Study & Simulation of a Potential Betting based Business on Premier League Games

*BANA 7030 – Simulation Modelling and Methods*

*By*

*Mohammed Nifaullah Sailappai (M13433497)*

Contents

Problem Statement……........................................................................................................................... 2 Data...…………….………..................................................................................................................... 2

Model………………….……………...................................................................................................... 2

Simulation Analyis…………..…………................................................................................................. 4

Alternate Approach…….……………….................................................................................................. 6

Conclusion………….………………...……………………..................................................................... 6

References.................................................................................................................................................. 7 Appendices……….………………...……………………........................................................................ 7

1. Problem Statement

Clifton Sports Entertainment (CSE) is a popular sport oriented business based out of Clifton, Cincinnati. CSE owns many sport companies including 10 popular sport bars in Cincinnati. Off late, CSE has realized, through various sporting avenues, that there is a growing interest for Premier League (Soccer League in UK) in Cincinnati, Ohio. Knowing the local communities history of interest in Sports betting, through their own online based sports betting platform, and a new season to start in 2 months times, CSE wonders if a similar profitable market exists for Premier League Games. Based on this intuition, CSE’s management commissioned a survey where they asked different questions to Premier League Fans with regards to premier league and betting in general, either through person or online. Results of this exercise were promising, through the questionnaire CSE estimated that if CSE were to start a betting business on Premier League Games in 2 months’ time they would able to generate at least $ 3000 worth of betting. However, before going full-on CSE would like to conduct a small analysis on how profitable the business would be. As a part of the Analysis, CSE requires the analyst to develop a betting based probabilistic model that provides betting on 3 categories 1. Premier League Winner – where a gambler picks his favorite to win the league, 2. Top 4 – where a gambler bets on a team to finish in one of the first 4 places, 3. Relegation – where a gambler bets on a team to be relegated (i.e. finishes on one of the last 3 positions). CSE, through publically available data and common knowledge, expects a full-fledged analysis on the Model’s performance on a wide variety scenarios. Also, CSE already has a web development which takes care of its other sports based betting, it assumes a fixed cost of $ 1000 for this task. Please note that in actuality profits/loss can be calculated only after 9 months (i.e. after the completion of the season)

2. Data

As there was no dataset provided by the company, we curated the data for the model from the openly available data at [Premier League](https://premierleague.com). We manually collected only last 5 season’s data because being a avid followers of the game we felt that this was the proper interval to average out any nuisance, anything longer or shorter than this would most likely skew the data. In some cases, where we didn’t have the last 5 years data we have averaged the final value over how many ever years the data was available for that team. From the data we deduced the probability of Wins, Draws or Losses over a 38 game season. We would then use this probabilities to generate all possible data for different teams. We also calculate the odds at which we have to pay back if the gambled event comes true, odds are a function of respective probabilities and predicted position.



3. Model

Parameterized by the probabilities estimated for each club in the earlier step we use multinomial distribution to generate Wins, Draws and Losses for each team by drawing from this distribution 10000 times and then averaging out to get the final result. From these results we can calculate the Expected Number of Points for each team and based on the Expected Number of Points we can calculate the Expected Position of the team. We then use the Expected Number of Points to calculate the probability of a team to finish on Top, we have use a simple approach of dividing the Expected Number of Points to the Maximum Number of points obtainable to get the probability. We then also create another column to get the normalized probability of a team winning league. After several trials and through common knowledge we set the threshold value at 85, we then divide the Expected Number of Points with this threshold value to get the probability of a club finishing in Top 4. Further we calculate the probability of relegation by subtracting one from its probability of finishing on Top (remember it was calculated against the Maximum Number of Points obtainable - 114). We further normalize probability for both Top 4 and Relegation.



3.1 Model Dictionary

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Variable** | **Description** | **Formula** |
| 1 | Teams | Name of the Club | NA |
| 2 | Win\_Prob | Winning Probabality across a 38 game season | Average( # wins in last 5 season)/38 |
| 3 | Draw\_Prob | Drawing Probabality across a 38 game season | Average( # draws in last 5 season)/38 |
| 4 | Loss\_Prob | Losing Probabality across a 38 game season | Average( # losses in last 5 season)/38 |
| 5 | Predicted\_Points | Expected Number of Points to obtained in the upcoming season | Averaged\_Wins \* 3 + Averaged\_Draws \* 1 |
| 6 | Predicted\_Position | Expected Position at the end of the season | Rank |
| 7 | winning\_league\_prob | Probability of winning the league | Expected # of points/maximum points obtainable(114) |
| 8 | winning\_league\_odds | Return on $1 waged on picking this particular team to win the league | 1 + ((1 - winning\_league\_prob) \* (1 + predicted\_position) |
| 9 | winning\_league\_risk | Normalized Probability of Winning the league | Winning\_league\_prob/Sum of all winning\_league\_probs |
| 10 | top\_4\_prob | Probability of finishing Top 4 | Expected # of points/85 |
| 11 | top\_4\_odds | Return on $1 waged on picking this particular team to finish in Top 4 | 1 + ((1 - top\_4\_prob) \* (1 + predicted\_position)/3 |
| 12 | top\_4\_risk | Normalized Probability of finishing in top 4 | top\_4\_prob/Sum of all top\_4\_\_probs |
| 13 | relegation\_prob | Probability of getting relegated from the league | 1 - winning\_league\_prob |
| 14 | relegation\_odds | Return on $1 waged on picking this particular team to be relegated | 1 + ((1 - relegation\_prob) \* (1 + (21 - predicted\_position)//3)) |
| 15 | relegation\_risk | Normalized Probability of relegation | relegation\_prob/Sum of all relegation\_\_probs |

4. Simulation Analysis

4.1 General Analysis

Below general analysis depicts the boundary conditions for both events. While the loss associated with worst case is very high, the profit at the best of time is not proportionate. Also it gives us a fair warning as to what might happen if worst happens.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Scenario** | **Winning League** | **Top 4** | **Relegation** | **Total Profit** | **Profitable From** |
| **Worst Case** | -14648.42 | -4556 | -5167.72 | -25372.85 | Never |
| **Best Case** | 1000 | 1000 | 1000 | 2000 | Always |

\

4.2 Risk Based Analysis – Clubs

In this set of analysis we analyze how the model performs to various risk sharing mechanism at club level.

