**Objective Questions**

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

***Answer*** : Int

***Approach****:* Used Information Schema to derive the output

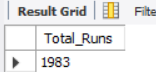
1. **What is the total number of run scored in 1st season by RCB (bonus : also include the extra runs using the extra runs table)**

***Answer*** : Total runs scored by RCB in 1st season is 1,983

***Approach****:* Created a View “ball\_bay\_ball\_seasons” wherein I have merged 3 tables “ball\_by\_ball”, “batsman\_scored”, “extra\_runs” & “seasons”and to extra the output I created 2 CTEs – 1 with the name Legal\_Runs and other with the name “Extra\_Runs\_New”, both of which provided output in numbers with the required filters. Later I added both the CTEs output to provide the final output

***Observation:*** Legal Runs was1,865 and extra runs were 118.

***Output:***



***Query:***

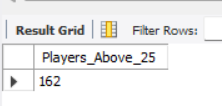
******

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

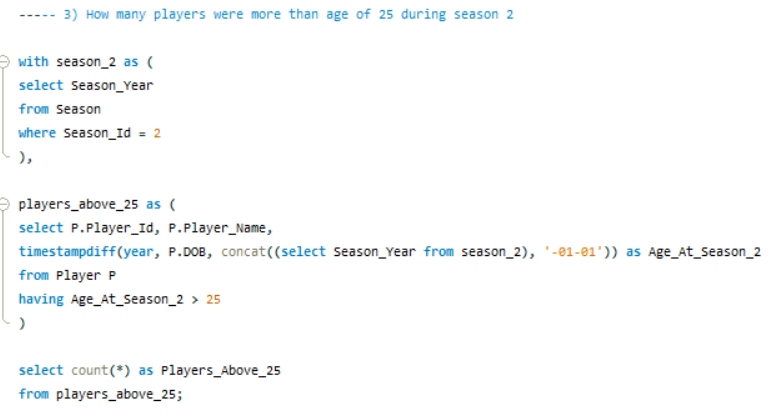
***Answer***: Total 162 players were above the age of 25 in Season 2

***Approach***: To reach the output I used the Season\_Year column to extract the year and then subtracted the Season\_Year with the DOB of the players to find the age using “Timestampdiff” function in SQL. Then used the count function to finally count the number of players falling under this category

***Output***:



***Query:***



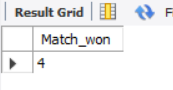
1. **How many matches did RCB win in season 1 ?**

***Answer***: RCB won 4 matches in Season 1

***Approach***: Used the Matches table to extract the output where I filtered the Season Id as 1st and then used the count function to calculate the matches won by RCB – Total matches played were 14.

***Observation***– Win% for RCB was very less in Season 1 i.e 4/14 = 29.9%

***Output:***



***Query:***

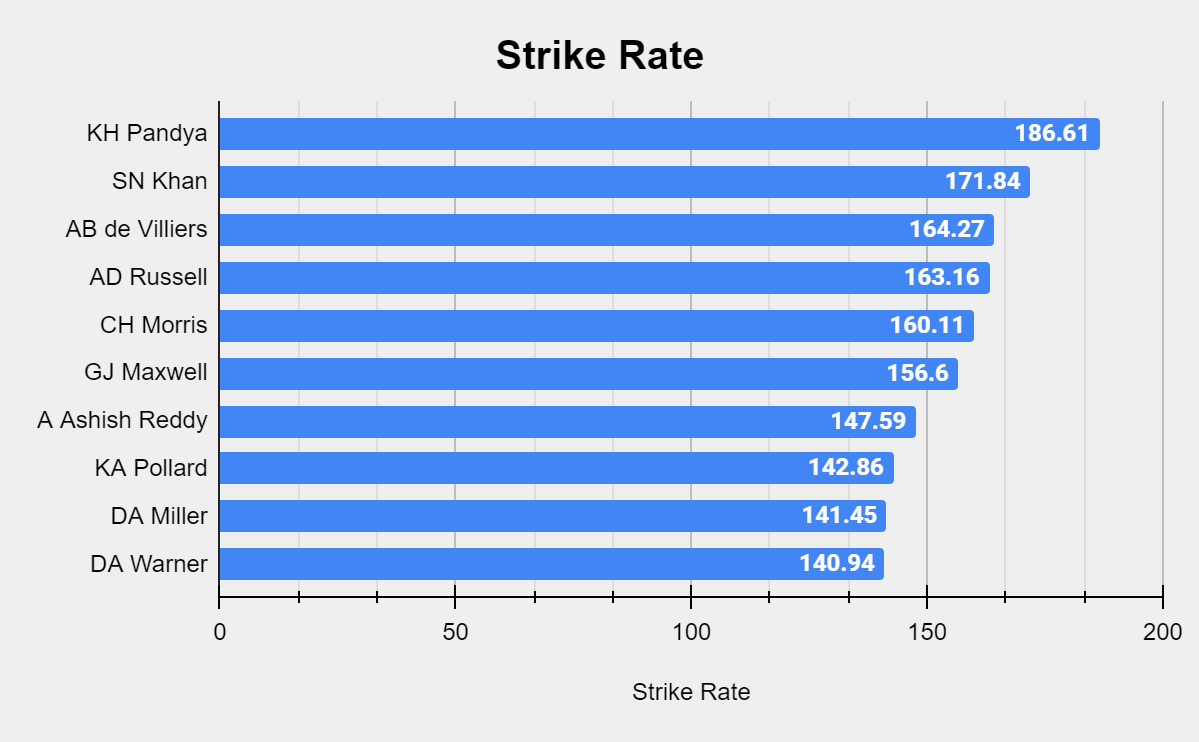


1. **List top 10 players according to their strike rate in last 4 seasons**

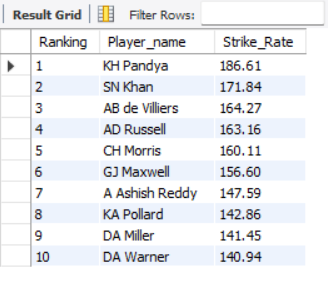
***Answer***: In below Bar Chart I have listed the top 10 players with highest striker rate

***Approach***: Used the view “ball\_bay\_ball\_seasons” and created a new CTE wherein Striker ID wise Striker Rate was plotted and then used join to connect with Player table to fetch the player name and used Rank window function to put the rank against each top strike rate and used limit to display the top 10 ranks only

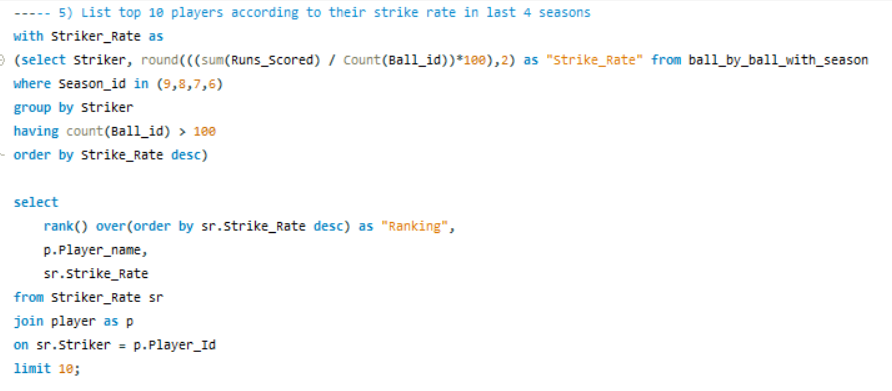
***Observation***– I have put a condition as ball\_faced > 100 so that it shows only those players who have faced more than 100 deliveries in 4 seasons combined, if this filter not added then bowlers who faced minimal balls will also reflect in the output.



**Output:**

****

**Query:**

****

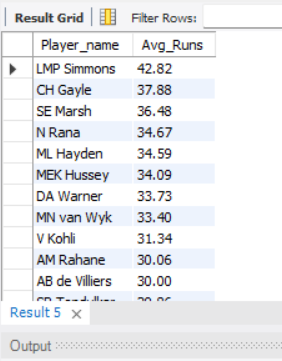
1. **What is the average runs scored by each batsman considering all the seasons?**

***Answer***: I have calculated batsman wise average run scored

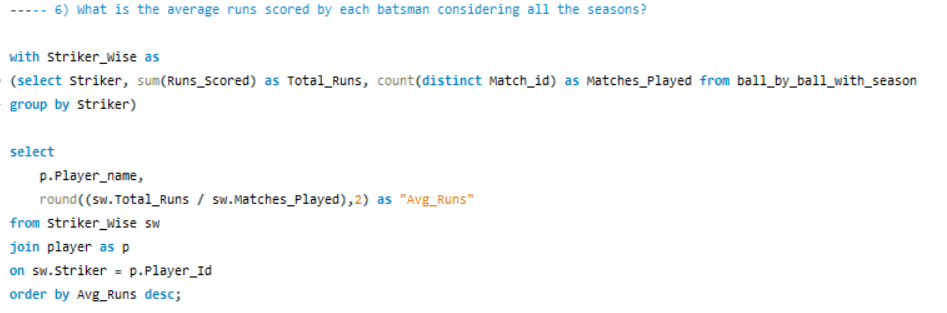
***Approach***: Used the view “ball\_bay\_ball\_seasons” and created a new CTE wherein Striker ID wise Avg Run was plotted and then used join to connect with Player table to fetch the player name and used order by to desc the records.

***Observation***– 400+ records were created

***Output***: (showing limited rows)



***Query:***

******

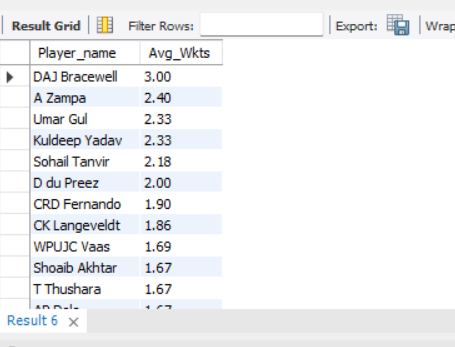
1. **What are the average wickets taken by each bowler considering all the seasons?**

***Answer***: I have calculated bowler wise average wkts scored

***Approach***: Used the view “ball\_bay\_ball\_seasons” and created a new CTE wherein Bowler ID wise Avg Wkts was plotted and then used join to connect with Player table to fetch the player name and used order by to desc the records.

***Observation***– 400+ records were created

***Output***: (showing limited rows)



***Query***:



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

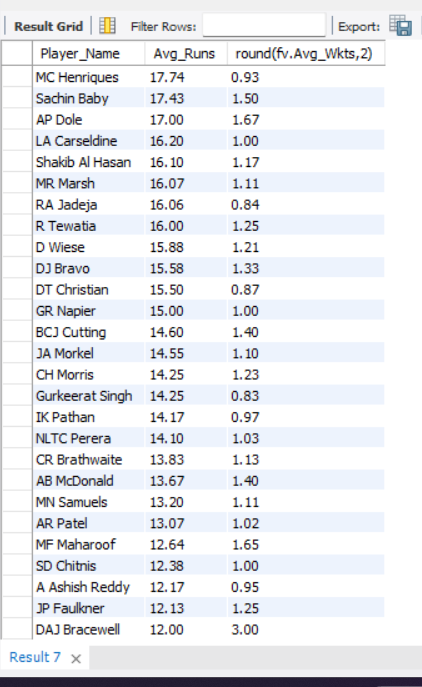
***Answer***: There are 41 such all-rounders whose

***Approach***: First I calculated Batsman wise Avg\_Runs scored under CTE Batsman\_Wise and then calculated Bowler wise Avg\_Wkts taken under CTE Bowler\_Wise. Then took the avg as Overall\_Avg\_Runs & Overall\_Avg\_Wkts respectively. Then joined both the CTEs to get the allrounders details whose Avg\_Runs > Overall\_Avg\_Runs and Avg\_Wkts > Overall\_Avg\_Wkts

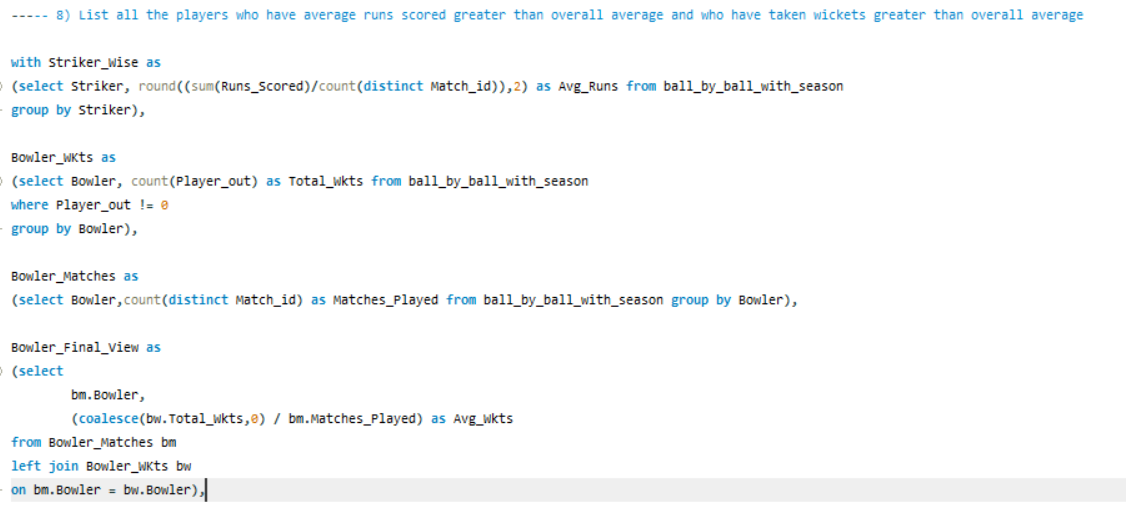
***Observation***– In this list I have used AND condition hence only all rounders are featuring here.

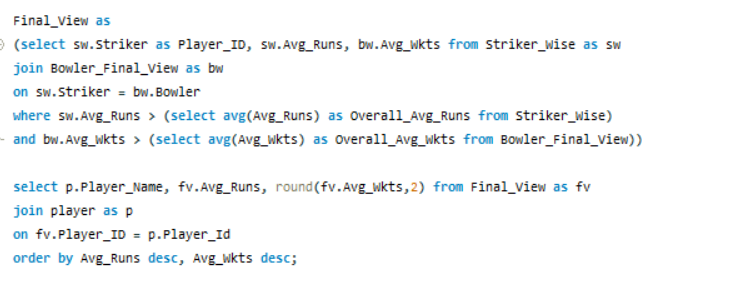
****

***Output:*** (showing limited rows)

******

***Query:***

******

******

***:***

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

***Answer***: Created a table named Venue\_Match\_Stats which shows Venue wise Matches Won and Lost by RCB across all seasons

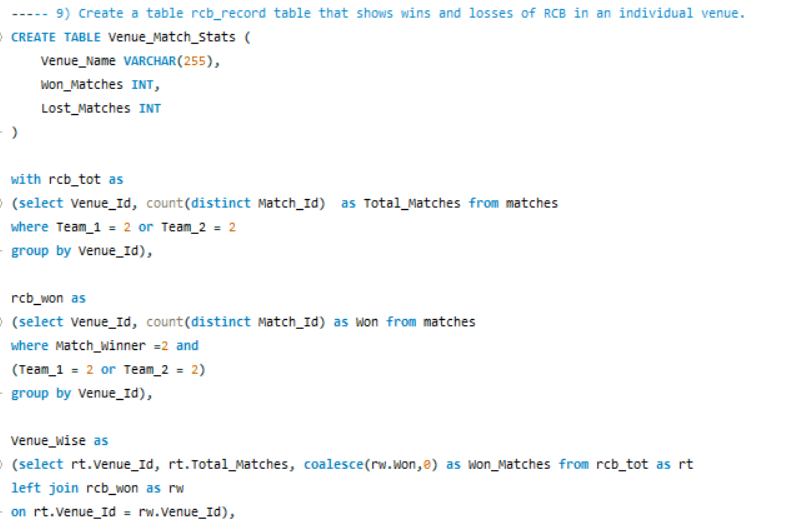
***Approach***: Calculated matches won in one view and then total matches played against each Venue and then used them in one CTE to calculate the matches lost.

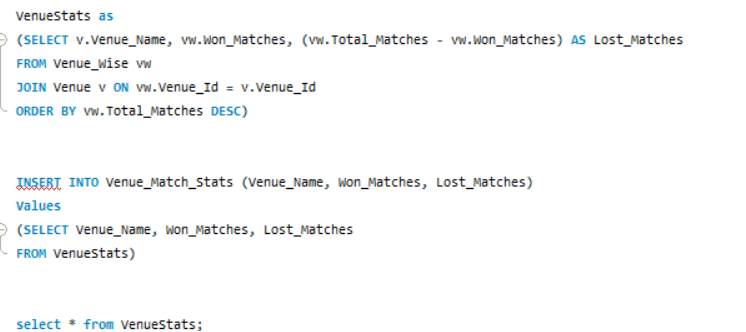
***Observation***– Below graph shows the Win% across the Venue, home ground has less win% but has the most numbers of matches played approx. 66 matches with a win% of 44 (1st from left in below chart).My SQL version is not accepting the Insert query but my syntax is proper.

