INFO 6205: Premier League Ranking System

By:

Karthik Balasubramanya, NUID:001374623

Ganesh Murthy Venkateshan, NUID:001087846

Chethan Tururvekere Sreenivas, NUID:001833160

INTRODUCTION:

Probability and Poisson Distribution Function:

Probability in general is the mathematical representation of how likely an event is going to result in a particular outcome. Probability always and should lie between 0 and 1.

Poisson Distribution function gives us the probability of the number of times an event results in the value. Poisson Distribution is used primarily for finding probability for discrete events. Also, when using Poisson distribution, we consider the events to be independent of each other and the number of occurrences is known.

In this case soccer goals x is a Poisson random variable and distribution of X is termed as Poisson distribution. In this case we assume that each goal scored is independent of other

Probability of occurrence of x goals is given by following formula:

Poisson Distribution Formula

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

where

x = 0, 1, 2, 3, ...

λ = mean number of occurrences in the interval

e = Euler's constant $\approx$ 2.71828

For Predicting a soccer match,

lambda – average goals number of goals expected from the team

x – probability of x goals in the match scored by that team

OBJECTIVE:

The Objective of this project is to:

1. Design and develop a ranking system for the English Premier League.
2. Calculate the Averages of a team's attack and defense for both home and away.
3. Predicting the outcome of a match played between two teams and calculating the win, loss and draw probability for each match.
4. Using the predictions, Simulate the remaining fixtures for the 2019-2020 season and come up with the final premier league table for the 2019/2020 season.

ALGORITHM AND IMPLEMENTATION:

To predict the result between two teams Home Team vs Away Team. We need to calculate the form of the home team at home ground (ability to score goals at home) and the form of the away team at away ground (ability to score goals away).

To calculate the mean (ability to score goals) we need to consider how each team has performed in their previous matches home or away. For our project, we have considered data from the 2015-2016 season up to the current season.

Data is of the format:

HomeTeam	AwayTeam	Full Time Home Goals	Full time Away Goals
Liverpool	Norwich	4	1
West Ham	Man City	0	5
Bournemouth	Sheffield Unite	1	1
Burnley	Southampton	3	0
Crystal Palace	Everton	0	0
Watford	Brighton	0	3
Tottenham	Aston Villa	3	1
Leicester	Wolves	0	0
Newcastle	Arsenal	0	1
Man United	Chelsea	4	0
Arsenal	Burnley	2	1
Aston Villa	Bournemouth	1	2
Brighton	West Ham	1	1
Everton	Watford	1	0
Norwich	Newcastle	3	1
Southampton	Liverpool	1	2
Man City	Tottenham	2	2
Sheffield United	Crystal Palace	1	0
Chelsea	Leicester	1	1
Wolves	Man United	1	1
Aston Villa	Everton	2	0
Norwich	Chelsea	2	3

Using the data from the previous seasons, we calculate the number of games played at home, number of games played away, the number of goals art home and the number of goals away for each team. We read from the CSV files and create a Club object and store it in a **hash map** with the name of the team as the key.

The Map looks like follows,

Team Name	homeMatchesPlayed	awayMatchesPlayed	homeGoals	awayGoals	homeConceded	awayConeceded
Fulham	19	19	22	12	36	45
Cardiff	19	19	21	13	38	31
Sunderland	38	38	39	38	54	77
Newcastle	71	72	89	61	78	123
Tottenham	90	91	183	160	71	105
Hull	19	19	28	9	35	45
Huddersfield	38	38	26	24	56	78
Crystal Palace	91	90	103	108	111	143
Watford	90	91	114	89	124	161
Middlesbrough	19	19	17	10	23	30
Arsenal	91	89	192	137	83	135
Chelsea	91	90	178	142	90	112
Stoke	57	57	66	51	78	101
Liverpool	91	90	218	162	72	101
Brighton	52	53	60	41	68	86
Aston Villa	32	34	32	29	59	73
Norwich	33	34	45	19	56	63
West Brom	57	57	68	40	77	78
Sheffield United	15	13	17	13	13	12
West Ham	90	91	132	115	136	152
Leicester	91	90	145	136	100	135
Burnley	72	71	86	68	88	114
Bournemouth	90	91	131	99	139	173
Man United	91	90	153	127	67	109
Southampton	91	90	119	98	130	132
Wolves	34	33	49	39	38	42
Everton	90	91	154	102	104	145
Swansea	57	57	64	51	78	100
Man City	89	91	235	185	76	85

