

# CUSTOMER RETENTION CASE STUDY

---

Problem Statement and understanding, EDA steps and visualizations

Customer satisfaction has emerged as one of the most important factors that guarantee the success of the 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 has been carried out to propose the models for customer activation and customer retention.

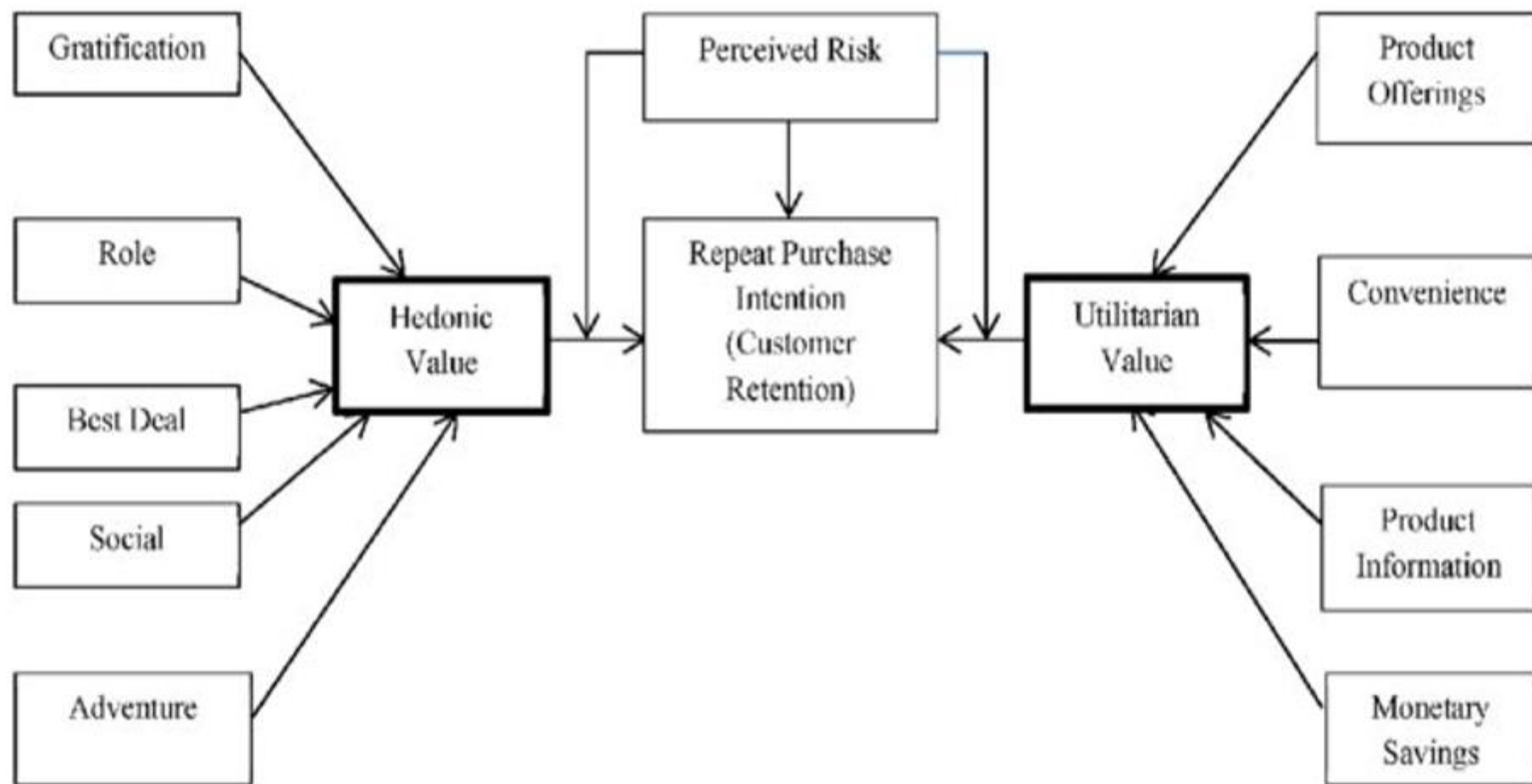
Five major factors that contributed to the success of an e-commerce store have been identified **service quality, system quality, information quality, trust, and net benefit.**

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

The combination of both **utilitarian value** and **hedonistic values** is needed to affect the repeat purchase intention (loyalty) positively.

The data is collected from Indian online shoppers. Results indicate the e-retail success factors, which are very much critical for customer satisfaction.

## FACTORS AFFECTING THE PURCHASE INTENTION OF THE CUSTOMERS



# HEDONIC VALUE

**Hedonic value** is defined as the **value** a customer receives based on the subject experience of fun and playfulness.

**Hedonic** goods are bought so that the consumer may gain pleasure and enjoyment from the good, and **value** experiences are also viewed as **hedonic** experiences.

With hedonic value, the value received is provided entirely from the actual experience and emotions associated with consumption, not because some other end is or will be accomplished.

For instance, a consumer visits your company page and happens upon an entertaining video you posted or a fun survey/poll you created.

Seldom does one look at pictures you posted in an effort to get a job done, yet features like these provide value that is emotional and subjective in nature and is an end in and of itself rather than a means to an end.

# Utilitarian Value

**Utilitarian value** is defined as the **value** that a customer receives based on a task-related and rational consumption behavior.

**Utilitarians** usually are products that add ease to your every day; e.g basic car, fridge, phone. **Utilitarian needs** are bought without second guessing and have little emotional and sensory attachment.

Another example of Utilitarian value can be, Hammer, the value of a hammer is solely depends upon the function that the hammer provides.

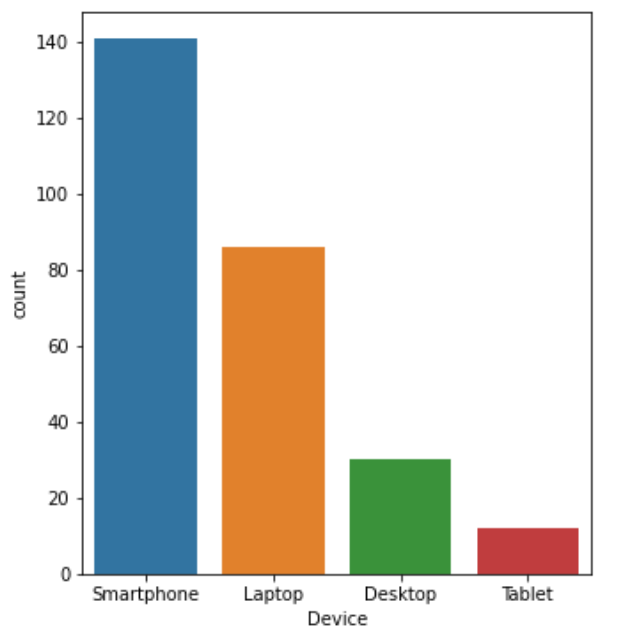
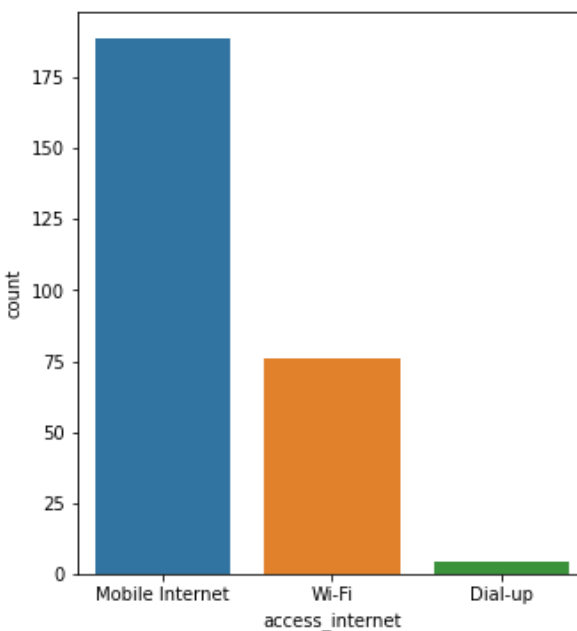
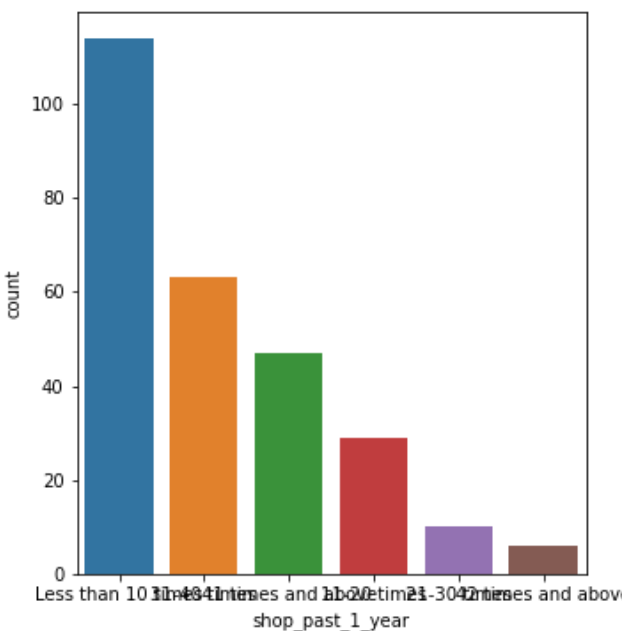
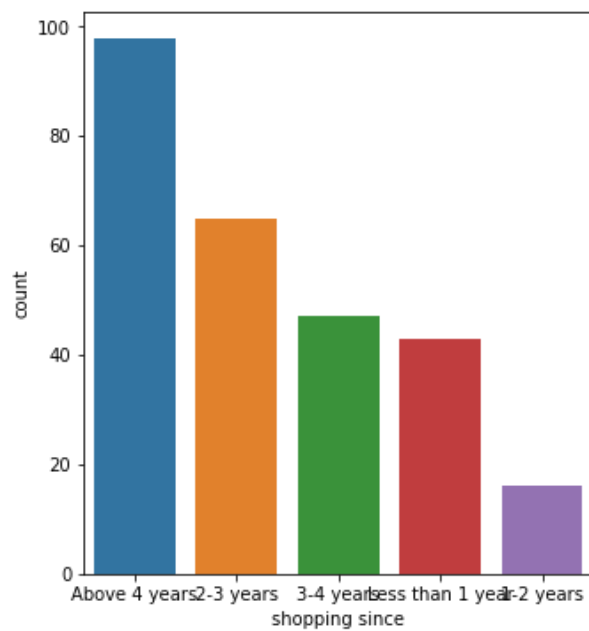
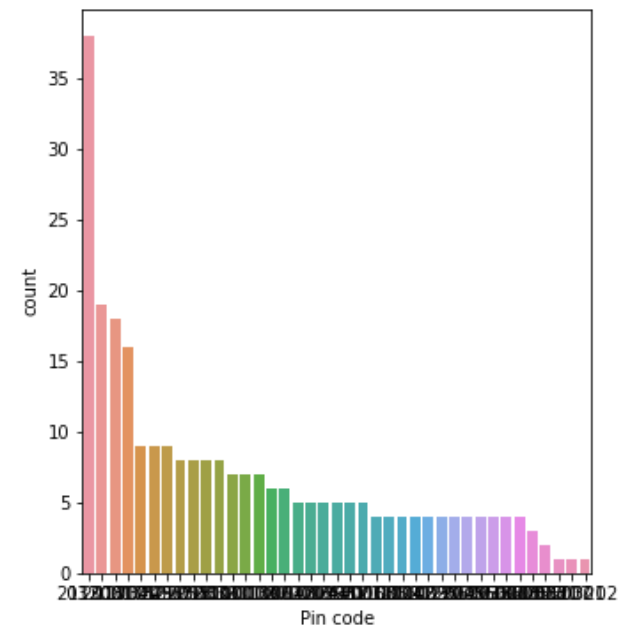
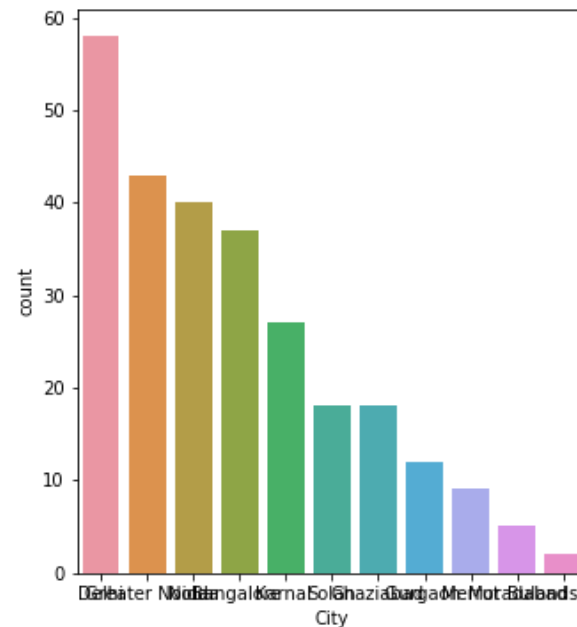
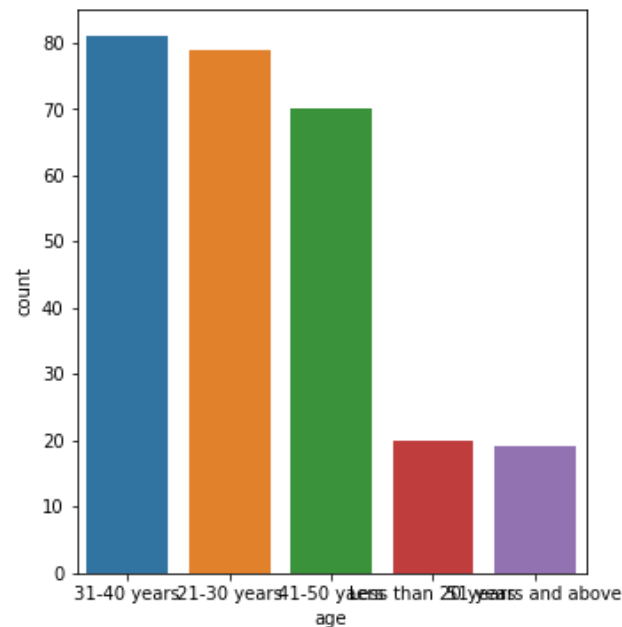
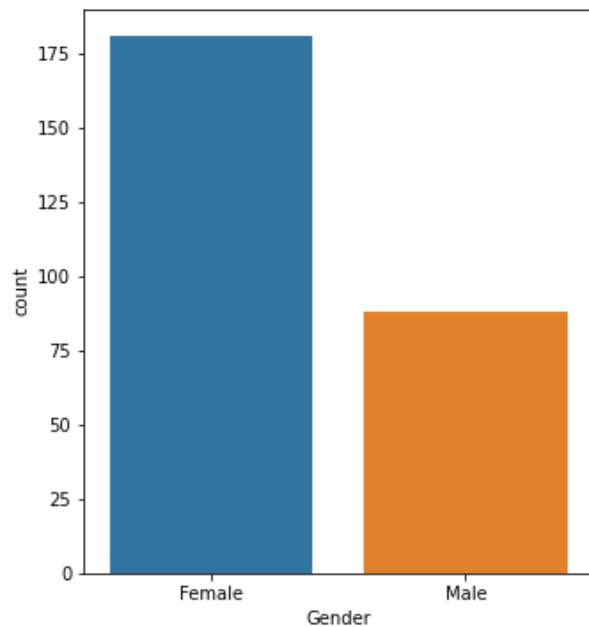
Car – The utilitarian value of a car is that it is a mode of transportation, however, if we look at this as a hedonic value then the luxury of the car will show its hedonic value

