

# Pandasy Premier League

Using Data Science to Maximise FPL Points

Matthew O'Donnell (18439746)  
Philip Byrne (19449216)

---

A report submitted in part fulfilment of COMP 30780

**Data Science in Practice**

**Demonstrator:** Eoghan Cunningham



UCD School of Computer Science  
University College Dublin

May 13, 2022

---

# Table of Contents

---

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Project Objectives</b>	<b>4</b>
2.1	FPL Game-Play	7
2.2	FPL on Social	12
2.3	FPL for Data Science	14
2.4	Objectives & Motivations	14
2.5	Research Questions	15
<b>3</b>	<b>Related Work</b>	<b>16</b>
<b>4</b>	<b>Data Considerations</b>	<b>17</b>
4.1	Data Collection	17
4.2	Data Preparation	19
<b>5</b>	<b>Results</b>	<b>21</b>
5.1	RQ1: How does formation play a part in overall points achieved?	21
5.2	RQ2: How to make best use of the budget?	25
5.3	RQ3: How transfers can affect a fantasy teams points?	31
5.4	RQ4: Can we predict which players will perform in a certain game week?	35
<b>6</b>	<b>Conclusions</b>	<b>38</b>
<b>7</b>	<b>Acknowledgements</b>	<b>40</b>

---

# Abstract

---

Fantasy Premier League (FPL) is a popular online game which casts you in the role of a fantasy manager of your chosen Premier League players. Each manager is allocated £100 million to build a squad of 15 virtual players, who are valued based on their real life ability and awarded points based on their weekly performances. Managers choose 11 starters each week in their preferred formation. They also have the option to make a transfer each week. The main aim of FPL is to achieve the most points possible over a season consisting of 38 game weeks. There has been a considerable interest in the rapid rise in FPL's popularity, the dynamics of its game-play, and the various strategies that managers can utilise to maximise points for the season. Mathematicians and Computer Scientists have examined various methods to build the ultimate FPL team and in this study we will review some of the recent work that has been done on the game. We describe the results of an in depth data analysis of the most fundamental aspects of FPL such as, the ideal formation to use, making best use of our limited budget, the effect of transfers on FPL points, and exploring predicting which players will perform in each game week. We have analysed FPL data of the full 2020/21 Premier League season in order to gain insight and make predictions for the ideal strategy to maximise points in succeeding seasons.

---

# Chapter 1: Introduction

---

Our Project is focused on using Data Science to build a winning Fantasy Premier League team. With over 9 million users this season and some fantastic prizes to be won, Fantasy Premier League is an incredibly competitive and entertaining online game where you own and manage a virtual team of Premier League super stars. The game allows you to create mini leagues, in which you can compete against your friends for prizes and bragging rights until the next season begins. The frustration and unpredictability of player performances can make it very hard for a Fantasy Manager to make the correct decisions each week in order to optimise performance. We were interested in finding a dataset that we could use to analyse player performances, looking for trends and predicted points for upcoming fixtures. We had found some articles looking at the use of Machine Learning to predict and build a winning Fantasy Football Team. The movie Moneyball was also an inspiration for this project. We will look more into related work in Chapter 3. The main aim of our project was to optimise the points of our virtual team for the season in order to come out as victors in our mini leagues and potentially overall.

What most managers don't know, is that the Official Fantasy Premier League provides an API giving access to player and game data each game week. This data is made available in a GitHub repository which is updated weekly. We planned to collect this data, download it and parse it into separate data frames which we could then analyse accordingly in order to answer our main research questions. These research questions will be discussed further in Chapter 2.

In Chapter 2 we will go into detail with why we chose our project, who will be interested by it, the importance of our project and what the main focus of the project will be. Chapter 3 involves the work related to our project, and aspects of these we took inspiration from. Following this we have Chapter 4 which discusses the datasets we chose to analyse and why we chose them. In Chapter 5 we will explain each of the main research questions describing the data used to answer the question, the methods used, and the results found. Chapter 6 will sum up the report, draw some conclusions about the work so far, and make some general observations about the work to come.

---

## Chapter 2: Project Objectives

---

Fantasy Premier League (FPL) is an online game which is simple to understand but very difficult to excel in. Each season, millions of people attempt to build the ultimate squad using their footballing knowledge. Fantasy sports actually date back to Oakland in 1963 when Bill Winkenbach and his friends created the first Fantasy Football draft (American Football not Soccer). This American Football fantasy became increasingly popular, but it wasn't until 1990 that the Soccer equivalent caught on. This was introduced in Italy in 1990 and the following year, the English followed suit, with the creation of what we now know as FPL. In 2002 the online version of FPL was created with 76,000 managers that first year [1]. This online game began growing year by year with over 2.5 million registered users after 10 years. Nearly 20 years after its inception, FPL has grown from 76,000 managers to over 9 million. These managers are spread across 190 countries of the world.

For instance, Figure 2.1 presents one version of the the level of interest in FPL, based on a website which keeps track of a range of Fantasy data [2] including the number of managers each season and the nationalities of these managers. It shows an increase of almost 5 million FPL accounts over a span of 8 years.

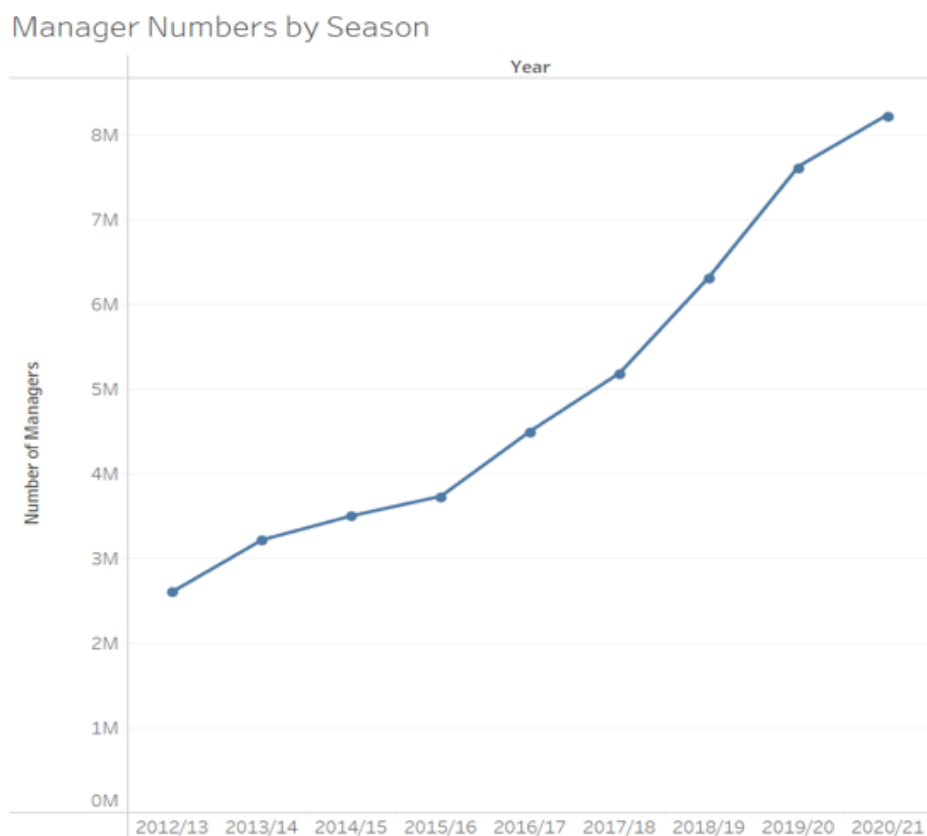


Figure 2.1: The Premier Fantasy Tools website shows the continuous rise in the popularity of FPL from 2 million accounts in 2012 to over 8 million in 2020.

In turn, Figure 2.2 shows the games regional popularity, using a website called My FPL analysis [3], and we can see that most managers come from England with over 2 million. There are also many managers from African countries like Egypt which has the second highest count. Ireland also contains a large amount of managers.

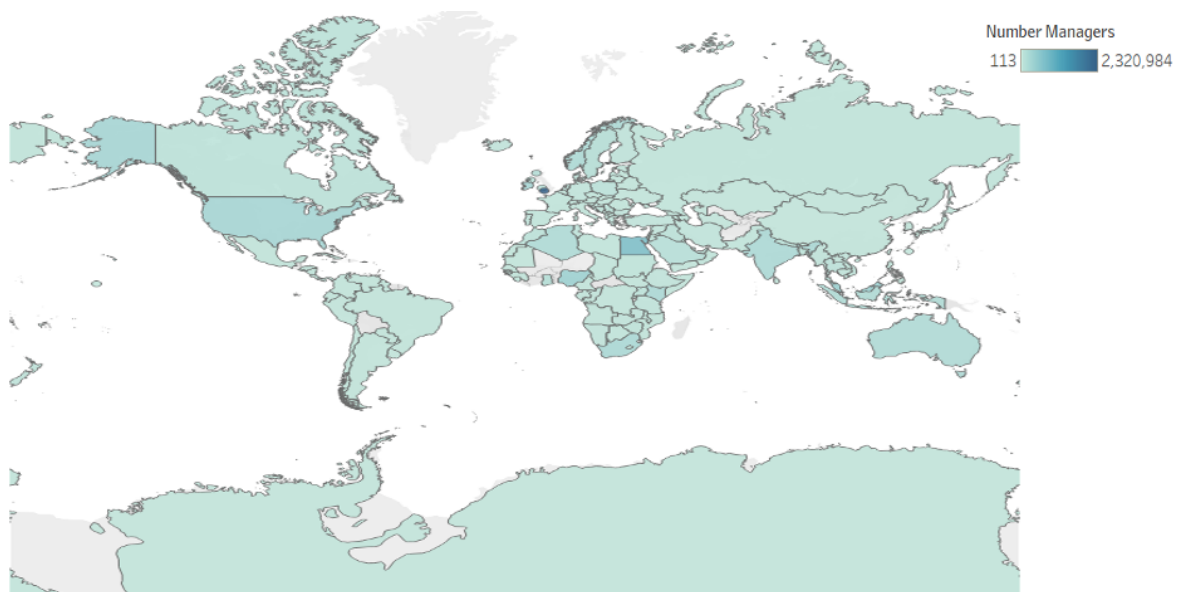


Figure 2.2: My FPL analysis data shows us that FPL managers are found in almost every country in the world, even in countries such as Antarctica.

Figure 2.3 gives us a more clear view of the top 10 countries for FPL managers. This is once again calculated using data from the My FPL analysis website, and we can see that the game is especially popular in Ireland and multiple African countries.

Number of Managers by Nation

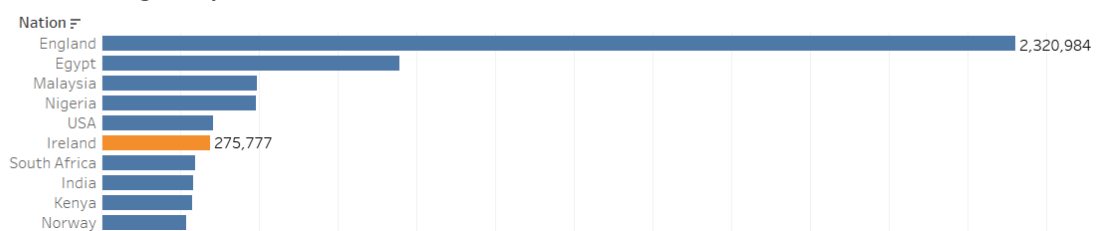


Figure 2.3: My FPL analysis data shows us that Ireland comes 6th in terms of ranking most managers with almost 300,000 accounts

Figure 2.4 shows us the game's regional popularity, this time based on Google searches, and we can see again that the game is especially popular in Ireland, but with Namibia leading the world in interest.

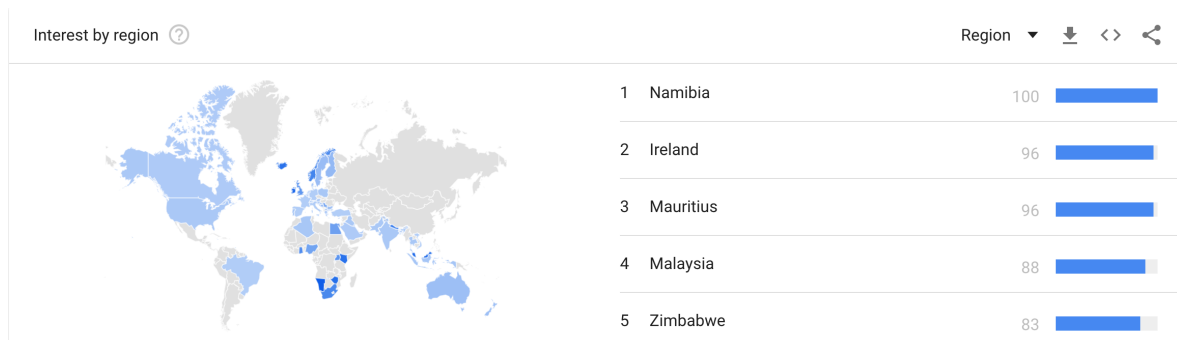


Figure 2.4: Google Trends data FPL popularity by region shows that Ireland is among the top countries with an interest in the game at the time of writing (May 2022).

