**Project Report**

Introduction -:

This is the project report of the Customer retention project. In this project we have to analyze the data set and perform EDA , Visualization techniques, Data Cleaning etc.

The dataset is of the e-commerce website which contains the 269 rows and 71 columns in which lots of information was there like the gender , device used,location , ratings from the user and many more.

We have analyzed the complete dataset and there are some questions which I have noticed in the dataset.

**Q1.** Whether the dataset is organized or not?

**Ans.** Yes, the dataset is organized but there is some problem in the column names which have been resolved with the Data Cleaning technique.

**Q2.** Is there any Null Values Present in the columns?

**Ans.** No, there were no null values present in the dataset.

**Q3.** How many Categorical and Continuous columns present in the dataset?

**Ans.** All the columns present in the dataset are categorical columns.

**Q4.** Whether the Outliers present in the dataset?

**Ans.** There were some outliers present in the dataset but these were the possible outcomes.

**Q5.** IsCorelation present between columns?

**Ans.** Yes, correlation is present between the columns and we have checked it with corr() function and with the help of heat map.

**Q6.** Is Multicollinearity present between the columns?

**Ans.** Yes, lot’s of multicollinearity present between the columns ,infacti there is perfect correlation between the columns i.e. ‘inf’.

**Data Section:-**

The data is organized and no null values present in the columns while we check the dataset. There are some outliers present in the columns but we did not remove them because those were th possible outcomes. While we check the correlation we found lots of correlation between the columns and after that we check for multicollinearity with the help of Variance inflation factor and we found that there is infinite correlation exists between the column.

**Methods Section:-**

We have analyzed the data dataset with the EDA techniques like Data Cleaning , Checking Null values , Checking the correlation , Visualization techniques . we have used count plot to visualize the data set so that we can know how many values lies in the particular category. We have also checked the milticollinearity with the help of the variance inflation factor.

**Analysis Section:-**

We have analyzed that the data set is organized and no null values present In the data set . There are some outliers present in the columns but we did not remove them because those were th possible outcomes. While we check the correlation we found lots of correlation between the columns and after that we check for multicollinearity with the help of Variance inflation factor and we found that there is infinite correlation exists between the column.

While visualizing the data set we have seen that majority of the customers are above 20 age and most of them are the females. The customers are generally from the NCR region i.e. Delhi , Noida, Gr. Noida, and the majority of the customers are shopping since 4 years. The max. Customers purchases 10 times in the last 1 year. The customers generally do the online shopping through the Smartphones by using mobile internet. Max customers are using Google chrome for the online purchasing that means they are reaching the retail store through the search engine but not through the application. The customers exploring the website more than 15 minutes and make their payments through the debit/credit cards. But most of the times the customers abandon the cart because they found better offers on the same product. Coming to the rating part ajority of the customers strogly agree on the content that means the content is good and the information on the similar products is also good. The navigation through the website is good and the website processing is good, the interface is user friendly and the customers found the convenient mode of payments. The customers trust the websites and the website provides the guarantee of the privacy of the customer. The website gives the benefits and discounts to the customer and they strongly agree on that. The Website provides the good replacement and return policy and the customers enjoy easy shopping because of that. Majority of the customers will shop from amazon, flipkart, myntra, paytm, snapdeal. Most of the customers found amazon and flipkart web page layout easy. Most of the customers found good discounts on amazon and flipkart. The best website and application according to the customers is amazon as it has the appealing layout and the processing is very fast. The payment is fast on amazon and the speedy order delivery is also available on the amazon. In the last most of the customers found the best website is Amazon and also suggest amazon to their friends.

**Conclusion**

After analyzing the dataset I have found that most of the customers found amazon as the best application/website due to various reasons i.e. amazon is easy to use, its interface is user friendly, fast in process, has the best product information, shows relevant information on the similar product, delivery is fast, better discounts than other websites, has more trust on amazon than any other, payments are secure, good return and replacement policies and many more . I also analyze that most of the users use the website rather than the application for that there can be two reasons i.e. either they don’t know about the application of the retail e-stores or they don’t know how to use the application. Most of the customers are of above 20 age and they use smartphones for the online purchase.