# ABOUT DATA

We have data from 269 users collected from a survey, in this data set users were asked about their suggestions for online shopping from different online sites like Amazon, Flipkart, Myntra, Paytm, Snapdeal, etc. Our data includes all the factors which are part of both Hedonic and Utilitarian values.

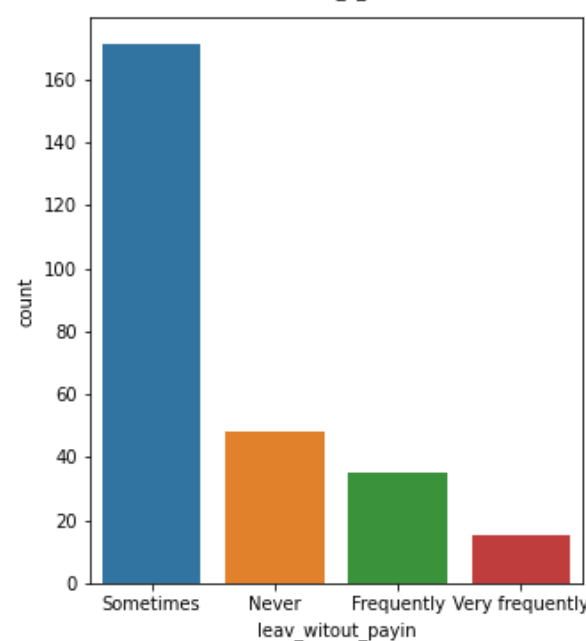
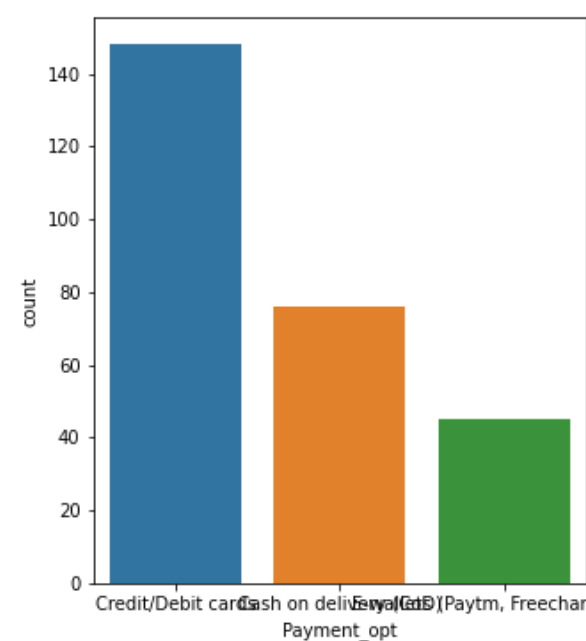
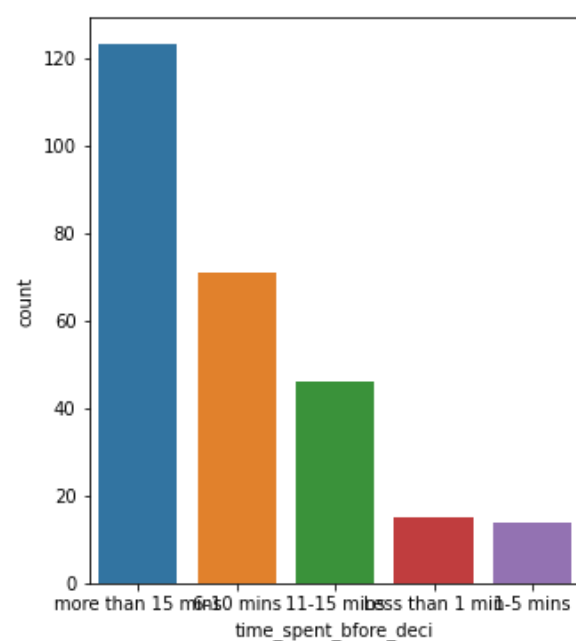
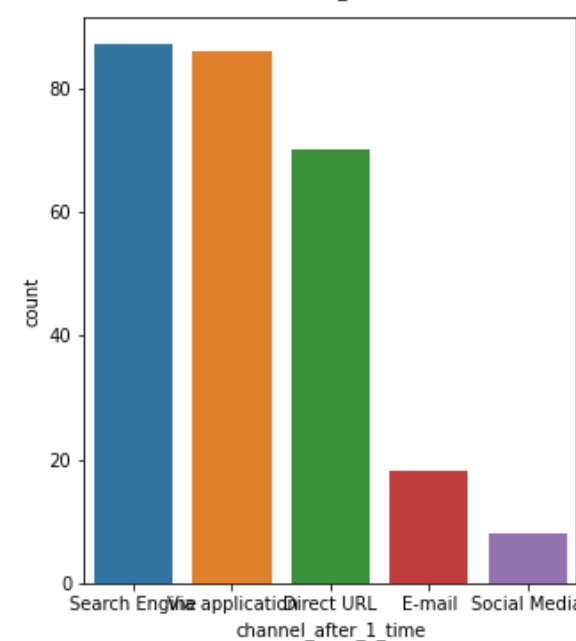
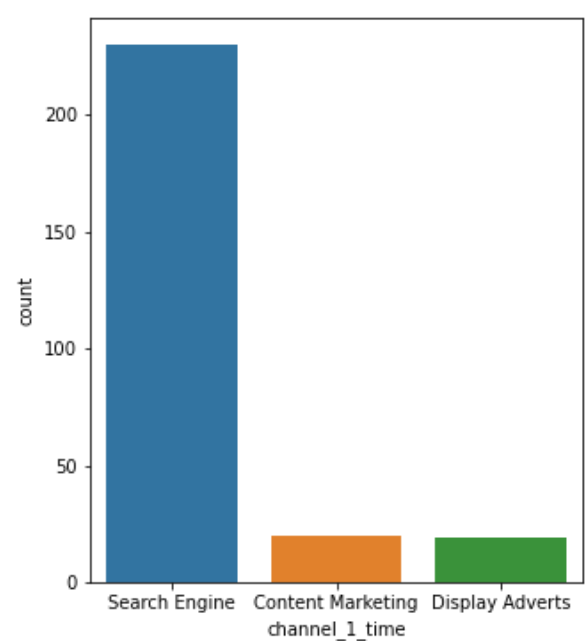
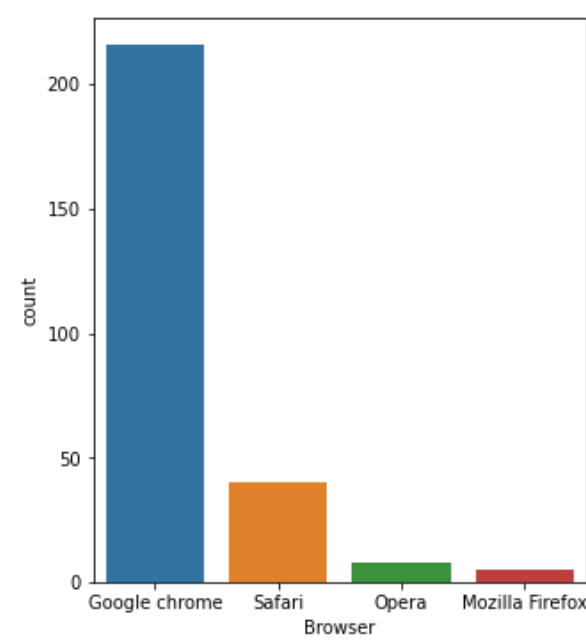
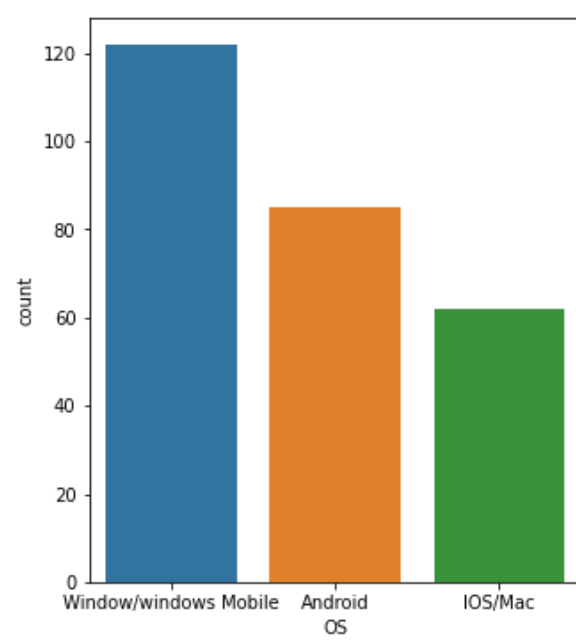
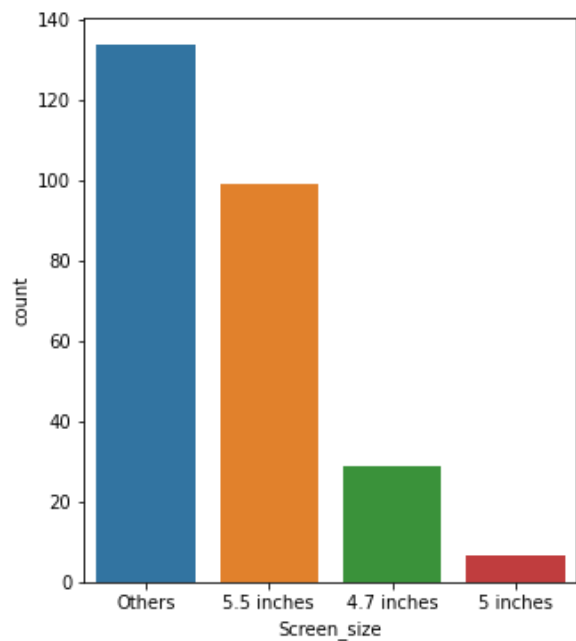
We have 71 variables and all of them are categorical in nature. These variables are used to check the purchase decision of the customers. Based on their experience how they rate online shopping for a particular site.