---

It's worth noting too that the popularity of FPL has prompted the creation of various websites, blogs and social media accounts dedicated to providing FPL managers with team news, transfer tips, captain picks, advice regarding the Premier League and also team planning and transfer tools [4]. FPL has also spawned the creation of thousands of YouTube Accounts, some of which have amassed over 200,000 subscribers. These accounts are generally run by consistently high ranking managers, who provide their subscribers with tips and tricks to succeed. Some of the most popular FPL sites and YouTube channels include:

1. *LiveFPL*<sup>1</sup>: Website which you can use to track your exact FPL rankings in real time, both overall and in your mini leagues and understand how to move up in rankings. You can also track top10k and overall averages, captaincy and chip statistics and plan your team for future gameweeks.
2. *FantasyFootballScout*<sup>2</sup>: This is another website which is described as the number one resource for those playing Fantasy Premier League (FPL) and other leading Fantasy Football games, including offerings from Sky Sports, FanTeam, and many others.
3. *FPLTeam*<sup>3</sup>: This is arguably the easiest FPL team/transfer planning website to use. It is very straight forward to use and, unlike many of the other sites, it doesn't require that you create an account.
4. *FPLtips*<sup>4</sup>: This YouTube channel was created in 2013 and has gained over 200,000 subscribers. It is one of the most popular FPL YouTube pages, as it provides quality FPL content each week and is the place to go for Gameweek Previews, Players to Watch, Team Selection, Q+A's and more.
5. *LetsTalkFPL*<sup>5</sup>: This YouTube channel was created in 2010 and has amassed 268,000 subscribers and 60 million video views. The channel owner has played FPL for the past 12 years, achieving some very high ranking finishes. The channel is a reliable source for viewing weekly Fantasy Premier League content, as well as pre-season videos, to bring you all the FPL Tips you need.
6. *FPLFocal*<sup>6</sup>: This is another popular FPL YouTube account, however it gained the majority of its subscribers in a very short period of time. This occurred as a result of the owner of the Account achieving the Number 1 rank in the FPL global league at the end of 2021 (1st out of 9 million for context). FPL Focal only created his YouTube account at the end of 2020 but has gained 71,000 subscribers in that period of time. This page is a reliable source for weekly FPL team selection content.

Even if you have no Premier league knowledge whatsoever, simply reading the weekly advice on these sites and following the team and transfer recommendations of these YouTubers, is likely to score you more points than the majority of FPL players.

---

<sup>1</sup><https://www.livefpl.net/>

<sup>2</sup><https://www.fantasyfootballscout.co.uk/>

<sup>3</sup><https://fpl.team/>

<sup>4</sup><https://www.youtube.com/user/FPLtips>

<sup>5</sup><https://www.youtube.com/c/LetsTalkFPL>

<sup>6</sup><https://www.youtube.com/c/FPLFocal>

## 2.1 FPL Game-Play

Let's now take a closer look at FPL and its game-play, and discuss what makes it especially suitable as a data science project candidate. As mentioned, FPL managers are allocated a budget of £100M to spend on an initial squad of 15 players, who are valued based on the number of points that they're projected to deliver for the season. This initial squad must consist of two goalkeepers, five defenders, five midfielders and three forwards. It is also important to note that a maximum of three players can be chosen from any one Premier League club.

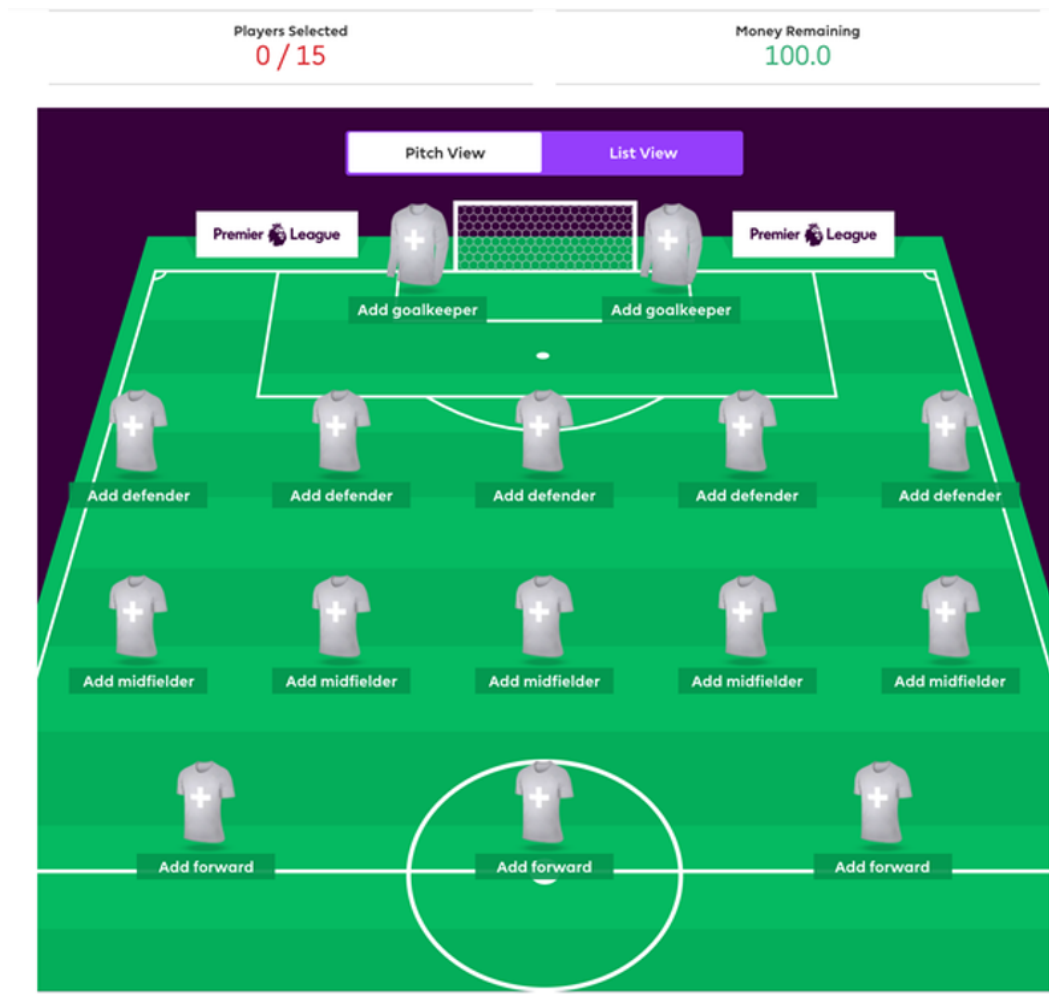


Figure 2.5: The empty squad which you are presented with when you sign up to FPL. The remaining budget and number of players chosen are shown above, and will be altered once you begin to add players to your squad

### 2.1.1 Selecting your starting XI

Each game week, managers are tasked with selecting 11 starting players from their 15 man squad, who they believe will score the most points. Managers can choose any formation for their starting XI, as long as 1 Goalkeeper, 3 defenders and 1 forward are selected at all times. Your substitutes, i.e. the four players not selected, provide cover for unseen events, replacing starting players who don't play in a game week. Starting players who register zero minutes are replaced by the highest priority substitute who did play that game week, assuming formation constraints are still satisfied. From your starting XI, you must also choose a captain and a vice captain. Your captain's points tally will be doubled for the game week, unless they register zero minutes, in which case the vice



captain fulfils the role of captain for the week. If neither your captain nor vice captain play in a game week, no player will score double points that game week. All changes to teams must be made by the game week deadline, which is typically 90 minutes before the kick off time of the first match in the game week.



Figure 2.6: Starting XI in a 3-4-3 formation, where players and their upcoming fixtures for the week are shown. Salah is chosen as our captain and Ronaldo as our vice captain. The star beside certain players indicates that they were one of the highest scoring players in the previous game week.

On occasion, some of your players may not be able to play a game week due to injuries, suspensions, health issues etc. A red flag appears beside players who are definitely ruled out of a game week, while a yellow flag will highlight players that are a doubt for the week due to injuries or health problems. Clicking on the flag will give you the reasoning behind a player's doubt for the week.

---

## 2.1.2 How Points are Scored

Points are scored in FPL based on a players performance in a game. Positive involvements such as goals, assists and clean sheets are rewarded, whereas negative involvements like yellow/red cards and goals conceded result in a deduction of points. Players are awarded points based on a number of criteria depending on their position. Some actions receive the same point gains or deductions regardless of player position e.g. assists, yellow/red cards, minutes played. Other actions differ based on the position of a player e.g. goals and clean sheets. As mentioned before, a team's points for the game week will be scored by their 11 starting players. If a starting player does not feature for their club, the points scored by the first player on your bench will automatically be counted instead. The same process occurs if two or three starting players fail to appear for their teams. Managers should therefore rank their bench in order of preference, to ensure that their first-choice reserve player is first in line. Managers must choose a captain and a vice-captain for the game week in their XI. Your captain's score is doubled, but if the selected skipper does not play in the game week then the vice-captain's score is doubled instead. Finally, bonus points are also awarded to the three best-performing players in each match, and are calculated using a separate bonus points system. There are also some game weeks in which players will have two fixtures. These are called double game weeks and they are particularly important game weeks for picking your ideal captain and making the most of the special chips to maximise points.

Action	Points
For playing up to 60 minutes	1
For playing 60 minutes or more (excluding stoppage time)	2
For each goal scored by a goalkeeper or defender	6
For each goal scored by a midfielder	5
For each goal scored by a forward	4
For each goal assist	3
For a clean sheet by a goalkeeper or defender	4
For a clean sheet by a midfielder	1
For every 3 shot saves by a goalkeeper	1
For each penalty save	5
For each penalty miss	-2
Bonus points for the best players in a match	1-3
For every 2 goals conceded by a goalkeeper or defender	-1
For each yellow card	-1
For each red card	-3
For each own goal	-2

Figure 2.7: During the season, your fantasy football players will be allocated points based on their performance in the Premier League

## 2.1.3 Making Transfers

After you have chosen your initial 15 man squad and the first deadline of the season has passed, managers are given one free transfer for each game week. This allows you to swap under performing, costly or injured players from your team with players who are in form and/or have favourable upcoming fixtures. This swap must be within the manager's budget. If managers are unhappy with multiple players or targeting multiple new players, they can make additional transfers in a

---

game week, however they will lose four points for each additional transfer they make. On the other hand, if managers do not use their free transfer, they can carry it over and have two free transfers for the following game week. Managers are unable to build up more than two free transfers.

## 2.1.4 Player Prices

Player prices are subject to change throughout the season, depending on their popularity among all the managers in the game. Prices fluctuate daily based on how many FPL managers are transferring the player in or out each day. On a single day, a player can increase or decrease by 0.1 million depending on the rise/fall in their overall ownership. To give an example, West Ham midfielder Jarrod Bowen cost £6.5 million at the beginning of the season. By game week 5, he had fallen to a price of £6.3 million as a result of a fall in ownership (likely due to poor performance), yet by game week 27, Bowen price increased to £7.1 million. Managers who own the player while his price is increasing will make a profit if they transfer him out at the higher price. However, they must pay a sell-on fee of 50 per cent of the profit made, rounded down to the nearest £0.1m. So in the example of the Jarrod Bowen who's price has risen from £6.5 to £7.1m, assuming you bought him at the start of the season, you would receive £6.8m when transferring him out. If you had bought him at his lowest cost of £6.3 million, you would receive £6.7 million for transferring him out. The fluctuation of player prices which can result in an addition to or deduction from the original £100 million budget, makes the initial selection of your FPL squad and the weekly transfers you make, that bit more important.

## 2.1.5 Special Chips

Another aspect that FPL managers must take into consideration is the five special chips allocated to them each season. These five chips include:

1. Two Wildcards - the most powerful chip, allowing managers to make any number of transfers in a game week free of any additional transfer charges. Essentially, this provides each manager the opportunity to make wholesale changes to their entire 15-man squad. One Wildcard is available for each half of the season.
2. Free Hit - allows you to change your team in its entirety for a single game week. This is useful to alter your team in a week of unfavourable fixtures. Once the game week is finished, your squad is reverted to the way it was before you applied the chip.
3. Triple Captain - your captain's points will be tripled for one game week.
4. Bench Boost - Every player in your squad, including your bench players, can score points when this chip is applied. This is useful for a week when you feel that all your players will score highly.

It is worth noting that only one chip can be played in each game week.

## 2.1.6 Leagues & Prizes

In FPL, you compete with the 9 million other managers in the global league and you also have the option to simultaneously compete against your friends in mini-leagues. These mini-leagues are your own private leagues where you compete against your friends. You can simply create a

---

league and then send out the unique code to allow other friends to join. You can compete in up to 25 invitational leagues. There is no limit on the number of teams in a single league. Managers can also join a public league of 20 randomly-assigned teams and can compete in up to five public leagues.

With regards to prizes, private league members can set their own buy-in costs and subsequent prizes, whereas for the more formal global league, the top 3 FPL managers and also the highest scoring managers each week and each month are awarded a variety of prizes.

---

## Prizes

### Overall Prizes

#### 1st Prize – 2021/22 Fantasy Premier League Champion

- 7-night break in the UK including VIP hospitality at two 2022/23 Premier League matches
- A week's worth of experiences at popular tourist attractions courtesy of Visit Britain
- Includes travel and 7 nights' accommodation
- Hublot connected watch
- Copy of FIFA 22 and a games console
- Nike manager jacket
- FPL goody bag consisting of a rucksack, t-shirt, mug, water bottle, stress ball, pen, pad and key ring

#### Runner-up

- VIP hospitality at one 2022/23 Premier League match
- Includes travel and 2 nights' accommodation
- Copy of FIFA 22 and a games console
- Tablet computer
- Bluetooth speaker
- Nike manager jacket
- FPL goody bag consisting of a rucksack, t-shirt, mug, water bottle, stress ball, pen, pad and key ring

#### 3rd place

- Copy of FIFA 22 and a games console
- Tablet computer
- Bluetooth speaker
- Nike manager jacket
- FPL goody bag consisting of a rucksack, t-shirt, mug, water bottle, stress ball, pen, pad and key ring

---

## Monthly Prizes

### Manager of the Month

- Tablet computer
- Bluetooth speaker
- Copy of FIFA 22
- Nike Flight match ball
- Nike manager jacket

Top 10 Managers each month (incl Manager of the Month) will receive:

- FPL goody bag consisting of a rucksack, t-shirt, mug, water bottle, stress ball, pen, pad and key ring

## Weekly Prizes

### Manager of the Week\*

- Copy of FIFA 22
- Nike Flight match ball
- FPL rucksack
- FPL mug

Top 20 Managers each week (incl Manager of the Week) will receive:

- FPL t-shirt, stress ball, pen, pad and key ring

---

\* To be eligible to win a weekly prize, you must have joined the game before the start of the season or a minimum of two clear Gameweeks before the one in which you are top of the weekly leaderboard. Players who have played a chip (including Bench Boost, Free Hit Triple Captain or Wildcard) will not be eligible to win a weekly prize for the Gameweek that such a chip has been used.

