



IPL SQL DATA ANALYSIS (2021-2023)

Unveiling Trends & Predictions

Presented by: Pakhi Singh Tak

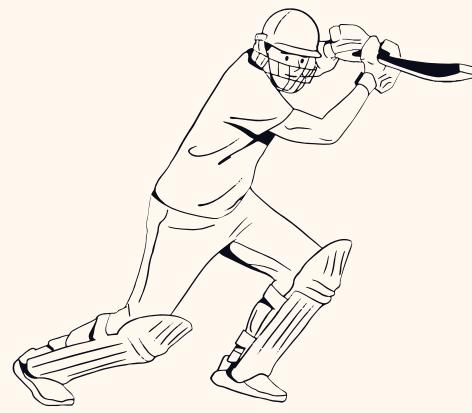


PROJECT OVERVIEW



- Sports Basics, a growing sports blog company, wanted to make an impact with IPL 2024.
- Chief Editor Tony Sharma believed analytics could create buzz.
- He approached me, a data driven cricket fan.
- Together, we created a **SQL Powered IPL Analysis** based on the last 3 years of data.

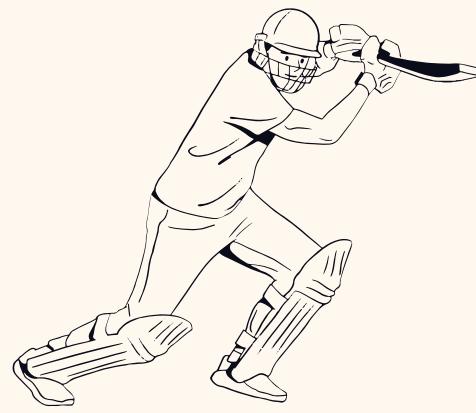
Facts, Form & Fantasy - Delivered through Data



PRIMARY ANALYSIS

Present SQL-backed insights for fans, teams & analysts

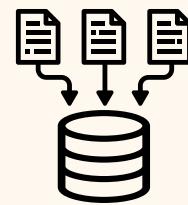
- Top 10 batsmen based on total runs scored in the last 3 years.
- Top 10 batsmen based on batting average in the last 3 years (min 60 balls faced)
- Top 10 batsmen based on strike rate in the last 3 years (min 60 balls faced)
- Top 10 bowlers based on total wickets taken in the last 3 years.
- Top 10 bowlers based on bowling average in the last 3 years (min 60 balls)
- Top 10 bowlers based on economy rate in the last 3 years (min 60 balls)
- Top 5 batsmen based on boundary percentage (fours and sixes) in the last 3 years.
- Top 5 bowlers based on dot ball % in the last 3 years.
- Top 4 teams based on winning % in the last 3 years.
- Top 2 teams with the highest number of wins achieved by chasing targets over the last 3 years.



SECONDARY ANALYSIS

Present SQL-backed insights for fans, teams & analysts

- **Prediction for Orange and Purple Cap players in the 2024 season.**
- **Prediction of the top 4 qualifying teams for 2024.**
- **Prediction of the winner and runner-up for 2024.**
- **Selection of the Best 11 players for your team based on positions, 3 years' performance data, and additional research.**
- **Selection of the top 3 all-rounders for your team.**



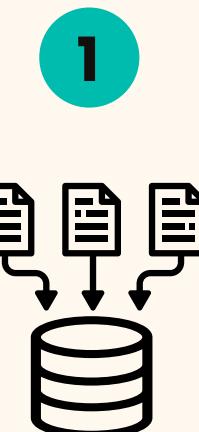
DATA OVERVIEW



The analysis utilizes 4 datasets:

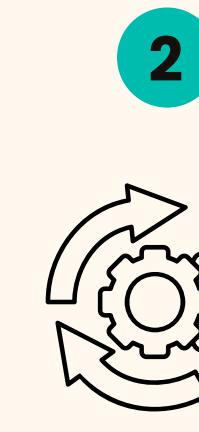
- **dim_match_summary:**
Team level match results
- **dim_players:**
Player info & roles
- **fact_bating_summary:**
Batting stats
- **fact_bowling_summary:**
Bowling stats

