

**National College of Ireland**

**Project Submission Sheet**

|  |  |  |  |
| --- | --- | --- | --- |
| **Student Name:** | Joseph Reddy | | |
| **Student ID:** | X22151508 | | |
| **Programme:** | Higher Diploma in Science in Data Analytics Information | **Year:** | 2 |
| **Module:** | Project (HDSDA\_SEPBL\_YR2) | | |
| **Lecturer:** | John Kelly | | |
| **Submission Due Date:** | 21/04/2024 | | |
| **Project Title:** | Can previous Fantasy Premier League data provide an optimised game plan for Fantasy Managers? | | |
| **Word Count:** | 4020 | | |

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the references section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

|  |  |
| --- | --- |
| **Signature:** | Joseph Reddy |
| **Date:** | 17/04/2024 |

**PLEASE READ THE FOLLOWING INSTRUCTIONS:**

1. Please attach a completed copy of this sheet to each project (including multiple copies).

2. Projects should be submitted to your Programme Coordinator.

3. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.

4. You must ensure that all projects are submitted to your Programme Coordinator on or before the required submission date. **Late submissions will incur penalties.**

5. All projects must be submitted and passed in order to successfully complete the year. **Any project/assignment not submitted will be marked as a fail.**

|  |  |
| --- | --- |
| **Office Use Only** | |
| Signature: |  |
| Date: |  |
| Penalty Applied (if applicable): |  |

AI Acknowledgement Supplement

# Project (HDSDA\_SEPBL\_YR2)

# Can previous Fantasy Premier League data provide an optimised game plan for Fantasy Managers?

|  |  |  |
| --- | --- | --- |
| **Your Name/Student Number** | **Course** | **Date** |
| **Joseph Reddy x22151508** | Higher Diploma in Science in Data Analytics Information | 17/04/2024 |

This section is a supplement to the main assignment, to be used if AI was used in any capacity in the creation of your assignment; if you have queries about how to do this, please contact your lecturer. For an example of how to fill these sections out, please click [here](https://libguides.ncirl.ie/useofaiinteachingandlearning/studentguide).

# AI Acknowledgment

This section acknowledges the AI tools that were utilized in the process of completing this assignment.

|  |  |  |
| --- | --- | --- |
| **Tool Name** | **Brief Description** | **Link to tool** |
| **N/A** | N/A | N/A |
|  |  |  |

# Description of AI Usage

This section provides a more detailed description of how the AI tools were used in the assignment. It includes information about the prompts given to the AI tool, the responses received, and how these responses were utilized or modified in the assignment. **One table should be used for each tool used**.

|  |  |
| --- | --- |
| **N/A** | |
| **N/A** | |
| **N/A** | N/A |

# Evidence of AI Usage

This section includes evidence of significant prompts and responses used or generated through the AI tool. It should provide a clear understanding of the extent to which the AI tool was used in the assignment. Evidence may be attached via screenshots or text.

# Additional Evidence:

[Place evidence here]

# Additional Evidence:

[Place evidence here]

***Can previous Fantasy Premier League data provide an optimised game plan for Fantasy Managers?***

Joseph Reddy   
*Higher Diploma in Science in Data Analytics*  
*National College of Ireland*x221151508  
x22151508@student.ncirl.ie

*Abstract*— With a near 11 million strong player base. Fantasy Premier League is the world's largest fantasy football competition. Highly competitive Managers look to gain any advantage available to them. This study looks at what advantages can be gained from analysing the data from previous seasons through statistical methods and machine learning. It was able to determine on average that the Goal Keeper was the highest scoring position; when to focus more budget on Forwards and less on Defenders; key game weeks for certain bonus chips to be played; and how well FPL’s ICT Index performs in predicting a player’s total points, and that a player’s Influence score is one of their key performance indicatiors.

Keywords—Fantasy Football, Statistics, Machine Learning.

