Analysis of ball by ball IPL data from the year 2008-2024

Github : https://github.com/jKunal24/IPL-Data-Analysis-With-S3-and-Spark

Storage : AWS S3

Environment : Databricks to use Spark for data processing and handle infrastructure side of things

Language : Python (as interface to use Apache Spark)

Visualization Libraries : Plotly, Matplotlib & Seaborn (Player-wise,Venue-wise and team-wise stats and insights)

Data Cleaning:

- Standardizing/Merging team names of the same Franchise with different strings in deliveries file.
  Eg : Royal Challengers Bangalore & Royal Challangers Bengaluru are the same franchise but both strings are used for different matches in the file.
- Standardizing/Merging venue/stadium names of the same stadiums with different strings in Matches file.
- Eg : Feroz Shah Kotla (old name)is the same stadium as Arun Jaitley Stadium, Delhi (new name)
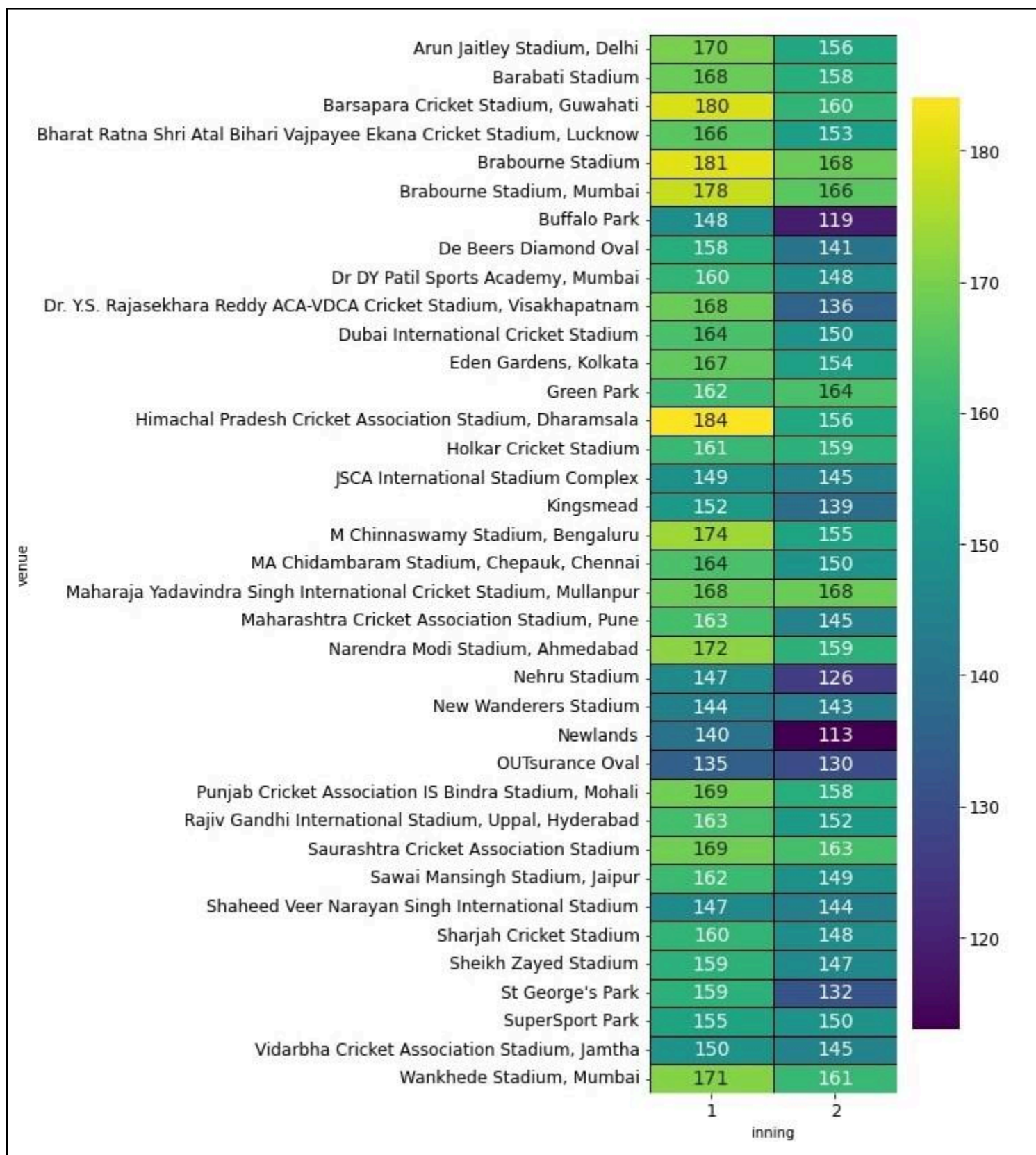 Zayed Cricket Stadium, Abu Dhabi is same as Sheikh Zayed Stadium

Schema :

Used custom schemas with **StuctType** & **Structfield** to avoid datatype errors in while performing transformations and actions.

```python
ball_by_ball_Schema = StructType([
    StructField("match_id",IntegerType(),True),
    StructField("inning",IntegerType(),True),
    StructField("batting_team",StringType(),True),
    StructField("bowling_team",StringType(),True),
    StructField("over",IntegerType(),True),
    StructField("ball",IntegerType(),True),
    StructField("batter",StringType(),True),
    StructField("bowler",StringType(),True),
    StructField("non_striker",StringType(),True),
    StructField("batsman_runs",IntegerType(),True),
    StructField("extra_runs",IntegerType(),True),
    StructField("total_runs",IntegerType(),True),
    StructField("extras_type",StringType(),True),
    StructField("is_wicket",IntegerType(),True),
    StructField("player_dismissed",StringType(),True),
    StructField("dismissal_kind",StringType(),True),
    StructField("fielder",StringType(),True)
])
```
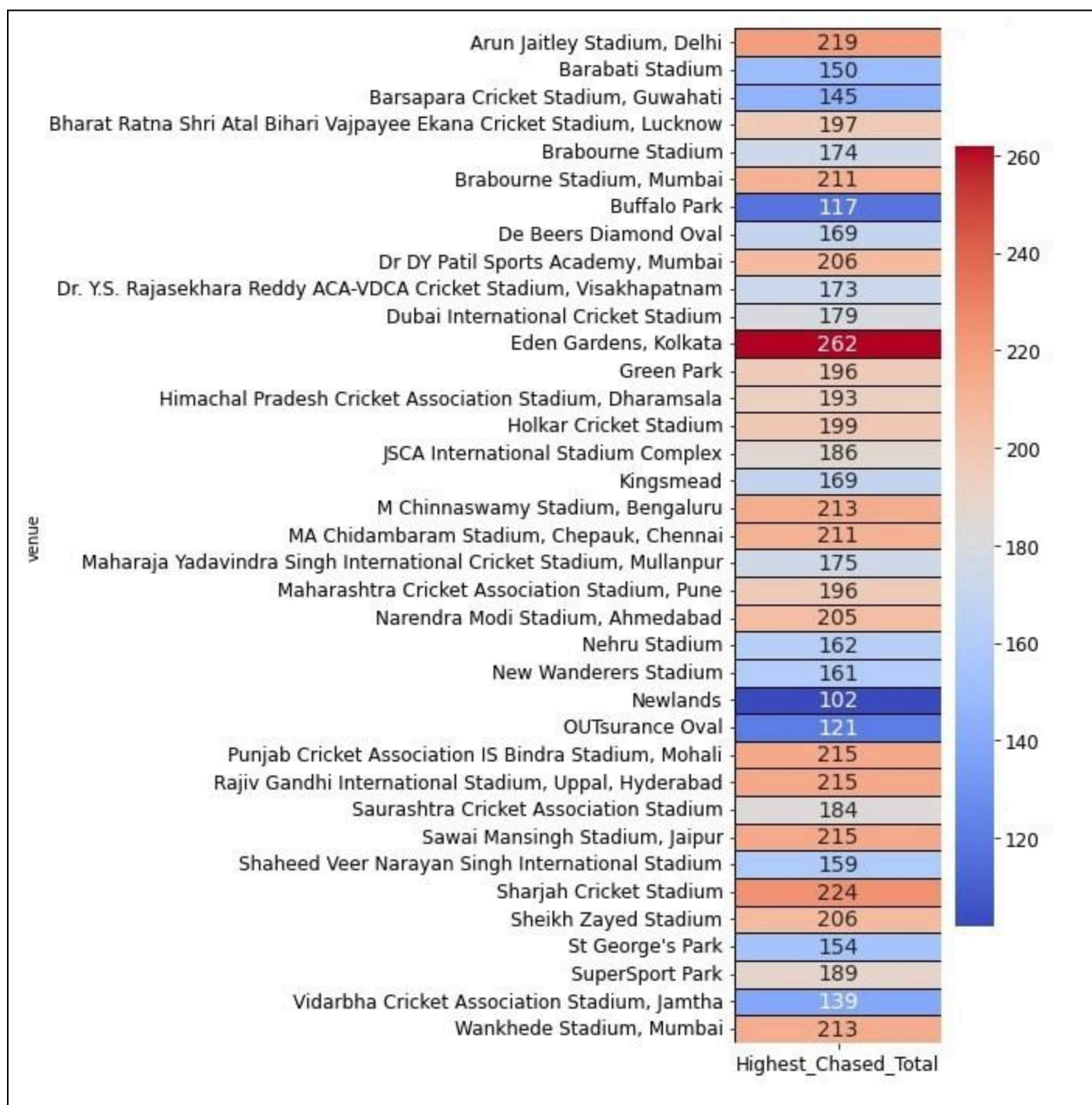
```python
matches_df_Schema = StructType([
    StructField("id",IntegerType(),True),
    StructField("season",StringType(),True),
    StructField("city",StringType(),True),
    StructField("date",DateType(),True),
    StructField("match_type",StringType(),True),
    StructField("player_of_match",StringType(),True),
    StructField("venue",StringType(),True),
    StructField("team1",StringType(),True),
    StructField("team2",StringType(),True),
    StructField("toss_winner",StringType(),True),
    StructField("toss_decision",StringType(),True),
    StructField("winner",StringType(),True),
    StructField("result",StringType(),True),
    StructField("result_margin",IntegerType(),True),
    StructField("target_runs",IntegerType(),True),
    StructField("target_overs",IntegerType(),True),
    StructField("super_over",StringType(),True),
    StructField("method",StringType(),True),
    StructField("umpire1",StringType(),True),
    StructField("umpire2",StringType(),True)
])
```

