**Issue 1**

IPL-Analysis



|  |  |
| --- | --- |

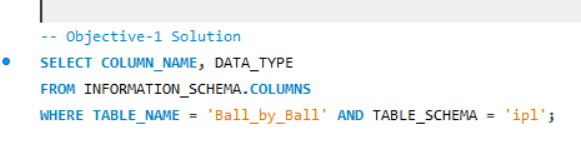
## Objective - Questions

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

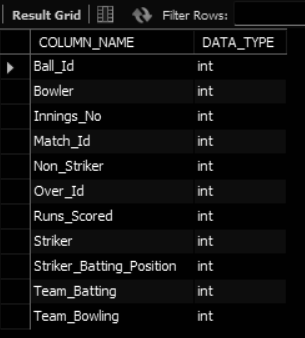
Ans.

Approach-:

* Retrieved the column names and data types from the Ball\_by\_Ball table to understand its structure before starting the analysis.



Result-:

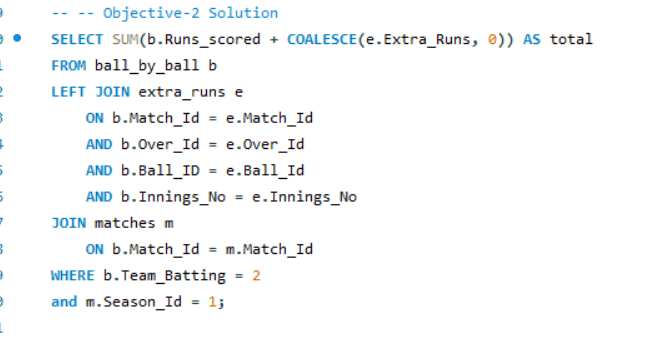


1. **What is the total number of runs scored in 1st season by RCB (bonus: also include the extra runs using the extra runs table).**

Ans.

Approach-:

* Calculated total runs scored by RCB in Season 1 by combining regular and extra runs using joins between ball\_by\_ball, extra\_runs, and matches tables.



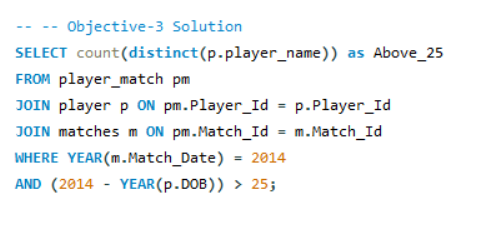
Note-: The Database has no records from season 1.

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

Ans.

Approach-:

* Counted distinct players above age 25 who played in matches in the 2014 season using joins on player\_match, player, and matches tables.



Result-:

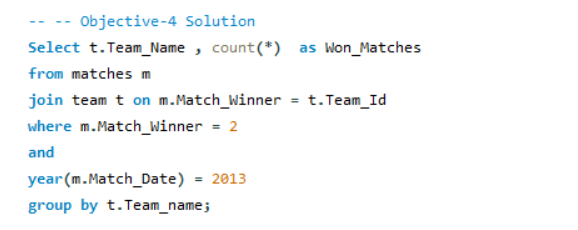


1. **How many matches did RCB win in 2013?**

Ans.

Approach-:

* Retrieved the number of matches won by Team 2 in 2013 by joining matches and team tables and grouping by team name.



Result-:

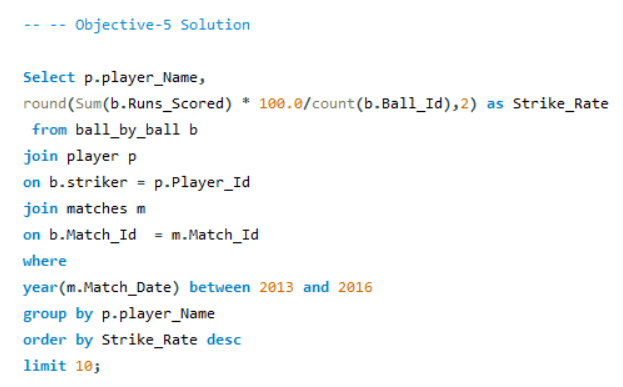


**5. List the top 10 players according to their strike rate in the last 4 seasons ?**

Ans.

Approach-:

* Calculated and ranked top 10 players by strike rate between 2013 and 2016 using data from ball\_by\_ball, player, and matches tables.



Result-:

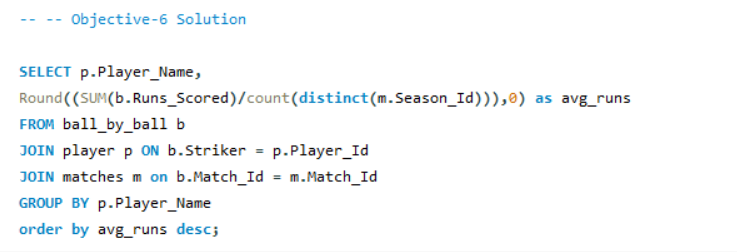


**6. What are the average runs scored by each batsman considering all the seasons?**

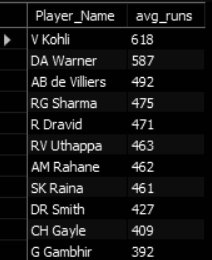
Ans.

Approach-:

* Calculated players' average runs per season by summing runs and dividing by distinct seasons played, using joins on ball\_by\_ball, player, and matches.



Result-:

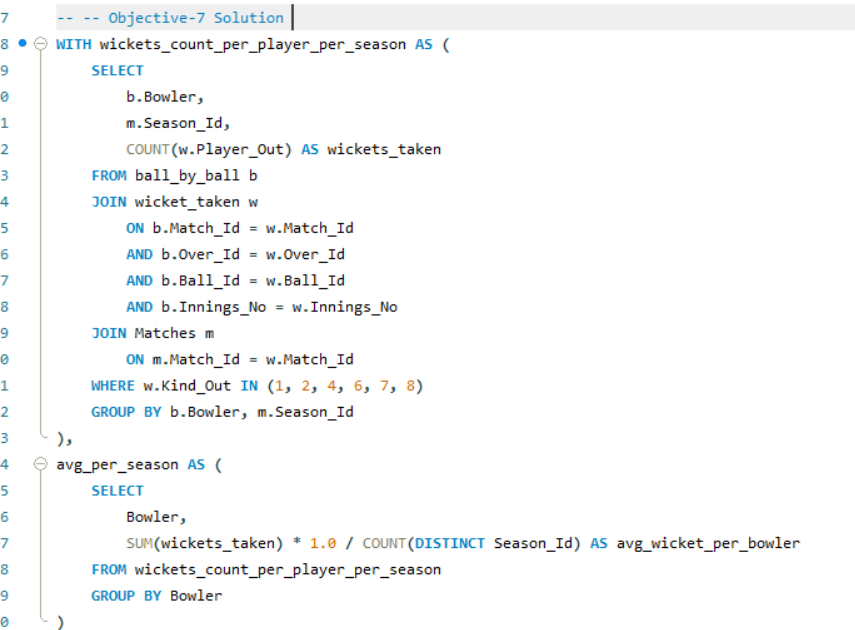


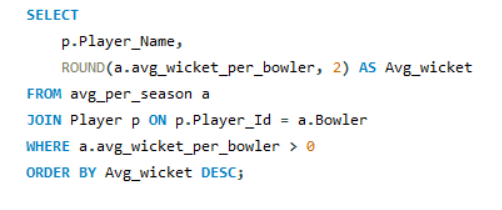
**7. What are the average wickets taken by each bowler considering all the seasons?**

Ans.

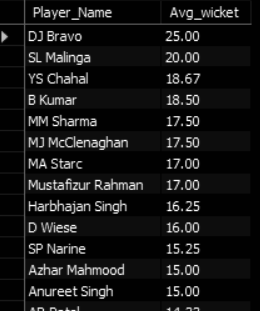
Approach-:

* Calculated average wickets per season for bowlers by filtering specific dismissal types and aggregating data season-wise using CTEs and joins.





