# CONSUMER RETENTION PROJECT

CONTEXT

1. Problem Statement

2. Understanding(of problem statement)

3. EDA steps and visualizations

4. Steps and Assumptions

5. Conclusions

1. 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, 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.
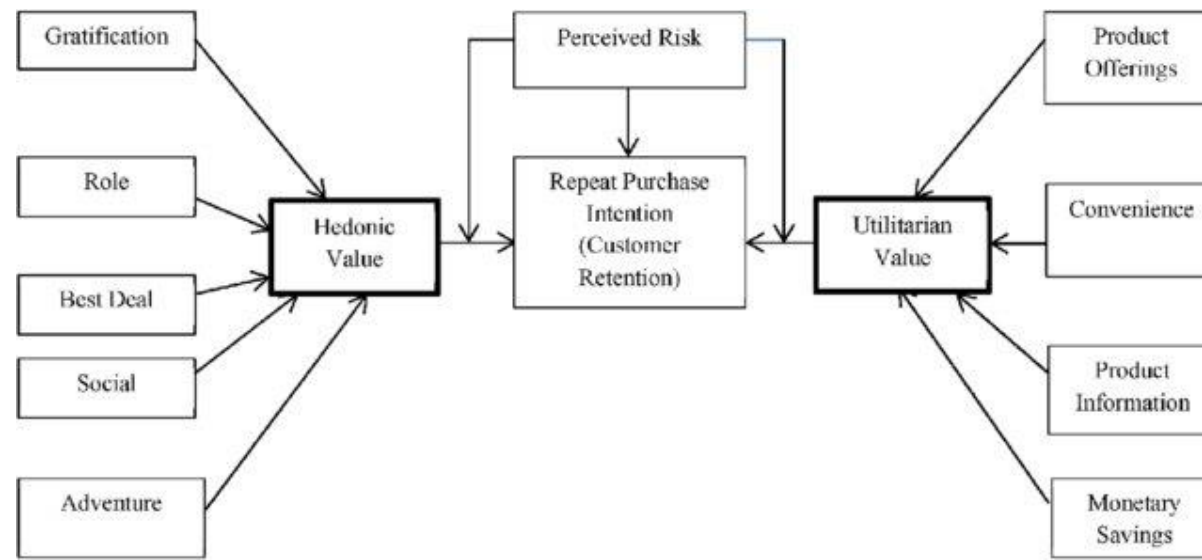
2. UNDERSTANDING(of problem statement):

Customer Retention used to control when your organization's data can be deleted and improve customer loyalty, policy. In simple terms, retention means to use or control of something or to take possession. The customers come back again and again for more and that makes customer retention increases your customers lifetime value. It also helps us build amazing relationships with our customers. They trust us with their money because we give them value in exchange. That makes customer satisfaction with the value and they trust our terms and conditions or policy.

