Answer to Objective Questions   
 Submitted by Tanupriya *Awasthi*

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

**Ans.** As per the schema, all the columns are declared with integer data type that means all the values are expected to be in numeric whole number.

|  |  |
| --- | --- |
| Column Name | Data type |
| Match\_Id | integer |
| Over\_Id | integer |
| Ball\_Id | integer |
| Innings\_No | integer |
| Team\_Batting | integer |
| Team\_Bowling | integer |
| Striker\_Batting\_Position | integer |
| Striker | integer |
| Non\_Striker | integer |
| Bowler | integer |

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

**Ans.** The total runs scored byRCB in 1st season are 1241592 which also include the extra runs.

**Query:   
select sum(Runs\_Scored+Extra\_Runs) as total\_runs  
from batsman\_scored b  
join extra\_runs e on b.Match\_Id=e.Match\_Id  
join player\_match pm on b.Match\_Id = pm.Match\_Id  
join matches m on e.Match\_Id = m.Match\_ID  
where m.Season\_Id =1 and pm.Team\_Id =2;**

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

**Ans. There were** 162 players which were more than age of 25 during season 2 which was in 2009.

**Strategy:** First looked when the season 2 was held then calculated the age of players in that year and then filtered the players whose age was greater than 25.

**with player\_age as   
(select \*,Timestampdiff(Year,DOB,'2009-01-01') as AGE  
from player)  
select count(\*) as no\_of\_players\_Season2  
from player\_age  
where AGE>25;**

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

**Ans.** RCB won **4 matches** in season 1.

**Strategy:** For this we used matches table where we applied the filter for season 1 and match winner as 2 because the team ID for RCB is 2.

**select count(\*) from matches  
where Season\_Id = 1 and Match\_Winner = 2;**

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

**Ans.** The below listed chart shows the top 10 players according to their strike rate in the last 4 season.

**6. What is the average run scored by each batsman considering all the seasons?  
Ans.** The below listed are some of the batsman with average runs scored considering all the seasons. For the full list kindly refer to the query below:

|  |  |
| --- | --- |
| Player Name | Avg. Runs |
| A Ashish Reddy | 282.32 |
| A Chandila | 284.08 |
| A Chopra | 241.29 |
| A Flintoff | 300.00 |
| A Kumble | 276.83 |
| A Mishra | 283.56 |
| A Mithun | 288.56 |
| A Mukund | 331.00 |

**Strategy:** For thisfirst we have joined the tables which contain the data like players name, runs scored. Then we have calculated the number of matches played and the total runs scored and on the basis which we have calculated the average runs by each batsman.  
 **Query:  
select p.Player\_Name, round(sum(Runs\_Scored)/Count(distinct m.Match\_Id), 2) as avg\_runs   
from batsman\_scored b  
join player\_match pm on b.Match\_Id = pm.Match\_Id  
join player p on pm.Player\_Id = p.Player\_Id  
join matches m on b.Match\_Id = m.Match\_ID  
group by 1;**

**7. What are the average wickets taken by each bowler considering all the seasons?  
Ans:** The below listed are some of the bowlers with average wickets taken considering all the seasons. For the full list kindly refer to the query below:

|  |  |
| --- | --- |
| Players Name | Average Wickets |
| A Ashish Reddy | 15 |
| A Chandila | 14.33 |
| A Chopra | 14.2 |
| A Flintoff | 14.14 |
| A Kumble | 14 |
| A Mishra | 13.54 |
| A Mithun | 13.4 |
| A Mukund | 13.33 |
| A Nehra | 13.14 |
| A Nel | 13.07 |
| A Singh | 13 |
| A Symonds | 13 |

**Query :**  
**select p.Player\_Name , round(count(Player\_Out)/Count(distinct m.Match\_Id),2) as avg\_wicket from wicket\_taken w  
join player\_match pm on w.Match\_Id=pm.Match\_Id  
join player p on pm.Player\_Id=p.Player\_Id  
join matches m on w.Match\_Id=m.Match\_Id  
group by 1;**

**8. List all the players who have average runs scored greater than overall average and who have taken wickets greater than overall average  
Ans.** With the help of the below written query we can get the player id of all the players who have average runs scored greater than overall average and who have taken wickets greater than overall average.

**Query:  
WITH overall\_averages AS (  
SELECT AVG(Runs\_Scored) AS overall\_avg\_runs,AVG(Player\_Out) AS overall\_avg\_wickets  
FROM batsman\_scored   
join wicket\_taken using (Match\_Id)),**

**player\_stats as (  
select Player\_Id round(count(Player\_Out)/Count(distinct m.Match\_Id),2) as avg\_wicket, avg(Runs\_Scored) as avg\_runs\_scored  
from wicket\_taken  
join batsman\_scored using(Match\_Id)  
join player\_match using(Match\_Id)  
group by 1)**

**Select Player\_Id,avg\_runs\_scored,overall\_avg\_runs,avg\_wickets,overall\_avg\_wickets  
from overall\_averages oa  
join player\_stats ps   
ON ps.avg\_runs\_scored > oa.overall\_avg\_runs   
AND ps.avg\_wickets > oa.overall\_avg\_wickets;**

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

**Ans.**

**Query:  
Create table rcb\_record (  
Venue\_Id int primary key,  
Venue\_name varchar(200),  
Wins int default 0,  
Losses int default 0);**

**insert into rcb\_record (Venue\_Id, Venue\_name, Wins, Losses)  
Select Venue\_Id , Venue\_name,  
sum(case when Match\_Winner=2 then 1 else 0 end) as Wins ,  
sum(case when(Match\_Winner!=2 and (Team\_1=2 or Team\_2=2)) then 1 else 0 end) as Losses  
from matches  
join venue using(Venue\_Id)  
where Team\_1= 2 or Team\_2=2  
group by 1,2;**

**10. What is the impact of bowling style on wickets taken.  
Ans.** From the below chart we can see the impact of bowling style. From this we can see that left arm fast bowlers take more wickets followed by left arm medium fast bowlers . We can say that left arm fast or medium fast bowlers have tendency to take more wickets compared to other bowler.

