**Objective Questions**

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

|  |  |
| --- | --- |
| COLUMN NAME | DATATYPE |
| Match\_Id | int PK |
| Over\_id | int PK |
| Ball\_Id | int PK |
| Innings\_No | int PK |
| Team\_Batting | int |
| Team\_Bowling | int |
| Striker\_Batting\_Position | int |
| Striker | int |
| Non\_Striker | int |
| Bowler | int |

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

*The total runs scored by RCB in 1st season including the extra runs is 1983*

*Runs scored = 1865, Extras = 118*

Query

WITH RCB\_Season1\_Matches AS (

SELECT \* FROM matches WHERE Season\_Id = 1

AND (Team\_1 = (SELECT Team\_Id FROM team WHERE Team\_Name = "Royal Challengers Bangalore")

OR Team\_2 = (SELECT Team\_Id FROM team WHERE Team\_Name = "Royal Challengers Bangalore"))

),

RCB\_Season1\_Matches\_And\_Innings AS (

SELECT Match\_Id, Toss\_Winner, Toss\_Decide,

CASE

WHEN Toss\_Winner = 2 AND Toss\_Decide = 2 THEN 1

WHEN Toss\_Winner = 2 AND Toss\_Decide = 1 THEN 2

WHEN Toss\_Winner != 2 AND Toss\_Decide = 2 THEN 2

WHEN Toss\_Winner != 2 AND Toss\_Decide = 1 THEN 1

END AS RCB\_Batting\_Innings\_No

FROM RCB\_season1\_matches

),

RCB\_Runs\_Scored AS (

SELECT rcbmi.Match\_Id, rcbmi.RCB\_Batting\_Innings\_No, SUM(bs.Runs\_Scored) AS RCB\_runs

FROM RCB\_Season1\_Matches\_And\_Innings rcbmi

JOIN batsman\_scored bs ON rcbmi.Match\_Id = bs.Match\_Id AND rcbmi.RCB\_Batting\_Innings\_No = bs.Innings\_No

GROUP BY 1,2

),

RCB\_Runs\_Scored\_and\_Extras AS (

SELECT rcbrs.Match\_Id, rcbrs.RCB\_Batting\_Innings\_No, rcbrs.RCB\_runs, SUM(er.Extra\_Runs) AS extras

FROM RCB\_Runs\_Scored rcbrs

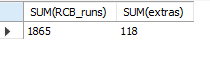
JOIN extra\_runs er ON rcbrs.Match\_Id = er.Match\_Id AND rcbrs.RCB\_Batting\_Innings\_No = er.Innings\_No

GROUP BY 1,2,3

)

SELECT SUM(RCB\_runs), SUM(extras) FROM RCB\_Runs\_Scored\_and\_Extras;

Result



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

*There were about 165 players above the age of 25 during IPL season 2*

Query

WITH Player\_Ages AS (

SELECT \*, YEAR(CURRENT\_DATE()) - YEAR(DOB) AS Current\_Age FROM player

),

Season2\_Players AS (

SELECT DISTINCT Player\_Id

FROM player\_match

WHERE Match\_Id IN (SELECT Match\_Id FROM matches WHERE Season\_Id = 2)

)

SELECT COUNT(pa.Player\_Name) AS players\_above\_25\_years\_of\_age\_in\_season2

FROM Season2\_Players s2p

JOIN Player\_Ages pa ON s2p.Player\_Id = pa.Player\_Id

WHERE Current\_Age > 25;

Result



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

*RCB won only 4 matches out of 14 in IPL Season 1*

Query

WITH RCB\_Season1\_Matches AS(

SELECT \* FROM matches WHERE Season\_Id = 1

AND (Team\_1 = (SELECT Team\_Id FROM team WHERE Team\_Name = "Royal Challengers Bangalore")

OR Team\_2 = (SELECT Team\_Id FROM team WHERE Team\_Name = "Royal Challengers Bangalore"))

)

SELECT COUNT(Match\_Winner) AS RCB\_Season1\_Wins FROM RCB\_Season1\_Matches WHERE Match\_Winner = (

SELECT Team\_Id FROM team WHERE Team\_Name = "Royal Challengers Bangalore"

);

Result



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

Query

WITH Last\_4Seasons\_Ball\_by\_Ball AS (

SELECT \* FROM Ball\_by\_ball

WHERE Match\_Id IN (SELECT Match\_Id FROM matches WHERE Season\_Id IN (6,7,8,9))

ORDER BY Match\_Id, Innings\_No, Over\_Id, Ball\_Id

),

Runs\_Scored\_by\_Player AS(

SELECT l4sb.Striker AS Player\_Id, SUM(bs.Runs\_Scored) AS Total\_Runs\_Scored

FROM Last\_4Seasons\_Ball\_by\_Ball l4sb

JOIN batsman\_scored bs ON bs.Match\_Id = l4sb.Match\_Id AND bs.Innings\_No = l4sb.Innings\_No AND bs.Over\_Id = l4sb.Over\_Id AND bs.Ball\_Id = l4sb.Ball\_Id

GROUP BY l4sb.Striker

HAVING SUM(bs.Runs\_Scored) > 0

ORDER BY Total\_Runs\_Scored DESC

),

Balls\_Faced\_by\_Player AS (

SELECT Striker AS Player\_Id, COUNT(Ball\_Id) AS Total\_Balls\_Faced

FROM Last\_4Seasons\_Ball\_by\_Ball

GROUP BY Striker

),

Strike\_rates AS (

SELECT rsbp.Player\_Id, ROUND((rsbp.Total\_Runs\_Scored/bfbp.Total\_Balls\_Faced)\*100,2) AS Strike\_Rate

FROM Runs\_Scored\_by\_Player rsbp

JOIN Balls\_Faced\_by\_Player bfbp ON rsbp.Player\_Id = bfbp.Player\_Id

)

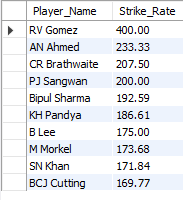
SELECT p.Player\_Name, sr.Strike\_Rate

FROM Strike\_rates sr

JOIN player p ON sr.Player\_Id = p.Player\_Id

ORDER BY Strike\_Rate DESC LIMIT 10;

Result



1. **What is the average of runs scored by each batsman considering all the seasons?**

Query

WITH Ball\_by\_Ball\_with\_Seasons AS (

SELECT bb.\*, m.Season\_Id

FROM matches m

JOIN Ball\_by\_ball bb ON m.Match\_Id = bb.Match\_Id

ORDER BY Match\_Id, Innings\_No, Over\_Id, Ball\_Id

),

Runs\_scored\_per\_season AS (

SELECT

bbbws.Striker AS Player\_Id,

bbbws.Season\_Id,

SUM(bs.Runs\_Scored) AS Runs\_Scored\_in\_Season

FROM batsman\_scored bs

LEFT JOIN Ball\_by\_Ball\_with\_Seasons bbbws

ON bbbws.Match\_Id = bs.Match\_Id

AND bbbws.Innings\_No = bs.Innings\_No

AND bbbws.Over\_Id = bs.Over\_Id

AND bbbws.Ball\_Id = bs.Ball\_Id

GROUP BY bbbws.Striker, bbbws.Season\_Id

),

Average\_runs\_scored\_by\_each\_player\_in\_seasons AS (

SELECT

Player\_Id,

ROUND(AVG(Runs\_Scored\_in\_Season), 2) AS Avg\_runs\_in\_seasons

FROM Runs\_scored\_per\_season

GROUP BY Player\_Id

HAVING COUNT(DISTINCT Season\_Id) > 1

),

Average\_Runs AS (

SELECT

p.Player\_Name,

ars.Avg\_runs\_in\_seasons AS Avg\_Runs\_Scored

FROM Average\_runs\_scored\_by\_each\_player\_in\_seasons ars

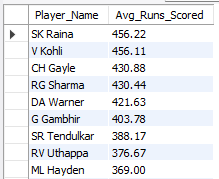
JOIN player p ON p.Player\_Id = ars.Player\_Id

ORDER BY Avg\_Runs\_Scored DESC

)

SELECT \* FROM Average\_Runs;

Result

****

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

Query

WITH Ball\_by\_Ball\_with\_Seasons AS (

SELECT bb.\*, m.Season\_Id

FROM matches m

JOIN Ball\_by\_ball bb ON m.Match\_Id = bb.Match\_Id

ORDER BY Match\_Id, Innings\_No, Over\_Id, Ball\_Id

),

Wickets\_taken\_per\_season AS (

SELECT

bbbws.Bowler AS Player\_Id,

bbbws.Season\_Id,

COUNT(\*) AS Wickets\_in\_Season

FROM wicket\_taken wt

LEFT JOIN Ball\_by\_Ball\_with\_Seasons bbbws

ON bbbws.Match\_Id = wt.Match\_Id

AND bbbws.Innings\_No = wt.Innings\_No

AND bbbws.Over\_Id = wt.Over\_Id

AND bbbws.Ball\_Id = wt.Ball\_Id

WHERE wt.Kind\_Out NOT IN (3, 5, 9)

