

**BC2410 - Prescriptive Analytics: From Data to Decision**

**Project Report:**

**Building Teams on Fantasy Premier League Using Linear Optimization**

**S01 Group 2:**

Tan Jun Hong (U1920048B)

Goh Jia Rui, Bryna (U1822465A)

Evelynn Tan Yee Lin (U1910714H)

Yap Teng Suan (U2011243E)

# Executive Summary

The Fantasy Premier League (FPL) is a game where regular people take on the role of a fantasy manager of Premier League players and earn points based on the real-life performances of those players in the Premier League. The game has proven to be attractive, especially for Premier League fans, given the quality of prizes given out to winners. However, the overwhelming number of players and statistics to consider, on top of adhering to the rules of the game, often cause new managers unfamiliar with the game to end up with less than ideal line-ups that do not maximise points. Our main interest thus lies in how we can use optimization to help a new manager unfamiliar with the rules of the game come up with a line-up that satisfies the basic rules of the game and maximises points earned.

In our study, we built models based on two user personas, the Passive Manager and the Active Manager, to cater to managers with different management styles. The Passive Manager is defined as one who adopts a hands-off approach and is unlikely to make changes to their player line-up after the first selection. Conversely, the Active Manager adopts a hands-on approach and will be managing the line-up on a weekly basis throughout the Premier League season. In both cases, binary optimization was used to determine which players to select for the respective line-ups. In these optimization models, the objective function was to maximise points earned and constraints were based on the rules of the game, including budget and position constraints.

Sensitivity analysis was performed from the game makers’ point of view by varying the budget constraint for the Passive Manager model. The analysis revealed that total points tended to plateau at a budget upward of £120 million. This justifies the existing £100 million budget constraint in the game as this limited budget causes a manager to make trade-offs between points and a player’s cost when playing the game. Such an analysis also emphasises the importance of game makers ensuring that constraints in-game remain relevant to keep the game interesting for managers.

While our existing models are valid, there is room for further work. To address the volatility of the game, more can be done, especially for the Active Manager model, to take into consideration current developments in the Premier League. Additionally, our model can also be further tweaked to be made interactive to provide more customization for managers. For the game makers, sensitivity analysis can be done by adjusting other constraints in the game to explore how the game can be made more interesting. Concepts of game theory can also be further taken into consideration to improve the model.

All in all, the FPL is affected by many factors, including luck and unforeseen circumstances. While the models in this study cannot guarantee that managers win the FPL, they cater to different manager types and provide valid suggestions that would be useful. This is especially so for new managers seeking a good line-up that would give them a good head start in building a good FPL run.

# Table of Contents

[**Executive Summary**](#_v42tinlljefh) **2**

[**Table of Contents**](#_oobbet1s6q7e) **3**

[**1. Introduction: What is the Fantasy Premier League?**](#_3sssgh73cl1t) **4**

[**2. The Problem**](#_bu70ouh1jot8) **4**

[**3. User Personas**](#_6evp2cjh19g0) **4**

[3.1 Type 1: Passive Manager](#_6vo6heqmljix) 4

[3.1.1 Data Preparation for Passive Manager](#_44h5c6q64jc8) 5

[3.1.2 Model Formulation for Passive Manager](#_r8z5azi3ewwa) 6

[3.2 Type 2: Active Manager](#_rj24d3ixfok7) 8

[3.2.1 Data Preparation for Active Manager](#_zb3qf4emjrp0) 8

[3.2.2 Model Formulation for Active Manager](#_ou0ljmh90fu3) 9

[**4. Sensitivity Analysis**](#_w6csucet60fk) **11**

[**5. Areas for Further Development**](#_nnhxawbkeew4) **12**

[**6. Conclusion**](#_jekx40cjaaaz) **13**

[**Appendices**](#_xpt3dvnercke) **14**

[Appendix A: Prizes for Winning the FPL](#_ns5by6tp2akv) 14

[Appendix B: Basic Rules of the Game](#_hyvivi1i3prw) 15

[Appendix C: Dataset Links](#_uxi8hwdugci0) 17

[**References**](#_ikz877u2tvb4) **18**

# 

# 1. Introduction: What is the Fantasy Premier League?

The Fantasy Premier League (FPL) is a game where regular people take on the role of a fantasy manager of Premier League players. As a manager, one has to select 15 players from the Premier League to form a squad. Managers will then earn points based on the real-life performances of these players in the Premier League. The objective of the game is to build and manage a team that would give the manager the highest cumulative points at the end of the season.

