[redacted] Fantasy Premier League Optimization

November 23rd, 2023

Keani Schuller [redacted]

Mohamed Elenany [redacted]

Niki Mahmood Zadeh [redacted]

Jessica Quansah [redacted]

McGill University

Desautels School of Management

Master in Management Analytics

1 Introduction

The English Premier League is a globally renowned football event with a potential TV audience of 4.7 billion people[1]. Fantasy Football is a popular online game where users create imaginary football teams from certain leagues with drafted real-life players and compete against other users' teams each week. Last year, the Fantasy Premier League was played by over 11 million users[2]. The Fantasy Premier League is the main pillar of marketing for the English Premier League, attracting notable sponsors to the game. In the 2021-2022 season, the Premier League generated 5.5 Billion Euros in revenue, which is more than the revenue generated by the Spanish and German leagues combined[3]. This success was credited to their fantasy football game.

This project aims to create an optimizer using Gurobi in Python that will reliably deliver the best possible Premier League Fantasy Football team composition. The Fantasy Premier League presents a unique challenge due to its dynamic nature. This is because the information needed to make decisions, such as a player getting injured, is constantly changing. This creates an opportunity for applying optimization techniques to constantly evaluate the best decision. What makes this particularly interesting is trying to simplify the complexity of real-world sports into one optimization problem. The unpredictability of player performance, coupled with budget constraints and the need to adapt strategies week-by-week, offers a rich and complex problem. The project aims to break this down into actionable insights, enabling users to make informed decisions and maximize their chances of success in Fantasy Football's competitive environment.

2 Problem Description and Formulation

The optimization problem is formulated as a linear mixed-integer optimization, considering various factors to maximize the total points accumulated by the team at the end of the season. The data used to achieve this is Premier League API data about the previous (2021/22) and next season's (2022-2023) individual player performances and fixtures to make a simulation for the next season (2022-2023). To accomplish this objective, strategic decisions, including selecting players who consistently score goals, assist others, maintain clean sheets, and earn bonus points, must be made, all while staying within budget constraints.

2.1 Decision Variables

The first set of decision variables for the problem consists of each individual player's selection, formulated as binary variables where 1 is if a player is selected and 0 is if the player is not. Next, captain selection and vice-captain selection are formulated as binary variables as well. To capture the initial starting team of 11 players, binary decision variables were used where 1 is if a player was in the starting 11 and 0 if not. There were also binary decision variables for all non-goalkeeper players to express whether or not they are a substitute. Lastly, one binary decision variable was used for goalkeeper substitutes. As such, the problem has 2891 decision variables overall. They can therefore be expressed as:

$$Players_i \in \{0,1\}$$

$$Captain_i \in \{0,1\}$$

$$ViceCaptain_i \in \{0,1\}$$

$$Starting11_i \in \{0,1\}$$

$$Substitutes_i \in \{0,1\}$$

where i is the player's index for i = 1 to 778

2.2 Objective Function

The objective function aims to create the best team composition to maximize total points to be potentially earned in the upcoming gameweeks. It has two parts: a utility function that scores player performance based on different player statistics and an expected number of points prediction for the next gameweek suggested by the Premier League based on the specific fixture for each player. Each player is given a score with 90% weight on their utility, which evaluates their overall performance till now, and 10% weight on their expected points, which evaluates their expected points for the upcoming gameweek. The expected points are very useful as they capture metrics like opponent difficulty, player form, and team form for the upcoming match, which helps us refrain from selecting players who have low form or hard upcoming matches. However, expected points have a lower weight in the objective function because putting a bigger emphasis on players' expected points for the upcoming gameweek can lead to short-sightedness as it would maximize points for the short term (one gameweek) instead of maximizing points for the next couple of gameweeks (long term).

