

Will Arsenal finish first?



Arsenal London is a football club that has a huge chance to win Premier League in the season 2022/2023. In this paper, I will try to explain why this London club has progressed so much during the last few months. For analysis data from fbref.com is used.

In my analysis I would like to answer the following questions:

1. What are the reasons that particular clubs became champions in the past?
2. How did Arsenal progress during the last 5 seasons?
3. Will Arsenal be the next Premier League champion?

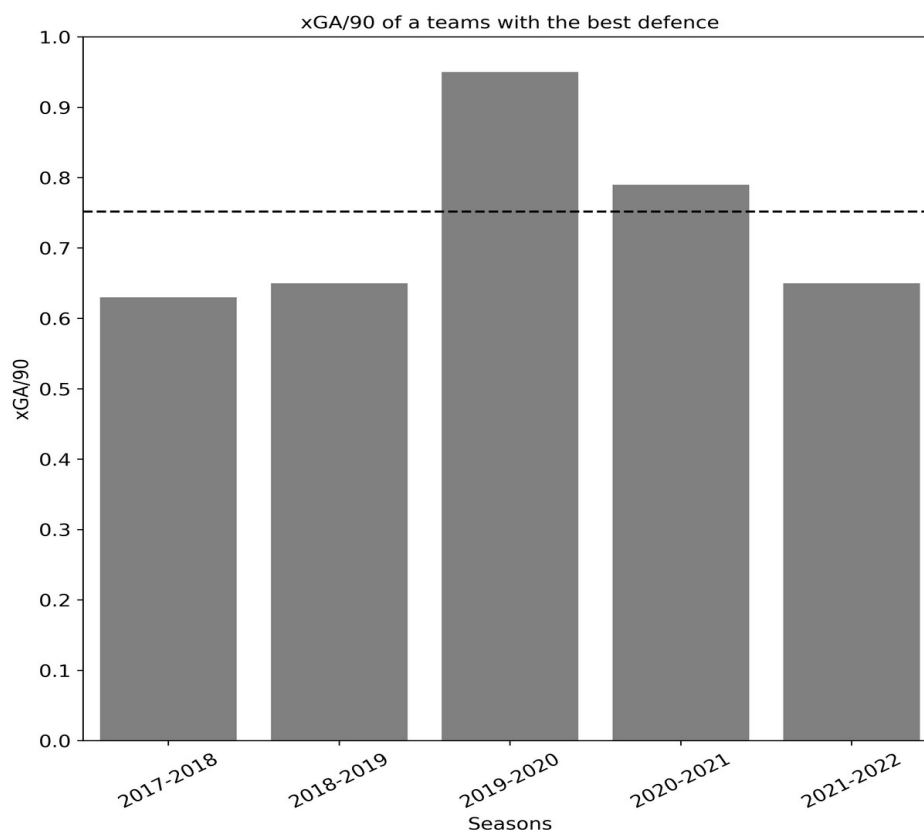
The general_data data frame has been used (sample below):

Rk	Squad	MP	W	D	L	GF	GA	GD	Pts	Pts/MP	xG	xGA	xGD	xGD/90	Last 5	Attendance	Top Team Scorer	Goalkeeper	Notes	xG/90	xGA/90	season
8	9Brentford	25	9	11	5	40	33	7	38	1.52	36.6	31.4	5.3	0.21	W D D W L	17076	Ivan Toney - 15	David Raya		1.46	1.26	2022-2023
0	1Arsenal	27	21	3	3	62	25	37	66	2.44	52.2	26.3	26	0.96	W W W W W	60196	Martinelli - 12	Aaron Ramsdale		1.93	0.97	2022-2023
7	8Fulham	27	11	6	10	38	37	1	39	1.44	33.7	44.9	-11.3	-0.42	W W D L L	23502	Aleksandar Mitrović - 11	Bernd Leno		1.25	1.66	2022-2023
1	2Manchester City	27	19	4	4	67	25	42	61	2.26	56.1	21.5	34.6	1.28	W D W W W	53219	Erling Haaland - 28	Ederson		2.08	0.8	2022-2023
15	16Leicester City	26	7	3	16	37	46	-9	24	0.92	32.9	43.1	-10.1	-0.39	W L L L L	31772	James Maddison - 9	Danny Ward		1.27	1.66	2022-2023

