Presentation on

# Customer retention

# project

Prepared by

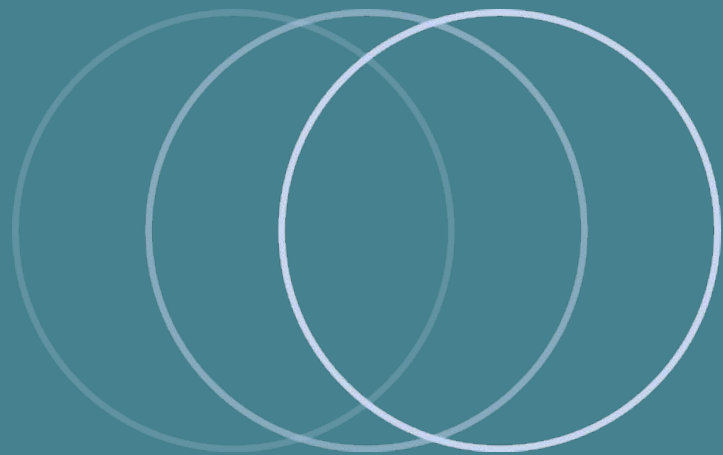Wesley Sirra

**FLIP ROBO**

FlipRobo
Technologies

# Matters on the Docket

A brief look at what
I will discuss on this report

# Introduction

## 1 What is Customer Retention?

A metric that measures customer loyalty

In addition to identifying the number of loyal customers, it can reflect or predict customer satisfaction, repurchase behavior, customer engagement and emotional ties to a brand.

## 2 e-Retail in India

- E-commerce is expanding steadily in India

- Customers have an increasing choice of products at competitive rates

- E-commerce is probably creating the biggest revolution in the retail industry, and this trend will continue in the years to come.

## 3 Objective of the Study

- To find out key factors influencing shoppers while choosing online retailers

- To study the consumer perceptions towards online retailers in India

- To compare the customer perceptions regarding selected online retailers

# Problem Statement

## 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.

# Data Analysis Process

**01**

## Loading Dataset and libraries

Importing basic libraries like Pandas, Numpy and loading the dataset which is in excel.

**02**

## Data Exploration

Checking the Data type of the Data, checking null values, unique and value counts, duplicate values and the Describe function.

**03**

## Data Cleaning

Replace duplicate values, Rename the columns, changing the datatypes.
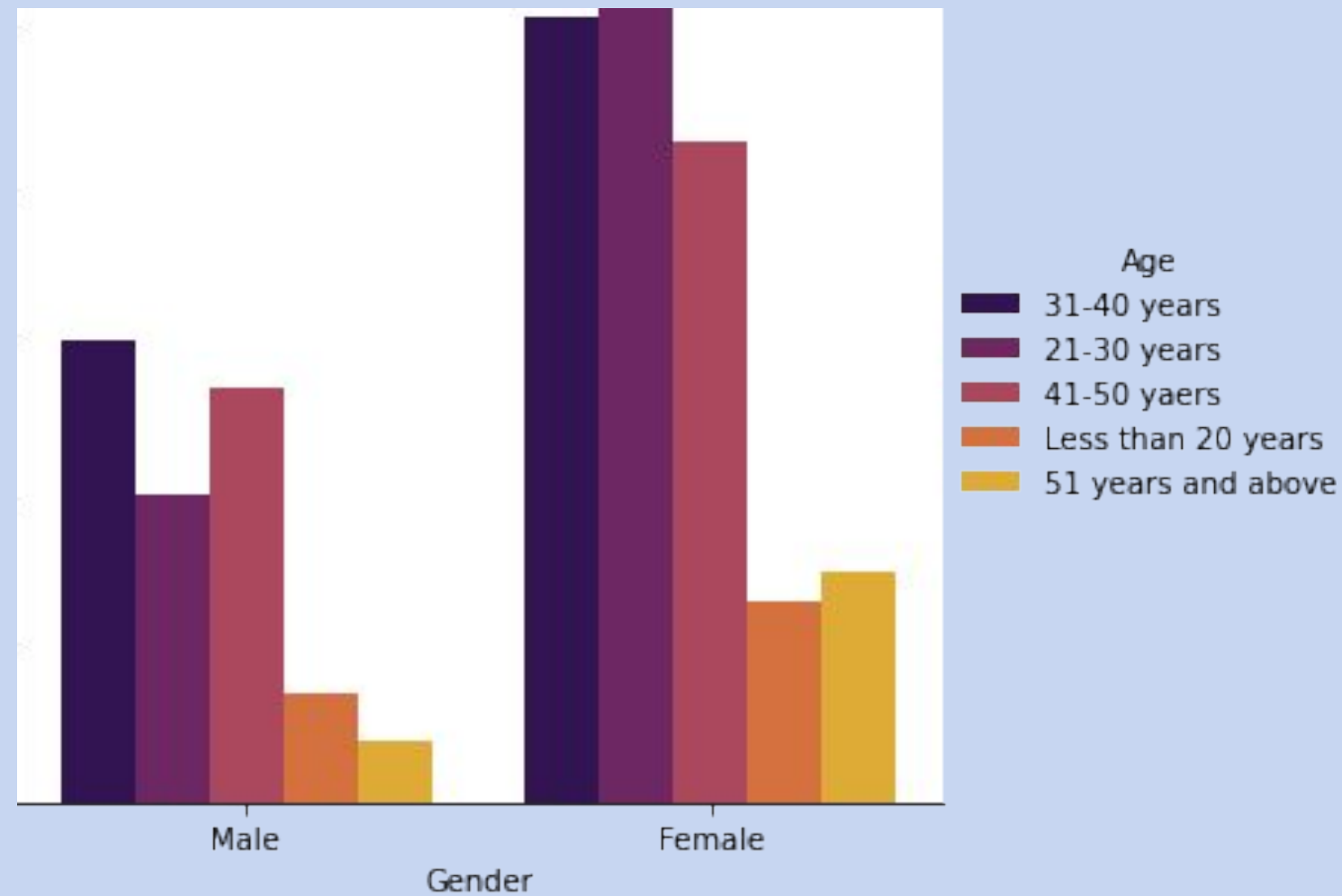
**04**

## Data Visualization

Analyze the data using Univariate analysis and Bivariate. Plotting the Data.

