IPL Data Analysis

CONTENTS
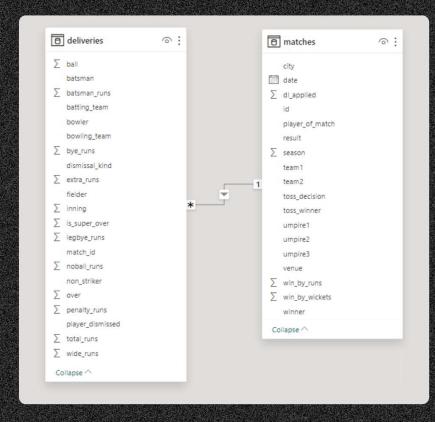
1. Problem Statement

2. Table Details

3. Data Insights

Problem Statement

1.Conduct an IPL data analysis using the "Matches" and "Deliveries" tables.

2.Utilize robust querying, data mining, and analytics techniques to extract extensive insights from the IPL dataset.

3.Focus on key aspects such as:

1. Calculating batting statistics

2. Deriving comprehensive bowling statistics

3. Profiling top player performances

4. Analyzing team-wise performance

5. Identifying impactful partnerships and individual
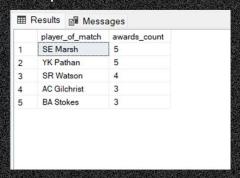
Table Details

Tables –

1. deliveries
2. matches

Data Insights

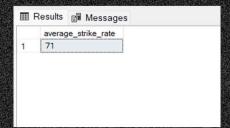Q. What are the top 5 players with the most player of the match awards?

_____

Query –

select top 5 player_of_match, count(*) as
awards_count from matches
group by player_of_match
order by awards_count desc

Output –

	player_of_match	awards_count
1	SE Marsh	5
2	YK Pathan	5
3	SR Watson	4
4	AC Gilchrist	3
5	BA Stokes	3

Q. What is the average strike rate of batsmen in the IPL dataset?
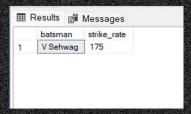
_____

Query –

select avg(strike_rate) as average_strike_rate
from (
Select batsman,
(sum(total_runs)/count(ball))*100 as strike_rate
from deliveries group by batsman) as batsman_stats

Output –

	average_strike_rate
1	71

Q. Which batsman has the highest strike rate (minimum 200
runs scored)?

_____

Query –

select top 1 batsman,
sum(batsman_runs)*100/count(*) as strike_rate from
deliveries
group by batsman
having sum(total_runs)>=200
order by strike_rate desc

Output –

	batsman	strike_rate
1	V Sehwag	175

Q. What is the percentage of boundaries (fours and sixes combined) runs by each batsman?

_____

Query –

```
with runs as (
select batsman, sum(batsman_runs) as total_runs
from deliveries
group by batsman),
bound_runs as (
select batsman, sum(batsman_runs) as boun_runs
from deliveries
where batsman_runs = 4 or batsman_runs = 6
group by batsman)

select r.batsman, r.total_runs, b.boun_runs,
b.boun_runs*100/r.total_runs as boundry_percent
from runs r
join bound_runs b on b.batsman = r.batsman
```
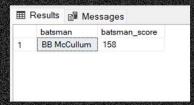
Output –

	batsman	total_runs	boun_runs	boundry_percent
1	SV Samson	386	242	62
2	ML Hayden	355	248	69
3	SB Bangar	49	22	44
4	Vishnu Vinod	19	10	52
5	K Goel	133	86	64
6	TG Southee	7	6	85
7	G Gambhir	1047	618	59
8	MEK Hussey	168	114	67
9	S Badrinath	199	132	66
10	RS Bopara	121	66	54
11	MM Sharma	51	22	43
12	M Rawat	38	18	47
13	AC Gilchrist	555	410	73
14	SP Narine	224	196	87
15	EJG Morgan	65	36	55
16	N Rana	333	198	59
17	SP Fleming	196	126	64
18	S Aravind	32	8	25
19	JH Kallis	300	180	60
20	DB Ravi Teja	126	78	61
21	KV Sharma	50	32	64
22	B Lee	17	14	82
23	DB Das	103	66	64
24	AB Agarkar	62	38	61
25	TM Dilshan	154	114	74

