Project Report : Indian-e-commerce Customer_Retention

#Project title,problem statement:

Project Title : Customer_Retention

Problem Statement:

   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.

3.steps to solve the dataset:

->Loading the essential libraries and then loading and reading the dataset.

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

import warnings

warnings.filterwarnings('ignore')

df=pd.read_excel("customer_retention_dataset.xlsx")

df

3)-Visualizing the dataset using charts, graphs, etc., to understand the data better.

   .Lineplot

   .Violinplot

   .Countplot

   .Stripplot

-Pre-processing the data using various manipulation techniques.

   .Separating the label from rest of the features

   .Encoding Categorical Features

   .using ordinal encoder for independent features

   .Scaling

Using various feature selection method to see which feature affects the most:

   .Using Feature importance of random forrest

   .plot graph of feature importances for better visualization

Using chi2 test: firstly we will import SelectKBest, chi2 and secondly fitting (x,y)after then

   .naming the dataframe columns

   .print10 best features

Identifying the best parameters for the classification models using hyperparameter tuning.

.Hyperparameter Tuning

.importing RandomizedSearchCV

.After fitting this we found best_estimator,best_params,best_score(max_depth=20, min_samples_leaf=4, min_samples_split=4,n_estimators=700)

Applying machine learning algorithm to train the models and evaluate their performance using metrics.

.Modelling Phase

.importing model(Random forest classifier),train_test_split,acc_score,confu_matrix,class_report,roc_auc_score,roc_curve

3)-We will analyse and adress our dataset like this:

-->First step:

we will do univariate analysis:

-->Second step:

-analysis on the basis of the following factors:

'Gender of respondent',

'How old are you?',

'Which city do you shop online from?',

'What is the Pin Code of where you shop online from?',

'Since How Long You are Shopping Online ?',

'How many times you have made an online purchase in the past year?'

3)-we will analyse personal info using these

.Lineplot

.Violinplot

.Countplot

.Stripplot

.we will check 'Loyalty of Customers' using Countplot figure.

.we will analyse 'Brand Image' using pie.plot figure:

-In the end we will know that from which site the customers purchase most,do shopping most continue shopping and with which site they are most loyal for shopping.

-->Third step:

-Processing the dataframe

-using model(RandomForestClassifier)-we will use Random Forest Classifier model to find best accuracy score.

--Hyperparameter Tuning using random forest classifier-we will use hyperparameter tunning to find accurate accuracy.

4)Dataset name,links,rows and columns,given attributes,null values,columns types,dtypes-int/categorical.

Dataset name: Customer_retention

Dataset is available on open-source platforms like Kaggle or Github.

Dataset customer_retention consists of 269 rows and 71 columns in the dataset.

Here we can see the given attributes in the columns below:

1.Gender of respondent

=>male/female

2. How old are you?

=>21-30years, 31-40years, lessthan20, 41-50years.

3. Which city do you shop online from?

=>Bangalore,Gaziabad,Solan,Delhi,Greater Noida,Karnal.

4. What is the Pin Code of where you shop online from?

=>110009/110030/530068/560010

5. Since How Long You are Shopping Online ?

=>Above1-2,2-3,3-4,4years.

6. How many times you have made an online purchase in the past 1

year?

=>Less than 10,10-20,31-40,41.

7. How do you access the internet while shopping on-line?

=>Dial-up,wi-fi,mobile-internet.

8. Which device do you use to access the online shopping?

=>Desktop,Smartphone,Laptop.

9. What is the screen size of your mobile device?\t\t\t\t\t\t

=>4.7inches,5.5inches,others.

10. What is the operating system (OS) of your device?\t\t\t\t

=>IOS/Mac,Android,Windows/windows Mobile

Longer time to get logged in (promotion, sales period)

=>Amazon.in/Flipcart.com/Snapdeal,myntra.com

Longer time in displaying graphics and photos (promotion, sales period)

=>Amazon.in/Flipcart.com/Snapdeal,myntra.com,paytm.com

Late declaration of price (promotion, sales period)

=>Amazon.in/Flipcart.com/Snapdeal,myntra.com,paytm.com

Longer page loading time (promotion, sales period)

Limited mode of payment on most products (promotion, sales period)

Longer delivery period   Change in website/Application design

Frequent disruption when moving from one page to another

Website is as efficient as before

Which of the Indian online retailer would you recommend to a friend?

Dataset customer_retention consists of 269 rows and 71 columns.

