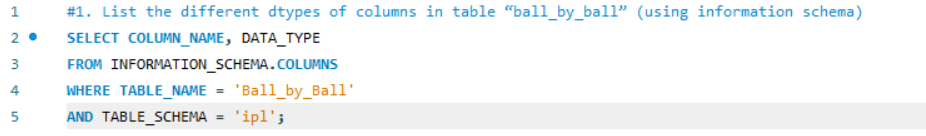
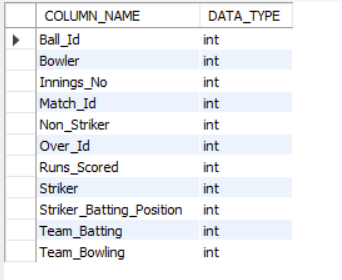
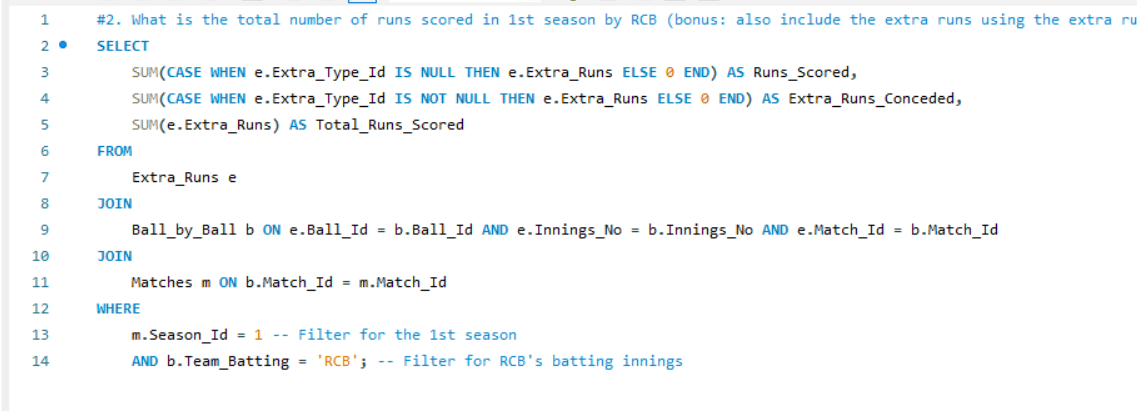
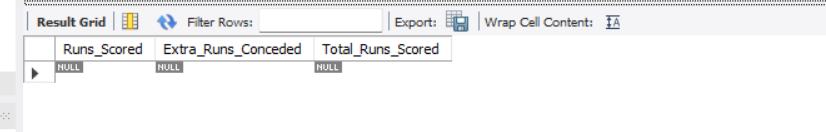
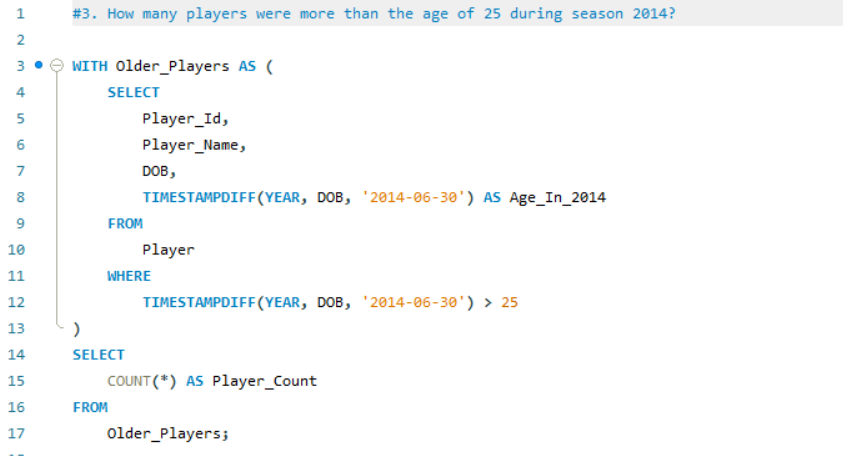
**Learners have to come up with a Report to support the answers to the following questions and suggestions**

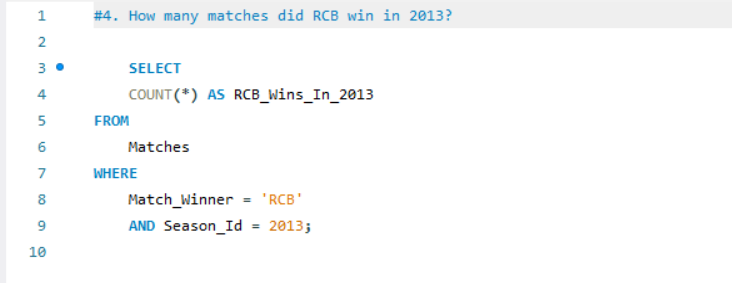
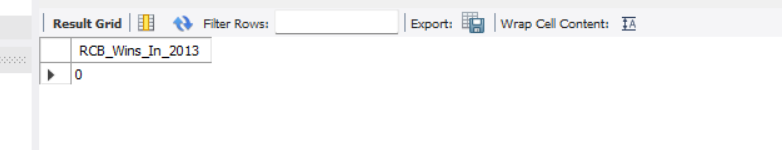
Objective Questions

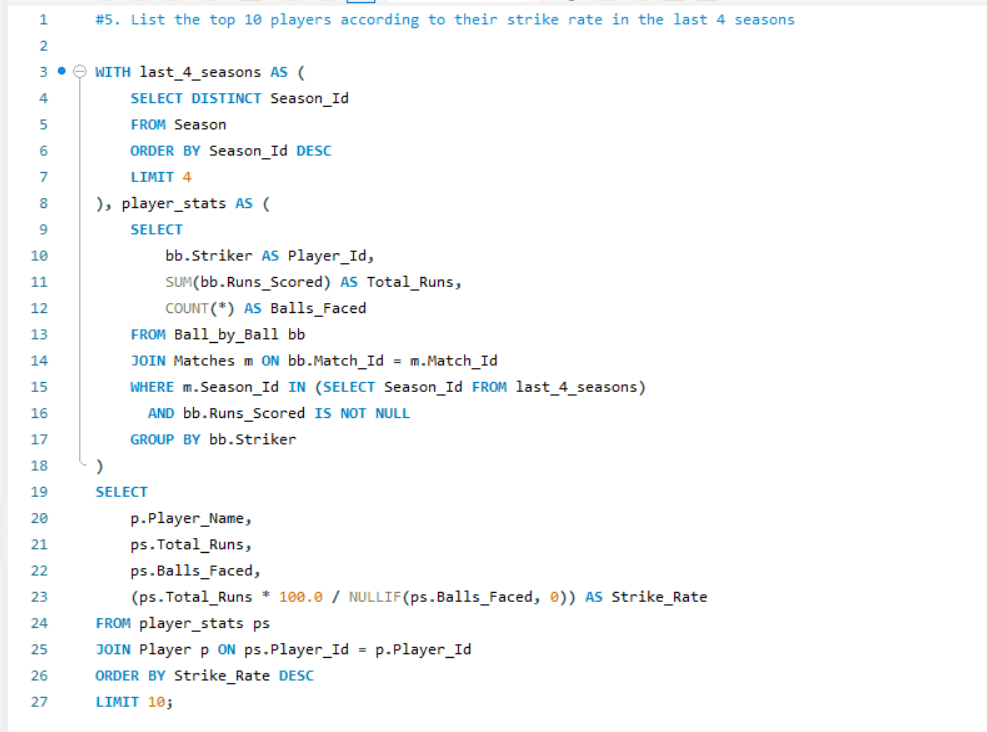
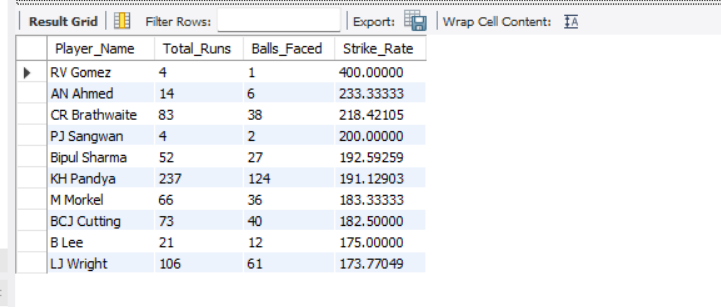
1. List the different dtypes of columns in table “ball\_by\_ball” (using information schema)  
     
   **Data Types:** Data types classify data and inform the compiler or interpreter how the programmer intends to use it.  
     
   

  
This lists all column names and their corresponding data types for the **'ball\_by\_ball'** table. It helps in understanding the table structure, which is useful when writing queries or performing data analysis.

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)  
     
     
     
     
     
   
2. How many players were more than the age of 25 during season 2014?  
      
   To count players older than 25 as of June 30, 2014, use:

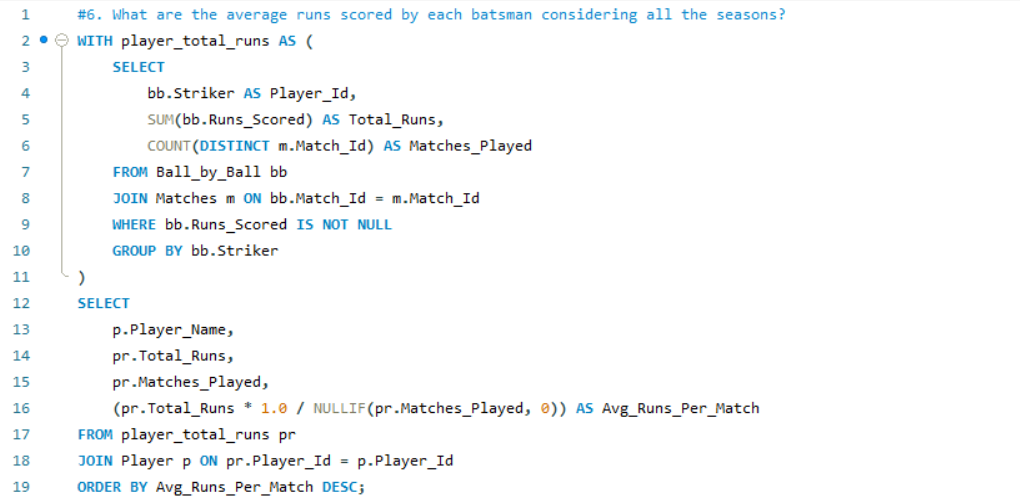
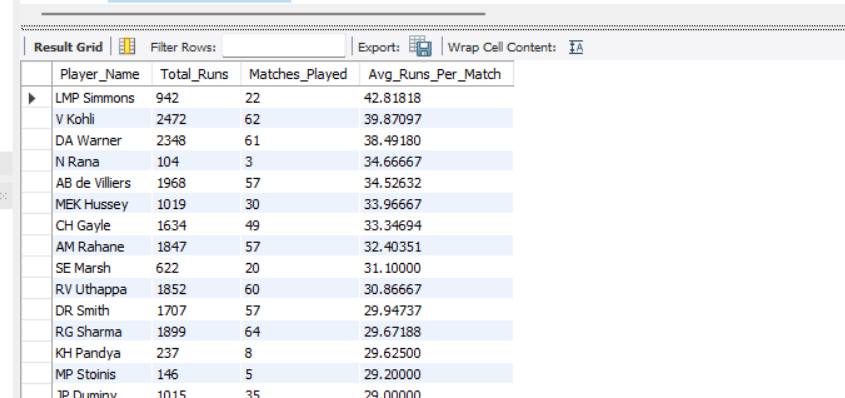
This calculates age using TIMESTAMPDIFF and filters players older than 25, returning the total count. For detailed results, use subqueries or CTEs.  
  
  


1. How many matches did RCB win in 2013?   
     
     
     
     
   1. SELECT COUNT(\*) AS RCB\_Wins\_In\_2013: Counts total rows where conditions are met, labels result as "RCB\_Wins\_In\_2013".  
   FROM Matches: Specifies the "Matches" table as the data source.  
   WHERE Match\_Winner = 'RCB' AND Season\_Id = 2013: Filters rows where RCB won in the 2013 season.  
   Output: Returns the total number of RCB wins in 2013.

1. List the top 10 players according to their strike rate in the last 4 seasons  
     
     
     
     
     
   last\_4\_seasons: Selects the last 4 distinct season IDs from the Season table, ordered descending.

player\_stats: Calculates total runs and balls faced by each player in matches from the last 4 seasons, excluding null runs.

