CUSTOMER RETENTION DATASET

Submitted by:

Seema Tomar

CONTENTS

INTRODUCTION

Customer Retention refers to the action and strategies a business uses to try and keep existing customers. To enable these actions, customer retention analytics provide predictive metrics of which customer might churn-which enable them to get ahead of it. 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.
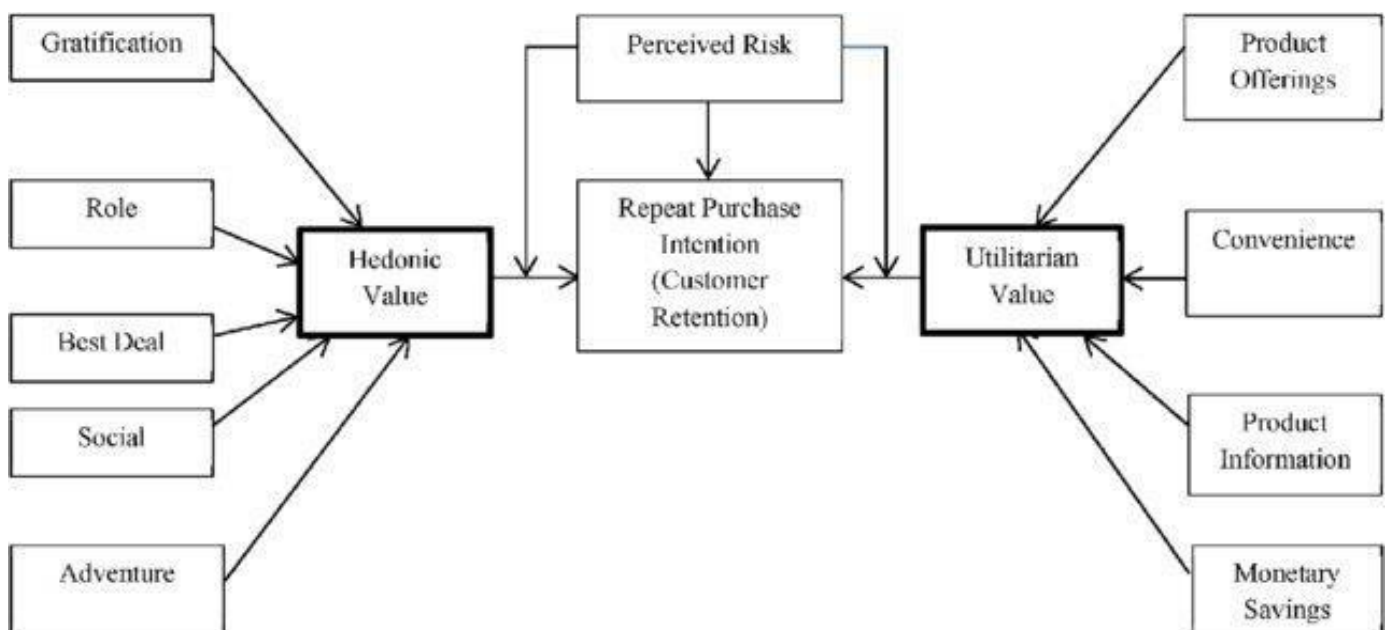
1. What is Customer Retention?

Customer Retention refers to a company's ability to turn customers into repeat buyers and prevent them from switching to a competitor. It indicates whether your product and the quality of the service please your existing customers. 3 Customer Retention strategies are the processes and initiatives businesses put in place to build customer loyalty and improve customer lifetime value.

2. Why is customer retention important?

Customer retention increases your customers' lifetime value and boosts your revenue. It also helps you build amazing relationship with your customers. You aren't just another website or store. They trust you

with their money because you give them value in exchange. According to the Harvard Business Review, acquiring a new customer can be 5 to 25 times more expensive than holding on to an existing one. You 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 the initial sale.

3. Customer Retention benefits:

• Retention is Cheaper than Acquisition

• Loyal Customers are more profitable.

• Your Brand will stand out from the crowd.

• Engage customers provide more feedback.

• Loyal customers are more forgiving.

• Customers will explore your brand

REVIEW OF LITERATURE

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 ecommerce 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.

Purchase intention: Purchase intention can be characterized as the probability that an online client will purchase an item and keep on buying items from a similar online retailer. This essential develop "customers purchase choice" is clarified in the model regard to its relationship with another key factors.

User satisfaction: Satisfaction has been characterized as an individual sensation of happiness or delight, because of contrasting the genuine result of shopping on the web as against their assumption. Satisfaction of on the web customers will emphatically affect their future intention (repeat purchase) straightforwardly or by implication.

Net advantages: This is utilized to portray the absolute advantages of got when shopping online by customers, as against the expenses (for instance, exertion, cash and time). A portion of the net advantages got from web-based shopping include: money related increase, item search and purchasing adequacy, accommodation, ideal conveyance, happiness, wide item reach, and adaptability. Agreeing to an examination by Childers, net advantage of shopping online incorporates both utilitarian worth and gluttonous qualities.

Data quality: This can be characterized as the customers' insights towards the show also, qualities of data introduced or showed in an online business site, web application or gateway. It takes the qualities of data into thought, for instance, practicality, pertinence reasonable, exactness, and fulfilment. One of the essential obligations of an e-posterior is to guarantee the accessibility of complete data on items being offered, the value-based cycle included, and administrations to be delivered. Showing excellent data on the site can invigorate a further developed change rate too as better customers' satisfaction.

Framework quality: This can be portrayed as the users' impression of the online retail site's capacity to proficiently give mentioned data just as its conveyance strategy. Framework quality examinations the exhibition of the site; a portion of the measurement incorporates; page load speed, simplicity of route between pages, web architecture, appearance, site accessibility, and site format. Innovation acknowledgment model (TAM) states that, an e-retail site which is viewed as being not difficult to utilize is most liable to achieve a 5 sensation of energy among the shoppers. Szymanski furthermore, likewise believed that the exhibition of a site plays out a critical job in choosing satisfaction of the customers while shopping on the site. At the point when customers purchase an item from a site or web application, usefulness obstacles   may bring about a helpless client user experience. Framework quality emphatically impacts online consumer loyalty. Administration quality Administration quality alludes to how well the administrations conveyed by an on the web retail location can coordinate with the assumptions for the client.

