Data Analysis Applied: E-Commerce Customer Activation and Retention

Introduction

India has an Internet user base of about 696.77million as of May 2020, about 40% of the population. Despite being the second-largest user base in world, only behind China (650 million, 48% of population), the penetration of e-commerce is low compared to markets like the United States (266 million, 84%), or France (54 M, 81%), but is growing, adding around 6 million new entrants every month. The industry consensus is that growth is at an inflection point.

In India, cash on delivery and credit/debit card payment is the most preferred payment method, accumulating 75% of the e-retail activities. Demand for international consumer products (including long-tail items) is growing faster than in-country supply from authorized distributors and e-commerce offerings. Long tail business strategy allows companies to realize significant profits by selling low volumes of hard-to-find items to many customers, instead of only selling large volumes of a reduced number of popular items. The term was first coined in 2004 by Chris Anderson.

In 2017, the largest e-commerce companies in India were Flipkart, Snapdeal and Amazon. In 2018, Amazon beat Flipkart and was recorded the biggest ecommerce in India in terms of revenue.

Problem definition

Still in such a marketing conditions, it can be tremendously difficult to know exactly how to retain a customer that will actually do well enough to generate revenue. There is so much information and so many ecommerce terms to sift through, so creating a plan of action is necessary to prioritize your resources as you develop your own ecommerce business plan.

This guide offers a streamlined and honest approach to subjects like ecommerce marketing, and provides pointers for sourcing products, along with other tips that are essential for those who are looking to launch a successful online store.

We've covered these things in detail, but this report synthesizes all the things you need for retaining and activating a customer. Without further ado, let's jump in.

Context of the project

   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.

Using the dataset, I have attempted to

1. Import the database
2. Clean the data
3. Analyze the data
4. Standardize the data
5. Sketch the needful insights

Insights of the Data.

   From all the Data available, we can bring out some neat insights or conclusions such as

- ✓ *Customer profiling*
- ✓ *E-Commerce profiling*
- ✓ *E-Commerce Customer Success Importance*
- ✓ *Customer Gratification and Satisfaction Importance*
- ✓ *How Customer Satisfaction promote the Brand?*
- ✓ *How a Customer can be a repeated buyer?*

Tools used

1. Python 3.8
2. NumPy
3. Pandas
4. Matplotlib
5. Seaborn
6. Data science
7. Machine learning

Exploratory Data Analysis.

Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations. Post importing the data from a excel file and converting it into a DataFrame I have done the following analysis.
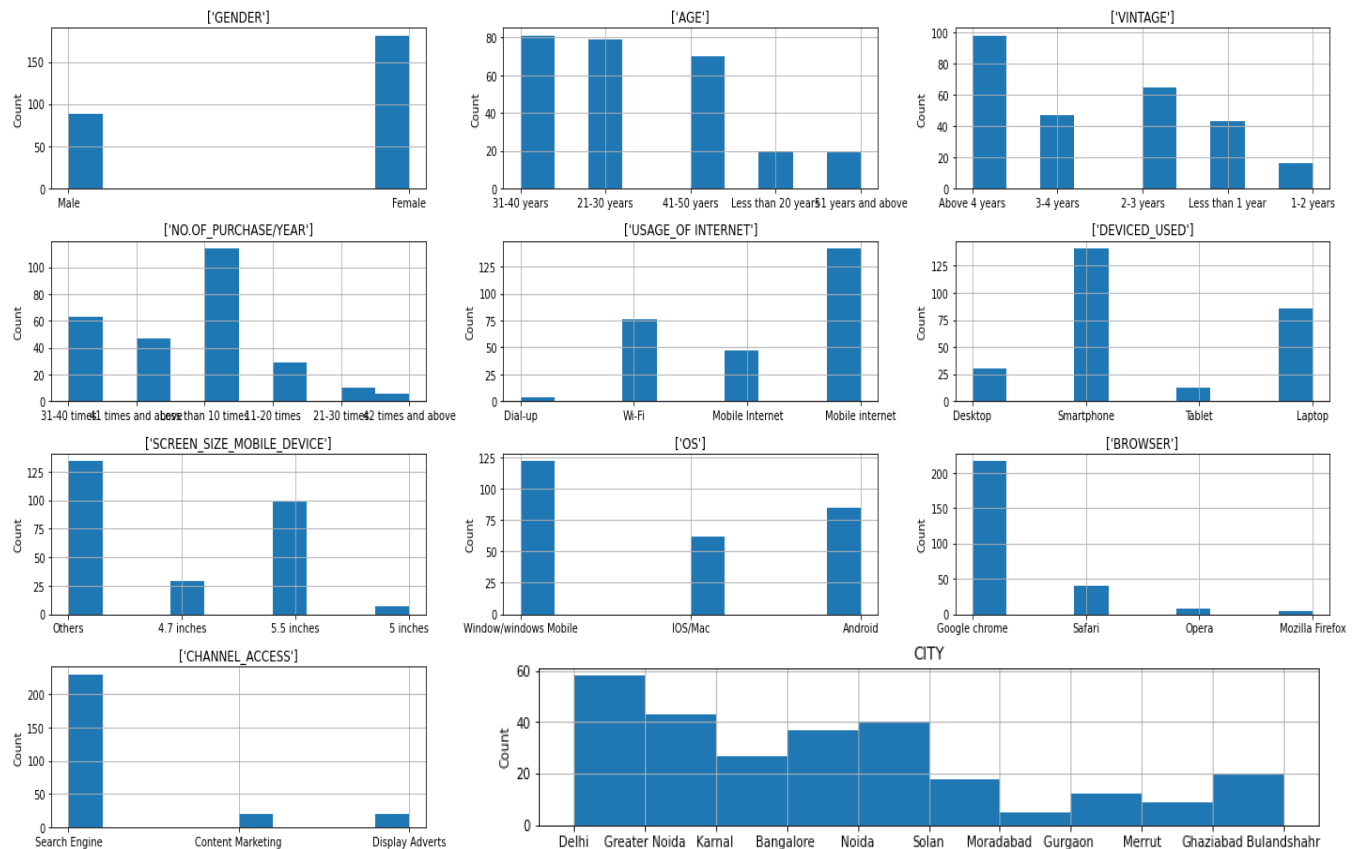
1. Univariate Analysis.
2. Multivariate Analysis.
3. Correlation of the dataset
4. Correlation with the Target variable.

Let's see these Analysis in details as follows.

Univariate Analysis.

I have split the large data into 6 DataFrames to make the analysis easier.

1) Customer and Customer usage details columns.

a. We can observe that the Female users are higher in number comparatively and highest population of the users are between 21-40 years of age and with the vintage above 4 years.
b. Highest No of purchase done/year is above 10.
c. Customers mostly use mobile internet in smartphones and all or any type of screen size.
d. Operating system is 'Windows' mostly on PC and 'Android' on smart phones and customer uses Chrome and search engines.

So, from above we can say that Female users mostly from Delhi aged between 20-40 years with vintage above 4 years of online shopping shop more with the use of android based smart phones.

## 2) E-commerce Merchant Website/Application details columns.
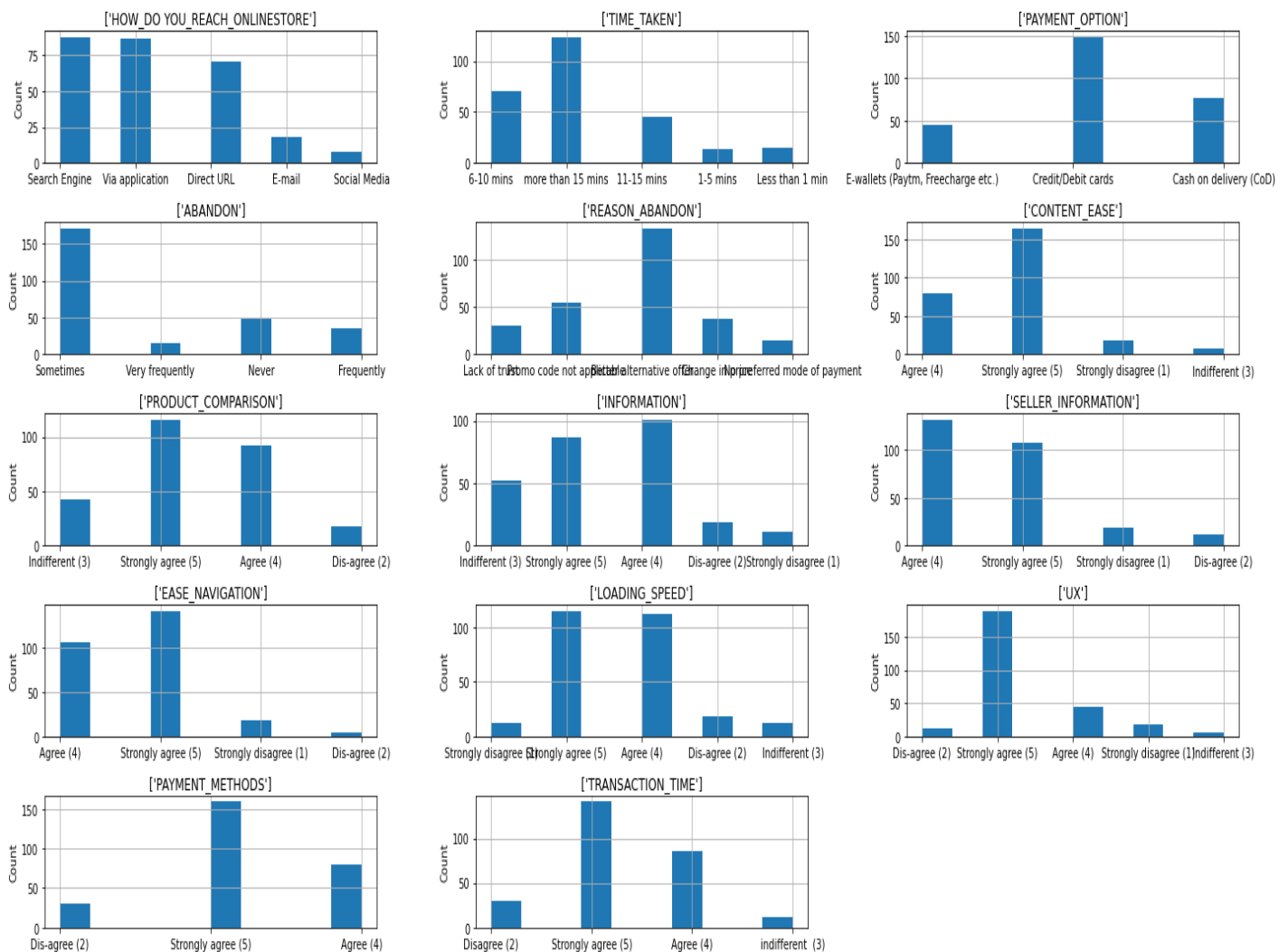
Key Observations.

a. Highest population reaches online by 'Search Engine' like google.com time spend on a purchase is more than 15 mins and people rarely abandon purchasing post entering the online store.

b. Most of the time highest reason for stopping transaction is better alternative offers from other merchants.

c. Payment options are mostly being Credit/Debit cards followed by COD.

d. Content information, Ease of navigation, product and vendor information clarity, site loading speed most of the population are between agree and strongly agree.

e. We also have mostly positive feedback with 'UX', 'Payment_options' and time taken for the transaction.

So, from above we can say people who are actively using online store are happy with the content of the site details and User experience.

