**SUBMISSION BY: TARUN KUMAR ANAND**

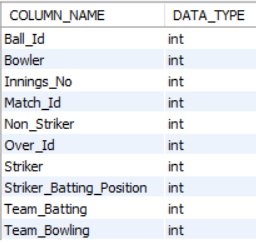
**PROJECT: IPL Strategy for RCB**

**BATCH: DATA SCIENCE COURSE OCTOBER 2024**  
  
  
**Learners have to come up with a Report to support the answers to the following questions and suggestions**

Objective Questions

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

**Ans: There are 10 columns with the data type integer.**



Desc Ball\_by\_Ball;

### Objective:

List the different data types (dtypes) of columns in the Ball\_by\_Ball table.

### Segregation and Documentation:

This shows the column data types in the Ball\_by\_Ball table, all of which are integers.

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: -

SELECT

SUM(ER.Extra\_runs) + SUM(BB.Runs\_Scored) AS total\_runs

FROM Extra\_runs ER

JOIN Ball\_by\_Ball BB ON ER.Match\_Id = BB.Match\_Id

WHERE BB.Innings\_No = 1

AND ER.Innings\_No = 1

AND BB.Team\_Batting = (SELECT Team\_Id FROM Team WHERE Team\_Id = 2);

### Objective:

Calculate the total runs scored by RCB in the 1st season, including extra runs.

### Query Review:

The query looks mostly correct but misses filtering for the RCB team. It should include a filter for RCB and check if the Match\_Id corresponds to the first season.

### Explanation:

SUM(ER.Extra\_runs): Adds extra runs.

SUM(BB.Runs\_Scored): Adds runs scored.

Team\_Batting: Filters for RCB team based on RCB Team ID.

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:**

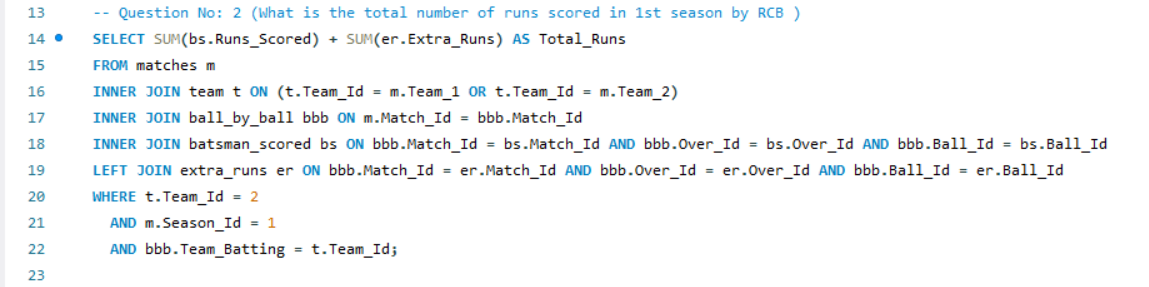
To find the total runs scored by Royal Challengers Bangalore (RCB) in the first IPL season, including extra runs, the following tables are needed:

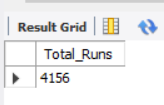
**Required Tables:**

1. **matches**: Contains match details, including the teams and season.
2. **team**: Provides team names and IDs.
3. **ball\_by\_ball**: Tracks each ball bowled, including the team batting.
4. **batsman\_scored**: Records runs scored by batsmen on each ball.
5. **extra\_runs**: Tracks additional runs like wides and no-balls.

**Query Approach:**

1. **Match Selection:** Filters matches where RCB played (Team\_Id = 2) in the first season (Season\_Id = 1).
2. **Joining Tables:** Joins relevant tables to link matches, balls, runs scored, and extra runs.
3. **Sum Runs:** Adds up runs scored by RCB's batsmen and any extra runs.
4. **Result:** The total runs scored by RCB in the first season.



****

1. How many players were more than age of 25 during season 2?

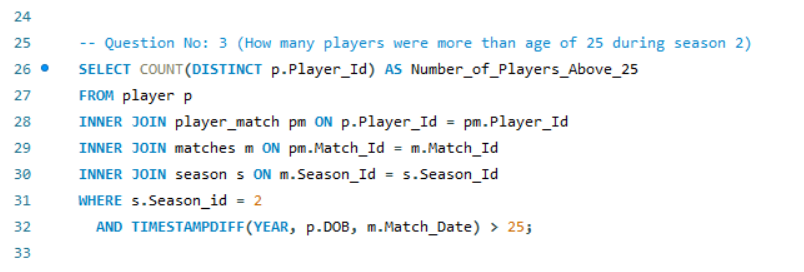
**Ans:**

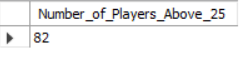
To find the number of players who were more than 25 years old during season 2, we need the following tables:

1. **player** - To get the player's date of birth.
2. **player\_match** - To find which players participated in matches during season 2.
3. **matches** - To find matches in season 2.
4. **season** - To identify season 2 and its matches.

**Query Approach:**

1. Identify all matches in season 2.
2. Find all players who played in these matches.
3. Calculate the age of each player during season 2.
4. Count the players who were older than 25.





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

**Ans:**

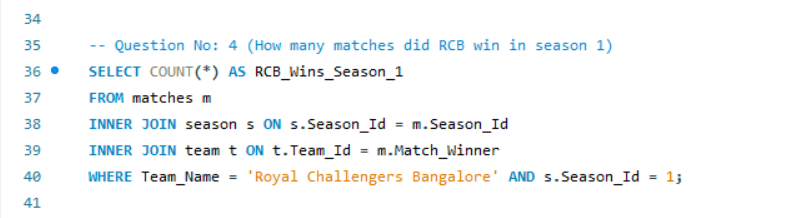
To determine how many matches Royal Challengers Bangalore (RCB) won in season 1, we need the following tables:

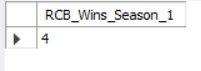
1. **matches** - To identify match details including the winning team.
2. **season** - To filter matches that occurred in the first season.
3. **team** - To match the winning team with RCB.

**Query Approach:**

1. Use the season table to identify matches from the first season.
2. Filter the matches for those where the Match\_Winner corresponds to RCB.
3. Count the number of matches where RCB was the winner.

This method provides the total number of matches won by RCB in season 1.



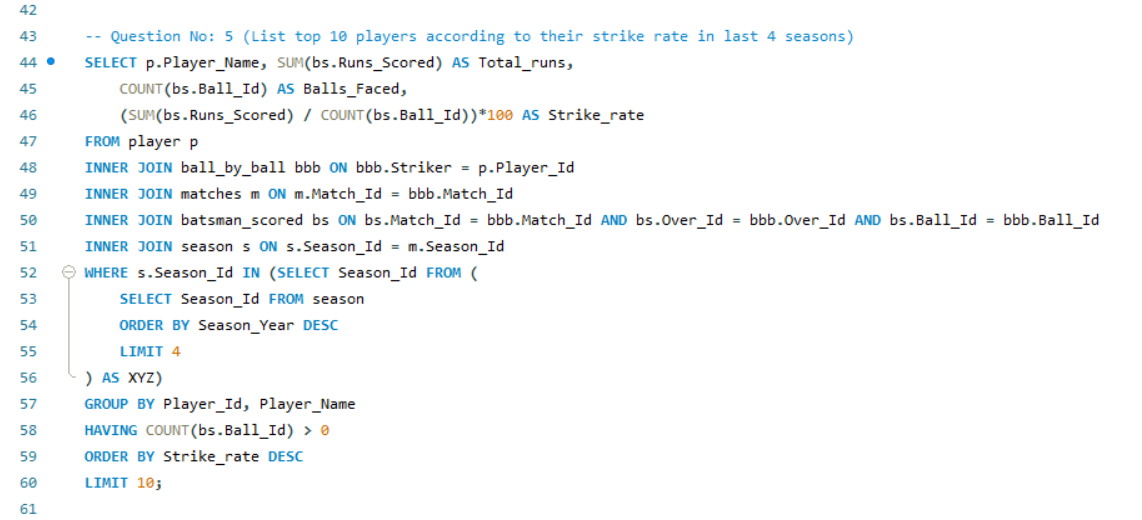
****

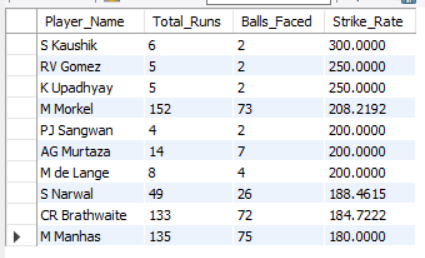
1. List top 10 players according to their strike rate in last 4 seasons

**Ans:**

To list the top 10 players by strike rate in the last 4 seasons, we will need the following tables:

* **Tables Required:**
* **player**: Provides the player's name (Player\_Name) and Player\_Id to identify each player uniquely.
* **matches**: Used to filter the matches that occurred in the last 4 seasons.
* **season**: Helps identify the last 4 seasons using Season\_Id.
* **ball\_by\_ball**: Provides the details of each ball faced by the player, allowing us to calculate the total balls faced.
* **batsman\_scored**: Provides the runs scored by the player on each ball, allowing us to calculate the total runs.
* **Query Approach:**
* We find the total runs and balls faced by each player.
* Then, calculate the strike rate.
* Finally, we list the top 10 based on this rate.
* **Insights:**
* Players like S Kaushik, RV Gomez, and K Upadhyay have shown exceptional strike rates, with over 200 in the last 4 seasons, indicating aggressive and efficient scoring in limited opportunities.
* Bowlers like M Morkel and PJ Sangwan also feature in the list, showcasing their ability to contribute valuable runs quickly despite not being specialist batsmen.
* Aggressive lower-order hitters, such as CR Brathwaite and M de Lange, can turn matches around with their quick scoring, which is critical in shorter formats like T20.
* The strike rate of over 180 for players in the list reflects their ability to accelerate the innings, making them game-changers in pressure situations.
* **Recommendations:**
* Teams should utilize these high-strike-rate players in critical overs, especially in the death overs of T20 matches, to maximize the team's total score.
* Strategic planning should focus on rotating these players in middle-to-lower batting orders to exploit their ability to score quickly in high-pressure situations.





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

**Ans:**

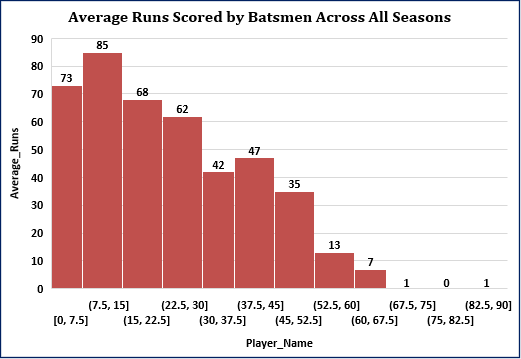
To find the average runs scored by each batsman considering all the seasons, we will need the following tables:

1. **player**: Provides player names and IDs.
2. **batsman\_scored**: Contains information about the runs scored by each batsman on each ball.
3. **ball\_by\_ball**: Links each run scored to a specific player.

**Query Approach:**

* **Joins**: The tables are joined to obtain the runs scored by each player.
* **Calculations**: The average of these runs is calculated for each player.
* **Sorting**: The results are sorted to show players with the highest averages first.





