**OBJECTIVE QUESTION**

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

**QUERY :** SELECT

COLUMN\_NAME,

DATA\_TYPE

FROM INFORMATION\_SCHEMA.COLUMNS

WHERE TABLE\_NAME = 'ball\_by\_ball';

|  |  |
| --- | --- |
| COLUMN\_NAME | DATA\_TYPE |
| Match\_Id | int |
| Over\_Id | int |
| Ball\_Id | int |
| Innings\_No | int |
| Team\_Batting | int |
| Team\_Bowling | int |
| Striker\_Batting\_Position | int |
| Striker | int |
| Non\_Striker | int |
| Bowler | int |
| Runs\_Scored | int |

*Table O\_1 - dtypes of columns in table ball\_by\_ball*

**Q2. What is the total number of runs scored in 1st season by RCB (bonus: also include the extra runs using the extra runs table)**

WITH extra\_run\_data AS (

SELECT

Team\_Batting AS team\_Id,

SUM(e.Extra\_Runs) as total\_extra

FROM ball\_by\_ball b

JOIN extra\_runs e ON e.Match\_Id = b.Match\_Id AND e.Innings\_No = b.Innings\_No AND e.Over\_Id = b.Over\_Id AND e.Ball\_Id = b.Ball\_Id

WHERE Team\_Batting = 2

AND b.Match\_Id IN(

SELECT distinct Match\_Id FROM matches WHERE Season\_Id = ( SELECT MIN(Season\_Id) as first\_season FROM Matches WHERE Team\_1 = 2 OR Team\_2 = 2))

),

run\_scored\_data AS (

SELECT

Team\_Batting AS team\_Id,

SUM(b.Runs\_Scored) AS total\_score

FROM ball\_by\_ball b

JOIN matches m ON m.Match\_Id = b.Match\_Id

WHERE Team\_Batting = 2 AND (Team\_1 = 2 OR Team\_2 = 2) AND m.Season\_Id = ( SELECT MIN(Season\_Id) as first\_season FROM Matches WHERE Team\_1 = 2 OR Team\_2 = 2)

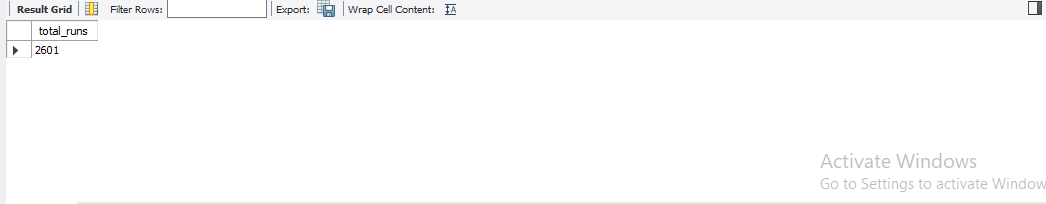
)

SELECT

(total\_score + total\_extra) AS total\_runs

FROM run\_scored\_data s

JOIN extra\_run\_data e ON e.team\_Id = s.team\_Id;



The total number of runs scored in 1st season by RCB including extra run is **2601.**

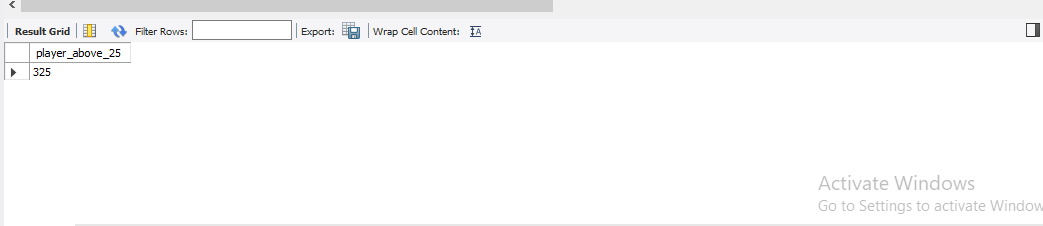
**Q3. How many players were more than the age of 25 during season 2014?**

SELECT

COUNT(\*) AS player\_above\_25

from player

WHERE TIMESTAMPDIFF(YEAR, DOB, (SELECT MIN(Match\_Date) FROM matches WHERE YEAR(Match\_Date) = 2014 )) > 25 ;



During season 2014 **325** player were more than the age of 25.

**Q4. How many matches did RCB win in 2013?**

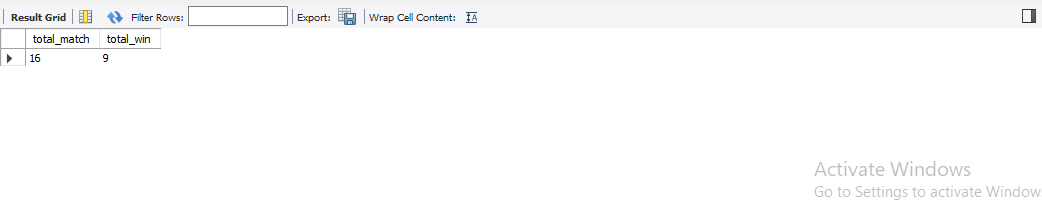
SELECT

COUNT(\*) as total\_match,

SUM( CASE WHEN Match\_Winner = 2 THEN 1 ELSE 0 END) AS total\_win

FROM matches

WHERE YEAR(Match\_Date) = 2013 AND (Team\_1 = 2 or Team\_2 = 2);



RCB played **16** Matches and they win **9** matches.

**Q5. List the top 10 players according to their strike rate in the last 4 seasons.**

WITH recent\_seasons AS (

SELECT s.Season\_Id

FROM season s

ORDER BY s.Season\_Id DESC

LIMIT 4

),

recent\_matches AS (

SELECT m.Match\_Id

FROM matches m

WHERE m.Season\_Id IN (SELECT Season\_Id FROM recent\_seasons)

)

SELECT

p.Player\_Id AS player\_id,

p.Player\_Name AS player\_name,

ROUND(100.0 \* SUM(b.Runs\_Scored) / COUNT(b.Runs\_Scored),2) AS strike\_rate

FROM player p

JOIN ball\_by\_ball b ON b.Striker = p.Player\_Id

WHERE b.Match\_Id IN (SELECT Match\_Id FROM recent\_matches)

AND NOT EXISTS (

SELECT 1

FROM extra\_runs e

WHERE b.Match\_Id = e.Match\_Id

AND b.Over\_Id = e.Over\_Id

AND b.Ball\_Id = e.Ball\_Id

AND e.Extra\_Type\_Id NOT IN(1,3)

AND b.Innings\_No = e.Innings\_No

)

GROUP BY p.Player\_Id, p.Player\_Name

HAVING COUNT(b.Runs\_Scored) >= 500

ORDER BY strike\_rate DESC

LIMIT 10;

|  |  |  |
| --- | --- | --- |
| player\_id | player\_name | strike\_rate |
| 110 | AB de Villiers | 167.52 |
| 305 | GJ Maxwell | 164.5 |
| 221 | KA Pollard | 148.92 |
| 187 | DA Warner | 145.18 |
| 320 | DA Miller | 145.1 |
| 162 | CH Gayle | 144.24 |
| 31 | YK Pathan | 144.15 |
| 20 | MS Dhoni | 142.51 |
| 8 | V Kohli | 139.1 |
| 21 | SK Raina | 137.57 |

*Table O\_2 – Top 10 player with their strike rate*

**Q6. What are the average runs scored by each batsman considering all the seasons?**

WITH grouped\_data AS (

SELECT

b.Match\_Id,

p.Player\_Name,

p.Player\_Id,

SUM(b.Runs\_Scored) AS batsman\_total\_score

FROM player p

JOIN ball\_by\_ball b ON b.Striker = p.Player\_Id

AND NOT EXISTS (

SELECT 1

FROM extra\_runs e

WHERE b.Match\_Id = e.Match\_Id

AND b.Over\_Id = e.Over\_Id

AND b.Ball\_Id = e.Ball\_Id

AND e.Extra\_Type\_Id NOT IN(1,3)

AND b.Innings\_No = e.Innings\_No

)

GROUP BY p.Player\_Id, p.Player\_Name, b.Match\_Id)

SELECT

Player\_Id,

Player\_Name,

ROUND(AVG(batsman\_total\_score), 2) AS avg\_score

FROM grouped\_data

GROUP BY Player\_Id, Player\_Name

HAVING SUM(batsman\_total\_score) > 500

ORDER BY avg\_score DESC;

|  |  |  |
| --- | --- | --- |
| Player\_Id | Player\_Name | avg\_score |
| 383 | LMP Simmons | 42.82 |
| 8 | V Kohli | 39.71 |
| 187 | DA Warner | 38.3 |
| 110 | AB de Villiers | 34.3 |
| 19 | MEK Hussey | 33.93 |
| 162 | CH Gayle | 32.94 |
| 85 | AM Rahane | 32.39 |
| 100 | SE Marsh | 30.9 |
| 46 | RV Uthappa | 30.78 |
| 147 | DR Smith | 29.86 |
| 57 | RG Sharma | 29.55 |
| 154 | JP Duminy | 28.94 |
| 21 | SK Raina | 28.28 |
| 254 | AJ Finch | 28.05 |
| 355 | Q de Kock | 27.88 |
| 42 | S Dhawan | 27.78 |
| 303 | F du Plessis | 26.94 |
| 2 | BB McCullum | 26.58 |
| 40 | G Gambhir | 26.41 |

*Table O\_3 – Player with avg runs scored in all season*

**Q7. What are the average wickets taken by each bowler considering all the seasons?**

WITH grouped\_data AS (

SELECT

m.Season\_Id,

b.Bowler AS Player\_Id,

p.Player\_Name,

COUNT(\*) AS total\_wicket

FROM wicket\_taken w

JOIN ball\_by\_ball b

ON w.Match\_Id = b.Match\_Id

AND w.Over\_Id = b.Over\_Id

AND w.Ball\_Id = b.Ball\_Id

AND w.Innings\_No = b.Innings\_No

JOIN player p ON p.Player\_Id = b.Bowler

JOIN matches m ON m.Match\_Id = w.Match\_Id

WHERE w.Kind\_Out != 3

GROUP BY b.Bowler, p.Player\_Name, m.Season\_Id

)

SELECT

Player\_Id,

Player\_Name,

ROUND(AVG(total\_wicket), 2) AS avg\_wicket

FROM grouped\_data

GROUP BY Player\_Id, Player\_Name

ORDER BY avg\_wicket DESC ;

|  |  |  |
| --- | --- | --- |
| Player\_Id | Player\_Name | avg\_wicket |
| 71 | DJ Bravo | 25 |
| 194 | SL Malinga | 20 |
| 382 | YS Chahal | 18.67 |
| 299 | B Kumar | 18.5 |
| 364 | MM Sharma | 17.5 |
| 403 | MJ McClenaghan | 17.5 |
| 378 | MA Starc | 17 |
| 460 | Mustafizur Rahman | 17 |
| 50 | Harbhajan Singh | 16.25 |
| 401 | D Wiese | 16 |
| 315 | SP Narine | 15.25 |
| 321 | Azhar Mahmood | 15 |
| 395 | Anureet Singh | 15 |
| 374 | AR Patel | 14.33 |
| 350 | CH Morris | 14 |
| 429 | BB Sran | 14 |
| 367 | Sandeep Sharma | 13.5 |
| 357 | MG Johnson | 13.25 |

*Table O\_4 – Player with avg wicket in all season*

**Q8. List all the players who have average runs scored greater than the overall average and who have taken wickets greater than the overall average.**

