Customer Retention

(Exploratory Data Analysis)

Submitted by:

Khushboo Khatri

ACKNOWLEDGMENT

I would like to appreciate FlipRobo mentors, for giving us the opportunity to do a project on Customer Retention (Exploratory data Analysis). Which indeed helps me doing lot of research and upscale my knowledge in the subject matter. The material and dataset provided by you guys really help us to understand the kind of observation to make data more understandable. The questionnaire based data collection is a perfect way to give an insight of any subject matter.

INTRODUCTION

E-retail factors for customer activation and retention: A case study from Indian e-commerce customers

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.

For this purpose the data is collected from the Indian online shoppers. Results indicate the e-retail success factors, which are very much critical for customer satisfaction. As a satisfied customer tents to retain with the same online store and proves to be a loyal customer.

Customer retention is a metric that measures customer loyalty, or the ability for an organization to keep its customers over time. It increases your customers' lifetime values and boosts the revenue. It also helps to build amazing relationships with the customers. The customer trust can only be obtain if online stores give them value in exchange for their money.

Review of Literature:

A Definition of Customer Retention

Customer retention refers to the activities and actions companies and organizations take to reduce the number of customer defections. The goal of customer retention programs is to help companies retain as many customers as possible, often through customer loyalty and brand loyalty initiatives. It is important to remember that customer retention begins with the first contact a customer has with a company and continues throughout the entire lifetime of the relationship.

Customer Retention Benefits

While most companies traditionally spend more money on customer acquisition because they view it as a quick and effective way of increasing revenue, customer retention often is faster and, on average, costs up to seven times less than customer acquisition. Selling to customers with whom you already have a relationship is often a more effective way of growing revenue because companies don't need to attract, educate, and convert new ones.

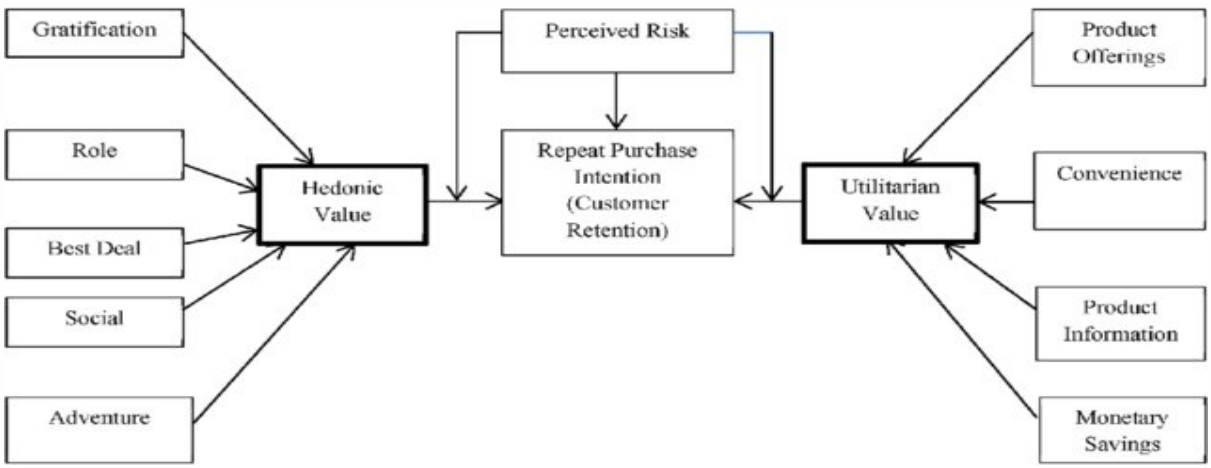
Companies that shift their focus to customer retention often find it to be a more efficient process because they are marketing to customers who already have expressed an interest in the products and are engaged with the brand, making it easier to capitalize on their experiences with the company. In fact, retention is a more sustainable business model that is a key to sustainable growth. The proof is in the numbers: according to studies done by Bain & Company, increasing customer retention by 5% can lead to an increase in profits of 25% – 95%, and the likelihood of converting an existing customer into a repeat customer is 60% – 70%, while the probability of converting a new lead is 5% – 20%, at best.

How to Improve Customer Retention

Obviously, established companies and organizations need to focus on customer retention. More important, companies are finding that customer profitability tends to increase over the life of a retained customer, so employing customer retention strategies is a worthwhile use of company resources. Some of the techniques to retain customers are as follows:

- Set customer expectations – Set customer expectations early and a little lower than you can provide to eliminate uncertainty about the level of your service and ensure you always deliver on your promises.
- Become the customers' trusted advisor – You need to be the expert in your particular field, so that you can gain customers' trust and build customer loyalty.
- Use relationships to build trust – Build relationships with customers in a way that fosters trust. Do this through shared values and fostering customer relationships.
- Take a proactive approach to customer service – Implement anticipatory service so that you can eliminate problems before they occur.
- Use social media to build relationships – Use LinkedIn, Twitter, and Facebook to connect and communicate with customers and give them a space for sharing experiences with your company, so they can become brand ambassadors.
- Go the extra mile – Going above and beyond will build strong relationships with customers and build long-term loyalty by paying attention to their needs and issues.
- Make it personal – Personalized service improves customer experience and is something customers are expecting and demanding. Make their experience personal to strengthen the bond with your brand.
- Predict your customer's wants and needs through a next best action marketing approach.

