IPL DATA ANALYSIS OVER THE YEARS

Project Work – Data Science

KEY AREAS OF ANALYSIS

- The main purpose of picking the top 10 record is to eliminate the extreme conditions(outliers) like batsman who have batted very less or bowlers who bowled very less. Main focus in on analyzing the data and get the best out of it in terms of players who have performed exceptionally well over a period of time.

- Overall performances of the IPL teams in terms of runs scored, wickets taken, win-loss ratio etc.

DATA VISUALIZATION AND PLOTTING

- Batsmen with most runs over the years
- Bowlers with most wickets
- Maximum sixes hit by players
- Maximum fours hit by players
- Maximum sixes by a player in a single match
- Maximum sixes by a player against a particular opposition
- Best economy rate with maximum numbers of over's bowled
- Most number of dots balls bowled by blowers with maximum over's bowled
- Overall performances of the team

- IPL teams win loss ratio
- Most MOM awards
- Bokeh Plotting of deliveries showing runs scored
- Swarm plots of balls bowled

PRE PROFILING

Warnings

`batsman` has a high cardinality: 488 distinct values `Warning`

`batsman_runs` has 65904 / 40.0% zeros `Zeros`

`bowler` has a high cardinality: 378 distinct values `Warning`

`bye_runs` has 164302 / 99.7% zeros `Zeros`

`bye_runs` is highly skewed (γ_1 = 30.052) `Skewed`

`dismissal_kind` has 156593 / 95.0% missing values `Missing`

`extra_runs` has 155872 / 94.6% zeros `Zeros`

`fielder` has 158832 / 96.4% missing values `Missing`

`fielder` has a high cardinality: 477 distinct values `Warning`

`legbye_runs` has 161989 / 98.3% zeros `Zeros`

`noball_runs` has 164093 / 99.6% zeros `Zeros`

`noball_runs` is highly skewed (γ_1 = 24.974) `Skewed`

`non_striker` has a high cardinality: 484 distinct values `Warning`

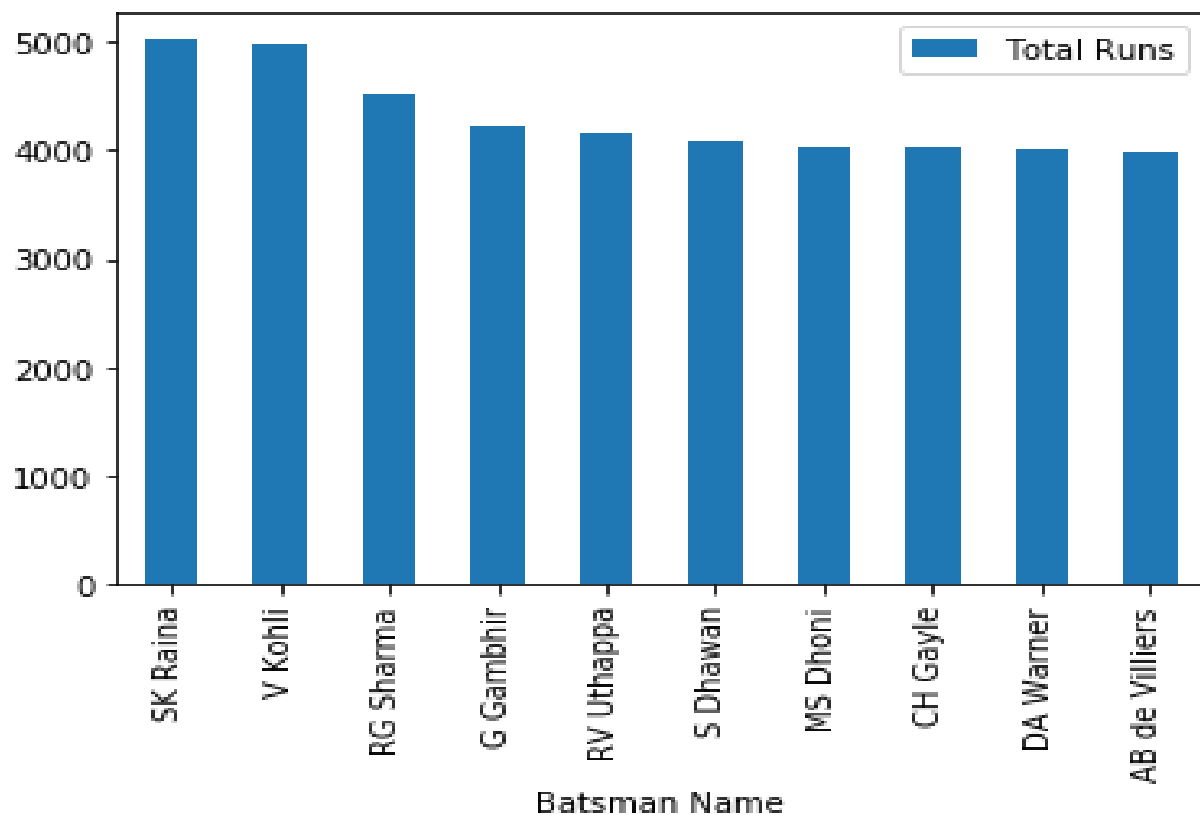`player_dismissed` has 156593 / 95.0% missing values `Missing`

