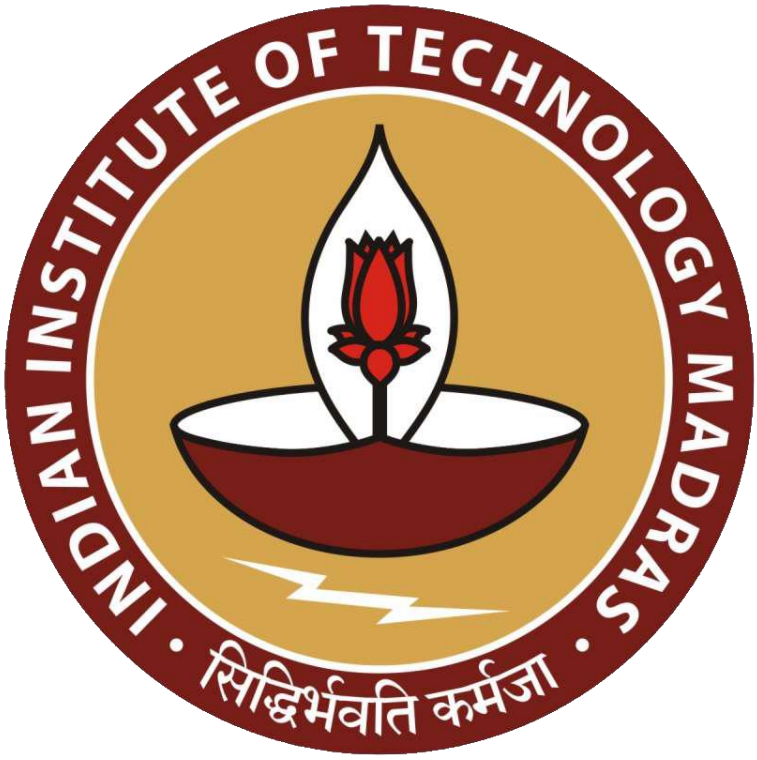
Enhancing Gym Business Sustainability: Overcoming Challenges and Boosting Customer Engagement

A Mid-Term Report for the BDM Capstone Project

Submitted by

Name: Harishankar Nagar

Roll Number: 21f2000681



IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

1 Executive Summary

In this capstone project, I have engaged with Mr. Amit Chaudhary, a multi-business owner, who is facing challenges in his gym business. Initially, his gym was successful in serving both college students and local residents, but it has since struggled due to increased competition from a new college gym. Additionally, Amit's lack of prior experience in the gym industry has hindered his ability to manage and grow the business effectively. A significant concern is that the gym heavily relies on college students as its primary customer base, making it vulnerable to fluctuations in customer numbers.

To address these issues, a data-driven approach was adopted, involving the collection and meticulous cleaning and processing of three key datasets: customer information, package renewal data, and expenditure data. This data analysis unearthed valuable insights and essential metrics, including the count of new members, monthly revenue generation, total revenue, and customer demographics. Notably, the analysis identified a substantial attrition rate in later months.

To delve deeper into the high attrition rate, a customer feedback survey is planned. This survey will enable me to understand the specific reasons behind customer churn, paving the way for more targeted solutions in the future.

2 Proof of Originality of the Data

2.1 About the Organization

Mr. Amit Chaudhary is a multi-business owner who has demonstrated an entrepreneurial spirit and a willingness to diversify his business ventures. Mr. Amit Chaudhary's first business venture was a general store, which catered to the daily needs of the local population, including people from nearby villages and college students. After that he open a restaurant “Food House and Fast-Food Restaurant”, which was successful idea. Recognizing an opportunity in the market, Mr. Chaudhary ventured into the fitness industry by opening a gym (Muscle Factory Gym) in the same area. This gym primarily targeted college students and residents who were looking for fitness and exercise services. The gym initially experienced a surge in popularity, but it faced challenges later. All his businesses are under the same roof and located near Central University of Rajasthan, Bandarsindari, pin code- 305817. Location

2.2 Images and Video of the Firm

Video Link:

