

Abstract:

Cricket is a religion in INDIA, Billion people follow cricket in INDIA. People follow cricketers as their role models, Now a days T20 cricket has gained lot of popularity and IPL is followed by every house member in India and overseas people also follow it since it has been a tournament with tremendous competition.

World's best player gather to form a team and compete against each other.

Here we are trying to analyze the data from the dataset provided of matches.csv and deliveries.csv

Analysis and Visualizations of dataset matches.csv and deliveries.csv are considered.

Here Visualizations are done with respect to Batsmen, Bowlers, Players of the match, Team winning the match on winning the Toss, and PIE chart plots are added.

Introduction:

Here matches.csv and deliveries.csv are the files used as dataset and is read using Pandas, matplotlib and seaborn are used to plot graphs for the various scenarios we have considered. This data analysis and plots help in evaluating performance of a player over the years , how consistent a player is , which All-rounder is better, which team has been outstanding throughout all the IPL seasons, if a new team is formed whom can be chosen as team members based on their previous performances and current form of the player. Visualization mainly summarizes the player performance and helps in choosing a team which performs better as per analytics. Visualization of predictions also help to know what factors to be considered to play against a team at a particular Venue.

Discussion and Results :

Deliveries.csv : 179079 rows.

Matches.csv: 757 rows.

Dataset Walk through.

Matches.csv

Id - Each IPL Match is associated with an id , this is a unique id for each IPL match. In matches.csv Id has been considered as serial number.

Season – Season is the IPL season year, informs the IPL data for a particular year.

City – Informs the date at which IPL match was played.

team1 – informs the first team participating in the IPL match.

team2 – informs the second team participating in the IPL match.

toss_winner – informs which team won the toss.

toss_decision – informs whether the team which won the toss decided to bat or field.

IPL Data Analysis

Result – if its win or loose then result is displayed as “normal”, if the match is drawn i.e if both the teams score the same number of runs then result is displayed as “tie”.

dl_applied – Duckworth Luis method applied or not, when the ongoing match is cancelled due to rain or bad whether then Duckworth Luis method is applied to inform who is the winner of the match, so this column informs whether dl method was applied in the match or not.

If Duckworth Luis method is not applied then 0.

If Duckworth Luis method is applied then 1.

Winner – Informs who won the match whether team1 or team2

win_by_runs – informs by how many runs the team won. This indicates team which batted first have won the match and the team batting second has lost the match.

win_by_wickets - – informs by how many wickets the team won. This indicates team which bowled first have won the match and the team bowled second has lost the match.

player_of_match – Informs the player who was responsible for the team win , that player is awarded as Player of the match and the same has been recorded in the “Player of match” column.

Venue – Informs where the match was played, which ground the match was organized.

umpire1 – is the umpire who is responsible for giving decisions in the match, same Umpire name has been mentioned in “umpire1” column

umpire2 – is the umpire who is responsible for giving decisions in the match suns as run outs etc., same Umpire name has been mentioned in “umpire2” column

umpire3 – is the umpire who is responsible for giving decisions in the match,umpire1 and umpire2 refer to the third umpire for the ones which they cannot decide.

Dataset Walk through.

deliveries.csv

Ball by Ball Data of the match.

match_id – Informs the unique id for a particular IPL match. Here match id is considered in sequential order.

Inning – informs whether its First Innings or Second Innings of the match. Value 1 indicates its first innings of the match and value 2 indicates its the second innings of the match.

batting_team – Team batting in the particular innings is displayed.

bowling_team – Team bowling in the particular Innings is displayed.

Over – which over of the match is being bowled.

Ball – There would be six balls in an over this will inform which ball in an over is mentioned.

IPL Data Analysis

Batsmen – One who is batting for the current over and ball is mentioned.

non_striker – One who is not batting and present at the bowlers end for current over and ball data is called non_striker.

is_super_over – When the match is tied we will bowl a Super Over, to achieve the result of the match the same is mentioned.

0 specifies no super over and 1 specifies there was a super over in the match.

wide_runs – Number of runs consumed by the team by bowling improper balls ,invalid ball. Here the ball is out of the reach of batsmen hence called Wide.

bye_runs – The runs obtained by flicking on body called bye runs. when over throw occurs while running between the wickets, these runs are considered as bye runs.

0 – no bye_runs

1,2,3,4 – bye_runs are counted for the particular ball.

legbye_runs – Rusn obtained by flicking a tghigh pad or pad are called legbye runs.

0 – no legbye_runs

1,2,3,4,5 – legbye_runs are counted for the particular ball.

noball_runs – Runs obtained by no_balls which happens on overstepping the crease while doing the bowling.

0 – no noball_runs

1,2,3,4,5 – noball_runs are counted for the particular ball.

penalty_runs – Suppose ball hits the helmet placed on ground when a ball is bowled , these runs are considered as penalty runs.

0 indicates no penalty runs and 5 indicates penalty runs.

batsman_runs – Runs scored by the batsmen, leaving bye_runs,legbye_runs.

0 – no runs scored by batsmen

1,2,3,4,5,6,7 – runs scored by batsmen for the particular ball.

Extra runs – Can eb considered as bye_runs or legbye_runs.

0 – no extra runs

1,2,3,4,5,6,7 – extra run for the particular ball.

total_runs – Total runs counted towards the score board for a particular ball.

0 – no runs counted towards the score board.

IPL Data Analysis

1,2,3,4,5,6,7,8,9,10 – Total runs counted towards the score board.

We will count in `bye_runs`, `legbye_runs`, `noball_runs`, `penalty_runs` batsman_runs, Extra runs while calculating `total_runs` for a scoreboard.

player_dismissed – informs the player being dismissed during a particular ball in an over.

dismissal_kind – Informs the way in which the player got dismissed.

Bowled, caught, caught and bowled, hit wicket, lbw, obstructing the field, retired hurt, run out, stumped.

Fielder – informs the name of the fielder who is responsible for out.

Results are attached in the form of plots.

Team names might have changed from the IPL playing years. Hence Team names and Ground name with minor changes are renamed to follow a common name or convention to avoid conflicts in the results computation. Missing or Nan values in city, Winner, Player of match, umpir1, umpire2 are followed by appropriate values to make sure we have meaningful data and further computation can be done.

I have merged matches.csv and deliveries .csv so that it will be easy to take the required data from one data frame of Pandas, List of IPL season is considered and iteration through the whole list is done which specifies the best bowler, best batsmen, best player, Match winning team in all IPL seasons.

Here I Plot

Top 20 batsmen from IPL season 2008 to IPL season 2019.

Top 20 bowler from IPL season 2008 to IPL season 2019.

Top 20 Player of the match from IPL season 2008 to IPL season 2019.

Team won on electing to bat and field from IPL season 2008 to IPL season 2019.

For further plotting I merge cells of my dataframe.

Innings, match, and corresponding runs being scored are merged and this is merged with IPL dataframe for Visualizations.

Number of wins by each team in particular city is plotted in bar graph and it is plotted for all the IPL seasons.

Player of the match data is collected from the Pandas Data frame and the Top 5 Player of the match are plotted in bar graph specifying the number of players of match award won by them.

Top 3 teams which has won maximum number of matches batting first are extracted and bar chart is plotted specifying the number of wins each team has got.

Percentage of wins each team has got batting first in the entire IPL season from 2008 to 2019. A pie chart is plotted for the data of percentage wins.

IPL Data Analysis

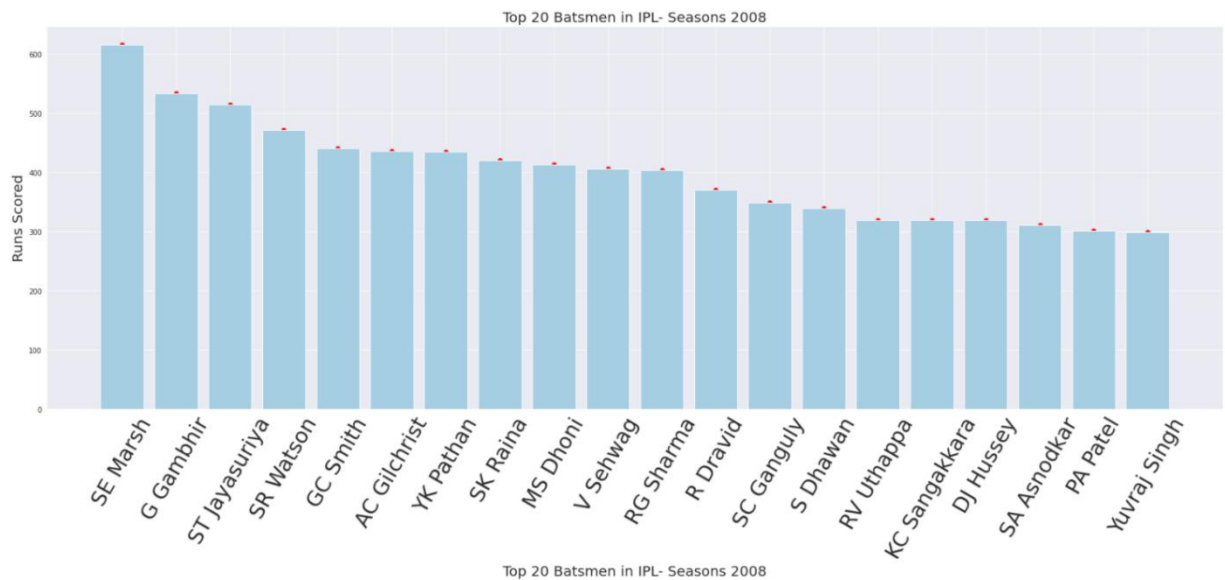
Plot Description:

Here Top 20 batsmen in IPL Season 2008 are mentioned.

x-axis represents the name of the Player and Y-axis represents the runs scored in IPL season 2008

The Top 10 Batsmen in 2008:

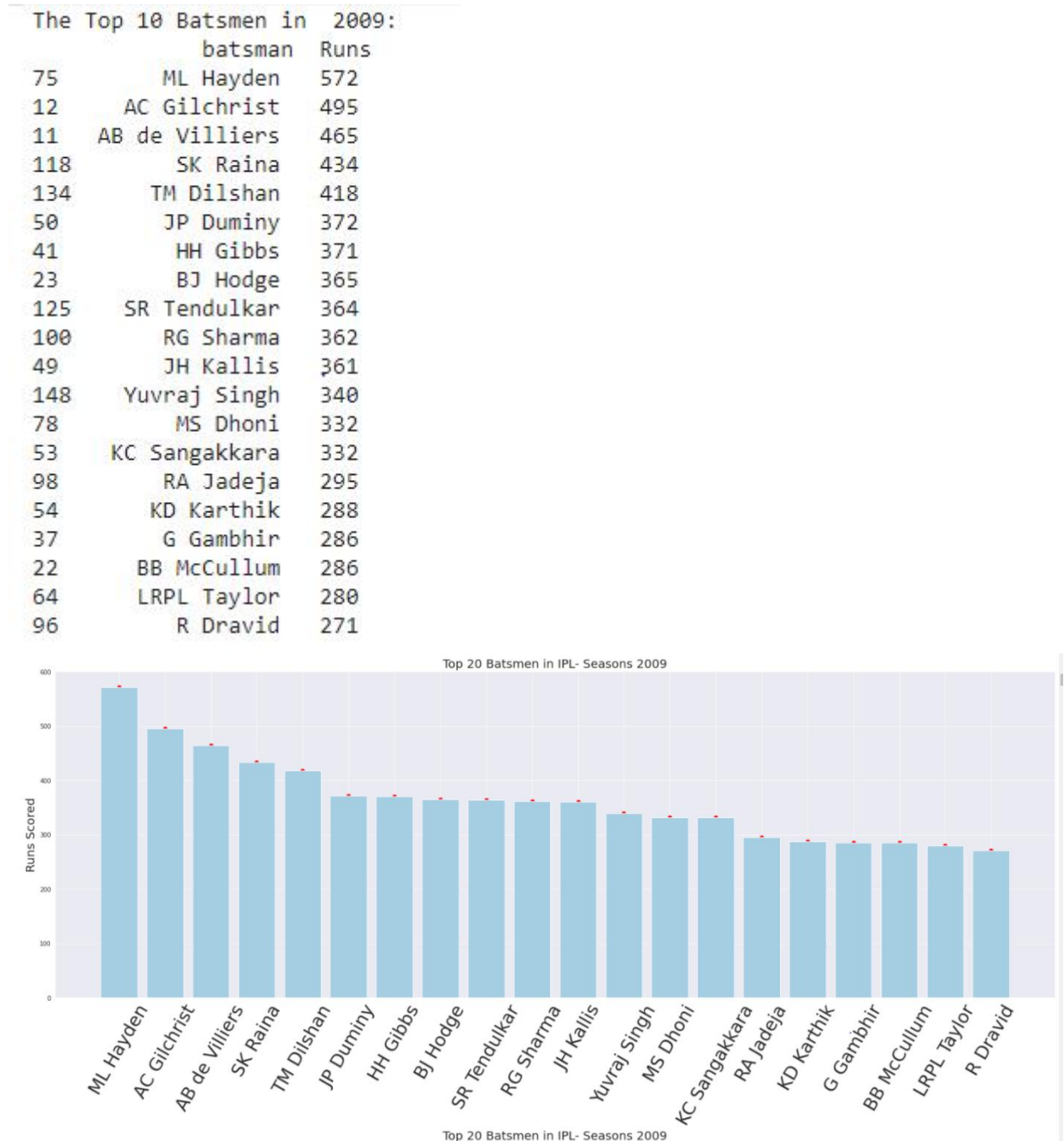
	batsman	Runs
115	SE Marsh	616
39	G Gambhir	534
126	ST Jayasuriya	514
124	SR Watson	472
40	GC Smith	441
10	AC Gilchrist	436
149	YK Pathan	435
116	SK Raina	421
76	MS Dhoni	414
139	V Sehwag	406
95	RG Sharma	404
92	R Dravid	371
113	SC Ganguly	349
105	S Dhawan	340
101	RV Uthappa	320
55	KC Sangakkara	320
29	DJ Hussey	319
109	SA Asnodkar	311
84	PA Patel	302
152	Yuvraj Singh	299



IPL Data Analysis

Here Top 20 batsmen in IPL Season 2009 are mentioned.

x-axis represents the name of the Player and Y-axis represents the runs scored in IPL season 2009



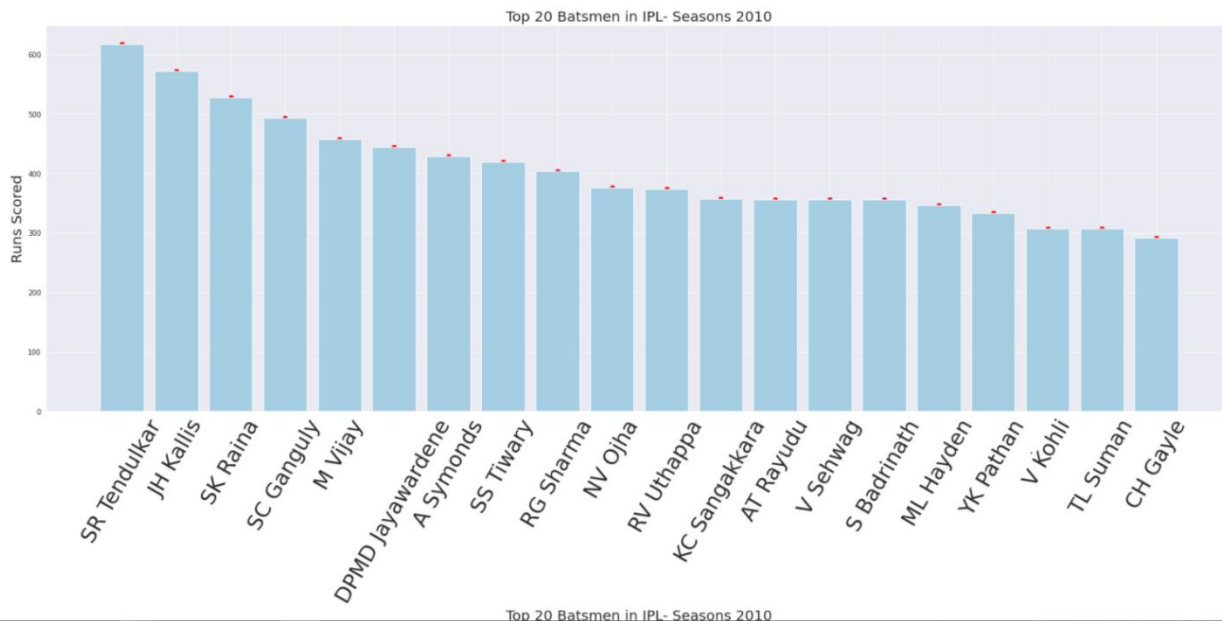
Here Top 20 batsmen in IPL Season 2010 are mentioned.

x-axis represents the name of the Player and Y-axis represents the runs scored in IPL season 2010

IPL Data Analysis

The Top 10 Batsmen in 2010:

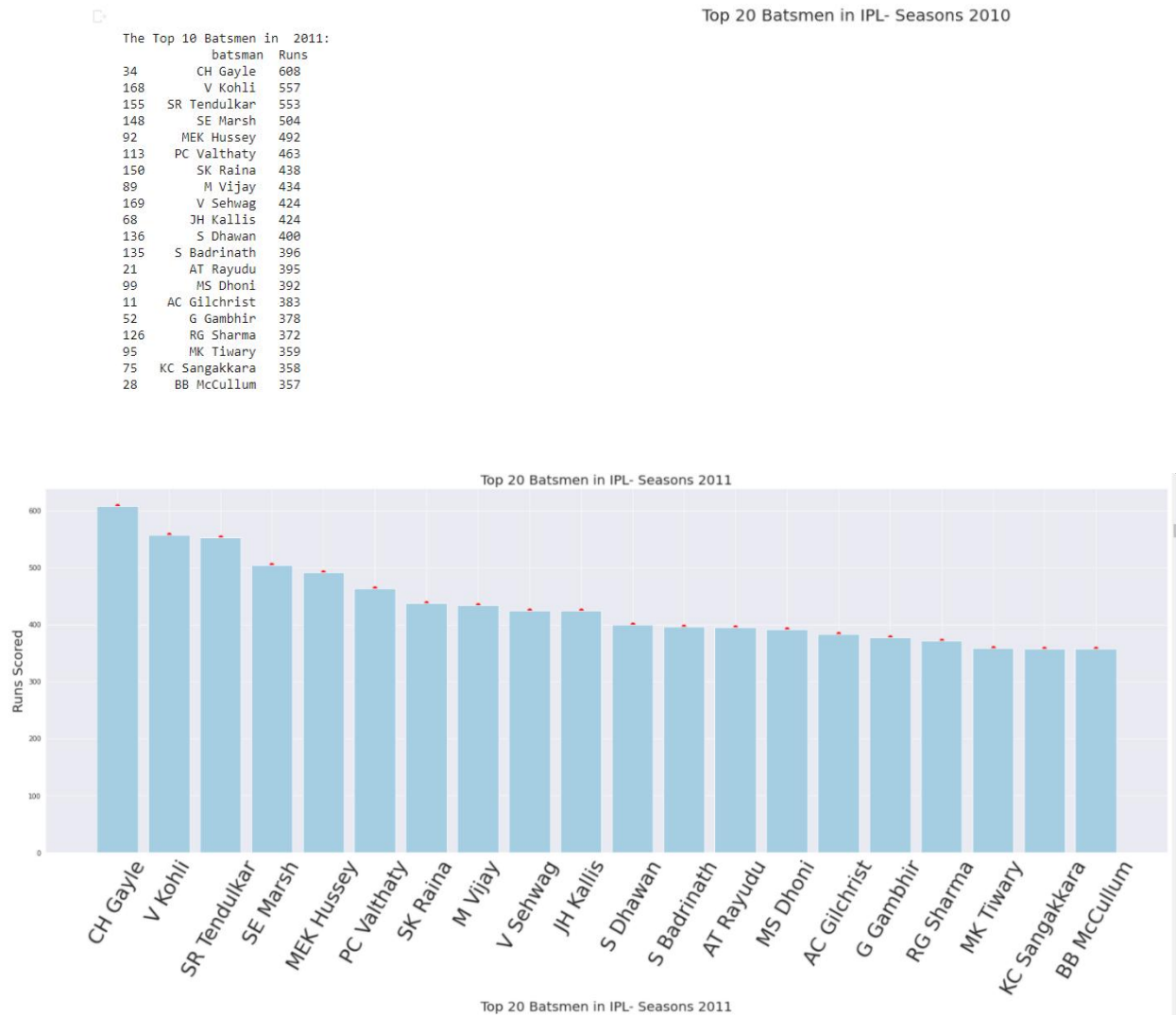
	batsman	Runs
142	SR Tendulkar	618
61	JH Kallis	572
137	SK Raina	528
133	SC Ganguly	493
82	M Vijay	458
44	DPMD Jayawardene	445
4	A Symonds	429
144	SS Tiwary	419
116	RG Sharma	404
98	NV Ojha	377
123	RV Uthappa	374
69	KC Sangakkara	357
24	AT Rayudu	356
153	V Sehwag	356
125	S Badrinath	356
90	ML Hayden	346
161	YK Pathan	333
152	V Kohli	307
149	TL Suman	307
34	CH Gayle	292



