

[**E-retail factors for customer activation and retention: A case study from Indian e-commerce customers**](https://www.researchgate.net/publication/346412647_E-retail_factors_for_customer_activation_and_retention_An_empirical_study_from_Indian_e-commerce_customers)

Submitted by:

ARUMOY SAHA

**ACKNOWLEDGMENT**

Following are the references used in this e-retail customer feedback project:

* <https://www.researchgate.net/publication/338305338_Online_Retail_in_India_A_Comparative_Analysis_of_top_Business_Players>
* <https://www.ibef.org/industry/ecommerce.aspx>
* <https://www.icommercecentral.com/open-access/sustainability-of-eretail-in-india.php?aid=58193>

**INTRODUCTION**

* Business Problem Framing:

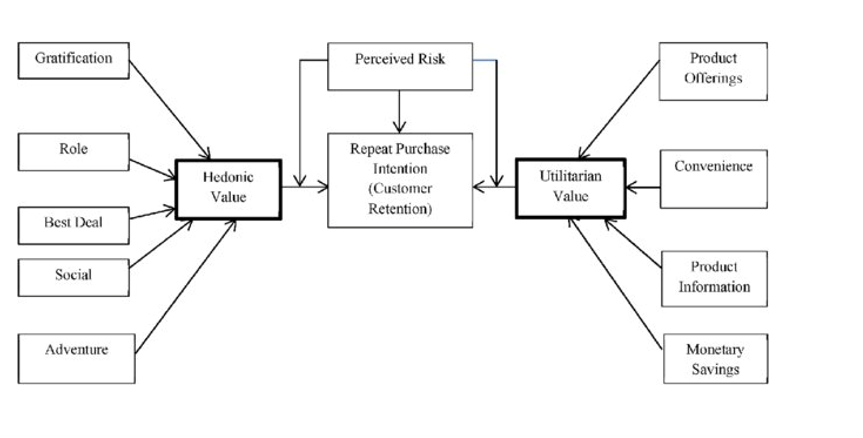
We are required to do the data analysis of the customer feedback with the available independent variables. This data analysis will be used by the management to understand how exactly the customer feel about products and customer experience journey while purchasing a product and post order feedback for 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 seaborn & matplotlib package to understand the customer satisfaction in a suitable manner.

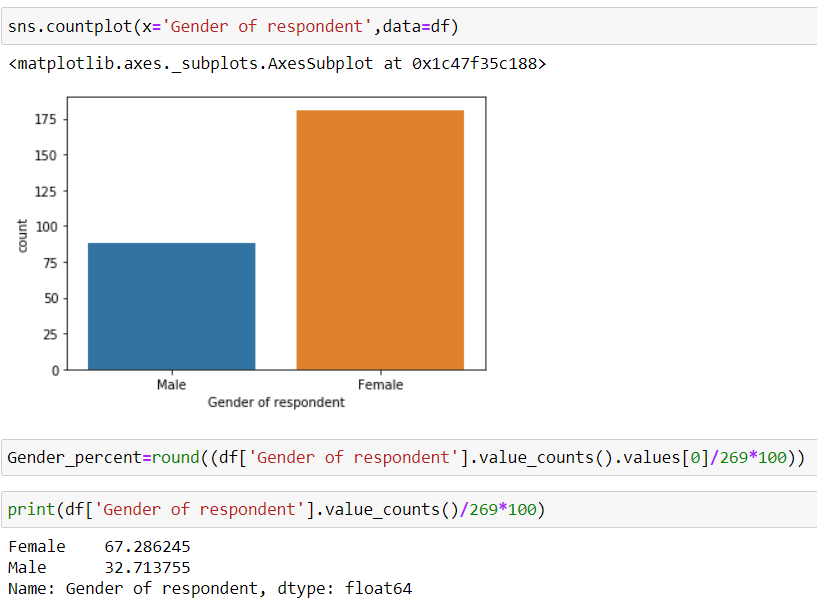
* Key Metrics for success in solving problem under consideration

