Scoring Efficiency in Elite Football: Identifying the Top Strikers of the 2022/23 Season

Muzaffer Cem Ates

Project Overview

This project aims to evaluate and compare the efficiency of top strikers from the Premier League, La Liga, and Bundesliga during the 2022/23 season. Using key performance metrics such as Goals per 90 minutes, Conversion Rate, and Shots on Target percentage, the analysis identifies the most efficient goal scorers across these elite competitions.

The methodology follows Google's data analysis process, structured into six distinct phases: Ask, Prepare, Process, Analyze, Share, and Act. This framework ensures a comprehensive approach to data handling, from defining the core questions to sharing actionable insights.

1. Ask Phase: Defining the Purpose and Questions

Purpose: The primary purpose of this analysis is to identify the most efficient strikers in the 2022/23 season from the Premier League, La Liga, and Bundesliga, using key performance metrics to evaluate and compare their effectiveness on the field.

Key Questions to Answer:

- Who was the most efficient striker in the 2022/23 season?
    - What defines "efficiency" for a striker? Is it goals per 90 minutes, conversion rate, or another metric?
- How do the top-performing strikers compare across the Premier League, La Liga, and Bundesliga?
    - Are there significant differences in striker performance between these leagues
- Which strikers demonstrate both high accuracy and efficiency in their scoring?
    - How do players rank in terms of "Shots on Target %" and "Conversion Rate"?
- What insights can be drawn about league difficulty and striker performance?
    - Do strikers in certain leagues tend to have higher or lower efficiency metrics?
- How can this analysis inform scouting, training, and tactical decisions?
    - Do strikers in certain leagues tend to have higher or lower efficiency metrics?

These questions will guide the rest of the analysis, ensuring that the project remains focused on answering relevant and impactful inquiries

2. Prepare

The second phase of the analysis involves preparing the data needed to address the questions defined in the Ask phase. This includes identifying and collecting relevant data, understanding its structure, and ensuring that it is suitable for the analysis. Proper data preparation ensures that the subsequent analysis is based on accurate and reliable information.

Data Source:

The dataset used in this analysis was sourced from [Kaggle](). For more details, you can check the website: https://www.kaggle.com/datasets/vivovinco/20222023-football-player-stats .It contains detailed player statistics for the 2022/23 season across multiple leagues, including the Premier League, La Liga, and Bundesliga. The dataset provides various metrics such as goals, assists, minutes played, shots, and shots on target, which are essential for evaluating player performance and efficiency.

Data Overview:

The some of the columns in the database:

- Player: Name of the player.
- Nation: Nationality of the player.
- Pos: Position played (e.g., FW for Forward).
- Comp: Competition or league (e.g., Premier League, La Liga, Bundesliga).
- Goals: Total goals scored.
- Min: Total minutes played.
- Shots: Total shots taken.
- SoT (Shots on Target): Number of shots on target.
- Goals per 90: Goals scored per 90 minutes played.
- Conversion Rate: Goals per shot taken.
- SoT%: Percentage of shots on target.

