Customer Retention Analysis

Submitted by:

Payal Gupta

ACKNOWLEDGMENT

Gratitude takes three forms-"A feeling from heart, an expression in words and a giving in return". We take this opportunity to express our feelings.

I express my gracious gratitude to our project guide **Sapna Verma**, SME of Flip Robo Technologies, for his valuable guidance and assistance. I am very thankful for his encouragement and inspiration that made project successful.

I express my gracious gratitude to our Trainer **Dr. Deepika Sharma**, Training Head of DataTrained - Data Analytics, Data Science Online Training, Bengaluru, for her valuable guidance and assistance throughout the PG Program. I am very thankful for her encouragement and inspiration that made project successful.

My special thanks to all the **Instructors** and **Subject Matter Experts** of DataTrained - Data Analytics, Data Science Online Training, Bengaluru, who helped me during the live sessions and during doubt clearing scenarios from which I received a lots of suggestions that improved the quality of the work.

I express my deep sense of gratitude to my family for their moral support and understanding without which the completion of my project would not have been perceivable.

INTRODUCTION

- Business Problem Framing

Our aim is to perform extensive data analysis on a given dataset and produce valuable insights that will help in customer retention.

- Conceptual Background of the Domain Problem

Capturing the audience's attention and converting them into paying customers is paramount for online businesses in a fiercely competitive e- commerce market. Driven by the need to succeed, more and more businesses are turning to artificial intelligence (AI) and data science applications in e-commerce to maximize customer retention and conversion.

- Motivation for the Problem Undertaken

Data science involves proficiently using underlying fields such as statistics, mathematics, and programming to develop an understanding of trends and patterns in structured and unstructured data. This data is deciphered into actions for defining customer acquisition and retention strategies by e- commerce businesses. Since increasing conversion propensity is the bottom line for e-commerce businesses, the insights on user behavior provided by customer analytics help target the right set of customers.

Analytical Problem Framing

- Mathematical/ Analytical Modeling of the Problem

    1. Data was cleaned by removing redundant values in the dataset.
    2. Column names were large, hence were changed to short names for convenience in analysis.

- Data Sources and their formats

    What are the data sources, their origins, their formats and other details that you find necessary? They can be described here. Provide a proper data description. You can also add a snapshot of the data.

    1. A csv file containing two sheets, one was encoded and the other was a raw data.
    2. It had 269 rows and 71 columns.

	1Gender of respondent	2 How old are you?	3 Which city do you shop online from?	4 What is the Pin Code of where you shop online from?	5 Since How Long You are Shopping Online ?	6 How many times you have made an online purchase in the past 1 year?	7 How do you access the internet while shopping on-line?	8 Which device do you use to access the online shopping?	9 What is the screen size of your mobile device? \t\t\t\t\t\t	10 What is the operating system (OS) of your device? \t\t\t\t	...	Longer time to get logged in (promotion, sales period)	Longer time in displaying graphics and photos (promotion, sales period)	Late declaration of price (promotion, sales period)
0	Male	31-40 years	Delhi	110009	Above 4 years	31-40 times	Dial-up	Desktop	Others	Window/windows Mobile	...	Amazon.in	Amazon.in	Flipkart.com
1	Female	21-30 years	Delhi	110030	Above 4 years	41 times and above	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	...	Amazon.in, Flipkart.com	Myntra.com	snapdeal.com
2	Female	21-30 years	Greater Noida	201308	3-4 years	41 times and above	Mobile Internet	Smartphone	5.5 inches	Android	...	Myntra.com	Myntra.com	Myntra.com
3	Male	21-30 years	Karnal	132001	3-4 years	Less than 10 times	Mobile Internet	Smartphone	5.5 inches	IOS/Mac	...	Snapdeal.com	Myntra.com, Snapdeal.com	Myntra.com
4	Female	21-30 years	Bangalore	530068	2-3 years	11-20 times	Wi-Fi	Smartphone	4.7 inches	IOS/Mac	...	Flipkart.com, Paytm.com	Paytm.com	Paytm.com

fig 1 : Sample Dataset

- Data Preprocessing Done

    1. There were few duplicate values in customer reviews. For example - Dis-agree and Diasgree. So made single value as both have the same meaning.
    2. Converted all long columns to short column names for easier access and convenience during analysis.

- Data Inputs- Logic- Output Relationships

    1. The data was cleaned.
    2. Inferential analysis is done.
    3. EDA was performed by creating valuable insights using various visualization libraries.