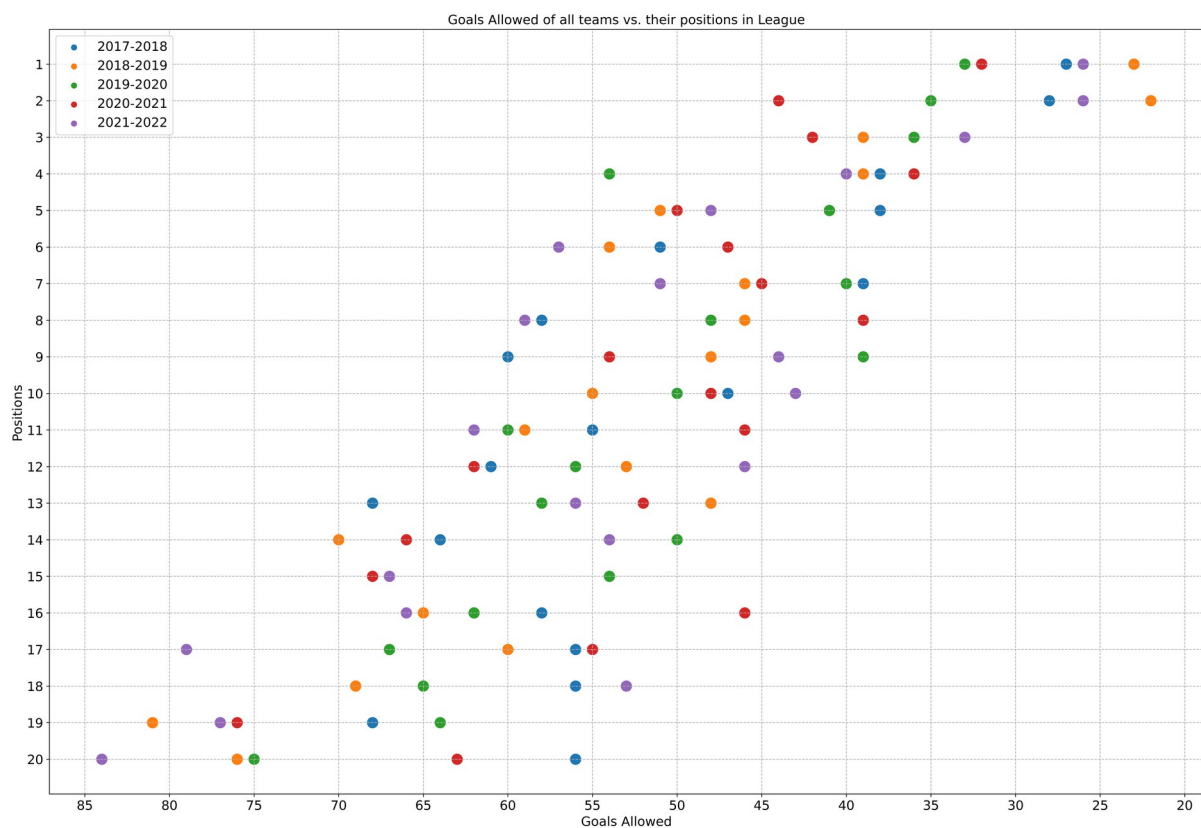
1. How do teams become champions?

Among all football proverbs the most overused one is that 'Attack wins you games, defence wins you title. Let's check then if it is true in the context of the Premier League.

Firstly, we should look at the defence of champions from season 2017 to 2022. To compare it we will use xGA/90 (expected goals allowed per match) and GA (goals allowed) statistics. Based on the discussed proverb, the best team defensively should always (or at least in most cases) win the title. Let's have a look at the below diagram and find out how to be the most effective defence in Premier League.

