

**Customer Retention: Data Analysis**

**DOCUMENTATION**

Submitted by:

-Pooja Mishra

**ACKNOWLEDGMENT**

I would also like to thank Flip Robo Technologies and Datatrained team who has given me such wonderful opportunity to learn more about Machine Learning and their mentors Shubham Yadav, Sajid Chaudhary for their constant encouragement, valuable suggestions, and cooperation.

This acknowledgement will remain incomplete if I fail to express my deep sense of obligation to my family members and God for their consistent blessings and believe in me.

Thank You.

**INTRODUCTION**

* Business Problem Framing

E-commerce is a term for any type of business that involves the exchange of information across electronic networks. It involves all aspects of an organization’s interaction or commercial transaction with suppliers, clients, stakeholders, and customers, etc. The best advantage is its huge reach across the global market and being able to conduct business 24x7x365 days by selling various kinds of goods and service.

Retaining on loyal customers creates a regular income, the goal of every seller is to transform occasional customers into loyal a one. Thus, it has urged sellers to create various retention strategies to increase profitability and gain market share by retaining customers. With this project we are analysing all important features that are necessary for customer retention.

* Conceptual Background of the Domain Problem

Presently, the number of people having access to the internet platform has been increased to a greater extent and this has led e-commerce business to grow rapidly. Since the e-commerce platform has enabled the customers to directly interact with sellers, it has led to a reduction in the supply chain process and accordingly customers also get to buy at a lower rate. It provides various choices for customers to choose among different sellers according to their preference for a product and budget ascertained. The objectives of the study were to know about the customer retention strategies in e-commerce business. The study was descriptive in nature with convenience sampling method employed. A well-structured questionnaire was distributed to the respondents.

* Review of Literature

Business enterprises across the globe are experiencing a significant change in the economy and turning their attention towards e-commerce business. By increasing the retention of customers and users. The significant growth of E-commerce industry has made the competition more intense among marketers. Maintaining customer retention strategies has been recognized as essential factors for business survival and growth.

The study has developed a e- satisfaction model which ties together factors representing major contexts namely convenience, merchandising, site design, security, and serviceability. The study identified that special attention to convenience, site design, age, gender, financial security produce most positive outcome pertaining to satisfaction.

This project comprises of well- structured questionnaire with closed ended questions was used to collect data from the respondents.

* Motivation for the Problem Undertaken

The main objective of this project is to analyse the features that are responsible for analysing the customer retention by the case study done on Indian e-commerce customers.

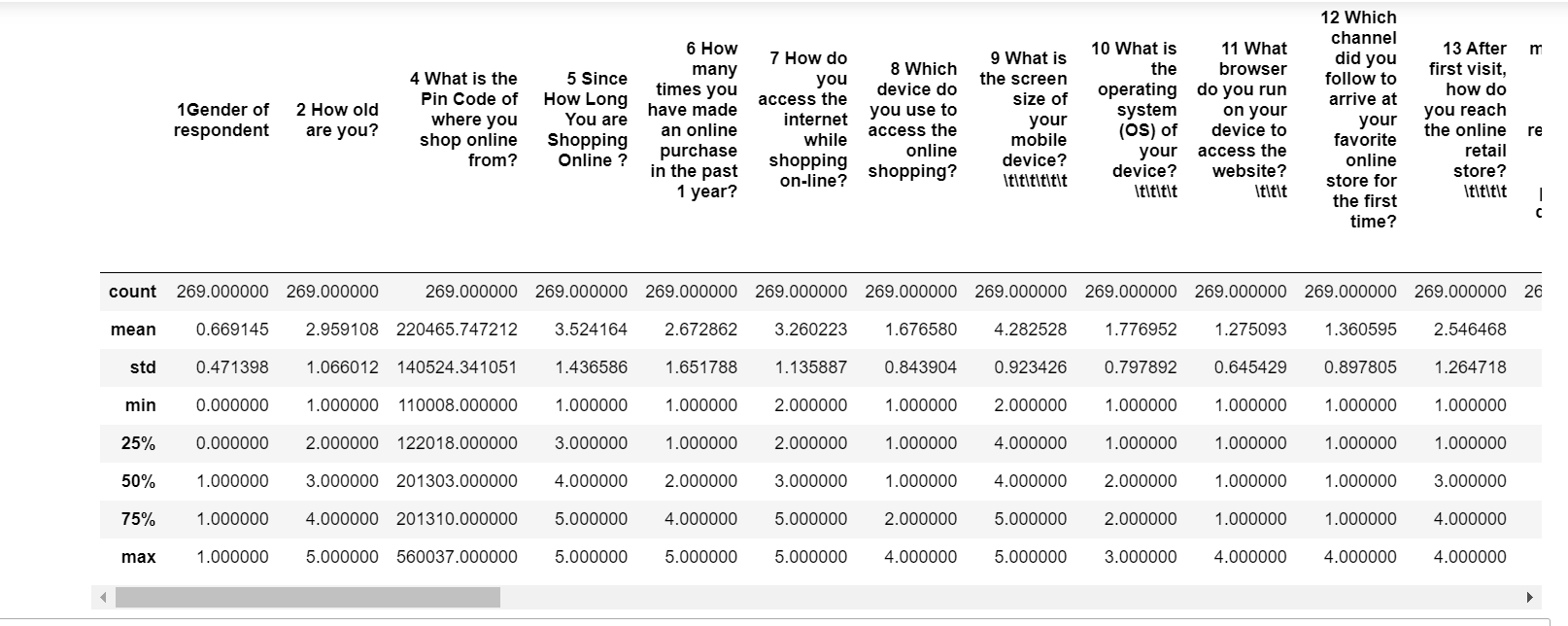
This project demonstrates the E-commerce industry overview the individual interest and several features that are related to online shopping. The research furthermore investigated the factors that influence the online customers repeat purchase intention. The combination of both utilitarian value and hedonistic 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.

