Analysis Report: Visualization of Goals Data (1872 to 2024)

Objective

The code aims to analyze and visually compare:
1. The proportion of own goals versus scored goals.
2. The proportion of penalty goals versus non-penalty goals.
This is achieved through pie charts using data from the 'Goalscorers.csv' file.

Process Overview

1. Data Loading:
- The code reads data from a CSV file named 'Goalscorers.csv'.
- Columns of interest are 'own_goal' and 'penalty'.

2. Data Preprocessing:
- Data from the 'own_goal' and 'penalty' columns is converted into NumPy arrays.
- The 'Counter' function from the 'collections' module is used to count the occurrences of each unique value in these arrays.
- Two dictionaries are created to store these counts for own goals and penalty data.

3. Visualization:
- Two pie charts are created:
  - Chart 1: Comparison of own goals vs scored goals.
  - Chart 2: Comparison of penalty goals vs non-penalty goals.
- Labels and percentages are added for better readability.

Insights from the Visualizations

1. Own Goals vs Scored Goals (1876 to 2024):
- Scored Goals dominate, comprising 98.1% of the total goals.
- Own Goals make up a small fraction, accounting for only 1.9%.
- This highlights the rarity of own goals in football over the observed period.

2. Penalty Goals vs Non-Penalty Goals (1876 to 2024):
- Non-Penalty Goals form the majority, representing 93.3% of the total goals.
- Penalty Goals constitute 6.7%, indicating penalties are less frequent compared to open-play goals.

Observations

- Data Accuracy: The code processes the data effectively, but the accuracy of insights depends on the completeness and correctness of the input data.
- Visualization:
  - Pie charts are an effective choice for presenting proportional data.
  - The use of percentages and clear labeling enhances understanding.

- Style: The 'fivethirtyeight' style improves aesthetics and ensures professional presentation.

Potential Improvements

1. Error Handling:
- Add checks to ensure the input CSV file exists and contains valid data.
2. Dynamic Titles:
- Automatically update chart titles to reflect the actual range of years based on the dataset.
3. Additional Insights:
- Include total counts of goals and penalties for context.
- Add a bar chart or trend line to visualize changes over time if year-wise data is available.

Conclusion

The analysis and visualizations provide a clear and concise summary of the data, highlighting key insights about goals in football from 1876 to 2024. The charts effectively showcase the dominance of scored goals over own goals and non-penalty goals over penalty goals.