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

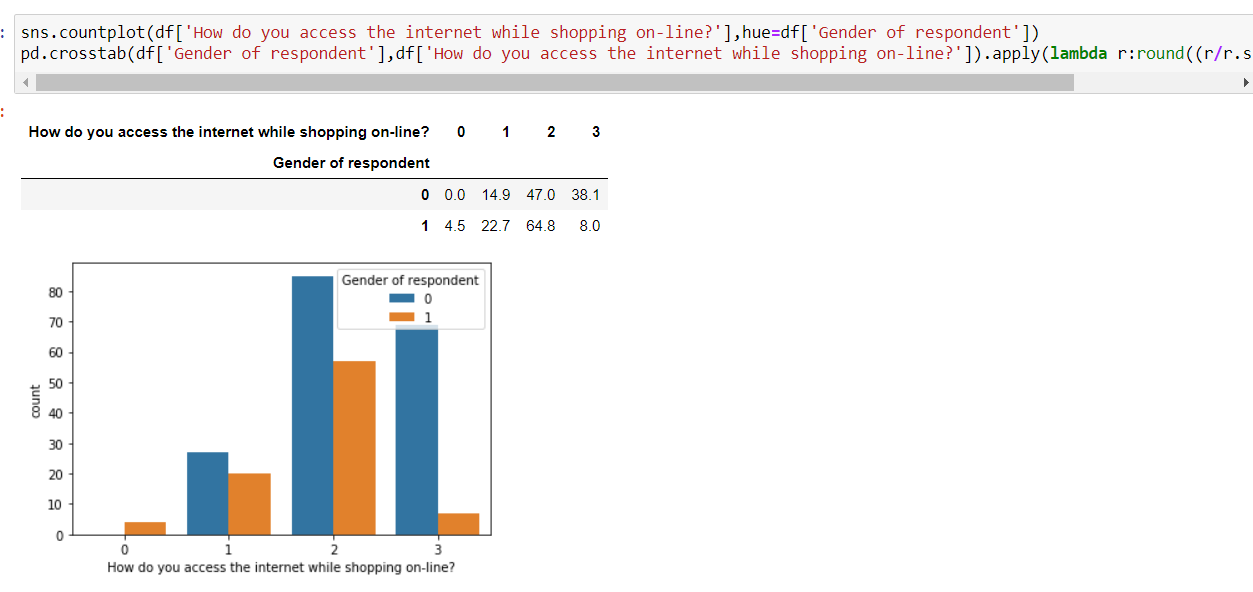
* Gender of the respondent
* What browser do you run on your device to access the website?
* Which device do you use to access the online shopping?
* Which of the Indian online retailer would you recommend to a friend?
* Quickness to complete purchase
* Speedy order delivery
* Presence of online assistance 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



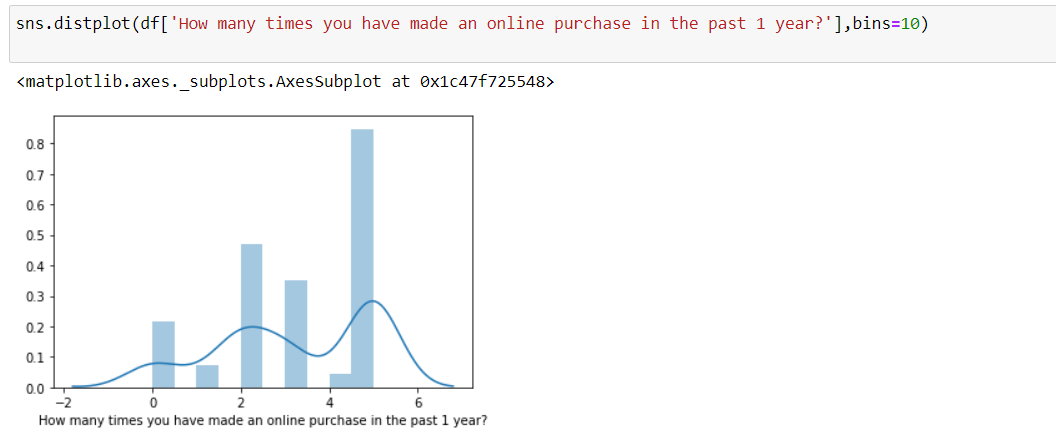
From the above count plot we got to know the respondent percentage of gender that is distributed across the whole dataset using the seaborn library.



