

Project Report on analysing European Football leagues and teams



SUBMITTED BY- Siddharth Jain

ROLL NO.- 045060

SECTION- H

GITHUB-LINK: https://github.com/wakeupsiddd/DEV1_project.git

Project Objectives:

The objective of this project is to conduct a comprehensive analysis of European football leagues and teams using key metrics such as Expected Goals (xG) values, xG difference (xG_diff), and team performance statistics including the sum and mean of wins, xG, goals scored. This analysis aims to provide insights into the performance, consistency, and tactical aspects of European football teams, with a focus on the following aspects:

1. Performance Evaluation:

Assessing the performance of European football teams across various leagues by analyzing their xG values and xG_diff over multiple seasons.

2. Consistency and Long-term Trends:

Analyzing the sum and mean of wins, xG, and goals scored by each team over multiple seasons to identify patterns of consistency or improvement.

3. Positional Changes:

Investigating the changes in league positions of European football teams over different years to understand their progress or regression.

4. Comparative Analysis:

Comparing the performance of teams from different European leagues to determine which leagues exhibit greater competitiveness or have teams with superior xG and performance statistics.

General Description of Data:

This dataset provides detailed information about football teams' performance in different leagues that are La Liga, EPL, Bunderliga, Serie A, Ligue 1, RFPL for the years 2015 to 2022 . It includes a range of performance metrics and statistics for each team over the course of the season. Here is a general description of the data columns:

Leagues: Indicates the name of the football league, in this case- La Liga, EPL, Bunderliga, Serie A, Ligue 1, RFPL

Year: Represents the year of the season, which is 2015 in this dataset.

Position: Denotes the final league position of the team at the end of the season.

Team: Specifies the name of the football team being analyzed.

Matches: Indicates the total number of matches played by the team during the season.

Wins: Represents the number of matches the team won.

Draws: Indicates the number of matches that ended in a draw for the team.

Loses: Denotes the number of matches the team lost.

Scored: Represents the total number of goals scored by the team in all matches played.

Missed: Indicates the total number of goals conceded by the team in all matches played.

Pts: Denotes the total number of points earned by the team during the season based on wins and draws.

Expected Goals (xG): measures the quality of a shot based on several variables such as assist type, shot angle and distance from goal, whether it was a headed shot and whether it was defined as a big chance.

xG_diff : difference between actual goals scored and expected goals.

Expected Assits (xGA): measures the likelihood that a given pass will become a goal assist. It considers several factors including the type of pass, pass end-point and length of the pass.

Expected Points (xPTS): measures the likelihood of a certaing game to bring points to the team.

ppda_coef : passes allowed per defensive action in the opposition half (power of pressure)

oppda_coef : opponent passes allowed per defensive action in the opposition half (power of opponent's pressure)

deep: Denotes the number of deep passes attempted by the team.

deep_allowed: Indicates the number of deep passes allowed by the team's defense.

xpts (Expected Points): Represents the expected number of points the team was expected to earn based on their performance.

xpts_diff (Expected Points Difference): Indicates the difference between the actual points earned and the expected points based on performance.

	Unnamed: 0	Unnamed: 1	position	team	matches	wins	\				
0	La_liga	2015	1	Barcelona	38	29					
1	La_liga	2015	2	Real Madrid	38	28					
2	La_liga	2015	3	Atletico Madrid	38	28					
3	La_liga	2015	4	Villarreal	38	18					
4	La_liga	2015	5	Athletic Club	38	18					
..					
907	RFPL	2022	12	Krylya Sovetov Samara	30	8					
908	RFPL	2022	13	Nizhny Novgorod	30	8					
909	RFPL	2022	14	Fakel	30	6					
910	RFPL	2022	15	Khimki	30	4					
911	RFPL	2022	16	Torpedo Moscow	30	2					
	draws	loses	scored	missed	pts	xG	xG_diff	npvG	xGA	xGA_diff	\
0	4	5	112	29	91	113.60	1.60	99.48	34.03	5.03	
1	6	4	110	34	90	90.45	-19.55	83.76	45.23	11.23	
2	4	6	63	18	88	54.93	-8.07	53.44	27.80	9.80	
3	10	10	44	35	64	40.49	-3.51	38.26	41.63	6.63	
4	8	12	58	45	62	53.90	-4.10	50.04	45.33	0.33	
..	
907	8	14	32	45	32	37.87	5.87	34.85	39.21	-5.79	
908	7	15	34	48	31	28.58	-5.42	22.54	45.24	-2.76	
909	12	12	36	48	30	31.31	-4.69	26.03	42.75	-5.25	
910	6	20	25	67	18	26.96	1.96	23.04	58.71	-8.29	
911	5	23	20	62	11	21.14	1.14	18.87	63.85	1.85	
	npvGA	npvGD	ppda_coef	oppda_coef	deep	deep_allowed	xpts	xpts_diff			
0	33.29	66.19	6.01	15.06	570	163	94.38	3.38			
1	42.26	41.50	9.25	14.57	404	211	79.09	-10.91			
2	27.05	26.39	8.83	9.05	261	170	72.28	-15.72			

Analysis:

For analysing Leagues

Football is a multifaceted sport, which is influenced by a multitude of factors that are xG (Expected Goals) and xG_diff (Expected Goals Difference). They offer valuable insights into the quality of goal-scoring opportunities and offensive efficiency but, they should not serve as the sole arbiters of a league's overall excellence.

For analysing Teams

One of the key objectives is to explore and identify the best-performing teams across different football leagues. To achieve this, we employ a multifaceted analysis that encompasses several essential metrics, including the number of wins, Expected Goals (xG), and total goals scored. We also calculate the mean values of these metrics to gain a deeper understanding of team consistency. Additionally, we utilize data visualization techniques to provide a clear and comprehensive overview of team performance within each league. Here's how we approach this analysis:

1. Grouping by Number of Wins, xG, and Score:

We start by grouping teams within each league based on their performance in terms of the number of wins, xG values, and total goals scored. This grouping helps us categorize teams into different performance tiers.

By examining these groupings, we can identify which teams stand out as the top performers in each league based on these fundamental metrics.

2. Calculating Mean Values:

In addition to identifying individual high-performing teams, we calculate the mean values of wins, xG, and scores for all teams within each league.

Mean values provide insights into the average team performance in terms of wins, goal-scoring ability (xG), and offensive prowess (score).

Comparing individual team performance to the league mean helps us assess how teams deviate from the league average.

3. Visualizing Team Wins:

Data visualization plays a crucial role in our analysis. We create visual representations, such as horizontal bar charts, to display the exact number of wins for each team within each league.

Visualizations allow us to quickly grasp which teams have the highest number of wins and how they compare to their league counterparts.