	Rk	Player	Nation	Pos	Squad	Comp	Age	Born	MP	Starts	...	Off	Crs	TklW	PKwon	PKcon	OG	Recov	AerWon	AerLost	AerWon%
0	198	Keidi Bare	ALB	MF	Espanyol	La Liga	25	1997	11	3	...	0.00	1.08	0.81	0.00	0.00	0.0	4.32	1.08	2.16	33.3
1	183	Ivan Balliu	ALB	DF	Rayo Vallecano	La Liga	31	1992	20	20	...	0.25	3.42	0.65	0.00	0.00	0.1	4.12	0.90	0.80	52.9
2	1360	Qazim LaCi	ALB	MF	Ajaccio	Ligue 1	27	1996	9	2	...	0.42	3.33	1.25	0.00	0.42	0.0	8.33	1.67	2.08	44.4
3	352	Armando Broja	ALB	FW	Chelsea	Premier League	21	2001	12	2	...	0.91	0.00	0.30	0.30	0.00	0.0	4.55	2.73	6.36	30.0
4	707	Berat Djimsiti	ALB	DF	Atalanta	Serie A	29	1993	10	6	...	0.00	0.00	1.16	0.00	0.00	0.0	5.36	2.03	1.30	60.9
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2684	587	Patson Daka	ZAM	FW	Leicester City	Premier League	24	1998	17	9	...	0.64	0.51	0.64	0.13	0.00	0.0	2.31	1.41	2.31	37.9
2685	189	Lameck Banda	ZAM	FWMF	Lecce	Serie A	22	2001	22	11	...	0.77	2.65	0.94	0.00	0.00	0.0	4.27	0.43	1.37	23.8
2686	1242	Tino Kadewere	ZIM	MFFW	Mallorca	La Liga	27	1996	5	0	...	0.00	1.33	0.00	0.00	0.00	0.0	4.00	6.00	2.00	75.0
2687	1739	Marshall Munetsi	ZIM	MF	Reims	Ligue 1	26	1996	19	16	...	0.13	0.69	0.94	0.00	0.00	0.0	5.75	1.75	1.31	57.1
2688	2665	Jordan Zemura	ZIM	DF	Bournemouth	Premier League	23	1999	16	14	...	0.00	1.04	1.48	0.00	0.07	0.0	4.67	0.44	0.74	37.5

2689 rows × 124 columns

Fig. 1: The database

Data Collection and Security Considerations:

The dataset was downloaded from Kaggle, a reputable data platform, and stored securely for analysis. As the dataset is publicly available and does not contain sensitive personal information, the primary focus was on ensuring data accuracy and integrity.

