



## **CUSTOMER RETENTION PROJECT**

Submitted by:

Shreya Jain

## ACKNOWLEDGMENT

I would like to thank Flip Robo Technologies for providing me with the opportunity to work on this project from which I have learned a lot. I am also grateful to Miss Khushboo Garg for her constant guidance and support. Some of the reference sources are as follows:

- Internet
- Coding Ninjas
- Medium.com
- Analytics Vidhya
- StackOverflow

***Special thanks to my point of contact Sajid Choudhary for solving all my doubts.***

## INTRODUCTION

- **Business Problem Framing**

Customer retention refers to the ability of a company or product to retain its customers over some specified period. High customer retention means customers of the product or business tend to return to, continue to buy or in some other way not defect to another product or business, or to non-use entirely. Selling organizations generally attempt to reduce customer defections. Customer retention starts with the first contact an organization has with a customer and continues throughout the entire lifetime of a relationship and successful retention efforts take this entire lifecycle into account. A company's ability to attract and retain new customers is related not only to its product or services, but also to the way it services its existing customers, the value the customers actually perceive as a result of utilizing the solutions, and the reputation it creates within and across the marketplace.

- **Conceptual Background of the Domain Problem**

All online e-commerce marketplaces encounter some amount of customer churn. While data analytics provides metrics on conversions and sales, the success of your e-commerce business also depends on your churn rate, which means the rate at which customers leave or cancel subscriptions.

The key to success for an e-commerce store lies in keeping the churn rate down, as on an average, the cost of attracting new customers is 6-7 times more than retaining existing ones. Once a customer buys on an e-commerce website there's a possibility that he/she would visit again, if satisfied with your products and service offerings.

All e-commerce platforms should consider implementing a churn model to add value to their businesses as it is a bare essential component for customer retention. Customer retention is the ability of a company to retain its customers over a specified period of time.

Although it's good to acquire new customers, existing customers bring in far more revenue in comparison. Another crucial aspect that underlines the dire need of retaining loyal customers is that they foster conversions by word of mouth publicity and referrals.

Churn model is the most commonly used data science technique to identify customers who are most likely to switch loyalties to another e-commerce website. The churn model calculates the churn rate by using different metrics depending on the nature of business. Primarily, there are three main variables that determine the churn rate: a specified period of time (month, quarter, or year), number of the customers at the beginning of the period, number of customers lost during the period.

- **Review of Literature**

The significance of customer retention was first quantified by Reichheld and Sasser (1990). They found that profits in service industries, including credit card companies, increased in direct proportion to the length of a customer's relationship. They noted the experience of MBNA America, citing its "customer defection 'swat' team staffed by some of the company's best telemarketers," which achieved a 50% success rate in persuading customers to retain their credit cards. At MBNA, a 5% improvement in customer retention increased average customer value by 125%. Reichheld and Sasser (1990) concluded that cutting defections in half could more than double the growth rate of the average company.

There is a great amount of literature about customer loyalty, customer satisfaction etc. But not many of researches are focus on the specific topic of customer retention. Customer retention should be a bigger topic. And also there are quite many different characters in the context of E-commerce compare to the regular businesses. So, this is the motivation and purpose of this thesis is to gain better understanding of how Ecommerce business should retain their customer.

- **Motivation for the Problem Undertaken**

Based on the empirical studies reviewed above, it is hypothesised that customer-perceived value, corporate image, and service quality influence customer retention through customer satisfaction, while switching barriers have a direct effect on customer retention. It is further hypothesised that customer satisfaction affects customer retention through customer trust. The conceptual framework is presentation.

## **Analytical Problem Framing**

- **Mathematical/ Analytical Modeling of the Problem**

Some of the columns have duplicate categories like, Mobile internet and mobile internet, one of the columns has categories 41 and above and 42 and above which need to be combined and made 41 and above.

- **Data Sources and their formats**

Data source were in Categorical Data. With String as the Datatype of all the columns.

There were 71 columns and few new columns were created for model prediction.