4. Monitoring Position Changes for Top 6 Teams:

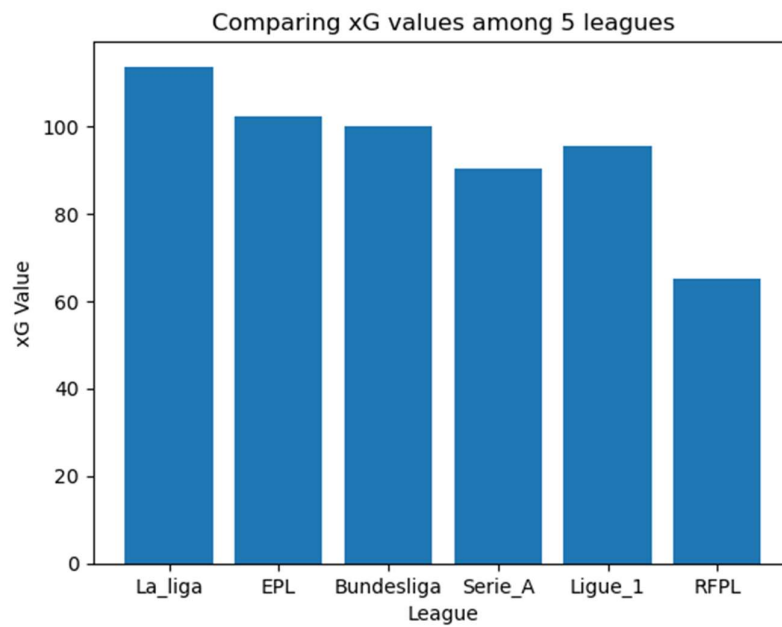
To understand the dynamics of league standings, we focus on tracking the position changes of the top 6 teams within each league across different years.

By analyzing these position changes, we can identify trends, such as which teams consistently maintain their top positions and which teams experience fluctuations in league standings over time.

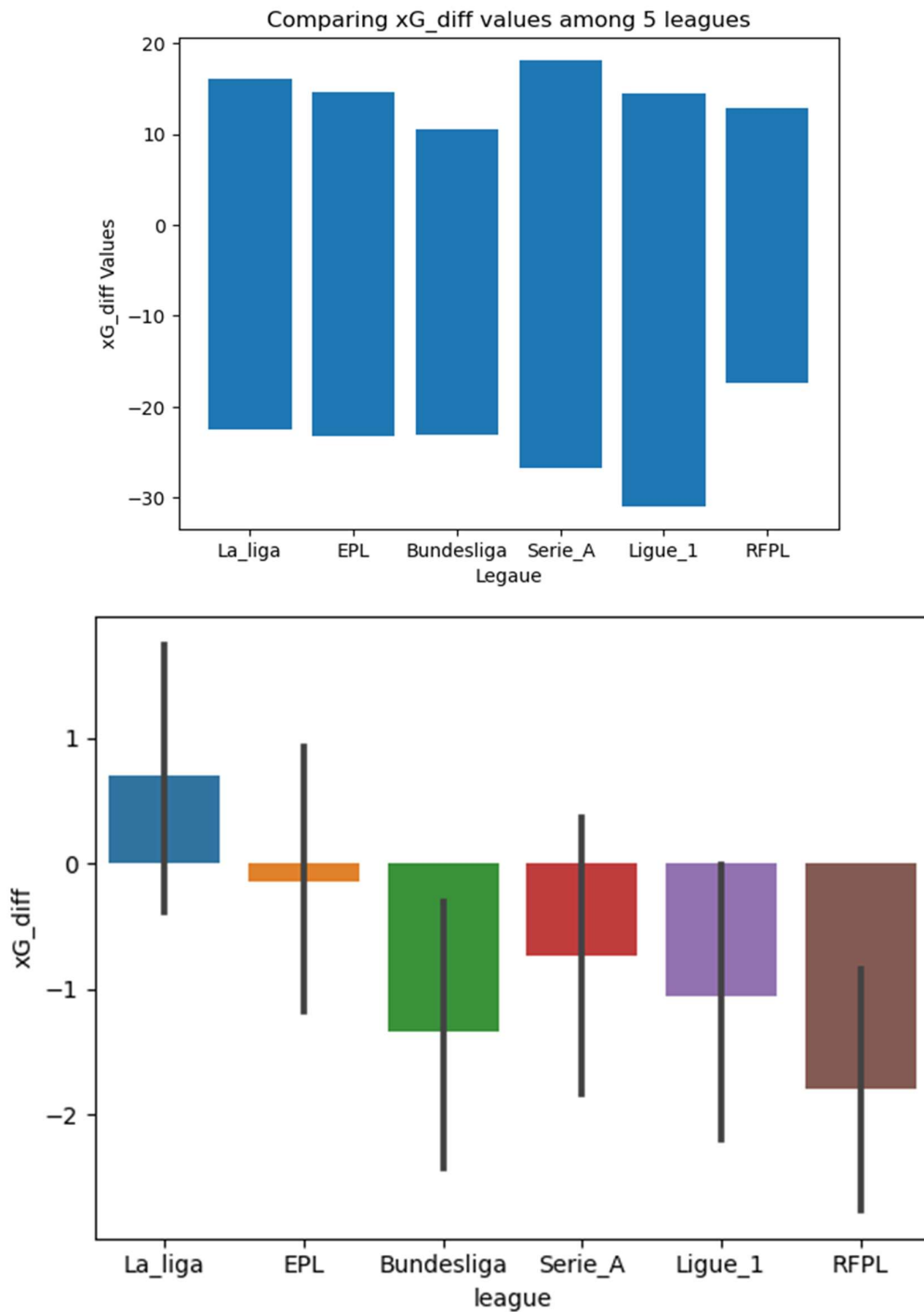
Through this comprehensive analysis, we aim to provide a holistic view of team performance within different football leagues. By considering a combination of wins, xG, and scores, we can identify the standout teams and assess their consistency. The visualizations further enhance our ability to interpret the data effectively, and monitoring position changes sheds light on the competitiveness and evolution of each league. Ultimately, this analysis will help stakeholders, football enthusiasts, and analysts gain valuable insights into the best-performing teams across various football leagues.

Findings & Inferences

With respect to leagues



xG value of La liga is the highest followed by the EPT. xG value of RFPL is the lowest.



These graphs show that EPL and La_liga are the best leagues only considering with xG value, xG_diff and score. Though, it is really hard to decide which league is the best only with xG or xG_diff values. It's because of that it's just rare to get many goals in football. It means that it can happen that we see very low goal score or xG value even in best league and one more reason is that it depends on the feature of league. While our analysis of xG and xG_diff values is indeed insightful, it is imperative that we adopt a more holistic approach encompassing a diverse spectrum of metrics and factors when endeavoring to gauge the caliber of football

leagues. Moreover, we should weigh auxiliary factors such as goal-scoring frequency, league competitiveness, and the international achievements of league teams to form a well-rounded perspective of which league may merit recognition as the premier one.

