Allison Biediger
DSC 680
Project 2-Milestone3

Optimizing Gym Offerings Through Member Data Analysis

Business Problem

Gyms are an essential business to support their clients' healthy decisions daily, however this is a competitive field where retaining current members and attracting new members is essential in their success. When gyms don't hold an understanding of gym member behavior trends and preferences they are more likely to have under attended classes, to have non essential equipment purchases and unhappy members. Through this project I aim to help gyms best optimize their offerings by recommending new classes, equipment, or strategies moving forward based on the patterns among current gym attendees. The goal of this is to create data-driven decisions that will help enhance overall member experience and help create operational efficiency.

Data Explanation

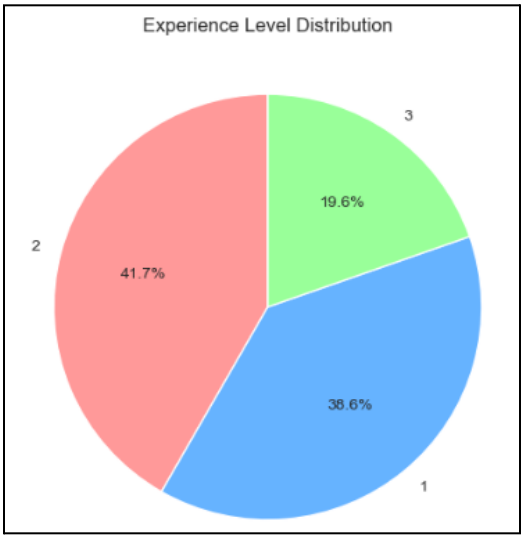
For this analysis I used the Gym Members Exercise Dataset from kaggle. This dataset contains information regarding gym members health data such as age, weight, gender, and height. As well as details regarding their workout habits such as calories burned in the workout, workout type, session duration, and workout frequency.

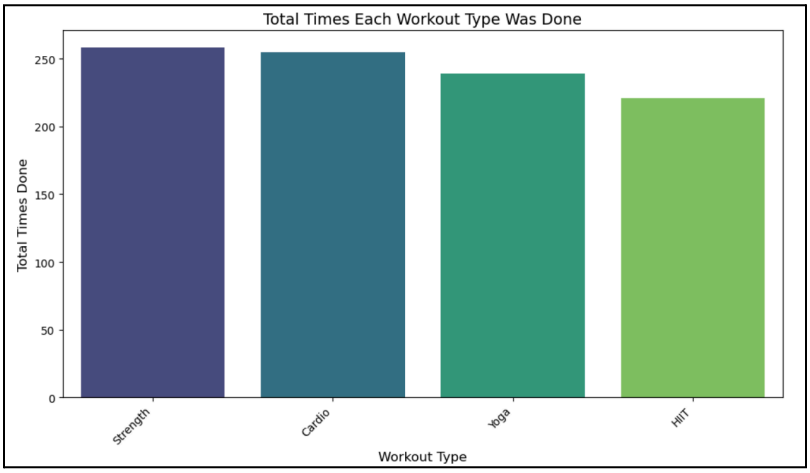
Methods

To start the analysis I focused on cleaning and understanding the dataset, checking and handling any missing variables. Next I focused on exploring the data to uncover potential trends, patterns and relationships within the data. I created visualizations to uncover patterns regarding workout type, demographic and session length. I also tried to create a model that could accurately represent the dataset. The model type that did the best was the random forest model, and I gathered further understanding of factor importance in regards to workout type decisions. Finally from these visualizations and learnings I found actionable insights for gym stakeholders.

