Simply Sport Foundation

Data Analyst Assignment

-Submitted by
Manjusha Madhukar Ghatke

Contents

Data Preprocessing- Python

The following data preprocessing steps have been carried out using python and Jupyter Notebooks:

- Missing Values Treatment
- Feature Engineering (e.g. extracting month, day from date column)
- Each of the 17 symptoms has been created into new columns with values 0 or 1 to indicate presence or absence of the particular symptom.
- Another table solely about symptoms (just in case if it would be necessary) has been created and has not been used in the dashboard.

Data Preprocessing- Excel

New columns created:

- Symptom score:
  - Sum of column values (1 or 0) of each of the 17 symptoms. Range is 0-17
- Mental health score:
  - Average of mental focus, stress management and positive motivation. Range: 0-4
- Period Day:
  - On_period column values (Yes / No) have been converted to 1 or 0 respectively
- Sleep duration:
  - >5 hrs is considered as 5 hrs; 5-7 hrs and <7 hrs has been considered as 6 hrs.
- Helper column:
  - Combination of date and athlete id, as a unique identifier

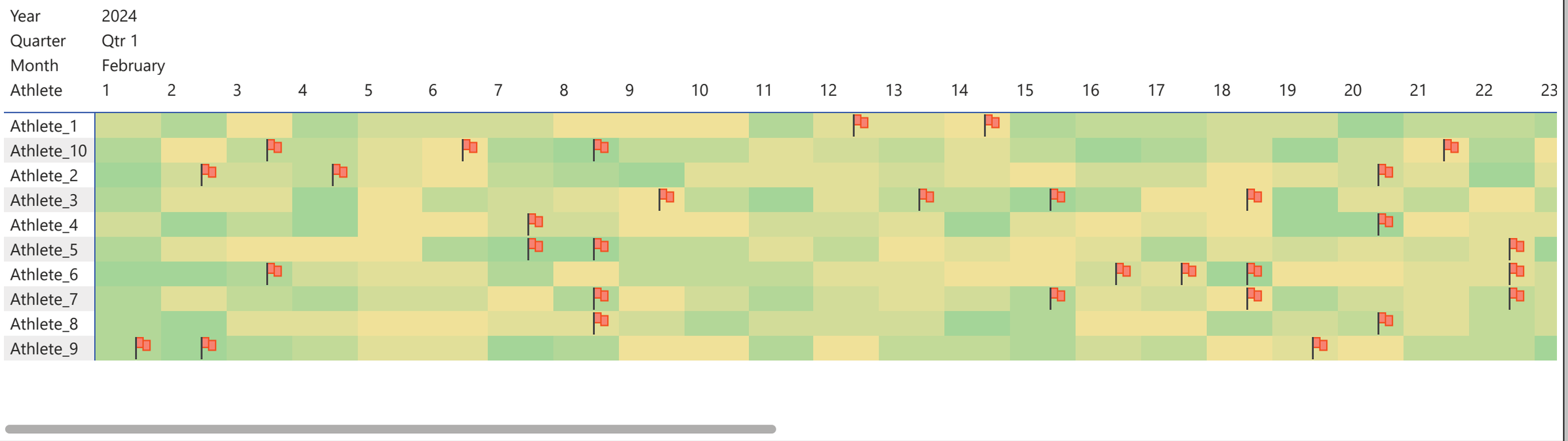
Data Preprocessing- Power Query (PowerBI)

- Week:
  - Week of the year
- Nutrition score:
  - Derived from columns: meals, water intake and pre-training food.
  - Values from each column were assigned numbers depending on the context. If meals were all nutritious , the entry was given a score of 3, if it was somewhat nutritious-2 , and not nutritious was given a score of 1.
  - All of these were added for each row and converted to a scale of 5.
- Recovery score:
  - Combination of positive factors (mental health score, sleep hours, nutrition score) and negative factors (symptom count, period pain, bleeding intensity, period days, and injury today). Range: 0-1
- Symptom severity:
  - Combination of symptom count and period pain, Range: 0-1
- Session Intensity to Recovery Score Ratio:
  - Ratio of session intensity to recovery score. Range: 0-1
- Performance Impact Factor:
  - Combination of positive factors (session intensity, mental health score, sleep hours, nutrition score) and negative factors (symptom severity and injury today). Range: 0-1

