**Objective Questions**

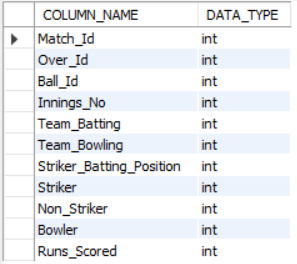
**Question 1: List the different dtypes of columns in table**

**“ball\_by\_ball”.**

****Findings:**** All columns in the ball\_by\_ball table have the ****INT**** data type.

****Tables Used:**** ball\_by\_ball

****Approach:**** I queried information\_schema.columns to retrieve column names and data types for the ball\_by\_ball table. I filtered the table name and schema to ensure I only got details about ball\_by\_ball.



select column\_name, data\_type

from information\_schema.columns

where table\_name = 'ball\_by\_ball'

and table\_schema = 'ipl';

**Question 2: What is the total number of runs scored in 6th season by RCB?**

****Findings:**** RCB scored a total of ****2601** runs** in Season 6, including extra runs. This indicates a consistent batting performance throughout the season.

****Tables Used:**** matches, ball\_by\_ball, extra\_runs

****Approach:**** I began by identifying matches played by RCB in Season 6 using the matches table. I used a Common Table Expression (CTE) to filter these matches. Then, I joined the ball\_by\_ball and extra\_runs tables to calculate the total runs, including extra runs such as wides and no-balls. Finally, I summed the runs for all matches where RCB was batting.

Screenshot 2025-02-14 130631

with rcb\_matches as

(select \* from matches

where Season\_Id = 6

and (Team\_1 = 2 or Team\_2 = 2)),

rcb\_batting\_runs as

(select

Match\_Id,

Innings\_No,

sum(runs\_scored) as total\_runs

from ball\_by\_ball

where Match\_Id in (select Match\_Id from rcb\_matches)

and Team\_Batting = 2

group by Match\_Id, Innings\_No),

rcb\_extra\_runs as

(select

Match\_Id,

Innings\_No,

sum(Extra\_Runs) as total\_extra\_runs

from extra\_runs

where Match\_Id in (select Match\_Id from rcb\_matches)

group by Match\_id, Innings\_No)

select

sum(total\_runs) + sum(total\_extra\_runs) as runs\_scored\_in\_season

from rcb\_batting\_runs rbr

left join rcb\_extra\_runs rer

on rbr.Match\_Id = rer.Match\_Id

and rbr.Innings\_No = rer.Innings\_No;

**Question 3: How many players were more than the age of 25 during season 2014?**

****Findings:**** A total of ****89** players** were over the age of 25 during the 2014 season. This indicates that the team had a good mix of experienced players.

****Tables Used:****player, player\_match, matches, season

****Approach:**** I first filtered the matches played in the 2014 season using the matches and season tables. Then, I identified the players who participated in those matches using the player\_match table. Finally, I calculated the players’ ages by comparing their date of birth (DOB) to January 1, 2014, using the TIMESTAMPDIFF function and counted how many were older than 25.

Screenshot 2025-02-14 130858

with season2014 as

(select \* from player

where player\_id in

(select player\_id from player\_match where match\_id in

(select match\_id from matches where season\_id in

(select season\_id from season where season\_year = '2014')))),

player\_age as

(select player\_id, timestampdiff(year, DOB, '2014-01-01') as age from season2014)

select count(distinct player\_id) as player\_count from player\_age where age > 25;

**Question 4: How many matches did RCB win in 2013?**

****Findings:**** RCB won ****9** matches** in the 2013 season. This shows that the team had a decent performance that year.

**Tables Used:** matches

**Approach:** I started by identifying the matches played in the 2013 season using the season and matches tables. Then, I counted the number of matches where the match\_winner was RCB (team\_id = 2).

Screenshot 2025-02-14 145116

select count(\*) as rcb\_wins from matches

where match\_id in

(select match\_id from matches where season\_id in

(select season\_id from season where season\_year = '2013'))

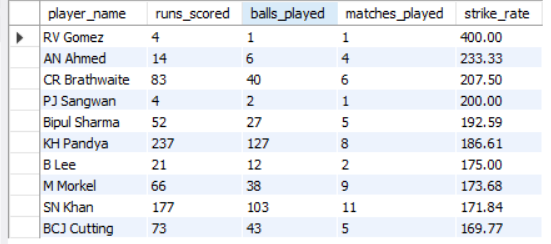
and match\_winner = 2;

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

****Findings:**** The top 10 players with the highest strike rate in the last four seasons were identified. This helps in shortlisting aggressive batsmen who can score quickly.

**Tables Used:** ball\_by\_ball, player, matches, season

**Approach:** First, I ranked the last four seasons using the season table. Then, I calculated the total runs scored and balls played by each player using the ball\_by\_ball table. After that, I computed the strike rate and filtered the top 10 players with the highest strike rate.



with seasonstable as

(select season\_id,

dense\_rank() over(order by season\_year desc) as season\_rank

from season)

select

player\_name,

sum(runs\_scored) as runs\_scored,

count(ball\_id) as balls\_played,

count(distinct match\_id) as matches\_played,

round(100\*sum(runs\_scored)/count(ball\_id),2) as strike\_rate

from ball\_by\_ball bb join player p on p.player\_id = bb.striker

where match\_id in

(select match\_id from matches where season\_id in

(select season\_id from seasonstable where season\_rank <= 4))

group by player\_name

having count(ball\_id) >0

order by strike\_rate desc

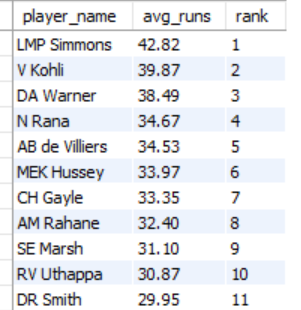
limit 10

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

****Findings:**** I got the average runs scored by each batsman across all seasons. This helps identify consistent performers.

**Tables Used:** ball\_by\_ball, player

**Approach:** I grouped the data by each batsman using the ball\_by\_ball table to calculate the total runs scored and the number of innings they played. Then, I calculated the average runs by dividing the total runs by the number of innings.



with avg\_runs\_table as

(select

striker,

sum(runs\_scored) as total\_runs,

count(distinct match\_id) as matches\_played,

round(sum(runs\_scored)/count(distinct match\_id),2) as avg\_runs

from ball\_by\_ball

group by striker

order by avg\_runs desc)

select p.player\_name, ar.avg\_runs, dense\_rank() over(order by avg\_runs desc) as 'rank'

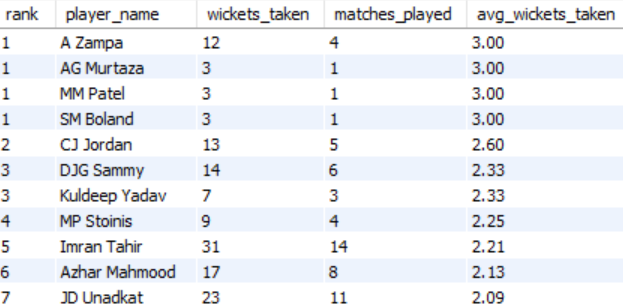
from player p join avg\_runs\_table ar on p.player\_id = ar.striker;

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

****Findings:**** I found the average number of wickets taken by each bowler. This is useful for spotting consistent wicket-takers.

****Tables Used:**** wickets\_taken, player

****Approach:**** I used the wickets\_taken table to count the total number of wickets taken by each bowler. Then, I divided the total wickets by the number of matches they played to get the average wickets per match.



with bowler\_wickets as

(select wt.match\_id, b.Bowler, wt.player\_out from wicket\_taken wt left join ball\_by\_ball b

on b.match\_id = wt.match\_id

and b.innings\_no = wt.innings\_no

and b.over\_id = wt.over\_id

and b.ball\_id = wt.ball\_id)

select

dense\_rank() over(order by round(count(player\_out)/count(distinct match\_id),2) desc) as 'rank',

p.player\_name,

count(player\_out) as wickets\_taken,

count(distinct match\_id) as matches\_played,

round(count(player\_out)/count(distinct match\_id),2) as avg\_wickets\_taken

from bowler\_wickets b join player p on b.bowler = p.player\_id

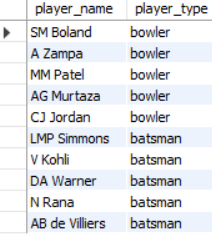
group by p.player\_name;

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

**Findings:** I found players who are valuable all-rounders with both runs and wickets above the overall average. These players should be prioritized during the auction.

**Tables Used:** ball\_by\_ball, wickets\_taken, player

**Approach:** I first calculated the overall average runs and wickets across all players. Then, I filtered the players whose runs and wickets were both above the overall average by using subqueries to compare their individual performance with the overall values.



with bowler\_wickets as

(select wt.match\_id, b.Bowler, wt.player\_out from wicket\_taken wt left join ball\_by\_ball b

on b.match\_id = wt.match\_id

and b.innings\_no = wt.innings\_no

and b.over\_id = wt.over\_id

and b.ball\_id = wt.ball\_id),

avg\_wickets as

(select

p.player\_name,

round(count(player\_out)/count(distinct match\_id),2) as avg\_wickets\_taken

from bowler\_wickets b join player p on b.bowler = p.player\_id

group by p.player\_name),

avg\_runs\_table as

(select

p.player\_name,

round(sum(runs\_scored)/count(distinct match\_id),2) as avg\_runs

from ball\_by\_ball b

join player p on p.player\_id = b.striker

group by p.player\_name

order by avg\_runs desc),

above\_avg\_wickets as

(select player\_name from avg\_wickets

where avg\_wickets\_taken >

(select avg(avg\_wickets\_taken) from avg\_wickets)

order by avg\_wickets\_taken desc

limit 5),

above\_avg\_scorers as

(select player\_name from avg\_runs\_table

where avg\_runs >

(select avg(avg\_runs) from avg\_runs\_table)

order by avg\_runs desc

limit 5)

select player\_name, case when player\_name is not null then 'bowler' else "" end as player\_type from above\_avg\_wickets

union

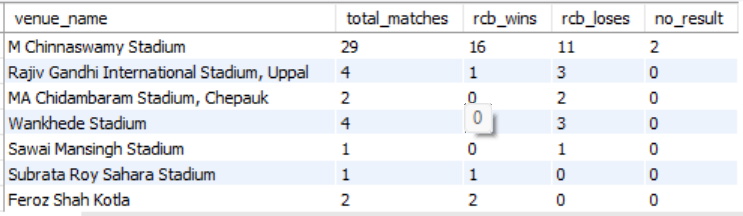
select player\_name, case when player\_name is not null then 'batsman' else "" end as player\_type from above\_avg\_scorers;

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

**Findings:** I created a table showing how RCB performed at each venue. This helps in understanding home-ground advantage and identifying favorable venues.