1. **\*\*\*Gender of respondent\*\*\*** - Here we can see that the number of female customers are more than of the males. Count of Female customers are more than 175 where male customers are ranging between 75-100.
2. **\*\*\*Age of customer\*\*\*** - we can see that the customer aged between 21 years to 50 years are more likely to shop online. 31-40 years shows the maximum number of shoppers followed by 21-30 and 41-50 years.
3. **\*\*\*City\*\*\*** - we can see that most of the online shopping is done in Delhi followed by Greater Noida and Noida, Moradabad and Bulandshahr shows minimum online shopping customers.
4. **\*\*\*Pin code\*\*\*** - we can see that 201308 which is the Pin code of Noida shows maximum online shopping customers, this is only if we compare with unique pin code otherwise Delhi has the most number of online customers.
5. **\*\*\*Shopping duration\*\*\*** - most of the customers are doing online shopping for more than 4 years.
6. **\*\*\*Shopping in last 1 year\*\*\*** - this shows that most of the customers shopped less than 10 times, the reason could be that they don't use online shopping for their basic needs like food, groceries. They use online shopping for the products which has high durability.
7. **\*\*\*Internet access\*\*\*** - customers usually use mobile internet the most for online shopping, the reason could be that using mobile app and mobile internet they can shop any where any time without having a wifi nearby.
8. **\*\*\*Device for online shopping\*\*\*** - Smartphone is most preferable by the customers as it provides ease in navigation and more attractive , easy to use. However we can see that laptop and desktop are classified uniquely here, if we merge these two then count of the customer using these 2 devices will be closer to the count of the customers using smartphones.

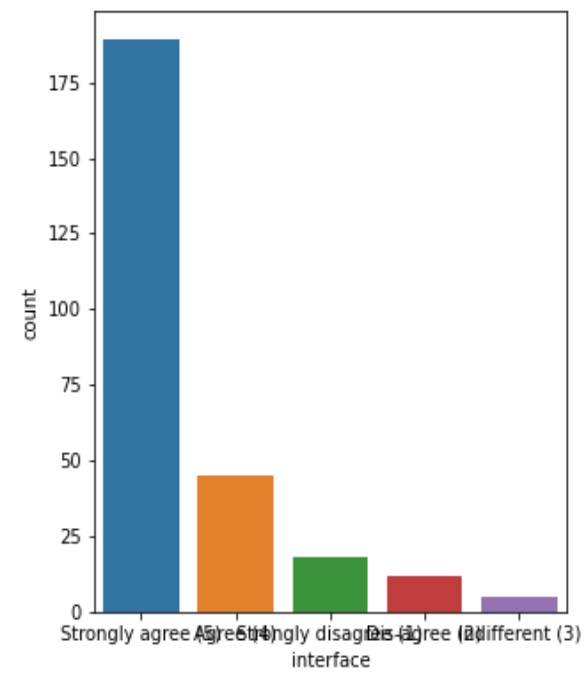
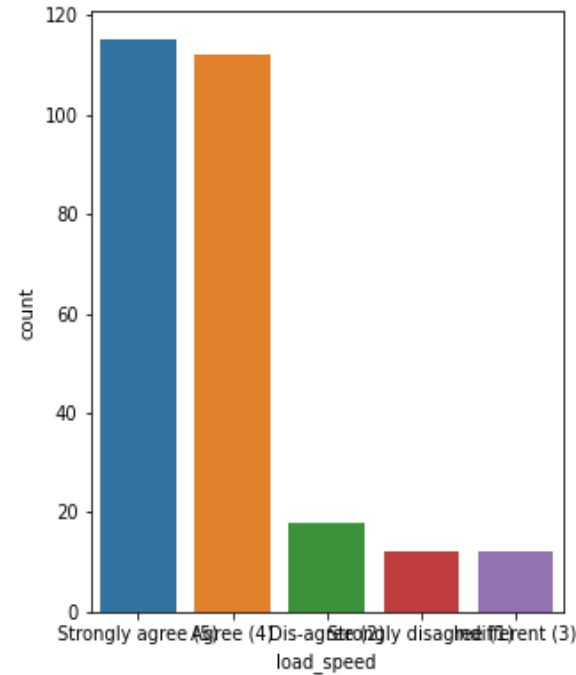
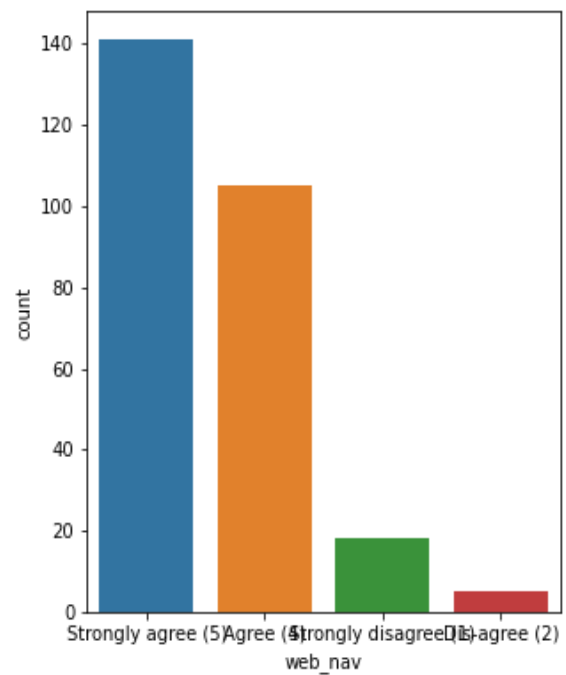
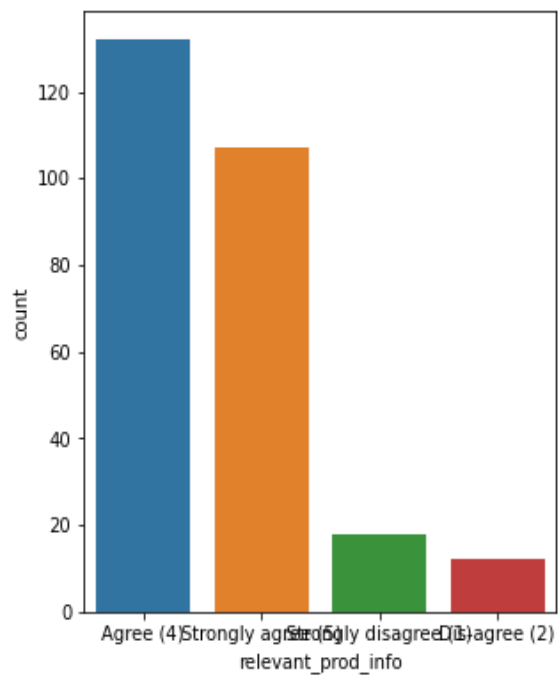
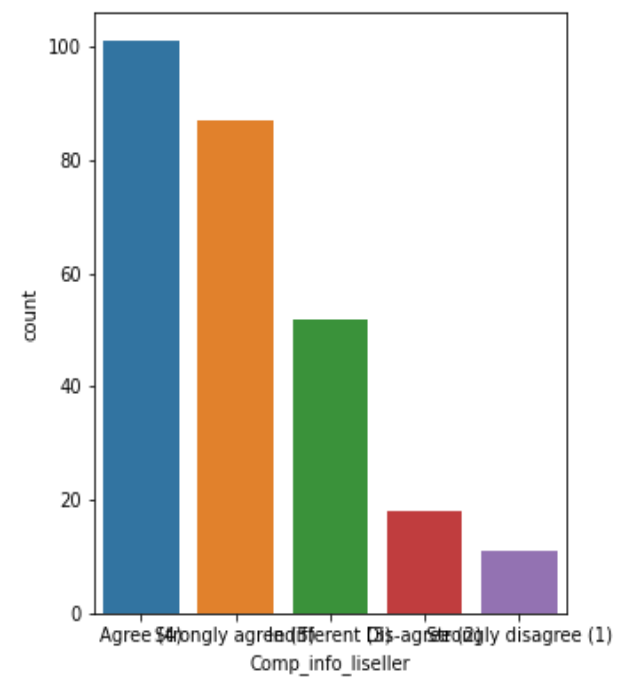
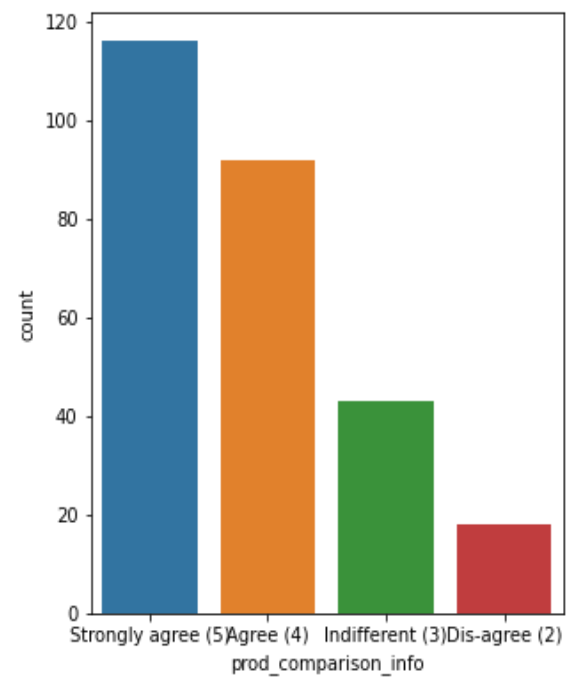
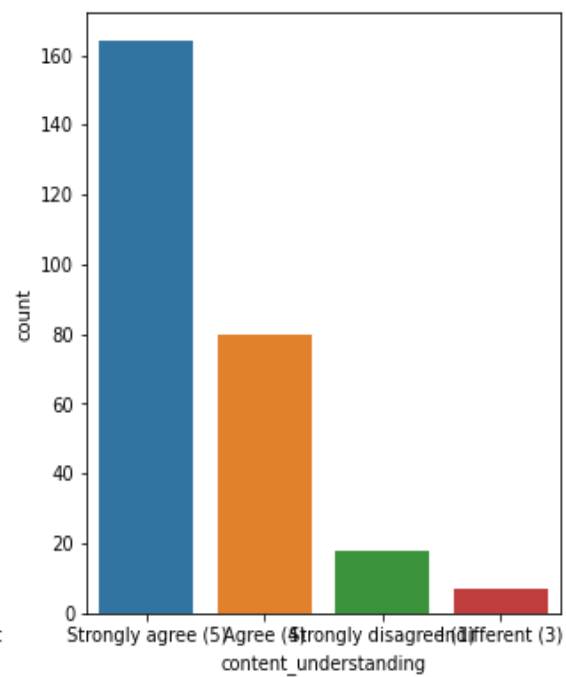
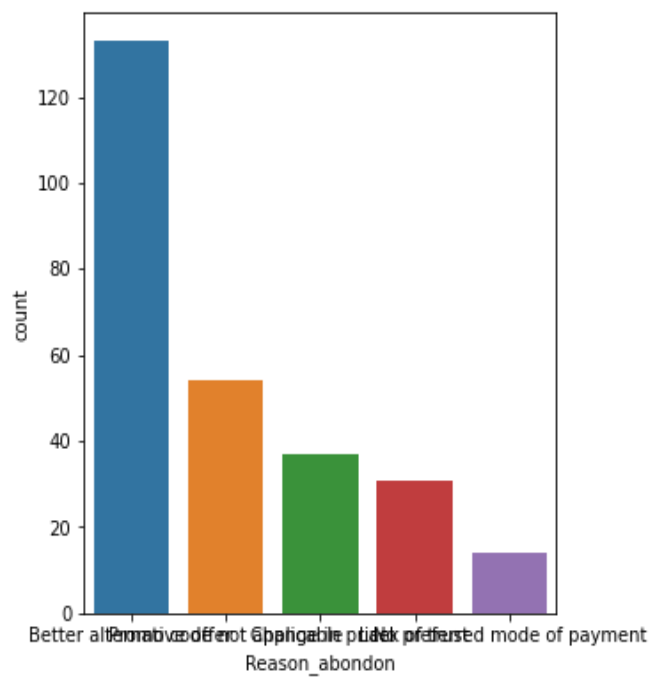