`player_dismissed` has a high cardinality: 465 distinct values `Warning`

`total_runs` is highly correlated with `batsman_runs` (ρ = 0.97674) `Rejected`

`wide_runs` has 159739 / 97.0% zeros `Zeros`

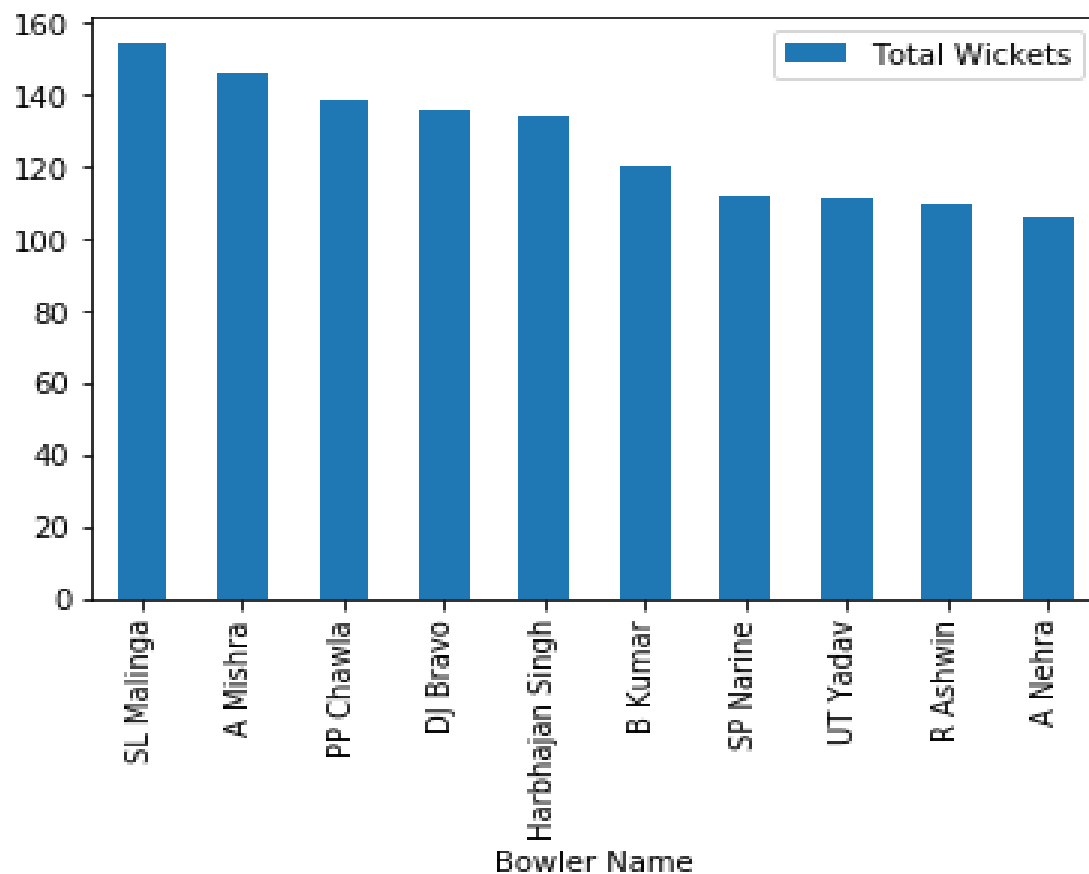Dataset has 5 duplicate rows `Warning`

BATSMEN WITH MOST RUNS (TOP 10)

	Batsman Name	Total Runs
0	SK Raina	5014
1	V Kohli	4962
2	RG Sharma	4504
3	G Gambhir	4223
4	RV Uthappa	4144
5	S Dhawan	4090
6	MS Dhoni	4041
7	CH Gayle	4037
8	DA Warner	4014
9	AB de Villiers	3974

BOWLERS WITH MOST WIKCETS (TOP 10)

	Bowler Name	Total Wickets
0	SL Malinga	154
1	A Mishra	146
2	PP Chawla	139
3	DJ Bravo	136
4	Harbhajan Singh	134
5	B Kumar	120
6	SP Narine	112
7	UT Yadav	111
8	R Ashwin	110
9	A Nehra	106

MAXIMUM SIXES HIT BY A PLAYER (TOP 10)
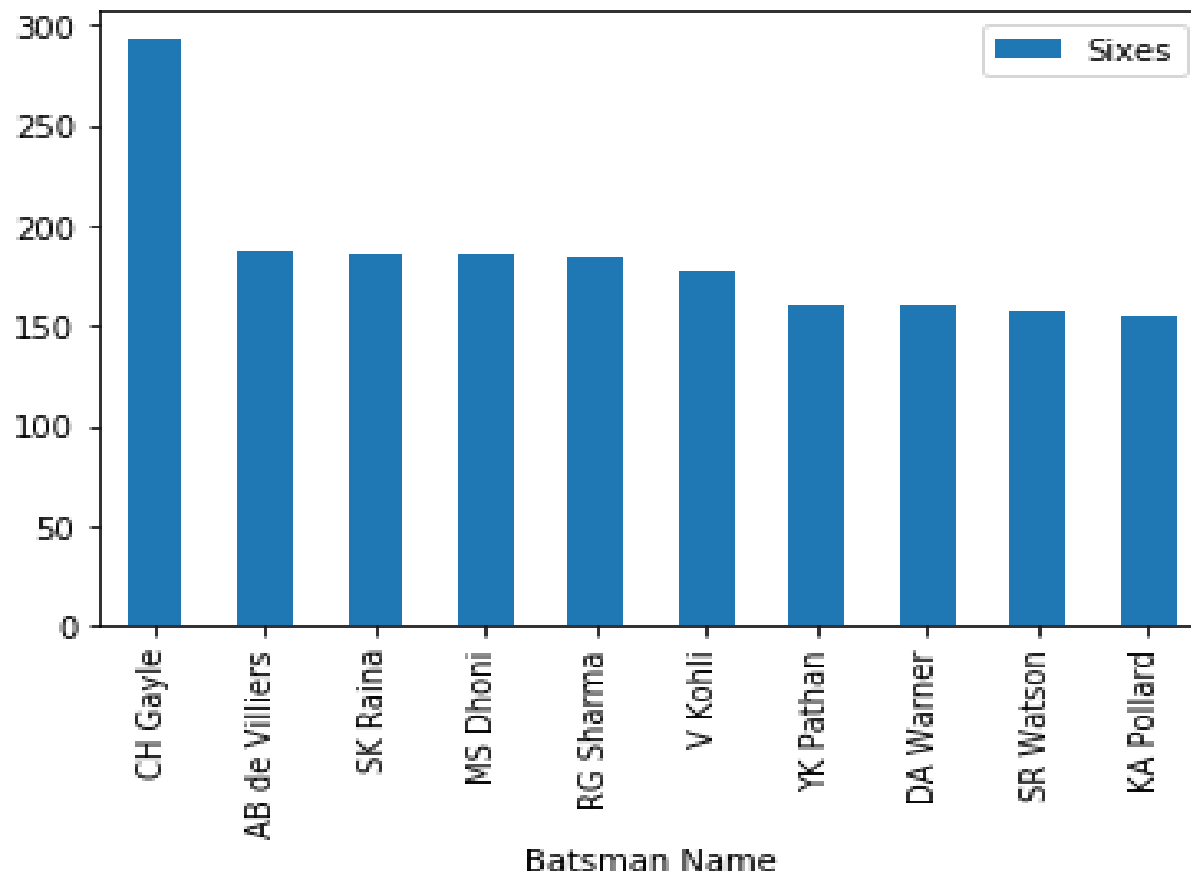
	Batsman Name	Sixes
0	CH Gayle	293
1	AB de Villiers	188
2	SK Raina	186
3	MS Dhoni	186
4	RG Sharma	185
5	V Kohli	178
6	YK Pathan	160
7	DA Warner	160
8	SR Watson	157
9	KA Pollard	155

