

Christmas Curse

A SQL ANALYSIS OF PREMIER LEAGUE HISTORY

Will Arsenal Finally Win?



MD Nowroz Imtiaz Rifat

Data analysis
Università degli Studi di Messina



Overview

In english football, they say if you are top at Christmas break, the title is yours. But is that actually true?

I used **30 years** of epl data (from 1993–2022) to test if the 'Christmas Number One' is a guarantee or a myth.

How?

Step 1: Ingested raw match data (CSV) into a database.

Step 2: Reconstructed historical league tables using SQL Window Functions from **11,113** matches data.

Step 3: Compared the "Christmas Leader" vs. the "Final Winner" for every season since 1993

The Toolkit

Engine: PostgreSQL

Interface: JetBrains DataGrip

Data Source: English Premier League Results (1993–2022) via Kaggle.

NEXT ►

Methodology

The Challenge

The raw dataset lists matches sequentially. To answer the question, I needed to 'freeze time' on December 25th for every single year since 1993, while simultaneously calculating the final season results. For that i also need to make a points table.

	season	match_date	hometeam	awayteam	fthg	ftag	ftr	hthg	htag	
1	1993-94	1993-08-14 00:00:00.000000	Arsenal	Coventry	0	3	A	NA	NA	N/
2	1993-94	1993-08-14 00:00:00.000000	Aston Villa	QPR	4	1	H	NA	NA	N/
3	1993-94	1993-08-14 00:00:00.000000	Chelsea	Blackburn	1	2	A	NA	NA	N/
4	1993-94	1993-08-14 00:00:00.000000	Liverpool	Sheffield Weds	2	0	H	NA	NA	N/
5	1993-94	1993-08-14 00:00:00.000000	Man City	Leeds	1	1	D	NA	NA	N/

The Solution: CTEs (Common Table Expressions)

I used CTEs to break the problem into logical modular steps. This avoided complex sub-queries and made the logic readable and reusable.

CTE 1

to calculate points in every match

Skill highlights

CASE...WHEN

UNION ALL

```
WITH team_points AS (  
  SELECT season, match_date, hometeam AS team,  
         CASE WHEN ftr = 'H' THEN 3 WHEN ftr = 'D' THEN 1 ELSE 0 END AS point  
  FROM matches  
  UNION ALL  
  SELECT season, match_date, awayteam AS team,  
         CASE WHEN ftr = 'A' THEN 3 WHEN ftr = 'D' THEN 1 ELSE 0 END AS point  
  FROM matches  
)
```

CTE 2&3

Using match point data ranking teams based on their total points. two tables for before xmas and in the end of the season

Skill highlights

Window Function CAST

RANK() MAKE_DATE()

```
xmas_standing AS (  
  SELECT season, team, sum(point) AS xmas_point,  
         rank() OVER (PARTITION BY season ORDER BY sum(point) DESC) as ranking  
  FROM team_points  
  WHERE match_date < MAKE_DATE(CAST(LEFT(season, 4) AS INT), 12, 25)  
  GROUP BY season, team  
)  
  
final_standing AS (  
  SELECT season, team, sum(point) AS final_point,  
         rank() OVER (PARTITION BY season ORDER BY sum(point) DESC) as ranking  
  FROM team_points  
  GROUP BY season, team  
)
```


