

Fitness Center Case Exploration

After preliminary investigation of 26 months of data from the fitness centers, in this document, we will list and review our findings, insights, discuss the data quality and concerns, and recommend a model suitable for the business problem. The Chief Marketing Officer has observed that there is a constant churning of customers, especially of those who enroll at the beginning of each year. We want to understand the factors that are influencing a customer to default on their membership and whether there is any economic benefit for the fitness center in attempting to prevent customers from defaulting.

To begin with, we find that 11.5% of customers churned. We observe below in figure 1, the number of customers enrolling overtime. While there is an overall increase in how many customers are enrolling, we observe seasonality in the trend—more customers tend to enroll at the start of the year and during the summer months—similar to the Chief Marketing Officer’s concern.

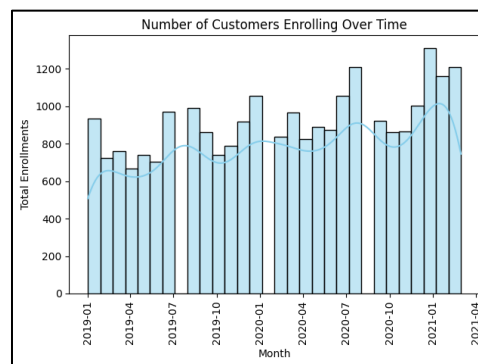


Fig. 1: Customer enrollment overtime has a seasonal trend.

In figure 2, we observe that customer defaulting seems to increase if customers are enrolling in mid-summer or at around the end and beginning of each year. On the other hand, there is a decrease in churning for enrollment dates in late summer and early fall. This is a potential area for further investigation. Are there promotions run during this time for other fitness centers that could cause customers to default in favor of a cheaper plan? Do customers start to lose motivation a few months later if they signed up at the start of the year, which would also cause them to stop visiting the fitness center?

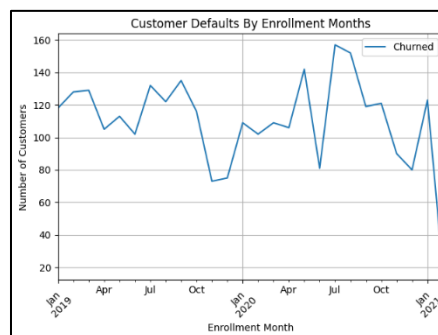


Fig. 2: Customer churning over time may potentially be correlated with enrollment month.

To continue our investigation, we observed the proportion of customers that defaulted based on their gender, payment type, usage level, and age. In figure 3, we found that there is a roughly equal proportion of both male and female customers among those who did not churn. However, of the customers who churned, a higher proportion were females.

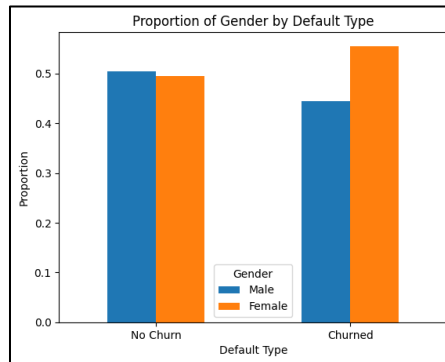


Fig. 3: More female customers churned than male customers.

We observed that approximately 63.18% of customers (from a total of 24,843 customers in the dataset) had low usage levels of either 0 or 1. This is a potential point of concern and further investigation is warranted into evaluating why a majority of customers do not fully use the fitness centers to their advantage. Furthermore, as shown in figure 4, for those customers who churned, most of them were also in lower usage levels and almost none of them had higher usage levels (7 or 8).

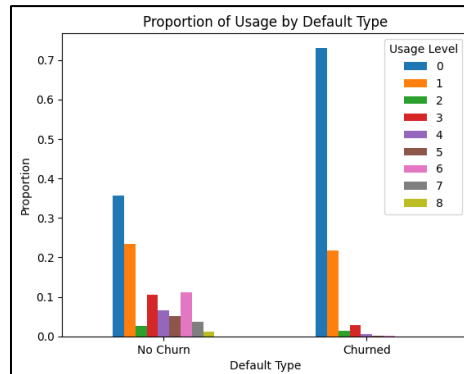


Fig. 4: Most customers that churned had low usage levels.

In figure 5, we observe that there is an equal proportion of each payment type among those who did not churn. However, of those who churned many of those customers used check or cash payment. It is possible that customers who use direct debit or credit are paying off the plan on a monthly basis so they are less likely to churn if they are currently paying for the plan, whereas, those who paid with cash or check have already paid of the plans so it may not matter to them if they churn or not. Further investigation into this could lead to additional interesting insights.

