

# Fantasy Football Analysis and Suggestion of Player Transfer

Rohit Thomas  
Student Number: 16210779  
School of Computing  
MSc. Computing (Data Analytics)  
Dublin City University  
[rohit.thomas8@mail.dcu.ie](mailto:rohit.thomas8@mail.dcu.ie)  
Date of Submission: 22 August 2017

**Disclaimer:** A report submitted to Dublin City University, School of Computing for module CA685 Practicum, 2016/2017.

I understand that the University regards breaches of academic integrity and plagiarism as grave and serious.

I have read and understood the DCU Academic Integrity and Plagiarism Policy. I accept the penalties that may be imposed should I engage in practice or practices that breach this policy.

I have identified and included the source of all facts, ideas, opinions, viewpoints of others in the assignment references. Direct quotations, paraphrasing, discussion of ideas from books, journal articles, internet sources, module text, or any other source whatsoever are acknowledged and the sources cited are identified in the assignment references.

I declare that this material, which I now submit for assessment, is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work. By signing this form or by submitting this material online I confirm that this assignment, or any part of it, has not been previously submitted by me or any other person for assessment on this or any other course of study.

By signing this form or by submitting material for assessment online I confirm that I have read and understood DCU Academic Integrity and Plagiarism Policy (available at: <http://www.dcu.ie/registry/examinations/index.shtml>)

Name: Rohit Thomas  
Date: 22 August, 2017

# Fantasy Football Analysis and Suggestion of Player Transfer

Rohit Thomas  
School of Computing  
MSc. Computing (Data Analytics)  
Dublin City University  
[rohit.thomas8@mail.dcu.ie](mailto:rohit.thomas8@mail.dcu.ie)

**Abstract-**Fantasy League is an online simulation game where simulation is based on real game events and updates. This paper documents the collection (Web-Scraping of data and transformation), Analysis and Prediction of the Fantasy Premier League (English Premier League) based on the performances of the 426 Registered players. The implementation of the model is done in Python with an aim of Predicting the players for a user based on certain factors using Linear Regression for Analysis.

**Keywords:** -Correlation matrix, Prediction Model, Web-Scraping, Data Modelling, Data Gathering, Linear Regression, Fantasy Premier League.

## I. INTRODUCTION

Fantasy League is an Online Simulation game which features selection of players based on the common standard fund which each User of the Game is given. The Fantasy League now implemented on wide array of Sports and Leagues for example NFL, English Premier League, Indian Premier League(Cricket). Fantasy Premier League based on the English Premier League is one of the most played after Fantasy Leagues. All users get a Standard 100 Million to select and buy players who have registered on real time basis to play in the Premier League. 100 Million is then used by the user to plan, allocate and buy 15 Players and it's done strategically with a mix of Star Players and players who have been performing well recently. The user must select 15 Players which should comprise of 15 Players which should have 2 Goalkeepers, 5 defenders, 5 Midfielders and 3 Strikers. Captain and Vice-Captain is nominated for the team by the user. The Purpose of nominating a Captain and Vice-Captain is very significant as Points collected by the captain gets Doubled. The Vice-Captain is only active in the event in which Captain in the team is not able to play, then the points accumulated by the Vice-Captain becomes doubled.

One more Significant event that is part of the Fantasy Premier League<sup>[1]</sup> is the option of offering

users a transfer every game-week to change one of his squad players and selecting a new one. The following points system is awarded to the players for their performances in the game week:

Scoring	
During the season, your fantasy football players will be allocated points based on their performance in the Barclays Premier League.	
Action	Points
For playing up to 60 minutes	1
For playing 60 minutes or more	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

Fig 1. Fantasy Premier League Player Performance Points Awards Chart.

The Fantasy Premier League Dashboard looks like below:



Fig 2. Fantasy Premier League Dashboard

The main aim of this study is to use Data Analytics and Machine Learning to achieve below mentioned goals: -

1. Gather and Collect the Data on which Analysis will be done.
2. Understand the Data and Understand the Factors helping Prediction.
3. Analysis and Prediction helping in suggestion of players.

The model employed in this Data Analysis project is done on different player positions like Attackers, Midfielders, Defenders and Goalkeepers individually as for all the player positions the factors and features affecting vary from position to position.

For Instance, in case of Attacking players the points accumulated by them and the bonus scored by them depends mainly on the goals scored and assists created by them. The points scored by the Defender or goalkeeper on the other hand depends more upon the tackles and saves that they make.