GROUP BY bbbws.Bowler, bbbws.Season\_Id

),

Avg\_wickets\_taken\_by\_each\_player\_in\_seasons AS (

SELECT

Player\_Id,

COUNT(DISTINCT Season\_Id) AS No\_of\_Seasons\_Played,

SUM(Wickets\_in\_Season) AS Total\_Wickets,

ROUND(AVG(Wickets\_in\_Season), 2) AS Avg\_wickets\_in\_seasons

FROM Wickets\_taken\_per\_season

GROUP BY Player\_Id

HAVING COUNT(DISTINCT Season\_Id) > 1

),

Average\_Wickets AS (

SELECT

p.Player\_Name,

awt.Avg\_wickets\_in\_seasons

FROM Avg\_wickets\_taken\_by\_each\_player\_in\_seasons awt

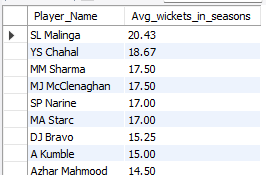
JOIN player p ON p.Player\_Id = awt.Player\_Id

ORDER BY awt.Avg\_wickets\_in\_seasons DESC

)

SELECT \* FROM Average\_Wickets;

Result



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

Query

WITH Ball\_by\_Ball\_with\_Seasons AS(

SELECT bb.\*, m.Season\_Id

FROM matches m

JOIN Ball\_by\_ball bb ON m.Match\_Id = bb.Match\_Id

ORDER BY Match\_Id, Innings\_No, Over\_Id, Ball\_Id

),

Average\_runs\_scored\_by\_each\_player\_in\_seasons AS (

SELECT

bbbws.Striker AS Player\_Id,

SUM(bs.Runs\_Scored) AS Total\_Runs,

COUNT(DISTINCT bbbws.Season\_Id) AS No\_of\_Seasons\_Played,

ROUND(SUM(bs.Runs\_Scored)/COUNT(DISTINCT bbbws.Season\_Id),2) AS Avg\_runs\_in\_seasons

FROM batsman\_scored bs

LEFT JOIN Ball\_by\_Ball\_with\_Seasons bbbws ON bbbws.Match\_Id = bs.Match\_Id AND bbbws.Innings\_No = bs.Innings\_No AND bbbws.Over\_Id = bs.Over\_Id AND bbbws.Ball\_Id = bs.Ball\_Id

GROUP BY bbbws.Striker

),

Playerwise\_Avg\_Runs AS (

SELECT p.Player\_Name, ars.Avg\_runs\_in\_seasons

FROM Average\_runs\_scored\_by\_each\_player\_in\_seasons ars

JOIN player p ON p.Player\_Id = ars.Player\_Id

ORDER BY 2 DESC

),

Overall\_avg\_runs\_scored AS (

SELECT ROUND(AVG(Avg\_runs\_in\_seasons),2) AS Overall\_Avg\_Runs FROM Playerwise\_Avg\_Runs

),

Avg\_wickets\_taken\_by\_each\_player\_in\_seasons AS (

SELECT

bbbws.Bowler AS Player\_Id,

COUNT(\*) AS wickets\_taken\_by\_bowler,

COUNT(DISTINCT bbbws.Season\_Id) AS No\_of\_Seasons\_Played,

ROUND(COUNT(\*)/COUNT(DISTINCT bbbws.Season\_Id),2) AS Avg\_wickets\_in\_seasons

FROM wicket\_taken wt

LEFT JOIN Ball\_by\_Ball\_with\_Seasons bbbws ON bbbws.Match\_Id = wt.Match\_Id AND bbbws.Innings\_No = wt.Innings\_No AND bbbws.Over\_Id = wt.Over\_Id AND bbbws.Ball\_Id = wt.Ball\_Id

WHERE wt.Kind\_Out NOT IN (3,5,9)

GROUP BY bbbws.Bowler

),

Playerwise\_Avg\_Wickets AS (

SELECT p.Player\_Name, awt.Avg\_wickets\_in\_seasons

FROM Avg\_wickets\_taken\_by\_each\_player\_in\_seasons awt

JOIN player p ON p.Player\_Id = awt.Player\_Id

ORDER BY 2 DESC

),

Overall\_avg\_wickets\_taken AS (

SELECT ROUND(AVG(Avg\_wickets\_in\_seasons),2) AS Overall\_Avg\_Wickets FROM Playerwise\_Avg\_Wickets

),

Players\_Above\_Avg AS (

SELECT p.Player\_Name,

ars.Avg\_runs\_in\_seasons,

awt.Avg\_wickets\_in\_seasons

FROM player p

LEFT JOIN Average\_runs\_scored\_by\_each\_player\_in\_seasons ars ON p.Player\_Id = ars.Player\_Id

LEFT JOIN Avg\_wickets\_taken\_by\_each\_player\_in\_seasons awt ON p.Player\_Id = awt.Player\_Id

WHERE ars.Avg\_runs\_in\_seasons > (SELECT Overall\_Avg\_Runs FROM Overall\_avg\_runs\_scored)

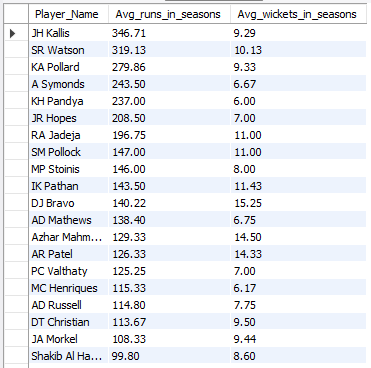
AND awt.Avg\_wickets\_in\_seasons > (SELECT Overall\_Avg\_Wickets FROM Overall\_avg\_wickets\_taken)

)

SELECT \* FROM Players\_Above\_Avg

ORDER BY Avg\_runs\_in\_seasons DESC, Avg\_wickets\_in\_seasons DESC;

Result

****

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

Query

CREATE TABLE rcb\_record AS

WITH RCB\_Matches AS (

SELECT \*

FROM matches

WHERE Team\_1 = (SELECT Team\_Id FROM team WHERE Team\_Name = "Royal Challengers Bangalore")

OR Team\_2 = (SELECT Team\_Id FROM team WHERE Team\_Name = "Royal Challengers Bangalore")

),

Venuewise\_wins\_and\_losses AS (

SELECT

Venue\_Id,

COUNT(CASE WHEN Match\_winner = (SELECT Team\_Id FROM team WHERE Team\_Name = "Royal Challengers Bangalore") THEN 1 END) AS Wins,

COUNT(CASE WHEN Match\_winner != (SELECT Team\_Id FROM team WHERE Team\_Name = "Royal Challengers Bangalore") THEN 1 END) AS Losses

FROM RCB\_Matches

WHERE Outcome\_Type IN (1,3)

GROUP BY Venue\_Id

)

SELECT

v.Venue\_Name,

vwl.Wins,

vwl.Losses

FROM Venuewise\_wins\_and\_losses vwl

JOIN venue v ON vwl.Venue\_Id = v.Venue\_Id;

SELECT \* FROM rcb\_record

Result



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

Query

WITH

Total\_wickets\_taken\_by\_each\_player\_in\_seasons AS (

SELECT

bbbws.Bowler AS Player\_Id,

COUNT(\*) AS wickets\_taken\_by\_bowler

FROM wicket\_taken wt

LEFT JOIN Ball\_by\_Ball bbbws ON bbbws.Match\_Id = wt.Match\_Id AND bbbws.Innings\_No = wt.Innings\_No AND bbbws.Over\_Id = wt.Over\_Id AND bbbws.Ball\_Id = wt.Ball\_Id

WHERE wt.Kind\_Out NOT IN (3,5,9)

GROUP BY bbbws.Bowler

)

SELECT bs.Bowling\_skill, SUM(twts.wickets\_taken\_by\_bowler) AS Wickets\_taken

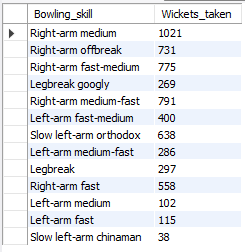
FROM Total\_wickets\_taken\_by\_each\_player\_in\_seasons twts

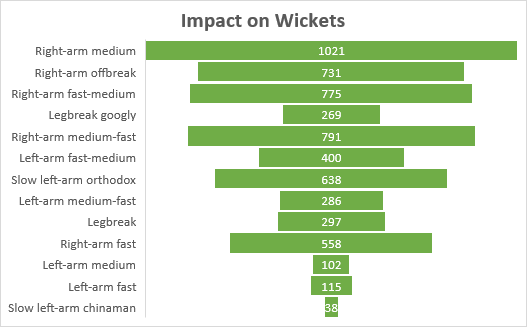
JOIN player p ON p.Player\_Id = twts.Player\_Id

