CONTENT:-

Introduction

Problem Statement

Importing Required Libraries

Importing DataSet

Preprocessing Of Data

Data Visualization

Model Building

Conclusion

Introduction

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.

A). What is Customer Retention?

Customer retention means the process of maintaining or keeping customers once you have acquired them. It's all the activities that a company must do in order to keep their customers around. The goal is to build a long-lasting relationship between the brand and consumers. Once a customer becomes loyal to your brand, not only he will buy more from you than a normal customer but he'll spread good words about your business, increase your reputation.

B). Why do we need Customer Retention?

1. Lower Marketing Costs.

2. Repeat Purchases Means Repeat Profits.

3. Gain Valuable Feedback.

4. Sell At Premimum Price.

5. Word Of Mouth Advertising.

Problem Statement

The data is collected from the Indian online shoppers. Results indicate the e-retail success factors, which are very much critical for customer satisfaction. There are two sheets (one is detailed) and second is encoded in the excel file. You may use any of them by extracting in separate excel sheet. The number of column(s) is more than 47.

The features in the dataset are as follows:

1. Gender 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?

10. What is the operating system (OS) of your device?

11. What browser do you run on your device to access the website?

8

12. Which channel did you follow to arrive at your favorite online store for the first time?

13. After first visit, how do you reach the online retail store?

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

15. What is your preferred payment Option?

16. How frequently do you abandon (selecting an items and leaving without making payment your shopping cart?

17. Why did you abandon the "Bag", "Shopping Cart"?

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

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.

21. All relevant information on listed products must be stated clearly 22 Ease of navigation in website

23. Loading and processing speed

24. User friendly Interface of the website

25. Convenient Payment methods

26. Trust that the online retail store will fulfill its part of the transaction at the stipulated time

9

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

28. Being able to guarantee the privacy of the customer

29. Responsiveness, availability of several communication channels (email, online rep, twitter, phone etc.)

30. Online shopping gives monetary benefit and discounts

31. Enjoyment is derived from shopping online

32. Shopping online is convenient and flexible

33. Return and replacement policy of the e-tailer is important for purchase decision

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

35. Displaying quality Information on the website improves satisfaction of customers

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

39. Offering a wide variety of listed product in several category 40 Provision of complete and relevant product information

41. Monetary savings

42. The Convenience of patronizing the online retailer

43. Shopping on the website gives you the sense of adventure

44. Shopping on your preferred e-tailer enhances your social status

10

45. You feel gratification shopping on your favorite e-tailer

46. Shopping on the website helps you fulfill certain roles

47. Getting value for money spent

Above are all the feature of or dataset.

Importing DataSet

As seen on above Dataset has also been imported and printed as shown in above image.

Preprocessing Of Data

1. First I checked the shape of dataset for better understanding of data. (From which I find out there are total of 71 feature and 269 no of entries(row) in each columns).

2. Then I checked different feature's present in our dataset and there datatype (now I have clear idea about the datatype of each features).

3. Next I checked for null values if any present in our dataset (And I found out there is no null value present in our dataset).

4. Then I checked the unique value each feature contains and there counts (Now I have clear idea about the unique value and the counts of each features).

5. Then I proceeded further and did some feature engineering.

Data Visualizaton

Since all the features are categorical we can use only categorical, ploting to get better insight. And particularly I have used Count plot for Gender Of Responce, Which city do you shop online from, How old are you? , Area Of Pincode, Screen Size, Shopping Frequency, Internet Access, Operating System, etc. columns I had done the count plot.

The Observation Are :-

1) There is double the number of women than men who have taken this survey.

2) Most of the people are in their 30's followed by 20's, teenagers and senior citizen are the least in number.

3) Most of the people belong from delhi, noida and banglore, ambiguity can also be seen as noida has two categories (noida and grater noida) which need to be handled.

4) Most of the people shopping online have been shopping from a long time.

5) Majority of people shop online 10 times a year.

