IPL Case Study

(using SQL)

Context: Let's not curb our cricket love and start analyzing the whole of IPL with this latest and complete Indian Premier League dataset. It contains the match descriptions, results, winners, player of the matches, ball by ball dataset and much more. So, stop thinking and start **analyzing**.

Content:

- This dataset consists of two separate CSV files : matches and deliveries. These files contain the information of each match summary and ball by ball details, respectively.
- I'll use Google BigQuery to treat these CSVs as tables and perfrom analysis using SQL.

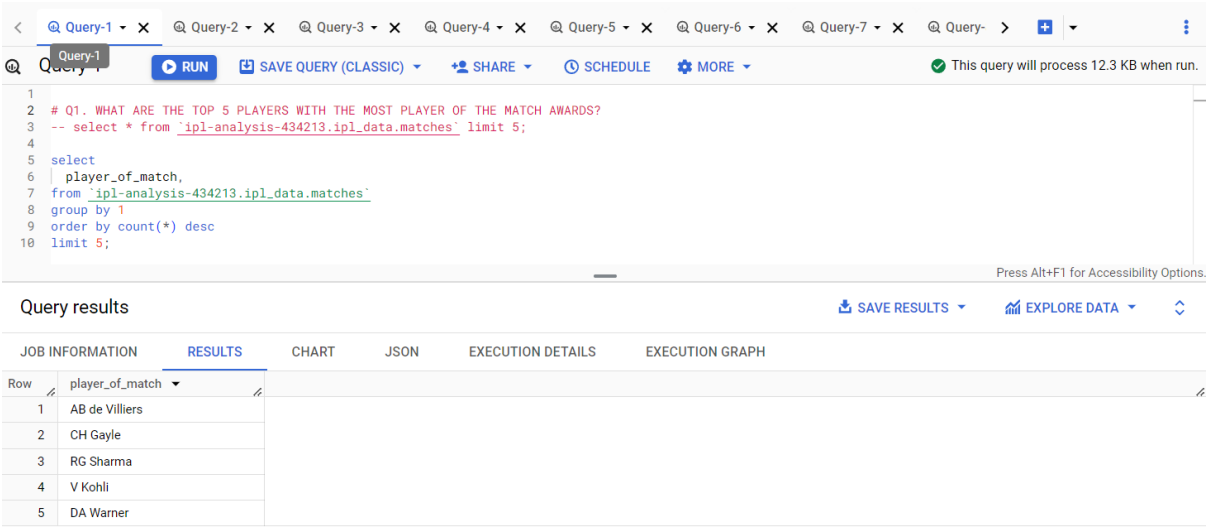
Dataset:

- Link: https://www.kaggle.com/datasets/patrickb1912/ipl-complete-dataset-20082020
- It contains data till the 2024 season of IPL.

Analysis-

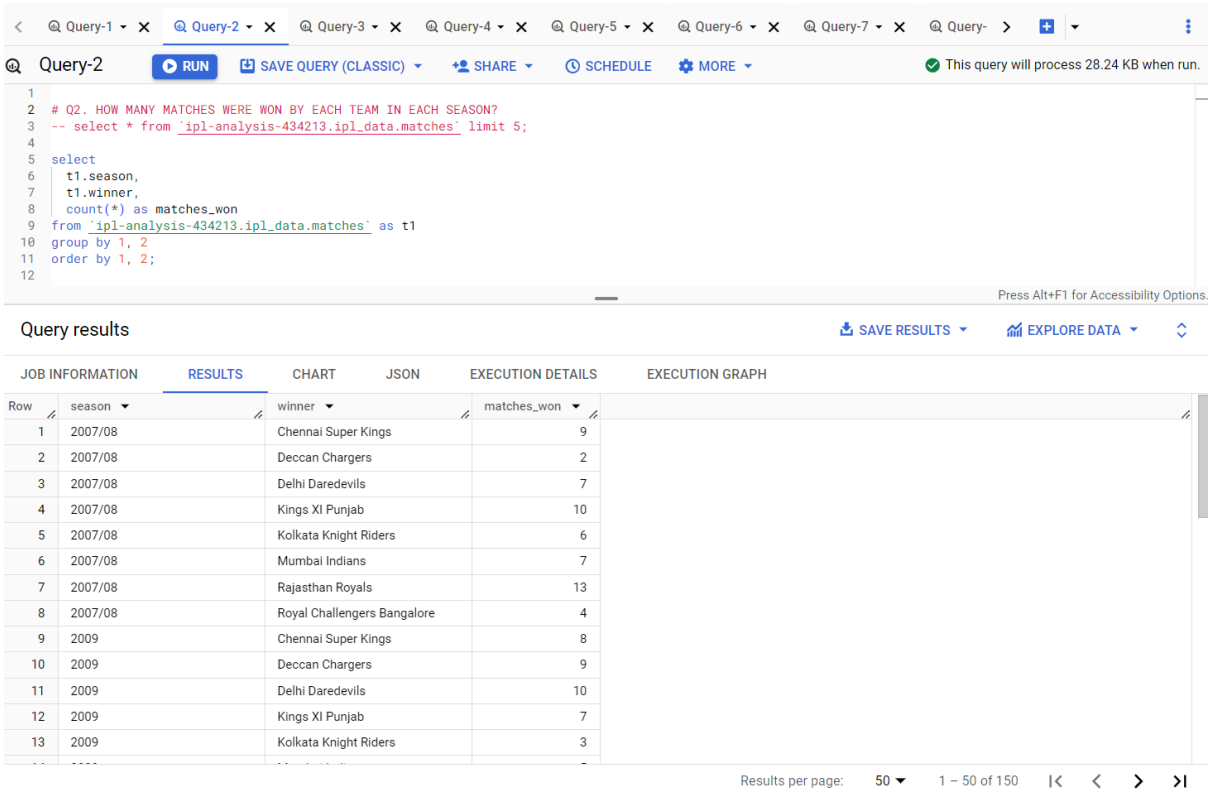
Query-1: WHAT ARE THE TOP 5 PLAYERS WITH THE MOST PLAYER OF THE MATCH AWARDS?

```
select
  player_of_match,
from `ipl-analysis-434213.ipl_data.matches`
group by 1
order by count(*) desc
limit 5;
```



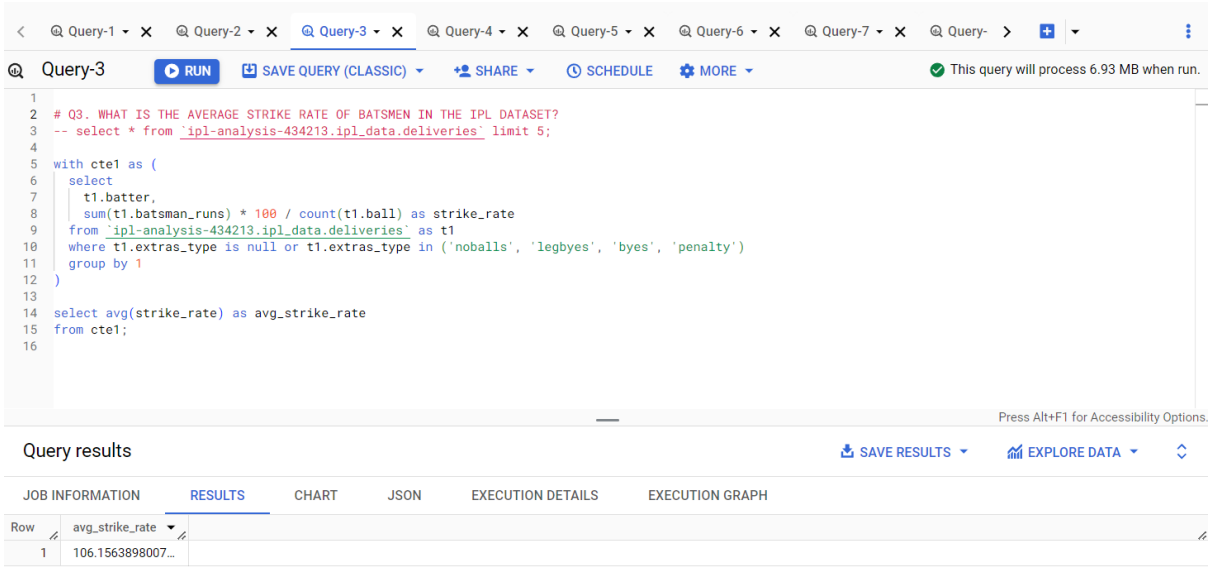
Query-2: HOW MANY MATCHES WERE WON BY EACH TEAM IN EACH SEASON?

```
select
  t1.season,
  t1.winner,
  count(*) as matches_won
from `ipl-analysis-434213.ipl_data.matches` as t1
group by 1, 2
order by 1, 2;
```



