



# **Customer Retention**

Submitted by:

Sachin Gupta

## **ACKNOWLEDGMENT**

It is a pleasant gratification to present Customer Retention report.

Working on this project was an incredible experience that will have a tremendous impact on my career.

I have completed this project by taking help of Google, Bing & various links.

Such as :

<https://www.ibef.org/industry/ecommerce.aspx>

[https://www.researchgate.net/publication/285927801\\_Customer\\_Retention](https://www.researchgate.net/publication/285927801_Customer_Retention)

# INTRODUCTION

Customer retention means the process of maintaining or keeping customers once you have acquired them. It's all the activities that a company must do in order to keep their customers around. The goal is to build a long-lasting relationship between the brand and consumers. Once a customer becomes loyal to your brand, not only he will buy more from you than a normal customer but he'll spread good words about your business, increase your reputation.

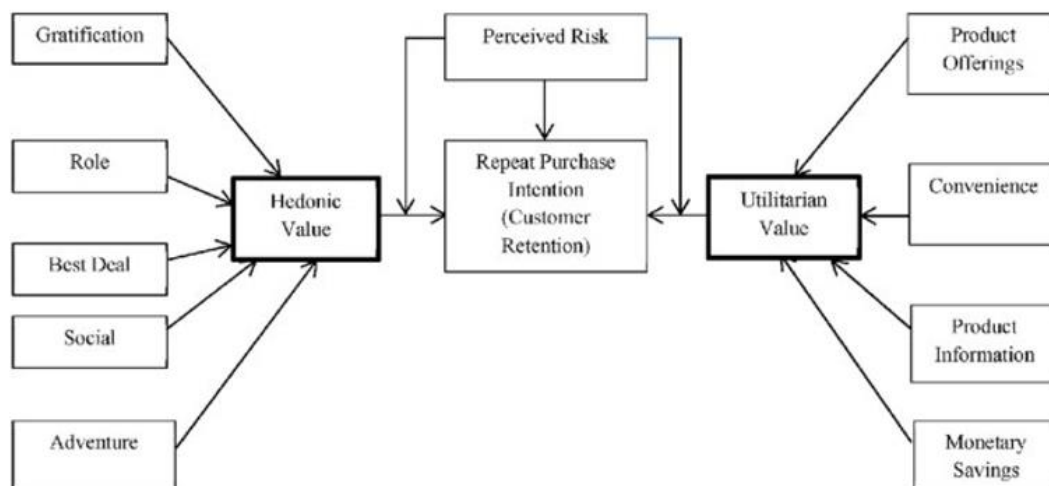
## ❖ Business Problem Framing:

We are required to do the data analysis of the customer feedback with the available independent variables. This data analysis will be used by the management to understand how exactly the customer feel about products and customer experience journey while purchasing a product and post order feedback for a better customer satisfaction. They can accordingly manipulate the strategy of the firm and concentrate on areas that will yield repeat purchase and better customer loyalty. Further, the data analysis will be a good way for the management to understand the customer better and rethink of a better strategy for customer retention.

## ❖ Conceptual Background of the Domain Problem:

Customer satisfaction has come out as one of the most important factors that guarantee the success of online store; it has been ranked as a key stimulant of purchase, repurchase intentions and customer loyalty. A whole review of the literature, theories and models have been taken into consideration for 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. The company wanted to a data analysis to understand various aspect of customer satisfaction and also make a better competitive strategy.

- ❖ **Review of Literature:** E-commerce market has been contributing to the significant growth for the GDP of the country. It has been continuously growing at more 8% CAGR on each year. And customer being a one of the stakeholder for the e-commerce players, every company wants to retain their valuable customer and grow customer loyalty. For improving the customer satisfaction each enterprise investing/spending a lot not only to improve their sales but improve the CSAT, that is customer satisfaction score.
- ❖ For this particular project we are provided with a data set to a data analysis to gather insights for the customer feedback and help us understand the customer in a better way



#### ❖ Motivation for the Problem Undertaken:

Since e-commerce growth has been significant so far at each industry for the last couple of years, understanding the customer who actually purchase your product and their journey post order in the platform plays a really crucial role for retaining a customer. From the dataset we got the feedback of each of the parameters for a customer and through the data analysis each significant parameter has been dig down and understand the deeper meaning of it. This project helps me understand the objective of customer feedback which at enterprise level has a greater focus.