6) Almighty can also be seen for range 42 times and above which needs to be handled.

Gender

Shopping_City

Age

PinCode

Shopping_Frequency

Screen_Size

Internet_Access

Operating_System
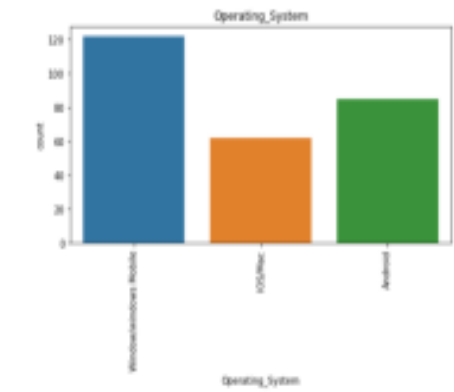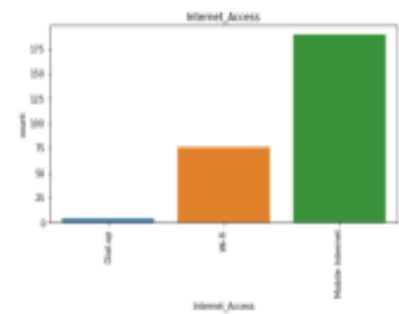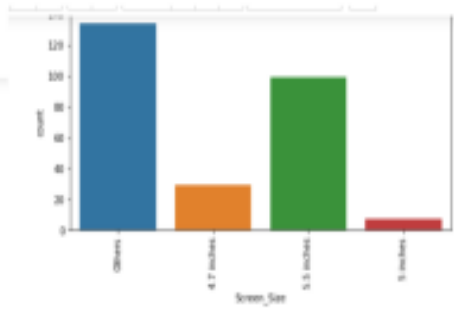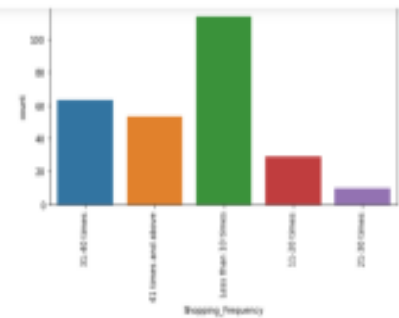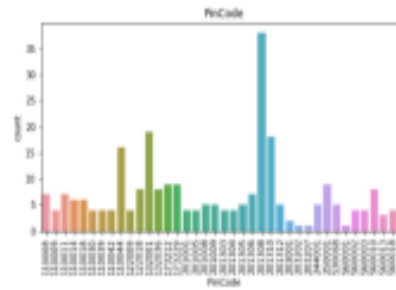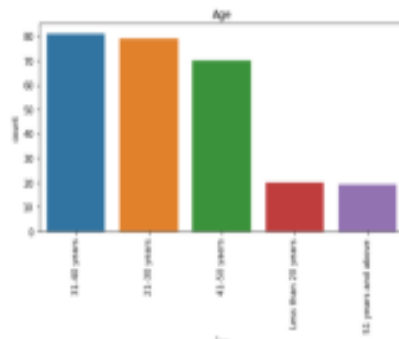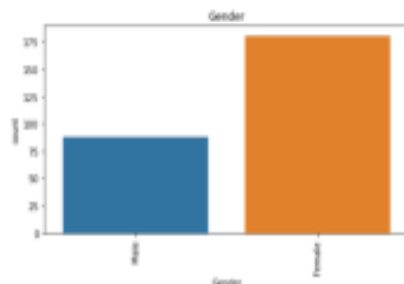
```
[5]: plt.figure(figsize=(15,8))
     sns.lineplot(x=data['city'],y=data['shopping_since'],hue=data['gender'])
```