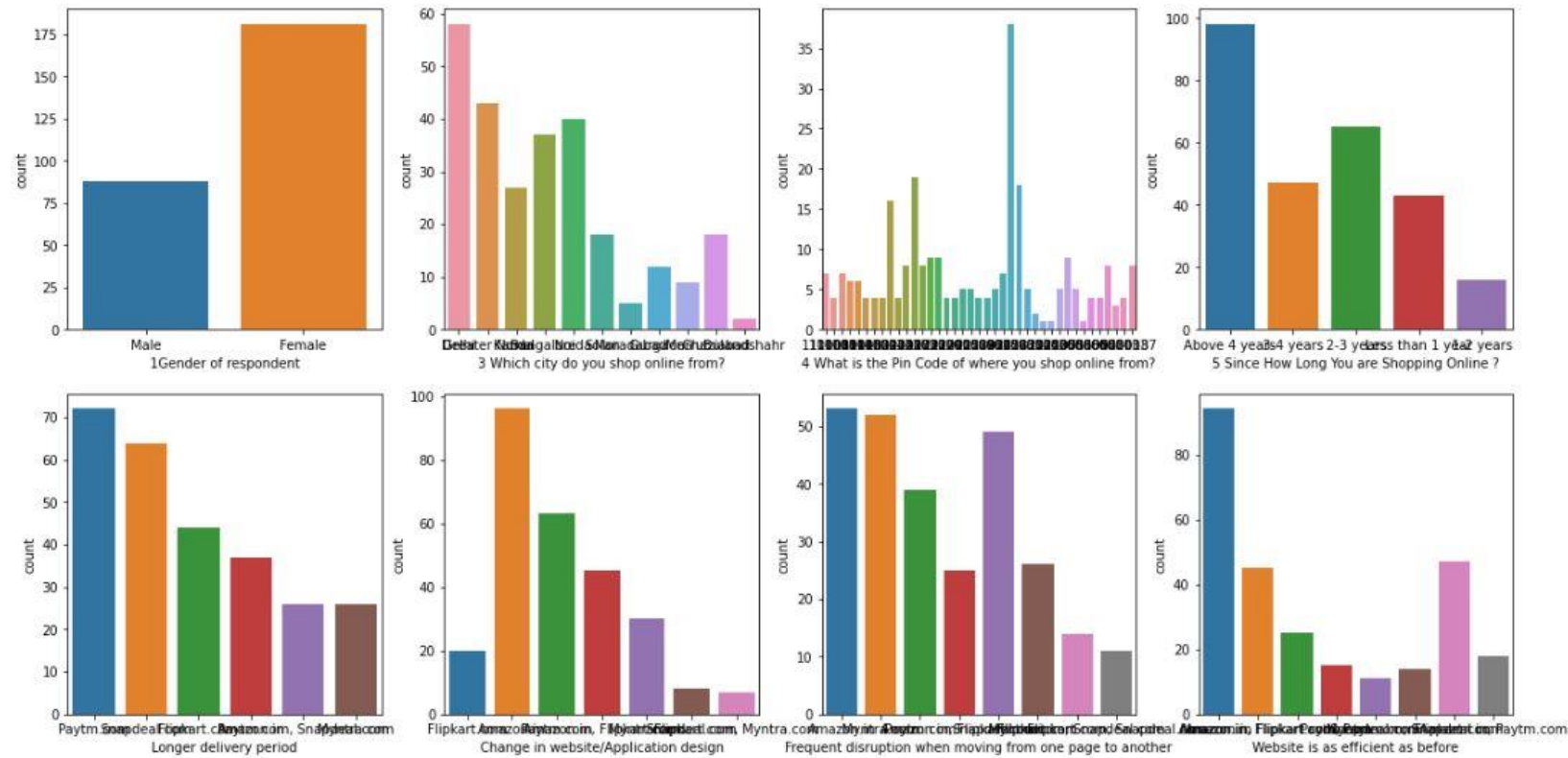# 3.EDA steps and visualizations

A.

```
1  def correlation_heatmap(df):
2      ax=plt.subplots(figsize=(10,5))
3      colormap=sns.diverging_palette(220, 10, as_cmap = True)
