Tracking Strength Progress Through Gym Consistency and Cross-Training Activities: A Self-Logged Weekly Analysis

Erica O. Galindo
College of Computing and Information Technologies
National University Philippines
Manila, Philippines
galindoeol@students.national-u.edu.ph

Abstract—This study analyzes a self-logged fitness routine to understand what drives a high training week and how strength changes over time. I tracked gym sessions and cross-training activities such as jogging, pickleball, basketball, walking, and swimming, recording minutes per session and basic lifting details when available. The data was aggregated at the weekly level to compute total training minutes, gym and cross-training minutes, session counts, and activity variety. I created visualizations to examine weekly training patterns and estimated strength progression using the Epley estimated one-repetition maximum (e1RM) for selected lower-body and whole-body exercises. I also ran a Pearson correlation test to assess the relationship between weekly gym sessions and total weekly training minutes. To model weekly training behavior, I trained two simple and interpretable machine learning models, Logistic Regression and a linear Support Vector Machine (SVM), to predict whether a week is a high training week, defined as having total minutes greater than or equal to the median. The results show that both gym minutes and cross-training minutes are strongly associated with high training weeks, while activity variety has a smaller influence. Overall, this project illustrates how personal self-tracking combined with basic data science methods can be used to quantify training patterns, monitor strength trends, and support better planning for fitness routines.

Index Terms—self-tracking, fitness routine, strength progress, gym consistency, cross-training, Logistic Regression, Support Vector Machine, Pearson correlation

I. INTRODUCTION

Personal fitness progress is not only about performing hard workouts. A large part of sustainable improvement depends on consistency, recovery, and the total amount of training accumulated over time. Many people feel that they are training "a lot," but without systematic tracking it is difficult to understand what is really happening. This study focuses on personal fitness tracking using self-logged daily data to understand training patterns and how gym consistency and cross-training activities relate to overall training volume and strength progress.

Tracking fitness habits is important because it supports more informed decision-making. If total training volume becomes too low, progress may slow down. If training becomes too high or irregular, it may lead to fatigue and poor recovery. By organizing and analyzing data, it becomes possible to identify patterns that are not obvious from memory alone. Instead of relying on subjective impressions, self-tracking enables an individual to measure routines, compare weeks, and observe gradual changes over time.

Prior self-tracking and quantified-self studies have shown that personal logs and wearable data can be used to understand behavioral patterns and outcomes such as consistency, performance, and lifestyle habits. Existing work has also demonstrated that simple and interpretable machine learning models, such as Logistic Regression and Support Vector Machines (SVMs), can help identify which factors are most influential by examining model coefficients. These studies support the idea that personal behavior data, when organized properly, can be analyzed using basic data science tools to reveal meaningful trends.

However, many previous studies rely on commercial wearable devices, mobile applications, or large population-level datasets, and fewer focus on simple self-logged fitness routines that combine structured gym training with multiple cross-training activities such as jogging, recreational sports, and walking. Because training can vary substantially from day to day, it can also be more informative to summarize daily logs into weekly aggregates when examining trends over time. In this project, a manually self-logged dataset of gym sessions and cross-training activities (including jogging, pickleball, basketball, walking, and swimming) is converted into weekly summaries to enable clearer comparison across weeks.

The goal of this project is to organize my personal workout logs, describe training trends over several weeks, and test relationships between gym consistency and overall training volume using statistical analysis. In addition, selected strength exercises are tracked using an estimated one-repetition maximum (e1RM) formula to observe changes in strength over time. Finally, simple and interpretable machine learning models Logistic Regression and a linear SVM are trained to predict whether a given week is a high training week, defined by total weekly minutes, and to identify which training factors are most strongly associated with these high training weeks.

II. LITERATURE REVIEW

A. What previous studies examined

Past studies on self-tracking, also called the quantified-self, examine how people record daily behaviors to understand patterns and improve habits. Lupton explains that self-tracking is often used to make routines more visible and measurable, especially for health and lifestyle behaviors such as exercise and activity levels [1]. Many studies in this area focus on physical activity tracking using simple metrics such as steps, movement time, and workout frequency because these are easy to log and compare over time [2]–[4].

Other research examines how tracking tools and fitness devices affect motivation and behavior. These studies investigate whether tracking helps people become more active and whether it supports long-term habit change [4], [5], [9]. In strength and resistance training, prior work focuses on how strength improves through training progression and how to measure progress safely using estimates rather than testing maximal lifts frequently [6]–[8].

