# FLIP ROBO

# Customer Retention Project

Submitted by:

EKANSH GUPTA

**ACKNOWLEDGMENT**

This includes mentioning of all the references, research papers, data sources, professionals and other resources that helped you and guided you in completion of the project.

# INTRODUCTION

- **Business Problem Framing**

  Customer satisfaction has emerged as one of the most important factors that guarantee the success of online shopping store. According to the research ,Customer satisfaction is affected by several factors which in turn contributes to the repeat purchase intention (customer retention).These factors consist of both **Hedonic and Utilitarian values.**

  Business problem is to identify how different factors contribute to the customer satisfaction based on the data from different Indian online shopping stores and to provide key findings and conclusion from the analysis of that data.

- **Conceptual Background of the Domain Problem**

  Customer satisfaction is the key to customer retention

- **Motivation for the Problem Undertaken**
  To understand real world problems where Data Analysis can be applied to help organizations in various domains from which one domain is eRetail industry.

# Analytical Problem Framing

- **Mathematical/ Analytical Modeling of the Problem**

  **Mathematical modelling:**

  ➢ The dataset contained too many features some of them were even irrelevant for analysis  like the customers pin code which was adding no value since there was an attribute showing the city name of the customers. So in my analysis I decided to choose only the relevant factors .

➢ The dataset consist of 71 features and 269 rows. All of the attributes were of 'object' type except the pin code (int).
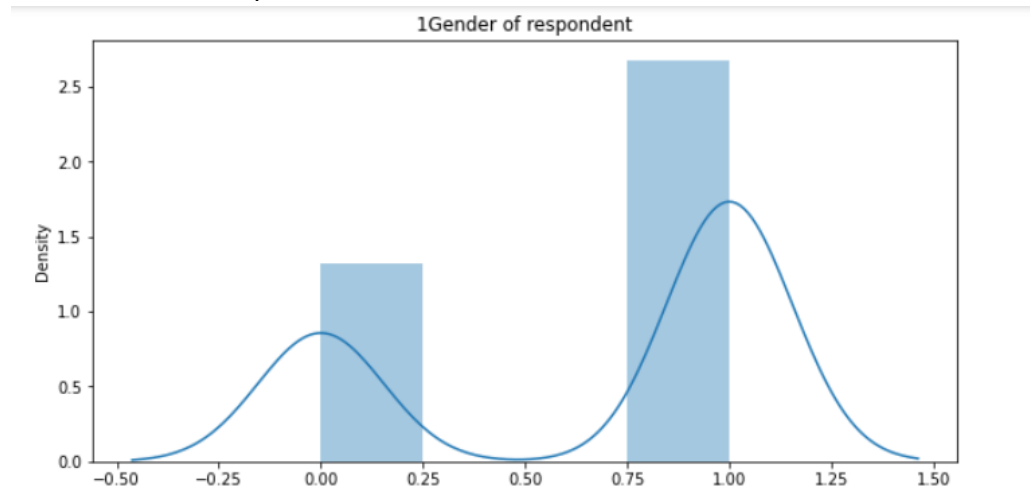➢ Dataset did not contained any null values. This was seen using below method:
**np.sum(Data.isnull().sum())**
where Data is the dataset.

**Analytical modelling :**

Univariate Analysis:

- I have first used seaborn's displot to see the distribution of each feature. Below is one example of one such attribute:



- After this to get '%' view of data I displayed value counts of all attributes and converted them to display percentage of each value present in each category. Below is the code used to achieve this along with some examples:

```
##### Displaying results in percentage format
for i in Data.columns:
    print(round(Data[i].value_counts()/2.69),2)
    print('\n')
```

```
Above 4 years        36.0
2-3 years            24.0
3-4 years            17.0
Less than 1 year     16.0
1-2 years             6.0
Name: 5 Since How Long You are Shopping Online ?, dtype: float64 2
```

```
Less than 10 times   42.0
31-40 times          23.0
41 times and above   17.0
11-20 times          11.0
21-30 times           4.0
42 times and above    2.0
Name: 6 How many times you have made an online purchase in the past 1 year?, dtype: float64 2
```

