A Presentation on Customer Retention

Submitted To: Khusboo Garg (SME of Internship Batch no-28)

Submitted By: Abhishek Behera
Internship Batch no-28

Content

- Introduction
- Problem Statement
- Objective
- Exploratory Data Analysis (EDA)
- Visualization
- Inference
- Conclusion & Future Work

Introduction

- *What is Customer Retention and is it really need for growing the business ?*

- "Customer retention " means a company's ability to turn it's primary customers into repeat buyers and prevent them from switching to their other competitors. In other words customer retention means – **"To maintain the existing customers from Switching"**

- This happens only if there exists a positive relation between the company and the customer.

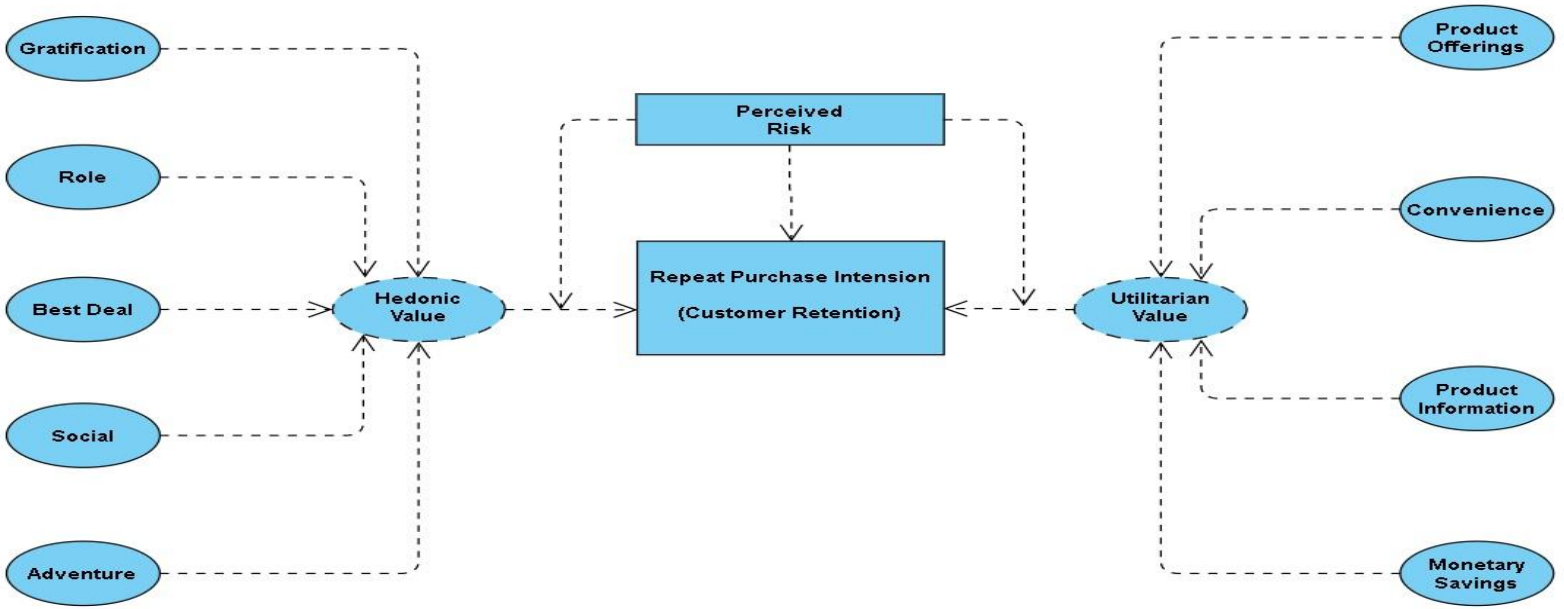
Benefits of Customer Retention

- Retained customers have a tendency to buy other services from same company.
- The probability of selling to an existing customer is 60-70% .
- The probability of selling to new prospect is 5-20% .
- Declined migration rates .
- It’s more expensive to acquire a new customer than to retain an old one.

Problem Statement

- 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 or 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 further more 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.

The problem statement can be represented in the form of below use case diagram as well.