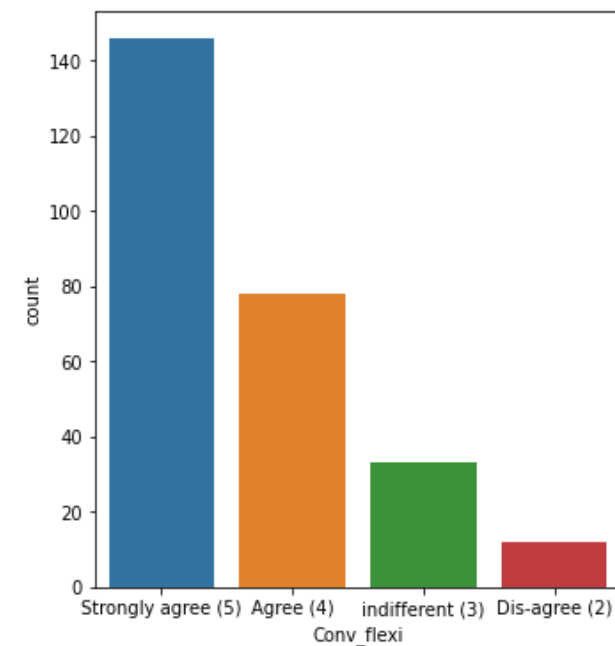
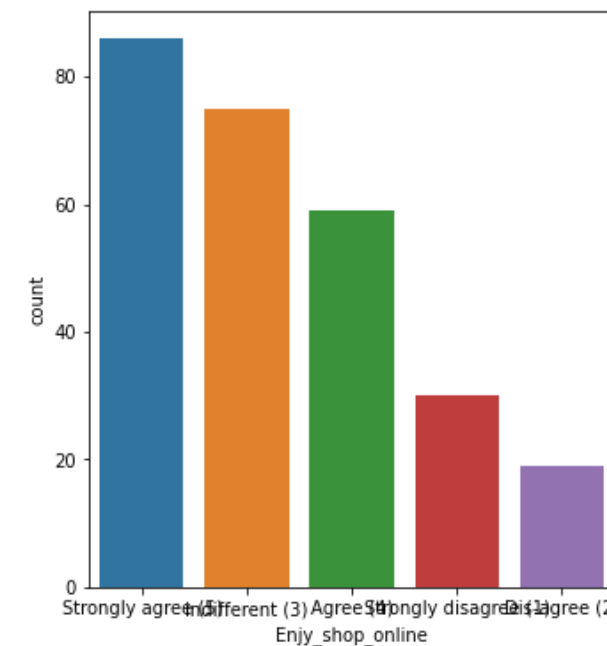
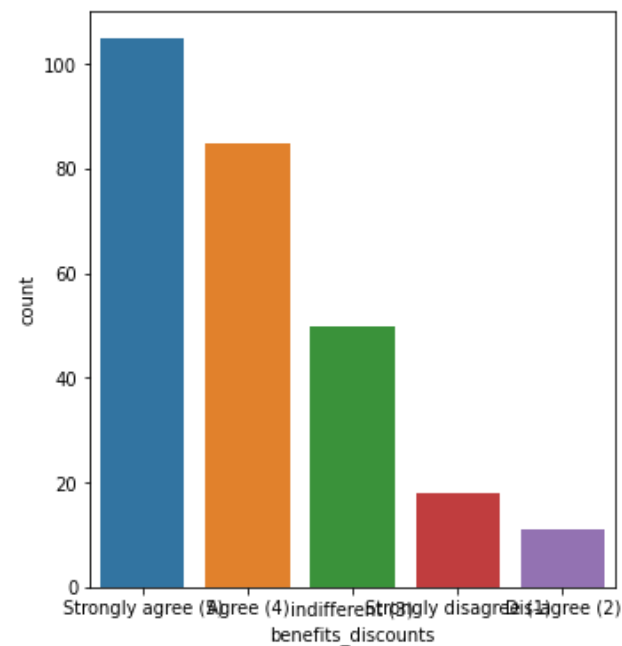
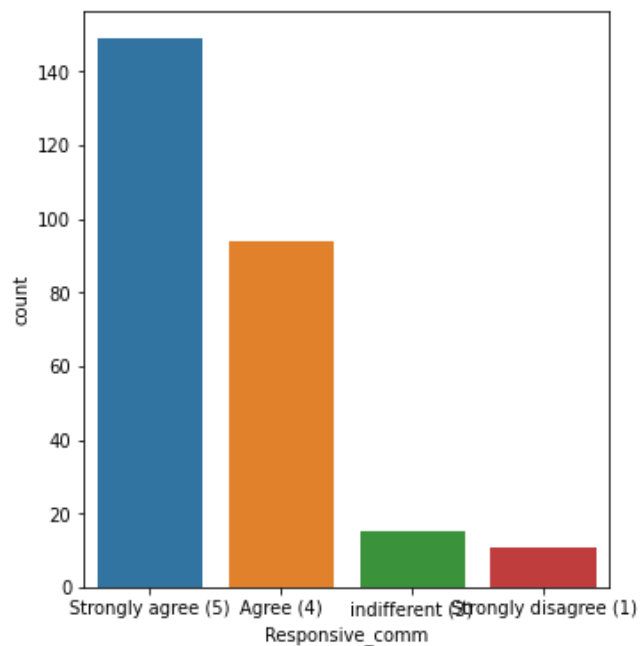
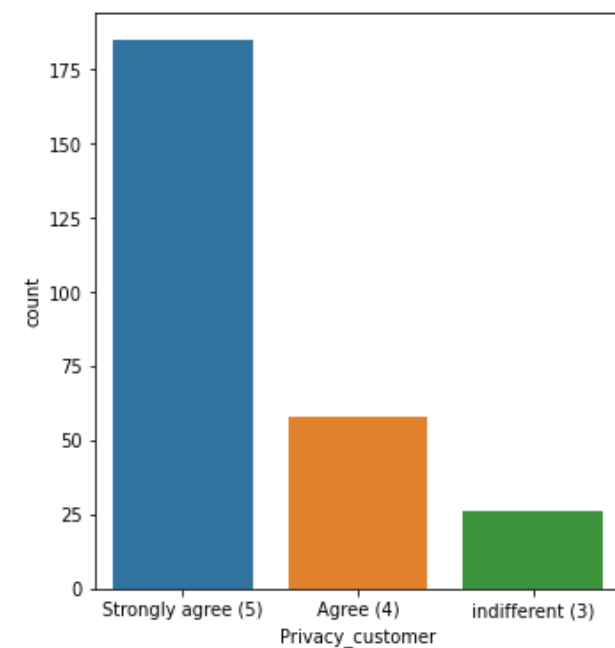
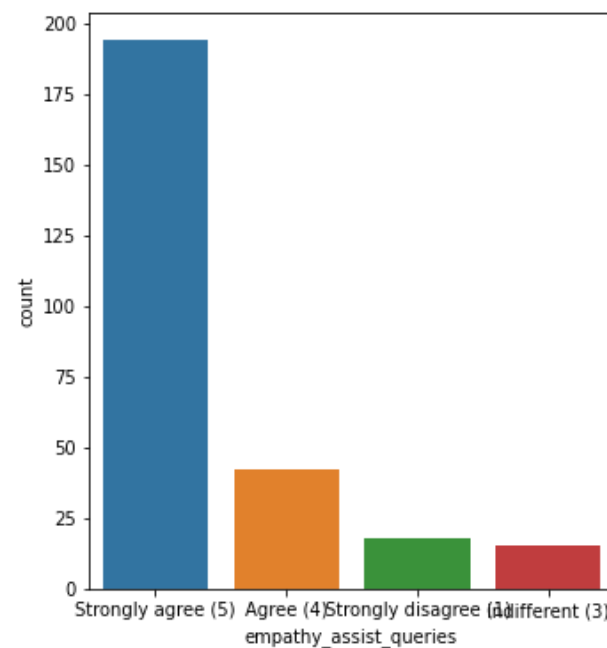
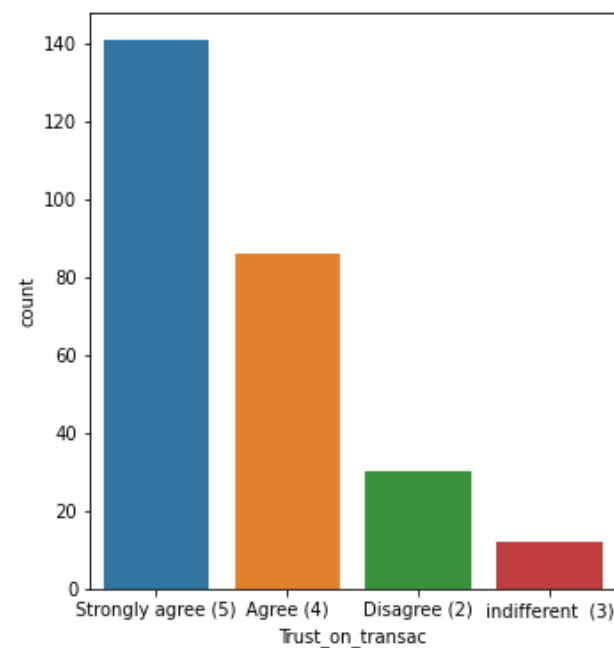
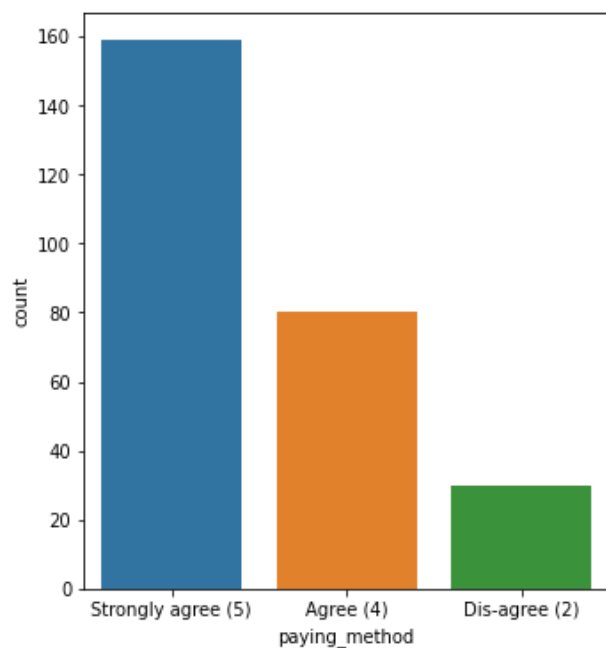
9. \*\*\***Screen size**\*\*\* - 'Others' here denotes the size of laptop or computer as usually their screen size is more than 10 inches. Others shows the maximum count here followed by 5.5 inches.
10. \*\*\***OS type**\*\*\* - windows shows the maximum count as this can include both smartphones and desktop/laptop. In India user base of android is more than of apple IOS that is why here we can see that android based customers are more than of the IOS.
11. \*\*\***Browser**\*\*\* - Google chrome is the most widely used web browser across world, An estimated 2.65 billion internet users globally use Chrome as their primary browser. That is why our data also shows that the customers are more likely to use Google chrome while they search for things online and shop online
12. \*\*\***Channel**\*\*\* - customer do online shopping using search engine as it will give them many options available.
13. \*\*\***Channel after 1 visit**\*\*\* - most customers still goes for using search engine followed by app users.
14. \*\*\***explore time**\*\*\* - most customers usually explore more than 15 mins, the time taken due to comparison of the product, reading product details, alternatives for the product chosen.
15. \*\*\***Preferred payment option**\*\*\* - Credit cards and Debit cards are most widely use of payment method , the reason is simple that on these payment mode customer gets many offer , some banks also partner with the online shopping site for the benefits of the customers.
16. \*\*\***abandon while shopping**\*\*\* - most of the customer abandon the shopping cart, the reason could be that they find a better alternative, promo code isn't working, delivery of product takes too long etc