**05**

## Feature Engineering and Model Building

Encoding the data, Scaling and data preprocessing for building the model. Hyperparameter tuned and saved the best Model
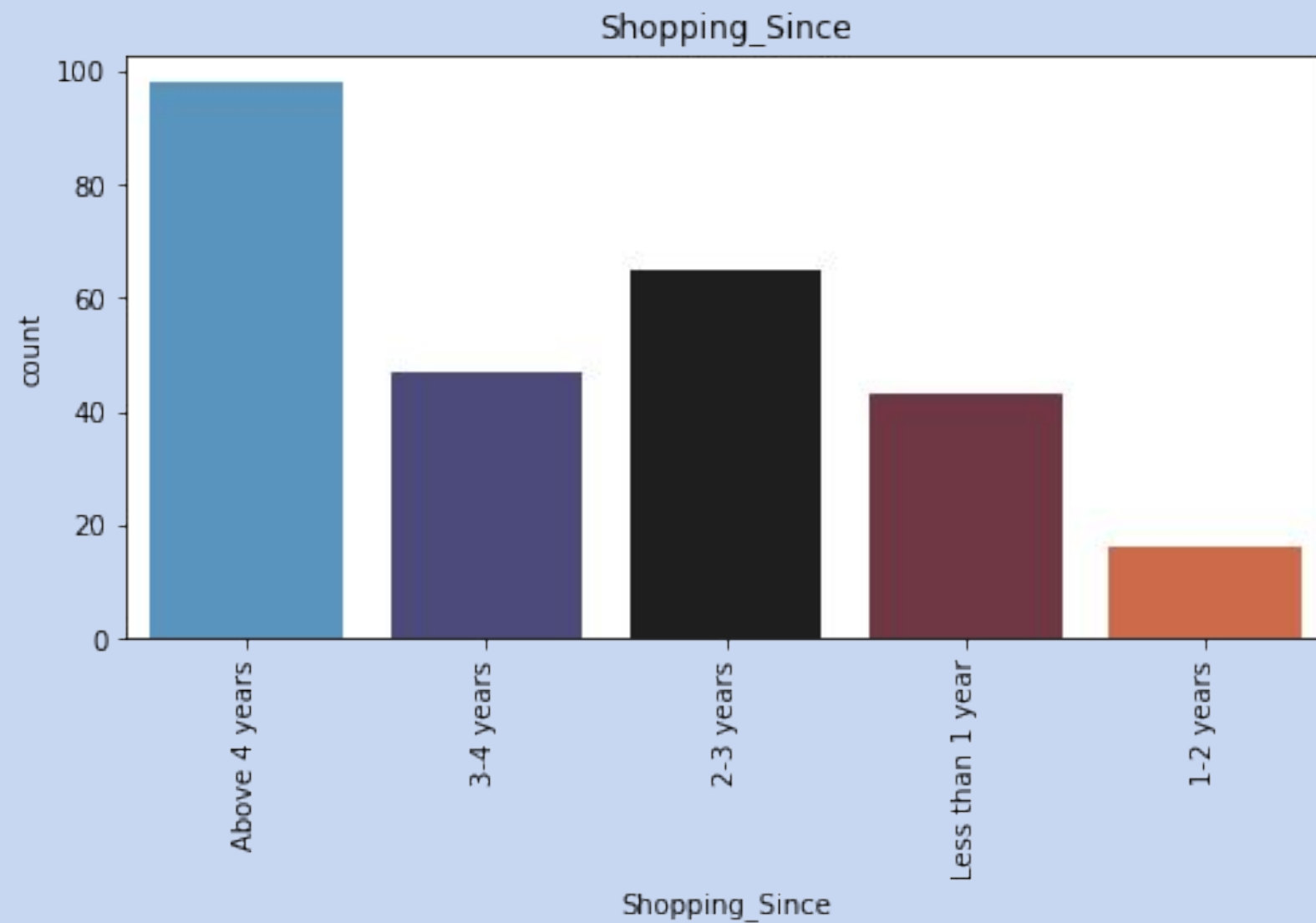
- Most of the online shoppers are FEMALES

- Males below 20 years and above 51 years uses less ecommerce websites.

- Females between the age of 21-30 shops more according to the data.

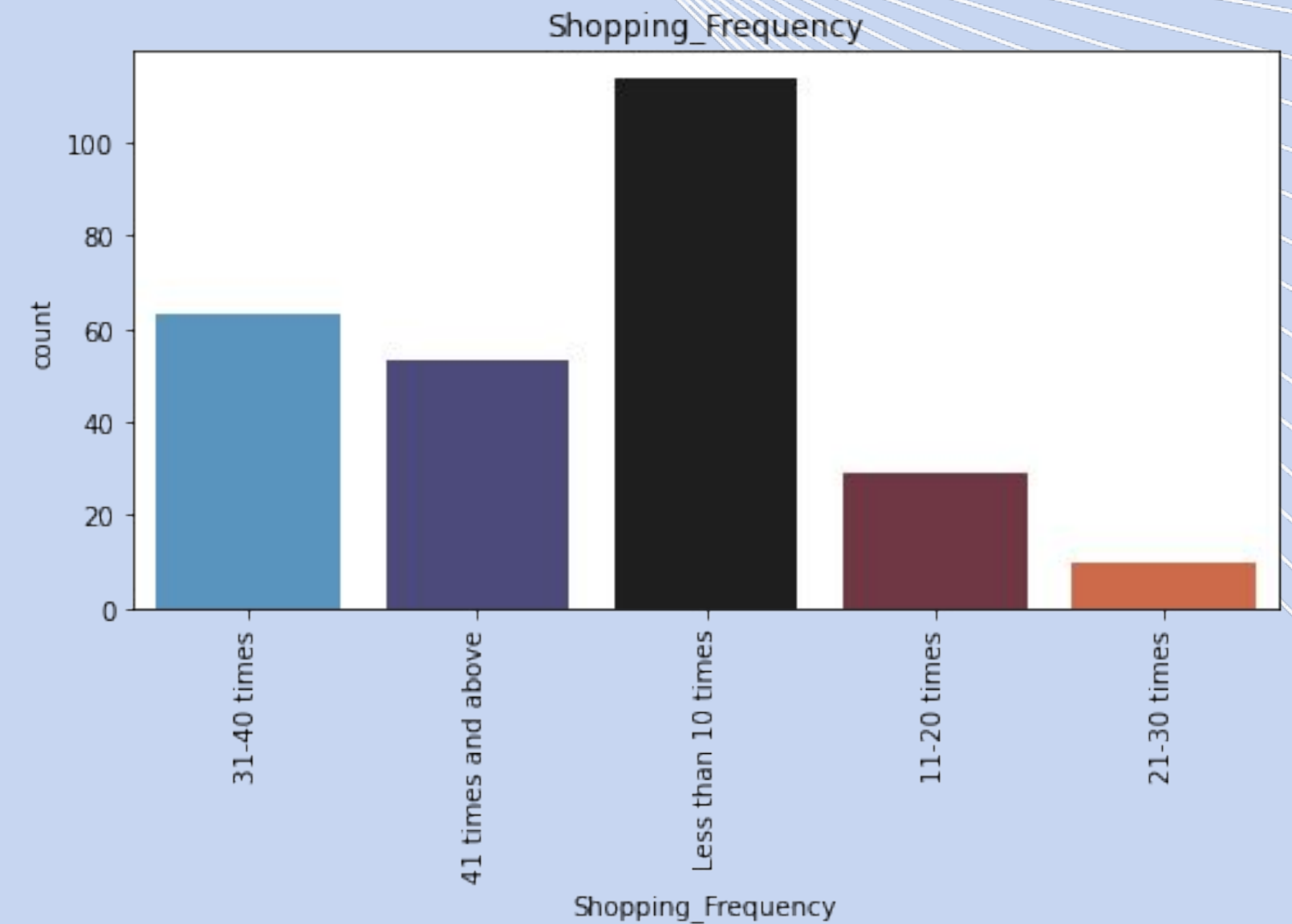- Males aged between 31-40 shops more among males category.

- According to the data, most shoppers shop from Delhi and greater Noida and the least shops from Bulandshahr.

# Data Visualization

Most shoppers are shopping online more than
4 years, followed by 2-3 years and the least number of
shoppers in the data started shopping for 1-2 years only