WITH player\_wickets AS (

SELECT

b.Bowler AS Player\_Id,

b.Match\_Id,

COUNT(\*) AS total\_wicket

FROM wicket\_taken w

JOIN ball\_by\_ball b

ON w.Match\_Id = b.Match\_Id

AND w.Over\_Id = b.Over\_Id

AND w.Ball\_Id = b.Ball\_Id

AND w.Innings\_No = b.Innings\_No

WHERE w.Kind\_Out != 3

GROUP BY b.Bowler, b.Match\_Id

),

player\_avg\_wicket AS (

SELECT

Player\_Id,

ROUND(AVG(total\_wicket), 2) AS avg\_wicket

FROM player\_wickets

GROUP BY Player\_Id

),

player\_runs AS (

SELECT

b.Match\_Id,

p.Player\_Name,

p.Player\_Id,

SUM(b.Runs\_Scored) AS batsman\_total\_score

FROM player p

JOIN ball\_by\_ball b ON b.Striker = p.Player\_Id

AND NOT EXISTS (

SELECT 1

FROM extra\_runs e

WHERE b.Match\_Id = e.Match\_Id

AND b.Over\_Id = e.Over\_Id

AND b.Ball\_Id = e.Ball\_Id

AND e.Extra\_Type\_Id NOT IN(1,3)

AND b.Innings\_No = e.Innings\_No

)

GROUP BY p.Player\_Id, p.Player\_Name, b.Match\_Id

) ,

player\_avg\_run AS (

SELECT

Player\_Id,

Player\_Name,

ROUND(AVG(batsman\_total\_score), 2) as avg\_run

FROM player\_runs pr

GROUP BY Player\_Id, Player\_Name

)

SELECT

r.Player\_Id,

Player\_Name,

r.avg\_run,

w.avg\_wicket

FROM player\_avg\_run r

JOIN player\_avg\_wicket w ON w.Player\_Id = r.Player\_Id

WHERE avg\_run > (SELECT ROUND(AVG(avg\_run),2) FROM player\_avg\_run )

AND avg\_wicket > (SELECT ROUND(AVG(avg\_wicket),2) FROM player\_avg\_wicket)

ORDER BY avg\_run DESC, avg\_wicket DESC;

|  |  |  |  |
| --- | --- | --- | --- |
| Player\_Id | Player\_Name | avg\_run | avg\_wicket |
| 162 | CH Gayle | 32.94 | 1.67 |
| 147 | DR Smith | 29.86 | 1.5 |
| 413 | KH Pandya | 29.63 | 1.5 |
| 409 | MP Stoinis | 29.2 | 2.67 |
| 154 | JP Duminy | 28.94 | 1.63 |
| 32 | SR Watson | 24.7 | 1.77 |
| 4 | DJ Hussey | 23.4 | 2 |
| 334 | AD Russell | 22.91 | 1.67 |
| 27 | Yuvraj Singh | 21.82 | 1.5 |
| 9 | JH Kallis | 21 | 1.54 |
| 321 | Azhar Mahmood | 18.36 | 2.14 |
| 161 | RS Bopara | 18.13 | 1.5 |
| 369 | Sachin Baby | 17.29 | 2 |
| 253 | MR Marsh | 16.33 | 1.57 |
| 401 | D Wiese | 15.75 | 1.78 |
| 280 | NLTC Perera | 15.15 | 1.71 |
| 362 | DJG Sammy | 14.75 | 1.83 |
| 385 | BCJ Cutting | 14.6 | 1.5 |
| 350 | CH Morris | 13.95 | 1.5 |
| 408 | CR Brathwaite | 13.83 | 1.75 |
| 35 | RA Jadeja | 13.34 | 1.82 |
| 374 | AR Patel | 13.07 | 1.48 |
| 329 | Gurkeerat Singh | 12.8 | 1.67 |

*Table O\_5 – Player with avg wicket & run in all season*

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

SELECT

v.Venue\_Id,

v.Venue\_Name,

COUNT(v.Venue\_Id) AS total\_matches,

SUM( CASE WHEN m.Match\_Winner = 2 THEN 1 ELSE 0 END ) AS total\_win,

SUM( CASE WHEN m.Match\_Winner != 2 THEN 1 ELSE 0 END ) AS total\_loss,

COUNT(v.Venue\_Id) - ( SUM( CASE WHEN m.Match\_Winner = 2 THEN 1 ELSE 0 END ) + SUM( CASE WHEN m.Match\_Winner != 2 THEN 1 ELSE 0 END )) AS `no\_result/draw`

FROM venue v

JOIN matches m ON m.Venue\_Id = v.Venue\_Id

WHERE (Team\_1 = 2 OR Team\_1 = 2)

GROUP BY v.Venue\_Id, v.Venue\_Name ;

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Venue\_Id | Venue\_Name | total\_matches | total\_win | total\_loss | no\_result/draw |
| 1 | M Chinnaswamy Stadium | 28 | 15 | 11 | 2 |
| 31 | Dubai International Cricket Stadium | 1 | 1 | 0 | 0 |
| 30 | Sharjah Cricket Stadium | 1 | 0 | 1 | 0 |
| 32 | Maharashtra Cricket Association Stadium | 1 | 1 | 0 | 0 |

*Table O\_6 – The wins & losses of RCB in individual venue*

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

WITH player\_bowlling\_Style AS (

SELECT

p.player\_Id,

p.player\_Name,

bs.Bowling\_Id,

bs.Bowling\_skill

FROM player p

JOIN bowling\_style bs ON bs.Bowling\_Id = p.Bowling\_skill

),

group\_bowler\_wicket AS (

SELECT

b.Bowler AS Player\_Id,

COUNT(\*) AS total\_wicket

FROM wicket\_taken w

JOIN ball\_by\_ball b

ON w.Match\_Id = b.Match\_Id

AND w.Over\_Id = b.Over\_Id

AND w.Ball\_Id = b.Ball\_Id

AND w.Innings\_No = b.Innings\_No

WHERE w.Kind\_Out != 3

GROUP BY b.Bowler

)

SELECT

pb.Bowling\_Id,

pb.Bowling\_skill,

SUM(gb.total\_wicket) AS total\_wicket,

ROUND( 100 \* SUM(gb.total\_wicket) / (SELECT SUM(total\_wicket) FROM group\_bowler\_wicket), 2) AS total\_wicket\_percentage

FROM group\_bowler\_wicket gb

JOIN player\_bowlling\_Style pb ON gb.Player\_Id = pb.Player\_Id

GROUP BY pb.Bowling\_Id, pb.Bowling\_skill

ORDER BY total\_wicket DESC;

|  |  |  |  |
| --- | --- | --- | --- |
| Bowling\_Id | Bowling\_skill | total\_wicket | total\_wicket\_percentage |
| 1 | Right-arm medium | 554 | 20.83 |
| 3 | Right-arm fast-medium | 333 | 12.52 |
| 5 | Right-arm medium-fast | 307 | 11.55 |
| 2 | Right-arm offbreak | 275 | 10.34 |
| 7 | Slow left-arm orthodox | 250 | 9.4 |
| 11 | Right-arm fast | 215 | 8.09 |
| 6 | Left-arm fast-medium | 163 | 6.13 |
| 4 | Legbreak googly | 157 | 5.9 |
| 10 | Legbreak | 143 | 5.38 |
| 9 | Left-arm medium-fast | 101 | 3.8 |
| 14 | Left-arm fast | 91 | 3.42 |
| 13 | Left-arm medium | 42 | 1.58 |
| 8 | Slow left-arm chinaman | 28 | 1.05 |

*Table O\_7 – Impact of bowling style in wicket taken*

**Q.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?**

WITH team\_run\_details AS (

SELECT

m.Season\_Id,

t.Team\_Id,

t.Team\_Name,

SUM(b.Runs\_Scored) AS total\_run

FROM ball\_by\_ball b

JOIN team t ON b.Team\_Batting = t.Team\_Id

JOIN matches m ON m.Match\_Id = b.Match\_Id

GROUP BY t.Team\_Id, t.Team\_Name, m.Season\_Id

ORDER BY t.Team\_Id, m.Season\_Id desc

),

team\_wicket\_details AS(

SELECT

b.Team\_Bowling AS team\_Id,

m.Season\_Id,

COUNT(\*) AS total\_wicket

FROM wicket\_taken w

JOIN ball\_by\_ball b

ON w.Match\_Id = b.Match\_Id

AND w.Over\_Id = b.Over\_Id

AND w.Ball\_Id = b.Ball\_Id

AND w.Innings\_No = b.Innings\_No

JOIN matches m ON m.Match\_Id = b.Match\_Id

GROUP BY b.Team\_Bowling, m.Season\_Id

ORDER BY team\_Id, m.Season\_Id DESC

),

grouped\_data AS (

SELECT

tr.Season\_Id AS season,

tr.Team\_Id,

tr.Team\_Name,

tr.total\_run,

COALESCE(LEAD(tr.total\_run) OVER(PARTITION BY tr.Team\_Id ORDER BY tr.Team\_Id, tr.Season\_Id DESC ), 0) AS previous\_season\_run,

tw.total\_wicket,

COALESCE(LEAD(tw.total\_wicket) OVER(PARTITION BY tr.Team\_Id ORDER BY tr.Team\_Id, tr.Season\_Id DESC),0) AS previous\_season\_wicket

FROM team\_wicket\_details tw

JOIN team\_run\_details tr ON tr.team\_Id = tw.team\_Id AND tr.Season\_Id = tw.Season\_Id

)

SELECT

season,

Team\_Id,

Team\_Name,

total\_run,

previous\_season\_run,

(CASE

WHEN total\_run > previous\_season\_run AND previous\_season\_run != 0 THEN "Better"

WHEN previous\_season\_run = 0 THEN "Not played Previous Season"

ELSE "Worse"

END ) AS batting\_status,

total\_wicket,

previous\_season\_wicket,

( CASE

WHEN total\_wicket > previous\_season\_wicket AND previous\_season\_wicket != 0 THEN "Better"

WHEN previous\_season\_wicket = 0 THEN "Not played Previous Season"

ELSE "Worse"

END ) AS bowling\_status,

( CASE

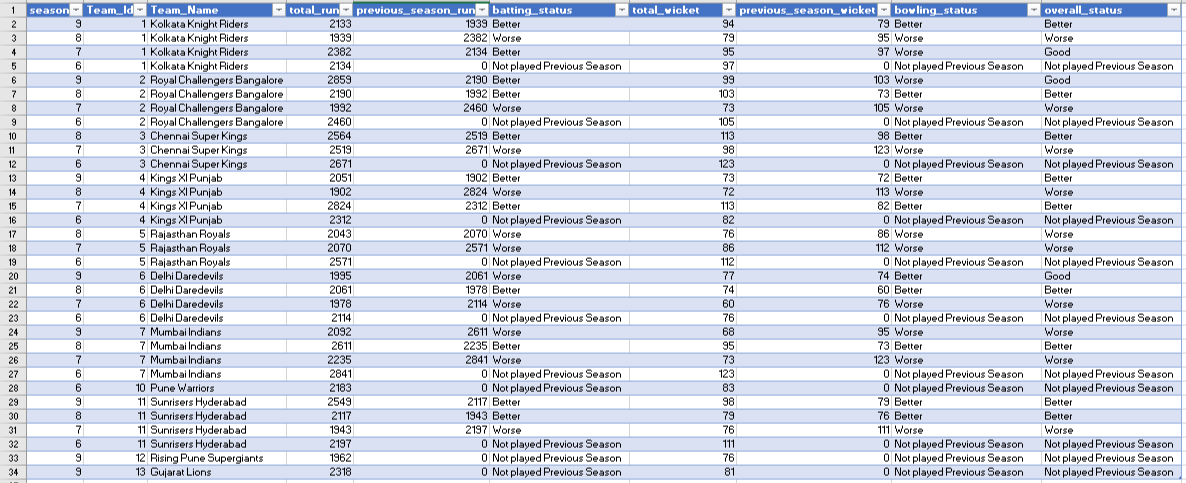
WHEN total\_wicket > previous\_season\_wicket AND total\_run > previous\_season\_run AND previous\_season\_run != 0 AND previous\_season\_wicket != 0 THEN "Better"

WHEN ((total\_wicket < previous\_season\_wicket) AND (total\_run > previous\_season\_run)) OR ((total\_wicket > previous\_season\_wicket) AND (total\_run < previous\_season\_run)) AND previous\_season\_run != 0 AND previous\_season\_wicket != 0 THEN "Good"

WHEN previous\_season\_wicket = 0 AND previous\_season\_run = 0 THEN "Not played Previous Season"

ELSE "Worse" END ) AS overall\_status

FROM grouped\_data;



*Table O\_8 – Analysis team performance with their previous performance*

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

Here are some key performance indicators (KPIs) that can be derived for the Royal Challengers Bangalore (RCB) team strategy:

**1. Top Order Stability**

* **Average Top Order Contribution**: Measures the average percentage contribution of top-order batsmen (positions 1-3) to the team's total runs in matches. A high percentage indicates strong top-order performance, contributing significantly to the team's overall score.