Cycle Mapping and Pattern Recognition

Period and Symptoms Count Tracker

Year 2024
Quarter Qtr 1
Month February



Insights

Athlete 1:

- 1st period cluster: 12-2-24 to 14-2-24 (3 days) | 2nd period cluster: 29-2-24 to 4-3-24 (5 days) | 3rd period cluster: 16-3-24 (1 day)
- Moderate to high symptom count before and during periods
- Avg cycle length: 17 days

Athlete 2:

- 1st period cluster: 2-2-24 to 4-2-24 (3 days) | 2nd period cluster: 20-2-24 (1 day) | 3rd period cluster: 2-3-24 (1 day) | 4th period cluster: 16-3-24 (1 day)
- Low to moderate symptom count during periods
- Avg cycle length: 14 days

Note:

- Period days have a red flag icon
- Cells have been conditionally formatted with a green to yellow gradient according to Symptom Score. Green being 0 symptom count and yellow being high symptom count

Athlete
Athlete_1 | Athlete_10 | Athlete_2

Week
5 | 6 | 7 | 8 | 9

Date
01-02-2024 16-03-2024

Dashboards- Cycle Mapping and Pattern Recognition

Insights

Athlete 1:

1st period cluster: 12-2-24 to 14-2-24 (3 days) | 2nd period cluster: 29-2-24 to 4-3-24 (5 days) | 3rd period cluster: 16-3-24 (1 day)
Moderate to high symptom count before and during periods
Avg cycle length: 17 days

Athlete 2:

1st period cluster: 2-2-24 to 4-2-24 (3 days) | 2nd period cluster: 20-2-24 (1 day) | 3rd period cluster: 2-3-24 (1 day) | 4th period cluster: 16-3-24 (1 day)
Low to moderate symptom count during periods
Avg cycle length: 14 days

Athlete 3:

1st period cluster: 9-2-24 to 18-2-24 (10 days) | 2nd period cluster: 29-2-24 (1 day)
Low to moderate symptom count before and during periods
Avg cycle length: 20 days

Athlete 4:

1st period cluster: 7-2-24 (1 day) | 2nd period cluster: 20-2-24 (1 day) | 3rd period cluster: 4-3-24 (1 day) | 4th period cluster: 14-3-24 t o16-3-24 (2 days)
Low symptom count during periods. Moderate symptom count around periods
Avg cycle length: 12 days

Athlete 5:

1st period cluster: 7-2-24 to 8-2-24 (2 days) | 2nd period cluster: 22-2-24 (1 day) | 3rd period cluster: 2-3-24 (1 day) | 4th period cluster: 13-3-24 t o16-3-24 (2 days)
Low symptom count during periods. Moderate symptom count around periods
Avg cycle length: 11 days

Dashboards- Cycle Mapping and Pattern Recognition

Insights

Athlete 6:

1st period cluster: 16-2-24 to 22-2-24 (7 days)
Low to moderate symptom count during and before periods
Avg cycle length: >30 days

Athlete 7:

1st period cluster: 15-2-24 to 22-2-24 (8 days) | (Not considering other period days as clusters as they are placed to close to the 1st period cluster)
Low symptom count during periods
Avg cycle length: >30 days

Athlete 8:

1st period cluster: 20-2-24 to 25-2-24 (6 days) | 2nd period cluster: 12-3-24 to 15-3-24 (4 days) | (Not considering other period days as clusters as they are placed to close to the 1st period cluster)
Low symptom count during periods. Moderate symptom count around periods
Avg cycle length: 21 days

Athlete 9:

1st period cluster: 1-2-24 to 2-2-24 (2 days) | 2nd period cluster: 19-2-24 (1 day) | 3rd period cluster: 5-3-24 to 13-3-24 (9 days)
Low symptom count during periods. Moderate symptom count around periods
Avg cycle length: 17 days

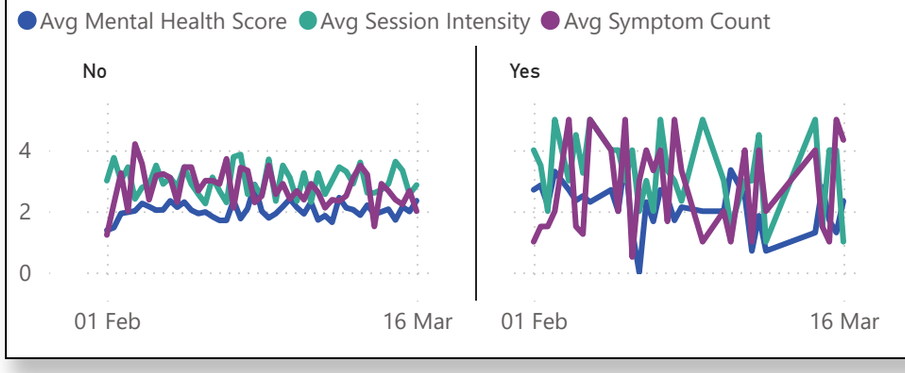
Athlete 10:

1st period cluster: 3-2-24 to 8-2-24 (6 days) | 2nd period cluster: 21-2-24 to 28-2-24 (8 days)
Low to moderate symptom count during periods
Avg cycle length: 18 days

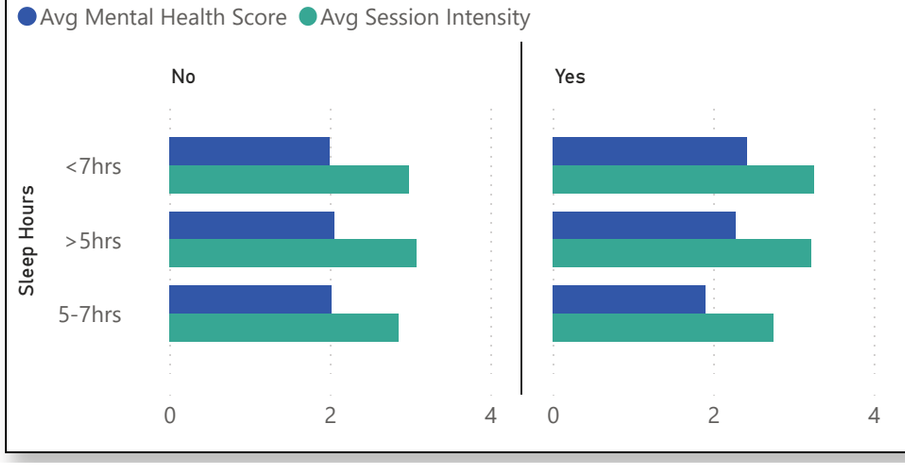
Mood, Sleep & Training Correlation

5.64	2.04	2.98	3.28
Avg Sleep Duration (hrs)	Avg Mental Health Score (0-4)	Avg Session Intensity (0-5)	Avg Nutrition Score (0-5)