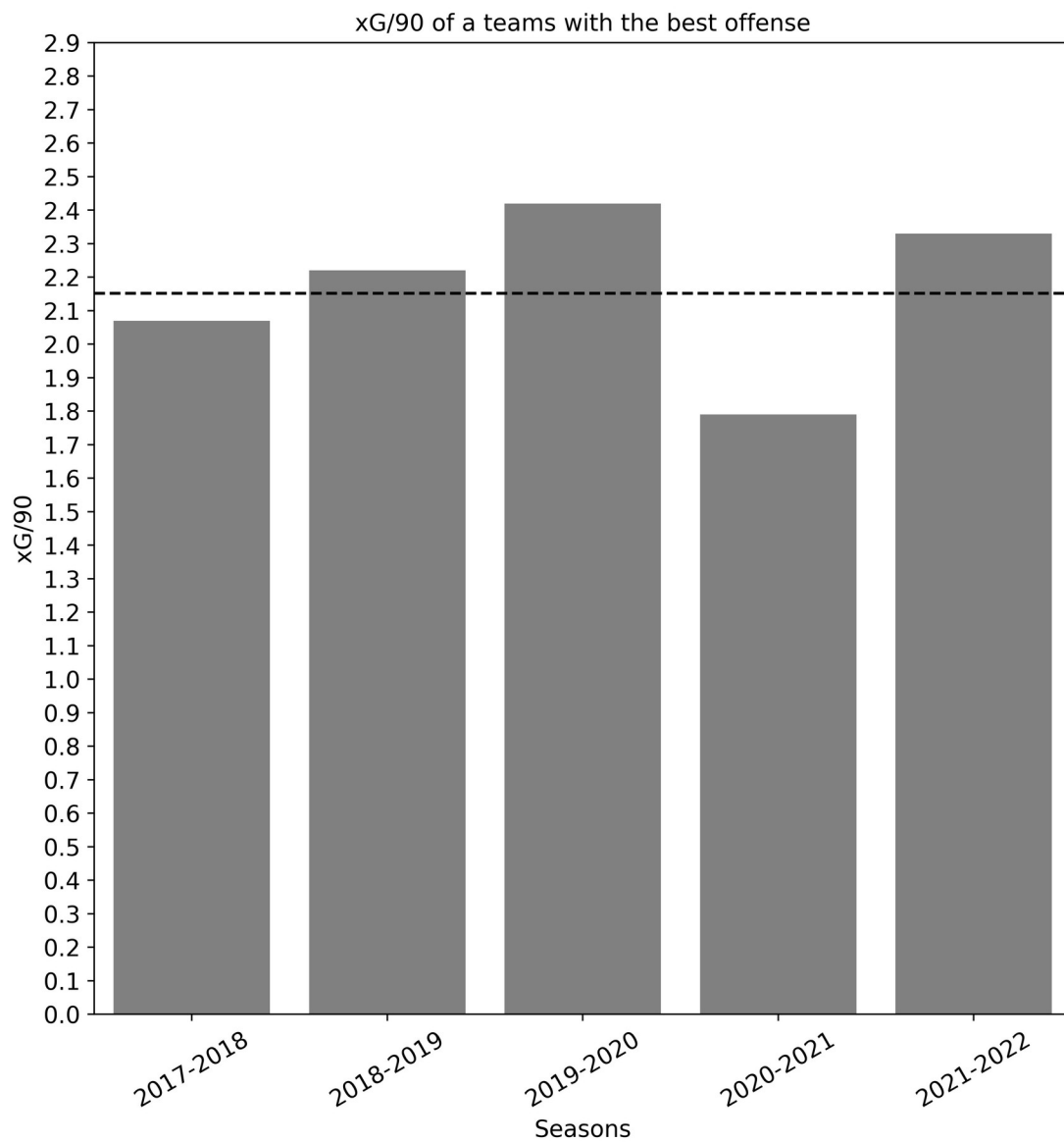


Based on the above diagram, to become the best defence in a league team needs to have an expected goal allowed per 90 minutes statistic on an average level of 0.75. It means that squad that is the best in a league allows for less than 1 offensive action that for 100% ends with a goal. What is even more interesting, during the last 5 seasons Manchester City won the title of the best defence in the league. Good job Pep. Nevertheless, let's see if it always guaranteed a trophy:

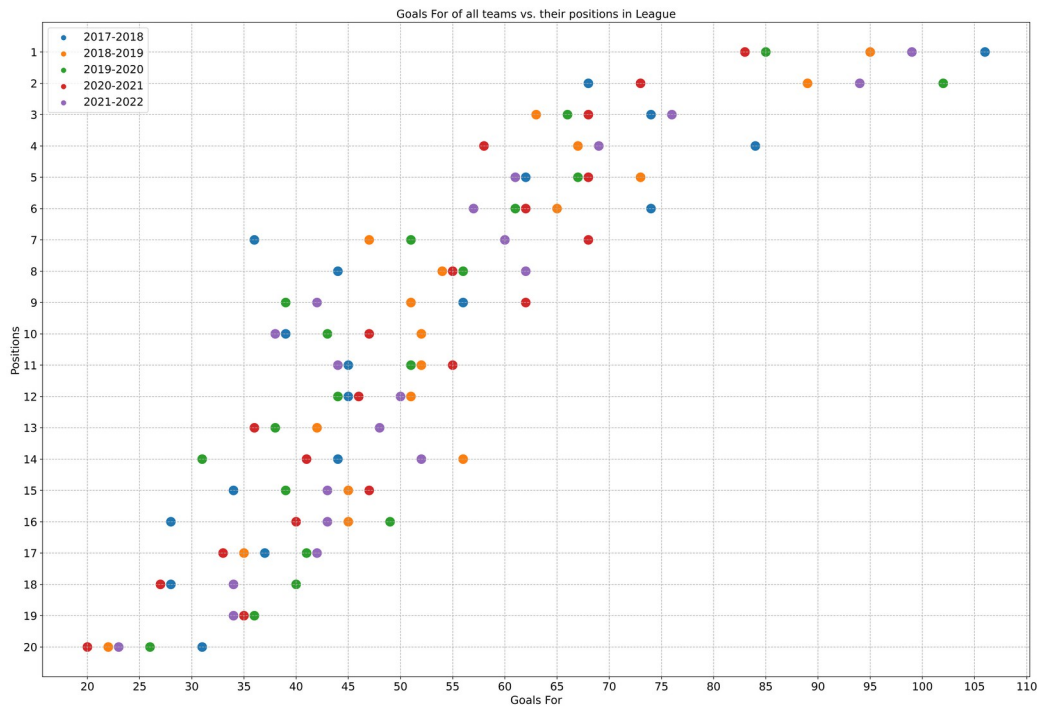


As you can see in 4 out of 5 seasons, the best defensive team won a title. The conclusion is that the best defence does not guarantee a championship, but it significantly increases the probability of winning it.

We went through the defensive aspects of a game, so now it is time to compare it with the offensive. During the last seasons a lot of articles, tv programs etc. were focused on the rivalry between Liverpool and Manchester City, so it is time to compare it with statistics. This time we will take a look at xG/90 (expected goals per 90 mins) of the best offensive team in each season and the GF (Goals For) of all teams in the League.



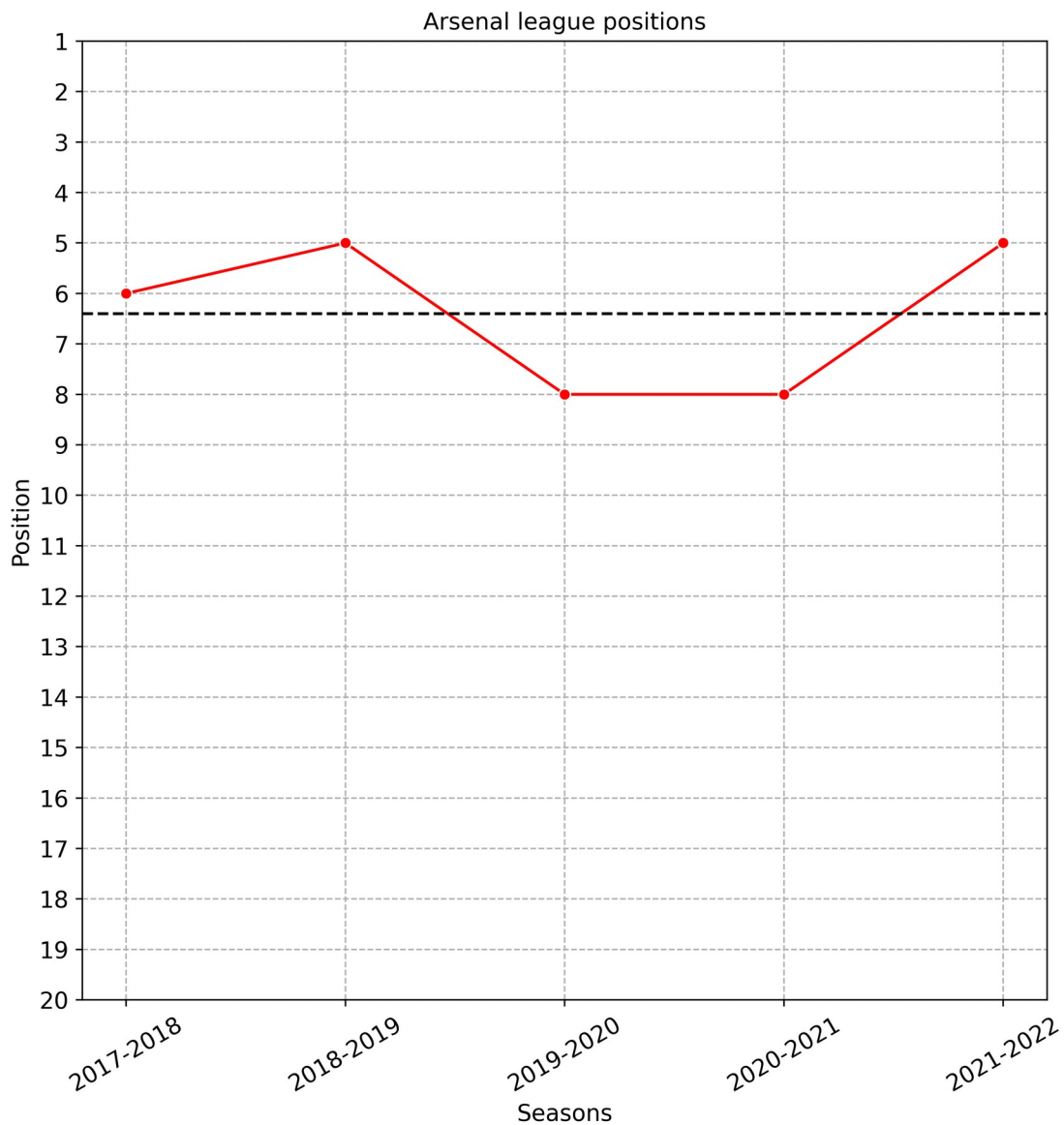
Starting with xG/90 to be the best offensive squad in the league, the team needs to create chances to score at least 2.15 goals per game on average. Although does it mean that scoring goals guarantees a title?



