# Optimizing Football Squad Planning: Balancing Competitive Success and Financial Sustainability

Pablo Galaz-Cares

## 1. Introduction

**Motivation**

The primary goal of professional football clubs is to achieve the best possible sporting results in all competitions they participate in. Consequently, the Sporting Director's role is to assemble the strongest squad, maximizing competitiveness and increasing the chances of winning championships through the proper selection of football players.

A squad refers to the group of players registered with a team during a season, represented through their federative rights, allowing the club to field them in various tournaments. Squad composition enables clubs to achieve sporting goals such as winning championships, qualifying for international tournaments, and securing monetary prizes. Each player possesses unique characteristics like age, nationality, position, height, contract length, and salary. Additionally, other key factors, such as quality and market value, are estimated, which may fluctuate over time.

The squad evolves due to changes in player contracts, typically defined by two main factors: 1) contract length, specifying how long the player will remain at the club, and 2) salary, representing the player's compensation. Changes in squad composition occur through player arrivals and departures, which might involve acquiring players to strengthen specific positions or releasing players (via transfers or contract termination) to create room for new arrivals. Players may also be promoted from the youth academy or acquired from the player market. If a player's contract is nearing expiration, the club must decide whether to renew it, often with new salary conditions. Consequently, the club's decision-making process must address the following:

- Which field positions should be improved?
- Should an expiring contract be renewed?
- Which players should be released?
- Which players should be loaned to other clubs?
- Should an offer to buy a player be accepted?
- Which young players should be promoted to the professional squad?
- Other decisions related to changes in players' current contracts

In football, these decisions are often made based on subjective evaluations, with no clear structure that integrates qualitative and quantitative analyses. Moreover, legal and financial restrictions, such as limits on foreign players or budget constraints for player signings and salaries, complicate these decisions.

**Objectives and Assumptions**

This work develops an analytical model to support professional football clubs in their squad planning. The objective is to define a methodology that helps the Sporting Director make optimal squad-related decisions, maximizing performance over multiple seasons, while balancing both sporting and financial goals. This is referred to as a multi-season squad planning strategy.

The model incorporates two key dimensions: sporting and business (legal & financial). In the sporting dimension, the club's main goal is success in competitions by consistently winning matches. While the aim is to win championships, other achievements also define success, such as qualifying for major international tournaments (e.g., Copa Libertadores or Champions League) or avoiding relegation.

On the business side, the primary objective is long-term financial sustainability, ensuring responsible spending and investments. This includes avoiding financial distress or bankruptcy and complying with financial fair play regulations and legal requirements imposed by governing bodies like FIFA and domestic leagues.

The model provides a strategic plan that informs decisions about current players and potential acquisitions from the market. It assumes that each player contributes to the club's sporting success and that building a stronger squad leads to better outcomes. The decision-making process considers not only the players' quality and market value evolution over time, but also trade-offs between sporting success and financial stability, ensuring that the club's competitive ambitions align with sound financial management.

**Contributions and Results**

This study aims to develop a methodology that enables football clubs to strategically plan and manage squad development over time, while considering both business constraints and tournament demands. The proposed model significantly advances current literature by integrating sporting and financial objectives into a multi-season squad planning framework, allowing clubs to achieve competitive success in alignment with their long-term vision.

One of the model's key contributions is its capacity to identify players in the market who can substantially impact the club's performance objectives. By evaluating players based on their potential to enhance the club's overall performance, the model helps clubs make data-driven acquisition decisions, ensuring that each player aligns with both the competitive and financial strategies.

Furthermore, the model determines the optimal timing for player transactions. This includes insights into when to sell a player to maximize financial returns or when to release a player to replace them with a more suitable alternative. These strategic decisions are critical for maintaining a balanced approach to both sporting and financial management.

The results demonstrate the model's effectiveness in guiding clubs through strategic squad development across different scenarios. By simulating various budget constraints and performance expectations, the

model provides clear insights into the trade-offs between developing homegrown talent and acquiring external players. For example, the sensitivity analysis revealed that larger budgets lead to better squad performance, but clubs can still maintain competitive levels with moderate budget reductions. However, significant cuts in the budget notably impact squad performance over time.

The model also highlights the importance of market timing in player transactions. It shows how strategic acquisitions can strengthen the squad without straining financial resources, allowing clubs to remain agile in the transfer market. Additionally, the model provides insights on when to sell players at peak market value, ensuring optimal financial returns while maintaining a high-performing squad.

Overall, the model underscores the importance of balancing competitive ambition with financial prudence. By investing selectively in high-impact players while fostering youth talent, clubs can achieve sustainable success without violating financial regulations. The results validate the model's practicality, offering football clubs a structured approach to long-term squad planning.

**Structure of the Article**

In the literature review section (Section 2), we examine related studies to provide context and background for our research. The model formulation section (Section 3) presents the proposed model, detailing its structure and the underlying assumptions. In the results section (Section 4), we demonstrate the model's effectiveness by analyzing the outcomes across various scenarios. Finally, in the conclusions section (Section 5), we discuss our findings, the implications of our assumptions, and offer our concluding thoughts.

---

# 2. Literature Review

**Workforce Planning**

Workforce planning is defined as the process of analyzing the current workforce, determining future needs, and identifying the gap between the present and the future. The goal is to ensure that an organization has the right people with the right skills at the right time. Factors such as people, skills, positions, and timing need to be aligned to fulfil strategic objectives. This topic is widely discussed in the literature [1].

In related studies, a four-step planning process is proposed, which includes: 1) supply analysis that studies the current and future composition of the workforce and considers turnover using metrics like the Attrition Rate Percentage, 2) demand analysis that examines future activities and organizational requirements, 3) gap analysis, and 4) solution analysis. The objectives of the last two steps involve defining the Workforce Shape, which concerns determining the required occupations, the timing for hiring, whether current personnel strategies suffice, and, if not, what alternative actions are necessary. The concept of the Workforce Shape could be particularly useful in structuring a football squad, providing a clear framework for determining the squad composition over time.

Managerial perspectives extend workforce planning by incorporating skills, which are classified as either hierarchical or categorical. Hierarchical skills suggest that increased skills correlate with greater work efficiency, such as age, experience, and expertise level, while categorical skills focus on what one can or cannot do, such as holding specific licenses or qualifications. The literature identifies five main elements in skill evaluation: labour costs, work efficiency, work quality, tasks that can be performed, and resulting flexibility. Among these, costs, efficiency, and quality are the most frequently discussed outcomes of hierarchical skills. This hierarchical structure could be analogous to the classification of players in sports, where higher-quality players generally incur higher labour costs and the categorial structure could be related with the field position of the players [2].

Adaptive skill pool decisions involve hiring or dismissing temporary or permanent workers, leading to a trade-off between flexibility and cost. Greater flexibility requires higher expenditure, including firing compensation and learning time for new hires. From a technical perspective, incorporating skills into workforce planning adds complexity due to heterogeneity, often addressed using mixed-integer problems, integer problems, and heuristic methods. The mixed-integer problem approach could be relevant when modelling decision-making processes for player selection and squad management in sports.

Existing literature, such as [2] identifies application areas for solving these kinds of problems, including services, production, and the military, but notably does not mention sports organizations. Furthermore, only a few studies have implemented these approaches with real-world data. A panel from the National Academy of Public Administration (NAPA) highlights that effective workforce planning must include a "Linkage to Other Plans," ensuring that workforce plans are integrated with strategic and financial plans, making them relevant to the strategic intent and financially viable given limited resources. This is particularly important in sports organizations where budget constraints are a critical factor [3].

**Workforce Planning in Sports Organizations**
Sports organizations face unique challenges in continually developing their workforce. They must design, develop, and sustain roles requiring advanced technical and cognitive skills, provide opportunities for higher-order skill development, facilitate skill transfer, and encourage the development of skills with long-term value beyond the employing organization. These requirements are more complex than traditional workforce planning scenarios due to the dynamic nature of sports and the need for continual performance optimization.