**2. Batting Performances**

* **Average Powerplay Runs**: The average runs scored by RCB during the powerplay overs (1-6). This indicates how effectively the team capitalizes on the initial overs.
* **Average Wickets Lost in Powerplay**: The average number of wickets lost during the powerplay, which highlights the stability and risk taken by the batting order early in the innings.
* **Average Death Overs Strike Rate**: The average strike rate during the death overs (17-20). This reflects how aggressively the team scores when the game is on the line.
* **Average Boundaries Per Match**: The average number of boundaries (fours and sixes) scored per match, indicating the team's ability to score quickly and keep the run rate high.
* **Boundary Per Ball**: This KPI measures the frequency of boundaries relative to total balls faced, providing insight into the batting aggression and efficiency.

**3. Bowling Performances**

* **Average Powerplay Economy**: The average runs conceded by RCB during the powerplay overs. A low economy rate here indicates effective bowling during the crucial early overs.
* **Average Middle Overs Economy**: The average runs conceded during the middle overs (7-15), which can show how well the team contains runs in the transitional phase of the game.
* **Average Death Overs Economy**: The average runs conceded in the death overs, which is crucial for understanding how well the bowlers perform under pressure.

**Strategic Insights:**

* **Batting Depth**: Analysis the contribution of top-order batsmen versus lower-order batsmen. If the top order consistently scores well, RCB might rely on them more heavily.
* **Powerplay Strategy**: High powerplay runs with low wickets lost suggest an aggressive approach that could be emphasized.
* **Aggression in Death Overs**: High strike rates in death overs with a reasonable economy could indicate a strategy focused on finishing strong, especially if they can also contain runs during earlier phases.
* **Bowling Consistency**: Economical bowling across all phases is critical; if there are spikes in the economy in middle or death overs, it could necessitate tactical adjustments.

By combining insights from these KPIs, RCB can make informed decisions on both batting and bowling strategies, tailoring their approach to maximize strengths and address weaknesses as needed throughout the season.**Query Strategy:**

**Q.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.**

WITH player\_per\_match\_wicket AS (

SELECT

b.Match\_Id,

b.Bowler AS Player\_Id,

p.Player\_Name,

COUNT(\*) AS total\_wicket

FROM wicket\_taken w

JOIN ball\_by\_ball b

ON w.Match\_Id = b.Match\_Id

AND w.Over\_Id = b.Over\_Id

AND w.Ball\_Id = b.Ball\_Id

AND w.Innings\_No = b.Innings\_No

JOIN player p ON p.Player\_Id = b.Bowler

WHERE w.Kind\_Out != 3

GROUP BY b.Bowler,b.Match\_Id

)

SELECT

m.Venue\_Id,

v.Venue\_Name,

p.Player\_Id,

p.Player\_Name,

ROUND(AVG(p.total\_wicket), 2) AS avg\_wicket,

ROW\_NUMBER() OVER(PARTITION BY m.Venue\_Id ORDER BY SUM(p.total\_wicket) DESC ,AVG(p.total\_wicket) DESC ) AS rnk

FROM player\_per\_match\_wicket p

JOIN matches m ON m.Match\_Id = p.Match\_Id

JOIN venue v ON m.Venue\_Id = v.Venue\_Id

GROUP BY m.Venue\_Id, p.Player\_Id, v.Venue\_Name, p.Player\_Name;

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Venue\_Id | Venue\_Name | Player\_Id | Player\_Name | avg\_wicket | rnk |
| 1 | M Chinnaswamy Stadium | 382 | YS Chahal | 1.76 | 1 |
| 1 | M Chinnaswamy Stadium | 32 | SR Watson | 2.29 | 2 |
| 1 | M Chinnaswamy Stadium | 378 | MA Starc | 2 | 3 |
| 1 | M Chinnaswamy Stadium | 81 | R Vinay Kumar | 1.71 | 4 |
| 1 | M Chinnaswamy Stadium | 270 | S Aravind | 1.71 | 5 |
| 1 | M Chinnaswamy Stadium | 61 | RP Singh | 2.2 | 6 |
| 1 | M Chinnaswamy Stadium | 434 | CJ Jordan | 2.5 | 7 |
| 1 | M Chinnaswamy Stadium | 73 | A Nehra | 2.67 | 8 |
| 1 | M Chinnaswamy Stadium | 349 | R Rampaul | 1.6 | 9 |
| 1 | M Chinnaswamy Stadium | 311 | HV Patel | 1.4 | 10 |
| 1 | M Chinnaswamy Stadium | 15 | Z Khan | 2 | 11 |
| 1 | M Chinnaswamy Stadium | 299 | B Kumar | 2 | 12 |
| 1 | M Chinnaswamy Stadium | 367 | Sandeep Sharma | 2 | 13 |
| 1 | M Chinnaswamy Stadium | 401 | D Wiese | 2 | 14 |
| 1 | M Chinnaswamy Stadium | 80 | DS Kulkarni | 2 | 15 |
| 1 | M Chinnaswamy Stadium | 346 | JD Unadkat | 1.5 | 16 |
| 1 | M Chinnaswamy Stadium | 27 | Yuvraj Singh | 2.5 | 17 |
| 1 | M Chinnaswamy Stadium | 296 | VR Aaron | 1.67 | 18 |
| 1 | M Chinnaswamy Stadium | 67 | PP Chawla | 1.67 | 19 |
| 1 | M Chinnaswamy Stadium | 392 | Karanveer Singh | 4 | 20 |

*Table O\_9 – Average wicket taken by each bowler in each venue*

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

**top 10 consistent Batsmen in all season with their run and strike rate:**

WITH batsman\_season\_runs AS (

SELECT

m.Season\_Id,

p.Player\_Id,

p.Player\_Name,

COUNT(DISTINCT m.Match\_Id) AS total\_Match,

SUM(b.Runs\_Scored) AS total\_run,

100 \* SUM(b.Runs\_Scored) / COUNT(b.Runs\_Scored) AS strike\_rate,

COUNT(DISTINCT m.Match\_Id) as match\_count

FROM ball\_by\_ball b

JOIN player p ON p.Player\_Id = b.Striker

JOIN matches m ON m.Match\_Id = b.Match\_Id

WHERE NOT EXISTS (

SELECT 1

FROM extra\_runs e

WHERE b.Match\_Id = e.Match\_Id

AND b.Over\_Id = e.Over\_Id

AND b.Ball\_Id = e.Ball\_Id

AND e.Extra\_Type\_Id NOT IN(1,3)

AND b.Innings\_No = e.Innings\_No

)

GROUP BY p.Player\_Name, p.Player\_Id, m.Season\_Id

)

SELECT

Player\_Id,

Player\_Name,

SUM(total\_Match) AS total\_match,

COUNT(DISTINCT Season\_Id) AS total\_season,

SUM(total\_run) AS total\_run,

ROUND(AVG(strike\_rate), 2) AS strike\_rate

FROM batsman\_season\_runs

GROUP BY Player\_Id, Player\_Name

HAVING COUNT(DISTINCT Season\_Id) > 3

ORDER BY total\_run DESC, strike\_rate DESC, total\_season DESC

LIMIT 10;

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Player\_Id | Player\_Name | total\_match | total\_season | total\_run | strike\_rate |
| 8 | V Kohli | 62 | 4 | 2462 | 136.1 |
| 187 | DA Warner | 61 | 4 | 2336 | 143.93 |
| 110 | AB de Villiers | 57 | 4 | 1955 | 167.03 |
| 57 | RG Sharma | 64 | 4 | 1891 | 134.33 |
| 46 | RV Uthappa | 60 | 4 | 1847 | 130.73 |
| 85 | AM Rahane | 57 | 4 | 1846 | 120.21 |
| 21 | SK Raina | 65 | 4 | 1838 | 136.67 |
| 147 | DR Smith | 57 | 4 | 1702 | 131.27 |
| 162 | CH Gayle | 49 | 4 | 1614 | 140.07 |
| 40 | G Gambhir | 59 | 4 | 1558 | 117.49 |

*Table O\_10 – Batsmen with total run and strike rate*

*Chart O\_1 – Batsmen with total run and strike rate*

**top 10 consistent Batsmen in all season with their run and strike rate:**

WITH bolwer\_season\_wicket AS (

SELECT

m.Season\_Id,

b.Bowler,

p.Player\_Name,

COUNT( DISTINCT b.Match\_Id) AS total\_matches,

COUNT( b.Match\_Id ) as total\_ball,

COUNT(w.Match\_Id) as total\_wicket

FROM ball\_by\_ball b

LEFT JOIN wicket\_taken w

ON w.Match\_Id = b.Match\_Id

AND w.Over\_Id = b.Over\_Id

AND w.Ball\_Id = b.Ball\_Id

AND w.Innings\_No = b.Innings\_No

AND w.Kind\_Out != 3

LEFT JOIN player p ON p.Player\_Id = b.Bowler

LEFT JOIN Matches m ON m.Match\_Id = b.Match\_Id

GROUP BY b.Bowler, m.Season\_Id, p.Player\_Name

)

SELECT

Bowler AS Player\_Id,

Player\_Name,

SUM(total\_matches) AS total\_match,

COUNT(DISTINCT Season\_Id) AS total\_season,

SUM(total\_wicket) AS total\_wicket,

ROUND(SUM(total\_ball) / SUM(total\_wicket), 2) AS bowling\_strike\_rate

FROM bolwer\_season\_wicket

GROUP BY Bowler, Player\_Name

HAVING COUNT(DISTINCT Season\_Id) >= 3

ORDER BY total\_wicket DESC ,total\_match DESC, bowling\_strike\_rate

LIMIT 10;

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Player\_Id | Player\_Name | total\_match | total\_season | total\_wicket | bowling\_strike\_rate |
| 71 | DJ Bravo | 49 | 3 | 75 | 14.19 |
| 299 | B Kumar | 61 | 4 | 74 | 19.23 |
| 364 | MM Sharma | 61 | 4 | 70 | 18.41 |
| 50 | Harbhajan Singh | 62 | 4 | 65 | 21.77 |
| 315 | SP Narine | 51 | 4 | 61 | 20.28 |
| 194 | SL Malinga | 42 | 3 | 60 | 17.05 |
| 382 | YS Chahal | 42 | 4 | 56 | 16.93 |
| 367 | Sandeep Sharma | 43 | 4 | 54 | 18.02 |
| 357 | MG Johnson | 43 | 4 | 53 | 19.68 |
| 201 | R Ashwin | 61 | 4 | 51 | 24.53 |