[5]: <AxesSubplot: xlabel='city', ylabel='shopping_since'>
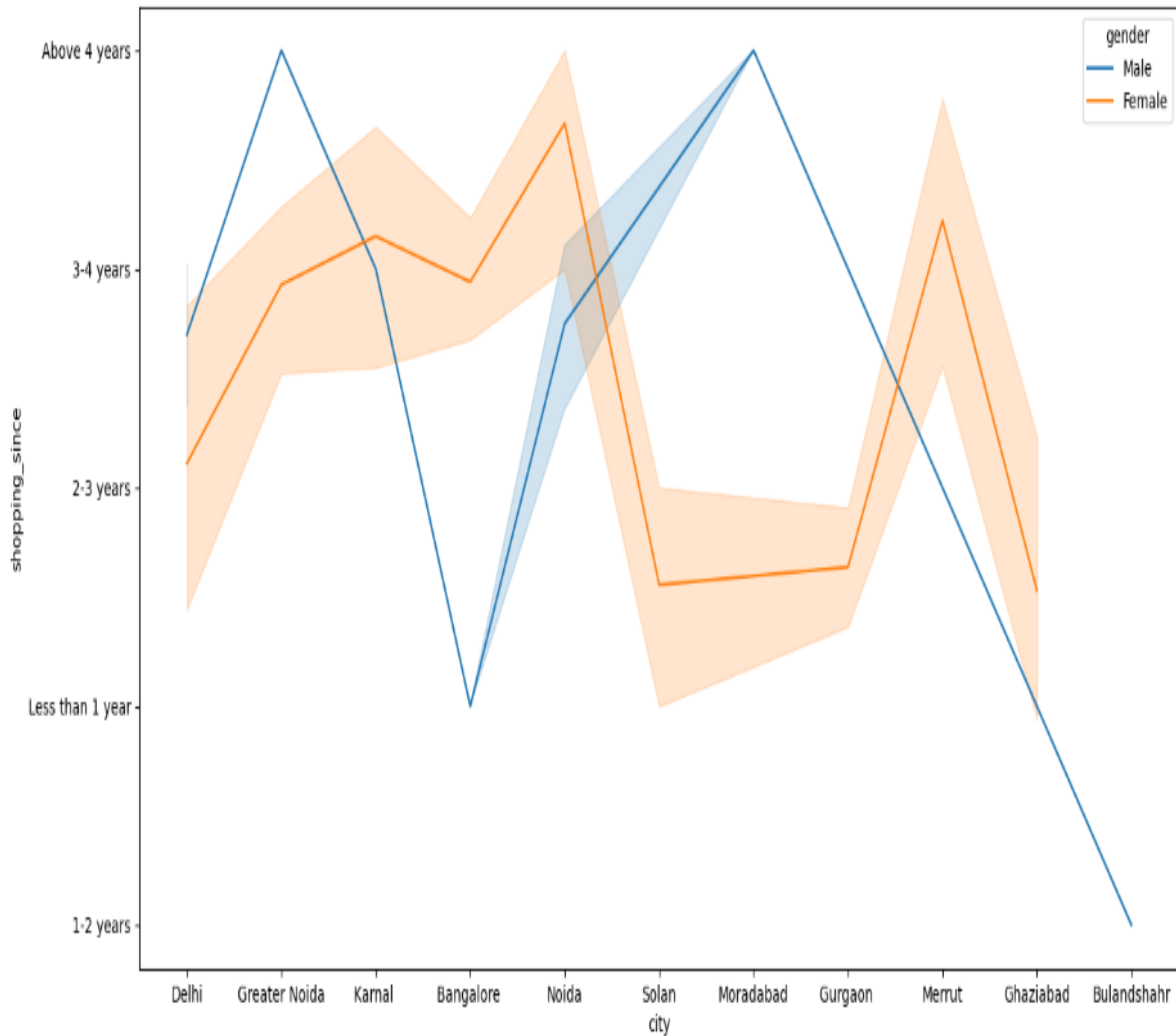
Highest number of people have been shopping online for above 4 years except for the age group below 20 years and above 50 years. People who are shopping online for 1-2 years does not include teenagers and elder people