***Output:***

****

**Query:**

****

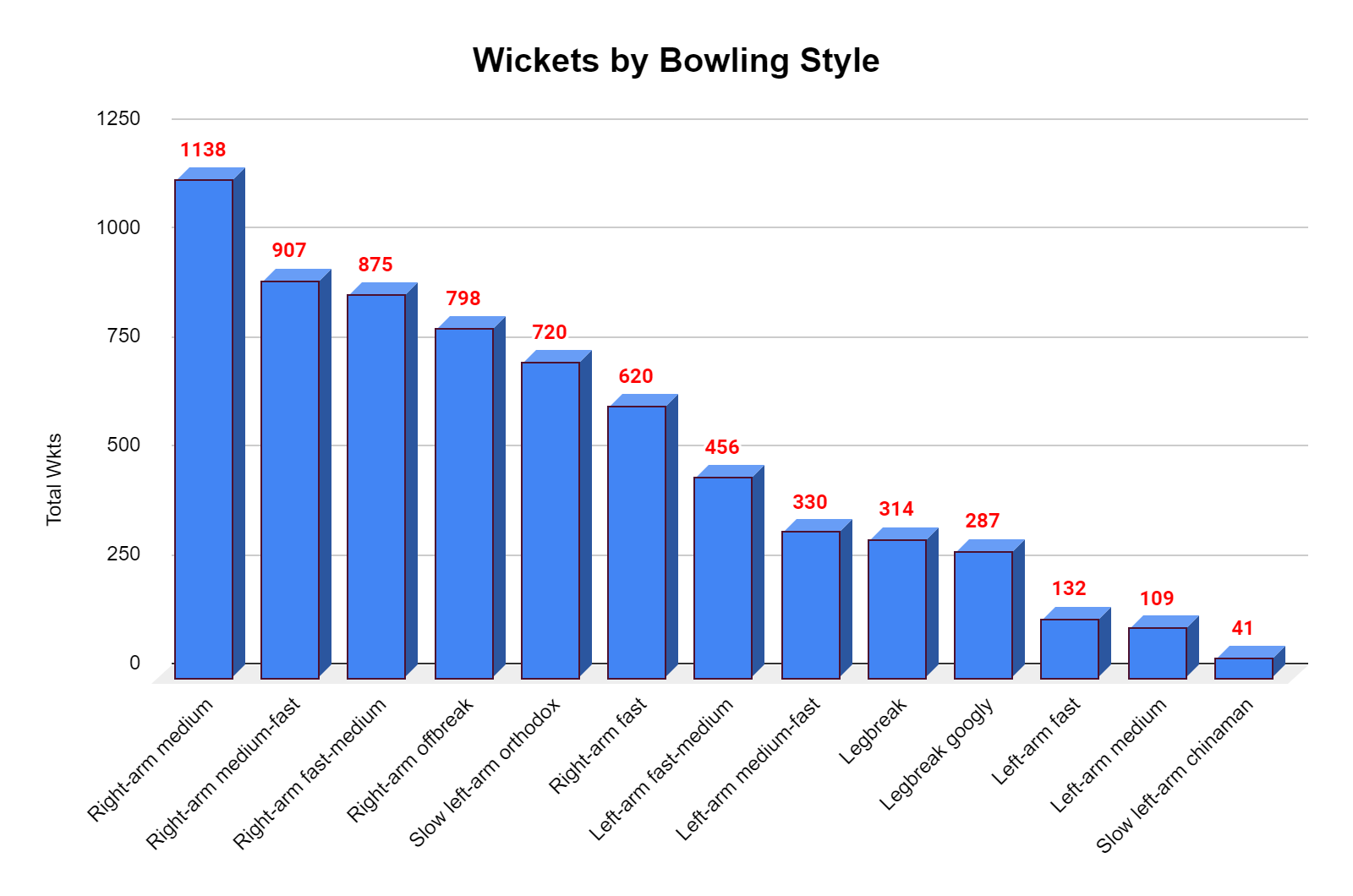
****

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

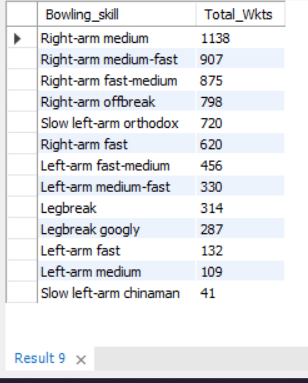
***Answer***: Right-arm Medium is the bowling style RCB should focus on more as this style of bowling has the most number of wickets.

***Approach***: Used the view “ball\_by\_ball\_with\_season” to calculate bowler wise total wickets taken and then used join to connect it with the Bowling style table.

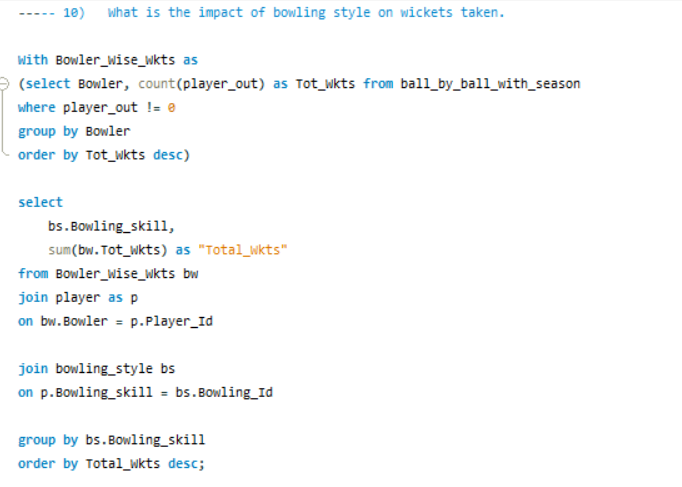
***Observation***– Below column chart shows that Right arm bowlers have dominance over the left arm bowlers.



***Output:***

******

***Query:***

******

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**

***Answer***: There has been 5 better performance year for RCB where the current season’s Runs and Wkts were more than the previous year and 3 Not Good performance years. In below table season wise understanding is shared

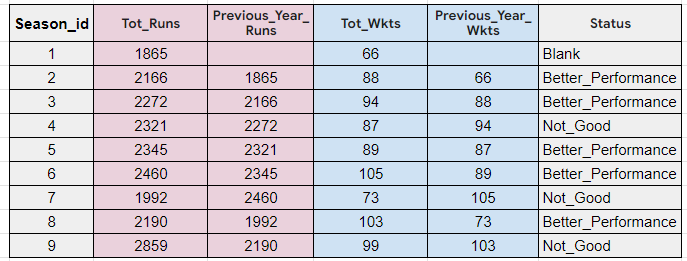
***Approach***: Created 2 CTEs with LAG function to pull the previous year data.

\* Run\_Chart which shows Season wise RCBs – Total Runs and Previous Years Run,

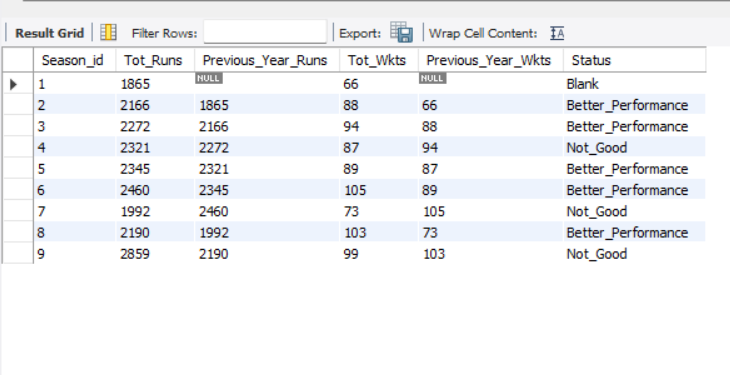
\* Bowl\_Chat which shows Season wise RCBs – Total Wkts and Previous Years Wkts

Used Join to connect both these CTEs and then used Case function to derive the “Status”

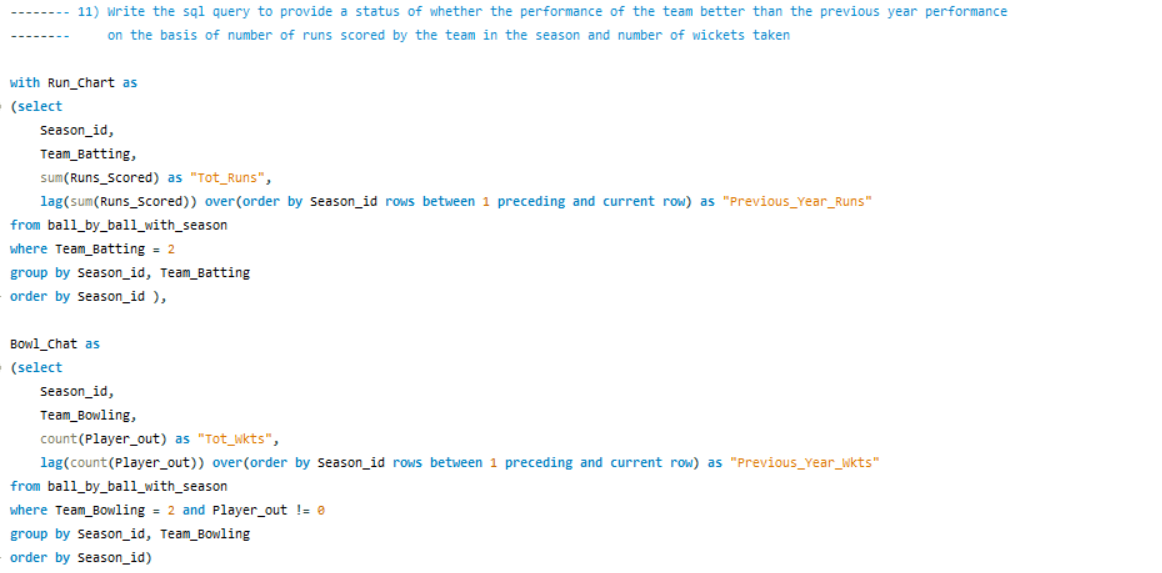
***Observation***– Most of the years has been better performing years

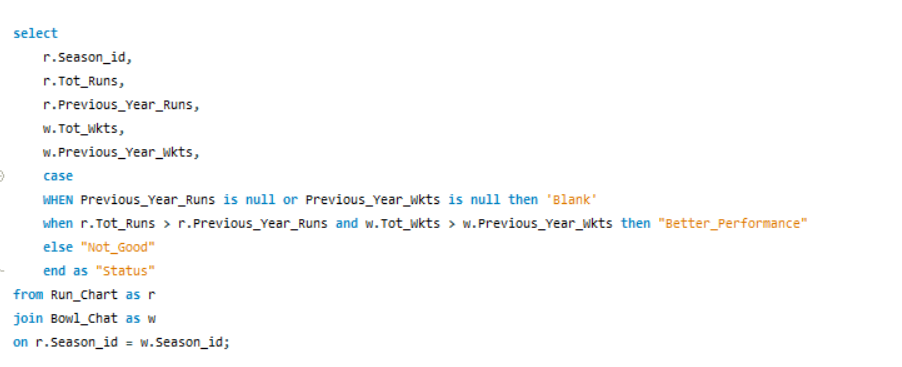
****

***Output:***

****

***Query*:**

****

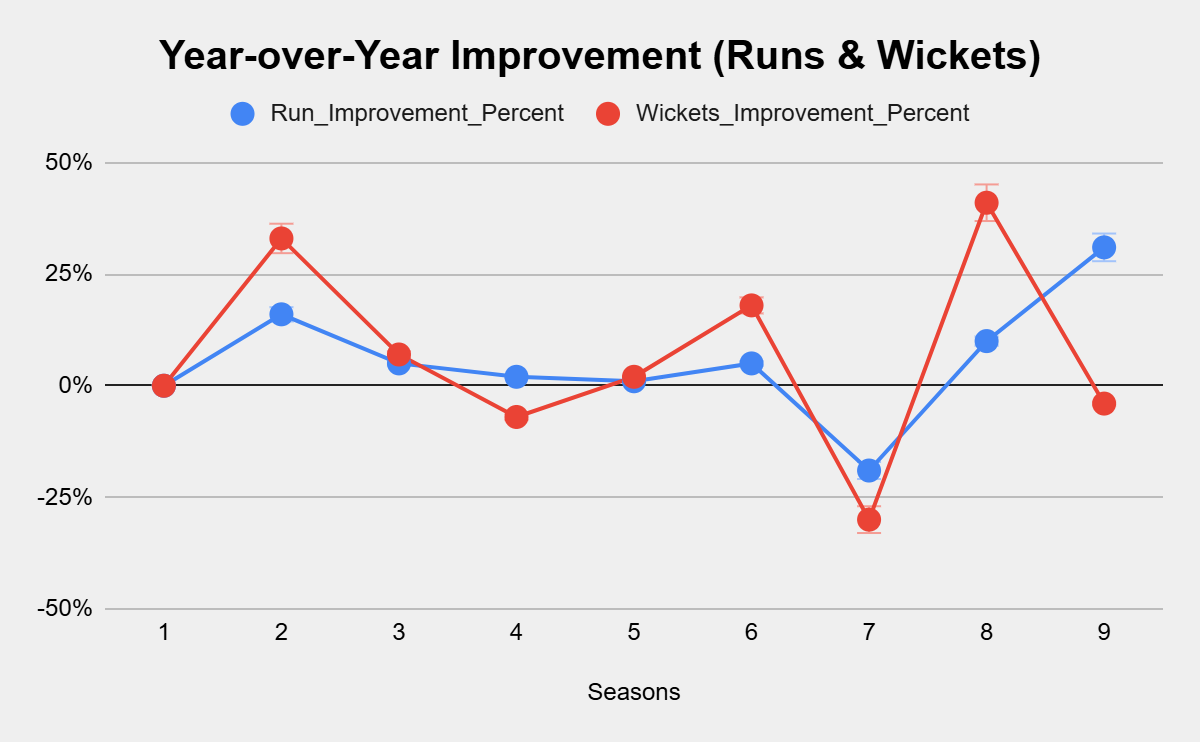
****

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

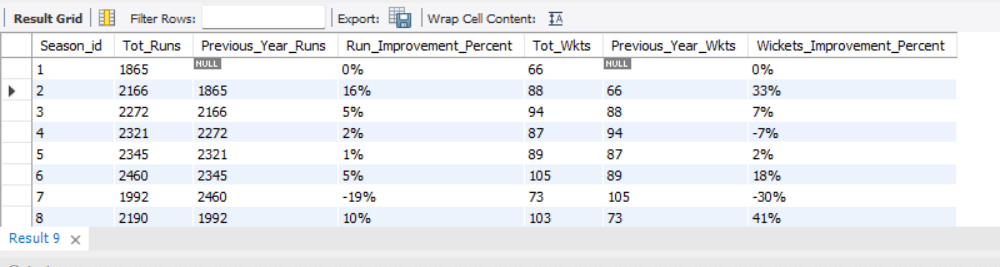
***Answer:***

1. ***Year-over-Year Improvement (Runs & Wickets)***

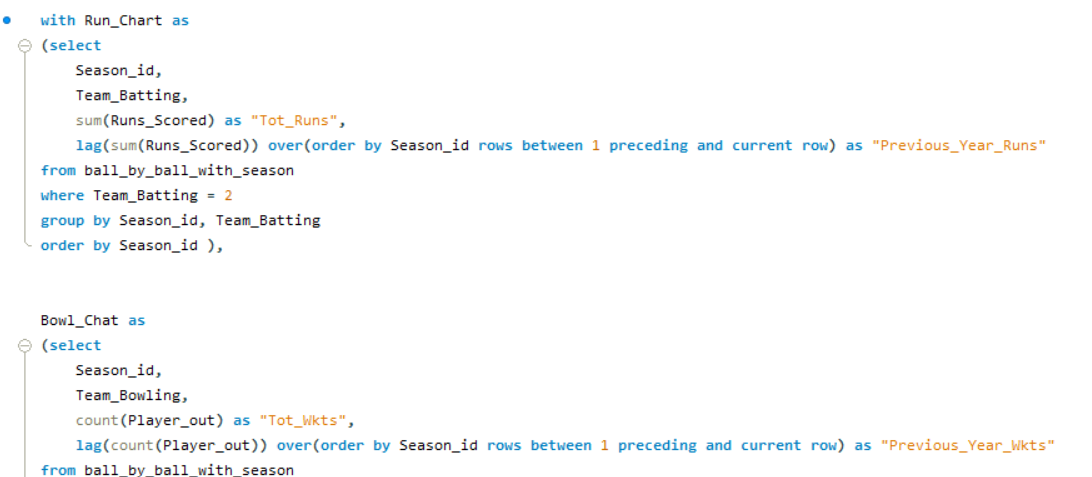
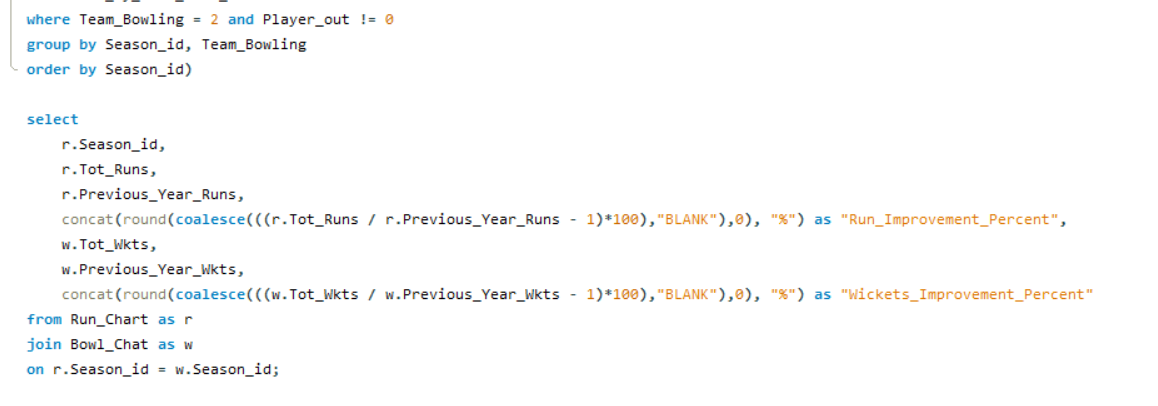
***Observation:*** Below Line chart shows season over season comparison for the run scored and wicket taken percentage growth or de-growth, the numbers in negative means we performed bad in that year as compared to last year



