Customer Retention Case Study Report

Prepared

by

Arpan Pattanayak

Data Science Intern at Flip Robo Technologies

SME Name: Shwetank Mishra

Acknowledgement

I would like to express my sincere thanks of gratitude to my mentors from Data Trained academy and Flip Robo company for letting me work on this project. Their suggestions and directions have helped me in the completion of this project successfully.

INTRODUCTION

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 or 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.

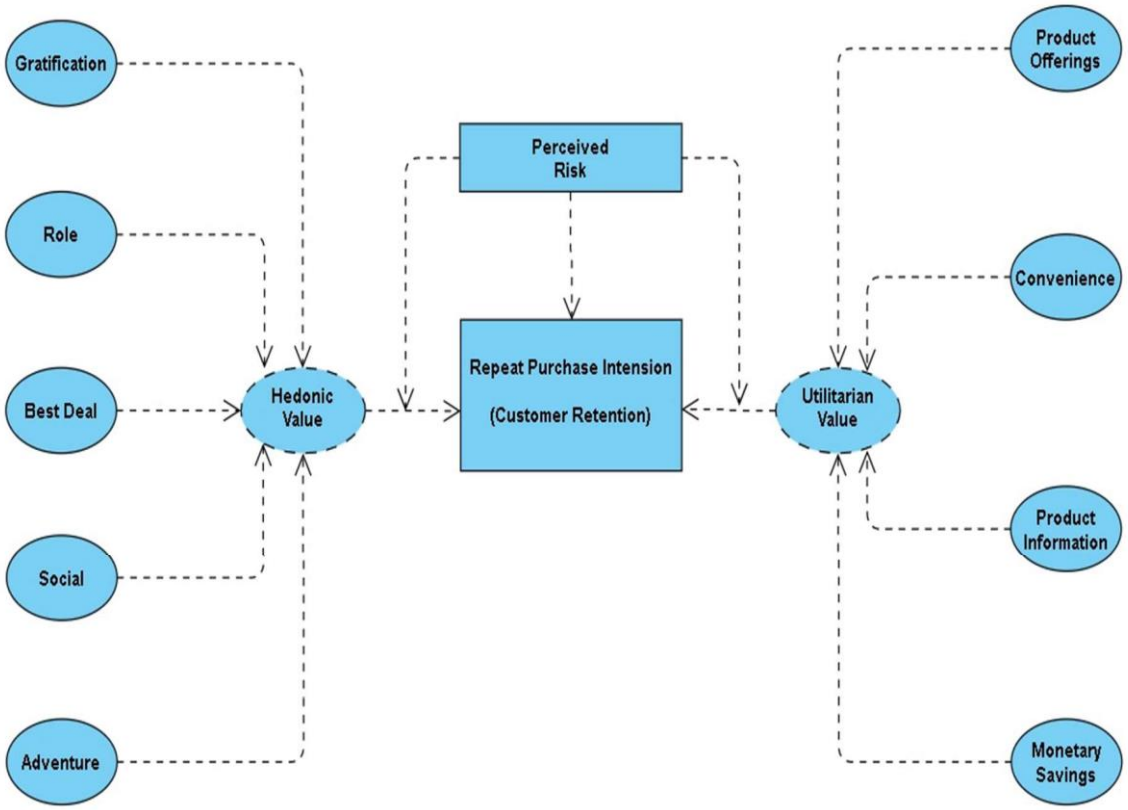
The research furthermore investigated the factors that influence the online customers repeat purchase intention.

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.

Use Case Diagram



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 re purchase 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 feedbacks to these websites on the basis of several positive and negative factors and also the details of the users on basis of factors like age, gender, city etc.

Benefits of Customer Retention:

1. Retention is cheaper than acquisition
    - While the 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, but it has also 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 will spread the word of a brand's exploits: 90% share their brand preferences online.

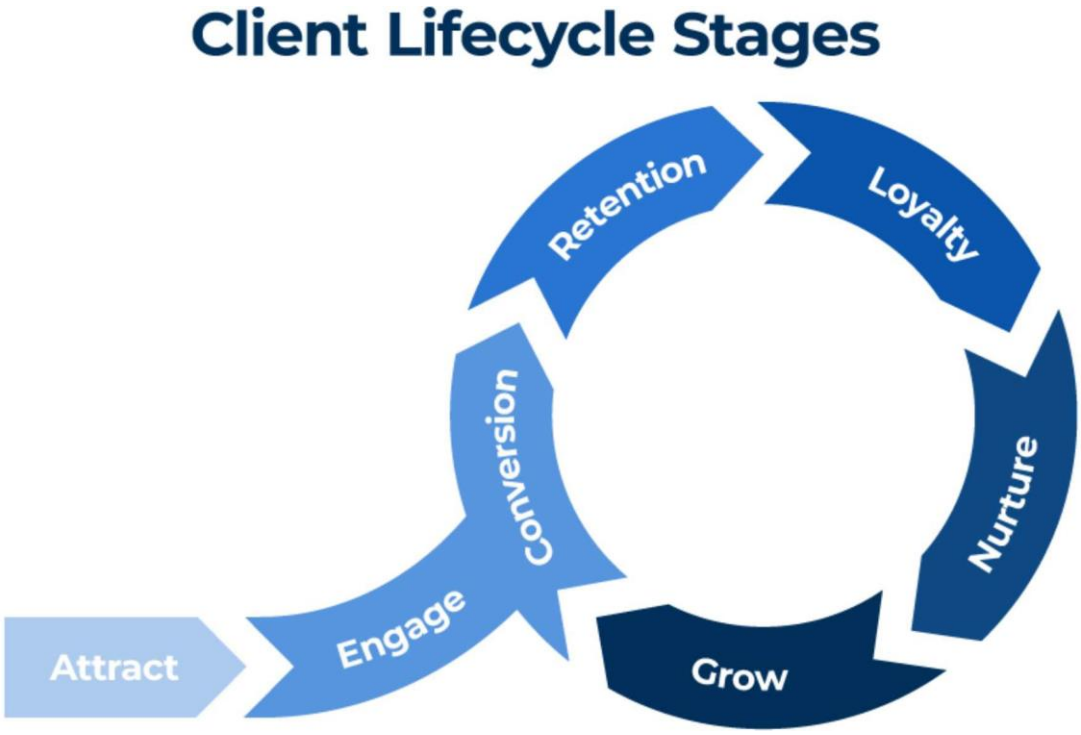
5. Engaged Customers Provide More Feedback
   - Feedback is critical to the success of any business.
   - Customers who provide feedbacks are often willing to give brands the benefit of the doubt.
   - They're telling you how to earn their business repeatedly. As research has shown, people who have complained 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.



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.

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.

Importing necessary Libraries

```
import pandas  as  pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

import os
import scipy as stats

import warnings
warnings.filterwarnings('ignore')
```

Then I separated the sheets present in our Excel spreadsheet and stored them in a dataframe variables.

```
#Importing the Dataset
df = pd.read_excel("customer_retention_dataset.xlsx")
pd.set_option("display.max_columns",None)
df
```

Exploratory Data Analysis (EDA):

After I got the dataset in our Jupyter Notebook I was able to notice that due to large number of rows and columns the information was truncated. Therefore, to overcome this challenge I used the pandas code.

I made sure to rename the column names that were ill formatted and quite long which made no sense to me. With the help of rename I was able to change the names of columns that were too lengthy and could have been accommodated in shorter formats.

I went ahead to look into each record information by making use of nunique methods.

```
# Checking number of unique values in each column
df.nunique().to_frame("No of Unique Values")
```

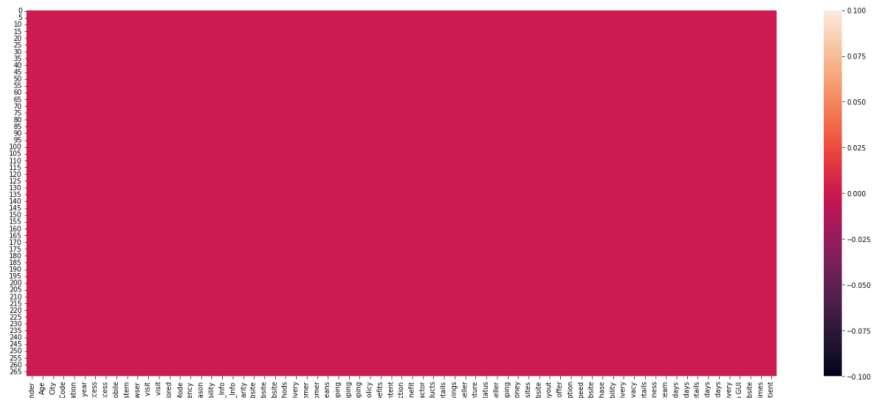
Now was the time to look at any kind of missing values or null value that might have been present in our dataset.

```
# Checking count of null values from the dataset.
df.isnull().sum()
```