- Video1
- Video2

2.3 Letter From Organization

Letter from organization

3 Metadata and Descriptive Statistics

3.1 Metadata

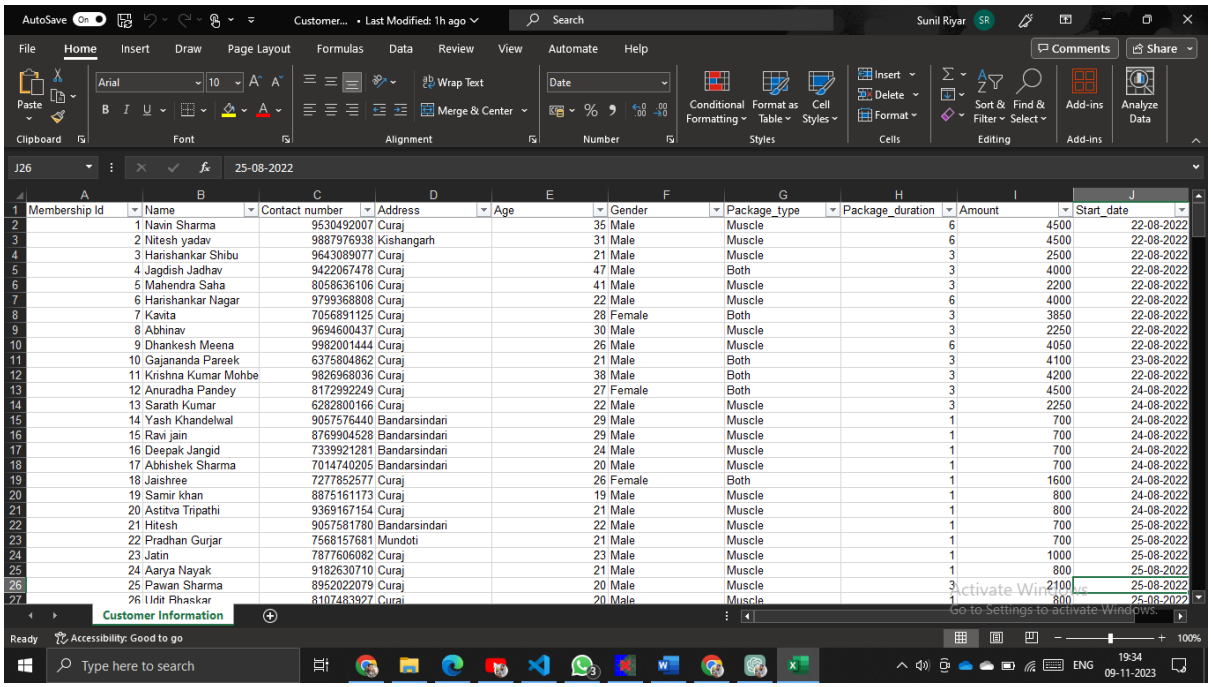
The metadata consists of three sheets. The description of the sheets are as follows.

1. The customer information sheet contains all the information regarding the customers. It has membership id, mobile number, package type, package duration, amount paid, date of start, and some demographics variables like gender, age.
    a. Membership_Id: This variable contains the unique identifier for all the customers, it is used to keep track of customer transactions.
    b. Mobile_number: I also collected customer's mobile number. It will be used to communicate feedback form.
    c. Package_type: The gym offers two types of packages the first one is cardio and the second one is muscle.
    d. Package_duration: The gym offered four plans in terms of duration i.e., one month, three months, six months, and one year.
    e. Amount: This variable contains the information about the amount customer paid at the time of registration. This variable will be helpful to perform revenue analysis.
    f. Start_date: This variable holds the information about the date on which consumer registered himself/herself. This variable will be helpful to assess the new customer count for each month.
    g. Gender and Age is self-explanatory.
2. Package Renewal dataset contains the information about the customers who has renewed their subscription. This dataset has the following columns-Membership_Id, Name, Pck_duration, and Amt_paid for each month.
3. Expenses dataset contains the information about fixed costs and variable costs.


