VINUNIVERSITY

Data Visualization EPL 2021-2022

Thai Ba Hung, Tran Le Hai, Dang Duc Dat

Short introduction

- **Title: Premier League Match Analysis (2021-2022)**

- **Why this match dataset?**
  - The Premier League is one of the most competitive football leagues.
  - We are passionate about football and want to analyze real match data.

- **What we aim to do:**
  - Q1: Identify the worst-performing team based on inefficiency and discipline.
  - Q2: Analyze player performance to determine which players should be sold.

Dataset overview

- **Data Source:**
  - Premier League Match Data 2021-2022 by **Evan Gower** on Kaggle.
  - EPL 21-22 Matches Players Dataset by **Azmine Toushik Wasi**.

- **Match Data (380 games, 22 columns):** Goals, shots, fouls, cards, results.

- **Player Stats (10 columns):** Goals, appearances, substitutions, disciplinary records.

- **Why Merge These Datasets?**
  - Match data helps analyze overall team performance.
  - Player data helps evaluate individual contributions.
  - Combining them allows **deeper insights** into team inefficiencies and underperforming players.

Q1. Which team playstyle is the least attractive?

Q1. Motivation

What audience like me should expect at a football match?

- A match with lots of goals and actions.
- A thrilling back-and-forth match.
- A fast-paced match with unpredictable result.

What does it mean for **“least attractive”** team?

- An unattractive playstyle will tend to be more practical.

  more defensive with lots of ways to interrupt the match.

- An optimal playstyle with few shots, less possession time, many fouls but still achieve acceptable results.

Q1. Processes

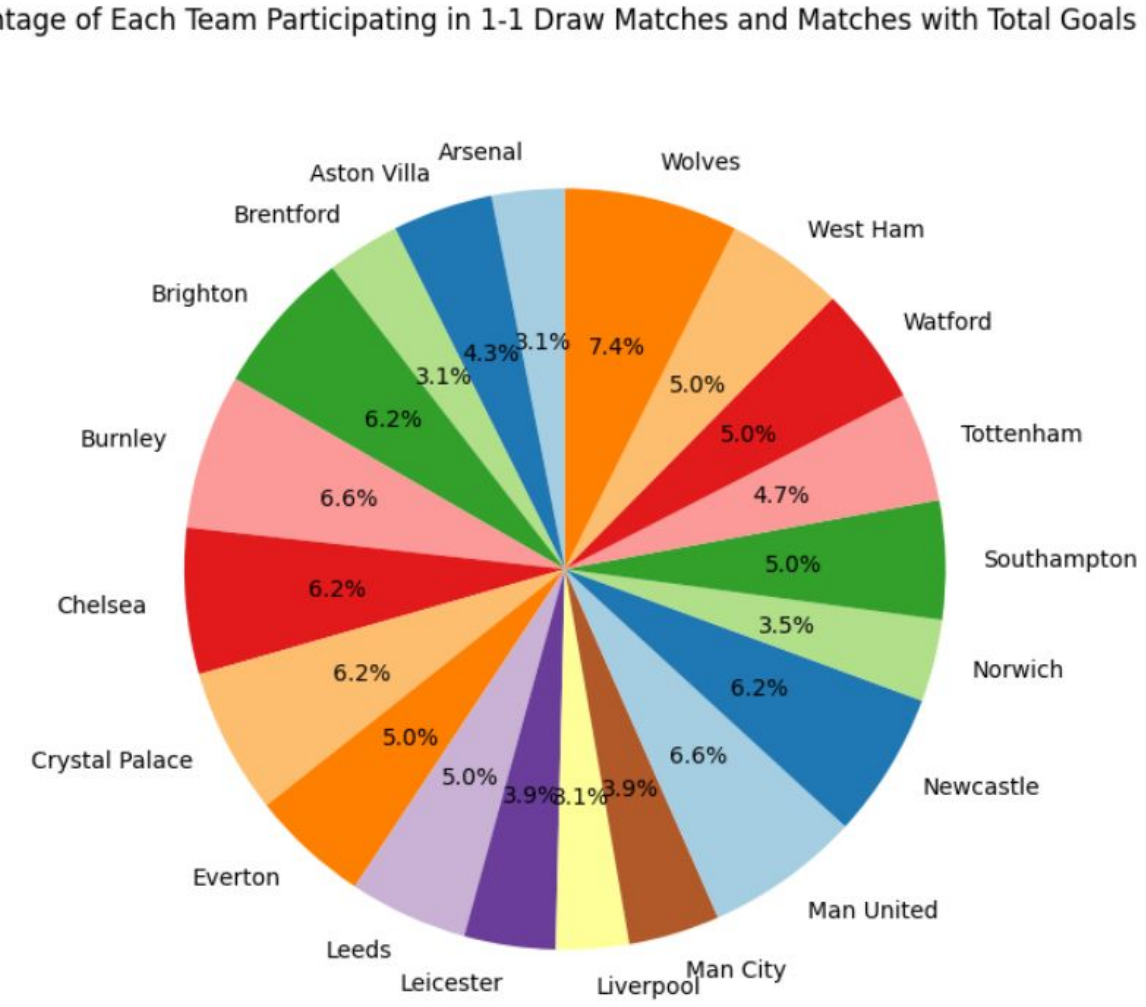
The variables should be considered from the two datasets to answer the questions:

- Full-time Result, Home/Away Goals, Home/Away Fouls, Home/Away Cards, Home/Away Shots.
- Weekly Ranking of Each Team.

The data visualization process that we use to answer the questions:

- Normalize the stats.
- View the least attractive matches
- View the least attractive weeks
- View the playstyle of each team as Home/Away
- Use a “dirty score” to decide which team is the “least attractive”.

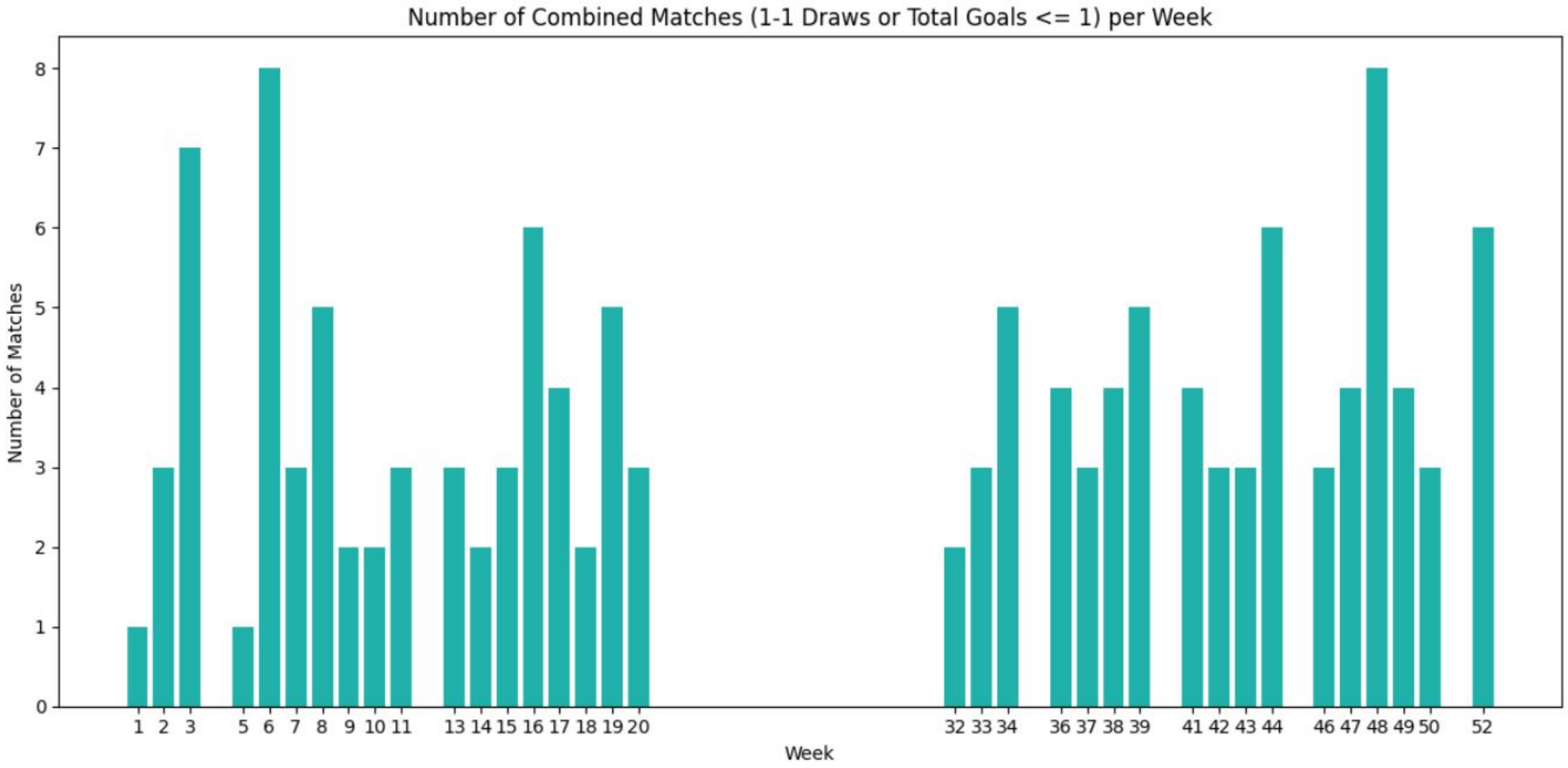
Q1. View the least attractive matches



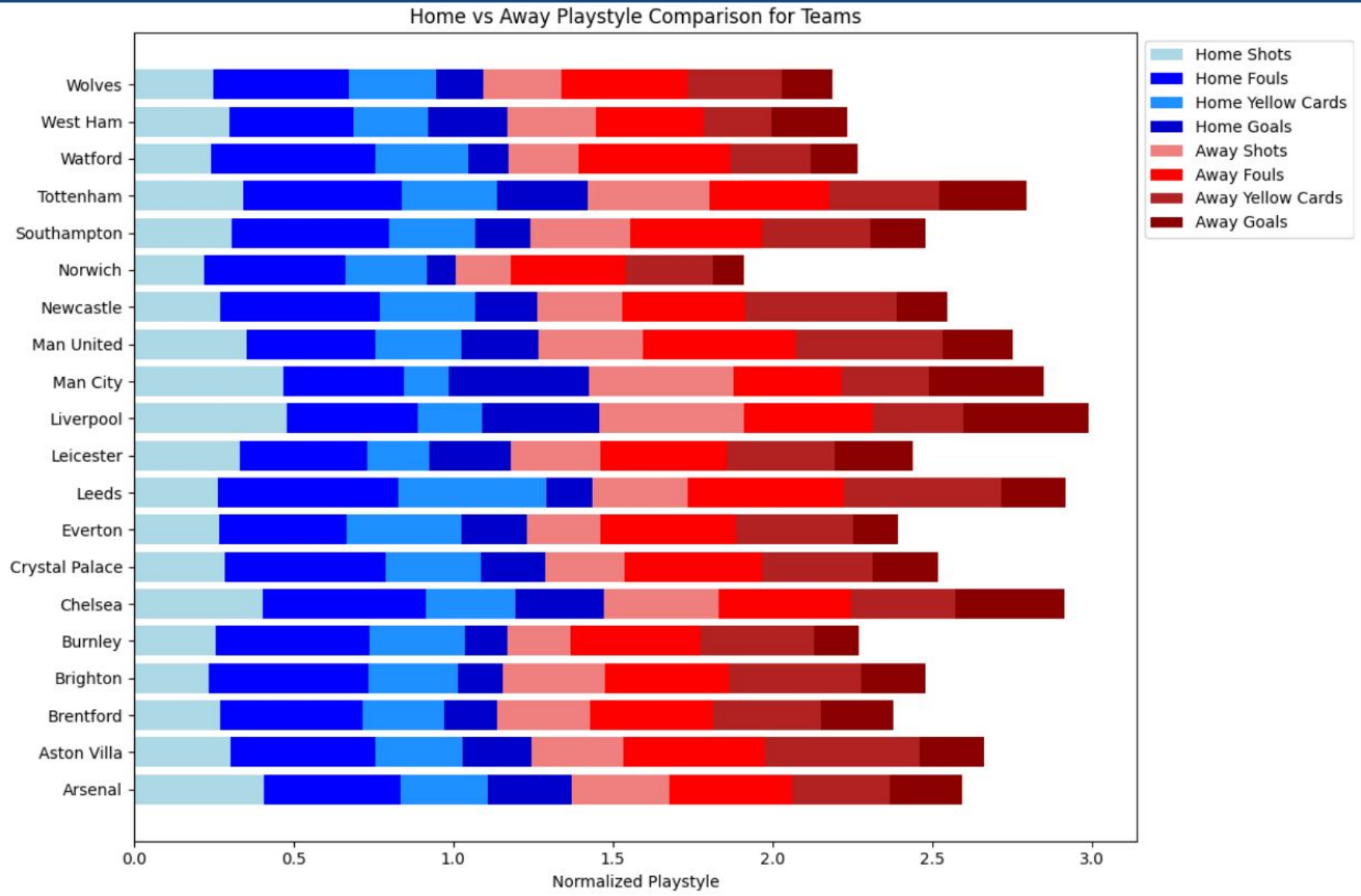
[32]:

	Draw 1-1 Percentage	Total Goals <= 1 Percentage
Brighton	23.684211	18.421053
Chelsea	21.052632	21.052632
Man United	21.052632	23.684211
Newcastle	21.052632	21.052632
Burnley	15.789474	28.947368
Crystal Palace	15.789474	26.315789
Leeds	15.789474	18.421053
Southampton	15.789474	18.421053
Leicester	13.157895	13.157895
West Ham	7.894737	26.315789

Q1. View the least attractive weeks



Q1. View the playstyle of each team as Home/Away

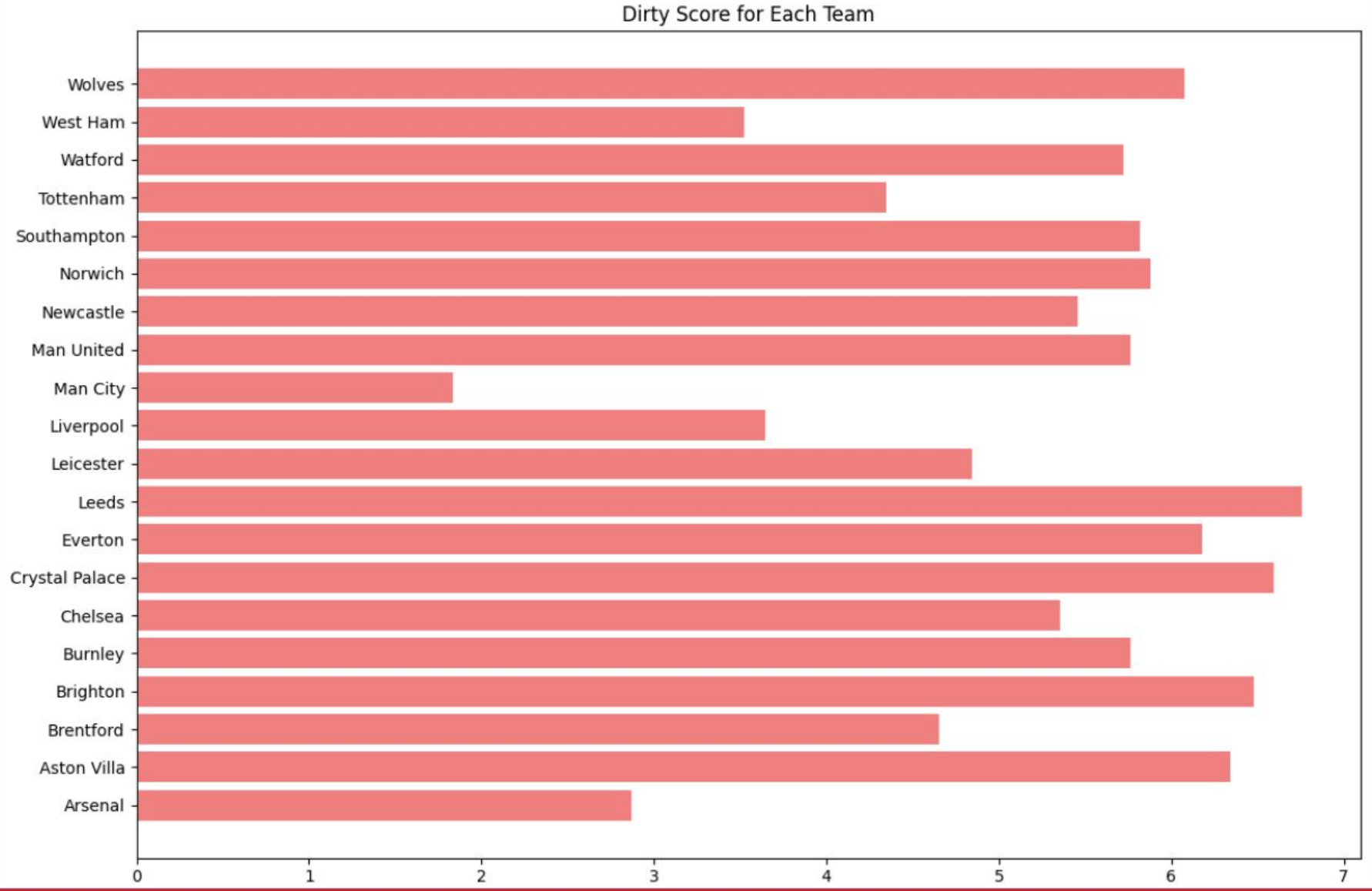


Q1. Decide the least attractive team

We use a metric call “dirty score”

Dirty Score=Normalized Fouls+Normalized Yellow Cards+(1−Normalized Shots) −Normalized Goals+Normalized 1-1 Matches+Normalized Opponent Strength

Q1. Decide the least attractive team



Q1. Decide the least attractive team

1.	Manchester City	38	29	6	3	99:26	73	**93**
2.	Liverpool	38	28	8	2	94:26	68	**92**
3.	Chelsea	38	21	11	6	76:33	43	**74**
4.	Tottenham	38	22	5	11	69:40	29	**71**
5.	Arsenal	38	22	3	13	61:48	13	**69**
6.	Manchester Utd	38	16	10	12	57:57	0	**58**
7.	West Ham	38	16	8	14	60:51	9	**56**
8.	Leicester	38	14	10	14	62:59	3	**52**
9.	Brighton	38	12	15	11	42:44	-2	**51**
10.	Wolves	38	15	6	17	38:43	-5	**51**
11.	Newcastle	38	13	10	15	44:62	-18	**49**
12.	Crystal Palace	38	11	15	12	50:46	4	**48**
13.	Brentford	38	13	7	18	48:56	-8	**46**
14.	Aston Villa	38	13	6	19	52:54	-2	**45**
15.	Southampton	38	9	13	16	43:67	-24	**40**
16.	Everton	38	11	6	21	43:66	-23	**39**
17.	Leeds	38	9	11	18	42:79	-37	**38**
18.	Burnley	38	7	14	17	34:53	-19	**35**
19.	Watford	38	6	5	27	34:77	-43	**23**
20.	Norwich	38	5	7	26	23:84	-61	**22**

Q2. Which players should be sold?

Reason for Analysis

- Analysing performance of each player (not only for fans but also coaches and clubs decision on giving chance to players)
- Expectancy of fans on their players
- Building a worst players squad

Forwards Analysis Metrics:

- **Goals & Scoring Efficiency:** Goals, Penalties, MinutesPerGoal
- **Team Attacking Metrics:** TeamGoals, TeamShotConversion
- **Player Participation:** Appearances, Substitutions
- **Custom Metric:** ForwardScore

Defenders Analysis Metrics:

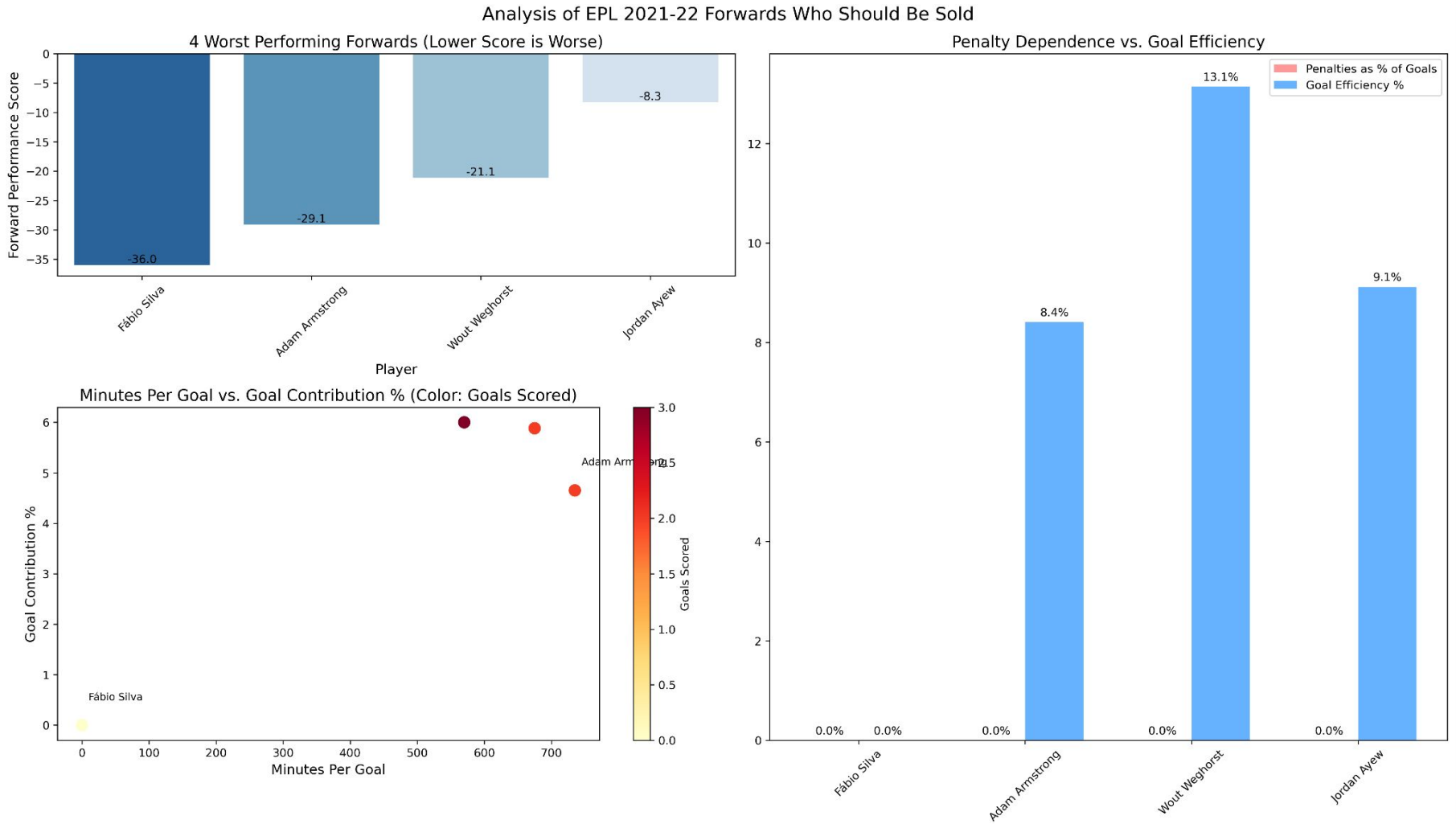
- **Defensive Actions:** Appearances, NumberofShotsAgainst, GoalConversion
- **Discipline:** Fouls, YellowCards, RedCards
- **Team Defense Metrics:** GoalsConceded, CleanSheets

Midfielderse Analysis Metrics:

- **Defensive Actions:** Appearances, NumberofShotsTaken, GoalConversion
- **Discipline:** Fouls, YellowCards, RedCards, CardRatio
- **Team Defense Metrics:** GoalsConceded, CleanSheets

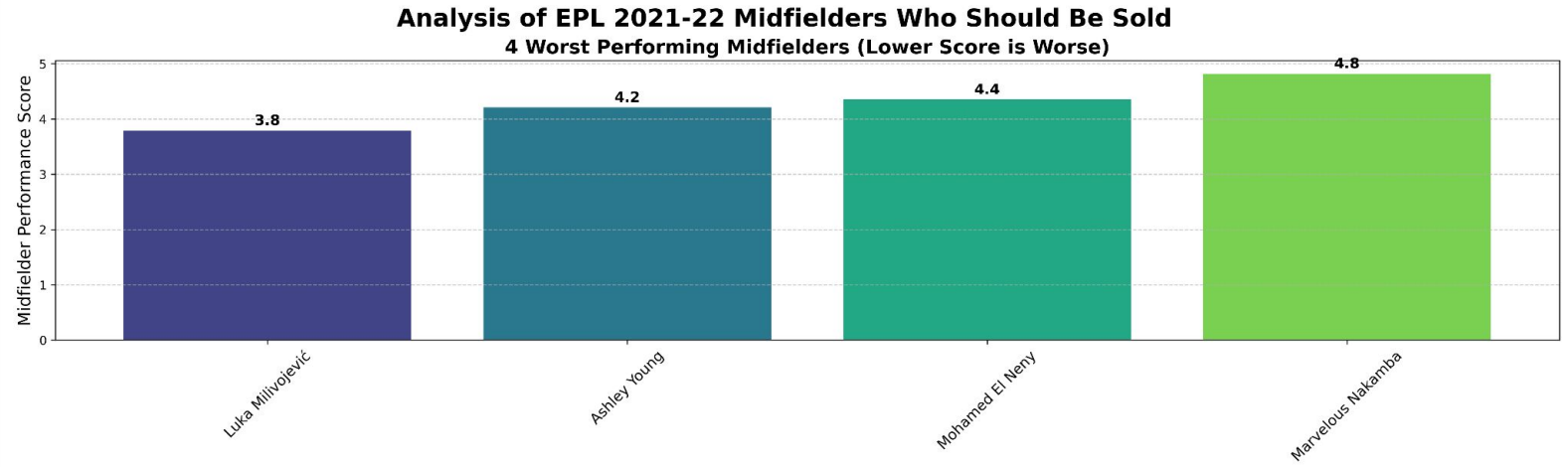
Q2. Which players should be sold?