Analysis

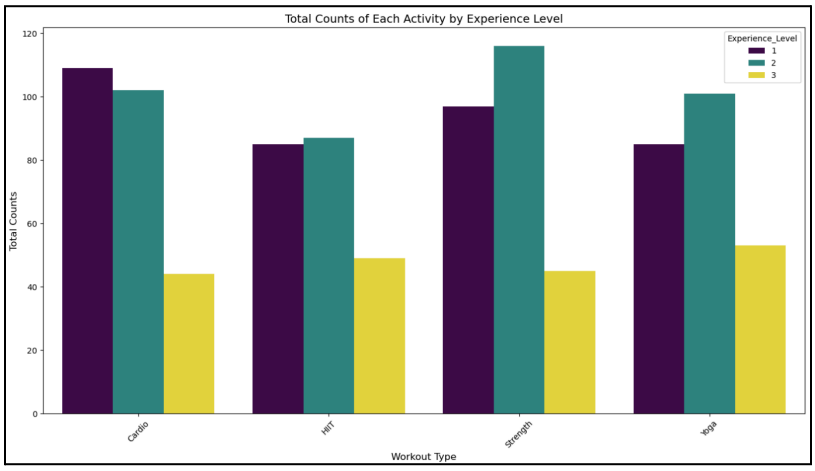
There were many important findings through the analysis of the gym member data. To begin to gain a better understanding of the member makeup in regards to workouts done and experience level.

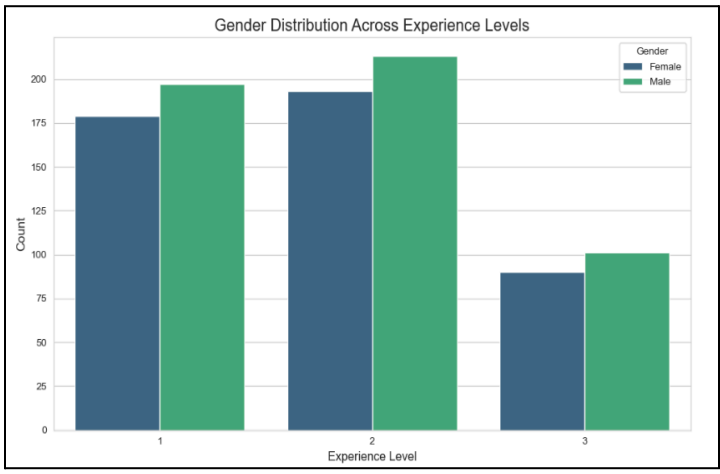




In these visualizations I found that strength workouts were the most popular overall while HIIT workouts were done the least among the members. I also found that the most number of members are considered experience level 2 and that the experience level 3 described the least amount of members.

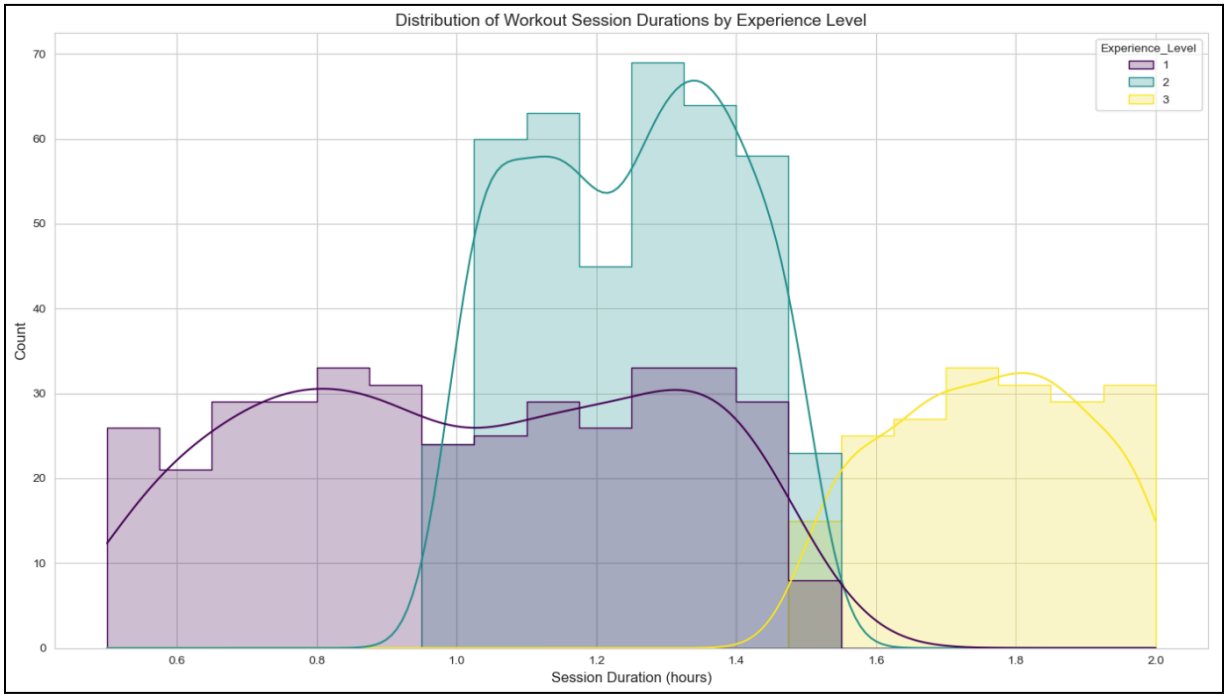
To better insights into the experience level makeup I created a couple of visualizations with workout types and experience levels and the experience levels and gender.