Within the objective function, weights are placed on players to ensure that the person selected as captain has the highest score of utility and expected points and the vice-captain is the second-best person. The smallest is placed on the substitutes to focus spending on the starting 11, and also as substitutes have a low probability of being substituted in. Using i as the player index, the above aforementioned objective function can be represented as follows:

$$
\begin{aligned}
Utility = &\sum_{i=1}^{n}(Utility_i) * (Starting11_i + 0.4 * Substitutes_i \\
&+ 2 * Captain_i + 1.5 * ViceCaptain_i) \\
ExpectedPoints(Xp) = &\sum_{i=1}^{n}(Xp_i) * (Starting11_i + 0.4 * Substitutes_i + \\
&1.5 * Captain_i + 1.4 * ViceCaptain_i)
\end{aligned}
$$

The Objective function can be expressed then as:

$$\max Utility * 0.9 + Xp * 0.1$$

2.3 Constraints

The constraints for the problem include game-specific rules such as a $100 million budget limitation, constraints on player allocation in each position, and a maximum number of players from each team. As such, the constraints for the problem can be formulated as follows:

1. The squad size must consist of 15 players:
$$\sum_{i=1}^{n} Players_i = 15$$
where n is the total number of players in the selected season

2. There must be 2 Goalkeepers, 5 Defenders, 5 Midfielders, and 3 Forwards on the team.

$$\sum_{i=\text{Goalkeeper Indices}} Players_i = 2$$

$$\sum_{i=\text{Defenders Indices}} Players_i = 5$$

$$\sum_{i=\text{Midfielders Indices}} Players_i = 5$$

$$\sum_{i=\text{Forward Indices}} Players_i = 3$$

3. The total value of your initial squad cannot exceed \$100 million:

$$\sum_{i=1}^{n} Cost_i * Players_i = 100$$

where $Cost_i$ = Cost of each player i in millions of dollars

4. Only 3 players from a single Premier League Team can be selected:

$$\sum_{i=\text{indices of players in the same team}} Players_i \leq 3$$

5. Only 11 players selected as the starting 11, and only players in our 15 man squad can be in the starting 11:

$$\sum_{i=1}^{n} Starting11_i = 11$$

$$\sum_{i=1}^{n} Starting11_i \leq Players_i$$

6. The active 11-player team can have between 3 and 5 Defenders, 2 and 5 Midfielders,

1 and 3 Forwards, and only 1 Goalkeeper:

$$\sum_{i=\text{Defenders Indices}} Starting11_i \geq 3$$

$$\sum_{i=\text{Defenders Indices}} Starting11_i \leq 5$$

$$\sum_{i=\text{Midfielders Indices}}^{n} Starting11_i \geq 2$$

$$\sum_{i=\text{Midfielders Indices}}^{n} Starting11_i \leq 5$$

$$\sum_{i=\text{Forward Indices}}^{n} Starting11_i \leq 3$$

$$\sum_{i=\text{Forward Indices}}^{n} Starting11_i \geq 1$$

$$\sum_{i=\text{Goalkeeper Indices}}^{n} Starting11_i = 1$$

7. One team captain has to be selected from this 11 member squad:

$$\sum_{i=1}^{n} Captain_i = 1$$

$$\sum_{i=1}^{n} Captain_i \leq Players_i$$

$$\sum_{i=1}^{n} Captain_i \geq Starting11_i$$

8. One vice-captain must be selected from the starting 11:

$$\sum_{i=1}^{n} ViceCaptain_i = 1$$

$$\sum_{i=1}^{n} ViceCaptain_i \leq Players_i$$

$$\sum_{i=1}^{n} ViceCaptain_i \geq Starting11_i$$

9. Vice captain cannot be the same player as captain:

$$ViceCaptain_i + Captain_i \leq 1$$

10. Four substitutes must be selected from the 15 player-team:

$$\sum_{i=1}^{n} Substitutes_i = 4$$

$$\sum_{i=1}^{n} Substitutes_i \leq Players_i$$

11. A player is either a substitute or in the starting 11:

$$Substitutes_i + Starting11_i \leq 1$$

3 Numerical Implication and Results

To create an optimized Fantasy Premier League (FPL) team, the problem incorporated various constraints, including player performance, budget limits, fixture schedules, and player availability. After formulating these constraints mathematically, preprocessing and exploratory data analysis was conducted on the data collected from the FPL API to get a better understanding of the data and how to best structure it to be able to generate optimization code using Gurobi in Python.

