Customer Retention

Introduction – The Customer retention dataset is about to know what the customer think about online store or online product selling website. The dataset has feedback as well as suggestion of the customer like – what they like, what they don't like, what need to be change, what more will make store good, what is lacking and many more.

Understanding – We need to analyze the customer response from the answer and feedback given by the customer and give output to the online store admin team to make change accordingly. So, that the purchase and traffic will increase on the online store.

Step taken to analyzing the dataset

- Checking the dataset.
- Checking the size of the dataset which results 71 columns.
- Dataset has column which is questions. So, I will change the column names with alias name which help in EDA. Alias name always small and easy to read and understand.
- Checked the null values in the dataset but there are no null values. So, I don't need to perform any data filling technique.
- As the dataset has no null values missing values. So, I moved forward to the next step EDA.

Exploratory Data Analysis (EDA)

- As the most of the columns has categorical values. So, I start with counplot.
- Plotted the Countplot to check the counting of the response for particular columns.

Features

- **Gender** – Mostly female doing more shopping comparison to the Male.
- **Age**- The age between 21 -50 years public doing more shopping.
- **City** – Metro cities public doing more shopping comparison to other cities.
- **Shopping Since** – This feature has mixed result. I conclude that customer base increasing gradually every year.
- **Purchase in past 1 Year**- Majority of less than 10 item is more but the count for 30 -40 purchase is also good.
- **Internet service used** – Mostly customer used mobile or wi-fi network.
- **Device**-Smartphone is most commonly device to do shopping.
- **Screen Size of used device**- This has mixed answer most of the user did give answer but mostly used 5.5 inches s creen.
- **Device OS** – Android and windows has majority in term of OS
- **Browser** – Google Chrome is most commonly used browser used for shopping.
- **Channel which direct to website** – Search engine help more to reach store.
- **How to reach store** – Most common is search engine, app and direct URLS.
- **Time taken to make decision** –The time range to make decision is 6 -15 minutes.
- **Payment Option** – Customer preferred credit or debit card payment.
- **Abandon** – Sometimes people leave the website without any specific reason.
- **Reason of Abandon** – Best and better deal with any other online store.
- **what Content easy to read?** – Strongly Agreed
- **Information need to highlight for comparison**- Strongly Agreed
- **Important information of seller** - Strongly Agreed
- **Information need to state clearly**- Strongly Agreed
- **Ease of Negative**- Strongly Agreed
- **Speed** – Customer Strongly Agreed
- **Interface**- Customer Strongly Agreed that interface should be good

- **Convenient Payment option**- Customer Strongly Agreed that payment should be more convenient.
- **Trust on transaction** - Customer Strongly Agreed that trust is important factor.
- **Empathy**- Customer Strongly Agreed
- **Privacy guarantee**- Customer Strongly Agreed that privacy should be guaranteed
- **Res, Avail of communication** - Customer Strongly Agreed
- **Benefit and Discounts** - Customer Strongly Agreed
- **Enjoyment**- Customer Strongly Agreed
- **Convenient and Flexible** - Customer Strongly Agreed
- **Is return, replace important** - Customer Strongly Agreed
- **Loyalty Programs Benefits** – Most of the Customer Agreed
- **Displaying Information** - Customer Strongly Agreed
- **Good Quality Website**- Customer Strongly Agreed
- **Net Benefit** - Customer Strongly Agreed
- **Satisfaction/Trust** - Customer Strongly Agreed
- **Wide variety**- Customer Strongly Agreed
- **Relevant product information** - Customer Strongly Agreed
- **Monetary savings**- Customer Strongly Agreed
- **Patronizing** - Customer Strongly Agreed
- **Sense of Adventure** – Indifferent for various
- **Preferred e-tailer**- Customer Agreed
- **Gratification**- Customer Strongly Agreed
- **Helps you fulfill certain roles**- Customer Strongly Agreed
- **Value for money spent** - Customer Strongly Agreed
- **Tick** – Mostly liked online stores are Amazon, flipkart, snapdeal, myntra
- **Website/Application** - Amazon, flipkart, snapdeal, myntra and flipkart has more likeable.
- **Web-page Layout**- Amazon, flipkart, snapdeal, myntra and flipkart has more likeable.
- **Wild variety** – Amazon and flipkart has more variety.
- **Product Information**- Amazon and flipkart has good information compared to other.
- **Website loading speed**- Amazon has good loading speed compared to other online store.
- **Reliability** - Amazon has good Reliability.
- **Quickness to purchase**- Amazon
- **several payment options** - Amazon and flipkart
- **Speedy order delivery** - Amazon has good speed of delivery.
- **Privacy** - Amazon
- **Security for financial info**- Amazon
- **Perceived Trustworthiness** - Amazon
- **Online assistance** - Amazon
- **Time to get logged**- Amazon and flipkart
- **Time in displaying graphics** – Amazon and flipkart
- **Late declaration of price**- Myntra has declare price late
- **Page loading time** - Myntra
- **Limited mode of payment**- Myntra
- **Delivery period** – Paytm has late delivery period.
- **Change in website/Application** – Customer need changes in amazon and flipkart
- **Frequent disruption**- Amazon and flipkart
- **efficient as before** – Amazon, flipkart and snapdeal has efficient as before
- **Recommend Indian online retailer** – Amazon is first choice and second is flipkart

Change Categorical values into Integers

As most of the features are categorical values and we cannot perform EDA with categorical values. So I changed the categorical values to integer with the help of Label Encoder.

Heat Map

I Used heatmap plotting to check multi co-linearity between features. Means if any feature is depend on any other feature and the dependency is greater than 70% then it will help us to take action and drop the column. This will also help in feature engineering.

As I plot Heatmap for customer retention, I found most of feature has multi co-linearity but I cannot drop any feature because most of the feature are Categorical and categorical values will be same but will put different effort on the target.

Density Plot

Density plot helps in understanding the value distribution of the feature. Density plot helped in identify that how these values of features will impact on model score because if the values distribution is normal then the model accuracy will automatically increase.

As I plot the density plot for the dataset. I found the all the feature has not normal distribution, all has skewness either left or right. I accept these distribution because most of the feature has categorical values and categorical values has these type distribution.

Box Plot

Box plot will help in identify that feature has outliers or not. Outlier are like noise, these can impact the model accuracy directly. We can also say that outlier are those values which does are not normal values like other values in the features. These value can be write by mistake or these can be exceptional values.

I plot the box plot for the dataset and found that only 6 feature has outliers, and other features has not any outliers. As only 6 feature has outliers and these feature has categorical values and categorical values will accept with some outliers which can be expectation values.

Conclusion

According to my analysis, the most frequent shopping online stores are Amazon, Flipkart, Snapdeal and Myntra and Females doing more shopping compared to the men. Customer need to more easy approach to the seller, clearer and complete product information, less delivery period and easy return and replace policy.

I can also conclude that the first choice for the online shopping platform is Amazon, customer like to do shopping from amazon and also ready to refer but only negative thing about amazon is customer need more attractive interface of the website.

We can conclude different thinks from this dataset like – Most popular Browser, OS, ISP Provider, Phone service and many more. It all depend on the target result.

With the help of this dataset I can also say that this analysis will also increase the sales of the online store because this dataset has clearly mention that what exactly customer need in the online stores.