# Analytical Problem Framing

## ❖ Mathematical/ Analytical Modelling of the Problem:

In this particular project I need to understand the customer satisfaction for each of the parameters provided. I have done the exploratory data analysis process and try to figure out the customer in a better way.

## ❖ Data Sources and their formats:

- ❖ Data sources are provided internally by the enterprise.

- ❖ Total datasets contain 269 rows and 71 columns and each of the features is extremely important to under the customer.

## ❖ Data Pre-processing:

In the data pre-processing stage, I have found out if there is any missing data in dataset, for a particular column if there are any outliers present and how to handle the outliers. I have also found the total shape of the data set. I have also found out the dataset description using describe method. So, in this pre-processing process I have mainly cleansed the data and prepared the right set of data for further processing.

## ❖ Data Inputs- Logic- Output Relationships:

To find out the relationship between all the input variable I have used correlation function and find out whether there is a positive/negative relationship between a pair of variables. From this describe function that also known as Five-point summary analysis if there are any outliers are present for a particular column.

## ❖ Hardware and Software Requirements and Tools Used:

For this particular dataset the Hardware is used Windows as operating system, and the software used are mainly Jupyter notebook for model building and various internal packages that are defined in the anaconda/jupyter notebook.

# Data Analysis Development and Evaluation

## ❖ Identification of possible problem-solving approaches (methods):

For this particular project I have done EDA to understand the outcome of this dataset. And also, I have used various visualization using sea born & matplotlib package to understand the customer satisfaction in a suitable manner.

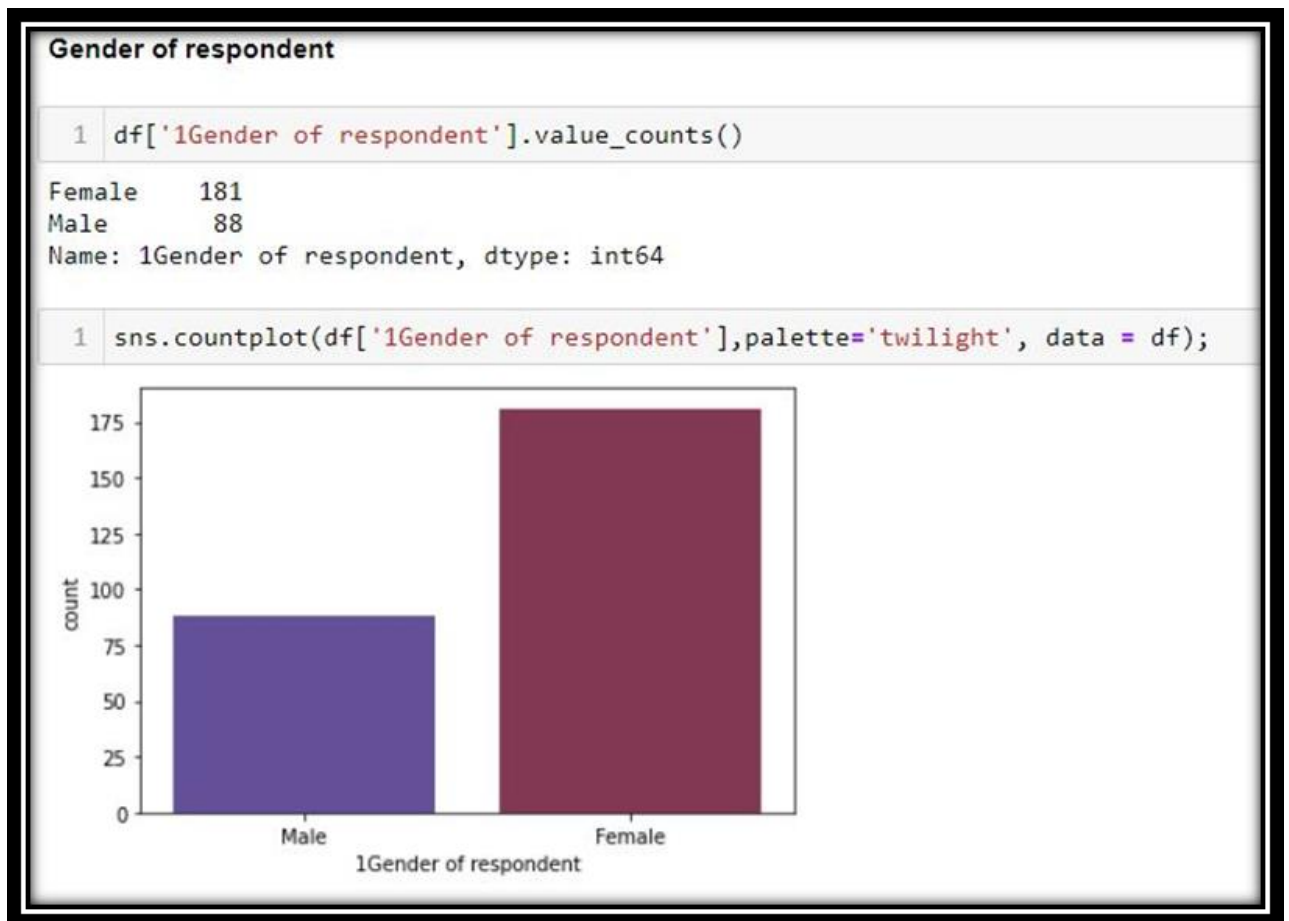
Key Metrics for success in solving problem under consideration