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

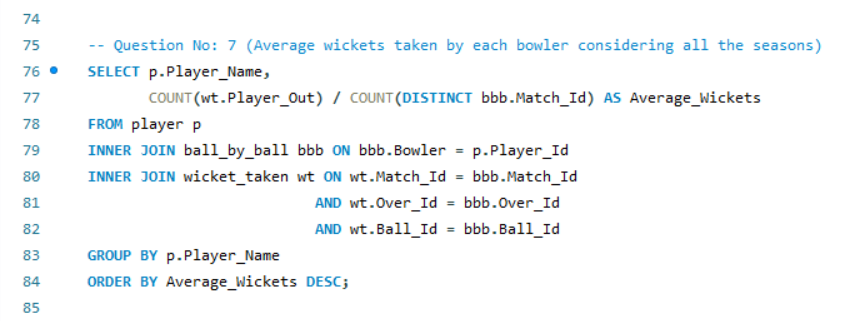
**Ans:**

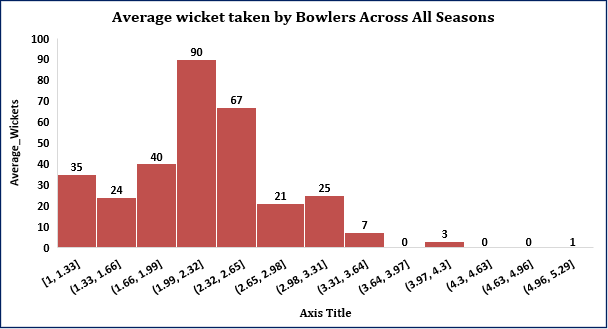
To find the average wickets taken by each bowler considering all the seasons, we will need the following tables:

1. **player**: Provides player names and IDs.
2. **wicket\_taken**: Contains information about the wickets taken by each bowler on each ball.
3. **ball\_by\_ball**: Links each ball to the bowler who delivered it.

**Query Approach:**

* **Joins**: The tables are joined directly to obtain the wickets taken by each bowler.
* **Calculations**: The total number of wickets is divided by the number of matches to calculate the average wickets taken by each player.
* **Sorting**: The results are sorted to show players with the highest averages first.



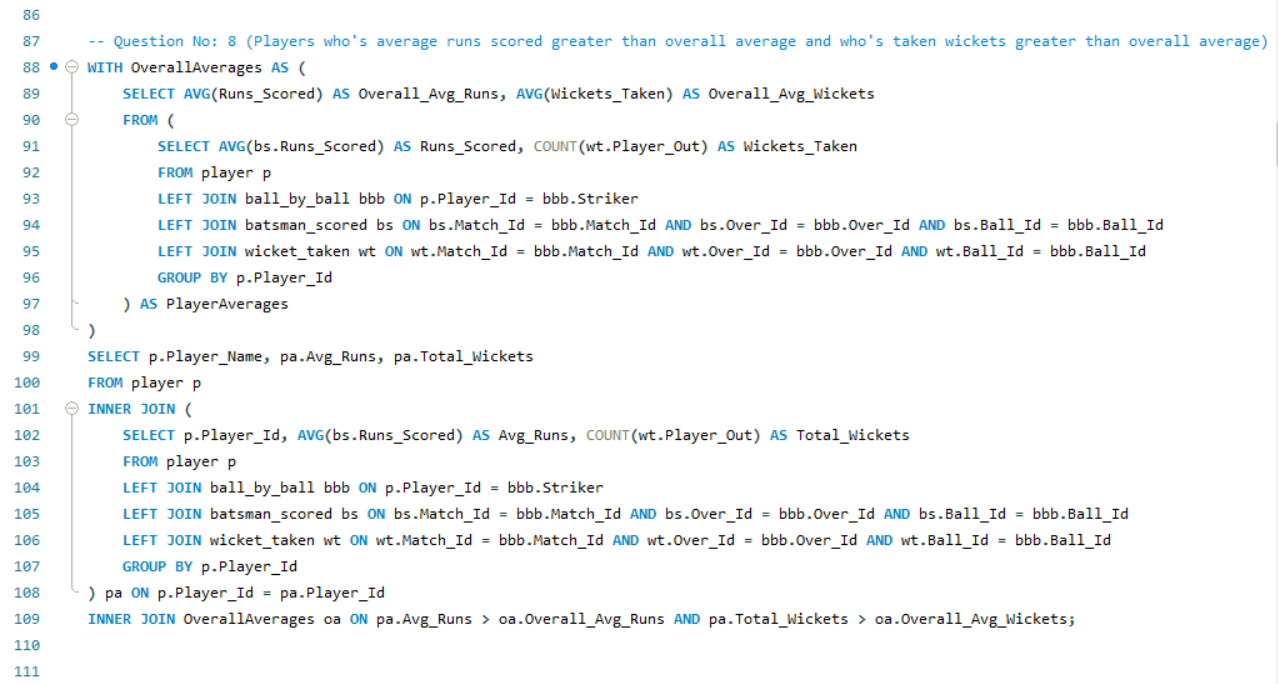
****

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

**Ans:**

To 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 will need the following tables:

* **Tables Required:**
* **player:** Provides player names and IDs.
* **batsman\_scored:** Contains information about the runs scored by each batsman.
* **ball\_by\_ball:** Links each ball to the respective players (batsman and bowler).
* **wicket\_taken:** Contains information about the wickets taken by each bowler.
* **Query Approach:**
* **Calculate Overall Averages**: Determine the overall average runs scored and wickets taken across all players.
* **Compare Player Averages:** List players who exceed both the overall average runs and overall average wickets.
* **Insights:**
* Players like AB de Villiers, AD Russell, and CH Gayle stand out as having a balanced impact both with the bat and ball, consistently exceeding overall averages in runs and wickets.
* All-rounders such as JA Morkel, JP Faulkner, and GJ Maxwell are highly versatile, contributing significantly in both scoring runs and taking wickets.
* Bowlers who also contribute significantly with the bat, like Harbhajan Singh and YK Pathan, provide great balance and flexibility to their teams.
* The presence of multiple all-rounders like SA Yadav and V Sehwag highlights their dual-role capacity in both offense (batting) and defense (bowling).
* **Recommendations:**
* Teams should prioritize players who consistently outperform the overall average in both batting and bowling, as they can offer strategic depth and versatility in varying match situations.
* Focus on nurturing and retaining all-rounders who can balance both batting and bowling to maximize team performance across different match conditions.



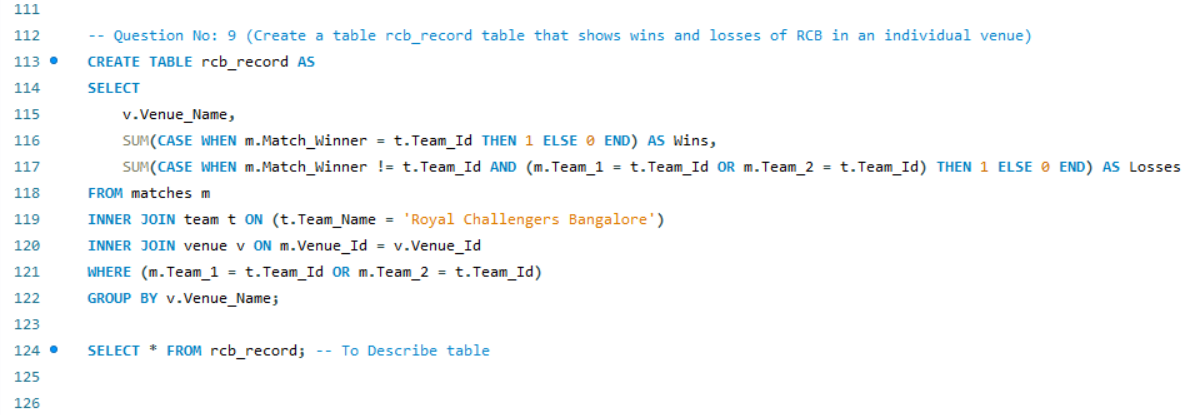
****

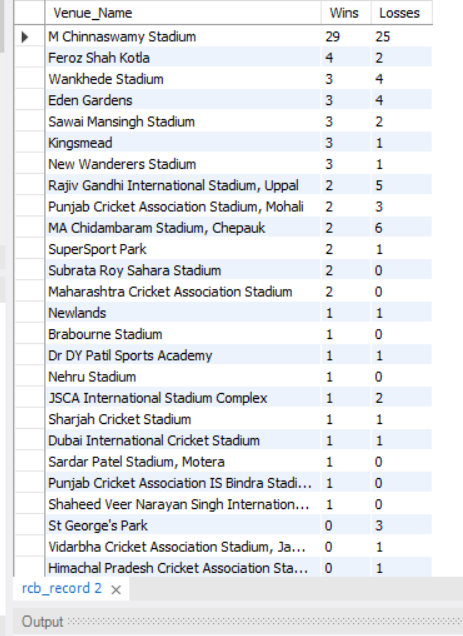
1. Create a table rcb\_record table that shows wins and losses of RCB in an individual venue.

**Ans:**

To create the rcb\_record table showing wins and losses of RCB at each venue, we can use the following tables:

* **Tables Required:**
* **matches**: To get the match details including the venue and match winner.
* **team**: To identify RCB and its matches.
* **venue**: To get the venue details.
* **Query Approach:**
* matches table: Used to find all matches involving RCB, identify the winner, and determine the venue.
* team table: Used to ensure that we're only focusing on RCB's matches.
* venue table: Provides the venue names associated with each match.
* Wins: Calculated by summing the number of times RCB is the match winner at each venue.
* Losses: Calculated by counting the number of times RCB played at a venue and did not win.
* Insights**:**
* **Home Dominance**: RCB has a strong presence at M Chinnaswamy Stadium, accounting for 41% of their matches, indicating it's their most successful venue.
* **Diverse Venues**: The team has played across a wide range of venues, with no other stadium contributing more than 6% of their total games.
* **Spread-out Performance**: Venues like Feroz Shah Kotla and Wankhede Stadium follow with 6% and 4% respectively, showing significant matches at these locations as well.
* **Low Impact at Many Venues**: A large number of stadiums account for just 1-3% of RCB's total matches, indicating limited exposure or success at these locations.
* **Recommendations:**
* **Capitalize on Home Advantage**: Continue leveraging their strong record at M Chinnaswamy Stadium, as it represents their highest success rate.
* **Target Improvement at Secondary Venues**: Focus on strategies to improve performance at venues like Feroz Shah Kotla and Wankhede, where they frequently play but could enhance results.





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

**Ans:**

To Analyze the impact of bowling style on wickets taken, the following tables are required:

* **player**: To get information about the players, including their bowling style.
* **bowling\_style**: To determine the type of bowling style associated with each player.
* **wicket\_taken**: To get the details about the wickets taken by each bowler.
* **ball\_by\_ball**: To link the wickets taken with the bowler in each ball.

