ANALYZING WORLD **CUP DATA**

OVERVIEW

This data analysis project aims to explore a dataset of international football matches from 1872 to 2022. The goal is to uncover insights about various aspects of international football, including:

- Identifying dominant teams and eras
- Analyzing trends in goals scored and home advantage
- Examining team performance based on playing location (home vs. neutral)
- Investigating the relationship between World Cup hosting and winning

GOALS

01
Understand historical trends in international football through data visualization and exploration.

02
Identify the most successful teams and countries based on goal scoring and winning records.

03
Analyze the impact of playing location (home vs. neutral) on team performance.

04
Investigate if hosting a World Cup provides an advantage for the host nation.

05
Explore the performance of teams from different continents in the World Cup.

DATASET
& SOURCE

The dataset consists of three CSV files:
- **results.csv:** Contains details of each match, including date, teams, scores, tournament, location, and neutral venue indicator.
- **shootouts.csv:** Details penalty shootout results (not used in this analysis).
- **winners.csv:** Lists World Cup winners by year and hosting country (with manual correction for 2018).

The data source is a compilation from various sources like Wikipedia, RSSSF.com, and individual football association websites.

DATA **FINDINGS**

01

The code explores the FIFA World Cup data within the **results.csv** file.

02

Functions are used to identify matches with the highest goal difference, most total goals, and top goal-scoring countries.

03

A choropleth map visualizes the total goals scored by each country.

04

Analysis of the **winners.csv** data reveals:
- Teams with the most World Cup wins.
- Success rates of South American and European teams based on hosting continent.

OUTPUTS

- Who is the greatest football team of all time?
- Which teams dominated different eras in football history?
- What trends have emerged in international football over the years, such as home advantage, total goals scored, and team strength distribution?
- Can we glean insights about geopolitics from football fixtures, such as changes in the number of participating countries or team rivalries?
- Which countries host the most matches without participating themselves?
- How does hosting a major tournament affect a country's chances in that tournament?
- Which teams frequently play friendlies and friendly tournaments, and does this benefit or hinder them?

OUTPUTS

- **Who is the greatest football team of all time?**
  - While the provided data and analysis offer valuable insights into the history of international football, definitively crowning the "greatest football team of all time" is subjective and often sparks passionate debates among fans. However, we can leverage the data to identify dominant teams in different eras and explore broader trends in the sport.

- **Which teams dominated different eras in football history?**
  - **1950s and 1960s:** Brazil emerged as a dominant force, winning the World Cup in 1958 and 1962. Their attacking flair and legendary players like Pelé defined an era.
  - **1970s and 1980s:** The Netherlands, with their "Total Football" philosophy, and West Germany, with their disciplined and efficient style, were among the top teams.
  - **1990s:** Germany, Italy, and Brazil continued their dominance, winning multiple World Cups and European Championships.
  - **2000s:** Brazil and France were consistently strong, winning World Cups and European Championships.
  - **2010s:** Spain, with their possession-based style, and Germany, with their tactical discipline, were dominant forces.

OUTPUTS

- **What trends have emerged in international football over the years, such as home advantage, total goals scored, and team strength distribution?**
  - **Home Advantage:** Historically, home teams have had a significant advantage, winning a higher proportion of matches. However, the advantage has decreased in recent years due to increased travel and more balanced playing fields.
  - **Total Goals Scored:** The overall trend in international football has been towards lower-scoring matches, particularly in recent decades. This can be attributed to defensive tactics, increased physicality, and more cautious approaches.
  - **Team Strength Distribution:** The gap between top-tier teams and weaker nations has narrowed over time. This is due to improved coaching, increased investment in youth development, and globalization of the sport.

- **Can we glean insights about geopolitics from football fixtures, such as changes in the number of participating countries or team rivalries?**
  - **Political Changes:** The dissolution of the Soviet Union and the breakup of Yugoslavia led to the emergence of new national teams and altered the geopolitical landscape of football.
  - **Team Rivalries:** Historical rivalries between countries, often rooted in political or cultural differences, have fueled intense matches and passionate fan bases.

OUTPUTS

- **Which countries host the most matches without participating themselves?**
  - While the data doesn't explicitly list countries hosting the most matches without participating, we can infer that major footballing nations like England, Germany, and France often host numerous international friendlies and tournaments.

