

Professional Sports Injury Analysis

Risk Management Project Report

Executive Summary

- * This report presents a comprehensive analysis of sports injury data collected from 30 athletes over a two-year period (May 2016 to April 2018).
- * The analysis focuses on identifying injury risk factors, creating individual athlete risk profiles, and establishing relationships between workload and injury occurrence.

Project Overview

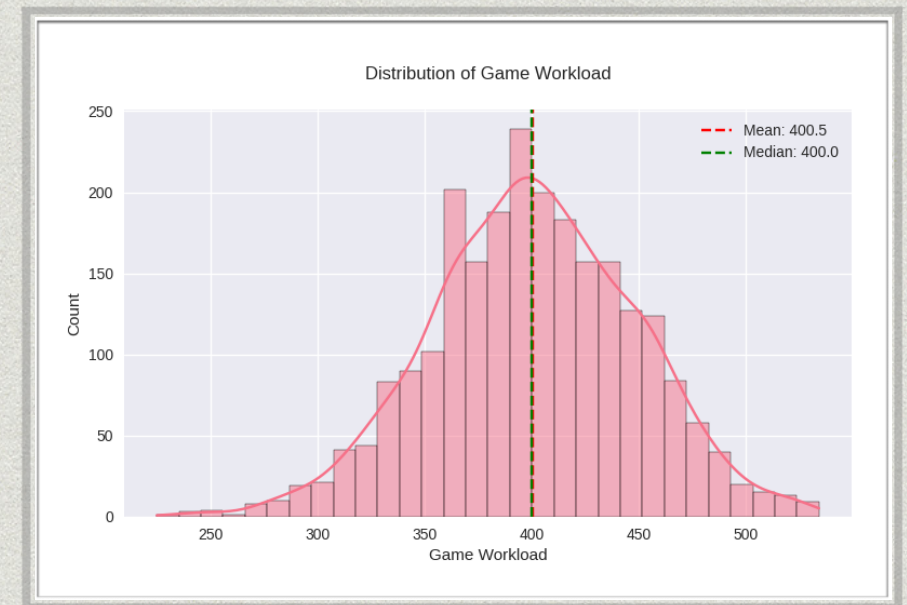
OBJECTIVES

- * Development of individual athlete injury risk profiles
- * Early identification of high-risk athletes
- * Analysis of the relationship between workload and injury occurrence
- * Creation of a predictive model for injury risk assessment

