Project Report on Data Analysis of **Customer Retention** in Ecommerce Sector

Internship work carried out at

FlipRobo, Bangalore

By

Ramesh Pyru

Project report submission as part of internship under the Supervision of **Ms.Khusboo Garg, SME FlipRobo, Bangalore**

ACKNOWLEDGEMENTS

I would like to thank FlipRobo team for their direction, assistance, and guidance. In particular Ms. Khushboo Garg (SME Flip Robo), her recommendations and suggestions have been invaluable for the project.

I also wish to thank "Data trained" which provided the opportunity for internship at FlipRobo.

Special thanks to my family & friends who helped me in many ways.

References use in this project:
1. SCIKIT Learn Library Documentation
2. Blogs from towardsdatascience, Analytics Vidya, Medium
3. Andrew Ng Notes on Machine Learning (GitHub)

TABLE OF CONTENTS

CHAPTER 1 – Introduction

1.1 Background

Currently many e-Commerce websites are available to customer's having tough competition for the continuous business growth & revenues where customer activation & retention is critical

Customer Satisfaction is identified as Key factor for repeat purchases & retention of customers .We need to further analyze the data collected from Indian e-commerce customers on how both utilitarian value and hedonistic values affects customer purchase intention

Other success factors □service quality, system quality, information quality, trust and net benefit.

1.1.1   Problem Definition

The Need for e-retail success factors which influence customer Satisfaction .Both Hedonic & Utilitarian values affect the customer repeat purchase retention & need to focus on the combination considering the perceived risks

1.1.2   Stages of Project

- Data Collection

- Data Cleaning--->Missing Values imputation--->Handling Outliers

  - 2.a. Univariate Analysis(Graphical representation-->Histograms (Bar Charts) , Box Plots, Bar &Group charts, Stem and Leaf Plots)

  - 2.b.Bivariate Analysis(Graphical representation-->Pai Plots, Bar Plot,CrossTab,Scatter Plots)

  - 2.c. Multivariate Analysis(Graphical representation-->HeatMap,LinePlot,PairPlot)

- Choosing the Right Statistical Methods

- Visualizing and Analyzing Results

The below figure depicts the different connectors/variables which influence the Customer retention Or Repeat purchase intention
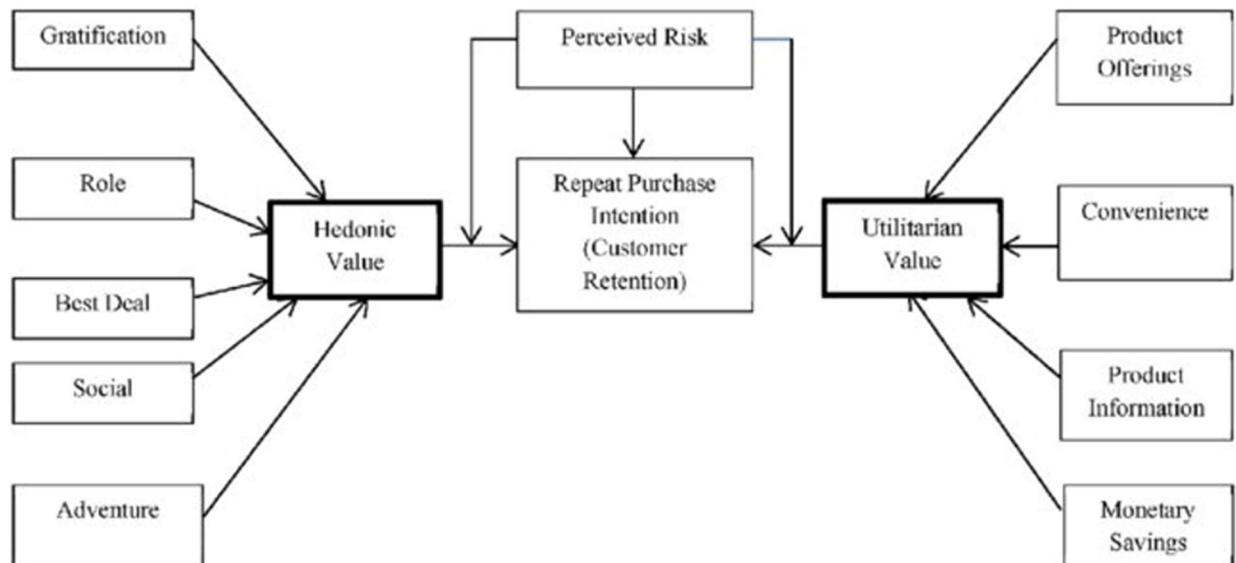

Fig 1.1 :- High Level Analysis

CHAPTER 2 – EDA Approach & Outcome

Exploratory Data Analysis (EDA) is an approach to analyze the data using visual techniques. It is used to discover trends, patterns, or to check assumptions with the help of statistical summary and graphical representations.

Steps involved

1) Importing a dataset
2) Understanding the big picture
3) Preparation
4) Understanding of variables
5) Study of the relationships between variables
6) Brainstorming

We will divide the data between Hedonic(HV) & utilitarian(UV) values to get the analysis with combinations

Hedonic Value(HV)---> is defined as value a customer receives based on the experience of products which are associated with fun, pleasure, and excitement, examples---> perfumes, flowers, luxury watches

- ❖ Psychological/Emotional experience**(HV)**
  - ✓ 1.Gratification
  - ✓ 2.Role
  - ✓ 3.Best Deal
  - ✓ 4.Social
  - ✓ 5.Adventure

Utilitarian Value(UV)--> is defined as value that a customer receives based on a task-related and rational consumption behavior,examples-->personal computers, detergents..

- ❖ Practical/Real experience**(UV)**
  - ✓ 1.Product Offerings
  - ✓ 2.Convenience
  - ✓ 3.product Information
  - ✓ 4.Monetary Savings

Out Come:

- ✓ Factors for Customer repeat purchase intention (loyalty)
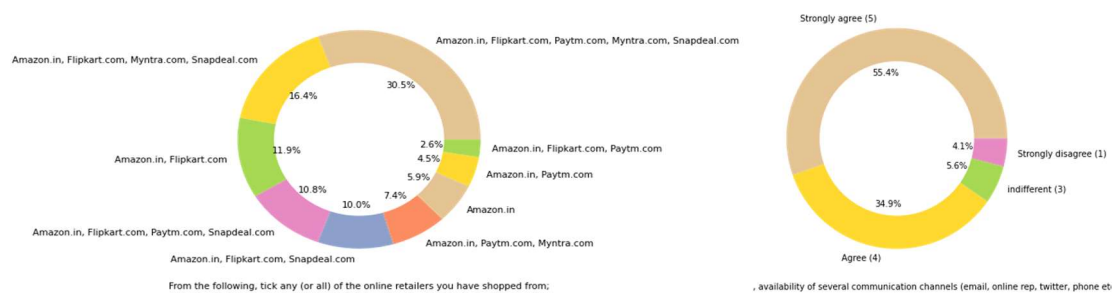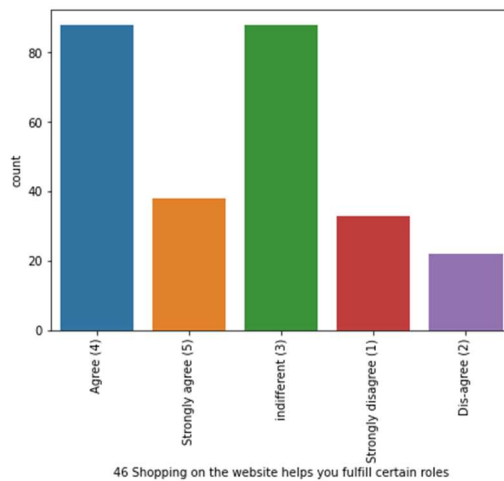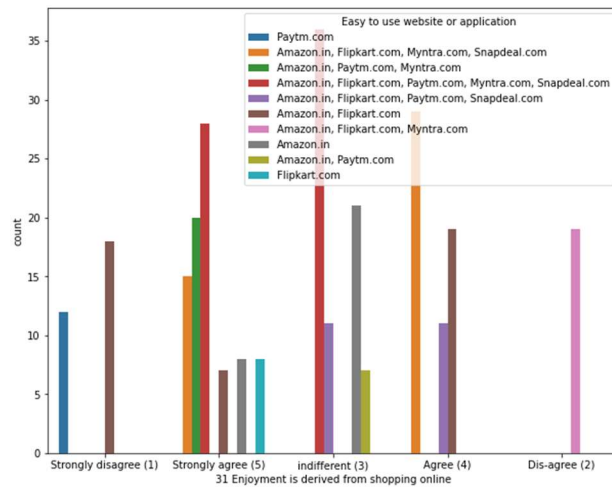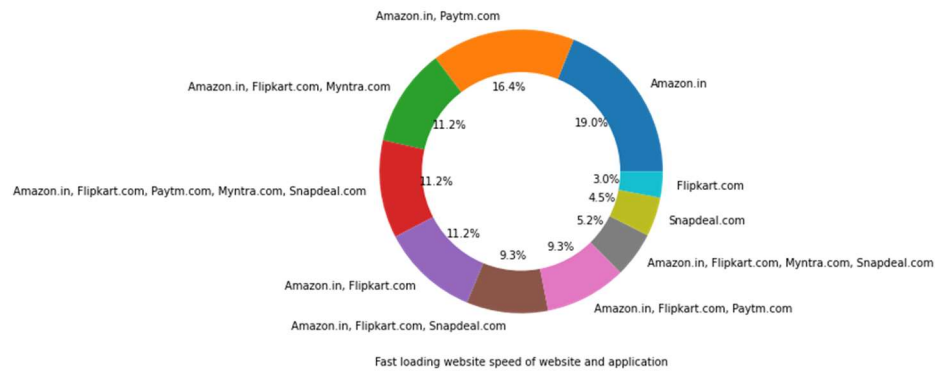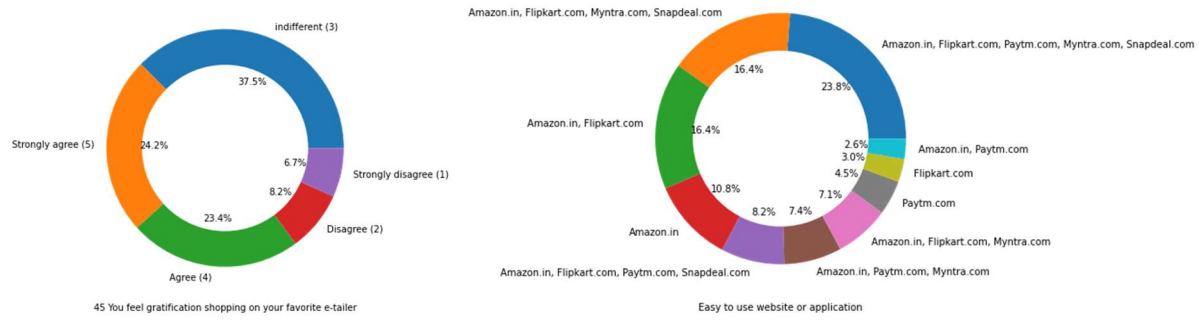- ✓ Factors for customer retention
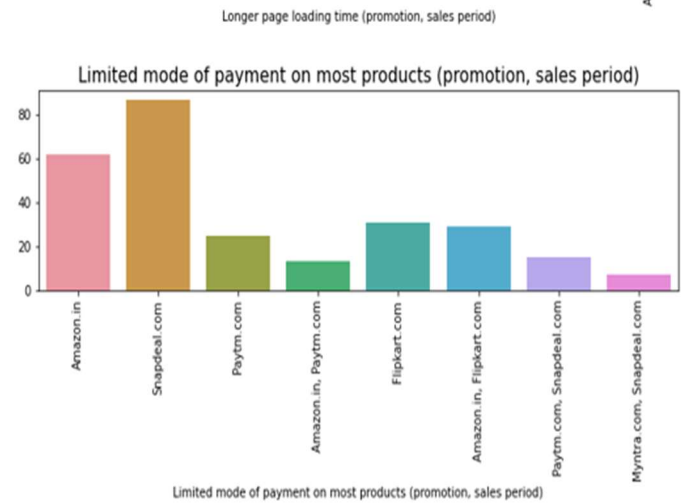
CHAPTER 3 – Hedonic Value(HV) Analysis

Variable are classified as Hedonic values(HV) based on Psychological/Emotional experience and divided for further analysis

- Feature_Gratifiction
- Feature_BestDeal
- Feature_social
- Feature_adventure

Visualization of the above are mentioned below from Univariate & Bi/Multi variate analysis→

6 How many times you have made an online purchase in the past 1 year?

5 Since How Long You are Shopping Online ?

31 Enjoyment is derived from shopping online

27 Empathy (readiness to assist with queries) towards the customers

22 Ease of navigation in website

24 User friendly Interface of the website

derive satisfaction while shopping on a good quality website or application

Net Benefit derived from shopping online can lead to users satisfaction

45 You feel gratification shopping on your favorite e-tailer

Easy to use website or application

Fast loading website speed of website and application

31 Enjoyment is derived from shopping online

46 Shopping on the website helps you fulfill certain roles

From the following, tick any (or all) of the online retailers you have shopped from;

, availability of several communication channels (email, online rep, twitter, phone et

Longer time to get logged in (promotion, sales period)

Longer page loading time (promotion, sales period)

Longer time in displaying graphics and photos (promotion, sales period)

Limited mode of payment on most products (promotion, sales period)

Late declaration of price (promotion, sales period)

14 How much time do you explore the e- retail store before making a purchase decision?

3.1 "Hedonic Value(HV)" Observations

> Pie plot also show that being able to guarantee the privacy of the customer is important silent feature for product selection. Another most important for product companies is availability of communication channels.

> 32% customer enjoyment from online shopping strongly matter

> 85 % of customer thinks, online shopping is convenient & flexible than physical shopping.

> 45.4% of customers strongly agree over fact that user satisfaction cannot exist without trust.

> 42.8 % customer strongly agree and 41.6 % customer agree over high loading & processing speed.

> Longer time to get logged in can annoyed customer. Amazon.in take longer time to logged in

> Most of the Customers spend more than 15 mins before making Purchase decision. Followed by 6-10 mins before making purchase decision.

> More people strongly agree to derive Enjoyment from shopping online based on the easiness of the application from Amazon,flipkart,paytm,SnapDeal,Myntra

> We can see different peoples have different opinions about connection between e-tailer & social status. Same with gratification on favorite e-tailer.

> Customers seem to be more loyal to amazon, flipkart and paytm as even though many of them have given negative remarks about them still they would recommend these platforms to their friend

CHAPTER 4 – Utilitarian Value(UV) Analysis

Variable are classified as Utilitarian values(UV) based on Practical/Real experience and divided for further analysis

- Feature_productOffer
- Feature_productInfo
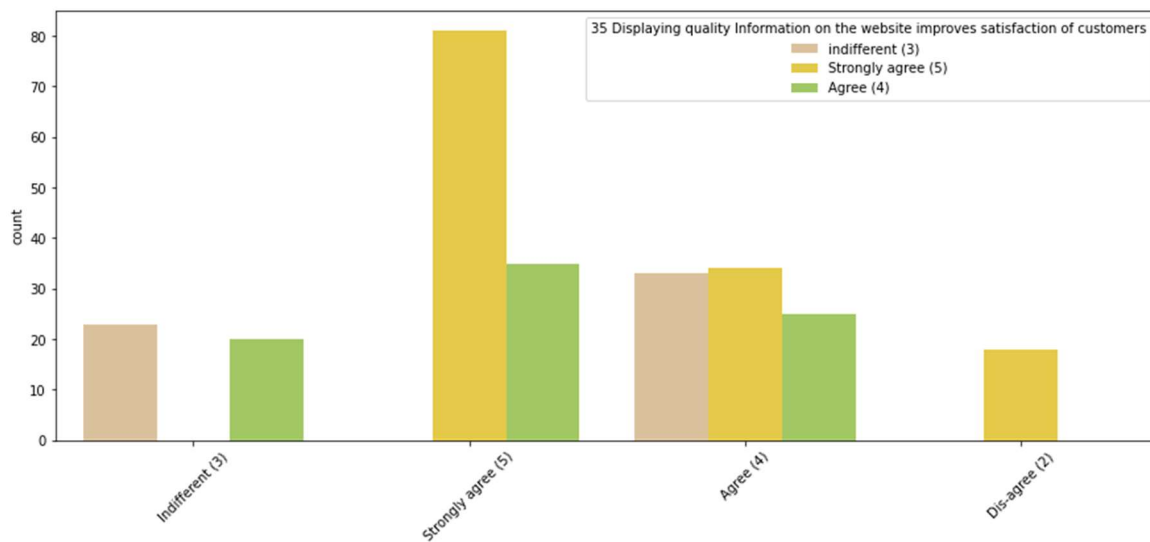- Feature_convenience
- Feature_Savings