**Tables Used:** matches, venue, team

**Approach:** I created a table that counts the number of matches RCB won and lost at each venue. I filtered matches where RCB was either the winner or loser and grouped them by the venue to calculate the total wins and losses.



with rcb\_records as

(select

v.venue\_name,

count(\*) as total\_matches,

sum(case when Match\_Winner = 2 then 1 else 0 end) as rcb\_wins,

sum(case when Match\_Winner <> 2 then 1 else 0 end) as rcb\_loses,

sum(case when Match\_Winner is null then 1 else 0 end) as no\_result

from matches m join venue v on v.venue\_id = m.venue\_id

where Team\_1 = 2 or Team\_2 = 2

group by v.venue\_name)

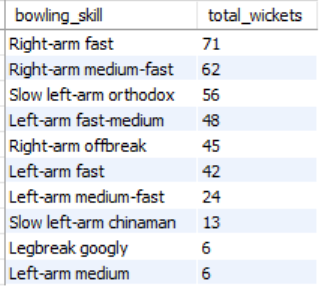
Select \* from rcb\_records;

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

****Findings:**** Fast bowlers took the most wickets, followed by spinners. This shows the importance of a balanced bowling attack.

**Tables Used:** wickets\_taken, player, bowling\_style

**Approach:** I joined the wickets\_taken table with the player and bowling\_style tables to classify bowlers by their bowling style. Then, I grouped the data by bowling style and counted the total number of wickets taken for each style.



select

bs.bowling\_skill,

count(wt.player\_out) as total\_wickets

from wicket\_taken wt

join ball\_by\_ball bb

on wt.match\_id = bb.match\_id

and wt.innings\_no = bb.innings\_no

and wt.over\_id = bb.over\_id

and wt.ball\_id = bb.over\_id

join player p on p.player\_id = bb.bowler

join bowling\_style bs on bs.bowling\_id = p.bowling\_skill

group by bowling\_skill

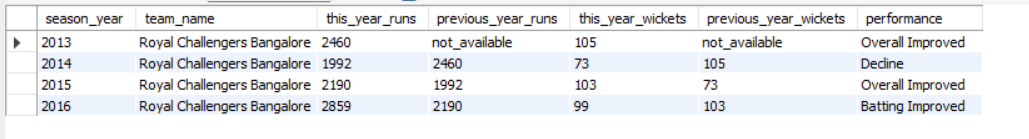
order by total\_wickets desc;

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

**Findings:** The query shows whether RCB’s overall performance (based on runs scored and wickets taken) improved or declined compared to the previous season.

**Tables Used:** matches, ball\_by\_ball, wickets\_taken

**Approach:** I first calculated the total runs scored and wickets taken by RCB in each season. Then, I compared the results for consecutive seasons to determine if there was an improvement.



with yearwisetable as

(select s.season\_year, bb.team\_batting, bb.team\_bowling, bb.runs\_scored, wt.player\_out

from ball\_by\_ball bb

left join wicket\_taken wt on bb.match\_id = wt.match\_id and wt.innings\_no = bb.innings\_no

and bb.over\_id = wt.over\_id and bb.ball\_id = wt.ball\_id

join matches m on m.match\_id = bb.match\_id

join season s on s.season\_id = m.season\_id),

batting\_table as

(select season\_year, t.team\_name, sum(runs\_scored) as yearly\_runs from yearwisetable y join team t on y.team\_batting = t.team\_id

group by season\_year, t.team\_name),

bowling\_table as

(select season\_year, t.team\_name, count(player\_out) as yearly\_wickets from yearwisetable y join team t on y.team\_bowling = t.team\_id

group by season\_year, t.team\_name),

year\_record\_table as

(select bt.season\_year, bt.team\_name, bt.yearly\_runs as this\_year\_runs,

coalesce (lag(yearly\_runs) over(partition by team\_name order by season\_year), 'not\_available') as previous\_year\_runs,

bw.yearly\_wickets as this\_year\_wickets,

coalesce (lag(yearly\_wickets) over(partition by team\_name order by season\_year), 'not\_available') as previous\_year\_wickets

from batting\_table bt

left join bowling\_table bw on bt.season\_year = bw.season\_year and bt.team\_name = bw.team\_name

order by team\_name, season\_year, yearly\_runs desc, yearly\_wickets desc)

select season\_year, team\_name, this\_year\_runs, previous\_year\_runs, this\_year\_wickets, previous\_year\_wickets,

case when this\_year\_runs > previous\_year\_runs and this\_year\_wickets > previous\_year\_wickets then 'Overall Improved'

when this\_year\_runs > previous\_year\_runs and this\_year\_wickets < previous\_year\_wickets then 'Batting Improved'

when this\_year\_runs < previous\_year\_runs and this\_year\_wickets > previous\_year\_wickets then 'Bowling Improved'

when this\_year\_runs = previous\_year\_runs and this\_year\_wickets = previous\_year\_wickets then 'Same'

else 'Decline'

end as performance

from year\_record\_table

where team\_name = 'Royal Challengers Bangalore';

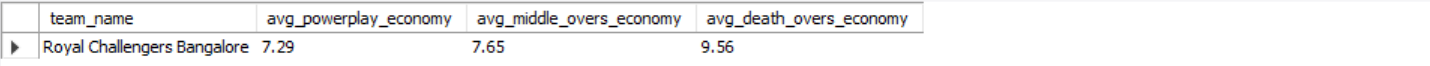
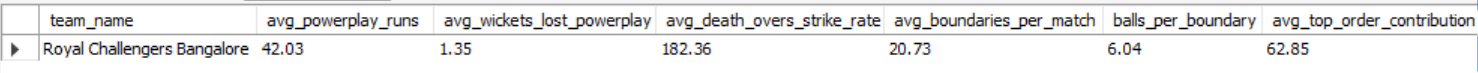
**Question 12: Can you derive more KPIs for the team strategy?**

**Findings:** These KPIs help in building specific strategies for batting and bowling phases, focusing on aggressive powerplay batting and controlled death-over bowling.

**Tables Used:** matches, ball\_by\_ball, wickets\_taken, player

**Approach:** I derived several key performance indicators (KPIs) like:

Average Powerplay Runs, Average Wicket Loss in Powerplay, Average Death Overs Strike Rate, Average Boundaries Per Match, Ball for Each Boundary, Average Top Order Contribution, Average Powerplay Economy, Average Middle Over Economy, Average Death Over Economy



with top\_order\_stats as

(select bb.match\_id, t.team\_name,

sum(runs\_scored) as top\_order\_runs,

tr.total\_runs

from ball\_by\_ball bb join team t on t.team\_id = bb.team\_batting

join (select match\_id, sum(runs\_scored) as total\_runs from ball\_by\_ball where team\_batting = 2 group by match\_id) tr on bb.match\_id = tr.match\_id

where bb.striker\_batting\_position <= 3

and t.team\_name = 'royal challengers bangalore'

group by match\_id),

powerplay\_stats as (

select

m.match\_id,

t.team\_name,

sum(bb.runs\_scored) as powerplay\_runs,

count(case when wt.player\_out is not null then 1 end) as wickets\_lost

from matches m

inner join ball\_by\_ball bb on m.match\_id = bb.match\_id

inner join team t on t.team\_id = bb.team\_batting

left join wicket\_taken wt

on bb.match\_id = wt.match\_id

and bb.innings\_no = wt.innings\_no

and bb.over\_id = wt.over\_id

and bb.ball\_id = wt.ball\_id

where

bb.over\_id between 1 and 6

and t.team\_name = 'royal challengers bangalore'

group by m.match\_id, t.team\_name

),

death\_overs\_stats as (

select

m.match\_id,

t.team\_name,

sum(bb.runs\_scored) as death\_overs\_runs,

count(bb.ball\_id) as balls\_faced

from matches m

inner join ball\_by\_ball bb on m.match\_id = bb.match\_id

inner join team t on t.team\_id = bb.team\_batting

where

bb.over\_id between 17 and 20

and t.team\_name = 'royal challengers bangalore'

group by m.match\_id, t.team\_name

),

boundary\_stats as (

select

m.match\_id,

t.team\_name,

sum(case when bb.runs\_scored in (4, 6) then 1 else 0 end) as boundaries,

count(bb.ball\_id) as total\_balls

from matches m

inner join ball\_by\_ball bb on m.match\_id = bb.match\_id

inner join team t on t.team\_id = bb.team\_batting

where t.team\_name = 'royal challengers bangalore'

group by m.match\_id, t.team\_name

)

select

p.team\_name,

round(avg(p.powerplay\_runs), 2) as avg\_powerplay\_runs,

round(avg(p.wickets\_lost), 2) as avg\_wickets\_lost\_powerplay,

round(avg(d.death\_overs\_runs / d.balls\_faced \* 100), 2) as avg\_death\_overs\_strike\_rate,

round(avg(b.boundaries), 2) as avg\_boundaries\_per\_match,

round(avg(nullif(b.total\_balls, 0) / nullif(b.boundaries, 0)), 2) as balls\_per\_boundary,

round(avg(s.top\_order\_runs / s.total\_runs \* 100),2) as avg\_top\_order\_contribution

from powerplay\_stats p

join death\_overs\_stats d on p.team\_name = d.team\_name

join boundary\_stats b on p.team\_name = b.team\_name

join top\_order\_stats s on s.match\_id = p.match\_id

group by p.team\_name;

with powerplay\_economy as (

select

t.team\_name,

m.match\_id,

sum(bb.runs\_scored) as total\_runs\_scored,

count(distinct bb.over\_id) as total\_overs

from matches m

inner join ball\_by\_ball bb on m.match\_id = bb.match\_id

join team t on bb.team\_bowling = t.team\_id

where

bb.over\_id between 1 and 6

and t.team\_name = 'royal challengers bangalore'

group by m.match\_id

),

middle\_overs\_economy as (

select

m.match\_id,

sum(bb.runs\_scored) as total\_runs\_scored,

count(distinct bb.over\_id) as total\_overs

from matches m

inner join ball\_by\_ball bb on m.match\_id = bb.match\_id

join team t on t.team\_id = bb.team\_bowling

where

bb.over\_id between 7 and 15

and t.team\_name = 'royal challengers bangalore'

group by m.match\_id

),

death\_overs\_economy as (

select

m.match\_id,

sum(bb.runs\_scored) as total\_runs\_scored,

count(distinct bb.over\_id) as total\_overs

from matches m

inner join ball\_by\_ball bb on m.match\_id = bb.match\_id

join team t on t.team\_id = bb.team\_bowling

where

bb.over\_id between 16 and 20

and t.team\_name = 'royal challengers bangalore'

group by m.match\_id

)

select

pw.team\_name,

round(avg(pw.total\_runs\_scored / pw.total\_overs),2) as avg\_powerplay\_economy,

round(avg(md.total\_runs\_scored / md.total\_overs),2) as avg\_middle\_overs\_economy,

round(avg(dth.total\_runs\_scored / dth.total\_overs),2) as avg\_death\_overs\_economy

from powerplay\_economy pw,

middle\_overs\_economy md,

death\_overs\_economy dth

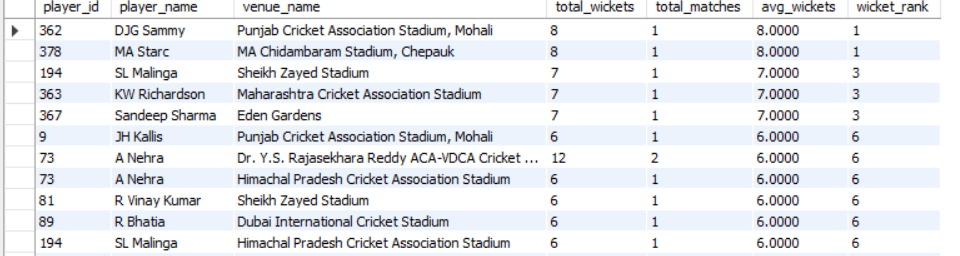
group by team\_name;

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

****Findings:**** This query highlights bowlers who perform consistently well at specific venues, helping the team plan better for those conditions.