B. What data collection and analysis methods they used

Most self-tracking studies use wearable devices (such as step counters and fitness watches) or phone-based applications to collect data automatically, typically measuring steps per day or total activity duration [3], [4]. Some studies also include self-monitoring or manual logging, especially when the behavior is more specific than steps, such as gym sessions, types of training, or workout structure [2], [9].

For analysis, many studies use basic descriptive statistics and hypothesis tests to examine relationships and trends, and some apply simple models to interpret which factors matter most. For example, structured behavior datasets can be analyzed using interpretable models where the influence of each variable can be inspected through model coefficients or feature weights [10]. In resistance training research, strength progress is commonly evaluated using training progression concepts and estimation formulas that use the load lifted and repetitions performed to estimate strength without requiring one-repetition maximum (1RM) testing in every session [6]–[8].

C. Main findings from previous work

A consistent finding is that tracking can help people increase awareness of their routines and support behavior change, especially for physical activity [2], [3]. Reviews and meta-analyses show that activity trackers can increase physical activity in many cases, although effects vary depending on the individual and on how consistently they engage with the tracking process [4], [9]. A randomized controlled trial also suggests that benefits from self-tracking devices are not guaranteed for everyone, and that real impact depends on user engagement and sustained use [5].

In strength-related work, training progression and consistency are linked with improvements over time when resistance training is performed in an organized manner [6]. Studies on strength estimation show that 1RM prediction equations can be useful for monitoring progress, but they still introduce error, and estimates can be affected by fatigue, technique, exercise type, and day-to-day variability [7], [8]. This supports using estimated measures as practical indicators of strength, while recognizing that they are approximations rather than exact measurements.

D. Limitations mentioned in prior work

A common limitation across self-tracking studies is data quality. Even when devices are used, there can be missing data (e.g., the device is not worn, batteries run out), and for manual logging, missing entries can occur because recording depends on time, memory, and effort [1], [2], [5], [9]. Another limitation is that self-tracking results are often highly personal. What works for one person may not work the same way for others, so findings may not generalize well to larger populations [1], [5].

In strength estimation, another limitation is that estimated 1RM or similar metrics do not perfectly represent true strength. Accuracy can vary depending on the specific formula, the exercise, and the individual [7], [8]. As a result, researchers typically treat estimated strength as a useful proxy rather than a definitive measurement.

E. How this project is similar or different

This project is similar to prior work in that it follows the same self-tracking idea: collect personal behavior data, summarize it over time, and analyze patterns using descriptive statistics, statistical tests, and simple interpretable models [1], [2], [9]. It also connects to physical activity tracking research because it studies training time and frequency as key variables related to behavior and consistency [3], [4].

At the same time, it differs from many prior studies because the dataset is fully self-logged and centered on a mixed fitness routine that combines structured gym training with multiple cross-training activities (jogging, pickleball, basketball, walking, and swimming). Instead of relying on wearable sensors, this project uses a structured daily log and then aggregates it at the weekly level to enable clearer comparisons across weeks. Finally, the project applies interpretable machine learning models to identify which training factors are most strongly related to a high training week, defined by total weekly training minutes, by comparing feature weights across Logistic Regression and linear SVM models, similar to how behavior-related features are interpreted using model coefficients in related work [10].

III. METHODOLOGY

A. Participants

I am the only participant in this study (n = 1). I am a college student who regularly engages in fitness activities, including gym training and sports-based cross-training. I did not collect or report any sensitive personal information.

B. Data Collection Methods

I collected my fitness data through daily self-logging. Each log entry represents one activity session done on a specific date. The activities include gym sessions and cross-training activities such as jogging, walking, pickleball, basketball, and swimming.

I initially recorded the logs manually, then transferred them into Google Sheets and exported the file as a CSV for analysis. The dataset covers multiple weeks, starting from November 18, 2025, up to January 17, 2026. I logged data only on days when I had a workout or activity; days without activity simply do not appear in the dataset.