Average inning wise score for all venues/stadiums

venue	inning 1	inning 2
Arun Jaitley Stadium, Delhi	170	156
Barabati Stadium	168	158
Barsapara Cricket Stadium, Guwahati	180	160
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cricket Stadium, Lucknow	166	153
Brabourne Stadium	181	168
Brabourne Stadium, Mumbai	178	166
Buffalo Park	148	119
De Beers Diamond Oval	158	141
Dr DY Patil Sports Academy, Mumbai	160	148
Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Stadium, Visakhapatnam	168	136
Dubai International Cricket Stadium	164	150
Eden Gardens, Kolkata	167	154
Green Park	162	164
Himachal Pradesh Cricket Association Stadium, Dharamsala	184	156
Holkar Cricket Stadium	161	159
JSCA International Stadium Complex	149	145
Kingsmead	152	139
M Chinnaswamy Stadium, Bengaluru	174	155
MA Chidambaram Stadium, Chepauk, Chennai	164	150
Maharaja Yadavindra Singh International Cricket Stadium, Mullanpur	168	168
Maharashtra Cricket Association Stadium, Pune	163	145
Narendra Modi Stadium, Ahmedabad	172	159
Nehru Stadium	147	126
New Wanderers Stadium	144	143
Newlands	140	113
OUTsurance Oval	135	130
Punjab Cricket Association IS Bindra Stadium, Mohali	169	158
Rajiv Gandhi International Stadium, Uppal, Hyderabad	163	152
Saurashtra Cricket Association Stadium	169	163
Sawai Mansingh Stadium, Jaipur	162	149
Shaheed Veer Narayan Singh International Stadium	147	144
Sharjah Cricket Stadium	160	148
Sheikh Zayed Stadium	159	147
St George's Park	159	132
SuperSport Park	155	150
Vidarbha Cricket Association Stadium, Jamtha	150	145
Wankhede Stadium, Mumbai	171	161

Highest chased target at all venues/stadiums

Venue	Highest_Chased_Total
Arun Jaitley Stadium, Delhi	219
Barabati Stadium	150
Barsapara Cricket Stadium, Guwahati	145
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cricket Stadium, Lucknow	197
Brabourne Stadium	174
Brabourne Stadium, Mumbai	211
Buffalo Park	117
De Beers Diamond Oval	169
Dr DY Patil Sports Academy, Mumbai	206
Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Stadium, Visakhapatnam	173
Dubai International Cricket Stadium	179
Eden Gardens, Kolkata	262
Green Park	196
Himachal Pradesh Cricket Association Stadium, Dharamsala	193
Holkar Cricket Stadium	199
JSCA International Stadium Complex	186
Kingsmead	169
M Chinnaswamy Stadium, Bengaluru	213
MA Chidambaram Stadium, Chepauk, Chennai	211
Maharaja Yadavindra Singh International Cricket Stadium, Mullanpur	175
Maharashtra Cricket Association Stadium, Pune	196
Narendra Modi Stadium, Ahmedabad	205
Nehru Stadium	162
New Wanderers Stadium	161
Newlands	102
OUTsurance Oval	121
Punjab Cricket Association IS Bindra Stadium, Mohali	215
Rajiv Gandhi International Stadium, Uppal, Hyderabad	215
Saurashtra Cricket Association Stadium	184
Sawai Mansingh Stadium, Jaipur	215
Shaheed Veer Narayan Singh International Stadium	159
Sharjah Cricket Stadium	224
Sheikh Zayed Stadium	206
St George's Park	154
SuperSport Park	189
Vidarbha Cricket Association Stadium, Jamtha	139
Wankhede Stadium, Mumbai	213

Distribution of per Innings Score across all IPL Season

- From the past 4 years the median score is constantly going up which shows the trend of batsmen dominated high scoring games with more than half of the games having per inning score in the 3rd & the 4th quartile in the year 2024.

Batsman dominated Face-offs

| batter	bowler	Matches	Balls	Runs	Wickets	Batter_SR	Bowler_Avg
DA Warner	SP Narine	16	123	195	2	158.54	97.5
V Kohli	R Ashwin	22	145	179	1	123.45	179
SK Raina	PP Chawla	16	101	175	4	173.27	43.75
RG Sharma	UT Yadav	17	118	170	5	144.07	34
V Kohli	UT Yadav	14	94	166	3	176.6	55.33
V Kohli	A Mishra	15	104	162	3	155.77	54
KL Rahul	DL Chahar	11	99	158	0	159.6	null
V Kohli	DJ Bravo	12	102	157	1	153.92	157
DA Warner	YS Chahal	11	96	156	2	162.5	78
V Kohli	RA Jadeja	19	143	156	3	109.09	52

- Interesting contest / one-to-one face-offs between the bat and the ball where the batters have had an upper edge and great success against a particular bowler.
- KL Rahul has scored 158 runs against Deepak Chahar at a whopping Strike rate of 160 and is not even once dismissed by Chahar.

Bowler dominated Face-offs

batter	bowler	Matches	Balls	Runs	Wickets	Batter_SR	Bowler_Avg
RG Sharma	SP Narine	21	133	142	8	106.77	17.75
RG Sharma	A Mishra	17	91	87	7	95.6	12.43
MS Dhoni	Z Khan	11	42	74	7	176.19	10.57
RR Pant	JJ Bumrah	14	46	55	7	119.57	7.86
AM Rahane	B Kumar	18	117	104	7	88.89	14.86
V Kohli	Sandeep Sharma	16	73	103	7	141.1	14.71
AT Rayudu	MM Sharma	14	49	71	7	144.9	10.14
V Kohli	A Nehra	10	54	60	6	111.11	10
RV Uthappa	R Ashwin	15	94	123	6	130.85	20.5
GJ Maxwell	RA Jadeja	11	51	70	6	137.25	11.67

- Interesting contest / one-to-one face-offs between the bat and the ball where the bowlers have had an upper edge and great success against a particular batsmen.
- Jasprit Bumrah has dismissed Rishab Pant 7 out of 14 times they have been against each other at an impressive average of 7.8.

Over-wise average run-rate for all teams

Over-Wise Average Run-rate across all games
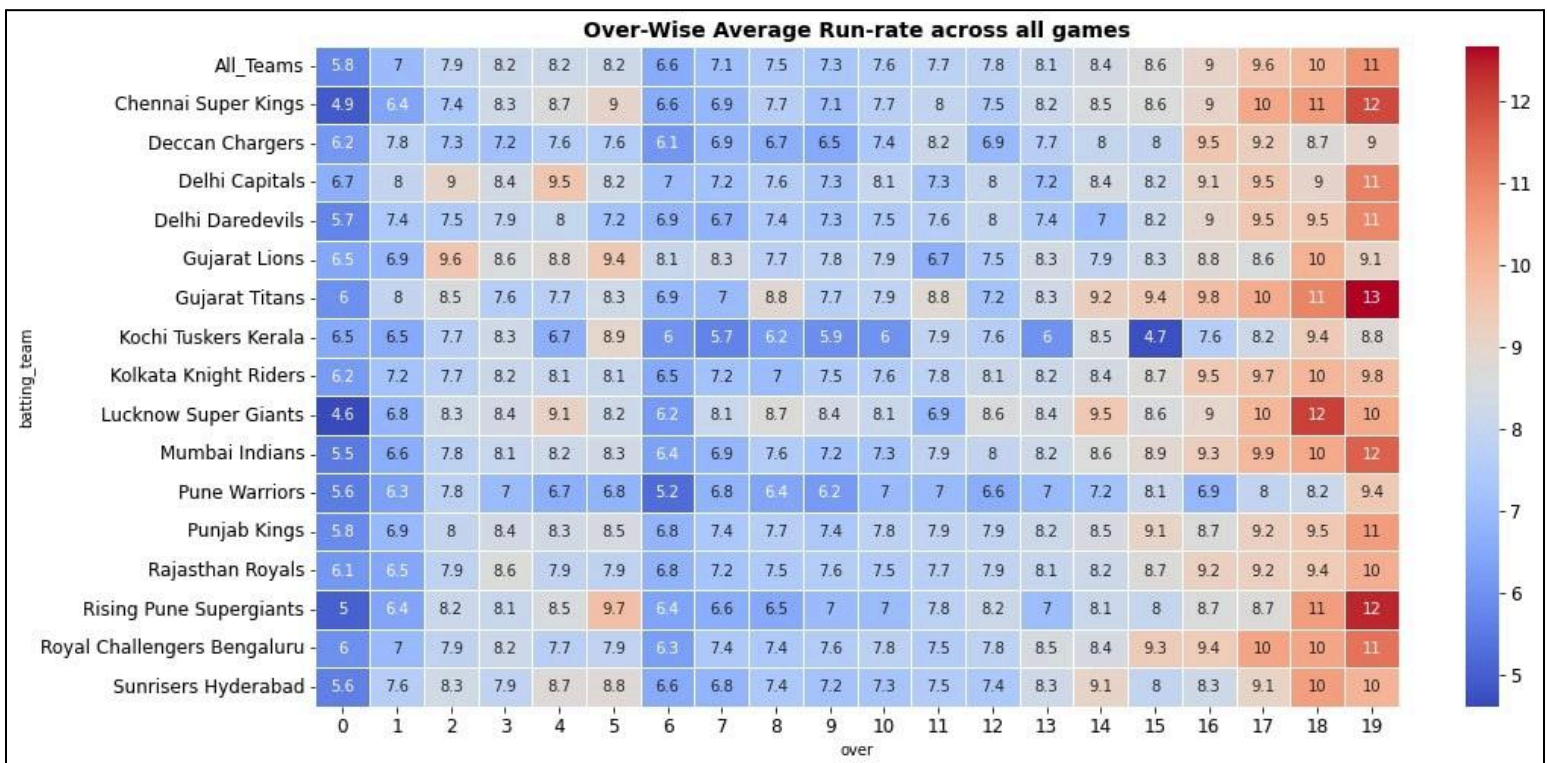
- Gujarat Titans, Royal Challengers Bengaluru and Mumbai Indians are the teams who have constantly kept their run rate above 9 in the death overs which shows that they have picked best of the finishers.
- This visualization also gives insights into how the T20 format has moved more towards a batter dominated format as the difference in run-rate for old ruled out franchises (like Pune, Gujarat and Kochi) is comparatively lower than the ones still contesting in the IPL.