- Hardware and Software Requirements and Tools Used

Hardware : 4GB Ram, Core-i5,8th Gen Software : Following libraries were used :

1. Pandas : To read the csv file, to convert the data into dataframe, for description and data type of data, to save the final output.
2. Matplotlib : To plot the graphs
3. Seaborn : To plot the graphs
4. LabelEncoder : To encode columns
5. Plotly : To plot the grpahs

Model/s Development and Evaluation

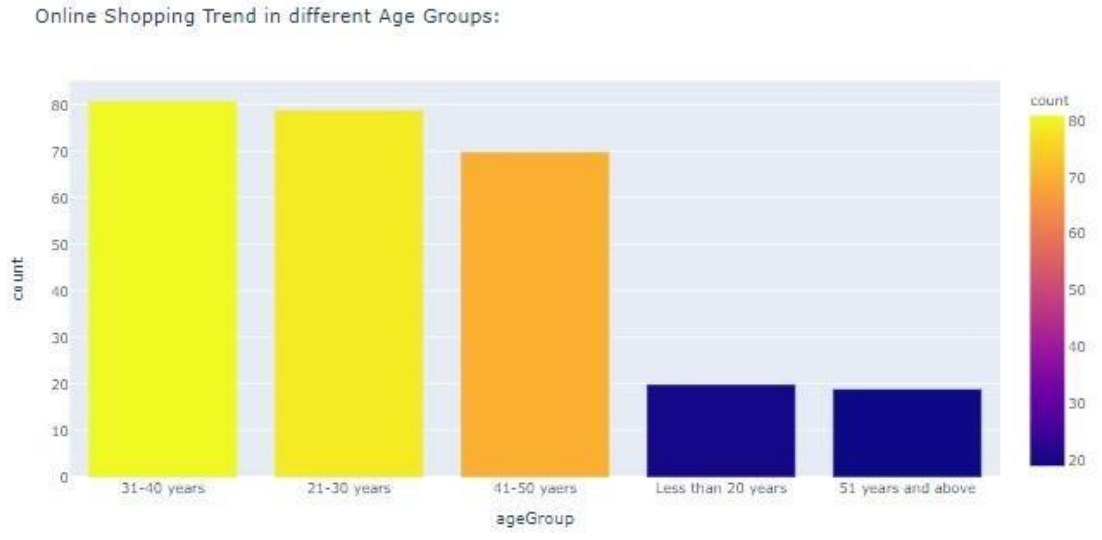
- Identification of possible problem soving approaches

  Firstly I imported all the required libraries, followed by loading the data. Then did a statistical analysis on the dataset and made note of all the cleaning that needs to be done. Analysed the data by performing EDA and tried to answer some of the questions that will help us for Customer Retention.

- Visualizations

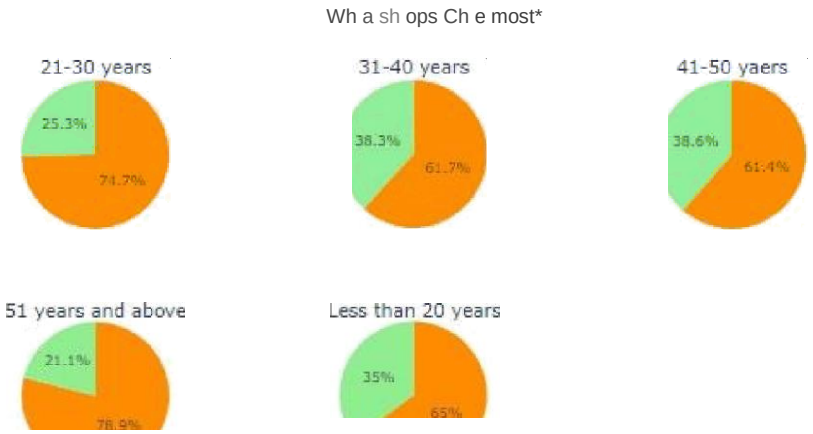
```
|: #Age groups which shops the most :
   age_count = pd.DataFrame(dataset['age'].value_counts()).reset_index()
   age_count.columns = ['ageGroup','count']


   fig = px.bar(age_count, x='ageGroup',y='count',
                color='count',
                title = 'Online Shopping Trend in different Age Groups: ')
   fig.show()
```