The shopping frequency for the past 1 year is
Less than 10 times for most of the shoppers.
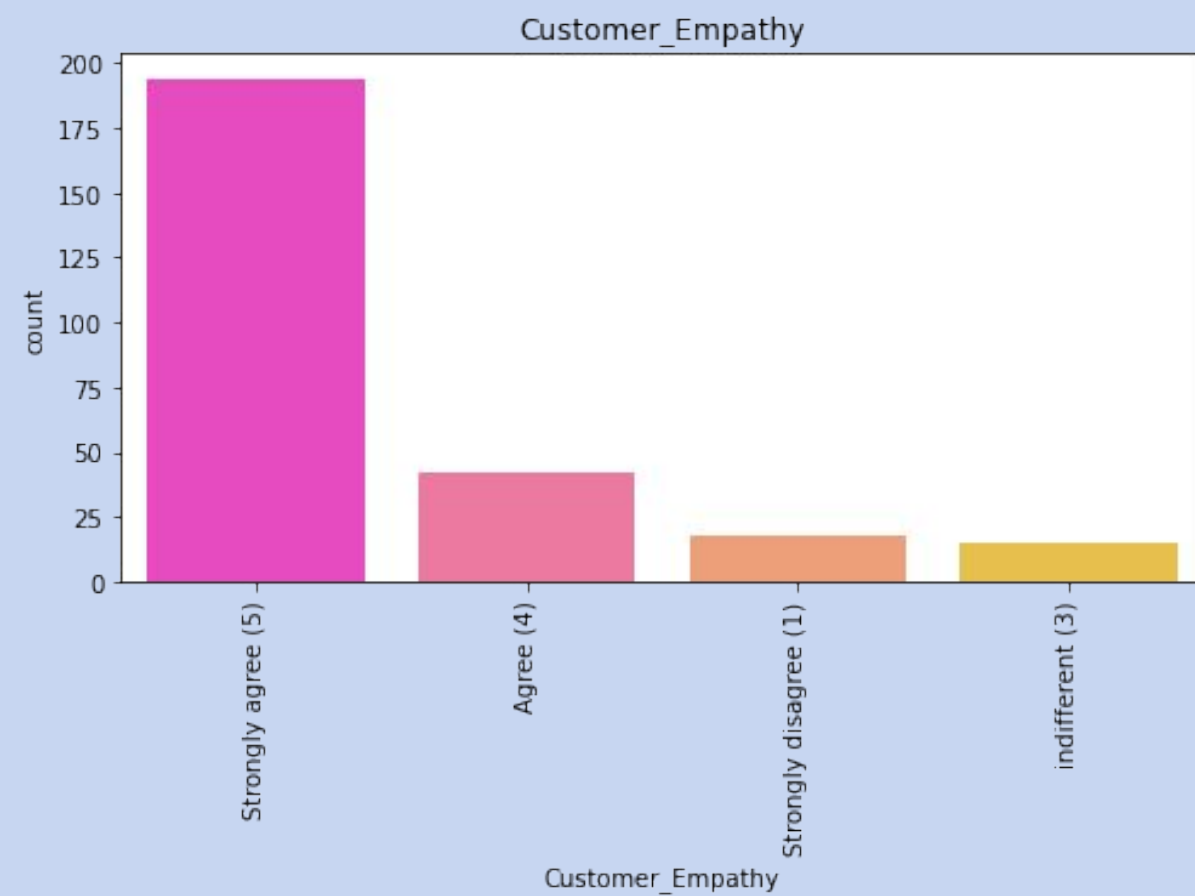
# Data Visualization

- Mobile Internet is used by most of the shoppers.
- Many use Smartphones followed by laptops.
- Operating systems used by most shoppers are windows and followed by Android.
- Most of the shoppers use Google chrome as their browser.
- Most of the shoppers came to the website for the first time through search engines.
- Most of the shopper's login to the website through Search Engines and Mobile Applications.

- Most shoppers spent more than 15 minutes before purchasing.
- A few of the shoppers notified that they used less than a minute for the purchase.
- E-retail customers mostly use Debit/credit card for purchasing.
- Most of the Shoppers abandon the cart sometimes and very few abandon very frequently.
- Most shoppers abandon the cart because they get a Better alternative Offer.
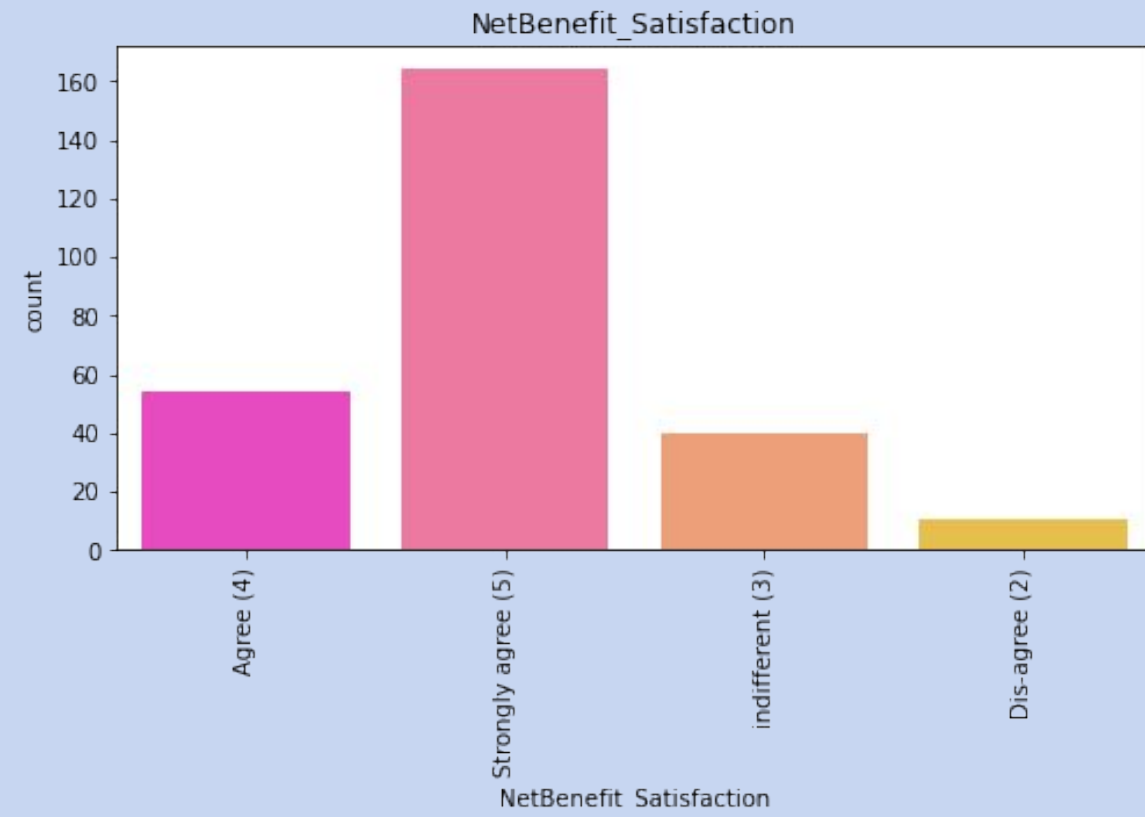
# Data Visualization