After reading from the csv files, we need to calculate the Average Home Attack Strength, Average Away Attack Strength, Average Home Defense Strength and the Average Away Defense Strength across all matches in the league.

We also know that the Home Attack Strength is the same as Away Defense Strength and Home Defense Strength is same as Away Attack Strength.

```
Home Attack Strength across the league = total Home Goals/total Games;
Away Attack Strength across the league = total Away Goals/total Games;
    Home Defense Strength across the league =Away Attack Strength;
    Away Defense Strength across the league =Home Attack Strength;
```

When predicting for a match between two teams Team1 (Home)vs Team2 (Away), we calculate the Team1 mean and Team2 mean.

First, we calculate Team1 mean at home:

To calculate the mean at home, we need to calculate Team1 Attacking power at home. Team1's Attacking power at home is

```
(Goals Scored by Team1 at Home/Total number of games played by Team1 at home)/Home Attack Strength across the league
```

Then, we calculate Team2's Defense power away which is

```
(Goals Conceded by team2 away/Total Number of games played by team2 away)/Away Defense Strength across the league
```

Similarly, we calculate Team2 mean:

Team2 Attacking power away:

```
(Goals Scored by Team2 away/Total number of games played by Team2 away)/Away Attack Strength across the league
```

Team1 Defense Power home:

```
(Goals Conceded by team1 home/Total Number of games played by team1 home)/Home Defense Strength across the league
```

Then, we calculate the means for team1 and team2

Mean of team1 = Attacking power of team1 at home × Defense power of team2 away × home Attack Strength across the league

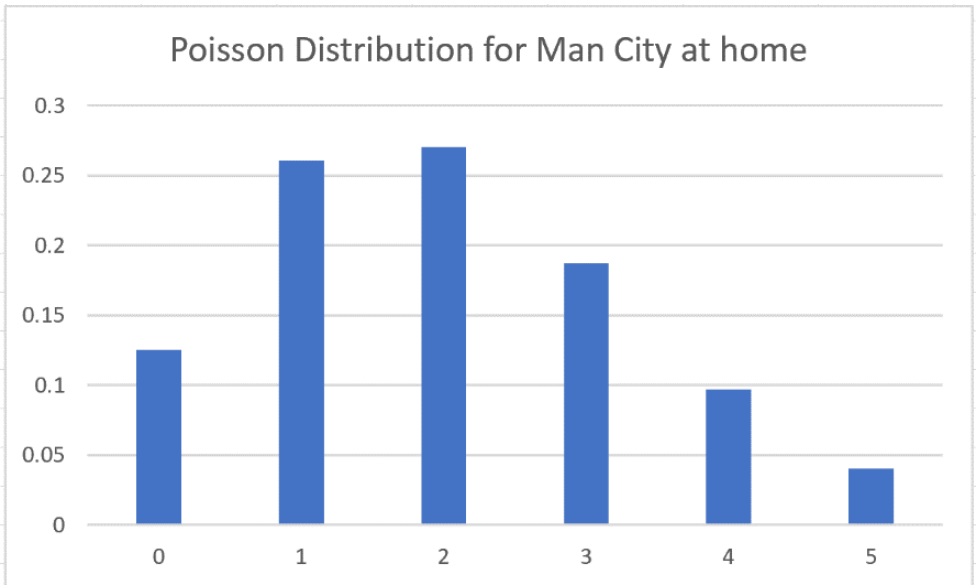
Mean of team2 = Attacking power of team2 away × Defense power of team1 home × away Strength across the league