Customer data.xlsx


Package Renewal.xlsx


Expenditure.xlsx

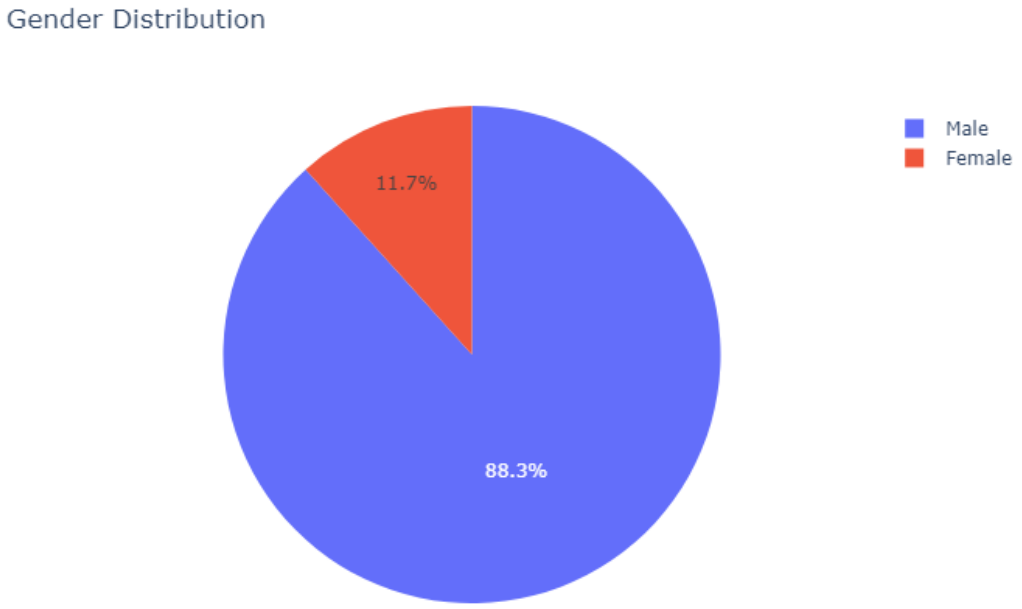


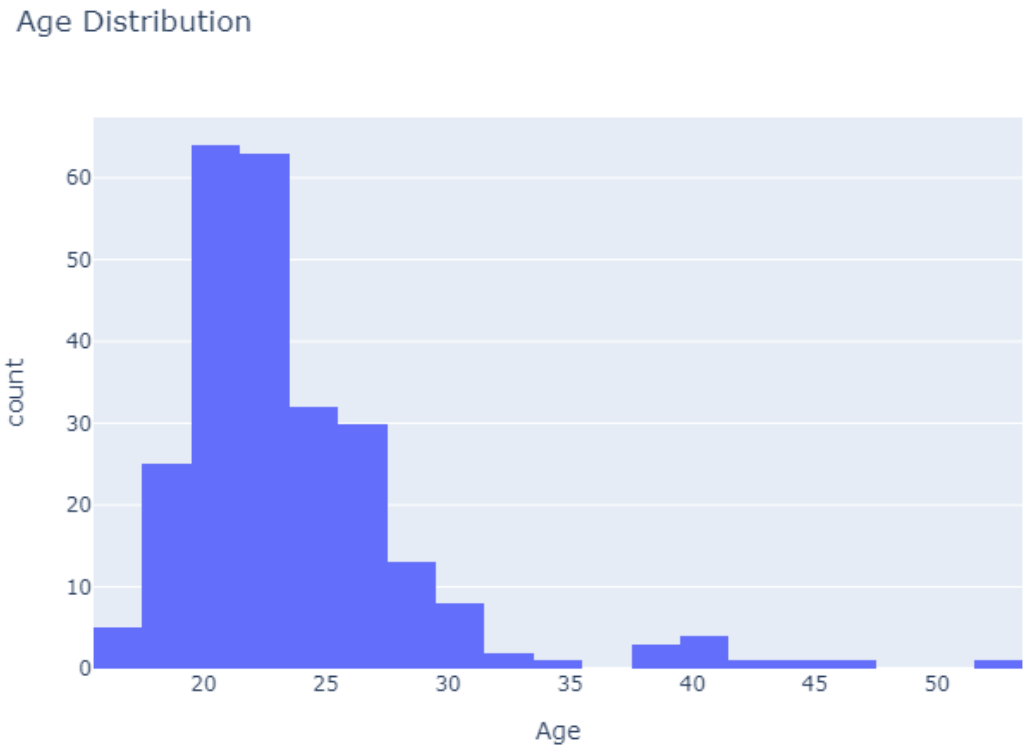
Membership Id	Name	Contact number	Address	Age	Gender	Package_type	Package_duration	Amount	Start_date
1	Navin Sharma	9530492007	Curaj	35	Male	Muscle	6	4500	22-08-2022
2	Nitesh yadav	9887976938	Kishangarh	31	Male	Muscle	6	4500	22-08-2022
3	Harishankar Shibu	9643089077	Curaj	21	Male	Muscle	3	2500	22-08-2022
4	Jagdish Jadhav	9422067478	Curaj	47	Male	Both	3	4000	22-08-2022
5	Mahendra Saha	8058636106	Curaj	41	Male	Muscle	3	2200	22-08-2022
6	Harishankar Nagar	9799368808	Curaj	22	Male	Muscle	6	4000	22-08-2022
7	Kavita	7056891125	Curaj	28	Female	Both	3	3850	22-08-2022
8	Abhinav	9694600437	Curaj	30	Male	Muscle	3	2250	22-08-2022
9	Dhankesh Meena	9982001444	Curaj	26	Male	Muscle	6	4050	22-08-2022
10	Gajananda Pareek	6375804862	Curaj	21	Male	Both	3	4100	23-08-2022
11	Krishna Kumar Mohbe	9826968036	Curaj	38	Male	Both	3	4200	22-08-2022
12	Anuradha Pandey	8172992249	Curaj	27	Female	Both	3	4500	24-08-2022
13	Sarath Kumar	6282800166	Curaj	22	Male	Muscle	3	2250	24-08-2022
14	Yash Khandelwal	9057576440	Bandarsindari	29	Male	Muscle	1	700	24-08-2022
15	Ravi jain	8769904528	Bandarsindari	29	Male	Muscle	1	700	24-08-2022
16	Deepak Jangid	7339921281	Bandarsindari	24	Male	Muscle	1	700	24-08-2022
17	Abhishek Sharma	7014740205	Bandarsindari	20	Male	Muscle	1	700	24-08-2022
18	Jaishree	7277852577	Curaj	26	Female	Both	1	1600	24-08-2022
19	Samir khan	8875161173	Curaj	19	Male	Muscle	1	800	24-08-2022
20	Astitva Tripathi	9369167154	Curaj	21	Male	Muscle	1	800	24-08-2022
21	Hitesh	9057581780	Bandarsindari	22	Male	Muscle	1	700	25-08-2022
22	Pradhan Gurjar	7568157681	Mundoti	21	Male	Muscle	1	700	25-08-2022
23	Jatin	7877606082	Curaj	23	Male	Muscle	1	1000	25-08-2022
24	Aarya Nayak	9182630710	Curaj	21	Male	Muscle	1	800	25-08-2022
25	Pawan Sharma	8952022079	Curaj	20	Male	Muscle	3	2100	25-08-2022
26	Udit Bhaskar	8107483927	Curaj	20	Male	Muscle	1	800	25-08-2022

3.2 Descriptive Statistics

I collected data from August 2022 to September 3, 2023, with the intention of conducting a comprehensive analysis. However, due to incomplete data for the last month of the period, I made the decision to exclude that data from the analysis. Instead, I focused my analysis on the data available for the rest of the time frame. This approach was chosen to ensure the integrity and accuracy of the results obtained from the available data.