**Output:**

****

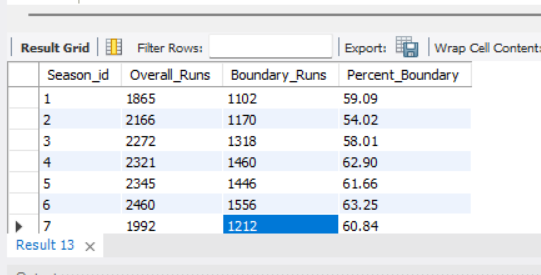
**Query:**

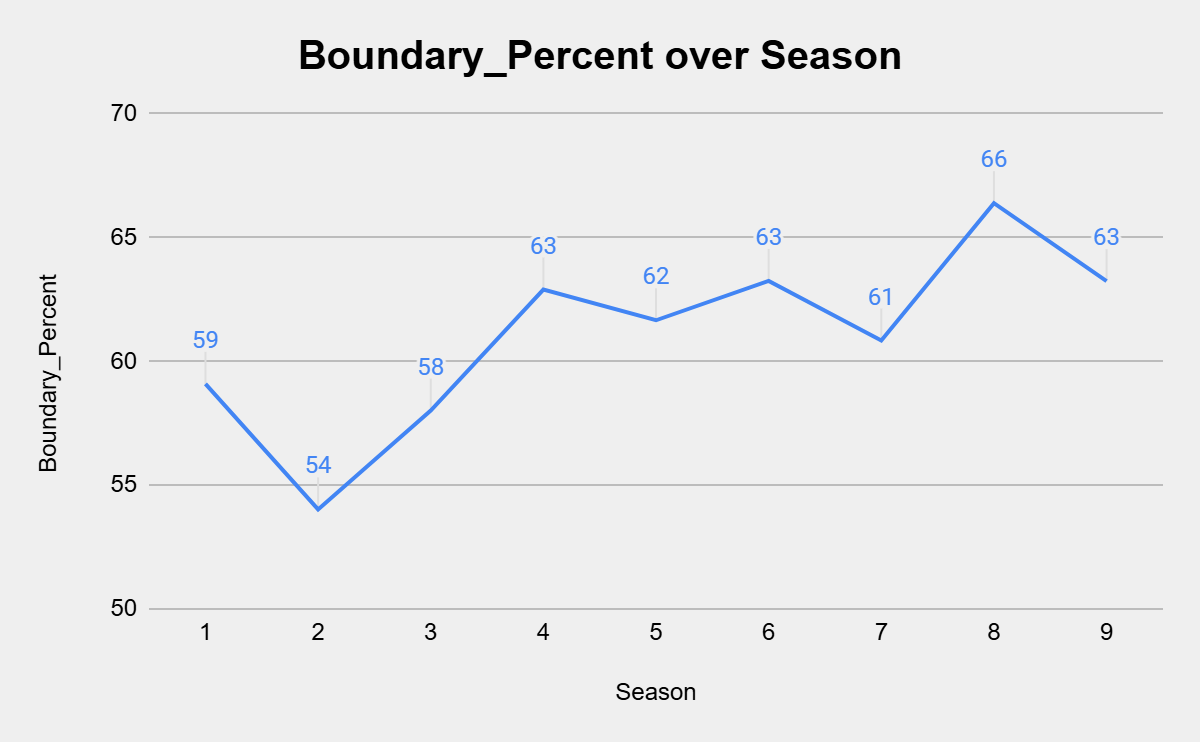
 

1. ***Boundary Performance Metrics***

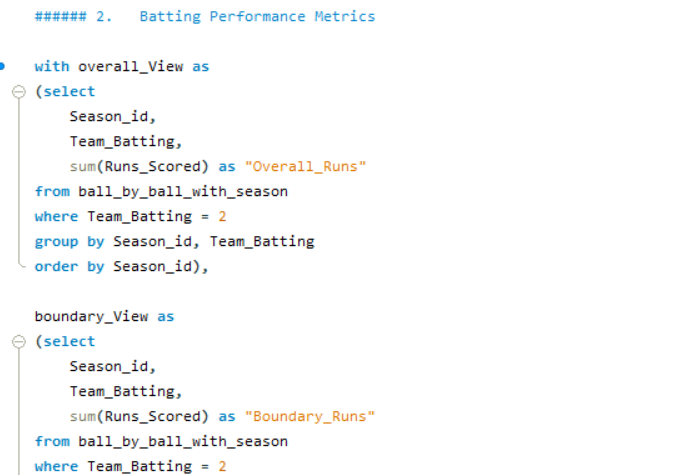
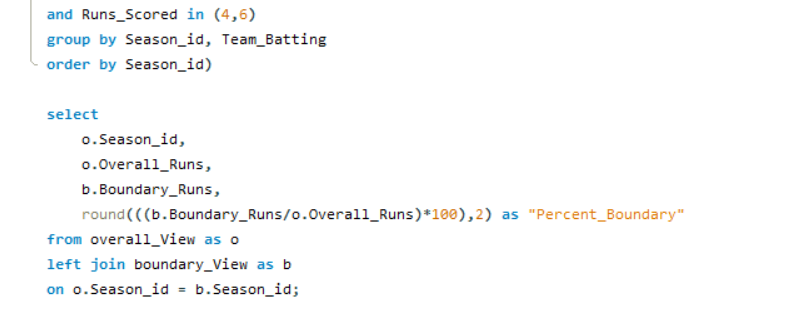
***Observation:*** Boundary Performance metrics shows the percentage of boundary contribution in overall runs scored, From Season 4 to Season 9 – our Boundary percent has improved and is hovering around 60%+. This shows the attacking approach of the team which is much required in T20 format.

***Output:***

******



***Query:***

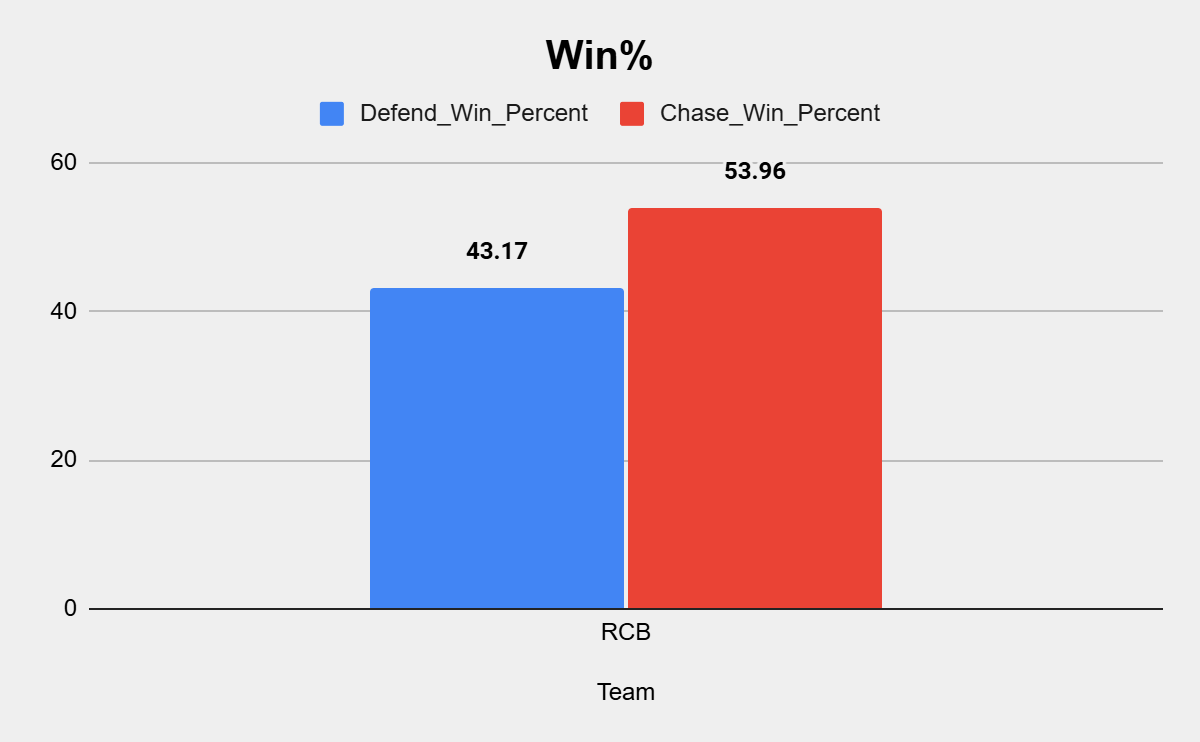
****** ******

**3. Match Impact Metrics**

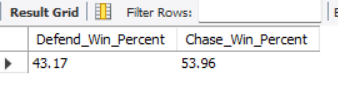
Win % in Chasing: Measures how successful the team is when chasing targets.

Win % in Defending: Measures success in defending targets.

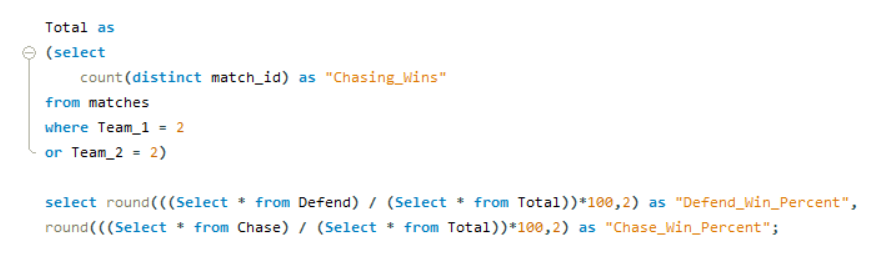
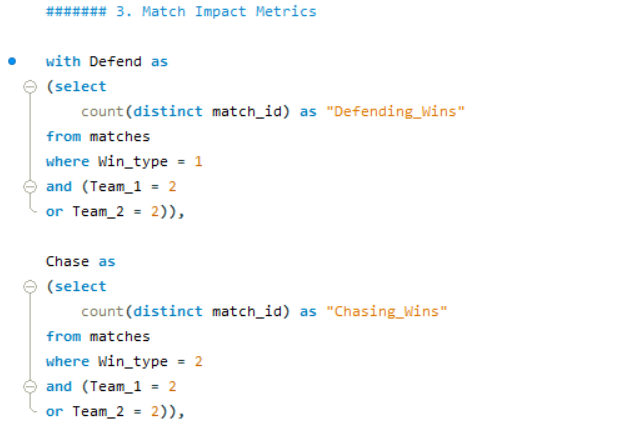
**Observation:** Team has won most of the matches by chasing.



**Output:**



**Query:**

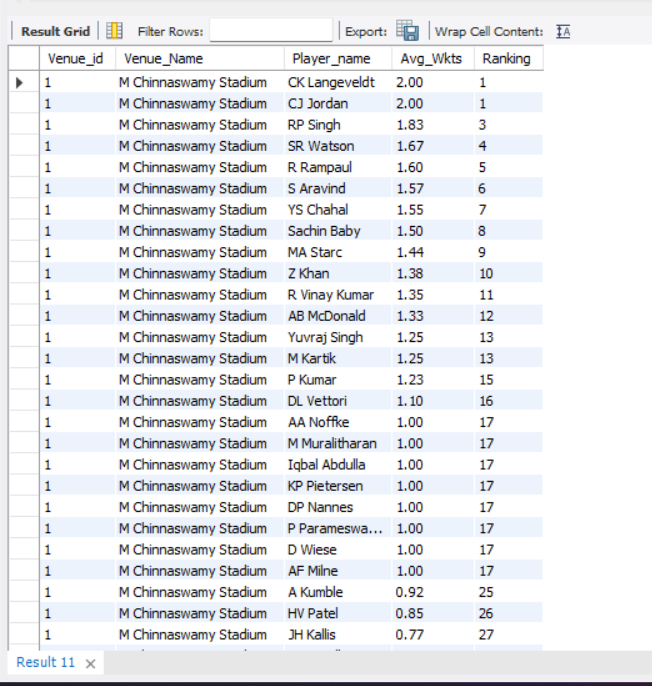
****

1. **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.**

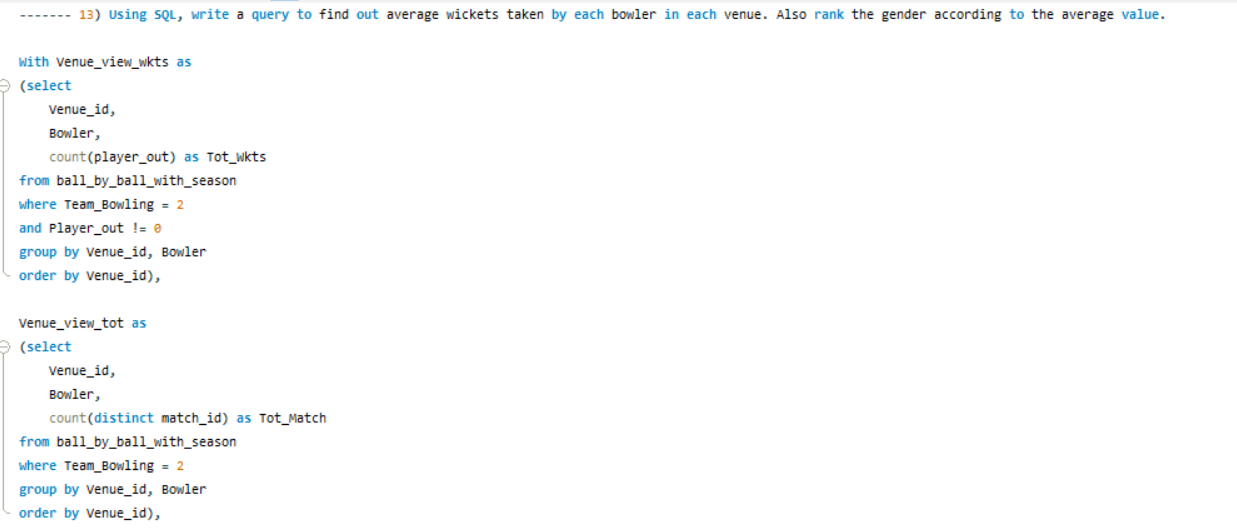
***Answer***: The query has been mentioned in the SQL file.