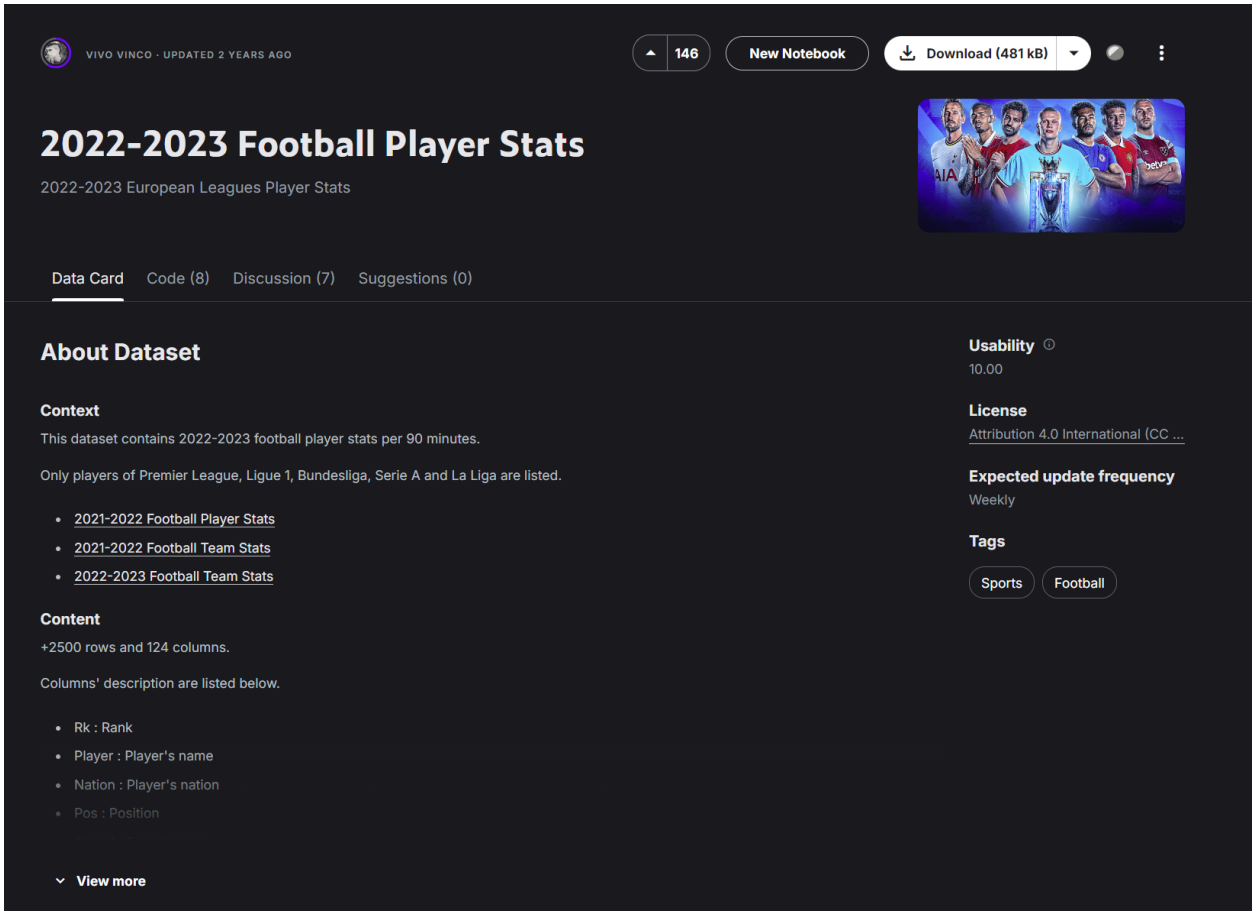

Table 1: Kaggle dataset page

3. Process

The third phase involves processing the data to ensure its accuracy and consistency. This step is crucial to identify and address any inaccuracies, errors, or inconsistencies in the dataset. Proper data processing prevents any issues that could affect the credibility of the analysis and ensures that the findings are based on reliable data.

Data Cleaning and Validation:

- **Handling Missing Values:**

After a thorough examination, it was found that there were no missing values in the dataset for any of the critical columns, such as Goals, Minutes Played, Shots, and Shots on Target. As a result, no changes were made to the dataset in this regard.

- **Removing Duplicates:**

There were two players named "Rodrigo" in the dataset, which could cause confusion during analysis. To resolve this, the last name "Sanchez" was added to one of the players, ensuring that each player has a unique identifier. This correction clarified the data and prevented any potential mix-ups.

- **Filtering Based on Playtime and Position:**
  - The dataset was filtered to include only players who have played more than 10 matches (at least 900 minutes) to ensure that only players with significant playing time were analyzed.
  - Additionally, only players in forward positions (FW, FWDF, FWMF) were considered, focusing the analysis on strikers.

```
[ ] main_data = filtered_player[(filtered_player["Min"]>=900) & (filtered_player["Goals"]>0) & (filtered_player["SoT"]>0) & (filtered_player["Shots"]>0)]
    main_data
```

