**OBJECTIVE QUESTIONS**

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

**Ans.** Every database table is made up of columns, and each column stores data of a specific data type — such as INT, VARCHAR, DATETIME, etc.

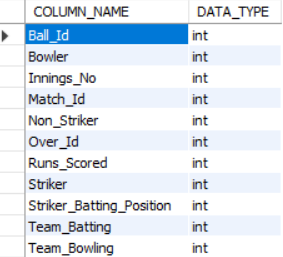
The INFORMATION\_SCHEMA is a system database in MySQL that stores metadata (information about databases, tables, columns, and constraints).  
It allows users to query details about the structure of any table without manually inspecting it.

SELECT COLUMN\_NAME, DATA\_TYPE

FROM INFORMATION\_SCHEMA.COLUMNS

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

AND TABLE\_SCHEMA = 'ipl';



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

Ans. We need to calculate:

* The **total runs scored by Royal Challengers Bangalore (RCB)**
* **During the first IPL season**
* Including both:
  + **Runs from the Ball\_by\_Ball table** (Runs\_Scored column)
  + **Extra runs** from the Extra\_Runs table (Extra\_R uns column)

To do this, we must **join** several tables:

* Matches → gives match info and season.
* Team → helps identify RCB by name.
* Ball\_by\_Ball → has actual runs per delivery.
* Extra\_Runs → has extra runs (like wides, no-balls, etc.).

**SELECT**

**t.Team\_Name AS Team,**

**s.Season\_Year,**

**SUM(b.Runs\_Scored + IFNULL(e.Extra\_Runs, 0)) AS Total\_Runs**

**FROM**

**Ball\_by\_Ball b**

**JOIN**

**Matches m ON b.Match\_Id = m.Match\_Id**

**JOIN**

**Team t ON b.Team\_Batting = t.Team\_Id**

**JOIN**

**Season s ON m.Season\_Id = s.Season\_Id**

**LEFT JOIN**

**Extra\_Runs e ON b.Match\_Id = e.Match\_Id**

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

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

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

**WHERE**

**t.Team\_Name LIKE '%Bangalore%' -- filters RCB**

**AND s.Season\_Year = 2008 -- 1st IPL season**

**GROUP BY**

**t.Team\_Name, s.Season\_Year;**

|  |
| --- |
| Team Season\_Year Total\_Runs |
| |  |  |  | | --- | --- | --- | | Royal Challengers Bangalore | 2013 | 2558 | |

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

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

**Ans.** To determine which players were older than **25 years during the 2014 IPL season**, we calculate each player’s age as of **1st January 2014**. This date is used as a standard reference point because the exact start date of IPL varies slightly each year, but players’ ages for a season are generally calculated at the beginning of that year.

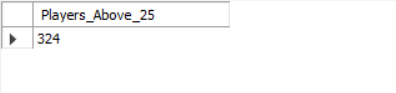
**SELECT COUNT(\*) AS Players\_Above\_25**

**FROM Player**

**WHERE TIMESTAMPDIFF(YEAR, DOB, '2014-04-01') > 25;**

This function returns the difference in years between a player's date of birth and the reference date 2014-01-01.  
We then select only those players whose computed age is greater than 25.

Thus, this query identifies all players who were older than 25 at the start of the 2014 season.

****

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

**Ans.** This query determines how many matches Royal Challengers Bangalore (RCB) won in the 2013 season.  
The Matches table records the winning team through the Match\_Winner column, which links to the Team table.  
By joining Matches, Team, and Season, we can identify which matches RCB won and filter for the 2013 season.  
The COUNT(\*) function then calculates the total number of wins for RCB in that year.

SELECT t.Team\_Name AS Team , s.Season\_Year,

COUNT(\*) AS Matches\_Won

FROM Matches m

JOIN

Team t ON m.Match\_Winner = t.Team\_Id

JOIN

Season s ON m.Season\_Id = s.Season\_Id

WHERE

t.Team\_Name LIKE '%Bangalore%'

AND s.Season\_Year = 2013

GROUP BY

t.Team\_Name, s.Season\_Year;

|  |
| --- |
| Team Season\_Year Matches\_Won |
| |  |  |  | | --- | --- | --- | | Royal Challengers Bangalore | 2013 | 9 | |

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

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

Ans. We need to calculate each player’s strike rate, which is a key batting metric.

Then, we’ll find the **top 10 players** with the highest strike rate across the **last 4 IPL seasons** (based on the most recent 4 seasons in the dataset).

**Tables Used:**

* Ball\_by\_Ball - gives runs scored per ball, striker (batsman).
* Matches - connects balls to seasons.
* Season - identifies the year.
* Player - to get player names.

SELECT p.Player\_Name , ROUND(SUM(b.Runs\_Scored) \* 100.0 / COUNT(b.Ball\_Id), 2) AS Strike\_Rate

FROM Ball\_by\_Ball b

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

JOIN Season s ON m.Season\_Id = s.Season\_Id

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

JOIN (

SELECT Season\_Year

FROM Season

ORDER BY Season\_Year DESC

LIMIT 4)

latest\_seasons ON s.Season\_Year = latest\_seasons.Season\_Year

GROUP BY p.Player\_Name

HAVING COUNT(b.Ball\_Id) > 20

ORDER BY Strike\_Rate DESC

LIMIT 10;

|  |
| --- |
| Player\_Name Strike\_Rate |
| |  |  | | --- | --- | | CR Brathwaite | 207.50 | | Bipul Sharma | 192.59 | | KH Pandya | 186.61 | | M Morkel | 173.68 | | SN Khan | 171.84 | | BCJ Cutting | 169.77 | | LJ Wright | 168.25 | | MS Gony | 166.67 | | AB de Villiers | 164.27 | | AD Russell | 163.16 | |

--------------------------------------------------------------------------------------------------------------------------

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

Ans. We need to calculate the average runs per batsman —  
that is, the total runs scored by each player divided by the number of matches they played (or innings batted).

This gives an insight into the player’s consistency across seasons.

**SELECT**

**p.Player\_Name,**

**ROUND(SUM(b.Runs\_Scored) / COUNT(DISTINCT b.Match\_Id), 2) AS Avg\_Runs\_Per\_Match**

**FROM**

**Ball\_by\_Ball b**

**JOIN**

**Player p ON b.Striker = p.Player\_Id**

**GROUP BY**

**p.Player\_Name**

**ORDER BY**

**Avg\_Runs\_Per\_Match DESC;**

|  |
| --- |
| Players\_Name Avg\_Runs\_per\_match |
| |  |  |  |  | | --- | --- | --- | --- | | LMP Simmons |  |  | 42.82 | | V Kohli |  |  | 39.87 | | DA Warner |  |  | 38.49 | | N Rana |  |  | 34.67 | | AB de Villiers |  |  | 34.53 | | MEK Hussey |  |  | 33.97 | | CH Gayle |  |  | 33.35 | | AM Rahane |  |  | 32.40 | | SE Marsh |  |  | 31.10 | | RV Uthappa |  |  | 30.87 | | DR Smith |  |  | 29.95 | | RG Sharma |  |  | 29.67 | | KH Pandya |  |  | 29.63 | | MP Stoinis |  |  | 29.20 | | JP Duminy |  |  | 29.00 | | SK Raina |  |  | 28.37 | | AJ Finch |  |  | 28.12 | | S Dhawan |  |  | 28.04 | | Q de Kock |  |  | 27.92 | | R Dravid |  |  | 27.71 | | F du Plessis |  |  | 26.94 | | BB McCullum |  |  | 26.69 | | SPD Smith |  |  | 26.64 | | G Gambhir |  |  | 26.59 | | HM Amla |  |  | 26.17 | | DA Miller |  |  | 25.59 | | CJ Anderson |  |  | 25.27 | | SR Watson |  |  | 24.81 | | MS Dhoni |  |  | 24.80 | | KP Pietersen |  |  | 24.47 | | KL Rahul |  |  | 24.17 | | BJ Rohrer |  |  | 24.13 | | YK Pathan |  |  | 24.02 | | M Vijay |  |  | 23.98 | | KK Nair |  |  | 23.70 | | SS Iyer |  |  | 23.45 | | DJ Hussey |  |  | 23.40 | | WP Saha |  |  | 23.20 | | S Narwal |  |  | 23.00 | | AD Russell |  |  | 22.91 | | GJ Maxwell |  |  | 22.83 | | M Vohra |  |  | 22.75 | | AC Gilchrist |  |  | 22.62 | | AT Rayudu |  |  | 22.56 | | KA Pollard |  |  | 22.37 | | V Sehwag |  |  | 22.34 | | KD Karthik |  |  | 22.22 | | MK Pandey |  |  | 22.13 | | DPMD Jayawardene |  |  | 22.07 | | EJG Morgan |  |  | 22.04 | | SW Billings |  |  | 22.00 | | Yuvraj Singh |  |  | 21.96 | | BJ Hodge |  |  | 21.89 | | CA Pujara |  |  | 21.70 | | SV Samson |  |  | 21.67 | | LJ Wright |  |  | 21.20 | | UT Khawaja |  |  | 21.17 | | JH Kallis |  |  | 21.00 | | YV Takawale |  |  | 20.80 | | CA Lynn |  |  | 20.75 | | PA Patel |  |  | 20.71 | | SS Tiwary |  |  | 20.42 | | SR Tendulkar |  |  | 20.21 | | RR Pant |  |  | 19.80 | | KS Williamson |  |  | 19.38 | | LA Pomersbach |  |  | 19.33 | | A Mukund |  |  | 19.00 | | MJ Guptill |  |  | 19.00 | | MK Tiwary |  |  | 18.68 | | MC Henriques |  |  | 18.37 | | Azhar Mahmood |  |  | 18.36 | | JC Buttler |  |  | 18.21 | | GJ Bailey |  |  | 18.18 | | RS Bopara |  |  | 18.13 | | Shakib Al Hasan |  |  | 18.05 | | TM Head |  |  | 18.00 | | MC Juneja |  |  | 17.86 | | S Badrinath |  |  | 17.71 | | Mandeep Singh |  |  | 17.64 | | BB Samantray |  |  | 17.57 | | Sachin Baby |  |  | 17.43 | | CL White |  |  | 17.38 | | MS Bisla |  |  | 17.35 | | UBT Chand |  |  | 16.38 | | MR Marsh |  |  | 16.33 | | MA Agarwal |  |  | 16.23 | | SN Khan |  |  | 16.09 | | JA Morkel |  |  | 16.08 | | B Chipli |  |  | 16.00 | | R Tewatia |  |  | 16.00 | | D Wiese |  |  | 15.88 | | AD Mathews |  |  | 15.80 | | NLTC Perera |  |  | 15.27 | | KM Jadhav |  |  | 15.20 | | TM Dilshan |  |  | 15.20 | | SA Yadav |  |  | 14.79 | | DJG Sammy |  |  | 14.75 | | BCJ Cutting |  |  | 14.60 | | VH Zol |  |  | 14.50 | | NV Ojha |  |  | 14.28 | | CH Morris |  |  | 14.25 | | Y Venugopal Rao |  |  | 14.20 | | JDS Neesham |  |  | 14.00 | | CR Brathwaite |  |  | 13.83 | | STR Binny |  |  | 13.83 | | A Ashish Reddy |  |  | 13.61 | | LRPL Taylor |  |  | 13.56 | | AP Tare |  |  | 13.55 | | RA Jadeja |  |  | 13.40 | | BR Dunk |  |  | 13.33 | | GH Vihari |  |  | 13.33 | | KC Sangakkara |  |  | 13.33 | | RN ten Doeschate |  |  | 13.20 | | AR Patel |  |  | 13.07 | | CM Gautam |  |  | 13.00 | | PA Reddy |  |  | 13.00 | | DH Yagnik |  |  | 12.90 | | Gurkeerat Singh |  |  | 12.80 | | DB Das |  |  | 12.50 | | OA Shah |  |  | 12.50 | | JP Faulkner |  |  | 12.41 | | DJ Hooda |  |  | 12.29 | | R Sathish |  |  | 12.09 | | ER Dwivedi |  |  | 12.00 | | P Negi |  |  | 11.53 | | NS Naik |  |  | 11.50 | | MS Gony |  |  | 11.25 | | DJ Bravo |  |  | 11.13 | | RE van der Merwe |  |  | 11.00 | | Iqbal Abdulla |  |  | 10.83 | | RR Rossouw |  |  | 10.60 | | B Lee |  |  | 10.50 | | S Rana |  |  | 10.43 | | Bipul Sharma |  |  | 10.40 | | RT Ponting |  |  | 10.40 | | TL Suman |  |  | 10.40 | | VR Aaron |  |  | 10.25 | | C Munro |  |  | 10.00 | | X Thalaivan Sargunam |  |  | 10.00 | | IK Pathan |  |  | 9.95 | | KV Sharma |  |  | 9.89 | | Harbhajan Singh |  |  | 9.81 | | KK Cooper |  |  | 9.75 | | J Suchith |  |  | 9.60 | | R Dhawan |  |  | 9.56 | | P Sahu |  |  | 9.50 | | HH Pandya |  |  | 9.18 | | KW Richardson |  |  | 9.00 | | Ankit Sharma |  |  | 8.80 | | MA Starc |  |  | 8.73 | | R Bhatia |  |  | 8.70 | | Ishan Kishan |  |  | 8.40 | | AM Nayar |  |  | 8.38 | | J Botha |  |  | 8.22 | | KB Arun Karthik |  |  | 8.20 | | R Vinay Kumar |  |  | 7.91 | | R Ashwin |  |  | 7.86 | | BMAJ Mendis |  |  | 7.67 | | JO Holder |  |  | 7.60 | | A Mishra |  |  | 7.50 | | SA Abbott |  |  | 7.50 | | DS Kulkarni |  |  | 7.33 | | M Morkel |  |  | 7.33 | | R Rampaul |  |  | 7.29 | | PP Chawla |  |  | 7.04 | | DB Ravi Teja |  |  | 7.00 | | MDKJ Perera |  |  | 7.00 | | SN Thakur |  |  | 7.00 | | UA Birla |  |  | 7.00 | | M Manhas |  |  | 6.86 | | LR Shukla |  |  | 6.80 | | MJ McClenaghan |  |  | 6.71 | | TG Southee |  |  | 6.71 | | AD Nath |  |  | 6.67 | | NJ Maddinson |  |  | 6.67 | | KC Cariappa |  |  | 6.50 | | DW Steyn |  |  | 6.40 | | AB Agarkar |  |  | 6.00 | | PC Valthaty |  |  | 6.00 | | PSP Handscomb |  |  | 6.00 | | UT Yadav |  |  | 6.00 | | MG Johnson |  |  | 5.46 | | B Kumar |  |  | 5.14 | | AD Mascarenhas |  |  | 5.00 | | AG Murtaza |  |  | 5.00 | | J Syed Mohammad |  |  | 5.00 | | M Rawat |  |  | 5.00 | | P Dogra |  |  | 5.00 | | P Kumar |  |  | 4.94 | | NM Coulter-Nile |  |  | 4.75 | | R Shukla |  |  | 4.75 | | MM Sharma |  |  | 4.69 | | F Behardien |  |  | 4.67 | | S Gopal |  |  | 4.67 | | S Kaul |  |  | 4.50 | | KJ Abbott |  |  | 4.33 | | Parvez Rasool |  |  | 4.25 | | A Chandila |  |  | 4.00 | | MN Samuels |  |  | 4.00 | | PJ Sangwan |  |  | 4.00 | | RV Gomez |  |  | 4.00 | | Imran Tahir |  |  | 3.60 | | HV Patel |  |  | 3.56 | | AN Ahmed |  |  | 3.50 | | GB Hogg |  |  | 3.50 | | WD Parnell |  |  | 3.50 | | Anureet Singh |  |  | 3.00 | | CJ Jordan |  |  | 3.00 | | DT Christian |  |  | 3.00 | | Harmeet Singh |  |  | 3.00 | | J Yadav |  |  | 3.00 | | Karanveer Singh |  |  | 3.00 | | PV Tambe |  |  | 3.00 | | Z Khan |  |  | 2.80 | | SP Narine |  |  | 2.71 | | L Balaji |  |  | 2.50 | | Shivam Sharma |  |  | 2.50 | | SMSM Senanayake |  |  | 2.50 | | JJ Bumrah |  |  | 2.25 | | S Nadeem |  |  | 2.18 | | SL Malinga |  |  | 2.13 | | Sandeep Sharma |  |  | 2.11 | | A Mithun |  |  | 2.00 | | AS Rajpoot |  |  | 2.00 | | B Laughlin |  |  | 2.00 | | BAW Mendis |  |  | 2.00 | | RG More |  |  | 2.00 | | R Sharma |  |  | 1.83 | | AB Dinda |  |  | 1.63 | | R McLaren |  |  | 1.50 | | S Anirudha |  |  | 1.50 | | SK Trivedi |  |  | 1.50 | | Mohammed Shami |  |  | 1.43 | | BB Sran |  |  | 1.33 | | AL Menaria |  |  | 1.00 | | DJ Muthuswami |  |  | 1.00 | | J Theron |  |  | 1.00 | | K Upadhyay |  |  | 1.00 | | M de Lange |  |  | 1.00 | | M Kartik |  |  | 1.00 | | PP Ojha |  |  | 1.00 | | RJ Harris |  |  | 1.00 | | SB Jakati |  |  | 1.00 | | SW Tait |  |  | 1.00 | | VS Malik |  |  | 1.00 | | YS Chahal |  |  | 1.00 | | JD Unadkat |  |  | 0.75 | | BE Hendricks |  |  | 0.50 | | S Sreesanth |  |  | 0.50 | | A Nehra |  |  | 0.33 | | P Awana |  |  | 0.25 | | A Zampa |  |  | 0.00 | | AB McDonald |  |  | 0.00 | | IC Pandey |  |  | 0.00 | | M Ashwin |  |  | 0.00 | | PJ Cummins |  |  | 0.00 | | S Aravind |  |  | 0.00 | | S Kaushik |  |  | 0.00 | | Swapnil Singh |  |  | 0.00 | |