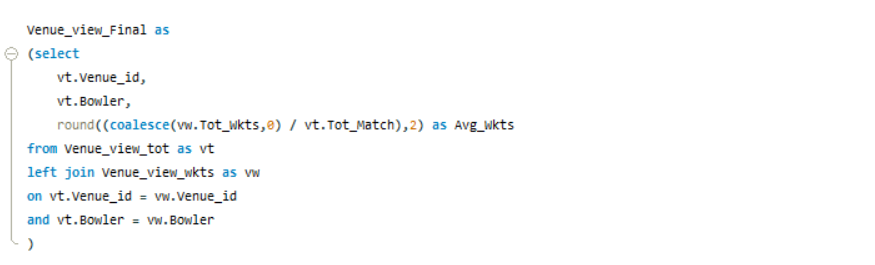
***Approach***: First found the Bowler & Venue Wise average wkts taken then used join to merge Venue and Player details and then used Rank in the final query to output the data Venue wise – Bowler avg\_wkts desc ranking

**Output:** (Showing limited rows)

****

***Query:***

****

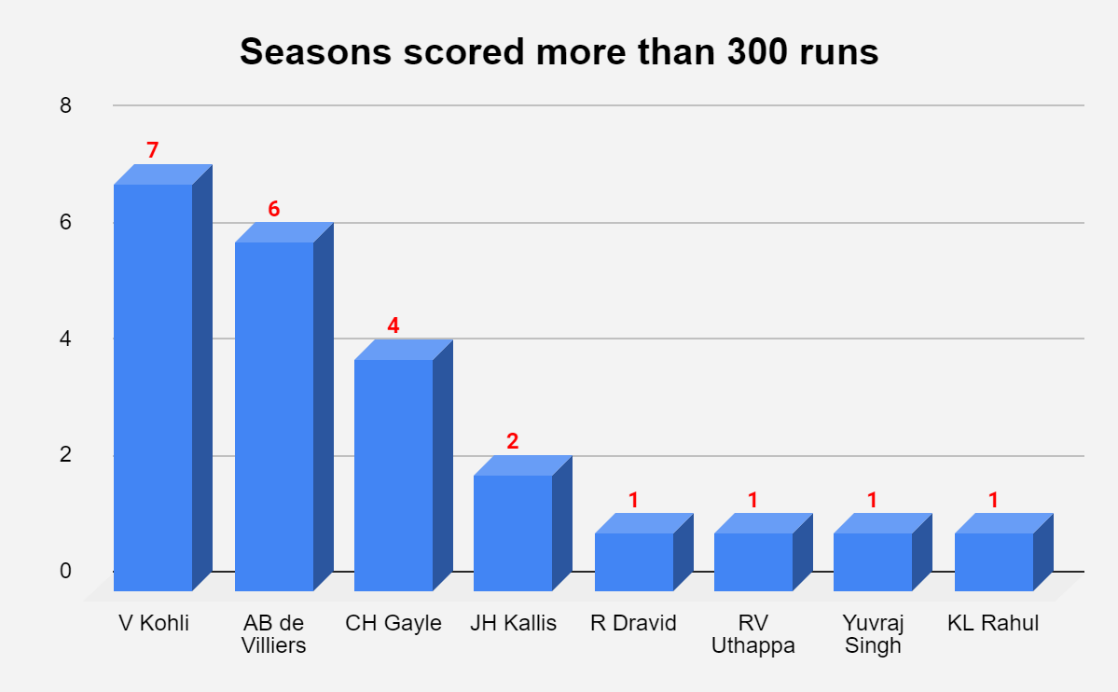
** **

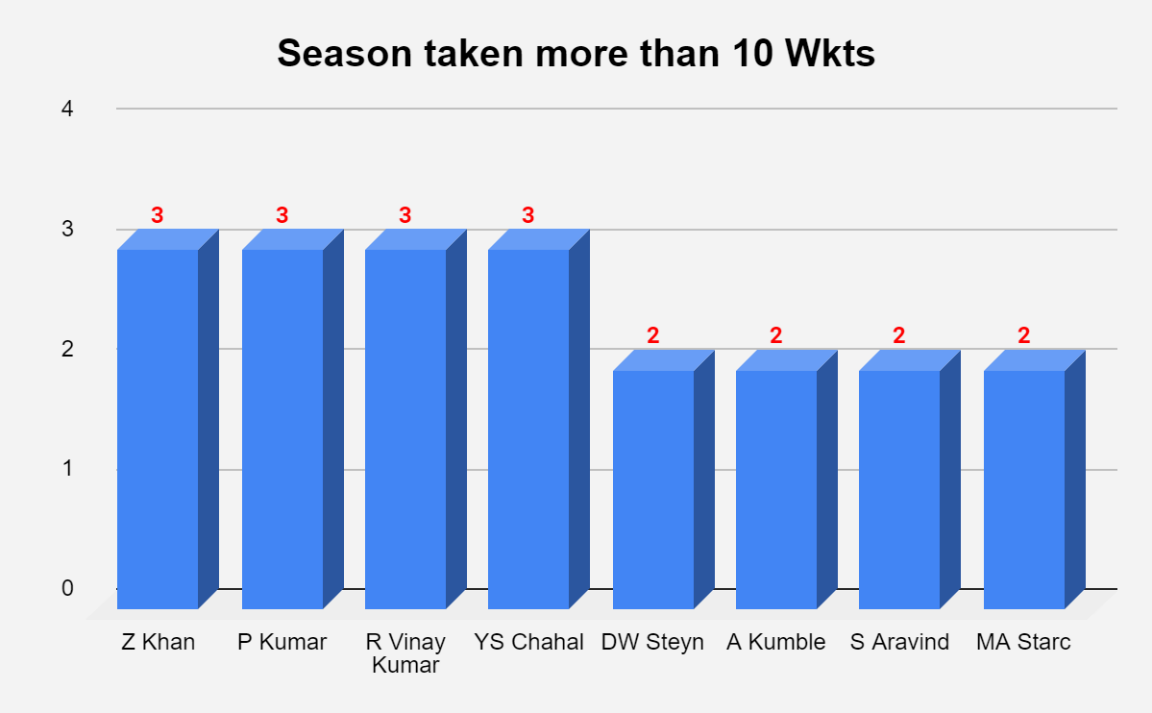
1. **Which of the given players have consistently performed well in past seasons? (will you use any visualisation to solve the problem)**

***Answer***: Below chart represents the answer, V Kohil and AB de Villers from batsman POV and Z Khan, P.Kumar and R Vinay Kumar from bowler’s POV

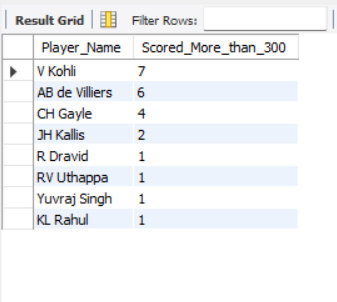
***Approach***: Created 2 CTEs one for the bowler and one for the batsman which computed the “Scored\_More\_than\_300” and “Taken\_more\_than\_10Wkts” which shows the count of Seasons where they have met the necessary conditions.

***Observation***– There are number of players who have achieved this milestone once or twice but I have selected the players who have consistently outperformed in most of the seasons.





***Output***



***Query:***

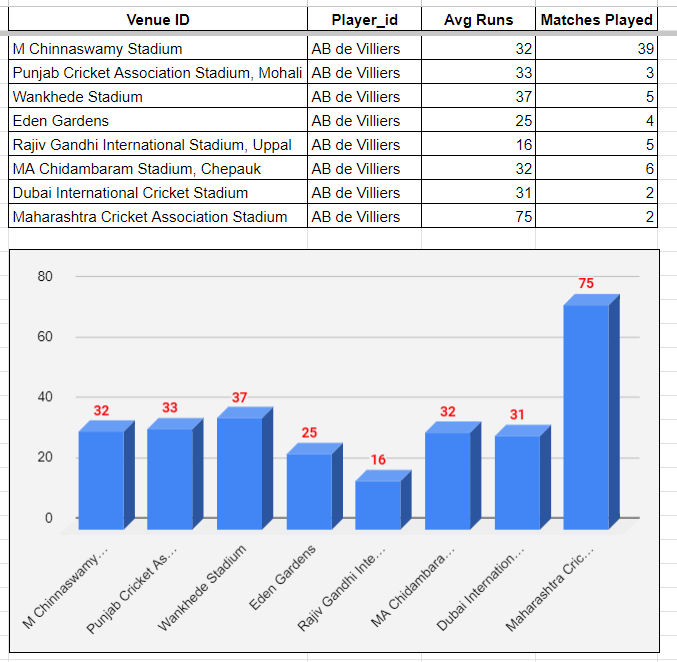
 

1. **Are there players whose performance is more suited to specific venues or conditions? (how would you present this using charts?)**

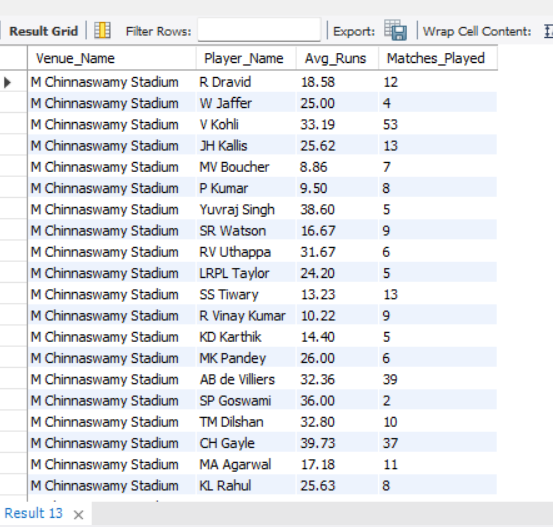
***Answer***: After viewing the entire data – I have found AB de Villiers to be one of the batsman who is more suited to a specific venue i.e MCA Stadium as shown in the below chart out of 2 matches played he averages around 75

***Approach***: Took Venue & Striker wise Avg runs scored for RCB players with condition that more than 50 balls should have been played

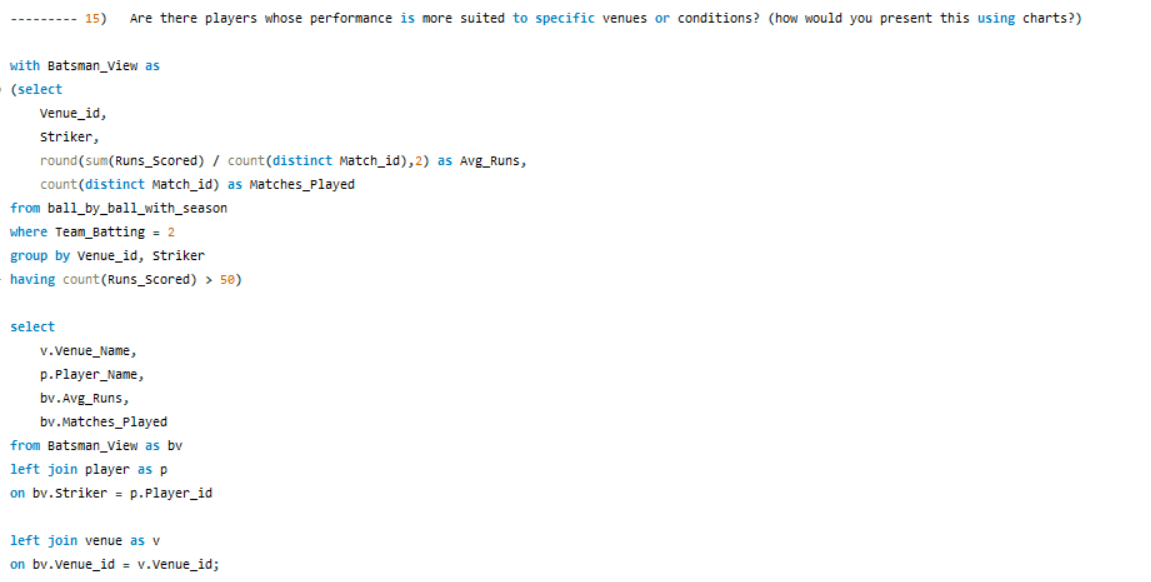
***Observation***– Rest all batsmen have almost similar average across venues.

****

***Output:*** (Showing limited rows)

******

***Query:***

******

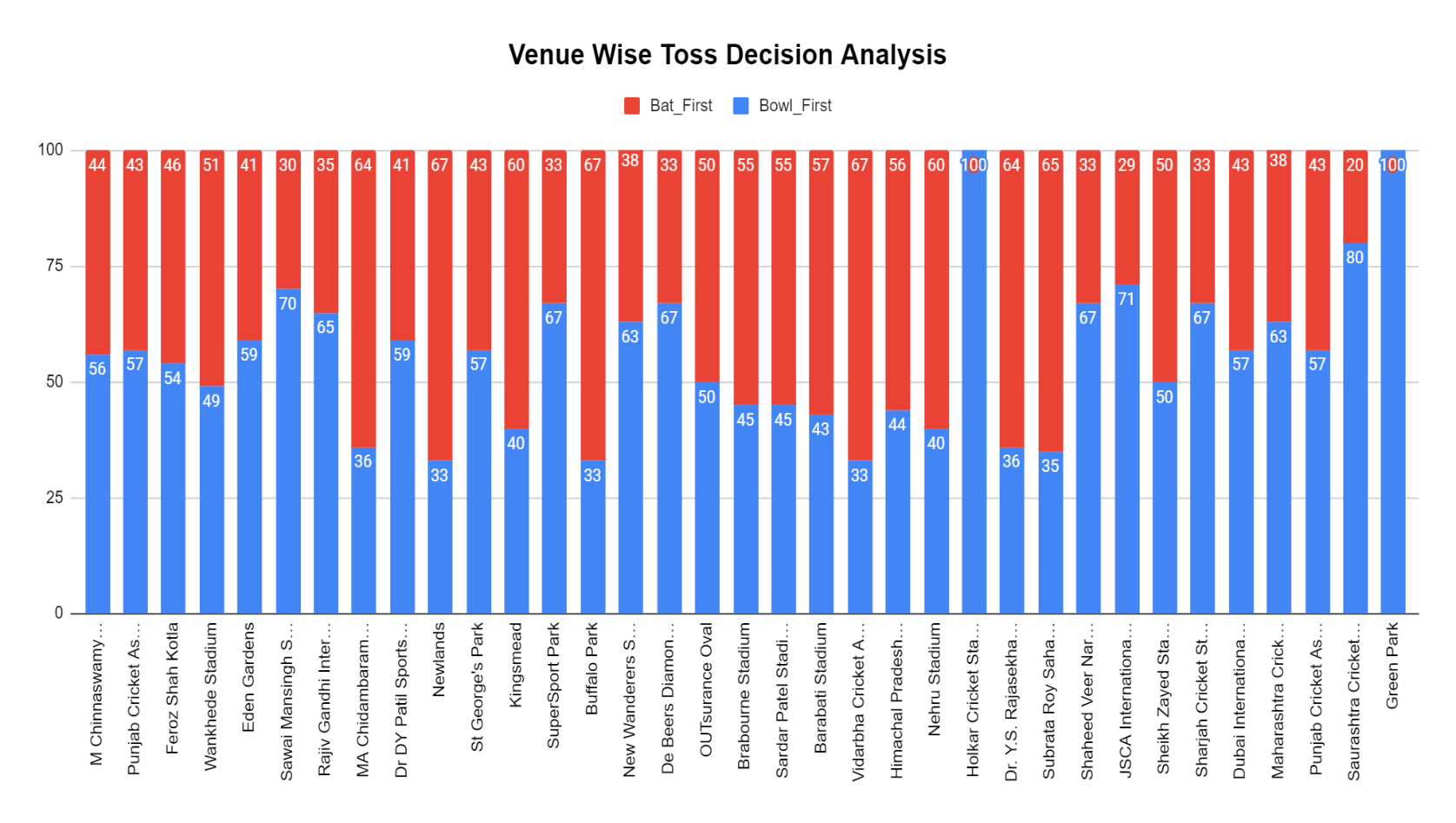
**Subjective Questions**

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?**

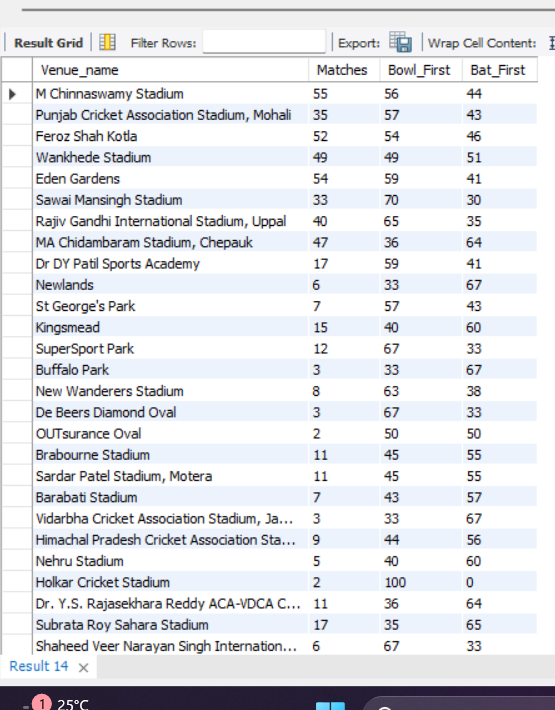
***Answer***: 50.43% of the matches are won by the team who wins the toss

***Approach***: There are 2 approaches used to answer this question – 1st is the Venue wise Bat\_First and Boiwl\_First win% which will help us understand based on each venue what should be our selection if we win the toss. 2nd is the Toss Win Match Win percent to understand the overall win% if teams win the toss.

