wodify

Identifying How to Use Free Trials and Discounts to Increase Customer Retention Rates and Customer Lifetime Value

Presented By:
Q CONSULTING
Elizabeth Ipe, Parsa Sedghi, Sebastian Decady,
William Procheska, and Madeline Whitt

Table of Contents

Letter From the Team

Q Consulting
1407 Van Munching Hall
University of Maryland
College Park, MD 20742

May 7th, 2024

Dear Steve, Barrett, and the Wodify Team,

Thank you so much for giving QUEST the opportunity to work with you this semester! Your guidance and support throughout this entire project has been instrumental to our success. We have greatly appreciated all of the time and effort you both have lent to us throughout this process.

With your trust in us, we were able to gain a unique experience working with real-life data on a subject that we all found truly interesting and exciting. Our team greatly enjoyed working with Wodify and we hope that you enjoyed working with QUEST as well!

In the following report, you will find a comprehensive overview of our work for this semester. We hope that our recommendations steer your clients, present and future, aiding them in enhancing their businesses. If you have any questions or if anything is unclear, please do not hesitate to contact us.

Sincerely,

Elizabeth Ipe
Parsa Sedghi
Sebastian Decady
William Procheska
Madeline Whitt

Executive Summary

Wodify is a developer and provider of premium customer relationship management (CRM) software that empowers gym owners and personal trainers to boost their productivity. Many gym owners commonly offer free trials and discounted memberships to promote their gyms and attract new members. Wodify wants to better understand the long-term impact of how these special offers relate to customer retention and lifetime value. They aim to provide data-driven insights and recommendations to their clients so that they can successfully increase their revenues and retain active gym members. Team Q Consulting partnered with Wodify to investigate these long-term impacts and help them achieve their goals.

This report provides detailed analysis pertaining to the impact of free trials and discounts on customer retention and lifetime value for Wodify's clients using their gym CRM software. The Q Consulting Team utilizes six different datasets in their analysis, including gym membership details, free trial information, and specific gym demographics. The data was cleaned and organized by categorizing clients into specific buckets based on whether they signed up for a free trial and/or a discounted membership. Throughout the report, Q Consulting provides valuable visualizations and analysis that display the key factors that influence higher customer retention rates and lifetime values.

Q Consulting's findings suggest that gym members who did not have a free trial but signed up for a discounted membership had a significantly higher retention rate—almost twice as much as other categories. The clients who were offered a discount would've already been comfortable with signing up for a membership, while free trials can often attract customers who are unwilling to continue a membership once the trial is complete. On the other hand, the customer lifetime value analysis shows that gym members who had a free trial then signed up for a membership had the highest median customer lifetime value. The data findings also show seasonal trends in membership sign-ups, with peaks at the beginning of the year that are likely related to fitness-related new year resolutions.

Based on the findings, Q Consulting recommends for gyms looking to enhance retention rates to consider offering more discounted memberships, as this average retention rate was notably higher than all other categories. The team also recommends that gyms continue to offer monthly memberships as they are the most popular, although it is important to consider that yearly memberships result in the highest retention rates. Finally, since the data shows that "free trial to membership" customers had the highest median customer lifetime value, Q Consulting suggests that gyms would benefit from encouraging these people with free trials to sign up for paid memberships.

Client

Wodify is a leading developer of fitness management tools that enable businesses to thrive. Their simple, elegant software suite connects the world to the future of fitness with engaging **performance tracking, dynamic event & competition planning, and comprehensive business management solutions**. To date, they've helped over two million members of 5000+ businesses in 100 countries get fitter, healthier, and happier.

Problem & Scope

Goals

Gyms commonly provide free trials and introductory offers for new sign-ups. Wodify is interested in understanding client retention by examining the **long-term impact of these free trials or discounts on memberships**. Analyzing the effectiveness of these offers can provide insights for recommending successful introductory strategies to gyms, in order to promote **client retention** and contribute to **revenue growth**.

Resources

We were given **6 different datasets**.

1) Four datasets were membership files with client membership information.
2) The fifth dataset was information about client free trials.
3) The last dataset had information about specific gyms.

Approach

Platforms

Our team eagerly took on this problem by cleaning and analyzing the data using Excel and the Python library pandas.

Data Preparation and Cleaning

The data needed to be cleaned of all duplicate LEAD_IDs. We opted to keep the first instance of all unique LEAD_IDs and sum the number of occurrences of that LEAD_ID to determine the retention rate.

Data Organization

We organized the data into five different categories that best described customer behavior across the gym clients. These categories were made by matching identifier information (LEAD_ID) in the free trials file and the memberships file. Here are the five categories:

Clients who had a free trial and then signed up for a membership

a) Matching LEAD_ID in both files

Clients who had a free trial and then signed up for a discounted membership

b) Matching LEAD_ID in both files + COMMITMENT_DISCOUNT is not 0

Clients who had a free trial and then did not sign up for a membership

c) LEAD_ID appears in free trials file, but not in memberships file(s)

Clients who did not have a free trial but signed up for a membership

d) No LEAD_ID in memberships file(s)

Clients who did not have a free trial but signed up for a discounted membership

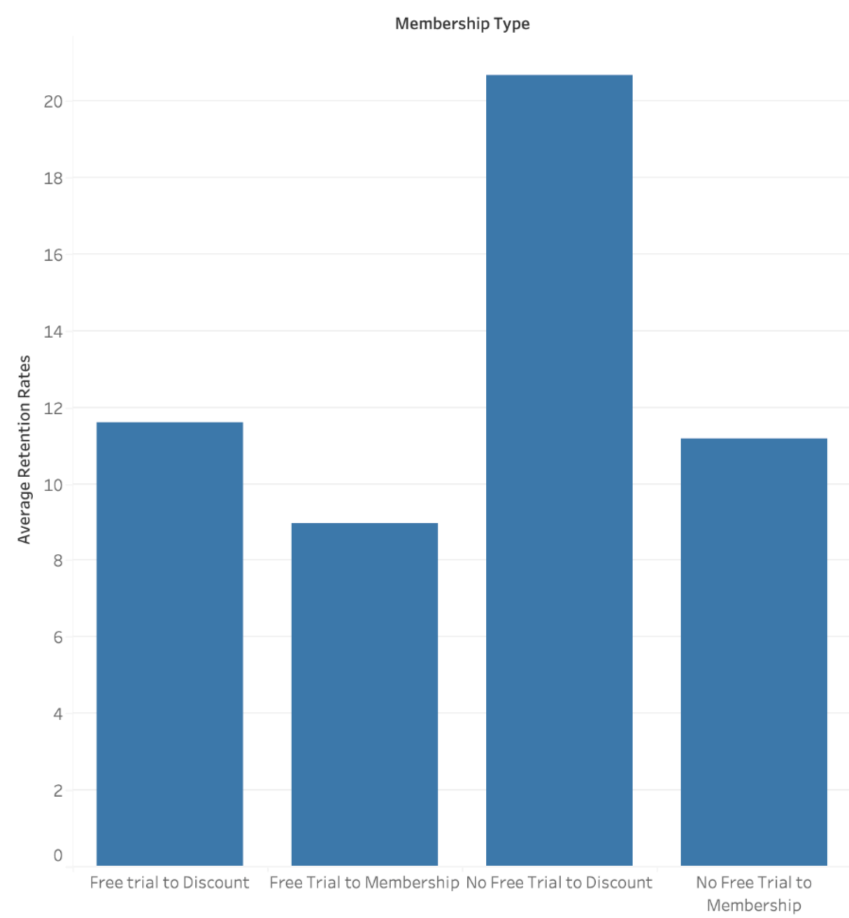
e) No LEAD_ID in memberships file(s) + COMMITMENT_DISCOUNT is not 0

