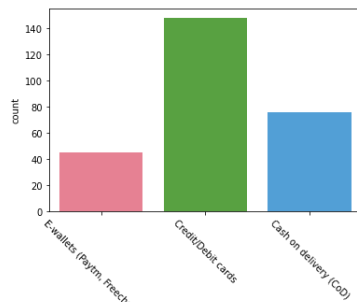


13 After first visit, how do you reach the online retail store?

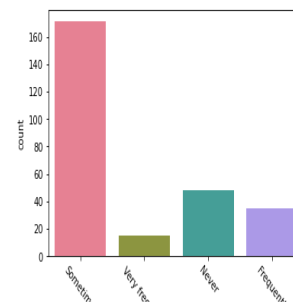


14 How much time do you explore the e- retail store before making a purchase decision?

People spend more 15mins before buying.

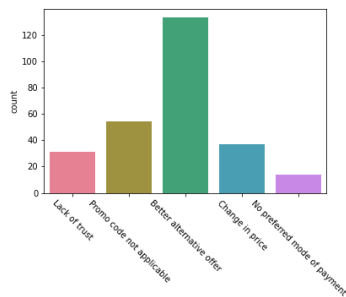


15 What is your preferred payment Option?

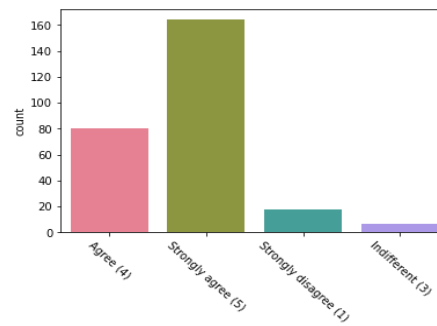


16 How frequently do you abandon (selecting an item and leaving without making payment) your shopping cart?

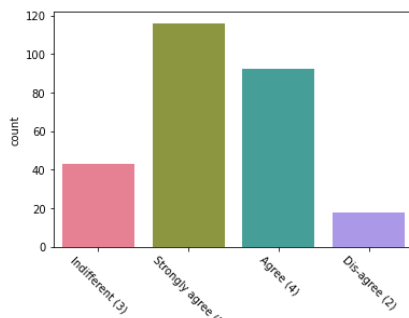
Most preferred payment is credit/debit card. Many a times after adding item to buy they discard it sometimes.



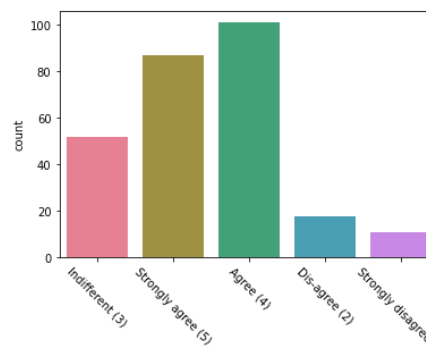
17 Why did you abandon the "Bag", "Shopping Cart"?