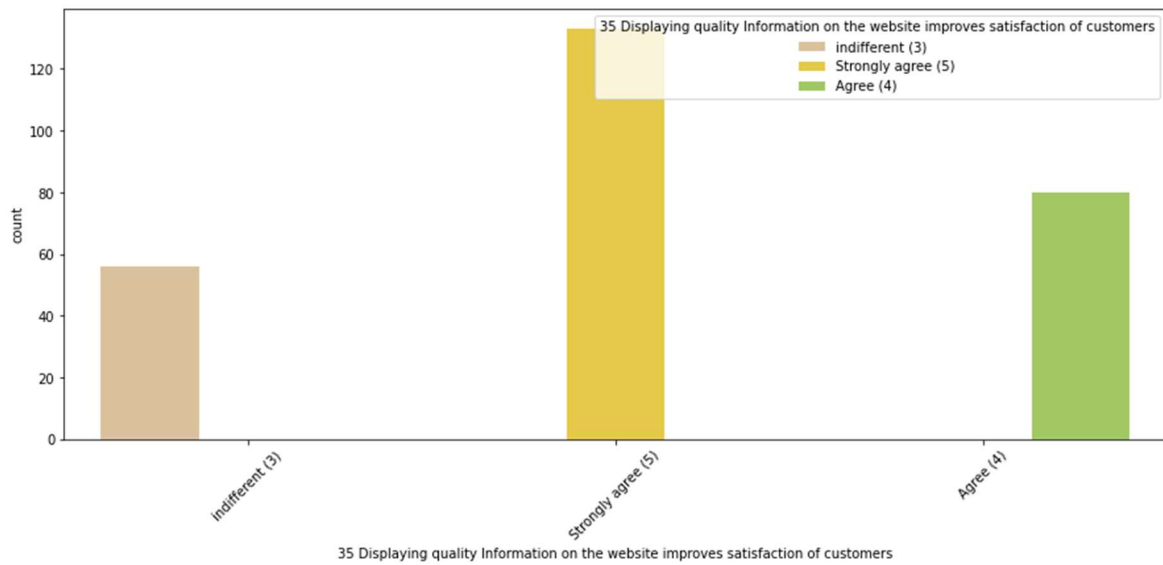Visualization of the above are mentioned below from Univariate & Bi/Multi variate analysis→
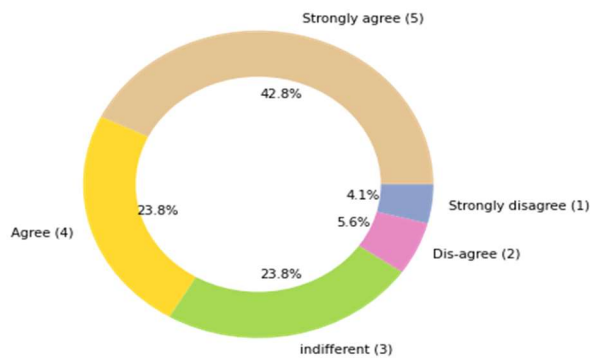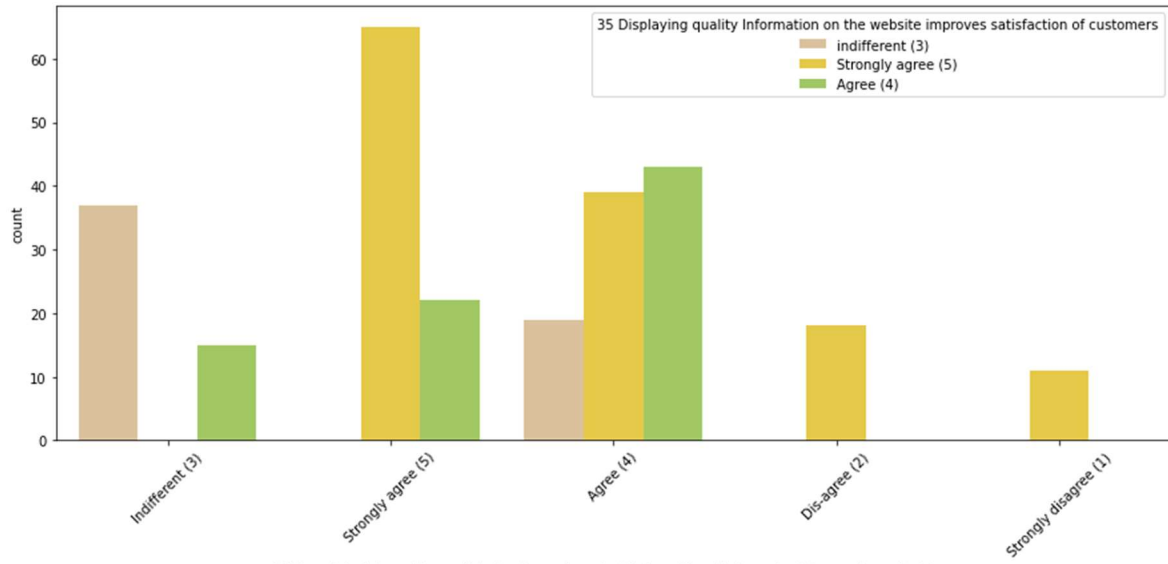

34 Gaining access to loyalty programs is a benefit of shopping online
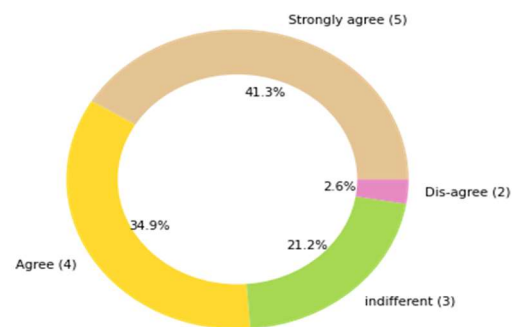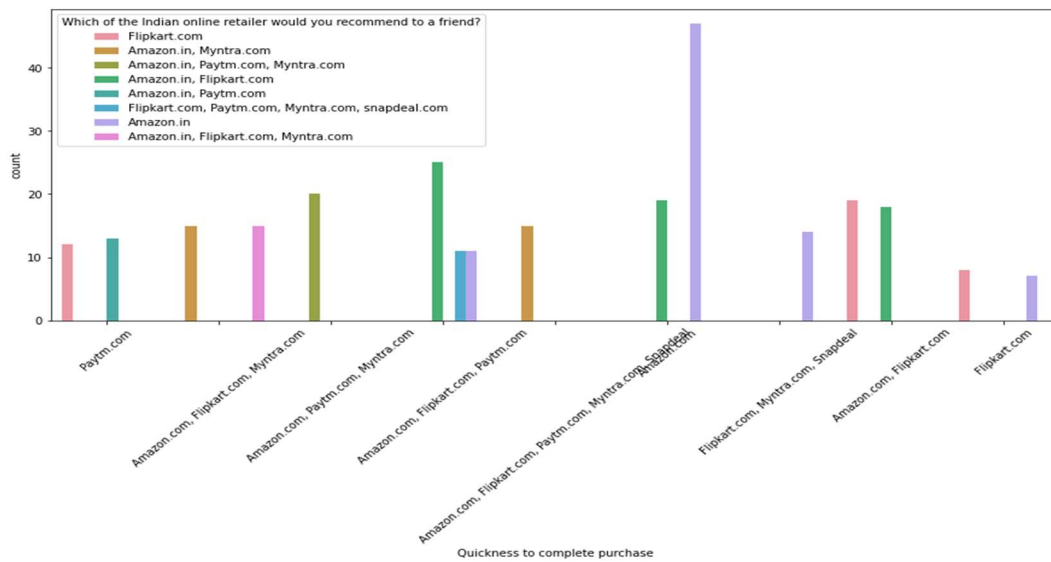

