



CUSTOMER RETENTION DATASET

Submitted by:

Trupti Badole

ACKNOWLEDGMENT

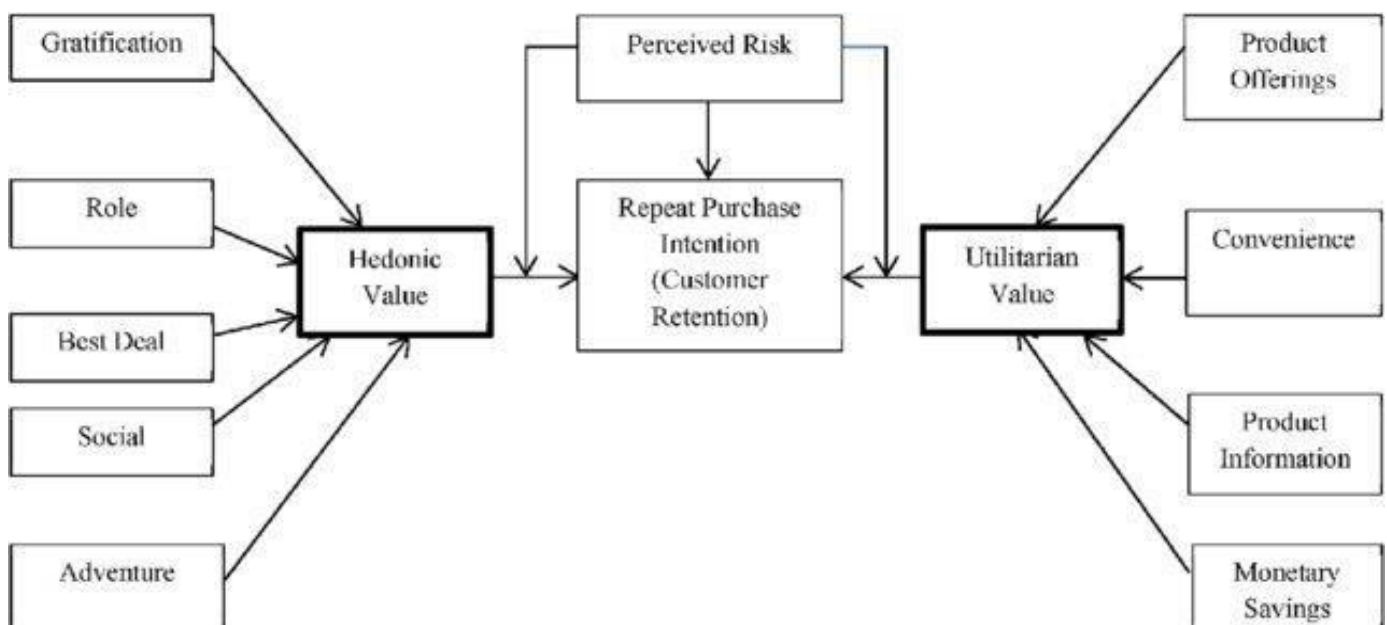
I would like to thank Flip Robo Technologies for providing me with the opportunity to work on this project from which I have learned a lot. I am also grateful to Swati Mahaseth for his constant guidance and support.

CONTENTS

1. What is Customer Retention?
2. Why is customer retention important?
3. Customer retention benefits.
4. Review of Literature

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.

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?	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	

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.

In [7]: OT

Out[7]:

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.
- Jupyter Notebook/Google chrome
- Libraries and packages used:

```
import pandas as pd
```

```
import numpy as np
```

```
import matplotlib.pyplot as plt
```

```
import seaborn as sn
```

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

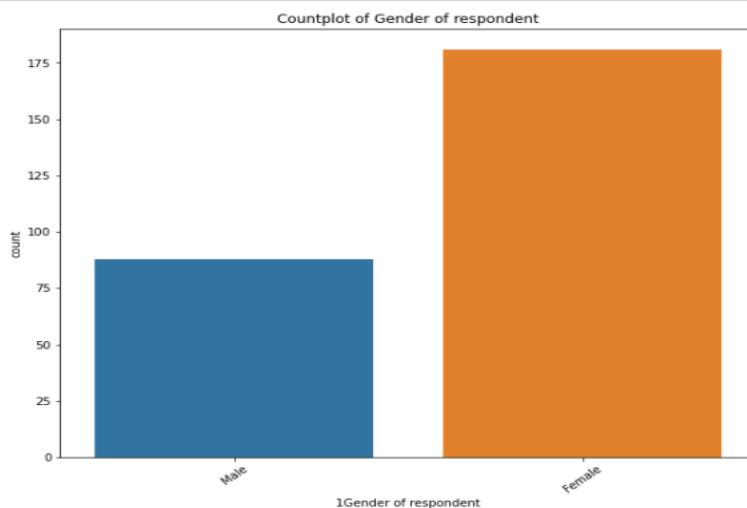
We will use Python through Jupyter notebook for data processing. Also we will use liabraries here is numpy, matplotlib, pandas and seaborn. The matpotplotlib and seaborn library has been used to make

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()
```

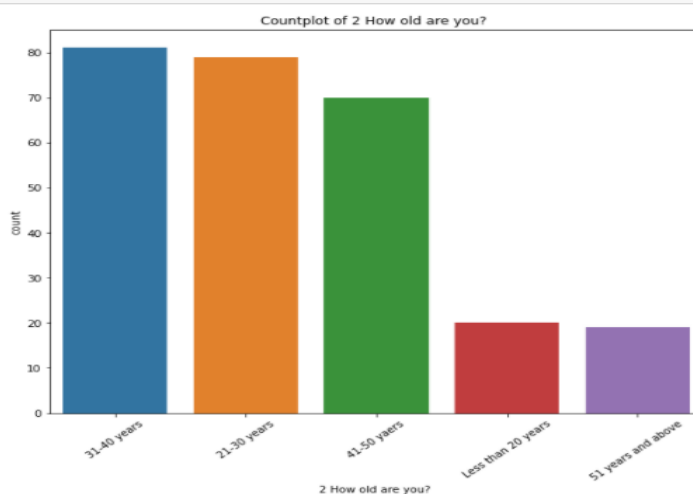


```
Out[20]: Female    181
Male         88
Name: 1Gender of respondent, dtype: int64
```

- 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()
```

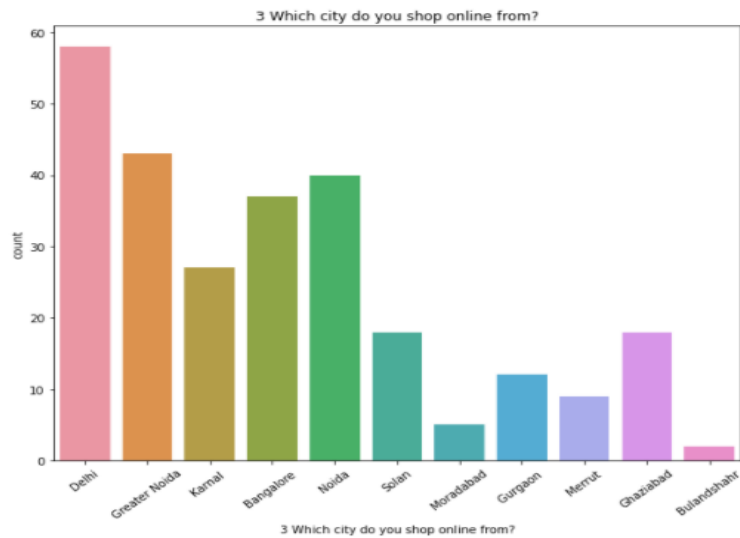


```
Out[21]: 31-40 years    81
21-30 years    79
41-50 yaers    70
Less than 20 years    20
51 years and above    19
Name: 2 How old are you? , dtype: int64
```

- 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()
```

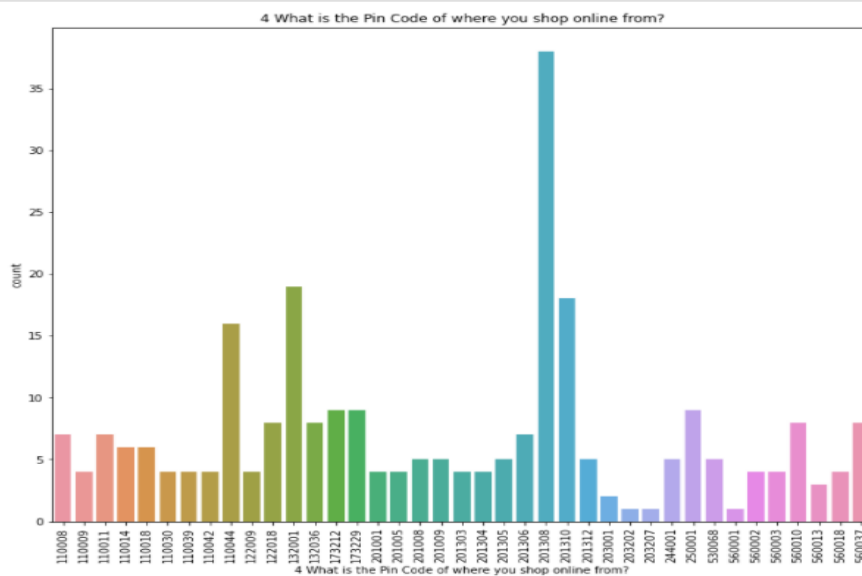


```
Out[22]: Delhi      58
Greater Noida    43
Noida            40
Bangalore       37
Karnal          27
Solan           18
Ghaziabad       18
Gurgaon         12
Merrut           9
Moradabad        5
Bulandshahr      2
Name: 3 Which city do you shop online from?, dtype: int64
```

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

```
In [23]: #checking the column 4 What is the Pin Code of where you shop online from?
plt.subplots(figsize=(12,10))
sns.countplot(x="4 What is the Pin Code of where you shop online from?", data=df)
plt.title("4 What is the Pin Code of where you shop online from?")
plt.xticks(rotation='vertical')
plt.xlabel('4 What is the Pin Code of where you shop online from?')
plt.ylabel("count")
plt.show()

df['4 What is the Pin Code of where you shop online from?'].value_counts()
```



```
Out[23]: 201308      38
132001      19
201310      18
110044      16
173229       9
173212       9
250001       9
122018       8
560037       8
132036       8
560010       8
```

- Majority, 38 of the customers placed an order from the pincode 201308.

