OPTIMAL PLAYER ROTATION AND MANAGEMENT FOR LONG-TERM INJURY REDUCTION IN FOOTBALL: A COMPREHENSIVE ANALYSIS

THIA SU MIAN

SUPERVISOR: LAM XIN YEE

EXAMINER: AARON LOW YI RUI

DEPARTMENT OF MATHEMATICS

NATIONAL UNIVERSITY OF SINGAPORE

Honors Year Project for Semester 1, AY2024/2025

Summary

This report presents a comprehensive analysis of optimal player rotation and management strategies in football, with a focus on long-term injury reduction. By leveraging machine learning and optimization techniques, the project aims to develop data-driven approaches that can assist coaches in making informed decisions regarding player selection and playing time allocation.

The study involves the collection and analysis of a large dataset encompassing player attributes, match data, and refereeing information. Machine learning models are employed to predict injury risk and identify factors influencing player vulnerability. Optimization techniques are then applied to formulate strategies for player rotation that minimize injuries while maximizing team performance.

The author's contributions to this report include:

1. Data collection and preprocessing: Collecting and cleaning a comprehensive dataset from multiple sources, including Transfermarkt and FBref [Section 2.1].

2. Feature engineering: Creating new features and transforming data into a suitable format for machine learning models [Section 2.2].

3. Machine learning model development: Experimenting with various machine learning algorithms, including linear regression, K-Nearest Neighbors, decision trees, random forest, gradient boosting, and neural networks [Section 3.3].

4. Optimization model formulation: Developing optimization models to allocate playing time while minimizing injury risk and maximizing team performance [Sections 4.3.1 and 4.4.1].

5. Result analysis and interpretation: Analyzing the results of the machine learning models and optimization process, drawing meaningful conclusions, and

identifying potential areas for improvement [Sections 4.3.3 and 4.4.2].

Key Findings and Contributions

The findings highlight the potential of data-driven approaches for injury prevention in football. The developed machine learning models demonstrate their ability to predict injury risk with reasonable accuracy, providing valuable insights for coaches and team managers. The optimization model offers a practical framework for allocating playing time while considering injury risk and team objectives.

This project contributes to the field of sports analytics by providing a comprehensive analysis of player rotation and injury management strategies. The findings can inform future research and practical applications in the realm of football and potentially other sports.

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Professor Lam Xin Yee, for her invaluable guidance, support, and mentorship throughout this project. Her expertise and encouragement have been instrumental in my success.

I would like to acknowledge the support provided by NUS Department of Mathematics, which made this project possible.

Finally, I would like to thank my family and friends for their unwavering support and encouragement. Their belief in me has been a constant source of motivation.

Abstract

This project investigates the problem of optimizing player rotation and management strategies in football to minimize long-term injuries. By leveraging machine learning and optimization techniques, we aim to develop a data-driven approach that can assist coaches in making informed decisions regarding player selection and playing time allocation.

We propose a novel prediction and optimization model that incorporates various factors, including player attributes, injury history, predicted days out due to injuries, and team performance metrics. The model aims to balance the competing objectives of minimizing injury risk while ensuring optimal team performance.

Through extensive experimentation and analysis, we demonstrate the effectiveness of our proposed approach in generating personalized recommendations for player rotation. Our results highlight the potential benefits of data-driven decision-making in reducing injuries and improving team performance.

Contents

List of Figures ix

List of Tables xi

1 Introduction **1**

1.1 Literature review 1

1.1.1 Injuries and their effects 2

1.1.2 Risk factors 2

1.1.3 Machine learning in football 3

1.2 Project objectives 3

1.3 Methodology 4

2 Data Collection **5**

2.1 Datasets 6

2.1.1 Player data 6

2.1.2 Team data 7

2.1.3 Refereeing data 8

2.2 Data processing and feature engineering 9

2.2.1 Data cleaning and standardization 9

2.2.2 Feature engineering 9

2.3 Key findings 10

2.3.1 Distribution 10

2.3.2 Injury rates 11

2.3.3 Correlations 13

3 Machine Learning 16

3.1 Exploratory data analysis 17

3.2 Evaluation metrics 18

3.2.1 Mean squared error 18

3.2.2 Mean absolute error 19

3.2.3 Normalised root mean squared error 19

3.3 Machine learning models 20

3.3.1 Linear regression 20

3.3.2 K-nearest neighbours 21

3.3.3 Decision tree 22

3.3.4 Random forest 23

3.3.5 Gradient boost 24

3.3.6 Voting regressor 25

3.3.7 Neural networks 26

3.4 Overall comparison 27

4 Optimization 30

4.1 Context 31

4.2 Assumptions 31

4.3 Using 'Total Days' 32

4.3.1 Objective function 33

4.3.2 Constraints 35

4.3.3 Results 36

4.4 Using 'Injury Count' and 'Average Injury Cost' 44

4.4.1 Objective function 44

4.4.2 Results 45
4.5 Conclusion on optimization models 47

5 Future Works 48
5.1 Data enrichment 48
5.2 Expanding the dataset 48
5.3 Validation and implementation 49
5.4 Broader applications 49
5.5 Conclusion 49

A Datasets 50
A.1 Player data 50
A.2 Team data 50
A.3 Referee data 51

B Machine Learning 52
B.1 Original dataset 52
B.2 Model comparison 52

C Optimization 53
C.1 Manchester City 53
C.2 Brighton & Hove Albion 55
C.3 Sheffield United 58
C.4 Real Madrid 60
C.5 CA Osasuna 63
C.6 Granada 66

References 70

List of Figures

2.1 Overall Age Distribution 11

2.2 Overall Position Distribution 12

2.3 Overall Injury Rates 12

2.4 Defensive Statistics 13

2.5 Possession Statistics 13

2.6 Passing Statistics 14

2.7 Shooting Statistics 14

2.8 Miscellaneous Statistics 14

2.9 Refereeing Statistics 14

3.1 PCA Analysis 17

3.2 Model Performance by total days 28

3.3 Model Performance by total games missed 28

3.4 Model Performance by injury count 29

3.5 Model Performance by average injury cost 29

4.1 Manchester City's Suggested Minutes vs. Weight when $\lambda = 0.9$ 37

4.2 Manchester City's Suggested Minutes vs. Weight when $\lambda = 0.2$ 38

4.3 Manchester City's Suggested Minutes vs. Weight as λ varies 39

4.4 Manchester City's Suggested Minutes vs. Total Days as λ varies 39

4.5 Kyle Walker's and Sergio Gomez's Suggested Minutes vs. Lambda 40

4.6 Brighton's Suggested Minutes vs. Weight as λ varies 41

4.7 Brighton's Suggested Minutes vs. Total Days as λ varies 41
4.8 Sheffield United's Suggested Minutes vs. Weight as λ varies 42
4.9 Sheffield United's Suggested Minutes vs. Total Days as λ varies 42
4.10 Jack Robinson's (Important) and Daniel Jebbison's (Less Important) Suggested Minutes vs. Lambda 43
4.11 Manchester City's Suggested Minutes vs. Weight as λ varies 46
4.12 Manchester City's Suggested Minutes vs. Average Injury Cost as λ varies 46

C.1 Simon Adingra's (Important) and Solly March's (Less Important) Suggested Minutes vs. Lambda 57
C.2 Real Madrid's Suggested Minutes vs. Weight as λ varies 61
C.3 Real Madrid's Suggested Minutes vs. Total Days as λ varies 61
C.4 Antonio Rüdiger's (Important) and Arda Güler's (Less Important) Suggested Minutes vs. Lambda 62
C.5 CA Osasuna's Suggested Minutes vs. Weight as λ varies 64
C.6 CA Osasuna's Suggested Minutes vs. Total Days as λ varies 64
C.7 Alejandro Catena's (Important) and Unai García's (Less Important) Suggested Minutes vs. Lambda 65
C.8 Granada's Suggested Minutes vs. Weight as λ varies 67
C.9 Granada's Suggested Minutes vs. Total Days as λ varies 68
C.10 Ignasi Miquel's (Important) and Jesús Vallejo's (Less Important) Suggested Minutes vs. Lambda 69

List of Tables

2.1 Sample of Manchester City's player data (part 1) 7

2.2 Sample of Manchester City's player data (part 2) 7

2.3 Sample of La Liga's team data collected 8

2.4 Sample of Bundesliga's refereeing data collected 8

2.5 Sample of Manchester City's player data with new columns 9

3.1 Linear regression results (original dataset) 21

3.2 Linear regression results (with 13 principle components) 21

3.3 K-nearest neighbours results (original dataset) 22

3.4 K-nearest neighbours results (with 13 principle components) 22

3.5 Decision tree results (original dataset) 23

3.6 Decision tree results (with 13 priniciple components) 23

3.7 Random forest results (original dataset) 24

3.8 Random forest results (with 13 principle components) 24

3.9 Gradient boosting results (original dataset) 25

3.10 Gradient boosting results (with 13 principle components) 25

3.11 Voting regressor results (original dataset) 26

3.12 Voting regressor results (with 13 principle components) 26

3.13 Neural network results (original dataset) 27

3.14 Neural network results (with 13 principle components) 27

4.1 Derivation of Min. playing minutes 35

4.2 Upper and lower bounds of minutes played by age group 35

4.3 Sample of Manchester City's results with $\lambda = 0.9$ 36

4.4 Sample of Manchester City's results with $\lambda = 0.2$ 37

4.5 Sample of Manchester City's results with $\lambda = 0.9$ 45

4.6 Sample of Manchester City's results with $\lambda = 0.2$ 45

C.1 Manchester City's results with $\lambda = 0.9$ 53

C.2 Manchester City's results with $\lambda = 0.2$ 54

C.3 Brighton & Hove Albion's results with $\lambda = 0.7$ 55

C.4 Brighton & Hove Albion's results with $\lambda = 0.1$ 56

C.5 Sheffield United's results with $\lambda = 0.8$ 58

C.6 Sheffield United's results with $\lambda = 0.2$ 59

C.7 Real Madrid's results with $\lambda = 0.6$ 60

C.8 Real Madrid's results with $\lambda = 0.2$ 60

C.9 CA Osasuna's results with $\lambda = 0.9$ 63

C.10 CA Osasuna's results with $\lambda = 0.4$ 63

C.11 Granada's results with $\lambda = 0.6$ 66

C.12 Granada's results with $\lambda = 0.1$ 67

Chapter 1

Introduction

The problem of long-term injuries in football has significant implications for both players and teams. Injuries can lead to lost playing time, decreased performance, and even career-ending consequences for players. For teams, injuries can disrupt squad balance, affect team morale, and impact overall performance.

While extensive research has been conducted on injury prevention from a medical perspective, there is a growing need for data-driven approaches that can provide objective insights and inform decision-making. Traditional methods often rely on subjective assessments and biological markers, which can be limited in their predictive power.

This project aims to explore the potential of data-driven methods to predict and prevent injuries in football. By leveraging player attributes, performance data, and other relevant factors, we seek to develop a systematic approach that can complement traditional medical assessments.

1.1 Literature review

This section explores the multifaceted nature of football injuries, their impact on players and teams, and underlying risk factors. Additionally, we investigate the

application of mathematical and machine learning techniques in sports analytics, highlighting their potential for predicting injuries and optimizing player management.

By critically evaluating existing studies and identifying gaps in the literature, this review establishes a strong foundation for the project and demonstrates the need for further exploration in the field of football injury prevention and player optimization.

1.1.1 Injuries and their effects

Previous research has consistently demonstrated the significant impact of injuries on football teams. Studies have found that lower injury rates are associated with higher league positions, increased points per match, and greater goal difference [Hägglund et al., 2013, Eirale et al., 2012]. Additionally, injuries can have substantial financial consequences, with teams losing millions of pounds due to injury-related performance decrements [Eliakim et al., 2020]. These findings underscore the importance of developing effective injury prevention strategies to enhance team success and minimize financial losses.

1.1.2 Risk factors

Studies have consistently demonstrated the prevalence of injuries in professional football, with an average of 14% of squads affected at any given time. These injuries can have substantial financial implications, as evidenced by the estimated €500,000 cost associated with a first-team player being out for one month [Ekstrand, 2013]. Research has identified various risk factors for injuries, including player attributes such as age, weight, and playing position, as well as training load and match schedules [Dut et al., 2023]. For example, studies have shown that players who are taller and heavier may be at increased risk of injuries, while excessive training hours can also

contribute to the likelihood of injuries. Understanding these factors is crucial for developing effective injury prevention strategies.

1.1.3 Machine learning in football

Recent studies have explored the application of machine learning techniques to predict injuries in football players. These studies have identified various biomarkers, such as muscle strength and stiffness, as reliable predictors of injury risk [Calderon-Díaz et al., 2023]. While our dataset may be limited to publicly available information, we aim to demonstrate the potential of data-driven approaches using more accessible features.

By leveraging data-driven approaches, it is possible to develop predictive models that can assist coaches in identifying players at risk and implementing preventive measures. For example, studies have successfully employed algorithms like XGBoost, Support Vector Machines (SVM), decision trees, and K-Nearest Neighbours (KNN) to predict injuries based on player attributes and performance data [Rossi et al., 2017]. However, the accuracy of these models can be influenced by the quality and quantity of data available. As more data is collected over time, the predictive power of these models is expected to improve.

1.2 Project objectives

Given the significant impact of injuries on football teams and the potential benefits of data-driven approaches, a natural question that arises is: **How can data-driven models and optimization techniques be applied to reduce long-term injury risk through optimal player rotation and management?**

This project aims to address this question by developing a predictive model and

exploring strategies for player rotation that can minimize injuries while maintaining team performance.

1.3 Methodology

We utilize a combination of machine learning techniques and optimization methods to develop a predictive model and player rotation strategies.

In line with our methodology, the following chapters of this report are organized as follows:

1. Data collection: Describes the sources of data, including player attributes, match data, and team-level factors.
2. Machine learning models: Details the algorithms used for injury prediction and comparison of results using a set of evaluation metrics.
3. Optimization: Details the formulation of a suitable optimization problem and discusses the sensibility and reliability of the results.
4. Future works: Highlights possible extensions for this project.

Finally, this project utilizes Python as the primary programming language and leverages the computational resources provided by Kaggle to facilitate machine learning and optimization tasks.[1]

[1]The source code is available at https://github.com/tsumian/MA4199-Injury-Prediction.

Chapter 2

Data Collection

To conduct a comprehensive analysis of injury prevention in football, we rely on a robust dataset encompassing player, team, and referee information. These data sources are essential for understanding the complex interplay of factors that contribute to injuries and for developing effective prevention strategies. While existing databases may provide some relevant information, we found it necessary to collect our own data using web scraping techniques to ensure the accuracy and completeness of the dataset.

Due to time and resource constraints, the current analysis focuses on data from the top 5 European leagues during the 2023/ 2024 season.[1] While this dataset provides a valuable foundation for the project, it is important to note that expanding the data to include multiple seasons and leagues could potentially enhance the reliability and robustness of our findings. Future studies may benefit from incorporating a broader dataset to gain a more comprehensive understanding of injury patterns and risk factors.

[1] Due to resource constraints, we were only able to look at men's leagues and chose to focus on outfield players only.

2.1 Datasets

The cornerstone of this research lies in a comprehensive dataset encompassing player, team, and referee information. To capture the most up-to-date and relevant data, we employed web scraping techniques to extract information from reputable websites like Transfermarkt and FBref. These platforms offer a wealth of player attributes, historical performance metrics, and team statistics essential for building robust machine learning models. Player data, including age, weight, injury history, and playing position, provides crucial insights into individual injury susceptibility. Team data allows us to understand the broader context about the intensity level of each team. Refereeing data, while potentially less directly linked to injury risk, could offer valuable clues regarding potential foul patterns that might contribute to injuries. By meticulously collecting and integrating data from these diverse sources, we lay the groundwork for developing powerful machine learning models capable of predicting injury risk and informing effective player rotation strategies.

2.1.1 Player data

To build a comprehensive dataset for our injury prediction models, we leveraged the wealth of information available on Transfermarkt.com. This platform provided essential player attributes[2] including:

- Basic demographics: Name, birthday, height
- Club and league affiliation: Current club, league, and previous clubs
- Performance metrics: Appearances, injuries, total days missed (due to injuries), and minutes played

[2]Full set of data columns for player data can be found in Appendix A.1

Table 2.1: Sample of Manchester City's player data (part 1)

Name	Birthday	Height	League	Club	Position	Appearances	Minutes Played Total 23/24:
Ederson	Aug 17, 1993 (30)	1,88 m	Premier League	Man City	Goalkeeper	43	3.718'
Stefan Ortega	Nov 6, 1992 (31)	1,85 m	Premier League	Man City	Goalkeeper	20	1.621'
Rúben Dias	May 14, 1997 (27)	1,87 m	Premier League	Man City	Centre-Back	45	3.817'
Manuel Akanji	Jul 19, 1995 (28)	1,88 m	Premier League	Man City	Centre-Back	48	3.861'
Nathan Aké	Feb 18, 1995 (29)	1,80 m	Premier League	Man City	Centre-Back	44	3.053'

Table 2.2: Sample of Manchester City's player data (part 2)

Name	Injuries	Total Days	Total Games Missed
Ederson	{'Eyebow fracture': {'Days': '18 days', 'Games missed': '3'}, 'Muscle injury': {'Days': '26 days', 'Games missed': '4'}, 'Knee injury': {'Days': '7 days', 'Games missed': '1'}}	51 days	8
Stefan Ortega	{}	0 days	0
Rúben Dias	{}	0 days	0
Manuel Akanji	{'unknown injury': {'Days': '24 days', 'Games missed': '4'}}	24 days	4
Nathan Aké	{'Calf injury': {'Days': '12 days', 'Games missed': '4'}}	12 days	4

Raw data from Transfermarkt often requires preprocessing to ensure its suitability for machine learning algorithms. In this case, players with incomplete or missing attributes (for e.g. made no appearances in the 2023/ 2024 season) were excluded from the dataset to maintain data quality. By carefully collecting and preprocessing this player data (Table 2.1 and Table 2.2), we were able to create a solid foundation for our injury prediction models, enabling us to explore the relationships between player attributes and injury risk.

2.1.2 Team data

To understand the relationship between team playing style and injury risk, we analyzed a comprehensive set of team-level statistics.[3] These statistics (Table 2.3),

[3]Full set of data columns for team data can be found in Appendix A.2

Table 2.3: Sample of La Liga's team data collected

Team	Players Tackled	Dribblers Tackled	Blocks	Interceptions	...	Fouls Committed	Fouls Drawn
Real Madrid	557	225	401	296	...	368	523
Barcelona	571	243	363	257	...	412	469
Girona	544	274	378	257	...	406	471
Atlético Madrid	654	327	460	256	...	438	447
Athletic Club	581	305	419	278	...	511	428

scraped from FBref, encompass shooting, passing, goal creation, defensive actions, possession, and miscellaneous metrics, providing a good picture of each team's playing style.

Analyzing these team-level statistics enables us to identify potential trends and associations between playing styles and injury risk, which can inform future research and strategies for injury prevention.

2.1.3 Refereeing data

To gain a more comprehensive understanding of the factors influencing injury risk, we incorporated refereeing data[4] into our analysis. By examining statistics from Table 2.4, we can roughly assess the level of protection afforded to each team.

Table 2.4: Sample of Bundesliga's refereeing data collected

Team	Fouls per 90	Fouls per Tackle	Penalties per 90	Yellow Cards per 90	Red Cards per 90	Fouls Committed	Fouls Drawn
Leverkusen	10.41	0.54	0.24	1.85	0.15	368	523
Stuttgart	11.56	0.65	0.24	2.21	0.09	412	469
Bayern Munich	9.38	0.44	0.15	1.76	0.09	406	471
RB Leipzig	11.91	0.71	0.21	2.59	0.15	438	447
Dortmund	11.94	0.63	0.18	2.06	0.03	511	428

While these statistics provide a surface-level understanding of refereeing intensity,

[4]Full set of data columns for team data can be found in Appendix A.3

they can offer valuable insights into the physical demands placed on teams. Teams facing more fouls and physical challenges may be at a higher risk of injuries. By analyzing these refereeing metrics, we can explore potential correlations between the level of physical contact and the incidence of injuries.

2.2 Data processing and feature engineering

To prepare our data for machine learning analysis, we conducted a comprehensive data processing and feature engineering pipeline.

Table 2.5: Sample of Manchester City's player data with new columns

Name	Categorized Position	Total Days Out	Total Games Missed	Age	Categorized Injuries	Injury Count	Average Injury Cost (million)
Rúben Dias	Defender	0	0	27	[]	0	0
Manuel Akanji	Defender	24	4	28	['Other']	1	0.17
Nathan Aké	Defender	12	4	29	['Calf/Shin']	1	0.17
John Stones	Defender	52	12	30	['Thigh/Hip']	1	0.17
Josko Gvardiol	Defender	22	6	22	['Ankle']	1	0.17

2.2.1 Data cleaning and standardization

We began by addressing any inconsistencies or missing values in the dataset. This involved identifying and handling missing data, such as players with no appearances in the season. Additionally, we standardized data formats, ensuring that all attributes were in a consistent numerical format suitable for machine learning algorithms.