## 3) E-commerce Merchant Customer Success details columns.
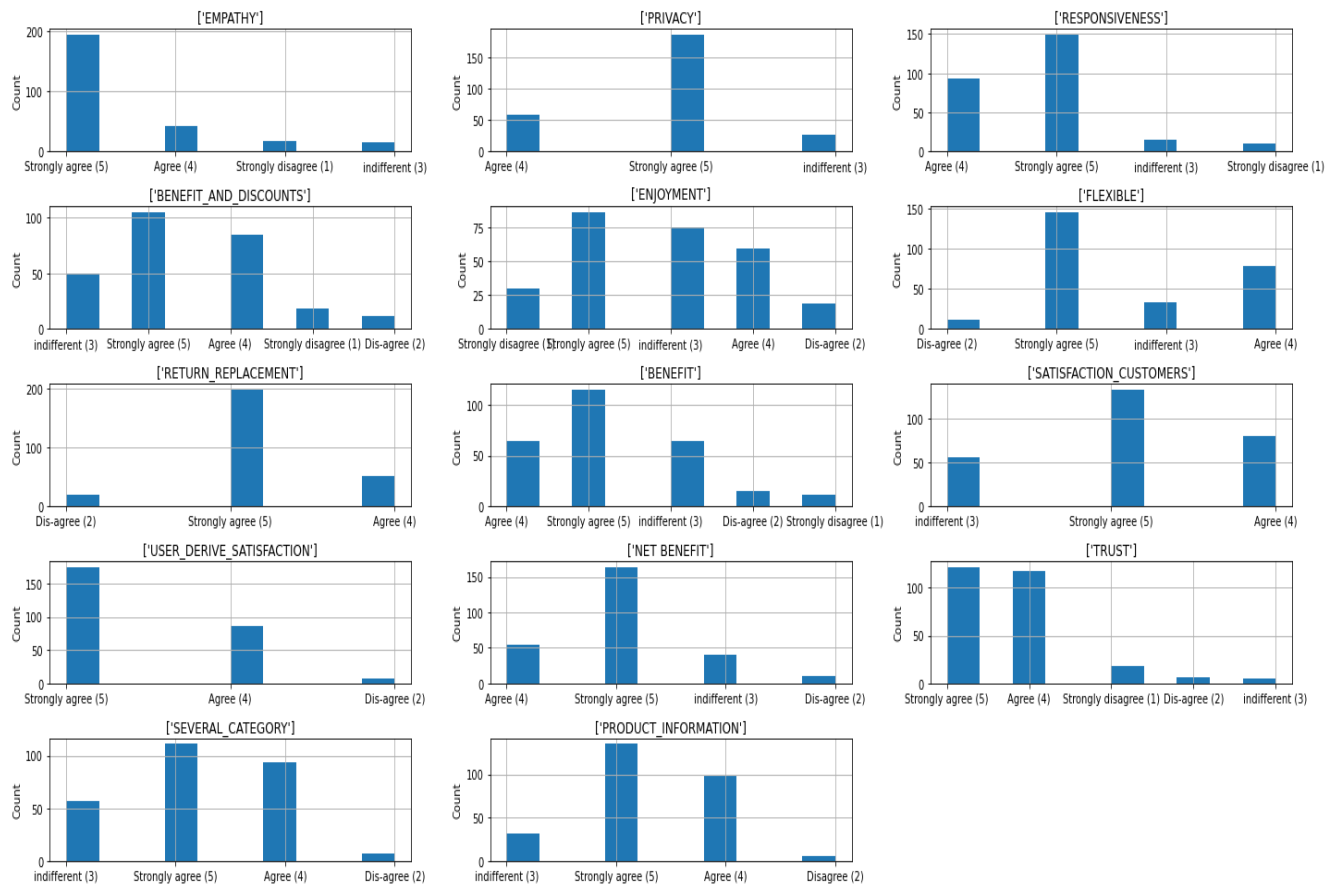
Key Observations.

a. From above we can see the major population of the people have voted for 'Strongly agree' on, Empathy, Privacy, Responsiveness, Benefits and Discounts, enjoyment, Flexibility, Return and replacement Grantee, Benefits, overall Customer satisfaction for services, Trust on the product, Product information in detailed, net benefits, Serval categories.

So, from the observation we can say that majority of the customers are happy with the services that they get from the e-commerce website, but however customer also voted for disagree majorly on benefits and discount and Enjoyment. So, to attract those customers E-Commerce merchant can also put-up persona-based offers.

4) E-commerce Merchant Customer Gratification details columns:
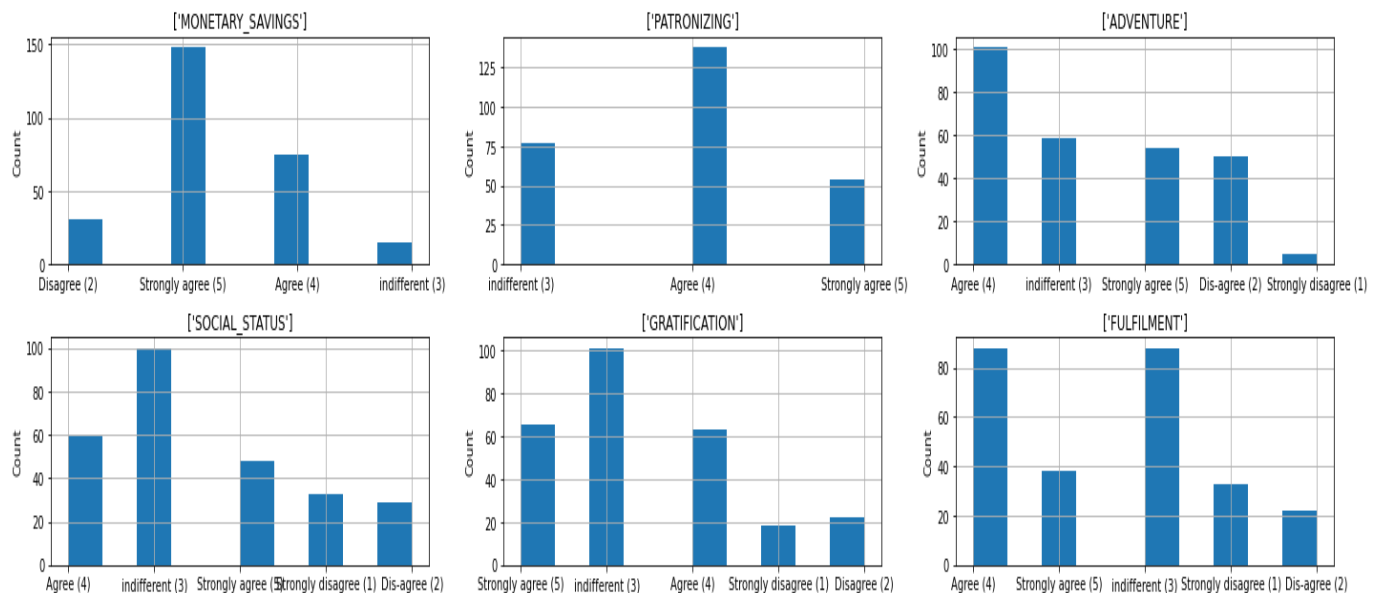
Key Observations.

a. Customer have voted strongly agree and agree for Monetary Savings, adventure and fulfillment.
b. Customer have voted indifferent and agree to social status, patronizing, gratification

This observation says that using the E-commerce website customers doesn't agree that social status, patronizing, gratification is improved but however the Monetary Savings, adventure and fulfillment of the purchase will be a guaranteed.
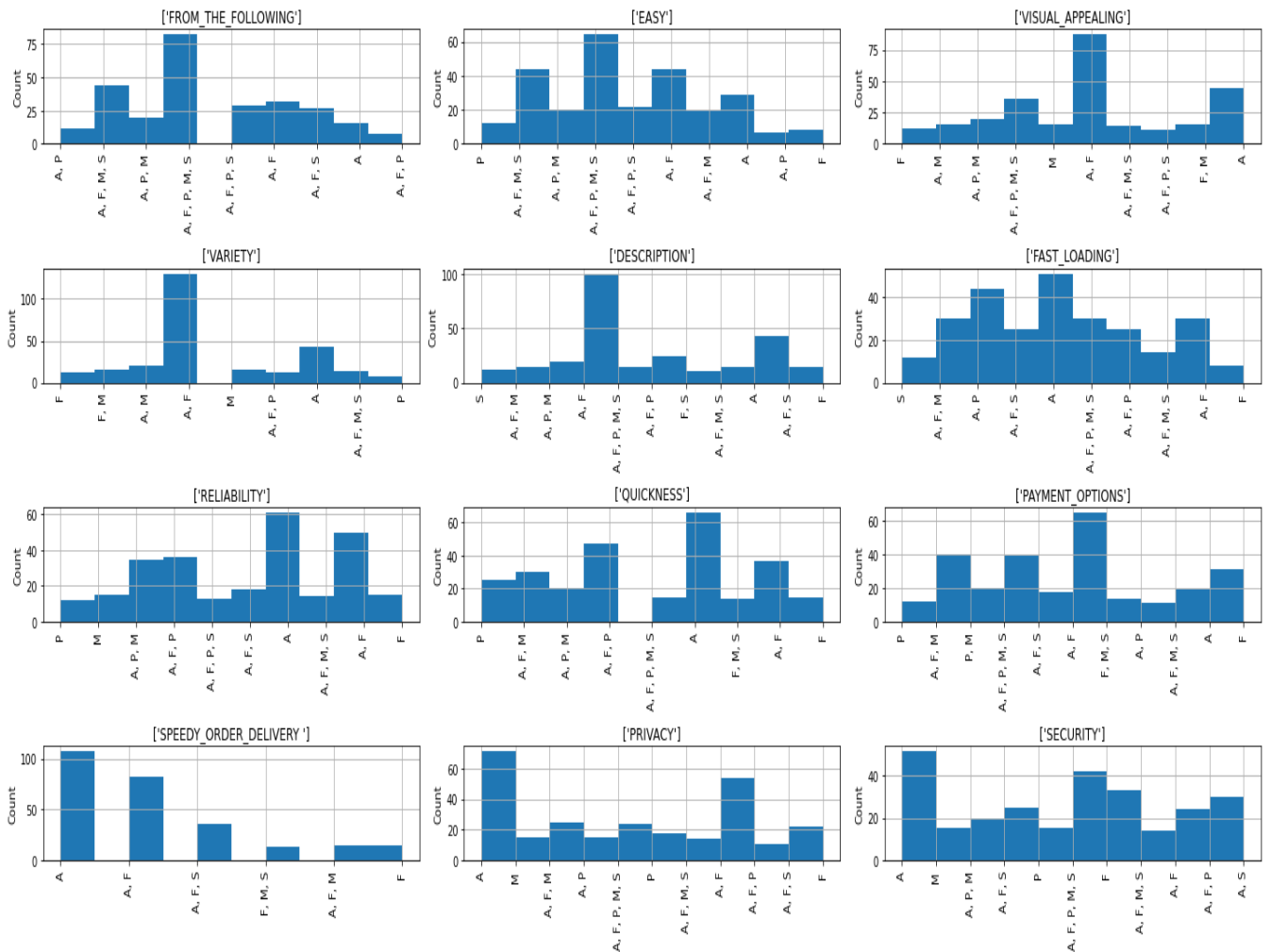
5) Customer Hedonic value column.
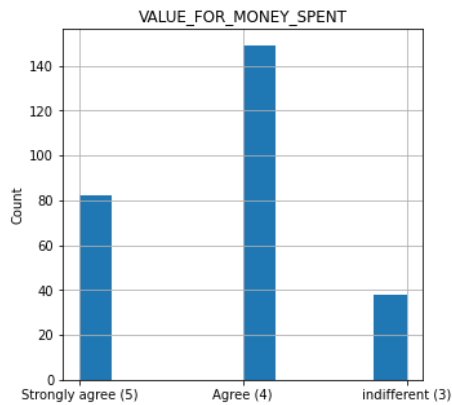
a.   Most of the Customers agrees with the hedonic value that the Ecommerce gives the customer and nearly half of the population declares they strongly agree.

6) E-commerce Customers Survey columns.
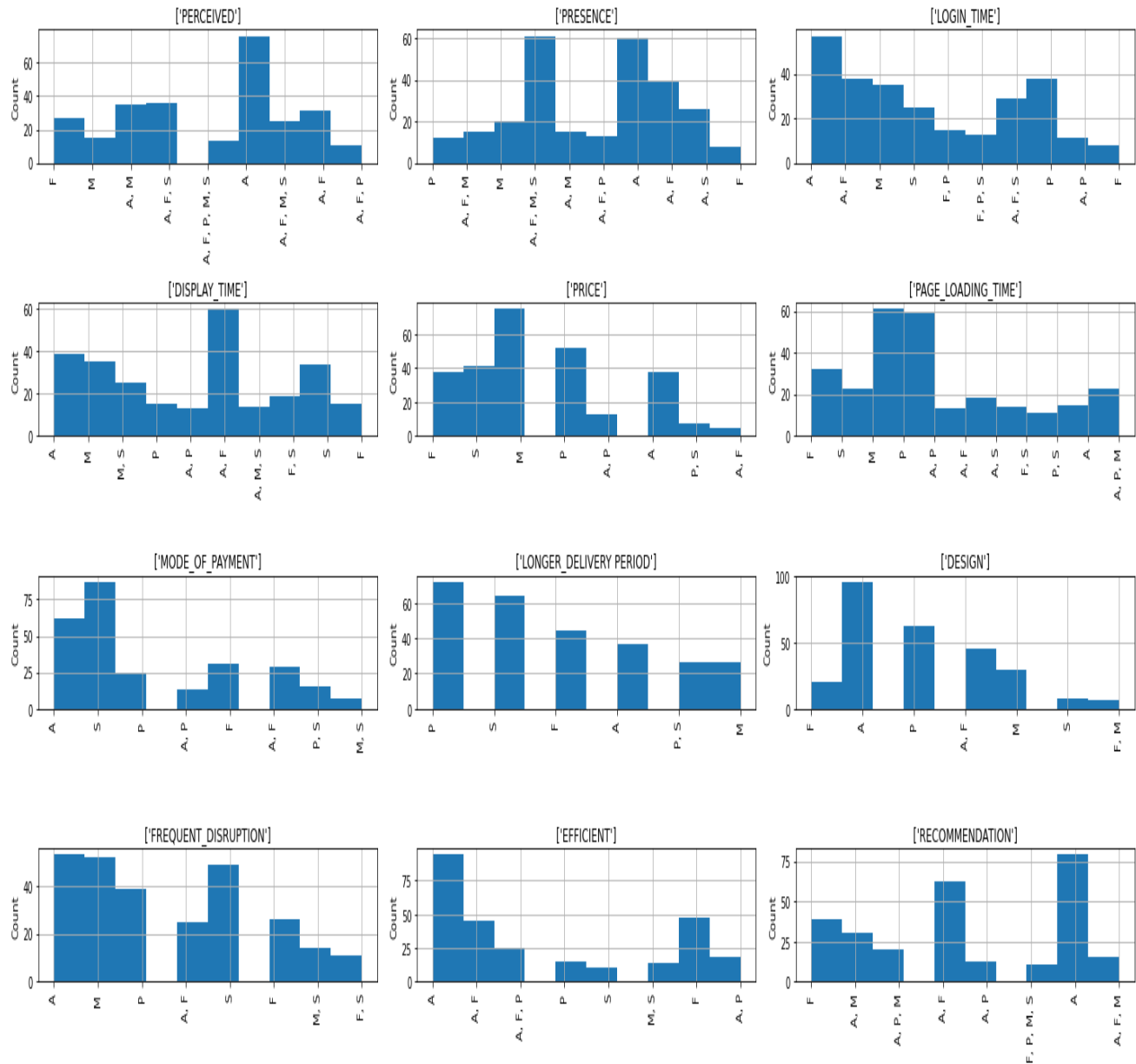
Key observation:

a. The observation says most of the people have voted Amazon at the top followed by Flipkart, Myntra, Snapdeal, Paytm for (EASY, VISUAL_APPEALING, PAYMENT_OPTIONS, SPEEDY_ORDER_DELIVERY, PRIVACY, SECURITY, LOGIN_TIME, DISPLAY_TIME, PRICE PAGE_LOADING_TIME, MODE_OF_PAYMENT, LONGER_DELIVERY, PERIOD, DESIGN, FREQUENT_DISRUPTION, EFFICIENT, RECOMMENDATION)

b. This says People enjoys using Amazon which tops all the category

c. The top 3 E-commerce vendors in India as per customer voting is Amazon, Flipkart, Myntra

d. Snapdeal and Paytm being the worst in delivery period.

