Customer Retention Case Study
Of
E-Commerce Websites.

Submitted by:

Ganesh Kumbhar

ACKNOWLEDGMENT

I would like to thank FLIP ROBO TECHNOLOGY for providing me the opportunity to do Data Analysis of CUSTOMER RETENTION CASE STUDY. It enhanced my data analytical capability. I also would like to thank Miss.Sapna varma madam for the guidance in this project.

To work on case study of CUSTOMER RETENTION PROJECT. I have used my basic knowledge of data analytics which I learned from DATA TRAINED institute of data science. Also, I have took the help of some famous websites like www.keggle.com & www.geeksforgeeks.com to learn & implement some useful concepts of data analysis in my research. I have also had the help of somefamous YouTube Channels of Data Science like: -

1. Edureka YouTube Channel.

2. CodeWithHarry YouTube Channel.

INTRODUCTION

Business Problem Framing

E-retail factors for 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 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.

- ## Conceptual Background of the Domain Problem

Customer Retention Case Study is a survey, which was organized by the organization to understand the core mind-set of the Indian customers who use to shop online frequently. The objective of the survey was to ask questions related to Indian online market structure, what are their likes and dislikes, what they want to improve, the usual troubles they experience, various facilities and comparison between different-different e-commerce websites and experience with them. These all questions in the survey form

was asked to fill by the respondents who use to shop online frequently, according to their experience. The survey form had 71 question. It started with the basic intro question about the shoppers, things they like while shopping online to the various features and facility comparison between some e-commerce websites. There was 5 websites which were mentioned in the survey. 1) Amazon, 2) Flipkart, 3) Paytm, 4)
Myntra & 5) Snapdeal. All these e-commerce websites are market giants & are very famous. Most of the Indian online shoppers use to hang around these websites only. Hence a study on customer's behaviour and experience on these particular websites will be helpful to understand, how to retain the customers.

- Review of Literature

The CUSTOMER RETENTION CASE STUDY is a type of project where we studied and researched the behaviour of online shopper. The motive of this research and analysis was to understand the behaviour & need of the customers so that e-retail websites can improve their services as per their customers and hence can retain their customers. We have done several statistical research as per the data was provided by the organisation, as we have already discussed the data comes from the survey which were taken by the organization direct from the online shoppers.

While research we have found many important notes and have documented step by step in jupyter notebook. Some of key points I am going to discuss here.

Since now-a-days internet is in every hand that is a very big reason why online shopping comes into trend. From the research we have found that online shopping is very famous among the females. Females are more online shopper than the males. 21-30 years female are high in numbers who do online shopping. Urban people are more online shoppers. One of the reason of doing online shopping is that people gets to see varieties of products of their need at one place with full descriptions, they don't need to move anywhere and at the same time sometimes they gets offers and deals which makes online shopping a great experience. People use to get attract to those e-commerce websites which provides them privacy, financial security, convenient payment methods, monetary benefits, online

assistance and a very important things trust. Also we found that content on the websites their graphics, detailed information matters a lot people frequently bent towards those websites who provides all these services and facilities. At the last we have done research on, some comparisons of 5 e-commerce websites on different topics like their services convenience provided by them different features and many more and we have found that Amazon and Flipkart are such an e-commerce companies people are attracted towards. At the end I can say from my experience of the research I have done on customer retention case study that people wants value for money those will provide them will retain them.

- Motivation for the Problem Undertaken

  This CUSTOMER RETENTION CASE STUDY project was provided to me by FLIP ROBO TECHNOGY during my internship with the company. This project is my first analytical project from which I have to derive the various conclusions of the research and come up with a detailed analysis results on the case study. The objective of doing this project was to gain information as much as possible on every aspects of online shopping, how the behaviours of customer get changes, affects with little things and how we can control them by studying these behaviour. Also use all the analytical skills, techniques and tools of data science so that I can grow my analytics skills in the field of data science. Studying the customers behaviour and their actions on various aspects at every steps was fun and it was giving information at every steps, this step by step analysis of each column and coming up with a result and solution was became the motivation of the project.

Analytical Problem Framing

- ## Mathematical/ Analytical Modeling of the Problem

  This project was taken care of under the Python programming Language's Jupyter Notebook, and all the mathematical, statistical and analytics modelling were done which comes under python's various libraries. The data which we got was in excel format and to read this in python we have used the Pandas Library, for the mathematical modelling we have used the Numpy(Numeric Python) Library. For the statistical and analytical modelling and visualization of the data of each columns we have used the Python's Matplotlib and Seaborn libraries, all the project's analysis were done under these tools only.