Result-:

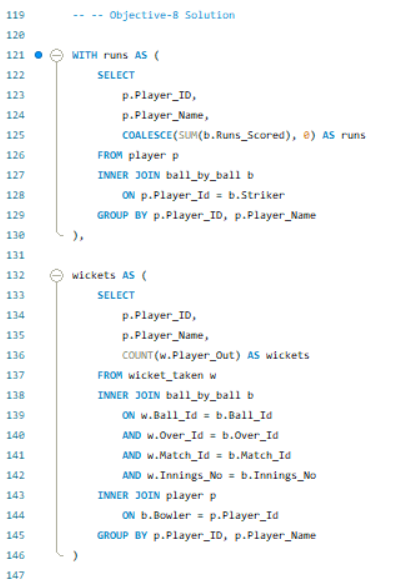


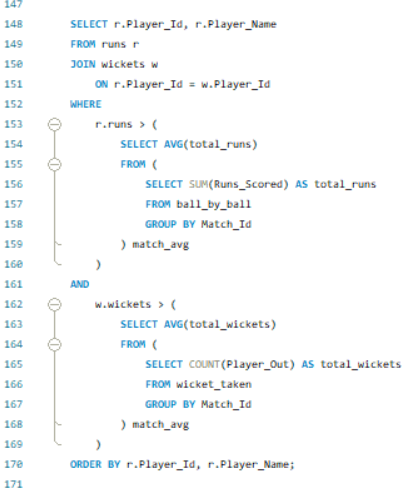
**8. List all the players who have average runs scored greater than the overall average and who have taken wickets greater than the overall average ?**

Ans.

Approach-:

* Identified top-performing all-rounders by selecting players with both runs and wickets above the match-wise average using CTEs and nested subqueries.





Result-:

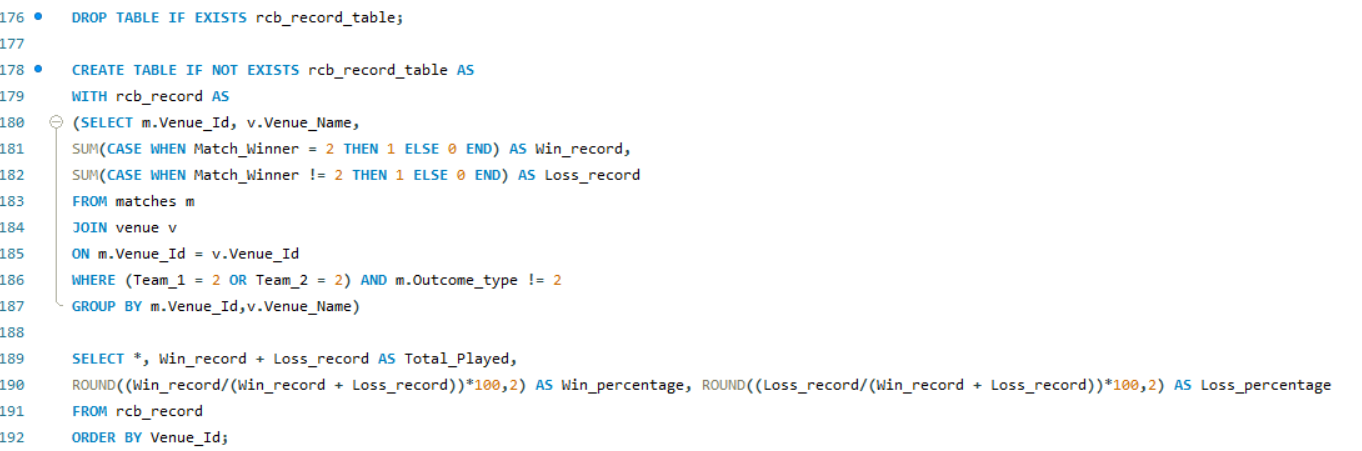
****

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

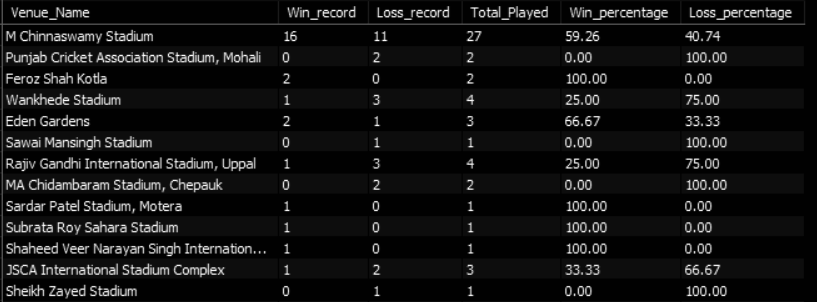
Ans.

Approach:-

* + Created a summary table for RCB’s venue-wise win/loss stats, including total matches and win/loss percentages, using CTEs and conditional aggregation.



Result-:

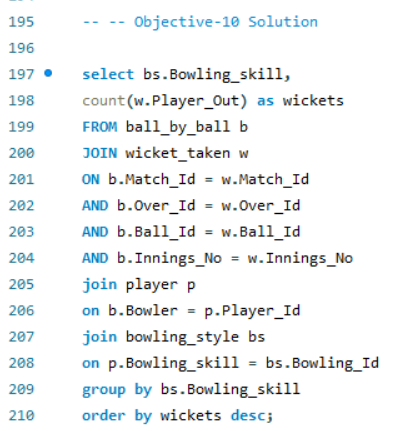


**10. What is the impact of bowling style on wickets taken?**

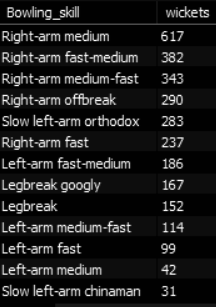
Ans.

Approach-:

* Analyzed wicket distribution by bowling style using joins on ball\_by\_ball, wicket\_taken, player, and bowling\_style tables.



Result-:

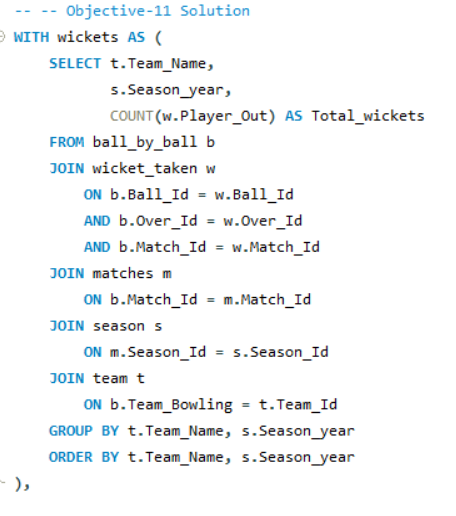


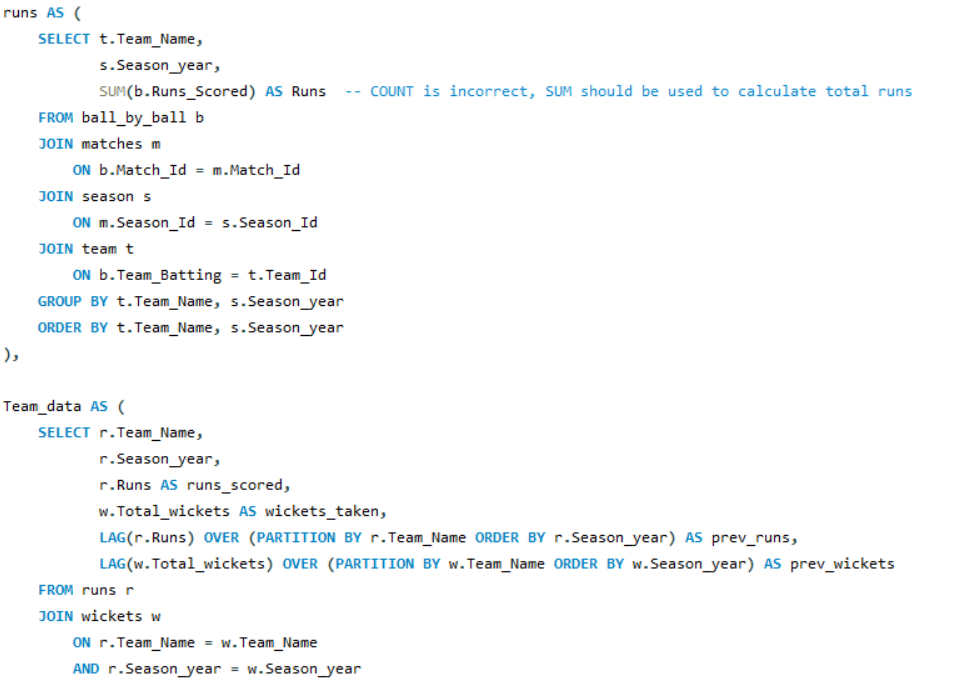
**11. Write the SQL query to provide a status of whether the performance of the team is better than the previous year's performance on the basis of the number of runs scored by the team in the season and the number of wickets taken?**