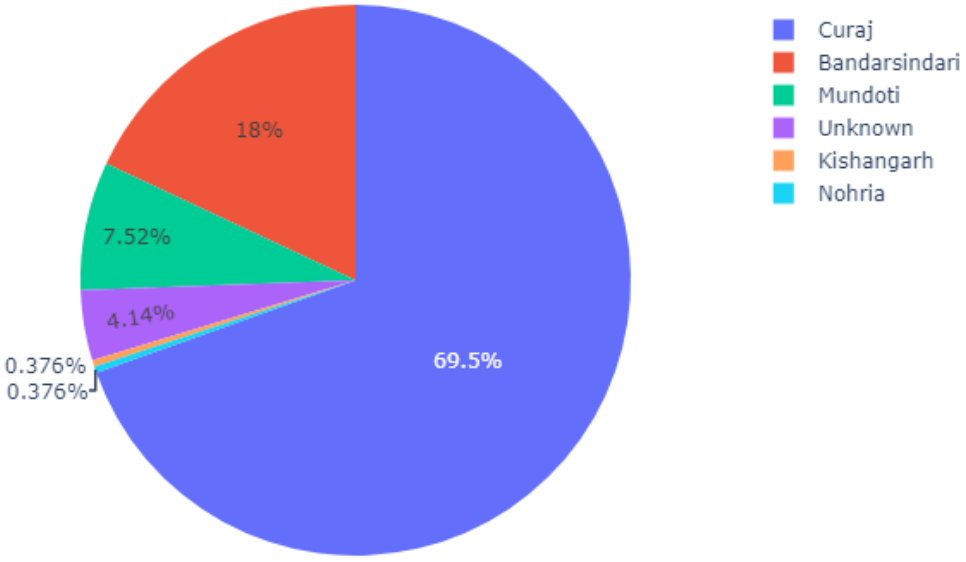
3.2.1 Customer Demography Analysis



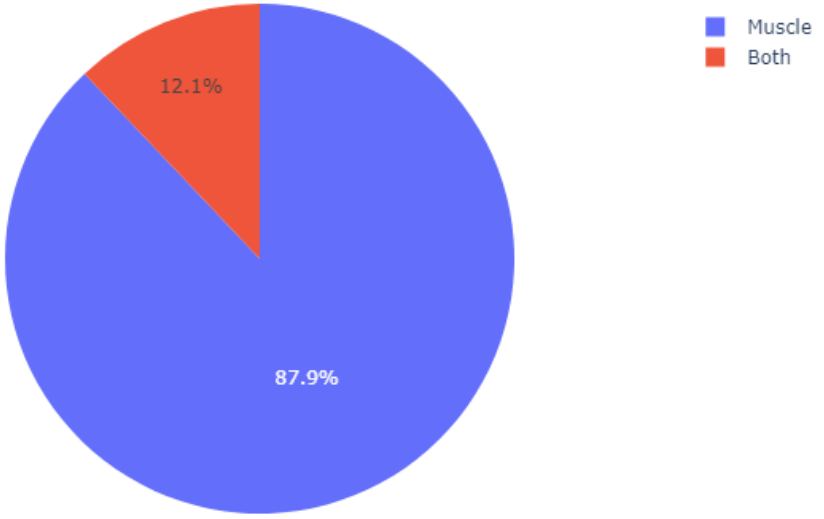


	Age
count	253
mean	23.72727
std	5.148656
min	16
25%	21
50%	23
75%	26
max	52

Address Distribution



Package Distribution



4 Detailed Explanation of Analysis Process/Method

The Analysis Process involved the following steps:

- Segmenting the Data
- Missing Value Removal
- Removing Redundant Data
- Validating the Data
- Creating Required Subsets for Analysis

4.1 Filtering Relevant Data

The "Package_renewal" dataset presented a unique challenge. It combined information about subscription renewals for multiple months, resulting in data entries with non-empty values only for the specific month in which a customer renewed their subscription. For all other months, the dataset contained empty or blank values in various columns. Additionally, the dataset included information about the amount paid by customers at the time of registration. This additional information posed a challenge when attempting to analyse data exclusively related to renewals. By segregating the data into monthly files, you've made the dataset more manageable and suitable for analysis.

4.2 Missing Value Removal

Upon segmenting the data into monthly files, you observed the presence of missing values, which had been inherited from the original "Package_renewal" dataset. To ensure data integrity and accuracy, I decided to remove the missing values from the renewal files, as it was crucial to have complete and reliable data for analysing subscription renewals.

In the case of the "Customer Information" dataset, I noticed that there were missing values in critical columns such as "Age," "Address," "Amount," and "Package_type." However, removing rows with missing values in this dataset would mean losing valuable information related to the amount paid at the time of registration and the details about the subscription start date. To address this challenge, I decided to replace the missing values in these columns with "unknown." This allowed me to retain the records and associated information, even in cases where certain details were not available.