3.1 Exploratory Data Analysis Insights

During the Exploratory Data Analysis (EDA) phase, the FPL data was analyzed to garner insights crucial for optimizing team selection. This analysis included a thorough examination of player statistics, team strengths, and matchday performances.

The average points-per-match ratio was critical in identifying high performers and potential differentiators in the team. For instance, players like Kevin De Bruyne and Mohamed Salah exhibited exceptionally high points-per-match ratios, making them primary candidates for selection. Players with less fluctuation in their performance across gameweeks were deemed more reliable and thus preferred for selection. The correlation between team strength and individual player performance was also explored. For example, players from teams with high defensive strength metrics were more likely to secure clean sheets, making them valuable picks for defenders (Appendix 6.5). Analysis of player performance across different matchdays revealed patterns and trends useful for predicting future performances. For instance, some players showed a tendency to perform exceptionally well against specific teams or in certain match conditions (Appendix 6.4-6.7). Players with high points-per-cost ratios were identified as potential underdogs, offering great value for their price. These players often went unnoticed in the general player pool but had the potential to provide high returns on investment. By comparing player performances against their costs, several undervalued players were identified who could contribute significantly to the team without exhausting the budget (Appendix 6.4.2). The impact of external factors such as player injuries, team formations, and player rotations was also considered. This holistic approach ensured that team selection was not just based on past performance but also accounted for potential future scenarios.

3.2 Lasso Regression Implementation

LASSO feature selection was used to weight and highlight crucial variables of player performance that correlate with the player's total points scored. This regression was applied distinctly for different positions; Goalkeepers (GK), Defenders (DEF), Midfielders (MID), and Forwards (FWD). Attributes like team strength, cost, minutes played, and goals scored were considered. The LASSO regression provided a set of weights to be placed on each variable, essential for formulating the utility function in our optimization model. For instance, the LASSO coefficients for goalkeepers highlighted "clean sheets" and "penalties saved" as significant factors, while for forwards, "goals scored" was a key attribute. To generate the utility function for each player, each player performance statistic was multiplied by its respective weight as suggested by LASSO, and the sum of those products was used as a utility score for each player.

3.3 Optimization Model Formulation

Using Gurobi, a binary model was made with variables representing the selection of players, captains, vice-captains, and starting eleven using the same structure as that in the mathematical formulation. The model's objective function aimed to maximize the combination of utility (informed by LASSO regression outcomes) and expected points for the upcoming gameweek (Xp), while placing the bigger emphasis (90%) on the player utilities.

3.4 Constraints

The model incorporated the constraints that mirror real-world scenario of Fantasy Premier League, and each of these played an important role in setting the problem boundaries:

- Squad Composition: Constraints on the number of players for each position, the number of allowed players selected from the same team, and starting l1 and substitutes selection.

- Budget Limitation: The total cost of selected players was restricted to £100 million, mirroring the game's financial constraints. Throughout the season, player values can increase/decrease based on player performance, and thus total budget changes. This is reflected through the iterations of the model throughout the season.

- Captain and Vice-Captain Selection: The rules surrounding the appointment of captain and vice-captain added strategic depth to the model. A Captain gets double points, while a Vice captain gets double points only if the captain did not play.

- Transfers per Gameweek: 1 free transfer per gameweek post initial team selection Wildcard Usage: Two unlimited transfers wildcards allowed, one in the first half of the season (gameweek 1-17) and the other in the second half (gameweek 18-38).