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

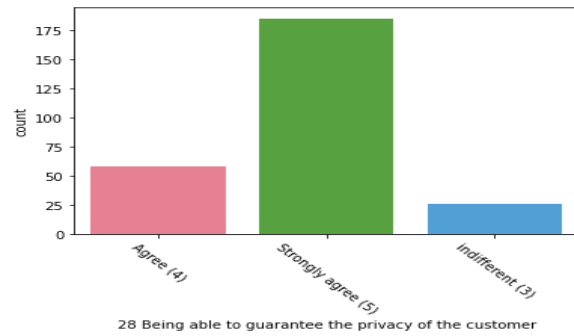
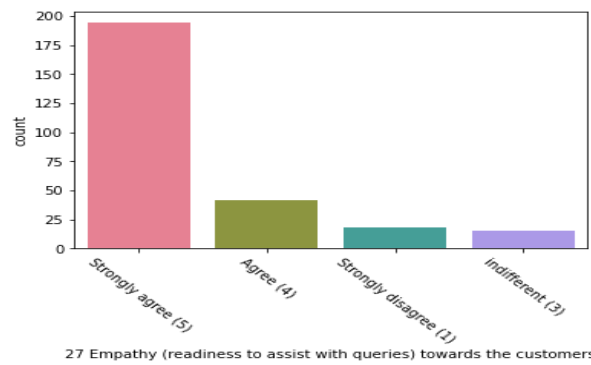
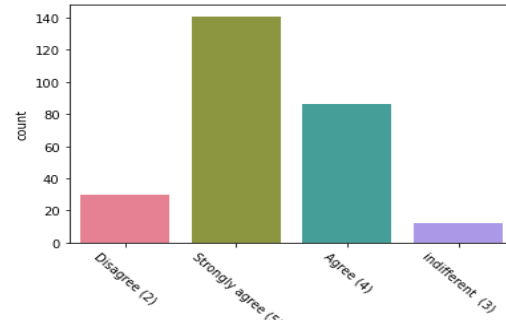
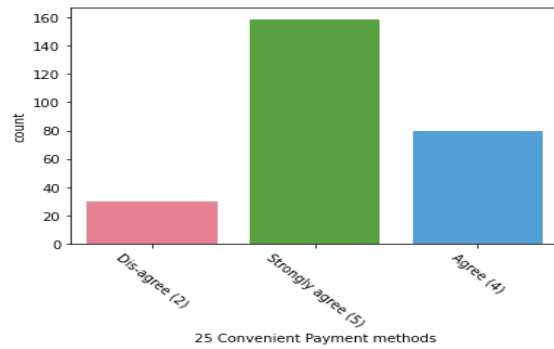
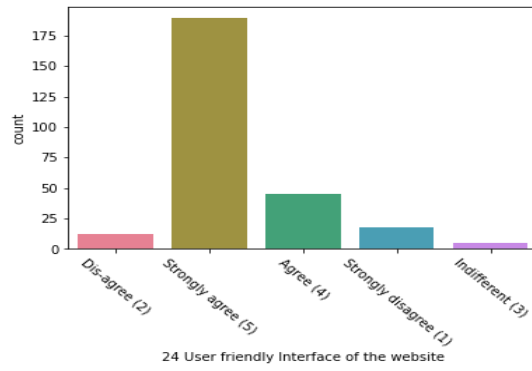
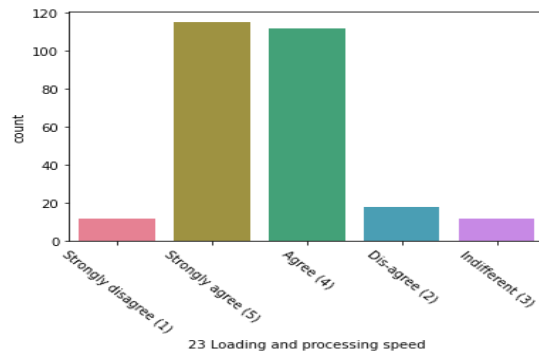
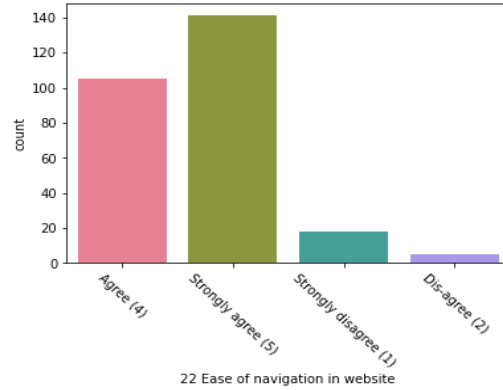
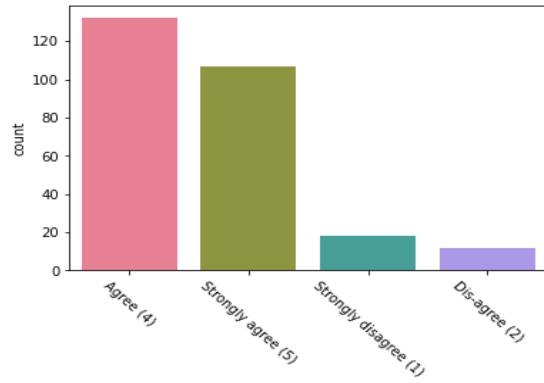


19 Information on similar product to the one highlighted is important for product comparison

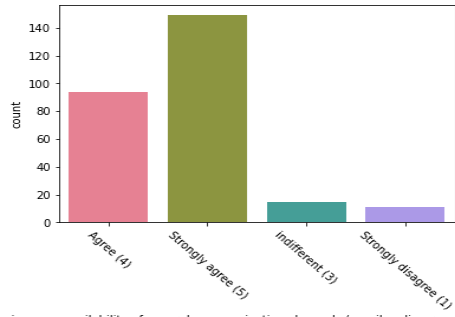


20 Complete information on listed seller and product being offered is important for purchase decision.

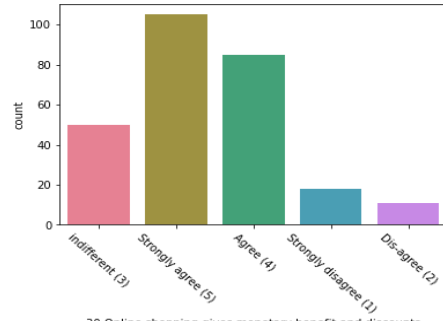
E-Retail: Customer Retention Case Study Data Visualization



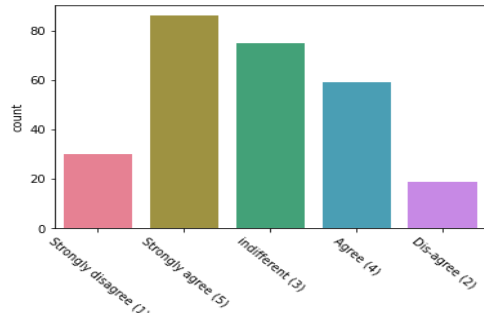
E-Retail: Customer Retention Case Study Data Visualization



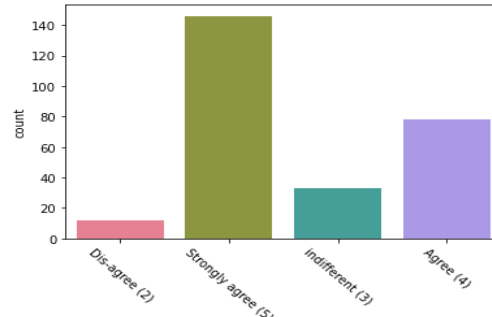
29 Responsiveness, availability of several communication channels (email, online rep, twitter, phone etc.)