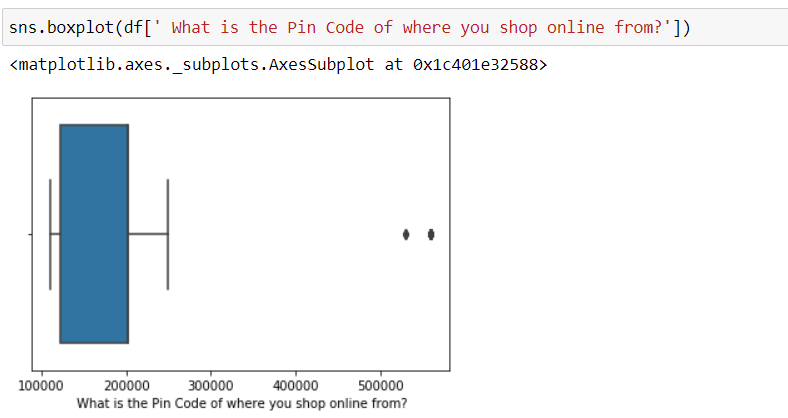
From the above countplot, we find out gender wise distribution of data for how the access the internet while shopping online using the crosstab function that build the data in a group.



From the above visualization we understand the device customers do for online shopping, so devices are like desktop, laptop, tablet, Smartphone etc.we have segregated it gender wise.



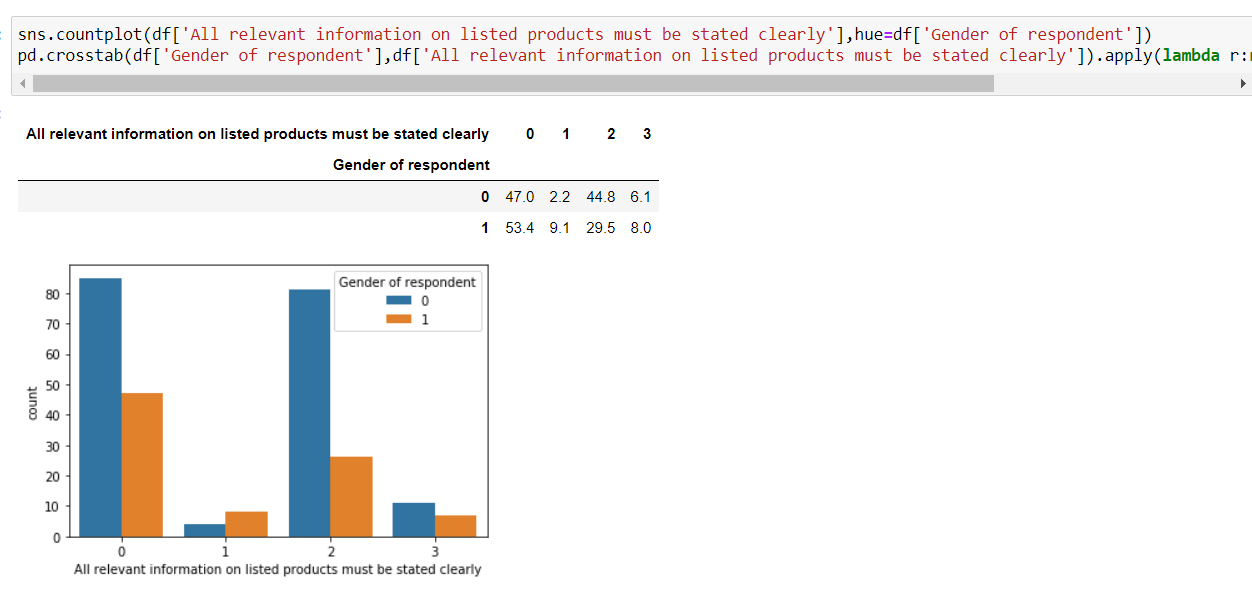
From the above distributed plot, we got understand how data are distributed for the purchase that was done for the past one year. From this we got to know that people mostly 31 to 40 times in the last one year and out of which 75 percent are female customers.