Analytical Problem Framing

The studies were identified using different keywords that are related mainly to customer retention and analysing other factors such as satisfaction, trust, and commitment in several searching processes until reaching a suitable number of related and targeted articles and studies to select and analyse. In the dataset there was no numerical columns and all the categorical columns has been analysed using all categorical plots to get better insight on the dataset. After visualizing the features it bought us a good insight that what actually customers are expecting from the e-commerce sellers.

```
In [7]: df
Out[7]:
```

	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	10 What is the operating system (OS) of your device? \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 the first time?	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 decision?	1
0	Male	31-40 years	Delhi	110009	Above 4 years	31-40 times	Dial-up	Desktop	Others	Window/windows Mobile	Google chrome	Search Engine	Search Engine	6-10 mins	F
1	Female	21-30 years	Delhi	110030	Above 4 years	41 times and above	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	Google chrome	Search Engine	Via application	more than 15 mins	C
2	Female	21-30 years	Greater Noida	201308	3-4 years	41 times and above	Mobile Internet	Smartphone	5.5 inches	Android	Google chrome	Search Engine	Via application	11-15 mins	F
3	Male	21-30 years	Karnal	132001	3-4 years	Less than 10 times	Mobile Internet	Smartphone	5.5 inches	IOS/Mac	Safari	Search Engine	Search Engine	6-10 mins	C
4	Female	21-30 years	Bangalore	530068	2-3 years	11-20 times	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	Safari	Content Marketing	Via application	more than 15 mins	C
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
264	Female	21-30 years	Solan	173212	1-2 years	Less than 10 times	Mobile Internet	Smartphone	5.5 inches	Android	Opera	Display Adverts	Social Media	more than 15 mins	
265	Female	31-40 years	Ghaziabad	201008	1-2 years	31-40 times	Mobile Internet	Smartphone	Others	Android	Google chrome	Search Engine	Via application	Less than 1 min	
		41-				Less than	Mobile			Window/windows	Google	Search	Search	more than	C

Data:

 In this project we are given excel file containing dataset of customer retention. There are 71 columns by which we can analyse how e-commerce sector can look after its customers to improve their service.

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?
Paytm.com	Amazon.in	Amazon.in	Flipkart.com	Flipkart.com	Amazon.in	Paytm.com	Flipkart.com	Amazon.in	Amazon.in	Flipkart.com
Amazon.in, Flipkart.com, Myntra.com	Amazon.in, Flipkart.com	Myntra.com	snapdeal.com	Snapdeal.com	Snapdeal.com	Snapdeal.com	Amazon.in	Myntra.com	Amazon.in, Flipkart.com	Amazon.in, Myntra.com
Myntra.com	Myntra.com	Myntra.com	Myntra.com	Myntra.com	Amazon.in	Paytm.com	Paytm.com	Paytm.com	Amazon.in	Amazon.in, Paytm.com, Myntra.com
Amazon.in, Flipkart.com, Myntra.com, Snapdeal	Snapdeal.com	Myntra.com, Snapdeal.com	Myntra.com	Paytm.com	Paytm.com	Paytm.com	Amazon.in, Flipkart.com	Amazon.in, Flipkart.com	Amazon.in, Flipkart.com, Paytm.com	Amazon.in, Flipkart.com
Amazon.in, Myntra.com	Flipkart.com, Paytm.com	Paytm.com	Paytm.com	Paytm.com	Snapdeal.com	Paytm.com	Amazon.in	Snapdeal.com	Paytm.com	Amazon.in, Myntra.com
...	...	...	...	...	...	...	...	...	...	...
Amazon.in	Amazon.in	Amazon.in	Amazon.in	Amazon.in	Amazon.in	Amazon.in	Amazon.in	Amazon.in	Amazon.in	Amazon.in
Flipkart.com	Flipkart.com	Flipkart.com	Flipkart.com	Flipkart.com	Flipkart.com	Flipkart.com	Flipkart.com	Flipkart.com	Flipkart.com	Flipkart.com
Amazon.in	Amazon.in	Snapdeal.com	Amazon.in	Snapdeal.com	Snapdeal.com	Snapdeal.com	Snapdeal.com	Snapdeal.com	Amazon.in	Amazon.in

All the variables in the dataset are of object type except pincode, however it is categorical in nature. The dataset contains 269 rows and 71 columns. The data contains no null value however there is a lot of unwanted characters in the columns.

Hardware and Software Requirements and Tools Used

• A mid level computer that runs on Intel i3- i5 8th generation, 4gb ram or A10/A11or any other equivalent chipset and a suitable processor.

• Juypter Notebook/Google chrome

• Libraries and packeges used:

import pandas as pd

import numpy as np

import matplotlib. pyplot as plt

import seaborn as sn

 import warnings warnings.filterwarnings("ignore")

We will use Python through Jupyter notebook for data processing. Also we will use liabraries here is numpy, matplotlib, pandas and seaborn. The matpotplotlib and seaborn library has been used to make charts to visualize and understand the problem, correlation, outliers and many other things, the pandas and numpy library issued to handel dataset and perform various tasks.

Exploratory Data Analysis:

```
In [20]:  #checking the column 1Gender of respondent
          plt.subplots(figsize=(10,8))
          sns.countplot(x="1Gender of respondent", data=df)
          plt.title("Countplot of Gender of respondent")
          plt.xticks(rotation=40)
          plt.xlabel('1Gender of respondent')
          plt.ylabel("count")
          plt.show()

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

Countplot of Gender of respondent
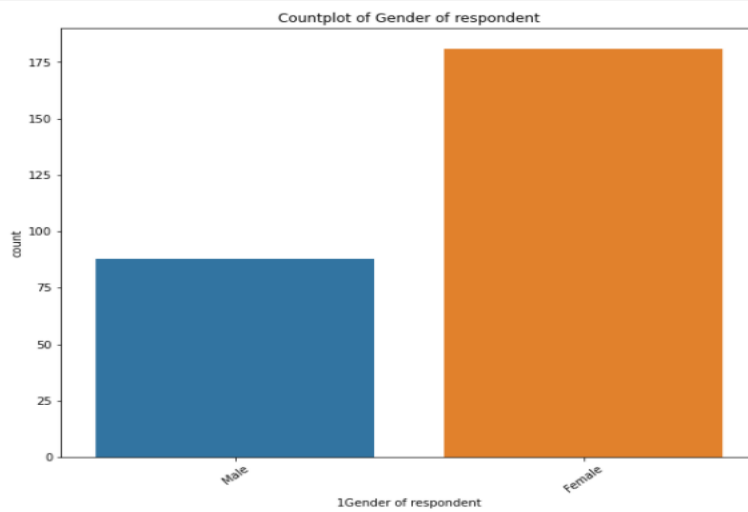