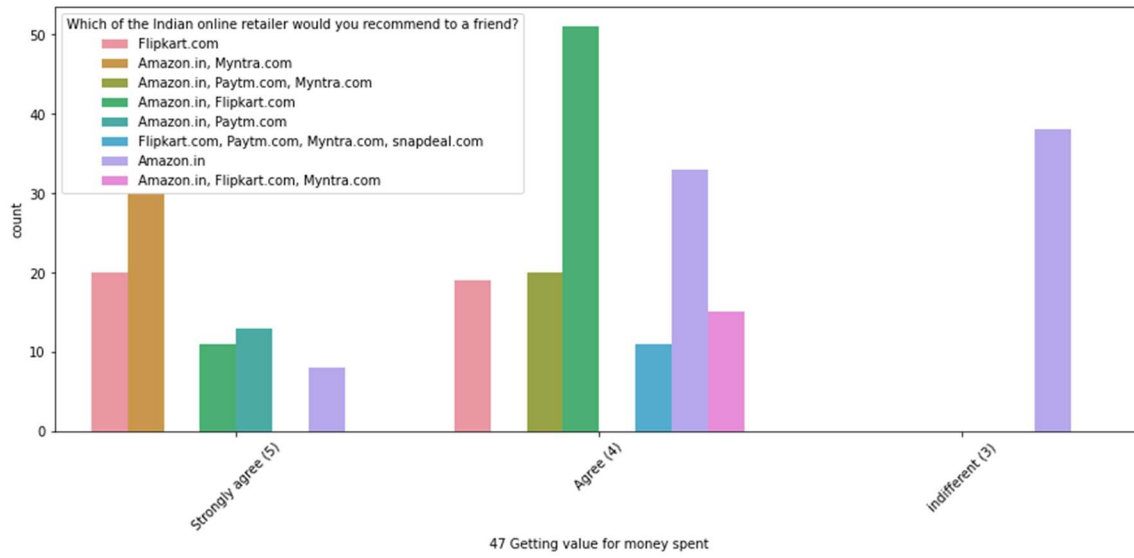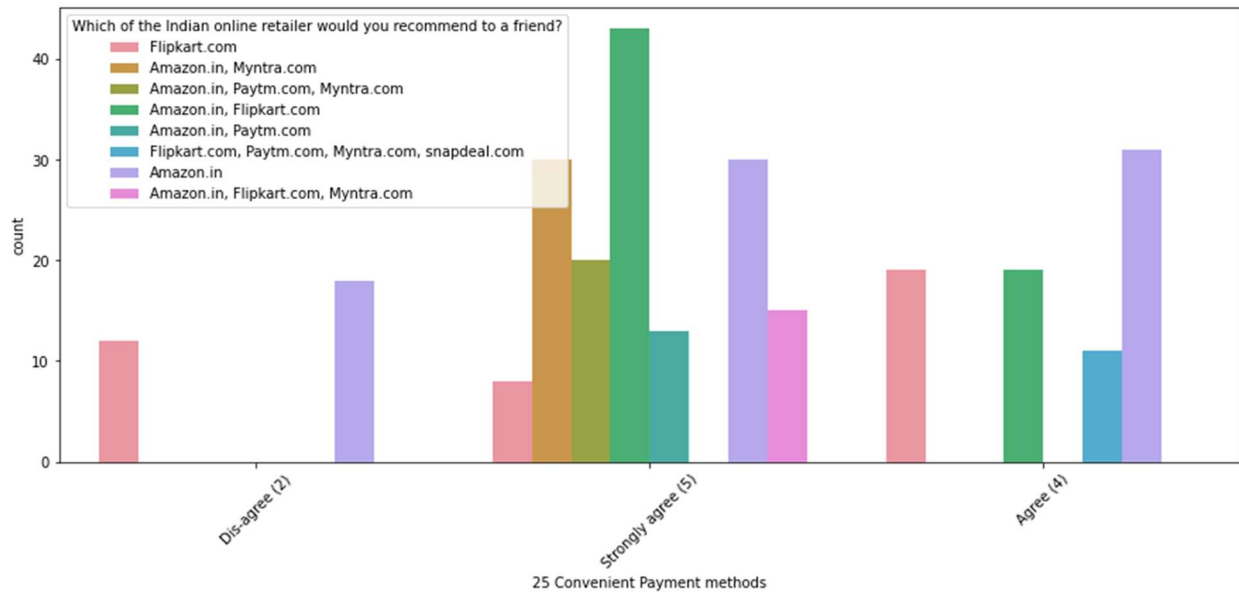39 Offering a wide variety of listed product in several category

19 Information on similar product to the one highlighted is important for product comparison

20 Complete information on listed seller and product being offered is important for purchase decision.

35 Displaying quality Information on the website improves satisfaction of customers

47 Getting value for money spent


Quickness to complete purchase

4.1 Utilitarian Value(UV)" Observations

➢ More customers look for Product offers while shopping online

➢ More than 60% of customer agree or strongly agree that complete information on listed seller and product being offered is important for purchase decision.

➢ Most of the Customers strongly agree thinks that companies readiness to assist customer queries related to product is important factor in purchase decision.

➢ Large portion of customers strongly agree that content on website must be easy to read and understand.

➢ Customer strongly agree that information on similar product to highlighted on website for product comparison.

➢ More than 90% of customer agree or strongly agree that all relevant information on listed products must be stated clearly.

➢ Majority of people also think that Amazon.in tops the chart in terms of quickness purchase process compare to others.

➢ Safe & Speed delivery very much deciding factor in terms of purchase.

➢ Most of the customer thinks that return & replacement policy of e-seller is important factor for making purchase decision.

➢ Around 83% customer pursue online shopping for Monetary Savings.

➢ 55.4% people strongly agree and 30.5% people agree that they prefer online shopping because they get value of money spent.

CHAPTER 5 – Perceived Risks Analysis

Visualization of the perceived risks are mentioned below from Univariate & Bi/Multi variate analysis→


Security of customer financial information

Privacy of customers' information
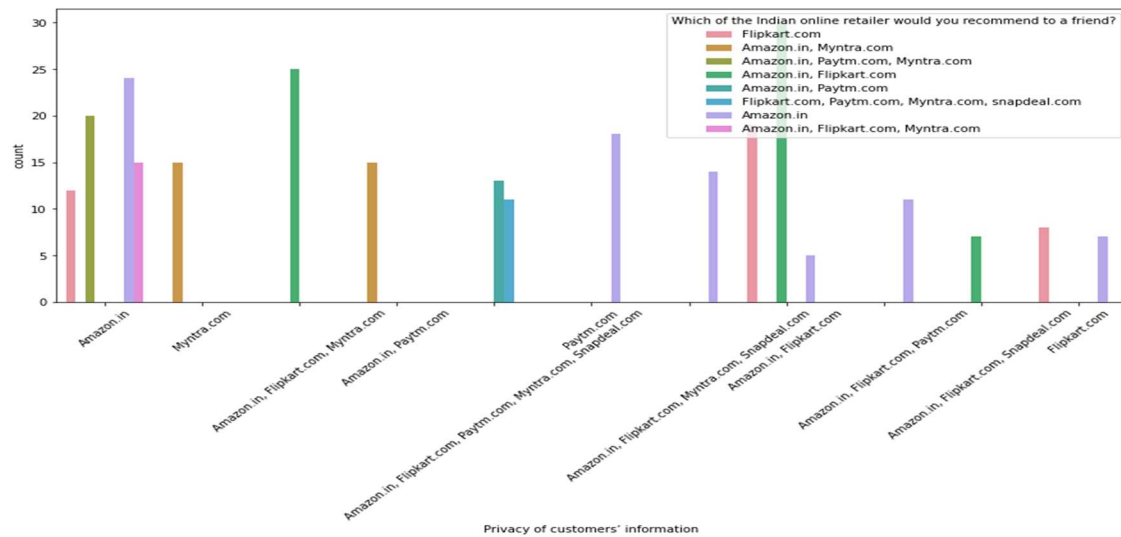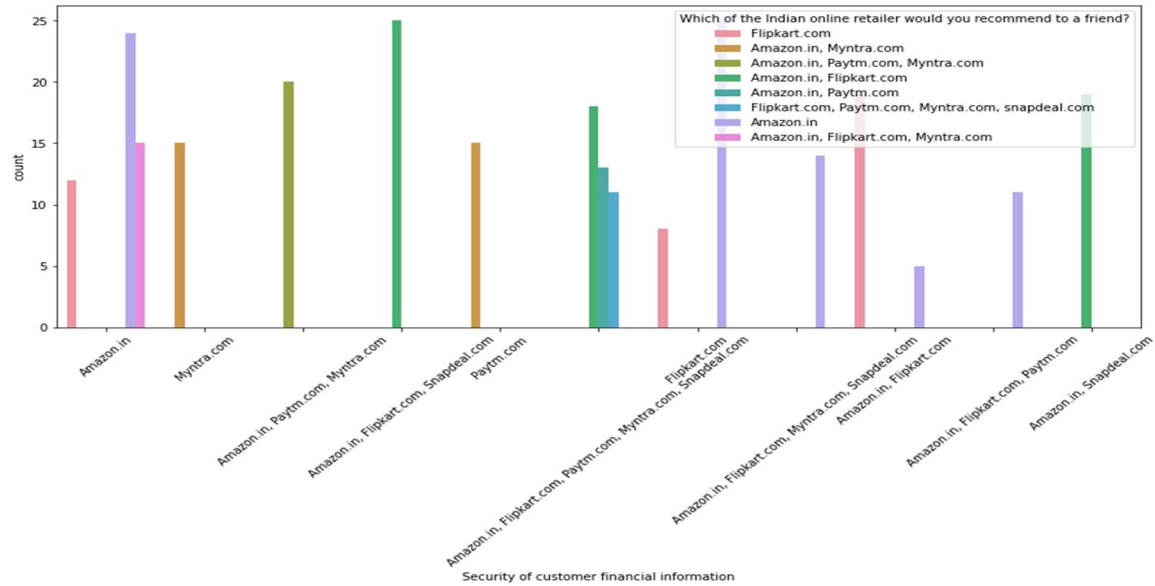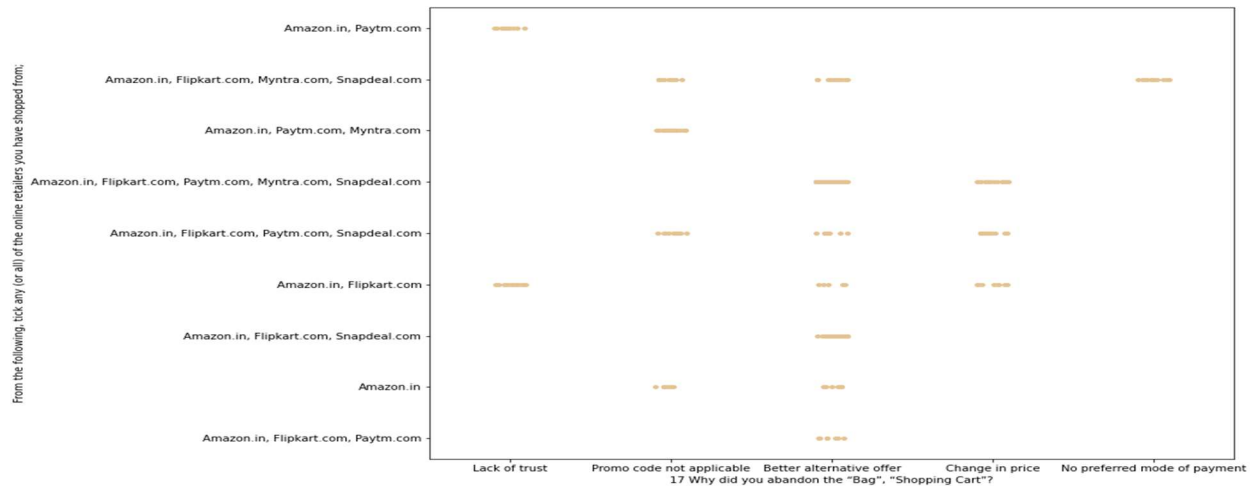
5.1 Perceived risks(UV)" Observations