MAXIMUM FOURS HIT BY A PLAYER (TOP 10)
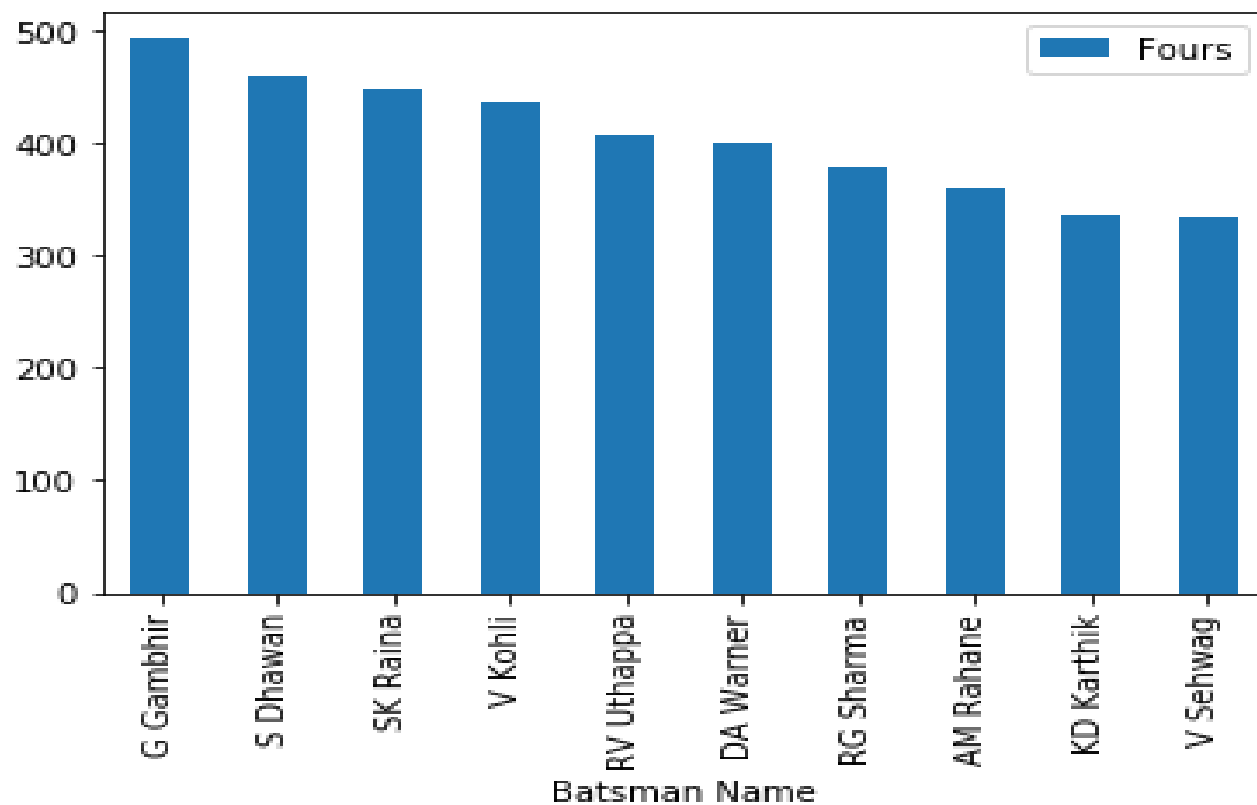
	Batsman Name	Fours
0	G Gambhir	492
1	S Dhawan	461
2	SK Raina	448
3	V Kohli	436
4	RV Uthappa	408
5	DA Warner	401
6	RG Sharma	379
7	AM Rahane	360
8	KD Karthik	336
9	V Sehwag	334