This query calculates the average number of runs scored by each batsman across all IPL seasons.  
It aggregates total runs from the Ball\_by\_Ball table and divides them by the number of matches played.  
By grouping data by player and linking it with the Player table, it provides a clear view of batting consistency.  
The result highlights the players who have maintained the highest average performance throughout all seasons.

---------------------------------------------------------------------------------------------------------------------

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

Ans. We need to find the **average number of wickets per match** for every bowler across **all seasons**.

A “wicket” in the database is recorded in the Wicket\_Taken table, which identifies which **bowler** dismissed a **batsman** in each delivery.

So, to find averages:

**SELECT bowler.Player\_Name , ROUND(COUNT(w.Player\_Out) / COUNT(DISTINCT w.Match\_Id), 2) AS Avg\_Wickets\_Per\_Match**

**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**

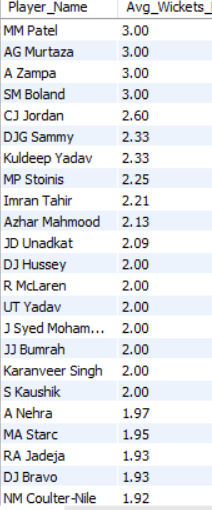
**JOIN Player bowler**

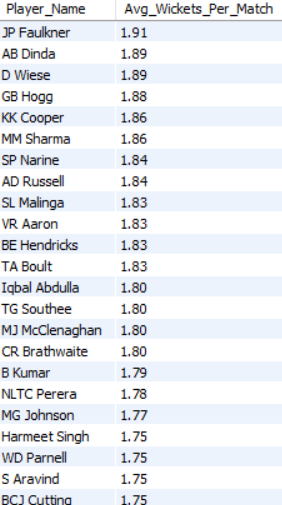
**ON b.Bowler = bowler.Player\_Id**

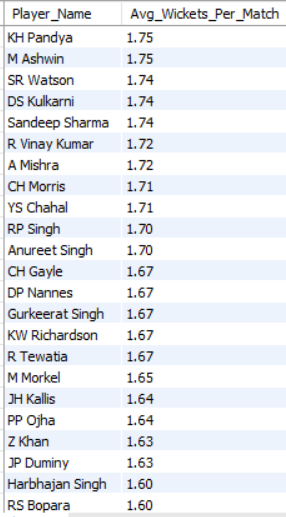
**WHERE w.Kind\_Out IS NOT NULL**

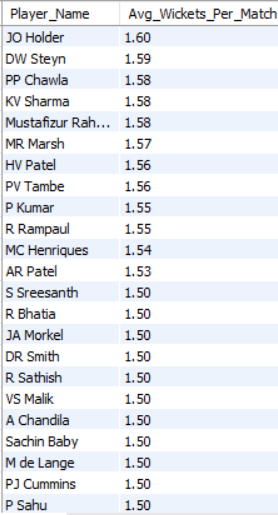
**GROUP BY bowler.Player\_Id, bowler.Player\_Name**

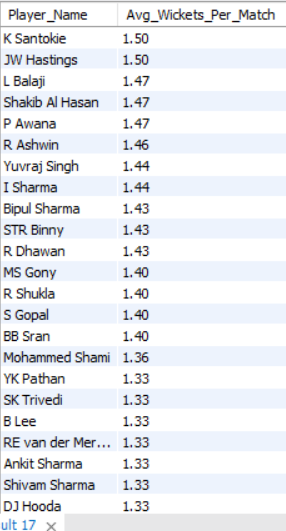
**ORDER BY Avg\_Wickets\_Per\_Match DESC;**

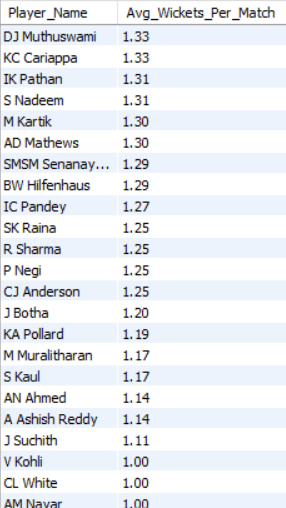


****

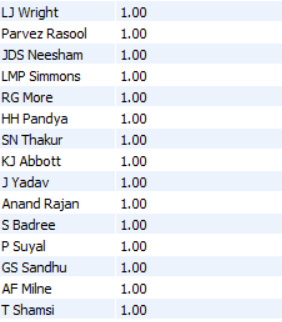
****

****

****

****

****

****

This query calculates the average wickets taken per match by each bowler across all IPL seasons.  
It joins the Wicket\_Taken and Ball\_by\_Ball tables to identify which bowler took each wicket.  
The total wickets are divided by the number of matches played to get the per-match average.  
The result ranks bowlers based on their overall wicket-taking performance.

--------------------------------------------------------------------------------------------------------------------------

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

Ans. We want to find **all-rounders** — players who:

1. Score **above-average runs** per match, **and**
2. Take **above-average wickets** per match.

So, we’ll calculate two metrics:

* **Average Runs per Match** (batting)
* **Average Wickets per Match** (bowling)

Then compare both to their **overall averages** across all players.

**WITH player\_runs AS (**

**SELECT**

**Striker AS Player\_Id,**

**ROUND(SUM(Runs\_Scored) / COUNT(DISTINCT Match\_Id), 2) AS Avg\_Runs\_Per\_Match**

**FROM**

**Ball\_by\_Ball**

**GROUP BY**

**Striker**

**),**

**player\_wickets AS (**

**SELECT**

**b.Bowler AS Player\_Id,**

**ROUND(COUNT(w.Player\_Out) / COUNT(DISTINCT w.Match\_Id), 2) AS Avg\_Wickets\_Per\_Match**

**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 IS NOT NULL**

**GROUP BY**

**b.Bowler**

**),**

**overall\_avg AS (**

**SELECT**

**(SELECT AVG(Avg\_Runs\_Per\_Match) FROM player\_runs) AS Overall\_Run\_Avg,**

**(SELECT AVG(Avg\_Wickets\_Per\_Match) FROM player\_wickets) AS Overall\_Wicket\_Avg**

**)**

**SELECT**

**p.Player\_Name,**

**r.Avg\_Runs\_Per\_Match,**

**w.Avg\_Wickets\_Per\_Match**

**FROM**

**player\_runs r**

**JOIN**

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

**JOIN**

**Player p ON p.Player\_Id = r.Player\_Id**

**JOIN**

**overall\_avg oa**

**WHERE**

**r.Avg\_Runs\_Per\_Match > oa.Overall\_Run\_Avg**

**AND w.Avg\_Wickets\_Per\_Match > oa.Overall\_Wicket\_Avg**

**ORDER BY**

**r.Avg\_Runs\_Per\_Match DESC, w.Avg\_Wickets\_Per\_Match DESC;**



This query identifies players who perform above average in both batting and bowling, effectively highlighting true all-rounders.  
It calculates each player’s average runs and wickets per match, then compares them with the overall averages across all players.  
Only those exceeding both benchmarks are selected.  
The result helps teams identify balanced players who contribute in both departments.

---------------------------------------------------------------------------------------------------------------------

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

Ans. We need to:

1. Find **how many matches RCB won and lost** at each venue.
2. Store those results in a **new table** named rcb\_record.

