**Goalscorer dataset**

**1. remove duplicates**

A screenshot of a computer error

Description automatically generated

Ensuring the datasets has no duplicate row is key to not repeating and ensuring its new data that is being analysed.

1. A screenshot of a computer

   Description automatically generated**Change date.**

The date changed from time/date format to date to provide a correct format.

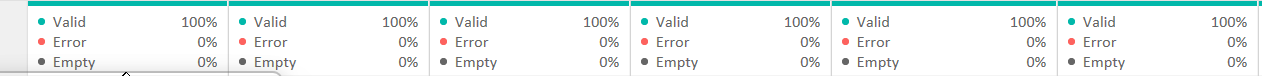
**3. Correct format**

**A screenshot of a computer

Description automatically generated**

For grammar purposes all rows with all capital letters have been transformed for it be analysed easier.

**4. Data Quality**

****

All rows have been quality checked and are all 100% are accurate.

**5. Rename Column**

****

Changes to column have been made to help ensure correct context is provided for that dataset and easier to analyse.

**Old Dataset**

A screenshot of a computer

Description automatically generated

**New dataset**

**A screenshot of a computer

Description automatically generated**

-- Question 1: Find the total number of matches played in the dataset

SELECT COUNT(\*) AS TotalMatches

FROM matches\_table;

-- Question 2: What is the average number of goals scored by home teams and away teams?

SELECT AVG(home\_score) AS AvgHomeGoals, AVG(away\_score) AS AvgAwayGoals

FROM matches\_table;

-- Question 3: Retrieve matches where the home team scored more than the away team.

SELECT \*

FROM matches\_table

WHERE home\_score > away\_score;

-- Question 4: Which team has the highest total number of goals scored as the home team?

SELECT home\_team, SUM(home\_score) AS TotalHomeGoals

FROM matches\_table

GROUP BY home\_team

ORDER BY TotalHomeGoals DESC

LIMIT 1;

-- Question 5: Find matches that ended in a draw (equal scores for home and away teams).

SELECT \*

FROM matches\_table

WHERE home\_score = away\_score;

-- Question 6: Retrieve matches held in Scotland.

SELECT \*

FROM matches\_table

WHERE city = 'Glasgow' OR hosting\_country = 'Scotland';

-- Question 7: Find the tournaments where England played as the home team.

SELECT DISTINCT tournament\_type

FROM matches\_table

WHERE home\_team = 'England';

-- Question 8: Count the number of matches where the hosting country was not the neutral ground.

SELECT COUNT(\*) AS NonNeutralMatches

FROM matches\_table

WHERE hosting\_country <> 'FALSE';

Penalty Shootout winner Dataset

**1. Remove duplicates**

A screenshot of a computer error

Description automatically generated

No duplicates rows provide originality and more accurate findings.

1. **A screenshot of a computer

   Description automatically generatedDelete column.**

This column has been removed from the database as its not providing any valuable information to the project or further contribution.

1. **Change date to correct format.**

A screenshot of a computer

Description automatically generated

Date format changed to fit standard calendar proceeding.

1. **Data quality**

A screenshot of a computer

Description automatically generated

Ensure all data is validated and to which no rows are empty nor contain errors.

1. **A screen shot of a computer

   Description automatically generatedRename Column**



The column has been removed to provide better context and understanding.

**Old Dataset**

**A screenshot of a computer

Description automatically generated**

**Data cleaned and prepared dataset.**

**A screenshot of a computer

Description automatically generated**

Results Dataset – First 795 rows – from 1872 to 1924 April

1. **Total number of matches played in the results table in that period.**

****

**A screenshot of a computer

Description automatically generated**

The total number of matches being analysed is 795 rows.

1. **The average number of goals scored by home and away teams from 1872 to 1924.**

****

**A screenshot of a computer

Description automatically generated**

The average number home goals scored is 2.2 goals per game and the average aways goals is 1.7, this finding shows that the home team is more likely to score a goal.

1. **A screenshot of a computer

   Description automatically generatedMatches where the home team scored more than the away team from 1872 to 1880**

This finding portrays Scotland where the most dominant home team in the 1872 - 1880 because they won all their home matches and scored more than the away team every time.

**3b) Matches where the home team scored more than the away team from 1880 to 1890.**

**A screenshot of a table

Description automatically generated**

Scotland is again the most dominant home team from 1880 to 1890 and won the most matches however England were a close second with a total eight wins.

**3c) Matches where the home team scored more than the away team from 1890 to 1899.**



**A screenshot of a computer

Description automatically generated**

In the 1890s, England became the most dominant home team and have won more matches at home than Scotland.

**3d) Matches where the home team scored more from 1990 to 1909.**



A screenshot of a computer

Description automatically generated

A table of sports teams

Description automatically generated

From these finding England and Hungary are joint home team winners in the first decade of the 1900s.

**3e) who won the most games in the 1910s.**



A screenshot of a computer

Description automatically generated

A table of names and numbers

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

A table of names and numbers

Description automatically generated

A screenshot of a computer

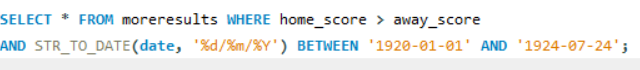
Description automatically generated

A table of names and numbers

Description automatically generated

In the 1910s the most successful national home team is Argentina however due the first world war the British national teams had to stop playing that potentially stopped from having more wins.

**3f) Home team winners from 1920 to 1924(exact date to the 795th row)**



A screenshot of a computer

Description automatically generated

A table with numbers and symbols

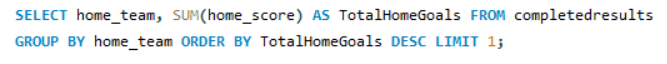
Description automatically generated

A table of names and numbers

Description automatically generated with medium confidence

Argentina once again is the most successful team once in the first 4 years of the 1920s.