```
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()
```

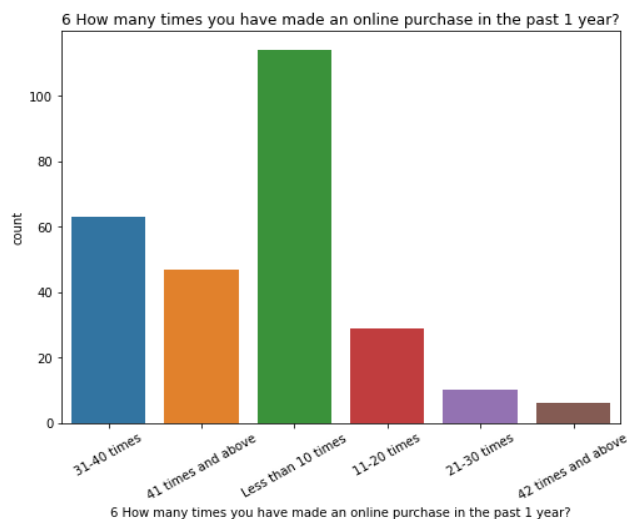


```
Out[24]: Above 4 years      98
         2-3 years         65
         3-4 years         47
         Less than 1 year   43
         1-2 years         16
         Name: 5 Since How Long You are Shopping Online ?, dtype: int64
```

- 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()
```

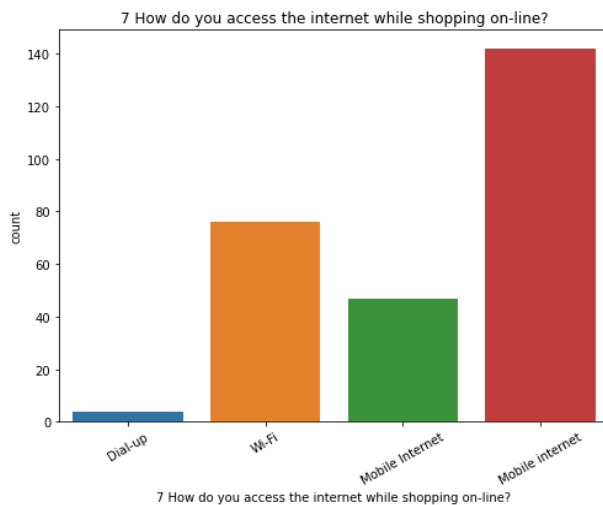


```
Out[25]: Less than 10 times    114
         31-40 times           63
         41 times and above    47
         11-20 times           29
         21-30 times           10
         42 times and above      6
         Name: 6 How many times you have made an online purchase in the past 1 year?, dtype: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



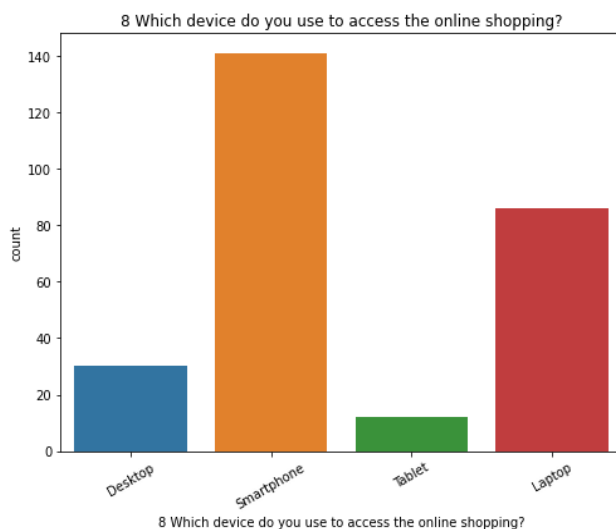
```
Out[26]: Mobile internet    142
Wi-Fi                      76
Mobile Internet           47
Dial-up                   4
Name: 7 How do you access the internet while shopping on-line?, dtype: int64
```

- 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()
```

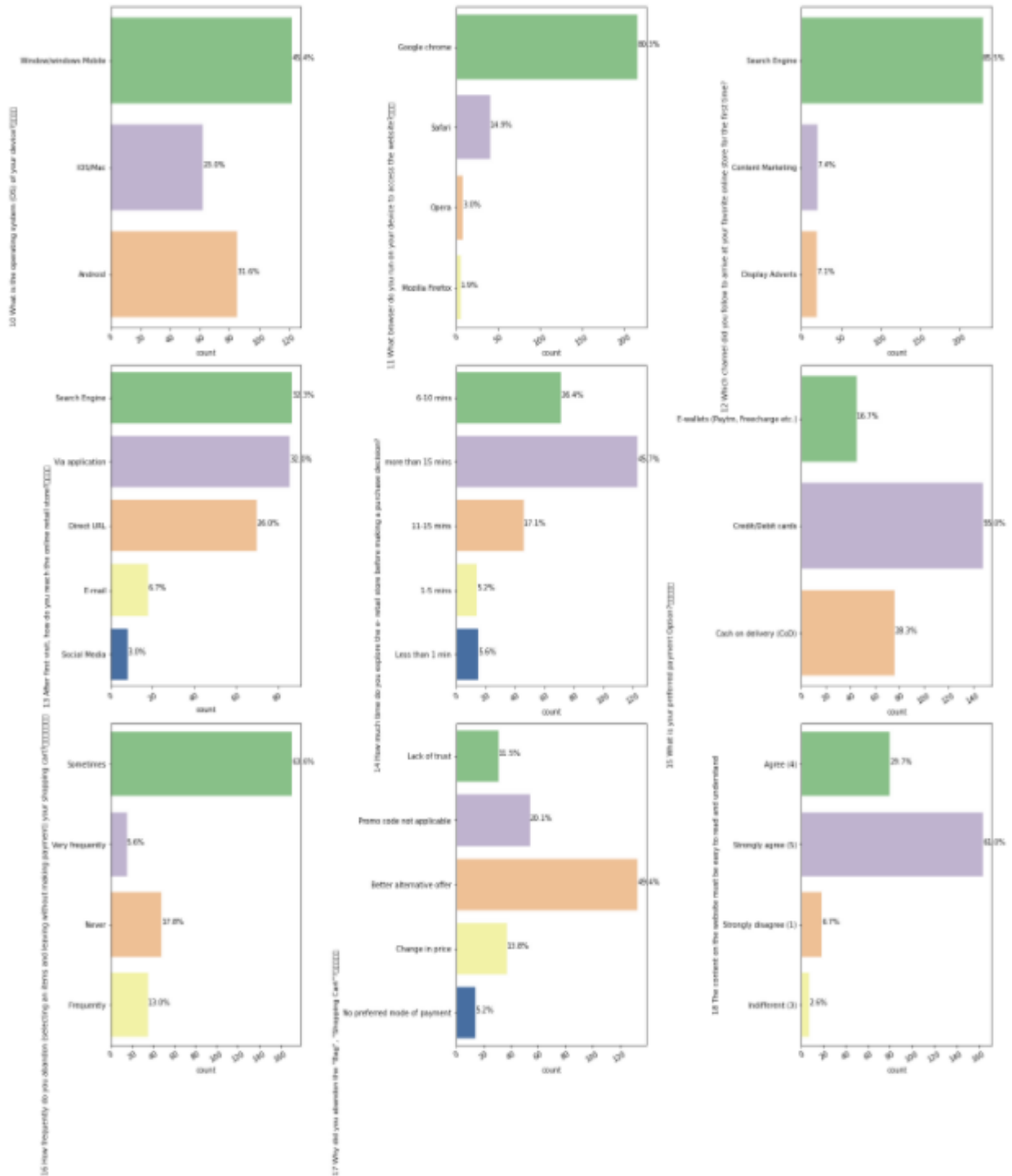
```
Out[27]: Smartphone    141
Laptop                86
Desktop              30
Tablet               12
Name: 8 Which device do you use to access the online shopping?, dtype: int64
```



- Majority, 141 customers smartphone while shopping online.

```
In [53]: 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)

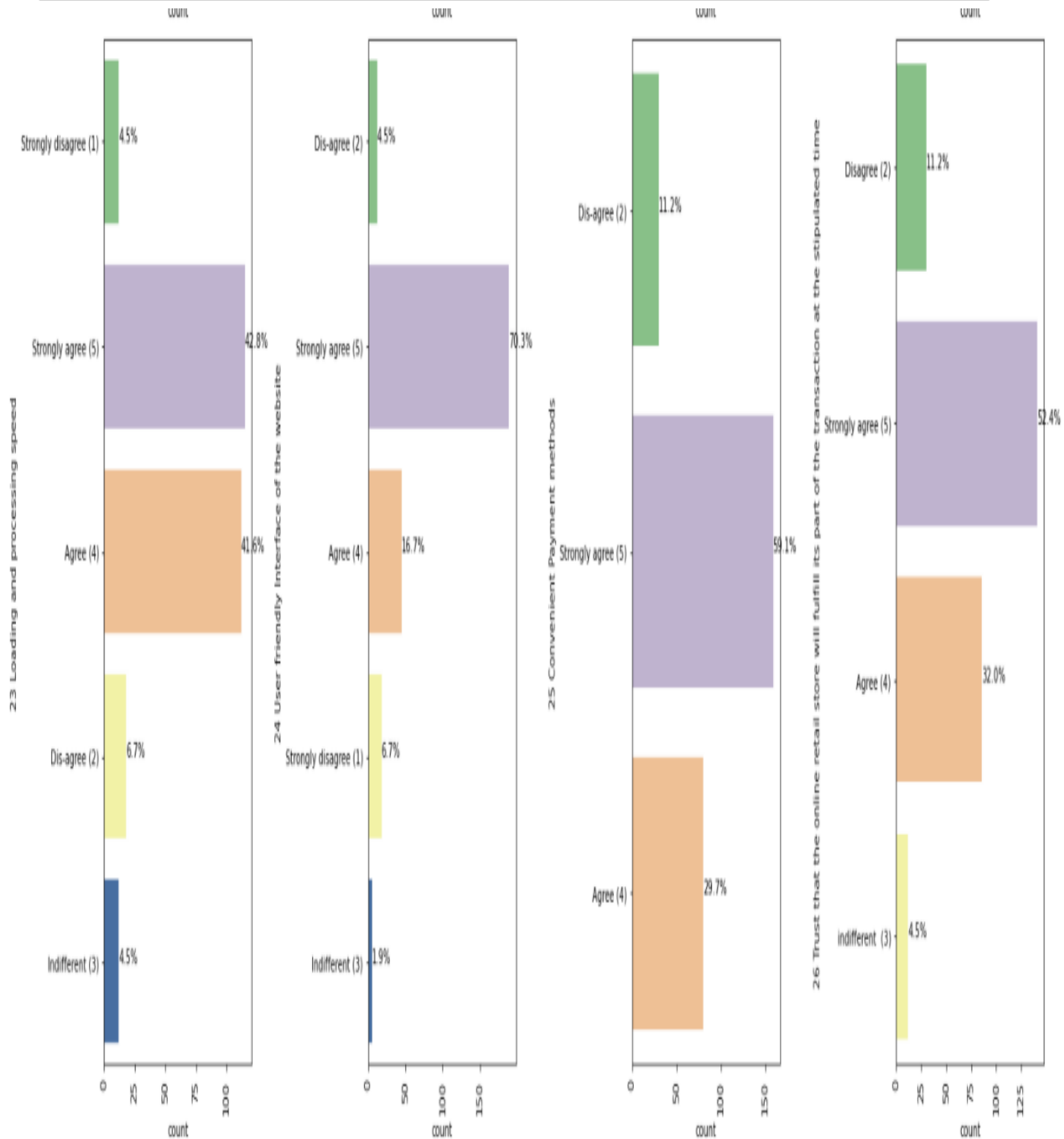
    total = len(df[i])
    for p in ax.patches:
        percentage = '{:.1f}%'.format(100 * p.get_width()/total)
        x = p.get_x() + p.get_width() + 0.02
        y = p.get_y() + p.get_height()/2
        ax.annotate(percentage, (x, y))
    plt.tight_layout()
```