We can easily understand this from below diagram, which explains the cycle of repeat purchase, or we can say that discuss various factor which affect customers decision to repeat the purchase from the same channel.



Our objective of this project is to do Exploratory Data Analysis, for the dataset provided. Let's have a look at our dataset columns so as to have clear picture of all the data which has been recorded. Following are columns we are dealing with:

```
'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\t
',
'10 What is the operating system (OS) of your device?\t\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 fo
r 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 purcha
se decision?
',
'15 What is your preferred payment Option?\t\t\t\t\t
',
'16 How 4 do you abandon (selecting an items and leaving without making pay
ment) your shopping cart?\t\t\t\t\t\t\t
',
'17 Why did you abandon the "Bag", "Shopping Cart"?\t\t\t\t\t
',
'18 The content on the website must be easy to read and understand',
'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 impo
rtant for purchase decision.',
'21 All relevant information on listed products must be stated clearly',
'22 Ease of navigation in website', '23 Loading and processing speed',
'24 User friendly Interface of the website',
'25 Convenient Payment methods',
'26 Trust that the online retail store will fulfill its part of the transac
tion at the stipulated time',
'27 Empathy (readiness to assist with queries) towards the customers',
'28 Being able to guarantee the privacy of the customer',
'29 Responsiveness, availability of several communication channels (email,
online rep, twitter, phone etc.)',
```

'30 Online shopping gives monetary benefit and discounts',
'31 Enjoyment is derived from shopping online',
'32 Shopping online is convenient and flexible',
'33 Return and replacement policy of the e-tailer is important for purchase decision',
'34 Gaining access to loyalty programs is a benefit of shopping online',
'35 Displaying quality Information on the website improves satisfaction of customers',
'36 User derive satisfaction while shopping on a good quality website or ap plication',
'37 Net Benefit derived from shopping online can lead to users satisfaction ',
'38 User satisfaction cannot exist without trust',
'39 Offering a wide variety of listed product in several category',
'40 Provision of complete and relevant product information',
'41 Monetary savings',
'42 The Convenience of patronizing the online retailer',
'43 Shopping on the website gives you the sense of adventure',
'44 Shopping on your preferred e-tailer enhances your social status',
'45 You feel gratification shopping on your favorite e-tailer',
'46 Shopping on the website helps you fulfill certain roles',
'47 Getting value for money spent',
'From the following, tick any (or all) of the online retailers you have sho pped from;
',

From here Multiple choice question starts:

- 'Easy to use website or application',
- 'Visual appealing web-page layout', 'Wild variety of product on offer ',
- 'Complete, relevant description information of products',
- 'Fast loading website speed of website and application',
- 'Reliability of the website or application',
- 'Quickness to complete purchase',
- 'Availability of several payment options', 'Speedy order delivery ',
- 'Privacy of customers' information',
- 'Security of customer financial information',
- 'Perceived Trustworthiness',
- '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 peri od)',
- '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? '

We got total of 47 base columns and remaining multiple choice columns

We can see that our data consist of two major types of columns: 1) Continuous column 2) Categorical columns.

Again when we dive deep into the data, we can observe that there are some informational column like Gender, age range, city, pin code, Time since shopping online, Device use, how you access internet, Screen size, Operating system, Browser run, Which channel used.

Then some columns are specific about website or portal use to make online shopping by the customer: How they access after first visit, time spend on website before making purchase decision, preferred payment method,

Then about customer not making purchase: how frequently they abandon their shopping cart, why they abandon.

Then there are rating columns from 18-47: which deal with rating regarding, website, its content, information provided, about monetary savings included discounts and offers, gratification, user friendly interface, trust, user satisfaction and so on.

And lastly our dataset has columns which are like multiple choice answer, where they talk about various website/ application used by them for online shopping, based on certain criteria or in other word ease of use, hassle free transition and overall performance including trustworthiness.

From here it is clear that certain columns are merely information like pin code, we already got city to cover it, so we can drop it in other to build the ML model. We will do some basic analysis and find some correlation among our data.