ANALYTICAL FRAMEWORK



Data Collection

Joining, filtering, and structuring with SQL



Data Transformation

Consistency in player roles, handling null values



Data Analysis

Execute queries to uncover Performance trends, wins, roles, run/wicket totals



Insight Generation

Summarize key findings, Strategic picks, top performers, predictions



KEY SQL ANALYSES



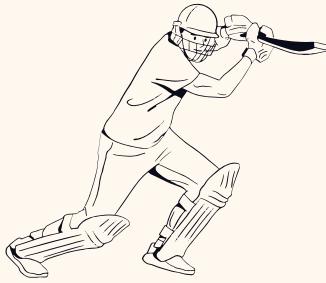
Q: Top 10 batsmen based on past 3 years total runs scored.



```
SELECT batsmanName, SUM(runs) AS total_runs
FROM fact_bating_summary
GROUP BY batsmanName
ORDER BY total_runs DESC
LIMIT 10;
```

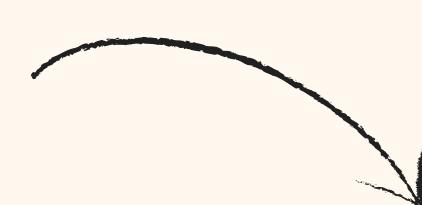


batsmanName	total_runs
ShubmanGill	1851
FafduPlessis	1831
RuturajGaikwad	1593
KLRahul	1516
JosButtler	1509
ShikharDhawan	1392
ViratKohli	1385
SanjuSamson	1304
SuryakumarYadav	1225
GlennMaxwell	1214



Q: Top 10 batsmen based on past 3 years batting average (min 60 balls faced in each season)

```
WITH s AS (
  SELECT batsmanName,
    EXTRACT(YEAR FROM STR_TO_DATE(d.matchDate, '%b %d, %Y')) AS yr,
    SUM(runs) AS r,
    SUM(balls) AS b,
    COUNT(CASE WHEN `out/not_out` = 'out' THEN 1 END) AS o
  FROM fact_bating_summary f
  JOIN dim_match_summary d ON f.match_id = d.match_id
  WHERE EXTRACT(YEAR FROM STR_TO_DATE(d.matchDate, '%b %d, %Y')) IN (2021, 2022, 2023)
  GROUP BY batsmanName, yr
  HAVING b >= 60
)
SELECT batsmanName,
  ROUND(SUM(r) / NULLIF(SUM(o), 0), 2) AS average
FROM s
GROUP BY batsmanName
HAVING COUNT(DISTINCT yr) = 3
ORDER BY average DESC
LIMIT 10;
```



batsmanName	average
KL Rahul	52.28
Faf du Plessis	43.60
David Miller	43.20
Jos Buttler	41.92
Shimron Hetmyer	40.67
Shubman Gill	40.27
Shikhar Dhawan	39.77
Ruturaj Gaikwad	38.22
David Warner	37.90
Suryakumar Yadav	35.00

Q: Top 10 batsmen based on past 3 years strike rate.



```
WITH s AS (
  SELECT batsmanName,
    EXTRACT(YEAR FROM STR_TO_DATE(d.matchDate, '%b %d, %Y')) AS yr,
    SUM(runs) AS r, SUM(balls) AS b
  FROM fact_bating_summary f
  JOIN dim_match_summary d ON f.match_id = d.match_id
  WHERE EXTRACT(YEAR FROM STR_TO_DATE(d.matchDate, '%b %d, %Y')) IN (2021, 2022, 2023)
  GROUP BY batsmanName, yr
  HAVING b >= 60
)
SELECT batsmanName, ROUND(SUM(r)*100.0 / SUM(b), 2) AS SR
FROM s
GROUP BY batsmanName
HAVING COUNT(DISTINCT yr) = 3
ORDER BY SR DESC
LIMIT 10;
```



batsmanName	SR
Glenn Maxwell	161.44
Suryakumar Yadav	160.55
Andre Russell	159.19
Shimron Hetmyer	157.27
Nicholas Pooran	157.11
Prithvi Shaw	153.20
Dinesh Karthik	152.64
Yashasvi Jaiswal	152.15
Jos Buttler	146.93
Shivam Dube	145.73

Q: Top 10 bowlers based on past 3 years total wickets taken.