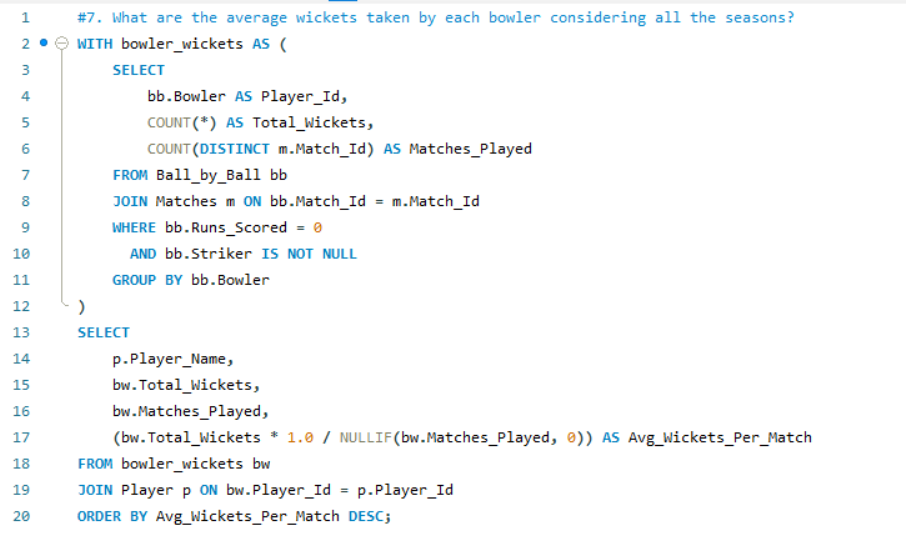
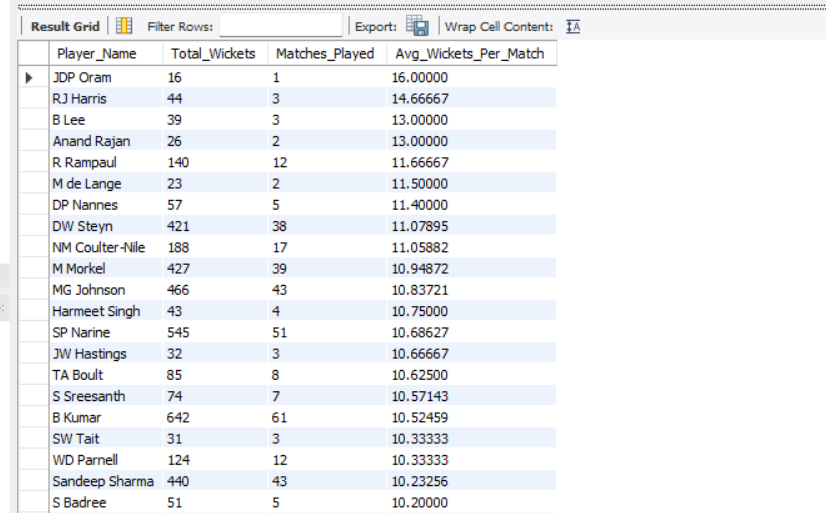
Final Query: Joins player\_stats with Player table, computes strike rate, and returns top 10 players by strike rate.

1. What are the average runs scored by each batsman considering all the seasons?  
     
     
     
     
   CTE Creation: player\_total\_runs calculates total runs and matches played per player using Ball\_by\_Ball and Matches tables.

Main Query: Joins player\_total\_runs with Player table to fetch player names.

Calculation: Computes average runs per match.

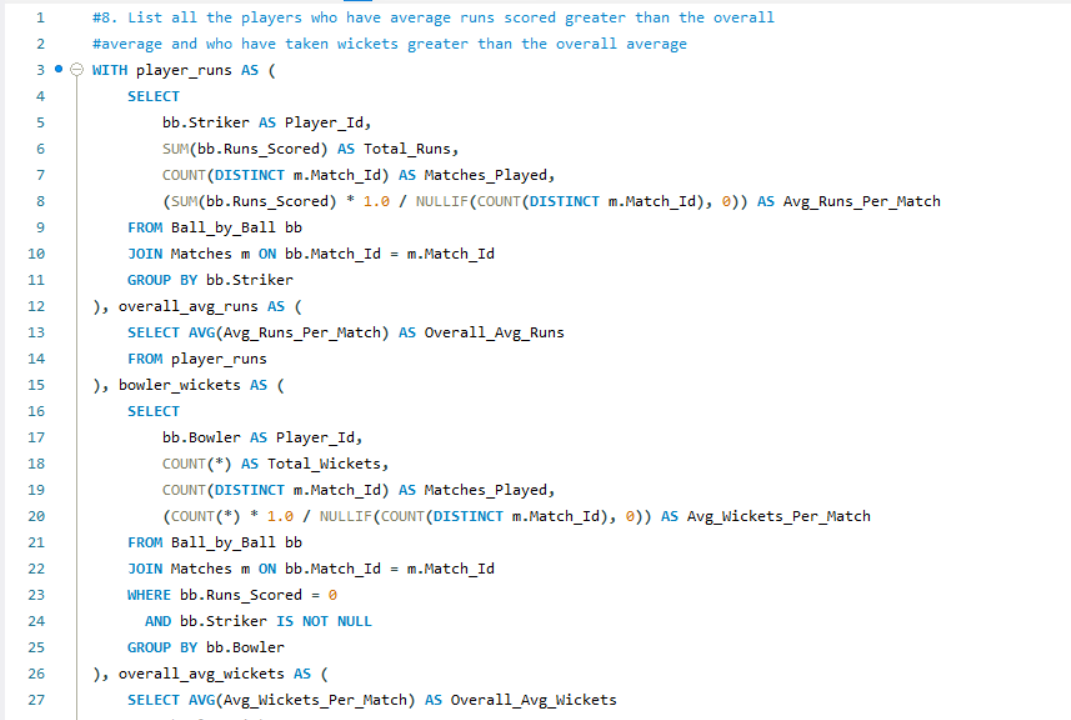
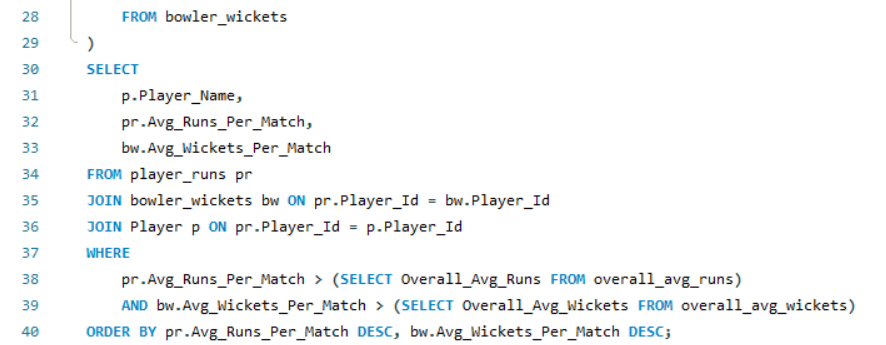
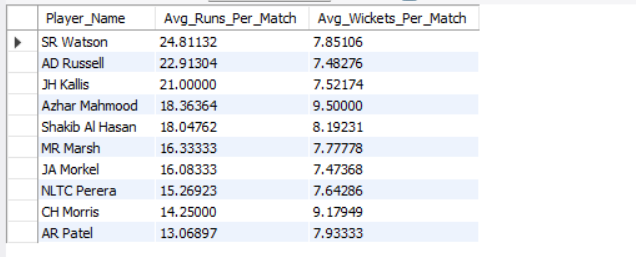
Sorting: Orders results by Avg\_Runs\_Per\_Match in descending order.

1. What are the average wickets taken by each bowler considering all the seasons?  
     
     
     
   CTE Creation: bowler\_wickets calculates total wickets and matches played for each bowler where runs scored are zero and striker is not null.

Main Query: Joins bowler\_wickets with Player table to fetch player names.

Calculation: Computes average wickets per match.

Ordering: Results are ordered by average wickets per match in descending order.