- Substitute Points: Substitute points only count when a player in the starting 11 does not play his match, thus getting substituted out. The substitutes need to be selected in order from a substitute ranking list; If a player in the starting squad does not play, they get substituted with the substitute player ranked number 1 by the user.

3.5 Results Intepretation

The model played the Fantasy Premier League game for the 2022-2023 season. Two different models were used within the game. The first model was used to formulate an initial team that used player performance data from the 2021-2022 season for the utility function. It placed a higher weight (20%) on the expected points for gameweek 1 to increase the chance of having a player that just transferred in this season into our selected team.

The second model, the one used to formulate all other midseason teams, used current season data for the utility. The initial team continued playing until gameweek 8, after which the unlimited transfers wild card was used to reallocate the team for gameweek 9 using player performance in the current season from gameweek 1 to 8. Using the same idea, the same team kept playing until gameweek 17. Then the 18 free transfers were used that were accumulating week by week to make a new team reallocation for free for gameweek 18. The game continued running the same way by making team reallocations once in gameweek 26 using the second free transfers wild card and a final team reallocation in gameweek 34 using the accumulated transfer points. To summarize, this took the approach of using the same team every week while making quarterly team reallocations based on player performances up until each reallocation gameweek.

In Gameweek 1, the team kicked off the season with a strong start, scoring 72 points, highlighting the model's ability to predict high-performing players accurately from the outset. Throughout the season, the team demonstrated stable and consistent performance, with points typically ranging between 50 and 70 in most gameweeks, indicating the model's effectiveness in capturing player potential. Additionally, this consistency underscores the model's adaptability and capability in maintaining performance levels despite the dynamic and unpredictable nature of the league.

The model's adaptability in response to player form, injuries, and other real-time factors ensured that the team's performance did not diminish as the season progressed as the usage of expected points allowed us to capture indirectly some of the aforementioned aspects. When there are extenuating circumstances where players have double matches in the same gameweek, the model was able to strategically select the players that have those double matches. This allowed the team to benefit from those players playing an extra game, which in turn allowed the team to score over 100 points in the gameweek that had those extra games being played.

The model's approach to selecting the captain and vice-captain was crucial in maximizing points. The algorithm's choices for these roles were based on in-depth analysis of player statistics and expected performance, contributing significantly to the overall points tally. Throughout each gameweek allocation, the model oscillated between Mohamed Salah and

Erling Halaand as team captain. This supported model accuracy in determining potential performers as they are both the top 2 highest point scoring players of the 2022-2023 season.

The model's team accumulated a total of 2,487 points over the season. This remarkable score reflects the model's proficiency and effectiveness in strategically selecting players and managing the team based on player utilities and expected points, ensuring a high-performance team within the budget limit. To put this into context, the highest overall score in the previous year's FPL globally was 2,776 points. The model's total points accumulation being close to this benchmark is a testament to its robustness and accuracy in predicting player performance.

The results from this model provide valuable insights into the effectiveness of integrating statistical models with optimization techniques in sports analytics. The model's ability to closely emulate the top global scores in FPL highlights its potential as a powerful tool for FPL enthusiasts and sports analysts, offering a data-driven approach to making informed decisions in fantasy sports management.

4 Problem Extension

The base model was expanded to capture some of the complexities of the game like including the free transfers wildcard, however, there are additional features within the game that add a layer of complexity and could have been further incorporated into our model. These include, but are not limited to, other special "chips" that participants can utilize throughout the game. For instance, the "Free Hit Chip" grants the ability to make unlimited transfers within one gameweek, with the caveat that the team reverts to its original state afterward; while the "Triple Captain Chip" enables managers to triple the points earned by a designated player in their team. These chips add another layer of complexity to the game, as their strategic deployment can significantly impact overall performance. Adding the option to use those chips would strongly enhance the model and allow for higher point scoring.