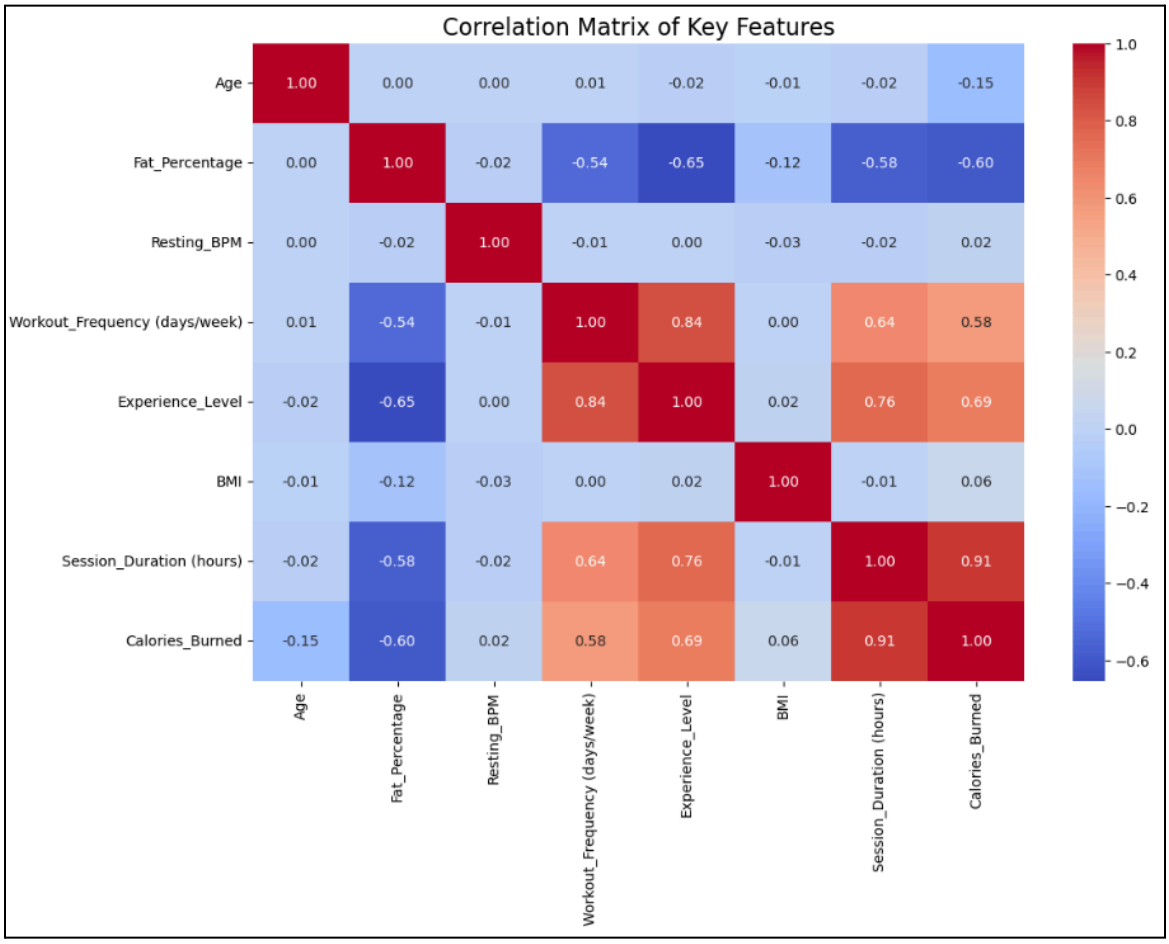
This showcased that experience level 2 members attended strength workouts the most while experience level 3 members attended the yoga workouts the most. Across all of the experience levels there were fewer women compared to men.