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

GROUP BY p.Bowling\_skill;

Result

****



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

Query

WITH RCB\_Matches AS(

SELECT \* FROM matches

WHERE Team\_1 = (SELECT Team\_Id FROM team WHERE Team\_Name = "Royal Challengers Bangalore")

OR Team\_2 = (SELECT Team\_Id FROM team WHERE Team\_Name = "Royal Challengers Bangalore")

),

RCB\_Matches\_And\_Innings AS(

SELECT Match\_Id, Toss\_Winner, Toss\_Decide,

CASE

WHEN Toss\_Winner = 2 AND Toss\_Decide = 2 THEN 1

WHEN Toss\_Winner = 2 AND Toss\_Decide = 1 THEN 2

WHEN Toss\_Winner != 2 AND Toss\_Decide = 2 THEN 2

WHEN Toss\_Winner != 2 AND Toss\_Decide = 1 THEN 1

END AS RCB\_Batting\_Innings\_No,

CASE

WHEN Toss\_Winner = 2 AND Toss\_Decide = 2 THEN 2

WHEN Toss\_Winner = 2 AND Toss\_Decide = 1 THEN 1

WHEN Toss\_Winner != 2 AND Toss\_Decide = 2 THEN 1

WHEN Toss\_Winner != 2 AND Toss\_Decide = 1 THEN 2

END AS RCB\_Bowling\_Innings\_No

FROM RCB\_matches

),

RCB\_Wickets\_Taken AS(

SELECT rcbmi.Match\_Id, rcbmi.RCB\_Bowling\_Innings\_No, COUNT(\*) AS RCB\_wickets\_taken

FROM RCB\_Matches\_And\_Innings rcbmi

JOIN wicket\_taken wt on rcbmi.Match\_Id = wt.Match\_Id AND rcbmi.RCB\_Bowling\_Innings\_No = wt.Innings\_No

GROUP BY 1,2

),

RCB\_Runs\_Scored AS(

SELECT rcbmi.Match\_Id, rcbmi.RCB\_Batting\_Innings\_No, SUM(bs.Runs\_Scored) AS RCB\_runs\_scored

FROM RCB\_Matches\_And\_Innings rcbmi

JOIN batsman\_scored bs ON rcbmi.Match\_Id = bs.Match\_Id AND rcbmi.RCB\_Batting\_Innings\_No = bs.Innings\_No

GROUP BY 1,2

),

RCB\_Runs\_Scored\_and\_Extras AS(

SELECT rcbrs.Match\_Id, rcbrs.RCB\_Batting\_Innings\_No, rcbrs.RCB\_runs\_scored, SUM(er.Extra\_Runs) AS extras

FROM RCB\_Runs\_Scored rcbrs

JOIN extra\_runs er ON rcbrs.Match\_Id = er.Match\_Id AND rcbrs.RCB\_Batting\_Innings\_No = er.Innings\_No

GROUP BY 1,2,3

),

RCB\_Total\_Runs\_Scored AS (

SELECT Match\_Id, RCB\_Batting\_Innings\_No, (RCB\_runs\_scored + extras) AS RCB\_runs\_total FROM RCB\_Runs\_Scored\_and\_Extras

)

SELECT

s.Season\_Year,

SUM(rcbtrs.RCB\_runs\_total) AS RCB\_runs\_total,

SUM(rcbwt.RCB\_wickets\_taken) AS RCB\_wickets\_taken,

COALESCE(LAG(SUM(rcbtrs.RCB\_runs\_total)) OVER (ORDER BY s.Season\_Year ASC),"-") AS Previous\_year\_runs\_total,

COALESCE(LAG(SUM(rcbwt.RCB\_wickets\_taken)) OVER (ORDER BY s.Season\_Year ASC),"-") AS Previous\_year\_wickets\_taken,

CASE

WHEN SUM(rcbtrs.RCB\_runs\_total) > LAG(SUM(rcbtrs.RCB\_runs\_total)) OVER (ORDER BY s.Season\_Year ASC)

AND SUM(rcbwt.RCB\_wickets\_taken) > LAG(SUM(rcbwt.RCB\_wickets\_taken)) OVER (ORDER BY s.Season\_Year ASC)

THEN "Improved"

WHEN SUM(rcbtrs.RCB\_runs\_total) > LAG(SUM(rcbtrs.RCB\_runs\_total)) OVER (ORDER BY s.Season\_Year ASC)

AND SUM(rcbwt.RCB\_wickets\_taken) < LAG(SUM(rcbwt.RCB\_wickets\_taken)) OVER (ORDER BY s.Season\_Year ASC)

THEN "Better in Scoring runs"

WHEN SUM(rcbtrs.RCB\_runs\_total) < LAG(SUM(rcbtrs.RCB\_runs\_total)) OVER (ORDER BY s.Season\_Year ASC)

AND SUM(rcbwt.RCB\_wickets\_taken) > LAG(SUM(rcbwt.RCB\_wickets\_taken)) OVER (ORDER BY s.Season\_Year ASC)

THEN "Better in Grabing wickets"

WHEN SUM(rcbtrs.RCB\_runs\_total) < LAG(SUM(rcbtrs.RCB\_runs\_total)) OVER (ORDER BY s.Season\_Year ASC)

AND SUM(rcbwt.RCB\_wickets\_taken) < LAG(SUM(rcbwt.RCB\_wickets\_taken)) OVER (ORDER BY s.Season\_Year ASC)

THEN "Bad"

ELSE "-"

END AS Status

FROM RCB\_Total\_Runs\_Scored rcbtrs

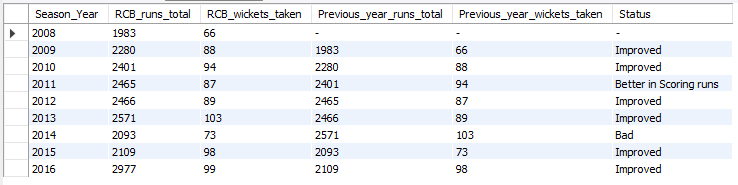
JOIN RCB\_Wickets\_Taken rcbwt ON rcbtrs.Match\_Id = rcbwt.Match\_Id

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

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

GROUP BY 1;

Result

****

1. **Derive more KPIs for the team strategy?**
   1. **Team Performance:**
   2. **Captain’s Win Rate**
   3. **Boundary Counts**
   4. **Seasonwise Wickets**
   5. **Seasonwise Runs**
2. **Using SQL, write a query to find out average wickets taken by each bowler in each venue. Also rank the gender according to the average value.**

Query

WITH Venuewise\_wickets AS (

SELECT

bbb.Bowler AS Player\_Id,

v.Venue\_Name,

COUNT(\*) AS wickets\_taken\_by\_bowler,

COUNT(DISTINCT wt.Match\_Id) AS matches\_played

FROM wicket\_taken wt

LEFT JOIN Ball\_by\_Ball bbb ON bbb.Match\_Id = wt.Match\_Id AND bbb.Innings\_No = wt.Innings\_No AND bbb.Over\_Id = wt.Over\_Id AND bbb.Ball\_Id = wt.Ball\_Id

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

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

WHERE wt.Kind\_Out NOT IN (3,5,9)

GROUP BY bbb.Bowler, v.Venue\_Name

ORDER BY bbb.Bowler

)

SELECT

p.Player\_Name AS Bowler,

Venue\_Name,

ROUND(vw.wickets\_taken\_by\_bowler/vw.matches\_played ,2) AS Avg\_wickets,

DENSE\_RANK() OVER (PARTITION BY Venue\_Name ORDER BY ROUND(vw.wickets\_taken\_by\_bowler/vw.matches\_played ,2) DESC) AS `Rank`

FROM Venuewise\_wickets vw

JOIN player p ON p.Player\_Id = vw.Player\_Id

ORDER BY Venue\_Name ASC;

Result



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

*Conditions for consistency in Batting:*

*Number of seasons played > 3*

*Total runs scored > 700*

*Total Batting\_average > 30*

|  |
| --- |
| **Consistent in Batting** |
| AB de Villiers |
| AC Gilchrist |
| AJ Finch |
| AM Rahane |
| AT Rayudu |
| BB McCullum |
| BJ Hodge |
| CH Gayle |
| CL White |
| DA Miller |
| DA Warner |
| DJ Hussey |
| DPMD Jayawardene |
| DR Smith |
| G Gambhir |
| JA Morkel |
| JH Kallis |
| JP Duminy |
| KA Pollard |
| KC Sangakkara |
| KP Pietersen |
| LRPL Taylor |
| M Vijay |
| MEK Hussey |
| MK Pandey |
| MK Tiwary |
| MS Dhoni |
| R Dravid |
| RG Sharma |
| RV Uthappa |
| S Badrinath |
| S Dhawan |
| SC Ganguly |
| SE Marsh |
| SK Raina |
| SPD Smith |
| SR Tendulkar |
| SR Watson |
| SS Tiwary |
| TM Dilshan |
| V Kohli |
| V Sehwag |
| WP Saha |
| YK Pathan |
| Yuvraj Singh |