IPL Data Analysis

Here Top 20 batsmen in IPL Season 2011 are mentioned.

x-axis represents the name of the Player and Y-axis represents the runs scored in IPL season 2011



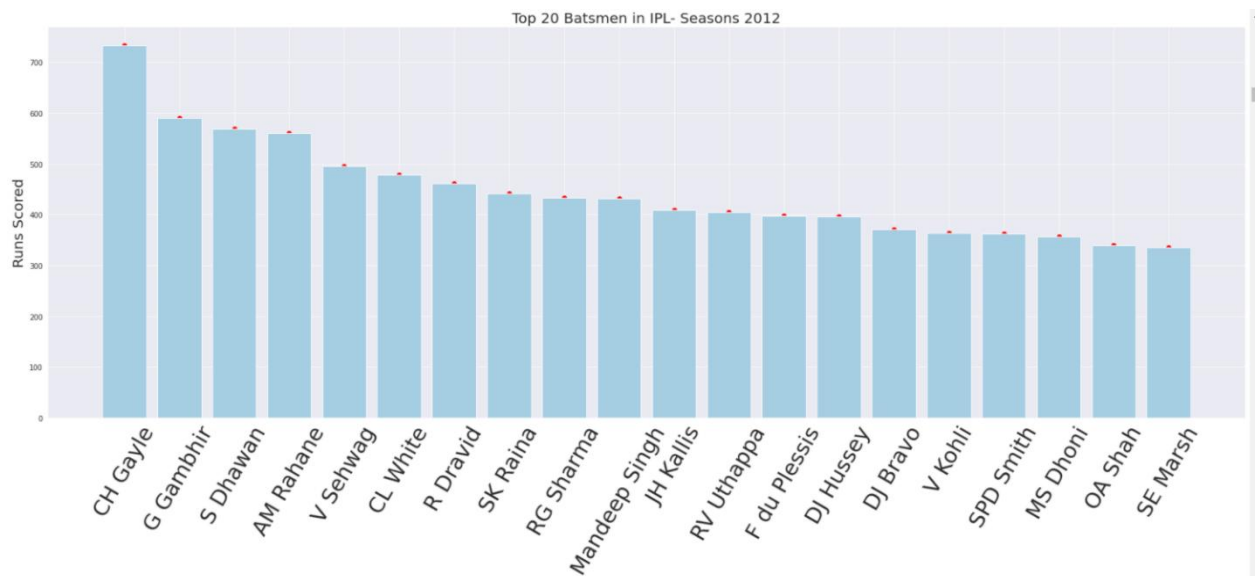
Here Top 20 batsmen in IPL Season 2012 are mentioned.

x-axis represents the name of the Player and Y-axis represents the runs scored in IPL season 2012

IPL Data Analysis

The Top 10 Batsmen in 2012:

	batsman	Runs
36	CH Gayle	733
58	G Gambhir	590
137	S Dhawan	569
20	AM Rahane	560
162	V Sehwag	495
39	CL White	479
122	R Dravid	462
144	SK Raina	441
127	RG Sharma	433
105	Mandeep Singh	432
76	JH Kallis	409
133	RV Uthappa	405
57	F du Plessis	398
49	DJ Hussey	396
47	DJ Bravo	371
160	V Kohli	364
149	SPD Smith	362
103	MS Dhoni	357
109	OA Shah	340
143	SE Marsh	336



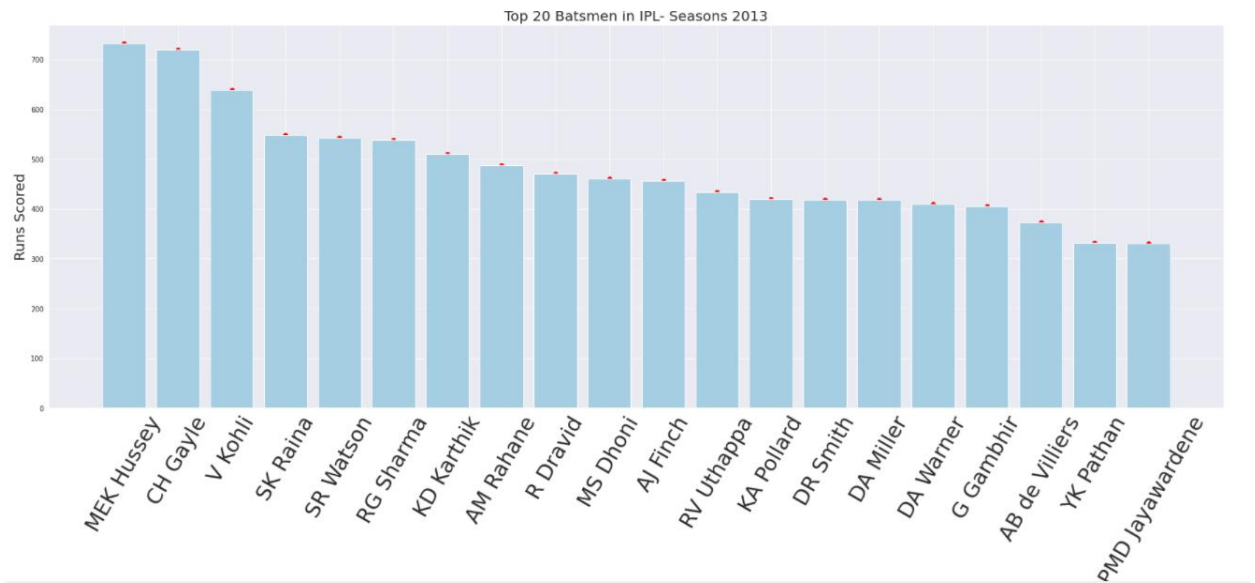
Here Top 20 batsmen in IPL Season 2013 are mentioned.

x-axis represents the name of the Player and Y-axis represents the runs scored in IPL season 2013

IPL Data Analysis

The Top 10 Batsmen in 2013:

	batsman	Runs
92	MEK Hussey	733
34	CH Gayle	720
165	V Kohli	639
146	SK Raina	548
153	SR Watson	543
132	RG Sharma	538
70	KD Karthik	510
18	AM Rahane	488
123	R Dravid	471
100	MS Dhoni	461
15	AJ Finch	456
137	RV Uthappa	434
67	KA Pollard	420
47	DR Smith	418
38	DA Miller	418
39	DA Warner	410
51	G Gambhir	406
9	AB de Villiers	373
170	YK Pathan	332
46	DPMD Jayawardene	331



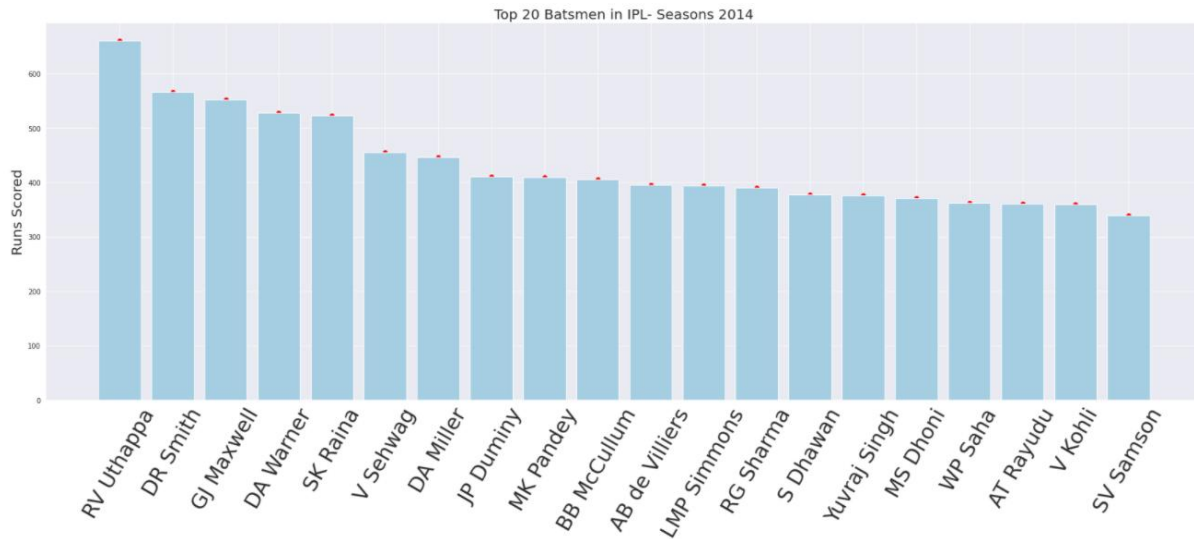
Here Top 20 batsmen in IPL Season 2014 are mentioned.

x-axis represents the name of the Player and Y-axis represents the runs scored in IPL season 2014

IPL Data Analysis

The Top 10 Batsmen in 2014:

	batsman	Runs
98	RV Uthappa	660
29	DR Smith	566
35	GJ Maxwell	552
24	DA Warner	528
107	SK Raina	523
121	V Sehwag	455
23	DA Miller	446
46	JP Duminy	410
72	MK Pandey	409
13	BB McCullum	405
2	AB de Villiers	395
59	LMP Simmons	394
95	RG Sharma	390
100	S Dhawan	377
131	Yuvraj Singh	376
76	MS Dhoni	371
126	WP Saha	362
10	AT Rayudu	361
120	V Kohli	359
113	SV Samson	339



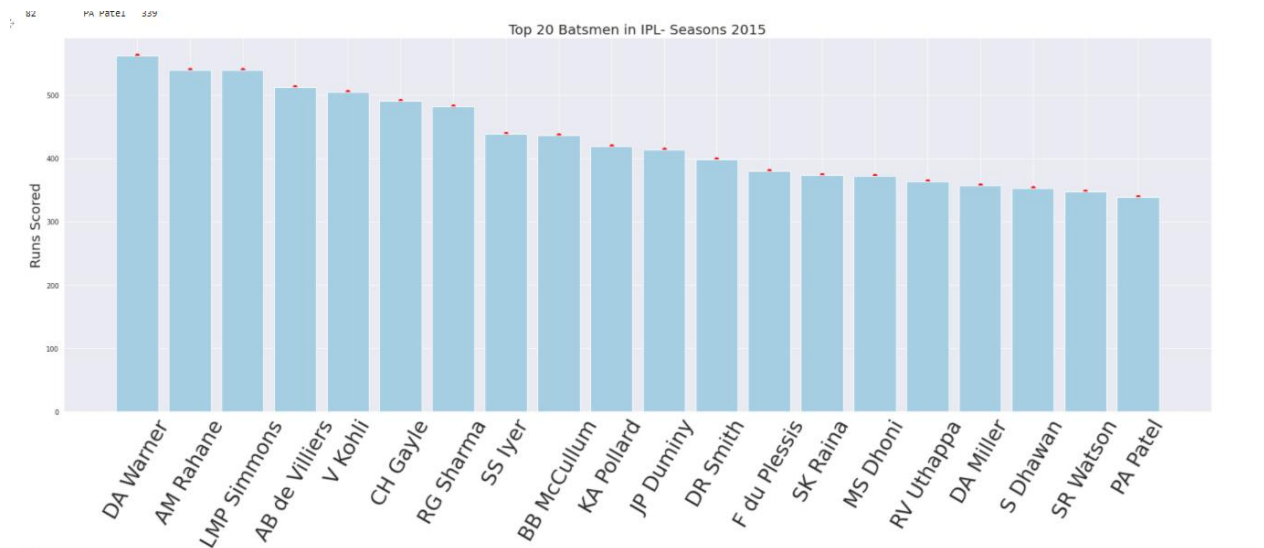
Here Top 20 batsmen in IPL Season 2015 are mentioned.

x-axis represents the name of the Player and Y-axis represents the runs scored in IPL season 2015

IPL Data Analysis

The Top 10 Batsmen in 2015:

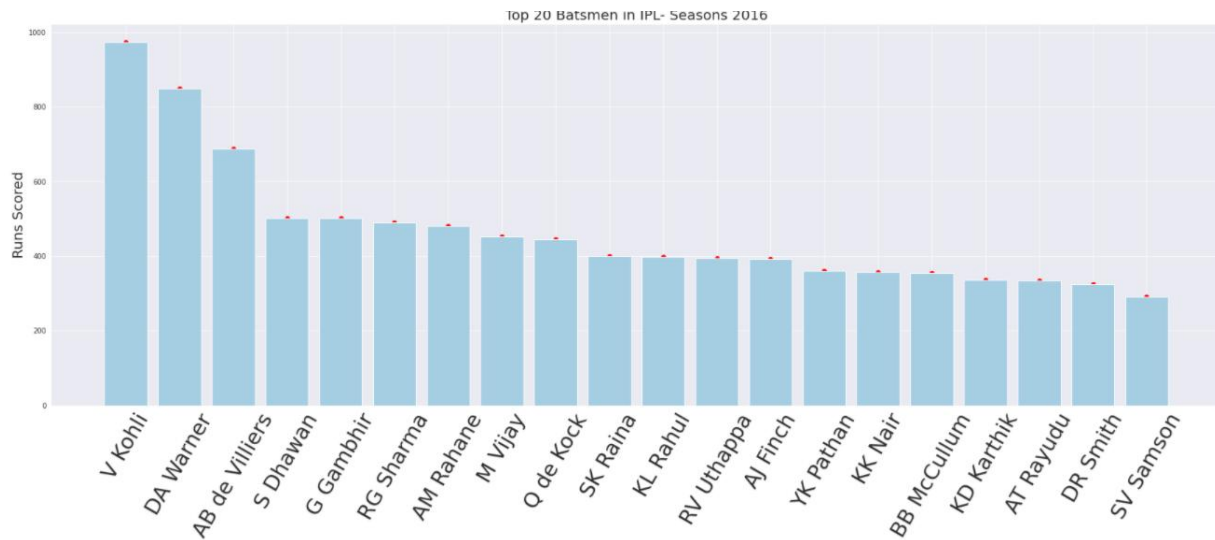
	batsman	Runs
25	DA Warner	562
7	AM Rahane	540
60	LMP Simmons	540
3	AB de Villiers	513
122	V Kohli	505
19	CH Gayle	491
94	RG Sharma	482
113	SS Iyer	439
16	BB McCullum	436
52	KA Pollard	419
50	JP Duminy	414
30	DR Smith	399
34	F du Plessis	380
106	SK Raina	374
74	MS Dhoni	372
98	RV Uthappa	364
24	DA Miller	357
100	S Dhawan	353
112	SR Watson	347
82	PA Patel	339



Here Top 20 batsmen in IPL Season 2016 are mentioned. x-axis represents the name of the Player and Y-axis represents the runs scored in IPL season 2016

IPL Data Analysis

The Top 10 Batsmen in 2016:		
	batsman	Runs
131	V Kohli	973
26	DA Warner	848
4	AB de Villiers	687
105	S Dhawan	501
36	G Gambhir	501
102	RG Sharma	489
8	AM Rahane	480
69	M Vijay	453
95	Q de Kock	445
112	SK Raina	399
60	KL Rahul	397
104	RV Uthappa	394
7	AJ Finch	393
133	YK Pathan	361
59	KK Nair	357
14	BB McCullum	354
56	KD Karthik	335
11	AT Rayudu	334
29	DR Smith	324
120	SV Samson	291

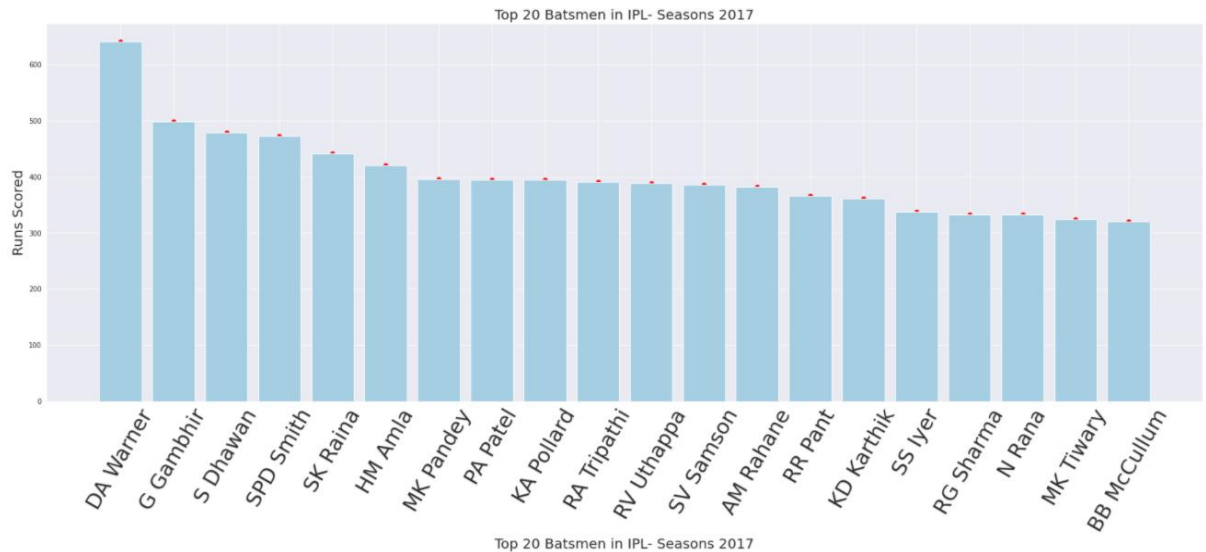


Here Top 20 batsmen in IPL Season 2017 are mentioned. x-axis represents the name of the Player and Y-axis represents the runs scored in IPL season 2017

IPL Data Analysis

The Top 10 Batsmen in 2017:

	batsman	Runs
34	DA Warner	641
43	G Gambhir	498
107	S Dhawan	479
117	SPD Smith	472
112	SK Raina	442
46	HM Amla	420
77	MK Pandey	396
92	PA Patel	395
60	KA Pollard	395
99	RA Tripathi	391
103	RV Uthappa	388
122	SV Samson	386
10	AM Rahane	382
102	RR Pant	366
62	KD Karthik	361
119	SS Iyer	338
101	RG Sharma	333
87	N Rana	333
78	MK Tiwary	324
21	BB McCullum	320