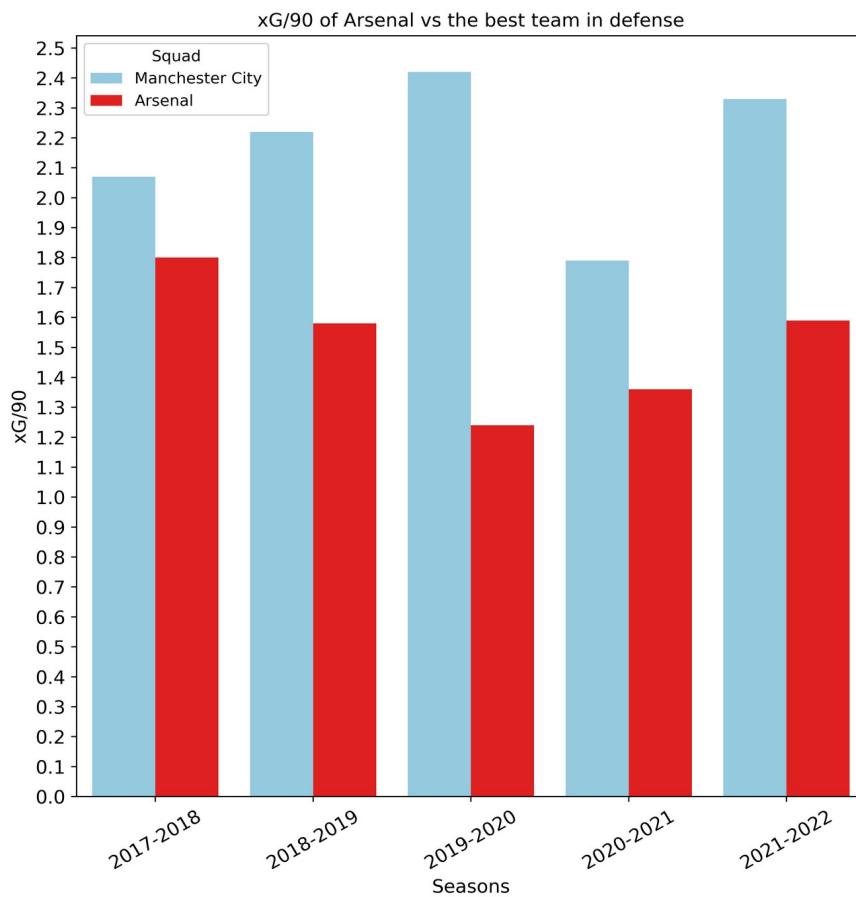
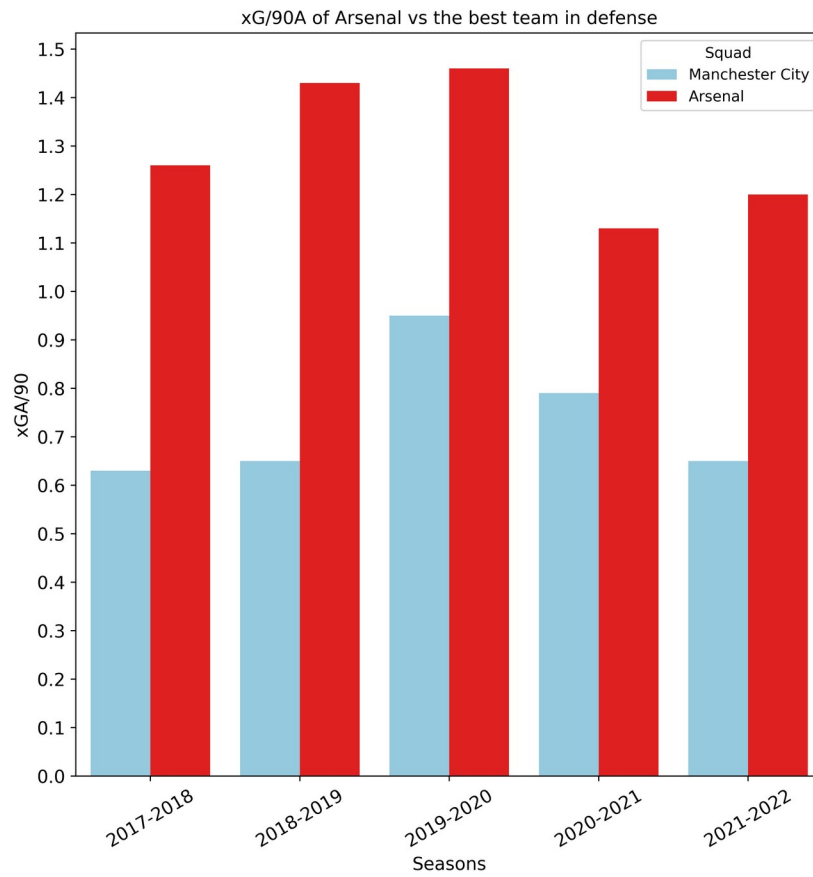
As you can see on the scatter plot in 4 out of 5 seasons team which won the championship did not score the most goals. Thus we can conclude (similarly to defensive statistics), that defence does not guarantee a championship, but it significantly increases the probability of winning it. What is even more interesting the ratio is the same, as for Goals Allowed. In 80% of seasons when a team scored more goals or conceded less, it won the title.