CREATE VIEW

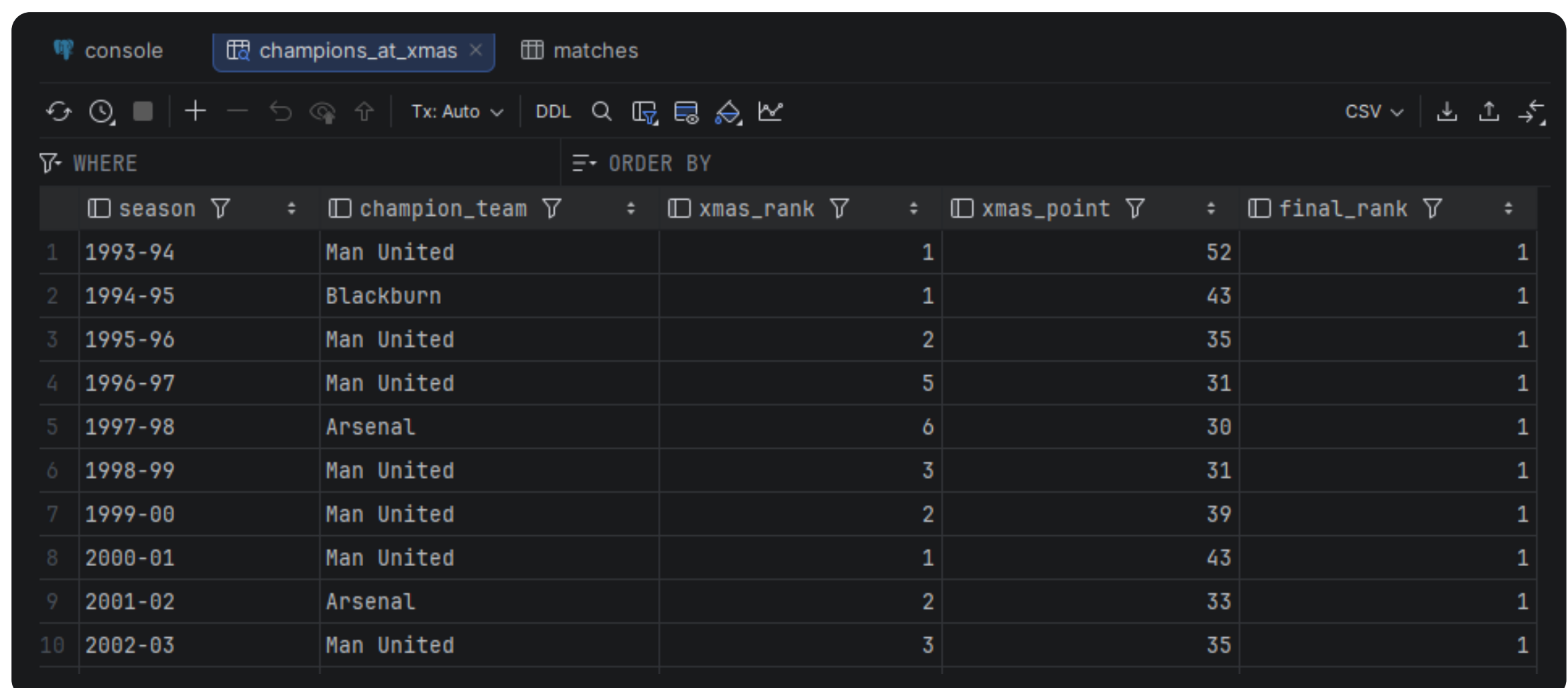
The complex CTE logic I wrote is powerful, but running it every time is inefficient. I needed a permanent, reusable way to access this 'Cleaned History'.

I encapsulated the entire logic into a Database View called `champions_at_xmas`. This acts as a virtual table, allowing me to query complex historical trends with simple one-line commands.

```
1 CREATE VIEW champions_at_xmas AS
24 SELECT f.season,
25       x.team AS champion_team,
26       x.ranking AS xmas_rank,
27       x.xmas_point,
28       f.ranking AS final_rank
29 FROM xmas_standing x
30 JOIN final_standing f ON x.season = f.season AND f.team = x.team
31 WHERE f.ranking = 1;
```

Result?

A clean table to work on



	season	champion_team	xmas_rank	xmas_point	final_rank
1	1993-94	Man United	1	52	1
2	1994-95	Blackburn	1	43	1
3	1995-96	Man United	2	35	1
4	1996-97	Man United	5	31	1
5	1997-98	Arsenal	6	30	1
6	1998-99	Man United	3	31	1
7	1999-00	Man United	2	39	1
8	2000-01	Man United	1	43	1
9	2001-02	Arsenal	2	33	1
10	2002-03	Man United	3	35	1