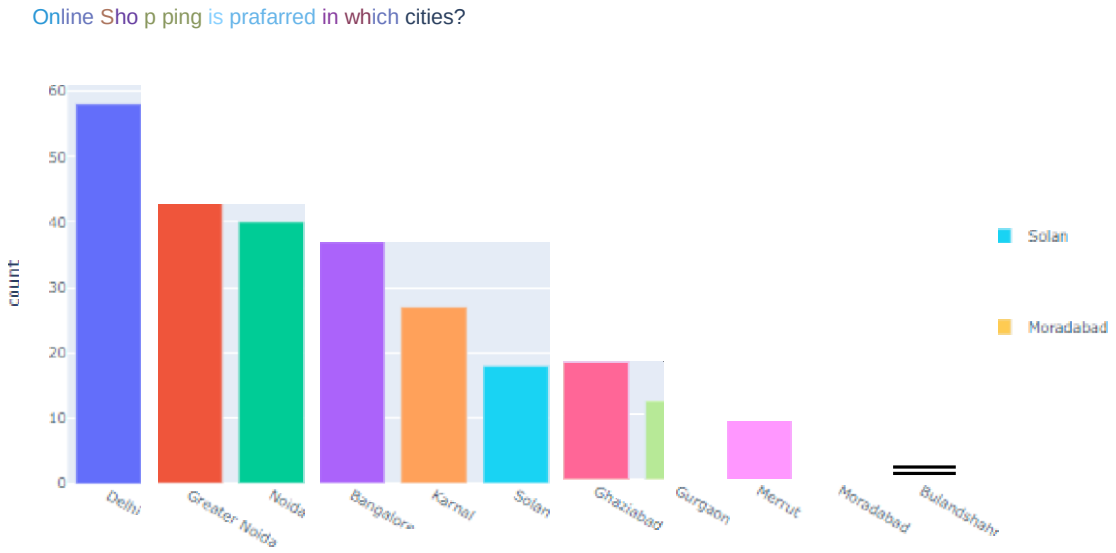


Observations :

- As expected, people who is between 20 to 50 years old, they prefer online shopping more than senior citizens or teenagers.
- We can also assume that, people with 31-40 years will not have much time to go out and shop, hence they will prefer online shopping more.

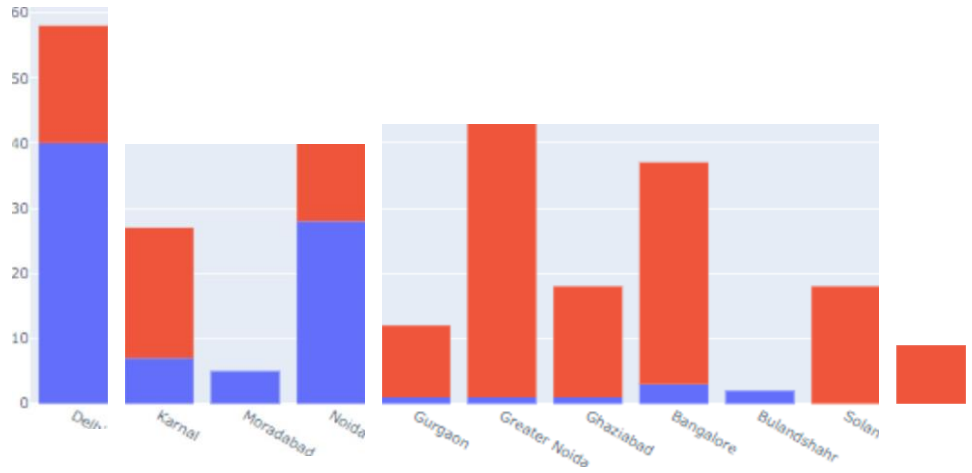


- Wé ca« obj from These pie charts that, in all age groups, mare than 50a of the sample, a is non+en who does more ordine shagping when compared



- De0fi shops If+e most a+ +e. folk›ered by' Nrñda a+d Bnagzdore.

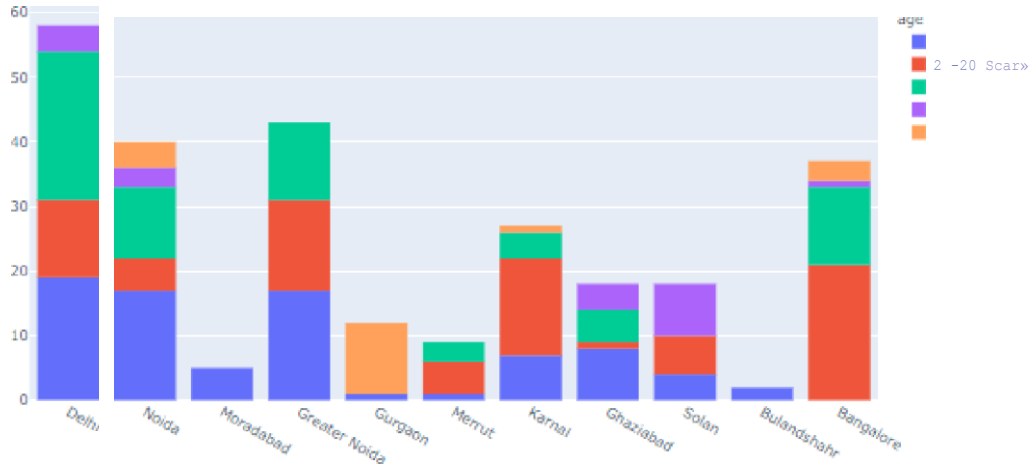
Online shopping in cities based on Gender:



Observations:

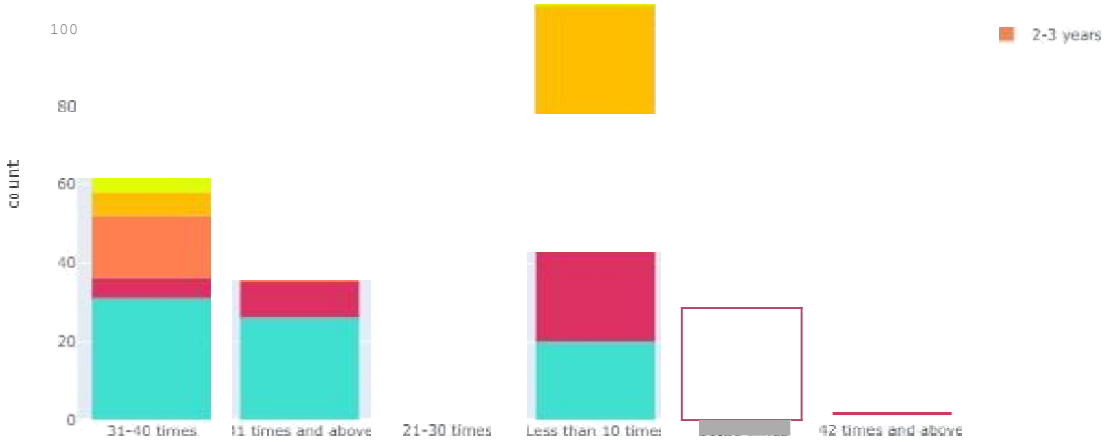
- Surprisingily. in Delhi and Noda, males prefe‹ online shoppng mo•e than females. So we know our tatget audience in Mese cities.right?
- B ulandshahr and Moradabad - both oties in UttarPradesh no female shoppers at all !!

Online shopping in cities based on various age groups :



- In Delhi, people aged b een II-TO yeais, prefér la shop on online more, vrtien compared to oi:her age groups.
- In Bangalore, n is pesple aged beMeen Z1-30 years of age.
- In the resl of the cities. on an average. people aged ber+naen 31-40 years prefer online sl+sppilg. exrept Gurgaon where senior citizens are more

S hoppin q fre quen cy of people in past one year :

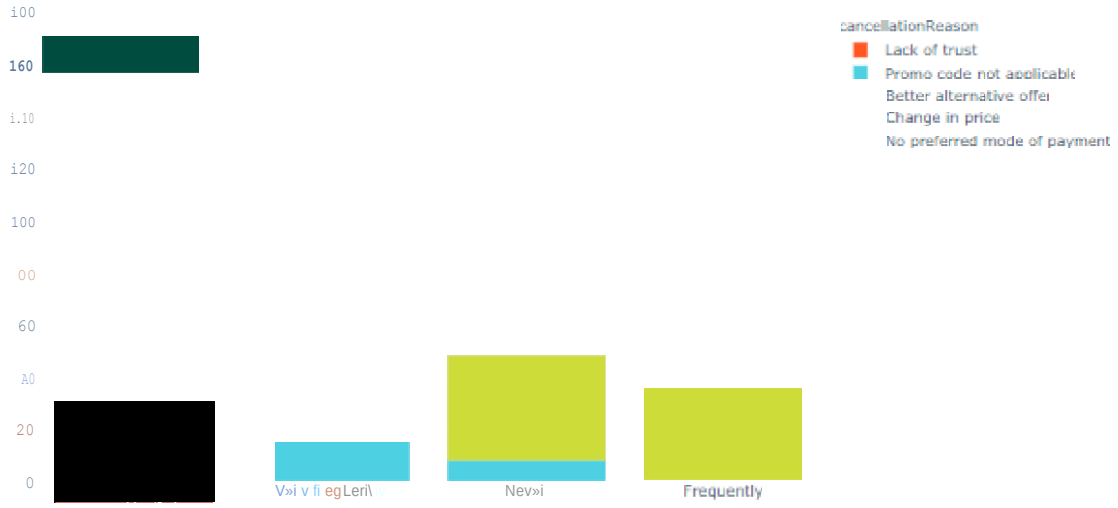


- In tñe past ‹ year, on an average people have stopped approximated 30—40 times.
- From the data. vie can see \ha\, people zcko fave Deen snapping for more thw« 3-4 years are he ones n'ha freqMenlly shops.