Multivariate Analysis.

I have split the large data into 6 DataFrames to make the analysis easier.

1) Customer Details Vs Hedonic Value.

VALUE_FOR_MONEY_SPENT = 3	VALUE_FOR_MONEY_SPENT = 4	VALUE_FOR_MONEY_SPENT = 5
USAGE_OF INTERNET	USAGE_OF INTERNET	USAGE_OF INTERNET
VALUE_FOR_MONEY_SPENT = 3	VALUE_FOR_MONEY_SPENT = 4	VALUE_FOR_MONEY_SPENT = 5
DEVICED_USED	DEVICED_USED	DEVICED_USED
VALUE_FOR_MONEY_SPENT = 3	VALUE_FOR_MONEY_SPENT = 4	VALUE_FOR_MONEY_SPENT = 5
SCREEN_SIZE_MOBILE_DEVICE	SCREEN_SIZE_MOBILE_DEVICE	SCREEN_SIZE_MOBILE_DEVICE
VALUE_FOR_MONEY_SPENT = 3	VALUE_FOR_MONEY_SPENT = 4	VALUE_FOR_MONEY_SPENT = 5
OS	OS	OS
VALUE_FOR_MONEY_SPENT = 3	VALUE_FOR_MONEY_SPENT = 4	VALUE_FOR_MONEY_SPENT = 5
BROWSER	BROWSER	BROWSER
VALUE_FOR_MONEY_SPENT = 3	VALUE_FOR_MONEY_SPENT = 4	VALUE_FOR_MONEY_SPENT = 5
CHANNEL_ACCESS	CHANNEL_ACCESS	CHANNEL_ACCESS

Correlation of Fetures with the target column

Key observation:

a. In Gender we observe 0 - Male 1 - Female, the highest population voted for Agree and strongly Agree.

b. In AGE (20-40), Vintage above 4 years and most of NO. OF_PURCHASE/YEAR (above 10) users have voted for Strongly Agree.

c. SCREEN_SIZE_MOBILE_DEVICE, BROWSER have given negative reviews/others.

d. Usage of Internet, DEVICED_USED, OS, CHANNEL_ACCESS, voted customers have voted randomly for all categories in 'VALUE_FOR_MONEY_SPENT'.

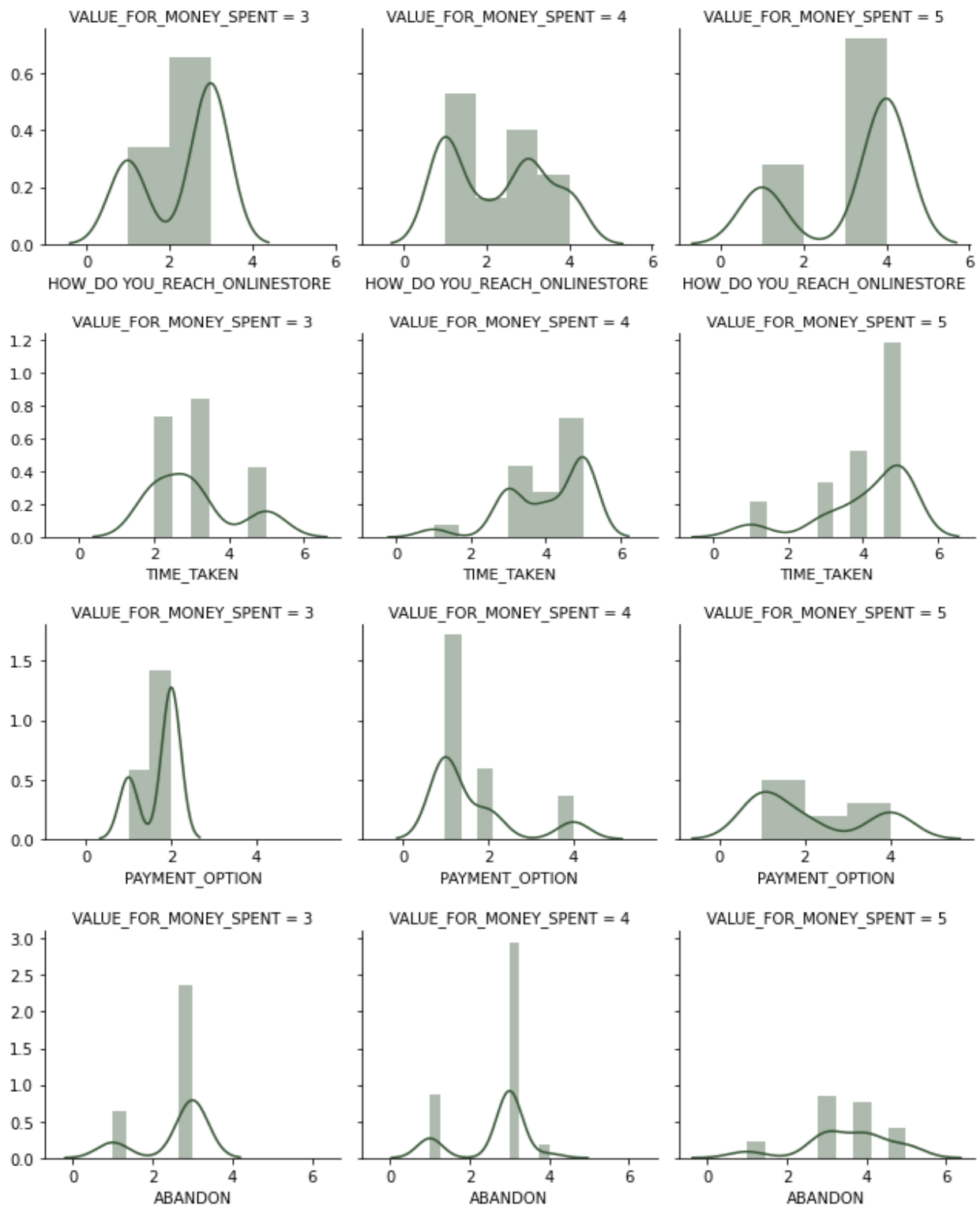e. Snapdeal and Paytm being the worst in delivery period.

From above we understand that people who are using the E-Commerce website more are strongly agreeing to the fact that its value for money. And these customers are not dependent on Usage of Internet, DEVICED_USED, SCREEN_SIZE_MOBILE_DEVICE, OS, BROWSER, CHANNEL_ACCESS.

2) eCommerce Details Vs Hedonic Value.