# Introduction

Fantasy Premier League (FPL) is a fantasy football game with nearly 11 million players (Managers) worldwide. It has spawned a sub-culture of die-hard fans looking to gain any competitive edge over their league rivals. There are now countless Columns, Podcasts, Forums, and Social Media accounts, both officially partnered and non-official, that Fantasy Managers subscribe to, to gain insight on how best to manage their teams. Initially, the strategy may seem simple – pick the highest scoring players and you will accumulate the most points. However, with budgets and limits on players from single teams, as well as bonus chips, and fixture calendars to navigate, there is much more to consider. The Author has a relevant background with the game, having played for over 13 years.

The goal of this study is to provide Managers with unique insights on when is best to transfer in certain players, where best to utilise team budget, and optimise strategy on when to play their bonus chips.

Managers are given an initial budget of £100M to build a squad, containing 2 goalkeepers (GK), 5 defenders (DEF), 5 midfielders (MID) and 3 forwards (FWD). Player’s prices vary depending on previous season performances, and net transfers. Managers must spread this budget amongst the 15 players.

Across the season, Managers have several bonus chips they can play. Wildcards appear twice in the season; allowing Managers to make unlimited transfers at no cost. The Triple Captain chip triples the points the designated team captain scores in a game week. The Bench Boost chip includes the points scored on the bench to the game week points total. The Free Hit chip allows Managers to change their team for one game week.

Choosing when to transfer in certain players, and when to utilise the bonus chips can be the difference between Manager’s topping their mini-leagues or missing out on victory.[1]

The intended solution to this problem is to use previous FPL data to identify if there are certain positions Managers should allocate more of their budget to, and to suggest when in the season they should play their bonus chips. This will be done in three stages. Firstly, by performing a Kruskal-Wallis test to determine if there is a statistically significant score between the four positions. Further to this, post-hoc testing will be performed to identify which (if any) positions differ from one another.

Secondly by performing a Time Series Analysis with a Holt Winters Triple Exponential Smoothing model to try and pinpoint when in the season certain positions should be prioritised over others and to try and identify when will best suit managers to play their bonus chips. Triple Exponential Smoothing will be used because it considers trend, seasonality, and level.

FPL provides an ICT Index to guide managers with player selection. FPL defines it’s index as *“[An] index… developed specifically to give a verdict on a player as an FPL asset. It uses match-event data to generate a single score for three key areas: Influence, Creativity and Threat. These figures then combine to create an individual’s ICT Index score… It condenses more than 40 match event statistics into four distinct scores. These offer a view on player performance for factors that are known to produce FPL points.”* [2] FPL does not provide an underlying methodology as to how exactly this index is calculated other than the combination of the Influence, Creativity, and Threat statistics, which then appear to be divided by 10 for the players’ final ICT score. With the vague explanation of the Index, can managers trust this model to suitably aid them in their team selection? The third stage of the study will be to build a Multiple Linear Regression model to determine if a player’s total points can be accurately predicted by the components of the ICT Index, and therefore if this, arguably underutilised statistic, performs well, and provides managers with useful information on which players to target or avoid for team selection. Further to this, the three components of the ICT Index will be examined to advise managers which has the most proportionate effect, this will be done so they can chose between players with a similar Index score.

The dataset used in this study comes from Kaggle.com [3] It contains data on all players’ performance, ranging back to the beginning of the 2016-17 Premier League season to the 2022-23 season. The data for the 2018-19 and 2019-20 seasons does not feature in the dataset. The dataset includes all statistics that score points within the game such as Goals Scored, Assists, Bonus Points, Yellow/Red Cards etc. as well as other statistics that are intended to aid managers eg, Players’ Influence, Creativity, Threat.

There are no ethical concerns about this data, as although it does contain individual’s names, these are the names of professional players playing in an open public forum.

# RELATED WORK

Currently the scope of academic research into the use of machine learning and statistical modelling for Fantasy sports games is small, but there is a selection of studies that can help inform this one. Many studies focus on the NFL and other American Sports, but the lessons learned from these studies can be transferred as they all focus on accurately predicting either game outcomes or team optimisation. Most studies tend to use Multiple-Linear Regression, Decision Trees, and Random Forests as the most common models.

