International Football Matches Dataset

1. INTRODUCTION

The dataset used to do this analysis contains results of 44232 men’s international football matches played since 1900 till end of 2022. The majority of the matches are international friendlies, and the other majority is formed by matches from high-profile football tournaments in the world (i.e., World Cup, Copa America, African Cup of Nations etc) and the qualification matches for these tournaments.

The purpose of the analysis is to attempt to generate insights on following matters:

* 1. Countries that have dominated world football
     + The best teams of all time
     + The teams that dominated different eras of football
  2. Countries that are likely to continue to be successful in immediate future
  3. Trends in international football
     + Home advantage
     + Goals scored and conceded
     + Unbeaten runs
  4. Rivalries in international football
     + Pairs that often play together
     + Face-to-face results
     + Rivalry over the years

1. DERIVED TABLES AND IMPORTANT DATA POINTS

The following table has been derived from the ‘results’ table that contains the original set of results:

* **teamlvlresults\_**: this table contains a dedicated row for every match played by a team, capturing opposition team name, score of the match, location, result of the match etc. It also assigns a unique ID to each entry of the table. This table will help in understanding performance of each team individually.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **match\_id** | **uniq\_match\_id** | **date** | **team** | **opposition** | **team\_score** | **opposition\_score** | **tournament** | **city** | **country** | **neutral** | **reg\_result** | **Era** |
| Aus28197waniaTa | iaAus28197waniaTaAu7 | 13-03-1977 | Australia | Taiwan | 3 | 0 | FIFA World Cup qualification | Ba | Fiji | TRUE | WIN | 4Modern Era |

Query:

CREATE OR REPLACE TABLE

  footballApr23Self.teamlvlresults\_ AS (

  SELECT

    DISTINCT match\_id,

    CONCAT(RIGHT(home\_team, 2), match\_id, LEFT(home\_team,2), home\_score+4) AS uniq\_match\_id,

    date,

    home\_team AS team,

    away\_team AS opposition,

    home\_score AS team\_score,

    away\_score AS opposition\_score,

    tournament,

    city,

    country,

    neutral,

    CASE

      WHEN regular\_result = home\_team THEN 'WIN'

      WHEN regular\_result = away\_team THEN 'WIN'

    ELSE 'TIE'

  END AS reg\_result,

    CASE

      WHEN EXTRACT(year FROM date) BETWEEN 1900 AND 1924 THEN "Vintage Era"

      WHEN EXTRACT(year FROM date) BETWEEN 1925 AND 1949 THEN "Old Era"

      WHEN EXTRACT(year FROM date) BETWEEN 1950 AND 1974 THEN "Classic Era"

      WHEN EXTRACT(year FROM date) BETWEEN 1975 AND 1999 THEN "Modern Era"

    ELSE "Postmodern Era"

  END AS Era

  FROM `footballApr23Self.results`

  UNION DISTINCT

  SELECT

    DISTINCT match\_id,

    CONCAT(RIGHT(away\_team, 2), match\_id, LEFT(away\_team,2), away\_score) AS uniq\_match\_id,

    date,

    away\_team AS team,

    home\_team AS opposition,

    away\_score AS team\_score,

    home\_score AS opposition\_score,

    tournament,

    city,

    country,

    neutral,

    CASE

      WHEN regular\_result = away\_team THEN 'WIN'

      WHEN regular\_result = home\_team THEN 'WIN'

    ELSE 'TIE'

  END AS reg\_result,

    CASE

      WHEN EXTRACT(year FROM date) BETWEEN 1900 AND 1924 THEN "Vintage Era"

      WHEN EXTRACT(year FROM date) BETWEEN 1925 AND 1949 THEN "Old Era"

      WHEN EXTRACT(year FROM date) BETWEEN 1950 AND 1974 THEN "Classic Era"

      WHEN EXTRACT(year FROM date) BETWEEN 1975 AND 1999 THEN "Modern Era"

    ELSE "Postmodern Era"

  END AS Era

  FROM `footballApr23Self.results`)

* **teamstats\_:** this table aggregates the important statistics pertaining to every team, making it easier to derive insights on different parameters. The different columns indicate number of matches played, goals scored, goals conceded, different outcomes, etc. corresponding to every team.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **team** | **total\_matches** | **total\_opponents** | **total\_scored** | **total\_conceded** | **reg\_result** | **count\_result** |
| Sweden | 1053 | 98 | 2062 | 1347 | WIN | 518 |
| Sweden | 1053 | 98 | 2062 | 1347 | TIE | 227 |
| Sweden | 1053 | 98 | 2062 | 1347 | LOSS | 308 |