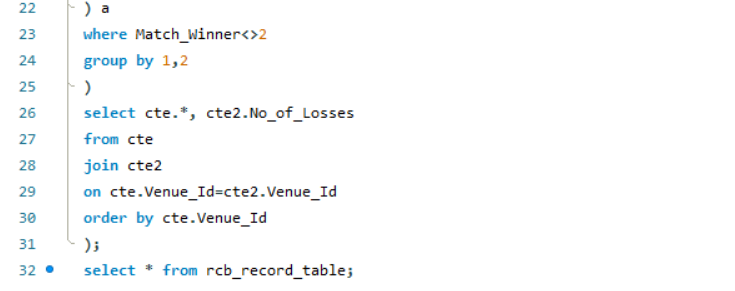
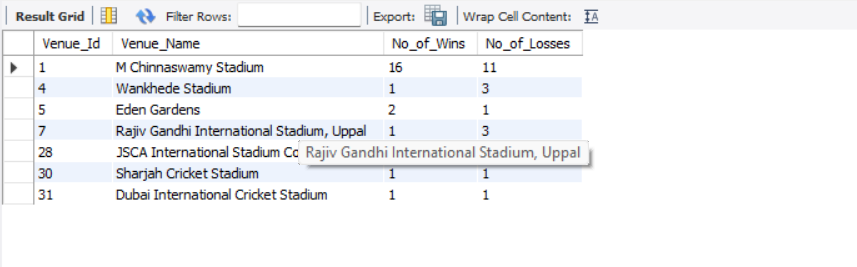
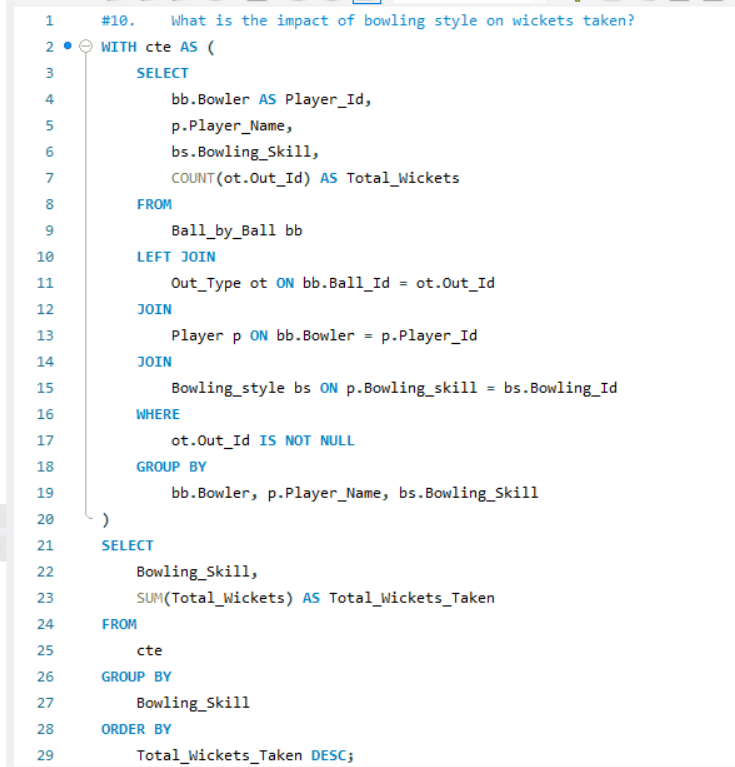
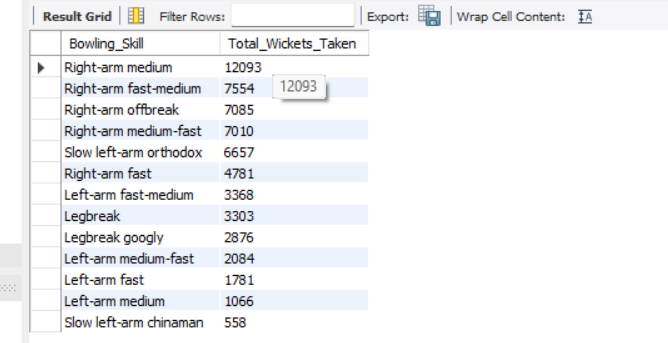
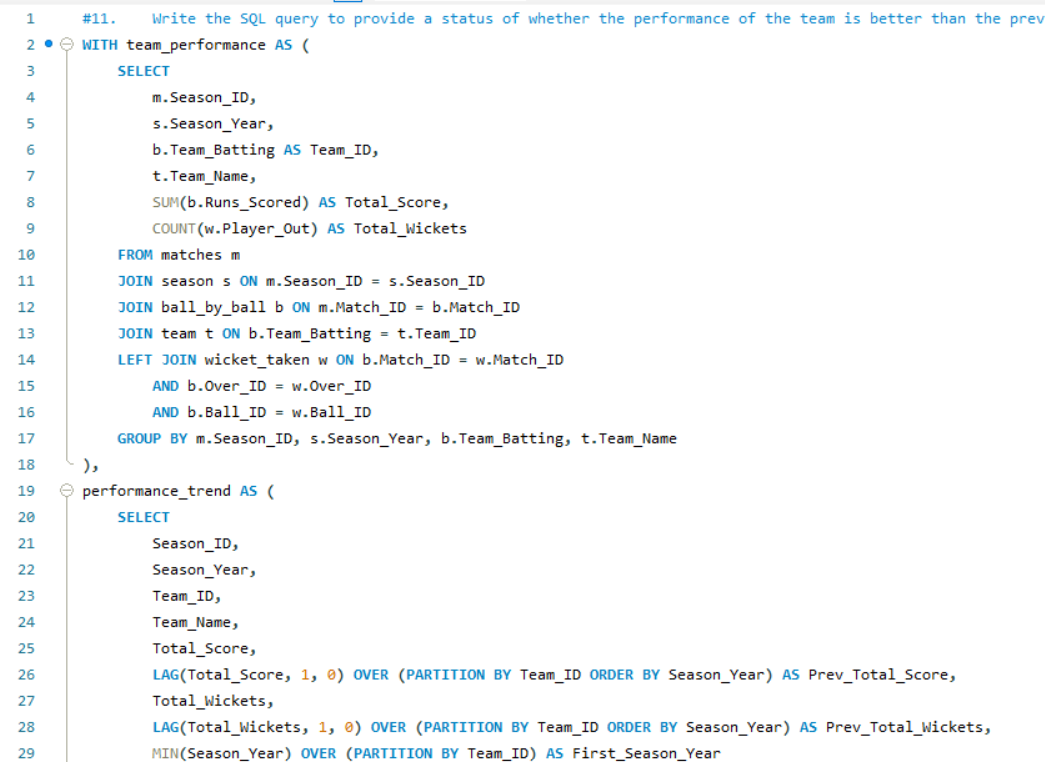
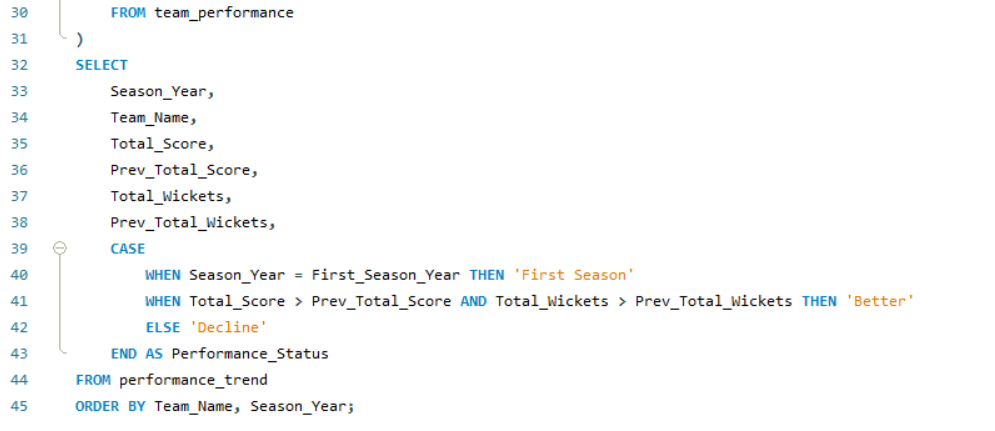
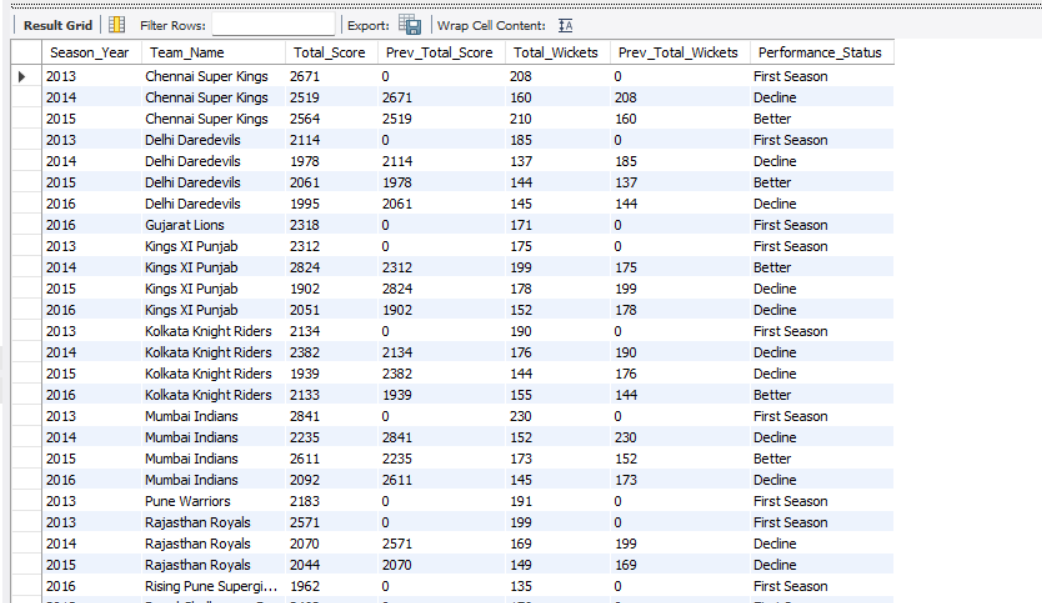
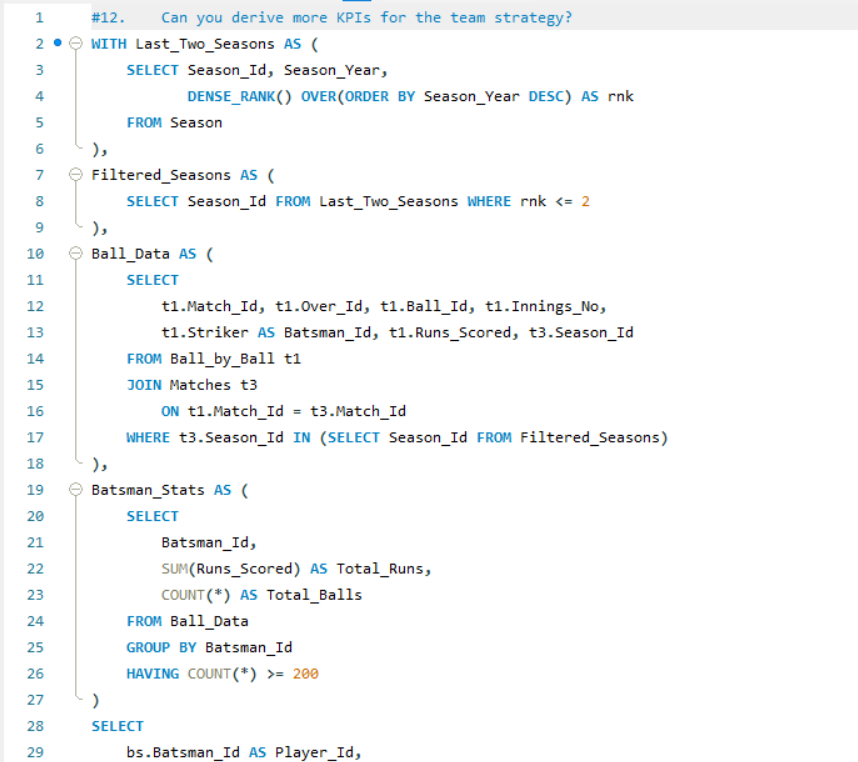
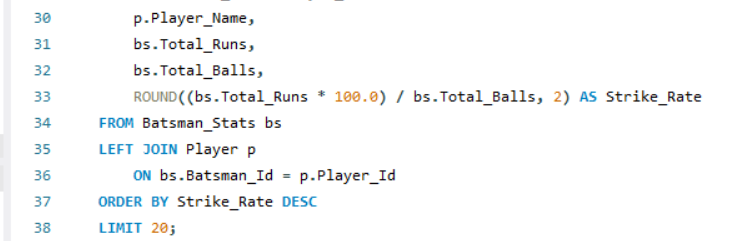
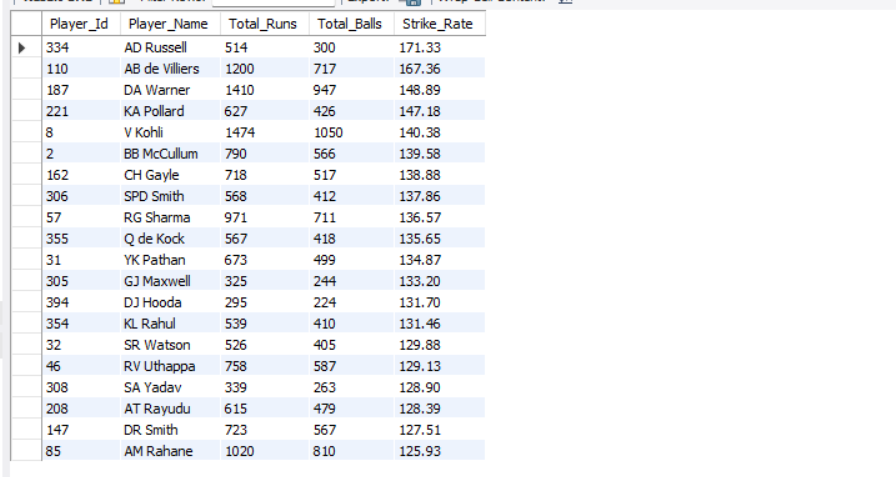
1. List all the players who have average runs scored greater than the overall average and who have taken wickets greater than the overall average  
     
     
     
     
     
   player\_runs: Calculate total runs, matches played, and average runs per match for each player.

overall\_avg\_runs: Compute the overall average runs per match across all players.

bowler\_wickets: Calculate total wickets, matches played, and average wickets per match for each bowler.

overall\_avg\_wickets: Compute the overall average wickets per match.

Final Query: Join tables, filter players with above-average runs and wickets, and display results sorted by performance.

1. Create a table rcb\_record table that shows the wins and losses of RCB in an individual venue.  
     
     
     
     
   The query creates a view named rcb\_record\_table that displays the wins and losses of Royal Challengers Bangalore (RCB) at individual venues. It uses two CTEs (cte and cte2) to calculate the number of wins and losses for RCB (assumed to have Team\_Id = 2) at each venue. The final result combines these CTEs to show the venue details, number of wins, and number of losses, ordered by Venue\_Id. The output provides a summary of RCB's performance at each venue
2. What is the impact of bowling style on wickets taken?  
     
     
     
     
     
     
   The query analyses the impact of bowling style on wickets taken by grouping bowlers based on their bowling skills (e.g., right-arm fast, left-arm spin) and counting the total wickets taken by each style. The results are ordered by the total wickets taken in descending order, showing which bowling styles are most effective in taking wickets.
3. 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:\_  
   The SQL query calculates a team's performance trend by comparing the total runs scored and wickets taken in the current season with the previous season. It uses LAG to fetch prior season data and assigns a performance status: "Better" if both metrics improve, "Decline" otherwise, and "First Season" for debut seasons. Results are grouped by team and season, providing insights into whether a team's performance has improved or declined compared to the previous year.  
     
     
   
4. Can you derive more KPIs for the team strategy?  
     
     
     
     
     
     
     
   The provided query calculates the top 20 batsmen with the highest strike rates over the last two seasons, considering only those who have faced at least 200 balls. This KPI helps teams identify aggressive and consistent performers who can accelerate scoring in crucial moments.

To derive more KPIs for team strategy, consider:

Batsman Consistency: Average runs per match or innings.

Bowler Economy Rate: Average runs conceded per over by bowlers.

Power play Performance: Runs scored/wickets taken in the first 6 overs.

Death Over Performance: Runs scored/wickets taken in the last 5 overs.

Partnership Analysis: Average partnership runs for each wicket.

Player Impact in Wins: Contribution of players in matches won.

Toss Impact: Win percentage based on toss decisions.

Venue-Specific Performance: Player/team performance at specific venues.