```
Out[20]:  Female    181
          Male       88
          Name: 1Gender of respondent, dtype: int64
```

- **Majority, 181 of the customers are Female whereas Male are 88.**

```
In [21]:  #checking the column 2 How old are you?
          plt.subplots(figsize=(10,8))
          sns.countplot(x="2 How old are you? ", data=df)
          plt.title("Countplot of 2 How old are you? ")
          plt.xticks(rotation=40)
          plt.xlabel('2 How old are you? ')
          plt.ylabel("count")
          plt.show()

          df['2 How old are you? '].value_counts()
```

Countplot of 2 How old are you?
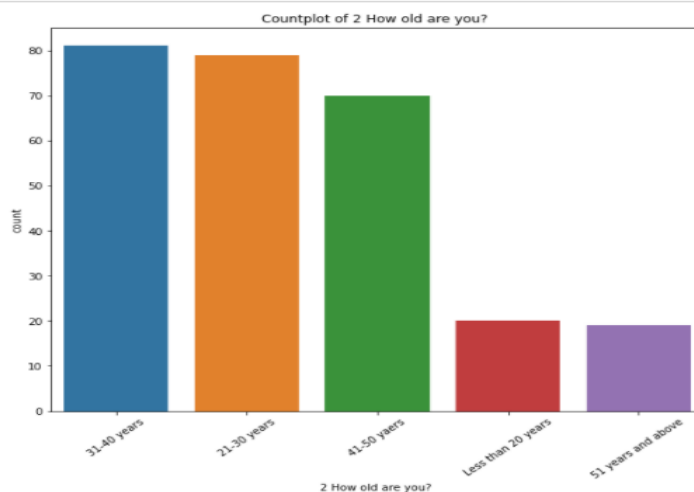

```
Out[21]:  31-40 years          81
          21-30 years          79
          41-50 yaers          70
          Less than 20 years   20
          51 years and above   19
          Name: 2 How old are you? , dtype: int64
```

- **Majority, 81 of the customers are from age group 31-40 years.**

```
In [22]: #checking the column 3 Which city do you shop online from?
         plt.subplots(figsize=(10,8))
         sns.countplot(x="3 Which city do you shop online from?", data=df)
         plt.title("3 Which city do you shop online from?")
         plt.xticks(rotation=40)
         plt.xlabel('3 Which city do you shop online from?')
         plt.ylabel("count")
         plt.show()

         df['3 Which city do you shop online from?'].value_counts()
```


3 Which city do you shop online from?
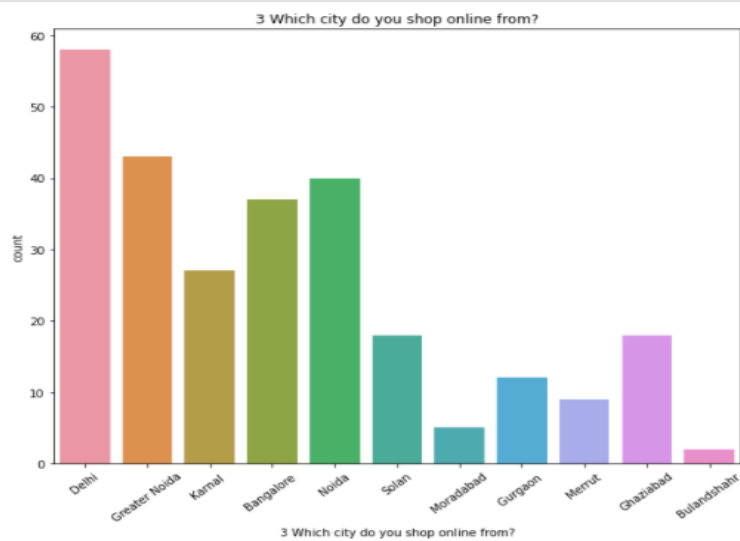
```
Out[22]: Delhi           58
         Greater Noida   43
         Noida           40
         Bangalore       37
         Karnal          27
         Solan           18
         Ghaziabad       18
         Gurgaon         12
         Merrut           9
         Moradabad        5
         Bulandshahr      2
         Name: 3 Which city do you shop online from?, dtype: int64
```

- **Majority, 58 of the customers placed the order at Delhi city.**

```
In [24]: #checking the column 5 Since How Long You are Shopping Online ?
         plt.subplots(figsize=(8,6))
         sns.countplot(x="5 Since How Long You are Shopping Online ?", data=df)
         plt.title("5 Since How Long You are Shopping Online ?")
         plt.xticks(rotation=30)
         plt.xlabel('5 Since How Long You are Shopping Online ?')
         plt.ylabel("count")
         plt.show()

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


5 Since How Long You are Shopping Online ?

```
Out[24]: Above 4 years      98
         2-3 years          65
         3-4 years          47
         Less than 1 year   43
         1-2 years          16
         Name: 5 Since How Long You are Shopping Online ?, dtype: int64
```
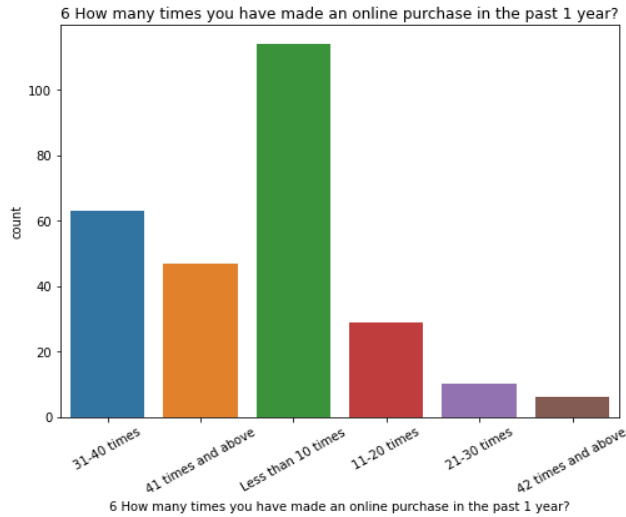
- **Majority, 98 customers are shopping since above 4 years.**

```
In [25]: #checking the column 6 How many times you have made an online purchase in the past 1 year?
         plt.subplots(figsize=(8,6))
         sns.countplot(x="6 How many times you have made an online purchase in the past 1 year?",data=df)
         plt.title("6 How many times you have made an online purchase in the past 1 year?")
         plt.xticks(rotation=30)
         plt.xlabel('6 How many times you have made an online purchase in the past 1 year?')
         plt.ylabel("count")
         plt.show()

         df['6 How many times you have made an online purchase in the past 1 year?'].value_counts()
```



```
Out[25]: Less than 10 times    114
         31-40 times            63
         41 times and above     47
         11-20 times            29
         21-30 times            10
         42 times and above      6
         Name: 6 How many times you have made an online purchase in the past 1 year?, dtype: int64
```