Q. Which batsman has scored the most runs in a match?

_____

Query –

select top 1 batsman, sum(batsman_runs) as
batsman_score from deliveries
group by match_id, inning, batsman
order by batsman_score desc

Output –

	batsman	batsman_score
1	BB McCullum	158

Results · Messages

Q. Which are the top 5 batsman with maximum sixes ?

_____

Query –
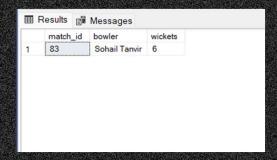
select top 5 batsman, sum(case when
batsman_runs = 6 then 1 else 0 end) as no_sixes
from deliveries
group by batsman
order by no_sixes desc

Output –

	batsman	no_sixes
1	RG Sharma	35
2	YK Pathan	35
3	BB McCullum	35
4	MS Dhoni	34
5	SK Raina	33

Q. Which bowler has the best bowling figures (most wickets taken) in a single match?

_____

Query –

select top 1 match_id, bowler, count(*) as wickets
from deliveries
where player_dismissed is not null and
dismissal_kind IN ('caught and bowled', 'bowled',
'stumped', 'caught', 'lbw')
group by match_id, bowler
order by wickets desc

Output –

	match_id	bowler	wickets
1	83	Sohail Tanvir	6

Q. Which are the top 5 bowlers according to economy rate who bowled more than 5 overs?

Query –

with ov as (
select bowler, over_no from deliveries
group by bowler, over_no), overs as (
select bowler, count(*) as total_overs from ov
group by bowler)

select top 5 d.bowler, o.total_overs,
sum(d.total_runs)/o.total_overs as econ_rate from
deliveries d
join overs o on o.bowler = d.bowler
group by d.bowler, o.total_overs
having o.total_overs > 5
order by econ_rate

Output –

	bowler	total_overs	econ_rate
1	R Tewatia	7	7
2	JD Ryder	9	7
3	Mohammad Nabi	8	7
4	KP Pietersen	9	7
5	BAW Mendis	8	7

Q. How many matches were won by each team in each season?

Query –

select season, winner as team, count(*)
as matches_won from matches
where winner is not null
group by season, winner
order by season

Output –

	season	team	matches_won
1	2008	Chennai Super Kings	9
2	2008	Deccan Chargers	2
3	2008	Delhi Daredevils	7
4	2008	Kings XI Punjab	10
5	2008	Kolkata Knight Riders	6
6	2008	Mumbai Indians	7
7	2008	Rajasthan Royals	13
8	2008	Royal Challengers Bangalore	4
9	2009	Chennai Super Kings	1
10	2009	Deccan Chargers	3
11	2009	Delhi Daredevils	2
12	2009	Kings XI Punjab	1
13	2009	Kolkata Knight Riders	1
14	2009	Mumbai Indians	1
15	2009	Royal Challengers Bangalore	1
16	2017	Delhi Daredevils	6
17	2017	Gujarat Lions	4
18	2017	Kings XI Punjab	7
19	2017	Kolkata Knight Riders	9
20	2017	Mumbai Indians	11
21	2017	Rising Pune Supergiant	10
22	2017	Royal Challengers Bangalore	3
23	2017	Sunrisers Hyderabad	8

Q. How many times did each team win the toss?

_____

Query –

select toss_winner, count(*) as win_tosses from
matches
group by toss_winner
order by win_tosses desc

Output –

	toss_winner	win_tosses
1	Mumbai Indians	19
2	Kolkata Knight Riders	18
3	Delhi Daredevils	16
4	Royal Challengers Bangalore	16
5	Kings XI Punjab	12
6	Rajasthan Royals	11
7	Deccan Chargers	11
8	Chennai Super Kings	7
9	Gujarat Lions	7
10	Rising Pune Supergiant	6
11	Sunrisers Hyderabad	5