MAXIMUM SIXES BY A PLAYER IN A SINGLE MATCH (TOP 10)
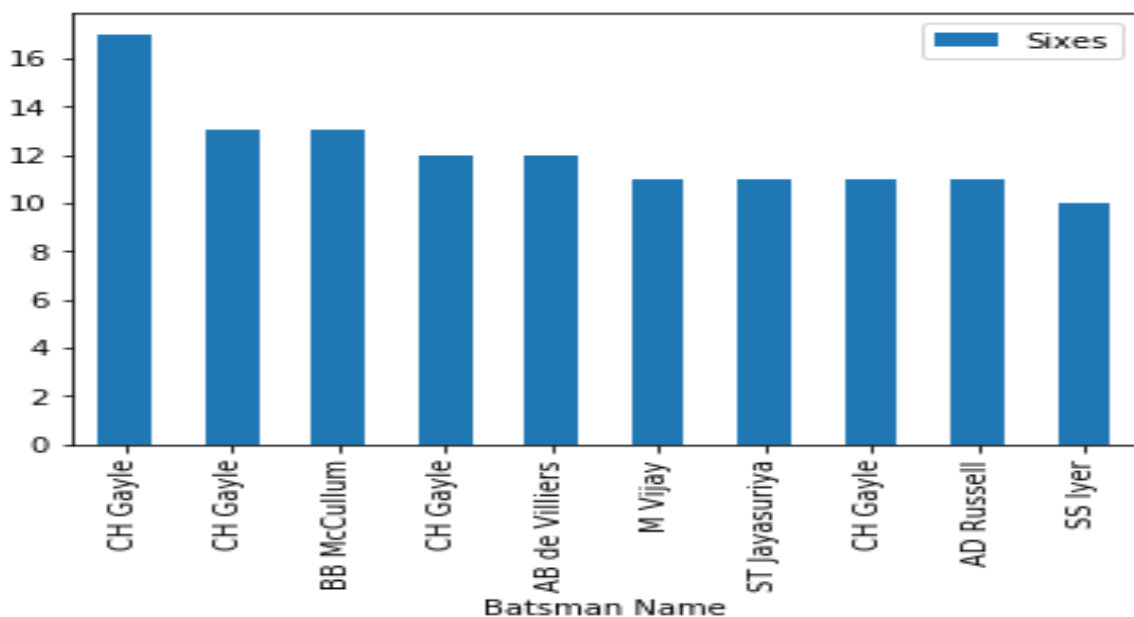
	Batsman Name	Match Number	Innings	Sixes	Opposition Name
0	CH Gayle	411	1	17	Pune Warriors
1	CH Gayle	372	1	13	Delhi Daredevils
2	BB McCullum	60	1	13	Royal Challengers Bangalore
3	CH Gayle	557	1	12	Kings XI Punjab
4	AB de Villiers	620	1	12	Gujarat Lions
5	M Vijay	206	1	11	Rajasthan Royals
6	ST Jayasuriya	96	2	11	Chennai Super Kings
7	CH Gayle	7909	1	11	Sunrisers Hyderabad
8	AD Russell	7898	1	11	Chennai Super Kings
9	SS Iyer	7919	1	10	Kolkata Knight Riders

# MAXIMUM SIXES BY A PLAYER AGAINST A PARTICULAR OPPOSITION (TOP 10)
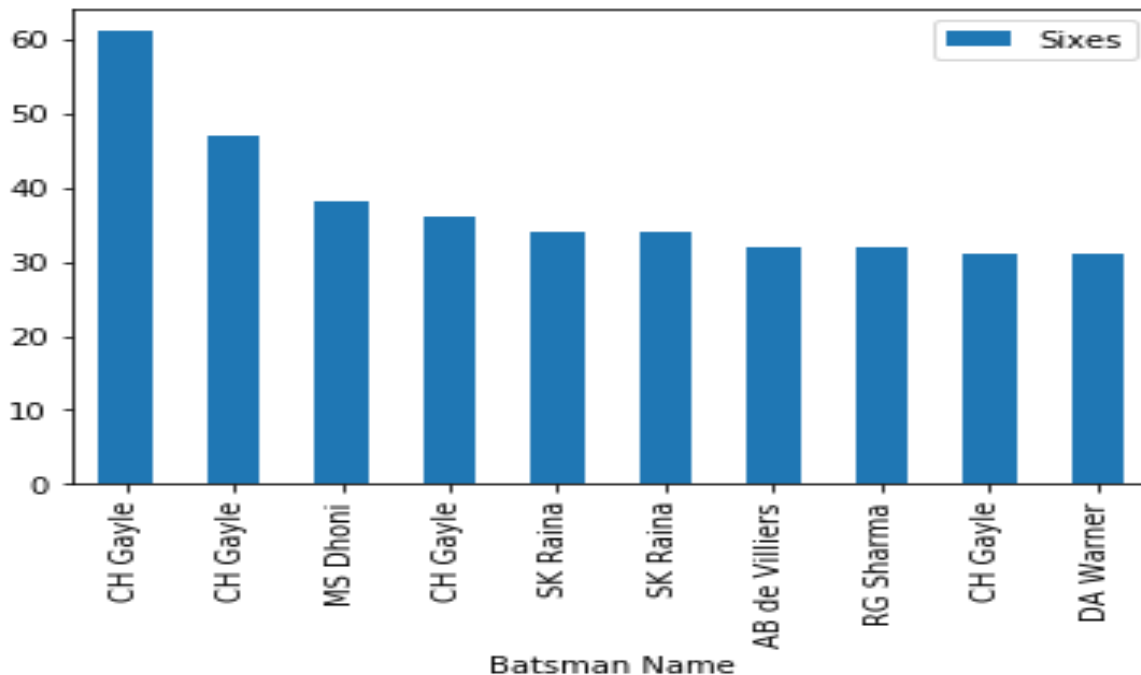
Batsman Name	Opposition Name	Sixes
CH Gayle	Kings XI Punjab	61
CH Gayle	Kolkata Knight Riders	47
MS Dhoni	Royal Challengers Bangalore	38
CH Gayle	Pune Warriors	36
SK Raina	Kings XI Punjab	34
SK Raina	Mumbai Indians	34
AB de Villiers	Kings XI Punjab	32
RG Sharma	Delhi Daredevils	32
CH Gayle	Sunrisers Hyderabad	31
DA Warner	Royal Challengers Bangalore	31