Figure 2.8: Prizes awarded to stand out managers in the global league made up of 9 million players

## 2.2 FPL on Social

For such a wildly accessible and increasingly huge game, fantasy football can really mess with your head. Millions of people have tried and failed at playing this simple but infuriating game – pick 15 players and amass points for every goal, assist, clean sheet and more. The official Fantasy Premier League game now boasts more than nine million players but Premier League outfits are beginning to ban it from dressing rooms due to the amount of attention and concentration it demands. To give evidence for the complexity of the game, one of the most consistent players, who was actually vying for the FPL title two years ago, is the five time reigning five-time World Chess Champion Magnus Carlsen. Magnus is consistently one of the world's best FPL players, putting his predictive strategy skills to extra-curricular use, yet even he is ranked at 19,000th this year and has failed to ever come out on top in the global league (his best season saw him placing 10th in 2019/20).

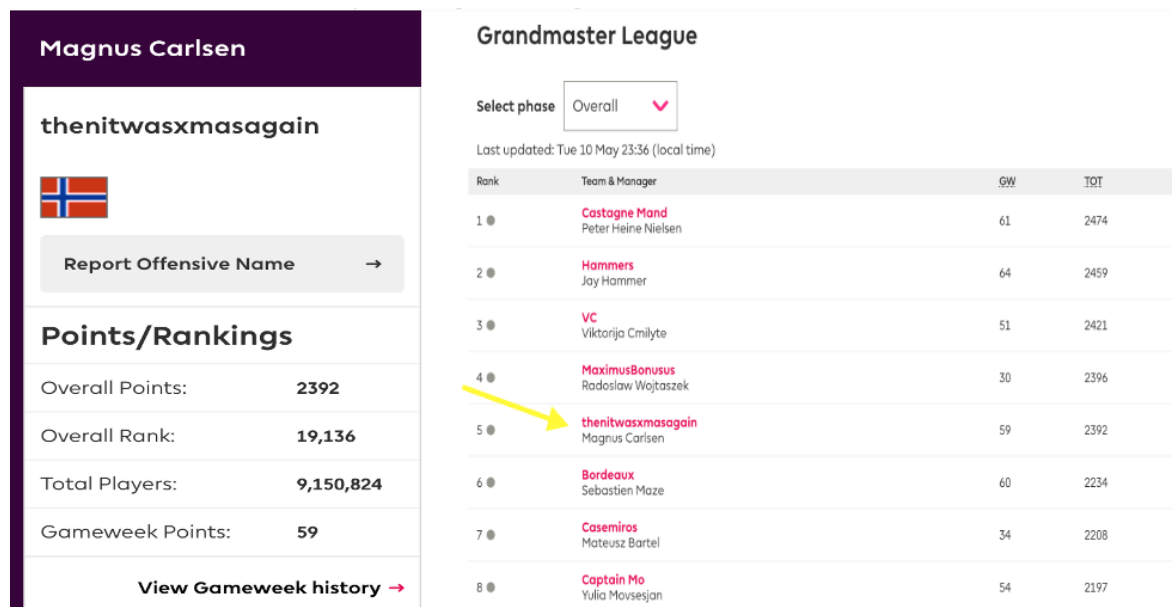


Figure 2.9: Magnus Carlsen's points and overall rank this year (left). Along with his rank in a table of 15 chess grand masters, 5 of whom score in the top 20,000 of all 9 million FPL players

Over the past decade, the rise of FPL has grown in tandem with the rise of social media engagement on platforms such as Twitter, Facebook and YouTube. Many football fans originally created fantasy football content websites to merely justify their interest in the game. However, in recent years these fun side projects have escalated into full-blown businesses. Fantasy Football Scout [5] is perhaps the biggest, with more than 250k Twitter followers, 70k subscribers on YouTube, and thousands of content subscribers who pay charges for an annual pass to stats tables, member articles, community chat and more. As well as the growth of its community, it now features a host of social content writers, presenters, and social media influencers.

As FPL players progressively turn more and more to private social networks to discuss topics and share content with friends, family and fellow football enthusiasts, the FPL content creators have capitalised on this opportunity, pivoting to whatever avenue works best and monetising it. Twitter still leads the way in FPL content as it provides a free discussion platform for fantasy football, with managers posting their team line ups each game week, and thousands of dedicated accounts providing detailed updates on Premier League fixture, club and player activity each day.



Figure 2.10: The official Fantasy Premier League Twitter Account [6] posts daily updates on player news and price changes. The account currently has over 3.5 million followers

The accessibility and immediacy of Twitter and football has given rise to many FPL influencers on the platform. Many content creators have large audiences such as Ben Crellin [7] with 200k followers, FPL General [8] with 180k followers and Ben Dinnery [9] with 150k followers. These creators would previously have spent time in forums offering their thoughts and opinions, but now they have thousands of people hanging on their every tweet, injury update, and transfer planning tip. In addition to Twitter, content creators on YouTube are also growing in popularity. Large dedicated audiences tune in several times a week to watch live broadcasts by creators like Andy@LetsTalkFPL [10] who has gathered more than 180k Twitter followers and 250k YouTube subscribers. Andy makes most of his money by way of his Patreon account, through which his supporters donate to him for his work. The Patreon platform has become a key content strategy for many creators in the FPL community, with creators offering exclusive content for audiences who pay monthly subscriptions. FPL has flourished because of social media and many creators have done a great job of reacting well to what works efficiently on social, adding features that are helping to build their businesses and help grow the popularity of FPL even more.

---

## 2.3 FPL for Data Science

What is it exactly that makes FPL so interesting to data scientists around the world? There are a number of different questions FPL managers have every year. On top of that there is a vast amount of data available on FPL, dating back more than 5 years. This data can be used to solve the answers to these questions, or at least look for some sort of answer to them.

Year on year, FPL managers look for ways to improve, and every year there is a new FPL machine learning project that looks at optimising game-play, be it for picking a team, or predicting how players will perform in a certain game week. Every manager questions themselves each week when making decisions. Questions such as "who should I captain?", "should I use my triple captain?" and "Should I transfer this player out?", spring to mind.

There is so much data available for us to answer the questions asked, and paired with data science, much sought after answers can be found to common questions posed by FPL managers. There are many machine-learning projects and scholarly articles that tackle these questions, providing scientific answers to them. The most common of all is a machine learning model that picks an optimal FPL team for a given season/game week.

Hence, we can see the relevance that FPL has in the world of data science and exactly why it is so suited to data science projects. It's a popular and very tactical game, with over 9 million managers looking to improve their game-play skills. There are also lots of questions to be asked with regards to game-play. All of this allows FPL to be explored in a detailed manner by data scientists. This is demonstrated extremely well by Joshua Bull, the winner of the 2019-2020 FPL season, beating over 8 million other players. He explains how he used data science to optimise his game-play in a YouTube video [\[11\]](#).

## 2.4 Objectives & Motivations

The main reason we decided to work on this particular project is the strong interest we have in Fantasy Premier League, and football itself. We enjoy playing FPL every year in our mini-leagues with friends as this adds some competition to the game. We believed this project could help us excel in FPL, thus giving us the advantage to maintain and enjoy 'bragging-rights' over friends, through dominating our mini-leagues. We are both passionate about FPL and are fascinated to find any potential advantages when it comes to playing the game and potentially winning the global league. Another factor in our decision to choose this project was the fact that, FPL also has such a huge following worldwide, as discussed earlier in the project [2.1](#). The fact that this game is of interest to over 9 million other people around the world, was part of the rationale behind selecting our project.

What we hope to achieve in completing this project is the "ultimate" FPL playing strategy. We hope that we can find ways to improve our current game-play to push us as far up the leader-board as possible. We will investigate all the aspects of the game in order to build our strategy. We will look into picking formation, using the budget, transfers, and predicting player performances. We will then use these findings to look into how we should use our chips. We hope that after all of our research, we will become much superior FPL managers, comfortably surpassing our ranks for previous seasons.



---

## 2.5 Research Questions

1. RQ1: How does formation play a part in overall points achieved? – Analysing the performance of different positions and formations in order to find the best formation to use. Formation is an important feature of an FPL team as it helps us out with picking our players and using the budget to maximise points.
2. RQ2: How to make best use of the budget? – Investigating where most of the budget should go and how best to use it. What types of players should we put into our team? How will this effect our points achieved? What players could be potential bargain bin players for the next season? All these questions will be examined.
3. RQ3: How transfers can affect a fantasy teams points? – Examine the effect that transfers have on the points achieved by teams to figure out the best way to use them. How should we use our transfers and will they be effective?
4. RQ4: Can we predict which players will perform in a certain game week? – Exploring different prediction methods to see how effective they can be and if we can predict how a player will perform in a given game week in order to maximise points. This will help with decision making when it comes to selecting our captain and playing our wildcards.



---

## Chapter 3: Related Work

---

When first starting this project, we knew that similar things had been done before. There were many examples online of Fantasy Premier League teams being built solely through machine learning, and other analyses focusing in on particular parts of the game such as transfers or performance predictions. There were also plenty of online accounts on Twitter and YouTube with channels that discussed FPL week by week. One thing that struck us was that it was very difficult to find any strategies that covered every part of the game. Inspired by the work of others, we decided that we would come up with a science-backed FPL playing strategy, not only to help improve our own FPL game-play, but to help others and new-comers to the game who have an interest in the mathematical side of things. This strategy would cover every part of FPL game-play, unlike previous projects.

A lot of the related work we came across focused mainly on performance predictions using different machine learning techniques. This was of interest to us, as the aim of the game is to score as many points as possible, and what better way to pick your players and captain than by knowing how many points they would score in a given game week. "Time Series Modeling for Dream Team in Fantasy Premier League" [12] looks into Autoregressive Integrated Moving Average (ARIMA) and Recurrent Neural Networks (RNNs) to predict players performances over the course of a season in order to pick an FPL team using a cost-point index to make the picks. Their model performed with pretty good accuracy, with an error of only 87 points over the course of an entire season for the team picked. An average error of 2.3 points per game week for the team. This showed us the possibilities when it came to points prediction.

Another article we came across, which was of great interest, looked into picking players. "How To Win Your Fantasy Football League, By Our Data Scientists" [13] delved deeply into safe-bet players, and 'maverick' players. Maverick players are classified as those with a high volatility of points. For example, they could score 16 points in one game week, and 3 in another. Safe-bet players score consistently throughout the season. It uses Monte Carlo Simulations to identify the best maverick players, then looks into probabilities of them scoring over 10 points, ultimately coming to the conclusion that you should take a gamble on maverick players as a strategy to build up your points if you are lagging behind others towards the end of the season. We took inspiration for this when it came to looking into putting together the 15 man squad.

There are also many FPL strategies to be found online, however these are not formed by data science. FPL experts compile these strategies based on their knowledge of the game to help other FPL managers. A website, [beatingbetting.co.uk](https://beatingbetting.co.uk), has it's own beginners strategy to FPL [14] containing information on every aspect of the game. It explores picking players, using chips, transfers, differential players and using the budget. This inspired us to investigate similar aspects of the game ourselves, however we would take a more scientific approach to things.

There is very little work published regarding the use of transfers and picking your teams formation. We decided that investigating these aspects of Fantasy Premier League could differentiate our project from previous work. This would also help to enhance our strategy by making it more useful for FPL managers, especially new-comers, as it covers all aspects of the game.

---

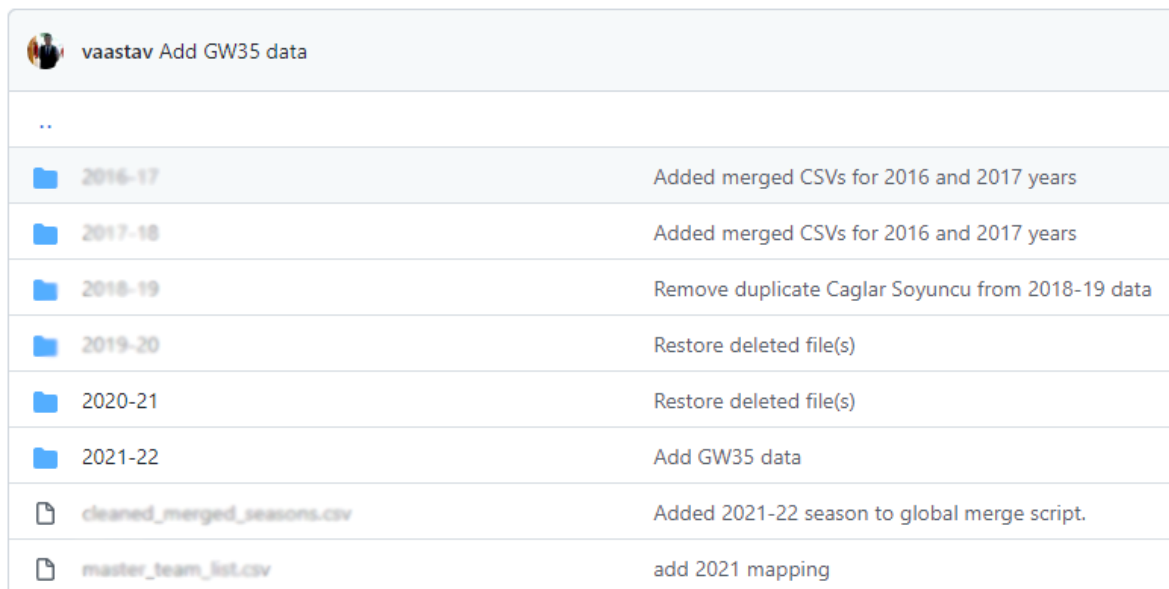
## Chapter 4: Data Considerations

---

Our dataset was available on a GitHub repository that was updated weekly, in line with FPL itself (<https://github.com/vaastav/Fantasy-Premier-League>)[15]. This made our data collection process very simple. The dataset contained weekly FPL data from the 2016/2017 season up until the current season (2021/22). This data continues to be updated weekly as the season progresses. Each season consists of a folder containing game weeks from that season, and information on Premier League teams and their fixtures. All of this data is in .csv format. This dataset was chosen as it contained all the relevant data to perform our analysis, while also being extremely easy to access and seemed to require minimal cleaning/preparation.