Objective

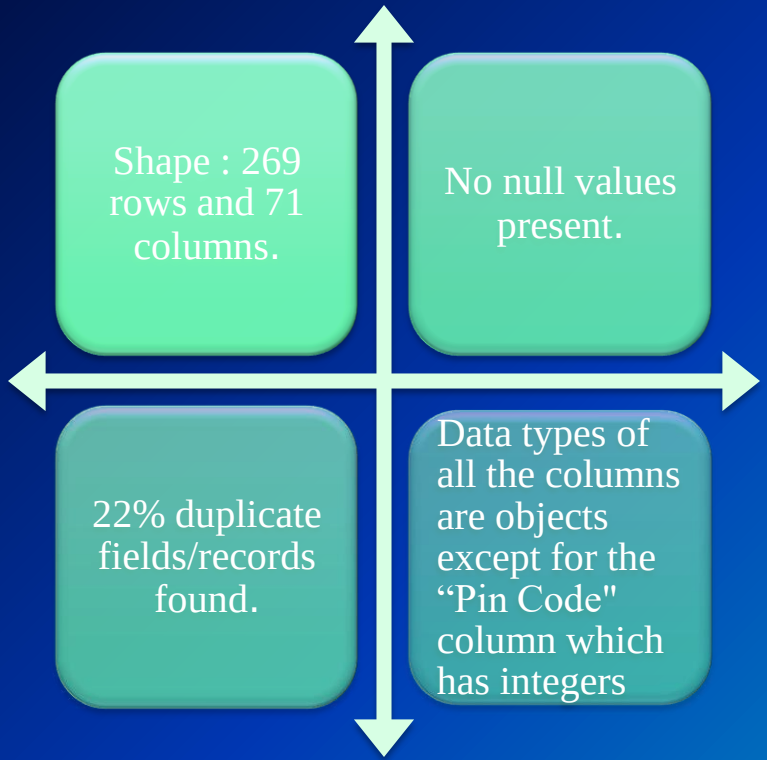
- The objective is to apply the analytical skills to provide findings and conclusion that would help to predict customer retention for a E-Retail company using their data on users provided over period of time.

- Using the model I was tasked with determining which features were most influential in loss of valuable customer and then making a plan for how the company could use this information to increase customer retention.

Exploratory Data Analysis (EDA)

Requirements	Tools Used
Hardware	RAM: 8 GB CPU : Intel(R) Core(TM) i5-5200U CPU @ 2.20GHz GPU : Intel(R) HD Graphics 5500and NVIDIA GeForce 940M
Software	Programming language : Python Distribution : Anaconda Navigator Browser based language shell : Jupyter Notebook
Libraries/Packages	Pandas NumPy Matplotlib seaborn scikit-learn

Process involved in EDA



- First I have imported all the necessary libraries and loaded the entire dataset in our Jupyter Notebook and renamed the columns.
- Then I checked the shape of our dataset and found that we have a total of 269 rows and 71 different columns.
- We don’t have any null values or missing values present in our dataset.
- There is 22% percent of duplicate records in our dataset however I have chosen to retain those information instead of removing them.
- By checking the data types I came to know that all the columns have ‘object’ data type except the column representing the Pin Code which has ‘integer’ data type.

Description of the dataset

- This data is collected from the Indian online shoppers. Our Dataset consists of reviews and feedbacks of customers on 5 top Indian Online Retailers : Amazon, Flipkart, Snapdeal, Myntra and Paytm.
- Questionnaire is formed on the basis of brand strength, brand empathy or commitment, overall customer satisfaction and perceived value for money with intention to recommend.
- Results indicate the e-retail success factors which are very much critical for customer satisfaction and retention

The top 5 Indian Online Retailers are:

- Amazon
- Flipkart,
- Snapdeal,
- Myntra
- Paytm

Data Visualization

Q. What is Data Visualization?

Ans. Data visualization is defined as a graphical representation that contains the information and the data and good data visualization is technique of visual art that grabs our interest and keeps our main focus on the message captured with the help of eyes.