Content_Readability
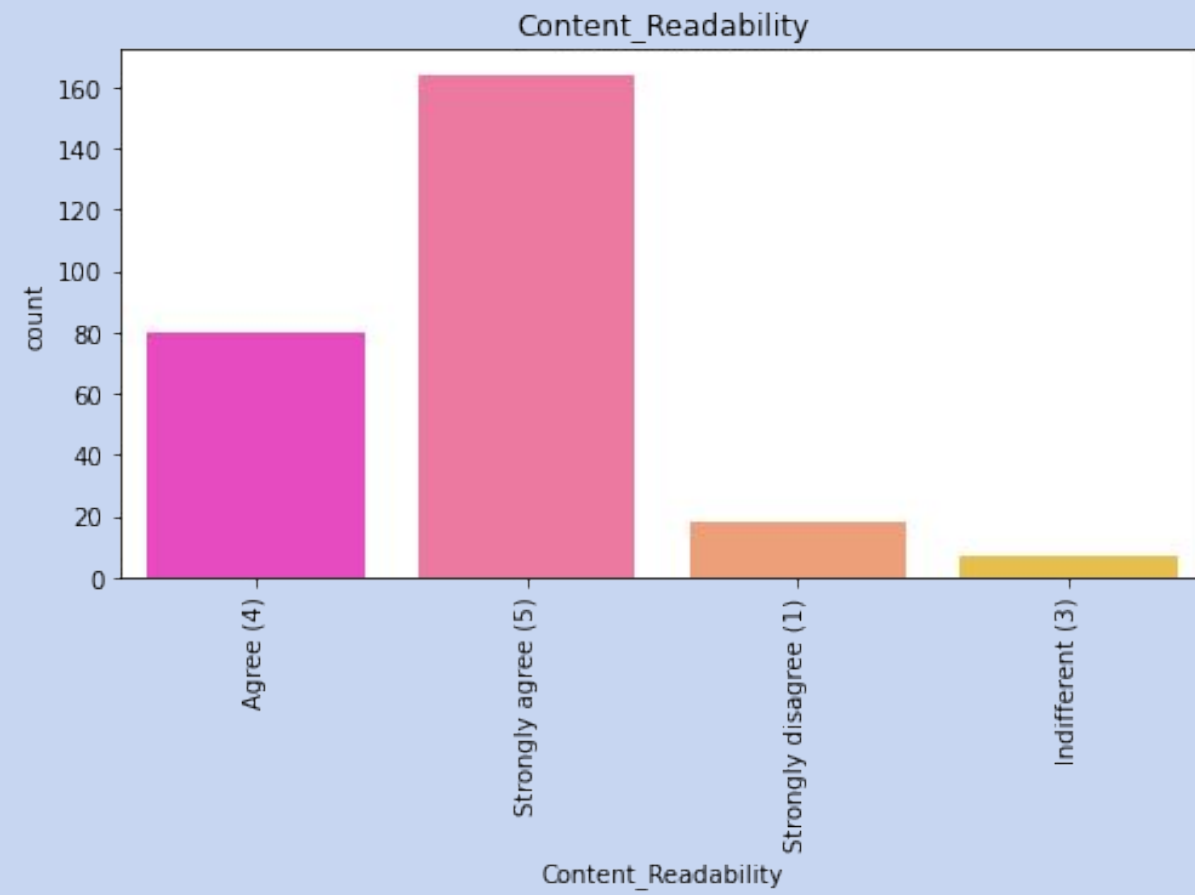
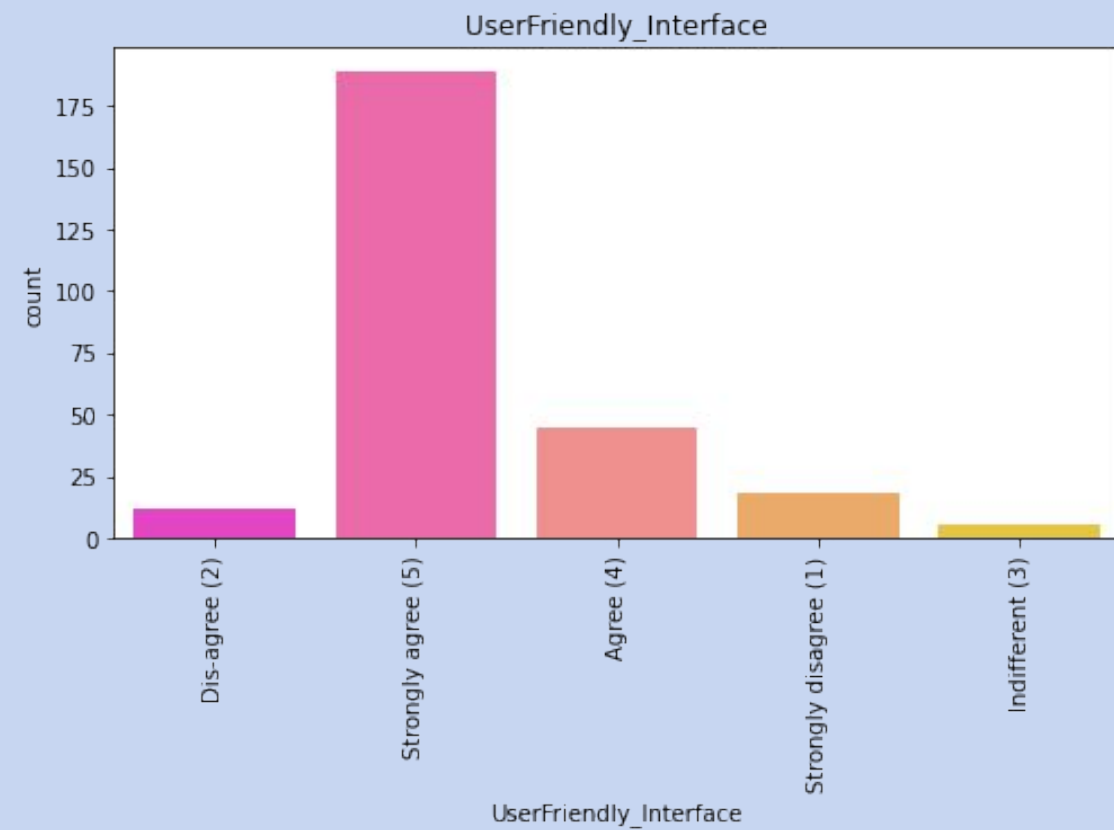- Most shoppers strongly agree that contents should be easy to read and understand



Customer_Empathy

- Most Shoppers strongly agree that they expect Empathy (readiness to assist with queries).

# Data Visualization

NetBenefit_Satisfaction

- Most Shoppers strongly agree that Net Benefit derived from shopping online can lead to users' satisfaction.



UserFriendly_Interface

- Most Shoppers strongly agrees that Friendly user interface is an important decision-making factor

# Data Visualization

- Most shoppers strongly agree that
  - *similar products to the content highlighted is important for product comparison.*
  - *seller and product complete information is important for purchase decisions*
  - *Trust that the online retail store will fulfill its part of the transaction at stipulated time*
  - *to have the ability to guarantee privacy of the customer*
  - *to have convenient payment methods*
- Most Shoppers strongly agree to have Responsiveness, availability of several communication channels
  - *(email, online rep, twitter, phone etc.).*

- Most Shoppers strongly agree that
  - *Online shopping gives monetary benefits and discounts.*
  - *Enjoyment is derived from shopping online.*
  - *Shopping online is convenient and flexible.*
  - *Return and replacement policy of the e-tailer is important for purchase.*
  - *Gaining access to loyalty programs is a benefit of shopping online.*
  - *Displaying quality Information on the website improves customer satisfaction.*