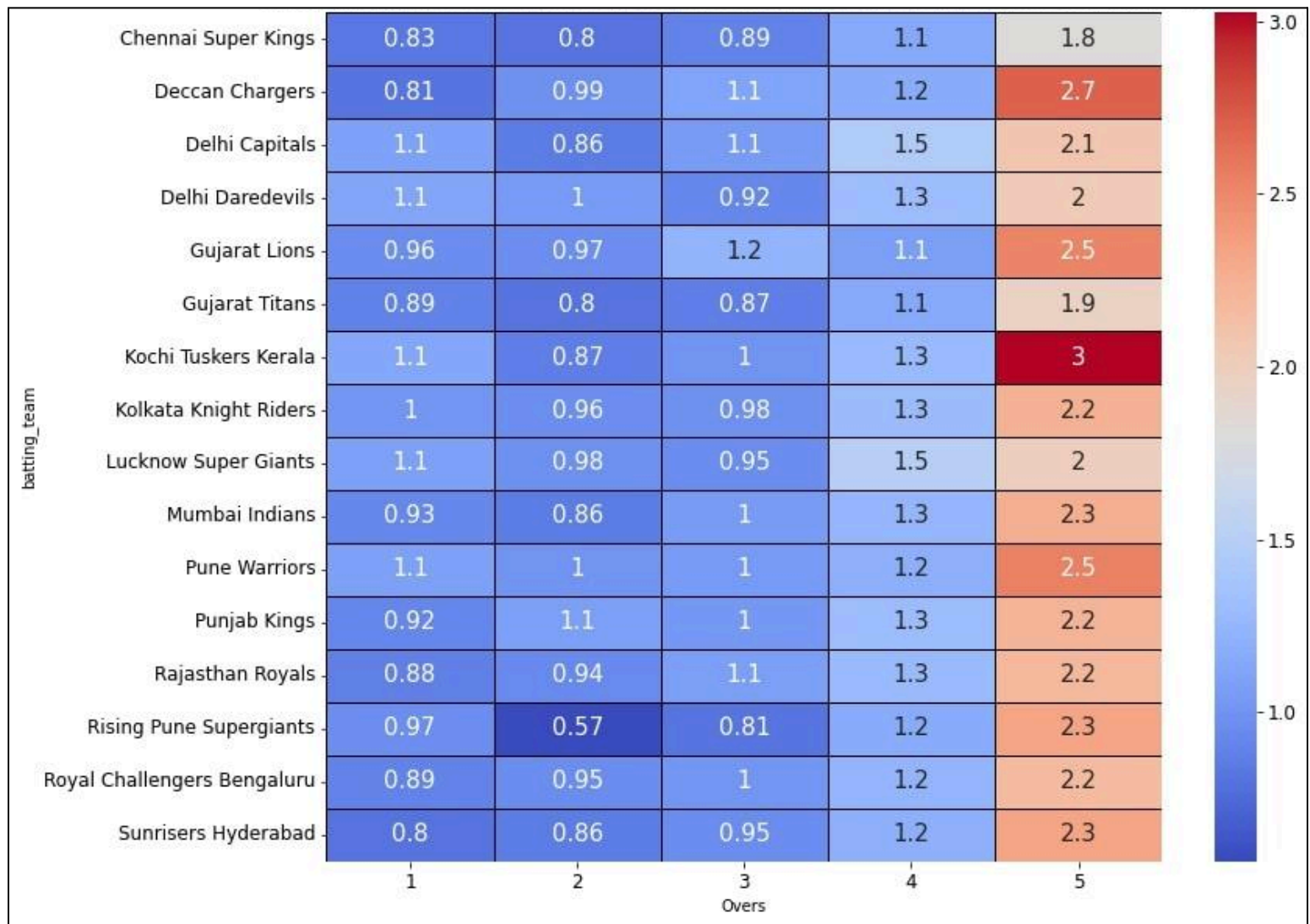- CSK has the sharpest fall in run-rate after powerplay to 6.9.Sunrisers Hyderabad has struggled in the middle overs as their run-rate is below the average run-rate of all teams in between overs 10-15.
- LSG has performed very well in the middle overs with run-rate constantly above 8 and KKR has not been able to aggressively increase their runs-rate in the death overs as compared to other teams.

Average No. of wickets taken in different parts of the game (overs : 1-4, 5-8, 9-12, 13-16, 17-20)

| bowling_team	1	2	3	4	5
Chennai Super Kings	0.93	0.97	1.1	1.3	2.3
Deccan Chargers	1.1	0.85	0.95	1.2	2.1
Delhi Capitals	0.86	1.1	1	1.4	2.2
Delhi Daredevils	0.9	0.92	1	1.1	2.2
Gujarat Lions	1.2	0.6	0.77	0.92	1.9
Gujarat Titans	1	1.2	0.88	1.5	2.4
Kochi Tuskers Kerala	1	0.86	0.79	1.3	1.9
Kolkata Knight Riders	0.81	0.95	1.1	1.3	2.2
Lucknow Super Giants	0.77	0.95	1.2	1.4	2.1
Mumbai Indians	0.99	0.95	0.97	1.3	2.2
Pune Warriors	0.8	0.83	0.86	1.2	1.8
Punjab Kings	0.89	0.89	0.95	1.2	2.3
Rajasthan Royals	1.1	0.87	0.93	1.2	2
Rising Pune Supergiants	0.8	1.1	1.1	1.2	2.5
Royal Challengers Bengaluru	0.93	0.92	1.1	1.2	2.2
Sunrisers Hyderabad	0.92	0.89	1	1.3	2.2

Overs

- CSK,GT and PBKS on average have been more effective as a balling unit in the death overs.
- GT,MI and RR on average have been successful in taking at least 1 wicket in the powerplay in every game which helps to slow down the game especially when there are only 2 players outside the 30-yard circle.

6

Average No. of wickets lost in different parts of the game (overs : 1-4, 5-8, 9-12, 13-16, 17-20)

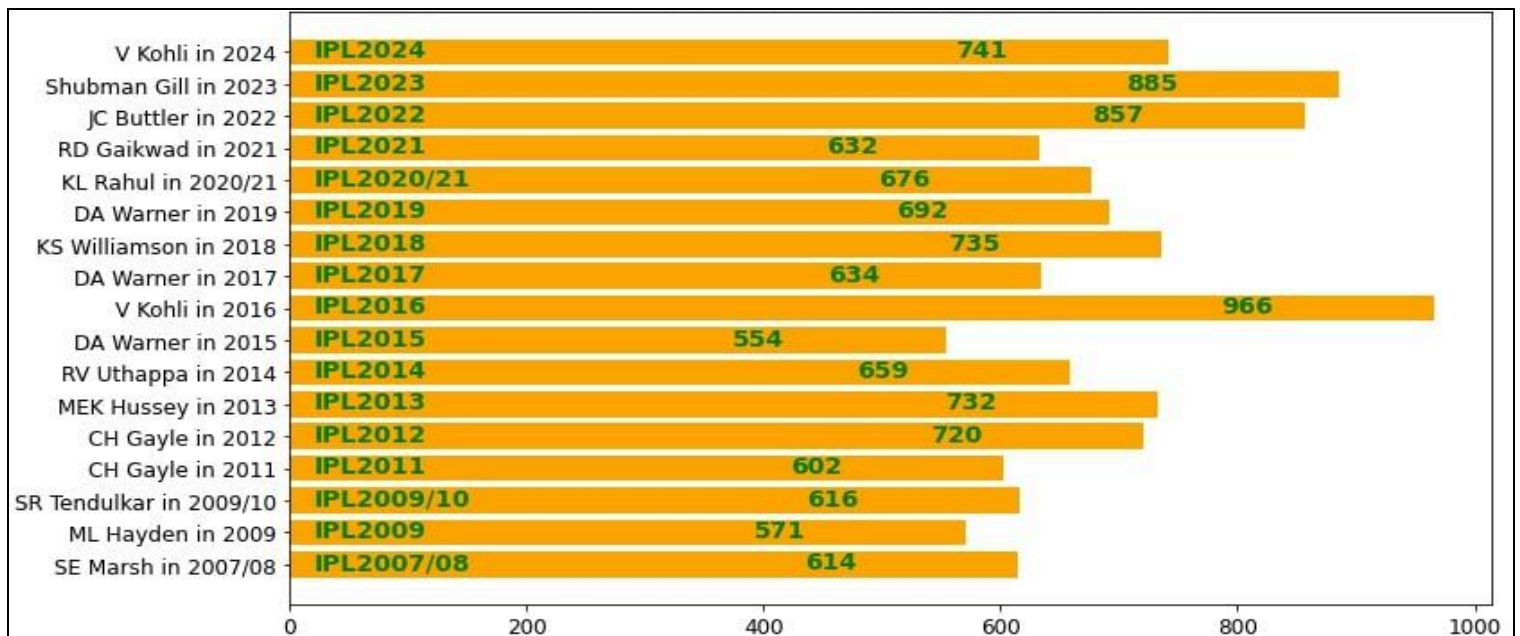| batting_team	1	2	3	4	5
Chennai Super Kings	0.83	0.8	0.89	1.1	1.8
Deccan Chargers	0.81	0.99	1.1	1.2	2.7
Delhi Capitals	1.1	0.86	1.1	1.5	2.1
Delhi Daredevils	1.1	1	0.92	1.3	2
Gujarat Lions	0.96	0.97	1.2	1.1	2.5
Gujarat Titans	0.89	0.8	0.87	1.1	1.9
Kochi Tuskers Kerala	1.1	0.87	1	1.3	3
Kolkata Knight Riders	1	0.96	0.98	1.3	2.2
Lucknow Super Giants	1.1	0.98	0.95	1.5	2
Mumbai Indians	0.93	0.86	1	1.3	2.3
Pune Warriors	1.1	1	1	1.2	2.5
Punjab Kings	0.92	1.1	1	1.3	2.2
Rajasthan Royals	0.88	0.94	1.1	1.3	2.2
Rising Pune Supergiants	0.97	0.57	0.81	1.2	2.3
Royal Challengers Bengaluru	0.89	0.95	1	1.2	2.2
Sunrisers Hyderabad	0.8	0.86	0.95	1.2	2.3