2.2.2 Feature engineering

To enrich the dataset and extract meaningful information, we created new features through feature engineering. One key feature was the creation of generalized playing positions, grouping specific positions into broader categories like "Defender," "Midfielder," and "Forward." This simplified the analysis and allowed for more meaningful

comparisons across different positions.

Furthermore, we calculated additional injury metrics, such as "Total Days Out" and "Total Games Missed," to quantify the impact of injuries on players and teams. These features provided valuable insights into the severity and duration of injuries.

To estimate the financial cost of injuries, we leveraged the Howden Group Injury Index [Howden, 2023]. By referencing the average injury costs for different positions and leagues, we calculated an estimated injury cost for each player based on their injury count. This provided a quantitative measure of the economic impact of injuries.

By effectively addressing data quality issues and creating relevant features, we laid the foundation for building accurate and informative machine learning models.

2.3 Key findings

To gain a deeper understanding of the collected data and assess its suitability for machine learning, we conducted a thorough exploratory data analysis. By visualizing the distribution of key variables and identifying potential relationships, we aimed to uncover valuable insights and assess the feasibility of building predictive models.

2.3.1 Distribution

The majority of players in our dataset fall within the age range of 22 to 33 from Table 2.1. This suggests that the prime playing years for professional football players are concentrated within this age bracket.

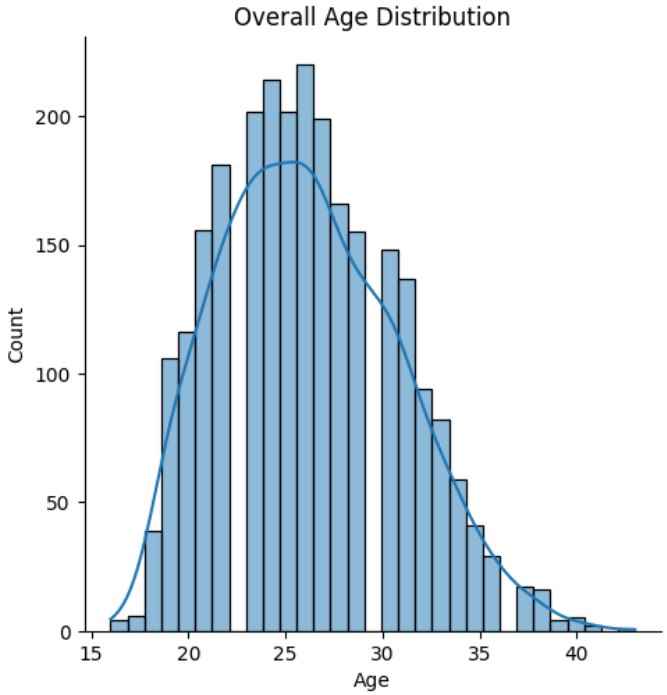


Figure 2.1: Overall Age Distribution

This makes sense because players in this age range are typically at their physical peak, with well-developed muscles and optimal athletic abilities. This peak performance may also come with increased demands on the body, potentially influencing injury risk.

Next, the majority of players in our dataset are defenders, followed by midfielders and forwards as shown in Figure 2.2. The distribution of players across positions provides valuable insights into the potential injury risks associated with different playing styles and physical demands.

2.3.2 Injury rates

A significant portion of the players in our dataset (53.5% from Figure 2.3) experienced one or more injuries during the season. This indicates a substantial number of injury

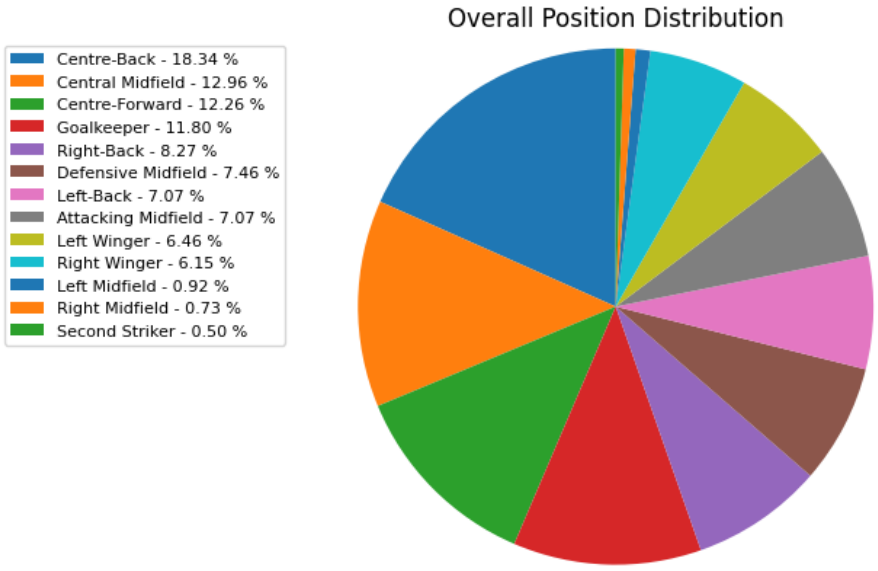


Figure 2.2: Overall Position Distribution

cases, providing a sufficient sample size for training and analysis.

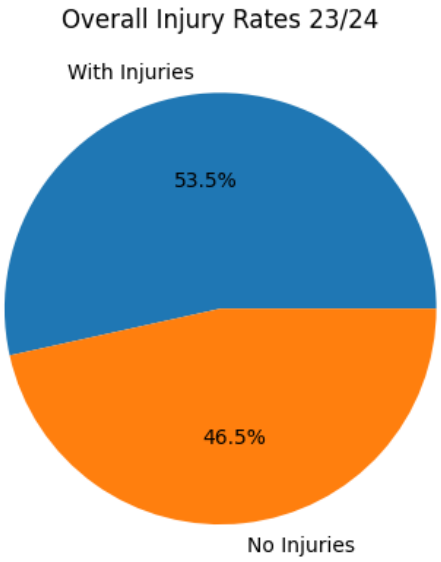


Figure 2.3: Overall Injury Rates

Having a representative sample of players with injury experience is crucial for

developing accurate predictive models. With a substantial number of injured players, we can effectively train our models to identify patterns and risk factors associated with injuries. This avoids the potential bias that could arise from under-representation of injured players.

2.3.3 Correlations

To understand the relationships between team performance and injury risk, we conducted a comprehensive correlation analysis. By examining the correlations between various features of our team dataset (Figures 2.4 to 2.9), we sought to identify significant associations that could inform our injury prediction models.

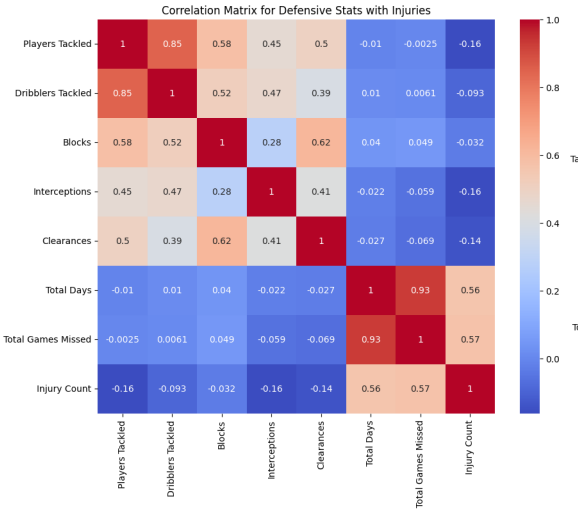


Figure 2.4: Defensive Statistics

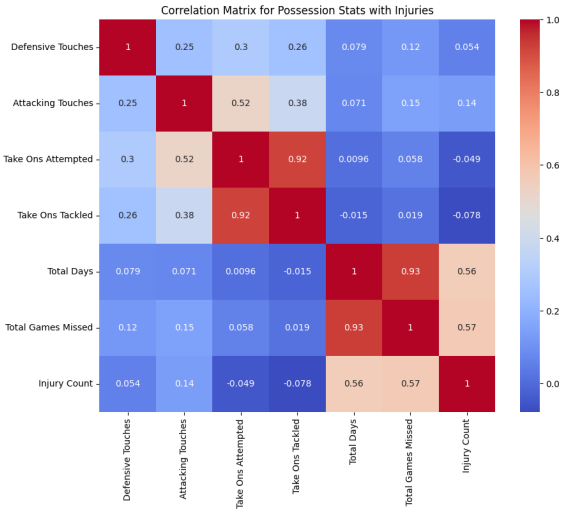


Figure 2.5: Possession Statistics

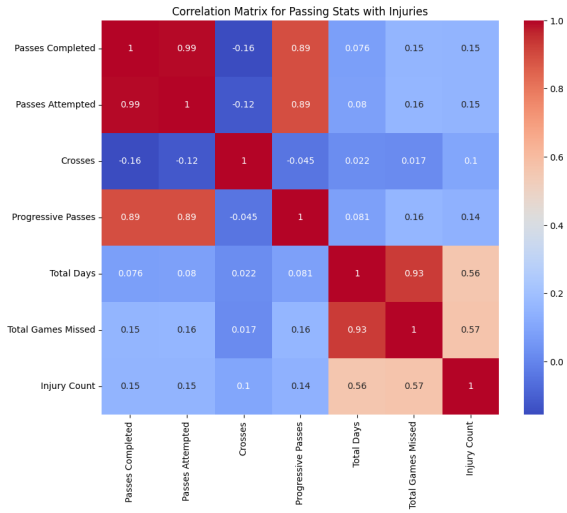


Figure 2.6: Passing Statistics

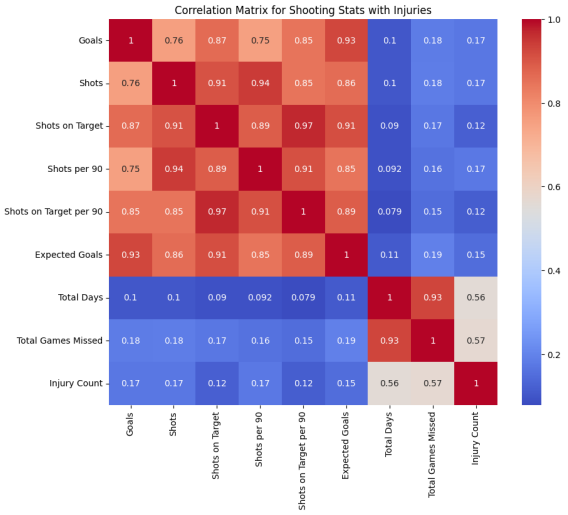


Figure 2.7: Shooting Statistics

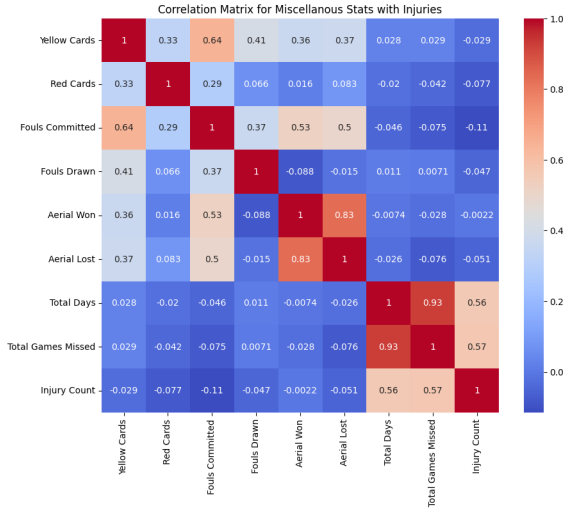


Figure 2.8: Miscellaneous Statistics

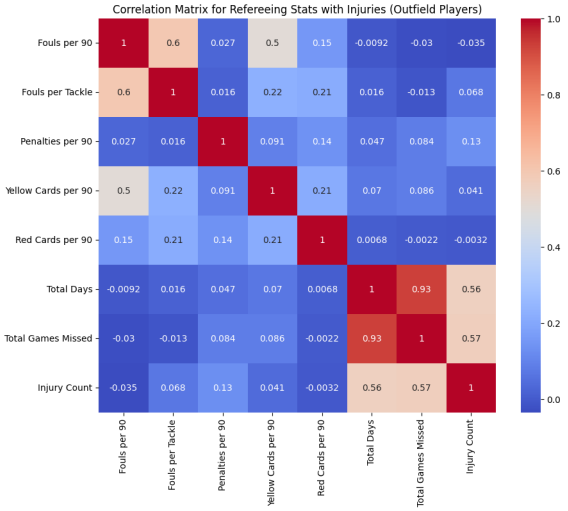


Figure 2.9: Refereeing Statistics

Unsurprisingly, the various play-style statistics such as passing statistics (from Figure 2.6) and shooting statistics (from Figure 2.7) were highly correlated with each other, indicating a consistent relationship between different shooting metrics. Similarly, injury indicators demonstrated strong correlations, reflecting the interconnected nature of various injury-related factors.

However, when examining the correlations between different categories of play-styles

and injury indicators, we observed a lack of strong linear relationships. This suggests that traditional play-style metrics may not be directly predictive of injuries.

While correlation analysis provides valuable insights into the relationships between variables, it's important to note that correlation does not imply causation. Further analysis, such as machine learning modeling, is necessary to explore deeper patterns and relationships within the data, potentially revealing factors that were not apparent through simple correlation measures.

In the following section, we will explore the algorithms and techniques that we employed to build predictive models capable of identifying players at risk.

Chapter 3

Machine Learning

This chapter delves into the application of machine learning techniques to predict injury risk in football. We focus on regression models, where the goal is to predict continuous variables such as **Total Days** (out due to injuries), **Total Games Missed** (due to injuries), **Injury Count**, and **Average Injury Cost**. By employing machine learning, we aim to uncover hidden patterns and relationships within the data that can inform effective injury prevention strategies.

Given the data collected, we can represent each player as a feature vector $\mathbf{x}$, where $\mathbf{x}$ is a collection of relevant attributes $(x_1, x_2, \ldots, x_n)^T$, for example:

$$
\begin{aligned}
x_1 &= \text{Height} \\
x_2 &= \text{Appearances} \\
x_3 &= \text{Age} \\
&\vdots \\
x_{n-1} &= \text{Yellow Cards per 90 minutes} \\
x_n &= \text{Red Cards per 90 minutes}
\end{aligned}
$$

The predicted variables, denoted by y, are continuous.

In addition to regression models, we also explored classification models like Random Forest and Decision Trees. While these models are typically used for categorical outcomes, their performances in our context were surprisingly promising.

By combining regression and classification techniques, we aim to gain a comprehensive understanding of injury risk and develop robust predictive models that can assist coaches and teams in making informed decisions regarding player management and injury prevention.

3.1 Exploratory data analysis

Before training the models, we conducted exploratory data analysis to understand the characteristics of our dataset and identify potential features relevant for injury prediction.

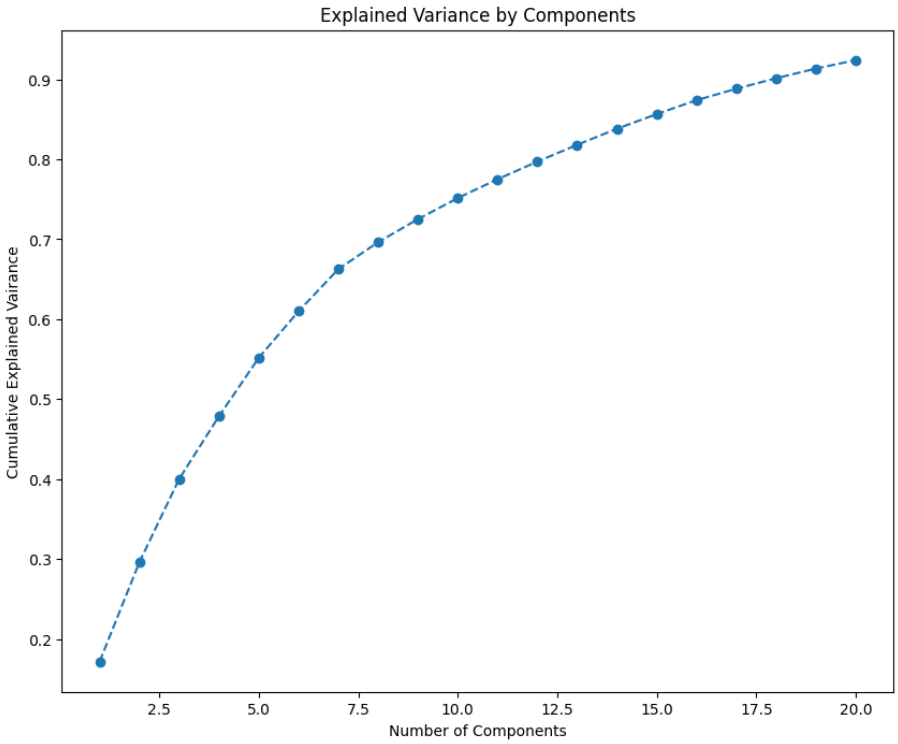


Figure 3.1: PCA Analysis

Principal Component Analysis (PCA) is a powerful technique used to reduce the dimensionality of a dataset while preserving the most important information. It transforms the original features into a new set of uncorrelated features called principal components. These components are ordered by their explained variance, with the first component capturing the most variance in the data.

From Figure 3.1, we need to keep about 13 principle components to retain about 80% of the variance of our dataset. By doing so, we aim to capture the most significant patterns and relationships within the data. This reduces the dimensionality of the dataset without sacrificing too much information.

3.2 Evaluation metrics

When evaluating the performance of machine learning models, it is crucial to employ standardized metrics that provide a common ground for comparison. These metrics allow us to objectively assess the accuracy, reliability, and effectiveness of different models.

3.2.1 Mean squared error

Mean Squared Error (MSE) is a widely used metric for evaluating the performance of regression models. It measures the average squared difference between the predicted values and the true values. A lower MSE indicates that the model's predictions are closer to the true values, suggesting better overall performance. The formula is

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

where y_i is the true value, $\hat{y}_i$ is the predicted value and n is the number of data points that we have.

Note that MSE is sensitive to outliers, as squaring the errors gives greater weight to larger differences. This can be a drawback if outliers are not representative of the true data distribution.

3.2.2 Mean absolute error

Mean Absolute Error (MAE) is another commonly used metric for evaluating regression models. It measures the average absolute difference between the predicted values and the actual values. A lower MAE indicates a better model fit, with predictions that are closer to the true values. The formula is

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

where y_i is the true value, $\hat{y}_i$ is the predicted value and n is the number of data points that we have.

Compared to MSE, MAE is less sensitive to outliers. This is because it takes the absolute value of the differences, which reduces the impact of large errors. This makes MAE a more robust metric in cases where the data contains outliers or extreme values.

3.2.3 Normalised root mean squared error

Normalized Root Mean Squared Error (NRMSE) is a variant of MSE that is scaled by the range of the target variable. A lower NRMSE indicates a better model fit, with predictions that are closer to the true values relative to the range of the target

variable. The formula is

$$NRMSE = \frac{1}{n}\sqrt{\sum_{i=1}^{n}\left(\log\frac{\hat{y}_i+1}{y_i+1}\right)^2}$$

where y_i is the true value, $\hat{y}_i$ is the predicted value and n is the number of data points that we have.

NRMSE inherits the sensitivity to outliers from MSE. However, by scaling the errors by the range of the target variable, it provides a more interpretable measure of model performance, especially when comparing models with different target variables.

3.3 Machine learning models

In this section, we will discuss the models that we experimented with and the results we obtained for each model with respect to the evaluation metrics above.

When training these models, we used different datasets[1] with different number of features: original dataset, dataset with 2, 3 and 13 principle components. This helped us gain insights into the impact of dimensionality reduction on model performance.

3.3.1 Linear regression

Linear regression serves as a baseline model for our analysis, establishing a benchmark against which we can compare more complex models.

[1]The original dataset can be found in Appendix B.1

Table 3.1: Linear regression results (original dataset)

Model	Target	Mean Square Error	Mean Absolute Error	Normalized Root Mean Square Error
Linear Regression	Total Days	3030.931202	36.9327342	1.783017462
Linear Regression	Total Games Missed	56.79102307	35.70268246	1.568959912
Linear Regression	Injury Count	1.074276793	36.57511234	1.830834629
Linear Regression	Average Injury Cost	0.0185591036	36.86357273	1.759536232

Table 3.2: Linear regression results (with 13 principle components)

Model	Target	Mean Square Error	Mean Absolute Error	Normalized Root Mean Square Error
Linear Regression (PCA n = 13)	Total Days	3569.128313	40.35383081	3.514180776
Linear Regression (PCA n = 13)	Total Games Missed	68.26398353	36.21451294	2.996935512
Linear Regression (PCA n = 13)	Injury Count	1.175746351	36.64422023	2.588815005
Linear Regression (PCA n = 13)	Average Injury Cost	0.01944930092	36.87100987	2.457467557

From Tables 3.1 and 3.2, simple linear regression generally produces satisfactory results across all metrics for each predicted variable. Although the MSE might appear relatively large, this can be attributed to the influence of outliers. Note that MSE for 'Total Days' and 'Total Games Missed' are larger because these predicted values could be large, hence squaring the difference will definitely result in larger values.