	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\t	10 What is the operating system (OS) of your device? \t\t\t\t	...	Longer time to get logged in (promotion, sales period)	Longer ti in display graphics a pho (promoti sales peri
0	0	3	Delhi	110009	5	4	4	3	5	1	...	Amazon.in	Amazo
1	1	2	Delhi	110030	5	5	2	1	2	3	...	Amazon.in, Flipkart.com	Myntra.c
2	1	2	Greater Noida	201308	4	5	3	1	4	2	...	Myntra.com	Myntra.c
3	0	2	Karnal	132001	4	1	3	1	4	3	...	Snapdeal.com	Myntra.c Snapdeal.c
4	1	2	Bangalore	530068	3	2	2	1	2	3	...	Flipkart.com, Paytm.com	Paytm.c

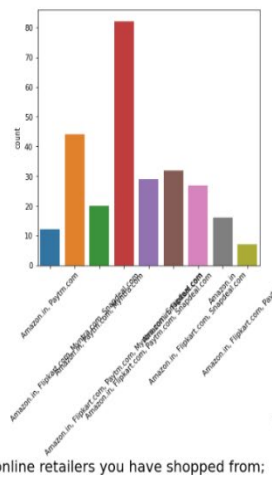
5 rows × 71 columns

0 What is the erating system (OS) of your levice? \t\t\t\t	...	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?
1	...	Amazon.in	Amazon.in	Flipkart.com	Flipkart.com	Amazon.in	Paytm.com	Flipkart.com	Amazon.in	Amazon.in	Flipkart.com
3	...	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
2	...	Myntra.com	Myntra.com	Myntra.com	Myntra.com	Amazon.in	Paytm.com	Paytm.com	Paytm.com	Amazon.in	Amazon.in, Paytm.com, Myntra.com
3	...	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
3	...	Flipkart.com, Paytm.com	Paytm.com	Paytm.com	Paytm.com	Snapdeal.com	Paytm.com	Amazon.in	Snapdeal.com	Paytm.com	Amazon.in, Myntra.com

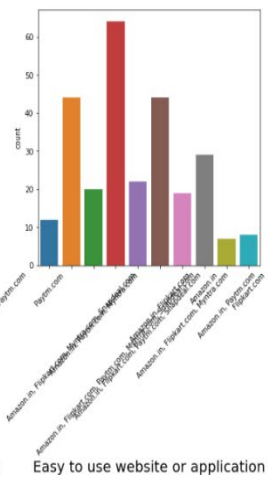
Our dataset consists of 47+ columns to be exact 71 columns. The second part i.e 48-71 column which we are seeing here is merely kind of survey taken to see where leading online shopping companies stand on various parameters, and gives us details about what website or application customers preferred over others and why.

While the first part from column 1-47 consist of actual data, which further segregate into two parts, one is customer information and choices from 1-17 column and other part from column 18-47 is their response towards the information provided by online shopping website or application. That information ranges from information about products, features, seller, accessibility, mode of payment, returns and replacement, customer service, variety, range of products and the list goes on.

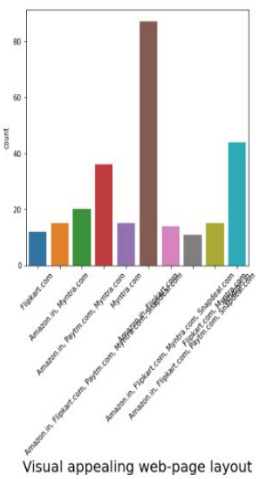
So, our analysis also divides into two categories. To begin with we analyse individually second part of the dataset i.e. from column 48-71, Just to see what's people opinion or rating for E-retail stores in Ind ia, these questionnaires are multiple choice type, where customer pick all their selections. The E- Reta il stores included here are Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com.



From the following, tick any (or all) of the online retailers you have shopped from;

