**FIFA WORLD CUP 2022 Analysis using SQL**

**The FIFA World Cup, an event cherished by football enthusiasts worldwide, brings together a tapestry of emotions and moments that are etched in the annals of sporting history. While millions tune in to watch the beautiful game from the comfort of their homes, there's an electrifying atmosphere that envelopes the stadiums. This is where dreams are realized, where heroes are born, and where the heartbeat of the sport echoes through the cheers of the fans.**

**In the realm of international football, talent and teamwork are paramount. Every player who graces the World Cup stage is a testament to years of dedication and unwavering commitment. Their skills are honed to perfection through countless hours of practice, and their ability to adapt and strategize can be the difference between glory and heartbreak.**

**Yet, in the world of competitive football, it's not just about how well a team executes its own game plan; it's also about understanding and countering the strategies of their opponents. This is where a deep analysis of both the opposition and one's own strengths and weaknesses becomes pivotal.**

**And this is where you come in.**

**As an integral part of the team behind the scenes, your role is crucial. You are the strategist, the data guru, the one who can turn numbers and statistics into insights that can give a team the upper hand. Your task is to analyze not only the tactics and tendencies of rival teams but also to assess the performance and potential of your own players.**

**Using the latest technology and the wealth of data at your disposal, you dive into the world of football analytics. You dissect the playing style of opposing teams, scrutinizing their past performances and identifying key players who pose a threat. You help your team develop strategies to exploit weaknesses and neutralize strengths.**

**As the tournament progresses, the pressure mounts, and the stakes become even higher. Your work becomes a silent but essential component of the team's success. Coaches, players, and fans may not see your efforts on the field, but they feel the impact in every well-executed play and every victory.**

**In the world of FIFA, you are the unsung hero, the one who helps turn dreams into reality. Your dedication to the game and your ability to transform data into actionable insights contribute to the magic of the World Cup, making every goal, every save, and every moment on the pitch that much more extraordinary.**

**Module 1**

**Task 1: Data Download, Import, and Database Connection.**

**In [1]:**

***# -- Load the sql extention ----***

**%load\_ext sql**

**​**

***# --- Load your mysql db using credentials from the "DB" area ---***

**%sql mysql+pymysql://bd83926d:Cab*#22se@localhost/bd83926d***

**Out[1]:**

**'Connected: bd83926d@bd83926d'**

**Module 2**

**Task 1: Counting Unique Customers.**

**We are conducting an analysis of the match data in our database to gain insights into the number of matches recorded. This information is crucial for understanding the scale and volume of matches in our records, which can provide valuable insights into various aspects of our business operations and performance.**

**In [2]:**

**%%sql**

**SELECT DISTINCT COUNT(match\_no) AS number\_of\_unique\_matches from match\_data;**

**UsageError: Cell magic `%%sql` not found.**

**Task 2: Obtaining a List of Unique Referees.**

**Continuing our comprehensive analysis of the match data in our database, we are now focusing on understanding the distinct referees who have officiated these matches. This aspect of our project complements our previous efforts to gauge the volume of matches, forming a holistic view of our sports-related operations.**

**In [8]:**

**%%sql**

**SELECT DISTINCT referee AS List\_of\_Unique\_Referees FROM match\_data;**

**\* mysql+pymysql://bd83926d:\*\*\*@localhost/bd83926d**

**29 rows affected.**

**Task 3: Finding the Hour with the Highest Frequency.**

**%%sql**

**SELECT hour, COUNT(match\_no) AS Number\_of\_Matches**

**FROM match\_data**

**GROUP BY hour**

**ORDER BY COUNT(match\_no) DESC**

**LIMIT 1;**

**Task 4: Fetching Data for Match Number 5.**[**¶**](https://jupyter.hicounselor.com:4957/notebooks/book.ipynb#Task-4:-Fetching-Data-for-Match-Number-5.)

