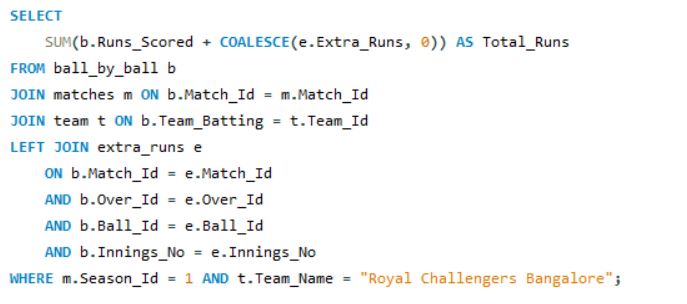
-- Objective Questions –

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

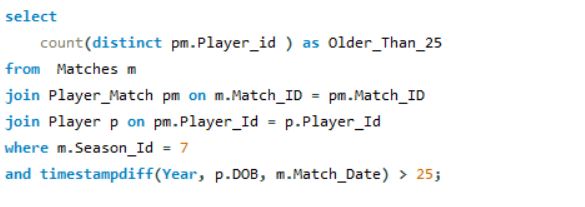
🡪 Match\_Id, Over\_id, Ball\_id, Innings\_No, Team\_Batting, Team\_Bowling , Striker\_Batting\_Position, Striker, Non\_Striker, Bowler, Runs\_Scored

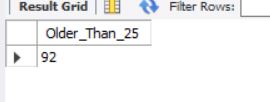
1. What is the total number of runs scored in 1st season by RCB (bonus: also include the extra runs using the extra runs table)



Ball\_by\_ball gives regular runs , extra\_runs gives extras and coalesce ensures Nulles become 0 and filtered by Season\_Id = 1 and Team\_Name = “Royal Challengers Banglaore”

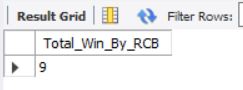
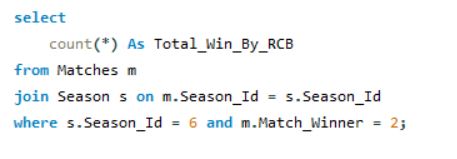
1. How many players were more than the age of 25 during season 2014?





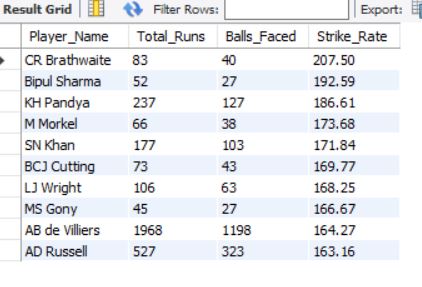
We use this approach because we join player, player\_match, and matches to access each player's DOB and the exact match date from season 7 (which is 2014), then use TIMESTAMPDIFF to calculate age at match time, ensuring accurate filtering of players older than 25 during that season.

1. How many matches did RCB win in 2013?



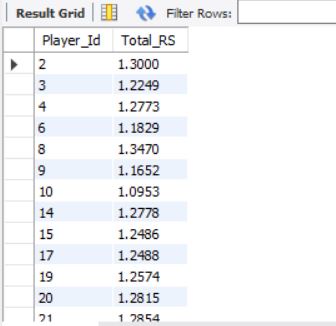
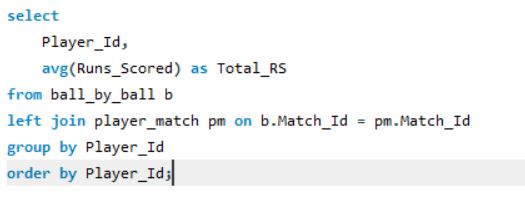
We use this approach because we join matches with season to filter only 2013 matches (Season\_Id = 6), then count the rows where Match\_Winner is RCB’s Team\_Id (2), which gives the total number of matches RCB won in that season.

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



We use this approach because we join ball\_by\_ball, matches, season, and player to get each striker’s performance in the last 4 seasons, then calculate strike rate as (total runs / balls faced) × 100, apply a minimum balls faced filter for reliability, and sort by strike rate to get the top 10 players.

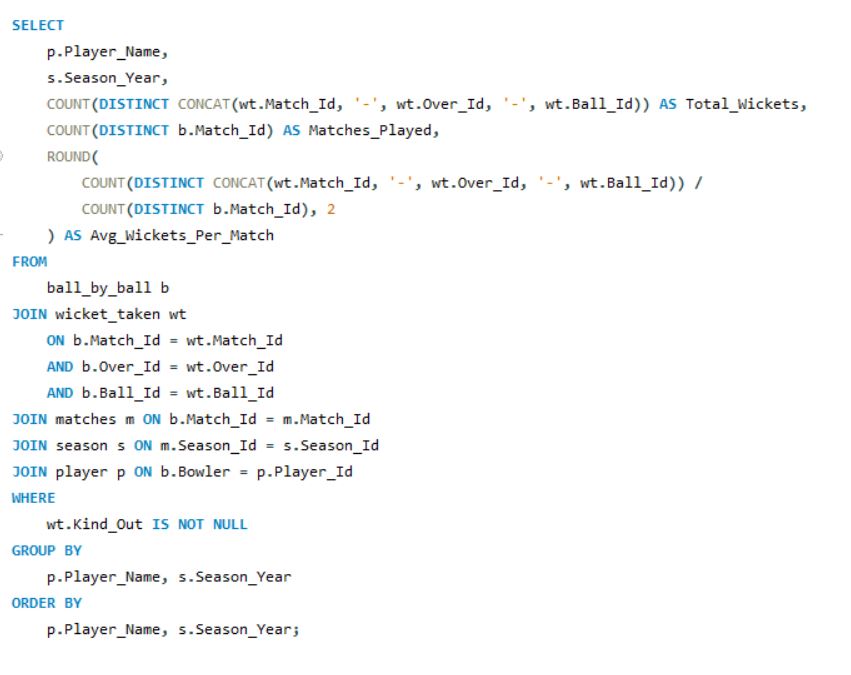
1. What are the average runs scored by each batsman considering all the seasons?



