Performance and Recovery Analysis of a 50-Day Bodyweight Training Dataset Using Power BI & R

Name: Daayum Mohsin ISPAS level 1

Date: 22nd March 2026

2283 words

Executive Summary:

This report examines a 50-day self-tracked bodyweight training challenge consisting of 100 push-ups, 100 sit-ups, and 100 squats completed daily, analysed through Power BI and supported by extended R-based statistical analysis. The central problem addressed in this study was whether a simple but highly consistent training protocol, performed during Ramadan under restricted nutrition, hydration, and recovery conditions, could still generate measurable improvements in performance and bodyweight, while also serving as a credible performance analytics case study.

The findings indicate that the challenge produced meaningful positive outcomes. Across the 50-day period, adherence remained at 100%, bodyweight decreased from 84.2 kg to 80.6 kg, and average exercise intensity remained relatively stable despite fluctuating recovery conditions. The dashboard showed that while short-term regression occurred, particularly around Day 40, the overall trend remained favourable. Recovery metrics such as heart rate variability, resting heart rate, sleep duration, and post-workout heart rate provided additional evidence that performance variability was influenced by recovery quality rather than motivation alone.

Extended R analysis strengthened these findings. A very strong negative relationship was identified between day and bodyweight ($r = -0.898$; $R^2 = 0.807$), suggesting that the structured programme was strongly associated with a consistent downward weight trend.

1. Introduction

This report analysed self-collected 50 day training dataset based on a bodyweight challenge consisting of 100 push-ups, 100 sit-ups and 100 squats completed daily. The purpose of the project was to investigate how consistency, recovery and physical output changed over time and to demonstrate how personal data can be transformed into meaningful analytical dashboard using Power BI.

The project was designed around the idea that even simple training routine can generate a rich dataset when supported by structured tracking. Rather than recording only repetitions completed, multiple dependent variables were monitored to provide a broader viewer of performance and recovery. These included weight, session intensity, sleep duration, exercise duration, resting heart rate, post workout heart rate, heart rate variability and estimated calorie intake. This expanded the analysis beyond simple fitness logging and into a more complete performance analytics framework.

The main output of the project is an interactive Power BI dashboard which combines KPI cards, time-series visualisations, conditional formatting, and a sparkline style trend element to communicate both performance outcomes and recovery patterns. The broader aim of the project is to demonstrate practical data analytics skills through a personal dataset, including data cleaning, modelling, DAX measure creation, dashboard design and Interpretation of findings.

2. Dataset and Methodology

The dataset consists of 50 daily observations recorded over the duration of the challenge with 14 columns. Each row represents one day of activity and includes both performance and recovery metrics. The core variables used in the dashboard were:

Variable	Description	Data Type	Measurement Type
Day / Date	Sequential day and calendar date	Integer / Date	Temporal
Weight (kg)	Body weight measurement	Decimal (Float)	Continuous
Push-ups Count	Number of push-ups completed	Integer	Discrete
Sit-ups Count	Number of sit-ups completed	Integer	Discrete
Squats Count	Number of squats completed	Integer	Discrete
Push-ups Intensity	Subjective intensity rating (1–10)	Integer	Ordinal
Sit-ups Intensity	Subjective intensity rating (1–10)	Integer	Ordinal
Squats Intensity	Subjective intensity rating (1–10)	Integer	Ordinal
Average Intensity	Mean intensity across exercises	Decimal (Float)	Continuous
Exercise Duration (min)	Total workout time per session	Integer	Continuous
Sleep Duration (hours)	Hours slept per night	Decimal (Float)	Continuous
Resting Heart Rate (bpm)	Heart rate at rest	Integer	Continuous
Post-Workout HR (bpm)	Heart rate after exercise	Integer	Continuous
Heart Rate Variability	HRV measurement	Integer	Continuous

Calorie Intake (kcal)	Estimated daily calorie consumption	Integer	Continuous

Table 1: Comprehensive Data Dictionary for Performance, Recovery, and Training Variables

The dataset includes a mix of temporal, continuous, discrete, ordinal, and qualitative variables, enabling multi-dimensional analysis of performance and recovery. Data preparation involved cleaning inconsistent formatting and converting measurements into a usable analytical structure. For example, exercise duration entries were standardised into minutes, weigh-ins were only kept on days where actual measurements were recorded, and estimated calorie values were assigned manually from the diet log where possible. Body fat percentage was removed from the main analysis because it was not measured consistently enough to support reliable interpretation.