Q

Query

CREATE OR REPLACE TABLE

  footballApr23Self.teamstats\_ AS

SELECT

  DISTINCT team,

  COUNT(uniq\_row\_id) OVER(PARTITION BY team) AS total\_matches,

  COUNT(DISTINCT opposition) OVER(PARTITION BY team) AS total\_opponents,

  SUM(team\_score) OVER(PARTITION BY team) AS total\_scored,

  SUM(opposition\_score) OVER(PARTITION BY team) AS total\_conceded,

  reg\_result,

  COUNT(uniq\_row\_id) OVER(PARTITION BY team, reg\_result) AS count\_result

FROM

  `footballApr23Self.teamlvlresults\_`

ORDER BY

  2 DESC,

  4 DESC,

  3 DESC,

  5 DESC,

  6 DESC,

  1

Some important observations:

* There is total 316 countries that have participated in international football in the duration of 1990 and 2022. If we divide the entire timespan roughly into 25-year partitions (eras), the number of teams and matches played between them is distributed as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| **era** | **years** | **number\_of\_teams** | **number\_of\_matches** |
| 1Vintage Era | 1900 to 1924 | 57 | 733 |
| 2Old Era | 1924 to 1949 | 113 | 2217 |
| 3Classic Era | 1950 to 1974 | 186 | 6323 |
| 4Modern Era | 1975 to 1999 | 247 | 13407 |
| 5Postmodern Era | 2000 onward | 302 | 21552 |

Query

SELECT

  era,

  CASE

    WHEN era = "1Vintage Era" THEN "1900 to 1924"

    WHEN era = "2Old Era" THEN "1924 to 1949"

    WHEN era = "3Classic Era" THEN "1950 to 1974"

    WHEN era = "4Modern Era" THEN "1975 to 1999"

  ELSE "2000 onward"

END AS years,

  COUNT(DISTINCT team) AS number\_of\_teams,

  COUNT(DISTINCT match\_id) AS number\_of\_matches

FROM

  `footballApr23Self.teamlvlresults\_`

GROUP BY

  1

ORDER BY

  1

* Out of 44232 matches, 17493 (about 40%) were friendly matches. Other significant percentages of matches are as follows.

|  |  |  |
| --- | --- | --- |
| **tournament** | **matches** | **percentage\_matches** |
| Friendly | 17493 | 40 |
| FIFA World Cup qualification | 7878 | 18 |
| UEFA Euro qualification | 2585 | 6 |
| African Cup of Nations qualification | 1932 | 4 |
| FIFA World Cup | 964 | 2 |
| Copa AmÃ©rica | 841 | 2 |
| AFC Asian Cup qualification | 764 | 2 |
| African Cup of Nations | 741 | 2 |

Query

SELECT

  DISTINCT tournament,

  COUNT(match\_id) OVER(PARTITION BY tournament) AS matches,

  ROUND(COUNT(match\_id) OVER(PARTITION BY tournament)\*100/44232) AS percentage\_matches

FROM

  `footballApr23Self.results`

ORDER BY

  2 DESC

* The matches have been hosted in about 264 different locations all over the world. The top locations are shown below.

|  |  |  |
| --- | --- | --- |
| **country** | **matches** | **percentage\_matches** |
| United States | 1265 | 2.86 |
| France | 830 | 1.88 |
| Malaysia | 755 | 1.71 |
| England | 686 | 1.55 |
| Sweden | 660 | 1.49 |
| Qatar | 633 | 1.43 |
| Germany | 608 | 1.37 |
| Brazil | 569 | 1.29 |
| Spain | 568 | 1.28 |
| United Arab Emirates | 557 | 1.26 |
| South Africa | 533 | 1.21 |
| Austria | 518 | 1.17 |
| Italy | 508 | 1.15 |
| Thailand | 501 | 1.13 |
| Switzerland | 497 | 1.12 |
| South Korea | 479 | 1.08 |
| Hungary | 462 | 1.04 |
| Argentina | 458 | 1.04 |

Query

SELECT

  DISTINCT country,

  COUNT(match\_id) OVER(PARTITION BY country) AS matches,

  ROUND(COUNT(match\_id) OVER(PARTITION BY country)\*100/44232, 2) AS percentage\_matches

FROM

  `footballApr23Self.results`

ORDER BY

  2 DESC

1. COUNTRIES THAT HAVE DOMINATED WORLD FOOTBALL

To analyse performance over the years, below parameters are calculated for each team. Each of the parameters are weighted suitably to arrive at the top teams that have displayed superlative performance over the time span under consideration. Certain assumptions are made for each parameter to ensure that outliers are eliminated.

The overall performance score is a combination of all of the defined parameters.

Using overall performance score, we find out the top 3 teams in each of the Eras and also for the entire period, and determine the best team of all time subsequently.

