E-RETAIL CUSTOMER ACTIVATION AND RETENTION

Submitted by: Yashna Shah

ACKNOWLEDGMENT

I, would like to express my gratitude to my SME Sapna Verma , who gave me the opportunity to work on project **" E-RETAIL CUSTOMER ACTIVATION AND RETENTION"** and also guiding me in gaining in-depth knowledge of Data Science and Machine Learning in deriving insights for Organizational goals and meeting Business needs.

Business Problem Framing

E – Retail Customer Activation and Retention

A case study from Indian e-commerce customers. 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 contributing 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 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.

Predictions:

We need to predict the E-retail factors for customer activation and retention

Conceptual Background of the Domain 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. Loyalty programs, sometimes referred to a customer retention program, are an effective way to increase purchase frequency because they motivate customers to purchase more often in order to earn v

ABSTRACT

Customer Retention:

In today's challenging economy and competitive business world, retaining their customer base is critical to organization success. If the company doesn't give their customer some good reason to stay, organization's competitors will give the customer a reason to leave. Customer retention and customer satisfaction drive profits. Its far less expensive to cultivate organization existing customer base and sell more service to the customer than to seek new, single transaction customers. Most surveys across industries shows that keeping one existing customer is five to seven times more profitable than attracting one new customer. In this era of intense competition, it is very important for any service company to understand that merely acquiring customer is not sufficient because there is a direct link between customer retention over time and profitability and growth. Customer retention to a great extent depends on service quality and customer satisfaction. Service recovery is the process of putting things right after something goes wrong in the service delivery. Customer retention is the mirror image of customer defection.

E-retailing is the sale of goods and services through the internet. Amazon.com is by far the largest online retailer providing consumer products and subscriptions through its website. E-retail has become the need of the hour.

This project focuses on the key factors of customer activation and retention.
The given dataset needs to be analysed in order to get the insights from the same and to formulate the strategies towards the customer activation and retention.

Customer Retention is all about how well you are building relationships and drawing existing customers back for subsequent purchase.

Motivation for the problem undertaken

In this competitive world, retaining the customers has become important part and parcel of the business activity. We are required to analyse India's E-retail business survey by bifurcating them into Hedonic and Utilitarian values. Understanding various influential factors customers encounter during online purchase, and customers choice across various platforms which makes them to repeat purchases on a e-retail platform.

This study will help us to understand the customers need, preferences and and what they require from the service station and also help the industry to improve its service standards.

Analytical Problem Framing

Mathematical/ Analytical Modeling of the Problem

Mathematical models are important in selecting the right one to answer the business questions that can bring tremendous value to the organization. Mathematics plays a vital role in the latest technologies like Machine learning, Data Science, Artificial Intelligence and Deep learning because every algorithm that is built with these latest technologies has a mathematical function behind it and also caters in identifying patterns.

Statistics and Probability are considered as the key for implementing such algorithms.

Notions include:

Regression

Understanding of various distributions like:

Gaussian Distribution, Bernoulli Distribution, Binomial Distribution and Bayes Theorem.

Statistics is further divided into 2 branches namely,

1. Inferential Statistics: it includes
    - Sampling Distributions
    - Confidence Interval
    - Chi Square Test
    - ANOVA
2. Descriptive Statistics: it includes
    - Mean, Median and Mode
    - IQR Percentiles
    - Standard Deviation and Variance
    - Normal Distribution
    - Z Statistics and T Statistics
    - Correlation
    - Linear Regression

Data Sources and their formats

1. Data contains 269 entries each having 71 variables.
2. Dataset doesn't contain any null values.
3. Extensive EDA has been performed.
4. The given dataset contains single numeric variable and others being categorical variables, which we converted using Label Encoder.
5. In our dataset we can see many categorical variables, the target variable (Recommended Retailer) being a categorical variable, it is a Classification Problem.

Data Analysis and Pre-processing Done

Since, we have no null values in our dataset, our data is clean now and we have segregated the target column.

Data Inputs- Logic- Output Relationships

In Classification problem, the target variable should be discrete.