Overs

- CSK and GT has the best finishers as they have lost less than 2 wickets on average in the death overs and have also been able to keep run-rate above 10 runs and over.
- KKR ,DC and LSG on average has lost at least 1 wicket in the first 4 overs in all of their games .

7

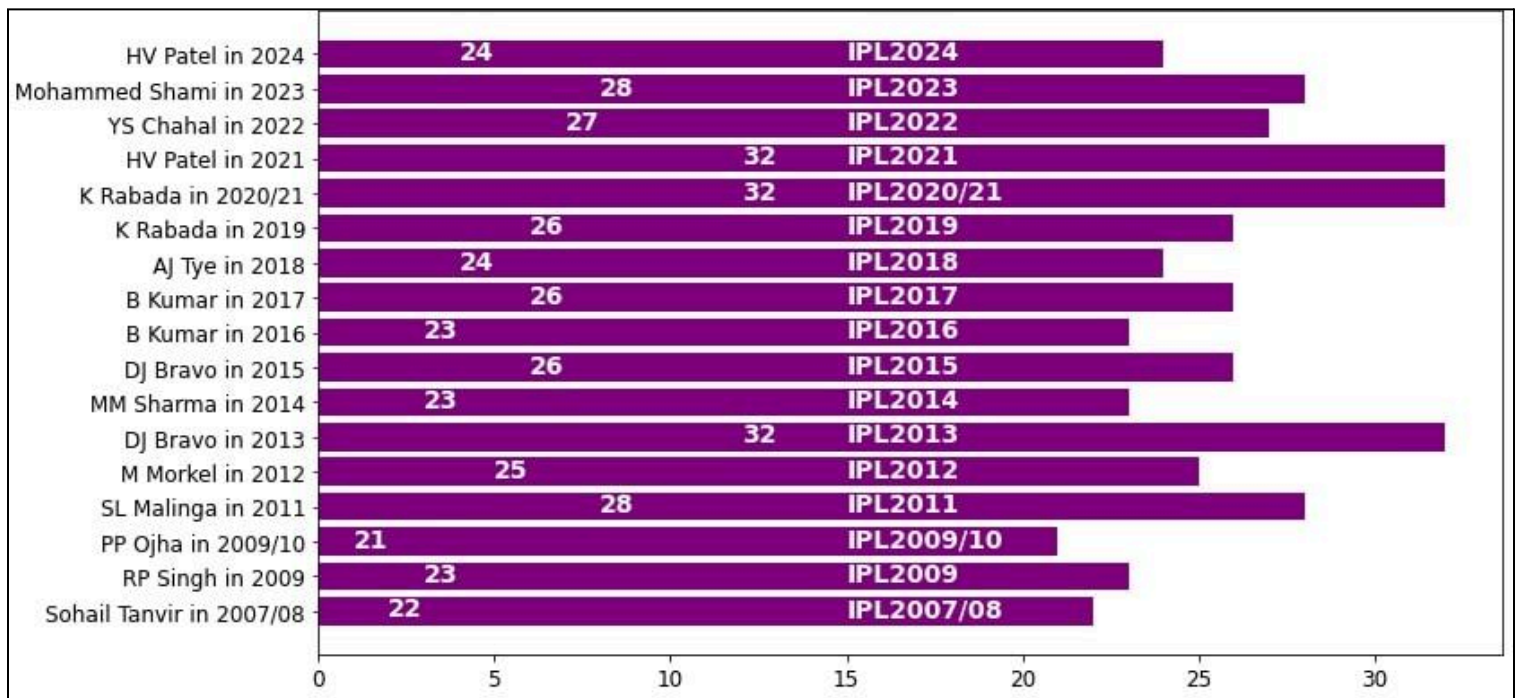## Over(1 to 20) v/s average run-rate for all teams across all games
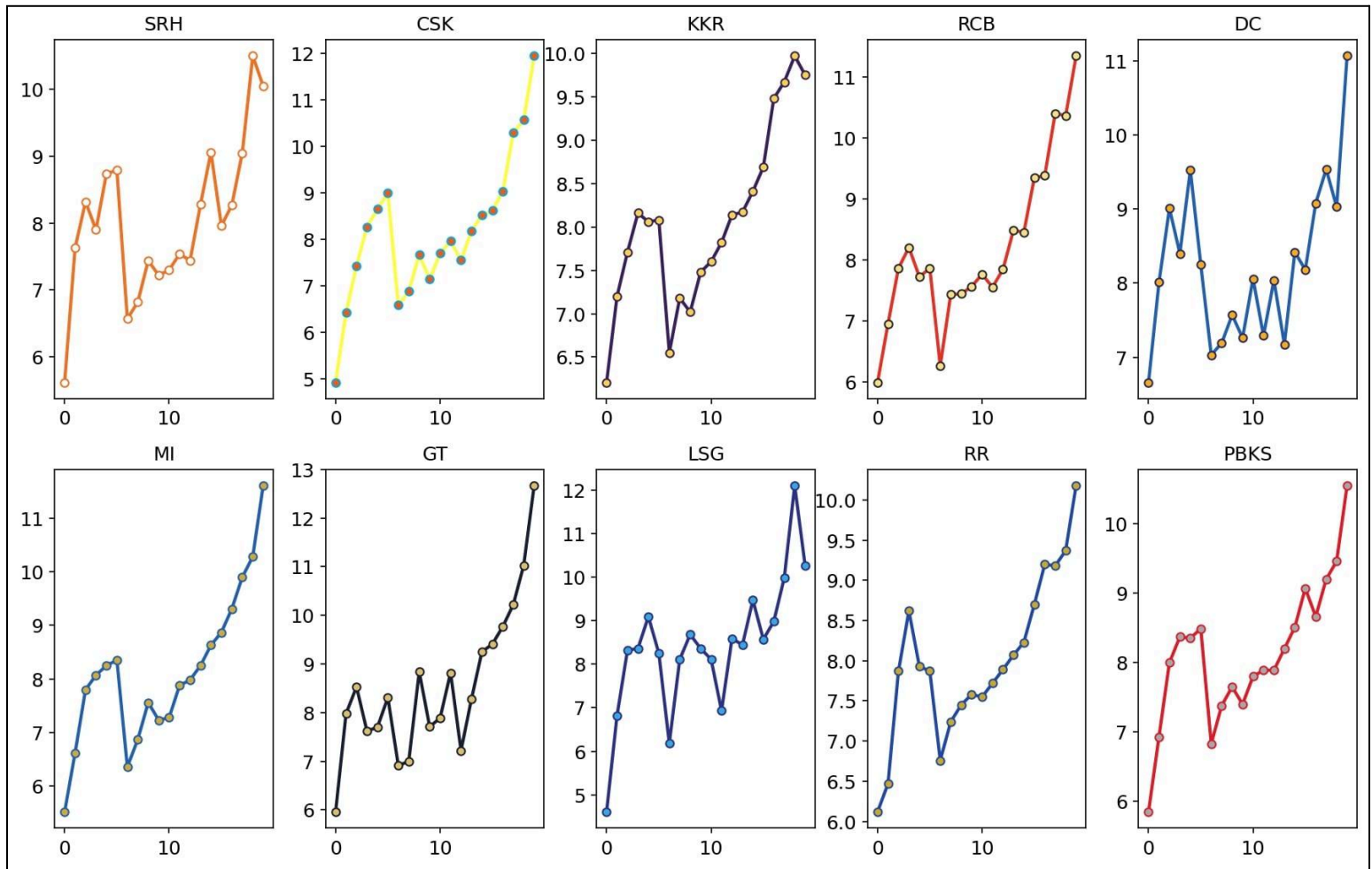
- Sharp fall in run-rate after 5th/6th over indicate an event i.e. the powerplay restrictions being lifted.
- MI and RR have shown continuous and steady rise in run-rate as the game progresses which sets up the platform for the finishers.