**4) Which team has the highest total number of goals scored as the home team?**

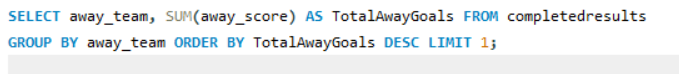
****

**A screenshot of a computer

Description automatically generated**

The highest goal scoring home team from 1872 to 1924 is Scotland.

**5) Which team has the highest total number of goals scored as the away team?**



A screenshot of a computer

Description automatically generated

The highest goal scoring away team is England, this finding present they are more likely to win away from home.

**6) matches that ended in a draw (equal scores for home and away teams) from 1872 to 1899**

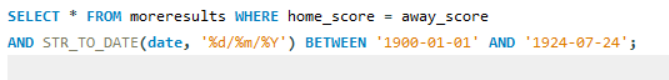
****

**A screenshot of a computer

Description automatically generated**

Scotland drew the most games within this frame.

**6b) matches that ended in a draw (equal scores for home and away teams) from 1900 to 1924.**



**A screenshot of a computer

Description automatically generated**

**A table of names and numbers

Description automatically generated with medium confidence**

**A screenshot of a computer screen

Description automatically generated**

**A table of text with black text

Description automatically generated with medium confidence**

**A table of numbers with black text

Description automatically generated with medium confidence**

**A screenshot of a computer screen

Description automatically generated**

From 1900 to 1924, the joint national teams that got the most draws were Argentina and Wales.

**7) Tournaments hosted by national teams not including friendlies**



A screenshot of a computer

Description automatically generated

Argentia hosted the most tournaments in South America and all British teams had a chance to host the tournament.

**8) Question 8: Count the number of matches where the hosting country was not the neutral ground.**

****

**A screenshot of a computer

Description automatically generated**

All matches were played on a neutral ground in the from 1872 to 1924.

**Results Dataset – Next 468 rows – from 1924 April – 1929 December**

1. **Total number of matches played in the results table in that period.**

****

**A screenshot of a number

Description automatically generated**

Within this period a total number of matches played which was 468 and will be analysed in this phase.

1. **The average number of goals scored by home and away teams from 1924 to 1929.**

****

**A screenshot of a computer

Description automatically generated**

The home team is twice as likely to score in each game, this finding portrays that the home side are twice as likely to score than the away team.

1. **Matches where the home team won from 1924 to 1929.**

****

****

A little more than half the home teams won in this period which highlights the home advantages in international matches.

**3b) The home team with the most wins in this period in descending order**

**A screenshot of a table

Description automatically generated**

Sweden has the most total wins in this period, the findings indicate they were the most successful home team in this period.

1. **Home team with the most goals scored.**

****

**A screenshot of a table

Description automatically generated**

In this find it shows that Sweden is the most dominant goal scoring team in the world between 1924 to 1929.

1. **Away team with most goals scored in this period.**

****

**A screenshot of a computer

Description automatically generated**

Sweden is the most successful away team in terms of scoring goals and the findings are for the Top 10 national teams in this period.

1. **Find matches that ended in a draw between 1924 to 1929.**

****

****

81 matches ended in a draw out 468 matches in total which shows that there is less than a 20% chance of a draw and the probability of winning a match is heightened whether you play home or away.

**6b) Most draws at home.**

**A screenshot of a computer

Description automatically generated**

**A table with numbers and names

Description automatically generated**

Argentina have drawn the most games at home, to potentially means they had less goals scorer in this period which led to this.

**6c) Most draws away from home.**

**A screenshot of a computer code

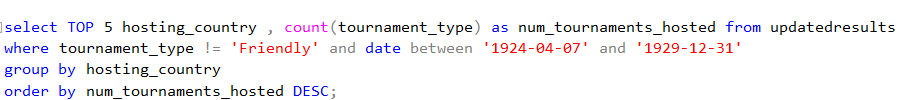
Description automatically generated**

**A screenshot of a table

Description automatically generated**

Argentina also has the most draws away from home in this period.

1. **Find top 10 hosting countries between 1924 to 1929.**

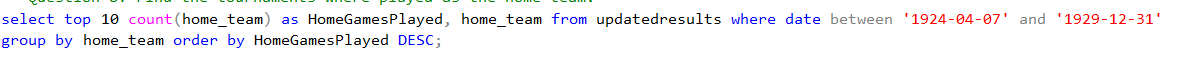


A screenshot of a computer

Description automatically generated

Argentina have hosted the most tournaments and Chile and Uruguay are joint second hosting countries, overall, there was more matches played in South America in this period as you can see the highest European teams only managed to host 8/9 tournaments at best.

1. **Select top 10 teams that played at home.**



A screenshot of a computer

Description automatically generated

Sweden played the most home games in this period there are six teams that won twenty games which is impressive.

1. **Select top 10 teams that played away from home.**



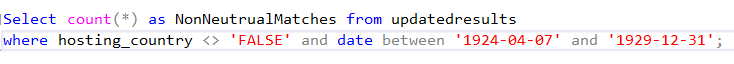
**A screenshot of a computer

Description automatically generated**

Sweden also have played the most away games however Argentina who played the second most games at home, are not in the top 10, this portrays the home advantage that Argentina possessed in this period.

****

1. **Matches played in a non-neutral game.**

****

**A screenshot of a computer

Description automatically generated**

Once again, all matches were played to a team’s advantage there was not a neutral ground for any of those games.