4.3 Removing Redundant Data

After effectively segmenting the data and addressing the missing values, the next challenge was to remove redundant data from the monthly files. These redundant entries contained information regarding the amount paid at both the time of registration and at the time of renewal. Moreover, there was no clear way to distinguish which rows corresponded to renewal transactions and which ones were related to registration.

To tackle this issue, I leveraged the "Customer_Id" from the "Customer Information" dataset. This unique identifier allowed me to link and cross-reference the payments with specific customers. By doing so, I could confidently eliminate the payments made at the time of registration, ensuring that only the relevant renewal transactions were retained for analysis. This approach not only streamlined the dataset but also enhanced its accuracy and appropriateness for analyzing renewal-related payment patterns while maintaining the essential context provided by the "Customer Information" dataset.

4.4 Validating the Data

Following the removal of redundant data and the utilization of Customer_Id for differentiation, the next crucial step was to ensure the integrity of the dataset through thorough data validation. During this process, I meticulously reviewed the data and identified minor errors, such as inconsistencies in the date column.

One specific issue I encountered was the sequencing of dates. For instance, I observed that "Dec-23" appeared before "January-23" in the data, which was clearly inaccurate. To rectify these types of date-related errors, I took the necessary corrective measures to ensure that the date column adhered to a logical and chronological order.

4.5 Creating Required Subsets for Analysis

Having successfully validated the data and resolved any integrity issues, the next step in the analysis process involved creating several valuable subsets from the dataset. These subsets were instrumental in gaining deeper insights and understanding various aspects of the data.

- Revenue Dataset: Capturing revenue from new registrations, renewals, and total revenue.
- Monthly Present Contingency Table: Analyzing categorical variables like Gender, Address, and Package_type for pattern recognition.
- Attrition Rate: Calculating customer churn rates for each month.
- New Customer Count for Each Month: Tracking new customer acquisitions for trend analysis.

These subsets facilitated in-depth analysis and informed decision-making based on revenue, customer demographics, attrition, and customer acquisition patterns.

5 Results and Findings

Through data visualization, I gained valuable insights into the trends within the business. Notably, I observed a significant decline in business revenue starting in February 2023. This decline was closely tied to the decreasing number of active customers, as evident in Figure 1, where we observed a rapid drop in customer numbers and an increase in the attrition rate, as evident in Figure 2. In Figure 3, we can see a promising start for the gym, but after January 2023, the count of active customers began to steadily decrease.

To delve deeper into the analysis, I plotted the active customer count against categorical variables such as “Address,” “Gender,” and “Package_type.” Figure 4 revealed that "Curaj" (college) was the primary source of active business customers. However, post-January, we observed a substantial decline in the number of active customers from "Curaj," which had a substantial impact on business revenue.

Intriguingly, when examining the relationship between active customers and gender in Figure 5, it became apparent that the number of active female customers was nearly non-existent.

Furthermore, I explored the connection between active customers and the “Package_type” variable in Figure 6. This visualization revealed that the number of active customers for the “Both(cardio+muscle)” package type dropped to almost zero in later months.

These visualizations provided critical insights into customer trends, demographic patterns, and package preferences. They shed light on the factors contributing to the revenue decline, helping to guide future strategies and decisions for the business.