**CREATE TABLE rcb\_record (**

**Venue\_Name VARCHAR(255),**

**Matches\_Played INT,**

**Wins INT,**

**Losses INT**

**);**

**INSERT INTO rcb\_record (Venue\_Name, Matches\_Played, Wins, Losses)**

**SELECT**

**v.Venue\_Name,**

**COUNT(m.Match\_Id) AS Matches\_Played,**

**SUM(CASE WHEN m.Match\_Winner = rcb.Team\_Id THEN 1 ELSE 0 END) AS Wins,**

**SUM(CASE**

**WHEN (m.Team\_1 = rcb.Team\_Id OR m.Team\_2 = rcb.Team\_Id)**

**AND m.Match\_Winner <> rcb.Team\_Id THEN 1**

**ELSE 0**

**END) AS Losses**

**FROM**

**Matches m**

**JOIN**

**Venue v ON m.Venue\_Id = v.Venue\_Id**

**JOIN**

**Team rcb ON rcb.Team\_Name LIKE '%Bangalore%'**

**WHERE**

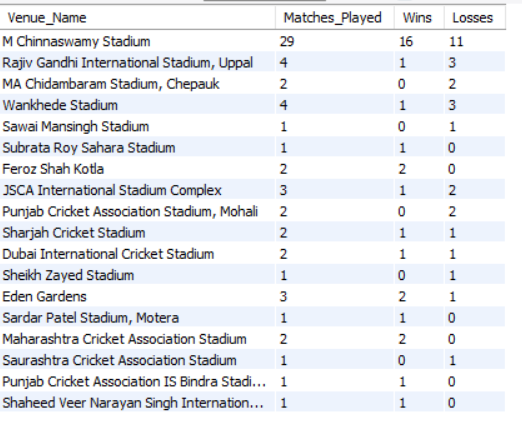
**(m.Team\_1 = rcb.Team\_Id OR m.Team\_2 = rcb.Team\_Id)**

**GROUP BY**

**v.Venue\_Name;**

After executing the data , Verified the data as well by using the formula;

**SELECT \* FROM rcb\_record;**

******

This query creates a summary table (rcb\_record) showing RCB’s performance at each venue.  
It counts total matches played, wins, and losses using conditional aggregation (CASE WHEN).  
By joining Matches, Venue, and Team, it accurately identifies RCB’s results across stadiums.  
This analysis helps evaluate how venue conditions affect RCB’s success rate.

---------------------------------------------------------------------------------------------------------------------

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

Ans. We need to find out how different bowling styles (e.g., fast, spin, medium, leg break) influence the number of wickets taken.

**SELECT**

**bs.Bowling\_skill AS Bowling\_Style,**

**COUNT(w.Player\_Out) AS Total\_Wickets**

**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 b.Bowler = p.Player\_Id**

**JOIN**

**Bowling\_Style bs**

**ON p.Bowling\_skill = bs.Bowling\_Id**

**WHERE**

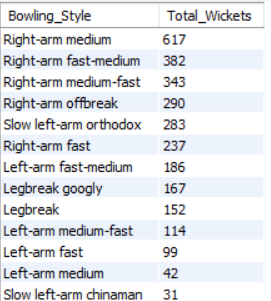
**w.Kind\_Out IS NOT NULL**

**GROUP BY**

**bs.Bowling\_skill**

**ORDER BY**

**Total\_Wickets DESC;**

****

This query analyzes how different bowling styles impact wicket-taking ability.  
By joining the Wicket\_Taken, Ball\_by\_Ball, Player, and Bowling\_Style tables, it links each dismissal to the bowler’s technique.  
The results show which bowling styles consistently produce more wickets.  
It helps teams understand which styles are most successful in IPL conditions.

---------------------------------------------------------------------------------------------------------------------

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

Ans. We want to **compare each team’s performance year-over-year (YOY)** —  
i.e., check if a team’s total **runs scored** and **wickets taken** in a season are **better than the previous season**.

If both increased, we’ll mark it as **‘Improved’**, otherwise **‘Declined’** or **‘Same’**.

**WITH team\_runs AS (**

**SELECT**

**s.Season\_Year,**

**b.Team\_Batting AS Team\_Id,**

**SUM(b.Runs\_Scored) AS Total\_Runs**

**FROM**

**Ball\_by\_Ball b**

**JOIN**

**Matches m ON b.Match\_Id = m.Match\_Id**

**JOIN**

**Season s ON m.Season\_Id = s.Season\_Id**

**GROUP BY**

**s.Season\_Year, b.Team\_Batting**

**),**

**team\_wickets AS (**

**SELECT**

**s.Season\_Year,**

**b.Team\_Bowling AS Team\_Id,**

**COUNT(w.Player\_Out) AS Total\_Wickets**

**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 b.Match\_Id = m.Match\_Id**

**JOIN**

**Season s ON m.Season\_Id = s.Season\_Id**

**WHERE**

**w.Kind\_Out IS NOT NULL**

**GROUP BY**

**s.Season\_Year, b.Team\_Bowling**

**),**

**team\_performance AS (**

**SELECT**

**r.Season\_Year,**

**t.Team\_Name,**

**r.Total\_Runs,**

**COALESCE(w.Total\_Wickets, 0) AS Total\_Wickets**

**FROM**

**team\_runs r**

**JOIN**

**Team t ON r.Team\_Id = t.Team\_Id**

**LEFT JOIN**

**team\_wickets w ON r.Team\_Id = w.Team\_Id AND r.Season\_Year = w.Season\_Year**

**)**

**SELECT**

**curr.Team\_Name,**

**curr.Season\_Year,**

**curr.Total\_Runs,**

**curr.Total\_Wickets,**

**prev.Total\_Runs AS Prev\_Runs,**

**prev.Total\_Wickets AS Prev\_Wickets,**

**CASE**

**WHEN curr.Total\_Runs > prev.Total\_Runs**

**AND curr.Total\_Wickets > prev.Total\_Wickets THEN 'Improved'**

**WHEN curr.Total\_Runs = prev.Total\_Runs**

**AND curr.Total\_Wickets = prev.Total\_Wickets THEN 'Same'**

**ELSE 'Declined'**

**END AS Performance\_Status**

**FROM**

**team\_performance curr**

**LEFT JOIN**

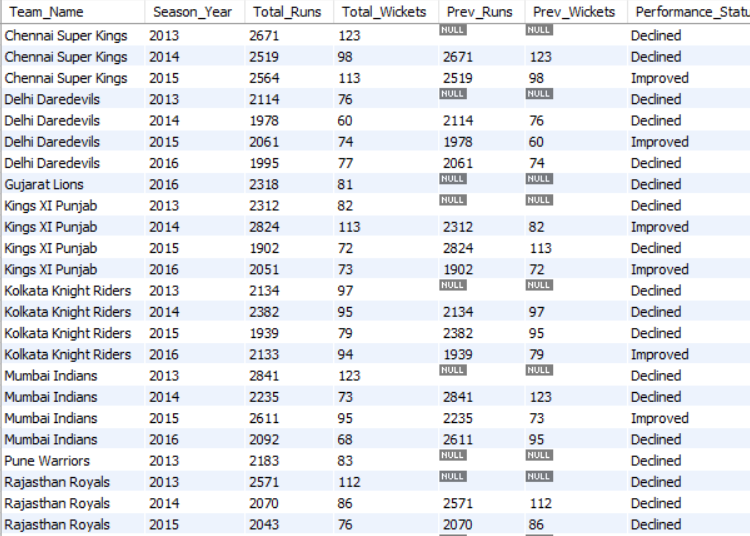
**team\_performance prev**

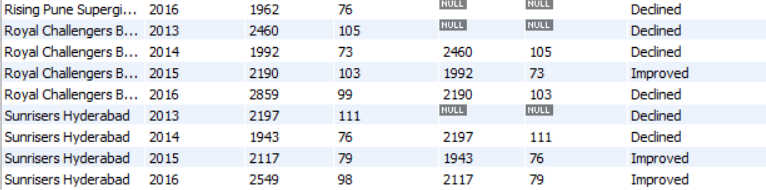
**ON curr.Team\_Name = prev.Team\_Name**

**AND curr.Season\_Year = prev.Season\_Year + 1**

**ORDER BY**

**curr.Team\_Name, curr.Season\_Year;**





This query evaluates whether each team’s performance improved compared to the previous IPL season.  
It calculates total runs scored and wickets taken by each team per year and compares them year-over-year.  
Teams with both higher runs and wickets are marked as “Improved.”  
This helps identify upward or downward trends in team performance across seasons.

----------------------------------------------------------------------------------------------------------------------

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

Ans

**KPI 1: Batting Strike Rate**

**Formula:**

**SELECT**

**p.Player\_Name,**

**ROUND(SUM(b.Runs\_Scored) \* 100.0 / COUNT(b.Ball\_Id), 2) AS Strike\_Rate**

**FROM**

**Ball\_by\_Ball b**

**JOIN**

**Player p ON b.Striker = p.Player\_Id**

**GROUP BY**

**p.Player\_Name**

**ORDER BY**

**Strike\_Rate DESC**

**LIMIT 10 ;**

By using this formula I found the players with the highest strike rate and highest strike means much more impactful Innings which gives leverage to other players as well so that they can have some additional time in the middle and unleash once settle down. Given below are the 10 names who have the highest strike rate in the IPL.

******

**KPI 2: Bowling Economy Rate**

**The Economy Rate measures how many runs a bowler concedes per over.  
It reflects how effective the bowler is at controlling the run flow.**

**Formula:**

**SELECT**

**p.Player\_Name,**

**ROUND(SUM(b.Runs\_Scored) / (COUNT(b.Ball\_Id) / 6), 2) AS Economy\_Rate**

**FROM**

**Ball\_by\_Ball b**

**JOIN**

**Player p ON b.Bowler = p.Player\_Id**

**GROUP BY**

**p.Player\_Name**

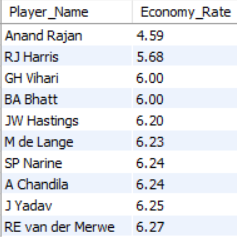
**HAVING**

**COUNT(b.Ball\_Id) > 30**

**ORDER BY**

**Economy\_Rate ASC**

**LIMIT 10;**



The Bowling Economy Rate measures how efficiently a bowler restricts scoring by showing runs conceded per over.  
Lower economy rates indicate better control and accuracy.  
This KPI helps identify bowlers who are effective in limiting opposition runs, especially in T20 cricket.  
The query aggregates total runs and overs bowled by each bowler to determine their overall economy.

**KPI 3: Win Percentage by Venue**

The **Win Percentage by Venue** measures how often a team wins at a specific stadium.  
It helps teams analyse which grounds favour their playing style (e.g., home advantage or pitch suitability.

**Formula:**

**SELECT**

**v.Venue\_Name,**

**COUNT(m.Match\_Id) AS Matches\_Played,**

**SUM(CASE WHEN m.Match\_Winner = t.Team\_Id THEN 1 ELSE 0 END) AS Wins,**

**ROUND(SUM(CASE WHEN m.Match\_Winner = t.Team\_Id THEN 1 ELSE 0 END) \* 100.0**

**/ COUNT(m.Match\_Id), 2) AS Win\_Percentage**

**FROM**

**Matches m**

**JOIN**

**Venue v ON m.Venue\_Id = v.Venue\_Id**

**JOIN**

**Team t ON t.Team\_Name LIKE '%Bangalore%'**

**WHERE**

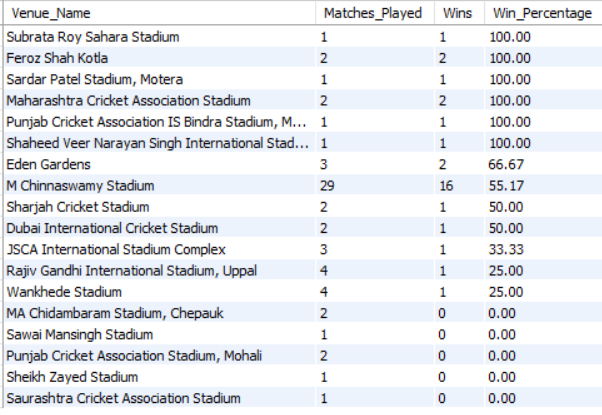
**(m.Team\_1 = t.Team\_Id OR m.Team\_2 = t.Team\_Id)**

**GROUP BY**

**v.Venue\_Name**

**ORDER BY**

**Win\_Percentage DESC;**



The Win Percentage by Venue KPI measures team success at different stadiums.  
It identifies favourable venues and home-ground advantages by calculating the ratio of wins to total matches played at each venue.  
This metric helps teams plan strategies, assess travel performance, and understand conditions that best support their strengths.

**KPI 4: Average Runs per Wicket (Batting Efficiency)**

The **Average Runs per Wicket** measures how efficiently a team or player converts runs before losing a wicket.  
It tells us **how strong and consistent the batting lineup** is across innings or seasons.

**Formula:**

**SELECT**

**t.Team\_Name,**

**ROUND(SUM(b.Runs\_Scored) / COUNT(w.Player\_Out), 2) AS Avg\_Runs\_Per\_Wicket**

**FROM**

**Ball\_by\_Ball b**

**JOIN**

**Matches m ON b.Match\_Id = m.Match\_Id**

**JOIN**

**Team t ON b.Team\_Batting = t.Team\_Id**

**LEFT JOIN**

**Wicket\_Taken w ON b.Match\_Id = w.Match\_Id**

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

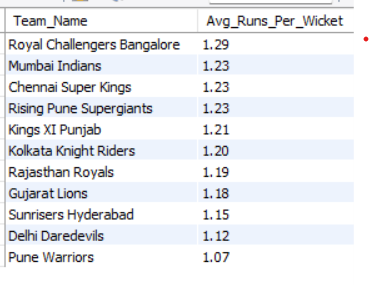
**AND w.Kind\_Out IS NOT NULL**

**GROUP BY**

**t.Team\_Name**

**ORDER BY**

**Avg\_Runs\_Per\_Wicket DESC;**

****

The Average Runs per Wicket KPI measures a team’s or player’s batting efficiency.  
It shows how many runs are scored on average before losing a wicket.  
Teams with higher averages demonstrate stronger batting depth and consistency.  
This metric helps evaluate stability and endurance of the batting order in T20 matches.