4      ax=sns.heatmap(df.corr(),cmap="spring",annot=True,linewidths=0.1,vmax=1.0,linecolor='white',annot_kws={'fontsize':12})
5      plt.title('Correlation',y=1.05,size=15)
6
7  correlation_heatmap(df)
```
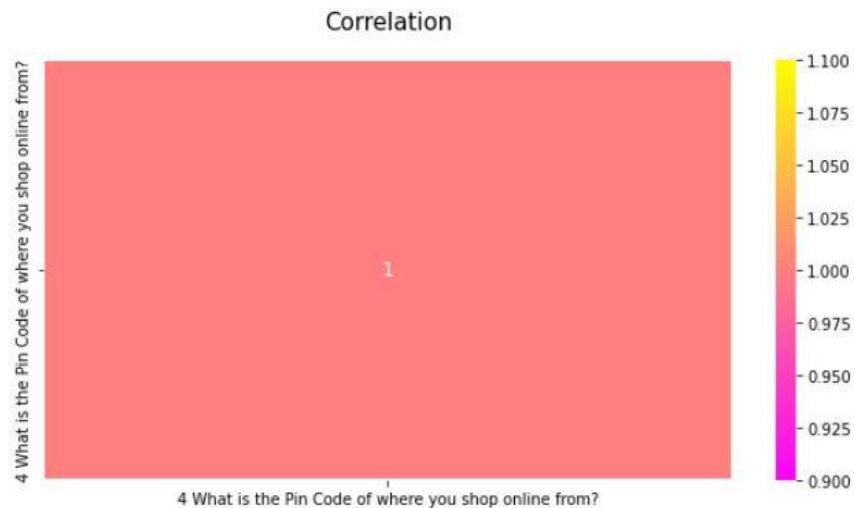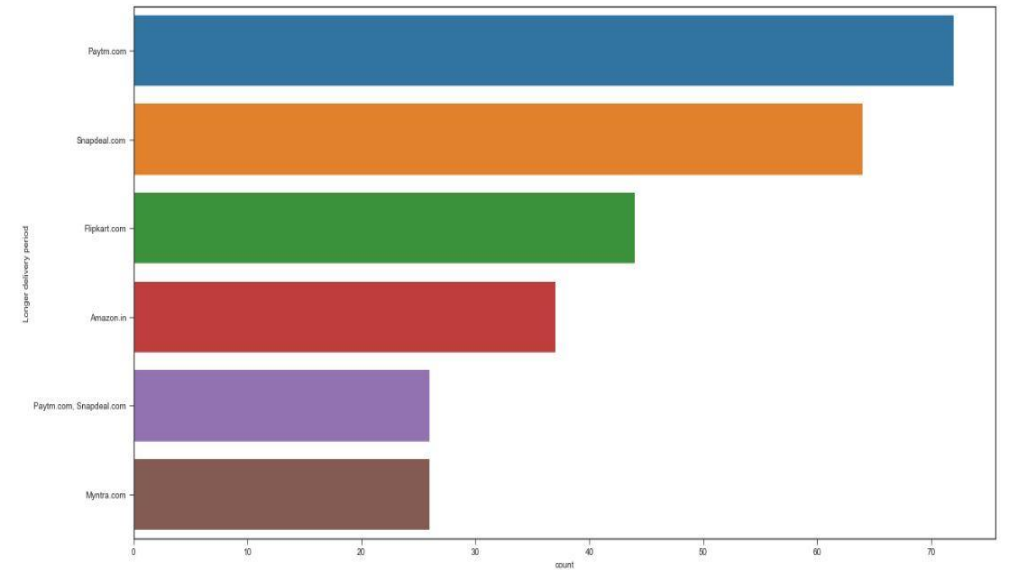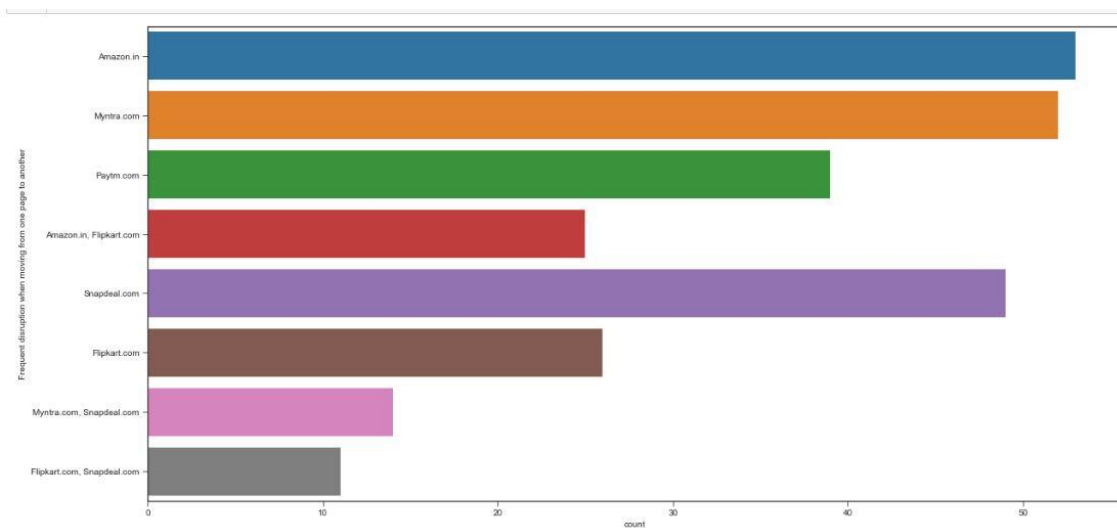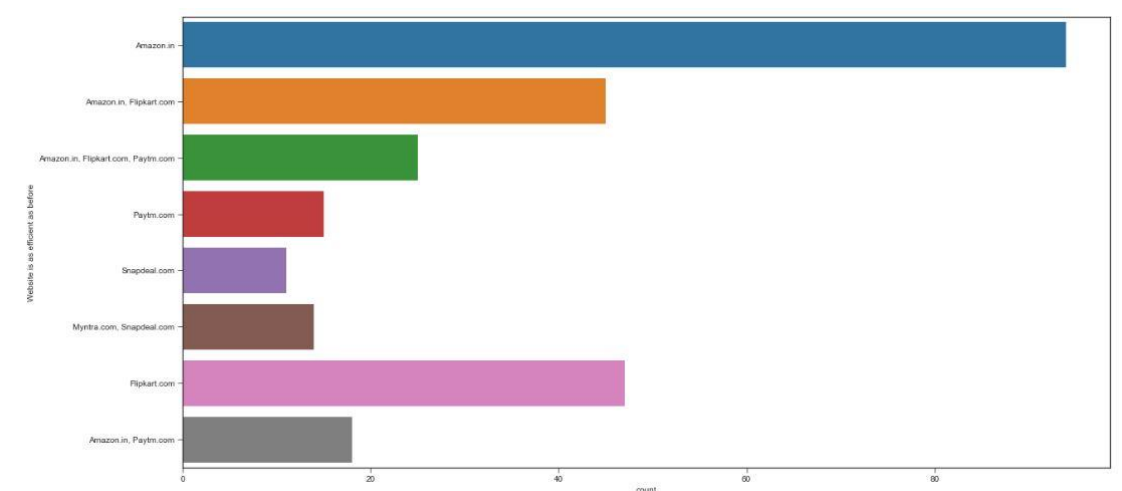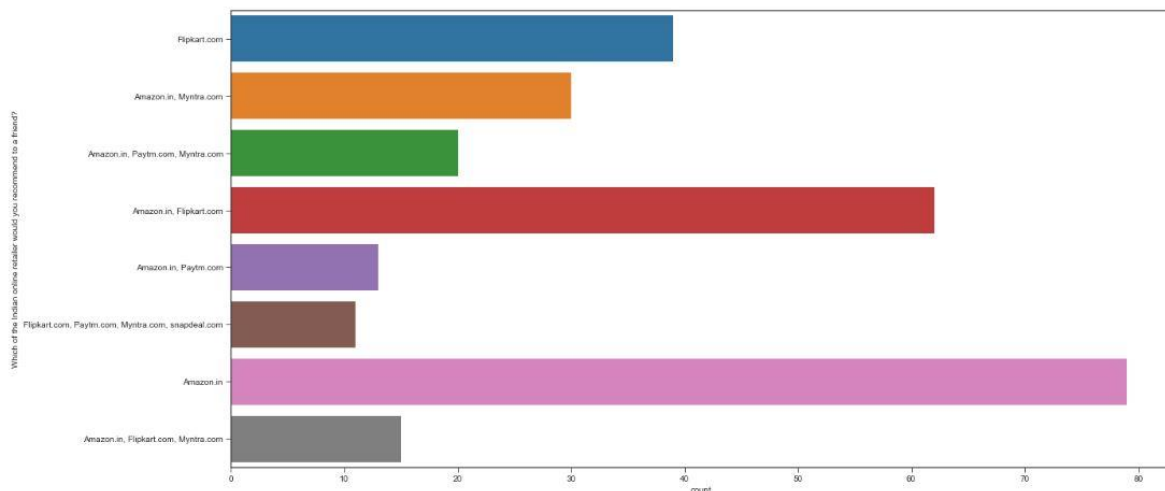