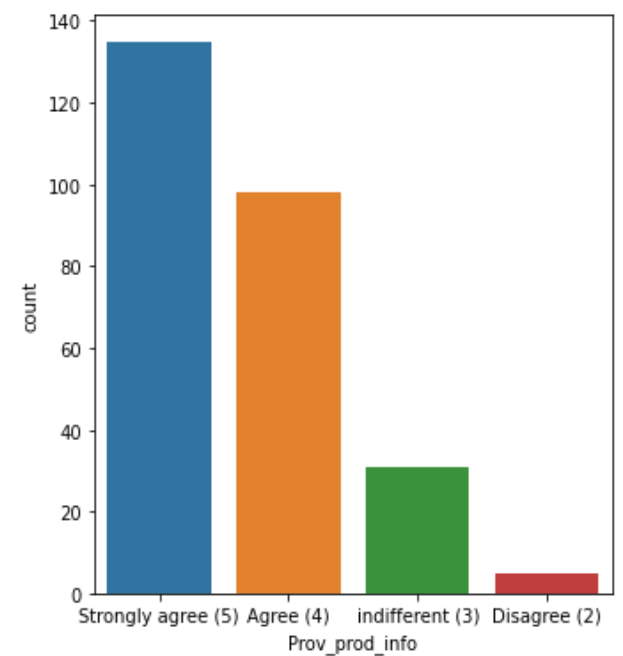
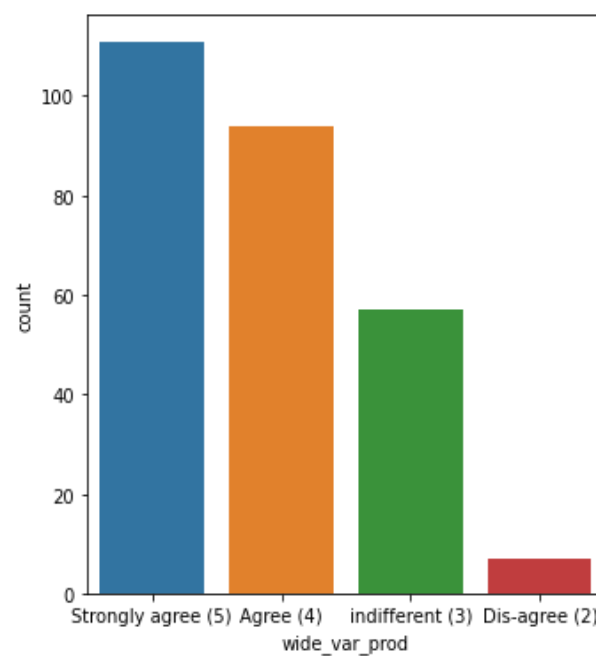
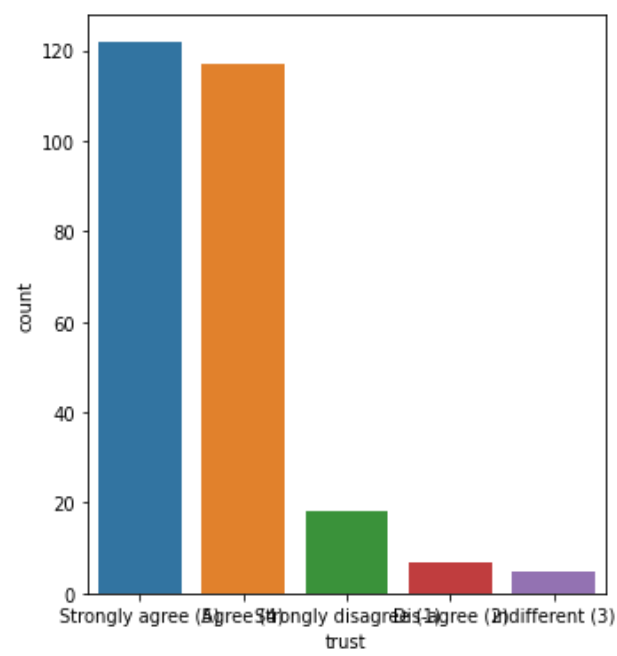
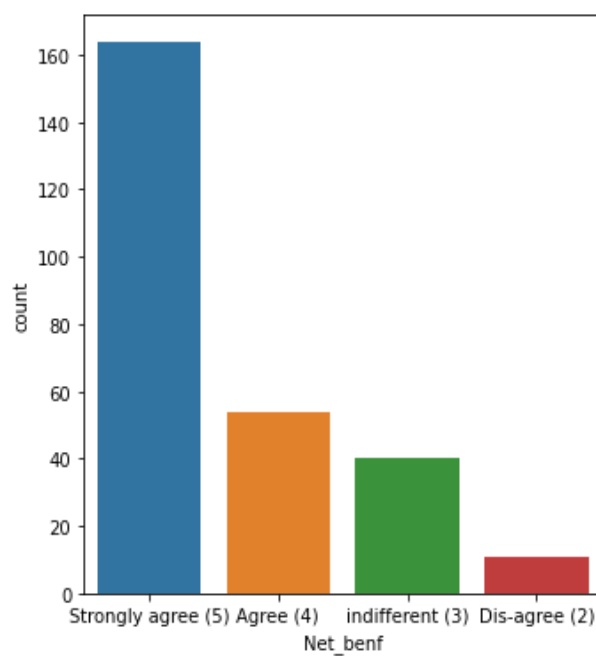
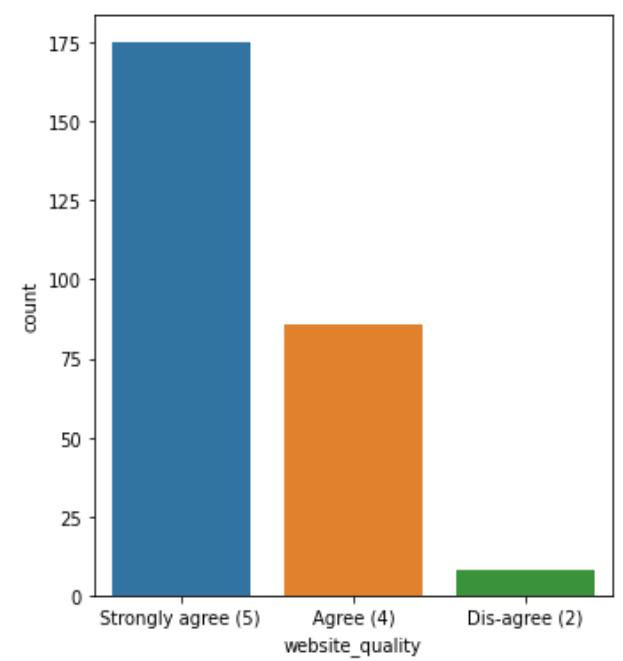
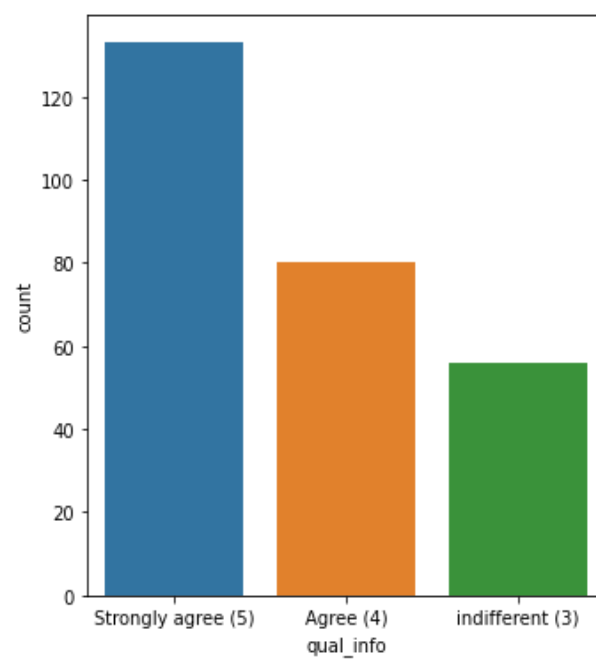
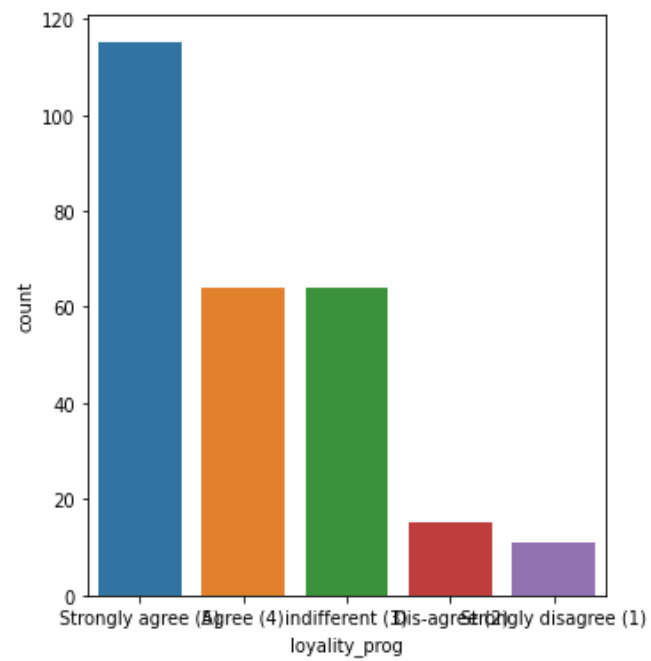
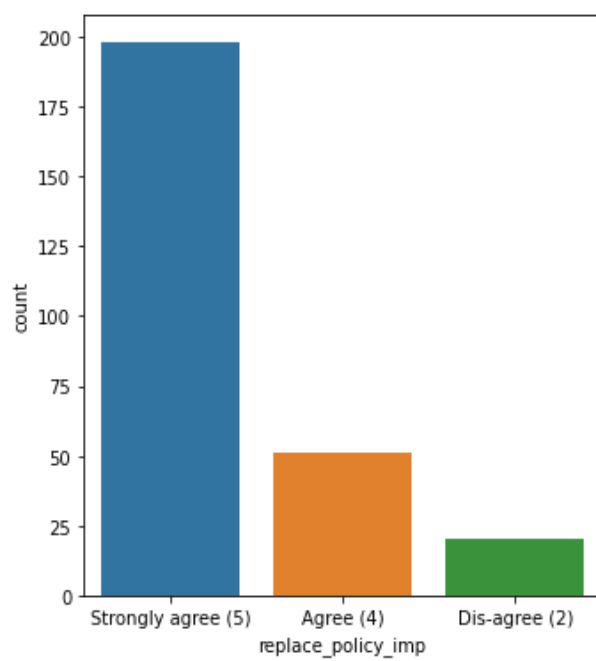
17. \*\*\***Abandoned cart reason**\*\*\* - The most common reason for abandoning the cart is having a better alternative offer, and the promo code not applicable.
18. \*\*\***website content is easy?**\*\*\* - Most customers agree that the content on the website is much easier to understand.
19. \*\*\***similar product information**\*\*\* - most of the customers find information about similar products useful for comparison.
20. \*\*\***complete product information**\*\*\* - this affects the purchase decision of the customer, Most of the customers find this important while they do online shopping.
21. \*\*\***product information**\*\*\* - most of the customers feel that the listed product should have complete details of it. This will affect the purchase decision of the customer.
22. \*\*\***Navigation on website**\*\*\* - Every customer wants to have easy navigation steps while doing online shopping.
23. \*\*\***Loading and processing speed**\*\*\* - most of the customers feel that loading and processing speed also affect their purchase decision.
24. \*\*\***Interface of website**\*\*\* - Most customers like having a good interface of the website, which attracts customers and affects their purchase decision.