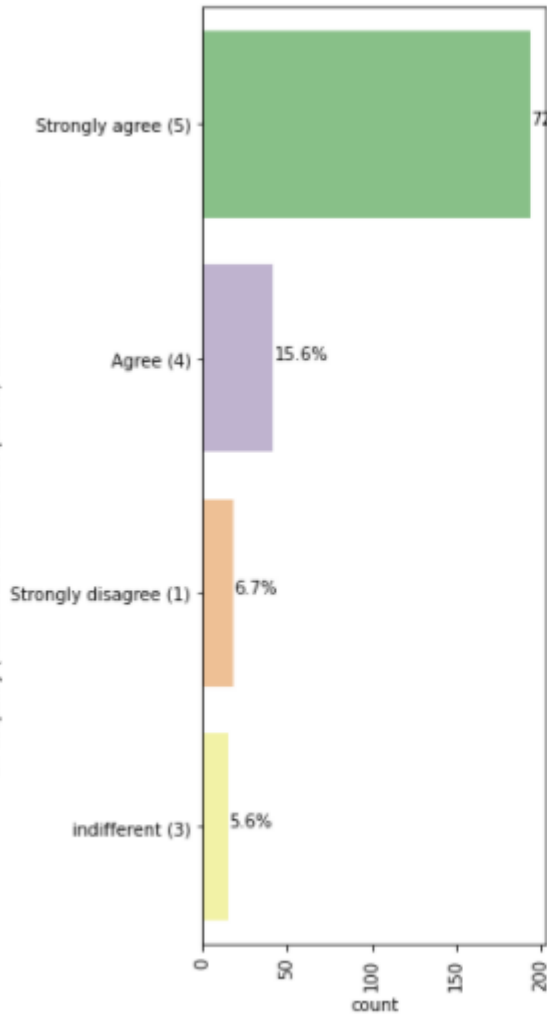
```
In [29]: plt.figure(figsize=(20,25))
c=1
for i in df.columns[18:28]:
    plt.subplot(3,4,c)
    c=c+1
    ax = sns.countplot(y = df[i],palette='Accent')
    plt.xticks(rotation=90)

    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))

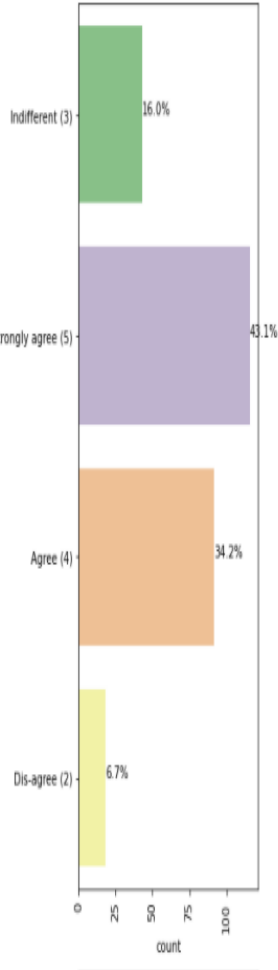
plt.tight_layout()
```



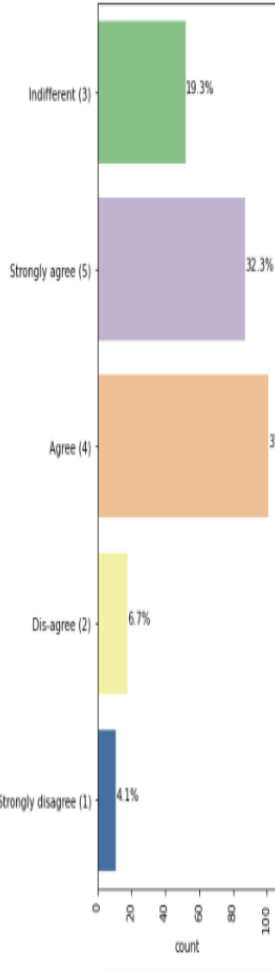
27 Empathy (readiness to assist with queries) towards the customers



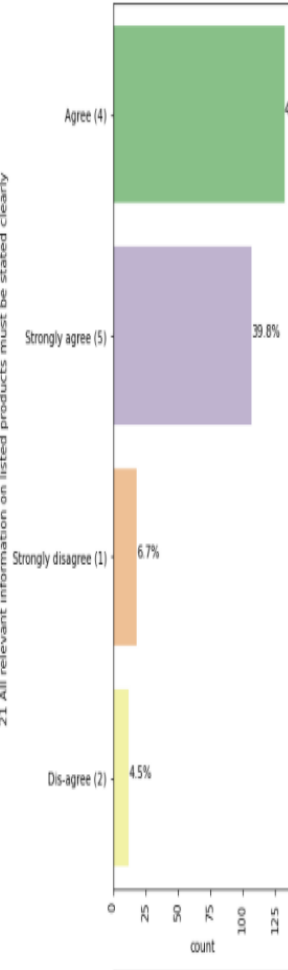
19 Information on similar product to the one highlighted is important for product comparison



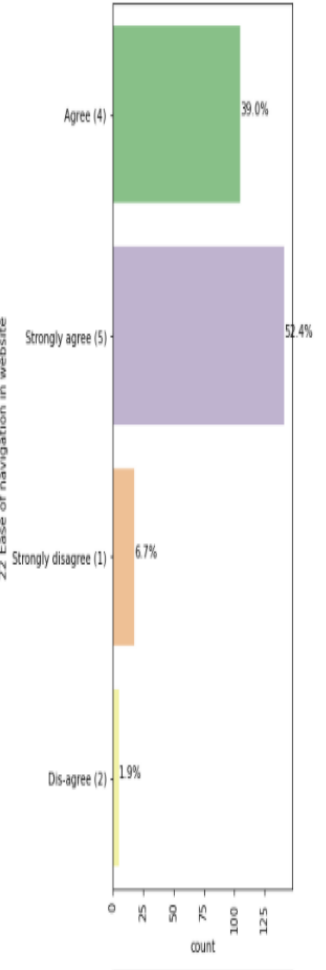
20 Complete information on listed seller and product being offered is important for purchase decision.



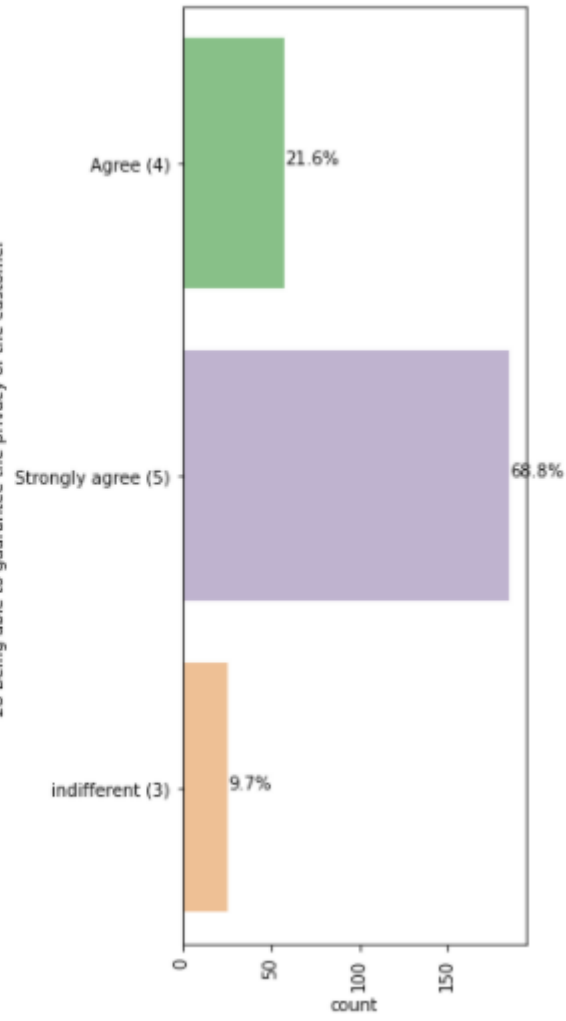
21 All relevant information on listed products must be stated clearly