Power BI was then used to build the model and create analytical measures. DAX was applied for calculations such as average HRV, average calories, heart rate recovery gap, checkpoint weight differences, and weight progression indicators. Filters were also created for intensity bands and duration bands so that the user could interact with the dashboard and observe how training behaviour changed under different conditions.

3. Dashboard Design and KPI Selection

The dashboard was designed to balance analytical depth with visual clarity. Rather than filling the page with unrelated charts, the layout was structured around three core questions: what was completed, how performance changed, and what outcomes were produced.

The top-right section of the dashboard contains KPI cards for total repetitions and average intensity across each exercise type. These metrics provide immediate context about volume and perceived effort. Because repetitions remained fixed, intensity became a more meaningful variable for understanding how demanding the sessions felt across time.

The lower-right section focuses on recovery and physiological response. Metrics such as average sleep duration, average HRV, average resting heart rate, average post-workout heart rate, and recovery gap were included because they provide insight into how the body responded to the workload. These KPIs are especially important in a project conducted during Ramadan, where recovery capacity may be affected by disrupted sleep and restricted nutrition timing.

The centre of the dashboard includes a time-series line chart for push-up, sit-up, and squat intensity. This allows performance patterns to be compared visually across the full challenge period. It shows whether intensity remained stable, increased, or declined, and whether different exercises followed similar or different patterns.

The lower left section was used to tell the weight progression story more explicitly. Checkpoint cards were created for Day 1, Day 10, Day 20, Day 30, Day 40, and Day 50, with directional indicators showing gains or losses between key points. This was paired with a sparkline-style trend visual to reinforce the overall bodyweight trajectory. This combination was chosen because a single final weight change number would not show the temporary regression phase around Day 40, which is important for honest analysis.

The dashboard also includes a short explanatory paragraph summarising the challenge and its context. This supports interpretation and makes the dashboard easier to understand for someone unfamiliar with the dataset. Overall, the design aims to combine personal branding, technical Power BI work, and analytical storytelling into one coherent output.