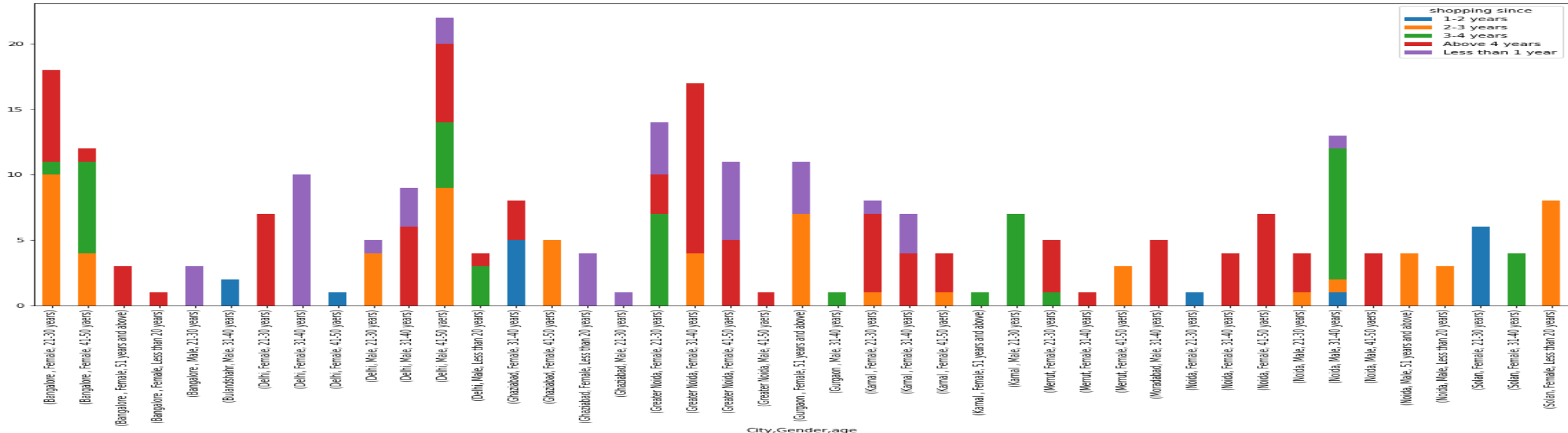
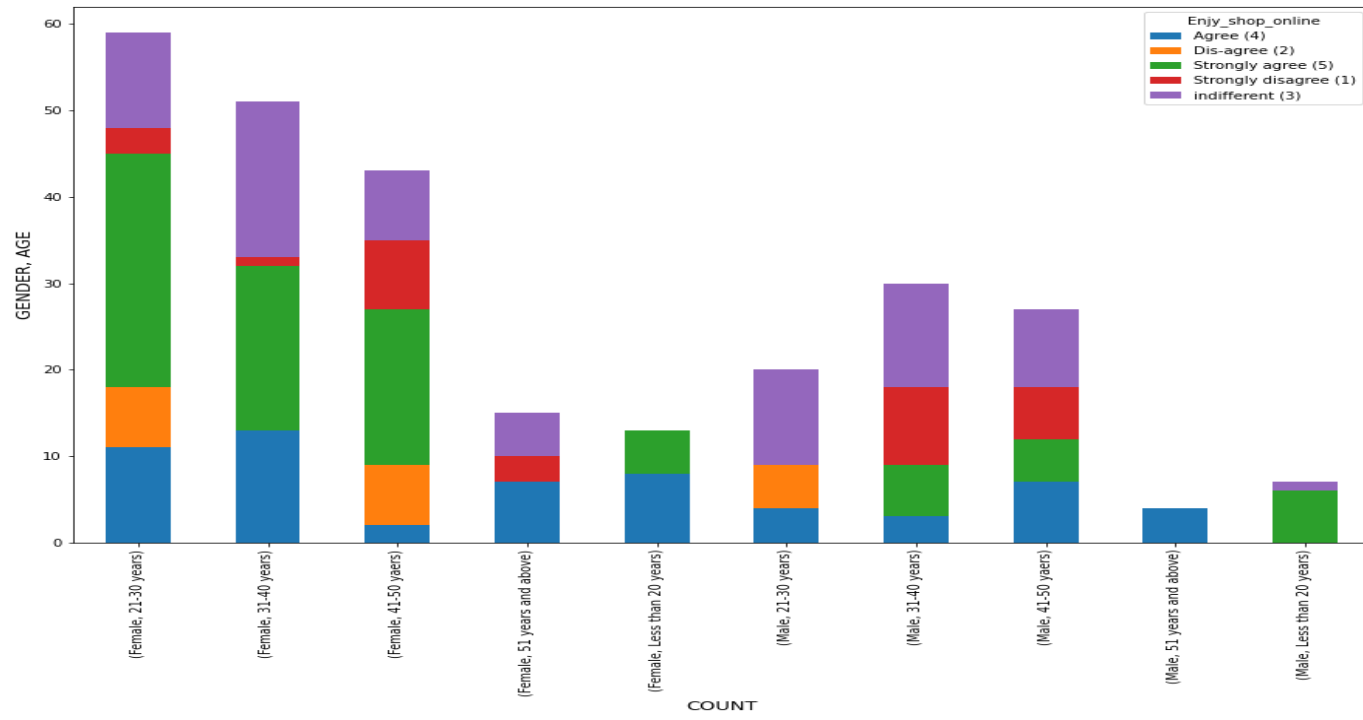
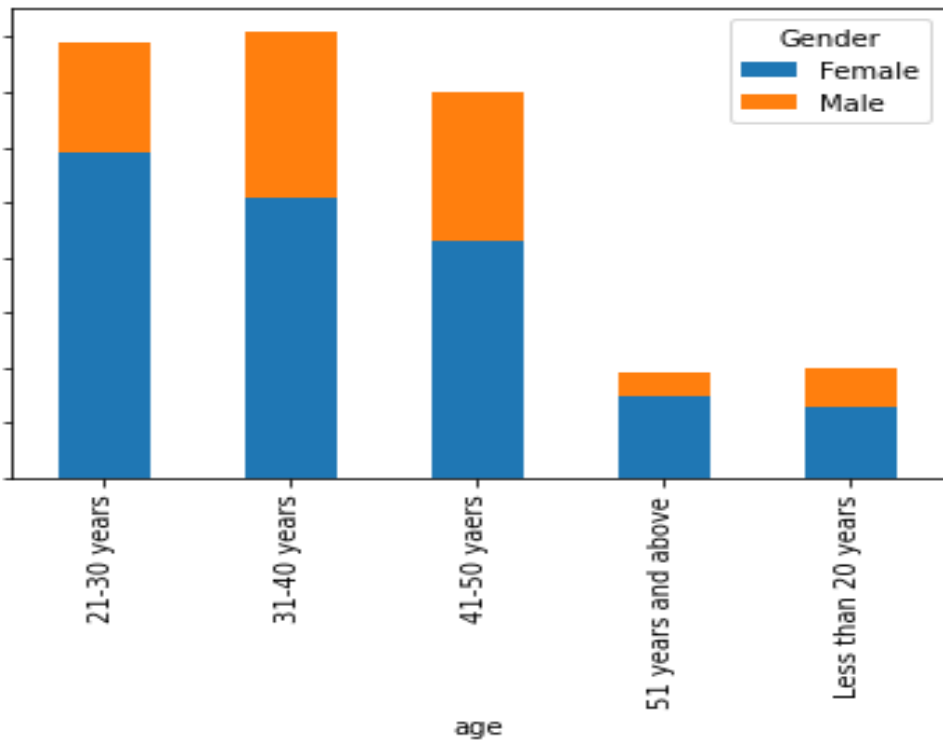
- 25. \*\*\*Payment convenience\*\*\*** - most of the Customer (almost every) wants a convenient payment method.
- 26. \*\*\*Trust on online retail store\*\*\*** - Most of the customers believes that the respective online retail store will full fill the transaction within given time by the merchant.
- 27. \*\*\*Empathy\*\*\*** - bases on query solving on a particular site like amazon, flipkart it is very important to resolve the customer's query as this will affect the customer purchase decision in future
- 28. \*\*\*Data Privacy\*\*\*** - This is becoming very important concern nowadays, as customer's details are very sensitive and customer will only provide data to those which have high data security., So most customers agree on this that there should be data security.
- 29. \*\*\*Communication channel\*\*\*** - There are various communication channels available like email support, chat support, phone support etc, customer wants communication medium of their choice. They agree that the online store should have various communication channels available.
- 30. \*\*\*Benefits and discounts\*\*\*** - Most of the customers agree that online shopping gives monetary discount and benefits, like cashback, reward points etc.
- 31. \*\*\*enjoyment\*\*\*** - most of the customers like shopping online, it is hassle free and time consuming, however there are some customer's who don't like online shopping as much. However most of the customers agree that they enjoy doing shopping online.
- 31. \*\*\*flexibility and convenience\*\*\*** - Online shopping is flexible and convenience, most of the customers agree on that.



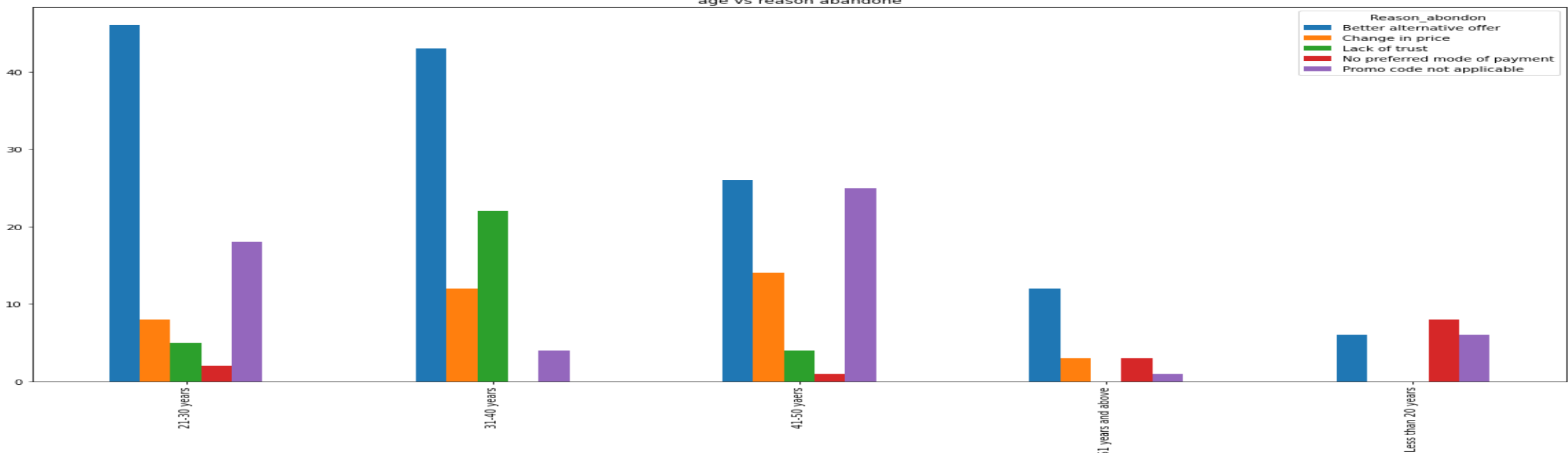
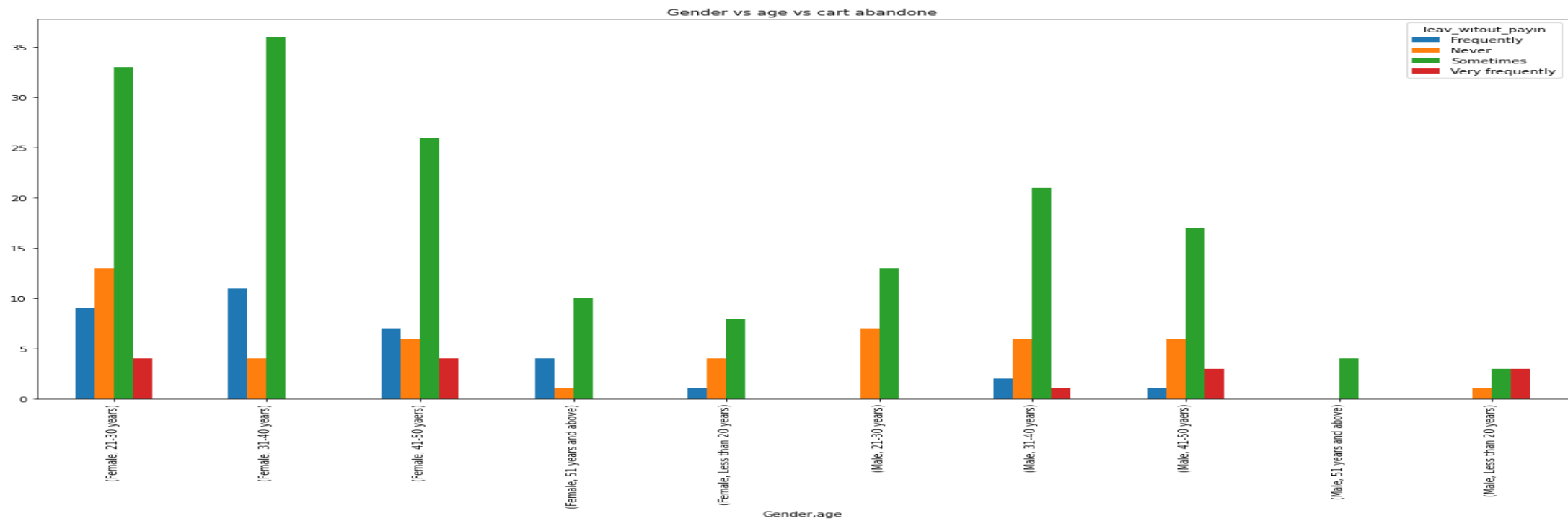
- 33. \*\*\*Return/replacment policy\*\*\*** - most of the costumers agree on having return and replacement policy.
- 34. \*\*\*Loyalty programs\*\*\*** - this refers to offers, rewards offered to customer so that the online retail store retain the customers. Example for this can be Flipkart's super coin, where for every purchase you will earn some amount of super coin that can be redeemed in future. Such policy attract customers and most of the customer's agree that some rewards option attract them.
- 35. \*\*\*quality information\*\*\*** - Most of the customers find this important.
- 36. \*\*\*website/application quality\*\*\*** - most of the customer's agree that a website or application quality should be good to retain the customers.
- 37. \*\*\*Net benefits\*\*\*** - Most of the customer's agree on net benefits can lead to customer's satisfaction.
- 38. \*\*\*trust vs satisfaction\*\*\*** - most of the customers agree on that trust results in customer's satisfaction.
- 39. \*\*\*variety of products\*\*\*** - most of the customers agree on that online shopping shows wide variety of the products available.
- 40. \*\*\*relevant product info\*\*\*** - most of the customers agree on this.