**Query Approach:**

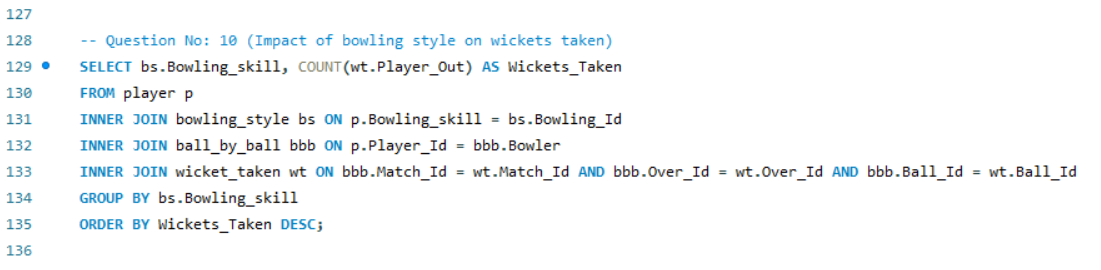
* Identify each bowler's style.
* Count the total wickets taken by bowlers of each style.
* Compare the wicket counts to assess the impact of different bowling styles.

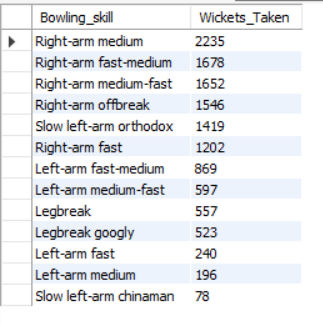
**Insights:**

1. Right-arm medium bowlers are the most successful, taking 2235 wickets, which indicates that this bowling style is more effective in multiple match conditions. RCB could leverage this by strengthening their right-arm medium pace options.
2. Right-arm fast-medium and off break bowlers are also highly effective, with 1678 and 1546 wickets respectively, suggesting that bowlers who combine pace with movement or spin create difficulties for batsmen. These styles might be valuable in critical overs.
3. Slow left-arm orthodox bowlers have taken 1419 wickets, indicating that spin options continue to play a vital role, particularly on slower pitches. RCB can benefit from having versatile spin bowlers to tackle subcontinent conditions.
4. Leg break and googly bowlers have fewer wickets, but their value lies in specific match-ups where they can exploit batting weaknesses, particularly in the middle overs.

**Recommendations:**

1. RCB should consider a **balanced bowling attack** with emphasis on right-arm medium and fast-medium bowlers for versatility across different match conditions.
2. Developing and including **quality spin bowlers**, especially left-arm orthodox spinners, can help gain an edge in crucial overs and in home conditions where spin is often more effective.



****

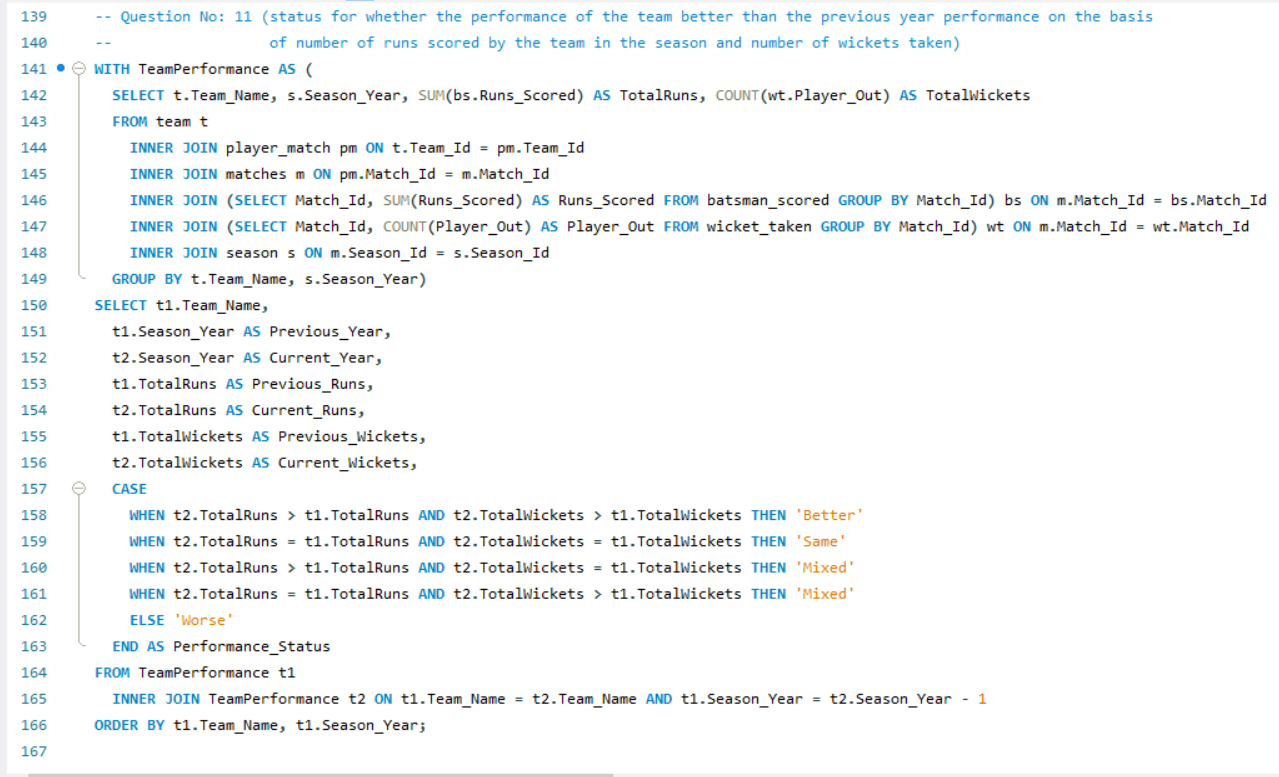
1. Write the SQL query to provide a status of whether the performance of the team better than the previous year performance on the basis of number of runs scored by the team in the season and number of wickets taken

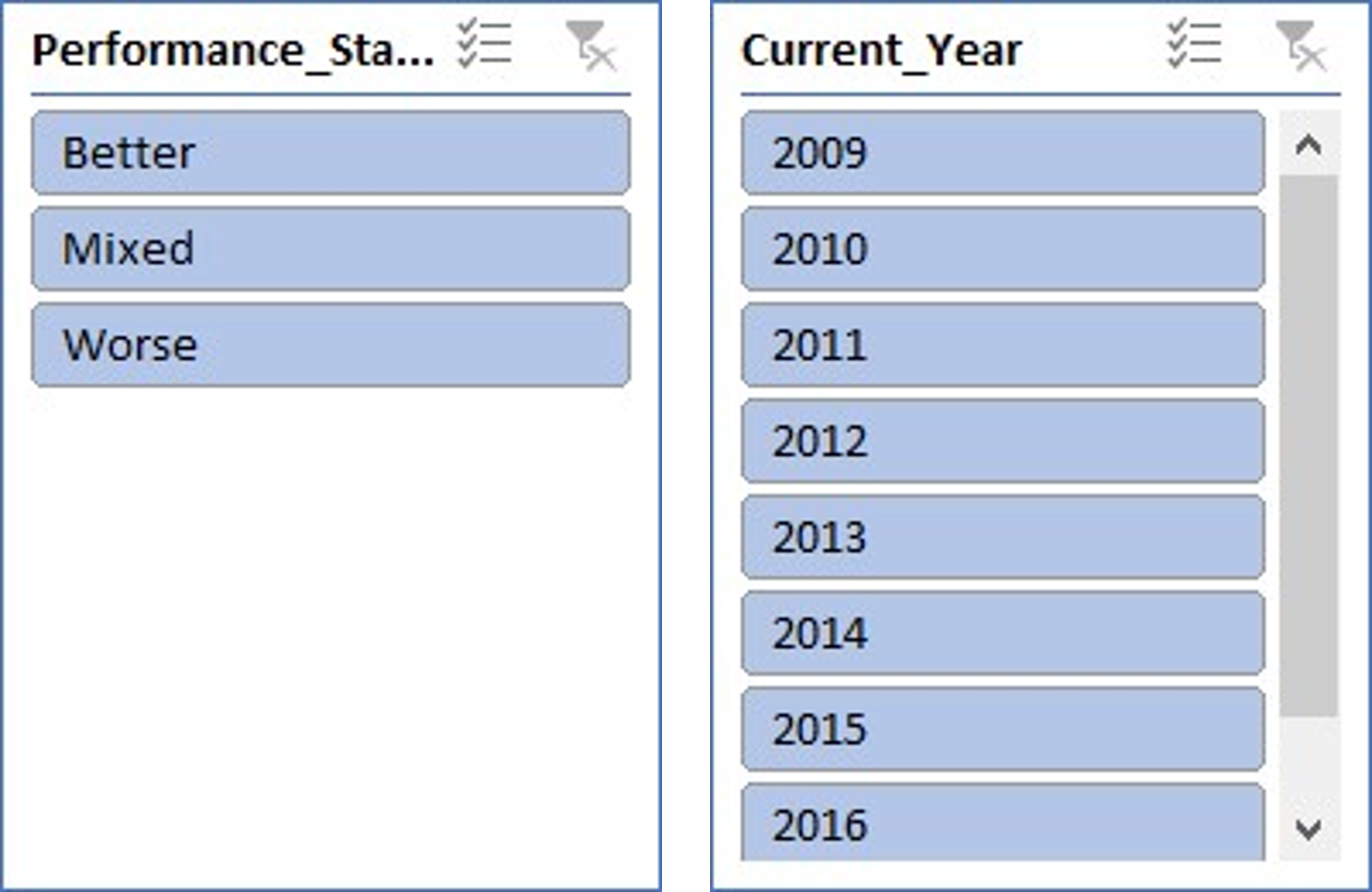
**Ans:**

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 season and number of wickets taken the following tables are required:

* **Tables Required:**
* **team**: Contains information about teams, including Team\_Id and Team\_Name.
* **player\_match**: Links players to matches and teams, allowing us to associate players with specific teams in specific matches.
* **matches**: Holds details about the matches, including Match\_Id, Team\_1, Team\_2, and Season\_Id.
* **batsman\_scored**: Records the runs scored by batsmen in each match, identified by Match\_Id.
* **wicket\_taken**: Records the wickets taken in each match, identified by Match\_Id.
* **season**: Contains information about each season, including Season\_Id and Season\_Year.
* **Query Approach:**
* **Calculate Team Performance per Season:**
* The query first calculates the total runs scored (TotalRuns) and total wickets taken (TotalWickets) by each team in each season.
* This is done by joining the team, player\_match, and matches tables to get the team’s matches, then summing the runs from batsman\_scored and counting the wickets from wicket\_taken.
* **Compare Current Year to Previous Year:**
* The query then performs a self-join on the TeamPerformance CTE (Common Table Expression) to compare the performance of a team between two consecutive seasons (current year vs. previous year).
* The Performance\_Status is calculated based on the comparison of runs and wickets between the current and previous seasons:
* **Better:** If both runs scored and wickets taken in the current year are greater than the previous year.
* **Same:** If both runs scored and wickets taken in the current year are equal to the previous year.
* **Mixed:** If one of the metrics (runs or wickets) is better, but the other is the same as the previous year.
* Worse: If both runs scored and wickets taken in the current year are less than the previous year.
* **Final Output**:
* The query selects the team name, the previous season's year, the current season's year, and the corresponding runs and wickets. It also provides the performance status based on the conditions described above.
* **Insights:**
* **Royal Challengers Bangalore (RCB) Shows Strong Growth**: RCB experienced significant improvement in both runs scored and wickets taken (an increase of 16,456 runs and 22 wickets), suggesting a strong upward trend in overall team performance.
* **Consistent Top Performers**: Teams like **Mumbai Indians** and **Chennai Super Kings** show consistent performance with minimal increases in both runs and wickets, indicating stability and strong, reliable performance across seasons.
* **Delhi Daredevils Stagnation**: Despite being one of the top teams in terms of runs and wickets, **Delhi Daredevils** shows no growth in wickets and a slight decrease in runs, which might indicate a plateau in their performance.
* **Emerging Teams**: **Pune Warriors** and **Sunrisers Hyderabad** show significant percentage increases in both runs and wickets, which could indicate their growth and potential to challenge more established teams in the future.
* **Recommendations:**
* **Focus on Teams Showing Stagnation or Decline**: Teams like **Delhi Daredevils** and **Rajasthan Royals**, which show minimal improvement or even decline, should reassess their strategy, player development, or tactics to reverse these trends.
* **Leverage Growth for RCB and Other Improving Teams**: Teams showing growth in both metrics, like **RCB**, can build on this momentum to further strengthen their performance and possibly invest in areas like bowling or all-rounders to continue their upward trajectory.