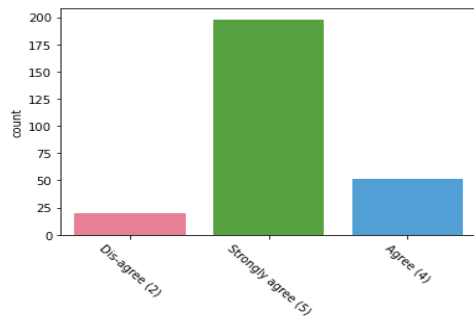
30 Online shopping gives monetary benefit and discounts



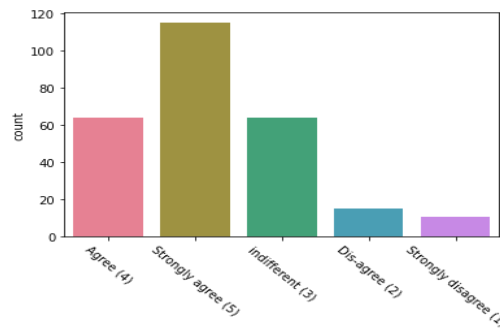
31 Enjoyment is derived from shopping online



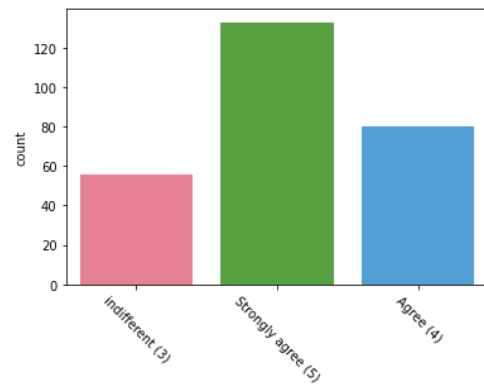
32 Shopping online is convenient and flexible



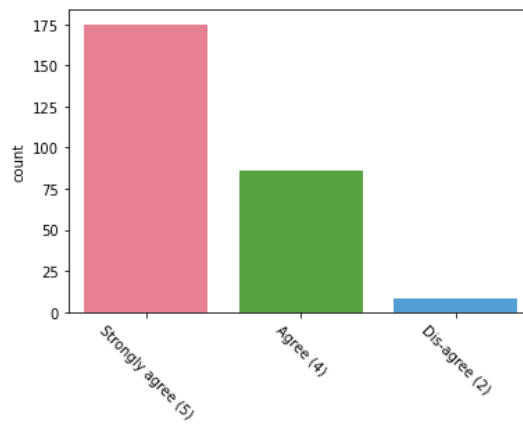
33 Return and replacement policy of the e-tailer is important for purchase decision



34 Gaining access to loyalty programs is a benefit of shopping online

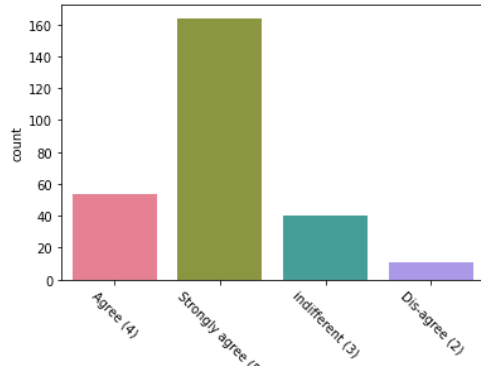


35 Displaying quality Information on the website improves satisfaction of customers

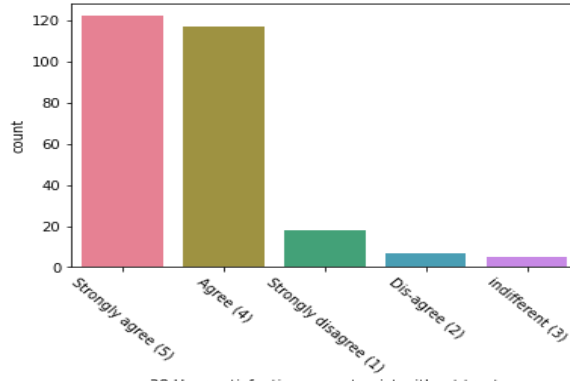


36 User derive satisfaction while shopping on a good quality website or application

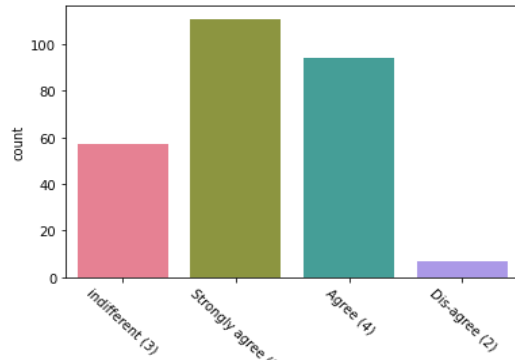
E-Retail: Customer Retention Case Study Data Visualization



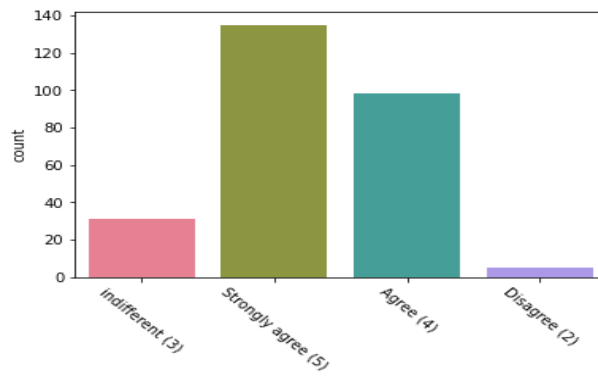
37 Net Benefit derived from shopping online can lead to users satisfaction