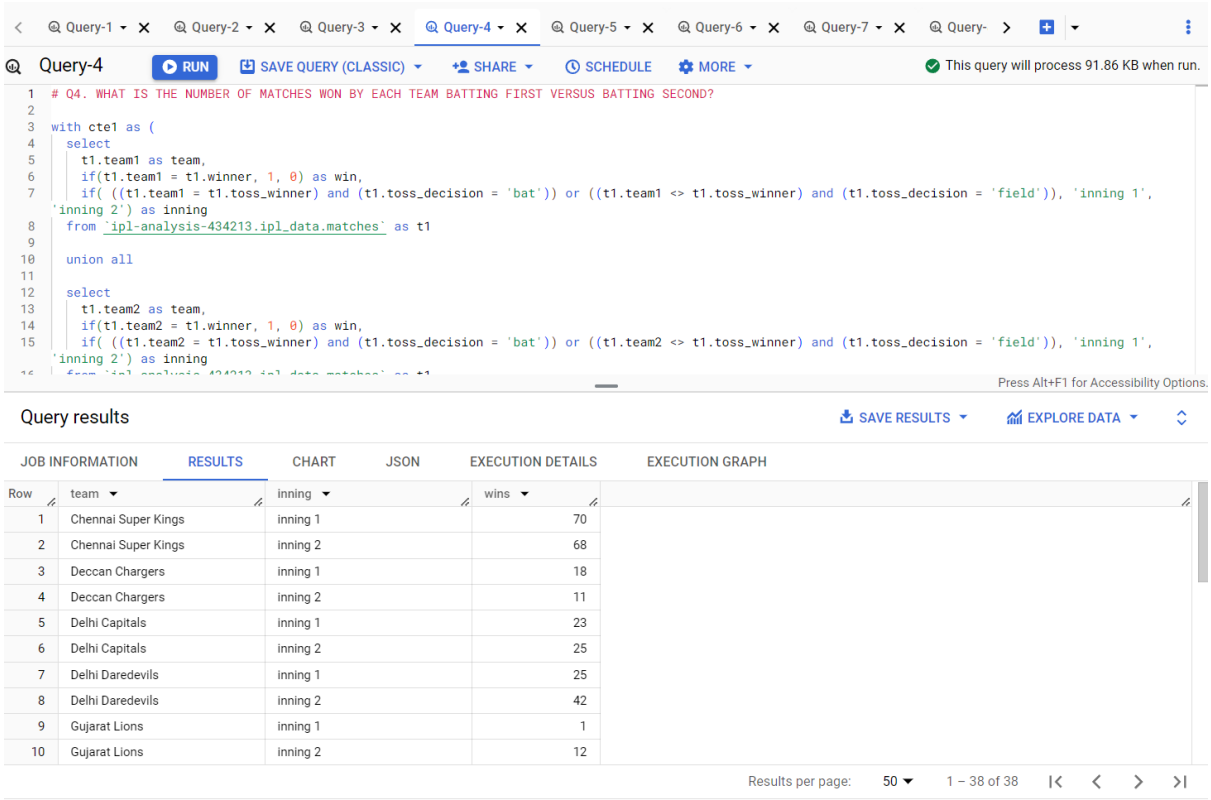
Query-3: WHAT IS THE AVERAGE STRIKE RATE OF BATSMEN IN THE IPL DATASET?

```
with cte1 as (
  select
    t1.batter,
    sum(t1.batsman_runs) * 100 / count(t1.ball) as strike_rate
  from `ipl-analysis-434213.ipl_data.deliveries` as t1
  where t1.extras_type is null or t1.extras_type in ('noballs', 'legbyes', 'byes', 'penalty')
  group by 1
)

select avg(strike_rate) as avg_strike_rate
from cte1;
```



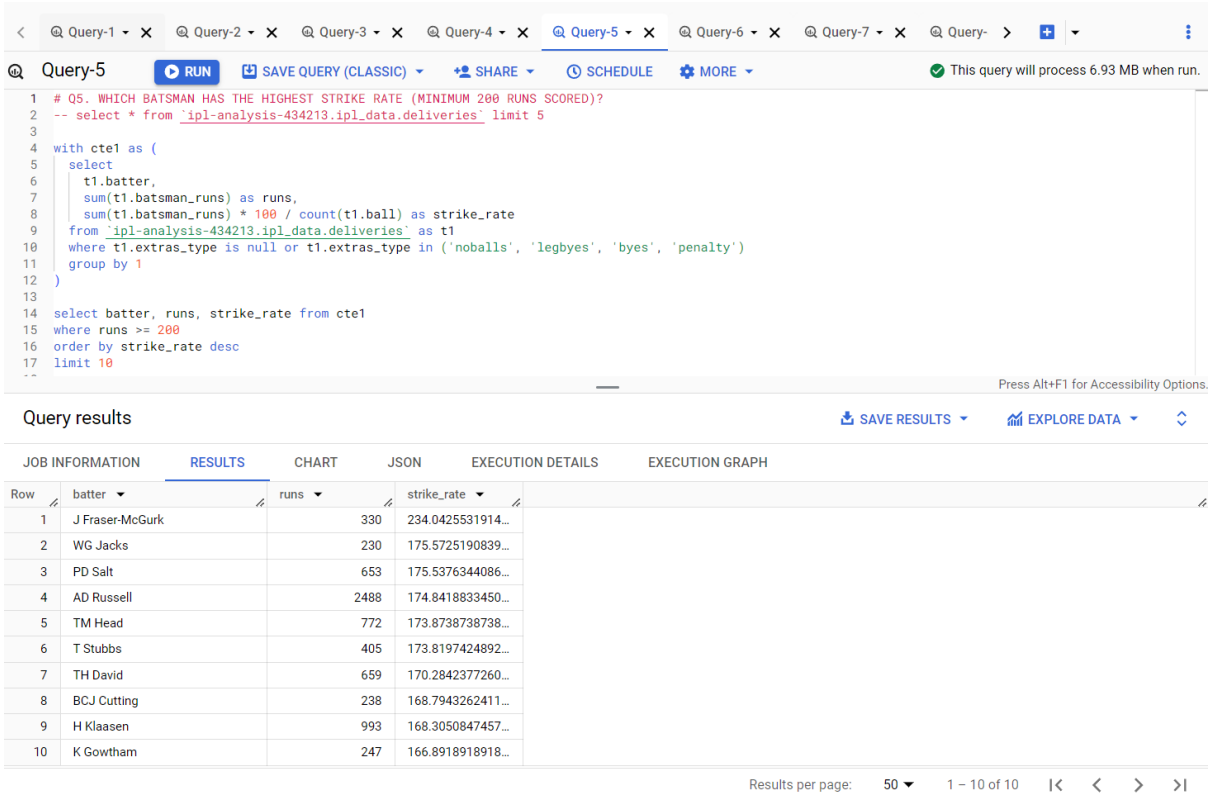
Query-4: # WHAT IS THE NUMBER OF MATCHES WON BY EACH TEAM BATTING FIRST VERSUS BATTING SECOND?

```
with cte1 as (
  select
    t1.team1 as team,
    if(t1.team1 = t1.winner, 1, 0) as win,
    if( ((t1.team1 = t1.toss_winner) and (t1.toss_decision = 'bat')) or ((t1.team1 <> t1.toss_winner) and
(t1.toss_decision = 'field')), 'inning 1', 'inning 2') as inning
  from `ipl-analysis-434213.ipl_data.matches` as t1

  union all

  select
    t1.team2 as team,
    if(t1.team2 = t1.winner, 1, 0) as win,
    if( ((t1.team2 = t1.toss_winner) and (t1.toss_decision = 'bat')) or ((t1.team2 <> t1.toss_winner) and
(t1.toss_decision = 'field')), 'inning 1', 'inning 2') as inning
  from `ipl-analysis-434213.ipl_data.matches` as t1
)

select
  c1.team, c1.inning, sum(win) as wins
from cte1 as c1
group by 1, 2
order by 1, 2;
```



