Fitness Centre EDA

**What do you consider the specific business problem and the matching goal of analysis?**

The primary business challenge centres on customer attrition in the chain of fitness centres. While new memberships are acquired throughout the year, with a notable surge in the early months, there is a consistent pattern of member disengagement and dropout. This turnover phenomenon has substantial financial implications for the business, warranting a comprehensive analysis. The core objectives of this analytical study are to:

1. Identify the key drivers contributing to customer churn

2. Discover specific patterns or customer segments exhibiting higher attrition rates

3. Quantify the financial impact of member disengagement

4. Generate actionable insights to inform leadership's strategic decisions on targeted marketing and retention initiatives

Data Insights and Observed Trends

Key Observations:

- There is an inverse correlation between the size of initial down payments and customer retention. Members making smaller upfront financial commitments are more prone to discontinuing their memberships, suggesting that lower initial investments may reduce the psychological barriers to cancellation.

- Along similar lines, although the annual price of the gym membership ranges up to $7800, most of the churned customers are observed to have low membership costs at ~$880.
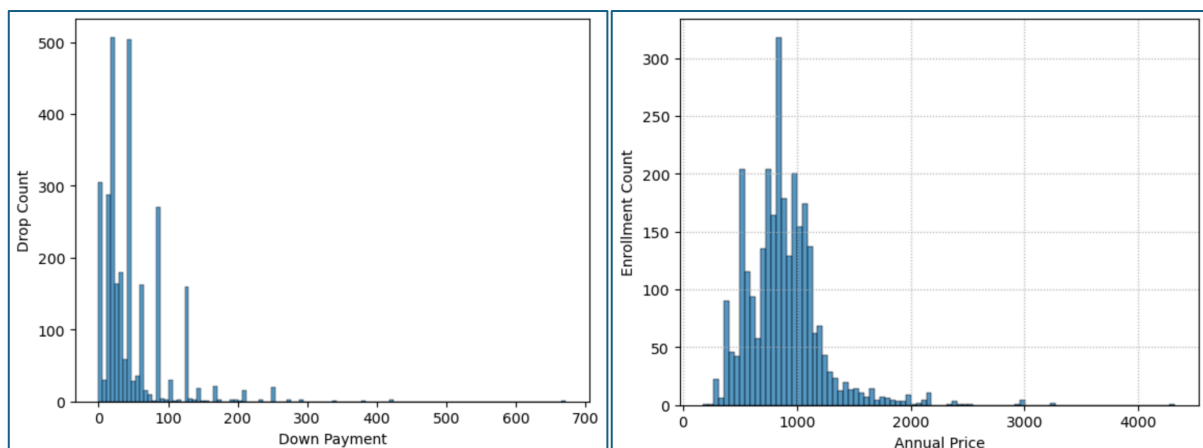
- Payment modality significantly affects retention rates, with cash and cheque users showing higher attrition. The absence of automatic debits for these methods allows easier discontinuation of memberships, potentially increasing churn rates.

- Additionally, the usage rate of customers is also impactful here. The lesser a customer uses the fitness center, more likely he/she is to drop out. We can clearly see a pattern where the red stack which represents the churned customers is higher when usage is low.
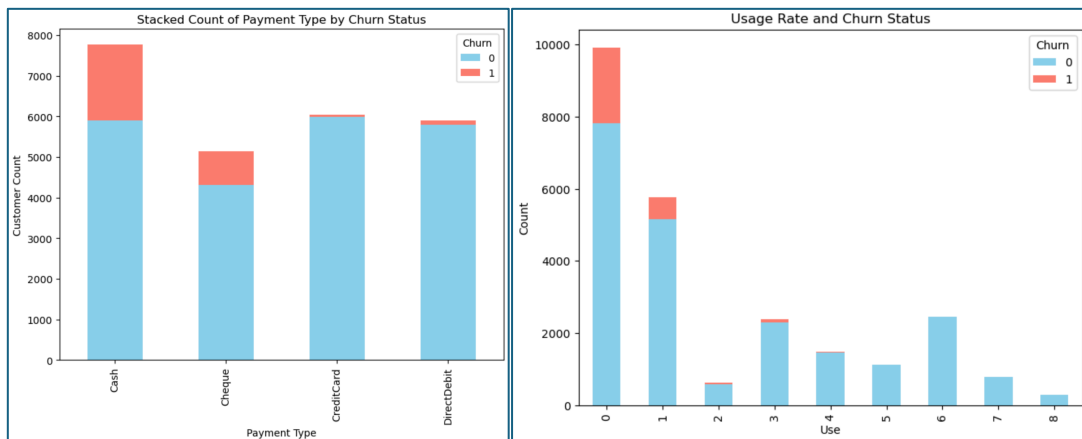
Potential Explanatory Factors:

- Individual health and fitness commitment levels

- Seasonal influences on member motivation and engagement

- External environmental factors affecting gym attendance and retention

What modelling approach do you suggest and what did you notice in the data that suggests it will be helpful here?

Given the nature of the data and the binary outcome of interest (churn vs. retention), a logistic regression model is recommended as the primary analytical model. This approach is well-suited for predicting customer churn probability and elucidating the factors that significantly influence member retention.

Data Preprocessing Steps:

1. Exclude data points with age values below 16 or above 80

2. Transform enrolment dates into a standardized datetime format

3. Implement categorical variables encoding using one-hot or dummy coding techniques for variables like 'payment type' and 'use'.

4. Address null values and missing data points, if any

5. Engineer new features to potentially enhance model performance

6. Eliminate highly correlated features (correlation coefficient > 0.90)

7. Mitigate class imbalance issues arising from the data skewness towards non-churned customers
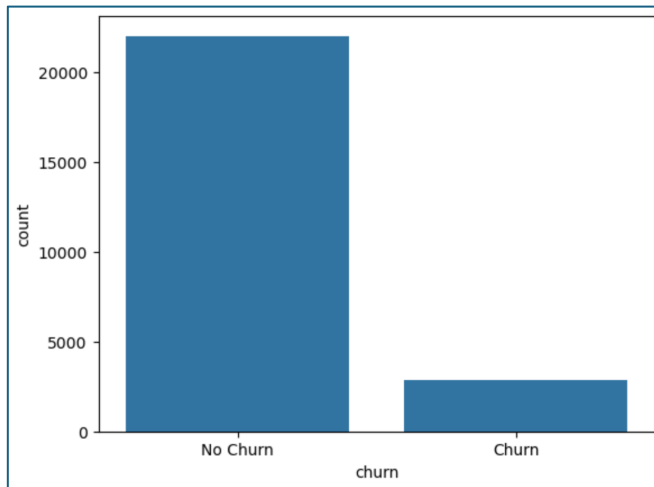
Model Development and Evaluation Framework:

1. Partition the dataset into training and test subsets

2. Calibrate model parameters using the training dataset

3. Generate predictions on the test dataset

4. Assess model efficacy through confusion matrix analysis and classification report

What questions, if any, do you have about the client organization, or its broader industry?

To develop a more nuanced understanding of the fitness centre industry and the client's specific organizational context, the following questions are proposed:

1. What are the distinguishing features or unique value propositions of this fitness centre compared to industry competitors?

2. Are systematic exit interviews or surveys conducted for departing members? What are the predominant reasons cited for discontinuation?

3. Are there any promotional campaigns or discount sales done over the year that could impact the customer enrolments? This will help us understand the possibility of some customers enrolling solely under the influence of targeted marketing.

4. What are the underlying seasonal factors driving the observed enrolment peaks in July-August and January?

5. Is personalized training available to members, and if so, what is the associated cost structure?

What questions, if any, do you have about the data?

To ensure data integrity and maximize the analytical value of the available information, the following questions should be addressed:

1. What is the most appropriate method for handling age data outside the 16-80 range? Should these records be excluded or should age values be imputed using average or median values?

2. Can you elaborate on the various data collection methodologies employed across different branches of your fitness centres? How do these methodologies impact data quality, consistency, and storage capacity?

3. Is it feasible to access historical data before 2019 to enhance our understanding of long-term trends and seasonality patterns?

4. Are there additional datasets available that provide contextual information such as geographical, economic, ethnic, or social demographics for the areas surrounding each fitness centre?

5. Do we have access to customer-centric data like their occupation, income category, how far they live away from the centre, do they have family/kids to attend to? These will help us identify the customer priorities and their personal challenges and if there is any way we can address them.

These questions will help us understand the customer base better and help us design strategies that are more impactful.