Bangdiwala et al. [4] have produced a study in this area focusing on predicting a fantasy player’s points using Multiple-Linear Regression, Decision Trees, and Random Forests. While Their study does claim to be successful in being able to accurately predict player’s FPL points, their methodology uses vague language in terms of how their models were constructed, e.g. exactly which dependant variables were used for their Multiple Linear Regression models, or which software was used, and focus more on providing a description of what the model’s are. There also appears to be some confusion with the conclusions in this report as they suggest that the Multiple Linear Regression model provides the best results, however their own figures dispute this.

Rajesh et al. [5] also use Machine Learning to name which specific players Managers should select for their teams, they implement a Random Forest model as well as a Gradient Boosting and are able to produce some well-performing models. However, their research appears to be somewhat redundant as they allow for two scenarios which don’t adhere to the rules of the game. Their first configuration adheres to spending/team selection limits, whereas their second (team selection limits ignored) and third (spending limits ignored) configurations outperform the first. Essentially stating that if managers could not follow the game’s rules, they would score more points – not a particularly helpful statement.

Random Forests seem to be the most used model amongst researchers. They implemented by Yang [6], alongside other machine learning models such as Naïve Bayes, K-Nearest Neighbour, and Support Vector Machines using similar data from Fantasy Premier League to predict real life match results based on starting lineups, achieving up to 82% accuracy on early stages of the season.

Other relevant studies include Beal et al.’s [7] which also aim to optimise fantasy sports teams, but for daily fantasy sports, which are typically played over much shorter time periods than FPL and provide successful players with financial reward. Their study focuses on NFL based fantasy games, and uses Linear Regression, Radial Basis Functions and Recurrent Neural Networks, to build on the work by Matthews et al[8], who established a baseline for machine learning methods in Fantasy Football. Using Bayesian Q-learning algorithm they were able to achieve a top 500th place finish, with their model that in practice modelled week by week sequential team selection. Beal’s work very impressively manages to improve on this model and simulates profit for players playing for financial reward in over 81% of game weeks.

Landers et al. [9] produce a study that again tries to simulate a successful team selection strategy, using Gradient Boosted Decision Trees as both regression and classification algorithms to choose an optimal roster in a traditional fantasy NFL league, and the winning teams in a spread competition. For the traditional league their algorithms best results achieved a success rate of 72%, the second algorithm achieved a 58% success rate, this was a successful enough figure to produce 7 individual wins across 3 fantasy seasons, outscoring the next best participant with 4, this impressive achievement was achieved by using an averaged perceptron with the Boosted Decision Tree.

While Beal’s, Mathews, and Lander’s studies are all impressive in what they can achieve, they do somewhat forget the spirit of the game, and relying on machine learning to select players, can be seen to ruin the fun. With this in mind, this study only aims to provide guidelines on which players to pick rather than have a computer do it for Managers.

A study by Kapania [10] tries to find a balance by trying to use Linear Regression which is a bit more in keeping with the spirit of the game by suggesting which players will do well and recommending them. They also make an interesting point that data can be improved by accounting for player’s age, and proneness to injury, sadly that is not included in this study’s dataset and can not be included as a relevant factor. Kapania’s study is also limited to the running-back position of NFL games so is more specific than this study.

Bonello et al. [11] agree with the points made by Kapania and expand that other external factors e.g. managerial decisions and other tournament fixtures can be significant factors in FPL performance and provide a study which introduces human feedback alongside Gradient Boosted Decision Trees by reading social media posts to inform their models. Initially, it would be easy to say that taking hundreds of Twitter posts would result in lots of noise that could incorporate lots of incorrect opinions. However, their methods are well tuned, and achieved an average of over 300 points per season which is an incredible result.