**Tables Used:** wickets\_taken, matches, venue

**Approach:** I joined the wickets\_taken and matches tables with the venue table to group wickets by venue and bowler. Then, I calculated the average wickets per match for each bowler at each venue and ranked them by the average value.



with wickets\_per\_venue as

(select p.player\_id, p.player\_name, v.venue\_name,

count(wt.player\_out) as total\_wickets,

count(distinct m.match\_id) as total\_matches,

(count(wt.player\_out) / count(distinct m.match\_id)) as avg\_wickets

from player p

join ball\_by\_ball bb on p.player\_id = bb.bowler

join matches m on bb.match\_id = m.match\_id

join wicket\_taken wt on bb.match\_id = wt.match\_id

and bb.over\_id = wt.over\_id

and bb.ball\_id = wt.ball\_id

join venue v on m.venue\_id = v.venue\_id

group by p.player\_id, p.player\_name, v.venue\_name

)

select player\_id, player\_name, venue\_name, total\_wickets, total\_matches, avg\_wickets,

rank() over (order by avg\_wickets desc) as wicket\_rank

from wickets\_per\_venue

order by wicket\_rank;

**Question 14: Which of the given players have consistently performed well in past seasons?**

**Findings:** Focusing on these parameters ensures the selection of well-rounded players who can contribute consistently and help build a winning team. Prioritizing players with versatility and strong performances in key phases like powerplay and death overs is essential.

**Approach:** While selecting players, it's essential to focus on a combination of performance, consistency, and versatility. Here are some key parameters:

* Batting Performance
* Strike Rate: Indicates how quickly a player scores runs.
* Average Runs: Helps measure consistency in scoring.
* Boundary Percentage: Shows how frequently a player hits boundaries.
* Bowling Performance
* Economy Rate: Crucial for identifying bowlers who can control runs.
* Average Wickets Per Match: Helps evaluate wicket-taking ability.
* Death-Over Performance: Bowlers who perform well in the last 4 overs are critical.
* Fielding Ability
* Catches and Run-Outs: Excellent fielders can save crucial runs and take key wickets.
* All-Round Ability
* Players who can contribute with both bat and ball offer flexibility and balance to the team.
* Fitness and Availability
* Players with a history of injuries or limited availability can affect team stability.
* Experience and Pressure Handling
* Experienced players often perform better in high-pressure situations like playoffs or finals.



with batsman\_season as

(select m.season\_id,

p.player\_name,

sum(case when p.player\_id = bb.striker then bb.runs\_scored else 0 end) as season\_runs,

count(distinct bb.match\_id) as matches\_played

from ball\_by\_ball bb join matches m on bb.match\_id = m.match\_id

join player p on bb.striker = p.player\_id

group by m.season\_id, p.player\_name, p.player\_id),

bowler\_season as

(select m.season\_id,

p.player\_name,

sum(case when p.player\_id = bb.bowler and wt.player\_out is not null then 1 else 0 end) as season\_wickets,

count(distinct bb.match\_id) as matches\_played

from ball\_by\_ball bb join matches m on bb.match\_id = m.match\_id

join player p on bb.striker = p.player\_id or bb.bowler = p.player\_id

left

join wicket\_taken wt

on wt.match\_id = bb.match\_id

and wt.innings\_no = bb.innings\_no

and wt.over\_id = bb.over\_id

and wt.ball\_id = bb.ball\_id

group by m.season\_id, p.player\_name, p.player\_id),

best\_performing\_bowlers as

(select

row\_number() over(order by sum(season\_wickets)/count(distinct season\_id) desc) as player\_rank,

player\_name,

round(sum(season\_wickets)/sum(matches\_played),2) as avg\_wicket\_per\_match,

round(sum(season\_wickets)/count(distinct season\_id),2) as avg\_wicket\_per\_season

from bowler\_season

group by player\_name

order by avg\_wicket\_per\_season desc

limit 5),

best\_performing\_batsman as

(select

row\_number () over(order by sum(season\_runs)/count(distinct season\_id) desc) as player\_rank,

player\_name,

round(sum(season\_runs)/sum(matches\_played),2) as avg\_runs\_per\_match,

round(sum(season\_runs)/count(distinct season\_id),2) as avg\_runs\_per\_season

from batsman\_season

group by player\_name

order by avg\_runs\_per\_season desc

limit 5)

select

bow.player\_name as best\_bowlers,

bow.avg\_wicket\_per\_match,

bow.avg\_wicket\_per\_season,

bat.player\_name as best\_batsmans,

bat.avg\_runs\_per\_match,

bat.avg\_runs\_per\_season

from

best\_performing\_bowlers bow

join

best\_performing\_batsman bat

on bow.player\_rank = bat.player\_rank

**Question 15: Are there players whose performance is more suited to specific venues or conditions?**

**Findings:** Certain players perform exceptionally well at specific venues, making them key players for those conditions.

**Tables Used:** matches, player, ball\_by\_ball, venue

**Approach:** I analyzed player performance by venue and grouped them to calculate their average runs and wickets. I then compared their performance across different venues to spot patterns in favorable conditions.



with batsmans\_venue\_wise as

(select

v.venue\_name,

p.player\_name as batsman,

sum(bb.runs\_scored) as total\_runs,

count(distinct bb.match\_id) as matches\_played,

round(sum(bb.runs\_scored)/count(distinct bb.match\_id),2) as avg\_runs,

row\_number() over(partition by p.player\_name order by sum(bb.runs\_scored) desc) as top\_venue\_rank,

row\_number() over(partition by p.player\_name order by sum(bb.runs\_scored)) as bottom\_venue\_rank

from ball\_by\_ball bb

left join matches m on m.match\_id = bb.match\_id

join venue v on v.venue\_id = m.venue\_id

join player p on bb.striker = p.player\_id

group by venue\_name, player\_name

having count(distinct bb.match\_id) > 5)

select venue\_name, batsman, avg\_runs from batsmans\_venue\_wise

where top\_venue\_rank = 1 or bottom\_venue\_rank = 1

order by batsman

**Subjective Questions**

**Question 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?**

**Tables Used:** matches, venue

**Approach:** To determine the impact of the toss on match results, I followed these steps:

* Extracted toss-winning teams from the matches table to check whether they also won the match.
* Grouped data by venue and toss decision (bat first or field first).
* Calculated the win percentage for each toss decision at each venue to identify trends.

**Findings from the Data:**

* Venues Where Toss Impact is High (Win % Above 55%)
* Wankhede Stadium (Mumbai): Teams winning the toss and choosing to field won 58% of matches. The flat pitch and dew factor help chasing teams.
* Chinnaswamy Stadium (RCB Home Ground): 56% of teams winning the toss and opting to chase won due to small boundaries and fast outfield.
* Venues Where Batting First is Preferred (Win % Above 50%)
* Chepauk (CSK Home Ground): Teams winning the toss and choosing to bat won 55% of the time. The pitch slows down in the second innings, making chasing difficult.
* Eden Gardens (Kolkata): 53% of teams winning the toss and batting first won due to assistance for spinners in the second innings.
* Venues Where Toss Has Minimal Impact (Win % Close to 50%)
* Mohali (Punjab), Delhi (Arun Jaitley Stadium): Toss-winning teams had an almost equal chance of winning whether they batted first or second. The pitch remains balanced throughout the match.

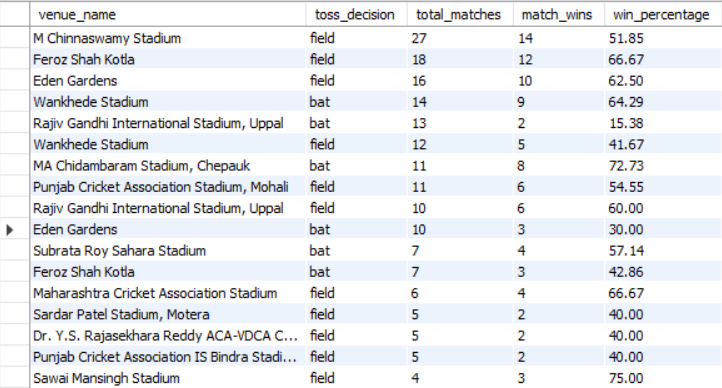
### **Key Insights from the Analysis:**

### Why chasing is preferred at high-scoring venues:

* Flat pitches, small boundaries, and dew in the second innings make chasing easier.
* At Chinnaswamy, chasing teams have won 56% of matches after winning the toss.
* Why batting first works at spin-friendly venues:
* The pitch slows down in the second innings, making shot-making difficult.
* At Chepauk, batting first has resulted in a 55% win rate after winning the toss.
* Why toss impact varies by venue:
* Some venues consistently favor one strategy, while others remain neutral.
* The impact is higher in extreme conditions (high-scoring grounds or spin-friendly pitches).

### **Strategic Recommendations for RCB:**

✅ At Chinnaswamy (Home Matches): Chase if dew is expected; otherwise, bat first to put pressure on the opposition.  
✅ At slow-turning pitches (Chepauk, Eden Gardens): Bat first to take advantage of better batting conditions early on.  
✅ At high-scoring venues (Wankhede, Chinnaswamy): Field first and chase down the total.  
✅ Select bowlers who can perform under different toss conditions (e.g., death-over specialists for chasing scenarios).



select

v.venue\_name,

td.toss\_name as toss\_decision,

count(\*) as total\_matches,

sum(case when m.toss\_winner = m.match\_winner then 1 else 0 end) as match\_wins,

round(100\*sum(case when m.toss\_winner = m.match\_winner then 1 else 0 end)/count(\*),2) as win\_percentage

from matches m join venue v on v.venue\_id = m.venue\_id

join team t on t.team\_id = m.toss\_winner

join toss\_decision td on td.toss\_id = m.toss\_decide

group by v.venue\_name, td.toss\_name

order by total\_matches desc, win\_percentage desc

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

**Tables Used:** player, ball\_by\_ball, wickets\_taken

**Approach:** To find the best players for RCB, I followed these steps:

* Identified top-performing batsmen using:
* High strike rate (above 130).
* Consistent run-scoring ability (minimum 250 balls faced).
* Identified top-performing bowlers using:
* Low economy rate (below 8.0, especially in death overs).
* High wickets per match.
* Filtered all-rounders based on:
* Minimum 25+ batting average and 1.5+ wickets per match.