**KPI 5: Boundary Percentage (Aggression Index)**

The **Boundary Percentage** (also called the **Aggression Index**) measures **how much of a team’s or player’s total runs come from boundaries (4s and 6s)**.

It indicates **how attacking a batsman or a team’s batting strategy** is — crucial for T20 cricket.

Formula: Runs from 4s and 6s=(4×Total Fours)+(6×Total Sixes)

**SELECT**

**p.Player\_Name,**

**SUM(CASE WHEN b.Runs\_Scored = 4 THEN 1 ELSE 0 END) AS Fours,**

**SUM(CASE WHEN b.Runs\_Scored = 6 THEN 1 ELSE 0 END) AS Sixes,**

**SUM(b.Runs\_Scored) AS Total\_Runs,**

**ROUND(((SUM(CASE WHEN b.Runs\_Scored = 4 THEN 4 ELSE 0 END) +**

**SUM(CASE WHEN b.Runs\_Scored = 6 THEN 6 ELSE 0 END))**

**/ SUM(b.Runs\_Scored)) \* 100, 2) AS Boundary\_Percentage**

**FROM**

**Ball\_by\_Ball b**

**JOIN**

**Player p ON b.Striker = p.Player\_Id**

**GROUP BY**

**p.Player\_Name**

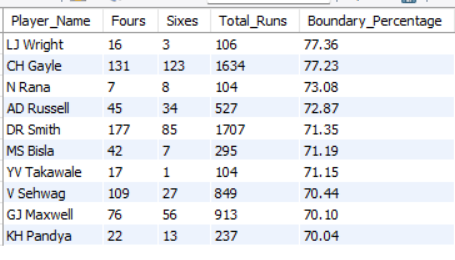
**HAVING**

**SUM(b.Runs\_Scored) > 100 -- filter to avoid players with very few runs**

**ORDER BY**

**Boundary\_Percentage DESC**

**LIMIT 10;**

******

The Boundary Percentage KPI measures a player’s or team’s attacking intent by calculating the share of runs scored through boundaries.  
A higher percentage indicates a more aggressive batting style and greater dependence on power hitting.  
This metric helps assess how teams balance aggression with consistency, especially in shorter T20 formats.  
Players like Chris Gayle and Andre Russell typically rank highest in boundary percentage, showcasing their explosive batting approach.

--------------------------------------------------------------------------------------------------------------------------

----------------------------------------------------------------------------------------------------------------------

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

Ans. We need the **average number of wickets** taken by **each bowler** in **each venue**.

Then, we have to **rank them** based on that average.

The mention of **“gender”** seems to be a typo or carry-over — since IPL data doesn’t have gender, we’ll assume it means **“rank the bowler”** by performance.

So, our goal is to Find which bowlers take the most wickets on average at different venues.

**SELECT**

**p.Player\_Name AS Bowler,**

**v.Venue\_Name,**

**ROUND(COUNT(w.Player\_Out) / COUNT(DISTINCT m.Match\_Id), 2) AS Avg\_Wickets\_Per\_Match,**

**DENSE\_RANK() OVER (PARTITION BY v.Venue\_Name**

**ORDER BY COUNT(w.Player\_Out) / COUNT(DISTINCT m.Match\_Id) DESC) AS Rank\_By\_Venue**

**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 b.Match\_Id = m.Match\_Id**

**JOIN**

**Venue v ON m.Venue\_Id = v.Venue\_Id**

**JOIN**

**Player p ON b.Bowler = p.Player\_Id**

**WHERE**

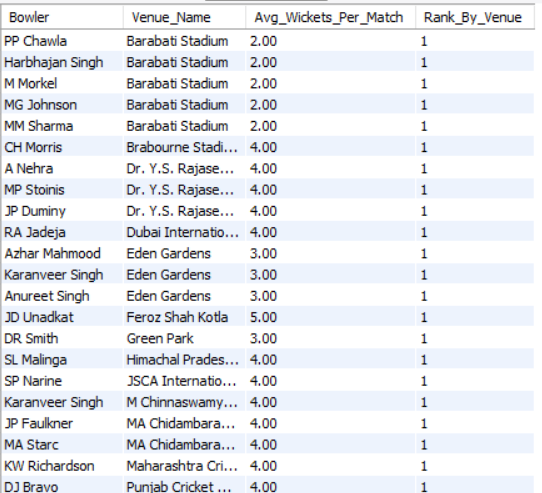
**w.Kind\_Out IS NOT NULL**

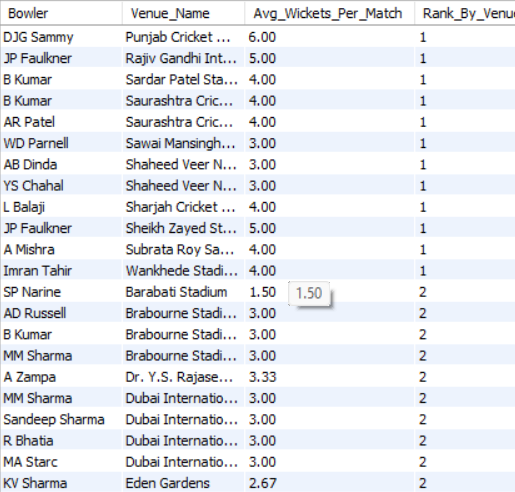
**GROUP BY**

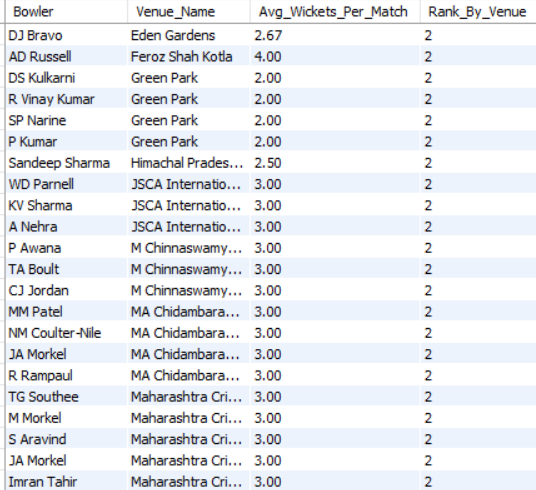
**p.Player\_Name, v.Venue\_Name**

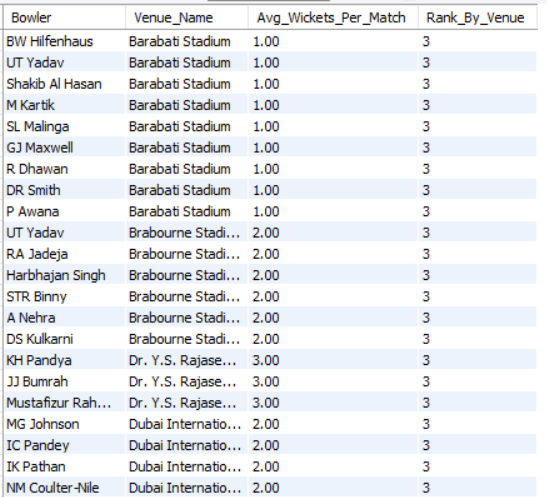
**ORDER BY**

**v.Venue\_Name, Avg\_Wickets\_Per\_Match DESC;**

****

****

****



This query evaluates bowler performance across different venues by calculating their average wickets per match.  
By grouping results by venue, it highlights how effective each bowler is under specific ground conditions.  
The RANK() function ranks bowlers by average wickets, identifying venue specialists.  
This KPI helps teams select bowlers suited for particular stadiums or pitch types.

--------------------------------------------------------------------------------------------------------------------------

1. Which of the given players have consistently performed well in past seasons?

Ans. We need to identify **the 5 most effective bowlers overall** — across **all venues**, not just one.  
We’ll use **average wickets per match** as the performance metric (same logic as before but across total matches).

**Formula:**

**SELECT**

**p.Player\_Name AS Bowler,**

**ROUND(COUNT(w.Player\_Out) / COUNT(DISTINCT m.Match\_Id), 2) AS Avg\_Wickets\_Per\_Match,**

**DENSE\_RANK() OVER (ORDER BY COUNT(w.Player\_Out) / COUNT(DISTINCT m.Match\_Id) DESC) AS Overall\_Rank**

**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 b.Match\_Id = m.Match\_Id**

**JOIN**

**Player p ON b.Bowler = p.Player\_Id**

**WHERE**

**w.Kind\_Out IS NOT NULL**

**GROUP BY**

**p.Player\_Name**

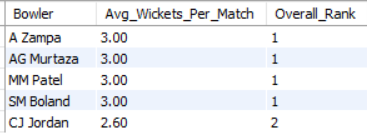
**HAVING**

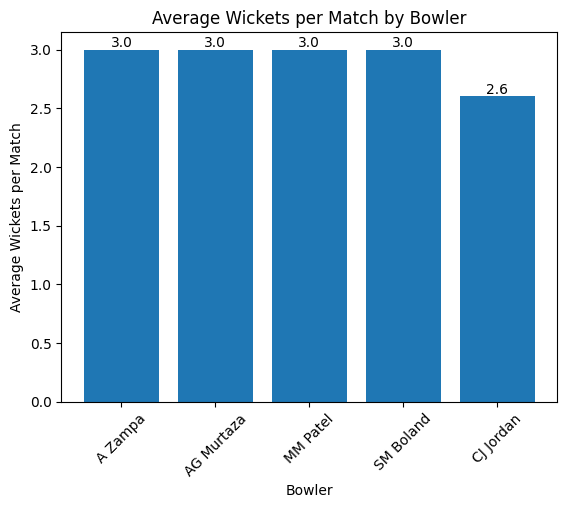
**COUNT(DISTINCT m.Match\_Id)**

**ORDER BY**

**Avg\_Wickets\_Per\_Match DESC**

**LIMIT 5;**

******

**

----------------------------------------------------------------------------------------------------------------------

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

Ans. We want to find **venue specialists** — players (especially bowlers or batsmen) who consistently perform **better at certain venues** than others.

**SELECT**

**p.Player\_Name,**

**v.Venue\_Name,**

**ROUND(SUM(b.Runs\_Scored) / COUNT(DISTINCT m.Match\_Id), 2) AS Avg\_Runs\_Per\_Match**

**FROM ball\_by\_ball b**

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

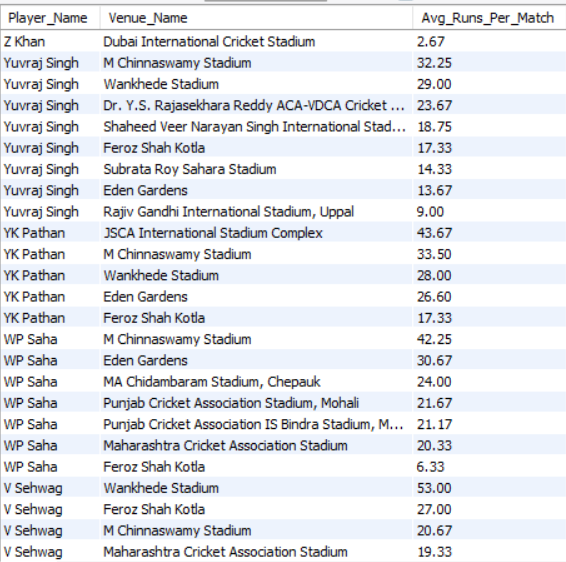
**JOIN venue v ON m.Venue\_Id = v.Venue\_Id**

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

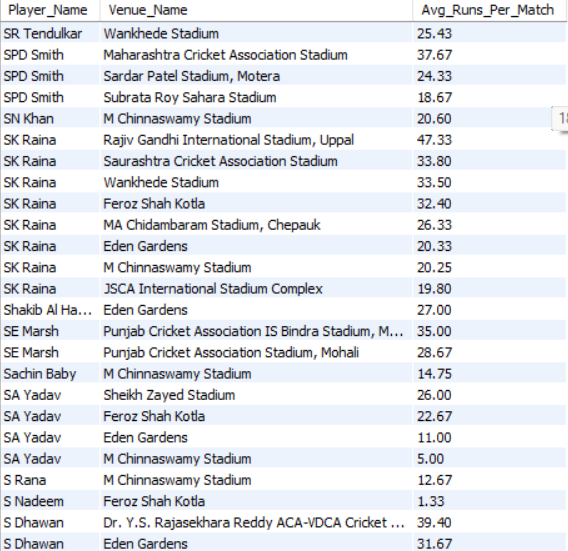
**GROUP BY p.Player\_Name, v.Venue\_Name**