➢ Customers seem to be more loyal to amazon, flipkart and paytm as even though many of them have given negative remarks about them still they would recommend these platforms to their friend

➢ Majority of customer trust Amazon.in followed by Flipkart.in over Security of their financial information.

➢ Another most common reason to abandon purchase decision is promo code not applicable on particular product. Followed by next most common reason is change in price.

➢ we can conclude that 84% customer abandon cart due to cost & discount tradeoff

CHAPTER 6 – Customer Analysis

Visualization of the customers are mentioned below from Univariate & Bi/Multi variate analysis➔
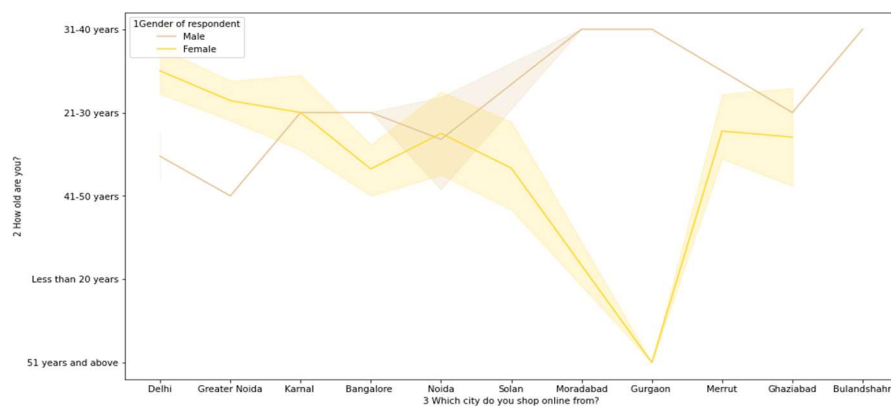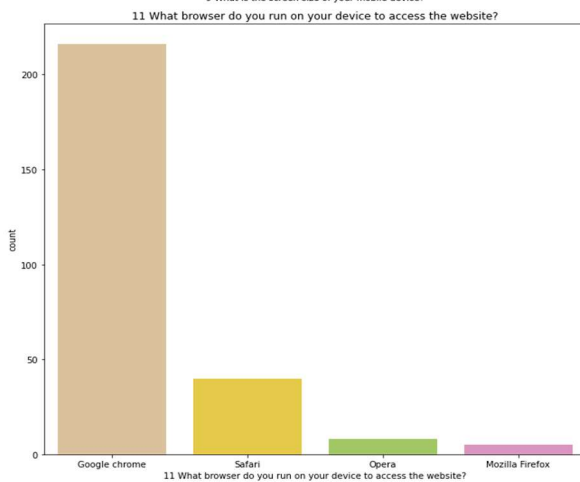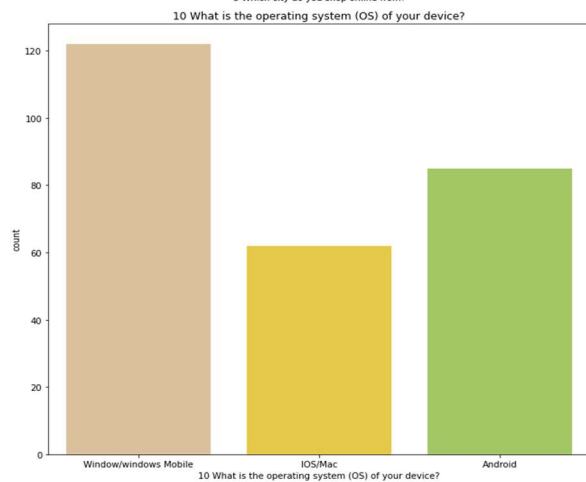
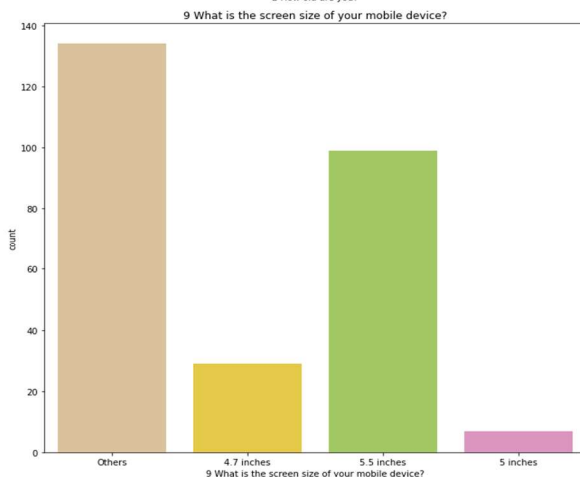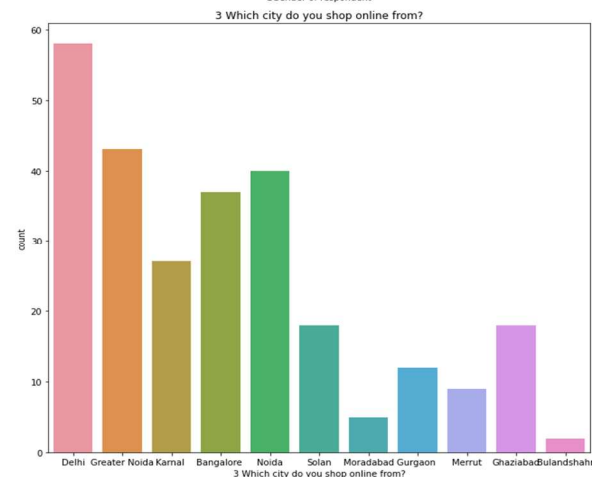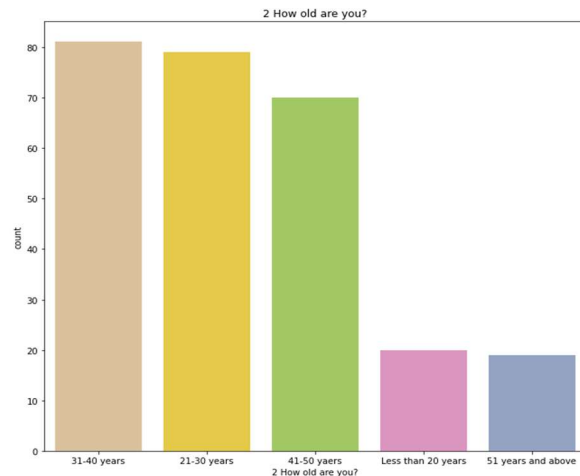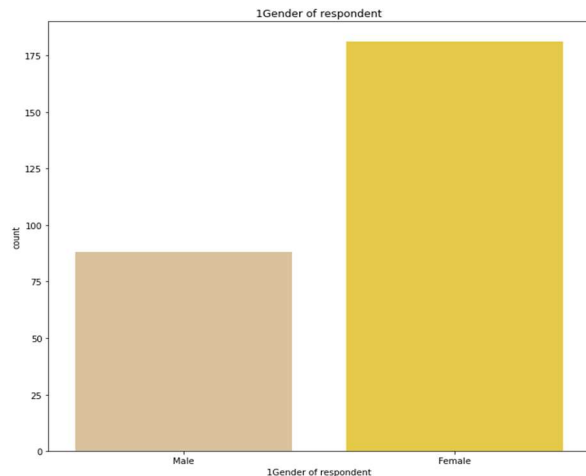**6.1 "Customer" Observations**