*Conditions for consistency in Bowling:*

*Number of seasons played > 3*

*Total wickets taken > 30*

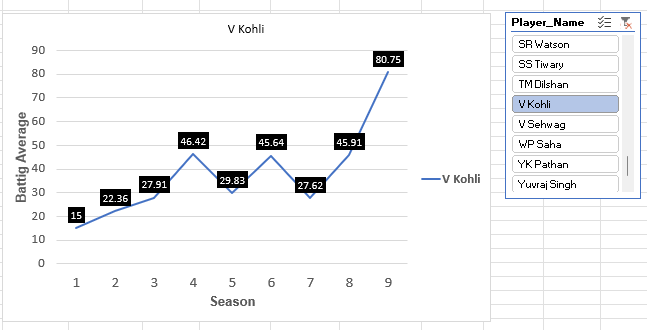
*Total Bowling\_average < 25*

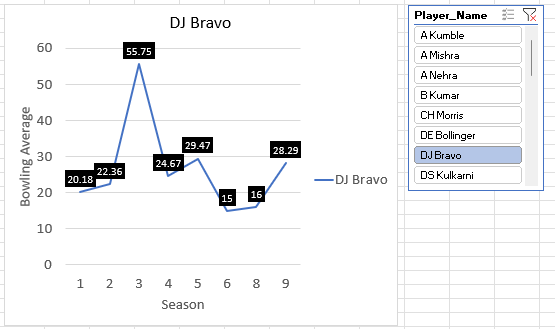
|  |
| --- |
| **Consistent in Bowling** |
| A Kumble |
| A Mishra |
| A Nehra |
| B Kumar |
| CH Morris |
| DE Bollinger |
| DJ Bravo |
| DS Kulkarni |
| DW Steyn |
| KK Cooper |
| MA Starc |
| MG Johnson |
| MJ McClenaghan |
| MM Patel |
| MM Sharma |
| R Ashwin |
| RJ Harris |
| RP Singh |
| S Aravind |
| Sandeep Sharma |
| Shakib Al Hasan |
| SK Warne |
| SL Malinga |
| SP Narine |
| YS Chahal |
|  |

***Line Chart****: Plotting each player's batting average across multiple seasons allows us to observe the trend over time. Players with relatively stable lines across seasons can be identified as more consistent, whereas players with fluctuating lines might be less consistent.*

*To achieve this, two pivot tables were created with the result of below queries and a line chart was plotted with a slicer for each player for both (batting and bowling)*

**Eg:**

****

****

Query

**-- For Seasonwise Batting\_averages of each player**

WITH Ball\_by\_Ball\_with\_Seasons AS (

SELECT bb.\*, m.Season\_Id

FROM matches m

JOIN Ball\_by\_ball bb ON m.Match\_Id = bb.Match\_Id

ORDER BY Match\_Id, Innings\_No, Over\_Id, Ball\_Id

),

Batting\_Averages AS (

SELECT

bbbws.Striker AS Player\_Id,

bbbws.Season\_Id,

SUM(bs.Runs\_Scored) AS Season\_Runs\_Scored,

COUNT(wt.Player\_Out) AS Season\_Outs,

ROUND(SUM(bs.Runs\_Scored)/

COUNT(wt.Player\_Out),2) AS Batting\_Average

FROM batsman\_scored bs

LEFT JOIN Ball\_by\_Ball\_with\_Seasons bbbws

ON bbbws.Match\_Id = bs.Match\_Id

AND bbbws.Innings\_No = bs.Innings\_No

AND bbbws.Over\_Id = bs.Over\_Id

AND bbbws.Ball\_Id = bs.Ball\_Id

LEFT JOIN wicket\_taken wt

ON bbbws.Striker = wt.Player\_Out

AND bbbws.Match\_Id = wt.Match\_Id

AND bbbws.Innings\_No = wt.Innings\_No

AND bbbws.Over\_Id = wt.Over\_Id

AND bbbws.Ball\_Id = wt.Ball\_Id

GROUP BY bbbws.Striker, bbbws.Season\_Id

ORDER BY bbbws.Striker, bbbws.Season\_Id

)

SELECT p.Player\_Name, ba.Season\_Id, ba.Batting\_Average

FROM Batting\_Averages ba

JOIN player p ON p.Player\_Id = ba.Player\_Id

WHERE ba.Batting\_Average IS NOT NULL

AND p.Player\_Id IN

(SELECT Player\_Id FROM Batting\_Averages

GROUP BY Player\_Id

HAVING COUNT(Season\_Id) > 3 AND SUM(Season\_Runs\_Scored) > 700 AND SUM(Season\_Runs\_Scored)/SUM(Season\_Outs) > 30

)

ORDER BY p.Player\_Name, ba.Season\_Id ASC;

**-- For Seasonwise Bowling\_averages of each player**

WITH Ball\_by\_Ball\_with\_Seasons AS (

SELECT bb.\*, m.Season\_Id

FROM matches m

JOIN Ball\_by\_ball bb ON m.Match\_Id = bb.Match\_Id

ORDER BY Match\_Id, Innings\_No, Over\_Id, Ball\_Id

),

Bowling\_Averages AS (

SELECT

bbbws.Bowler AS Player\_Id,

bbbws.Season\_Id,

SUM(bs.runs\_scored) AS Season\_runs\_given,

SUM(CASE WHEN wt.Kind\_Out IS NOT NULL AND wt.Kind\_Out NOT IN (3, 5, 9) THEN 1 ELSE 0 END) AS Season\_wickets,

ROUND(SUM(bs.runs\_scored)/

SUM(CASE WHEN wt.Kind\_Out IS NOT NULL AND wt.Kind\_Out NOT IN (3, 5, 9) THEN 1 ELSE 0 END),2) AS Bowling\_Average

FROM Ball\_by\_Ball\_with\_Seasons bbbws

LEFT JOIN wicket\_taken wt

ON bbbws.Match\_Id = wt.Match\_Id

AND bbbws.Innings\_No = wt.Innings\_No

AND bbbws.Over\_Id = wt.Over\_Id

AND bbbws.Ball\_Id = wt.Ball\_Id

LEFT JOIN batsman\_scored bs

ON bbbws.Match\_Id = bs.Match\_Id

AND bbbws.Innings\_No = bs.Innings\_No

AND bbbws.Over\_Id = bs.Over\_Id

AND bbbws.Ball\_Id = bs.Ball\_Id

WHERE bbbws.Bowler IS NOT NULL

GROUP BY bbbws.Bowler, bbbws.Season\_Id

ORDER BY Bowling\_Average ASC

)

SELECT p.Player\_Name, ba.Season\_Id, ba.Bowling\_Average

FROM Bowling\_Averages ba

JOIN player p ON p.Player\_Id = ba.Player\_Id

WHERE ba.Bowling\_Average IS NOT NULL

AND p.Player\_Id IN

(SELECT Player\_Id FROM Bowling\_Averages

GROUP BY Player\_Id

HAVING COUNT(Season\_Id) > 3 AND SUM(Season\_wickets) > 30 AND SUM(Season\_runs\_given)/SUM(Season\_wickets) < 25

)

ORDER BY p.Player\_Name, ba.Season\_Id;

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

Some players do exhibit exceptionally high batting average in some venues.

This can be visualized using a clustered column chart below.

*1. Top Venue-Specific Performers*

MK Pandey at SuperSport Park: Batting Avg = 114.00, Avg Difference = 86.53

GC Smith at Chepauk: Batting Avg = 91.00, Avg Difference = 67.38

*2. Players with High Consistency*

JD Ryder (Feroz Shah Kotla): Avg = 86.00, Diff = 62.58

SR Watson (Subrata Roy Sahara Stadium): Avg = 90.00, Diff = 62.61

GJ Maxwell (Sharjah Stadium): Avg = 92.00, Diff = 64.33

*3. Best Performers at M Chinnaswamy Stadium*

RS Bopara: Batting Avg = 77.00, Diff = 51.40

TL Suman: Batting Avg = 78.00, Diff = 58.73

*4. Players with Avg Difference > 60*

MK Pandey, GJ Maxwell, JD Ryder, SR Watson, S Dhawan show exceptional venue-specific boosts.

Players like MK Pandey and GC Smith excel at specific venues, while others consistently outperform their averages at select stadiums.

