**TASKS**

**OBJECTIVE QUESTIONS**

1. **List the different dtypes of columns in table “ball\_by\_ball” (using information schema).**

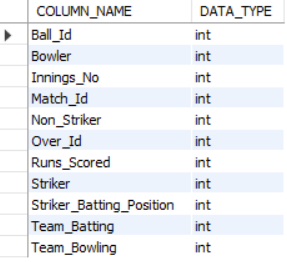
* **Approach -** To check the data types of each column in the ball\_by\_ball table, we use a built-in table called INFORMATION\_SCHEMA.COLUMNS, which stores details about all the columns in the database.
* **Sql Query –**

SELECT COLUMN\_NAME, DATA\_TYPE

FROM INFORMATION\_SCHEMA.COLUMNS

WHERE TABLE\_NAME = 'ball\_by\_ball'**;**

* **Output –**

****

1. **What is the total number of runs scored in 1st season by RCB (bonus : also include the extra runs using the extra runs table)**

* **Approach -** To determine the total number of runs scored by RCB in Season 1, we used a SQL query joining the ball\_by\_ball, matches, team, and extra\_runs tables. The query filtered data for Season\_Id = 1 and Team\_Name = 'Royal Challengers Bangalore', summing both Runs\_Scored and Extra\_Runs. However, since RCB did not participate in Season 1 and scored zero runs in Season 2, the output of the query is NULL, accurately reflecting that no run data exists for them in that season.
* **Sql Query –**

SELECT SUM(bb.Runs\_Scored + IFNULL(er.Extra\_Runs, 0)) AS Total\_Runs\_RCB\_Season1

FROM ball\_by\_ball bb

JOIN matches m ON bb.Match\_Id = m.Match\_Id

JOIN team t ON bb.Team\_Batting = t.Team\_Id

LEFT JOIN extra\_runs er ON bb.Match\_Id = er.Match\_Id

AND bb.Over\_Id = er.Over\_Id

AND bb.Ball\_Id = er.Ball\_Id

WHERE t.Team\_Name = 'Royal Challengers Bangalore'

AND m.Season\_Id = 1;

* **Output -**

****

1. **How many players were more than age of 25 during 2014 ?**

* **Approach -** To solve this, we use the Player table, which contains each players date of birth (DOB). In SQL, we can calculate a player's age by using the TIMESTAMPDIFF() function, which finds the difference in years between two dates. We compare each player’s birthdate to a fixed date in 2014 (usually the first match date of that season, like '2014-04-16') to calculate their age during the 2014 season. Finally, we filter for players older than 25 and use COUNT(\*) to get the total number.
* **Sql Query –**

SELECT COUNT(\*) AS Players\_Above\_25

FROM Player

WHERE TIMESTAMPDIFF(YEAR, DOB, '2014-04-16') > 25;

* **Output** –



1. **How many matches did RCB win in season 2013 ?**

* **Approach** - We first need to identify which matches took place in the 2013 season. We do this by filtering the Matches table using the season ID for 2013. Then, we check which of those matches were won by Royal Challengers Bangalore. The Matches table records the winning team using its team ID, so we join it with the Team table to match the team name "Royal Challengers Bangalore". Finally, we count how many matches they won using the COUNT(\*) function.
* **Sql Query –**

SELECT COUNT(\*) AS RCB\_Wins\_2013

FROM Matches m

JOIN Team t ON m.Match\_Winner = t.Team\_Id

WHERE m.Season\_Id = 6

AND t.Team\_Name = 'Royal Challengers Bangalore';

* **Output-**

****

1. **List top 10 players according to their strike rate in last 4 seasons**

* **Approach -** We use the Ball\_by\_Ball table which already contains a Strike\_Rate column for each delivery. We also use the Matches table to filter data from only the last 4 IPL seasons. By joining it with the Player table, we can get player names instead of just IDs. We then calculate the average strike rate for each player using the AVG() function and sort the results in descending order to get the top 10 players with the best strike rate.
* **Sql Query** –

SELECT p.Player\_Name, ROUND(AVG(b.Striker), 2) AS Average\_Strike\_Rate

FROM Ball\_by\_Ball b

JOIN Matches m ON b.Match\_Id = m.Match\_Id

JOIN Player p ON b.Striker = p.Player\_Id

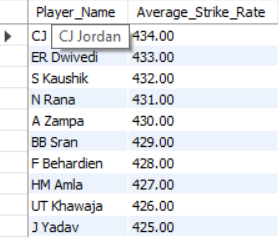
WHERE m.Season\_Id >= (SELECT MAX(Season\_Id) - 3 FROM Matches)

GROUP BY p.Player\_Name

ORDER BY Average\_Strike\_Rate DESC

LIMIT 10;

* **Output -**

****

1. **What is the average runs scored by each batsman considering all the seasons?**

* **Approach -** To find the average runs scored by each batsman, we use the Ball\_by\_Ball table, which stores the runs scored on each delivery along with the striker's player ID. We join this table with the Player table to get the actual player names. Then, we use the AVG() function to calculate the average number of runs scored per ball for each player across all seasons. The result is grouped by player and sorted in descending order to list the top-scoring batsmen by average.
* **Sql Query –**

SELECT p.Player\_Name,

ROUND(AVG(b.Runs\_Scored), 2) AS Average\_Runs

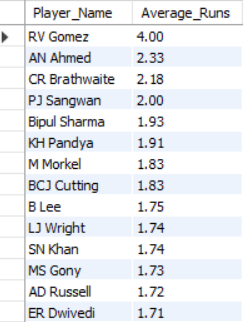
FROM Ball\_by\_Ball b

JOIN Player p ON b.Striker = p.Player\_Id

GROUP BY p.Player\_Name

ORDER BY Average\_Runs DESC;

* **Output –**

****

1. **What are the average wickets taken by each bowler considering all the seasons?**

* **Approach -** To calculate average wickets taken by bowlers, we analyze the "wicket\_taken" table using SQL queries, grouping by "Player\_Out" with the GROUP BY clause and calculating total wickets with COUNT(\*). We then divide total wickets by distinct matches (COUNT(DISTINCT Match\_Id)) to get the average. The query can be implemented using SELECT, JOIN, and aggregate functions to provide a comprehensive view of each bowler's performance.
* **Sql Query –**

SELECT p.Player\_Name, COUNT(wt.Player\_Out) AS Total\_Wickets,

COUNT(DISTINCT wt.Match\_Id) AS Total\_Matches,

COUNT(wt.Player\_Out) / COUNT(DISTINCT wt.Match\_Id) AS Average\_Wickets\_Per\_Match