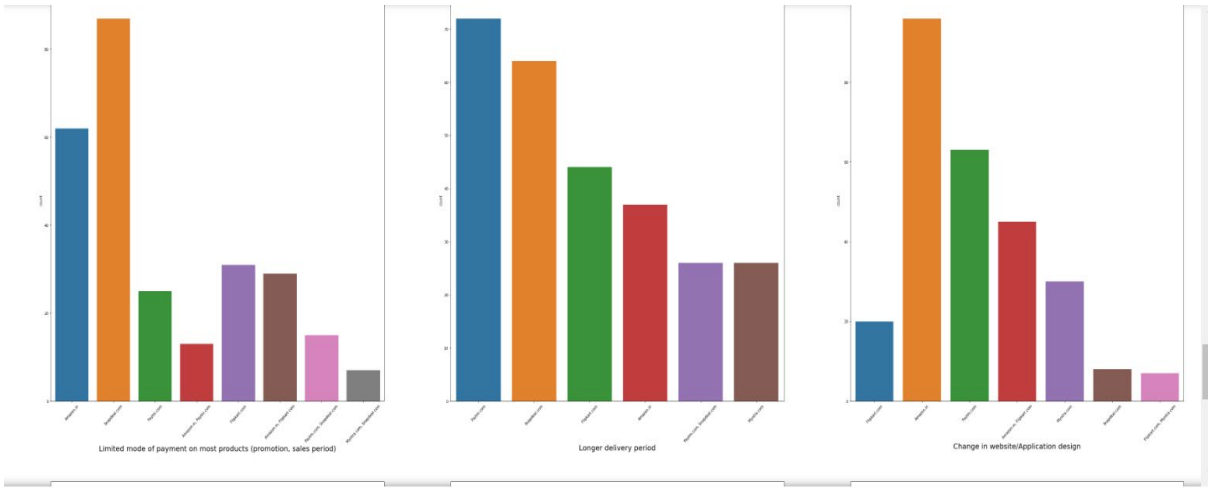


Easy to use website or application

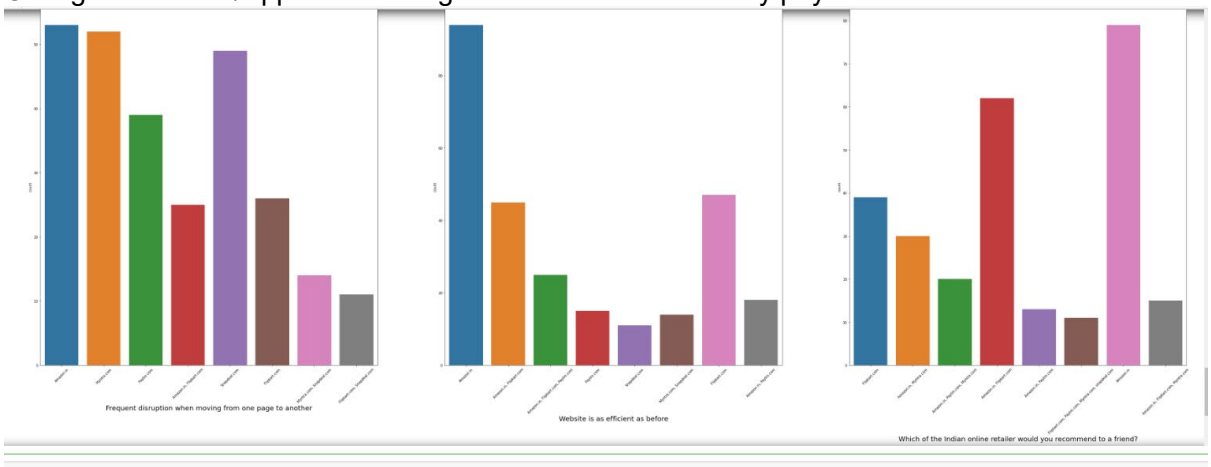


Visual appealing web-page layout

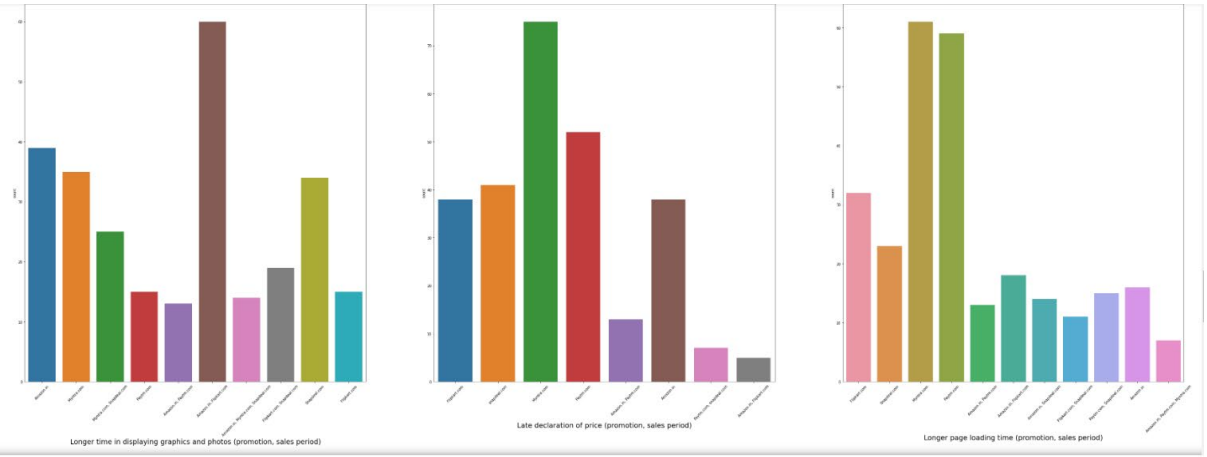
From here we can observe that customer basically shop from Amazon.in, Flipkart.com, Paytm.com, Myntra.com and Snapdeal.com .
Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com are easy to access websites or appl ication.
Amazon.in, Flipkart.com has highest visual appeal in there web layout.



From here following observation can be made :-
Limited mode of payment on most products (promotion, sales period) : Snapdeal.com
Longer delivery period : paytm.com followed by Snapdeal.com.
Change in website/Application design : Amazon.in followed by paytm.com



From here we can observe that Frequent disruption when moving from one page to another : Am azon.in followed by Myntra.com.
Website is as efficient as before: Amazon.in.
Which of the Indian online retailer would you recommend to a friend?: Well Amazon.in got highes t recommendation followed by Flipkart.