**11. 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 team\_performance as  
(select m.Season\_Id,t.Team\_Name, sum(bs.Runs\_Scored) as total\_runs, sum(w.Player\_Out) as total\_wickets  
 from team t join matches m on t.Team\_Id in (m.Team\_1,m.Team\_2)  
 left join ball\_by\_ball b on b.Match\_Id=m.Match\_Id and t.Team\_Id=b.Team\_Batting  
left join batsman\_scored bs on bs.Match\_Id=b.Match\_Id and bs.Over\_Id=b.Over\_Id and bs.Ball\_Id=b.Ball\_Id and bs.Innings\_No=b.Innings\_No  
 left join wicket\_taken w on w.Match\_Id=b.Match\_Id and w.Over\_Id=b.Over\_Id and w.Ball\_Id=b.Ball\_Id and w.Innings\_No=b.Innings\_No  
group by 1,2,3),**

**performance\_comparison as (  
select Season\_Id,Team\_Name, total\_runs,total\_wickets, lag(total\_runs) over(partition by Team\_Id order by Season\_Id) as prev\_year\_runs,  
lag(total\_wickets) over(partition by Team\_Id order by Season\_Id)as prev\_year\_wickets   
from team\_performance)**

**select Season\_Id,Team\_Name,total\_runs, prev\_year\_runs,total\_wickets,prev\_year\_wickets,   
case   
when total\_runs > prev\_year\_runs and total\_wickets > prev\_year\_wickets then 'Better'   
 when total\_runs < prev\_year\_runs and total\_wickets < prev\_year\_wickets then 'Poor'  
else 'Mixed’  
end as Performance\_status from performance\_comparison   
 where prev\_year\_runs is not null and prev\_year\_wickets is not null   
 order by Team\_Name,Season\_Id;**

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

**Ans.** The more KPIs for the team strategy are :

**Bowling KPI –** Economy Rate, Average wickets per match, Dot balls, Power play runs, Death over runs

**Batting KPIs** – Strike Rate, Average runs per match, Boundary, Dot ball

We have compared the strike rate with previous year strike rate , wickets taken with previous etc. and calculate the performance status of the teams whether their performance is getting better every year or declining or its constant.

From these KPI the team should look for the openers who have a high strike rate and has the ability to score freely in the first 6 and 7 overs.

The team should focus on death bowling specialist who have a track record of low economy rate.

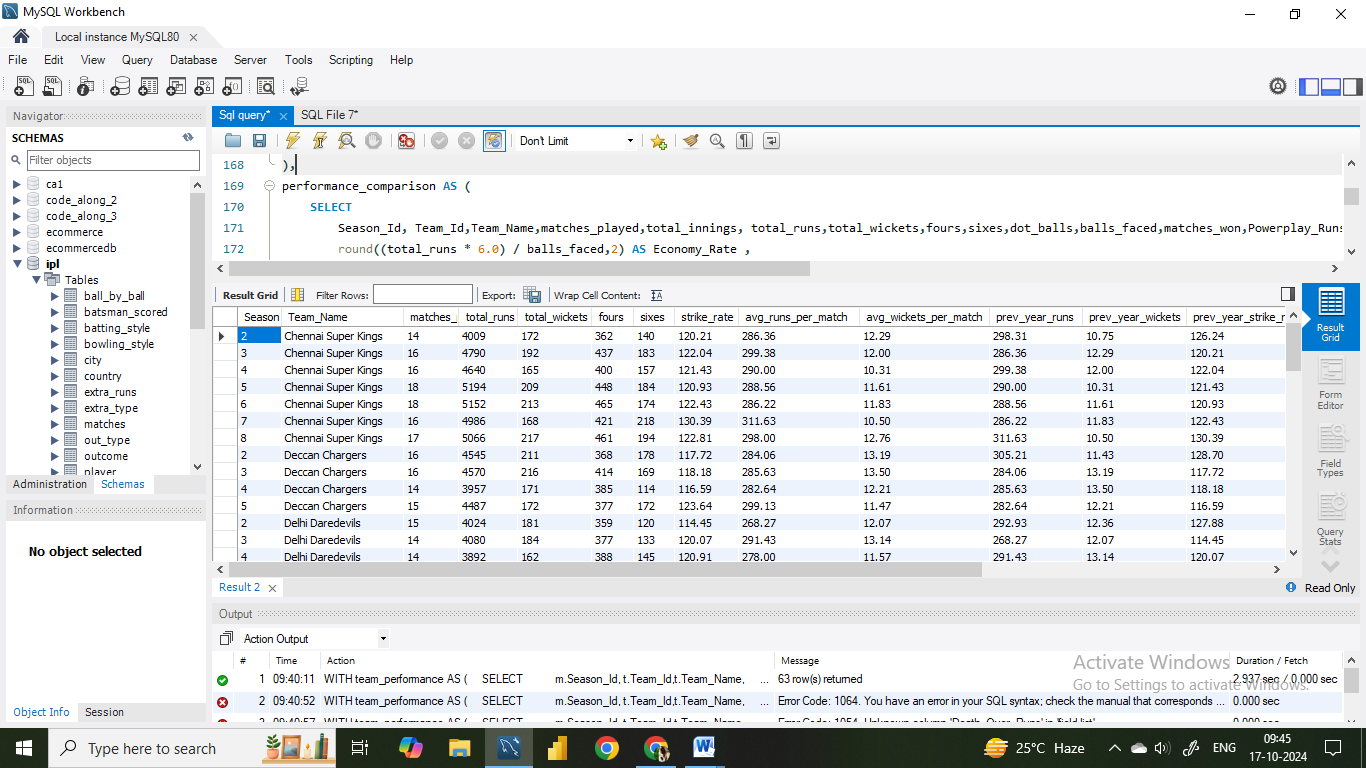
Can also look for players who have average or high strike as well as low economy rate.

