Documentation for Sub-Project: Analysis and Visualization of Own Goals and Penalty Goals (1872 to 2024)

Overview

This sub-project is a part of the broader analysis and visualization of international football results from 1872 to 2024. The objective is to analyze the trends of own goals and penalty goals during this period, using a dataset of goal scorers and presenting the insights via visualizations.

Dataset Information

- **File Format**: CSV
- **Dataset Name**: Goalscorers.csv
- **Relevant Columns**:
    - own_goal: Indicates whether a goal is an own goal.
    - penalty: Indicates whether a goal is a penalty goal.

Steps for Data Processing and Visualization

1. Importing Required Libraries

The following Python libraries are used in this project:

- **matplotlib.pyplot**: For generating visualizations.
- **pandas**: For data manipulation and reading the CSV file.
- **numpy**: For handling array data.
- **collections.Counter**: For counting occurrences of values in data.

2. Data Reading

The dataset is loaded using pandas.read_csv() from the file path:

pd.read_csv("Data(file_format_CSV)\Goalscorers.csv")

3. Data Extraction and Transformation

- Convert the own_goal and penalty columns to NumPy arrays for efficient processing.
- Use the Counter class to calculate the frequency of unique values in these columns.
- Separate the data into meaningful categories:

- own_goal: **Scored goals** and **Own goals**
- penalty: **Non-penalty goals** and **Penalty goals**

4. Data Preparation for Visualization

The categories and their respective frequencies are stored in lists:

- new_list_OwnGoalsData_check contains labels for own goals and scored goals.
- new_list_PenaltyData_check contains labels for penalty and non-penalty goals.
- Corresponding frequencies are stored in new_list_OwnGoalsData_popn and new_list_PenaltyData_popn.

5. Visualization

- **Figure Layout**: A 1x2 grid is created using matplotlib for side-by-side pie charts.
- **Pie Charts**:
  - **Chart 1**: Distribution of own goals vs scored goals.
  - **Chart 2**: Distribution of penalty goals vs non-penalty goals.
- **Customization**:
  - Labels, autopct, and legends are added for clarity.
  - Titles are set to contextualize each chart.

6. Styling

The fivethirtyeight style is used for a professional look:

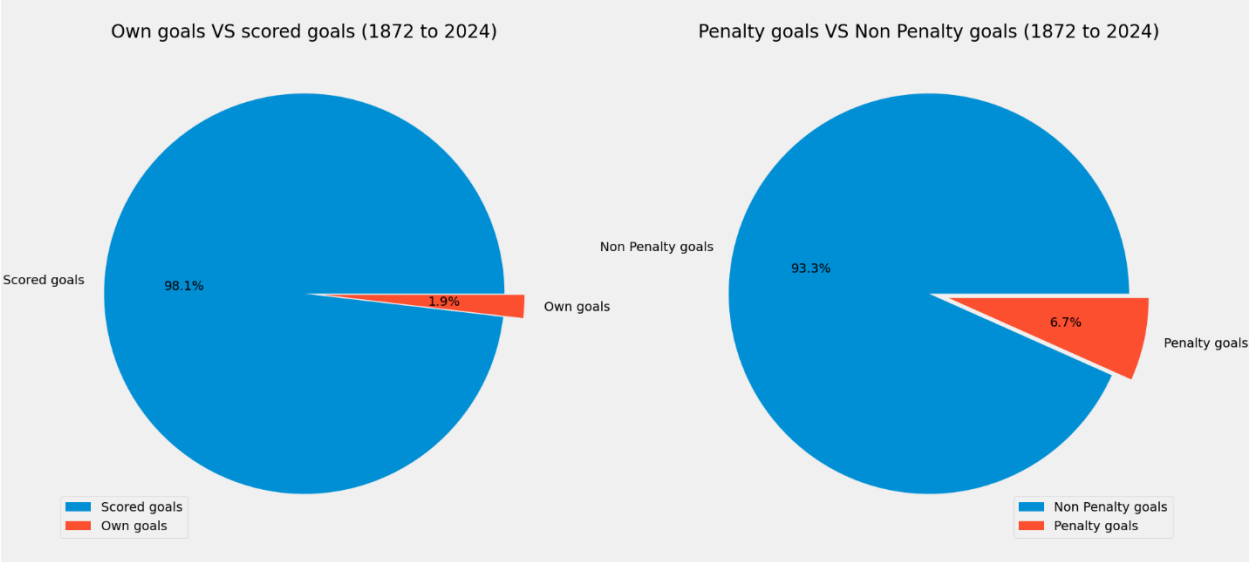
plt.style.use("fivethirtyeight")

Visualization Output

The output consists of two pie charts:

1. **Own Goals vs Scored Goals**:
   - Percentage distribution of own goals and regular goals scored from 1876 to 2024.
2. **Penalty Goals vs Non-Penalty Goals**:
   - Percentage distribution of goals scored from penalties vs those scored through open play or other means.

Visualization:



Code Walkthrough

Below is the Python code used to generate the visualizations:

```
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
from collections import Counter

plt.style.use("fivethirtyeight")

data = pd.read_csv("Data(file_format_CSV)\Goalscorers.csv")
OwnGoalsData = np.array(data["own_goal"], dtype = str)
Penaltydata = np.array(data["penalty"], dtype = str)

list_OwnGoalsData = dict(Counter(OwnGoalsData))
list_PenaltyData = dict(Counter(Penaltydata))

new_list_OwnGoalsData_check = ['Scored goals', 'Own goals']
```

```
new_list_OwnGoalsData_popn = []

for items in list_OwnGoalsData:
    new_list_OwnGoalsData_popn.append(list_OwnGoalsData.get(items))

new_list_PenaltyData_check = ['Non Penalty goals', 'Penalty goal']
new_list_PenaltyData_popn = []

for items in list_PenaltyData:
    new_list_PenaltyData_popn.append(list_PenaltyData.get(items))

fig, axs = plt.subplots(1, 2, figsize=(22,15))
axs[0].pie(new_list_OwnGoalsData_popn, labels = new_list_OwnGoalsData_check, autopct = '%1.1f%%')
axs[0].legend(loc = 'lower left')
axs[0].set_title('Own goals VS scored goals (1876 to 2024)')

axs[1].pie(new_list_PenaltyData_popn, labels = new_list_PenaltyData_check, autopct = '%1.1f%%')
axs[1].legend(loc = 'lower right')
axs[1].set_title('Penalty goals VS Non Penalty goals (1876 to 2024)')

plt.tight_layout()
plt.show()
```

Conclusion

This analysis highlights:

- The proportion of own goals relative to total goals scored.
- The prevalence of penalty goals over non-penalty goals in international football history.

Future Scope

- Extend this analysis to include trends over specific time periods.
- Correlate own goals and penalty goals with match outcomes.
- Enhance visualizations with interactive dashboards.