# METHODOLOGY

## Technologies used

The software which will be used in this project are as follows: IBM SPSS V. 28.0.1.1, Python 3.9.12, Jupyter Notebook 6.4.8. Various python libraries will also be implemented, these include: Pandas, Seaborn, SciKitLearn, MatPlotLib, Statsmodels, and NumPy. This was run on an HP 14s-dq2 laptop with an Intel Core i7-1165G7 processor, and Windows 11 Operating System.

## Methodology

To make sure the data that is being used is relevant, it was necessary to remove some of the noisy data within the dataset. The first step to doing this was to remove any rows that indicate that a player has not made an appearance. The dataset contains rows of players that did not feature during matches, they may have been named in the team’s squad, but not selected for one or more matches. To filter these unnecessary rows out, Excel’s filter tool which allows data from certain columns to be omitted, was used and any row that recorded minutes (i.e. player did not make an appearance) with a value equal to zero was filtered out. The remaining data (43,946 rows) was then copied and saved into a new .csv file ready for use. The decision was made to perform this data cleaning step in excel, so the same remaining data could be copied into SPSS and read into the Jupyter Notebook file as the same and wouldn’t require being cleaned in two different programs.

To prepare the data for the Kruskall-Wallis test in SPSS. There was further data preparation that was performed in Excel. As SPSS would not accept the playing position as groups in their alphabetic form the way they appear in the original data (GK, DEF etc.), Excel’s find and replace tool was used to convert the alphabetic form into numbers. GK, DEF, MID, and FWD were converted to 1, 2, 3, and 4 respectively. The position and total points columns were then in a suitable format to copy and paste into an SPSS workbook, where the value labels for position were then re-converted to their alphabetic form so the output produced could be more clearly interpreted.

Once the data was copied into SPSS, it was inspected using the descriptive statistics option. The data produced a skewness value of 1.867 meaning it was highly skewed, as well as a Kurtosis value of 4.14, comparing this to the produced histogram reveals that the data had a Leptokurtic distribution. This indicates non-normally distributed data, so a test for normality was performed to determine which statistical test was appropriate between ANOVA and Kruskal-Wallis. As this was a very large sample size, a Kolmogorov-Smirnov normality test was performed. The test returned a significance value of < 0.05 and so it was determined that that the data was not normally distributed and a Kruskal-Wallis test would be used to test for differences between positions.

The test was conducted at a significance level of 0.05 and determined with a H statistic. This statistic was compared to a χ2 distribution table to determine whether there was a significant difference between the positions.

The Hypotheses for the Kruskal-Wallis test were as follows:

H0: There is no difference between the positions.

H1: There is a difference between at least one of the positions.

The results and their evaluation will be discussed in Section IV.

The next set of data analysis conducted was a Time Series Analysis. The purpose of this is to forecast the potential points scored by positions over a season, to indicate to managers when specifically, to transfer that type of player into their team, and when they can potentially capitalise on playing bonus chips. Further data preparation was required for this analysis, as the original dataset did not order the rows chronologically. The data was once again filtered in Excel. Firstly, by season, then by position, and then by each game week within the season sequentially. Then the average score was taken from that game week’s total points. That score was then saved in a new .csv file, with one column for each of the four positions. This process was repeated until there was an average score for each position for each game week contained within the filtered dataset. A Jupyter Notebook file was then created, where Pandas, MatPlotLib, and Statsmodels libraries were imported. The new .csv file containing the average game week scores was then read in as a pandas data-frame. The column listing game weeks was set as the index. The columns containing the average scores were then fit individually to Statsmodels’ Holt Winters Exponential Smoothing models. Each position used the same settings, with the ‘trend’ and ‘seasonal’ parameters set to additive, and the ‘seasonal period’ option set to 38. This is because an FPL season contains 38 game weeks. After this, the next 38 values for the four average points columns were forecasted using Statsmodels .forecast method to represent the next season of FPL. MatPlotLib was then implemented to plot a line graph for each of the forecasts with their original data (Figs. B, C, D E) and another was produced (Fig. F) to just show the forecasted data together. This was done to visualise the models, to help explain the forecasts to Managers, and will be discussed in Section IV.