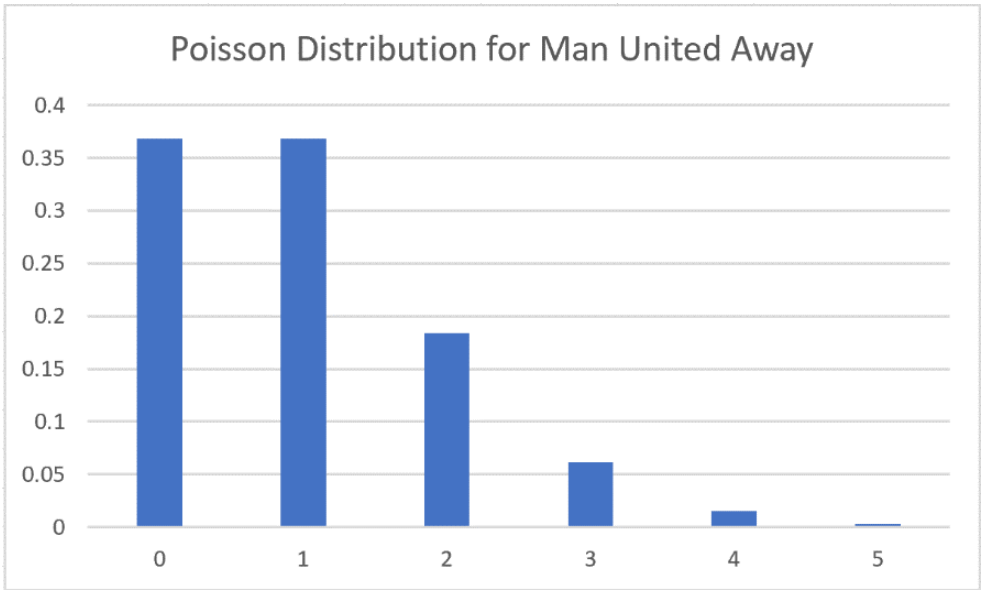
We now calculate our Poisson Distribution for home team and the away team using the means from above.

Using the Poisson distributions from the two teams, we calculate what could be the outcome of the match. We consider that each team can score from 0-5 goals. The probability of teams scoring more than 5 goals is very unlikely. The Poisson Distribution is calculated by importing the apache.commons.math3.distribution.PoissonDistribution class.

So, we calculate the probability that the team will score goals (0-5). For example, for

Man City vs Man United, the probabilities are going to be:





After calculating the Poisson Distribution for home and the away team, we us a **2-D array** to predict the score where the row represents the number of goals the away team can score, and the column represents the number of goals the home team can score.

We populate the table using:

```
table[i][j] =
(poissonDistributionAway.probability(i)*poissonDistributionHome.probability(j))*100
```

So, for example the probability of the score being 1-2, we multiply the Poisson probability of the away team scoring two goals and the home team scoring one goal

```
(poissonDistributionAway.probability(2) *poissonDistributionHome.probability(1))*100
```

Table for two teams looks as follows:

	0	1	2	3	4	5
0	5.441	8.59	6.79	3.57	1.41	0.44
1	7.24	10.02	10.444	4.76	1.88	0.59
2	4.8	7.61	6.01	3.16	1.25	0.39
3	2.139	3.380	2.670	1.9	0.55	0.17
4	0.712	1.125	0.88	0.46	0.18	0.05
5	0.18	0.29	0.23	0.12	0.049	0.015

From the above table we find out that at table [1][2], the probability is the highest. So, the most probable score is 2-1.