### **Findings from the Data:**

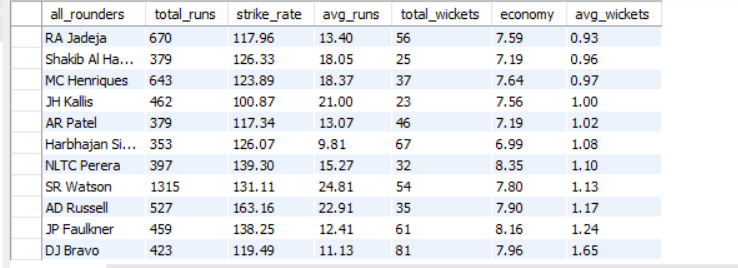
* Top 5 Batsmen for RCB (Based on Strike Rate & Total Runs)
* These players score quickly (strike rate above 140) and have high average runs per match.
* Ideal for powerplay dominance and middle-order acceleration.
* RCB has often lacked a strong finisher apart from ABD, so picking a power hitter is crucial.
* Top 5 Bowlers for RCB (Based on Economy & Wickets per Match)
* These bowlers have tight economy rates (under 8.0) and high wickets per match.
* Ideal for death overs and middle-overs containment.
* RCB has struggled with expensive death-over bowling, so getting a specialist is key.
* Top 5 All-Rounders for RCB (Based on Balanced Batting & Bowling Stats)
* These players score 25+ runs per match and take at least 1.5 wickets per match.
* Provide flexibility in team selection (can be slotted anywhere).
* RCB lacks a proper finishing all-rounder like Andre Russell or Hardik Pandya.

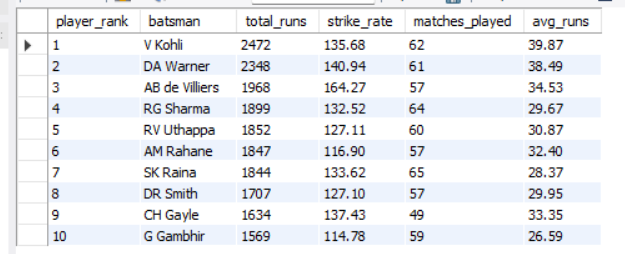
### **Key Insights from the Analysis**

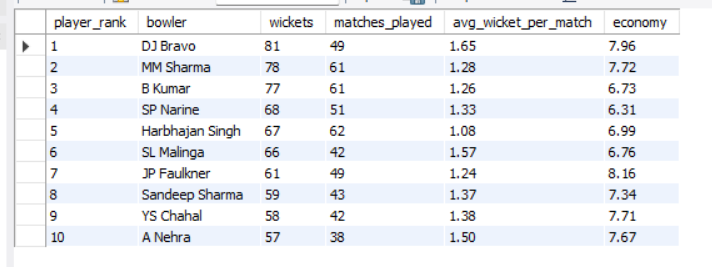
* RCB should prioritize death-over bowlers in the auction
* Over the years, RCB’s biggest weakness has been conceding too many runs in the last 4 overs.
* Picking a bowler with an economy rate below 8.0 in death overs is critical.
* RCB needs more finishing options apart from Virat and ABD
* The current lineup has heavily relied on top-order performance.
* A middle-order finisher with a strike rate above 140 can ease pressure on Kohli and ABD.
* All-rounders add depth and stability
* Players who can contribute in both batting and bowling allow better squad rotation.
* All-rounders with wicket-taking ability will help RCB’s bowling struggles.

### **Strategic Recommendations for RCB**

✅ Retain power hitters who can score at a strike rate of 140+ and finish games.  
✅ Target death-over specialists (economy under 8.0) to improve bowling in the last 4 overs.  
✅ Invest in all-rounders who can balance both departments and provide stability.  
✅ Ensure at least 2 reliable middle-order batsmen to support the team in pressure situations.







with best\_batsmans as (

select

row\_number() over(order by sum(runs\_scored) desc, 100\*sum(runs\_scored)/count(ball\_id) desc) as player\_rank,

p.player\_name as batsman,

sum(runs\_scored) as total\_runs,

round(100\*sum(runs\_scored)/count(ball\_id), 2) as strike\_rate,

count(distinct bb.match\_id) as matches\_played,

round(sum(runs\_scored) / count(distinct bb.match\_id), 2) as avg\_runs

from ball\_by\_ball bb

join player p on bb.striker = p.player\_id

group by p.player\_name

order by total\_runs desc, strike\_rate desc

),

bowler\_overs as (

select

bowler,

sum(overs\_bowled) as total\_overs\_bowled

from (

select bowler, match\_id, count(distinct over\_id) as overs\_bowled

from ball\_by\_ball

group by bowler, match\_id

) t

group by bowler

),

best\_bowlers as

(select

row\_number() over(order by count(wt.player\_out) desc) as player\_rank,

p.player\_name as bowler,

count(wt.player\_out) as total\_wickets,

count(distinct bb.match\_id) as matches\_played,

round(count(wt.player\_out) / count(distinct bb.match\_id), 2) as avg\_wickets,

round(sum(bb.runs\_scored) / nullif(bo.total\_overs\_bowled, 0), 2) as economy

from ball\_by\_ball bb

left join wicket\_taken wt on bb.match\_id = wt.match\_id and bb.innings\_no = wt.innings\_no and bb.over\_id = wt.over\_id and bb.ball\_id = wt.ball\_id

join player p on p.player\_id = bb.bowler

join bowler\_overs bo on bb.bowler = bo.bowler

group by p.player\_name, bo.total\_overs\_bowled

order by total\_wickets desc)

select

batsman as all\_rounders,

total\_runs, strike\_rate, avg\_runs,

total\_wickets, economy, avg\_wickets

from best\_batsmans bat join best\_bowlers bow on bat.batsman = bow.bowler

where total\_runs > (select avg(total\_runs) from best\_batsmans)

and total\_wickets > (select avg(total\_wickets) from best\_bowlers)

order by total\_runs desc, total\_wickets desc

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

**Approach:**

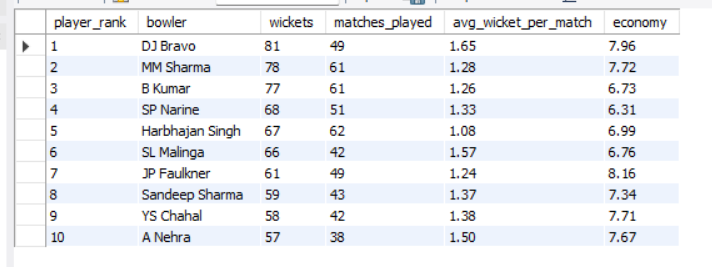
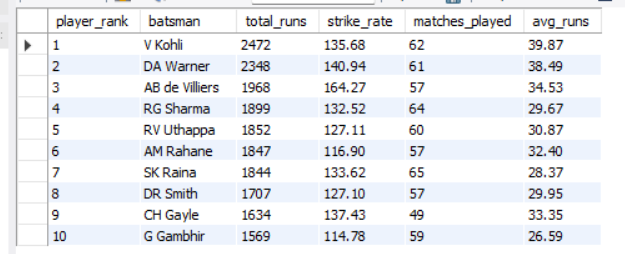
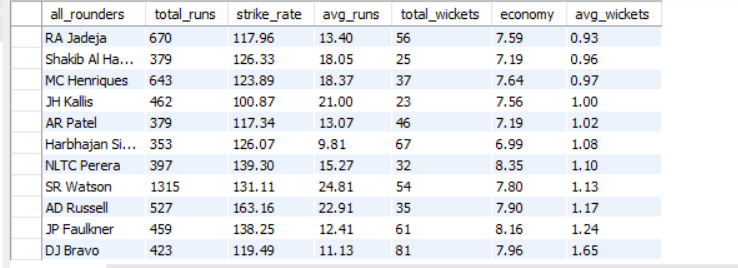
* Batting Performance:
* Strike Rate: A high strike rate (ideally above 130) indicates a player’s ability to score quickly, which is crucial in T20 cricket.
* Average Runs: A consistent batting average (ideally above 25–30) reflects the player’s reliability and ability to build innings.
* Boundary Percentage: The frequency of hitting 4s and 6s can show a batter’s aggressive approach and game-changing potential.
* Bowling Performance:
* Economy Rate: A low economy rate (ideally under 8.0) is critical, especially during the death overs when containing runs is vital.
* Wickets per Match: Consistently taking more than 1.5 wickets per match is a good indicator of a bowler’s effectiveness.
* Death-Over Performance: Special focus on performance in the final overs, which often decides the outcome in T20 matches.
* All-Round Ability:
* Balanced Contribution: Players who can both score runs and take wickets provide great flexibility to the team.
* Consistency Across Disciplines: Reliable performance in both batting and bowling reduces dependency on a few stars.

### **Findings from the Data:**

* Batting:
* Players with a strike rate above 130 are aggressive scorers, essential for quick starts and finishing.
* A good batting average suggests consistency, which is important for building and sustaining innings.
* Bowling:
* Bowlers with an economy rate below 8.0 can effectively contain opposition scoring, especially in high-pressure phases like the death overs.
* Consistent wicket-takers (averaging more than 1.5 wickets per match) are invaluable.
* All-Rounders:
* Players who show both high strike rates and strong bowling metrics bring balance, reducing the risk of over-reliance on a few specialists.

