



TAN TAO UNIVERSITY
FROM KNOWLEDGE TO THE STARS

Course: Data Visualization

FINAL REPORT

Football Match Analysis Report

Project overview

This report examines international men's and women's football match results from 1872 to the present day. The project involved collecting, cleaning, and analyzing match data to understand team performance metrics across different continents and tournaments.

Data Collection and Preparation

Data Sources

Primary match data was obtained from a CSV file containing international men's football results.

	Unnamed: 0	date	home_team	away_team	home_score	away_score	tournament
0	0	1872-11-30	Scotland	England	0	0	Friendly
1	1	1873-03-08	England	Scotland	4	2	Friendly
2	2	1874-03-07	Scotland	England	2	1	Friendly
3	3	1875-03-06	England	Scotland	2	2	Friendly
4	4	1876-03-04	Scotland	England	3	0	Friendly

Men's football match results

	Unnamed: 0	date	home_team	away_team	home_score	away_score	tournament
0	0	1969-11-01	Italy	France	1	0	Euro
1	1	1969-11-01	Denmark	England	4	3	Euro
2	2	1969-11-02	England	France	2	0	Euro
3	3	1969-11-02	Italy	Denmark	3	1	Euro
4	4	1975-08-25	Thailand	Australia	3	2	AFC Championship

Women's football match results

Continent mapping data was sourced from a JSON file that associates countries with their respective continents

Data Cleaning Steps:

Do same things for both 2 datasets for men and women

- Removed unnecessary columns: Dropped the "Unnamed: 0" column which served no analytical purpose
- Added continent information: Mapped both home and away teams to their continents using the continent mapping file
- Goal difference (absolute difference between home and away scores)
- Match winner (home team, away team, or draw)
- Extracted year from date for temporal analysis

Result after cleaning steps

	date	home_team	away_team	home_score	away_score	tournament	home_continent	away_continent	goal_difference	winner	year
0	1872-11-30	Scotland	England	0	0	Friendly	Europe	Europe	0	Draw	1872
1	1873-03-08	England	Scotland	4	2	Friendly	Europe	Europe	2	England	1873
2	1874-03-07	Scotland	England	2	1	Friendly	Europe	Europe	1	Scotland	1874
3	1875-03-06	England	Scotland	2	2	Friendly	Europe	Europe	0	Draw	1875
4	1876-03-04	Scotland	England	3	0	Friendly	Europe	Europe	3	Scotland	1876

Men's football match results

	date	home_team	away_team	home_score	away_score	tournament	home_continent	away_continent	goal_difference	winner	year
0	1969-11-01	Italy	France	1	0	Euro	Europe	Europe	1	Italy	1969
1	1969-11-01	Denmark	England	4	3	Euro	Europe	Europe	1	Denmark	1969
2	1969-11-02	England	France	2	0	Euro	Europe	Europe	2	England	1969
3	1969-11-02	Italy	Denmark	3	1	Euro	Europe	Europe	2	Italy	1969
4	1975-08-25	Thailand	Australia	3	2	AFC Championship	Asia	Oceania	1	Thailand	1975

Women's football match results

Afterward, I calculate total matches played both home and away as well as their percentage and saved as a dataframe

	country	home_win_percent	away_win_percent	home matches	away_matches	total matches
0	Scotland	56.57	38.76	396	418	814
1	England	62.10	51.72	525	524	1049
2	Wales	38.42	25.64	341	351	692
3	Northern Ireland	33.23	18.39	325	348	673
4	United States	54.71	27.21	446	283	729

Men's football match results

	country	home_win_percent	away_win_percent	home matches	away_matches	total matches
0	Italy	62.39	48.72	109	117	226
1	France	64.76	48.60	105	107	212
2	Denmark	59.17	43.71	120	151	271
3	England	67.59	49.50	108	101	209
4	Thailand	70.59	34.00	51	50	101

Women's football match results

Next, I calculate total matches played for each country as well as total matches wins, losses and draws

	country	tournament	total_matches	wins	losses	draws
0	Germany	Friendly	577	307	148	122
1	Hungary	Friendly	571	267	172	132
2	Sweden	Friendly	549	249	175	125
3	Poland	Friendly	533	217	172	144
4	France	Friendly	490	218	173	99

