



CUSTOMER RETENTION DATA ANALYSIS REPORT

ABSTRACT

This article is about detailed report on data analysis of collected data of survey on Indian e-commerce shopping website i.e. identifying the factor that influence the customer repeat purchase intention.

Konatala Mohit

Internship 23, ID-34

Context

1. Introduction

2. Objective

3. Descriptive Data Analysis

4. Data Visualization

5. Observation of Univariate Analysis

6. Feature Selection

7. Preparing the data for Model training

8. Conclusion

Introduction:

In today's digital era there has been boom in E-commerce and its market is still accelerating at a sky rocketing speed. E-commerce is a steppe platform for many rapidly evolving and growing companies to sell commodities. E-retailers expand their business by mean of increment in number of customers to their platform. For any E-retailer to be successful there are many factors to surge the sales but the key to the solution is very simple i.e. to satisfy the customer. Satisfying the customer leads to retaining them back to E-retail platform and the cycle of purchase is repeated. There have been many researches carried out to explore what exactly retention the customer and how to extract loyalty out of all the customers. There majorly five features that influence the success of an E-commerce, they are service and system quality, information quality and to sum it up trust and net benefit are the final pillar of a profound successful E-retailer. In this project exploratory analysis is done on the data collected from the survey of Indian customers regarding the e-commerce shopping.

Objective:

Analysis of the data set and summarizing the factor that effect the customers recommendation regarding e-retail store.

Exploratory Data Analysis:

1.Descriptive Analysis:

- The size of the dataset is pretty small.
- Shape of the Dataset: 269 rows and 71 columns
- In the dataset other than 'Pin Code of where you shop online from' column (i.e. column number 3 is integer type) all the data in the columns are in ordinal form.
- The dataset doesn't contain any null values in any of the columns.
- Checking the number of unique elements in each column. Third column i.e. address pin code column has maximum unique values in the whole dataset.
- Checking the count of each repeating individual element in the every column of the dataset.

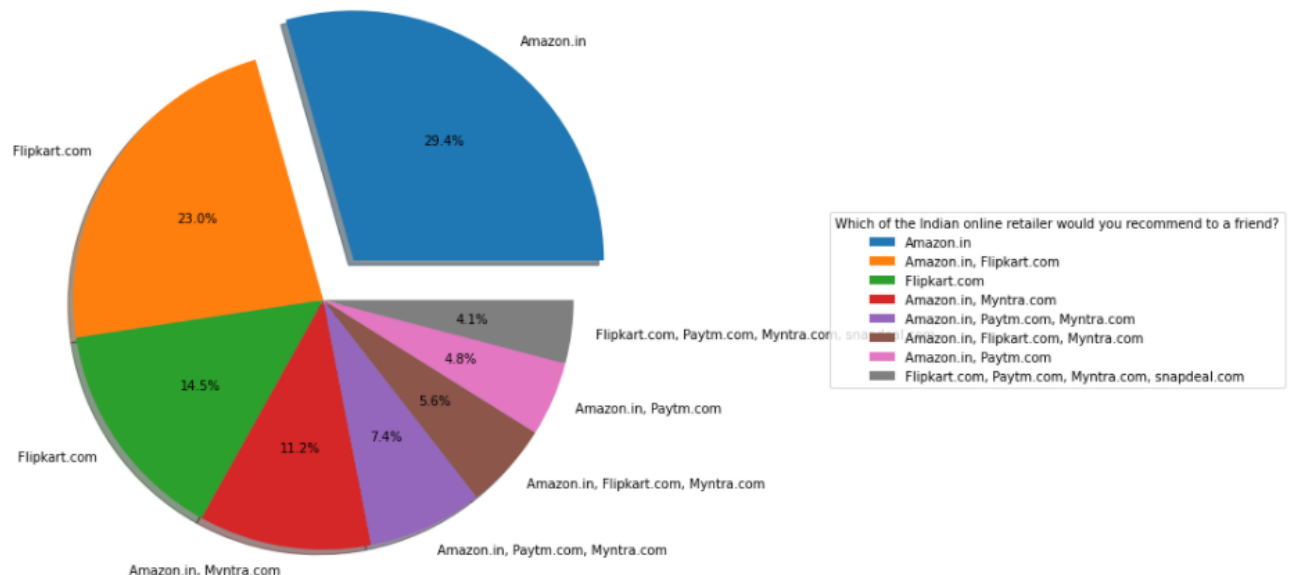
Data Visualization:

Univariate Data Analysis:

Visualizing each column data using pie plot get clear picture of the data.

