Final Project Report

Narek Sahakyan

December 5, 2019

The goal of this small project is to analyze the play off round games in champions league starting from its founding season. Except the usual ways of determining a winner in a 2 match leg there are two more rules to determine who goes to the next round, one of them is away goal rule, according which is applied in cases when the teams have the same amount of goals scored against each other in 2 games, in such cases if one team has scored more goals while playing away it goes forward. When even the number of goals scored in away games is the same for both teams, the teams play an extra time in the second game of the leg I believe that the away goal rule strongly affects the quality of the game and teams tend to play more deffensive and static, in order to not lose their home advantage. The same applies to extra times, as the teams are tired and the weaker ones tend to build strong deffensive structure to last the 30 minutes and then win the game by penalties. Through this project, I want to find out if some adjustements to these rules and situations would change the picture. One modification is "cancellation" of away goal rule, let's suppose that if teams have the same amount goals, but one of them has scored more, than a game in neutral stadium would be played, so that neither would have a home advantage. The modification applied to extra time games is an old cancelled rule. It is the famous rule of golden goal rule according to which the team, that is first to score in the extra time is the winner.

I have three main data frames, that data of which was initially scrapped and later cleaned and prepared for analysis.

This is the general overview of the format of games table. Except the obvious variables it also contains some other variables which were used for the analysis.

COMP	SEASON	ROUND	LEG	DATE	HOMETEAM	AWAYTEAM	FTHG	FTAG
UCL	1994-1995	R8	1	1995-03-01	FC Barcelona	Paris Saint-Germain	1	1
UCL	1994-1995	R8	2	1995-03-15	Paris Saint-Germain	FC Barcelona	2	1
UCL	1994-1995	R8	1	1995-03-01	AC Milan	SL Benfica	2	0
UCL	1994-1995	R8	2	1995-03-15	SL Benfica	AC Milan	0	0
UCL	1995-1996	R8	1	1996-03-05	Borussia Dortmund	AFC Ajax	0	2
UCL	1995-1996	R8	2	1996-03-20	AFC Ajax	Borussia Dortmund	1	0
UCL	1995-1996	R8	1	1996-03-05	FC Nantes	Spartak Moskva	2	0
UCL	1995-1996	R8	2	1996-03-20	Spartak Moskva	FC Nantes	2	2
UCL	1995-1996	R8	1	1996-03-05	Real Madrid	Juventus	1	0
UCL	1995-1996	R8	2	1996-03-20	Juventus	Real Madrid	2	0

The legs information data frame contains information about the leg's games. The dates for both games, the winner, the round, the number of goals scored by both teams and the rule through which the winner was determined.

COMP	SEASON	ROUND	TEAM1	TEAM2	WINNER
UCL	1996-1997	R8	AFC Ajax	Atlético Madrid	AFC Ajax
UCL	1997-1998	R8	Bayern München	Borussia Dortmund	Borussia Dortmund
UCL	1999-2000	R8	Chelsea FC	FC Barcelona	FC Barcelona
UCL	2002-2003	R8	Juventus	FC Barcelona	Juventus
UCL	2004-2005	R8	Olympique Lyon	PSV Eindhoven	PSV Eindhoven
UCL	2004-2005	R8	AC Milan	Inter	AC Milan
UCL	2006-2007	R4	Chelsea FC	Liverpool FC	Liverpool FC

COMP	SEASON	ROUND	TEAM1	TEAM2	WINNER
UCL	2007-2008	R16	FC Schalke 04	FC Porto	FC Schalke 04
UCL	2007-2008	R16	Fenerbahçe	Sevilla FC	Fenerbahçe
UCL	2008-2009	R16	Arsenal FC	AS Roma	Arsenal FC

I also got the goal scoring minutes of the games that ended up in an extra time.

| HOMETEAM	AWAYTEAM	HG	AG	MINUTE	WS	LEG_ID	GAME_ID
APOEL Nikosia	Olympique Lyon	1	0	9	APOEL Nikosia	ET11	22
Chelsea FC	SSC Napoli	1	0	28	Chelsea FC	ET12	24
Chelsea FC	SSC Napoli	2	0	47	Chelsea FC	ET12	24
Chelsea FC	SSC Napoli	2	1	55	SSC Napoli	ET12	24
Chelsea FC	SSC Napoli	3	1	75	Chelsea FC	ET12	24
Chelsea FC	SSC Napoli	4	1	105	Chelsea FC	ET12	24
Chelsea FC	Paris Saint-Germain	1	0	81	Chelsea FC	ET13	26
Chelsea FC	Paris Saint-Germain	1	1	86	Paris Saint-Germain	ET13	26
Chelsea FC	Paris Saint-Germain	2	1	96	Chelsea FC	ET13	26
Chelsea FC	Paris Saint-Germain	2	2	114	Paris Saint-Germain	ET13	26