Q. What is the highest partnership (runs) for each team in each season?

Query –

with part1 as (
select season, batting_team, batsman, non_striker, sum(total_runs)
as runs_bat from deliveries, matches
where deliveries.match_id = matches.id
group by season, batting_team, batsman, non_striker),
part as (select p1.season as season, p1.batting_team as team,
p1.batsman as batsman1, p1.non_striker as batsman2, p1.runs_bat
as bat1_run, p2.runs_bat as bat2_run from part1 p1
join part1 p2 on (p2.season = p1.season and p2.batting_team =
p1.batting_team and p2.batsman = p1.non_striker and
p2.non_striker = p1.batsman)),
high_part as (select season, team, batsman1, batsman2,
(bat1_run+bat2_run) as partnership, ROW_NUMBER() over(partition
by season, team order by (bat1_run+bat2_run) desc) as ro from part)

select season, team, batsman1, batsman2, partnership from
high_part
where ro=1order by season, team

Output –

	season	team	batsman1	batsman2	partnership
1	2008	Chennai Super Kings	SK Raina	PA Patel	231
2	2008	Deccan Chargers	HH Gibbs	AC Gilchrist	288
3	2008	Delhi Daredevils	V Sehwag	G Gambhir	529
4	2008	Kings XI Punjab	SE Marsh	JR Hopes	266
5	2008	Kolkata Knight Riders	SC Ganguly	DJ Hussey	299
6	2008	Mumbai Indians	ST Jayasuriya	SR Tendulkar	400
7	2008	Rajasthan Royals	SA Asnodkar	GC Smith	418
8	2008	Royal Challengers Bangalore	LRPL Taylor	W Jaffer	129
9	2009	Chennai Super Kings	PA Patel	ML Hayden	163
10	2009	Deccan Chargers	HH Gibbs	AC Gilchrist	133
11	2009	Delhi Daredevils	MK Tiwary	AB de Villiers	74
12	2009	Kings XI Punjab	RS Bopara	K Goel	128
13	2009	Kolkata Knight Riders	CH Gayle	BB McCullum	98
14	2009	Mumbai Indians	SR Tendulkar	JP Duminy	95
15	2009	Rajasthan Royals	GC Smith	YK Pathan	56
16	2009	Royal Challengers Bangalore	V Kohli	R Dravid	80
17	2017	Delhi Daredevils	SV Samson	RR Pant	226
18	2017	Gujarat Lions	KD Karthik	SK Raina	318
19	2017	Kings XI Punjab	HM Amla	GJ Maxwell	293
20	2017	Kolkata Knight Riders	RV Uthappa	G Gambhir	428
21	2017	Mumbai Indians	PA Patel	JC Buttler	341
22	2017	Rising Pune Supergiant	RA Tripathi	AM Rahane	322
23	2017	Royal Challengers Bangalore	V Kohli	CH Gayle	275
24	2017	Sunrisers Hyderabad	DA Warner	S Dhawan	646

Q. What are average scores in each city?

Query –

with cte as (select m.city as city, m.id, d.inning,
sum(d.total_runs) as score from deliveries d
join matches m on m.id = d.match_id
group by m.city, m.id, d.inning)

select city, avg(score) as avg_score from cte
group by city
order by avg_score desc

Output –

	city	avg_score
1	Kanpur	176
2	Indore	170
3	Chandigarh	167
4	Chennai	166
5	Delhi	166
6	Hyderabad	163
7	Durban	158
8	Pune	152
9	Mumbai	152
10	Jaipur	151
11	Rajkot	149
12	Kolkata	144
13	Bangalore	138
14	Port Elizabeth	133
15	Cape Town	110

Q. What is the average number of boundaries hit by each team
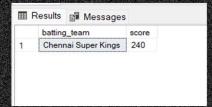in each season?

_____

Query –

select season, batting_team, avg(fours+sixes) as
average_boundaries from (
select season, match_id, batting_team, sum(case
when batsman_runs=4 then 1 else 0 end) as fours,
sum(case when batsman_runs=6 then 1 else 0 end)
as sixes from deliveries, matches
where deliveries.match_id = matches.id
group by season, match_id, batting_team) as
team_bounsaries
group by season, batting_team
order by season