Interestingly, the original dataset slightly outperformed the PCA-reduced dataset with 13 principal components. This could be due to the loss of subtle relationships or nuances in the data when dimensionality is reduced.

3.3.2 K-nearest neighbours

K-Nearest Neighbors (KNN) is a versatile machine learning algorithm that can be used for both classification and regression tasks. In regression tasks, KNN predicts a continuous value by taking the average of the target variable values of its k nearest neighbors.

Table 3.3: K-nearest neighbours results (original dataset)

Model	Target	Mean Square Error	Mean Absolute Error	Normalized Root Mean Square Error
KNN	Total Days	5393.664488	45.26797386	1.463407457
KNN	Total Games Missed	109.2156863	6.62745098	1.315849147
KNN	Injury Count	2.289760349	1.065359477	1.282483154
KNN	Average Injury Cost	0.03995729847	0.134379085	1.248227883

Table 3.4: K-nearest neighbours results (with 13 principle components)

Model	Target	Mean Square Error	Mean Absolute Error	Normalized Root Mean Square Error
KNN (PCA n = 13)	Total Days	5642.37037	44.61873638	1.300773248
KNN (PCA n = 13)	Total Games Missed	122.6100218	6.649237473	1.277926806
KNN (PCA n = 13)	Injury Count	2.148148148	1.002178649	1.286455355
KNN (PCA n = 13)	Average Injury Cost	0.03583703704	0.1247494553	1.250781756

From Tables 3.3 and 3.4, we observe that KNN outperformed linear regression in terms of MAE, indicating better robustness to outliers, while its MSE was worse. KNN, being a distance-based algorithm, can be influenced by outliers, leading to larger errors in MSE.

The similarity in NRMSE between KNN and linear regression suggests that the scaling of errors is comparable between the two models. This indicates that the relative performance differences observed in MSE and MAE are primarily due to the sensitivity of these metrics to outliers.

3.3.3 Decision tree

While decision trees are primarily known for their use in classification tasks, they can also be adapted for regression problems. In regression, decision trees partition the data into subsets based on feature values and predict a continuous target variable for each subset.

Table 3.5: Decision tree results (original dataset)

Model	Target	Mean Square Error	Mean Absolute Error	Normalized Root Mean Square Error
Decision Tree	Total Days	5340.220044	45.03485839	1.3405706
Decision Tree	Total Games Missed	102.9324619	6.348583878	1.189878233
Decision Tree	Injury Count	2.047930283	1.015250545	1.238192476
Decision Tree	Average Injury Cost	0.0367130719	0.1246840959	1.208593826

Table 3.6: Decision tree results (with 13 priniciple components)

Model	Target	Mean Square Error	Mean Absolute Error	Normalized Root Mean Square Error
Decision Tree (PCA n = 13)	Total Days	5935.568627	46.40087146	1.246360727
Decision Tree (PCA n = 13)	Total Games Missed	113.0544662	6.77124183	1.353566799
Decision Tree (PCA n = 13)	Injury Count	1.928104575	0.9389978214	1.222831813
Decision Tree (PCA n = 13)	Average Injury Cost	0.03310653595	0.1235511983	1.224697517

Tables 3.5 and 3.6 highlight that the results are similar to those in KNN. The similarity in MAE and NRMSE suggests that decision trees, like KNN, are relatively robust to outliers and provide comparable predictions in terms of overall accuracy and scaling of errors.

The higher MSE observed with decision trees compared to linear regression could be attributed to the potential for overfitting, especially in complex decision tree structures. Overfitting can lead to the model memorizing the training data too closely, resulting in poor generalization performance.

3.3.4 Random forest

Random Forest is an ensemble learning algorithm that combines multiple decision trees to improve prediction accuracy and reduce overfitting.

Table 3.7: Random forest results (original dataset)

Model	Target	Mean Square Error	Mean Absolute Error	Normalized Root Mean Square Error
Random Forest	Total Days	3039.627834	36.37949891	1.816287218
Random Forest	Total Games Missed	55.77260349	5.156732026	1.571665664
Random Forest	Injury Count	1.055317865	0.7923965142	1.641215136
Random Forest	Average Injury Cost	0.01801293031	0.1027026144	1.647714571

Table 3.8: Random forest results (with 13 principle components)

Model	Target	Mean Square Error	Mean Absolute Error	Normalized Root Mean Square Error
Random Forest (PCA n = 13)	Total Days	5567.24183	44.99782135	1.244698388
Random Forest (PCA n = 13)	Total Games Missed	116.0196078	6.830065359	1.300644878
Random Forest (PCA n = 13)	Injury Count	2.008714597	0.962962963	1.204061873
Random Forest (PCA n = 13)	Average Injury Cost	0.03143681917	0.1204575163	1.215501504

An interesting observation from Table 3.7 is that this model outperformed all previous models mentioned for the original dataset. However, with reduced dimensionality (Table 3.8), it had similar performances as KNN and decision tree.

Random forest's superior performance on the original dataset highlights its ability to extract meaningful patterns from complex data. However, the performance degradation on the reduced dimensionality datasets suggests that the original features may contain important information that is lost during the dimensionality reduction process. This underscores the importance of carefully considering the trade-offs between dimensionality reduction and model performance.

3.3.5 Gradient boost

Gradient boost is another ensemble learning technique that combines multiple weak learners to create a strong predictive model. It works by iteratively training decision trees to predict the residuals of the previous ensemble, focusing on areas where the current model is making errors. This process helps the model gradually improve its performance over time.

Table 3.9: Gradient boosting results (original dataset)

Model	Target	Mean Square Error	Mean Absolute Error	Normalized Root Mean Square Error
Gradient Boosting	Total Days	3423.009121	36.54281557	1.699736427
Gradient Boosting	Total Games Missed	62.48764771	5.180833406	1.508073771
Gradient Boosting	Injury Count	1.147289281	0.8196845944	1.575486844
Gradient Boosting	Average Injury Cost	0.01863637064	0.1022556757	1.547830807

Table 3.10: Gradient boosting results (with 13 principle components)

Model	Target	Mean Square Error	Mean Absolute Error	Normalized Root Mean Square Error
Gradient Boosting (PCA n = 13)	Total Days	3451.384901	38.96538616	1.803568884
Gradient Boosting (PCA n = 13)	Total Games Missed	68.44115946	5.667885497	1.683646025
Gradient Boosting (PCA n = 13)	Injury Count	1.241811039	0.8366481227	1.624394201
Gradient Boosting (PCA n = 13)	Average Injury Cost	0.0206566609	0.1074367941	1.541002023

From Tables 3.9 and 3.10, the results are similar to linear regression, even with dimensionality reduction. This could suggest that the underlying relationships between the features and the target variables may be relatively linear. However, it is worth exploring gradient boosting as it can potentially capture non-linear relationships and improve model performance in certain scenarios.

3.3.6 Voting regressor

Voting Regressor is yet another ensemble method that combines the predictions of multiple base models. In this case, we combined linear regression, gradient boost, and random forest. The final prediction is made by aggregating the predictions of the individual models. This can help to improve overall accuracy and reduce the risk of overfitting.

Table 3.11: Voting regressor results (original dataset)

Model	Target	Mean Square Error	Mean Absolute Error	Normalized Root Mean Square Error
Voting Regressor	Total Days	2993.668443	35.23294078	1.823565994
Voting Regressor	Total Games Missed	54.32948819	5.041526781	1.624642859
Voting Regressor	Injury Count	1.025086209	0.7786292923	1.769106612
Voting Regressor	Average Injury Cost	0.01713976029	0.09957030844	1.725589935

Table 3.12: Voting regressor results (with 13 principle components)

Model	Target	Mean Square Error	Mean Absolute Error	Normalized Root Mean Square Error
Voting Regressor (PCA n = 13)	Total Days	3305.422299	38.35535264	2.396646356
Voting Regressor (PCA n = 13)	Total Games Missed	63.7722389	5.603110044	2.163806635
Voting Regressor (PCA n = 13)	Injury Count	1.149268535	0.8193088437	1.992938905
Voting Regressor (PCA n = 13)	Average Injury Cost	0.01882809941	0.1053186931	1.890503114

As seen from Table 3.11, this model produces the best result so far in terms of MSE and MAE. This could be because voting regressor effectively combines the strengths of different models, potentially offsetting the weaknesses of individual algorithms, reduces overfitting, hence leading to improved generalization performance.

3.3.7 Neural networks

Neural networks are powerful machine learning models that can learn complex patterns and relationships within data. However, their performance can be sensitive to various factors, including architecture, hyperparameters, and data quality.

To optimize the performance of our neural network model, we conducted a comprehensive hyperparameter tuning process. We experimented with various loss functions, including Mean Squared Error (MSE), Mean Absolute Error (MAE), Huber Loss, and log-cosh, to evaluate the model's sensitivity to outliers and overall accuracy. Additionally, we explored different activation functions, such as ReLU, Sigmoid, and Tanh, to capture non-linear relationships within the data. To optimize the learning process, we compared various optimizers, including Stochastic Gradient Descent

(SGD) and Adam. Furthermore, we experimented with different numbers of hidden layers (1 and 2) and epochs (50, 100, and 200) to determine the optimal network architecture and training duration.

By systematically evaluating the impact of these hyperparameters, we were able to identify the best-performing model configuration for each evaluation metric and each predicted variable, resulting in the outcomes presented in the following tables.

Table 3.13: Neural network results (original dataset)

Model	Target	Mean Square Error	Mean Absolute Error	Normalized Root Mean Square Error
Neural Network	Total Days	4331.870115	35.09874388	10.03603321
Neural Network	Total Games Missed	102.6343262	5.295950077	92.27793471
Neural Network	Injury Count	2.014282401	0.9376731022	13.19282619
Neural Network	Average Injury Cost	0.04246237371	0.1382144151	1.091839017

Table 3.14: Neural network results (with 13 principle components)

Model	Target	Mean Square Error	Mean Absolute Error	Normalized Root Mean Square Error
Neural Network (PCA n = 13)	Total Days	4545.480899	35.82502081	16.28914305
Neural Network (PCA n = 13)	Total Games Missed	103.1732673	5.30270803	121.8907564
Neural Network (PCA n = 13)	Injury Count	1.99575218	0.9435354433	10.57424559
Neural Network (PCA n = 13)	Average Injury Cost	0.03558144202	0.1331184142	1.145902536

From Tables 3.13 and 3.14, we see that neural networks performed worse than linear regression in terms of MSE and NRMSE, but better in terms of MAE. This could suggest that the neural network may be struggling to capture the underlying relationships in the dataset.

3.4 Overall comparison

To establish a common ground for comparison across all models[2], we primarily focused on MSE as our main evaluation metric. Based on MSE, the voting regressor

[2]Comparison of all models trained can be found in Appendix B.2

model without dimensionality reduction consistently outperformed the other models we considered for all predicted variables as shown in Figures 3.2, 3.3, 3.4 and 3.5. This suggests that the ensemble approach of combining multiple models was particularly effective in capturing the underlying relationships within our data and achieving accurate predictions.

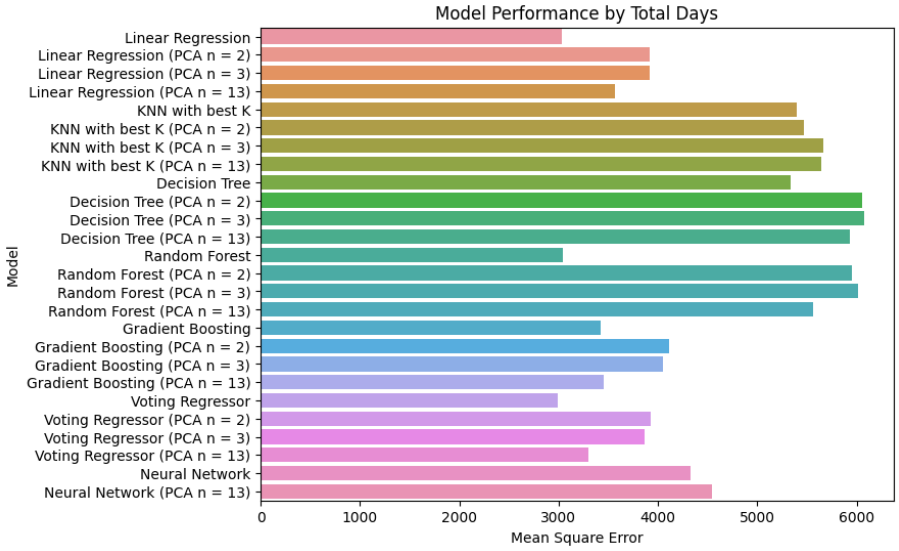


Figure 3.2: Model Performance by total days

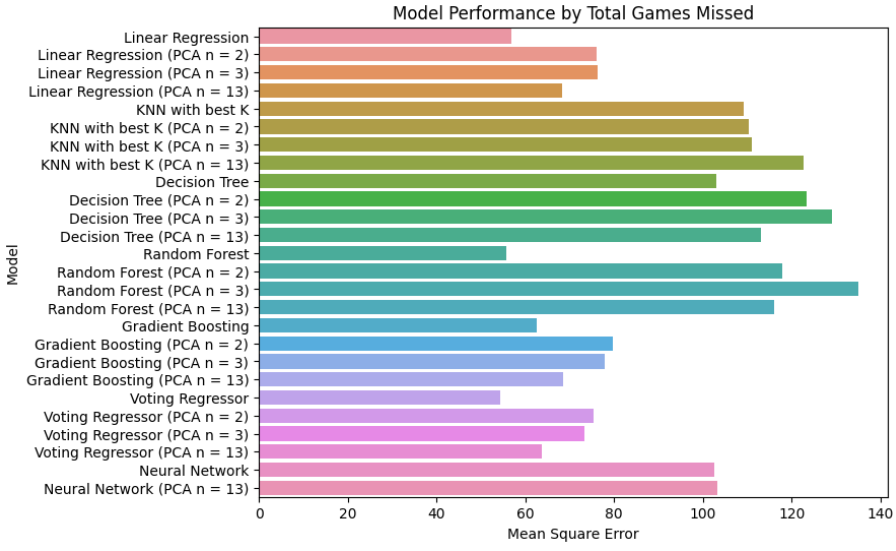


Figure 3.3: Model Performance by total games missed

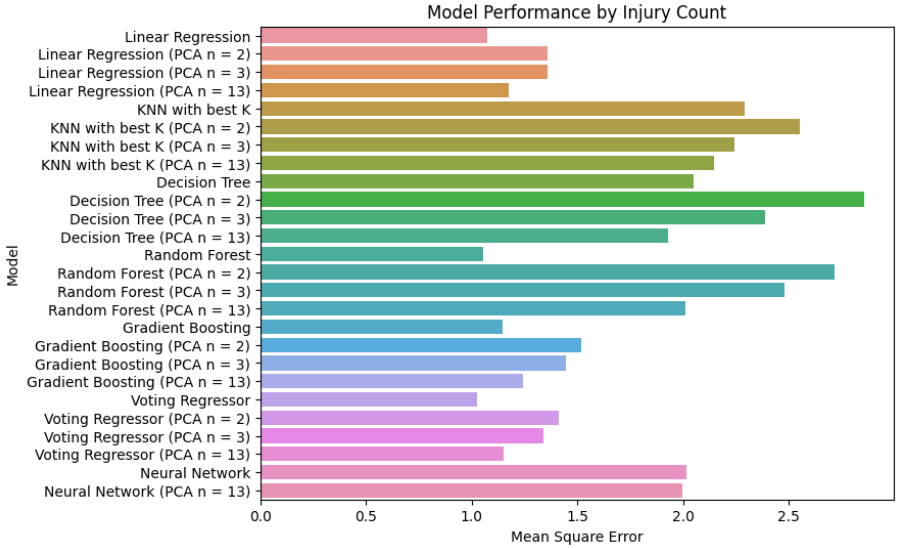


Figure 3.4: Model Performance by injury count

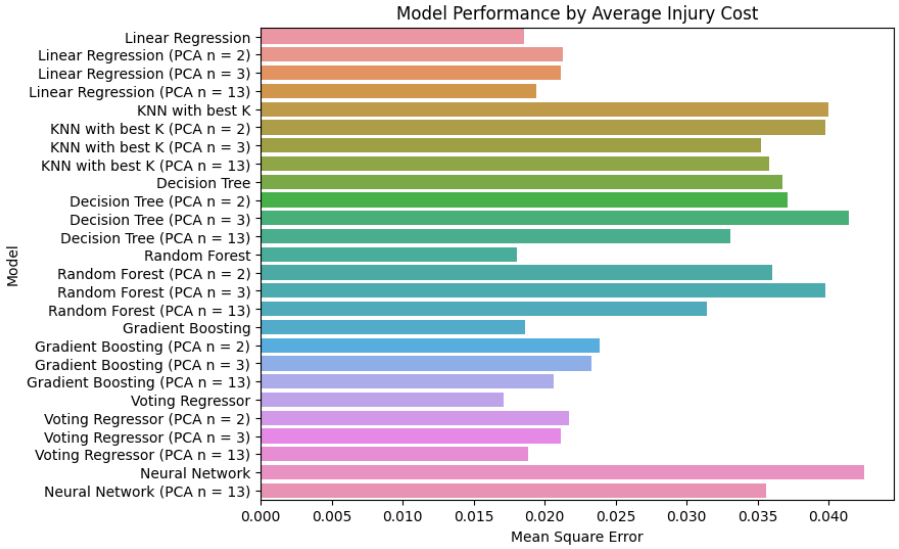


Figure 3.5: Model Performance by average injury cost

In the following chapter on optimization, we will leverage the voting regressor as our primary predictive model to produce the estimated injury risk for each player. These predicted values will serve as inputs to our optimization problem, where we aim to allocate playing time while minimizing injury risk and maximizing team performance.

Chapter 4

Optimization

This chapter focuses on formulating and solving optimization problems to determine the optimal playing time allocation for individual players within a team. By leveraging the predicted variables of '**Total Days**', '**Injury Count**' and '**Average Injury Cost**', we aim to minimize injury risk while maximizing team performance.

There are some key considerations when it comes to formulating the optimization problem:

- Individual team optimization: The optimization problem is tailored to a specific team, considering the unique characteristics and objectives of that team.
- Variable team sizes: Different teams may have varying numbers of players, influencing the allocation of playing time.
- Season-specific goals: Teams may have diverse objectives for a season, ranging from winning all games to simply avoiding relegation. These objectives can impact the prioritization of injury prevention and performance optimization.

While the optimization framework provides a valuable tool for decision-making, it's important to acknowledge the limitations. Validating the results in real-world scenarios is challenging, as implementing the suggested playing time allocations would require significant coordination and cooperation. However, sensitivity analysis

can help assess the reasonableness and feasibility of the proposed solutions.

By considering these factors and addressing the challenges, we aim to develop a robust optimization framework that can provide valuable insights for coaches and managers in managing player rotation and minimizing injury risk.

4.1 Context

To provide context for our analysis, it's essential to understand the typical structure and dynamics of a football season. A professional football team typically consists of around 25 players, forming a squad that is managed by a manager.

The number of matches played in a season varies across different leagues, but generally ranges from 40 to 70 matches, especially with the recent changes to the UEFA Champions League format. Each match involves two teams of 11 players, with one goalkeeper and ten outfield players. The game is played for 90 minutes, with additional stoppage time added at the referee's discretion.

To manage player workload and tactical adjustments, teams are allowed to make up to five substitutions during a match. This means that not all players will necessarily play the full 90 minutes, allowing for strategic rotations and rest periods.

Understanding these fundamental aspects of football provides a valuable foundation for appreciating the complexities of player rotation and injury management in the sport.

4.2 Assumptions

To simplify our analysis, we have made several assumptions:

1. No additional stoppage time: We assume that matches are played for exactly 90 minutes, neglecting any additional stoppage time.

2. Static squads: We assume that the team squads remain unchanged throughout the season, without considering winter transfers or promotions from youth teams.

3. Only outfield players: Since we only did predictions on outfield players, we will exclude the goalkeepers in our optimization problem, and focus on the remaining 10 outfield players.

4. Prioritizing experienced players: We assume that teams prioritize protecting younger players from injuries and rely more heavily on experienced players.

5. Balancing injury prevention and performance: We acknowledge that teams face a trade-off between minimizing injury risk and maintaining competitive performance.