C. Operational Definitions

I tracked the following variables in my daily dataset:

- date – the calendar date when I did the activity session
- session_id – a unique ID assigned per session to avoid duplicates and to track sessions clearly
- category – the general type of session (gym or cross_training)
- activity_type – the specific activity performed (gym, jog, walk, pickleball, basketball, swimming)

- minutes – the duration of the session in minutes; most gym sessions were 60 minutes because I trained with a coach
- distance_km – the distance in kilometers when applicable (primarily for jogging)
- notes – short text describing session details such as workout structure, exercises, loads, or training focus

For analysis, I converted the daily logs into weekly summaries using the ISO calendar week (year–week) based on the date. For each week, I derived the following variables:

- total_sessions – total number of sessions in that week
- total_minutes – total training minutes in that week
- gym_sessions – number of gym sessions in that week
- gym_minutes – total gym minutes in that week
- cross_sessions – number of cross-training sessions in that week
- cross_minutes – total cross-training minutes in that week
- activity_variety – number of unique activity types in that week

For modeling, I also created a binary label:

- high_training_week – a week labeled 1 if its total_minutes was greater than or equal to the median weekly total_minutes, and 0 if it was below the median.

In a separate compact dataset, I also tracked selected strength-related exercises (e.g., barbell deadlift, barbell Romanian deadlift, and leg extension) with the following variables:

- date – date of the gym session
- exercise – exercise name
- weight_kg_total – total load in kilograms used for the exercise set
- reps – repetitions performed per set
- sets – number of sets performed

From this strength dataset, I computed an estimated one-repetition maximum (e1RM) per exercise using the Epley equation:

$$e1RM = \text{weight\_kg\_total} \times \left(1 + \frac{\text{reps}}{30}\right)$$

This allowed me to monitor strength trends over time for key lower-body and whole-body movements.

D. Data Cleaning and Preprocessing

I imported the CSV file into Python using the pandas library and converted the date column to datetime format. I also converted minutes and distance_km to numeric types. I inspected missing values for each column. Since distance is only relevant for some activities (mainly jogging), missing values in distance_km were treated as “not applicable” rather than errors.

I removed any rows with missing session_id, since these represented incomplete or invalid entries. Then, using the date column, I extracted the ISO calendar year and week number to create year and week variables. The daily sessions were then grouped by (year, week) to compute the weekly summary variables described above. Weekly total minutes and category-specific minutes were filled with zero when needed to ensure numeric consistency.

For machine learning, I selected the following weekly features:

- gym_sessions
- gym_minutes
- cross_sessions
- cross_minutes
- activity_variety

Before training the models, I standardized these numeric features using z-score scaling (StandardScaler) so that variables were on a comparable scale. The data was then split into training and test sets using a stratified train–test split.

E. Statistical Analysis and Visualization

I used both descriptive analysis and simple statistical and modeling techniques.

1. Descriptive statistics and visualizations:
   I summarized weekly training behavior using descriptive statistics (mean, median, standard deviation, minimum, and maximum) for the main weekly variables. I also created visualizations, including:
- A line chart of weekly total training minutes
- A line chart comparing weekly gym minutes and cross-training minutes
- A histogram showing the distribution of total weekly minutes
- A correlation matrix heatmap for the weekly variables

These visualizations allowed me to examine trends, distributions, and relationships in training volume and activity composition over time.

1. Pearson correlation:
   I used Pearson correlation to test the relationship between **gym_sessions** and **total_minutes** per week, to determine whether gym consistency is associated with overall weekly training time. I reported the Pearson correlation coefficient **r** and the corresponding p-value. Pearson correlation was selected because both variables are numeric and the goal was to assess a linear relationship.

2. Hypotheses:

   H0 (Null hypothesis): There is no significant linear relationship between gym_sessions and total_minutes per week.

   H1 (Alternative hypothesis): There is a significant linear relationship between gym_sessions and total_minutes per week.

3. Machine learning models (interpretable):
   I trained two simple and interpretable binary classification models to predict whether a week is a

high training week:

- Logistic Regression
- Linear Support Vector Machine (linear SVM)

The input features were the weekly variables listed above, and the output label was high_training_week. I used a train–test split with stratification, standardized the inputs, and then trained each model. I reported model performance using accuracy and a classification report (precision, recall, and F1-score).

To compare feature importance across the two models, I extracted the feature weights (coefficients) from both Logistic Regression and the linear SVM. I then created a combined table and computed the average weight per feature across the two models as a simple “mutual agreement” score, indicating which factors were consistently influential in predicting high training weeks.

4. Strength progression analysis:
   For the strength dataset, I computed e1RM values using the Epley equation for selected exercises and plotted e1RM over time per exercise. This provided a visual representation of estimated strength progress across the study period.

F. Bias and Measurement Considerations

Because this is self-logged data, it may include logging bias such as missing entries, uneven detail in notes, or occasional delays between doing the activity and recording it. In addition, many of my gym sessions were fixed at approximately 60 minutes, which reduces variation in gym_minutes and can make gym_sessions and gym_minutes closely related.