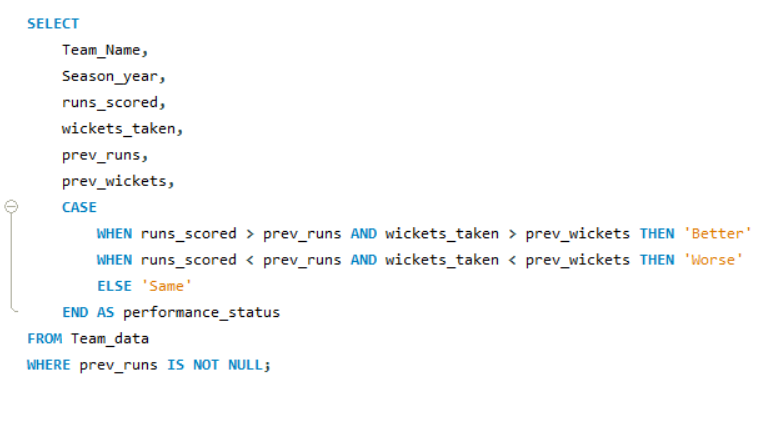
Ans.

Approach-:

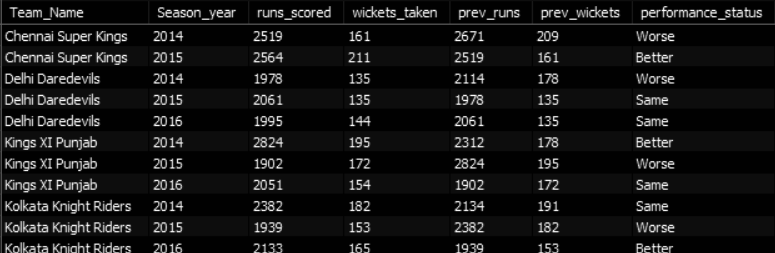
* Analyzed year-over-year team performance trends by comparing seasonal changes in runs scored and wickets taken, using window functions and CTEs.







Result-:



**12. Can you derive more KPIs for the team strategy?**

Ans.

Approach -:

**Powerplay Analysis:**

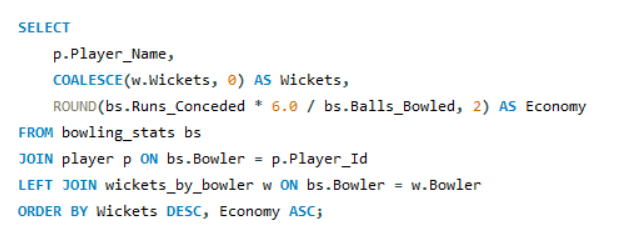
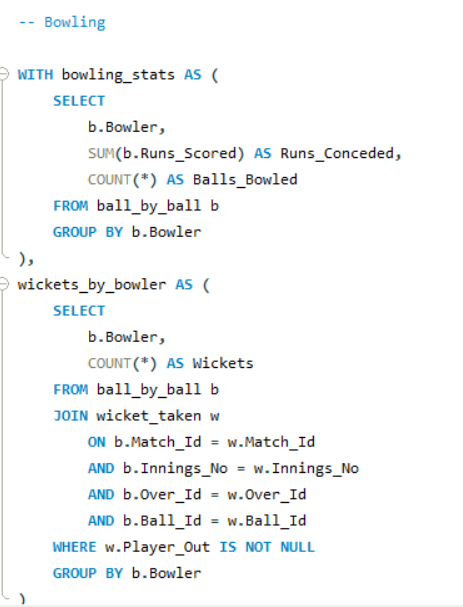
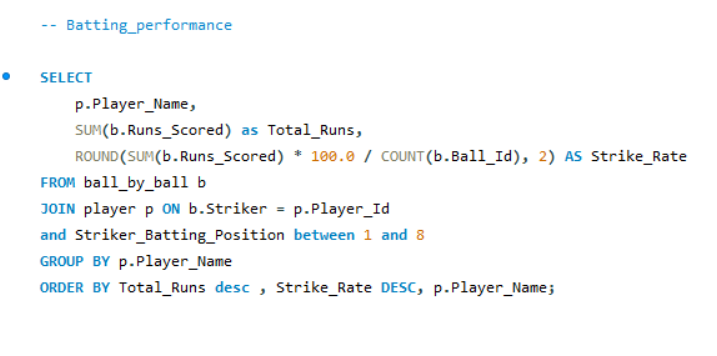
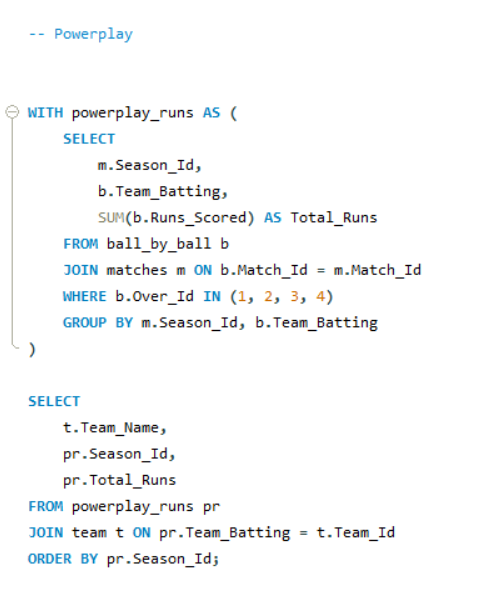
* Calculated total runs scored by each team in the first four overs of every season to assess powerplay performance.

**Batting Performance:**

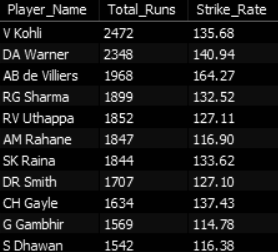
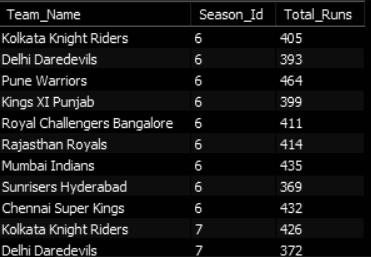
* Evaluated top-order batsmen (positions 1–8) based on total runs and strike rate to highlight consistent scorers.

**Bowling Performance:**

* Analyzed bowlers based on wickets taken and economy rate using total runs conceded and balls bowled.



Result-:

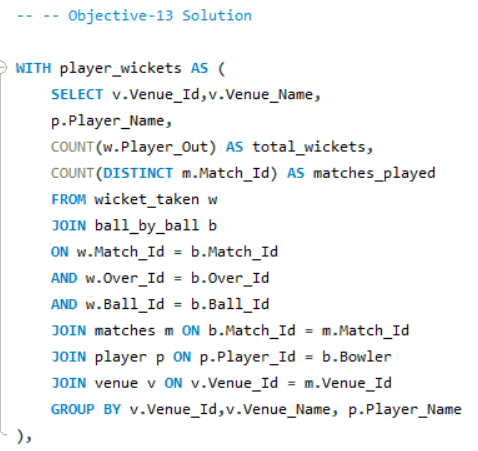


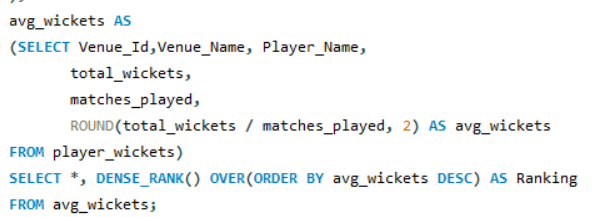
**13. Using SQL, write a query to find out the average wickets taken by each bowler in each venue. Also, rank the gender according to the average value.**

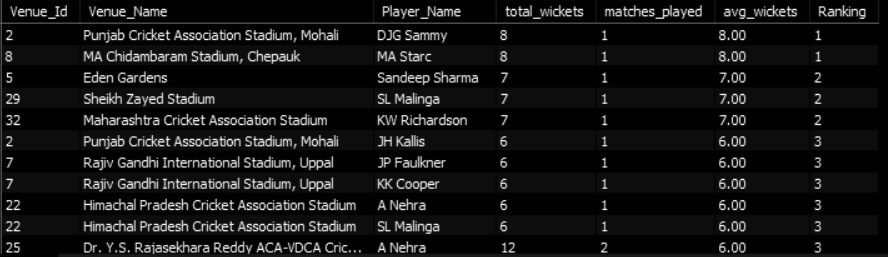
Ans.