- Different Types of Analysis for Data Visualization are

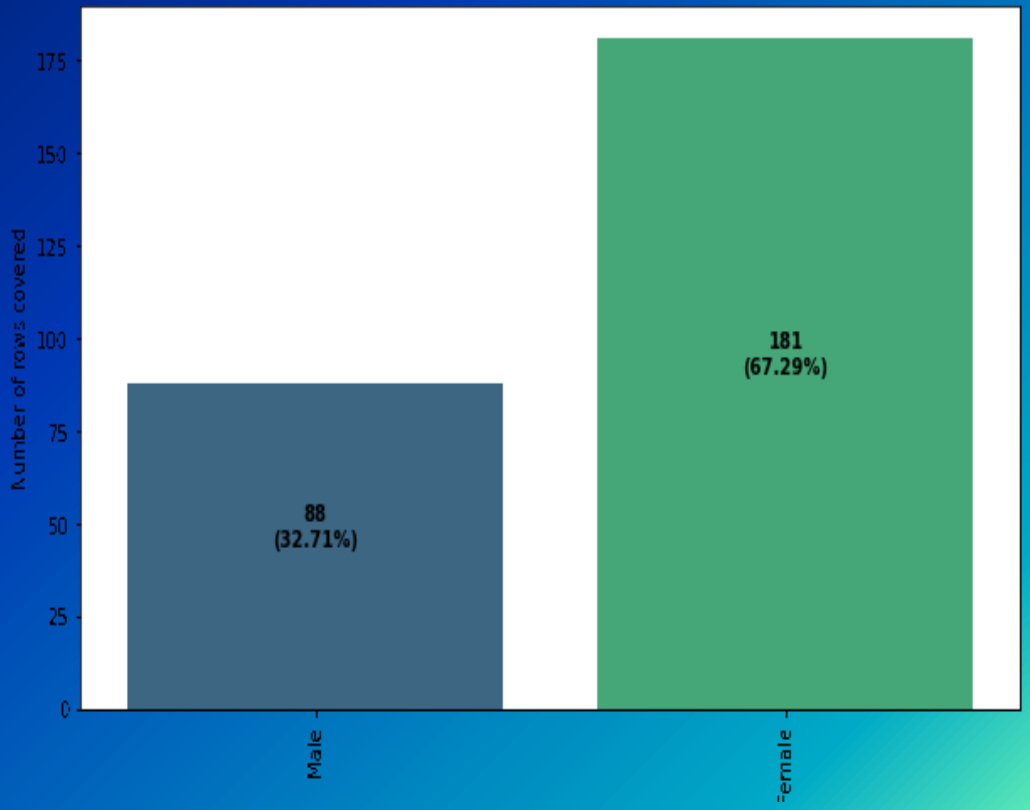
Univariate Analysis: In univariate analysis, we will be using a single feature to analyze almost all of its properties.

Bivariate Analysis: When we compare the data between exactly 2 features then it is known as bivariate analysis.

Multivariate Analysis: In the multivariate analysis, we will be comparing more than 2 variables.

Univariate Analysis

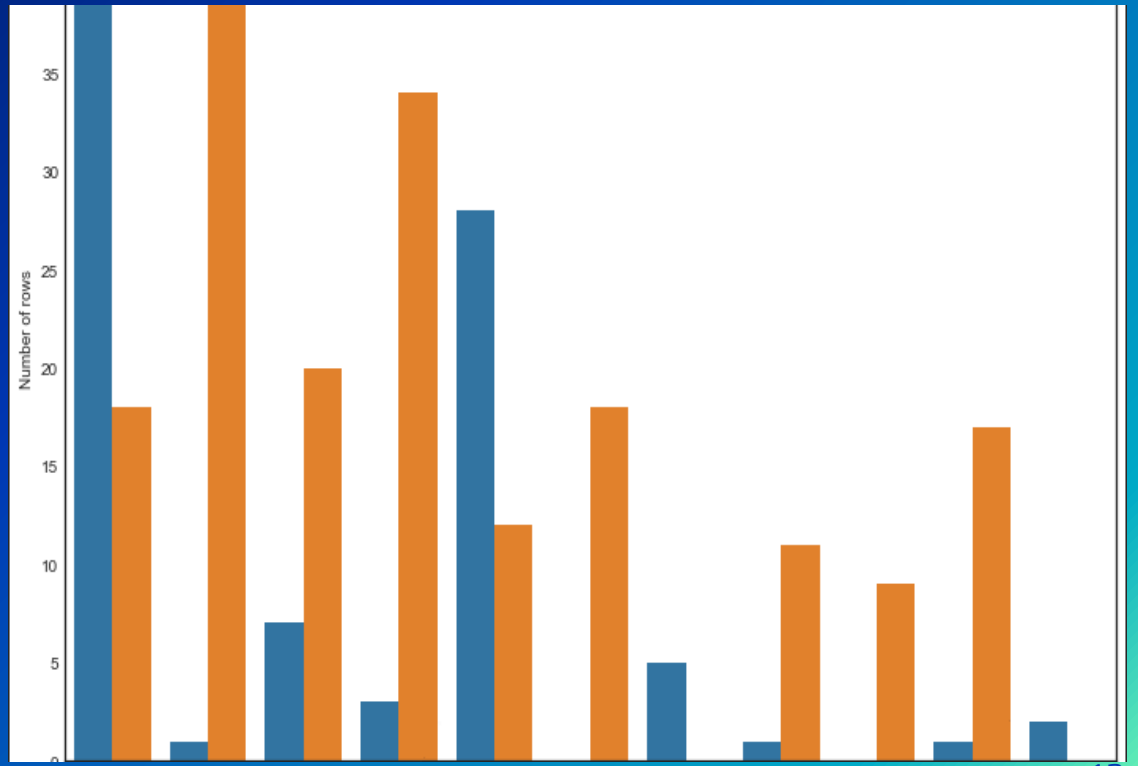
- With the help of count plots I am able to get the total number of rows covered by each unique categorical value present in all the columns of our dataset.
- I ensured that along with the total row number, the percentage of data coverage is also visible.