- ## Data Sources and their formats

  The data was provided to the FLIP ROBO TECHNOGY by their client to do a detail research and analysis. During the internship this data was provided to interns to work upon so that interns can get experience of real life projects. We have already discussed that this data was originated by the organization by taking a survey on e-commerce market. A total of 269 people had taken the participation on this survey a total of 71 questions were asked to the participants and their answers were recorded for research and analysis. The data was recorded in Microsoft Excel format and was used in excel for further analysis.

- ## Data Preprocessing Done

  The excel file of the data which were provided us was already sorted, we had not found any missing values or disrupted data in the dataset hence we didn't have to do any cleaning in the dataset to analyse it further. The data had 71 columns and 269 records of people, all of the data in the columns were categorical data and in string format (Object datatype) only the pin code column's data was in integer format. So we only changed the Pin codes data from integer to string format. From column no. 48 to 71, these column were the review questions of the e-commerce websites, to answer those questions people would have to choose the best or worst websites

according to questions, from the 5 websites, according to their experience. People had options to choose one or multiple websites if they feel so and people had done the same which were making the unorganised categories in the column of different combination of 5 websites. So instead of using these unorganised categorical column we have made 5 column for 5 websites of each question having the direct answers in Yes or No related to each websites and deleted the original column so that we can analyse the direct review of the websites by customers. By doing this we have added 120 extra columns in the data set and deleted 24 original column from the dataset.

- Data Inputs- Logic- Output Relationships

  Since we have discussed that the data of customer retention case study were recorder by organizing a survey. The data that was collected by the responders in a survey form were most likely a choose option type of form in which the options were given to them to choose and these options are either statement, sentence or words that is why these data are in string format hence the data input in the dataset is also categorical and object data type. When we analyse the data with the help of statistical tools like Matplotlib and seaborn we have different concepts to use in according to data, datatype and it's format. Since all the columns are in categorical form we used count plot method to visualise these data along with Panda's value counts method. In Seaborn and Matplotlib the input data will come up in output in the same format as we provide them. Some time we have to change them like we did in Pin Code columns to better understand and plot it in seaborn.

- State the set of assumptions (if any) related to the problem under consideration

  We have taken assumptions in last 24 columns from 48 to 71. All these columns were either the statements or questions about the e-commerce websites and the responders had to choose websites according to the statements or the questions, responders had the options to choose 1 or more than 1 websites, if the feel that the statement or question is true for all the websites they could have choose all the websites as per their views and experience with the websites. When we analysed these data columns we

found that the questions of statements was for 5 websites but the columns had more than 5 categories of the combination of the websites because many of the responders had chosen multiple websites so to sort this out we have made separate column for each websites of each statement or question so this way we have created 5 column of 5 websites for each statements having direct answer in (Yes and No) by assuming that if responders had not chosen any website in their answers for that website column we marked their answer as No and if chosen any website in the answer for the statement or the question for that website's column it was marked as Yes. This way we had direct analysis of each questions and statement to the each website.

- Hardware and Software Requirements and Tools Used

In 21st century the data is being recorded digitally and is being used digitally as well hence to analyse these data or to use in any case we need computer and computer knowledge to work upon. Hence we used Computer and some of computer's hardware. Talking about the software, we are into data analysis and it is a part of data science which is done with the help of programming language because sometimes the data are very large in size which we cannot handle in any other software. Here we are using Python Programming Language in Jupyter Notebook which is available in Anaconda software. We have used some library and packages to work upon the project. To read the dataset in the python programming language we have used Pandas Library and all the understanding of data structure data types etc. we have taken care of with the help of Pandas Library. We have used the Numpy library for some basic mathematical function and its implementation in the data set. To do the Statistical analysis of the data set and without it data analysis can't be completed. We have used Matplotlib and seaborn library package and with the help of these libraries we visualized the data of each column and understand the story behind the data.

Model/s Development and Evaluation

Identification of possible problem-solving approaches (methods)

After reading the dataset through pandas library at the very first & very basic we went for finding the basics of the dataset. We found that dataset has no cleanliness required all the columns and rows were fully filled with the genuine data, no missing values were there. We found that the data has 269 rows and 71 column. After completion of basic research about the dataset we were ready for the data analysis. We found that all the columns in the dataset is categorical hence we immediately realize that count plot method would be the best fit for the data analyzation. Count plot with value counts method of pandas gave us the major information like the number of males and females, which city has highest online shoppers how many people are going with which categories, things like that because in this customer retention case study we had to study the customer behaviour and what are things towards majority of people are going with, because that is the only way to understand the pattern of customer's behaviour and implementing the similar approach we can retain the customers.