4. Results: Power BI Dashboard

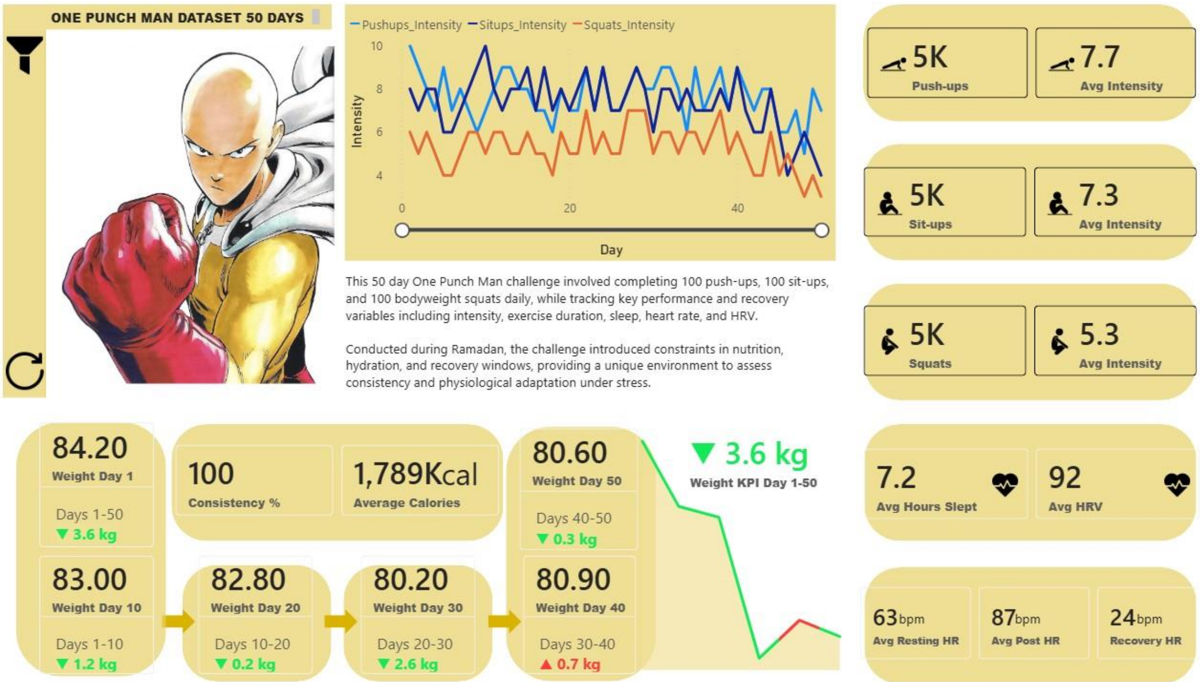

Figure 1: Final Power BI Dashboard Overview

The final dashboard showed that the challenge was completed with 100% adherence across all 50 days. This is one of the most significant findings, as consistency formed the foundation of all other outcomes observed in the project.

In terms of bodyweight, the recorded checkpoints showed a reduction from 84.2 kg on Day 1 to 80.6 kg on Day 50. This represents a net loss of 3.6 kg across the challenge period. However, the weight trend was not perfectly linear. The checkpoint view showed a steady decline from Day 1 to Day 30, followed by a temporary increase at Day 40, before falling again by Day 50. This indicates that short-term regression can occur even within an overall positive trend.

Average calorie intake was estimated at approximately 1,789 kcal per day based on the logged diet entries. This suggests that the challenge took place within a relatively constrained energy intake context, especially considering the daily training volume and Ramadan fasting conditions. The average sleep duration was around 7.2 hours, which is reasonable overall, though variability between days likely contributed to fluctuations in recovery markers.

Recovery-related KPIs also produced useful patterns. Average HRV was approximately 92, while average resting heart rate and average post-workout heart rate were around 63 bpm and 87 bpm respectively. The recovery gap measure provided a simple way to summarise the difference between baseline and post-session cardiac response. These indicators, combined with the session intensity chart, suggest that performance was not static but responded to changes in recovery quality.

The intensity chart showed that average effort remained relatively strong across the challenge, particularly for push-ups and sit-ups, although there was visible variability. Squat intensity remained lower on average than the upper-body exercises, which may reflect differences in muscular fatigue, familiarity, or pacing strategy. Despite these fluctuations, the overall pattern suggests that stable performance was maintained across the full challenge period.

5. R studio Extended Analysis

Relationship	r (Correlation)	R^2 (Variance Explained)	Interpretation
Day vs Weight	-0.898	0.807 (~80.7%)	Very strong negative relationship
Calories vs Weighted Volume	0.108	0.012 (~1.16%)	Weak positive relationship
Time vs Average Intensity	-0.042	0.002 (~0.18%)	Extremely weak negative relationship

Table 2: Correlation Analysis and Variance Explained Across Key Performance Variables