Reasons for noC purchasing any prodLrct :

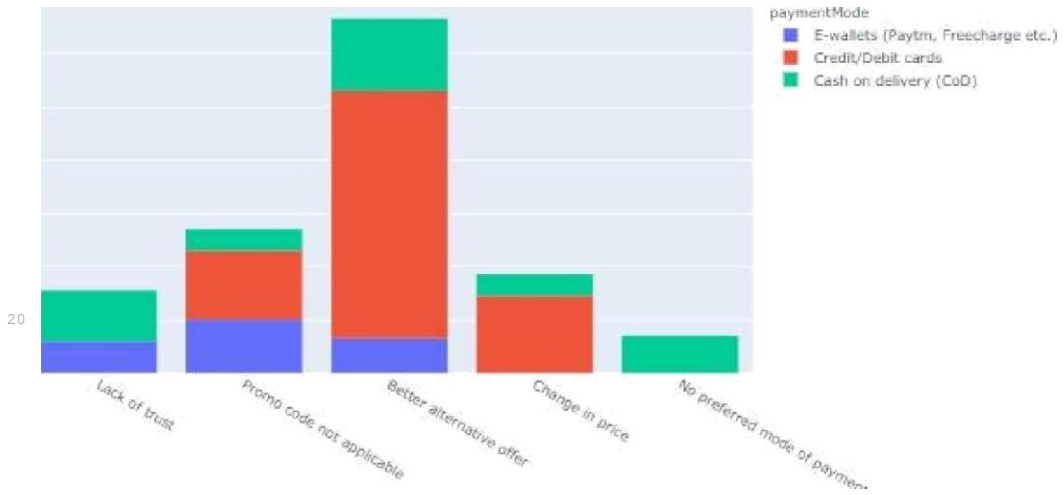
- t'L'e car. observe thaf most of the people. abandceied their car as th.ey were having bear aiieii auue oJfers.
  *eople even abadon tneir casts. because tney do *na!* h ave their pre'lérred mcd= of payment.
  5o whec the preferred oayment mode of people who cancel Jneir c-at was ana Used,* appears that they preferred CoD, but was not available for tbat product, Vert th can also imps that the customer may have scme difñculo/ in musing the ieLsil store?
- Lack of must is aha a reason for can ellég they product

Why a nd how often do people abandon their cartel



ca ncelli ngFrequency

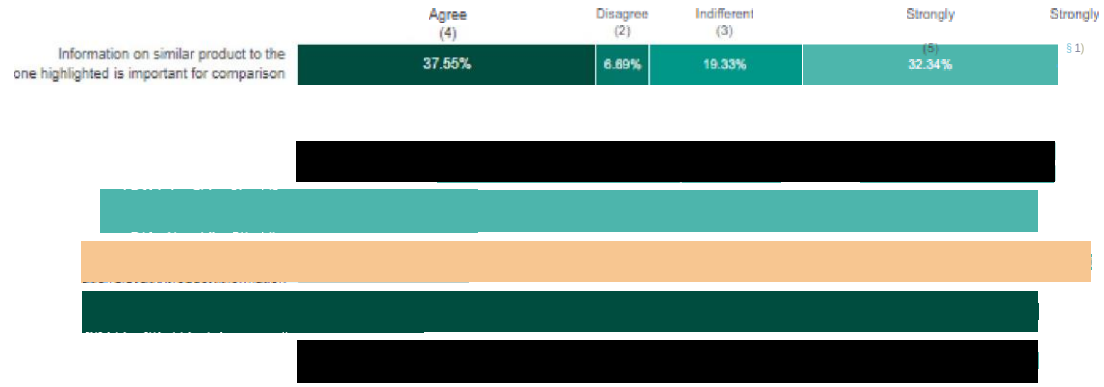
Payment Td o de vs Cancellation Reasons :



czr:cellationRe ason

- to when the preferred payment mode of peogle wño cancel tfieir cart eras analysed , it appears that they preferred CoD. but 'rJas not a'railable for fJ›af product. vih*n nan also imply that the customer nay have scme di1 cull in trusang the refail store?
- Lace: of trust is agso a reason for cancellmg 'her produm. and the paymen• mcde is e-+valle\s, vJn*h they believe must not be reliabe?