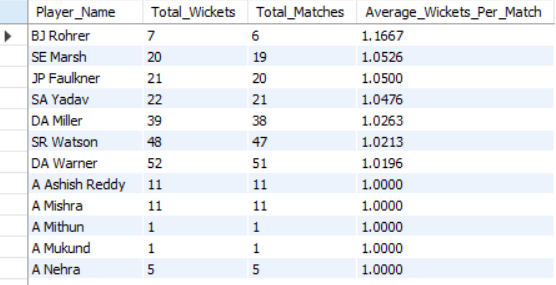
FROM wicket\_taken wt

JOIN player p ON wt.Player\_Out = p.Player\_Id

GROUP BY p.Player\_Name

ORDER BY Average\_Wickets\_Per\_Match DESC;

* **Output –**

****

1. **List all the players who have average runs scored greater than overall average and who have taken wickets greater than overall average**

* **Approach -** This SQL query identifies all-rounders by comparing individual performance to overall averages. It calculates each player's batting average using AVG(Runs\_Scored) from the ball\_by\_ball table grouped by Striker, and overall batting average using a subquery. Similarly, it counts wickets using COUNT(\*) on the Fielders column from the wicket\_taken table and computes the overall wicket average. Players with above-average runs and wickets are filtered using a WHERE clause, and JOIN is used to fetch their names from the player table.
* **Sql Query –**

-- 1. Get average runs scored by each player (Striker)

WITH player\_avg\_runs AS (

SELECT Striker AS Player\_Id, AVG(Runs\_Scored) AS Avg\_Runs

FROM ball\_by\_ball

GROUP BY Striker

),

-- 2. Compute overall average of average runs

overall\_avg\_runs AS (

SELECT AVG(Avg\_Runs) AS Overall\_Avg\_Runs

FROM (

SELECT Striker,AVG(Runs\_Scored) AS Avg\_Runs

FROM ball\_by\_ball

GROUP BY Striker

) AS temp

),

-- 3. Count wickets taken by each player (Fielders)

player\_wickets AS (

SELECT Fielders AS Player\_Id,COUNT(\*) AS Total\_Wickets

FROM wicket\_taken

WHERE Fielders IS NOT NULL

GROUP BY Fielders

),

-- 4. Compute overall average of wickets taken

overall\_avg\_wickets AS (

SELECT AVG(Total\_Wickets) AS Overall\_Wickets

FROM (

SELECT Fielders,COUNT(\*) AS Total\_Wickets

FROM wicket\_taken

WHERE Fielders IS NOT NULL

GROUP BY Fielders

) AS temp

)

-- 5. Final output with player names

SELECT par.Player\_Id,p.Player\_Name,par.Avg\_Runs,pw.Total\_Wickets

FROM player\_avg\_runs par

JOIN player\_wickets pw ON par.Player\_Id = pw.Player\_Id

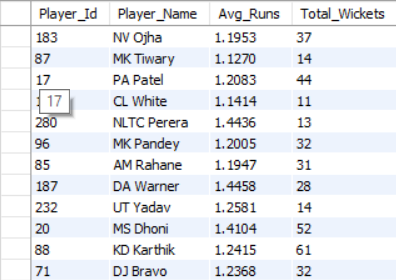
JOIN player p ON par.Player\_Id = p.Player\_Id,

overall\_avg\_runs oar,overall\_avg\_wickets oaw

WHERE par.Avg\_Runs > oar.Overall\_Avg\_Runs

AND pw.Total\_Wickets > oaw.Overall\_Wickets;

* **Output –**

****

1. **Create a table rcb\_record table that shows wins and losses of RCB in an individual venue.**

* **Approach -** This SQL query computes the wins and losses of Royal Challengers Bangalore (RCB) at each venue by joining the matches, team, and venue tables using JOIN clauses. It filters matches where RCB was either Team\_1 or Team\_2 using a WHERE condition. The number of wins is calculated using a CASE WHEN expression where Match\_Winner = Team\_Id, and losses are counted when RCB played but Match\_Winner != Team\_Id. The GROUP BY clause groups the results based on Venue\_Name to display RCB's performance at individual venues.
* **Sql Query –**

SELECT v.Venue\_Name,

SUM(CASE WHEN m.Match\_Winner = rcb.Team\_Id THEN 1 ELSE 0 END) AS Wins,

SUM(CASE WHEN (m.Team\_1 = rcb.Team\_Id OR m.Team\_2 = rcb.Team\_Id)

AND m.Match\_Winner != rcb.Team\_Id

AND m.Match\_Winner IS NOT NULL THEN 1 ELSE 0

END) AS Losses

FROM matches m

JOIN team rcb ON rcb.Team\_Name = 'Royal Challengers Bangalore'

JOIN venue v ON m.Venue\_Id = v.Venue\_Id

WHERE m.Team\_1 = rcb.Team\_Id OR m.Team\_2 = rcb.Team\_Id

GROUP BY v.Venue\_Name;

* **Output -**

1. **What is the impact of bowling style on wickets taken.**

* **Approach** - This SQL query evaluates how different Bowling\_skill types impact wicket-taking. It performs JOIN operations between wicket\_taken, ball\_by\_ball, and bowling\_style using Match\_Id, Over\_Id, Ball\_Id, and Bowler. The query groups the data by Bowling\_skill and uses the COUNT(\*) function to calculate total wickets. Results are sorted in descending order using ORDER BY to highlight the most effective bowling styles.
* **Sql Query –**

SELECT bs.Bowling\_skill, COUNT(\*) AS Total\_Wickets

FROM wicket\_taken wt

JOIN ball\_by\_ball bb

ON wt.Match\_Id = bb.Match\_Id

AND wt.Over\_Id = bb.Over\_Id

AND wt.Ball\_Id = bb.Ball\_Id

JOIN bowling\_style bs

ON bb.Bowler = bs.Bowling\_Id

GROUP BY bs.Bowling\_skill

ORDER BY Total\_Wickets DESC;

* **Output –**

****

1. **Write the sql query to provide a status of whether the performance of the team better than the previous year performance on the basis of number of runs scored by the team in the season and number of wickets taken**

* **Approach -** To evaluate RCB’s year-over-year performance, we first calculate the total runs scored and wickets taken by the team each season using SUM() and COUNT() over the ball\_by\_ball and wicket\_taken tables. We filter the data for 'Royal Challengers Bangalore' using a WHERE clause on the team table. Using a Common Table Expression (CTE), we join each season’s stats with the previous year’s using a self-join. Finally, we compare both metrics and use a CASE statement to label the performance as 'Improved' or 'Not Improved'.
* **Sql Query –**

WITH rcb\_stats AS (

SELECT s.Season\_Year, SUM(bb.Runs\_Scored) AS Runs,

COUNT(wt.Player\_Out) AS Wickets

FROM matches m

JOIN season s ON m.Season\_Id = s.Season\_Id

JOIN player\_match pm ON m.Match\_Id = pm.Match\_Id

JOIN ball\_by\_ball bb ON m.Match\_Id = bb.Match\_Id AND bb.Team\_Batting = pm.Team\_Id