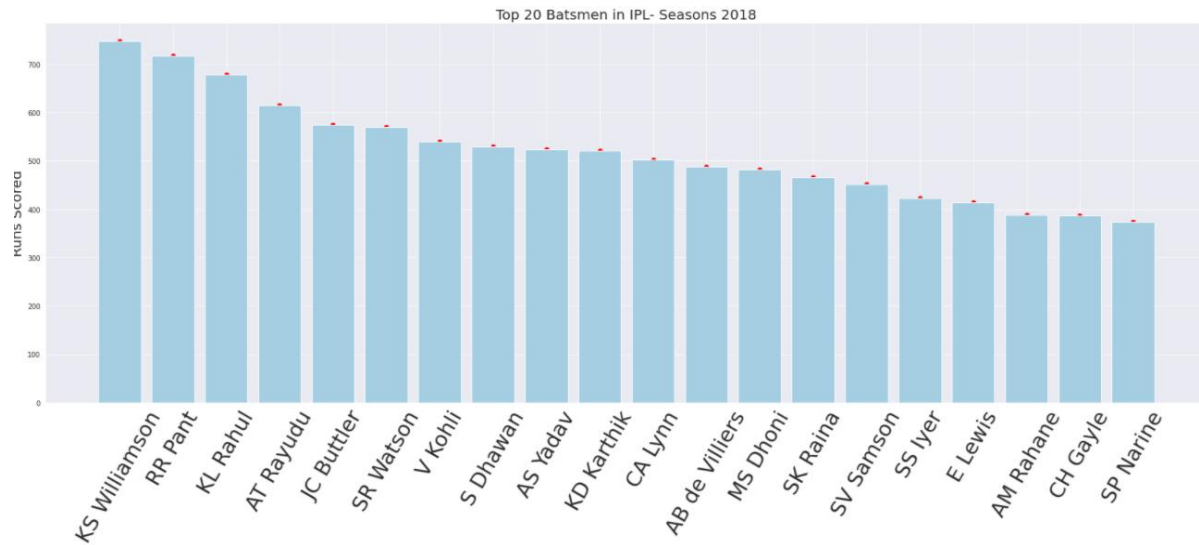


Here Top 20 batsmen in IPL Season 2018 are mentioned. x-axis represents the name of the Player and Y-axis represents the runs scored in IPL season 2018

IPL Data Analysis

The Top 10 Batsmen in 2018:

	batsman	Runs
64	KS Williamson	747
106	RR Pant	717
62	KL Rahul	678
12	AT Rayudu	615
52	JC Buttler	574
120	SR Watson	569
131	V Kohli	539
109	S Dhawan	529
11	AS Yadav	524
59	KD Karthik	520
25	CA Lynn	502
3	AB de Villiers	488
82	MS Dhoni	481
115	SK Raina	466
123	SV Samson	451
121	SS Iyer	423
39	E Lewis	414
8	AM Rahane	388
26	CH Gayle	386
119	SP Narine	373

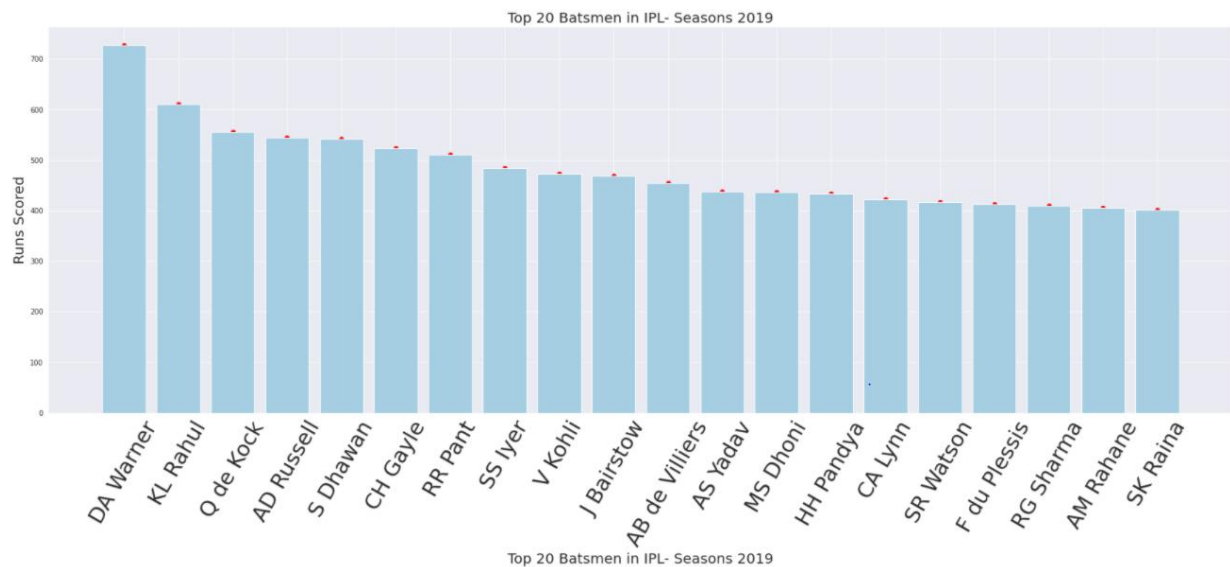


Here Top 20 batsmen in IPL Season 2019 are mentioned. x-axis represents the name of the Player and Y-axis represents the runs scored in IPL season 2019

IPL Data Analysis

The Top 10 Batsmen in 2019:

	batsman	Runs
27	DA Warner	727
62	KL Rahul	610
96	Q de Kock	555
5	AD Russell	545
111	S Dhawan	542
22	CH Gayle	523
107	RR Pant	511
128	SS Iyer	484
136	V Kohli	472
47	J Bairstow	468
3	AB de Villiers	454
9	AS Yadav	438
80	MS Dhoni	436
40	HH Pandya	433
21	CA Lynn	422
127	SR Watson	417
33	F du Plessis	412
106	RG Sharma	410
7	AM Rahane	405
122	SK Raina	401

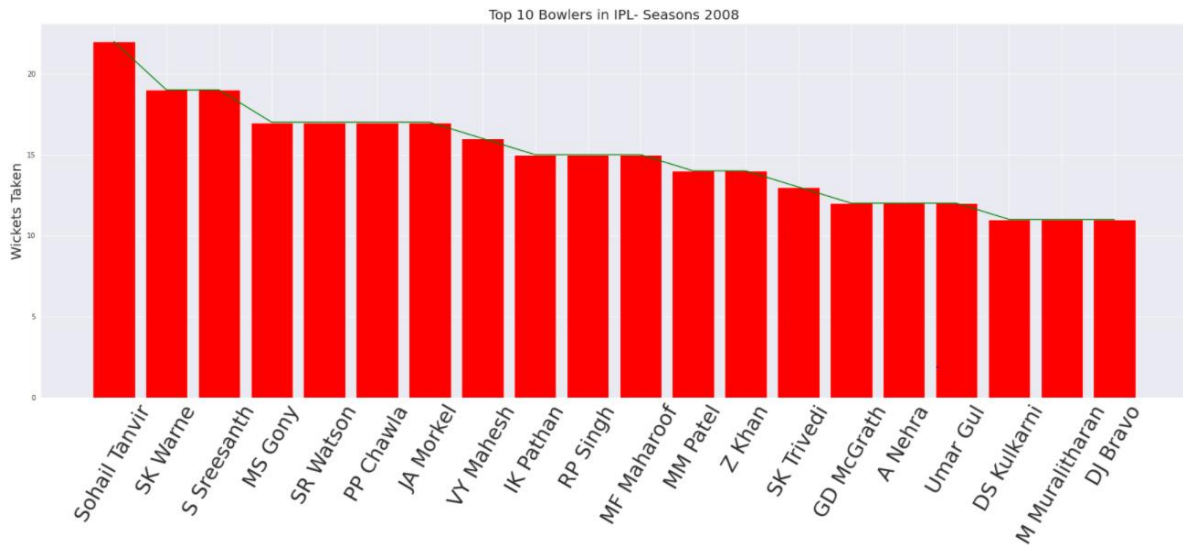


Here Top 20 Bowlers in IPL Season 2008 are mentioned. x-axis represents the name of the Player and Y-axis represents the wickets taken in IPL season 2008

IPL Data Analysis

The Top Wicket Takers in the Seasons 2008 are:

	bowler	Wickets
75	Sohail Tanvir	22
68	SK Warne	19
59	S Sreesanth	19
44	MS Gony	17
70	SR Watson	17
51	PP Chawla	17
31	JA Morkel	17
81	VY Mahesh	16
30	IK Pathan	15
56	RP Singh	15
42	MF Maharroof	15
43	MM Patel	14
87	Z Khan	14
67	SK Trivedi	13
26	GD McGrath	12
2	A Nehra	12
77	Umar Gul	12
24	DS Kulkarni	11
39	M Muralitharan	11
17	DJ Bravo	11

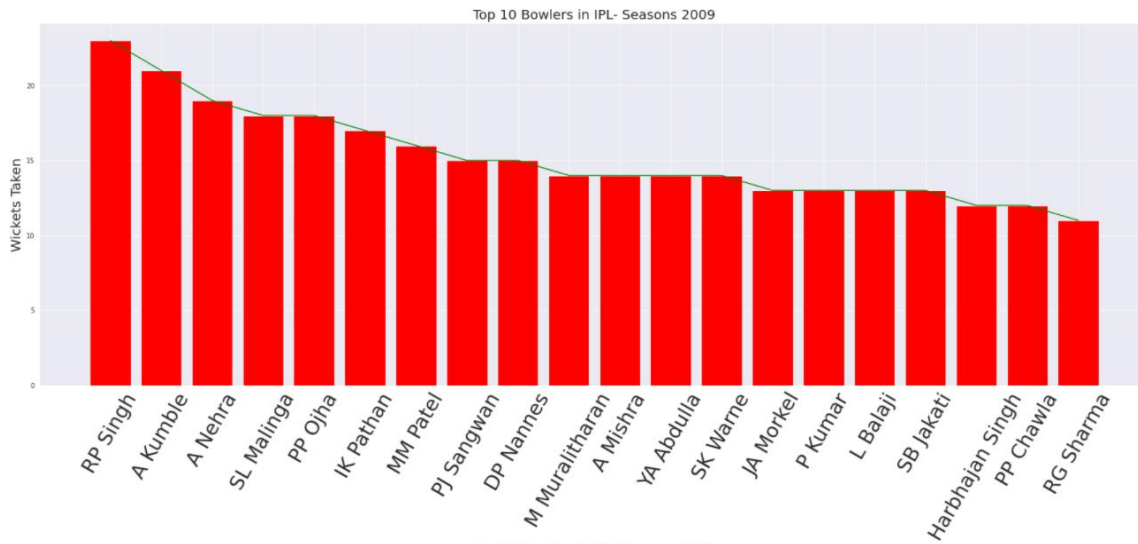


Here Top 20 Bowlers in IPL Season 2009 are mentioned. x-axis represents the name of the Player and Y-axis represents the wickets taken in IPL season 2009

IPL Data Analysis

The Top Wicket Takers in the Seasons 2009 are:

	bowler	wickets
64	RP Singh	23
1	A Kumble	21
3	A Nehra	19
75	SL Malinga	18
55	PP Ojha	18
32	IK Pathan	17
50	MM Patel	16
53	PJ Sangwan	15
23	DP Nannes	15
47	M Muralitharan	14
2	A Mishra	14
90	YA Abdulla	14
74	SK Warne	14
34	JA Morkel	13
52	P Kumar	13
43	L Balaji	13
69	SB Jakati	13
29	Harbhajan Singh	12
54	PP Chawla	12
62	RG Sharma	11

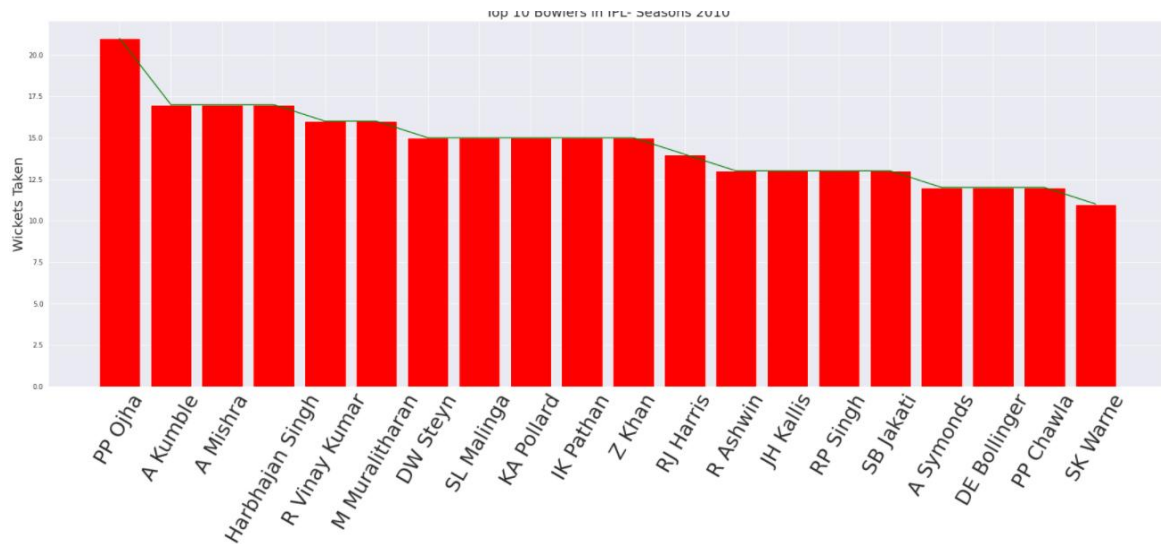


Here Top 20 Bowlers in IPL Season 2010 are mentioned. x-axis represents the name of the Player and Y-axis represents the wickets taken in IPL season 2010

IPL Data Analysis

The Top Wicket Takers in the Seasons 2010 are:

	bowler	wickets
59	PP Ojha	21
0	A Kumble	17
1	A Mishra	17
26	Harbhajan Singh	17
66	R Vinay Kumar	16
47	M Muralitharan	16
25	DW Steyn	15
84	SL Malinga	15
39	KA Pollard	15
29	IK Pathan	15
99	Z Khan	15
69	RJ Harris	14
61	R Ashwin	13
34	JH Kallis	13
70	RP Singh	13
77	SB Jakati	13
4	A Symonds	12
19	DE Bollinger	12
58	PP Chawla	12
83	SK Warne	11

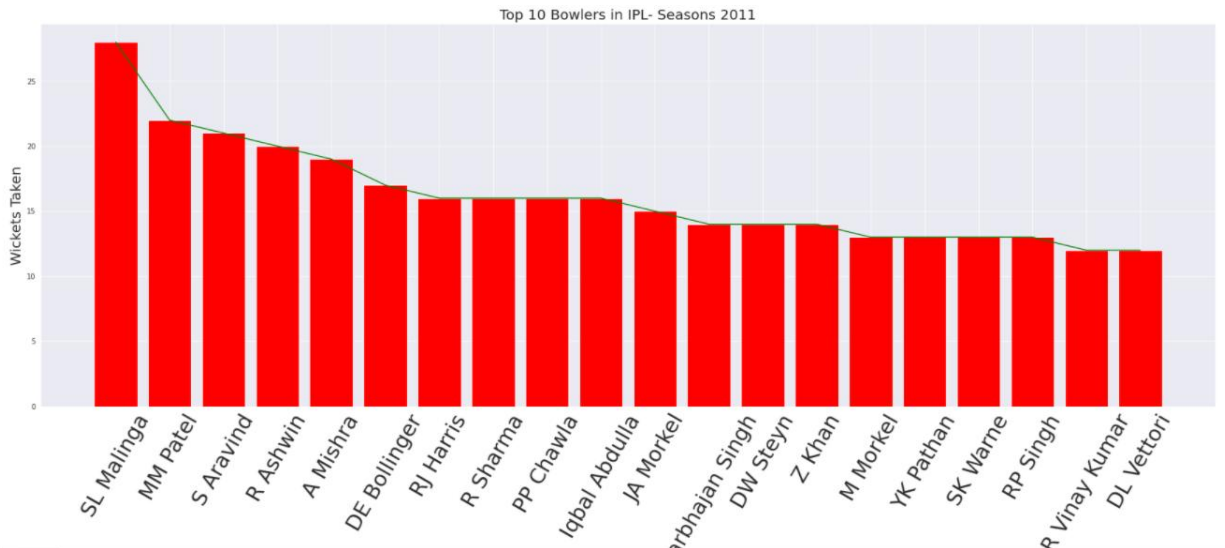


Here Top 20 Bowlers in IPL Season 2011 are mentioned. x-axis represents the name of the Player and Y-axis represents the wickets taken in IPL season 2011

IPL Data Analysis

The Top Wicket Takers in the Seasons 2011 are:

	bowler	Wickets
85	SL Malinga	28
52	MM Patel	22
74	S Aravind	21
62	R Ashwin	20
0	A Mishra	19
20	DE Bollinger	17
70	RJ Harris	16
66	R Sharma	16
60	PP Chawla	16
32	Iqbal Abdulla	16
35	JA Morkel	15
28	Harbhajan Singh	14
27	DW Steyn	14
99	Z Khan	14
50	M Morkel	13
97	YK Pathan	13
84	SK Warne	13
71	RP Singh	13
67	R Vinay Kumar	12
23	DL Vettori	12

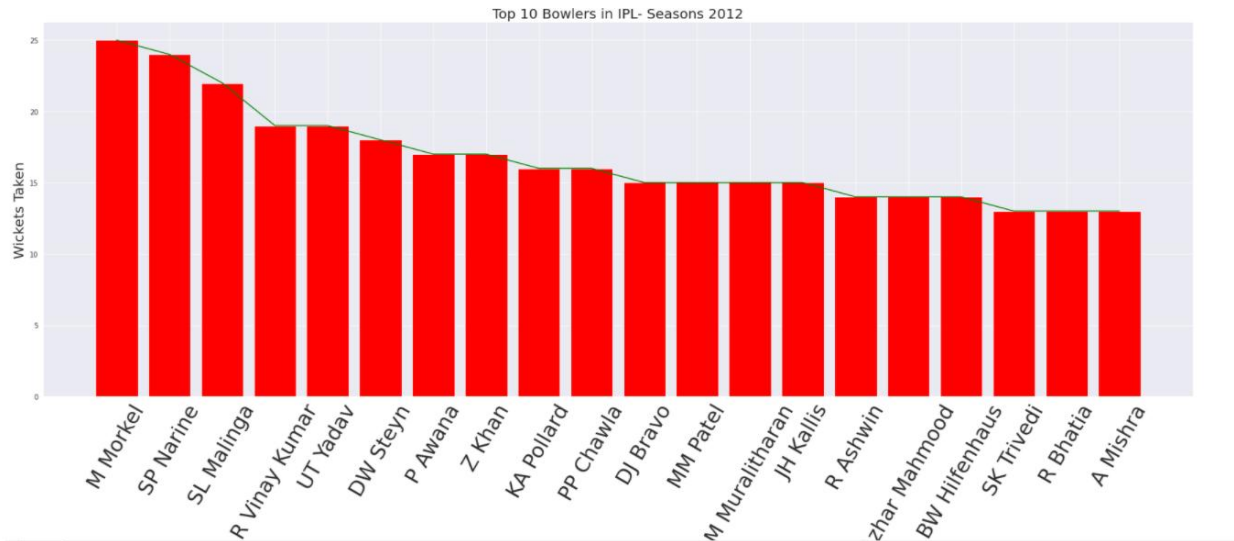


Here Top 20 Bowlers in IPL Season 2012 are mentioned. x-axis represents the name of the Player and Y-axis represents the wickets taken in IPL season 2012