22 Ease of navigation in website



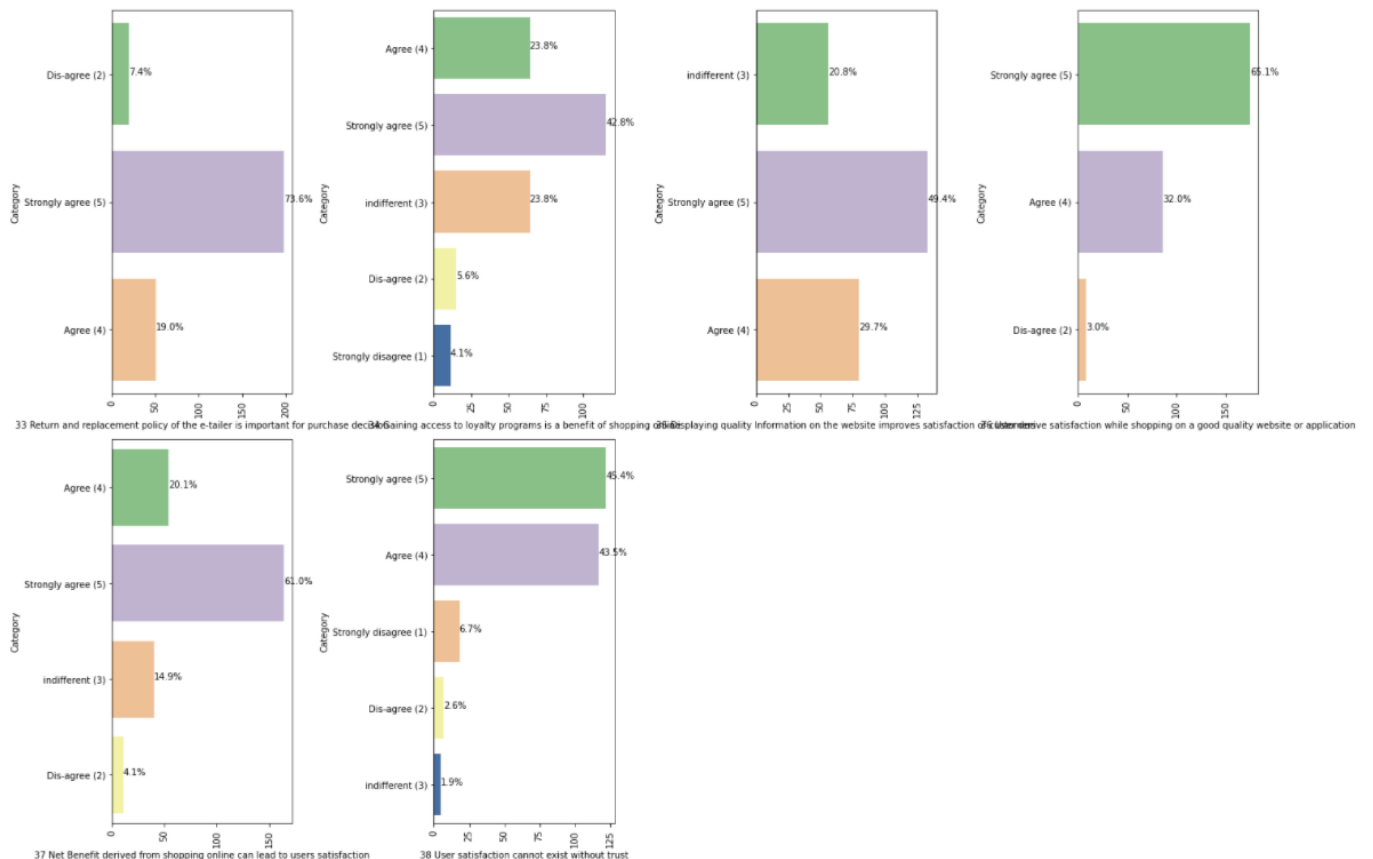
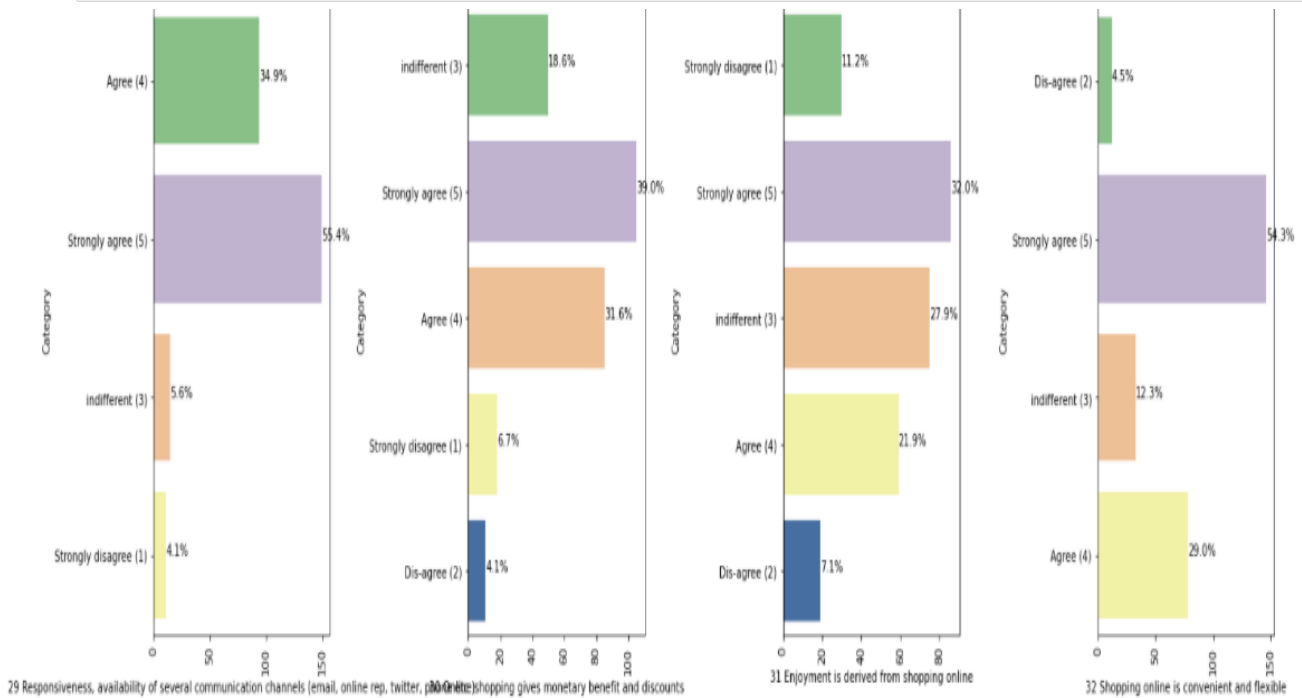
28 Being able to guarantee the privacy of the customer



```
In [30]: plt.figure(figsize=(20,20))
c=1
for i in df.columns[28:38]:
    plt.subplot(3,4,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()
```