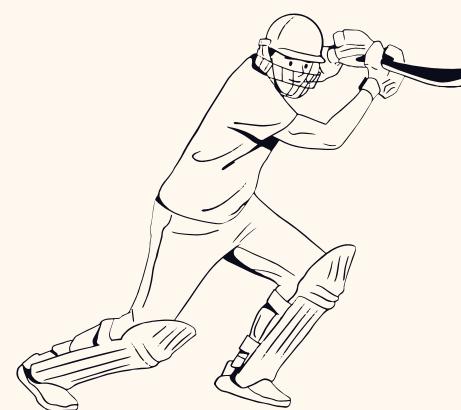
```
SELECT bowlerName, SUM(wickets) AS total_wickets
FROM fact_bowling_summary f
JOIN dim_match_summary d ON f.match_id = d.match_id
WHERE EXTRACT(YEAR FROM STR_TO_DATE(d.matchDate, '%b %d, %Y')) IN (2021, 2022, 2023)
GROUP BY bowlerName
ORDER BY total_wickets DESC
LIMIT 10;
```



bowlerName	total_wickets
MohammedShami	67
YuzvendraChahal	66
HarshalPatel	65
RashidKhan	63
AveshKhan	47
KagisoRabada	45
ArshdeepSingh	45
VarunChakravarthy	44
ShardulThakur	43
TrentBoult	42

Q: Top 10 bowlers based on past 3 years bowling average.

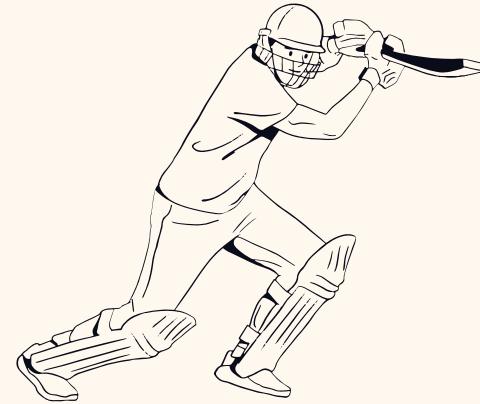
```
WITH s AS (
  SELECT bowlerName,
    EXTRACT(YEAR FROM STR_TO_DATE(d.matchDate, '%b %d, %Y')) AS yr,
    SUM(runs) AS r, SUM(overs)*6 AS balls, SUM(wickets) AS w
  FROM fact_bowling_summary f
  JOIN dim_match_summary d ON f.match_id = d.match_id
  WHERE EXTRACT(YEAR FROM STR_TO_DATE(d.matchDate, '%b %d, %Y')) IN (2021, 2022, 2023)
  GROUP BY bowlerName, yr
  HAVING balls >= 60
)
SELECT bowlerName, ROUND(SUM(r) / NULLIF(SUM(w), 0), 2) AS bowling_avg
FROM s
GROUP BY bowlerName
HAVING COUNT(DISTINCT yr) = 3
ORDER BY bowling_avg
LIMIT 10;
```



A curved arrow points from the text "Top 10 bowlers based on past 3 years bowling average." to the "bowlerName" column of the table.

bowlerName	bowling_avg
AndreRussell	18.23
YuzvendraChahal	20.20
RashidKhan	20.21
HarshalPatel	20.35
MohammedShami	20.54
AveshKhan	23.72
KagisoRabada	23.76
MoeenAli	23.86
AnrichNortje	24.77
UmranMalik	26.10