With respect to Teams

1. EPL

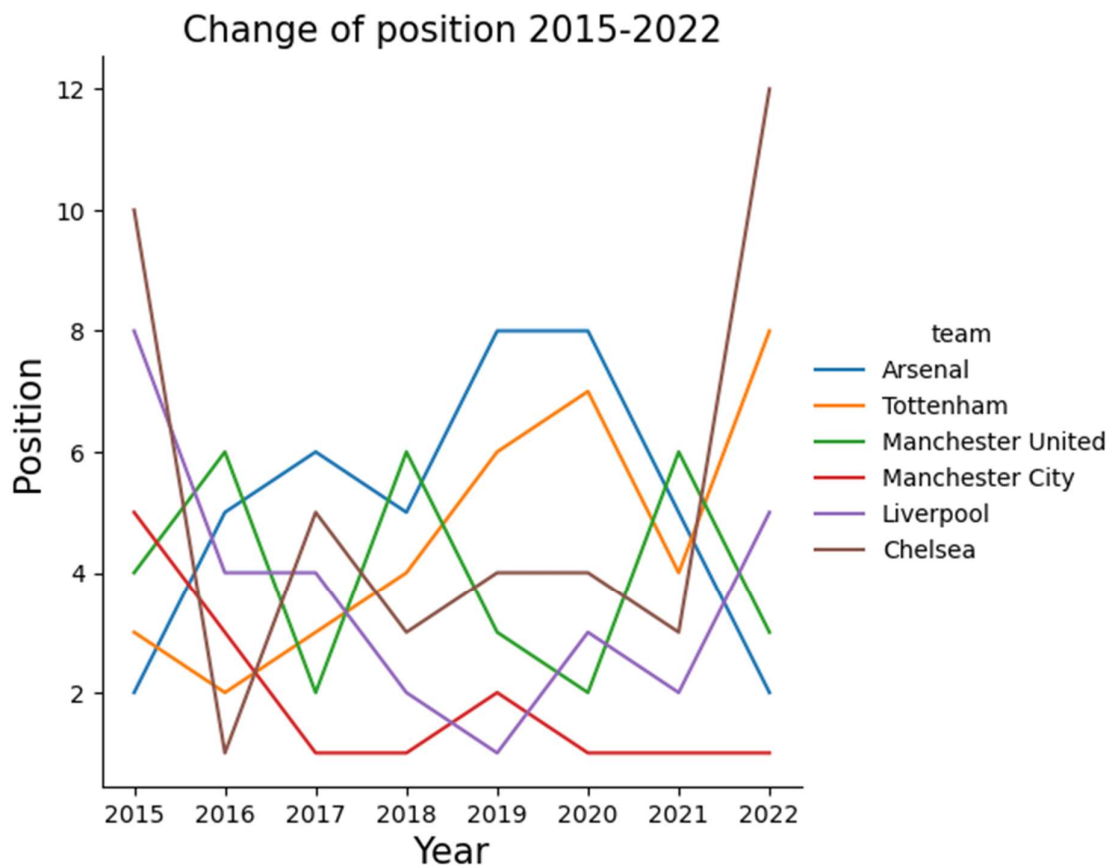
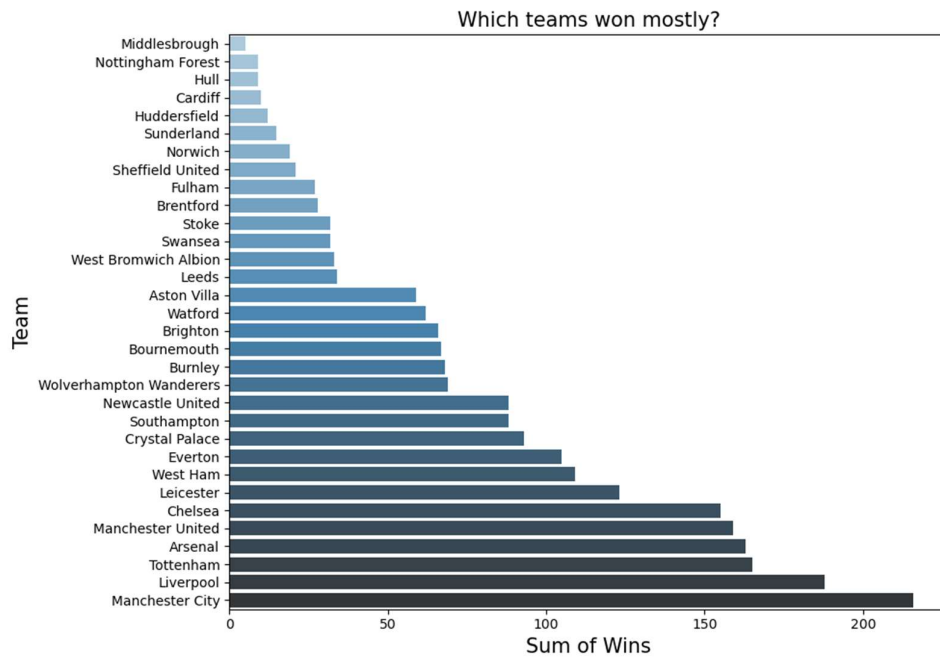
```
In [19]: epl.groupby('team')[['wins', 'xG', 'scored']].sum().sort_values(by='wins', ascending=False).head()
Out[19]:
```

	wins	xG	scored
Manchester City	216	688.85	730
Liverpool	188	599.14	636
Tottenham	165	495.63	564
Arsenal	163	517.15	549
Manchester United	159	489.13	490

```
In [20]: epl.groupby('team')[['wins', 'xG', 'scored']].mean().sort_values(by='wins', ascending=False).head()
Out[20]:
```

	wins	xG	scored
Manchester City	27.00	86.11	91.25
Liverpool	23.50	74.89	79.50
Tottenham	20.62	61.95	70.50
Arsenal	20.38	64.64	68.62
Manchester United	19.88	61.14	61.25

In EPL, Manchester City has secured the most victories with an impressive 216 wins. They also lead in several statistical categories: their xG (expected goals) is the highest at 688.85, their total score of 730 points is the highest among all teams, and they top the charts for average wins with 27, mean xG with 86.11, and mean goals scored with 91.25. Liverpool holds the second position, followed by Tottenham in third, Arsenal in fourth, and Manchester United in fifth place. Middlesbrough, on the other hand, finished last in 32nd place.



This shows that Manchester City is getting better from 2016 to 2022 but Tottenham and Chelsea are getting worse. And it seems that Arsenal and Manchester City were very competitive for the league title in 2022.

When you are talking about football, it is all about, which team has won the game. It doesn't matter how many goals the team scored, when they lose the game, then they are not winner.

And even considering other conditions like xG values and goal score, we can say that, the top 4 teams in EPL are :

Manchester City

Liverpool

Chelsea

Tottenham

2. La Liga

```
In [24]: la_liga.groupby('team')[['wins', 'xG', 'scored']].sum().sort_values(by='wins', ascending=False).head()
```

```
Out[24]:
```