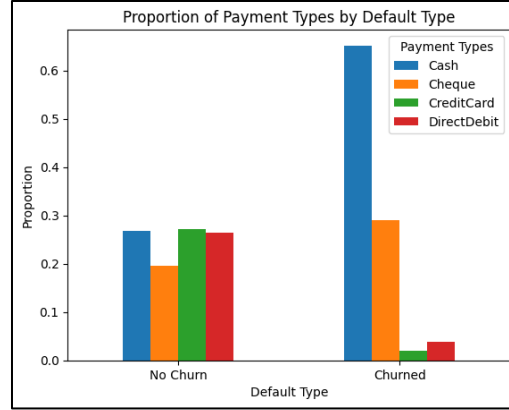


Fig. 5: A majority of customers that churned paid with cash or check.

Lastly, we found that most of the customers in the dataset were between the ages of 22-36 years old. In figure 6, we observed that there is an increase in churning among customers younger than 20 years old, and a steady decline in churning among customers older than 20 years old.

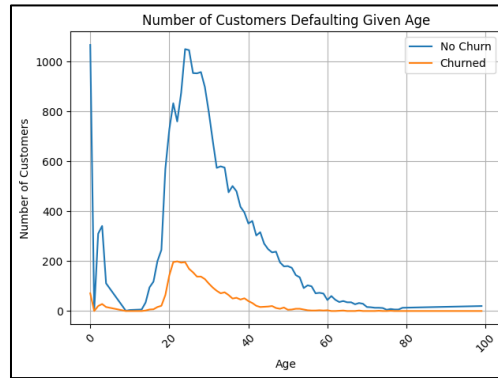


Fig. 6: There is an increase in churning among customers younger than 20 years old.

After our preliminary investigation, we find that there is a potential correlation between default and the following variables: age, payment type, usage level, and gender. We hypothesize that female customers, customers that paid in cash or check, younger customers under the age of 20, and those who don't frequently use the fitness centers are all more likely to churn respectively.

As the target is to identify what factors are influencing churning, we propose a logistic regression model to initiate the analysis. Since default is a binary variable (the customer either churns or they do not), logistic regression would work best compared to linear regression which is better for a linear regression. Logistic regression predicting whether or not a customer will default using the predictors of age, payment type, usage level, and gender will allow us to identify whether these predictor variables truly influence customer defaulting. Furthermore, this model can be extended to a classifier to predict if a new customer will default given their gender, age, choice of payment type, and previous usage history at other fitness centers. This

classification would be helpful in preventing loss to the business by preemptively identifying customers at high risk of churning and not finishing their plan payments.

An additional model that could be helpful in analyzing the economic benefit to the fitness center of determining influencing factors on churning is a linear regression model predicting price given the variables age, payment type, usage level, and gender. This could be used to implement a dynamic pricing model and optimize a plan price per customer to minimize their likelihood of churning. Since payment price is a continuous variable, linear regression is the best model for this use case.

Part of our initial investigation included an assessment of data quality. Although there are no missing values in the entire dataset, some variables in the data take on values that warrant clarification. 4.58% (1,138 customers) in the dataset have a reported age of 0 years old. Is this due to an omission of age by the customer during enrollment? Additionally, 71 customers have a reported price plan of \$0. Were these customers part of a promotion or was this due to a free trial? This is a similar case for the down payment variable, where 243 customers have a reported \$0 down payment. Were they exempt from paying a down payment.

Another variable of concern is the month due variable. Would it be possible to clarify what this variable represents? It was reported that this variable corresponds to the monthly payment, however, for each customer, the monthly payment across 12 months plus the downpayment does not equal the annual plan price.

Lastly, there are a few continuous variables with skewed distributions. For age, there are a few outliers with customer ages greater than 80 years old, while most customers are between 20-40 years old. For both downpayment and price, the distributions have a few outliers above the \$2000 range. Why are there customers paying downpayments or annual prices greater than \$2000? Clarifying the meaning of the monthdue variable and identifying the reason behind outliers in the down payment, age, and price variables will allow for a more complete and comprehensive analysis of the data and a more accurate implementation of modeling going forward.

We have a few questions about your business that will help us better understand the context of this business problem and methods to remediate the number of customers defaulting:

- What other businesses do you compete with?
- Where are these fitness centers located? What amenities does each center offer?
- Are there any consequences if a customer churns? Does the customer have to pay off the entire plan and an additional fee?
- What are the operational costs of each facility?
- Have any promotional offers been conducted in the past? Is there any data from these promotions that we can access?

Gaining more information on the operation costs, competition, and past promotional offers will be helpful with analyzing the economic benefit of remediating customer churning. Thank you for taking the time to read our proposal, please email us if you have any follow-up concerns or questions.