- I then used seaborn countplot to view the frequency distribution of each imprtant attribute. Below is the code for that with an example:

```
for i in Data.columns[:47]:
    plt.subplots(figsize=(10,5))
    sb.countplot(x=i,data=Data)
    #print(Data[i].value_counts())
```
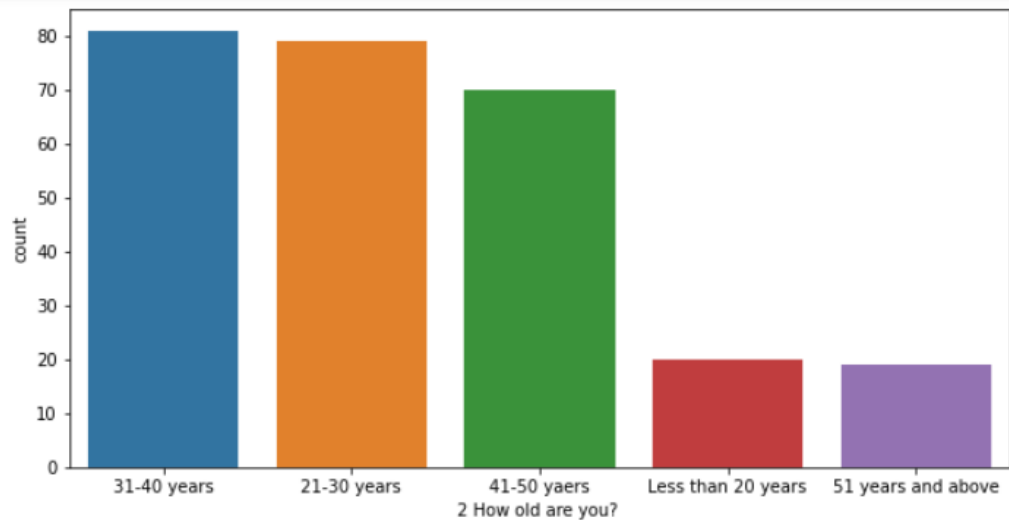


After this it was time to do bivariate analysis:

Bivariate Analysis:

- Since the motive was to do the analysis of customer retention factors therefore in my analysis I decided to view all attributes against 'the number of times a customer has done online shopping' since this was the attribute which contained the purchase frequency of customers and hence can be utilized for analysis. Below is the code for that with example:

```
data1=Data.iloc[:,[0,1,2,4,5,13,14,16,26,28,31,32,33,34,36,38,39,40,46]]
for i in data1.columns:
    plt.subplots(figsize=(12,5))
    sb.countplot(x='6 How many times you have made an online purchase in the past 1 year?',hue=i,data=data1)
```



- I further realized that some insights can be taken using only the data of high purchase frequency customers so I made a separate dataset which contained

data of only the customers who have purchased more than 20 times in last one year. Below is the code:

**data3=Data[(Data[Data.columns[5]]=='31-40 times') | (Data[Data.columns[5]]=='41 times and above') | (Data[Data.columns[5]]=='21-30 times') ]**
**where data3 is the dataset containing only high purchase frequency customers.**

- Using seaborn countplot I then utilized the above dataset for analysis of various attributes. Below is the code with example:

```
data_3=data3.iloc[:,[5,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70]]
for i in data_3.columns:
    plt.subplots(figsize=(12,5))
    sb.countplot(hue='6 How many times you have made an online purchase in the past 1 year?',x=i,data=data_3)
    plt.xticks(rotation=75)
```
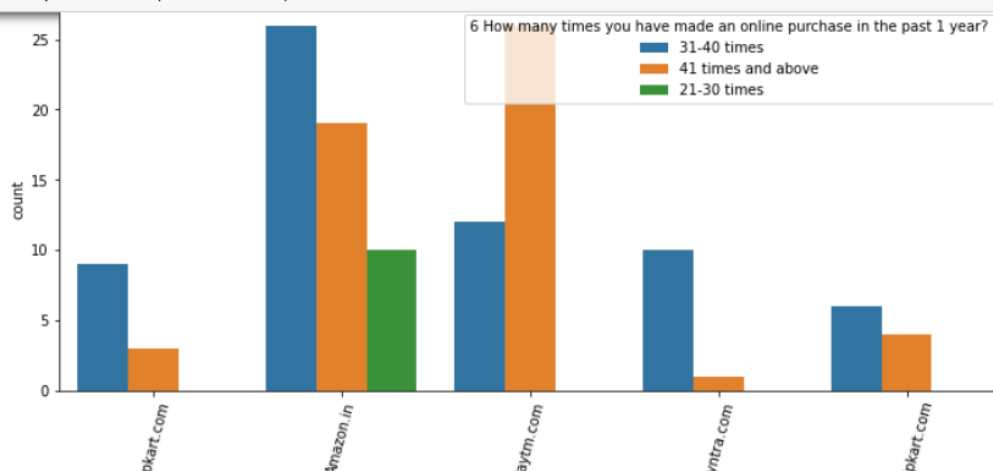