From the table, we can find the win probability, draw probability and the loss probability (other team's win probability). To calculate the probability of the home team winning, we add the probabilities in all the cells that depict the home team winning (All the cells with color green), to calculate the probability of the match resulting in a draw, we add the probabilities in all the cells that show the draw match(All the cells with color yellow) and finally to calculate the loss probability, we add the probabilities in all the cells that depict the home team losing or the away team winning (All the cells with color red).

Along with the probabilities, we return top three probable scores for simulating the league. To retrieve the top 3 sores, we have used a **priority queue** and we remove the maximum value three times.

SIMULATING THE LEAGUE:

When Simulating the Premier league, we initially need to import the premier league table and the list of upcoming fixtures that needs to be predicted through csv files. When importing the table, we create an object for each row in the table and store it in a Map with the name of the team as the key.

For each fixture, we run the prediction and we get our win probabilities, draw probability and loss probability. Along with the probabilities we return the top 3 scores with the highest probability.

For updating the premier league table, we consider the score with the highest probability.

General rule to update the premier league table:

Win: +3 points

Draw: +1 point

Loss: +0 points.

Additionally, we also update the goal difference with respect to each result.

After simulating all the matches, we write the final table back to the csv file.

OUTPUTS:

Simulation of a Premier League Match between Arsenal vs Tottenham:

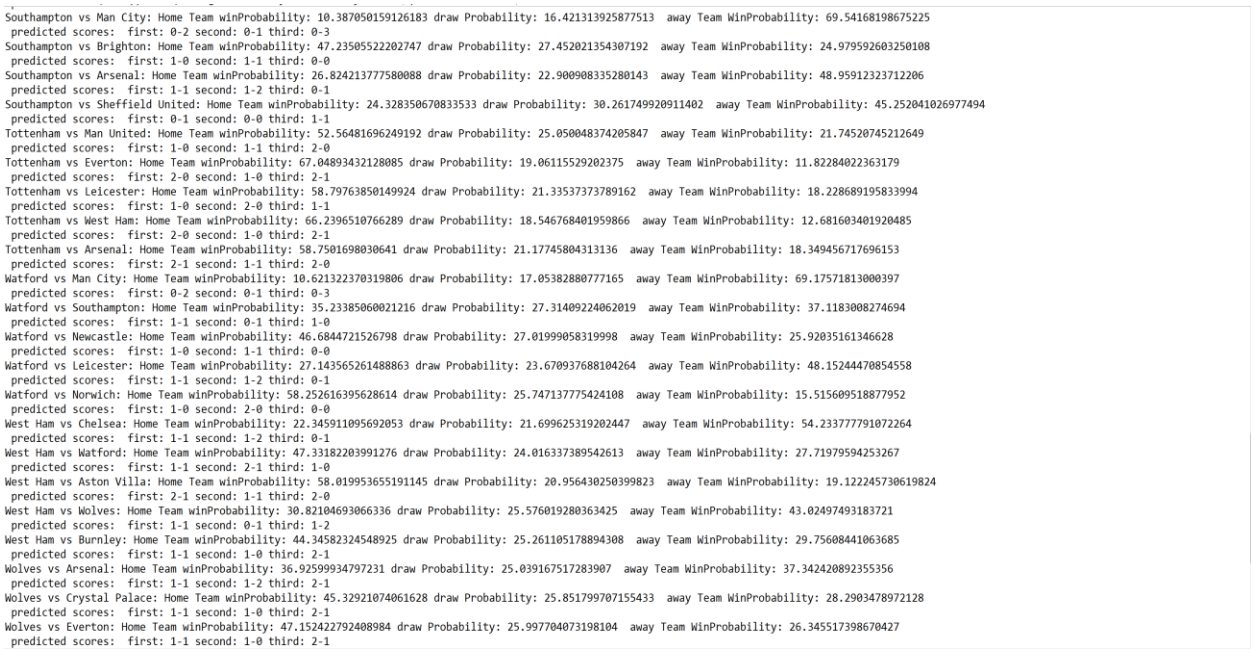
Poisson Distribution for both the teams:

```
Arsenal Form: 1.5798834772825605
Tottenham Form: 1.3312384141554972
Arsenal Poisson Distribution:
Probability of scoring 0 goals :0.2059991003814234
Probability of scoring 1 goals :0.32545457502768244
Probability of scoring 2 goals :0.25709015284612646
Probability of scoring 3 goals :0.13539082821788115
Probability of scoring 4 goals :0.05347543311925793
Probability of scoring 5 goals :0.016896990645128843
Tottenham Poisson Distribution:
Probability of scoring 0 goals :0.26414993164229844
Probability of scoring 1 goals :0.3516465360987764
Probability of scoring 2 goals :0.23406268852970444
Probability of scoring 3 goals :0.1038644140970853
Probability of scoring 4 goals :0.03456707447744841
Probability of scoring 5 goals :0.009203403481870678
```