2. How did Arsenal progress during the last 5 seasons?

The main topic of this paper is assessing Arsenal's chance to win the title this season. Thus we need to focus on their progress throughout the last 5 years. Firstly, let's have a look at the positions of The Gunners from 2018 to 2022:



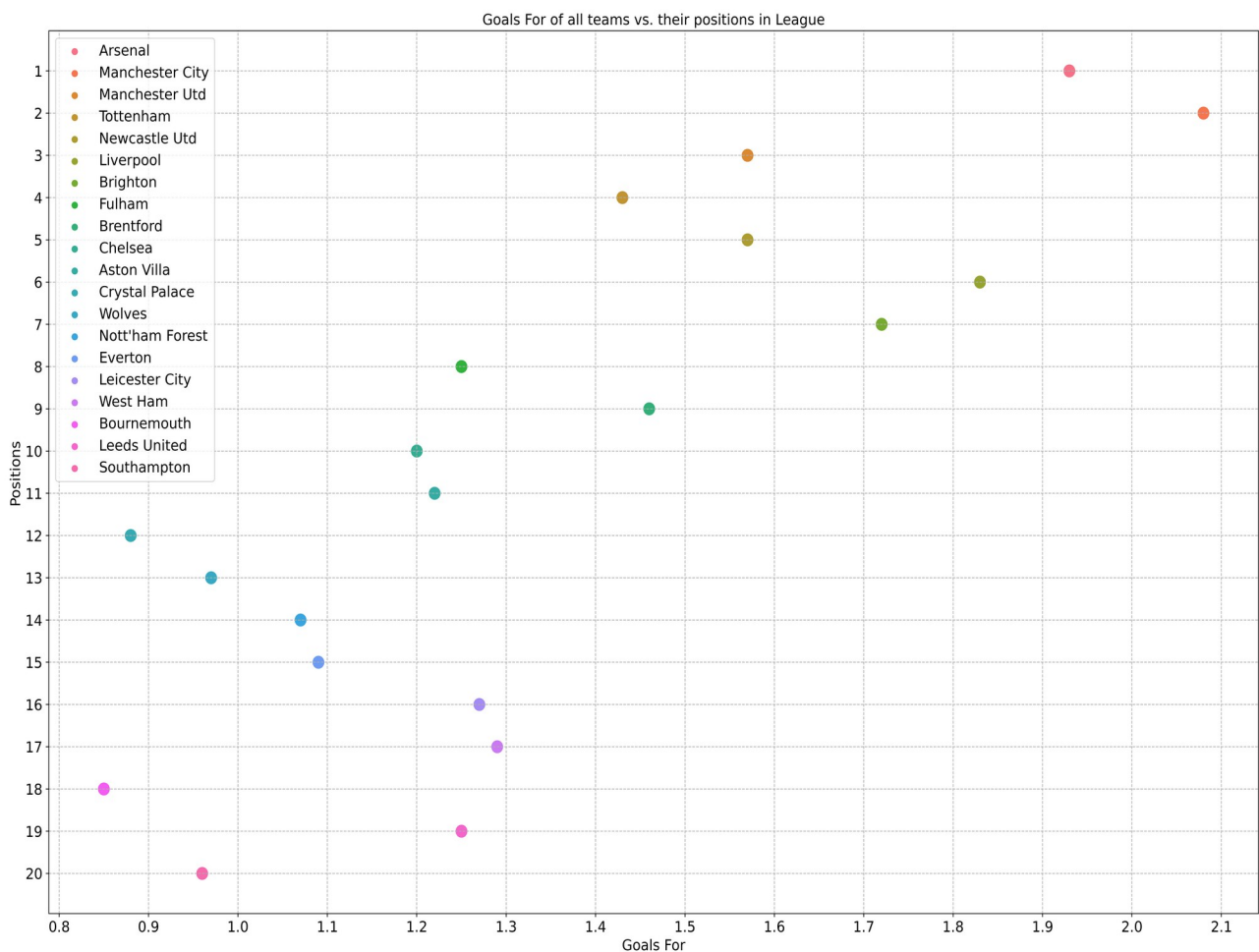
As you can see positions fluctuate between 8 and 5. Arsenal has never reached the top 4 that guarantees participation in the Champions League. Moreover, when we see the comparison of $xG/90$ and $xGA/90$ between Arsenal and the best team in these categories (always Manchester City), there is a huge difference.



