**Fantasy PL Project 2021-2022**

**Ethan Mitten**

**June 27th 2021**

Today is the start of the Fantasy Premier League Analytics Project. I have found an amazing source for all of the statistics of Premier League data over the last five years. I have created a GitHub repository that credits the repository with all of the data I will use in this analysis. I do not plan to do anything with this, but do it for fun and for my own enjoyment as a project to do on the side as I love watching soccer. I am going to stay consistent with this blog to keep an account of what has been going on with the project. My stopping point today is setting up my GitHub and starting this blog and doing a little rough draft outline in Notes to set myself off on the right foot. I am also after I finish up this blog going to go ahead and make a database in mySQL for all of the data I will be using.

**June 29th 2021**

I have been working on importing all the data into mySQL and have been running into some issues. I have tried doing a lot of data cleaning in Excel and have been running into issues. It only wants to import the first 250ish rows of my data when I have 750 rows. I am going to keep working on troubleshooting this issue, but I do not want to just avoid SQL because importing is causing trouble. SQL will be vital for looking at all this data across all of these tables, so this is a very crucial step that I need to work through. The next approach is to see if I can export my Excel file into a text file and see if SQL likes that better.

I am not happy with Caglar Soyuncu at the moment! I was searching up troubleshooting problems for the last couple days and found that the reason my import was not working correctly is because I did not change over Caglar yet. Part of the cleaning process was going in and altering the accents on player’s names and I thought I got all of the names, so was starting to get frustrated with importing until I started importing another dataset and it did not work again, so I went back through and manually looked through player names and finally found that Soyuncu still had an accent on his name. Problem fixed and all the data for the 2020-2021 season has now been imported into mySQL. That is definitely enough for today.

**June 30th 2021**

Not a lot of time invested into the PL project today. I want to make sure I have everything planned out for when I start analyzing so I am not just blindly making analysis without a story/goal. I need to direct focus towards statistics that will directly affect score. This, as layed out in my notes, could be anything from clean sheets for a keeper to who takes penalties for a team. These are important statistics and picking the starting squad weighs heavily on the future outcome. There is the wildcard to dump the whole team, but we obviously want to get off to a good start.

**July 2nd 2021**

Today I am looking over some general information for the 2020-2021 season and will record them in a separate document. I got through the most common statistics such as goals, assists, saves etc. What I have found is that the teams that did not perform that well relatively had the goalkeepers with the most saves. Furthermore, Harry Kane seemed to be the most involved player for getting goals in the back of the net. The other noteable player for goal contributions was Bruno Fernandes. Next time when going through the general information I would like to look through penalties because that is a big part of scoring goals and getting points. I’ll need to find some data on made penalties because there is not any in the tables that I have. Next time I would also like to go through and start looking at the statistics specific to fantasy points such as most fantasy points and influence.

**July 10th 2021**

Started up the Premier League project again. Started doing more of the general statistics just for a little bit. Fantasy Premier League is about budgeting as much as it is about getting the players who will get the most points and with that I have an idea to do some scraping of my own. I do not have a price column in my dataset so I am going to be researching and getting some price info on all the players.

**July 11th 2021**

I am currently going through updating my database so that I have player prices and can divide a player’s total points by their price to get informative data on who were the best bang for buck players and then bang for buck players at each position. I was stuck on read only mode and was trying to figure out how to get into a mode I could edit my table in mySql. I finally hovered over the info icon next to Read only and realized I needed to set a Primary Key for my table. I also made a delete statement for the players on teams that have been relegated as it obviously does not do anything for me to know information on player’s who I cannot select next year.

**July 12th 2021**

Today was a productive day as the 2020-2021 season general information has been completed. We went through yesterday and added a price to all the players and got to see good budget players and also a bunch of different statistics to determine who were the solid players of the season. Although a lot can happen in terms of transfers, new managers, etc. and that can shape a player’s outlook just as much as the previous year, so when I list some players these are considerations just based on general information from the previous season. It will still have to be weighed out with historical data and some user judgement (although not a lot as using data analytics to select the squad is the most important part of this project). From our general information we definitely had some stand out players. They were:

* Harry Kane
* Bruno Fernandes
* Stuart Dallas
* Patrick Bamford
* Jesse Lingard
* Emiliano Martinez
* John Stones
* Heung-Min Son
* Mohamed Salah

Other Players that were also shocking or intriguing through the general information process were:

* Jack Harrison
* Matt Targett
* Ollie Watkins
* Callum Wilson
* Ben Chilwell
* Illan Meslier
* Aaron Cresswell
* James Ward-Prowse

Today marked the end of a very productive phase of the project which was looking at last season’s general information. From here we are going to look at some historical trends and then we will be about ready to make some candidates for our starter squad.