- Gender: Female have majority proportion compared to male.
- 31-40 followed by 21-30 years old are more into e-shopping. 51 and above age are less into e-shopping.
- City: Delhi has huge e-shopping population and Bulandshahr has the least.
- Pin code: doesn't seem make in the data set as we already have city column in the column.
- Majority of the public have been using e-shopping more 3 years and above.
- Less than 10 online purchases have been made by majority of the people.
- For maximum people the source of internet for online shopping is mobile internet.
- Smartphone are main access for online shopping for most people, tablets are least used for e-shopping.
- Screen size column data has others data
- Operating System who are on windows platform high proportion.
- Google Chrome is the highly used browser to access the e-shopping website, and barely anyone is using mozilla firefox
- Most user are using search engine to their online favourite e-shopping.
- Search engine, Via application, Direct URL is how most of the user are accessing the online platform.
- Interestingly most people are exploring more than 15 mins before making any purchase.
- Most of the transactions are made via Credit and Debit cards, e-wallets the least.
- Large ratio sometimes add the goods to the cart and abandon them without paying and the reason behind abandoning is better alternative option for that particular good.
- Content of the website must be easy to read and understand is what most user strongly agreed on.
- Most user are highly in search for similar product info that to highlighted with product comparisons.
- For user seller info is important before making any purchases.
- User want all the relevant information regarding the product and it must be stated very clearly.
- Easy navigation option is what most user are in looking for on an e-shopping store.
- Loading and processing time of the website must be high is what most users are agreeing on.
- Majority of the user are strongly agreeing on easy UI of the website.
- Payment must be convenient for user is what the data is saying and also trust on transaction is more when the online e-commerce shop can complete the payment in minimal amount of time.
- Highly quality customer service is what majority of the users are looking for from the e-shopping website.
- For majority of the customers privacy matters the most on an e-shopping platform.
- Customers seeking major interest in monetary benefits and discounts.
- Maximum customers seek enjoyment out of online shopping and also they want the shopping to be convenient and flexible.

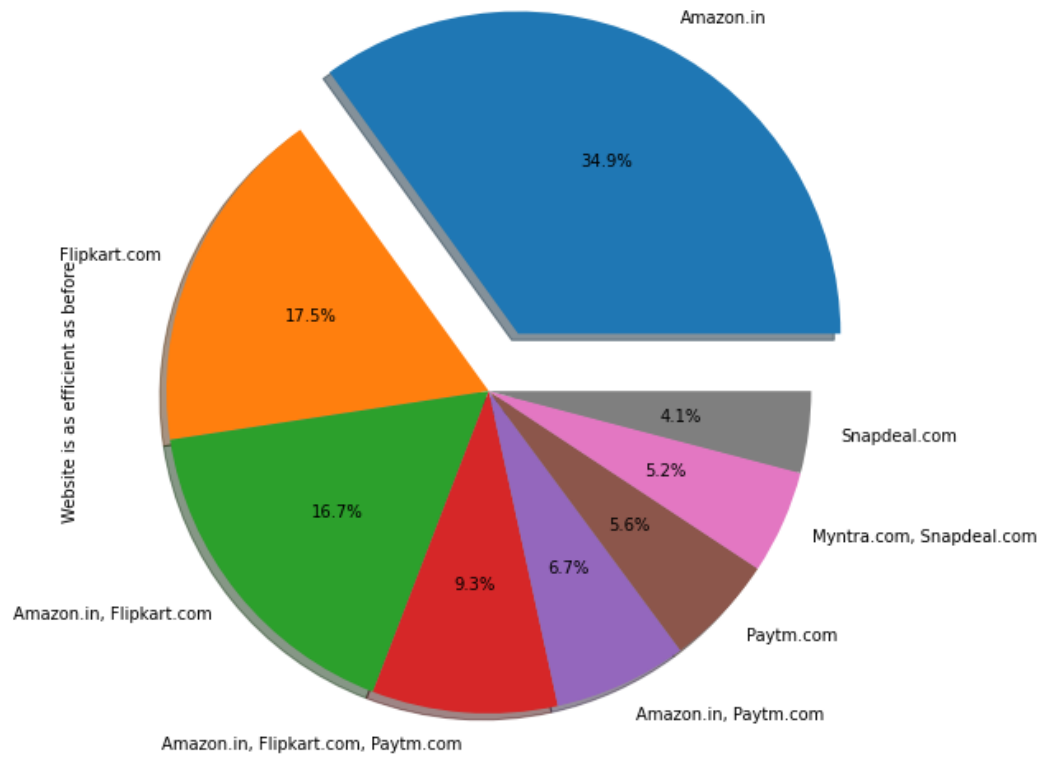
- Return and replacement of the product play an important role for a user in approaching a particular e-store also exclusive access to the loyalty programs
- User want the e-shopping website to display quality information so as to match customers satisfaction and also quality of application and website.
- For customers to prefer online e-shopping it should net benefits.
- User prefer the e-store to huge diversity in the type of product they sell.
- People gratified when they shop on their favourite website.
- Value on purchase is what most user look for on a online shopping platform
- 30% of the users have shopped from all the mentioned online shops.
- User want the product to be delivered as early as possible after placing an order.
- Amazon website seem to more of user trust compared any other e-commerce store.
- Amazon also has multiple online assistance channels.
- Paytm and snapdeal have longer delivery period compared to other e-stores.
- Amazon seems to constantly modifying the page to meet the customers satisfaction.
- Amazon And flipkart have the efficient website.
- Most users would recommend Amazon, Flipkart and Myntra to their friends based their experience with the e-platform.