In xGA statistic breakthrough took place in the season 2020/21 when Mikel Arteta took over the managerial position. It is said that great defence is a cornerstone of success during this season. Although, based on observations from the previous season, no trend can be observed that explains clearly why Arsenal is at the top of the table this season.

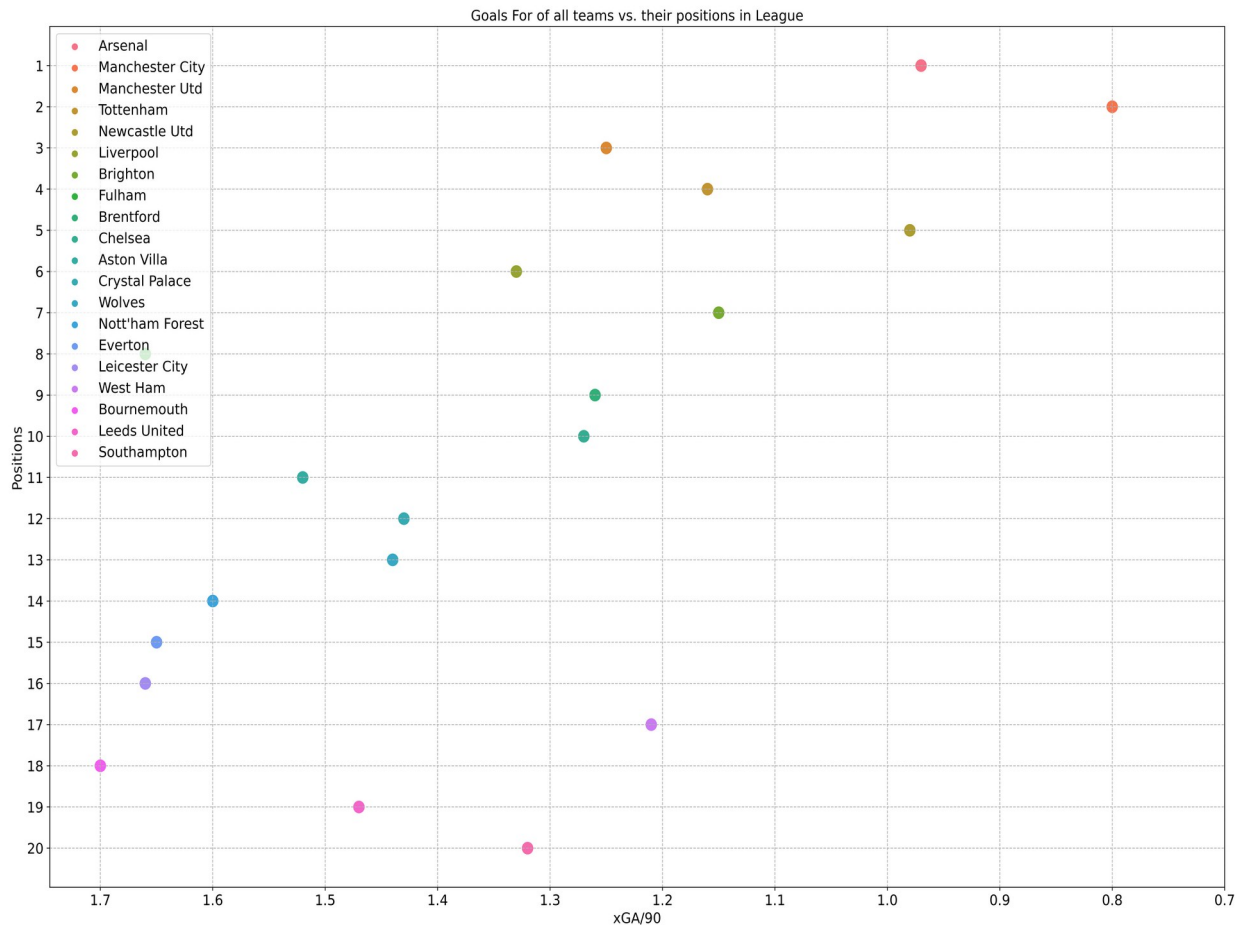
3. Will Arsenal be the next Premier League champion?

