



CUSTOMER RETENTION PROJECT

Submitted by:
Sumitra

ACKNOWLEDGEMENT

The sources that helped me out in completion of the project are, I usually take reference from my personal notes, internet sources and some websites and all these sources helped me out to complete this project.

Introduction

Business problem Framing

As customers are the backbone of any business, business without customers would not be able to sustain their performance. This is because businesses are believed to have no revenues, no profits and therefore no market value. Accordingly, managing customers is deemed to be very crucial business agenda in which the key focus has been switched in recent years from attracting new customers to preserving existing ones. Every practitioner has discovered that it is much easier and cheaper to retain the existing customers than investing on the potential customers. A decent customer retention level is believed to be a significant contributor towards improvement in the overall business performance. A survey of forty (40) customers showed that customer service and quality of product have significant influence on customer retention which in turn has direct impact on business performance. Hence, it is advisable for businesses to pursue an effective and efficient customer service and ensure high and consistent quality of their product.

Conceptual Background of the Problem

Customer retention is the collection of activities a business uses to

increase the number of repeat customers and to increase the profitability of each existing customer.

Customer retention strategies enable you to both provide and extract more value from your existing customer base. You want to ensure the customers you worked so hard to acquire stay with you, have a great customer experience, and continue to get value from your products.

Review of Literature Retention analysis is the process of analyzing user metrics to understand how and why customers churn. Retention analysis is key to gain insights on how to maintain a profitable customer base by improving retention and new user acquisition rates.

How customer retention analysis works

Conducting a customer retention analysis is much more than calculating the churn rates. To get a clear picture, we need to figure out where, how and why our customers leave.

Identify where customers churn

An efficient retention analysis can help us understand at which point in the journey, our customers are more likely to churn.

By studying our existing customers' behavior and the retention rate fluctuations, we can find out which stage of the new user journey is highly volatile.

Once we have this information, we can adjust the marketing strategy to improve the retention efforts at the most vulnerable part of the customer's lifecycle and increase the number of users that stick with our product.

The key to identifying the customers churn is the continuous analysis of their behavior, right from onboarding, with retention and churn rates in mind.

Why customers churn

Retention analysis shows us the number of customers who churn and why. By doing this, we can identify behaviors associated with churning and acquire direct feedback.

Figuring out when customers leave us, can help us identify weak spots in our strategy. Meanwhile, feedback offers a direct explanation of why a customer decides to churn.

By adjusting our business strategy to battle the common reasons for churning, we can increase retention rates quickly and efficiently.

Find out how to improve retention A

smart customer retention analysis allows us to understand what we've done wrong. Knowing where and why customers churn can help you single out weak parts of your marketing and retention strategies and fix them timely.

A variety of retention improvement tactics exist. Many companies make a mistake of employing them without analyzing the situation first. With retention analysis results in hand, we can pinpoint strategies that work specifically for your business.

How to conduct retention analysis The

key to getting the most of our retention analysis is conducting it regularly. Customer behavior changes over time and so does our product offering. Seeing the full picture is only possible when we analyze customer cohorts consistently and continuously.

Calculate retention rate

Customer retention rate is the starting point of any retention analysis. We need to know how many customers are staying and how many are leaving. Even if the rate is satisfactory, we still need the analysis to continue keeping it this way.

Motivation of the problem undertaken

My only objective behind this project is to find out the solution for the customer retention in any business project. The motivation that comes is with helping the industry to come out of this problem, by providing them a complete solution to their customer retention

Analytical Problem Framing

- **Mathematical /Analytics Modeling of the Problem** Mathematical /Analytical functions and methods are used such as `df.shape ()` to check out the number of rows and columns in the dataset. `df.dtypes()` is used to check out the data type of the columns we have in the dataset. `df.isnull()` is used to check whether null values are present or not, we have also checked it by using Heatmap from seaborn. And the other one is checking out the statistical summary of the dataset by using `df.describe()` method. We got lots of fresh information from **statistical summary** is like:

There is certain difference in mean and median which shows that outliers are present.

There is little difference in 75% percentile and max column.

We can see the large difference in 75% percentile and max in columns like `city_pincode`, `browser_used`, `channel follow` and `abandon`.

We can see the little difference in columns like `device used`, `ph_scr_size`, `payment option`, `content`, `info_similar_pr` etc.

- **Data Sources and their formats**

The data source is private, the datafile is provided in `xlsx`. format,

We have imported the dataset in the file, the dataset has 269 rows and 71 columns. We can see the snapshot of the dataset below:

	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?
0	0	3	Delhi	110009	5	4	4	3
1	1	2	Delhi	110030	5	5	2	1
2	1	2	Greater Noida	201308	4	5	3	1
3	0	2	Karnal	132001	4	1	3	1
4	1	2	Bangalore	530068	3	2	2	1
...
264	1	2	Solan	173212	2	1	3	1
265	1	3	Ghaziabad	201008	2	4	3	1
266	1	4	Bangalore	560010	3	1	3	2
267	1	1	Solan	173229	3	1	2	1
268	1	4	Ghaziabad	201009	3	4	3	1

269 rows × 9 columns

- Data Preprocessing Done

Following are the steps followed for data cleaning:

Missing Values if we have missing values then treat them by using suitable method, but we have no missing data in our dataset.

Label Encoding, we have used label encoding on the columns which were in string format.

Removing the outliers, we have few outliers present in the dataset, we can remove these outliers by using zscore or IQR, but as we have few outliers, we can ignore them.

Removing the highly correlated columns, we have from the correlation matrix that few columns are highly negatively correlated, thus we have dropped those columns.

Removing Skewness, we have observed that skewness is present in many columns, So we have removed the skewness by using power transform method.

- **Data Inputs-Logic–Output Relationships**

The input columns of the data frame express the choice of customer in each field according to which we give the output to them. As in this project we recommend an online shopping website to a friend according to all his choices.

- **Hardware and Software Requirements and Tools used**

The hardware device used is only the pc. In software devices so many applications and libraries are used to for the completion of the project. I have used Jupyter software for using jupyter notebook. The libraries used are pandas, NumPy, matplotlib, seaborn these libraries are used for analysis purpose. And from sklearn power transform is used for removing skewness.

CONCLUSION

- • Key findings

Till here we have done all the Data Preprocessing we have seen the outliers present in the dataset, skewness present in many columns.

In Data Cleaning we have removed the skewness.

The data is ready for training testing and model development.