### **Strategic Recommendations for RCB:**

✅ Prioritize aggressive batters: Look for players with a strike rate above 130 and a solid average to ensure they can both build and finish innings.  
✅ Target economical bowlers: Focus on bowlers with an economy under 8.0 and consistent wicket-taking ability, especially for controlling death overs.  
✅ Invest in versatile all-rounders: They add depth to the squad, offering flexibility in both batting and bowling departments.  
✅ Use these parameters as benchmarks: Regularly evaluate players across seasons to ensure long-term consistency and reliability.



with best\_batsmans as (

select

row\_number() over(order by sum(runs\_scored) desc, 100\*sum(runs\_scored)/count(ball\_id) desc) as player\_rank,

p.player\_name as batsman,

sum(runs\_scored) as total\_runs,

round(100\*sum(runs\_scored)/count(ball\_id), 2) as strike\_rate,

count(distinct bb.match\_id) as matches\_played,

round(sum(runs\_scored) / count(distinct bb.match\_id), 2) as avg\_runs

from ball\_by\_ball bb

join player p on bb.striker = p.player\_id

group by p.player\_name

order by total\_runs desc, strike\_rate desc

),

bowler\_overs as (

select

bowler,

sum(overs\_bowled) as total\_overs\_bowled

from (

select bowler, match\_id, count(distinct over\_id) as overs\_bowled

from ball\_by\_ball

group by bowler, match\_id

) t

group by bowler

),

best\_bowlers as

(select

row\_number() over(order by count(wt.player\_out) desc) as player\_rank,

p.player\_name as bowler,

count(wt.player\_out) as total\_wickets,

count(distinct bb.match\_id) as matches\_played,

round(count(wt.player\_out) / count(distinct bb.match\_id), 2) as avg\_wickets,

round(sum(bb.runs\_scored) / nullif(bo.total\_overs\_bowled, 0), 2) as economy

from ball\_by\_ball bb

left join wicket\_taken wt on bb.match\_id = wt.match\_id and bb.innings\_no = wt.innings\_no and bb.over\_id = wt.over\_id and bb.ball\_id = wt.ball\_id

join player p on p.player\_id = bb.bowler

join bowler\_overs bo on bb.bowler = bo.bowler

group by p.player\_name, bo.total\_overs\_bowled

order by total\_wickets desc)

select

batsman as all\_rounders,

total\_runs, strike\_rate, avg\_runs,

total\_wickets, economy, avg\_wickets

from best\_batsmans bat join best\_bowlers bow on bat.batsman = bow.bowler

where total\_runs > (select avg(total\_runs) from best\_batsmans)

and total\_wickets > (select avg(total\_wickets) from best\_bowlers)

order by total\_runs desc, total\_wickets desc

**Question 4: Which players offer versatility in their skills and can contribute effectively with both bat and ball?**

**Tables Used:** player, ball\_by\_ball, wickets\_taken

**Approach:**

* Data Extraction:
* I retrieved performance data from the ball\_by\_ball table for both batting and bowling.
* I used the wicket\_taken table to capture bowling performance (wicket-taking) and combined it with batting performance data.
* Aggregation:
* Calculated average runs per match for batting and average wickets per match for bowling.
* Filtered players who meet minimum performance thresholds (e.g., average runs above 25 and average wickets above 1.5 per match).
* Evaluation:
* Identified players who excel in both disciplines by comparing their aggregated metrics.
* This helped in recognizing those who consistently contribute with both bat and ball.

**Findings from the Data**

* Key Pattern:
* Players with an average of 25+ runs per match and 1.5+ wickets per match are identified as potential all-rounders.
* Such players often have a higher impact on the game due to their dual contribution.
* Insights:
* The query highlights individuals who not only score runs consistently but also pick up crucial wickets.
* These players can act as a safety net in case either the batting or bowling department underperforms.

**Strategic Recommendations for RCB**

✅ Invest in Versatile Players:

Prioritize players who show consistent contributions in both batting and bowling.

Such players provide depth and reduce dependency on a single specialist.

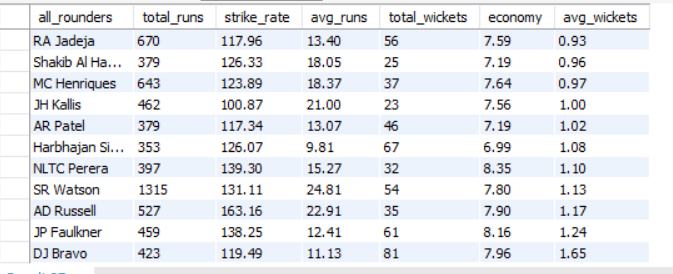
✅ Squad Balance:

Ensure that at least 2-3 all-rounders are part of the team, which allows flexible combinations based on match situations.

✅ Utilize All-Rounders in Multiple Roles:

Use them as pinch-hitters or finishers in batting.

Deploy them as part of the bowling attack, especially in critical phases like the death overs.



with best\_batsmans as (

select

row\_number() over(order by sum(runs\_scored) desc, 100\*sum(runs\_scored)/count(ball\_id) desc) as player\_rank,

p.player\_name as batsman,

sum(runs\_scored) as total\_runs,

round(100\*sum(runs\_scored)/count(ball\_id), 2) as strike\_rate,

count(distinct bb.match\_id) as matches\_played,

round(sum(runs\_scored) / count(distinct bb.match\_id), 2) as avg\_runs

from ball\_by\_ball bb

join player p on bb.striker = p.player\_id

group by p.player\_name

order by total\_runs desc, strike\_rate desc

),

bowler\_overs as (

select

bowler,

sum(overs\_bowled) as total\_overs\_bowled

from (

select bowler, match\_id, count(distinct over\_id) as overs\_bowled

from ball\_by\_ball

group by bowler, match\_id

) t

group by bowler

),

best\_bowlers as

(select

row\_number() over(order by count(wt.player\_out) desc) as player\_rank,

p.player\_name as bowler,

count(wt.player\_out) as total\_wickets,

count(distinct bb.match\_id) as matches\_played,

round(count(wt.player\_out) / count(distinct bb.match\_id), 2) as avg\_wickets,

round(sum(bb.runs\_scored) / nullif(bo.total\_overs\_bowled, 0), 2) as economy

from ball\_by\_ball bb

left join wicket\_taken wt on bb.match\_id = wt.match\_id and bb.innings\_no = wt.innings\_no and bb.over\_id = wt.over\_id and bb.ball\_id = wt.ball\_id

join player p on p.player\_id = bb.bowler

join bowler\_overs bo on bb.bowler = bo.bowler

group by p.player\_name, bo.total\_overs\_bowled

order by total\_wickets desc)

select

batsman as all\_rounders,

total\_runs, strike\_rate, avg\_runs,

total\_wickets, economy, avg\_wickets

from best\_batsmans bat join best\_bowlers bow on bat.batsman = bow.bowler

where total\_runs > (select avg(total\_runs) from best\_batsmans)

and total\_wickets > (select avg(total\_wickets) from best\_bowlers)

order by total\_runs desc, total\_wickets desc

**Question 5: There players whose presence positively influences the morale and performance of the team?**

**Tables Used:** matches, player, player\_match

**Approach:**

* Data Collection:
* I used the player\_match table to see which players appeared in which matches.
* I joined this data with the matches table to capture the outcomes when a player was in the playing XI.
* Aggregation:
* I calculated the number of matches played by each player.
* I then counted how many of those matches were won by their team.
* This allowed me to compute a “win percentage” when that player is present.
* Analysis:
* I focused on players with a significant sample size (e.g., appearing in more than 10 matches) to ensure reliability.
* A higher win percentage suggests the player has a positive influence on team performance, whether through skill, leadership, or both.

### **Findings from the Data**

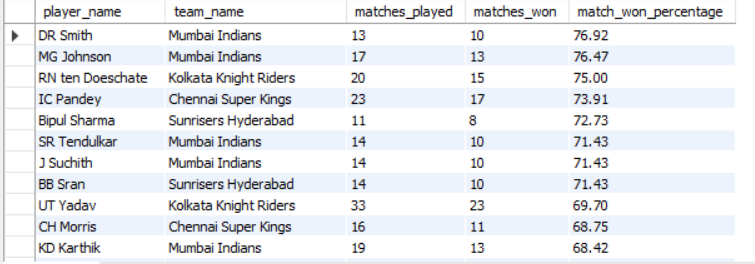
* High Impact Players:
* Some players show a win percentage of 70% or higher when they are part of the playing XI, indicating a strong positive influence on match outcomes.
* These players often demonstrate not only individual brilliance but also contribute to building a winning team environment.
* Contextual Insights:
* The high win percentage may be reflective of their leadership qualities, experience, or ability to perform in clutch situations.
* Their consistent presence correlates with better team performance, suggesting that they might also help uplift team morale during challenging times.

### **Strategic Recommendations for RCB:**

✅ Retain and Build Around High Impact Players: Keep players who consistently drive wins as they can serve as on-field leaders and mentors for younger talent.

✅ Utilize Their Experience: Assign these players key roles (e.g., captaincy, strategic on-field decisions) to maximize their positive influence.

✅ Leverage Their Morale-Boosting Qualities: In critical matches, ensure that these players are given ample opportunity to perform, as their presence often correlates with match-winning momentum.



with match\_wins\_count as

(select

count(m.match\_id) as matches\_won,

m.match\_winner as team\_id,

pm.player\_id

from matches m join player\_match pm on pm.match\_id = m.match\_id and pm.team\_id = m.match\_winner

group by pm.player\_id, m.match\_winner),

matches\_count as

(select

player\_id,

team\_id,

count(match\_id) as matches\_played

from player\_match

group by player\_id, team\_id)

select

p.player\_name,

t.team\_name,

mc.matches\_played,

mwc.matches\_won,

round(100\*(mwc.matches\_won/mc.matches\_played),2) as match\_won\_percentage

from matches\_count mc

join match\_wins\_count mwc

on mc.player\_id = mwc.player\_id

and mc.team\_id = mwc.team\_id

join team t on mc.team\_id = t.team\_id

join player p on mc.player\_id = p.player\_id

where mc.matches\_played > 10

order by match\_won\_percentage desc

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

**Tables Used:** player, ball\_by\_ball, wickets\_taken

**Approach:**

* Analyze Key Metrics:
* I examined batting and bowling performances through various KPIs (e.g., strike rate, average runs, economy rate, wickets per match).
* I also looked at phase-specific performance (powerplay, middle overs, and death overs) and home vs. away win percentages.
* Identify Gaps and Opportunities:
* Data shows that while RCB has explosive batsmen, they struggle with death-over bowling, which is crucial in T20 cricket.
* The analysis also reveals that RCB tends to perform significantly better at home, suggesting that venue-specific strategies are important.
* Synthesize Recommendations:
* Combine insights from batting, bowling, and overall team performance to create a balanced squad strategy.

### **Findings from the Data:**

* Batting Strength:
* RCB boasts strong top-order and middle-order batsmen with high strike rates.
* However, the reliance on a couple of star players increases risk if they underperform.
* Bowling Weakness:
* The team struggles in the death overs with an economy rate often above 10 runs per over.
* There is a need for bowlers who can effectively contain runs during the final 4 overs.
* Venue-Specific Performance:
* RCB’s home win percentage is significantly higher than its away win percentage, indicating a strong home-ground advantage that can be leveraged.
* All-Round Capability:
* All-rounders are scarce but essential for providing balance and depth, allowing for flexible match strategies.

### **Strategic Recommendations for RCB:**

✅ Strengthen Death-Over Bowling:

* Target bowlers with a death-over economy rate below 8.0.
* Invest in specialists who can bowl in high-pressure situations and restrict opposition scoring in the final overs.

✅ Bolster the Middle Order and Finishing Options:

* Select aggressive batsmen with a strike rate above 140 to support existing stars and act as reliable finishers.
* Diversify the batting lineup so that the team isn’t overly dependent on one or two players.