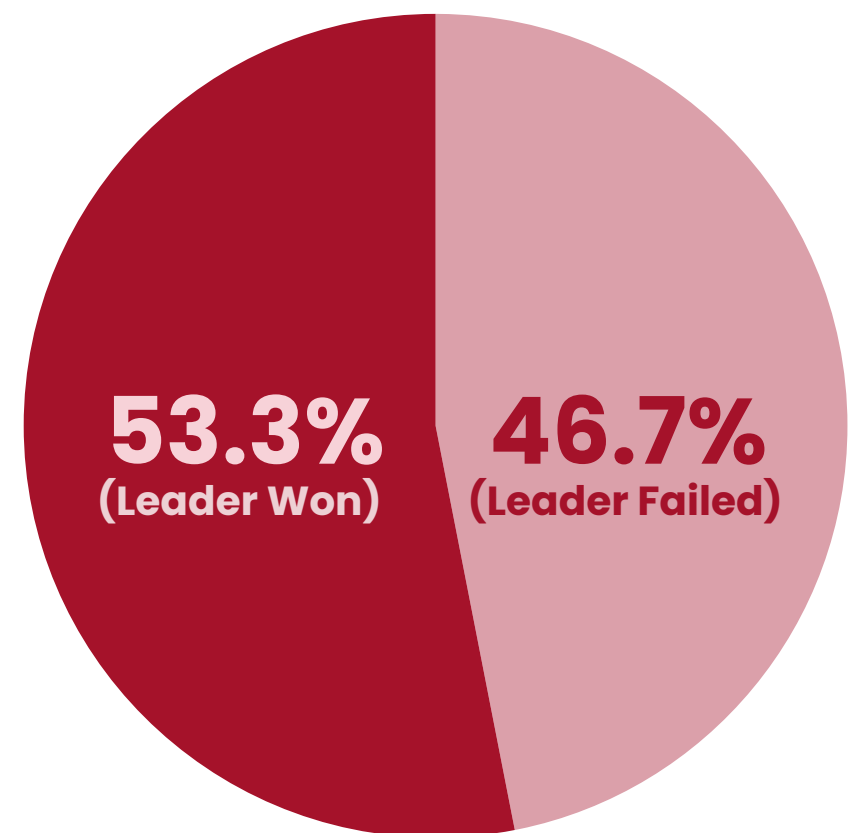
NEXT ▶

The Verdict (Myth Busted)

#1: Is Being Top at Christmas a Guarantee?

Historically, holding the top spot on Dec 25th is effectively a coin flip.

Contrary to popular belief, leading the pack at Christmas is not a guarantee. In fact, nearly half of all Christmas leaders (47%) fail to lift the trophy in May. The pressure of being the 'Hunted' often proves too much

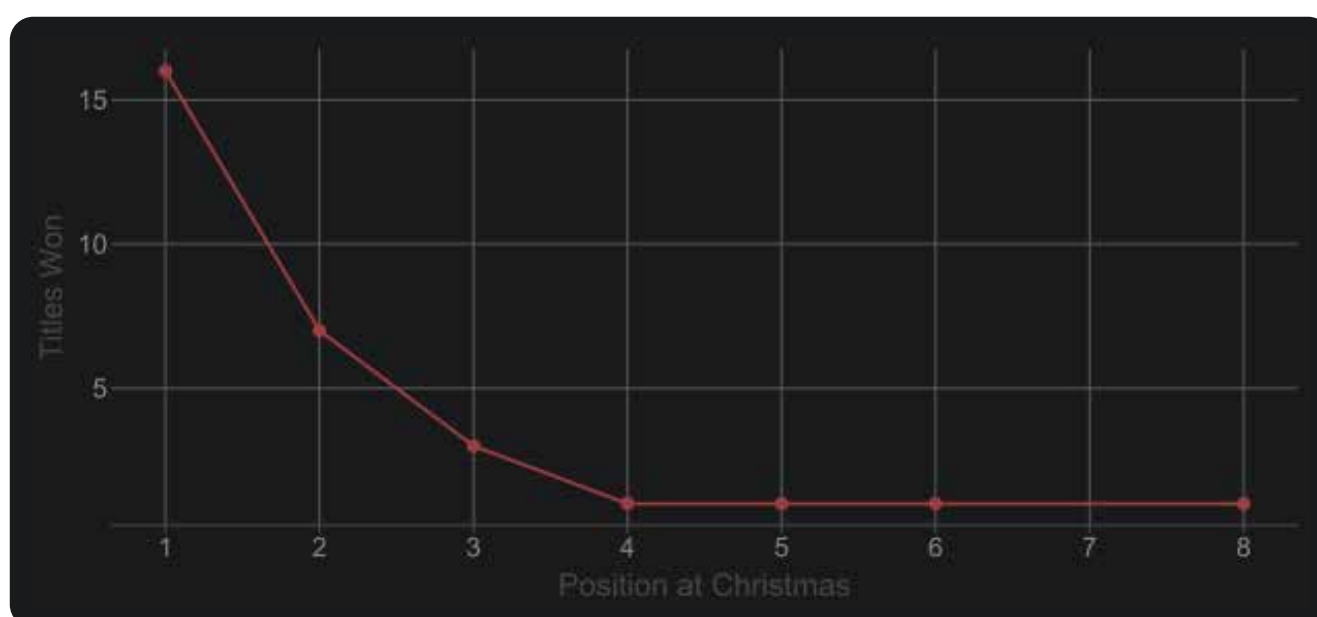


SQL query

```
SELECT ROUND(AVG(CASE WHEN xmas_rank = 1 THEN 1 ELSE 0 END) * 100, 1) AS leader_won
FROM champions_at_xmas;
-- Result: 53.3%
```

#2: The Safe Zone

If you aren't 1st, you better be 2nd. A staggering 23 out of 30 Premier League champions come from the Top 2 spots at Christmas



```
SELECT
    xmas_rank AS "Position at Christmas",
    COUNT(*) AS "Titles Won"
FROM champions_at_xmas
GROUP BY xmas_rank
ORDER BY xmas_rank ASC;
```

