Documentation: Visualizing Top Goal-Scoring Countries of All Time

Overview

This script visualizes the top 4 goal-scoring countries of all time using a Pie Chart and a Horizontal Bar Chart. The data is sourced from a CSV file containing records of matches played by various teams, and the visualizations highlight the most successful nations in terms of scoring goals.

Features

Data Aggregation:

Combines data for both home_team and away_team to calculate the total goals scored by each country.

Top 4 Analysis:

Extracts the top 4 goal-scoring countries based on their total goal count.

Dual Visualizations:

A Pie Chart displays the proportion of goals contributed by each of the top 4 countries.

A Horizontal Bar Chart ranks the countries by the total number of goals scored.

Customization:

Enhanced readability through styles, labels, titles, and an "exploded" slice in the pie chart for better emphasis.

Dependencies

The script requires the following libraries:

matplotlib: For creating visualizations.

pandas: For reading and handling CSV data.

collections.Counter: For counting and aggregating data efficiently.

Code Breakdown

1. Import Required Libraries

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

matplotlib.pyplot: For plotting.

pandas: To read and manipulate the CSV file.

Counter: To efficiently count occurrences of teams in both home and away matches.

2. Set Plotting Style

```
plt.style.use('fivethirtyeight')
```

Uses the "fivethirtyeight" style to make the plots visually appealing and professional.

3. Read the Data

```
data = pd.read_csv('Data(file_format_CSV)\Goalscorers.csv')
home_team = data['home_team']
away_team = data['away_team']
```

Reads the CSV file containing match data.

Extracts home_team and away_team columns for goal-scoring data.

4. Aggregate Goal Data

```
count_nations = Counter(home_team)
count_nations.update(away_team)
```

Counts the number of occurrences of each team in the home_team column.

Updates the count by including goals scored by teams in the away_team column.

5. Identify Top 4 Goal-Scoring Countries

```
Top3_count_nations = count_nations.most_common(4)
countries = [items[0] for items in Top3_count_nations]
scores = [items[1] for items in Top3_count_nations]
```

Retrieves the top 4 countries with the highest goal counts.

Splits the data into two lists:

countries: Names of the top 4 countries.

scores: Corresponding total goals scored.

6. Create Subplots

```
fig, axs = plt.subplots(1,2, figsize=(12,24))
```

Creates two side-by-side subplots with a specified figure size.

7. Plot Pie Chart

```
explode = [0.06, 0, 0, 0]
axs[0].pie(scores, labels=countries, explode=explode, autopct="%1.1f%%", shadow=True)
axs[0].legend(loc='upper left')
axs[0].set_title('Top 4 Goal scoring Countries of all Time\nPie Chart')
```

explode: Highlights the first country (largest contributor) by separating it slightly from the rest.

autopct: Displays percentage contributions up to one decimal place.

shadow: Adds a shadow effect for better aesthetics.

Legend: Adds a legend and title to the chart.

8. Plot Horizontal Bar Chart

```
countries.reverse()
scores.reverse()
axs[1].barh(countries, scores)
axs[1].set_xlabel('No of Goals')
axs[1].set_title('Top 4 Goal scoring Countries of all Time\nBar Chart')
```

Reverses the countries and scores lists to rank countries in descending order of goals scored.

Creates a horizontal bar chart using barh().

Adds an x-axis label and a title for better clarity.

9. Display the Plots

```
plt.show()
```

Renders both the Pie Chart and Horizontal Bar Chart side by side.

Visualization Output

Pie Chart:

Shows the percentage contribution of goals scored by the top 4 countries.

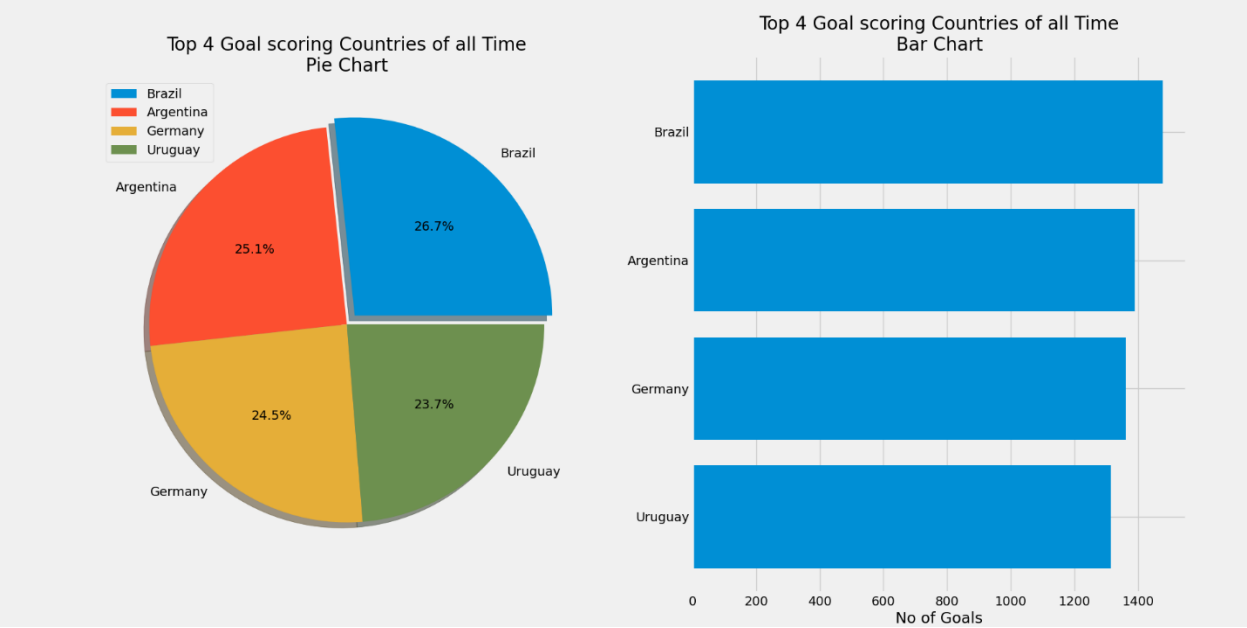
The first country is highlighted for emphasis.

Horizontal Bar Chart:

Displays the exact goal counts of the top 4 countries in descending order.

Provides a clear comparison of total goals.

Visualization:



Key Highlights

The Pie Chart provides a high-level summary of goal distribution among the top 4 countries.

The Bar Chart offers a detailed view, making it easy to compare the goal-scoring statistics.

Effective use of styling, labels, legends, and titles enhances readability and visual appeal.

How to Use

Replace 'Data(file_format_CSV)\Goalscorers.csv' with the path to your CSV file.

Run the script to generate the visualizations.

Conclusion

This script demonstrates how to perform basic data analysis and visualization using Matplotlib and Pandas. It effectively communicates the dominance of specific countries in goal-scoring history through aesthetically pleasing and informative charts. Perfect for exploratory data analysis (EDA) tasks in sports analytics!