- **How does hosting a major tournament affect a country's chances in that tournament?**
  - Hosting a major tournament can provide a significant boost to a country's footballing infrastructure and national pride. It can also lead to increased investment in youth development and improved facilities. However, the impact on a host nation's performance in the tournament itself is mixed. Some host nations have enjoyed success, while others have underperformed.

- **Which teams frequently play friendlies and friendly tournaments, and does this benefit or hinder them?**
  - Friendly matches serve various purposes, including building team cohesion, testing new tactics, and providing opportunities for younger players. While they can be beneficial, excessive friendly schedules can disrupt the rhythm of top teams and potentially lead to fatigue. ■

THE NUMBER OF MATCHES FOR EACH TOURNAMENT

tournament	num_matches
Friendly	17427
FIFA World Cup qualification	7774
UEFA Euro qualification	2593
African Cup of Nations qualification	1932
FIFA World Cup	900
...	...
AFF Championship qualification	2
TIFOCO Tournament	1
FIFA 75th Anniversary Cup	1
Copa Confraternidad	1
Real Madrid 75th Anniversary Cup	1

139 rows × 1 columns

The matches cover various events, including the FIFA World Cup, FIFI Wild Cup, and regular friendly matches. It strictly contains men's full international matches and excludes Olympic Games or matches involving B-teams, U-23 teams, or league select teams.

The Number Matches in Each World Cup

year	num_matches
2018	64
2014	64
2010	64
2006	64
2002	64
1998	64
1982	52
1994	52
1990	52
1986	52
1978	38
1974	38
1958	35
1970	32
1966	32
1962	32
1954	26
1950	22
1938	18
1930	18
1934	17

Identify Games with the Highest Goal Difference

	date	home_team	away_team	home_score	away_score	tournament	city	country	neutral	year	goal_difference
3667	1954-06-17	Hungary	South Korea	9	0	FIFA World Cup	Zürich	Switzerland	True	1954	9
9208	1974-06-18	Yugoslavia	DR Congo	9	0	FIFA World Cup	Gelsenkirchen	Germany	True	1974	9
12555	1982-06-15	Hungary	El Salvador	10	1	FIFA World Cup	Elche	Spain	True	1982	9

Game with the Highest Total Goals

	date	home_team	away_team	home_score	away_score	tournament	city	country	neutral	year	total_goals
3680	1954-06-26	Switzerland	Austria	5	7	FIFA World Cup	Lausanne	Switzerland	False	1954	12

Identify the Country with
the Most Goals

team	total_goals
Brazil	229
Germany	226
Argentina	137
Italy	128
France	120
...	...
Canada	0
Trinidad and Tobago	0
China PR	0
Indonesia	0
DR Congo	0
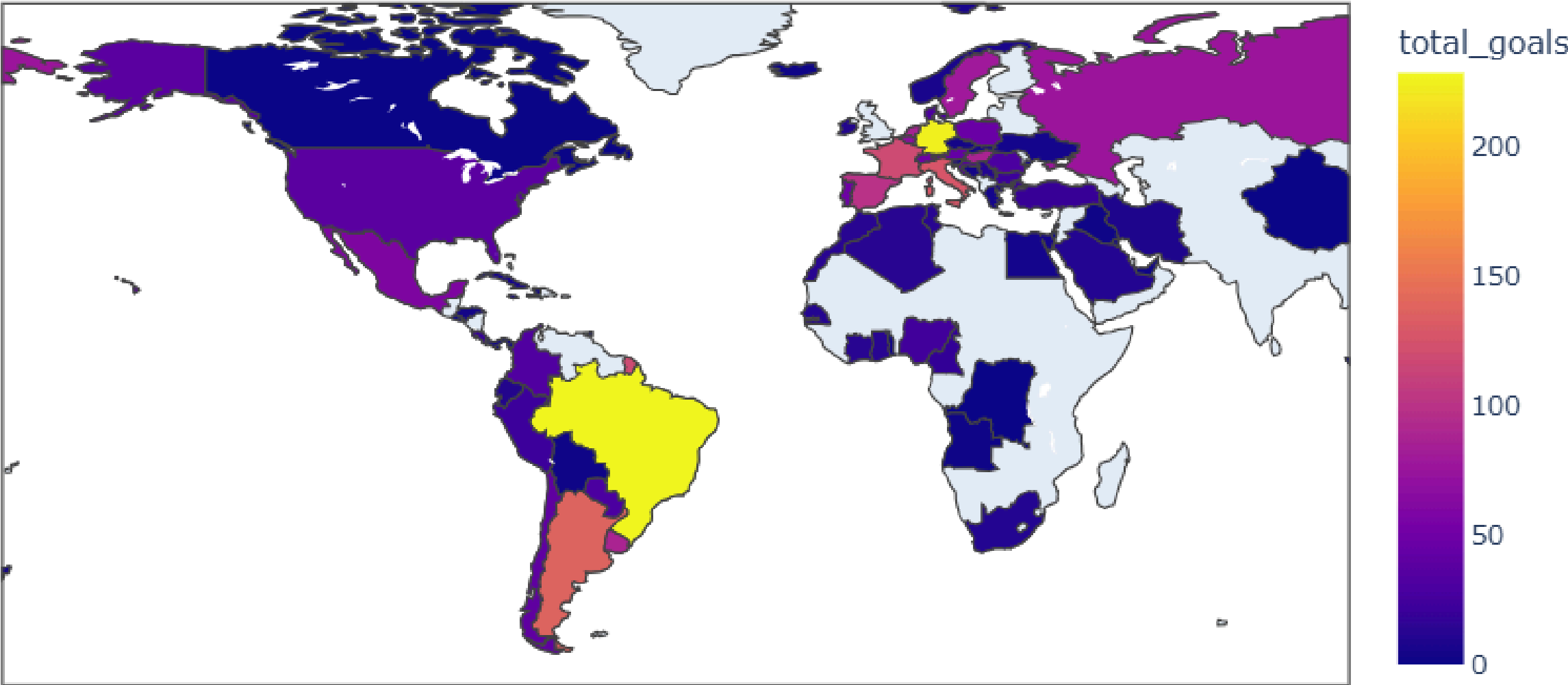
Does Playing Close to Home Matter?