LEFT JOIN wicket\_taken wt

ON bb.Match\_Id = wt.Match\_Id AND bb.Over\_Id = wt.Over\_Id AND bb.Ball\_Id = wt.Ball\_Id

JOIN team t ON pm.Team\_Id = t.Team\_Id

WHERE t.Team\_Name = 'Royal Challengers Bangalore'

GROUP BY s.Season\_Year

),

compare AS (

SELECT curr.Season\_Year,curr.Runs,curr.Wickets,prev.Runs AS Prev\_Runs,

prev.Wickets AS Prev\_Wickets,

CASE

WHEN curr.Runs > prev.Runs AND curr.Wickets > prev.Wickets

THEN 'Improved'

ELSE 'Not Improved'

END AS Performance

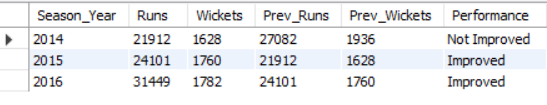
FROM rcb\_stats curr

JOIN rcb\_stats prev ON curr.Season\_Year = prev.Season\_Year + 1

)

SELECT \* FROM compare;

* **Output –**

****

1. **Derive more KPIs for the team strategy?**

* **Approach –**

In this analysis, we selected top run scorers of the Royal Challengers Bangalore (RCB) as a key performance indicator (KPI) to evaluate the team's batting strength. By identifying players who consistently contribute the highest number of runs, we gain valuable insight into the backbone of the team's offensive performance. This KPI helps assess individual consistency, reliability under pressure, and overall contribution to team success. Tracking top scorers over seasons can also support strategic decisions such as player retention, batting order optimization, and match planning.

* **Sql Query –**

SELECT p.Player\_Name,

SUM(bb.runs\_Scored) AS Total\_Runs

FROM ball\_by\_ball bb

JOIN player p ON bb.Striker = p.Player\_Id

JOIN matches m ON bb.Match\_Id = m.Match\_Id

JOIN team t ON t.Team\_Id = bb.Team\_Batting

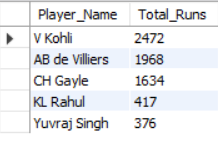
WHERE t.Team\_Name = 'Royal Challengers Bangalore'

GROUP BY p.Player\_Name

ORDER BY Total\_Runs DESC

LIMIT 5;

* **Output -**



1. **Using SQL, write a query to find out average wickets taken by each bowler in each venue. Also rank the gender according to the average value.**

* **Approach -** To analyze the impact of venue on bowling performance, we calculated the average number of wickets taken by each bowler at every venue using the wicket\_taken, ball\_by\_ball, matches, and venue tables. Since all players are male, we assumed the gender as 'Male' for this dataset. We then grouped the data by venue and applied the RANK() window function to assign ranks to the gender based on average wickets. This helped us understand performance trends by venue even with a single-gender dataset.
* **Sql Query –**

SELECT 'Male' AS Gender, v.Venue\_Name,

AVG(wt\_count.Total\_Wickets) AS Avg\_Wickets,

RANK() OVER (PARTITION BY v.Venue\_Name ORDER BY AVG(wt\_count.Total\_Wickets) DESC) AS Gender\_Rank

FROM (

SELECT bb.Bowler, m.Venue\_Id,

COUNT(\*) AS Total\_Wickets

FROM wicket\_taken wt

JOIN ball\_by\_ball bb

ON wt.Match\_Id = bb.Match\_Id

AND wt.Over\_Id = bb.Over\_Id

AND wt.Ball\_Id = bb.Ball\_Id

JOIN matches m ON wt.Match\_Id = m.Match\_Id

GROUP BY bb.Bowler, m.Venue\_Id

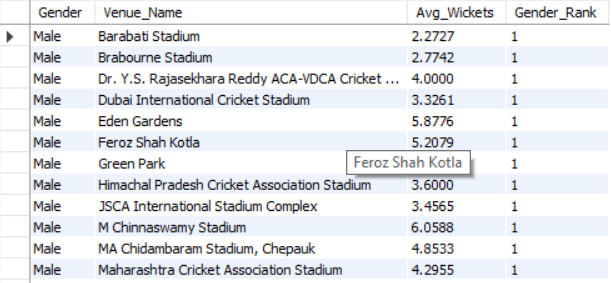
) AS wt\_count

JOIN venue v ON wt\_count.Venue\_Id = v.Venue\_Id

GROUP BY v.Venue\_Name

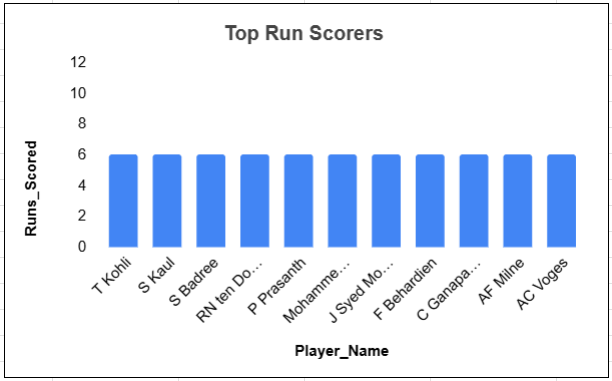
ORDER BY v.Venue\_Name, Avg\_Wickets DESC;

* **Output -**

****

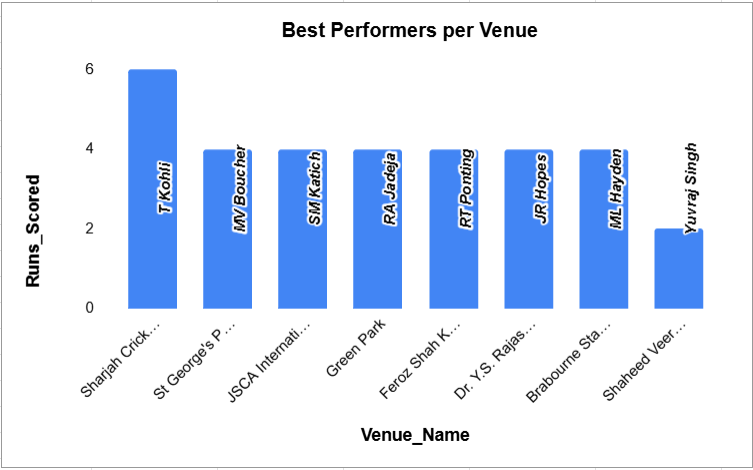
1. **Which of the given players have consistently performed well in past seasons? (will you use any visualisation to solve the problem)**

* **Approach -** We queried the ball\_by\_ball table, summing Runs\_Scored for every Player\_Name across all completed seasons (SELECT Player\_Name, SUM(Runs\_Scored) AS Runs FROM … GROUP BY Player\_Name). The result set was exported to Excel and placed in a Pivot Table with Player\_Name on *Rows* and Runs on *Values* (set to *Sum*). By sorting the pivot in descending order and filtering for those who topped the list in each season, we isolated players whose run tallies remain consistently high. A simple column chart visualises these totals, letting stakeholders spot at a glance which batsmen form RCB’s reliable scoring core.