These assumptions allow us to focus on the core aspects of player rotation and injury management, while recognizing that real-world scenarios may involve additional complexities.

4.3 Using 'Total Days'

In this section, we look at one way we can use the predicted variable 'Total Days' (out due to injuries) in an optimization problem.

4.3.1 Objective function

Given a squad of n players, and that the team plays g number of games across the season, we can formulate the following objective function

$$\begin{aligned}
\min_{x_i} \quad & \sum_{i=1}^{n} \left(\lambda \delta_1 y_i^2 x_i + (1-\lambda)\delta_2 c_i^2\right) \\
\text{s.t.} \quad & \sum_{i=1}^{n} x_i = 90 \times 10 \times g \\
& l_i \times g \leq x_i \leq u_i \times g \\
& x_i \geq 0 \\
& c_i \geq 0 \\
& c_i \geq \text{ min. playing time} - x_i
\end{aligned}$$

where:

- x_i is the total number of minutes player i plays across the season, which is the variable we are optimizing over,
- y_i is the predicted 'Total Days' (out due to injuries) for player i,
- l_i and u_i are the minutes lower and upper bound of player i, determined by his age group (Table 4.2),
- $c_i = \max(0,\ \text{min. playing time} - x_i)$,
- λ is the trade-off term between minimizing injury risk $\delta_1 \sum_{i=1}^{n} y_i^2 x_i$ and performance $\delta_2 c_i^2$,
- and δ_1, δ_2 controls the severity of each term. In our experiments we found that $\delta_1 = 100$ and $\delta_2 = 1000$ proved to be effective.

The first term

$$\delta_1 y_i^2 x_i$$

is regarded as the injury susceptibility term. We want to propose a suitable number of minutes x_i for player i, given his predicted number of days out y_i (squared to make sure that this term is significant enough to make a difference to the optimization results), such that this product is small enough to indicate a low injury susceptibility. At the same time, we recognize that teams will still want to perform well. To do so, they have to play their best players which gives rise to the second term

$$\delta_2 c_i^2 = \delta_2 \times \max(0,\ \text{min. playing time} - x_i)^2$$

which penalizes the objective function only if player i plays less than the minimum playing time that is suggested. c_i can be expressed as

$$c_i = \begin{cases} 0, & \text{if player } i \text{ plays more than min. playing time} \\ \text{min. playing time} - x_i, & \text{otherwise.} \end{cases}$$

This suggested minimum playing time is derived based on the observation of the number of minutes played in our original dataset. If we gauge the importance of a player by the number of minutes played, then Table 4.1[1] shows a rough guide of the suggested minimum playing time for players of difference importance to the team.

[1]'Base playing minutes' is the average number of minutes that players in the 50% percentile played

Table 4.1: Derivation of Min. playing minutes

Actual Playing Time Percentile	Min. Playing Minutes
Greater than 90%	1.6 * Base playing minutes
Between 75% to 89%	1.3 * Base playing minutes
Between 60% to 74%	1.0 * Base playing minutes
Between 45% to 59%	0.8 * Base playing minutes
Between 30% to 44%	0.6 * Base playing minutes
Below 30%	0

Furthermore, to address assumption 4, it is natural for us to set lower and upper bounds on minutes played for different players based on their age as shown in Table 4.2. Intuitively, we want younger players to be more protected and also not exert older players too much, hence justifying the reduced minutes.

Table 4.2: Upper and lower bounds of minutes played by age group

Age group	Lower bound (per match)	Upper bound (per match)
16 to 22	15	60
23 to 35	70	90
35 and above	15	70

4.3.2 Constraints

The first constraint

$$\sum_{i=1}^{n} x_i = 90 \times 10 \times g$$

ensures that the total playing time of all the players adheres to the rules of the game. The second constraint

$$90 \times l_i \times g \leq x_i \leq 90 \times u_i \times g$$

enforces the upper and lower bounds in Table 4.2. The third constraint $x_i \geq 0$ ensures that playing minutes are non-negative and the last two constraints are used to simplify the function $c_i = \max(0,\ \text{min. playing time} - x_i)$ for the optimization problem.

Finally, note that $n,\ g,\ \lambda,\ ,\delta_1,\ \delta_2$, age group constraints $(l_i,\ u_i)$ and suggested minimum playing time can all be customized based on the needs and characteristics of the team and the manager.

4.3.3 Results

In this section, we discuss some of the results obtained by applying and solving this optimization problem on selected teams. We will look at some values of λ to simulate a team trying to prioritize performance over injury prevention and vice versa.

Manchester City

Manchester City is currently one of the top teams in Europe which plays a lot of games per season (about 60 to 70). It is vital that they are able to prevent injuries to their key players but still maintain a certain level of performance. If we set $\lambda = 0.9$, we simulate prioritizing injury prevention over results.

Table 4.3: Sample of Manchester City's results with $\lambda = 0.9$

Name	Club	Age	Minutes Played Total 23/24:	Weight	Predicted Days Out	Min Playing Time	Lambda	**Suggested Minutes**
Rúben Dias	Man City	27	3817	0.061609	11.715614	3524.3	0.9	**2086**
Manuel Akanji	Man City	28	3861	0.062319	12.118954	3524.3	0.9	**2086**
Josko Gvardiol	Man City	22	3479	0.056154	27.903629	2711.0	0.9	**2306**
Sergio Gómez	Man City	23	445	0.007183	48.224480	0.0	0.9	**2517**
Rico Lewis	Man City	19	1606	0.025922	23.911762	0.0	0.9	**893**
Kyle Walker	Man City	34	4083	0.065903	19.281223	4337.6	0.9	**3545**
Rodri	Man City	27	4327	0.069841	4.260078	4337.6	0.9	**4190**
Kevin De Bruyne	Man City	32	1750	0.028246	101.475521	0.0	0.9	**3227**
Erling Haaland	Man City	23	3745	0.060447	24.129783	2711.0	0.9	**2288**

From Table 4.3[2], we see that compared to 'Minutes Played Total 23/24', the optimal 'Suggested Minutes' is comparatively lower for important players and higher for less important players ('Weight[3]' column represents each player's importance). This makes sense because the model is trying to prioritise injury prevention and balancing match load.

Table 4.4: Sample of Manchester City's results with $\lambda = 0.2$

Name	Club	Age	Minutes Played Total 23/24:	Weight	Predicted Days Out	Min Playing Time	Lambda	**Suggested Minutes**
Rúben Dias	Man City	27	3817	0.061609	11.715614	3524.3	0.2	**3351**
Manuel Akanji	Man City	28	3861	0.062319	12.118954	3524.3	0.2	**3392**
Josko Gvardiol	Man City	22	3479	0.056154	27.903629	2711.0	0.2	**2059**
Sergio Gómez	Man City	23	445	0.007183	48.224480	0.0	0.2	**2168**
Rico Lewis	Man City	19	1606	0.025922	23.911762	0.0	0.2	**1028**
Kyle Walker	Man City	34	4083	0.065903	19.281223	4337.6	0.2	**4742**
Rodri	Man City	27	4327	0.069841	4.260078	4337.6	0.2	**3748**
Kevin De Bruyne	Man City	32	1750	0.028246	101.475521	0.0	0.2	**2178**
Erling Haaland	Man City	23	3745	0.060447	24.129783	2711.0	0.2	**2134**

Now comparing this with Table 4.4, where we set $\lambda = 0.2$ to simulate prioritizing performance, we see that important players will consequently play more. In fact for some players like Kyle Walker, who is an important player for Manchester City, his optimal 'Suggested Minutes' is higher than 'Minutes Played Total 23/24' because the model is trying to let important players play more.

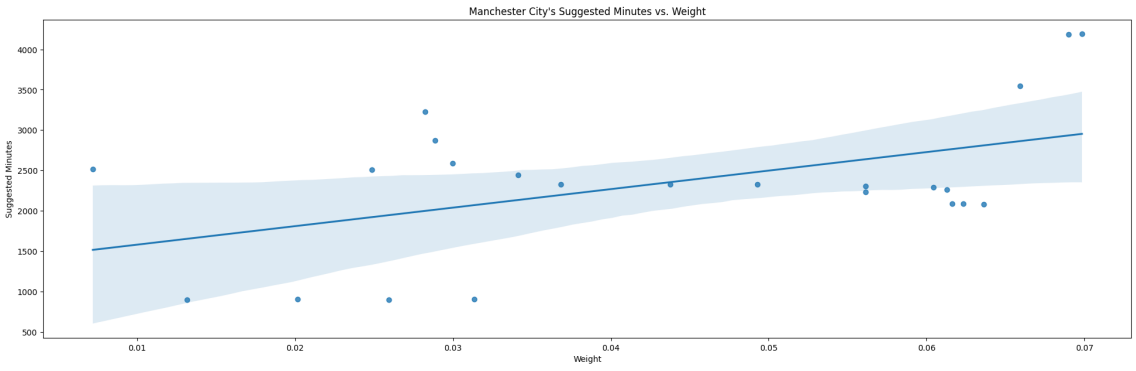


Figure 4.1: Manchester City's Suggested Minutes vs. Weight when $\lambda = 0.9$

[2]Full table for Manchester City can be found in C.1

[3]'Weight' is derived based on the ratio of minutes played by the player compared to team total

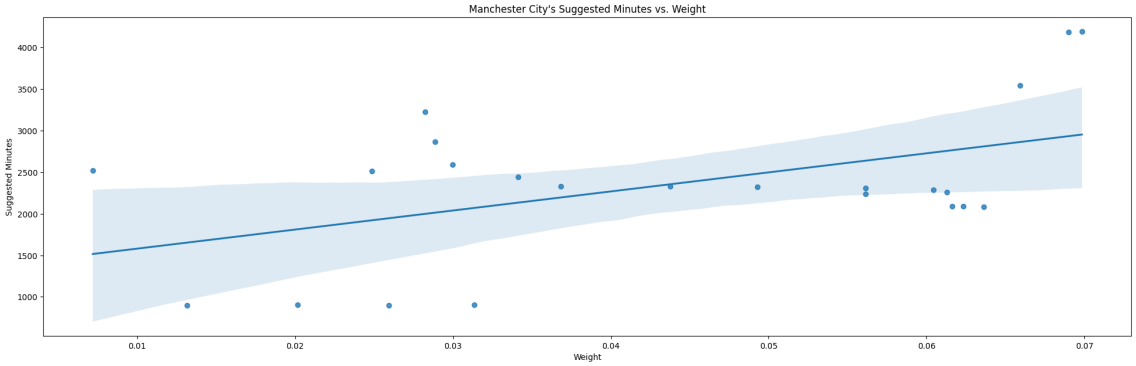


Figure 4.2: Manchester City's Suggested Minutes vs. Weight when $\lambda = 0.2$

From Figures 4.1 and 4.2, we also observed that regardless of whether $\lambda = 0.9$ or 0.2, as the 'Weight' (importance) of the player increases, the model will still try to let him play more to ensure a certain level of team performance.

We performed some sensitivity analysis by varying the values of λ to observe any fluctuations in the general trend of the results.

From Figure 4.3, as λ varies from 0.1 to 0.9, optimal suggested minutes of players changes from favouring important players to become more balanced. From Figure 4.4, we also observe that optimal suggested minutes changes from favouring those who are predicted to have fewer days out from injuries to become more balanced as well. This highlights that our model is doing the intended thing which is to gradually distribute playing load evenly.

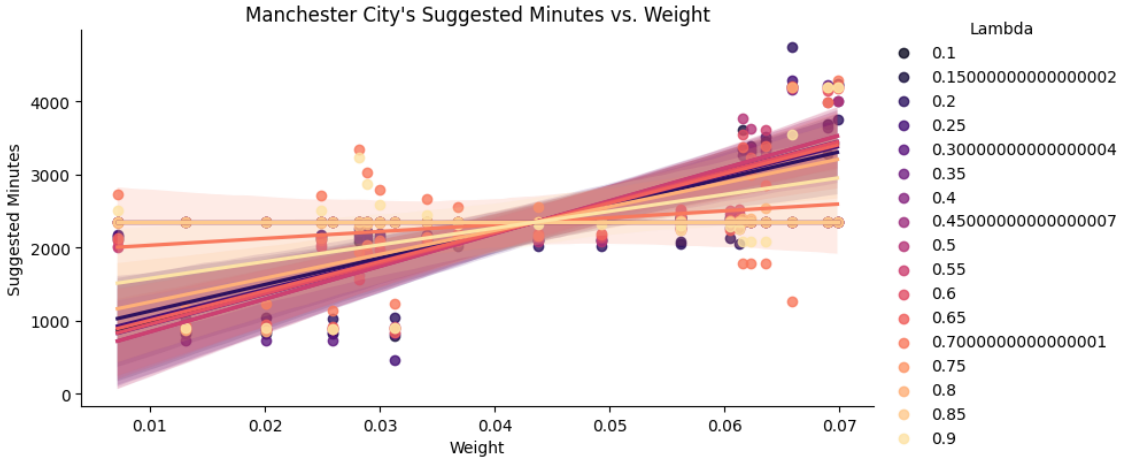


Figure 4.3: Manchester City's Suggested Minutes vs. Weight as λ varies

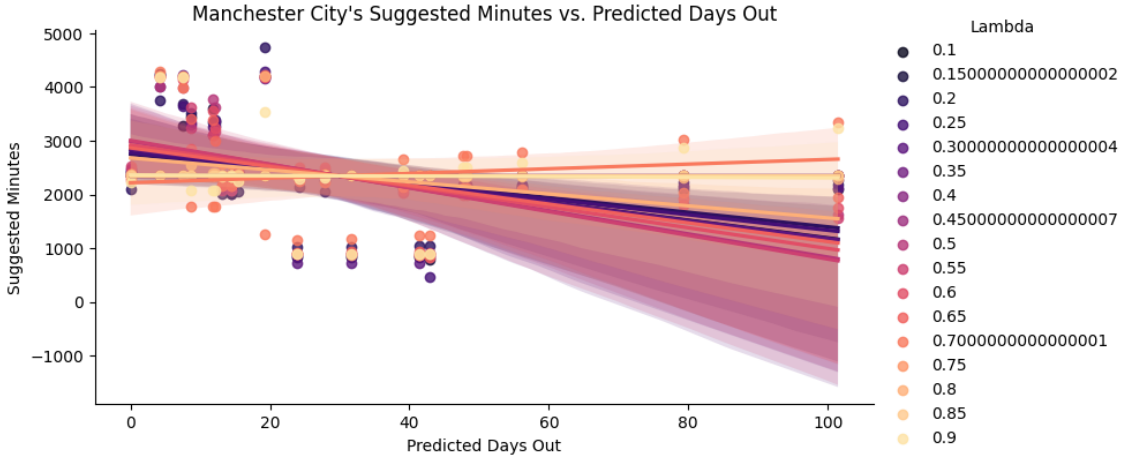


Figure 4.4: Manchester City's Suggested Minutes vs. Total Days as λ varies

Furthermore, if we zoom into the suggested minutes of Kyle Walker in Figure 4.5, we see that it decreases as λ increases to prioritise injury prevention. In contrast, for Sergio Gomez, who is not as important, his optimal suggested minutes increases as λ increases. This increase is to reduce the stress on other important players so that the injury susceptibility remains low.

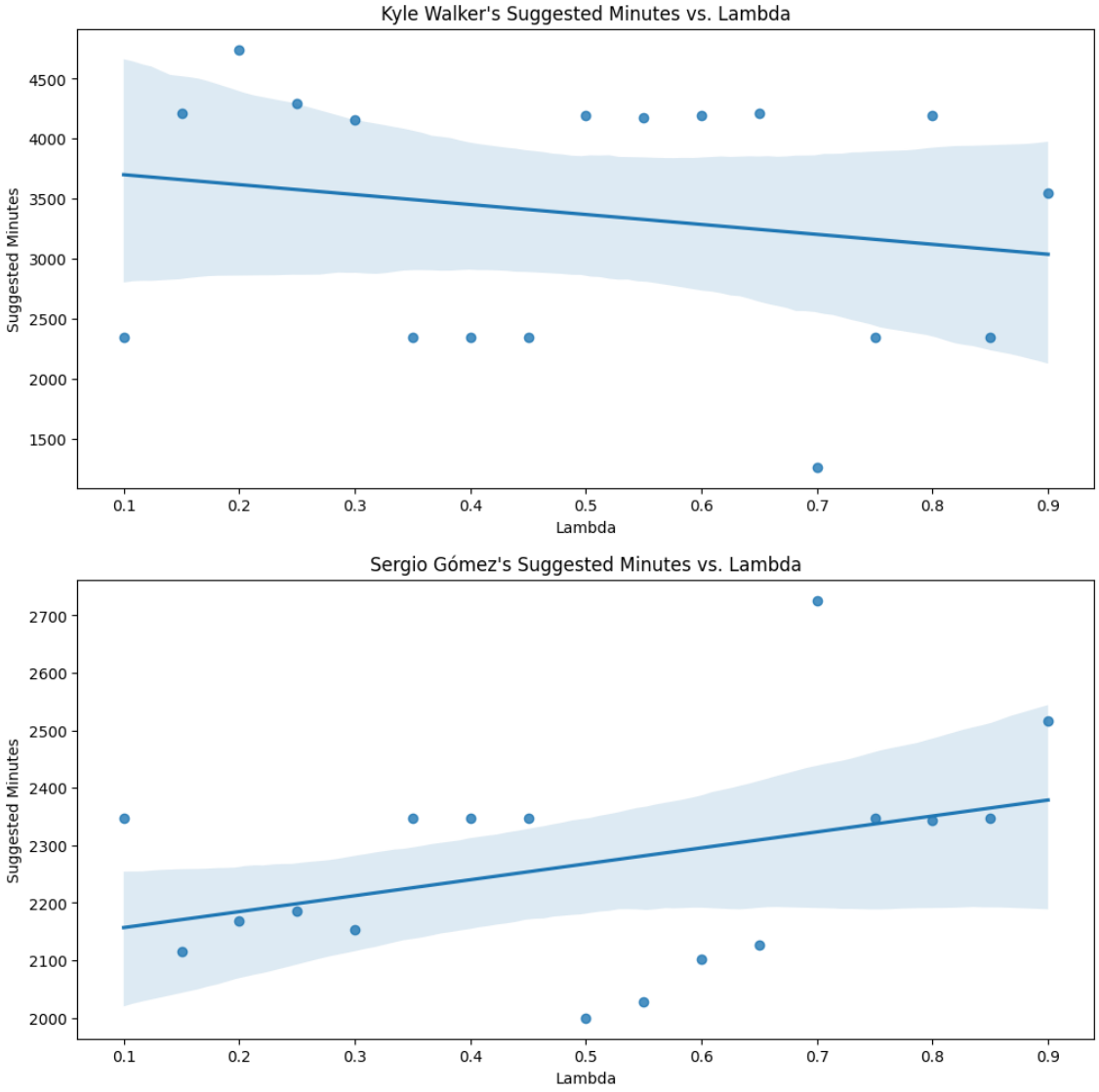


Figure 4.5: Kyle Walker's and Sergio Gomez's Suggested Minutes vs. Lambda

Brighton & Hove Albion and Sheffield United

To further test the robustness of our optimization model, we need to test it on different teams. Although Brighton & Hove Albion[4] plays in the same league as Manchester City, they were a mid-table team, and plays fewer matches (about 50) across the season with different season objectives compared to Manchester City.