	Rk	Player	Nation	Pos	Squad	Comp	Age	Born	MP	Starts	...	Off	Crs	TklW	PKwon	PKcon	OG	Recov	AerWon	AerLost	AerWon%
1674	897	Chris Fuhrich	GER	FWMF	Stuttgart	Bundesliga	25	1998	19	13	...	0.15	2.82	0.92	0.00	0.0	0.0	4.66	0.08	0.31	20.0
1669	2412	Simon Terodde	GER	FW	Schalke 04	Bundesliga	34	1988	20	16	...	1.55	0.14	0.49	0.07	0.0	0.0	2.18	3.38	5.49	38.1
1698	900	Niclas Fullkrug	GER	FW	Werder Bremen	Bundesliga	30	1993	19	19	...	0.92	0.81	0.22	0.00	0.0	0.0	3.35	5.14	3.68	58.3
1697	744	Marvin Ducksch	GER	FW	Werder Bremen	Bundesliga	28	1994	20	20	...	0.32	5.19	0.16	0.05	0.0	0.0	3.46	1.03	1.08	48.7
1655	370	Marius Bulter	GER	FWMF	Schalke 04	Bundesliga	29	1993	20	17	...	0.24	4.04	0.30	0.00	0.0	0.0	4.34	1.69	2.29	42.4
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
617	1544	Solly March	ENG	FWDF	Brighton	Premier League	28	1994	21	21	...	0.45	4.63	0.85	0.00	0.0	0.0	4.08	1.14	1.54	42.6
604	2310	Dominic Solanke	ENG	FWMF	Bournemouth	Premier League	25	1997	17	16	...	0.31	0.25	0.25	0.00	0.0	0.0	3.08	1.57	2.58	37.9
1755	151	Jordan Ayew	GHA	FWMF	Crystal Palace	Premier League	31	1991	22	18	...	0.06	1.14	1.81	0.00	0.0	0.0	5.48	1.69	3.55	32.2
613	2452	Ivan Toney	ENG	FW	Brentford	Premier League	26	1996	20	20	...	0.85	0.85	0.50	0.15	0.0	0.0	2.96	3.47	3.72	48.3
1727	1083	Kai Havertz	GER	FWMF	Chelsea	Premier League	23	1999	21	19	...	1.36	0.28	0.62	0.00	0.0	0.0	2.99	2.88	2.15	57.3

129 rows × 124 columns

Table 2: The cleared and filtered dataset

4. Analyze

The fourth phase focuses on analyzing the data to uncover relationships, trends, and patterns that can provide meaningful insights into striker performance. This phase involves thinking analytically about the data, formatting it for better understanding, and identifying key findings that address the questions defined in the Ask phase.

The metrics decided in order to examine the most efficient striker of the 2022/23 season in Premier League, La liga and Bundesliga:

- **Goals per 90 Minutes: (Goals / Minutes Played) * 90:** This metric is crucial for evaluating efficiency because it accounts for different playing times. A player who scores 10 goals in 900 minutes (1 goal per 90 minutes) is more efficient than a player who scores 10 goals in 1800 minutes (0.5 goals per 90 minutes).
- **Shots on Target Percentage (SoT%):** (Shots on Target / Total Shots) * 100 : A higher SoT% suggests that the player is good at directing their shots towards the goal, which is a key skill for a striker. However, it doesn't account for the quality of those shots—just whether they hit the target.
- **Goals per Shot (G/Sh): Goals / Shots :** This metric directly measures shooting efficiency. A higher G/Sh value indicates that the player is more effective in converting their shot attempts into goals.
- **Goals per Shot on Target (G/SoT):** Goals / Shots on Target : This metric goes a step further than G/Sh by evaluating the quality of shots that are on target. A high G/SoT indicates that the player is very effective at converting their on-target shots into goals, reflecting strong finishing skills.
- **Conversion Rate :(Goals / Total Shots) * 100 :** It offers a more intuitive view, especially when comparing multiple players. A higher conversion rate suggests a player is more clinical and converts a greater proportion of their opportunities into goals.
- **Minutes per Goal (Min/Goal): Minutes Played / Goals :** A lower Min/Goal value indicates greater efficiency, as the player needs less time on the pitch to score. It is particularly useful when comparing players with different playing times.

In this analysis, I included only those players who have participated in more than 10 games (equivalent to over 900 minutes of playing time) and have recorded at least one shot, one goal, and one shot on target. This criterion ensures that all players analyzed have a meaningful level of participation and performance, and it prevents division by zero errors during metric calculations.