The strength estimates are based on an indirect formula (e1RM) rather than true maximal testing and are influenced by factors such as fatigue, technique, and exercise form. Finally, this dataset reflects only my own routine over a limited period, so the findings are intended to describe my personal patterns and are not meant to be generalized to other individuals or populations.

IV. RESULTS

A. Descriptive Statistics of Weekly Training

The weekly training variables are summarized in Table 1.

Table 1. Summary statistics of weekly training variables (mean, median, SD).

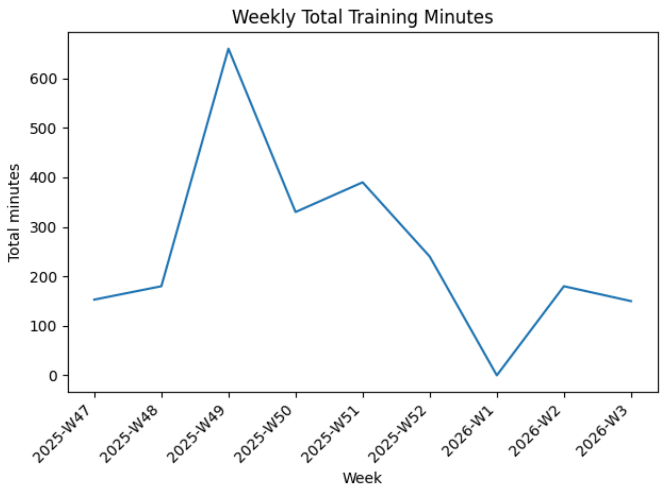
Variable	Mean	Median	SD
Total sessions	3.33	3.0	1.41
Total minutes	253.67	180.0	189.01
Gym sessions	1.67	2.0	1.12
Gym minutes	100.00	120.0	67.08
Cross-training sessions	1.67	1.0	1.41
Cross-training minutes	153.67	33.0	201.27
Activity variety	2.11	2.0	1.05

On average, there were about 3.33 training sessions and 253.67 minutes of training per week, but the large SD for total minutes (189.01) indicates substantial variability between weeks. Gym sessions averaged 1.67 per week with 100 minutes, while cross-training sessions also averaged 1.67 per week but with a higher mean of 153.67 minutes. This suggests that cross-training often contributes more to total training volume than gym sessions, and that weekly training patterns are not uniform across time.

B. Time-Series Trends in Training Volume

Weekly total training minutes over the study period are shown in Fig. 1.

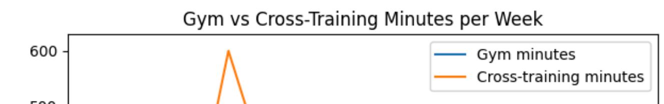
Fig. 1. Weekly total training minutes across the study period.



Total training minutes increase from late November (around Week 47) to a peak near Week 49, where weekly volume reaches roughly 660 minutes. After this peak, total minutes drop sharply, reaching a low level around Week 52 and Week 1 of 2026, and then rise again in Weeks 2 and 3 of 2026. This pattern indicates alternating periods of high and low training load rather than a steady, continuous increase.

Weekly gym minutes and cross-training minutes are compared in Fig. 2.

Fig. 2. Weekly gym minutes and cross-training minutes across the study period.

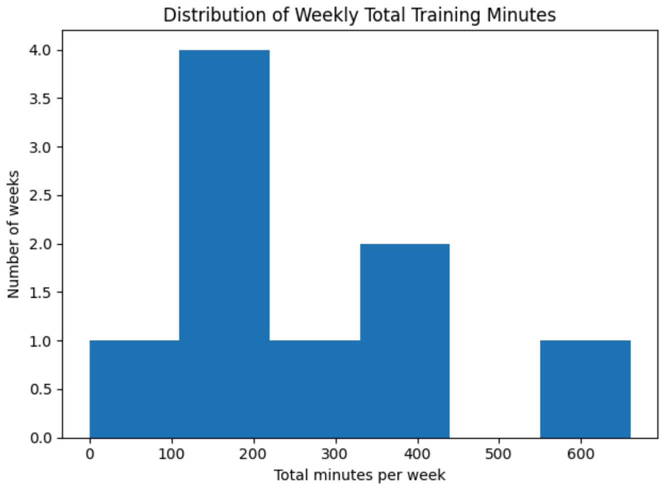


Gym minutes remain within a relatively narrow range, reflecting the mostly fixed 60-minute duration of gym sessions. Cross-training minutes fluctuate much more, with a large spike around Week 48 driven by long pickleball sessions and several weeks with near-zero cross-training in early 2026. These trends show that week-to-week variation in total training volume is mainly driven by changes in cross-training duration, while gym training provides a more stable baseline.