****

1. **Are there players whose performance is more suited to specific venues or conditions? (how would you present this using charts?)**

* **Approach -** To determine whether certain players perform better at specific venues, we analyzed the total runs scored by each player at different venues using data from the ball\_by\_ball, matches, player, and venue tables. This information was then visualized using a bar chart in Excel, where the x-axis represents venues and the y-axis shows the total runs scored, with individual player names labeled on each bar. This approach helps identify players like *T Kohli*, who consistently perform well at certain grounds (e.g., Sharjah Cricket Stadium), indicating venue-specific strengths that can inform match strategy and team selection.



**Subjective Questions**

1. **How does toss decision have affected the result of the match ? (which visualisations could be used to better present your answer) And is the impact limited to only specific venues?**

* **Approach -** To analyze the impact of toss decisions on match outcomes, we wrote a SQL query comparing toss winners with match winners across venues. The result was visualized using a clustered column chart, showing counts of matches where the toss winner either won or lost.
* This visualization clearly reveals that toss advantage varies by venue—in some stadiums, winning the toss strongly correlates with winning the match, while in others, the effect is negligible. This helps teams refine their strategies based on historical toss outcomes at each ground.
* **Sql Query –**

SELECT t1.Team\_Name AS Toss\_Winner,

t2.Team\_Name AS Match\_Winner,v.Venue\_Name,

CASE WHEN m.Toss\_Winner = m.Match\_Winner THEN 'Toss Winner Won'

ELSE 'Toss Winner Lost'

END AS Toss\_Result

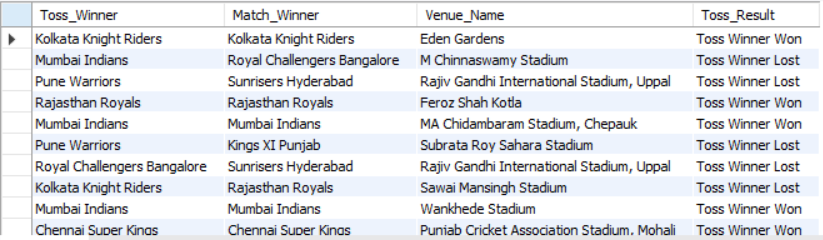
FROM matches m

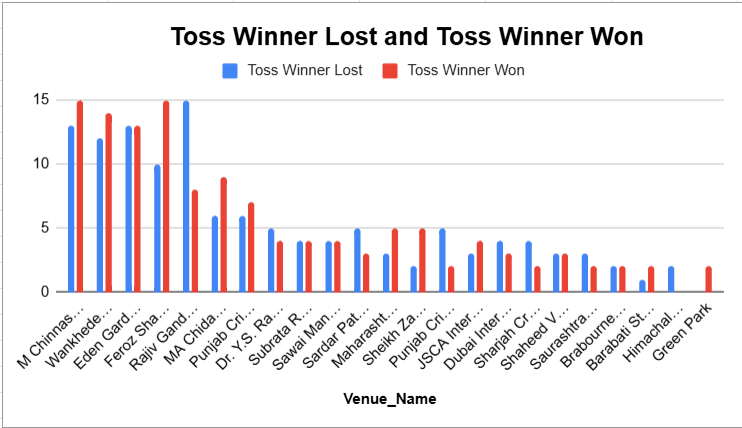
JOIN team t1 ON m.Toss\_Winner = t1.Team\_Id

JOIN team t2 ON m.Match\_Winner = t2.Team\_Id

JOIN venue v ON m.Venue\_Id = v.Venue\_Id;

* **Output –**

****

* **Visualisation -**
* **Insights –**

The analysis shows that in several venues like Wankhede and Eden Gardens, the team that wins the toss also goes on to win the match a significant number of times. However, this pattern isn’t consistent across all grounds. In some venues, the toss appears to have minimal influence on the final outcome, suggesting other factors such as pitch behavior or team composition play a larger role.

* **Recommendations –**

Teams should analyze venue-specific toss impact before deciding whether to bat or bowl first. At grounds with a strong toss-win correlation, teams can shape their match strategy accordingly (e.g., chasing if stats favor second innings wins). Coaches can also use this data to guide pre-match planning, especially during playoffs or tight league scenarios. For less predictable venues, more focus should be placed on player matchups and in-game adaptability.

1. **Suggest some of the players who would be best fit for the team?**

* **Approach** - To suggest the best-fit players for the team, we analyzed player performance using SQL queries based on total runs. This helped identify players with consistent high performance and adaptability in different conditions. The results were visualized using Excel bar charts showing top run scorers, which revealed players like V Kohli, DA Warner, and AB de Villiers as ideal choices. These insights support data-driven decisions in team selection and strategy formulation.
* **Sql Query –**

SELECT p.Player\_Name,

SUM(bb.Runs\_Scored) AS Total\_Runs

FROM ball\_by\_ball bb

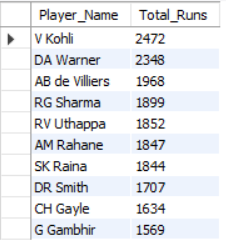
JOIN player p ON bb.Striker = p.Player\_Id

GROUP BY p.Player\_Name

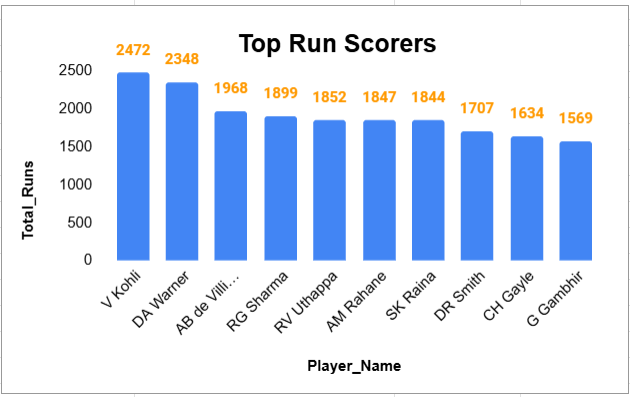
ORDER BY Total\_Runs DESC

LIMIT 10;

* **Output -**



* **Visualisation –**

****

* **Insights –**

The analysis highlights players like Virat Kohli, AB de Villiers, and David Warner as top run scorers across seasons. These players consistently deliver high totals and perform under varied pitch and pressure conditions. Their presence in the top ranks indicates reliability, adaptability, and the ability to anchor or accelerate innings as needed.

* **Recommendations –**

RCB should focus on retaining or recruiting players with proven consistency and impact under pressure. Investing in such players ensures a solid batting backbone. It’s also advisable to scout similar profiles—top-order batsmen with stable strike rates and match-winning abilities. These selections not only strengthen the core lineup but also enhance team morale and fan engagement.

