**Data Exploration of World Cup 2022**

This project uses three different datasets downloaded from Kaggle, a full fixtures list of the tournament, all player details, and all coach details. The aim of this project is to explore a variety of insights into the World Cup 2022 based on the ten questions below. I will using MySQL and will be using a variety of different SQL queries including WHERE, ORDER BY, GROUP BY, CASE, creating views, subqueries and joins.

1. *Which team scored the most goals?*
2. *Which team had the highest number of assists?*
3. *Which team conceded the most goals?*
4. *Show a breakdown of the tournaments win-loss record.*
5. *Which team had the most shots on target?*
6. *Calculate the goal conversion rate of each team.*
7. *Which team had the highest average possession and average passes completed throughout the tournament?*
8. *How many teams have a national coach? Where do most of the coaches come from?*
9. *What's the competition accuracy goal rate for each team and goal conversion for each game?*
10. *Find the total number of players registered per team and how many of them play in their national league.*

Answers:

1. A screenshot of a computer

   Description automatically generatedSELECT team, SUM(goal\_scored) as goals\_scored

FROM fixtures

GROUP BY team

ORDER BY 2 DESC

LIMIT 1;

= France - 16

1. A screenshot of a computer

   Description automatically generatedSELECT team, SUM(assists) as assists

FROM fixtures

GROUP BY team

ORDER BY 2 DESC

LIMIT 1;

= France - 12

1. A screenshot of a computer

   Description automatically generatedSELECT against,

SUM(goal\_scored) AS goals\_conceded

FROM fixtures

GROUP BY against

ORDER BY 2 DESC;

= Costa Rica conceded the most goals. 64% of their goals conceded were conceded to their 7-0 defeat to Spain.

1. SELECT team,

COUNT(\*) AS total\_matches,

SUM(CASE WHEN points = 3 THEN 1 ELSE 0 END) AS total\_wins,

SUM(CASE WHEN points = 1 THEN 1 ELSE 0 END) AS total\_draws,

SUM(CASE WHEN points = 0 THEN 1 ELSE 0 END) AS total\_losses

FROM fixtures

GROUP BY team

ORDER BY total\_wins DESC;

A screenshot of a computer

Description automatically generated

1. A screenshot of a computer

   Description automatically generatedSELECT team,   
    SUM(shots\_on\_target) as shots\_on\_target

FROM fixtures

GROUP BY team

ORDER BY shots\_on\_target DESC;

*=* Argentina. I found this result quite interesting as Morocco came 4th overall in the tournament but are ranked 9th in the number of shots on target whereas Germany was knocked out of the group stages but came 7th highest in their shots on target. After seeing these two results, I was interested in seeing both their goal conversion rates.

1. SELECT team,

SUM(goal\_scored) AS goals\_scored,  
 SUM(shots\_on\_target) AS shots\_on\_target,   
 ROUND((SUM(goal\_scored) / NULLIF(SUM(shots\_on\_target), 0)) \* 100, 2) AS conversion\_rate

FROM fixtures

GROUP BY team

ORDER BY conversion\_rate DESC;

*=* Netherlands scored the highest goal conversion rate achieving 62.5%. Further to point 5, Morocco’s goal conversion rate was 14th highest (35.29%) and Germany ranked 23rd overall (26.09%).

A screenshot of a computer

Description automatically generated

1. SELECT team,

CONCAT(ROUND(AVG(possession),0),'%') AS highest\_avg\_possession,

CONCAT(ROUND(SUM(passes\_completed) / SUM(passes)\*100,0),'%') AS avg\_passes\_completed

FROM fixtures

GROUP BY team

ORDER BY highest\_avg\_possession DESC;

*=* Spain ranked 1st in both the metrics. (I used CONCAT to add in the % as well as ROUND with the results to show a better view of the results).

A screenshot of a computer

Description automatically generated

1. SELECT (

-- This is the first subquery that calculates teams that have a national coach.

SELECT COUNT(\*)

FROM coachdetails

WHERE country = coach\_nationality) AS team\_with\_national\_coach,

-- This is the outermost part of the query calculation the count of total teams.

COUNT(\*) AS total\_teams,

-- This is the second subquery that calculates what percentage of managers make up the total number of national managers managing their own nation.

CONCAT(ROUND((SELECT COUNT(\*)

FROM coachdetails

WHERE country = coach\_nationality) / COUNT(\*) \* 100, 2),'%') AS percentage\_total

FROM coachdetails;

A screenshot of a computer

Description automatically generated

= Out of 32 different managers, 22 (69%) are managers representing their national teams. The most popular countries of origin for coaches are Argentina, France, Spain, and Portugal, all four countries also had a national coach as well. See below:

SELECT coach\_name, country, coach\_nationality,

COUNT(coach\_nationality) OVER(PARTITION BY coach\_nationality) AS coach\_nationality

FROM Coachdetails

ORDER BY coach\_nationality DESC

A screenshot of a computer

Description automatically generated

1. SELECT team, against, goal\_scored, total\_attempts,

CASE

WHEN total\_attempts = 0 THEN '0'

ELSE CONCAT(ROUND(goal\_scored/total\_attempts,2),'%')

END AS goal\_conversion,

SUM(goal\_scored) OVER(PARTITION BY team) AS comp\_goals\_scored,

SUM(total\_attempts) OVER(PARTITION BY team) AS comp\_total\_attempts,

CONCAT(ROUND(SUM(goal\_scored) OVER(PARTITION BY team) / SUM(total\_attempts) OVER(PARTITION BY team),2),'%') AS competition\_accuracy

FROM fixtures

ORDER BY competition\_accuracy DESC;

A screenshot of a computer

Description automatically generated

To display the results, I used case statement to pull to goal\_conversion which calculates the number of goals scored divided by total number of attempts. Then competition\_accuracy calculates the total number of goals scored during the tournament divided by the tournaments total number of attempts on goal. The actual number of goals scored, and total attempts are shown to look at how accuracy compares to total number of attempts. Costa Rica has the highest accuracy rate (27%) with a relatively low total number of goals scored (3). Top 2 scoring countries are however second and third in teams of accuracy rate 24% for both England and Spain with 9 total goals scored by each.

1. -- First I am checking how many players are playing in the same nation they are representing.

SELECT player\_name, nation, country\_league

FROM squaddetails

WHERE nation = country\_league;

-- This first view is the count of players that play in their national league. (Use N in join query).

CREATE VIEW national\_league AS

SELECT nation, COUNT(\*) as players

FROM squaddetails

WHERE nation = country\_league

GROUP BY nation

ORDER BY players DESC

-- This second view is the count of players registered for their national team. (Use R in join query).

CREATE VIEW registered\_players AS

SELECT nation, COUNT(\*) AS total\_players

FROM squaddetails

GROUP BY nation

ORDER BY total\_players DESC

-- Creating a left join using both the views to achieve desired results:

SELECT

n.nation,

n.players AS playing\_in\_their\_nation,

r.total\_players AS total\_players\_registered,

CONCAT(ROUND(n.players/total\_players \* 100,0),"%") AS percentage\_players – this percetage is showing how the percentage of players

FROM national\_league n

JOIN registered\_players r

A screenshot of a computer

Description automatically generated ON n.nation = r.nation