The key metrics that were mainly taken into consideration were the followings:

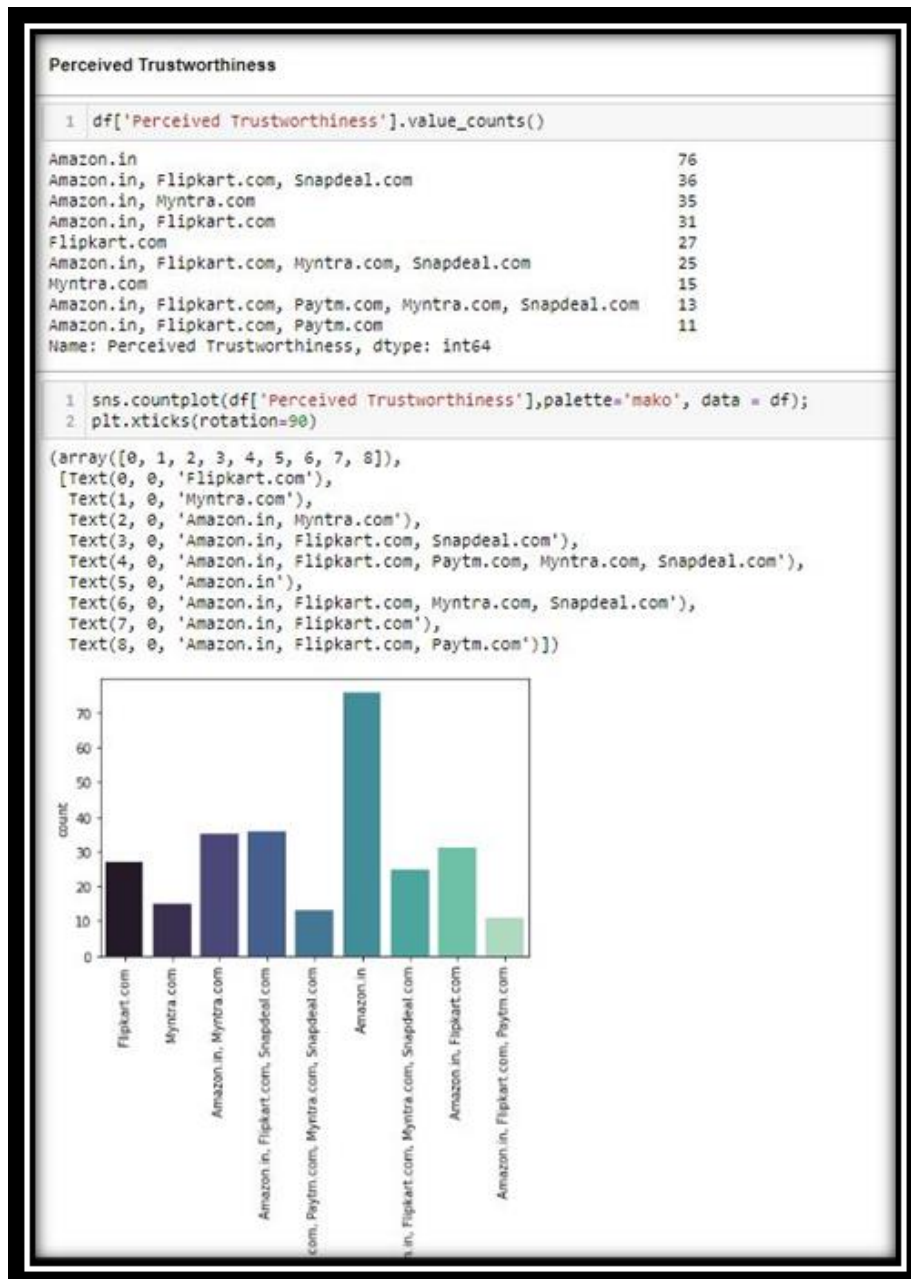
- Gender of the respondent
- What browser do you run on your device to access the website?
- Which device do you use to access the online shopping?
- Which of the Indian online retailer would you recommend to a friend?
- Quickness to complete purchase
- Speedy order delivery
- Presence of online assistance through multi-channel
- Getting value for money spent
- Return and replacement policy of the e-tailer is important for purchase decision
- Shopping online is convenient and flexible
- Convenient Payment methods
- User friendly Interface of the website