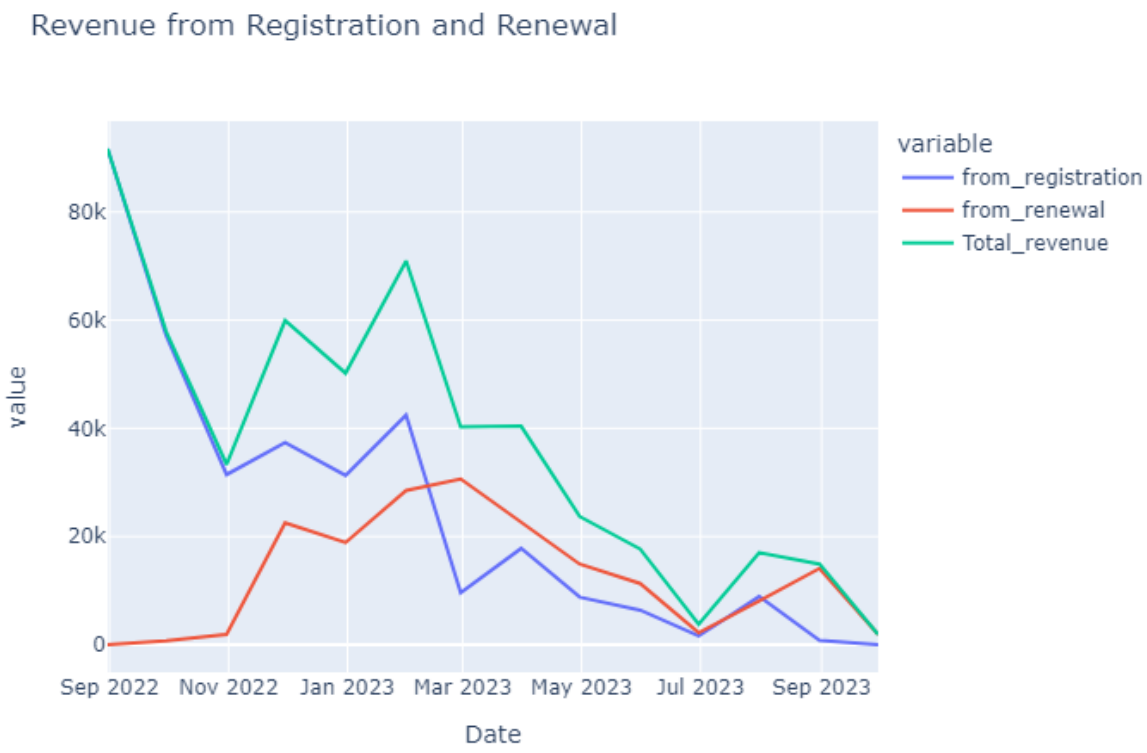


Figure 1 Revenue Plot

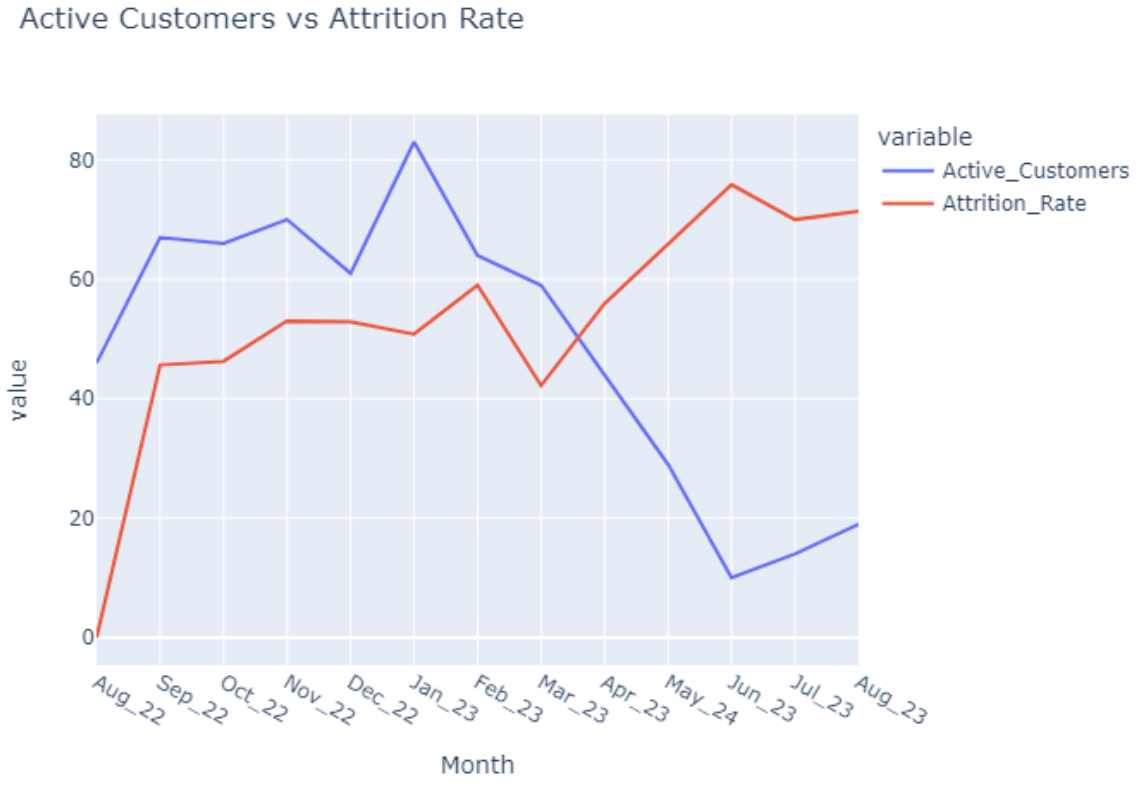


Figure 2 Active Customers vs Attrition Rate

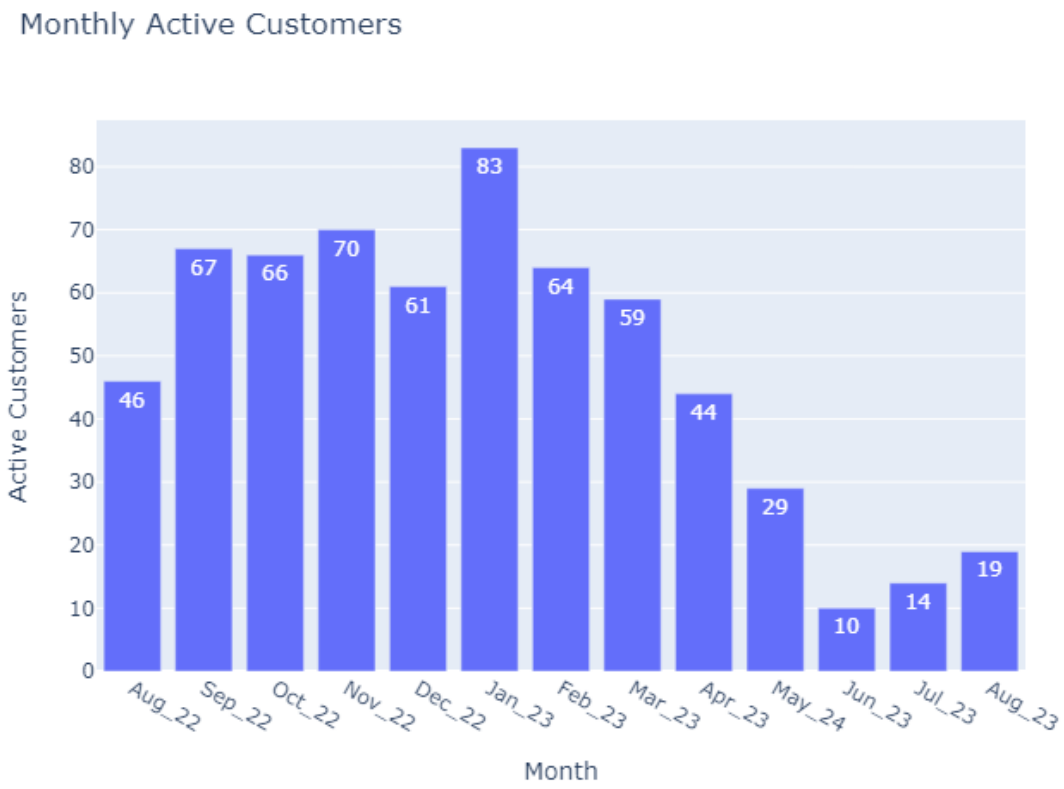