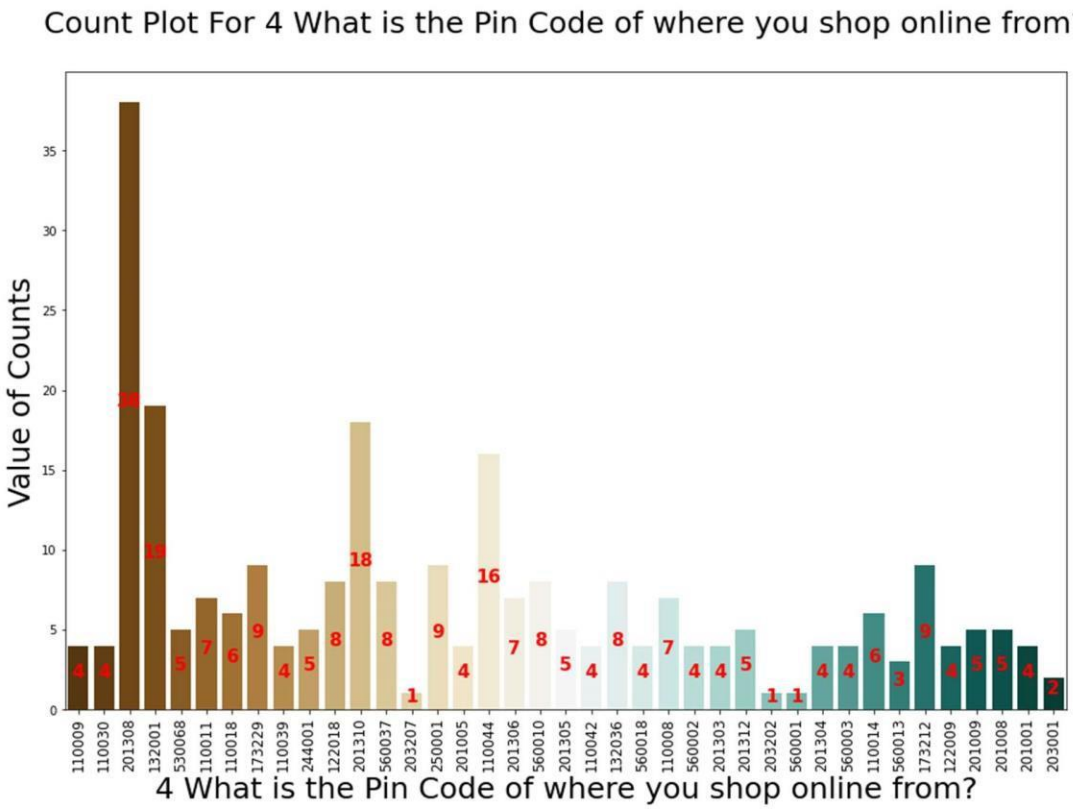
The very first problem came when we were plotting the count plot of the pin code column it was giving some key error for the pin code 110008 than we realise that the data type of this column is integer we assumed that maybe changing in data type would help hence we changed the data type the column from integer to string and the problem was solved.