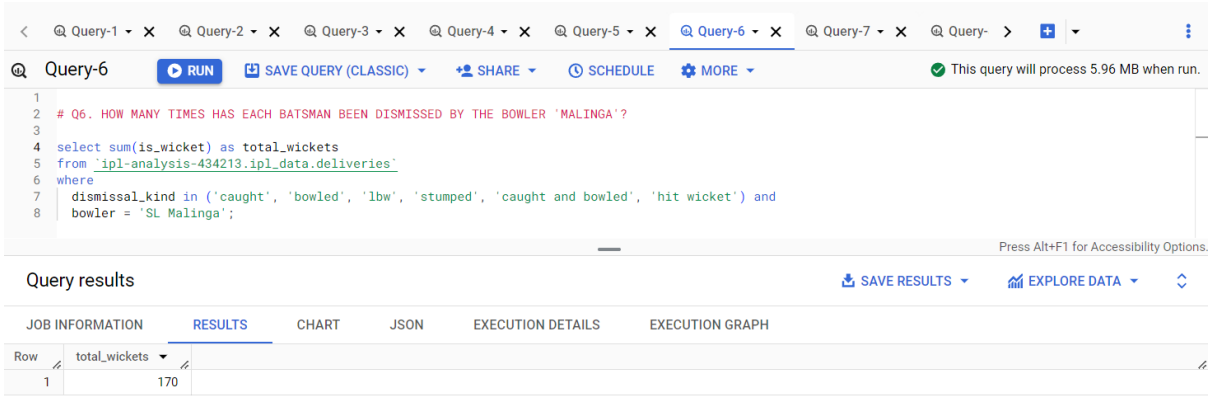
Query-5: WHICH BATSMAN HAS THE HIGHEST STRIKE RATE (MINIMUM 200 RUNS SCORED)?

```
with cte1 as (
  select
    t1.batter,
    sum(t1.batsman_runs) as runs,
    sum(t1.batsman_runs) * 100 / count(t1.ball) as strike_rate
  from `ipl-analysis-434213.ipl_data.deliveries` as t1
  where t1.extras_type is null or t1.extras_type in ('noballs', 'legbyes', 'byes', 'penalty')
  group by 1
)

select batter, runs, strike_rate from cte1
where runs >= 200
order by strike_rate desc
limit 10;
```



Query-6: HOW MANY TIMES HAS EACH BATSMAN BEEN DISMISSED BY THE BOWLER 'MALINGA'?

```
select sum(is_wicket) as total_wickets
from `ipl-analysis-434213.ipl_data.deliveries`
where
  dismissal_kind in ('caught', 'bowled', 'lbw', 'stumped', 'caught and bowled', 'hit wicket') and
  bowler = 'SL Malinga';
```

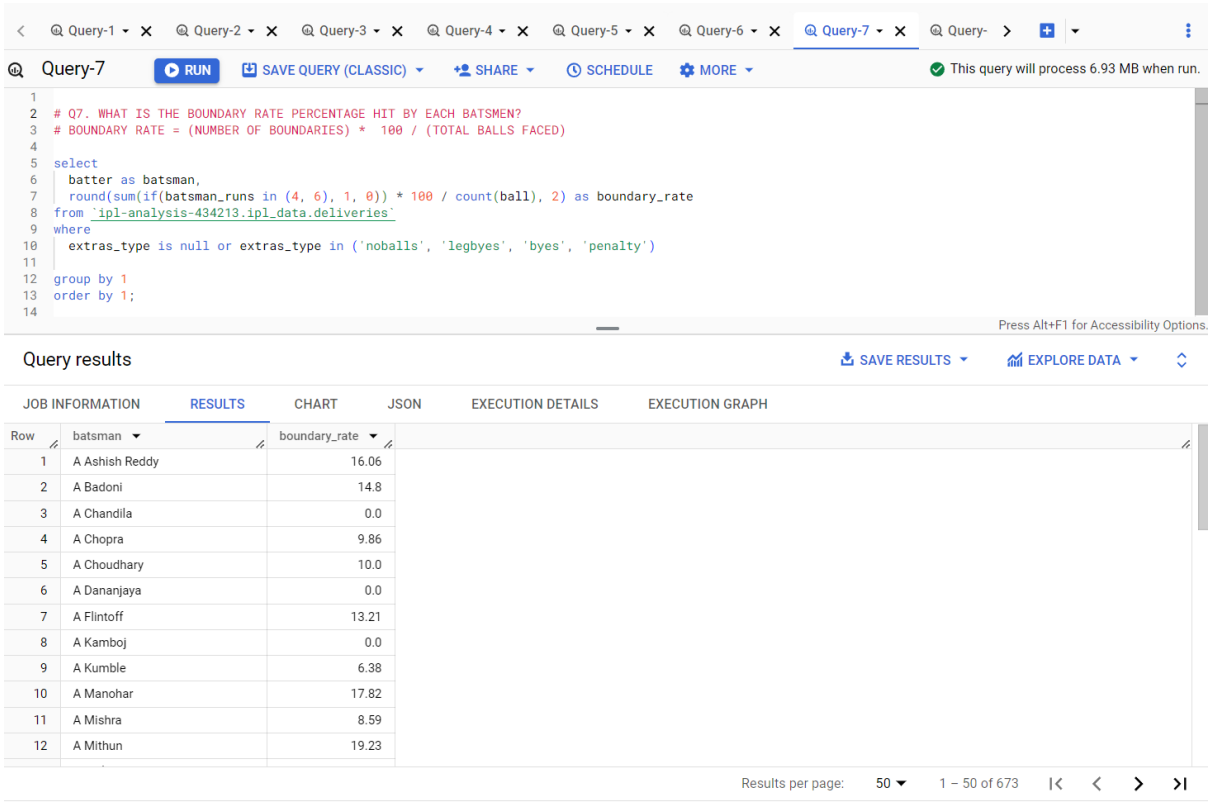


Query-7: WHAT IS THE BOUNDARY RATE PERCENTAGE HIT BY EACH BATSMEN?