DETERMINING LEG'S WINNER IN UCL PO ROUNDS
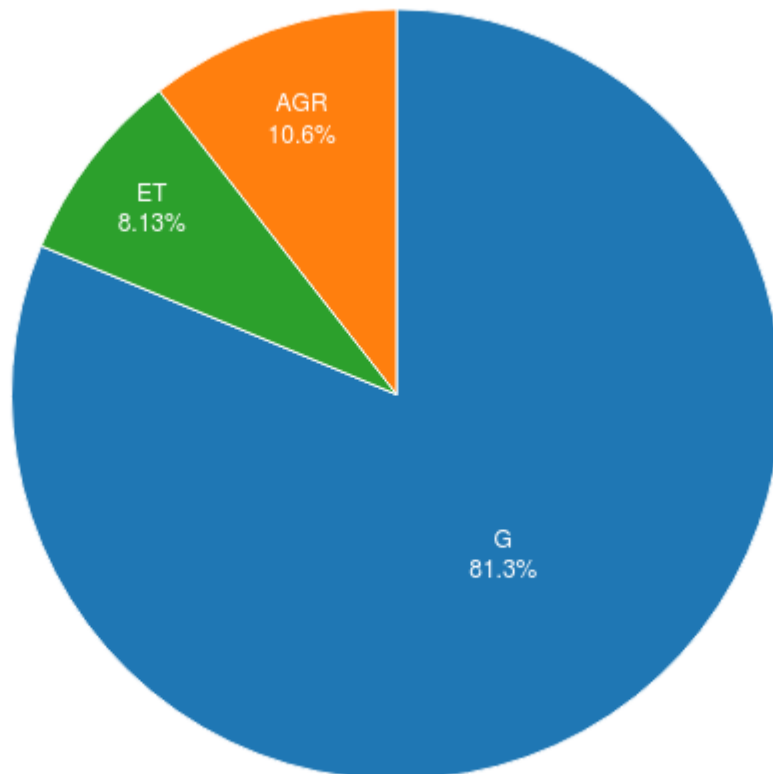
As expected majority of the leg winners were decided through the standard rule of who scored more We can see there were more legs where the away goal rule was applied and the application of extra time was a rare event in playoff rounds, occuring only in 8 percent of the legs. Having around 500 games in the playoff

rounds, divided equally into 250 legs, we can observe that the results of around 50 of the legs were "suspicios" and "not expected" In order to predict the expected winners of the legs, in which away goal rule was used, I will get the elo ratings of the teams based on the second game's date and won't treat any of them as a home team. For the games in extra time, I will apply golden goal rule to see if it would change anything.

NUMBER OF PO MATCHES BY TEAMS
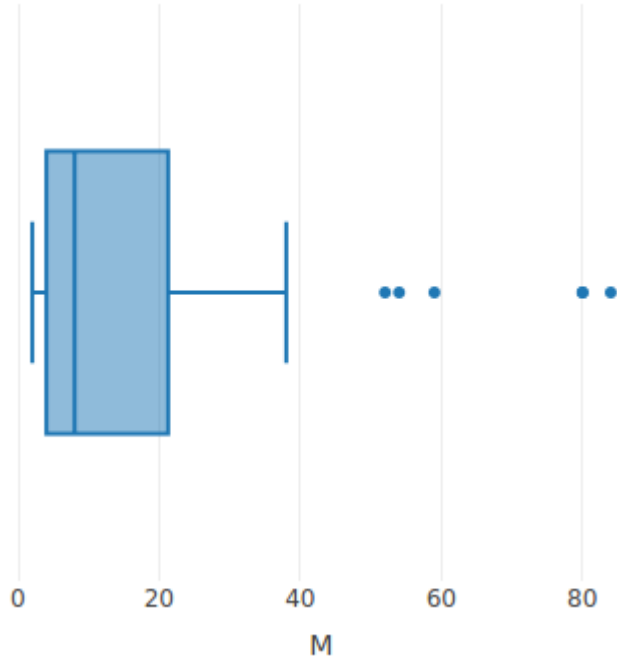
As we can see there are only a few teams, that have a high number of games played in the play off round. Those teams are Real Madrid,Fc Barcelona, Bayern Munich, MU and teams of similar ranking. There is no outlying behavior in the amount of games played by the other teams as except top teams, most of the teams in the play off round are usually repeating with some different teams appearing in high stages of the championship by chance or by good draw. Let's build a standings table for the teams based on all games that they played in the playoff rounds In order to have meaningful predictions, I have removed some teams played in less than one playoff leg over those 25 years, and unfortunately Lazio was among those :(, but Mr.Madoyan please don't leave the paper as you will enjoy the later foundings.I also removed the teams, for which I was not able to get the elo ratings from the data I managed to find. So this is how the CL playoff rounds table looks like