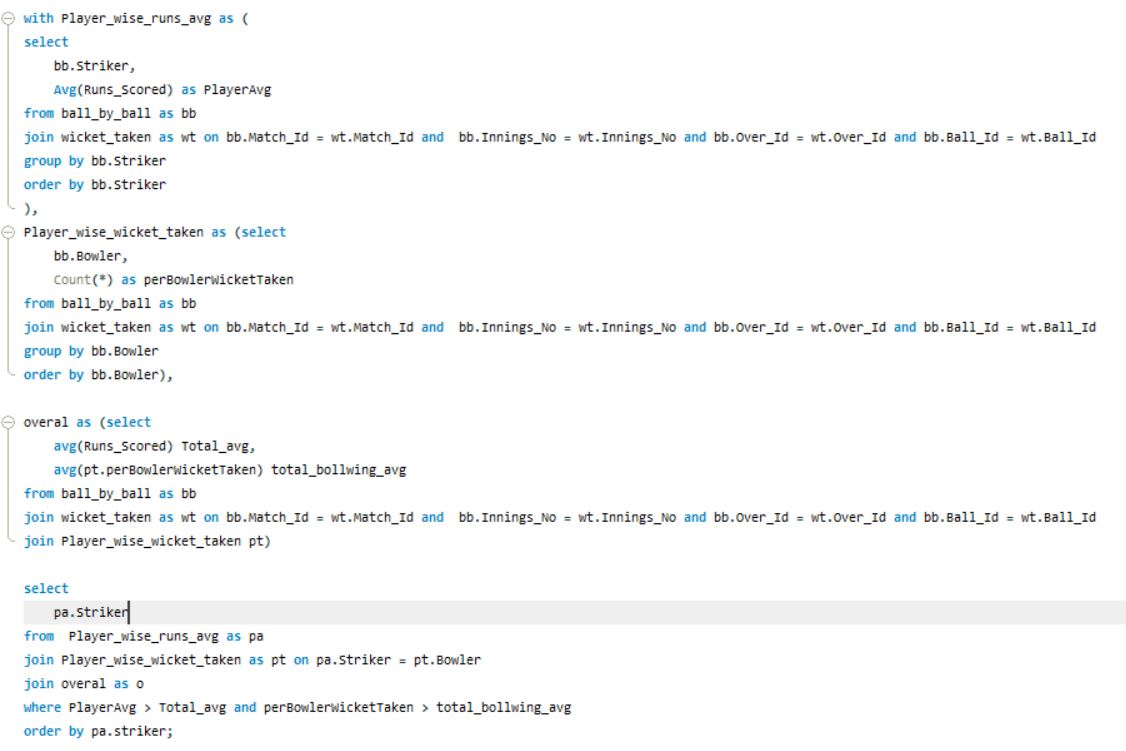
We use this approach because we calculate the average runs scored per ball for each Player\_Id from ball\_by\_ball, grouping by player, which gives us a rough average scoring rate over all seasons, although joining with player\_match here is unnecessary unless we need extra player info.

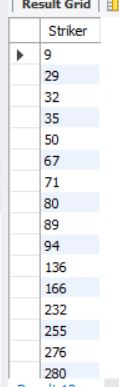
1. What are the average wickets taken by each bowler considering all the seasons?



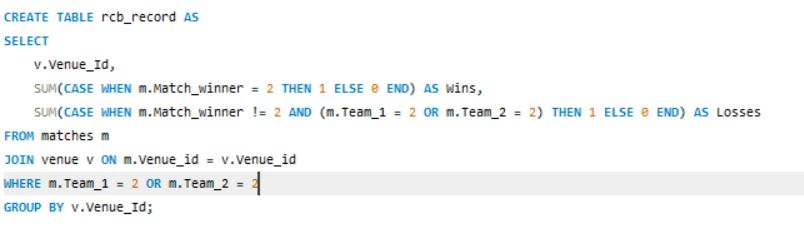


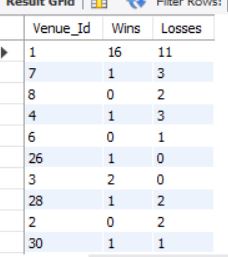
1. List all the players who have average runs scored greater than the overall average and who have taken wickets greater than the overall average:



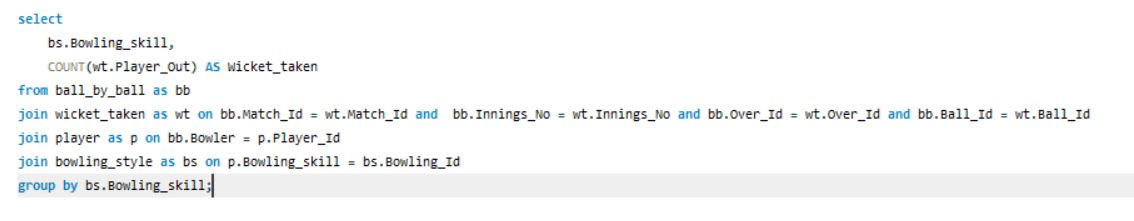
We use this approach because we calculate each player's average runs and total wickets using CTEs, compute the overall average for both metrics, and then filter players who exceed both averages to find true all-rounders who perform better than average with both bat and ball.

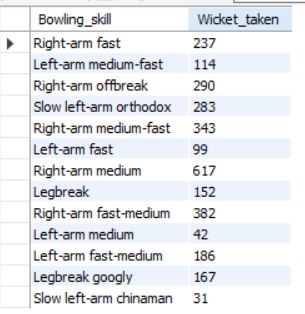
1. Create a table rcb\_record table that shows the wins and losses of RCB in an individual venue.



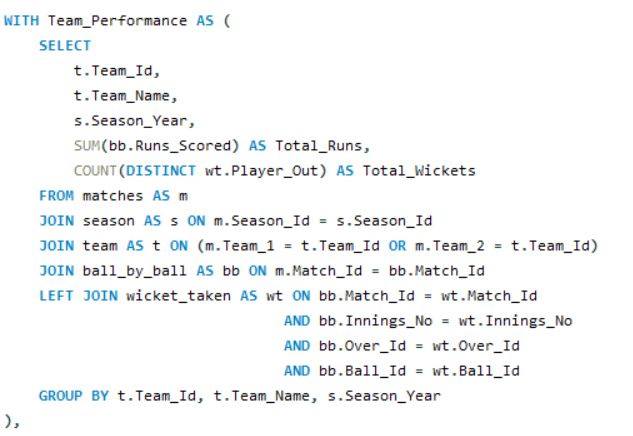
We use this approach because we filter matches where RCB played, then group them by venue and count wins when RCB is the match winner, and losses when RCB played but didn’t win, giving us a venue-wise win-loss record for RCB and storing it in the rcb\_record table.

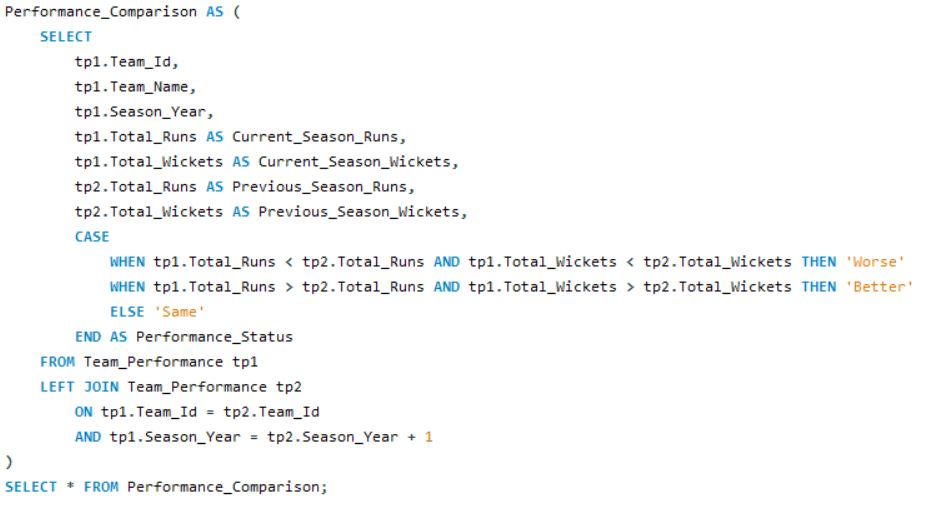
1. What is the impact of bowling style on wickets taken?