### 4.1 Data Collection

The data collection process was short and simple. We downloaded the data from the 2020/2021 and 2021/2022 seasons and saved this data in separate folders based on the season.












 vaastav	Add GW35 data
..	
 2016-17	Added merged CSVs for 2016 and 2017 years
 2017-18	Added merged CSVs for 2016 and 2017 years
 2018-19	Remove duplicate Caglar Soyuncu from 2018-19 data
 2019-20	Restore deleted file(s)
 2020-21	Restore deleted file(s)
 2021-22	Add GW35 data
 cleaned_merged_seasons.csv	Added 2021-22 season to global merge script.
 master_team_list.csv	add 2021 mapping

Figure 4.1: We downloaded the 2020/2021 and 2021/2022 data.

vaastav Restore deleted file(s)		
..		
gws		Fix merged_gw with new data
players		Restore deleted file(s)
understat		Add GW38 data
best_players.csv		Add script for best players
cleaned_players.csv		Fix gw38 numbers
fixtures.csv		Add GW38 data
player_idlist.csv		Add GW38 data
players_raw.csv		Fix gw38 numbers
teams.csv		Add GW31 data

Figure 4.2: From this we kept the gws folder and the fixtures, player\_idlist and teams files.

vaastav Fix merged_gw with new data		
..		
gw1.csv		Add missing GW1 folders
gw10.csv		Add GW10 data
gw11.csv		Add GW11 data
gw12.csv		Add GW12 data
gw13.csv		Add GW13 data
gw14.csv		Add GW14 data
gw15.csv		Add GW15 data
gw16.csv		Add GW16 data

Figure 4.3: Each game week contained 38 separate .csv files with info on each separate game week.

---

## 4.2 Data Preparation

Now that all the data we needed was downloaded and saved, the next step was to prepare our data for analysis. This involved the gws folder and the player\_idlist, fixtures and teams files.

The player\_idlist contained players first name, surname and player id. This was prepared in order to add player ids to the gws data. The first name and surname were combined in each row so it could be joined on 'name' in the gws data. This new dataframe was saved as a pickle file.

	first_name	second_name	id
0	Mesut	Özil	1
1	Sokratis	Papastathopoulos	2
2	David	Luiz Moreira Marinho	3
3	Pierre-Emerick	Aubameyang	4
4	Cédric	Soares	5

	id	name
0	1	Mesut Özil
1	2	Sokratis Papastathopoulos
2	3	David Luiz Moreira Marinho
3	4	Pierre-Emerick Aubameyang
4	5	Cédric Soares

The gws folder had to be parsed and merged into one large dataframe to make it more manageable. In order to do this, a list of all files in the folder had to be firstly ordered by the corresponding game week number, then parsed into an array of dataframes. Then the array of dataframes could be flattened into one large dataframe. Then the data could be cleaned, removing any players that played 0 minutes all season and setting all forwards clean sheets to 0, as they cannot be awarded points for clean sheets. This dataframe was then saved and stored as a pickle file.

Some of the code used in preparing game week data:

```
#Create and sort list of files in folder
all_files_gws20_21 = glob(gws20_21)
all_files_gws20_21.sort(key=lambda f: int(re.sub('\D', '', f)))

#Create array of DFs
gws20_21_dfs = []
for i in all_files_gws20_21:
    logger.info((i))
    gws20_21_dfs.append(pd.read_csv(i, parse_dates = True))

#Flatten array of DFs
gws20_21_dfs = pd.concat(gws20_21_dfs)

#Remove non-players and remove clean sheets for FWDs
non_player20_21 = gws20_21_dfs.groupby('id')['minutes'].sum().to_frame()
non_player20_21 = non_player20_21[non_player20_21.minutes == 0]
non_player20_21 = non_player20_21.drop(['minutes'], axis = 1)
non_player20_21 = non_player20_21.index.to_list()
len(non_player20_21)
```

---

```
gws20_21_dfs = gws20_21_dfs[~gws20_21_dfs.id.isin(non_player20_21)]
gws20_21_dfs.loc[gws20_21_dfs['position'] == 'FWD', 'clean_sheets'] = 0
```

The teams data was then prepared. It didn't require much cleaning. A few unnecessary columns were removed in order to reduce the file size and make the data more manageable.

```
#Removing unnecessary columns
teams20_21_df = teams20_21_df.drop(["draw", "form",
                                     "loss", "played", "points", "position",
                                     "team_division", "unavailable", "win"],
                                   axis=1).set_index('id', drop=False)
```

Fixture data was cleaned and prepared however it was not used in any analysis.

---

## Chapter 5: Results

---

All of the findings of our research questions help us to gain an insight into different aspects of the game of Fantasy Premier League. Each result presented us with a scientific way to maximise points in FPL. These results are then used to put together our FPL playing strategy.

### 5.1 RQ1: How does formation play a part in overall points achieved?

For this research question we take on a topic that has been the subject of considerable debate among FPL players: what formation should we use each game week. Remember, in FPL you are tasked with building a squad of 15 players including 2 Goalkeepers (GK), 5 Defenders (DEF), 5 Midfielders (MID) and 3 Forwards (FWD). From this 15-man squad, manager's must choose a starting XI in a formation that satisfies a number of constraints. Formation's must contain one GK, a minimum of three DEF and at least two FWD. These constraints leave us with seven formations to choose from: 3-4-3, 3-5-2, 4-3-3, 4-4-2, 4-5-1, 5-2-3, 5-3-2. There are no hints to guide you as to what formation to opt for each week, so most players generally pick the 11 players they believe are most likely to start that game week, without investigating which combination of positions, i.e. formation, is more likely to yield the most points in a particular game week. For example, the player's you have who are facing weaker opposition on a given week are usually a good bet, as regardless of whether a manager has chosen a strong squad at the beginning of the season, players of every position are likely to score higher points against weaker teams. But the question we pose is do certain positions take greater advantage of weaker opponents? Do Defenders score high against weak teams, where clean sheets are likely, thus making a 5 at the back formation more favourable if your defenders have easier fixtures than your other positions? What factor do team strength, opponent strength, player ability etc. play in achieving points? Is there a certain combination of positions (formation) which consistently scores the highest points regardless of the 'ability' of players in your squad. If so, what formation is this and how reliable is it throughout the season.

#### 5.1.1 Data & Method

We will use data from the most recently completed Premier League season (2020/21) to perform our analysis. The data we used was from the cleaned gws20\_21.pkl and teams20\_21.pkl files containing information on each game week and also corresponding teams data, from which we used the opponent strength and team strength indexes.

In order to get an idea of which formation to target, first we needed to investigate the relative points performance of each position. It was obvious that we needed to distinguish between position points performance for 'average' players and 'top' players. Average players would take every players points into account whereas the top/elite players would be the players which registered the highest points over the entire season. We were especially interested to see if a particular position had a higher percentage out of the elite players.

We had to compare and contrast overall position performance for the season by analysing the total points for the season of every player i.e. the sum of their points for the 38 game weeks where  $\text{points}(\text{gw})$  is the total points in game week gw.

$$\text{PositionPoints}(\text{position}) = \sum_{\text{gw}=1}^{38} \text{points}(\text{gw}, \text{position})$$

The use of box plots was especially useful to find the distribution of total points for all players in each position. This helped to distinguish whether or not there was a group of outliers for a particular position, thus skewing the data. Another thing we had to take into account, was that there was a much higher quantity of defenders and midfielders, than goalkeepers and forwards. For elite players we analysed the top 20 players for midfielders and top 20 for defenders. Whereas for forwards and goalkeepers, we analysed the top 10 points performers from each. This observation led us even further when we investigated average position points for players for the season. This was calculated using the following equation:

$$\text{AvgPoints}(\text{position}) = \frac{\text{TotalPoints}(\text{position})}{\text{TotalPlayers}(\text{position})}$$

Once player position points and average position points for both average players (all players) and elite players were calculated across all the game weeks for our season, we wanted, ultimately, to look at the corresponding formation performance over the season. We calculated this for the season where GK is goalkeeper, D is defenders, M is midfielders, F is forwards, N is the numerical amount and  $\text{points}(N(D)-N(M)-N(F))$  is the total points for the formation in the form numDefenders-numMidfielders-numForwards:

$$\text{points}(N(D) - N(M) - N(F)) = (\text{AvgPoints}(\text{GK}) * 1) + (\text{AvgPoints}(\text{D}) * N(D)) + (\text{AvgPoints}(\text{M}) * N(M)) + (\text{AvgPoints}(\text{F}) * N(F)) \quad (5.1)$$

## 5.1.2 Results

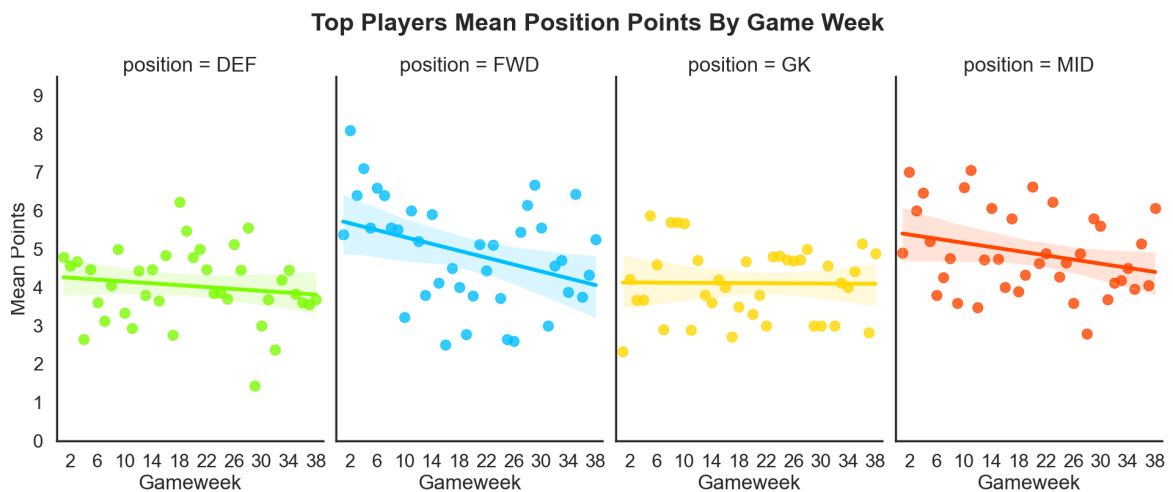


Figure 5.1: Regression plot showing the mean points by position each game week for the elite FPL players.

We can see from figure 5.1 that midfielders and forwards are the highest performers in terms of scoring points throughout the season.

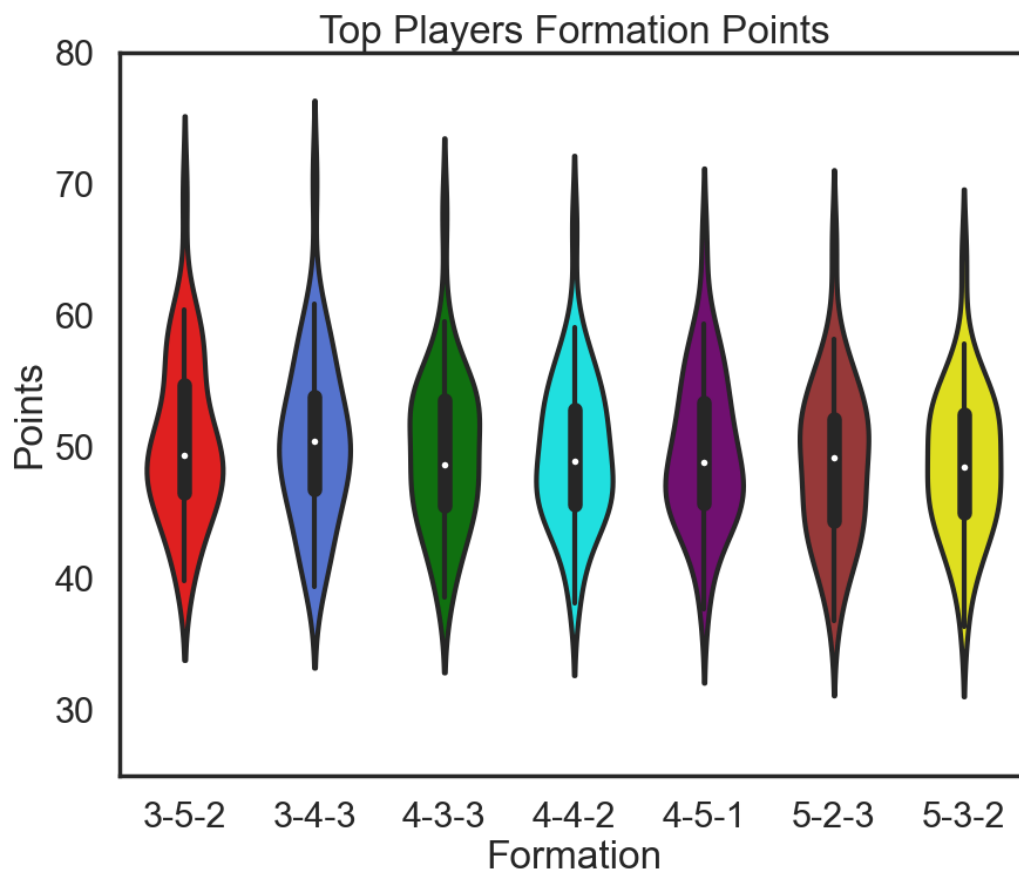


Figure 5.2: Violin plot showing the distribution of Elite players formation points across the 38-gameweek season.