Orange Cap Holders in all IPL Seasons (Highest Run Scorers)

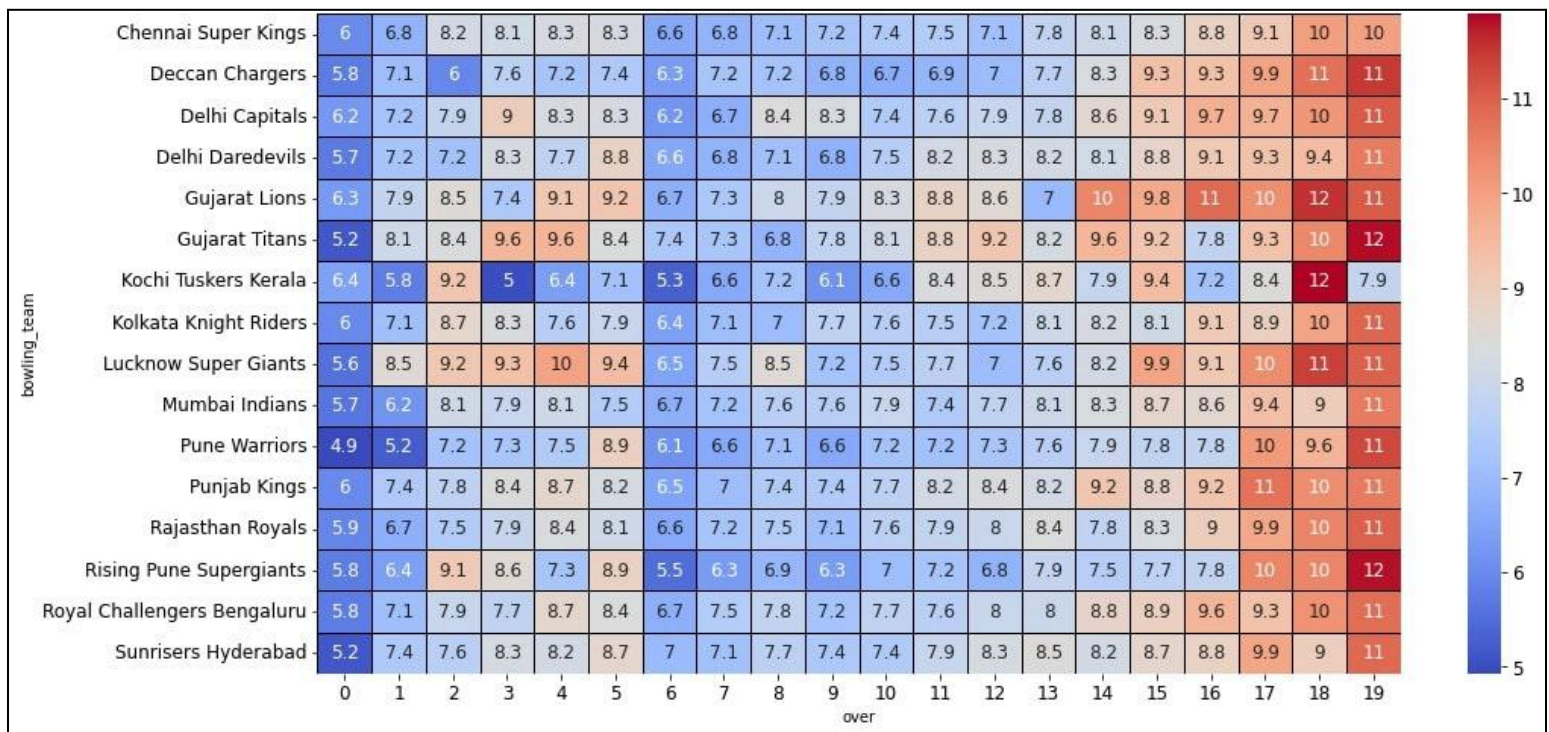| Player	Season	Runs
V Kohli in 2024	IPL2024	741
Shubman Gill in 2023	IPL2023	885
JC Buttler in 2022	IPL2022	857
RD Gaikwad in 2021	IPL2021	632
KL Rahul in 2020/21	IPL2020/21	676
DA Warner in 2019	IPL2019	692
KS Williamson in 2018	IPL2018	735
DA Warner in 2017	IPL2017	634
V Kohli in 2016	IPL2016	966
DA Warner in 2015	IPL2015	554
RV Uthappa in 2014	IPL2014	659
MEK Hussey in 2013	IPL2013	732
CH Gayle in 2012	IPL2012	720
CH Gayle in 2011	IPL2011	602
SR Tendulkar in 2009/10	IPL2009/10	616
ML Hayden in 2009	IPL2009	571
SE Marsh in 2007/08	IPL2007/08	614

- Virat Kohli and Chris Gayle have won the Orange cap twice and David Warner has won the orange cap thrice.

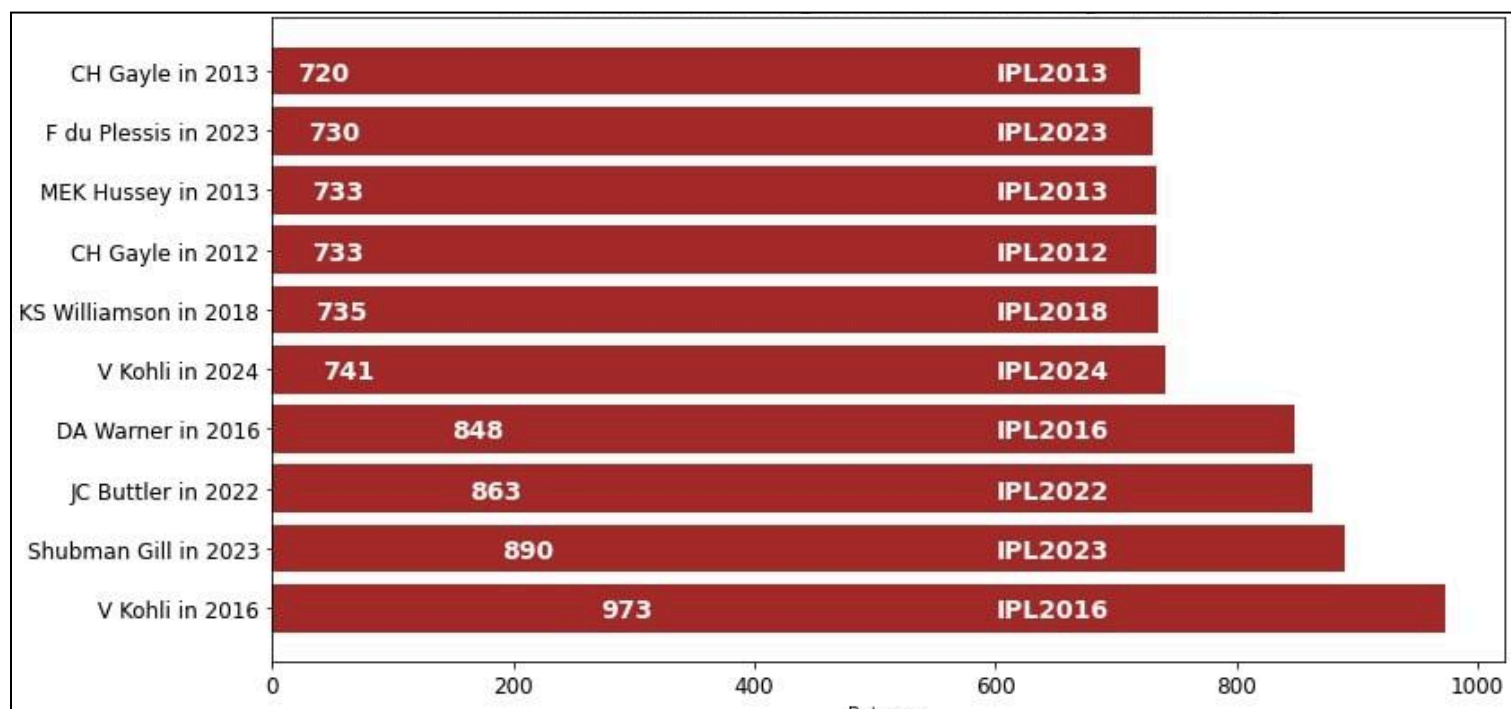## Purple Cap Holders in all IPL Seasons(Highest Wicket takers)

Player	Season	Wickets
HV Patel in 2024	IPL2024	24
Mohammed Shami in 2023	IPL2023	28
YS Chahal in 2022	IPL2022	27
HV Patel in 2021	IPL2021	32
K Rabada in 2020/21	IPL2020/21	32
K Rabada in 2019	IPL2019	26
AJ Tye in 2018	IPL2018	24
B Kumar in 2017	IPL2017	26
B Kumar in 2016	IPL2016	23
DJ Bravo in 2015	IPL2015	26
MM Sharma in 2014	IPL2014	23
DJ Bravo in 2013	IPL2013	32
M Morkel in 2012	IPL2012	25
SL Malinga in 2011	IPL2011	28
PP Ojha in 2009/10	IPL2009/10	21
RP Singh in 2009	IPL2009	23
Sohail Tanvir in 2007/08	IPL2007/08	22

- Harshal Patel, Dwayne Bravo, Kagiso Rabada and Bhuvneshwar Kumar have won the purple cap twice.