I then focused on the distribution of workout session duration by experience level across the different experience levels.



In this I found that experience level 2 members had sessions that typically lasted 1-1.5 hours, and experience level 3 had sessions that lasted 1.5-2 hours. There was a clear progression in session length as the experience level increased.

I then focused on analyzing the correlations between session duration and workout frequency.



I found that experience level had the strongest impact on workout frequency (0.84). I also found the workout frequency correlated highly with calories burned (0.91) and experience level (0.76).

Finally I attempted a few different models to predict workout type however each method was not successful nor accurate. This is likely due to lack of other important variables at play.

Conclusion

In conclusion strength workouts and experience level 2 dominated the member activity. There was a high correlation between session duration and experience level. Across all of the different experience levels had consistently less women. With these findings there is an opportunity to cater to under represented groups and less popular activities.

Assumptions

There were many assumptions at play through the analysis including that the data set was an accurate representation for the gym's members. In addition to this it was assumed that the members were accurate and consistent in recording their workouts, and that workout habits remained consistent overtime.

Limitations

Among the limitations is the potential for bias in self reporting, as well as the missing insights into other qualitative factors (such as member satisfaction, motivation and date/time). In addition the lack of further insight into the current classes marketing and offerings also limited deeper understandings in this analysis.

Challenges

This analysis had many difficulties and challenges including any data quality issues. As this is not a dataset that I collected or have the ability to ask clarifying questions about, it can be challenging to manage any inconsistent or missing values. In addition to this another challenge will arise with potential changing preferences of members. These can include potential changes in personal interest or seasonal trends. By anticipating these potential challenges I was able to better address them as they arose.

Future Uses

In the future we can gather further data with qualitative insights that were lacking from this analysis (such as surveys on client preferences). In addition tracking long-term trends in client progression and retention would add further insights and predictability for future trends and needs.

Recommendations

From this analysis I recommend creating more strength focused classes and yoga classes. As well as designing programs that target the experience level 2 members, and to conduct outreach to attract more experience level 3 members and women.

Ethical Considerations

In this analysis there are many ethical considerations at play including data privacy, avoiding bias in health and workout data, and data accuracy. To start since this data set is representative of active members of the gym it is essential to keep the data confidential. In addition to this it is important to avoid any bias in regards to preconceived notions about weight, BMI etc. This also applies to favoring specific demographics with biased recommendations. Lastly I will focus on avoiding any misinterpretations of trends to ensure both fair and useful decisions for all members.

Implementation Plan

To implement these recommendations the gym should continue to use a data driven program approach. They should use insights to create a targeted new class schedule and workout plans. They should also begin targeted marketing campaigns focusing on women and on experience level 3 members. In addition the gym should use these insights to manage their resource allocation , adding more equipment and trainers for strength and yoga classes. Finally they should do their best to encourage a deeper member feedback loop, gathering further information through surveys and other feedback opportunities.