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

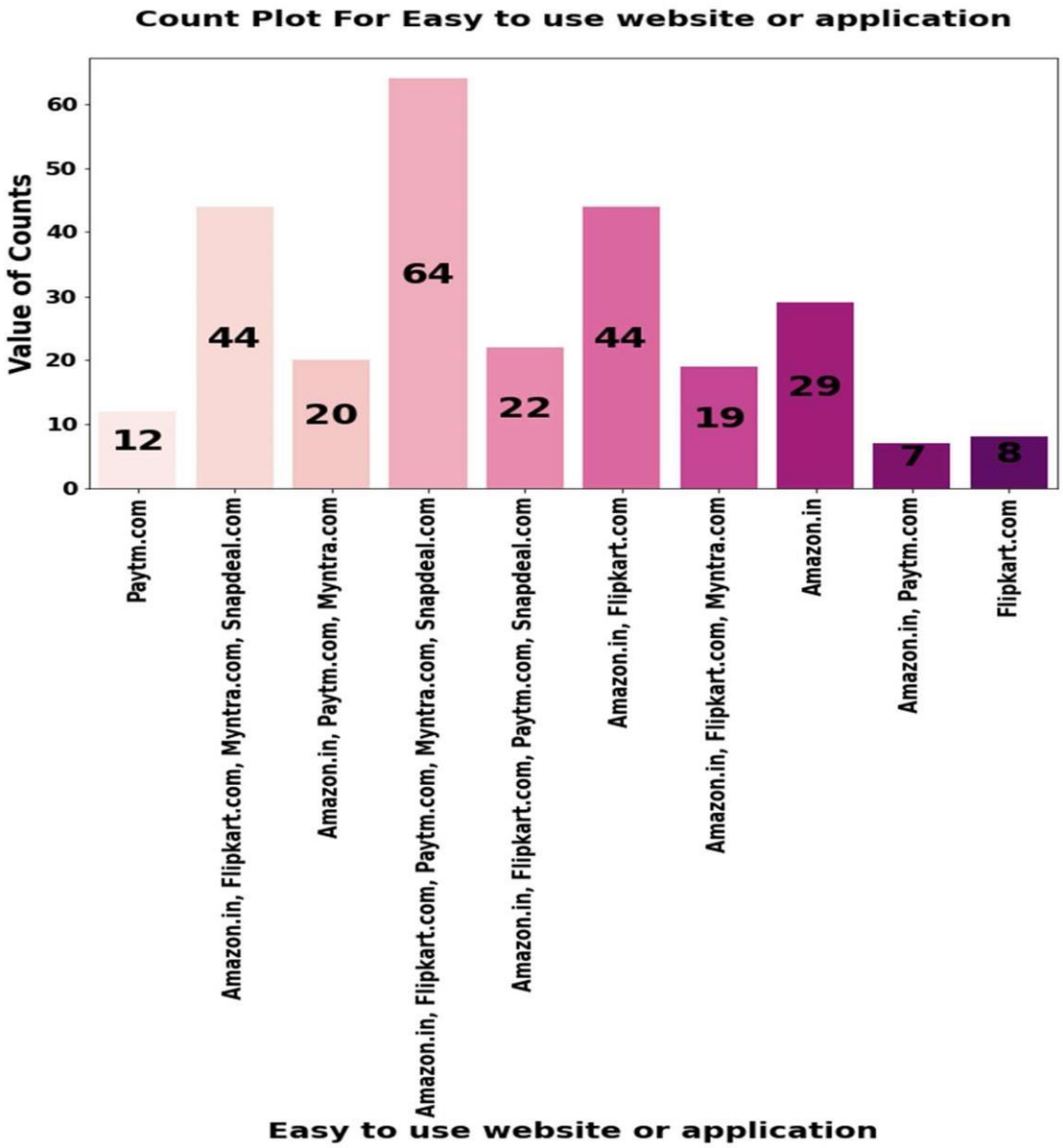
```
In [12]: ting pin code data into string data type
         What is the Pin Code of where you shop online from?']=data['4 What is the Pin Code of where you shop online from?'].astype('str')

In [13]: data['4 What is the Pin Code of where you shop online from?'].unique()

Out[13]: array(['110009', '110030', '201308', '132001', '530068', '110011',
                '110018', '173229', '110039', '244001', '122018', '201310',
                '560037', '203207', '250001', '201005', '110044', '201306',
                '560010', '201305', '110042', '132036', '560018', '110008',
                '560002', '201303', '201312', '203202', '560001', '201304',
                '560003', '110014', '560013', '173212', '122009', '201009',
                '201008', '201001', '203001'], dtype=object)
```

Also before that we wanted to plot all the categorical column's count plot in one go with the help of python's for loop function. But we realised that the size of column's name were very large and some unwanted spaces were in there which were causing the image size of the plots. We could have renamed all the columns, reduced its size and could have plotted it but there were 71 columns and renaming all would have taken time so instead of renaming we individually plotted each and every column one by one step by step and the benefits for doing so was we were able to concentrate on each and every column's data and it's story and to come up with some conclusion.

```
In [14]: plt.figure(figsize = (15,10))
col_name = '4 What is the Pin Code of where you shop online from?'
values = data[col_name].value_counts()
index = 0
ax = sns.countplot(data[col_name], palette = 'BrBG')

for i in ax.get_xticklabels():
    ax.text(index, values[i.get_text()]/2, values[i.get_text()],
            horizontalalignment = 'center', fontweight = 'bold', fontsize = 15, color = 'r')
    index += 1

plt.title(f'Count Plot For {col_name}\n', fontsize = 25)
plt.xlabel(col_name, fontsize = 25)
plt.ylabel(f'Value of Counts', fontsize = 25)
plt.xticks(rotation = 90, fontsize = 12)
plt.show()
```



We followed the same procedure till the column no. 47. From column no. 48 the type columns were changed and it was for the analysis of 5 major Indian e-commerce websites and what people think about that. We found that columns have unnecessary categories of different combinations of the websites which were chosen by the responders according to their views hence we decided to change the columns format from website categories to Yes and No category. These Yes and No categories were the direct answers of the statements or question related to each and every websites. We used Python's Oops Oriented Lambda function to change the category in Yes and No and used dot map function python's to extract the data from each rows.