5. **Share**

The fifth phase of the analysis involves sharing the insights derived from the data. This phase aims to communicate findings effectively, using visualizations and narratives that can be easily understood by stakeholders. In this project, three dashboards were created to present the key insights:

Scoring Efficiency Analysis

Graphs Used:

- **Goals Per 90 Minutes by Player**
- **Conversion Rate vs. Goals per 90**
- **Top Players by Conversion Rate**

Interpretation: This dashboard highlights the efficiency of strikers in terms of scoring relative to their playing time and shot conversion rate. Erling Haaland emerges as the clear leader across all three graphs. Despite scoring a high number of goals, his efficiency metrics, such as Goals per 90 and Conversion Rate, are significantly higher than any other player, demonstrating his superior scoring ability. According to the Conversion Rate vs. Goals per 90 graph, Haaland is far ahead of the second-most efficient striker, with Ivan Toney, Vincenzo Grifo, and Harry Kane following as the other top performers. This dashboard illustrates that merely scoring a large number of goals does not necessarily equate to efficiency. The true measure of efficiency combines both scoring frequency and shot effectiveness, which Haaland excels at.

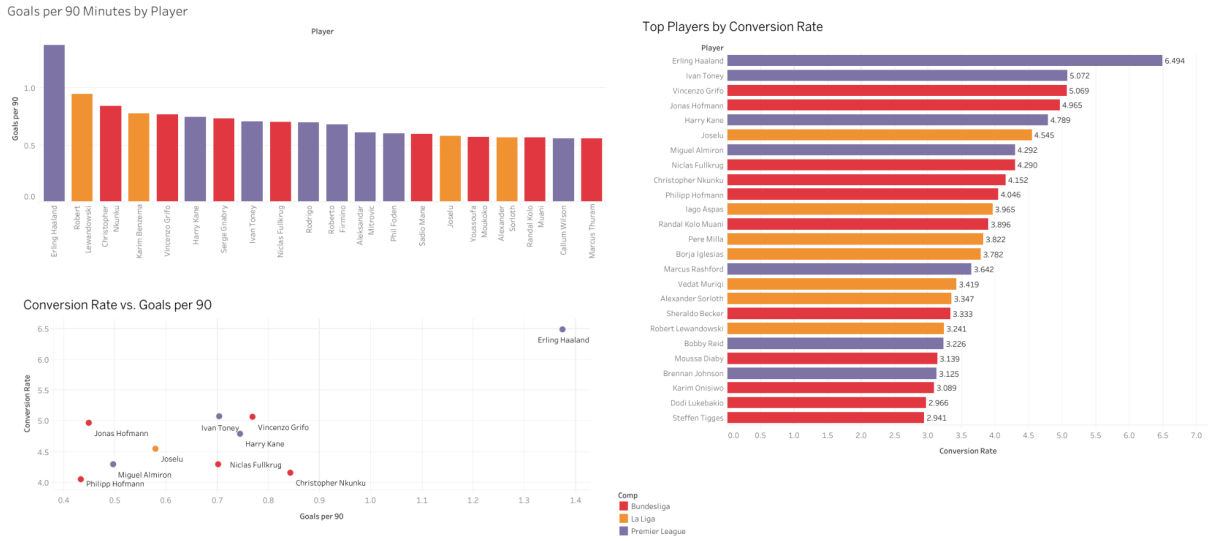

Fig 2: Scoring Efficiency Analysis

Player Accuracy and Efficiency Dashboard

Graphs Used:

- **Conversion Rate vs. Shots on Target Percentage**
- **Shots on Target per Goal vs. Goals per 90**

Interpretation: This dashboard focuses on the relationship between shooting accuracy and efficiency. It demonstrates the clustering of players based on their performance in these metrics. Once again, Erling Haaland stands out as the most efficient and accurate striker. The scatter plots show that Haaland consistently scores with fewer shots on target, indicating a high level of efficiency. His position on the graphs is noticeably separated from other players, confirming his dominance. The closest competitors in this regard still trail significantly, making Haaland's performance even more impressive.