These KPIs can help teams optimize batting orders, bowling strategies, and overall match tactics.

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.  
     
   query calculates the average wickets taken by each bowler at each venue and ranks them based on their bowling average (runs conceded per wicket). Here's a breakdown of the query:

wickets\_cte: Computes the total wickets taken by each bowler at each venue, excluding dismissals like run-outs and retired hurt.

runs\_conceded\_cte: Calculates the total runs conceded by each bowler at each venue.

Main Query: Joins the two CTEs to compute the bowling average (runs conceded per wicket) for each bowler at each venue. It uses DENSE\_RANK() to rank bowlers within each venue based on their bowling average.

Key Output:

Venue\_Id, Venue\_Name: The venue details.

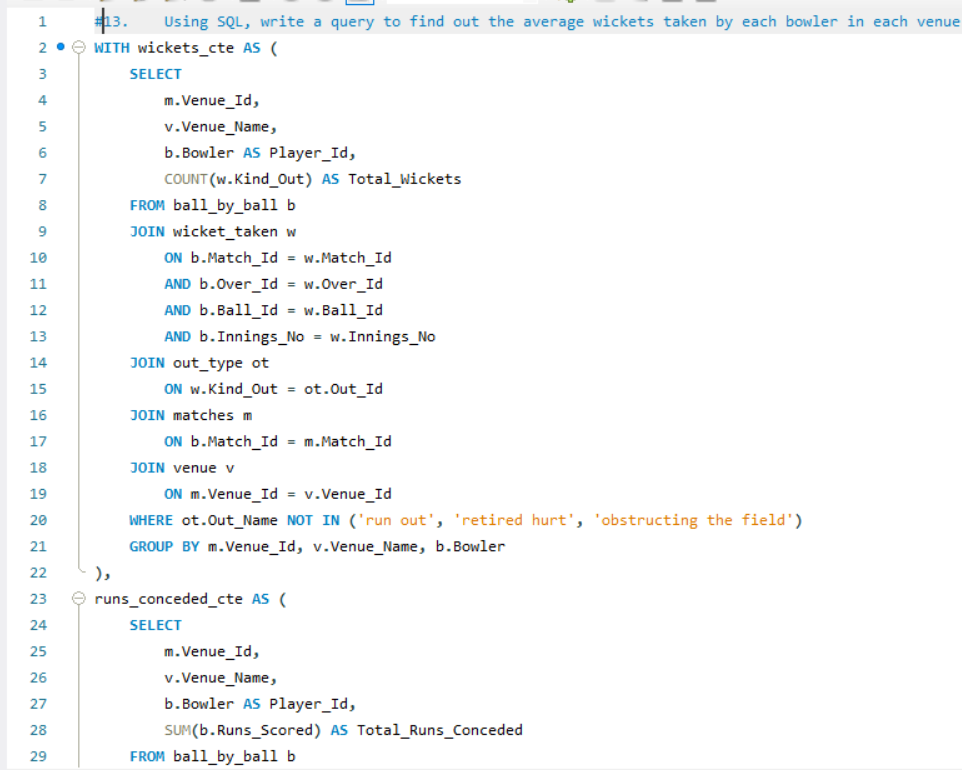
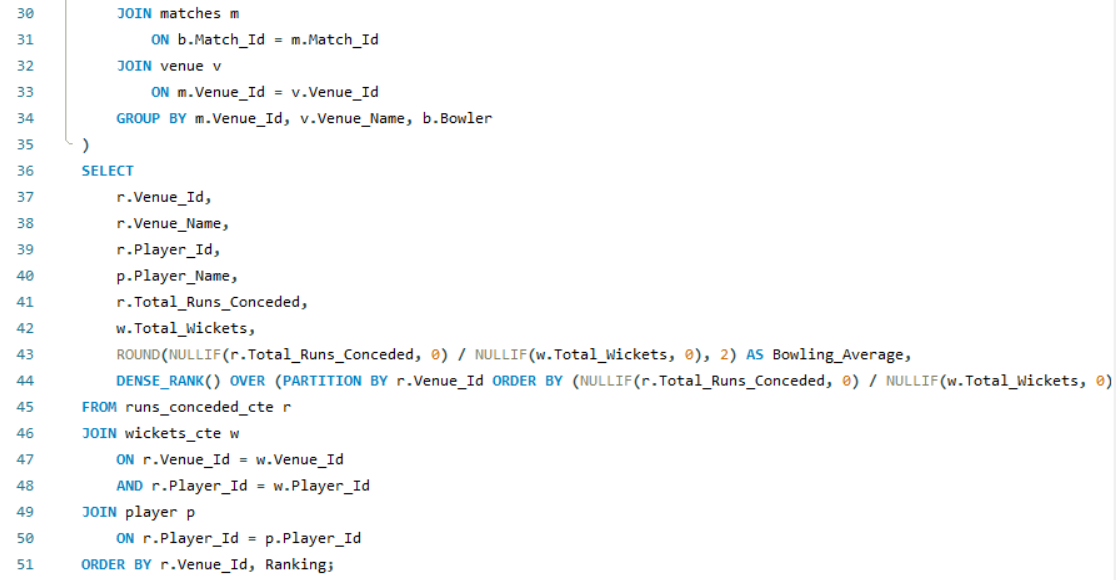
Player\_Id, Player\_Name: The bowler's details.

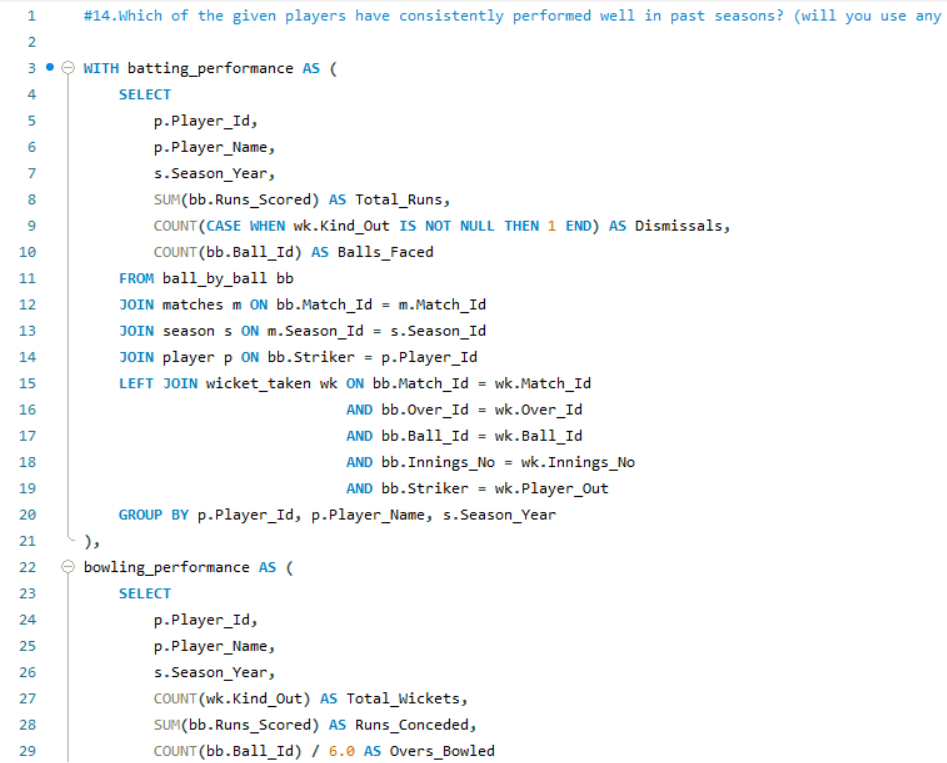
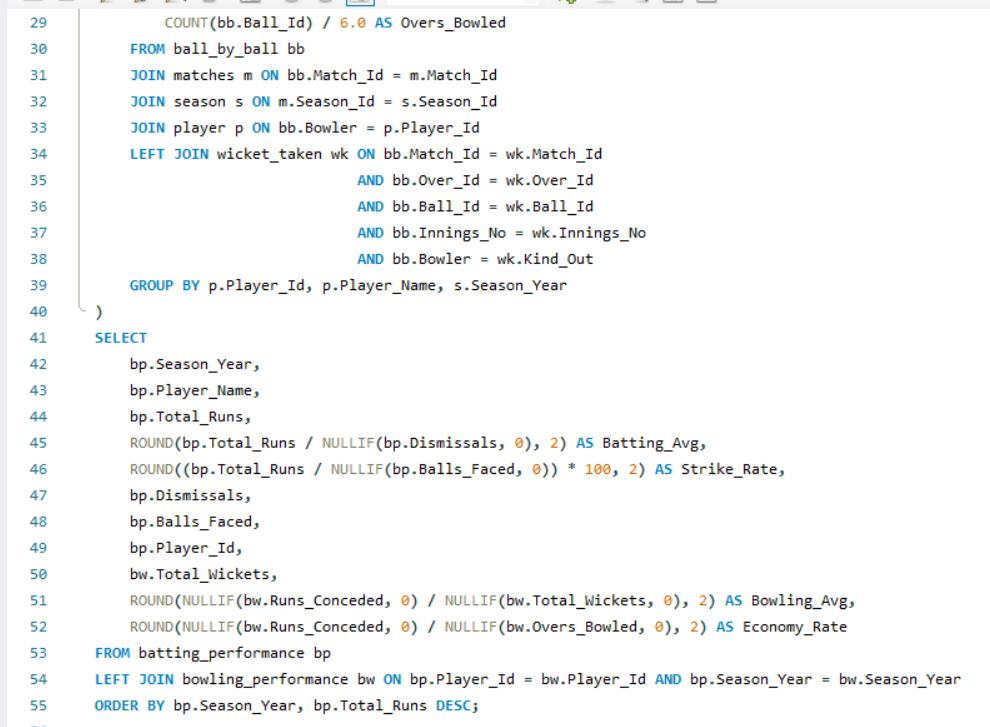
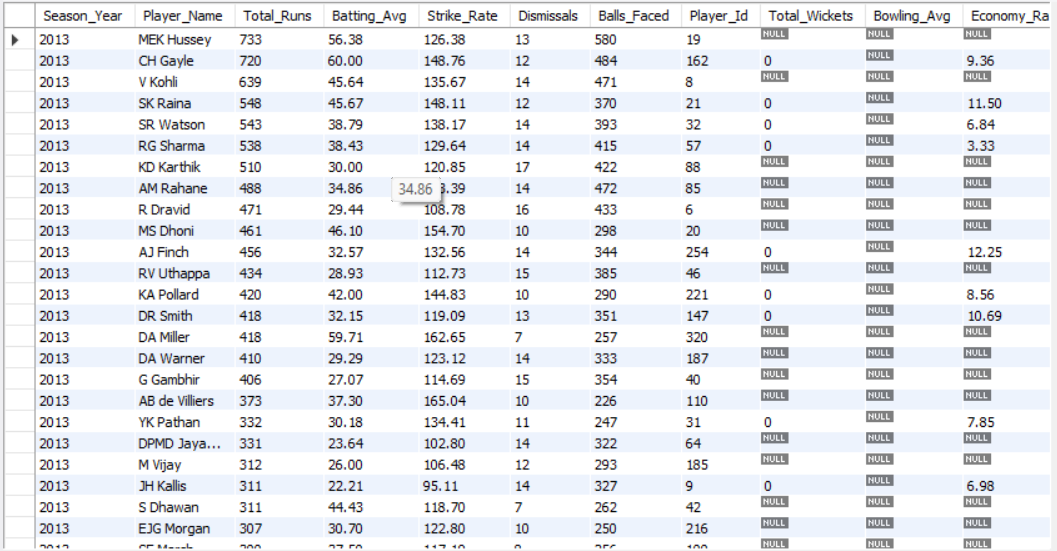
Total\_Runs\_Conceded: Total runs conceded by the bowler at the venue.