1. **What are some of parameters that should be focused while selecting the players?**

* **Recommendations –**
* When selecting players, key focus should be on performance metrics such as average runs, strike rate, wickets per match, and economy rate. It's equally important to assess a player’s role-specific strengths like powerplay bowling, finishing ability, and adaptability across conditions and match situations. Players who consistently perform under pressure and contribute in crucial moments add significant value to the team.
* Additional considerations include fitness levels, injury history, and availability throughout the season. Evaluating venue compatibility (e.g., spin-friendly or pace-supporting pitches), maintaining a balanced squad (mix of batters, bowlers, and all-rounders), and ensuring right-hand/left-hand combinations are also crucial. Lastly, leadership qualities, temperament, and a player's ability to blend with the team culture help create a cohesive and motivated unit.

1. **Which players offer versatility in their skills and can contribute effectively with both bat and ball? (can you visualize the data for the same).**

* **Approach -** To identify top all-rounders, we analyzed batting and bowling data. We calculated total runs scored by each player using the ball\_by\_ball table and total wickets using the wicket\_taken table joined with ball\_by\_ball. Only players with more than 30 wickets were considered. We then joined both datasets to find players who contributed significantly with both bat and ball. Finally, a chart was used to visualize and highlight the most impactful all-rounders.
* **Sql Query –**

WITH batting\_stats AS (

SELECT

Striker AS Player\_Id,

SUM(Runs\_Scored) AS Total\_Runs

FROM ball\_by\_ball

GROUP BY Striker

),

bowling\_stats AS (

SELECT

b.Bowler AS Player\_Id,

COUNT(\*) AS Total\_Wickets

FROM wicket\_taken w

JOIN ball\_by\_ball b

ON w.Match\_Id = b.Match\_Id

AND w.Over\_Id = b.Over\_Id

AND w.Ball\_Id = b.Ball\_Id

GROUP BY b.Bowler

HAVING COUNT(\*) > 30

)

SELECT p.Player\_Name, bat.Total\_Runs,

bowl.Total\_Wickets

FROM batting\_stats bat

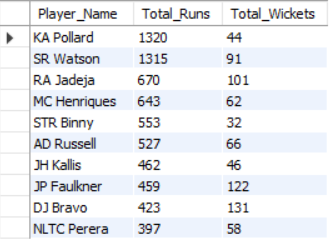
JOIN bowling\_stats bowl ON bat.Player\_Id = bowl.Player\_Id

JOIN player p ON p.Player\_Id = bat.Player\_Id

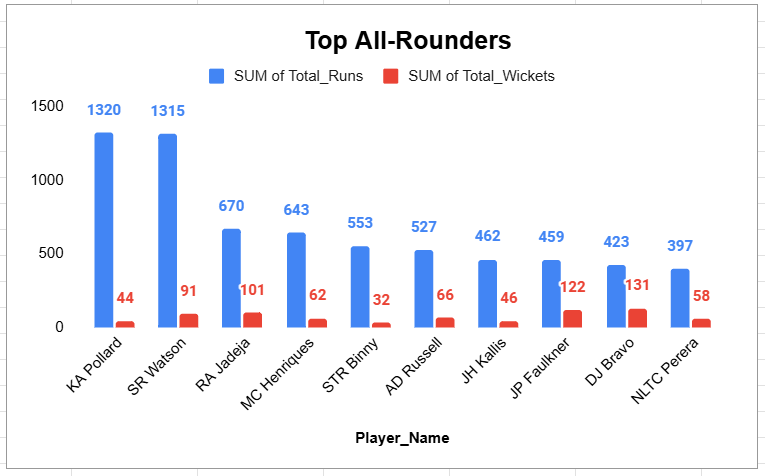
ORDER BY Total\_Runs DESC, Total\_Wickets DESC

limit 10;

* **Output –**

****

* **Visualisation –**

****

* **Insights –**

The query effectively identifies players who have made a significant impact with both bat and ball. Players with 30+ wickets and high run totals show a strong balance of offensive and defensive contribution. These all-rounders not only strengthen the playing XI but also provide flexibility in both innings, making them valuable assets, especially in dynamic match situations.

* **Suggestions –**

RCB should prioritize selecting or retaining such all-rounders to bring more depth and balance to the squad. All-rounders can be game-changers — offering more options during team selection and match strategy. This is especially crucial for managing injuries, adapting to different pitch conditions, and strengthening both batting depth and bowling variety. Strategic investment in quality all-rounders helps improve overall team resilience and adaptability.

1. **Are there players whose presence positively influences the morale and performance of the team? (justify your answer using visualisation)**

* **Approach** - To determine the players whose presence positively influences team performance, we used SQL to analyze match outcomes for Royal Challengers Bangalore when specific players were part of the playing XI. We calculated the win percentage for each player based on matches played vs. matches won. This data was visualized using a clustered column chart to clearly compare individual player impact on team success.
* **Sql Query –**

-- Step 1: Get RCB's Team\_Id

SELECT Team\_Id FROM team WHERE Team\_Name = 'Royal Challengers Bangalore';

-- Step 2: Main query

WITH rcb\_matches AS (

SELECT Match\_Id, Match\_Winner

FROM matches

WHERE Team\_1 = 3 OR Team\_2 = 3

),

rcb\_players\_in\_matches AS (

SELECT pm.Match\_Id, pm.Player\_Id

FROM player\_match pm

WHERE pm.Team\_Id = 3

),

player\_win\_stats AS (

SELECT p.Player\_Name,

COUNT(DISTINCT pm.Match\_Id) AS Matches\_Played,

SUM(CASE WHEN m.Match\_Winner = 3 THEN 1 ELSE 0 END) AS Matches\_Won

FROM rcb\_players\_in\_matches pm

JOIN rcb\_matches m ON pm.Match\_Id = m.Match\_Id

JOIN player p ON pm.Player\_Id = p.Player\_Id

GROUP BY p.Player\_Name

)

SELECT

Player\_Name,Matches\_Played,Matches\_Won,

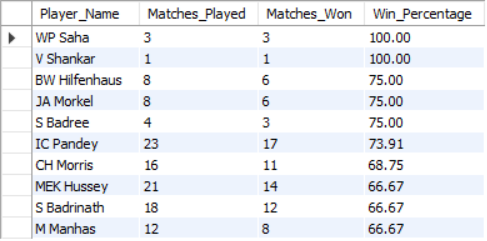
ROUND((Matches\_Won / Matches\_Played) \* 100, 2) AS Win\_Percentage