VALUE_FOR_MONEY_SPENT = 3 VALUE_FOR_MONEY_SPENT = 4 VALUE_FOR_MONEY_SPENT = 5

PAYMENT_METHODS

VALUE_FOR_MONEY_SPENT = 3 VALUE_FOR_MONEY_SPENT = 4 VALUE_FOR_MONEY_SPENT = 5

TRANSACTION_TIME
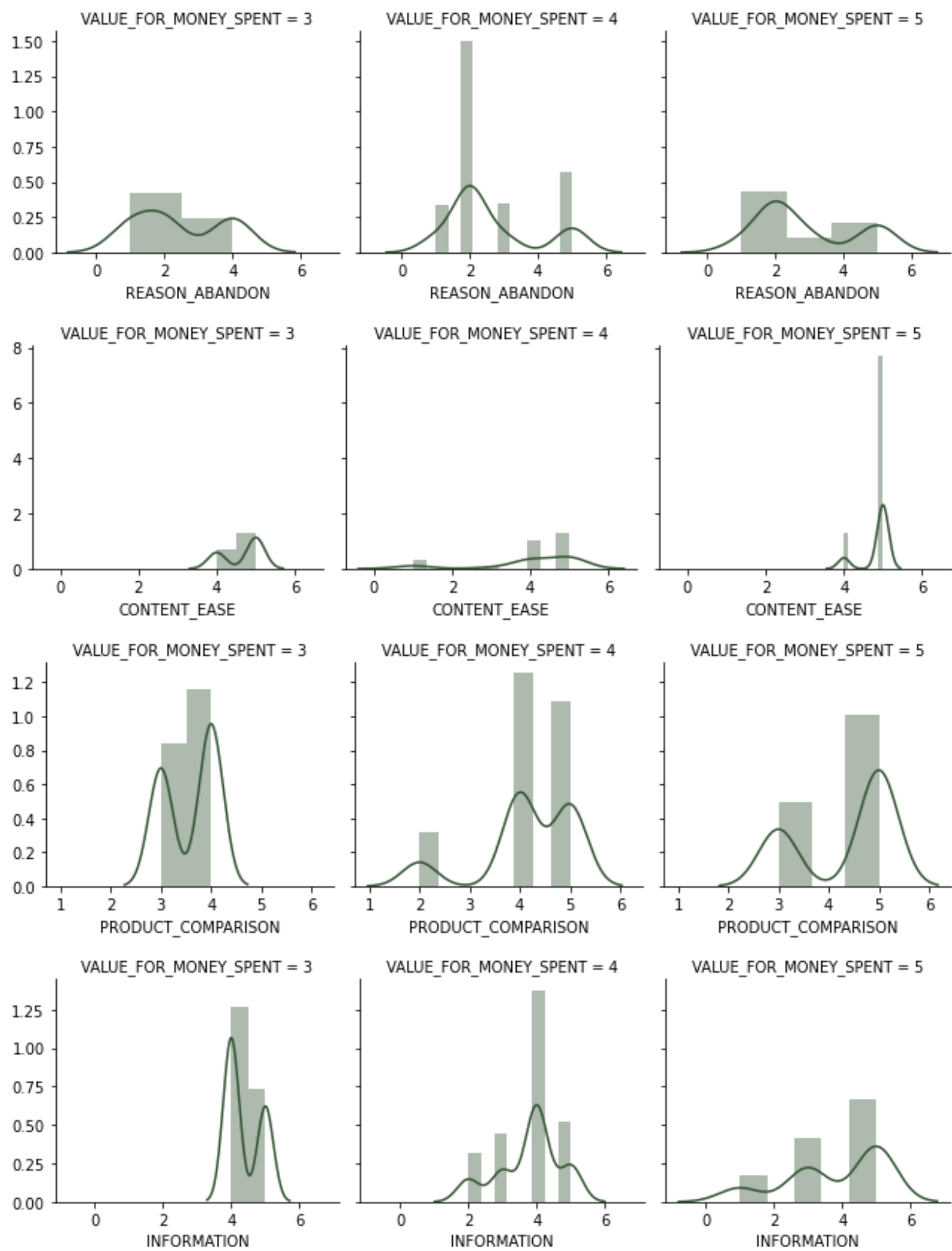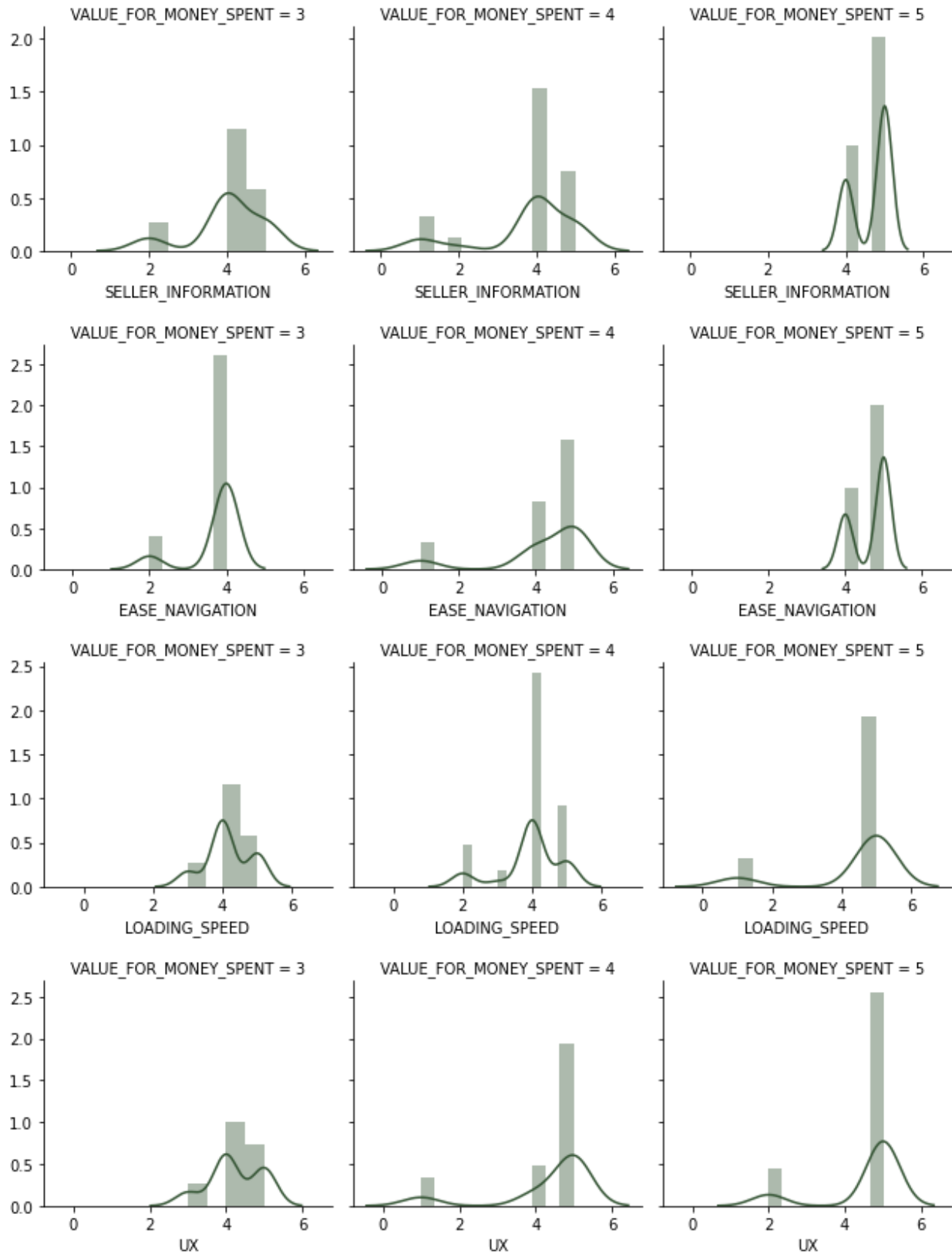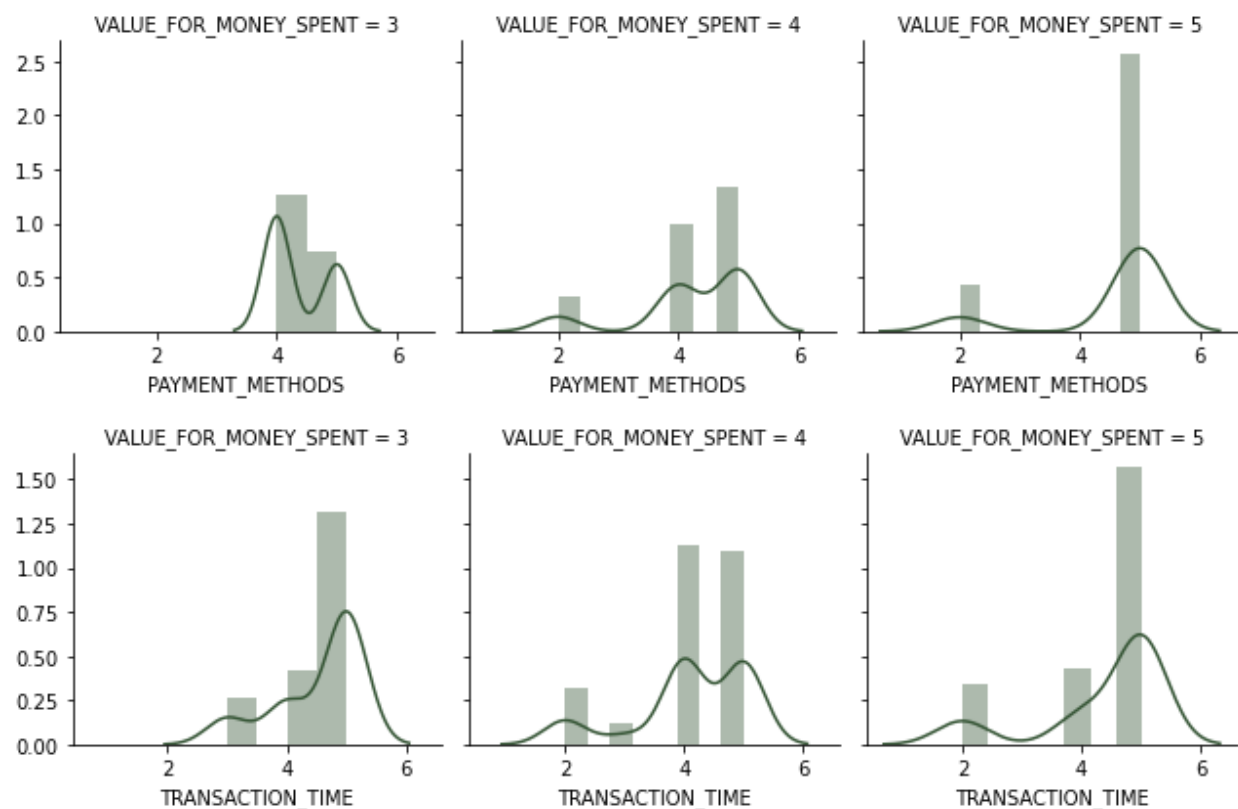
Correlation of Fetures with the target column

a. People who voted for HOW_DO YOU REACH ONLINE, SELLER INFORMATION, PRODUCT COMPARISON, TIME_TAKEN, PAYMENT_OPTIONS, CONTANT EASE, LODING_SPEED, UX, REASON_ABANDON, PAYMENT_METHOD, TRANSACTION TIME voted for almost all the values in values spent for money.

b. ABANDON, EASE OF NAVIGATION voted customers have voted mostly Strongly Agree for Value for Money spent.

To summarize, people are not interested in facilities or the information provided by the E-commerce merchant but those whose are interested in knowing about good deals and abandoning one-site and goes to another gives feedback as value for money.

3) Customer Success Vs Hedonic Value.