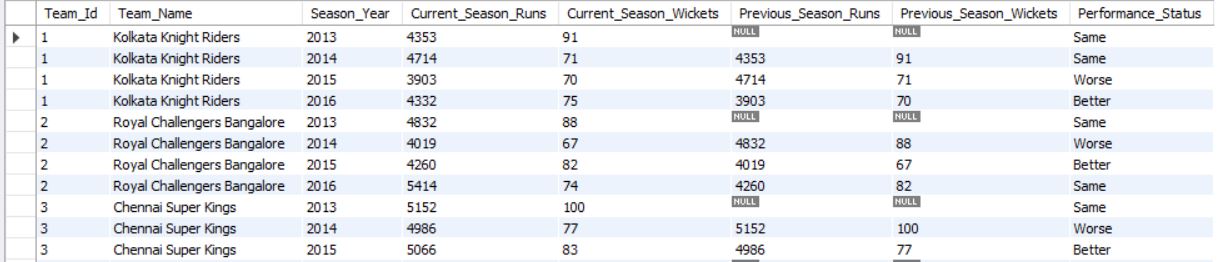


We use this approach because we join ball\_by\_ball, wicket\_taken, player, and bowling\_style to link each dismissal to a bowler’s style, then group by bowling skill and count wickets to analyze which bowling styles lead to more dismissals, showing their impact on performance.

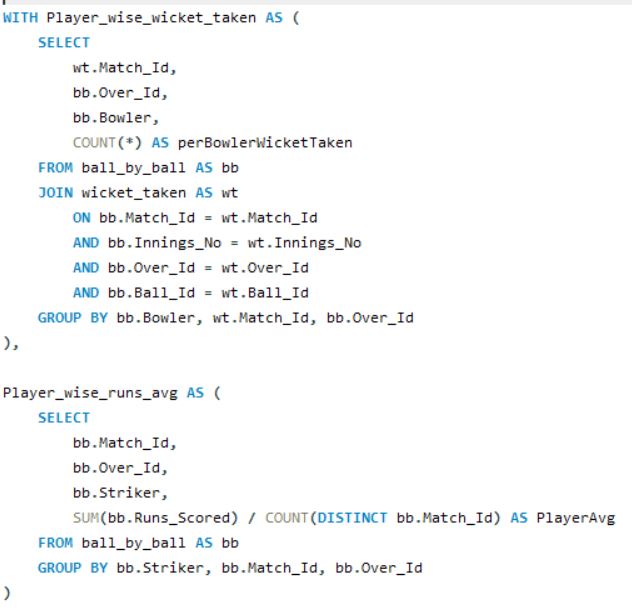
1. Write the SQL query to provide a status of whether the performance of the team is better than the previous year's performance on the basis of the number of runs scored by the team in the season and the number of wickets taken :

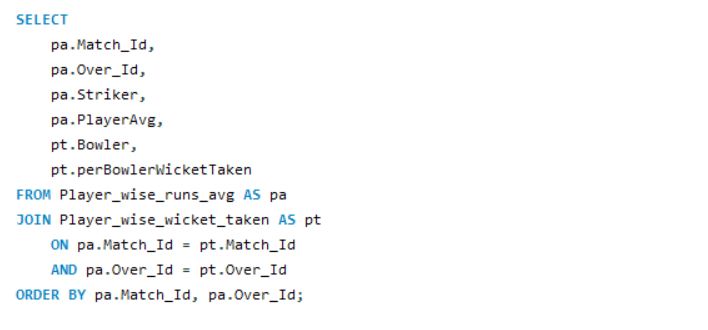






We use this approach because we first calculate each team’s total runs and wickets per season using a CTE, then in a second CTE, we self-join the performance data to compare current season stats with the previous year’s, and use conditional logic to classify performance as ‘Better’, ‘Worse’, or ‘Same’ based on improvements in both runs and wickets.

1. Can you derive more KPIs for the team strategy? 

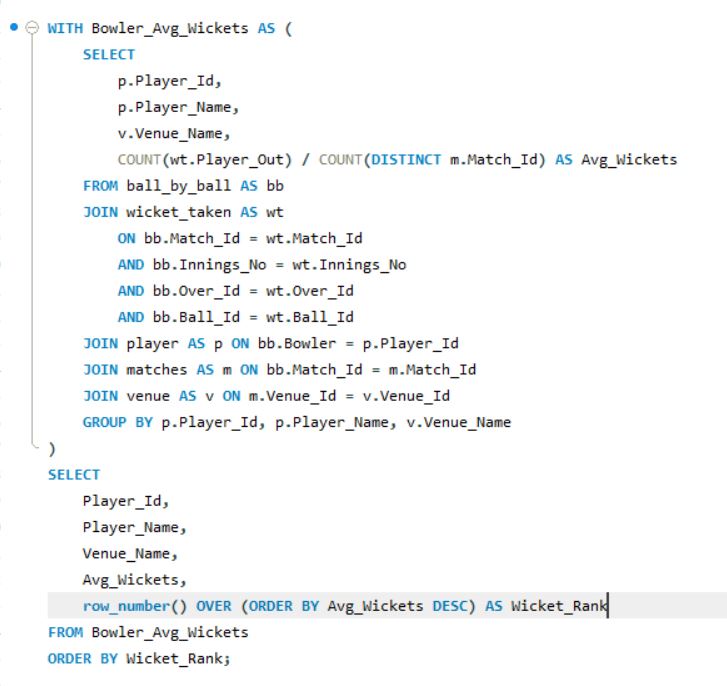




We use this approach because we first calculate each team’s total runs and wickets per season using a CTE, then in a second CTE, we self-join the performance data to compare current season stats with the previous year’s, and use conditional logic to classify performance as ‘Better’, ‘Worse’, or ‘Same’ based on improvements in both runs and wickets.

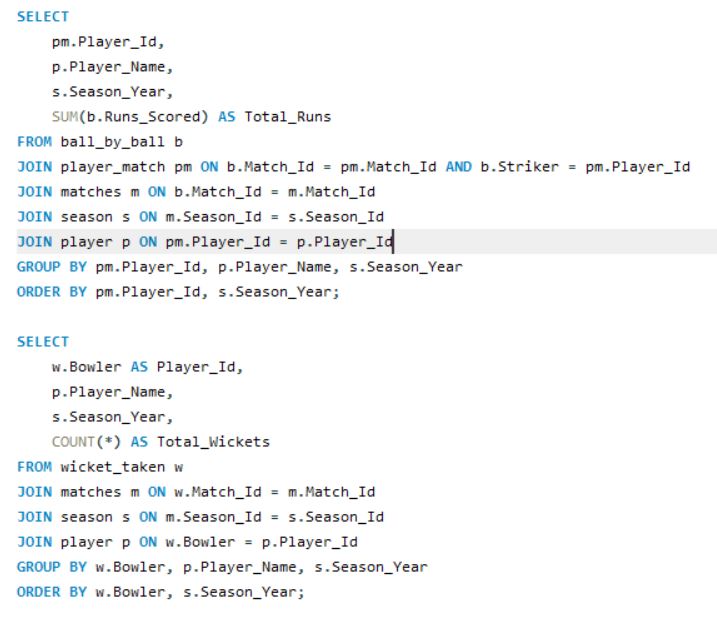
1. Using SQL, write a query to find out the average wickets taken by each bowler in each venue. Also, rank the gender according to the average value.