IPL Data Analysis

The Top Wicket Takers in the Seasons 2012 are:

	bowler	wickets
55	M Morkel	25
87	SP Narine	24
86	SL Malinga	22
73	R Vinay Kumar	19
93	UT Yadav	19
32	DW Steyn	18
62	P Awana	17
99	Z Khan	17
48	KA Pollard	16
67	PP Chawla	16
26	DJ Bravo	15
59	MM Patel	15
56	M Muralitharan	15
45	JH Kallis	15
70	R Ashwin	14
17	Azhar Mahmood	14
22	BW Hilfenhaus	14
85	SK Trivedi	13
71	R Bhatia	13
2	A Mishra	13

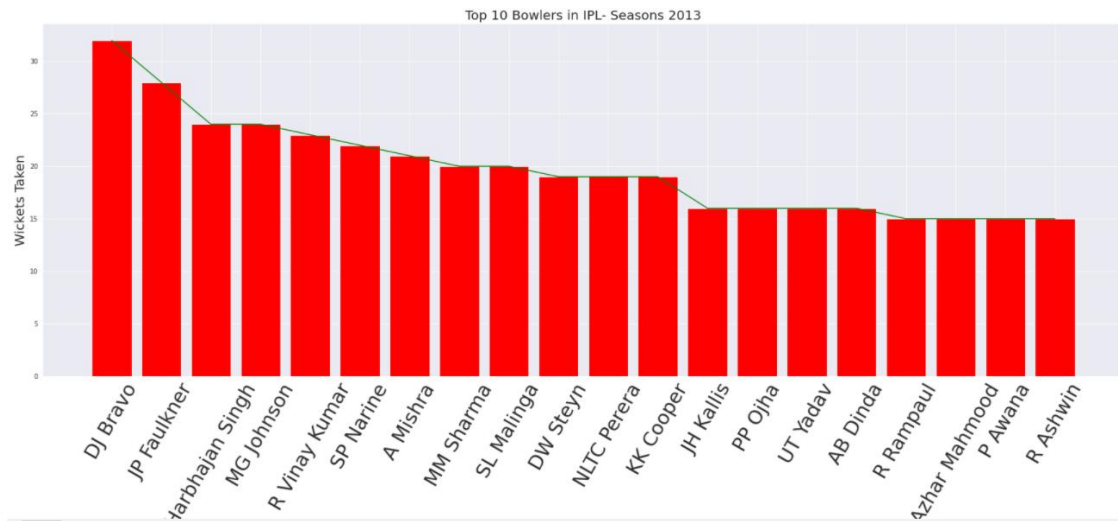


Here Top 20 Bowlers in IPL Season 2013 are mentioned. x-axis represents the name of the Player and Y-axis represents the wickets taken in IPL season 2013

IPL Data Analysis

The Top Wicket Takers in the Seasons 2013 are:

	bowler	Wickets
27	DJ Bravo	32
48	JP Faulkner	28
34	Harbhajan Singh	24
60	MG Johnson	24
83	R Vinay Kumar	23
96	SP Narine	22
2	A Mishra	21
62	MM Sharma	20
94	SL Malinga	20
32	DW Steyn	19
67	NLTC Perera	19
50	KK Cooper	19
45	JH Kallis	16
73	PP Ojha	16
102	UT Yadav	16
6	AB Dinda	16
80	R Rampaul	15
16	Azhar Mahmood	15
69	P Awana	15
76	R Ashwin	15

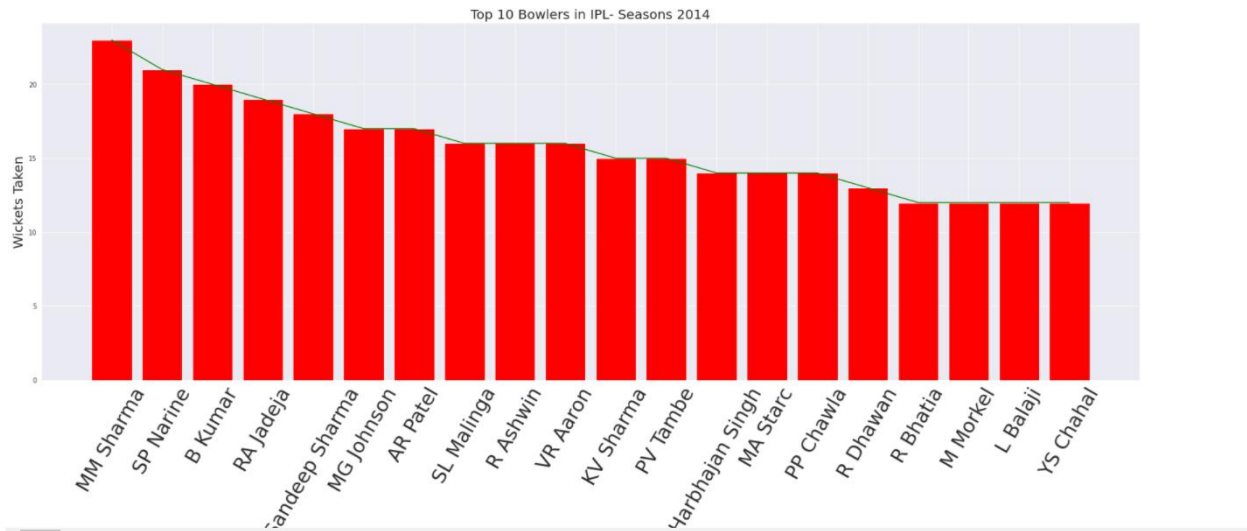


Here Top 20 Bowlers in IPL Season 2014 are mentioned. x-axis represents the name of the Player and Y-axis represents the wickets taken in IPL season 2014

IPL Data Analysis

The Top Wicket Takers in the Seasons 2014 are:

	bowler	Wickets
49	MM Sharma	23
77	SP Narine	21
7	B Kumar	20
68	RA Jadeja	19
80	Sandeep Sharma	18
48	MG Johnson	17
5	AR Patel	17
76	SL Malinga	16
60	R Ashwin	16
84	VR Aaron	16
36	KV Sharma	15
58	PV Tambe	15
19	Harbhajan Singh	14
46	MA Starc	14
56	PP Chawla	14
62	R Dhawan	13
61	R Bhatia	12
43	M Morkel	12
39	L Balaji	12
87	YS Chahal	12

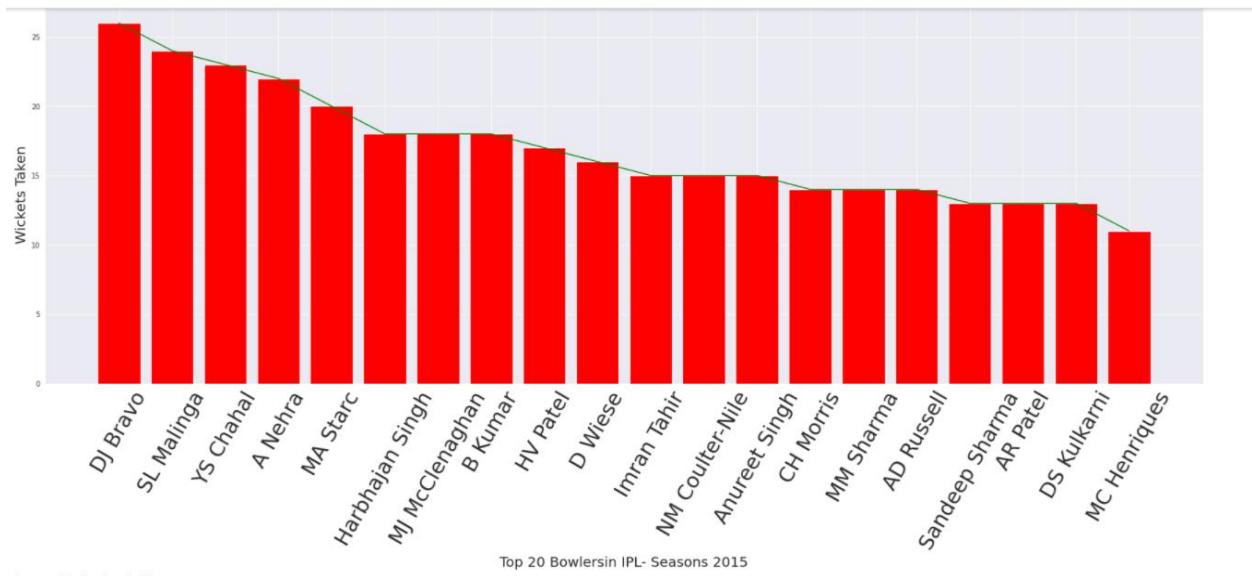


Here Top 20 Bowlers in IPL Season 2015 are mentioned. x-axis represents the name of the Player and Y-axis represents the wickets taken in IPL season 2015

IPL Data Analysis

The Top Wicket Takers in the Seasons 2015 are:

	bowler	Wickets
15	DJ Bravo	26
68	SL Malinga	24
81	YS Chahal	23
2	A Nehra	22
45	MA Starc	20
27	Harbhajan Singh	18
48	MJ McClenaghan	18
9	B Kumar	18
26	HV Patel	17
14	D Wiese	16
30	Imran Tahir	15
51	NM Coulter-Nile	15
8	Anureet Singh	15
12	CH Morris	14
49	MM Sharma	14
4	AD Russell	14
73	Sandeep Sharma	13
6	AR Patel	13
19	DS Kulkarni	13
46	MC Henriques	11

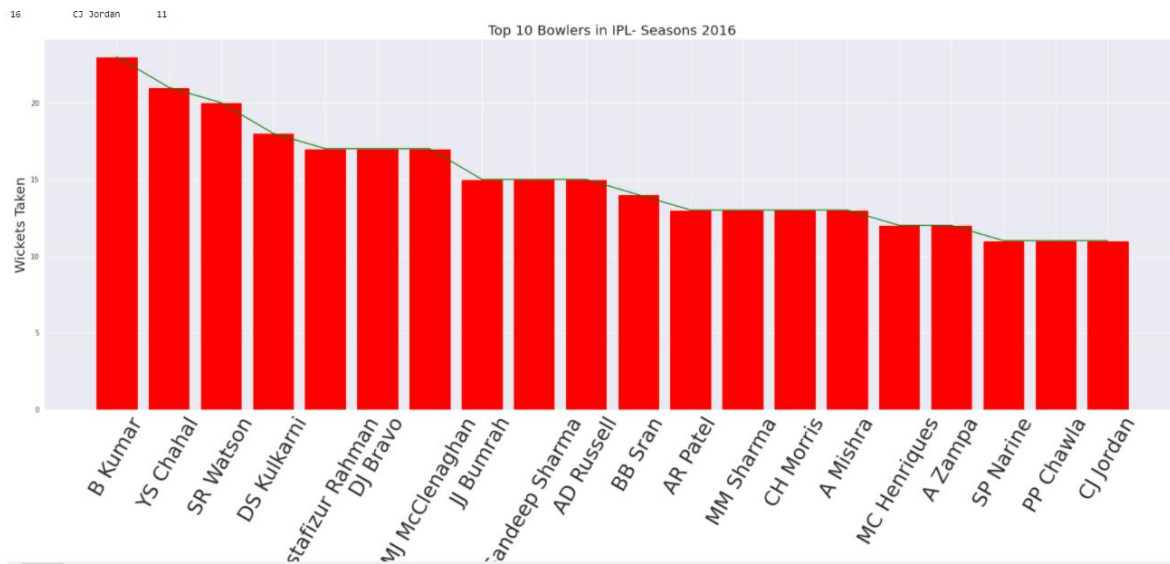


Here Top 20 Bowlers in IPL Season 2016 are mentioned. x-axis represents the name of the Player and Y-axis represents the wickets taken in IPL season 2016

IPL Data Analysis

The Top Wicket Takers in the Seasons 2016 are:

	bowler	wickets
10	B Kumar	23
84	YS Chahal	21
73	SR Watson	20
21	DS Kulkarni	18
51	Mustafizur Rahman	17
18	DJ Bravo	17
46	MJ McClenaghan	17
32	JJ Bumrah	15
76	Sandeep Sharma	15
5	AD Russell	15
11	BB Sran	14
7	AR Patel	13
47	MM Sharma	13
15	CH Morris	13
1	A Mishra	13
44	MC Henriques	12
3	A Zampa	12
72	SP Narine	11
57	PP Chawla	11
16	CJ Jordan	11

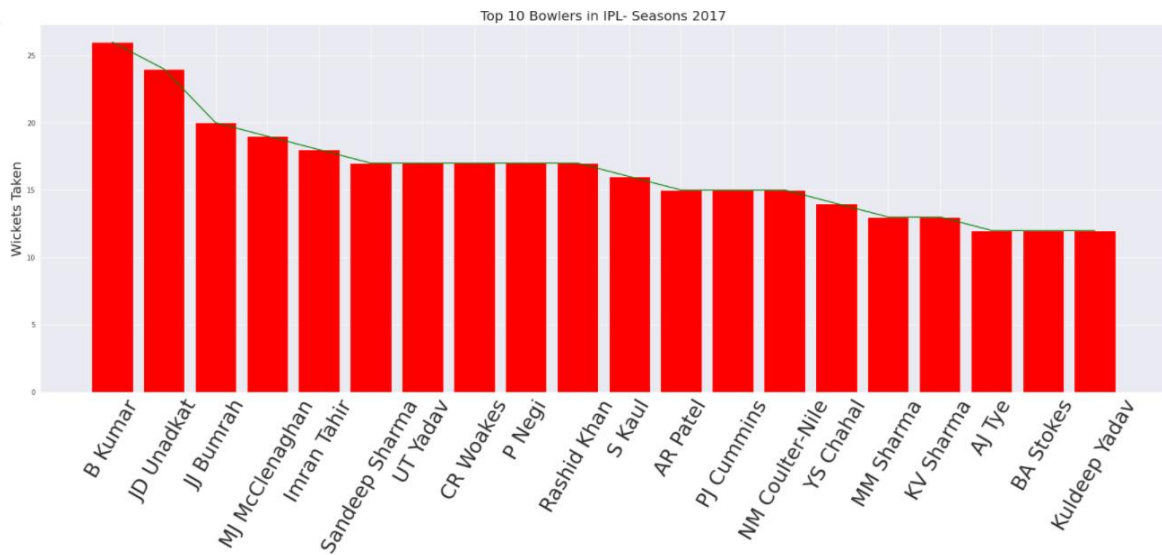


Here Top 20 Bowlers in IPL Season 2017 are mentioned. x-axis represents the name of the Player and Y-axis represents the wickets taken in IPL season 2017

IPL Data Analysis

The Top Wicket Takers in the Seasons 2017 are:

	bowler	Wickets
11	B Kumar	26
34	JD Unadkat	24
35	JJ Bumrah	20
46	MJ McClenaghan	19
32	Imran Tahir	18
77	Sandeep Sharma	17
84	UT Yadav	17
22	CR Woakes	17
56	P Negi	17
65	Rashid Khan	17
68	S Kaul	16
7	AR Patel	15
57	PJ Cummins	15
54	NM Coulter-Nile	15
88	YS Chahal	14
48	MM Sharma	13
40	KV Sharma	13
6	AJ Tye	12
13	BA Stokes	12
41	Kuldeep Yadav	12

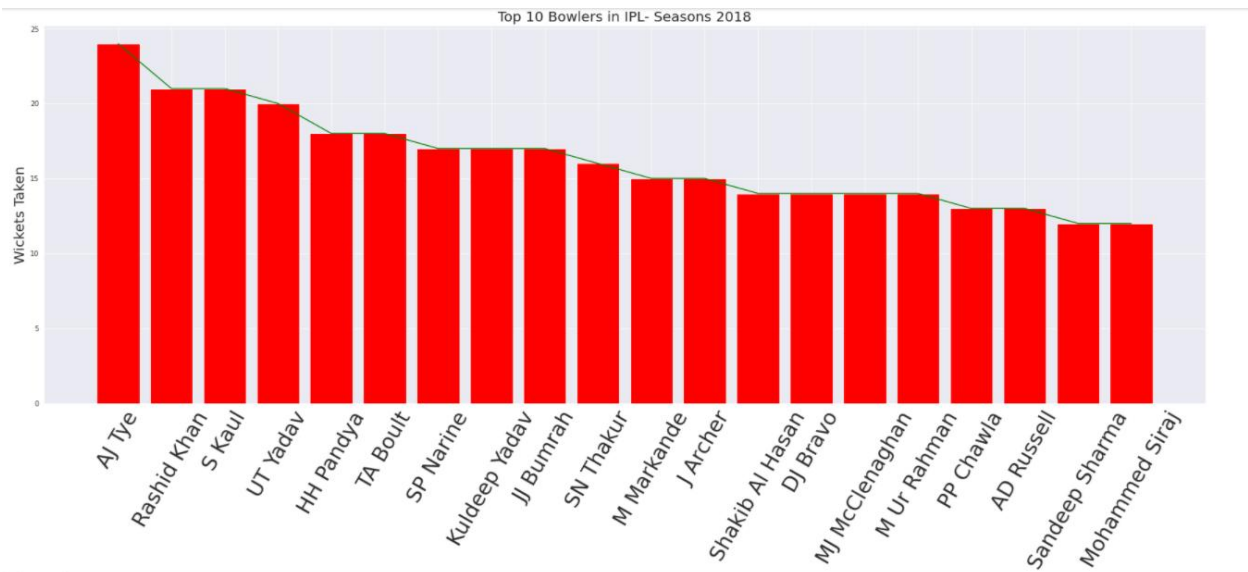


Here Top 20 Bowlers in IPL Season 2018 are mentioned. x-axis represents the name of the Player and Y-axis represents the wickets taken in IPL season 2018

IPL Data Analysis

The Top Wicket Takers in the Seasons 2018 are:

	bowler	Wickets
2	AJ Tye	24
63	Rashid Khan	21
65	S Kaul	21
77	UT Yadav	20
26	HH Pandya	18
75	TA Boult	18
70	SP Narine	17
40	Kuldeep Yadav	17
34	JJ Bumrah	17
69	SN Thakur	16
45	M Markande	15
31	J Archer	15
73	Shakib Al Hasan	14
21	DJ Bravo	14
48	MJ McClenaghan	14
46	M Ur Rahman	14
58	PP Chawla	13
1	AD Russell	13
72	Sandeep Sharma	12
53	Mohammed Siraj	12

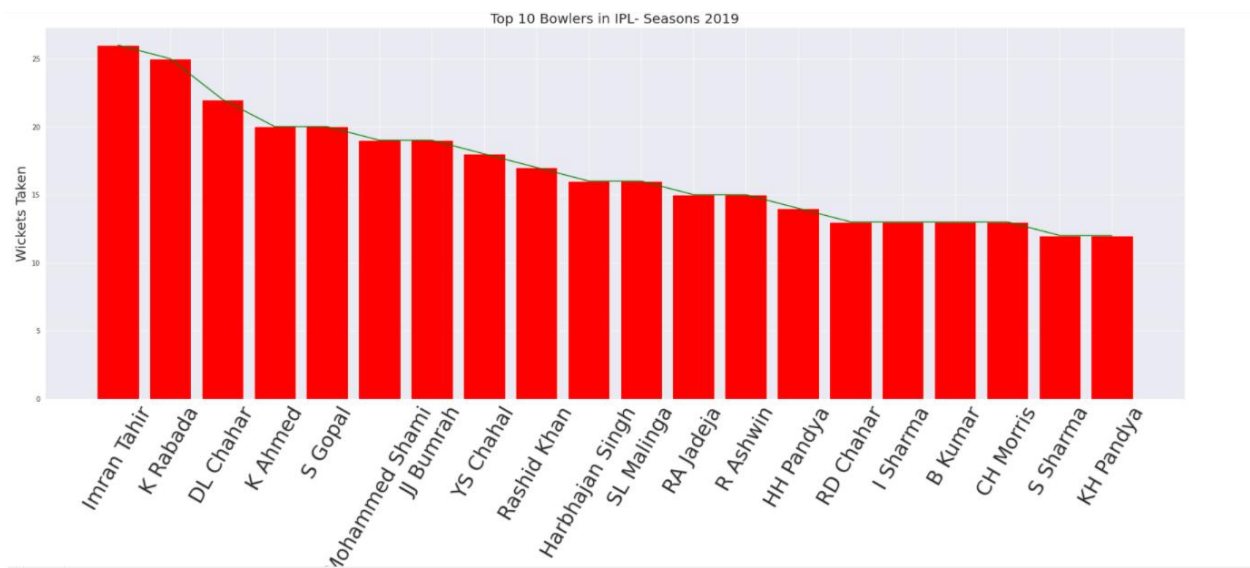


Here Top 20 Bowlers in IPL Season 2019 are mentioned. x-axis represents the name of the Player and Y-axis represents the wickets taken in IPL season 2019