C. Distribution of Weekly Training Minutes

The distribution of total weekly training minutes is shown in Fig. 3.

Fig. 3. Distribution of weekly total training minutes.

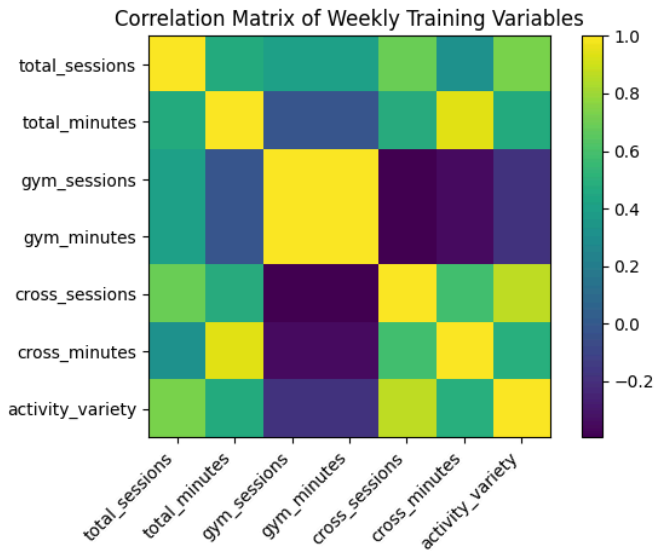


Most weeks fall between roughly 100 and 250 minutes of training, with only a few weeks exceeding 300–400 minutes. This skewed distribution indicates that a small number of high-volume weeks account for much of the total training time, while typical weeks are more moderate.

D. Correlations Among Weekly Training Variables and Statistical Test

Pairwise correlations among the weekly variables are visualized in Fig. 4.

Fig. 4. Correlation matrix of weekly training variables.



Total weekly minutes are most strongly aligned with cross-training minutes, suggesting that high-volume weeks are often driven by longer cross-training sessions such as pickleball or jogging. Gym sessions and gym minutes are strongly correlated with each other, which is expected given the relatively fixed structure of gym workouts. Activity variety shows positive correlations with cross-training sessions and minutes, indicating that weeks with more different activity types also tend to include more cross-training.

A Pearson correlation test was used to examine the linear relationship between gym sessions per week and total weekly training minutes. The results are presented in Table 2.

Table 2. Pearson correlation between gym sessions per week and total training minutes per week.

Test	**Variable** 1	**Variable** 2	**r**	**p_value**
Pearson correlation	gym_sessions	total_minutes	-0.011	0.977

The correlation coefficient is approximately r=−0.011 with a p-value of 0.977, indicating essentially no linear relationship between the number of gym sessions in a week and total weekly training minutes. Statistically, this effect size is effectively zero and not significant, confirming that total training volume is not determined by gym frequency alone and that cross-training plays a major role.

E. Classification of High-Training Weeks and Feature Importance

Weekly training patterns were modeled using Logistic Regression and a linear Support Vector Machine (SVM) to predict whether a week was a high training week, defined as having total minutes greater than or equal to the median. The feature weights from both models and their mutual average are given in Table 3.

Table 3. Feature weights from Logistic Regression and SVM, and mutual average weight for identifying high training weeks.

feature	LR_weight	SVM_weight	Mutual_avg _weight
cross_minutes	0.663	0.869	0.766
gym_sessions	0.530	0.642	0.586
gym_minutes	0.530	0.642	0.586
cross_sessions	0.104	0.132	0.118
activity_variety	0.071	-0.071	0.000