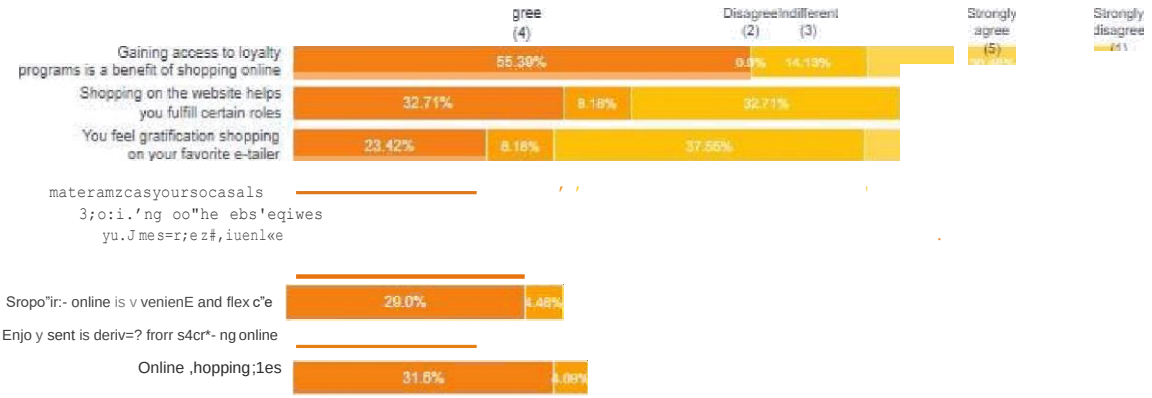
Purchase decisions on online Retail Store are based on following factors :



- We can obau'ze that mostly agree wilh the fact lhat the product they ae pr•chasñ+g fzm Jhe app ar w+KzB+ie, they hope to have complete infazriation regar+:¥ng the product.
- Mcel of them zgjrees that lhe fmctioring+ app e4icierr0y is also a majcs facts wh h he¥ss in enhance user exgerierce while doing online shop|fing.

Purchase.decisions on oniirie Retaii Eitore are. based on following factors :

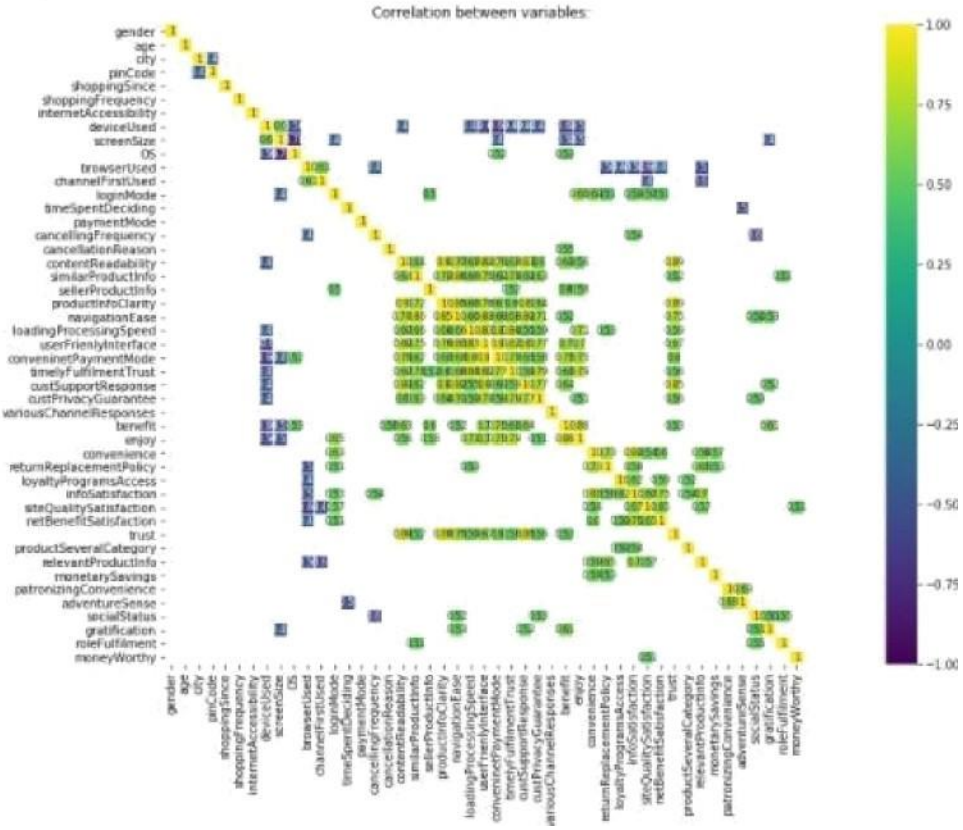
Benefits of Shopping Onli ne :



- To some group of people. ils not nly how the app is vvorldrig. bul \vlial benefits il amves v/ilh the purchase.
- 5Ome people just shop online just because Ihey enjoy il.
- Moslly because, it is convenient and flaible people prefer online shopping.