For the Multiple Linear Regression model, it is wanted to determine whether total points can accurately be predicted by the components of FPL’s ICT Index. To create the model, several libraries were imported into a Jupyter Notebook file, these were: Pandas, Matplotlib, Seaborn, and Numpy. Next the .csv file containing the filtered FPL data was read in as a panda’s data-frame. Some basic Exploratory Data Analysis was performed to confirm the data types in the columns. This determined that 1 Column contained Boolean variables, 6 columns contained objects, the rest came back as integer and floats. All columns apart from ‘Influence’, ‘Creativity’, ‘Threat’, ‘ICT Index’, and ‘Total Points’ were then dropped as they are irrelevant to the intended analysis. Using Seaborn a heat plot was generated to see which of the columns featured highly correlated variables. This was done as highly correlated variables will not produce an accurate Regression model. The variables all appeared to be weakly to moderately correlated.

The data was then normalised so that all values were on a scale between 0 and 1, as this helps build more accurate Regression Models when columns contain different metrics.

The next part of the process was to import Sklearn’s Train\_Test\_Split, and LinearRegression utilities. The data was then split with 80% used to train the data and 20% used to test. By default, the model’s shuffle parameter is set to True and this was left at the default setting. The random\_state parameter was set at 42 so that results would be reproduceable. Before setting the final model parameter which is whether or not to use stratified sampling, a count plot was made with Seaborn to determine whether the data in the target y variable (total points) was imbalanced. Fig. G shows that the data was highly imbalanced and so it was hoped the model could use stratified sampling so that data was evenly distributed, however the least populated class in y had only 1 member, which is too few, and so the model had to be built without stratified sampling. The trained data was then fitted to the model. The results and their evaluation will be discussed in the next section.

# RESULTS AND EVALUATION

A Kruskal-Wallis test was performed in SPSS to determine whether there was a statistically significant difference between the total points scored by different positions, using a total of 43,922 players over 5 seasons.

A H statistic of 676.631 was cross referenced against a χ2 distribution table with 3 degrees of freedom at a significance level of 0.05. The H statistic was larger than the χ2 value of 7.815. This means the decision of the test is to reject the null hypothesis and conclude that there appears to be a significant difference between at least two of the positions. The SPSS output on the Kruskal-Wallis test provides a pairwise comparison table to determine specifically which positions differ from one another. The output indicates that all positions significantly differ from one another as all produced a significance value of <0.05.

A means plot (Fig. A) was also generated in SPSS to better visualise the results. From this it can be seen that, on average, GK is the highest scoring position followed by FWD, then DEF, and finally, MID. From the results produced, managers could be advised to allocate more of their budget to the GK and FWD positions as these contain higher scoring players.

A Holt Winters Time Series Analysis with triple exponential smoothing was then performed, and visualisations were created to examine whether there was a certain point in the season that Managers should prioritise transferring in players from specific positions, and when to play bonus chips. When examining the average forecasted values for the different positions it can be seen that (with the exception of a couple of early game weeks) the GK position is predicted to consistently outperform the other positions, so managers should make sure they have a reliable player in that position throughout. Other distinctions that can be drawn from the visualisation are that FWDs will perform best in the final third of the season, whilst DEFs will drop in performance, so managers should prioritise transfers and team setups accordingly, perhaps with their second wildcard. The Triple Captain could be best used by Manager’s Goalkeeper in game week 36 as this was the highest forecasted value of 4.14. The game week with the highest sum of average values (11.34) was game week 1, as such this may be a beneficial week for managers to play their Bench Boost chip as they will avail of the points of the players on their bench.

In order to evaluate how well the Multiple Linear Regression model performed, a number of performance metrics were imported from sklearn and were run, their scores are as follows: Mean Squared Error: 0.003, Root Mean Squared Error: 0.057, R2: 0.539, and Max Residual Error: 0.399.