In this simulation, it was assumed that the team stayed the same throughout every set of gameweeks. However, in reality, players almost always change the captain and vicecaptain allocations every week based on the fixture each player will play. Additionally, weekly transfers are utilized most of the time, specifically when player injuries happen. In this simulation, the model is stuck with that injured player until the next team reallocation. Since the model does not harness these abilities, it loses a significant portion of its power, which is reflected in losing some points between gameweeks.

Additionally, there is an additional constraint that each transfer made over the amount of free transfers a user has per week would cost a 4 point deduction from the gameweek's points. In the model, it does not implement this constraint as all transfers made are free, either by wild cards or by accumulated points. If it were to implement week by week changes, then this extra constraint must be taken into consideration.

A current weakness in the model is the initial team selection. Currently, the model selects players for the current season using data from the previous season while using a slightly larger weight on expected points to detect any promising players. However, if a new player just transferred to the Premier League, that player would have no chance of being selected for the initial squad as he would be assigned a utility score of 0 due to not having any previous data. This is a huge drawback because the Premier League is the most attractive league in the world and draws most of the exciting and top players towards it each year.

For example, Erling Halaand, the player of the season in 2022-2023, joined the Premier League that season and earned the most points. He even scored 10 goals in his first 7 gameweeks, which is about 25% of his total goals in the season. However, because the model had no previous data about him, he was not selected in the initial team, which made it lose at least 60 extra points from gameweek 1 to 7. A suggestion to improve this is using a generative AI that could assign a utility score for new players based on previous performances in other leagues. Another could also be using the player's media attractiveness to guide the model to select potentially attractive players in the initial round. By detecting those promising prospects from the beginning, the model does not lose the opportunity cost of not selecting new transfers.

By incorporating the aforementioned factors into the model, this could develop an improved version of the current model to excel in the Fantasy Premier League. A thorough understanding of the constraints on transfers and strategic utilization of the available chips would provide the model with a competitive edge, allowing it to navigate the complexities of the Fantasy Premier League and increasing its chances of outperforming opponents.

5 Recommendations and Conclusions

This optimization model illustrates the power of statistical techniques and mathematical programming in sports analytics. The model provided a comprehensive strategy for FPL team management by systematically analyzing player data and applying optimization techniques. The numerical results underscore the model's potential in guiding FPL enthusiasts in making informed decisions, ultimately enhancing their performance in fantasy leagues. This approach, bridging data science and sports, opens new avenues for exploration in predictive analytics within the realm of fantasy sports.

The results obtained from the model demonstrate its significant efficacy. Notably, the model successfully achieves a delicate balance between previous player performance and upcoming fixtures, resulting in a remarkably competitive team selection. Remarkably, running the model throughout the 2022-2023 season, it accumulated a total of 2,487 points by the end of the season. This achievement is particularly impressive considering it falls only 289 points short of the highest-scoring team in the 2022-2023 season, who scored 2,776[4].

These results provide a solid foundation for future research and the continued optimization of player selection in the Fantasy Premier League; the model, however, is far from reaching its full potential. Incorporating strategies related to the optimal utilization of chips, making weekly transfers, and improving initial team selection would further refine our model and improve its ability to outperform competing teams. In conclusion, this project has successfully demonstrated the power of optimization in a global entertainment activity such as the Fantasy Premier League. It has the potential to effectively improve the decision-making of over 11-million participants. With the potential for further enhancements and the incorporation of additional strategies and perhaps personal preferences per participant, even greater success in optimizing player selection and transfers can be anticipated.

6 Appendix

6.1 Positions Data Table

The positions data table is an extract from the Fantasy Premier League (FPL) API, which details the available positions for players within the league. This data is integral for team formation constraints within the optimization model.

6.1.1 Data Structure

The table comprises four columns:

- position_ID: A unique identifier for each position.
- position: The shorthand notation for the player position (e.g., GKP for Goalkeeper).
- selected_per_team: The number of players that can be selected for each position per team.
- squad_min_play: The minimum number of players required to play for each position in a squad.
- squad_max_play: The maximum number of players allowed to play for each position in a squad.

