

Q. State the problem described by the client & desired outcomes.

Our client is the Chief Marketing Officer of a Fitness Club. The CMO's main concern is customer churn. It was noticed that majority of the clients signing up for monthly subscriptions especially in the beginning of the year right after holiday season, tend to terminate their membership in a couple of months. The CMO is concerned about this high customer churn rate and wants to devise a solution aiming for customer retention. He would like to identify a pattern in this customer group in order to understand the driving factors of churn.

The client is also interested in understanding the economic benefit that the Fitness Club would reap in the event of reducing the number of people who leave. We need to help the client devise a good marketing strategy aimed at curbing churn and retaining customers for a longer duration.

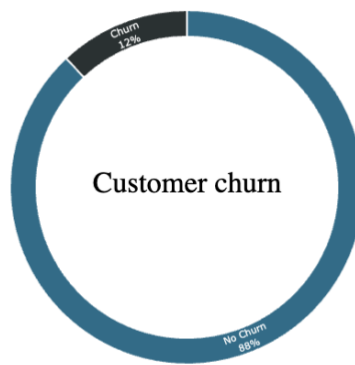
Using the SMART FRAMEWORK, the current state is that the Fitness club is witnessing high customer churn and the desired future state is developing a marketing strategy by identifying the factors for churn and increase customer retention. The timelines of the project and the team involved from the client's end need to be further discussed with the client

Q. Describe your hypothesis identifying the business drivers (influences).

From the EDA, on studying the distribution of variables we can infer the following and draw a few hypotheses.

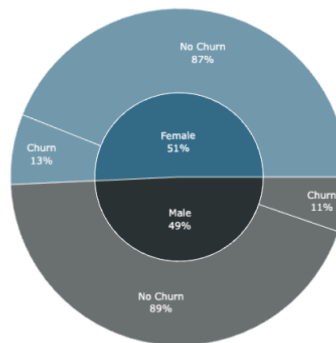
Only 12% of churn is observed across the 3 years 2019, 2020 and 2021 with 9043, 10174 and 3362 customers enrolling in the corresponding years. Only 3 months of data is available for the year 2021.

Though the client mentioned that more people sign up at the start of the year, the data shows that maximum number of people enrolled in the months of July, August and September.



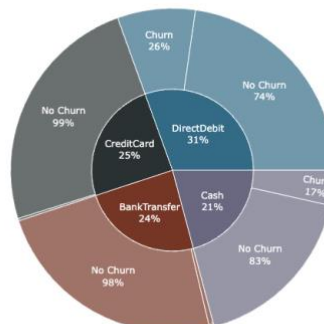
Gender - Gender has no significant effect on the attrition rate.

Effect of gender on customer churn



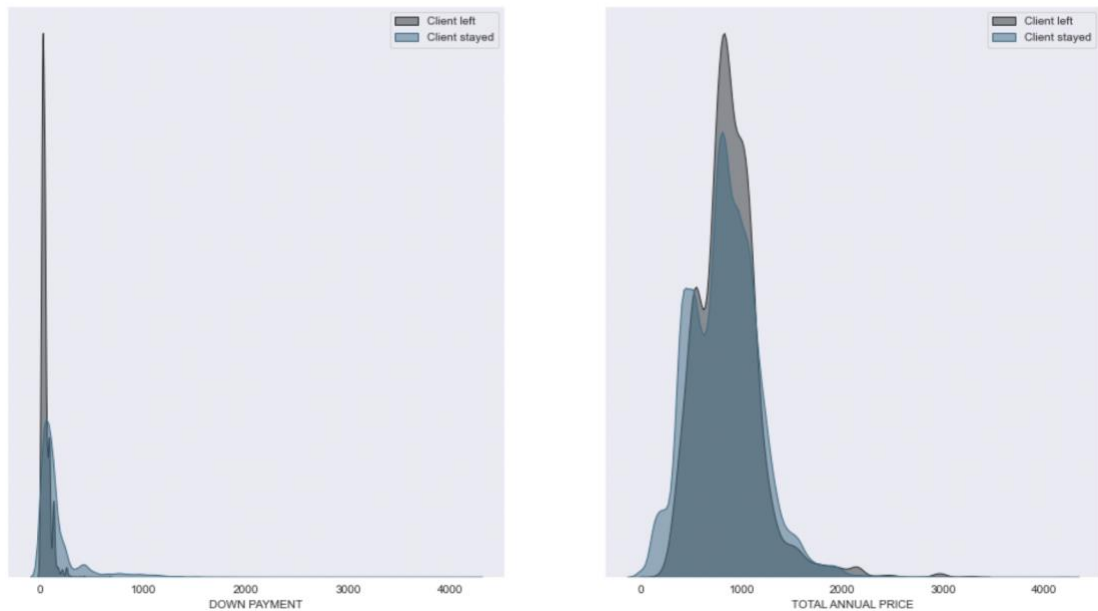
Mode of payment - Customers using payment options Cash and DirectDebit are more likely to churn.