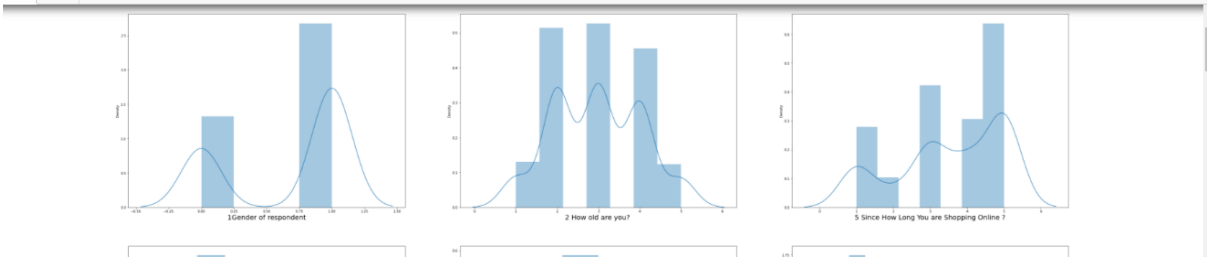
Following observation can be made from above graphs :-
Longer time in displaying graphics and photos (promotion, sales period): Amazon.in, Flipkart.com
Late declaration of price (promotion, sales period) : Myntra.com
Longer page loading time (promotion, sales period) : Myntra.com followed by paytm.com.

We did plot some more graphs in this category in our Jupyter Notebook, and come out with the observation that Amazon and Flipkart are the most trustworthy, competent, efficient and secure website which customer believes. Then later on followed by Myntra.com, paytm.com, snapdeal.com.

Now, let's deal with first part of the dataset, to start with we draw distribution plot to see how our data is distributed, as our data is basically discrete in value we are not able to see some definite pattern. However, we are able to observe the inclination of customers. As from column 18-47 we have rating questionnaire, which clearly helps us to predict few things. Here I am sharing my observation, you can refer the graph from Jupyter Notebook.

The below mention code is use to visualize the dataset from column 1-47, with glimpse of how data look in first three columns.

```
# visualizing continuous column data
plt.figure(figsize=(50,200))
plotnumber=1
for column in df[conti_col]:
    if plotnumber<=47:
            ax=plt.subplot(16,3,plotnumber)
            sns.distplot(df[column])
            plt.xlabel(column,fontsize=20)
            plotnumber+=1
plt.show()
```



The observation which I made are as follows:

1. The distribution of the data clearly shows that almost all the columns have discrete values, which is quite clear as this is coded version of categorical data.
2. As we know that 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 and from above categorical visual analysis amazon stand correct in almost every field.

When we observe above continuous data visualization we can see that preferred payment option is mostly credit/debit card.

3. People Sometimes also abandon their cart as they get better alternative option/offer.

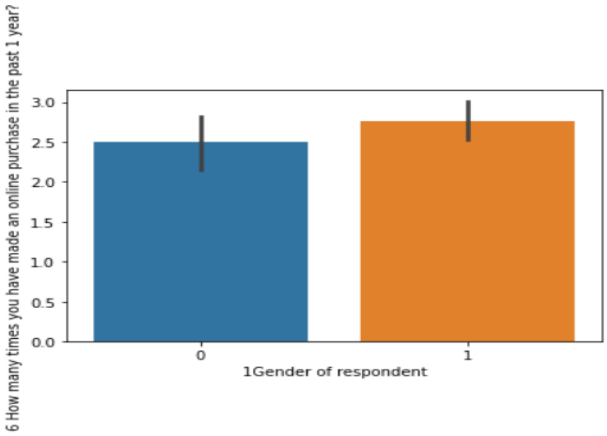
4. Customer strongly agree that the content of the website must be easy to read and understand
5. We can see that customer strongly agrees on security, shopping convenience , monetary benefits and discount i.e utilitarian values are highly important for customers. As they strongly agree on the utilitarian features.
6. Customers also seeks great satisfaction on good quality website, net benefit derived from shopping, wide variety, complete relevant information about product.

7. Customer also shows indifferent opinion about gratification, it enhance your social status.
8. And agrees through online shopping they get the value for money spent. It gives them sense of adventure i.e. Hedonic values also plays important role.

From here we also observe that same information is collected by twisting the question but as it is more of behavioural approach, it gives clearer outlook about the customer how he deals in particular situation.

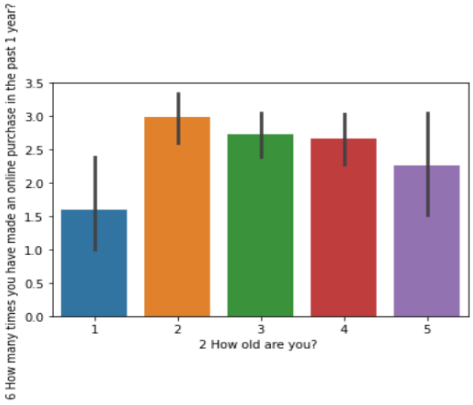
The above-mentioned analysis is the observation which comes under the part of Univariate analysis. Now we want to know How data is connected to each other, and their impact. So we perform some multivariate analysis. Also I assume that the label for this dataset is Number of times customer made purchase in the past 1 year, and treating rest as features.

In the below graph, we want to observe how different gender group shop online, here 0 denotes Male while 1 denotes Female.