Luckily, I was able to see that there were no missing values in our entire dataset that is prominently visible in the matrix visual below.

```
# Let's visualize the null values clearly through Heatmap
plt.figure(figsize=(25,10))
sns.heatmap(df.isnull())
```

```
<AxesSubplot:>
```



I used a for loop to look at all the unique values present in the categorical columns covering the number of rows in the dataset.

```
Catg_data = []
for x in df.dtypes.index:
    if df.dtypes[x] == 'object':
        Catg_data.append(x)
Catg_data
```

Visualization:

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:

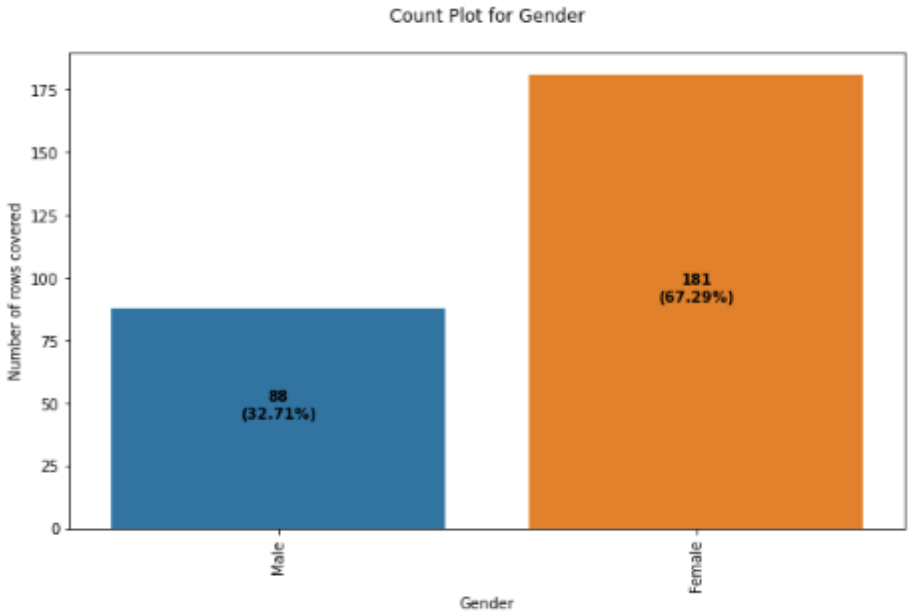
1. Univariate Analysis: In the univariate analysis, we will be using a single feature to analyze almost all of its properties.
2. Bivariate Analysis: When we compare the data between exactly 2 features then it is known as bivariate analysis.
3. Multivariate Analysis: In the multivariate analysis, we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

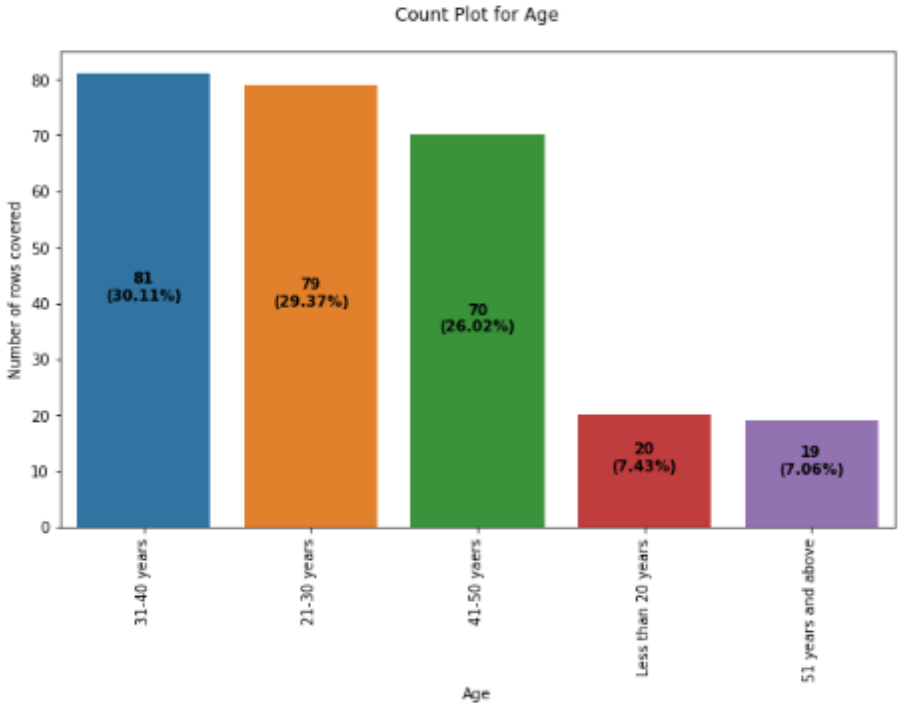
Univariate Analysis:

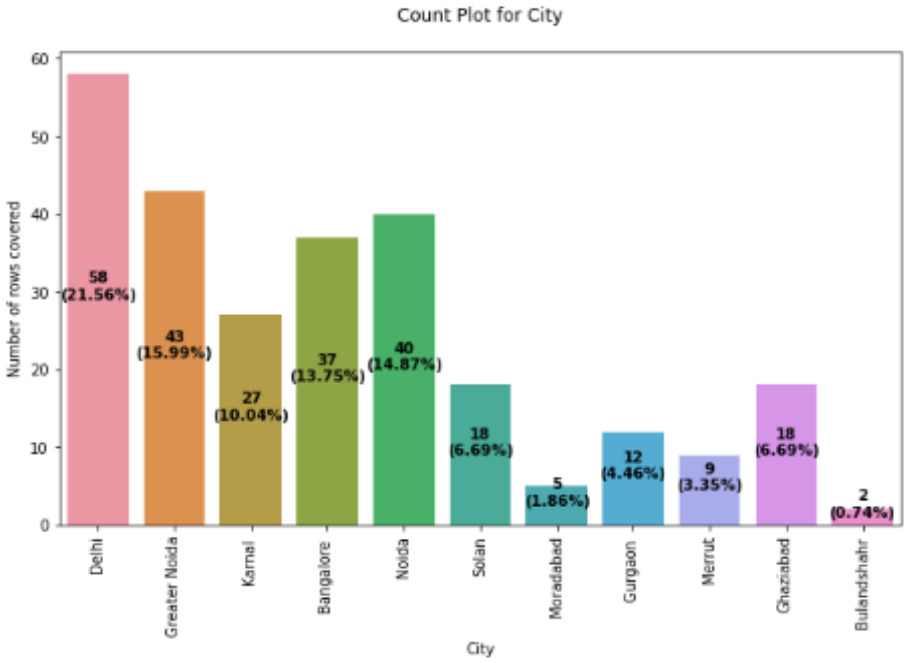
I made use of 2 for loops to generate count plots for all our columns showing the percentage of data coverage.

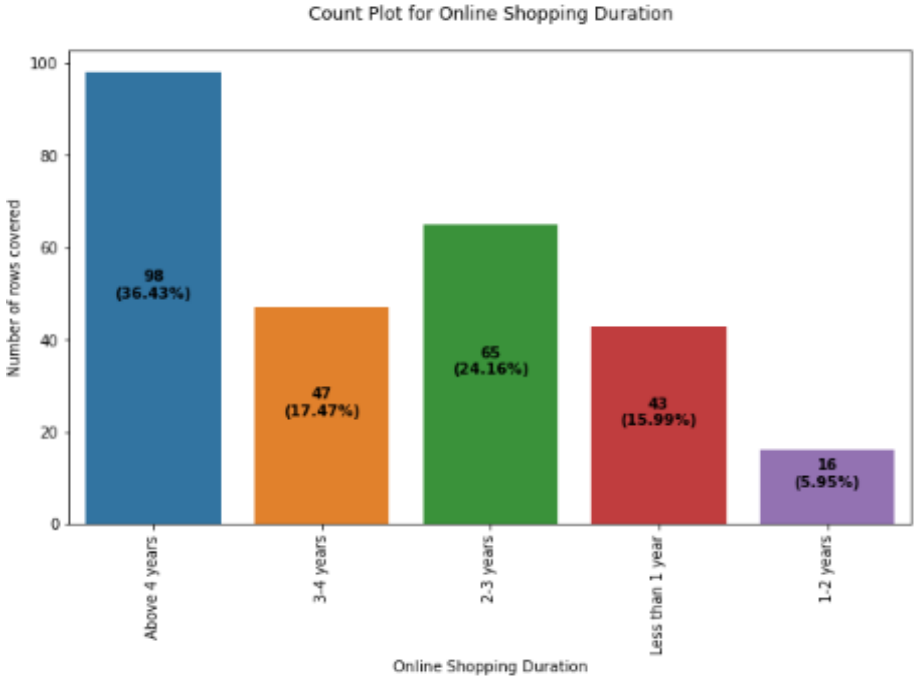
```
for col in df[Catg_data]:
    plt.figure(figsize=(10,6))
    col_name = col
    values = df[col_name].value_counts()
    index = 0
    ax = sns.countplot(df[col_name])

    for i in ax.patches:
        h = i.get_height() # getting the count of each value
        t = len(df[col_name]) # getting the total number of records using length
        s = f"{h}\n({round(h*100/t,2)}%)" # making the string for displaying in count bar
        plt.text(index, h/2, s, ha="center", fontweight="bold")
        index += 1

    plt.title(f"Count Plot for {col_name}\n")
    plt.xlabel(col_name)
    plt.ylabel(f"Number of rows covered")
    plt.xticks(rotation=90)
    plt.show()
```

