PROJECT REPORT- Customer Retention Case study

Richard Prabhakar

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. The data is collected from the Indian online shoppers. Results indicate the e-retail success factors, which are very much critical for customer satisfaction.

PROJECT REPORT-Customer Retention Case study

Problem statement.

* This dataset comprises of respondents from online shoppers , they have shared thier shopping experience.

* Using the combined responses of theyr experience we are going to see the utilitarian as well as the hedonistic value weightage.

* Since we do not have a target variable we will assume that the higher values in both the utilitarian as wel as the hedonistic value and feature which compliments that will increase customer retention

* We will also find out which group of the respondents fall into the category of having highest customer retention

Steps :

- check the no of rows and columns
- check the names of the columns
- Checking for null values
- Check to see the data type if numeric or not
- Draw graph to visualize if null values present
- Checking to see the mean and std deviation values of all numeric columns as we as the 25%, 50%, and 75% quanile value
    - We see that the most of the columns are encoded with values 1 to 5
    - 5 is highest and 1 is lowest
    - Plotting a graph to see if there is any deviation in the data or extreme values
    - Found all the values with the constraint on 0 to 10000
- Printing all the unique values within each column
    - *Columns which have numeric values are in range of 1 to 5 , Most of them are ratings for questions based on the type of common feature in online shopping
    - *5 is the highest while 1 is the lowest and 3 is indifferent whch means they are not againest it or for it
- Splitting the columns into numeric features as well as categorical features

Libraries used

pandas as pd

numpy as np

matplotlib.pyplot as plt

import seaborn as sns

%matplotlib inline

DATA SET :

Cust_retention_unencode.csv

Cust_retention_encode.csv

CONTENTS:-

269 rows containing responses from Online shoppers.

71 columns:-

- 1to 18 has user specific feature
- 18 to 47 has platform specific data .
- The other columns have user experience by way of “ratings:
- 5 highest and 1 is lowest

- We have 46 numerical features
- We have 25 categorical features

- Visualization of the data
  - Plotting count plot as well as the number of values in the data
    - We see the max of users are females
    - We see that people between 21 to 40 are the highest age group
    - We see most of user are from Delhi
    - Most of the users have been shopping online for more than 4 years
    - And they have purchased more than 10 times in the past 1 year
    - Most users have mobile internet to use the online shopping platform
    - Most users are iphone or Mac users
    - Most users have 5.5 inches or more devices
    - Majority of users use google chrome
    - Most of the users use search engine to arrive at the product or platform website
    - Most reach the product page and also the app page after the search
    - Most spend more than 15 min in page
    - Most pay with debit or credit card
  - After checking user features we move to actions on page
    - We see most leave the page after checkout sometimes
    - We see they find better offer s they abandon
  - Then rating columns begin(we will touch on most but as we see they all are choosing that all the features are very important for customer retention
  - 18 The content on the website must be easy to read and understand
  - Strongly agree (5) 164
  - 19 Information on similar product to the one highlighted is important for product comparison
  - Strongly agree (5) 116
  - 20 Complete information on listed seller and product being offered is important for purchase decision
  - Agree (4) 101
  - 21 All relevant information on listed products must be stated clearly
  - Agree (4) 132
  - 22 Ease of navigation in website
  - Strongly agree (5) 141
  - 23 Loading and processing speed
  - Strongly agree (5) 115
  - 24 User friendly Interface of the website
  - Strongly agree (5) 189
  - 25 Convenient Payment methods
  - Strongly agree (5) 159
  - 26 Trust that the online retail store will fulfill its part of the transaction at the stipulated time
  - Strongly agree (5) 141
  - 27 Empathy (readiness to assist with queries) towards the customers
  - Strongly agree (5) 194
  - 28 Being able to guarantee the privacy of the customer
  - Strongly agree (5) 185
  - 29 Responsiveness, availability of several communication channels (email, online rep, twitte r, phone etc.)
  - Strongly agree (5) 149

- 30 Online shopping gives monetary benefit and discounts
- Strongly agree (5) 105
- 31 Enjoyment is derived from shopping online
- Strongly agree (5) 86
- 32 Shopping online is convenient and flexible
- Strongly agree (5) 146
- 28 Being able to guarantee the privacy of the customer
- Strongly agree (5) 185
- 34 Gaining access to loyalty programs is a benefit of shopping online
- Strongly agree (5) 115
- 35 Displaying quality Information on the website improves satisfaction of customers
- Strongly agree (5) 133
- 36 User derive satisfaction while shopping on a good quality website or application
- Strongly agree (5) 175
- 37 Net Benefit derived from shopping online can lead to users satisfaction
- Strongly agree (5) 164
- 38 User satisfaction cannot exist without trust
- Strongly agree (5) 122
- 39 Offering a wide variety of listed product in several category
- Strongly agree (5) 111
- 40 Provision of complete and relevant product information
- Strongly agree (5) 135
- 41 Monetary savings
- Strongly agree (5) 148
- We see that the rest of columns show similar traits so we plot a gragh collectively to show the Columns

- Instantiate the rating columns in object : df_new
- Insatiate the rating columns which are encoded to object df_new2
- Plotting a countplot again to see the graph for the rating columns
- Creating a function to check the values in all the columns
- Creating another object for all ratings column as rating_col
- Checking correlation between features
- Creating heatmap to understand the correlation better
- Creating pivot table to see relationship between the user features as well as the ratings columns
- Creating heatmap to see relationship between the ratings columns
- Plotting a distplot to see the spread , find similar findings
- Plotting box plot to see outliers and find all data within 4 to 5
- Plotting Stripplot and see the most density in 5 ratings

Conclusion:-

- We have done Exploratory Data Analysis , In our analysis we have seen -
- There are 71 columns and 269 rows of data.
- We have seen that the values from 1 to 17 shows the characteristics of the user or respondents
- We see that the columns 18 to 47 show rating of the features of the online shopping experience.
- We see that the rest columns form 48 to 71 show the similar features of the user experience
- we have plotted countplot,barplot,stripplot,boxplot
- we have done correlation matrix with the whole data as well as the ratings columns alone

- we have done correlation matrix with the whole data as well as the ratings columns alone
- we did feature engineering by checking nulls as well as making separate objects for rating and non-numeric col
- we have plotted pivot table to check the relationship btw the user characteristics as well as the rating columns

We see that from our analysis the maximum no of users feel that all the features are very important in their online experience, Be it Utilitarian value and hedonistic, all of them have been selected by the respondents as the highest importance