Over-wise (from 1-20) avg runs conceded

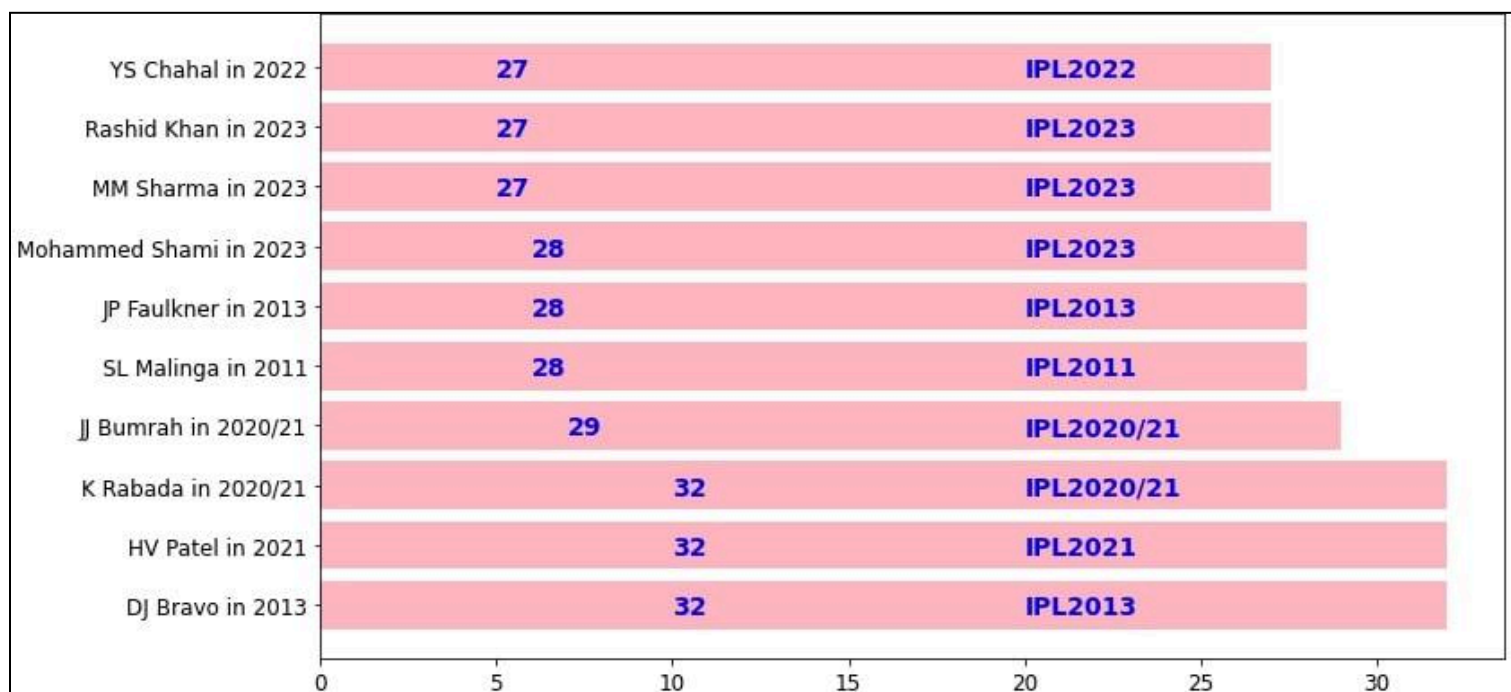| bowling_team \ over	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Chennai Super Kings	6	6.8	8.2	8.1	8.3	8.3	6.6	6.8	7.1	7.2	7.4	7.5	7.1	7.8	8.1	8.3	8.8	9.1	10	10
Deccan Chargers	5.8	7.1	6	7.6	7.2	7.4	6.3	7.2	7.2	6.8	6.7	6.9	7	7.7	8.3	9.3	9.3	9.9	11	11
Delhi Capitals	6.2	7.2	7.9	9	8.3	8.3	6.2	6.7	8.4	8.3	7.4	7.6	7.9	7.8	8.6	9.1	9.7	9.7	10	11
Delhi Daredevils	5.7	7.2	7.2	8.3	7.7	8.8	6.6	6.8	7.1	6.8	7.5	8.2	8.3	8.2	8.1	8.8	9.1	9.3	9.4	11
Gujarat Lions	6.3	7.9	8.5	7.4	9.1	9.2	6.7	7.3	8	7.9	8.3	8.8	8.6	7	10	9.8	11	10	12	11
Gujarat Titans	5.2	8.1	8.4	9.6	9.6	8.4	7.4	7.3	6.8	7.8	8.1	8.8	9.2	8.2	9.6	9.2	7.8	9.3	10	12
Kochi Tuskers Kerala	6.4	5.8	9.2	5	6.4	7.1	5.3	6.6	7.2	6.1	6.6	8.4	8.5	8.7	7.9	9.4	7.2	8.4	12	7.9
Kolkata Knight Riders	6	7.1	8.7	8.3	7.6	7.9	6.4	7.1	7	7.7	7.6	7.5	7.2	8.1	8.2	8.1	9.1	8.9	10	11
Lucknow Super Giants	5.6	8.5	9.2	9.3	10	9.4	6.5	7.5	8.5	7.2	7.5	7.7	7	7.6	8.2	9.9	9.1	10	11	11
Mumbai Indians	5.7	6.2	8.1	7.9	8.1	7.5	6.7	7.2	7.6	7.6	7.9	7.4	7.7	8.1	8.3	8.7	8.6	9.4	9	11
Pune Warriors	4.9	5.2	7.2	7.3	7.5	8.9	6.1	6.6	7.1	6.6	7.2	7.2	7.3	7.6	7.9	7.8	7.8	10	9.6	11
Punjab Kings	6	7.4	7.8	8.4	8.7	8.2	6.5	7	7.4	7.4	7.7	8.2	8.4	8.2	9.2	8.8	9.2	11	10	11
Rajasthan Royals	5.9	6.7	7.5	7.9	8.4	8.1	6.6	7.2	7.5	7.1	7.6	7.9	8	8.4	7.8	8.3	9	9.9	10	11
Rising Pune Supergiants	5.8	6.4	9.1	8.6	7.3	8.9	5.5	6.3	6.9	6.3	7	7.2	6.8	7.9	7.5	7.7	7.8	10	10	12
Royal Challengers Bengaluru	5.8	7.1	7.9	7.7	8.7	8.4	6.7	7.5	7.8	7.2	7.7	7.6	8	8	8.8	8.9	9.6	9.3	10	11
Sunrisers Hyderabad	5.2	7.4	7.6	8.3	8.2	8.7	7	7.1	7.7	7.4	7.4	7.9	8.3	8.5	8.2	8.7	8.8	9.9	9	11

- LSG and GT can have better strategies for the powerplay overs as they have conceded lots of runs due to which the batting team has an upper hand and get the much needed momentum.
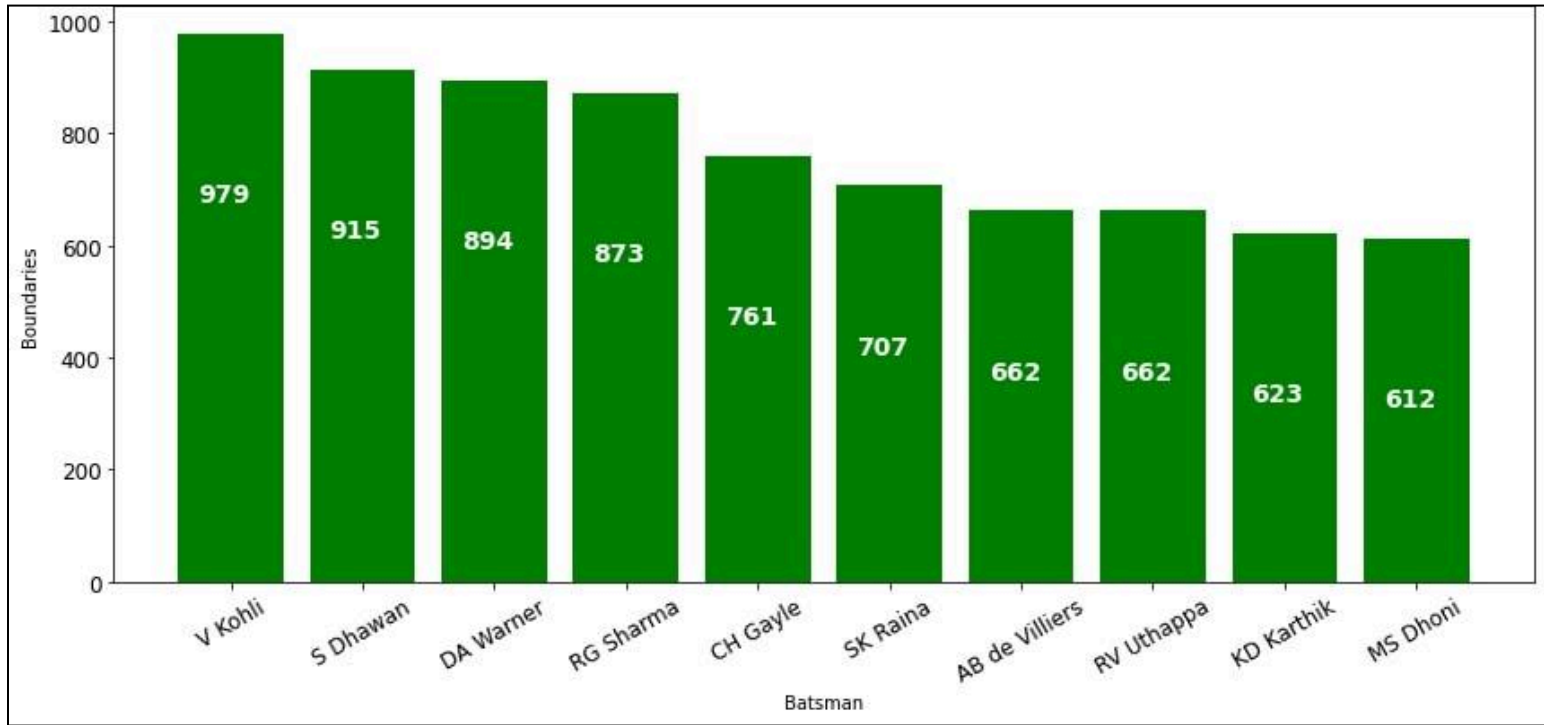
Most Runs in a Single edition of IPL (Top 10)