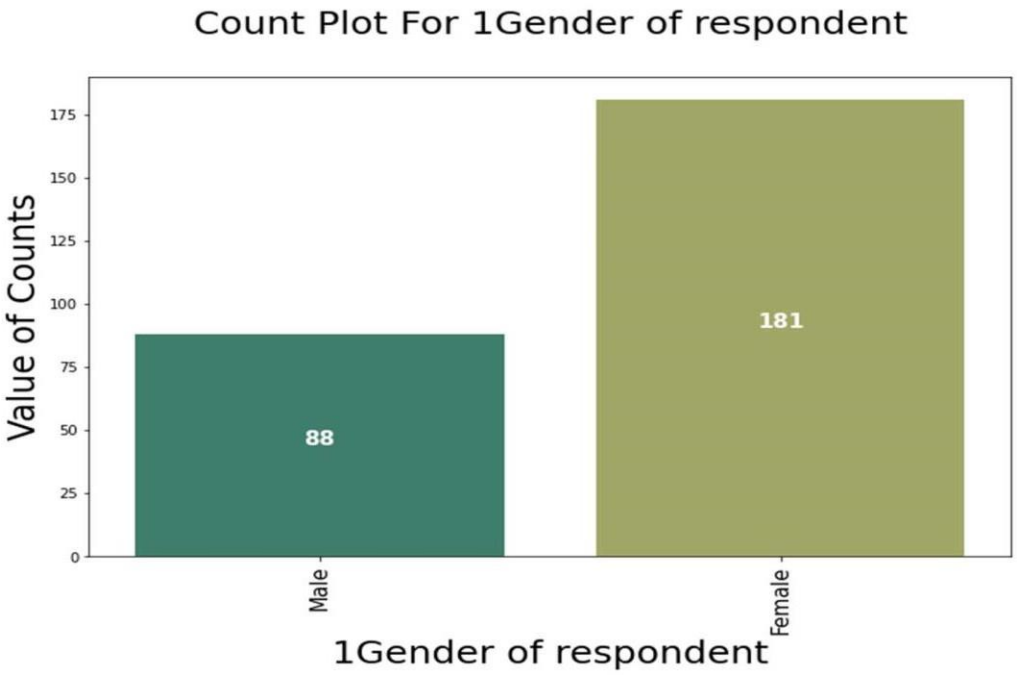


In [72]:
```
# making 5 new column of 5 websites stating that if a particular website is easy to use or not
website1 = 'Amazon.in'
website2 = 'Flipkart.com'
website3 = 'Paytm.com'
website4 = 'Myntra.com'
website5 = 'Snapdeal.com'

# making new columns mentioning a particular website having data in YES or NO as the answer if a particular website is easy to us
data['If Amzon.in is easy to use or not?'] = data['Easy to use website or application'].map(lambda x : 'Yes' if website1 in x els
data['If Flipkart.com is easy to use or not?'] = data['Easy to use website or application'].map(lambda x : 'Yes' if website2 in x
data['If Paytm.com is easy to use or not?'] = data['Easy to use website or application'].map(lambda x : 'Yes' if website3 in x el
data['If Myntra.com is easy to use or not?'] = data['Easy to use website or application'].map(lambda x : 'Yes' if website4 in x e
data['If Snapdeal.com is easy to use or not?'] = data['Easy to use website or application'].map(lambda x : 'Yes' if website5 in x

# droping the original column from the data as now we don't need.
data = data.drop(columns = ['Easy to use website or application'], axis = 1)
```