Research by [4] explores the relationship between workforce diversity and organizational performance, specifically examining primary identity dimensions like nationality and age. The study assessed the performance of football clubs based on league points and standings for the 2016/17 season. Interestingly, the results indicated no significant relationship between age diversity and performance, suggesting that other factors may play a more critical role in determining success in sports organizations.

**Squad and Players Selection in Sports**

In the context of sports, player selection involves optimizing the composition of a team from a pool of available players. For football, [5] proposes an optimization model to select players from the market to form an effective lineup. This problem is approached as an integer optimization model, focusing primarily on assignment constraints. However, the problem becomes more complex in other sports, such as volleyball, due to additional factors like positional rotations and special player roles (e.g., the libero). For these cases, mixed-integer optimization problems are developed and tested using real data. In the football context, additional constraints, such as budget limitations and managerial considerations, should be incorporated into the model to reflect the complexities of real-world decision-making.

In cricket, [6] presents a multi-objective optimization model for selecting a final team from the available market. The methodology is generic and could easily be adapted for other sports like football or baseball. The model includes budget constraints and a predefined squad structure (e.g., selecting a captain, a team composition, and nine additional players). A sensitivity analysis of the budget is also performed to evaluate how different financial scenarios affect team selection. This approach could be extended to football by considering various squad structures, such as depth (two players per position) or flexible groupings (15 competitive players capable of switching positions on the field).

**Measuring Football Players Performance**

Early studies on football player performance, such as [7], focused primarily on physical dimensions, citing the complexity of the game as a barrier to evaluating technical and tactical aspects. However, with the rise of big data, more comprehensive datasets have become available, allowing for richer analysis. For instance, public access to player event data from the top five European leagues, provided by platforms like [8], has opened new avenues for research. This data, sourced from Opta, enables more detailed performance evaluations than previously possible with video analysis tools like Wyscout or Instats.

Expanding on this data, researchers have explored relationships between player market value and performance using econometric models that incorporate aggregated statistics (e.g., pass accuracy, defensive ground duels) from sources like WhoScored. These models often utilize a rating scale (0-10) to assess player performance, with higher scores indicating better performance. The literature on market value estimation for players frequently employs linear regression models and classification algorithms, although much of this work relies on artificial datasets from games like FIFA or Football Manager [9].

Despite advancements, few studies focus on player performance itself. For example, [10] develops two algorithms to estimate overall performance scores and market value projections for players, using linear regressions on data from Sofifa and the FIFA video game. Similarly, [11] uses an artificial neural network to assess player attributes like technique, speed, physicality, and resistance based on data from the PES football game, providing an overall score to inform team-building decisions. Incorporating a policy of selection criteria based on these scores could be valuable for optimizing squad composition.

As highlighted by [12], there are several challenges in this area of research: the need for standardized datasets for consistent experimentation, comparative studies to evaluate different machine learning

techniques, and a clear understanding of which analyses are most critical for football analytics. This study addresses these challenges by leveraging real data, employing advanced algorithms, and aligning the model with the strategic goals of sports managers.

This work addresses several critical challenges highlighted by gaps in the existing literature. One significant challenge is the lack of tailored optimization models for sports organizations, particularly football clubs, that integrate both budget constraints and managerial considerations. Many models used in other sectors, such as services and production, do not account for the unique dynamics of sports teams, including the need for a flexible squad structure and the specific financial constraints faced by clubs. Additionally, there is a need for more sophisticated approaches that combine hierarchical skill evaluations with adaptive decision-making, as current models often fall short in incorporating these elements effectively. This study also aims to bridge the gap in empirical applications by using real-world data to validate the model, addressing the scarcity of practical implementations in sports contexts. By tackling these gaps, the research provides a more relevant and actionable framework for football clubs to optimize their squad planning and financial management.

# 3. Model Formulation/Methodology

The goal of the proposed methodology is to provide recommendations to the sporting director regarding current and future decisions related to squad composition for the upcoming seasons. The decisions addressed pertain to the contractual relationship that a player may have with the club. To achieve this, a finite set of states representing this relationship is defined. These states include: the player being part of the club or having no affiliation with the club. These decisions are projected over future seasons, allowing for better management of current player contracts and potential contracts offers to players belonging to other clubs.

To represent this set of decisions, an Integer Programming (IP) model is presented. For each player in the selected universe, the model determines the type of affiliation they should have with the club for all subsequent seasons within a predetermined finite horizon. This optimization model requires that each player in the selected universe be represented by their basic characteristics, as well as by certain attributes that will be projected over time, such as their performance score and market value. This set of characteristics will be referred to as the "player profile over time."

To approach this problem holistically, a methodology with different modules is proposed to address each of the modelling challenges required for this management problem. Figure 1 illustrates the five proposed modules, which range from the input data needed for the entire process to the final optimization model that provides recommendations on contractual decisions for each player studied. The following sections explain each of these modules in detail.
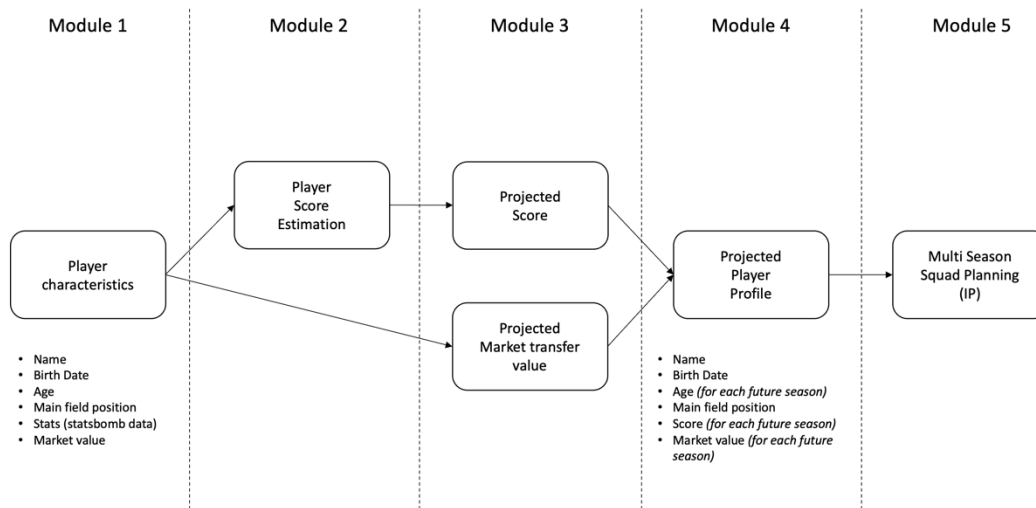
Figure 1: Modules

### Module 1 - Player Characteristics

This module gathers certain player characteristics necessary for the subsequent mathematical models. The characteristics are divided into three subsets: basic data, performance statistics, and financial attributes. The first subset includes data such as name, date of birth, nationality, and primary playing position. The second subset contains data related to the player's sports performance statistics. The data to be used are from the last two seasons of the Premier League (2022/2023 and 2023/2024) and include event data provided by StatsBomb. The third subset will seek the market players for the seasons which were obtained through a Web Scrapping to TransferMarkt web site.

### Module 2 – Player Score Estimation

In this module, player performance statistics are used to estimate a score for each player for the seasons. The focus will be on the On-Ball Value (OBV) variable of actions recorded in the provided event data from StatsBomb. The output of this module will be a score value for each player for the 2022/2023 and 2023/2024 seasons based on the quality of their recorded actions.