*Table O\_11 – Bowler with total wicket and bowling strike rate*

*Chart O\_2 – Bowler with total wicket and bowling strike rate*

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

**Batsmen performance in different venues or conditions:**

SELECT

m.Venue\_Id,

v.Venue\_Name,

b.Striker AS player\_Id,

p.Player\_Name,

SUM(b.Runs\_Scored) AS total\_run,

COUNT( DISTINCT m.Match\_Id) AS total\_match,

ROUND(SUM(b.Runs\_Scored) / COUNT( DISTINCT m.Match\_Id), 2) AS avg\_venue\_run,

ROUND(100 \* SUM(b.Runs\_Scored) / COUNT(b.Runs\_Scored), 2) AS strike\_rate

FROM ball\_by\_ball b

JOIN matches m ON b.Match\_Id = m.Match\_Id

JOIN player p ON p.Player\_Id = b.Striker

JOIN venue v ON v.Venue\_Id = m.Venue\_Id

WHERE NOT EXISTS (

SELECT 1

FROM extra\_runs e

WHERE b.Match\_Id = e.Match\_Id

AND b.Over\_Id = e.Over\_Id

AND b.Ball\_Id = e.Ball\_Id

AND e.Extra\_Type\_Id NOT IN(1,3)

AND b.Innings\_No = e.Innings\_No

)

GROUP BY b.Striker, m.Venue\_Id, p.Player\_Name

HAVING SUM(b.Runs\_Scored) > 200

ORDER BY total\_run DESC, avg\_venue\_run DESC, total\_match DESC;

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Venue\_Id | Venue\_Name | player\_Id | Player\_Name | total\_run | total\_match | avg\_venue\_run | strike\_rate |
| 1 | M Chinnaswamy Stadium | 8 | V Kohli | 1191 | 29 | 41.07 | 149.81 |
| 1 | M Chinnaswamy Stadium | 110 | AB de Villiers | 956 | 27 | 35.41 | 172.88 |
| 1 | M Chinnaswamy Stadium | 162 | CH Gayle | 946 | 26 | 36.38 | 155.34 |
| 4 | Wankhede Stadium | 57 | RG Sharma | 825 | 25 | 33 | 143.98 |
| 7 | Rajiv Gandhi International Stadium, Uppal | 187 | DA Warner | 712 | 16 | 44.5 | 158.93 |
| 5 | Eden Gardens | 40 | G Gambhir | 638 | 23 | 27.74 | 120.15 |
| 4 | Wankhede Stadium | 221 | KA Pollard | 636 | 25 | 25.44 | 164.34 |
| 5 | Eden Gardens | 46 | RV Uthappa | 593 | 17 | 34.88 | 135.08 |
| 7 | Rajiv Gandhi International Stadium, Uppal | 42 | S Dhawan | 533 | 20 | 26.65 | 114.13 |
| 5 | Eden Gardens | 31 | YK Pathan | 531 | 20 | 26.55 | 156.18 |
| 8 | MA Chidambaram Stadium, Chepauk | 20 | MS Dhoni | 467 | 15 | 31.13 | 151.62 |
| 4 | Wankhede Stadium | 208 | AT Rayudu | 433 | 21 | 20.62 | 138.78 |
| 3 | Feroz Shah Kotla | 154 | JP Duminy | 399 | 12 | 33.25 | 125.87 |
| 8 | MA Chidambaram Stadium, Chepauk | 21 | SK Raina | 395 | 15 | 26.33 | 130.36 |
| 5 | Eden Gardens | 96 | MK Pandey | 377 | 16 | 23.56 | 128.23 |
| 2 | Punjab Cricket Association Stadium, Mohali | 320 | DA Miller | 373 | 11 | 33.91 | 157.38 |
| 4 | Wankhede Stadium | 383 | LMP Simmons | 334 | 11 | 30.36 | 117.61 |
| 4 | Wankhede Stadium | 88 | KD Karthik | 318 | 10 | 31.8 | 134.18 |
| 8 | MA Chidambaram Stadium, Chepauk | 19 | MEK Hussey | 315 | 7 | 45 | 136.96 |

*Table O\_12 - Batsmen performance in different venues or conditions*

*Chart O\_3 – Total runs at each venue*

**Bowler performance in different venues or conditions:**

SELECT

v.Venue\_Id,

v.Venue\_Name,

b.Bowler AS Player\_Id,

p.Player\_Name,

COUNT( DISTINCT b.Match\_Id) AS total\_matches,

COUNT(w.Match\_Id) as total\_wicket

FROM ball\_by\_ball b

LEFT JOIN wicket\_taken w

ON w.Match\_Id = b.Match\_Id

AND w.Over\_Id = b.Over\_Id

AND w.Ball\_Id = b.Ball\_Id

AND w.Innings\_No = b.Innings\_No

AND w.Kind\_Out != 3

LEFT JOIN player p ON p.Player\_Id = b.Bowler

LEFT JOIN Matches m ON m.Match\_Id = b.Match\_Id

LEFT JOIN venue v ON v.Venue\_Id = m.Venue\_Id

GROUP BY b.Bowler, v.Venue\_Id, v.Venue\_Name, p.Player\_Name

ORDER BY total\_wicket DESC, total\_matches DESC;

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Venue\_Id | Venue\_Name | Player\_Id | Player\_Name | total\_matches | total\_wicket |
| 1 | M Chinnaswamy Stadium | 382 | YS Chahal | 20 | 30 |
| 4 | Wankhede Stadium | 50 | Harbhajan Singh | 25 | 29 |
| 4 | Wankhede Stadium | 194 | SL Malinga | 19 | 29 |
| 8 | MA Chidambaram Stadium, Chepauk | 71 | DJ Bravo | 15 | 27 |
| 5 | Eden Gardens | 315 | SP Narine | 17 | 23 |
| 4 | Wankhede Stadium | 357 | MG Johnson | 11 | 18 |
| 7 | Rajiv Gandhi International Stadium, Uppal | 163 | MC Henriques | 14 | 17 |
| 3 | Feroz Shah Kotla | 136 | A Mishra | 13 | 17 |
| 5 | Eden Gardens | 67 | PP Chawla | 15 | 16 |
| 8 | MA Chidambaram Stadium, Chepauk | 201 | R Ashwin | 13 | 16 |
| 8 | MA Chidambaram Stadium, Chepauk | 364 | MM Sharma | 13 | 16 |
| 1 | M Chinnaswamy Stadium | 32 | SR Watson | 12 | 16 |
| 3 | Feroz Shah Kotla | 386 | Imran Tahir | 10 | 16 |
| 7 | Rajiv Gandhi International Stadium, Uppal | 299 | B Kumar | 16 | 15 |
| 5 | Eden Gardens | 334 | AD Russell | 13 | 15 |
| 7 | Rajiv Gandhi International Stadium, Uppal | 136 | A Mishra | 12 | 15 |
| 4 | Wankhede Stadium | 403 | MJ McClenaghan | 10 | 14 |
| 3 | Feroz Shah Kotla | 232 | UT Yadav | 9 | 13 |
| 4 | Wankhede Stadium | 80 | DS Kulkarni | 7 | 13 |

*Table O\_13 - Bowler performance in different venues or conditions*

*Chart O\_4 – Total wickets at each venue*

**SUBJECTIVE QUESTION**

**Q1. 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?**

SELECT

m.Venue\_Id,

v.Venue\_Name,

t.Toss\_Name,

COUNT(\*) AS total\_match,

SUM(CASE WHEN Toss\_Winner = Match\_Winner THEN 1 ELSE 0 END) AS total\_win,

ROUND(100\* SUM(CASE WHEN Toss\_Winner = Match\_Winner THEN 1 ELSE 0 END) / COUNT(\*), 2) AS win\_percentage

FROM matches m

JOIN toss\_decision t ON t.Toss\_Id = m.Toss\_Decide

JOIN venue v ON v.venue\_Id = m.Venue\_Id

GROUP BY t.Toss\_Name, m.Venue\_Id, v.Venue\_Name

ORDER BY m.Venue\_Id, win\_percentage DESC;

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Venue\_Id | Venue\_Name | Toss\_Name | total\_match | total\_win | win\_percentage |
| 1 | M Chinnaswamy Stadium | Field | 27 | 14 | 51.85 |
| 1 | M Chinnaswamy Stadium | Bat | 3 | 1 | 33.33 |
| 2 | Punjab Cricket Association Stadium, Mohali | Field | 11 | 6 | 54.55 |
| 2 | Punjab Cricket Association Stadium, Mohali | Bat | 2 | 1 | 50 |
| 3 | Feroz Shah Kotla | Field | 18 | 12 | 66.67 |
| 3 | Feroz Shah Kotla | Bat | 7 | 3 | 42.86 |
| 4 | Wankhede Stadium | Bat | 14 | 9 | 64.29 |
| 4 | Wankhede Stadium | Field | 12 | 5 | 41.67 |
| 5 | Eden Gardens | Field | 16 | 10 | 62.5 |
| 5 | Eden Gardens | Bat | 10 | 3 | 30 |
| 6 | Sawai Mansingh Stadium | Field | 4 | 3 | 75 |
| 6 | Sawai Mansingh Stadium | Bat | 4 | 1 | 25 |
| 7 | Rajiv Gandhi International Stadium, Uppal | Field | 10 | 6 | 60 |
| 7 | Rajiv Gandhi International Stadium, Uppal | Bat | 13 | 2 | 15.38 |
| 8 | MA Chidambaram Stadium, Chepauk | Bat | 11 | 8 | 72.73 |
| 8 | MA Chidambaram Stadium, Chepauk | Field | 4 | 1 | 25 |
| 18 | Brabourne Stadium | Bat | 1 | 1 | 100 |
| 18 | Brabourne Stadium | Field | 3 | 1 | 33.33 |
| 19 | Sardar Patel Stadium, Motera | Field | 5 | 2 | 40 |
| 19 | Sardar Patel Stadium, Motera | Bat | 3 | 1 | 33.33 |
| 20 | Barabati Stadium | Field | 3 | 2 | 66.67 |
| 22 | Himachal Pradesh Cricket Association Stadium | Field | 2 | 0 | 0 |
| 25 | Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Stadium | Bat | 4 | 2 | 50 |
| 25 | Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Stadium | Field | 5 | 2 | 40 |
| 26 | Subrata Roy Sahara Stadium | Bat | 7 | 4 | 57.14 |
| 26 | Subrata Roy Sahara Stadium | Field | 1 | 0 | 0 |
| 27 | Shaheed Veer Narayan Singh International Stadium | Field | 3 | 2 | 66.67 |
| 27 | Shaheed Veer Narayan Singh International Stadium | Bat | 3 | 1 | 33.33 |
| 28 | JSCA International Stadium Complex | Field | 4 | 3 | 75 |
| 28 | JSCA International Stadium Complex | Bat | 3 | 1 | 33.33 |
| 29 | Sheikh Zayed Stadium | Bat | 4 | 3 | 75 |
| 29 | Sheikh Zayed Stadium | Field | 3 | 2 | 66.67 |
| 30 | Sharjah Cricket Stadium | Field | 4 | 2 | 50 |
| 30 | Sharjah Cricket Stadium | Bat | 2 | 0 | 0 |

*Table s\_1 - Percentage Win in Venue with respective toss decision*

*Chart s\_1 - Percentage Win in Venue with respective toss decision*

1. **Overall Insights on Toss Choices:**

* **M Chinnaswamy Stadium:** Fielding has a win percentage of 51.85%, while batting has only 33.33%.
* **Eden Gardens:** Fielding has a win percentage of 62.5%, compared to batting's 30%.
* **MA Chidambaram Stadium, Chepauk:** Batting has a high win percentage of 72.73%, while fielding lags at 25%.
* **Wankhede Stadium:** Batting win percentage is higher at 64.29%, compared to 41.67% for fielding.