Approach-:

* Ranked bowlers by average wickets per match at each venue using CTEs and DENSE\_RANK, combining venue-wise match data with player stats.





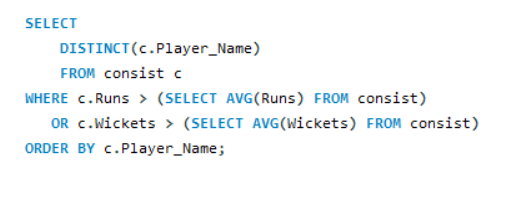
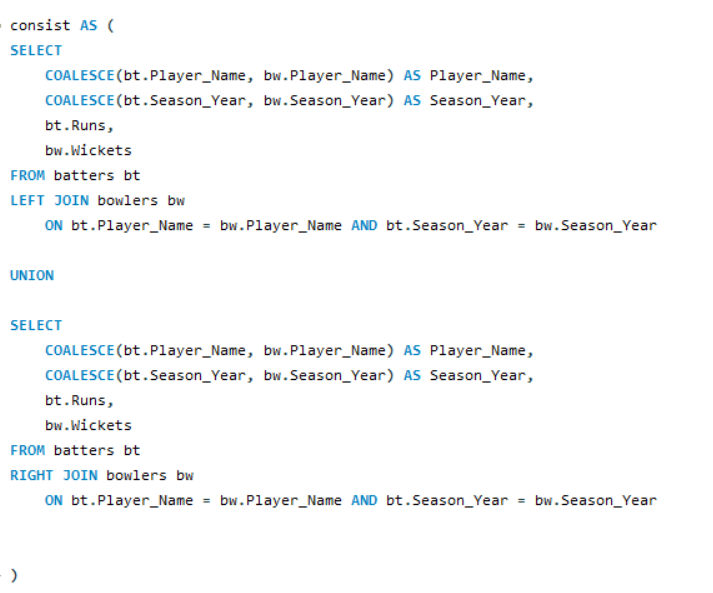
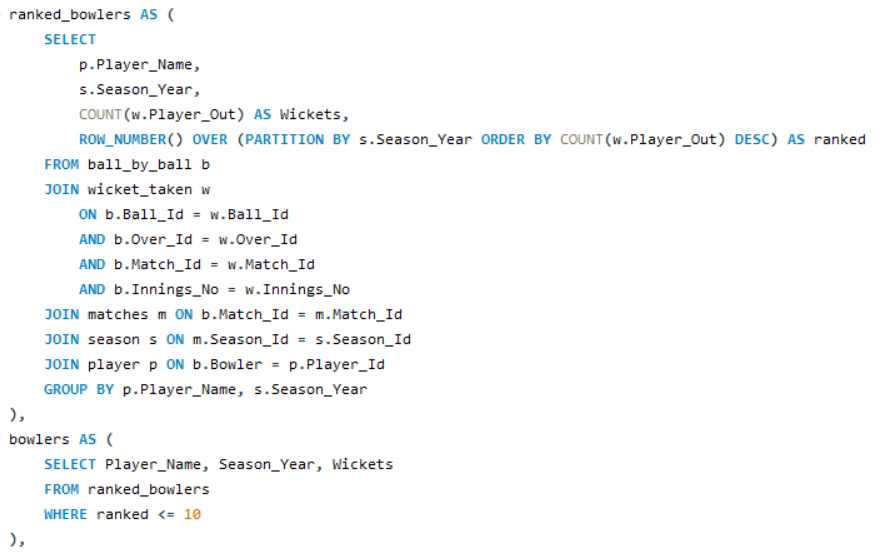
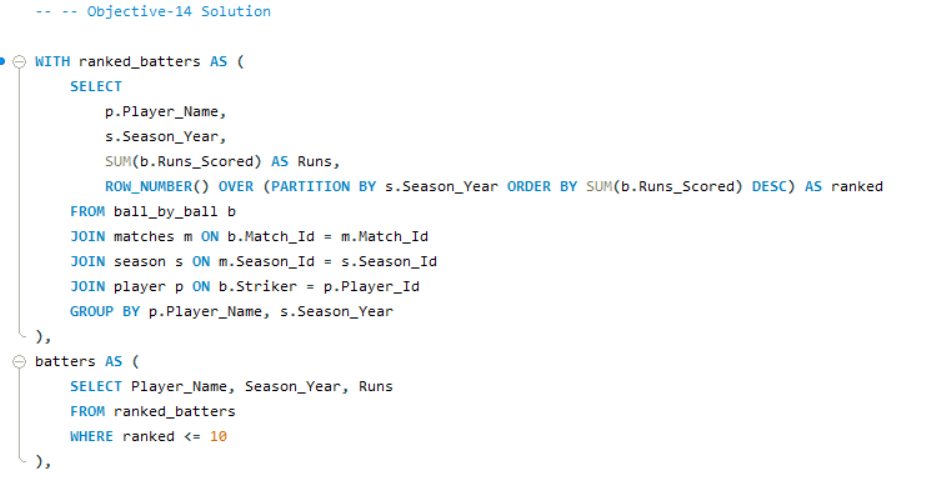


**14. Which of the given players have consistently performed well in past seasons? (will you use any visualization to solve the problem)**

Ans.

Approach-:

* Identified top 10 batters and bowlers per season, then filtered players who consistently performed above average in either runs or wickets using ranking functions and set operations.



Result-:

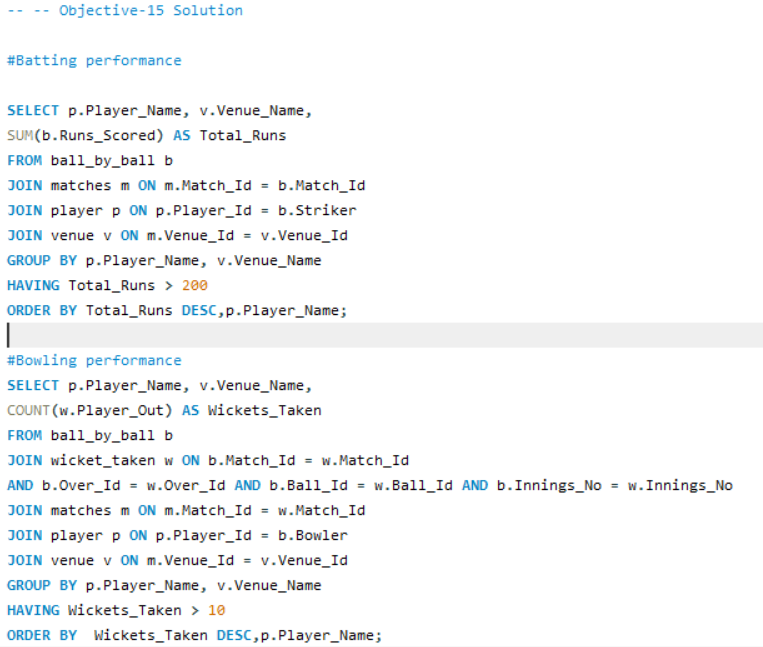


**15. Are there players whose performance is more suited to specific venues or conditions? (how would you present this using charts?)**

Ans.