[4]Full table for Brighton & Hove Albion can be found in C.2

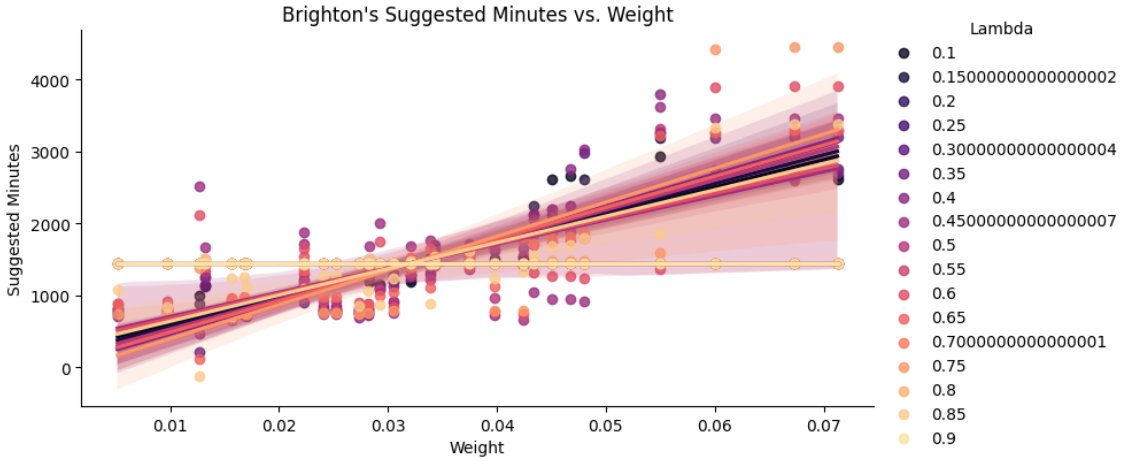


Figure 4.6: Brighton's Suggested Minutes vs. Weight as λ varies

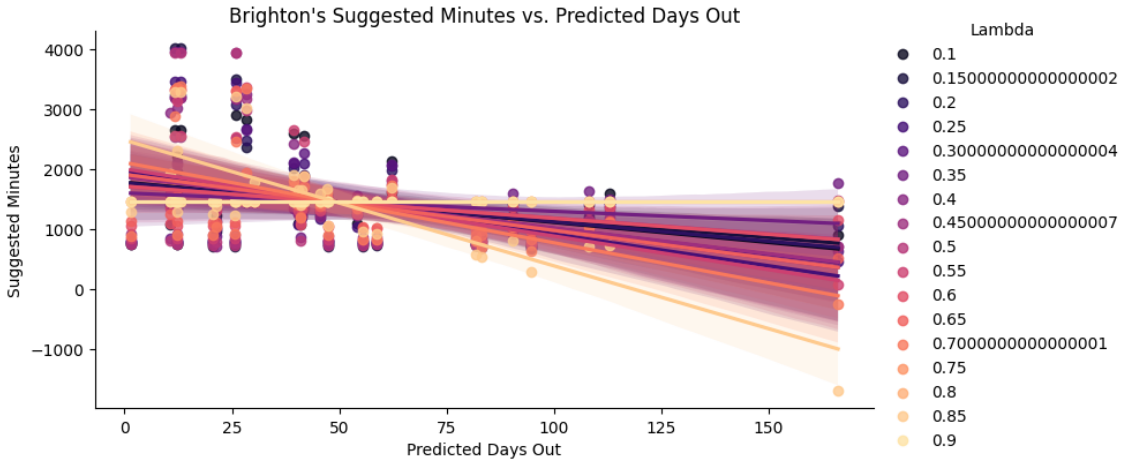


Figure 4.7: Brighton's Suggested Minutes vs. Total Days as λ varies

From Figures 4.6 and 4.7 we observe the same general trend as that of Manchester City's despite playing fewer matches.

Similarly, important players for Brighton like Simon Adingra will see a decrease in optimal suggested minutes, and less important players like Solly March will see a slight increase in optimal playing minutes as λ increases (refer to Appendix C.2).

Sheffield United[5] also played in the Premier League with Manchester City and Brighton. However, they finished at the bottom of the table. Across the season they played about 40 games.

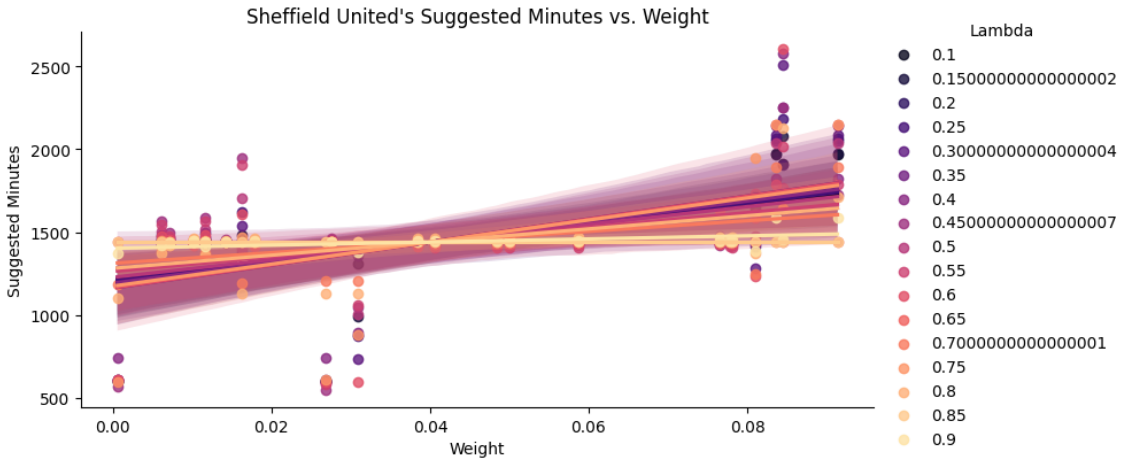


Figure 4.8: Sheffield United's Suggested Minutes vs. Weight as λ varies

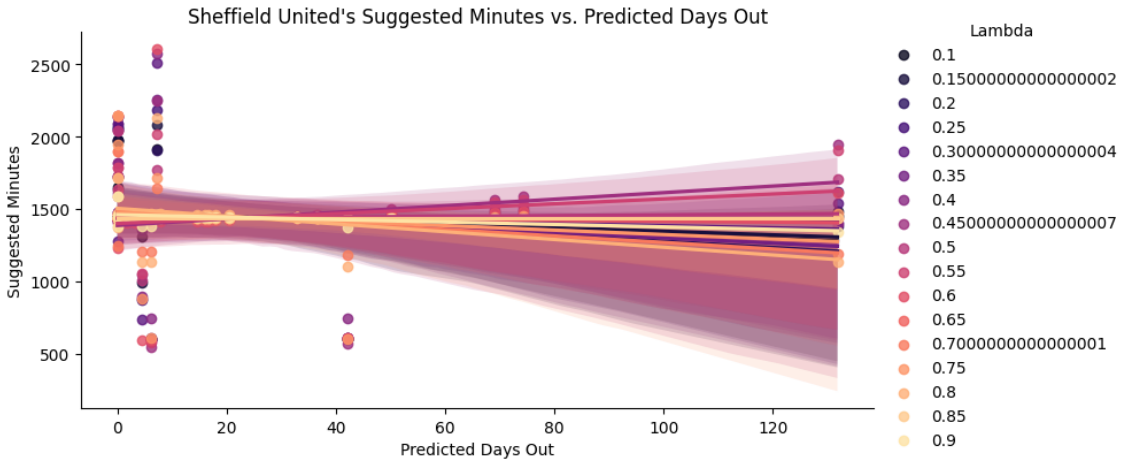


Figure 4.9: Sheffield United's Suggested Minutes vs. Total Days as λ varies

[5]Full table for Sheffield United can be found in C.3

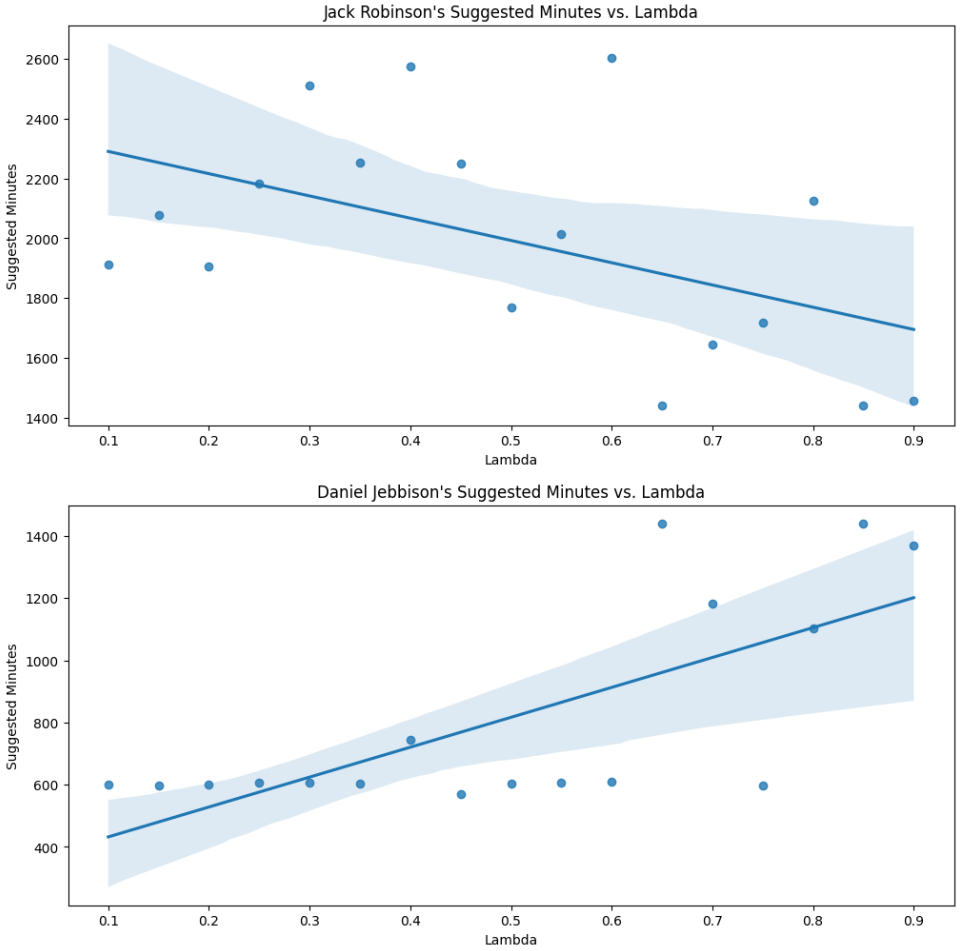


Figure 4.10: Jack Robinson's (Important) and Daniel Jebbison's (Less Important) Suggested Minutes vs. Lambda

From Figures 4.8, 4.9 and 4.10, we see that the result is still consistent even for very different teams which further substantiates the robustness of our optimization model.

La Liga teams

To further showcase the robustness of the model, we performed the same optimization task for teams in La Liga (Spain). We also, conducted optimization for the top (Real Madrid in Appendix C.4), mid-table (Osasuna in Appendix C.5) and bottom (Granada in Appendix C.6) teams. The results are consistent with those teams in

the Premier League.

4.4 Using 'Injury Count' and 'Average Injury Cost'

We can also make use of other predicted variables like 'Injury Count' and 'Average Injury Cost' to formulate another optimization problem. In this case, we still want to minimize injury susceptibility, but from a economic point of view.

4.4.1 Objective function

We have the following objective function,

$$\begin{aligned}
\min_{x_i} \ & \sum_{i=1}^{n} \left(\lambda\delta_1(\alpha_i\beta_i)^2 x_i + (1-\lambda)\delta_2 c_i^2\right) \\
\text{s.t.} \ & \sum_{i=1}^{n} x_i = 90 \times 10 \times g \\
& l_i \times g \leq x_i \leq u_i \times g \\
& x_i \geq 0 \\
& c_i \geq 0 \\
& c_i \geq \text{ min. playing time} - x_i
\end{aligned}$$

where:

- x_i is the still total number of minutes player i plays across the season, which is the variable we are optimizing over,
- α_i is the predicted 'Injury Count' for player i,
- β_i is the predicted 'Average Injury Cost' for player i
- $n, g, l_i, u_i, c_i, \lambda, \delta_1$ and δ_2 are as previously defined,

The constraints over x_i remain unchanged. The key difference here is that we use variables α_i and β_i to quantify injury risk in an economic sense via the term $\lambda\delta_1(\alpha_i\beta_i)^2x_i$ instead.

4.4.2 Results

Manchester City

When we performed the optimization on Manchester City with $\lambda = 0.9$, we observe from Table 4.5 that the optimal suggested minutes for some important players will be fewer than what they actually played in 2023/ 2024 season and vice versa for less important players. This is similar to what we observed in Section 4.3.3.

Table 4.5: Sample of Manchester City's results with $\lambda = 0.9$

Name	Club	Age	Minutes Played Total 23/24:	Weight	Predicted Injury Count	Predicted Average Injury Cost	Min Playing Time	Lambda	**Suggested Minutes**
Rúben Dias	Man City	27	3817	0.061609	0.418027	0.077455	3524.3	0.9	**2445**
Manuel Akanji	Man City	28	3861	0.062319	0.599638	0.114154	3524.3	0.9	**2441**
Josko Gvardiol	Man City	22	3479	0.056154	0.790046	0.125385	2711.0	0.9	**2348**
Sergio Gómez	Man City	23	445	0.007183	1.006816	0.182618	0.0	0.9	**2344**
Rico Lewis	Man City	19	1606	0.025922	0.407508	0.073580	0.0	0.9	**895**
Kyle Walker	Man City	34	4083	0.065903	0.820027	0.154623	4337.6	0.9	**4200**
Rodri	Man City	27	4327	0.069841	0.250853	0.058731	4337.6	0.9	**4200**
Kevin De Bruyne	Man City	32	1750	0.028246	2.345621	0.423790	0.0	0.9	**2344**
Erling Haaland	Man City	23	3745	0.060447	0.620467	0.143507	2711.0	0.9	**2348**

Table 4.6: Sample of Manchester City's results with $\lambda = 0.2$

Name	Club	Age	Minutes Played Total 23/24:	Weight	Predicted Injury Count	Predicted Average Injury Cost	Min Playing Time	Lambda	**Suggested Minutes**
Rúben Dias	Man City	27	3817	0.061609	0.418027	0.077455	3524.3	0.2	**3660**
Manuel Akanji	Man City	28	3861	0.062319	0.599638	0.114154	3524.3	0.2	**3660**
Josko Gvardiol	Man City	22	3479	0.056154	0.790046	0.125385	2711.0	0.2	**2576**
Sergio Gómez	Man City	23	445	0.007183	1.006816	0.182618	0.0	0.2	**1869**
Rico Lewis	Man City	19	1606	0.025922	0.407508	0.073580	0.0	0.2	**908**
Kyle Walker	Man City	34	4083	0.065903	0.820027	0.154623	4337.6	0.2	**3916**
Rodri	Man City	27	4327	0.069841	0.250853	0.058731	4337.6	0.2	**3916**
Kevin De Bruyne	Man City	32	1750	0.028246	2.345621	0.423790	0.0	0.2	**1869**
Erling Haaland	Man City	23	3745	0.060447	0.620467	0.143507	2711.0	0.2	**2576**

When we set $\lambda = 0.2$ and simulate prioritizing performance, the optimal suggested minutes in Table 4.6 are closer to the actual playing time in 2023/ 2024, which highlights that the optimization model can produce realistic results in the case for

teams who want to perform at the highest level while keeping injury cost manageable.

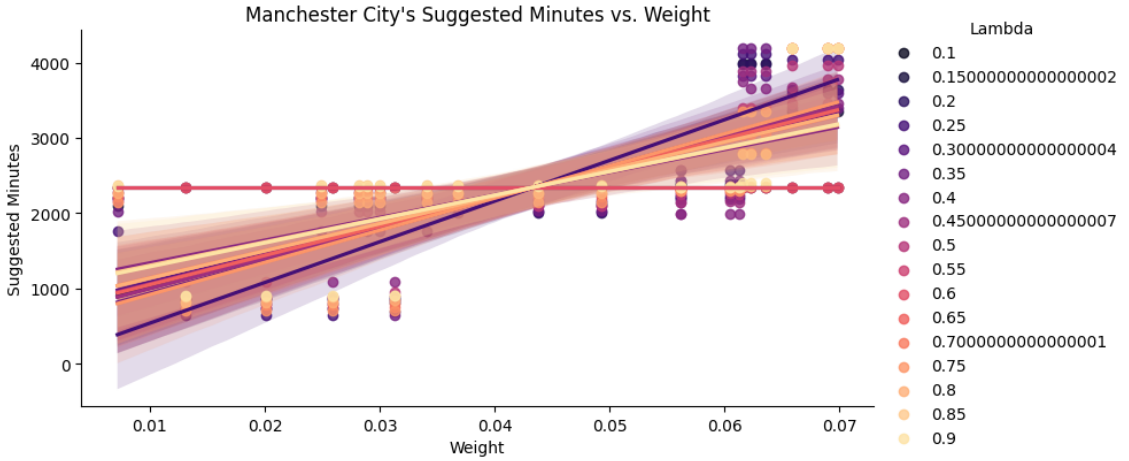


Figure 4.11: Manchester City's Suggested Minutes vs. Weight as λ varies

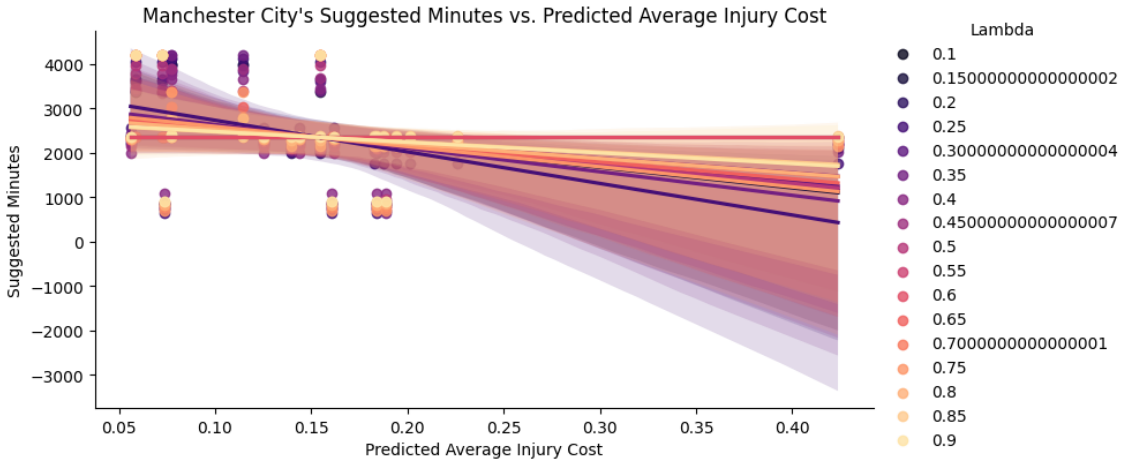


Figure 4.12: Manchester City's Suggested Minutes vs. Average Injury Cost as λ varies

Performing the same sensitivity analysis by varying λ and observing the effect on optimal suggested minutes, we see the same general trend. In Figure 4.12, we plot 'Suggested Minutes' against the predicted variable 'Average Injury Cost' instead. When we are trying to minimize the injury cost (by setting λ to be larger) the model still tries to even out the playing time across the squad, while still being cautious

that player who incur higher injury cost should play fewer minutes as indicated by the negative slopes.

While the general trends may be similar, the choice of predicted variables can influence the specific recommendations generated by the optimization model, as shown by the difference in numerical values of 'Suggested Minutes'. By considering multiple perspectives and evaluating the trade-offs between different variables, we can make more informed decisions regarding player rotation and injury management.

4.5 Conclusion on optimization models

The optimization models developed in this study has demonstrated their ability to produce results that align with our expectations regarding player rotation and injury management. While it may seem intuitive to allocate playing time based on experience and injury history, the models offers a quantitative approach that can be customized to meet the specific needs of different teams. These recommendations can serve as valuable guidelines for coaches and managers, assisting them in making informed decisions regarding player rotation and injury prevention.

While further research and practical implementation are necessary to fully validate the model's effectiveness, the results presented in this study offer promising insights into the potential benefits of data-driven approaches for optimizing player management in football.

Chapter 5

Future Works

Given more time and resources, we believe this project has a lot of potential in the industry. In this chapter, we will explore potential extensions to this project.

5.1 Data enrichment

While our current model demonstrates promising results using publicly available data, incorporating additional biomarkers and training data could further enhance its predictive power. Access to biomarkers, such as muscle strength and stiffness, could provide valuable insights into individual player vulnerability. Additionally, detailed training schedules and intensity data would enable a more precise analysis of player workload and its impact on injury risk.

5.2 Expanding the dataset

Extending the dataset to include multiple seasons would allow for a more comprehensive analysis of injury patterns and trends over time. By incorporating data from past seasons, we could identify long-term trends, seasonality effects, and potential changes in injury risk factors.

5.3 Validation and implementation

To validate the effectiveness of our optimization model, future research should focus on implementing it in real-world scenarios. By observing the outcomes of player rotations based on the model's recommendations, we can assess its impact on injury rates and team performance. This practical validation would provide valuable insights into the model's effectiveness and identify areas for further refinement.

5.4 Broader applications

The principles and techniques employed in this research can be extended to other sports that require player rotation and injury management. Basketball, American football, and other sports with similar demands on athletes could benefit from data-driven approaches to optimize player usage and minimize injuries. By adapting our models to the specific characteristics of these sports, we can contribute to improving player health and performance across various athletic disciplines.

5.5 Conclusion

This project has demonstrated the potential of machine learning and optimization techniques to address the complex problem of injury prevention in football. By leveraging data-driven approaches, we have developed a predictive model that can provide valuable insights for coaches and team managers. While further research and validation are needed to fully realize the benefits of this approach, the findings presented in this project offer a promising foundation for future advancements in injury prevention and player management.

Appendix A

Datasets

A.1 Player data

The full set of data columns in player data include: Name, Birthday, Height, League, Club, Position, Appearances, Injuries, Total Days, Total Games Missed, Minutes Played Total 23/24:, Age, Type of Injury, Categorized Injuries, Injury Count, Categorized Position, Average Injury Cost.