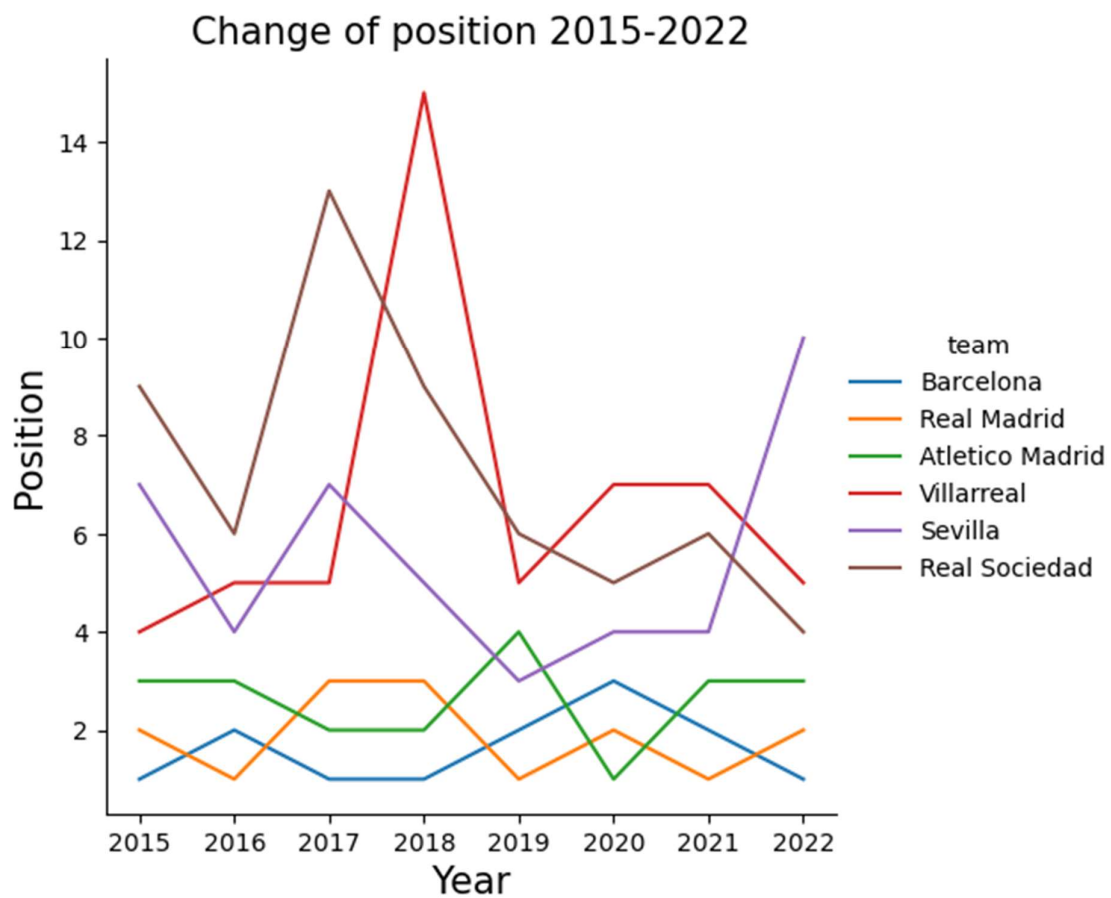
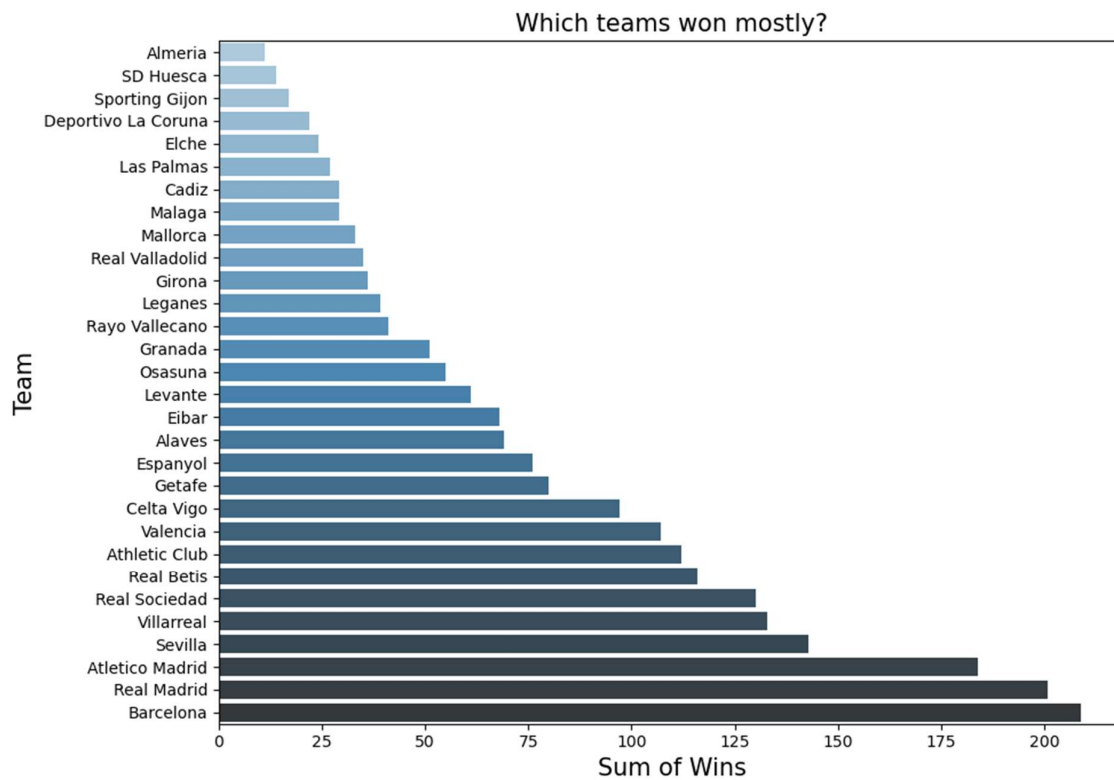
	wins	xG	scored
team			
Barcelona	209	690.22	726
Real Madrid	201	638.10	665
Atletico Madrid	184	462.24	499
Sevilla	143	461.29	438
Villarreal	133	462.54	451

```
In [25]: la_liga.groupby('team')[['wins', 'xG', 'scored']].mean().sort_values(by='wins', ascending=False).head()
```

```
Out[25]:
```

	wins	xG	scored
team			
Barcelona	26.12	86.28	90.75
Real Madrid	25.12	79.76	83.12
Atletico Madrid	23.00	57.78	62.38
Sevilla	17.88	57.66	54.75
Villarreal	16.62	57.82	56.38

In La Liga, Barcelona has secured the most victories with an impressive 209 wins. They also lead in several statistical categories: their xG (expected goals) is the highest at 690.22, their total score of 726 points is the highest among all teams, and they top the charts for average wins with 26.17, mean xG with 86.28, and mean goals scored with 90.75. Real Madrid holds the second position, followed by Atletico Madrid in third, Sevilla in fourth, and Villarreal in fifth place. Almeria, on the other hand, finished last.



This shows that Real Sociedad is getting better from 2016 to 2022 but Sevilla is getting worse. And it seems that Real Madrid, Barcelona and Atletico Madrid were very competitive for the league title from 2015 to 2022.

When you are talking about football, it is all about, which team have won the game. It doesn't matter how many goals the team scored, when they lose the game, then they are not winner.

And even considering other conditions like xG values and goal score, we can say that, the top 4 teams in La Liga are :

Barcelona

Real Madrid

Atletico Madrid

Servilla

3. Bundesliga

```
In [33]: bundesliga.groupby('team')[['wins', 'xG', 'scored']].sum().sort_values(by='wins', ascending=False).head()
```

```
Out[33]:
```