TEAM	M	POINTS	W	D	L	FTGF	FTGA	WPCT	EWPCT
Bayern München	80	140	40	20	20	152	95	0.5875000	0.6466440
FC Barcelona	80	139	39	22	19	135	87	0.5750000	0.6375919
Real Madrid	84	139	41	16	27	144	99	0.5476190	0.6181663
Manchester United	59	95	26	17	16	88	64	0.5423729	0.6009585
Juventus	54	87	23	18	13	82	56	0.5370370	0.6201897
Chelsea FC	52	68	16	20	16	67	64	0.4423077	0.5147210
Liverpool FC	33	66	19	9	5	64	33	0.6666667	0.7009167
AC Milan	38	51	13	12	13	46	46	0.4473684	0.5000000
Atlético Madrid	30	46	11	13	6	32	25	0.5000000	0.5786922
Arsenal FC	36	41	11	8	17	43	56	0.3888889	0.4158969
Borussia Dortmund	28	35	10	5	13	33	36	0.4285714	0.4720598
Paris Saint-Germain	26	34	9	7	10	37	40	0.4230769	0.4749607
FC Porto	30	33	8	9	13	28	48	0.3666667	0.3333595

TEAM	M	POINTS	W	D	L	FTGF	FTGA	WPCT	EWPCT
AFC Ajax	21	31	8	7	6	32	23	0.4761905	0.6045876
Inter	26	31	9	4	13	22	33	0.3846154	0.3725395

As the table clearly shows, most of the top teams were expected to win more games based on their goal differences.

```
##
##  Pearson's product-moment correlation
##
## data:  wpct_table_1$WPCT and wpct_table_1$FTGD
## t = 4.4782, df = 44, p-value = 5.285e-05
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3214546 0.7310840
## sample estimates:
##       cor
## 0.5595379
```

As we can see the correlation between Win Percentage and Goal Difference exists, but is not that high compared to previous measurements done during class practice. For example in La Liga this coefficient was around 0.87. This is a good point for starting to make some conclusions. As the Goal Difference is mostly used for measuring team's "ranking" and "class". Having this information, let's analyze this connection in Champions League Play off rounds. Let's make very rough estimations. Having a correlation coefficient of 0.55 we can have general assumptions that only about the half of the teams get what they deserve based on their "class"

TEAM	M	POINTS	W	D	L	FTGF	FTGA	WPCT	EWPCT
FC Porto	30	33	8	9	13	28	48	0.3666667	0.3333595
Inter	26	31	9	4	13	22	33	0.3846154	0.3725395
AS Roma	20	24	8	0	12	25	40	0.4000000	0.3533560
FC Schalke 04	18	17	5	2	11	24	42	0.3333333	0.3274941
Tottenham Hotspur	12	17	5	2	5	15	16	0.5000000	0.4792664
SL Benfica	14	15	4	3	7	11	18	0.3571429	0.3467819
Galatasaray	8	11	3	2	3	11	16	0.5000000	0.3818337
Panathinaikos	8	11	3	2	3	8	9	0.5000000	0.4622115
Shakhtar Donetsk	10	11	3	2	5	11	22	0.4000000	0.2908555
Girondins Bordeaux	4	9	3	0	1	5	4	0.7500000	0.5712402

Those are the lucky teams. I hope you find some interesting names here :).