Firstly, I use correlation heatmap of dataframe to show interdependence of variable quantities. Here, in this correlation only one column is showing from the dataframe because of the statistical relationship between random variables.
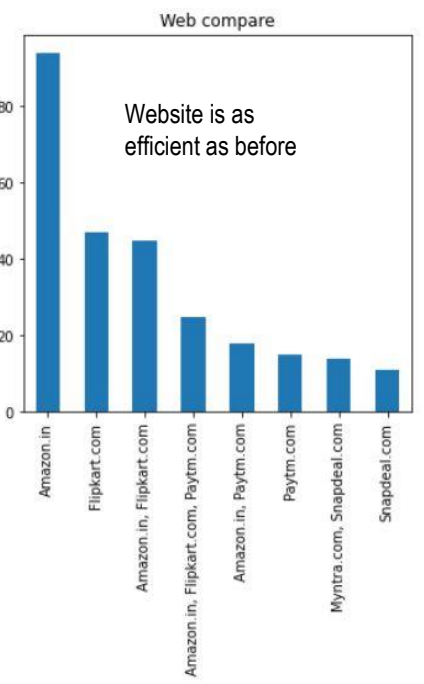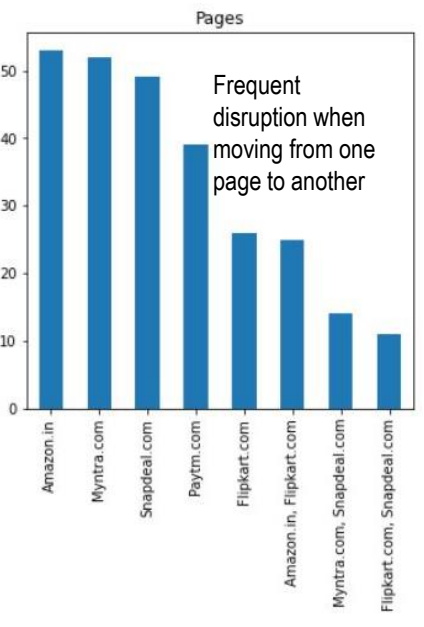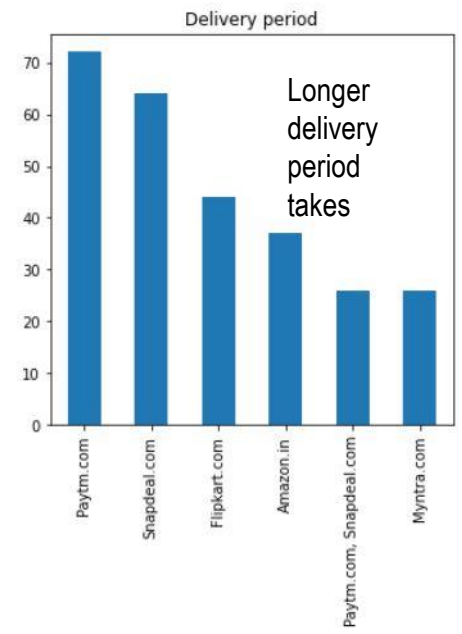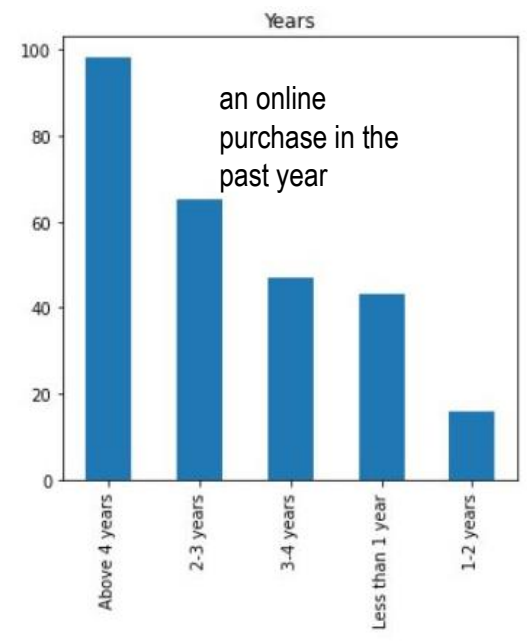
B.