Men's football match results

	country	tournament	total_matches	wins	losses	draws
0	United States	Friendly	111	90	4	17
1	Denmark	Algarve Cup	106	42	50	14
2	Portugal	Algarve Cup	106	26	63	17
3	Italy	UEFA Euro qualification	104	71	14	19
4	Norway	Algarve Cup	104	53	37	14

Women's football match results

I spent a lot of time on this, so after cleaning the data, I saved them as DataFrames—not only for the entire world, but also split by individual continents.

```
# summarize
```

```
world = calculator.country_match_summary(df)
asia = calculator.country_match_summary(asia_df)
africa = calculator.country_match_summary(africa_df)
europe = calculator.country_match_summary(europe_df)
north_america = calculator.country_match_summary(north_america_df)
south_america = calculator.country_match_summary(south_america_df)
oceania = calculator.country_match_summary(oceania_df)
```

```
# save
```

```
df.to_csv('men_results.csv', index=False)

win_percent.to_csv('men_win_percent.csv', index=False)

world.to_csv('men_world.csv', index=False)
asia.to_csv('men_asia.csv', index=False)
africa.to_csv('men_africa.csv', index=False)
europe.to_csv('men_europe.csv', index=False)
north_america.to_csv('men_north_america.csv', index=False)
south_america.to_csv('men_south_america.csv', index=False)
oceania.to_csv('men_oceania.csv', index=False)
```

Analysis and Visualization

General

	date	home_team	away_team	home_score	away_score	tournament	home_continent	away_continent	goal_difference	winner	year
0	1872-11-30	Scotland	England	0	0	Friendly	Europe	Europe	0	Draw	1872
1	1873-03-08	England	Scotland	4	2	Friendly	Europe	Europe	2	England	1873
2	1874-03-07	Scotland	England	2	1	Friendly	Europe	Europe	1	Scotland	1874
3	1875-03-06	England	Scotland	2	2	Friendly	Europe	Europe	0	Draw	1875
4	1876-03-04	Scotland	England	3	0	Friendly	Europe	Europe	3	Scotland	1876

Men's football match results

	date	home_team	away_team	home_score	away_score	tournament	home_continent	away_continent	goal_difference	winner	year
0	1969-11-01	Italy	France	1	0	Euro	Europe	Europe	1	Italy	1969
1	1969-11-01	Denmark	England	4	3	Euro	Europe	Europe	1	Denmark	1969
2	1969-11-02	England	France	2	0	Euro	Europe	Europe	2	England	1969
3	1969-11-02	Italy	Denmark	3	1	Euro	Europe	Europe	2	Italy	1969
4	1975-08-25	Thailand	Australia	3	2	AFC Championship	Asia	Oceania	1	Thailand	1975

Women's football match results

It loads match data from a CSV file hosted at

http://115.78.93.252/TTU/pool/donghuynh0/data_visualization_2025/final_proeject/ready/

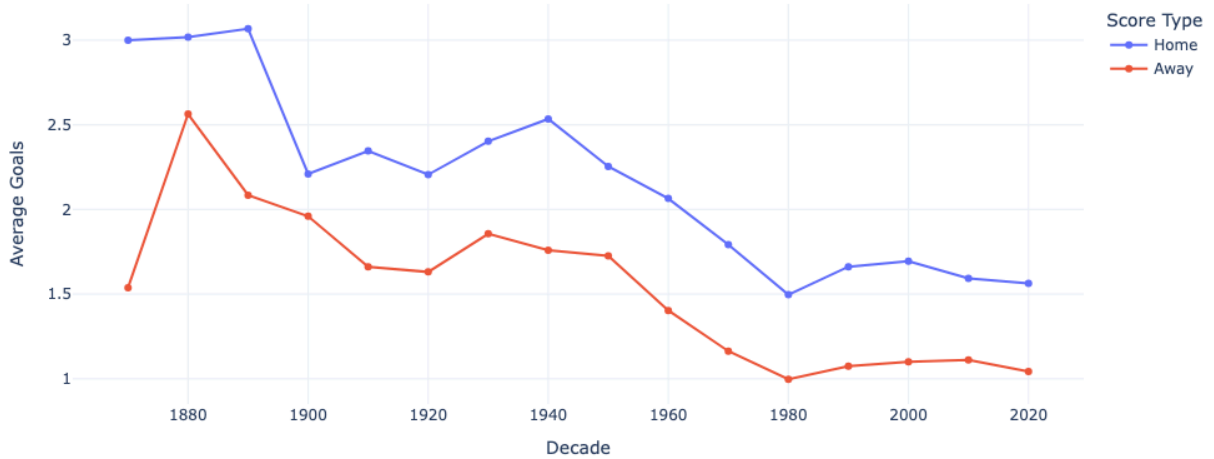
The DataFrame contains **44,353** rows (matches) with **11** columns:

1. **date**: Match date
2. **home_team**: Home team name
3. **away_team**: Away team name
4. **home_score**: Home team goals
5. **away_score**: Away team goals
6. **tournament**: Type of match (Friendly, FIFA World Cup, etc.)
7. **home_continent**: Continent of home team
8. **away_continent**: Continent of away team
9. **goal_difference**: Difference in goals (home - away)
10. **winner**: Match result (Home team, Away team, or Draw)
11. **year**: Year extracted from date

The data spans 150 years of international football history, from the first recorded international match (Scotland vs England, 1872) to recent World Cup matches (2022).

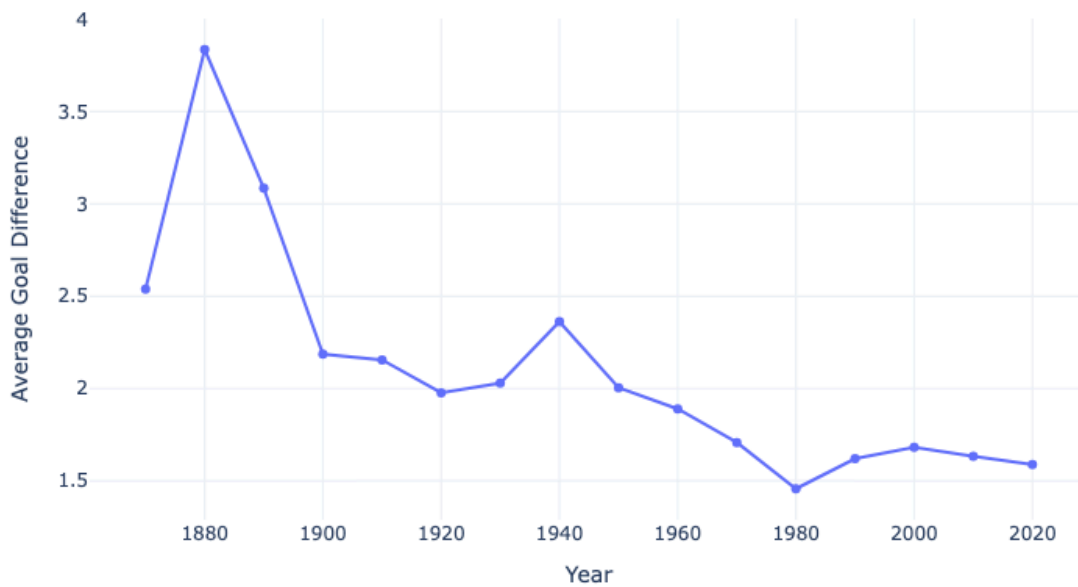
Men

Average Goals per Match by Decade