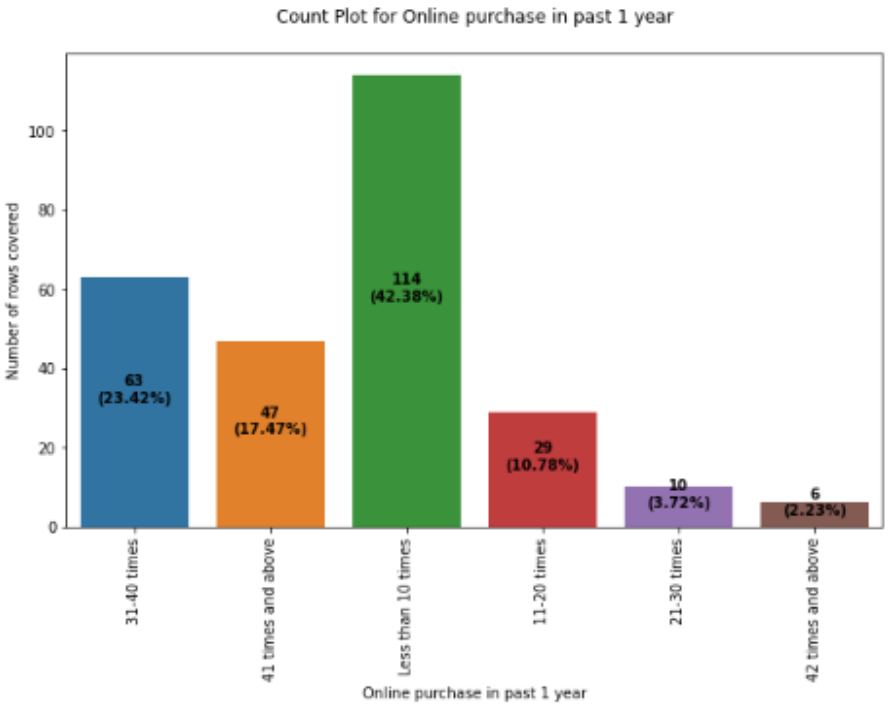
This piece of code generated multiple count plot images as displayed below (Only few images mentioned here).

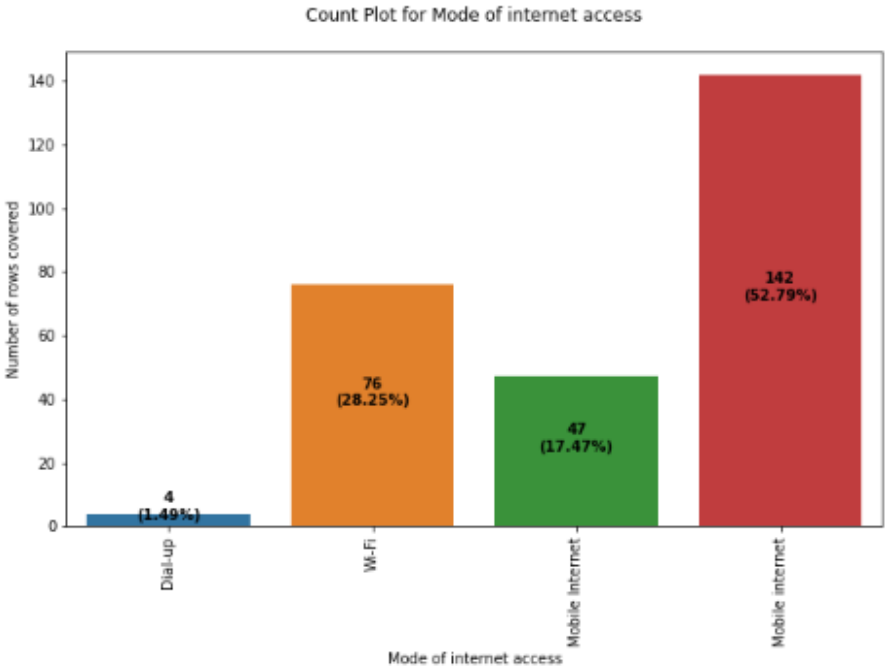

Count Plot for Gender
Number of rows covered
0
25
50
75
100
125
150
175
88
(32.71%)
181
(67.29%)
Male
Female
Gender

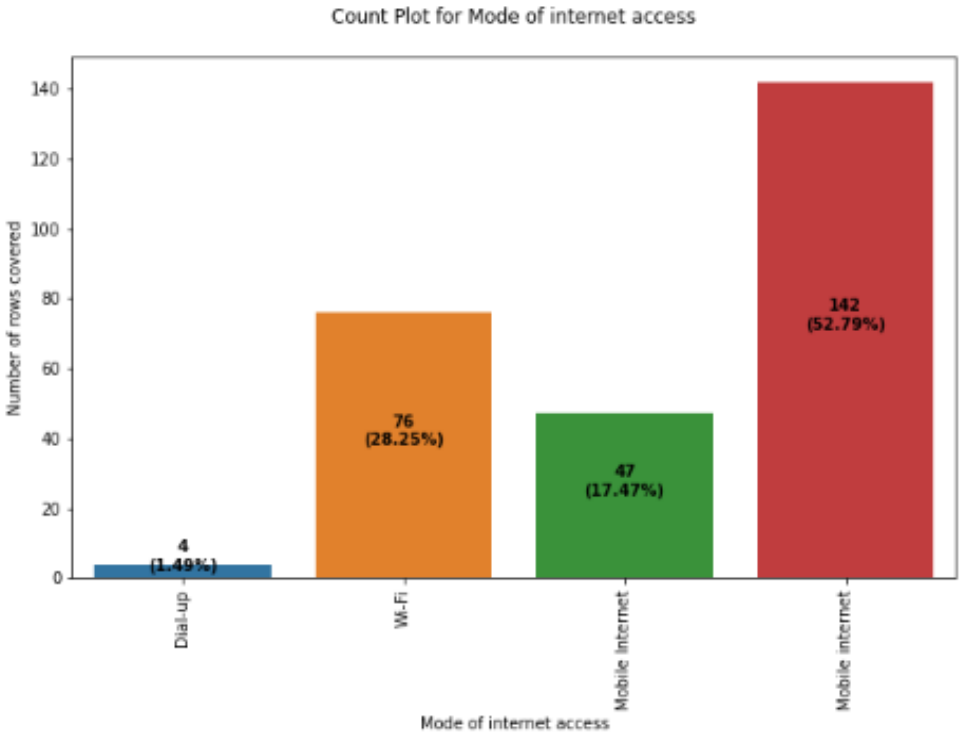

Count Plot for Age
Number of rows covered
0
10
20
30
40
50
60
70
80
81
(30.11%)
79
(29.37%)
70
(26.02%)
20
(7.43%)
19
(7.06%)
31-40 years
21-30 years
41-50 years
Less than 20 years
51 years and above
Age

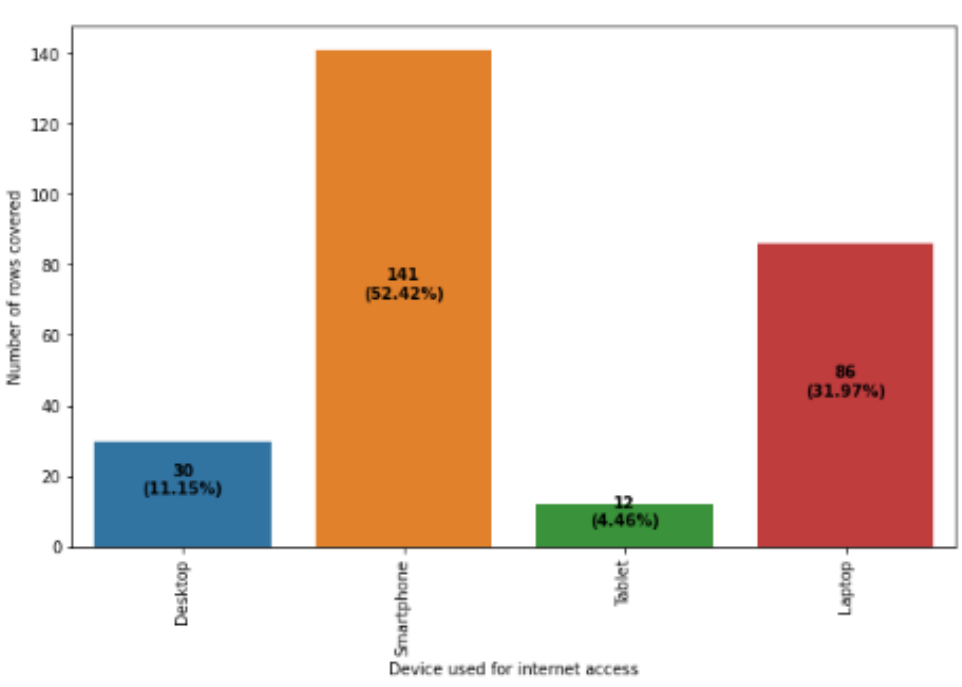

Count Plot for City
Number of rows covered
City
Delhi
Greater Noida
Karnal
Bangalore
Noida
Solan
Moradabad
Gurgaon
Meerut
Ghaziabad
Bulandshahr
58 (21.56%)
43 (15.99%)
27 (10.04%)
37 (13.75%)
40 (14.87%)
18 (6.69%)
5 (1.86%)
12 (4.46%)
9 (3.35%)
18 (6.69%)
2 (0.74%)