The score ranges from 0 to 100 and provides insight into a player's contribution both in generating goal-scoring opportunities for their team and in preventing them for the opposition. To capture these two effects, we use the On-Ball Value (OBV) metrics available from StatsBomb. This score is created as a linear combination of both effects. In Formula 1, we present the mathematical formulation, where the OBV values represent the percentile rank of the player relative to other players in the league playing in the same position, for each season for which data is available. The weights $w_{For}^{f_p}$ and $w_{Against}^{f_p}$ used for this linear combination vary by field position $f_p$. Some positions place greater importance on creating goal-scoring opportunities for the team, while others emphasize preventing goal-scoring chances for the opposition.

$$Score_p^t = w_{For}^{f_p} \cdot OBV_{p,t}^{For} + w_{Against}^{f_p} \cdot OBV_{p,t}^{Against} \quad \forall p \in P, t \in \text{T} \qquad (1)$$

The $OBV_{For}$ and $OBV_{Against}$ metrics represent the player's percentile rank in these respective facets. These values are also calculated as a linear combination of each player's relative contribution in each event, compared to players in the league who play in the same position. To compute this, we use a set of 10 events, detailed in Table 1. These events are weighted by $v_{e,f_p}^{For}$ and $v_{e,f_p}^{Against}$ differently depending on whether they are $OBV_{For}$ or $OBV_{Against}$ and, vary by field position $f_p$ and event $e$. As such, certain events, such as shots or dribbles, are more highly valued for offensive positions, while others, such as blocks or recoveries, are more valued for defensive positions. Formulas 2 present the formulation for OBV For and Formula 3 for OBV Against.

$$OBV_{p,t}^{For} = \sum_e v_{e,f_p}^{For} \cdot \overline{OBV_{p,t,e}^{For}} \quad \forall p \in P, t \in T \text{ and } e \in Events \quad (2)$$

$$OBV_{p,t}^{Against} = \sum_e v_{e,f_p}^{Against} \cdot \overline{OBV_{p,t,e}^{Against}} \quad \forall p \in P, t \in T \text{ and } e \in Events \quad (3)$$

Finally, the percentile for each action type is calculated as the player's relative position compared to players in the same position. The index is constructed as: 1) the average contribution per action for each player minus 2) the average contribution per action of players in the league who play in the same position. Formulation is presented in Formulas 4 and 5. Examples of this calculation can be found in the results section.

$$\overline{OBV_{p,t,e}^{For}} = \overline{OBV_{p,t,e}^{PlayerFor}} - \overline{OBV_{t,e}^{LeagueFor}} \quad \forall p \in P, t \in T, e \in E \qquad (4)$$

$$\overline{OBV_{p,t,e}^{Against}} = \overline{OBV_{p,t,e}^{PlayerAgainst}} - \overline{OBV_{t,e}^{LeagueAgainst}} \quad \forall p \in P, t \in T, e \in Events \quad (5)$$

Player position is determined by the mode of minutes played, meaning that each player is assigned to a single position for this analysis, corresponding to the position in which they played the most minutes during the season. Considered positions are in Table 2 and their field distribution are shown in Image 2.

| Events (10) | |
|---|---|
| Shot | Block |
| Pass | Clearance |
| Carry | Interception |
| Dribble | Duel |
| Foul Committed | Goalkeeper Save |

Table 1: Considered Events

| Field Positions (11) | |
| --- | --- |
| Goalkeeper (GK) | Defensive Midfield (DM) |
| Left Center Back (LCB) | Center Midfield (CM) |
| Right Center Back (RCB) | Offensive Midfield (OM) |
| Left Back (LB) | Left Wing (LW) |
| Right Back (RB) | Right Wing (RW) |
| | Center Forward (CF) |

Table 2: Considered Positions
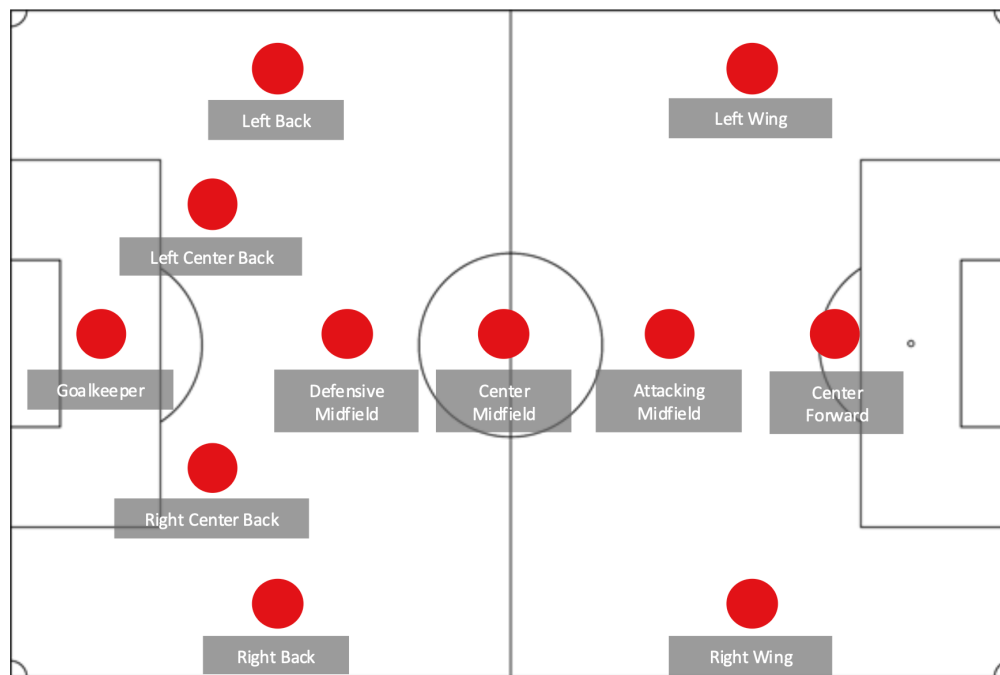


Image 2: Field positions distribution.

The considered weights were proposed and discussed between Sporting Director and the club's staff. For more information, Appendix A and Appendix B.

**Module 3 - Machine Learning Prediction Models**
This module utilizes machine learning models to project two player characteristics for future seasons, thus creating a Projected Player Profile. These two characteristics are: 1) performance score, and 2) market value. The following sections will detail the data and models used for projecting each characteristic.

1. **Performance Score:** as it was explained before, the Player Score is estimated with the model of Module 2. To estimate the performance score for the next seasons the following model is proposed.

$$\Delta score_{t,t-1} \sim score_{t-1} + age_t + age_t^2 + position \qquad (6)$$

2. **Market Value:** using the Web Scrapped data from TransferMarkt, the following model is proposed.

$$\Delta log(market\ value)_{t,t-1} \sim log(market\ value)_{t-1} + age_t + age_t^2 + position \qquad (7)$$

To estimate the Performance Score and Market Value of each player for each season, a set of models are going to be tested and measure their performance. Models are shown in Table 3. To measure the performance of the models and make them comparable, we are going to check two measure errors: RMSE (root mean squared error) and MAE (mean absolute error). Algo, Shap Values are evaluated to understand the importance of each variable to each model.

| Models (8) | |
|---|---|
| Linear Regression | Random Forest |
| Ridge Regression | xG Boost |
| LASSO Regression | Support Vector Regression |
| Decision Tree | Artificial Neural Networks |

Table 3: Considered models to calibrate and evaluate the predictions.

**Module 4 - Projected Player Profile (PPP)**

By combining the basic characteristics of the player that do not change over time (Module 1) with the projected characteristics (Module 3), we can form the projected player profile. This module aims to consolidate all the previous information to provide a well-processed input for the Integer Programming model, which will be explained in the following module. In Table 3 are shown all characteristics considered to create the Projected Player Profile.

| Projected Player Profile Attributes | |
|---|---|
| General | For Each Season |
| Name | Age |
| Date of Birth | Is Youth (< 21 y.o.) |
| Position | Score |
| | Market Value |

Table 3: Player Characteristics. Some change through the seasons, others don't.

