NAME OF THE PROJECT

Customer Retention Project

Submitted by:

DOLYPONA DAS

ACKNOWLEDGMENT

I would like to thank our SME (Shubham Yadav) for his expert advice and encouragement throughout this project.

INTRODUCTION

• Business Problem Framing

- 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.
- The research furthermore investigated the factors that influence the online customers repeat purchase intention.
- The data is collected from the Indian online shoppers. Results indicate the e-retail success factors, which are very much critical for customer satisfaction.
- We are required to apply our analytical skills to give findings and conclusions in detailed data analysis. Only data analysis is required for the given data.

Analytical Problem Framing

• Data Preprocessing Done

- Importing the libraries needed for my project.
- Loading the dataset/Creating my dataframe .
- Checking the head of the dataset(first five rows).
- Checking the shape of the dataset, there are 71 columns where we have to work upon.
- Checking the datatypes
- Checking for null values if any
- I have renamed almost all the columns of the dataset and minimise the lengthy names of the columns.

- I have checked the value counts of all the columns to get some valuable information.
- In the dataset , almost all the columns are object type ,so I have used Label Encoder to encode those columns.
- For data visualization, I have used histogram .

• Hardware and Software Requirements and Tools Used

Hardware requirements:
Processor: Intel(R) Celeron(R) CPU N3050 @1.60 GHz 1.60 GHz
RAM: 3.92 GB
System type: 64-bit operating system, x64-based processor
Software requirements:
Python: One of the most used programming languages
Tools used:
Jupyter notebook: Jupyter is a free, open-source,interactive web tool known as a computational notebook where I have written my python codes.
NumPy: NumPy is an open-source numerical Python library. NumPy contains a multi-dimensional array and matrix data structures.
Pandas: Pandas is an open source Python package that is most widely used for data science/data analysis and machine learning tasks. It is built on top of another package named Numpy, which provides support for multi-dimensional arrays.
Matplotlib: It is a Python library used for plotting graphs with the help of other libraries like Numpy and Pandas. It is a powerful tool for visualizing data in Python.
Seaborn: It is also a Python library used for plotting graphs with the help of Matplotlib, Pandas, and Numpy.

Evaluation

- Visualizations
  - Visualization of our dataset is done using histogram and using those graphs we can see how data analysis is done.

- A histogram is the most commonly used graph to show frequency distributions. It looks very much like a bar chart, but there are important differences between them. This helpful data collection and analysis tool is considered one of the seven basic quality tools.

• Interpretation of the Results

- When I started analysing my data, I have observed that my data is all about Indian online shoppers and customer satisfaction.
- I have observed that most of the columns are object type, which needs to be encoded so that machine can process it .
- As columns are of object type, there is no need for outliers removal. We generally go for outliers removal for continuous type data.
- I also observed that among all the online shopping websites ,amazon.in is considered best among the customers in many aspects.