Count Plot for Online Shopping Duration
Number of rows covered
Online Shopping Duration
Above 4 years
3-4 years
2-3 years
Less than 1 year
1-2 years
98 (36.43%)
47 (17.47%)
65 (24.16%)
43 (15.99%)
16 (5.95%)

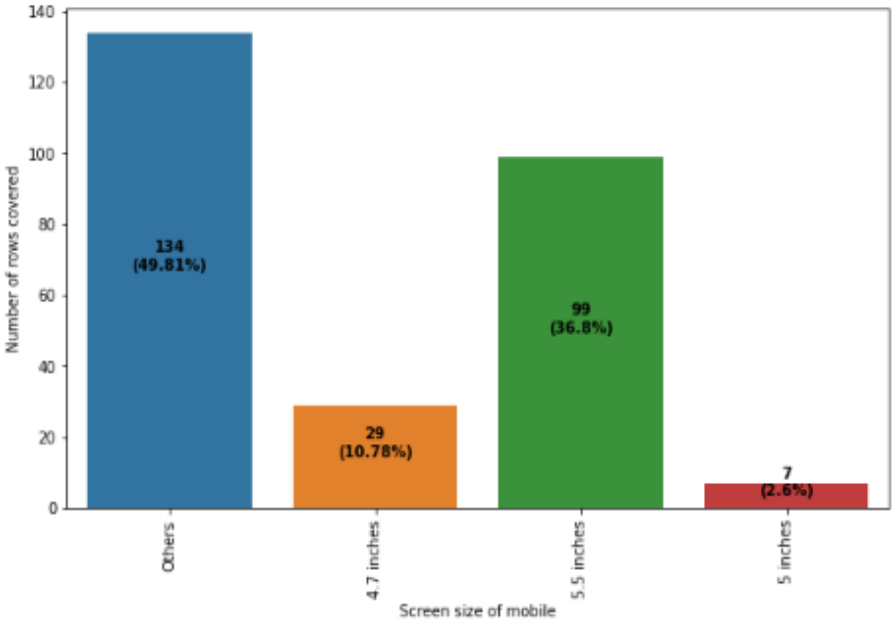

Count Plot for Online purchase in past 1 year
Number of rows covered
0
20
40
60
80
100
63
(23.42%)
47
(17.47%)
114
(42.38%)
29
(10.78%)
10
(3.72%)
6
(2.23%)
31-40 times
41 times and above
Less than 10 times
11-20 times
21-30 times
42 times and above
Online purchase in past 1 year


Count Plot for Mode of internet access
Number of rows covered
0
20
40
60
80
100
120
140
4
(1.49%)
76
(28.25%)
47
(17.47%)
142
(52.79%)
Dial-up
Wi-Fi
Mobile internet
Mobile internet
Mode of internet access