VALUE_FOR_MONEY_SPENT = 3   VALUE_FOR_MONEY_SPENT = 4   VALUE_FOR_MONEY_SPENT = 5

TRUST

VALUE_FOR_MONEY_SPENT = 3   VALUE_FOR_MONEY_SPENT = 4   VALUE_FOR_MONEY_SPENT = 5

SEVERAL_CATEGORY

VALUE_FOR_MONEY_SPENT = 3   VALUE_FOR_MONEY_SPENT = 4   VALUE_FOR_MONEY_SPENT = 5
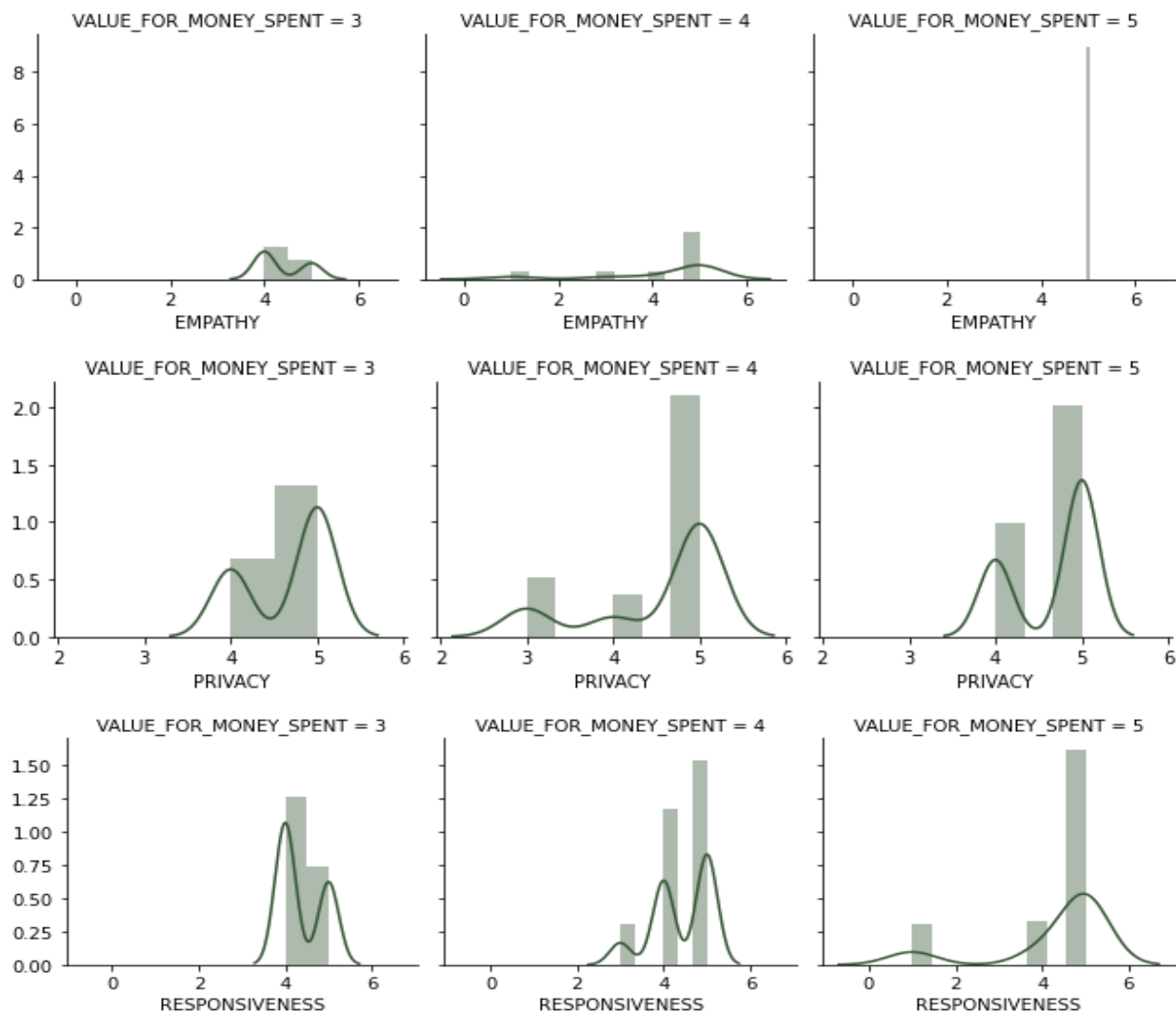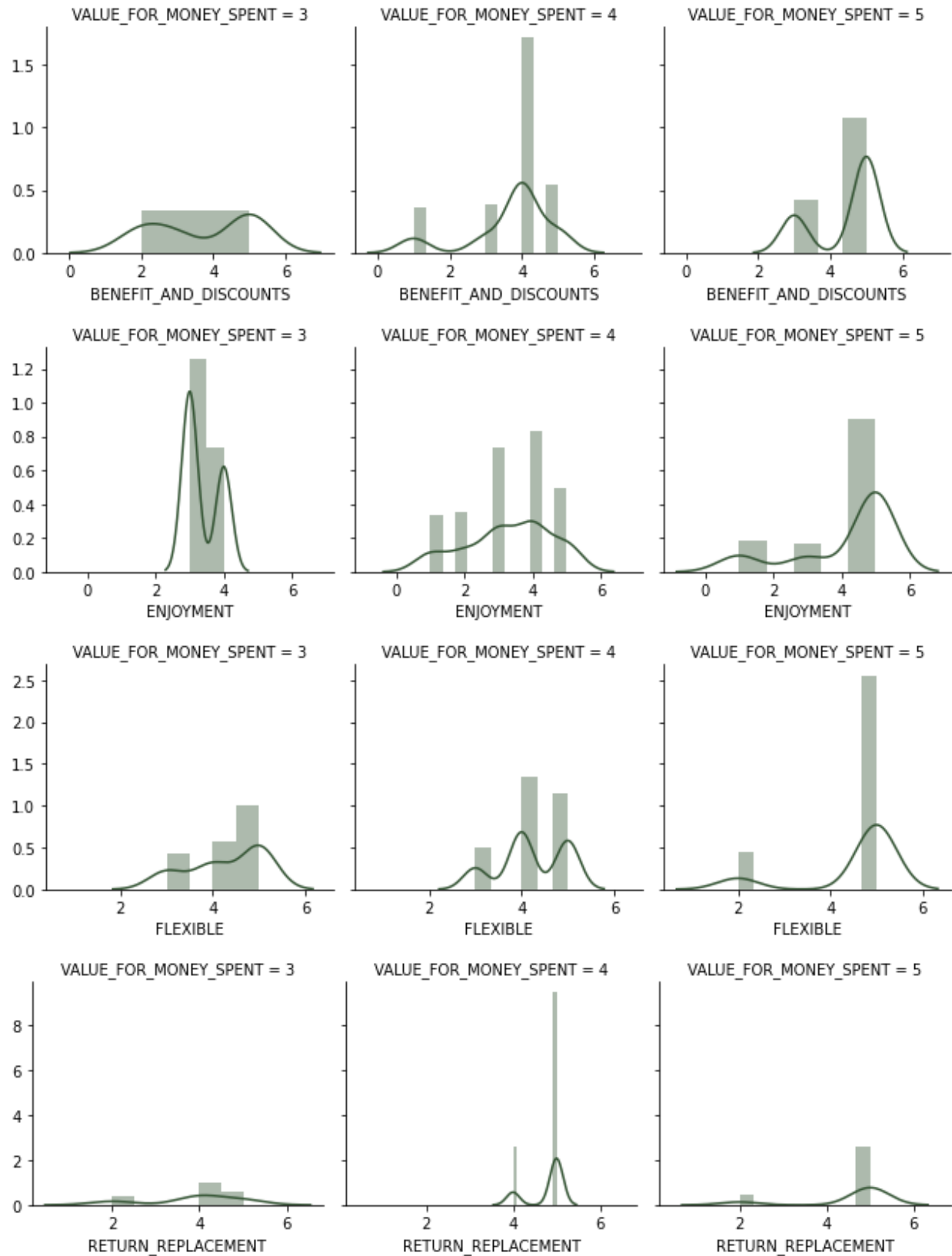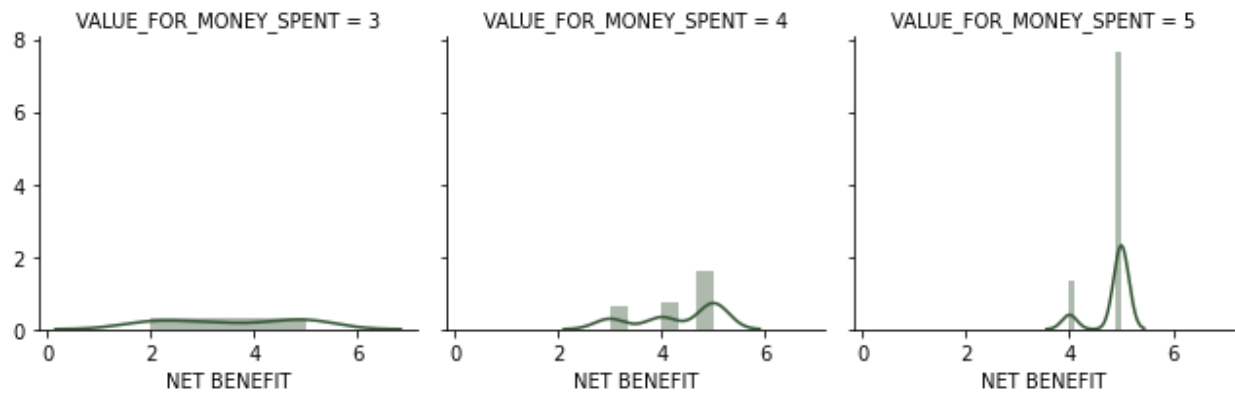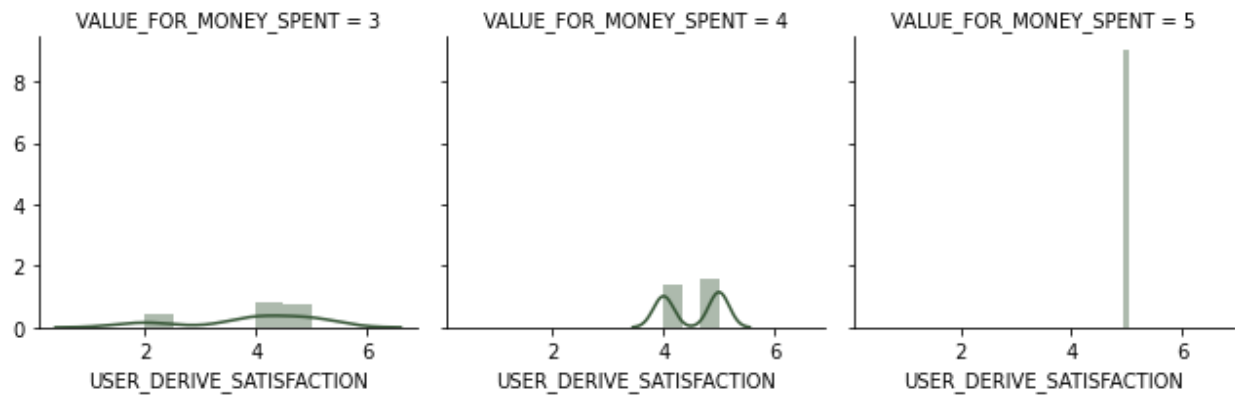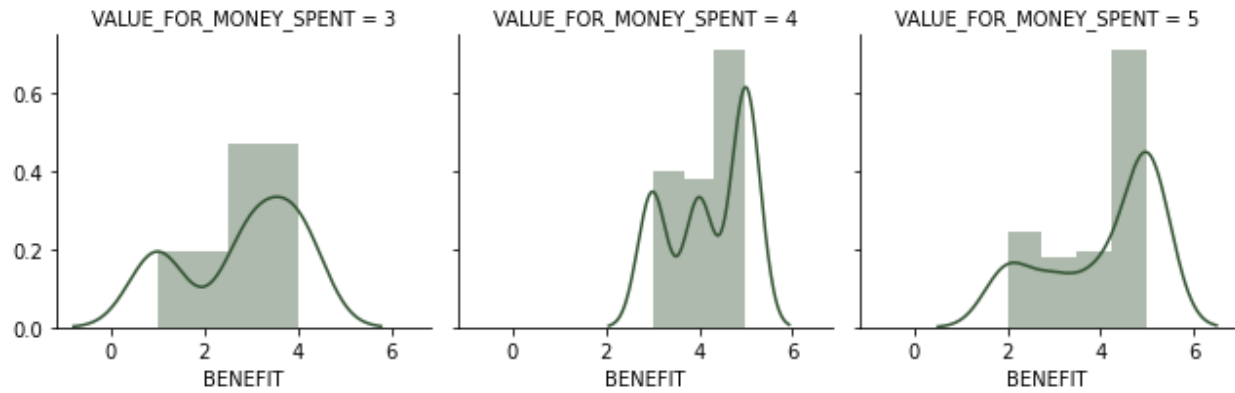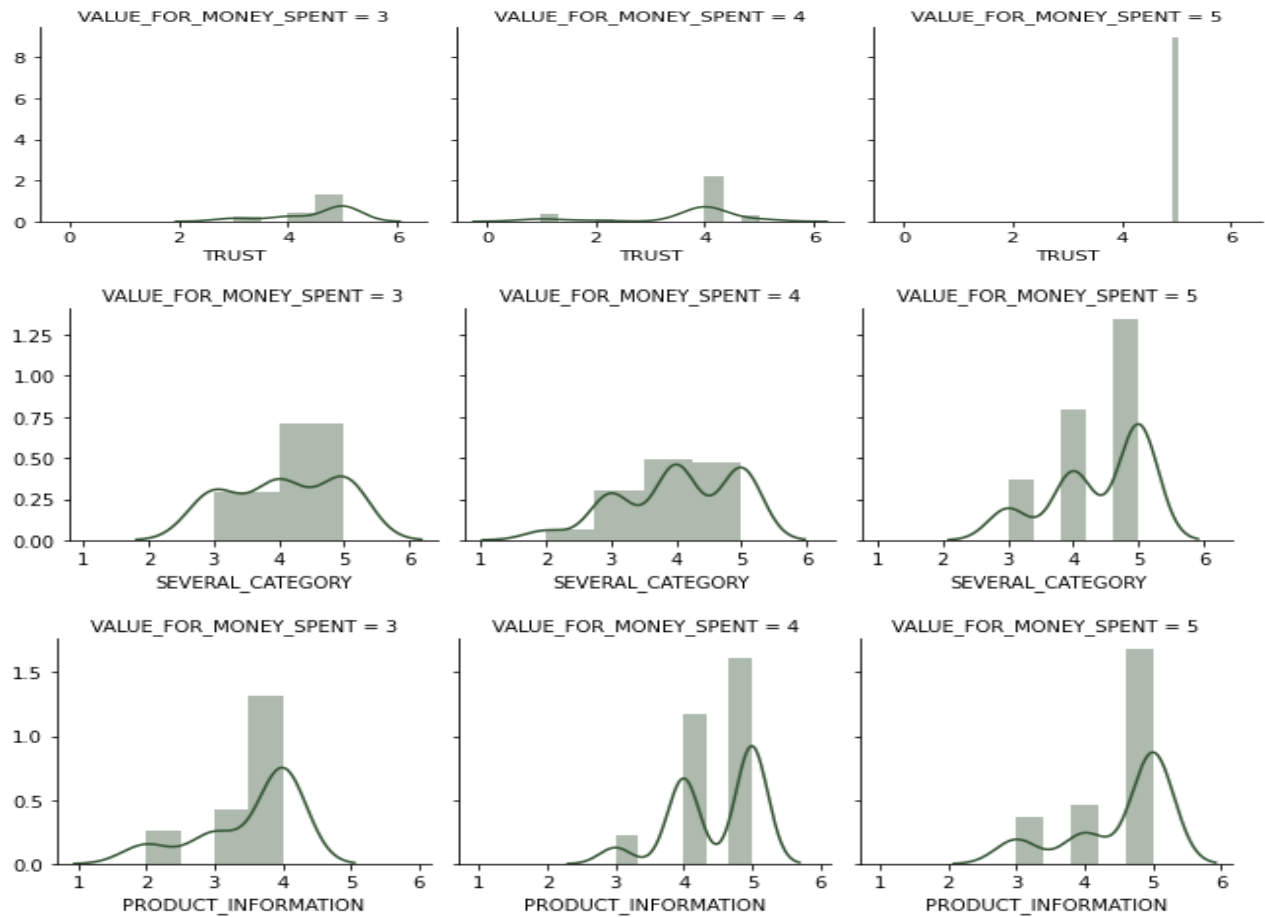
PRODUCT_INFORMATION

Correlation of Fetures with the target column

a. USER_DERIVE_SATISFACTION, NET BENEFIT, PRODUCT_INFORMATION, TRUST, BENEFIT, EMPATHY, BENEFIT_AND_DISCOUNTS, ENJOYMENT voted customers have voted mostly agree for value for money spent.