BOUNDARY RATE = (NUMBER OF BOUNDARIES) * 100 / (TOTAL BALLS FACED)

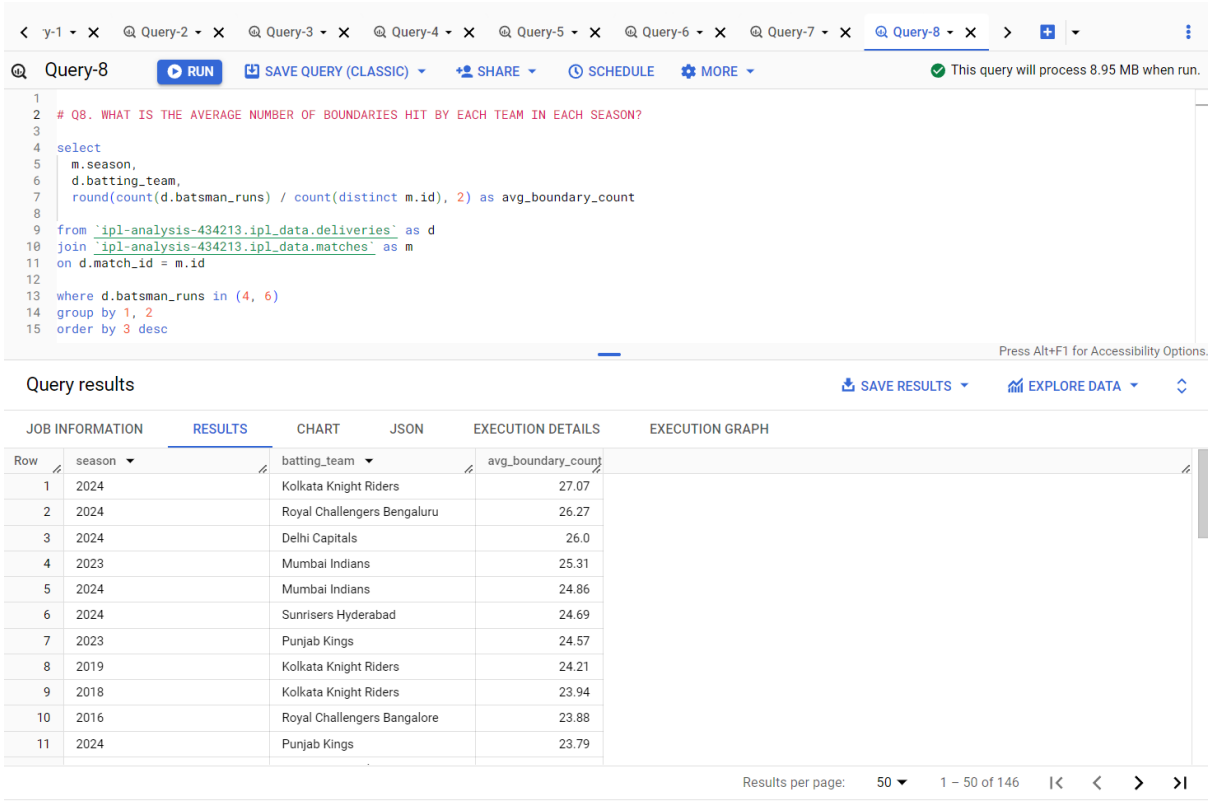
```
select
  batter as batsman,
  round(sum(if(batsman_runs in (4, 6), 1, 0)) * 100 / count(ball), 2) as boundary_rate
from `ipl-analysis-434213.ipl_data.deliveries`
```

```
where
  extras_type is null or extras_type in ('noballs', 'legbyes', 'byes', 'penalty')

group by 1
order by 1;
```



Query-8: WHAT IS THE AVERAGE NUMBER OF BOUNDARIES HIT BY EACH TEAM IN EACH SEASON?

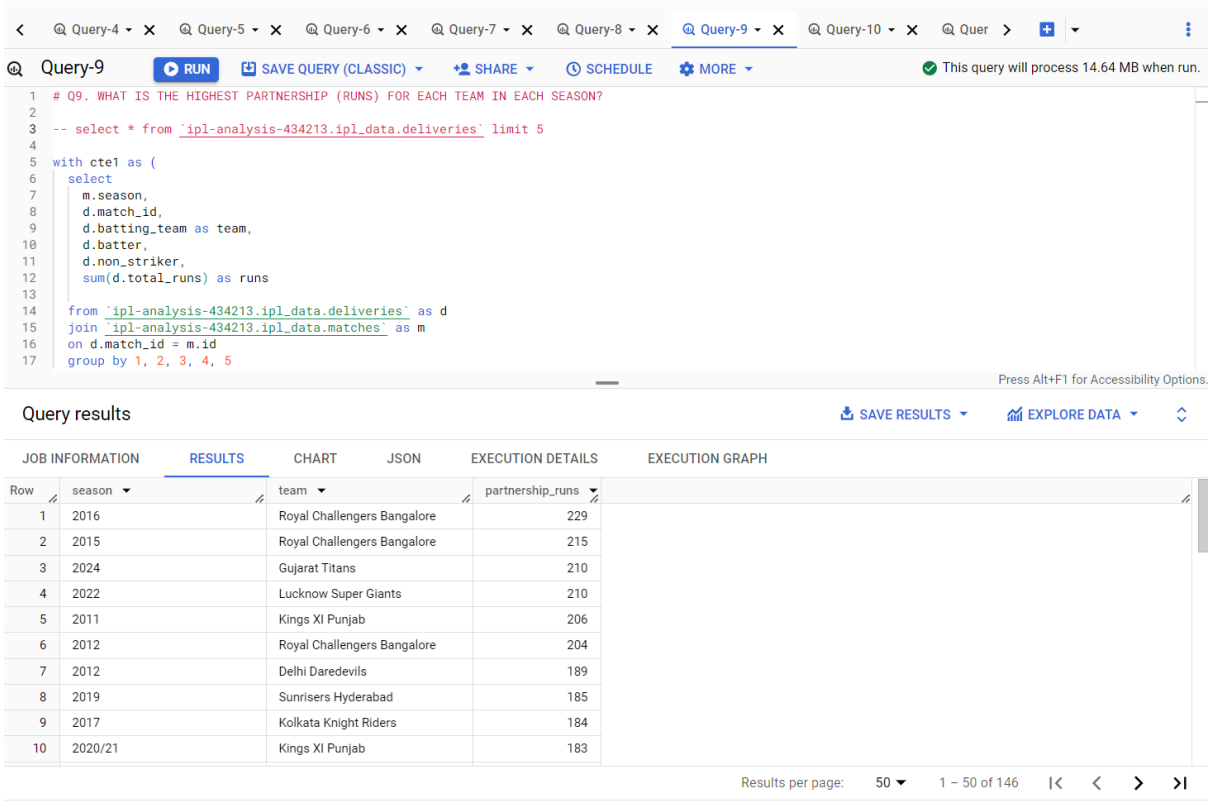
```
select
  m.season,
  d.batting_team,
  round(count(d.batsman_runs) / count(distinct m.id), 2) as avg_boundary_count

from `ipl-analysis-434213.ipl_data.deliveries` as d
join `ipl-analysis-434213.ipl_data.matches` as m
on d.match_id = m.id

where d.batsman_runs in (4, 6)
group by 1, 2
order by 3 desc;
```



Query-9: WHAT IS THE HIGHEST PARTNERSHIP (RUNS) FOR EACH TEAM IN EACH SEASON?