Findings

Customer Retention Rate

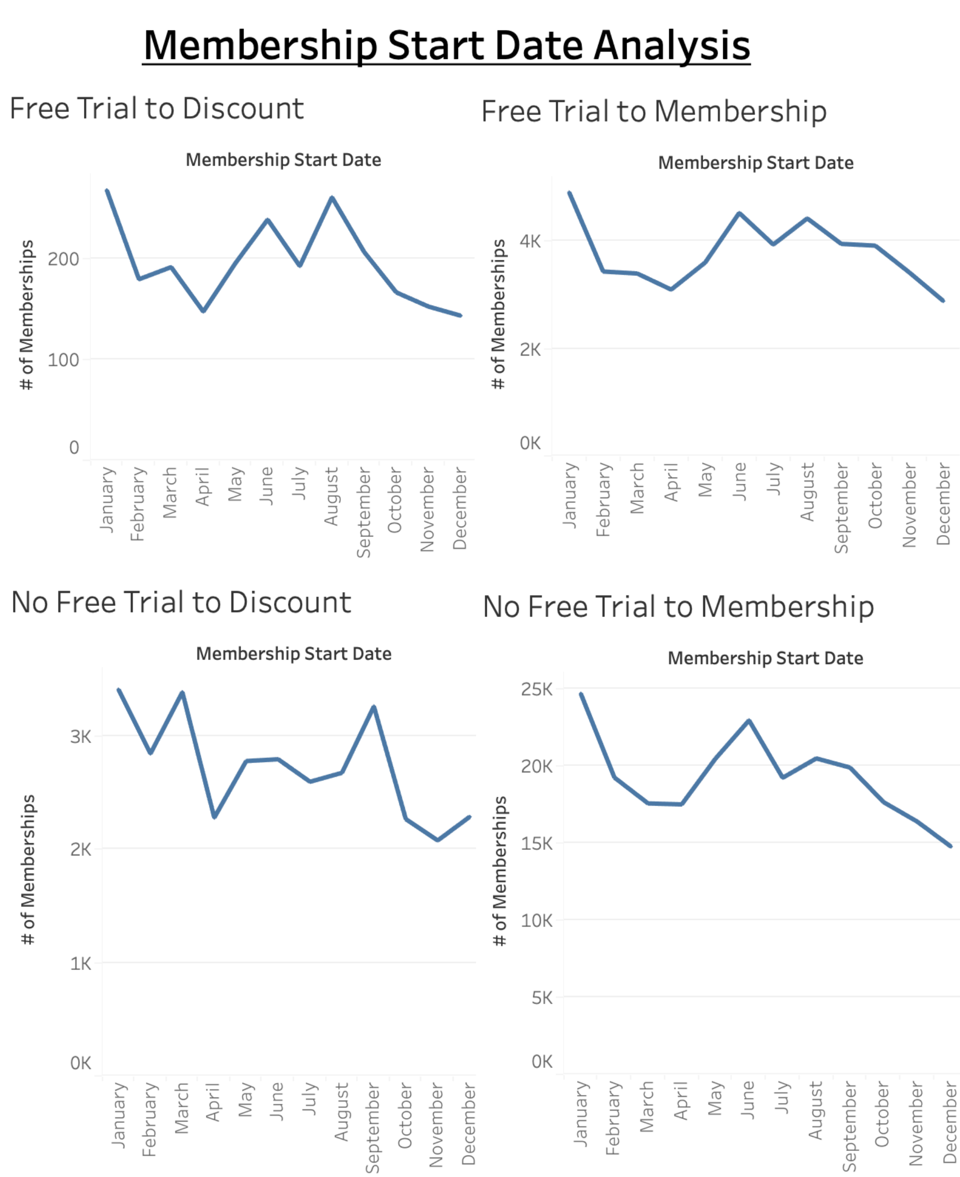
We were able to find the retention rate for each category, represented by MEMBERSHIP_COUNT. We opted to keep the first instance of all unique LEAD_IDs and sum the number of occurrences of that LEAD_ID to determine the retention rate, and this is the MEMBERSHIP_COUNT.

The **average retention rate** for clients who **did not have a free trial and signed up for a membership with a discount** is almost **twice as much** as other categories. This could be because clients who were already comfortable with signing up for a membership without a free trial were offered a discount. These discounts could have been offered at different times during the year at the various gyms.



Start Date Analysis

Using the retention rate, we were able to find trends over time analyzing when customers signed up for memberships.



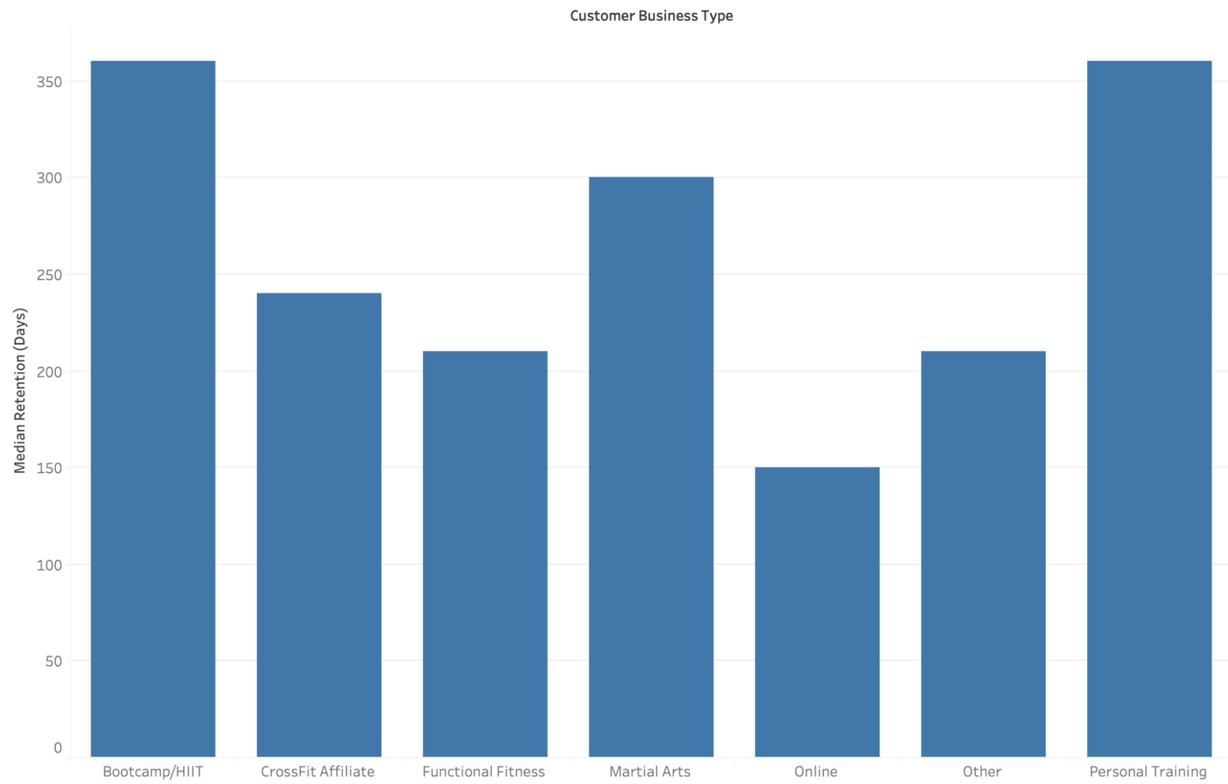
It seems as though a lot of customers **sign up for a membership at the beginning of the year in January**. This is probably because of the tendency for many people to approach the new year with a new goal, which usually tends to be a fitness goal. Colloquially, this is known as a "New year's resolution," and even more informally, "new year, new me!" There is then a huge drop from January to February, and a spike again before the summer in around May, when people start training to get their fitness goal for the summer. The total sample size of "Free Trial to Discount" customers was notably lower than the rest of the observed samples.