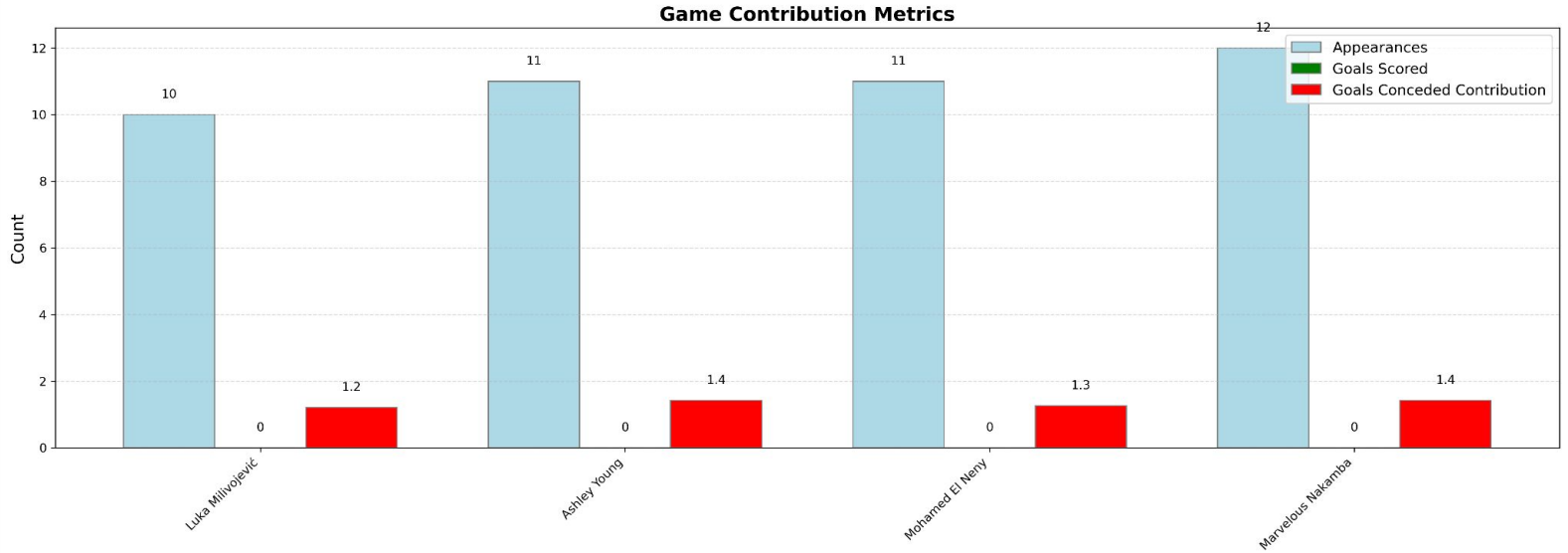
```
forwards['ForwardScore'] = (
    forwards['Goals'] * 10 +
    forwards['GoalContribution'] * 2 +
    (forwards['GoalEfficiency'] - 1) * 20 -
    forwards['MinutesPerGoal'] * 0.05 -
    forwards['SubstitutionRate'] * 0.1 -
    forwards['Penalties'] / forwards['Goals'].clip(lower=1) * 10
)
```



Q2. Which players should be sold?