Parameters:

* **Win ratio** = matches won / total matches played

*[assumption: teams which have played more than 100 international football matches are considered]*

* **Loss ratio** = matches lost/ total matches played

*[assumption: teams which have played more than 100 international football matches are considered]*

* **Goals scored per match** = goals scored / total matches played
* **Goals conceded per match** = goals scored / total matches played
* **Goal ratio** = total goals scored / total goals conceded
* **Overall performance score = Win ratio \* 10 – Loss ratio \* 10 + Goals scored per match – Goals conceded per match + Goal ratio**

Results:

All time performance (1900 to 2022)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **team** | **total\_matches** | **total\_opponents** | **win\_ratio** | **loss\_ratio** | **goals\_scored\_per\_match** | **goals\_conceded\_per\_match** | **goals\_ratio** | **overall\_performance\_score** |
| Brazil | 1023 | 90 | 0.6 | 0.2 | 2.2 | 0.9 | 2.5 | 8.6 |
| Spain | 733 | 91 | 0.6 | 0.2 | 2 | 0.9 | 2.3 | 7.3 |
| England | 982 | 87 | 0.6 | 0.2 | 2.1 | 0.9 | 2.2 | 7.1 |
| Iran | 525 | 121 | 0.6 | 0.2 | 1.9 | 0.8 | 2.3 | 7 |
| Germany | 986 | 88 | 0.6 | 0.2 | 2.2 | 1.2 | 1.9 | 6.8 |

Vintage-era performance (1900 to 1924)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **team** | **total\_matches** | **total\_opponents** | **win\_ratio** | **loss\_ratio** | **goals\_scored\_per\_match** | **goals\_conceded\_per\_match** | **goals\_ratio** | **overall\_performance\_score** |
| England | 76 | 9 | 0.6 | 0.2 | 2.5 | 1 | 2.4 | 8.2 |
| Denmark | 43 | 7 | 0.6 | 0.3 | 2.7 | 1.2 | 2.2 | 7.2 |
| Scotland | 61 | 4 | 0.5 | 0.2 | 1.9 | 0.9 | 2.1 | 6.2 |
| Netherlands | 51 | 10 | 0.5 | 0.2 | 2.7 | 1.8 | 1.5 | 5.4 |
| Hungary | 98 | 12 | 0.5 | 0.3 | 2.6 | 1.7 | 1.6 | 5.3 |

Old-era performance (1925 to 1949)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| team | total\_matches | total\_opponents | win\_ratio | loss\_ratio | goals\_scored\_per\_match | goals\_conceded\_per\_match | goals\_ratio | overall\_performance\_score |
| Argentina | 126 | 14 | 0.7 | 0.1 | 2.7 | 1.1 | 2.4 | 9.4 |
| Italy | 114 | 22 | 0.6 | 0.2 | 2.4 | 1.4 | 1.7 | 7.3 |
| England | 117 | 22 | 0.6 | 0.3 | 2.9 | 1.4 | 2.1 | 7 |
| Germany | 142 | 26 | 0.6 | 0.2 | 2.9 | 1.6 | 1.9 | 6.8 |
| Scotland | 83 | 15 | 0.6 | 0.3 | 2.2 | 1.3 | 1.7 | 6 |

Classic-era performance (1950 to 1974)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **team** | **total\_matches** | **total\_opponents** | **win\_ratio** | **loss\_ratio** | **goals\_scored\_per\_match** | **goals\_conceded\_per\_match** | **goals\_ratio** | **overall\_performance\_score** |
| Brazil | 260 | 43 | 0.7 | 0.2 | 2.3 | 1 | 2.2 | 8.4 |
| Ghana | 118 | 24 | 0.6 | 0.2 | 2.6 | 1.2 | 2.2 | 7.5 |
| Costa Rica | 102 | 19 | 0.6 | 0.2 | 2.2 | 1 | 2.1 | 6.9 |
| Hungary | 214 | 38 | 0.6 | 0.2 | 2.5 | 1.3 | 1.9 | 6.8 |
| England | 235 | 37 | 0.6 | 0.2 | 2.2 | 1.2 | 1.9 | 6.8 |

Modern-era performance (1975 to 1999)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **team** | **total\_matches** | **total\_opponents** | **win\_ratio** | **loss\_ratio** | **goals\_scored\_per\_match** | **goals\_conceded\_per\_match** | **goals\_ratio** | **overall\_performance\_score** |
| Brazil | 325 | 62 | 0.6 | 0.1 | 2.1 | 0.8 | 2.8 | 9.4 |
| Iraq | 189 | 51 | 0.6 | 0.2 | 1.9 | 0.8 | 2.6 | 7.8 |
| Germany | 286 | 60 | 0.6 | 0.2 | 1.9 | 0.9 | 2.2 | 7.6 |
| South Korea | 353 | 80 | 0.6 | 0.2 | 1.9 | 0.8 | 2.4 | 7.3 |
| France | 225 | 56 | 0.6 | 0.2 | 1.7 | 0.8 | 2.1 | 6.8 |