These are the prime metrics under consideration, but there are factors too can be considered for solving the house price prediction.

## ❖ Visualizations



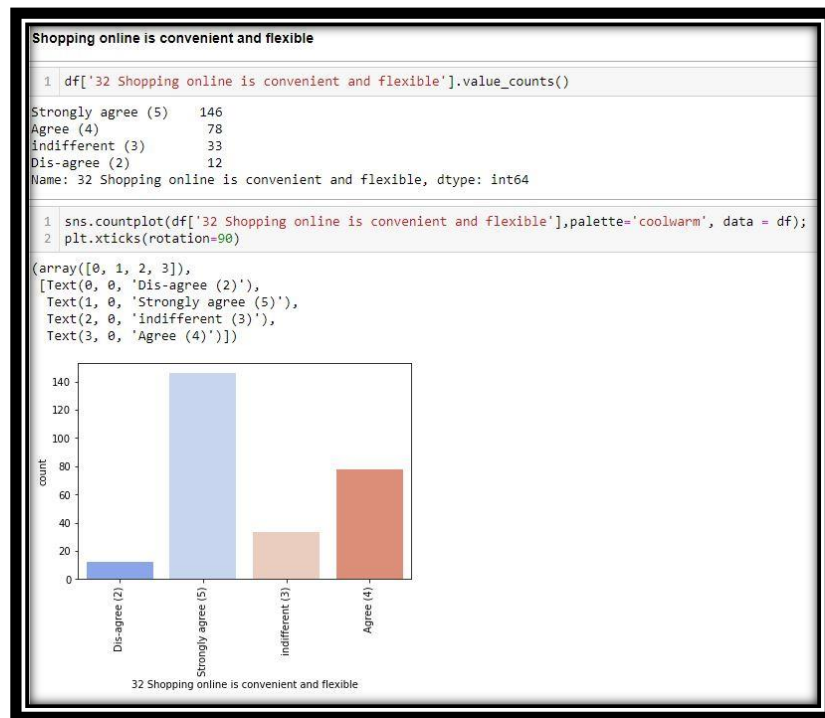
From the above count plot I can say that female prefer online shopping more than male as the number of female for online shopping is almost twice than male.



Trust plays the vital role in any brand.

The above graph I can say that amazon have huge brand value as it is trust worthy.





In online shopping customer are strongly agree with the fact that Shopping online is convenient and flexible.

## Label Encoder Process

Label Encoder

```

1 # Using Label encoder
2
3 le = LabelEncoder()
4 df = df.apply(LabelEncoder().fit_transform)
5 df.head()

```

	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?	...	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)
0	1	1	2	1	3	2	0	0	3	2	...	0	0	3	5
1	0	0	2	5	3	3	3	2	0	1	...	1	6	7	10
2	0	0	4	23	2	3	1	2	2	0	...	7	6	4	7
3	1	0	6	11	2	5	1	2	2	1	...	9	7	4	8
4	0	0	0	31	1	0	3	2	0	1	...	5	8	5	8

Applied Label Encoder to convert all the string columns into integers.

1	df.describe()												
	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?	...	Longer time to get logged in (promotion, sales period)	Longer time in displaying graphics and photos (promotion, sales period)
count	269.000000	269.000000	269.000000	269.000000	269.000000	269.000000	269.000000	269.000000	269.000000	269.000000	...	269.000000	269.000000
mean	0.327138	1.330855	4.494424	18.163569	2.323420	3.237918	2.078067	1.501859	2.256506	1.137546	...	4.044610	4.063197
std	0.470042	1.183774	3.187687	10.343865	1.176357	1.739331	0.715919	0.751240	0.944942	0.867985	...	3.343218	3.177536
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000
25%	0.000000	0.000000	2.000000	10.000000	1.000000	2.000000	2.000000	1.000000	2.000000	0.000000	...	1.000000	1.000000
50%	0.000000	1.000000	4.000000	19.000000	3.000000	3.000000	2.000000	2.000000	2.000000	1.000000	...	3.000000	4.000000
75%	1.000000	2.000000	7.000000	24.000000	3.000000	5.000000	3.000000	2.000000	3.000000	2.000000	...	7.000000	7.000000
max	1.000000	4.000000	10.000000	38.000000	4.000000	5.000000	3.000000	3.000000	3.000000	2.000000	...	9.000000	9.000000
8 rows × 71 columns													