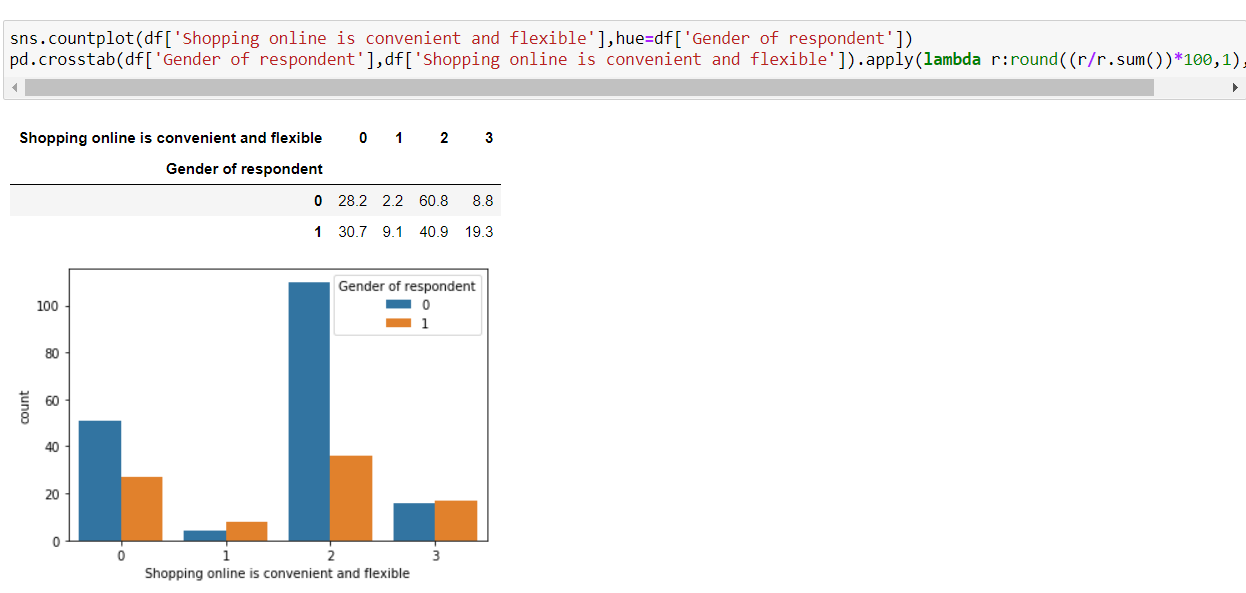
From the above box plot we find out the pin code location of the customers and more than 95 percentage of customers belong to North India. (NCR region).



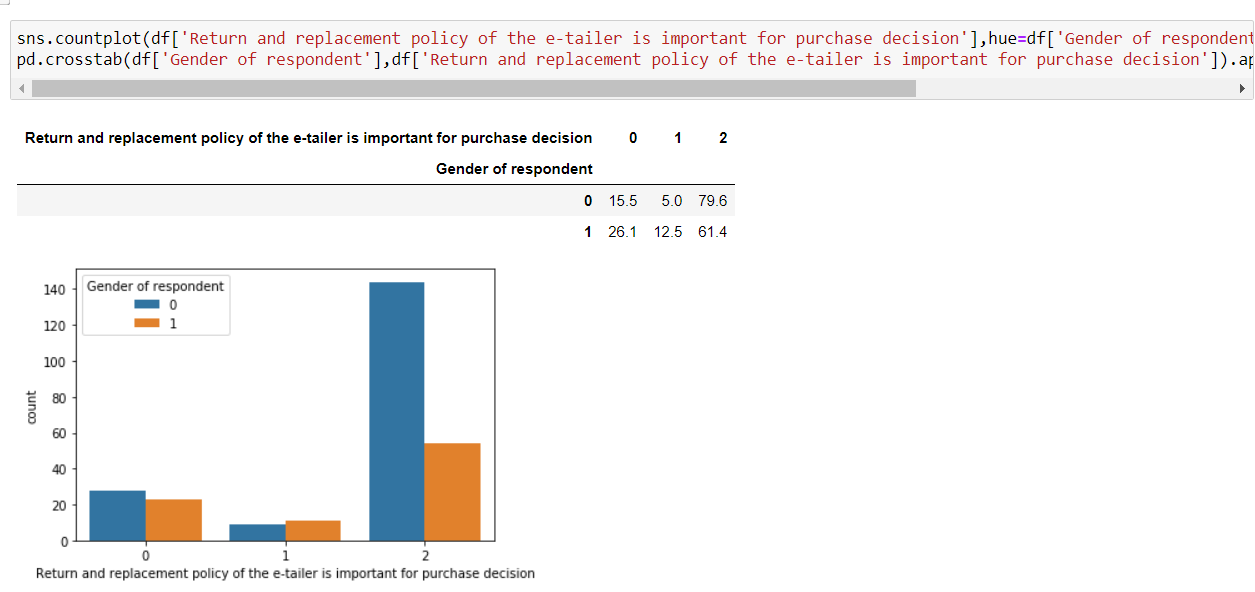
From the above distribution plot using the seaborn library we find how long they are shopping online, and we got know 35 percentage of the overall dataset of customers are shopping online for more than 4 years and out of which more than 70 percent are female customers. So, ease of access and convenient are one of primary reason to shop online.