The Mean Squared Error is the most popular metric when evaluating Regression problems, and on a scale of 0 to ∞ a score of 0.003 would indicate that the model performs very well. However, MSE is prone to any remaining outliers in the dataset, Whereas the next metric RMSE is less prone to outliers, returning a value of 0.057 indicates that the model performs well. This is also backed up by the low MAE value 0.04, as once again this metric is stronger against the influence of any residual outliers. An R2 Statistic has also been calculated to determine how much of the total variation in the target variable is explained by the regression line, and we can see that the model does a reasonable job capturing around 54% of the overall variation. [12]

The coefficient values for the ICT components were also produced and were as follows: Influence: 0.61, Creativity: 0.02, Threat: 0.11. The coefficient values explain which features best explain Total Points. From the produced coefficients the study can recommend to managers using the ICT index to pick players for their team are hypothetically choosing between players with similar ICT Index score, that it is advised to choose the player with the highest score for Influence as this was the best performing feature for the model.

# CONCLUSIONS AND FUTURE WORK

## Conclusions

This Study set out to determine whether previous FPL data could assist in providing an optimised game plan for Fantasy Managers. In order to answer this question, an Kruskal-Wallis test, with post-hoc testing was performed, A Holt Winters Time Series Analysis was conducted, and a Multiple Linear Regression Model was created. The Kruskall-Wallis test provided very compelling evidence that there was a significant difference between the positions, with GK outscoring the others. The Holt Winters Time Series Analysis predicted that the trend for GK to outperform the others would continue, as well as providing a few insights on where managers could most effectively play their bonus chips. The Multiple Linear Regression model showed that a player’s total points could be accurately predicted by FPL’s ICT Index and recommends players with high Influence scores over others.

So, in conclusion: Can previous data provide managers with an optimised game plan? Yes, it is believed that the study has shown some key insights from previous FPL data that would help enhance future scores. Specifically, by picking players from the demonstrably higher performing positions, playing bonus chips at times highlighted by the Time Series Analysis, and picking players with the best performing attributes.

## Future Work

In future, it is believed that the study would benefit from additional data. Specifically, the data from the missing seasons 2018-19 and 2019-20 would be particularly useful. This would hopefully benefit the Multiple Linear Regression model and provide enough correct data to satisfy the stratified sampling parameter as it is believed this would help to create a better model as the data would be more evenly distributed. The study could also be furthered by implementing more models such as…

From a methodology perspective, in future it would be more efficient to automate the data scraping process for the Time Series Analysis using Python, as performing it manually was particularly time consuming.