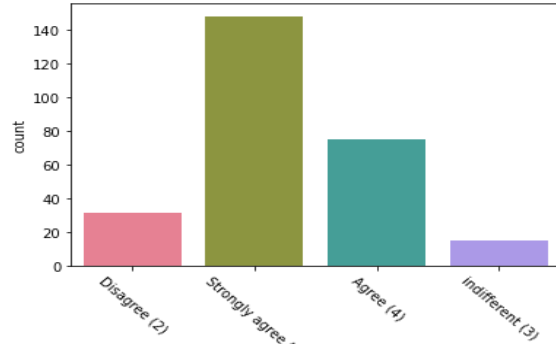
38 User satisfaction cannot exist without trust



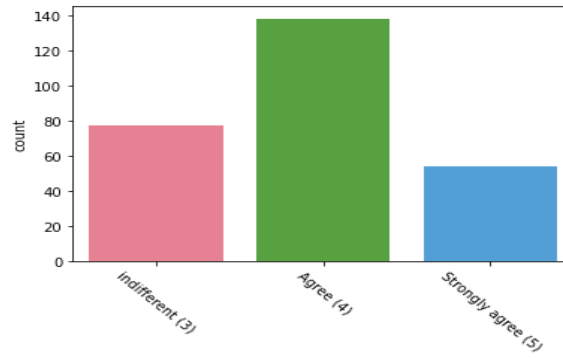
39 Offering a wide variety of listed product in several category



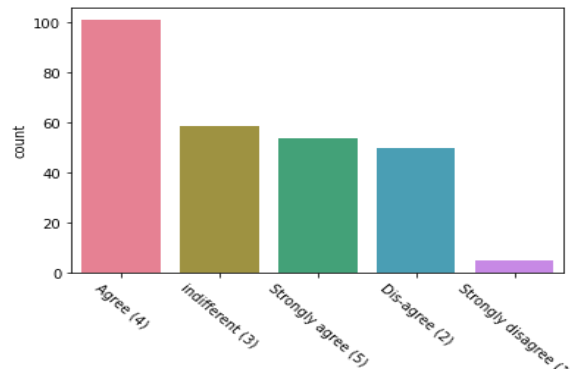
40 Provision of complete and relevant product information



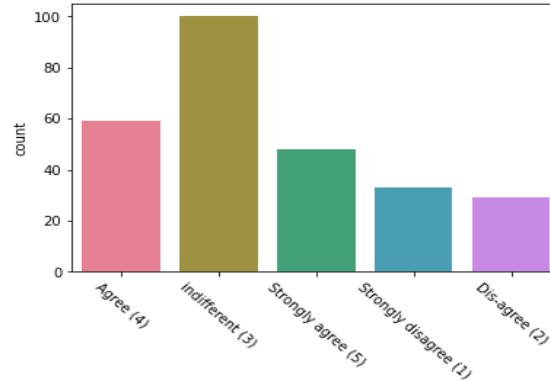
41 Monetary savings



42 The Convenience of patronizing the online retailer

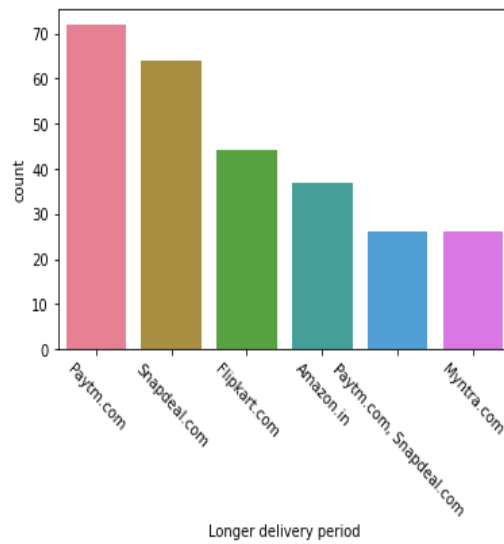
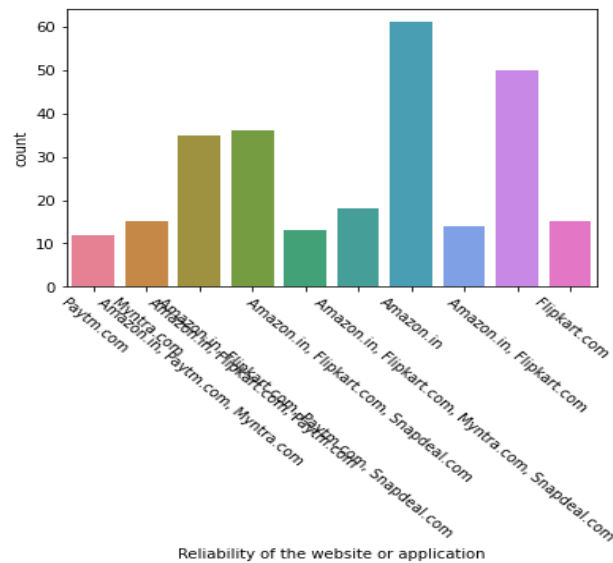
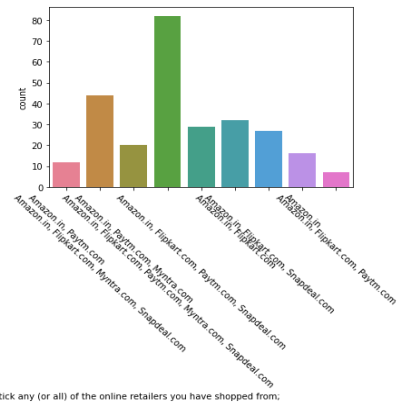
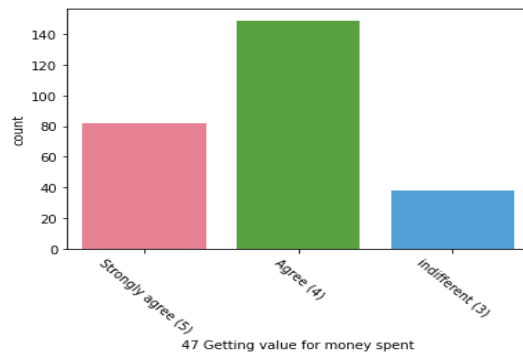
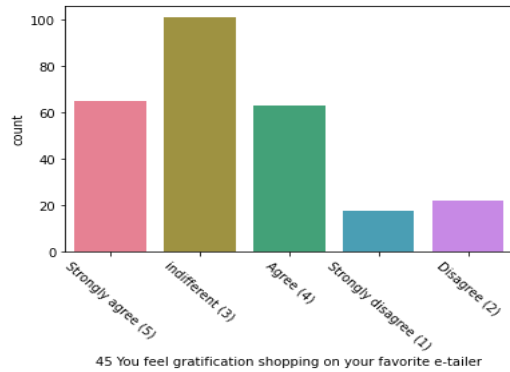