IPL Data Analysis

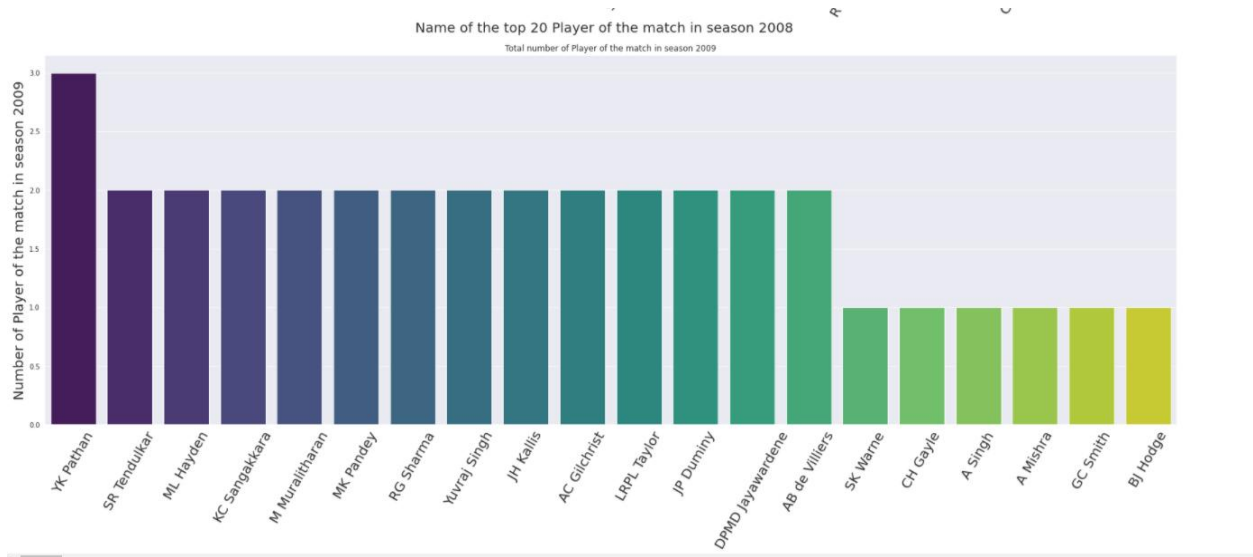
The Top Wicket Takers in the Seasons 2019 are:

	bowler	Wickets
24	Imran Tahir	26
35	K Rabada	25
14	DL Chahar	22
31	K Ahmed	20
65	S Gopal	20
49	Mohammed Shami	19
30	JJ Bumrah	19
85	YS Chahal	18
63	Rashid Khan	17
21	Harbhajan Singh	16
73	SL Malinga	16
61	RA Jadeja	15
58	R Ashwin	15
19	HH Pandya	14
62	RD Chahar	13
22	I Sharma	13
9	B Kumar	13
12	CH Morris	13
71	S Sharma	12
36	KH Pandya	12

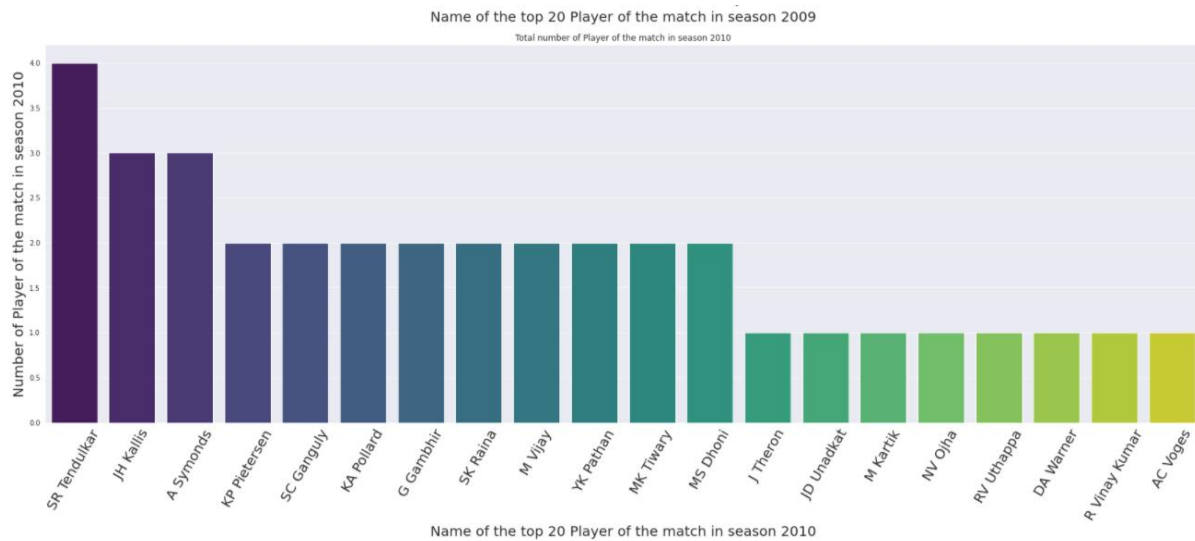


Here Top 20 Player of the match in IPL Season 2008 are mentioned. x-axis represents the name of the Player of the match and Y-axis represents the Number of matches in IPL season 2008

IPL Data Analysis

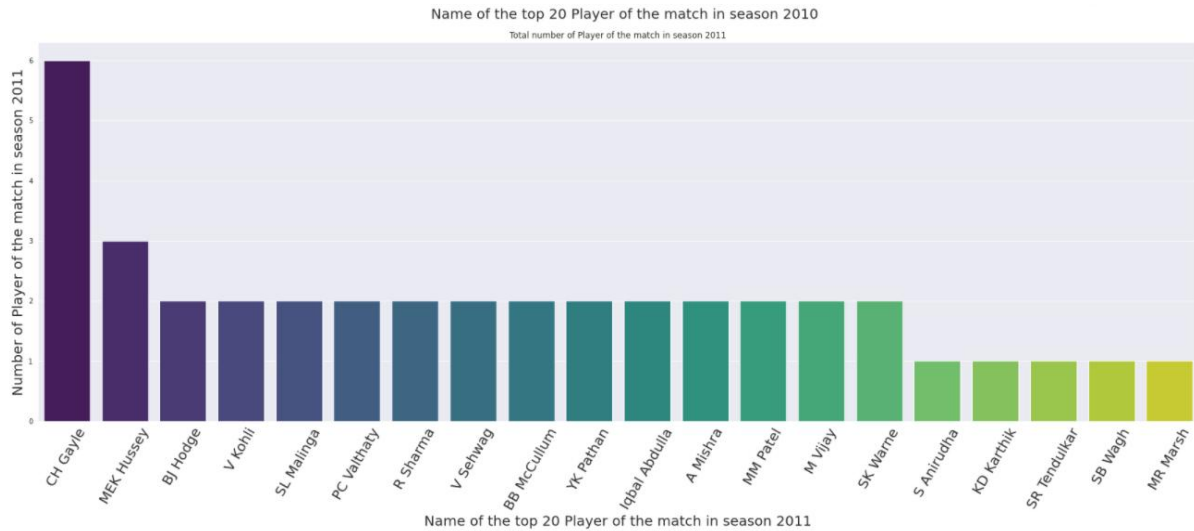


Here Top 20 Player of the match in IPL Season 2009 are mentioned. x-axis represents the name of the Player of the match and Y-axis represents the Number of matches in IPL season 2009

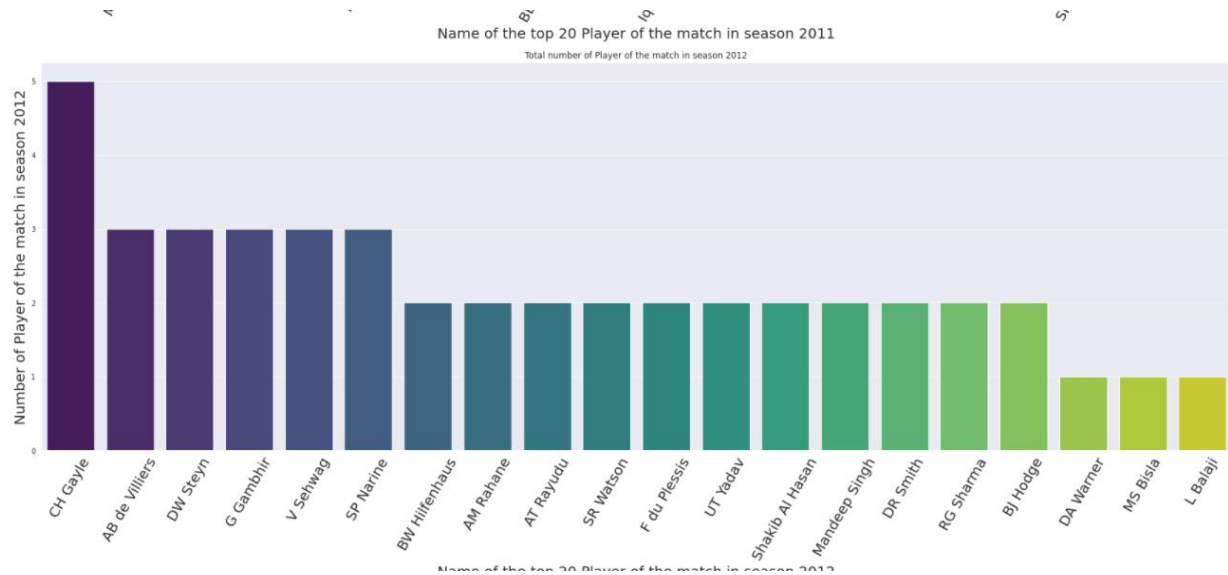


Here Top 20 Player of the match in IPL Season 2010 are mentioned. x-axis represents the name of the Player of the match and Y-axis represents the Number of matches in IPL season 2010

IPL Data Analysis

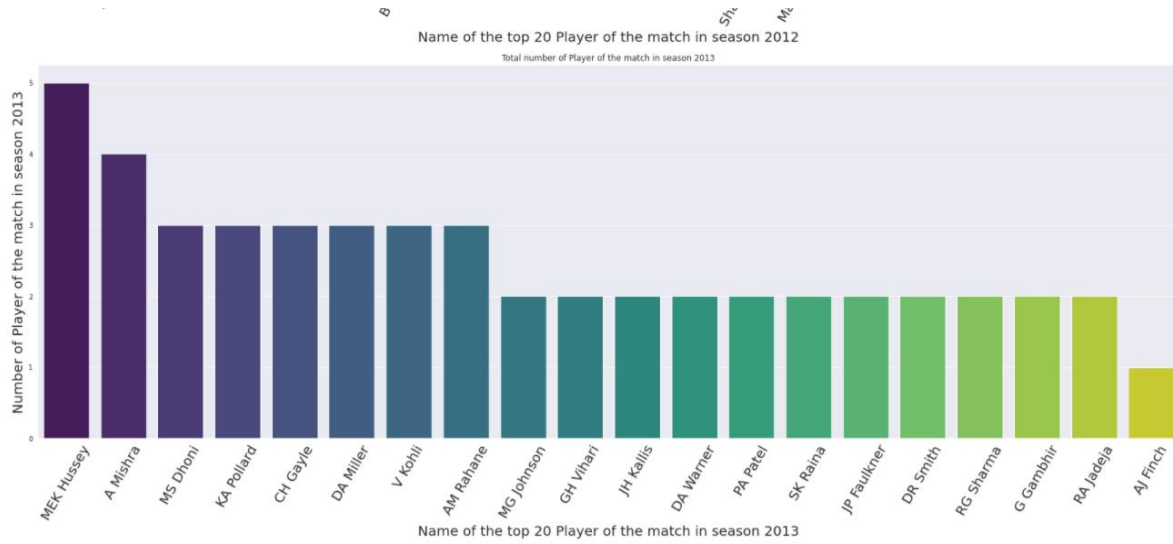


Here Top 20 Player of the match in IPL Season 2011 are mentioned. x-axis represents the name of the Player of the match and Y-axis represents the Number of matches in IPL season 2011

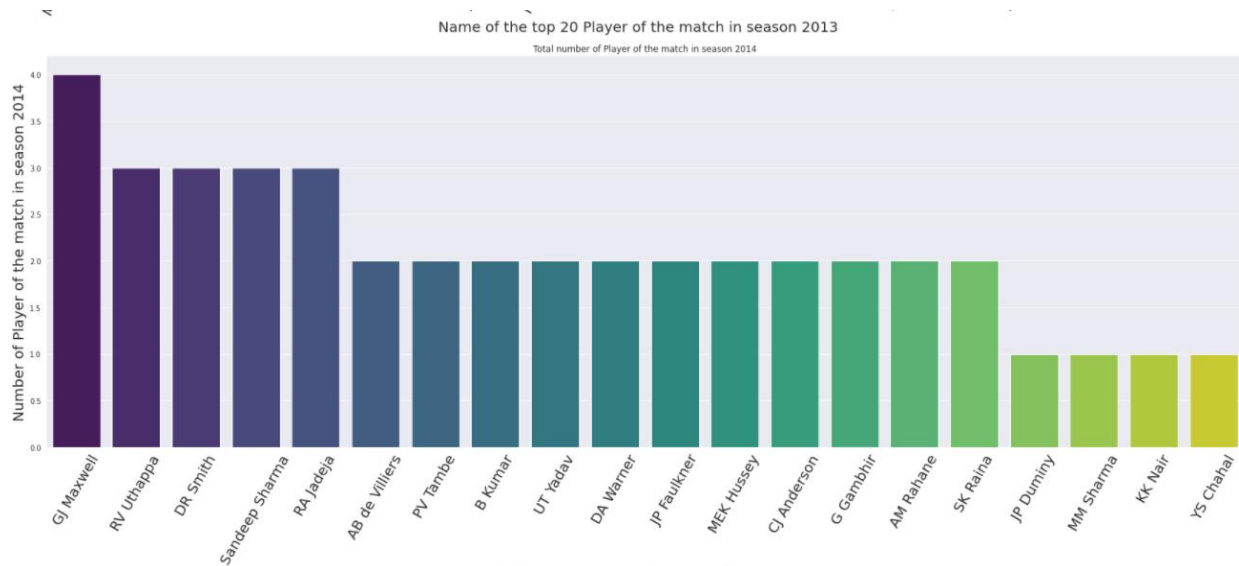


Here Top 20 Player of the match in IPL Season 2012 are mentioned. x-axis represents the name of the Player of the match and Y-axis represents the Number of matches in IPL season 2012

IPL Data Analysis

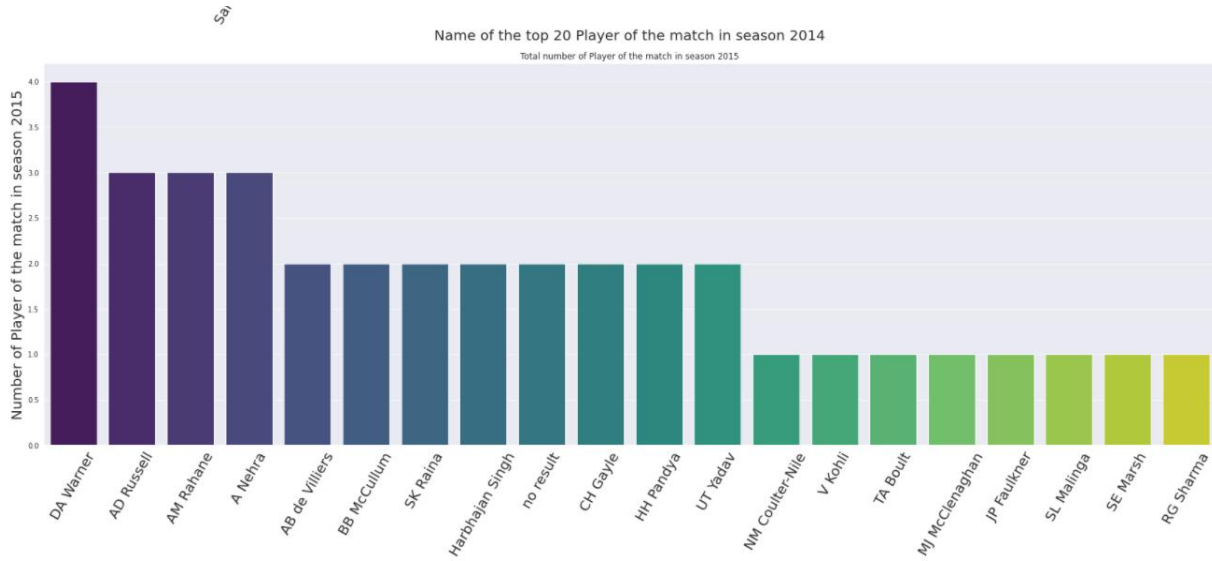


Here Top 20 Player of the match in IPL Season 2013 are mentioned. x-axis represents the name of the Player of the match and Y-axis represents the Number of matches in IPL season 2013

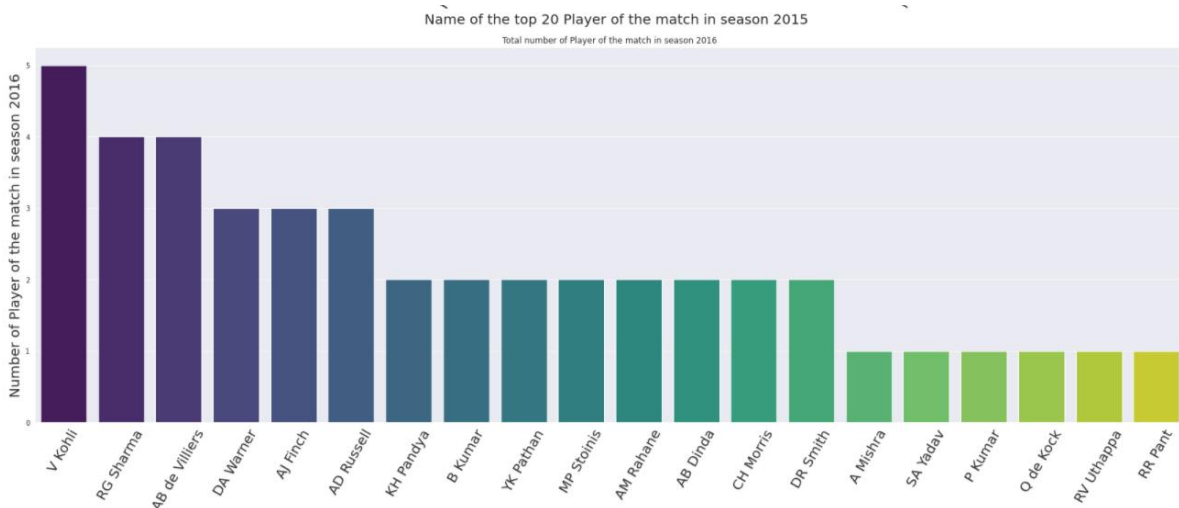


Here Top 20 Player of the match in IPL Season 2014 are mentioned. x-axis represents the name of the Player of the match and Y-axis represents the Number of matches in IPL season 2014