This paper also explains the Python Web-Scraping of the Fantasy Premier League dataset from the premier league website. The data is extracted from the Fantasy premier league API. The data is iterated individually and populated for all the registered players in the premier league. The Data retrieved here is stored in JSON format which using Dictionary is converted into CSV format.

This paper deals with how graphical representations and visualisation using Tableau helped in gaining insight into the data. Excel and open refine were tools which were used for the data cleaning and retrieval of the necessary features. Python using Anaconda Prompt was used for the model development purposes. The Web scraping of the data from the API and the development of analysis models were achieved by using python.

In this paper, Data Analytics life cycle has been followed. This paper deals from collection of data, identifying key features and Analysis of model on the testing data.

## II. LITERATURE REVIEW

The main motivation behind the creation of this model is to suggest players on the basis of analysis and prediction of the training data and testing it on the future data (Testing data). The model will generate a system which predicts the points that will be scored by the players.

The following existing Literature helped in achieving goal like Collection of data using web scraping techniques to models and analysis done English premier league and other fantasy league data.

### A. The Fantasy Premier League Website

The Fantasy Premier Website<sup>[2]</sup> provides an interactive dashboard of all the results which has happened over the game week, points scored by players and the other information like the market value of the player and current availability status (Injured or suspended). This website houses information of all the player throughout all the game weeks and performance of the users in the current game week and in all the past weeks.

### B. PL Fantasy Blog

PL Fantasy blog<sup>[3]</sup> is an information dashboard created using tableau software. The dashboard is accessed to show the performances of the player. The dashboard provides a graphical interface to the user.

### C. Extract and Analyse Player Performance Data from the Fantasy Premier League Website

This paper<sup>[4]</sup> deals with the minute Analysis of Fantasy Premier League data on a Qlik dashboard. The Analysis and data collection and scraping gave a tremendous insight into the fantasy league domain and working and functioning of fantasy premier league. The Web scraping used in the paper also helped implementing the Web-scraping model in python and collect the dataset from the fantasy premier league website.

## III. OBJECTIVE AND HYPOTHESIS

### A. Objective

The main objective of the thesis is to predict the number of points that will be scored by the players in the Fantasy Premier League which thereafter will be used to Suggest the players to the users.

### B. Hypothesis

*H0*(Null Hypothesis): It is possible to predict the total number of points scored by the Player based on each position (Attacker, Midfielder, Defender, Goalkeeper).

*H1*(Alternative Hypothesis): It is not possible to predict the total number of points scored by the Player based on each position (Attacker, Midfielder, Defender, Goalkeeper).

#### IV. IMPLEMENTATION

##### A. Data Gathering

###### 1. Database Fantasy BootStrap Generator

The Database Fantasy bootstrap Generator <sup>[5]</sup> Generates the Fantasy Data and statistics. The data generated is consolidated statistics to the point of that season for all the players along with the Current Game-Weeks Figures.

###### 2. Web scraping from Fantasy Premier League

Web Scraping of the Data from the Fantasy Premier League involves collecting records of all the players. The URL provides information when added a number at the end of the URL <sup>[5]</sup>. the number added is the registered player number of player in the premier league. The Python code scripts and generates the statistics and data for all 426 players and uses Python functionality “DictWriter” to store all the values in a key which in turn is converted into a CSV file in the future.

##### B. Data Cleaning

###### 1. Feature Selection and JSON to CSV Generation (Python)

There was Data Cleaning involved as part of this paper. The Dataset generated by the Static API<sup>[6]</sup> for the fantasy Premier League generated a mixed and unnecessary information which was not to be part of the analysis phase. The JSON file was converted to a csv file using python functionality of CSV.Dictwriter and this CSV file was cleaned ahead of being used for the analysis phase. 150 redundant and unnecessary columns which was not to be used for analysis phase ahead was able to be removed. Google Refine and Excel were Applications that were used to perform the Data Cleaning Process.

###### 2. Filtering of Data According to Player Position

The Dataset has players from all the positions i.e. Attackers, Midfielders, Defenders, Goalkeepers. In order for the model to perform and function better, the dataset is filtered based on the position of the players. Therefore, analysis is done separately for Attackers and Midfielders. This Data Selection and Filtration process will help increase the accuracy and performance of the model. For Instance, in the dataset the players in the role midfielder were selected

and exported to a new csv in the format “Midfielders\_Gameweek.csv”.

The new Streamlined Dataset based on position was used for Analysis and Prediction phase.