Gyms and their Individual Retention Rates

We were able to calculate the retention rates for each gym over the course of the data as well. This is a separate file just for Wodify so they can share customer retention rates with their clients. This will be handed over as a separate file.

Retention by Business Type

Median Customer Retention by Gym Type

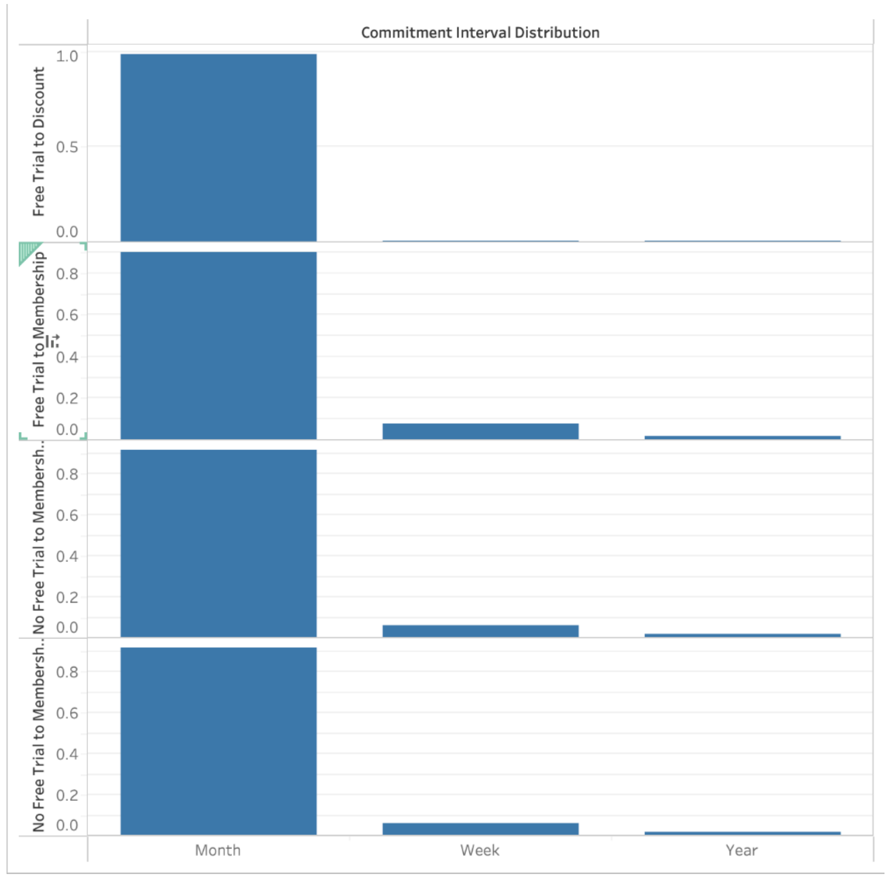


Although personal training has the highest retention rate in days, it is important to note that there are only ~250 personal training clients in the dataset, as opposed to over 200,000 CrossFit affiliate clients and nearly 100,000 functional fitness clients.

* Note that this is essentially an "average membership renewal duration" chart.

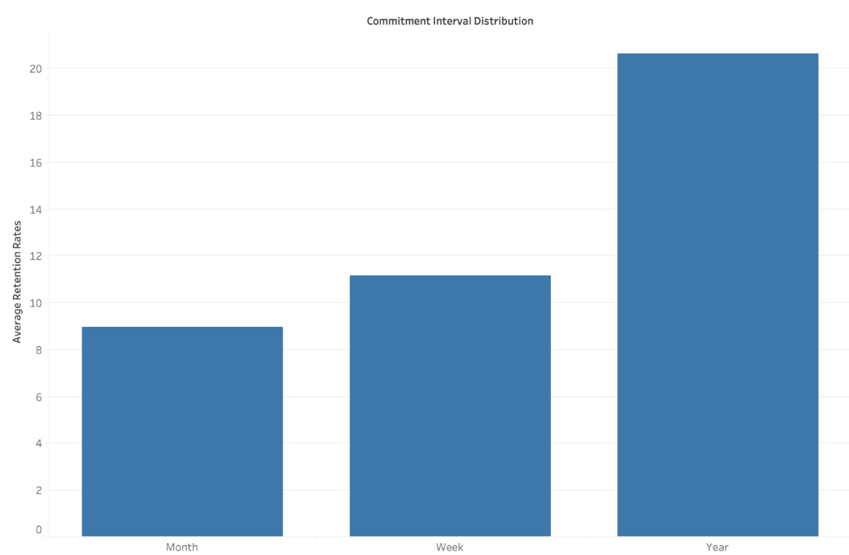
Commitment Interval Distribution

We were able to find the commitment interval distribution for each category.