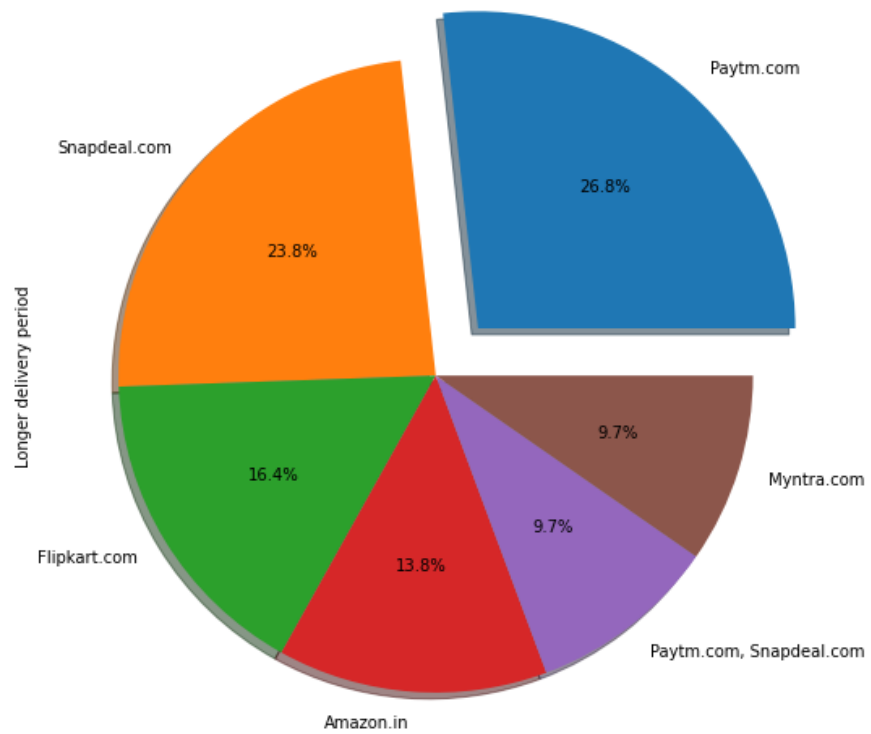
The above Image Indicates the User recommendation of the e-platform to their friend.

Below Plot are the few of factor plot that are effecting the above recommendation pie plot i.e. based the on observation of univariate analysis visualization.