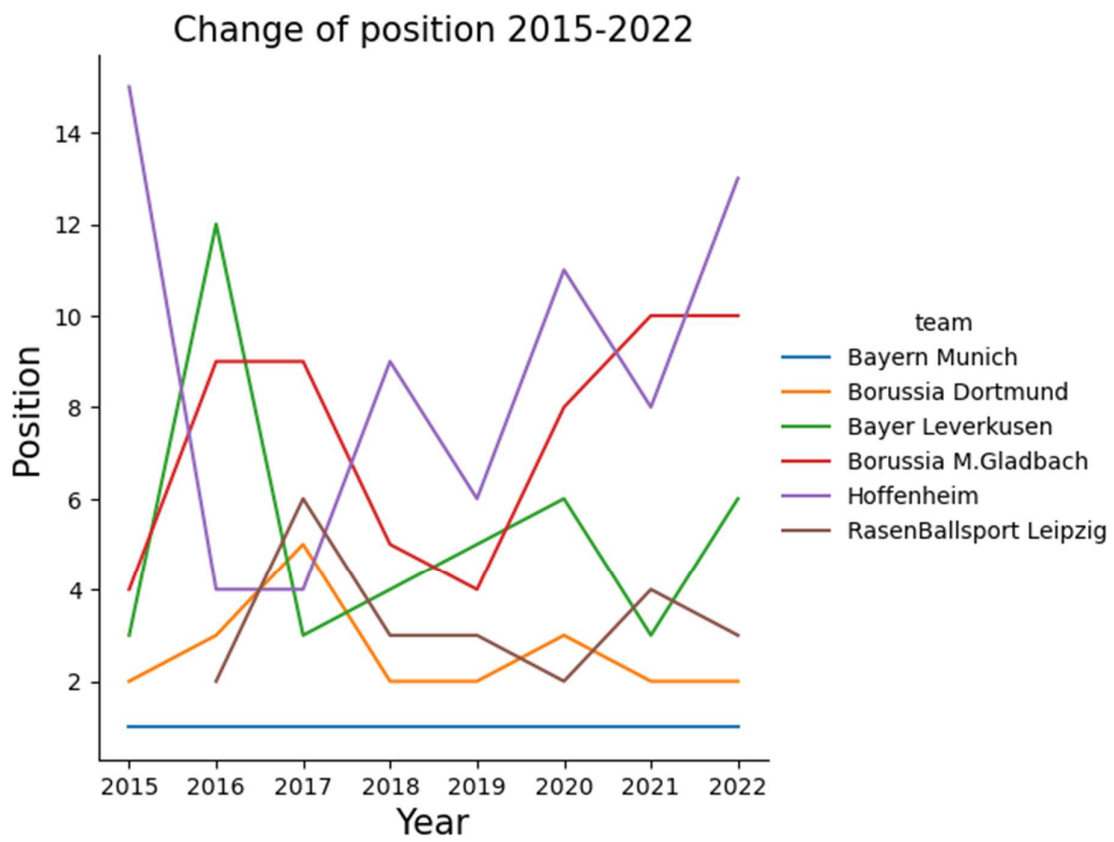
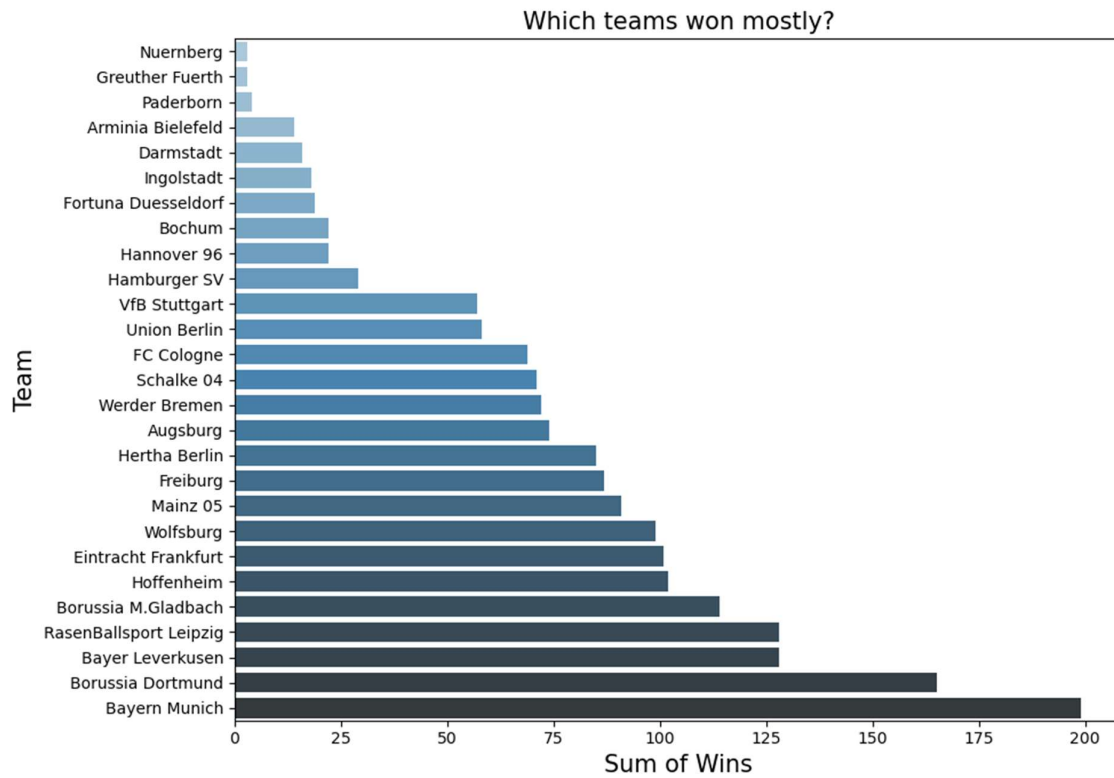
	wins	xG	scored
team			
Bayern Munich	199	673.55	737
Borussia Dortmund	165	565.24	626
Bayer Leverkusen	128	453.64	487
RasenBallsport Leipzig	128	454.27	463
Borussia M.Gladbach	114	446.11	450

```
In [34]: bundesliga.groupby('team')[['wins', 'xG', 'scored']].mean().sort_values(by='wins', ascending=False).head()
```

```
Out[34]:
```

	wins	xG	scored
team			
Bayern Munich	24.88	84.19	92.12
Borussia Dortmund	20.62	70.66	78.25
RasenBallsport Leipzig	18.29	64.90	66.14
Bayer Leverkusen	16.00	56.71	60.88
Union Berlin	14.50	44.46	48.00

In Bundeliga, Buyern Munich has secured the most victories with 199 wins. They also lead in several statistical categories: their xG (expected goals) is the highest at 673.55, their total score of 737 points is the highest among all teams, and they top the charts for average wins with 24.88, mean xG with 84.19, and mean goals scored with 92.12. Borussia Dortmund holds the second position, followed by RasenBallsport Lepzig in third, Bayer Leverkusen in fourth, and Union Berlin in fifth place. Nuernberg, on the other hand, finished last.



This shows that Borussia Dortmund is getting better from 2015 to 2022 but Hoffenheim is getting worse. And it seems that Borussia Dortmund and RasenBallsport Leipzig were very competitive for positions but Bayern Munich won all league title from 2015 to 2022.

When you are talking about football, it is all about, which team have won the game. It doesn't matter how many goals the team scored, when they lose the game, then they are not winner.