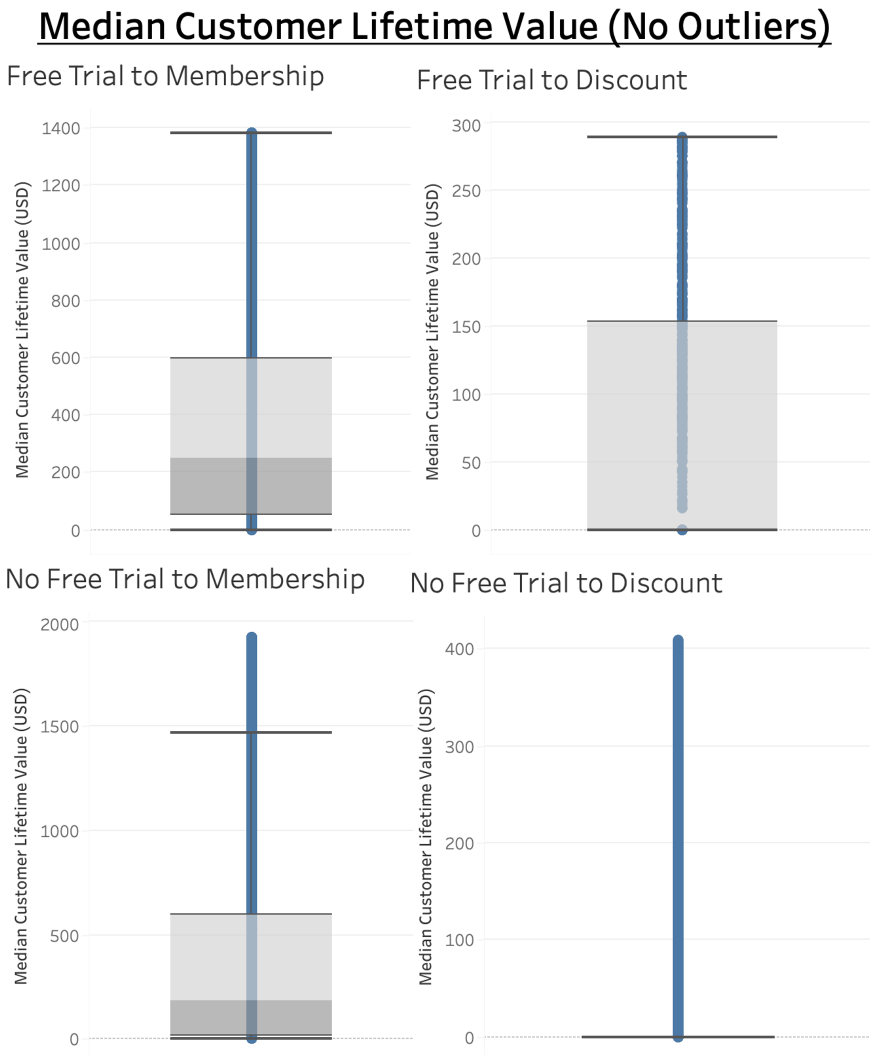


This is the proportion of clients who chose to sign up for a weekly, monthly, or yearly membership.
It is shown that the **majority of memberships were monthly memberships**. This is typically the most widely offered option for a gym membership.

When paired with the average retention rate, those who signed up for a **yearly membership** had the **highest retention rate**.



Customer Lifetime Value

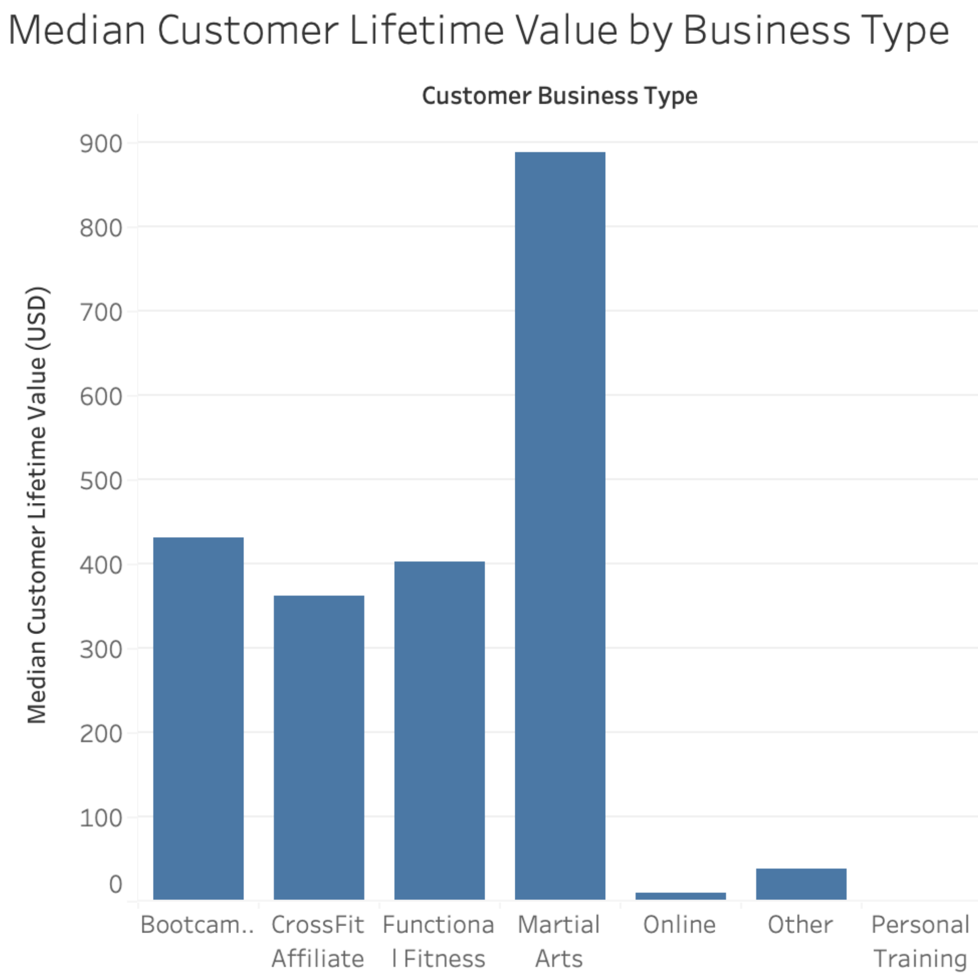


The most notable observation is that the median customer lifetime value of customers who had no free trial and then signed up for a discount is $0.

The highest median is the customer lifetime value of customers who had a free trial and then signed up for a membership, with a value $250. The second highest median is the group of customers who had no free trial and signed up for a membership at a value of $185.

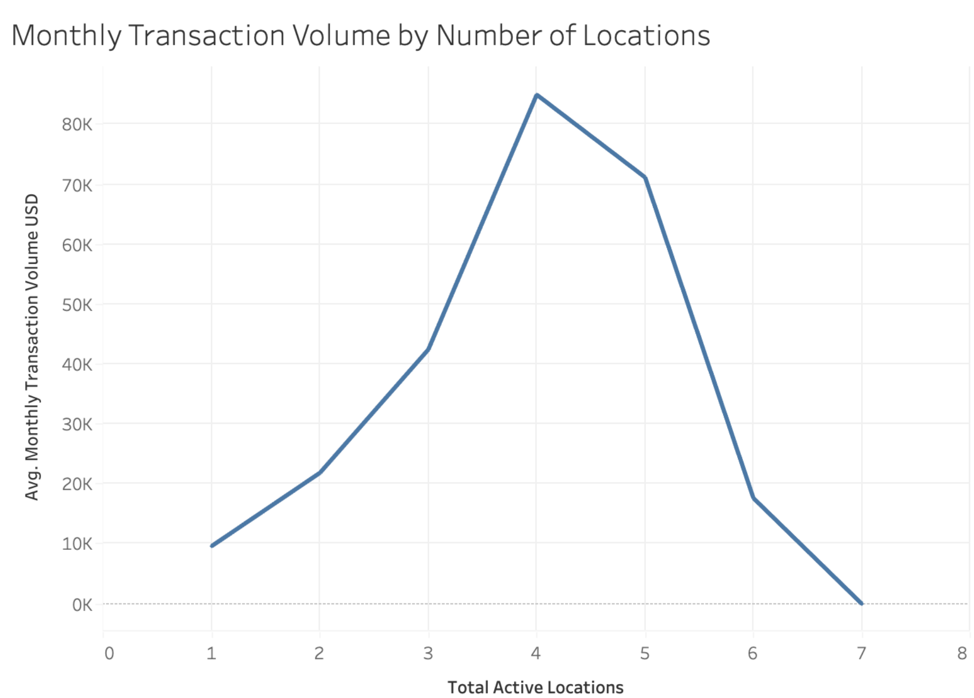
* These distributions were skewed, so outliers were removed. Even without outliers, the distributions are skewed, so analysis was primarily based on the median of the distribution.

Customer Lifetime Value by Gym Type



Yet again, it is important to note that there are only ~250 personal training clients in the dataset, as opposed to over 200,000 CrossFit affiliate clients and nearly 100,000 functional fitness clients.