IPL Data Analysis

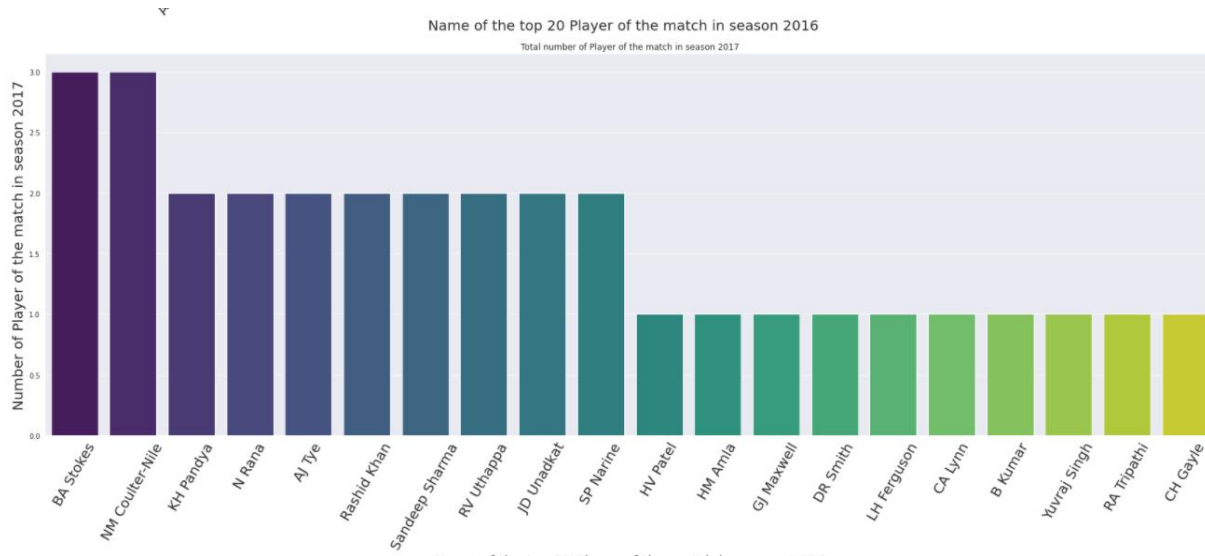


Here Top 20 Player of the match in IPL Season 2015 are mentioned. x-axis represents the name of the Player of the match and Y-axis represents the Number of matches in IPL season 2015

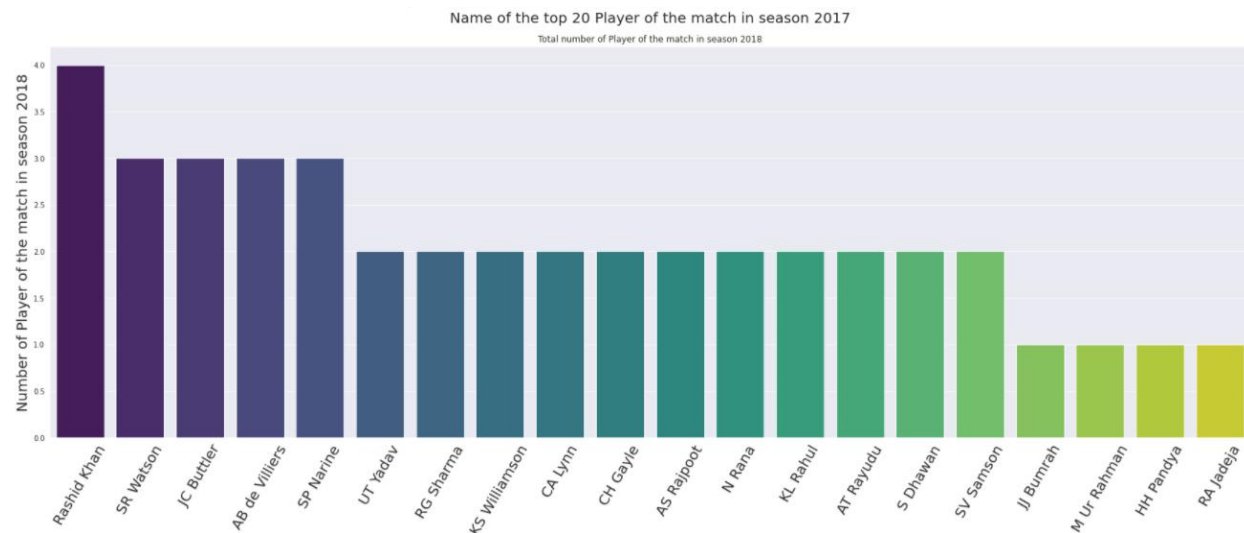


Here Top 20 Player of the match in IPL Season 2016 are mentioned. x-axis represents the name of the Player of the match and Y-axis represents the Number of matches in IPL season 2016

IPL Data Analysis

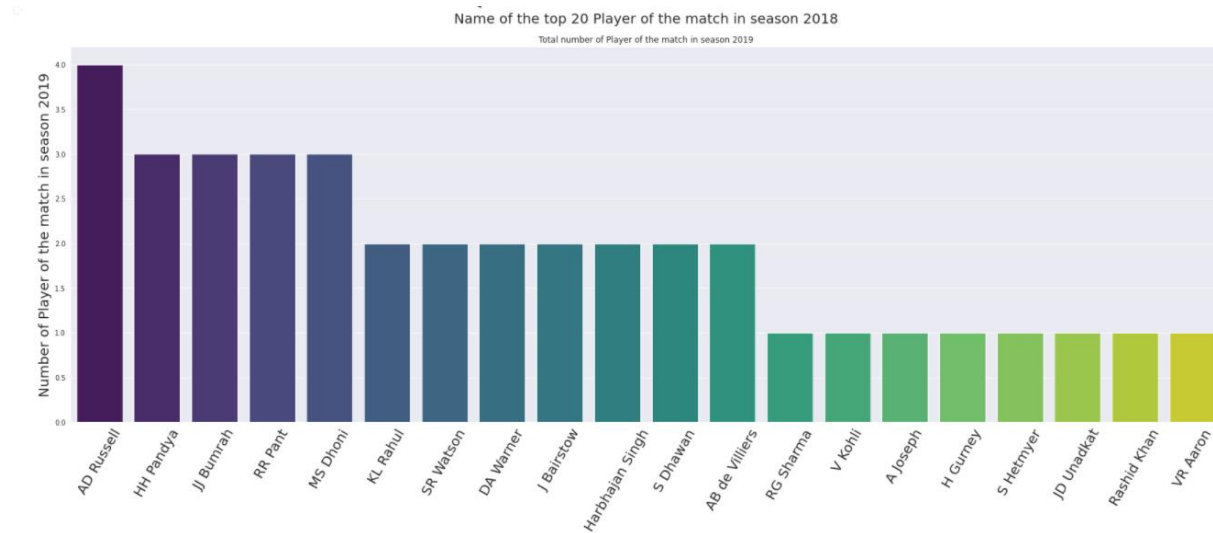


Here Top 20 Player of the match in IPL Season 2017 are mentioned. x-axis represents the name of the Player of the match and Y-axis represents the Number of matches in IPL season 2017

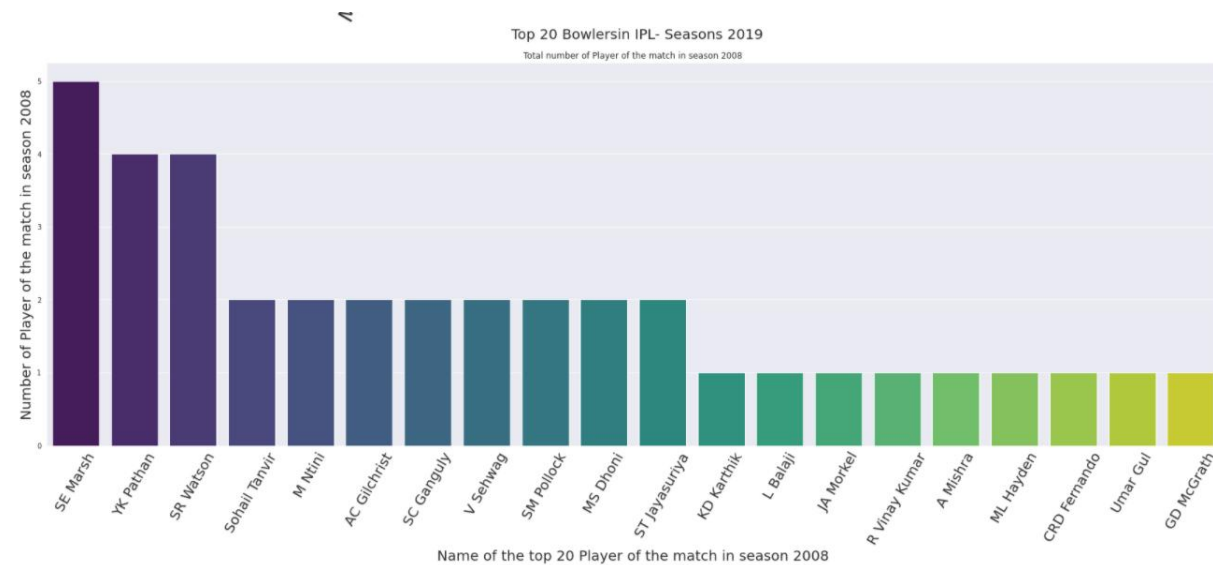


Here Top 20 Player of the match in IPL Season 2018 are mentioned. x-axis represents the name of the Player of the match and Y-axis represents the Number of matches in IPL season 2018

IPL Data Analysis



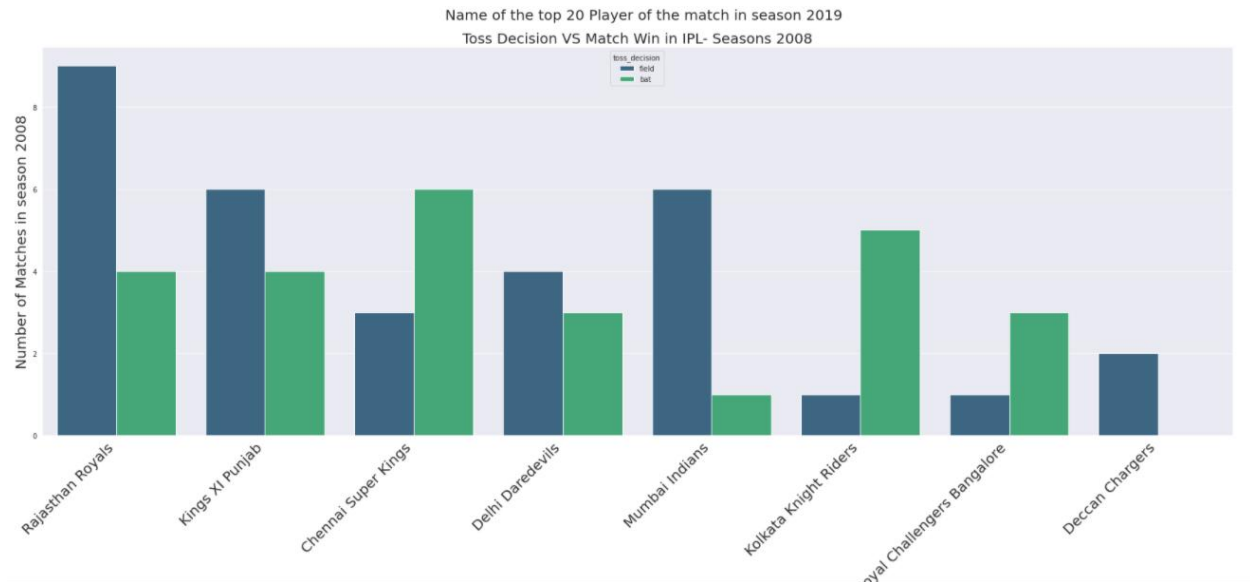
Here Top 20 Player of the match in IPL Season 2019 are mentioned. x-axis represents the name of the Player of the match and Y-axis represents the Number of matches in IPL season 2019



This plot indicates the number of wins when selected to bat first and number of wins when selected to field first in IPL season 2008.

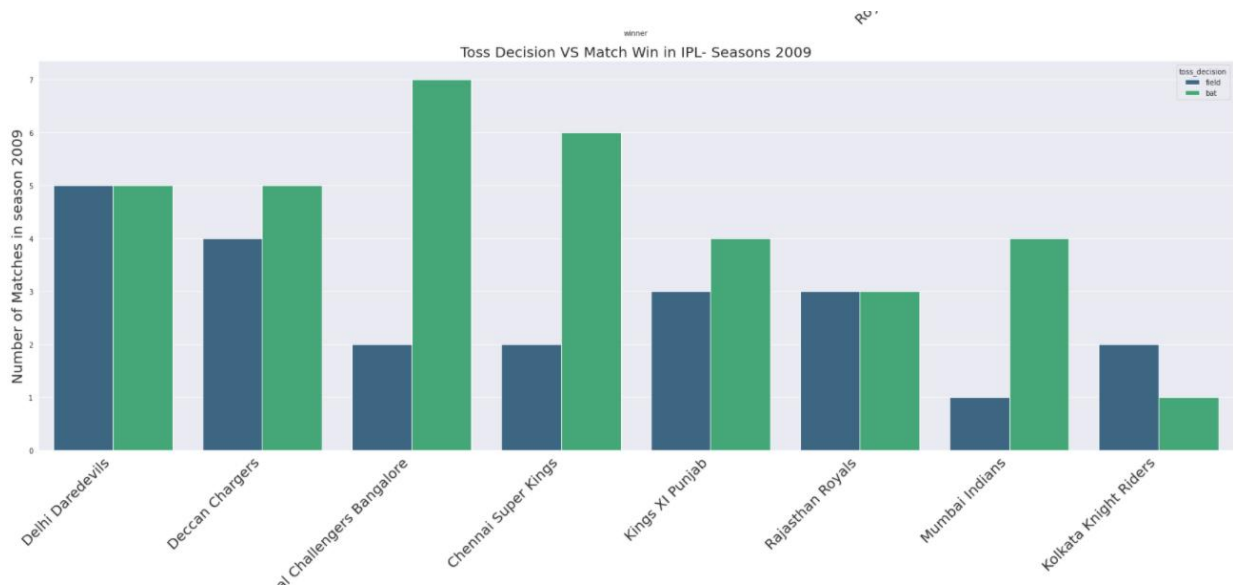
Here x-axis indicates the IPL teams and Y axis indicates the Number of matches won by IPL team in season 2008.

IPL Data Analysis



This plot indicates the number of wins when selected to bat first and number of wins when selected to field first in IPL season 2009.

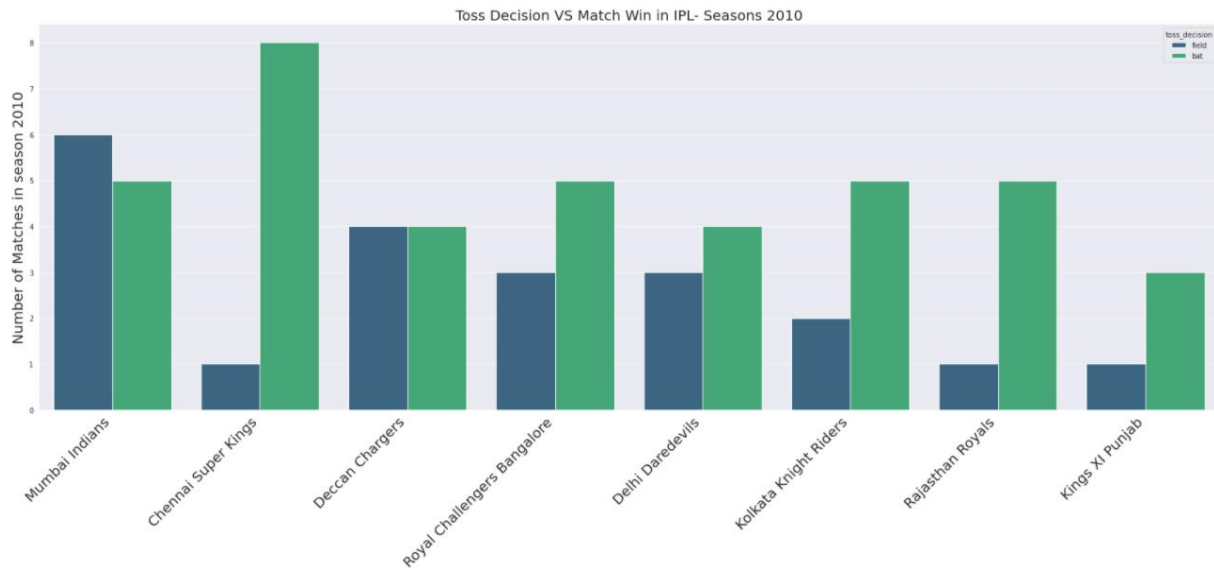
Here x-axis indicates the IPL teams and Y axis indicates the Number of matches won by IPL team in season 2009.



This plot indicates the number of wins when selected to bat first and number of wins when selected to field first in IPL season 2010.

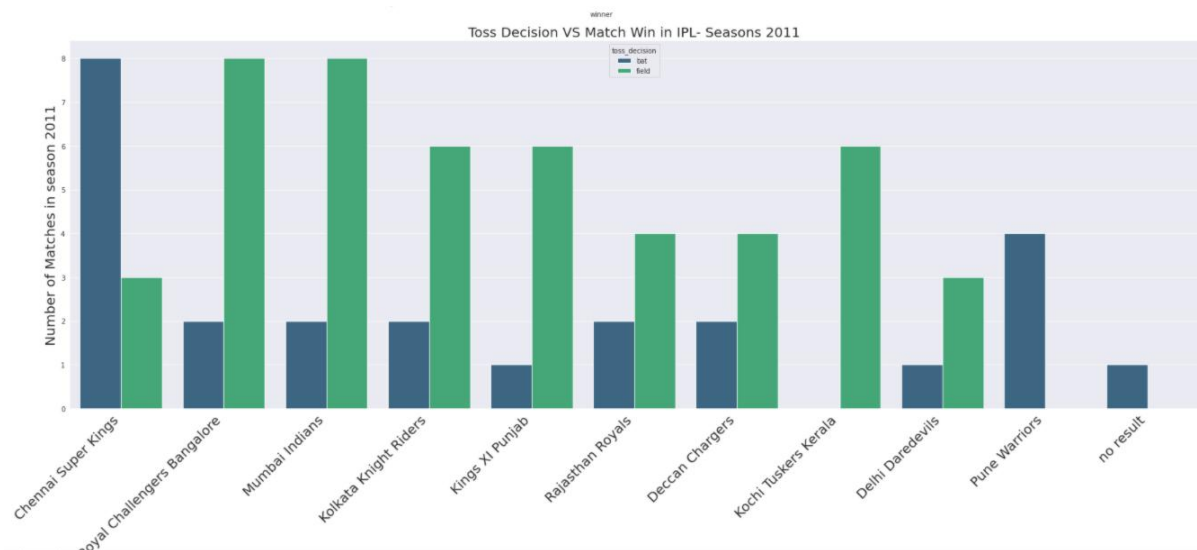
IPL Data Analysis

Here x-axis indicates the IPL teams and Y axis indicates the Number of matches won by IPL team in season 2010.