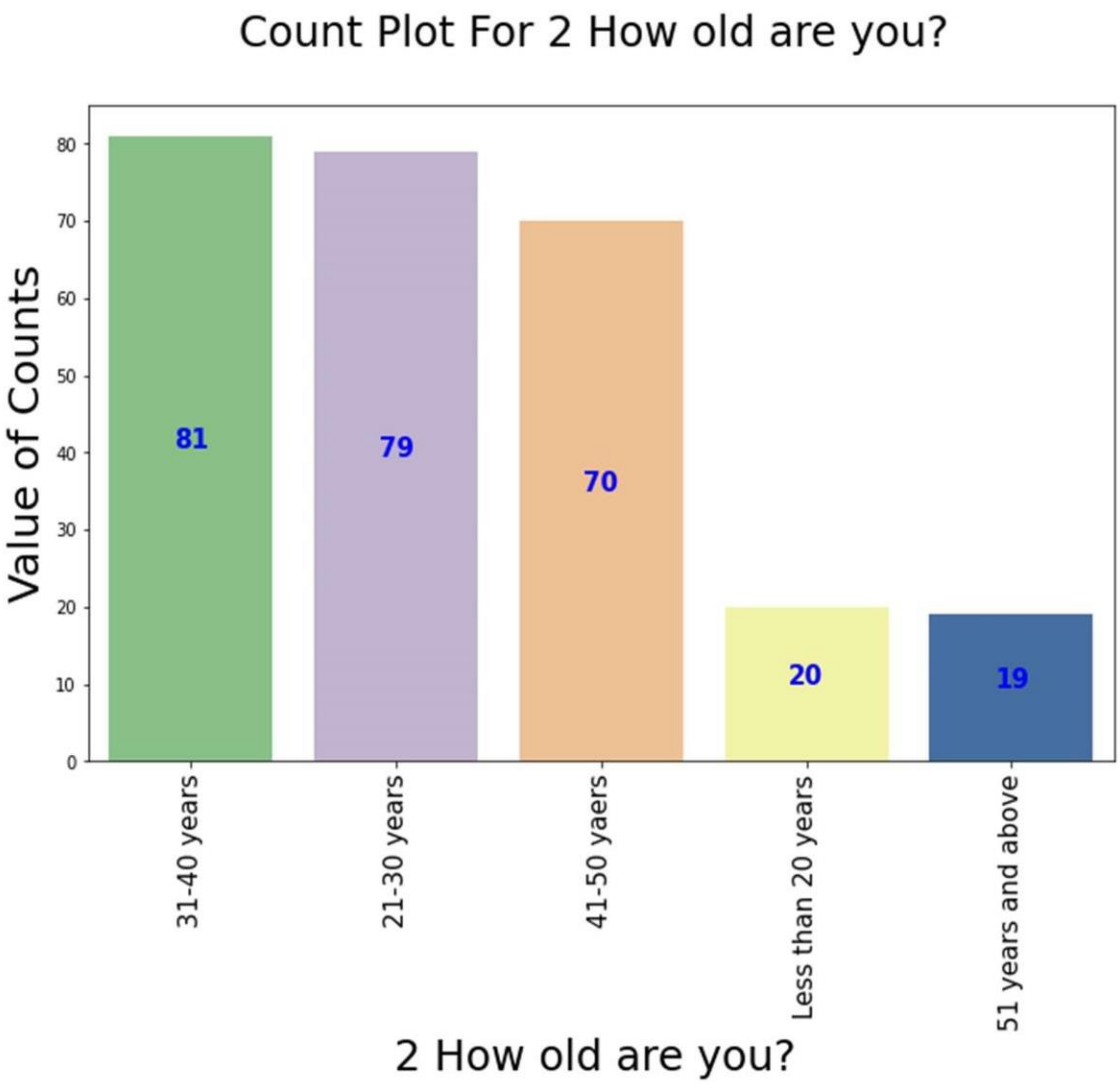
Successfully created 5 new columns out of `Easy to use website or application` column and deleted it after extracting the usefull column and data from it. Let's analyze the columns which we have created.

Visualizations

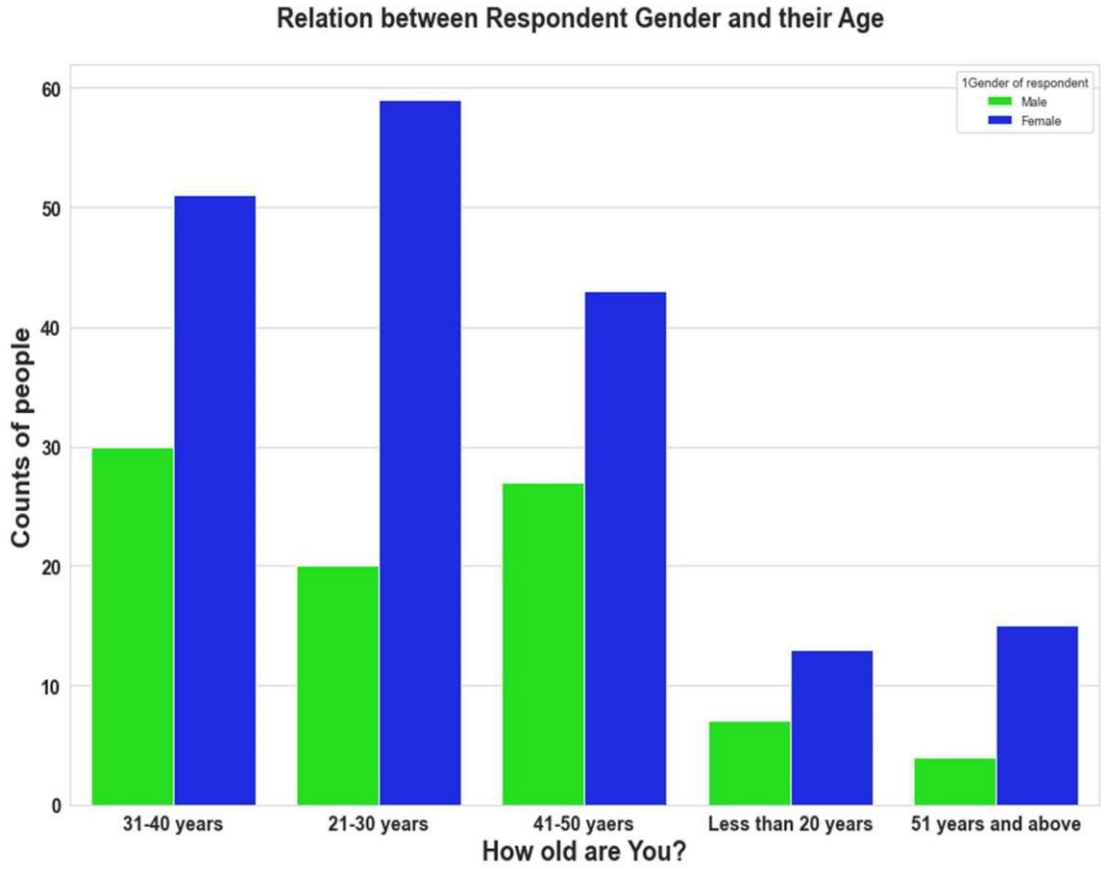
We had used Only Count Plot for our Data Visualization the reason for using Only Count Plot was that all the columns were categorical column and the requirement of the case study was to understand the pattern of the customer behaviour for the different scenarios hence for this we needed to observe the data distribution for the categories that column were holding, which were clearly stating that majority of people are going with what. This way we have done all the analysis. We also used the Count Plot for the checking of the relations between 2 similar columns Like Pin Code column and City Column.