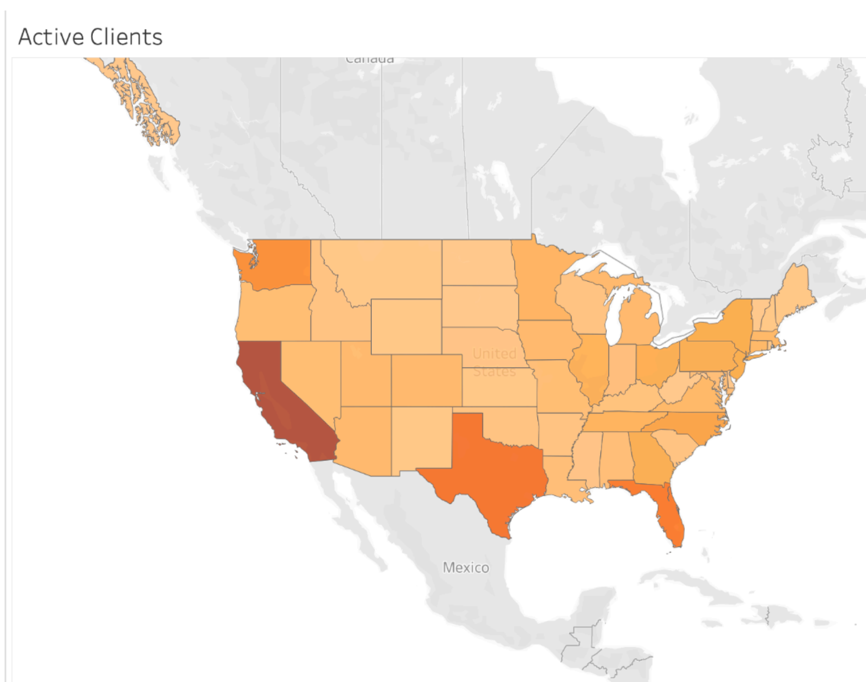
Monthly Transaction Volume



The monthly transaction volume is calculated based on the gym size. The gym size is determined by how many active locations the gym has. On average, gyms with 4 active locations are producing the greatest monthly transaction volume. It makes the argument that those gym owners with only one gym can expand their transaction volume exponentially until the decline starts after 4 active locations. That could possibly be due to business cannibalization as they expand their expansion and could take business from themselves.

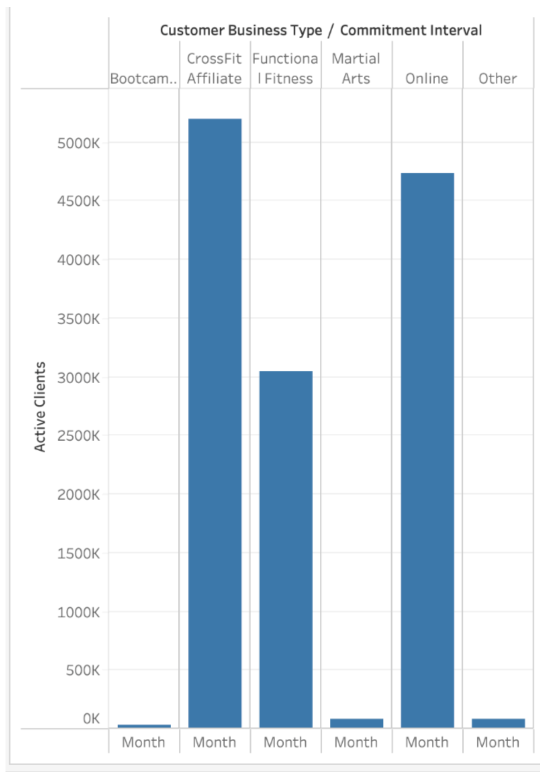
Active Client Heat Map (US)

As you can see by the heat map California, Texas, Florida, and Washington have a higher concentration of active clients within the US. This could be due to the sheer number of people living in these locations, but it can also be safe to assume that people in these states are focused on their fitness goals. In addition, despite having a lot of demographic data, California & Texas



have on average a lower median age which can contribute to the amount of people working out in their specific states in accordance with the Census Bureau.

Customer Business Type by Commitment Interval



These active clients with monthly membership based on business types Crossfit, online, and function fitness dominate in terms of active clients amongst wodify gyms due to the higher gym count of these types of business along with being in my populated/active regions in the us.

Recommendations & Conclusions

General Recommendations

Increasing Customer Retention Rate

Presence of Free Trial and Discount

The highest retention rate as shown in the previous section was among clients who had no free trial and signed up for a membership when there was a discount. The retention rate for this category is twice as high as all the other combinations of a no/free trial and no/discount. We recommend that if gyms want to increase their retention rate, that they offer a discounted membership and encourage their customers to **sign up for the discounted membership directly**.

Business Type

The highest median retention rate by business type is for Bootcamp/HIIT and Personal Training. However, these businesses are less prevalent than CrossFit Affiliate gyms by a large margin as stated above. We recommend that gyms opt for **CrossFit Affiliate, followed by Bootcamp/HIIT and Personal Training**.

Commitment Interval

Over 90% of all customers across all categories signed up for a monthly membership. We recommend that **gyms offer monthly memberships**, since they are the most popular among customers.
However, it is important to note that customers who sign up for a **yearly membership have a higher retention rate**. This could be due to the fact that gym-goers who sign up for a yearly membership are already committed to their fitness, as seen through their willingness to sign up for a year-long commitment to a gym.

Increasing Customer Lifetime Value

Presence of Free trial and Discount

The best way to increase customer lifetime value when considering free trials and discounts is to offer **free trials to customers and then encourage them to sign up for a membership.** This showed the highest median overall.

Business Type

The highest median customer lifetime value by business type was the Martial Arts business. This business type produced almost 2 times the median of all other business types. Therefore, we recommend that in order to produce the highest median customer lifetime value, Wodify clients consider integrating or creating **Martial Arts gyms**.

Regression Analysis

To delve deeper into the data provided by Wodify and uncover valuable insights for improving **customer retention and lifetime value (CLV)**, we performed regression analyses using a systematic and structured approach. The primary aim was to **identify the relationships** between key factors, such as **free trials**, **discounts**, **gym types**, and **membership plans**, and pivotal outcomes like **customer retention rates and CLV**.

The process began with data collection and cleaning. Six datasets provided by Wodify included information on memberships, free trials, and gym details. After removing duplicate LEAD_IDs and standardizing the data across all relevant files, we categorized customers based on their engagement with memberships and free trials. This led to the five primary groups mentioned in the Data Organization section.

For the analysis, we chose various regression models based on the data structure and desired outcomes. We used an **ANOVA analysis** to evaluate the impact of categorical predictors (such as regions, gym types, and membership plans) on **COMMITMENT_TOTAL**.

Anova Table

	sum_sq	df	F	PR(>F)
C(REGION)	1.035580e+06	3.0	25.265544	2.544922e-16
C(CUSTOMER_BUSINESS_TYPE)	1.521015e+06	2.0	55.663424	7.311542e-25
C(GYM_SIZE)	6.901288e+06	2.0	252.561104	1.251104e-109
Residual	4.786695e+08	35035.0	NaN	NaN

Based on the anova table that we have generated, we can conclude that **each of the predictors have significant impacts on our output (the commitment total) for each customer**. This is proven by the extremely **low p-values** for each of these predictors (REGION, CUSTOMER_BUSINESS_TYPE_, GYM_SIZE) and the **large F-values**.