There is no null value present in the dataset.

All the columns are of categorical types. There are no identifier or constant columns

All the columns are of object datatype except for pincode column which is of int type.

5)import pandas as pd

.Firstly, I downloaded   this   Customer_Retention file then extract the dataset after then drag the dataset here

->displaying the dataset:

df=pd.read_excel("customer_retention_dataset.xlsx")

df

6)We have taken some steps to adress the problem in my project:

1.Step: we have used pie.plot method to display in figure personal information of customer like:

'Gender of respondent',

'How old are you?',

'Which city do you shop online from?',

'What is the Pin Code of where you shop online from?',

'Since How Long You are Shopping Online ?',

'How many times you have made an online purchase in the past year?'

2.step: we have used Lineplot method to display in figure

.How many times customers have made an online purchase in the

past year?

.From which online retailers they have shopped ?

3.step: we have used violinplot method to display in figure

.From which online retailers they have shopped most?

.They feel gratification shopping on your favorite e-tailer

4.Step:we have used violinplot method to display in figure

.From which online retailers they have shopped most ?

.Gaining access to loyalty programs is a benefit of shopping online

5.step: We have used countplot method to display in figure

.Since How Long You are Shopping Online ?(you-customers)

.How old are you?

6.step:Converting Years to numbers for better analysis

7.step:we have used lineplot method to display in figure

.Which city do you shop online from?

.Average years of shopping online

.Gender of respondent

8.step:we have used countplot method to display in figure

.Since How Long You are Shopping Online

.After first visit, how do you reach the online retail store?

9.step:we have used plot.pie method to display in figure

.Brand image:

performance=['Easy to use website or application',

'Visual appealing web-page layout', 'Wild variety of product on offer',

'Complete, relevant description information of products',

'Fast loading website speed of website and application',

'Reliability of the website or application',

'Quickness to complete purchase',

'Availability of several payment options', 'Speedy order delivery',

'Privacy of customers' information',

'Security of customer financial information',

'Perceived Trustworthiness',

'Presence of online assistance through multi-channel'])

.which websites have maintained a very good brand image

10.step: we have used stripplot method to display in figure

.Why did you abandon the "Bag", "Shopping Cart

.From the following, tick any (or all) of the online retailers you have shopped from

11.step: we will check Loyalty of customers even though these bad quality can be seen in website services

.Collecting all the negative remarks about a brand

bad=['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)',

'Limited mode of payment on most products (promotion, sales period)',

'Longer delivery period', 'Change in website/Application design',

'Frequent disruption when moving from one page to another']

-we have used countplot method to display in figure

.Which of the Indian online retailer would you recommend to a friend?

7)We have taken some steps to adress the problem and found these results after applying these steps in my project :

1.Step: we have used pie.plot method to display in figure personal information of customer like:

'Gender of respondent',

'How old are you?',

'Which city do you shop online from?',

'What is the Pin Code of where you shop online from?',

'Since How Long You are Shopping Online ?',

'How many times you have made an online purchase in the past year?'

-we found this result in figure after using Pie.plot method:

result: There is double the number of women than men who have taken this survey.

-Most of the people are in their 30's followed by 20's, teenagers and senior citizen are the least in number.

-Most of the people belong from delhi, noida and banglore, ambiguity can also be seen as noida has two categories

(noida and grater noida) which need to be handled -Most of the people shopping online have been shopping from a long time.

-Majority of people shop online 10 times a year, amiguity can also be seen for range 42 times and above which needs to be

handled.

2.step: we have used Lineplot method to display in figure

.How many times customers have made an online purchase in the

past year?

.From which online retailers they have shopped ?

-we found this result in figure after using method:

Result: -Heavy shoppers who shop more than 41 times a year shop from all the online brands, some of the people who shop

for 32-40 and less than 10 times a year seem to exclude myntra. People shop from Amazon and flipkart whatever be the case.

3.step: we have used violinplot method to display in figure

.From which online retailers they have shopped most?

.They feel gratification shopping on your favorite e-tailer

-we found this result in figure after using method:

Result: Almost all the people who have shopped from amazon, flipkart and paytm are satisfied.

-People who shop from a more number of online brands dosent seem to be satisfied.

4.Step:we have used violinplot method to display in figure

.From which online retailers they have shopped most ?