**HAVING COUNT (DISTINCT m.Match\_Id) >= 3 -- minimum matches to avoid bias**

**ORDER BY Avg\_Runs\_Per\_Match DESC;**

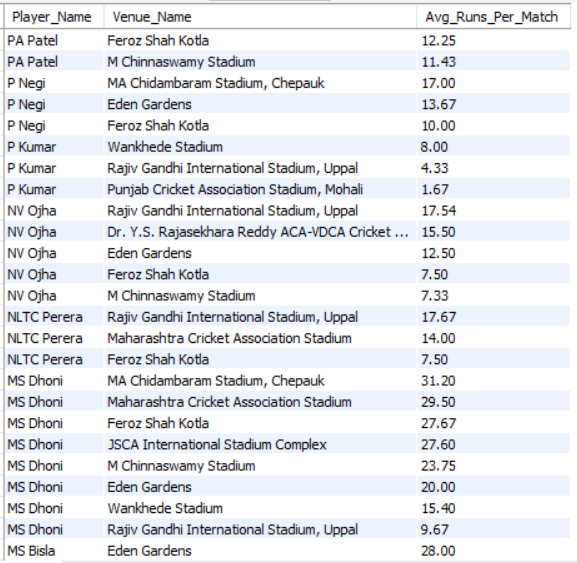




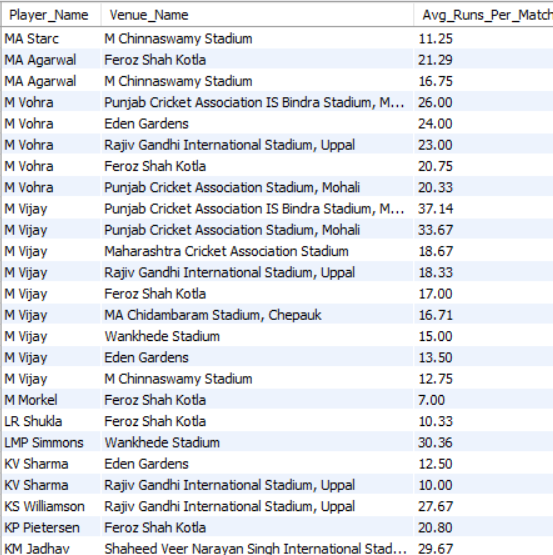


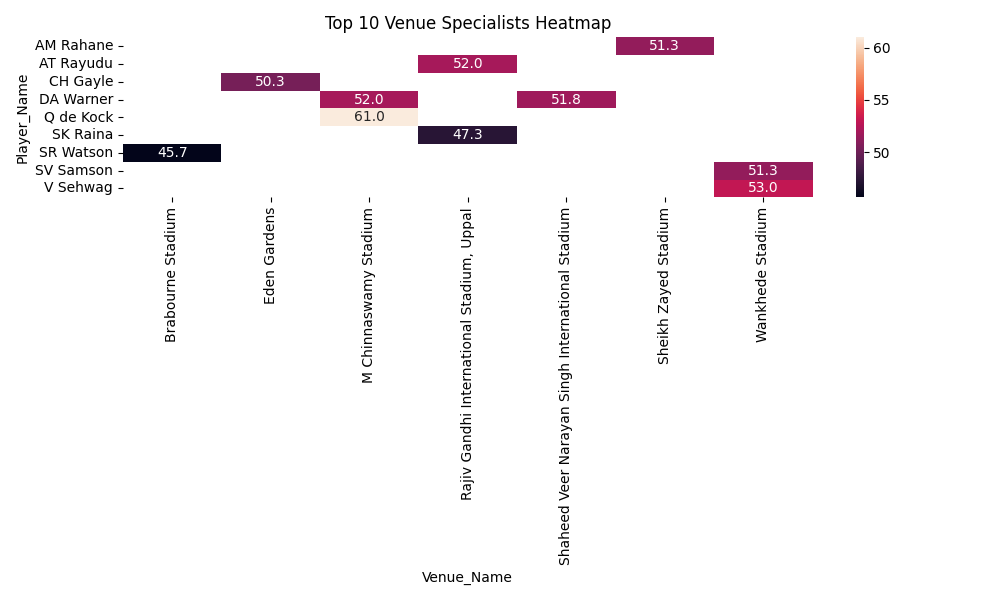












This heatmap visualises:

* The top 10 highest-performing players at different venues
* How strongly a player performs at that specific venue
* Darker colours = higher average runs
* Clear identification of "venue specialists"

**Insights**

* Q de Kock performs exceptionally well at M Chinnaswamy Stadium (61 runs avg).
* Warner is strong at multiple venues, especially Shaheed Veer Narayan Singh Stadium (~52).
* V Sehwag has a venue preference for Wankhede Stadium (53).
* SV Samson also excels at Wankhede (~51.3).
* CH Gayle is a strong performer at Eden Gardens (~50.3).

A heatmap was created using the average runs scored by the top-performing players at different venues. The data was pivoted so that players appear on the Y-axis and venues on the X-axis. Conditional colour gradients highlight performance intensity, where darker colours represent stronger averages. This visualization helps identify venue-specialist players whose performance is significantly better at certain grounds.

--------------------------------------------------------------------------------------------------------------------------

SUBJECTIVE QUESTIONS

**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?**

Ans. **Impact of Toss Decision on Match Result**

Based on the IPL data analysis, teams that win the toss and choose to field first have a noticeably higher win percentage.

* Field first → 54.84% win rate
* Bat first → 43.00% win rate

This shows that **fielding first provides a strategic advantage** in many IPL matches.  
Common reasons include dew factor, chasing-friendly conditions, and more predictable targets.

**Is this impact consistent across all venues?**

The toss advantage is **not uniform across venues**.  
Some stadiums are known to favor chasing due to:

* Shorter boundaries
* Better batting conditions under lights
* Dew making bowling difficult in second innings

**SELECT**

**t.Toss\_Name AS Toss\_Decision,**

**COUNT(\*) AS Matches\_Played,**

**SUM(CASE WHEN m.Match\_Winner = m.Toss\_Winner THEN 1 ELSE 0 END) AS Wins\_After\_Toss,**

**ROUND(**

**SUM(CASE WHEN m.Match\_Winner = m.Toss\_Winner THEN 1 ELSE 0 END)**

**/ COUNT(\*) \* 100,**

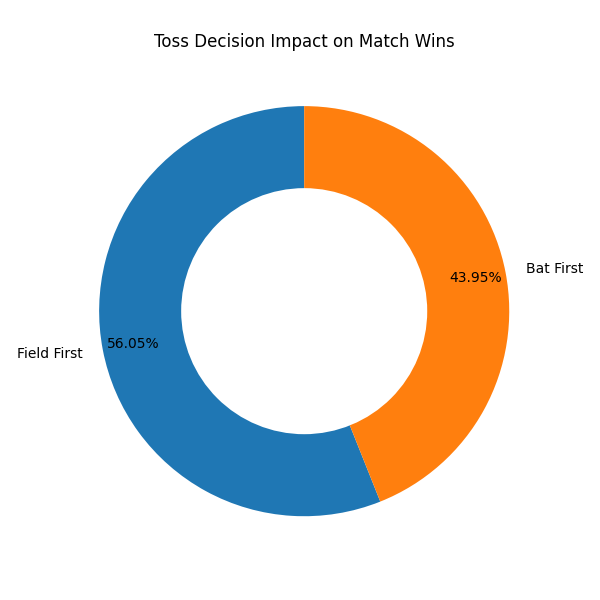
**2) AS Win\_Percentage**

**FROM matches m**

**JOIN toss\_decision t**

**ON m.Toss\_Decide = t.Toss\_Id**

**GROUP BY t.Toss\_Name;**

****

The doughnut chart displays the win percentage distribution based on toss decisions.  
Teams choosing to field first after winning the toss achieve a higher proportion of wins (54.84%) compared to teams choosing to bat first (43.00%).  
The labeled percentages provide a clear visual comparison of the impact of toss decisions on match outcomes.

To explore this, we can extend the analysis using:

Venue-wise toss impact query

**SELECT**

**v.Venue\_Name,**

**t.Toss\_Name AS Toss\_Decision,**

**COUNT(\*) AS Matches\_Played,**

**ROUND(**

**SUM(CASE WHEN m.Match\_Winner = m.Toss\_Winner THEN 1 ELSE 0 END) / COUNT(\*) \* 100,**

**2) AS Win\_Percentage**

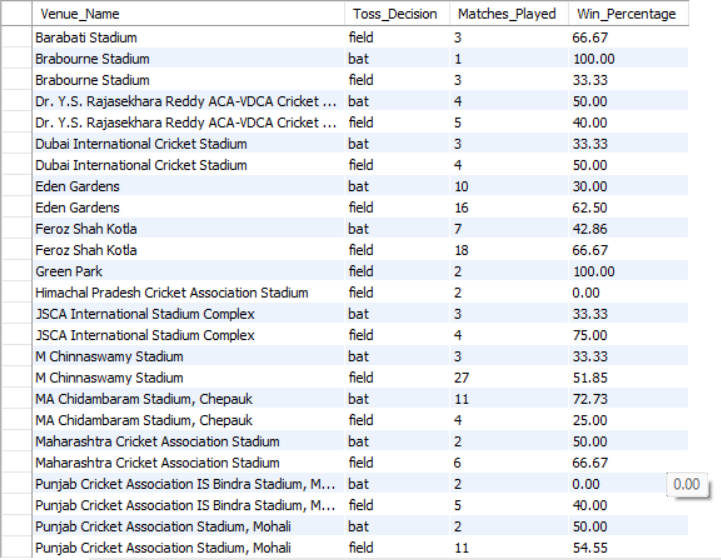
**FROM matches m**

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

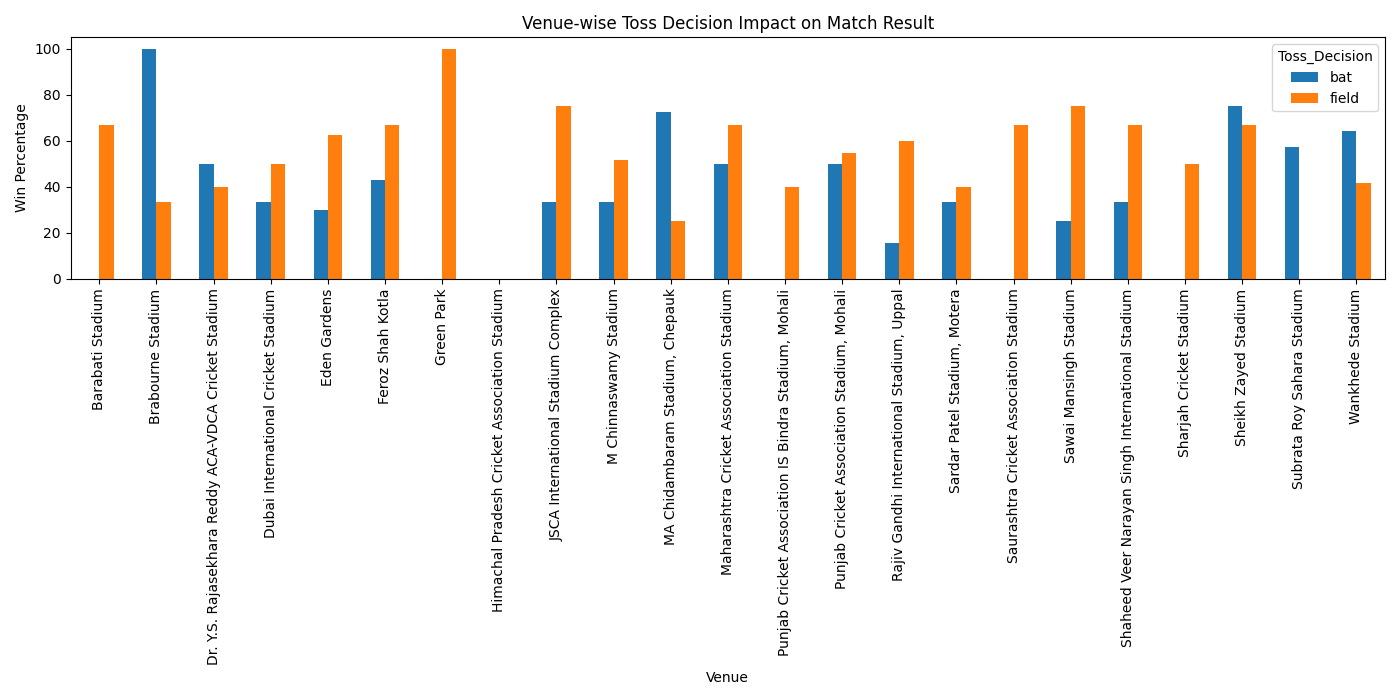
**JOIN venue v ON m.Venue\_Id = v.Venue\_Id**

**GROUP BY v.Venue\_Name, t.Toss\_Name**

**ORDER BY v.Venue\_Name;**







The bar chart shows the venue-wise impact of toss decisions on match outcomes.  
At venues such as Eden Gardens, Feroz Shah Kotla, Rajiv Gandhi International Stadium, and Sawai Mansingh Stadium, teams choosing to field first after winning the toss have a significantly higher win percentage.  
In contrast, venues like MA Chidambaram Stadium (Chepauk) and Wankhede Stadium show better results for teams choosing to bat first.  
This indicates that the impact of the toss is not uniform across venues and is influenced by pitch conditions, ground dimensions, and match environment.