# BEST ECONOMY RATE WITH MAXIMUM NUMBERS OF OVER'S BOWLED (TOP 10)
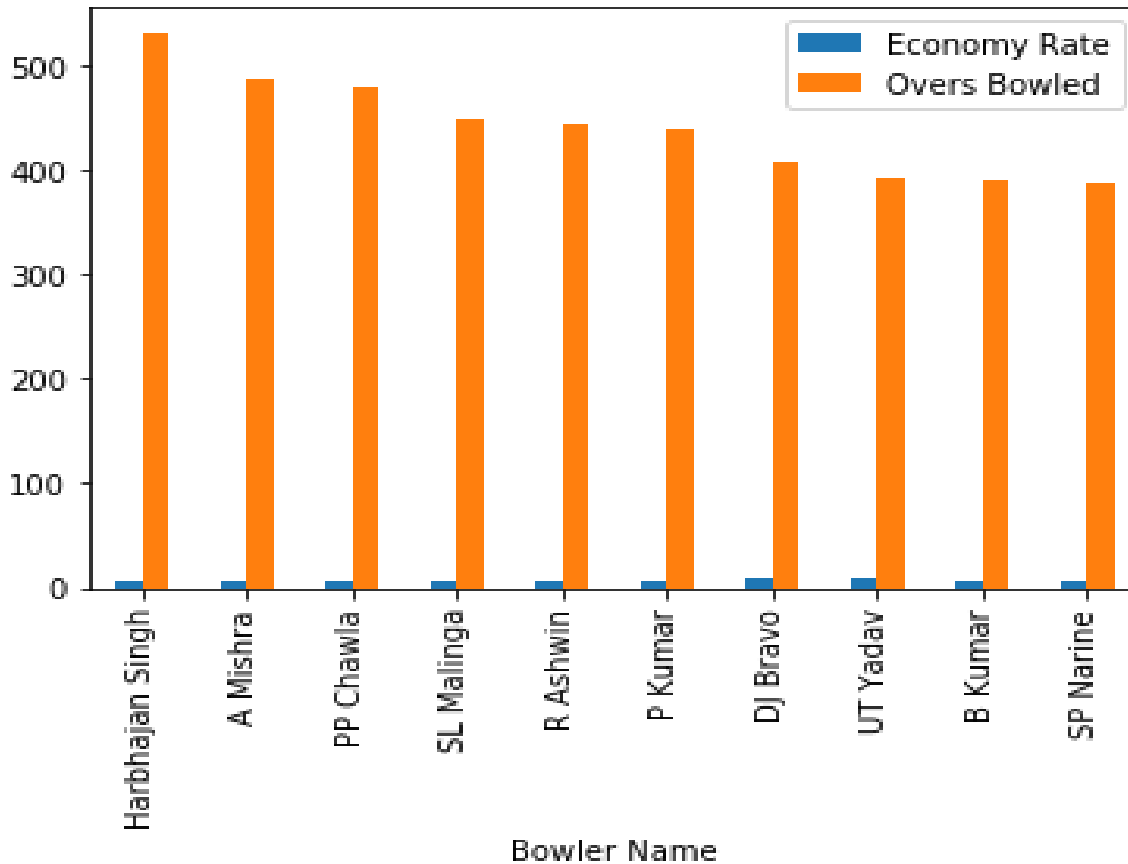
	Bowler Name	Overs Bowled	Economy Rate
0	Harbhajan Singh	530	7.0283
1	A Mishra	488	7.32787
2	PP Chawla	481	7.77963
3	SL Malinga	449	6.75724
4	R Ashwin	445	6.68539
5	P Kumar	439	7.61276
6	DJ Bravo	408	8.27206
7	UT Yadav	394	8.24619
8	B Kumar	390	7.12564
9	SP Narine	387	6.63824