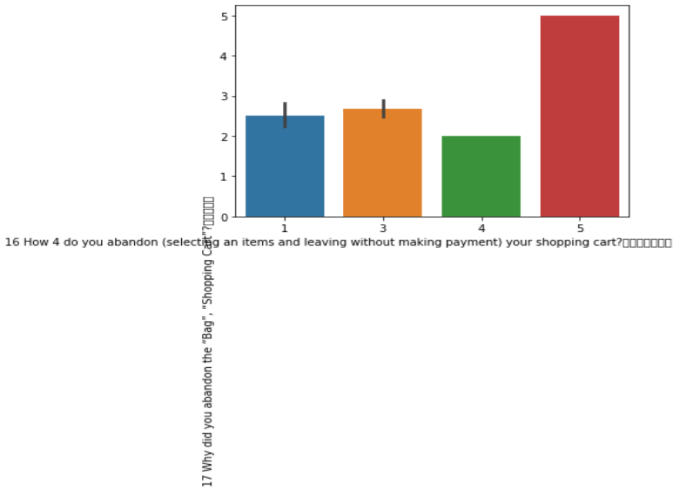
And we can clearly see that Female did more online shopping than male.

From here we can observe that the Age range between 21-30 did more online shopping, while customer below 20 are way less than rest of them



we can observe that 21-30 age group people do alot of online shopping while least online shopping is done by people less than 20.

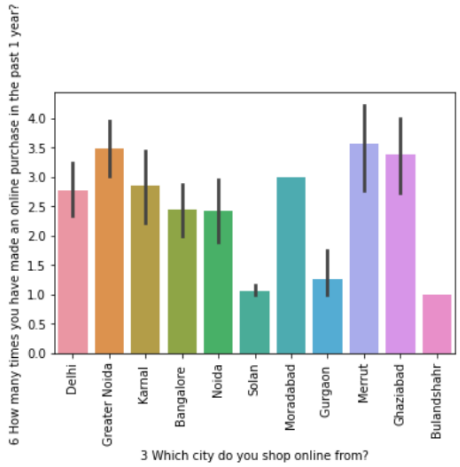
Here we try to observe why customer abandon their cart or bags and how frequently they did that.



From here we can observe that people very frequently abandon carts because there promo code is not applicable. And the lowest value belongs to frequently and the reason is better alternative offer.

From above graph we can clearly see that one of the major reasons why customer abandon their cart/ bags in online shopping is because the promo code is not valid for that purchase. And least amount of customer abandons their cart as they find better option.

Below we can see which city customers makes how much purchase.



We can observe that over past 1 year more number of are made from Greater Nodia, followed by Meerut and Gaziabad. Its interesting to observe that sub part(cities touching Delhi) of Delhi are more in online shopping than Delhi itself.

Statistical Analysis:

Statistical analysis is the collection and interpretation of data in order to uncover patterns and trends. It is a component of data analytics. It helps to get some insight information about the data. Let's have a look at our statistical description.

	1Gender of respondent	2 How old are you?	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\t	10 What is the operating system (OS) of your device? \t\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?	...	38 User satisfaction cannot exist without trust	39 Offerin a wic variety liste product sever categor
count	269.000000	269.000000	269.000000	269.000000	269.000000	269.000000	269.000000	269.000000	269.000000	269.000000	...	269.000000	269.0000
mean	0.669145	2.959108	3.524164	2.672862	3.260223	1.676580	4.282528	1.776952	1.275093	1.360595	...	4.182156	4.1486
std	0.471398	1.066012	1.436586	1.651788	1.135887	0.843904	0.923426	0.797892	0.645429	0.897805	...	1.072162	0.8421
min	0.000000	1.000000	1.000000	1.000000	2.000000	1.000000	2.000000	1.000000	1.000000	1.000000	...	1.000000	2.0000
25%	0.000000	2.000000	3.000000	1.000000	2.000000	1.000000	4.000000	1.000000	1.000000	1.000000	...	4.000000	4.0000
50%	1.000000	3.000000	4.000000	2.000000	3.000000	1.000000	4.000000	2.000000	1.000000	1.000000	...	4.000000	4.0000
75%	1.000000	4.000000	5.000000	4.000000	5.000000	2.000000	5.000000	2.000000	1.000000	1.000000	...	5.000000	5.0000
max	1.000000	5.000000	5.000000	5.000000	5.000000	4.000000	5.000000	3.000000	4.000000	4.000000	...	5.000000	5.0000