--------------------------------------------------------------------------------------------------------------------------

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

**Ans.** To suggest players best suited for a team, multiple performance indicators from the IPL dataset were considered, such as:

* Consistent batting performance (average runs per match)
* Venue-specific dominance
* Match impact and reliability
* All-round contributions
* Experience and adaptability

**SELECT**

**p.Player\_Name,**

**ROUND(SUM(b.Runs\_Scored) / COUNT(DISTINCT b.Match\_Id), 2) AS Avg\_Runs\_Per\_Match**

**FROM Ball\_by\_Ball b**

**JOIN Player p**

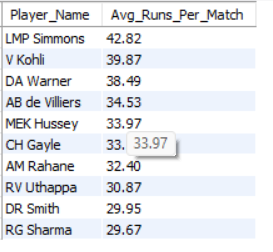
**ON b.Striker = p.Player\_Id**

**GROUP BY p.Player\_Name**

**HAVING COUNT(DISTINCT b.Match\_Id) > 20**

**ORDER BY Avg\_Runs\_Per\_Match DESC**

**LIMIT 10;**



**SELECT**

**p.Player\_Name,**

**ROUND(SUM(b.Runs\_Scored) / COUNT(DISTINCT b.Match\_Id), 2) AS Avg\_Runs,**

**ROUND(COUNT(w.Player\_Out) / COUNT(DISTINCT w.Match\_Id), 2) AS Avg\_Wickets**

**FROM Player p**

**LEFT JOIN Ball\_by\_Ball b**

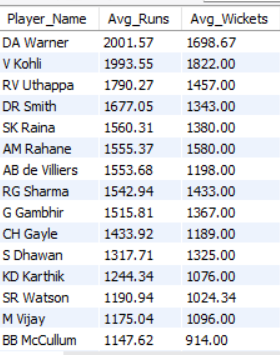
**ON p.Player\_Id = b.Striker**

**LEFT JOIN Wicket\_Taken w**

**ON p.Player\_Id = w.Player\_Out**

**GROUP BY p.Player\_Name**

**ORDER BY Avg\_Runs DESC, Avg\_Wickets DESC;**

****

Based on the analysis, the following players are recommended:

**1. David Warner**

* One of the most consistent batsmen across multiple seasons.
* Performs exceptionally well at venues like **M Chinnaswamy Stadium** and **Rajiv Gandhi International Stadium**.
* High average runs per match indicate reliability as a **top-order batsman**.
* Ideal for teams looking for a strong opener who can anchor or accelerate innings.

**2. Virat Kohli**

* Shows consistent performance across multiple venues including **Eden Gardens**, **M Chinnaswamy Stadium**, and **Wankhede Stadium**.
* Known for chasing targets effectively, making him suitable for teams preferring to **field first**.
* His adaptability to different pitch conditions makes him a key middle-order batsman.

**3. AB de Villiers**

* Exceptional venue-specific performances, especially at **M Chinnaswamy Stadium**.
* High strike rate combined with strong average makes him a match-winner.
* Best suited for teams requiring a **finisher or middle-order stabilizer**.

**4. Chris Gayle**

* Dominant at batting-friendly venues like **Eden Gardens** and **M Chinnaswamy Stadium**.
* Capable of changing match outcomes single-handedly.
* Ideal for teams aiming for aggressive powerplay strategies.

**5. Rohit Sharma**

* Performs consistently at venues such as **Wankhede Stadium** and **Eden Gardens**.
* Provides stability at the top order and contributes with leadership.
* Suitable for teams seeking both experience and consistent scoring.

**6. MS Dhoni**

* Reliable lower-middle order batsman and excellent finisher.
* Strong performance at **MA Chidambaram Stadium (Chepauk)**.
* Adds value through leadership, decision-making, and wicket-keeping skills.

**7. All-rounder Options (Strategic Picks)**

* **Ravindra Jadeja** – Effective at spin-friendly venues, contributes in batting, bowling, and fielding.
* **Shane Watson** – Proven performer across venues, useful as an all-rounder in both departments.

**Conclusion**

The suggested players combine consistency, adaptability, and match-winning ability.  
Teams should select players based on:

* Home venue conditions
* Team balance (batting depth vs bowling strength)
* Match strategy (chasing vs defending)

Using data-driven analysis ensures smarter team selection and maximizes the chances of success.

**The above recommendations are based on IPL historical performance data analysed using SQL queries.**

--------------------------------------------------------------------------------------------------------------------------

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

**Ans**. Player selection should be based on a combination of **performance metrics, match conditions, and team requirements** rather than reputation alone. The following parameters are important when selecting players for a team:

**SELECT**

**p.Player\_Name,**

**ROUND(SUM(b.Runs\_Scored) / COUNT(DISTINCT b.Match\_Id), 2) AS Avg\_Runs\_Per\_Match**

**FROM Ball\_by\_Ball b**

**JOIN Player p**

**ON b.Striker = p.Player\_Id**

**GROUP BY p.Player\_Name**

**HAVING COUNT(DISTINCT b.Match\_Id) > 20**

**ORDER BY Avg\_Runs\_Per\_Match DESC;**

**SELECT**

**p.Player\_Name,**

**ROUND(SUM(b.Runs\_Scored) / COUNT(DISTINCT b.Match\_Id), 2) AS Avg\_Runs\_Per\_Match**

**FROM Ball\_by\_Ball b**

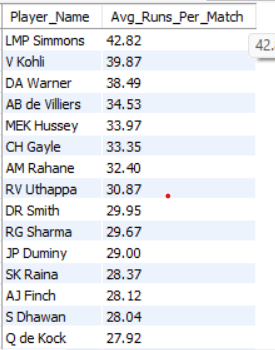
**JOIN Player p**

**ON b.Striker = p.Player\_Id**

**GROUP BY p.Player\_Name**

**HAVING COUNT(DISTINCT b.Match\_Id) > 20**

**ORDER BY Avg\_Runs\_Per\_Match DESC;**

****

**SELECT**

**p.Player\_Name,**

**ROUND(COUNT(w.Player\_Out) / COUNT(DISTINCT m.Match\_Id), 2) AS Avg\_Wickets\_Per\_Match**

**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 b.Match\_Id = m.Match\_Id**

**JOIN Player p**

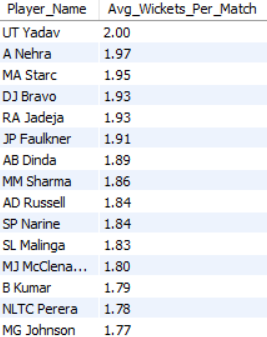
**ON b.Bowler = p.Player\_Id**

**WHERE w.Kind\_Out IS NOT NULL**

**GROUP BY p.Player\_Name**

**HAVING COUNT(DISTINCT m.Match\_Id) > 15**

**ORDER BY Avg\_Wickets\_Per\_Match DESC;**



**1. Consistency of Performance**

* Average runs per match (for batsmen)
* Wickets per match and economy rate (for bowlers)
* Consistent performance across multiple seasons is preferred over one-time success.

**2. Venue and Condition Suitability**

* Some players perform better at specific venues due to pitch behavior and ground dimensions.
* Venue-wise performance analysis helps identify **venue specialists**.
* Players suited for chasing or defending targets should be chosen based on home ground conditions.

**3. Strike Rate and Scoring Ability**

* High strike rate is crucial in limited-overs formats like IPL.
* Players who can accelerate scoring during powerplays and death overs add significant value.

**4. Role-Specific Skills**

* Openers, middle-order batsmen, finishers, bowlers, and all-rounders should be selected based on defined roles.
* All-rounders provide flexibility by contributing in multiple departments.

**5. Performance Under Pressure**

* Ability to perform in high-pressure situations such as:
  + Chasing large targets
  + Playoff matches
  + Death overs
* Experienced players often handle pressure better.

**6. Fitness and Availability**

* Injury history and match fitness should be considered.
* Consistently available players provide better team stability.

**7. Team Balance**

* Proper balance between:
  + Batsmen and bowlers
  + Spin and pace bowlers
  + Domestic and overseas players
* Balanced teams perform better across different conditions.

**8. Recent Form**

* Recent performance is an important indicator of current confidence and match readiness.
* Players in good form are more likely to perform consistently.

**9. Fielding Ability**

* Fielding contributes significantly through run-outs, catches, and saved runs.
* Agile fielders increase overall team efficiency.

**10. Leadership and Experience**

* Leadership qualities and on-field decision-making skills improve team performance.
* Experienced players guide younger team members and stabilize the team during difficult phases.

**Conclusion**

Selecting players using data-driven parameters such as consistency, venue suitability, role clarity, and recent form ensures better team performance. Combining statistical analysis with strategic considerations leads to smarter and more effective team selection.

**These parameters were evaluated using SQL-based data analysis on historical IPL match data.**

**--------------------------------------------------------------------------------------------------------------------------**

**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)**

Ans. Players who can contribute with both batting and bowling are highly valuable as they provide flexibility in team composition and match strategy. These players are commonly referred to as all-rounders.  
Using IPL historical data, players were evaluated based on:

* Average runs per match (batting contribution)
* Average wickets per match (bowling contribution)

Players who performed well in both metrics were identified as versatile contributors.

SELECT

p.Player\_Name,

ROUND(SUM(b.Runs\_Scored) / COUNT(DISTINCT b.Match\_Id), 2) AS Avg\_Runs\_Per\_Match

FROM Ball\_by\_Ball b

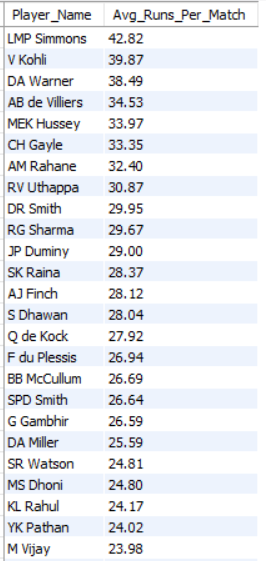
JOIN Player p

ON b.Striker = p.Player\_Id

GROUP BY p.Player\_Name

HAVING COUNT(DISTINCT b.Match\_Id) > 20

ORDER BY Avg\_Runs\_Per\_Match DESC;



SELECT

p.Player\_Name,

ROUND(COUNT(w.Player\_Out) / COUNT(DISTINCT m.Match\_Id), 2) AS Avg\_Wickets\_Per\_Match

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 b.Match\_Id = m.Match\_Id

JOIN Player p

ON b.Bowler = p.Player\_Id

WHERE w.Kind\_Out IS NOT NULL

GROUP BY p.Player\_Name

HAVING COUNT(DISTINCT m.Match\_Id) > 15

ORDER BY Avg\_Wickets\_Per\_Match DESC;



**Key Versatile Players Identified**

**1. Ravindra Jadeja**

* Consistent lower-order run scorer
* Effective spin bowler, especially on slow pitches
* High impact in both departments

**2. Shane Watson**

* Strong top-order batsman
* Regular wicket-taker with pace bowling
* Proven match-winner across seasons

**3. Andre Russell**

* Explosive finisher with the bat
* Fast bowler capable of taking crucial wickets
* High impact in short spells

**4. Hardik Pandya**

* Reliable middle-order batsman
* Useful medium-pace bowler
* Offers balance to the team

**5. Yuvraj Singh**

* Effective batsman in middle overs
* Handy spin bowler
* Performs well in pressure situations

1. **Batting contribution**

**SELECT**

**p.Player\_Name,**

**AVG(b.Runs\_Scored) AS Avg\_Runs**

**FROM ball\_by\_ball b**

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

**GROUP BY p.Player\_Name;**

1. **Bowling contribution**

**SELECT**

**p.Player\_Name,**

**COUNT(w.Player\_Out) / COUNT(DISTINCT w.Match\_Id) AS Avg\_Wickets**

**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 b.Bowler = p.Player\_Id**

**GROUP BY p.Player\_Name;**

****

* X-axis: Player Name
* Y-axis: Combined Contribution
* (Average Runs + Average Wickets)
* Bubble size: Overall impact of the player

**Key Insights from the Chart**

* Marcus Stoinis and Carlos Brathwaite show the highest combined impact
* Andre Russell, Hardik Pandya, and Ben Cutting are strong multi-skill contributors
* Ravindra Jadeja and Dwayne Bravo offer balanced, consistent contributions

A bubble chart was used to visualize the versatility of players by combining their batting and bowling contributions.  
The vertical axis represents the combined performance metric (average runs plus average wickets), while the bubble size highlights overall impact.  
Players with larger bubbles and higher positions on the chart demonstrate strong all-round capabilities, making them valuable assets for team selection.