MOST DOT DELIVERIES BOWLED IN IPL VS OVER'S COMPARISON
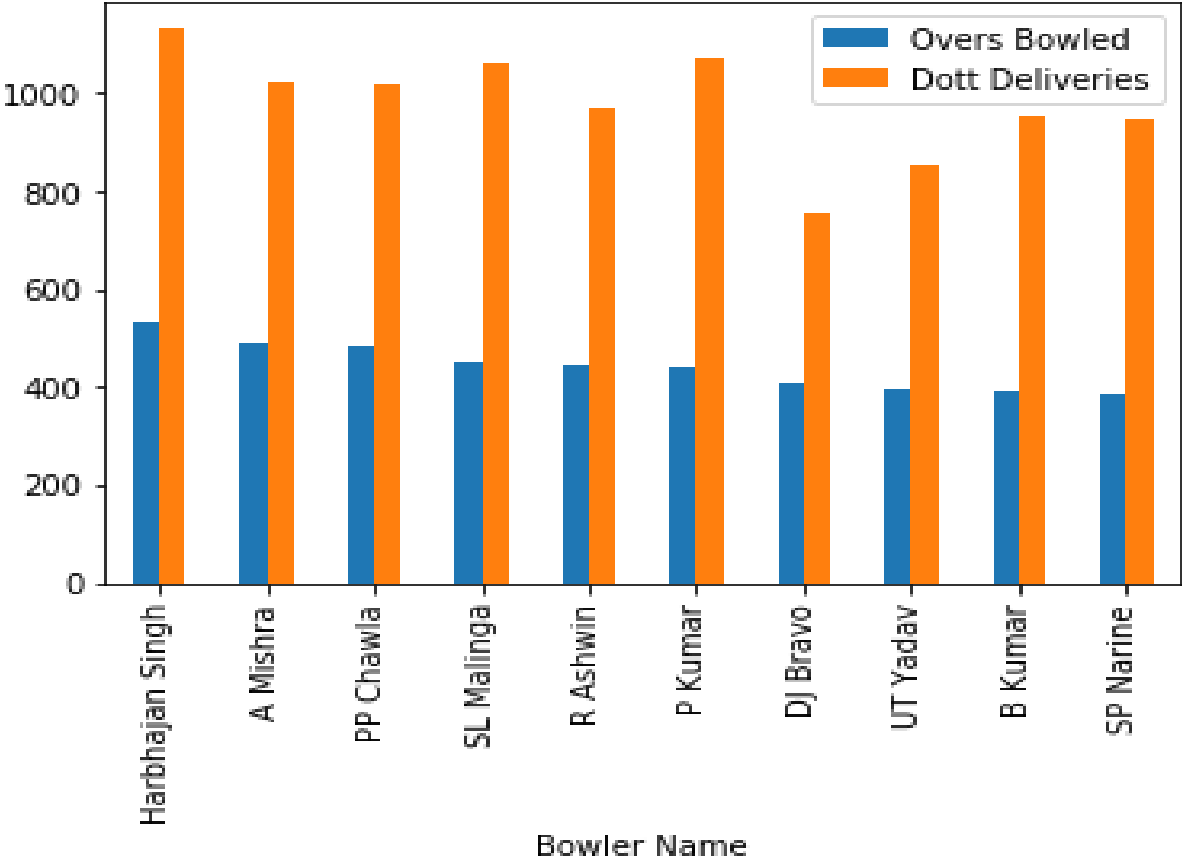
	Bowler Name	Overs Bowled	Dott Deliveries
0	Harbhajan Singh	530	1130
1	A Mishra	488	1026
2	PP Chawla	481	1016
3	SL Malinga	449	1060
4	R Ashwin	445	967
5	P Kumar	439	1075
6	DJ Bravo	408	754
7	UT Yadav	394	855
8	B Kumar	390	953
9	SP Narine	387	948