Figure 3 Monthly Active Customers

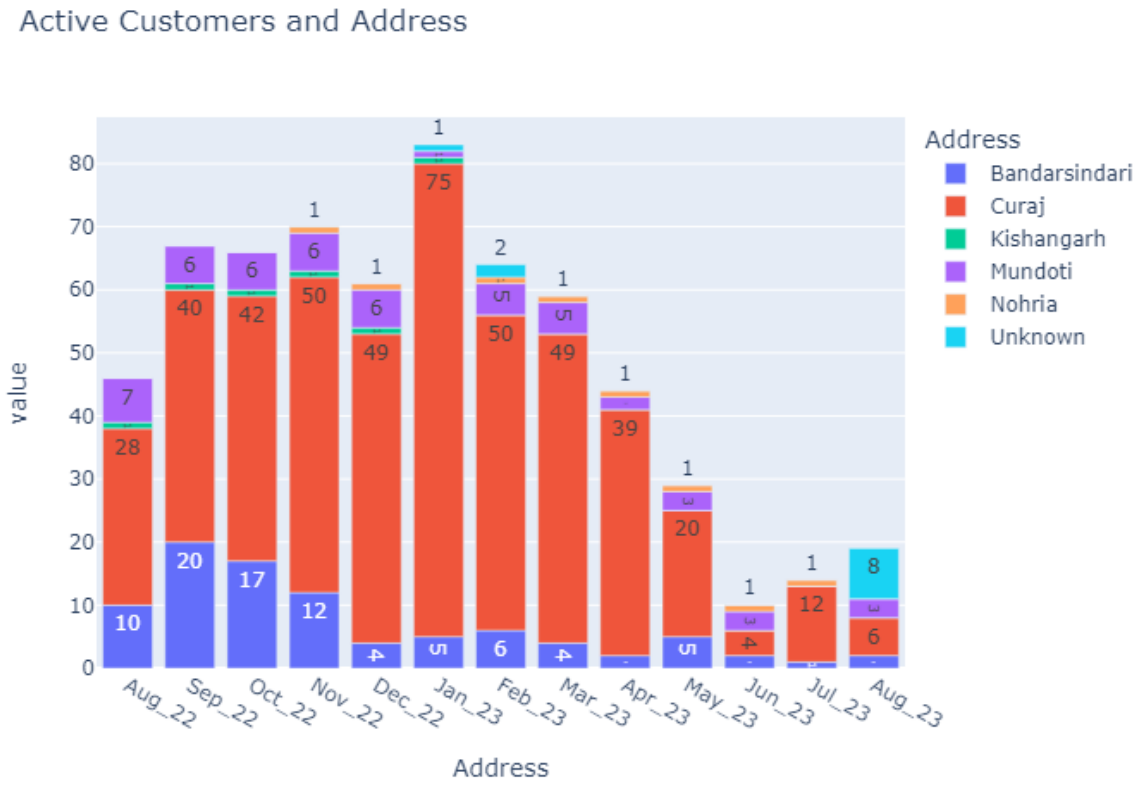


Figure 4 Stacked Bar Plot of Active Customers with Address

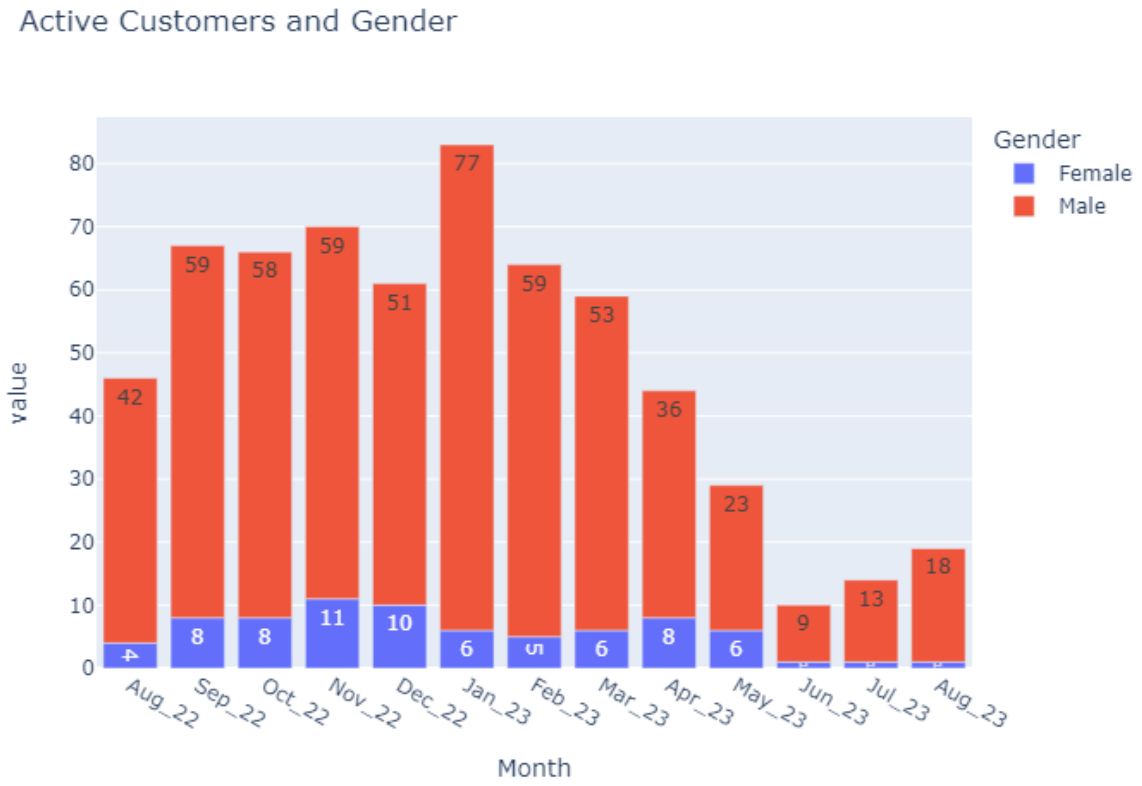