1. **Venues with High Toss Influence on Outcomes:**

* **Sheikh Zayed Stadium:** Batting and fielding show high success rates (75% and 66.67%, respectively).
* **JSCA International Stadium Complex:** Fielding shows strong success (75%), compared to 33.33% for batting.
* **Barabati Stadium:** Fielding has a solid win rate of 66.67%.

1. **Venues with Extremely Favorable or Unfavorable Toss Outcomes:**

* **Brabourne Stadium:** Batting has a 100% win percentage in limited matches (1 win).
* **Himachal Pradesh Cricket Association Stadium:** Fielding has a win percentage of 0%, indicating challenges for fielding sides.

1. **Balanced Venues:**

* **Punjab Cricket Association Stadium, Mohali:** Fielding has a win percentage of 54.55%, while batting is close at 50%.
* **Sharjah Cricket Stadium:** Both toss outcomes have a 50% success rate for fielding.

1. **Venues with Very Low Batting Success:**

* **Sharjah Cricket Stadium:** Batting teams have a 0% win percentage across two matches.
* **Saurashtra Cricket Association Stadium:** Batting teams also show 0% success.

**Conclusion:**

This analysis suggests that teams should consider venue-specific statistics when deciding to bat or field after winning the toss. For example:

* At **M Chinnaswamy Stadium**, fielding may be more advantageous.
* At **MA Chidambaram Stadium**, batting appears to be a stronger choice.

**Q2. Suggest some of the players who would be best fit for the team.**

**Analysis of Top Batsmen:**

SELECT

p.Player\_Id,

p.Player\_Name,

COUNT(DISTINCT m.Match\_Id) AS total\_Match,

SUM(b.Runs\_Scored) AS total\_run,

ROUND( 100 \* SUM(b.Runs\_Scored) / COUNT(b.Runs\_Scored), 2) AS strike\_rate

FROM ball\_by\_ball b

JOIN player p ON p.Player\_Id = b.Striker

JOIN matches m ON m.Match\_Id = b.Match\_Id

GROUP BY p.Player\_Name, p.Player\_Id

ORDER BY total\_run DESC, strike\_rate DESC

LIMIT 10;

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Player\_Id | Player\_Name | total\_Match | total\_run | strike\_rate |
| 8 | V Kohli | 62 | 2472 | 138.57 |
| 187 | DA Warner | 61 | 2348 | 144.58 |
| 110 | AB de Villiers | 57 | 1968 | 167.77 |
| 57 | RG Sharma | 64 | 1899 | 134.2 |
| 46 | RV Uthappa | 60 | 1852 | 130.42 |
| 85 | AM Rahane | 57 | 1847 | 119.47 |
| 21 | SK Raina | 65 | 1844 | 137.61 |
| 147 | DR Smith | 57 | 1707 | 130.01 |
| 162 | CH Gayle | 49 | 1634 | 143.21 |
| 40 | G Gambhir | 59 | 1569 | 118.15 |

*Table s\_2 – Top 10 Batsman with their stats*

* **AB de Villiers (Player\_Id: 110)**: Total Runs: 1968, Strike Rate: 167.77

He has the highest strike rate among the top batsmen, making him an explosive option.

* **DA Warner (Player\_Id: 187)**: Total Runs: 2348, Strike Rate: 144.58

Combines consistency with aggressive scoring.

* **V Kohli (Player\_Id: 8)**: Total Runs: 2472, Strike Rate: 138.57

Most reliable batsman based on runs scored.

* **CH Gayle (Player\_Id: 162)**: Total Runs: 1634, Strike Rate: 143.21

Known for explosive starts and match-winning innings.

* **SK Raina (Player\_Id: 21)**: Total Runs: 1844, Strike Rate: 137.61

Reliable middle-order batsman.

* **RG Sharma (Player\_Id: 57)**: Total Runs: 1899, Strike Rate: 134.2

Consistent opener with stability.

* **AM Rahane (Player\_Id: 85)** and **G Gambhir (Player\_Id: 40)**:

While they are reliable, their strike rates (119.47 and 118.15, respectively) suggest they are better suited for anchoring innings.

**Analysis of Top Bowlers :-**

WITH bolwer\_season\_wicket AS (

SELECT

m.Season\_Id,

b.Bowler,

p.Player\_Name,

COUNT( DISTINCT b.Match\_Id) AS total\_matches,

COUNT( b.Match\_Id ) as total\_ball,

COUNT(w.Match\_Id) as total\_wicket

FROM ball\_by\_ball b

LEFT JOIN wicket\_taken w

ON w.Match\_Id = b.Match\_Id

AND w.Over\_Id = b.Over\_Id

AND w.Ball\_Id = b.Ball\_Id

AND w.Innings\_No = b.Innings\_No

AND w.Kind\_Out != 3

LEFT JOIN player p ON p.Player\_Id = b.Bowler

LEFT JOIN Matches m ON m.Match\_Id = b.Match\_Id

GROUP BY b.Bowler, m.Season\_Id, p.Player\_Name

)

SELECT

Bowler AS Player\_Id,

Player\_Name,

SUM(total\_matches) AS total\_match,

SUM(total\_wicket) AS total\_wicket,

ROUND(SUM(total\_ball) / SUM(total\_wicket), 2) AS bowling\_strike\_rate

FROM bolwer\_season\_wicket

GROUP BY Bowler, Player\_Name

HAVING COUNT(DISTINCT Season\_Id) >= 3

ORDER BY total\_wicket DESC ,total\_match DESC, bowling\_strike\_rate

LIMIT 10;

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Player\_Id | Player\_Name | total\_match | total\_wicket | bowling\_strick\_rate |
| 71 | DJ Bravo | 49 | 75 | 14.19 |
| 299 | B Kumar | 61 | 74 | 19.23 |
| 364 | MM Sharma | 61 | 70 | 18.41 |
| 50 | Harbhajan Singh | 62 | 65 | 21.77 |
| 315 | SP Narine | 51 | 61 | 20.28 |
| 194 | SL Malinga | 42 | 60 | 17.05 |
| 382 | YS Chahal | 42 | 56 | 16.93 |
| 367 | Sandeep Sharma | 43 | 54 | 18.02 |
| 357 | MG Johnson | 43 | 53 | 19.68 |
| 201 | R Ashwin | 61 | 51 | 24.53 |

*Table S\_3 – top 10 bowler with their stats*

* **DJ Bravo (Player\_Id: 71)**: Total Wickets: 75 (Highest among all players), Bowling Strike Rate: 14.19 (Best efficiency).

A standout performer who excels in wicket-taking and efficiency.

* **B Kumar (Player\_Id: 299)**: Total Wickets: 74, Bowling Strike Rate: 19.23

A reliable bowler with a strong performance.

* **MM Sharma (Player\_Id: 364)**: Total Wickets: 70, Bowling Strike Rate: 18.41

Highly consistent and effective.

* **SL Malinga (Player\_Id: 194)**: Total Wickets: 60, Bowling Strike Rate: 17.05

Known for his lethal yorkers and match-winning spells.

* **YS Chahal (Player\_Id: 382)**: Total Wickets: 56, Bowling Strike Rate: 16.93

Exceptional in the middle overs for taking crucial wickets.

* **Sandeep Sharma (Player\_Id: 367)**: Total Wickets: 54, Bowling Strike Rate: 18.02

Dependable in the powerplay overs.

* **SP Narine (Player\_Id: 315)**: Total Wickets: 61, Bowling Strike Rate: 20.28

Effective as a spinner in the middle overs.

* **R Ashwin (Player\_Id: 201)** and **Harbhajan Singh (Player\_Id: 50)**: While experienced, their bowling strike rates (24.53 and 21.77, respectively) are on the higher side, making them less efficient compared to the others.

**Conclusion:**

Based on the above analysis, the recommended batsmen and bowlers are perfectly suited to form a well-balanced and competitive team. Together, this combination of players offers the ideal blend of power, reliability, and match-winning potential for the team.

**Q3. What are some of the parameters that should be focused on while selecting the players?**

In the IPL, the contribution during the death overs plays a pivotal role in determining the outcome of matches. Therefore, while selecting players, I focused on the following key parameters:

1. **Batsman with High Impact:**  
   A batsman who can consistently score runs at a high strike rate, particularly during the death overs, is essential. Such players can accelerate the innings, finish strong, and put immense pressure on the opposition.
2. **Bowler with Efficiency and Precision:**  
   A bowler who can maintain a low economy rate while consistently taking wickets during the death overs is crucial. These bowlers not only prevent the opposition from scoring freely but also create opportunities to break partnerships and turn the game around.

* **Batsman with High Impact:**

WITH run\_group\_data AS (

SELECT

b.Striker,

p.Player\_Name,

m.Season\_Id,

m.Match\_Id,

SUM(b.Runs\_Scored) AS total\_runs,

SUM(CASE WHEN b.Over\_Id BETWEEN 15 AND 20 THEN b.Runs\_Scored ELSE 0 END) AS death\_overs\_runs,

COUNT(CASE WHEN b.Over\_Id BETWEEN 15 AND 20 THEN 1 END) AS total\_ball,

ROUND(100 \* SUM(CASE WHEN b.Over\_Id BETWEEN 15 AND 20 THEN b.Runs\_Scored ELSE 0 END) / COUNT(CASE WHEN b.Over\_Id BETWEEN 15 AND 20 THEN 1 END), 2) AS strike\_rate

FROM ball\_by\_ball b

JOIN matches m ON m.Match\_Id = b.Match\_Id

JOIN player p ON p.Player\_Id = b.Striker

WHERE NOT EXISTS (

SELECT 1

FROM extra\_runs e

WHERE b.Match\_Id = e.Match\_Id

AND b.Over\_Id = e.Over\_Id

AND b.Ball\_Id = e.Ball\_Id

AND e.Extra\_Type\_Id NOT IN(1,3)

AND b.Innings\_No = e.Innings\_No

)

GROUP BY m.Match\_Id, m.Season\_Id, b.Striker,p.Player\_Name

)

SELECT

Striker AS Player\_Id,

Player\_Name,

SUM(total\_runs) AS total\_runs,

ROUND(SUM(death\_overs\_runs), 2) AS avg\_death\_overs\_runs,

ROUND(AVG(strike\_rate), 2) AS strike\_rate

FROM run\_group\_data

GROUP BY Player\_Id, Player\_Name

HAVING SUM(total\_runs) > 100 AND sum(total\_ball) > 50

ORDER BY strike\_rate DESC;

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Player\_Id | Player\_Name | total\_runs | avg\_death\_overs\_runs | strike\_rate |
| 162 | CH Gayle | 1614 | 241 | 220.97 |
| 110 | AB de Villiers | 1955 | 808 | 200.7 |
| 187 | DA Warner | 2336 | 423 | 197.17 |
| 407 | JC Buttler | 254 | 144 | 195.65 |
| 413 | KH Pandya | 237 | 96 | 182.06 |
| 308 | SA Yadav | 497 | 275 | 179.88 |
| 19 | MEK Hussey | 1018 | 160 | 179.2 |
| 8 | V Kohli | 2462 | 605 | 174.24 |
| 57 | RG Sharma | 1891 | 718 | 166.7 |