**Module 5 - Multi-Season Squad Planning**

After all data collection and processing to generate the Projected player profile, we move to this module, which consists of an Integer Programming (IP) optimization problem. This model aims to maximize squad quality subject to budget constraints for player acquisitions and salaries, tournament regulations such as the maximum number of foreign players or a minimum number of young players, and squad structure requirements, such as ensuring a certain number of players for each position. The decisions represented concern the type of affiliation the club has with its players. The model formulation is presented below:

**Sets**

- $p \; in \; P \;= \; set \; of \; considered \; players$
- $t \; in \; T \;= \; set \; of \; considered \; seasons$
- $f \; in \; F \;= \; set \; of \; field \; positions \; performed \; by \; players$

**Parameters:**

- $q_{p,t}$ = quality of player $p$ in season t
- $m_{p,t}$ = market value of player p in season t
- $f_p$ = field position of player $p$
- $h_{p,t} = \begin{cases} 1, & if \; player \; p \; is \; considered \; youth \; in \; season \; t \\ 0, otherwise \end{cases}$
- $CAP_{hire}$ = annual budget to hire players
- $k_f$ = required number of players of field position $f$ in the squad
- $num_{youth}$ = min number of youth players required in the squad

**Decision Variables:**

- $x_{p,t} = \begin{cases} 1, & if \; player \; p \; is \; selected \; in \; the \; squad \; in \; season \; t \; (In) \\ 0, otherwise \end{cases}$

- $y_{p,t} = \begin{cases} 1, & if \; player \; p \; is \; not \; related \; with \; the \; club \; in \; season \; t \; (Out) \\ 0, otherwise \end{cases}$

- $hire - aux_{p,t} = \begin{cases} 1, & if \; player \; p \; is \; hired \; by \; the \; club \; in \; season \; t \; (Hired) \\ 0, otherwise \end{cases}$

**Constraints:**

- **Budget constraints:**
  - Budget to hire players

$$\sum_p hire - aux_{p,t} \cdot m_{p,t} \leq CAP_{hire} \quad \forall t \in T$$

- **Regulation constraints**
  - Min number of youth players required by regulation

$$\sum_{p} x_{p,t} \cdot h_p \geq n_{youth} \quad \forall t \in T$$

- **Squad structure constraints**
  - Number of required players per field position

$$\sum_{p} x_{p,t} \cdot f_p \leq k_f \quad \forall f \in F, t \in T$$

  - Max number of players in the squad

$$\sum_{p} x_{p,t} = \sum_{f} k_f \quad \forall t \in T$$

  - Max age to be hired

$$x_{p,t} \cdot age_{p,t} \leq \max_{age} \quad \forall p \in P, t \in T$$

  - Each season the overall score must increase

$$\sum_{p,t} x_{p,t} \cdot q_{p,t} \leq \sum_{p,t+1} x_{p,t+1} \cdot q_{p,t+1} \quad \forall t \in T$$

- **Variable relation**
  - Each player must be in one status in each season

$$x_{p,t} + y_{p,t} = 1 \quad \forall p \text{ in } P, t \text{ in } T$$

  - Hire dummy variable

$$y_{p,t} - y_{p,t+1} \leq hire - aux_{p,t+1} \quad \forall p \text{ in } P, t \text{ in } T$$

**Objective Function**

$$Max \sum_{p,t} x_{p,t} \cdot q_{p,t}$$

In summary, the proposed methodology provides a comprehensive framework for optimizing squad planning over multiple seasons. It integrates a variety of data inputs, including player characteristics, performance metrics, and financial attributes, to construct a detailed player profile. This profile is further enhanced with projected attributes using machine learning models, allowing for informed decision-making regarding player affiliations and contracts. The final Integer Programming model synthesizes these elements to generate optimal squad configurations, balancing quality, financial constraints, and regulatory requirements. By employing this multi-faceted approach, the methodology offers strategic guidance to the sporting director, facilitating effective management of both current and future squad composition.

## 4. Results

This section presents examples of the Performance Score calculations (Module 2). Also, it presents the comparative performance of various machine learning models employed for predictive analysis of Performance Score in Table 4 and Market Value in Table 5 (Module 3). The models were evaluated using RMSE and MAE. Following the comparison, a sensitivity analysis was conducted on the integer optimization problem (Module 5), assessing the robustness of the solution to changes in key parameters. The results offer insights into the models' predictive reliability and the optimization problem's sensitivity to parameter variation, highlighting trade-offs between overall score of the team and budget constraints in Table 6.

**Module 2 – Performance Score**

Image 3 shows the total score of two players, in addition, the OBV For & OBV Against score is shown. The graphs also show what the percentiles are for each event within the analysis for each player. In this case, the comparison is between two Winger. Another example could be found in Image 4 with the analysis of another two top players.
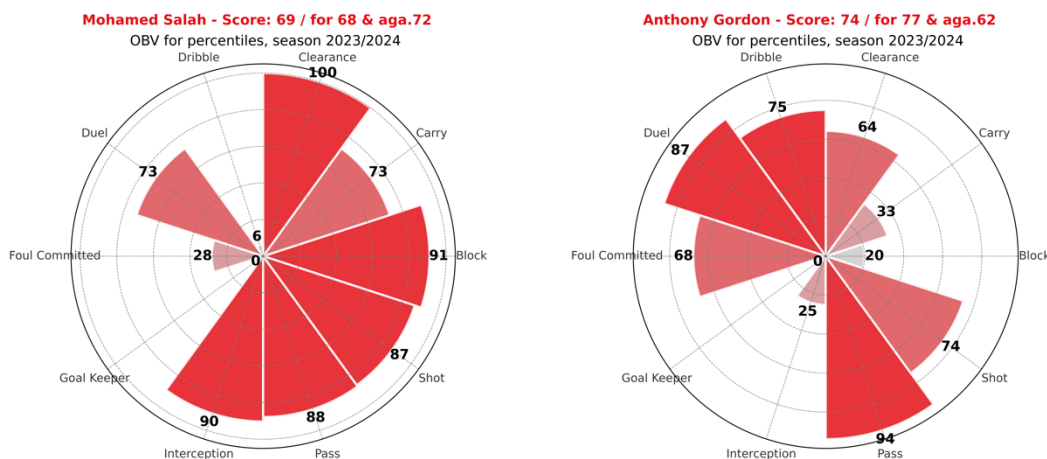


Image 3: Examples of Score. In the title of each plot is the Score and the OBV For & Against. The radar plot shows the OBV For Percentile of Mohammed Salah (Liverpool FC) and Anthony Gordon (Newcastle United FC) for each action.
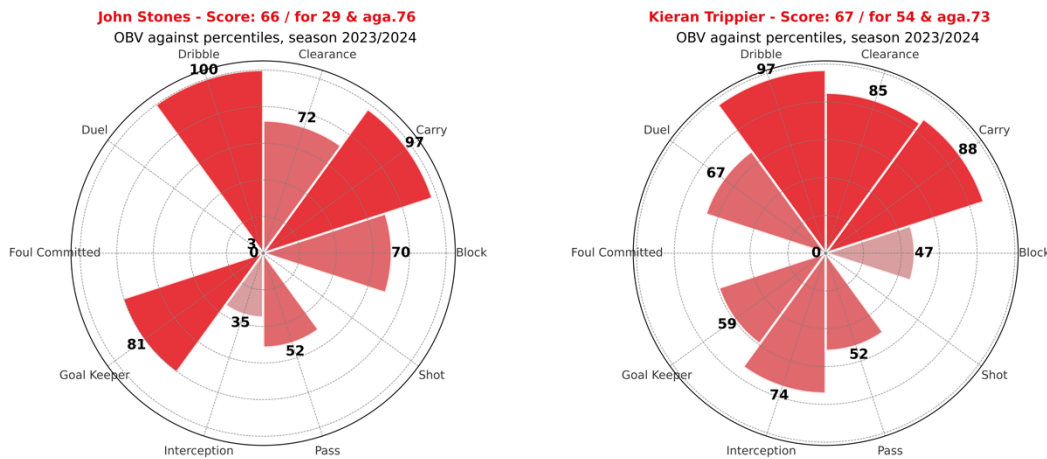
Image 4: Examples of Score. In the title of each plot is the Score and the OBV For & Against. The radar plot shows the OBV For Percentile of John Stones (Manchester City FC) and Kieran Trippier (Newcastle United FC) for each action.