- After doing bivariate analysis and noting down the key findings/conclusion I then proceeded towards multivariate analysis . Since all attributes were categorical in nature it was difficult to do multivariate analysis using visualisation techniques (plots/graphs etc) , I utilized grouping technique on key features to do multivariate analysis. Below is one example showing data of only high purchase frequency customer first grouped by Gender and then on city to get insight of which gender group people retended most from which city:

```
data3.groupby([Data.columns[5],Data.columns[0],Data.columns[2]]).count().iloc[:,:1]
```

| 6 How many times you have made an online purchase in the past 1 year? | 1Gender of respondent | 3 Which city do you shop online from? | 2 How old are you? |
|---|---|---|---|
| 21-30 times | Male | Delhi | 1 |
| | | Moradabad | 5 |
| | | Noida | 4 |
| 31-40 times | Female | Bangalore | 11 |
| | | Ghaziabad | 10 |
| | | Greater Noida | 10 |
| | | Karnal | 10 |
| | | Merrut | 6 |
| | Male | Delhi | 12 |
| | | Greater Noida | 1 |
| | | Gurgaon | 1 |
| | | Noida | 2 |
| 41 times and above | Female | Bangalore | 2 |
| | | Delhi | 7 |
| | | Ghaziabad | 3 |
| | | Greater Noida | 16 |
| | | Karnal | 5 |
| | | Merrut | 1 |
| | | Noida | 5 |
| | Male | Delhi | 9 |
| | | Noida | 5 |

- **Data Sources and their formats**

I got the data from FlipRobo technologies in .xlsx format , the origin of data is from Indian online shopping stores.

There was one main file containing two sheets ,where the first sheet showed the data as it is without categories being encoded and second one was encoded version of the first sheet.

The dataset did not contain any null values.

## Data Preprocessing Done

While viewing percentage count I observed some values were repeated with no change in meaning. For example in the attribute 'How do you access the Internet' the value **Mobile internet** and **Mobile Internet** were two different values with same meaning. I replaced all such values accordingly. Below is the code:

```
## Replacing same values with same name
Data.iloc[:,5].replace('42 times and above','41 times and above',inplace=True)
Data.iloc[:,6].replace('Mobile internet','Mobile Internet',inplace=True)
```

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

The input data was raw containing unrealistic values in some features which I have described how I dealt with. Also the data contained many outliers, outliers removal has also been discussed earlier. Mainly all the important features were int64 and float64 type , I observed that in all features the min and max values were in the range of int32 and float32 type so to avoid unnecessary memory consumption and efficient memory use I converted the features from int64 to int32 and float64 to float32. And then data was scaled before feeding the data to machine learning model.

- In my analysis I have assumed that Indonesian Telecom providers do not keep their customers account balance (main or data account) in negative.

- **Hardware and Software Requirements and Tools Used**

  **Hardware used**: I prepared this complete project on 8 Gb RAM machine to ensure good processing of algorithms since we have huge data. Nothing else was required related to hardware stuff.

  **Software used**: Anaconda installation was the  must requirement since it installs two main components needed to develop any datascience project i.e Python programming and Jupyter notebook (a tool used to write python codes ,typically used by all datascientists)

  **Libraries used**: Below are the python library used in this project.

  **Pandas**:  To read the Data file in form of data frame and do various data manipulations using pandas Data frame functions.

  **Matplotlib**: This library is typically used to plot the figures for better visualisation of data.

  **Seaborn**: A advanced version of Matplotlib to do data visualisation.

## Model/s Development and Evaluation

- No model was prepared since only data analysis was required

**CONCLUSION**

- **Key Findings and Conclusions of the Study**

  Below are the key findings/conclusions obtained from all analysis helpful in understanding customer retention:

  **Univariate Analysis findings:**

  * Female customers(67%)are more than male customers
  * Customers between 20 to 50 years shopped more from the store. 75 % of customers are between 20 to 50
  * NCR and Bangalore (67%) are the cities with high number of buyers
  * 36% customers were found shopping online for more than 4 years , 24% were shopping for 2-3 years
  * IN last one year 42% customers have purchased online less than 10 times,23% cutomers purchased 31-40 times only 2 percent customers purchased more than 42 times
  * highest 70% customers used mobile internet for online purchase
  * 52% customers used Smartfone for online purchase, 32% used laptop least 4% used tablet
  * 45% customers had windows OS on their device of purchase, 32% had Android
  * 80% customers used Google chrome for for online purhcase
  * 86% customers uses Search engine to reach their favorite online store
  * After first visit 32 % customers used search engine to reach online store and 32% used application
  * 46% customers take at least 15 minutes before making a purchase, 26% take 6-10 minutes only
  * 55% customers prefer Credit/Debit cards to make payments ,28% preferred COD and 17 % preferred Ewallets
  * 64% customers not usually abandon their shopping cart, 19% abandon their cart frequently
  * 49% customres abandon their bag due to better alternative offer, 20% abandon due to promocode not applicable
  * 61% customers have strongly agreed to have easy website content which is easy to understand
  * 43% customers likes the information shared in context to similat product ,which is useful in comparison
  * 70% customers like to have seller and complete product info to make buying decisions
  * 90% customers like to have relevant information on listed products to be started clearly.
  * 90% cusotmers like the website to be easily navigable
  * 85% cusotmers had no issues with the loading and processing speed
  * 87% customers like the website interface
  * 90% customers liked the paying methods.

* 52% customers trust that online store will fufill its part of transaction at stipulated time
* 85% customers like the organization's readiness to assist with queries
* 90% customers belive the retail store to guarantee their privacy
* 90% customers liked the responsiveness of several communication channels
* 70% agree that online shopping gives monetary benefit and discounts to the customer
* 60% customers enjoys online shopping
* 84% customers agree that online shopping is convenient and flexible
* 90% customers agree that return and replacement policy helps them making purchase decision
* aPPROX 50% customers agree that gaining access to loyality program is a benefit of shopping online
* 80% customers like to have quality information on the website
* 95% customers are satisfied while shopping on a good quality website
* 80% customers agree that net benefit derived can improve user satisfaction
* 85% customers agree that customer satisfaction cannot be built without trust
* 70% like to have a wide variety of listed products in several category.
* 86% customers like to have complete and relevant information
* 80% customers agree to receive monetary savings while shopping online
* 50% agree with the convenience of patronizing the online retailer
* 50% customers agree that online shopping give the sense of adventure
* only 30% customers agree that online shopping enhances their social status
* 45% customers feel gratified while shopping with their favorite retailer
* 45% customers feel that shopping onine helps them fulfill certain roles
* 85% customers agree they get value for their money while shopping online
* Most customers finds Amazon.in weblayout most appealing followed by Flipkart and Myntra
* 48% customers says flipkart, amaon shows wide variety of products
* 37% customers like flipkart and amazon in terms of displaying complete and relevant information of the products
* In terms of speed Amazon.in is liked by 90% customers , 60 % like Flipkart , 27% Myntra
* 78% customers trust Amazon to be reliable , 55% likes flipkart , 25% Myntra
* 80% customers likes Amazon's quickness to complete the purchase , 60% likes Flipart's, 30% likes myntra's
* 86% likes Amazon's delivery speed , 56% likes flipkart's , 5% likes myntra's
* 82% customers trust amazon in terms of keeping the privacy of their data, 56% trust flipkart 20% trust myntra , 12 % trust snapdeal
* 70% trust amazon in  terms of keeping their financial information secured, 56% flipakrt, 34% myntra , 21% snapdeal , 22@ paytm
* 84% customers believe Amazon perceived trustworthiness, 43% - Flipkart , 33% - Myntra ,snapdeal 27% snapdeal

* 86% customers like Amazon in terms of online assistance through multi channel , 51% -Flipkart ,42% myntra
* 50% customers agree that Amazon takes longer time to log them in while in sales period/promotions , 39% goes with flipkart, 13%  goes with myntra , 16%-snapdeal
* 46% customers says that amazon takes londer time in dsiplaying photos in sales/promition , 35%- flipkart, 21% snapdeal, 14% myntra, 11% paytm
* In terms of late declaration of price in promotion/sales --28% goes with  myntra, 21% amazon , 16% flipkart , 18% snapedeal, 23% paytm
* In terms of longer page loading in promotion/sales -- 26% customer goes with myntra, 23% -flipkart,23% -amazon,24% snapdeal
* In terms of limited mode of payment on most products during sales period/promotions -- 41% goes with snapdeal, 40% -amazon,23% flipkart, 20% paytm and only 3% myntra . Hence myntra gives most payment options during sales/promotion.
* In terms of time taken in product delivery-- paytm has highest votes of 37%, snapdeal 24%, flipkart 16%, amazon 14% and  snapdeal and myntra 10%. hence snapdeal and myntra takes minimum time for delivery among all.
* 20% customers dislikes disruptions while moving to another page on amazon, 24% on myntra, 27% snapdeal, 14% paytm,23% snapdeal
* 61% customers says that amazon website is as efficient as before, 43% for flipkart,22% for paytm,9% for snapdeal
* 81% customers would like to recommend amazon to a friend, 43% would like flipkart , 22% would like for myntra ,9% for paytm and least 4 % would like to recommend snapdeal.