**July 13th 2021**

Today’s work was spent on preparing and outlining the next step of our process. The next step we will want to cover is the relationship between players and the teams they were on. Does being on a better team lead to higher total points? Does having better teammates lead to higher points? These are the kind of questions that are being aimed at answering. Tomorrow we will go ahead and start that process in SQL. Once we get the code down for one of our queries it will be relatively easy for the rest of them. A lot of GROUP BY will be involved.

**July 14th 2021**

I went and started work on the player’s relation to team query. For that I needed to modify my spreadsheet for the teams. I went ahead and updated the wins, losses, draws and placement of all the teams before running queries. In SQL I ran some queries for total fantasy points by team and then by position. Some interesting results out of this were that total points seems to be fairly related to how a team finishes. Burnley who finished 17th had the highest total points for attackers. The midfield seemed really related to the performance of the team as 8 out of the 10 top finishing teams had midfields in the top 10 for total points and the teams that finished in the top 3 aligned respectively for midfield points (Man City first, Man Utd second…). Chelsea’s defense acquired them a lot of points this year which was a little surprising. It was also just surprising that Chelsea finished 2nd in a lot of these rankings, but finished 4th in the league whereas Manchester City finished first in a lot and won the league. This might lead to some important decisions about how the structure of my fantasy team will be picked. It looks as if I should be picking my midfield based on teams I believe will perform the best and not necessarily completely off individual performance.

**July 16th 2021**

Today’s session was not as informative as I was hoping for. However, if I had not done the process I did I may not have understood that statistically I have not found a lot there. Today’s look was at the top 10 players and how much teammates contributed to their success. This was done by looking at teammates goals, assists and total points to get a small understanding of their contribution to the team. I was not able to find a lot of trends and these stats varied from player to player. For example, Kane and Son were the only big contributers at Tottenham and both had a lot of points, but at Manchester United there were multiple contributers alongside Bruno Fernandes. In conclusion, today was informative in the respect that these statistics are not super helpful in deciding whether teammates make a difference for a player.

**July 17th 2021**

I like the direction things are going for this research now. We have done the preliminary statistics research and seeing how players are doing and now are going to start analyze trends in Tableau. This piece is going to be very informative on putting the rest of my side together. We were able to determine that ICT index is actually a pretty important measurement for determining total points and with that one player that has stood out in a team that stood out is Ollie Watkins. I now feel confident about three of my selections. As I get a little further into this I will need to be careful about salary. For now the current salary we have given out is:

Bruno Fernandes -> 12.0 (100-12 = 88)

Ollie Watkins -> 7.5 (88-7.5 = 80.5)

Emiliano Martinez -> 5.5 (80.5 -5.5 = 75)

A picture containing text, indoor, green, colorful

Description automatically generated

**July 24th 2021**

At this stage something of great importance is not comparing different positions to each other, but finding trends within specific positions. We know that midfielders are the ones that are collecting the most total points and we want to select midfielders from the higher placed teams. We need more information crucially on what defenders I need to pick and what are distinguishing defenders in terms of points. John Stones from the data is a stand out and has been a stand out in some of the other analytics we have run so John Stones will be added to our side as well. For the rest of the players we will need to make a shortlist of good candidates and then watch how the transfer window pans out. We are selecting our team August 10th to give teams enough time to set up with who they want, but we need to see from the data a list of the best choices.