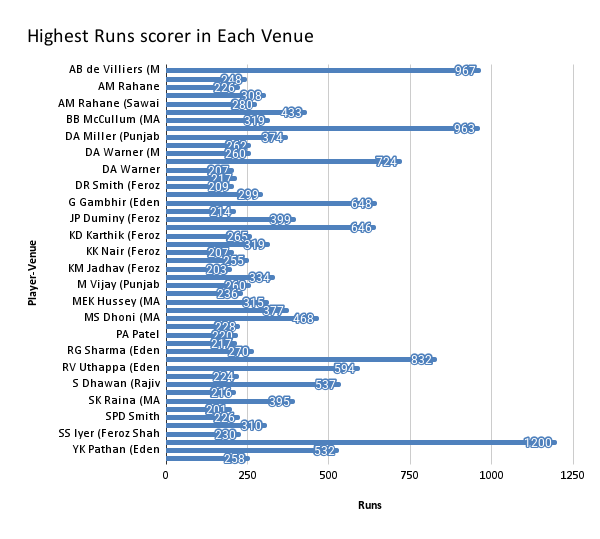
Approach-:

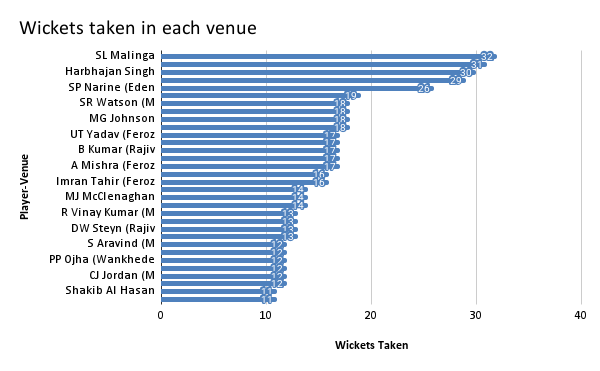
* Analyzed player performance per venue by filtering batters with 200+ runs and bowlers with 10+ wickets, highlighting top venue-wise contributors.



Result-:







## 

## 

## Subjective- Questions

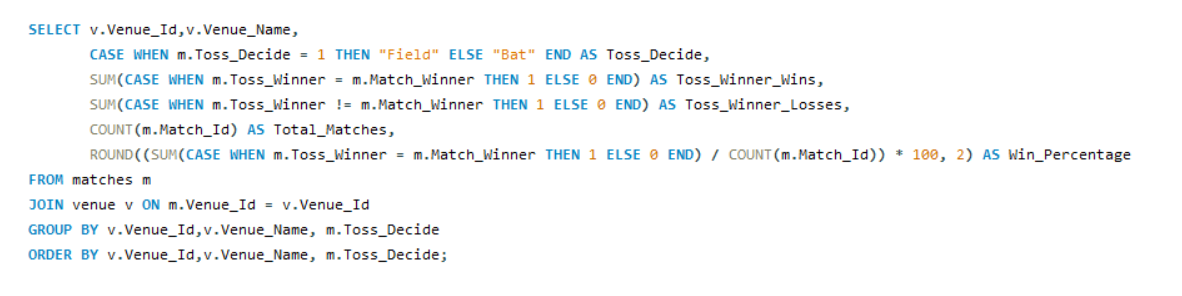
1. **How does the toss decision affect the result of the match? (which visualizations could be used to present your answer better) And is the impact limited to only specific venues?**

Ans.

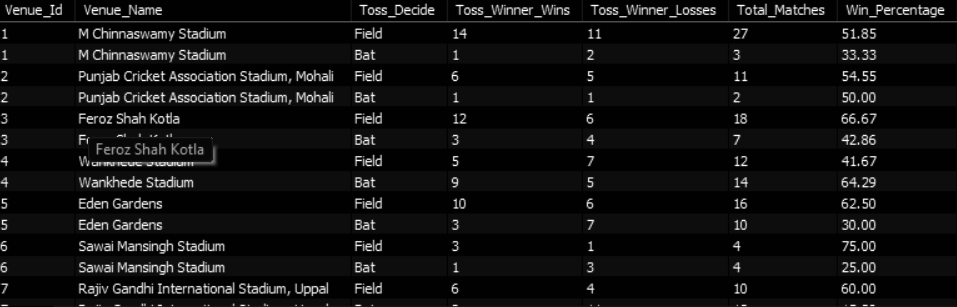
**Approach**:  
 The toss decisions across venues and seasons by checking if the team that won the toss went on to win the match. Also compared win percentages when choosing to bat vs bowl.

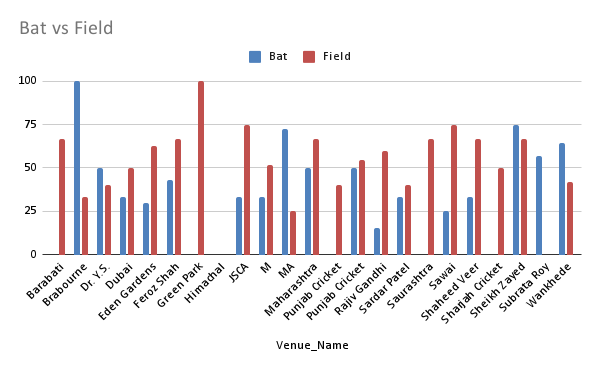
**Insights**:

* The chasing team tends to have a higher win percentage, especially at venues like Chinnaswamy (RCB’s home ground) and Wankhede, where dew in the second innings plays a role.
* In venues with larger boundaries or slower pitches, batting first was slightly more favorable.
* The impact of the toss was venue-dependent, not universal.



Result-:





**Recommendations**:

* RCB should opt to chase at Chinnaswamy when possible.
* Visualization like bar charts comparing win percentages based on toss decision per venue would be effective in presenting this insight.

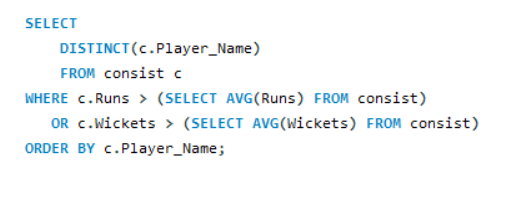
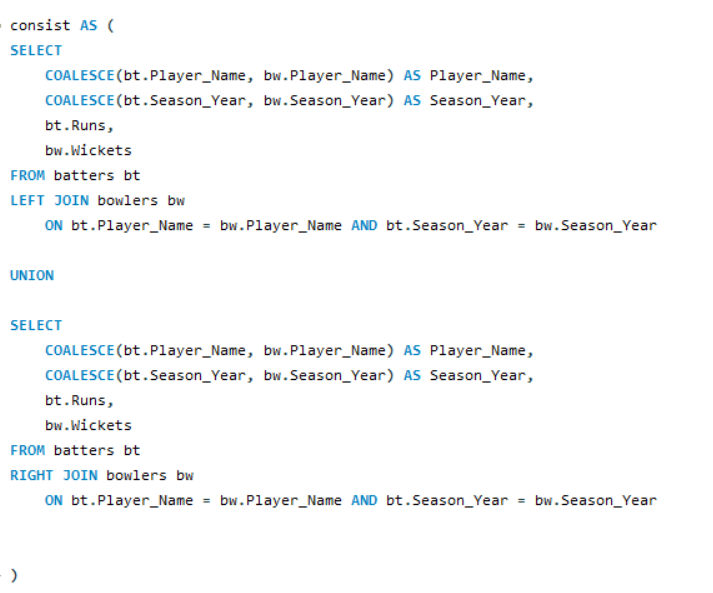
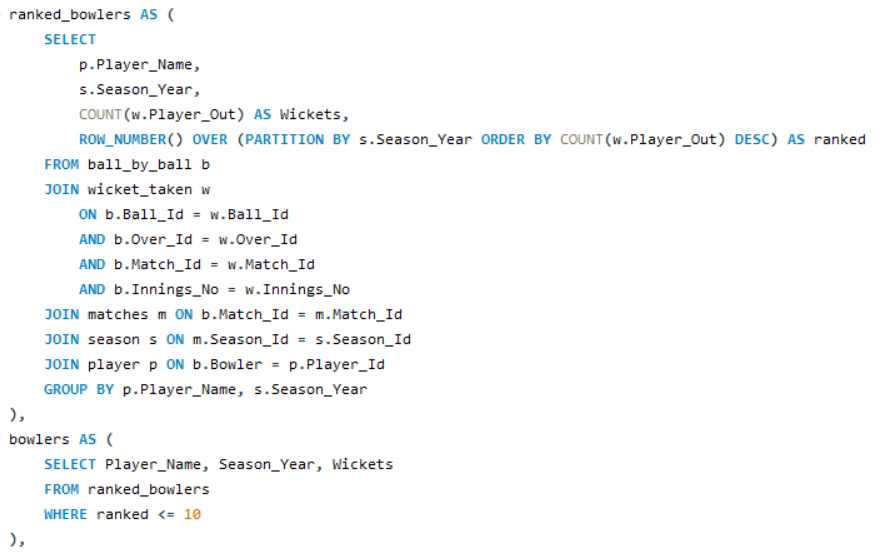
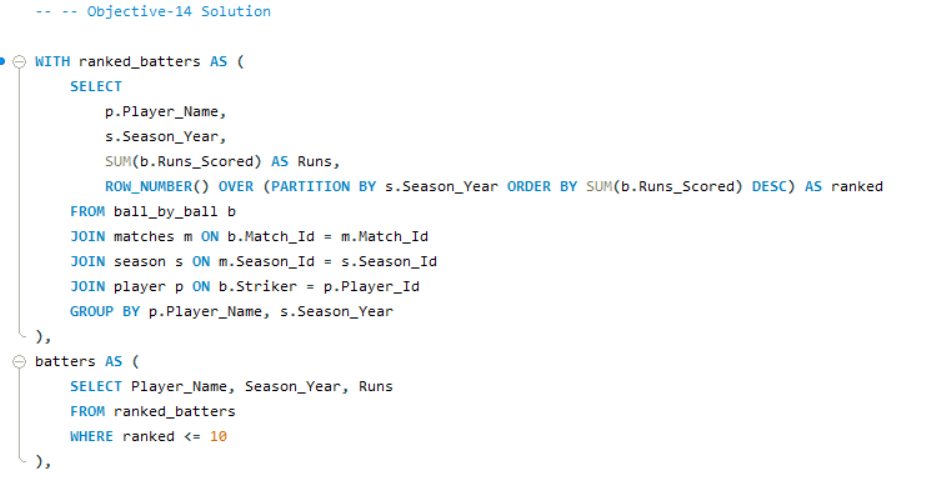
**2. Suggest some of the players who would be best fit for the team.**