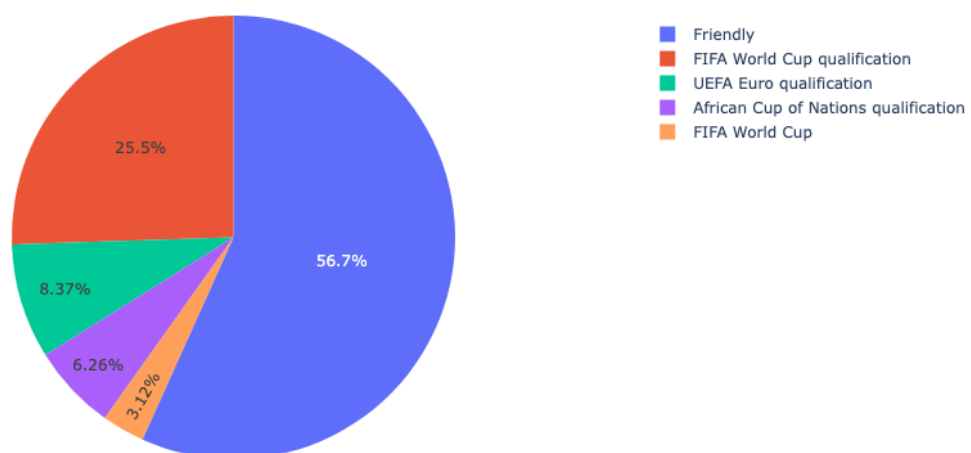
Average goals scored by both home and away teams have generally declined. Home teams consistently score more than away teams, but the gap has narrowed, especially after 1960. This suggests a long-term decrease in goal-scoring and a gradual reduction in home-field advantage over time.

Average Goal Difference by Decade



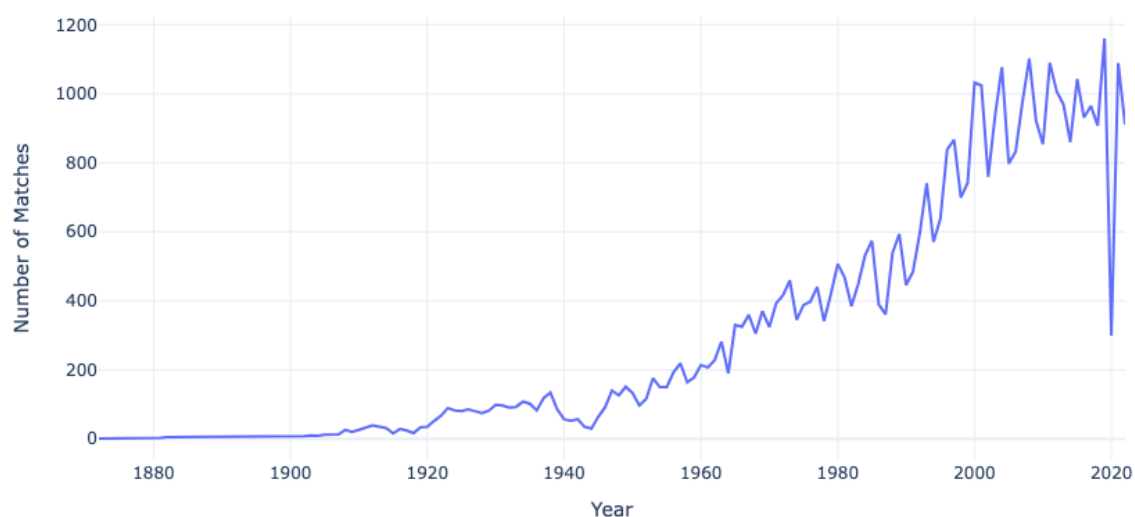
The average goal difference per match has steadily declined from the 1880s to the 2020s. In the early decades, matches tended to have larger goal margins—particularly peaking in the 1880s - indicating more uneven contests. Over time, the decreasing goal difference suggests that matches have become more competitive and balanced, especially since the 1960s.