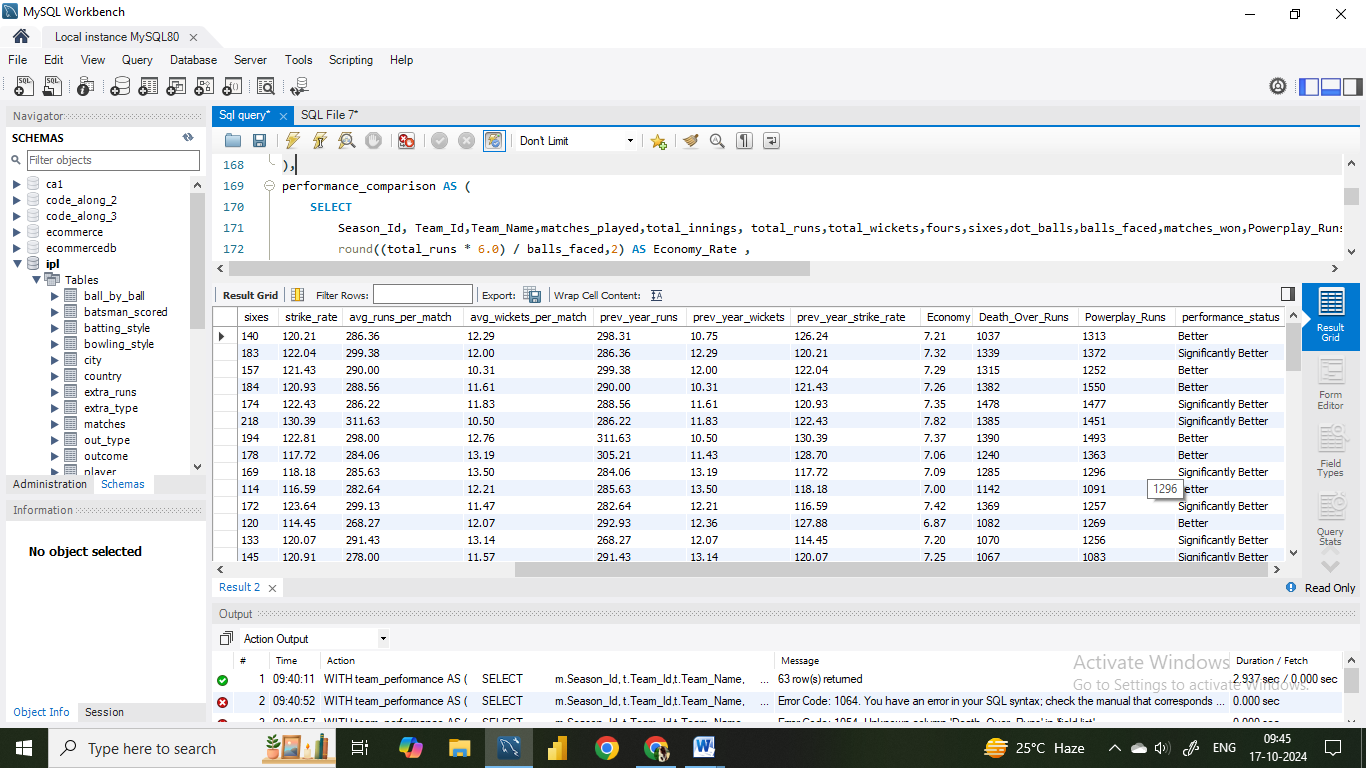
**Query:** The more KPI can be the number of sixes and foursa player had number of dot balls he faced, strikr rate, average runs per match etc.

**WITH team\_performance AS   
( SELECT m.Season\_Id, t.Team\_Id,t.Team\_Name,   
 COUNT(DISTINCT m.Match\_Id) AS matches\_played,  
 COUNT(b.Innings\_No) AS total\_innings,   
 SUM(bs.Runs\_Scored) AS total\_runs,  
 COUNT(w.Player\_Out) AS total\_wickets,  
 SUM(CASE WHEN bs.Runs\_Scored = 4 THEN 1 ELSE 0 END) AS fours,  
 SUM(CASE WHEN bs.Runs\_Scored = 6 THEN 1 ELSE 0 END) AS sixes,   
 SUM(CASE WHEN bs.Runs\_Scored = 0 THEN 1 ELSE 0 END) AS dot\_balls,   
 COUNT(b.Ball\_Id) AS balls\_faced,  
 SUM(CASE WHEN b.Over\_Id >= 16 THEN bs.Runs\_Scored ELSE 0 END) AS Death\_Over\_Runs,  
 SUM(CASE WHEN b.Over\_Id < 7 THEN bs.Runs\_Scored ELSE 0 END) AS Powerplay\_Runs,  
 COUNT(DISTINCT CASE WHEN m.Win\_Type = t.Team\_Id THEN m.Match\_Id END) AS matches\_won FROM team t   
JOIN matches m ON t.Team\_Id IN (m.Team\_1, m.Team\_2)  
 LEFT JOIN ball\_by\_ball b ON b.Match\_Id = m.Match\_Id  
LEFT JOIN batsman\_scored bs ON b.Match\_Id = bs.Match\_Id AND b.Over\_Id = bs.Over\_Id AND b.Ball\_Id = bs.Ball\_Id AND b.Innings\_No = bs.Innings\_No  
LEFT JOIN wicket\_taken w ON w.Match\_Id = b.Match\_Id AND w.Over\_Id = b.Over\_Id AND w.Ball\_Id = b.Ball\_Id AND w.Innings\_No = b.Innings\_No  
GROUP BY 1, 2, 3),**

**performance\_comparison AS (  
 SELECT Season\_Id, Team\_Id,Team\_Name,matches\_played,total\_innings, total\_runs, total\_wickets,fours,sixes,dot\_balls,balls\_faced,matches\_won, Death\_Over\_Runs, Powerplay\_Runs,  
round((total\_runs \* 6.0) / balls\_faced,2) AS Economy\_Rate ,  
ROUND(total\_runs\*100/balls\_faced , 2) AS strike\_rate,  
ROUND(total\_runs/matches\_played, 2) AS avg\_runs\_per\_match,  
ROUND(total\_wickets/matches\_played, 2) AS avg\_wickets\_per\_match,  
 ROUND(dot\_balls\*100/matches\_played , 2) AS dot\_balls\_percentage,  
 LAG(total\_runs) OVER (PARTITION BY Team\_Id ORDER BY Season\_Id) AS prev\_year\_runs, LAG(total\_wickets) OVER (PARTITION BY Team\_Id ORDER BY Season\_Id) AS prev\_year\_wickets, LAG(ROUND(total\_runs/balls\_faced, 2)) OVER (PARTITION BY Team\_Id ORDER BY Season\_Id) AS prev\_year\_strike\_rate  
 FROM team\_performance)**