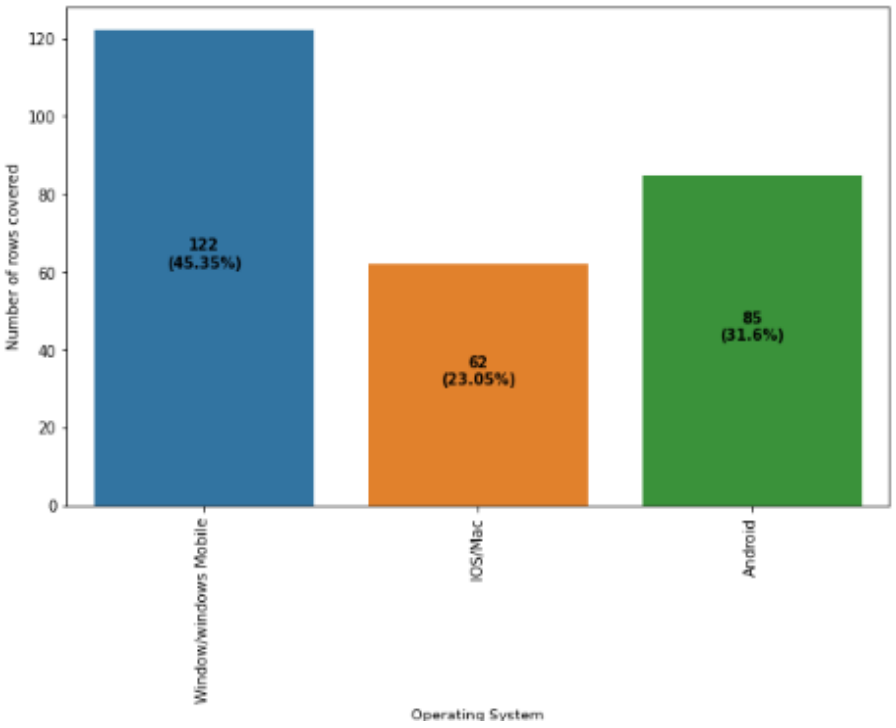


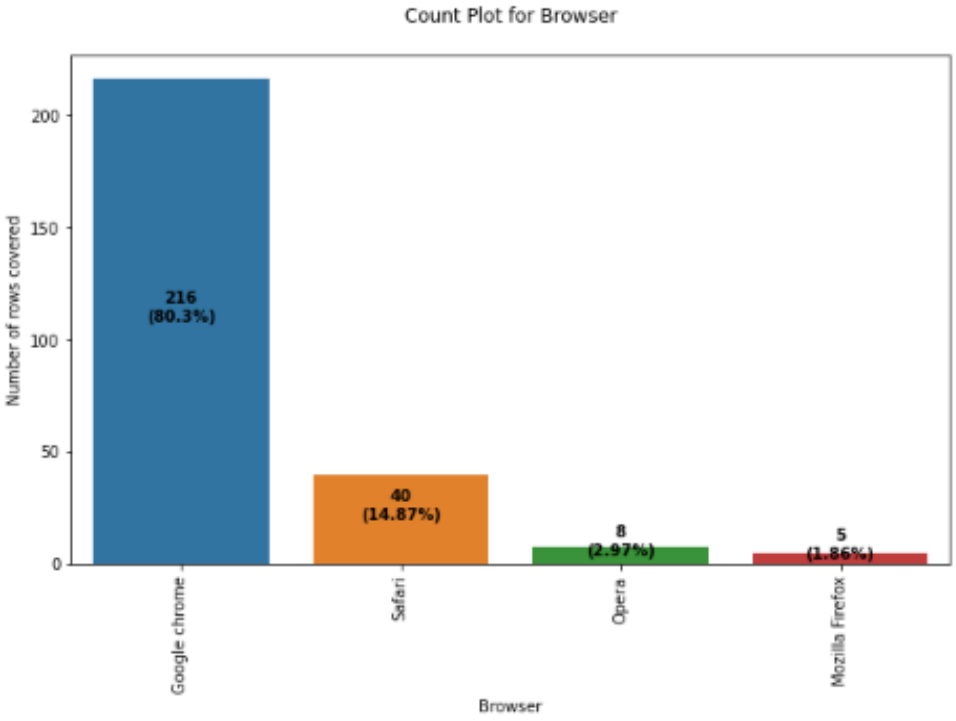


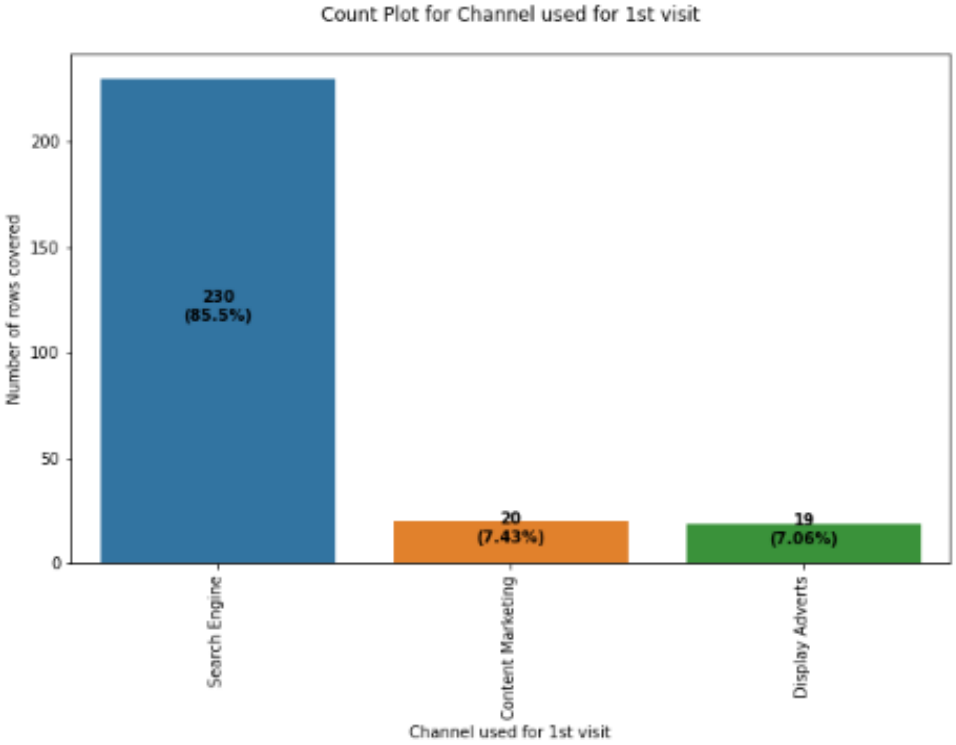
Count Plot for Screen size of mobile



Count Plot for Operating System




Count Plot for Browser
Number of rows covered
200
150
100
50
0
216
(80.3%)
40
(14.87%)
8
(2.97%)
5
(1.86%)
Google chrome
Safari
Opera
Mozilla Firefox
Browser


Count Plot for Channel used for 1st visit
Number of rows covered
200
150
100
50
0
230
(85.5%)
20
(7.43%)
19
(7.06%)
Search Engine
Content Marketing
Display Adverts
Channel used for 1st visit

Observations from the above plots:

o Female buyers are more in number than male in the Dataset.

o Buyer Age between 31 to 40 years is more followed by 21-30 years for online shopping.

o Delhi city has a maximum number of buyers followed by greater Noida & Noida. Bulandshahr city has a minimum number of buyers for online shopping.

o Tenure wise maximum buyers are preferred online shopping since 4+ years.

o Maximum number of online buyers purchased less than 10 products in past one year.

o Using mobile phone internet is more for online shopping.

o Smart Phone has been used more for online shopping followed by laptop.

o Smartphones having screen sizes other than 4.7, 5.5 & 5 inches had been used more times for online shopping.

o Devices having a windows operating system has been used often more for online shopping followed by android.

o Google Chrome browser is used maximum times to access the website as compared to other browsers.

o Search engine is more useful for the potential online buyers, also after the 1st visit to the store, the search engine has been used more to locate the website.

o Usually buyers explore the particular website more than 15 times before making a purchase decision.

o Credit/Debit has been used more than other options for online purchases.

o Maximum buyer chooses to add the product to the cart & leave without making payment because of a better alternative offer.

o Content & all relevant information on the website is easy to read & understand, also easy to navigate for a maximum number of the online buyer.

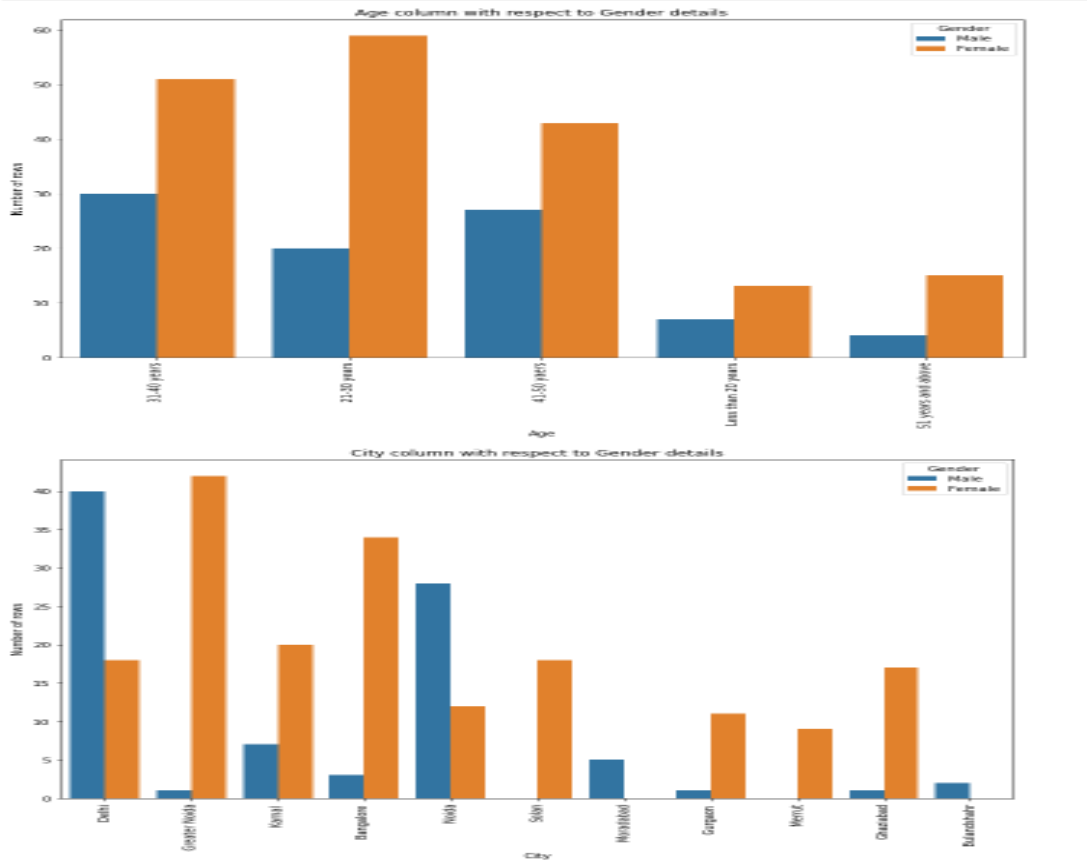
Bivariate Analysis:

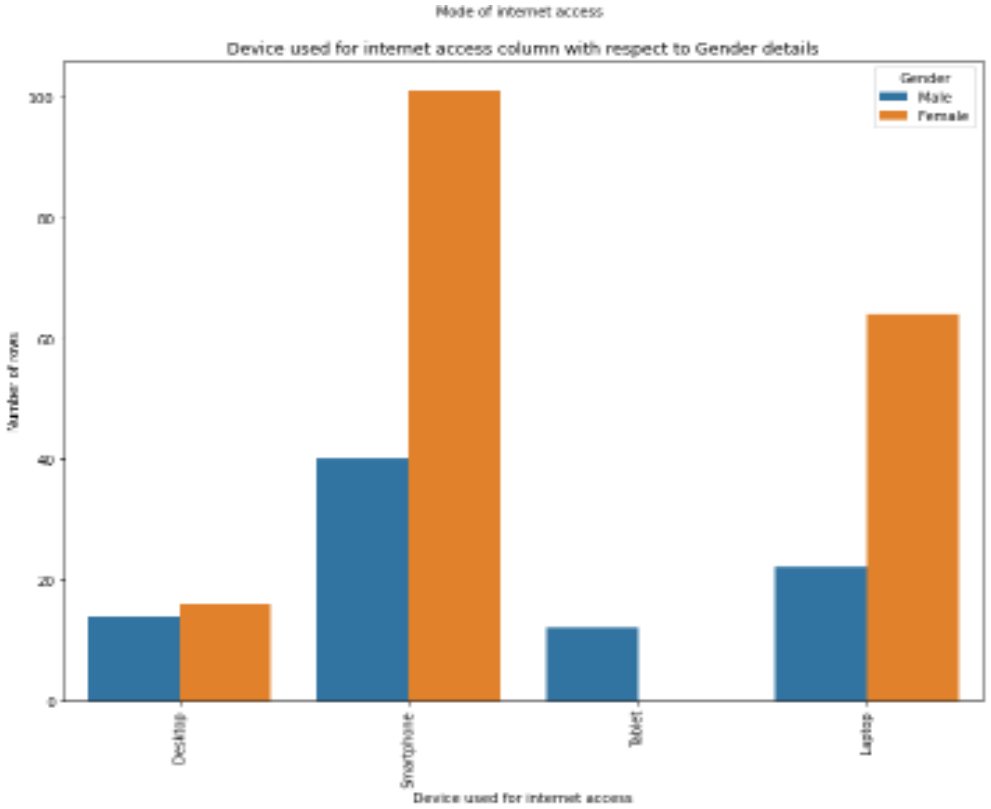
I performed bivariate analysis using count plots again and changing the hue format. Please refer the code and the outputs in GIF formats below.