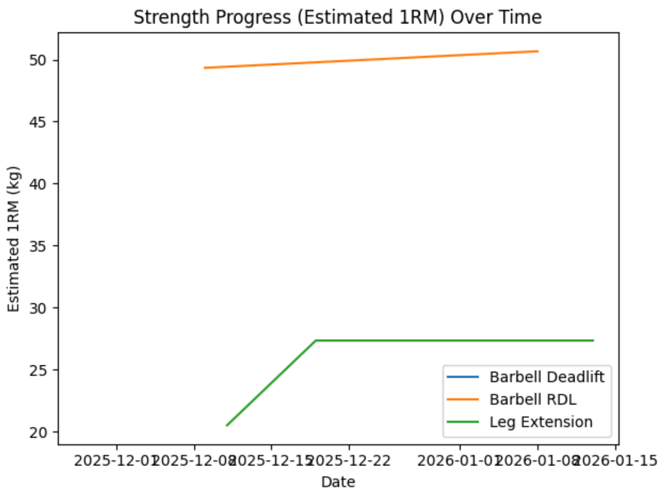
Across both models, cross_minutes has the largest positive mutual average weight (0.766), indicating that weeks with more cross-training minutes are the most likely to be classified as high training weeks. gym_sessions and gym_minutes also have relatively large positive mutual average weights (0.586), showing that higher gym involvement contributes meaningfully to high training weeks, though slightly less than cross-training minutes.

cross_sessions has a smaller but still positive mutual average weight (0.118), suggesting that total cross-training duration is more important than the number of cross-training sessions alone. activity_variety has a mutual average weight of 0.000 because Logistic Regression and SVM assigned weights with opposite signs, implying that activity variety does not have a consistent effect on whether a week is high volume in this dataset.

F. Strength Progress Based on Estimated 1RM

Strength progression was evaluated using the Epley estimated one-repetition maximum (e1RM) for selected lower-body exercises. The e1RM trends for barbell Romanian deadlift (RDL) and leg extension are shown in Fig. 5.

Fig. 5. Strength progress (estimated 1RM) for barbell RDL and leg extension over time.



Barbell RDL e1RM increases slightly from about 49 kg in early December to just over 51 kg in early January, indicating a small but steady improvement in hip-hinge pulling strength during the study period. Leg extension e1RM rises more clearly, from approximately 20.5 kg in mid‑December to about 27.3 kg later in December and mid‑January, suggesting a noticeable gain in quadriceps strength. The barbell deadlift appears only once and therefore does not show a trend. Although the strength dataset is small, the upward e1RM changes for RDL and leg extension are consistent with gradual strength improvements alongside regular gym training.

G. Summary of Key Findings

1. What does the dataset look like overall?
 Several weeks of self‑logged gym and cross‑training data show an average of about 3–4 sessions and 250 minutes of training per week, with substantial variability in total minutes (Table 1, Fig. 3).
2. What patterns or trends appear in the descriptive statistics?
 Training volume peaks in late 2025 and dips in early 2026, with cross‑training spikes creating the highest‑volume weeks (Figs. 1–2).
3. What relationships were found between variables?
 Total minutes are most closely related to cross‑training minutes, and activity variety is positively related to cross‑training involvement (Fig. 4).
4. Did any statistical tests show significant results?
 The Pearson correlation between gym_sessions and total_minutes was not significant (Table 2), with $r \approx -0.011$ and $p = 0.977$ indicating no meaningful linear effect.
5. What graphs best illustrate the findings?
 Time‑series plots highlight weekly changes in training volume and composition (Figs. 1–2); the histogram emphasizes that most weeks are moderate in volume (Fig. 3); the correlation matrix visualizes relationships among training variables (Fig. 4); and the e1RM plot illustrates strength progress for key exercises (Fig. 5).

V. DISCUSSION

This section interprets the results in the context of my personal training behavior and relates them to previous self-tracking and physical activity research. It also describes the main limitations of the study and suggests directions for future work.

A. Interpretation of Results

The descriptive statistics and histograms show that my weekly training activity is moderately consistent overall, but with notable variability. Most weeks fall between approximately 100 and 250 minutes of total training, indicating a generally regular routine. However, a few weeks reach very high volumes above 300–400 minutes, mainly driven by longer cross-training sessions such as extended pickleball play.

The time-series line charts highlight that gym minutes remain relatively stable across weeks, while cross-training minutes fluctuate sharply. This is likely because gym sessions are planned, scheduled, and typically fixed at 60 minutes, whereas cross-training depends more on external factors such as free time, social opportunities, or energy level on a given day.

The Pearson correlation analysis showed that the number of gym sessions per week is essentially unrelated to total weekly training minutes (approximately $r = -0.011$, $p = 0.977$).

This means that simply going to the gym more often does not automatically translate into higher total training volume in this dataset. Instead, the correlation matrix and feature weights from the classification models indicate that cross-training minutes and gym minutes together are more important drivers of high-volume weeks.