**Module 3 - Machine Learning Prediction Models**

In this module, two models were calibrated to predict continuous numerical variables such as Performance Score and Market Value. Tables 4 and 5 show the prediction errors for each model. Each model was calibrated with 80% of the data and the predictions and their respective error calculations were performed with the remaining 20% of the data.

| Model | RMSE | MAE |
|---|---|---|
| Linear Regression | 13.7 | 11.0 |
| Ridge Regression | 13.6 | 10.9 |
| LASSO Regression | 13.2 | 10.6 |
| Decision Tree | 16.0 | 11.7 |
| Random Forest | 14.6 | 11.4 |
| xG Boost | 15.8 | 12.2 |
| Support Vector Regression | 14.8 | 11.3 |
| Artificial Neural Networks | 14.4 | 12.0 |

Table 4: Performance Score prediction errors.

LASSO Regression, with an **RMSE** of 13.2 and an **MAE** of 10.6, demonstrates the best performance in predicting the players' performance score. The relatively small gap between RMSE and MAE suggests that the model provides consistent and reliable predictions, with fewer extreme outliers. Other linear models

like **Ridge Regression** and **Linear Regression** also show good performance, with slightly higher errors than LASSO, whereas more complex models like **Decision Tree** and **xG Boost** perform worse. Given that the performance score is on a 0-100 scale, the errors indicate that predictions, on average, deviate by about 13 points, which represents a moderate level of accuracy. This suggests LASSO effectively balances model simplicity and predictive accuracy in this task.

| Model | RMSE | MAE |
|---|---|---|
| Linear Regression | 0.43 | 0.32 |
| Ridge Regression | 0.42 | 0.32 |
| LASSO Regression | 0.41 | 0.31 |
| Decision Tree | 0.54 | 0.35 |
| Random Forest | 0.47 | 0.33 |
| xG Boost | 0.50 | 0.34 |
| Support Vector Regression | 0.42 | 0.31 |
| Artificial Neural Networks | 0.46 | 0.34 |

Table 5: Market Value performance errors.

For the log-transformed market value, LASSO Regression also performs best, with an **RMSE** of 0.41 and an **MAE** of 0.31. The close alignment between these two-error metrics indicates a high-quality prediction with minimal extreme outliers. Similarly, other linear models like **Support Vector Regression** and **Ridge Regression** perform well, while more complex models such as **Random Forest** and **Artificial Neural Networks** produce higher errors. In the context of a log-log transformation, these errors translate into a relatively small deviation in market value predictions when back-transformed, making LASSO well-suited for this task. The errors suggest that LASSO captures the underlying relationships effectively while maintaining low variability in its predictions.

The SHAP analysis for Score Prediction (Model 1) and Market Value Prediction (Model 2) models illustrates the impact and direction of each feature's influence on the LASSO regression models' predictions, with SHAP values along the x-axis showing whether a feature drives the prediction positively or negatively. In Model 1, $age$, $age^2$, and $score_{t-1}$ emerge as the most influential features, with high values contributing positively to predictions. Model 2 similarly highlights $age^2$ and $\log(market\ value)_{t-1}$ as primary drivers, especially with high $age^2$ values positively impacting the model's output. In both models, positional features display minimal impact, clustering around zero. This suggests that age-related factors and season information are more significant in determining predictions, while positional attributes contribute less prominently. See Imagen 5 to more details.
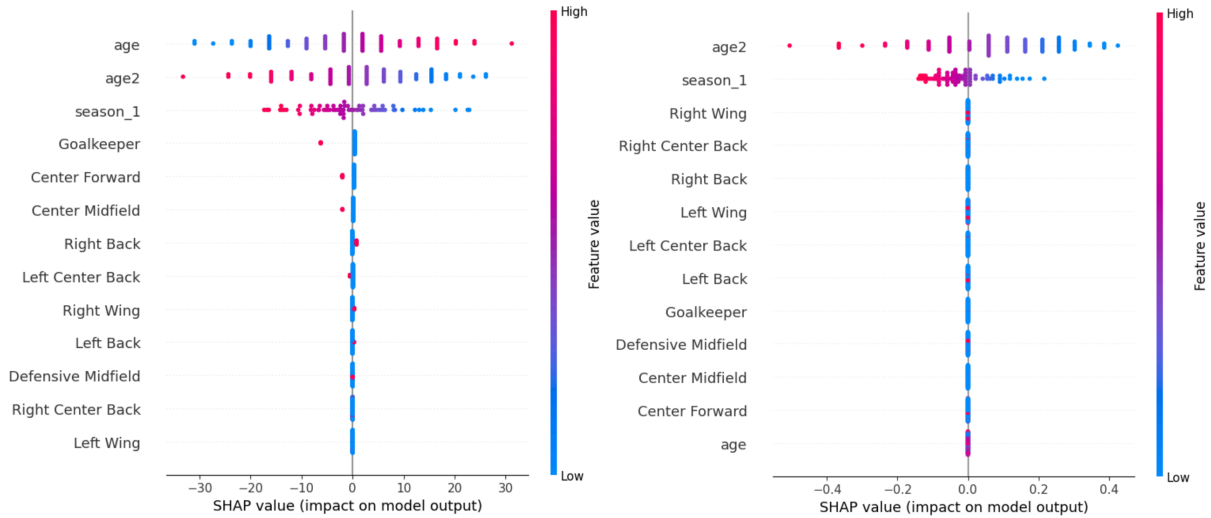
Image 5: Shape Values of Score Prediction Model and Market Value Prediction model.

## Module 5 - Multi-Season Squad Planning

We will analyze an instance of the integer programming model with parameters specified in Table 6. This will give us a first impression of how the model is performing. In this case we consider the planning of the Crystal Palace team with a hiring budget of 100 million euros. Other parameters were also used.

| Parameters | |
|---|---|
| **Team** | Crystal Palace |
| **Budget to Hire** | €100.000.000 |
| **Min. U21 Players** | 3 |
| **Players per position (excepts GK)** | 2 |
| **Goalkeepers** | 3 |
| **Max. Age** | 35 |
| **Min. Contract Length** | 2 |

Table 6: Parameters of first instance

**Crystal Palace 2023/2024 Squad**



**Center Forward**
Jean-Philippe Mateta (a.26) - 67
Odsonne Édouard (a.25) - 62

**Left Wing**
Jeffrey Schlupp (a.30) - 64

**Right Wing**
Jordan Ayew (a.31) - 31
Michael Olise (a.21) - 63

**Attacking Midfield**
Eberechi Eze (a.25) - 64

**Center Midfield**

**Defensive Midfield**
Jaïro Riedewald (a.26) - 47
Will Hughes (a.28) - 51
Naouirou Ahamada (a.21) - 56
Jefferson Andrés Lerma Solís (a.28) - 47
David Ozoh (a.18) - 39
Cheick Oumar Doucoure (a.23) - 45

**Left Back**
Tyrick Mitchell (a.23) - 51

**Right Back**
Joel Ward (a.33) - 55
Nathaniel Edwin Clyne (a.32) - 49

**Left Center Back**
Marc Guehi (a.22) - 35
Chris Richards (a.23) - 46

**Right Center Back**
Joachim Andersen (a.27) - 50
James Tomkins (a.34) - 24
Rob Holding (a.27) - 0

**Goalkeeper**
Dean Henderson (a.26) - 52
Sam Johnstone (a.30) - 16

Image 6: Crystal Palace's squad composition for the 2023/24 season, according to data from Statsbomb and TransferMarkt.