43 Shopping on the website gives you the sense of adventure



44 Shopping on your preferred e-tailer enhances your social status

E-Retail: Customer Retention Case Study Data Visualization

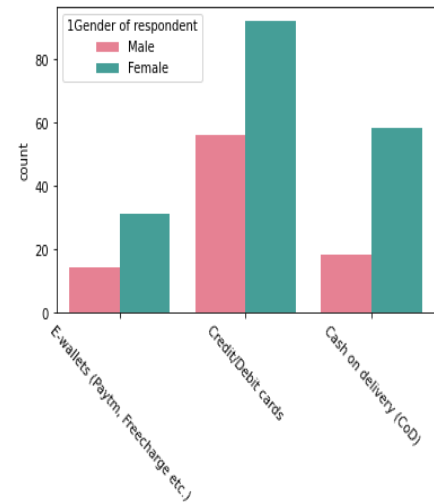
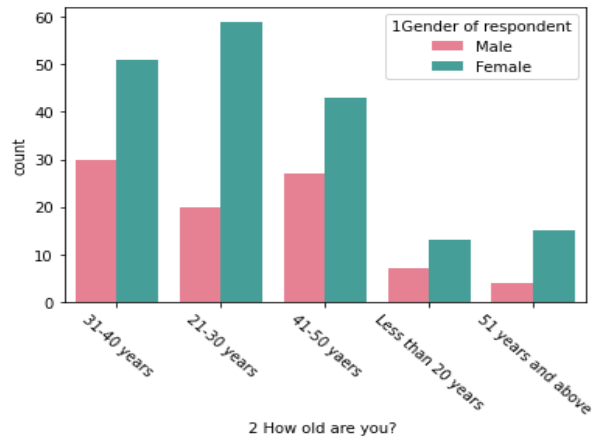


Looking at the data most preferred online shopping platform is Amazon and flipkart.

2.4 Count Plot based on Gender

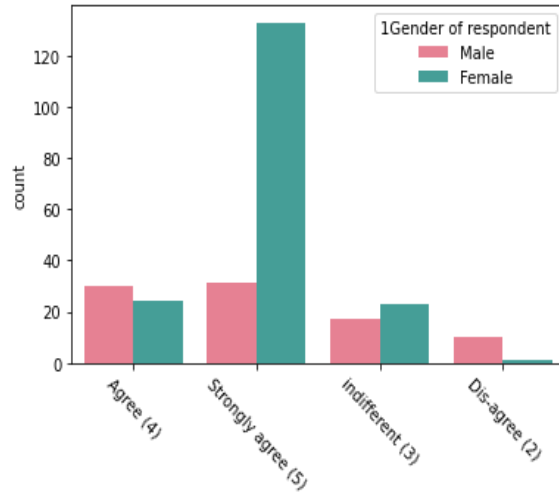
E-Retail: Customer Retention Case Study Data Visualization

Below plot clearly shows customer data with respect to gender. Only few plots are shown here and all other plots can be seen in jupyter notebook file. In most of the cases female customers are more prominent.