**Subjective Questions**

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

Query

WITH toss\_winner\_as\_match\_winner AS (

SELECT

Venue\_Id,

COUNT(CASE WHEN Toss\_Decide = 1 THEN 1 END) AS Fielding\_first,

COUNT(CASE WHEN Toss\_Decide = 2 THEN 1 END) AS Batting\_first

FROM matches

WHERE Toss\_Winner = Match\_Winner

GROUP BY Venue\_Id

)

SELECT

v.Venue\_Name,

Fielding\_first+Batting\_first AS No\_of\_matches,

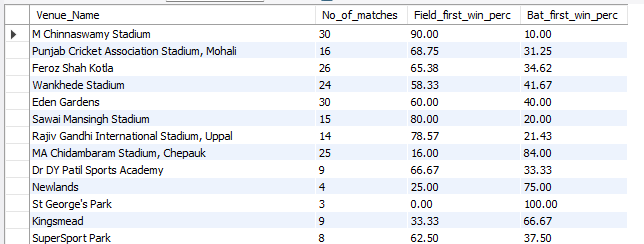
ROUND((Fielding\_first/(Fielding\_first+Batting\_first))\*100,2) AS Field\_first\_win\_perc,

ROUND((Batting\_first/(Fielding\_first+Batting\_first))\*100,2) AS Bat\_first\_win\_perc

FROM toss\_winner\_as\_match\_winner twmw

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

Result



Analysis & Suggestions

High Impact Venues:

* M Chinnaswamy Stadium: Field-first win percentage is very high (90%), indicating a clear advantage in fielding first.
* Sawai Mansingh Stadium and Rajiv Gandhi International Stadium: Field-first win percentages are also very high (80% and 78.57%, respectively), suggesting a strong field-first advantage.
* MA Chidambaram Stadium, Chepauk: The opposite is observed, with an 84% bat-first win rate, indicating that batting first is beneficial at this venue.

Balanced Venues:

* Venues like De Beers Diamond Oval, OUTsurance Oval, Sardar Patel Stadium, Motera, and Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Stadium have a 50-50 win distribution between batting and fielding first, suggesting that the toss decision might not significantly affect match outcomes at these venues.

Significant Biases:

* Some venues like St George's Park and Subrata Roy Sahara Stadium have a 100% win-rate for batting first, while others like New Wanderers Stadium and Holkar Cricket Stadium show a 100% win rate for fielding first. This indicates a venue-specific preference in toss outcomes.

Observations on Specific Venues:

* MA Chidambaram Stadium, Chepauk consistently shows a higher bat-first advantage (84%), which could be due to the pitch’s tendency to deteriorate, making batting easier earlier on.
* Venues in South Africa (like Newlands and Kingsmead) tend to show a higher win rate for batting first, possibly due to weather or pitch conditions favoring early batting stability.

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

|  |  |  |  |
| --- | --- | --- | --- |
| **Exceptionally Consistent Batsmen** | **Exceptionally Consistent Bowlers** | **Highly effective All-Rounders** | **Young Talents** |
| V Kohli | DS Kulkarni | AD Russell | SV Samson |
| MEK Hussey | A Mishra | DJ Bravo | AR Patel |
| MS Dhoni | SL Malinga | DR Smith | VH Zol |
| SK Raina | SP Narine | JA Morkel | SS Iyer |
| A Symonds | MM Patel | KA Pollard | DJ Hooda |
| AM Rahane | Sandeep Sharma | MC Henriques | J Suchith |
| SE Marsh | DJ Bravo | RA Jadeja | SN Khan |
| AB de Villiers | RJ Harris | SR Watson | Ishan Kishan |
| JP Duminy | A Nehra | Yuvraj Singh | NS Naik |
| KP Pietersen | S Aravind |  | RR Pant |
| CH Gayle | B Kumar |  | KC Cariappa |
| DA Warner | YS Chahal |  | S Kaushik |
| F du Plessis |  |  | Kuldeep Yadav |
| SPD Smith |  |  |  |
| DA Miller |  |  |  |

*Exceptionally Consistent Batsmen*

*No of seasons played > 3*

*Total Runs scored > 800*

*Batting Average > 35*

*Exceptionally Consistent Bowlers*

*No of seasons played > 3*

*Total Wickets taken > 35*

*Bowling Average > 23*

*Highly effective All Rounders*

*Number of seasons played > 3*

*Total runs scored > 500*

*Total Batting\_average > 20*

*Total wickets taken > 20*

*Total Bowling\_average < 30*

*Young Talents*

*Country – India*

*Age – 18 – 23*

From the above list, RCB might target any 3-4 batsmen, 2-3 bowlers, 1-2 all-rounders and invest in 1-2 Indian youngsters for the Mega-Auction. These players will best fit for the team.

Query

-- For Exceptionally Consistent Batsmen

WITH Ball\_by\_Ball\_with\_Seasons AS (

SELECT bb.\*, m.Season\_Id

FROM matches m

JOIN Ball\_by\_ball bb ON m.Match\_Id = bb.Match\_Id

ORDER BY Match\_Id, Innings\_No, Over\_Id, Ball\_Id

),

Batting\_Averages AS (

SELECT

bbbws.Striker AS Player\_Id,

bbbws.Season\_Id,

SUM(bs.Runs\_Scored) AS Season\_Runs\_Scored,

COUNT(wt.Player\_Out) AS Season\_Outs,

ROUND(SUM(bs.Runs\_Scored)/

COUNT(wt.Player\_Out),2) AS Batting\_Average

FROM batsman\_scored bs

LEFT JOIN Ball\_by\_Ball\_with\_Seasons bbbws

ON bbbws.Match\_Id = bs.Match\_Id

AND bbbws.Innings\_No = bs.Innings\_No

AND bbbws.Over\_Id = bs.Over\_Id

AND bbbws.Ball\_Id = bs.Ball\_Id

LEFT JOIN wicket\_taken wt

ON bbbws.Striker = wt.Player\_Out

AND bbbws.Match\_Id = wt.Match\_Id

AND bbbws.Innings\_No = wt.Innings\_No

AND bbbws.Over\_Id = wt.Over\_Id

AND bbbws.Ball\_Id = wt.Ball\_Id

GROUP BY bbbws.Striker, bbbws.Season\_Id

ORDER BY bbbws.Striker, bbbws.Season\_Id

),

Exceptionally\_consistent\_batsmen AS (

SELECT p.Player\_Name, ba.Season\_Id, ba.Batting\_Average

FROM Batting\_Averages ba

JOIN player p ON p.Player\_Id = ba.Player\_Id

WHERE ba.Batting\_Average IS NOT NULL

AND p.Player\_Id IN

(SELECT Player\_Id FROM Batting\_Averages

GROUP BY Player\_Id

HAVING COUNT(Season\_Id) > 3 AND SUM(Season\_Runs\_Scored) >= 800 AND SUM(Season\_Runs\_Scored)/SUM(Season\_Outs) > 35

)

ORDER BY p.Player\_Name, ba.Season\_Id ASC

)

SELECT DISTINCT Player\_Name FROM Exceptionally\_consistent\_batsmen;

-- For Exceptionally Consistent Bowlers

WITH Ball\_by\_Ball\_with\_Seasons AS (

SELECT bb.\*, m.Season\_Id

FROM matches m

JOIN Ball\_by\_ball bb ON m.Match\_Id = bb.Match\_Id

ORDER BY Match\_Id, Innings\_No, Over\_Id, Ball\_Id

),

Bowling\_Averages AS (

SELECT

bbbws.Bowler AS Player\_Id,

bbbws.Season\_Id,

SUM(bs.runs\_scored) AS Season\_runs\_given,

SUM(CASE WHEN wt.Kind\_Out IS NOT NULL AND wt.Kind\_Out NOT IN (3, 5, 9) THEN 1 ELSE 0 END) AS Season\_wickets,

ROUND(SUM(bs.runs\_scored)/

SUM(CASE WHEN wt.Kind\_Out IS NOT NULL AND wt.Kind\_Out NOT IN (3, 5, 9) THEN 1 ELSE 0 END),2) AS Bowling\_Average

FROM Ball\_by\_Ball\_with\_Seasons bbbws

LEFT JOIN wicket\_taken wt

ON bbbws.Match\_Id = wt.Match\_Id

AND bbbws.Innings\_No = wt.Innings\_No

AND bbbws.Over\_Id = wt.Over\_Id

AND bbbws.Ball\_Id = wt.Ball\_Id

LEFT JOIN batsman\_scored bs

ON bbbws.Match\_Id = bs.Match\_Id

AND bbbws.Innings\_No = bs.Innings\_No

AND bbbws.Over\_Id = bs.Over\_Id

AND bbbws.Ball\_Id = bs.Ball\_Id

WHERE bbbws.Bowler IS NOT NULL

GROUP BY bbbws.Bowler, bbbws.Season\_Id

ORDER BY Bowling\_Average ASC

),

Exceptionally\_consistent\_bowlers AS (

SELECT p.Player\_Name, ba.Season\_Id, ba.Bowling\_Average

FROM Bowling\_Averages ba

JOIN player p ON p.Player\_Id = ba.Player\_Id

WHERE ba.Bowling\_Average IS NOT NULL

AND p.Player\_Id IN

(SELECT Player\_Id FROM Bowling\_Averages

GROUP BY Player\_Id

HAVING COUNT(Season\_Id) > 3 AND SUM(Season\_wickets) >= 35 AND SUM(Season\_runs\_given)/SUM(Season\_wickets) <= 23