The correlation matrix reveals that total minutes are most strongly associated with cross-training minutes, while gym sessions and gym minutes are very highly related to each other due to the fixed session length. Activity variety is positively related to cross-training sessions and minutes, suggesting that weeks with more different activity types tend to be weeks when I do more cross-training.

The classification models add another perspective: both Logistic Regression and the linear SVM assign the largest positive weight to cross-training minutes, followed by gym sessions and gym minutes. Activity variety receives near-zero mutual weight, indicating that in this small dataset, how many different activities I do matters less than how long I actually spend training. Taken together, these findings suggest that my total training volume is driven more by a combination of cross-training duration and consistent gym time than by gym frequency alone.

The strength analysis based on estimated 1RM (e1RM) shows small but meaningful improvements in lower-body strength. Barbell RDL e1RM increases slightly over the study period, while leg extension e1RM rises more clearly from mid-December to mid-January. Even though there are only a few strength data points, the upward trend is consistent with the idea that relatively regular resistance training can support gradual strength gains.

B. Comparison to Related Work

These patterns align with previous self-tracking and quantified-self research showing that personal schedules, context, and motivation shape physical activity levels [1]–[3]. Similar to prior studies that used self-logged or smartphone-based data to reveal daily and weekly activity patterns, my data shows clear fluctuations across weeks rather than a constant level of training. The finding that cross-training contributes substantially to total volume is also consistent with work emphasizing that different activity modes (e.g., sports, walking, informal exercise) can add up to significant overall activity time [2], [3].

Research on self-tracking tools and activity monitors has reported that logging behavior can increase awareness and sometimes support behavior change, although effects vary by individual [4], [5], [9]. In my case, organizing the data and seeing the visualizations helped me recognize how a few long cross-training sessions can dominate some weeks, and how my gym routine remains more stable. This type of insight is similar to what prior studies describe: self-tracking does not automatically make behavior perfect, but it makes patterns more visible and easier to evaluate.

The use of simple, interpretable models such as Logistic Regression and SVM to inspect feature weights follows approaches in prior behavior and health analytics research, where model coefficients are used to identify which variables are most influential [10]. My results are consistent with that idea: the models clearly highlight cross-training minutes and gym involvement as the main predictors of high training weeks.

At the same time, this project differs from many large-scale or device-based studies because it uses fully self-logged data from a single participant and focuses on a mixed routine that includes gym training and multiple cross-training activities. The findings are therefore highly personal and context-specific, but they illustrate how even a small, manually collected dataset can be analyzed with the same types of methods used in larger studies.

C. Limitations

This study has several important limitations. First, the sample size is very small ($n = 1$) and all data come from my own training. As a result, the findings cannot be generalized to other people, and any patterns observed may reflect my personal preferences, schedule, and fitness level.

Second, the data are self-reported, which introduces potential logging bias. Some entries might be incomplete or slightly inaccurate because they were recorded manually and sometimes after the sessions. While I tried to be consistent, recall errors and uneven detail in the notes are possible.

Third, the dataset contains some missing or non-applicable values, especially in distance-related fields. Although these were handled during data cleaning and are not central to the main analyses, they still represent a limitation in data quality. In addition, many gym sessions have a fixed 60-minute duration, which reduces variation in gym_minutes and makes some variables highly correlated by design.

Fourth, the data collection period is relatively short, spanning only several weeks across late 2025 and early 2026. The limited time window makes it difficult to assess longer-term trends, seasonal effects, or the sustainability of observed strength improvements.

Because of these limitations, the results should be interpreted as a case study of one individual rather than evidence that applies broadly to other populations.

D. Recommendations and Future Work

Several improvements could be made in future work. First, extending the data collection period to several months or a full year would allow for the analysis of longer-term trends, seasonal changes, and periods of planned progression or deloading.

Second, combining manual self-logging with automated tools such as fitness trackers, smartwatches, or mobile applications could reduce entry errors and provide additional variables (for example, heart rate, step counts, or sleep duration). This would improve data completeness and make the analysis more robust.

Third, future studies could incorporate more contextual variables, such as sleep hours, perceived fatigue, stress level, academic workload, or nutrition. These factors may help explain why some weeks are high or low in training volume and how recovery interacts with performance and strength changes.

Fourth, including multiple participants would make it possible to compare different training styles and examine whether patterns seen in this study such as the importance of cross-training minutes are common or unique. This would improve the external validity of the findings.