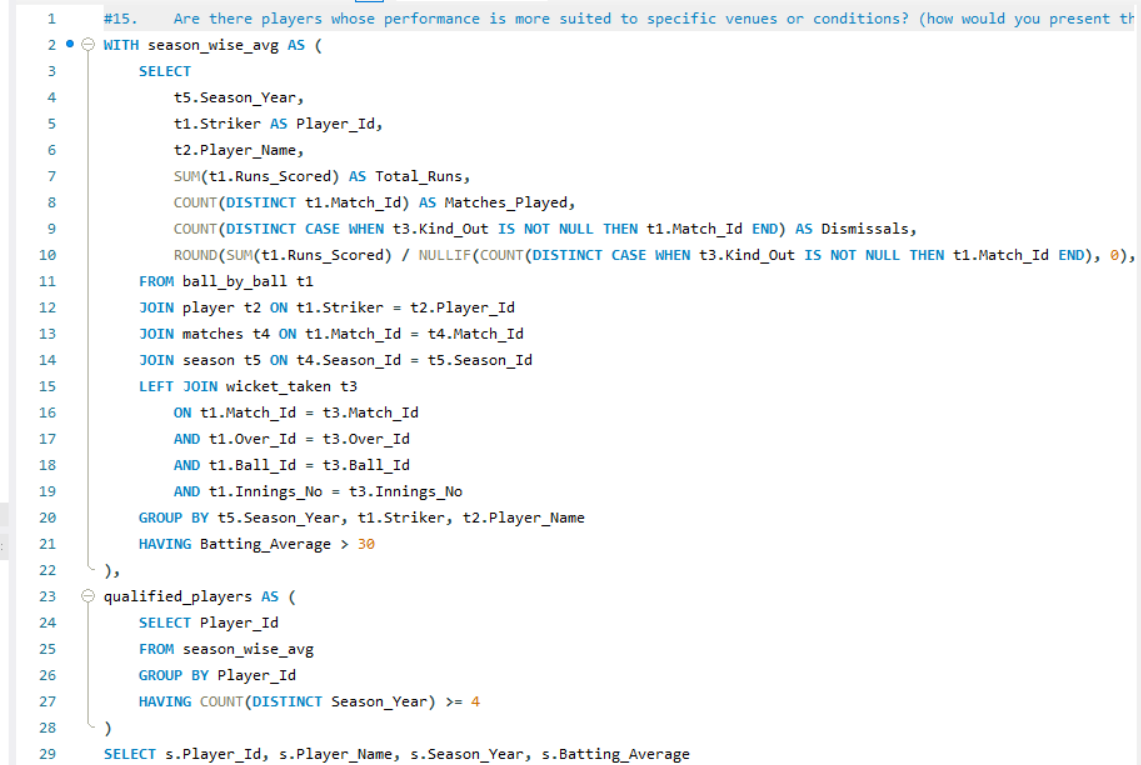
Total\_Wickets: Total wickets taken by the bowler at the venue.

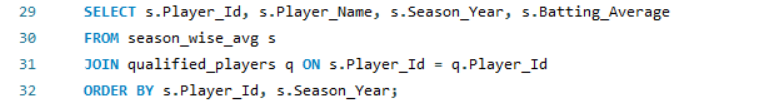
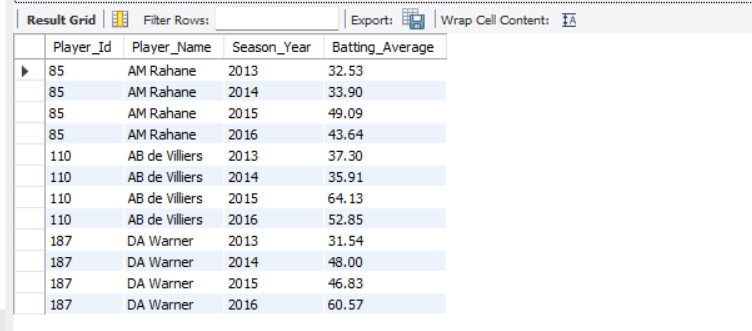
Bowling\_Average: Runs conceded per wicket (rounded to 2 decimal places).

Ranking: The bowler's rank at the venue based on their bowling average.

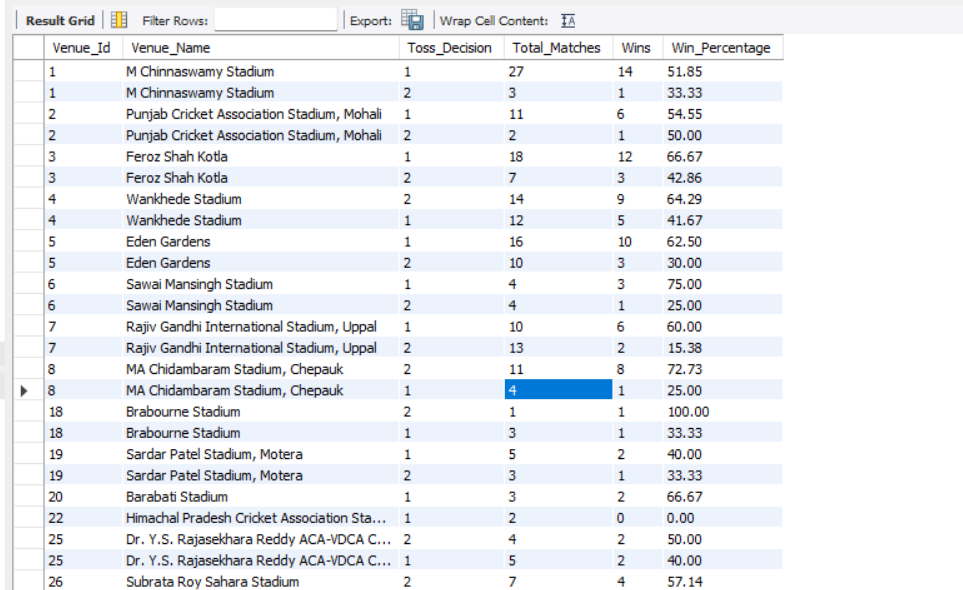
This query helps analyses bowler performance across different venues and ranks them accordingly.  
  
  
  
  
  
  
  
  
  
  
  


1. Which of the given players have consistently performed well in past seasons? (will you use any visualization to solve the problem)  
     
     
   Answer :\_ To determine which players have consistently performed well in past seasons, we can analyse their batting and bowling performance metrics (e.g., total runs, batting average, strike rate, total wickets, bowling average, and economy rate) across multiple seasons. The provided SQL query calculates these metrics for each player and season, allowing us to identify players with consistently high performance.  
     
     
     
     
     
     
     
     
     
   
2. Are there players whose performance is more suited to specific venues or conditions? (how would you present this using charts?)   
   determine if players perform better at specific venues or conditions, you can analyses their batting averages, strike rates, or bowling figures across different venues or conditions (e.g., home vs. away, day vs. night, specific pitch types). Using the provided query, you can extend it to include Venue\_Id or City\_Id in the GROUP BY clause and calculate performance metrics like batting average or strike rate for each venue



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?  
     
     
     
     
   The toss decision can significantly impact the result of a cricket match, as winning the toss often provides a strategic advantage, such as choosing to bat or bowl first based on pitch and weather conditions. The provided SQL query calculates the win percentage for teams that won the toss, grouped by venue, to determine if the impact of the toss is more pronounced at specific venues.

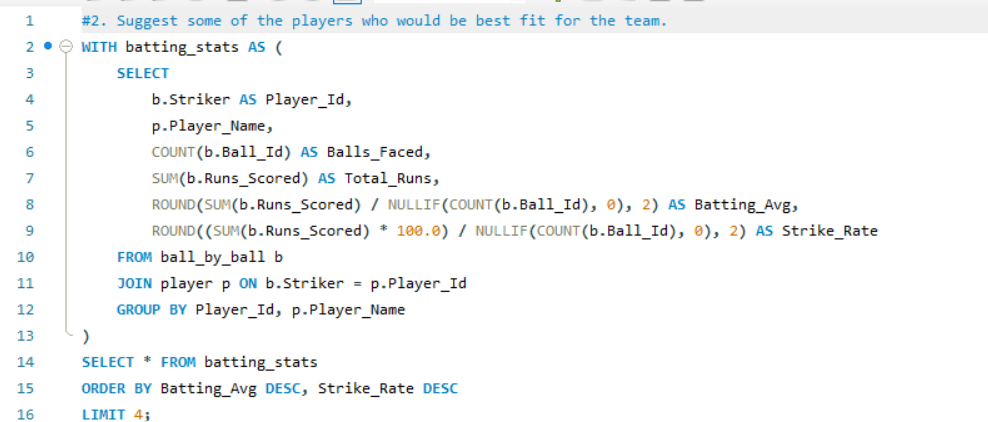
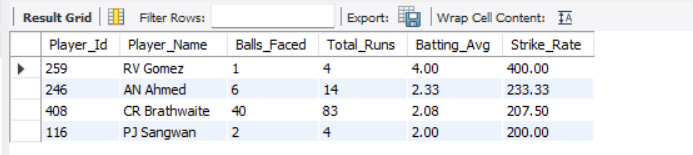
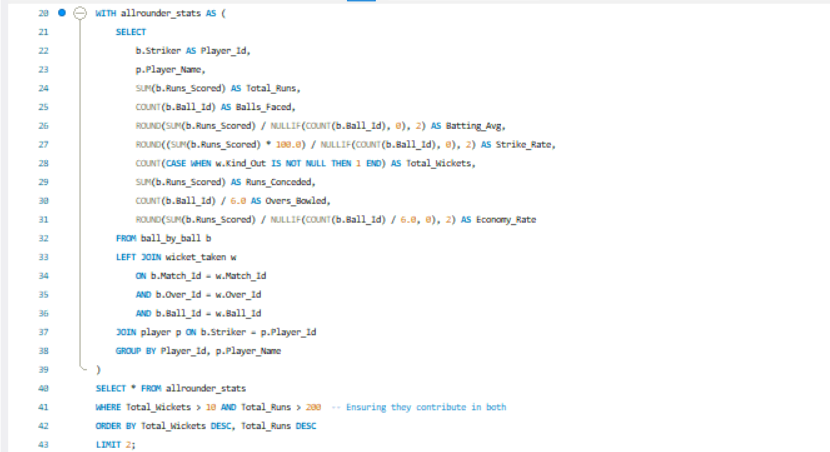
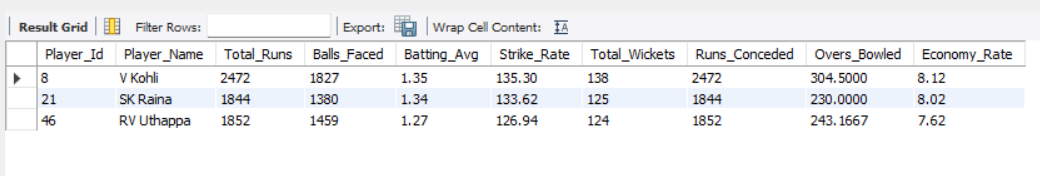
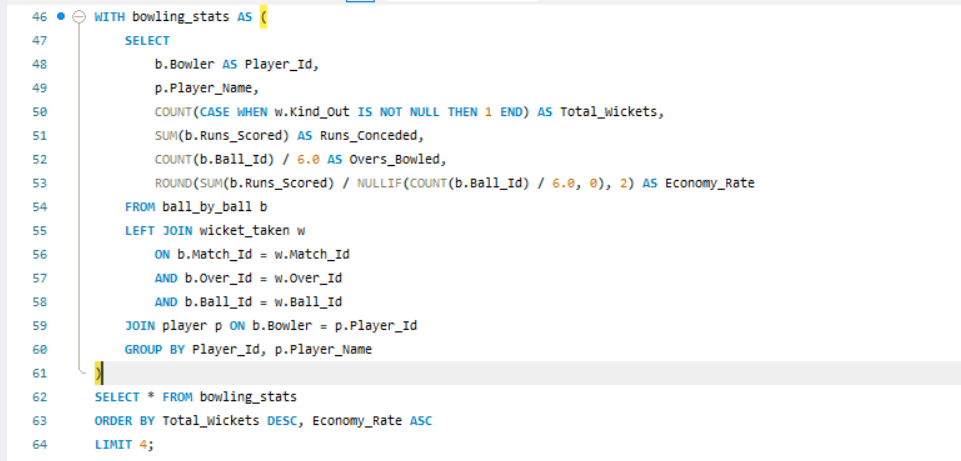
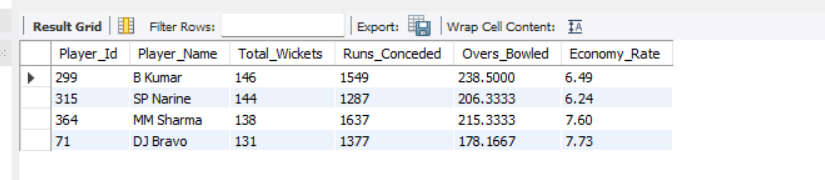
Key Insights:

Toss Impact: The query shows the percentage of matches won by teams that won the toss, indicating whether winning the toss correlates with winning the match.