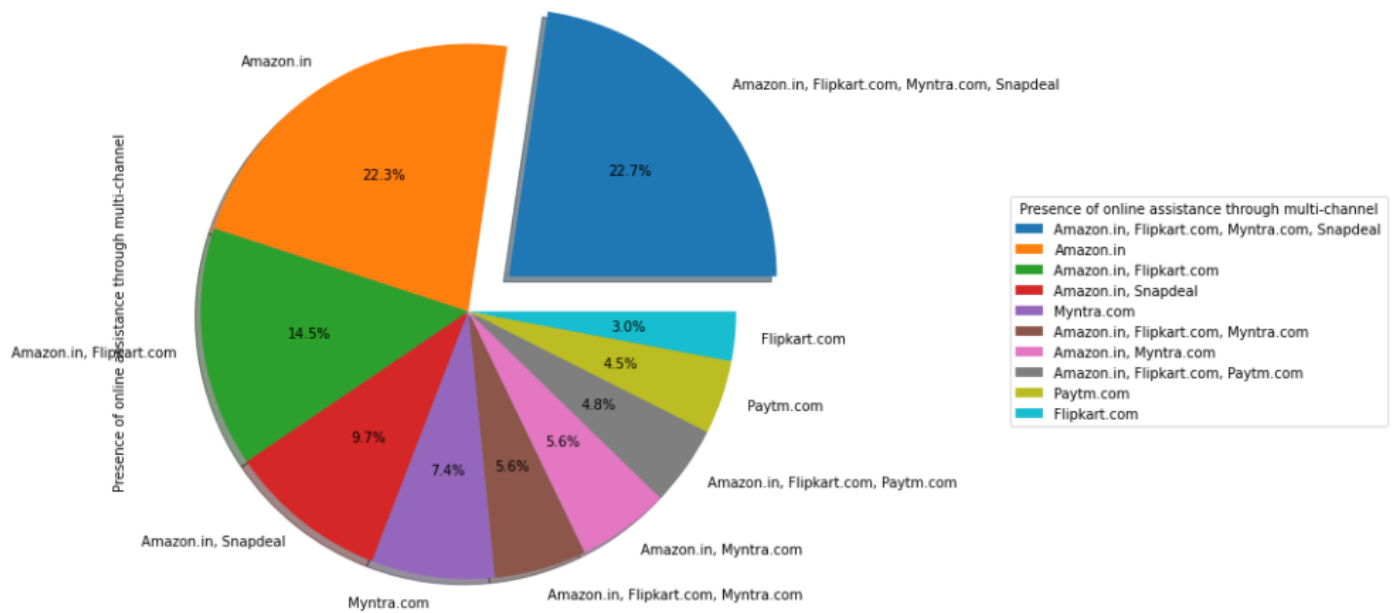
Website is as efficient as before



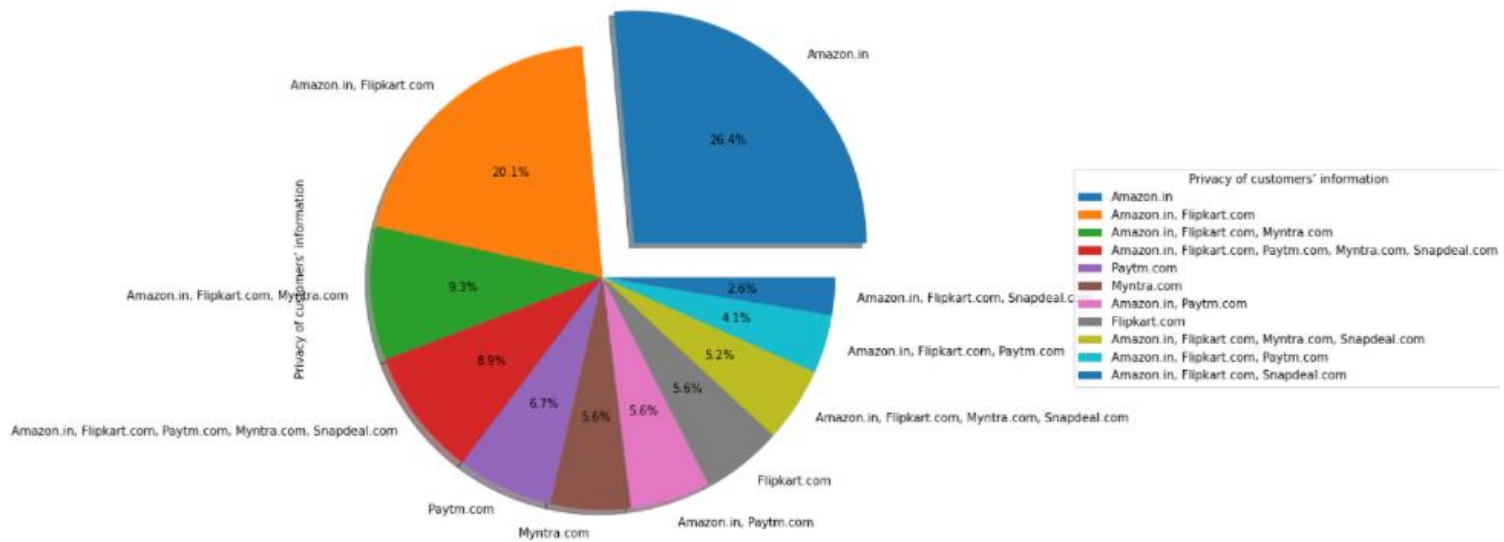
Longer delivery period



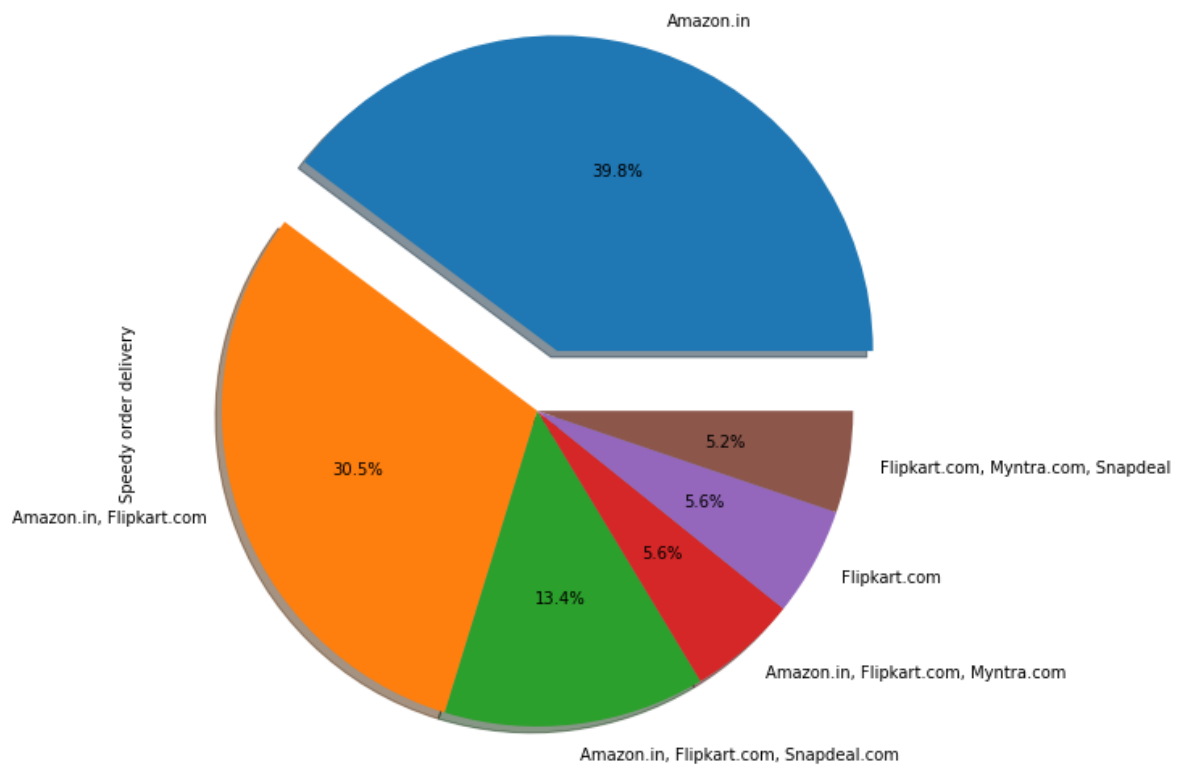
Presence of online assistance through multi-channel