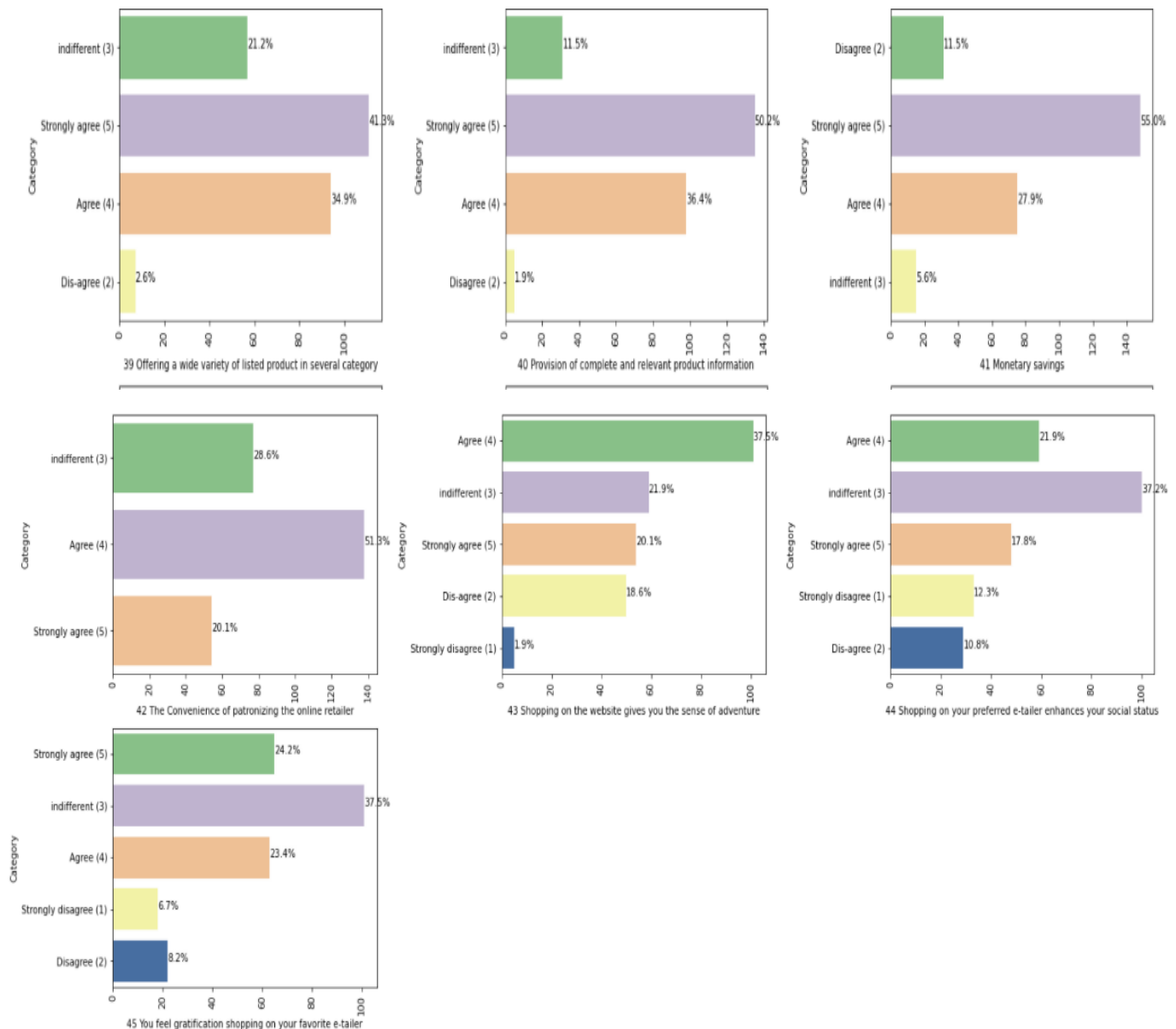
```

In [31]: plt.figure(figsize=(20,20))
c=1
for i in df.columns[38:45]:
    plt.subplot(5,3,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()

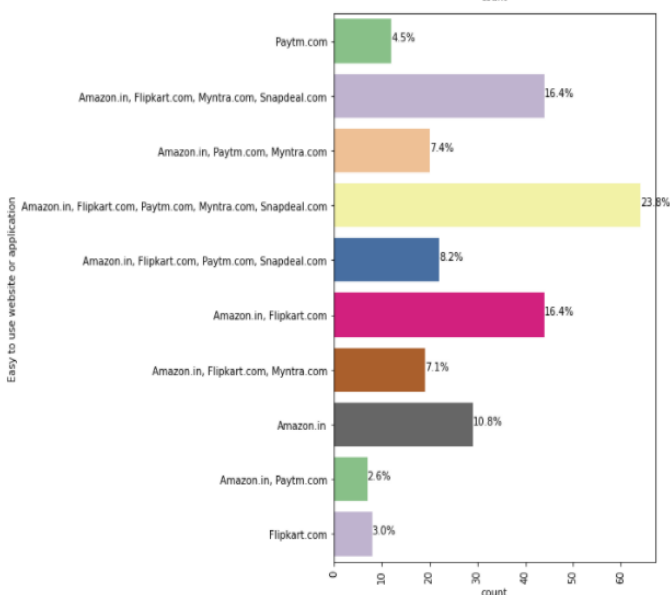
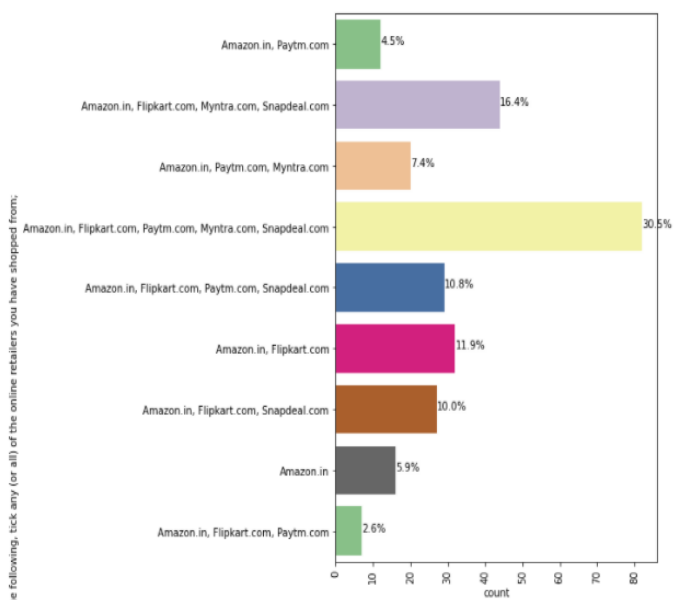
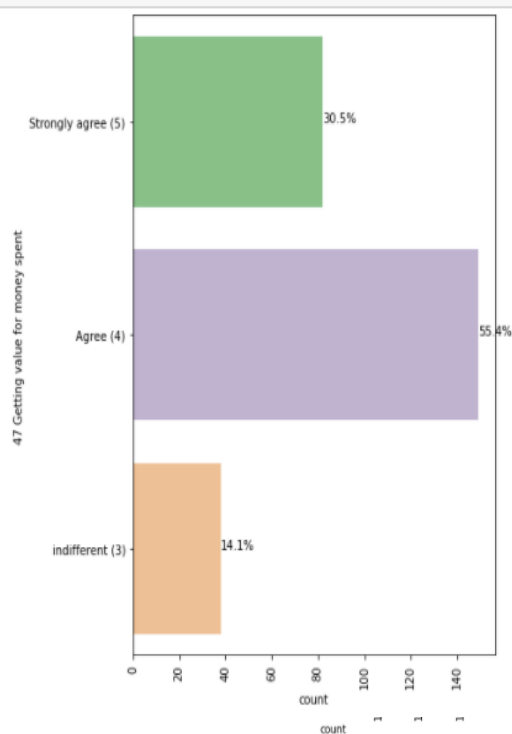
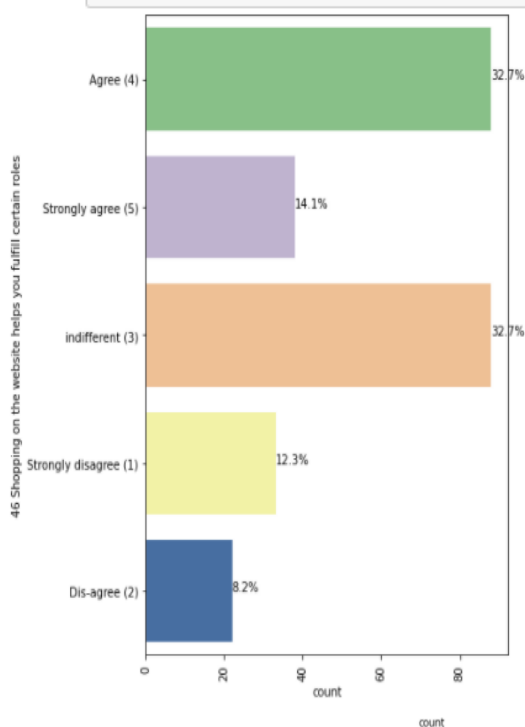
```



```
In [32]: plt.figure(figsize=(20,38))
c=1
for i in df.columns[45:52]:
    plt.subplot(5,2,c)
    c=c+1
    ax = sns.countplot(y = df[i],palette='Accent')
    plt.xticks(rotation=90)

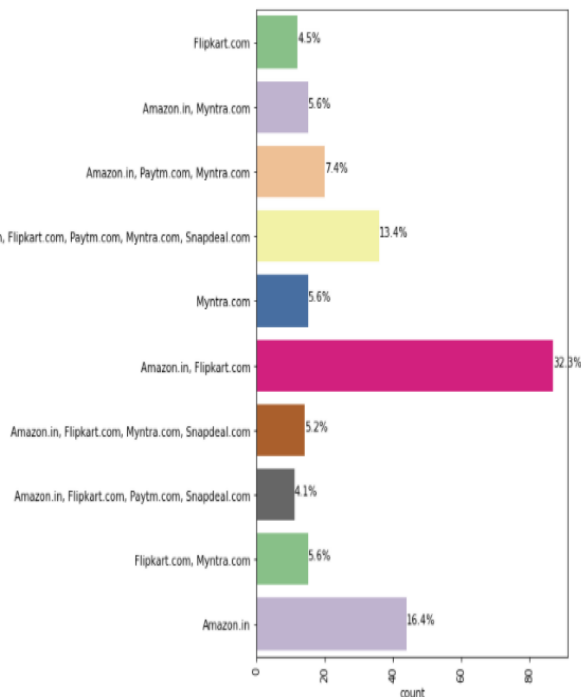
    total = len(df[i])
    for p in ax.patches:
        percentage = '{:.1f}%'.format(100 * p.get_width()/total)
        x = p.get_x() + p.get_width() + 0.02
        y = p.get_y() + p.get_height()/2
        ax.annotate(percentage, (x, y))

    plt.tight_layout()
```