Figure 5 Stacked Bar Plot of Active Customers with Gender

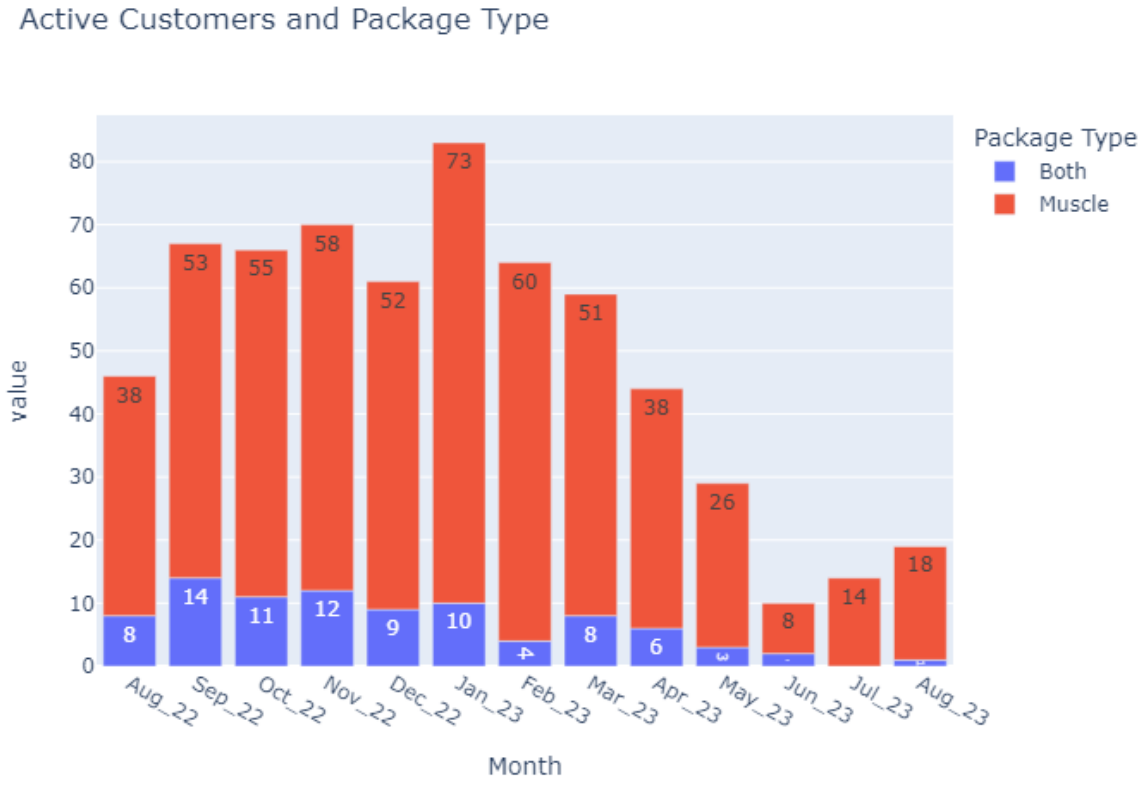


Figure 6 Stacked Bar Plot of Active Customers with Package Type