###In lines, we can see that density of female customers is more than male. Men living in banglore and ghaziabad shop have shopped online for less than 1 year. Highest number of men shopping online belong from delhi and noida, while men from moradabad have been shopping online for the longest. Women from meerut and noida have shopped the longest.

```
[43]: plt.figure(figsize=(20,15))
      sns.stripplot(x=data['cancel_reason'],y=data['convinient_accessibility'])
```

[43]: <AxesSubplot: xlabel='cancel_reason', ylabel='convinient_accessibility'>
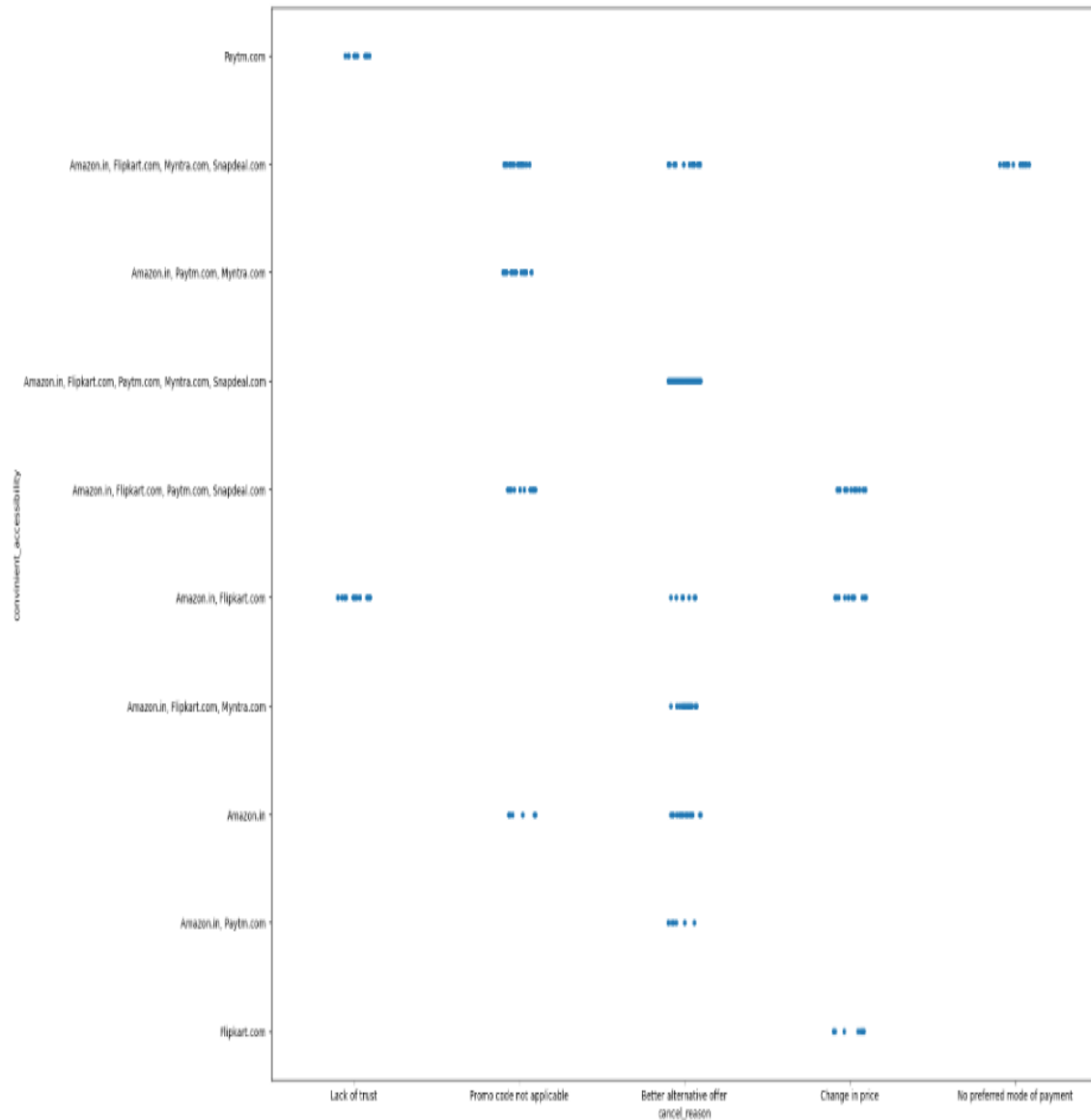
'''We can clearly see that most of the time people abandon the bag is beacuse they get a better alternative offer or promo code not applicable. There is also lack of trust seen in amazon, flipkart and paytm by some people.'''
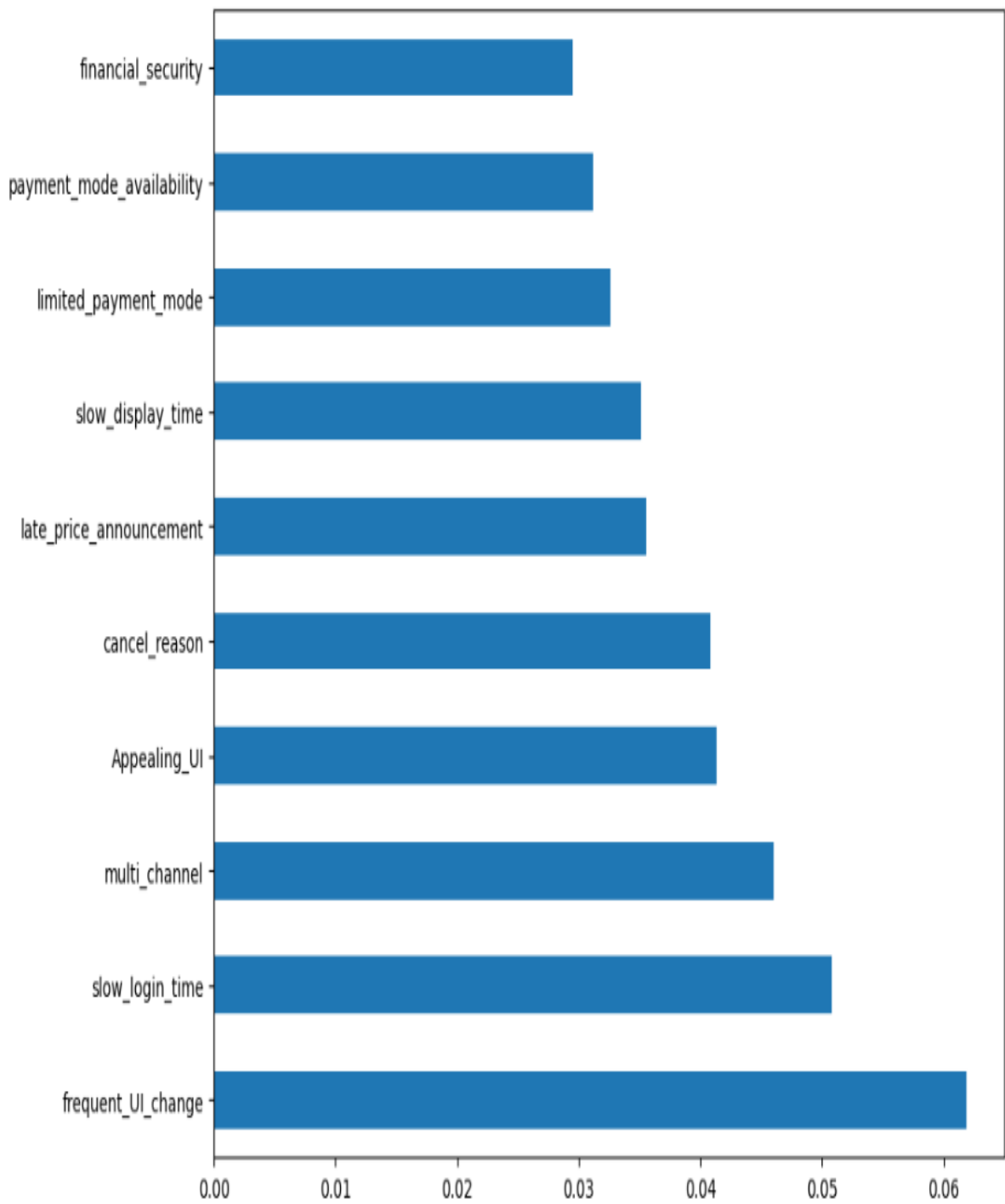
'''In the above chart we can see that above features are of most importance in determining whhich platform will a ciustomer recommend to his friend'''

```