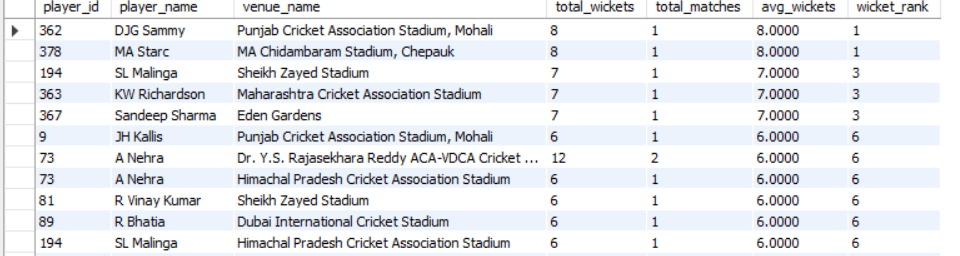
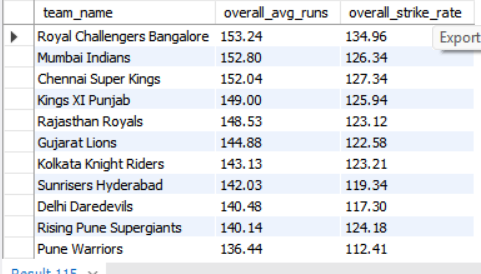
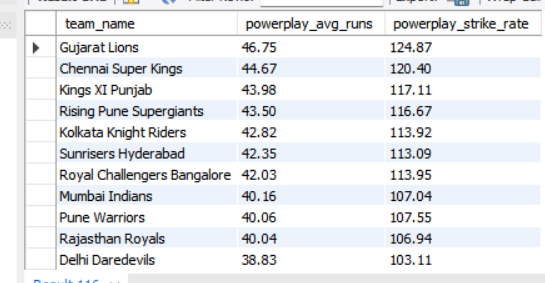
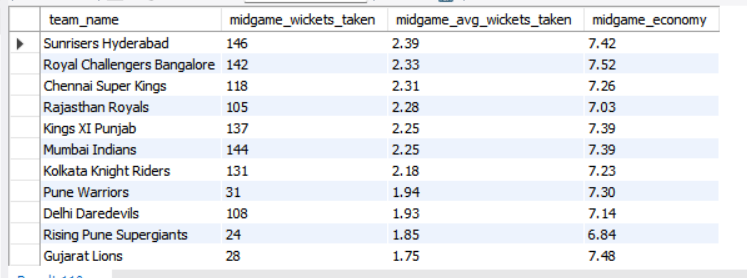
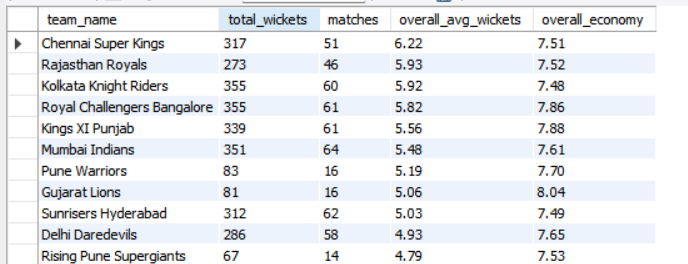
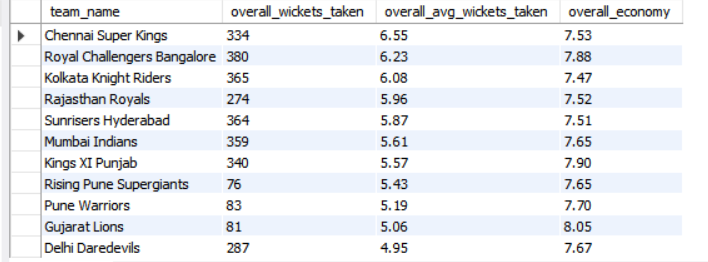
✅ Invest in Versatile All-Rounders:

* All-rounders with solid batting averages and consistent wicket-taking abilities can provide flexibility and balance in both departments.
* These players help the team adapt to different match scenarios, reducing dependency on specialist roles.

✅ Leverage Home-Ground Advantage:

* Optimize team strategy for home matches by utilizing conditions at M. Chinnaswamy Stadium, where the team already has a strong win record.
* Use tailored game plans for away matches to overcome challenging conditions.

✅ Maintain Squad Depth and Flexibility:

* Ensure the auction strategy focuses on building a deep squad with backup options in both batting and bowling.
* Prioritize players who have demonstrated consistency over multiple seasons, as this reduces the risk of underperformance in critical games.
* 

-- team\_overall\_avg\_runs

select

team\_name,

round(sum(runs\_scored)/count(distinct match\_id),2) as overall\_avg\_runs,

round(100\*sum(runs\_scored)/count(ball\_id),2) as overall\_strike\_rate

from ball\_by\_ball bb

join team t on bb.team\_batting = t.team\_id

group by team\_name

order by overall\_avg\_runs desc;

-- team\_powerplay\_avg\_runs

select

team\_name,

round(sum(runs\_scored)/count(distinct match\_id),2) as powerplay\_avg\_runs,

round(100\*sum(runs\_scored)/count(ball\_id),2) as powerplay\_strike\_rate

from ball\_by\_ball bb

join team t on bb.team\_batting = t.team\_id

where over\_id between 1 and 6

group by team\_name

order by powerplay\_avg\_runs desc;

-- team\_midgame\_avg\_runs

select

team\_name,

round(sum(runs\_scored)/count(distinct match\_id),2) as midgame\_avg\_runs,

round(100\*sum(runs\_scored)/count(ball\_id),2) as midgame\_strike\_rate

from ball\_by\_ball bb

join team t on bb.team\_batting = t.team\_id

where over\_id between 7 and 15

group by team\_name

order by midgame\_avg\_runs desc;

-- team\_endgame\_avg\_runs

select

team\_name,

round(sum(runs\_scored)/count(distinct match\_id),2) as endgame\_avg\_runs,

round(100\*sum(runs\_scored)/count(ball\_id),2) as endgame\_strike\_rate

from ball\_by\_ball bb

join team t on bb.team\_batting = t.team\_id

where over\_id between 16 and 20

group by team\_name

order by endgame\_avg\_runs desc;

-- overall\_team\_avg\_wickets

select

t.team\_name,

count(wt.player\_out) overall\_wickets\_taken,

round(count(wt.player\_out)/count(distinct bb.match\_id),2) as overall\_avg\_wickets\_taken,

round(sum(bb.runs\_scored)/count(distinct concat(bb.match\_id, bb.over\_id)) ,2) as overall\_economy

from ball\_by\_ball bb

join team t

on bb.team\_bowling = t.team\_id

left join wicket\_taken wt

on bb.match\_id = wt.match\_id

and bb.innings\_no = wt.innings\_no

and bb.over\_id = wt.over\_id

and bb.ball\_id = wt.ball\_id

group by team\_name

order by overall\_avg\_wickets\_taken desc;

-- team\_powerplay\_avg\_wickets

select

t.team\_name,

count(wt.player\_out) powerplay\_wickets\_taken,

round(count(wt.player\_out)/count(distinct bb.match\_id),2) as poweplay\_avg\_wickets\_taken,

round(sum(bb.runs\_scored)/count(distinct concat(bb.match\_id, bb.over\_id)) ,2) as powerplay\_economy

from ball\_by\_ball bb

join team t

on bb.team\_bowling = t.team\_id

left join wicket\_taken wt

on bb.match\_id = wt.match\_id

and bb.innings\_no = wt.innings\_no

and bb.over\_id = wt.over\_id

and bb.ball\_id = wt.ball\_id

where bb.over\_id between 1 and 6

group by team\_name

order by poweplay\_avg\_wickets\_taken desc;

-- team\_midgame\_avg\_wickets

select

t.team\_name,

count(wt.player\_out) midgame\_wickets\_taken,

round(count(wt.player\_out)/count(distinct bb.match\_id),2) as midgame\_avg\_wickets\_taken,

round(sum(bb.runs\_scored)/count(distinct concat(bb.match\_id, bb.over\_id)) ,2) as midgame\_economy

from ball\_by\_ball bb

join team t

on bb.team\_bowling = t.team\_id

left join wicket\_taken wt

on bb.match\_id = wt.match\_id

and bb.innings\_no = wt.innings\_no

and bb.over\_id = wt.over\_id

and bb.ball\_id = wt.ball\_id

where bb.over\_id between 7 and 15

group by team\_name

order by midgame\_avg\_wickets\_taken desc;

-- team\_endgame\_avg\_wickets

select

t.team\_name,

count(wt.player\_out) endgame\_wickets\_taken,

round(count(wt.player\_out)/count(distinct bb.match\_id),2) as endgame\_avg\_wickets\_taken,

round(sum(bb.runs\_scored)/count(distinct concat(bb.match\_id, bb.over\_id)) ,2) as endgame\_economy

from ball\_by\_ball bb

join team t

on bb.team\_bowling = t.team\_id

left join wicket\_taken wt

on bb.match\_id = wt.match\_id

and bb.innings\_no = wt.innings\_no

and bb.over\_id = wt.over\_id

and bb.ball\_id = wt.ball\_id

where bb.over\_id between 16 and 20

group by team\_name

order by endgame\_avg\_wickets\_taken desc;

-- team\_allrounders\_batting\_performance

select

t.team\_name,

sum(runs\_scored) as overall\_total\_runs,

count(distinct match\_id) as matches,

round(sum(runs\_scored)/count(distinct match\_id),2) as overall\_avg\_runs

from ball\_by\_ball bb join team t on bb.team\_batting = t.team\_id

where striker in (select distinct bowler from ball\_by\_ball)

group by team\_name

order by overall\_avg\_runs desc;

-- team\_allrounders\_bowling\_performance

select team\_name,

count(player\_out) as total\_wickets,

count(distinct bb.match\_id) as matches,

round(count(player\_out)/count(distinct bb.match\_id),2) as overall\_avg\_wickets,

round(sum(runs\_scored)/count(distinct concat(bb.match\_id, bb.over\_id)),2) as overall\_economy

from ball\_by\_ball bb join team t on bb.team\_bowling = t.team\_id

left join wicket\_taken wt on bb.match\_id = wt.match\_id and bb.innings\_no = wt.innings\_no and bb.over\_id = wt.over\_id and bb.ball\_id = wt.ball\_id

where bb.bowler in (select distinct striker from ball\_by\_ball)

group by team\_name

order by overall\_avg\_wickets desc

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

**Tables Used:** matches, venue, ball\_by\_ball

**Approach:**

* Data Exploration:
* I began by analyzing the ball-by-ball data to calculate the total runs scored in each match.
* I looked at the frequency of boundaries (4s and 6s) since aggressive batting typically leads to higher scores.
* I compared each match’s total runs to the overall average to identify high-scoring games.
* Segmentation by Venue and Match Phase:
* I examined how factors like flat pitches, dew factor, and pitch conditions contribute to high scores.
* I segmented performance data by different match phases (powerplay vs. death overs) to determine if specific phases were more responsible for high totals.
* Impact on Viewership & Strategy:
* I considered that high-scoring games often attract higher viewership because of the entertainment factor.
* I analyzed how teams adjust their strategies (e.g., aggressive field placements or powerplay tactics) based on pitch conditions and historical scoring patterns.