6.1.2 Data Table Content

Below is the content of the positions data table detailing the constraints on player selection for an FPL team:

position_ID	position	selected_per_team	squad_min_play	squad_max_play
1	GKP	2	1	1
2	DEF	5	3	5
3	MID	5	2	5
4	FWD	3	1	3

6.2 Teams Data Table

The teams data table consolidates information about all the teams participating in the Premier League for the seasons 22/23 and 21/22. This data is crucial for modeling team strength and home-away advantage in the optimization model.

6.2.1 Data Structure

The table includes five columns:

- id: A unique identifier for each team.
- name: The official name of the Premier League team.
- strength: An overall strength rating of the team.
- strength_overall_home: The team's strength rating for home matches.
- strength_overall_away: The team's strength rating for away matches.

6.2.2 Data Table Content

The content of the teams data table for the season 22-23 is displayed below, showing the strength ratings that contribute to team selection and match outcome predictions:

id	name	strength	strength_overall_home	strength_overall_away
1	Arsenal	4	1245	1285
2	Aston Villa	3	1070	1100
...	...	...	...	...
20	Wolverhampton	3	1080	1090

6.3 Fixutres Data Table

The fixtures data table is an essential component for analyzing match outcomes and team performances across different gameweeks in the Premier League for seasons 22-23 and 21-22. This data underpins the optimization model, influencing player selection based on upcoming fixtures.

6.3.1 Data Structure

The table contains the following columns:

- game_id: A unique identifier for each fixture.
- date: The date and time when the match was played.
- Gameweek: The number of the gameweek in the Premier League season.
- Home Team: The name of the home team.
- Away Team: The name of the away team.
- Home Team ID: The unique identifier for the home team.
- Away Team ID: The unique identifier for the away team.
- Home Team Score: The number of goals scored by the home team.
- Away Team Score: The number of goals scored by the away team.

6.3.2 Data Table Content

Below is an excerpt from the fixtures data table for the season 22-23, illustrating the structure and type of data collected: (380 rows)

game_id	date	gameweek	home team	away team	home team ID	away team ID	home team score	away team score
1	..	1	Team A	Team B	7	1	2	2
..	...	...	...	...	...	...	...	...

6.4 Correlation Matrix for Goalkeeper Points

The correlation matrix for goalkeepers' points provides a quantitative assessment of the relationship between various factors that contribute to the total points accumulated by goalkeepers in the Fantasy Premier League (FPL).

6.4.1 Data Anlysis

A correlation analysis was performed to understand the strength and direction of the relationship between goalkeeper performance metrics and their total points. The following metrics were included in the analysis: Total points, Cost, Minutes played, Goals conceded, Creativity, Influence, ICT index (Influence, Creativity, Threat), Clean sheets, Saves, Penalties saved and Expected goals conceded

6.4.2 Insights from Correlation Matrix

The correlation matrix revealed several key insights:

- Influence and clean sheets showed a strong positive correlation with the total points, indicating that goalkeepers who actively contribute to preventing goals tend to score higher in FPL.
- Conversely, goals conceded had a strong negative correlation with total points, as expected.
- The cost of a goalkeeper showed a moderate positive correlation with total points, suggesting that higher-priced goalkeepers generally accumulate more points, likely due to playing for stronger teams.
- Metrics such as saves and penalties saved also showed positive correlations, highlighting their importance in a goalkeeper's point-scoring potential in FPL.

6.5 Correlation Matrix for Defender Points

The correlation matrix for defenders' points provides a quantitative exploration of the relationships between various gameplay statistics and the total points accumulated by defenders in the Fantasy Premier League (FPL).