4.2.1 Uniform Risk Sharing

In this analysis we uniformly share the risk to each team irrespective of its odds and probabilities. We can see that the model has already seen a marked improvement and is already seeing a reduced in worst case. Best case may be not as good as previous one but still is substantially higher.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Scenario** | **Winning League** | **Top 4** | **Relegation** | **Total Profit** | **Profitable From** |
| **Worst Case** | 217.58 | -2.6 | 136.4 | -648.62 | 3060 dollars |
| **Best Case** | 923.82 | 760.23 | 784.75 | 1468.8 | 400 dollars |

4.2.2 Probability based Risk Sharing

Here we share the revenue based on probability of it happening, this follows a simple hypothesis that most people invest in low risk low return ideas. It is evident from the below table that this is a highly profitable model given the fact that it is able to churn profit even in the worst case scenario. Please note that these results are averaged after at least 10000 rounds of simulation and In-between analysis – I is profit after simulation for 50000 rounds accommodating for not the worst losses, whereas In-between Analysis – II is a mixed bag simulation of 30000 rounds

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Scenario** | **Winning League** | **Top 4** | **Relegation** | **Total Profit** |
| Worst Case | 498.85 | 302.08 | 480.35 | 281.28 |
| Best Case | 880.43 | 659.95 | 724.89 | 1265.27 |
| In-Between Analysis - I | 809.14 | 520.89 | 602.6 | 932.63 |
| In-Between Analysis - II | 857.7 | 644.22 | 706.19 | 1208.11 |

4.3 Risk Based Analysis – Categories

In this set of analysis we analyze how the model performs to various risk sharing mechanism at category level. Remember that the club level risk are still shared based on the probability of a outcome.

4.3.1 Individual Categories

Here we assume that all the revenues come from only one category.

4.3.1.1 Winning League

Model is profitable in both the scenarios given that we manage to get a bet of at least $ 3000.

|  |  |  |
| --- | --- | --- |
| **Scenario** | **Winning League** | **Total Profit** |
| Worst Case | 1493.51 | 493.51 |
| Best Case | 2642 | 1642 |

4.3.1.2 Top 4

Model is slightly losing some money in the worst case scenario.

|  |  |  |
| --- | --- | --- |
| **Scenario** | **Top 4** | **Total Profit** |
| Worst Case | 907.64 | -92.36 |
| Best Case | 1980.24 | 980.24 |

4.3.1.3 Relegation

Model is profitable in both the scenarios given that we manage to get a bet of at least $ 3000.

|  |  |  |
| --- | --- | --- |
| **Scenario** | **Relegation** | **Total Profit** |
| Worst Case | 1441 | 441 |
| Best Case | 2175 | 1175 |

4.3.2 Probability based Risk Sharing

Continuing on the same idea as above only thing is here category based risk sharing is also assigned based on probability of a outcome we share the revenue based on probability of it happening, this follows a simple hypothesis that most people invest in low risk low return ideas. Overall after several thousand rounds of simulation we can see that the model is profitable by a good margin

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Scenario** | **Winning League** | **Top 4** | **Relegation** | **Total Profit** |
| Worst Case | 187.25 | 454.29 | 541.87 | 183.41 |
| Best Case | 330.12 | 989.9 | 815.44 | 1135.46 |
| In-Between Analysis - I | 303.48 | 782.2 | 678.4 | 764.08 |
| In-Between Analysis - II | 321.63 | 966.61 | 794.68 | 1082.92 |

5. Alternate Approach

There’re many other ways one could have approached this problem, but one approach that I would have tried is to predict the probabilities using the Logistic Regression Model. I would most likely collect more data & build a more robust Logistic Regression model which would account for team improvement through transfers, injuries, new managers, new owners and many other factors. Although our model is good and seems to be doing pretty well in most scenarios but the Logistic Regression with all those factors would have been more prone to any noise in the data.

6. Conclusion

Baring few extreme scenarios through the above simulation analysis and modelling we were able to demonstrate on most cases that the model is profitable provided that on an average revenue (betting) on each category is at least 700 dollars (for most cases). A deep dive into different scenarios revealed some interesting insights, for instance if we can manage gamblers to uniformly spend the revenue on the Premier League Winner Category then profits increase by a higher rate compared to others. Overall Top 4 is the safest option for the Gamblers given that they've a high probability of getting it right, and likewise profits for the betting company does not increase with the same rate as it increases for PL Winner Category. Overall our model is a highly profitable model if we manage to at least purchase gambles worth of 1000 dollars for each category.

7. References:

1. <http://www.bestbettingonline.com/strategy/create-model/>
2. <http://www.football-data.co.uk/>
3. <https://tolstoy.newcastle.edu.au/R/e8/help/att-6544/dixoncoles97.pdf>

8. Appendices:

This analysis was done using Jupyter Notebook and Python Programing.

 