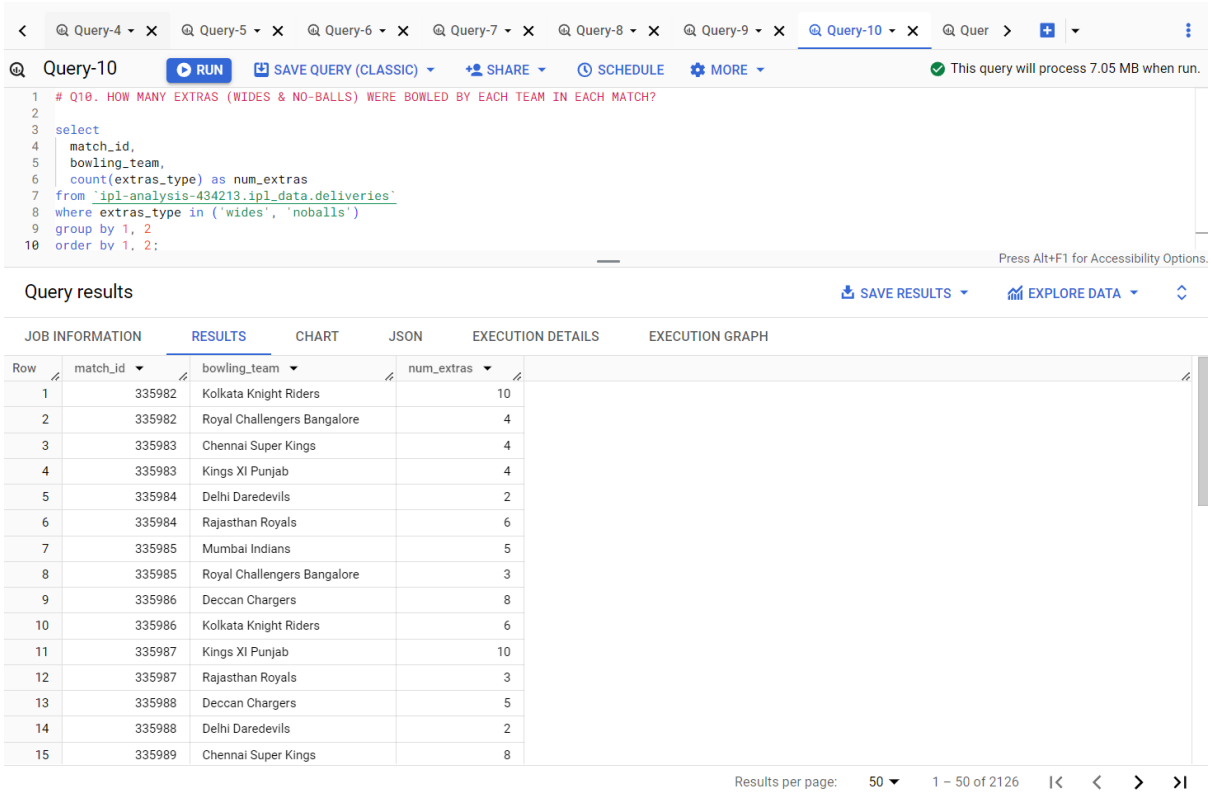
```
with cte1 as (
  select
    m.season,
    d.match_id,
    d.batting_team as team,
    d.batter,
    d.non_striker,
    sum(d.total_runs) as runs

  from `ipl-analysis-434213.ipl_data.deliveries` as d
  join `ipl-analysis-434213.ipl_data.matches` as m
  on d.match_id = m.id
  group by 1, 2, 3, 4, 5
),
cte2 as (
  select
    c1.season,
    c1.team,
    c1.batter as player1,
    c1.non_striker as player2,
    (c1.runs + c2.runs) as partnership_runs
  from cte1 as c1
  join cte1 as c2
  on c1.match_id = c2.match_id and c1.batter = c2.non_striker and c1.non_striker = c2.batter
)

select
```

```
 season, team, max(partnership_runs) as partnership_runs

from cte2
group by 1, 2
order by 3 desc;
```



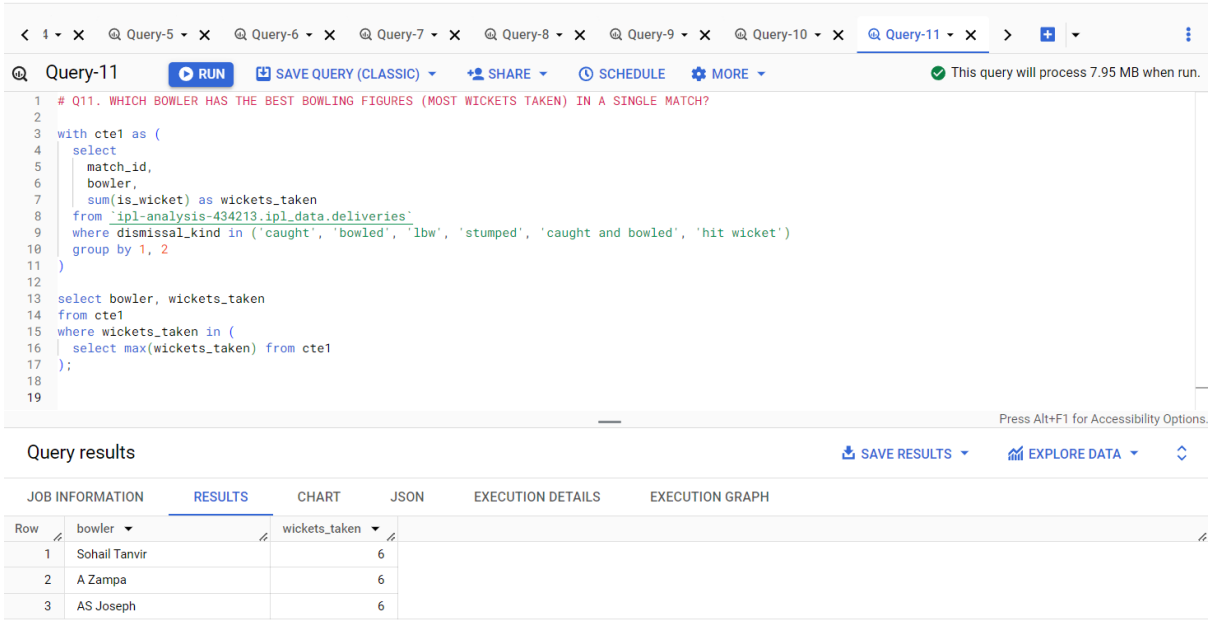
Query-10: HOW MANY EXTRAS (WIDES & NO-BALLS) WERE BOWLED BY EACH TEAM IN EACH MATCH?

```
select
  match_id,
  bowling_team,
  count(extras_type) as num_extras
from `ipl-analysis-434213.ipl_data.deliveries`
where extras_type in ('wides', 'noballs')
group by 1, 2
order by 1, 2;
```