Match Distribution Among Top 5 Tournaments



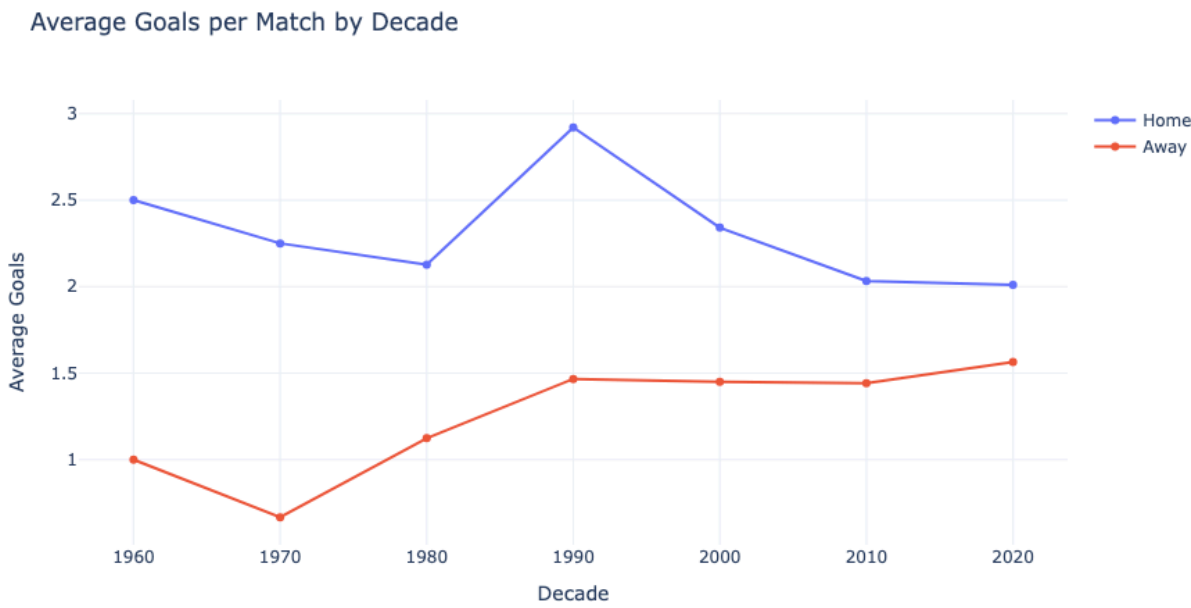
The chart shows that most international matches are friendlies, making up **56.7%** of games. Competitive tournaments like World Cup qualifiers and continental qualifiers follow, while the actual FIFA World Cup accounts for a small portion.

Total Football Matches Played per Year (1872-2022)

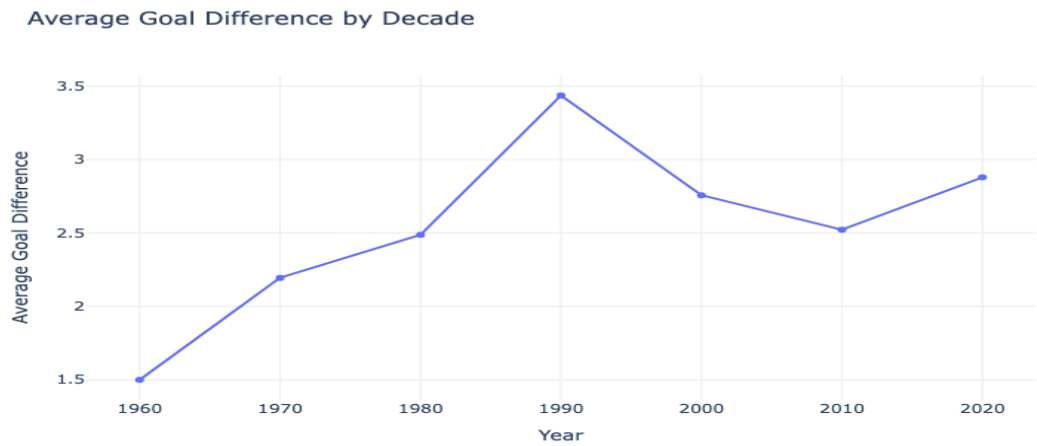


The line chart shows a steady increase in the number of football matches played annually from **1872 to 2022**, with significant growth starting around the **1950s**. The number of matches peaked around the **2010s**, reflecting the global expansion of international football. A noticeable dip around **2020** likely corresponds to the **COVID-19 pandemic**, which disrupted global sporting events. Overall, the data indicates a strong upward trend in match frequency over the last century.