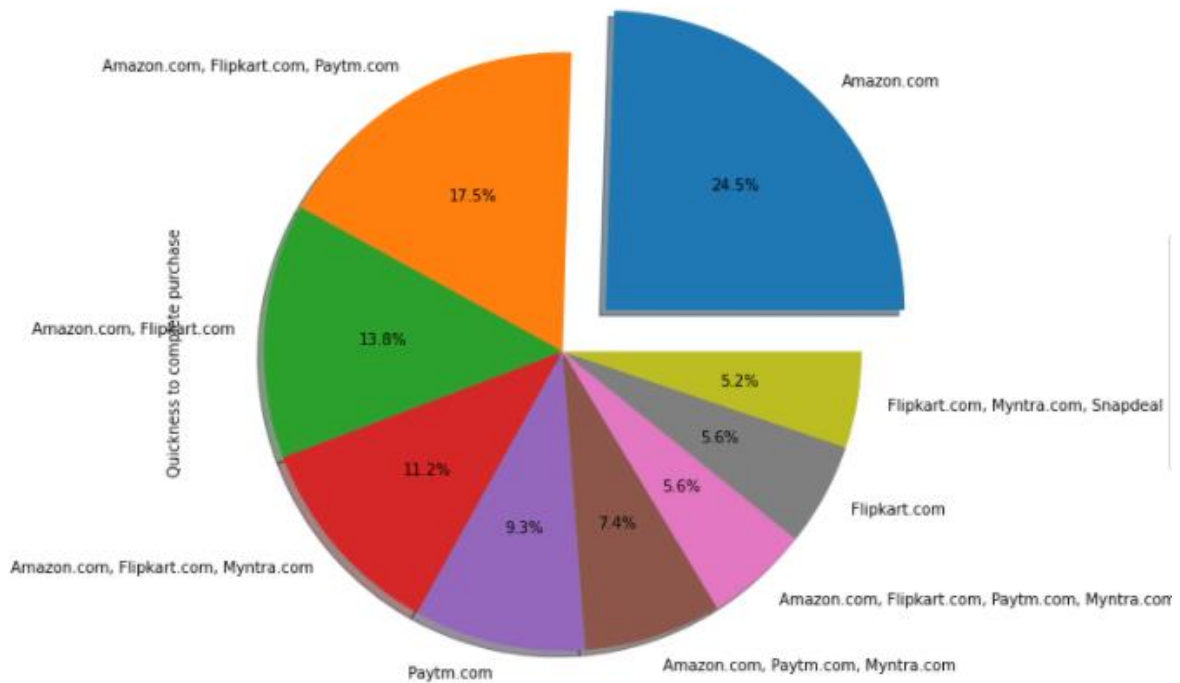
Privacy of customers' information



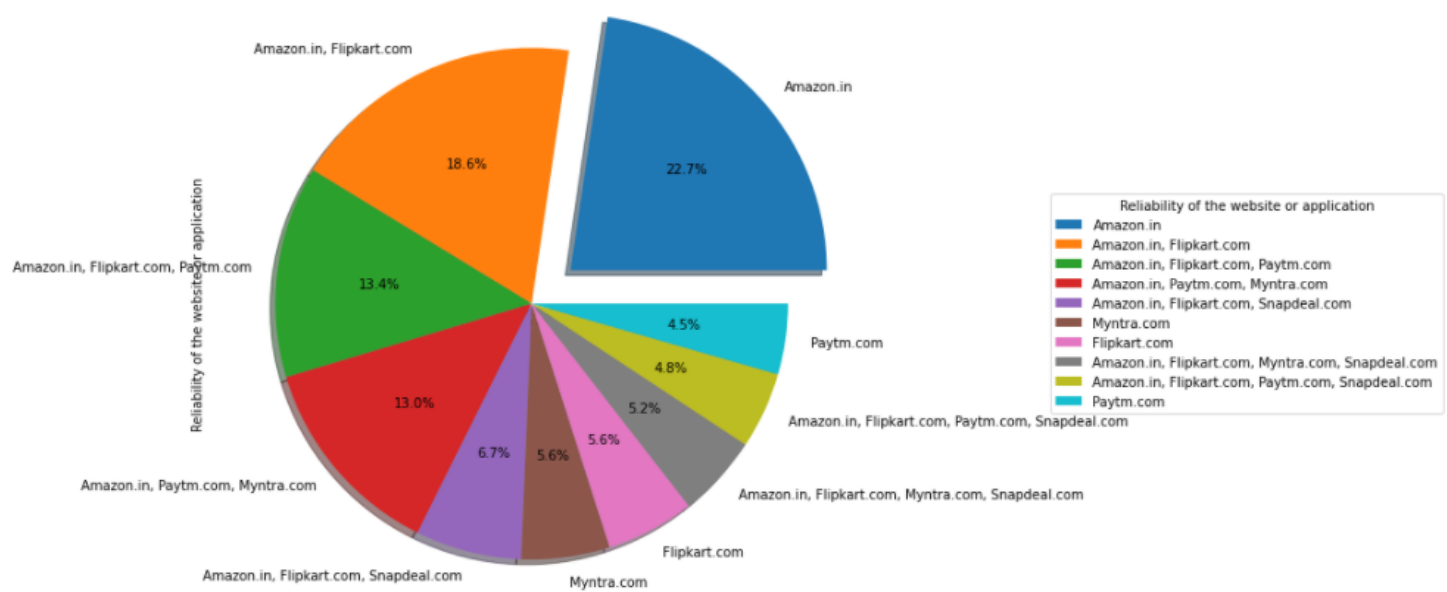
Speedy order delivery



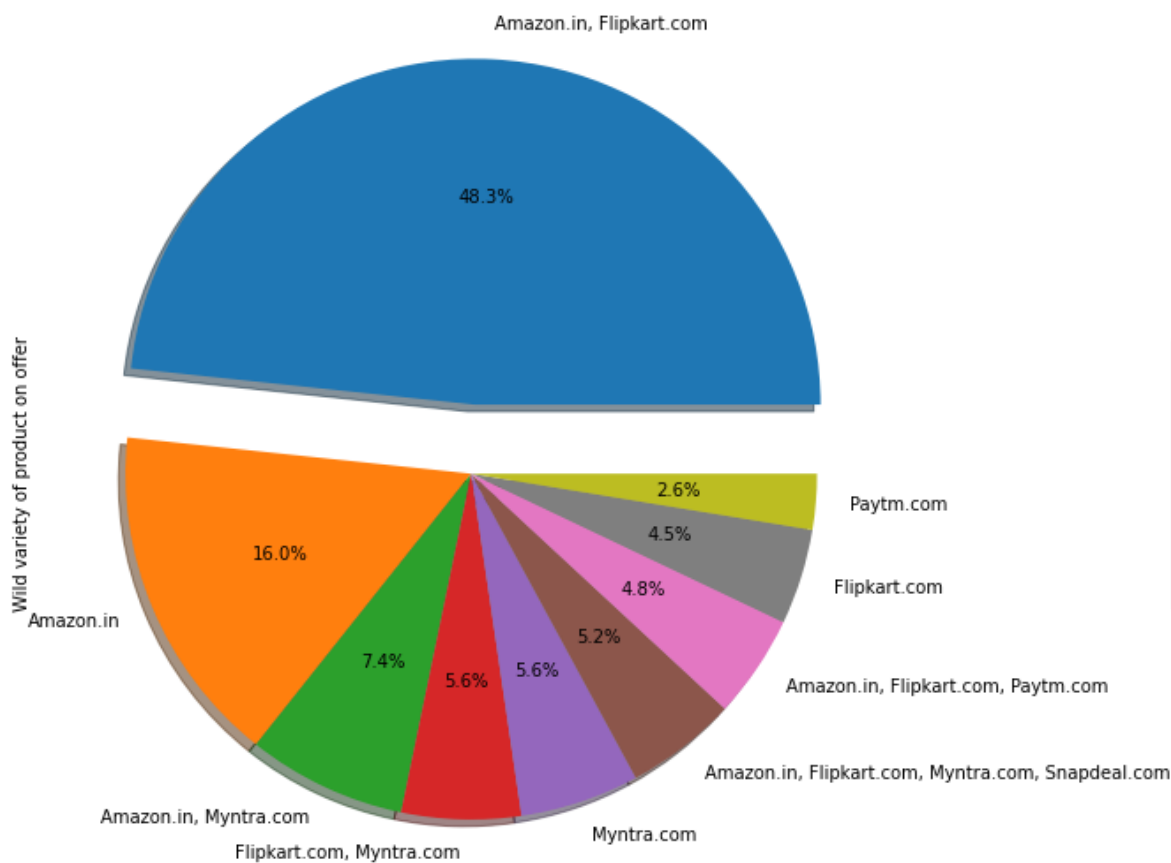
Quickness to complete purchase



Reliability of the website or application



Wild variety of product on offer



Observation from univariate analysis:

Factors influencing the customer retention based on the univariate analysis is as follows:

1. Easy User Interface
2. Information of similar product highlighted with the current searched product also the seller information of the product.
3. Clear information of the product.
4. Loading and processing speed of the application or the website.
5. Convenient payment method.
6. Customer Service and privacy.
7. Net benefits and discounts.
8. Return and Replacement policy.
9. Wide variety of products availability on the website.
10. Value for money.
11. All the filtered data is ordinal

Dropping the Columns based on the observation. All the columns before 49th column are the review columns given by the users so dropping them. Now number of columns is equal to 23.