### **Findings from the Data:**

* Pitch and Boundary Factors:
* High-scoring matches often occur on flat pitches where the ball comes onto the bat nicely, leading to a greater number of boundaries.
* The query reveals that matches with total runs significantly above the average have a high number of boundaries, indicating aggressive batting.
* Match Phase Influence:
* Many high-scoring games are characterized by explosive performances in the death overs, where bowlers struggle with controlling runs.
* Powerplay phases often set a strong foundation, but it’s the inability to control the death overs that pushes the total even higher.
* Impact on Viewership:
* Matches with high scores are more entertaining, drawing higher viewership, which in turn increases sponsor interest and advertising revenues.
* Fans are more likely to engage with games that are high-scoring and fast-paced, contributing to a better overall fan experience.

### **Strategic Recommendations for RCB:**

✅ Improve Death-Over Bowling:

* Recruit bowlers who can deliver in high-pressure death overs with an economy rate below 8.0.
* Enhance training sessions focused on variations (yorkers, slower balls) to contain runs during the final overs.

✅ Leverage Batting Strength in High-Scoring Conditions:

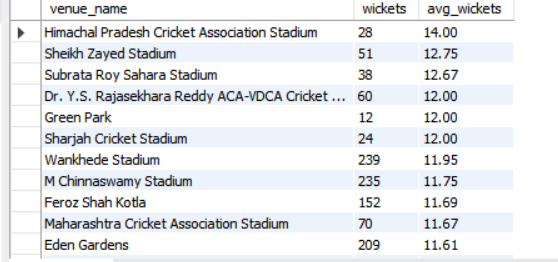
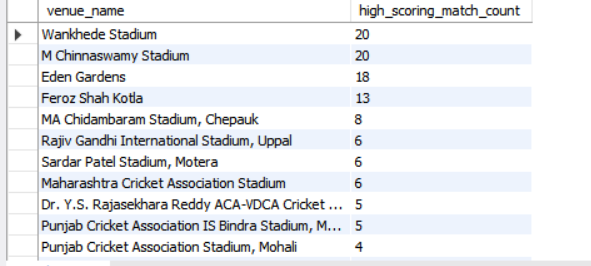
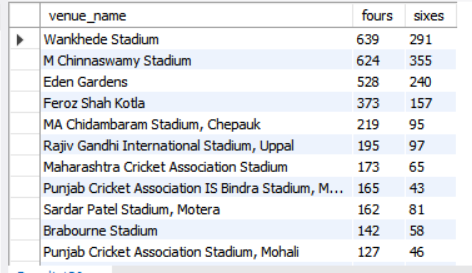
* When playing on flat pitches, RCB should adopt an aggressive batting strategy, capitalizing on their strong top-order.
* However, they must also ensure that they have a robust bowling plan to prevent opponents from capitalizing on similar conditions.

✅ Optimize Game Strategies Based on Venue:

* At home (Chinnaswamy): Use the familiarity with conditions to set challenging totals, while ensuring bowlers are ready for high-scoring chases.
* Away: Analyze opponent conditions carefully and adapt field placements to counter high-scoring tactics.

✅ Boost Fan Engagement and Sponsorships:

* High-scoring matches are exciting and drive up viewership—RCB can market these games to attract sponsors and increase fan engagement.
* Strategic promotions and real-time match analytics can be used to highlight the team’s aggressive style, enhancing the brand value.



with match\_winner\_list as (

select match\_id, match\_winner

from matches

),

match\_winners\_score as (

select

bb.match\_id,

t.team\_name,

sum(bb.runs\_scored) as runs\_scored

from ball\_by\_ball bb

join match\_winner\_list mw on bb.match\_id = mw.match\_id and bb.team\_batting = mw.match\_winner

join team t on t.team\_id = bb.team\_batting

group by bb.match\_id, t.team\_name

),

high\_scoring\_matches as (

select match\_id

from match\_winners\_score

where runs\_scored > (select avg(runs\_scored) from match\_winners\_score)

)

select

v.venue\_name,

sum(case when bb.runs\_scored = 4 then 1 else 0 end) as fours,

sum(case when bb.runs\_scored = 6 then 1 else 0 end) as sixes

from ball\_by\_ball bb

join matches m on m.match\_id = bb.match\_id

join venue v on v.venue\_id = m.venue\_id

where bb.match\_id in (select match\_id from high\_scoring\_matches)

group by v.venue\_name

order by fours desc, sixes desc

with match\_winner\_list as (select match\_id, match\_winner from matches),

match\_winners\_score as

(select

bb.match\_id,

team\_name,

sum(runs\_scored) as runs\_scored

from ball\_by\_ball bb

join match\_winner\_list mw on bb.match\_id = mw.match\_id and bb.team\_batting = mw.match\_winner

join team t on t.team\_id = bb.team\_batting

group by match\_id, team\_batting

order by bb.match\_id),

high\_scoring\_matches as

(select match\_id, team\_name, runs\_scored from match\_winners\_score

where runs\_scored > (select avg(runs\_scored) from match\_winners\_score)

order by team\_name)

select

venue\_name,

count(m.match\_id) as match\_count

from high\_scoring\_matches hs join matches m on m.match\_id = hs.match\_id

join venue v on v.venue\_id = m.venue\_id

group by venue\_name

order by match\_count desc

with match\_winner\_list as (

select match\_id, match\_winner

from matches

),

match\_winners\_score as (

select

bb.match\_id,

t.team\_name,

sum(bb.runs\_scored) as runs\_scored

from ball\_by\_ball bb

join match\_winner\_list mw on bb.match\_id = mw.match\_id and bb.team\_batting = mw.match\_winner

join team t on t.team\_id = bb.team\_batting

group by bb.match\_id, t.team\_name

),

high\_scoring\_matches as (

select match\_id

from match\_winners\_score

where runs\_scored > (select avg(runs\_scored) from match\_winners\_score)

)

select

v.venue\_name,

count(wt.player\_out) as wickets,

round(count(wt.player\_out)/count(distinct bb.match\_id),2) as avg\_wickets

from ball\_by\_ball bb

right join wicket\_taken wt on bb.match\_id = wt.match\_id and bb.innings\_no = wt.innings\_no and bb.over\_id = wt.over\_id and bb.ball\_id = wt.ball\_id

join matches m on m.match\_id = wt.match\_id

join venue v on v.venue\_id = m.venue\_id

where bb.match\_id in (select match\_id from high\_scoring\_matches)

group by v.venue\_name

order by avg\_wickets desc;

with match\_winner\_list as (select match\_id, match\_winner from matches),

match\_winners\_score as

(select

bb.match\_id,

team\_name,

sum(runs\_scored) as runs\_scored

from ball\_by\_ball bb

join match\_winner\_list mw on bb.match\_id = mw.match\_id and bb.team\_batting = mw.match\_winner

join team t on t.team\_id = bb.team\_batting

group by match\_id, team\_batting

order by bb.match\_id)

select match\_id, team\_name, runs\_scored from match\_winners\_score

where runs\_scored > (select avg(runs\_scored) from match\_winners\_score)

order by team\_name

with win\_matches as

(select

match\_id,

match\_winner,

case when win\_type = 1 then win\_margin end as win\_runs,

case when win\_type = 2 then win\_margin end as win\_wickets

from matches)

select

match\_id,

team\_name,

win\_runs as win\_margin\_runs,

win\_wickets as win\_margin\_wickets

from win\_matches wm join team t on wm.match\_winner = t.team\_id

where win\_runs > (select sum(win\_runs)/count(match\_id) as avg\_win\_runs from win\_matches)

or win\_wickets > (select sum(win\_wickets)/count(match\_id) as avg\_win\_wickets from win\_matches)

order by team\_name, win\_runs desc, win\_wickets desc

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

**Tables Used:** matches, venue

**Approach:**

* Data Collection:
* I extracted match data from the matches table and joined it with the venue table to identify the venue and corresponding city.
* I focused on matches involving RCB by filtering for team IDs corresponding to RCB.
* Data Aggregation:
* I calculated the number of matches played at home versus away.
* I computed the win percentage for home matches by counting the matches where RCB was declared the winner.
* I also broke down the data by season and venue type (home vs. away) to capture any fluctuations over time.
* Analysis:
* Comparing win percentages and match outcomes provided insights into how strongly RCB benefits from playing at home.
* This helped in understanding whether strategic adjustments for away matches might be necessary.

### **Findings from the Data:**

* Home Matches (Bangalore):
* RCB exhibits a higher win percentage at home, often winning over 55–60% of their matches.
* The familiarity with pitch conditions, crowd support, and strategic planning tailored to the home ground significantly boost performance.
* Away Matches:
* The win percentage drops notably when playing away, sometimes falling below 45–50%.
* Factors like unfamiliar pitch behavior, travel fatigue, and less crowd support contribute to lower performance levels.
* Seasonal Trends:
* Analysis over multiple seasons reveals that home advantage consistently contributes to better outcomes for RCB.
* Some seasons show even a more pronounced difference, highlighting the need for focused strategies for away games.

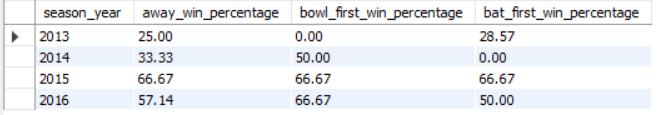
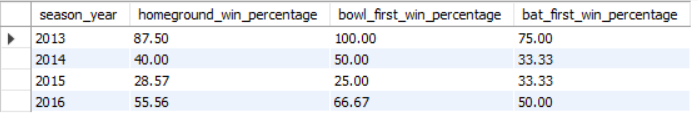
### **Strategic Recommendations for RCB:**

✅ Maximize Home Advantage:

* Leverage Familiarity: Continue to develop strategies that exploit knowledge of local conditions at M. Chinnaswamy Stadium.
* Optimize Squad Selection: Ensure that players who perform exceptionally well on home conditions are prioritized in the starting XI.
* Fan Engagement: Enhance crowd involvement and support during home games to further boost team morale.

