**E-Commerce Customer Churn**

**Cindy Jeon(student ID : 922779082)**

1. **Introduction**

This dataset focuses on customer churn within an ecommerce company. By analyzing this data, I intend to identify patterns in customer behavior that reveal whether they are likely to churn or remain engaged.

1. **Goals**

My goal is to identify specific customer behaviors that signal a likelihood of churn. By understanding these patterns, we can proactively address potential churn, implementing strategies to retain customers before they discontinue their engagement with the company.

Additionally, through analysis, I aim to uncover the customer behaviors strongly correlated with high sales. These insights will empower companies to optimize their strategies, fostering an environment that not only prevents churn but also strategically boosts overall sales.

1. **Discussion of how you converted the dataset into tables.**

I structured database by creating seven tables to capture different aspects of my data. The tables and the characteristics of the tables are as below.

This structure will allow for a comprehensive representation of the dataset, enabling me to analyze different dimensions of customer behavior and business performance. The relationships between these tables, established through keys, will be essential for querying and gaining insights across the various aspects of the data.

1) Customer Table

- Contains personal information such as gender, marriage status, tenure, churn status.

2) User Interaction Table

- Captures data related to how users interact with the app and the number of devices registered on the site.

3, 4) Satisfaction and Complain Table

- Stores information about customer satisfaction and any complaints they may have.

5) Location Table

- Includes city tier for customers' houses and the distance of the warehouse to customers' homes.

6). Sales Order Table

- Provides details on the number of orders each customer made and tracks sales increases compared to the previous year.

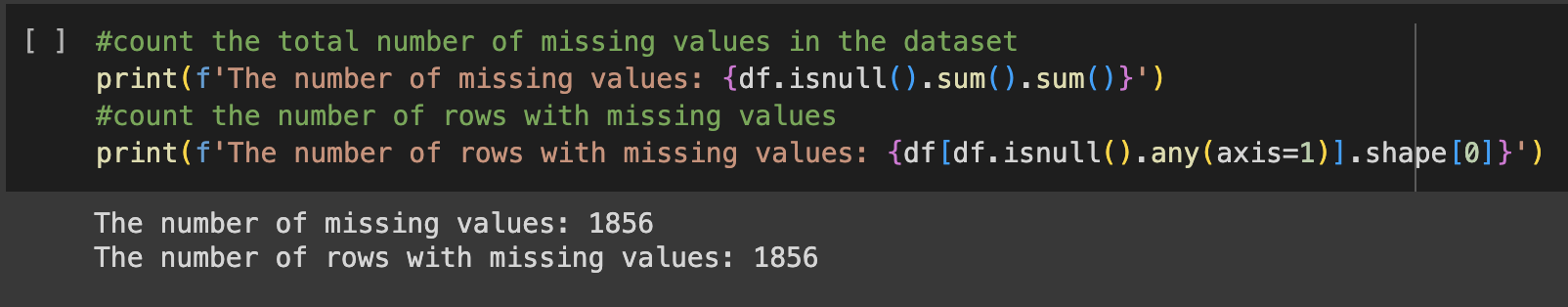
7). Payment Information Table

- Contains data related to payment methods preferred by customers, coupon usage, and cashback received with purchases.

1. **Challenges faced during importing of your data and how did you overcome these data importation challenges.**
2. Data type error when importing to SQL
   * First of all, my computer is mac, so I had to change xlxs. file to JSON type. However, when I converted the file type to JSON, the data were transferred to the empty cells, resulting in the data being corrupted. It turned out that I had many cells without data.
3. Dealing with missing values
   * I had to update those empty cells with ‘NULL’ values instead of removing them.
   * There were 1856 missing values and the rows with the number of missing values were exactly the same as 1856 which means that all the missing values are in the different rows.
   * Total number of raws of the data was 5630 but the rows with missing values were 33% which is one third of the data; therefore, I did not delete the rows with missing values but decided to keep them for the sake of reliability of the data.