In this visualizations, all the plot are in the countplot which are showing how are they higher at one point of time and also to comparison between those plots. Next slide is about them only which show differently.

C.

Why I use four countplot instead of all? Because all of them are having some information. Here, one plot is about the people or customer recommend online retailer to another customer that amazon.in is at higher accuracy as compared to others, second is about the website how efficient they are to satisfy the online customer that here also amazon balanced its position as efficient work in the field, third is about the frequent disruption when customer moving from one page to another that here also the same as before but myntra is in the competitions with amazon.in and last one is longer 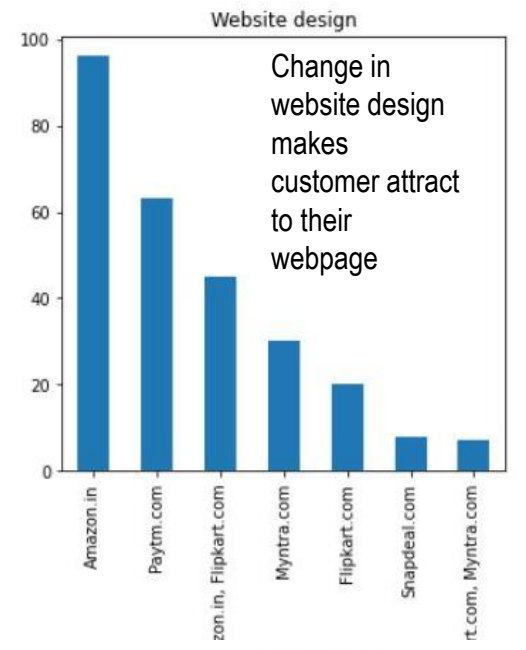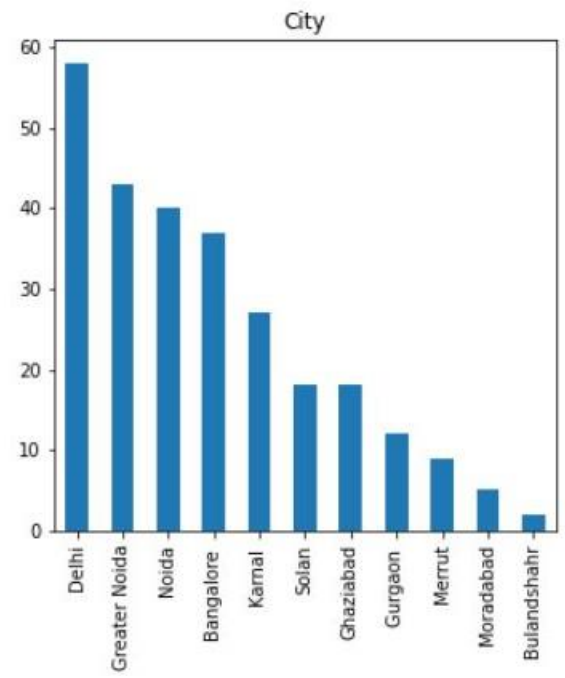delivery period that amazon makes no difference here it takes longer delivery period but snapdeal also into the range to compete with amazon in delivery. Each showing some correlation between them. That's why I choose them to show what all customer wants or thinks.

D.

Scatter_matrix

Gender

The no. of female customer = higher as compared to male customer

City

Website design

Change in website design makes customer attract to their webpage

Pin Code

Higher accuracy is where the people do online shopping

Years

an online purchase in the past year

Delivery period

Longer delivery period takes

Pages

Frequent disruption when moving from one page to another

Web compare

Website is as efficient as before

4. STEPS AND ASSUMPTIONS

- First, need to extract the excel file into CSV file .
  And then by importing libraries, find some datasets insights.
- By importing pandas there I have to take CSV file to dataframe.
- Datasets has 269 rows, 71columns.
- By info() method, found about columns and their data types.
- By handling missing values found that the datasets have no missing values ; why we do this because it can occur when no information is provided for one or more items or for a whole unit.
- Data visualization : the process of analyzed data in the form of graphs or maps making it a lot easier to understand the trends or patterns in the data. 3 data visualizations-
  1.Univariate analysis, 2.Bi-variate analysis, 3.Multi-variate analysis
- The bar plot can be used for univariate data visualization plot on a two-dimensional axis.
- The box plot can also be used for univariate and bivariate analyses.
- The joint plot can do both univariate as well as bivariate analyses.
- The histogram can be used for both univariate and bivariate analysis.
- The pair plot can be used for multiple pairwise bivariate distributions.
- The handling outliers deviates significantly from the rest of the (so-called normal)objects. They can be caused by measurement or execution errors. The analysis for outlier detection is referred to as outlier mining.
- Last , need to remove the outliers: it will detect the outliers using IQR and then we will remove them and also draw the boxplot to see if the outliers are removed or not, the old shape: 269 columns and 71 rows , the new shape: 232 columns and 71 columns.

CONCLUSIONS

Customer Retention is a critical activity in today's world or surroundings. It helps in providing control whenever the organization lose their control over the customer. Its objective really helps to overcome this problem by 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. By using various methods in EDA from the datasets by importing various libraries to analyse such visualization of datasets.

THANK YOU