TEAM	M	POINTS	W	D	L	FTGF	FTGA	WPCT	EWPCT
Bayern München	80	140	40	20	20	152	95	0.5875000	0.6466440
FC Barcelona	80	139	39	22	19	135	87	0.5750000	0.6375919
Real Madrid	84	139	41	16	27	144	99	0.5476190	0.6181663
Manchester United	59	95	26	17	16	88	64	0.5423729	0.6009585
Juventus	54	87	23	18	13	82	56	0.5370370	0.6201897
Chelsea FC	52	68	16	20	16	67	64	0.4423077	0.5147210
Liverpool FC	33	66	19	9	5	64	33	0.6666667	0.7009167
AC Milan	38	51	13	12	13	46	46	0.4473684	0.5000000

| TEAM	M	POINTS	W	D	L	FTGF	FTGA	WPCT	EWPCT
Arsenal FC	36	41	11	8	17	43	56	0.3888889	0.4158969

EXPECTED WIN PERCENTAGE VS ACTUAL
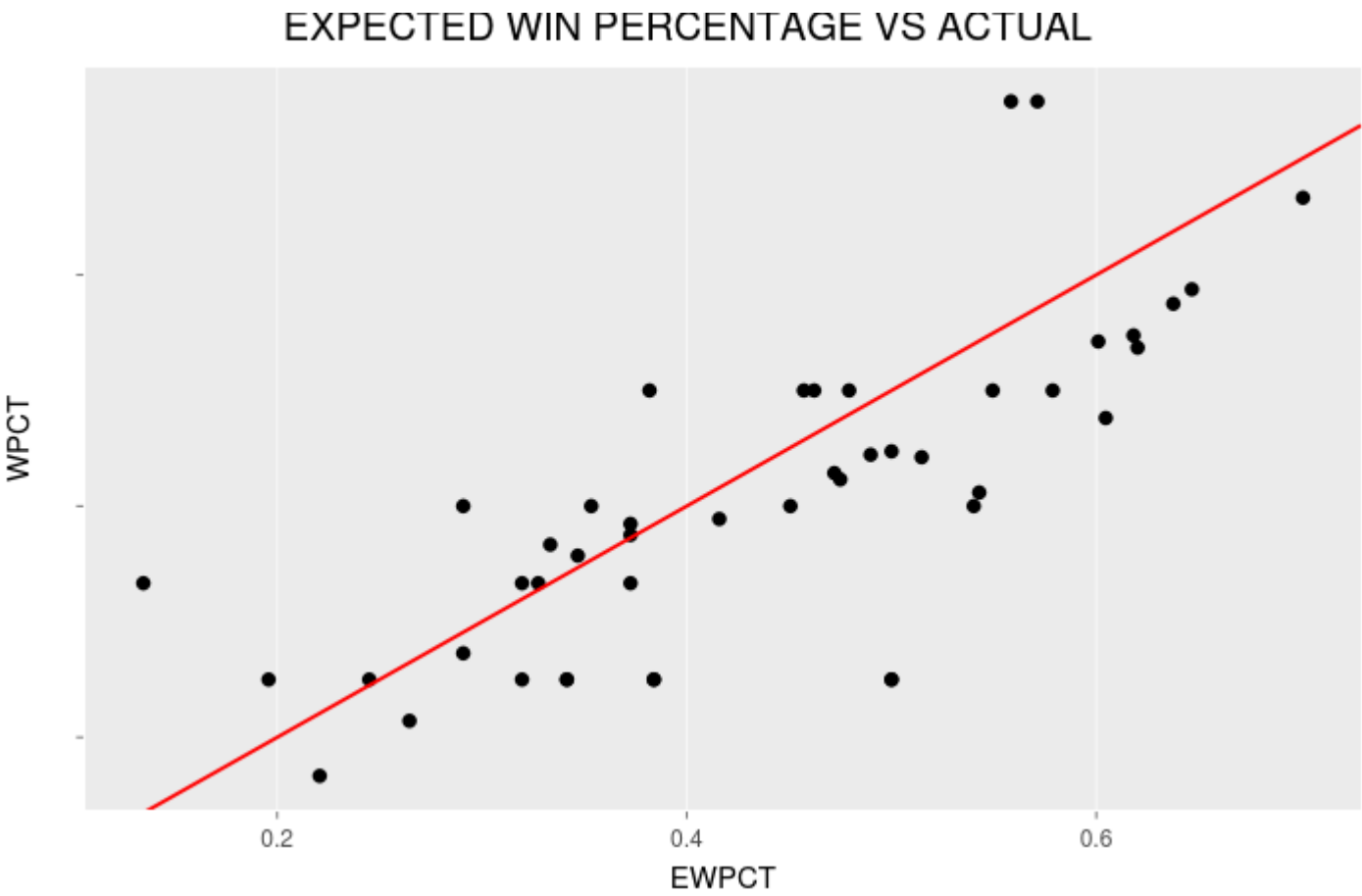
This visualization shows the overall picture. As we can see there are more teams who win less than they were expected rather than teams that win more than expected