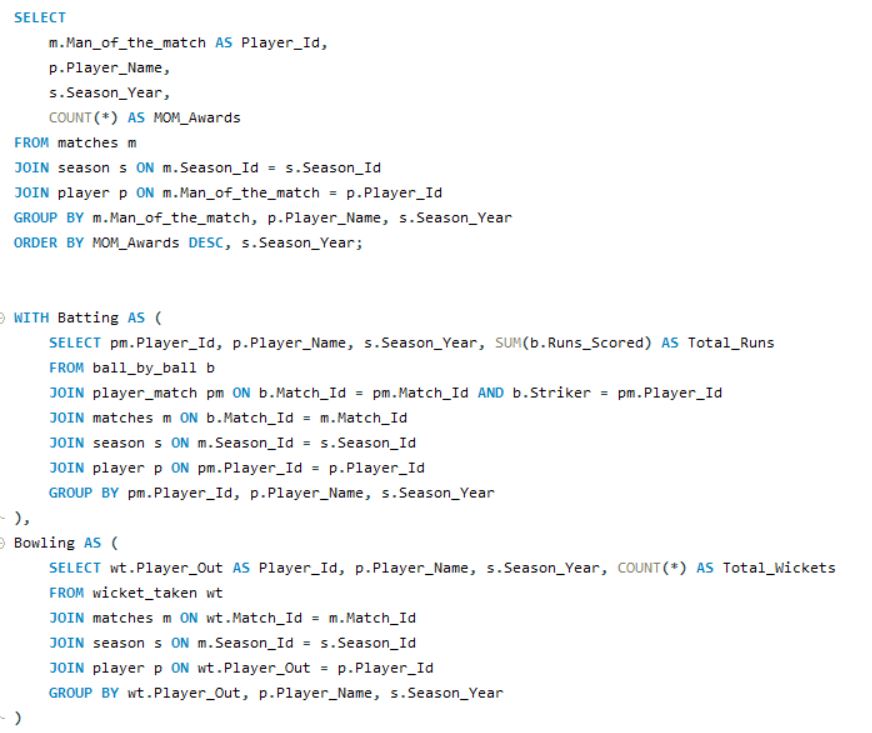
🡪 It links detailed ball-level data (ball\_by\_ball) with wicket events (wicket\_taken) to track exactly when and where a bowler took wickets. It uses the matches and venue tables to understand which venue each match was played in. The performance is grouped by each bowler and each venue, calculating the average wickets taken by a bowler per match at that particular venue. Finally, it ranks these performances to highlight the top wicket-takers by average at any ground.

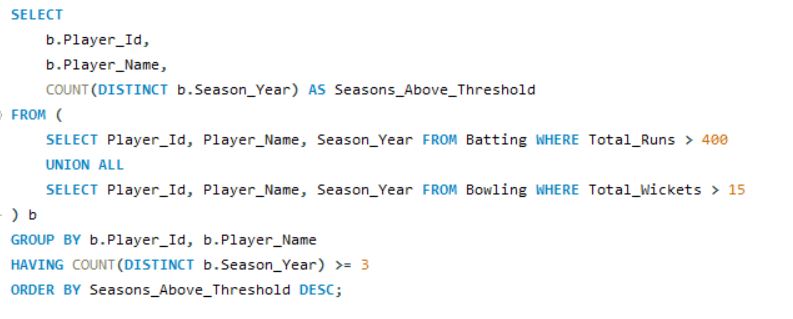




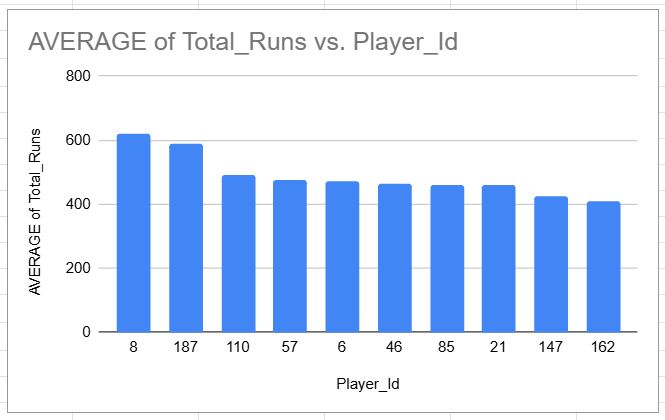
1. Which of the given players have consistently performed well in past seasons? (will you use any visualization to solve the problem)



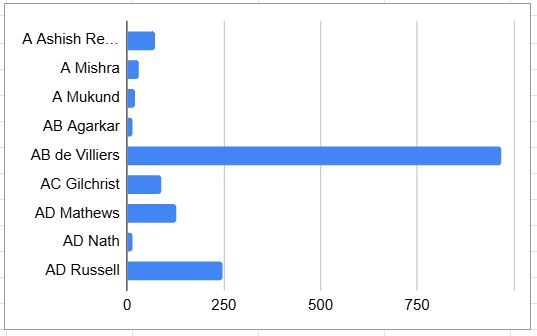


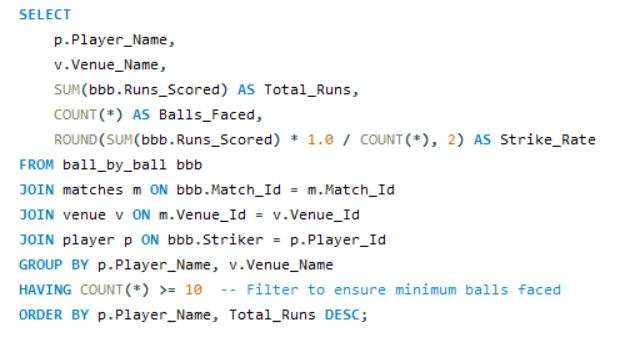


We use this approach to identify consistent top performers by first calculating season-wise total runs and wickets for each player using separate CTEs (Batting and Bowling), then filtering players who scored over 400 runs or took more than 15 wickets in a season. We use UNION ALL to merge both sets and group by player to count how many seasons they crossed the threshold, finally selecting those with 3 or more such seasons to highlight long-term impactful players.