➢ Most Online Shopping Customer belong to Metro Cities. and most of them are Female customer

➢ We can conclude that in Metro city like Delhi, Male have more tendency of online shopping. So shopping platform can target this population in marketing.

➢ In Tier 2 & 3 Cities Majority of online shopping customers are Females.

➢ We can target these customer population in these cities.

➢ Highest number of people have been shopping online for above 4 years except for the age group below 20 years and above 50 years.

➢ People who are shopping online for 1-2 years does not include teenagers and elder people.

CHAPTER 7 – Other factor Analysis

Visualization of the other factors(service quality, system quality, information quality, trust and net benefit) are mentioned below from Univariate & Bi/Multi variate analysis➔

21 All relevant information on listed products must be stated clearly

aying quality Information on the website improves satisfaction of custor

22 Ease of navigation in website

18 The content on the website must be easy to read and understand

Perceived Trustworthiness

Net Benefit derived from shopping online can lead to users satisfaction
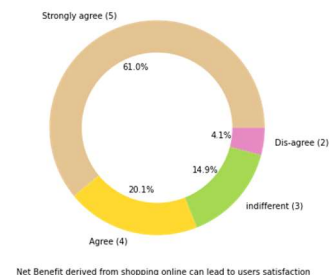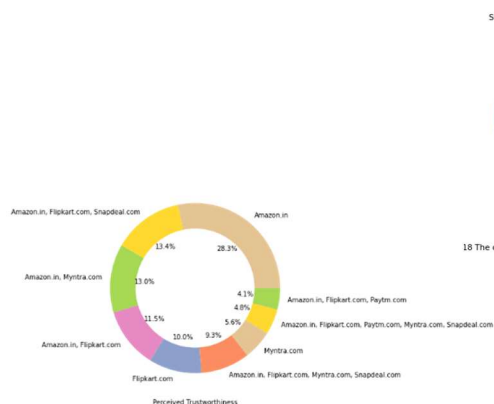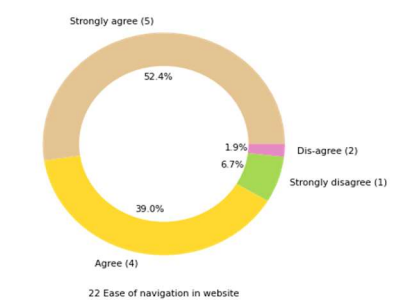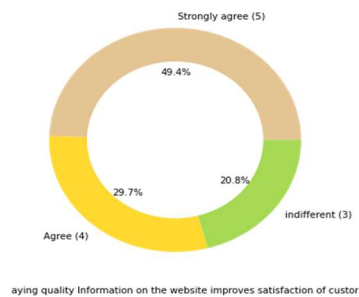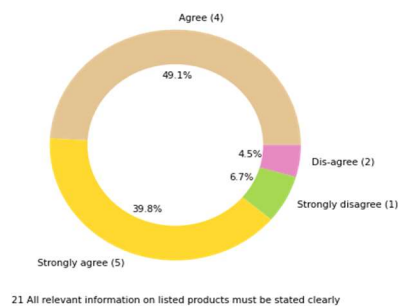
7.1 "Other Factors" Observations

- Most of people want shorter delivery time frame, majority customer agree that Paytm.com takes longest time for delivery compare to others.
- 50.2% Customer strongly agree & 36.4% customer agree that for good online shopping complete and relevant product information.
- 49.4% customer strongly agree and 29.7% customer agree that displaying quality information on website helps in decision making in turn improves customer satisfaction.
- 70.3 % customer strongly agree and 16.7 % customer agree that website should be user friendly.
- 52.4% of customers strongly agree that website should be easy for navigation

CHAPTER 8 – Correlation with HV & Observations

Dython will automatically find which features are categorical and which are numerical, compute a relevant measure of association between each and every feature, and plot it all as an easy-to-read heat-map we can find the correlation matrix of categorical variables.

The correlation matrix really helps us in identifying the features which are suitable for our model training.

- An effect scores closer to 0 translates to there being no relationship.