The full dataset can be found here.

A.2 Team data

The full set of data columns in team data include: Team, Goals Against Per 90, Shots On Target Against, Saves Percentage, GK Passes attemped, GK Throws Attempted, GK Crosses Faced, Sweeper per 90, Players Tackled, Dribblers Tackled, Blocks, Interceptions, Clearances, Goals, Shots, Shots on Target, Shots per 90, Shots on Target per 90, Expected Goals, Passes Completed, Passes Attempted, Crosses, Progressive Passes, Shot Creating Actions, Shot Creating Actions per 90, Goal Creating Actions, Goal Creating Actions per 90, Defensive Touches, Attacking Touches, Take Ons Attempted, Take Ons Tackled, Yellow Cards, Red Cards, Fouls

Committed, Fouls Drawn, Aerial Won, Aerial Lost.

The full dataset can be found here:

- Premier League
- La Liga
- Bundesliga
- Serie A
- Ligue 1

A.3 Referee data

The full set of data columns in referee data include: Team, Fouls per 90, Fouls per Tackle, Penalties per 90, Yellow Cards per 90, Red Cards per 90.

The full dataset can be found here:

- Premier League
- La Liga
- Bundesliga
- Serie A
- Ligue 1

Appendix B

Machine Learning

B.1 Original dataset

The original data without dimensionality reduction can be found here:

- **X** training set
- **y** training set
- **X** test set
- **y** test set

B.2 Model comparison

We consolidated the performances of all models trained in Section 3.3 (including datasets with 2 and 3 principle components for comparison) which can be found here. Note that the R-squared evaluation metric was included in the comparison but not useful in our context.

Appendix C

Optimization

C.1 Manchester City

Table C.1: Manchester City's results with $\lambda = 0.9$

Name	Club	Age	Minutes Played Total 23/24:	Weight	Predicted Days Out	Min Playing Time	Lambda	**Suggested Minutes**
Rúben Dias	Man City	27	3817	0.061609	11.715614	3524.3	0.9	**2342**
Manuel Akanji	Man City	28	3861	0.062319	12.118954	3524.3	0.9	**2342**
Nathan Aké	Man City	29	3053	0.049278	13.162987	2168.8	0.9	**2349**
John Stones	Man City	30	1858	0.029990	56.180477	1355.5	0.9	**2357**
Taylor Harwood-Bellis	Man City	22	3942	0.063627	8.670015	3524.3	0.9	**2342**
Josko Gvardiol	Man City	22	3479	0.056154	27.903629	2711.0	0.9	**2348**
Sergio Gómez	Man City	23	445	0.007183	48.224480	0.0	0.9	**2354**
Josh Wilson-Esbrand	Man City	21	1248	0.020144	41.408318	0.0	0.9	**896**
Rico Lewis	Man City	19	1606	0.025922	23.911762	0.0	0.9	**901**
Kyle Walker	Man City	34	4083	0.065903	19.281223	4337.6	0.9	**4383**
Issa Kaboré	Man City	23	1788	0.028860	79.272749	0.0	0.9	**2365**
Rodri	Man City	27	4327	0.069841	4.260078	4337.6	0.9	**4201**
Matheus Nunes	Man City	25	1540	0.024857	47.740331	0.0	0.9	**2354**
Mateo Kovacic	Man City	30	2711	0.043758	14.511348	2168.8	0.9	**2349**
James McAtee	Man City	21	1942	0.031345	42.963722	1355.5	0.9	**900**
Bernardo Silva	Man City	29	3797	0.061286	15.459842	2711.0	0.9	**2347**
Kevin De Bruyne	Man City	32	1750	0.028246	101.475521	0.0	0.9	**2376**
Jérémy Doku	Man City	22	2281	0.036817	14.484747	2168.8	0.9	**2349**
Jack Grealish	Man City	28	2114	0.034122	39.032298	1355.5	0.9	**2352**
Phil Foden	Man City	24	4274	0.068986	7.540023	4337.6	0.9	**4200**
Oscar Bobb	Man City	20	815	0.013155	31.661233	0.0	0.9	**899**
Erling Haaland	Man City	23	3745	0.060447	24.129783	2711.0	0.9	**2348**
Julián Álvarez	Man City	24	3479	0.056154	0.000000	2711.0	0.9	**2346**

Table C.2: Manchester City's results with $\lambda = 0.2$

Name	Club	Age	Minutes Played Total 23/24:	Weight	Predicted Days Out	Min Playing Time	Lambda	**Suggested Minutes**
Rúben Dias	Man City	27	3817	0.061609	11.715614	3524.3	0.2	**3481**
Manuel Akanji	Man City	28	3861	0.062319	12.118954	3524.3	0.2	**3620**
Nathan Aké	Man City	29	3053	0.049278	13.162987	2168.8	0.2	**2042**
John Stones	Man City	30	1858	0.029990	56.180477	1355.5	0.2	**2067**
Taylor Harwood-Bellis	Man City	22	3942	0.063627	8.670015	3524.3	0.2	**3568**
Josko Gvardiol	Man City	22	3479	0.056154	27.903629	2711.0	0.2	**2236**
Sergio Gómez	Man City	23	445	0.007183	48.224480	0.0	0.2	**2072**
Josh Wilson-Esbrand	Man City	21	1248	0.020144	41.408318	0.0	0.2	**841**
Rico Lewis	Man City	19	1606	0.025922	23.911762	0.0	0.2	**843**
Kyle Walker	Man City	34	4083	0.065903	19.281223	4337.6	0.2	**4207**
Issa Kaboré	Man City	23	1788	0.028860	79.272749	0.0	0.2	**2049**
Rodri	Man City	27	4327	0.069841	4.260078	4337.6	0.2	**3701**
Matheus Nunes	Man City	25	1540	0.024857	47.740331	0.0	0.2	**2072**
Mateo Kovacic	Man City	30	2711	0.043758	14.511348	2168.8	0.2	**2042**
James McAtee	Man City	21	1942	0.031345	42.963722	1355.5	0.2	**894**
Bernardo Silva	Man City	29	3797	0.061286	15.459842	2711.0	0.2	**2241**
Kevin De Bruyne	Man City	32	1750	0.028246	101.475521	0.0	0.2	**2026**
Jérémy Doku	Man City	22	2281	0.036817	14.484747	2168.8	0.2	**2042**
Jack Grealish	Man City	28	2114	0.034122	39.032298	1355.5	0.2	**2076**
Phil Foden	Man City	24	4274	0.068986	7.540023	4337.6	0.2	**4489**
Oscar Bobb	Man City	20	815	0.013155	31.661233	0.0	0.2	**843**
Erling Haaland	Man City	23	3745	0.060447	24.129783	2711.0	0.2	**2254**
Julián Álvarez	Man City	24	3479	0.056154	0.000000	2711.0	0.2	**2294**

C.2 Brighton & Hove Albion

Table C.3: Brighton & Hove Albion's results with $\lambda = 0.7$

Name	Club	Age	Minutes Played Total 23/24:	Weight	Predicted Days Out	Min Playing Time	Lambda	**Suggested Minutes**
Adam Webster	Brighton	29	1360	0.022310	90.331617	0.0	0.7	**978**
Jan Paul van Hecke	Brighton	23	3352	0.054988	28.319009	2851.2	0.7	**3275**
Igor	Brighton	26	2284	0.037468	30.341437	1782.0	0.7	**1632**
Lewis Dunk	Brighton	32	3658	0.060008	25.941292	2851.2	0.7	**3338**
Ed Turns	Brighton	21	2424	0.039764	41.009800	1782.0	0.7	**1410**
Pervis Estupiñán	Brighton	26	1782	0.029233	108.209049	1425.6	0.7	**807**
Valentín Barco	Brighton	19	315	0.005167	55.572379	0.0	0.7	**802**
Tariq Lamptey	Brighton	23	1361	0.022326	54.140801	891.0	0.7	**1327**
Joël Veltman	Brighton	32	1956	0.032087	45.477262	1425.6	0.7	**1389**
Carlos Baleba	Brighton	20	1540	0.025263	1.405206	891.0	0.7	**875**
Billy Gilmour	Brighton	22	2925	0.047983	12.135381	2316.6	0.7	**2883**
Jack Hinshelwood	Brighton	19	1666	0.027330	81.701664	891.0	0.7	**727**
Jakub Moder	Brighton	25	808	0.013255	40.314118	0.0	0.7	**1414**
Pascal Groß	Brighton	32	4104	0.067324	12.942845	2851.2	0.7	**3382**
Yasin Ayari	Brighton	20	1023	0.016782	20.674368	0.0	0.7	**865**
James Milner	Brighton	38	955	0.015666	94.749782	0.0	0.7	**640**
Marc Leonard	Brighton	22	4344	0.071261	11.627335	2851.2	0.7	**3384**
Jensen Weir	Brighton	22	1725	0.028298	41.132645	1425.6	0.7	**1412**
Julio Enciso	Brighton	20	590	0.009679	83.173095	0.0	0.7	**708**
Facundo Buonanotte	Brighton	19	1860	0.030512	21.314211	1425.6	0.7	**952**
Andrew Moran	Brighton	20	2589	0.042471	10.500671	1782.0	0.7	**1823**
Adam Lallana	Brighton	36	951	0.015601	25.540740	0.0	0.7	**860**
Kaoru Mitoma	Brighton	27	2064	0.033859	113.052193	1782.0	0.7	**917**
Ansu Fati	Brighton	21	1037	0.017011	47.355927	0.0	0.7	**822**
Jeremy Sarmiento	Brighton	21	1470	0.024115	12.242050	891.0	0.7	**872**
Simon Adingra	Brighton	22	2851	0.046769	39.235844	2316.6	0.7	**1872**
Solly March	Brighton	29	770	0.012631	166.144182	0.0	0.7	**-309**
Abdallah Sima	Brighton	22	2642	0.043341	62.189812	2316.6	0.7	**1731**
Evan Ferguson	Brighton	19	1720	0.028216	58.809519	891.0	0.7	**797**
João Pedro	Brighton	22	2744	0.045014	41.733583	2316.6	0.7	**1860**
Danny Welbeck	Brighton	33	2089	0.034269	47.052974	1782.0	0.7	**1555**

Table C.4: Brighton & Hove Albion's results with $\lambda = 0.1$

Name	Club	Age	Minutes Played Total 23/24:	Weight	Predicted Days Out	Min Playing Time	Lambda	**Suggested Minutes**
Adam Webster	Brighton	29	1360	0.022310	90.331617	0.0	0.1	**1053**
Jan Paul van Hecke	Brighton	23	3352	0.054988	28.319009	2851.2	0.1	**2831**
Igor	Brighton	26	2284	0.037468	30.341437	1782.0	0.1	**1702**
Lewis Dunk	Brighton	32	3658	0.060008	25.941292	2851.2	0.1	**2894**
Ed Turns	Brighton	21	2424	0.039764	41.009800	1782.0	0.1	**1570**
Pervis Estupiñán	Brighton	26	1782	0.029233	108.209049	1425.6	0.1	**1174**
Valentín Barco	Brighton	19	315	0.005167	55.572379	0.0	0.1	**740**
Tariq Lamptey	Brighton	23	1361	0.022326	54.140801	891.0	0.1	**1092**
Joël Veltman	Brighton	32	1956	0.032087	45.477262	1425.6	0.1	**1222**
Carlos Baleba	Brighton	20	1540	0.025263	1.405206	891.0	0.1	**758**
Billy Gilmour	Brighton	22	2925	0.047983	12.135381	2316.6	0.1	**2217**
Jack Hinshelwood	Brighton	19	1666	0.027330	81.701664	891.0	0.1	**753**
Jakub Moder	Brighton	25	808	0.013255	40.314118	0.0	0.1	**1102**
Pascal Groß	Brighton	32	4104	0.067324	12.942845	2851.2	0.1	**2644**
Yasin Ayari	Brighton	20	1023	0.016782	20.674368	0.0	0.1	**747**
James Milner	Brighton	38	955	0.015666	94.749782	0.0	0.1	**724**
Marc Leonard	Brighton	22	4344	0.071261	11.627335	2851.2	0.1	**2645**
Jensen Weir	Brighton	22	1725	0.028298	41.132645	1425.6	0.1	**1224**
Julio Enciso	Brighton	20	590	0.009679	83.173095	0.0	0.1	**729**
Facundo Buonanotte	Brighton	19	1860	0.030512	21.314211	1425.6	0.1	**1222**
Andrew Moran	Brighton	20	2589	0.042471	10.500671	1782.0	0.1	**1573**
Adam Lallana	Brighton	36	951	0.015601	25.540740	0.0	0.1	**746**
Kaoru Mitoma	Brighton	27	2064	0.033859	113.052193	1782.0	0.1	**1587**
Ansu Fati	Brighton	21	1037	0.017011	47.355927	0.0	0.1	**742**
Jeremy Sarmiento	Brighton	21	1470	0.024115	12.242050	891.0	0.1	**758**
Simon Adingra	Brighton	22	2851	0.046769	39.235844	2316.6	0.1	**2599**
Solly March	Brighton	29	770	0.012631	166.144182	0.0	0.1	**910**
Abdallah Sima	Brighton	22	2642	0.043341	62.189812	2316.6	0.1	**2138**
Evan Ferguson	Brighton	19	1720	0.028216	58.809519	891.0	0.1	**755**
João Pedro	Brighton	22	2744	0.045014	41.733583	2316.6	0.1	**2548**
Danny Welbeck	Brighton	33	2089	0.034269	47.052974	1782.0	0.1	**1603**

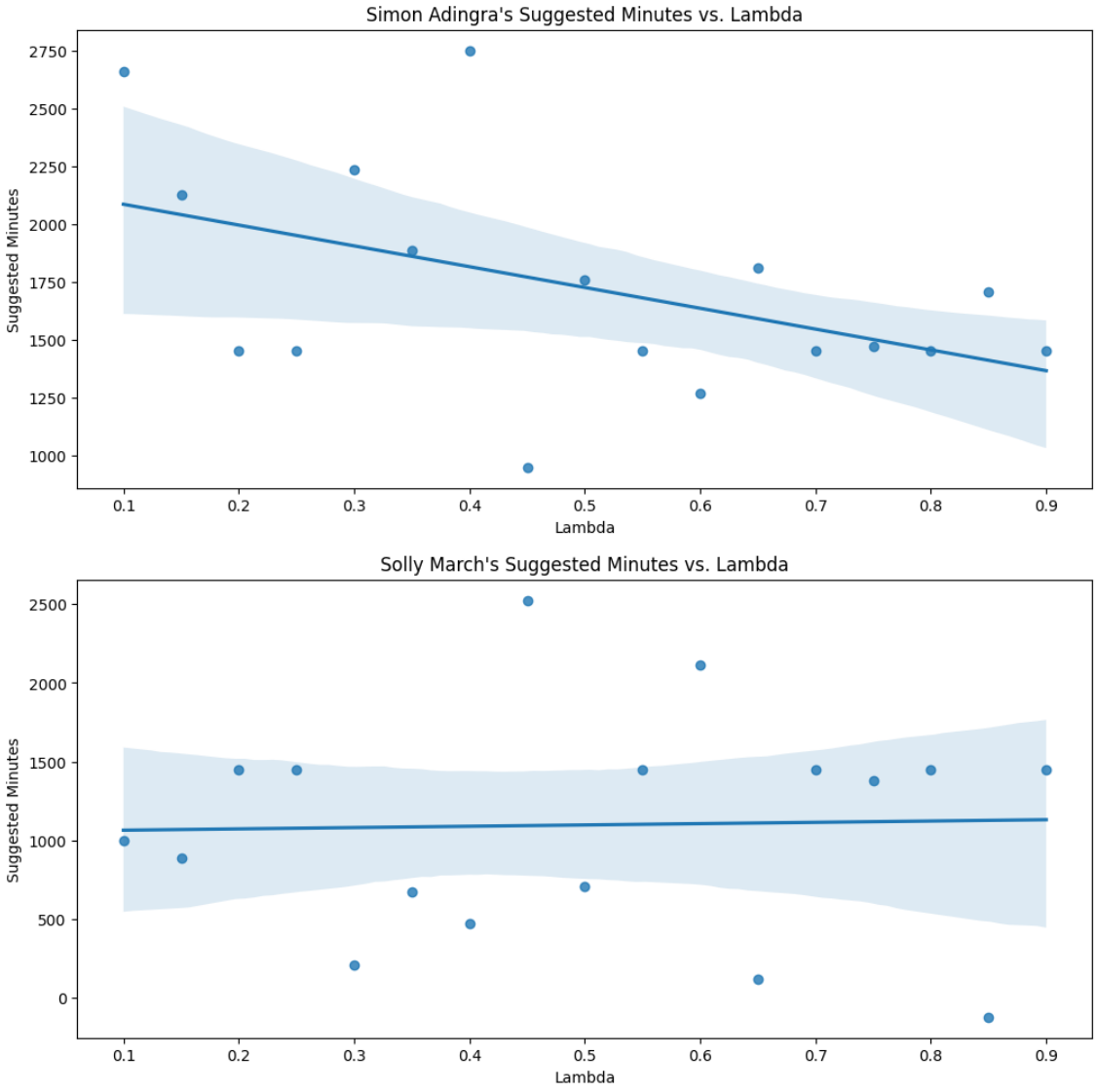


Figure C.1: Simon Adingra's (Important) and Solly March's (Less Important) Suggested Minutes vs. Lambda

C.3 Sheffield United

Table C.5: Sheffield United's results with $\lambda = 0.8$

Name	Club	Age	Minutes Played Total 23/24:	Weight	Predicted Days Out	Min Playing Time	Lambda	**Suggested Minutes**
Anel Ahmedhodzic	Sheff Utd	25	2739	0.077788	16.574005	1414.4	0.8	**1479**
Auston Trusty	Sheff Utd	25	2752	0.078157	6.146398	1414.4	0.8	**1486**
John Egan	Sheff Utd	31	574	0.016302	131.780905	544.0	0.8	**961**
Jack Robinson	Sheff Utd	30	2976	0.084519	7.161556	1740.8	0.8	**2681**
Chris Basham	Sheff Utd	35	408	0.011587	74.240789	0.0	0.8	**1320**
Yasser Larouci	Sheff Utd	23	629	0.017864	14.656405	544.0	0.8	**1481**
Max Lowe	Sheff Utd	27	499	0.014172	32.914369	0.0	0.8	**1455**
Rhys Norrington-Davies	Sheff Utd	25	251	0.007128	50.033798	0.0	0.8	**1412**
Jayden Bogle	Sheff Utd	23	2946	0.083667	0.000000	1740.8	0.8	**1897**
George Baldock	Sheff Utd	31	969	0.027520	36.415216	870.4	0.8	**1447**
Femi Seriki	Sheff Utd	22	360	0.010224	20.474682	0.0	0.8	**1475**
Vini Souza	Sheff Utd	24	2696	0.076567	0.000000	1414.4	0.8	**1488**
Gustavo Hamer	Sheff Utd	26	3217	0.091363	0.000000	1740.8	0.8	**1897**
Tom Davies	Sheff Utd	25	214	0.006078	69.063339	0.0	0.8	**1343**
Anis Slimane	Sheff Utd	23	411	0.011672	36.641268	0.0	0.8	**1447**
Andre Brooks	Sheff Utd	20	1088	0.030899	4.425573	870.4	0.8	**815**
Oliver Norwood	Sheff Utd	33	1703	0.048366	1.746051	1088.0	0.8	**1491**
Ben Osborn	Sheff Utd	29	1430	0.040612	7.761583	870.4	0.8	**1486**
Oliver Arblaster	Sheff Utd	20	2850	0.080941	0.000000	1414.4	0.8	**1418**
Ben Brereton Díaz	Sheff Utd	25	1759	0.049956	0.000000	1088.0	0.8	**1488**
Cameron Archer	Sheff Utd	22	2065	0.058646	3.308637	1088.0	0.8	**1488**
Oli McBurnie	Sheff Utd	27	1354	0.038454	15.935488	870.4	0.8	**1480**
Rhian Brewster	Sheff Utd	24	358	0.010167	17.855917	0.0	0.8	**1478**
William Osula	Sheff Utd	20	943	0.026781	6.153313	544.0	**0.8**	**810**
Daniel Jebbison	Sheff Utd	20	20	0.000568	42.138616	0.0	**0.8**	**780**