# Data Visualization
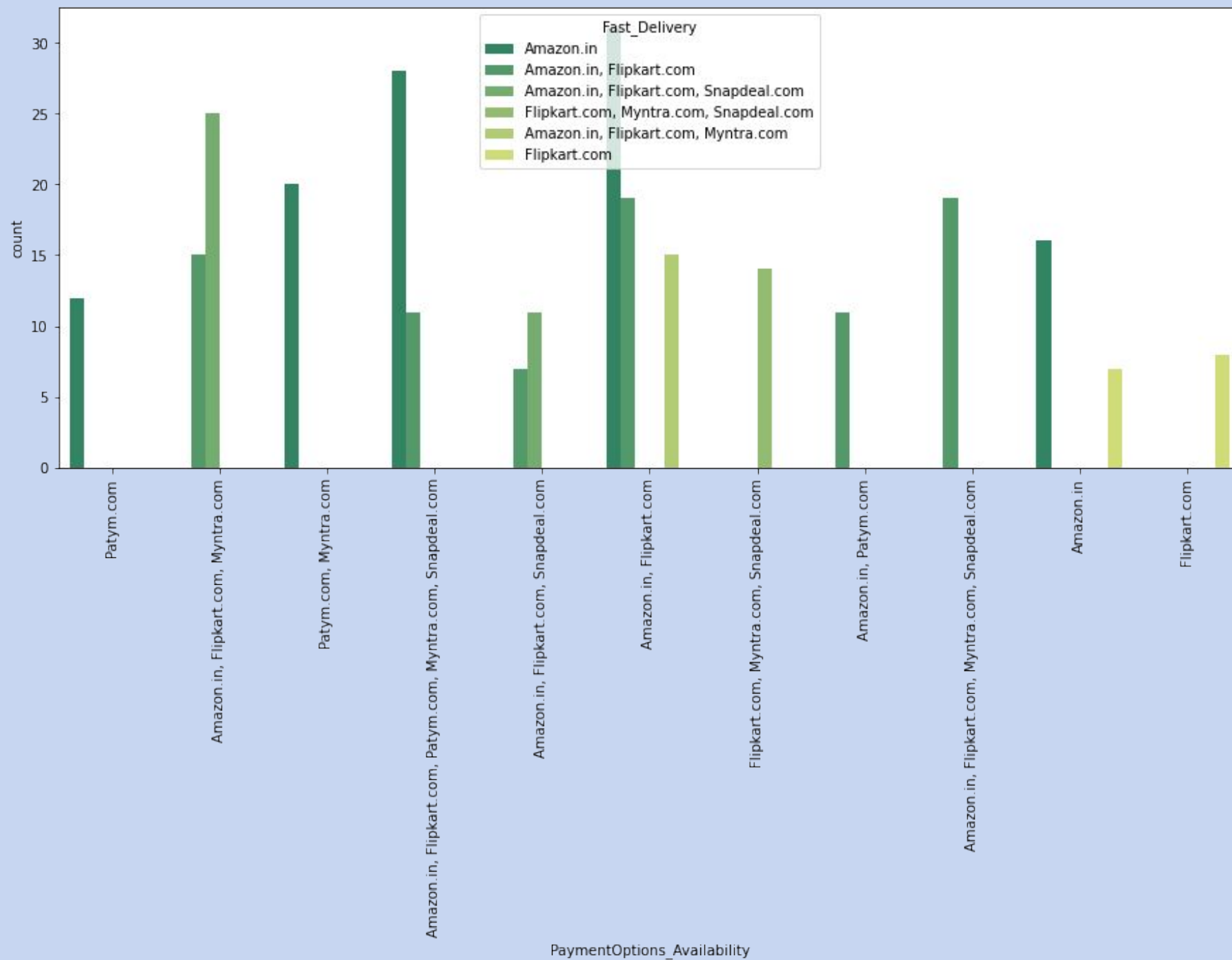
- Payment Option Availability Vs Fast Delivery

- Best Payment option and faster delivery is Amazon and Flipkart by most of the shoppers.
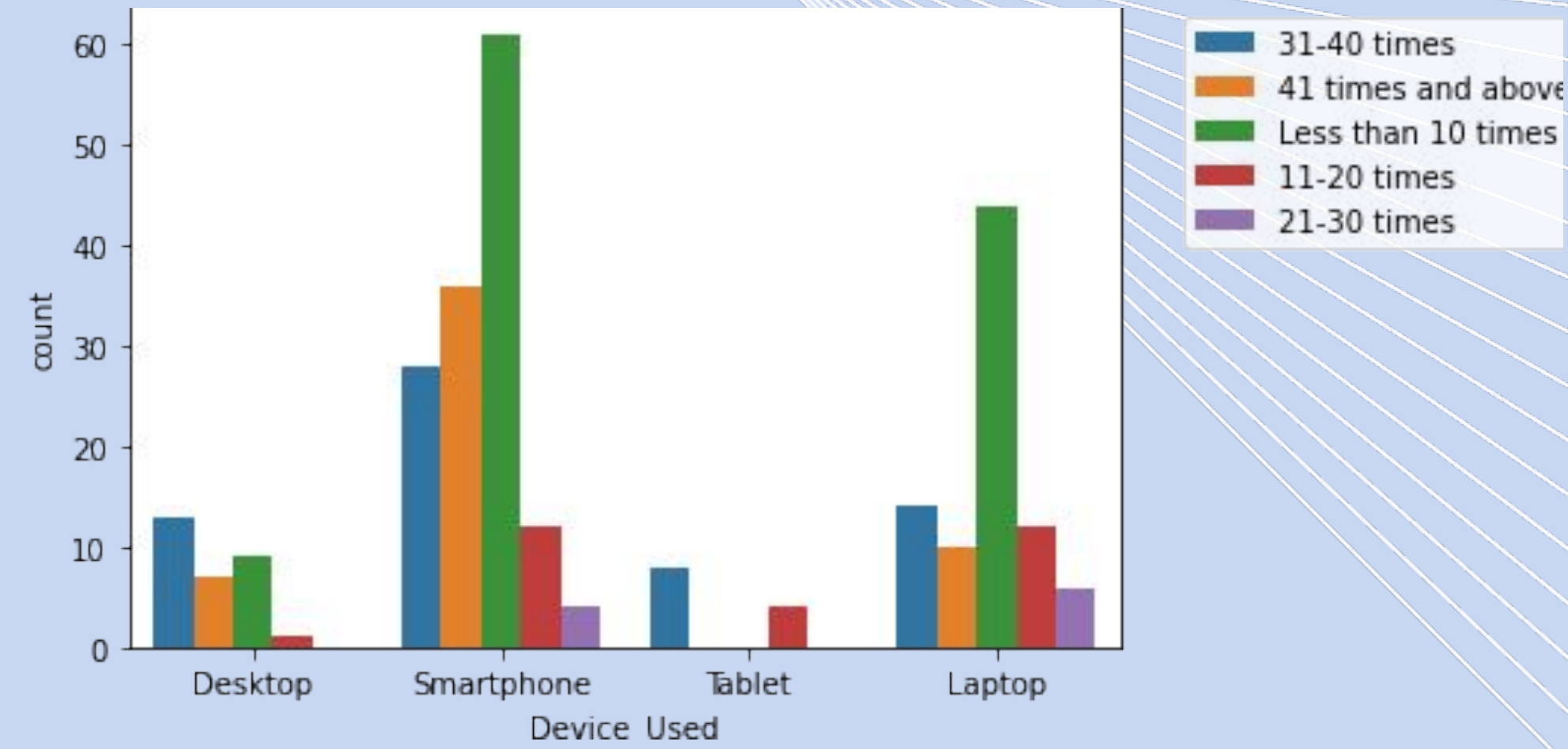


- Device Used Vs Frequency of shopping

- Most people used smartphones and shopping frequency was less than 10. Second most shopping frequency is above 40 times for smartphone users.

- Tablets users' shopping frequency is 10-20 and 30-40.