**SELECT Season\_Id, Team\_Name, matches\_played,total\_runs,total\_wickets,fours,sixes,strike\_rate,avg\_runs\_per\_match,avg\_wickets\_per\_match, prev\_year\_runs, prev\_year\_wickets, prev\_year\_strike\_rate, Death\_Over\_Runs, Powerplay\_Runs,  
 CASE WHEN total\_runs > prev\_year\_runs AND total\_wickets > prev\_year\_wickets AND strike\_rate > prev\_year\_strike\_rate THEN 'Significantly Better'  
WHEN total\_runs > prev\_year\_runs AND total\_wickets > prev\_year\_wickets THEN 'Better'  
WHEN total\_runs < prev\_year\_runs AND total\_wickets < prev\_year\_wickets AND strike\_rate < prev\_year\_strike\_rate THEN 'Significantly Worse'  
 WHEN total\_runs < prev\_year\_runs AND total\_wickets < prev\_year\_wickets THEN 'Worse'  
WHEN strike\_rate > prev\_year\_strike\_rate THEN 'Improved strike\_rate '  
ELSE 'Mixed'  
END AS performance\_status   
FROM performance\_comparison   
WHERE prev\_year\_runs IS NOT NULL AND prev\_year\_wickets IS NOT NULL   
ORDER BY Team\_Name, Season\_Id;**

****

****

**13. 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.  
Ans.**

**Query:  
Select p.Player\_Id,Player\_Name,Venue\_Name,Round(avg(Player\_Out),2) as average\_wickets,  
 dense\_rank() over(partition by Venue\_Name order by avg(Player\_Out) desc) as `rank`  
from wicket\_taken w  
join matches m on w.Match\_Id=m.Match\_Id  
join venue v on v.Venue\_Id=m.Venue\_Id  
join player\_match pm on m.Match\_Id=pm.Match\_Id  
join player p on pm.Player\_ID=p.Player\_Id  
group by 1,2,3;**

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

**Ans.** The low standard deviation indicates that the players are consistent in their performance across the season.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Player Id | Player Name | Min Runs | Max Runs | Avg Runs | Std dev |
| 460 | Mustafizur Rahman | 4704 | 4704 | 4704.0000 | 0 |
| 407 | JC Buttler | 4180 | 4180 | 4180.0000 | 0 |
| 118 | GD McGrath | 4101 | 4101 | 4101.0000 | 0 |
| 413 | KH Pandya | 3596 | 3596 | 3596.0000 | 0 |
| 49 | SM Pollock | 3557 | 3557 | 3557.0000 | 0 |
| 70 | Shahid Afridi | 3071 | 3071 | 3071.0000 | 0 |
| 102 | Sohail Tanvir | 3042 | 3042 | 3042.0000 | 0 |
| 319 | N Saini | 3012 | 3012 | 3012.0000 | 0 |
| 420 | RR Pant | 2949 | 2949 | 2949.0000 | 0 |
| 434 | CJ Jordan | 2939 | 2939 | 2939.0000 | 0 |
| 107 | SP Fleming | 2833 | 2833 | 2833.0000 | 0 |
| 418 | M Ashwin | 2811 | 2811 | 2811.0000 | 0 |
| 143 | DP Vijaykumar | 2793 | 2793 | 2793.0000 | 0 |
| 230 | AC Voges | 2597 | 2597 | 2597.0000 | 0 |
| 117 | Mohammad Asif | 2518 | 2518 | 2518.0000 | 0 |
| 326 | V Pratap Singh | 2500 | 2500 | 2500.0000 | 0 |
| 108 | S Vidyut | 2418 | 2418 | 2418.0000 | 0 |
| 353 | BJ Rohrer | 2397 | 2397 | 2397.0000 | 0 |
| 122 | M Ntini | 2395 | 2395 | 2395.0000 | 0 |

**15. Are there players whose performance is more suited to specific venues or conditions? (how would you present this using charts?)   
Ans.** Yes, there are players whose performance is more suited to specific venues. The column chart below shows the top 10 players with their average runs at venue and overall average runs.

Answer to 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?**

**Ans.** We can see that at some stadiums toss decision really do not matter at all in the result of the match.

But in some stadium the toss decision effect the result of the match.

* In Rajiv Gandhi International Stadium we can see that batting after winning the toss yields very poor results and fielding is therefore much favourable after winning the toss.
* In Eden Gardens also we can see that teams that fielded first won a significantly higher number of matches. Same as withPunjab Cricket Association Stadium fielding first yields a favourable outcome.