And even considering other conditions like xG values and goal score, we can say that, the top 4 teams in Bundeliga are :

Bayern Munich

Borussia Dortmund

RasenBallsport Leipzig

Bayer Leverkusen

4. Serie A

```
In [39]: Serie_a.groupby('team')[['wins', 'xG', 'scored']].sum().sort_values(by='wins', ascending=False).head()
```

```
Out[39]:
```

	wins	xG	scored
team			
Juventus	207	533.36	574
Napoli	197	559.96	623
Inter	179	565.02	570
Roma	167	562.68	554
AC Milan	159	494.17	487

```
In [40]: Serie_a.groupby('team')[['wins', 'xG', 'scored']].mean().sort_values(by='wins', ascending=False).head()
```

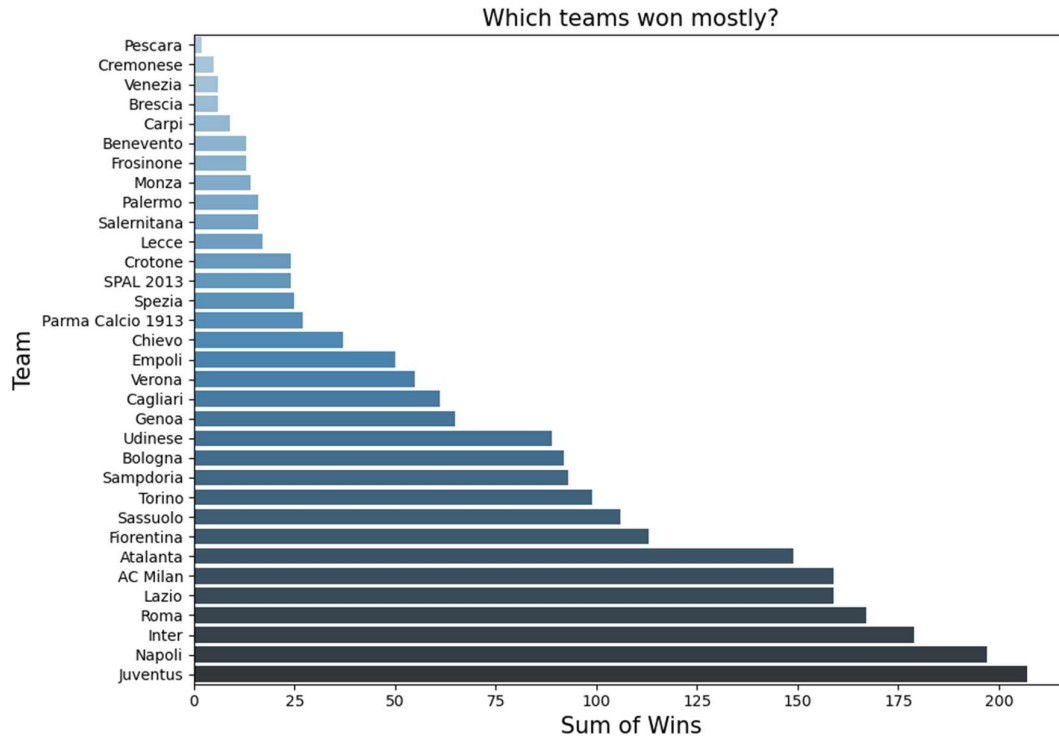
```
Out[40]:
```

	wins	xG	scored
team			
Juventus	25.88	66.67	71.75
Napoli	24.62	69.99	77.88
Inter	22.38	70.63	71.25
Roma	20.88	70.34	69.25
AC Milan	19.88	61.77	60.88

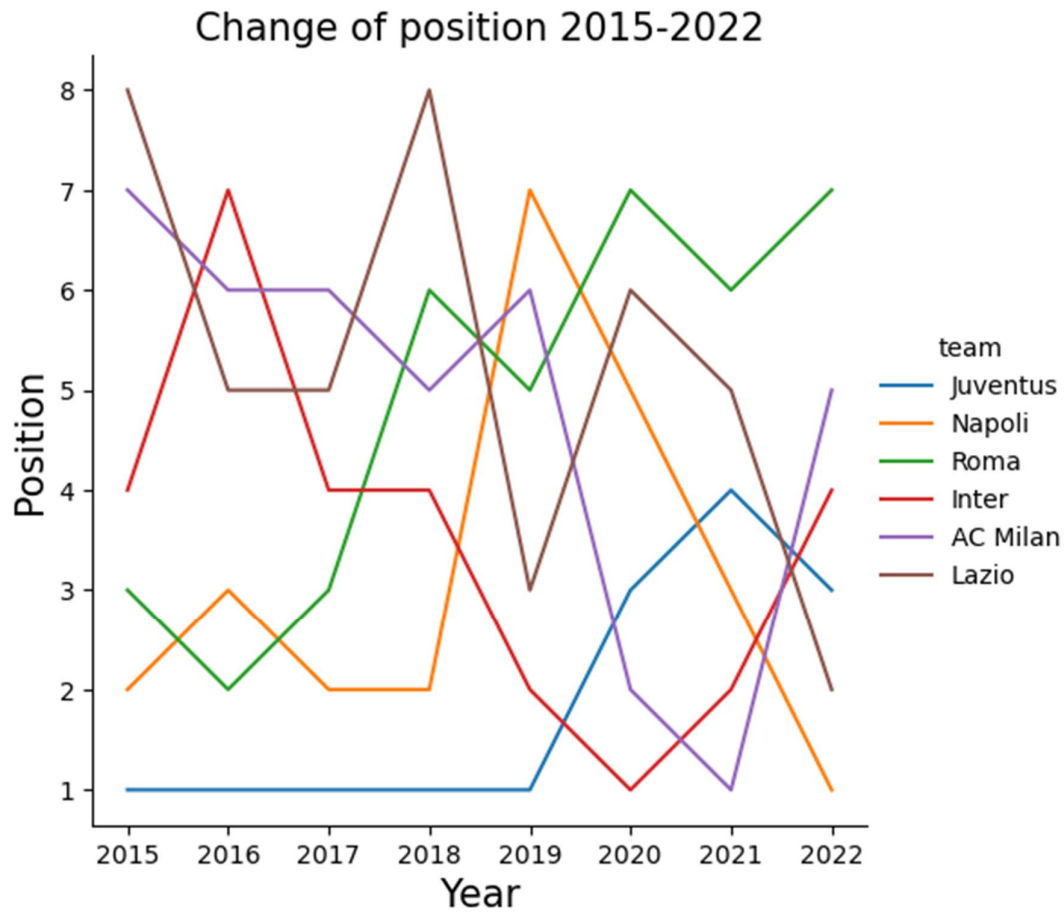
In Serie A, Juventus has secured the most victories with an impressive 207 wins.

But Inter in xG (expected goals) is the highest at 559.96, and Napoli has the highest of total score of 623 points among all teams. Juventus top the charts for average wins with 25.88, but Inter top in mean xG with 69.66, and Napoli top in mean goals scored with 77.88.

This means that the performance of Napoli and Inter is not constant, in some matches they perform extraordinarily well but does not perform well in some matches. Where as performance of Juventus is consistant and less deviated. They perform equally well in most of the matches.



In Wins Napoli holds the second position after Juventus, followed by Inter in third, Roma in fourth, and AC Milan in fifth place. Pescara, on the other hand, finished last.



This shows that AC Milan is getting better from 2016 to 2021 but Roma is getting worse. When you are talking about football, it is all about, which team have won the game. It doesn't matter how many goals the team scored, when they lose the game, then they are not winner.