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

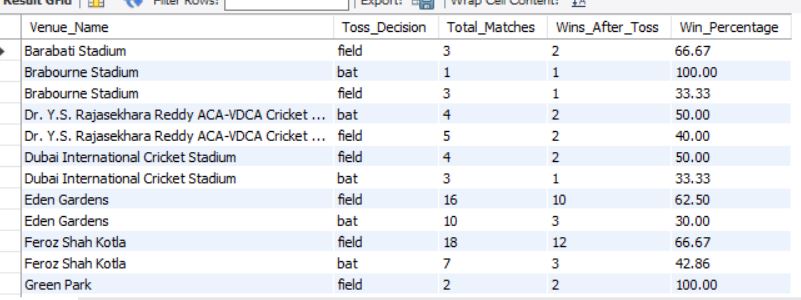
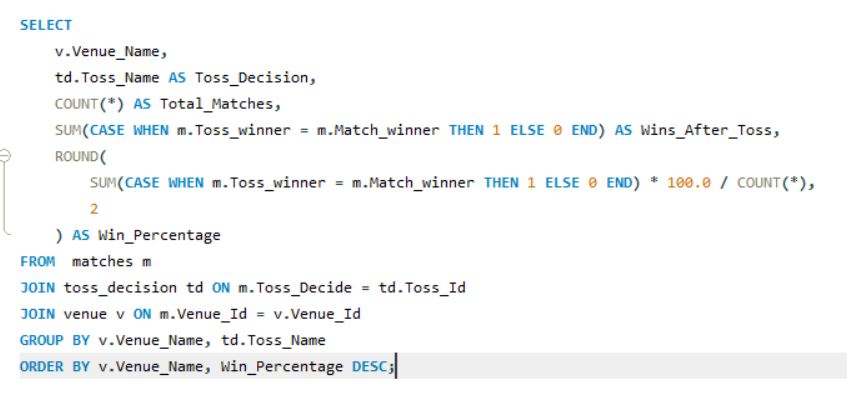




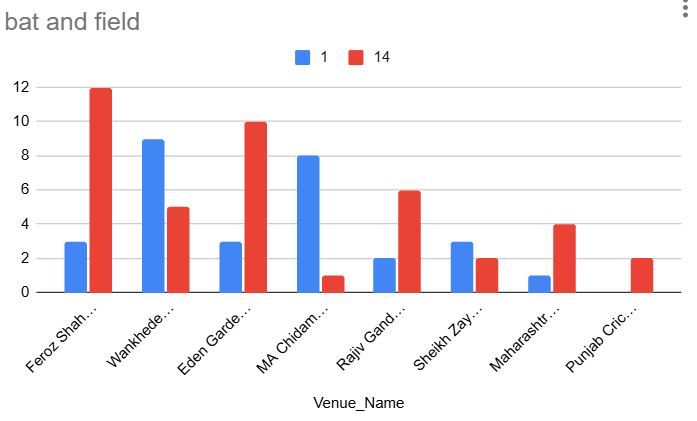
We use this approach to analyze player performance at different venues by grouping data by Player\_Name and Venue\_Name, calculating total runs, balls faced, and strike rate per venue, which helps identify venue-specific strengths. Using a minimum balls filter ensures reliable stats.

-- Subjective Questions—

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

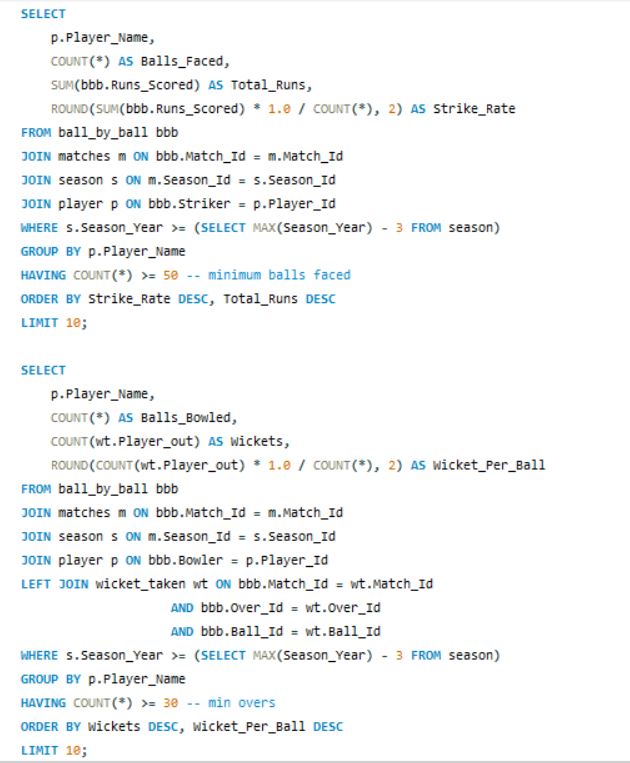


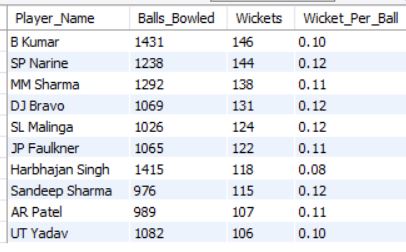
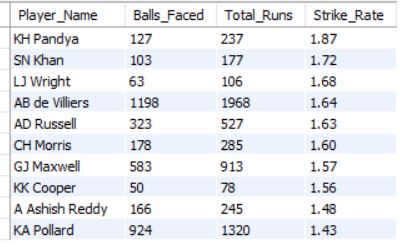
We use this approach to evaluate how toss decisions impact match outcomes by comparing whether the toss winner also won the match, grouped by venue and decision type (bat or field), and calculating the win percentage for each scenario. This helps assess if certain decisions give a strategic advantage, especially in specific venues.



2) Suggest some of the players who would be best fit for the team.

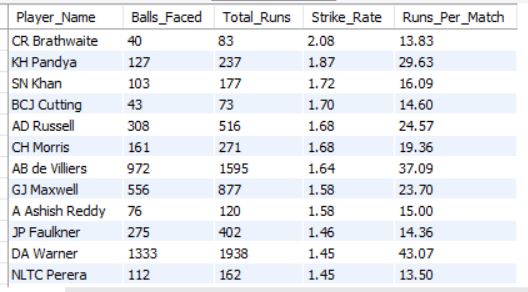
We use this approach to **identify top-performing players** from recent seasons who consistently deliver with bat or ball. By filtering the last 3 seasons and applying minimum thresholds (50 balls faced or 30 balls bowled), we ensure only **active and impactful players** are considered. Sorting by **strike rate and total runs** for batters, and **wickets and wicket per ball** for bowlers helps us shortlist **balanced and high-impact candidates** suitable for team selection.





3) What are some of the parameters that should be focused on while selecting the players?