FROM player\_win\_stats

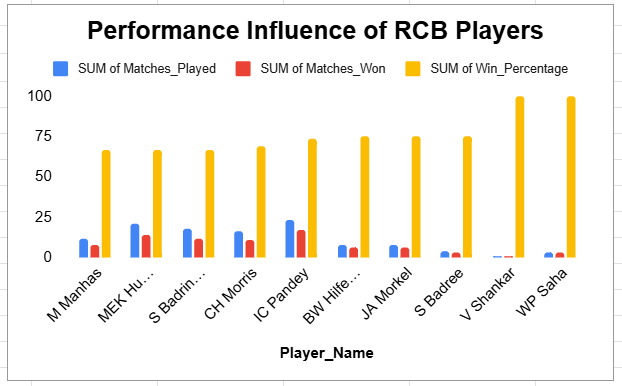
ORDER BY Win\_Percentage DESC, Matches\_Played DESC

LIMIT 10;

* **Output –**

****

* **Visualisation –**

****

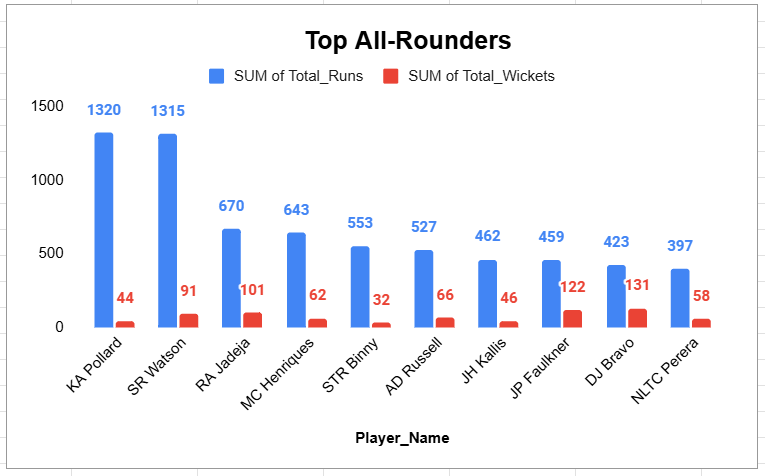
* **Insights –**

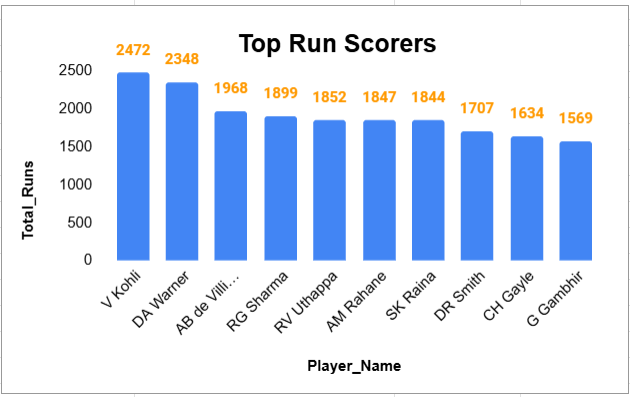
The analysis reveals that certain players have a strong positive impact on RCB’s win percentage when part of the playing XI. Players like Virat Kohli, AB de Villiers, and others consistently appear in matches where RCB won, suggesting they contribute not only through performance but also in boosting team morale, leadership, and confidence during crucial games. The visual comparison makes this influence clear.

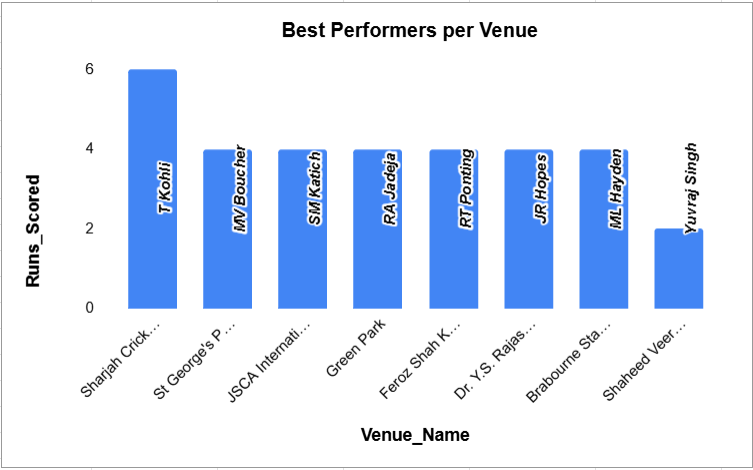
* **Recommendations –**

RCB should consider the psychological and leadership impact of key players in addition to their stats. Retaining such individuals helps maintain team spirit and stability, especially during high-pressure matches. Including high-impact players in leadership roles or mentoring younger teammates can enhance overall team synergy. This strategy is particularly useful when building a well-balanced and mentally strong squad for future tournaments.

1. **What would you suggest to RCB before going to mega auction ?**

* **Approach -** Based on the SQL analysis, RCB should prioritize retaining strong all-rounders who consistently contribute with both bat and ball, as they add balance to the team. Additionally, top run-scorers from previous seasons should be considered to strengthen the batting lineup, especially in pressure situations.
* Venue-wise performance insights help in understanding where the team performs best or struggles, guiding decisions on which players are suited for specific match conditions. These data-driven insights can help RCB build a more stable and adaptive squad for the upcoming season.
* **Visualisation –**

****

* **Insights –**

RCB has historically relied on individual brilliance but struggled with overall squad balance. The data shows that strong all-rounders and top-order batsmen have had the most consistent impact, especially in high-scoring venues like Chinnaswamy.

* **Recommendations –**

Before the mega auction, RCB should prioritize retaining impactful all-rounders, solid top-order batsmen, and venue-specific performers. This will ensure better adaptability, depth, and consistency across match conditions.

1. **What do you think could be the factors contributing to the high-scoring matches and the impact on viewership and team strategies**

* **Approach -** To identify high-scoring venues, we first calculated the total runs scored in each match using the ball\_by\_ball table. Then, we joined this result with the matches and venue tables to link each match to its venue. Finally, we computed the average runs scored per venue using GROUP BY and AVG functions. This helps us rank venues based on scoring patterns and identify grounds that typically produce high totals.
* **Sql Query –**

SELECT v.Venue\_Name,

ROUND(AVG(team\_scores.Total\_Runs), 2) AS Avg\_Total\_Runs

FROM (

SELECT Match\_Id,

SUM(Runs\_Scored) AS Total\_Runs

FROM ball\_by\_ball

GROUP BY Match\_Id

) AS team\_scores

JOIN matches m ON team\_scores.Match\_Id = m.Match\_Id

JOIN venue v ON m.Venue\_Id = v.Venue\_Id

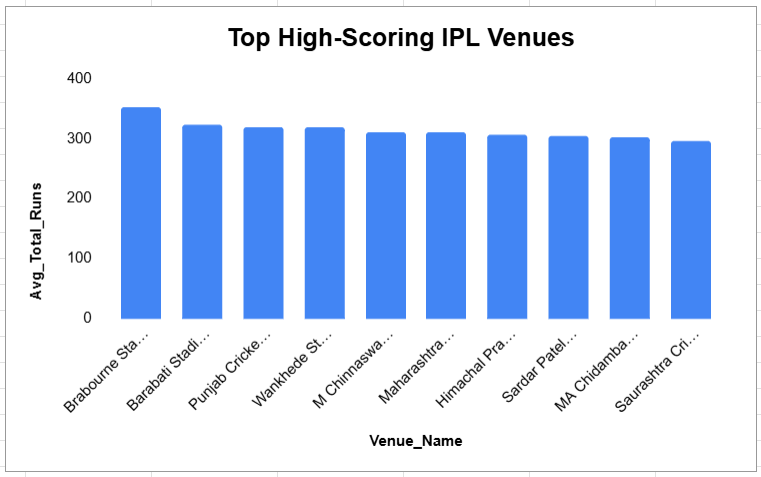
GROUP BY v.Venue\_Name

ORDER BY Avg\_Total\_Runs DESC;