DATA SOURCES

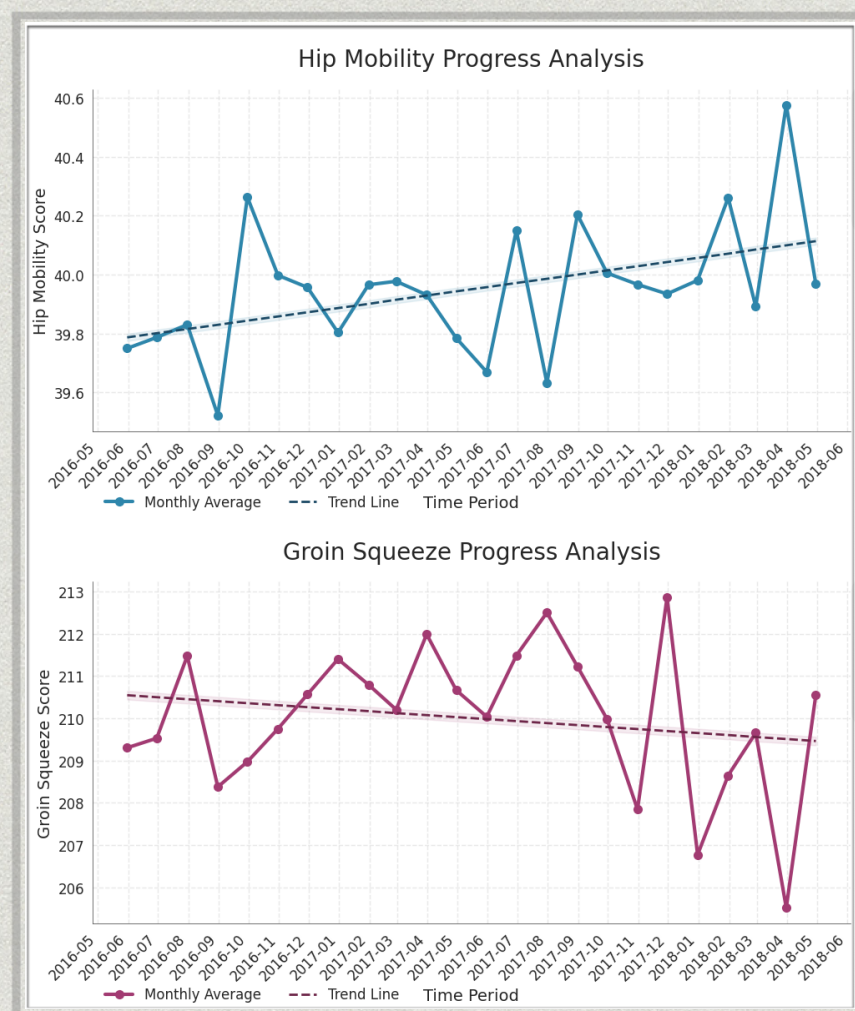
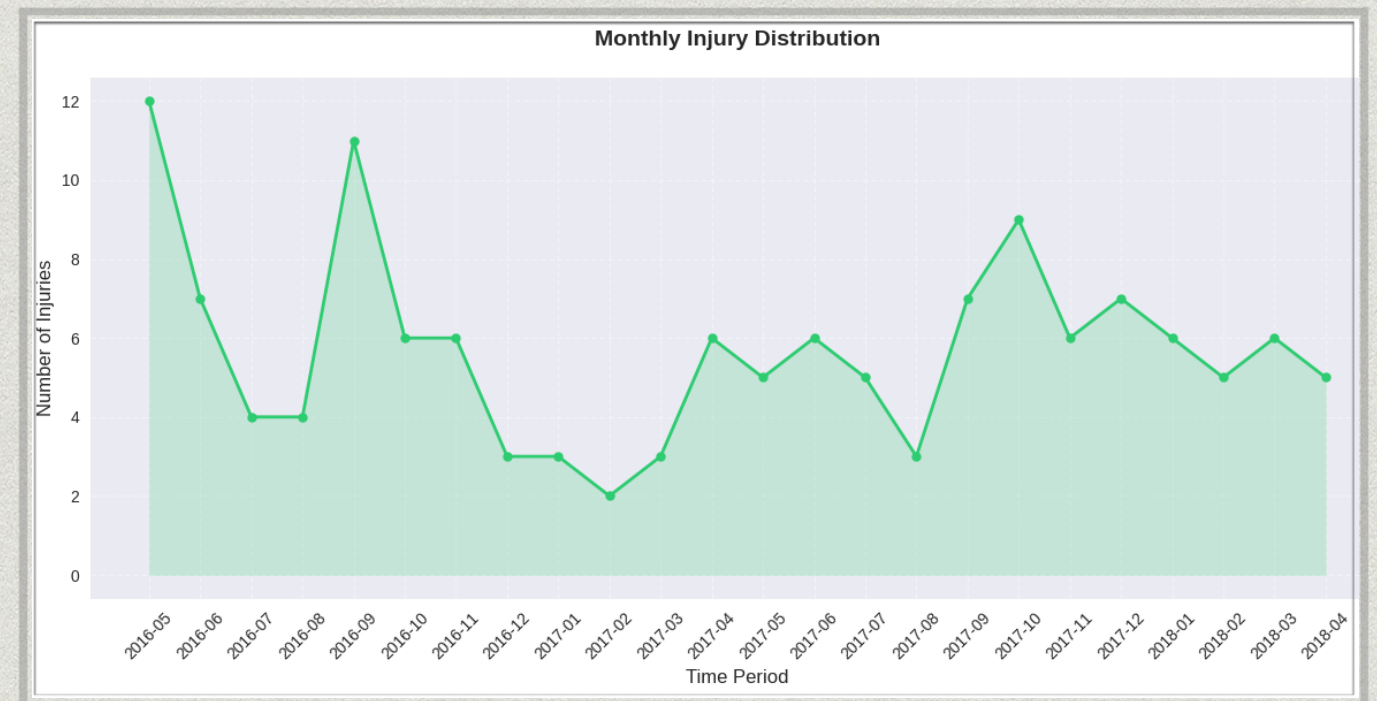
- * Metrics Dataset - Hip mobility and groin squeeze measurements for 30 athletes (2016-2018)
- * Workload Dataset - Game-specific performance and exertion metrics
- * Injuries Dataset - Complete injury tracking system

- * Shows a normal-like distribution centered around 400 units
- * The distribution is slightly right-skewed
- * Range appears to be between ~250-500 units



- * Median workloads generally fall between 350-450 units
- * Several outliers present, particularly in lower ranges
- * Relatively consistent interquartile ranges across athletes

- * Peak of 12 injuries in one month
- * Minimum of 2 injuries in another month
- * Average of 5.7 injuries per month



- * Hip Mobility: Positive trend (slope: 0.0142), significant (p: 0.0312), weak correlation (R^2 : 0.1941).
- * Groin Squeeze: Negative trend (slope: -0.0471), not significant (p: 0.3699), very low correlation (R^2 : 0.0367).
- * High variability; hip mobility shows fluctuations, groin squeeze ranges between 206-213 units.

Methodology

Data Integration and Preprocessing

DATASET MERGING

- * Merging DataFrames: metrics dataset, workload dataset, and injuries dataset were merged using athlete_id and date columns.
- * Handling NaN Values: NaN values in game_workload were filled with 0, treating them as rest days.
- * Adding Injury Status: An injuries_status variable was created, assigning "injured" in injuries dataset and filling others with "non_injured"
- * Creating New Column: A new "resting" column was added, indicating 1 for days with missing workload and 0 for others.
- * Pivoting Data: After these steps, the hip_mobility and groin_squeeze metrics were used to create new columns via a pivot table, reducing duplicate dates for each athlete ID to a single entry.

FEATURE ENGINEERING

* Training/Match Load Monitoring Features:

- workload_7d: Cumulative 7-day workload
- acwr: Acute/Chronic Workload Ratio
- workload_change: Daily workload change rate

* Health and Risk Indicators:

- mobility_trend, rest_days_7d: Fatigue and recovery indicators
- injuries_30d, days_since_injury: Injury history features
- workload_risk, overall_risk: Weighted risk scores combining multiple factors

Analysis Techniques

STATISTICAL ANALYSIS

- * Correlation analysis between metrics and injury occurrence
- * Time series analysis of workload patterns
- * Distribution analysis of injury frequencies