The steps include:

- Data Pre- processing and feature Engineering- which includes converting the data type using Label Encoder.
- Checking the correlation among the variables and plotting a correlation matrix for graphical representation.
- Correlation helps in feature selection and making the feature ready for the model.
- In this project, we can observe from the matrix that the highly positively correlated column is ‘getting value for the money spent’. The most negatively correlated column is ‘Longer time in displaying’

Hardware and Software Requirements and Tools Used

Hardware: Intel i5

Software : Jupyter Notebook (Anaconda 3)

Language : Python

Libraries :

1. Pandas - Used for loading the data and doing analysis
2. Numpy - Used for loading the data and doing analysis.
3. Matplotlib - Used for data visualization.
4. Seaborn - Also used for plotting library and is more advanced than matplotlib.
5. SkLearn - Used for data Pre processing
6. Scipy
7. Statsmodel - Provides functions for calculating mathematical statistics of numeric data.
8. Performance Metrics used : Accuracy, Confusion Matrix, Precision, Recall

Model/s Development and Evaluation

Identification of possible problem-solving approaches (methods)

- We had no null values in our dataset.
- Renamed all the column names
- Identified skewness and outliers in the dataset and removed the outliers with the help of z-score method.
- Splitting the dataset and scaling the data with Standard Scaler method.

Testing of Identified Approaches (Algorithms)

The algorithms used for model building and predictions are listed below:

1. Random Forest Classifier
2. Decision Tree
3. Logistic Regression

   These algorithms have been used for both training and testing the model and the model has been evaluated with the help of classification metrics like confusion matrix, precision and recall.

Run and Evaluate selected models

Logistic Regression: It is used to describe data and to explain the relationship between dependent and independent variable. It is a classification model, and is used when the dependent variable is categorical.

Random Forest Classifier: It is a classification algorithm which consists of many decision trees. It can perform both regression and classification tasks. Random Forest produces good predictions that can be understood easily. It searches for the best feature among a random subset of features.

Decision Tree Classifier: The main advantage is its ability to use different feature subsets and decision rules at different stages of classification. It often involves higher time to train the model. The tree can be explained by two entities, namely decision nodes and leaves.

Key Metrics for success in solving problem under consideration

Will go with Random Forest Classifier.

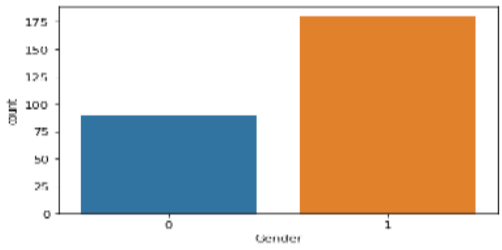
Reasons

1. Random Forest reduces overfitting in decision tree and helps to improve accuracy.
2. It is flexible for both classification and regression tasks.
3. It also works well with both categorical and continuous values.
4. It is a rule based approach.
5. It automates missing values present in the data.

Visualizations

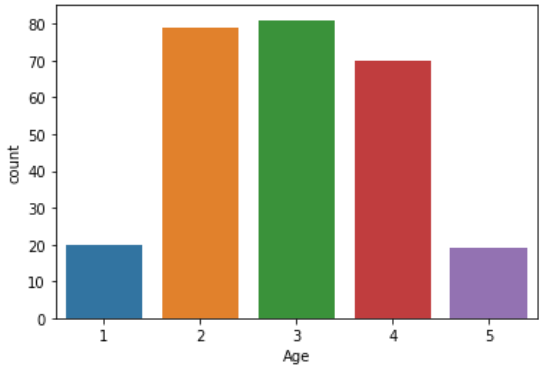
GENDER

Here we can see there are more no of female shoppers than male.



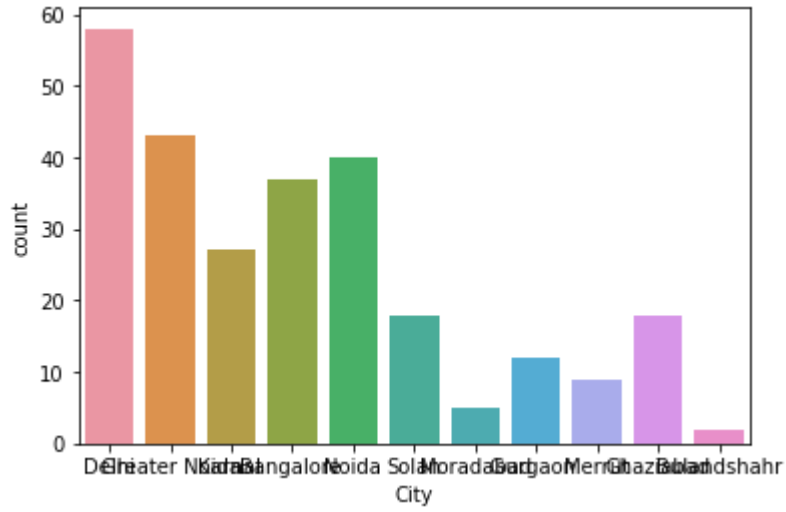
AGE

Maximum online shoppers are from the age group 31-40 Years. More than 75 percent of online shoppers lies between 21-50 years old. Numbers for less than 20 years and above 51 are less, which indicates these age group rarely use online shopping.