The above figure 5.2 shows that a midfielder/forward dominant formation such as 3-5-2 and 3-4-3 score higher points throughout the season. We can see that 3-5-2 and 3-4-3 are more dense around the 60 point mark and contain the highest outliers. Each distribution contains the formation points for the 38 game weeks. 3-4-3 has the highest median and has the most even distribution. Formations with more defenders like 5-2-3 and 5-3-2 have the lowest scoring game weeks being more dense around the 40 point mark. Note that figure 5.2 looks at the elite players here as those are the players we are much more likely to be choosing from, however let's look next at both average and elite players when analysing which formation generates the most "wins" for the season.



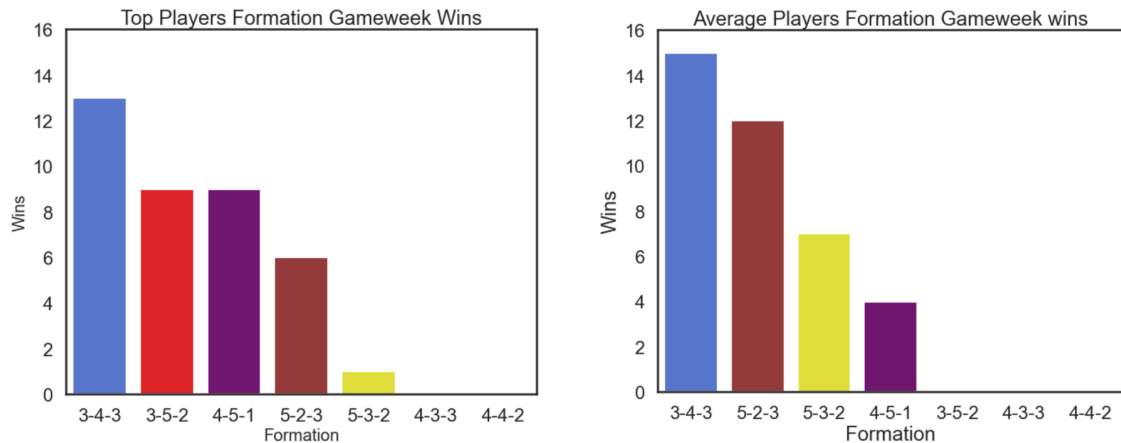


Figure 5.3: Bar Chart showing Formation wins for Elite Players and for Average Players.

In figure 5.3 we are analysing formation "wins" for the season. A win can be defined as the formation which accrues the largest sum of points in a particular game week. For both average players and elite players, 3-4-3 comes out on top with 15 and 13 wins respectively. Realistically we will be opting for elite players when picking our squad but we see here that even with an average squad 3-4-3 is the best performing formation. 3-5-2 and 4-5-1 also gather 8 wins each for top players.

### 5.1.3 Discussion

The drop off in points as the season progresses is notable especially for midfielders and forwards. This drop as seen in figure 5.1 is because Champions League Fixtures began for the four top Premier League teams after game week 6. This results in much more congested fixtures for the elite players as a large proportion of elite players play for the top teams. Elite players are often rested for Premier League games, taken off early or they are often played so often (due to their ability) that injuries and fatigue start to set in with the pile up in fixtures and therefore, points performances take a hit. The reason that goal keepers are so consistent throughout the season is that keepers are rarely rested as fatigue is less of an issue, injuries are much more rare and they are unlikely to receive yellow/red cards and thus avoid suspensions. Defenders suffer less of a decline in points performance as the season progresses as top defenders are unlikely to be rotated and don't fatigue as often as midfielders and forwards.

What we can take from these graphs in figure 5.1 is that we should definitely target formations which are midfielder/forward heavy, especially early in the season. It looks like a 3-4-3 would be best for the first half of the season, while near the end of the season a formation with more midfielders like 3-5-2 would make more sense. Three at the back seems like our most logical formation throughout the season, however if in a particular week our defenders are playing against weaker opponents than midfielders or forwards, we should look to potentially extend to four or potentially five at the back.

Figure 5.2 confirms our theory that a midfielder/forward dominant formation like 3-5-2 and 3-4-3 scores higher points throughout the season. We can see that 3-5-2 and 3-4-3 are more dense around the 60 point mark and contain the highest outliers. 3-4-3 has the highest median and has the most even distribution, showing its consistency for the season. Formations with more defenders like 5-2-3 and 5-3-2 look like ones you should avoid as they have the lowest scoring game weeks and are more dense around the 40 point mark.

---

Once again figure 5.3 shows us much of what we would've expected. For both average players and elite players, 3-4-3 comes out on top with 15 and 13 wins respectively. Realistically we will be opting for elite players when picking our squad but we see here that even with an average squad 3-4-3 is the formation to go for, as it actually registers even more wins for average players. This confirms our theory that we should always try to get a balance of midfielders and forwards, yet near the end of the season we should start to look more at midfielder dominant formations as forwards start to drop off with fixture build up and also we should always opt for the positions which have the lowest opponent strength on a given week.

## 5.2 RQ2: How to make best use of the budget?

In FPL we are given a starting budget of 100M. This budget must be spent on 15 players: 2 Goalkeepers (GK), 5 Defenders (DEF), 5 Midfielders (MID) and 3 Forwards (FWD). From these 15 we must have 11 starters: 1 GK, at least 3 DEF and at least one FWD. The other 4 players are benched. A maximum of 3 players can be selected from any one club. We will use the data provided to deeply analyse each player in order to figure out which players are the best possible players to have in your starting team. These players will be analysed in order to fit the 100M budget perfectly. Data from the 2020/2021 season will be used as this is the most recently completed season.

### 5.2.1 Data & Method

The data used was the cleaned gws20\_21.pkl and teams20\_21.pkl files containing information on each game week and also corresponding teams data, from which we used the team strength indexes.

The methodology behind investigating ways of using the budget involved firstly examining the relationship between points achieved and value based on players. This was then broken down into separate positions to see if there are certain positions which are more beneficial to invest our budget into. We also must keep in mind that we can only pick 3 players from each team, so we would have to look into various teams and teams of different strengths as well.

First of all, we explored the relationship between points and overall player value for each team, which was further investigated by analysing each position. This was done using scatter-plots and the calculating correlation coefficient for each plot. In order to create these plots we had to calculate the mean points and value for each team, grouping these by position, from which we could then calculate the correlation coefficients. This provided us insight into what positions we should invest most of our budget into.

Next we investigated the return on investment (ROI) of players in order to spot 'bargain-bin' players. This was calculated using the following equation:

$$ROI = \frac{PlayerValue(cost)}{TotalPoints(return)}$$

This allowed us to create two 15 man squads in line with the FPL rules of 2 GK, 5 DEF, 5 MID and 3 FWD: One 15 man squad based on the highest ROI and one 15 man squad based on the highest valued players. We wanted to compare how each squad performed over the season. The best possible 15 man squad over the season was also included. Budget was not taken into account for any of these squads. When picking the squad with the highest player value, both Virgil van

Dijk and Sergio Agüero were dropped as they were injured for the majority of the season. The squads created are shown:

	name	position	team	value
Based on ROI:	Ederson Santana de Moraes	GK	Man City	6.1
	Alisson Ramses Becker	GK	Liverpool	6.0
	Trent Alexander-Arnold	DEF	Liverpool	7.8
	Andrew Robertson	DEF	Liverpool	7.4
	Benjamin Chilwell	DEF	Chelsea	6.2
	Lucas Digne	DEF	Everton	6.2
	Kyle Walker	DEF	Man City	6.2
	Mohamed Salah	MID	Liverpool	12.9
	Sadio Mané	MID	Liverpool	12.1
	Kevin De Bruyne	MID	Man City	12.0
	Pierre-Emerick Aubameyang	MID	Arsenal	12.0
	Bruno Miguel Borges Fernandes	MID	Man Utd	11.6
	Harry Kane	FWD	Spurs	11.9
	Jamie Vardy	FWD	Leicester	10.3
	Gabriel Fernando de Jesus	FWD	Man City	9.5

	name	position	team	value
Based on Player Value:	Emiliano Martínez	GK	Aston Villa	5.4
	Illan Meslier	GK	Leeds	4.8
	Stuart Dallas	DEF	Leeds	5.5
	Matt Targett	DEF	Aston Villa	5.1
	Vladimir Coufal	DEF	West Ham	4.8
	Lewis Dunk	DEF	Brighton	5.0
	Aaron Cresswell	DEF	West Ham	5.9
	Jack Harrison	MID	Leeds	5.6
	Tomas Soucek	MID	West Ham	5.4
	Matheus Pereira	MID	West Brom	6.0
	Ilkay Gündogan	MID	Man City	6.2
	James Ward-Prowse	MID	Southampton	6.3
	Patrick Bamford	FWD	Leeds	6.9
	Ollie Watkins	FWD	Aston Villa	6.6
	Che Adams	FWD	Southampton	6.1

---

	name	position	team	value
Best Possible Squad:	Emiliano Martínez	GK	Aston Villa	5.4
	Ederson Santana de Moraes	GK	Man City	6.1
	Stuart Dallas	DEF	Leeds	5.5
	Andrew Robertson	DEF	Liverpool	7.4
	Trent Alexander-Arnold	DEF	Liverpool	7.8
	Aaron Cresswell	DEF	West Ham	5.9
	Aaron Wan-Bissaka	DEF	Man Utd	5.8
	Bruno Miguel Borges Fernandes	MID	Man Utd	11.6
	Mohamed Salah	MID	Liverpool	12.9
	Heung-Min Son	MID	Spurs	9.8
	Sadio Mané	MID	Liverpool	12.1
	Marcus Rashford	MID	Man Utd	9.6
	Harry Kane	FWD	Spurs	11.9
	Patrick Bamford	FWD	Leeds	6.9
	Jamie Vardy	FWD	Leicester	10.3

From this we designed a bar chart which plotted the total points scored by each squad, in order to help understand how we should pick our team in line with the 100M budget. Their points totals were calculated by adding up total points from each player, not taking into account captains or wildcards.

Team strength indexes were also explored in order to see what strength teams score the most points, and the differences between points scored for each strength level. One thing we had to note was that there was only one team of strength level 5, Manchester City. Box plots were used to plot the distribution of points by teams of different strength. This helped us develop on the previous findings of this research question.

## 5.2.2 Results

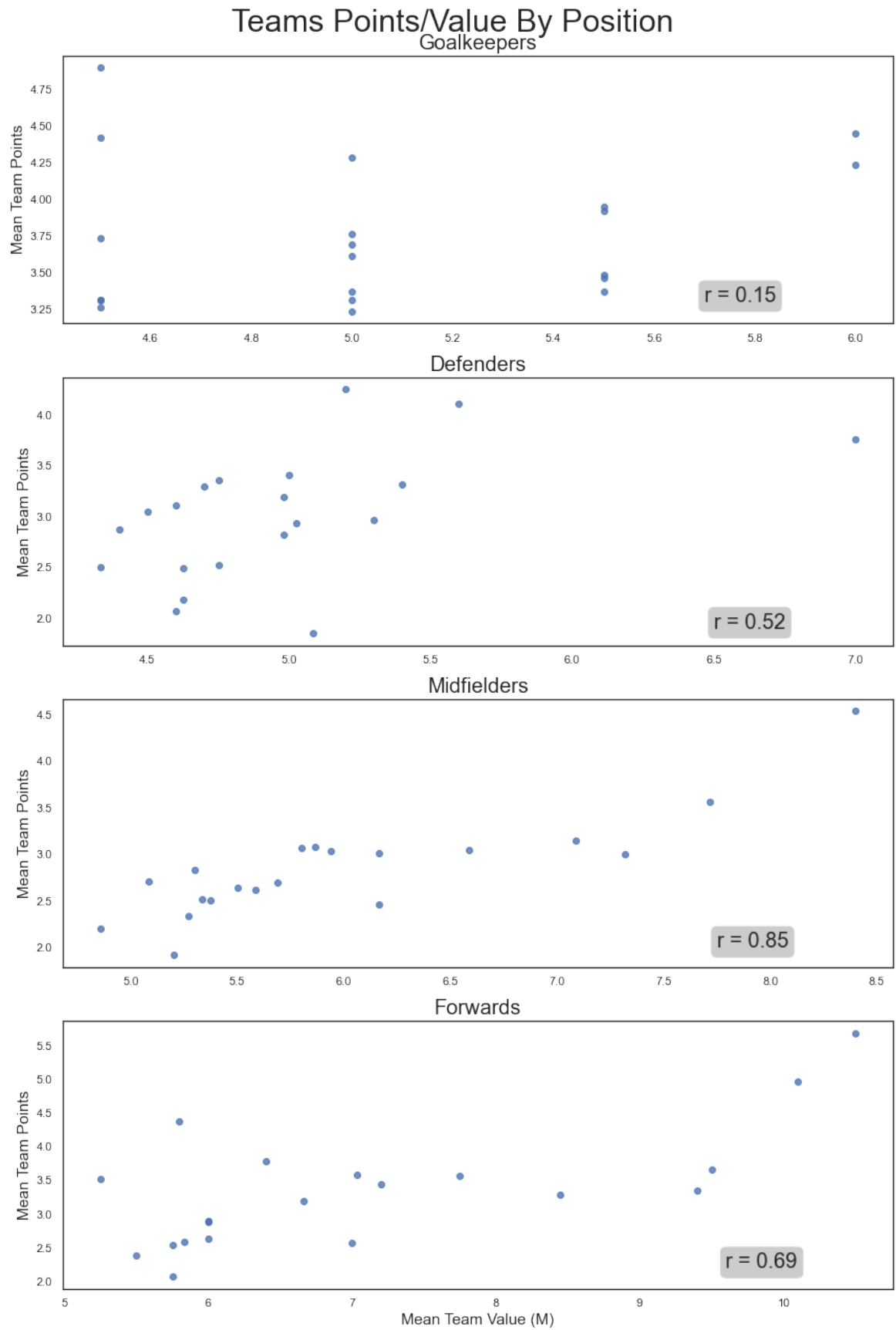


Figure 5.4: The correlation between points scored by teams and team value based on position.