plt.ylim(0.0,1.1)
plt.plot(xi, yi, marker='o', linestyle='--', color='b')

plt.xlabel('Number of Components')
plt.xticks(np.arange(0, 72, step=1)) #change from 0-based array index to 1-based human-readable label
plt.ylabel('Cumulative variance (%)')
plt.title('The number of components needed to explain variance')

plt.axhline(y=1, color='r', linestyle='-')
plt.text(0.5, 0.85, '100% cut-off threshold', color = 'red', fontsize=16)

ax.grid(axis='x')
plt.show()
```
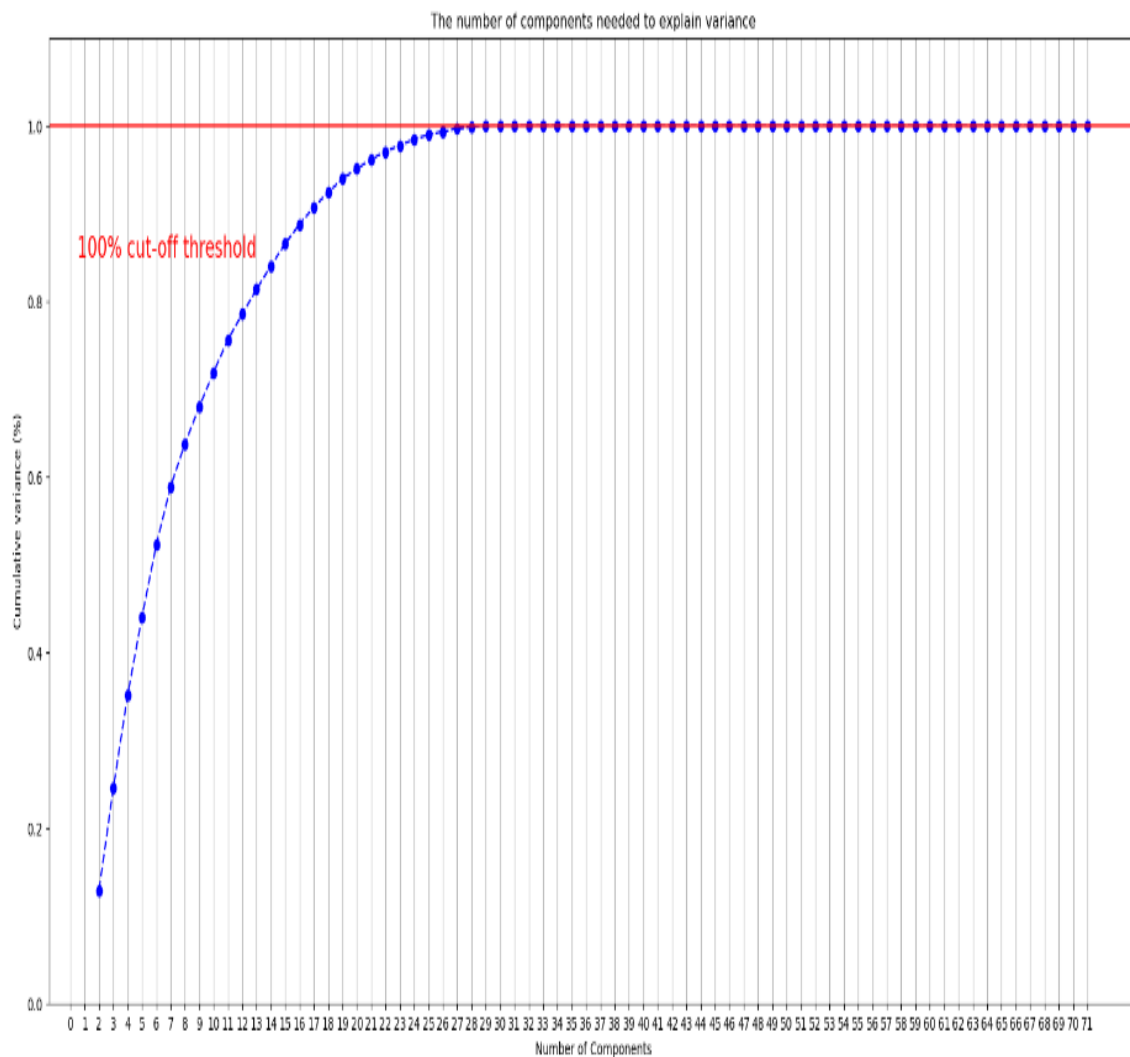