Bivariate Analysis

- Using the count plot along with the hue parameter I was able to generate analysis details comparing 2 columns of the dataset where the hue remained constant while the other one kept changing inside a loop.

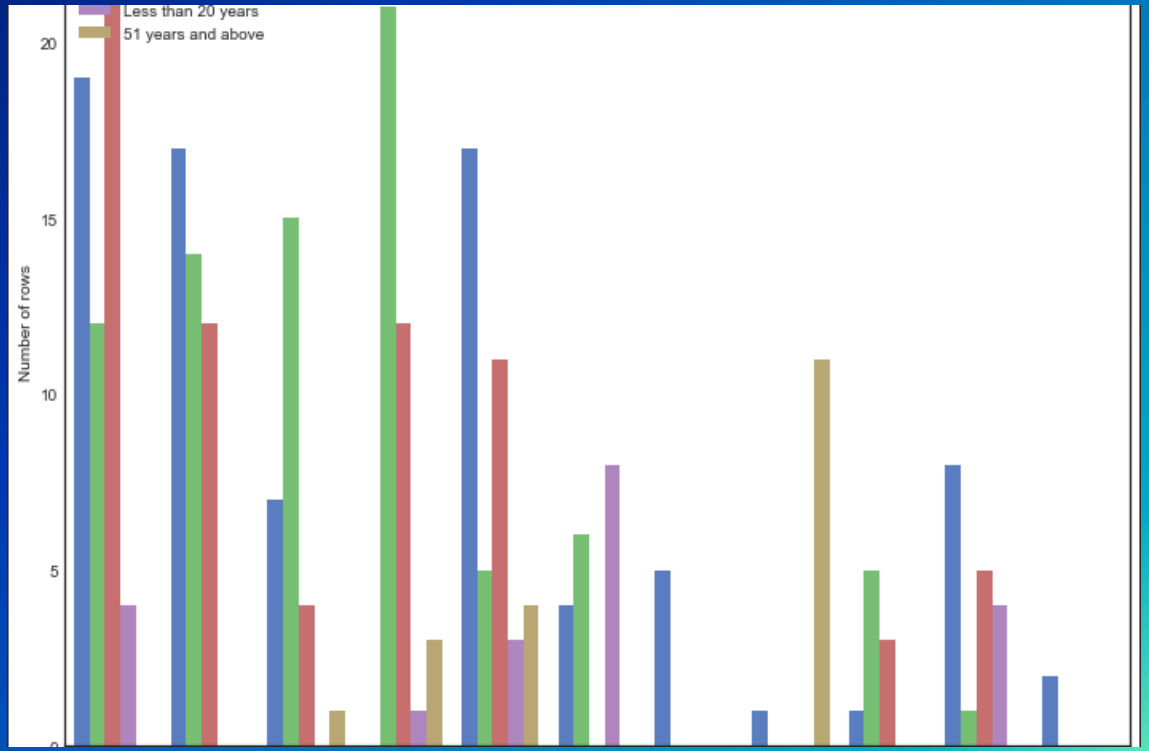
- The hue used in this plot is the “Gender” column showing the legend accordingly



Bivariate Analysis

- Using the count plot along with the hue parameter I was able to generate analysis details comparing 2 columns of the dataset where the hue remained constant while the other one kept changing inside a loop.