The above figure 5.4 shows the scatter plots of mean team points and mean team value for each position, as well as the correlation coefficient for each. For each position we can clearly see how strong/weak the relationship is between value and points. For Goalkeepers there is no relationship ( $r=0.15$ ), for defenders there is a moderate relationship ( $r=0.52$ ), for midfielders there is a strong relationship ( $r=0.85$ ) and there is a moderate to strong relationship ( $r=0.69$ ) for forwards.

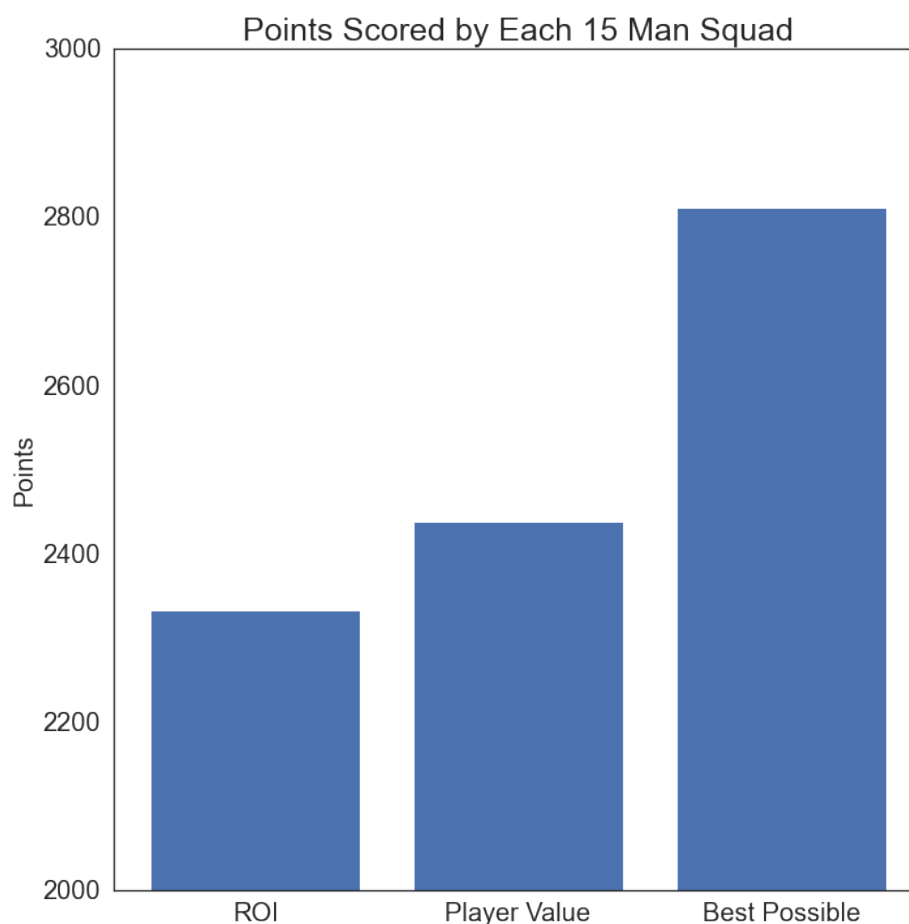


Figure 5.5: Performance of the best teams picked based on ROI, value and total points for the season

We can clearly see from figure 5.5 that the team picked based on the highest player values performs better than the team of the players with the highest ROI. The Player Value squad out performs the ROI squad by roughly 200 points. If captains and chips were utilised, the gap between these would increase even further. One thing to also bear in mind is that the ROI squad would be difficult to pick at the start of the season as it consists of the top 'bargain-bin' players. This helps us to conclude that players of the highest value score the highest points.

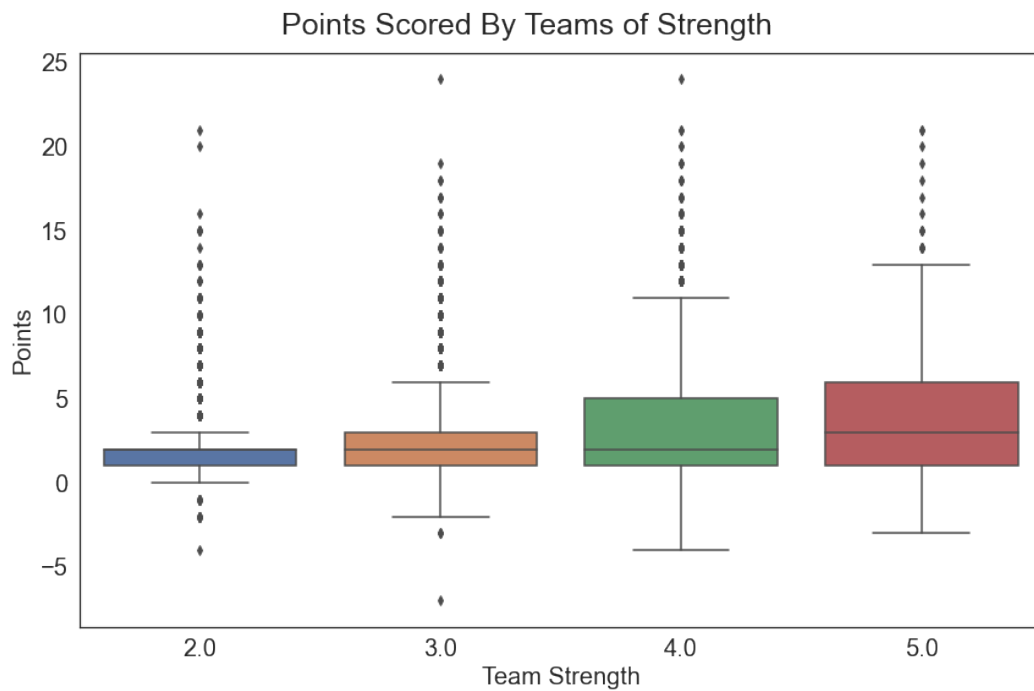


Figure 5.6: Points scored by players from teams of different strengths. Only one team of Strength 5 (Man City)

strength	
team	
Arsenal	4
Aston Villa	3
Brighton	3
Burnley	2
Chelsea	4
Crystal Palace	3
Everton	3
Fulham	3
Leicester	4
Leeds	3
Liverpool	4
Man City	5
Man Utd	4
Newcastle	2
Sheffield Utd	2
Southampton	3
Spurs	4
West Brom	2
West Ham	4
Wolves	3

The points distribution in figure 5.6 shows that Man City score the most points followed by teams of strength 4. There is then a dramatic drop off in points scored between teams of strength 3 and 4, while teams of strength 2 score the lowest points. One thing to note is that teams of strength 3 have the same median value as teams of strength 4, however the upper quartile for teams of strength 4 is much higher. Teams of strength 3 have a large number of high-scoring outliers which is down to their bargain-bin players.

---

### 5.2.3 Discussion

In figure 5.4, we can see the relationship between points scored and value for each team broken down into positions. These correlation coefficients tell us the strength of the relationship between the two for each position. As the correlation for goalkeepers is so low, we can effectively pick a cheap goalkeeper who will perform just as well as an expensive one. For defenders we have a moderate correlation, so we can save some of our budget here too as the price of a defender doesn't have a huge effect on how they perform. As we found in RQ1, midfielders can pick up the most points if chosen wisely, and we build on this here with a strong correlation for midfielders. This means we should choose expensive midfielders as value has a huge effect on their points scored. Forwards have a moderate to strong correlation. We should consider picking expensive forwards, but we could also risk it for a bargain in this position.

We delve deeper into this in figure 5.5 as we look at how teams picked based on the highest ROI and the highest value perform over the course of a season. One thing to bear in mind with this graph is that players with high ROI could be difficult to pick at the start of the season and the squad picked based on ROI is the best of the best. This would mean in a more realistic situation, the squad based on ROI would score lower. It is clear that the more expensive players perform better than those of the highest ROI, hinting that we should fill our squad with expensive, high scorers first before we go looking in the "bargain bin" to populate the rest of our squad.

To further investigate, we looked into the points scored by teams of different strength indexes in figure 5.6. This graph tells us that we should have 3 Man City players in our squad. It also tells us that the majority of players we pick should be from teams of strength 4. As there are a large number of outliers for teams of strength 2 and 3, we could potentially pick out "bargain bin" players from these teams.

## 5.3 RQ3: How transfers can affect a fantasy teams points?

Each game week, every FPL manager is given one free transfer to use, with each additional transfers made by a manager resulting in a four point deduction from their total for the week. Transfers may only be made when budget and team constraints are met. Transfers can be used to bring in a new in-demand player or to get rid of under-performing/injured/suspended players, however they do not have to be used. We will explore transfers made in each game week and how they had an effect on a team's points in order to figure out the best transfer-making strategy.

### 5.3.1 Data & Method

The data used was the cleaned gws20\_21.pkl and teams20\_21.pkl files containing information on each game week and also corresponding teams data, from which we used the team strength indexes.

Scatter-plots were used in order to plot the number of transfers-in in each game week alongside the points scored in each corresponding game week for each player. This scatter-plot 5.7 made it clear that it would be very difficult to find a relationship between transfers-in and points as many players had very few transfers-in in a lot of game weeks as they are the top players who most managers already had in their squad.



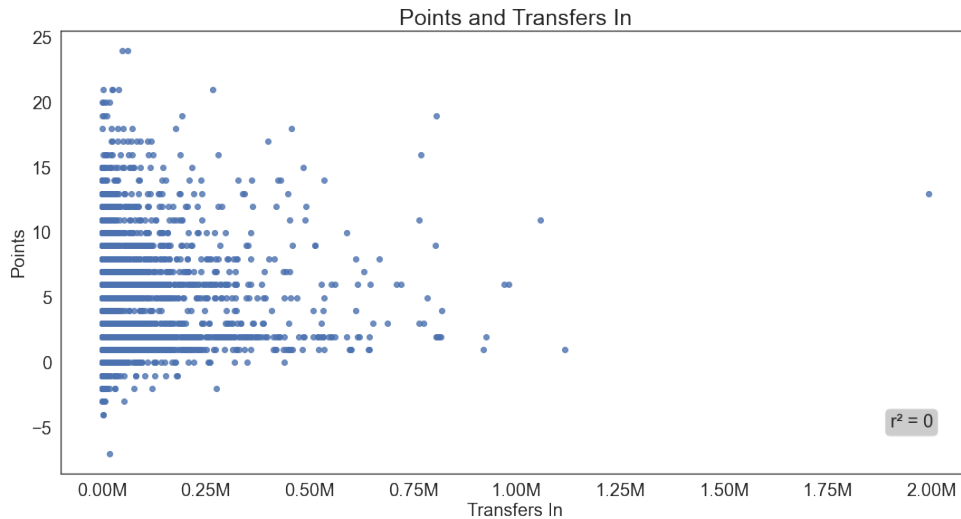


Figure 5.7: Scatter-plot of points and transfers-in for each player in each game week.

We had to find a way to work around this. The solution we came up with was the calculate z-scores of number of transfers-in for each player. This would allow us to see when players who aren't transferred-in often had game weeks when they were transferred-in a lot relative to individual players. This allowed us to create a new scatter-plot with the z-scores for transfers-in, in the most transferred players, to investigate the relationship between points scored and transfers-in.

```
#Calculating transfers z-scores
t_in = gws_info.reset_index(drop=True)
t_in = t_in.groupby(['name'], as_index=True)['transfers_in'].apply(
    lambda df: df.reset_index(drop=True)).unstack()
k = list(range(38,40))
t_in = t_in.drop(t_in.columns[k], axis = 1)
t_in = t_in.fillna(0)
t_in = t_in.T
t_in = t_in.apply(zscore,axis=1)
t_in.index += 1
t_in = t_in.stack().reset_index(drop=False)
t_in = t_in.rename(columns = {'level_0':'gw', 0:'z_score_transin'})
```

A line graph was then used to visualise the transfers-in and points scored by the top 20 most transferred-in players in each game week with regards to z-scores. This allowed us to easily see if a high amount of transfers-in indicated high points, or whether following transfer trends is a strategy that should be avoided.

After calculating the z-scores for number of transfers-in for each player, we decided to perform an anomaly detection to find the reasons why the players who scored highly after having a high number of transfers-in in a game week, were able to achieve the points that they did. This was done by taking all players who had a z-score for transfers-in of greater than 2 who also scored 6 or more points and then investigating their performances, focusing in on opponent strength and points distributions.