*Table S\_4 – Batsmen with run and strike rate in death over*

* **Bowler with Efficiency and Precision:**

WITH economy\_grouped\_data AS (

SELECT

b.Bowler,

p.Player\_Name,

m.Season\_Id,

m.Match\_Id,

COUNT( DISTINCT b.Over\_Id) AS total\_over,

SUM(b.Runs\_Scored) AS total\_runs\_conceded,

SUM(b.Runs\_Scored) / COUNT( DISTINCT b.Over\_Id) as overall\_economy,

SUM(CASE WHEN b.Over\_Id BETWEEN 15 AND 20 THEN b.Runs\_Scored ELSE 0 END) AS runs\_conceded\_death\_overs,

SUM(CASE WHEN b.Over\_Id BETWEEN 15 AND 20 THEN b.Runs\_Scored ELSE 0 END) / COUNT( distinct CASE WHEN b.Over\_Id BETWEEN 15 AND 20 THEN b.Over\_Id END ) AS death\_over\_economy FROM ball\_by\_ball b

JOIN matches m ON m.Match\_Id = b.Match\_Id

JOIN player p ON p.Player\_Id = b.Bowler

GROUP BY b.Bowler, m.Season\_Id, m.Match\_Id, p.Player\_Name

)

SELECT

Bowler,

Player\_Name,

ROUND(AVG(total\_runs\_conceded), 2) AS avg\_total\_runs\_conceded,

ROUND(AVG(overall\_economy), 2) AS avg\_overall\_economy,

ROUND(AVG(runs\_conceded\_death\_overs), 2) AS avg\_runs\_conceded\_death\_overs,

ROUND(AVG(death\_over\_economy), 2) AS avg\_death\_over\_economy

FROM economy\_grouped\_data

GROUP BY Bowler, Player\_Name

HAVING SUM(total\_over) > 50

ORDER BY avg\_death\_over\_economy, avg\_overall\_economy;

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Bowler | Player\_Name | runs\_conceded\_death\_overs | avg\_total\_runs\_conceded | avg\_overall\_economy | avg\_runs\_conceded\_death\_overs | avg\_death\_over\_economy |
| 201 | R Ashwin | 232 | 21.57 | 7.19 | 3.8 | 6.49 |
| 67 | PP Chawla | 240 | 24.41 | 7.77 | 4.71 | 6.8 |
| 194 | SL Malinga | 547 | 26.71 | 6.74 | 13.02 | 7.19 |
| 109 | JA Morkel | 72 | 23.05 | 8.02 | 3.79 | 7.19 |
| 315 | SP Narine | 634 | 25.24 | 6.3 | 12.43 | 7.37 |
| 378 | MA Starc | 268 | 25.23 | 6.88 | 10.31 | 7.52 |
| 293 | S Nadeem | 113 | 25 | 7.03 | 3.9 | 7.53 |
| 94 | DW Steyn | 375 | 25.95 | 7.04 | 9.87 | 7.73 |
| 166 | KV Sharma | 125 | 23.96 | 7.87 | 2.72 | 7.77 |
| 429 | BB Sran | 114 | 28.47 | 8.26 | 7.6 | 7.83 |
| 460 | Mustafizur Rahman | 229 | 25.69 | 6.79 | 14.31 | 7.88 |
| 309 | KK Cooper | 283 | 29.26 | 7.74 | 14.89 | 8.06 |
| 255 | STR Binny | 56 | 11.35 | 7.13 | 1.3 | 8.2 |
| 9 | JH Kallis | 173 | 24.65 | 7.76 | 7.52 | 8.21 |

*Table S\_5 – List of economical bowler in death over*

WITH wicket\_grouped\_data AS (

SELECT

m.Match\_Id,

m.Season\_Id,

b.Bowler,

p.Player\_Name,

COUNT(\*) AS total\_wickets,

COALESCE(SUM(CASE WHEN b.Over\_Id BETWEEN 15 AND 20 THEN 1 END), 0) AS wickets\_death\_overs

FROM ball\_by\_ball b

JOIN matches m ON m.Match\_Id = b.Match\_Id

JOIN wicket\_taken w

ON w.Match\_Id = b.Match\_Id

AND w.Over\_Id = b.Over\_Id

AND w.Ball\_Id = b.Ball\_Id

AND w.Innings\_No = b.Innings\_No

JOIN player p ON p.Player\_Id = b.Bowler

GROUP BY m.Match\_Id, m.Season\_Id, b.Bowler, p.Player\_Name

)

SELECT

Bowler,

Player\_Name,

SUM(total\_wickets) AS total\_wicket,

ROUND(AVG(total\_wickets), 2) AS avg\_total\_wickets,

ROUND(AVG(wickets\_death\_overs), 2) AS avg\_wickets\_death\_overs

FROM wicket\_grouped\_data

GROUP BY Bowler, Player\_Name

HAVING SUM(total\_wickets) >= 50

ORDER BY avg\_wickets\_death\_overs DESC, avg\_total\_wickets DESC;

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Bowler | Player\_Name | total\_wicket | avg\_total\_wickets | avg\_wickets\_death\_overs |
| 71 | DJ Bravo | 81 | 1.93 | 1.48 |
| 232 | UT Yadav | 56 | 2 | 1.29 |
| 194 | SL Malinga | 66 | 1.83 | 1.28 |
| 310 | JP Faulkner | 61 | 1.91 | 1.22 |
| 315 | SP Narine | 68 | 1.84 | 1.14 |
| 32 | SR Watson | 54 | 1.74 | 1.13 |
| 73 | A Nehra | 57 | 1.97 | 1.07 |
| 81 | R Vinay Kumar | 50 | 1.72 | 1.03 |
| 299 | B Kumar | 77 | 1.79 | 1 |
| 364 | MM Sharma | 78 | 1.86 | 0.95 |
| 357 | MG Johnson | 55 | 1.77 | 0.74 |
| 35 | RA Jadeja | 56 | 1.93 | 0.62 |
| 136 | A Mishra | 55 | 1.72 | 0.56 |
| 367 | Sandeep Sharma | 59 | 1.74 | 0.53 |
| 201 | R Ashwin | 54 | 1.46 | 0.43 |
| 382 | YS Chahal | 58 | 1.71 | 0.41 |
| 50 | Harbhajan Singh | 67 | 1.6 | 0.14 |

*Table S\_6 – List of bowler with their death over wickets*

**Q4. Which players offer versatility in their skills and can contribute effectively with both bat and ball? (can you visualize the data for the same)**

WITH player\_wickets AS (

SELECT

b.Bowler AS Player\_Id,

COUNT(w.Match\_Id) AS total\_wicket,

ROUND(COUNT(b.Match\_Id) / COUNT(w.Match\_Id), 2) AS Bowling\_Strike\_Rate

FROM ball\_by\_ball b

LEFT JOIN wicket\_taken w

ON w.Match\_Id = b.Match\_Id

AND w.Over\_Id = b.Over\_Id

AND w.Ball\_Id = b.Ball\_Id

AND w.Innings\_No = b.Innings\_No

AND w.Kind\_Out != 3

GROUP BY b.Bowler

),

player\_runs AS (

SELECT

p.Player\_Name,

p.Player\_Id,

SUM(b.Runs\_Scored) AS total\_run,

ROUND(100 \* SUM(b.Runs\_Scored) / COUNT(b.Runs\_Scored),2) AS batting\_strike\_rate

FROM player p

JOIN ball\_by\_ball b ON b.Striker = p.Player\_Id

WHERE NOT EXISTS (

SELECT 1

FROM extra\_runs e

WHERE b.Match\_Id = e.Match\_Id

AND b.Over\_Id = e.Over\_Id

AND b.Ball\_Id = e.Ball\_Id

AND e.Extra\_Type\_Id NOT IN(1,3)

AND b.Innings\_No = e.Innings\_No

)

GROUP BY p.Player\_Id, p.Player\_Name

)

SELECT

r.Player\_Id,

r.Player\_Name,

r.total\_run,

w.total\_wicket,

r.batting\_strike\_rate,

w.Bowling\_Strike\_Rate

FROM player\_runs r

JOIN player\_wickets w ON w.Player\_Id = r.Player\_Id

WHERE r.total\_run >= (SELECT AVG(total\_run) FROM player\_runs )

AND w.total\_wicket >= (SELECT AVG(total\_wicket) FROM player\_wickets )

ORDER BY r.total\_run DESC, w.total\_wicket DESC

LIMIT 10;

*Chart S\_2 – Player’s batting & bowling stats*

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

Identify players whose presence has a measurable positive impact on their team’s performance, particularly by evaluating the team's win percentage when they are included in matches. The goal is to determine if certain players are correlated with higher win rates, suggesting their influence on team morale and success.

WITH player\_win\_percentage AS (

SELECT

p.Player\_Id,

p.Player\_Name,

pm.Team\_Id,

t.Team\_Name,

COUNT(m.Match\_Id) AS Total\_Matches,

SUM(CASE WHEN m.Match\_Winner = pm.Team\_Id THEN 1 ELSE 0 END) AS Matches\_Won,

ROUND( 100 \* SUM(CASE WHEN m.Match\_Winner = pm.Team\_Id THEN 1 ELSE 0 END) / COUNT(m.Match\_Id), 2) AS win\_percentage

FROM player p

JOIN player\_match pm ON p.Player\_Id = pm.Player\_Id

JOIN matches m ON pm.Match\_Id = m.Match\_Id

JOIN team t ON t.Team\_Id = pm.Team\_Id

WHERE m.Outcome\_type = 1

GROUP BY p.Player\_Id, p.Player\_Name, pm.Team\_Id, t.Team\_Name

HAVING count(m.Match\_Id) > 10

-- ORDER BY win\_percentage DESC

)

SELECT

Player\_Id,

Player\_Name,

SUM(Total\_Matches) AS Total\_Matches,

SUM(Matches\_Won) AS Matches\_Won,

ROUND(AVG(win\_percentage), 2) AS win\_percentage

FROM player\_win\_percentage

GROUP BY Player\_Id, Player\_Name

ORDER BY win\_percentage DESC;

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Player\_Id | Player\_Name | Total\_Matches | Matches\_Won | win\_percentage |
| 284 | RN ten Doeschate | 20 | 15 | 75 |
| 359 | IC Pandey | 23 | 17 | 73.91 |
| 228 | Bipul Sharma | 11 | 8 | 72.73 |
| 133 | SR Tendulkar | 14 | 10 | 71.43 |
| 398 | J Suchith | 14 | 10 | 71.43 |
| 429 | BB Sran | 14 | 10 | 71.43 |
| 383 | LMP Simmons | 22 | 15 | 68.18 |
| 19 | MEK Hussey | 21 | 14 | 66.67 |
| 23 | S Badrinath | 18 | 12 | 66.67 |
| 171 | M Manhas | 12 | 8 | 66.67 |
| 403 | MJ McClenaghan | 26 | 17 | 65.38 |
| 39 | SK Trivedi | 17 | 11 | 64.71 |
| 190 | M Morkel | 28 | 18 | 64.29 |
| 346 | JD Unadkat | 11 | 7 | 63.64 |
| 369 | Sachin Baby | 11 | 7 | 63.64 |
| 163 | MC Henriques | 30 | 19 | 63.33 |
| 96 | MK Pandey | 40 | 25 | 62.5 |
| 303 | F du Plessis | 32 | 20 | 62.5 |
| 460 | Mustafizur Rahman | 16 | 10 | 62.5 |

*Table S\_7 – List of players with their win percentage*

*Chart S\_3 – players with their win percentage*

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

Before RCB heads to the mega auction, the following strategic recommendations can guide them in making valuable selections:

1. **Focus on Multi-skilled Players**:

* Prioritize players who contribute both with bat and ball, as they add depth to the team and offer flexibility in squad balance. Visualizations, such as a scatter plot comparing players’ batting and bowling performances (e.g., runs scored vs. wickets taken), can highlight versatile players.