.Gaining access to loyalty programs is a benefit of shopping online

-we found this result in figure after using method:

Result: People shopping from amazon and paytm are getting benefits from the loyalty points,

-flipkart and sanpdeal also seem to give such benefits but people who shop from almost everywhere disagree with

this statement too.

5.step: We have used countplot method to display in figure

.Since How Long You are Shopping Online ?(you-customers)

.How old are you?

-we found this result in figure after using method:

Online Retailing Result: Highest number of people have been shopping online for above 4 years except for the age group below 20

years and above 50 years.People who are shopping online for 1-2 years does not include teenagers and elder people.

6.step:Converting Years to numbers for better analysis.

7.step:we have used lineplot method to display in figure

      .Which city do you shop online from?

      .Average years of shopping online

      .Gender of respondent

-we found this result in figure after using method:

Result:In lines, we can see that density of female customers is more than male.

-Men living in banglore and ghaziabad shop have shopped online for less than 1 year.

-Highest number of men shopping online belong from delhi and noida,

-while men from moradabad have been shopping online for the longest.

-Women from meerut and noida have shopped the longest.

8.step:we have used countplot method to display in figure

      .Since How Long You are Shopping Online

      .After first visit, how do you reach the online retail store?

-we found this result in figure after using method:

Result:Even though people who are shopping online for more than 3 years donot use the application rather use search engine and

-direct url's in large number which indicates that online brands should update all their platforms rather than just application.

9.step:we have used plot.pie method to display in figure

.Brand image:

performance=['Easy to use website or application',

'Visual appealing web-page layout', 'Wild variety of product on offer',

'Complete, relevant description information of products',

'Fast loading website speed of website and application',

'Reliability of the website or application',

'Quickness to complete purchase',

'Availability of several payment options', 'Speedy order delivery',

'Privacy of customers' information',

'Security of customer financial information',

'Perceived Trustworthiness',

'Presence of online assistance through multi-channel'])

.which websites have maintained a very good brand image

-we found this result in figure after using Pie.plot method:

Result:-Amazon, Flipkart have been had the highest votes for having all the positive points and have maintained a very good

brand image followed by paytm and the myntra.

10.step: we have used stripplot method to display in figure

.Why did you abandon the "Bag", "Shopping Cart

.From the following, tick any (or all) of the online retailers you have shopped from

-we found this result in figure after using method:

Result: We can clearly see that most of the time people abandon the bag is beacuse they get a better alternative offer

-or promo code not applicable. There is also lack of trust seen in amazon, flipkart and paytm by some people.

11.step: we will check Loyalty of customers even though these bad quality can be seen in website services

.Collecting all the negative remarks about a brand

bad=['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)',

'Limited mode of payment on most products (promotion, sales period)',

'Longer delivery period', 'Change in website/Application design',

'Frequent disruption when moving from one page to another']

-we have used countplot method to display in figure

.Which of the Indian online retailer would you recommend to a friend?

-we found this result in figure after using method:

Result: Customers seem to be more loyal to amazon, flipkart and paytm as even though many of them have given negative remarks

about them still they would recommend these platforms to their friend.

****

----> My reason to choose certain features over others:

   .I choose these methods for analysing my dataset Because these methods gives best and acurate result as we can see in

   figures which are showing us proper info and result of all about problems clearly.

8)Conclusion:

-The cost of the product, the reliability of the E-commerce company and the return policies all play an equally important role

-in deciding the buying behaviour of online customers. The cost is an important factor as it was the basic criteria used by

-online retailers to attract customers. The reliability of the E-commerce company is also important, as it is even required

-in offline retail. It is important because customers are paying online, so they need to be sure of security of the online

-transaction. The return policies are important because in online retail customer does not get to feel the product.

-Thus, he wants to be sure that it will be possible to return the product if he does not like it in real. Whereas,

-the logistics factor, which included Cash on delivery option, One day delivery and the quality of packaging plays a secondary

-role in this process though these are Must-be-quality. This is so because these all does not interfere with the real product

-and people believe that this is the basic value that E-commerce websites provide.

-All the websites were not equally preferred by online customers. Amazon was the most preferred followed by Flipkart.

-This can be explained easily by previous result that we got. These two companies are most trusted in the industry and hence,

-have a huge reliability. Also, the sellers listed on these websites are generally from Tier 1 cities as compared to Snapdeal

-and PayTM which have more sellers from tier 2 and 3 cities. Also, these websites have the most lenient return policies as

-compared to others and also the time required to process a return is low for these.