GOAL SCORING MINUTES IN ALL ROUNDS OF PO ET GAMES
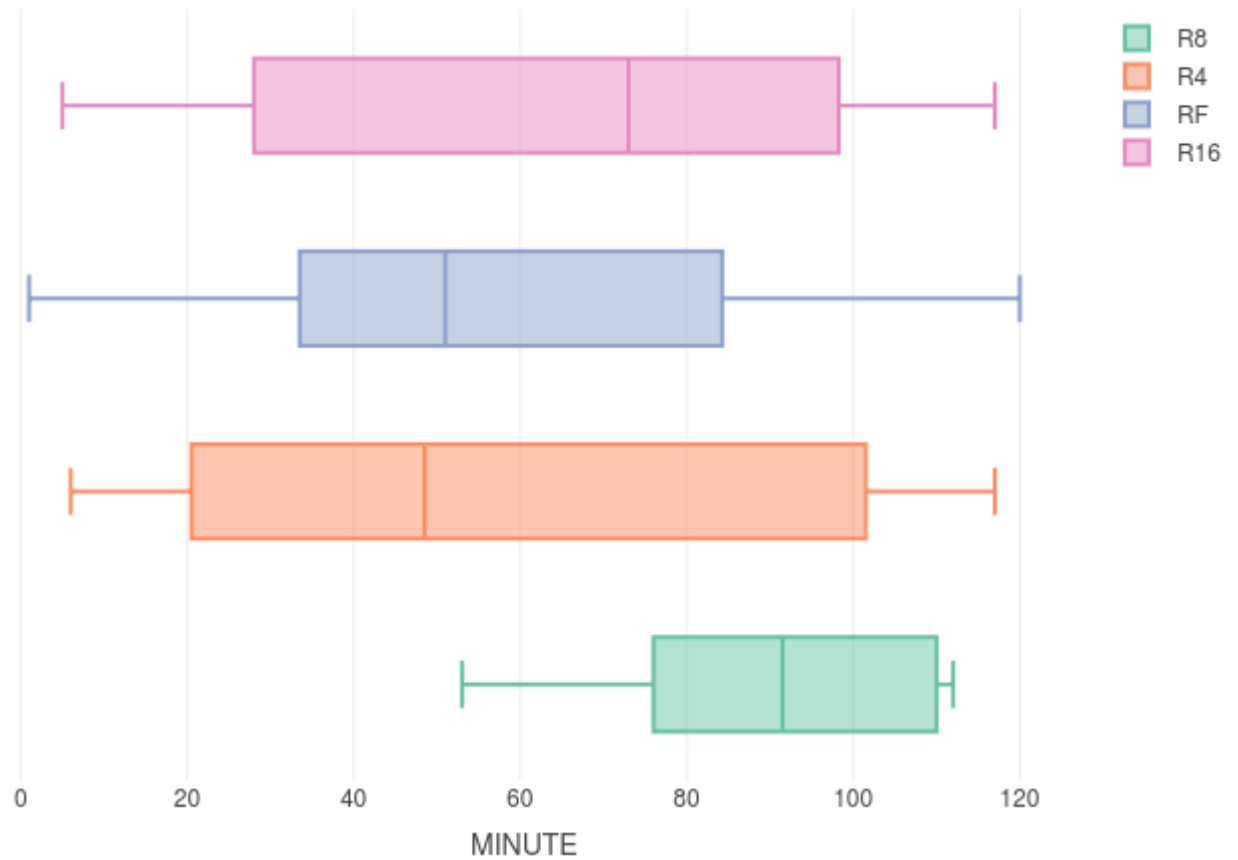
As we can see the goals scored in the late stages of the games are mostly occuring in the extra time games of quarter finals, where most of the goals occur in minutes from around 75 to 110, in this round the first goals are usually scored after the first half. This round is the only one having this property Semi final games also had some late goals, but most of them before 100th minute and earliest of them in the first half beginning periods. In the final games, which ended up in an extra time, most of them went to penalties as the teams play extra "carefully" in the extra time period of final games. Some teams can get to the round of 16 by chance, and win their leg by "killing" the game. Those "lucky" teams can still occur in the quarter finals, but the chance is lower as they are mostly expected to be beaten by strong teams in the round of 16. The "lucky" teams mostly get to higher stages due to a good draw. To cap it all the density of goals scored in the late stages of the game is highest in R4 extra time games. As this is the round where most of the "lucky" and game "killing" teams are mostly eliminated, but can still occur.