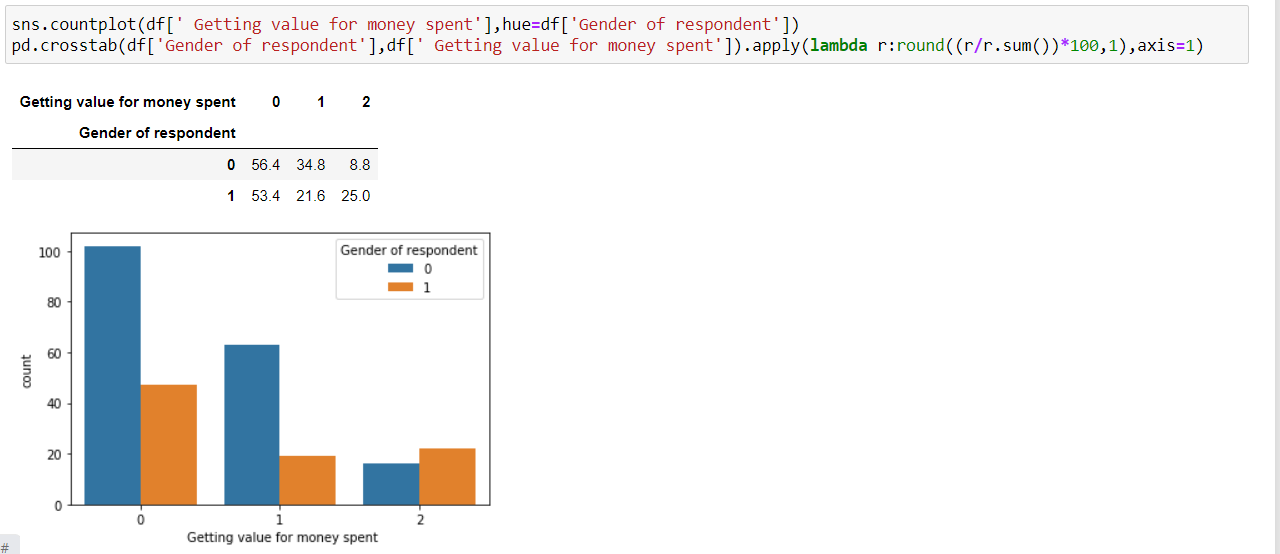
From the above visualization we understand while buying online people do read product listed information before taking the decision whether to go ahead with buying or not, so more than 45% of the shoppers agreed that information should be precisely defined which help us understand the customer knowing the product better.



From the above count plot, we got to know shopping online is convenient and flexible and more than 18 percentage of the people strongly agreed for it.



Return a product is an import aspect from any eCommerce platform and mostly customer expect a hassle-free return and more than 70 percentage of customer strongly agreed for it for this sample dataset.



From the above count plot we got the value for money aspect for the purchase that is made online and more than 50 percentage of the people agreed for it.

* Interpretation of the Results
* I have used various visualization tool to understand the data in a better way.
* I have used a few methods for finding the missing value, finding the total shape of the dataset and also used describe method for five-point summary analysis.
* I have also used label encoder technique and convert all the data into numerical form to do the data analysis in an easier way.

**CONCLUSION**

* 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 eCommerce 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.