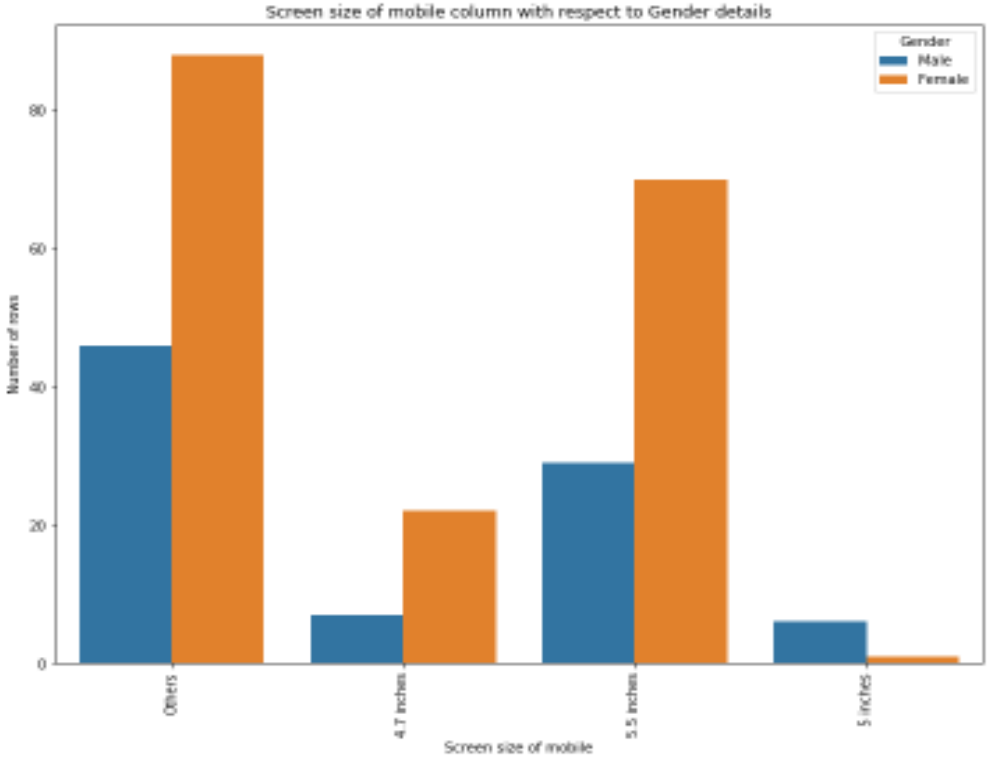
Code:

```
for col in df:
    if col == "Gender":
        pass
    elif col == "Pin Code":
        pass
    else:
        plt.figure(figsize=(10,8))
        sns.countplot(x=col, data=df, hue="Gender")
        plt.title("{} column with respect to Gender details".format(col))
        plt.tight_layout()
        plt.xticks(rotation=90)
        plt.ylabel("Number of rows")
        plt.show()
```

Output:




Mode of internet access
Device used for internet access column with respect to Gender details
Gender
Male
Female
Number of rows
0
20
40
60
80
100
Desktop
Smartphone
Tablet
Laptop
Device used for internet access


Screen size of mobile column with respect to Gender details
Gender
Male
Female
Number of rows
0
20
40
60
80
Others
4.7 inches
5.5 inches
5 inches
Screen size of mobile

Observations from the above plots:

• Female of Age 21-30 yers are much more in numbers than other age group.

• Delhi and Noida have many respondents are male than female.

• Both men and women shopping from desktop count are almost same. However, more women shop from either smartphone or laptop.

• Google Chrome has the highest usage for browsing shopping site.

• Most of the women come back to shopping website by using search engine.

• Most men prefer to use search engine and Url and app little less, however women prefer to use search engine or app, rather than direct URL

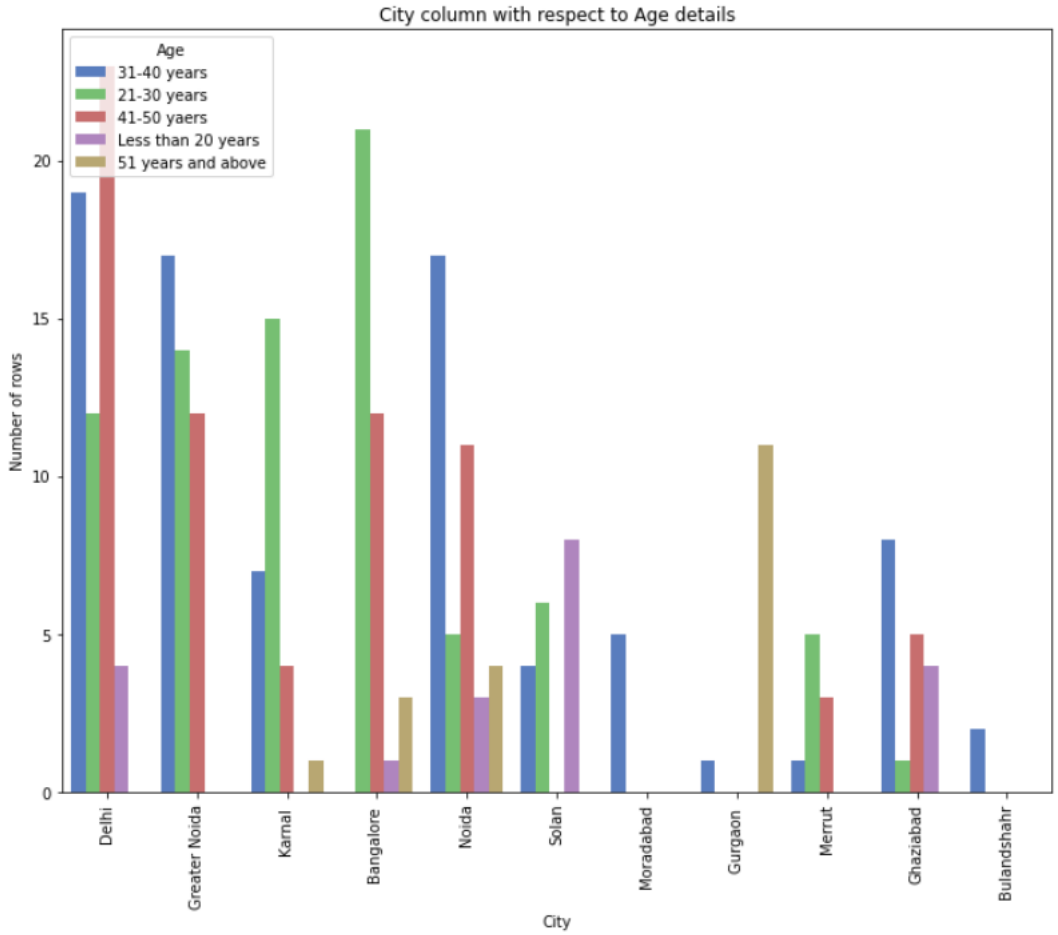
• Women spend more time than men during online shopping and the time is mostly more than 15 mins, followed by 11-15 mins.

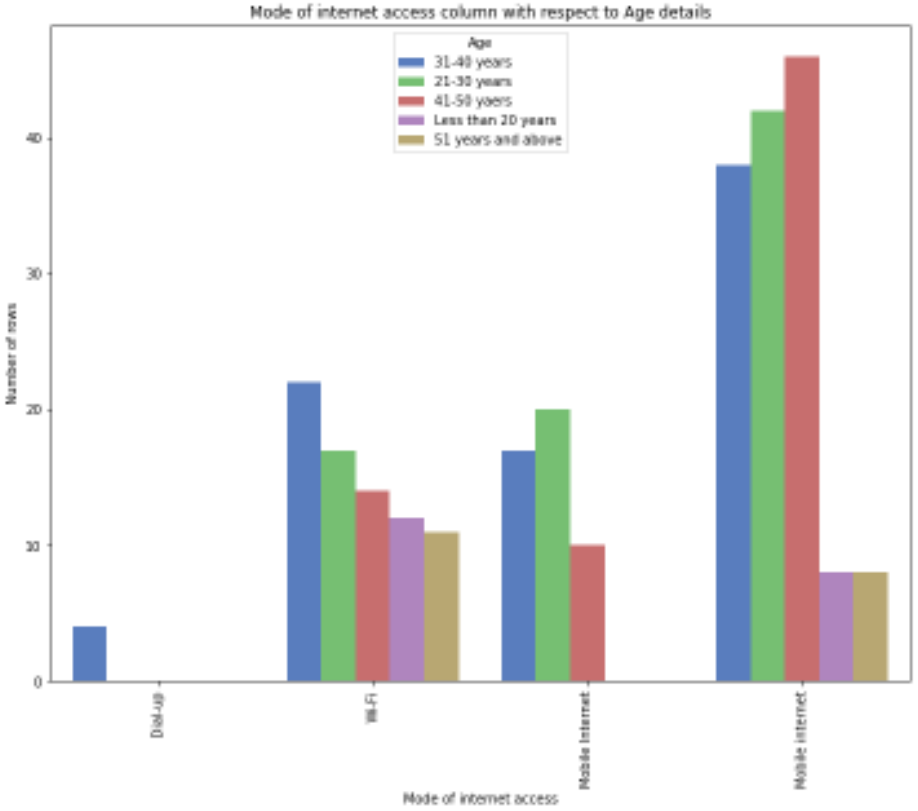
• Credit / Debit card is mostly preferred for payments