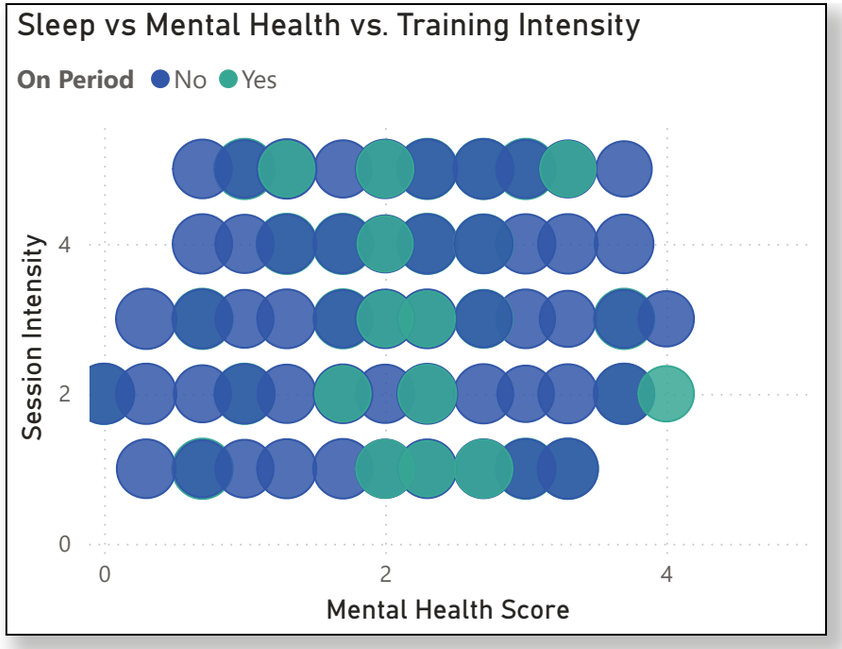
Training, Mental Health & Period Trends



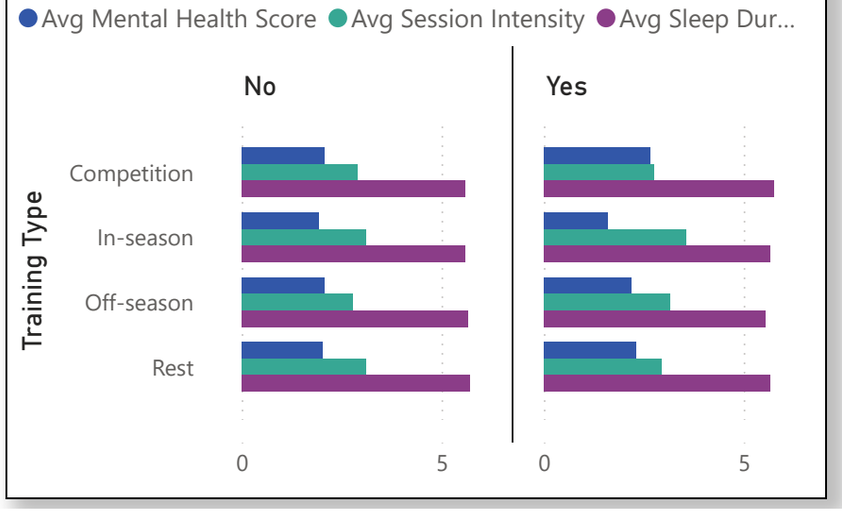
Sleep Quality Impact

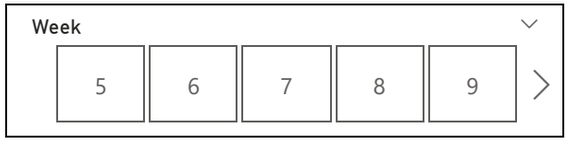


Sleep vs Mental Health vs. Training Intensity



Training Type Impact





Note: Sleep, Mental Health and Training Intensity scatter plot has bubble size according to sleep duration.

Insights

Overview:

- Average sleep duration of 5.64 hours, indicating poor sleep health.
- Below average mental health with a score of 2.04.
- Good session intensity with an average of 3
- Good nutrition score of 3.28

Training, Mental Health & Period Trends:

- Increased variation in mental health score, symptom count and training intensity on period days.

Sleep Quality Impact:

- Period days don't seem to have a negative effect on training intensity and mental health score with 5-6 hrs of sleep.

Dashboards- Mood, Sleep & Training Correlation

Insights

Overview:

- Average sleep duration of 5.64 hours, indicating poor sleep health.
- Below average mental health with a score of 2.04.
- Good session intensity with an average of 3
- Good nutrition score of 3.28

Training, Mental Health & Period Trends:

- Increased variation in mental health score, symptom count and training intensity on period days.