**Analytical Problem Framing**

* Mathematical/ Analytical Modelling of the Problem

This project consists of two Datasets (one is detailed) and second is encoded. In this study encoded dataset has been taken and analysis done it to understand the nature and behaviour of Indian customers. The mathematical approach behind this project is to get the strategical overview of the dataset. The dataset has been categories into categorial and numerical features to perform the further analysis.

We can observe the count, mean, Standard deviation, min and max range, mean, median, first and third quantile values of each features present in the dataset which will help to get more information on dataset.



* Data Sources and their formats

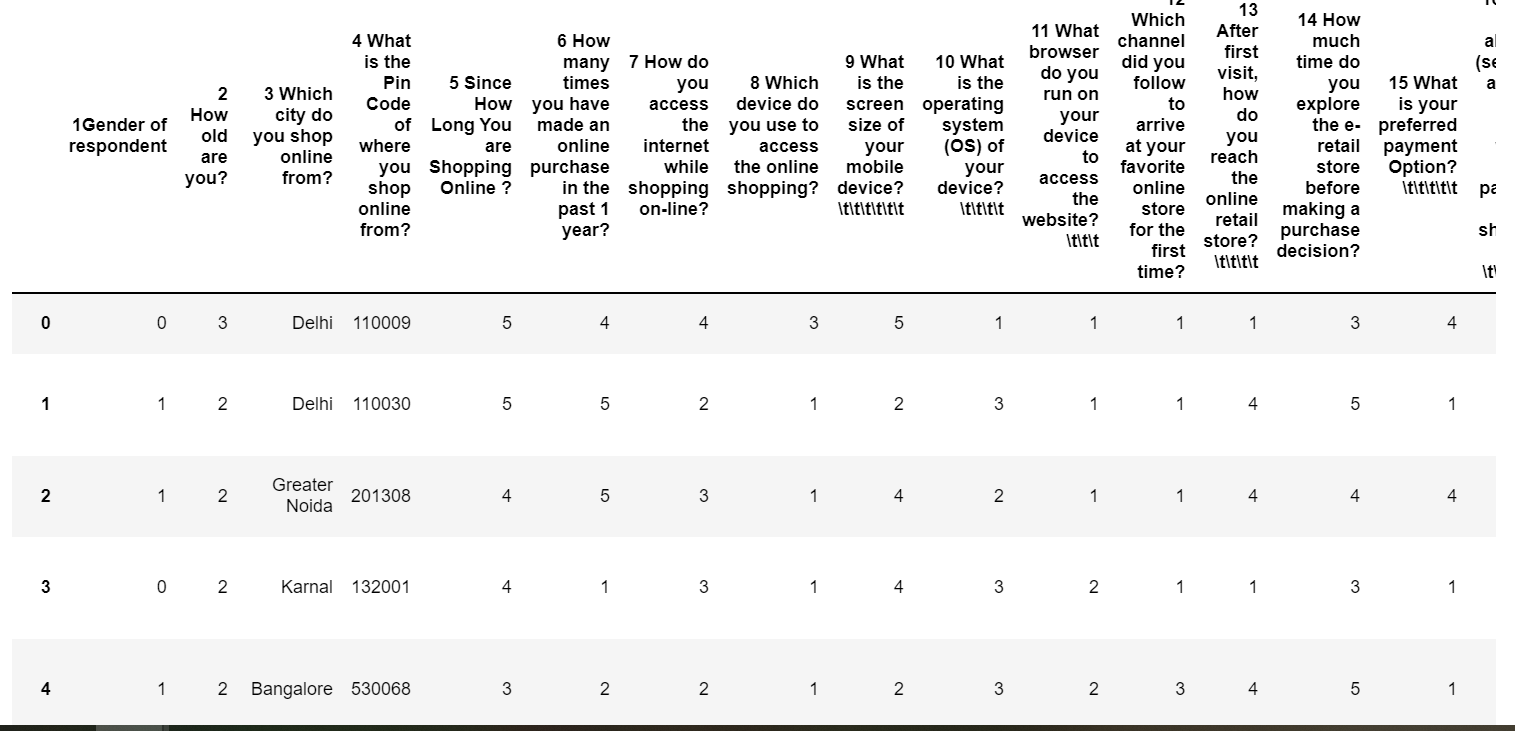
This dataset is collected from the Indian online shoppers. Results indicate the e-retail success factors, which are very much critical for customer satisfaction.