- Here we can see that number of female customers in all age categories are more than of male, from age 21-30 years count of female customers is the most.
- Data has more female customers.
- Bulandshahar consists of only Male customers age between 31-40 years have been shopping online for 1-2 years.
- Moradabad consists of only Male customers age between 31-40 years have been shopping for more than 4 years.
- Gender ration is approx. equal in Delhi.
- Bangalore shows maximum female customers age between 21-30 years.
- Delhi shows maximum Male customers age between 41-50.
- Female with age 21-30 are strongly agreeing that enjoying shopping.



# OBSERVATIONS

- Females are more likely to abandon the online shopping cart.
- Customers below 20 years old don't abandon that much.
- Most the of the customers age between 21-50 years abandon the online shopping cart due to better alternative offers available to them. These are the customer tends to explore more while shopping online.

## CHECKING IF AGE AND REASON TO ABANDON ARE DEPENDENT OR NOT

To check the dependencies of independent categorical variable I have used chi square test. And based upon my test score I found out that these 2 variables are highly dependent.

My null hypothesis was that the variables are independent.

Using chi square test I have calculated the critical value and Chi statistic value and compared the scores

To check if null hypothesis is acceptable or not.

Here we can see that the age of the customer and the Reason for abandoning the online shopping cart is very highly dependent with each other.

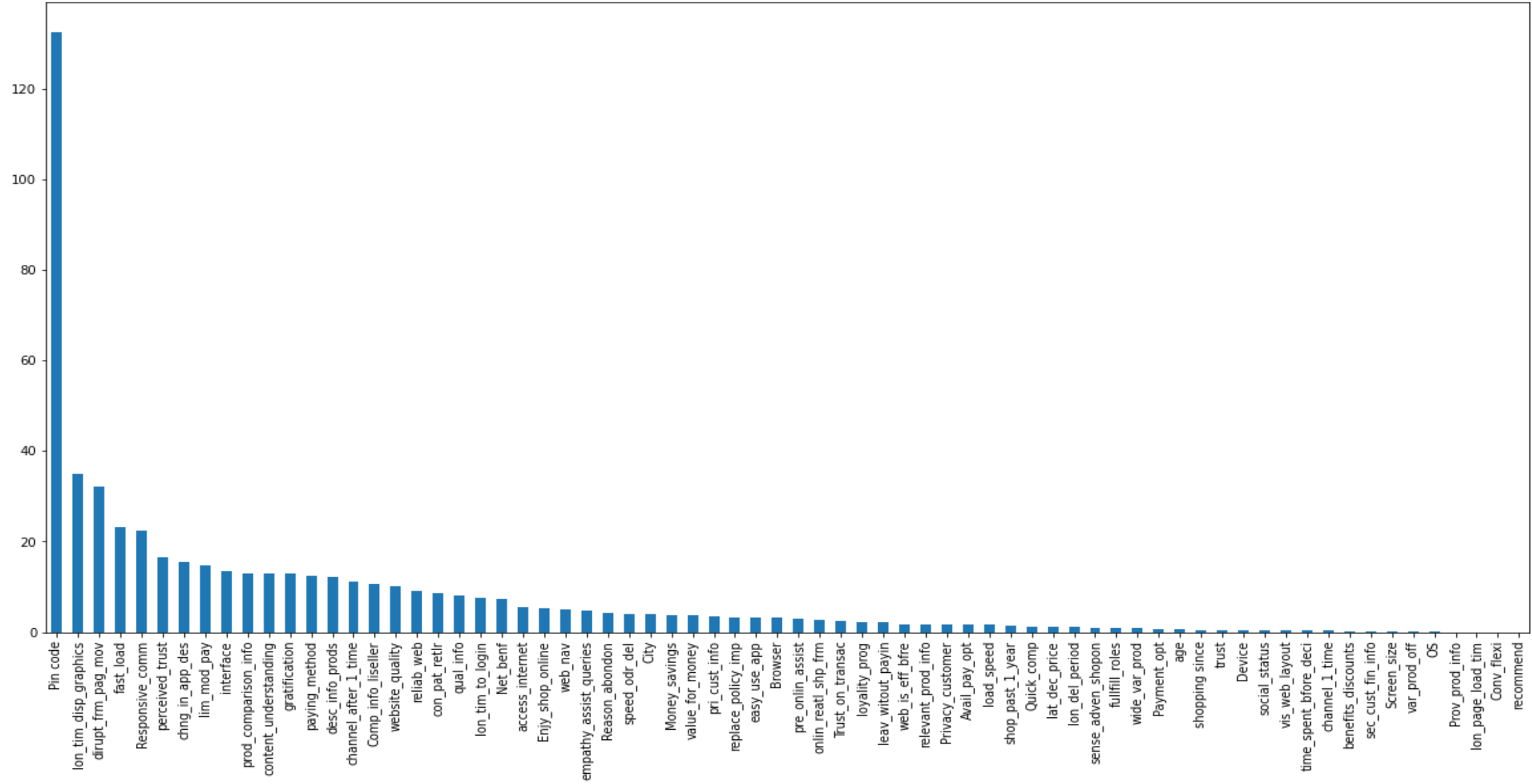
```
In [97]: chi2_param(pd.crosstab(df['age'],df['Reason_abandon'])))
```

```
degree of freedom is = 16
stats is = 115.99491738190628
p-value is = 3.2225346232096846e-17
expected values are:
[[39.05947955 10.866171    9.10408922  4.11152416 15.85873606]
 [40.04832714 11.14126394  9.33457249  4.21561338 16.26022305]
 [34.60966543  9.62825279  8.0669145   3.64312268 14.05204461]
 [ 9.39405204  2.6133829   2.18959108  0.98884758  3.81412639]
 [ 9.88847584  2.75092937  2.30483271  1.04089219  4.01486989]]
```

```
critical value is = 26.29622760486423
chi2 value we calculated is = 115.99491738190628
Is the chi2 calculated value more than the critical value? : True
Dependent (reject H0)
```

# FEATURE IMPORTANCE USING CHI-SQUARE

Feature selection using chi2 test



I have used chi square test to check the importance of each independent variable towards the target variable Gender.

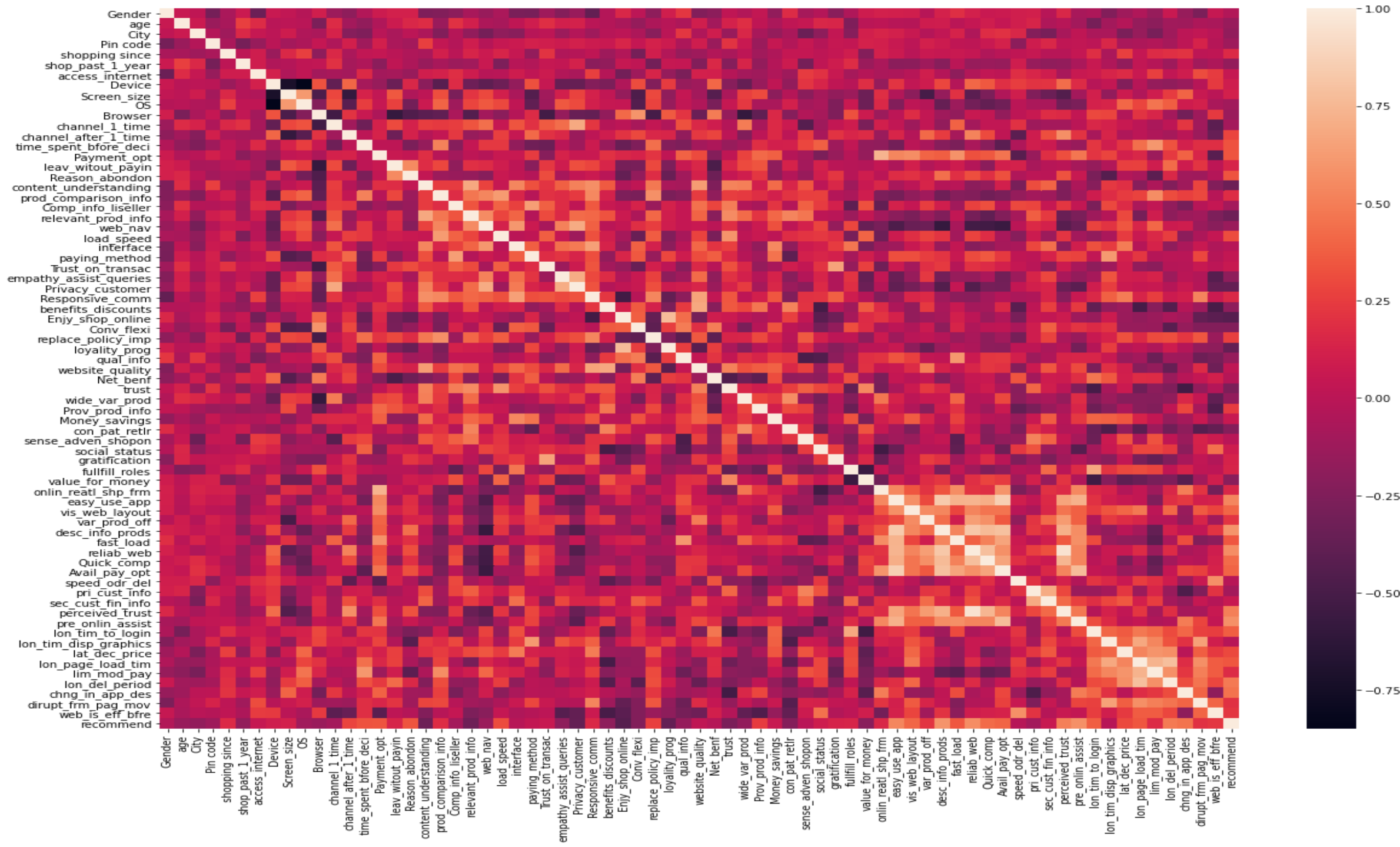
As we have categorical variable data so the chi- square test is the best to check the variable importance. I have calculated the variable importance depending upon the F-score of each variable. Higher the F-score means higher the importance of variable.