Women

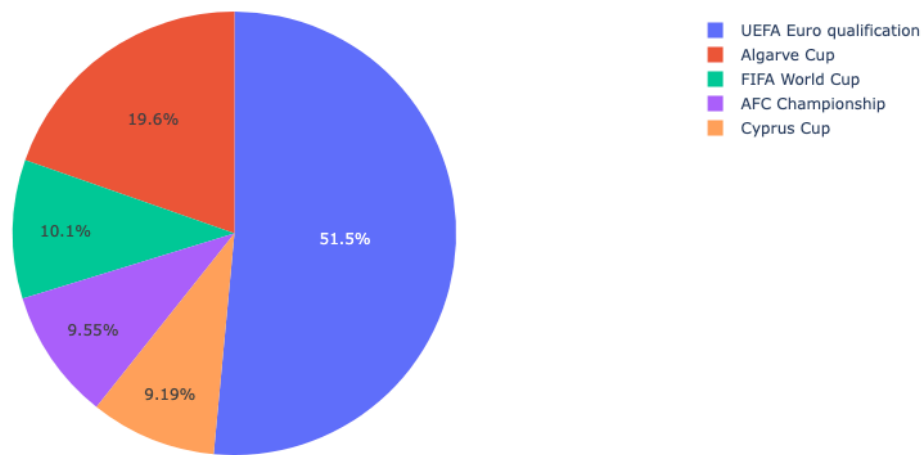


Over the decades, the average goals scored by home teams have declined slightly, while away team scores have steadily increased. This indicates a reduction in home advantage and suggests that football matches have become more balanced and competitive over time.



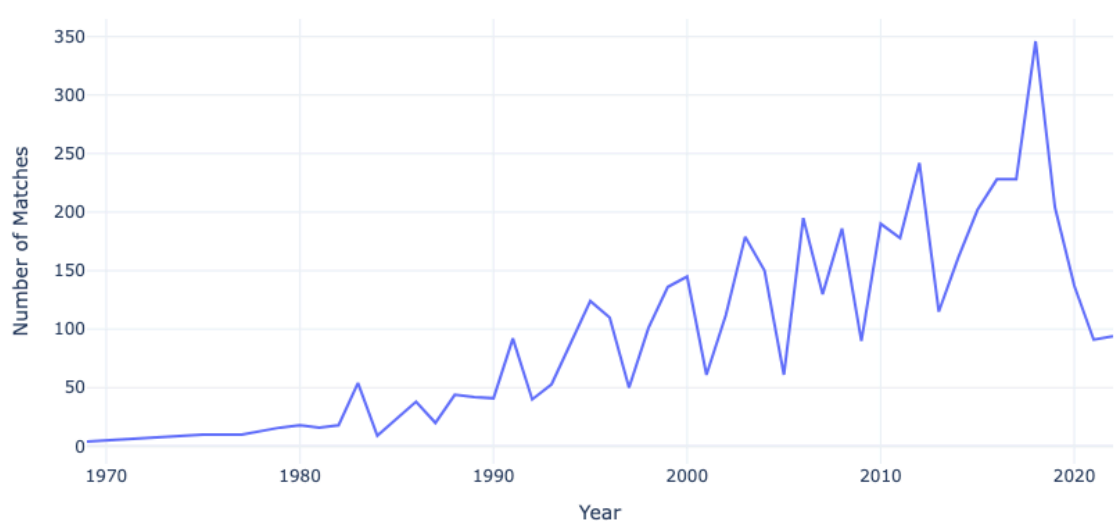
From 1960 to 2020, the average goal difference per match showed an overall increase, peaking sharply in the 1990s. This indicates a period where matches were less evenly contested. Although the difference decreased in the 2000s and 2010s, it slightly rose again in the 2020s, suggesting recent fluctuations in match competitiveness.

Match Distribution Among Top 5 Tournaments



The chart shows that regional qualification tournaments, especially in Europe, host a significantly larger number of matches compared to global or invitational tournaments.