)

ORDER BY p.Player\_Name, ba.Season\_Id

)

SELECT DISTINCT Player\_Name FROM Exceptionally\_consistent\_bowlers;

-- For Young Talents from India

SELECT Player\_Name AS Youngsters

FROM player

WHERE 2017 - YEAR(DOB) BETWEEN 18 AND 23

AND Country\_Name = (SELECT Country\_Id FROM country WHERE Country\_Name = 'India');

-- For Highly effective All-Rounders

WITH Ball\_by\_Ball\_with\_Seasons AS (

SELECT bb.\*, m.Season\_Id

FROM matches m

JOIN Ball\_by\_ball bb ON m.Match\_Id = bb.Match\_Id

ORDER BY Match\_Id, Innings\_No, Over\_Id, Ball\_Id

),

Batting\_Averages AS (

SELECT

bbbws.Striker AS Player\_Id,

bbbws.Season\_Id,

SUM(bs.Runs\_Scored) AS Season\_Runs\_Scored,

COUNT(wt.Player\_Out) AS Season\_Outs,

ROUND(SUM(bs.Runs\_Scored)/

COUNT(wt.Player\_Out),2) AS Batting\_Average

FROM batsman\_scored bs

LEFT JOIN Ball\_by\_Ball\_with\_Seasons bbbws

ON bbbws.Match\_Id = bs.Match\_Id

AND bbbws.Innings\_No = bs.Innings\_No

AND bbbws.Over\_Id = bs.Over\_Id

AND bbbws.Ball\_Id = bs.Ball\_Id

LEFT JOIN wicket\_taken wt

ON bbbws.Striker = wt.Player\_Out

AND bbbws.Match\_Id = wt.Match\_Id

AND bbbws.Innings\_No = wt.Innings\_No

AND bbbws.Over\_Id = wt.Over\_Id

AND bbbws.Ball\_Id = wt.Ball\_Id

GROUP BY bbbws.Striker, bbbws.Season\_Id

ORDER BY bbbws.Striker, bbbws.Season\_Id

),

Seasonwise\_batting\_averages AS (

SELECT p.Player\_Name, c.Country\_Name, ba.Season\_Id, ba.Season\_Runs\_Scored, ba.Batting\_Average

FROM Batting\_Averages ba

JOIN player p ON p.Player\_Id = ba.Player\_Id

JOIN country c ON p.Country\_Name = c.Country\_Id

WHERE ba.Batting\_Average IS NOT NULL

AND p.Player\_Id IN

(SELECT Player\_Id FROM Batting\_Averages

GROUP BY Player\_Id

HAVING COUNT(Season\_Id) > 3 AND SUM(Season\_Runs\_Scored) > 500 AND SUM(Season\_Runs\_Scored)/SUM(Season\_Outs) > 20

)

ORDER BY p.Player\_Name, ba.Season\_Id ASC

),

Bowling\_Averages AS (

SELECT

bbbws.Bowler AS Player\_Id,

bbbws.Season\_Id,

SUM(bs.runs\_scored) AS Season\_runs\_given,

SUM(CASE WHEN wt.Kind\_Out IS NOT NULL AND wt.Kind\_Out NOT IN (3, 5, 9) THEN 1 ELSE 0 END) AS Season\_wickets,

ROUND(SUM(bs.runs\_scored)/

SUM(CASE WHEN wt.Kind\_Out IS NOT NULL AND wt.Kind\_Out NOT IN (3, 5, 9) THEN 1 ELSE 0 END),2) AS Bowling\_Average

FROM Ball\_by\_Ball\_with\_Seasons bbbws

LEFT JOIN wicket\_taken wt

ON bbbws.Match\_Id = wt.Match\_Id

AND bbbws.Innings\_No = wt.Innings\_No

AND bbbws.Over\_Id = wt.Over\_Id

AND bbbws.Ball\_Id = wt.Ball\_Id

LEFT JOIN batsman\_scored bs

ON bbbws.Match\_Id = bs.Match\_Id

AND bbbws.Innings\_No = bs.Innings\_No

AND bbbws.Over\_Id = bs.Over\_Id

AND bbbws.Ball\_Id = bs.Ball\_Id

WHERE bbbws.Bowler IS NOT NULL

GROUP BY bbbws.Bowler, bbbws.Season\_Id

ORDER BY Bowling\_Average ASC

),

Seasonwise\_bowling\_averages AS(

SELECT p.Player\_Name, c.Country\_Name, ba.Season\_Id, ba.Season\_wickets, ba.Bowling\_Average

FROM Bowling\_Averages ba

JOIN player p ON p.Player\_Id = ba.Player\_Id

JOIN country c ON p.Country\_Name = c.Country\_Id

WHERE ba.Bowling\_Average IS NOT NULL

AND p.Player\_Id IN

(SELECT Player\_Id FROM Bowling\_Averages

GROUP BY Player\_Id

HAVING COUNT(Season\_Id) > 3 AND SUM(Season\_wickets) > 20 AND SUM(Season\_runs\_given)/SUM(Season\_wickets) < 30

)

ORDER BY p.Player\_Name, ba.Season\_Id

)

SELECT sboa.Player\_Name, sboa.Country\_Name, sbaa.Season\_Id, sbaa.Season\_Runs\_Scored, sbaa.Batting\_Average, sboa.Season\_wickets, sboa.Bowling\_Average

FROM Seasonwise\_bowling\_averages sboa

JOIN Seasonwise\_batting\_averages sbaa ON sboa.Player\_Name = sbaa.Player\_Name AND sboa.Season\_Id = sbaa.Season\_Id;

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

Batting Metrics

* Average Runs Scored: Measures a player's run-scoring reliability.
* Strike Rate: Assesses scoring speed, especially valuable in limited-overs formats.
* Standard Deviation of Runs: Indicates consistency; lower values mean more consistent performance.
* Boundary Count (4s and 6s): Shows power-hitting capability, useful in T20 and ODI formats.
* Performance in Death Overs: Measures ability to score quickly in the final overs of an innings.
* Batting Position Versatility: Evaluates adaptability to bat in different positions based on team needs.

Bowling Metrics

* Average Wickets per Match/Season: Reflects a bowler’s wicket-taking ability.
* Economy Rate: Indicates how effectively a bowler restricts runs.
* Standard Deviation of Economy Rate: Measures consistency in maintaining low run rates.
* Strike Rate (Balls per Wicket): Shows efficiency in taking wickets relative to the number of balls bowled.
* Death Over Economy Rate: Specific to performance in high-pressure, final overs of the innings.
* Bowling Style and Specialization: Fast, spin, or medium pace; useful for tailoring strategy against specific opponents.

Further parameters if available might include

Other International Match Performances

* Performance in High-Stakes Matches: Assess past performances in playoffs, finals, and tournaments.
* Venue-Specific Performance: Tracks performance at different international venues (e.g., spin-friendly, bouncy pitches).
* Performance in Different Conditions: Reflects adaptability in various weather and pitch conditions (e.g., swinging pitches vs. flat tracks).
* Match-winning Innings/Bowling Spells: Considers past impactful performances that directly contributed to wins.
* Performance Against Top Teams: Analyzes success rate against strong teams or bowlers/batsmen.

Player Experience

* Number of Matches Played: Provides a general gauge of experience.
* Experience in International Tournaments: Important for high-pressure tournament play.
* Leadership and Mentorship Skills: Assesses potential to mentor younger players or serve as captain/vice-captain.

Fitness & Availability

* Injury History: Assesses risk of injuries and downtime, impacting availability.
* Endurance and Stamina: Important for maintaining high performance through long tournaments.
* Physical Fitness Level: Indicators like agility, speed (especially for fielders), and strength.