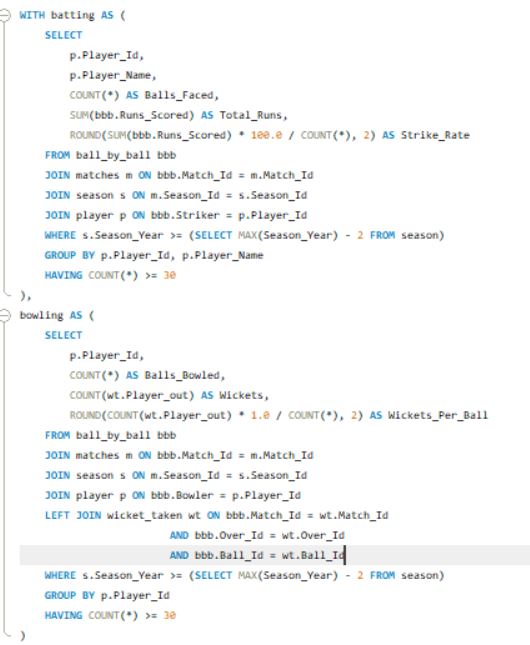


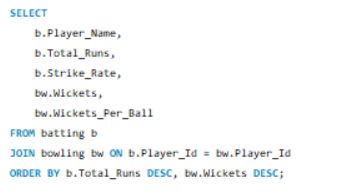


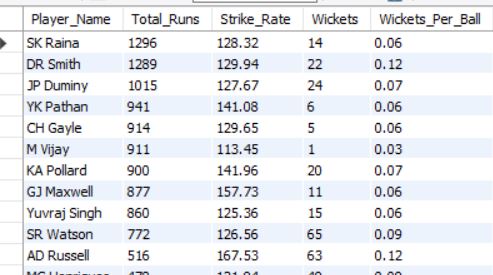
We use this approach to **evaluate recent batting performance** by focusing on key metrics: **strike rate**, **total runs**, and **runs per match**, which together indicate a player's **scoring ability, consistency, and impact per game**. Filtering players with at least 30 balls faced in the **last 2 seasons** ensures we assess only **active and reliable performers**, and sorting by strike rate and runs helps shortlist those most effective in high-pressure T20 environments.

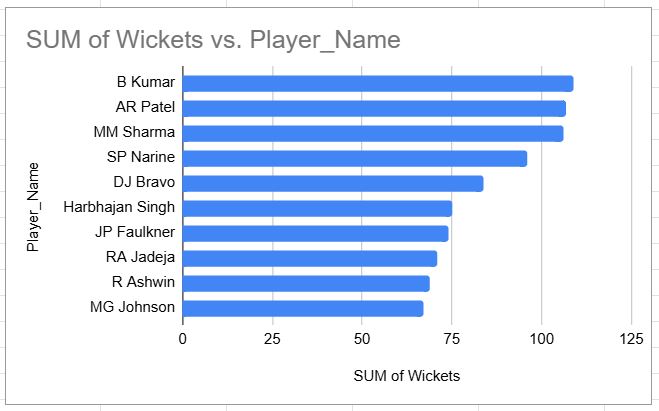
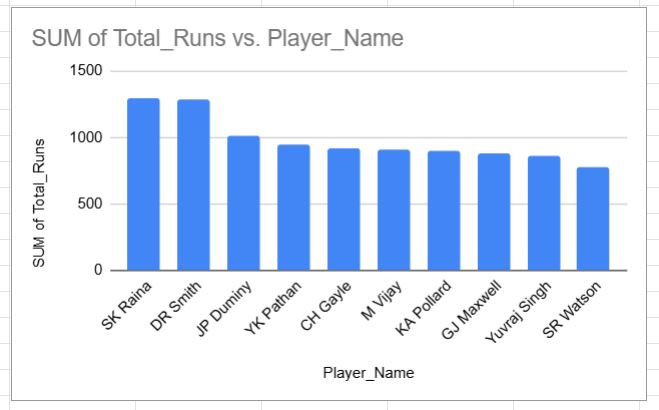
4) Which players offer versatility in their skills and can contribute effectively with both bat and ball? (Can you visualize the data for the same)

We use this approach to **identify all-rounders**—players who significantly contribute in both batting and bowling—by combining metrics like **Total Runs**, **Strike Rate**, **Wickets Taken**, and **Wickets per Ball** from the **last 2 seasons**. Filtering out players with fewer than 30 deliveries ensures we evaluate only those with a meaningful sample size. Joining both skillsets helps shortlist **versatile players** who can add value in multiple roles, making them strategic assets in team selection.









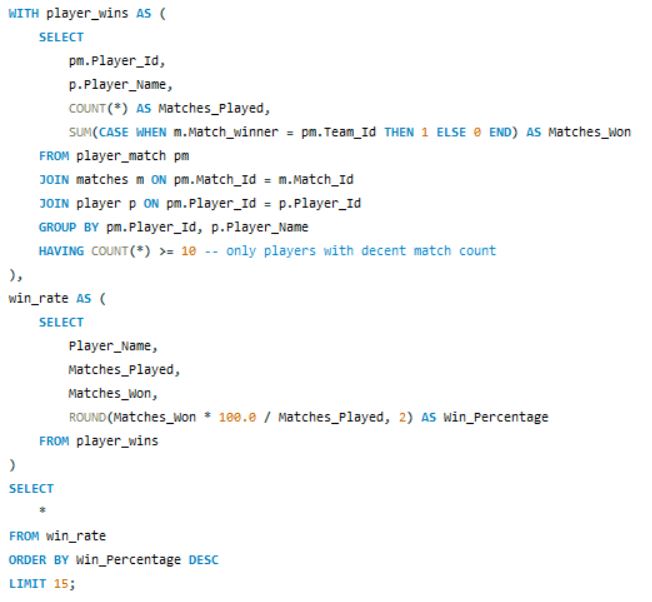
5) Are there players whose presence positively influences the morale and performance of the team? (Justify your answer using visualization)

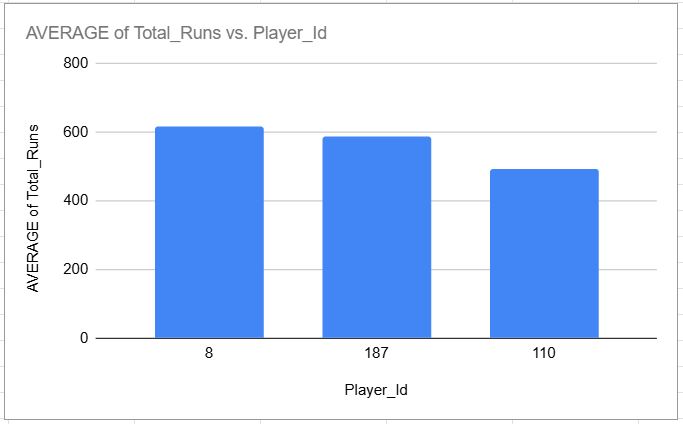
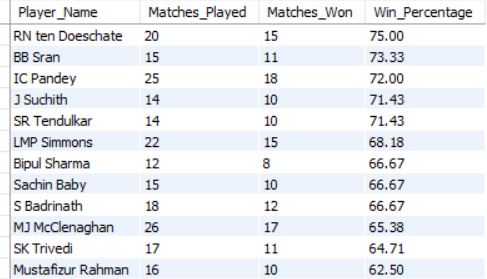
**Player-Team Relationship Insight**:  
By analyzing each player's **matches played vs. matches won** by their team when they were part of the playing XI, we can infer if the player’s presence tends to boost team success.

**Filtering for Relevance**:  
A threshold (HAVING COUNT(\*) >= 10) ensures we consider only players with a **decent sample size**, avoiding skewed results from one-off performers.