Player	Runs	Edition
CH Gayle in 2013	720	IPL2013
F du Plessis in 2023	730	IPL2023
MEK Hussey in 2013	733	IPL2013
CH Gayle in 2012	733	IPL2012
KS Williamson in 2018	735	IPL2018
V Kohli in 2024	741	IPL2024
DA Warner in 2016	848	IPL2016
JC Buttler in 2022	863	IPL2022
Shubman Gill in 2023	890	IPL2023
V Kohli in 2016	973	IPL2016

Most Wickets in a Single edition of IPL (Top 10)

| Player	Wickets	Edition
YS Chahal in 2022	27	IPL2022
Rashid Khan in 2023	27	IPL2023
MM Sharma in 2023	27	IPL2023
Mohammed Shami in 2023	28	IPL2023
JP Faulkner in 2013	28	IPL2013
SL Malinga in 2011	28	IPL2011
JJ Bumrah in 2020/21	29	IPL2020/21
K Rabada in 2020/21	32	IPL2020/21
HV Patel in 2021	32	IPL2021
DJ Bravo in 2013	32	IPL2013

Most no. of boundaries in IPL (Top 10)

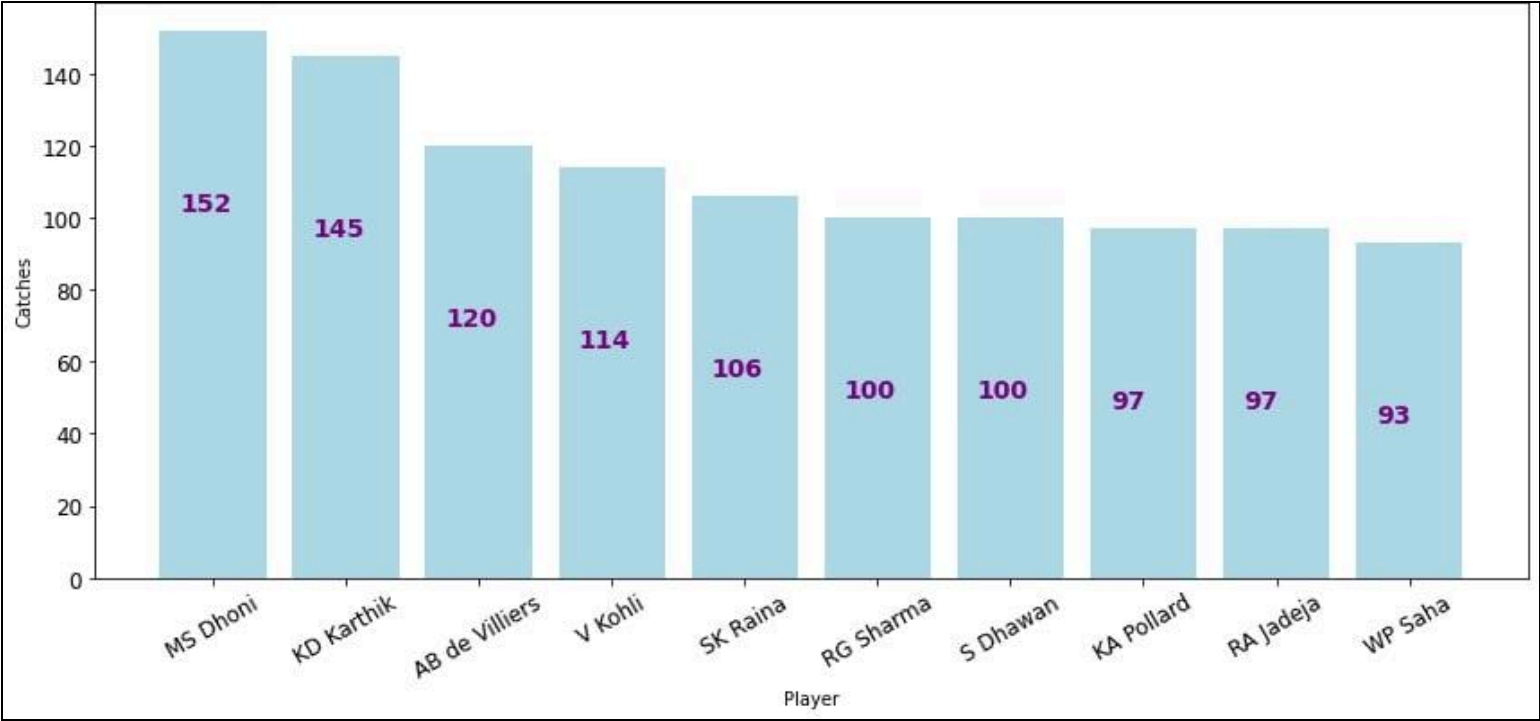## Most Economical bowlers in powerplay (Top 10) **min 50 overs bowled in powerplay
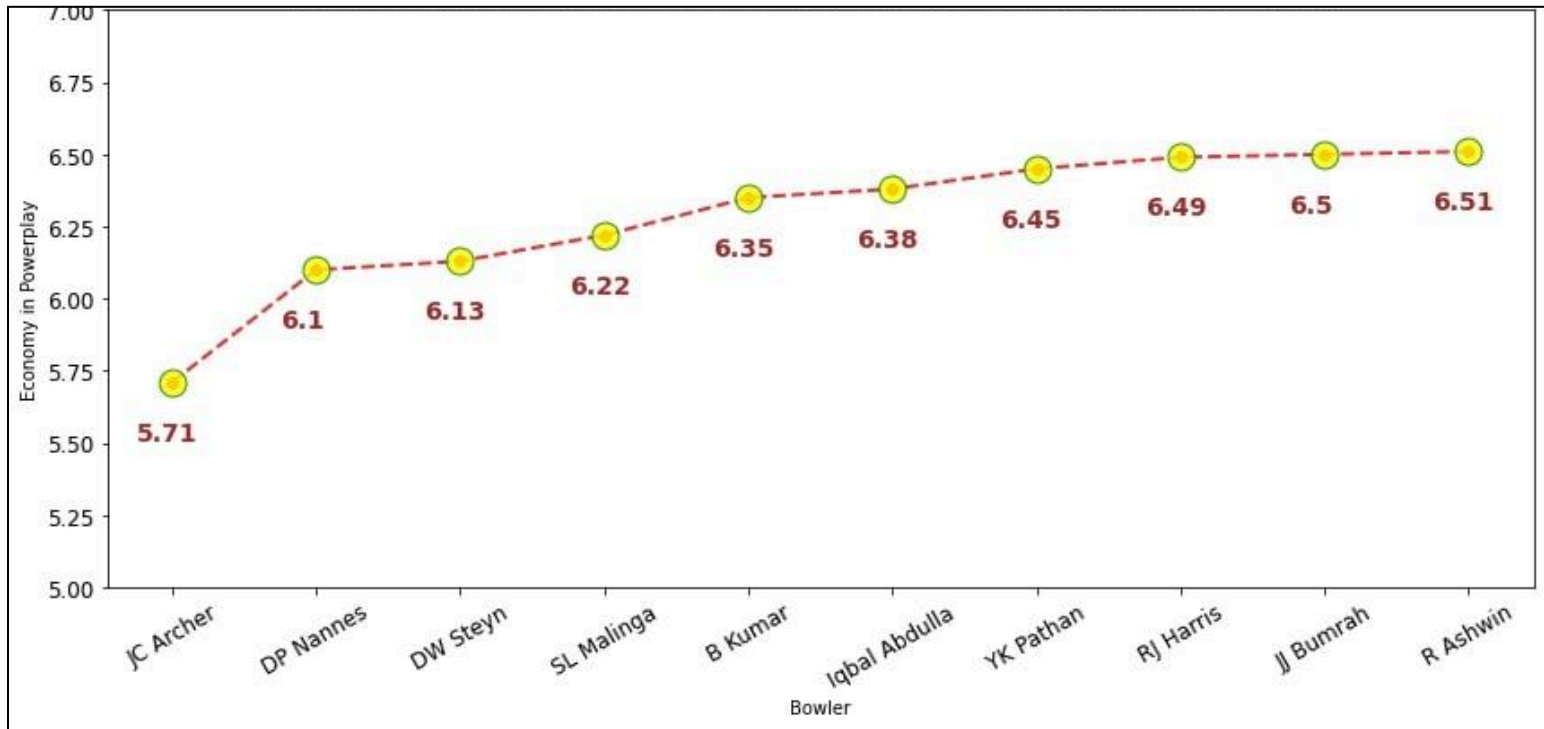
Highest run scorers in IPL