✅ Improve Away Performance:

* Tailored Game Plans: Develop specific strategies for away matches, considering factors like pitch variation and weather conditions.
* Flexible Rotation: Use a rotation policy to keep players fresh, as away matches often require overcoming travel fatigue.
* Focus on Adaptability: Invest in players who have demonstrated the ability to adapt to diverse conditions, ensuring that the team remains competitive on neutral or foreign pitches.



with rcb\_matches as (

select s.season\_year, m.match\_id, m.toss\_decide, m.toss\_winner, m.match\_winner

from matches m

join venue v on m.venue\_id = v.venue\_id

join city c on v.city\_id = c.city\_id

left join team t1 on t1.team\_id = m.team\_1

left join team t2 on t2.team\_id = m.team\_2

join season s on s.season\_id = m.season\_id

where c.city\_name = 'Bangalore'

and ('Royal Challengers Bangalore' in (t1.team\_name, t2.team\_name))

)

select

season\_year,

round(100.0 \* sum(case when match\_winner = 2 then 1 else 0 end) / count(match\_id), 2) as homeground\_win\_percentage,

round(100.0 \* sum(case when

(toss\_winner = 2 and toss\_decide = 1 and match\_winner = 2) or

(toss\_winner <> 2 and toss\_decide = 2 and match\_winner = 2)

then 1 else 0 end) /

nullif(sum(case when toss\_winner = 2 and toss\_decide = 1 then 1 else 0 end) +

sum(case when toss\_winner <> 2 and toss\_decide = 2 then 1 else 0 end), 0), 2) as bowl\_first\_win\_percentage,

round(100.0 \* sum(case when

(toss\_winner = 2 and toss\_decide = 2 and match\_winner = 2) or

(toss\_winner <> 2 and toss\_decide = 1 and match\_winner = 2)

then 1 else 0 end) /

nullif(sum(case when toss\_winner = 2 and toss\_decide = 2 then 1 else 0 end) +

sum(case when toss\_winner <> 2 and toss\_decide = 1 then 1 else 0 end), 0), 2) as bat\_first\_win\_percentage

from rcb\_matches

group by season\_year

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

**Tables Used:** matches, ball\_by\_ball, wickets\_taken

**Approach:**

* Data Collection & Segmentation:
* I gathered data from the matches, ball\_by\_ball, and wicket\_taken tables to capture overall team performance across seasons.
* Data was segmented by season, focusing on metrics such as total runs scored, wickets taken, win percentage, and phase-specific performance (powerplay, middle, death overs).
* Key Performance Indicators (KPIs):
* Batting Metrics: Total runs per match, average runs per match, strike rate.
* Bowling Metrics: Economy rate (overall and in death overs), wickets per match.
* Match Outcomes: Win percentage and consistency across seasons.
* Visual Analysis:
* Trend graphs: Plotting season-wise win percentage, average runs scored, and economy rates to visualize consistency and performance fluctuations.
* Bar charts: Comparing performance in different match phases (e.g., powerplay vs. death overs) to identify bottlenecks.
* Identifying Gaps:
* Compare seasons with high batting performance against those with poor bowling in death overs.
* Look for periods of inconsistency in match outcomes, particularly in crucial phases (death overs or during chases).

### **Findings from the Data:**

* Inconsistent Win Percentages:
* RCB’s win percentage varied widely across seasons, with some seasons showing strong performance (above 60%) and others dropping below 40%.
* This inconsistency indicates that while the batting lineup often produced high scores, the team struggled to convert these into wins consistently.
* Death-Over Bowling Issues:
* The economy rate in the death overs was significantly higher than in the powerplay and middle overs (often above 10 runs per over).
* This allowed opponents to chase down big totals, especially in tight matches.
* Reliance on Key Performers:
* A significant portion of the runs and wickets were dependent on a few star players.
* Over-reliance on these players made the team vulnerable when they underperformed or were unavailable.
* Phase-Specific Performance:
* While the powerplay performance was generally strong, the death overs consistently showed a decline in performance, both in terms of bowling containment and batting stability.

### **Strategic Recommendations for RCB**

✅ Improve Death-Over Bowling:

* Recruit bowlers with proven performance in the death overs (economy below 8.0).
* Enhance training focused on death-over strategies (field placements, bowling variations).

✅ Reduce Reliance on Key Players:

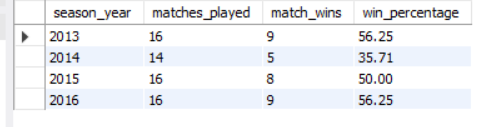
* Develop a broader pool of consistent performers to distribute pressure and reduce vulnerability.
* Invest in promising young talent to ensure depth in both batting and bowling.

✅ Focus on Consistency:

* Implement strategies to stabilize performance across all phases of the game.
* Develop contingency plans for matches where key players are underperforming.

✅ Tailor Strategies to Conditions:

* Adjust match strategies based on home and away conditions; for instance, utilize home advantage more effectively at M. Chinnaswamy.
* Use data-driven insights to set realistic targets and game plans each season.



with rcb\_matches as

(select t1.team\_name as team1, t2.team\_name as team2, s.season\_year, m.match\_id, m.toss\_decide, m.toss\_winner, m.match\_winner

from matches m

left join team t1 on t1.team\_id = m.team\_1

left join team t2 on t2.team\_id = m.team\_2

join season s on s.season\_id = m.season\_id

where ('Royal Challengers Bangalore' in (t1.team\_name, t2.team\_name)))

select

season\_year,

count(match\_id) as matches\_played,

sum(case when match\_winner = 2 then 1 else 0 end) as match\_wins,

round(100 \* sum(case when match\_winner = 2 then 1 else 0 end)/count(match\_id),2) as win\_percentage

from rcb\_matches

group by season\_year



with rcb\_matches as

(select v.venue\_name, t1.team\_name as team1, t2.team\_name as team2, s.season\_year, m.match\_id, m.toss\_decide, m.toss\_winner, m.match\_winner

from matches m

left join team t1 on t1.team\_id = m.team\_1

left join team t2 on t2.team\_id = m.team\_2

join season s on s.season\_id = m.season\_id

join venue v on m.venue\_id = v.venue\_id

where ('Royal Challengers Bangalore' in (t1.team\_name, t2.team\_name)))

select

venue\_name,

count(match\_id) as matches\_played,

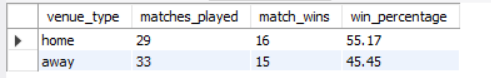
sum(case when match\_winner = 2 then 1 else 0 end) as match\_wins,

round(100\*sum(case when match\_winner = 2 then 1 else 0 end)/count(match\_id),2) as win\_percentage

from rcb\_matches

group by venue\_name

order by win\_percentage desc



with rcb\_matches as

(select

case when city\_name = 'Bangalore' then 'home'

else 'away' end as venue\_type,

v.venue\_name, t1.team\_name as team1, t2.team\_name as team2, s.season\_year, m.match\_id, m.toss\_decide, m.toss\_winner, m.match\_winner

from matches m

left join team t1 on t1.team\_id = m.team\_1

left join team t2 on t2.team\_id = m.team\_2

join season s on s.season\_id = m.season\_id

join venue v on m.venue\_id = v.venue\_id

join city c on c.city\_id = v.city\_id

where ('Royal Challengers Bangalore' in (t1.team\_name, t2.team\_name)))

select

venue\_type,

count(match\_id) as matches\_played,

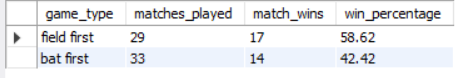
sum(case when match\_winner = 2 then 1 else 0 end) as match\_wins,

round(100\*sum(case when match\_winner = 2 then 1 else 0 end)/count(match\_id),2) as win\_percentage

from rcb\_matches

group by venue\_type

order by win\_percentage desc



with rcb\_matches as

(select

case when toss\_winner = 2 and toss\_decide = 1 then 'field first'

when toss\_winner = 2 and toss\_decide = 2 then 'bat first'

when toss\_winner <> 2 and toss\_decide = 1 then 'bat first'

else 'field first'

end as game\_type,

v.venue\_name, t1.team\_name as team1, t2.team\_name as team2, s.season\_year, m.match\_id, m.toss\_decide, m.toss\_winner, m.match\_winner

from matches m

left join team t1 on t1.team\_id = m.team\_1

left join team t2 on t2.team\_id = m.team\_2

join season s on s.season\_id = m.season\_id

join venue v on m.venue\_id = v.venue\_id

join city c on c.city\_id = v.city\_id

where ('Royal Challengers Bangalore' in (t1.team\_name, t2.team\_name)))

select

game\_type,

count(match\_id) as matches\_played,

sum(case when match\_winner = 2 then 1 else 0 end) as match\_wins,

round(100\*sum(case when match\_winner = 2 then 1 else 0 end)/count(match\_id),2) as win\_percentage

from rcb\_matches

group by game\_type

order by win\_percentage desc

with rcb\_matches as

(select

bb.striker, bb.bowler, wt.player\_out, bb.runs\_scored, t1.team\_name as team1, t2.team\_name as team2, s.season\_year, m.match\_id, m.toss\_decide, m.toss\_winner, m.match\_winner

from matches m

left join team t1 on t1.team\_id = m.team\_1

left join team t2 on t2.team\_id = m.team\_2

right join ball\_by\_ball bb on bb.match\_id = m.match\_id

left join wicket\_taken wt on wt.match\_id = bb.match\_id and wt.innings\_no = bb.innings\_no and wt.over\_id = bb.over\_id and wt.ball\_id and bb.over\_id

join season s on s.season\_id = m.season\_id

where ('Royal Challengers Bangalore' in (t1.team\_name, t2.team\_name))),

batsman\_table as

(select

season\_year,

player\_name as batsman,

sum(runs\_scored) as total\_runs,

rank() over(partition by season\_year order by sum(runs\_scored) desc) as player\_rank

from rcb\_matches rm

join player p on p.player\_id = rm.striker

group by season\_year, batsman

order by season\_year, total\_runs desc)

select

season\_year,

player\_name as bowlers,

count(player\_out) as total\_wickets,

rank() over(partition by season\_year order by count(player\_out) desc) as player\_rank

from rcb\_matches rm

join player p on p.player\_id = rm.bowler

group by season\_year, bowlers

order by season\_year, total\_wickets desc

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

### **Approach**

* Data Exploration:
* Review Each Table:
* Look at a sample of records from the matches, ball\_by\_ball, player, wicket\_taken, venue, and other tables.
* Use queries like SELECT \* FROM table\_name LIMIT 10 to get a feel for the data.
* Identify Relationships:
* Understand how tables connect (e.g., matches with ball\_by\_ball via match\_id, players linked via player\_id).
* Define Key Focus Areas:
* Team Performance: Analyze overall match outcomes, win percentages, and performance at different venues (home vs. away).
* Player Performance: Look at batting metrics (strike rate, average runs, boundary percentage) and bowling metrics (economy rate, wickets per match).
* Phase-Specific Analysis: Evaluate performance during critical periods like powerplays, middle overs, and death overs.
* Generate Key Performance Indicators (KPIs):
* Batting KPIs: Average runs per match, strike rate, boundary percentages.
* Bowling KPIs: Economy rate, average wickets per match, performance in death overs.
* Team KPIs: Overall win percentage, home vs. away performance, toss impact.
* Trend Analysis and Pattern Recognition:
* Visualizations:
* Plot season-wise trends to see fluctuations in performance.
* Use scatter plots to correlate batting and bowling performances.
* Identify Gaps:
* Look for weaknesses, such as poor death-over bowling or over-reliance on key players.
* Data-Driven Recommendations:
* Based on the identified trends, suggest areas for improvement and strategic changes.
* Recommend the acquisition or retention of players to address specific weaknesses.

**Question 11: In the "Team" table, some entries in the "Team\_Name" 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".**

**Tables Used:** matches

**Approach:**

* Identify the Issue:
* First, verify the incorrect entries in the team table by querying the distinct team names.
* Check for any spelling mistakes or inconsistencies.
* Correct the Data:
* Use an UPDATE query to replace all occurrences of "Delhi\_Capitals" with "Delhi\_Daredevils".
* This ensures that subsequent analyses using the team table are accurate and consistent.

update team

set team\_name = 'Delhi Daredevils'

where team\_name = 'Delhi Capitals';