- **Data Pre-processing Done**

- Finding Missing Values
  - No missing Values found in the Data.
- Finding outliers
  - As all the columns are categorical columns, outliers in these columns would not be considered as outliers.
- Label Encoding of all the categorical Data with Label Encoder.

- Feature Selection with PCA, to remove noise component.

- **State the set of assumptions (if any) related to the problem under consideration**

Assumed that the problem statement is to find, which online retail store would a customers recommend, this is a multi-output prediction problem where a customer can recommend more than one retail store and at least 1 store.

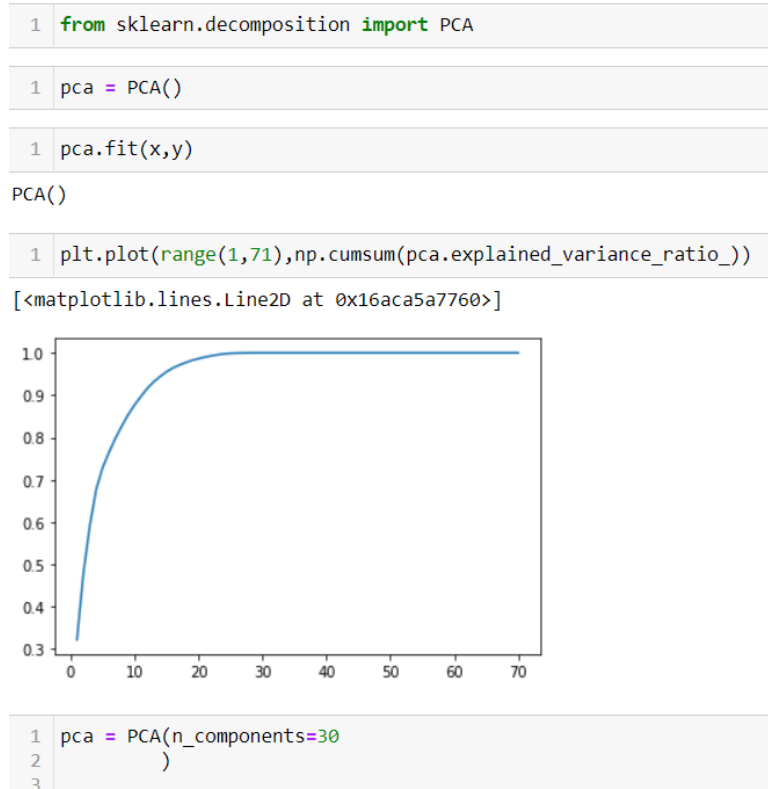
- **and Tools Used**

- Python Jupyter notebook for model visaulization to model preiction.
- Libraries used are pandas, numpy, sklearn.preprocssing , sklearn.model\_selection, sklearn.emsemble, seaborn, matplotlib.

## **Model/s Development and Evaluation**

- **Identification of possible problem-solving approaches (methods)**
- Data Visualisation to understand patterns with the Features.

- Identifying insights from the visualization that influence the outputs.
- Correlation table to understand correlation.
- PCA to understand how many features required for model prediction, out of 71 features, upto 30 features are giving 100% information about the targets, telling us that noise component is high and we need PCA for feature selection.



- Multi Class Problem.

- **Testing of Identified Approaches (Algorithms)**

Ensembling techniques used as multi class model evaluation.

- **Run and Evaluate selected models**

RANDOM FOREST \_ MULTI CLASS OUTPUT PREDICTION

Evaluation metric: Accuracy Score.

Training metrics: Cross validation score.

Validation Score with Accuracy Score: 1.0

Training Score with Cross Validation Score: [1. 1. 1. 1. 1.]