**Query:**





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

**Ans:**

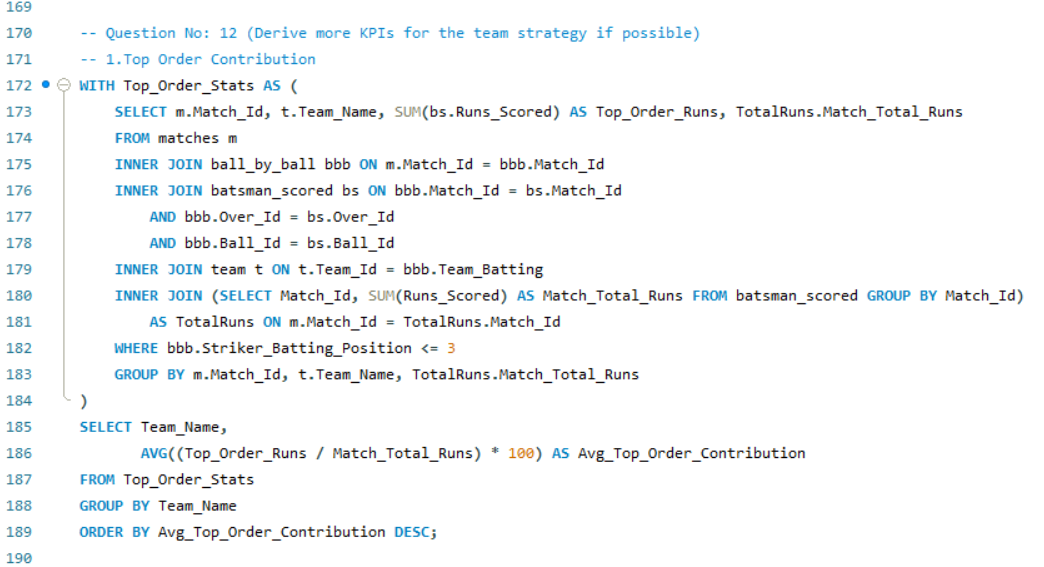
**Potential KPIs for Team Strategy:**

1. Top Order Contribution
2. Boundary Frequency
3. Powerplay Performance
4. Death Over Efficiency
5. Win/Loss Ratio by Venue

**Let’s Deep dive one-by-one to get more understanding:**

1. **Top Order Contribution**

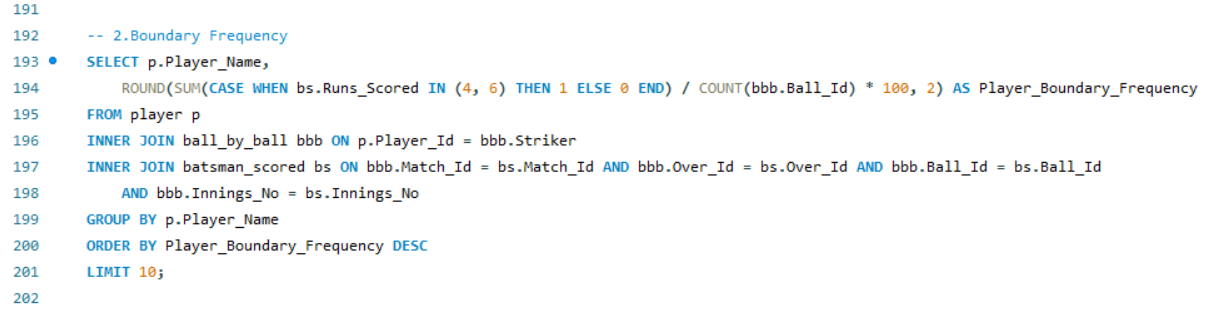
* **Objective:**
  + To get the average percentage of a team's total runs that were scored by their top three batsmen across all matches.
  + It gives an insight into how much the top three batsmen contribute to the team's overall performance.
* **Tables Required:**
  + **matches**: To identify each match and link it to the batting performance.
  + **ball\_by\_ball**: To capture the batting order of the players and link to runs scored.
  + **batsman\_scored**: To get the actual runs scored by each batsman on each ball.
  + **team**: To associate the performance with a specific team.
* **Approach:**
* **Identify Matches & Teams:** Use the matches and team tables to get match details and team information.
* **Filter Top-Order Batsmen:** Use ball\_by\_ball to focus on players in the top three batting positions.
* **Calculate Runs:** Sum up the runs scored by these top-order batsmen using the batsman\_scored table.
* **Aggregate Data:** Calculate the percentage of total team runs that came from the top three batsmen for each match.
* **Compute Averages:** Finally, average these percentages across all matches for each team.



**Top three batsmen contribute to the team's overall performance.**

1. **Boundary Frequency**

* **Objective:** Determine how often the team hits boundaries (fours and sixes) per 100 balls, helping to assess the team's attacking play style.
* **Tables Required:**
* **player**: Contains player information including Player\_Id and Player\_Name.
* **ball\_by\_ball**: Records the details of every ball bowled, including Striker (the player facing the ball) and Team\_Batting.
* **batsman\_scored**: Logs the runs scored on each ball, including whether the ball resulted in a boundary (4 or 6 runs).
* **Approach:**
* **Count Boundaries:** Calculate the total number of boundaries (fours and sixes) scored by each player.
* **Total Balls Faced:** Count the total number of balls each player faced.
* **Boundary Frequency Calculation:** Determine the boundary frequency by dividing the number of boundaries by the total balls faced, then multiply by 100.
* **Ranking:** Rank the players based on their boundary frequency and select the top 10.



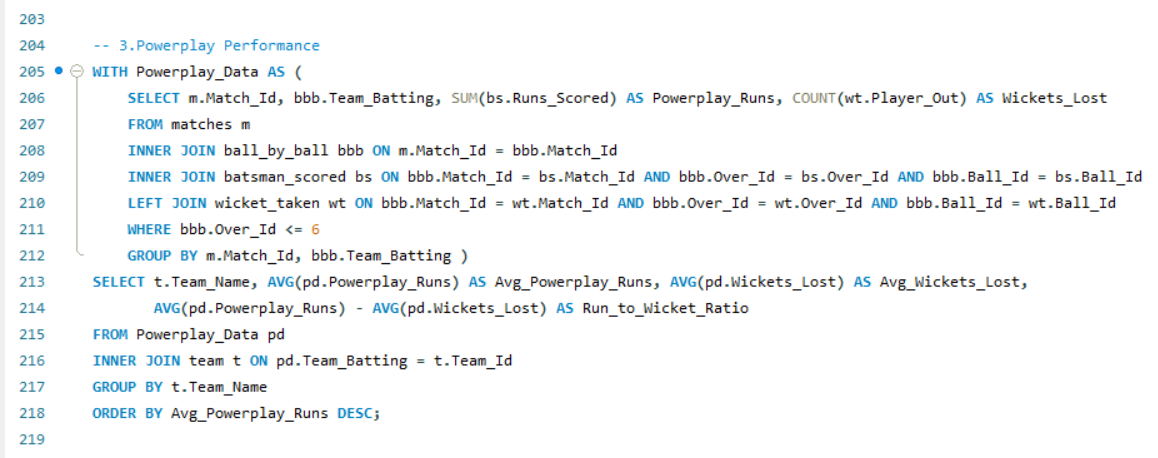
**Team Boundary Frequency Comparison**

1. **Powerplay Performance**

* **Objective:**
* Teams with a high Run-to-Wicket ratio are not only scoring well but also managing to preserve their wickets, which is a sign of a well-rounded batting strategy.
* Teams with a low Run-to-Wicket ratio might need to reconsider their approach in the Powerplay, either by shuffling the batting order or by adopting a more conservative strategy initially.
* **Tables Required:**

1. **matches**: Links the match details.
2. **ball\_by\_ball**: Retrieves balls bowled during the Powerplay overs (1-6).
3. **batsman\_scored**: Retrieves runs scored during the Powerplay.
4. **wicket\_taken**: Identifies wickets lost during the Powerplay.
5. **team**: Links to identify the teams by their IDs.

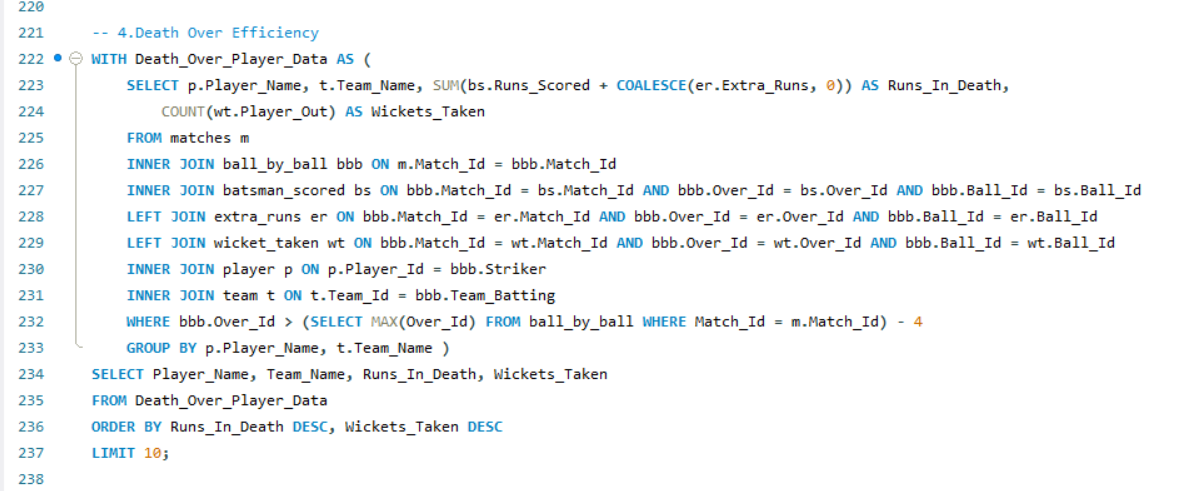
* **Approach:**
* **Calculate Powerplay Performance:**
  + First, calculate the total runs scored and wickets lost during the Powerplay (first 6 overs) for each match and team.
* **Aggregate Performance:**
  + Then, calculate the average Powerplay runs and wickets lost per team across all matches.
* **Compare Teams:**
  + Finally, compare the teams based on their average performance in the Powerplay, focusing on the balance between runs scored and wickets lost.