- Laptop users shopping frequency is more at 0-10

# Data Visualization
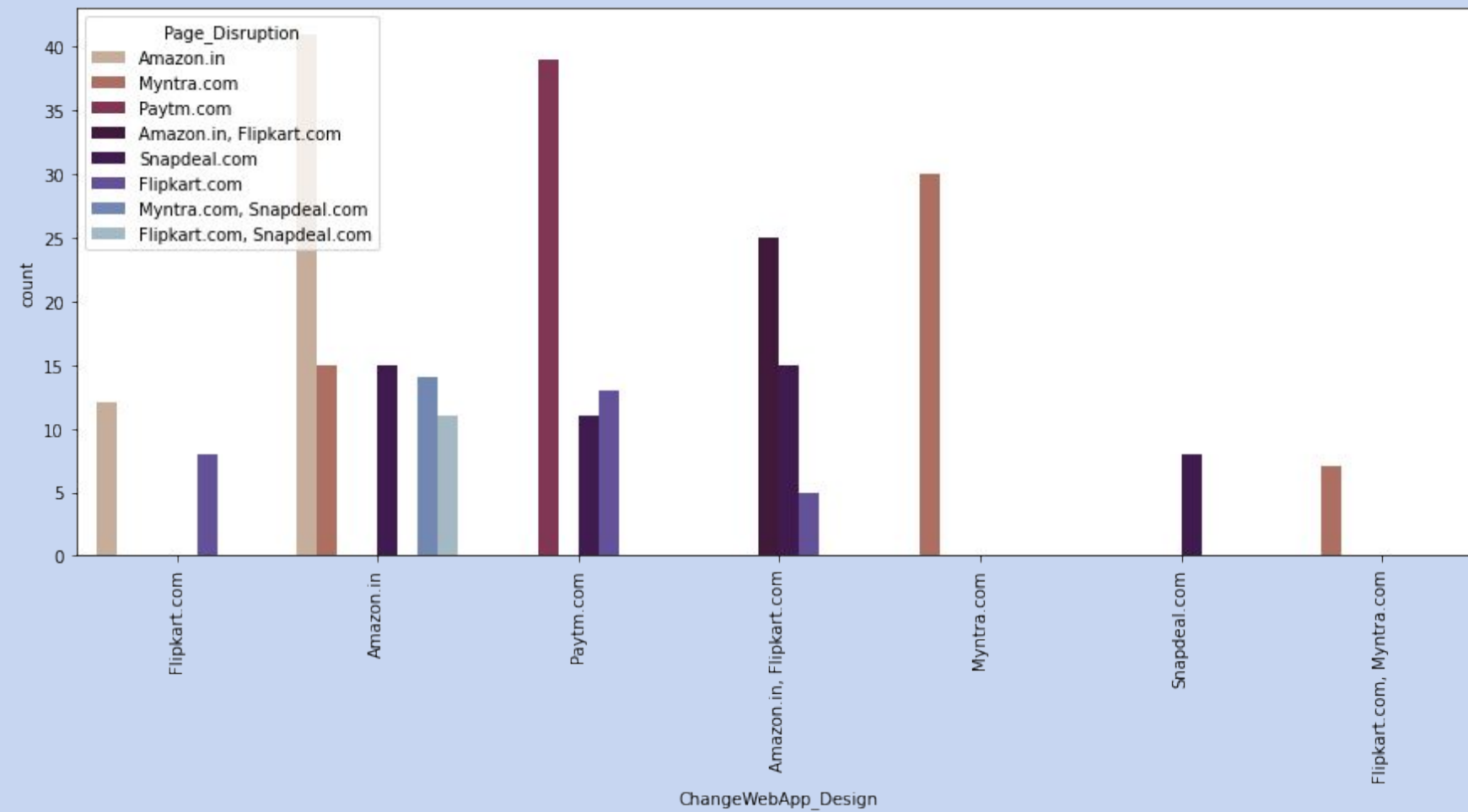
- Change web/App design Vs Page Disruption
- Most shoppers say Amazon has the best design for the website and a major drawback is that it is also the Most page disrupted.
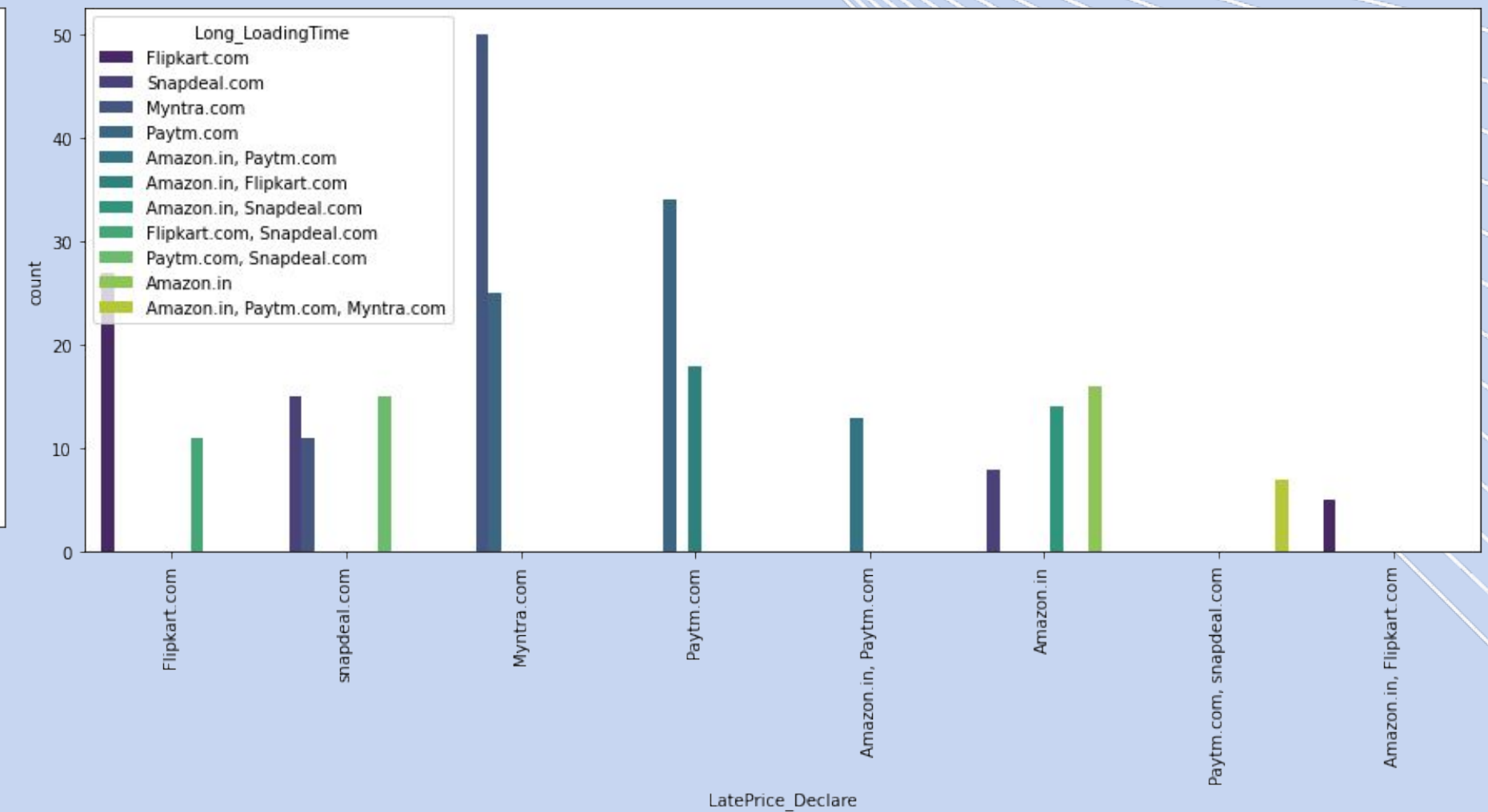
- Late Price declaration Vs Long Loading Time
- Major drawback for Myntra.com is Late declaration of price and Long Loading time during sale and Promotion time.

# Data Visualization

- Most people say that Snapdeal has the least payment options followed by Amazon. Even then, Amazon is the most recommended but Snapdeal comes in least recommended.
- Even though Login time is more for Amazon, It is the most recommended.
- We can see that the density of female customers is more than male.
- Most People abandon cart due to Better alternates.
- Very Frequently has been marked by few shoppers and the reason is Promo code not applicable.
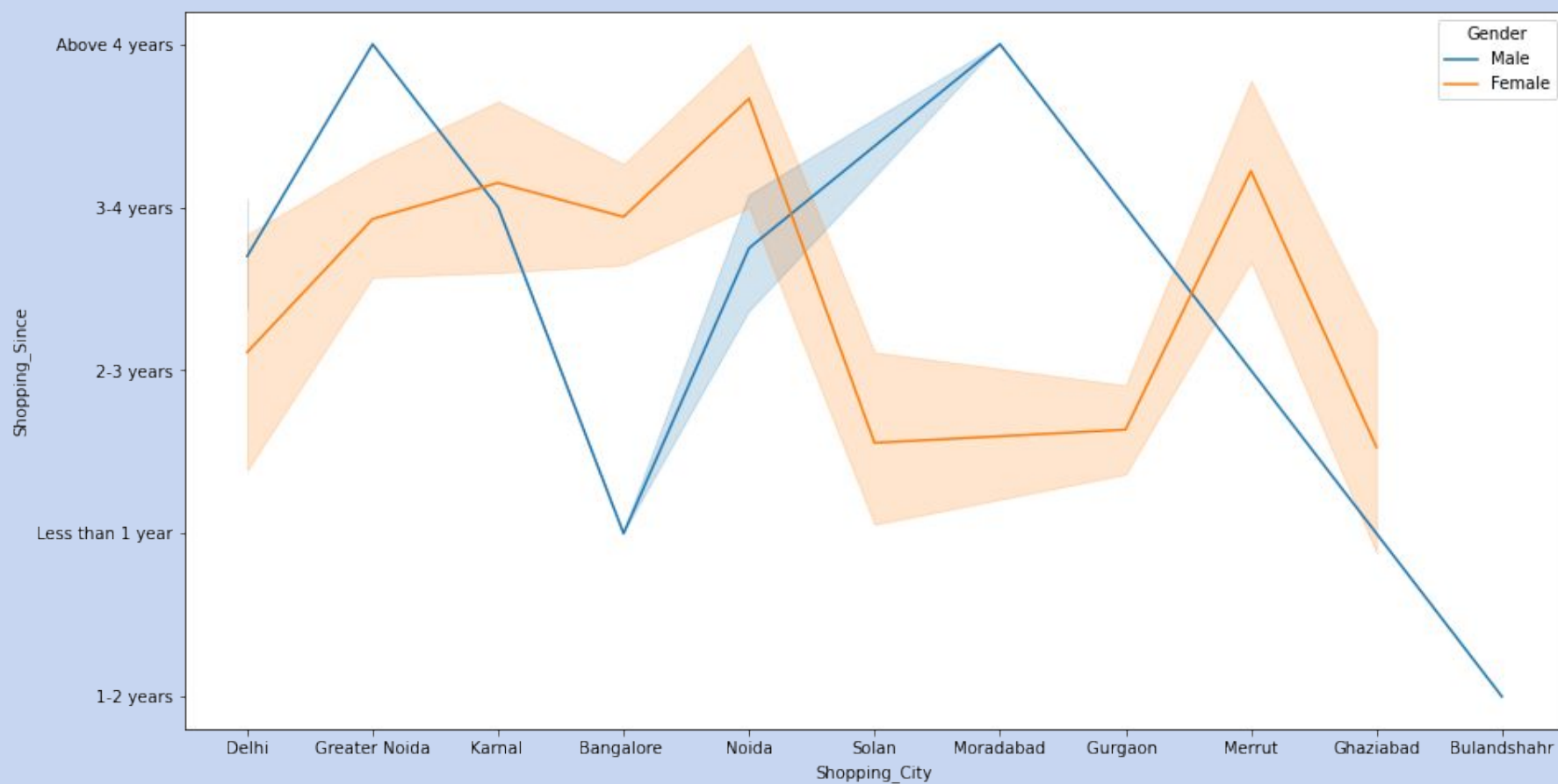