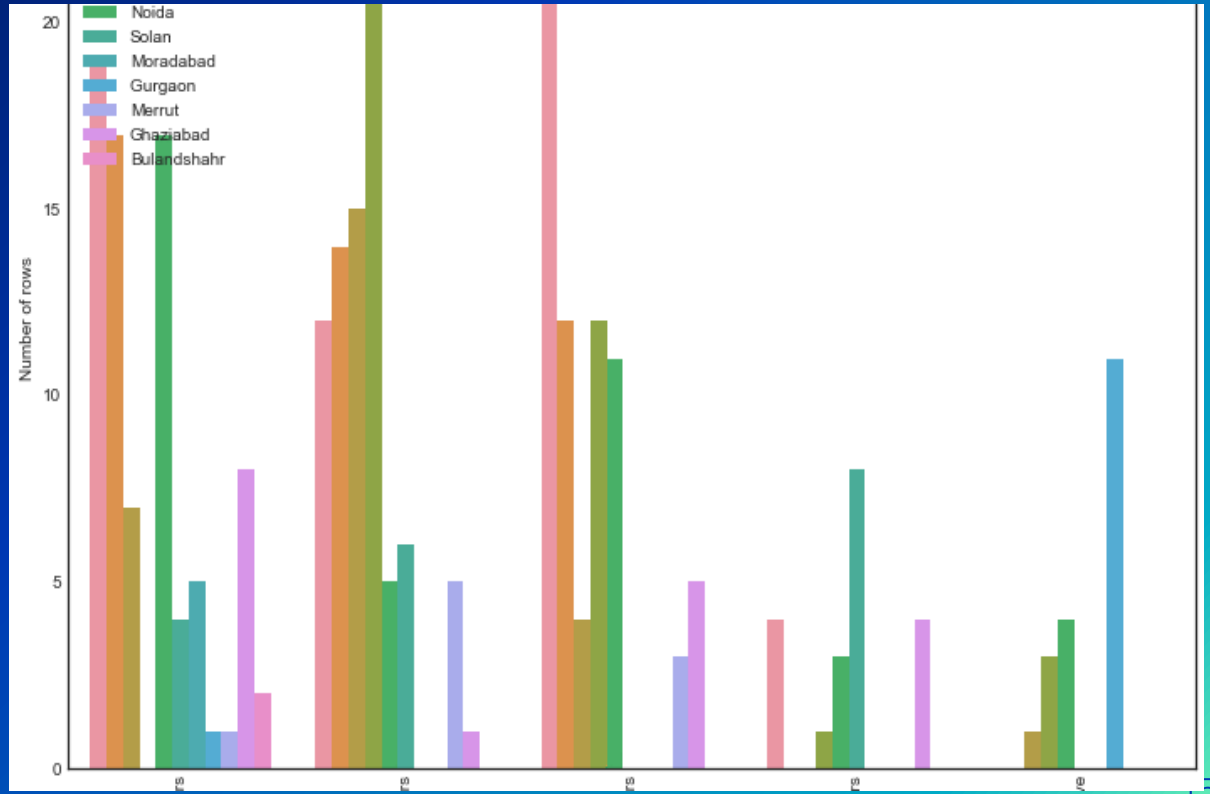
- The hue used in this plot is the “Age” column showing the legend accordingly.