The prizes (Appendix A) for winning the game (i.e. having the highest points) are extremely attractive, especially for a Premier League fan. On top of an air ticket to travel to the UK for a holiday, winners of the top overall prizes can even get to go through the VIP experience at Premier League Matches. Other prizes are also attractive, with winners entitled to technological gadgets, Nike products, FIFA games and FPL merchandise. These prove to be a huge draw for people to play the FPL.

# 2. The Problem

Before beginning the game, a manager must first understand the rules of the game. In the FPL, these rules fall under the categories of “Selecting Your Initial Squad”, “Managing your Squad”, “Transfers”, “Chips”, “Deadlines”, “Scoring” etc. While rules for a manager to select their initial squad are easy to understand and adhere to, rules in subsequent categories add to the complexity of the game. This, in addition to the overwhelming number of players and statistics to consider, often causes new managers unfamiliar with the game to end up with less-than-ideal line-ups that do not maximise points. Our main aim is thus to help a new manager unfamiliar with the rules of the game come up with a line-up that satisfies the basic rules of the game and maximises the points earned.

# 3. User Personas

To address the needs of new managers, we came up with two user personas for our project. Both users have different management styles, and our models will cater to each of their needs.

## 3.1 Type 1: Passive Manager

The Passive Manager adopts a hands-off approach and is unlikely to make any changes to their player line-up. As such, when building the model for Passive Managers, we consider players’ performance over the past few seasons and generate only one strong line-up, which is expected to be kept the same throughout the upcoming season.

### 3.1.1 Data Preparation for Passive Manager

Our main dataset comes from a [GitHub repository](https://github.com/vaastav/Fantasy-Premier-League) containing FPL statistics. The repository contains data from over 6 seasons, from the 2016/2017 season to the 2021/2022 season, that is currently being updated. As the 2021/2022 season is still in progress, we mainly used data from the earlier 5 seasons. We decided to use *players\_raw.csv* from each season as it was available for all seasons and had comprehensive statistics on each player’s performance. Additionally, it also had some of the columns that we were interested in, namely the player’s name (*first\_name* and *second\_name*), the player’s club (*team*), total points earned by the player that season (*total\_points*), and the cost of the player in the last game week of that season (*now\_cost*).

For each season’s *players\_raw.csv* from 2016/2017 to 2020/2021, we first extracted the relevant columns: *first\_name*, *second\_name*, *team*, and *total\_points*, then standardised NAs from all columns into *numpy.nan*, and combined *first\_name* and *second\_name* into a single column *full\_name*. To combine the historical data from all seasons into a single dataframe, we merged each season’s data, keeping the columns of the more recent season, while taking the average of the numerical columns (*total\_points*). In particular, we made the assumption that more recent statistics are more representative of a player’s current performance as compared to older statistics. Thus, we merged our data from the earliest season onwards: 2016/2017 data was first merged with 2017/2018 data, and the resulting dataset was then merged with 2018/2019, and so on. This resulted in a final dataset that weighted the most recent data from 2020/2021 as 50%, 2019/2020 as 25%, 2018/2019 as 12.5%, and so on.

However, we were still missing one column we require for our models: a player’s position. This data was not readily available in *players\_raw.csv*. As such, we decided to look for alternative sources of data to supplement our FPL data, and eventually settled on using EPL data and 3 different FIFA data from Kaggle (*See Appendix C: Dataset Links*). First, we filtered the FIFA data to only include Premier League teams. Then, for both EPL and FIFA data, we cleaned the player’s position so that it can only be one of the following values: DF, FW, GK, or MF representing defender, forward, goalkeeper, and midfielder respectively. If a player has more than one position, only the first position stated is considered. From these datasets, we extracted the *Name* and *Position* columns.

To merge our position data (from EPL and FIFA) with our merged dataset (from FPL), we had to merge on the Name column. As the EPL dataset generally had similar naming for their players as the FPL dataset, we initially did a merge on names using an exact match. However, many NA rows were still observed after this. To fill in these rows, we used our FIFA data. As the FIFA datasets recorded player names slightly differently, we performed a fuzzy match for rows with no position to still match successfully on the Name column. We first set an accuracy threshold of 90% for all 3 supplementary datasets (*players\_21.csv*, *FIFA21\_official\_data.csv* and *fifa21\_male2.csv*), but realised that there were too many null rows. We then ran again at 80%, and then 70%, to reduce the number of null rows. This was done to ensure that the more reliable matching was done first, before dropping the accuracy in attempts to match more names. We stopped at 70% because the matching became very inaccurate below that.