6.5.1 Data Anlysis

Correlation analysis was performed to gauge the strength and directionality of associations between defenders' performance metrics and their total points. The metrics considered in this analysis included Total points, Assists, Clean sheets, Goals scored, Goals conceded, ICT index (Influence, Creativity, Threat), Threat, Creativity, Influence, Expected goals, Expected assists, Expected goal involvements and Expected goals conceded.

6.5.2 Insights from Correlation Matrix

The correlation matrix revealed several insights into defenders' performance:

- A strong positive correlation was observed between clean sheets and total points, emphasizing the importance of selecting defenders from teams with a strong defensive record.
- Goals scored also displayed a moderate positive correlation with total points, indicating that defenders who are also scoring threats can greatly enhance a team's point total.
- Negative correlations were observed between goals conceded and total points, underscoring the detrimental impact of a team's poor defensive performance on a defender's point potential in FPL.
- Influence and ICT index were positively correlated with total points, suggesting that defenders who are heavily involved in the gameplay tend to score higher.

6.6 Correlation Matrix for Midfielder Points

This appendix presents the correlation analysis of midfielder points within the Fantasy Premier League (FPL), highlighting the relationships between key performance metrics and the total points earned by midfielders.

6.6.1 Analytical Approach

A Pearson correlation matrix was constructed to evaluate the linear relationships between various performance indicators for midfielders and their impact on total FPL points. The analysis focused on the following metrics: Total points, Minutes played, Assists, Clean sheets, Goals scored, Goals conceded, ICT index (Influence, Creativity, Threat), Threat, Creativity, Influence, Expected goals, Expected assists, Expected goal involvements, Expected goals conceded

6.6.2 Key Findings

The correlation matrix provided the following insights for midfielders:

- A strong positive correlation was observed between total points and minutes played, highlighting the importance of selecting midfielders who are regular starters for their teams.

- Goals scored and assists also showed a positive correlation with total points, emphasizing the value of midfielders who contribute to offensive play.

- Clean sheets had a positive impact on total points, indicating that midfielders from teams with strong defensive records can provide additional points.

- ICT index, especially the creativity component, showed a strong positive correlation, suggesting that midfielders who frequently create chances are more likely to earn points.

- Negative correlations were noted for goals conceded, underscoring the negative impact of a midfielder's defensive duties on their point haul.

6.7 Correlation Matrix for Forward Points

This appendix provides a detailed correlation analysis for the forward players' performance metrics in the context of the Fantasy Premier League (FPL) data, elucidating the relationships between various in-game statistics and the total FPL points accrued by forwards.

6.7.1 Methodology

Pearson correlation coefficients were computed to quantify the linear relationships between several performance indicators and the total FPL points scored by forwards. The analysis encompassed metrics such as: Total points, Minutes played, Assists, Clean sheets, Goals scored, Goals conceded, ICT index (Influence, Creativity, Threat), Threat, Creativity, Influence, Expected goals, Expected assists, Expected goal involvements and Expected goals conceded

6.7.2 Observations

The correlation matrix offered several significant insights:

- A strong positive correlation between total points and both goals scored and assists, confirming the forwards' primary role in scoring and creating goals.
- An inverse relationship between goals conceded and total points, indicating that forwards from defensively stronger teams might not always contribute to high FPL points.
- The ICT index, especially the threat component, had a strong correlation with total points, suggesting the importance of forwards who are more likely to score.
- Clean sheets showed a moderate positive correlation, reflecting the added bonus for forwards participating in games without conceding goals.

6.8 Objective Function Decisions and LASSO Regression Analysis

6.8.1 Objective Function Formulation

The objective function of the Fantasy Premier League (FPL) optimization problem was crucial in determining the best team composition. The formulation process involved selecting relevant player performance metrics that significantly impact their point-scoring ability in the FPL context.

6.8.2 LASSO Regression

LASSO regression, which stands for Least Absolute Shrinkage and Selection Operator, was applied to identify the most predictive features for the total points scored by goalkeepers. This method helps in both variable selection and regularization to enhance predictive accuracy while preventing overfitting.