- Shopping City Vs Shopping Since, Hue-Gender
- Men living in Bangalore and Ghaziabad have shopped online for less than 1 year.
- More Number of men shopping online from greater Noida and Moradabad are using online portals for more than 4 years.
- Women from Meerut and Noida have shopped the longest.

Data Visualization

# Model Building

- Feature Engineering

- Data Preprocessing

- HyperParameter Tuning

- Saving the Best Model

- Feature Importance

```
Final_Model=RandomForestClassifier(max_depth=20, min_samples_leaf=3,min_samples_split=4,n_estimators=100)
Final_Model.fit(xtrain,ytrain)
pred=Final_Model.predict(xtest)
acc= accuracy_score(pred,ytest)
score=cross_val_score(Final_Model,x,y,cv=10)
print('The Accuracy score is:', acc*100)
print('The cross validation score', (score.mean()*100))

The Accuracy score is: 100.0
The cross validation score 100.0

print('Confusion Matrix')
print(confusion_matrix(pred,ytest))

Confusion Matrix
[[25  0  0  0  0  0  0  0]
 [ 0 21  0  0  0  0  0  0]
 [ 0  0  4  0  0  0  0  0]
 [ 0  0  0  6  0  0  0  0]
 [ 0  0  0  0  4  0  0  0]
 [ 0  0  0  0  0  5  0  0]
 [ 0  0  0  0  0  0 12  0]
 [ 0  0  0  0  0  0  0  4]]
```

Encoding and Scaling has been done before building the Model

- Using the best Parameters, I have built the Final Model, Found the Accuracy Score, Cross validation Score and Built the confusion Matrix.
- Achieved Accuracy score of 100% and CV Score of 100% which is the best Score.
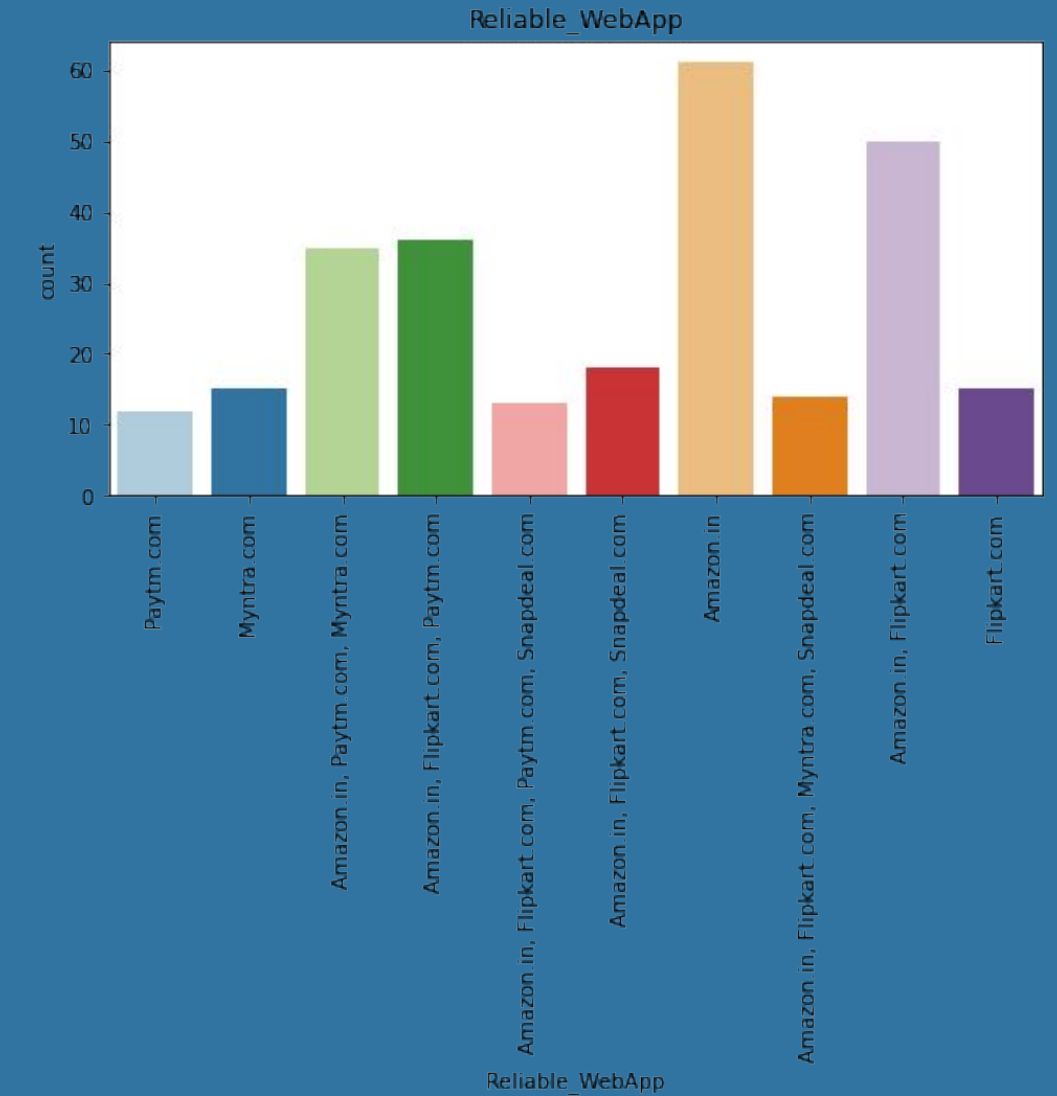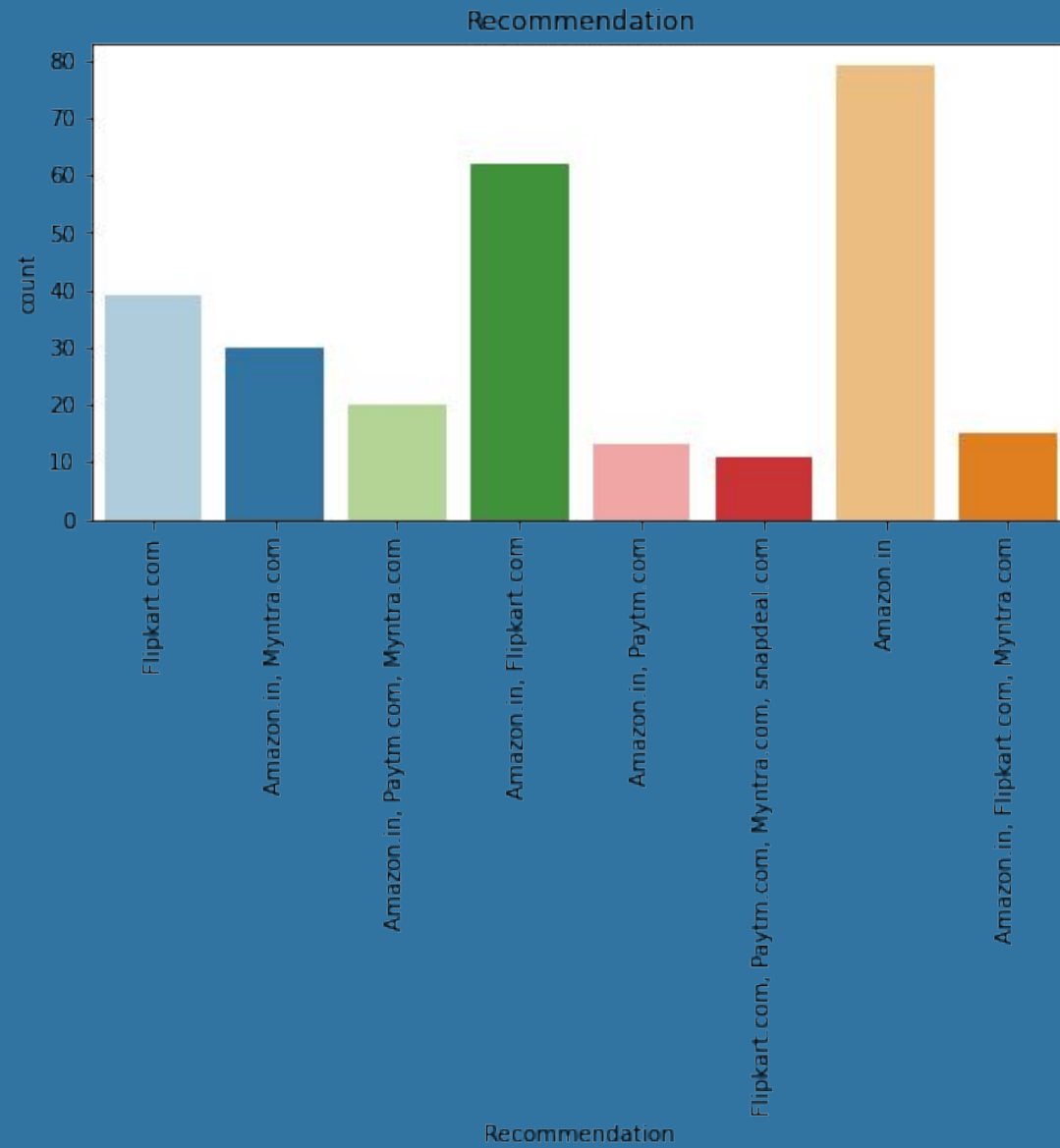- Saved the Final Model and found the Feature Importance.

```python
feat_importances = pd.Series(Final_Model.feature_importances_, index=a.columns)
feat_importances.nlargest(70).plot(kind='barh',figsize=(10,18))
```