"'We can clearly see that with 29 features all the information can be retained'"

```
In [68]: print('Accuracy',np.round(accuracy_score(p,ytest),4))
         print('---------------------------------------------------------')
         print('Mean of Cross Validation Score',np.round(s.mean(),4))
         print('---------------------------------------------------------')
         print('Confusion Matrix')
         print(confusion_matrix(p,ytest))
         print('---------------------------------------------------------')
         print('Classification Report')
         print(classification_report(p,ytest))
```

```
Accuracy 1.0
---------------------------------------------------------
Mean of Cross Validation Score 1.0
---------------------------------------------------------
Confusion Matrix
[[26  0  0  0  0  0  0  0]
 [ 0 22  0  0  0  0  0  0]
 [ 0  0  4  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  7  0  0]
 [ 0  0  0  0  0  0 11  0]
 [ 0  0  0  0  0  0  0  2]]
---------------------------------------------------------
Classification Report
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        26
           1       1.00      1.00      1.00        22
           2       1.00      1.00      1.00         4
           3       1.00      1.00      1.00         4
           4       1.00      1.00      1.00         5
           5       1.00      1.00      1.00         7
           6       1.00      1.00      1.00        11
           7       1.00      1.00      1.00         2

    accuracy                           1.00        81
   macro avg       1.00      1.00      1.00        81
weighted avg       1.00      1.00      1.00        81
```

Conclusion:

The results of this study suggest following outputs which might be useful for E-commerce websites to extend their business. The cost of the product, the reliability of the E-commerce company and the return policies all play an equally important role in deciding the buying behaviour of online customers. The cost is an important factor as it was the basic criteria used by online retailers to attract customers. The reliability of the E-commerce company is also important, as it is even required in offline retail. It is important because customers are paying online, so they need to be sure of security of the online transaction. The return policies are important because in online retail customer does not get to feel the product. Thus, he wants to be sure that it will be possible to return the product if he does not like it in real. Whereas, the logistics factor, which included Cash on delivery option, one day delivery and the quality of packaging plays a secondary role in this process though these are Must-be-quality. This is so because these all does not interfere with the real product and people believe that this is the basic value that E-commerce websites provide. All the websites were not equally preferred by online customers. Amazon was the most preferred followed by Flipkart. This can be explained easily by previous result that we got. These two companies are most trusted in the industry and hence, have a huge reliability. Also, the sellers listed on these websites are generally from Tier 1 cities as compared to Snapdeal and Pay TM which have more sellers from tier 2 and 3 cities. Also, these websites have the most lenient return policies as compared to others and also the time required to process a return is low for these.