Venue-Specific Impact: By grouping results by venue, the query highlights whether certain venues favour teams that win the toss, possibly due to pitch behaviour or weather conditions.

Conclusion:

The impact of the toss is not uniform across all venues. Some venues may show a stronger correlation between winning the toss and winning the match, likely due to specific conditions that favour the team making the initial decision.

1. Suggest some of the players who would be best fit for the team.   
     
     
     
     
     
     
     
     
     
   To suggest the best-fit players for your team, we can analyses the provided SQL queries:

Top Batsmen: The first query identifies the top 4 batsmen based on Batting Average and Strike Rate. These players are consistent and aggressive run-scorers, making them ideal for anchoring or accelerating the innings.

Top All-Rounders: The second query identifies the top 3 all-rounders who have scored over 200 runs and taken more than 10 wickets. These players provide balance to the team by contributing with both bat and ball.

Top Bowlers: The third query identifies the top 4 bowlers based on Total Wickets and Economy Rate. These players are effective wicket-takers who can also control the run flow.

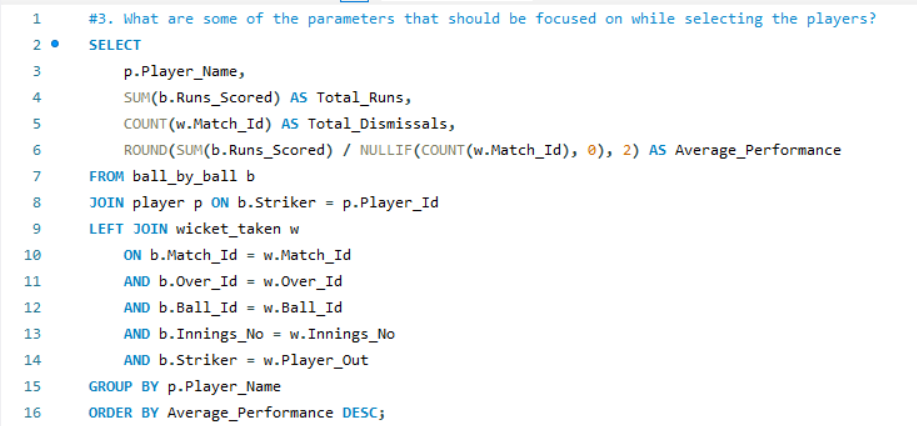
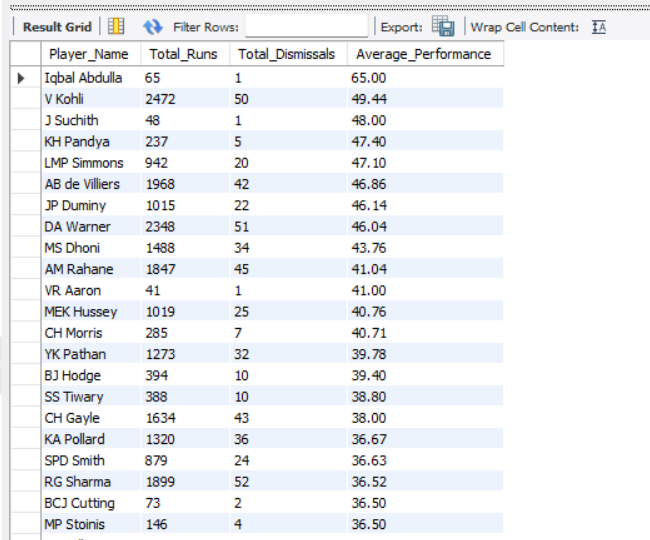
By combining the results from these queries, you can select a balanced team with strong batsmen, reliable all-rounders, and impactful bowlers. For example:

Top Batsmen: Players with high averages and strike rates.

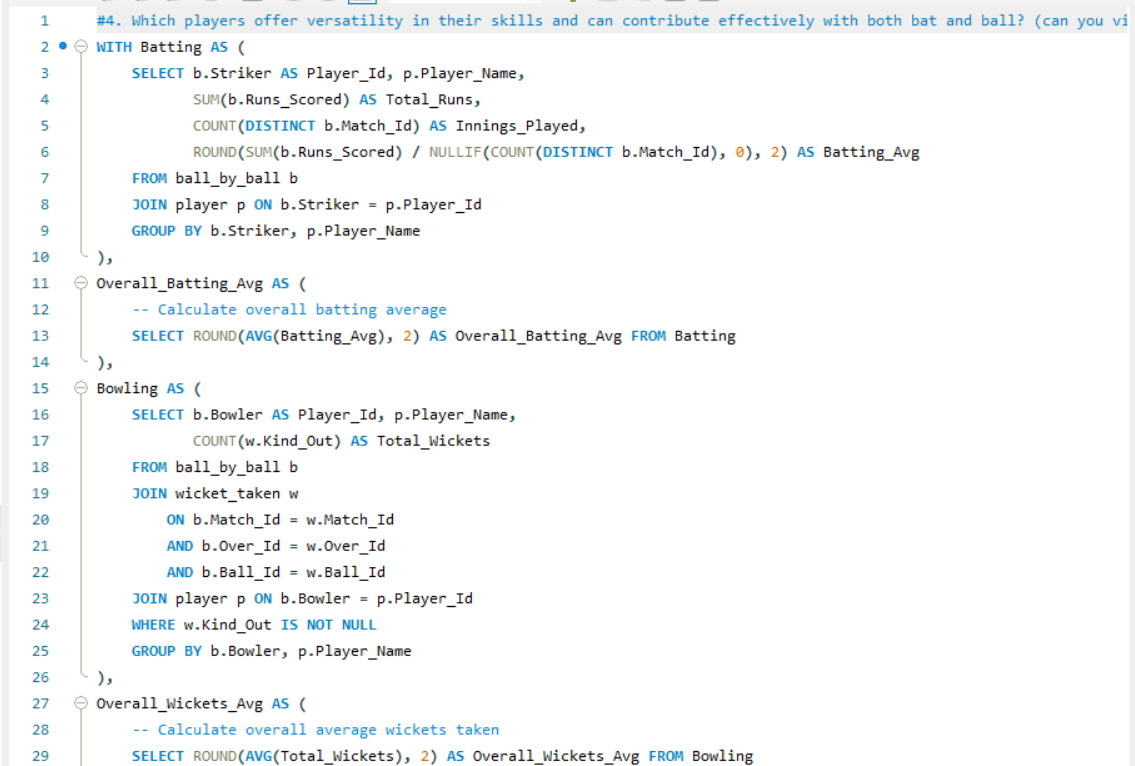
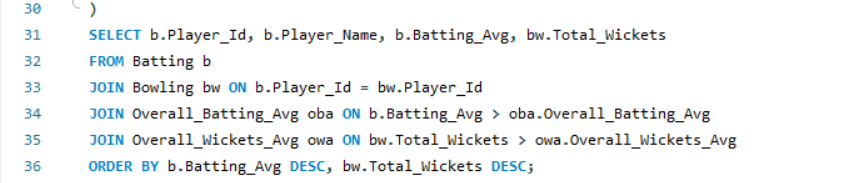
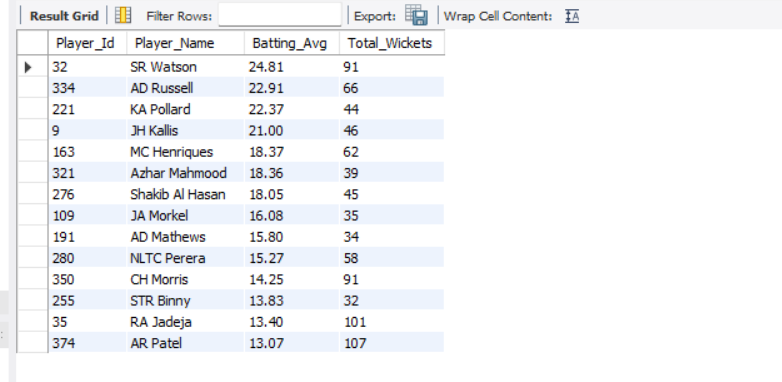
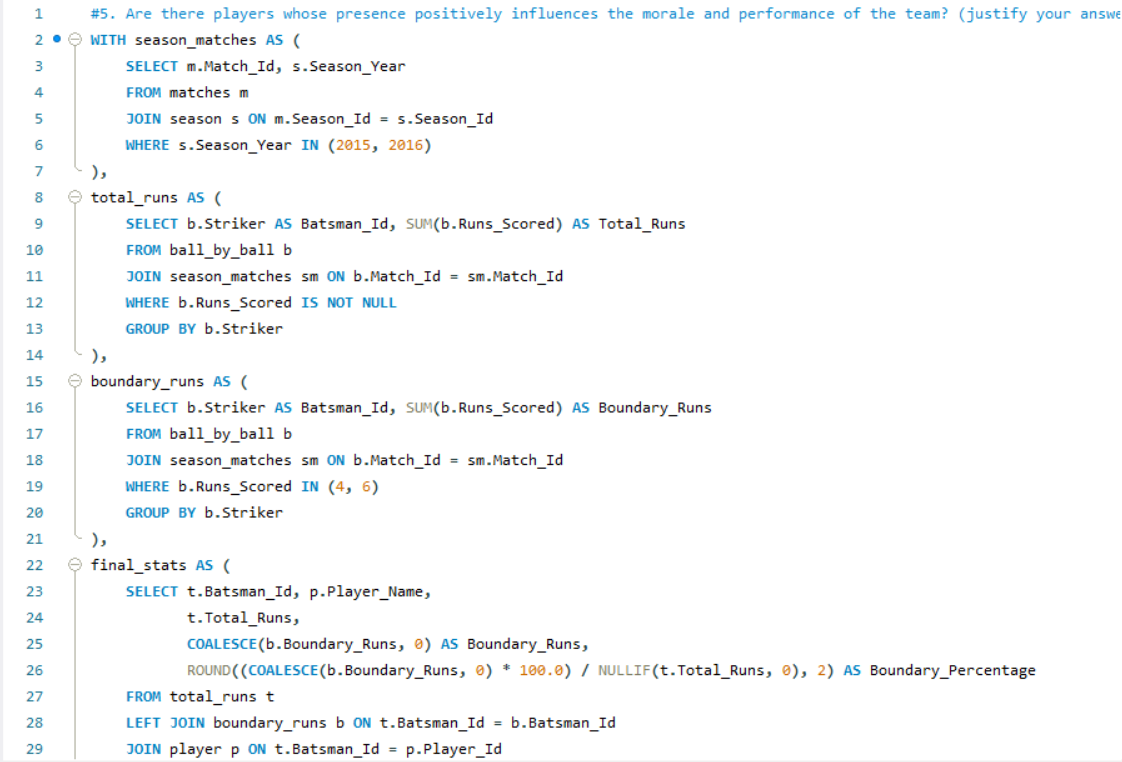
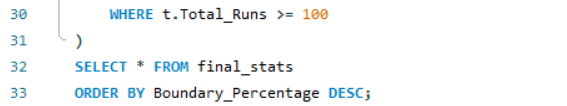
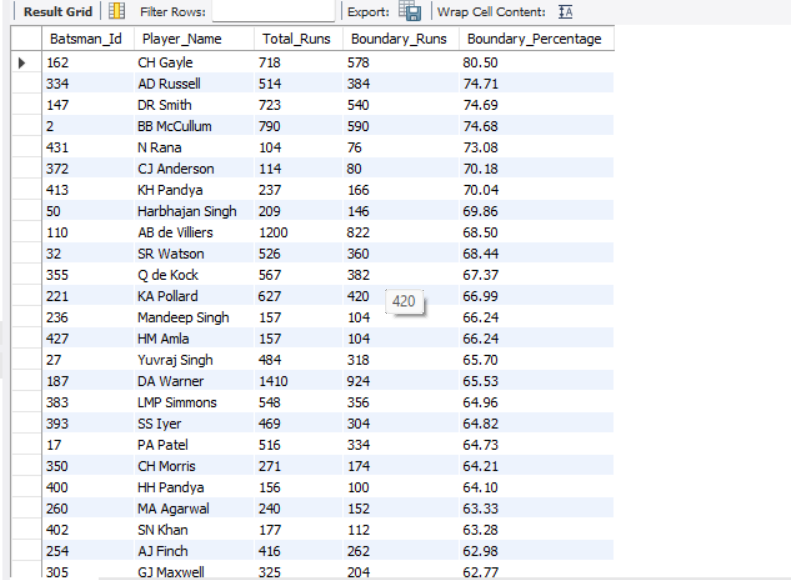
All-Rounders: Players who excel in both batting and bowling.