Ethical Assessment

In order to maintain trust and to ensure ethical use of member data there are many factors to keep in mind. To begin, both the data privacy and the data security are essential in maintaining the members' trust. In order to successfully manage this concern all member data needs to be securely stored and anonymized. Next it is essential to be transparent and clear in communication on how the member data is used to inform decisions. Finally gaining clear and informed consent is necessary before both data collection and data analysis.

Potential Questions:

1) *How did you decide what features to focus on in this analysis?*
    a) *For this analysis I decided to focus on features that a typical client of a gym would already fit between to start. This was to best navigate the offerings that would best suit each group and adjust current programing. I also decided to focus on features that would directly relate to programming such as time and experience level to help streamline these classes to best fit the population.*
2) *Why do you think strength training is the most popular workout type?*
    a) *This could be due to many factors, strength training is great at building muscle which can make other workout options easier or more manageable. This is also a type of workout that is easily tailored to the individual, making it easier for the client to manage alone and feel successful.*
3) *What would you recommend in order to address the lower HIIT participation?*
    a) *I would recommend approaching the advertising of this class. By focusing on the advertising you can incorporate both internal and external advertising pointed for groups less represented at the class. Externally creating advertisements that appeal to experience level 3 individuals may also help address this gap.*
    b) *In addition I would recommend trying out different length sections. Seen in the visualizations above, clients in experience level 1 and 2, which are the majority of HIIT attendees, have a good overlap when it comes to time of workout session. By doing this they will be able to gather data as the most popular time lengths for their clients and optimize this class.*

4) *How will the gym be able to reach out to women specifically?*
    a) *By creating women focused classes the gym would better reach this group. One option for this could be creating a womens only class to help clients that may be more uncomfortable in a gym setting. Focusing on inclusion and safety in advertisements would also help reach out to women specifically.*
5) *How can the gym use these insights to market new classes?*
    a) *There are many ways these insights can be used, when creating new classes the insights into their current clients will help direct the decisions made. For example making classes that best fit with experience level 1 and 2 since that is the majority of their client base.*
    b) *In addition these findings highlight opportunity zones of where the gym can grow, for example there are fewer women on every experience level as well as a much lower number of clients that consider themselves an experience level 3.*
6) *What other data points would help deepen this analysis?*
    a) *Further data in regards to current equipment use would help give further insight in a typical workout for the clients and help navigate any gaps that may exist. In addition, data on the frequency each member attends a class or visits the gym would help give further insights in maintaining highly engaged clients and opportunities for those that visit or join less frequently.*
7) *How can the gym retain members with low attendance or short session durations?*
    a) *To help maintain members that have a low attendance rate or short session durations the gym can consider incentives to attend. This could include prizes or exclusive access to classes, this could help encourage further attendance if it increases the likelihood for winning. In addition, making and maintaining consistent classes that fall within the shorter session duration allotted time would be a good way to maintain those that do not have a lot of time to set aside for the gym.*
8) *What is the next step for implementation?*
    a) *Next these data driven insights will be used to expand marketing to women and experience level 3 individuals. The gym can also begin updating schedules and offering classes to best fit the findings and adding more equipment and trainers for strength and yoga classes. The gym can continue to gather feedback through surveys and different feedback opportunities.*
9) *Could this analysis predict peak workout times?*
    a) *This analysis is not able to predict peak workout times as this was not a measure of data that was measured. However if this is desired the gym can keep a log data set of clients scanning in or out to focus on daily time trends.*
10) *Could these insights be used for changes in a chain of gyms instead of a singular location?*
    a) *Yes, these can be used for a chain of gyms to address similar concerns however there are some things to keep in mind if this is how it is used. There may be differences in behavior based on location, specific clients and specific offerings. The best way to have direct and optimized findings would be to focus on individual locations.*