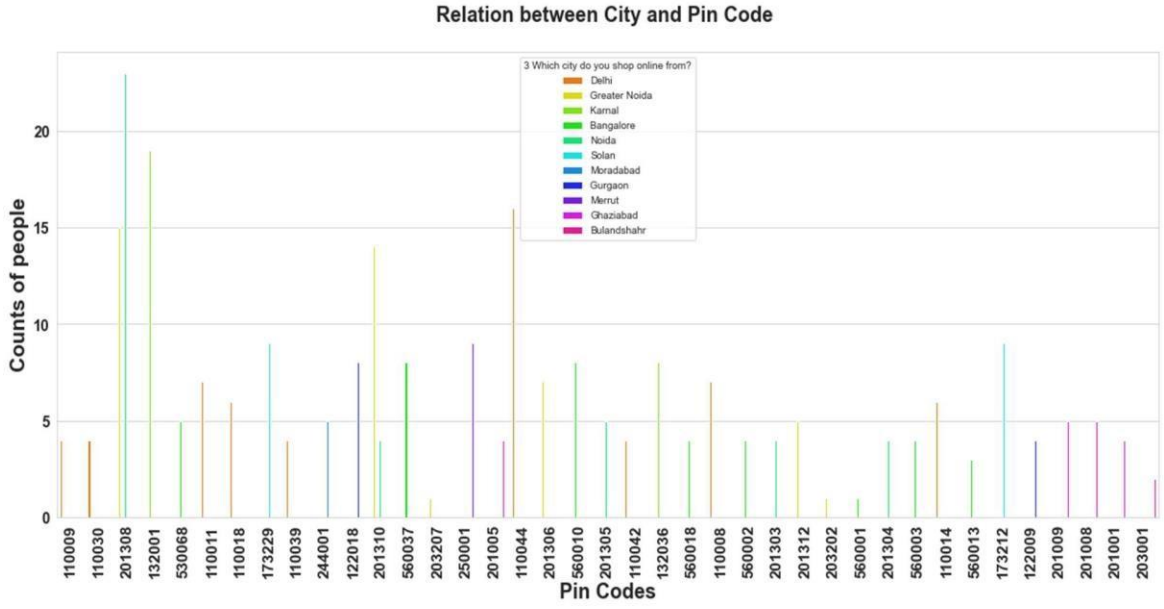
Like in above Count Plot of the gender of respondents we can clearly see the distribution of the genders. The number of female respondents are much higher than the Male respondents. This is giving us the clear story that women likes to shop more than the men and if e-commerce will focus on their female customers and provides them the necessary attention this could lead their more sales and retention of their customer.



Here we can see in above image the age distribution of the shoppers and we are clearly seeing that 21-30 & 31-40 age range people are the highest online shopper and this makes sense because these are age categories who use to be young and at the same time independents hence they spent more on their selves' focus on these age category people will lead to retention of the customer.



The above count plot of the relation between Age and Gender of the respondents are telling us that 21-30 age females are more online shopper than anyone else and like same 31-40 years males are more online shopper in males category. Females in all age range are more online shopper than males.