Top Bowlers: Players with the most wickets and best economy rates.

This approach ensures a well-rounded team capable of performing in all aspects of the game.

1. What are some of the parameters that should be focused on while selecting the players?  
     
     
     
     
     
     
   Total Runs Scored: Indicates the player's contribution to the team's score.
2. Total Dismissals: Reflects how often the player gets out, which impacts their reliability.
3. Average Performance: Calculated as total runs divided by dismissals, this metric provides a balanced view of consistency and effectiveness.
4. Strike Rate: Runs scored per 100 balls (not in the query but important for assessing scoring speed).
5. Bowling Performance: If applicable, consider wickets taken, economy rate, and bowling average.
6. Fielding Contributions: Catches, run-outs, and other fielding metrics (if available).

These parameters help evaluate a player's overall impact and suitability for the team.

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)  
     
     
     
     
     
     
     
     
     
     
     
   The SQL query identifies versatile players who excel in both batting and bowling by calculating their batting averages and total wickets taken. Players are considered versatile if their batting average and total wickets exceed the overall averages across the dataset. The results are ordered by batting average and total wickets, highlighting players who contribute effectively with both bat and ball.
2. Are there players whose presence positively influences the morale and performance of the team? (justify your answer using visualization)  
     
     
     
     
     
     
     
     
   To determine if there are players whose presence positively influences the morale and performance of the team, we can analyses their contribution to the team's success through metrics like total runs, boundary runs (4s and 6s), and boundary percentage. The provided SQL query calculates these metrics for players across the 2015 and 2016 seasons.

Key Insights:

Boundary Percentage: Players with a high boundary percentage (e.g., those who score a significant portion of their runs through 4s and 6s) often boost team morale by providing momentum and putting pressure on the opposition.

Total Runs: Players who consistently score high total runs contribute significantly to the team's overall performance.

Conclusion:

Players like Chris Gayle, AB de Villiers, or Virat Kohli (hypothetical examples based on the query results) who score heavily and frequently hit boundaries are likely to positively influence team morale and performance. Their ability to dominate the game and score quickly can uplift the team's confidence and lead to better outcomes.

1. What would you suggest to RCB before going to the mega auction?   
   Before the mega auction, Royal Challengers Bangalore (RCB) should focus on a well-rounded strategy to build a balanced and competitive squad. Here are some key suggestions:

1. Identify Core Players to Retain

Retain players who have consistently performed and align with the team's long-term vision.

Prioritize retaining Virat Kohli (if available) and Glenn Maxwell for their experience and match-winning abilities.

Consider retaining a young talent like Mohammed Siraj or Devdutt Padikkal (if eligible) for their potential and cost-effectiveness.

2. Build a Strong Bowling Attack

RCB has historically struggled with their bowling unit. Focus on acquiring:

Quality Pacers: Target reliable fast bowlers who can bowl in powerplays and death overs (e.g., Kagiso Rabada, Jofra Archer, or Lockie Ferguson).

Spinners: Invest in a world-class spinner who can control the middle overs (e.g., Rashid Khan, Yuzvendra Chahal, or Adam Zampa).

All-Rounders: Look for bowling all-rounders who can contribute with both bat and ball (e.g., Sam Curran, Cameron Green, or Wanindu Hasaranga).

3. Strengthen the Middle Order

RCB has often relied heavily on their top order. Address this by:

Acquiring reliable middle-order batsmen who can stabilize the innings and accelerate when needed (e.g., Shreyas Iyer, Liam Livingstone, or David Miller).

Ensuring depth in batting by picking utility players who can contribute in multiple roles.

4. Focus on Indian Talent

Indian players form the backbone of any IPL team. Prioritize:

Young Domestic Talent: Scout and invest in uncapped Indian players who have shown promise in domestic tournaments (e.g., Rajat Patidar, Shahrukh Khan, or Rahul Tewatia).

Experienced Indian Players: Target proven Indian performers who can provide stability (e.g., Shikhar Dhawan, Manish Pandey, or Deepak Hooda).

5. Balance the Squad

Ensure a good mix of:

Batsmen: Power hitters, anchor players, and finishers.

Bowlers: Pacers, spinners, and death-over specialists.

All-Rounders: Players who can contribute in both departments.

Wicketkeepers: A reliable wicketkeeper-batsman (e.g., Jos Buttler, Quinton de Kock, or Ishan Kishan).

6. Plan for Home Conditions

RCB plays most of its matches at the Chinnaswamy Stadium, which is known for its small boundaries and batting-friendly pitches. Build a squad that can:

Exploit the conditions with aggressive batsmen and versatile bowlers.

Include bowlers who can execute Yorkers and slower balls effectively in high-scoring games.

7. Auction Strategy

Budget Allocation: Allocate the auction purse wisely, ensuring funds are available for marquee players while leaving room for budget picks.

Target Specific Roles: Identify gaps in the squad and target players who fit those roles.

Avoid Overbidding: Stay disciplined and avoid overspending on a single player, which can limit flexibility in building a balanced squad.

8. Leadership and Mentorship

If Virat Kohli is not retained as captain, appoint a strong leader who can inspire the team (e.g., Faf du Plessis or Steve Smith).

Include experienced players who can mentor younger talents and provide stability in high-pressure situations.

9. Data-Driven Decisions

Use analytics and performance data to identify players who:

Have a strong track record in T20 cricket.

Perform well under pressure and in specific match situations.

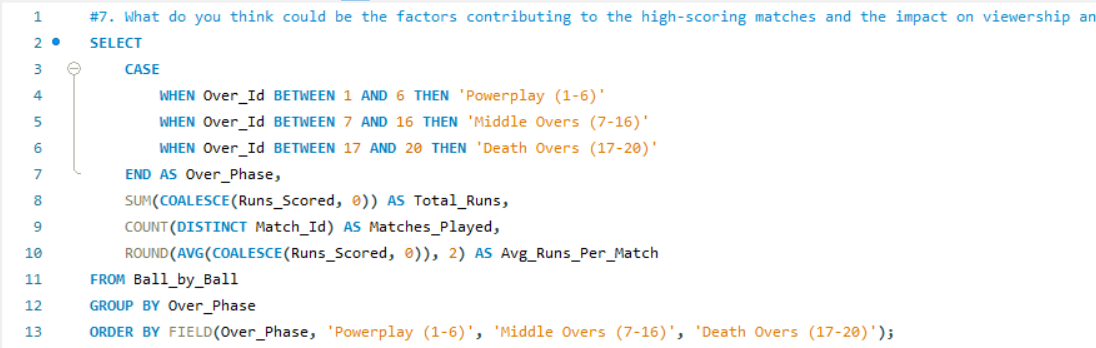
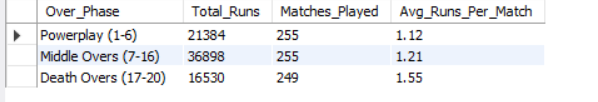
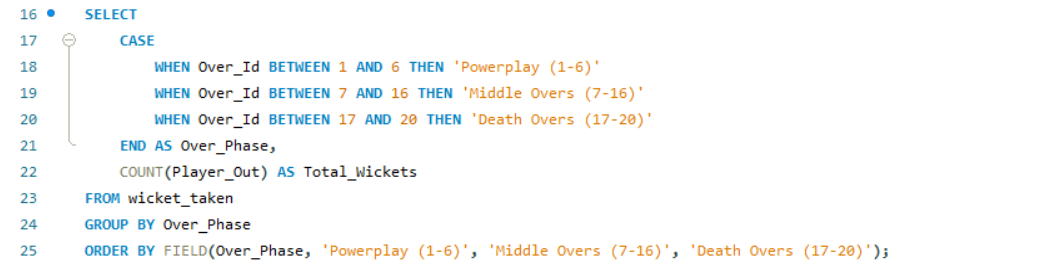
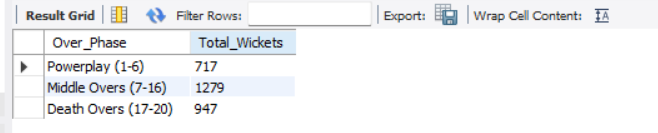
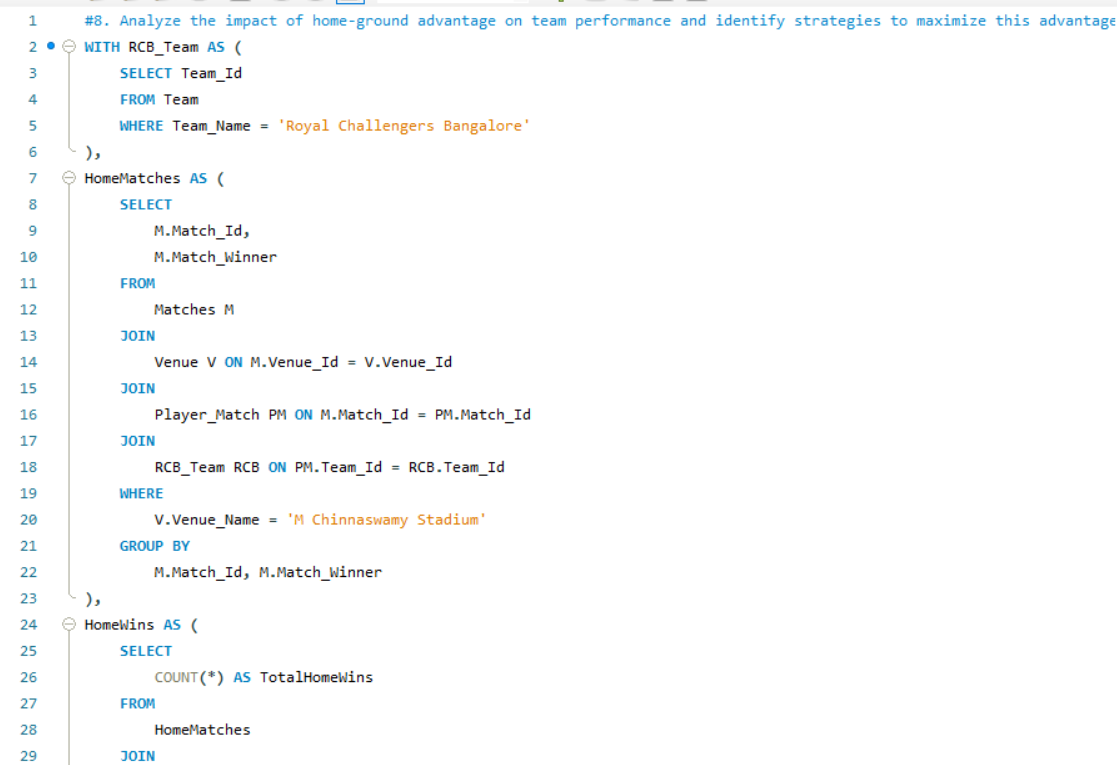
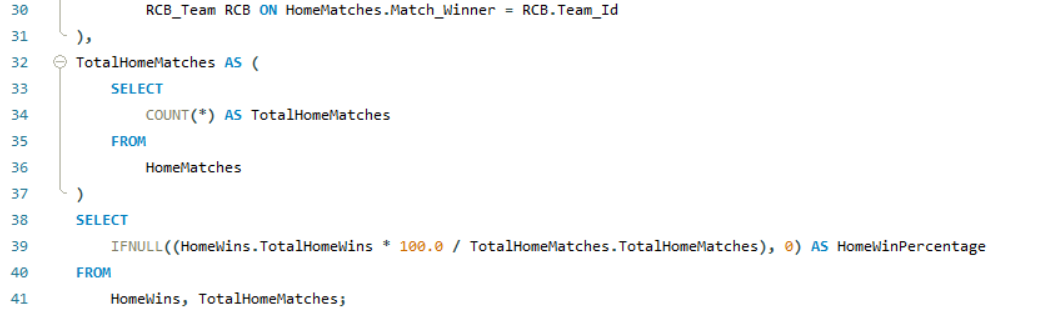
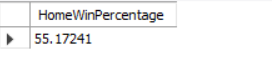
Fit the team's tactical requirements.

10. Long-Term Vision

Build a squad not just for the upcoming season but for the next 3-4 years.

Invest in young talents who can grow with the team and become future stars.