Out[82]: <AxesSubplot:>
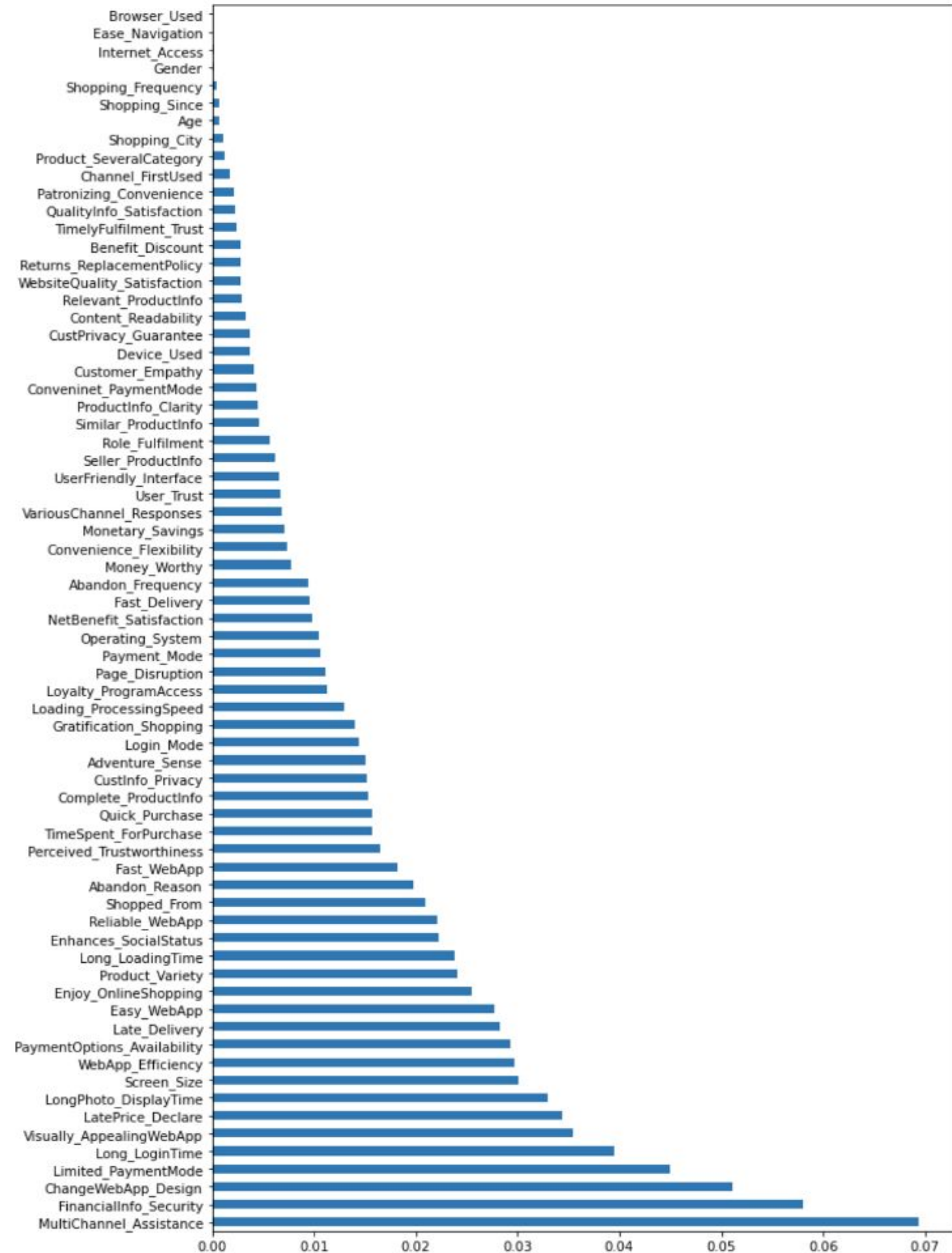


# Feature Importance

# Conclusions



- **There is high risk of customer churn with Myntra.com, Snapdeal.com and Paytm.com**
- The reasons why Customer retention is low for Myntra, Snapdeal and Paytm can be clearly inferred from above plots.
  - Low Customer Information Privacy
  - Late Delivery time
  - Long Loading time than other websites.
  - Price declaration is very late during Sale and promotions.
  - Low Payment Option

- **Amazon was the most preferred followed by Flipkart.**
- This can be explained easily by the previous result that we got. These two companies are most trusted in the industry and hence, have a huge reliability. These websites have the most lenient return policies as compared to others and the time required to process a return is low for these companies.

# Analysis

The results of this study suggest following outputs which might be useful for

E-commerce websites to extend their business:

- Females are more prone to shopping and so more feminine related products attract more females and hence improves the Customer retention.

- The shoppers want to be sure that it will be possible to return the product if he does not like it in real life.

- It is crucial for E-commerce to consider their customer satisfaction because this will retain customer loyalty as well as attract potential customers.

- The Financial Security, Perceived trustworthiness, the reliability of the Ecommerce website, all play an equally important role in deciding the buying behavior of online customers.

- All the websites were not equally preferred by online customers.

- The logistics factor, which includes Cash on delivery option, one day delivery and Descriptive factors like the Product information and Loading factors like Long Loading Time, Price declaration etc. plays a secondary role in this process though these are Must-be-quality.

- To conclude, Customer's Trust on the Company is the most important factor for Customer Retention. Factors like, Return Policy, Refund Policy, Fast Delivery, Wide Payment Channels, etc. helps the high-risk E Commerce websites to improve their Customer Retention Score.