**Query:**

**select Venue\_Name,Toss\_Name,Sum(case when Toss\_Winner=Match\_Winner then 1 else 0 end ) as**

**wins, Sum(case when Toss\_Winner!=Match\_Winner then 1 else 0 end) as Losses**

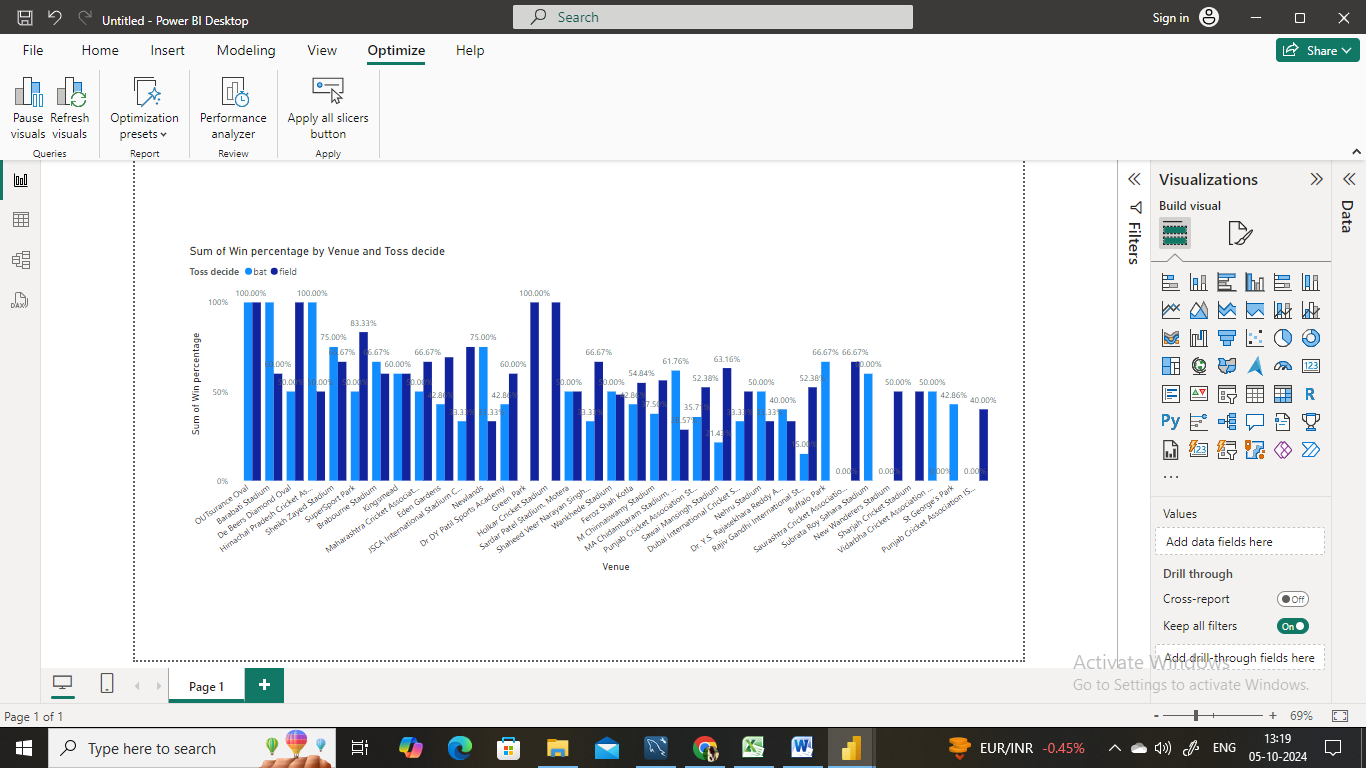
**from matches m**

**join venue v on m.Venue\_Id=v.Venue\_Id**

**join toss\_decision t on m.Toss\_Decide=t.Toss\_id**

**group by 1,2**

**order by 1;**



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

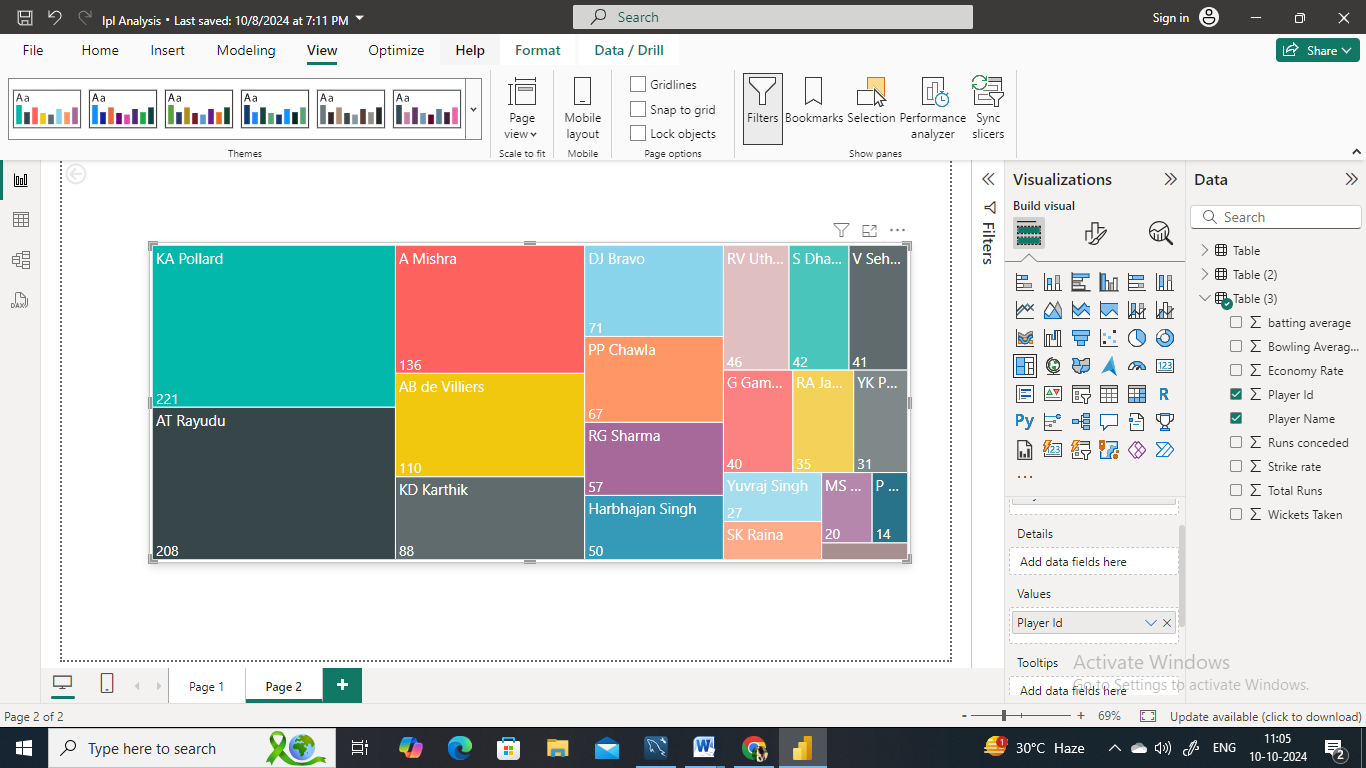
**Ans.** On the basis of the performance of the players we can choose the players who would be the best fit.

While selecting the players we will analyse the performance on the basis of the average runs scored,

Batting average, their strike rate ,their bowling average, bowling strike rate etc.

Below mentioned tree map shows the top 20 players on the basis of their performance which is

being analysed on the above discussed criteria.



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

**Ans.** The following are some of the parameters that we should focus while selecting the players :

* **Batting Performance:**

**1) Total runs scored:** It is a direct measure of batsman productivity. The higher the runs, the valuable the batsman will be for his team. Consistent run-scorers are very valuable and can give good opening to its team.

**2) Batting Average :** The batting average indicates the consistency of a batsman. The higher batting average shows that the batsman scores more before getting out. For example if the batting average of a batsman is 100 it means they score 100 runs every time they bat before getting out.

**3) Strike rate:** The strike rate indicates the speed at which the batsman scores run. A higher strike rate indicates an aggressive and a high scoring player, who can be good performer in limited over match. The higher strike rate shows the effectiveness of a batsman in rotating the side and hitting the boundaries.