- Most of the user think all of the website are easy to use but overall amazon is most easy to website or application
- Amazon's and Flipkart's page layout is visually the most appealing.
- Amazon and Flipkart are which sell wild variety of products on offer also they have the most relevant and detail description of the product.
- Amazon is standing on top when it comes to loading the page instantly.
- Amazon is the most reliable website also quickest when it comes to purchase.
- Amazon and Flipkart offer the more varieties of payment options and Snapdeal has the least.
- Amazon has the fastest delivery of the product from the time of order also the safest when it comes to customer privacy. Paytm is late when it comes to delivery
- Most users trust amazon website.
- To many users notice Amazon, Flipkart, Myntra and Snapdeal have presence of online assistance via multi-channel still overall amazon and flipkart are dominating the assistance overall.
- Amazon has the longer login time compared all the other website.
- Myntra lags in declaring the pricing of the product also it takes more time to load the page
- Amazon has frequent disruption in its page but the website is the most efficient of the lot
- Most the users recommended Amazon to their friends

Converting the Ordinal data into numerical data using label encoder to project dist-plot and boxplot to visualize the outliers and skewness.

Most of the Data is right skewed and outliers are absent in our dataset.

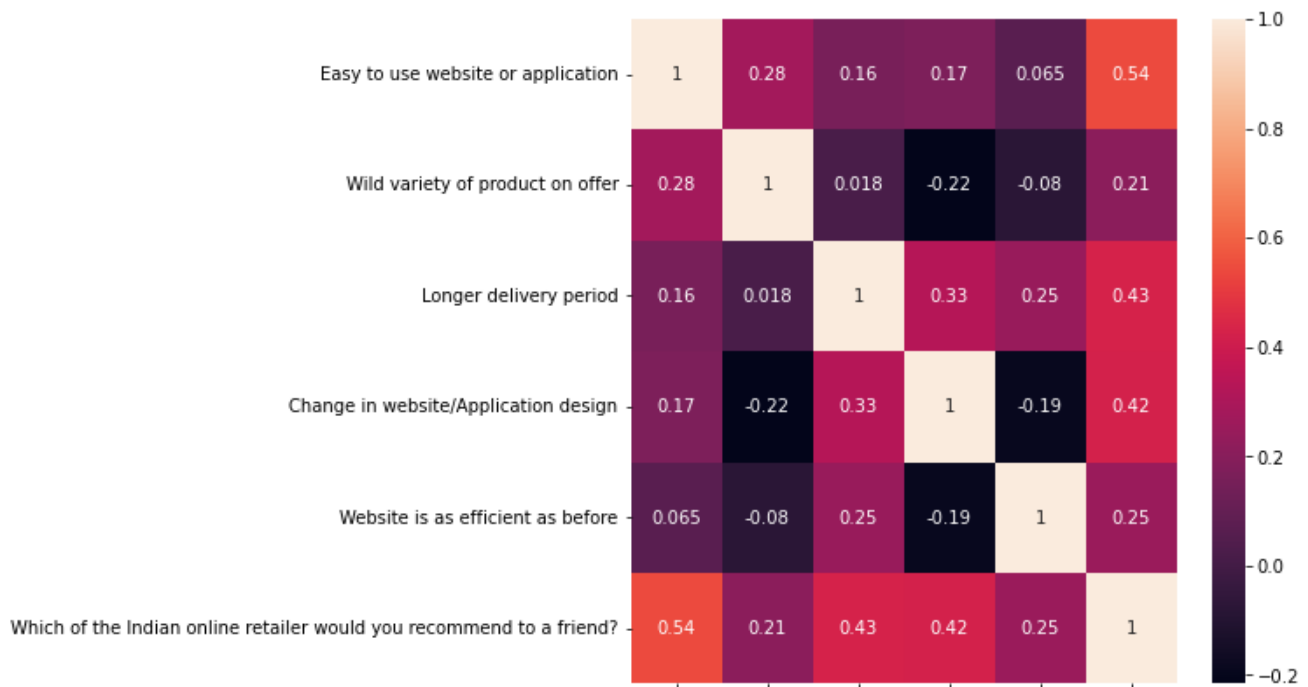
Wild variety of product on offer column has some skewness.

Feature Selection:

Checking the Corelation between the independent variables:

Using seaborn heatmap to get visual data i.e. co relation between the variables. It can be observed that the variables in the dataset have multi-collinearity between.