- **Majority 114 of the customers have made less than 10 times online purchase in the past 1 year.**

```
In [26]: #checking the column 7 How do you access the internet while shopping on-line?
         plt.subplots(figsize=(8,6))
         sns.countplot(x="7 How do you access the internet while shopping on-line?", data=df)
         plt.title("7 How do you access the internet while shopping on-line?")
         plt.xticks(rotation=30)
         plt.xlabel('7 How do you access the internet while shopping on-line?')
         plt.ylabel("count")
         plt.show()

         df['7 How do you access the internet while shopping on-line?'].value_counts()
```
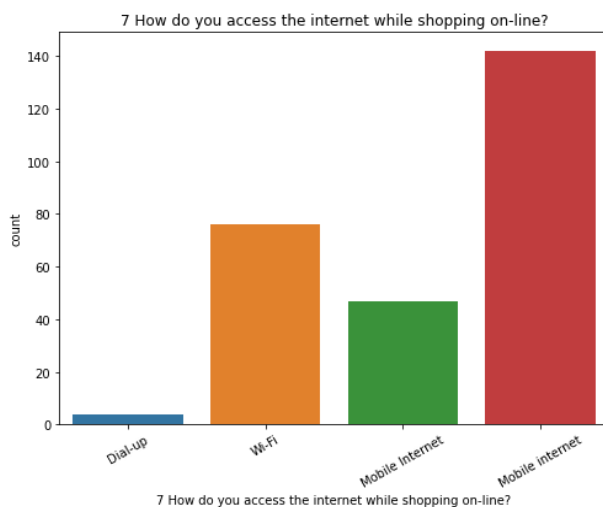


```
Out[26]: Mobile internet    142
         Wi-Fi               76
         Mobile Internet     47
         Dial-up              4
         Name: 7 How do you access the internet while shopping on-line?, dtype: int64
```

- **Majority, 189 customers use Mobile internet while shopping online.**

```
In [27]: #checking the column 8 Which device do you use to access the online shopping?
         plt.subplots(figsize=(8,6))
         sns.countplot(x="8 Which device do you use to access the online shopping?", data=df)
         plt.title("8 Which device do you use to access the online shopping?")
         plt.xticks(rotation=30)
         plt.xlabel('8 Which device do you use to access the online shopping?')
         plt.ylabel("count")
         plt.show

         df['8 Which device do you use to access the online shopping?'].value_counts()
```

```
Out[27]: Smartphone    141
         Laptop         86
         Desktop        30
         Tablet         12
         Name: 8 Which device do you use to access the online shopping?, dtype: int64
```