After cleaning the data, we proceeded to split the data into two different parts: We split the data into **training (80%)** and **testing (20%)** sets to validate our models. We will be using the training data to form our models and the testing data set to make sure that our model predicts values accurately. The analysis process involved thorough **feature engineering** to ensure all relevant variables were captured, including newly derived features like **GYM_SIZE** (categorized into small, medium, and large). **Categorical predictors were transformed into dummy variables** for better analysis. We then proceeded to conduct a **logistic regression** that helped determine factors that influenced **binary outcomes**, such as **whether a customer signed up for a membership or not**.

The logistic regression model achieved **accuracy rates exceeding 70% in predicting free trial and discount sign-ups**, while the linear regression model explained variances in monthly revenue with reasonable accuracy. Although 70% is relatively accurate, we wanted to obtain a more precise model for our categorial results.

To improve the accuracy of our predictive models, we sought a more sophisticated approach using **polynomial predictors**. These predictors are capable of revealing intricate relationships between features, which enhances the model's ability to capture complex patterns in the data. We began by transforming **categorical predictors like REGION and CUSTOMER_BUSINESS_TYPE into dummy variables**, allowing for numerical analysis. Additionally, we grouped gyms based on **ACTIVE_CLIENTS_WITH_MEMBERSHIPS** into **three categories**: 'small,' 'medium,' and 'large,' and converted these into dummy variables to reflect their potential impact.

It was crucial to handle **missing data and to standardize the data**. We utilized the **SimpleImputer** to replace missing values in **MONTHLY_TRANSACTION_VOLUME_USD** with the mean, while **StandardScaler** normalized feature distributions, ensuring **consistent scaling and eliminating unit disparitie**s.

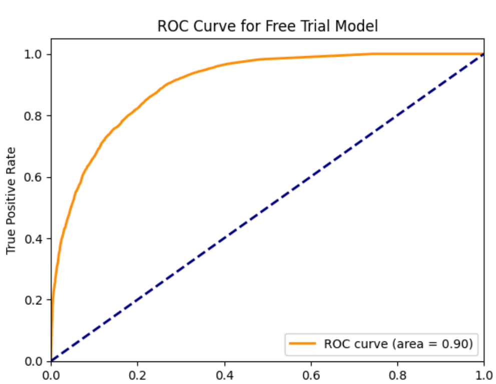
Polynomial transformations were applied to generate interaction terms and non-linear feature transformations, enabling the model to capture intricate relationships. **Principal Component Analysis (PCA)** was employed to reduce the data's dimensionality while retaining significant variability. By keeping **95% of the variance**, we reduced model complexity without losing critical information. For classification, we used a **RandomForestClassifier** due to its ability to handle non-linear data and high accuracy. This ensemble method builds multiple decision trees and averages their predictions to minimize overfitting.

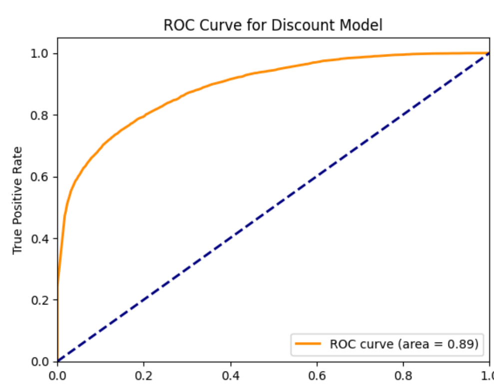
To ensure proper training and evaluation, two separate pipelines were created: **one for free trial recommendations and another for discount recommendations**. Both pipelines incorporated preprocessing, **SMOTE (Synthetic Minority Oversampling Technique) oversampling** to address class imbalance, and **RandomForestClassifier** for classification.

The results were promising from this model. The **free trial recommendation model** achieved an **accuracy of 84.1%**, while the **discount recommendation model obtained 82.2%**. The ROC curves demonstrated strong separability between positive and negative classes, with an Area **Under the Curve (AUC) of 0.90 for free trials and 0.89 for discounts**.

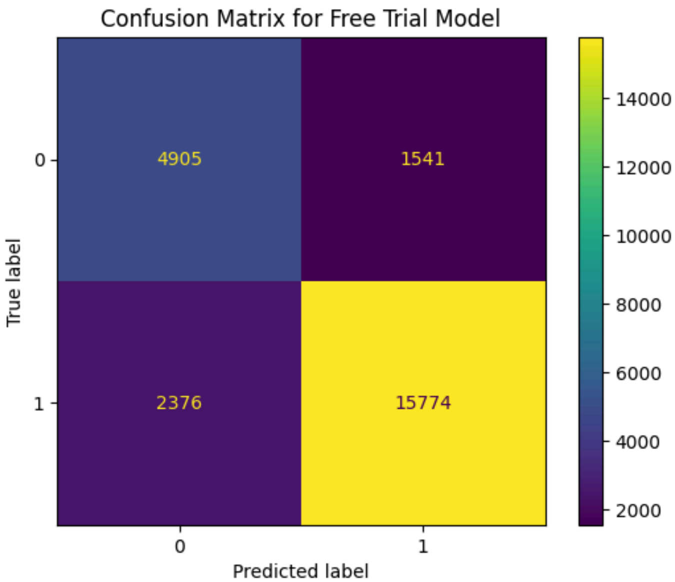
Confusion matrices revealed that the **free trial model tended to be conservative** in its predictions, with a **high count of true positives but some false negatives**. The **discount model**, however, had a higher count of **false positives**, suggesting a tendency to overpredict discounts.

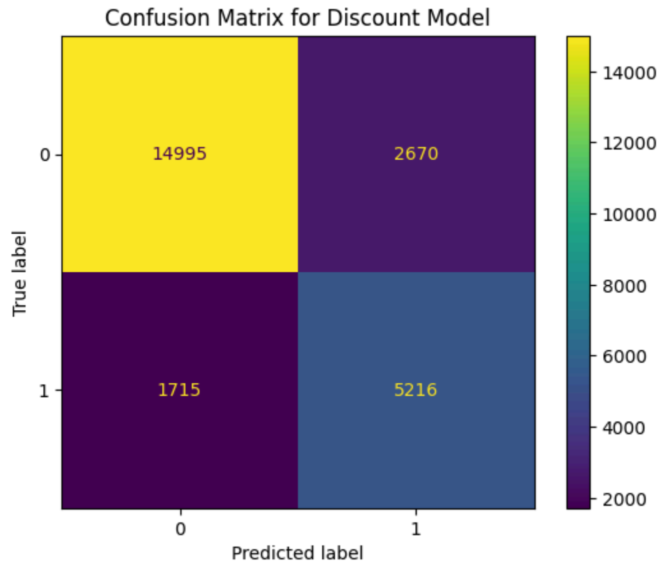
Here are the ROC curves for the RandomForestClassifier





Here are the Confusion Matrices for the RandomForestClassifier





Overall, the enhanced model using polynomial predictors and RandomForestClassifier provided robust predictions for both free trial and discount recommendations. The high accuracy and AUC scores reflect the reliability of these models that our client can use.

Next Steps

Recommendation Calculator

We hope to use the regression model in an interactive way for Wodify to use. We currently have a **front-end**, and are going to **integrate the back-end**, which is the regression model from the previous section. The goal of the application is to provide an interface for Wodify employees to input certain characteristics about gyms. It will output recommendations about free trial or discount options for memberships based on the previous data that we analyzed. We hope to get this deliverable to Wodify within a month. Wodify employees will be able to navigate the application with ease.