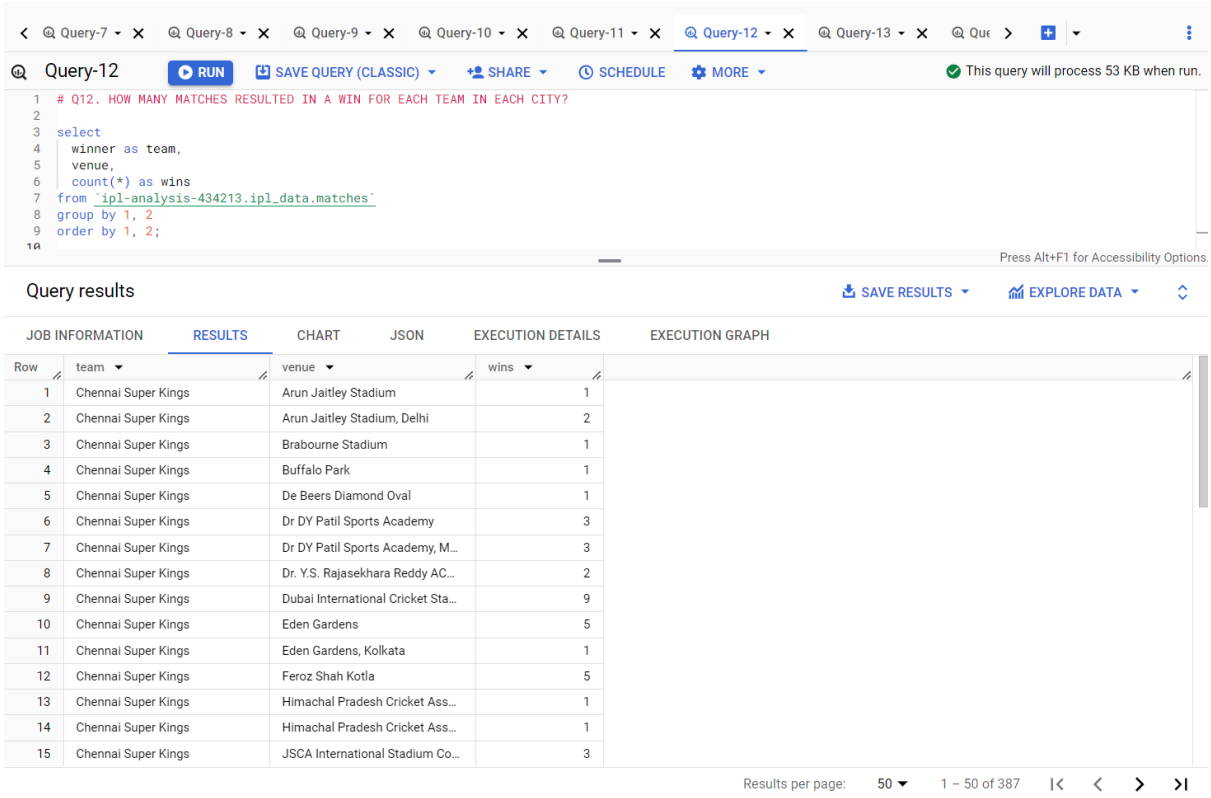
Query-11: WHICH BOWLER HAS THE BEST BOWLING FIGURES (MOST WICKETS TAKEN) IN A SINGLE MATCH?

```
with cte1 as (
  select
    match_id,
    bowler,
    sum(is_wicket) as wickets_taken
  from `ipl-analysis-434213.ipl_data.deliveries`
  where dismissal_kind in ('caught', 'bowled', 'lbw', 'stumped', 'caught and bowled', 'hit wicket')
  group by 1, 2
)

select bowler, wickets_taken
from cte1
where wickets_taken in (
  select max(wickets_taken) from cte1
);
```



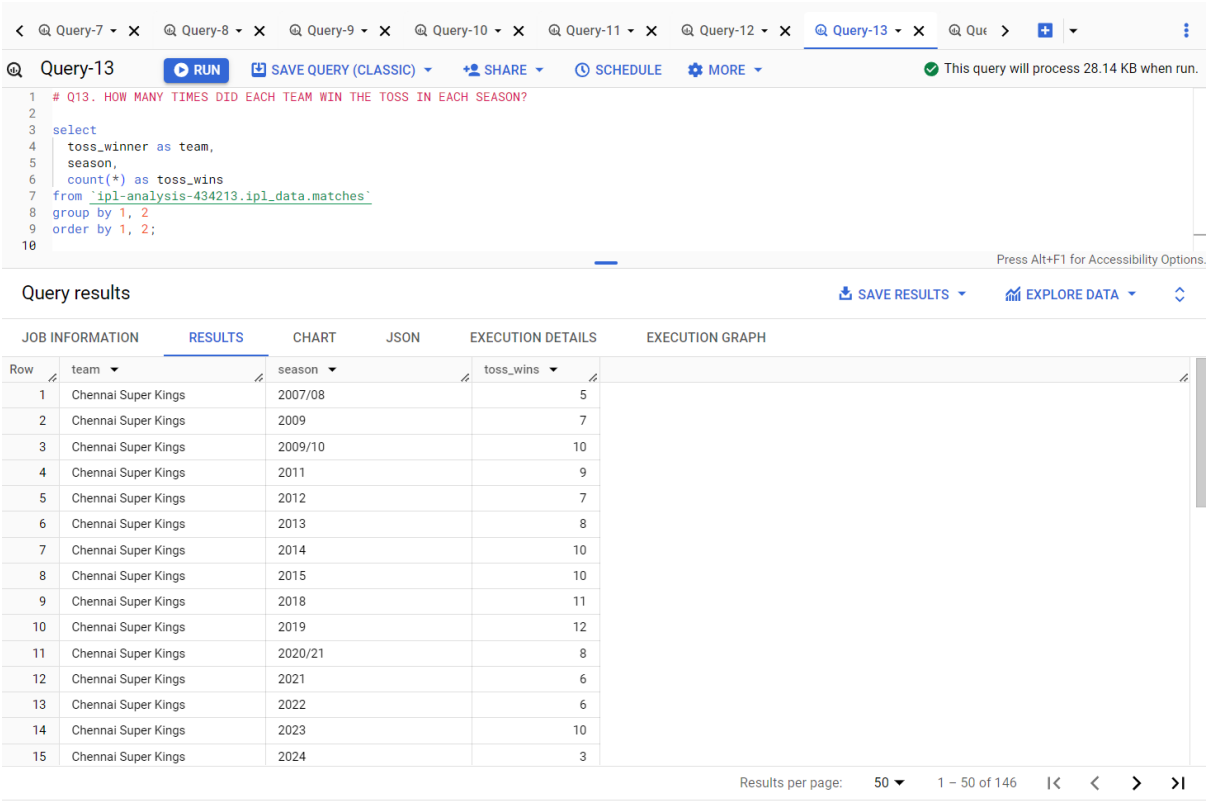
Query-12: HOW MANY MATCHES RESULTED IN A WIN FOR EACH TEAM IN EACH CITY?

```
select
  winner as team,
  venue,
  count(*) as wins
from `ipl-analysis-434213.ipl_data.matches`
group by 1, 2
order by 1, 2;
```