This merged dataset’s columns were renamed to *Total Points, Name, Position,* and *Club* and was then exported as a CSV file. We will refer to this as our overall historical dataset. When this dataset was later used in our models, we further performed one-hot encoding on the Position and Club columns, as they are categorical variables. The encoding resulted in 2D matrices representing if a player plays in a certain position or belongs to a certain team.

Lastly, to ensure that our model result can be used in the current season, we retrieved current player costs from the 2021/2022 *players\_raw.csv* dataset from GitHub. The *now\_cost* column was extracted, and the *first\_name* and *last\_name* columns were combined to a *full\_name* column to match our overall historical dataset. The dataset columns were renamed to *Cost* and *Name*, and exported as another CSV file. We will refer to this as our costs dataset.

### 3.1.2 Model Formulation for Passive Manager

To help our managers choose the best-performing player line-up, while still adhering to game rules, we selected the Binary Optimisation Model for our use case.

The following are our assumptions for the Passive Manager model:

* A player’s historical performance is a good indicator of their overall performance
* Player costs remain the same from week to week
* Each player can only hold one position throughout the season
* Each player remains in the same club throughout the entire season
* Our Passive Manager will not perform any player transfers throughout the entire season

Before moving on to the model formulation, we first identified the objective function, decision variable, and data required.

Objective: Maximise the points the squad can earn

Decision Variable:

Yi: Binary variable that indicates whether player i is selected for our squad

(Yi = 1 ⇒ Player selected, Yi = 0 ⇒ Player not selected), i = 1, 2, 3…, 382

Data:

Xi: Total average points scored by player i in previous seasons, i = 1, 2, 3…, 382

Pi: Cost of player i, i = 1, 2, 3…, 382

Mip: Binary data indicating whether player i holds the position p, where p = 1 (DF), 2 (FW), 3 (GK), 4 (MF), i = 1, 2, 3…, 382

Tik: Binary data indicating whether player i is in club k, k = 1, 2, …, 20

The model below selects 15 players into the squad:

|  | **Description** | **Formulation** |
| --- | --- | --- |
| **Objective Function** | Maximise the points earned by the squad |  |
| **Constraint 1** | Budget must not exceed £100 million |  |
| **Constraint 2** | Up to 15 players in the squad |  |
| **Constraint 3** | Position Requirements  5 Defenders, 3 Forwards,  2 Goalkeepers, 5 Midfielders, |  |
| **Constraint 4** | Select up to only 3 players from a single Premier League club |  |
| **Constraint 5** | Binary Variable |  |

Following the next step in the rules of the game, we will proceed to select the Captain and 11 players out of the 15 for our starting line-up. The 11 players selected for the starting line-up will contribute to a manager’s points for the Gameweek while the unselected players will serve as automatic substitutions in the event that any players in the starting line-up do not play in the Premier League for that Gameweek. Per the rules of the game, the player that gets selected as the Captain will have their points doubled in our model. Hence, another binary optimization model will be generated for this purpose.

The assumptions for the model for the selection of 11 players and the captain are the same as those made previously, with the addition of the following assumption:

* All players will play more than 0 minutes during the gameweek so that there will be no automatic substitutions. This ensures that all 15 players are eligible to be selected.

Objective: Maximise the points the starting line-up can earn

Decision Variable:

Yi: Binary variable that indicates whether player i is selected as part of the Starting 11

Zi: Binary variable that indicates whether player i is selected as the Captain

Data:

Xi: Total average points scored by player i in previous seasons, i = 1, 2, 3, …, 15

Gi, Di , Fi: Binary data indicating whether player i’s position is GK, DF, or FW, i = 1, 2, 3, …, 15

|  | **Description** | **Formulation** |
| --- | --- | --- |
| **Objective Function** | Maximise the points earned by Starting 11 |  |
| **Constraint 1** | Select 11 players |  |
| **Constraint 2** | Position Requirements  1 Goalkeepers, at least 3 Defenders, at least 1 Forward |  |
| **Constraint 3** | There can only be 1 Captain |  |
| **Constraint 4** | Binary Variables |  |

As seen in the table above, the objective function now includes Z so that the player that gets selected as a captain will have their points doubled in our model — the captain’s points will be counted once with Y, and again with Z.