The above relational Count Plot of City and Pin Codes are describing clearly that Pin Codes with same initials belongs to the same city. For eg. We can see that red colour bar represents the Delhi city and we can see that all the pin codes which starts (1100__) are having the red bars. So it's proved that the pin codes which starts with 11 are belongs to Delhi city.

Interpretation of the Results

From the analysis of the data we interpreted many of thing from all the columns here I am discussing some of the key interpretations.

Females are more online shopper than the males so focus on female customers more will lead to benefit to the e-commerce companies. He have also interpreted that the 21-40 range people are highly active on online purchase may be because these people are independents earning good money and taking their decisions and taking care of their necessity and needs by tem selves.

Noida and Delhi has the highest online shoppers, so here we interpreted that most of the online shoppers belongs to urban area also out of 11 cities 5 cities were of Delhi NCR. Which are the most urban cities in India.

From the analysis we also got to know that people won't compromise in certain things like Privacy, Financial securities, flexible payment methods etc. those who lacks in these facilities people won't even consider those channels.

We have seen that people are strongly agree on certain things like content on the websites must be easy to read and understand, user interface must be user friendly, similar product high lightening is a must for product comparison, payment method should be convenient and smooth, customer service and trustworthiness has no compromise.

Customers were asked and given certain questions & statements about 5 major online shopping websites. Amazon.in, Flipkart.com, Paytm.com, Myntra.com, Snapdeal.com. And we found that people's no. 1 choice is Amazon.com after that Flipkart.com's response was also good but Paytm.com & Snapdeal.com were totally neglected by the customers.

CONCLUSION

Key Findings and Conclusions of the Study

From the analysis of the CUSTOMER RETENTION CASE STUDY we have concluded the certain factors that will be the key factors to the success of an e-commerce websites and it will gain the activations & retention of customers.

The websites or the e-commerce companies which will provide the key services like convenient payment methods, flexible payment options, on time delivery, better replacement and refund policies. Systems like better visual graphics, better user interface, multiples of varieties on product, suggestions of better product or offer, privacy guarantee, Financial securities, trusts of not being looted or not getting victim of frauds, better offers and discounts, loyalty points etc. Those companies will provide all of the above categories of service benefits will be the company whol gain the new customers and retain the old customers.

In the data set there were 24 columns in which customers had given their views & reviews in the comparison of 5 major websites or companies which were Amazon, Flipkart, Paytm, Myntra & Snapdeal. From the analysis of these comparison by the customers we found that Amazon is the ideal e-commerce company and most people are attractive towards amazon followed by the Flipkart. While Snapdeal and Paytm are the least choice of the people, we found that at some instance Myntra's performance was good but overall it was not that good.

At the last question of recommending the websites to the family and friend we noted that Amazon is most recommended website, Flipkart were also recommended at a good level but other than these all three websites recommendations were very poor. Which again proves that the quality which we mentioned above was found in Amazon at highest level and hence the Amazon is this much famous and favourite website.

- Learning Outcomes of the Study in respect of Data Science

  While working on the CUSTOMER REYENTION CASE STUDY project, I have learnt several things in respect of Data Science. I learnt that Data analysis is the complementary of Data Science, without Data analysis Data Science & it's models are not possible. The fundamentals of any dataset is very important to know before working on any projects of any datasets, like the data structure, data shape, data types etc. sometimes we have to adjust, convert from one form to another, add, remove & replace these fundamentals as per the requirements & objectives of the projects.

  Statistical visualization is the very important factor in data analysis. Statistical visualization clears the pictures of data and it conveys the story to us what data set wants to tell us. The power of visualization techniques are such that while using visualization we automatically use to get idea what we need more to clean the data pre-process the data & get ready the data for its final objectives.

- Limitations of this work and Scope for Future Work

  As soon as the time passes the value of this data will be reduced because the companies comes up with new things every day, companies use to change and develop their process, services according to the situations with new ideas just to attract more and more customers. As the time passes new trends and new visions come and also with increasing technologies innovations are booming like anything and so the old thing feds up and companies develops themselves accordingly hence the old data and analysis based on these data loses its value day by day hence the value of the solution provided will no longer be useful to implement.

  Future is more bright for the e-commerce companies as the new technologies are coming every day the needs of people are getting change, people wants to be in comfort, economy of the country is growing bigger and bigger hence there should be new ideas and their implementations to please the customers and their experience. Today's world is of Artificial Intelligence, the software and application works nowa-days as per the user.

With the help of Machine Learning applications itself gathers data, learns from the data and self-old behaviour and works as per user thinks this will tells the companies what is the need of markets And follows the same path will lead to grow the business.

Thank You !!