1. Easy to use website has high co-relation with visual appealing, relevant description product info, loading speed of the website, reliability of the page, quick purchase and several payment option, trustworthiness and assistance through multi-channels. Keeping only easy to use website column and dropping other highly correlated column as it has good corelation with website recommendation.
2. Visually appealing has multi-collinearity with longer time displaying the graphics and photos, late price declaration and longer page loading time (dropping these three, as Visually appealing has more collinearity with recommendation column)
3. Change in website design is co-related with Privacy of the customer also privacy is co-related to security of customer financial information. Therefore dropping privacy of customer column.
4. Dropping longer time to get logged and frequency disruption it has multi-collinearity with change in website or application.
5. Removing Speedy delivery column as it share co-relation with website efficient before and also the column data is right skewed.
6. Deleting Limited mode of payments as it has co-relation with Longer delivery period. Also now looking at the new co-relation heatmap Security is correlated with Wild variety of product on offer column therefore dropping Security column as it less corelation compared to wild variety column.

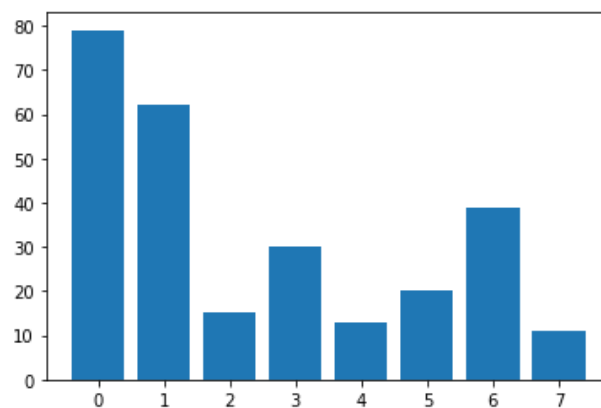


Preparing the data for Model training:

- Splitting the features from the prediction variable and storing both of them as two separate lists.
- Normalizing the feature variable to make the data standard and completely unbiased.
- Checking the variation inflation factor for cross checking the multi collinearity before feeding it to the model.
- The VIF score are below 5 for all column so there is no multi-collinearity between the existing columns.
- Using SMOTE to balance the data.

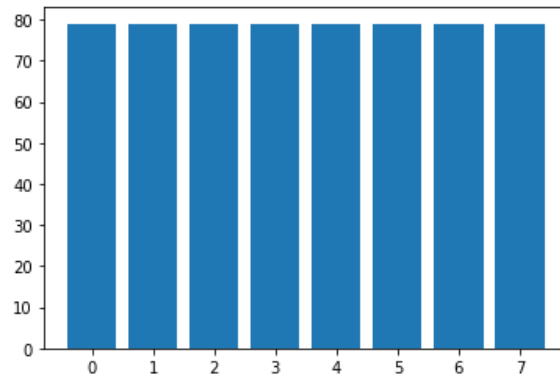
Before SMOTE:

```
Class=6, n=39 (14.498%)
Class=3, n=30 (11.152%)
Class=5, n=20 (7.435%)
Class=1, n=62 (23.048%)
Class=4, n=13 (4.833%)
Class=7, n=11 (4.089%)
Class=0, n=79 (29.368%)
Class=2, n=15 (5.576%)
```



After SMOTE:

```
Class=6, n=79 (12.500%)  
Class=3, n=79 (12.500%)  
Class=5, n=79 (12.500%)  
Class=1, n=79 (12.500%)  
Class=4, n=79 (12.500%)  
Class=7, n=79 (12.500%)  
Class=0, n=79 (12.500%)  
Class=2, n=79 (12.500%)
```



Conclusion:

After analysing the customer retention dataset one can conclude that there vast too many factors to fetch the willingness of a customer to make another purchase on the e-shopping website or application though there few major factors by which one can retain the customers, they are: The website page or application's user interface must be easy to use. Visualization always grabs the customer attention therefore having an appealing page or application designed is always beneficial . Privacy is the most important when it comes to retaining a customer as creating a safe personal environment for the customer can lead to gaining the trust worthiness of the customer. The loading speed of the page should be as minimum as possible so that the customer doesn't get annoyed by the waiting period for each time of interaction with page or application. Speed of delivery, ease of ordering and online shopping benefits can make the customer use the e-store site as often as possible. Having a vast variety of commodities can increase the chance of customer to buy a product. Brands are the other things that customer are more interested in these days as few of the brands already have good recognizing it leads to blind purchase in some scenarios. Lastly the page must always show similar product highlighting the key specifications as the item the customer is searching for so that customer can spend maximum amount of time on the website or application which can lead the customer having a better in-site at the product and increase the chance of purchase. Below the keys point in Summarizing the data analysis of customer retention data

- Size of the Dataset is size.
- No missing values in the dataset.
- Dataset contains ordinal data
- Converted ordinal data into numerical data.
- Dataset has no outliers.
- Feature selection of independent data is done via heatmap and VIF value.
- Separate features and prediction variables into two separate list.
- Used SMOTE to oversample the data and balance it.