Highest Runs Scorers in IPL { Y-axis : Total_Runs / 100 , Z-axis : Batting_Average}
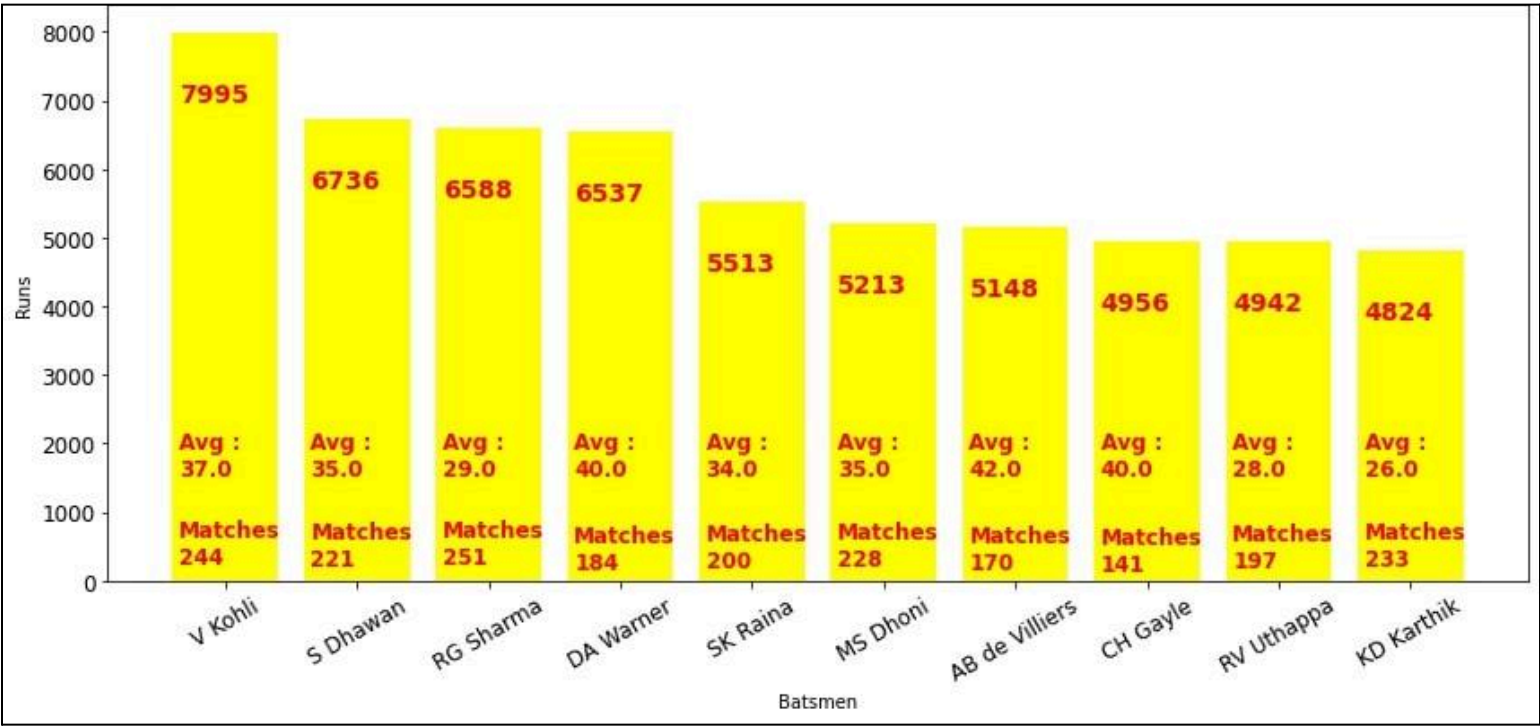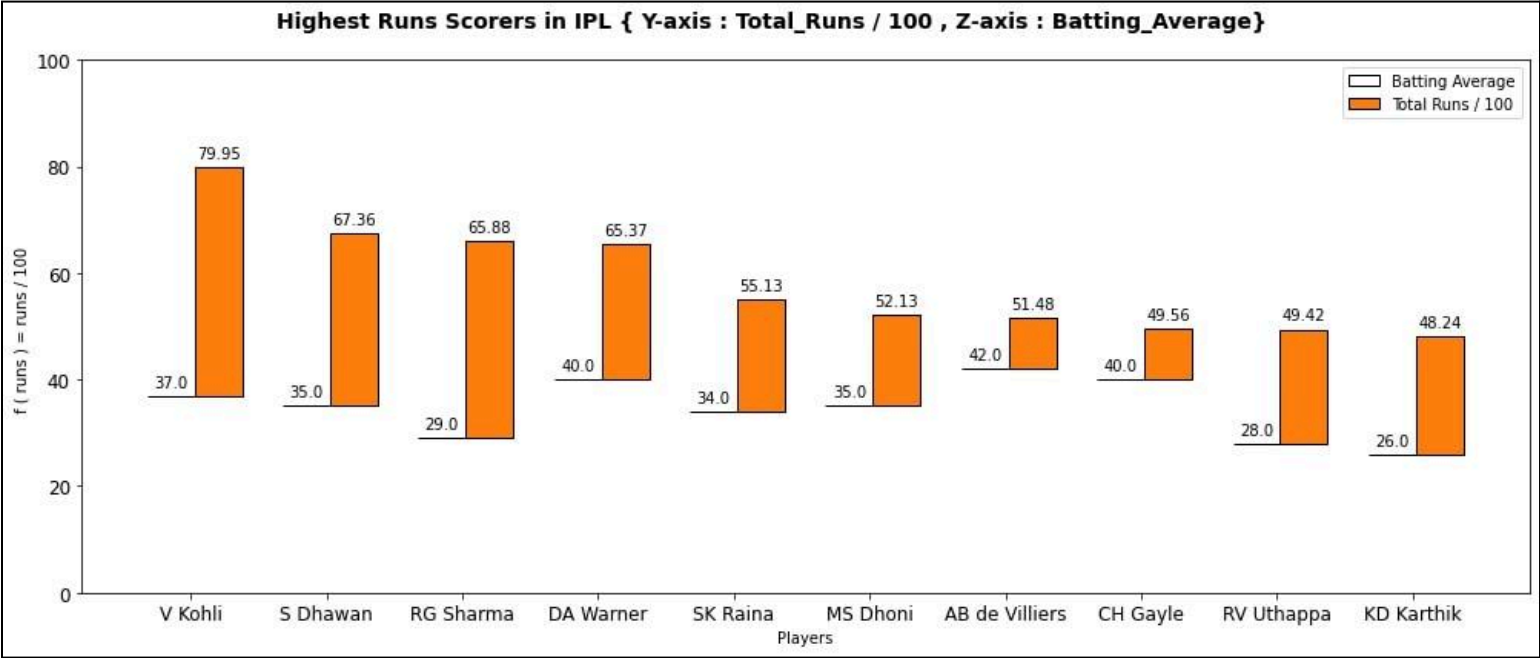
Highest wicket takers in IPL (Top 10)
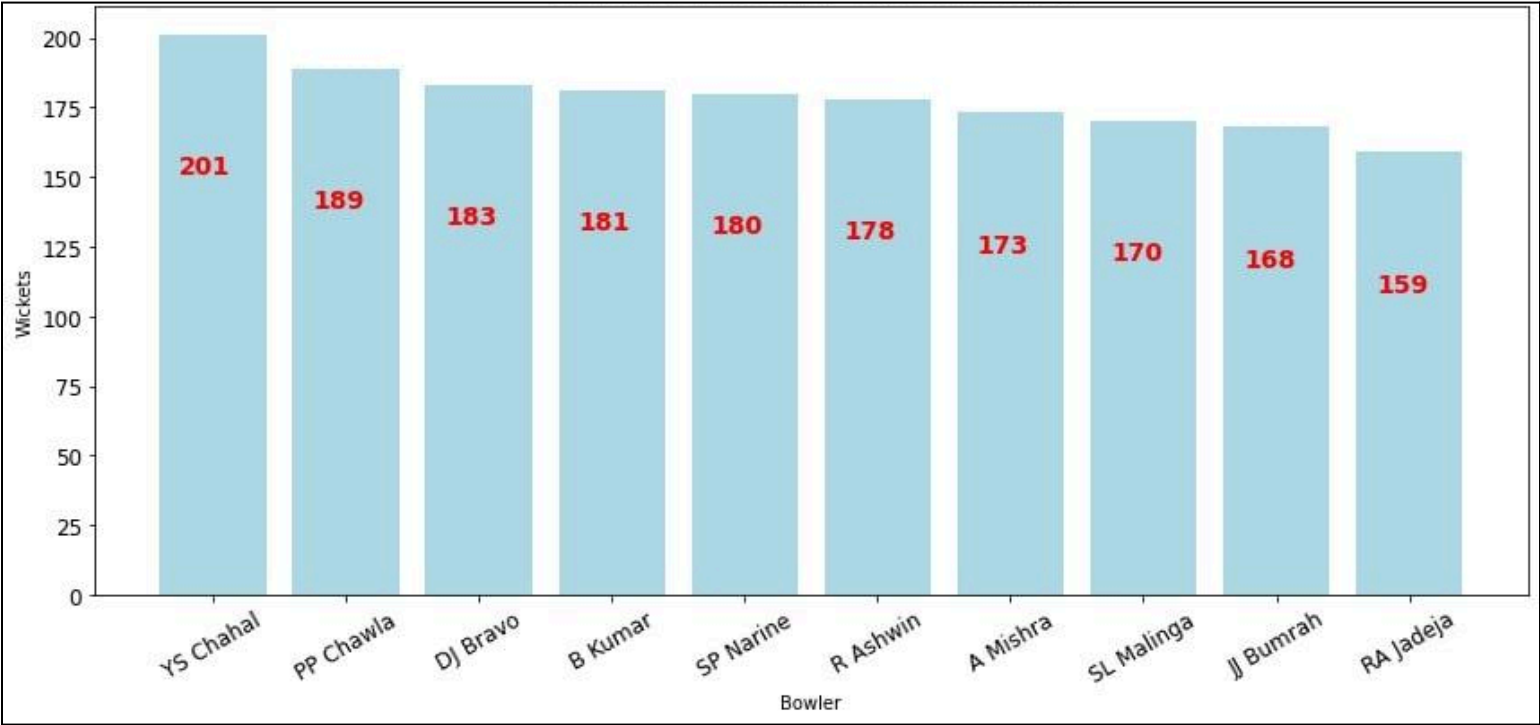
Most no. of catches in IPL (Top 10 fielders)

Most no. of stumpings (Top 10 Wicketkeepers)