8 Which device do you use to access the online shopping?
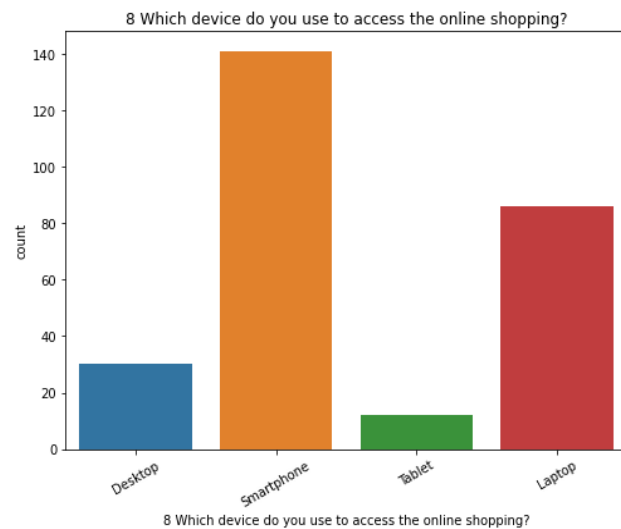
- **Majority, 141 customers smartphone while shopping online.**

```python
plt.figure(figsize=(20,20))
c=1
for i in df.columns[9:18]:
    plt.subplot(3,3,c)
    c=c+1
    ax = sns.countplot(y = df[i],palette='Accent')
    plt.xticks(rotation=30)


    total = len(df[i])
    for p in ax.patches:
        percentage = '{:.1f}%'.format(100 * p.get_width()/total)
        x = p.get_x() + p.get_width() + 0.02
        y = p.get_y() + p.get_height()/2
        ax.annotate(percentage, (x, y))
        plt.tight_layout()
```
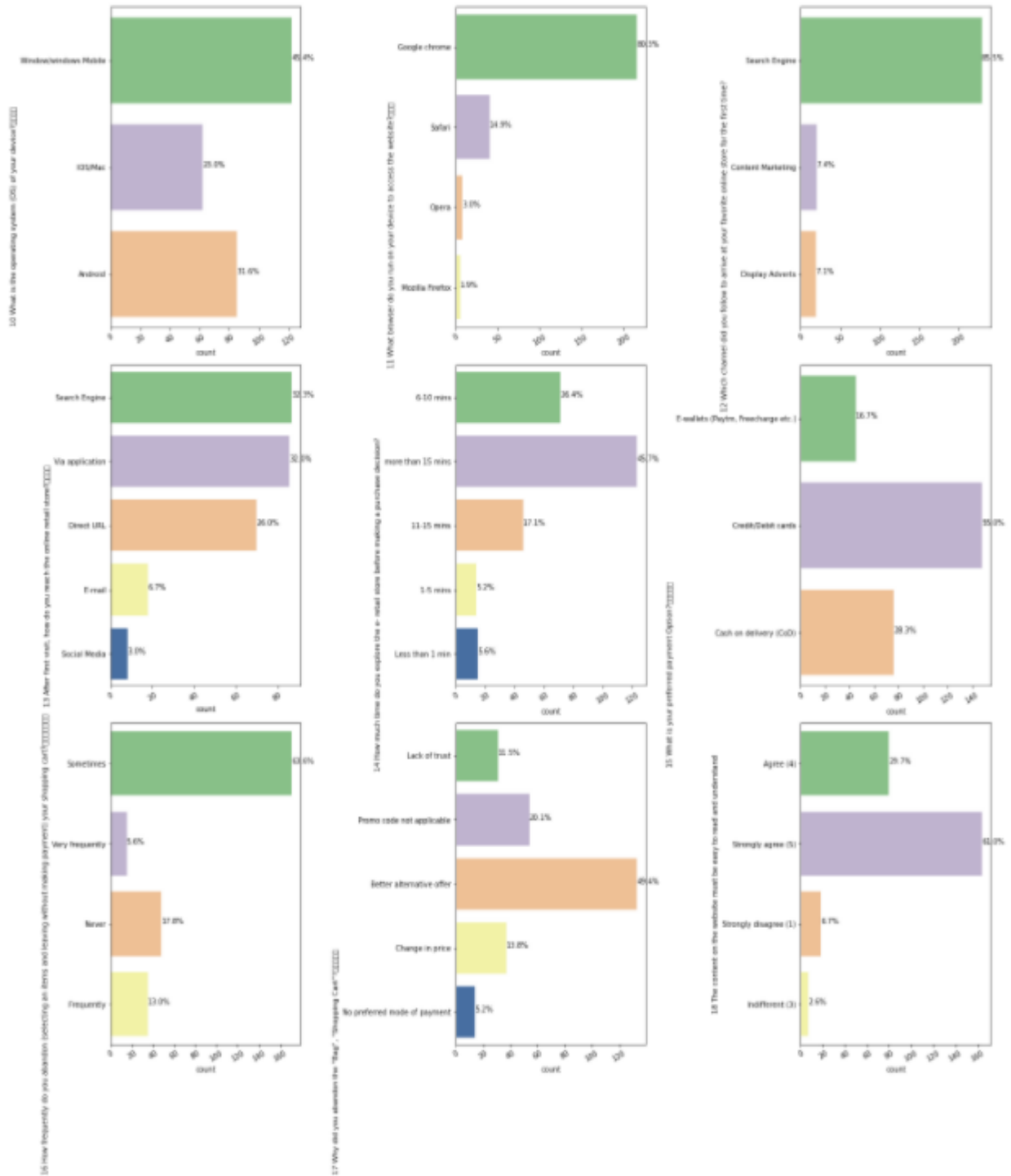
```
In [54]: plt.figure(figsize=(20,38))
         c=1
         for i in df.columns[65:72]:
             plt.subplot(5,2,c)
             c=c+1
             ax = sns.countplot(y = df[i],palette='Accent')