**Crystal Palace 2024/2025 Squad**



**Center Forward**
Jean-Philippe Mateta (a.27) - 54
Odsonne Édouard (a.26) - 52

**Left Wing**
Kaoru Mitoma (a.27) - 57
Jeffrey Schlupp (a.31) - 55

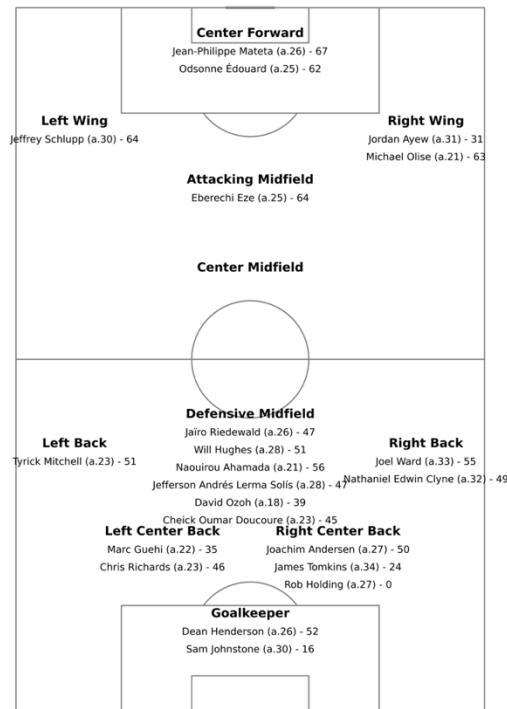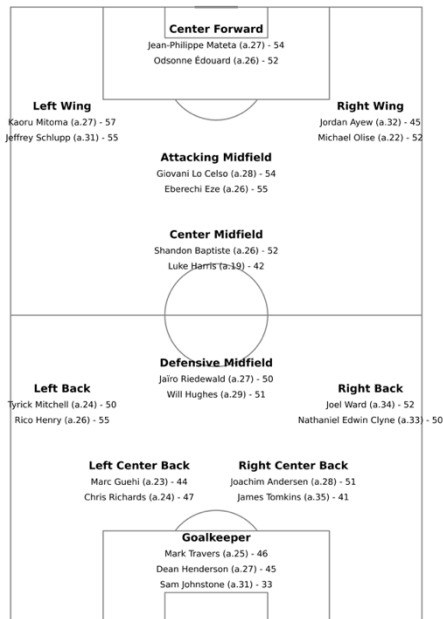**Right Wing**
Jordan Ayew (a.32) - 45
Michael Olise (a.22) - 52

**Attacking Midfield**
Giovani Lo Celso (a.28) - 54
Eberechi Eze (a.26) - 55

**Center Midfield**
Shandon Baptiste (a.26) - 52
Luke Harris (a.19) - 42

**Defensive Midfield**
Jaïro Riedewald (a.27) - 50
Will Hughes (a.29) - 51

**Left Back**
Tyrick Mitchell (a.24) - 50
Rico Henry (a.26) - 55

**Right Back**
Joel Ward (a.34) - 52
Nathaniel Edwin Clyne (a.33) - 50

**Left Center Back**
Marc Guehi (a.23) - 44
Chris Richards (a.24) - 47

**Right Center Back**
Joachim Andersen (a.28) - 51
James Tomkins (a.35) - 41

**Goalkeeper**
Mark Travers (a.25) - 46
Dean Henderson (a.27) - 45
Sam Johnstone (a.31) - 33

**Crystal Palace 2025/2026 Squad**



**Center Forward**
Jean-Philippe Mateta (a.28) - 50
Odsonne Édouard (a.27) - 49

**Left Wing**
Kaoru Mitoma (a.28) - 53
Jeffrey Schlupp (a.32) - 52

**Right Wing**
David Brooks (a.27) - 53
Michael Olise (a.23) - 50

**Attacking Midfield**
Giovani Lo Celso (a.29) - 52
Eberechi Eze (a.27) - 52

**Center Midfield**
Shandon Baptiste (a.27) - 49
Luke Harris (a.20) - 41

**Defensive Midfield**
Jaïro Riedewald (a.28) - 51
Will Hughes (a.30) - 51

**Left Back**
Tyrick Mitchell (a.25) - 50
Rico Henry (a.27) - 53

**Right Back**
Timothy Castagne (a.29) - 55
Nathaniel Edwin Clyne (a.34) - 50

**Left Center Back**
Marc Guehi (a.24) - 47
Chris Richards (a.25) - 49

**Right Center Back**
Mason Holgate (a.28) - 53
Joachim Andersen (a.29) - 51

**Goalkeeper**
Mark Travers (a.26) - 42
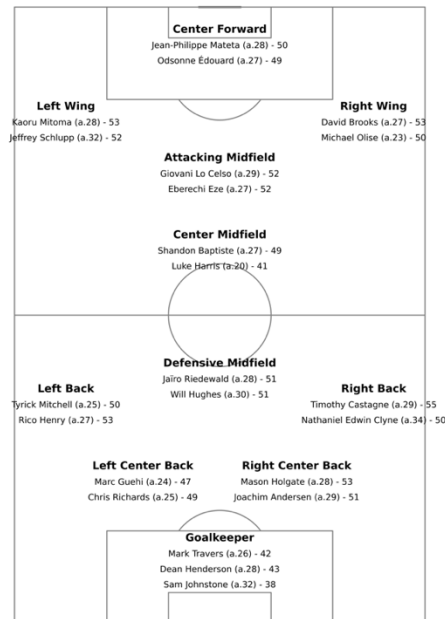Dean Henderson (a.28) - 43
Sam Johnstone (a.32) - 38

Image 7: The new composition of Crystal Palace's squads suggested by the full scheduling model for the 2024/2025 and 2025/2026 seasons is shown.

The results of the first instance show that the problem is solvable in just 5 seconds with an optimality gap of 0.1%. This indicates that the model can handle further complexity by incorporating additional constraints and expanding the pool of players in the study, which in practice allows for scouting across more leagues. Additionally, the generated squads provide better organization and balance across different positions, reflecting the effectiveness of the constraints applied (see Image 6 and 7).
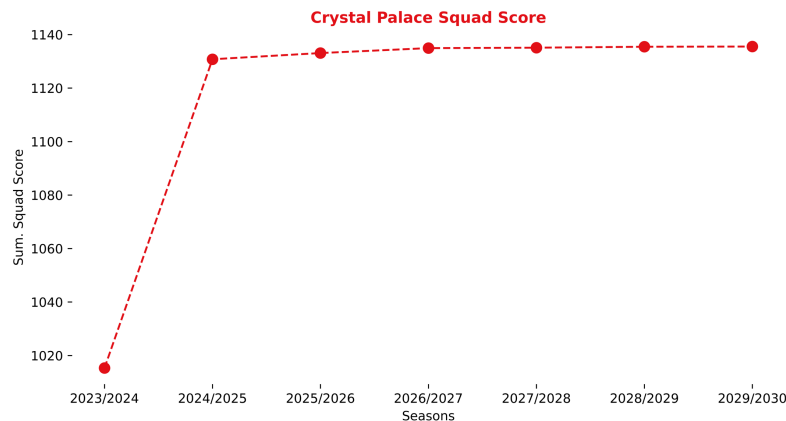


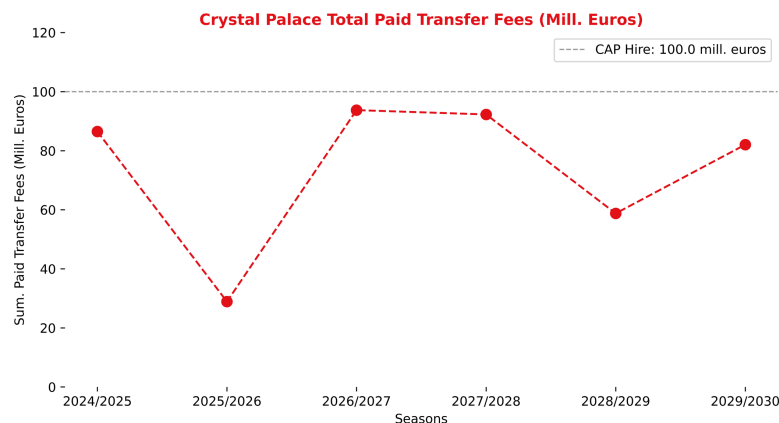Image 8: Crystal Palace's score over the next seasons