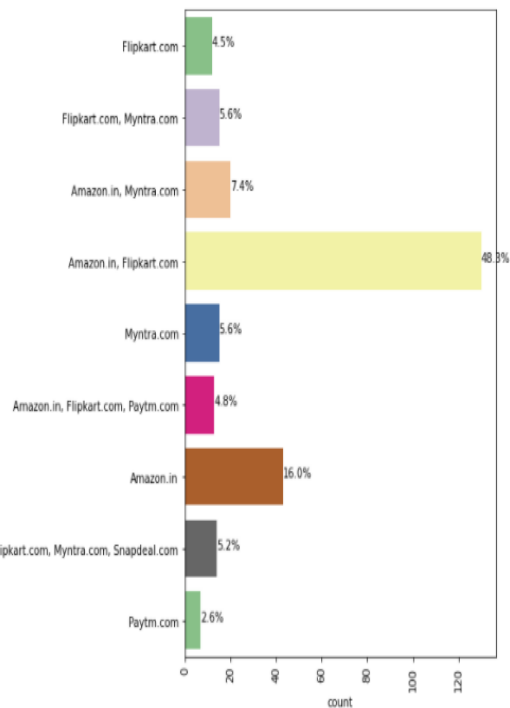


From the

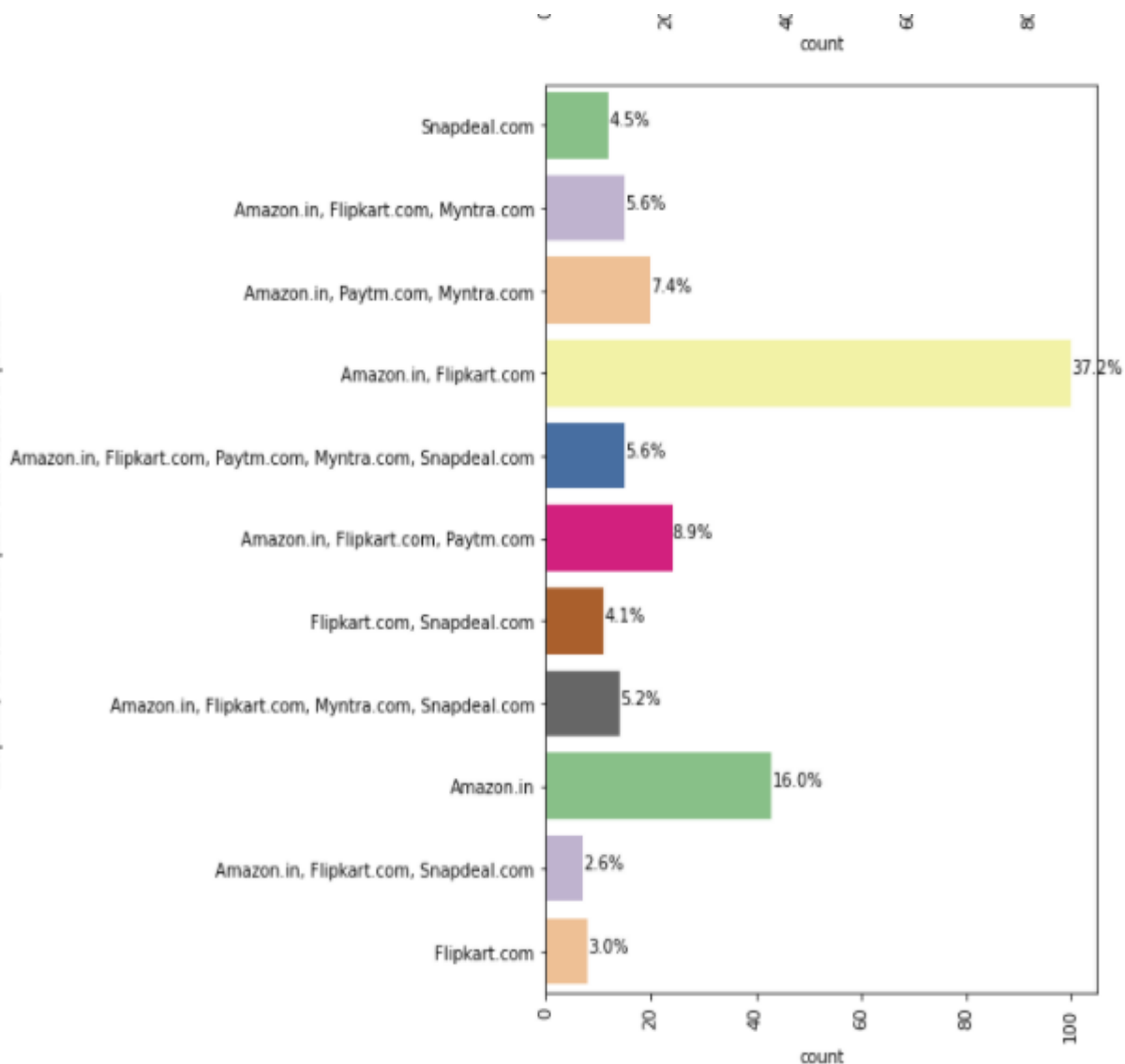
Visual appealing web-page layout



Wild variety of product on offer



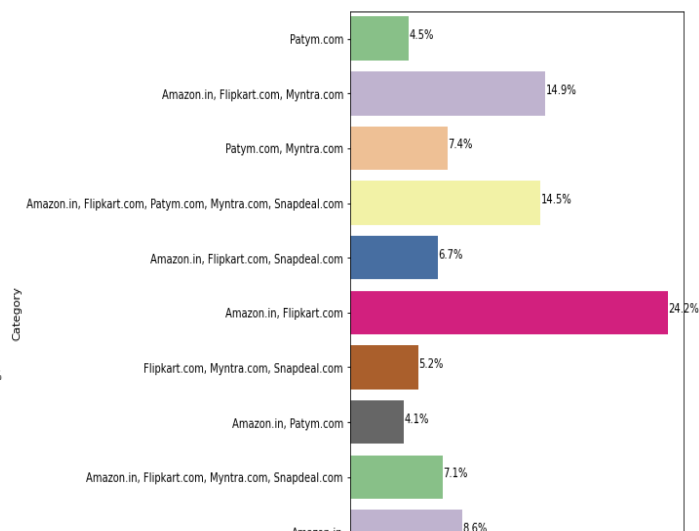
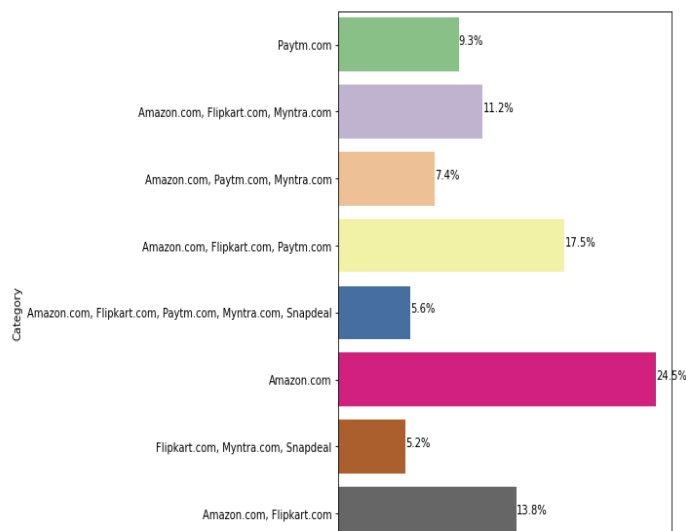
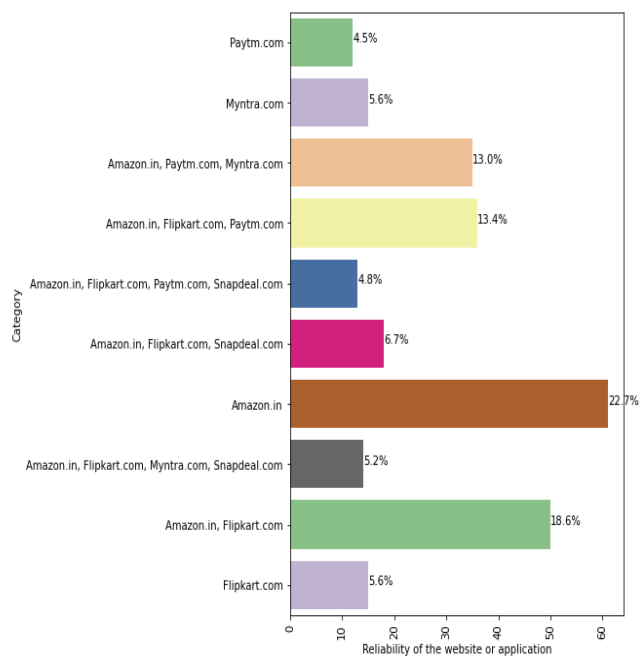
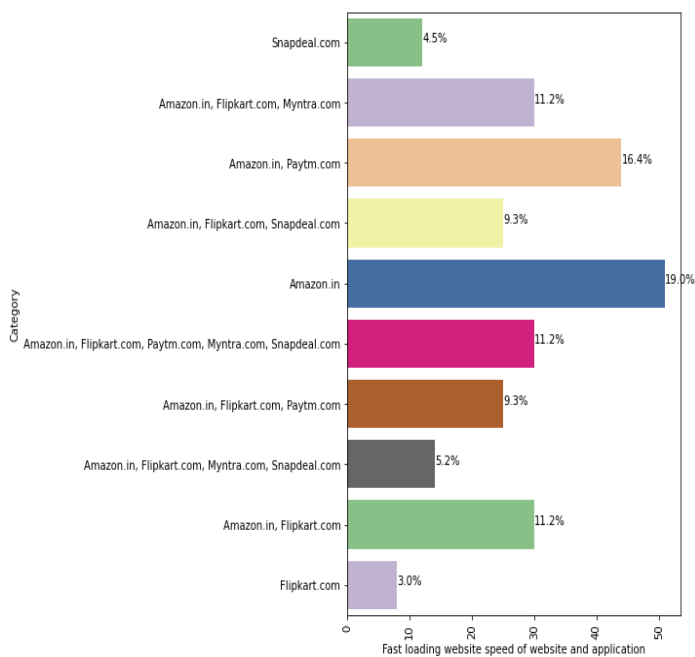
Complete, relevant description information of products

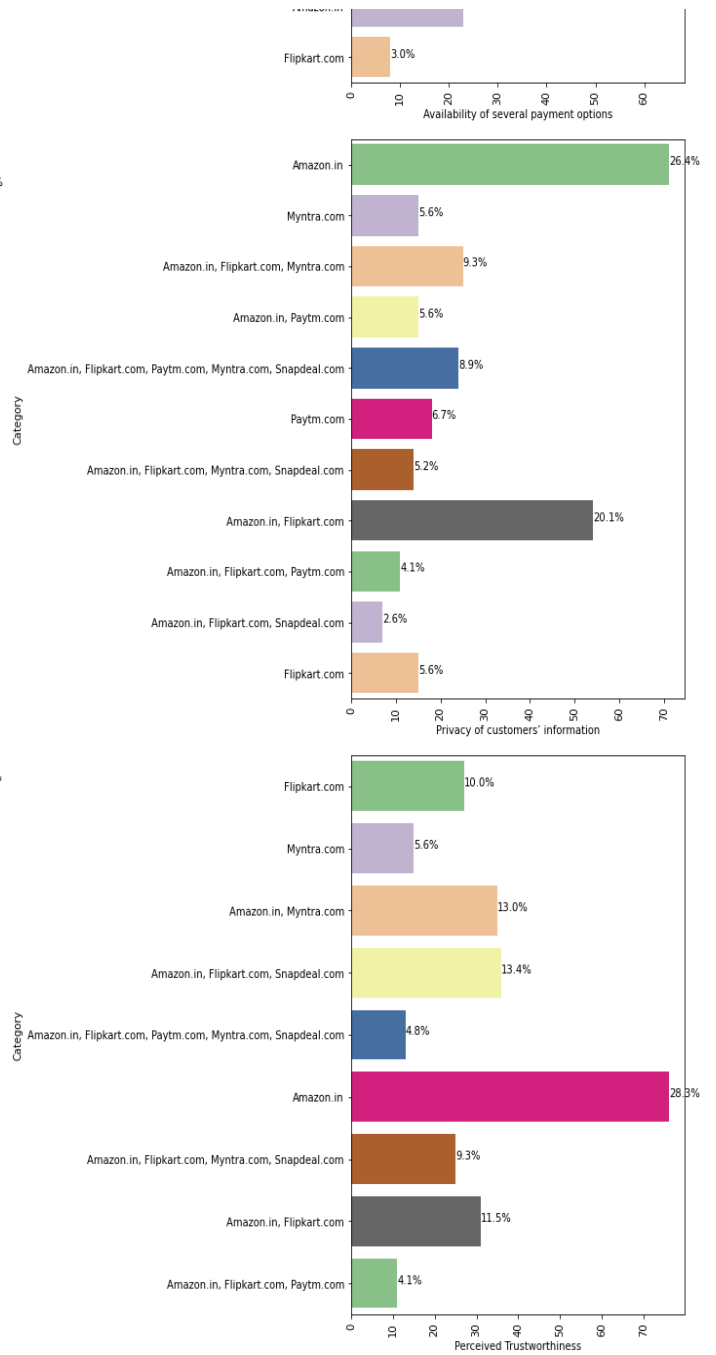
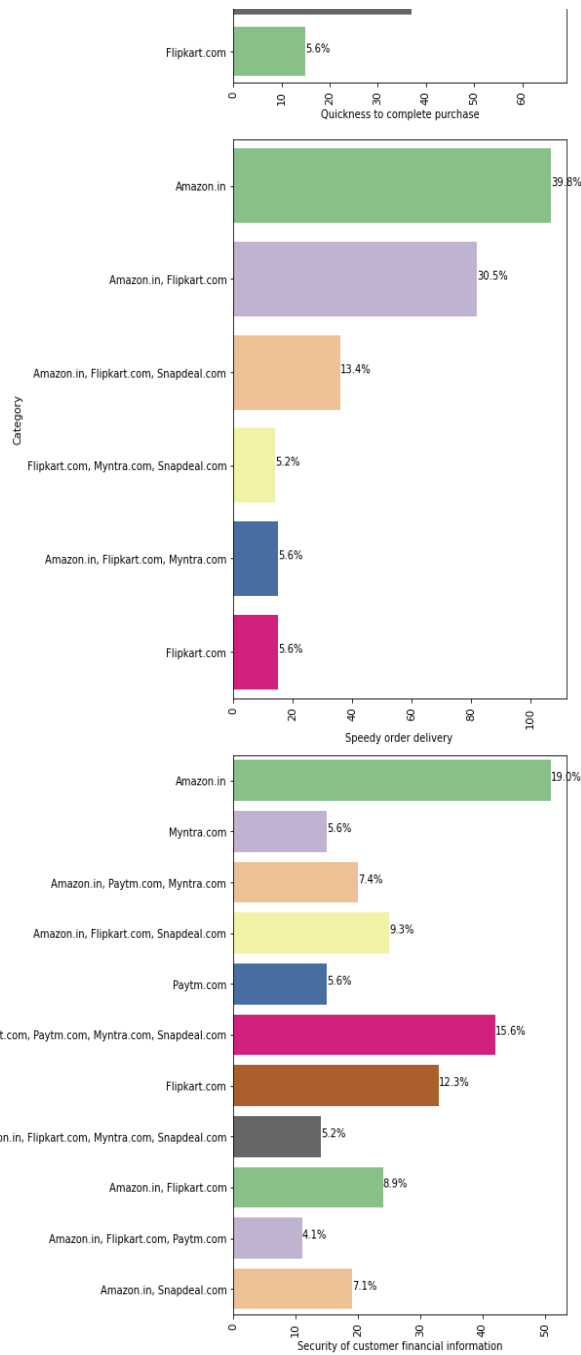



```
In [33]: plt.figure(figsize=(20,38))
c=1
for i in df.columns[52:60]:
    plt.subplot(5,2,c)
    c=c+1
    ax = sns.countplot(y = df[i],palette='Accent')
    plt.xticks(rotation=90)
    plt.xlabel(i)
    plt.ylabel('Category')

    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()
```