Postmodern-era performance (1975 to 1999)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **team** | **total\_matches** | **total\_opponents** | **win\_ratio** | **loss\_ratio** | **goals\_scored\_per\_match** | **goals\_conceded\_per\_match** | **goals\_ratio** | **overall\_performance\_score** |
| Spain | 301 | 85 | 0.7 | 0.1 | 2.1 | 0.7 | 3.1 | 10.1 |
| Brazil | 334 | 76 | 0.7 | 0.1 | 2.1 | 0.7 | 3.1 | 9.8 |
| Netherlands | 280 | 73 | 0.6 | 0.2 | 2.1 | 0.8 | 2.5 | 8.2 |
| France | 310 | 74 | 0.6 | 0.2 | 1.8 | 0.8 | 2.5 | 8.1 |
| Iran | 313 | 94 | 0.6 | 0.2 | 1.9 | 0.8 | 2.5 | 7.9 |

Query

CREATE OR REPLACE TABLE

  footballApr23Self.teamstats\_ AS

SELECT

  DISTINCT team,

  COUNT(uniq\_row\_id) OVER(PARTITION BY team) AS total\_matches,

  COUNT(DISTINCT opposition) OVER(PARTITION BY team) AS total\_opponents,

  SUM(team\_score) OVER(PARTITION BY team) AS total\_scored,

  SUM(opposition\_score) OVER(PARTITION BY team) AS total\_conceded,

  reg\_result,

  COUNT(uniq\_row\_id) OVER(PARTITION BY team, reg\_result) AS count\_result

FROM

  `footballApr23Self.teamlvlresults\_`

ORDER BY

  2 DESC,

  4 DESC,

  3 DESC,

  5 DESC,

  6 DESC,

  1;

WITH

  base AS (

  SELECT

    t1.\*,

    t2.reg\_result AS reg\_result1,

    t2.count\_result AS count\_result1,

    t3.reg\_result AS reg\_result2,

    t3.count\_result AS count\_result2

  FROM

    `footballApr23Self.teamstats\_` t1

  JOIN

    `footballApr23Self.teamstats\_` t2

  ON

    (t1.team = t2.team

      AND t1.reg\_result = 'WIN'

      AND t2.reg\_result = 'TIE'

      AND t1.reg\_result <> t2.reg\_result)

  JOIN

    `footballApr23Self.teamstats\_` t3

  ON

    (t1.team = t3.team

      AND t3.reg\_result = 'LOSS'

      AND t1.reg\_result <> t3.reg\_result

      AND t2.reg\_result <> t3.reg\_result))

SELECT

  team,

  total\_matches,

  total\_opponents,

  ROUND(count\_result/total\_matches,1) AS win\_ratio,

  ROUND(count\_result2/total\_matches,1) AS loss\_ratio,

  ROUND(total\_scored/total\_matches,1) AS goals\_scored\_per\_match,

  ROUND(total\_conceded/total\_matches,1) AS goals\_conceded\_per\_match,

  ROUND(total\_scored/total\_conceded,1) AS goals\_ratio,

  ROUND((count\_result/total\_matches\*10 - count\_result2/total\_matches\*10 + total\_scored/total\_matches - total\_conceded/total\_matches + total\_scored/total\_conceded),1) AS overall\_performance\_score

FROM

  base

WHERE

  total\_matches > 200

ORDER BY

  9 DESC,

  4 DESC,

  6 DESC,

  1;

Overall assessment:

A limitation of above analysis is that the quality of opponents is not taken into account but only results of the matches are analysed, hence the outcome can be biased towards teams that compete with other teams that rank low.

Looking at the all-time performance and also performance in different eras, it can be said that Brazil has been the most consistent and superlative performer in international football.

Some other teams which have consistently done well across eras are England, Spain, and Germany.

1. COUNTRIES THAT ARE LIKELY TO CONTINUE TO BE SUCCESSFUL IN IMMEDIATE FUTURE

To understand which countries are the most likely to continue to be successful in immediate future, we will perform an analysis similar to ‘RFM analysis (Recency, Frequency, Monetary)’ that is often performed on customer bases of product companies.

We will be applying three parameters – Recency, Frequency, and Totality – in total wins by different teams, and understand the quartiles where each team falls in each of these parameters.

This will give us a model to comment on which teams would be most likely to perform well in immediate future given the available data.