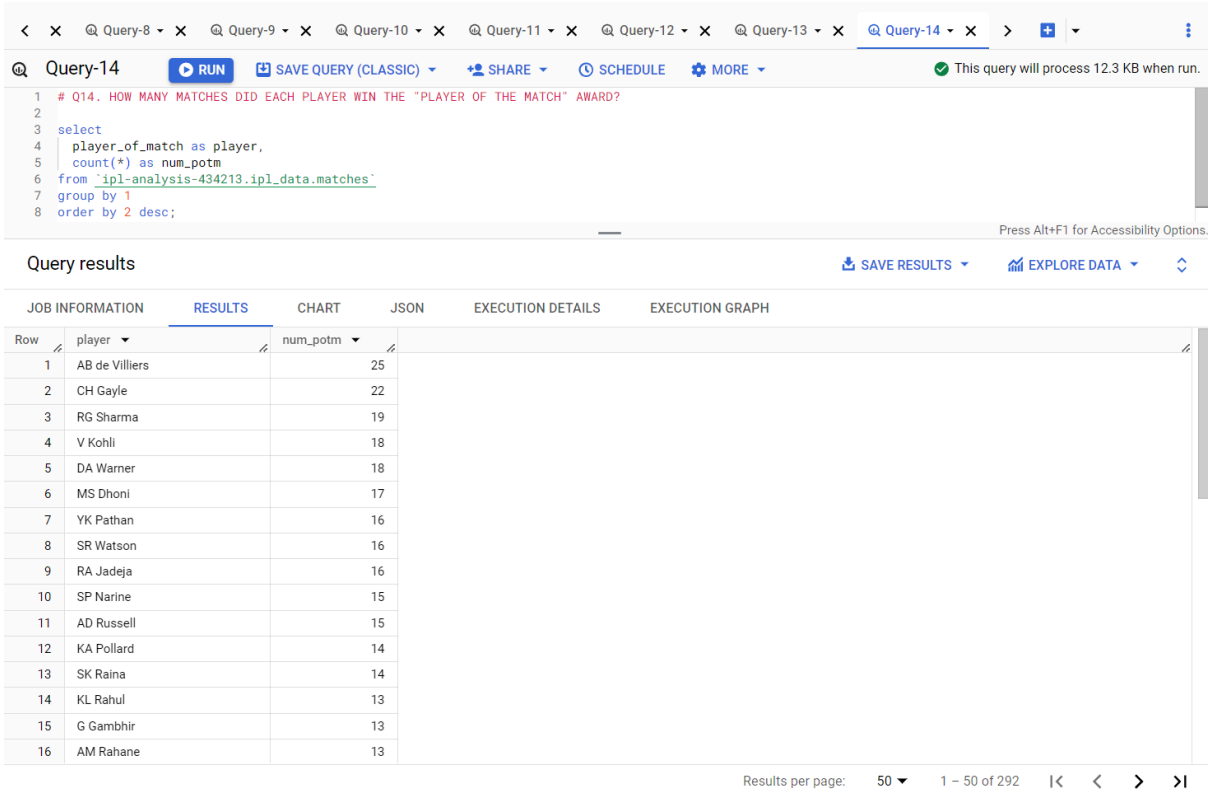
Query-13: HOW MANY TIMES DID EACH TEAM WIN THE TOSS IN EACH SEASON?

```
select
  toss_winner as team,
  season,
  count(*) as toss_wins
from `ipl-analysis-434213.ipl_data.matches`
group by 1, 2
order by 1, 2;
```



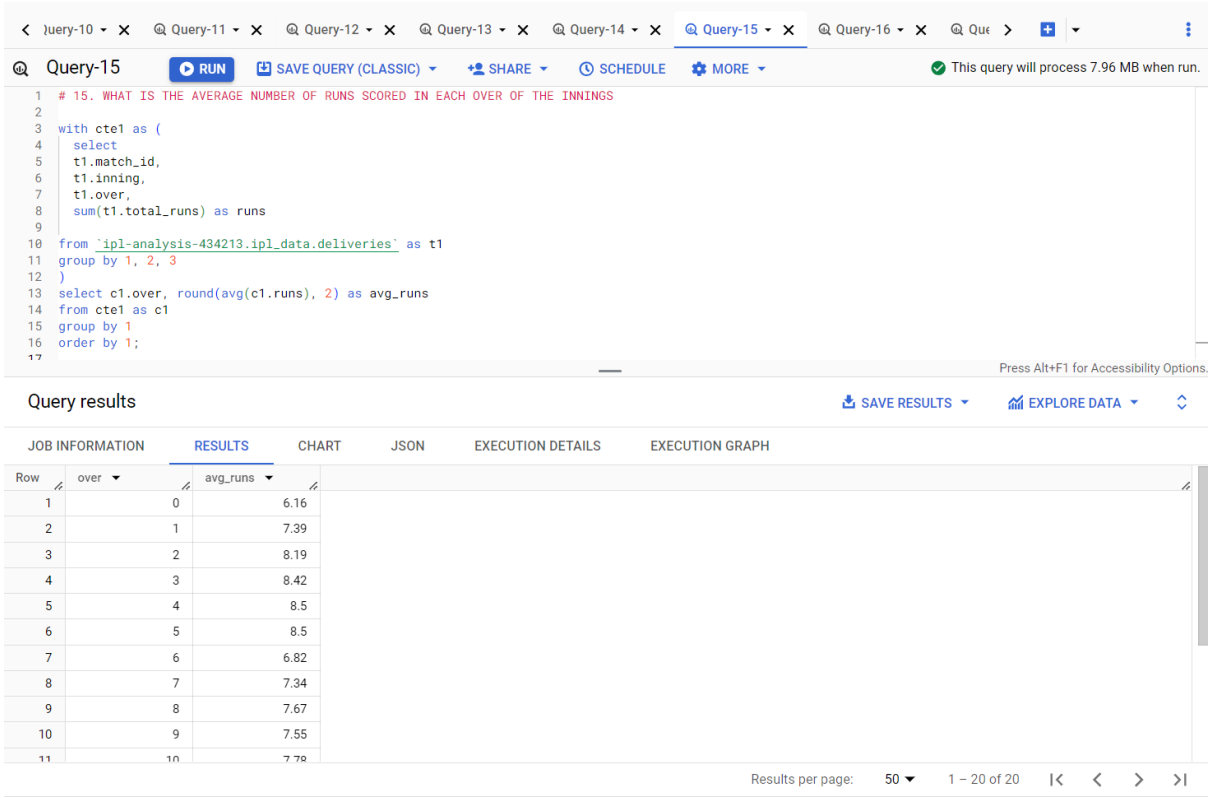
Query-14: HOW MANY MATCHES DID EACH PLAYER WIN THE "PLAYER OF THE MATCH" AWARD?

```
select
  player_of_match as player,
  count(*) as num_potm
from `ipl-analysis-434213.ipl_data.matches`
group by 1
order by 2 desc;
```