##### 3. Removing the Fringe Players

On Observing the dataset in the data description phase, finding and identifying players who have played very less or not played at all led to the considerable difference in performance and functioning of the model. The Minutes Played feature in the dataset is the key feature used here to remove the players who have not played a single minute up till now.

The consequence to this feature being implemented is that “What if a star player out Injured is coming back to play?”. If the player in any point in season has played, the model remains unaffected and the player is still in consideration for the analysis. If the player has not played a single minute up till now then the model does not pick up the player for analysis for that upcoming week. The player would be available for analysis if starting that week would be available for analysis in the future weeks and the model functions properly.

##### C. Data Exploration

In the phase of data exploration, visualizations helped understand and study the dataset in a much broader perspective. It has been explored through below diagrams created using Python and Tableau:

###### 1. Mean total points distributed amongst the different clubs

The below visualization depicts the total points scored by players from the following different clubs. This Visualization helps us understand that players from Chelsea have scored more points than the other teams.

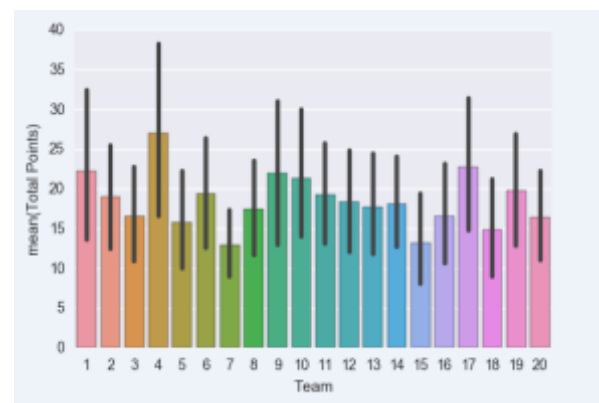


Fig 3. Club Id vs Mean Total points

## 2. Points per game earned by different player position types

The image below depicts the mean points scored per game by players from different position types:

Goalkeeper = 1  
Defender=2  
Midfielder = 3  
Attacker = 4

As seen from the image midfielders have scored maximum points per game.

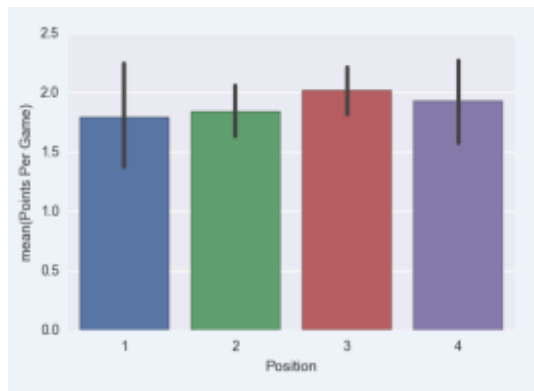


Fig 4. Distribution of player position vs mean total points.

## 3. Mean cost of the player based on player position

The visualization below depicts the mean cost of player based on the player position. The mean cost of an attacker as depicted here is the highest.

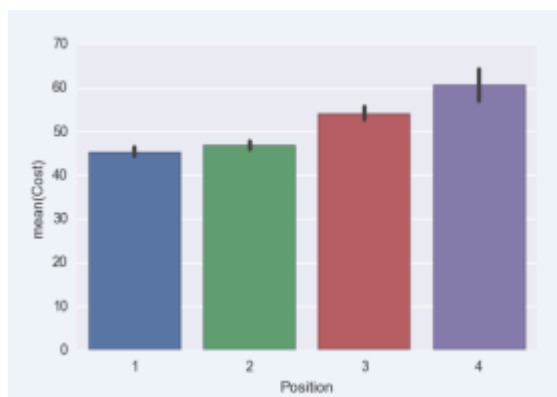


Fig 5. Distribution of player position vs mean cost of the player.

## 4. Points scored by dream team player to another player based on player position

The following visualization was a key in understanding the performance and points scored by a dream team player to that of a normal player. As the visualization clearly depicts that Dream team

player have scored much more points than compared to a regular player.