The dataset contains two sheets (one is detailed) and second is encoded. In this dataset the number of column(s) is more than 47.

Here, we will provide a brief description of dataset features. Since there are 71 features thus describing each could be tough so mentioning each datatype here to understand the dataset.

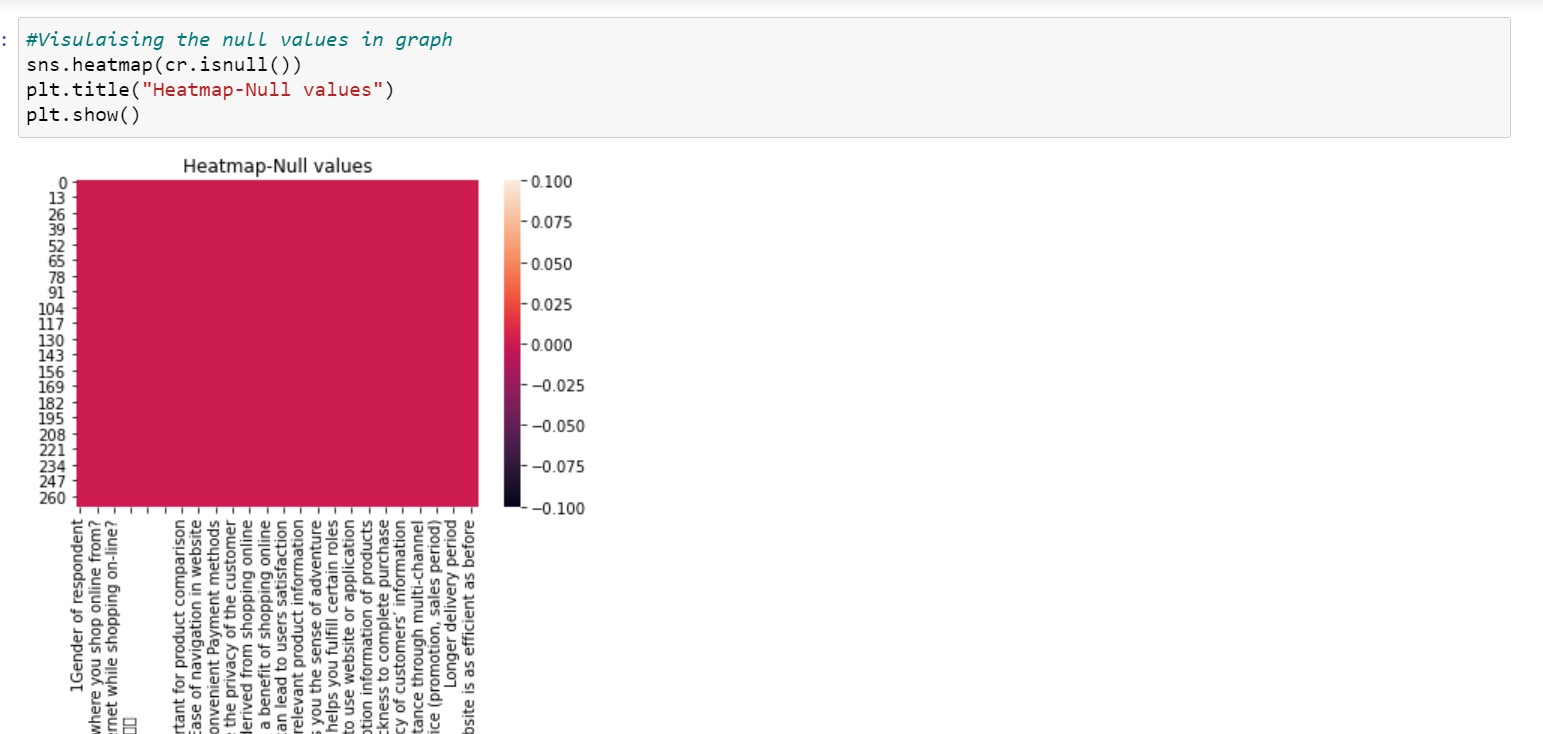
Detailed Dataset Description: This detailed dataset consists of 71 features describing every aspect of the Indian customers related to online shopping. In this data, the features are described in detailed format that are understandable at first look.