**Bivariate analysis findings:**

* customers with age between 21 to 50 have shopped more number of times from online store
* NCR region contains customers who have shopped dmost number of times
* Most customers who are shopping online from 1 to 4 years have shopped less than 10 times and customers shoppping online for above 4 years have highest purchase count in all more than 10 times group
* Most customers spends more than 15 minutes before making decision irrespective of the purchase count
* Most cutomers who purchased less than 10 times preferred COD as payment option along with credit/debit cards.All customers who purchased more than 10 times preferred credit/Debit cards for payment.
* Most customers who purchased less than 10 times abandoned the cart due to better alternative offer.cutomers purchasing more than 41 times abandoned due to promo code not applied Or better alternative offer

* It can be seen clearly that all customers irrespective of the purchase count strongly agree with the empathy which measn empathy is very important for customer satisfaction and hence retention
* It can be seen clearly that all customers irrespective of the purchase count strongly agree with availability of several communication channels
* All customers irrespective of purchase count agree with online shopping being flexible
* It can be seen that return and replacement policy plays an important role in customer satisfaction
* Customers who have purchased many times (more than 30) agrees that they get benefit of gaining access to loyality programs while shopping online
* Most of the users who have made numerous online purchase agree to have quality information displayed on the site

* Customers who have shopped more than 30 times strongly agree that net benefit from online purchase leads to user satisfaction Hence net benefit derived is an important factor for customer retention
* Customers having purchase frequency of more than 30 times strongly agree to get wide variety of listed prodcuts in several category.
* High purchase frequency customers ( >30) strongly agree to have complete and relevant product information
* Cutomers with high purchase (>30) agree to get monetary savings

* Almost all customers with high purchase agree that they get value for money spent.it indicates that quality of product sold plays an important role in customer retention
* Highest number of customers have shopped from amazon, paytm and myntra
* Highets number of customers like amazon, paymt and myntra web page layout
* Highest number of customers like amazon and flipkart in terms of variety of product offered
* Highest number of customers like amazon and flipkart in terms of relevant and complete description of products
* High proportion of high purchase frequency customers like amazon and paytm in terms od website speed
* Highest relaibility is obtianed by amazon and flipkart
* Highest number of customers like Amazon in terms of quickness to complete purchase
* In terms of payment option Amazon, flipkart and myntra are liked most by the high purchase frequency customers
* Amazon stands apart in terms of speed of order delivery and privacy of customers information
* Highest number of customers like amazon,flipkart in terms of security of customer private inforamtion
* Amazon perceived highest trustworthyness of high purchase frequency customers

* Most of customers like amazon in terms of online assistance through multichannel
* Most user claims flipkart to take maximumm login time during sales/promotion
* Highest customers claims myntra to take maximum time in displaying graphics and photos

**Multivariate analysis findings:**

**\*** Although we have seen that female customers are the most retended customers , specifically female customers from Greater noida stands highest in retention followed by males of Delhi
* Female's of age group 20 to 50 have contributed most in retention
* Out of most retented customers ,Most time taken by Female before making pruchase decision was more than 15 min
* Out of most retended customers ,Females choose to opt most for ewallets followed by credit/Deit cards
* most retended Female customers abandoned their bag due to better alternative offer,promocode not applicable
* Highest retended customers are from Delhi and of age group 21-30, second highest are again from Delhi of age group 41-50
* Highest retended customers are from Delhi and prefer credit/debit cards for payments, second highest are from Ghaziabad and gr noida pereferring COD and ewallets respectively. hence payment option preferrence varies for retened customers along with city

- **Learning Outcomes of the Study in respect of Data Science**

  I found visualisation a very useful technique to infer insights from dataset.

  I faced some difficulty in visualization of data while doing multivariate analysis since all attributes were categorical in nature.

- **Limitations of this work and Scope for Future Work**

  Since the data contained less number of rows only 269. The analysis will be limited in scope .More data of such customers will definitely help increasing the probability of findings drawn above.