1. **Prioritize Power Hitters for Middle Overs**:

* RCB has traditionally relied heavily on top-order performance. They should target reliable power hitters for the middle order who can maintain a strong strike rate under pressure. A bar chart showing potential players' strike rates and average scores in overs 10-15 could help identify impactful players for this role.

1. **Strengthen Death Bowling**:

* Death bowling has been a weak spot for RCB in past seasons. They should aim to secure bowlers with high death-over wicket-taking ability and economical performance in the last 5 overs. Line graphs showing economy rates and wickets taken in death overs for target players can visually aid in choosing the right candidates.

1. **Seek Players with Proven Performance at Home Venues**:

* Players who perform well in the Chinnaswamy Stadium’s conditions (high-scoring pitch, smaller boundaries) will be an asset. Filtering players based on past performance at this venue and visualizing their averages and strike rates at similar high-scoring venues could help pinpoint ideal candidates.

1. **Consider Players with Strong Morale-boosting Presence**:

* Team dynamics are crucial for RCB's consistent performance. Identifying players whose presence has positively influenced past teams’ win percentages can improve team morale. Win percentage charts of players with their respective teams over multiple seasons would be beneficial for selecting such players.

1. **Balance Experience with Emerging Talent**:

* RCB could invest in emerging players with high potential while retaining a core of experienced players. A radar chart displaying experience factors (matches played, years of experience) alongside key performance metrics (average, economy, strike rate) could help achieve this balance.

These data-backed strategies would support RCB in creating a well-rounded, resilient team structure in the auction.

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

Several factors contribute to high-scoring matches in cricket, especially in T20 leagues like the IPL, which impact both viewership and team strategies. Here are key contributing elements and their effects:

1. **Pitch Conditions**

* **Flat Pitches**: Favor batsmen, leading to higher scores. Flat pitches, particularly in venues like Chinnaswamy Stadium in Bangalore, support stroke play due to low seam movement and predictable bounce.
* **Impact**: High scores make matches more exciting, attracting more viewers. Teams are more likely to strategize around batting depth and choose power hitters for flat-pitch venues.

2. **Smaller Boundaries**

* **Boundary Dimensions**: Smaller stadiums make it easier for batsmen to clear the boundary, resulting in more fours and sixes, which quickly increase the total score.
* **Impact**: This leads to more thrilling, high-scoring games, capturing viewers' attention. Teams often build squads with hard-hitters capable of exploiting small boundaries and target high boundary percentages in their game plans.

3. **Player Skill Levels and Batting Depth**

* **Strong Batting Line-ups**: Teams with power hitters and deep batting orders can post big totals, maintaining aggression from start to end.
* **Impact**: With stronger batting line-ups, teams may adopt aggressive strategies, aiming for high run rates. Viewers are drawn to this intensity, particularly in nail-biting finishes.

4. **Advances in Batting Techniques and Equipment**

* **Bat Technology and Shots Innovation**: Modern bats and techniques (like switch hits and scoops) allow batsmen to score runs faster.
* **Impact**: This adds excitement and unpredictability, boosting viewership. Teams may prioritize players skilled in innovative shots, especially in death overs.

5. **High Altitude and Weather Factors**

* **Altitude Effects**: In venues at higher altitudes, the ball travels farther, aiding batsmen in hitting boundaries more easily.
* **Weather Conditions**: Humid weather or dew in night games makes bowling and fielding tougher, often favouring the batting side.
* **Impact**: Teams adjust strategies, like choosing to bowl first in dewy conditions, to maintain control. This adaptability appeals to viewers, who enjoy seeing how teams tackle varying conditions.

6. **Shorter Game Format**

* **Nature of T20**: In T20, every ball count, pushing batsmen to be aggressive and increasing the overall scoring rate.
* **Impact**: This format’s fast-paced nature ensures constant excitement, making it ideal for high viewership. Teams may select players skilled in maintaining high strike rates and set aggressive powerplay and death-over strategies.

**Impact on Viewership and Team Strategies:**

* **Increased Viewership**: High-scoring matches with thrilling chases and frequent boundaries are highly appealing, keeping fans engaged for the duration of the match.
* **Enhanced Fan Engagement**: High scores and close chases result in more engaging, memorable moments, boosting fan engagement on social media and through in-game events.
* **Strategic Shift Towards Batting Depth**: Teams may focus on recruiting big-hitters and developing a long batting line-up to capitalize on favorable scoring conditions.
* **Dynamic Use of Analytics**: Teams use data to optimize batting strategies based on venue conditions and opponent weaknesses, targeting ideal line-ups and shot selections.

SELECT

m.Match\_Id,

SUM(b.Runs\_Scored) AS match\_total\_run

FROM ball\_by\_ball b

JOIN matches m ON b.Match\_Id = m.Match\_Id

JOIN venue v ON v.Venue\_Id = m.Venue\_Id

GROUP BY m.Match\_Id

HAVING SUM(b.Runs\_Scored) >= 350

ORDER BY match\_total\_run DESC;

*Chart S\_4 – Match with total runs in both innings*

**Q8. Analyze the impact of home-ground advantage on team performance and identify strategies to maximize this advantage for RCB.**

WITH filtered\_data AS (

SELECT

(

CASE

WHEN v.Venue\_Id = 5 THEN 1 WHEN v.Venue\_Id = 1 THEN 2

WHEN v.Venue\_Id = 8 THEN 3 WHEN v.Venue\_Id = 2 THEN 4

WHEN v.Venue\_Id = 6 THEN 5 WHEN v.Venue\_Id = 3 THEN 6

WHEN v.Venue\_Id = 4 THEN 7 WHEN v.Venue\_Id = 7 THEN 11

END ) AS Team\_Id ,

v.Venue\_Id,

v.Venue\_Name,

m.Match\_Winner

FROM venue v

JOIN matches m ON m.Venue\_Id = v.Venue\_Id

WHERE ((m.Team\_1 = 1 OR m.Team\_2 = 1 ) AND v.Venue\_Id = 5) OR ((m.Team\_1 = 2 OR m.Team\_2 = 2 ) AND v.Venue\_Id = 1)

OR ((m.Team\_1 = 3 OR m.Team\_2 = 3 ) AND v.Venue\_Id = 8) OR ((m.Team\_1 = 4 OR m.Team\_2 = 4 ) AND v.Venue\_Id = 2)

OR ((m.Team\_1 = 5 OR m.Team\_2 = 5 ) AND v.Venue\_Id = 6) OR ((m.Team\_1 = 6 OR m.Team\_2 = 6 ) AND v.Venue\_Id = 3)

OR ((m.Team\_1 = 7 OR m.Team\_2 = 7 ) AND v.Venue\_Id = 4) OR ((m.Team\_1 = 11 OR m.Team\_2 = 11 ) AND v.Venue\_Id = 7)

)

SELECT

Venue\_Id,

Venue\_Name,

f.Team\_Id,

t.Team\_Name,

count(Match\_Winner) as total\_match,

SUM( CASE WHEN f.Team\_Id = f.Match\_Winner THEN 1 ELSE 0 END) AS total\_win,

ROUND(100 \* SUM( CASE WHEN f.Team\_Id = f.Match\_Winner THEN 1 ELSE 0 END) / count(Match\_Winner), 2) AS win\_percentage

FROM filtered\_data f

JOIN team t ON t.Team\_Id = f.Team\_Id

GROUP BY Venue\_Id, Venue\_Name, f.Team\_Id, t.Team\_Name

ORDER BY win\_percentage DESC;

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Venue\_Id | Venue\_Name | Team\_Id | Team\_Name | total\_match | total\_win | win\_percentage |
| 6 | Sawai Mansingh Stadium | 5 | Rajasthan Royals | 8 | 8 | 100 |
| 8 | MA Chidambaram Stadium, Chepauk | 3 | Chennai Super Kings | 15 | 12 | 80 |
| 4 | Wankhede Stadium | 7 | Mumbai Indians | 25 | 19 | 76 |
| 5 | Eden Gardens | 1 | Kolkata Knight Riders | 23 | 17 | 73.91 |
| 7 | Rajiv Gandhi International Stadium, Uppal | 11 | Sunrisers Hyderabad | 23 | 14 | 60.87 |
| 1 | M Chinnaswamy Stadium | 2 | Royal Challengers Bangalore | 27 | 16 | 59.26 |
| 2 | Punjab Cricket Association Stadium, Mohali | 4 | Kings XI Punjab | 13 | 6 | 46.15 |
| 3 | Feroz Shah Kotla | 6 | Delhi Daredevils | 21 | 6 | 28.57 |

*Table S\_7 – Team home ground win analysis*

#### **Home-Ground Performance Analysis:**

* The data highlights that teams like Rajasthan Royals and Chennai Super Kings have leveraged their home-ground advantage exceptionally well, with win percentages of **100%** and **80%**, respectively.
* Mumbai Indians and Kolkata Knight Riders also demonstrate a strong home advantage with win percentages of **76%** and **73.91%**.
* Royal Challengers Bangalore (RCB), despite playing the most home matches (27) at M. Chinnaswamy Stadium, has a relatively moderate win percentage of **59.26%**, indicating room for improvement in capitalizing on their home advantage.

#### **Strategies for RCB to Maximize Home-Ground Advantage:**

1. **Leverage Chinnaswamy’s Batting-Friendly Conditions:**

* Develop a strong core of aggressive batsmen who can capitalize on the high-scoring nature of the Chinnaswamy pitch.
* Focus on power hitters and strike rotators to set or chase big totals.

1. **Enhance Death Bowling:**

* Invest in death-over specialists to address the high-scoring nature of the venue.
* Bowlers with variations like yorkers, slower balls, and cutters can be crucial in containing runs during the death overs.

1. **Optimize Field Placements:**

* Tailor fielding strategies to the shorter boundaries, ensuring fielders are positioned to cut off potential boundary opportunities.
* Use analytical insights to predict opposition batters' hitting zones.

1. **Utilize Spin Variations:**

* Although Chinnaswamy traditionally favors batsmen, spinners who can vary pace and turn could be more effective, especially in middle overs.

1. **Focus on Building Partnerships:**

* Encourage top-order batsmen to form steady partnerships to set a solid foundation and utilize the last overs effectively.

1. **Leverage Crowd Energy:**

* Engage fans and create an electrifying atmosphere that boosts player morale and puts pressure on the opposition.