. The number of missing values: 1856

. The number of rows with missing values: 1856

. The percentage of the missing values out of the total data : 33%

1. The number of limit rows in mySQL
   * Since this is my first time using SQL, I did not realize there was a limits in the rows, but I was able to change the limit from 1000 to 10000 since my data has 55630 rows.
2. Issues with selecting primary keys
   * Although the primary key is expected to be a unique identifier, I initially overlooked this requirement, resulting in the selecting of columns that lacked this uniqueness attribute. Consequently, I encountered difficulties storing data in the tables I had created. To fix this issue, I modified the primary keys to the unique identifier criterion, specifically using 'CustomerID' in certain tables. This adjustment successfully resolved the problem. Additionally, I learned that I could create new columns as primary keys and that I can assign numerical values to it. This experience was very intriguing yet vital lesson in managing SQL tables.
3. **A complete data dictionary for every table in your database.** 
   * 1. **Customer table**

|  |  |  |  |
| --- | --- | --- | --- |
| **Key** | **Variable** | **Data type** | **Discerption** |
| PK | CustomerID | INT | Unique customer ID |
|  | Churn | INT | Churn Flag |
|  | Tenure | INT | Tenure of customer in organization |
|  | CityTier | INT | City tier |
|  | Gender | VARCHAR(30) | Gender of customer |
|  | MaritalStatus | VARCHAR(30) | Marital status of customer |
|  | NumberOfAddress | INT | Total number of added added on particular customer |

* + 1. User\_interaction table

|  |  |  |  |
| --- | --- | --- | --- |
| **Key** | **Variable** | **Data type** | **Discerption** |
| PK,FK1 | CustomerID | INT | Unique customer ID |
|  | PreferredLoginDevice | VARCHAR(30) | Preferred login device of customer |
|  | HourSpendOnApp | INT | Number of hours spend on mobile application or website |
|  | NumberOfDeviceRegistered | INT | Total number of deceives is registered on particular customer |
|  | DaySinceLastOrder | INT | Day Since last order by customer |

* + 1. Satisfaction

|  |  |  |  |
| --- | --- | --- | --- |
| **Key** | **Variable** | **Data type** | **Discerption** |
| PK | CustomerID | INT | Unique customer ID |
|  | SatisfactionScore | INT | Satisfactory score of customer on service |

* + 1. Complain

|  |  |  |  |
| --- | --- | --- | --- |
| **Key** | **Variable** | **Data type** | **Discerption** |
| PK | CustomerID | INT | Unique customer ID |
|  | Complain | INT | Any complaint has been raised in last month |

* + 1. Location

|  |  |  |  |
| --- | --- | --- | --- |
| **Key** | **Variable** | **Data type** | **Discerption** |
| Pk | CustomerID | INT | Unique customer ID |
|  | CityTier | INT | City tier |
|  | WarehouseToHome | INT | Distance in between warehouse to home of customer |

* + 1. Sales\_Order

|  |  |  |  |
| --- | --- | --- | --- |
| **Key** | **Variable** | **Data type** | **Discerption** |
| PK | Order ID | INT | Unique order ID |
| FK1 | CustomerID | INT | Unique customer ID |
|  | PreferedOrderCat | VARCHAR(30) | Preferred order category of customer in last month |
|  | OrderAmountHikeFromlastYear | INT | Percentage increases in order from last year |
|  | OrderCount | INT | Total number of orders has been places in last month |

* + 1. Payment\_information

|  |  |  |  |
| --- | --- | --- | --- |
| **Key** | **Variable** | **Data type** | **Discerption** |
| PK | PaymentID | INT | Unique payment ID |
| FK1 | CustomerID | INT | Unique customer ID |
|  | PreferredPaymentMode | VARCHAR(30) | Preferred payment method of customer |
|  | CouponUsed | INT | Total number of coupon has been used in last month |
|  | CashbackAmount | INT | Average cashback in last month |

1. **The list of business questions.**

There are 8 business questions that I would like to figure out through this dataset.

In general, I would like to find out what factors influence on the churn.

I plan to conduct a thorough analysis to explore the interconnections between

satisfaction, complaints, location, marital status, and churn in the dataset. Additionally, I

will examine how these factors relate to order status, incorporating other relevant

variables. The goal is to gain insights into the complex relationships within these

elements and their potential impact on overall outcomes.

Based on each table category, I would ask 8 business questions below.

|  |  |
| --- | --- |
| **No.** | **Questions** |
| **1** | The relationship between **Churn and complain / satisfaction level** |
| **2** | The relationship between **tenure & Churn status** |
| **3** | The relationship between **tenure & customer satisfaction** |
| **4** | The relationship between **preferred order category and Churn** |
| **5** | The relationship between **Marital status and churn** |
| **6** | The relationship between **Marital status&product category and churn** |
| **7** | The relationship of **Churn and DaySinceLastOrder** |
| **8** | The relationship of **Churn and warehouse to house?** |