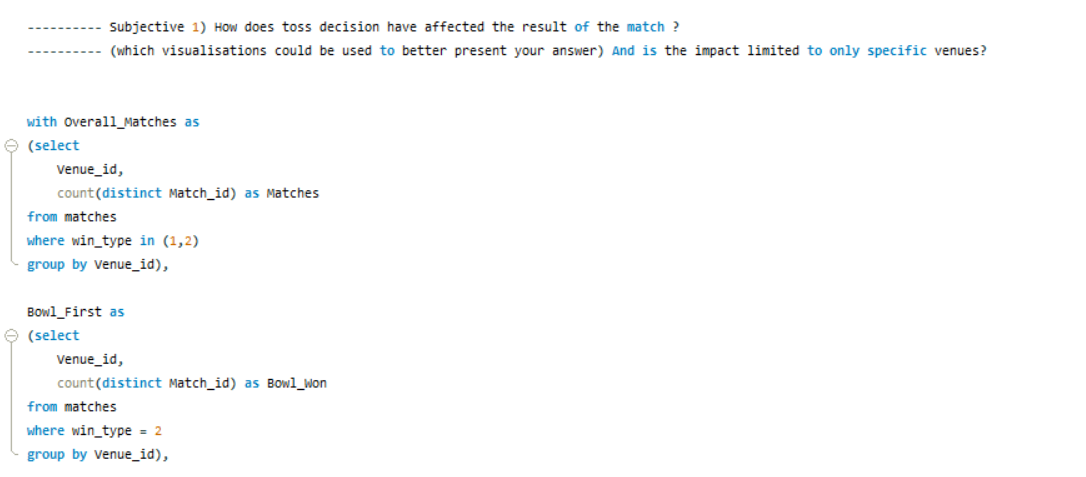
***Observation***– Below chart shows venue wise what should be our selection post winning the toss.

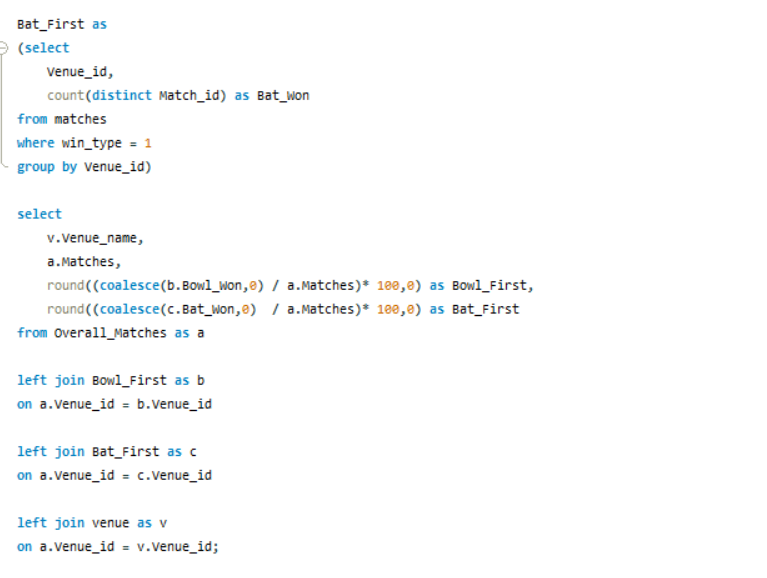


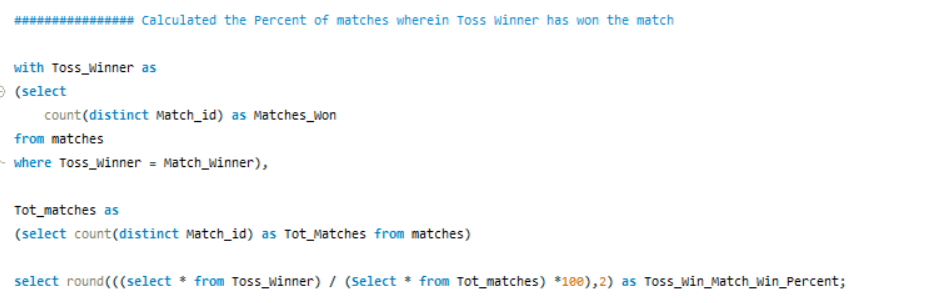
**Output:**

****

**Query:**

****

****

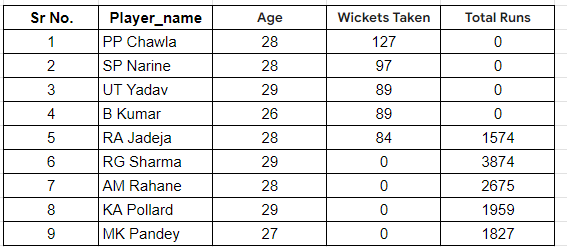
****

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

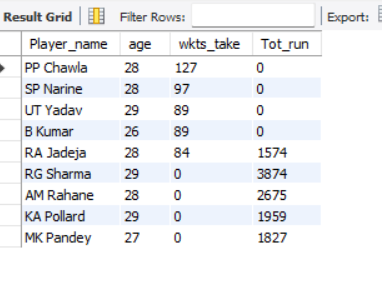
***Answer***: Young Highest Run\_Scorer and Highest Wkt\_Taker should be the players who would be best fit for the team

***Approach***: First I identified the young player by keeping the age criteria < 30 and the found the players who are currently not playing in RCB and found the Highest Run getter (Runs > 1500) and Wickets taker (Wkts > 80) of the tournament across all seasons and below is the table of players who should be in our to-get list

***Observation***– RA Jadeja is the only all-rounder who is performing best in both domains.

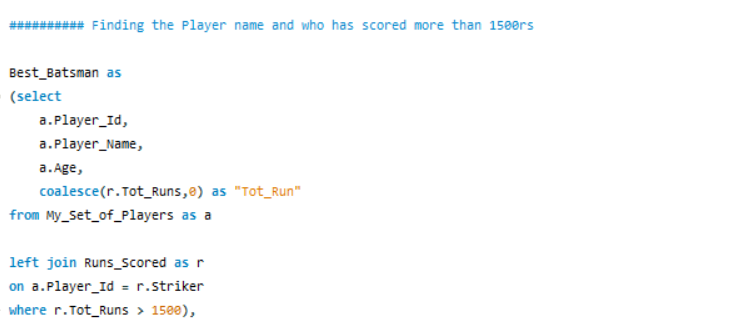
****

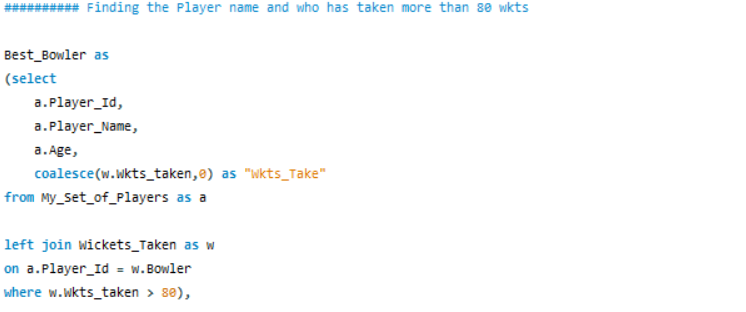
***Output:***

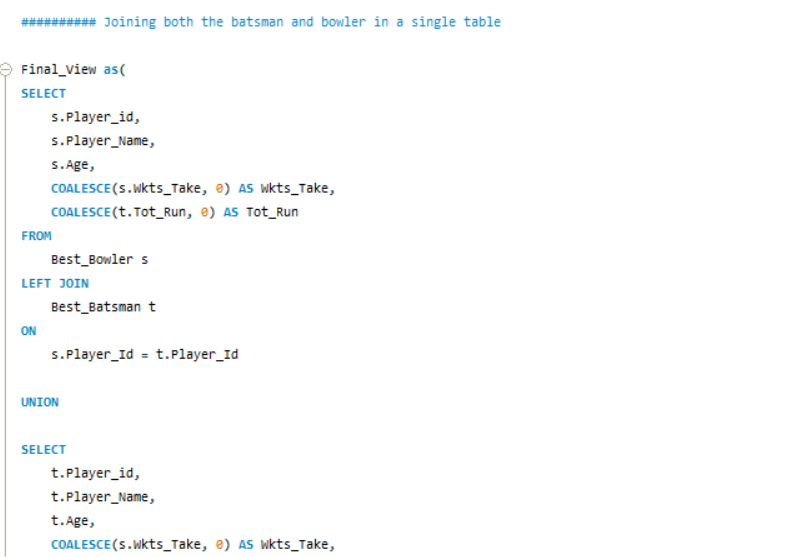
******

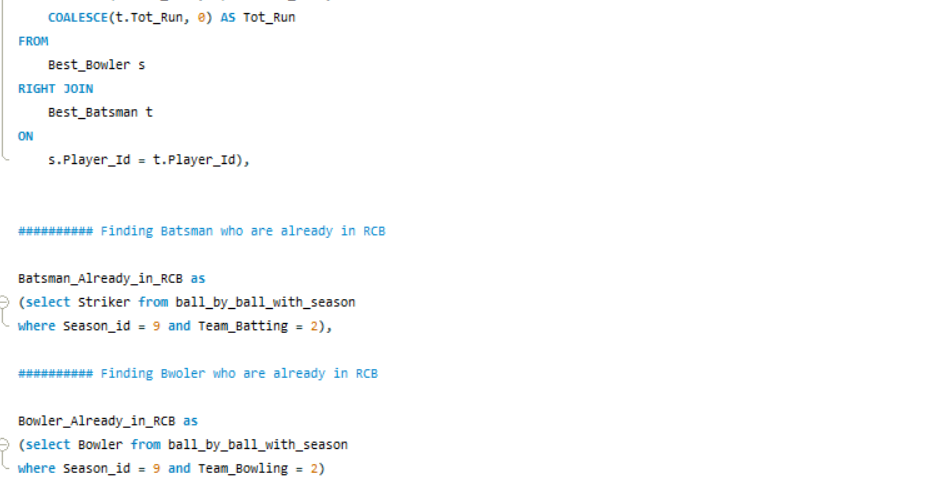
***Query:***

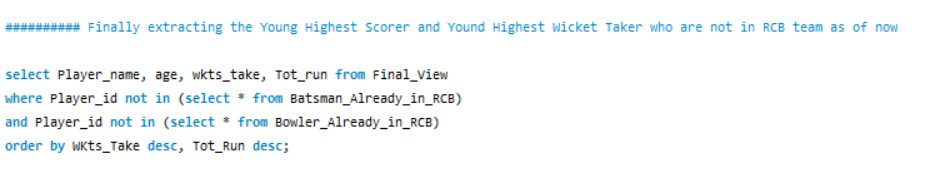
******

******

******

******

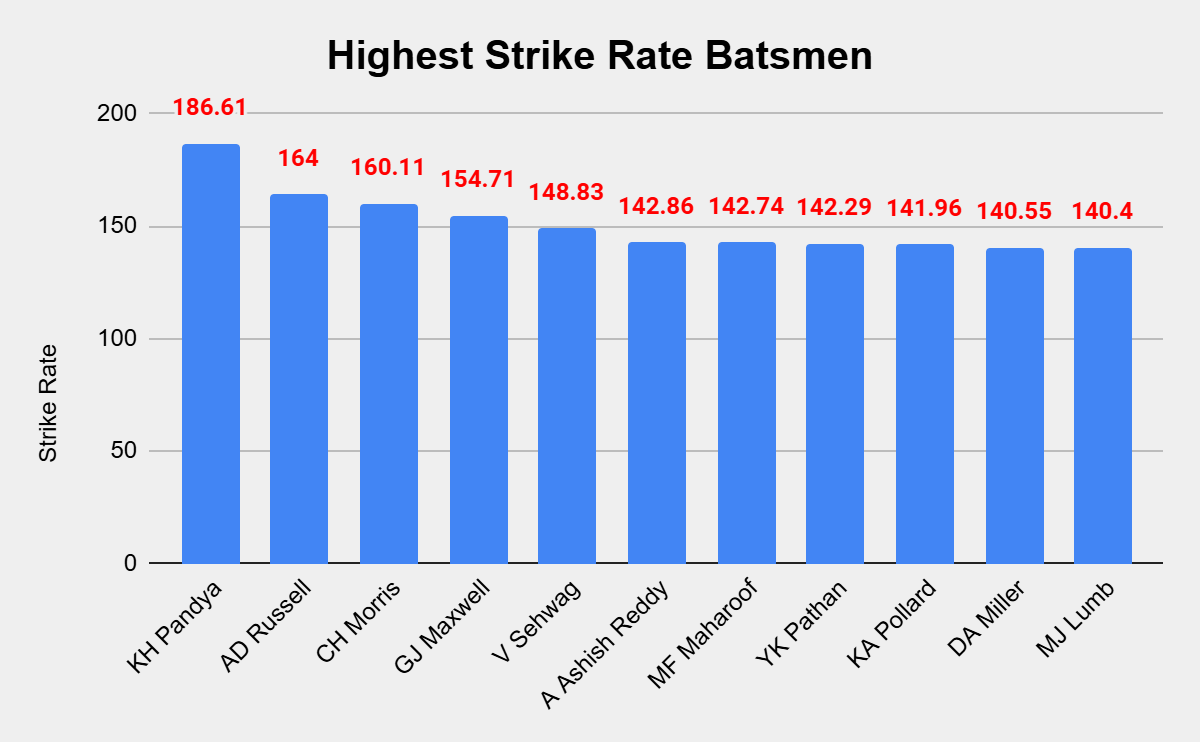
******

******

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

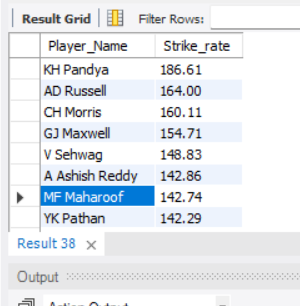
***Answer*:**

* Batsmen with High Strike Rates:

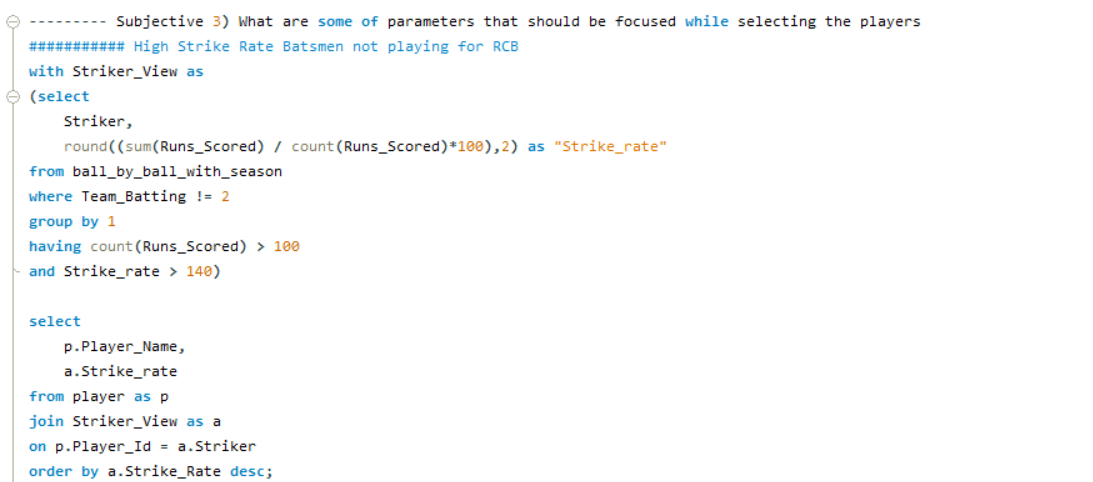


***Observation:*** Above chart shows the player’s name who should be looked out for, as they posses playing this game with high strike rates and these players are not playing for RCB currently.