Histogram for Goal Minutes
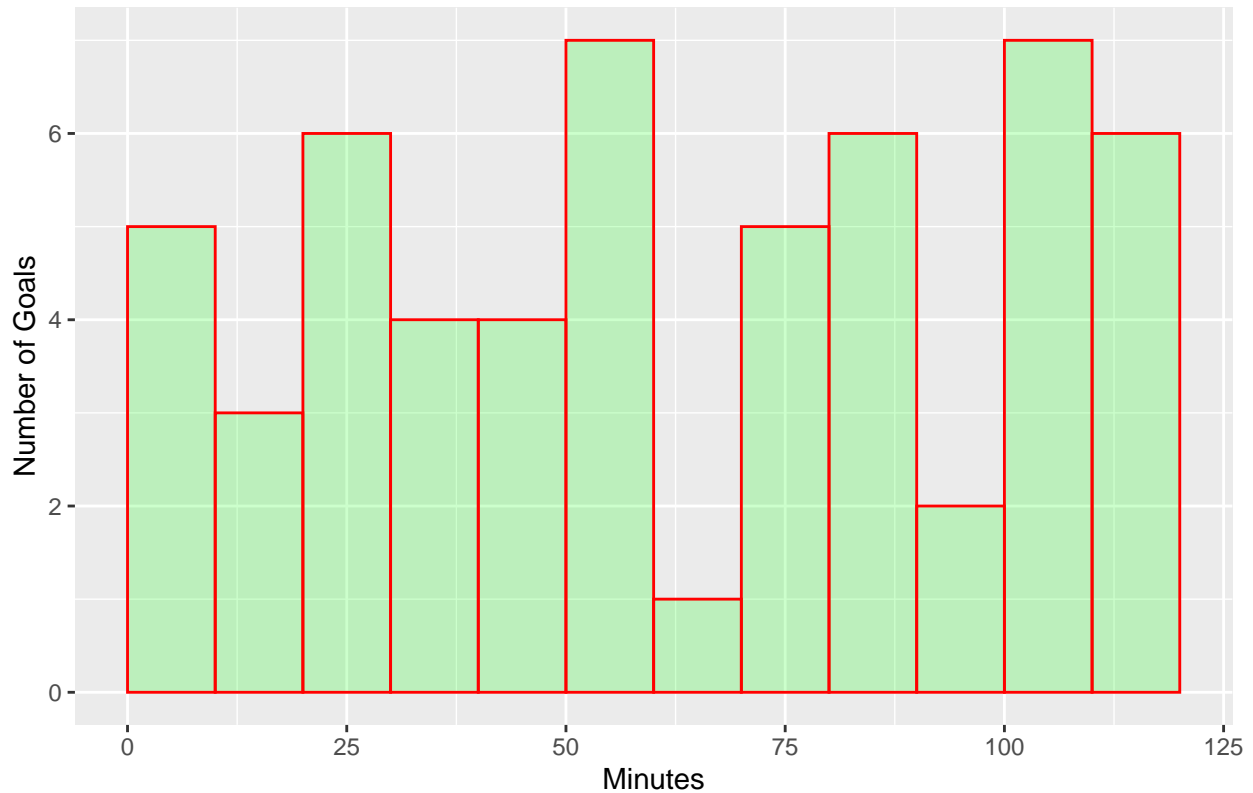
The histogram shows the general picture of the goal's scoring minutes. The highest frequencies of goals in extra time occur in the following ranges of the game time, 50 - 60 (second half "game engine" time as some time is passed from the start and there is still plenty of time left till the final whistle), 100 - 110, 110 -120(intrigue killing goals, as the times are completely tired and almost no time is left), 70 - 80(bringing intrigue as there is a chance for a comeback for the team that allowed the goal), 80 - 90 (intrigue killing goals or goals leading to extra time)