And even considering other conditions like xG values and goal score, we can say that, the top 4 teams in Serie A are :

Juventus
Napoli
Inter
Roma

5. Ligue 1

```
In [44]: ligue_1.groupby('team')[['wins', 'xG', 'scored']].sum().sort_values(by='wins', ascending=False).head()
```

```
Out[44]:
```

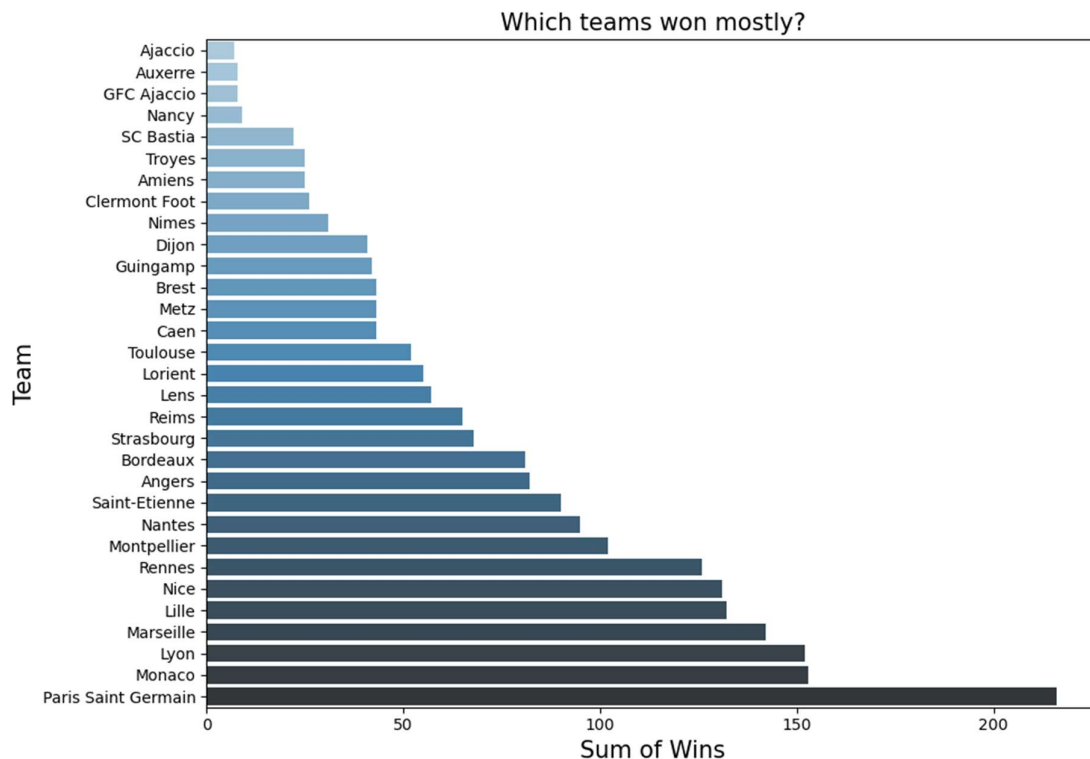
	wins	xG	scored
Paris Saint Germain	216	682.42	738
Monaco	153	490.92	542
Lyon	152	533.84	558
Marseille	142	455.52	470
Lille	132	401.41	400

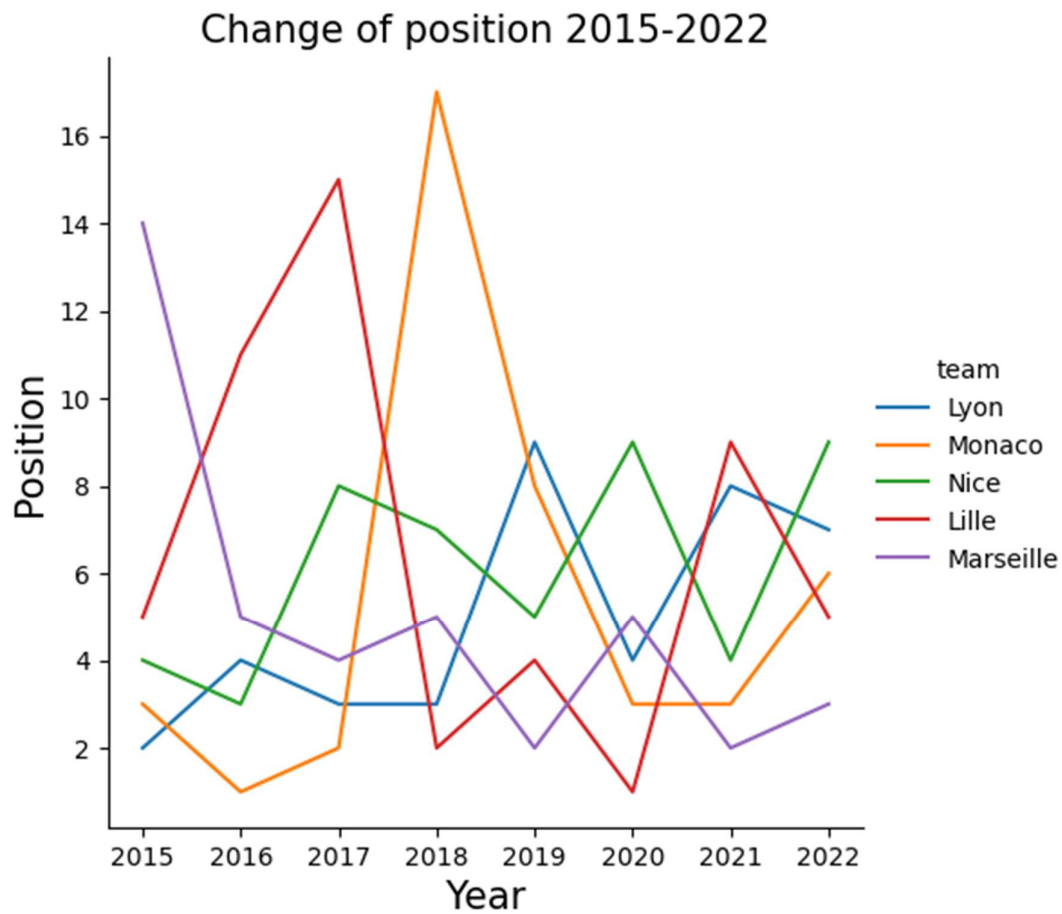
```
In [45]: ligue_1.groupby('team')[['wins', 'xG', 'scored']].mean().sort_values(by='wins', ascending=False).head()
```

```
Out[45]:
```

	wins	xG	scored
Paris Saint Germain	27.00	85.30	92.25
Monaco	19.12	61.36	67.75
Lyon	19.00	66.73	69.75
Lens	19.00	59.51	61.67
Marseille	17.75	56.94	58.75

In Ligue 1, Paris Saint Germain has secured the most victories with an impressive 216 wins. They also lead in several statistical categories: their xG (expected goals) is the highest at 682.42, their total score of 738 points is the highest among all teams, and they top the charts for average wins with 27, mean xG with 85.30, and mean goals scored with 92.25. Monaco holds the second position, followed by Lyon in third, Marseille in fourth, and Lille in fifth place. Ajaccio, on the other hand, finished last.





This shows that Marseille is getting better from 2015 to 2022 but Lyon is getting worse. When you are talking about football, it is all about, which team have won the game. It doesn't matter how many goals the team scored, when they lose the game, then they are not winner.