User Saitisfaction from Online shopping :

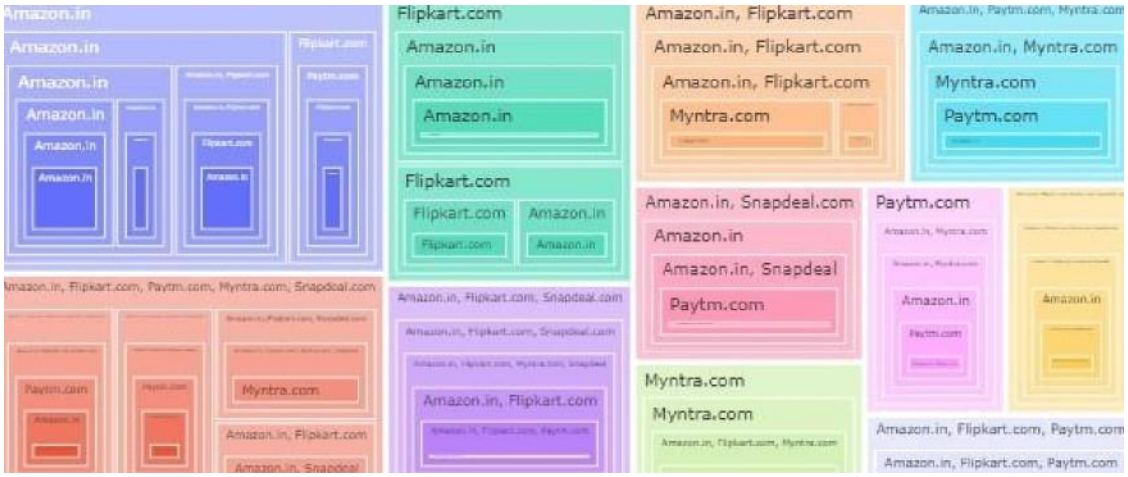
```
Text(0.5, 1, 'Correlation between variables: ')
```



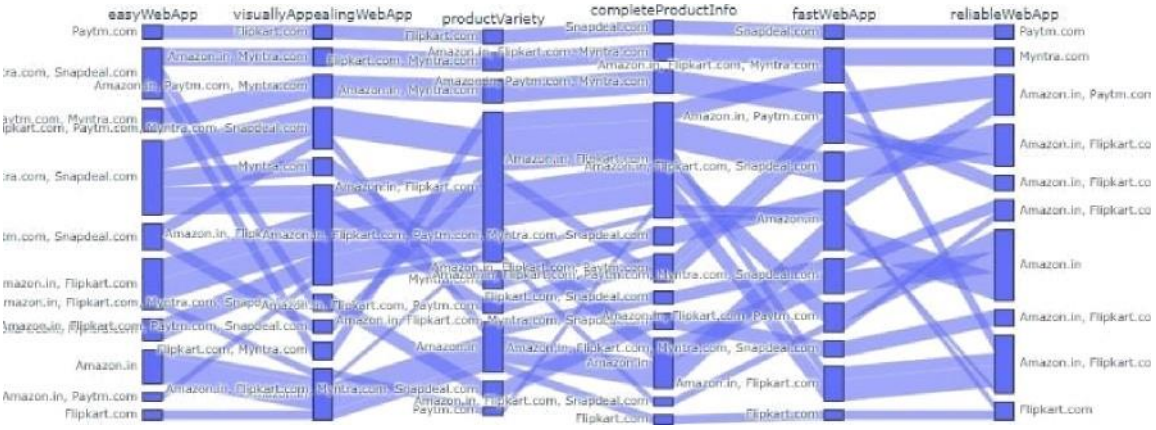
Observation:

- From we data we can see Inat. data relstat \o hovi the person is ac<essi tne app er websif= Yes not maker as a has negative correaaon.
- Customer mention can be done. majorly xdh customer ie ‘——and by finding eut if Ihey are a1isFed vJeh tfie qualify oF product and ezperien ce delivered a Ifiern.