Evaluating the goal scoring matrix for Arsenal vs Tottenham:

```
[5.441464828412797, 8.596880374623561, 6.791034630021237, 3.5763478018747477, 1.4125532002994285, 0.44633389238713406]
[7.243887008859167, 11.444497396598386, 9.040486171344535, 4.760971576236238, 1.8804450822768837, 0.5941768230852987]
[4.821670326997643, 7.6176772825271755, 6.0175212369677, 3.1689941254940637, 1.251660364618391, 0.3954955058460122]
[2.1395975865643204, 3.3803148750466123, 2.670251809549303, 1.406228904696935, 0.5554194529519596, 0.17549960333602385]
[0.7120786245172034, 1.125001253400822, 0.8886854460850648, 0.4680064842570921, 0.18484892793472005, 0.05840795340749178]
[0.18958928377126194, 0.2995289769000514, 0.2366104407858705, 0.12460564198338023, 0.0492155987364321, 0.015550982253651507]
Arsenal winProbability: 42.704706557803 draw Probability: 24.510112276864188 Tottenham away Team Probability: 31.9667206399814
```

Predicting the league by simulating the upcoming fixtures:



```
Southampton vs Man City: Home Team winProbability: 10.387050159126183 draw Probability: 16.421313925877513  away Team WinProbability: 69.54168198675225
 predicted scores:  first: 0-2 second: 0-1 third: 0-3
Southampton vs Brighton: Home Team winProbability: 47.23505522202747 draw Probability: 27.452021354307192  away Team WinProbability: 24.979592603250108
 predicted scores:  first: 1-0 second: 1-1 third: 0-0
Southampton vs Arsenal: Home Team winProbability: 26.824213777580088 draw Probability: 22.900908335280143  away Team WinProbability: 48.95912323712206
 predicted scores:  first: 1-1 second: 1-2 third: 0-1
Southampton vs Sheffield United: Home Team winProbability: 24.328350670833533 draw Probability: 30.261749920911402  away Team WinProbability: 45.252041026977494
 predicted scores:  first: 0-1 second: 0-0 third: 1-1
Tottenham vs Man United: Home Team winProbability: 52.56481696249192 draw Probability: 25.050048374205847  away Team WinProbability: 21.74520745212649
 predicted scores:  first: 1-0 second: 1-1 third: 2-0
Tottenham vs Everton: Home Team winProbability: 67.04893432128085 draw Probability: 19.06115529202375  away Team WinProbability: 11.82284022363179
 predicted scores:  first: 2-0 second: 1-0 third: 2-1
Tottenham vs Leicester: Home Team winProbability: 58.79763850149924 draw Probability: 21.33537373789162  away Team WinProbability: 18.228689195833994
 predicted scores:  first: 1-0 second: 2-0 third: 1-1
Tottenham vs West Ham: Home Team winProbability: 66.2396510766289 draw Probability: 18.546768401959866  away Team WinProbability: 12.681603401920485
 predicted scores:  first: 2-0 second: 1-0 third: 2-1
Tottenham vs Arsenal: Home Team winProbability: 58.7501698030641 draw Probability: 21.17745804313136  away Team WinProbability: 18.349456717696153