And even considering other conditions like xG values and goal score, we can say that, the top 4 teams in Ligue 1 are :

Paris Saint Germain
Monaco
Lyon
Marseille

6. RFPL

```
In [49]: rfpl.groupby('team')[['wins', 'xG', 'scored']].sum().sort_values(by='wins', ascending=False).head()
```

Out[49]:

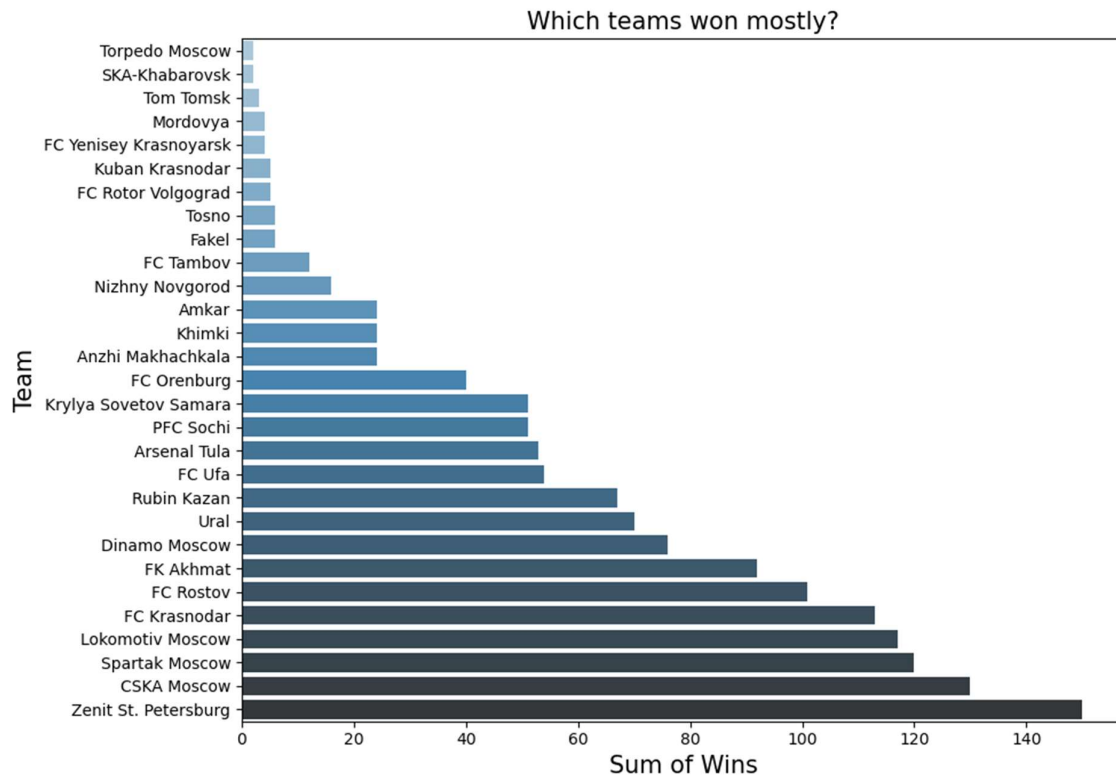
	wins	xG	scored
team			
Zenit St. Petersburg	150	447.09	495
CSKA Moscow	130	362.98	385
Spartak Moscow	120	365.02	369
Lokomotiv Moscow	117	320.62	351
FC Krasnodar	113	343.75	400

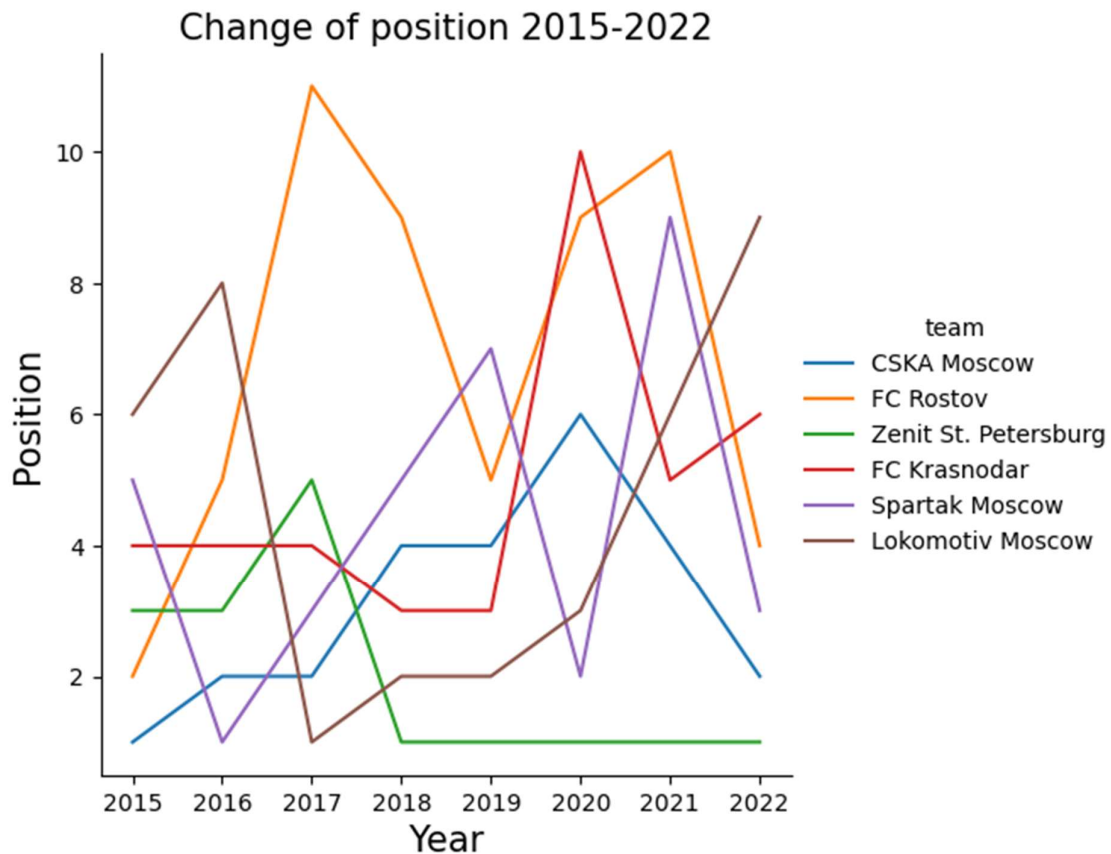
```
In [50]: rfpl.groupby('team')[['wins', 'xG', 'scored']].mean().sort_values(by='wins', ascending=False).head()
```

Out[50]:

	wins	xG	scored
team			
Zenit St. Petersburg	18.75	55.89	61.88
CSKA Moscow	16.25	45.37	48.12
Spartak Moscow	15.00	45.63	46.12
Lokomotiv Moscow	14.62	40.08	43.88
FC Krasnodar	14.12	42.97	50.00

In RFPL, Zenit St. Petersburg has secured the most victories with 150 wins. They also lead in several statistical categories: their xG (expected goals) is the highest at 447.09, their total score of 495 points is the highest among all teams, and they top the charts for average wins with 18.75, mean xG with 55.89, and mean goals scored with 61.88. CSKA Moscow holds the second position, followed by Spartak Moscow in third, Lokomotiv Moscow in fourth, and KC Krasnodar in fifth place. Torpedo Moscow, on the other hand, finished last.





This shows that Zenit St. Petersburg is getting better from 2016 to 2022

When you are talking about football, it is all about, which team have won the game. It doesn't matter how many goals the team scored, when they lose the game, then they are not winner.

And even considering other conditions like xG values and goal score, we can say that, the top 4 teams in RFPL are :

Zenit St. Petersburg
CSKA Moscow
Spartak Moscow
Lokomotiv Moscow