Ans.

**Approach**:  
 Filtered players with consistent performance in both batting (high average and strike rate) and bowling (low economy and good wicket-taking ability).

**Insights**:

* Players like DJ Bravo, Glenn Maxwell offer great all-round value.
* Players like AB de villiers, Chris gayle have done the job consistently for RCB.
* Players like Bhuwneshwar kumar, Amit mishra , DJ Bravo can be the best bowling options



Result-:



The above SQL snippet highlights the most consistent players in the IPL across previous seasons. In a high-stakes and fast-paced tournament like the IPL, where teams play numerous matches in a short span, player consistency becomes a crucial asset. Consistent performers provide teams with reliability under pressure, stepping up during critical phases to help turn games around or maintain control when needed.

Moreover, having a variety of dependable players across roles—batsmen, bowlers, all-rounders—gives team management greater flexibility in forming strategies. It enables them to rotate the squad efficiently, manage player workloads, and plan backups in case of injuries or dips in form. Ultimately, a strong core of consistent players contributes significantly to a team’s chances of progressing deep into the tournament and potentially winning the title.

**Recommendations**:

* RCB should target multi-skilled players who can contribute in both departments, especially for the middle and lower order.
* Consistency in performance should be prioritized over big names.

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

Ans.

The following parameters are very important while selecting the players

Strike-Rate- : It is most important in fast paced matches such as IPL because it leverages the team with high scoring opportunities.

Average-: Having a good batting average provides a dependency over the batsman so that the other batsmen can take risks and keep the scoreboard rolling.

Bowling Average-: Just like batting average, bowling average is also equally important so that the team can force the opposition to take risks for scoring big runs.

Economy -: Some venues such as Bangalore have a short boundary to overcome this having a bowling attack with economical bowlers is important to limit the score at a below par score.

**Insights**:

* Fitness and availability: Players who are injury-prone or play partial seasons can disrupt momentum.
* Matchups: Some players perform better against specific teams or at particular venues , should be considered for selection.

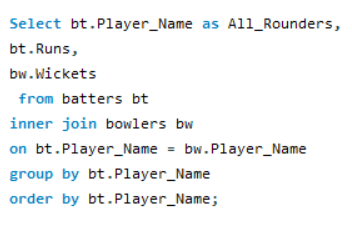
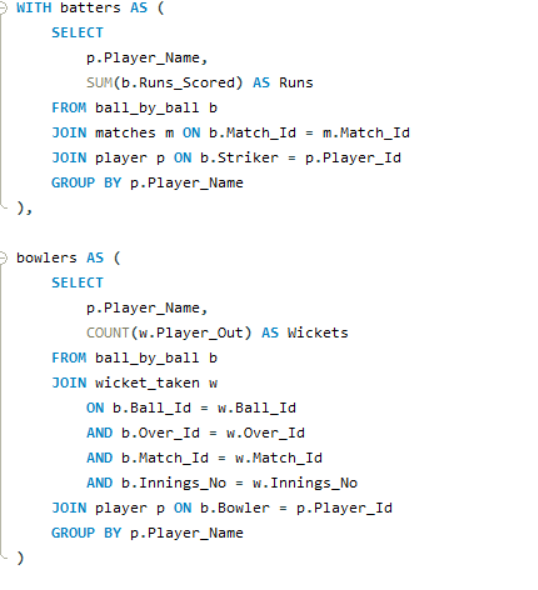
**4. Which players offer versatility in their skills and can contribute effectively with both bat and ball? (can you visualize the data for the same)**

Ans.

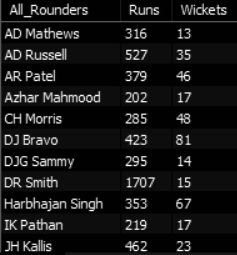
**Approach**:  
 Used combined metrics runs scored and wickets taken to shortlist all-rounders across seasons.

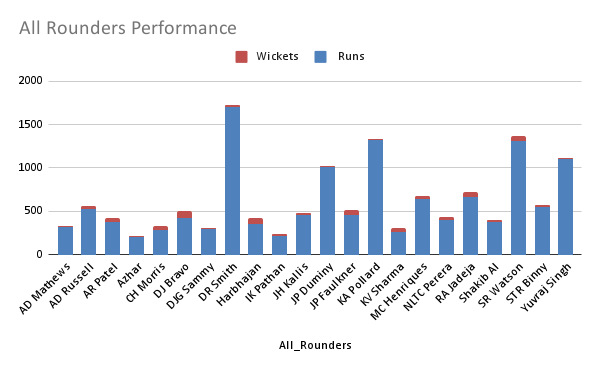
**Insights**:

* Andre Russel, Ravindra Jadeja, Glenn Maxwell, DJ Bravo have consistently contributed in both departments.
* Players like Axar Patel and DJ Bravo provide control and economy along with useful batting down the order.



Result-:





**Recommendation**:

Players who contribute significantly with both bat and ball, are valuable assets for any IPL team. Their dual capabilities offer strategic flexibility, allowing teams to adjust their lineup based on match conditions, balance the side with an extra batter or bowler, and cover for potential underperformance in either department. These all-rounders bring much-needed stability to the team, especially in pressure situations, by stepping up when either the batting or bowling unit is struggling.

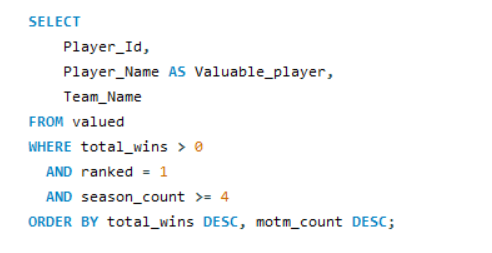
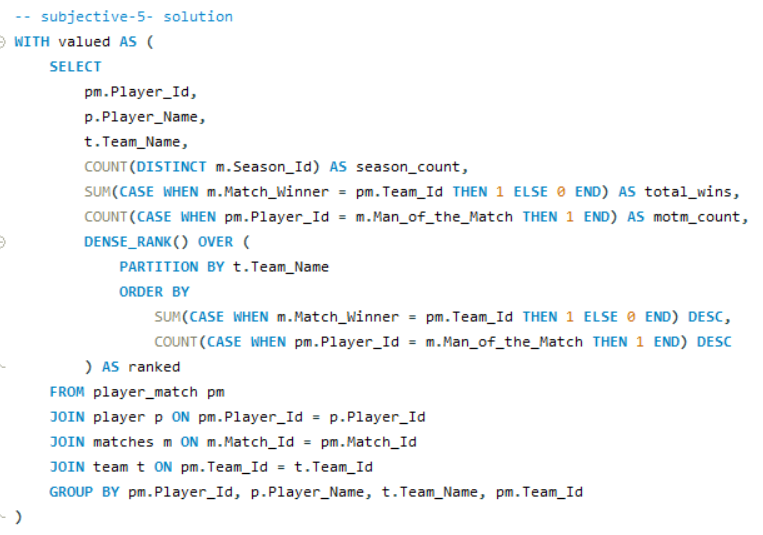
**5. Are there players whose presence positively influences the morale and performance of the team? (justify your answer using visualization) ?**

Ans.