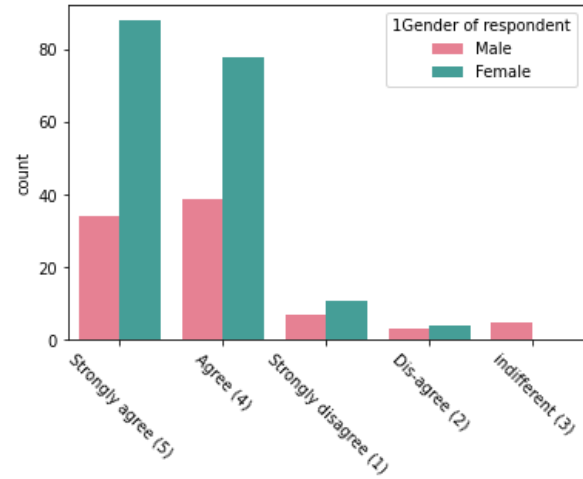


15 What is your preferred payment Option? ☐ ☐ ☐ ☐ ☐

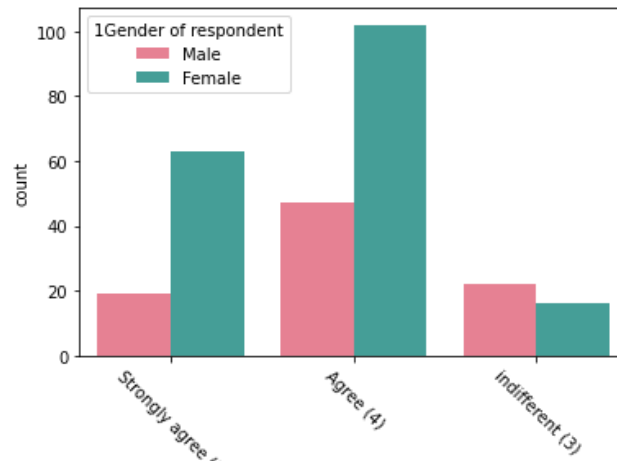
E-Retail: Customer Retention Case Study Data Visualization



37 Net Benefit derived from shopping online can lead to users satisfaction



38 User satisfaction cannot exist without trust



47 Getting value for money spent

2.5 Group By function

```
In [14]: group1 = cr_df.groupby(['3 Which city do you shop online from?', '1Gender of respondent', 'Perceived Trustworthiness'])
aa = pd.DataFrame(group1['Perceived Trustworthiness'].count())
```

```
In [15]: aa.head()
```

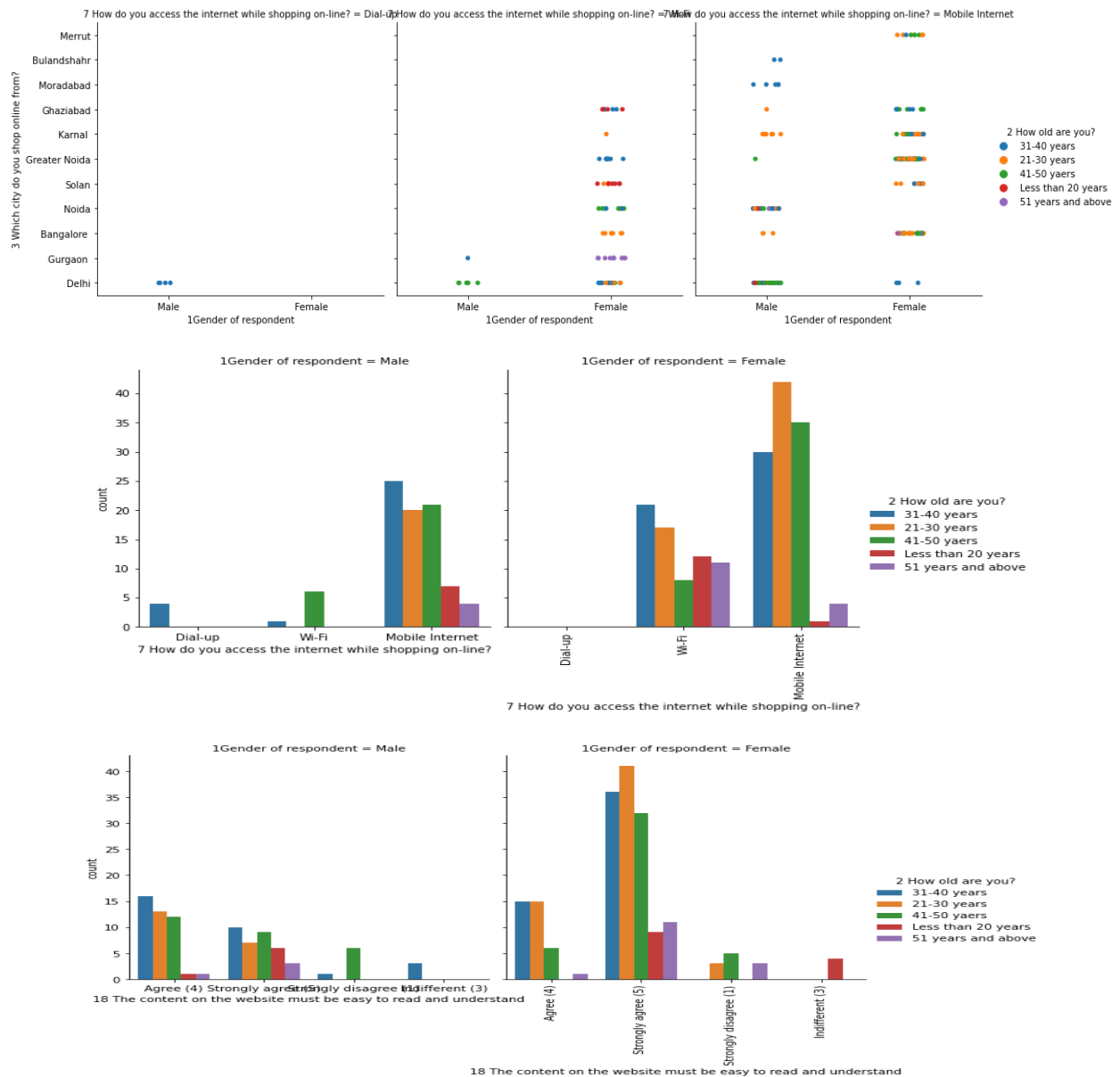
Out[15]:

3 Which city do you shop online from? 1Gender of respondent		Perceived Trustworthiness	
Bangalore	Female	Amazon.in	11
		Amazon.in, Flipkart.com	2
		Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com	4
		Amazon.in, Flipkart.com, Snapdeal.com	1
		Amazon.in, Myntra.com	9

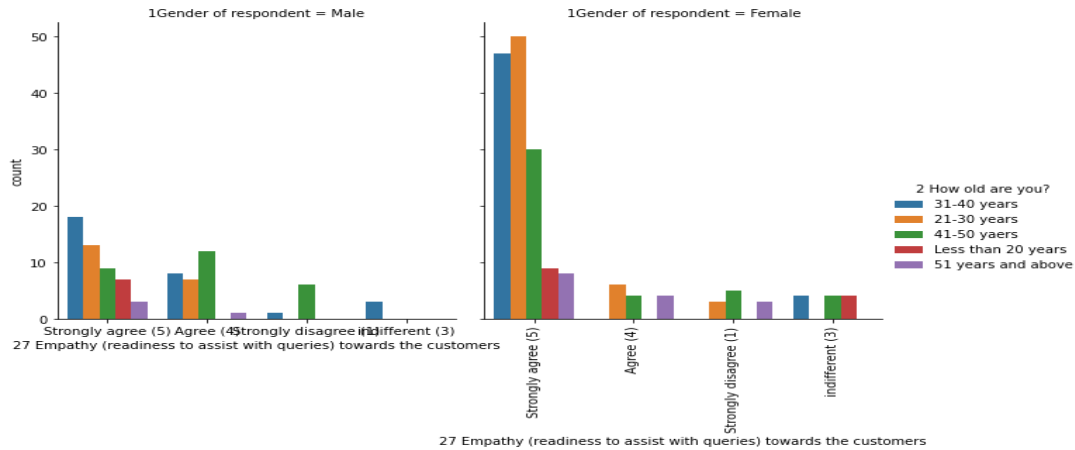
2.6 Catplot

```
In [16]: sns.catplot(data = cr_df, x = '1Gender of respondent', y = '3 Which city do you shop online from?',
hue = '2 How old are you? ',
col = '7 How do you access the internet while shopping on-line?')
# This plot gave us for different age groups of male and female respondents mode of access of
# internet while shopping
```

```
Out[16]: <seaborn.axisgrid.FacetGrid at 0x18c59c23580>
```



E-Retail: Customer Retention Case Study Data Visualization



3. Model Building

In the given dataset I have consider the gender of respondents as the target variable and the remaining columns as dependent variables. So, this becomes a classification problem. From the data available we will be able to predict whether buyer is female or male.

3.1 Label Encoding

Label Encoding

```
In [22]: from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
```

```
In [23]: for k in cr_df.columns:
          cr_df[k]=le.fit_transform(cr_df[k])
```

```
In [24]: cr_df.head()
```

Out[24]:

	1 Gender of respondent	2 How old are you?	3 Which city do you shop online from?	4 What is the Pin Code of where you shop online from?	5 Since How Long You are Shopping Online ?	6 How many times you have made an online purchase in the past 1 year?	7 How do you access the internet while shopping on-line?	8 Which device do you use to access the online shopping?	9 What is the screen size of your mobile device?	10 What is the operating system (OS) of your device?	11 What browser do you run on your device to access the website?	12 Which channel did you follow to arrive at your favorite online store for the first time?	13 After first visit, how do you reach the online retail store?	14 How much time do you explore the e-retail store before making a purchase decision?	15 What is your preferred payment Option?	16 How frequently do you abandon (selecting an item and leaving without making payment you shopping cart)
0	1	1	2	1	3	2	0	0	3	2	0	2	2	2	2	;
1	0	0	2	5	3	3	2	2	0	1	0	2	4	4	1	;
2	0	0	4	23	2	3	1	2	2	0	0	2	4	1	2	;

3.2 Skewness

```
In [27]: # Before removing skewness
print("Total count of numeric features: ",len(cr_df.skew()))
print("count of features which are significantly skewed: ",len(cr_df.skew().loc[abs(cr_df.skew())>0.5]))
```

Total count of numeric features: 71
count of features which are significantly skewed: 30

```
In [28]: for index in cr_df.skew().index:
        if cr_df.skew().loc[index]>0.5:
            cr_df[index]=np.log1p(cr_df[index])
        if cr_df.skew().loc[index]<-0.5:
            cr_df[index]=np.square(cr_df[index])
```

```
In [29]: # After removing skewness
print("Total count of numeric features: ",len(cr_df.skew()))
print("count of features which are significantly skewed: ",len(cr_df.skew().loc[abs(cr_df.skew())>0.5]))
```

Total count of numeric features: 71
count of features which are significantly skewed: 7

3.3 Scaling

Data is first split into x and y before applying scaling. Scaling is been done for both x and y.

```
In [30]: # Splitting the data in x and y
x = cr_df.drop(columns=["1Gender of respondent"])
y = cr_df[["1Gender of respondent"]]
```

```
In [31]: x.shape, y.shape
```

```
Out[31]: ((269, 70), (269, 1))
```