3 rows × 45 columns

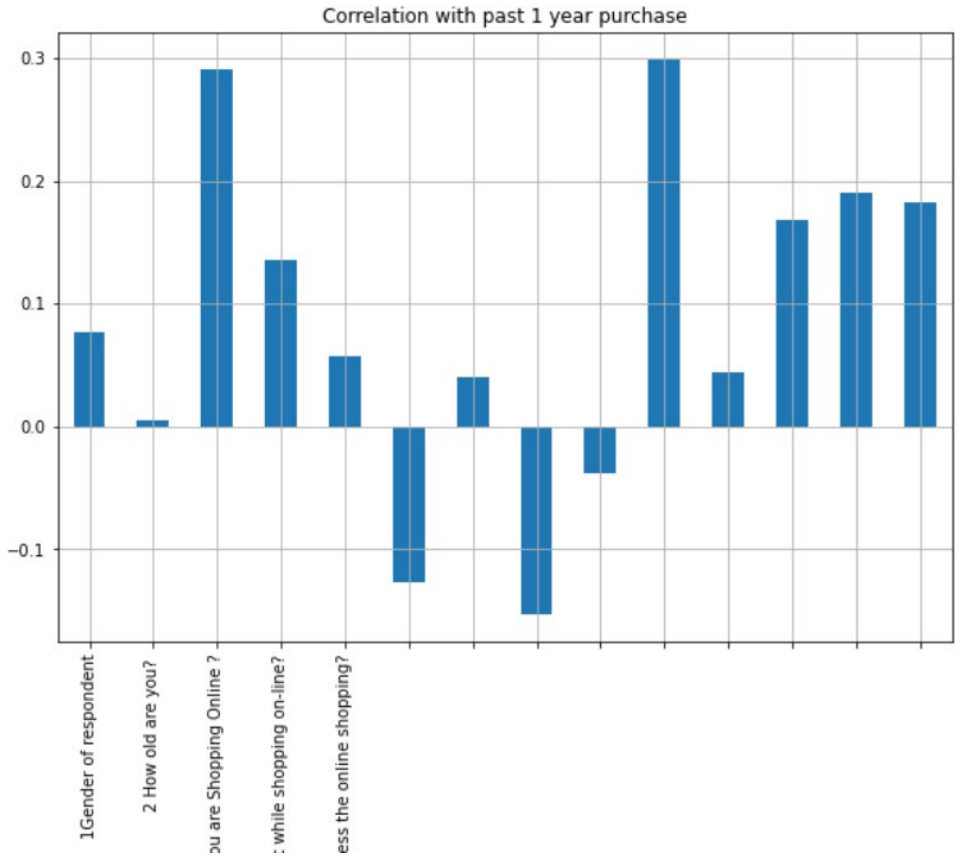
We can observe that this piece belongs to informational set of data, one thing to note here is that statistical analysis can only be perform on numerical data. The mean shows that where most of the response lies for our dataset column wise.

is he ng im of ur e? 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?	...	38 User satisfaction cannot exist without trust	39 Offering a wide variety of listed product in several category	40 Provision of complete and relevant product information	41 Monetary savings	42 The Convenience of patronizing the online retailer	43 Shopping on the website gives you the sense of adventure	44 Shopping on your preferred e-tailer enhances your social status	45 You feel gratification shopping on your favorite e-tailer	46 Shopping on the website helps you fulfill certain roles	47 Getting value for money spent
00	269.000000	269.000000	...	269.000000	269.000000	269.000000	269.000000	269.000000	269.000000	269.000000	269.000000	269.000000	269.000000
52	1.275093	1.360595	...	4.182156	4.148699	4.349442	4.263941	3.914498	3.553903	3.223048	3.501859	3.282528	4.163569
92	0.645429	0.897805	...	1.072162	0.842110	0.755953	1.000485	0.693879	1.065869	1.219581	1.141564	1.178995	0.648773
00	1.000000	1.000000	...	1.000000	2.000000	2.000000	2.000000	3.000000	1.000000	1.000000	1.000000	1.000000	3.000000
00	1.000000	1.000000	...	4.000000	4.000000	4.000000	4.000000	3.000000	3.000000	3.000000	3.000000	3.000000	4.000000
00	1.000000	1.000000	...	4.000000	4.000000	5.000000	5.000000	4.000000	4.000000	3.000000	3.000000	3.000000	4.000000
00	1.000000	1.000000	...	5.000000	5.000000	5.000000	5.000000	4.000000	4.000000	4.000000	4.000000	4.000000	5.000000
00	4.000000	4.000000	...	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000

We can observe that the mean value for from column 18-42, 47 are around 4 which means most of the customers agrees with the statement given, while from column 43-46 we can responses mean as 3-3.5 meaning customers do have indifferent opinion about online shopping as in getting sense of adventure, enhance social status, fulfil certain roles, basically some of the hedonic values. While they feel strongly in favour about their utilitarian values.

Let’s see how informational features are correlated with past 1 year purchase.

We got dataset where we are applying the same rule to predict the outcome, Lets see how our dataset looks like.