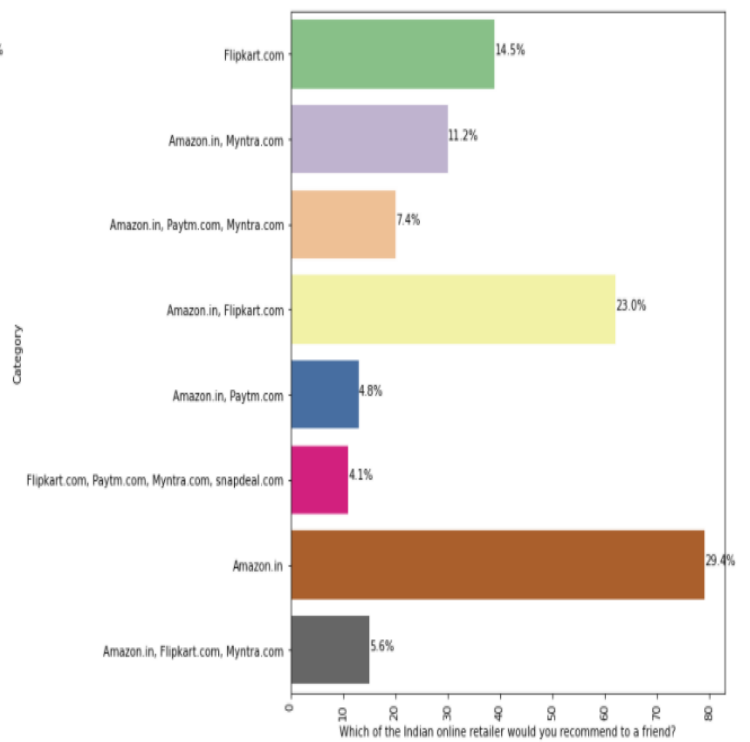
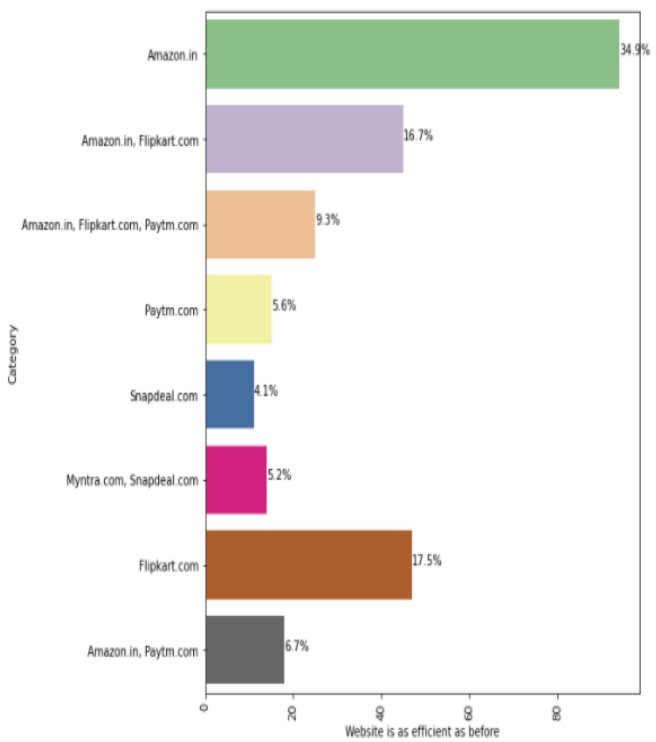
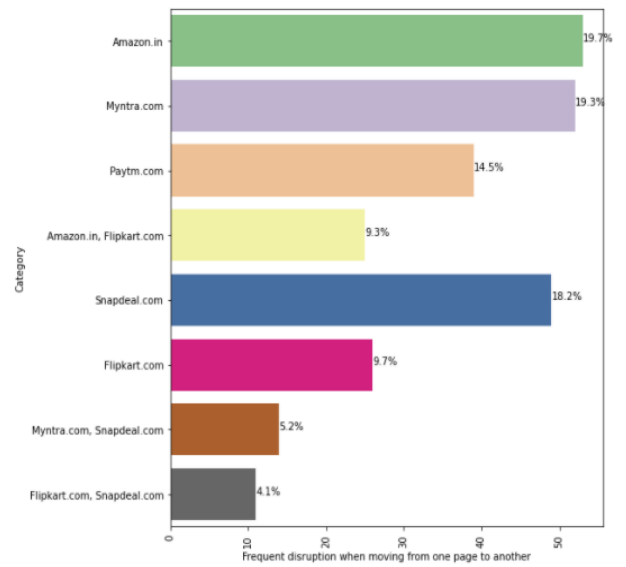
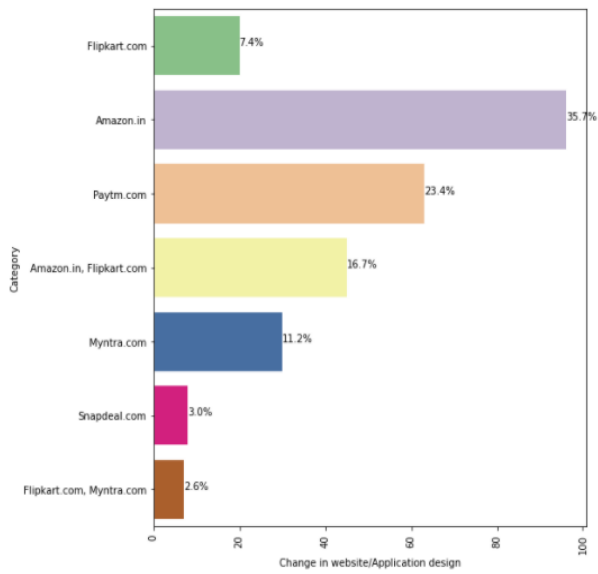
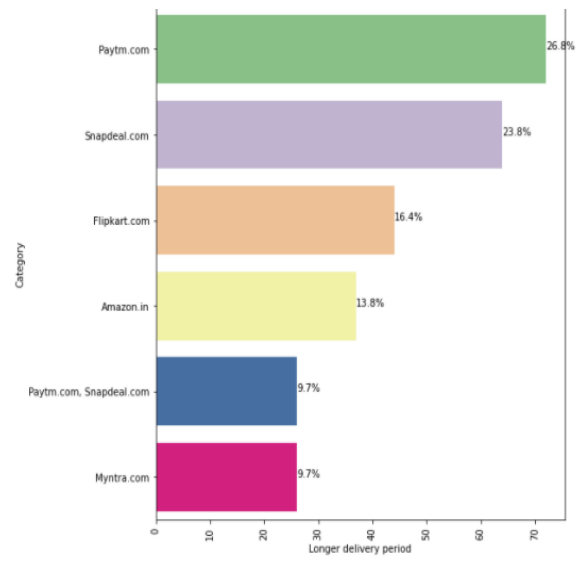
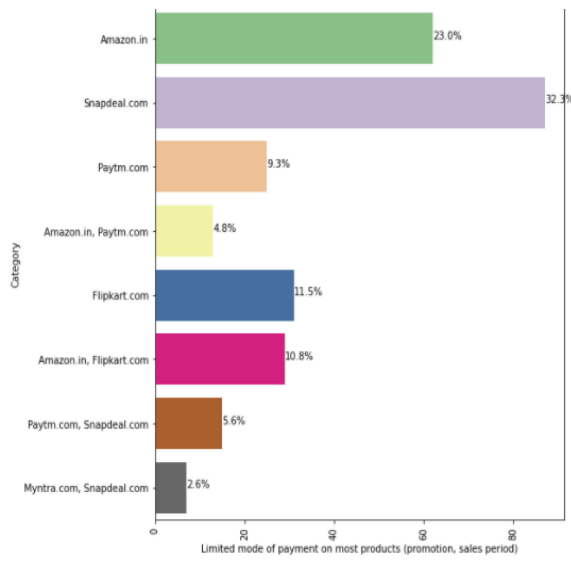