I have used Chi2 module/library from sklearn.feature\_selection and then plotted the bar graph in descending order to show the variable importance visually.

## OBSERVATIONS

- Based on chi-square test I can observe that hedonic type of variables are showing more importance to the target variable Gender.
- We will not consider Pin code here as it will not make sense here to classify the gender of the user, however based on pin code it can be checked the distribution of users (male and female) at a particular location.
- Interface, trust, responsive communication etc, all these types refers to the hedonic values. These variables showing more importance.

# CORRELATION HEATMAP



I have used spearman correlation technique to check the correlation between all the variables because most of the variables are Ordinal in nature.

We can see that there are many variables which are highly positive and negative correlated with each other.

We can remove one of the highly correlated variables and then feed the data for model building.

```
In [152]: so[so.values>.8].drop_duplicates()

Out[152]:
```

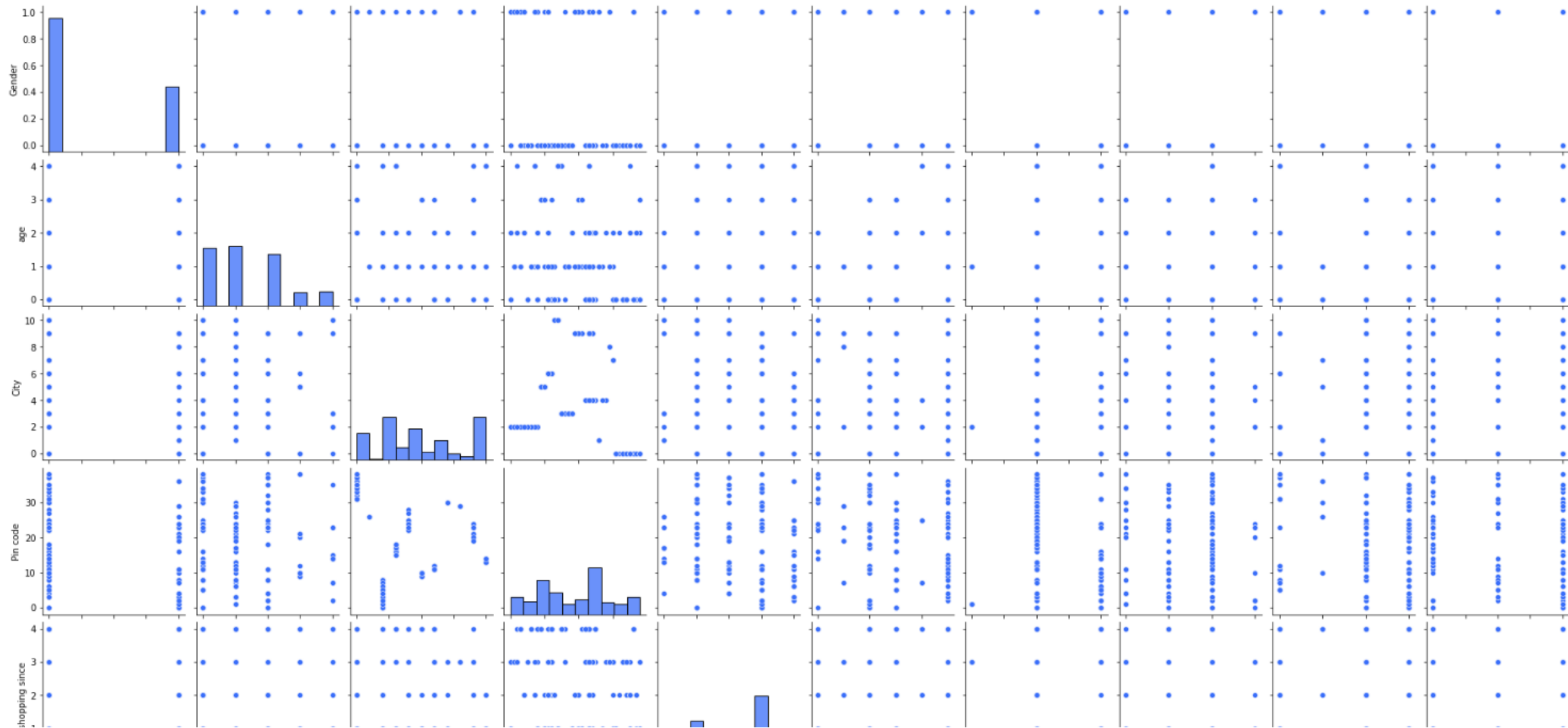
reliab_web	desc_info_prods	0.824962
desc_info_prods	easy_use_app	0.839925
easy_use_app	Avail_pay_opt	0.857014
reliab_web	perceived_trust	0.953129
Gender	Gender	1.000000

dtype: float64

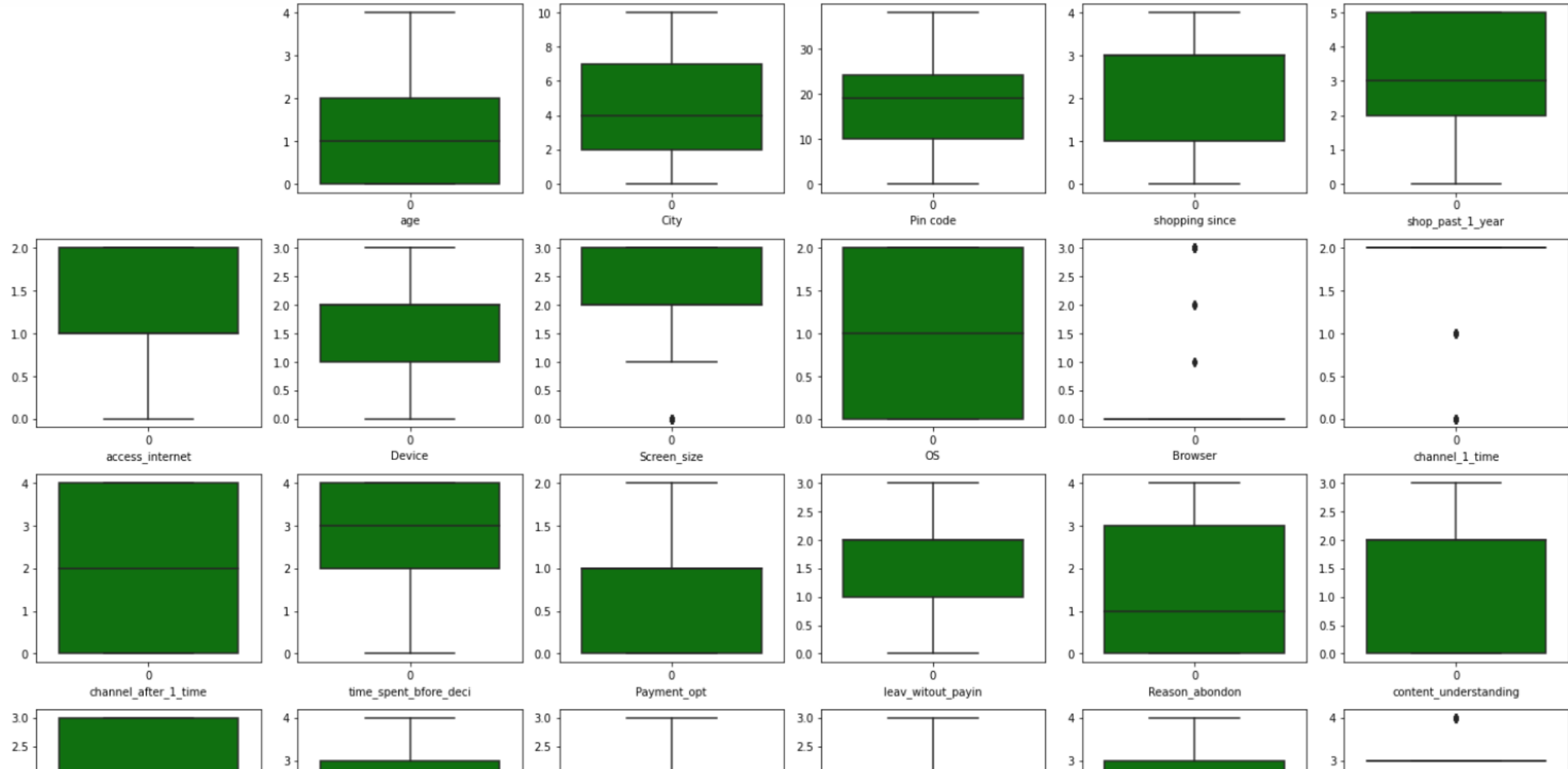
Above are some highly positive correlated variables, I took the threshold as 0.8

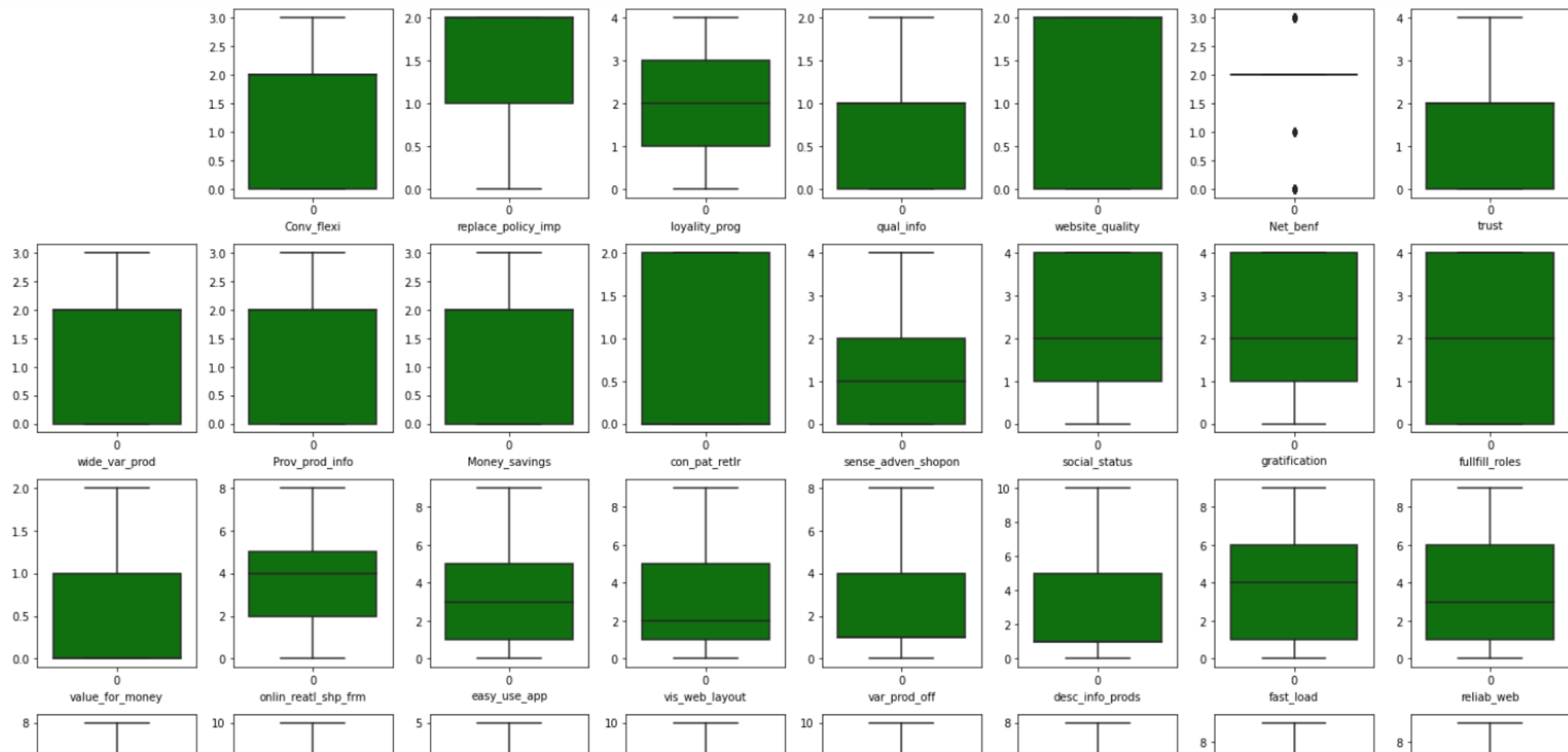


Pair plot for data frame



## Checking outliers





THANK YOU