CLASSIFICATION REPORT				precision	recall	f1-score	support
0	1.00	1.00	1.00	1.00	64		
1	1.00	1.00	1.00	1.00	37		
2	1.00	1.00	1.00	1.00	29		
3	1.00	1.00	1.00	1.00	13		
4	1.00	1.00	1.00	1.00	3		
micro avg	1.00	1.00	1.00	1.00	146		
macro avg	1.00	1.00	1.00	1.00	146		
weighted avg	1.00	1.00	1.00	1.00	146		
samples avg	1.00	1.00	1.00	1.00	146		

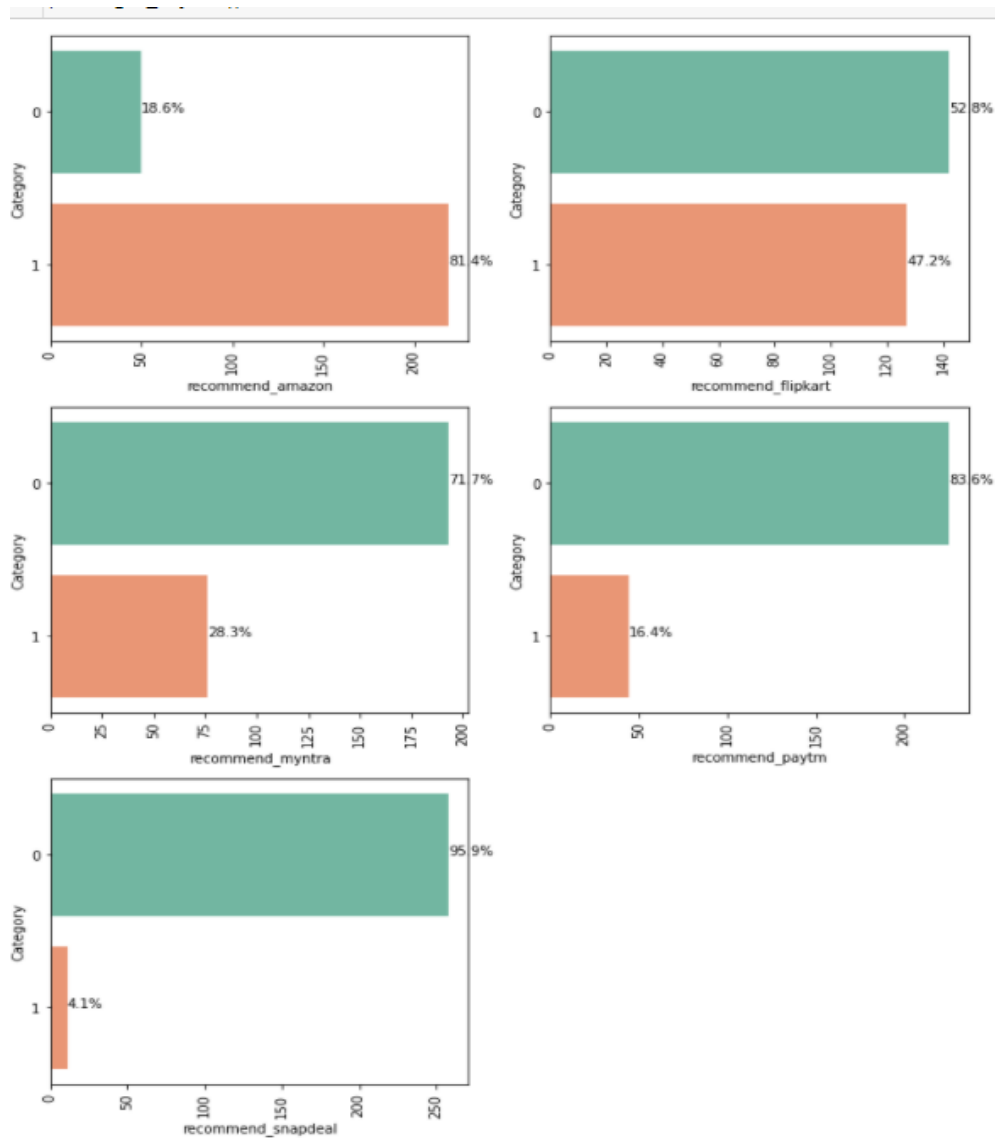
- **Key Metrics for success in solving problem under consideration**

This problem statement is more involved in data visualization of all the parameters, For model building with various ml algorithms are Accuracy Score for validation score and training score metrics used is Cross Validation Score . We Can see the Precision and Recall Score in the Classification Report Above.