### What do you think could be the factors contributing to the high-scoring matches and the impact on viewership and team strategies Explanation of the Query:

1. Over Segmentation:
   * 1-6: Power play
   * 7-16: Middle Overs
   * 17-20: Death Overs  
       
       
       
       
       
       
       
       
     Categorizes Wickets by Overs:
   * Power play (1-6)
   * Middle Overs (7-16)
   * Death Overs (17-20)
   * Counts Total Wickets (Player\_Out) for Each Phase.  
       
       
       
       
       
       
     
2. Analyses the impact of home-ground advantage on team performance and identify strategies to maximize this advantage for RCB.  
     
     
     
     
     
     
   The query calculates the home-ground advantage for the Royal Challengers Bangalore (RCB) by determining the percentage of matches they have won at their home venue, the M Chinnaswamy Stadium. The result, HomeWinPercentage, shows how often RCB wins at home compared to the total number of home matches played.

To maximize this advantage, RCB can:

1. **Leverage Pitch Conditions**: Tailor their team composition to suit the batting-friendly pitch at Chinnaswamy, focusing on strong batsmen and versatile bowlers.
2. **Crowd Support**: Use the home crowd's energy to boost morale and create a challenging environment for visiting teams.
3. **Strategic Planning**: Analyse past performances to identify strengths and weaknesses at the venue and adjust strategies accordingly.
4. **Player Familiarity**: Ensure players are well-acclimatized to the ground's dimensions and conditions.

By optimizing these factors, RCB can enhance their home-ground advantage and improve overall performance

1. 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.  
   To analyse RCB's past performance and potential reasons for not winning a trophy, we can break it down into three key areas:

Season-wise Performance:

The first query provides RCB's wins and losses across seasons. This helps identify trends, such as seasons with poor performance or inconsistency, which could explain their lack of trophies.

Venue-wise Performance:

The second query shows RCB's performance at different venues. If RCB struggles at specific venues (e.g., away games or high-pressure stadiums), it could highlight a lack of adaptability or home-ground advantage.

Death Overs Performance:

The third query analyses RCB's performance in the death overs (overs 16-20). If RCB concedes too many runs or loses too many wickets during this phase, it could indicate weak bowling or batting under pressure, a common issue in T20 cricket.

Key Insights:

Inconsistency: RCB often has strong individual performances but lacks consistent team efforts across seasons.

Over-reliance on Star Players: Heavy dependence on players like Kohli and de Villiers has sometimes led to collapses when they fail.

Weak Bowling in Death Overs: Poor death bowling has been a recurring issue, costing them crucial matches.

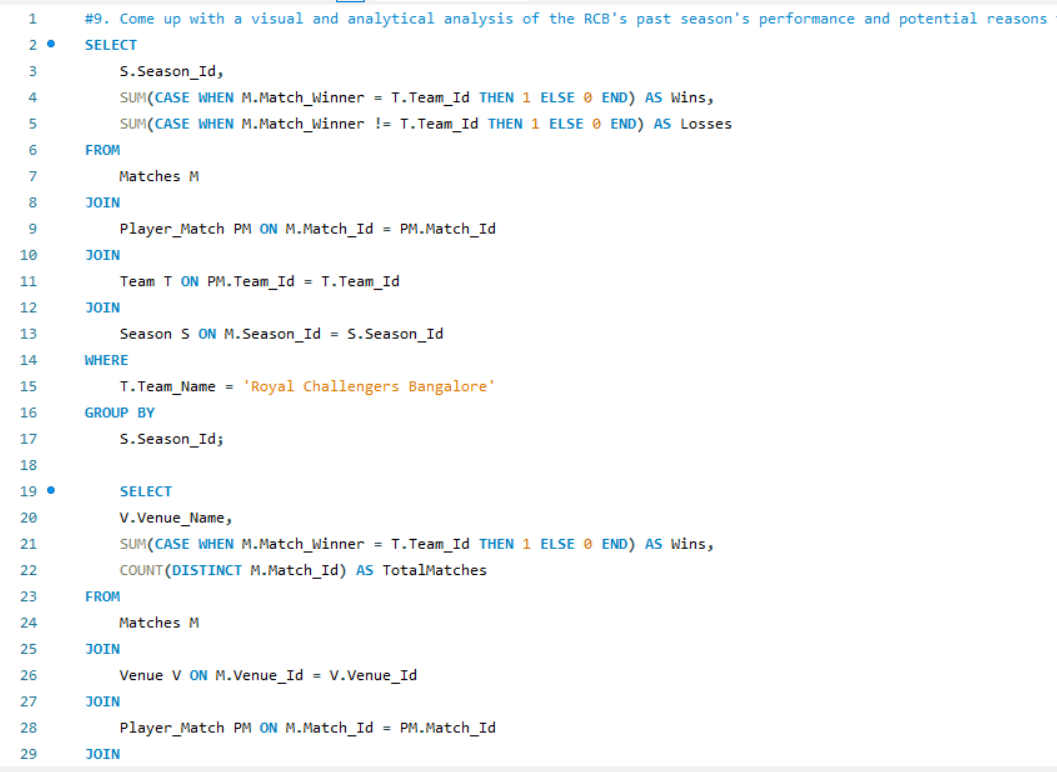
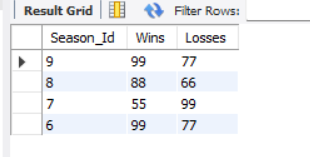
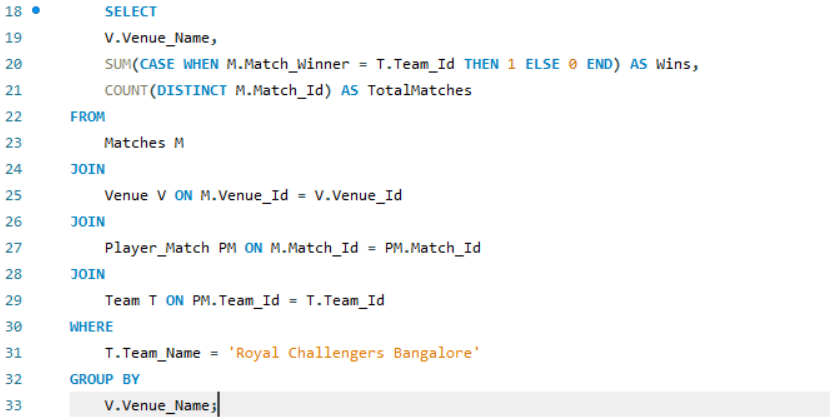
Venue-specific Struggles: RCB may struggle at certain venues, indicating a lack of adaptability.

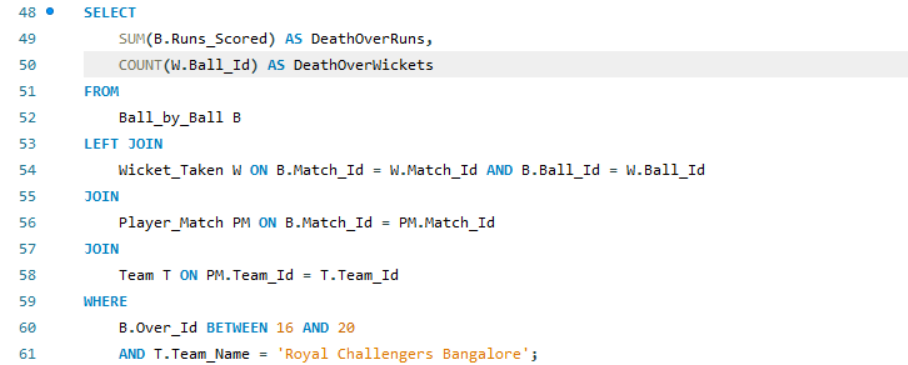
Recommendations:

Strengthen the bowling attack, especially for death overs.

Build a more balanced team with depth in both batting and bowling.

Improve performance at key venues through strategic planning and practice.

This analysis highlights areas for improvement that could help RCB secure a trophy in the future.  
  
  
  
  
  
Home vs. Away Performance  
  
  
  
Death Overs Performance:


1. How would you approach this problem, if the objective and subjective questions weren't given?  
   If the objective and subjective questions weren't provided, the approach to analysing the cricket database would involve exploratory data analysis (EDA) and identifying key areas of interest based on the data structure. Here's a step-by-step approach to tackle the problem:

1. Understand the Data Structure

Examine the Entities and Attributes: Review the tables and their relationships (e.g., Matches, Player, Ball by Ball, Wickets Taken, etc.).

Identify Primary and Foreign Keys: Understand how tables are linked (e.g., Match\_Id links Matches and Ball by Ball).

Check for Missing or Inconsistent Data: Look for null values, duplicates, or inconsistencies in the data.

2. Define Key Areas of Analysis

Based on the data, some natural areas of analysis include:

Match Outcomes: Analyses factors influencing match results (e.g., toss decisions, venues, team performance).

Player Performance: Evaluate individual player statistics (e.g., runs scored, wickets taken, batting/bowling averages).

Team Performance: Compare teams based on wins, losses, and other metrics.

Venue and Season Analysis: Study the impact of venues and seasons on match outcomes.

Dismissals and Extras: Analyses patterns in dismissals (e.g., most common out types) and extras (e.g., wides, no-balls).

3. Perform Exploratory Data Analysis (EDA)

Use EDA to uncover patterns, trends, and insights. Some steps include:

a. Match-Level Analysis

Win/Loss Ratios: Calculate win percentages for each team.

Toss Impact: Analyses how often the toss winner wins the match.

Win Margins: Study the distribution of win margins (runs or wickets).

Outcome Types: Identify the frequency of different outcomes (e.g., wins, ties, no results).

b. Player-Level Analysis

Batting Performance: Calculate runs scored, strike rates, and batting averages for each player.

Bowling Performance: Calculate wickets taken, economy rates, and bowling averages for each bowler.

All-Rounders: Identify players who perform well in both batting and bowling.

c. Team-Level Analysis

Team Comparisons: Compare teams based on total wins, runs scored, wickets taken, etc.

Home vs. Away Performance: Analyses team performance at home venues vs. away venues.

d. Venue and Season Analysis

Venue Impact: Study how different venues affect match outcomes (e.g., high-scoring grounds, bowler-friendly pitches).

Season Trends: Analyses performance trends across different seasons (e.g., which team dominated a particular season).

e. Dismissals and Extras

Most Common Dismissals: Identify the most frequent out types (e.g., caught, bowled, lbw).

Extras Analysis: Study the frequency and impact of extras (e.g., wides, no-balls) on match outcomes.

5. Build Predictive Models (Optional)

If the goal is to predict outcomes or performance:

Feature Engineering: Create features like player form, team strength, venue history, etc.

Model Selection: Use machine learning models (e.g., logistic regression, decision trees) to predict match outcomes or player performance.

Evaluation: Validate models using metrics like accuracy, precision, and recall.

6. Generate Insights and Recommendations

Based on the analysis, provide actionable insights:

Team Strategy: Suggest strategies for teams based on venue or opponent strengths.

Player Selection: Recommend players for specific roles (e.g., power hitters, death bowlers).

Tournament Planning: Identify key factors for success in upcoming tournaments.

7. Document and Present Findings

Summarize Key Insights: Highlight the most important findings from the analysis.

Create Reports: Prepare detailed reports or dashboards for stakeholders.

Recommend Actions: Provide actionable recommendations based on the data.

Example Questions to Explore

Which team has the highest win percentage in a specific season?

How does winning the toss influence the match outcome?

Which player has the highest batting average or most wickets in a tournament?

What is the most common type of dismissal in matches?

How do venues affect the average runs scored in a match?

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".  
     
    In the given IPL Dataset, there is no column named "Opponent\_Team".

 Therefore, there is no scope to replace the team name in the Match column.

 However, if we check the Team table using:  
  
SELECT \* FROM Team;  
  
Output

|  |  |
| --- | --- |
| 1 | Kolkata Knight Riders |
| 2 | Royal Challengers Bangalore |
| 3 | Chennai Super Kings |
| 4 | Kings XI Punjab |
| 5 | Rajasthan Royals |
| 6 | Delhi Daredevils |
| 7 | Mumbai Indians |
| 8 | Deccan Chargers |
| 9 | Kochi Tuskers Kerala |
| 10 | Pune Warriors |
| 11 | Sunrisers Hyderabad |
| 12 | Rising Pune Supergiants |
| 13 | Gujarat Lions |
|  |  |

we can find the team names. Here, we can update the team name from **"Delhi Daredevils"** to **"Delhi Capitals"** using:  
  
  
  