OVERALL PERFORMANCES OF THE TEAM
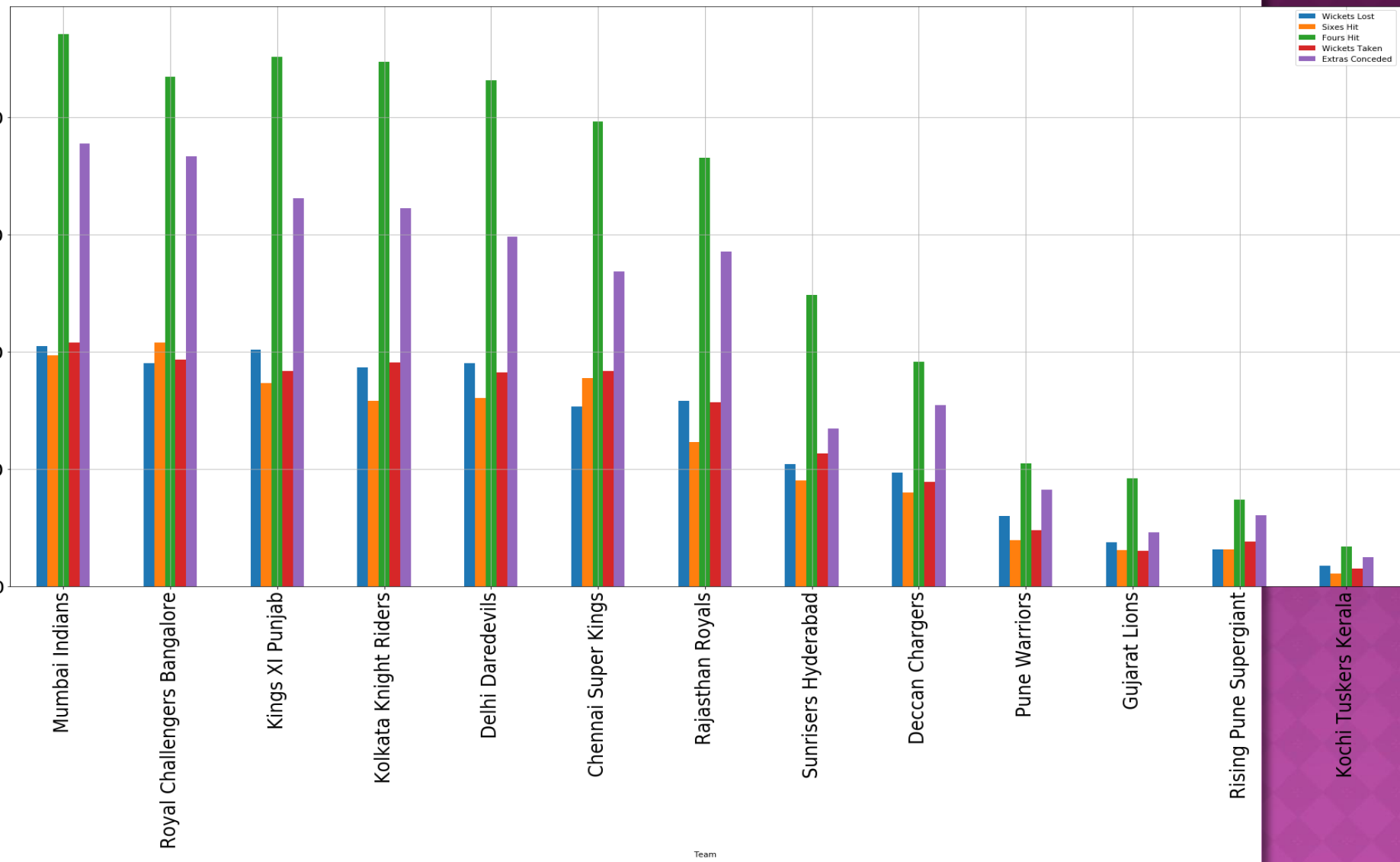
IPL TEAMS WIN LOSS COMPARISON
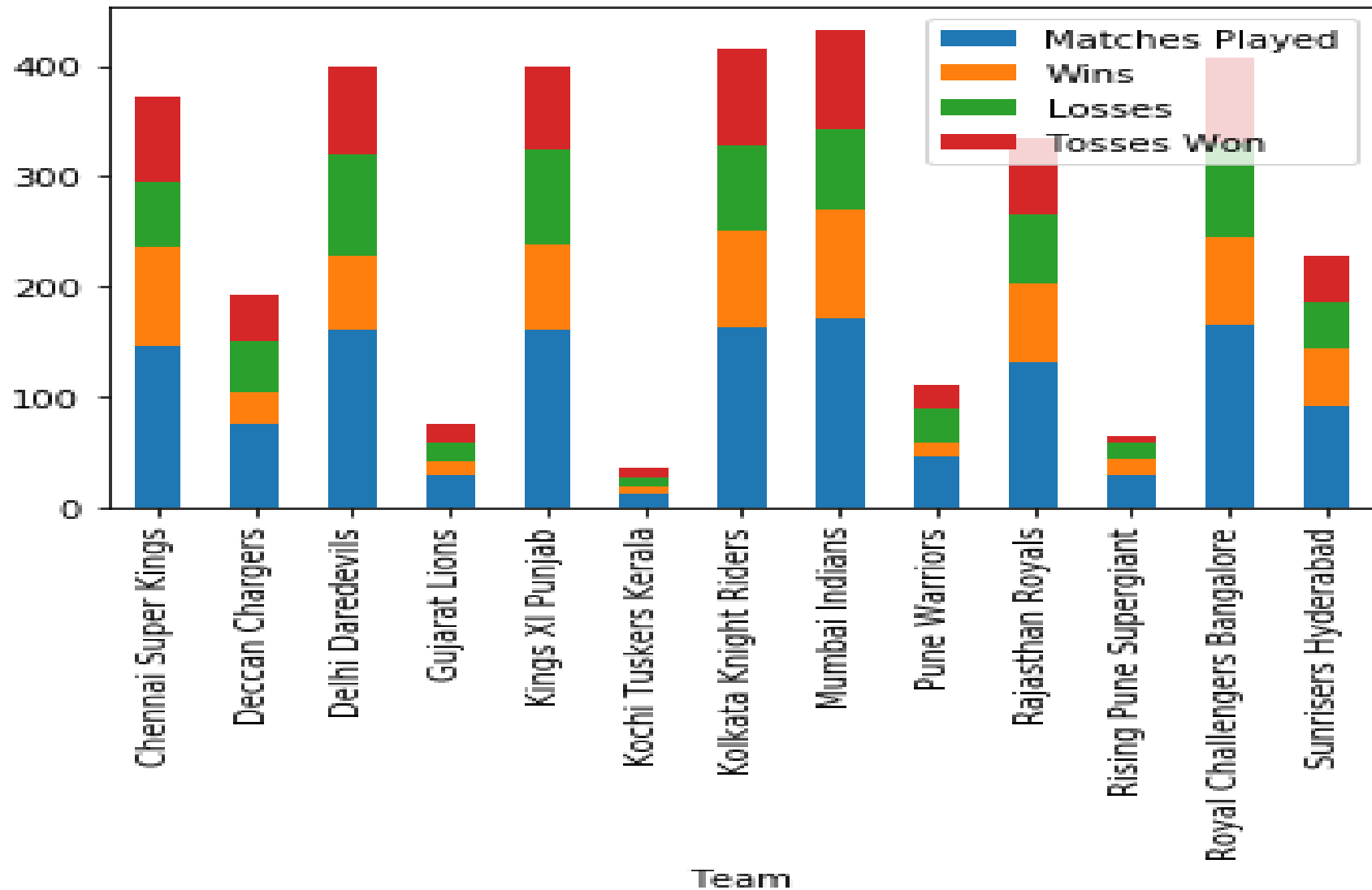
MOST MOM AWARDS
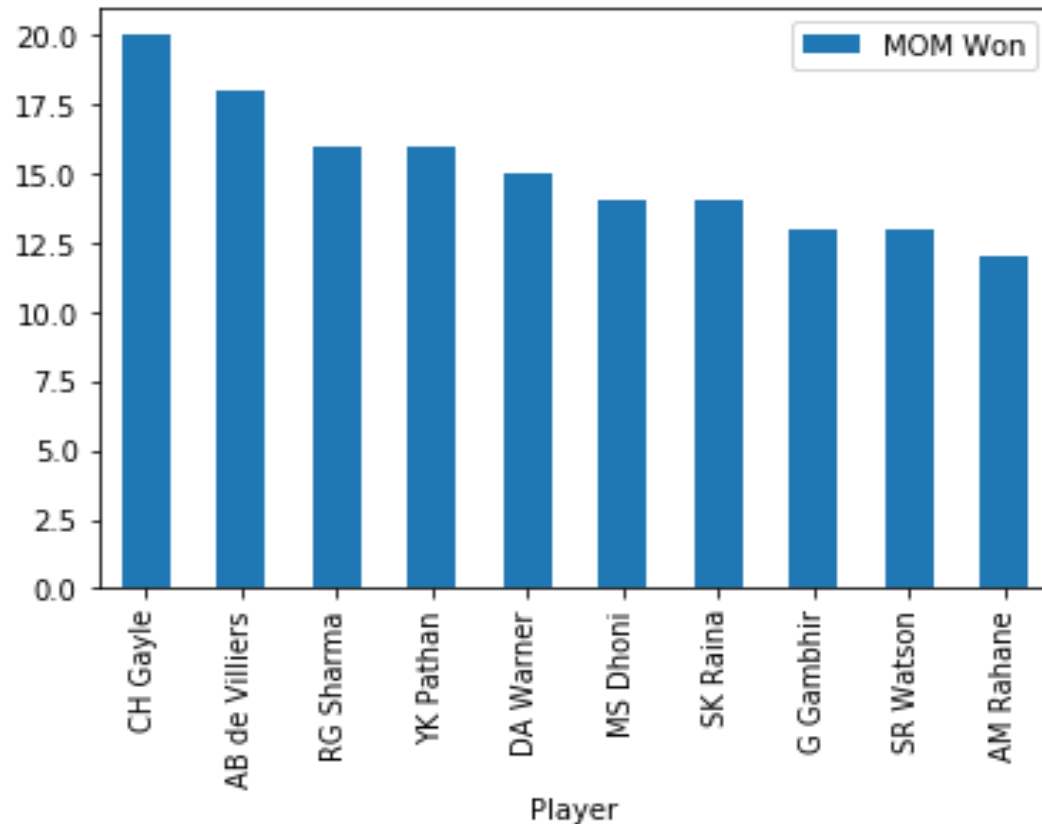
	Player	MOM Won
0	CH Gayle	20
1	AB de Villiers	18
2	RG Sharma	16
3	YK Pathan	16
4	DA Warner	15
5	MS Dhoni	14
6	SK Raina	14
7	G Gambhir	13
8	SR Watson	13
9	AM Rahane	12