From the above describe method :

I find out that how each parameters are distributed across the dataset & by using

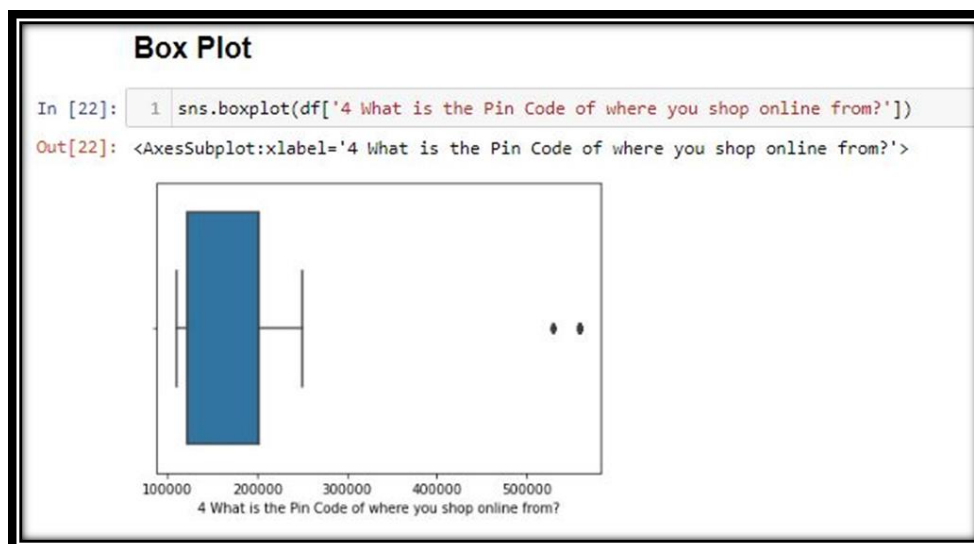
this method I got to know the five point summary analysis like mean, max and the quartiles

## Distribution Plot



From the above Distribution Plot I can say that Most of the columns are Categorical in nature & the dataset is normally distributed.

## Box Plot



There is no outliers in any columns.

## **Interpretation of the Results**

- ❖ I have used various visualization tool such as Count Plot, Box Plot & Distribution Plot to understand the data in a better way.
- ❖ I have also used label encoder technique and convert all the data into numerical form to do the data analysis in an easier way.
- ❖ I used describe method for five-point summary analysis and also found the number of rows and columns in dataset.
- ❖ There is no missing values in the dataset and no outliers detected in the dataset.

# CONCLUSION

From this project I have research several analysis:

a) Having the right customer retention strategy will keep your company grows if you know how to take advantage of it.

b) Your customers will find their way back and continue buying stuff from you.

c) At the end of the day, what you need to do is fulfilling your customers' needs, listen to their complaints and provide professional solutions.

d) If your customers feel like they are valued, they will become increasingly loyal to your brand.

e) According to my research & analysis Customer are the best wealth wisher of a company if they are certified with ultimate useful good products.

## ❖ Key Findings and Conclusions of the Study:

- I used various visualization methods and understand the EDA in a better way
- This customer satisfaction can be used as an impact of e-commerce market development as well as for economic development of the country.

## ❖ Learning Outcomes of the Study in respect of Data Science:

As per as learning outcomes is concerned, I have learnt the following things:

- Visualization helps us understand the data graphically.
- From describe method we can get some knowledge related to outliers present in the particular columns (large difference between 75th percentile and maximum percentile)
- I also understand the about reading various related features and importance of them in the whole dataset

### Challenges:

- It was difficult to read each columns and comparing with others on the dataset in notebook as it took some time to understand and analyse it a proper way.

### ❖ Limitations of this work and Scope for Future Work:

Since I have only used a sample dataset, hence sometimes it is difficult to understand the overall impact of this customer feedback at a larger scale.