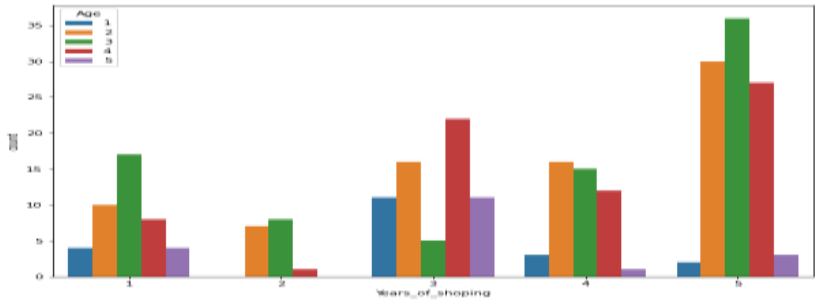
CITY

Delhi has highest number of customers who shop online. Graph also shows that most of the customers are from Delhi, Greater Noida, Noida, Banglore and Karnal.

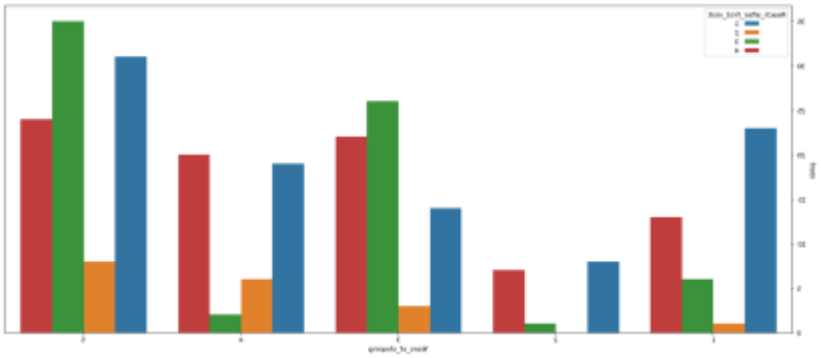


ONLINE RETAILING

Maximum no of people are shopping online for more than 4 years, excluding the age group of less than 20 years and above 50years.

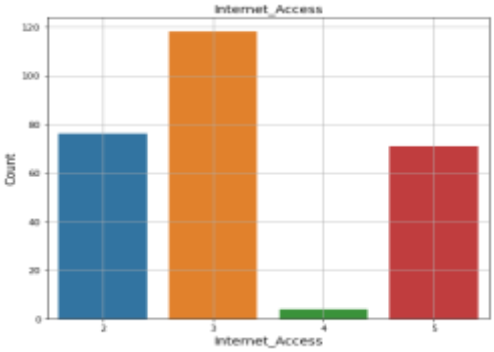


Here we can observe that people shopping more than 3 yrs use link,url but not the app for shopping. This shows that online brands should update their platforms.



INTERNET ACCESS

Majority of the shoppers are using Mobile internet for shopping online.



INFERENCES

1. We can observe that Amazon and Flipkart are the most preferred sites for shopping by the customers.
2. Amazon is considered as the most customer friendly website as compared to others.
3. Amazon is the most shopped website for years and customers use debit / credit card among other payment methods.
4. Amazon performs well in Customer Retention with utilitarian value.
5. The websites which perform well in customer retention are Amazon and Flipkart as compared to other websites.

CONCLUSION

- Female shoppers more than males overall
- Most shopping age is 31-40 years
- Most people are shopping from delhi in overall people shopping around 4 years
- People made purchase in past years in less than 10 times
- People mostly use mobile internet and smartphone buy from amazon
- Screen size of mobile is 5.5 and others category and has windows OS mostly
- People use google chrome and search engine channel to explore first mostly
- People purchase via application and direct url mostly .
- Customer use credit/debit cards mostly for their transactions in amazon
- They abandon purchase sometimes only
- They quit purchase because of better alternative offer and promo code if not available
- People strongly agree that the content of the website should be readable easily.
- People mostly strongly agree to the both hedonic and utilitarian values provided by the e-retail
- Amazon and flipkart provides both hedonic and utilitarian values to the customers to make activation and retention. And they do best among others here
- Myntra and paytm has longer page loading time
- Snapdeal and paytm has longer delivery speed
- Limited mode of payments in snapdeal
- Amazon changed its webpage to upgraded one
- Amazon is efficient in web design as befor