After analyzing previous seasons, now there is time to take a look at the current season. At first, we need to take a closer look at xG/90 of Arsenal vs the rest of the teams during the 2022/2023 league campaign.

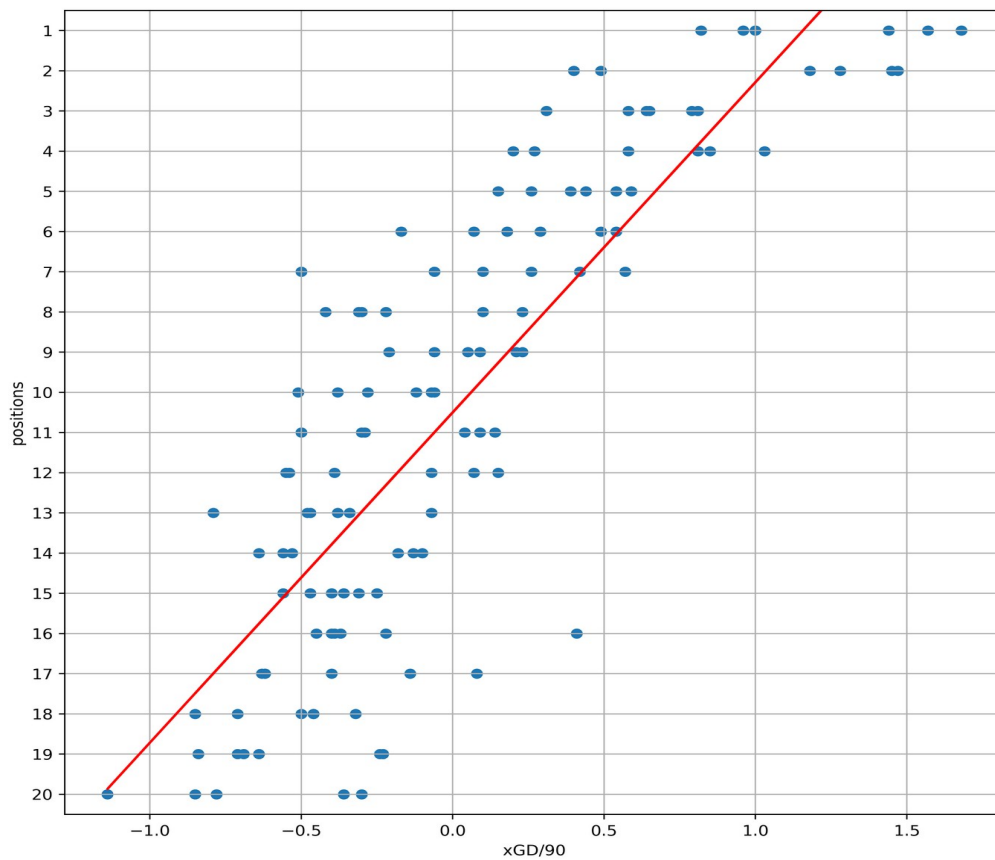


Based on the scatter plot it is clear that Arsenal does not have the best xG/90 in the league this season (Manchester City is still at the top). Nevertheless, as we proved in the previous paragraph it

is not the only thing that decides the title. We should take into consideration also xGA/90, so the most important defensive parameter.



Unfortunately, Arsenal is not leading in defensive statistics as well (Manchester City is also the best team here). As a last part of the analysis, we need to take a look at a model that can help us to predict if Arsenal has any chance to win a title besides that it is not the best team in offensive or defence. Below linear regression is presented. xGD/90 (expected goals difference per 90 minutes) is the explanatory variable and the positions of teams are the dependent variable. Model score = 70%.



Based on the calculation for the 2022/2023 season xGD/90 for Arsenal (0.96), the chance that this club will win Premier League is really low. Nevertheless, Liverpool in season 2019/2020 proved that not always the best team statistically win the title.

Parameters Glossary:

Rk- league position

Squad- team name

MP- matches played

W- wins

D- draws

L- loses

GF- goals forwarded (scored)

GA- goal allowed

GD- goal difference (GF-GA)

Pts- points

Pts/MP- points per match

xG- expected goals

xGA- expected goals allowed

xGD- expected goals difference

xGD/90- expected goals difference per 90 minutes

Last 5- form in last 5 matches

Attendance

Top Team Scorer

Goalkeeper

Notes- comments

xG/90- expected goals per 90 mins

xGA/90- expected goals allowed per 90 mins

season