	year	hosting_country	winning_country
0	1930	Uruguay	Uruguay
1	1934	Italy	Italy
2	1938	France	Italy
3	1950	Brazil	Uruguay
4	1954	Switzerland	Germany
5	1958	Sweden	Brazil
6	1962	Chile	Brazil
7	1966	England	England
8	1970	Mexico	Brazil
9	1974	Germany	Germany
10	1978	Argentina	Argentina
11	1982	Spain	Italy
12	1986	Mexico	Argentina
13	1990	Italy	Germany
14	1994	USA	Brazil
15	1998	France	France
16	2002	Korea/Japan	Brazil
17	2006	Germany	Italy
18	2010	South Africa	Spain
19	2014	Brazil	Germany
0	2018	Russia	France

Who Has the Most Wins?

```
winning_country
England        1
Spain          1
Argentina      2
France         2
Uruguay        2
Germany        4
Italy          4
Brazil         5
dtype: int64
```

	year	hosting_country	winning_country	winning_continent	hosting_continent
0	1930	Uruguay	Uruguay	South America	South America
1	1934	Italy	Italy	Europe	Europe
2	1938	France	Italy	Europe	Europe
3	1950	Brazil	Uruguay	South America	South America
4	1954	Switzerland	Germany	Europe	Europe
5	1958	Sweden	Brazil	South America	Europe
6	1962	Chile	Brazil	South America	South America
7	1966	England	England	Europe	Europe
8	1970	Mexico	Brazil	South America	North America
9	1974	Germany	Germany	Europe	Europe
10	1978	Argentina	Argentina	South America	South America
11	1982	Spain	Italy	Europe	Europe
12	1986	Mexico	Argentina	South America	North America
13	1990	Italy	Germany	Europe	Europe
14	1994	USA	Brazil	South America	North America
15	1998	France	France	Europe	Europe
16	2002	Korea/Japan	Brazil	South America	Asia
17	2006	Germany	Italy	Europe	Europe
18	2010	South Africa	Spain	Europe	Africa
19	2014	Brazil	Germany	Europe	South America
20	2018	Russia	France	Europe	Europe

Analyze South American wins

- South American teams won 9 world cups in total
- 4 of these were won in South America
- 5 of these were won elsewhere
- 4/5 of South America hosted world cups were won by South American teams.

Analyze European wins

- European teams won 12 world cups in total
- 10 of these were won in Europe
- 2 of these were won elsewhere
- 10/11 of Europe hosted world cups were won by European teams.

INSIGHTS

01.

Football fans
Understanding historical dominance of teams and eras in international football.

02.

Analysts
Identifying trends in goal scoring, home advantage, and team performance based on location.

03.

National teams
Evaluating the potential benefits and drawbacks of hosting major tournaments.