- **Visualizations**

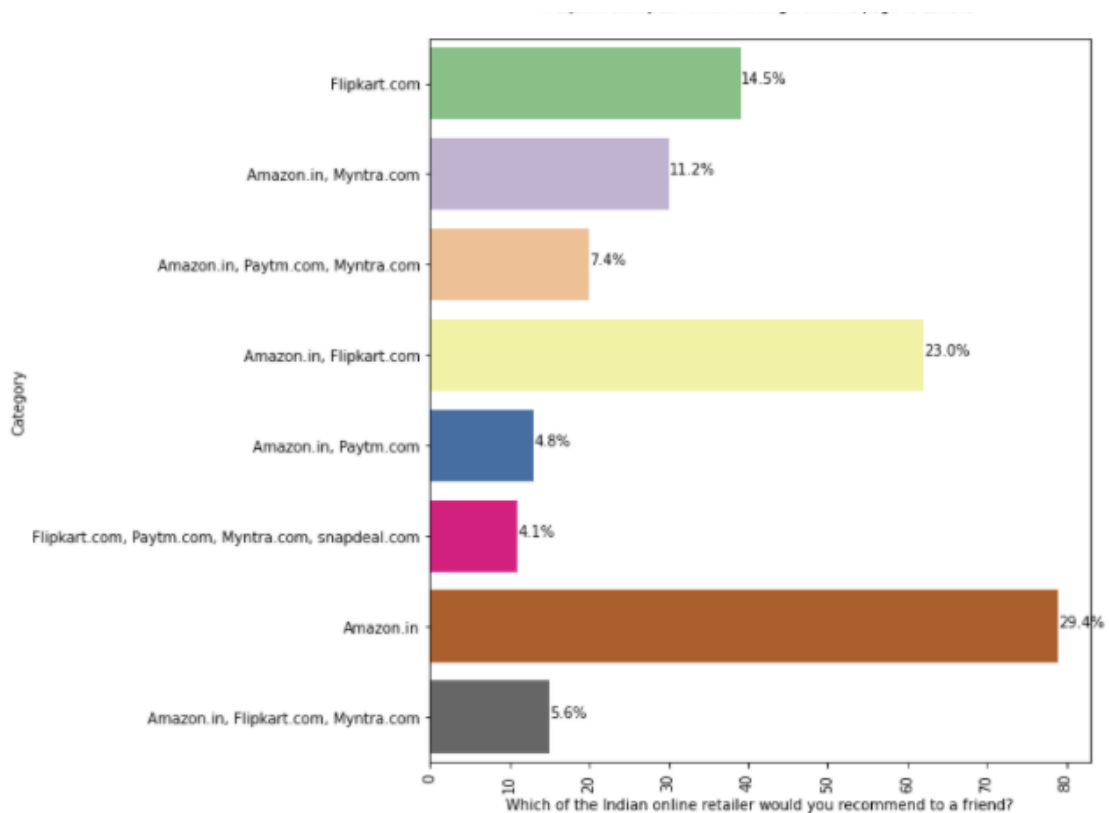


- It shows that there are no null values in the dataset.
- Some insights from visualization about which online retailer is suggested by customers.



## AMAZON IS THE MOST RECOMMENDED WEBSITE BY CUSTOMERS

Amazon.in	29.368030 %
Amazon.in, Flipkart.com	23.048327 %
Flipkart.com	14.498141 %
Amazon.in, Myntra.com	11.152416 %
Amazon.in, Paytm.com, Myntra.com	7.434944 %
Amazon.in, Flipkart.com, Myntra.com	5.576208 %
Amazon.in, Paytm.com	4.832714 %
Flipkart.com, Paytm.com, Myntra.com, snapdeal.com	4.089219 %



We can observe, amazon is most recommended by the customers, then flipkart.