From above we can understand customer will interested in E-Commerce when the customer is valued and pleased with the product, customer service and product purchased.
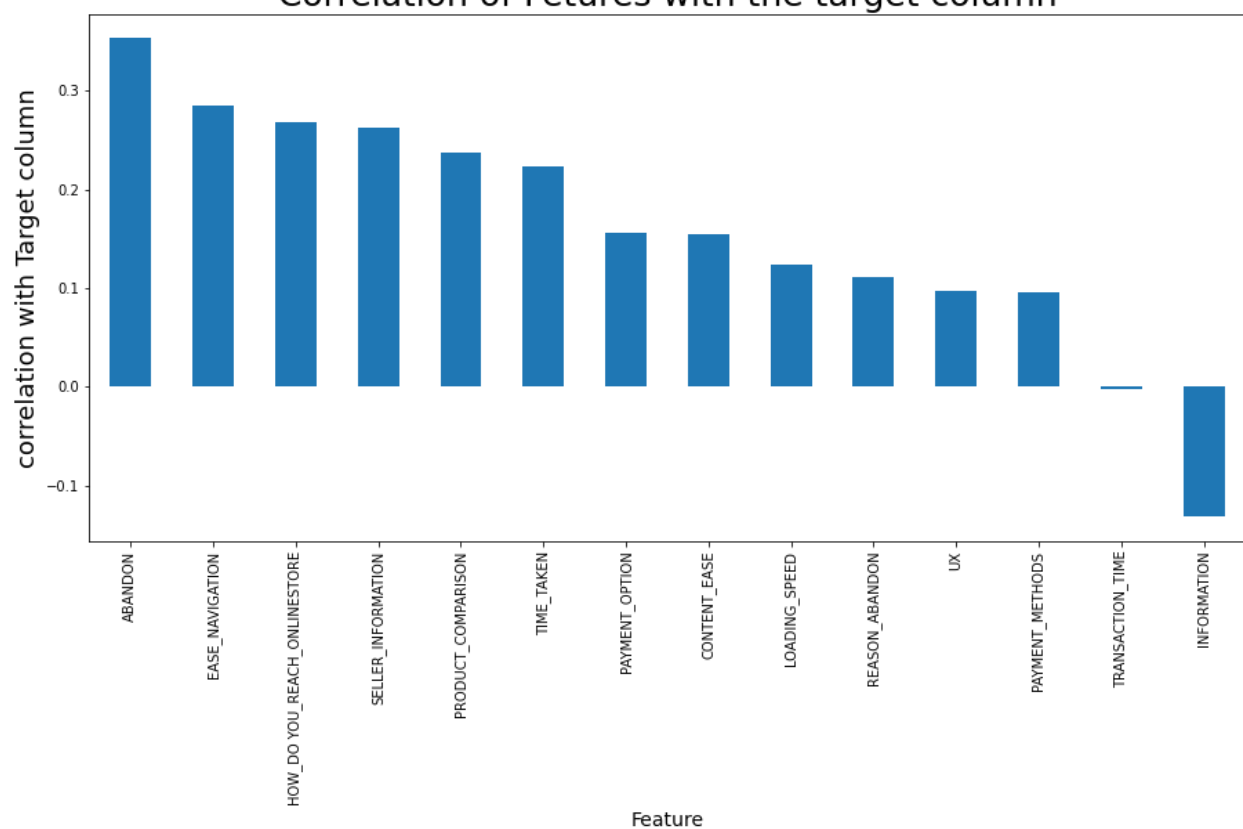
4) Customer Gratification Vs Hedonic Value.
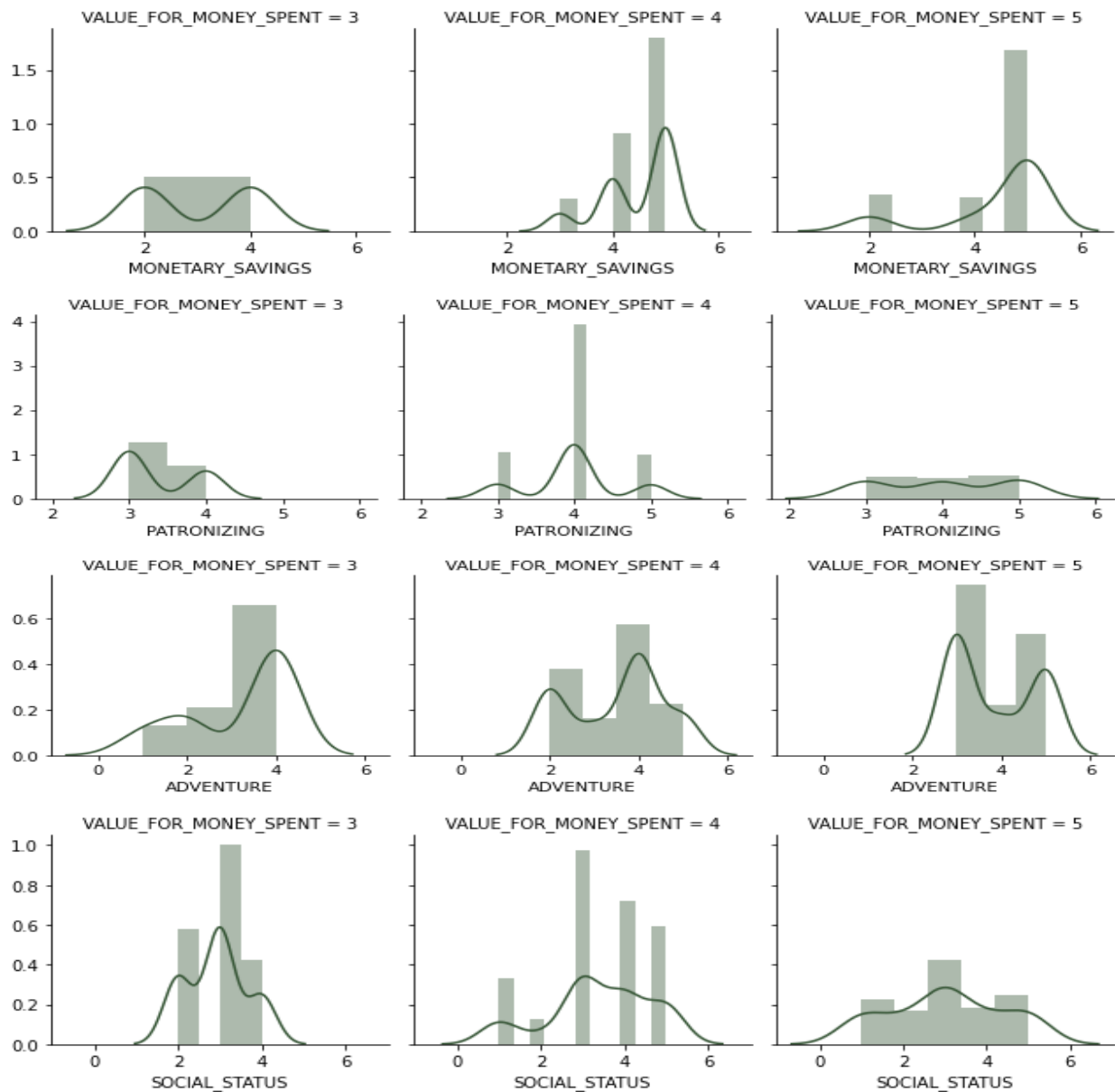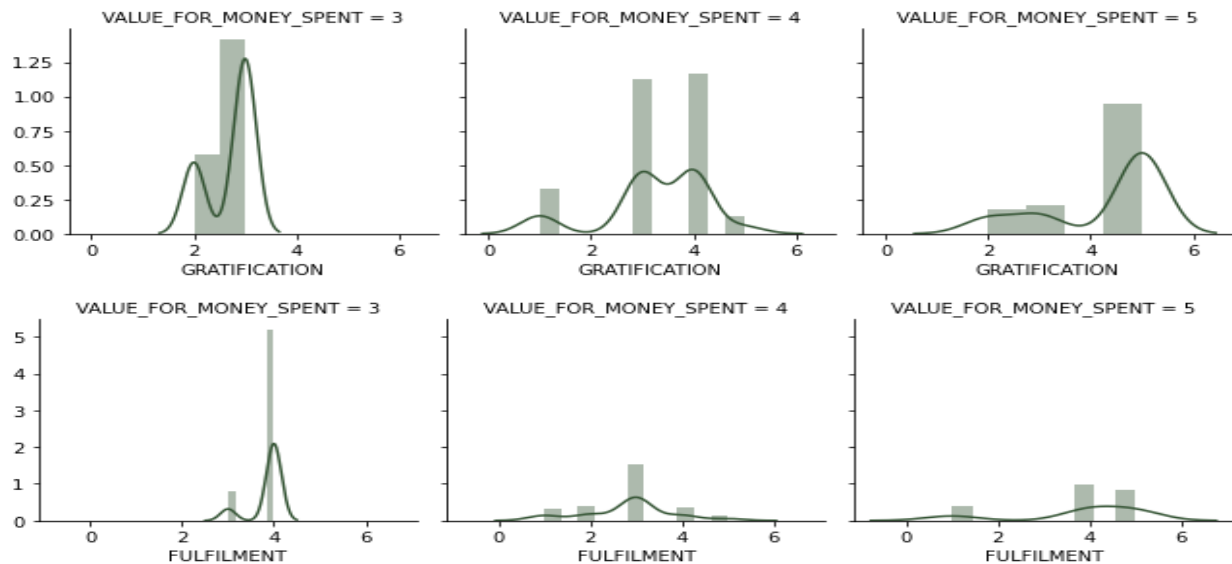
Correlation of Fetures with the target column
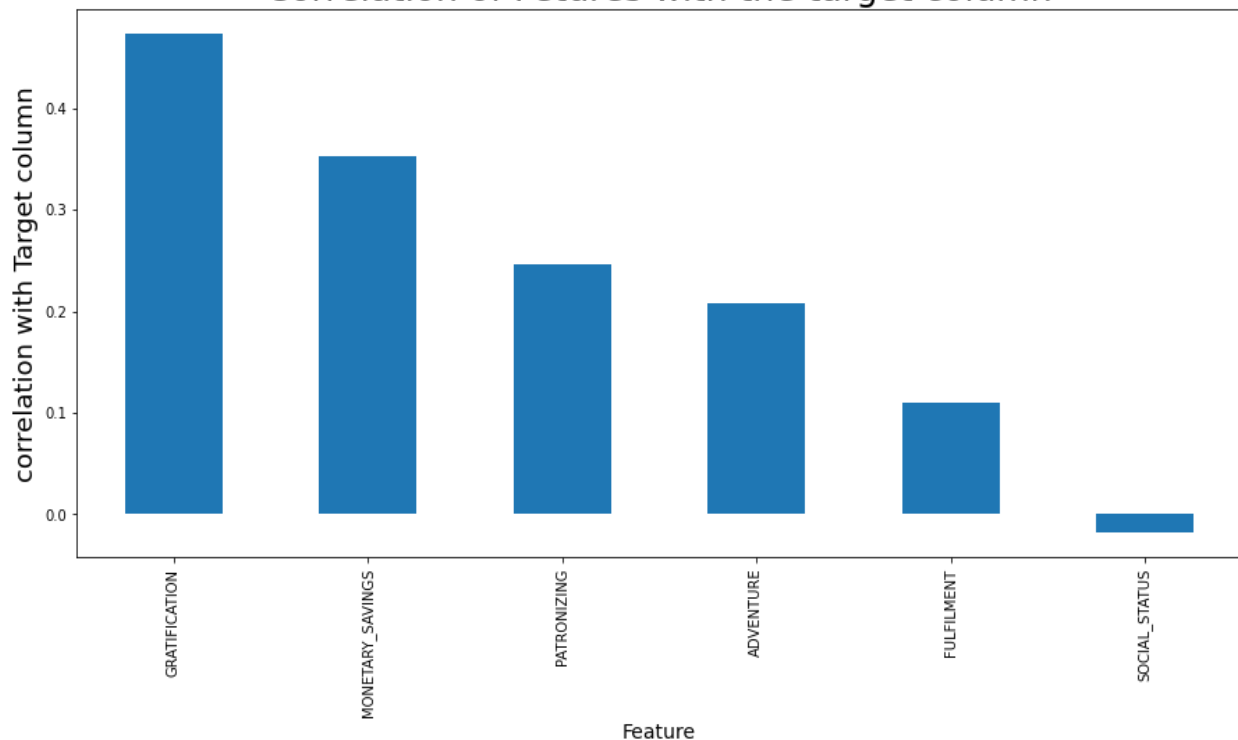
Key observation:

    a.   GRATIFICATION, MONETARY_SAVINGS customers have voted for strongly agree in VALUE_FOR_MONEY_SPENT

   b.   In PATRONIZING, ADVENTURE, FULFILMENT, SOCIAL_STATUS customer voted for all the Fields more are less equally

The above observation says that the customer with good gratification and monetary savings tends to give feedback as value for money spend.

Amazon have most of the votes in all the parameters followed by Flipkart and Myntra.

Correlation of Dataset.

We will observe the correlation of the data as per data in the data frame split. And Explanation at the last.

Correlation matrix 1

	MONETARY_SAVINGS	PATRONIZING	ADVENTURE	SOCIAL_STATUS	GRATIFICATION	FULFILMENT
MONETARY_SAVINGS	1	0.4	0.23	0.043	0.13	-0.21
PATRONIZING	0.4	1	0.68	-0.026	-0.19	-0.11
ADVENTURE	0.23	0.68	1	0.088	-0.13	0.12
SOCIAL_STATUS	0.043	-0.026	0.088	1	0.51	0.55
GRATIFICATION	0.13	-0.19	-0.13	0.51	1	0.41
FULFILMENT	-0.21	-0.11	0.12	0.55	0.41	1