HOMETEAM	AWAYTEAM	ROUND	GGR.WINER	WINNER	MINUTE
Chelsea FC	SSC Napoli	R16	Chelsea FC	Chelsea FC	105
Chelsea FC	Paris Saint-Germain	R16	Chelsea FC	Chelsea FC	96
Bayern München	Juventus	R16	Bayern München	Bayern München	108
FC Porto	AS Roma	R16	FC Porto	FC Porto	117
Real Madrid	Bayern München	R8	Real Madrid	Real Madrid	105
Chelsea FC	Liverpool FC	R4	Chelsea FC	Chelsea FC	98
Real Madrid	Atlético Madrid	RF	Real Madrid	Real Madrid	110

As we can see, we have only rare cases of games in extra time where the Golden Goal rule can be applied and it wont change anything, as the actual winners won't change for any of those 7 games. Maybe people in UEFA knew something, when they cancelled the rule :) Now let's apply the modification for away goal rule and simulate neutral games among the teams.

TEAM1	TEAM2	ACTUAL_WINNER	NG_WINNER	WINS.1	WINS.2
AC Milan	Inter	AC Milan	Inter	0.4913662	0.5086338
Real Madrid	AS Monaco	AS Monaco	Real Madrid	0.5924662	0.4075338
Inter	Villarreal CF	Villarreal CF	Inter	0.6413901	0.3586099

TEAM1	TEAM2	ACTUAL_WINNER	NG_WINNER	WINS.1	WINS.2
Olympique Marseille	Inter	Olympique Marseille	Inter	0.4612219	0.5387781
Shakhtar Donetsk	AS Roma	AS Roma	Shakhtar Donetsk	0.6212967	0.3787033
Manchester United	Paris Saint-Germain	Manchester United	Paris Saint-Germain	0.1961815	0.8038185
Paris Saint-Germain	Chelsea FC	Chelsea FC	Paris Saint-Germain	0.8399507	0.1600493
FC Barcelona	AS Roma	AS Roma	FC Barcelona	0.7105400	0.2894600
Tottenham Hotspur	Manchester City	Tottenham Hotspur	Manchester City	0.4299472	0.5700528
Atlético Madrid	Bayern München	Atlético Madrid	Bayern München	0.4899276	0.5100724
Tottenham Hotspur	AFC Ajax	Tottenham Hotspur	AFC Ajax	0.3679300	0.6320700

This modification changed the outcome of 11 legs of away goal rule. Having in general 46 legs with away goal rule application we can observe that 23 % of the game outcomes could have been different, which in case could change the later game results too.

```
## [1] 0.2389815


## Confusion Matrix and Statistics
##
##           Reference
## Prediction 0 1
##          0 7 5
##          1 6 7
##
##                Accuracy : 0.56
##                  95% CI : (0.3493, 0.756)
##     No Information Rate : 0.52
##     P-Value [Acc > NIR] : 0.422
##
##                   Kappa : 0.1214
##
##  Mcnemar's Test P-Value : 1.000
##
##             Sensitivity : 0.5385
##             Specificity : 0.5833
##          Pos Pred Value : 0.5833
##          Neg Pred Value : 0.5385
##              Prevalence : 0.5200
##          Detection Rate : 0.2800
##    Detection Prevalence : 0.4800
##       Balanced Accuracy : 0.5609
##
##        'Positive' Class : 0
##
```

The brier score and the confustion matrix metrics, show that the predictions based on the team's ratings are not highly accurate for champions league play off rounds games, meaning that most of them are unpredictible, which is partly the answer for my research question. As most of the games are unpredictible it means that the strong teams applying their full potential of skills can lose to weaker teams in terms of both game metrics and UCL current rules. Golden Goal rule won't change anything, but the Neutral Game rule has a potential of making some huge changes in the league.

TO CONCLUDE

This was a little project, the reasons behind which were mostly enthusiastic rather than scientific. I spent most of the time on scrapping, cleaning and preparing the data for analysis. However, as the scope of the games that fit my project goal(Extra time legs or Away goal rule application), the results are based on a small set of observations. If I had enough time and opportunity, I would include more similar games to have higher set of information. Also, in case of time availability I would get more information on goal scoring minutes, to build flexible models for estimating the teams chances to winbased on the time when it allowed a goal, in order to have more observations for golden goal rule application. Nevertheless, I enjoyed the time spent on the project and the whole class, and I will surely dive deeper into similar projects.