Sleep Quality Impact:

- Period days don't seem to have a negative effect on training intensity and mental health score with 5-6 hrs of sleep.

Sleep vs Mental Health vs Training Intensity:

- While on period, the athletes seem to go through sessions of varying intensity, periods don't seem to affect training intensity.
- While on period, the mental health score ranges from 1.3 to 4, while clustering around 2
- Sleep duration of 5-6 hrs spans across different mental health scores and session intensities
- Overall, there is no significant correlation

Training Type Impact:

- Period days don't seem to have a negative impact on training intensity or type.
- Mental health score is slightly lower during in-season training
- Training intensity is highest during in-season training.

Athlete 10 Deep Dive Report: Weekly Analysis

0.15	2.59	3.28	0.31
Avg Symptom Severity (0-1)	Avg Session Intensity to Recovery Ratio (0-1)	Avg Nutrition Score (0-5)	Avg Performaance Impact Factor (0-1)

Athlete Metrics Summary

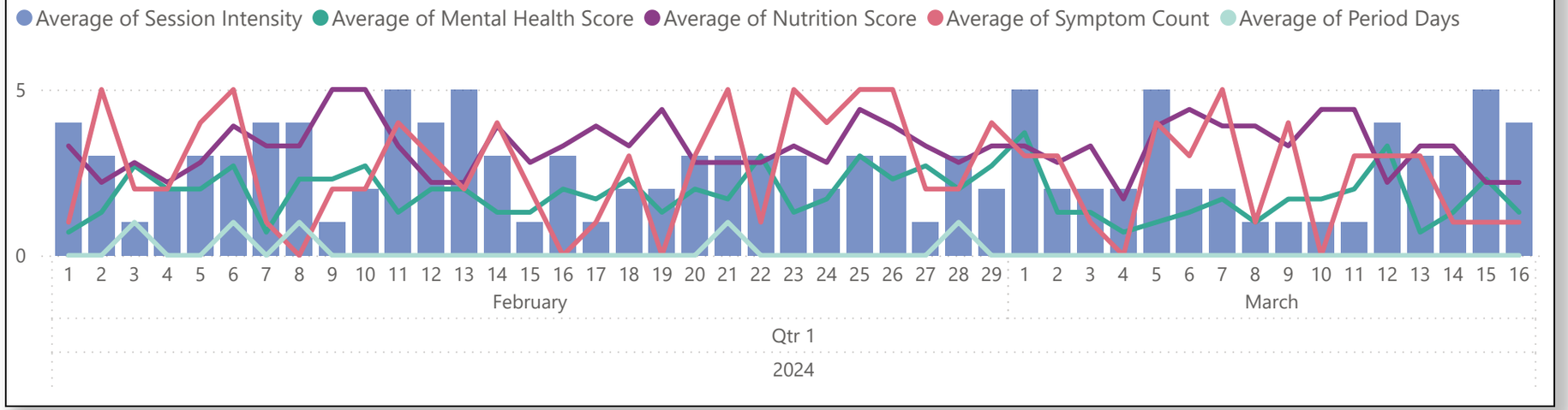
Athlete	Week	Avg Mental Health Score	Avg Session Intensity	Avg
Athlete_10	5	1.57	2.67	
Athlete_10	6	2.10	2.71	
Athlete_10	7	1.66	3.14	
Athlete_10	8	1.90	2.57	
Total		**1.85**	**2.71**	