**Output:**



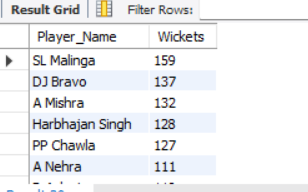
***Query:***

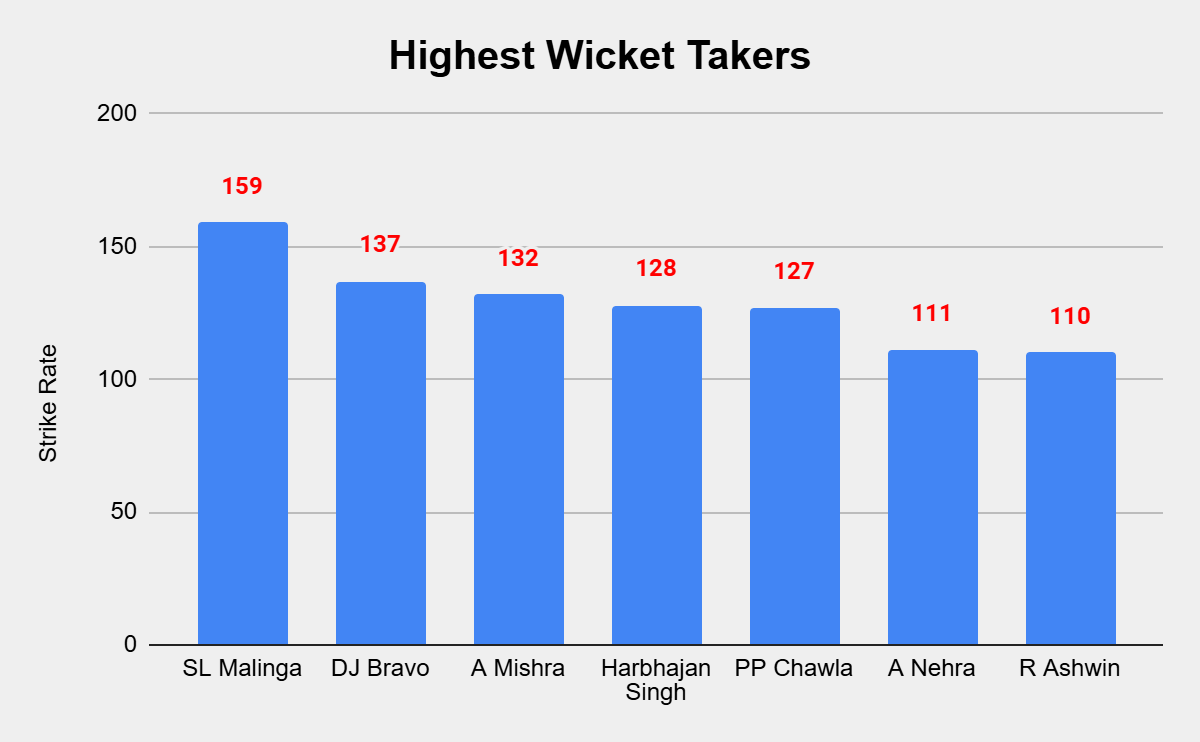


* Bowlers with Most Wickets in this Tournament:

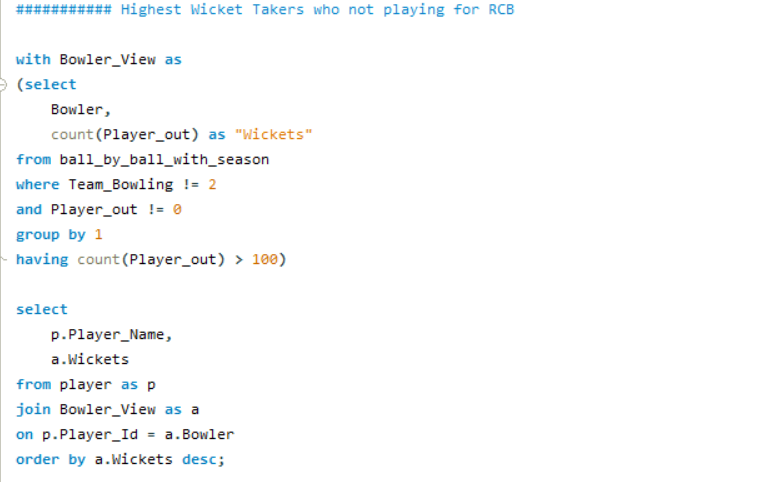
***Observation:*** These are the recommended list of bowlers who are currently not playing for RCB but are having highest wicket taking ability in this tournament.

***Output:***





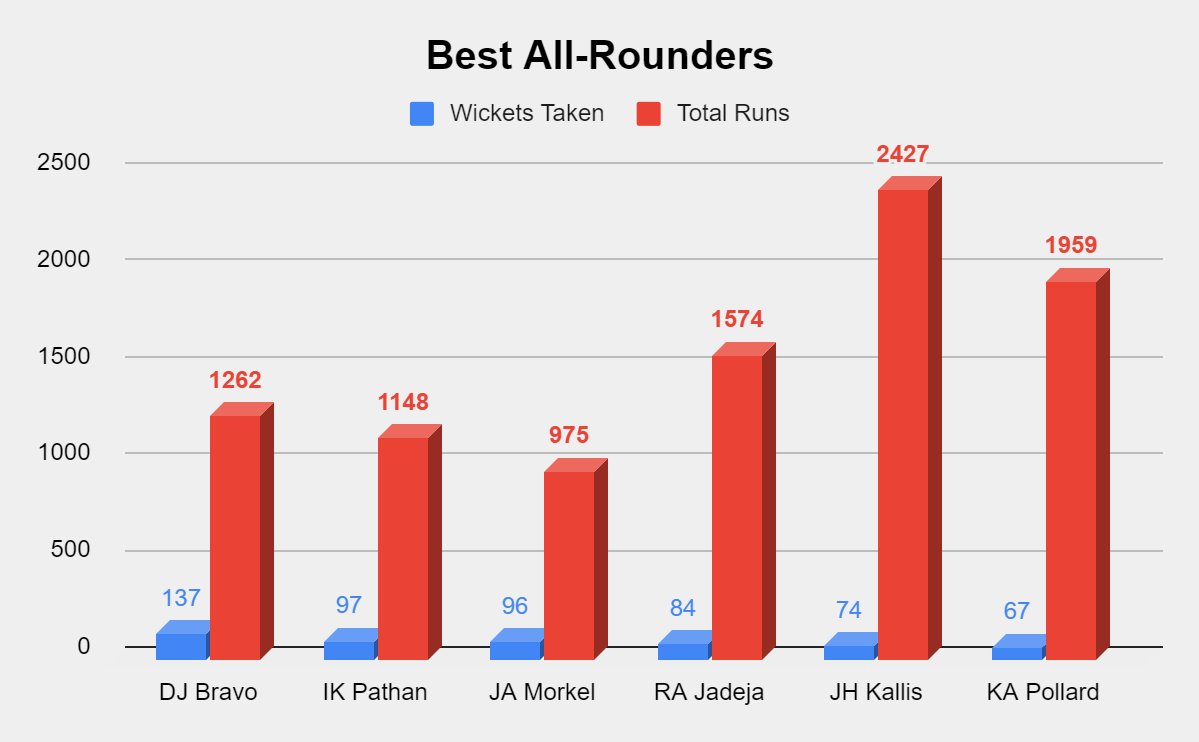
***Query:***



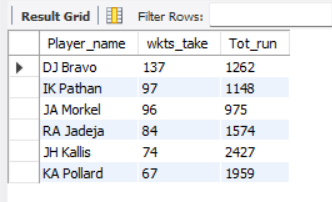
1. **Which players offer versatility in their skills and can contribute effectively with both bat and ball? (can you visualize the data for the same)**

***Answer***: The players who have good batting runs and good number of wickets in this tournament needs to be considered here

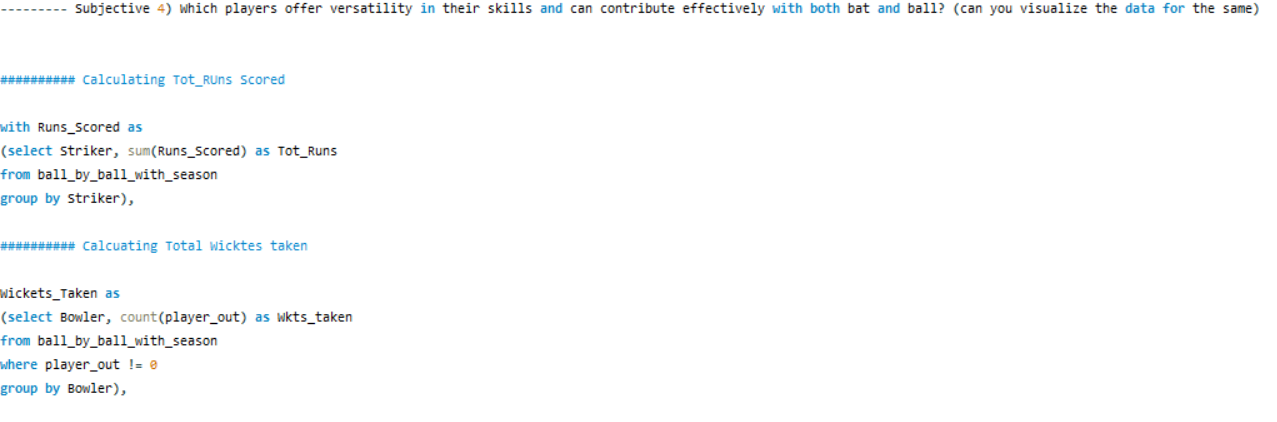
***Approach***: First I found the players who are currently not playing in RCB and found the Highest Run getter (Runs > 800) and Wickets taker of the tournament (Wkts > 50) across all seasons and below is the table of players who should be in our to-get list

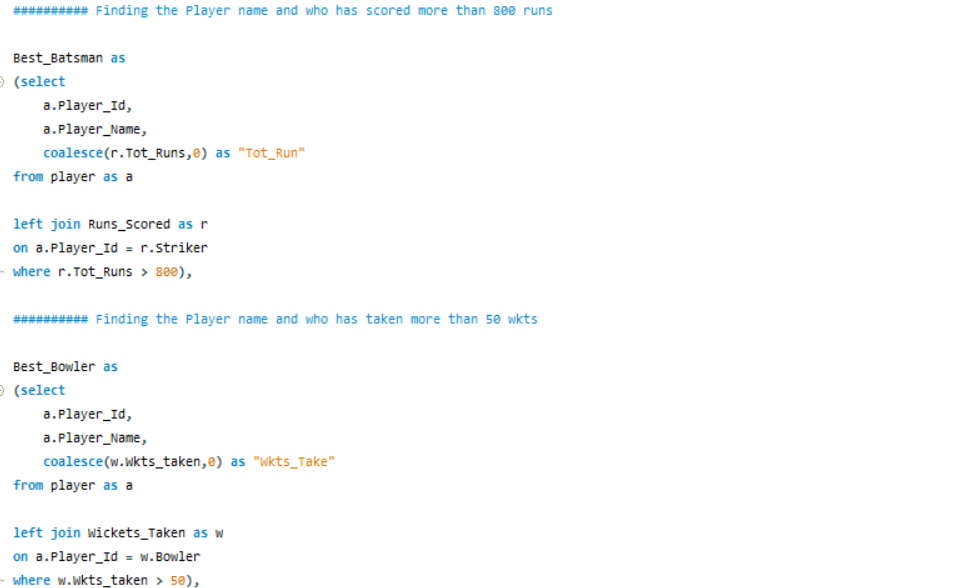


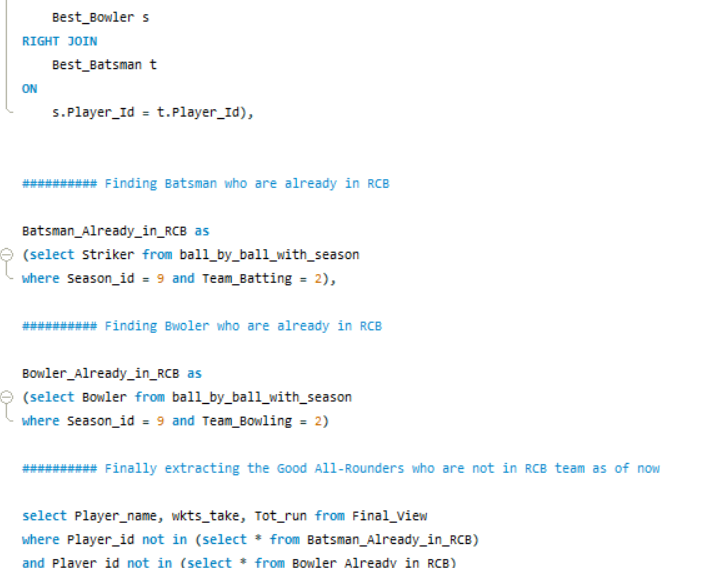
***Output:***

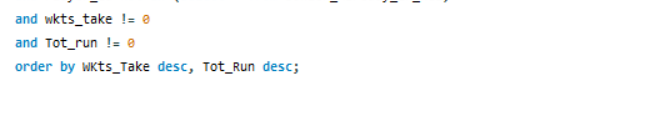
******

***Query:***

******

******

******

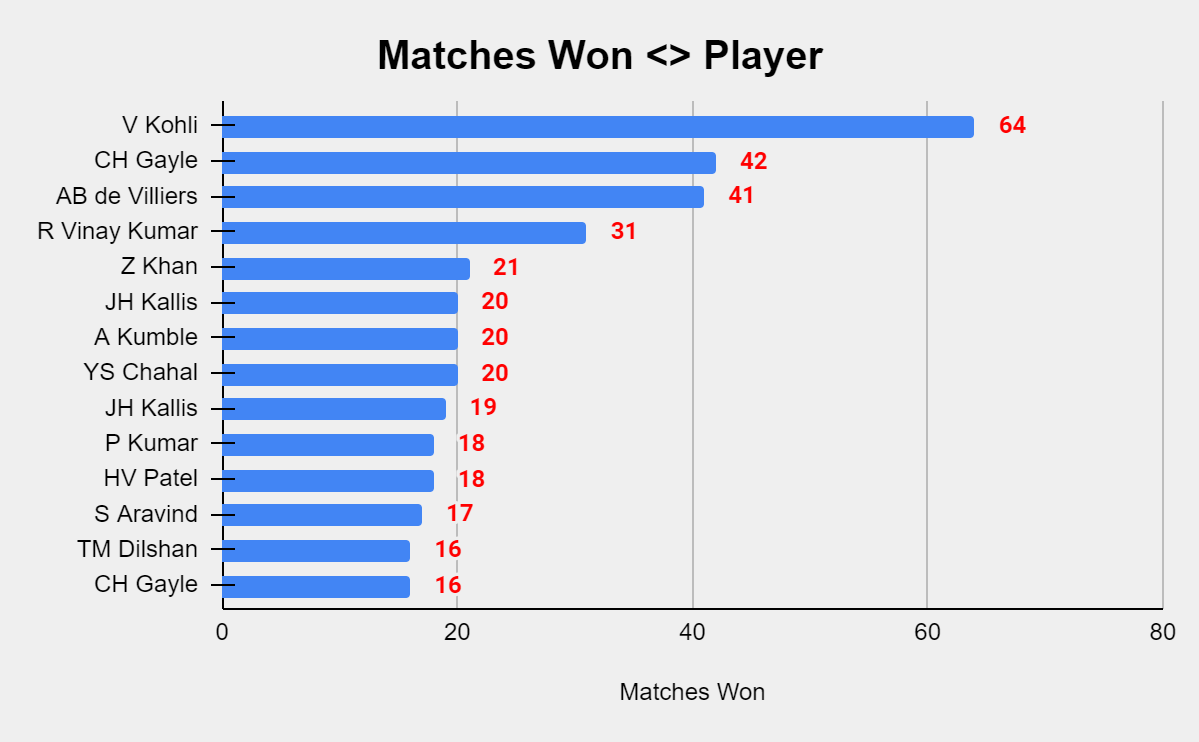
******

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

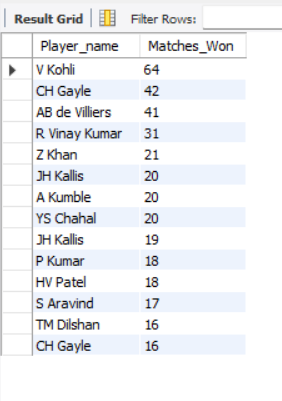
***Answer***: The players whose presence positively influence the morale are the players who pushes the team members to achieve their best and eventually win matches for them

***Approach***: Found the player wise matches won and used the filter Matches Won greater than 15 to get the set of players who have contributed the most in the wins.