## 3.2 Type 2: Active Manager

The Active Manager adopts a hands-on approach, and they are very likely to invest time to make the necessary changes to their player line-up to win the game. Therefore, when building the model for Active Managers, we consider players’ performance over the past and current seasons to generate different squads (on a weekly basis) that will maximise the managers’ chances of earning the most points and eventually winning the FPL grand prize. A key difference between the Active and Passive Manager is that the former has a more dynamic management style that adapts to the recent performance of the players, while the latter only considers players’ performance from past seasons.

### 3.2.1 Data Preparation for Active Manager

We decided to use *merged\_gw.csv* from each season as it provided data for all players in that season at the Gameweek level. For our Active Manager, we needed to prepare a weighted dataset, to assign different levels of importance to a player’s historical performance as compared to their more recent performance.

We first merged Gameweek data from all 4 seasons before 2020/2021, grouped by player names, to get each player’s historical data. 2020/2021 data was kept separate to explore how more recent Gameweek data could be weighed differently. The data revealed that some players did not play in some Gameweeks. To fix this, dummy Gameweek rows with 0 total points were appended for players that did not play in a particular Gameweek.

With the historical data and 2020/2021 data, we could start to prepare our weighted dataset. Assuming that more recent statistics are more representative of a player’s current performance compared to older statistics, we performed the following steps for each player’s data:

1. For all Gameweek data before 2020/2021, we took the mean of the player’s points as our starting point.
2. For a player’s first Gameweek statistic, a weighted average was calculated using 0.9 for the player’s historical mean (calculated in Step 1), and 0.1 for the player’s performance in the first Gameweek of 2020/2021.
3. For subsequent Gameweeks, a weighted average was calculated using 0.9 for the player’s performance up to that Gameweek (i.e. for Gameweek 2, the value obtained in Step 2 is weighted 0.9), and 0.1 for the player’s performance in the corresponding Gameweek of 2020/2021.

The resulting dataset has 38 Gameweek statistics for each player, where each Gameweek statistic is the weighted average of the player’s performance in previous Gameweeks. To experiment with our model, we also tried a weight distribution of 0.5 for historical data and 0.5 for recent data.

Both datasets were then exported as CSV files, with the columns *Name*, *GW*, *Position*, *Club*, *Total Points*, and *Cost*. When these datasets were later used in our models, we performed one-hot encoding to convert our categorical Position and Team columns into 2D matrices, similar to what was previously done for our Passive Manager model.

### 3.2.2 Model Formulation for Active Manager

Unlike the Passive Manager, the model formulation for the Active Manager generates new line-ups every week while considering the penalty of 4 points for every player transferred in after the first transfer (as the first transfer for each Gameweek is free). Other constraints for the budget, team size, team diversity and player position remain the same as in the case of the Passive Manager.

The following are our assumptions for the Active Manager model:

* A player’s historical performance is a good indicator of their overall performance
* Each player can only hold one position throughout the season
* Each player remains in the same club throughout the entire season

For this model, the assumptions of player costs remaining the same from week to week and that of the manager not performing any player transfers throughout the season, as introduced in our Passive Manager model, will not hold. Given that our Active Manager will be making changes to the squad on a weekly basis, changing player costs will have to be taken into consideration to make the model more useful. Naturally, the latter assumption will also not hold in the case of the Active Manager as our Active Manager is expected to perform the necessary player transfers week-to-week.

The following model formulated represents one Gameweek, and is run for all 38 Gameweeks of the season:

Objective: Maximise the points the squad can earn

Decision Variable:

Yi: Binary variable indicating whether player i was selected in current Gameweek, i = 1, 2, …, 712

Ci: Binary variable indicating whether player i was transferred into squad in current Gameweek, i = 1, 2, …, 712

Z: Derived decision variable representing the transfer penalties in current Gameweek, i = 1, 2, …, 712

Data:

Ai: Binary variable indicating whether player i was selected in previous Gameweek (based on result from previous Gameweek), i = 1, 2, …, 712

Xi: Total points scored by player i in current Gameweek, i = 1, 2, …, 712

Pi : Cost of player i in any week, i = 1, 2, …, 712

Mip: Binary data indicating whether player i holds the position p, where p = 1 (DF), 2 (FW), 3 (GK), 4 (MF), i = 1, 2, …, 712