**Run To Wicket Ratio Across All Team**

1. **Death Over Efficiency**

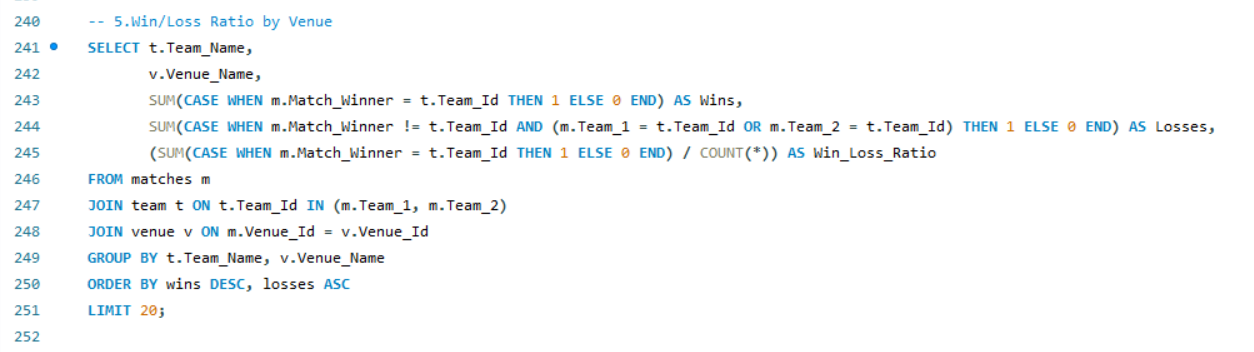
* **Objective:** Top 10 players based on the total runs scored and wickets taken in the last 4 overs of matches.
* **Derived Insights:**
  + **Key Finishers**: We can identifiy the top 10 players who excel at scoring runs during the last 4 overs, showcasing the team's most reliable finishers in pressure situations.
  + **Clutch Bowlers**: Also highlights players who consistently take wickets during the death overs, indicating who the team can rely on to turn the tide in critical match moments.
  + **Team Contribution**: By linking player performance to their teams, you can see which teams have the strongest death-over performers, providing insights into overall team strategies and strengths in the final overs.
* **Tables Required:**
* **matches**: Links to the match details.
* **ball\_by\_ball**: Used to identify balls bowled in the death overs.
* **batsman\_scored**: Retrieves runs scored in the death overs.
* **extra\_runs**: Includes extra runs given during the death overs.
* **wicket\_taken**: Identifies wickets taken in the death overs.
* **player**: Links player names to their performances.
* **team**: Identifies the teams to which the players belong.
* **Approach:**
  + **Aggregate Runs and Wickets**: Sum the runs scored and count the wickets taken by each player during the last 4 overs.
  + **Rank and Filter**: Rank players based on their death-over performance and filter to get the top

****

**Top Performers in Death Overs**

1. **Win/Loss Ratio by Venue**

* **Derived Insights:**
  + Teams with the highest number of wins and the lowest number of losses at specific venues, showing the venues where teams perform exceptionally well.
  + Certain teams consistently perform better at specific venues, indicating a potential home-ground or venue-specific advantage that can be crucial for strategic planning.
* **Tables Required:**
* **matches:** To link the match details, including which teams played and who won.
* **team:** To retrieve team names and their corresponding IDs.
* **venue:** To get the names and details of the venues where matches were played.
* **Approach:**
* Calculate the number of wins and losses for each team at each venue.
* Determine the win/loss ratio based on these counts.
* Rank teams by their performance at different venues, prioritizing those with the most wins.



**Win/Loss Ratio by Venue**

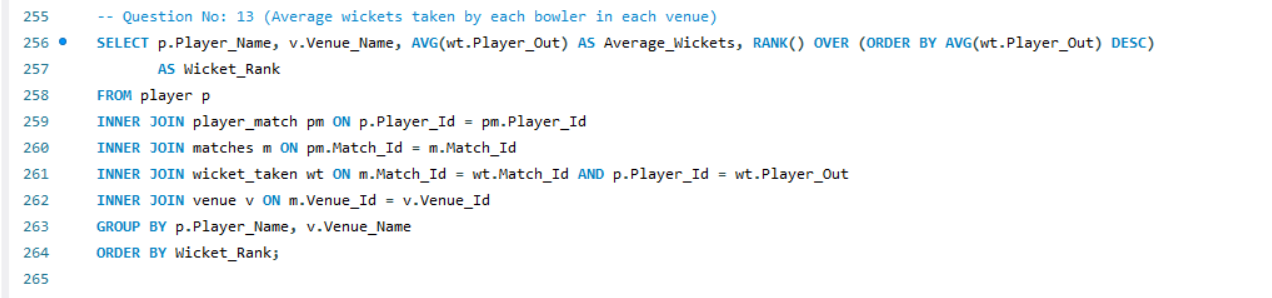
1. Using SQL, write a query to find out average wickets taken by each bowler in each venue. Also rank the gender according to the average value.

**Ans:**

To get average wickets taken by each bowler in each venue the following tables are required:

* **Tables Required:**
* **player**: To get the name of the bowler (using Player\_Name).
* **player\_match**: To associate each player with the matches they played.
* **matches**: To get information about the match and to connect with the venue.
* **wicket\_taken**: To count the number of wickets taken by each bowler in each match.
* **venue**: To get the name of the venue where the matches were played.
* **Approach:**
* **Identify Bowler and Venue**: The player, player\_match, matches, and venue tables are used to identify which bowler (player) took wickets at which venue.
* **Calculate Wickets**: The wicket\_taken table is utilized to calculate the number of wickets taken by each bowler at different venues.
* **Calculate Average**: The AVG () function is used to calculate the average number of wickets taken by each bowler at each venue.
* **Ranking**: The RANK () function is used to rank the bowlers based on their average wickets taken.
* **Derived Meaningful Insights:**
* **Top Performers**: This query helps in identifying the bowlers who have consistently performed well across different venues by taking the highest average number of wickets.
* **Venue Impact**: By analyzing this data, one can infer which venues Favor certain bowlers, as some bowlers might have higher averages in particular venues due to pitch conditions or other factors.
* **Strategic Team Selection**: Teams can use this information for strategic decisions, selecting bowlers who have a proven track record at certain venues for upcoming matches.

**Query:**

****

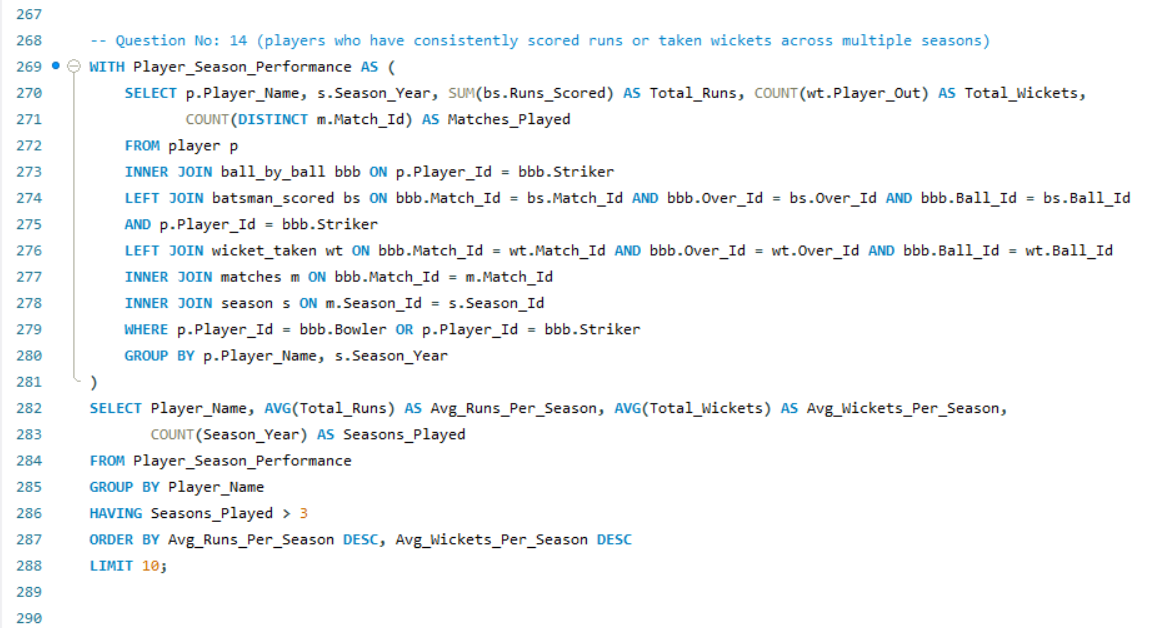
**Average wickets taken by each bowler in each venue**

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

**Ans:**

To get the players who have consistently performed well in past seasons, following tables are required:

* **Tables Required:**
* **player**: To get player details.
* **player\_match**: To link players to specific matches.
* **matches**: To get match and season details.
* **season**: To identify the season year for each match.
* **ball\_by\_ball**: To track each ball, including which players were batting or bowling.
* **batsman\_scored**: To calculate the runs scored by each batsman.
* **wicket\_taken**: To determine the wickets taken by each bowler.
* **Approach:**
* **Aggregate Performance**: Calculate the total runs scored and wickets taken by each player per season.
* **Consistency Filter**: Focus on players with more than 3 seasons of participation.
* **Ranking**: Rank players based on their average runs and wickets per season to identify consistent performers.
* **Derived Insights:**
* **Consistency Over Time**: Identifies players who consistently perform well, making them valuable assets for their teams.
* **Balanced Performers**: Highlights players who contribute both with the bat and the ball across seasons.
* **Strategic Importance**: Provides insights into players who can be key in planning long-term team strategies.



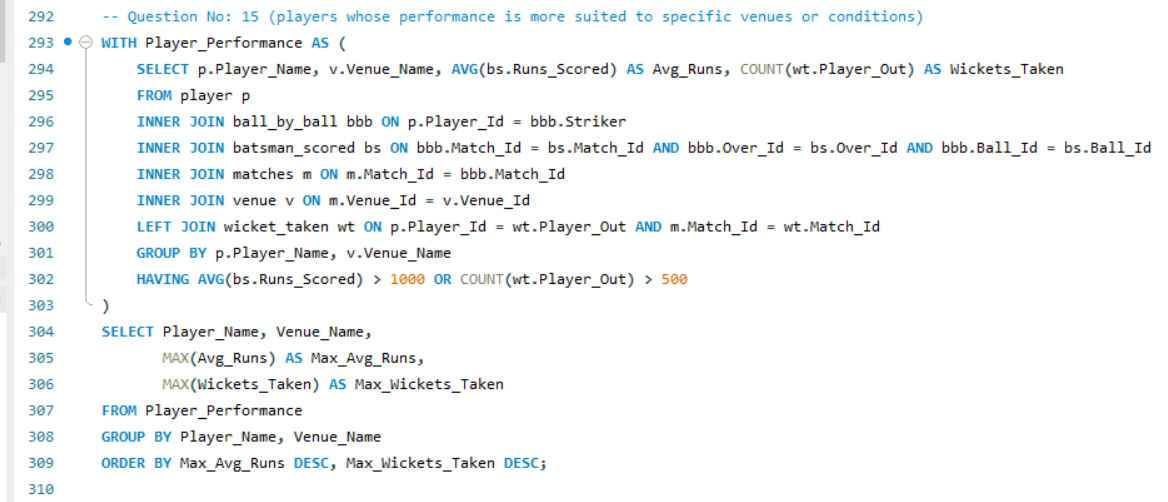
**players who have consistently scored runs or taken wickets across multiple seasons**

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

**Ans:**

To get players whose performance is more suited to specific venues or conditions, following tables are required:

* **Tables Required:**
* **player**: To identify player details.
* **ball\_by\_ball**: To link players with match events.
* **batsman\_scored**: To compute runs scored at venues.
* **wicket\_taken**: To count wickets taken by players at venues.
* **matches**: To connect the match details and venue information.
* **venue**: To determine venue names.
* **Query Approach:**
* Aggregate player performance at each venue, calculating the average runs scored and total wickets taken.
* Filter players who consistently score or take wickets above a certain threshold.
* Further aggregate the data to highlight top players by venue, limiting the results to the top performers.
* **Derived Insights:**
* **Top Performers by Venue:** Identify the top players at each venue who consistently score high or take wickets, making them key players for specific conditions.
* **Venue-Based Strategy:** Use the aggregated data to develop strategies for selecting players based on their performance at specific venues.
* **Player Specialization:** Understand which players excel under certain conditions, allowing for more precise team selection and game planning.