**%%sql**

**SELECT \* FROM match\_data**

**WHERE match\_no=5;**

**Task 5: Fetching Position Data for Match Number 5.**[**¶**](https://jupyter.hicounselor.com:4957/notebooks/book.ipynb#Task-5:-Fetching-Position-Data-for-Match-Number-5.)

**%%sql**

**SELECT 1\_poss, 2\_poss**

**FROM match\_data**

**WHERE match\_no=5;**

**Task 6: Retrieving Goal Prevention Data for Match Number 5.**

**%%sql**

**SELECT 1\_goal\_prevented, 2\_goal\_prevented FROM match\_data**

**WHERE match\_no=5;**

**Task 7: Finding Peak Performance.**

**%%sql**

**SELECT \* FROM match\_data**

**WHERE 1\_ontarget=(SELECT MAX(1\_ontarget) FROM match\_data);**

**Identifying Team 2's Top On-Target Performance.**

**%%sql**

**SELECT second\_team,first\_team,1\_ontarget,2\_ontarget,2\_goals,(2\_goals/2\_ontarget\*100) AS percentage FROM match\_data**

**WHERE 2\_ontarget-2\_goals=(SELECT MAX(2\_ontarget-2\_goals) FROM match\_data);**

**Task 9: Identifying Match with Maximum Attendance.**

**%%sql**

**SELECT match\_no,first\_team, second\_team,venue, attendance FROM match\_data**

**WHERE attendance=(SELECT MAX(attendance) FROM match\_data);**

**Task 10: Analyzing Team Performance at Al Janoub Stadium.**[**¶**](https://jupyter.hicounselor.com:4957/notebooks/book.ipynb#Task-10:-Analyzing-Team-Performance-at-Al-Janoub-Stadium.)

**%%sql**

**SELECT first\_team, second\_team, 1\_poss,2\_poss, 1\_goals,2\_goals,(1\_goals/1\_poss\*100),(2\_goals/2\_poss\*100) FROM match\_data**

**WHERE venue="Al Janoub Stadium"**

**GROUP BY match\_no,first\_team, second\_team,1\_poss,2\_poss, 1\_goals,2\_goals;**

**Task 11: Analyzing Penalty Area Goals at Different Venues.**

**%%sql**

**SELECT venue, SUM(1\_goal\_inside\_penalty\_area+2\_goal\_inside\_penalty\_area) AS Toalgoalsinside,SUM(1\_goal\_outside\_penalty\_area+2\_goal\_outside\_penalty\_area) AS tOTALGOALSOUTSIDE**

**FROM match\_data**

**GROUP BY venue;**

**​**

**​**

**Task 12: Comparing Goal Success Rates.**

**%%sql**

**WITH goals\_summary AS (**

**SELECT venue, SUM(1\_goal\_inside\_penalty\_area+2\_goal\_inside\_penalty\_area) AS Totalgoalsinside,SUM(1\_goal\_outside\_penalty\_area+2\_goal\_outside\_penalty\_area) AS Totalgoalsoutside,**

**SUM(1\_attempts\_inside\_penalty\_area+2\_attempts\_inside\_penalty\_area) AS Totalattemptsinside,SUM(1\_attempts\_outside\_penalty\_area+2\_attempts\_outside\_penalty\_area) AS Totalattemptsoutside**

**FROM match\_data**

**GROUP BY venue**

**)**

**SELECT venue,(Totalgoalsinside/Totalattemptsinside)\*100,(Totalgoalsoutside/Totalattemptsoutside)\*100**

**FROM goals\_summary**

**;**

**​**

**Stored Procedure for Extracting Match Data for Two Teams.**[**¶**](https://jupyter.hicounselor.com:4957/notebooks/book.ipynb#Task-13:-Stored-Procedure-for-Extracting-Match-Data-for-Two-Teams.)