Tik: Binary data indicating whether player i is in club k, k = 1, 2, …, 20

The model below selects 15 players into the squad, on a weekly basis:

|  | **Description** | **Formulation** |
| --- | --- | --- |
| **Objective Function** | Maximise the points earned by the squad |  |
| **Constraint 1** | Linearisation of Z |  |
| **Constraint 2** | Linearisation of Ci |  |
| **Constraint 3** | Up to 15 players in the squad |  |
| **Constraint 4** | Budget must not exceed £100 million |  |
| **Constraint 5** | Position Requirements  5 Defenders, 3 Forwards,  2 Goalkeepers, 5 Midfielders |  |
| **Constraint 6** | Select up to only 3 from each Premier League club |  |
| **Constraint 7** | Binary Variable |  |

As seen from *Section 3.2.1 Data Preparation for Active Manager*, the Xi data point is derived using a weighted average of the players’ points, at 90% from previous gameweeks and 10% from the most recent gameweek in the current season.

To evaluate our models, we ran our model-chosen line-ups with the latest season data, to calculate the points a manager would have earned by following our line-ups. As the current 2021/2022 season is currently only on Gameweek 32 (out of 38 weeks), we only use the provided data up till Gameweek 32:

| **Passive Manager**  **(Initial Squad)** | **Active Manager**  **(90-10 split)** | **Active Manager**  **(50-50 split)** |
| --- | --- | --- |
| 1381 | 1398 | 792 |

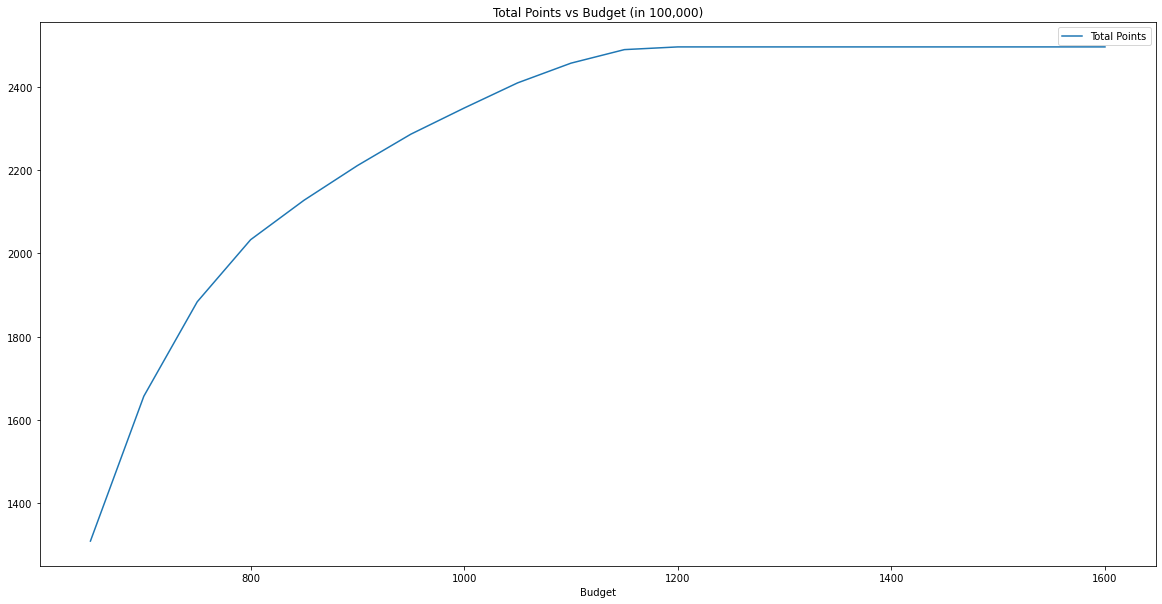
From these results, we can see that the Active Manager with 90-10 split performs slightly better than our Passive Manager. This can be because the Active Manager places slightly more importance on a player’s performance in the more recent Gameweeks, allowing the model to also consider players that have improved over the season break.

However, we note that our Active Manager with 50-50 split performs far worse, and this can likely be attributed to the volatility of a player’s performance — a player’s performance in one week is not dependent solely on their performance of the previous week. A 50-50 split places too much importance on their latest performance, causing our Active Manager to perform poorly.