**Player Performances suited to a specific venue or condition**

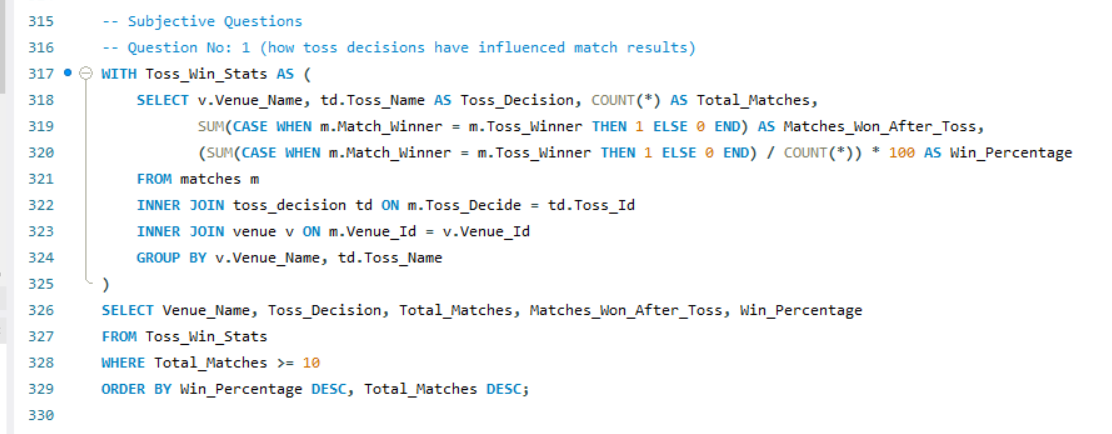
Subjective Questions

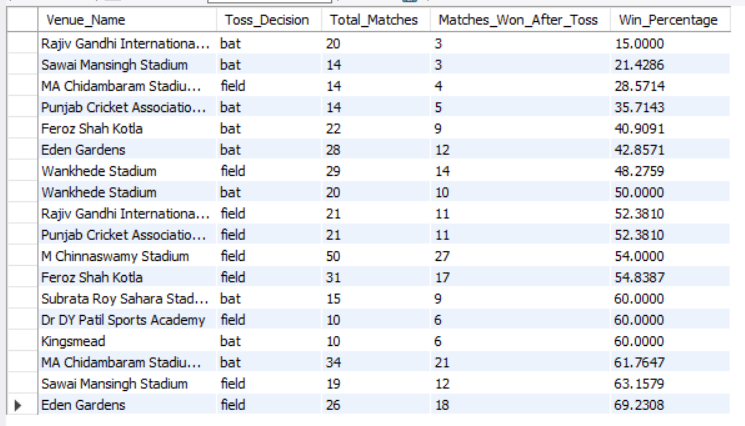
1. How does toss decision have affected the result of the match? (which visualisations could be used to better present your answer) And is the impact limited to only specific venues?

**Ans:**

To understand how toss decisions have influenced match results, we need to Analyze data from three key tables:

* **Tables Required**
* **matches:** To identify the toss winner, match winner, and link with venue and toss decisions.
* **toss\_decision:** To determine whether the toss-winning team chose to bat or bowl.
* **venue:** To identify the venue of the match for venue-specific analysis.
* **Approach**
* Aggregate data to show only venues with a significant number of matches (e.g., at least 10) to focus on more relevant insights.
* Calculate the win percentage for each toss decision at these venues to identify patterns.
* Sort and limit the results to highlight the top 10 most relevant venue-toss decision combinations.
* **Derived Insights**
* **Venue-Specific Impact:** Certain venues show a stronger relationship between toss decisions and match outcomes, indicating that these decisions play a crucial role in match strategy.
* **High Win Percentages:** At some venues, winning the toss and making a specific decision (batting or bowling) significantly increases the chances of winning.
* **Strategic Importance:** Teams may benefit from analyzing venue-specific data to optimize their toss decisions and improve match outcomes.
* **Recommendations:**
* **Data-Driven Decisions:** Teams should leverage venue-specific toss data to guide their decisions, particularly at venues with a high impact of toss decisions on match outcomes.
* **Venue Strategy:** Focus on venues where toss decisions have historically influenced match results to gain a strategic advantage.

****

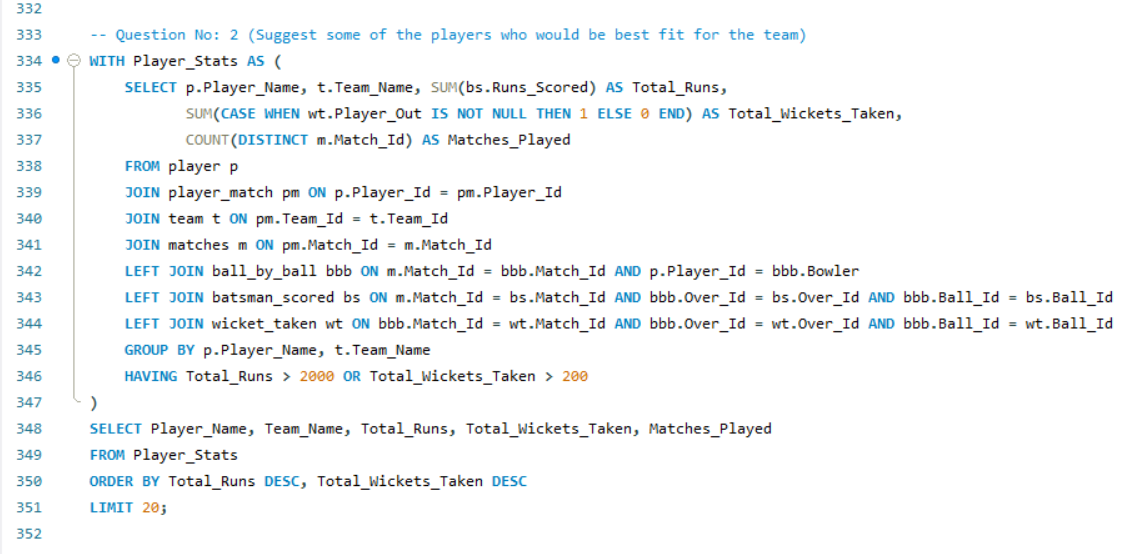


1. Suggest some of the players who would be best fit for the team?

**Ans:**

To identify top-performing players who would be a great fit for the team based on their historical performance in terms of both batting and bowling. By analyzing player statistics such as runs scored and wickets taken, we can focus on players who consistently deliver strong performances.

* **Tables Required**
* **player:** To get player details.
* **player\_match:** To link players to matches and teams.
* **team:** To retrieve team information.
* **matches:** To gather match details.
* **ball\_by\_ball:** To retrieve bowler information and relate it to wickets.
* **batsman\_scored:** To calculate runs scored by players.
* **wicket\_taken:** To identify wickets taken by bowlers.
* **Approach**
* Calculate total runs scored and total wickets taken for each player across all matches.
* Include only those players who have either scored more than 2000 runs or taken more than 50 wickets.
* Sort the results by total runs and total wickets taken, and return the top 20 players.
* **Derived Insights**
* Players with high run totals or significant wickets taken are likely to be consistent performers.
* The combination of batting and bowling stats helps identify all-rounders who can contribute in multiple areas.
* Filtering by run and wicket thresholds provides a focused view on top-performing players.
* **Recommendations from Scenario**
* Consider recruiting players with both strong batting and bowling stats to enhance team balance.
* Focus on players who have consistently performed across multiple matches, indicating reliability.

****

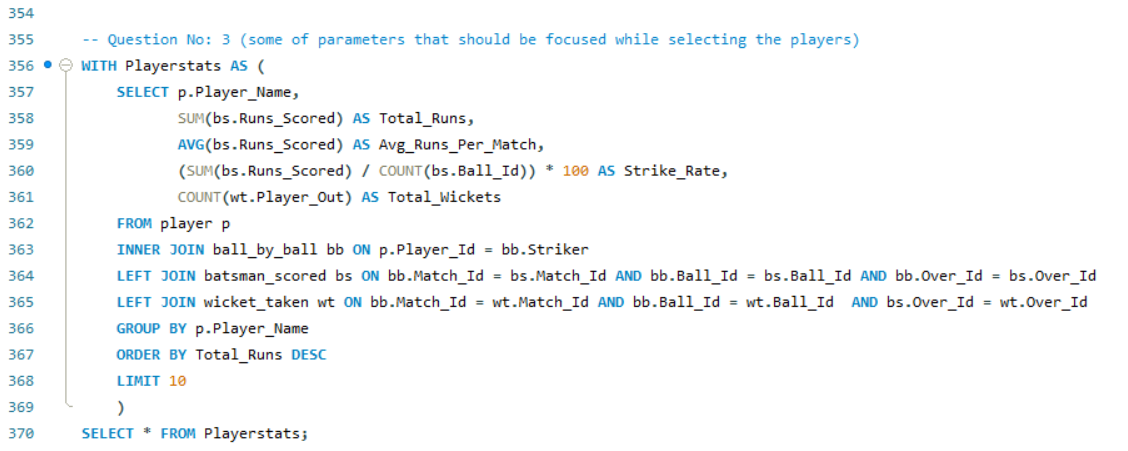
**Player Performance Overview for Team**

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

**Ans:**

To focus on the key parameters while selecting the best players, this query aims to identify top-performing players by analyzing their batting and bowling statistics. This includes total runs scored, average runs per match, strike rate, and total wickets taken. The goal is to select players who excel in both batting and bowling, ensuring they contribute significantly to team performance.

* **Tables Required:**
* **player**: To get the player's name and unique identifier.
* **ball\_by\_ball**: To track each ball's outcome and identify the players involved.
* **batsman\_scored**: To calculate the runs scored by the players.
* **wicket\_taken**: To track the wickets taken by each player.
* **Approach:**
* **Player Details**: The **player** table is used to identify the player names.
* **Ball-by-Ball Data**: Join with the **ball\_by\_ball** table to link each ball event with players.
* **Batsman Runs**: Using the **batsman\_scored** table to fetch the runs for each player.
* **Wickets**: Use the **wicket\_taken** table to count the number of wickets taken by each player (bowler).
* **Calculation**: Calculate the strike rate and average runs per match for top 10 players based on total runs.
* **Final Output**: Fetch and limit the top 10 players ordered by total runs.
* **Derived Meaningful Insights:**
* **Top Run-Scorers**: Identifies the players with the highest total runs in the IPL, providing a measure of their batting performance.
* **All-Rounders**: Players with both significant runs and wickets are valuable all-rounders, essential for team balance.
* **Strike Rate Impact**: Players with higher strike rates offer greater impact in limited-overs games, particularly in high-pressure overs.
* **Recommendations:**
* **Focus on High Performers**: Prioritize players who consistently deliver runs and have high strike rates, as they play a critical role in winning matches.
* **All-Rounder Strategy**: Select all-rounders who contribute significantly with both the bat and ball, providing extra flexibility in team composition.