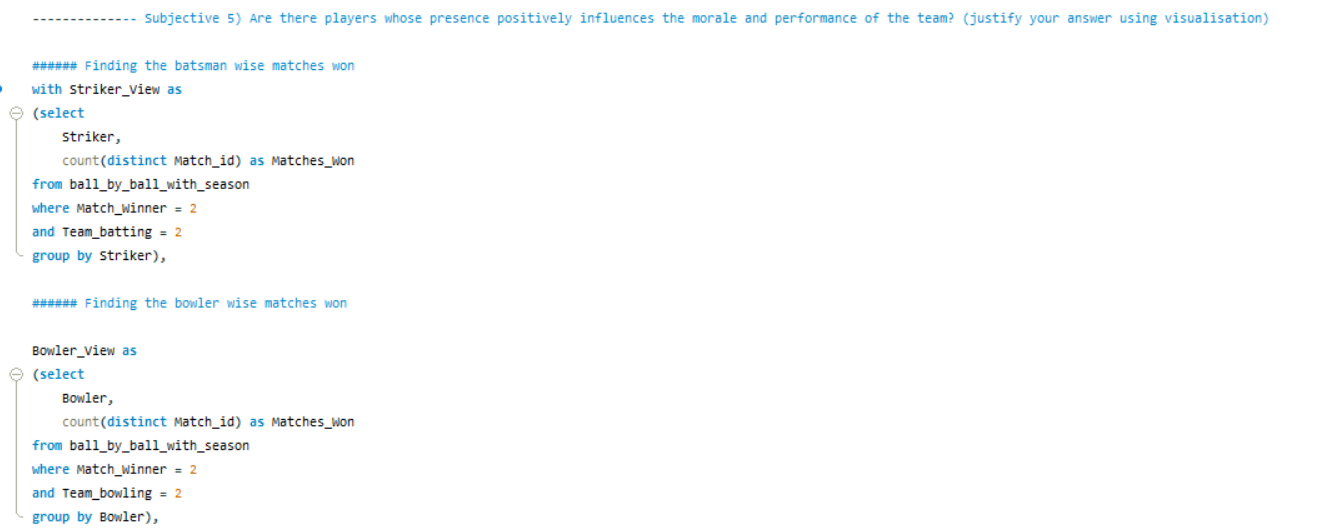
***Observation:*** Senior players like V Kohli, C Gayle and AB de Villers contribute heavily in boosting the morale.

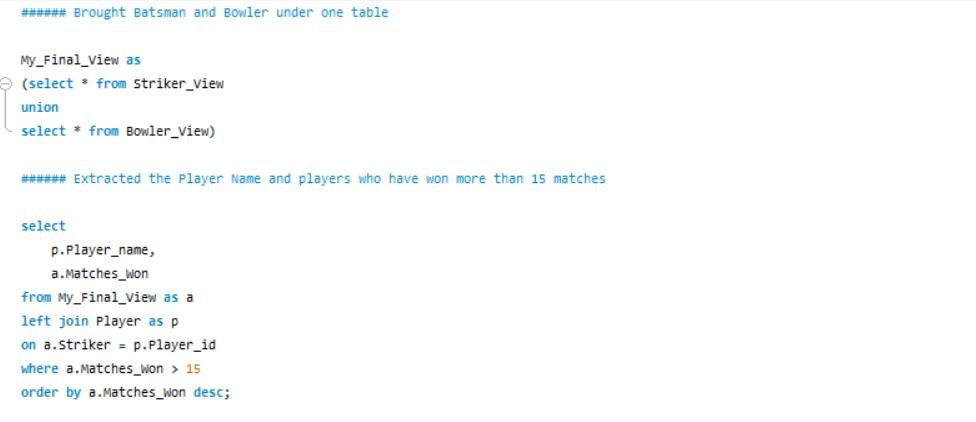


***Output:***

****

**Query:**

****

****

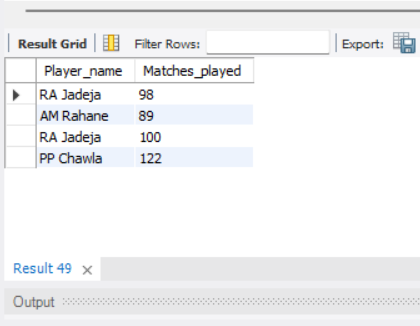
1. **What would you suggest to RCB before going to mega auction?**

***Answer*:**

* I would suggest to go out for young players as they will be the future of our team, proper grooming will help us achieve great success

**Observation**: There are 4 players who are less than 29 years old and played more than 80 matches in this tournament

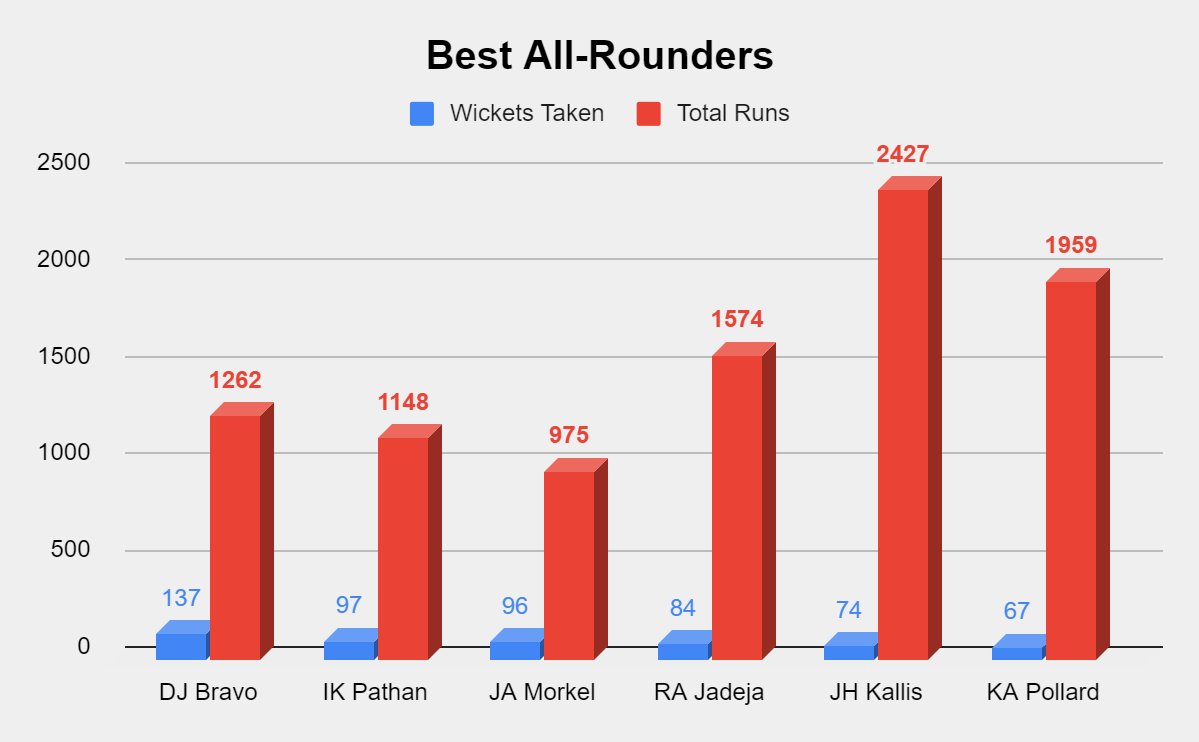
***Output***



***Query:***

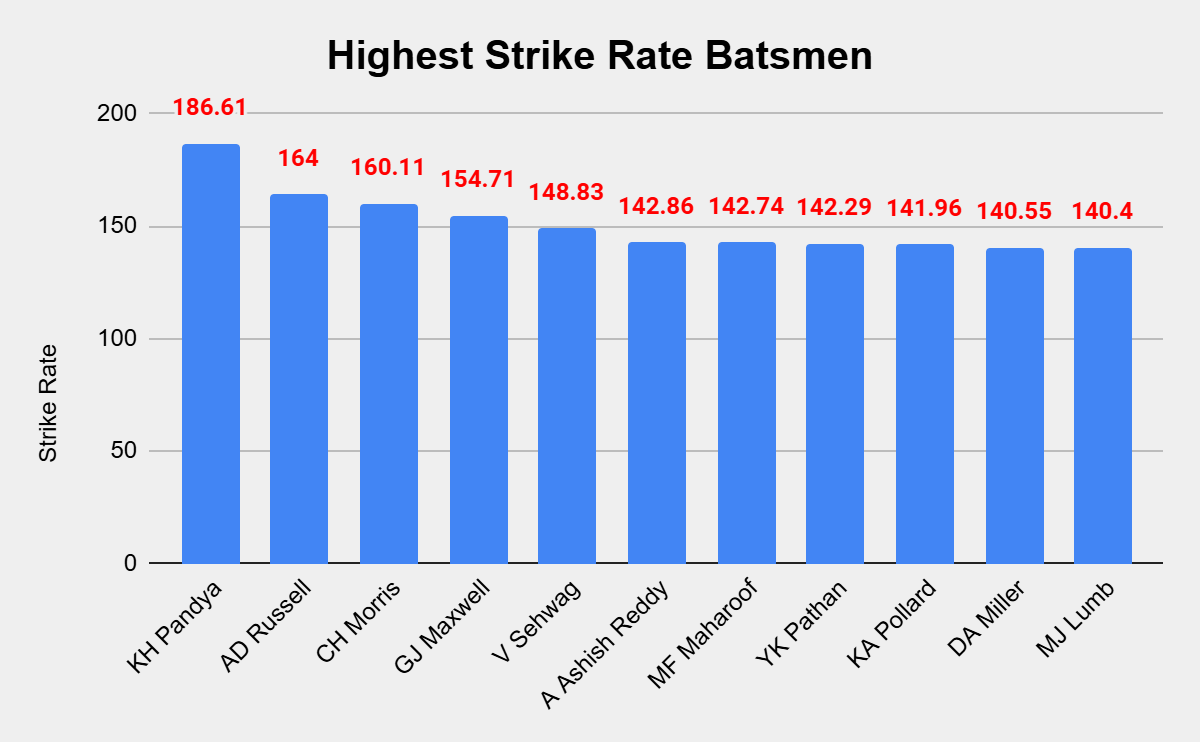


* Focus on all rounders as game has levelled up and we should focus on players who can play both ways.



Query & Output mention in Subjective Q4

* High Strike rate players to be targeted.

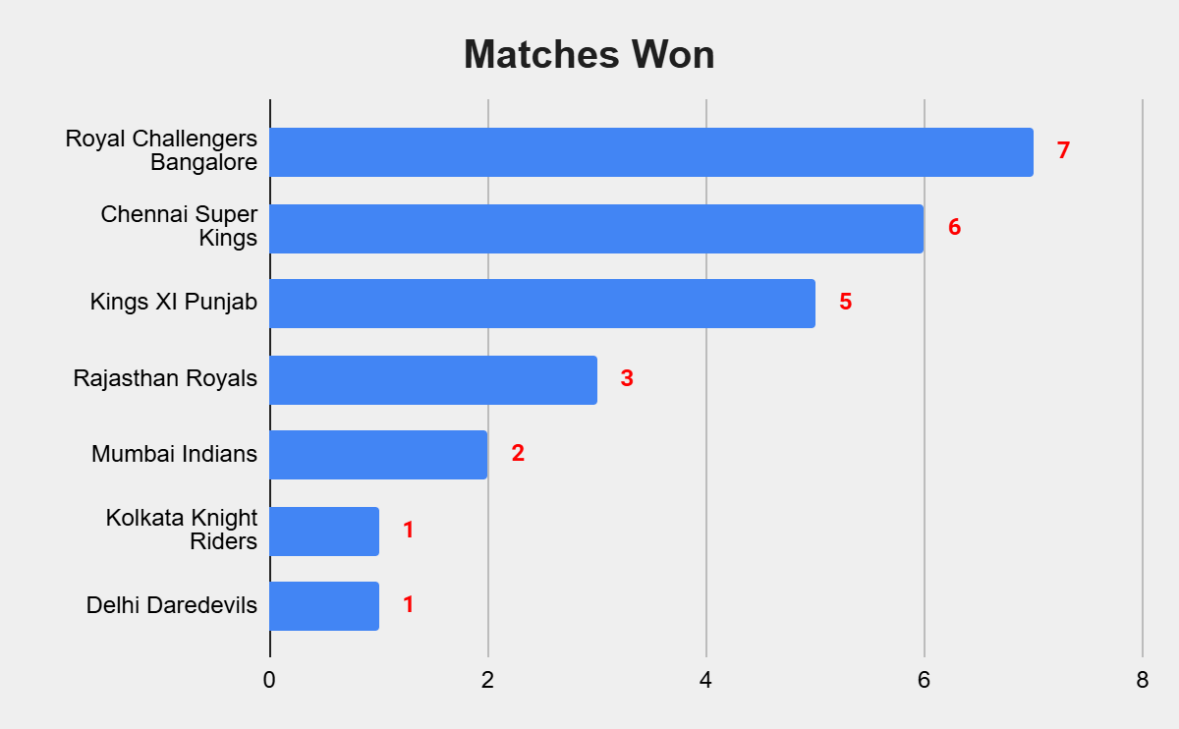


Query & Output mention in Subjective Q3

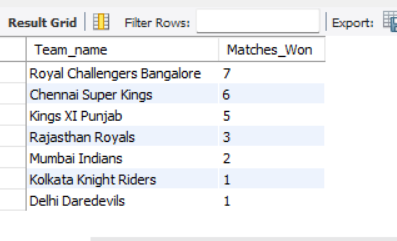
1. **What do you think could be the factors contributing to the high-scoring matches and the impact on viewership and team strategies**

***Answer*:** BelowChartrepresents the number of matches team has won crossing 200 runs – in such high scoring matches viewership is at peak

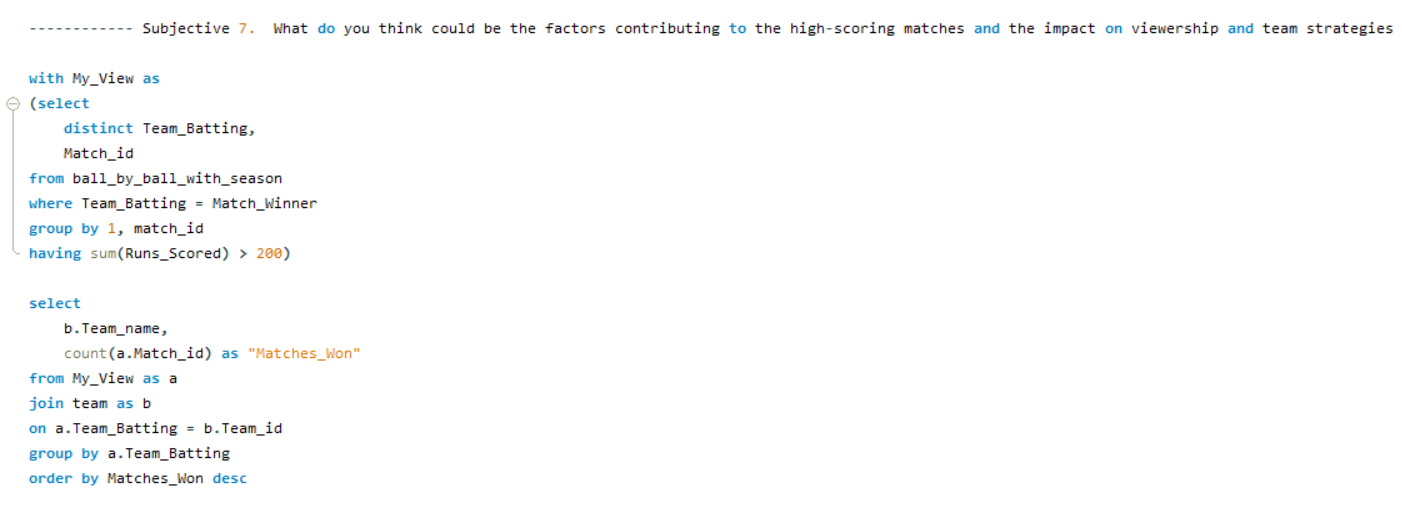
Further the factor behind such high scoring matches can be a specific team as it can be seen that matches against RCB are mostly the high run scoring matches and RCB has won a lot of such matches.