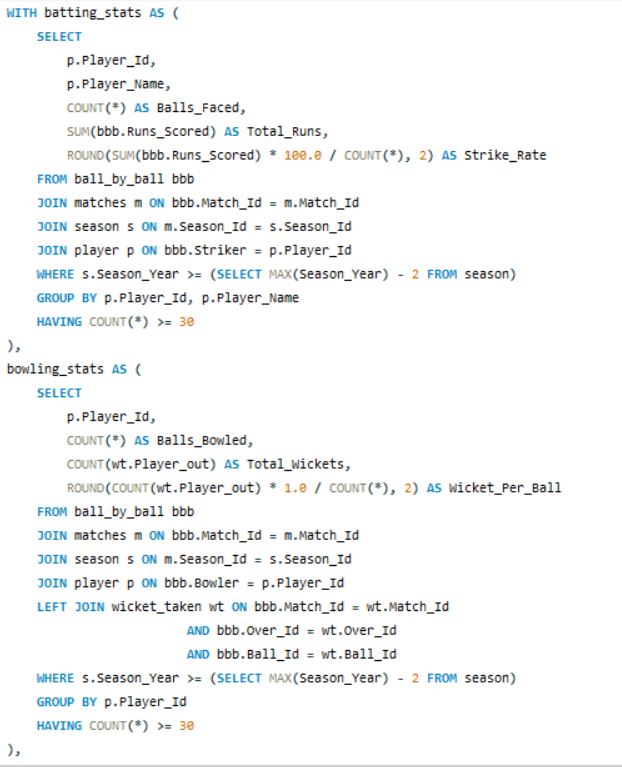
**Win Percentage as Impact Indicator**:  
We calculate Win\_Percentage to measure **how frequently teams win when a specific player is in the lineup**, which may hint at:

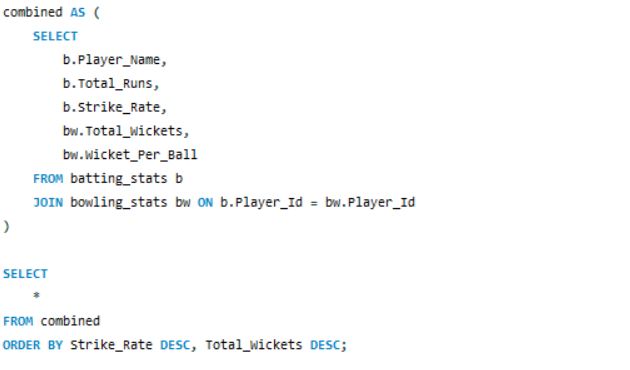
Leadership qualities, Performance under pressure, Influence on team dynamics

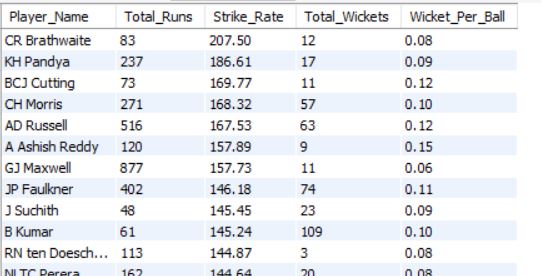




6) What would you suggest to RCB before going to the mega auction?







* Focus on Recent Performance

We only consider data from the last two seasons to ensure we’re evaluating current form, not past reputation.

This gives RCB a clearer picture of who’s in top shape heading into the next season.

* All-Rounder Identification

Two subqueries collect key metrics:

Batting: Total runs, strike rate, and minimum balls faced

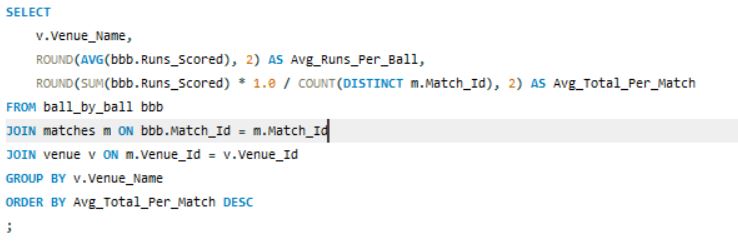
Bowling: Total wickets, wickets per ball, and minimum balls bowled

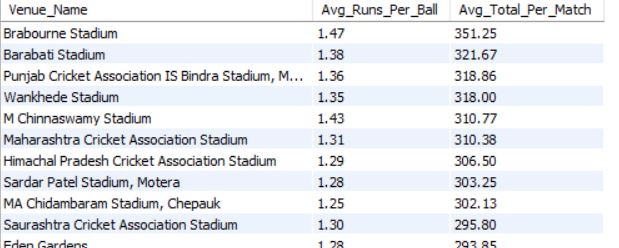
We then join these datasets to find players who excel at both skills, making them versatile assets.

* Filters Ensure Quality

We only include players who have had at least 30 deliveries faced or bowled, which filters out irregulars or fluke performances.

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





**To Understand Venue Behavior**

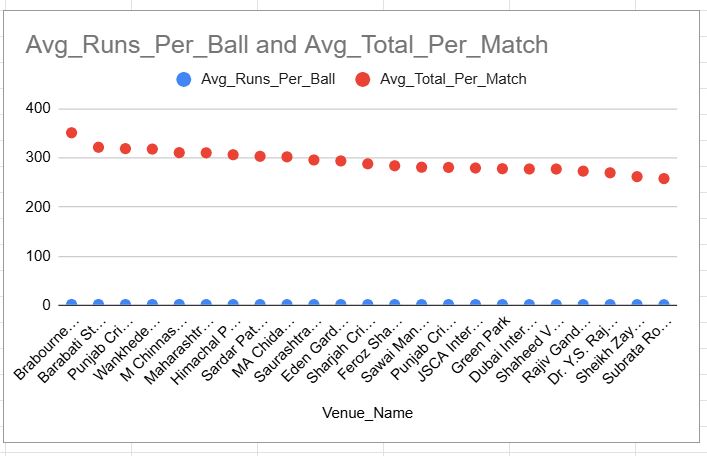
* By calculating:
  + **Average Runs per Ball** — reflects how batting-friendly a pitch is.
  + **Average Total Score per Match** — gives a full-picture view of the scoring environment.
* This helps us detect **batting paradises** like Chinnaswamy or Wankhede, where run-fests are common.

**To Aid Strategic Planning**

* Teams can **adjust lineups** based on venue type:
  + More **power-hitters** and fewer bowlers in high-scoring venues.
  + More **specialist bowlers** and balanced strategy in bowling-friendly ones.
* Pitch-aware strategies improve winning chances.

**To Drive Auction Decisions**

* Teams could target players who **excel in high-pressure, high-scoring scenarios**, especially if they play many matches at such venues.



8) Analyze the impact of home-ground advantage on team performance and identify strategies to maximize this advantage for RCB.