Effect of payment type on customer churn

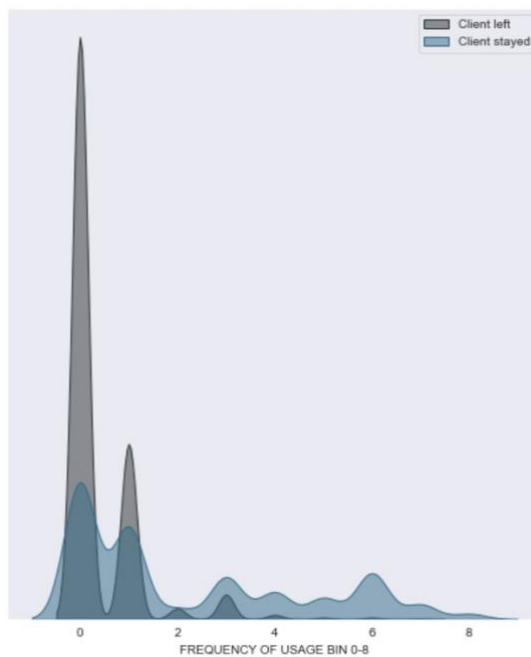


Down payment - Customers who pay low down payment are more susceptible to stop coming to the gym regularly.

Annual rate – Customers who signed up for short term membership plans are more loyal compared to customers who sign up for long terms. Most members are likely to terminate their membership after a few months because they give up on their goals or they get burnt out



Gym visit frequency – Regular customers have high retention rates compared to irregular customers as they are more disciplined and focussed on their fitness regime and are more likely to be consistent with their workout. Regular and long term customers have more brand affinity.



Some external business drivers are :

Infrastructure – Most members usually tend to change gyms because of the insufficient infrastructure. Improving infrastructure with top grade workout equipment can help increase customer retention rates

Trends – Changing trends with the introduction of new workout routines like Zumba dance routines, MMA, Boxing, Pilates, Spin classes and other such group and instructor led workouts have influenced people to move out of conventional gyms to keep up with the times. The fitness club can expand services in order to retain customers

Customer service - Increasing the level of customer service each member receives through behavioural change techniques can help improve retention

Competition – A price war with the competition or competitor fitness clubs offering extremely inviting discounts on membership plans can influence customers to leave and can impact member acquisition as well

Pandemics – Unforeseen external factors like pandemics and full outbreaks can impact the fitness club by reducing footfall as customers would prefer to work out from the comfort of their home with high regard to safety and sanitation

Q. Identify data issues that might influence model selection

The data provided contains no null values but some duplicates were encountered

The *age* data had a number of outliers and this was treated by removing the customers with age less than 16 and older than 99 intuitively

We can analyse the data and compute retention rates across fixed periods of time and for consistency and in order to avoid bias we might have to choose data with enrolment dates in 2019 and 2020 as the data provided for the year 2021 is insufficient

Client needs to provide more clarity on the *monthdue* column and the *price* (annual cost) columns as we would need more additional information into why different customers have different annual rates. Different plans and programs are also essential features that could influence the model

We have information regarding enrolment dates along with Churn or No churn but we do not have the data to help us understand the timeframe. This would be essential in drawing patterns amongst customers based on how long they stayed with the gym before churning

The *default* column giving is information about Churn or No churn is categorical and since we have labelled data we would be implementing a **Supervised Machine Learning Algorithm**

Q. Identify all factors that might influence model selection process

The model would be selected based on the following factors

Timeline of the project – We need to first align with the client on the project timelines and the timeline of data available for analyses because the complexity of the model could be based off of the project timelines

Data Quality and Availability – Preliminary analysis and extended data acquisition has to be done both from internal and external sources once all the business drivers are identified. The model is heavily dependent on the amount of usable data available

Client Analytics Maturity – We need a thorough understanding of how analytics ready the client is and is the client looking forward to a complex analytical solution to enhance their business or for ready to implement business insights. This ties back to data quality mainly because a analytically mature client would have more structured and relevant data which in turn will translate to ease of model implementation

Life Cycle View – Are the clients looking for a solution that helps them retain existing customers or to understand the pattern of churned customers and use it to identify potential customers and devise a strategy to retain them or draw in new customers to make up for high retention? Do the clients need a marketing strategy to identify target customers for a one-time discount program or are they looking to have a long term plan? These questions need to be answered and the relevant model would be selected

Essentially, we would be building a **Binary classification** model with the target feature being *customers who are most likely to leave (tagging customers with Churn or No churn)*. In the presence of additional data, we could also select the *customers leaving month* as the target feature and build a model to identify it. We would need to implement a logistic regression model or a random forest classifier and measure their accuracy and precision on train and validation data sets.

Additionally, we could also train a clustering model with the K-means clustering algorithm in order to identify user clusters by setting aside the *default* column (containing Churn or No churn tags). On identifying the clusters, we can compute the churn rates for each cluster. This will help us identify the cluster that has customers that are more prone to leaving and loyal customers.