Q: Top 10 bowlers based on past 3 years economy rate.



```
SELECT bowlerName, ROUND(AVG(avg_economy), 2) AS economy
FROM (
  GROUP BY bowlerName
  HAVING COUNT(DISTINCT Season) = 3
  ORDER BY economy asc
  LIMIT 10;
```

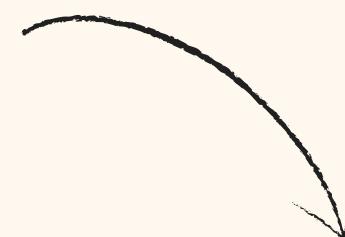


bowlerName	economy
SunilNarine	6.72
RashidKhan	7.05
AxarPatel	7.22
RavichandranAshwin	7.61
RaviBishnoi	7.61
RahulChahar	7.68
YuzvendraChahal	7.75
RavindraJadeja	7.78
BhuvneshwarKumar	7.86
MohammedShami	7.87

Q: Top 5 batsmen based on past 3 years boundary % (fours and sixes).



```
SELECT batsmanName,
       ROUND((SUM(`4s`) + SUM(`6s`)) * 100.0 / SUM(balls), 2) AS boundary_pct
  FROM fact_bating_summary f
  JOIN dim_match_summary d ON f.match_id = d.match_id
 WHERE EXTRACT(YEAR FROM STR_TO_DATE(d.matchDate, '%b %d, %Y')) IN (2021, 2022, 2023)
 GROUP BY batsmanName
 HAVING SUM(balls) > 60
 ORDER BY boundary_pct DESC
 LIMIT 5;
```



batsmanName	boundary_pct
PatCummins	30.00
RashidKhan	25.88
PhilSalt	25.56
YashasviJaiswal	25.27
PrithviShaw	24.44

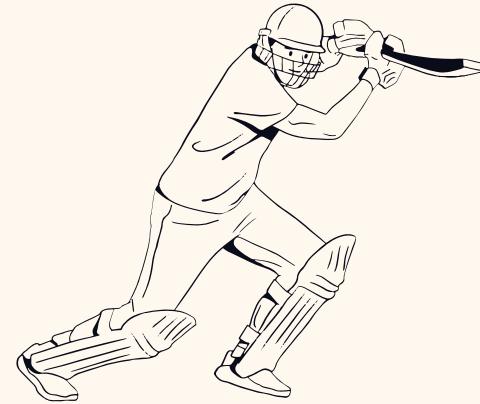
Q: Top 5 bowlers based on past 3 years dot ball %



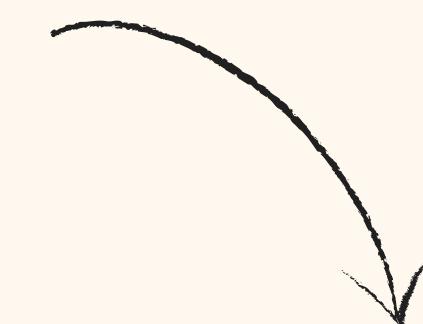
```
SELECT bowlerName,
       ROUND(SUM(`0s`) * 100.0 / (SUM(overs) * 6), 2) AS dot_pct
  FROM fact_bowling_summary f
  JOIN dim_match_summary d ON f.match_id = d.match_id
 WHERE EXTRACT(YEAR FROM STR_TO_DATE(d.matchDate, '%b %d, %Y')) IN (2021, 2022, 2023)
 GROUP BY bowlerName
 HAVING SUM(overs)*6 >= 180
 ORDER BY dot_pct DESC
 LIMIT 5;
```

bowlerName	dot_pct
MohsinKhan	50.39
PrasidhKrishna	48.12
MohammedShami	47.99
MohammedSiraj	47.71
TrentBoult	46.53