# 4. Sensitivity Analysis

In both models, the budget could not exceed £100 million based on the rules set for the FPL. To study the game from the game makers’ point of view, sensitivity analysis was performed by varying this constraint to observe how the total points earned by our chosen line-up changed. To observe the changes, we ran our Passive Manager model with a range of different budget constraints. Specifically, the range of £65 million to £165 million was explored in increments of £5 million.

The graph below shows how total points change as the budget constraint varies:



From this graph, we can note several points. Firstly, total points plateau off as the budget constraint increases beyond £120 million, meaning any extra budget from this point will not increase the total points. This could be because the model is limited by another binding constraint, or that the best possible players have already been chosen. Secondly, the current budget constraint of £100 million comes rather close to the plateau but still does not achieve the maximum total points possible with no budget constraint. This shows that the limited budget causes trade-offs to be made between a manager’s points and cost when choosing players. Having to make the choice between this trade-off makes the game interesting for the managers — for example, with the budget constraint, the manager can choose between a team with a high point player with a high cost and a low point player with a low cost, or a team with two average point players with average costs. Such an observation reiterates the importance of game makers in ensuring that constraints in-game are kept relevant to maintain an engaging gaming experience. To illustrate — if player costs have been dropping, game makers should ensure that the budget constraint still remains relevant such that the players still have to decide between players.

# 5. Areas for Further Development

Although our models can provide suggestions that can aid a new manager to start playing the FPL, there are many improvements that we can explore to build on our existing models.

Firstly, we can consider current developments in the Premier League in our models. As the Premier League season spans 10-months every year, many things can happen. This can include injuries, red cards, the addition of new players, existing players changing teams or positions, or even getting suspended or banned from playing. To deal with these changes, we can extend our Active Manager model to also consider current developments when deciding which players to choose for each Gameweek. For example, if a player has been injured in the previous Gameweek, the model should be able to remove the player from consideration for our chosen team for this Gameweek.

Secondly, our models can also be made interactive to provide more customisation for our managers. Specifically, we can look into allowing managers to select a specific age range, specific teams, and specific chips they wish to use each week. With their customisation choices, new constraints can be added to our model, and more tailor-made line-ups can be suggested to our managers. This can be useful if our managers have specific preferences, such as being loyal to certain Premier League clubs.

Thirdly, we can also look into more ways to help the game makers. As explored previously in *Section 4: Sensitivity Analysis*, game makers can adjust the budget constraints to make the game more challenging and interesting for managers. Moving forward, we can also look into varying other constraints, such as adjusting the position requirements, and the number of players that can be selected from each club. By performing sensitivity analysis for these other constraints, we can explore ways to improve the existing game.

Lastly, we can also apply the concepts of game theory to our model. As we are competing against other players, there has to be a Nash equilibrium, where each player is assumed to know the equilibrium strategies of the other players, and no player has anything to gain by changing only their own strategy. We can incorporate this idea to further analyse the manager’s way of thinking and take into account what other managers will do, before deciding on our own.

# 6. Conclusion

To conclude, there are many things that can affect the game that we cannot foresee. A player’s performance can be affected by many different factors, such as their current mindset, their luck, or current game conditions.

Thus, our models cannot guarantee that a manager will win the FPL. However, our models can still cater to different types of managers, and provide suggestions to our managers on which players to pick, which can help a new manager get started in the FPL.

All in all, the applications of optimization modelling in the FPL are clear. While our models may not guarantee prizes for managers due to the volatility of the game, they can help provide an enhanced FPL experience for managers. These models also open up room for game makers in exploring how the game can be further developed to attract more players, which in turn generates a stronger following for the Premier League.

Looking beyond the Premier League, these models can also be applied to other fantasy sports games to enhance the playing experience and build a more immersive fan experience.