***Output:***



***Query:***

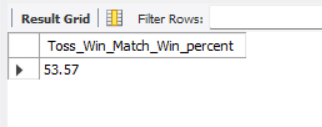


1. **Analyze the impact of home ground advantage on team performance and identify strategies to maximize this advantage for RCB.**

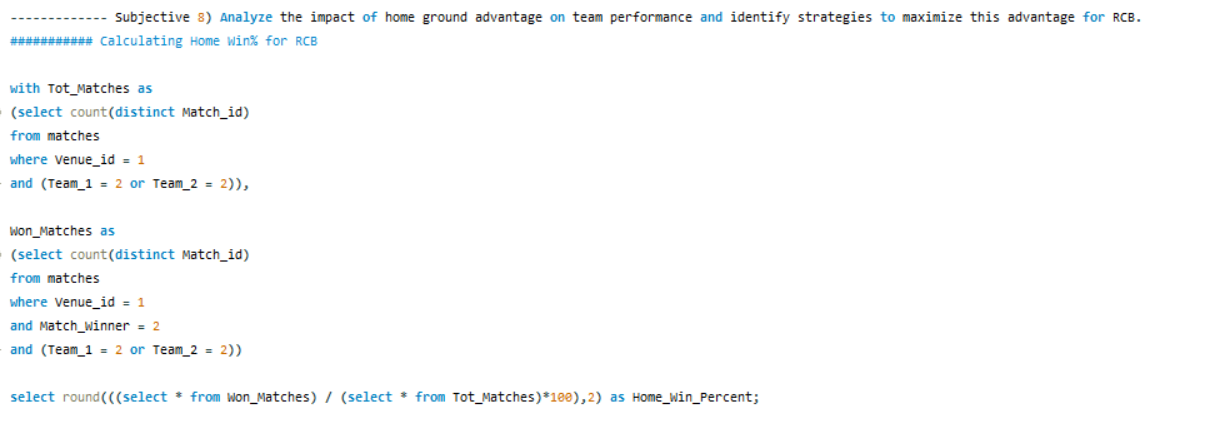
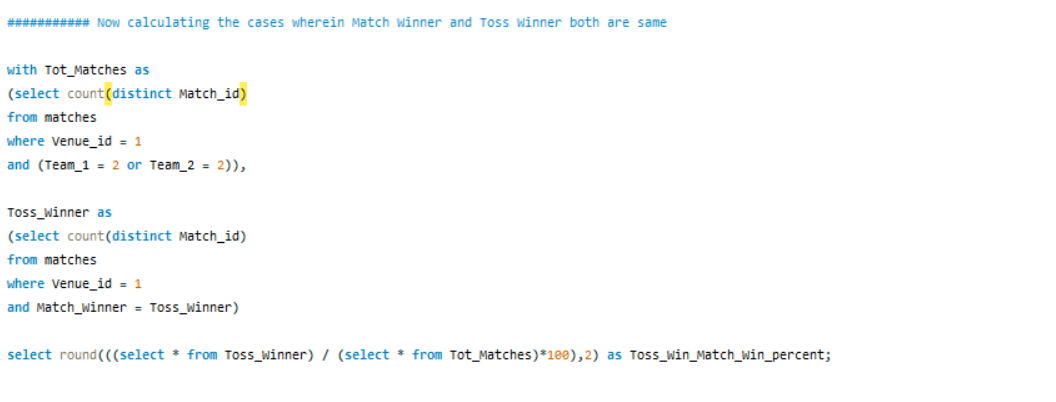
***Answer***: RCB Home Win% is 51.79%

Strategy to Improve Home Advantage for RCB: As I have checked in RCB’s home ground M.Chinnaswamy Stadium – the team who has won the toss has won 53.57% of the matches – which means just by winning toss we can gain a 2% growth in win% , additionally we have already seen in the data that Bowl\_First as 56% win% in this Stadium.

***Output:***



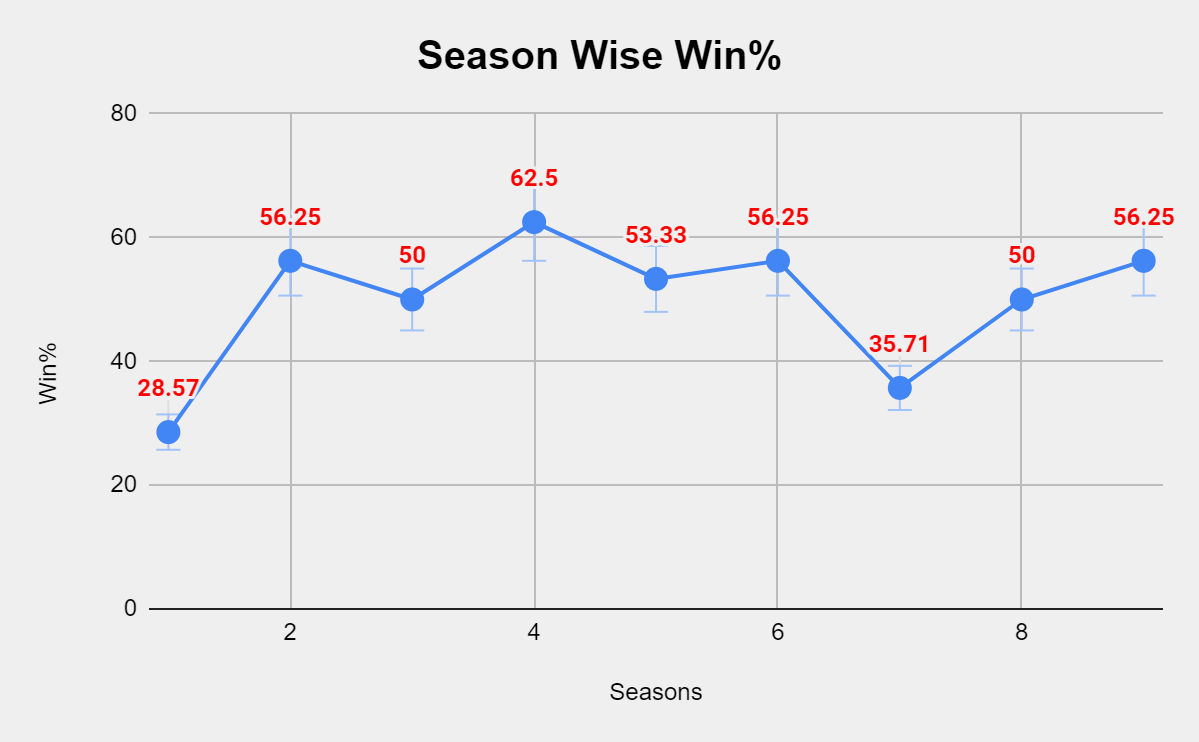
***Query:***

**** ****

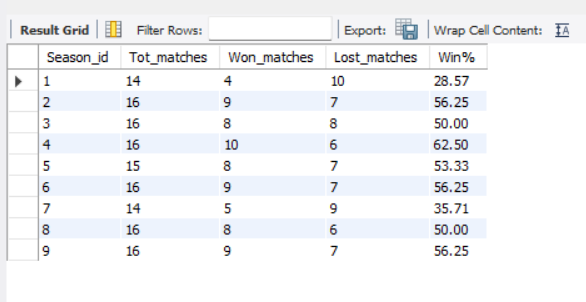
1. **Come up with a visual and analytical analysis with the RCB past seasons performance and potential reasons for them not winning a trophy.**

***Answer***: Below chat shows the win% of RCB across seasons. As it can be seen only in Season 4 RCB was able to cross 60%. Other important factor to be taken into consideration is that RCB looses 51% of the matches Chasing – this is one of the areas of improvement for RCB

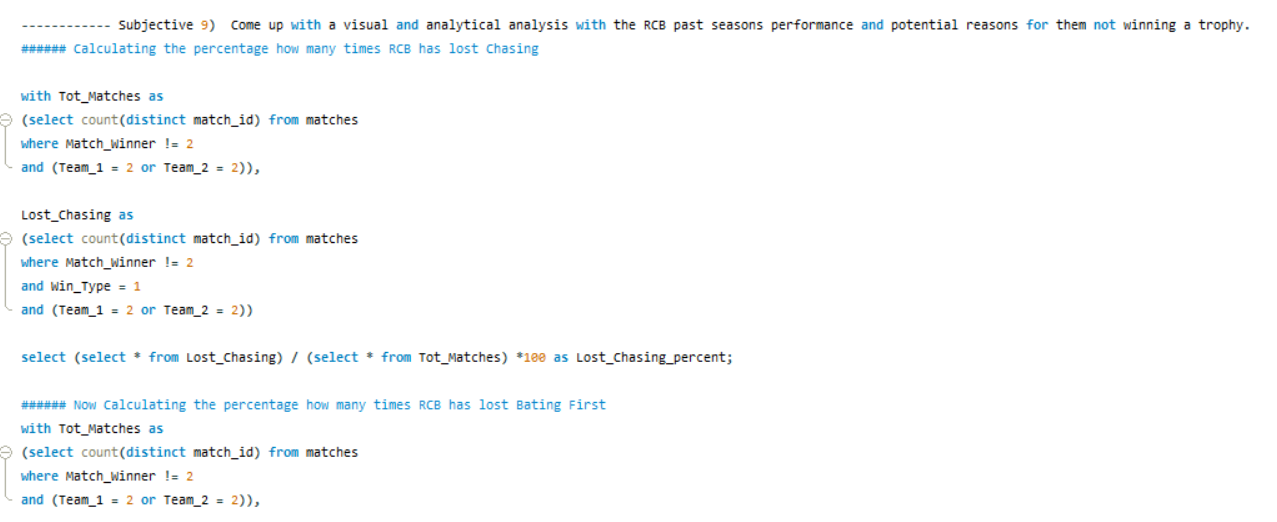
***Approach***: First found the Season wise RCB record with SQL query and then plotted the Chasing or Bating\_First ratio of losing the matches

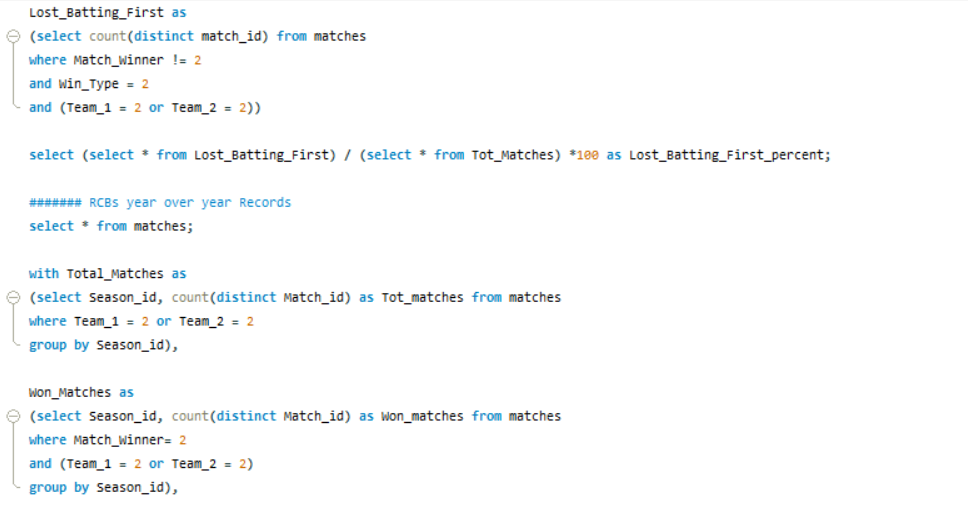


***Output:***

****

***Query:***

****

**** ****

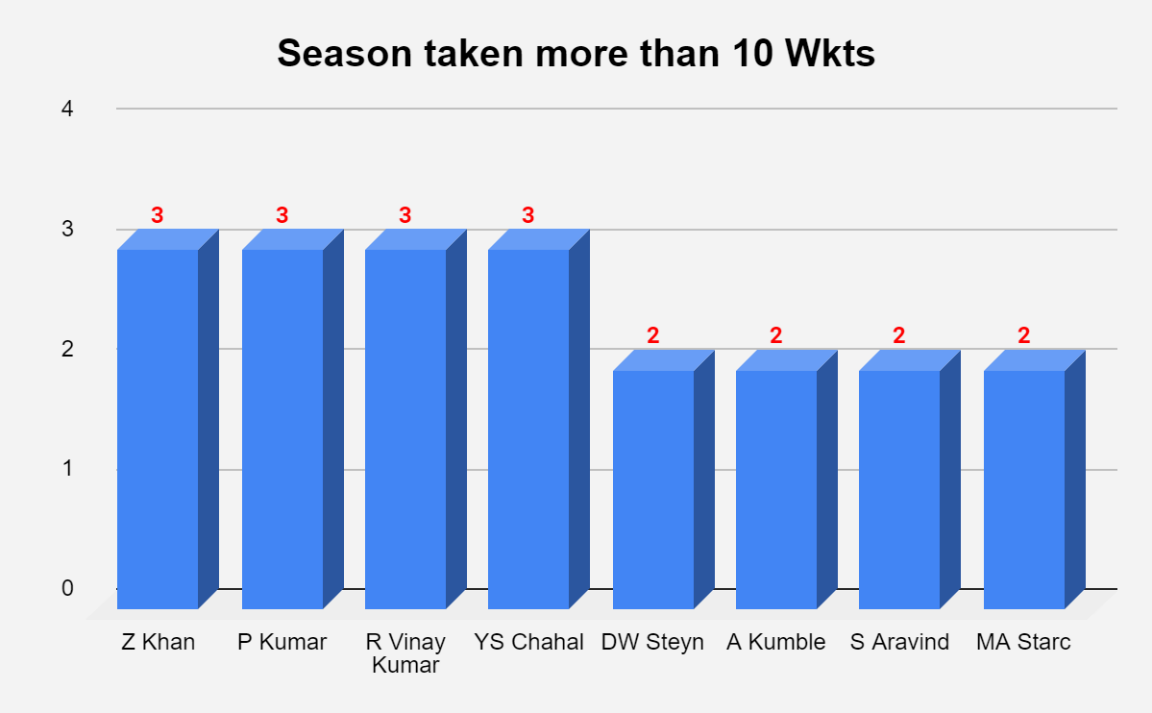
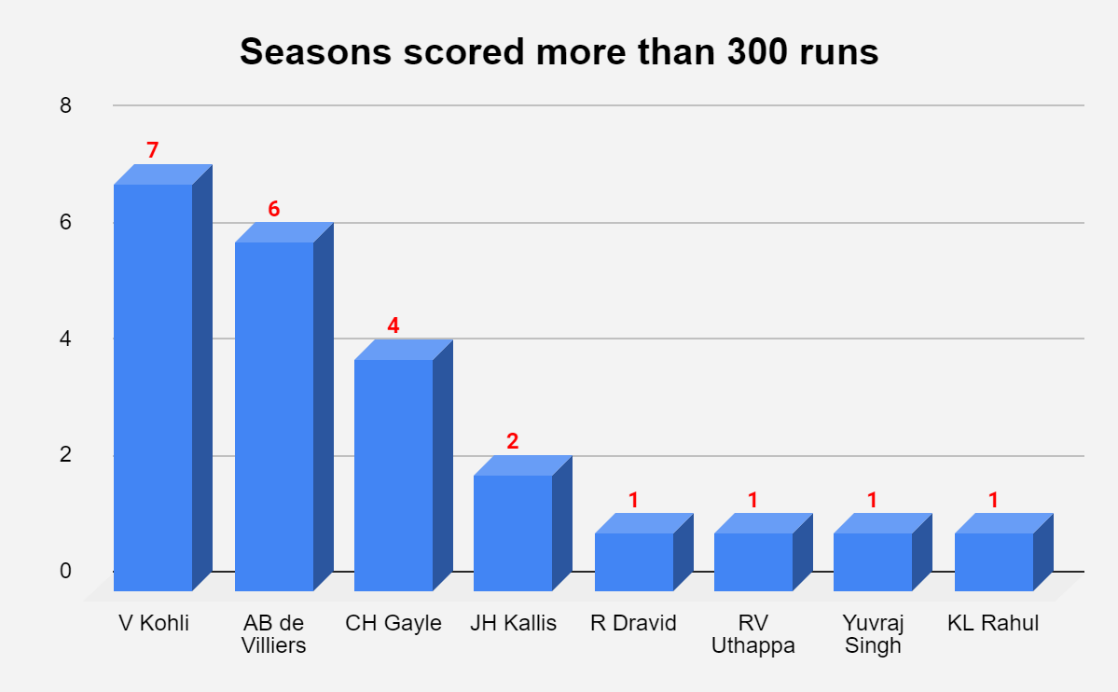
1. **How would you approach this problem, if the objective and subjective questions weren't given?**

**Answer:**

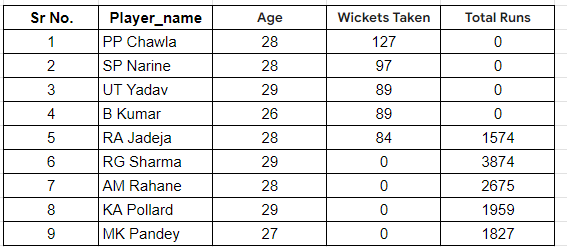
First out of my current team players, I would have picked the players who should be retained.

Now below view shows the RCB batsmen who has been consistent over season and has scored more than 300 runs in multiple seasons.

Similarly, bowler view shows the RCB bowler who has taken more than 10 wickets in multiple seasons. (Query & Output view is available in Objective Q14)



Then I would have focused on players to look out for, below players would have been my recommendation as they have scored more than 1500 runs across reasons and taken over 80 wickets, so below is the combo of Batsmen and Bowlers who can be the right choice for the team. (Query & Output view is available in Subjective Q2)

****

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".**

***Answer*:**

Update Matches

Set Opponent\_Team = Replace (Opponent\_Team, 'Delhi\_Capitals', 'Delhi\_Daredevils')

where Opponent\_Team = 'Delhi\_Capitals';