             plt.xticks(rotation=90)
             plt.xlabel(i)
             plt.ylabel('Category')

             total = len(df[i])
             for p in ax.patches:
                 percentage = '{:.1f}%'.format(100 * p.get_width()/total)
                 x = p.get_x() + p.get_width() + 0.02
                 y = p.get_y() + p.get_height()/2
                 ax.annotate(percentage, (x, y))

             plt.tight_layout()
```
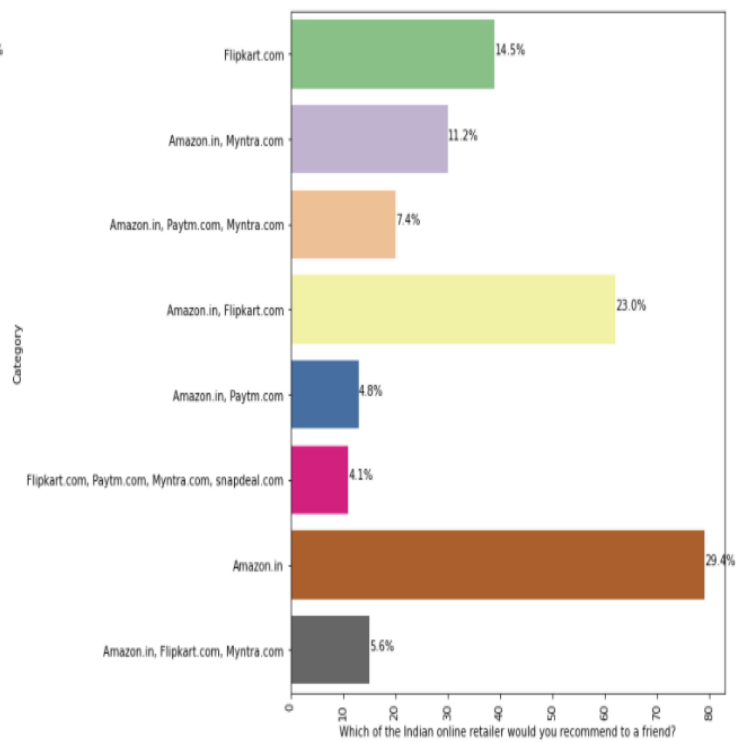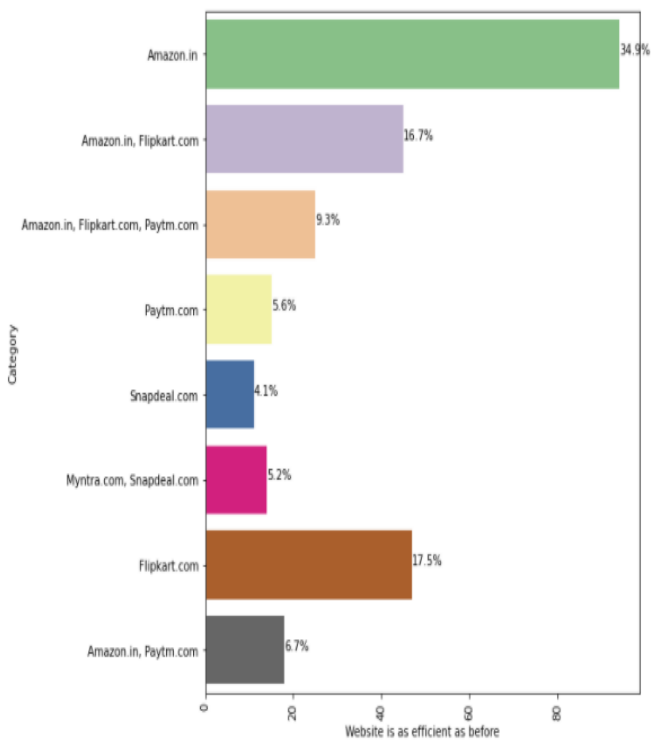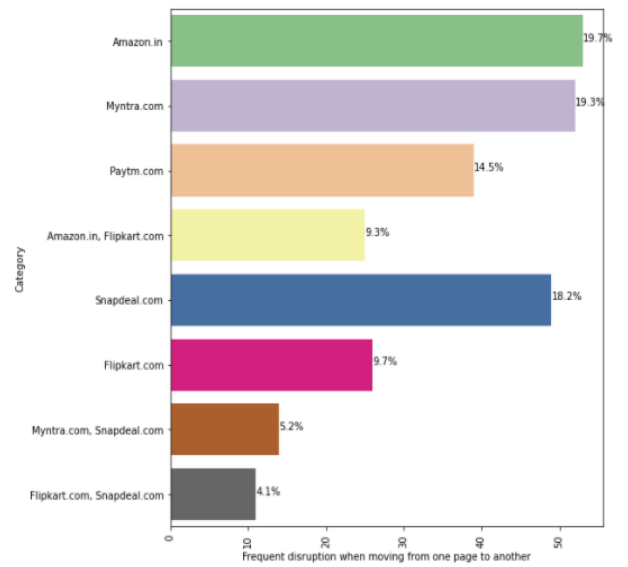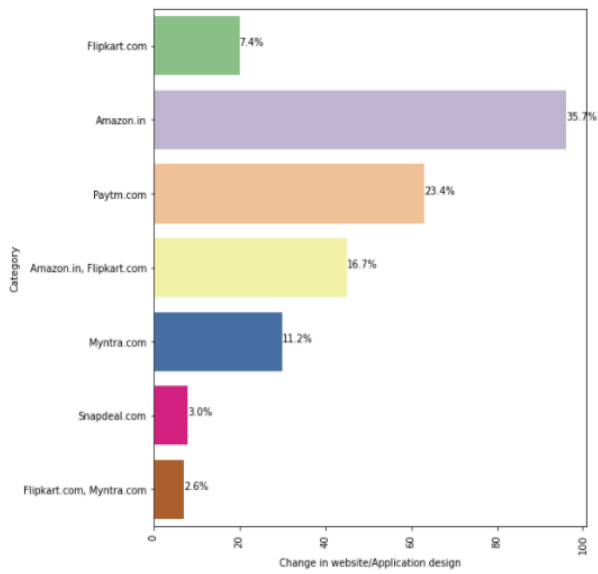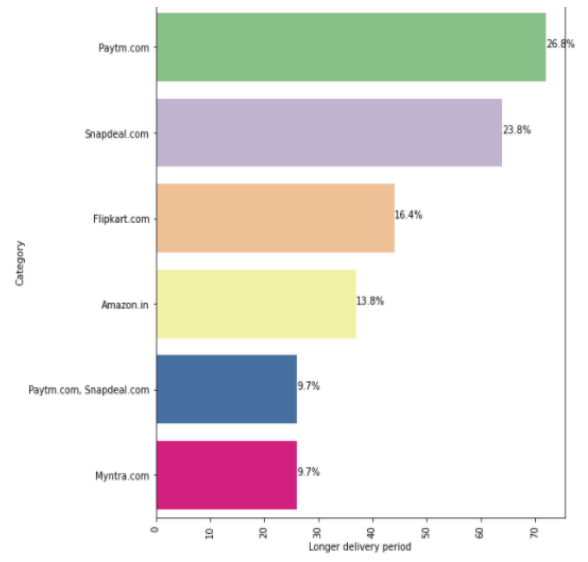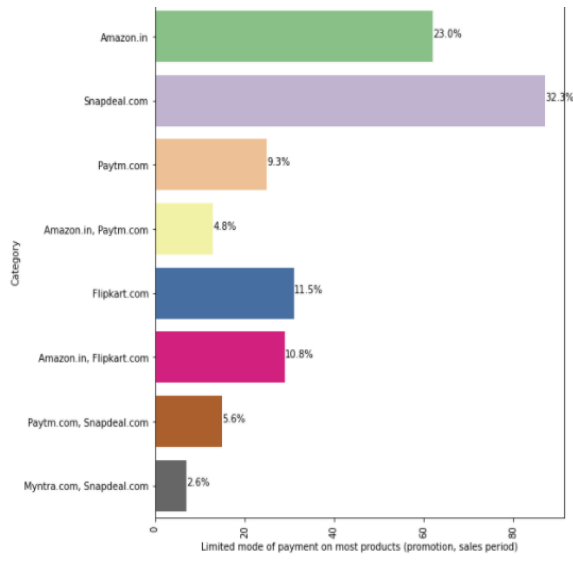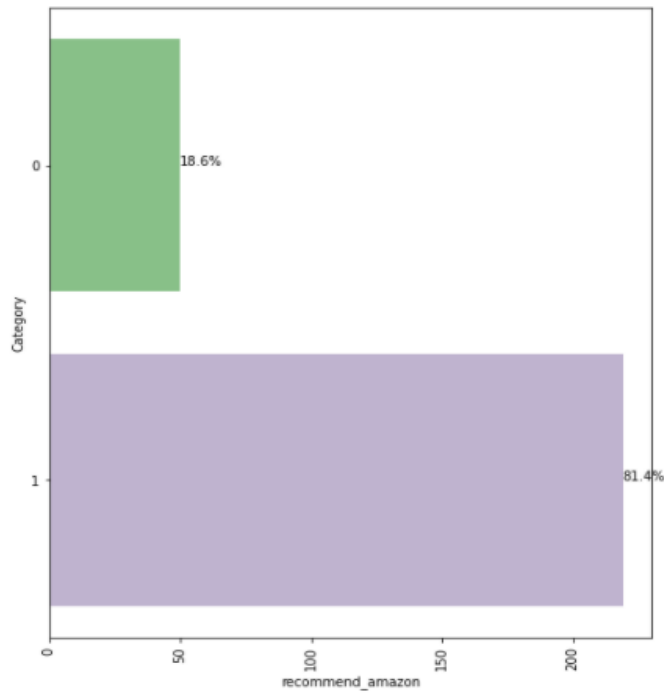
Chart 1 (top left): Limited mode of payment on most products (promotion, sales period)