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

Query

WITH Ball\_by\_Ball\_with\_Seasons AS (

SELECT bb.\*, m.Season\_Id

FROM matches m

JOIN Ball\_by\_ball bb ON m.Match\_Id = bb.Match\_Id

ORDER BY Match\_Id, Innings\_No, Over\_Id, Ball\_Id

),

Batting\_Averages AS (

SELECT

bbbws.Striker AS Player\_Id,

bbbws.Season\_Id,

SUM(bs.Runs\_Scored) AS Season\_Runs\_Scored,

COUNT(wt.Player\_Out) AS Season\_Outs,

ROUND(SUM(bs.Runs\_Scored)/

COUNT(wt.Player\_Out),2) AS Batting\_Average

FROM batsman\_scored bs

LEFT JOIN Ball\_by\_Ball\_with\_Seasons bbbws

ON bbbws.Match\_Id = bs.Match\_Id

AND bbbws.Innings\_No = bs.Innings\_No

AND bbbws.Over\_Id = bs.Over\_Id

AND bbbws.Ball\_Id = bs.Ball\_Id

LEFT JOIN wicket\_taken wt

ON bbbws.Striker = wt.Player\_Out

AND bbbws.Match\_Id = wt.Match\_Id

AND bbbws.Innings\_No = wt.Innings\_No

AND bbbws.Over\_Id = wt.Over\_Id

AND bbbws.Ball\_Id = wt.Ball\_Id

GROUP BY bbbws.Striker, bbbws.Season\_Id

ORDER BY bbbws.Striker, bbbws.Season\_Id

),

Seasonwise\_batting\_averages AS (

SELECT p.Player\_Name, c.Country\_Name, ba.Season\_Id, ba.Season\_Runs\_Scored, ba.Batting\_Average

FROM Batting\_Averages ba

JOIN player p ON p.Player\_Id = ba.Player\_Id

JOIN country c ON p.Country\_Name = c.Country\_Id

WHERE ba.Batting\_Average IS NOT NULL

AND p.Player\_Id IN

(SELECT Player\_Id FROM Batting\_Averages

GROUP BY Player\_Id

HAVING COUNT(Season\_Id) > 3 AND SUM(Season\_Runs\_Scored) > 500 AND SUM(Season\_Runs\_Scored)/SUM(Season\_Outs) > 20

)

ORDER BY p.Player\_Name, ba.Season\_Id ASC

),

Bowling\_Averages AS (

SELECT

bbbws.Bowler AS Player\_Id,

bbbws.Season\_Id,

SUM(bs.runs\_scored) AS Season\_runs\_given,

SUM(CASE WHEN wt.Kind\_Out IS NOT NULL AND wt.Kind\_Out NOT IN (3, 5, 9) THEN 1 ELSE 0 END) AS Season\_wickets,

ROUND(SUM(bs.runs\_scored)/

SUM(CASE WHEN wt.Kind\_Out IS NOT NULL AND wt.Kind\_Out NOT IN (3, 5, 9) THEN 1 ELSE 0 END),2) AS Bowling\_Average

FROM Ball\_by\_Ball\_with\_Seasons bbbws

LEFT JOIN wicket\_taken wt

ON bbbws.Match\_Id = wt.Match\_Id

AND bbbws.Innings\_No = wt.Innings\_No

AND bbbws.Over\_Id = wt.Over\_Id

AND bbbws.Ball\_Id = wt.Ball\_Id

LEFT JOIN batsman\_scored bs

ON bbbws.Match\_Id = bs.Match\_Id

AND bbbws.Innings\_No = bs.Innings\_No

AND bbbws.Over\_Id = bs.Over\_Id

AND bbbws.Ball\_Id = bs.Ball\_Id

WHERE bbbws.Bowler IS NOT NULL

GROUP BY bbbws.Bowler, bbbws.Season\_Id

ORDER BY Bowling\_Average ASC

),

Seasonwise\_bowling\_averages AS(

SELECT p.Player\_Name, c.Country\_Name, ba.Season\_Id, ba.Season\_wickets, ba.Bowling\_Average

FROM Bowling\_Averages ba

JOIN player p ON p.Player\_Id = ba.Player\_Id

JOIN country c ON p.Country\_Name = c.Country\_Id

WHERE ba.Bowling\_Average IS NOT NULL

AND p.Player\_Id IN

(SELECT Player\_Id FROM Bowling\_Averages

GROUP BY Player\_Id

HAVING COUNT(Season\_Id) > 3 AND SUM(Season\_wickets) > 20 AND SUM(Season\_runs\_given)/SUM(Season\_wickets) < 30

)

ORDER BY p.Player\_Name, ba.Season\_Id

)

SELECT sboa.Player\_Name, sboa.Country\_Name, sbaa.Season\_Id, sbaa.Season\_Runs\_Scored, sbaa.Batting\_Average, sboa.Season\_wickets, sboa.Bowling\_Average

FROM Seasonwise\_bowling\_averages sboa

JOIN Seasonwise\_batting\_averages sbaa ON sboa.Player\_Name = sbaa.Player\_Name AND sboa.Season\_Id = sbaa.Season\_Id;

Analysis

Above query results in season-wise data of each player who are contributing effectively with both bat and ball throughout the matches. This data was used in excel to create a pivot table in order to get the names of all these players which is listed below.

For these ALL-ROUNDER players,

*Effectiveness in Batting is measured by:*

*Number of seasons played > 3*

*Total runs scored > 500*

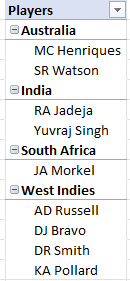
*Total Batting\_average > 20*

*Effectiveness in Bowling is measured by:*

*Number of seasons played > 3*

*Total wickets taken > 20*

*Total Bowling\_average < 30*

\

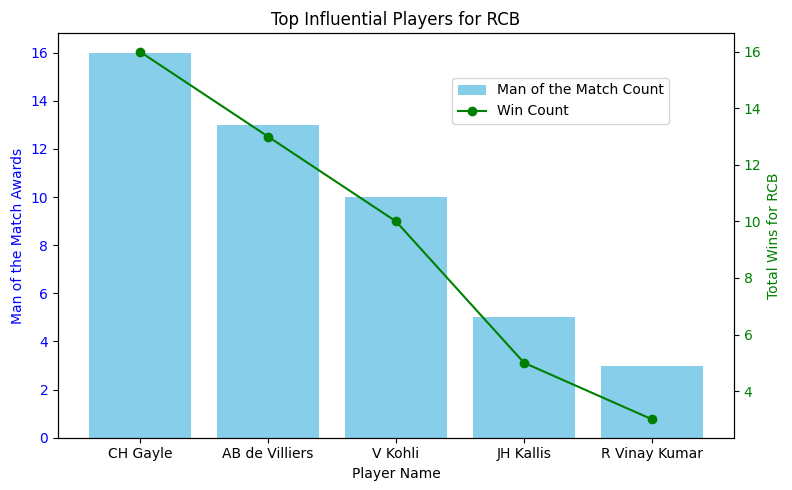
This list highlights players who have made significant contributions with both bat and ball.

Notably, the **West Indies** lead in producing the highest number of all-rounders who have proven highly effective for the team.

Following closely are **Australia and India**, which also feature a strong presence of versatile players who excel in both departments.

(Also the count of 4’s and 6’s could have been great measures to consider for a batting all-rounders)

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



This Visualisation clearly shows

*Top Influencers:*

CH Gayle: Leads with 16 MOM awards, suggesting his consistent match-winning performances significantly boost team morale.

AB de Villiers and Virat Kohli: Follow closely with 13 and 10 awards, highlighting their reliability as top performers.

*Influence on Team Performance:*

Players like CH Gayle, AB de Villiers, and V Kohli regularly perform well when the team wins, indicating that their presence and contributions positively influence both morale and results.

*Lower Influencers:*

While JH Kallis and R Vinay Kumar have fewer MOM awards, they still contribute positively during critical matches.

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

Retain Proven Match-Winners:

Focus on players who have consistently delivered impactful performances, especially in critical moments. Prioritize individuals who can win games single-handedly with bat or ball.

Enhance Squad Flexibility:

Invest in versatile players who can adapt to different roles, such as all-rounders who can contribute in both batting and bowling departments.

Address Bowling Weaknesses:

Strengthen the bowling attack, particularly with reliable death bowlers. Analyze performance data to identify bowlers who excel under pressure and in high-scoring conditions.

Reinforce the Middle Order:

Secure middle-order batsmen who can anchor the innings, stabilize after early setbacks, and accelerate during the final overs. Focus on experienced finishers.

Invest in Emerging Talent:

Target young, promising players from domestic circuits or T20 leagues who can bring fresh energy to the squad. Look for potential future stars to groom.

Balance Experience and Youth:

Maintain a healthy mix of seasoned players and dynamic youngsters. Experienced players provide leadership, while younger players bring agility and long-term potential.

Optimize Home-Ground Advantage:

Choose players who thrive in high-scoring venues like M. Chinnaswamy Stadium. For example, power-hitters and bowlers with excellent control in challenging conditions.

Prioritize Stability:

Build a core team that can perform together over multiple seasons, ensuring continuity and better team dynamics.

Target Big-Game Performers:

Look for players with a strong track record in playoffs and high-pressure matches to enhance the squad’s ability to perform when it matters most.

Enhance Net Run Rate(NRR):

Aggressive Batting - Set ambitious targets in each phase to increase total scores and positively impact NRR.

Maintain Scoring Rate - Avoid periods of low scoring; encourage calculated risks to keep the scoreboard moving.