 predicted scores:  first: 2-1 second: 1-1 third: 2-0
Watford vs Man City: Home Team winProbability: 10.621322370319806 draw Probability: 17.05382880777165  away Team WinProbability: 69.17571813000397
 predicted scores:  first: 0-2 second: 0-1 third: 0-3
Watford vs Southampton: Home Team winProbability: 35.23385060021216 draw Probability: 27.31409224062019  away Team WinProbability: 37.1183008274694
 predicted scores:  first: 1-1 second: 0-1 third: 1-0
Watford vs Newcastle: Home Team winProbability: 46.6844721526798 draw Probability: 27.01999058319998  away Team WinProbability: 25.92035161346628
 predicted scores:  first: 1-0 second: 1-1 third: 0-0
Watford vs Leicester: Home Team winProbability: 27.143565261488863 draw Probability: 23.670937688104264  away Team WinProbability: 48.15244470854558
 predicted scores:  first: 1-1 second: 1-2 third: 0-1
Watford vs Norwich: Home Team winProbability: 58.252616395628614 draw Probability: 25.747137775424108  away Team WinProbability: 15.515609518877952
 predicted scores:  first: 1-0 second: 2-0 third: 0-0
West Ham vs Chelsea: Home Team winProbability: 22.345911095692053 draw Probability: 21.699625319202447  away Team WinProbability: 54.233777791072264
 predicted scores:  first: 1-1 second: 1-2 third: 0-1
West Ham vs Watford: Home Team winProbability: 47.33182203991276 draw Probability: 24.016337389542613  away Team WinProbability: 27.71979594253267
 predicted scores:  first: 1-1 second: 2-1 third: 1-0
West Ham vs Aston Villa: Home Team winProbability: 58.019953655191145 draw Probability: 20.956430250399823  away Team WinProbability: 19.122245730619824
 predicted scores:  first: 2-1 second: 1-1 third: 2-0
West Ham vs Wolves: Home Team winProbability: 30.82104693066336 draw Probability: 25.576019280363425  away Team WinProbability: 43.02497493183721
 predicted scores:  first: 1-1 second: 0-1 third: 1-2
West Ham vs Burnley: Home Team winProbability: 44.34582324548925 draw Probability: 25.261105178894308  away Team WinProbability: 29.75608441063685
 predicted scores:  first: 1-1 second: 1-0 third: 2-1
Wolves vs Arsenal: Home Team winProbability: 36.92599934797231 draw Probability: 25.039167517283907  away Team WinProbability: 37.342420892355356
 predicted scores:  first: 1-1 second: 1-2 third: 2-1
Wolves vs Crystal Palace: Home Team winProbability: 45.32921074061628 draw Probability: 25.851799707155433  away Team WinProbability: 28.2903478972128
 predicted scores:  first: 1-1 second: 1-0 third: 2-1
Wolves vs Everton: Home Team winProbability: 47.152422792408984 draw Probability: 25.997704073198104  away Team WinProbability: 26.345517398670427
 predicted scores:  first: 1-1 second: 1-0 third: 2-1
```

Simulation of the Premier League considering data from 2015-2016 Season to 2019-2020 Season