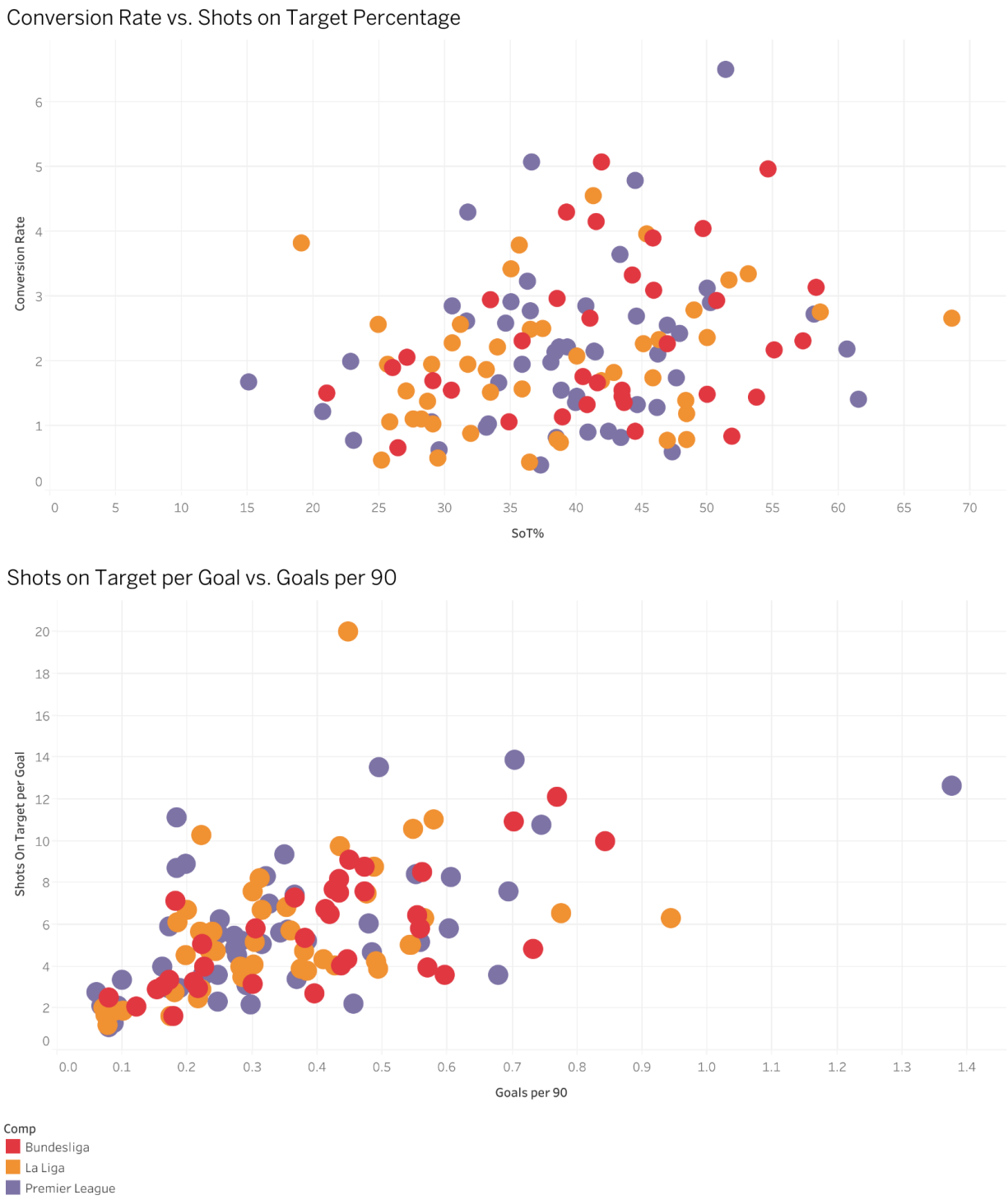

Fig 3: Player Accuracy and Effectiveness

League and Player Performance Dashboard

Graphs Used:

- **Minutes per Goal by Competition**
- **Total Goals Scored by Player**

Interpretation: This dashboard explores the difficulty of scoring goals across different leagues and how players perform within these contexts. The `Minutes per Goal by Competition` graph indicates that the Premier League is the toughest league to score in, as evidenced by the higher `Minutes per Goal` values compared to La Liga and the Bundesliga. Despite this, Erling Haaland's performance remains exceptional, highlighting his ability to maintain high efficiency in a challenging environment. This context amplifies the significance of his achievements and places his efficiency metrics in perspective relative to the competitive nature of the Premier League.