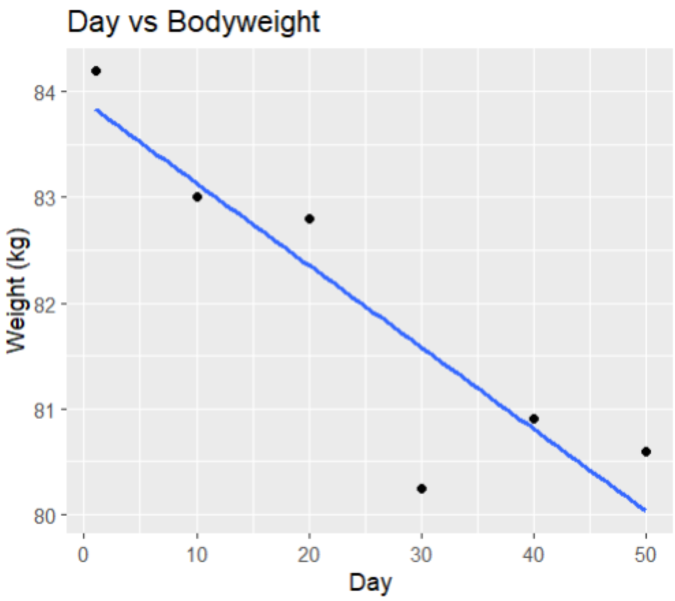

Figure 2: Bodyweight measured over a period of 50 days.

Correlation analysis identified a very strong negative relationship between day and bodyweight (r = -0.898), indicating a consistent reduction in bodyweight over the duration of the training programme.

The coefficient of determination ($R^2 = 0.807$) suggests that approximately 80.7% of the variation in bodyweight can be explained by time, demonstrating the effectiveness of the programme in achieving weight loss.

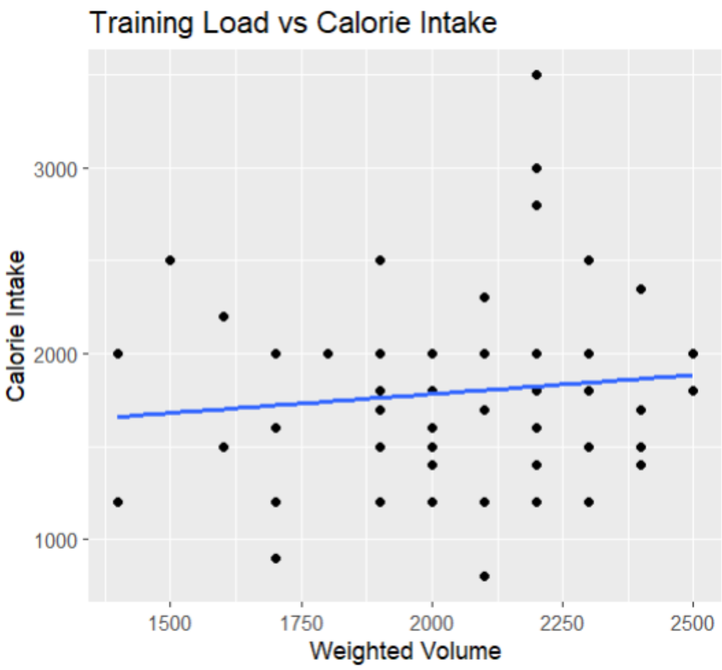

Figure 3: Training Load vs Calorie Intake (R Output)

A weak positive relationship was observed between caloric intake and weighted training volume ($r = 0.108$), indicating that higher caloric intake may be associated with slightly increased training output. However, the low R^2 value (= 1.16%) suggests that this relationship has limited practical significance.