### 5.3.2 Results

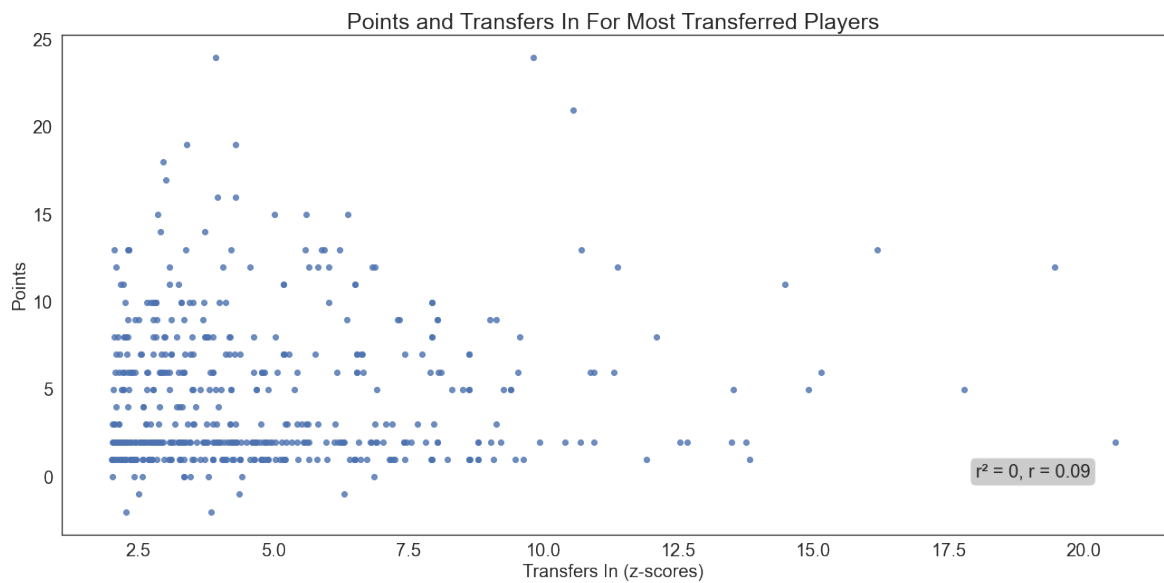


Figure 5.8: Scatter-plot of points and z-scores of transfers-in for the 20 most transferred-in players in each game week. Includes  $r^2$  and correlation coefficient.

From figure 5.8 we can see the relationship between high transferred-in players and the points they scored. The graph clearly shows no relationship between both transfers-in and points, and the  $r^2$  value (0) and correlation co-efficient (0.09) back this up. The following figure 5.9 investigates this further.

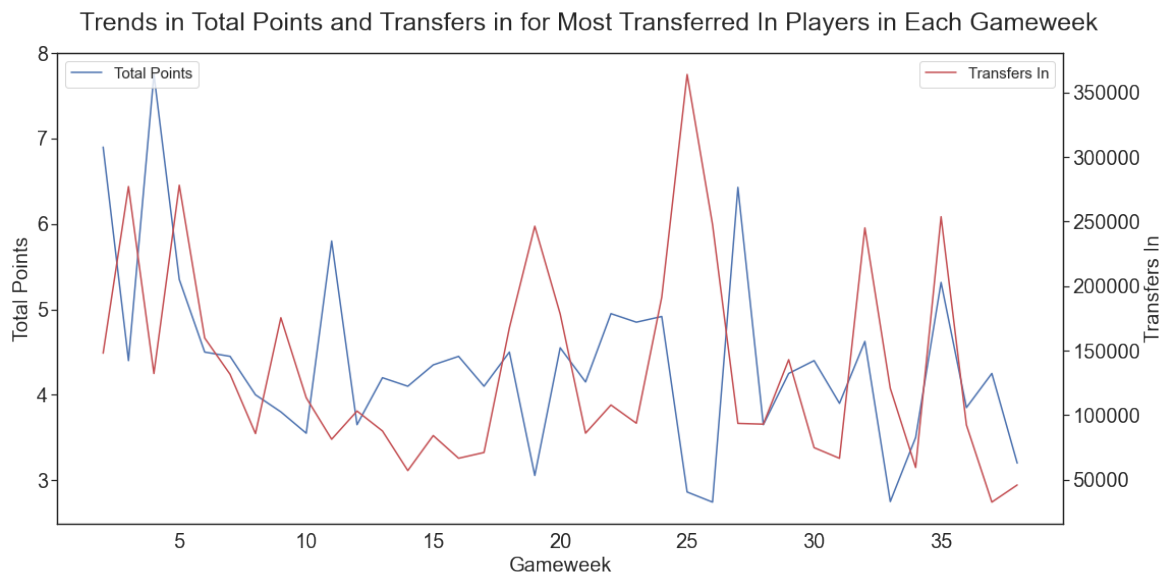


Figure 5.9: Points scored and number of transfers-in for the 20 most transferred-in players in each game week.

Figure 5.9 delves further into the relationship between transfers-in and points. It takes the top 20 most transferred-in players in each game week and shows the mean points and mean transfers for each as separate lines. Ideally, if high transfers-in were to imply high points, we would see the peaks of each line match up in each game week, however most peaks are mirrored by a dip, showing that high transfers-in don't necessarily indicate high points. We also found that only 33%

of these highly transferred-in players scored high points in that game week, while 50% of these highly transferred-in players score 2 points or less.

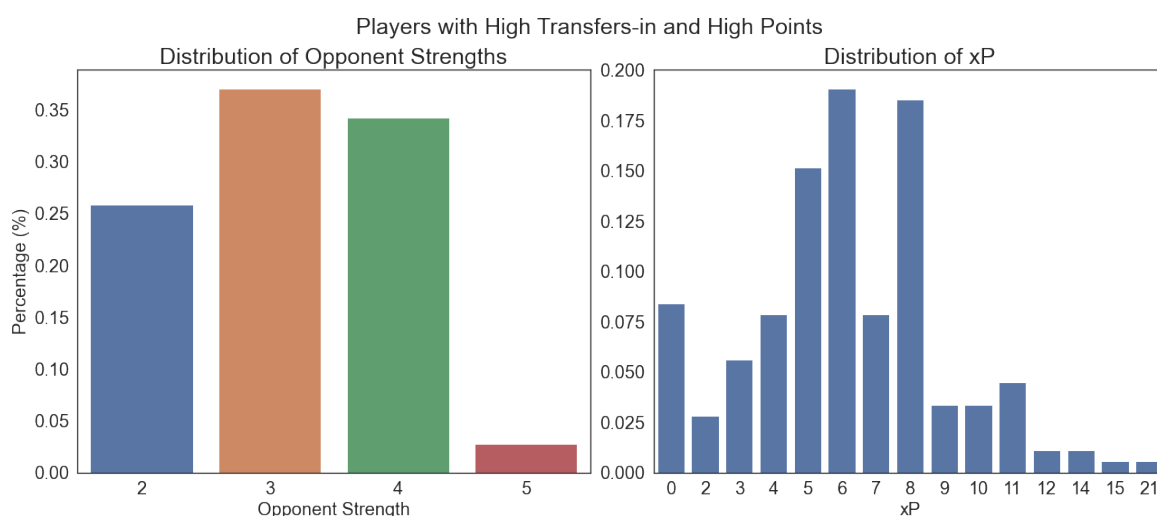


Figure 5.10: Opponent Strength and Expected Points distributions for players with high points ( $\geq 6$ ) and high transfers ( $z\text{-score} > 2$ ) in each game week. Only one team of strength 5 (Man City).

This figure 5.10 displays the findings of the anomaly detection. It looks into the opponent strength and expected points for the players that scored high points while having a high amount of transfers-in each game week, to try and gain an understanding as to why these players performed how they did. One thing that stood out is that these players played against Man City. The distribution of points scored between teams of strength 2, 3 and 4 doesn't vary that much. Surprisingly players who scored highly while also having a high number of transfers-in played less against opponents of strength 2. The xP distribution shows the majority of players who did well when transferred in had an xP of 5 or more.

### 5.3.3 Discussion

As in figure 5.8, we see that there is no relationship between transfers-in and points scored. This suggests that we shouldn't follow the transfer trends of other managers, which is investigated further in figure 5.9. In this figure, we see that the peaks don't line up, showing again that there is no relationship between transfers-in and points scored. This indicates further that we shouldn't make transfers unless necessary due to things such as injuries, poor performers, non-players or suspensions. However, an anomaly detection was performed to see if there was any evidence of reasons for a good transfer.

Figure 5.10 shows us the strengths of opponents faced when a player was transferred in a lot and scored highly, as well as the xP these players had. A small percentage of these good transfers were against teams of strength 5 (Man City). This tells us that if we need to transfer a player in, it should be a player who is not playing against Man City. Also these good transfers mainly had an xP of 5 or more, so again, if we need to make a transfer, we should make sure to bring in a player with an xP of 5 or more points. As there is no real stand out results to back up why these transfers were good, we can conclude from our findings that we should not follow transfer trends of other managers, and we should only make transfers if necessary. However, if we are faced with an issue and need to bring in a player, we should make sure that this player is not playing against Man City, and also has an xP of 5 or more.

---

## 5.4 RQ4: Can we predict which players will perform in a certain game week?

An important part of FPL is picking your captain and using wildcards wisely. If we can predict the performances of our players in a game week it improves our chances of maximising points from our captain and wildcards. Different prediction methods will be used such as: multiple linear regression, neural network, previous game week points and expected points (xP) (provided by our data).

### 5.4.1 Data & Method

The data used was the cleaned gws20\_21.pkl and teams20\_21.pkl files containing information on each game week and also corresponding teams data, from which we used the opponent strength and team strength indexes.

Different methods for making predictions were explored when it came to picking ways to predict players performances in a given game week. A multiple linear regression and a neural network stood out as two adequate prediction models to use. We also came across ARMA and ARIMA models which we attempted to build, however they were too difficult to perfect and fit in the end due to the complications of the parameters needed. We also decided to use a player's previous game week points and their expected points (xP) to see how accurate these could be for making predictions. The xP for a player is already provided by the FPL data, and previous game week points were acquired by shifting the data for each player by a game week.

The neural network and multiple linear regressions were both global models so they were not fit to specific players, which could be done in future work. Both of these models were trained and tested on an 80/20 train-test split. The inputs for both were Opponent Strength and Previous Game Week Points, with an output of Total Points. The neural network contains a single layer.

As well as just general points prediction, we investigated the accuracy of points prediction when it predicted 6 or more points (high points) in a game week. This was done as high point predictions are extremely important for managers when it comes to picking captains and playing your chips. We used the xP provided, previous game week points and a neural network for this. Again, the neural network is a single layered global network trained on an 80/20 split. For previous game week points we tested the prediction of high points for when a player scored 5 or more points in the previous game week. For xP, if the xP was 5.5 or greater, we tested the prediction of high points.

The accuracy of all of these predictions models were explored and graphed on bar charts.

```
#Multiple Linear Regression model
x = linear.drop('total_points',axis=1)
x = x[['opponent_strength', 'prev_points']]
y = linear['total_points']
x_train, x_test, y_train, y_test = train_test_split(x, y,
                                                    test_size = 0.2, random_state = 42)

LR = LinearRegression()
LR.fit(x_train,y_train)

y_prediction = LR.predict(x_test)
y_prediction = y_prediction.round(decimals = 0)
```

---

```

#Basic Neural Network model
x = nn_data.drop('total_points',axis=1)
x = x[['opponent_strength', 'prev_points']]
y = nn_data['total_points']
x_train, x_test, y_train, y_test = train_test_split(x, y,
                                                    test_size = 0.2, random_state = 42)

x1 = tf.convert_to_tensor(x_train)
y1 = y_train.to_numpy()

model1 = Sequential()
model1.add(Dense(2, input_shape=(2,)))
model1.compile(loss='mean_squared_error',
               optimizer='sgd', metrics=['accuracy'])
model1.fit(x1, y1, epochs=20, batch_size=10000)

```

## 5.4.2 Results

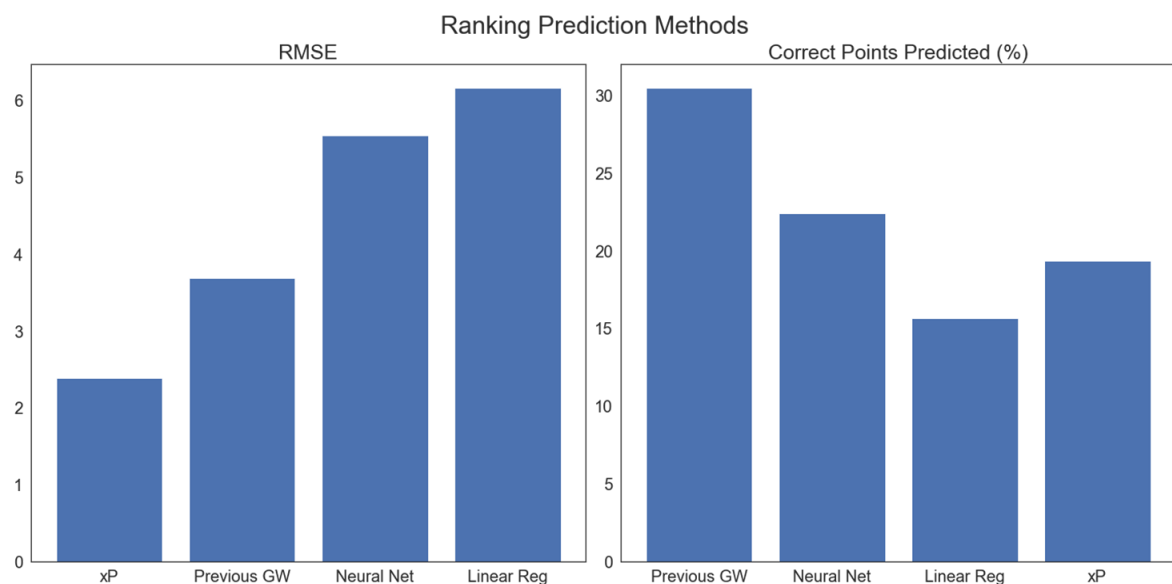


Figure 5.11: Ranking prediction methods based on root mean square error and accuracy of correct points predicted.

The above figure 5.11 shows us how our prediction models performed with regards to root mean square error (RMSE), and the percentage of times it predicted a player's points correctly. The RMSE values are quite high, ranging from roughly 2.4 to 6.1. This shows us that the error in our predictions are roughly 2.4-6.1 points off each time. Expected points (xP) is the best performer with regards to RMSE, however it still isn't particularly accurate. Our correct points prediction accuracy's range from 15%-30%. A player's previous game week's points is the most accurate points predictor, being correct 30% of the time.