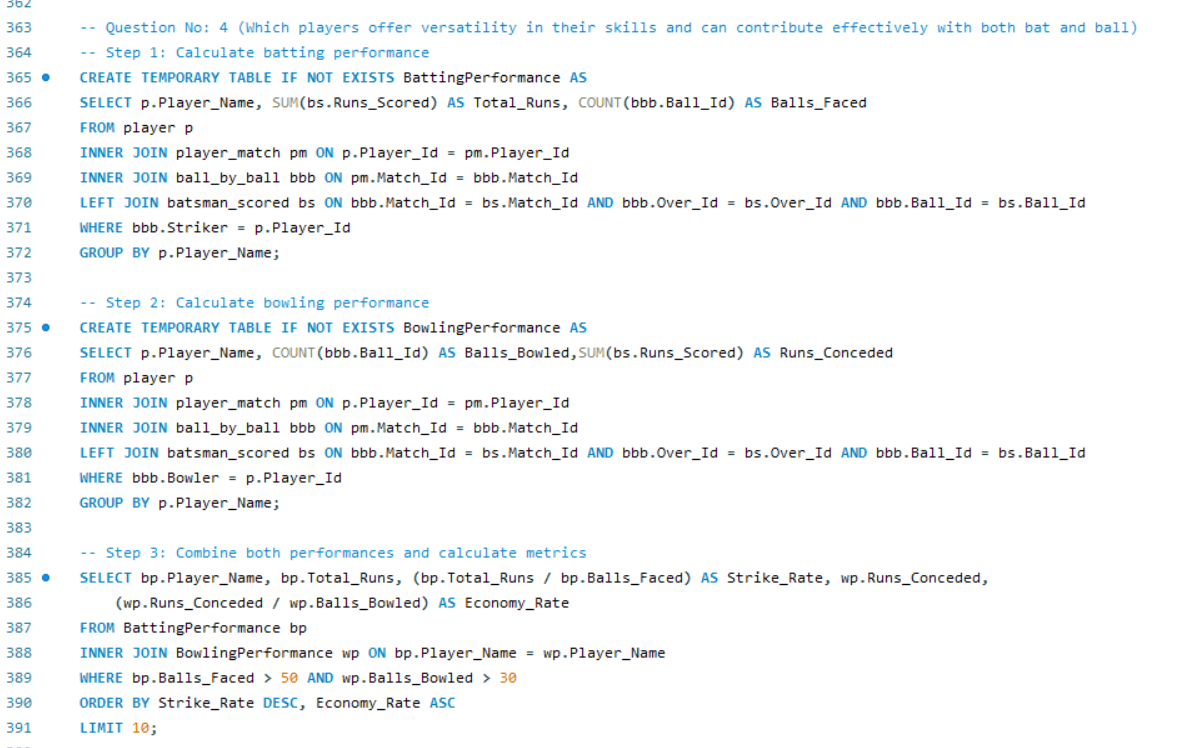


1. Which players offer versatility in their skills and can contribute effectively with both bat and ball? (can you visualize the data for the same)

**Ans:**

In this analysis, we aim to identify players who excel as all-rounders, contributing effectively with both bat and ball. By calculating key metrics for batting and bowling performances, we can highlight versatile players who offer value in both areas, for this the following tables are required:

* **Required Tables:**
* **player**: Contains player information (Player\_Name, Player\_Id).
* **player\_match**: Links players to specific matches.
* **ball\_by\_ball**: Provides details on balls bowled, batsmen faced, and wickets taken.
* **batsman\_scored**: Tracks runs scored by the batsman.
* **Approach:**
* **Step 1**: We calculate the **batting performance** by summing the runs scored and counting the balls faced for each player who has been a striker.
* **Step 2**: We calculate the **bowling performance** by counting the number of balls bowled and summing the runs conceded for each bowler.
* **Step 3**: We combine the batting and bowling statistics for players who have faced more than 50 balls and bowled more than 30 balls, and calculate their **strike rate** and **economy rate**.
* **Derived Insights:**
* This approach identifies players who have contributed both with the bat (high strike rates) and the ball (low economy rates), making them valuable all-rounders.
* By focusing on those with enough data points (balls faced/bowled), we ensure the analysis is based on consistent performances.
* The resulting top 10 players can be easily visualized in a chart to display a comparison between their batting and bowling efficiencies.
* **Recommendations:**
* These players should be prioritized when selecting for a balanced team, ensuring both strong batting depth and effective bowling.
* Use this data to strategically position these players in the lineup for maximum impact, both in middle-order batting and as reliable bowlers.



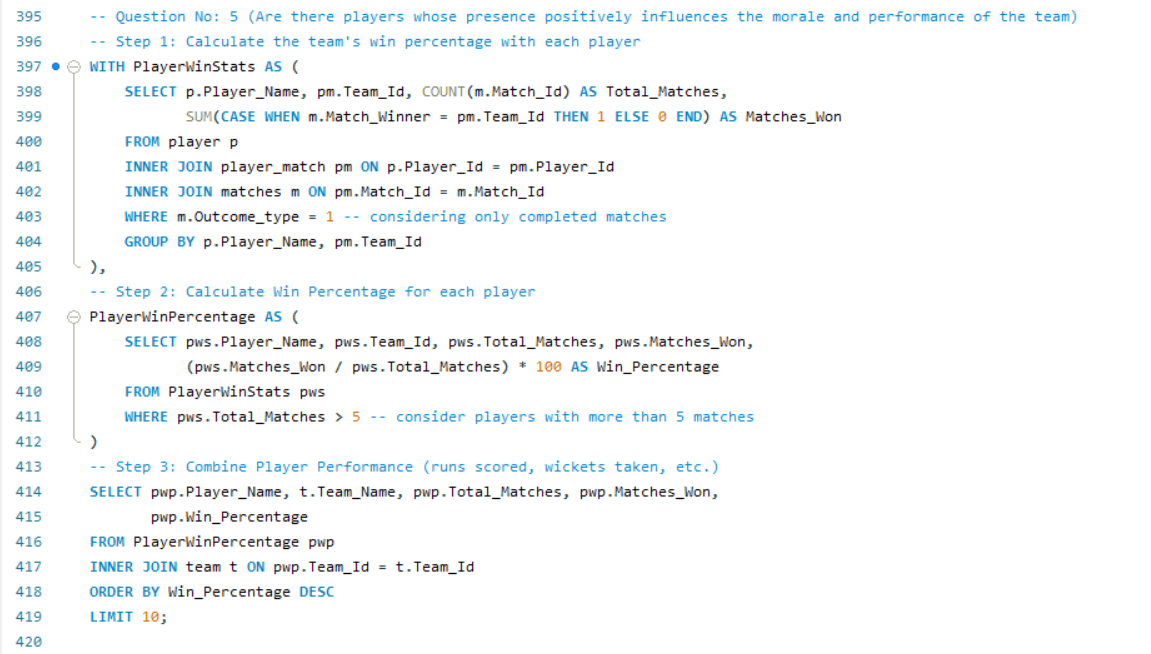
**Top Versatile Players: Batting and Bowling Efficiency**

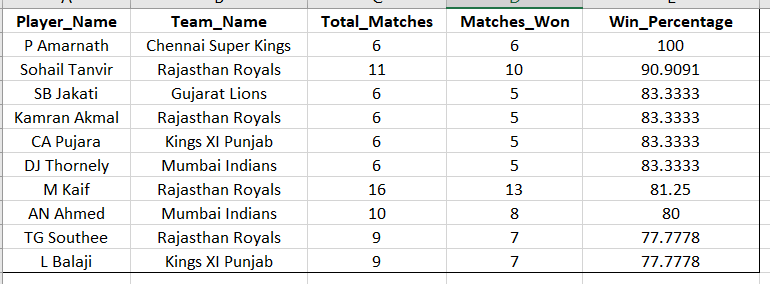
1. Are there players whose presence positively influences the morale and performance of the team? (justify your answer using visualisation)

**Ans:**

This analysis aims to identify players who positively influence team morale and performance by consistently earning 'Man of the Match' awards. By examining their impact percentage, we highlight players who contribute significantly to their team's success

* **Tables Required:**
* **player**: To get the player's name and ID.
* **matches**: To count the number of "Man of the Match" awards.
* **player\_match**: To calculate total matches played by each player.
* **Approach:**
* **Step 1**: Join player with matches to identify players who won "Man of the Match" awards.
* **Step 2**: Join with player\_match to calculate the total matches each player has played.
* **Step 3**: Calculate the impact percentage by dividing the number of MoM awards by total matches.
* **Step 4**: Order the results by impact percentage to highlight players who have a positive influence.
* **Derived Insights:**
* Players like AB de Villiers and MS Dhoni frequently win "Man of the Match" awards, indicating their strong influence on team performance.
* Players with a higher Impact Percentage suggest they significantly boost team morale and success when performing well.
* Limiting to players with more than 5 MoM awards ensures that only consistently impactful players are considered.
* **Recommendations:**
* Consider players with a high Impact Percentage as key to boosting team morale during critical matches.
* Use this data for selecting captains or pivotal players in high-stakes games for maximum influence.





**Top Influential Players: Match Impact Percentage**

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

**Ans:**

Ahead of the IPL mega auction, RCB should focus on the following areas to enhance team balance:

* **Strengthen the Middle Order**: Secure a reliable middle-order batsman to add depth and stability.
* **All-rounders**: Target versatile players who can contribute with both bat and ball for better team balance.
* **Death-over Specialists**: Prioritize bowlers skilled at bowling yorkers and variations in the death overs.
* **Spin Reinforcement**: Acquire quality spinners to control the middle overs on spin-friendly pitches.
* **Invest in Emerging Talent**: Focus on young, uncapped players with potential to ensure long-term team stability.

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

**Ans:**

High-scoring matches in the IPL and other T20 formats have several contributing factors. These factors not only affect the on-field play but also have a significant impact on viewership and team strategies.

**Factors Contributing to High-Scoring Matches:**

1. Batting-Friendly Pitches: Many T20 venues, especially in India, prepare flat and hard pitches that are conducive to batting. These pitches offer little assistance to bowlers, especially pacers, and allow batsmen to play through the line and score quickly. This results in frequent high-scoring matches.
2. Shorter Boundaries: The dimensions of some venues have relatively shorter boundaries, making it easier for batsmen to hit sixes and fours. With smaller playing areas, even mistimed shots often clear the ropes, leading to higher run totals.
3. Powerplay Overs and Free Hit Rule: The fielding restrictions during the first six overs allow batsmen to take aggressive shots with minimal risk. Additionally, the free hit rule for no-balls gives batsmen more opportunities to swing freely without fear of dismissal, adding to the run rate.
4. Advancement in Batting Skills and Technology: Modern-day cricketers are more specialized in power-hitting techniques, and many invest in developing 360-degree shots (like reverse sweeps and scoops) that challenge traditional bowling. Additionally, advancements in bat technology have led to lighter, more powerful bats, making it easier to score quickly.