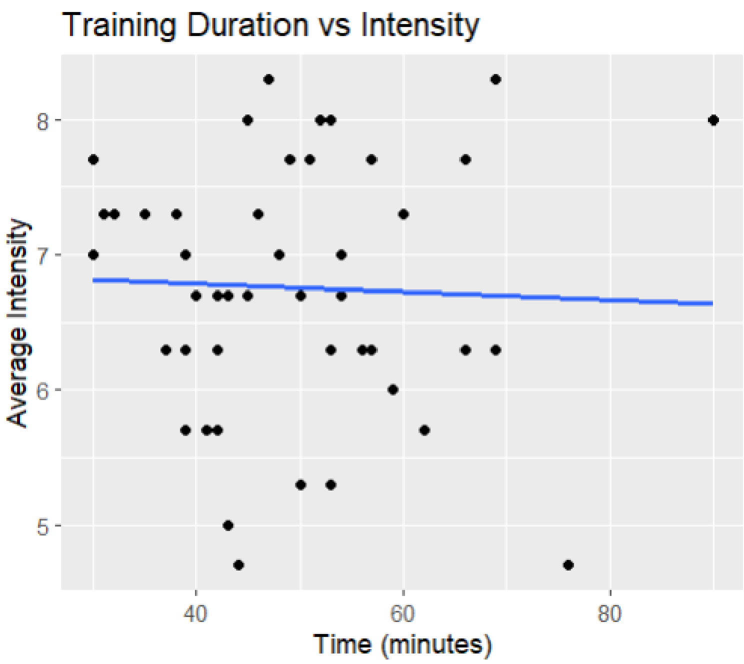

Figure 4: Time vs Average Intensity (R Output)

In contrast, the relationship between training duration and average intensity was negligible (r = -0.042), indicating that session length does not meaningfully influence training intensity within this dataset.

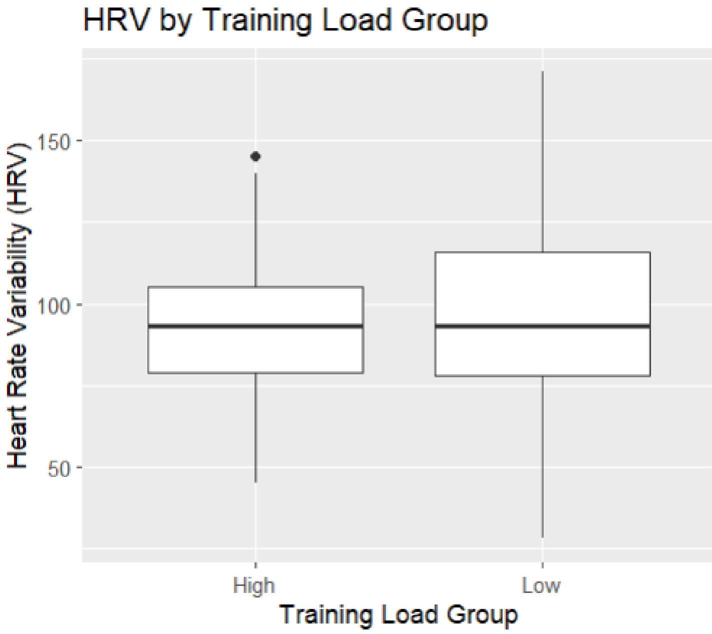

Figure 5: Heart Rate Variability by Training Load Group Boxplot Analysis

To further investigate the effect of training load on recovery, a Welch two-sample t-test was conducted comparing HRV between high and low training load groups. The results showed no statistically significant difference between the groups (p = 0.754), with mean HRV values of 92.96 for the high load group and 90.36 for the low load group. The confidence interval (-14.00 to 19.20) crossed zero, indicating that any observed difference is likely due to random variation rather than a true effect of training load.