Image 9: Crystal Palace's transfer fees over the next seasons

Building on these results, we focus on the team's overall score and transfer spending within the study horizon. One key observation is that the constraint of 'improving the team's overall score season by season' is being met. The most significant increase in the overall score occurs in the first season, as the model is explicitly optimizing for this (the objective function), while balancing the goal of building a strong squad and adhering to budget constraints (Image 8). Regarding spending, most of it is concentrated in the first season (Image 9), where there is the greatest potential for improvement since the 2023/2024 squad was not assembled using this methodology. In the following seasons, the aim is to target specific reinforcements that will continue to improve the squad while meeting the legal and squad structure constraints.
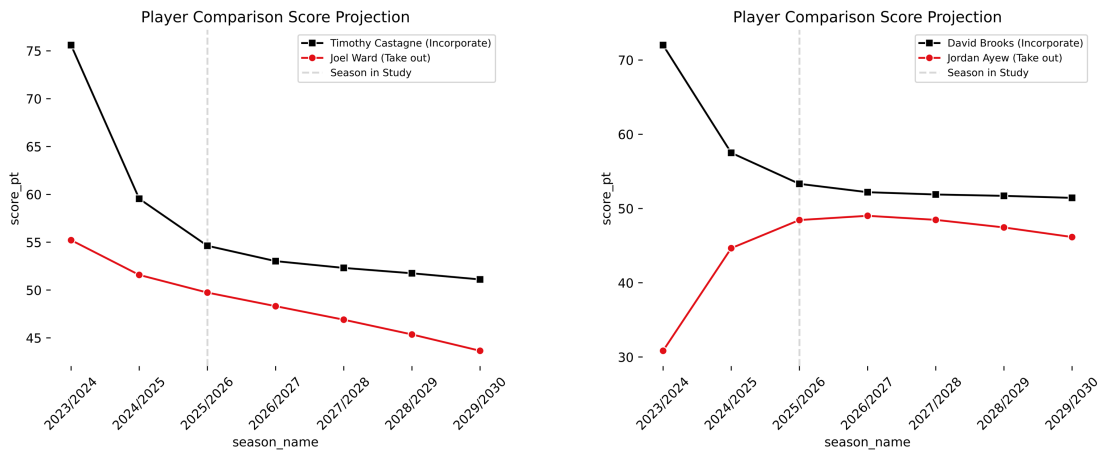
Image 10: Examples of player's replacement. Each plot shows the player who was removed from the team (red dotted line) and the player who was added (black squared line), as well as each player's score over the seasons. The season in which the change occurred is also shown (vertical gray line).

Now, we analyze two cases of player replacements. In both cases, the incoming player has a higher performance score than the outgoing player, both in the season of the replacement and in the following seasons. One question that arises is why this replacement wasn't made earlier. The hypothesis is that in the previous season (2024/2025 first of the study horizon), the model was addressing squad structure issues, and it wasn't until the 2025/2026 season that it began selecting players who excel in their performance

| Parameters | | | | |
|---|---|---|---|---|
| **Team** | Crystal Palace | Crystal Palace | Crystal Palace | Crystal Palace |
| **Budget to Hire** | €100.000.000 | €50.000.000 | €20.000.000 | €10.000.000 |
| **Min. U21 Players** | 3 | 3 | 3 | 3 |
| **Players per position (excepts GK)** | 2 | 2 | 2 | 2 |
| **Goalkeepers** | 3 | 3 | 3 | 3 |
| **Max. Age** | 35 | 35 | 35 | 35 |
| **Min. Contract Length** | 2 | 2 | 2 | 2 |
| **Total Overall Score** | **7820** | **7810** | **7792** | **7751** |

Table 6: Sensitivity analysis of the Multi Season Squad Planning (IP)

The sensitivity analysis (see Table 6), adjusting only the budget constraint, reveals a clear relationship between financial flexibility and the squad's overall performance score over time. With a budget of €100 million, the total overall score reaches 7820, the highest among all scenarios. As the budget decreases, we observe a gradual decline in performance, with scores of 7810, 7792, and 7751 for budgets of €50 million, €20 million, and €10 million, respectively. This indicates that while performance remains relatively stable with moderate budget cuts, significant reductions, particularly at the €10 million level, noticeably impact the team's performance over time.

As conclusion of results section, the analysis provides valuable insights into squad optimization, performance prediction, and budgetary constraints in multi-season planning. The machine learning models, particularly LASSO Regression, demonstrate strong predictive capability for both performance score and market value, offering consistent and reliable predictions. The integer optimization model proves effective in constructing balanced squads, as demonstrated in the case of Crystal Palace, achieving a rapid solution time and a minimal optimality gap. The improvements in overall team score, particularly in the first season, highlight the importance of addressing initial squad structure issues. However, the sensitivity analysis reveals that significant budget reductions can hinder long-term performance gains, with notable drops in overall performance score when financial flexibility is constrained. These findings underscore the need for a strategic balance between budget management and performance objectives in squad planning.

## 5. Conclusions

This study presents a comprehensive methodology for optimizing football squad planning, integrating sporting, financial, and strategic considerations. By adopting workforce planning concepts and adapting them to the specific dynamics of football, this research addresses key challenges faced by sports organizations in managing player contracts, performance evaluations, and financial constraints over multiple seasons.

The methodology developed includes five interconnected modules that work together to provide actionable insights for club managers. The model begins by compiling and analyzing player characteristics, followed by the estimation of player performance scores through advanced statistical methods such as the On-Ball Value (OBV). These scores are then projected into future seasons using machine learning models, alongside player market values. This information feeds into an Integer Programming model that optimizes squad composition while balancing quality, financial limitations, and regulatory requirements.

The novelty of this study lies in its ability to adapt workforce planning frameworks to football, where the concepts of skill heterogeneity, squad structure, and dynamic contractual decisions must be carefully considered. The integration of machine learning models to project both performance scores and market values allows for a data-driven approach to squad planning. Moreover, the sensitivity analysis conducted on budgetary constraints offers valuable insights into how financial limitations impact both short-term and long-term squad development.

From a practical perspective, this methodology provides sporting directors with a powerful tool for decision-making. It enables clubs to maximize squad quality, optimize financial resources, and adhere to regulations governing player acquisitions and roster compositions. In doing so, the model supports long-term competitiveness and financial sustainability, two critical objectives in professional football.

This research contributes to the growing body of literature on sports analytics by addressing several gaps. First, it provides an actionable model that combines player performance metrics with financial constraints, something that has been underexplored in football management studies. Second, by utilizing real-world data from StatsBomb and TransferMarkt, this study bridges the gap between theoretical models and their practical application in football club management. Lastly, it highlights the importance of a holistic approach to squad planning, where both current performance and future projections play essential roles in contract and transfer decisions.

In conclusion, this methodology offers a robust framework for football clubs to make more informed, strategic decisions regarding player contracts, squad composition, and resource allocation. By integrating data-driven insights with strategic goals, it facilitates more efficient squad planning and management, ensuring clubs remain competitive while maintaining financial discipline.

# References

[1] Anderson, M. (2004). The metrics of workforce planning. Public Personnel Management. 33. 364-379. 10.1177/009102600403300402.

[2] Bruecker, Philippe & Van den Bergh, Jorne & Beliën, Jeroen & Demeulemeester, Erik. (2015). Workforce planning incorporating skills: State of the art. European Journal of Operational Research. 243. 10.1016/j.ejor.2014.10.038.