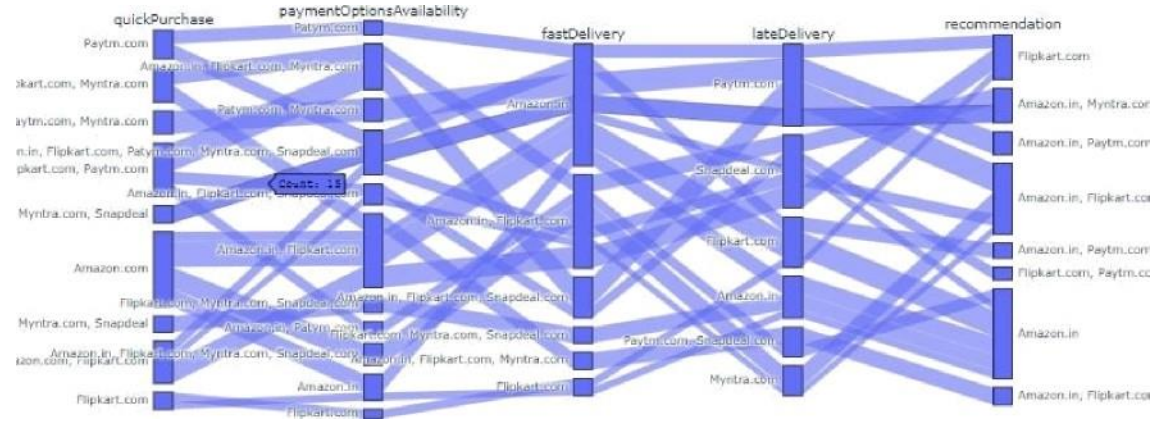
Indian Online Reatil Stores : People's Choice :



Online Retail store Comparisons



Online Retail gtore Comparisons



O bservafio ns:

- With all the data represented above, we can observe Ihat Amazon and Flipkart have extremely good reviews and most of the people prefer these Eyo app
- Myntra is also quite famous in a smaller group of people.
- Snap deal and Paylm is ogled less comparatively.
- People recommend Amazon and Flipkart, majorly.

CONCLUSION

- Key Findings and Conclusions of the Study

1. People who is between 20 to 50 years old, they prefer online shopping more than senior citizens or teenagers.
2. We can also assume that, people with 31-40 years will not have much time to go out and shop, hence they will prefer online shopping more.
3. We can observe from these pie charts that, in all age groups, more than 50% of the sample, it is women who does more online shopping when compared to men.
4. Delhi shops the most online, followed by Noida and Bangalore.
5. Surprisingly, in Delhi and Noida, males prefer online shopping more than females. So we know our target audience in these cities,right?
6. Bulandshahr and Moradabad - both cities in UttarPradesh no female shoppers at all !!
7. In Delhi, people aged between 41-50 years, prefer to shop on online more, when compared to other age groups.
8. In Bangalore, it is people aged between 21-30 years of age.
9. In the rest of the cities, on an average, people aged between 31-40 years prefer online shopping, except Gurgaon where senior citizens are more involved.
10. In the past 1 year, on an average people have shopped approximately 30-40 times.
11. From the data, we can see that, people who have been shopping for more then 3-4 years are the ones who frequently shops.
12. We can observe that most of the people, abandoned their cart as they were having better alternative offers.
13. People even abandon their carts, because they do not have their preferred mode of payment.
14. So when the preferred payment mode of people who cancel their cart was analysed, it appears that they preferred

CoD, but was not available for that product, which can also imply that the customer may have some difficulty in trusting the retail store?

15. Lack of trust is also a reason for cancelling their product.
16. We can observe that, mostly agree with the fact that the product they are purchasing from the app or website, they hope to have complete information regarding the product.
17. Most of them agrees that the functioning of app efficiently is also a major factor which helps in enhance user experience while doing online shopping.
18. To some group of people, its not only how the app is working, but what benefits it arrives with the purchase.
19. SOme people just shop online just because they enjoy it.
20. Mostly because, it is convenient and flexible, people prefer online shopping.
21. From the data we can see that, data related to how the person is accessing the app or website does not matter as it has negative correlation.
22. Customer retention can be done, majorly with customer reviews and by finding out if they are satisfied with the quality of product and experience delivered to them.
23. With all the data represented above, we can observe that Amazon and Flipkart have extremely good reviews and most of the people prefer these two app.
24. Myntra is also quite famous in a smaller group of people.
25. Snap deal and Paytm is opted less comparatively.
26. People recommend Amazon and Flipkart, majorly.

- Learning Outcomes of the Study in respect of Data Science

For any given dataset, the EDA process is extremely important as well as beneficial in order to build an effective model. Visualizationshelps us to analyze the data patterns, outliers, and various information of the events that occurred. It will also help in the data cleaning process. Data cleaning and manipulation is the next big step which will bring out the best in the data. While working with this dataset, initially domain knowledge was a challenge. I read more articles on the internet, understood the problem carefully andtried various ways to bring out the best method and executed this project.

- Limitations of this work and Scope for Future Work

The dataset could have had more information regarding users, regarding the purchase they made. Time could have been provided and prices could also have been mentioned. It could have helped to provide better and more detailed information regarding thecustomers and helped us to retain the customers.

THANK YOU