**%%sql**

**CREATE PROCEDURE getmatchdata2 (IN firstteam TEXT,**

**IN secondteam TEXT)**

**BEGIN**

**SELECT**

**venue,**

**score,**

**1\_poss,**

**2\_poss,**

**1\_panelties\_scored,**

**2\_panelties\_scored,**

**1\_defensive\_pressure\_applied,**

**2\_defensive\_pressure\_applied**

**FROM match\_data**

**WHERE first\_team = firstteam AND second\_team = secondteam;**

**END;**

**Task 14: Calling the Created Procedure.**

**%%sql**

**CALL getmatchdata2('ARGENTINA','FRANCE');**

**​**

**​**

**Module 3**[**¶**](https://jupyter.hicounselor.com:4957/notebooks/book.ipynb#Module-3)

**Task 1: Top 5 Goal-Scoring Players with Dribbling Success Rates.**

**%%sql**

**SELECT**

**st.player,**

**st.assists,**

**st.cards\_yellow,**

**st.cards\_red,**

**st.goals,**

**CASE**

**WHEN p.dribbles IS NULL OR p.dribbles = 0 THEN 'None'**

**WHEN p.dribbles\_completed=0 AND p.dribbles>0 THEN '0.0'**

**ELSE (p.dribbles\_completed) / (p.dribbles)\*100**

**END AS dribblespercent**

**FROM player\_stats AS st**

**LEFT JOIN player\_possession AS p ON st.player = p.player**

**ORDER BY**

**st.goals DESC**

**LIMIT 5;**

**Task 2: Analyzing Players: Goals vs. Progressive Passes Received.**

**%%sql**

**SELECT st.player,st.goals,p.progressive\_passes\_received,**

**ROUND((st.goals/p.progressive\_passes\_received)\*100,2) AS Ratio\_goals\_passes**

**FROM player\_stats AS st**

**LEFT JOIN**

**player\_possession AS p**

**ON st.player=p.player**

**ORDER BY st.goals DESC;**

**​**

**Task 3: Analyzing Player Performance Criteria for Portugal.**

**%%sql**

**SELECT st.player,**

**st.goals,**

**p.touches\_def\_pen\_area,**

**p.touches\_def\_3rd,**

**p.touches\_att\_3rd,**

**p.touches\_att\_pen\_area,**

**p.dribbles\_completed**

**FROM player\_stats AS st**

**LEFT JOIN**

**player\_possession AS p**

**ON st.player=p.player**

**WHERE p.team='Portugal' AND st.games\_starts<2 AND st.goals>0;**

**Task 4: Analyzing Club Dribbling Performance.**

**%%sql**

**SELECT**

**SUM(p.dribbles\_completed),**

**MAX(p.dribbles\_completed), %%sql**

**SELECT**

**COUNT(player),SUM(goals), SUM(goals)/COUNT(player)\*100 AS ratio**

**FROM player\_stats**

**WHERE age<25;**

**​**

**​**

**st.club**

**​**

**FROM player\_stats AS st**

**INNER JOIN**

**player\_possession AS p**

**ON st.player=p.player**

**WHERE st.club IS NOT NULL**

**GROUP BY st.club**

**ORDER BY MAX(p.dribbles\_completed) DESC;**

**​**

**Task 5: Analyzing Goal Scoring Performance of Players Under 25.**

**%%sql**

**SELECT**

**COUNT(player),SUM(goals), SUM(goals)/COUNT(player)\*100 AS ratio**

**FROM player\_stats**

**WHERE age<25;**

**​**

**​**

**Task 6: Analyzing Goal Scoring Performance of Players Age 25 and Older.**[**¶**](https://jupyter.hicounselor.com:4957/notebooks/book.ipynb#Task-6:-Analyzing-Goal-Scoring-Performance-of-Players-Age-25-and-Older.)