Bivariate Analysis

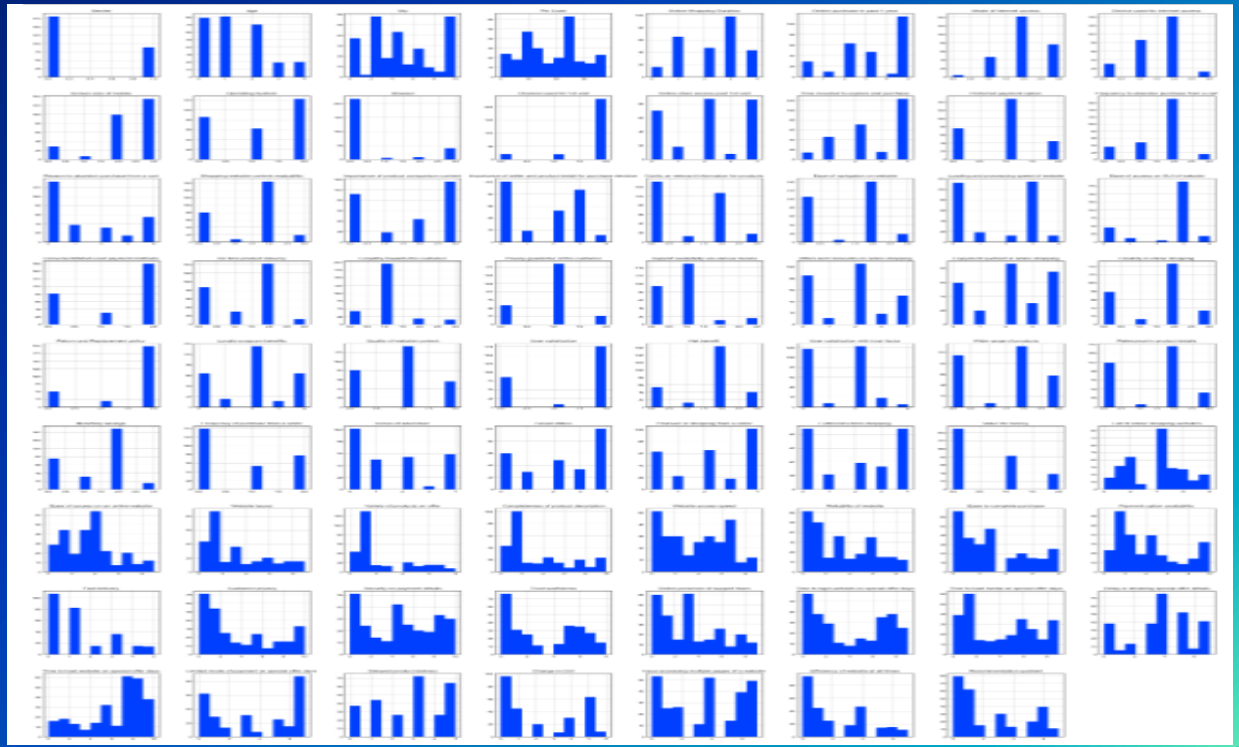
- Using the count plot along with the hue parameter I was able to generate analysis details comparing 2 columns of the dataset where the hue remained constant while the other one kept changing inside a loop.

- The hue used in this plot is the “City” column showing the legend accordingly.