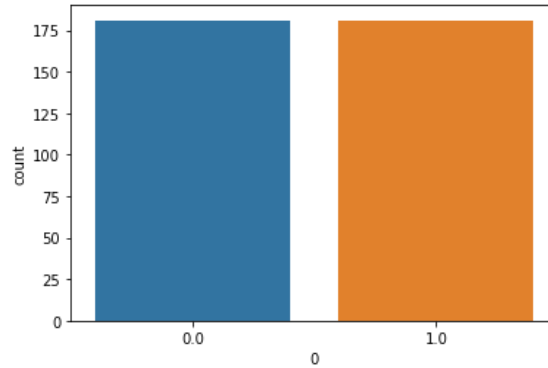
```
In [32]: # Applying MinMaxScaler before training the model for both feature and target variable
from sklearn.preprocessing import MinMaxScaler
mms_cr = MinMaxScaler()
X = mms_cr.fit_transform(X)
X
```

```
Out[32]: array([[0.43067656, 0.2      , 0.02631579, ..., 0.      , 0.      ,
                0.93578497],
                [0.      , 0.2      , 0.13157895, ..., 0.57142857, 0.33333333,
                0.66666667],
                [0.      , 0.4      , 0.60526316, ..., 0.85714286, 0.      ,
                0.86165417],
                ...,
                [0.68260619, 0.      , 0.92105263, ..., 1.      , 0.      ,
                0.      ],
                [1.      , 1.      , 0.36842105, ..., 0.71428571, 0.86165417,
                0.      ],
                [0.68260619, 0.3      , 0.47368421, ..., 0.      , 0.      ,
                0.      ]])
```

```
In [33]: y = mms_cr.fit_transform(y)
y
```

3.4 SMOTE for imbalance

Gender is the target variable which we have considered which is imbalanced. So, we need to balance it before applying any algorithm for training. After applying SMOTE we can see the data is balanced now as shown below.



3.5 Model Training

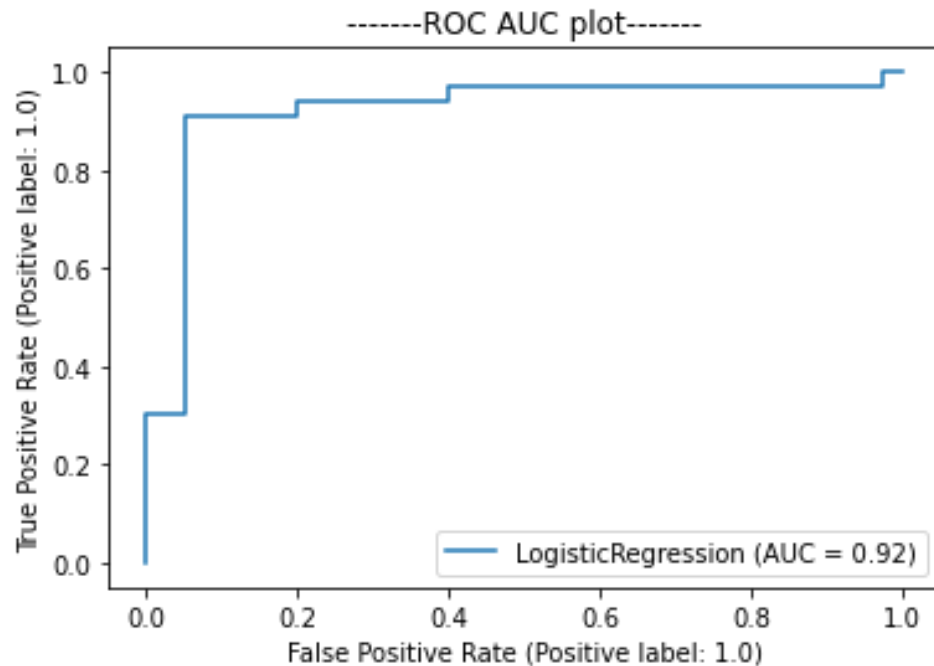
Only one algorithm is applied here i.e. Logistic Regression. No cross validation, neither hyper parameter is applied. Only with one model training the algorithm is giving good results as shown below with all the output (accuracy, training and testing score, confusion matrix and classification report) for classification problem along with roc curve.

```

----- LogisticRegression() -----
Accuracy score of LogisticRegression() is: 0.8904109589041096
Training score of LogisticRegression() is 0.9134948096885813
Testing score of LogisticRegression() is 0.9134948096885813
Confusion matrix:
[[35  5]
 [ 3 30]]
Classification Report:

```

	precision	recall	f1-score	support
0.0	0.92	0.88	0.90	40
1.0	0.86	0.91	0.88	33
accuracy			0.89	73
macro avg	0.89	0.89	0.89	73
weighted avg	0.89	0.89	0.89	73



Data frame of Predicted value and original value.

```
In [61]: pred_cr_df = pd.DataFrame([lr.predict(x_test), y_test], index=['Predicted', 'Original'])
pred_cr_df
```

Out[61]:

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Predicted	0.0	1.0	0.0	0.0	1.0	1.0	1.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0
Original	1.0	1.0	0.0	1.0	0.0	1.0	1.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0

Conclusion

From EDA of the data there are more female customers than male ones. Count plot individually as well as with respect gender was plotted. Summary statistics whose table is given above tells many things about all the columns present in the dataset. This summary will let us know all important aspects of customer response, opinion, choice and interest for online shopping websites. Group by function was used to see the trust worthiness of both male and female from various cities from where they shop online. Categorical plot also was plotted to give more clear picture.

Later, after label encoding, skewness and scaling the data was trained with Logistic Regression which gave an accuracy score of 0.9134 or 91.34%.

References

- [1] [Online]. Available: <https://www.dreamstime.com/stock-photo-customer-retention-image17628620>.