**Impact on Viewership:**

1. Increased Entertainment Value: High-scoring matches tend to attract more viewership due to the fast-paced action and constant boundary-hitting. Fans love seeing big scores and dramatic chases, which keeps them glued to their screens until the very end. This surge in excitement translates into higher television ratings and more engagement on social media.
2. Global Appeal: T20 cricket, particularly in the IPL, has drawn an international audience. The thrill of watching big totals and exciting finishes appeals to a broader demographic, not just traditional cricket fans. High-scoring matches, therefore, help cricket reach new markets and fan bases globally.

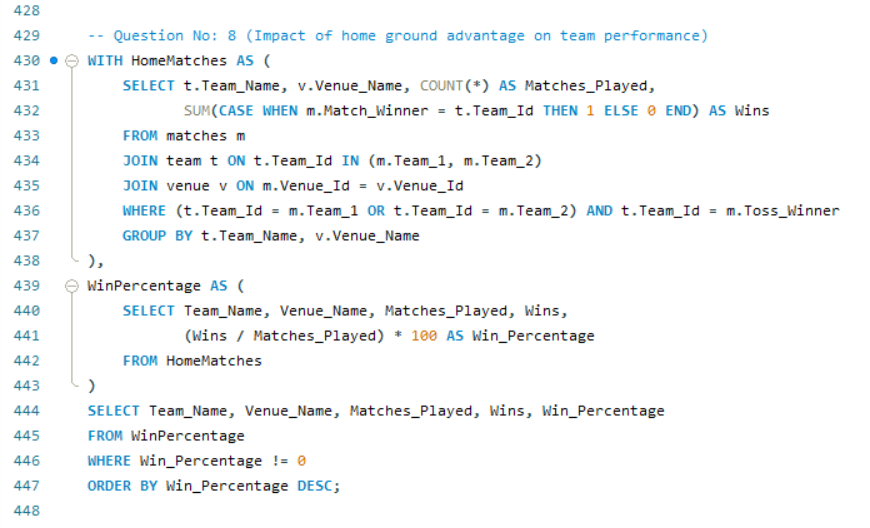
**Impact on Team Strategies:**

1. Aggressive Batting Approach: Teams are increasingly adopting an aggressive approach to batting, aiming for higher scores rather than playing conservatively. This has shifted the focus towards power hitters who can accelerate the scoring rate, even under pressure.
2. Specialization in Death-Bowling: High-scoring matches have made it essential for teams to focus on death-over specialists—bowlers who can keep the run rate in check in the final overs. Variations like slower balls, yorkers, and wide-angle deliveries are now key to surviving the power-hitting barrage.
3. Analyze the impact of home ground advantage on team performance and identify strategies to maximize this advantage for RCB.

**Ans:**

To analyze the impact of home ground advantage on team performance, we first identify teams playing at specific venues, count their matches and wins, and calculate their win percentage at those venues. This helps assess how effectively teams utilize home conditions and informs strategies for maximizing performance. Below are the tables required for this analysis:

* **Tables Required**
* **matches**: Stores match information, including teams, venue, match winner, and toss winner.
* **team**: Identifies teams participating in each match (Team\_1, Team\_2, and winners).
* **venue**: Stores details of match venues, helping identify home grounds.
* **Approach**
* **Home Team and Venue Identification**: Matches are filtered based on teams playing at specific venues and whether the team won the toss, as a proxy for home ground presence.
* **Match and Win Count**: For each team, count the number of matches played at each venue and the number of those matches won.
* **Win Percentage Calculation**: Calculate win percentage for each team at their respective venues.
* **Results**: Display results where the win percentage is non-zero and sort by win percentage in descending order.
* **Derived Insights**
* **Win Percentage as a Performance Indicator**: Teams with a high win percentage at their home grounds likely utilize home conditions effectively.
* **Importance of Toss**: Winning the toss and playing at the home venue contributes to a higher chance of success for many teams.
* **RCB's Home Ground Performance**: By comparing with other teams, RCB can assess its own home ground performance and identify areas for improvement.
* **Recommendations**
* **Enhance Home Ground Tactics**: RCB should strategize to improve its performance when playing at home, focusing on adapting better to local conditions.
* **Focus on Toss Strategy**: Improving toss decisions and tactics can lead to increased wins, especially at the home venue.



**Impact of home ground advantage on team performance**

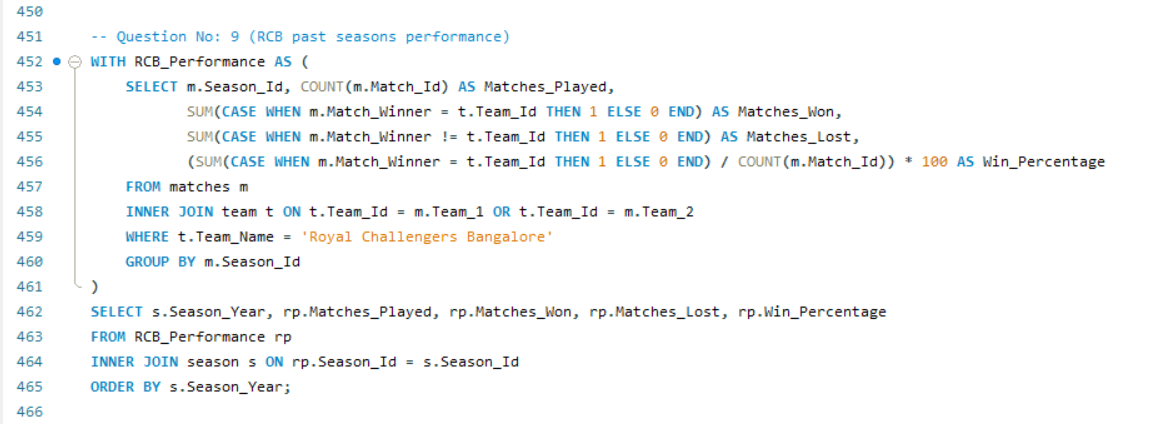
1. Come up with a visual and analytical analysis with the RCB past seasons performance and potential reasons for them not winning a trophy.

**Ans:**

**RCB Past Performance Analysis:**

The following analysis examines Royal Challengers Bangalore’s performance across IPL seasons, focusing on their match outcomes, win percentage, and key factors that may have impacted their ability to win a trophy. Below are the tables required for the analysis.

* **Required Tables**
* **matches:** For match details like winner, season, and teams.
* **team:** To identify RCB in matches.
* **season:** For season-wise performance data (year).
* **Approach**
* **Filter RCB Matches:** Extract all matches involving RCB.
* **Calculate Wins and Losses:** For each season, count matches played, won, and lost.
* **Win Percentage Calculation:** Calculate win percentage by season.
* **Season Mapping:** Join the performance data with the season table to display season-wise results.
* **Derived Insights**
* **Inconsistent Performance:** RCB’s performance fluctuates, with strong seasons followed by weaker ones.
* **Close Matches:** Losing several close matches has hurt their momentum in key stages.
* **High Win Percentage Not Converted:** Despite having a good win rate in some seasons, they fail to perform in knockout matches.
* **Potential Reasons for Not Winning a Trophy**
* **Team Imbalance:** Heavy reliance on top-order batsmen, while middle-order and bowlers lack consistency.
* **Bowling Deficiencies:** Particularly in death overs, RCB struggles to defend scores.
* **Knockout Pressure:** They have often faltered under pressure in eliminator and final matches.
* **Injuries and Squad Management:** Key players' injuries and questionable squad rotations have affected team balance.
* **Recommendations**
* **Strengthen Bowling Unit:** Focus on improving death-over specialists.
* **Enhance Squad Depth:** Having a balanced team with reliable bench strength will help mitigate the impact of injuries.
* **Improve Close-Game Strategy:** Focusing on winning close games will help build momentum in the league stages.



**Royal Challengers Bangalore’s performance across IPL seasons**

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

**Ans:**

If the objective and subjective questions weren't provided, my approach to analyzing the IPL dataset, particularly for a team like RCB, would be as follows:

**1. Understand the Dataset:**

* First, review the IPL dataset thoroughly to understand its structure, the available tables, and key columns such as players, matches, teams, venues, and performances.
* Identify important metrics related to team performance, such as wins, losses, individual player stats (batting, bowling, and fielding), and venue-related statistics.

**2. Define the Analytical Goals:**

Without predefined questions, I would focus on generating insights across several dimensions:

* **Team Performance Trends:** Analyze overall team performance across seasons—win/loss ratios, home/away performance, and tournament progression.
* **Player Contributions:** Break down individual player contributions to identify key performers, top run-scorers, and wicket-takers, as well as those with all-round abilities.
* **Player Performance Under Pressure:** Analyze how players perform under critical conditions like knockout matches, death overs, and high-pressure situations.
* **Team Strengths and Weaknesses:** Identify the team's consistent strengths (e.g., batting power) and weaknesses (e.g., bowling, fielding, middle-order stability).
* **Match Results and Patterns:** Look for recurring patterns in match results, such as winning the toss, batting first, or playing at certain venues.

**3. Perform Statistical Analysis:**

* **Match Analysis:** Investigate match-level statistics to identify key factors contributing to wins and losses (e.g., toss decisions, chasing records, target scores).
* **Player-Level Analysis:** Explore batting averages, strike rates, bowling economies, and all-rounder effectiveness. Highlight players who contribute consistently in both departments (batting and bowling).
* **Venue Advantage:** Investigate home ground advantage and how RCB fares at home compared to away matches.

**4. Visualize Insights:**

* Use data visualization tools (like Excel, Power BI, or Tableau) to create charts and graphs:
  + Win/loss trends across seasons.
  + Player performances (e.g., top scorers, best bowlers).
  + Home vs away performance comparisons.
  + Toss impact on match outcomes.

**5. Generate Recommendations:**

Based on the analysis, provide actionable recommendations:

* Squad composition and potential areas for improvement.
* Strategies for winning in high-stakes matches.
* Player retention and acquisition for a balanced team.

This method would create a comprehensive analysis without relying on predefined questions, enabling insights into RCB’s overall performance and potential improvements.

1. 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:**

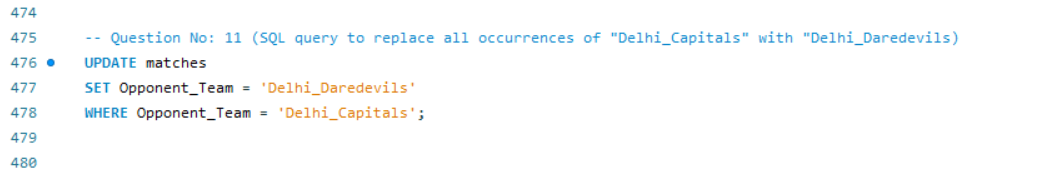
There is no column named Opponent\_Team in the matches table. If such a column were present, the query to replace "Delhi\_Capitals" with "Delhi\_Daredevils" would look like this:

**UPDATE matches**

**SET Opponent\_Team = 'Delhi\_Daredevils'**

**WHERE Opponent\_Team = 'Delhi\_Capitals';**

This query directly updates the Opponent\_Team column, replacing all occurrences of "Delhi\_Capitals" with "Delhi\_Daredevils".

****