##### References

1. Reddy J., (2024), *Can previous Fantasy Premier League data provide an optimised game plan for Fantasy Managers?* – Project Proposal, National College of Ireland.
2. Premier League, *How the ICT index in Fantasy Works* (10 Jul 2023) *Premier League*. Available at: <https://www.premierleague.com/news/65567>
3. Arvidsson, J. (2023) *Fantasy football*, *Kaggle*. Available at: <https://www.kaggle.com/datasets/joebeachcapital/fantasy-football>
4. Bangdiwala, Malhar & Choudhari, Rutvik & Hegde, Adwait & Salunke, Abhijeet. (2022). *Using ML Models to Predict Points in Fantasy Premier League.* Available at: <https://www.researchgate.net/publication/364322490_Using_ML_Models_to_Predict_Points_in_Fantasy_Premier_League>
5. Rajesh V., Arjun P., Jagtap K. R., Prakash S. C. M and J., "Player Recommendation System for Fantasy Premier League using Machine Learning," *2022 19th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, Bangkok, Thailand, 2022, pp. 1-6 Available at: <https://ieeexplore.ieee.org/document/9836260>
6. Yang R., (2019). Using Supervised Learning to Predict English Premier League Match Results from Starting Line-up Player Data. M.Sc. in Computing (Data Analytics). Technological University Dublin. Available at: <https://arrow.tudublin.ie/cgi/viewcontent.cgi?article=1192&context=scschcomdis>
7. Beal, R., Norman, T.J. and Ramchurn, S.D. (2020) ‘Optimising daily fantasy sports teams with Artificial Intelligence’, *International Journal of Computer Science in Sport*, 19(2), pp. 21–35 Available at: <https://eprints.soton.ac.uk/445995/1/DFS_IJCSS.pdf>
8. Matthews T., Ramchur, S., & Chalkiadakis G. (2021). *Competing with Humans at Fantasy Football: Team Formation in Large Partially-Observable Domains*. Proceedings of the AAAI Conference on Artificial Intelligence, 26(1), 1394-1400. Available at: <https://doi.org/10.1609/aaai.v26i1.8259>
9. Landers J. R., and Duperrouzel b., "Machine Learning Approaches to Competing in Fantasy Leagues for the NFL," in *IEEE Transactions on Games*, vol. 11, no. 2, pp. 159-172, June 2019 Available at: <https://ieeexplore.ieee.org/document/8367900>
10. Kapania, N. (2012). *Predicting Fantasy Football Performance with Machine Learning Techniques*. Stanford University, Available at: <http://cs229.stanford.edu/proj2012/KapaniaFantasyFootballAndMachineLearning.pdf>
11. Bonello N., Beel J., Lawless S., and Debattista J., (2019), *Multi-streamData Analytics for Enhanced Performance Prediction inFantasy Football*. 27th AIAI Irish Conference on Artificial Intelligence and Cognitive Science. Available at: [1912.07441.pdf (arxiv.org)](https://arxiv.org/ftp/arxiv/papers/1912/1912.07441.pdf)
12. Reddy J., (2024), *Machine Learning CA 2*, National College of Ireland.

APPENDIX

A graph with a line

Description automatically generated

Fig. A. Means plot of total points by playing position created in SPSS.

A graph showing a sound wave

Description automatically generated

Fig. B. Time Series Forecast for GK, created with MatPlotLib.

A graph showing a sound wave

Description automatically generated

Fig. C. Time Series Forecast for DEF, created with MatPlotLib.

A graph showing a sound wave

Description automatically generated

Fig. D. Time Series Forecast for DEF, created with MatPlotLib.

A graph showing a graph of a graph

Description automatically generated with medium confidence

Fig. E. Time Series Forecast for DEF, created with MatPlotLib.

A graph of different colored lines

Description automatically generated

Fig. F. Time Series Forecast for DEF, created with MatPlotLib.

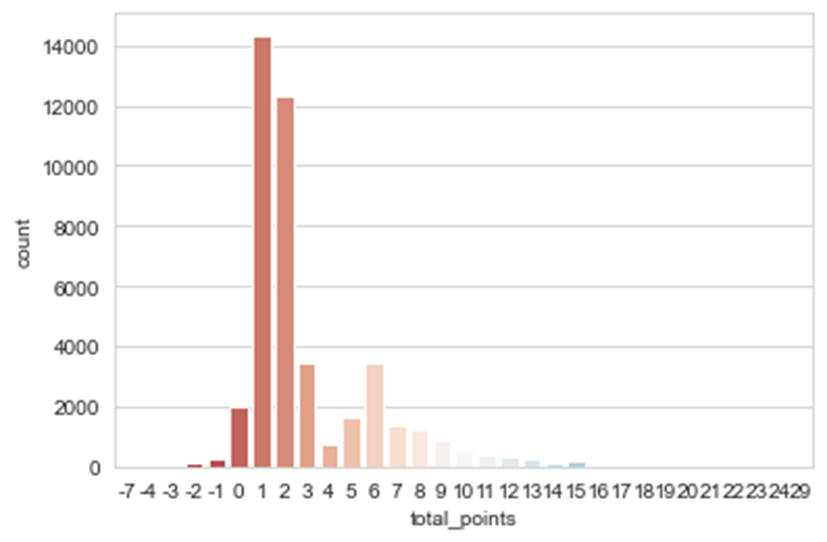


Fig. G. Count plot of total points distribution made with Seaborn.