Athlete: Athlete_10

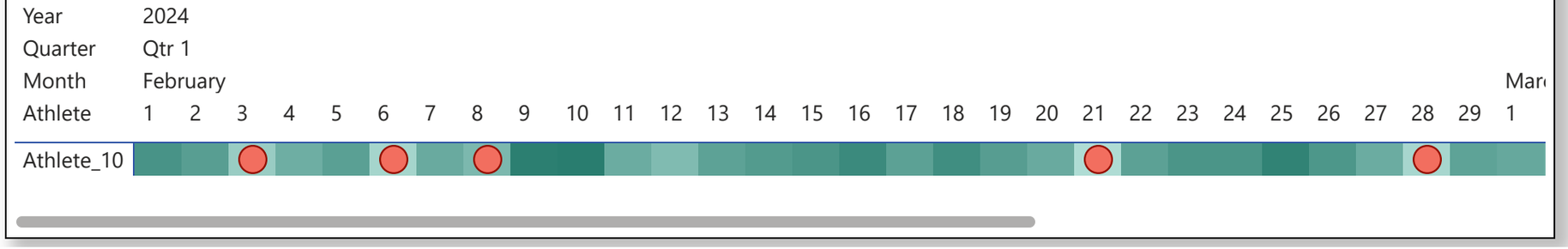
Week: 5 6 7 8 9

Date: 01-02-2024 16-03-2024

Performance Influencer



Athlete Recovery Pattern Analysis



Note: Period days are marked with a red circle in the Recovery Pattern Analysis visualization

Insights

Weeks 5-6 (Early February)

- **Recovery**: Inconsistent recovery patterns with two period days (red circles) visible
- **Mental Health**: Fluctuating mental health scores (dropping as low as 0.70 and peaking at 2.70)
- **Nutrition**: Generally good nutrition scores averaging around 3.3 (on a 0-5 scale)
- **Training**: Variable training intensity with some high-intensity (4.0) sessions

Weeks 7-8 (Mid February)

- **Recovery**: Improved recovery pattern with more consistent green days
- **Mental Health**: More stable mental health with values around 1.30-2.30
- **Nutrition**: Excellent nutrition on Feb 19 (4.40 score). Avg (3.13)

Dashboards- Athlete 10 Deep Dive Report: Weekly Analysis

Insights

Weeks 5-6 (Early February)

- **Recovery**: Inconsistent recovery patterns with two period days (red circles) visible
- **Mental Health**: Fluctuating mental health scores (dropping as low as 0.70 and peaking at 2.70)
- **Nutrition**: Generally good nutrition scores averaging around 3.3 (on a 0-5 scale)
- **Training**: Variable training intensity with some high-intensity (4.0) sessions

Weeks 7-8 (Mid February)

- **Recovery**: Improved recovery pattern with more consistent green days
- **Mental Health**: More stable mental health with values around 1.30-2.30
- **Nutrition**: Excellent nutrition on Feb 19 (4.40 score). Avg (3.13)
- **Training**: Moderate intensity (2.0-3.0) training sessions

Week 9 (Late February)

- **Recovery**: One period day (Feb 28) impacting recovery
- **Mental Health**: Generally good mental health (2.0-3.0)
- **Nutrition**: Good nutrition scores (3.3-4.4)
- **Training**: Fluctuating training intensity

Weeks 10-11 (Early-Mid March)

- **Recovery**: Good and consistent recovery
- **Mental Health**: Extremely variable (from 0.70 to 3.70)
- **Training**: Notable high intensity session (5.0) on March 1 and March 15
- **Symptom Count**: Periods of increased symptoms, particularly March 7

Insights

Menstrual Cycle Impact:

Period days (red circles in the Recovery Pattern Analysis) on Feb 3, 6, 8, 21, and 28 correspond with:

- Increased Symptom count values (5 for Feb 6 and 21)
- Decreased Mental Health scores on most period days
- Training intensity often drops during and immediately around period days

Mental Health-Performance Correlation:

Instances were mental health shows significant correlation with training intensity:

- March 1 shows the athlete's highest mental health score (3.70) coinciding with maximum training intensity (5.0)
- Low mental health days (scores below 1.0) almost always coincide with reduced training intensity

Nutrition Impact:

- The athlete maintains consistently good nutrition (average 3.28 on a 0-5 scale)
- Peak nutrition days (4.4 score on Feb 19 and Feb 25) don't always correlate with peak performance
- However, the lowest nutrition score (1.70 on March 4) coincides with a low mental health score (0.70)

