

Analysis Document for Shootout Data Visualization

Objective

The objective of the analysis is to identify and visualize the top 10 countries with the highest number of shootout wins based on the provided dataset.

Dataset Description

The dataset shootouts.csv contains information about shootout events, including the country that won each shootout. The column of interest in the dataset is winner, which records the name of the country that won each shootout.

Data Processing and Analysis Steps

1. Data Import

The dataset is imported using the pandas library:

```
import pandas as pd
```

```
data = pd.read_csv("Data(file_format_CSV)/shootouts.csv")
```

- The dataset is read into a pandas DataFrame named data for further processing.

2. Counting Shootout Wins

The winner column is extracted, and the number of occurrences of each country is counted using the Counter from the collections module:

```
from collections import Counter
```

```
shootout = data["winner"]
```

```
count_shootout = Counter(shootout)
```

- This results in a dictionary-like object count_shootout, where the keys are country names and the values are the corresponding counts of shootout wins.

3. Extracting Top 10 Countries

The top 10 countries with the highest number of shootout wins are identified:

```
Top_10_count_shootout = count_shootout.most_common(10)
```

- This generates a list of tuples, Top_10_count_shootout, where each tuple contains a country name and its count of shootout wins.

4. Preparing Data for Visualization

The data is restructured into a pandas DataFrame for visualization:

```
countries = [item[0] for item in Top_10_count_shootout]
```

```
no_of_goals = [item[1] for item in Top_10_count_shootout]
```

```
dft = {  
    'countries': countries,  
    'no_of_goals': no_of_goals  
}
```

```
df = pd.DataFrame(dft)
```

- Two lists, countries and no_of_goals, are created to store the country names and their corresponding counts of shootout wins.
- A dictionary dft is created and converted into a DataFrame df.

5. Data Visualization

A bar chart is created using Plotly Express:

```
import plotly.express as px
```

```
fig = px.bar(df, x='countries', y='no_of_goals', color='countries')
```

```
fig.show()
```

- The bar chart visualizes the number of shootout wins for each of the top 10 countries.
- Each country is represented by a bar, and the bars are colored differently for visual distinction.

Analysis and Insights

Key Findings

- The bar chart shows the distribution of shootout wins among the top 10 countries.
- Each bar represents a country's total number of shootout wins, with the height indicating the count.
- The color-coding of bars helps differentiate between the countries visually.

Observations

- The countries are ranked based on the number of shootout wins, with the country having the highest number at the top.
- There is a visible difference in the number of shootout wins among the countries, with some having significantly higher counts than others.

Implications

- The analysis provides insight into which countries have historically performed better in shootouts.

- This information can be useful for understanding patterns in shootout performance and may assist in making strategic decisions in future shootout scenarios.

Conclusion

The analysis successfully identified the top 10 countries with the highest number of shootouts wins and visualized the data in a clear and interpretable bar chart. The visualization highlights the distribution of shootout wins among these countries, providing valuable insights into their historical performance in shootouts.