6.8.3 Methodology

The LASSO regression was performed on data split by player type with the following steps: Subsetting the dataset to include only player type-specific metrics(e.g. Goalkeepers would have assists, cleansheets, goals saved etc. in the model). Defining the LASSO model with a chosen alpha value that controls the strength of the penalty applied to the features. Fitting the LASSO model to the data and obtaining the coefficients for each predictor.

6.8.4 Coefficients of Determination:

The coefficients obtained from the LASSO model for each player type are as follows:

Predictor	**Goalkeepers**	**Defenders**	**Midfielders**	**Forwards**
Team Strength	0.473848	1.103897	-0.101506	1.189882
Cost	0.402883	0.530888	-0.553001	0.429152
minutes	0.022563	0.000000	0.023928	0.031191
Clean Sheets	3.830484	6.770899	0.881174	-0.729885
Goals Conceded	-0.649044	0.000848	-0.100615	-0.120736
Saves	0.252022	0.000000	0.000000	0.000000
Yellow Cards	-1.065981	0.000000	0.000000	0.000000
Red Cards	-1.279828	0.000000	0.000000	0.000000
Penalties Saved	6.501081	0.000000	0.000000	0.000000
ICT index	1.266702	0.915476	-0.000000	-4.287843
Threat	1.620136	0.096743	0.139610	-0.333383
Creativity	-2.906320	0.546564	0.351276	0.677184
Influence	1.146190	0.935122	-0.429995	0.800906
Expected Goals	0.000000	-0.000000	-0.000000	0.000000
Expected Assists	-0.000000	-0.000000	0.000000	-0.000000
Expected Goal Involvements	0.000000	-0.000000	-0.000000	-0.000000
Expected Goals Conceded	-1.825277	7.908730	-1.189891	3.039047
Assists	0.000000	2.630392	3.646533	3.193599
Goals Scored	0.000000	6.008930	5.948781	5.040497

6.8.5 Insights and Interpretation

The regression analysis demonstrated that clean sheets, saves, and penalties saved are highly influential in predicting the total points scored by goalkeepers, which is consistent with the scoring rules of FPL. Also, team strength is positively correlated with goalkeeper performance, suggesting that goalkeepers from stronger teams tend to score more FPL points. The negative coefficients for goals conceded, yellow cards, and red cards indicate that these negatively impact a goalkeeper's total FPL points. Features like expected goals, expected assists, and expected goal involvements had coefficients of zero, suggesting they do not influence the points for goalkeepers. For defenders, team strength, clean sheets, goals scored, expected goals conceded, assists, ICT index, influence and creativity were the top influences of total points scored, again matching the expectations as observed from the game. Similar results were observed for midfielders and forwards as our regression results concluded that clean sheets, threat, creativity and influence, goals scored and assists were the features that mattered most in the total points achieved. Interestingly, while team strength and ICT index mattered more for forwards, they had smaller weights for midfielders.

References

[1] Premier League. (2023, August 16). Opening weekend brings record audiences. Retrieved from https://www.premierleague.com/news/3636215

[2] Sim, J. (2023, June 12). How FPL has become 'one of the most important pillars' of the Premier League's digital strategy. Retrieved from https://www.sportspromedia.com/insights/fantasy-premier-league-alexandra-willis-value-global-marketing-promotion/?zephr_sso_ott=RCCaS8

[3] Salley, E. (2023, June 14). Premier League leads European football market with revenue double La Liga's in 2021-22. Retrieved from https://www.bbc.com/sport/football/65901519#:~:text=Premier%20League%20establishes%20itself%20as,euros%20(%C2%A35.5bn)

[4] Rigga, S. (2023, May). FLP 2022/23 winner: Who was the overall champion. Fantasy Football Scout. Retrieved from https://www.fantasyfootballscout.co.uk/2023/05/28/fpl-2022-23-winner-who-was-the-overall-champion/