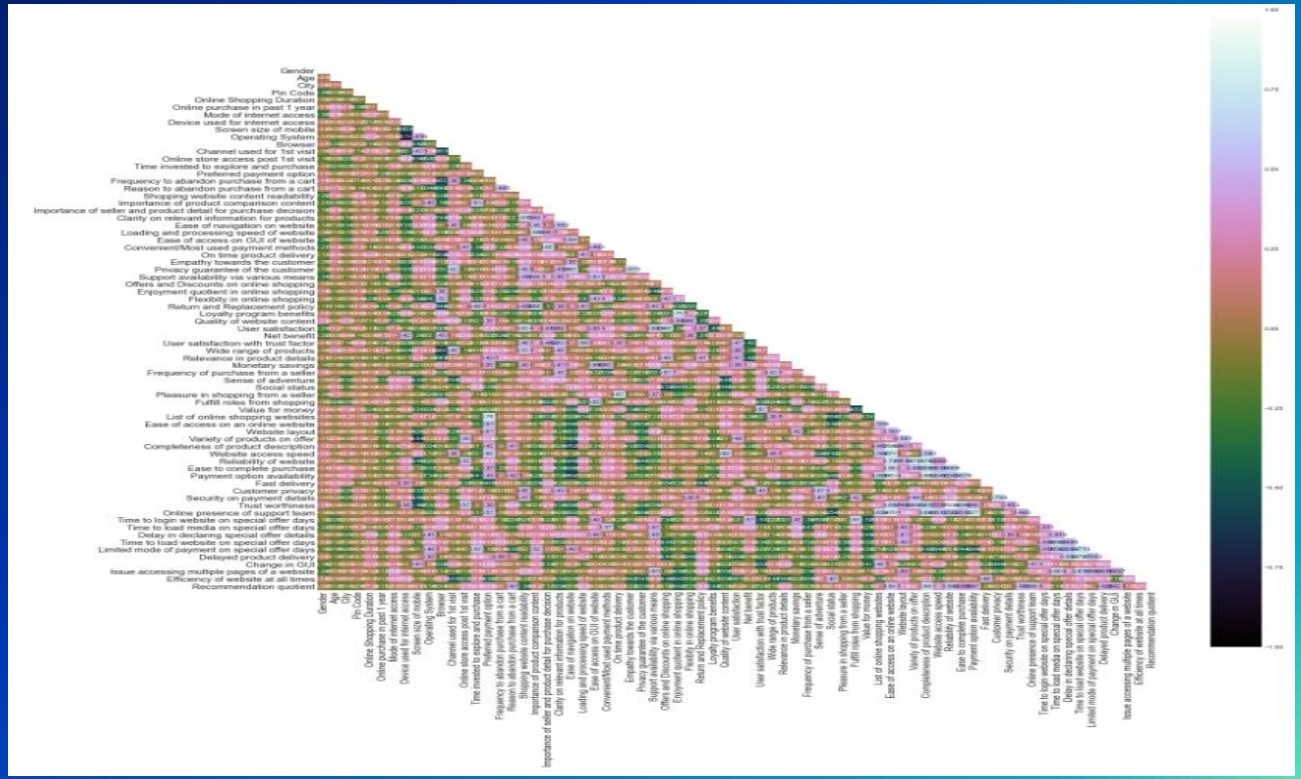
Multivariate Analysis

- I used the histogram after encoding all the columns of our dataset.
- Since histogram only understand numeric values it would not have generated a diagrammatic representation of object data type columns.
- Once the encoding was complete the histograms showed the information.



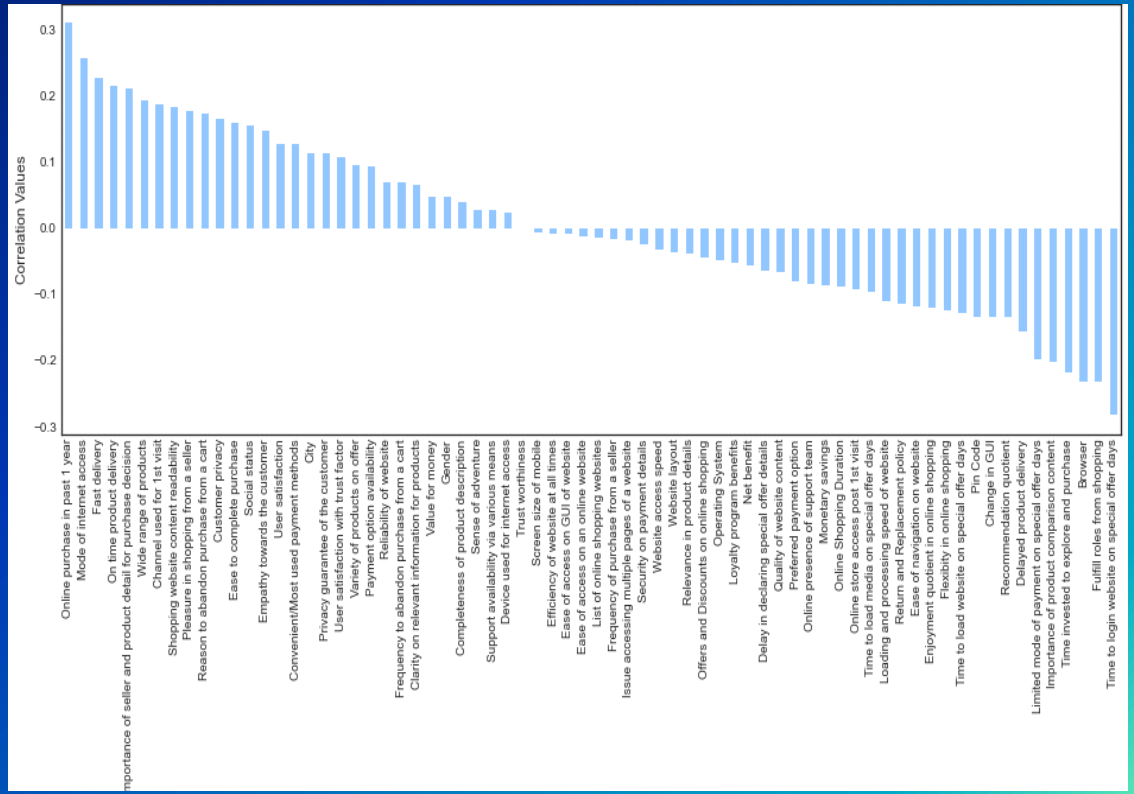
Multivariate Analysis

- I used the heat map on the encoded dataset to see the correlation details between the columns.
- Even on the Jupyter Notebook the picture was too tiny however seeing the color combinations I was able to figure out that there is no multi co linearity concern between the columns.