**%%sql**

**SELECT**

**COUNT(player),SUM(goals), SUM(goals)/COUNT(player)\*100 AS ratio**

**FROM player\_stats**

**WHERE age>=25;**

**Task 7: Top 5 Clubs with Most Players Under 25.**

**%%sql**

**SELECT COUNT(player),club**

**FROM player\_stats**

**WHERE age<25**

**GROUP BY club**

**ORDER BY COUNT(player) DESC**

**LIMIT 5;**

**Module 4**

**Task 1: Top 10 Players with Longest Average Shot Distance.**

**%%sql**

**SELECT player, average\_shot\_distance, goals,shotS**

**FROM player\_shooting**

**ORDER BY average\_shot\_distance DESC**

**LIMIT 10;**

**​**

**​**

**Task 2: Analyzing Player Shot Accuracy**

**%%sql**

**SELECT player,shots,shots\_on\_target,goals,(shots\_on\_target / shots) \* 100 AS shot\_accuracy**

**​**

**FROM player\_shooting**

**​**

**ORDER BY shot\_accuracy DESC;**

**Task 3: Identifying High-Performing Players with Precision Shooting**

**%%sql**

**SELECT ps.player,st.club, ps.shots, ps.shots\_on\_target, ps.goals,(ps.shots\_on\_target / ps.shots)\* 100 AS shot\_accuracy**