This plot indicates the number of wins when selected to bat first and number of wins when selected to field first in IPL season 2011.

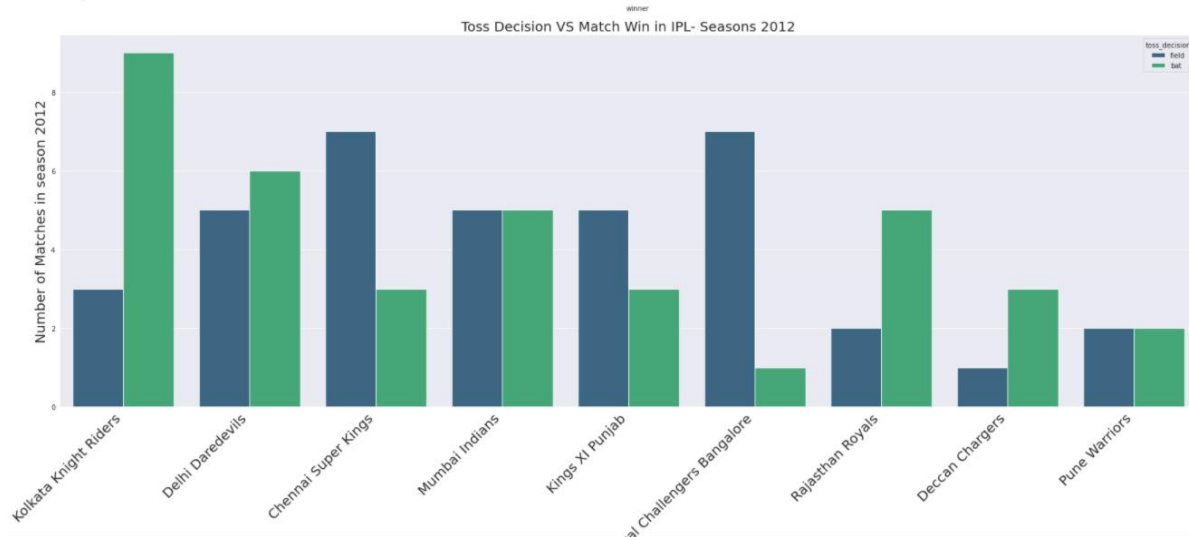
Here x-axis indicates the IPL teams and Y axis indicates the Number of matches won by IPL team in season 2011.



IPL Data Analysis

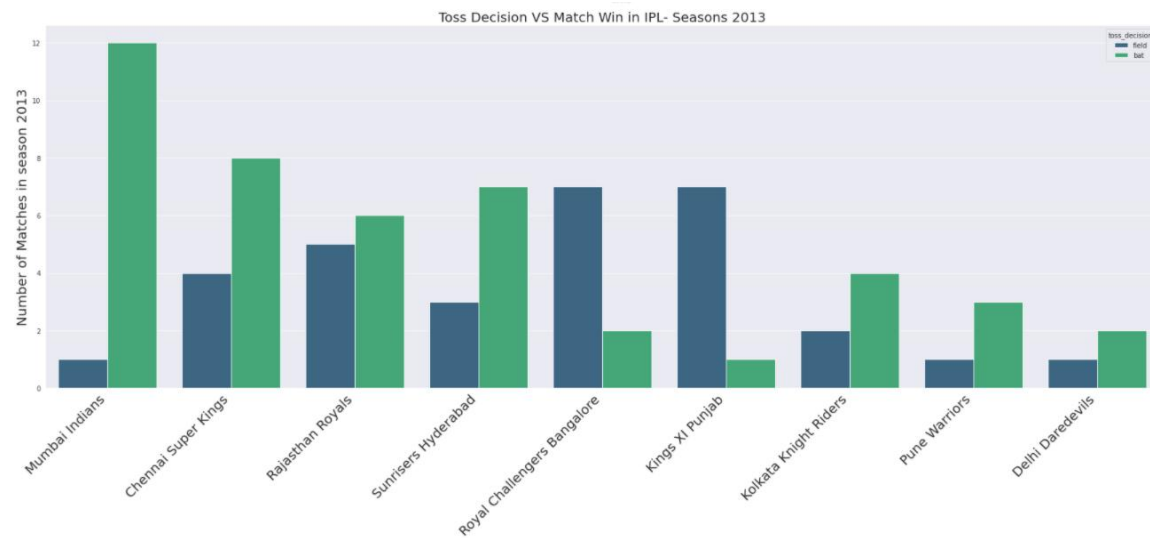
This plot indicates the number of wins when selected to bat first and number of wins when selected to field first in IPL season 2012.

Here x-axis indicates the IPL teams and Y axis indicates the Number of matches won by IPL team in season 2012.



This plot indicates the number of wins when selected to bat first and number of wins when selected to field first in IPL season 2013.

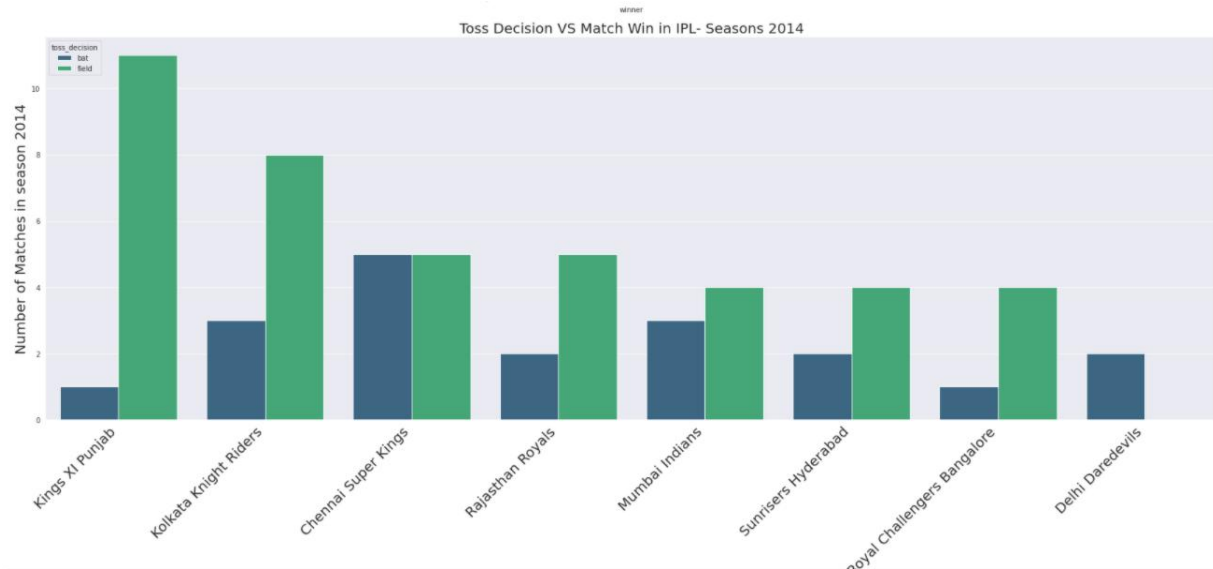
Here x-axis indicates the IPL teams and Y axis indicates the Number of matches won by IPL team in season 2013.



IPL Data Analysis

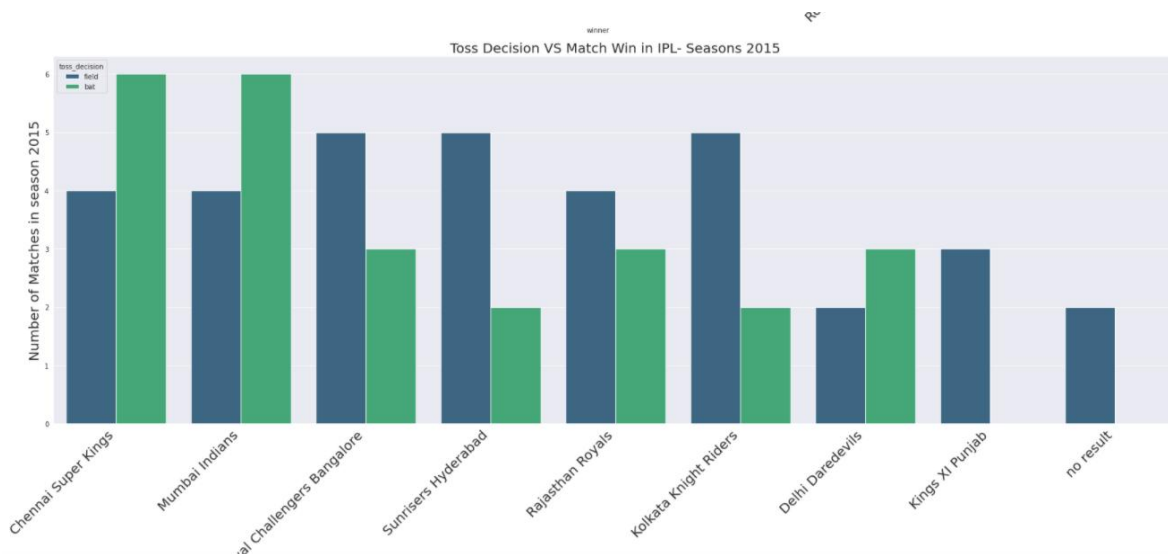
This plot indicates the number of wins when selected to bat first and number of wins when selected to field first in IPL season 2014.

Here x-axis indicates the IPL teams and Y axis indicates the Number of matches won by IPL team in season 2014.



This plot indicates the number of wins when selected to bat first and number of wins when selected to field first in IPL season 2015.

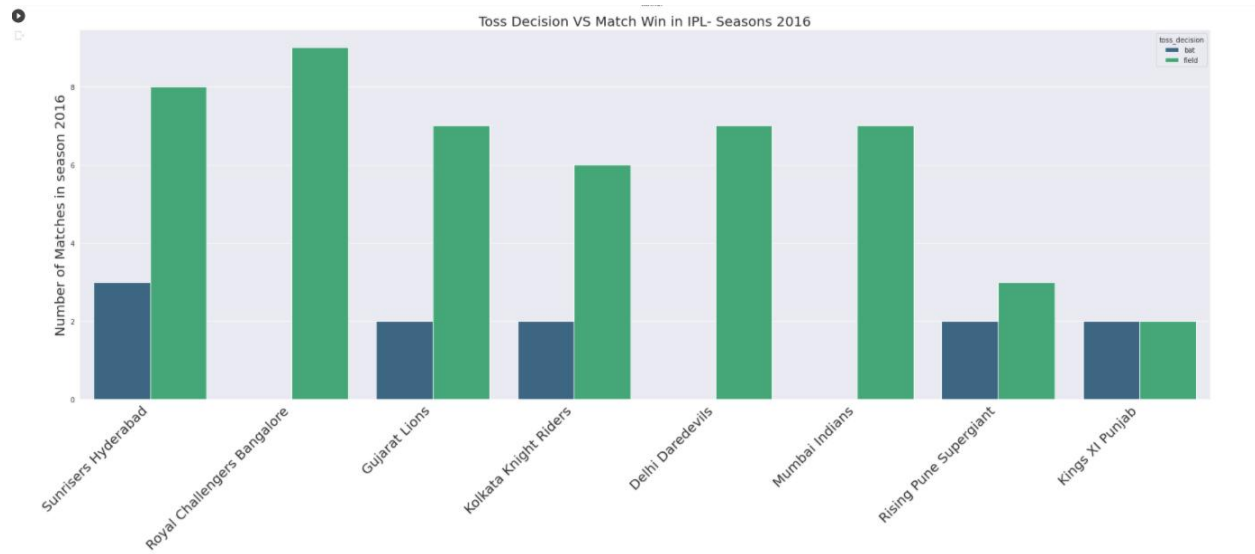
Here x-axis indicates the IPL teams and Y axis indicates the Number of matches won by IPL team in season 2015.



IPL Data Analysis

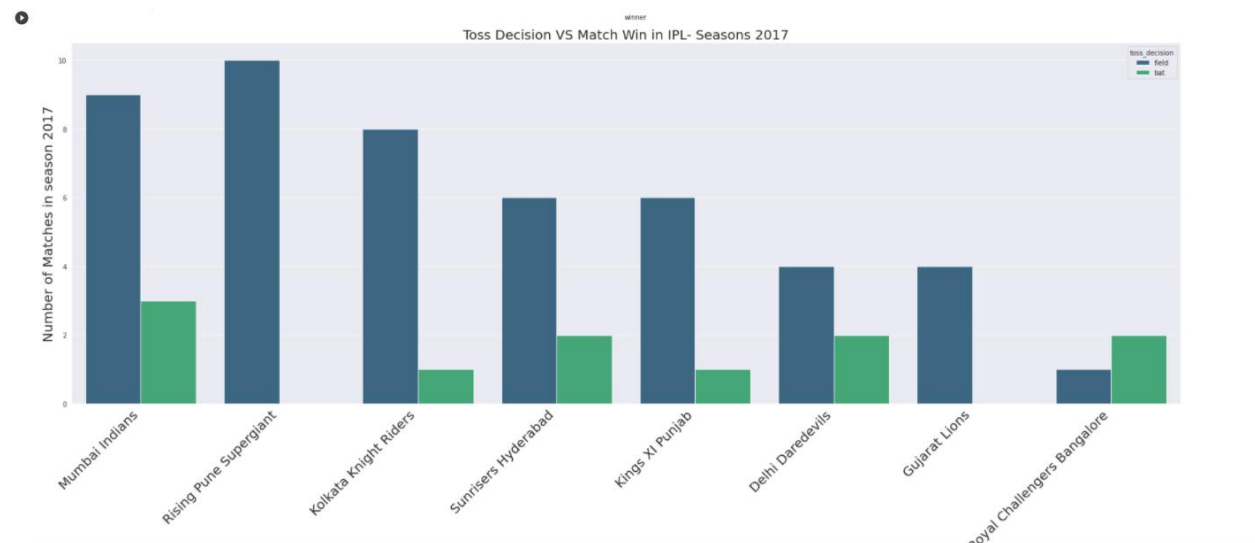
This plot indicates the number of wins when selected to bat first and number of wins when selected to field first in IPL season 2016.

Here x-axis indicates the IPL teams and Y axis indicates the Number of matches won by IPL team in season 2016.



This plot indicates the number of wins when selected to bat first and number of wins when selected to field first in IPL season 2017.

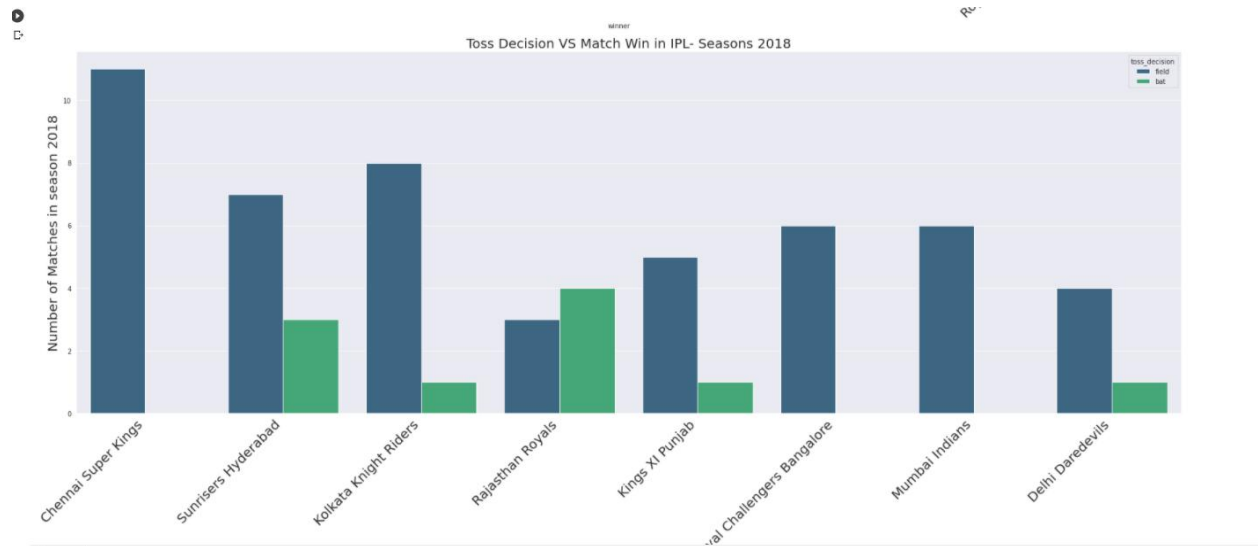
Here x-axis indicates the IPL teams and Y axis indicates the Number of matches won by IPL team in season 2017.



IPL Data Analysis

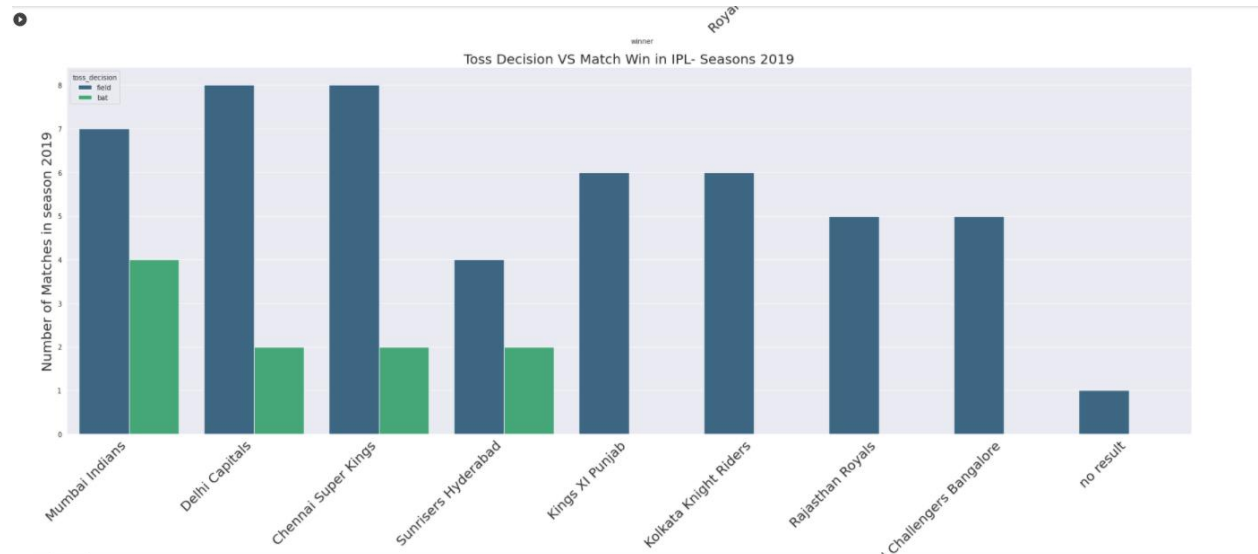
This plot indicates the number of wins when selected to bat first and number of wins when selected to field first in IPL season 2018.

Here x-axis indicates the IPL teams and Y axis indicates the Number of matches won by IPL team in season 2018.



This plot indicates the number of wins when selected to bat first and number of wins when selected to field first in IPL season 2019.