Table C.6: Sheffield United's results with $\lambda = 0.2$

Name	Club	Age	Minutes Played Total 23/24:	Weight	Predicted Days Out	Min Playing Time	Lambda	**Suggested Minutes**
Anel Ahmedhodzic	Sheff Utd	25	2739	0.077788	16.574005	1414.4	0.2	**1436**
Auston Trusty	Sheff Utd	25	2752	0.078157	6.146398	1414.4	0.2	**1436**
John Egan	Sheff Utd	31	574	0.016302	131.780905	544.0	0.2	**1438**
Jack Robinson	Sheff Utd	30	2976	0.084519	7.161556	1740.8	0.2	**1910**
Chris Basham	Sheff Utd	35	408	0.011587	74.240789	0.0	0.2	**1437**
Yasser Larouci	Sheff Utd	23	629	0.017864	14.656405	544.0	0.2	**1436**
Max Lowe	Sheff Utd	27	499	0.014172	32.914369	0.0	0.2	**1436**
Rhys Norrington-Davies	Sheff Utd	25	251	0.007128	50.033798	0.0	0.2	**1436**
Jayden Bogle	Sheff Utd	23	2946	0.083667	0.000000	1740.8	0.2	**2051**
George Baldock	Sheff Utd	31	969	0.027520	36.415216	870.4	0.2	**1436**
Femi Seriki	Sheff Utd	22	360	0.010224	20.474682	0.0	0.2	**1436**
Vini Souza	Sheff Utd	24	2696	0.076567	0.000000	1414.4	0.2	**1436**
Gustavo Hamer	Sheff Utd	26	3217	0.091363	0.000000	1740.8	0.2	**2051**
Tom Davies	Sheff Utd	25	214	0.006078	69.063339	0.0	0.2	**1437**
Anis Slimane	Sheff Utd	23	411	0.011672	36.641268	0.0	0.2	**1436**
Andre Brooks	Sheff Utd	20	1088	0.030899	4.425573	870.4	0.2	**1303**
Oliver Norwood	Sheff Utd	33	1703	0.048366	1.746051	1088.0	0.2	**1436**
Ben Osborn	Sheff Utd	29	1430	0.040612	7.761583	870.4	0.2	**1436**
Oliver Arblaster	Sheff Utd	20	2850	0.080941	0.000000	1414.4	0.2	**1638**
Ben Brereton Díaz	Sheff Utd	25	1759	0.049956	0.000000	1088.0	0.2	**1436**
Cameron Archer	Sheff Utd	22	2065	0.058646	3.308637	1088.0	0.2	**1436**
Oli McBurnie	Sheff Utd	27	1354	0.038454	15.935488	870.4	0.2	**1436**
Rhian Brewster	Sheff Utd	24	358	0.010167	17.855917	0.0	0.2	**1436**
William Osula	Sheff Utd	20	943	0.026781	6.153313	**544.0**	0.2	**596**
Daniel Jebbison	Sheff Utd	20	20	0.000568	42.138616	**0.0**	0.2	**599**

C.4 Real Madrid

Table C.7: Real Madrid's results with $\lambda = 0.6$

Name	Club	Age	Minutes Played Total 23/24:	Weight	Predicted Days Out	Min Playing Time	Lambda	**Suggested Minutes**
Éder Militão	Real Madrid	26	505	0.009740	161.049497	0.0	0.6	**1459**
David Alaba	Real Madrid	31	1407	0.027136	137.460755	0.0	0.6	**1736**
Antonio Rüdiger	Real Madrid	31	3986	0.076876	13.708890	4390.4	0.6	**4473**
Nacho Fernández	Real Madrid	34	3035	0.058534	15.343626	2744.0	0.6	**2563**
Ferland Mendy	Real Madrid	28	2744	0.052922	51.527754	2195.2	0.6	**2376**
Fran García	Real Madrid	24	1686	0.032517	35.197406	0.0	0.6	**2432**
Daniel Carvajal	Real Madrid	32	3261	0.062893	25.899997	3567.2	0.6	**2978**
Lucas Vázquez	Real Madrid	32	1780	0.034330	16.854470	0.0	0.6	**2470**
Aurélien Tchouaméni	Real Madrid	24	2809	0.054176	25.589066	2744.0	0.6	**2546**
Federico Valverde	Real Madrid	25	4190	0.080810	4.016790	4390.4	0.6	**4530**
Eduardo Camavinga	Real Madrid	21	2733	0.052710	10.931446	2195.2	0.6	**1830**
Toni Kroos	Real Madrid	34	3094	0.059672	8.981627	3567.2	0.6	**3001**
Luka Modric	Real Madrid	38	2238	0.043163	5.119133	1372.0	0.6	**1834**
Dani Ceballos	Real Madrid	27	862	0.016625	43.860810	0.0	0.6	**2405**
Jude Bellingham	Real Madrid	20	3559	0.068640	33.097226	3567.2	0.6	**2315**
Brahim Díaz	Real Madrid	24	2064	0.039807	3.557439	1372.0	0.6	**2481**
Arda Güler	Real Madrid	19	440	0.008486	122.694315	0.0	0.6	**1242**
Nico Paz	Real Madrid	19	2757	0.053173	10.600582	2195.2	0.6	**1831**
Vinicius Junior	Real Madrid	23	2994	0.057743	43.992237	2744.0	0.6	**2496**
Rodrygo	Real Madrid	23	3616	0.069740	13.143800	4390.4	0.6	**4524**
Joselu	Real Madrid	34	2090	0.040309	9.265594	1372.0	0.6	**2478**

Table C.8: Real Madrid's results with $\lambda = 0.2$

Name	Club	Age	Minutes Played Total 23/24:	Weight	Predicted Days Out	Min Playing Time	Lambda	**Suggested Minutes**
Éder Militão	Real Madrid	26	505	0.009740	161.049497	0.0	0.2	**2305**
David Alaba	Real Madrid	31	1407	0.027136	137.460755	0.0	0.2	**2315**
Antonio Rüdiger	Real Madrid	31	3986	0.076876	13.708890	4390.4	0.2	**3860**
Nacho Fernández	Real Madrid	34	3035	0.058534	15.343626	2744.0	0.2	**2356**
Ferland Mendy	Real Madrid	28	2744	0.052922	51.527754	2195.2	0.2	**2339**
Fran García	Real Madrid	24	1686	0.032517	35.197406	0.0	0.2	**2341**
Daniel Carvajal	Real Madrid	32	3261	0.062893	25.899997	3567.2	0.2	**4197**
Lucas Vázquez	Real Madrid	32	1780	0.034330	16.854470	0.0	0.2	**2343**
Aurélien Tchouaméni	Real Madrid	24	2809	0.054176	25.589066	2744.0	0.2	**2352**
Federico Valverde	Real Madrid	25	4190	0.080810	4.016790	4390.4	0.2	**3879**
Eduardo Camavinga	Real Madrid	21	2733	0.052710	10.931446	2195.2	0.2	**1585**
Toni Kroos	Real Madrid	34	3094	0.059672	8.981627	3567.2	0.2	**4699**
Luka Modric	Real Madrid	38	2238	0.043163	5.119133	1372.0	0.2	**929**
Dani Ceballos	Real Madrid	27	862	0.016625	43.860810	0.0	0.2	**2340**
Jude Bellingham	Real Madrid	20	3559	0.068640	33.097226	3567.2	0.2	**2985**
Brahim Díaz	Real Madrid	24	2064	0.039807	3.557439	1372.0	0.2	**2343**
Arda Güler	Real Madrid	19	440	0.008486	122.694315	0.0	0.2	**717**
Nico Paz	Real Madrid	19	2757	0.053173	10.600582	2195.2	0.2	**1588**
Vinicius Junior	Real Madrid	23	2994	0.057743	43.992237	2744.0	0.2	**2342**
Rodrygo	Real Madrid	23	3616	0.069740	13.143800	4390.4	0.2	**3843**
Joselu	Real Madrid	34	2090	0.040309	9.265594	1372.0	0.2	**2343**

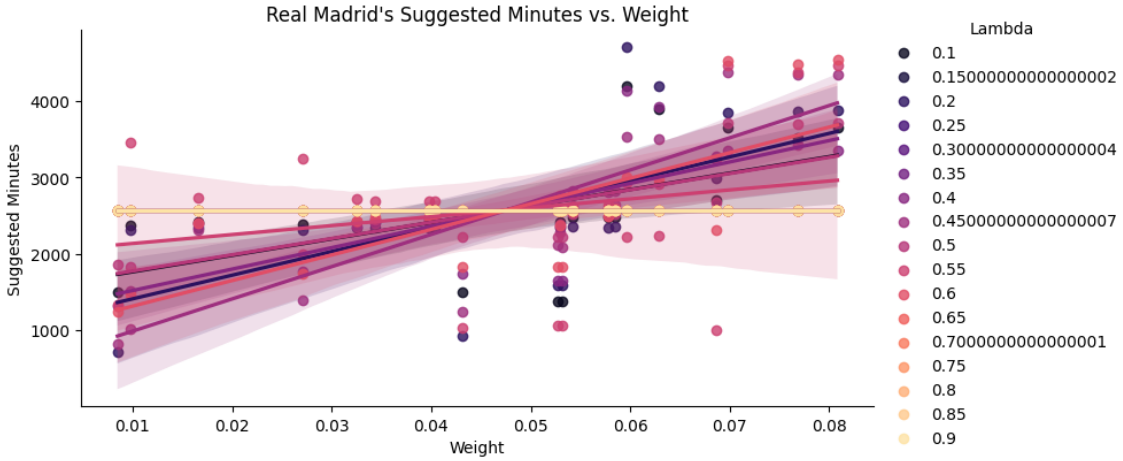


Figure C.2: Real Madrid's Suggested Minutes vs. Weight as λ varies

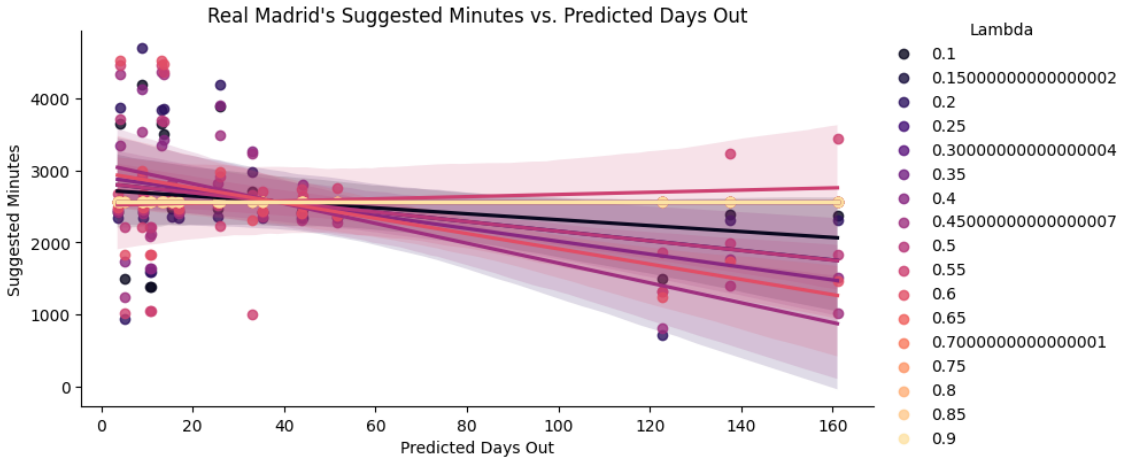


Figure C.3: Real Madrid's Suggested Minutes vs. Total Days as λ varies

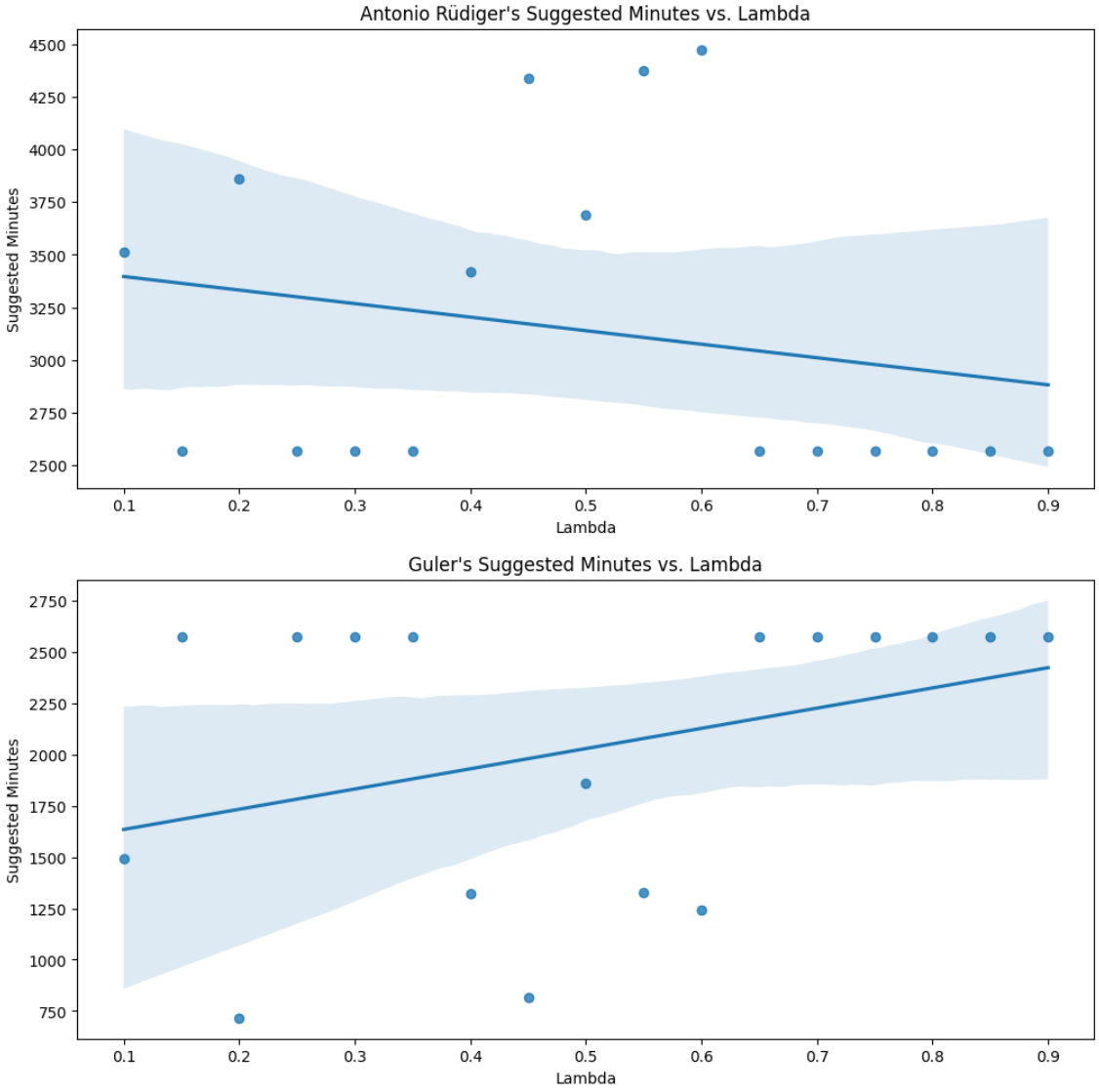


Figure C.4: Antonio Rüdiger's (Important) and Arda Güler's (Less Important) Suggested Minutes vs. Lambda

C.5 CA Osasuna

Table C.9: CA Osasuna's results with $\lambda = 0.9$

Name	Club	Age	Minutes Played Total 23/24:	Weight	Predicted Days Out	Min Playing Time	Lambda	**Suggested Minutes**
David García	CA Osasuna	30	2816	0.075073	41.183196	2737.8	0.9	**1900**
Alejandro Catena	CA Osasuna	29	2862	0.076300	17.944567	3369.6	0.9	**2333**
Unai García	CA Osasuna	32	559	0.014903	88.584142	0.0	0.9	**1801**
Jorge Herrando	CA Osasuna	23	1764	0.047027	26.861534	1053.0	0.9	**1889**
Juan Cruz	CA Osasuna	31	2040	0.054385	22.919017	1053.0	0.9	**1891**
Johan Mojica	CA Osasuna	31	2298	0.061264	19.712879	2106.0	0.9	**1899**
Jesús Areso	CA Osasuna	24	2899	0.077286	0.000000	3369.6	0.9	**2805**
Rubén Peña	CA Osasuna	32	2411	0.064276	29.084247	2106.0	0.9	**1893**
Lucas Torró	CA Osasuna	29	2141	0.057078	10.951659	1684.8	0.9	**1897**
Iker Muñoz	CA Osasuna	21	1923	0.051266	8.601684	1053.0	0.9	**603**
Jon Moncayola	CA Osasuna	26	2106	0.056145	6.632598	1684.8	0.9	**1897**
Pablo Ibáñez	CA Osasuna	25	1259	0.033564	0.000000	0.0	0.9	**1898**
Aimar Oroz	CA Osasuna	22	2655	0.070781	18.211386	2737.8	0.9	**1918**
Moi Gómez	CA Osasuna	29	2252	0.060037	32.691527	2106.0	0.9	**1890**
Rubén García	CA Osasuna	30	1757	0.046841	14.552830	0.0	0.9	**1895**
José Manuel Arnáiz	CA Osasuna	29	1136	0.030285	24.628379	0.0	0.9	**1890**
Kike Barja	CA Osasuna	27	397	0.010584	34.614847	0.0	0.9	**1883**
Ante Budimir	CA Osasuna	32	2761	0.073607	18.895933	2737.8	0.9	**1918**
Raúl García	CA Osasuna	23	1474	0.039296	0.000000	0.0	0.9	**1898**

Table C.10: CA Osasuna's results with $\lambda = 0.4$

Name	Club	Age	Minutes Played Total 23/24:	Weight	Predicted Days Out	Min Playing Time	Lambda	**Suggested Minutes**
David García	CA Osasuna	30	2816	0.075073	41.183196	2737.8	0.4	**1995**
Alejandro Catena	CA Osasuna	29	2862	0.076300	17.944567	3369.6	0.4	**2133**
Unai García	CA Osasuna	32	559	0.014903	88.584142	0.0	0.4	**1782**
Jorge Herrando	CA Osasuna	23	1764	0.047027	26.861534	1053.0	0.4	**1830**
Juan Cruz	CA Osasuna	31	2040	0.054385	22.919017	1053.0	0.4	**1832**
Johan Mojica	CA Osasuna	31	2298	0.061264	19.712879	2106.0	0.4	**1875**
Jesús Areso	CA Osasuna	24	2899	0.077286	0.000000	3369.6	0.4	**3429**
Rubén Peña	CA Osasuna	32	2411	0.064276	29.084247	2106.0	0.4	**1872**
Lucas Torró	CA Osasuna	29	2141	0.057078	10.951659	1684.8	0.4	**1834**
Iker Muñoz	CA Osasuna	21	1923	0.051266	8.601684	1053.0	0.4	**540**
Jon Moncayola	CA Osasuna	26	2106	0.056145	6.632598	1684.8	0.4	**1835**
Pablo Ibáñez	CA Osasuna	25	1259	0.033564	0.000000	0.0	0.4	**1835**
Aimar Oroz	CA Osasuna	22	2655	0.070781	18.211386	2737.8	0.4	**2004**
Moi Gómez	CA Osasuna	29	2252	0.060037	32.691527	2106.0	0.4	**1871**
Rubén García	CA Osasuna	30	1757	0.046841	14.552830	0.0	0.4	**1834**
José Manuel Arnáiz	CA Osasuna	29	1136	0.030285	24.628379	0.0	0.4	**1831**
Kike Barja	CA Osasuna	27	397	0.010584	34.614847	0.0	0.4	**1827**
Ante Budimir	CA Osasuna	32	2761	0.073607	18.895933	2737.8	0.4	**2004**
Raúl García	CA Osasuna	23	1474	0.039296	0.000000	0.0	0.4	**1835**