Fielding Efficiency - Minimize runs conceded through strong fielding, leading to dot balls and wickets that lower the opposition's scoring rate.

Chasing Targets - Develop clear plans based on required run rates, adapting approaches as needed to chase efficiently without losing too many wickets.

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

1. Pitch Conditions

* Factor: Pitches that are flat and well-prepared give batsmen an advantage, leading to higher scores.
* Impact on Viewership: Fans love seeing big hits and boundaries, so these types of pitches tend to keep viewers excited and engaged.
* Impact on Team Strategies: Teams often go for aggressive batting, especially when setting or chasing a big target, knowing the pitch supports high scores.

2. Weather Conditions

* Factor: Hot and humid weather, along with dew in the evening, can make the ball easier to hit and harder for bowlers to control.
* Impact on Viewership: These conditions usually result in higher scores, making the game more exciting and drawing in more viewers.
* Impact on Team Strategies: Teams tend to play more aggressively, especially in the latter stages, and captains tweak their fielding setups to contain runs.

3. Batting Strength and Top-Order Performances

* Factor: Teams with strong top-order batsmen can get off to fast starts, putting pressure on the opposition.
* Impact on Viewership: A powerful opening partnership or standout performances from key batsmen keep fans hooked and excited.
* Impact on Team Strategies: Teams may focus on having hard-hitting openers and finishers to capitalize on the early momentum.

4. Quality of Bowlers

* Factor: When bowlers are either out of form or fatigued, they tend to give away more runs. Strong bowlers, on the other hand, can control the game.
* Impact on Viewership: High-scoring games with fewer wickets tend to be more enjoyable for casual fans, as they can get caught up in the action.
* Impact on Team Strategies: Teams may rotate bowlers frequently to keep them fresh and use their strongest bowlers during key moments to stop the flow of runs.

5. Fielding

* Factor: Poor fielding, like misfields or dropped catches, can lead to easy runs, while strong fielding can restrict scoring.
* Impact on Viewership: Fans love watching brilliant fielding, but mistakes like dropped catches can be frustrating. Good fielding can really elevate the excitement.
* Impact on Team Strategies: Teams focus on strategic field placements to limit boundaries and keep pressure on the batsmen, knowing that even a single misfield can cost runs.

6. Innovations in Batting

* Factor: Modern-day batsmen use creative shots like reverse sweeps or lofted drives, making it easier to score quickly.
* Impact on Viewership: These innovative shots are exciting to watch, and fans love the unpredictability it brings to the game.
* Impact on Team Strategies: Teams often promote aggressive players and adjust field settings to counter these unconventional batting tactics, making the game more dynamic.

1. Analyse the impact of home ground advantage on team performance and identify strategies to maximize this advantage for RCB.

RCB Home Ground Performance

Win Percentage = Total Wins/Total Played ×100 = 50 %

Batting First Win Percentage = Wins Batting First/Total Played Batting First × 100 = 46.4 %

Fielding First Win Percentage = Wins Fielding First/Total Played Fielding First × 100 = 60 %

**Trends Observed:**

Strength in Chasing: RCB performs better when fielding first(60% Wins) with a high winning average of 7 wickets.

Weaker Defending: The win percentage while batting first is consistently lower(46.4% Wins) with again high winning average of 57 runs.

**Strategy Optimization:**   
Focus on fielding first to capitalize on RCB’s strength in defending targets. This approach allows RCB to set a challenging total for the opposition while relying on their bowlers to defend the score effectively.

**Batting First Improvement:**Develop targeted strategies to enhance performance when batting first, such as adjusting the batting order or setting specific target scores for each phase of the innings.  
For example:

Powerplay (1-6 overs): Aim for a quick 75+ runs in the powerplay while ensuring crucial wickets are not lost.

Middle Overs (7-14 overs): Target a steady 60+ runs during the middle overs to build momentum for the final attack.

Death Overs (16-20 overs): Focus on power hitting to accelerate and reach a total of around 200 runs, setting a competitive target.

**Toss Decisions:**In favourable conditions (such as dry pitches or when there is dew), prioritize fielding first to gain an early advantage and put pressure on the opposing team to set a target. When batting first, ensure the conditions are optimal for a high-scoring total.

**Improving Fielding First Win Percentage:**

Analyse opponent weaknesses when defending targets. Focus on bowling strategies that capitalize on the opposition’s vulnerabilities (e.g., top-order weaknesses, chasing struggles).

Enhance fielding performance, particularly in high-pressure situations, by strengthening death-over fielding and placing fielders in key positions to cut off boundary options.

Work on strategies to restrict big partnerships by rotating bowlers and maintaining pressure through constant field adjustments.

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

The team improved significantly from a win percentage of 28.57% in Season 1 to 62.50% in Season 4.

After peaking in Season 4, the team experienced fluctuations, notably dropping to 35.71% in Season 7.

Seasons 2, 5, 6, and 9 showed stable performance with win percentages between 53.33% and 56.25%.

The highest percentile of 1 in Season 4 indicates they were the top-performing team that season, while the lowest of 0 in Season 1 reflects a poor start.

Fluctuations in win percentages may suggest varying performances based on match conditions or locations.

The drop in win percentage during Season 7 highlights the need for introspection and strategies to prevent similar declines.

High-performing seasons likely depended on key players, while dips may indicate underperformance or injuries among them.

To achieve sustained success, the team must focus on maintaining consistency and strengthening both batting and bowling departments.

Key Factors Contributing to RCB's Inability to Win a Trophy

**Inconsistent Middle Order:**

RCB has often relied heavily on star players like Virat Kohli and AB de Villiers for runs, leading to an over-reliance on their performances.

For instance, in 2013, despite having strong top-order batsmen, the middle order's failure to contribute significantly led to several match losses.

**Bowling Weaknesses:**

RCB’s bowling attack has historically been inconsistent, particularly in death overs.

In 2016, despite reaching the finals, they struggled with bowling depth, which became evident in crucial matches.

**Pressure Situations:**

RCB has faced immense pressure during high-stakes matches, particularly in playoffs.

In seasons where they reached the finals (like 2009, 2011, and 2016), the inability to handle pressure often led to disappointing results.

**Injuries and Player Form:**

Injuries to key players have severely impacted team performance across multiple seasons.

**Coaching and Strategy:**

Frequent changes in coaching staff have disrupted team dynamics and strategies.

The lack of a stable strategy contributed to their inability to build a cohesive team capable of competing for titles.

**Over-reliance on Star Players:**

While having star players like Kohli and de Villiers is an asset, an over-reliance on them can backfire if they fail to perform or are out of form.

In recent seasons (notably from 2015 onwards), when Kohli or other stars faltered, the team struggled to find adequate replacements or support within the squad.

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

To approach the problem without predefined questions, I would:

1. Understand the Problem

Objective: Help RCB identify top-performing, reliable players while optimizing auction spending for the 2017 mega auction.

Constraints: Limited budget, team composition rules, and specific needs like power-hitters or death bowlers.

2. Gather and Analyse Data

Auction Data: Player base price, final price trends, and value-for-money patterns.

Performance Data: Metrics like runs scored, strike rate, economy rate, and consistency.

Team Fit: Players’ roles, adaptability, and experience at M. Chinnaswamy Stadium (a batting-friendly venue).

3. Identify Key Metrics

Performance:

For batsmen: Average, strike rate, and clutch performances.

For bowlers: Wickets taken, economy rate, and death-over effectiveness.

Value for Money: Cost-per-run or wicket, potential long-term impact.

4. Develop Strategies

Prioritize undervalued players with consistent performance.

Target players who fill gaps in team composition (e.g., finishers or death bowlers).

Build a mix of experienced performers and promising young talent.

5. Validate and Communicate

Shortlist players based on performance and auction trends.

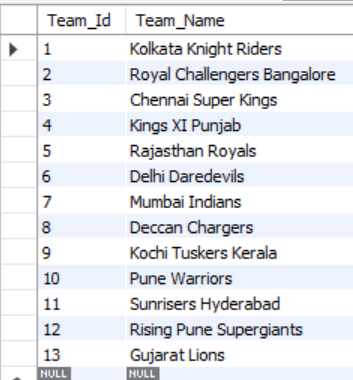
Use visual comparisons (charts showing cost vs. performance) to explain insights.

Simulate auction scenarios to ensure budget is effectively allocated.

This approach focuses on blending performance insights with financial efficiency to help RCB strengthen its team while staying competitive in the auction.

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

*There were no such entries in the “Opponent Team” column of “Match table” where “Delhi Daredevils” spelled as “Delhi Capitals”.*



*If it were present, it can be updated by the below query.*

Query

UPDATE Match

SET Opponent\_Team = REPLACE(Opponent\_Team, 'Delhi\_Capitals', 'Delhi\_Daredevils')

WHERE Opponent\_Team = 'Delhi\_Capitals';