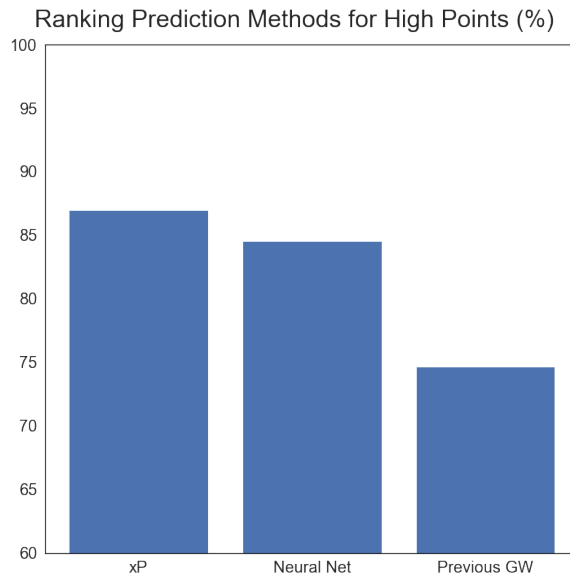


Figure 5.12: Ranking prediction methods based on accuracy of high points predicted.

Figure 5.12 ranks the accuracy of different methods of predicting high points in a game week. We classified high points as 6 or more points. The xP of a player is the most accurate, predicting high points 87% of the time, while the neural network was the 2nd most accurate with an accuracy of 84%.

### 5.4.3 Discussion

The takeaway from exploring these different prediction methods is that it's actually very difficult to predict how a player will perform in a given game week as shown in figure 5.11, where the lowest RMSE is above 2 and the most accurate we can get with predicting a players exact points is 30%. This is down to the high variance in the data from game week to game week.

However, we made one key discovery when it came to predicting high points for a player in a given game week as in figure 5.12. Expected points was correct 87% of the time when predicting that a player would score 6 or more points. This would be particularly important to FPL players when it comes to picking their captain for the game week, or even when considering the use of chips as the player will score 6 or more points almost 9 times out of 10.

---

## Chapter 6: Conclusions

---

### 6.0.1 Use of Chips

We put together this section based on our previous knowledge of the game, as well as our findings from our research.

- Triple Captain - Use when an elite, high valued player has a double game week, and high expected points in both games.
- Wildcard - Use when you have lots of injured players, lots of difficult fixtures or lots of chosen players are under-performing.
- Bench Boost - Use when most of the 15 players in the team have high or relatively high expected points in a game week. Could be especially useful in a double game week.
- Free Hit - Use when a lot of your players are non-players in a game week, due to injuries, suspensions or no fixtures.

### 6.0.2 The Strategy

All the findings from our research questions were used in order to help us build our strategy for maximising points in Fantasy Premier League.

- Use 3-4-3 as your formation - as seen in our research, 3-4-3 is the best performing formation.
- Choose your midfielders wisely - they are capable of scoring extremely highly, and the more expensive midfielders tend to perform well.
- Pick expensive players first, then fill in the squad with bargain-bin players - as the most expensive players score the highest points.
- Most of the budget should go towards midfielders and forwards - the more expensive midfielders and forwards score the highest.
- Have 3 Manchester City players in your team - they score the most points of all team strengths.
- Pick players mainly from teams of strength 4 - they score highly when compared to player5s from teams of strength 2 and 3.
- Do not make transfers unless you have to (injuries, suspensions, not playing) - there is no relationship between an increase in points and making a lot of transfers.
- If you do make a transfer, do not transfer in a player who is playing Man City, and make sure they have an xP of 5 or more - this maximises the points return from the transfer made as in figure 5.10.
- Pick your captain based on high xP (5.5 or more) - they score 6 or more points 87% of the time.
- When using chips, consult the Use of Chips section 6.0.1.

---

### 6.0.3 Future Work

There are many aspects of this project we could explore in future work. We could delve deeper into the area of prediction models and work to construct a much more accurate model. Transfers could be analysed further to find out what makes a good transfer. And finally, more work could be done on picking out players in line with the budget.

With regards to prediction models, we could work more on the ARIMA models and see how accurate these can get. The neural network could also be trained using more data and fit to individual players. More inputs could also be used for the neural network such as a moving average of the player's points or their team's form. These improvements as future work could really enhance our FPL game-play.

When looking into transfers, we could look more deeply into why some transfers were good, and why transfers were bad. This would allow us to add to our strategy. This could be difficult to do as there was no real data to show us as to why some transfers were good and some were bad.

To pick out players in line with the budget we could potentially create a team picking simulator that simulated team picks using up the majority of the budget. This simulator would be fed in restraints based on different calculations made for players and eventually produce an optimal starting squad for the FPL season.



---

## Chapter 7: **Acknowledgements**

---

Major thanks to Barry Smyth, Eoghan Cunningham and Ciara Feely for their help and suggestions along the way. We would also like to thank Fareed Idris for helping out with Kreoh. Finally we would like to thank our classmates for the ideas and support they provided along the way.

---

# Bibliography

---

1. surya. FPL and Fantasy Football History. *AllaboutFPL*. <https://allaboutfpl.com/2020/05/fpl-history-and-origin-of-fantasy-football/> (May 2020).
2. <https://www.premierfantasytools.com/whats-my-all-time-rank/>.
3. [https://myfplanalysis.co.in/top10k\\_random](https://myfplanalysis.co.in/top10k_random).
4. <https://allaboutfpl.com/fpl-recommended-sites/>.
5. <https://www.fantasyfootballscout.co.uk/>.
6. <https://twitter.com/OfficialFPL>.
7. <https://twitter.com/BenCrellin>.
8. <https://twitter.com/FPLGeneral>.
9. <https://twitter.com/BenDinnery>.
10. [https://twitter.com/LetsTalk\\_FPL](https://twitter.com/LetsTalk_FPL).
11. <https://www.youtube.com/watch?v=LzEuweGrHvc>.
12. Gupta, A. Time Series Modeling for Dream Team in Fantasy Premier League. *CoRR* **abs/1909.12938**. arXiv: 1909.12938. <http://arxiv.org/abs/1909.12938> (2019).
13. <https://www.frontier-economics.com/media/3962/how-to-win-your-fantasy-football-league-by-our-data-scientists.pdf>.
14. <https://www.beatingbetting.co.uk/fantasy-premier-league/beginners-guide/>.
15. <https://github.com/vaastav/Fantasy-Premier-League>.

---

# Appendices

---

## Appendix A – Individual Contributions

- RQ1: How does formation play a part in overall points achieved? - Matthew
- RQ2: How to make best use of the budget? - Philip
- RQ3: How transfers can affect a fantasy teams points? - Philip
- RQ4: Can we predict which players will perform in a certain game week? - Philip

## Appendix B – Datasets & Notebooks

### Datasets Used

We mainly used datasets from the 2020/2021 season, however 2021/2022 data was used aswell as 2020/2021 data when building prediction models. For each dataset we had separate folders: 20\_21 and 21\_22 from which we used these datasets:

1. gws – the main dataset for all research questions, containing information on players performances. this dataset is a folder containing separate .csv files for corresponding game weeks (gw1.csv, gw2.csv etc.).
2. player\_idlist.csv - data set containing player names and their corresponding player id. Used for preparing gw data.
3. teams.csv - Used for RQ2, RQ3 and RQ4. It contains information on teams, most importantly their different strength indexes.

### Notebooks Used

1. (PB) 0000\_cleaning\_player\_idlist.ipynb - Cleans and prepares player ids to be combined with game weeks data.
2. (PB) 0100\_cleaning\_preparing\_gws.ipynb - Creates two separate dataframes for the 20/21 and 21/22 seasons containing all game weeks in each.
3. (PB) 0200\_cleaning\_teams.ipynb - Cleans and prepares teams data, removing unnecessary columns.
4. (PB) 0300\_cleaning\_fixtures.ipynb - Cleans fixtures data.
5. (MoD) 1000\_rq1\_the\_best\_formation.ipynb - Investigates RQ1.

- 
6. (PB) 2000\_rq2\_use\_of\_budget.ipynb - Investigates RQ2.
  7. (PB) 3000\_rq3\_transfers.ipynb - Investigates RQ3.
  8. (PB) 4000\_rq4\_performance\_prediction.ipynb - Investigates RQ4.

## Appendix C – RQ1 Results



Figure 7.1: Pie chart showing the percentages of players from each position in the top 20 points scorers for the entire season.

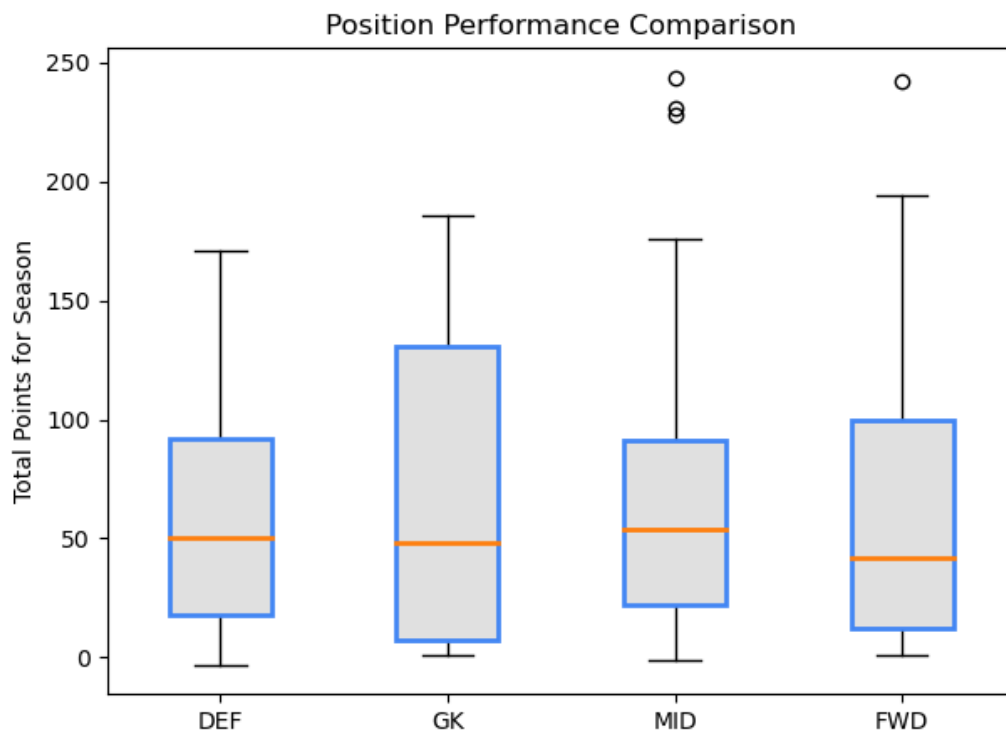


Figure 7.2: Box Plot showing the distribution of the sum of total points (across the 38 game week season) for each player, separated into positions.

## Appendix D – RQ2 Results

### Potential Bargain Bin Players in Each Position

Best Players from 2020/2021 Season Based on ROI

name	position	team
Emiliano Martínez	GK	Aston Villa
Illan Meslier	GK	Leeds
Sam Johnstone	GK	West Brom
Alphonse Areola	GK	Fulham
Hugo Lloris	GK	Spurs
Edouard Mendy	GK	Chelsea
Ederson Santana de Moraes	GK	Man City
Bernd Leno	GK	Arsenal
Lukasz Fabianski	GK	West Ham
Nick Pope	GK	Burnley

---

name	position	team
Stuart Dallas	DEF	Leeds
Matt Targett	DEF	Aston Villa
Vladimir Coufal	DEF	West Ham
Lewis Dunk	DEF	Brighton
Aaron Cresswell	DEF	West Ham
Michael Keane	DEF	Everton
Aaron Wan-Bissaka	DEF	Man Utd
Ezri Konsa Ngoyo	DEF	Aston Villa
Ezgjani Alioski	DEF	Leeds
John Stones	DEF	Man City
Tyrone Mings	DEF	Aston Villa
Ben White	DEF	Brighton
Rúben Santos Gato Alves Dias	DEF	Man City
Matthew Lowton	DEF	Burnley
Rob Holding	DEF	Arsenal
Luke Shaw	DEF	Man Utd
Benjamin Chilwell	DEF	Chelsea
João Pedro Cavaco Cancelo	DEF	Man City
Luke Ayling	DEF	Leeds
Tosin Adarabioyo	DEF	Fulham

---

name	position	team
Jack Harrison	MID	Leeds
Tomas Soucek	MID	West Ham
Matheus Pereira	MID	West Brom
Ilkay Gündogan	MID	Man City
James Ward-Prowse	MID	Southampton
Raphael Dias Belloli	MID	Leeds
Heung-Min Son	MID	Spurs
Bertrand Traoré	MID	Aston Villa
Leandro Trossard	MID	Brighton
Jarrod Bowen	MID	West Ham
Jorge Luiz Frello Filho	MID	Chelsea
Pierre-Emile Højbjerg	MID	Spurs
Bruno Miguel Borges Fernandes	MID	Man Utd
Ademola Lookman	MID	Fulham
Stuart Armstrong	MID	Southampton
Eberechi Eze	MID	Crystal Palace
John McGinn	MID	Aston Villa
Bukayo Saka	MID	Arsenal
Pedro Lomba Neto	MID	Wolves
Youri Tielemans	MID	Leicester

---

---

name	position	team
Patrick Bamford	FWD	Leeds
Ollie Watkins	FWD	Aston Villa
Che Adams	FWD	Southampton
Chris Wood	FWD	Burnley
Dominic Calvert-Lewin	FWD	Everton
Harry Kane	FWD	Spurs
Callum Wilson	FWD	Newcastle
Christian Benteke	FWD	Crystal Palace
David McGoldrick	FWD	Sheffield Utd
Jamie Vardy	FWD	Leicester
Michail Antonio	FWD	West Ham
Kelechi Iheanacho	FWD	Leicester
Danny Welbeck	FWD	Brighton
Bobby Decordova-Reid	FWD	Fulham
Neal Maupay	FWD	Brighton