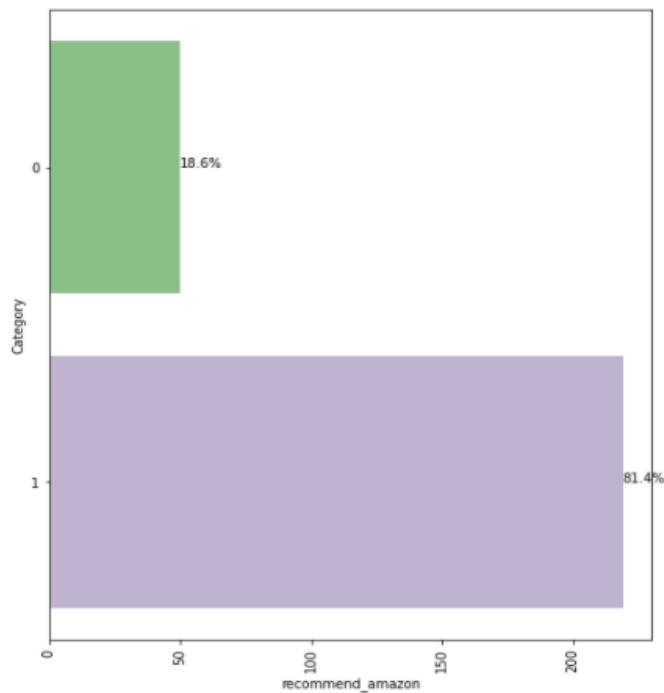


```
In [54]: plt.figure(figsize=(20,38))
c=1
for i in df.columns[65:72]:
    plt.subplot(5,2,c)
    c=c+1
    ax = sns.countplot(y = df[i],palette='Accent')
    plt.xticks(rotation=90)
    plt.xlabel(i)
    plt.ylabel('Category')

    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()
```





```
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() == 'amazon') else 0)
df['recommend_flipkart'] = df['Which of the Indian online retailer would you recommend to a friend?'].apply(lambda x :1 if (x.lower() == 'flipkart') else 0)
df['recommend_mynta'] = df['Which of the Indian online retailer would you recommend to a friend?'].apply(lambda x :1 if (x.lower() == 'mynta') else 0)
df['recommend_paytm'] = df['Which of the Indian online retailer would you recommend to a friend?'].apply(lambda x :1 if (x.lower() == 'paytm') else 0)
df['recommend_snapdeal'] = df['Which of the Indian online retailer would you recommend to a friend?'].apply(lambda x :1 if (x.lower() == 'snapdeal') else 0)
```

In [42]: df[0:10]

Out[42]:

	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 mode of payment?
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	Free (E-commerce)
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	Free (E-commerce)
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 years	Noida	201308	Above 4 years	41 times and above	Wi-Fi	Smartphone	5.5 inches	Android	Google chrome	Search Engine	Via application	more than 15 mins	Free (E-commerce)

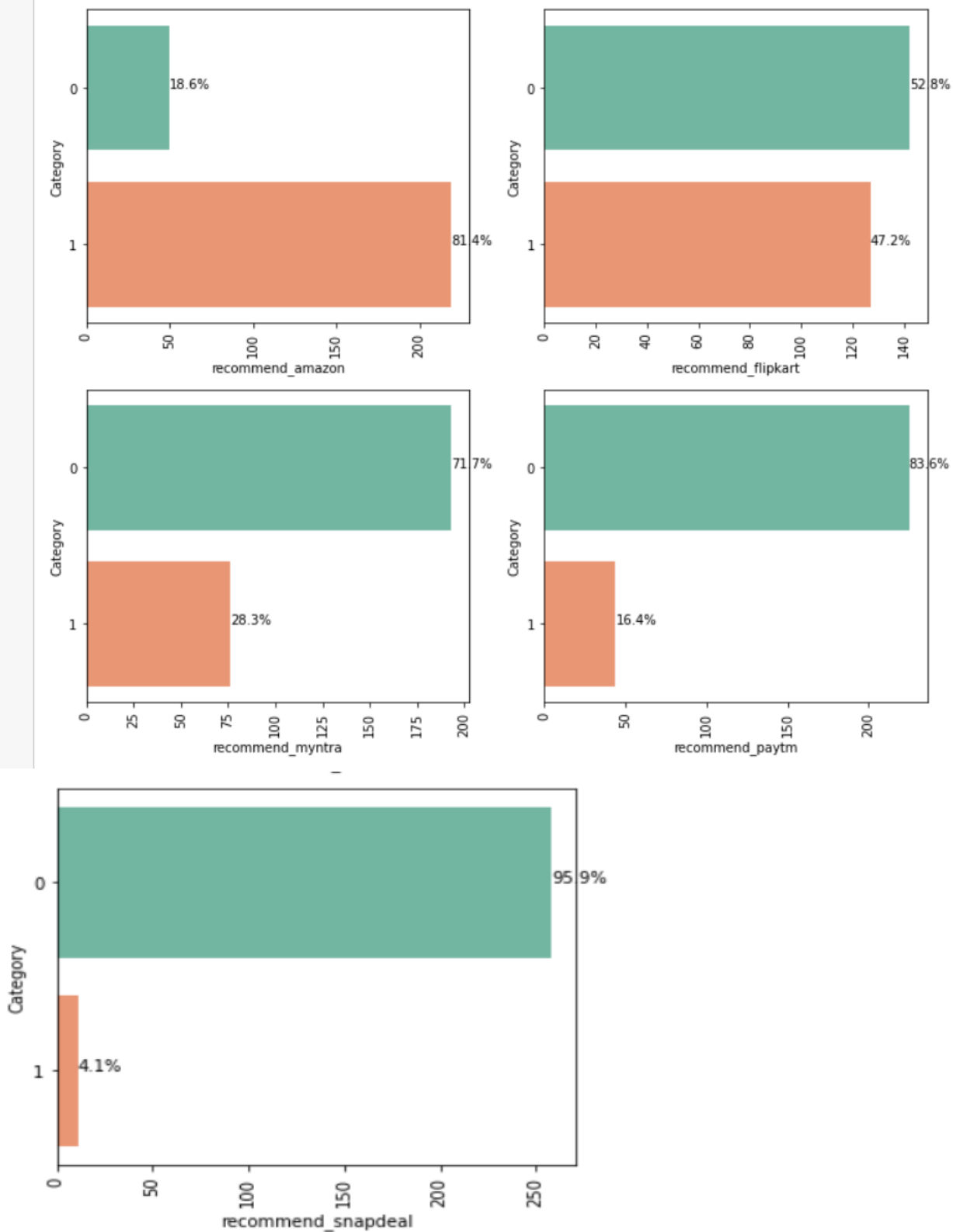
```

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))

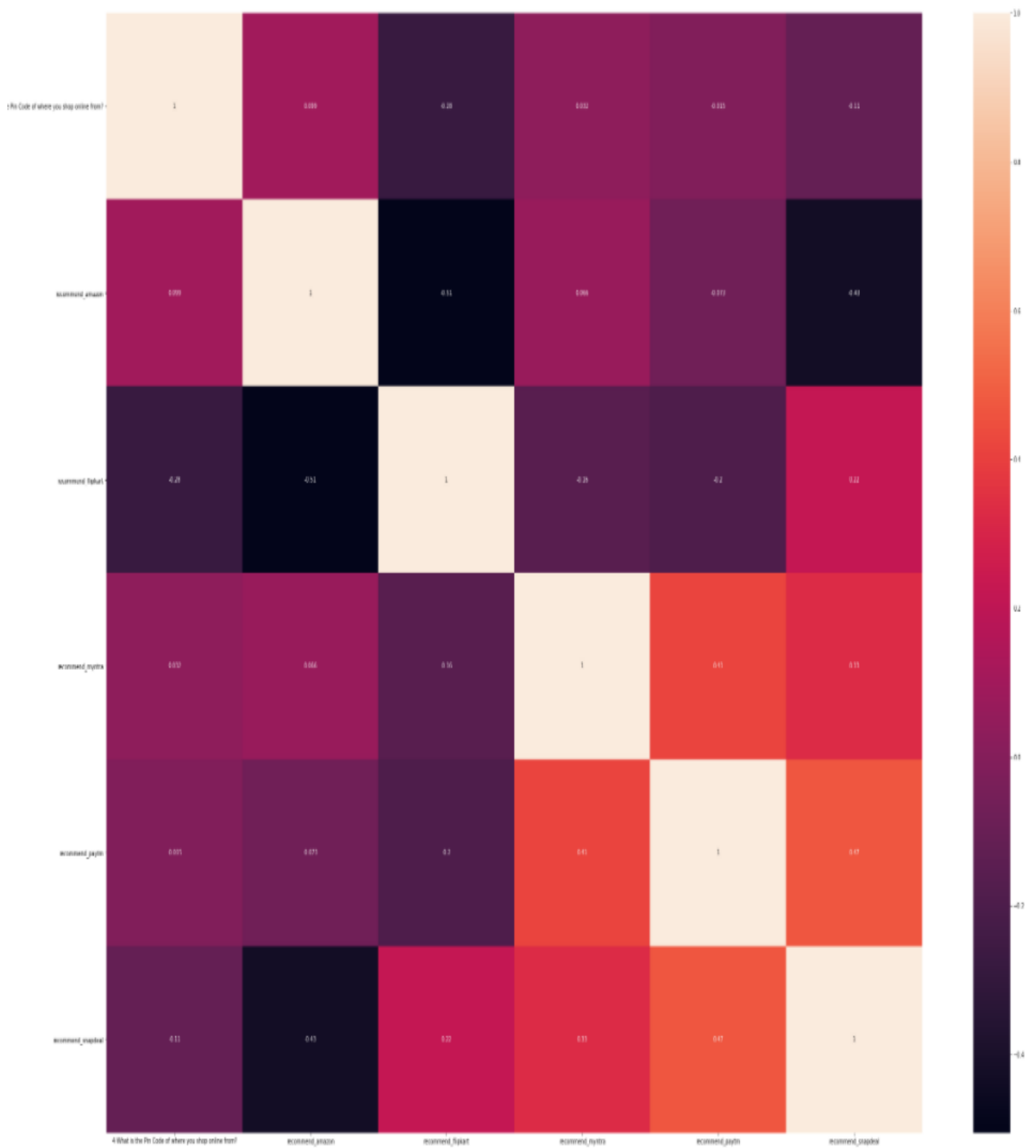
plt.tight_layout()

```



```
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

As in the final conclusion in which user were asked which online retailer they would recommend to a friend. Most of the respondents says Amazon. in because it is providing all the features that users want. Website is efficient and it is fast loading, It give complete, relevant description and information of products. It is reliable and quick to complete the purchase. Amazon give speedy delivery to its customers and there is several payment option available, on the website. It provide online assistance through multi channels. Providing good deals on products. Amazon have a user friendly interface and has visual appealing webpage layout. Amazon also offers wide variety of products and its application is easy to use. lastly the main thing why user recommend it is because of its Trustworthiness and also its robust Security in protecting customer financial information and their Privacy information. They are some cons like the amazon website topping the list In Frequent disruption when moving from one page to another this con company should see and improve it to give overall best experience to the users. This paper investigated the factors that influence the online customers repeat purchase intention. During the process various data processing methods has been used to clean the data. The project contains extensive EDA considering every aspect. The major finding is Complete relevant description of products and reliability of the website increases chances of customer retention. However, if the content on the website in not easy to read and understand or can't guarantee the privacy of the customer will lead to chances of customer retention. This project has increased my understanding of the concept. During the research I came across various challenges and while solving them I learned a lot of new things. How to plot different charts. For example, I learned how to plot subplot. How to handle legends manually. How to group data and visualize that. The limitation of the solution provided is that the data carried a lot of unrealistic values.

Thank You