**Approach**-:  
 Observed win ratios and team performances in matches with and without certain players.

**Insights**:

* Gautam Gambhir , Virat Kohli, David Miller, Shikhar Dhawan and Rohit Sharma clearly influence team morale — win % was higher when they played.
* Presence of leaders or clutch performers correlates with better handling of pressure situations.



Result-:





The above sql query presents the most valuable players for different Teams in the league who boost the Morale of the Team and at the same time performs up to the mark as well.

**Recommendation**:

Teams should not only focus on raw statistics like runs scored or wickets taken, but also consider a player's leadership qualities and the impact they have on the team’s morale and performance. Players with a strong presence such as experienced captains or those known for composure under pressure can influence the dressing room atmosphere, guide younger teammates, and make critical decisions on the field. Their ability to motivate others and maintain a positive mindset during challenging moments often proves to be the difference between a narrow loss and a match winning performance. These attributes, not always visible in data, play a vital role in building a resilient and cohesive team.

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

Ans.

Mega Auction is an opportunity where every team has a chance to turn around their faiths. For the upcoming mega auction RCB should focus on the following Analytical strategies.

**Maximising Power-Plays -:**  Rcb can bid for the players having a strike rate of 150+ in the Top order to maximise the power play when there is a big scoring opportunity since only 2 Boundary riders are available for the feilding.

**Middle order**-: Sometimes there is a chance of early fall of wickets in the Powerplay. There comes the need of a good middle order which can play a little defensive early to soak the pressure and then accelerate following up a big score.

**Lower-Order**-: It should be a combination of batsmen and all-rounder having a strike rate of 180+ in death overs to make the score bigger if already in a good situation or finishing the match early to increase the net run-rate.

**Economical Bowlers-:** Every team plays 50% of matches in the home ground if a team wins all the matches in the home ground it gives them a big edge over the other teams, In case of RCB it’s a tough task as the boundaries are shorter so a good wicket-taking bowling line-up with less economy is very important.

**Variety in bowling -:** Rcb should opt for different bowling options such as left arm pacer, right arm pacer, chinamen, leg-spin as well as off spin to make effective plans against each batsmen.

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

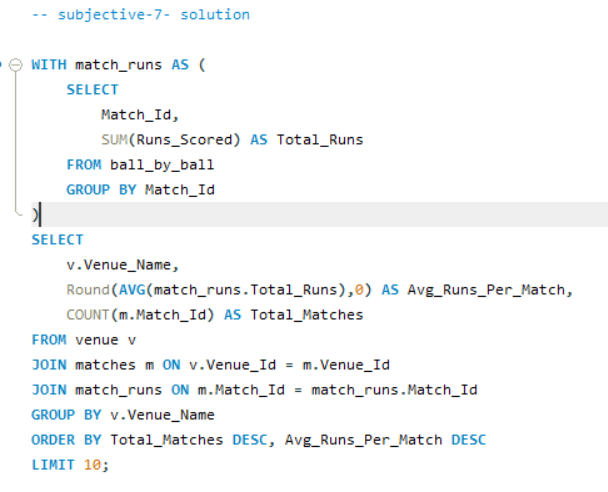
Ans.

**Approach**- :

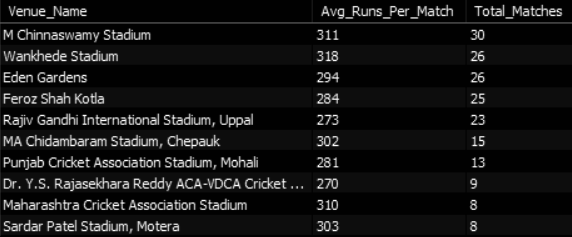
Calculated the average runs scored per match at each venue. Created a (CTE) to get the total runs scored in each match and joined this with the matches and venue tables to compute the average runs per match at each venue and ranked the venues based on match count and scoring rate.

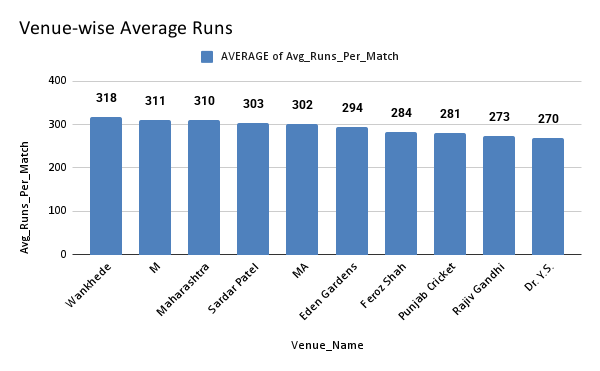
**Insights**:

* Pitches with no lateral movement and short boundaries drive up scores.
* Viewership spikes when matches are thrilling, and high scores amplify that.
* High-scoring matches push teams to draft more explosive batters and death-over specialists.



Result-:





Venue plays an important role in a high-scoring match the pitches that don't offer any purchase to the bowlers are mainly good for a high scoring match and also the venues that have shorter boundaries are also favourable for high scoring matches.

The mentioned venues offer a high scoring match most of the time.

A high scoring match has a variety of effects that are as follows-:

In a fast moving world the audience don’t really invest their complete time to watch all the innings but if in a match there is a lot of enjoyment and thrill to the audience ball after ball resulting in increased viewership.

**Recommendations:-**

The team should be ready with a good set of batsmen and all-rounders with a good strike-rate to score high. Also the team must have a set of bowling units that can take wickets and concede less runs; this is the only way to outplay the opposition in these kinds of venues.

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

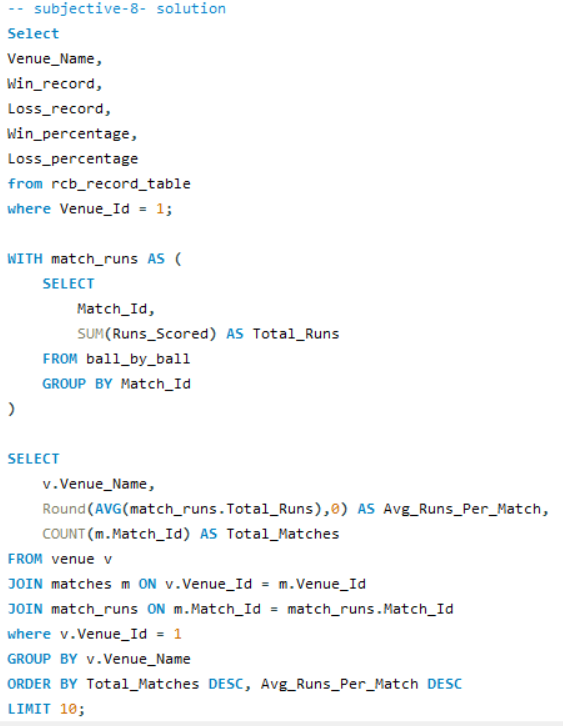
Ans.

**Approach**-:

To analyze RCB’s home-ground advantage, used two queries. The first retrieved RCB’s win-loss record and percentages at a specific venue (Venue\_Id = 1), which corresponds to their home ground (Chinnaswamy Stadium). The second query calculated the average runs scored per match at that venue using a CTE to aggregate total match runs. By combining performance metrics (win/loss ratios) with scoring conditions (average runs), one can understand how venue characteristics influence outcomes .

**Insight**:-

* RCB can simulate pressure situations in home conditions during training to get used to high-scoring scenarios.
* Invest in data-backed decisions for bowling changes and field settings at Chinnaswamy.