Finally, more advanced analytical methods, such as time-series forecasting, clustering of weekly patterns, or mixed-effects models, could be applied to identify distinct training behaviors over time and to better model the relationship between training load and strength outcomes.

Overall, this project shows that even a small, self-logged dataset can provide useful insights into personal training behavior, and it highlights several directions for future research on self-tracked fitness and lifestyle habits.

VI. CONCLUSION

This study analyzed my personal fitness behavior by tracking gym consistency and cross-training activities using self-logged data over several weeks. The main goal was to understand how different types of physical activities contribute to overall weekly training volume, how consistent my routine is over time, and how these patterns relate to strength progress.

The results showed that most training weeks fall within a moderate range of roughly 100 to 250 minutes, with a few high-volume weeks driven primarily by longer cross-training sessions such as pickleball and jogging. Gym sessions were stable in both frequency and duration, reflecting structured 60-minute workouts, while cross-training minutes varied much more across weeks. The Pearson correlation test indicated that the number of gym sessions per week was essentially unrelated to total weekly training time, suggesting that cross-training duration and overall minutes, rather than gym frequency alone, are the main drivers of weekly training volume.

Model-based analyses using Logistic Regression and a linear SVM further supported this conclusion. Feature weights from both models consistently highlighted cross-training minutes and gym involvement (gym sessions and gym minutes) as the strongest predictors of high training weeks, while activity variety had little consistent effect. The strength analysis using estimated one-repetition maximum (e1RM) for barbell Romanian deadlift and leg extension showed small but meaningful improvements over the study period, consistent with the effects of regular resistance training.

Through this project, I learned that my routine is generally consistent but becomes more intense during specific weeks when I accumulate substantial cross-training time. Simply increasing the number of gym visits does not automatically translate into higher total training volume; instead, the type and duration of activities are more important. These insights can be applied in practice by monitoring weekly patterns, planning cross-training more deliberately, and aiming to avoid long inactive periods while maintaining a sustainable workload.

Overall, this project demonstrates that self-tracking combined with basic data science methods such as descriptive statistics, visualization, simple statistical tests, and interpretable machine learning models can provide useful insights into personal fitness habits and strength trends. Analyzing my own data helped me better understand my behavior and identify areas for improvement, illustrating how small-scale, personal data science projects can support more informed and healthier lifestyle decisions.

REFERENCES

[1] D. Lupton, The Quantified Self: A Sociology of Self-Tracking. Cambridge, UK: Polity Press, 2016.

[2] E. J. Page, K. M. Free, and A. T. Kwasnicka, “The use of self-monitoring and technology to increase physical activity: A systematic review,” International Journal of Environmental Research and Public Health, 2020.

[3] D. M. Bravata, C. Smith-Spangler, V. Sundaram, et al., “Using pedometers to increase physical activity and improve health: A systematic review,” JAMA, vol. 298, no. 19, pp. 2296–2304, 2007, doi: 10.1001/jama.298.19.2296.

[4] T. Ferguson, A. R. Rowlands, et al., “Effectiveness of wearable activity trackers to increase physical activity: A systematic review and meta-analysis,” The Lancet Digital Health, 2022.

[5] B. Stiglbauer, S. Weber, and B. Batinic, “Does your health really benefit from using a self-tracking device? Evidence from a longitudinal randomized control trial,” Computers in Human Behavior, vol. 94, pp. 131–139, 2019, doi: 10.1016/j.chb.2019.01.018.

[6] American College of Sports Medicine, “Progression models in resistance training for healthy adults,” Medicine & Science in Sports & Exercise, vol. 41, no. 3, pp. 687–708, 2009, doi: 10.1249/MSS.0b013e3181915670.

[7] T. M. Wood, G. F. Maddalozzo, and R. A. Harter, “Accuracy of seven equations for predicting 1-RM performance,” Measurement in Physical Education and Exercise Science, 2002.

[8] M. A. Nascimento, et al., “Validation of the Brzycki equation for the estimation of 1-RM,” Revista Brasileira de Medicina do Esporte, 2007.

[9] D. Jin and others, “Self-tracking behaviour in physical activity: A systematic review of drivers and outcomes of fitness tracking,” Behaviour & Information Technology, 2022.

[10] R. Wang, G. M. Harari, P. Hao, X. Zhou, and A. T. Campbell, “SmartGPA: How smartphones can assess and predict academic performance of college students,” in Proc. ACM UbiComp, 2015, doi: 10.1145/2750858.2804251.