Appendices

Appendix A: Data Dictionary

About this data:

- Free Trials
  - One record for every free trial booked by a Lead (potential client) at a Wodify-powered business
  - This table can be joined to memberships on the Lead ID
- Memberships (4 divided csv files due to file size)
  - One record for each membership a Client (customer of a Wodify-powered business, the "gym-goer")
  - Clients can have multiple memberships
  - Each row will have a Client ID (and possibly a Lead ID)
    - If a person has both Client ID and Lead ID, they once were a Lead and converted to a Client (possibly did a free trial one point)
    - If a person only has a Client ID, they were never a Lead (and never did a free trial)

Data Dictionary:
Free Trials columns:

- Customer ID - unique identifier for business using Wodify
- Free Trial ID - unique identifier for the free trial booking
- Lead ID - unique identifier for lead profile of Wodify business
- Free Trial Booked - date free trial was booked

Memberships columns:

- Customer ID - unique identifier for business using Wodify
- Client ID - unique identifier for client of Wodify business
- Lead ID - unique identifier for lead profile of Wodify business
- Membership ID - unique identifier for a membership
- Start Date - membership start date
- Expiration Date - membership expiration date
- Membership Number - ordered membership number for each client
- Gross Commitment Total - pre discount total in USD
- Commitment Discount - discount in USD
- Commitment Total - net total in USD

- Commitment Length - quantity of commitment length
- Commitment Interval - unit of commitment length
- Membership Type - Class Plan, Class Session, or Appointment Session
- Attendances - total attendances using given membership
- Attendance Limit - if applicable, number of allowed attendances
- Attendance Limit Unit - unit of attendance limit
- Attendance Limitation - description of attendance limit

Here is additional data on each of the businesses! Included is a data dictionary for the columns that are Wodify specific.

1. Business Type - primary function of business; Bootcamp/HIIT, CrossFit Affiliate, Functional Fitness, Health Club, Martial Arts, Online, Other, Personal Training, Spin Studio, Yoga; note that Jiu-Jitsu is not a business type and these businesses would fall under Martial Arts
2. Jiu Jitsu Customer - Yes/No field that can be used to identify our Jiu-Jitsu customers
3. Hubspot Business Offerings - All of the different programs this business offers; if null this data is not in Hubspot and thus not pulled into the data
4. Customer Package - Wodify subscription cost; Legacy (old pricing billed at $2/member/month), Engage (cheapest package - $99/month), Grow (middle - $199/month), and Promote (highest - $299/month)
5. Payment Processor - who the gym uses to process their billing (all platforms are integrated through Wodify)
6. Monthly Transaction Volume - how much money the business processes through Wodify; this is a good estimate for their monthly revenue
7. Customer Value Tier - 0 to 100; how valuable the business is to Wodify; think of a Tier 99 as an Enterprise customer and Tier 12 as a smaller operation (even though all of our customers are pretty small in the scheme of things)

Appendix B: Average Retention Rate

```
# 1. Retention for each category - represented by MEMBERSHIP_COUNT in your datasets
print("Average retention for free_trials_to_memberships:", free_trials_to_memberships['MEMBERSHIP_COUNT'].mean())
print("Average retention for free_trials_to_discount:", free_trials_to_discount['MEMBERSHIP_COUNT'].mean())
print("Average retention for no_free_trial_to_memberships:", no_free_trial_to_memberships['MEMBERSHIP_COUNT'].mean())
print("Average retention for no_free_trial_to_discount:", no_free_trial_to_discount['MEMBERSHIP_COUNT'].mean(), "\n")
```

This is the calculation for the retention rates for each category discussed in the “Data Cleaning and Organizing” section.

Appendix C: Commitment Interval Distribution

```
#analyze memberhsip type for each category

print("Membership Type Distribution for free_trials_to_memberships:\n", free_trials_to_memberships['MEMBERSHIP_TYPE'].value_counts(normalize=True))

#analyze commitment interval distribution for each of the categories
print("\n Commitment Interval Distribution for free_trials_to_memberships:\n", free_trials_to_memberships['COMMITMENT_INTERVAL'].value_counts(normalize=True))
print("\n Commitment Interval Distribution for free_trials_to_discounts:\n", free_trials_to_discount['COMMITMENT_INTERVAL'].value_counts(normalize=True))
print("\n Commitment Interval Distribution for no_free_trial_to_memberships:\n", no_free_trial_to_memberships['COMMITMENT_INTERVAL'].value_counts(normalize=True))
print("\n Commitment Interval Distribution for no_free_trial_to_discount:\n", no_free_trial_to_discount['COMMITMENT_INTERVAL'].value_counts(normalize=True))
```

This is the calculation for the commitment interval distributions (what kind of plan clients preferred) for each category discussed in the "Data Cleaning and Organizing" section.

Appendix D: Gyms and their Individual Retention Rates

```
# Map customers to their gyms by using CUSTOMER_ID
# Count and average the commitment numbers to account for different numbers of occurrences for the same gym

# Category one
fttm_gym_attendance = free_trials_to_memberships.groupby('CUSTOMER_ID_x')['MEMBERSHIP_COUNT'].mean().reset_index(name='TOTAL_GYM_ATTEND')
print(fttm_gym_attendance)

# Category two
fttd_gym_attendance = free_trials_to_discount.groupby('CUSTOMER_ID_x')['MEMBERSHIP_COUNT'].mean().reset_index(name='TOTAL_GYM_ATTEND')
print(fttd_gym_attendance)

# Category four
nfttm_gym_attendance = no_free_trial_to_memberships.groupby('CUSTOMER_ID')['MEMBERSHIP_COUNT'].mean().reset_index(name='TOTAL_GYM_ATTEND')
print(nfttm_gym_attendance)

# Category five
nfttd_gym_attendance = no_free_trial_to_discount.groupby('CUSTOMER_ID')['MEMBERSHIP_COUNT'].mean().reset_index(name='TOTAL_GYM_ATTEND')
print(nfttd_gym_attendance)
```

This is the calculation for the retention rate for specific gyms.

Appendix E: Memberships by Gym Size

```
#This is for free trials to memberships
fttm_by_gym_size = free_trials_to_memberships_with_locations.groupby('Total Locations')['MEMBERSHIP_ID'].count()

print("\nFree Trials to Memberships - Memberships by Gym Size:\n", fttm_by_gym_size)

# This is for free trials to discount
fttd_by_gym_size = free_trials_to_discount_with_locations.groupby('Total Locations')['MEMBERSHIP_ID'].count()

print("\nFree Trials to Discounted Memberships - Memberships by Gym Size:\n", fttd_by_gym_size)


#This is for no free trial to memberships
nftm_by_gym_size = no_free_trial_to_memberships_with_locations.groupby('Total Locations')['MEMBERSHIP_ID'].count()

print("\nNo Free Trial to Memberships - Memberships by Gym Size:\n", nftm_by_gym_size)

#This is for no free trial to discount
nftd_by_gym_size = no_free_trial_to_discount_with_locations.groupby('Total Locations')['MEMBERSHIP_ID'].count()

print("\nNo Free Trial to Discounted Memberships - Memberships by Gym Size:\n", nftd_by_gym_size)
```

This is the calculation for the number of memberships by the gym size for each category discussed in the "Data Cleaning and Organizing" section.