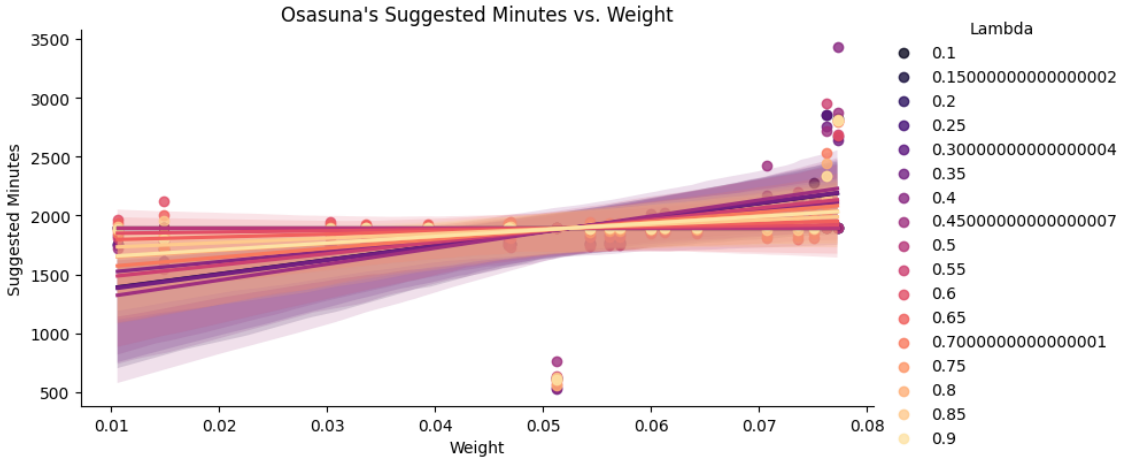


Figure C.5: CA Osasuna's Suggested Minutes vs. Weight as λ varies

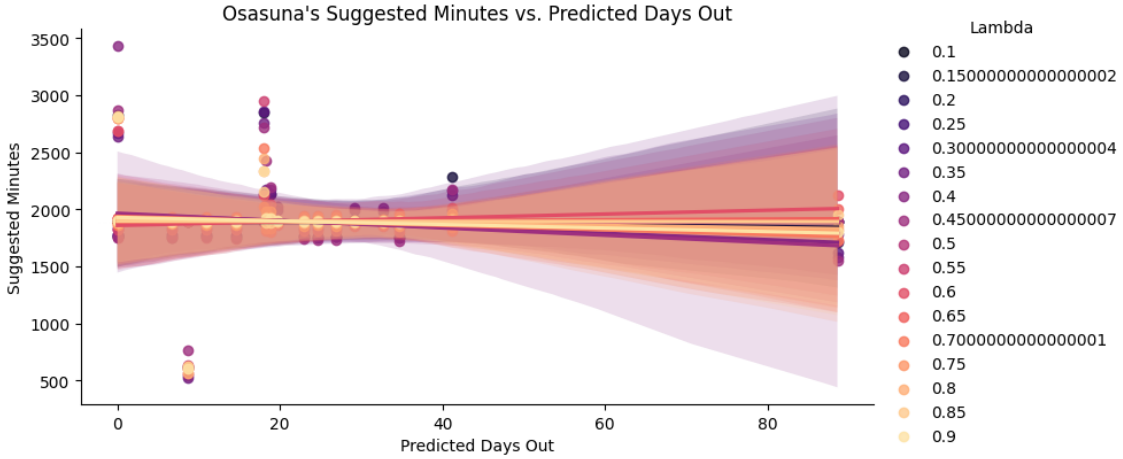


Figure C.6: CA Osasuna's Suggested Minutes vs. Total Days as λ varies

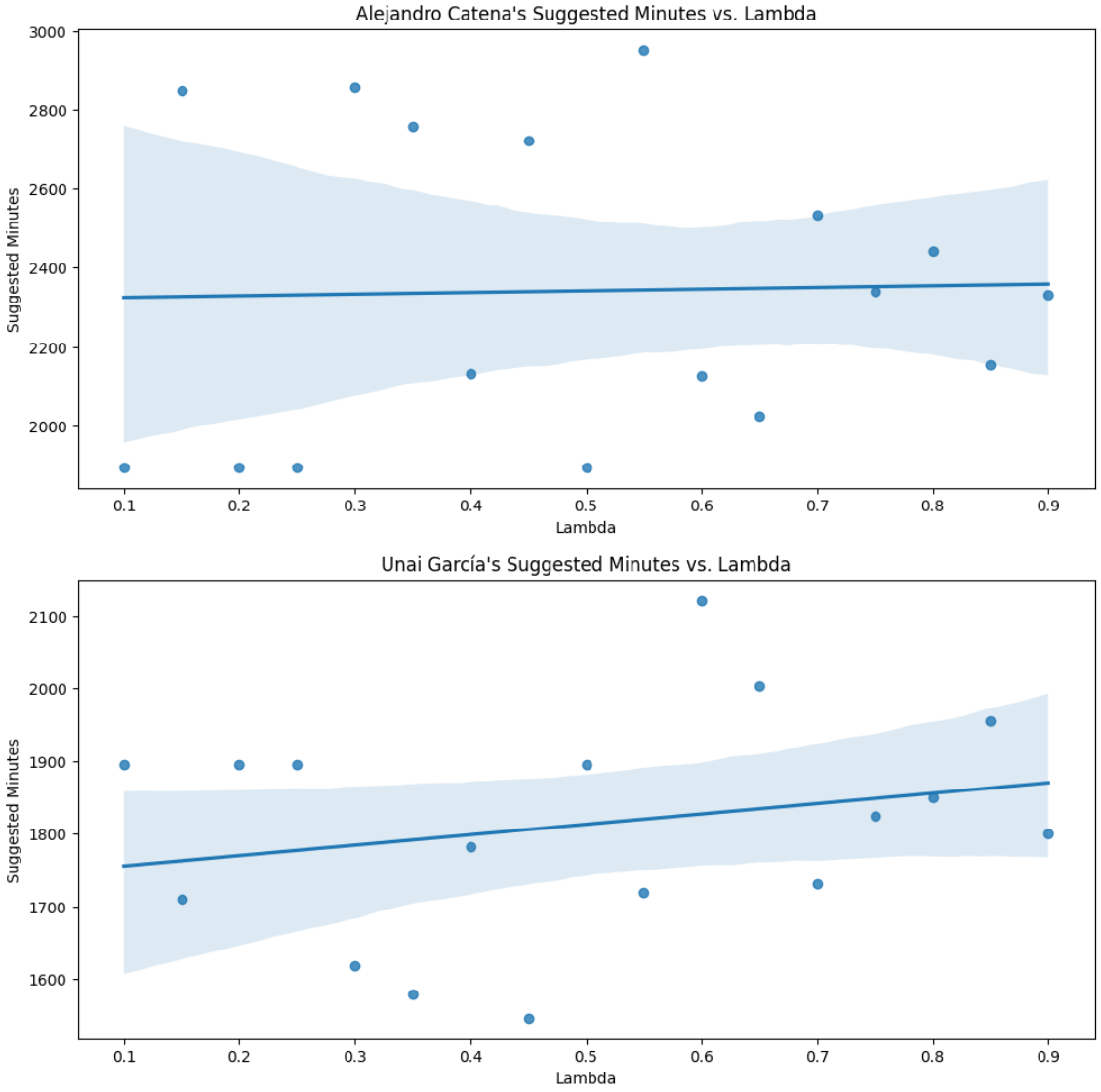


Figure C.7: Alejandro Catena's (Important) and Unai García's (Less Important) Suggested Minutes vs. Lambda

C.6 Granada

Table C.11: Granada's results with $\lambda = 0.6$

Name	Club	Age	Minutes Played Total 23/24:	Weight	Predicted Days Out	Min Playing Time	Lambda	**Suggested Minutes**
Bruno Méndez	Granada CF	24	1336	0.037744	24.527861	1091.2	0.6	**1565**
Kamil Piatkowski	Granada CF	23	1077	0.030427	8.325076	682.0	0.6	**1565**
Raúl Torrente	Granada CF	22	1250	0.035315	16.771774	682.0	0.6	**1565**
Jesús Vallejo	Granada CF	27	106	0.002995	163.750577	0.0	0.6	**1565**
Ignasi Miquel	Granada CF	31	3006	0.084925	2.807759	2182.4	0.6	**1565**
Miguel Rubio	Granada CF	26	1560	0.044073	17.193781	1091.2	0.6	**1565**
Faitout Maouassa	Granada CF	25	645	0.018222	16.771791	0.0	0.6	**1565**
Carlos Neva	Granada CF	27	3156	0.089163	4.111828	2182.4	0.6	**1565**
Ricard Sánchez	Granada CF	24	2299	0.064951	0.000000	1364.0	0.6	**1565**
Víctor Díaz	Granada CF	35	252	0.007119	53.306429	0.0	0.6	**1565**
Gerard Gumbau	Granada CF	29	2714	0.076675	0.000000	1773.2	0.6	**1565**
Martin Hongla	Granada CF	26	1805	0.050994	0.742741	1364.0	0.6	**1565**
Gonzalo Villar	Granada CF	26	1841	0.052012	0.000000	1364.0	0.6	**1565**
Sergio Ruiz	Granada CF	29	2519	0.071166	0.352159	1773.2	0.6	**1565**
Óscar Melendo	Granada CF	26	1126	0.031812	4.209096	682.0	0.6	**1565**
Myrto Uzuni	Granada CF	29	2459	0.069471	4.142879	1773.2	0.6	**1565**
Kamil Jóźwiak	Granada CF	26	572	0.016160	32.505333	0.0	0.6	**1565**
Facundo Pellistri	Granada CF	22	1587	0.044836	10.349203	1091.2	0.6	**1565**
Antonio Puertas	Granada CF	32	715	0.020200	7.608153	0.0	0.6	**1565**
Theo Corbeanu	Granada CF	22	1364	0.038535	15.981357	1091.2	**0.6**	**1565**
José Callejón	Granada CF	37	881	0.024890	11.393583	0.0	**0.6**	**1565**
Matías Arezo	Granada CF	21	359	0.010142	24.833905	0.0	**0.6**	**1565**
Lucas Boyé	Granada CF	28	2767	0.078173	12.804128	2182.4	**0.6**	**1565**

Table C.12: Granada's results with $\lambda = 0.1$

Name	Club	Age	Minutes Played Total 23/24:	Weight	Predicted Days Out	Min Playing Time	Lambda	**Suggested Minutes**
Bruno Méndez	Granada CF	24	1336	0.037744	24.527861	1091.2	0.1	**1528**
Kamil Piatkowski	Granada CF	23	1077	0.030427	8.325076	682.0	0.1	**1530**
Raúl Torrente	Granada CF	22	1250	0.035315	16.771774	682.0	0.1	**1529**
Jesús Vallejo	Granada CF	27	106	0.002995	163.750577	0.0	0.1	**1455**
Ignasi Miquel	Granada CF	31	3006	0.084925	2.807759	2182.4	0.1	**1918**
Miguel Rubio	Granada CF	26	1560	0.044073	17.193781	1091.2	0.1	**1529**
Faitout Maouassa	Granada CF	25	645	0.018222	16.771791	0.0	0.1	**1529**
Carlos Neva	Granada CF	27	3156	0.089163	4.111828	2182.4	0.1	**2561**
Ricard Sánchez	Granada CF	24	2299	0.064951	0.000000	1364.0	0.1	**1530**
Víctor Díaz	Granada CF	35	252	0.007119	53.306429	0.0	0.1	**1522**
Gerard Gumbau	Granada CF	29	2714	0.076675	0.000000	1773.2	0.1	**1625**
Martin Hongla	Granada CF	26	1805	0.050994	0.742741	1364.0	0.1	**1530**
Gonzalo Villar	Granada CF	26	1841	0.052012	0.000000	1364.0	0.1	**1530**
Sergio Ruiz	Granada CF	29	2519	0.071166	0.352159	1773.2	0.1	**1625**
Óscar Melendo	Granada CF	26	1126	0.031812	4.209096	682.0	0.1	**1530**
Myrto Uzuni	Granada CF	29	2459	0.069471	4.142879	1773.2	0.1	**1625**
Kamil Jóźwiak	Granada CF	26	572	0.016160	32.505333	0.0	0.1	**1527**
Facundo Pellistri	Granada CF	22	1587	0.044836	10.349203	1091.2	0.1	**1530**
Antonio Puertas	Granada CF	32	715	0.020200	7.608153	0.0	0.1	**1530**
Theo Corbeanu	Granada CF	22	1364	0.038535	15.981357	**1091.2**	0.1	**1529**
José Callejón	Granada CF	37	881	0.024890	11.393583	**0.0**	0.1	**587**
Matías Arezo	Granada CF	21	359	0.010142	24.833905	**0.0**	0.1	**587**
Lucas Boyé	Granada CF	28	2767	0.078173	12.804128	**2182.4**	0.1	**2613**

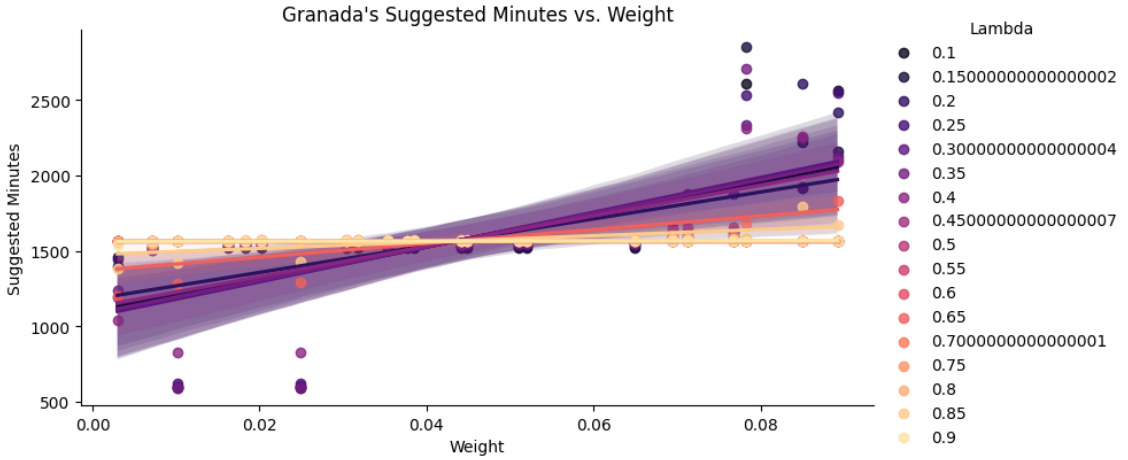


Figure C.8: Granada's Suggested Minutes vs. Weight as λ varies

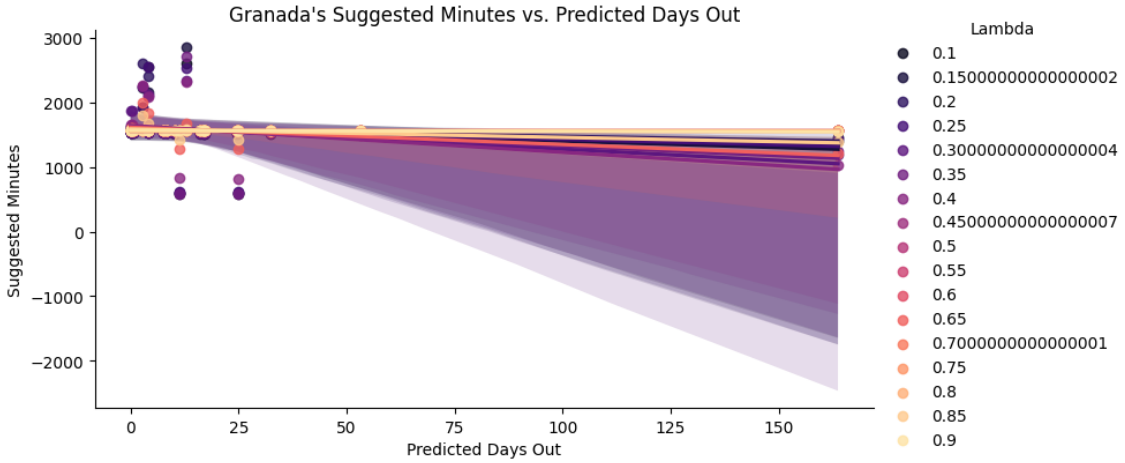


Figure C.9: Granada's Suggested Minutes vs. Total Days as λ varies

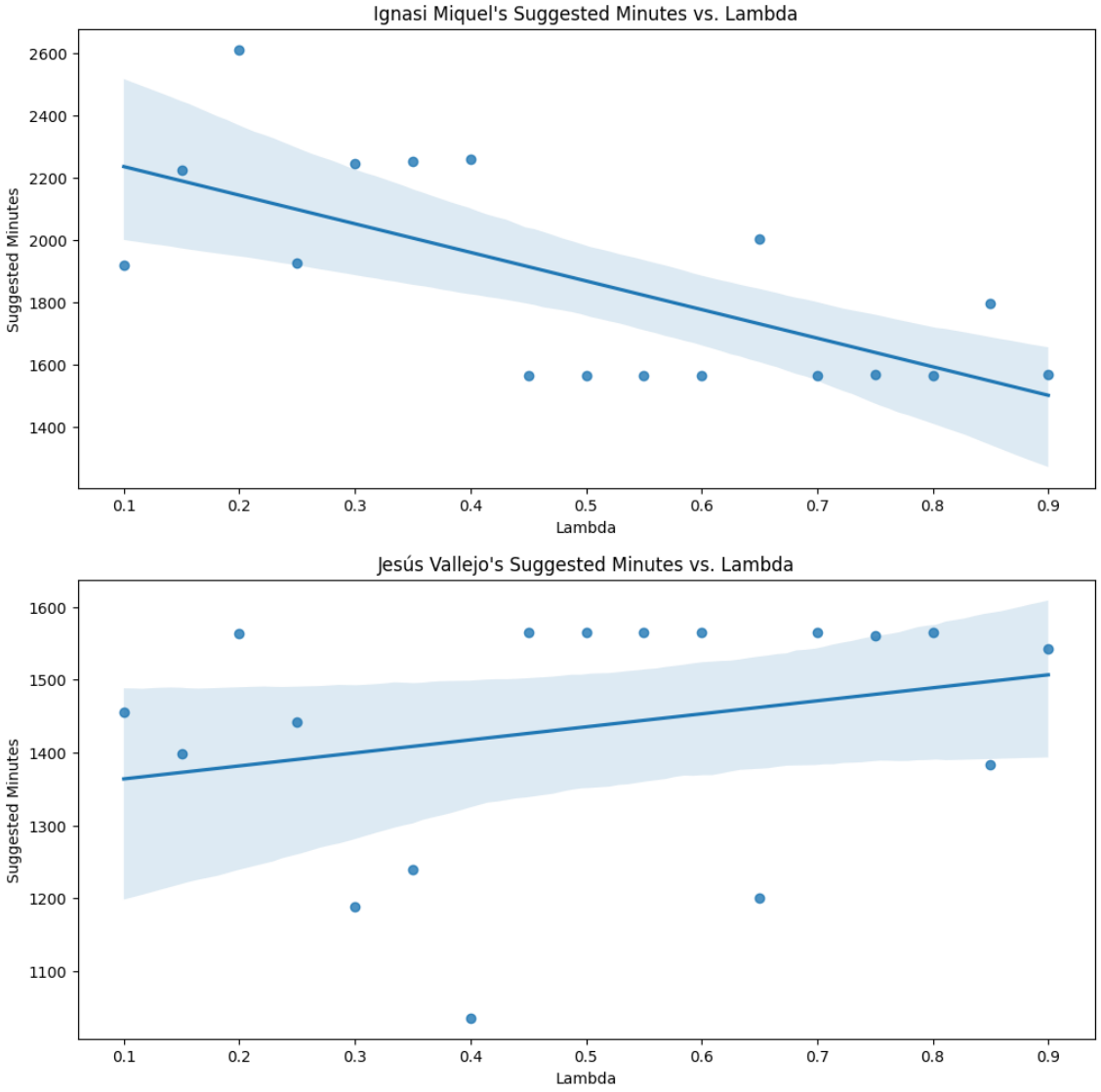


Figure C.10: Ignasi Miquel's (Important) and Jesús Vallejo's (Less Important) Suggested Minutes vs. Lambda

References

[Calderon-Díaz et al., 2023] Calderon-Díaz, M., Silvestre, R., Vásconez Hurtado, J., Yáñez, R., Roby, M., Querales, M., and Salas, R. (2023). Explainable machine learning techniques to predict muscle injuries in professional soccer players through biomechanical analysis. *Sensors (Basel, Switzerland)*, 24.

[Dut et al., 2023] Dut, R., Akgül, S., Dönmez, G., Ulkar, B., Kanbur, N., and derman, O. (2023). Adolescent male soccer players have higher growth rates and risk of injury is associated with biological maturity. *The Turkish Journal of Pediatrics*, 65:990.

[Eirale et al., 2012] Eirale, C., Tol, J., Farooq, A., Smiley, F., and Chalabi, H. (2012). Low injury rate strongly correlates with team success in qatari professional football. *British journal of sports medicine*, 47.

[Ekstrand, 2013] Ekstrand, J. (2013). Keeping your top players on the pitch: The key to football medicine at a professional level. *British Journal of Sports Medicine*, 47:723–724.

[Eliakim et al., 2020] Eliakim, E., Morgulev, E., Lidor, R., and Meckel, Y. (2020). Estimation of injury costs: Financial damage of english premier league teams' underachievement due to injuries. *BMJ Open Sport Exercise Medicine*, 6:e000675.

[Howden, 2023] Howden (2023). Clubs in europe's top five leagues spent nearly €40m on a total of 501 covid-19 absences: Howden uk.

[Hägglund et al., 2013] Hägglund, M., Waldén, M., Hedevik, H., Kristenson, K., Bengtsson, H., and Ekstrand, J. (2013). Injuries affect team performance negatively in professional football: An 11-year follow-up of the uefa champions league injury study. *British journal of sports medicine*, 47.

[Rossi et al., 2017] Rossi, A., Pappalardo, L., Cintia, P., Fernández, J., Iaia, F., and Medina, D. (2017). Who is going to get hurt? predicting injuries in professional soccer. In *Who is going to get hurt? Predicting injuries in professional soccer*.