Correlation between the columns

- With the help of correlation technique I generated separate bar plots to check the positive and negative correlation details in our dataset.
- Positive correlation - A correlation of +1 indicates a perfect positive correlation, meaning that both variables move in the same direction together.
- Negative correlation - A correlation of –1 indicates a perfect negative correlation, meaning that as one variable goes up, the other goes down.



Inference

- Based on overall observations the first 47 features provide insights on how e-retail is helpful and growing based on customer inputs. The data explained how the online platform has been used more often in which CITY, PIN CODE, AGE etc. It also showed us that in some factors there is less importance given to contribute to the success of an e-commerce store, so based on that we could remove those factors and keep all the important factors. Also we could improve on some factors that influence the online customers repeat purchase intention.
- Apart from the first 47 features the rest of the features showed which online platform has been used more based on the success factors. Based on the case study for customer activation and retention, Amazon is the most reliable and has been fulfilled all the customer requirements. After Amazon the data showed Flipkart has been used more for online shopping.
- The case study from Indian e-commerce customers showed Amazon and Flipkart has been used mostly for Online Shopping and most recommended by Friends. So based on the research factors Amazon and Flipkart are the e-commerce platform which are having the combination of both utilitarian and hedonistic values to keep the repeat purchase intention (loyalty) positively.

1. Amazon.in

- **To be improved:**

1. During promotions, try to give a disturbance free shopping experience to customers.
2. Give more payment options to customers.
3. Try to give price early during promotion.
4. Reduce the delivery time of the products.

- **Positive feedback summary:**

1. Convenient to use and also a good website for shopping.
2. Fast delivery of products.
3. Availability of complete information of the products.
4. Presence of online assistance through multi-channels.
5. Reliable website or app, perceived trustworthiness.

- ***Online e-commerce Company:***

2. Flipkart.com

To be improved:

1. During promotions, try to give a disturbance free shopping experience to customers.
2. Give more payment options to customers.
3. Try to give the price early during promotion.
4. Reduce the delivery time of the products.
5. Flipkart and Amazon almost share the same feedbacks with varying percentages as the only difference.

Positive feedback summary:

1. Convenient to use and also a good website for shopping.
2. Fast delivery of products.
3. Availability of complete information of the products.
4. Presence of online assistance through multi-channels.
5. Reliable website or app, perceived trustworthiness.
6. Wild variety of products to offer.

Online e-commerce Company:

3. Myntra.com

- **To be improved:**

1. During promotions, try to give a disturbance free shopping experience to customers.
2. Try to give the price early during promotions.
3. Reduce the delivery time of the products during promotions.

- **Positive feedback summary:**

1. Convenient to use and also a good website.
2. Availability of several payment options.
3. Faster products delivery.
4. Complete information of products available.
5. Reliable website or app, perceived trustworthiness.
6. Wild variety of product to offer

- *Online e-commerce Company:*

4. Paytm.com

To be improved:

1. Reduce the delivery time of the products during promotions.
2. Try to give the price early during promotion.
3. During promotions, try to give a disturbance free shopping experience to customers.
4. Late declaration of price and discounts.
5. Frequent disturbance is occurring while moving from one page to another.

.

Positive feedback summary:

1. Convenient to use and a good website.
2. Quickness to complete a purchase.
3. About 64% of the customers feel that either web or app is reliable.
4. Around 20% of the customers believe that Paytm has a wild variety of products on offer.

Online e-commerce Company:

5. Snapdeal.com

To be improved:

1. Reduce the delivery time of the products during promotions.
2. Try to give the price early during promotion.
3. During promotions, try to give a disturbance free shopping experience to customers.
4. Late declaration of price and discounts.
5. No one has expressed to recommend Snapdeal to a contact as it has the most negative feedbacks among all other websites.

Online e-commerce Company:

Positive feedback summary:

1. Convenient to use.
2. 54% of the customers are happy about the availability of financial information security.

Conclusion & Future Work

- I will need to perform some preprocessing on the data for example using the scaling techniques.
- Since I have mostly categorical data present in the dataset I am not going to worry about removing outliers or skewness.
- Need to build some unsupervised machine learning models.
- Will have to verify the clustering or association algorithm details that can be used on the dataset
- Some algorithms that I intend to work upon are k-means clustering, k-nearest neighbors for unsupervised machine learning, hierarchal clustering, apriori algorithm and neural networks.


Thank you...