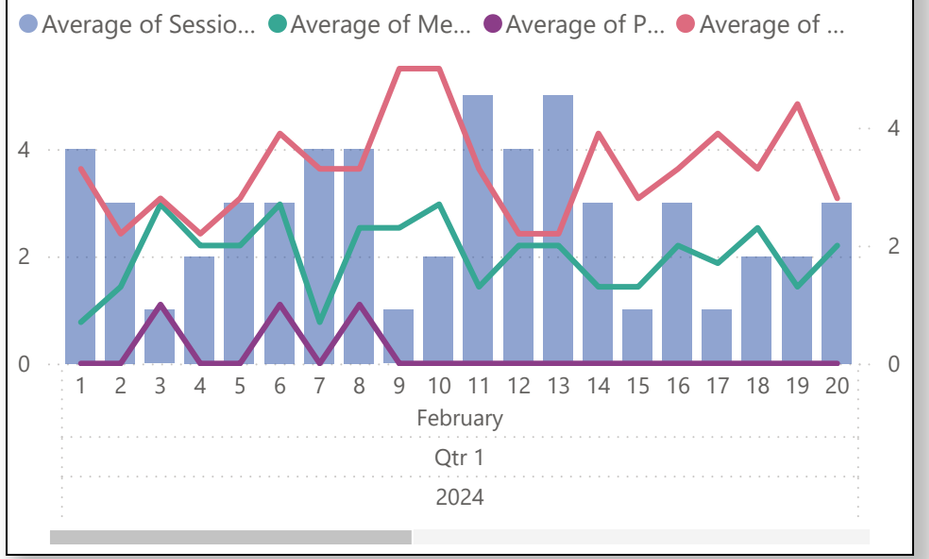
Recovery Capacity:

- The Session Intensity to Recovery Ratio is 2.59, suggesting training intensity may occasionally exceed recovery capacity
- The Performance Impact Factor is relatively low at 0.31 (on a 0-1 scale), indicating room for optimization

Athlete 10 Deep Dive Report: Mood, Sleep & Training

5.67	1.85	2.71
Avg Sleep Duration	Avg Mental Health Score (0-4)	Avg Session Intensity (0-5)

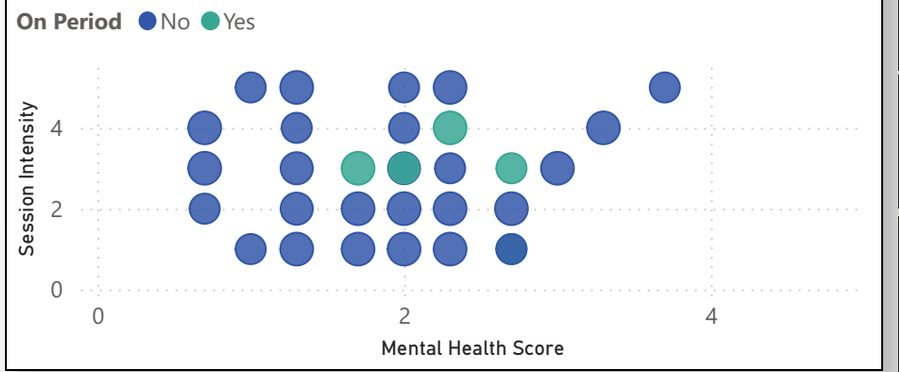
Training, Mental Health, Nutrition & Period Trends