1	Liverpool	played=38	won=32	drew=5	lost=1	goalsFired=79	goalsAcquired=25	goalDifference=54	points=101
2	Man City	played=38	won=26	drew=5	lost=7	goalsFired=88	goalsAcquired=34	goalDifference=54	points=83
3	Chelsea	played=38	won=20	drew=9	lost=9	goalsFired=63	goalsAcquired=42	goalDifference=21	points=69
4	Tottenham	played=38	won=20	drew=8	lost=10	goalsFired=60	goalsAcquired=42	goalDifference=18	points=68
5	Man United	played=38	won=19	drew=10	lost=9	goalsFired=54	goalsAcquired=32	goalDifference=22	points=67
6	Leicester	played=38	won=17	drew=11	lost=10	goalsFired=65	goalsAcquired=36	goalDifference=29	points=62
7	Sheffield United	played=38	won=15	drew=13	lost=10	goalsFired=34	goalsAcquired=28	goalDifference=6	points=58
8	Arsenal	played=38	won=13	drew=17	lost=8	goalsFired=54	goalsAcquired=45	goalDifference=9	points=56
9	Wolves	played=38	won=12	drew=19	lost=7	goalsFired=48	goalsAcquired=40	goalDifference=8	points=55
10	Everton	played=38	won=12	drew=12	lost=14	goalsFired=46	goalsAcquired=55	goalDifference=-9	points=48
11	Burnley	played=38	won=13	drew=8	lost=17	goalsFired=38	goalsAcquired=49	goalDifference=-11	points=47
12	Crystal Palace	played=38	won=11	drew=13	lost=14	goalsFired=31	goalsAcquired=41	goalDifference=-10	points=46
13	Southampton	played=38	won=11	drew=9	lost=18	goalsFired=41	goalsAcquired=61	goalDifference=-20	points=42
14	Newcastle	played=38	won=10	drew=11	lost=17	goalsFired=28	goalsAcquired=49	goalDifference=-21	points=41
15	West Ham	played=38	won=8	drew=12	lost=18	goalsFired=43	goalsAcquired=60	goalDifference=-17	points=36
16	Watford	played=38	won=8	drew=12	lost=18	goalsFired=32	goalsAcquired=54	goalDifference=-22	points=36
17	Brighton	played=38	won=7	drew=13	lost=18	goalsFired=35	goalsAcquired=49	goalDifference=-14	points=34
18	Bournemouth	played=38	won=7	drew=11	lost=20	goalsFired=36	goalsAcquired=61	goalDifference=-25	points=32
19	Norwich	played=38	won=5	drew=11	lost=22	goalsFired=30	goalsAcquired=65	goalDifference=-35	points=26
20	Aston Villa	played=38	won=7	drew=5	lost=26	goalsFired=36	goalsAcquired=73	goalDifference=-37	points=26

Simulation of the Premier League Considering data from only 2019-2020 season:

1	Liverpool	played=38	won=35	drew=2	lost=1	goalsFired=82	goalsAcquired=23	goalDifference=59	points=107
2	Man City	played=38	won=26	drew=5	lost=7	goalsFired=88	goalsAcquired=34	goalDifference=54	points=83
3	Leicester	played=38	won=24	drew=6	lost=8	goalsFired=70	goalsAcquired=30	goalDifference=40	points=78
4	Chelsea	played=38	won=19	drew=9	lost=10	goalsFired=64	goalsAcquired=47	goalDifference=17	points=66
5	Wolves	played=38	won=15	drew=17	lost=6	goalsFired=52	goalsAcquired=40	goalDifference=12	points=62
6	Sheffield United	played=38	won=16	drew=12	lost=10	goalsFired=36	goalsAcquired=29	goalDifference=7	points=60
7	Man United	played=38	won=16	drew=11	lost=11	goalsFired=51	goalsAcquired=34	goalDifference=17	points=59
8	Tottenham	played=38	won=15	drew=12	lost=11	goalsFired=55	goalsAcquired=44	goalDifference=11	points=57
9	Arsenal	played=38	won=12	drew=16	lost=10	goalsFired=47	goalsAcquired=45	goalDifference=2	points=52
10	Burnley	played=38	won=12	drew=11	lost=15	goalsFired=39	goalsAcquired=49	goalDifference=-10	points=47
11	Everton	played=38	won=12	drew=10	lost=16	goalsFired=43	goalsAcquired=54	goalDifference=-11	points=46
12	Crystal Palace	played=38	won=10	drew=14	lost=14	goalsFired=28	goalsAcquired=39	goalDifference=-11	points=44
13	Newcastle	played=38	won=10	drew=12	lost=16	goalsFired=27	goalsAcquired=48	goalDifference=-21	points=42
14	Southampton	played=38	won=11	drew=8	lost=19	goalsFired=41	goalsAcquired=63	goalDifference=-22	points=41
15	Brighton	played=38	won=8	drew=15	lost=15	goalsFired=38	goalsAcquired=48	goalDifference=-10	points=39
16	Watford	played=38	won=8	drew=11	lost=19	goalsFired=31	goalsAcquired=53	goalDifference=-22	points=35
17	West Ham	played=38	won=8	drew=10	lost=20	goalsFired=42	goalsAcquired=63	goalDifference=-21	points=34
18	Bournemouth	played=38	won=7	drew=10	lost=21	goalsFired=33	goalsAcquired=59	goalDifference=-26	points=31
19	Aston Villa	played=38	won=7	drew=7	lost=24	goalsFired=40	goalsAcquired=72	goalDifference=-32	points=28
20	Norwich	played=38	won=5	drew=10	lost=23	goalsFired=30	goalsAcquired=63	goalDifference=-33	points=25