[3] NAPA (1999) Building the workforce of the future to achieve organizational success.

[4] Tworek, Katarzyna & Zgrzywa-ziemak, Anna & Kamiński, Robert. (2020). Workforce diversity and organizational performance – a study of European football clubs. Argumenta Oeconomica. 2020. 189-211. 10.15611/aoe.2020.2.08.

[5] Boon, Bart & Sierksma, Gerard. (2003). Team formation: Matching quality supply and quality demand. European Journal of Operational Research. 148. 277-292. 10.1016/S0377-2217(02)00684-7.

[6] Ahmed, F., Jindal, A., Deb, K. (2011). Cricket Team Selection Using Evolutionary Multi-objective Optimization. In: Panigrahi, B.K., Suganthan, P.N., Das, S., Satapathy, S.C. (eds) Swarm, Evolutionary, and Memetic Computing. SEMCCO 2011. Lecture Notes in Computer Science, vol 7077. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-27242-4_9

[7] Rösch, D & Hodgson, Roy & Peterson, T & Graf-Baumann, Toni & Junge, Astrid & Chomiak, Jiri & Dvorak, Jiri. (2000). Assessment and Evaluation of Football Performance. The American journal of sports medicine. 28. S29-39. 10.1177/28.suppl_5.s-29.

[8] Pappalardo, L., Cintia, P., Rossi, A. *et al.* A public data set of spatio-temporal match events in soccer competitions. *Sci Data* **6**, 236 (2019). https://doi.org/10.1038/s41597-019-0247-7

[9] He, Miao & Cachucho, Ricardo & Knobbe, Arno. (2015). Football player's performance and market value.

[10] Pariath, Richard & Shah, Shailin & Surve, Aditya & Mittal, Jayashri. (2018). Player Performance Prediction in Football Game. 1148-1153. 10.1109/ICECA.2018.8474750.

[11] Onwuachu, Uzochukwu & Enyindah, Promise. (2022). A Machine Learning Application for Football Players' Selection.

[12] Sayan, V. H., & Hançer, E. (2022). A Survey on Football Player Performance and Value Estimation Using Machine Learning Techniques. Scientific Journal of Mehmet Akif Ersoy University, 5(2), 57-62.

# Appendix

## Appendix A: OBV Weights per Field Position ($w_{For}^{f_p}$ & $w_{Against}^{f_p}$)

| Event | OBV For | OBV Against |
|---|---|---|
| Goalkeeper (GK) | 0.1 | 0.9 |
| Left Center Back (LCB) | 0.2 | 0.8 |
| Right Center Back (RCB) | 0.2 | 0.8 |
| Left Back (LB) | 0.3 | 0.7 |
| Right Back (RB) | 0.3 | 0.7 |
| Defensive Midfield (DM) | 0.4 | 0.6 |
| Center Midfield (CM) | 0.5 | 0.5 |
| Offensive Midfield (OM) | 0.7 | 0.3 |
| Left Wing (LW) | 0.8 | 0.2 |
| Right Wing (RW) | 0.8 | 0.2 |
| Center Forward (CF) | 0.9 | 0.1 |

## Appendix B: OBV Weights per Field Position & Event ($v_{e,f_p}^{For}$ & $v_{e,f_p}^{Against}$)

### B.1 Goalkeeper

| Event | OBV For | OBV Against |
|---|---|---|
| Shot | 0.0 | 0.0 |
| Pass | 0.3 | 0.1 |
| Carry | 0.1 | 0.0 |
| Dribble | 0.0 | 0.0 |
| Foul Committed | 0.1 | 0.1 |
| Block | 0.0 | 0.1 |
| Clearance | 0.0 | 0.05 |
| Interception | 0.1 | 0.05 |
| Duel | 0.0 | 0.1 |
| Goalkeeper Save | 0.4 | 0.5 |

**B.2 Right & Left Center Back**

| Event | OBV For | OBV Against |
|---|---|---|
| Shot | 0.0 | 0.0 |
| Pass | 0.3 | 0.1 |
| Carry | 0.1 | 0.0 |
| Dribble | 0.0 | 0.0 |
| Foul Committed | 0.1 | 0.1 |
| Block | 0.0 | 0.1 |
| Clearance | 0.0 | 0.05 |
| Interception | 0.1 | 0.05 |
| Duel | 0.0 | 0.1 |
| Goalkeeper Save | 0.4 | 0.5 |

**B.3 Right & Left Back**

| Event | OBV For | OBV Against |
|---|---|---|
| Shot | 0.1 | 0.0 |
| Pass | 0.3 | 0.1 |
| Carry | 0.3 | 0.1 |
| Dribble | 0.1 | 0.0 |
| Foul Committed | 0.0 | 0.1 |
| Block | 0.0 | 0.2 |
| Clearance | 0.1 | 0.1 |
| Interception | 0.0 | 0.2 |
| Duel | 0.1 | 0.2 |
| Goalkeeper Save | 0.0 | 0.0 |

**B.4 Defensive Midfield**

| Event | OBV For | OBV Against |
|---|---|---|
| Shot | 0.0 | 0.0 |
| Pass | 0.4 | 0.2 |
| Carry | 0.1 | 0.0 |
| Dribble | 0.05 | 0.0 |
| Foul Committed | 0.05 | 0.1 |
| Block | 0.1 | 0.2 |
| Clearance | 0.1 | 0.1 |
| Interception | 0.1 | 0.2 |
| Duel | 0.1 | 0.2 |
| Goalkeeper Save | 0.0 | 0.0 |

**B.5 Center Midfield**

| Event | OBV For | OBV Against |
|---|---|---|
| Shot | 0.1 | 0.0 |
| Pass | 0.3 | 0.2 |
| Carry | 0.2 | 0.1 |
| Dribble | 0.05 | 0.1 |
| Foul Committed | 0.05 | 0.1 |
| Block | 0.0 | 0.1 |
| Clearance | 0.1 | 0.1 |
| Interception | 0.1 | 0.2 |
| Duel | 0.1 | 0.1 |
| Goalkeeper Save | 0.0 | 0.0 |

**B.6 Offensive Midfield**

| Event | OBV For | OBV Against |
|---|---|---|
| Shot | 0.2 | 0.0 |
| Pass | 0.3 | 0.2 |
| Carry | 0.2 | 0.1 |
| Dribble | 0.2 | 0.1 |
| Foul Committed | 0.0 | 0.1 |
| Block | 0.0 | 0.0 |
| Clearance | 0.0 | 0.1 |
| Interception | 0.05 | 0.2 |
| Duel | 0.05 | 0.2 |
| Goalkeeper Save | 0.0 | 0.0 |

**B.7 Right & Left Wing**

| Event | OBV For | OBV Against |
|---|---|---|
| Shot | 0.3 | 0.0 |
| Pass | 0.3 | 0.2 |
| Carry | 0.1 | 0.1 |
| Dribble | 0.2 | 0.0 |
| Foul Committed | 0.0 | 0.1 |
| Block | 0.0 | 0.0 |
| Clearance | 0.0 | 0.1 |
| Interception | 0.0 | 0.2 |
| Duel | 0.1 | 0.3 |
| Goalkeeper Save | 0.0 | 0.0 |

**B.8 Center Forward**

| Event | OBV For | OBV Against |
|-------|---------|-------------|
| Shot | 0.4 | 0.0 |
| Pass | 0.1 | 0.1 |
| Carry | 0.1 | 0.1 |
| Dribble | 0.2 | 0.1 |
| Foul Committed | 0.0 | 0.1 |
| Block | 0.0 | 0.0 |
| Clearance | 0.0 | 0.2 |
| Interception | 0.1 | 0.2 |
| Duel | 0.1 | 0.2 |
| Goalkeeper Save | 0.0 | 0.0 |