* **Bowling Performance:**

**1) Total wickets taken:** It indicates the bowler’s productivity. It shoes that the bowler is good in taking the wickets of the key players.

**2) Bowling Average:** The bowling average indicates the consistency of a bowler. It shows how much runs a bowlers concedes for each wicket taken. The lower bowling average indicates that bowler is not only taking the wickets but also concedes the fewer runs and this shows the bowlers efficiency.

**3) Economy Rate:** Economy rate indicates how many runs a bowler concedes per over. A bowler with a lower economy rate is much crucial in limited over match as he can keep check on the opposition’s run rate.

**4) Strike Rate:** The strike rate in bowling shoes how many deliveries it takes on an average to take a wicket. A lower strike rate indicates that a bowler takes wickets more frequently thus creating pressure on the opposition team.

**4. 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:**

**Query:**

**SELECT p.Player\_Name,**

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

**ROUND(SUM(bs.Runs\_Scored) \* 1.0 / NULLIF(COUNT(DISTINCT CASE WHEN b.Striker= p.Player\_Id THEN m.Match\_Id END), 0), 2) AS Batting\_Average,**

**COUNT(w.Player\_Out) AS Wickets\_Taken,**

**ROUND(SUM(CASE WHEN b.Bowler = p.Player\_Id THEN bs.Runs\_Scored ELSE 0 END) \* 1.0 / NULLIF(COUNT(w.Player\_Out), 0), 2) AS Bowling\_Average,**

**ROUND(((SUM(bs.Runs\_Scored) \* 1.0 / NULLIF(COUNT(DISTINCT CASE WHEN b.Striker = p.Player\_Id THEN m.Match\_Id END), 0)) +**

**(COUNT(w.Player\_Out) \* 25.0 / NULLIF(COUNT(DISTINCT CASE WHEN b.Bowler= p.Player\_Id THEN m.Match\_Id END), 0))**

**), 2) AS All\_Rounder\_Score**

**FROM**

**player p**

**LEFT JOIN**

**ball\_by\_ball b ON p.Player\_Id IN (b.Striker, b.Non\_Striker, b.Bowler)**

**LEFT JOIN**

**batsman\_scored bs ON b.Match\_Id = bs.Match\_Id AND b.Over\_Id = bs.Over\_Id AND b.Ball\_Id = bs.Ball\_Id**

**LEFT JOIN**

**wicket\_taken w ON b.Match\_Id = w.Match\_Id AND b.Over\_Id = w.Over\_Id AND b.Ball\_Id = w.Ball\_Id**

**JOIN**

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

**GROUP BY**

**p.Player\_Id, p.Player\_Name**

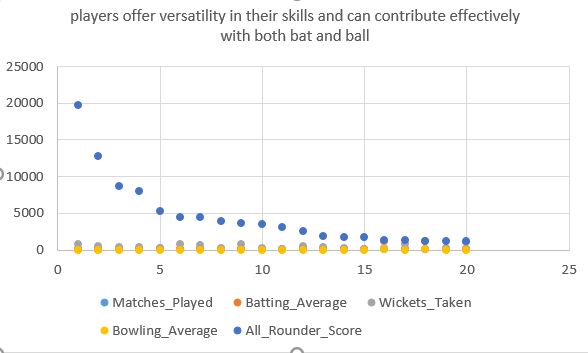
**HAVING**

**Matches\_Played >= 20 AND Wickets\_Taken >= 10 AND Batting\_Average > 20**

**ORDER BY**

**All\_Rounder\_Score DESC0**

**LIMIT 20;**

****

**5. Are there players whose presence positively influences the morale and performance of the team?**

**Ans.**  Yes, there are some players whose presence influences the morale and performance of the team.

**Query:**

**with player\_matches as (select m.Match\_Id, t.Team\_Id, t.Team\_Name,**

**p.Player\_Id,p.Player\_Name,m.Match\_Winner,m.Win\_Margin,   
 CASE WHEN w.Win\_Type='runs' then 1 else 0 end as Won\_Batting\_First,**

**CASE WHEN w.Win\_Type='wickets' then 1 else 0 end as Won\_Bowling\_First**

**from player\_match pm**

**join player p on pm.Player\_Id=p.Player\_Id**

**join matches m on pm.Match\_Id=m.Match\_Id**

**join team t on t.Team\_Id in (m.Team\_1,m.Team\_2)**

**join win\_by w on w.Win\_Id=m.Win\_Type),**

**team\_performance as (select pm1.Team\_Id, pm1.Team\_Name, pm1.Player\_Id,pm1.Player\_Name,**

**count(distinct pm1.Match\_Id) as matches\_played,**

**sum(case when pm1.Match\_Winner =pm1.Team\_Id then 1 else 0 END) AS Matches\_Won,**

**avg(case when pm1.Match\_Winner =pm1.Team\_Id then pm1.Win\_Margin else null end) as avg\_win\_margin,**

**sum(pm1.Won\_Batting\_First) as wins\_batting\_first, sum(pm1.Won\_Bowling\_First) as wins\_bowling\_first**

**from player\_matches pm1**

**group by 1,2,3,4),**

**team\_overall\_performance as (select t1.Team\_Id, count(distinct m1.Match\_Id) as total\_matches, sum(case when m1.Match\_Winner=t1.Team\_Id then 1 else 0 end) as total\_wins**

**from matches m1**

**join team t1 on t1.Team\_Id in (m1.Team\_1,m1.Team\_2)**

**group by t1.Team\_Id)**

**SELECT tp.Team\_Id,tp.Team\_Name, tp.Player\_Name, tp.matches\_played,tp.Matches\_Won,**

**round(tp.Matches\_Won \* 100.0/tp.matches\_played,2) as win\_percentage,**

**round(tp.avg\_win\_margin ,2) as avg\_win\_margin,tp.wins\_batting\_first,tp.wins\_bowling\_first,top.total\_matches,top.total\_wins,**

**round(top.total\_wins \* 100.0/top.total\_matches, 2) as team\_overall\_win\_percentage,**

**round((tp.Matches\_Won \* 100.0/tp.matches\_played)-(top.total\_wins \* 100.0 /top.total\_matches),2) as win\_percentage\_difference**

**from team\_performance tp join team\_overall\_performance top on tp.Team\_Id=top.Team\_Id**

**where tp.matches\_played >=10**

**order by win\_percentage\_difference desc, win\_percentage desc**

**limit 20;**

The below given charts represents the win percentage when a player has played a match and overall team win percentage. Here we can see that those players when included in the team, then the winning percentage of a team increases that influences the morale and the performance of the team gets even more better.