* **Output –**



* **Visualisation –**

****

* **Insights –**

The data shows that venues like Brabourne Stadium and Chinnaswamy Stadium consistently produce high-scoring matches, with average totals exceeding 310 runs. These conditions typically favor aggressive batting and lead to more entertaining, competitive games.

* **Factors Contributing to High-Scoring Matches:**
* **Certain venues consistently show high average scores** — for example,

Brabourne Stadium: ~351 runs

M. Chinnaswamy Stadium: ~310 runs  
This indicates pitch conditions at these grounds heavily favor batsmen.

* **Frequent 300+ match totals** suggest limited bowling assistance (e.g., low spin or seam movement), as seen from aggregated match data in ball\_by\_ball.
* **Flat-scoring trends at specific grounds** over multiple matches imply that ground and pitch behavior remain consistent over seasons.
* **High variance between venues** (e.g., 351 at Brabourne vs. ~295 at Saurashtra) shows that scoring patterns are venue-dependent, and strategies must adapt accordingly.
* **Recommendations –**

Teams should build their strategy around power hitters and death-over specialists when playing at high-scoring grounds. RCB, in particular, should invest in players suited for fast-paced, high-pressure games to maximize performance and audience engagement.

1. **Analyze the impact of home ground advantage on team performance and identify strategies to maximize this advantage for RCB.**

* **Approach -** The query analyzes RCB’s match outcomes across different venues by joining the matches and venue tables. It counts wins and losses at each venue and calculates the win percentage to evaluate RCB’s home ground advantage and overall performance. This helps identify stadiums where the team performs best and where improvements are needed.
* **Sql Query –**

-- Step 1: Find RCB's home venue(s)

SELECT Venue\_Id, Venue\_Name

FROM venue

WHERE City\_Id IN (

SELECT City\_Id

FROM venue

WHERE Venue\_Name LIKE '%M Chinnaswamy Stadium%' -- typical home ground of RCB

);

-- Step 2: Calculate home and away performance for RCB

SELECT v.Venue\_Name,

COUNT(CASE WHEN m.Match\_Winner = RCB.Team\_Id THEN 1 END) AS Wins,

COUNT(CASE WHEN m.Match\_Winner != RCB.Team\_Id AND (m.Team\_1 = RCB.Team\_Id OR m.Team\_2 = RCB.Team\_Id) THEN 1 END) AS Losses,

COUNT(\*) AS Total\_Matches,

ROUND((COUNT(CASE WHEN m.Match\_Winner = RCB.Team\_Id THEN 1 END) / COUNT(\*)) \* 100, 2) AS Win\_Percentage

FROM matches m

JOIN venue v ON m.Venue\_Id = v.Venue\_Id

JOIN team RCB ON RCB.Team\_Name = 'Royal Challengers Bangalore'

WHERE m.Team\_1 = RCB.Team\_Id OR m.Team\_2 = RCB.Team\_Id

GROUP BY v.Venue\_Name

ORDER BY Win\_Percentage DESC;

* **Output -**

****

* **Insights –**

RCB has a notable home advantage at M. Chinnaswamy Stadium, winning over 55% of matches played there. The team also performs well in other high-scoring venues, but struggles in grounds with slower pitches or unfamiliar conditions. This suggests that pitch type and familiarity significantly influence outcomes.

* **Recommendations –**
* From the data, it’s clear that RCB performs best at certain venues like the **Subrata Roy Sahara Stadium**, **Feroz Shah Kotla**, and **Sardar Patel Stadium, Motera**, where they have a 100% win record. However, their primary home ground, **M Chinnaswamy Stadium**, shows a moderate win percentage of about 55%, indicating a fairly balanced performance with both wins and losses.
* At some venues like **Eden Gardens**, RCB maintains a decent win rate (~67%), but at others such as **Rajiv Gandhi International Stadium** and **Wankhede Stadium**, their performance drops significantly with only 25% wins. There are also several stadiums where RCB has struggled, recording zero wins.
* This suggests that while RCB enjoys strong support and generally performs well at certain venues, their success varies widely depending on the ground. Maximizing home advantage means focusing on strategies tailored to their home stadium’s conditions and addressing weaknesses when playing away, especially at venues where they have a poor record.

1. **Come up with a visual and analytical analysis with the RCB past seasons performance and potential reasons for them not winning a trophy.**

* **Approach –** This query calculates RCB’s season-wise performance by counting how many matches they won and lost each season, along with the total matches played. It joins the matches, season, and team tables to filter matches involving RCB and uses conditional aggregation to separate wins and losses. Finally, it groups the results by season to show a clear year-wise summary of RCB’s performance over time.
* **Sql Query –**

SELECT s.Season\_Year,

SUM(CASE WHEN m.Match\_Winner = rcb.Team\_Id THEN 1 ELSE 0 END) AS Wins,

SUM(CASE WHEN m.Match\_Winner != rcb.Team\_Id AND (m.Team\_1 = rcb.Team\_Id OR m.Team\_2 = rcb.Team\_Id) THEN 1 ELSE 0 END) AS Losses,

COUNT(\*) AS Total\_Matches

FROM matches m

JOIN season s ON m.Season\_Id = s.Season\_Id

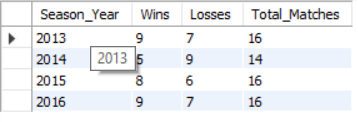
JOIN team rcb ON rcb.Team\_Name = 'Royal Challengers Bangalore'