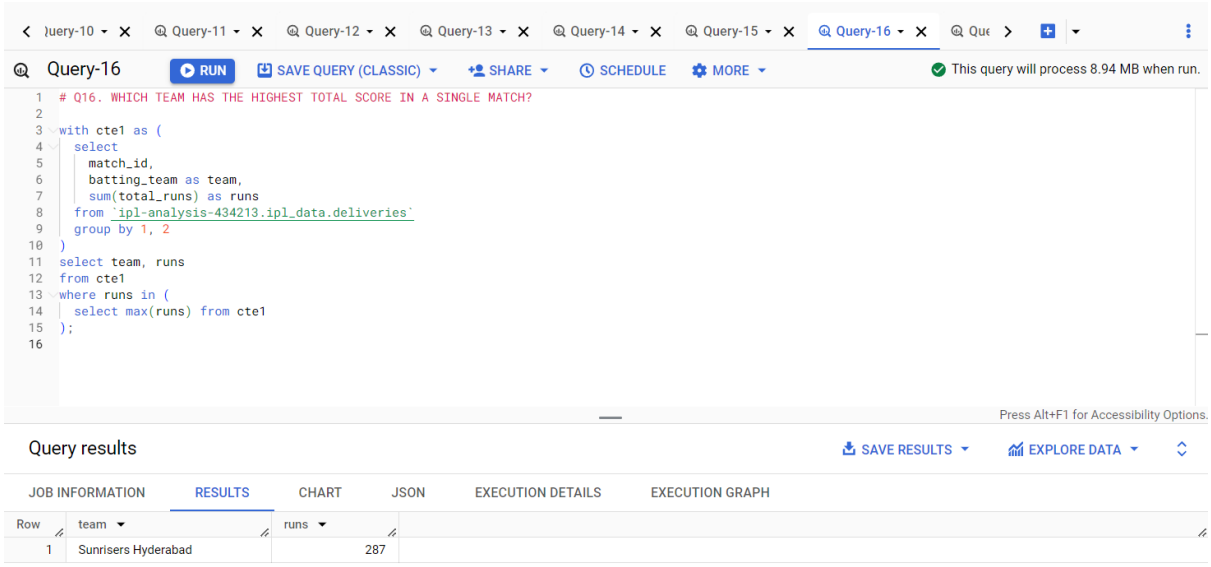
Query-15: WHAT IS THE AVERAGE NUMBER OF RUNS SCORED IN EACH OVER OF THE INNINGS

```
with cte1 as (
  select
  t1.match_id,
  t1.inning,
  t1.over,
  sum(t1.total_runs) as runs

from `ipl-analysis-434213.ipl_data.deliveries` as t1
group by 1, 2, 3
)
select c1.over, round(avg(c1.runs), 2) as avg_runs
from cte1 as c1
group by 1
order by 1;
```



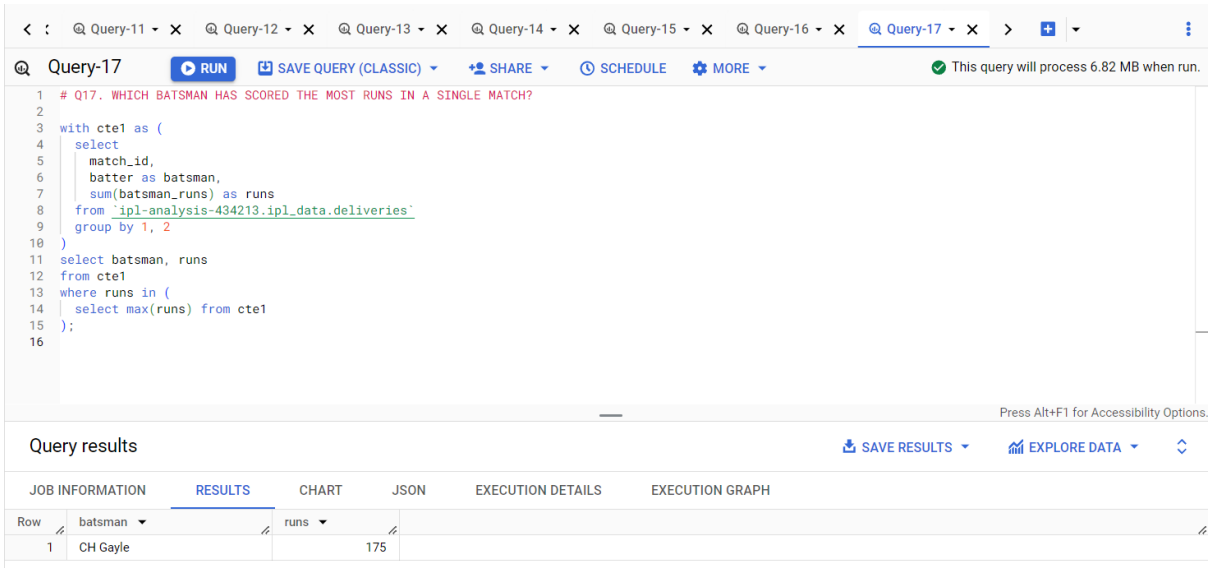
Query-16: WHICH TEAM HAS THE HIGHEST TOTAL SCORE IN A SINGLE MATCH?

```
with cte1 as (
  select
    match_id,
    batting_team as team,
    sum(total_runs) as runs
  from `ipl-analysis-434213.ipl_data.deliveries`
  group by 1, 2
)
select team, runs
from cte1
where runs in (
  select max(runs) from cte1
);
```



Query-17: WHICH BATSMAN HAS SCORED THE MOST RUNS IN A SINGLE MATCH?

```
with cte1 as (
  select
    match_id,
    batter as batsman,
    sum(batsman_runs) as runs
  from `ipl-analysis-434213.ipl_data.deliveries`
  group by 1, 2
)
select batsman, runs
from cte1
where runs in (
  select max(runs) from cte1
);
```



Insights:

1. **Team Performance by Batting Order:**
    - **Batting First:** Mumbai Indians (72), Chennai Super Kings (70) lead in wins.
    - **Batting Second:** Kolkata Knight Riders (78), Mumbai Indians (72) lead in wins.
2. **Seasonal Wins:**
    - Mumbai Indians (2013) and Rajasthan Royals (2007-08) hold the record for most wins in a single season with 13.
3. **Strike Rates:**
    - Average strike rate across all batsmen is 106.
    - Top performers include J Fraser-McGurk (234), WG Jacks (175.57), and PD Salt (175.54).
4. **Boundary Rates:**
    - J Fraser-McGurk (42.55%) and PD Salt (29.03%) top the charts for boundary percentages.
5. **Team Boundary Performance:**
    - Kolkata Knight Riders (27.07 in 2024) and Royal Challengers Bangalore (26.27 in 2024) had the highest boundary averages in a season.
6. **Highest Partnerships:**
    - Royal Challengers Bangalore holds the top spots with 229 in 2016 and 215 in 2015.
7. **Best Bowling Figures:**
    - Sohail Tanvir, Adam Zampa, and Alzarri Joseph have all taken 6 wickets in a match.
8. **City-Specific Wins:**
    - Kolkata Knight Riders (45 in Eden Gardens) have the most wins at a single venue.
9. **Player of the Match Awards:**
    - AB de Villiers (25) and Chris Gayle (22) lead in Player of the Match awards.
10. **High Scoring Overs:**
    - The 20th, 19th, and 18th overs are the most productive in terms of runs.
11. **Highest Team Totals:**
    - Sunrisers Hyderabad recorded the highest team total with 287 in a match.
12. **Highest Individual Score:**
    - Chris Gayle's 175 is the highest individual score in a match.

Actionable Recommendations:

1. **Strategize Around Batting Order:**
    - Teams should analyze their batting order effectiveness, considering the success of Mumbai Indians and Kolkata Knight Riders in specific batting scenarios.
2. **Focus on Key Players for Strike Rate:**
    - Invest in players with consistently high strike rates and boundary percentages like J Fraser-McGurk and WG Jacks to boost team totals.
3. **Maximize Boundary Opportunities:**
    - Teams should strategize to increase boundary rates, particularly in the overs identified as most productive, to maximize their scoring potential.
4. **Venue-Specific Strategies:**

    - Develop tailored strategies for different venues, especially focusing on maximizing performance in historically successful locations like Eden Gardens for Kolkata Knight Riders.
5. **Bowler Utilization:**
    - Bowlers who have demonstrated the ability to take multiple wickets in a single match should be strategically utilized during critical overs.
6. **Leverage High-Performing Partnerships:**
    - Teams should focus on building strong partnerships, which have historically led to high team totals and match victories.
7. **Recognize and Reward Top Performers:**
    - Players with a history of winning Player of the Match awards should be key players in high-pressure matches, given their proven impact on game outcomes.