# Appendices

## Appendix A: Prizes for Winning the FPL

| Prizes | **Overall Prizes** | | | | **Monthly Prizes** | | **Weekly Prizes** | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Champion | Runner-Up | 3rd Place | Cup Winner | Manager of the Month | Top 10 | Manager of the Week | Top 20 |
| VIP hospitality at Premier League match | x2 | x1 |  | x1 |  |  |  |  |
| Travel & accommodation in the UK | 7 nights | 2 nights |  | 2 nights |  |  |  |  |
| Hublot connected watch |  |  |  |  |  |  |  |  |
| Copy of FIFA 22 |  |  |  |  |  |  |  |  |
| Games console |  |  |  |  |  |  |  |  |
| Tablet computer |  |  |  |  |  |  |  |  |
| Bluetooth speaker |  |  |  |  |  |  |  |  |
| Nike manager jacket |  |  |  |  |  |  |  |  |
| Nike Flight match ball |  |  |  |  |  |  |  |  |
| FPL goody bag |  |  |  |  |  |  |  |  |

Source: <https://fantasy.premierleague.com/prizes>

## Appendix B: Basic Rules of the Game

**1. Selecting Your Initial Squad**

**Squad Size**

To join the game select a fantasy football squad of 15 players, consisting of:

* 2 Goalkeepers
* 5 Defenders
* 5 Midfielders
* 3 Forwards

**Budget**

The total value of your initial squad must not exceed £100 million.

**Players Per Team**

You can select up to 3 players from a single Premier League team.

**2. Managing Your Squad**

**Choosing Your Starting 11**

* From your 15 player squad, select 11 players by the Gameweek deadline to form your team.
* All your points for the Gameweek will be scored by these 11 players, however if one or more doesn't play they may be automatically substituted.
* Your team can play in any formation providing that 1 goalkeeper, at least 3 defenders and at least 1 forward are selected at all times.

**Selecting a Captain and a Vice-Captain**

* From your starting 11 you nominate a captain and a vice-captain. Your captain's score will be doubled.
* If your captain plays 0 minutes in the Gameweek, the captain will be changed to the vice-captain.
* If both captain and vice-captain play 0 minutes in a Gameweek, then no player's score will be doubled.

**Prioritising Your Bench For Automatic Substitutions**

* Your substitutes provide cover for unforeseen events like injuries and postponements by automatically replacing starting players who don't play in a Gameweek.
* Playing in a Gameweek means playing at least 1 minute or receiving a yellow / red card.
* Based on the priorities you assign, automatic substitutions are processed at the end of the Gameweek as follows:
* If your Goalkeeper doesn't play in the Gameweek, he will be substituted by your replacement Goalkeeper, if he played in the Gameweek.
* If any of your outfield players don't play in the Gameweek, they will be substituted by the highest priority outfield substitute who played in the Gameweek and doesn't break the formation rules (eg. If your starting team has 3 defenders, a defender can only be replaced by another defender).

**3. Transfers**

After selecting your squad you can buy and sell players in the transfer market. Unlimited transfers can be made at no cost until your first deadline.

After your first deadline you will receive 1 free transfer each Gameweek. Each additional transfer you make in the same Gameweek will deduct 4 points from your total score (Classic scoring) and match score (Head-to-Head scoring) at the start of the next Gameweek.

If you do not use your free transfer, you are able to make an additional free transfer the following Gameweek. If you do not use this saved free transfer in the following Gameweek, it will be carried over until you do. You can never have more than 1 saved transfer.

You are limited to 20 transfers in any single Gameweek. This rule does not apply when playing a Wildcard or a Free Hit Chip.

**Player Prices**

Player prices change during the season depending on the popularity of the player in the transfer market. Player prices do not change until the season starts.

The price shown on your transfers page is a player's selling price. This selling price may be less than the player's current purchase price as a sell-on fee of 50% (rounded up to the nearest £0.1m) will be applied on any profits made on that player.

Source: <https://fantasy.premierleague.com/help/rules>

## Appendix C: Dataset Links

GitHub Repository with our FPL data

* <https://github.com/vaastav/Fantasy-Premier-League>

EPL Dataset:

* <https://www.kaggle.com/datasets/rajatrc1705/english-premier-league202021>

FIFA Datasets:

* <https://www.kaggle.com/datasets/stefanoleone992/fifa-21-complete-player-dataset?select=players_21.csv>
* <https://www.kaggle.com/datasets/bryanb/fifa-player-stats-database?select=FIFA21_official_data.csv>
* <https://www.kaggle.com/datasets/stefanoleone992/fifa-22-complete-player-dataset>

# 

# References

Premier League. (n.d.). *Fantasy Football Game Prizes*. Fantasy Premier League. Retrieved April 16, 2022, from https://fantasy.premierleague.com/prizes

Premier League. (n.d.). *Official Fantasy Football Game of the Premier League*. Fantasy Premier League. Retrieved April 16, 2022, from https://fantasy.premierleague.com/help/rules