- **Interpretation of the Results**

- myntra and paytm each take longer time to load by 22% of the customers each.
- limited mode of payment for most products, 32% customers find limited payment method in snapdeal, only 9% find limited payments in paytm.
- Longer delivery period, 26% of the customers find PAYTM to take longer time to deliver, 23% of the customers find snapdeal to take longer time.
- Website is as efficient as before, 36% of the customers find the amazon website still efficient as before, 16.7% of the customers go with flip kart and amazon both, least customers % for only Snapdeal 7%.
- OS Of the device used for shopping by customers is 45% Windows Os, then 31% with android and 23% with MAC IOS.
- Web browser used to access the website by customers is 80% Chrome, then 15% Safari, then Opera 3.9% and Least used Browser is Mozilla with 1.9%.
- Which Channel do you follow to arrive to the website for the first time, 85.5% of the customers came to the website through search engine, and others are through advertisement displays and content marketing which comprises of around 7% of the customers each.
- After the First Visit how to reach the online retail store, 32% of the customers reach the retail store via search engine and via application each, then 26% of the customers reach the store via direct URL, and at the least 3% of the Customers reach

via Social Media, and other 7% of the Customer reach through email.

- The factors have positive impact on customer retention are Reliability, Responsiveness, Contact Points, Convenience, Merchandising, Site Design, Security, Serviceability, Contact Interactivity, Customization, Care, Community
- How Much time spend in the e-retail store before making a purchase, 45% of the Customers take more than 15 mins before making a purchase, then 26% of the customers take 6-10 mins and 17% of the customers take 11-15mins before making a purchase. and least percent of customers take less than a min, 1-5 mins before making a purchase are 5-6% in percent of customers.
- Payment Option: 55% of the customers prefer credit/Debit card and AS Payment Option, then 28% of the customer prefer COD and other 16% of the customers prefer e-wallets payment option.

## CONCLUSION

- **Key Findings and Conclusions of the Study**

Key Finding about the problem statement from the survey from customers that which online retailer is recommended on majority basis. The Survey can be used for future prediction based on their likes and dislikes, kind of products they like, according to that which online retailer website should be recommended to a particular customer.

The Study showed Amazon is a website where you can find wide variety of products in their website, and also majority of the products are on offer and contains all the payment methods, reliable, speedy delivery and is the most recommended online retailer. After Amazon, Flipkart is recommended, then Myntra, then Paytm and Snapdeal is the Least recommended website.

- **Learning Outcomes of the Study in respect of Data Science**

Customer Retention Analysis deals with the survey from customers to know which online retailer customers like according various parameters and which online retailers will face customer retention. As online retailer recommended to other(friends/family) can be one or more than one online retail websites, which tells us a particular data row can have one class target or more, which is Multi Class Classifier problems. For Multi Class problems we can use ensembling techniques, we have used Random Forest as all the techniques are giving 100% accuracy.

In the survey we asked the customers what they think these factors could affect their willing to come back to this website to purchase again. Most of these factors have a very positive impact. Most of the factors influencing E-satisfaction and E-service quality got high agree rate. The highest is the factor of contact points, it has a 92.3% agree rate. All together 8 factors for E-satisfaction and E-service quality 5 of them got more than 90 % and also most of the factors are support both of them in the concept framework. So E-satisfaction and the E-service quality are quite important for the online customer retention. Especially the item of E-Service quality the 5 factors which have more than 90% agree rate contribute to it. Also between these two constructs E-service quality has a very strong influence on the E-satisfaction. The ability to fulfill the customer's needs could bring the customer's satisfaction.

- **Limitations of this work and Scope for Future Work**

During the research process of this thesis some more interesting research areas are come out. The relationship between E-satisfaction, E-service quality, E-loyalty and Switch barriers. How exactly they affect each other. I have some little discussion in this thesis but it still need further investigate and prove. How to use member community to improve customer relationship. Switching barrier design strategy