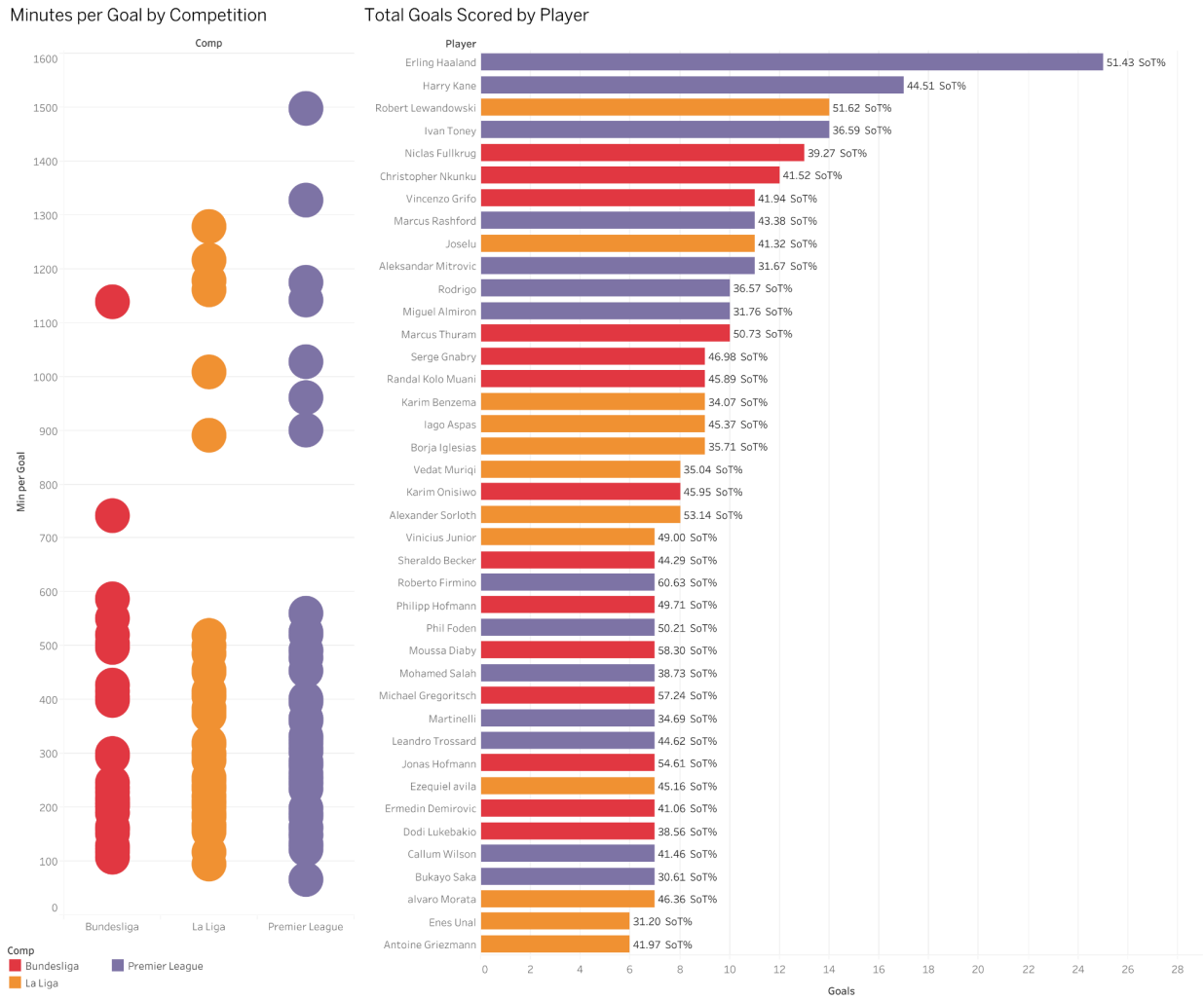

Fig 4: League and Player Performance Dashboard

6. Act

The final phase of the analysis is to act upon the insights gained. This involves making data-driven decisions or recommendations based on the findings. In this project, the analysis highlighted the top-performing strikers in Europe's elite leagues, providing valuable insights into their efficiency and accuracy.

The analysis identified Erling Haaland as the most efficient striker of the 2022/23 season, demonstrating his dominance across multiple key metrics. Other top performers include Ivan Toney, Vincenzo Grifo, and Harry Kane, who have also showcased exceptional scoring efficiency and accuracy.

	Player	Nation	Squad	Pos	Comp	Age	Min	Goals	SoT	Shots	Goals_per_90	SoT%	Conversion_Rate	G_per_ShotOnTarget	Min_per_Goal
2261	Erling Haaland	NOR	Manchester City	FW	Premier League	22	1636	25	1.98	3.85	1.375306	51.428571	6.493506	12.626263	65.440000
613	Ivan Toney	ENG	Brentford	FW	Premier League	26	1789	14	1.01	2.76	0.704304	36.594203	5.072464	13.861386	127.785714
1830	Vincenzo Grifo	ITA	Freiburg	FWMF	Bundesliga	29	1289	11	0.91	2.17	0.768037	41.935484	5.069124	12.087912	117.181818
1624	Jonas Hofmann	GER	M'Gladbach	FW	Bundesliga	30	1402	7	0.77	1.41	0.449358	54.609929	4.964539	9.090909	200.285714
729	Harry Kane	ENG	Tottenham	FW	Premier League	29	2055	17	1.58	3.55	0.744526	44.507042	4.788732	10.759494	120.882353
898	Joselu	ESP	Espanyol	FW	La Liga	32	1710	11	1.00	2.42	0.578947	41.322314	4.545455	11.000000	155.454545