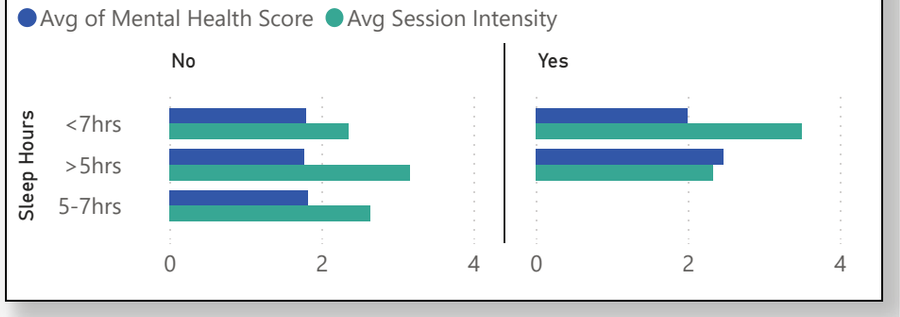
Period and Symptoms Count Tracker

Year	2024
Quarter	Qtr 1
Month	February
Athlete	1 2 3 4 5 6 7 8
Athlete_10	

Sleep vs Mental Health vs. Training Intensity



Sleep Quality Impact



Training Type Impact

Athlete

Athlete_10

Week

5 6 7 8 9

Date

01-02-2024 16-03-2024

Insights

- **Sleep Quality**: Average sleep duration of 5.67 hours
- **Mental Health**: Fluctuating mental health scores with an overall average of 1.85
- **Training**: Variable intensity (average 2.71 on a 0-5 scale) with some high-intensity (4.0) sessions

Training, Mental Health, Nutrition & Period Trends:

- On most period days Training Intensity < 4. and Mental Health Score is < 2.7
- No clear correlation between Mental Health Score and Training Intensity
- No clear correlation between nutrition score and training intensity

Periods and Symptom Count Tracker:

- High symptom count observed on and around period days
- Period clusters seem to occur around 18 days apart

Dashboards- Athlete 10 Deep Dive Report: Mood, Sleep & Training Correlation

Insights

Sleep Quality: Average sleep duration of 5.67 hours
Mental Health: Fluctuating mental health scores with an overall average of 1.85
Training: Variable intensity (average 2.71 on a 0-5 scale) with some high-intensity (4.0) sessions

Training, Mental Health, Nutrition & Period Trends:

On most period days Training Intensity < 4. and Mental Health Score is < 2.7
No clear correlation between Mental Health Score and Training Intensity
No clear correlation between nutrition score and training intensity

Periods and Symptom Count Tracker:

High symptom count observed on and around period days
Period clusters seem to occur around 18 days apart

Sleep vs Mental Health vs Training Intensity:

Sleep Duration- Mostly 5-6 hours (small, similar bubbles)
Mental Health Score- Clustered between 1 and 3
Training Intensity- Spread across all mental health scores (no clear trend)
Period Status- Both groups distributed throughout; no clear pattern
Correlation- No strong correlation between mental health, intensity, and sleep duration

Sleep Quality Impact

We can't say that increased sleep duration correlates with increased training intensity or mental health score.
When on period, higher mental health scores and training intensity are observed

Training Type Impact:

Competition training shows the highest disparity between period and non-period days
Rest days show similar training intensities regardless of period status
In-season training intensity is more consistent across period days

Recommendations

Data Collection:

- Menstruation:
  - Have a column only for start of period instead of on period, this can help calculate different phases of the menstrual cycle and menstrual cycle length more accurately.
  - Ensure athletes understand the various reasons for bleeding (like spotting, activities like bicycle riding), other than bleeding during menstrual phase, to avoid marking additional days as period days.
- Sleep:
  - Collect approximate number of sleep hours instead of ranges
  - The existing options (>5 , 5-7, <7) provide a very narrow range for analysis.
  - Collect sleep quality data as well.

Recommendations

Athlete 10 Deep Dive Report: Recommendations for coach

- Menstrual Cycle:
  - Plan reduced training intensity few days before and during the initial days of menstruation and when there are more symptoms to enable recovery
  - High-intensity training during non-period days where mental health scores naturally trend higher
- Mental Health:
  - Implement mental wellness check-ins
  - Build on the conditions where good mental health scores are observed
- Training & Recovery:
  - Leverage the athlete's ability to handle high-intensity sessions after proper recovery
- Nutrition:
  - Maintain the excellent nutrition practices and address the drop in nutrition score to avoid decrease in mental health and performance
- Sleep:
  - Prioritize 7.5+ hours of sleep, especially during competition phases

THE END