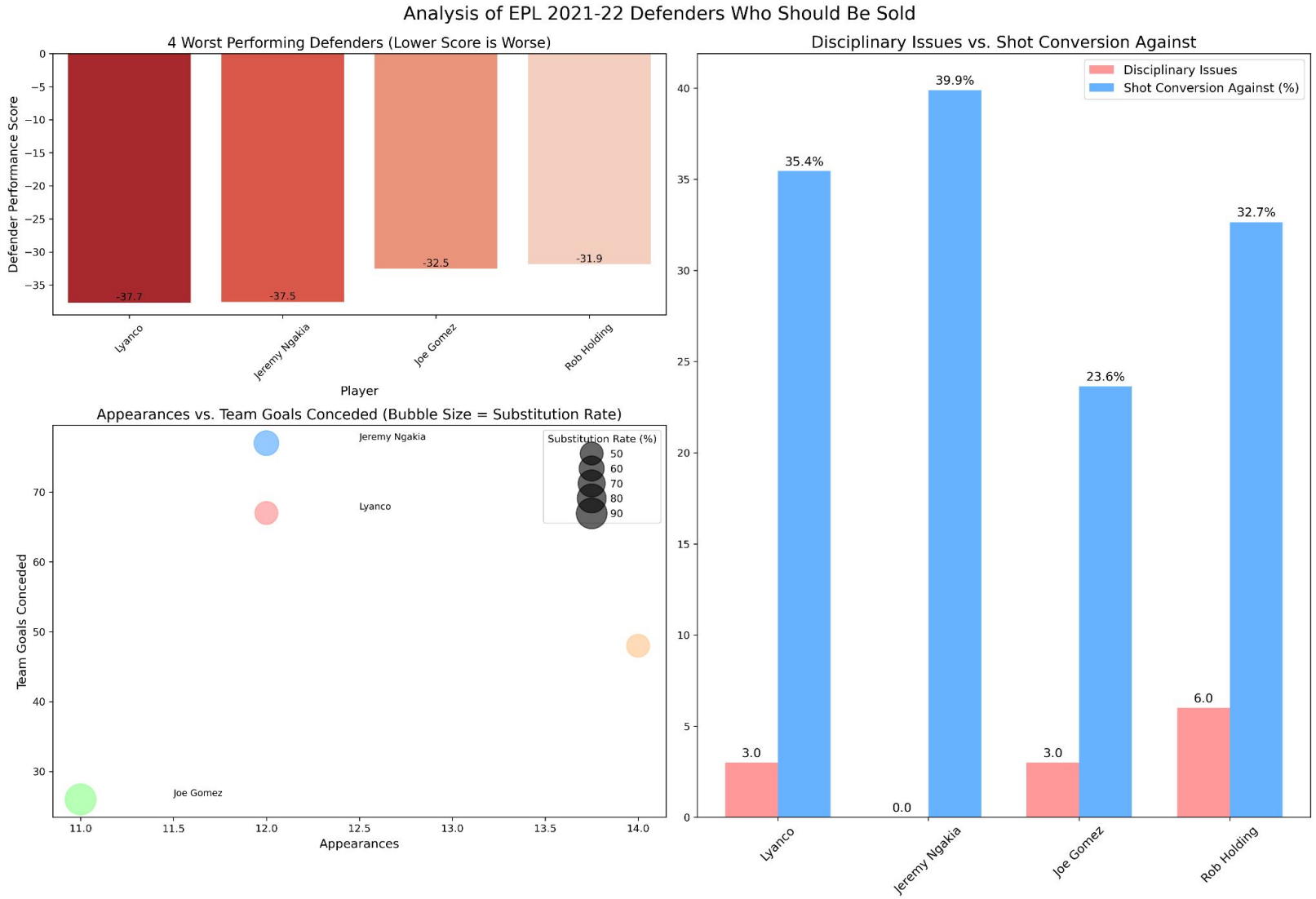
```
midfielders['MidfielderScore'] = (
    midfielders['Goals'] * 5 -
    midfielders['PlayerDefensiveContribution'] * 10 +
    midfielders['Apearances'] * 0.5
)
```





Q2. Which players should be sold?

```
defenders['DefenderScore'] = (
    defenders['AppearancePercentage'] * 0.3 +
    defenders['Goals'] * 5 -
    defenders['SubstitutionRate'] * 0.2 -
    defenders['DisciplinaryIssues'] * 2 -
    (defenders['GoalsConceded'] / 5) -
    defenders['ShotConversionAgainst'] * 0.5
)
```



Worst Squad of EPL 21-22

THANK YOU!