Total Football Matches Played per Year (1872-2022)



The line chart shows a steady increase in the number of football matches played annually from **1872 to 2022**. A noticeable dip around **2020** likely corresponds to the **COVID-19 pandemic**, which disrupted global sporting events. Overall, the data indicates a strong upward trend in match frequency over the last

Detail

Dataframes are used to analyze:

	country	tournament	total_matches	wins	losses	draws
0	Germany	Friendly	577	307	148	122
1	Hungary	Friendly	571	267	172	132
2	Sweden	Friendly	549	249	175	125
3	Poland	Friendly	533	217	172	144
4	France	Friendly	490	218	173	99

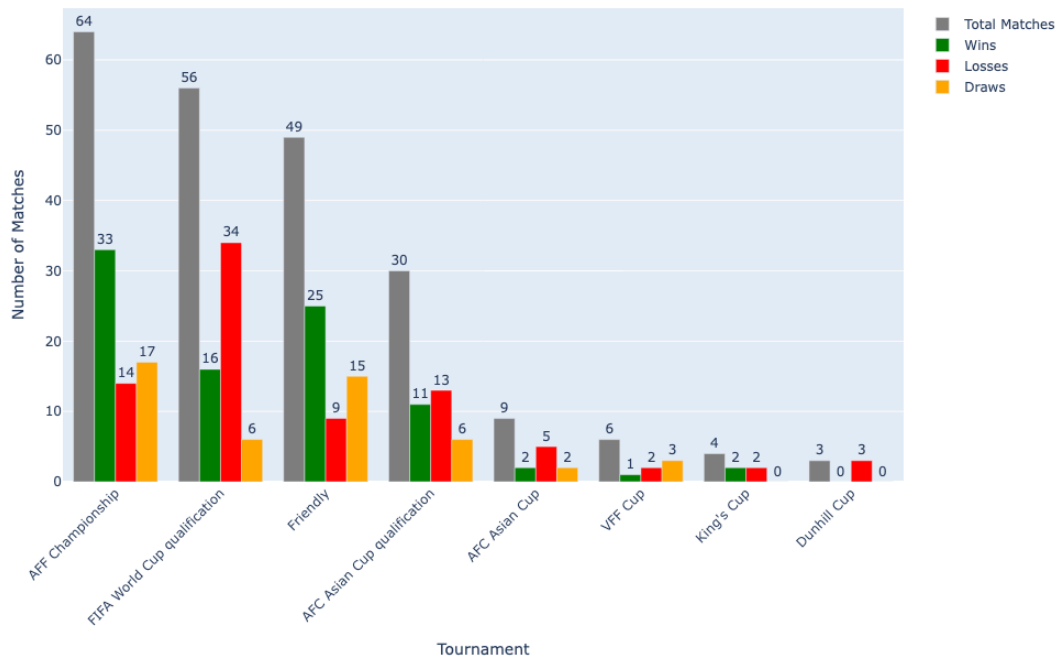
	country	home_win_percent	away_win_percent	home matches	away_matches	total matches
0	Sweden	58.50	40.59	506	547	1053
1	England	62.10	51.72	525	524	1049
2	Brazil	71.72	53.40	594	427	1021
3	Argentina	65.96	39.06	570	448	1018
4	Germany	62.17	53.70	526	460	986

Both men's and women's football matches are analyzed in the same way and include:

A country's match statistics across different tournaments

The chart shows a country's performance across different tournaments through total matches played, wins, losses, and draws. Based on this chart, I can explore a country's football strength both within and outside its region.