NEXT ▶

#3 The Outliers

Too soon to loose hope. beacuse miracles do happen like Man Utd (1996-97): Won from 5th, Arsenal (1997-98): Won from 6th, Man City (2020-21): The record-holders, winning from 8th.



```
SELECT
    season,
    champion_team AS team,
    xmas_rank,
    xmas_point,
    final_rank
FROM champions_at_xmas
WHERE xmas_rank >= 4
ORDER BY xmas_rank DESC;
```

	season	team	xmas_rank	xmas_point	final_rank
1	2020-21	Man City	8	23	1
2	1997-98	Arsenal	6	30	1
3	1996-97	Man United	5	31	1
4	2008-09	Man United	4	32	1

The Twist: A Clash of Personalities

On the surface, the history books favor the Christmas leader. But is there a twist hiding in the data?

By isolating specific 'Bottling' (losing the lead) and 'Hunting' (winning from behind) events in the modern era, we can profile the mental resilience of the top teams. The results reveal a stark contrast in personality between the two current title rivals

	team	"Bottled_Lead"	"Won_From_Behind"
1	Man United	1	4
2	Man City	0	3
3	Arsenal	3	2
4	Liverpool	4	0
5	Chelsea	0	0

The raw data was looking surprisingly kind to Arsenal... until I fixed it.

NEXT ▶

Because my dataset was limited, I manually injected the stats from the last two title races. Why? Because you cannot tell the story of the modern Premier League without mentioning the 'Elephants in the Room': The 2023 and 2024 collapses.

With these corrections, the chart transforms. It's no longer just a rivalry; it's a psychological profile of the Best Chaser (City) vs. the Ultimate Bottler (Arsenal)

	team	"Bottled_Lead"	"Won_From_Behind"
1	Man City	0	5
2	Man United	1	4
3	Arsenal	5	2
4	Liverpool	4	0
5	Chelsea	0	0

SQL query

```
95 SELECT
96     team,
97     -- ● Calculated from data + 2 Manual additions for Arsenal (22/23, 23/24)
98     COUNT(*) FILTER (WHERE xmas_rank = 1 AND final_rank > 1)
99     + CASE
100         WHEN team = 'Arsenal' THEN 2
101         ELSE 0
102     END AS "Bottled_Lead",
103
104     -- ● Calculated from data + 2 Manual additions for City (22/23, 23/24)
105     COUNT(*) FILTER (WHERE xmas_rank > 1 AND final_rank = 1)
106     + CASE
107         WHEN team = 'Man City' THEN 2
108         ELSE 0
109     END AS "Won_From_Behind"
110
111 FROM rankings
112 WHERE season >= '2000-01'
113     AND team IN ('Arsenal', 'Man City', 'Liverpool', 'Man United', 'Chelsea')
114 GROUP BY team
115 ORDER BY "Won_From_Behind" DESC, "Bottled_Lead" DESC;
```

NEXT ▶

Three Key Findings

#1: The "Coin Flip" Myth

"Being top at Christmas is not a golden ticket. With a 53.3% conversion rate, leading the pack is statistically equivalent to a coin flip. The pressure of being 'The Hunted' causes nearly half of all leaders to fail."

#2: The "Safe Zone" Rule

"While 1st place is precarious, the Top 2 is non-negotiable. 23 out of 30 champions (77%) came from the Top 2 spots at Christmas."

The Warning: If you are 3rd or lower, you are praying for a statistical anomaly (an outlier)."

#3: The "Personality" Factor

"History isn't just about position; it's about psychology. Our 'Hunter vs. Bottler' analysis proves that Man City thrives on the chase (The Hunter), while Arsenal statistically struggles to defend a lead (The Bottler)."

The Final Verdict: The Path to Glory

"For Arsenal to break their drought, history suggests they are safer chasing from 2nd than leading from 1st."

The Data Formula: To win the league, a team typically needs to be in the Top 2 at Christmas AND possess a 'Hunter' mentality.

Currently, the data favors the blue side of Manchester."

Analyst's Note: Football is complex. A perfect trend analysis would require data like xG (Expected Goals), Squad Depth, and Injury Reports, which are outside the scope of this project.

Ultimately, this was a fun challenge designed to flex my SQL muscles—using CTEs, Window Functions, and Views to turn raw match stats into a compelling data story

NEXT ▶

Thank you

I hope you enjoyed this deep dive into Premier League history. I am currently open to Data Analyst opportunities where I can apply my SQL, visualization, and storytelling skills.

MD Nowroz Imtiaz Rifat



nourose.rifat@gmail.com



github.com/n-im-Rifat