Appendix F: Customer Lifetime Value

```
fttm_commitment_mean = free_trials_to_memberships['COMMITMENT_TOTAL'].describe()
plt.boxplot(free_trials_to_memberships['COMMITMENT_TOTAL'])
plt.show()
print("Commitment total summary (free trial to membership):\n", fttm_commitment_mean)

fttd_commitment_mean = free_trials_to_discount['COMMITMENT_TOTAL'].describe()
plt.boxplot(free_trials_to_discount['COMMITMENT_TOTAL'])
plt.show()
print("Commitment total summary (free trial to discount):\n", fttd_commitment_mean)

nfttm_commitment_mean = no_free_trial_to_memberships['COMMITMENT_TOTAL'].describe()
plt.boxplot(no_free_trial_to_memberships['COMMITMENT_TOTAL'])
plt.show()
print("Commitment total summary (no free trial to membership):\n", nfttm_commitment_mean)

nfttd_commitment_mean = no_free_trial_to_discount['COMMITMENT_TOTAL'].describe()
plt.boxplot(no_free_trial_to_discount['COMMITMENT_TOTAL'])
plt.show()
print("Commitment total summary (no free trial to discount):\n", nfttd_commitment_mean)
```

This is the calculation for the customer lifetime value for each category discussed in the “Data Cleaning and Organizing” section.

Appendix G: Monthly Transaction Volume

```
# free trials to memberships
fttm_avg_transaction_by_gym_size = free_trials_to_memberships_with_locations.groupby('Total Active Locations')['Monthly Transaction Volume USD'].mean()
print("Free Trials to Memberships - Average Monthly Transaction Volume by Gym Size:\n", fttm_avg_transaction_by_gym_size)

# no free trials to memberships
nftm_avg_transaction_by_gym_size = no_free_trial_to_memberships_with_locations.groupby('Total Active Locations')['Monthly Transaction Volume USD'].mean()
print("No Free Trial to Memberships - Average Monthly Transaction Volume by Gym Size:\n", nftm_avg_transaction_by_gym_size)
```

Appendix H: Monthly Retention per Gym

```
# Assuming 'MEMBERSHIP_COUNT' represents retention
fttm_avg_retention_per_gym = free_trials_to_memberships_with_locations.groupby('Customer ID')['MEMBERSHIP_COUNT'].mean()

print("Free Trials to Memberships - Average Retention per Gym:\n", fttm_avg_retention_per_gym)

free_trials_to_discount_with_locations = free_trials_to_discount_with_locations.loc[:, ~free_trials_to_discount_with_locations.columns.duplicated()]

fttd_avg_retention_per_gym = free_trials_to_discount_with_locations.groupby('Customer ID')['MEMBERSHIP_COUNT'].mean()

print("Free Trials to Discounted Memberships - Average Retention per Gym:\n", fttd_avg_retention_per_gym)

nftm_avg_retention_per_gym = no_free_trial_to_memberships_with_locations.groupby('Customer ID')['MEMBERSHIP_COUNT'].mean()

print("No Free Trial to Memberships - Average Retention per Gym:\n", nftm_avg_retention_per_gym)

nftd_avg_retention_per_gym = no_free_trial_to_discount_with_locations.groupby('Customer ID')['MEMBERSHIP_COUNT'].mean()

print("No Free Trial to Discounted Memberships - Average Retention per Gym:\n", nftd_avg_retention_per_gym)
```

Appendix I: Customer Lifetime Value by Gym Type

```
# average commitment length
# break it by gym type
# Category one: Free trial to membership
free_trials_to_memberships_with_locations['TOTAL_MEMBERSHIP_PRODUCT'] = free_trials_to_memberships_with_locations['COMMITMENT_TOTAL'] * free_trials_to_memberships_with_locations['MEMBERSHIP_COUNT']
fttm_clv = free_trials_to_memberships_with_locations.groupby('Customer Business Type')['TOTAL_MEMBERSHIP_PRODUCT'].describe()
print("Customer lifetime value (free trial to membership):\n", fttm_clv)

# Category two: Free trial to discount
free_trials_to_discount_with_locations['TOTAL_MEMBERSHIP_PRODUCT'] = free_trials_to_discount_with_locations['COMMITMENT_TOTAL'] * free_trials_to_discount_with_locations['MEMBERSHIP_COUNT']
fttd_clv = free_trials_to_discount_with_locations.groupby('Customer Business Type')['TOTAL_MEMBERSHIP_PRODUCT'].describe()
print("Customer lifetime value (free trial to discount):\n", fttd_clv)

# Category three is skipped because the customer did not sign up for a membership

# Category four: No free trial to membership
no_free_trial_to_memberships_with_locations['TOTAL_MEMBERSHIP_PRODUCT'] = no_free_trial_to_memberships_with_locations['COMMITMENT_TOTAL'] * no_free_trial_to_memberships_with_locations['MEMBERSHIP_COUNT']
nfttm_clv = no_free_trial_to_memberships_with_locations.groupby('Customer Business Type')['TOTAL_MEMBERSHIP_PRODUCT'].describe()
print("Customer lifetime value (no free trial to membership):\n", nfttm_clv)

# Category five: No free trial to membership
no_free_trial_to_discount_with_locations['TOTAL_MEMBERSHIP_PRODUCT'] = no_free_trial_to_discount_with_locations['COMMITMENT_TOTAL'] * no_free_trial_to_discount_with_locations['MEMBERSHIP_COUNT']
nfttd_clv = no_free_trial_to_discount_with_locations.groupby('Customer Business Type')['TOTAL_MEMBERSHIP_PRODUCT'].describe()
print("Customer lifetime value (no free trial to discount):\n", nfttd_clv)
```

This is the calculation for the customer lifetime value for each gym type.

Appendix J: Start Date Analysis

```
# 2. Start date analysis for each category
free_trials_to_memberships['STARTDATE'] = pd.to_datetime(free_trials_to_memberships['STARTDATE'])
free_trials_to_discount['STARTDATE'] = pd.to_datetime(free_trials_to_discount['STARTDATE'])
no_free_trial_to_memberships['STARTDATE'] = pd.to_datetime(no_free_trial_to_memberships['STARTDATE'])
no_free_trial_to_discount['STARTDATE'] = pd.to_datetime(no_free_trial_to_discount['STARTDATE'])

# Analyze start dates by year and month for trends
fttm_start_counts = free_trials_to_memberships['STARTDATE'].dt.month.value_counts().sort_index()
fttd_start_counts = free_trials_to_discount['STARTDATE'].dt.month.value_counts().sort_index()
nfttm_start_counts = no_free_trial_to_memberships['STARTDATE'].dt.month.value_counts().sort_index()
nfttd_start_counts = no_free_trial_to_discount['STARTDATE'].dt.month.value_counts().sort_index()
```

This is the calculation for the average start dates that memberships started for each category discussed in the “Data Cleaning and Organizing” section.

Appendix K: Data Cleaning and Organizing

```
# Merging the two to find the similar LEAD_ID's first instance
# Drop the repeated free trials based on LEAD_ID
free_trials_unique = free_trials_df.drop_duplicates(subset=['LEAD_ID'], keep='first')
memberships_a_unique = memberships_a.drop_duplicates(subset=['LEAD_ID'], keep='first') # Finding the first instance of LEAD_ID

# Merging the two datasets on LEAD_ID to find common entries (only the first instance though for now :P)
free_trials_to_memberships = pd.merge(free_trials_unique, memberships_a_unique, on='LEAD_ID', how='inner')
```

This is the data cleaning for category one, clients who had a free trial and signed up for a membership. We opted to keep the first instance of all unique LEAD_IDs and sum the number of occurrences of that LEAD_ID to determine the retention rate, and this is the MEMBERSHIP_COUNT.