OBSERVATIONS:

1. Newly Promoted teams have less data to predict. Promoted Teams that have performed well this season will have better win probabilities. But, considering real life seasons, consistency to perform is also important and we don't know how they would perform against a strong team considering that the fixture is to be played.
2. Teams that have played in the premier league for all 5 seasons have more data and hence their form is fine tuned and by using more data, we can know how consistent the team's performance is.
3. For Simulating the current premier league table, though the form of a team is fine tuned by considering past previous seasons, the form at present is of utmost importance as the table is calculated only based on how well you perform in the current season rather than in previous seasons.
   For example: Leicester city's overall performance across the five seasons is average (1st in 2015/2016, 12th in 2016/2017, 9th in 2017/2018,9th in 2018/2019, currently 3rd). So, when we predict a game vs a strong opponent that has finished in the top four in previous seasons (Tottenham) but are struggling now, Leicester city's chances of winning may look low although they are pretty good this season.
   When predicting Leicester's position considering data from five seasons, Leicester finishes 6th, but considering only this season Leicester finishes 3rd.
4. For Simulating the premier league, we consider the score to have the highest probability. In majority of the cases when the win percentage is higher, the score will be a win. But in some cases, though the win percentage is higher, the score with the highest probability maybe a draw.
   For example: Wolves vs Everton:

```
Wolves winProbability: 47.152422792408984, draw Probability:
25.997704073198104, Everton winProbability: 26.345517398670427
predicted scores:  first: 1-1 second: 1-0 third: 2-1
```

CONCLUSIONS AND FURTHER IMPROVEMENTS:

- We have Successfully implemented a ranking system for the premier league which predicts matches between two teams using Poisson distribution along with considering their home and away form and simulates the remaining fixtures and finally give out the simulated premier league table.
- Final Premier league table gives us pretty good estimate of how the premier league table would look like for the 2019-2020 season.
- Predicting the result of a soccer match does not only depend on the ability to score goals, but also the form of the players, tactical changes in the team, red card and yellow card bookings that could change the game. So, this is one of the disadvantages of using Poisson's Distribution.
- We can also use concepts of machine learning, to constantly update the form of the team after a prediction.

REFERENCES:

1. https://help.smarkets.com/hc/en-gb/articles/115001457989-How-to-calculate-Poisson-distribution-for-football-betting
2. https://www.hackerearth.com/blog/developers/football-betting-odds-work-using-poisson-distribution/
3. https://towardsdatascience.com/the-poisson-distribution-and-poisson-process-explained-4e2cb17d459
4. https://talksport.com/football/689392/opta-premier-league-simulated-table-liverpool-man-city-tottenham-arsenal-brighton-west-ham/#Echobox=1586086157