Documentation for Code and Visualization

Objective

The purpose of this script is to visualize the countries that have hosted the most football matches from 1872 to 2024 using data from a CSV file. The visualization highlights the top 10 countries and their hosting counts with an interactive bar chart.

Code Breakdown

1. Importing Necessary Libraries

```
import plotly.express as px

import pandas as pd

from collections import Counter
```

plotly.express: Used for creating interactive visualizations (e.g., bar charts).

pandas: A data manipulation and analysis library to work with structured data.

collections.Counter: Helps count occurrences of elements (e.g., country names in this case).

2. Reading the Data

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

Reads the CSV file (results.csv) containing football match data.

Assumes that the CSV contains a column named country, which represents the host country of each football match.

3. Extracting and Counting Host Occurrences

```
data_country =  data['country']

count_noof_hostings_by_countries = Counter(data_country)
```

data['country']: Extracts the column country from the dataset.

Counter(data_country): Counts the occurrences of each country to determine the number of matches hosted.

4. Identifying the Top 10 Countries

top4_count_noof_hostings_by_countries = count_noof_hostings_by_countries.most_common(10)

most_common(10): Retrieves the 10 countries with the highest number of matches hosted.

5. Preparing the Data for Visualization

countries = [item[0] for item in top4_count_noof_hostings_by_countries]

no_of_hostings = [item[1] for item in top4_count_noof_hostings_by_countries]

dft = {

   'countries': countries,

   'no_of_hostings': no_of_hostings

}

df = pd.DataFrame(dft)

Extracts the country names (item[0]) and their corresponding hosting counts (item[1]) from the most_common result.

Stores this data in a dictionary (dft) and converts it into a DataFrame for visualization.

6. Creating the Bar Chart

```
fig = px.bar(

    df,

    x = 'countries',

    y = 'no_of_hostings',

    color = 'countries',

    title = "Countries Hosting The Most Number of Football Matches from 1872 to 2024"

)

fig.show()
```

px.bar: Generates an interactive bar chart:

   x='countries': The country names are displayed along the X-axis.

    y='no_of_hostings': The number of matches hosted is represented on the Y-axis.

    color='countries': Adds unique colors for each country to enhance visual clarity.

    title: Specifies the chart title.

  fig.show(): Displays the bar chart.

Output Visualization

Countries Hosting The Most Number of Football Mathces from 1872 to 2024
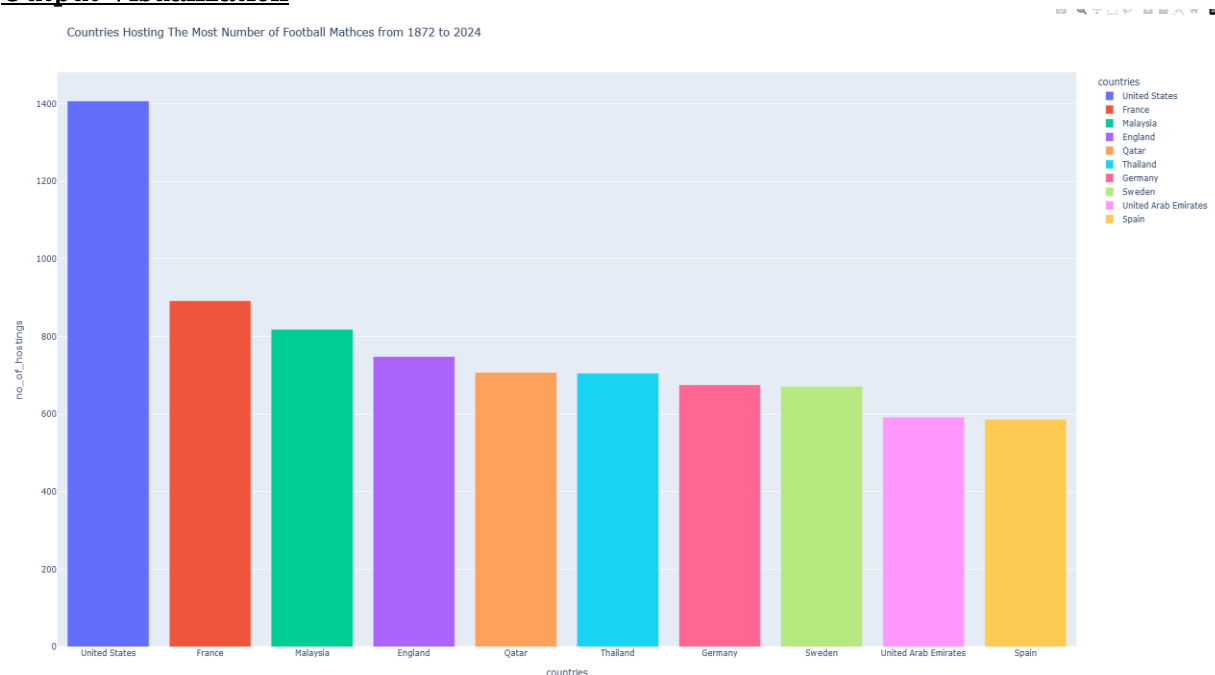
The bar chart generated by the code is interactive and contains the following features:

X-Axis: Displays the top 10 countries hosting the most football matches.

Y-Axis: Shows the count of matches hosted by each country.

Color Encoding: Each country is represented by a distinct color, making it easy to distinguish.

Interactive Features: Users can hover over bars to see precise counts, zoom in/out, and pan the chart.

Insights from the Visualization

The United States has hosted the highest number of football matches, significantly more than other countries.

France, Malaysia, and England follow as countries with substantial hosting counts.

The remaining countries in the top 10 include Qatar, Thailand, Germany, Sweden, the UAE, and Spain, each contributing to the global distribution of football match hosts.

Enhancements and Suggestions

Data Validation: Ensure the country column in the dataset is free from null or incorrect values.

Dynamic Top N Selection: Allow users to specify how many top countries they want to visualize (e.g., top 5, top 20).

Additional Insights: Overlay population or stadium capacity data to provide context for hosting capabilities.

Save Visualization: Export the chart as an image or HTML file for further sharing and analysis.