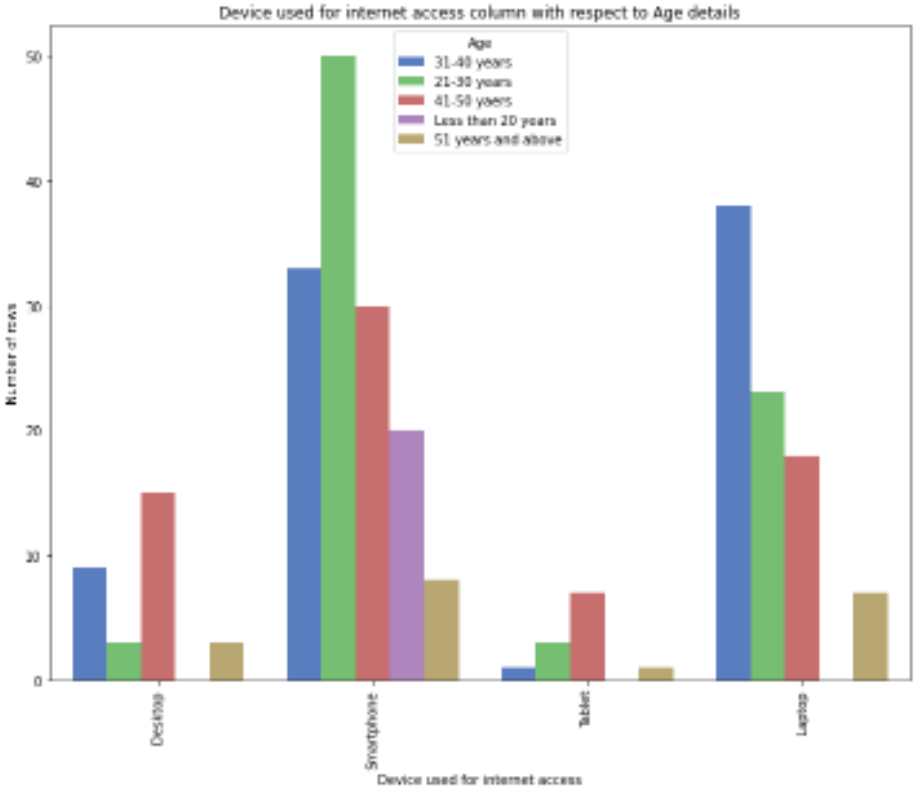
Code:

```
for col in df:
    if col == "Age":
        pass
    elif col == "Pin Code":
        pass
    else:
        plt.style.use('seaborn-muted')
        plt.figure(figsize=(10,8))
        sns.countplot(x=col, data=df, hue="Age")
        plt.title("{} column with respect to Age details".format(col))
        plt.tight_layout()
        plt.xticks(rotation=90)
        plt.ylabel("Number of rows")
        plt.show()
```

Output:




Mode of internet access column with respect to Age details
Age
31-40 years
21-30 years
41-50 years
Less than 20 years
51 years and above
Number of rows
0
20
40
Dial-up
Wi-Fi
Mobile Internet
Mobile internet
Mode of internet access


Device used for internet access column with respect to Age details
Age
31-40 years
21-30 years
41-50 years
Less than 20 years
51 years and above
0
20
40
50
Desktop
Smartphone
Tablet
Laptop
Device used for internet access

Observations from the above plots:

• From gender, female shoppers are more than male.

• From age 21-30 and 31-40 years of people use to do more shopping and peoples who are 50 years old and above they do very less shopping

• Delhi has the highest number of shoppers, however age group of 41-50 are more in Delhi than another city.

• Mobile internet usages are more for shopping and very few use dialup

• People use their smartphones more for shopping than laptops and very few uses tablet

• Mostly people use window phones then android and then less people use ios for inline shopping

• Maximum people use google chrome and very less people use firefox

• From After first visit, how do you reach the online retail store I can say people used search engine or app for this

• From How much time do you explore the e- retail store before making a purchase decision i can say mostly people use to take 15 min and very few people take 1 min only

• From What is your preferred payment Option i can say mostly people use credit/debit card then 2nd preference of people are cash on delivery then least they use E-wallet

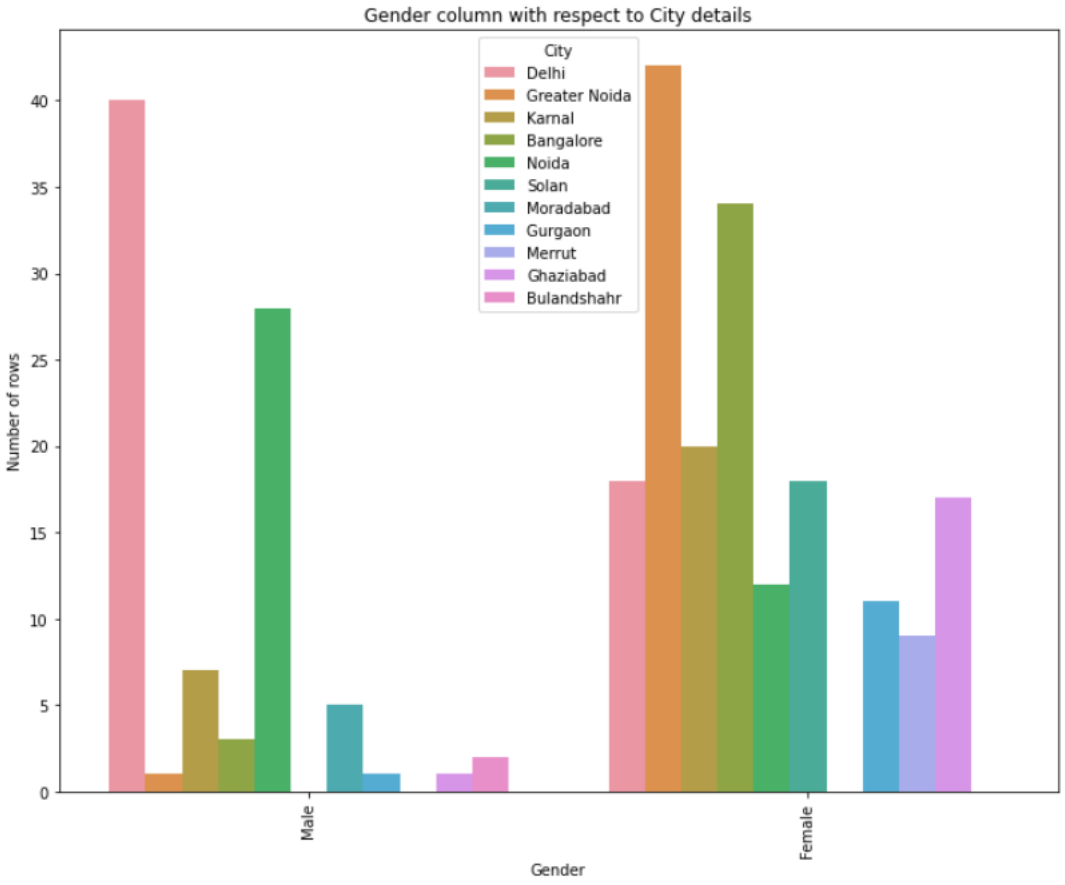
• Flipkart is mostly used by age group of 31-40