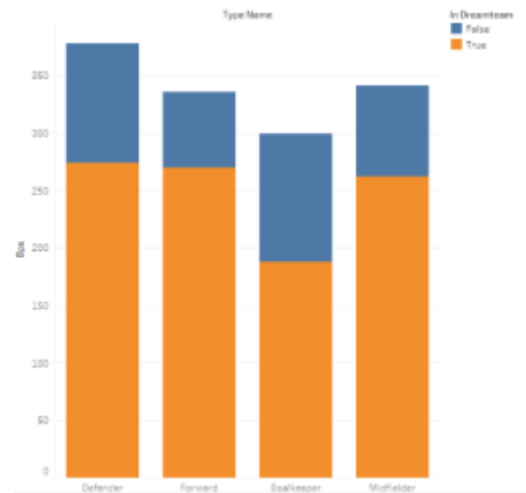


Fig 6. Bar graph plot depicting player position vs ratio of the bonus points scored by dream team player to another player

## D. Analysis and Prediction

The analysis used on the Fantasy League Dataset to determine which model is able to give better results was one of the looming questions. After study of prior papers, the model's Linear regression was chosen to train and test upon the data on. The output of the model which analyses the Fantasy premier dataset is Continuous so Linear Regression was chosen as the prediction model to be utilised for the prediction of the points of the players.

The splitting of the dataset into training and testing dataset was done using numpy's "randn" function. The dataset for the month of November was taken in as the training dataset and datasets for the December 1, December 4, December 9 game-weeks was taken as the testing dataset.

The Dataset, descriptive statics and the code for Web scraping and analysis on the fantasy league are available to access in the GitLab Repository<sup>[7]</sup>.

The Data Exploration and visualization which has been explained above and domain knowledge gained thorough the prior work done in this field helped in selecting below features for further exploration. The selected features were analysed in a correlation matrix to identify the collinearity amongst them.

Correlation matrix was analysed on different player type positions to understand the impact of correlation and multicollinearity amongst the features vs the player position type.

Variables	Column Description
<b>Now cost</b>	The cost of a player in a game week and this factor is affected by the form of the player.
<b>Selected by</b>	The percentage of user that have chosen / represented the player in their Fantasy League Team.
<b>BPS</b>	Its an Index denoting the ranking of the players on the basis of the Bonus points that player accumulates.
<b>Event total</b>	The points in total scored by a player in a week.
<b>Form</b>	A Parameter representing the current form of the player.
<b>Ea index</b>	Index created by the EA org. and used for ranking/comparing players, and its also changes based on the current performances of the player
<b>Value Season</b>	It's a parameter which which compared the form of the player in the previous rounds to the value of the player

Fig 7. Model Feature Description

The following Correlation Matrix shows a significant correlation between the above-mentioned features. Shown below are the correlation matrix generated for each of the player type positions.

The correlation matrix shows very high collinearity between EA index and bps. The collinearity of form and value form is also very high.