6. Slicers and Interactivity

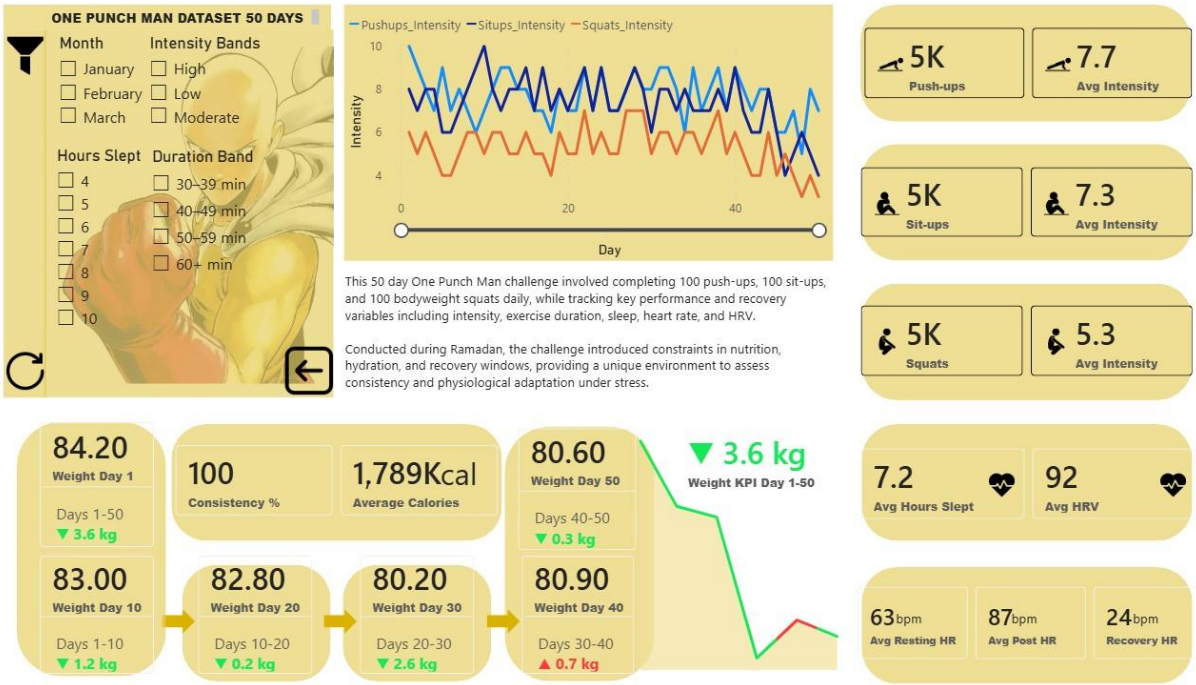

Figure 6: The User Interface Including Slicers from a dropdown menu

To enhance user interaction and enable dynamic data exploration, multiple slicers were incorporated into the dashboard. These slicers allow users to filter the dataset based on key behavioural and temporal variables, making the analysis more flexible and user driven.

The primary slicers implemented include:

- **Month Filter:** Enables selection between January, February, and March, allowing users to observe how performance and recovery metrics varied across different phases of the challenge.
- **Intensity Bands:** Categorises sessions into High(8-10), Moderate(6-7), and Low intensity (Less than or equal to 5). This allows users to isolate performance under different effort levels and assess how intensity impacts recovery variables such as HRV and heart rate.
- **Sleep Duration (Hours Slept):** Allows filtering based on sleep quantity (e.g. 4–10 hours). This is particularly useful for analysing the relationship between sleep and performance variability.
- **Exercise Duration Bands:** Groups sessions into time intervals (30–39 min, 40–49 min, 50–59 min, and 60+ min). This enables users to evaluate training efficiency and determine whether shorter sessions maintained similar performance outputs.

These slicers were positioned on the left hand side of the dashboard to ensure easy accessibility without disrupting the main visual flow. Their design follows a consistent structure, maintaining visual clarity while supporting quick filtering.

From an analytical perspective, the slicers play a key role in transforming the dashboard from a static report into an interactive analytical tool. Users can explore relationships between variables such as intensity, report and performance outcomes. This interactivity enhances the dashboard's analytical value by enabling real time segmentation of performance data, supporting deeper insight into the interaction between training load, recovery, and outcomes.

6. Limitations

Although the project produced useful insights, it also revealed several limitations that should be acknowledged. First, the dataset is based on a single participant, meaning these findings cannot be generalised to a wider population. As such, the analysis is best interpreted as a case study rather than a universal model of performance and recovery.

Second, calorie intake values were estimated rather than measured precisely. While some meals were clearly quantified, others were recorded more generally, reducing the overall accuracy of the calorie intake KPI. Therefore, this metric should be interpreted as an approximate indicator rather than an exact nutritional measurement.

Third, weight measurements were not recorded daily, but instead at selected checkpoints. This limits the resolution of the time-series analysis, as short-term fluctuations may not be fully captured. A more consistent weigh-in schedule would have enabled a more detailed and reliable trend analysis.

Fourth, body fat percentage was excluded from the final dashboard due to inconsistent measurement. While this decision improved data reliability, it reduced the ability to assess changes in body composition beyond total bodyweight.