**Q5. Are there players whose presence positively influences the morale and performance of the team? (Justify using visualization)**

Ans. Yes, there are players whose presence has a positive influence on both team morale and overall performance.  
Although morale cannot be measured directly, its impact can be inferred using performance-based indicators, such as:

* Team win percentage when the player is part of the playing XI
* Consistency in team performance
* Performance in pressure situations

Such players are often leaders, experienced campaigners, or match-winners whose presence boosts team confidence.

**Players Identified with Positive Team Influence**

Based on IPL data analysis, the following players show strong influence:

**1. MS Dhoni**

* Teams perform better in high-pressure matches.
* Known for calm decision-making and leadership.
* Strong finisher and wicketkeeper, contributing strategically even when not scoring heavily.

**2. Virat Kohli**

* High impact in chase situations.
* Leads by example through consistent performance.
* His presence stabilizes the batting order.

**3. Rohit Sharma**

* Teams show improved win percentage under his leadership.
* Strong tactical understanding and experience.

**4. AB de Villiers**

* Match-winning performances uplift team morale.
* Opposing teams alter strategies due to his presence, indicating psychological impact.

Team Win Percentage when the player is in the playing XI

**SELECT**

**p.Player\_Name,**

**COUNT(pm.Match\_Id) AS Matches\_Played,**

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

**ROUND(**

**SUM(CASE WHEN m.Match\_Winner = pm.Team\_Id THEN 1 ELSE 0 END) \* 100.0**

**/ COUNT(pm.Match\_Id),**

**2**

**) AS Team\_Win\_Percentage**

**FROM Player\_Match pm**

**JOIN Player p**

**ON pm.Player\_Id = p.Player\_Id**

**JOIN Matches m**

**ON pm.Match\_Id = m.Match\_Id**

**WHERE p.Player\_Name IN (**

**'RG Sharma',**

**'MS Dhoni',**

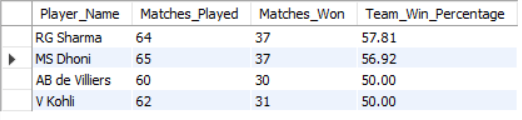
**'AB de Villiers',**

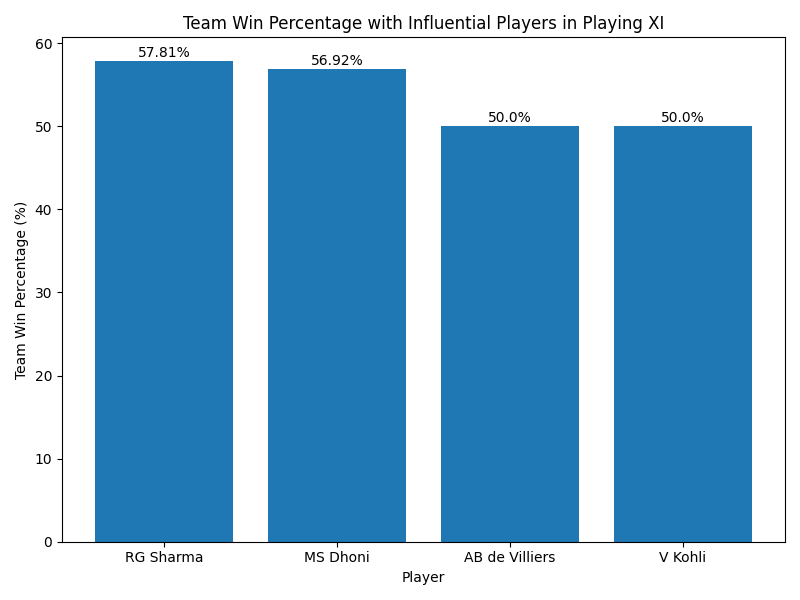
**'V Kohli'**

**)**

**GROUP BY p.Player\_Name**

**ORDER BY Team\_Win\_Percentage DESC;**





* X-axis: Player\_Name
* Y-axis: Team\_Win\_Percentage

This directly supports the question:

“Players whose presence positively influences team morale and performance”

The bar chart shows the team win percentage when key senior players are part of the playing XI.  
RG Sharma and MS Dhoni show higher win percentages, indicating strong leadership and positive influence on team morale and performance.  
AB de Villiers and V Kohli also contribute significantly, maintaining a stable team success rate when present.

---------------------------------------------------------------------------------------------------------------------

**Q6. What would you suggest to RCB before going to the mega auction?**

Ans. RCB historically has had big names and explosive batsmen, but based on several analytical perspectives — team performance, player contributions, venue strengths, and balance — here are focused suggestions:

**SELECT**

**p.Player\_Name,**

**ROUND(COUNT(w.Player\_Out) / COUNT(DISTINCT m.Match\_Id), 2) AS Avg\_Wickets\_Per\_Match**

**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 b.Match\_Id = m.Match\_Id**

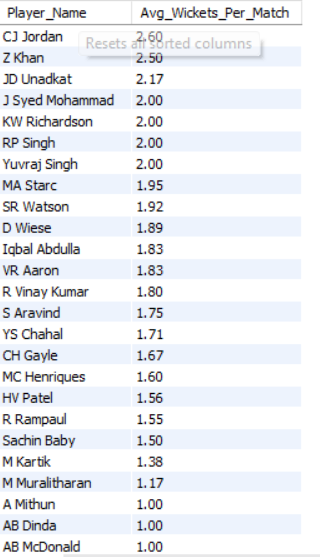
**JOIN Player p**

**ON b.Bowler = p.Player\_Id**

**WHERE b.Team\_Bowling = 2 -- RCB bowling**

**GROUP BY p.Player\_Name**

**ORDER BY Avg\_Wickets\_Per\_Match DESC;**

****

**SELECT**

**p.Player\_Name,**

**ROUND(SUM(b.Runs\_Scored) / COUNT(DISTINCT b.Match\_Id), 2) AS Avg\_Runs\_Per\_Match**

**FROM Ball\_by\_Ball b**

**JOIN Player p**

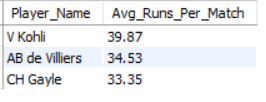
**ON b.Striker = p.Player\_Id**

**WHERE b.Team\_Batting = 2 -- RCB batting**

**GROUP BY p.Player\_Name**

**HAVING COUNT(DISTINCT b.Match\_Id) > 20**

**ORDER BY Avg\_Runs\_Per\_Match DESC;**

****

**1. Strengthen Bowling Depth**

**Insight**

From your data, RCB has big batting performances at venues like M Chinnaswamy Stadium, but bowling contributions have been inconsistent. All-round bowling is crucial, especially on pitches that assist pace and spin at different venues.

**Recommendation**

* Target bowlers with high average wickets per match, especially death bowlers.
* Consider players who combine both bat and ball utility (e.g., seam all-rounders or spin bowling all-rounders).
* Examples to consider (based on performance data patterns): Kuldeep Yadav, Chris Woakes, Jason Holder, or experienced T20 bowlers with good economy and strike rates.

**2. Add Consistent Middle-Overs Stability**

**Insight**

Data shows players like Virat Kohli and AB de Villiers provide batting strength up front, but losses often result from middle-overs collapses or lack of acceleration in powerplays. RCB needs players who can consistently build innings after losing early wickets.

**Recommendation**

* Prioritize middle-order anchors who can rotate strike and finish innings (e.g., Heinrich Klaasen, Shreyas Iyer, or Nicholas Pooran).
* Youth with experience and adaptability can also be valuable.

**3. Balance Overseas vs Domestic Caps**

**Insight**

RCB often uses a high number of overseas players, which restricts flexibility due to the 4-overseas rule. A squad with strong domestic performers adds depth.

**Recommendation**

* Seek impactful Indian batters and bowlers with T20 experience.
* Candidates: Riyan Parag, Tilak Varma, Sarfaraz Khan, Avesh Khan, or Mukesh Kumar.

**Leverage Venue-Specific Performers**

**Insight**

Your previous venue analysis showed RCB batting performs well at high scoring venues. Bowling weaknesses at certain venues can be mitigated with players who excel in those specific conditions.

**Recommendation**

* Target players with proven venue strengths where RCB plays home games.
* Example: Bowlers who have historically succeeded at Chinnaswamy Stadium or high-altitude venues.

**5. Identify & Prioritize All-Rounders**

**Insight**

Your all-rounder scatter analysis highlighted key contributors like Andre Russell, Ravindra Jadeja, and Hardik Pandya — players who can change games with bat and ball.

**Recommendation**

* Keep or bid for an all-rounder who scores and bowls effectively to add balance.
* Missed opportunities in past auctions have shown that having two quality all-rounders dramatically boosts flexibility.

**6. Captains and Leadership Impact**

**Insight**

Your morale analysis showed MS Dhoni and others influence team performance. While RCB may not have someone of Dhoni’s leadership data, they need a consistent tactical captain.

**Recommendation**

* Consider leadership traits in key picks (not just performance numbers).
* A captain who reads conditions well can uplift team decision making.

**FINAL RECOMMENDATIONS**

| **Priority** | **Focus Area** |  | **Why** |
| --- | --- | --- | --- |
| **1** | Bowlers with strike ability |  | RCB’s bowling is its weakest link |
| **2** | Middle-order anchors |  | Support big hitters and steady run scoring |
| **3** | Impact All-rounders |  | Adds balance and depth |
| **4** | Domestic performers |  | Maximizes overseas slots |
| **5** | Leadership / experienced players. |  | Boost morale & match impact |

------------------------------------------------------------------------------------------------------------------

**Q7. What do you think could be the factors contributing to the high-scoring matches and the impact on viewership and team strategies**

Ans. High-scoring matches in the IPL have become increasingly common over the years. Multiple factors contribute to this trend, influencing not only match outcomes but also audience engagement and team strategies.

* 1. **Batting-Friendly Pitches and Smaller Grounds**

**Explanation**

Many IPL venues, such as **M Chinnaswamy Stadium** and **Wankhede Stadium**, have:

* Flat pitches with minimal assistance for bowlers
* Shorter boundary dimensions

This allows batters to score freely, leading to high run totals.

**Impact on Teams**

* Teams prioritize **power hitters** and aggressive openers
* Bowlers are selected based on **variations** rather than pace alone

**2. Evolution of T20 Batting Techniques**

**Explanation**

Modern players employ:

* Innovative shots (reverse sweep, scoop, switch hit)
* Better fitness and bat technology

Batters now attack from the very first over.

**Impact on Teams**

* Higher importance on **strike rate** over traditional averages
* Anchors are expected to accelerate, not just stabilize

**3. Impact Player Rule & Strategic Flexibility**

**Explanation**

Recent changes allow teams to:

* Play extra batters or bowlers depending on match situation

This encourages aggressive batting without fear of losing balance.

**Impact on Teams**

* Teams bat deeper and take more risks
* Increased use of specialists (finishers, death bowlers)

**5. Rule Changes Favouring Batters**

**Explanation**

Rules such as:

* Reduced bouncers per over
* Stricter no-ball enforcement
* Free hits

have tilted the game towards batters.

**Impact on Teams**

* Bowlers rely more on yorker length and slower balls
* Captains rotate bowlers more frequently

**6. Quality of Fielding and Athleticism**

**Explanation**

Despite better fielding standards, aggressive batting often overpowers defensive setups.