### 1. Attackers

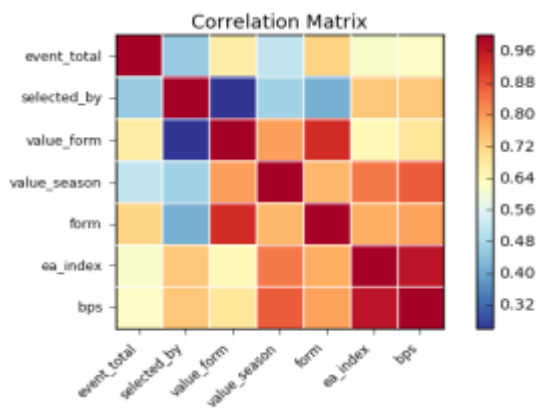


Fig 8. Correlation matrix for Attackers

### 2. Midfielders

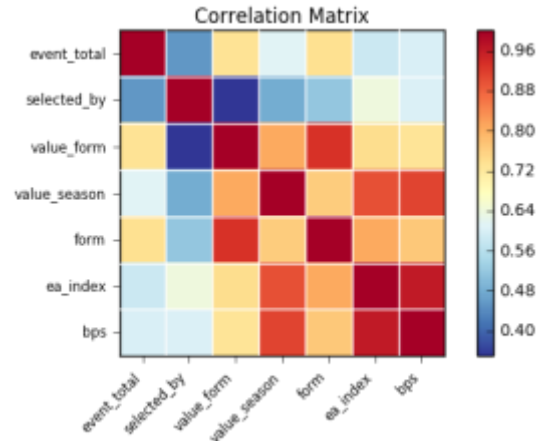


Fig 9. Correlation matrix for Midfielders

### 3. Defenders

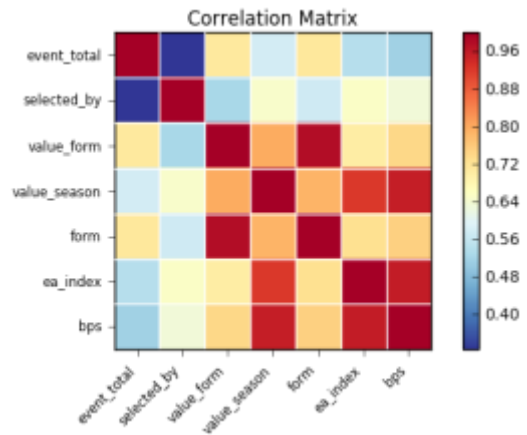


Fig 10. Correlation matrix for Defenders

### 4. Goalkeepers

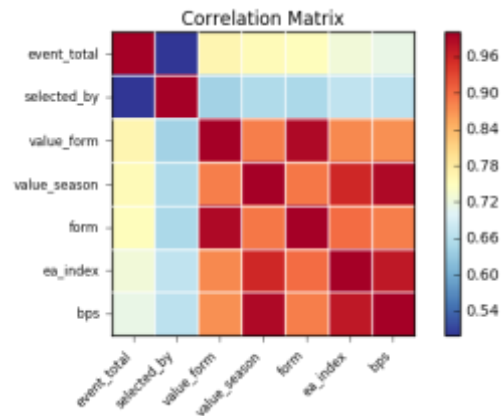


Fig 11. Correlation matrix for Goalkeepers

## V. RESULT

The Linear Regression model is analysed on the dataset on all the player positions. The model was separately deployed on Attackers, Midfielders, defenders and Goalkeepers based on the Features Value Season, BPS (Bonus Points Scored), Form, Event total, EA index.

The Model is trained on the November Dataset and is tested on the December 1, December 4 and December 9 weeks. The accuracy of the model is determined by the root mean square value:

$$RMS\ Error = \sqrt{\frac{1}{n} \sum_{i=1}^n (predicted_i - observed_i)^2}$$

The Linear Regression Model Generates the following model for the Fantasy Premier League Player Point Prediction with below mentioned RMSE:

TABLE 1. RMSE FOR EACH POSITION

Player Position	RMSE		
	December 1 Week	December 4 Week	December 9 Week
Attackers	2.3746	2.8351	3.0285
Midfielders	1.7222	1.907	1.4177
Defenders	2.4425	2.2145	2.40596
Goalkeepers	2.7991	2.1821	2.9641

The above table shows that the root mean square error for the midfielders is very less which means that the model has worked well for the Midfielders type of player position and predicts the points scored by the midfielders as accurate as possible.

The RMSE value for attackers is slightly on the higher side which means the model didn't work as well for the attackers, defenders and Goalkeepers. The accuracy of the RMSE generated by the model proves that the model is functioning well and

predicts the points scored by the players as accurately as it can.

## VI. CONCLUSION

The aim of the paper of the paper is suggest the players for selection to the user by prediction and analysis Fantasy Premier League Dataset. The analysis shows that the past week performances of a player does impact and influence the future predictions. The Web-scraping of the data from the API was implemented to help with easier access to analyse the data.

In future, the scope of this model can be extended to a web application with dashboard which can show live analysis and player suggestion functionality from a front-end point of view. The fine tuning of the analysis.

## ACKNOWLEDGEMENT

I am very grateful for the help and guidance provided by my supervisor Dr. Geoff Hamilton. The guidance of Dr. Geoff Hamilton helped me to gain quite a lot of insight into this new domain and help me implement and develop the model better.

## REFERENCES

- [1] Premier League, 2014. Fantasy Premier League Rules. [Online] Available at: <http://fantasy.premierleague.com/rules/> [Accessed 20 April 2014].
- [2] Fantasy Premier League, <https://fantasy.premierleague.com/>
- [3] Tableau, 2014. Tableau. [Online] Available at: <http://public.tableausoftware.com/views/PlayerPage/Dashboard1?:showVizHome=no> [Accessed 20 April 2014].
- [4] Gibbons, B., 2014. Project Proposal – “Extract and Analyse Player Performance Data from the Fantasy Premier league Website”, Dublin: s.n.
- [5] Premier League, 2014. Fantasy Premier League API. [Online] Available at: <http://fantasy.premierleague.com/web/api/element/s/180> [Accessed 25 February 2014].
- [6] <https://fantasy.premierleague.com/drf/bootstrap-static>
- [7] Gitlab DCU Code and Data, <https://gitlab.com/computing.dcu.ie/thomasm8/Fantasy-Premier-League>