Fifth, the original "One Punch Man" concept includes a daily 10 km run; however, this component was intentionally excluded from the dataset. This decision was made to reduce the risk of overuse injury and excessive physiological strain, particularly given the high frequency training and additional constraints imposed during Ramadan.

Research suggests that sudden increases in training volume and repetitive high impact activity are strongly associated with increased injury risk (Nielsen et al., 2014). Additionally, inadequate recovery and cumulative fatigue can further elevate the likelihood of musculoskeletal injury (Meeusen et al., 2013).

7. Future Work

While this project provides a strong descriptive analysis of performance and recovery over a 50-day period, several opportunities exist to extend and improve the analytical depth of the study.

First, future work could incorporate continuous and higher-frequency data collection, including daily weight measurements and more precise nutritional tracking. This would allow for more robust time-series modelling and reduce variability introduced by missing or estimated values.

Second, the integration of wearable technology data (e.g. continuous heart rate monitoring, sleep stages, and step count) could enhance the accuracy of recovery and performance metrics. Previous research has demonstrated that wearable devices can provide valuable physiological insights and improve the monitoring of training load and recovery (Peake et al., 2018).

Third, the analysis could be extended by applying statistical and machine learning techniques. For example, correlation analysis could be used to quantify relationships between HRV, sleep, and performance, while regression models could help predict performance outcomes based on recovery variables. More advanced approaches, such as time-series forecasting, could also be applied to predict future weight trends or performance levels.

Fourth, expanding the dataset to include multiple participants would allow for comparative analysis and improve the generalisability of findings. This would enable identification of broader patterns and reduce the limitations associated with single-subject analysis.

Finally, future versions of the dashboard could incorporate a composite performance score, combining variables such as intensity, duration, HRV, and sleep into a single KPI.

References

- Meeusen, R., Duclos, M., Foster, C., Fry, A., Gleeson, M., Nieman, D., Raglin, J., Rietjens, G., Steinacker, J. and Urhausen, A. (2013) ‘Prevention, diagnosis and treatment of the overtraining syndrome: Joint consensus statement of the European College of Sport Science and the American College of Sports Medicine’, *European Journal of Sport Science*, 13(1), pp. 1–24.
- Nielsen, R.O., Buist, I., Sørensen, H., Lind, M. and Rasmussen, S. (2014) ‘Training errors and running related injuries: A systematic review’, *International Journal of Sports Physical Therapy*, 9(3), pp. 408–417.
- Peake, J.M., Kerr, G. and Sullivan, J.P. (2018) ‘A critical review of consumer wearables, mobile applications, and equipment for providing biofeedback, monitoring stress, and sleep in physically active populations’, *Frontiers in Physiology*, 9, p. 743.

Appendices

DAX Measures

Avg HRV = ROUND(AVERAGE('one_punch_man_50_day_dataset'[HRV]),0)

Avg Post HR = ROUND(AVERAGE('one_punch_man_50_day_dataset'[PostHR_bpm]),0)

Avg Resting HR = ROUND(AVERAGE('one_punch_man_50_day_dataset'[RestingHR_bpm]),0)

HR Recovery Gap = ROUND(AVERAGE('one_punch_man_50_day_dataset'[PostHR_bpm]),0)

- ROUND(AVERAGE('one_punch_man_50_day_dataset'[RestingHR_bpm]),0)

Weight Day 1 = CALCULATE(MAX('one_punch_man_50_day_dataset'[Weight_kg]), 'one_punch_man_50_day_dataset'[Day] = 1)

Weight Day 10 = CALCULATE(MAX('one_punch_man_50_day_dataset'[Weight_kg]), 'one_punch_man_50_day_dataset'[Day] = 10)

Weight Day 20 = CALCULATE(MAX('one_punch_man_50_day_dataset'[Weight_kg]), 'one_punch_man_50_day_dataset'[Day] = 20)

Weight Day 30 = CALCULATE(MAX('one_punch_man_50_day_dataset'[Weight_kg]), 'one_punch_man_50_day_dataset'[Day] = 30)