- A score closer to 1 or -1 is a positive or negative relationship

- A perfect score of 1 is a direct correlation
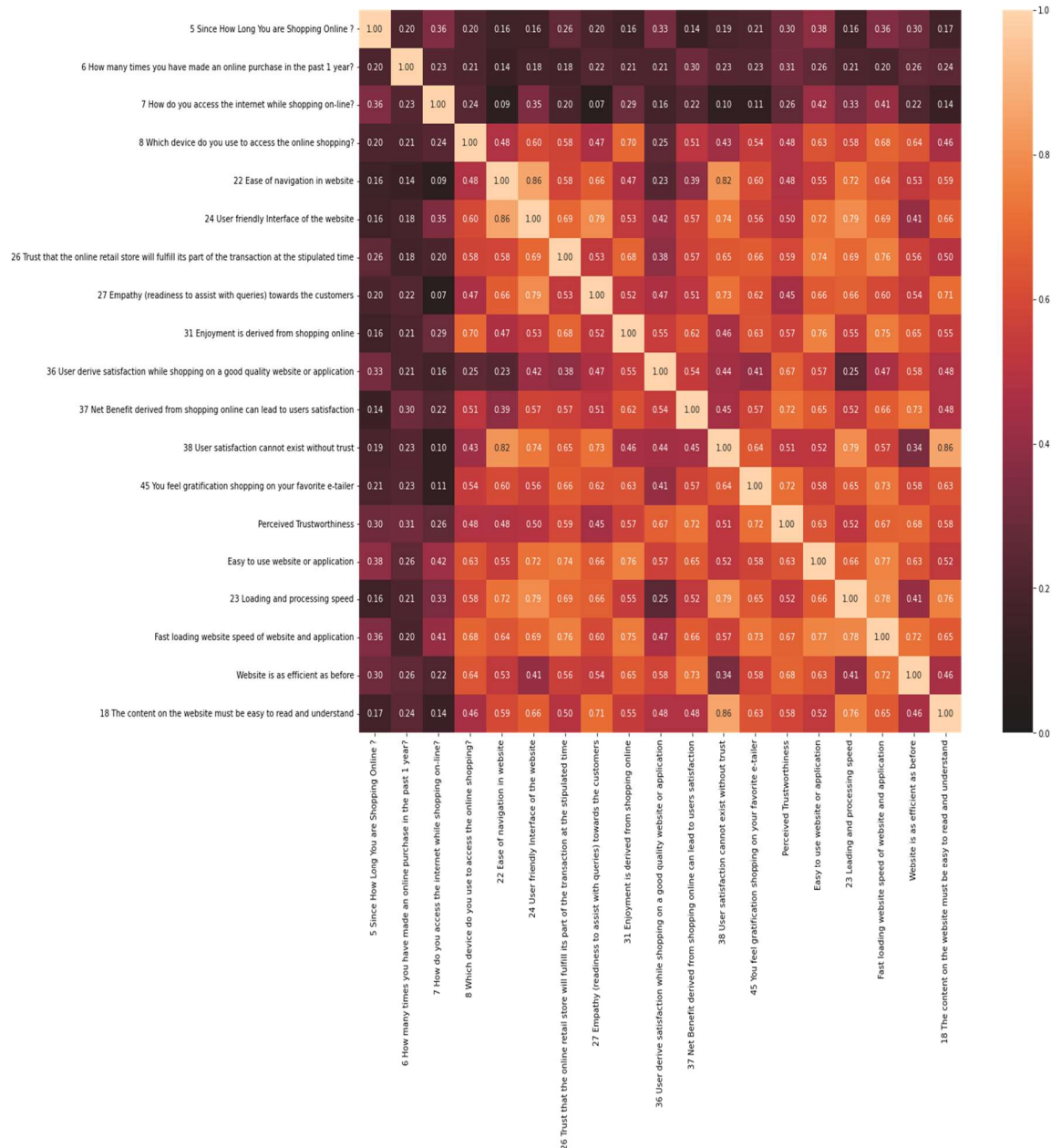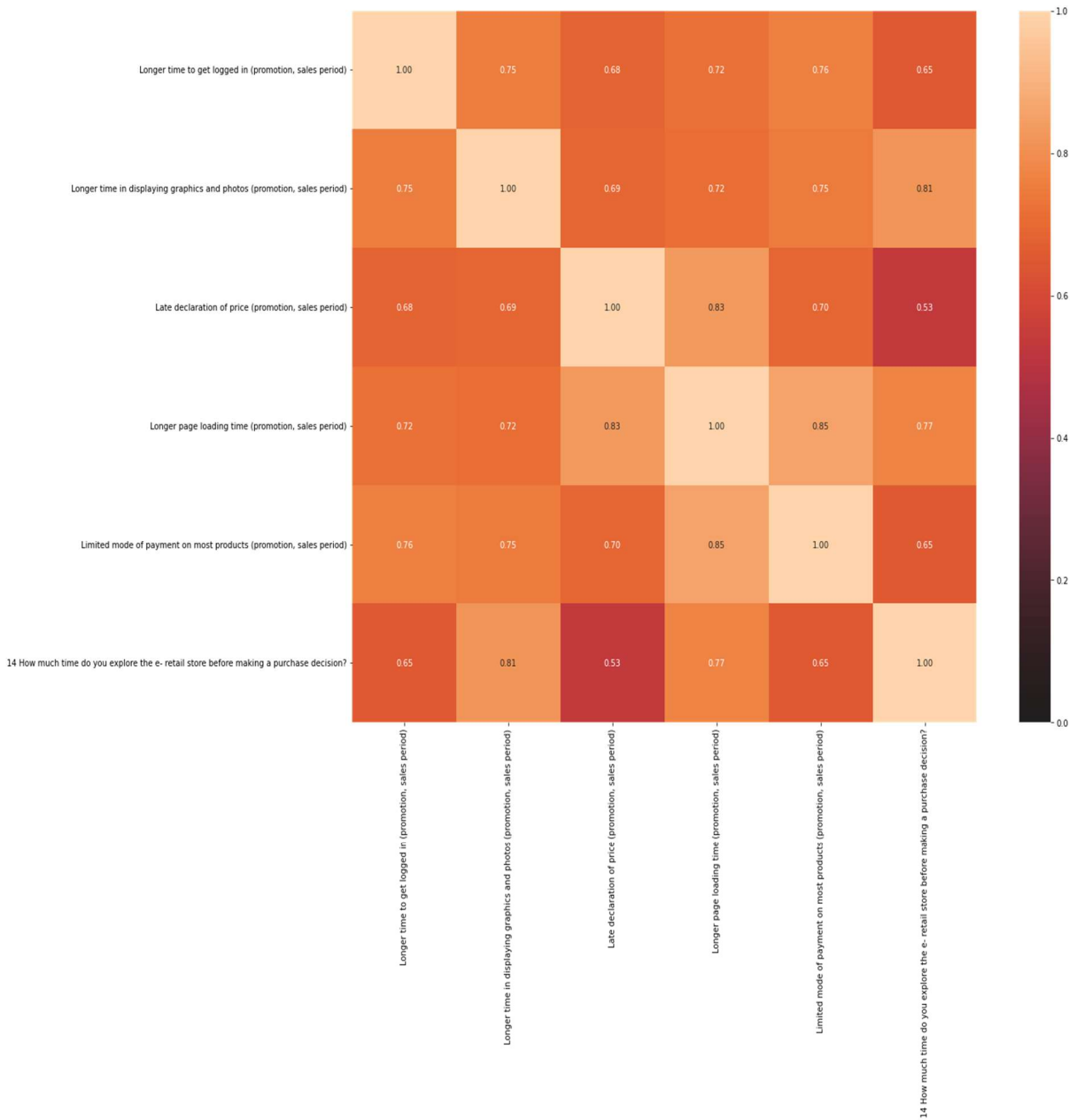
Below are highly correlated based on the scores ~1

> 18 The content on the website must be easy to read and understand-->(0.86)-->38 User satisfaction cannot exist without trust

> 22 Ease of navigation in website--(0.82) --> 38 User satisfaction cannot exist without trust

> 23 Loading and processing speed->(0.79)-->24 User friendly Interface of the website

> 27 Empathy (readiness to assist with queries) towards the customers---(0.79) -->24 User friendly Interface of the website

> Limited mode of payment on most products (promotion, sales period)--->(0.85)-->Longer page loading time (promotion, sales period)

> Presence of online assistance through multi-channel-->(0.83)-->Which of the Indian online retailer would you recommend to a friend?

> 14 How much time do you explore the e- retail store before making a purchase decision?--->(0.77)-->Longer page loading time (promotion, sales period)