**Home vs. Away Performance Comparison**

* We **classify matches as 'Home' or 'Away'** based on whether RCB played at Chinnaswamy Stadium.
* Then, we calculate:
  + **Matches Played**
  + **Matches Won**
  + **Win Percentage**

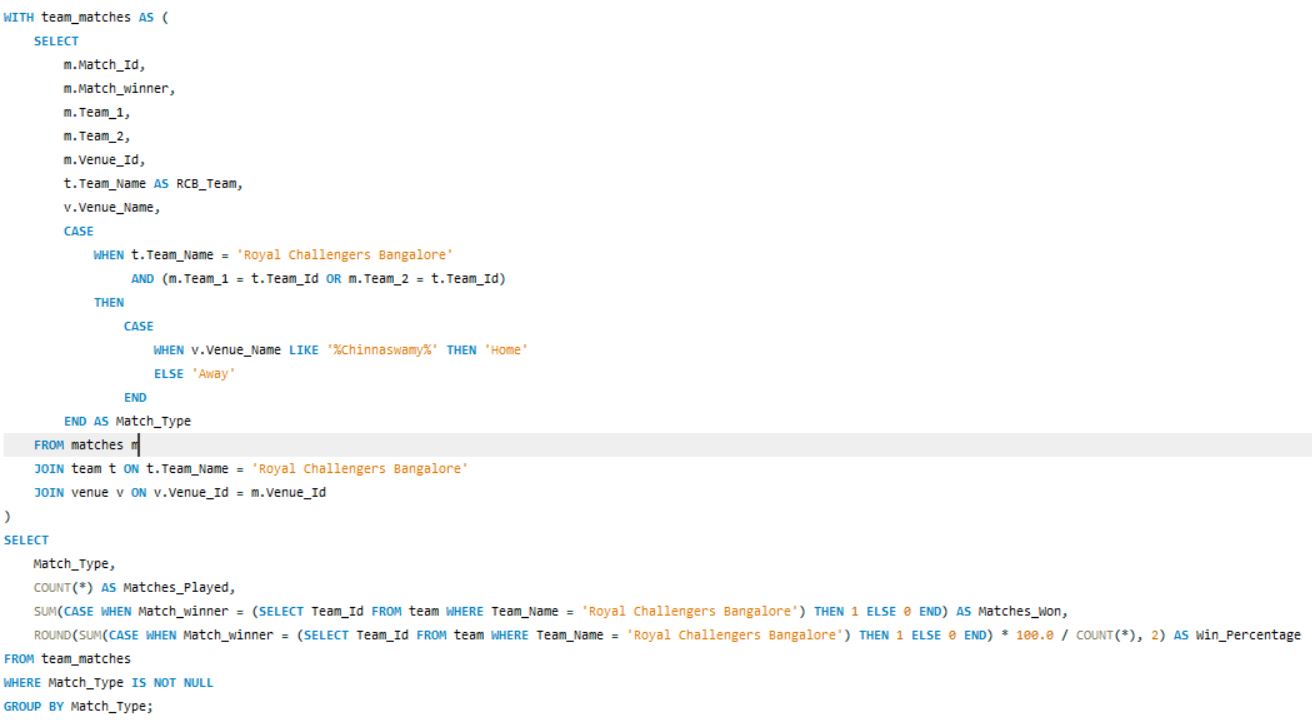
This gives a **clear statistical picture** of whether RCB performs better at home.

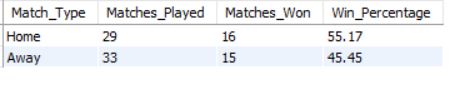
**To Leverage Familiar Conditions**

* If the **Win % is significantly higher at home**, it indicates that RCB benefits from:
  + Familiar **pitch conditions** (batting-friendly surface at Chinnaswamy).
  + **Crowd support** boosting morale.
  + Better **adaptation to local weather and ground dimensions**.

**Strategic Implications:**

* **Team Composition**: Prioritize big hitters and death-over bowlers at Chinnaswamy (high-scoring venue).
* **Toss Decisions**: Use historical data to decide whether to bat or bowl first at home.
* **Training Focus**: Simulate away conditions in practice sessions to improve performance on the road.
* **Player Selection**: Retain/acquire players who historically perform well in home conditions.





9) Come up with a visual and analytical analysis of the RCB's past season's performance and potential reasons for them not winning a trophy.

### 1. ****Season-wise Performance Overview****

* By calculating:
  + **Matches Played**
  + **Matches Won**
  + **Win Percentage**
* We gain a **season-by-season snapshot** of RCB’s consistency (or lack of it).
* It helps answer: Was RCB a consistently strong team or were they sporadically good?

### 2. ****Trend Identification****

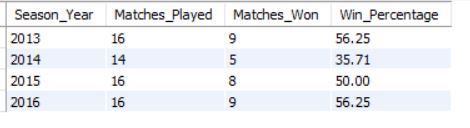
* A **line chart** or **bar graph** of Win % over seasons can visually show:
  + Performance spikes (like playoffs/semi-finals years)
  + Declines or inconsistencies
* This provides **evidence-based insight** rather than just fan sentiment.

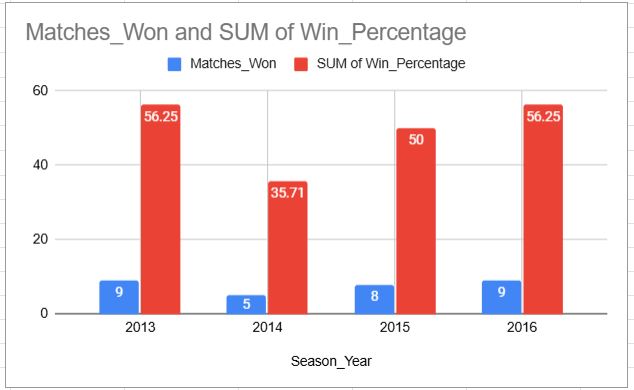
### 3. ****Root Cause Analysis of Trophy Drought****

Using this analysis, we can further explore:

* **If RCB started seasons strong but failed in knockouts**.
* **Impact of key players being out or inconsistent**.
* Whether **high-scoring games were offset by poor bowling**.
* **Lack of finishing strength or middle-order collapses**.







Q10) How would you approach this problem, if the objective and subjective questions weren't given?

### ****Initial Exploration of Team Performance Over Time****

* When analyzing IPL data from scratch, a good starting point is: "How has a specific team performed across seasons?"
* This query gives a **season-wise breakdown** of:
  + Matches played
  + Matches won
  + Win percentage
* It serves as a **baseline** to evaluate the team’s consistency and progress.

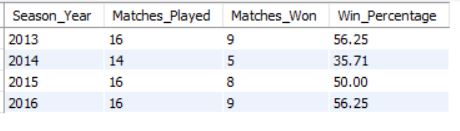
### 2. ****Helps Spot Trends and Patterns****

* Without knowing what to look for initially, this kind of breakdown helps you ask further questions like:
  + Why did the win percentage drop in a particular year?
  + Which players played in high/low-performing seasons?
  + Were there any management or squad changes in those years?

### 3. ****Foundation for Deeper Analysis****

* Once we see **which seasons were successful or poor**, we can:
  + Dive into **player-level performance**
  + Evaluate **venue impacts** or **toss decisions**
  + Analyze **match conditions** in specific years
* So this acts like a **launchpad** for targeted analysis.





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".

UPDATE matches

SET

Team\_1 = CASE WHEN Team\_1 = 'Delhi\_Capitals' THEN 'Delhi\_Daredevils' ELSE Team\_1 END,

Team\_2 = CASE WHEN Team\_2 = 'Delhi\_Capitals' THEN 'Delhi\_Daredevils' ELSE Team\_2 END

WHERE Team\_1 = 'Delhi\_Capitals' OR Team\_2 = 'Delhi\_Capitals';