Encoded Dataset Description: This Encoded dataset consists of 71 features describing every aspect of the Indian customers related to online shopping. In this data, the features are described in coded format which need data analysis to understand it.

­

**Data Cleaning:**

There are no null values in this dataset.



* Data Analysis

In this project, to analyse the different features that are influencing the online customers retention. We’ll use that to identify high risk customers and inform retention strategies and marketing experiments.

A well- structured questionnaire with closed ended questions was used to collect data from the respondents.

Analysis on Numerical Features taken from coded dataset:

1 Gender: Here we have analysed that rate of females is much higher than males for online shopping. from the graph we have observed that

females count is much higher than males count.

2 Age Analysis: From the graph we analysed that online shopping is higher for age between 31-40 years and least with age 51 and above years.

3 Pin Code Analysis: From the graph we analysed that Delhi Pincodes are maximum for online shopping and Bulandshar pincodes are count is least

4 Since how long you're shopping: Its observed from this dataset that people are shopping since 2-3 years and above 4 years are maximum and least with 1-2 years.

5 Online Purchase in last 1 year: Online purchase in last 1 year is maximum with 'less than 10 times' and minimum with 21-30 times.

6 While shopping mode of internet Access: Maximum with Mobile Internet and dial-up is least to access internet while shopping.

7 Screen size of mobile device: With help of graph visualization, we analysed that maximum number of users who are using others category for their mobile screen size that are around count of 135. After that 5.5 inch mobile screen size used by customers for online shopping that are around count of 100. Least count is for 5 inch and 4.7 inch of mobile device users.

8 Operating System of device: We analysed that maximum users doing online shopping are using Window/Windows mobile that are around 123 and after the maximum users are Android OS users that are around count of 86. Least number of users whose mobile OS is IOS/Mac.

9 What browser do you run on device to access website: The maximum users who are using Google chrome browser to access the website that count is around 217. The least count is of Mozilla Firefox browser users.

The rest column data analysis can be collected from the data visualization.

Categorical features data analysis can be getting from jupyter notebox.

https://github.com/github-pooja/Internship/blob/main/Customer%20Retention%20Project.ipynb

The Libraries are as listed:

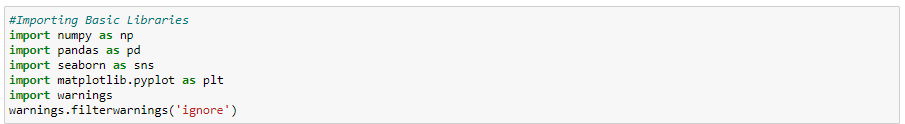
For Data loading and Visualisation:

Numpy - NumPy is very useful for handling linear algebra, Fourier transforms, and random numbers.

Pandas - Pandas are turning up to be the most popular Python library that is used for data analysis with support for fast, flexible, and expressive data structures designed to work on both “relational” or “labeled” data.

Matplotlib - The library helps to generate histograms, plots, error charts, scatter plots, bar charts with just a few lines of code.

Seaborn – Used for visualization.



Conclusion

The study has found that clear state of return policy, quality guarantee, monetary benefits and refund policy are important for the customers while making online purchase. These are useful findings which organized e-commerce websites need to keep in mind while conducting their business through websites. E-commerce websites must be convenient for customers to find what they want complexity in using a website may lead customers in switching over to another websites. The study also points out serviceability and web design factor while trading through online. E-commerce websites should ensure about Prompt delivery, complaint handling and good packaging. Because, this factor is considered as a significant factor in retaining customers. They should ensure about website designs, easy navigation, updated information and easy to recall avoiding discomfort.

.

Thank You.