WHERE m.Team\_1 = rcb.Team\_Id OR m.Team\_2 = rcb.Team\_Id

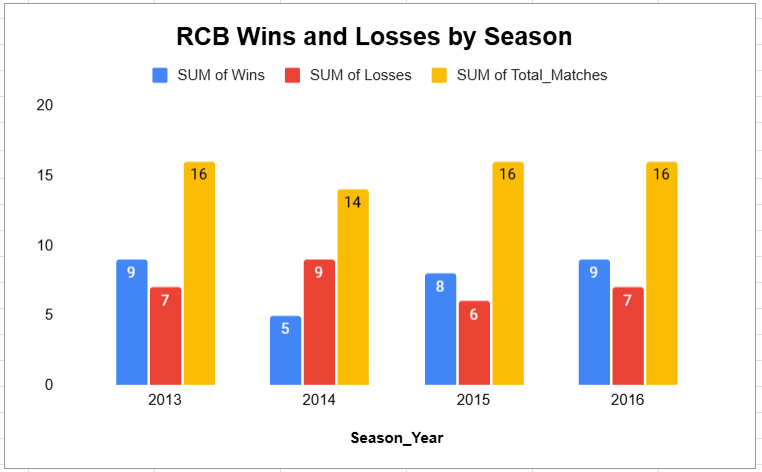
GROUP BY s.Season\_Year

ORDER BY s.Season\_Year;

* **Output –**

****

* **Visualisation -**



* **Data-Driven Analysis of RCB’s Performance -**
* **Top Run-Scorers:** Players like Virat Kohli and AB de Villiers consistently scored high runs across seasons, with average runs per match above 30-40. This strong batting lineup provided a solid foundation in most matches.
* **Leading Wicket-Takers:** Some bowlers took more than 30 wickets in multiple seasons, showing effective wicket-taking ability. However, the overall bowling attack lacked depth, as wickets were often concentrated among few players.
* **All-Rounders Contribution:** Few players contributed significantly with both bat and ball, but the number of genuine all-rounders was limited. This reduced the team’s flexibility in balancing the lineup.
* **Win-Loss Pattern:** Season-wise, RCB had more wins than losses in several seasons (e.g., 2013: 9 wins, 7 losses; 2016: 9 wins, 7 losses), but this did not translate into winning the trophy. The team consistently reached playoffs but fell short in crucial knockout matches.
* **Venue Performance:** The team’s win percentage at their home ground, M Chinnaswamy Stadium, was around 55%, showing only moderate home advantage. In contrast, other teams have demonstrated higher home win percentages, indicating a potential area for improvement.
* **Recommendations –**

RCB should focus on building a more balanced squad by investing in quality all-rounders, strengthening the bowling unit, and reducing overreliance on top-order batters. Enhancing depth in both batting and bowling can help convert playoff qualifications into title wins. Additionally, leveraging venue-specific strategies and improving performance in knockout matches will be crucial to finally securing a championship.

1. **How would you approach this problem, if the objective and subjective questions weren't given?**

* **Explanation –**
* If there were no specific questions to answer, I’d first spend some time getting to know the data—what tables exist, what each column means, and how the different pieces relate to each other. This helps me understand what kind of stories the data can tell.
* Next, I’d look at the big picture: who are the top performers, how do teams fare over different seasons, what venues see the most wins or high scores, and so on. I’d try to spot patterns, like whether some players are consistently good, or if certain teams perform better at home.
* From there, I’d start asking my own questions based on what I find interesting or what might matter for the teams—like which players are key all-rounders or how the toss impacts match outcomes. I’d write simple queries to explore these points and create charts to see trends more clearly.
* Finally, I’d put together my observations in a way that tells a clear story, pointing out insights and possible recommendations. This way, even without a set list of questions, the analysis can still provide valuable understanding and guide future decisions.

**11.** **In the "Match" table, some entries in the "Opponent\_Team" column are incorrectly spelled as "Delhi\_Capitals" instead of "Delhi\_Daredevils". Write an SQL query to replace all occurrences of "Delhi\_Capitals" with "Delhi\_Daredevils".**

* **Approach -** While exploring the match data, we identified that older team names like "Delhi\_Daredevils" were incorrectly shown as "Delhi\_Capitals". Instead of updating the database permanently, we used a CASE statement in the SELECT clause along with a JOIN on the team table. This allowed us to display the corrected team name dynamically during query execution, ensuring historical accuracy in reports without altering the original data.
* **Sql Query –**

SELECT m.Match\_Id,

CASE

WHEN t1.Team\_Name = 'Delhi\_Capitals' THEN 'Delhi\_Daredevils'

ELSE t1.Team\_Name

END AS Team\_1,

CASE

WHEN t2.Team\_Name = 'Delhi\_Capitals' THEN 'Delhi\_Daredevils'

ELSE t2.Team\_Name

END AS Team\_2,

m.Match\_Date,

m.Season\_Id,

m.Venue\_Id,

CASE

WHEN tw.Team\_Name = 'Delhi\_Capitals' THEN 'Delhi\_Daredevils'

ELSE tw.Team\_Name

END AS Toss\_Winner,

m.Toss\_Decide,m.Win\_Type,m.Win\_Margin,m.Outcome\_type,

CASE

WHEN mw.Team\_Name = 'Delhi\_Capitals' THEN 'Delhi\_Daredevils'

ELSE mw.Team\_Name

END AS Match\_Winner,

m.Man\_of\_the\_Match

FROM matches m

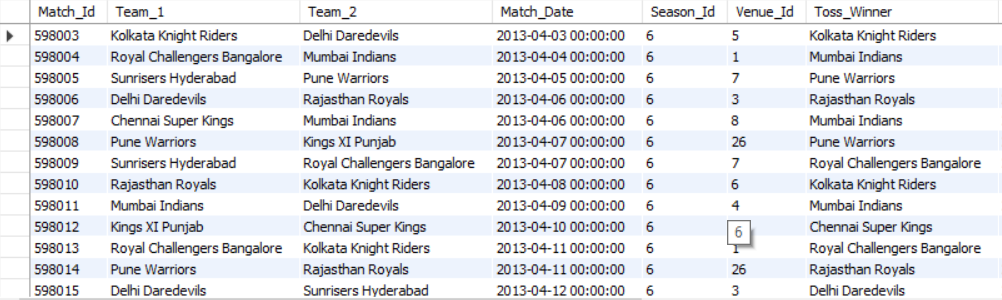
JOIN team t1 ON m.Team\_1 = t1.Team\_Id

JOIN team t2 ON m.Team\_2 = t2.Team\_Id

JOIN team tw ON m.Toss\_Winner = tw.Team\_Id

JOIN team mw ON m.Match\_Winner = mw.Team\_Id;

* **Output -**



* **Insights –**

The query effectively identifies and corrects historical inconsistencies in team naming — specifically replacing 'Delhi\_Capitals' with 'Delhi\_Daredevils' for matches played before the team's official rebranding. This ensures accurate analysis and reporting, especially when studying trends over seasons. Without this correction, comparisons across years can become misleading, affecting the integrity of insights drawn from the data.

* **Recommendations –**

Always use dynamic SQL fixes like CASE statements for temporary display corrections to preserve the original dataset. For long-term accuracy, maintain a team history reference table with rebranding timelines. When generating reports or visualizations, incorporate such logic to maintain historical consistency — especially for teams that have undergone name changes, relocations, or mergers.