Result-:



**Recommendations-:**

In the home ground there is a familiarity with the pitch and conditions of the venue so it is somehow easier for the teams to perform. Also when a team plays matches in the homeground they have local crowd support as well to boost the morale as well.

To increase this advantage Rcb should take following steps -:

* Play with good striking ability batters
* Play economical bowlers
* Every 2nd batsmen should be an attacking batsman.
* Try to score a big total
* When chasing should keep run rate slightly above the required one.
* The toss decision at Chinnaswamy favours the chasing team to try to win the toss and chase down the score for that having attacking lower order is required.

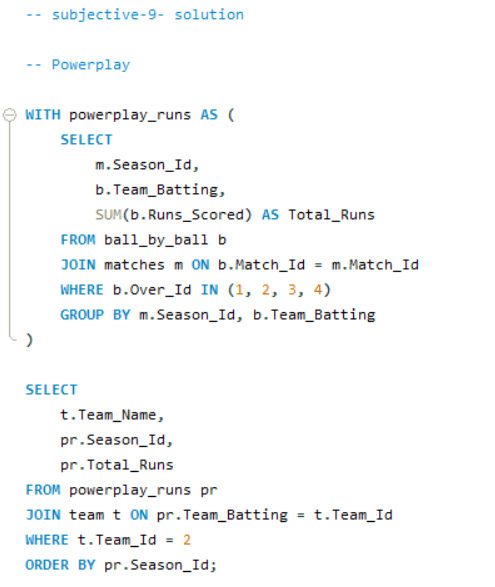
**9. Come up with a visual and analytical analysis of the RCB's past season's performance and potential reasons for them not winning a trophy.**

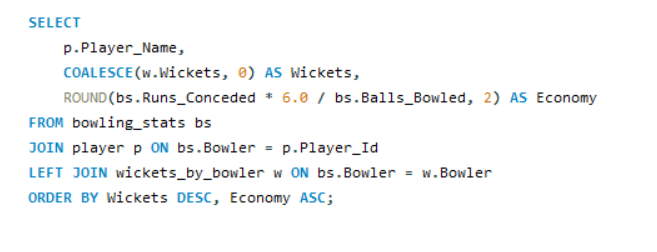
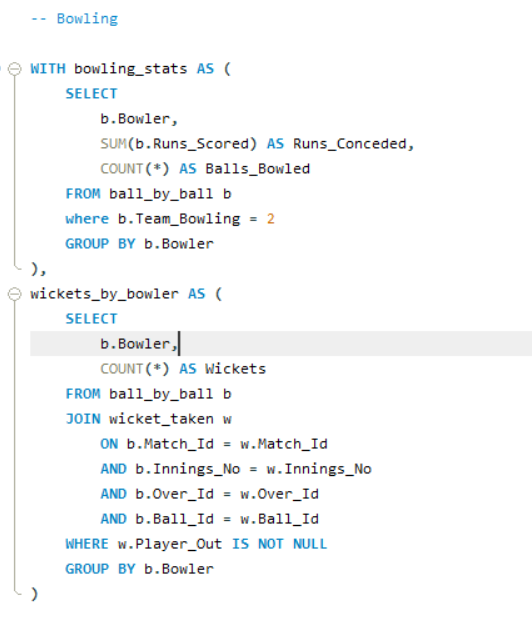
Ans.

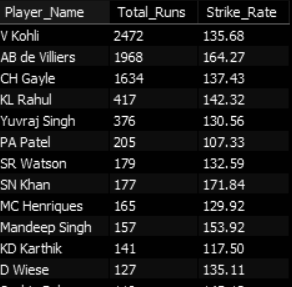
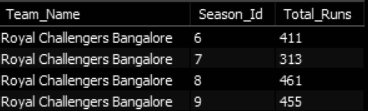
**Approach**-:

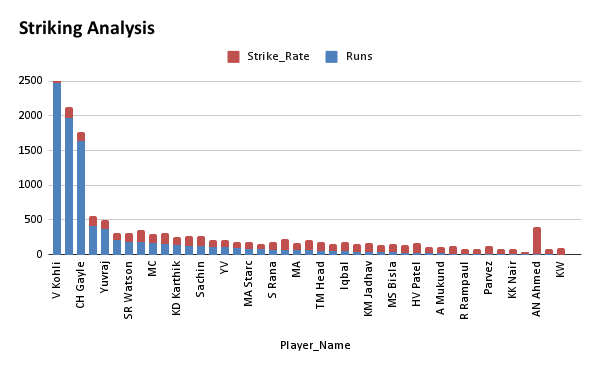
**Insights**:

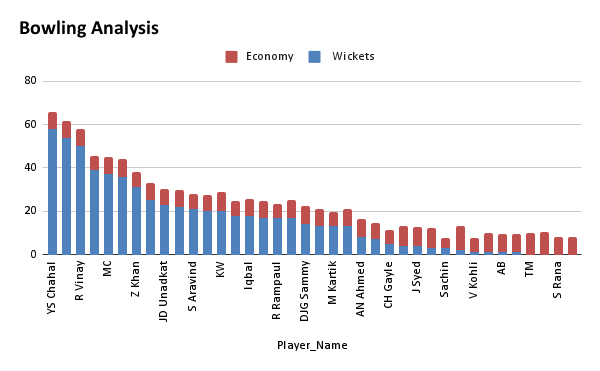
* Over-reliance on top 3 batters, low contribution from middle order.
* Bowling lacked consistency, with high economy rates at death.
* Powerplay scoring is below league average.











The above query provides a brief about multiple aspects which are the major reason of not winning a trophy yet

* If we have a look at the batting performance it clearly shows that RCB is too much depending on their top order if we ignore the top 3 batters there is a dip in the average runs as well as the strike-rate which results in a loss if the top order fails.To improve this RCB must buy some dependable batting options in the middle order and lower order to strengthen the batting depth.
* Similarly we have a look at the bowling metrics RCB has a lot of good wicket-taking bowling options but they have a certainly high economy rate which can result in putting a big total on board by opposition which makes both chasing as well as defending hard especially when the home ground is a high scoring venue. So they should look for the economical wicket taking options to transfer the pressure back on the opposition.
* The powerplay is something which can enable a team to take a lead in the game but RCB has never scored more than 500 runs in this which could be a reason of not winning a trophy as well because every team plays with each other twice one home one away game that makes it a minimum of 12-14 games that means RCB is scoring close to 45 runs in the powerplay a increase in the score by 10 runs each game can make team take a marginal lead in the game.

These were some of the major reasons that RCB can focus upon for winning their first title

**Recommendations**:

* Strengthen the lower middle order with all-rounders.
* Using the resources at their best possible batting positions.
* Drafting the economical wicket-takers in the Auction.

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

Ans.

* **Study the Database Structure** Starting by understanding the schema and relationships between key tables such as matches, ball\_by\_ball, player, venue, etc. This step is essential to understand how the data is organized and what joins are required for analysis.
* **Define Key Performance Metrics** Establishing KPIs  
  + **Batting**: Strike rate, batting average, total runs
  + **Bowling**: Wickets taken, economy rate, bowling average
  + **Team**: Win-loss records, run rates, consistency metrics
* **Use SQL Aggregations** Applying SQL functions like SUM(), AVG(), COUNT(), and using GROUP BY to derive meaningful aggregations across players, matches, seasons, and venues.
* **Conduct Player Performance Analysis** Analyzing individual performance over multiple seasons for both batters and bowlers to spot trends, consistency, and standout contributions.
* **Extract Venue-Based Insights** Comparing team and player performance across different venues, especially to understand home vs. away dynamics and high/low scoring grounds.
* **Evaluate Toss Decision Impact** Exploring how the decision to bat or bowl first affects match outcomes, especially at specific venues or in certain seasons.
* **Identify Consistent Performers** Using techniques like window functions or season-by-season comparisons to shortlist players with consistently high performance.
* **Find Role-Specific Strengths** Analyzing player performance based on match phases (e.g., powerplay, middle overs, death overs) to understand who excels in which role.
* **Build RCB-Specific Insights** Deep dive into RCB’s data by evaluating their past performance, best venues, reliable players, and weaknesses to tailor strategies specific to the team.
* **Formulate Strategy Recommendations** Finally, using the findings to suggest actionable steps like:
* Ideal player retention/acquisition
* Game-day tactics (batting order, bowling combinations)
* Auction planning based on gaps and conditions

**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".**

Ans.