Output –

	season	batting_team	average_boundaries
1	2008	Chennai Super Kings	20
2	2008	Deccan Chargers	21
3	2008	Delhi Daredevils	19
4	2008	Kings XI Punjab	21
5	2008	Kolkata Knight Riders	18
6	2008	Mumbai Indians	19
7	2008	Rajasthan Royals	21
8	2008	Royal Challengers Bangalore	17
9	2009	Chennai Super Kings	18
10	2009	Deccan Chargers	18
11	2009	Delhi Daredevils	16
12	2009	Kings XI Punjab	16
13	2009	Kolkata Knight Riders	13
14	2009	Mumbai Indians	19
15	2009	Rajasthan Royals	11
16	2009	Royal Challengers Bangalore	17
17	2017	Delhi Daredevils	20
18	2017	Gujarat Lions	22
19	2017	Kings XI Punjab	19
20	2017	Kolkata Knight Riders	19
21	2017	Mumbai Indians	20
22	2017	Rising Pune Supergiant	17
23	2017	Royal Challengers Bangalore	16
24	2017	Sunrisers Hyderabad	20

Q. Which team has the highest total score in a single match?

_____

Query –

select top 1 batting_team, sum(total_runs) as
score from deliveries
group by batting_team, match_id
order by score desc

Output –

Q. What is the number of matches won by each team batting first
versus batting second?

_____

Query –

```
with bat as (
select winner, count(*) as bat_first from matches
where (winner = toss_winner and toss_decision =
'bat') or (winner != toss_winner and toss_decision =
'field')
group by winner
), bowl as (
select winner, count(*) as bat_second from matches
where (winner = toss_winner and toss_decision =
'field') or (winner != toss_winner and toss_decision =
'bat')
group by winner)

select b.winner as team, b.bat_first, w.bat_second
from bat b
join bowl w on w.winner = b.winner
```

Output –

	team	bat_first	bat_second
1	Chennai Super Kings	5	5
2	Deccan Chargers	2	3
3	Delhi Daredevils	6	9
4	Kings XI Punjab	8	10
5	Kolkata Knight Riders	6	10
6	Mumbai Indians	7	12
7	Rajasthan Royals	4	9
8	Rising Pune Supergiant	5	5
9	Royal Challengers Bangalore	6	2
10	Sunrisers Hyderabad	5	3

Q. What are the number of wins by city while batting first and

batting second?

_____

Query –

with bat as(select city, count(*) as bat_first from
matches
where (winner = toss_winner and toss_decision =
'bat') or (winner != toss_winner and toss_decision =
'field')
group by city), bowl as (
select city, count(*) as bat_second from matches
where (winner = toss_winner and toss_decision =
'field') or (winner != toss_winner and toss_decision =
'bat')
group by city)

select b.city, b.bat_first, w.bat_second from bat b
join bowl w on w.city = b.city

Output –

	city	bat_first	bat_second
1	Bangalore	6	9
2	Cape Town	3	2
3	Chandigarh	6	5
4	Chennai	4	3
5	Delhi	6	7
6	Durban	2	2
7	Hyderabad	8	7
8	Jaipur	2	5
9	Kolkata	5	9
10	Mumbai	7	11
11	Pune	2	5
12	Rajkot	2	2

Q. How many extras (wide's & no-balls) were bowled by each
team in each match?

_____

Query –

select bowling_team, sum(case when wide_runs
!= 0 or noball_runs !=0 then 1 else 0
end)/count(distinct match_id) as avg_extra from
deliveries
group by bowling_team

Output –

	bowling_team	avg_extra
1	Rising Pune Supergiant	3
2	Kolkata Knight Riders	4
3	Delhi Daredevils	4
4	Rajasthan Royals	4
5	Chennai Super Kings	4
6	Royal Challengers Bangalore	4
7	Sunrisers Hyderabad	2
8	Mumbai Indians	4
9	Deccan Chargers	4
10	Gujarat Lions	3
11	Kings XI Punjab	4

THANKS

Linkedin - www.linkedin.com/in/vaibhav-mehendale-725371194

Github Link - https://github.com/Vaibs24?tab=repositories