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

Ans: Before going to mega auction I would like to suggest the RCB to look for the batsman who can perform well in death overs and focus on their strike rate .

In case when choosing for the bowlers the RCB should look for the bowlers whose economy rate is low as the lower economy rate indicate a bowler is more effective at preventing the opposing team from scoring the runs.

Not only batsman and bowlers the RCB should have at least 4 all rounders whose batting as well as bowling average is good and economy rate is also low.

The RCB should have experienced players as well as the younger players because the experience can be used in building the strategy to win the match and the young players have different bowling and batting style and have new strategy .

Also we see historically the RCB has always rely on the top order batsman and the middle order keep struggling under pressure. So RCB should strengthen his middle over. For this RCB should look for the players with the consistent batting average and strike rate under pressure.

We also see that the team often succeed in middle overs has effective spinners. So RCB should look for spinners who have a low economy rate in middle overs.

As per me , these are some of the important points that RCB should focus on before going to mega auction.

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

**Viewership and team strategies.**

**Ans.** The factors contributing to high scoring matches are the power play runs and death over runs.

The team that scores high had a good start and comparatively score more runs in power play and also in the death overs.

**Query:**

**WITH match\_scores AS (select m.Match\_Id,t.Team\_Name AS Batting\_Team,SUM(bs.Runs\_Scored) AS Total\_Runs, v.Venue\_Name,c.City\_Name, co.Country\_Name, td.Toss\_Name, m.Season\_Id, m.Outcome\_type,**

**SUM(CASE WHEN b.Over\_Id >= 16 THEN bs.Runs\_Scored ELSE 0 END) AS Death\_Over\_Runs,**

**SUM(CASE WHEN b.Over\_Id < 7 THEN bs.Runs\_Scored ELSE 0 END) AS Powerplay\_Runs**

**FROM matches m**

**join player\_match pm on m.Match\_Id=pm.Match\_Id join ball\_by\_ball b on m.Match\_Id=b.Match\_Id**

**join batsman\_scored bs on b.Match\_Id=bs.Match\_Id and b.Over\_Id=bs.Over\_Id and b.Ball\_Id=bs.Ball\_Id**

**join toss\_decision td on td.Toss\_Id=m.Toss\_Winner**

**join team t on pm.Team\_Id=t.Team\_Id join venue v on m.Venue\_Id=v.Venue\_Id join city c on v.City\_Id=c.City\_Id**

**join country co on c.Country\_id=co.Country\_Id where bs.Runs\_Scored is not null**

**group by m.Match\_Id, t.Team\_Name, v.Venue\_Name, c.City\_Name, co.Country\_Name, td.Toss\_Name, m.Season\_Id, m.Outcome\_type**

**having sum(bs.Runs\_Scored) >300),**

**venue\_analysis as**

**( select v.Venue\_Name,c.City\_Name, co.Country\_Name, AVG(ms.Total\_Runs) AS Avg\_Runs,AVG(ms.Death\_Over\_Runs) AS Avg\_Death\_Over\_Runs,**

**AVG(ms.Powerplay\_Runs) AS Avg\_Powerplay\_Runs,**

**COUNT(ms.Match\_Id) AS High\_Scoring\_Matches from match\_scores ms join venue v on ms.Venue\_Name=v.Venue\_Name**

**join city c on c.City\_Id=v.City\_Id**

**join country co on c.Country\_id=co.Country\_id**

**group by v.Venue\_Name,c.City\_Name, co.Country\_Name),**

**team\_performance as**

**(select t.Team\_Name,AVG(ms.Total\_Runs) AS Avg\_Team\_Runs, AVG(ms.Death\_Over\_Runs) AS Avg\_Team\_Death\_Over\_Runs,**

**AVG(ms.Powerplay\_Runs) AS Avg\_Team\_Powerplay\_Runs, COUNT(ms.Match\_Id) AS High\_Scoring\_Matches**

**from match\_scores ms join team t on ms.Batting\_Team=t.Team\_Name**

**group by t.Team\_Name)**

**select va.Venue\_Name,**

**va.City\_Name,va.Country\_Name,va.Avg\_Runs AS Avg\_Venue\_Runs,va.Avg\_Death\_Over\_Runs AS Avg\_Venue\_Death\_Over\_Runs,**

**va.Avg\_Powerplay\_Runs AS Avg\_Venue\_Powerplay\_Runs,va.High\_Scoring\_Matches AS High\_Scoring\_Matches\_Venue,**

**tp.Team\_Name,tp.Avg\_Team\_Runs AS Avg\_Team\_Runs,tp.Avg\_Team\_Death\_Over\_Runs AS Avg\_Team\_Death\_Over\_Runs,**

**tp.Avg\_Team\_Powerplay\_Runs AS Avg\_Team\_Powerplay\_Runs,tp.High\_Scoring\_Matches AS High\_Scoring\_Matches\_Team**

**from venue\_analysis va JOIN team\_performance tp on tp.Avg\_Team\_Runs > 300**

**order by va.High\_Scoring\_Matches DESC,tp.Avg\_Team\_Runs DESC;**

**8. Analyze the impact of home ground advantage on team performance and identify strategies to**

**maximize this advantage for RCB.**

**Query:**

**with team\_performance as (SELECT t.Team\_Name,v.Venue\_Name,m.Match\_Id,SUM(bs.Runs\_Scored) AS Total\_Runs,COUNT(w.Player\_Out) AS Wickets\_Taken,**

**SUM(CASE WHEN b.Over\_Id >= 16 THEN bs.Runs\_Scored ELSE 0 END) \* 6.0 / COUNT(CASE WHEN b.Over\_Id >= 16 THEN 1 END) AS Death\_Over\_Economy,**