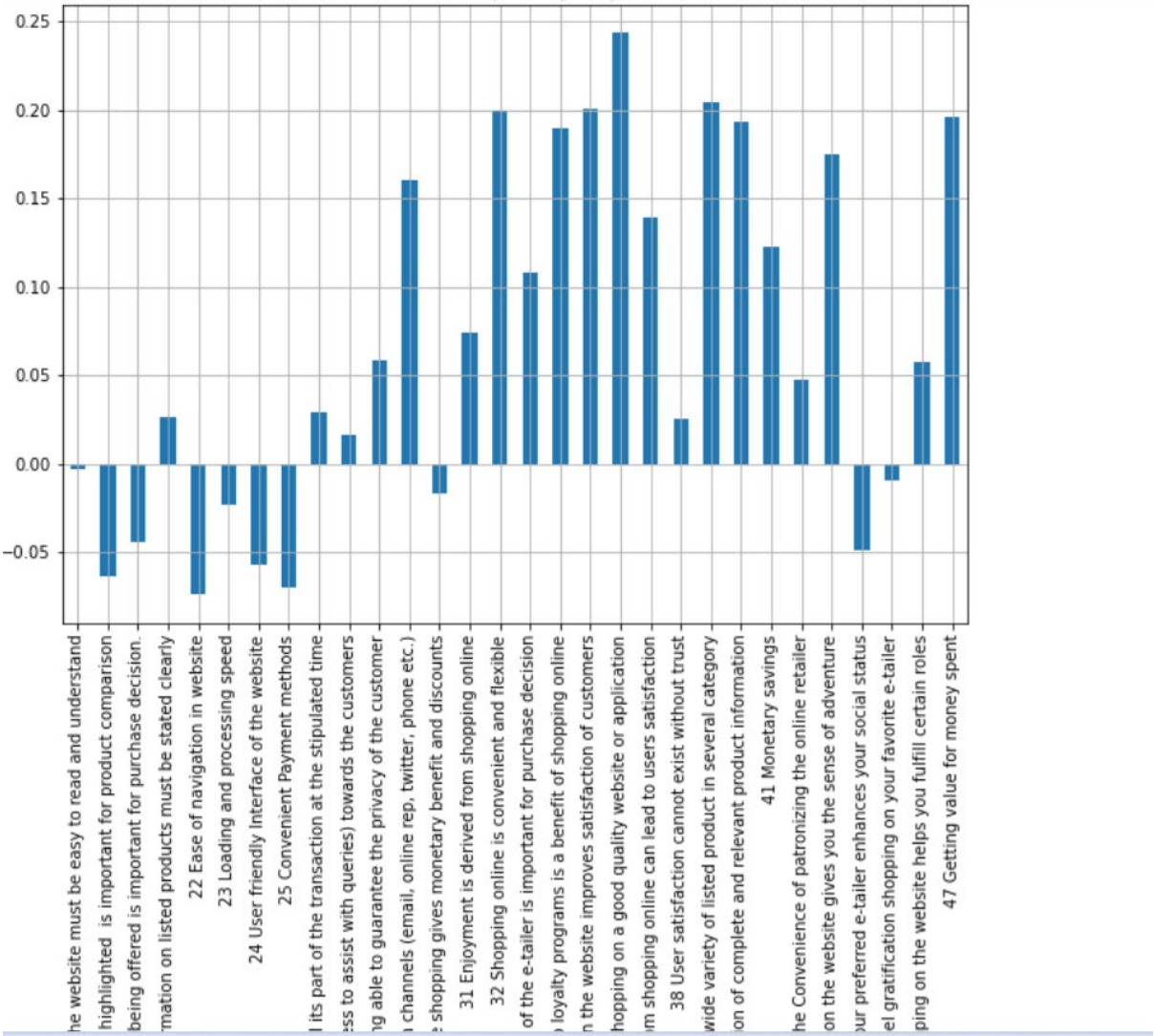


In this correlation we include all the first 17 columns of our dataset, where purchase column become our dependent variable and others all are independent variable. While talking about correlation we basically see two types of correlation- 1) Positive correlation 2) Negative correlation.

From here we observe that 'After first visit, how do you reach the online retail store?' is highly correlated to repeat purchase, and second feature which also shows very strong correlation is 'Since How long you are shopping online?' while the least correlated feature is 'How old are you?'.

Some of the features are positively correlated i.e. they are directly proportional relation with each other, while some feature are negatively correlated i.e. inversely proportional with each other, one increases the other decreases. Both positively and negatively correlated features are equally important.

We will do some more plotting with the rating columns to understand how much they are correlated to past 1 year purchase.



From here we can say that customer retention is highly correlated with quality information which leads to customer satisfaction. Also, we notice that, website must be easy to read and understand is least correlated with past 1 year purchase.

Conclusion:

Customer retention is highly correlated with customer satisfaction, a happy customer will retain with the app or website for future purchases. For that we can divide the factors into major two categories Utilitarian and Hedonic values.

Under Utilitarian values we include product information, product offerings, monetary benefits, and convenience. These are the major factors which can affect customer decision to make purchase, while Hedonic values basically related to outer self, idealism related to sense of adventure, role fulfilment, social status, gratification and list goes on. These are basically things we did for sake of others.

We also see that females are move likely to do online shopping than male, also young people in age range of 21-30 are likely to do more online shopping than others.

In short, we can say that out data definitely provide us some insight about the customer retention on online platforms, and in second part of data we also observe that out of all websites mentioned, Amazon.in followed by Flipkart full fill most of the tick marks with flying colours. customer trust on these platforms with their information, security of data and overall customer service/convenience offered by them.