**Q.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.**

**Season-Wise Analysis of RCB’s Batting Performance:**

WITH batting\_group\_data AS (

SELECT

m.Season\_Id,

m.Match\_Id,

SUM(b.Runs\_Scored) AS total\_runs,

SUM(CASE WHEN b.Over\_Id BETWEEN 1 AND 6 THEN b.Runs\_Scored ELSE 0 END) AS first\_powerplay\_runs,

SUM(CASE WHEN b.Over\_Id BETWEEN 7 AND 14 THEN b.Runs\_Scored ELSE 0 END) AS middle\_overs\_runs,

SUM(CASE WHEN b.Over\_Id BETWEEN 15 AND 20 THEN b.Runs\_Scored ELSE 0 END) AS death\_overs\_runs

FROM ball\_by\_ball b

JOIN matches m ON m.Match\_Id = b.Match\_Id

WHERE (m.Team\_1 = 2 OR m.Team\_2 = 2) AND b.Team\_Batting = 2

GROUP BY m.Season\_Id, m.Match\_Id

)

SELECT

Season\_Id,

ROUND(AVG(total\_runs), 2) AS avg\_total\_runs,

ROUND(AVG(first\_powerplay\_runs), 2) AS avg\_first\_powerplay\_runs,

ROUND(AVG(first\_powerplay\_runs) / 6, 2) AS avg\_first\_powerplay\_economy,

ROUND(AVG(middle\_overs\_runs), 2) AS avg\_middle\_overs\_runs,

ROUND(AVG(middle\_overs\_runs) / 8, 2) AS avg\_middle\_overs\_economy,

ROUND(AVG(death\_overs\_runs), 2) AS avg\_death\_overs\_runs,

ROUND(AVG(death\_overs\_runs) / 6, 2) AS avg\_death\_overs\_economy

FROM batting\_group\_data

GROUP BY Season\_Id;

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Season\_Id | avg\_total\_runs | avg\_first\_powerplay\_runs | avg\_first\_powerplay\_economy | avg\_middle\_overs\_runs | avg\_middle\_overs\_economy | avg\_death\_overs\_runs | avg\_death\_overs\_economy |
| 6 | 153.75 | 42.88 | 7.15 | 57.75 | 7.22 | 53.13 | 8.85 |
| 7 | 142.29 | 35.21 | 5.87 | 54.57 | 6.82 | 52.5 | 8.75 |
| 8 | 136.88 | 43 | 7.17 | 52.94 | 6.62 | 40.94 | 6.82 |
| 9 | 178.69 | 46.19 | 7.7 | 70 | 8.75 | 62.5 | 10.42 |

*Table S\_9 - Season-Wise Analysis of RCB’s Batting Performance*

*Chart S\_5 - Season-Wise Analysis of RCB’s Batting Performance*

* RCB has shown steady improvement in average powerplay runs over seasons, peaking at **46.19 runs** in Season 9.
* The first powerplay economy has remained around the **7.0** & in season 7 first powerplay economy is very bad **5.87**  performance.
* The death overs have consistently been a strong scoring phase, with Season 9 recording the highest average runs (**62.5**).
* However, the death-over economy increased to **10.42** in Season 9, highlighting potential areas for improvement in restricting runs during this phase.

**Bowling Performance Analysis of RCB (Season-Wise):**

WITH economy\_grouped\_data AS (

SELECT

m.Season\_Id,

m.Match\_Id,

SUM(b.Runs\_Scored) AS total\_runs\_conceded,

SUM(CASE WHEN b.Over\_Id BETWEEN 1 AND 6 THEN b.Runs\_Scored ELSE 0 END) AS runs\_conceded\_first\_powerplay,

SUM(CASE WHEN b.Over\_Id BETWEEN 7 AND 14 THEN b.Runs\_Scored ELSE 0 END) AS runs\_conceded\_middle\_overs,

SUM(CASE WHEN b.Over\_Id BETWEEN 15 AND 20 THEN b.Runs\_Scored ELSE 0 END) AS runs\_conceded\_death\_overs

FROM ball\_by\_ball b

JOIN matches m ON m.Match\_Id = b.Match\_Id

WHERE (m.Team\_1 = 2 OR m.Team\_2 = 2) AND b.Team\_Bowling = 2

GROUP BY m.Season\_Id, m.Match\_Id

)

SELECT

Season\_Id,

ROUND(AVG(total\_runs\_conceded), 2) AS avg\_total\_runs\_conceded,

ROUND(AVG(runs\_conceded\_first\_powerplay), 2) AS avg\_runs\_conceded\_first\_powerplay,

ROUND(AVG(runs\_conceded\_first\_powerplay) / 6, 2) AS first\_powerplay\_economy,

ROUND(AVG(runs\_conceded\_middle\_overs), 2) AS avg\_runs\_conceded\_middle\_overs,

ROUND(AVG(runs\_conceded\_middle\_overs) / 8, 2) AS middle\_overs\_economy,

ROUND(AVG(runs\_conceded\_death\_overs), 2) AS avg\_runs\_conceded\_death\_overs,

ROUND(AVG(runs\_conceded\_death\_overs) / 6, 2) AS death\_overs\_economy

FROM economy\_grouped\_data

GROUP BY Season\_Id;

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Season\_Id | avg\_total\_runs\_conceded | avg\_runs\_conceded\_first\_powerplay | first\_powerplay\_economy | avg\_runs\_conceded\_middle\_overs | middle\_overs\_economy | avg\_runs\_conceded\_death\_overs | death\_overs\_economy |
| 6 | 148.25 | 41.44 | 6.91 | 56.06 | 7.01 | 50.75 | 8.46 |
| 7 | 144.79 | 41.79 | 6.96 | 52.79 | 6.6 | 50.21 | 8.37 |
| 8 | 138 | 46.13 | 7.69 | 56.73 | 7.09 | 35.13 | 5.86 |
| 9 | 159.69 | 45.56 | 7.59 | 60.56 | 7.57 | 53.56 | 8.93 |

*Table S\_10 – Season wise bowling economy*

*Chart S\_6 – Season wise bowling economy*

* The average total runs conceded increased significantly in **Season 9 (159.69)** compared to the previous seasons.
* The best bowling performance was in **Season 8**, where the average runs conceded were lowest at **138**.
* RCB showed consistent performance in the powerplay in Seasons 6 and 7, conceding around **41 runs** with an economy rate below **7.0**.
* The middle overs economy has been steady across seasons, staying between **6.6** and **7.5**.
* RCB’s best death-over performance was in **Season 8**, with an economy rate of **5.86** and average runs conceded of **35.13**.

**Wicket Analysis of RCB by Season:**

WITH wicket\_grouped\_data AS (

SELECT

m.Match\_Id,

m.Season\_Id,

COUNT(\*) AS total\_wickets,

COALESCE(SUM(CASE WHEN b.Over\_Id BETWEEN 1 AND 6 THEN 1 END), 0) AS wickets\_first\_powerplay,

COALESCE(SUM(CASE WHEN b.Over\_Id BETWEEN 7 AND 14 THEN 1 END), 0) AS wickets\_middle\_overs,

COALESCE(SUM(CASE WHEN b.Over\_Id BETWEEN 15 AND 20 THEN 1 END), 0) AS wickets\_death\_overs

FROM ball\_by\_ball b

JOIN matches m ON m.Match\_Id = b.Match\_Id

JOIN wicket\_taken w

ON w.Match\_Id = b.Match\_Id

AND w.Over\_Id = b.Over\_Id

AND w.Ball\_Id = b.Ball\_Id

AND w.Innings\_No = b.Innings\_No

WHERE (m.Team\_1 = 2 OR m.Team\_2 = 2) AND b.Team\_Bowling = 2

GROUP BY m.Match\_Id, m.Season\_Id

)

SELECT

Season\_Id,

ROUND(AVG(total\_wickets), 2) AS avg\_total\_wickets,

ROUND(AVG(wickets\_first\_powerplay), 2) AS avg\_wickets\_first\_powerplay,

ROUND(AVG(wickets\_middle\_overs), 2) AS avg\_wickets\_middle\_overs,

ROUND(AVG(wickets\_death\_overs), 2) AS avg\_wickets\_death\_overs

FROM wicket\_grouped\_data

GROUP BY Season\_Id;

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Season\_Id | avg\_total\_wickets | avg\_wickets\_first\_powerplay | avg\_wickets\_middle\_overs | avg\_wickets\_death\_overs |
| 6 | 6.56 | 1.75 | 1.75 | 3.06 |
| 7 | 5.21 | 1.5 | 2.07 | 1.64 |
| 8 | 6.87 | 1.53 | 2.33 | 3 |
| 9 | 6.19 | 1.63 | 2 | 2.56 |

*Table S\_11 - Wicket Analysis of RCB by Season*

*Chart S\_7 – Wicket Analysis of RCB by Season*

* The average total wickets per season per match between **5**  to **6** is very low.
* In first powerplay avg wicket is around **1**  which is very low in all season**.**
* Death-over wickets often prevent opposition teams from accelerating and scoring heavily in the final few overs

**Suggestions for RCB:**

#### **1. Strengthen the Bowling Unit:**

* Invest in bowlers with strong **death-over skills** (e.g., Yorkers, slower balls, wide lines).
* Focus on recruiting **spinners or pacers** who can maintain control in the **middle overs** and pick crucial wickets to disrupt partnerships.
* Build a **bench of versatile bowlers** to rotate the lineup and avoid burnout during long tournaments.

#### **2. Address Middle Overs Batting:**

* Include a **dependable anchor batsman** in the middle order to stabilize innings and rotate the strike.
* Introduce versatile **all-rounders** who can accelerate if required but also ensure stability in the middle.

#### **3. Develop Strategies for High-Scoring Matches:**

* M. Chinnaswamy is a **batting-friendly ground**. RCB needs to adapt by focusing on:
* Bowling strategies to **contain opposition batsmen**.
* Strengthening their lower-middle order to post big totals consistently.

#### **4. Enhance Bench Strength:**

* Build a strong **backup for key players** to maintain consistency during injuries or dips in form.
* Focus on grooming young talent to develop a more balanced squad.

#### **5. Improve Team Balance:**

* Avoid over-reliance on a few individuals. A strong team effort is needed, with every department contributing equally.
* Recruit all-rounders who can provide depth in both batting and bowling.

**Conclusion:**

RCB’s struggles to win trophies can be attributed to **inconsistency in batting and bowling**, a lack of depth in the squad, and an over-reliance on star players. By focusing on **death-over bowling, middle-order batting stability, home-ground strategies, and bench strength**, RCB can significantly improve their performance and work towards lifting their first IPL trophy.

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

If I were to approach this problem independently, without being provided with specific objective or subjective questions, my strategy would involve a structured and data-driven approach as outlined below:

1. **Database Analysis:**

* Analyze historical data (player stats, match outcomes, auction results).
* Focus on past auction strategies: player performance vs. objectives.

1. **Performance Metrics:**

* **Toss Analysis**: Toss Win-Match Loss (TWML) & Toss Loss-Match Win (TLMW).
* **Batting**: Compare win rates when batting first (e.g., avg. total: 148.25, Season 6) vs. chasing.
* **Bowling**: Death overs economy (e.g., 8.46 in Season 6, 10.42 in Season 9) to identify weaknesses.

1. **Venue Statistics:**

* **Home vs. Away**: Win/loss margins, adaptability to different conditions.
* Key player performances at venues (e.g., Kohli: Avg 41.07 at M. Chinnaswamy, Strike Rate 149.81).

1. **Opponent Analysis:**

* Head-to-head win/loss records.
* Identify players who excel against rivals or underperform.

1. **Recommendations:**

* **Bowling**: Focus on death-over specialists to reduce economy rates (e.g., <8).
* **Batting**: Strengthen middle-order stability and finishing power.
* **Tactical Flexibility**: Reduce toss dependency and improve decision-making.
* **Leverage Key Metrics**: Build around strong venue performers like Kohli, ABD, and Gayle.

### ****Why RCB Hasn't Won the Trophy****

1. **Weak Death Bowling**: High economy rates (e.g., 10.42 in Season 9).
2. **Over-Reliance on Star Players**: Top-heavy batting with weak middle order.
3. **Inconsistent Decision-Making**: Over-dependence on toss outcomes.
4. **Pressure Handling**: Struggles in knockout stages.

Improving team balance, adaptability, and focusing on key metrics can transform RCB’s performance and lead to a trophy.

**Q. 11 In the "Match" table, some entries in the "Opponent\_Team" column are incorrectly spelled as "Delhi\_Capitals" instead of "Delhi\_Daredevils". Write an SQL query to replace all occurrences of "Delhi\_Capitals" with "Delhi\_Daredevils".**

Disabled the safe mode and then updated the Team Name

SET SQL\_SAFE\_UPDATES = 0;

UPDATE Team

SET Team\_Name = "Delhi Daredevils"

WHERE Team\_Name = "Delhi Capitals";