**FROM player\_shooting AS ps**

**LEFT JOIN**

**player\_stats AS st**

**ON ps.player=st.player**

**WHERE (ps.shots\_on\_target / ps.shots)\* 100 >20 AND ps.goals> 2**

**ORDER BY shot\_accuracy DESC;**

**Task 4: Top 10 Clubs with Young and High-Performing Players.**

**%%sql**

**SELECT**

**st.club,COUNT(ps.player) AS young\_players\_count**

**FROM**

**player\_shooting AS ps**

**LEFT JOIN**

**player\_stats AS st**

**ON**

**ps.player = st.player**

**WHERE**

**(ps.shots\_on\_target/ ps.shots)\*100 >20 AND**

**ps.goals > 0 AND**

**ps.age < 28**

**GROUP BY**

**st.club**

**ORDER BY**

**young\_players\_count DESC,**

**st.club**

**LIMIT 10;**

**Task 5: Top 10 Teams with Young and High-Performing Players.**

**%%sql**

**SELECT**

**st.team,COUNT(ps.player) AS young\_players\_count**

**FROM**

**player\_shooting AS ps**

**LEFT JOIN**

**player\_stats AS st**

**ON**

**ps.player = st.player**

**WHERE**

**(ps.shots\_on\_target/ ps.shots)\*100 >20 AND**

**ps.goals > 0 AND**

**ps.age < 28**

**GROUP BY**

**st.team**

**ORDER BY**

**young\_players\_count DESC,**

**st.team**

**LIMIT 10;**

**Task 6: Analyzing Player Performance in Terms of Goals, Shots, and Assists.**

**%%sql**

**SELECT**

**ps.player,**

**ps.shots,**

**st.goals,**

**ps.shots\_on\_target,**

**st.assists\_per90,**

**st.goals\_per 90**

**FROM**

**player\_shooting AS ps**

**RIGHT JOIN**

**player\_stats AS st**

**ON**

**ps.player = st.player**

**WHERE**

**st.minutes\_90s> 2 AND st.assists IS NOT NULL**

**ORDER BY**

**st.goals DESC,**

**ps.player;**

**Module 5**

**Task 1: Analyzing Player Performance in the Opponent's Penalty Area.**

**%%sql**

**SELECT ps.player, st.club,ps.touches\_att\_pen\_area**

**FROM player\_possession AS ps**

**LEFT JOIN player\_stats AS st**

**ON ps.player=st.player**

**WHERE ps.touches\_att\_pen\_area>=1**

**ORDER BY**

**ps.touches\_att\_pen\_area DESC,**

**ps.player;**

**​**

**​**

**Task 2: Count of Players with Touches in Opponent's Penalty Area by Club**

**%%sql**

**​**

**SELECT st.club, COUNT(DISTINCT st.player)**

**FROM player\_stats AS st**

**LEFT JOIN**

**player\_possession AS ps**

**ON st.player=ps.player**

**WHERE ps.touches\_att\_pen\_area>=1**

**GROUP BY**

**st.club**

**ORDER BY COUNT(st.player) DESC;**

**Task 3: Average Player Touches in Different Field Areas.**

**%%sql**

**SELECT**

**AVG(touches\_def\_pen\_area),**

**AVG(touches\_def\_3rd),**

**AVG(touches\_mid\_3rd),**

**AVG(touches\_att\_3rd),**

**AVG(touches\_att\_pen\_area)**

**FROM player\_possession;**

**Task 4: Average Player Touches in Field Areas for Names Starting with 'A'.**[**¶**](https://jupyter.hicounselor.com:4957/notebooks/book.ipynb#Task-4:-Average-Player-Touches-in-Field-Areas-for-Names-Starting-with-'A'.)

**%%sql**

**SELECT player,**

**AVG(touches\_def\_pen\_area),**

**AVG(touches\_def\_3rd),**

**AVG(touches\_mid\_3rd),**

**AVG(touches\_att\_3rd),**

**AVG(touches\_att\_pen\_area)**

**FROM player\_possession**

**WHERE player LIKE "A%"**

**GROUP BY**

**player;**

**​**

**Task 5: Count of Players with Goals by Position.**

**%%sql**

**SELECT**

**COUNT(goals),**

**position**

**FROM player\_stats**

**WHERE goals>0**

**GROUP BY**

**position;**

**Task 6: Top-Scoring Defender in the Dataset.**

**%%sql**

**SELECT st.player, st.goals**

**FROM player\_stats AS st**

**LEFT JOIN player\_possession AS ps**

**ON st.player=ps.player**

**WHERE ps.position='DF'AND st.player = 'Nahuel Molina'**

**ORDER BY goals DESC**

**LIMIT 1;**

**​**

**Task 7: Top-Scoring Midfielder**

**%%sql**

**SELECT st.player, st.goals**

**FROM player\_stats AS st**

**LEFT JOIN player\_possession AS ps**

**ON st.player=ps.player**

**WHERE ps.position='MF'**

**ORDER BY goals DESC**

**LIMIT 1;**

**Task 8: Top-Scoring Forward.**[**¶**](https://jupyter.hicounselor.com:4957/notebooks/book.ipynb#Task-8:-Top-Scoring-Forward.)

**%%sql**

**SELECT st.player, st.goals**

**FROM player\_stats AS st**

**LEFT JOIN player\_possession AS ps**

**ON st.player=ps.player**

**WHERE ps.position='FW'**

**ORDER BY goals DESC**

**LIMIT 1;**

**​**

**Task 9: Ranking Players by Shooting Accuracy and Progressive Passes Received.**[**¶**](https://jupyter.hicounselor.com:4957/notebooks/book.ipynb#Task-9:-Ranking-Players-by-Shooting-Accuracy-and-Progressive-Passes-Received.)

**%%sql**

**WITH Top\_200\_Players AS (**

**SELECT**

**pp.player,**

**pp.progressive\_passes\_received**

**FROM**

**player\_possession AS pp**

**ORDER BY**

**pp.progressive\_passes\_received DESC**

**LIMIT 200**

**),**

**Player\_Shots AS (**

**SELECT**

**ps.player,**

**ps.shots\_on\_target**

**FROM**

**player\_shooting AS ps**

**),**

**Player\_Stats AS (**

**SELECT**

**t.player,**

**t.progressive\_passes\_received,**

**COALESCE(s.shots\_on\_target, 0) AS shots\_on\_target**

**FROM**

**Top\_200\_Players AS t**

**LEFT JOIN**

**Player\_Shots AS s**

**ON**

**t.player = s.player**

**),**

**Player\_Ratios AS (**

**SELECT**

**player,**

**(shots\_on\_target / NULLIF(progressive\_passes\_received, 0)\*100) AS ratio**

**FROM**

**Player\_Stats**

**)**

**SELECT**

**player,**

**ratio**

**FROM**

**Player\_Ratios**

**ORDER BY**

**ratio DESC**

**LIMIT 200;**