BOKEH PLOTTING OF DELIVERIES BASED ON SAMPLE
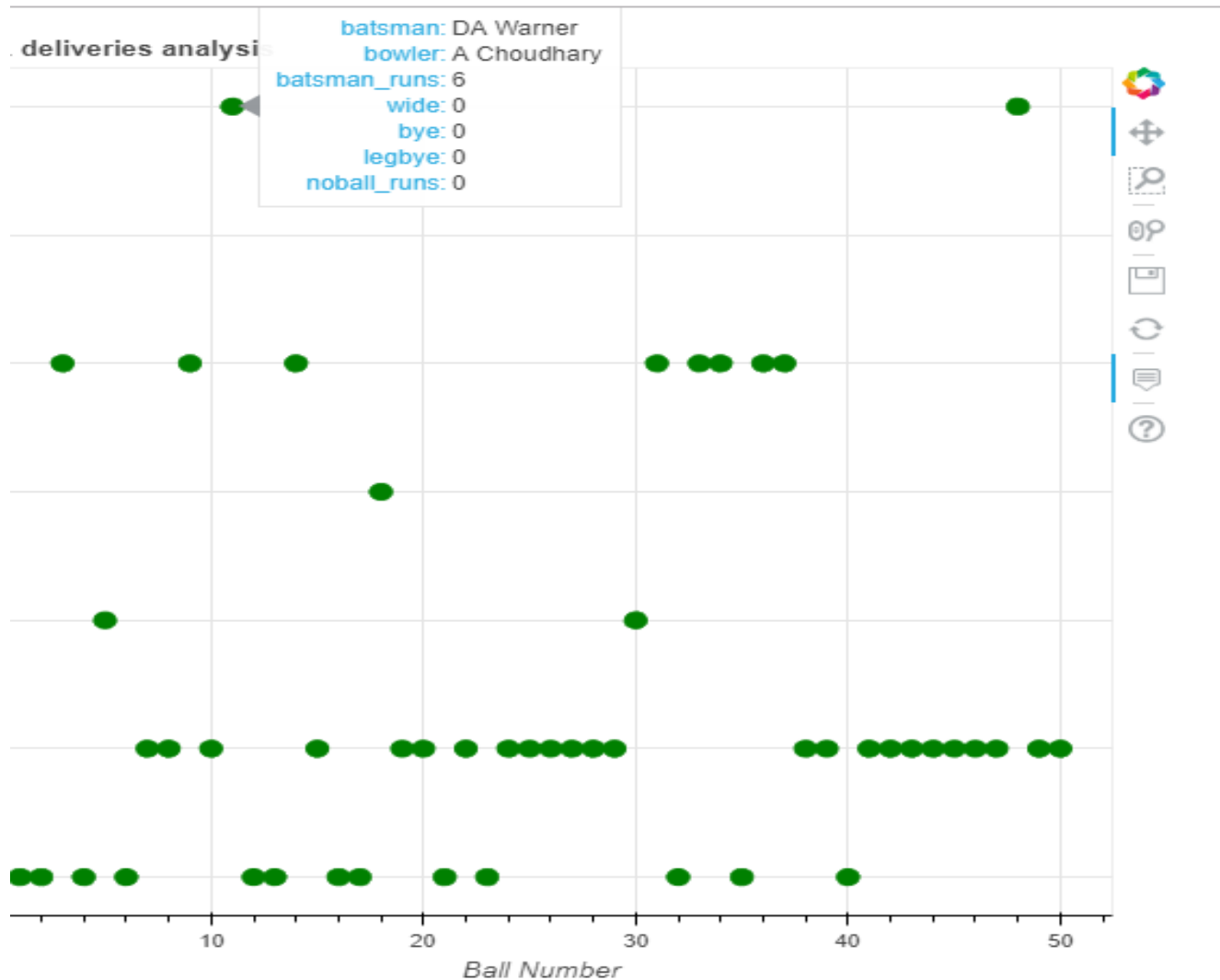
SWARM PLOT OF BALLS BOWLED BASED ON SAMPLE