CHAPTER 9 – Correlation with UV & Observations

Correlation with Utilitarian Values→
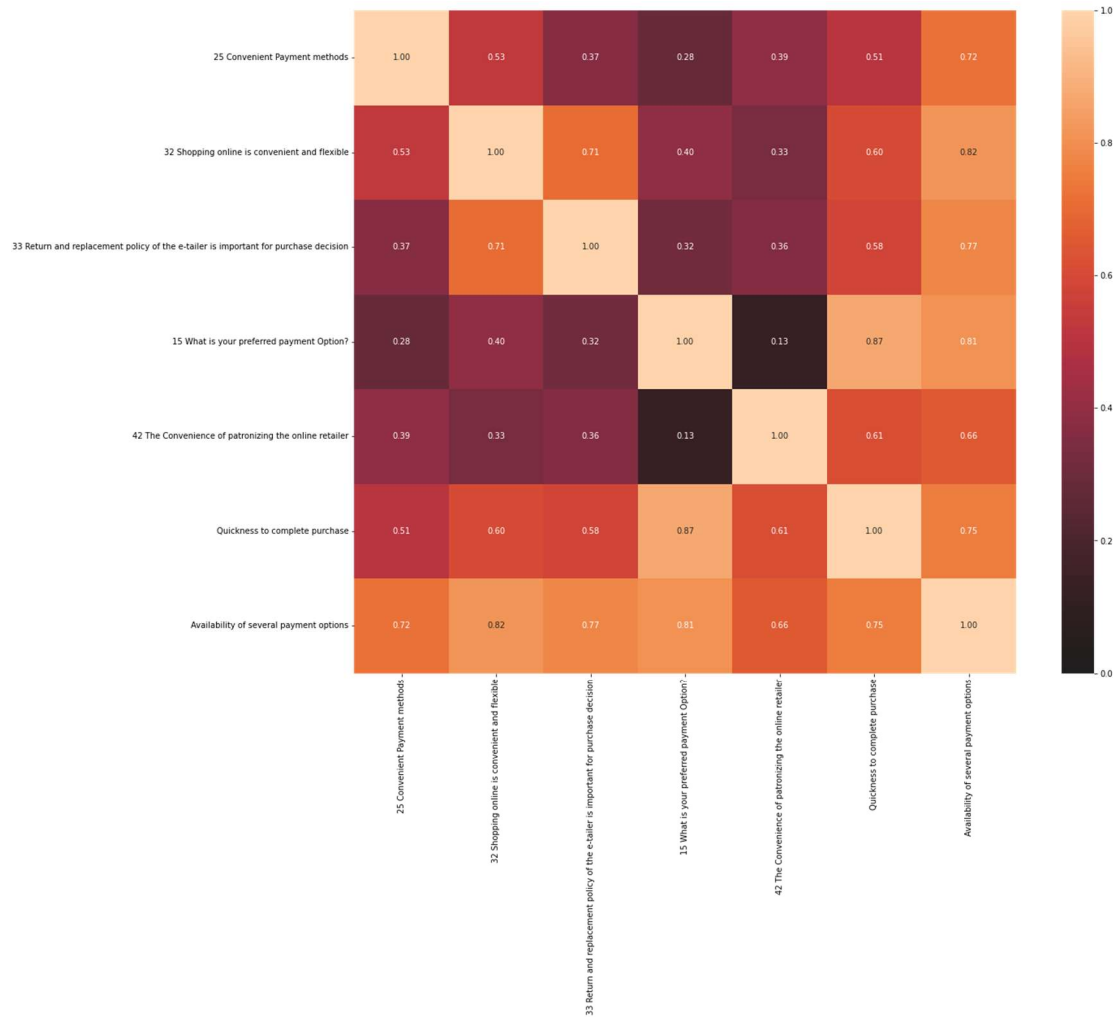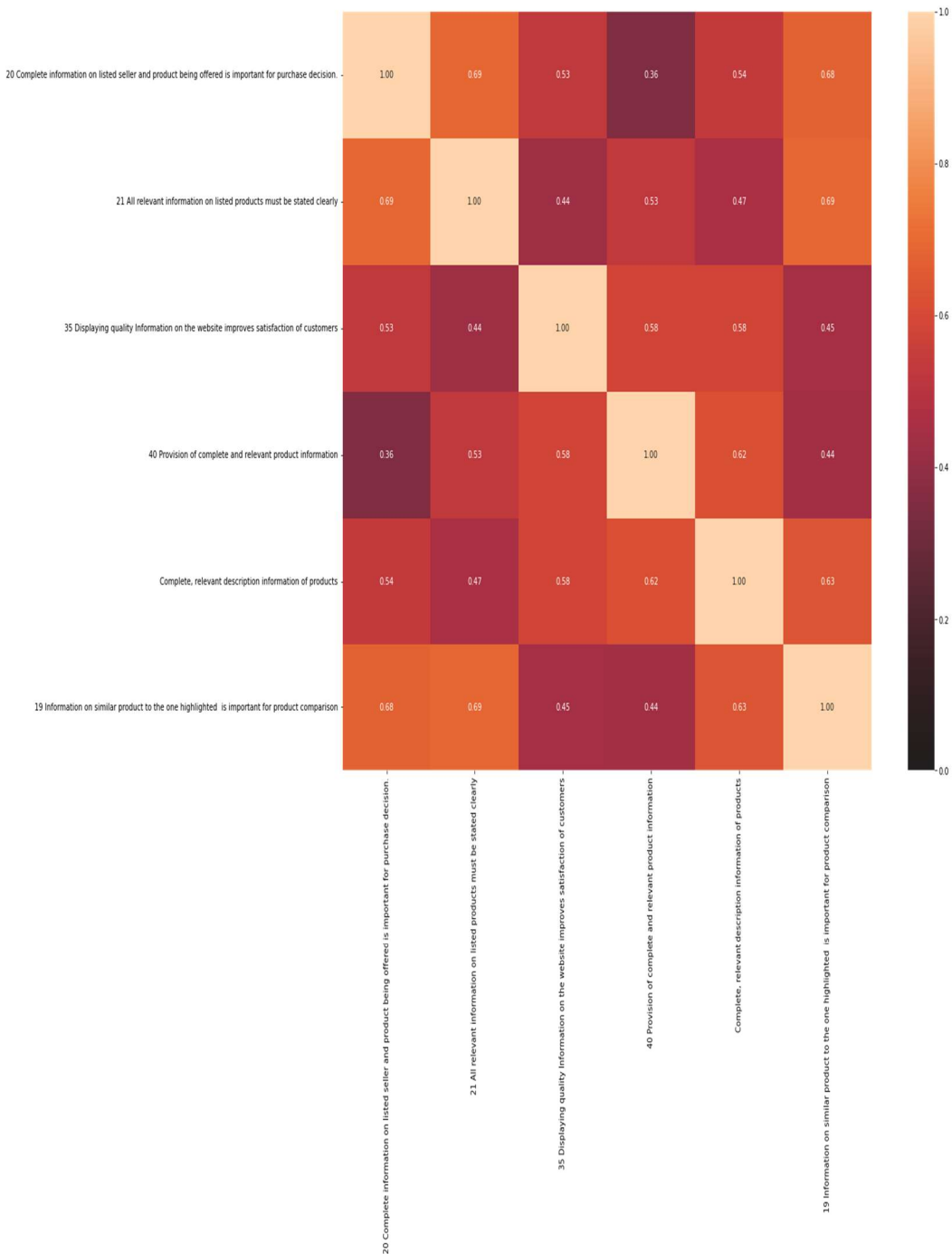
Below are **HV & UV** highly correlated based on the scores ~1

- ➢ Quickness to complete purchase -->(0.87)-->15 What is your preferred payment Option?

- ➢ 32 Shopping online is convenient and flexible-->(0.82)-->Availability of several payment options

- ➢ Availability of several payment options -->(0.81)-->15 What is your preferred payment Option?.

CHAPTER 10 – Correlation with HV ++UV & Observations

HV & UV Correlation "Observations are done based on chi-square tests

- ➢ Feature_Savings (UV) && Feature_social(HV) are highly correlated based on p-Value of chi-Square test
- ➢ Feature_productOffer(UV) && Feature_Gratifiction(HV) are highly correlated based on p-Value of chi-Square test

Feature_productOffer = ['34 Gaining access to loyalty programs is a benefit of shopping online',
                                   'Wild variety of product on offer',
                                   '39 Offering a wide variety of listed product in several category']

Feature_Gratifiction= ['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?',
              '22 Ease of navigation in website',
              '24 User friendly Interface of the website',
              '26 Trust that the online retail store will fulfill its part of the transaction at the stipulated time',
              '27 Empathy (readiness to assist with queries) towards the customers',
              '31 Enjoyment is derived from shopping online',
              '36 User derive satisfaction while shopping on a good quality website or application',
              '37 Net Benefit derived from shopping online can lead to users' satisfaction',
              '38 User satisfaction cannot exist without trust',
              '45 You feel gratification shopping on your favorite e-tailer',
              'Perceived Trustworthiness',
              'Easy to use website or application',
              '23 Loading and processing speed',
              'Fast loading website speed of website and application',
              'Website is as efficient as before',
              '18 The content on the website must be easy to read and understand']

Feature_Savings = ['30 Online shopping gives monetary benefit and discounts',
                              '41 Monetary savings',
                              '47 Getting value for money spent']

Feature_social = ['12 Which channel did you follow to arrive at your favorite online store for the first time?',
                '29 Responsiveness, availability of several communication channels (email, online rep, twitter, phone etc.)',
                '44 Shopping on your preferred e-tailer enhances your social status',
                'Presence of online assistance through multi-channel',
                'From the following, tick any (or all) of the online retailers you have shopped from;',
                'Which of the Indian online retailer would you recommend to a friend?']

ANALYSIS

HV & UV Intention of Repeat purchase

- ➢ Heavy shoppers who shop more than 41 times a year shop from all the online brands,

- ➢ some of the people who shop for 32-40 and less than 10 times a year seem to exclude Myntra.

- ➢ People shop from Amazon and flipkart

UV Based Intention of purchase

- ➢ People shopping from amazon and paytm are getting benefits from the loyalty points

- ➢ flipkart and Snapdeal also seem to give such benefits but people who shop from almost everywhere disagree with this statement

HV Based Intention of purchase

- ➢ Almost all the people who have shopped from amazon, flipkart and paytm are satisfied

- ➢ People who shop from a greater number of online brands doesn't seem to be satisfied.

SUMMARY (How we got here!!)

EDA

- ❖ Data divided with PV, UV & risks
- ❖ Features with PV performed Visualizations
- ❖ Featured with UV performed with Visualizations
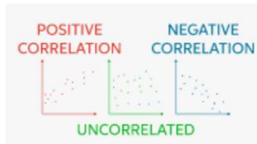
Correlation

- ❖ Used Dython library for correlation of Categorical Features
- ❖ Chi-Square test (Nominal Values)with P-Value

Analysis

- ❖ Count plot, Pie Chart ,Line plots provided insights
- ❖ Combination of PV & UV features are derived

CONCLUSION

Customer retention (UV+HV)

➢ People shopping from amazon and paytm are getting benefits from the loyalty points
➢ People who shop from a greater number of online brands doesn't seem to be satisfied.
➢ Loyal customers prefer buying and tend to spend more money on shopping in your store. Statistics show that engaged consumers purchase more frequently
➢ People shop from Amazon and flipkart whatever be the case.

Customer Retention (reduce risks)

➢ Online brands should update all their platforms rather than just application

DIRECTIONS OF FUTURE WORK

❖ Dataset with 269 rows is small sample size but can have more correct insight with large dataset

❖ Appropriate Machine Learning model can be built predict orientation of customer.

❖ More features around purchase details, product price can be available to help in best recommendation.