Correlation matrix 2

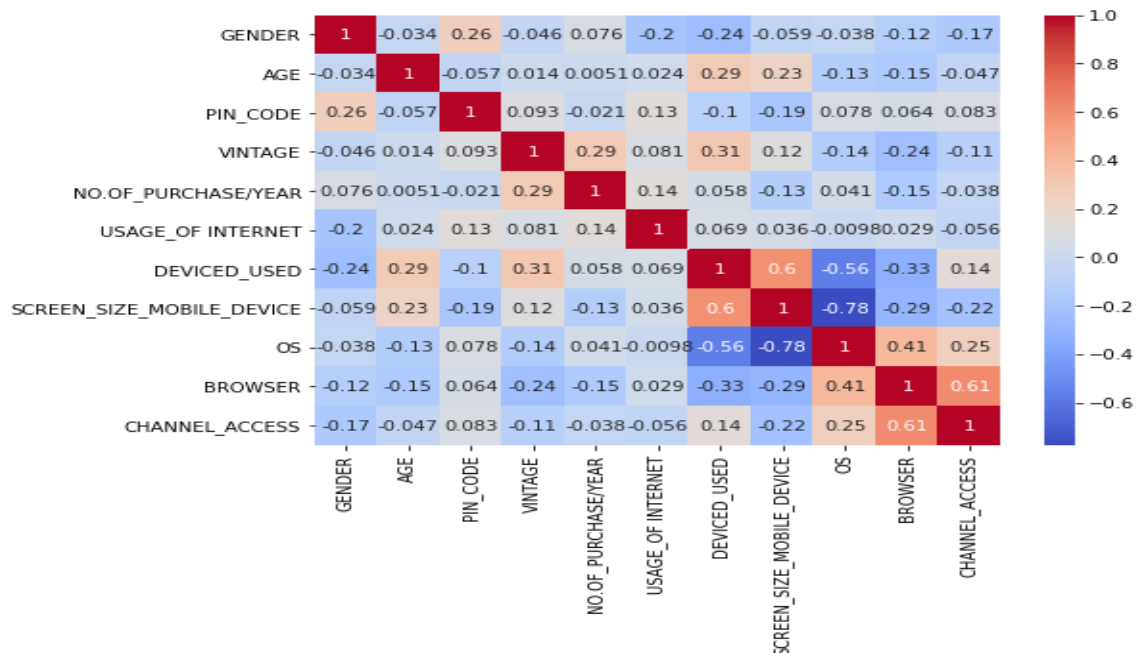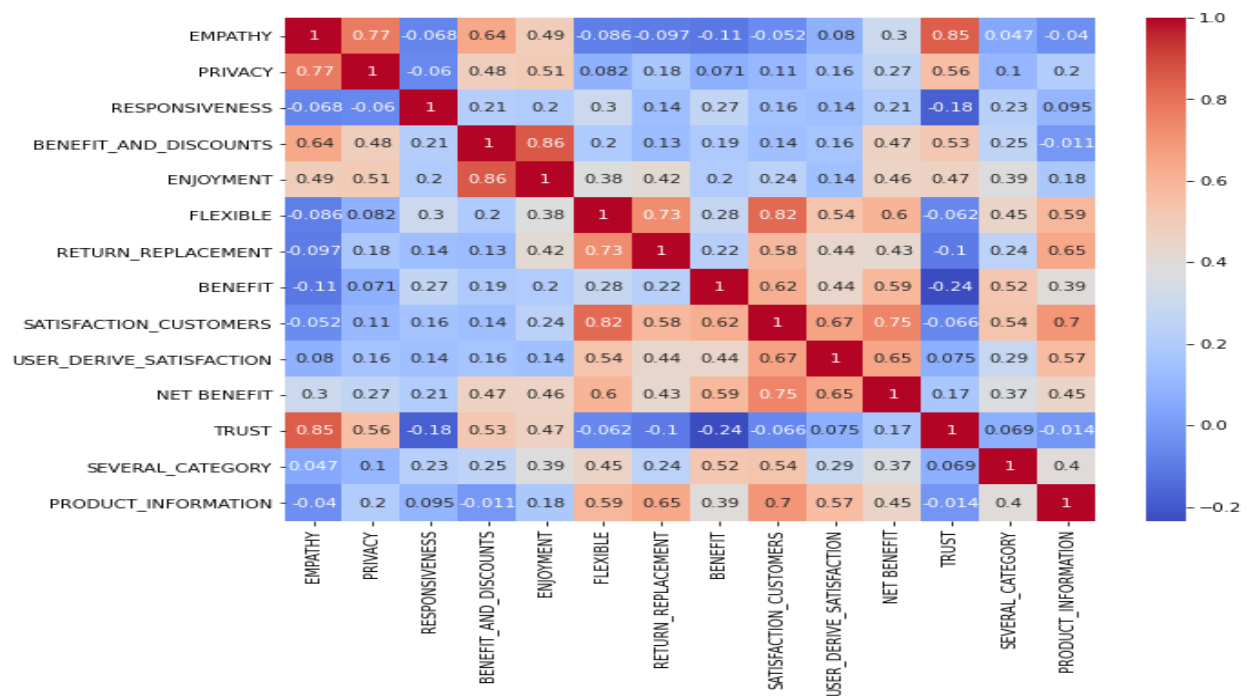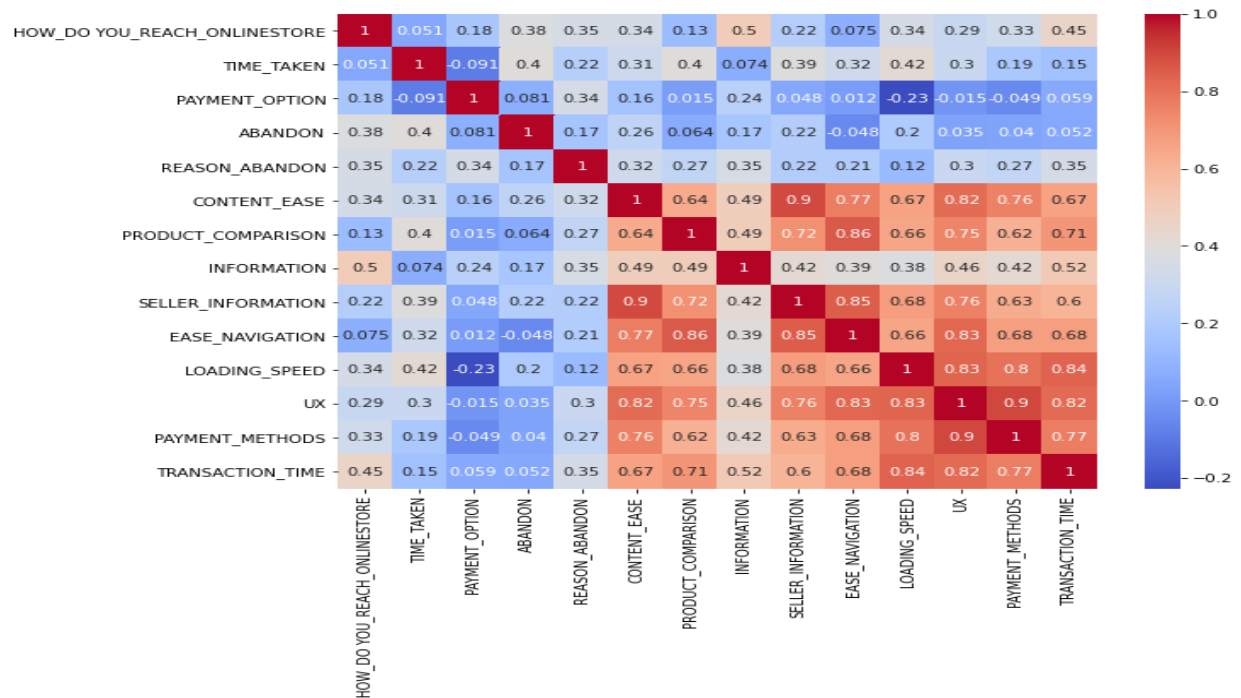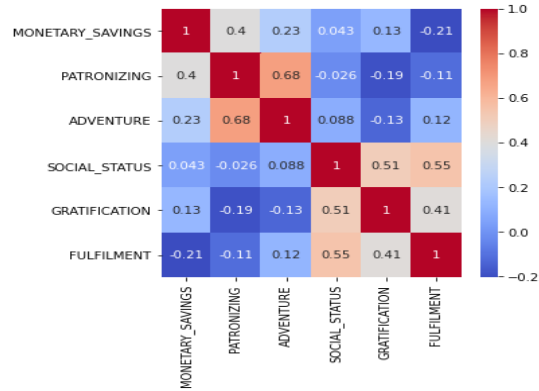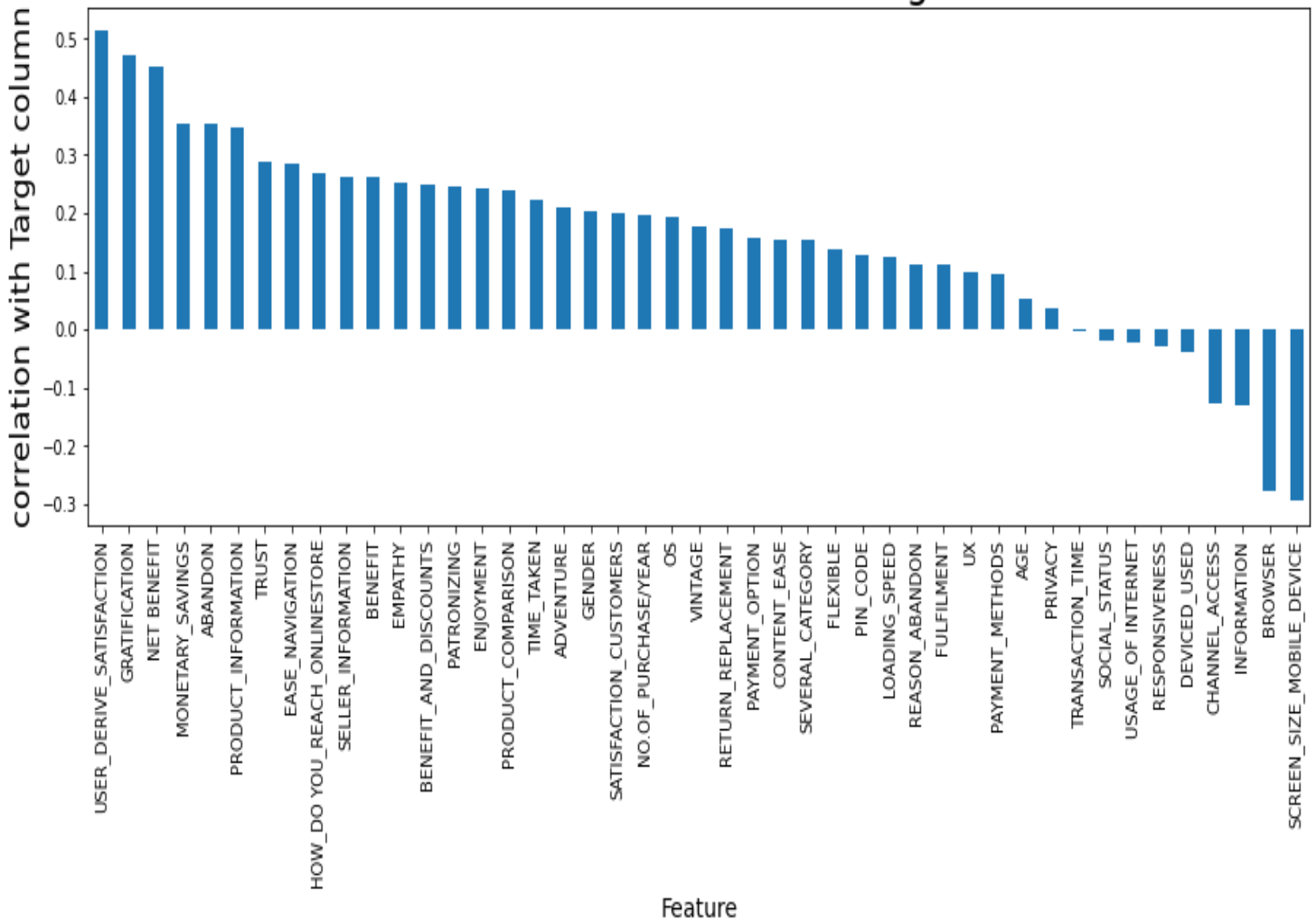|	HOW_DO YOU_REACH_ONLINESTORE	TIME_TAKEN	PAYMENT_OPTION	ABANDON	REASON_ABANDON	CONTENT_EASE	PRODUCT_COMPARISON	INFORMATION	SELLER_INFORMATION	EASE_NAVIGATION	LOADING_SPEED	UX	PAYMENT_METHODS	TRANSACTION_TIME
HOW_DO YOU_REACH_ONLINESTORE	1	0.051	0.18	0.38	0.35	0.34	0.13	0.5	0.22	0.075	0.34	0.29	0.33	0.45
TIME_TAKEN	0.051	1	-0.091	0.4	0.22	0.31	0.4	0.074	0.39	0.32	0.42	0.3	0.19	0.15
PAYMENT_OPTION	0.18	-0.091	1	0.081	0.34	0.16	0.015	0.24	0.048	0.012	-0.23	-0.015	-0.049	0.059
ABANDON	0.38	0.4	0.081	1	0.17	0.26	0.064	0.17	0.22	-0.048	0.2	0.035	0.04	0.052
REASON_ABANDON	0.35	0.22	0.34	0.17	1	0.32	0.27	0.35	0.22	0.21	0.12	0.3	0.27	0.35
CONTENT_EASE	0.34	0.31	0.16	0.26	0.32	1	0.64	0.49	0.9	0.77	0.67	0.82	0.76	0.67
PRODUCT_COMPARISON	0.13	0.4	0.015	0.064	0.27	0.64	1	0.49	0.72	0.86	0.66	0.75	0.62	0.71
INFORMATION	0.5	0.074	0.24	0.17	0.35	0.49	0.49	1	0.42	0.39	0.38	0.46	0.42	0.52
SELLER_INFORMATION	0.22	0.39	0.048	0.22	0.22	0.9	0.72	0.42	1	0.85	0.68	0.76	0.63	0.6
EASE_NAVIGATION	0.075	0.32	0.012	-0.048	0.21	0.77	0.86	0.39	0.85	1	0.66	0.83	0.68	0.68
LOADING_SPEED	0.34	0.42	-0.23	0.2	0.12	0.67	0.66	0.38	0.68	0.66	1	0.83	0.8	0.84
UX	0.29	0.3	-0.015	0.035	0.3	0.82	0.75	0.46	0.76	0.83	0.83	1	0.9	0.82
PAYMENT_METHODS	0.33	0.19	-0.049	0.04	0.27	0.76	0.62	0.42	0.63	0.68	0.8	0.9	1	0.77
TRANSACTION_TIME	0.45	0.15	0.059	0.052	0.35	0.67	0.71	0.52	0.6	0.68	0.84	0.82	0.77	1