Category	Percentage
Amazon.in	23.0%
Snapdeal.com	32.3%
Paytm.com	9.3%
Amazon.in, Paytm.com	4.8%
Flipkart.com	11.5%
Amazon.in, Flipkart.com	10.8%
Paytm.com, Snapdeal.com	5.6%
Myntra.com, Snapdeal.com	2.6%

Chart 2 (top right): Longer delivery period

Category	Percentage
Paytm.com	26.8%
Snapdeal.com	23.8%
Flipkart.com	16.4%
Amazon.in	13.8%
Paytm.com, Snapdeal.com	9.7%
Myntra.com	9.7%

Chart 3 (middle left): Change in website/Application design

Category	Percentage
Flipkart.com	7.4%
Amazon.in	35.7%
Paytm.com	23.4%
Amazon.in, Flipkart.com	16.7%
Myntra.com	11.2%
Snapdeal.com	3.0%
Flipkart.com, Myntra.com	2.6%

Chart 4 (middle right): Frequent disruption when moving from one page to another

Category	Percentage
Amazon.in	19.7%
Myntra.com	19.3%
Paytm.com	14.5%
Amazon.in, Flipkart.com	9.3%
Snapdeal.com	18.2%
Flipkart.com	9.7%
Myntra.com, Snapdeal.com	5.2%
Flipkart.com, Snapdeal.com	4.1%

Chart 5 (bottom left): Website is as efficient as before

Category	Percentage
Amazon.in	34.9%
Amazon.in, Flipkart.com	16.7%
Amazon.in, Flipkart.com, Paytm.com	9.3%
Paytm.com	5.6%
Snapdeal.com	4.1%
Myntra.com, Snapdeal.com	5.2%
Flipkart.com	17.5%
Amazon.in, Paytm.com	6.7%

Chart 6 (bottom right): Which of the Indian online retailer would you recommend to a friend?

Category	Percentage
Flipkart.com	14.5%
Amazon.in, Myntra.com	11.2%
Amazon.in, Paytm.com, Myntra.com	7.4%
Amazon.in, Flipkart.com	23.0%
Amazon.in, Paytm.com	4.8%
Flipkart.com, Paytm.com, Myntra.com, snapdeal.com	4.1%
Amazon.in	29.4%
Amazon.in, Flipkart.com, Myntra.com	5.6%

In [40]: #Creating one Hot encode for Recommend online retailers

df['recommend_amazon'] = df['Which of the Indian online retailer would you recommend to a friend?'].apply(lambda x :1 if (x.lower

df['recommend_flipkart'] = df['Which of the Indian online retailer would you recommend to a friend?'].apply(lambda x :1 if (x.lo

df['recommend_myntra'] = df['Which of the Indian online retailer would you recommend to a friend?'].apply(lambda x :1 if (x.lower

df['recommend_paytm'] = df['Which of the Indian online retailer would you recommend to a friend?'].apply(lambda x :1 if (x.lower(

df['recommend_snapdeal'] = df['Which of the Indian online retailer would you recommend to a friend?'].apply(lambda x :1 if (x.lo

In [42]: df[0:10]

Out[42]:

	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	10 What is the operating system (OS) of your device? \t\t\t	11 What browser do you run on your device to access the website? \t\t	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? \t\t\t	14 How much time do you explore the e-retail store before making a purchase decision?	15 W pr pa O
0	Male	31-40 years	Delhi	110009	Above 4 years	31-40 times	Dial-up	Desktop	Others	Window/windows Mobile	Google chrome	Search Engine	Search Engine	6-10 mins	E- (Free
1	Female	21-30 years	Delhi	110030	Above 4 years	41 times and above	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	Google chrome	Search Engine	Via application	more than 15 mins	Cred
2	Female	21-30 years	Greater Noida	201308	3-4 years	41 times and above	Mobile Internet	Smartphone	5.5 inches	Android	Google chrome	Search Engine	Via application	11-15 mins	E- (Free
3	Male	21-30 years	Karnal	132001	3-4 years	Less than 10 times	Mobile Internet	Smartphone	5.5 inches	IOS/Mac	Safari	Search Engine	Search Engine	6-10 mins	Cred
4	Female	21-30 years	Bangalore	530068	2-3 years	11-20 times	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	Safari	Content Marketing	Via application	more than 15 mins	Cred
5	Female	31-40	Noida	201308	Above 4 years	41 times and	Wi-Fi	Smartphone	5.5 inches	Android	Google chrome	Search Engine	Via application	more than 15 mins	E- (Free

```
In [49]: plt.figure(figsize=(10,20))
         c=1
         for i in df.columns[71:]:
             plt.subplot(5,2,c)
             c=c+1
             ax = sns.countplot(y = df[i],palette='Set2')
             plt.xticks(rotation=90)
             plt.xlabel(i)
             plt.ylabel('Category')

             total = len(df[i])
             for p in ax.patches:
                 percentage = '{:.1f}%'.format(100 * p.get_width()/total)
                 x = p.get_x() + p.get_width() + 0.02
                 y = p.get_y() + p.get_height()/2
                 ax.annotate(percentage, (x, y))

         plt.tight_layout()
```
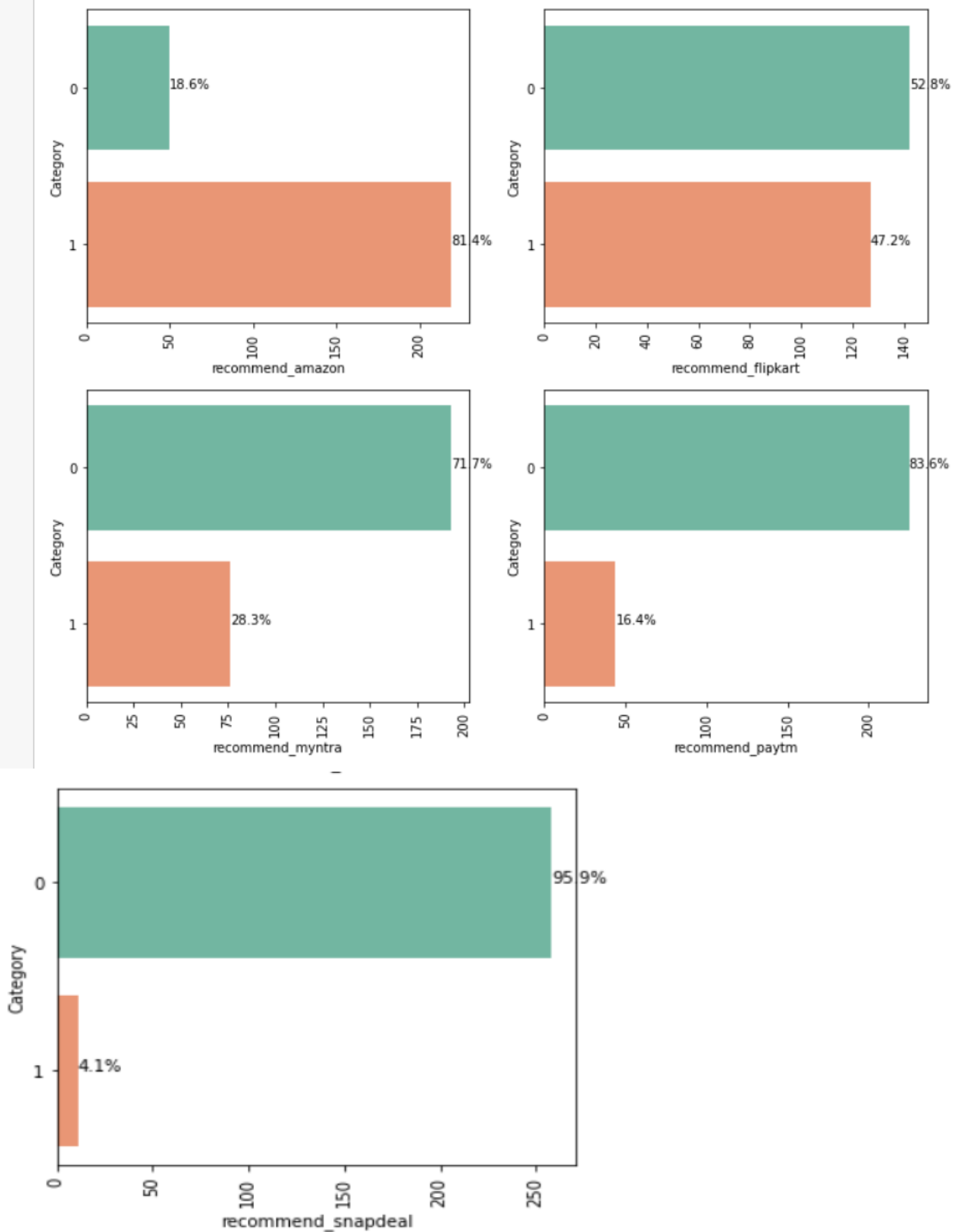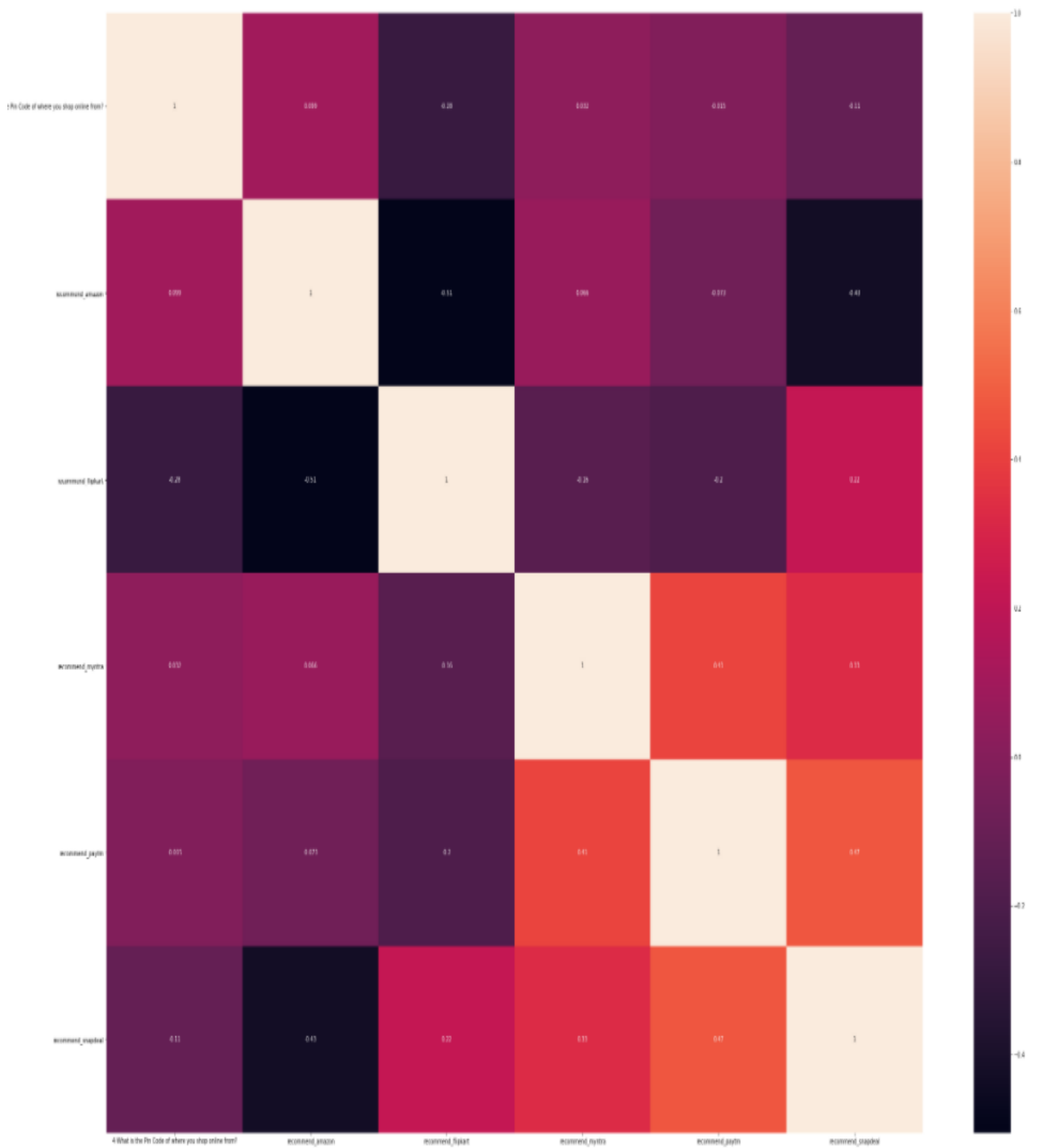
```
In [46]: plt.figure(figsize=(40,30))
         sns.heatmap(df.corr(),annot=True)
```

Out[46]: <AxesSubplot:>

CONCLUSION

- **Amazon is recommended by 81.4% of the customers.**

- **Snapdeal is recommended by 4.1% of the customers.**

- **Flipkart is recommended by 47.2 of the customers.**

- **Myntra is recommended by 28% of the customers.**

- **Paytm is recommended by 16% of the customers**

Thank You!!