2277	Miguel Almiron	PAR	Newcastle Utd	FW	Premier League	29	1816	10	0.74	2.33	0.495595	31.759657	4.291845	13.513514	181.600000
1698	Niclas Fullkrug	GER	Werder Bremen	FW	Bundesliga	30	1667	13	1.19	3.03	0.701860	39.273927	4.290429	10.924370	128.230769
1157	Christopher Nkunku	FRA	RB Leipzig	FWMF	Bundesliga	25	1282	12	1.20	2.89	0.842434	41.522491	4.152249	10.000000	106.833333
1534	Philipp Hofmann	GER	Bochum	FW	Bundesliga	29	1456	7	0.86	1.73	0.432692	49.710983	4.046243	8.139535	208.000000

Table 3: Top 10 Most Efficient Strikers of the 2022/23 Season

These players consistently ranked highly across metrics such as `Goals per 90`, `Conversion Rate`, and `Shots on Target Percentage`. Their performances provide valuable insights for teams, scouts, and analysts in making informed decisions regarding player development, scouting, and tactical planning.

The dashboards provide a comprehensive view of striker performance across key metrics. Erling Haaland's dominance in terms of both scoring frequency and efficiency is evident throughout the analysis. His ability to perform at such a high level in one of the toughest leagues underscores his exceptional talent. Other notable performers include Ivan Toney, Vincenzo Grifo, and Harry Kane, who also demonstrate high efficiency and accuracy in their scoring.

These insights can be valuable for understanding player performance, scouting, and tactical planning. By sharing this information through well-structured dashboards, stakeholders can gain a clear and actionable understanding of striker efficiency in the 2022/23 season.