Correlation matrix 3

	EMPATHY	PRIVACY	RESPONSIVENESS	BENEFIT_AND_DISCOUNTS	ENJOYMENT	FLEXIBLE	RETURN_REPLACEMENT	BENEFIT	SATISFACTION_CUSTOMERS	USER_DERIVE_SATISFACTION	NET BENEFIT	TRUST	SEVERAL_CATEGORY	PRODUCT_INFORMATION
EMPATHY	1	0.77	-0.068	0.64	0.49	-0.086	-0.097	-0.11	-0.052	0.08	0.3	0.85	0.047	-0.04
PRIVACY	0.77	1	-0.06	0.48	0.51	0.082	0.18	0.071	0.11	0.16	0.27	0.56	0.1	0.2
RESPONSIVENESS	-0.068	-0.06	1	0.21	0.2	0.3	0.14	0.27	0.16	0.14	0.21	-0.18	0.23	0.095
BENEFIT_AND_DISCOUNTS	0.64	0.48	0.21	1	0.86	0.2	0.13	0.19	0.14	0.16	0.47	0.53	0.25	-0.011
ENJOYMENT	0.49	0.51	0.2	0.86	1	0.38	0.42	0.2	0.24	0.14	0.46	0.47	0.39	0.18
FLEXIBLE	-0.086	0.082	0.3	0.2	0.38	1	0.73	0.28	0.82	0.54	0.6	-0.062	0.45	0.59
RETURN_REPLACEMENT	-0.097	0.18	0.14	0.13	0.42	0.73	1	0.22	0.58	0.44	0.43	-0.1	0.24	0.65
BENEFIT	-0.11	0.071	0.27	0.19	0.2	0.28	0.22	1	0.62	0.44	0.59	-0.24	0.52	0.39
SATISFACTION_CUSTOMERS	-0.052	0.11	0.16	0.14	0.24	0.82	0.58	0.62	1	0.67	0.75	-0.066	0.54	0.7
USER_DERIVE_SATISFACTION	0.08	0.16	0.14	0.16	0.14	0.54	0.44	0.44	0.67	1	0.65	0.075	0.29	0.57
NET BENEFIT	0.3	0.27	0.21	0.47	0.46	0.6	0.43	0.59	0.75	0.65	1	0.17	0.37	0.45
TRUST	0.85	0.56	-0.18	0.53	0.47	-0.062	-0.1	-0.24	-0.066	0.075	0.17	1	0.069	-0.014
SEVERAL_CATEGORY	0.047	0.1	0.23	0.25	0.39	0.45	0.24	0.52	0.54	0.29	0.37	0.069	1	0.4
PRODUCT_INFORMATION	-0.04	0.2	0.095	-0.011	0.18	0.59	0.65	0.39	0.7	0.57	0.45	-0.014	0.4	1

Correlation of Fetures with the target column

Key observation:

a. From all the above observations we can derive the following points.

b. DEVISE_USED, OS, SCREEN_SIZE, WEB BROWSER, usage will directly be related to each other which will also affect the usage of internet or e commerce application or website.

c. CONTENT_EASE, PRODUCT_COMPARISON, INFORMATION, SELLER_INFORMATION, EASE_NAVIGATION, LODING_SPEED, UX, TRANSACTION_TIME are more correlated to each other which will also affect the usage of internet or e commerce application or website.

d. Customer satisfaction and customer care with customer success directly affect the usage. Happy customers tend to purchase again.

Correlation of Dataset
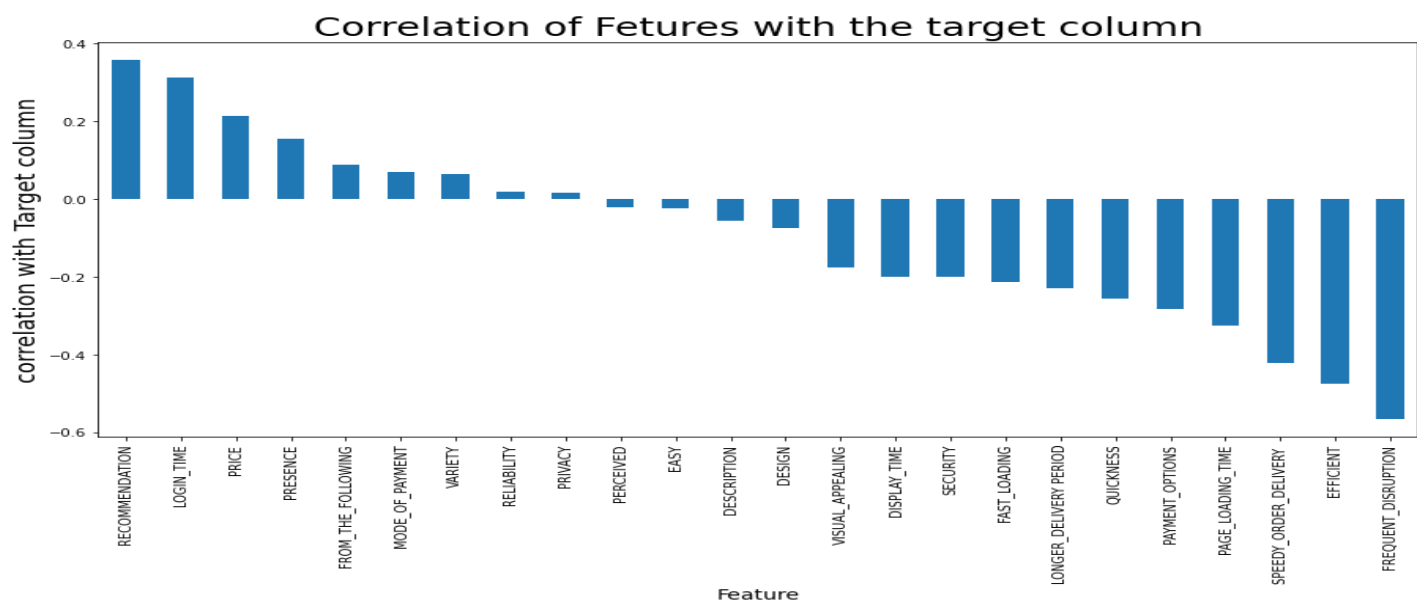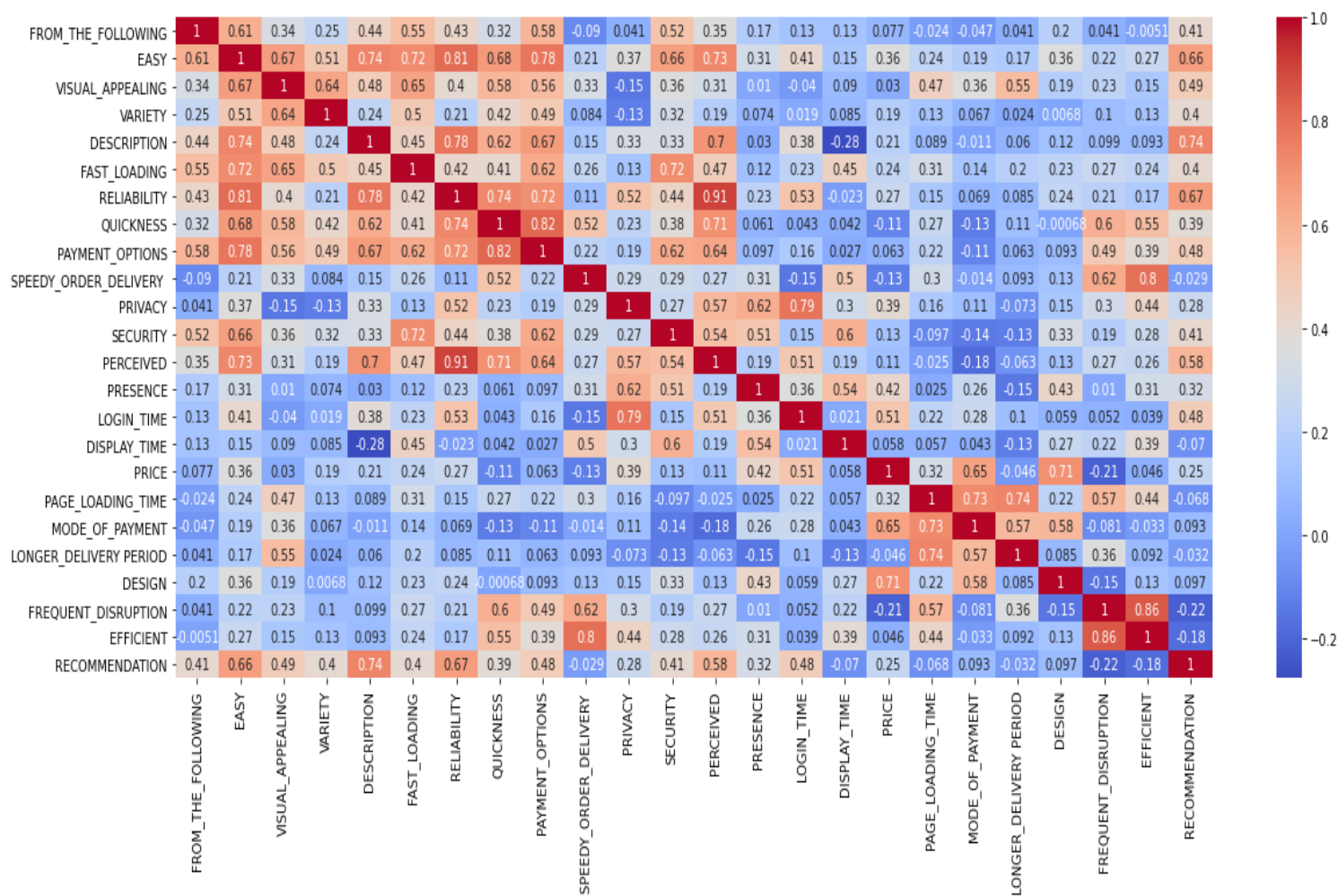
Correlation of Fetures with the target column

   a. EFFICIENT, SPEEDY_ORDER_DELIVERY, PAGELODING, PAYMENT_OPTIONS, RECOMENDATION, LOGIN TIME, PRICE Directly affect the Value for money spent.

   b. This observation says that customers are more interested in faster delivery ease of access product details over other factors which also makes the customers to purchase again from the same vendor.

Summary.
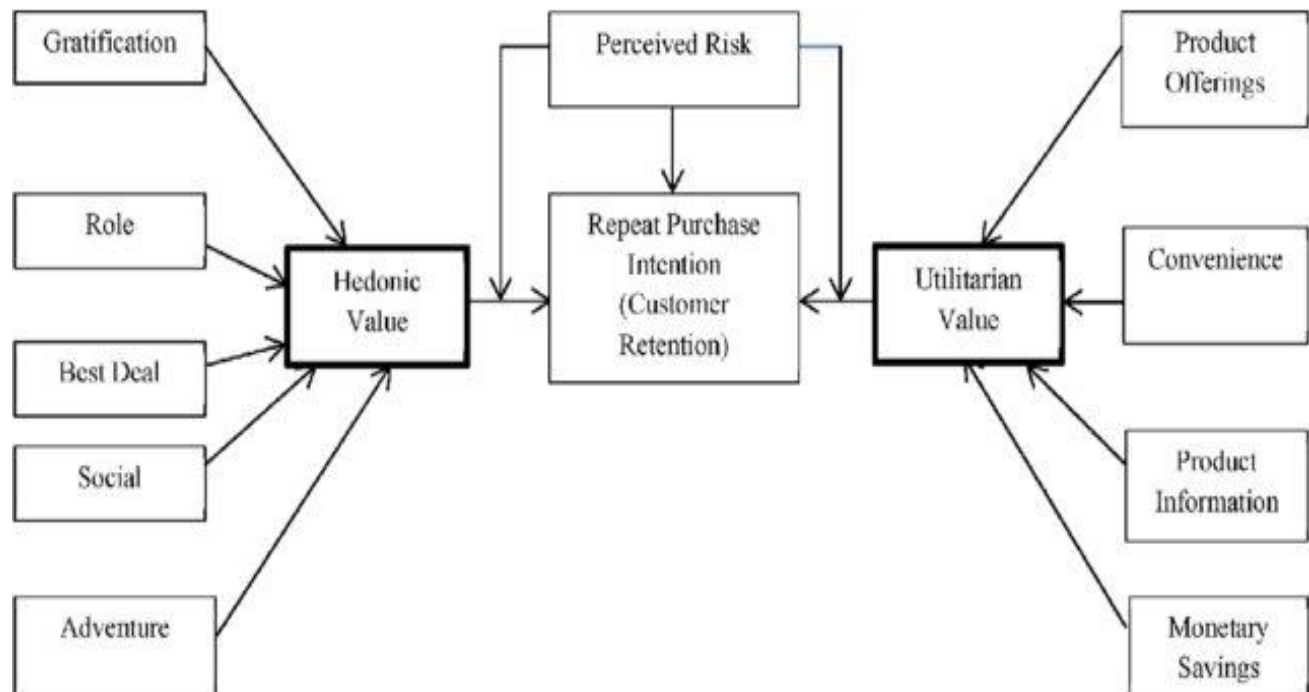
   The ease of usage of internet which includes the device used browser and internet the attractiveness of the web application/mobile application with convenience of ordering the product along with the Customer Support with Gratification makes a customer to be a repeated user.

The Customer Gratification that can be increased to make the customer feel that he is valued which directly increases the repeated usage.

Conclusion.

   E-commerce has transformed the way business is done in India. The Indian E-commerce market is expected to grow to US dollars 200 billion by 2026 from US dollars 38.5 billion as of 2021. Much of the growth for the industry has been triggered by an increase in internet and smartphone penetration, in such a market a customer can be a repeated user in depending on the factors as seen above and these factors are summarized into two such as Hedonic value and Utilitarian value. Improvement on these factors a Organization will be a leader in E-Commerce business in India.