Influence seems to be something that is crucial for finding a good defender as the trend line is positive almost to a one slope surprisingly. Even after some really quality graphs looking at different trends of total points for defenders nothing really stands out too much to me which is a little disappointing. We will need to keep working on finding patterns within our data and see if we can make a break through. The fun part of this analytical work, detective work. I am going to go back on my earlier statement and I do not want to select John Stones just yet because as I was doing all of these graphs on trends for total points I was not seeing John Stones very involved in the top except for points per game which tells me that if he plays more this year he would be solid. We might have to see how Manchester City’s back line is looking before we decide anything. For now our selections stay the same. One thing that worries me about our picks however is the news of Grealish transferring away from Aston Villa. From our statistics Grealish had tremendous influence on the game and it is interesting to see how much that would affect Watkins and Martinez. Bruno is still an untouchable, but feeling a little more hesistant than before with the news of Grealish to Manchester City.

A picture containing text, indoor, green, colorful

Description automatically generated

**July 25th 2021**

Today is kind of a data to step away from the data side and work with what salary allows me to do. I am going to go through a couple mock teams and build some combinations.

Graphical user interface

Description automatically generated

Here we have kind of a massive amount of star players with a decent side around them. We were able to get Watkins, Fernades and Martinez like we want and also were able to add Salah and Kane as well. If Kane goes to Manchester City he is a must have with the creative side that Manchester City have. We are still able to get quality, but lacking a little bit with Origi as our third striker and none of our defenders are real point machines.

Graphical user interface

Description automatically generated

In this edition the deletion of Salah led to a spread in the wealth with the addition of Calvert-Lewin, Mount and Chilwell who all statistically have indications they will continue to contribute.

Graphical user interface

Description automatically generated

With the next team above we brought in one cheap player for each group, so we could have just a little bit more money to spend on a midfielder. Although Gundogan was the one that was picked the increase of funds to just 7.5 makes the options tough as the difference in statistics between the 7.5 and below players are hard to distinguish which dampens the thought of going with something like this. It is interesting to see where salary allows me and I really would like the idea of working Salah, Fernandes and if he goes to Manchester City, Kane into the side because that is consistent almost statistically guaranteed great point return.

**July 28th 2021**

Yesterday was a good day for this project as I went in a great direction in terms of the analyzing what I needed. I decided to break up my player data by positions in separate spreadsheets. So I duplicated a modified spreadsheet and made a goalkeeper spreadsheet etc. From there I took my spreadsheets into Spyder and then created a correlation matrix to see what was correlating the most to total points for each position. Today I will go ahead and analyze the results and do a little write up here about what the results mean.

For the goalkeepers correlation matrix… For total points we see that some parts are very highly correlated to total points which is an encouraging lead to follow. Surprisingly the most correlated is the bps. Bps is simply a measurement of the amount of bonus points that a player obtains. It is interesting that bps seems to be more impactful than saves and clean sheets. Another important factor is minutes. Overall, we should do one final analysis for goalkeepers based on who played the most minutes and who had the highest bps. Also we will want to check out influence as is it high on the list as well. (**BPS, Minutes, Influence)**

For the defenders correlation matrix… There are two spots of high correlation. Again surprisingly we see that BPS is highly correlated to Total Points. The other highly correlated statistic is influence. Other factors include clean sheets and minutes. (**BPS, Influence, Minutes)**

For the midfielder correlation matrix… we have two standout features. The first being influence and the second being the ict index statistic. Interestingly we do not see as many highly correlated attributes in this matrix as the other ones. One other that is pretty high as well is BPS again. (**Influence, ICT\_Index, BPS)**

For the forward correlation matrix… we have a lot of attributes that are fairly highly correlated to total points. We will try to weed out the most relevant ones though. The first being BPS. Besides that the other two standout statistics are influence and ict\_index. (**BPS, Influence, ICT\_Index)**

What an interesting find. When a started this analysis a month ago I would have stated that having the most goals and assists is going to be the biggest source of points for players and although that may be true for like the top 5 players it is interesting to see the trend on what is considered important. We will need to go in depth and look at BPS, ICT\_Index and Influence for each position as they are considered very correlated to total points.

**July 29th 2021**

Today I went through and produced SQL queries for the attributes we were looking at yesterday and found some standout players from each position. We will continue this further in the coming days to see how a team could be put together to best put in as many of these players as. I can. The main standouts were Emiliano Martinez, Luke Ayling, Bruno Fernades and Harry Kane. Ayling is a very suprising one and he is quite cheap as well.