**case when v.Venue\_Name='M Chinnaswamy Stadium' then 'Home' else 'Away' end as 'Ground\_type',**

**case when m.Outcome\_type=2 then 1 else 0 end as win**

**from matches m join player\_match pm on m.Match\_Id=pm.Match\_id**

**join team t on pm.Team\_Id=t.Team\_Id**

**join ball\_by\_ball b on m.Match\_Id = b.Match\_Id**

**left join batsman\_scored bs on b.Match\_Id = bs.Match\_Id AND b.Over\_Id = bs.Over\_Id AND b.Ball\_Id = bs.Ball\_Id**

**left join wicket\_taken w on b.Match\_Id = w.Match\_Id AND b.Over\_Id = w.Over\_Id AND b.Ball\_Id = w.Ball\_Id**

**join venue v on m.Venue\_Id=v.Venue\_Id where t.Team\_Name='Royal Challengers Bangalore' group by t.Team\_Name,v.Venue\_Name,m.Match\_Id),**

**team\_home\_away\_stats AS (**

**SELECT**

**Ground\_Type,**

**COUNT(DISTINCT Match\_Id) AS Matches\_Played,**

**SUM(Win) AS Wins,**

**ROUND(SUM(Win) \* 100.0 / COUNT(DISTINCT Match\_Id), 2) AS Win\_Percentage,**

**AVG(Total\_Runs) AS Avg\_Runs\_Scored,**

**AVG(Wickets\_Taken) AS Avg\_Wickets\_Taken,**

**ROUND(AVG(Death\_Over\_Economy), 2) AS Avg\_Death\_Over\_Economy**

**FROM**

**team\_performance**

**GROUP BY**

**Ground\_Type)**

**SELECT**

**Ground\_Type AS Venue\_Type,**

**Matches\_Played,**

**Wins,**

**Win\_Percentage,**

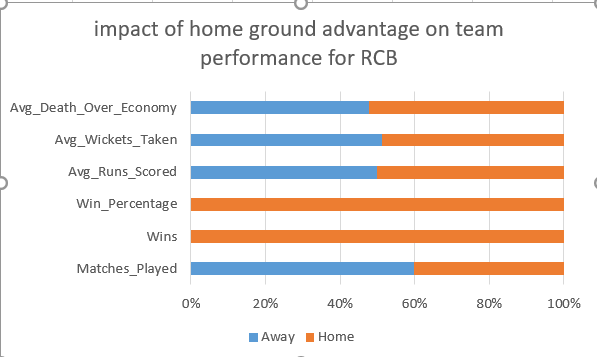
**Avg\_Runs\_Scored,**

**Avg\_Wickets\_Taken,**

**Avg\_Death\_Over\_Economy**

**FROM**

**team\_home\_away\_stats;**



**9. Come up with a visual and analytical analysis with the RCB past seasons performance and potential**

**reasons for them not winning a trophy.**

**Ans.**

**Query:**

**WITH rcb\_performance AS (**

**SELECT**

**m.Match\_Id,**

**m.Season\_Id,**

**m.Match\_Winner,**

**SUM(CASE WHEN (m.Team\_1 = 2 OR m.Team\_2 = 2) AND b.Striker = 2 THEN bs.Runs\_Scored ELSE 0 END) AS Runs\_Scored,**

**SUM(CASE WHEN (m.Team\_1 = 2 OR m.Team\_2 = 2) AND b.Striker != 2 THEN bs.Runs\_Scored ELSE 0 END) AS Runs\_Conceded**

**FROM**

**matches m**

**JOIN**

**ball\_by\_ball b ON m.Match\_Id = b.Match\_Id**

**JOIN**

**batsman\_scored bs ON b.Match\_Id = bs.Match\_Id AND b.Over\_Id = bs.Over\_Id AND b.Ball\_Id = bs.Ball\_Id**

**WHERE**

**m.Team\_1 = 2 OR m.Team\_2 = 2**

**GROUP BY**

**m.Match\_Id, m.Season\_Id, m.Match\_Winner**

**)**

**SELECT**

**Season\_Id,**

**COUNT(CASE WHEN Match\_Winner = 2 THEN 1 END) AS Wins,**

**COUNT(CASE WHEN Match\_Winner != 2 AND Match\_Winner IS NOT NULL THEN 1 END) AS Losses,**

**ROUND(AVG(Runs\_Scored), 2) AS Avg\_Runs\_Scored,**

**ROUND(AVG(Runs\_Conceded), 2) AS Avg\_Runs\_Conceded**

**FROM**

**rcb\_performance**

**GROUP BY**

**Season\_Id**

**ORDER BY**

**Season\_Id;**

RCB’s inability to win the IPL trophy stems from a combination of factors: over-reliance on key players, inconsistent bowling, particularly at the death, frequent middle-order collapses, and fielding lapses. Addressing these areas in the upcoming mega auction can provide them with a balanced team capable of finally lifting the trophy.

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

**Ans.** Understand the data: First, I'd examine the database schema to understand what data is available.

This typically includes tables for matches, players, teams, ball-by-ball data, and possibly player

statistics

* Define key areas of analysis: Even without specific questions, we can identify important aspects of IPL analysis- Team performance, Player performance, Tournament trends, Match statistics, Advantages of home ground.
* Come up with some of the players who would be best fit for the team and some of parameters that should be focused while selecting the players
* Dive deeper into specific areas: Based on the initial findings, I'd dive deeper into interesting trends or patterns: 1) Team analysis: Performance over seasons, home vs away performance, player retention impact. 2) Player analysis: Consistency, performance in different phases of the game, impact players. 3) Match analysis: Factors affecting match outcomes (toss, venue, etc.), close matches. 4) Economic analysis: Player auction trends, team budget utilization.
* Visualization: Create visualizations to better understand and communicate findings: 1) Win-loss charts for teams. 2) Performance trend lines for players. 3) Heat maps for batting/bowling performances.
* Identify areas for further research: Based on the analysis, identify questions that require more in-depth investigation or additional data.

**11. In the "Match" table, some entries in the "Opponent\_Team" column are incorrectly spelled as**

**"Delhi\_Capitals" instead of "Delhi\_Daredevils". Write an SQL query to replace all occurrences of**

**"Delhi\_Capitals" with "Delhi\_Daredevils".**

**Query:**

**update `Match `**

**Set Opponent\_Team = 'Delhi\_Daredevils'**

**where Opponent\_Team = 'Delhi\_Capitals';**

**select \* from matches**

**where Opponent\_Team = 'Delhi\_Daredevils'**