Weight Day 40 = CALCULATE(MAX('one_punch_man_50_day_dataset'[Weight_kg]), 'one_punch_man_50_day_dataset'[Day] = 40)

Weight Day 50 = CALCULATE(MAX('one_punch_man_50_day_dataset'[Weight_kg]), 'one_punch_man_50_day_dataset'[Day] = 50)

Consistency % = DIVIDE(COUNTROWS('one_punch_man_50_day_dataset'),50) * 100

Weight KPI Days 1-50 =

UNICHAR(10) &

UNICHAR(9660) & " " & FORMAT(ABS([Weight Day 50] - [Weight Day 1]), "0.0") & " kg"

Weight KPI Day 10 =

UNICHAR(10) &

UNICHAR(9660) & " " & FORMAT(ABS([Weight Day 10] - [Weight Day 1]), "0.0") & " kg"

Weight KPI Day 20 =

UNICHAR(10) &

UNICHAR(9660) & " " & FORMAT(ABS([Weight Day 20] - [Weight Day 10]), "0.0") & " kg"

Weight KPI Day 30 =

UNICHAR(10) &

```
UNICHAR(9660) & " " & FORMAT(ABS([Weight Day 30] - [Weight Day 20]), "0.0") & " kg"

Weight KPI Day 40 =

UNICHAR(10) &

UNICHAR(9660) & " " & FORMAT(ABS([Weight Day 40] - [Weight Day 30]), "0.0") & " kg"

Weight KPI Day 50 =

UNICHAR(10) &

UNICHAR(9660) & " " & FORMAT(ABS([Weight Day 50] - [Weight Day 40]), "0.0") & " kg"
```

R Output

```
OPM_enriched <- OPM %>%
  select(Day, Weight_kg, Total_Volume, Avg_Intensity, Sleep_hr, HRV) %>%
  mutate(Calories_est = OPM$Calories_est)


OPM <- OPM %>%
  mutate(Total_Volume = Pushups_Count + Situps_Count + Squats_Count)


OPM <- OPM %>%
  mutate(Weighted_Volume =
         (Pushups_Count * Pushups_Intensity) +
         (Situps_Count * Situps_Intensity) +
         (Squats_Count * Squats_Intensity))


ggplot(OPM, aes(y = HRV, x = Sleep_hr)) +
    geom_point() +
    geom_smooth(method = "lm", se = FALSE) +
    labs(title = "Weighted Volume vs Duration(m) ")


#Correlation Analysis
ggplot(OPM, aes(x = Weighted_Volume, y = Calories_est)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
```

```
  labs(title = "Training Load vs Calorie Intake",
       x = "Weighted Volume",
       y = "Calorie Intake")

r_Calories <- cor(OPM$Weighted_Volume, OPM$Calories_est, use = "complete.obs")
r_Calories
r_Calories^2

ggplot(OPM, aes(x = Weighted_Volume, y = HRV)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Training Load vs Recovery (HRV)",
       x = "Weighted Volume",
       y = "HRV")

r_load_hrv <- cor(OPM$Weighted_Volume, OPM$HRV, use = "complete.obs")
r_load_hrv
r_load_hrv^2
ggplot(OPM, aes(x = Day, y = Weight_kg)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Day vs Bodyweight",
       x = "Day",
       y = "Weight (kg)")

r_day_weight <- cor(OPM$Day, OPM$Weight_kg, use = "complete.obs")
r_day_weight
r_day_weight^2
```

```
# Correlation
ggplot(OPM, aes(x = Time_min, y = Avg_Intensity)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Training Duration vs Intensity",
       x = "Time (minutes)",
       y = "Average Intensity")

r_time_intensity <- cor(OPM$Time_min, OPM$Avg_Intensity, use = "complete.obs")
r_time_intensity
r_time_intensity^2

ggplot(OPM, aes(x = Load_Group, y = HRV)) +
  geom_boxplot() +
  labs(title = "HRV by Training Load Group",
       x = "Training Load Group",
       y = "Heart Rate Variability (HRV)")

t.test(HRV ~ Load_Group, data = OPM)
```