Vietnam Match Stats by Tournament



An example of Vietnam

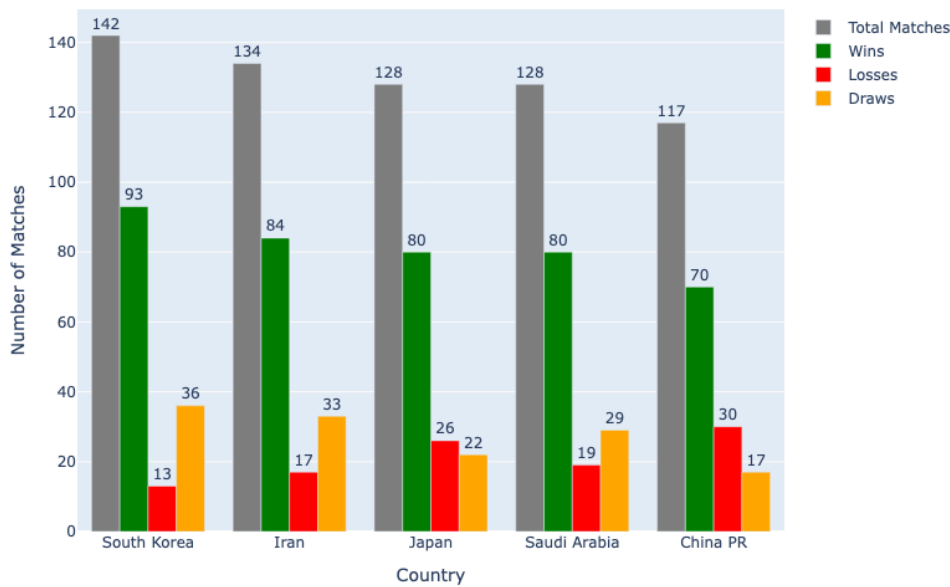
The chart shows that Vietnam has participated most frequently in the AFF Championship, FIFA World Cup qualification, and Friendly matches. The AFF Championship is where Vietnam has performed best, with 33 wins out of 64 matches, indicating strong regional dominance. In contrast, performance in FIFA World Cup qualification and AFC Asian Cup qualification has been more challenging, with a higher number of losses than wins. Vietnam's draw rate is notably high in the AFF Championship. Overall, Vietnam performs best in regional tournaments but faces tougher competition on the broader Asian and world stages.

Top countries in different continents based on tournament qualifications

I can explore the top-performing countries in a specific tournament by analyzing the total matches played, as well as the number of wins, losses, and draws.

From this, I can conclude which country has been the most dominant in that tournament.

Top 5 Countries Stats – FIFA World Cup qualification



Top Asian Teams in World Cup Qualifiers

South Korea leads in both total matches (**142**) and wins (**93**), demonstrating strong and consistent performance in World Cup qualifiers. Iran follows closely with **84** wins out of **134** matches, while Japan and Saudi Arabia each show identical win counts (**80**), despite Saudi Arabia having slightly fewer losses. China PR ranks lowest among the top five in wins (**70**) and has the highest number of losses (**30**), suggesting room for improvement. Overall, all five countries show strong regional competitiveness, with South Korea standing out as the most dominant.

A country's home vs away performance

I wonder if playing at home versus away impacts performance. I compared the performance of teams when playing at home and away.

Japan Match Outcomes



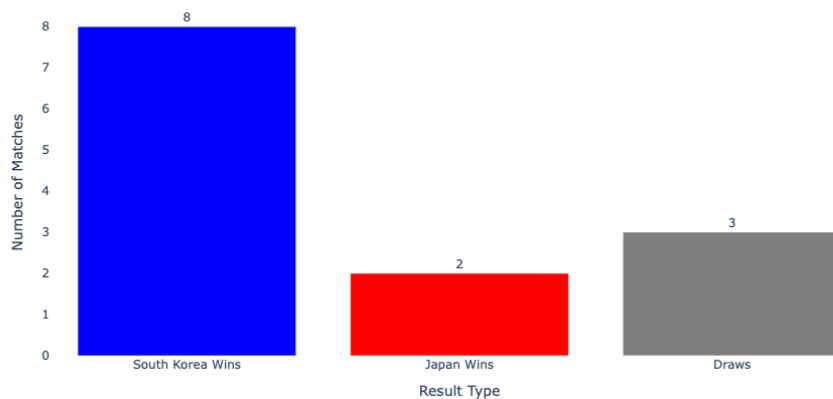
An example of Japan

Japan performs better in home matches with a higher win rate (**52.1%**) compared to away matches (**42.2%**).

Head-to-head comparison based on tournament qualifications

I compared the performance of 2 countries via various tournaments

South Korea vs Japan in FIFA World Cup qualification
Total Matches: **13**



In 13 encounters, South Korea won 8 out of 13 matches. I can conclude that in the FIFA World Cup qualifiers, South Korea has been stronger than Japan.