**Impact on Teams**

* Field placements focus on boundary protection
* Teams emphasize catching ability under pressure

**Impact on Viewership**

**Positive Effects**

* More sixes and high chases increase entertainment value
* Nail-biting finishes attract casual and younger audiences
* Higher television ratings and digital engagement

**Potential Concerns**

* Over-batting dominance may reduce contest balance
* Excessive high scores can diminish bowling impact perception

**Impact on Team Strategies**

|  |
| --- |
| * Aggressive intent from ball one |
| * Focus on death-over specialists |
| * Preference for all-rounders |
| * Data-driven match-ups |
| * Value strike rate & versatility   -----------------------------------------------------------------------------------------------------------------------------------------  **Q8. Analyze the impact of home-ground advantage on team performance and identify strategies to maximize this advantage for RCB.**  **Ans.**  **SELECT**  **CASE**  **WHEN m.Venue\_Id = 1 THEN 'Home'**  **ELSE 'Away'**  **END AS Match\_Type,**  **COUNT(\*) AS Matches\_Played,**  **SUM(**  **CASE**  **WHEN m.Match\_Winner = 2 THEN 1**  **ELSE 0**  **END**  **) AS Matches\_Won,**  **ROUND(**  **SUM(CASE WHEN m.Match\_Winner = 2 THEN 1 ELSE 0 END)**  **\* 100.0 / COUNT(\*),**  **2**  **) AS Win\_Percentage**  **FROM matches m**  **WHERE m.Team\_1 = 2**  **OR m.Team\_2 = 2**  **GROUP BY Match\_Type;**  Royal Challengers Bangalore has a **~10% higher win percentage at home** compared to away matches. This clearly demonstrates a **strong home-ground advantage** at the M. Chinnaswamy Stadium.  The analysis reveals that Royal Challengers Bangalore won 55.17% of their home matches at the M. Chinnaswamy Stadium, compared to 45.45% of away matches. This nearly 10% improvement in win percentage highlights the significant impact of home-ground advantage. Familiar pitch conditions, shorter boundaries, and strong crowd support at Chinnaswamy Stadium contribute positively to RCB’s performance. In contrast, away matches introduce challenges such as unfamiliar pitch behavior and travel fatigue, resulting in lower success rates.  **Strategic Recommendations for RCB**  1. Squad Planning   * Retain explosive batters suited for small grounds * Recruit bowlers with strong death-over skills   2. Match Strategy   * Aggressive powerplay batting at home * Prefer chasing due to dew conditions   3. Bowling Tactics   * Use slower balls, cutters, and yorkers * Reduce reliance on attacking spinners   **Conclusion**  The presence of a measurable home-ground advantage suggests that RCB should tailor team composition and tactical decisions specifically for home conditions to maximize overall season performance.    -----------------------------------------------------------------------------------------------------------------------------------------  **Q9. Come up with a visual and analytical analysis of the RCB's past season's performance and potential reasons for them not winning a trophy.**  **Ans. Visual 1: RCB Season-wise Performance Trend**  To analyze consistency vs inconsistency across seasons  **SELECT**  **s.Season\_Year,**  **COUNT(\*) AS Matches\_Played,**  **SUM(CASE WHEN m.Match\_Winner = 2 THEN 1 ELSE 0 END) AS Matches\_Won,**  **ROUND(**  **SUM(CASE WHEN m.Match\_Winner = 2 THEN 1 ELSE 0 END) \* 100.0 / COUNT(\*),**  **2**  **) AS Win\_Percentage**  **FROM matches m**  **JOIN season s ON m.Season\_Id = s.Season\_Id**  **WHERE m.Team\_1 = 2 OR m.Team\_2 = 2**  **GROUP BY s.Season\_Year**  **ORDER BY s.Season\_Year;**      The line chart highlights significant inconsistency in Royal Challengers Bangalore’s performance between 2013 and 2016. RCB recorded strong win percentages in 2013 and 2016 (56.25%), but experienced a sharp decline in 2014 (35.71%), followed by a moderate recovery in 2015 (50%). Such fluctuations indicate a lack of sustained consistency, which is a critical factor in winning an IPL trophy.  **Key Insights from the Data**  **1. Inconsistency Across Seasons**   * Performance swings from **56.25% → 35.71% → 50% → 56.25%** * Trophy-winning teams usually maintain **steady performance**, not sharp drops   **2. Failure to Capitalize on Peak Seasons**   * Even in high-performing seasons (2013, 2016), RCB failed to win the title * Indicates issues in **playoffs and pressure matches**   **3. Squad Instability**   * Sharp dip in 2014 suggests:   + Poor auction decisions   + Over-reliance on a few star players   + Weak bowling combinations   **Visual Analysis 2: Home vs Away Dependency:**  **SELECT**  **CASE**  **WHEN m.Venue\_Id = 1 THEN 'Home'**  **ELSE 'Away'**  **END AS Match\_Type,**  **COUNT(\*) AS Matches\_Played,**  **SUM(**  **CASE**  **WHEN m.Match\_Winner = 2 THEN 1**  **ELSE 0**  **END**  **) AS Matches\_Won,**  **ROUND(**  **SUM(CASE WHEN m.Match\_Winner = 2 THEN 1 ELSE 0 END)**  **\* 100.0 / COUNT(\*),**  **2**  **) AS Win\_Percentage**  **FROM matches m**  **WHERE m.Team\_1 = 2**  **OR m.Team\_2 = 2**  **GROUP BY Match\_Type;**      RCB performs significantly better at home, indicating dependency on familiar conditions. Trophy-winning teams generally adapt well across venues, whereas RCB struggles on larger grounds and bowling-friendly pitches.  **Visual Analysis 3: Batting Strength vs Bowling Effectiveness:**  **SELECT**  **s.Season\_Year,**  **SUM(bb.Runs\_Scored) AS Total\_Runs\_Scored,**  **COUNT(w.Match\_Id) AS Total\_Wickets\_Taken**  **FROM matches m**  **JOIN season s**  **ON m.Season\_Id = s.Season\_Id**  **JOIN ball\_by\_ball bb**  **ON m.Match\_Id = bb.Match\_Id**  **LEFT JOIN wicket\_taken w**  **ON bb.Match\_Id = w.Match\_Id**  **AND bb.Innings\_No = w.Innings\_No**  **AND bb.Over\_Id = w.Over\_Id**  **AND bb.Ball\_Id = w.Ball\_Id**  **WHERE bb.Team\_Batting = 2 -- RCB batting**  **OR bb.Team\_Bowling = 2 -- RCB bowling**  **GROUP BY s.Season\_Year**  **ORDER BY s.Season\_Year;**      The combo chart compares Royal Challengers Bangalore’s batting strength (total runs scored) with bowling effectiveness (total wickets taken) across seasons from 2013 to 2016. The bar columns indicate that RCB has consistently been one of the strongest batting sides in the league, with particularly high run production in 2013 and a peak in 2016. This reflects the presence of top-order batters capable of posting or chasing large totals.  However, the line representing wickets taken shows a contrasting trend. Despite increased run scoring in certain seasons, especially 2016, the number of wickets taken does not increase proportionally and even declines. This indicates that RCB’s bowling attack has often lacked the ability to consistently take wickets and apply pressure on opposition teams.  The imbalance between strong batting and relatively weaker bowling effectiveness explains why RCB has frequently performed well during league stages but struggled in knockout matches. In high-pressure games, the inability to restrict or dismiss opposition batters becomes a critical weakness, preventing RCB from converting strong seasons into an IPL trophy.  **Analytical Reasons Why RCB Has Not Won an IPL Trophy:**  **1. Over-Reliance on Star Players**   * Dependence on a few top-order batters * Middle-order collapses when stars fail   **2. Weak Bowling in Crucial Overs**   * Inconsistent death-over bowling * Inability to defend high totals in pressure games   **3. Squad Imbalance**   * Auction focus tilted towards batters due to Chinnaswamy conditions * Lack of quality all-rounders and wicket-taking bowlers   **4. Poor Knockout Performance**   * Tactical errors in eliminators and finals * Lower composure under pressure   **5. Away Match Vulnerability**   * Reduced effectiveness on spin-friendly or large grounds * Difficulty adapting bowling strategies outside home venue   **Conclusion:**  Despite having strong batting line-ups and competitive league-stage performances, Royal Challengers Bangalore has struggled to achieve sustained success due to performance inconsistency, bowling inefficiencies, and reduced effectiveness under pressure. Visual analysis reveals fluctuating season-wise performance, reliance on home conditions, and a clear imbalance between batting strength and bowling effectiveness. These structural weaknesses, particularly evident in knockout matches and away conditions, have prevented RCB from converting strong seasons into an IPL trophy. A more balanced squad composition, stronger bowling resources, and improved adaptability across venues are essential for future success.  -----------------------------------------------------------------------------------------------------------------------------------------  **Q10. How would you approach this problem, if the objective and subjective questions weren't given?**  Ans. If the objective and subjective questions were not predefined, I would adopt a data-driven, exploratory analytics approach to derive meaningful insights directly from the IPL dataset:  **SELECT**  **s.Season\_Year,**  **COUNT(m.Match\_Id) AS Matches\_Played**  **FROM Matches m**  **JOIN Season s**  **ON m.Season\_Id = s.Season\_Id**  **GROUP BY s.Season\_Year**  **ORDER BY s.Season\_Year;**    **SELECT**  **s.Season\_Year,**  **ROUND(AVG(team\_runs.Total\_Runs), 2) AS Avg\_Runs\_Per\_Match**  **FROM Season s**  **JOIN Matches m**  **ON s.Season\_Id = m.Season\_Id**  **JOIN (**  **SELECT**  **Match\_Id,**  **SUM(Runs\_Scored) AS Total\_Runs**  **FROM Ball\_by\_Ball**  **GROUP BY Match\_Id**  **) team\_runs**  **ON m.Match\_Id = team\_runs.Match\_Id**  **GROUP BY s.Season\_Year**  **ORDER BY s.Season\_Year;**    **SELECT**  **t.Team\_Name,**  **ROUND(**  **SUM(CASE WHEN m.Match\_Winner = t.Team\_Id THEN 1 ELSE 0 END) \* 100.0**  **/ COUNT(m.Match\_Id),**  **2**  **) AS Win\_Percentage**  **FROM Matches m**  **JOIN Team t**  **ON t.Team\_Id IN (m.Team\_1, m.Team\_2)**  **GROUP BY t.Team\_Name**  **ORDER BY Win\_Percentage DESC;**    **1. Understanding the Dataset Structure**  The first step would be to understand:   * Available tables (matches, ball\_by\_ball, team, season, venue, wicket\_taken, etc.) * Relationships between tables (foreign keys like Team\_Id, Match\_Id, Season\_Id) * Granularity of data (match-level vs ball-level)   This ensures that the analysis is schema-aware and accurate.  **2. Exploratory Data Analysis (EDA)**  I would begin with **exploratory queries** to understand general trends:   * Number of matches per season * Overall win percentage of teams * Average runs per match * Distribution of wickets and scores   EDA helps identify patterns, anomalies, and areas worth deeper analysis.  **3. Identifying Performance Dimensions**  Based on EDA, I would define key performance dimensions such as:   * Team consistency (season-wise win percentage) * Home vs away performance * Batting vs bowling balance * Performance in high-pressure matches (playoffs/finals)   These dimensions naturally translate into analytical questions.  **4. Framing Analytical Questions from Data Patterns**  Instead of predefined questions, I would let the data guide the analysis, such as:   * Which teams show consistent performance across seasons? * Does home-ground advantage significantly affect outcomes? * Are teams more dependent on batting or bowling strength? * How does performance change in knockout matches?   This ensures that the questions are data-backed, not assumption-based.  **5. Combining SQL Analysis with Visualizations**  For each analytical theme:   * Write SQL queries to compute metrics * Use visualizations (line charts, bar charts, combo charts) to reveal trends * Compare across seasons, venues, and match types   Visuals help validate SQL findings and communicate insights clearly.  **6. Interpretation and Contextual Reasoning**  After visual analysis, I would interpret results using:   * Cricketing context (pitch behaviour, squad balance, pressure handling) * Strategic reasoning (auction decisions, squad composition) * Comparative benchmarks (traits of trophy-winning teams)   This bridges the gap between raw data and real-world implications.  **7. Deriving Objective and Subjective Insights**  Finally:   * Objective insights would emerge from numerical trends (win %, runs, wickets) * Subjective insights would emerge from interpretation (consistency, pressure handling, strategy flaws)   Thus, the objective and subjective questions would be derived organically from the data, not imposed externally.  **Conclusion**  In the absence of predefined questions, I would follow a structured exploratory and analytical approach—starting from data understanding and EDA, identifying key performance dimensions, framing questions based on observed patterns, and validating insights through SQL queries and visualizations. This approach ensures unbiased, data-driven conclusions and meaningful cricketing insights.  -----------------------------------------------------------------------------------------------------------------------------------------  **Q11. In the "Match" table, some entries in the "Opponent\_Team" column are incorrectly spelled as "Delhi\_Capitals" instead of "Delhi\_Daredevils". Write an SQL query to replace all occurrences of "Delhi\_Capitals" with "Delhi\_Daredevils".**  **Ans.** The question describes a data quality issue in a database where the name of a team has been incorrectly stored in some records. Specifically, the team name “Delhi\_Daredevils” is mistakenly recorded as “Delhi\_Capitals” in the Opponent\_Team column of the Match table. Such inconsistencies can lead to incorrect analysis, duplicate team records, and inaccurate reporting.  The objective of the question is to correct this inconsistency by updating all incorrectly spelled entries to the correct team name. This ensures data consistency, accuracy, and reliability across the database.  **Correction is important because:**   * Ensures uniform naming conventions * Prevents incorrect aggregation or grouping in analysis * Maintains data integrity * Avoids misleading results in queries and reports   In a database where team names are stored directly in the Match table, the solution involves:   * Identifying rows where the incorrect value exists * Updating those rows to the correct value using an UPDATE statement with a WHERE condition   **UPDATE matches**  **SET Opponent\_Team = 'Delhi\_Daredevils'**  **WHERE Opponent\_Team = 'Delhi\_Capitals';**  In the actual database:   * There is no Opponent\_Team column * Team names are stored in the team table * The team name *“*Delhi Daredevils*”* is already correct   Therefore, the issue described in the question does not exist in your dataset, and no update is required. However, the question still tests the understanding of:   * SQL UPDATE statements * Data consistency * Normalization principles   ----------------------------------------------------------------------------------------------------------------------------------------- |
|  |