Q: Top 4 teams based on past 3 years winning %.



```
SELECT team,
       COUNT(*) AS matches_played,
       SUM(CASE WHEN team = winner THEN 1 ELSE 0 END) AS matches_won,
       ROUND(SUM(team = winner) * 100.0 / COUNT(*), 2) AS win_pct
  FROM (
    SELECT team1 AS team, winner, matchDate FROM dim_match_summary
    UNION ALL
    SELECT team2, winner, matchDate FROM dim_match_summary
  ) AS all_matches
 WHERE EXTRACT(YEAR FROM STR_TO_DATE(matchDate, '%b %d, %Y')) IN (2021, 2022, 2023)
 GROUP BY team
 ORDER BY win_pct DESC
 LIMIT 4;
```

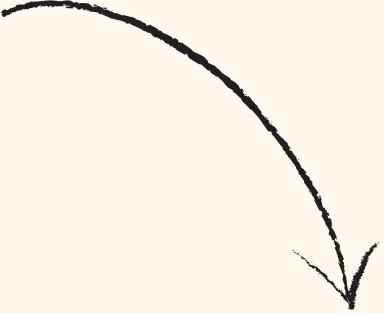


team	matches_played	matches_won	win_pct
Titans	32	23	71.88
Super Giants	29	17	58.62
RCB	45	25	55.56
Super Kings	44	24	54.55

Q: Top 4 teams based on past 3 years winning %.



```
SELECT winner, COUNT(*) AS chase_wins
FROM dim_match_summary
WHERE EXTRACT(YEAR FROM STR_TO_DATE(matchDate, '%b %d, %Y')) IN (2021, 2022, 2023)
  AND margin LIKE '%wickets'
GROUP BY winner
ORDER BY chase_wins DESC
LIMIT 2;
```



winner	chase_wins
KKR	14
Titans	14

IPL 2024 Prediction

Orange Cap



Ruturaj Gaikwad

Purple Cap



Yuzvender Chahal



Shubhman Gill



Mohammed Shami



IPL 2024 Prediction

PLAYOFFS TEAMS



WINNER



TOP 3 ALL Rounders Prediction

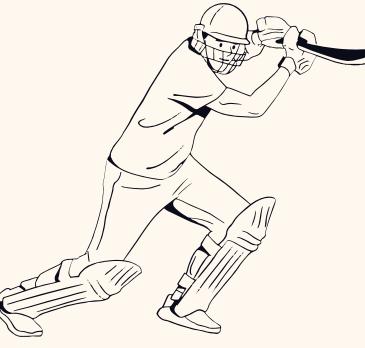
NAME	IMAGE	PLAYING ROLE	TOTAL RUNS	TOTAL WICKETS	STRIKE RATE
Andre Russell		All Rounder	745	35	130.04
Ravindra Jadeja		All Rounder	533	37	143.5
Axar Patel		Bowling All Rounder	505	30	129.53

Criteria: Allrounder should have played and bowled min 60 balls in all 3 seasons.

BEST 11 PLAYERS

NAME	PLAYING ROLE	TEAM
Shubhman Gill	Opening Batter	Gujrat Titans
Virat Kohli	Opening Batter	RCB
Shikhar Dhawan	Top order Batter	Delhi Capitals
Faf Du Plessis	Middle order Batter	RCB
KL Rahul	Wicketkeeper Batter	LSG
Rashid Khan	Bowling All rounder	Gujrat Titans
Sunil Narine	Bowling All rounder	KKR
Mohammed Shami	Bowler	Gujrat Titans
Yuzvendra Chahal	Bowler	Rajasthan Royals
Harshal Patel	Bowler	Punjab Kings
Avesh Khan	Bowler	Rajasthan Royals

Key Takeaways



- 1 **Highest Run Scorer:** **Shubhman Gill** 
- 2 **Top Wicket-Taker:** **Mohammed Shami** 
- 3 **Predicted Winner:** **Rajasthan Royals** 
- 2 **Predicted Runner-up:** **Chennai Super Kings** 
- 1 **Most Valuable All-Rounder:** **Andre Russell** 

Actionable Insights for IPL Teams



- Prioritize consistent all-rounders at auction
- Use performance data for mid-season strategy 
- Rotate players based on matchups & past data 
- Blend data & instinct to dominate IPL 2024