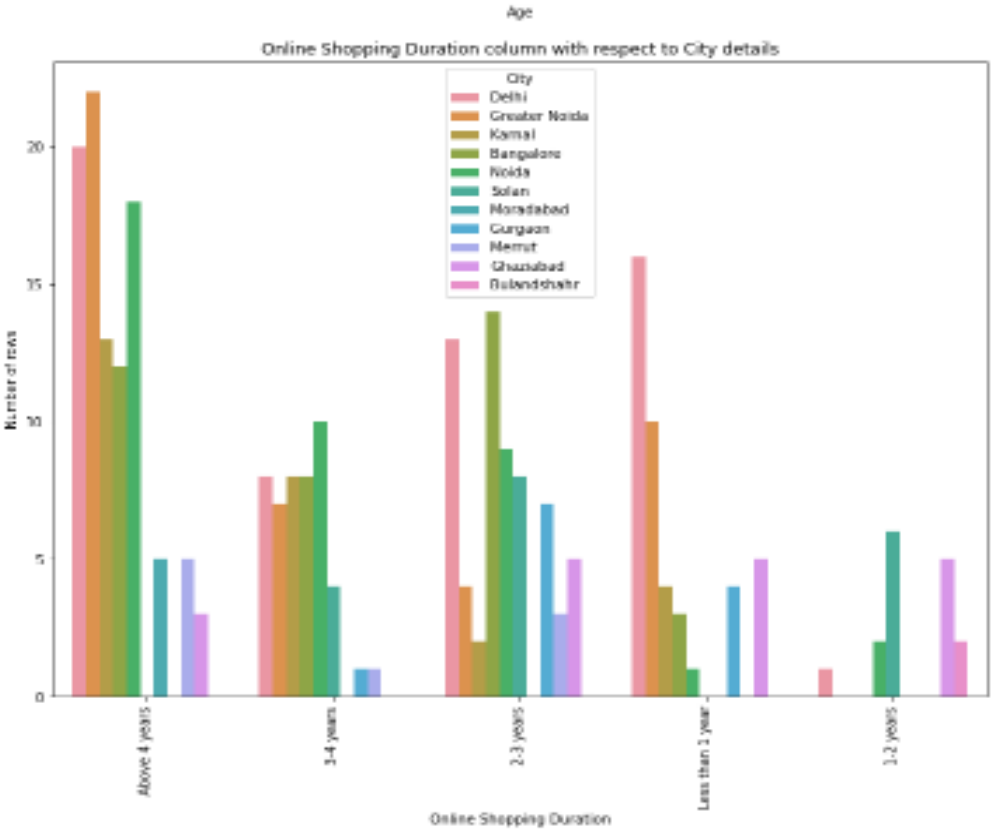
• Amazon is mostly preferred by 41-50 years of Age group

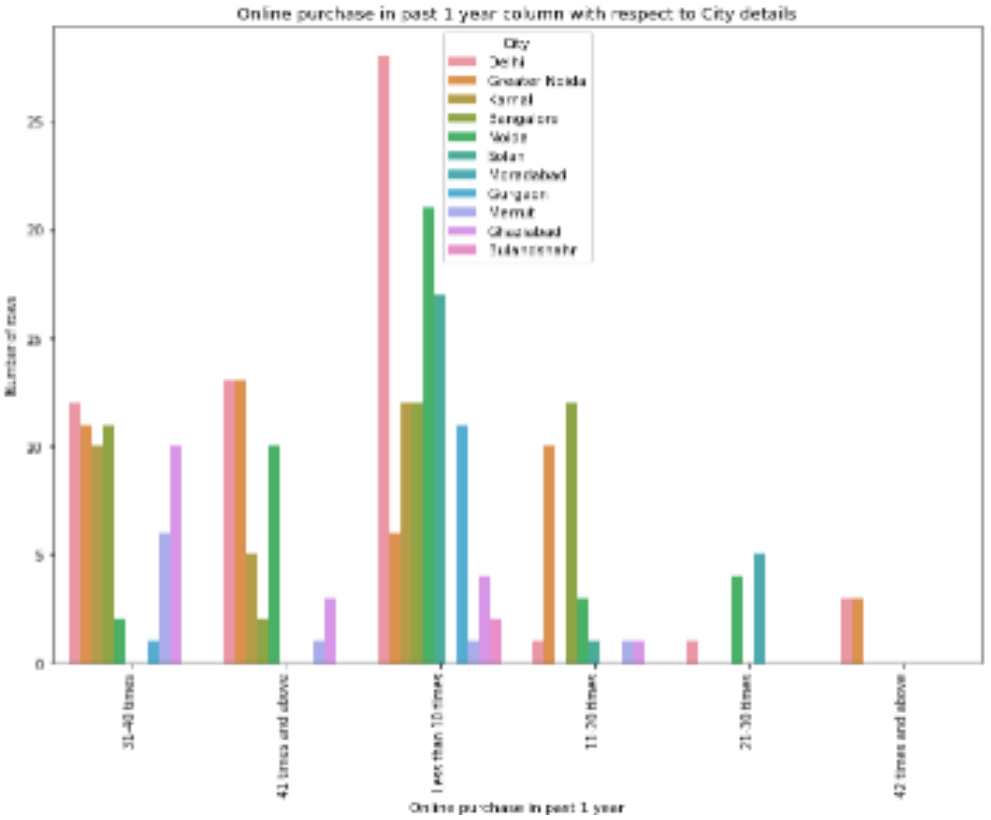
Code:

```
for col in df:
    if col == "City":
        pass
    elif col == "Pin Code":
        pass
    else:

        plt.figure(figsize=(10,8))
        sns.countplot(x=col, data=df, hue="City")
        plt.title("{} column with respect to City details".format(col))
        plt.tight_layout()
        plt.xticks(rotation=90)
        plt.ylabel("Number of rows")
        plt.show()
```

Output:




Age
Online Shopping Duration column with respect to City details
City
Delhi
Greater Noida
Karnal
Bangalore
Noida
Solan
Moradabad
Gurgaon
Meerut
Ghaziabad
Bulandshahr
Number of men
Above 4 years
3-4 years
2-3 years
Less than 1 year
1-2 years
Online Shopping Duration


Online purchase in past 1 year column with respect to City details
City
Number of men
31-40 times
41 times and above
Less than 10 times
11-20 times
21-30 times
42 times and above
Online purchase in past 1 year

Observations from the above plots:

• More Female are from Greater Noida followed by Bangalore. Whereas most men shoppers are from Delhi.

• Amongst the respondents, the major class targeted is between 21-40 years, followed by 41-50 and less than 20 years.

• The respondents are majorly residing in cities like Delhi, Greater Noida, Noida and Bangalore.

• Majority are shopping online for more than 4 years. There are considerable people who are shopping online since less than one year and it shows that many new customers are being added every year.

• We can observe that many have shopped less than 10 times in the past year.

• Many of them use mobile to shop online, followed by laptop, desktop, and tablet.

• Windows constitute the major OS of the customer device, followed by Android and Mac.

• Google Chrome is majorly used to access the shopping website.

• People are becoming customers of their favourite stores by using the search engine. Content marketing or display advertisements are not that impactful when coming to online marketing. So, companies should spend more on advertising on search engines.

• For repeated visits, people use search engine first, followed by app and direct URL. We can see that difference between app and search engine is small.

• Majority of the people spend more than 15 minutes before making a purchase, followed by 6-10 minutes.

• The major payment method used by all is credit/debit cards, followed by COD and e-wallets.

• People have mentioned that sometimes they would leave the cart without purchasing and the major reason they have mentioned is that they are finding

some better alternative offer. It means that people are comparing from many online websites before making any purchase.

• Customers strongly agree that content of website must be easy to read and understandable.

• Majority of customers want information of similar products to make purchase.

• Majority of the customers want complete information on listed sellers and their products being offered.

• Customers want all relevant information on the listed products and very less customers disagree to that.

• The customers wanted the websites to be easily navigated.

Then I performed Label Encoding on all the object datatype columns before I could proceed with any kind of multivariate analysis.

Code:

```
from sklearn.preprocessing import LabelEncoder

le=LabelEncoder()
for i in df.columns:
    if df[i].dtypes=="object":
        df[i]=le.fit_transform(df[i])
```

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

Finally, I have checked Outliers through Box plot and Correlation through Heatmap.

Interpretation of the Results:

The results that were interpreted from the visualization are as follows:

From the survey we found that the customers agreed with certain things like ecommerce websites have empathy towards them, content must be easy to read and understand, similar products should be highlighted for product comparison, payment convenience, trustworthy and they felt gratified while shopping etc.

The respondents were asked to give ratings and feedback regarding certain statements that may be used to describe their ideal online store. We have found answers from customers like which websites among listed sites they are familiar with, they find easy to use, reliable, secured, about their delivery, change in web design, etc.

From the visualization we have found Amazon.in was the most used and favorite websites for the customers and Flip Kart was the second favourite webapp while snapdeal.com and Paytm.com were the least used websites by the customers.

Conclusion:

• The dataset is to identify the retention factors of customers in online shopping

• In this project, I have renamed the columns to a shorter name, found no null values in columns. Visualized and analyse the data using count plot and encoded the object data into numerical using label encoding method. Checked the statistical summary of the dataset, checked the corelation between features, outliers, and skewness in the dataset.

• Based on overall observations, the first few 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 Gender wise, City, Age group etc. It also showed that in some factors there is less importance given to contribute to the success of an e-commerce store, so based on that we could remove those factors & keep all the important factors.

• Last few 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 been 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 has been used mostly for Online Shopping. So, based on the research factors, Amazon & Flipkart are the e-commerce platform, which are having the combination of both utilitarian and hedonistic values to keep the repeat purchase intention (loyalty) positively.

• It was also observed that online shopping is not trustworthy and reliable to some consumers due to online payment system and personal privacy. In addition, online security is a major concern for the consumer particularly in terms of fraud, privacy, and hacking.