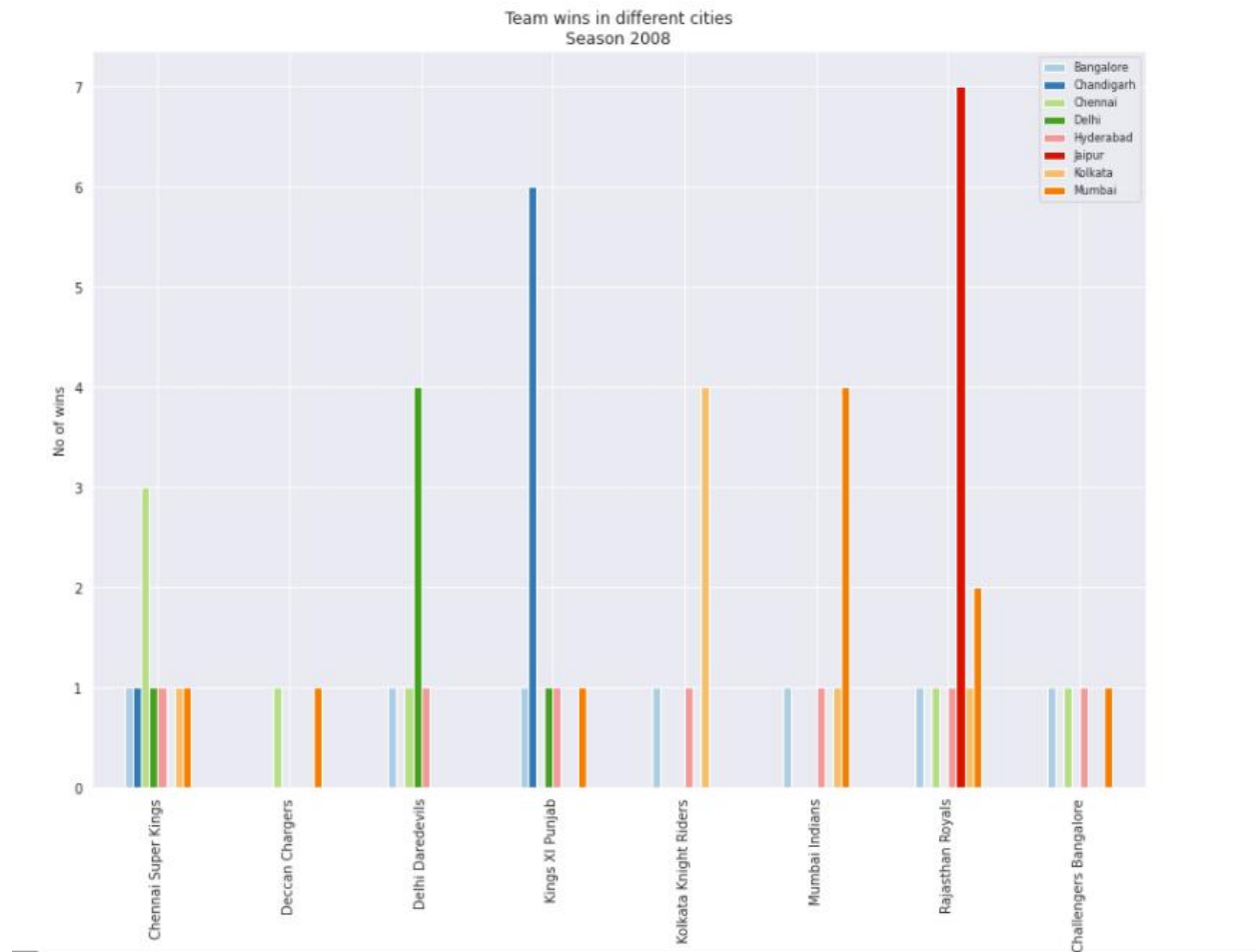
Here x-axis indicates the IPL teams and Y axis indicates the Number of matches won by IPL team in season 2019.



IPL Data Analysis

This plot indicates the number of wins for different IPL teams in different venues in IPL season 2008.

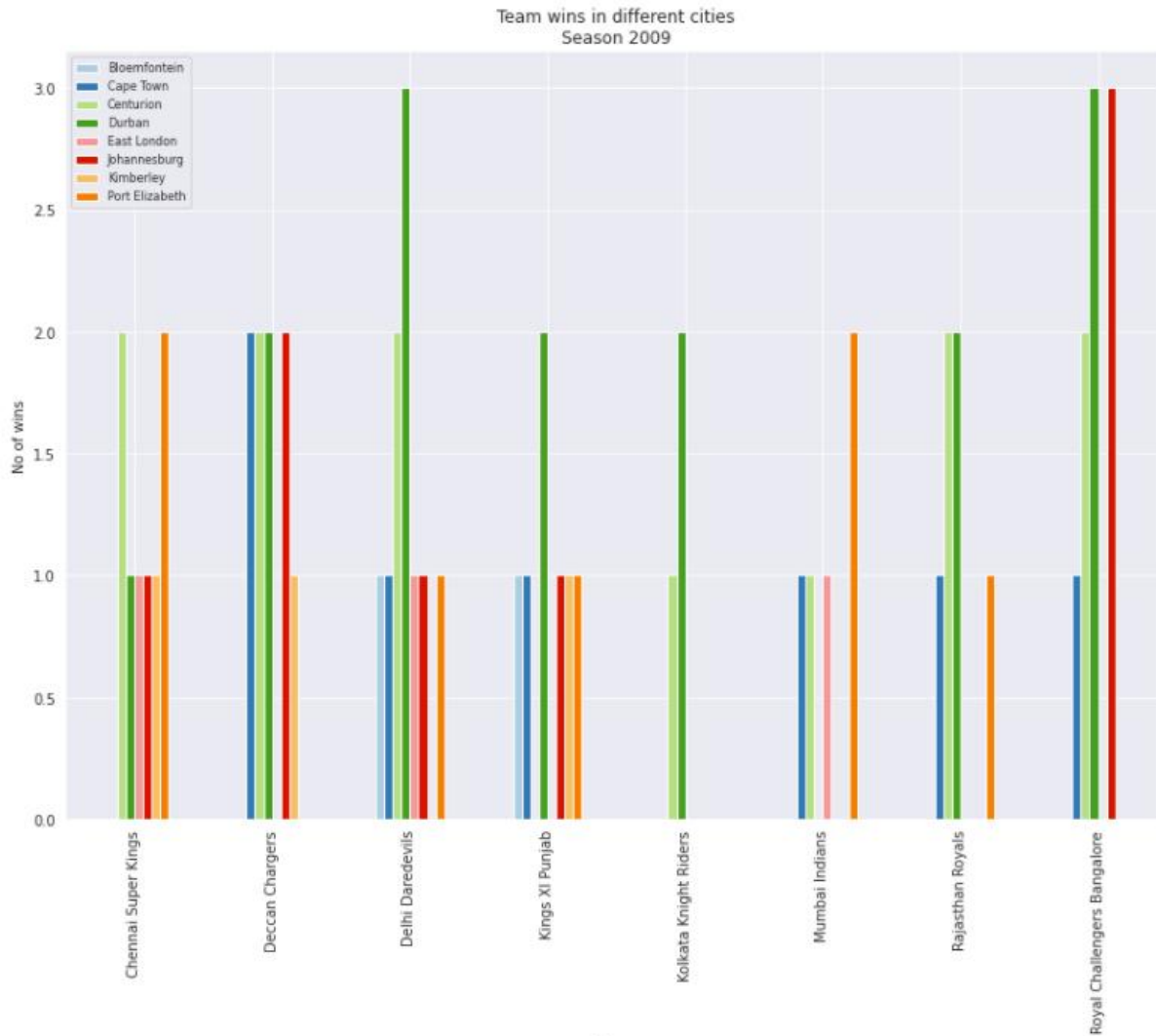
Here x-axis indicates the IPL teams and Y axis indicates the Number of matches won by IPL team in season 2008. Add different city is represented by different colors and its plotted.



This plot indicates the number of wins for different IPL teams in different venues in IPL season 2009.

Here x-axis indicates the IPL teams and Y axis indicates the Number of matches won by IPL team in season 2009. Add different city is represented by different colors and its plotted.

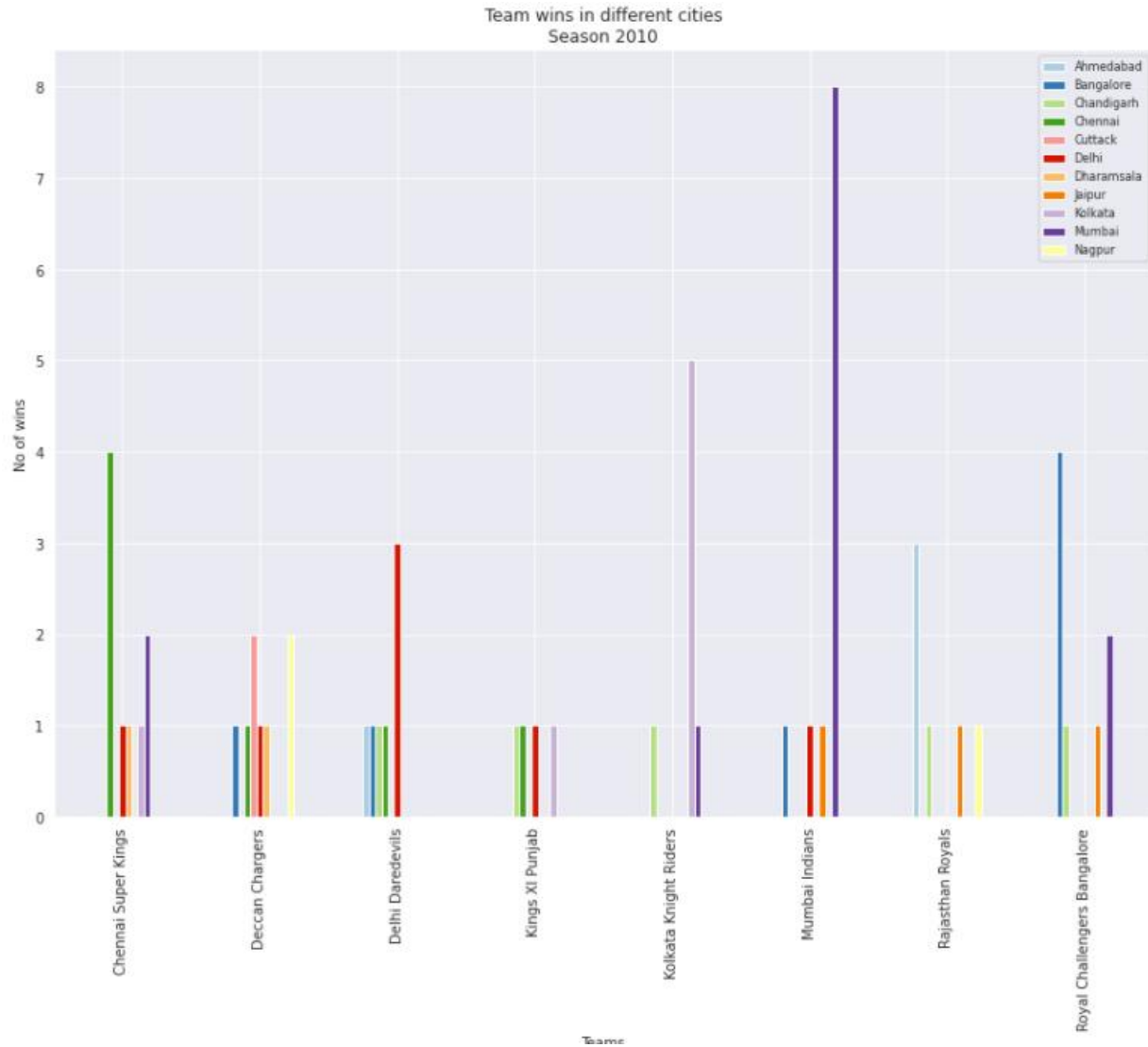
IPL Data Analysis



This plot indicates the number of wins for different IPL teams in different venues in IPL season 2010.

Here x-axis indicates the IPL teams and Y axis indicates the Number of matches won by IPL team in season 2010. Add different city is represented by different colors and its plotted.

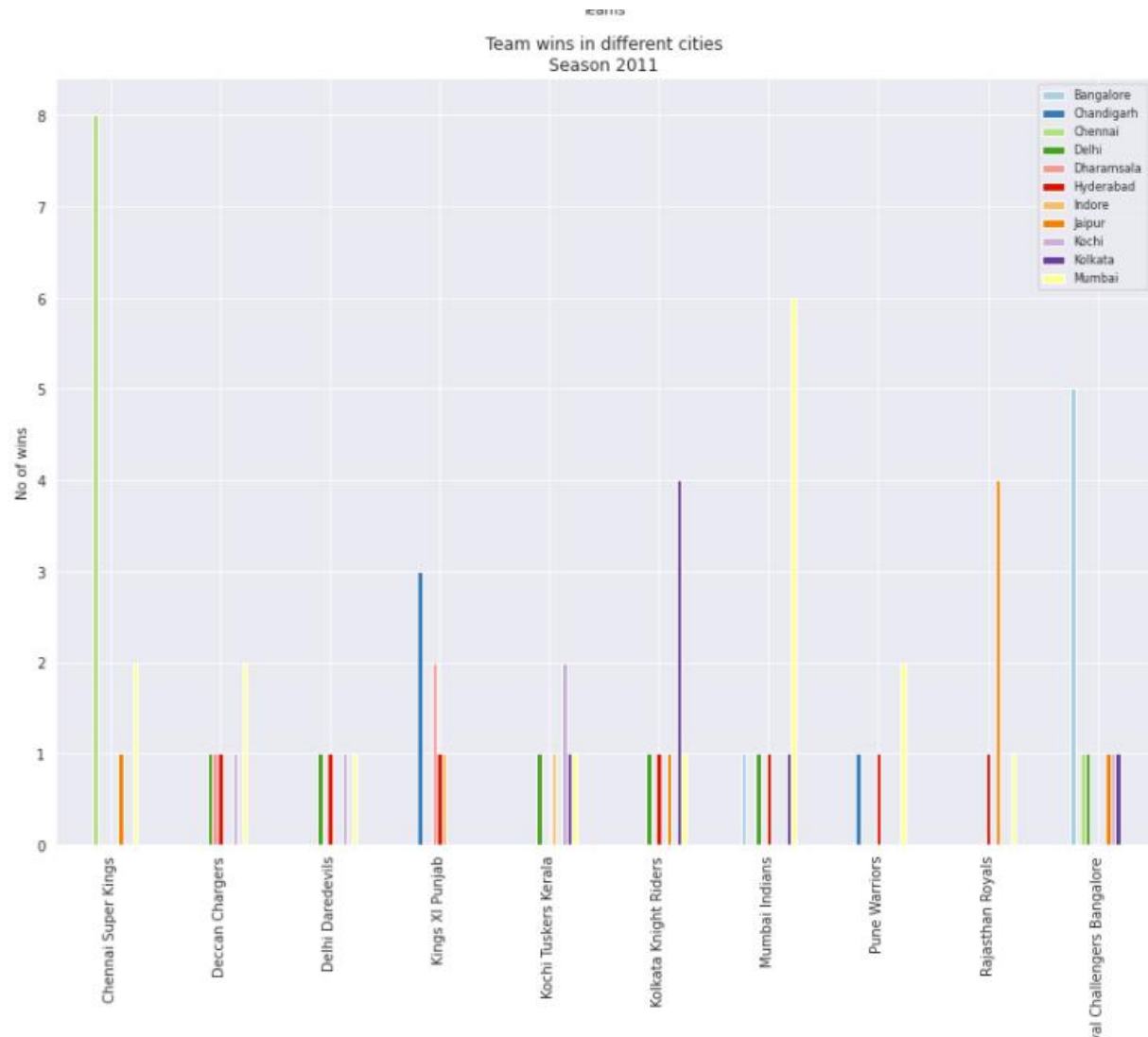
IPL Data Analysis



This plot indicates the number of wins for different IPL teams in different venues in IPL season 2011.

Here x-axis indicates the IPL teams and Y axis indicates the Number of matches won by IPL team in season 2011. Add different city is represented by different colors and its plotted.

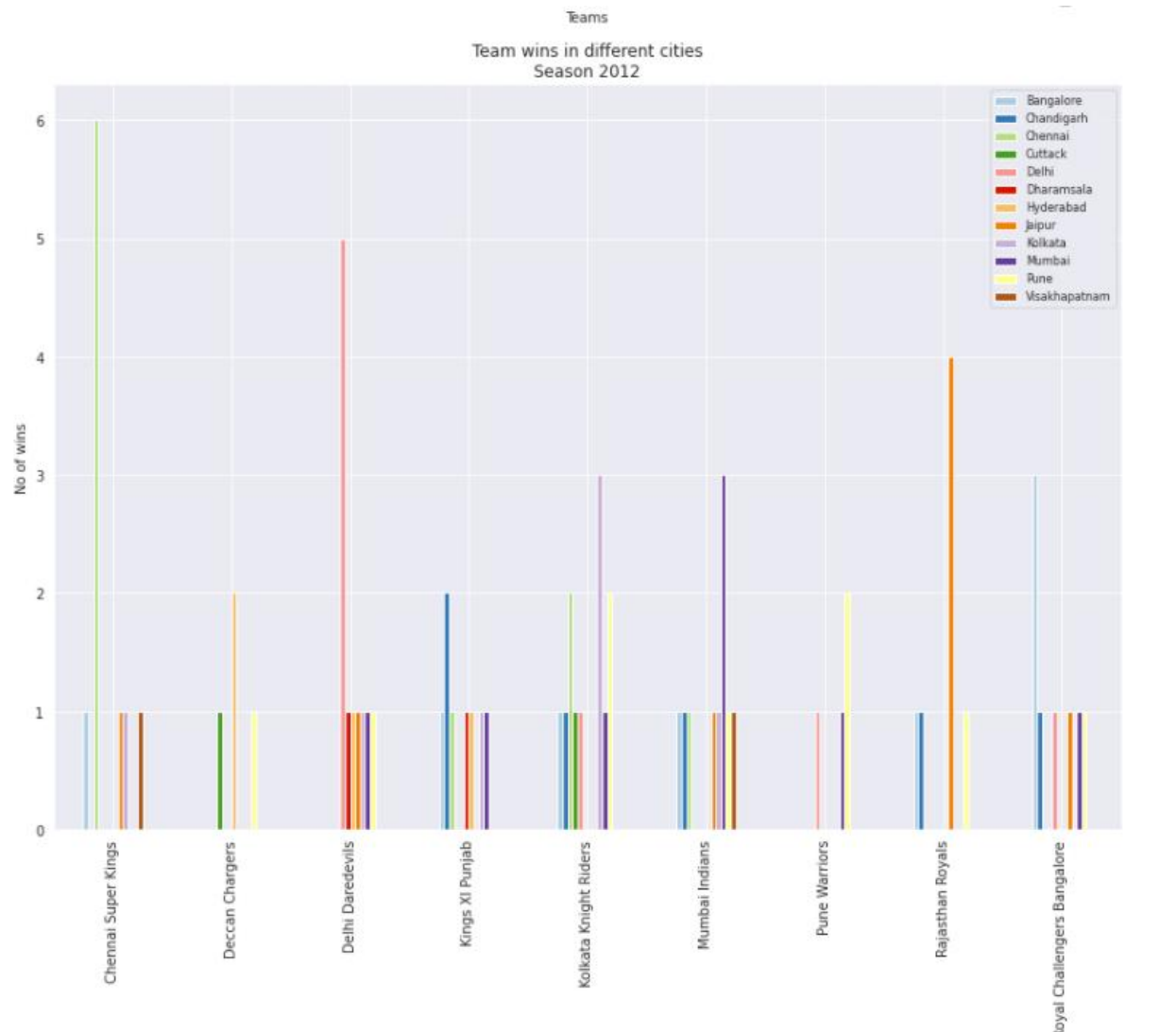
IPL Data Analysis



This plot indicates the number of wins for different IPL teams in different venues in IPL season 2012.

Here x-axis indicates the IPL teams and Y axis indicates the Number of matches won by IPL team in season 2012. Add different city is represented by different colors and its plotted.

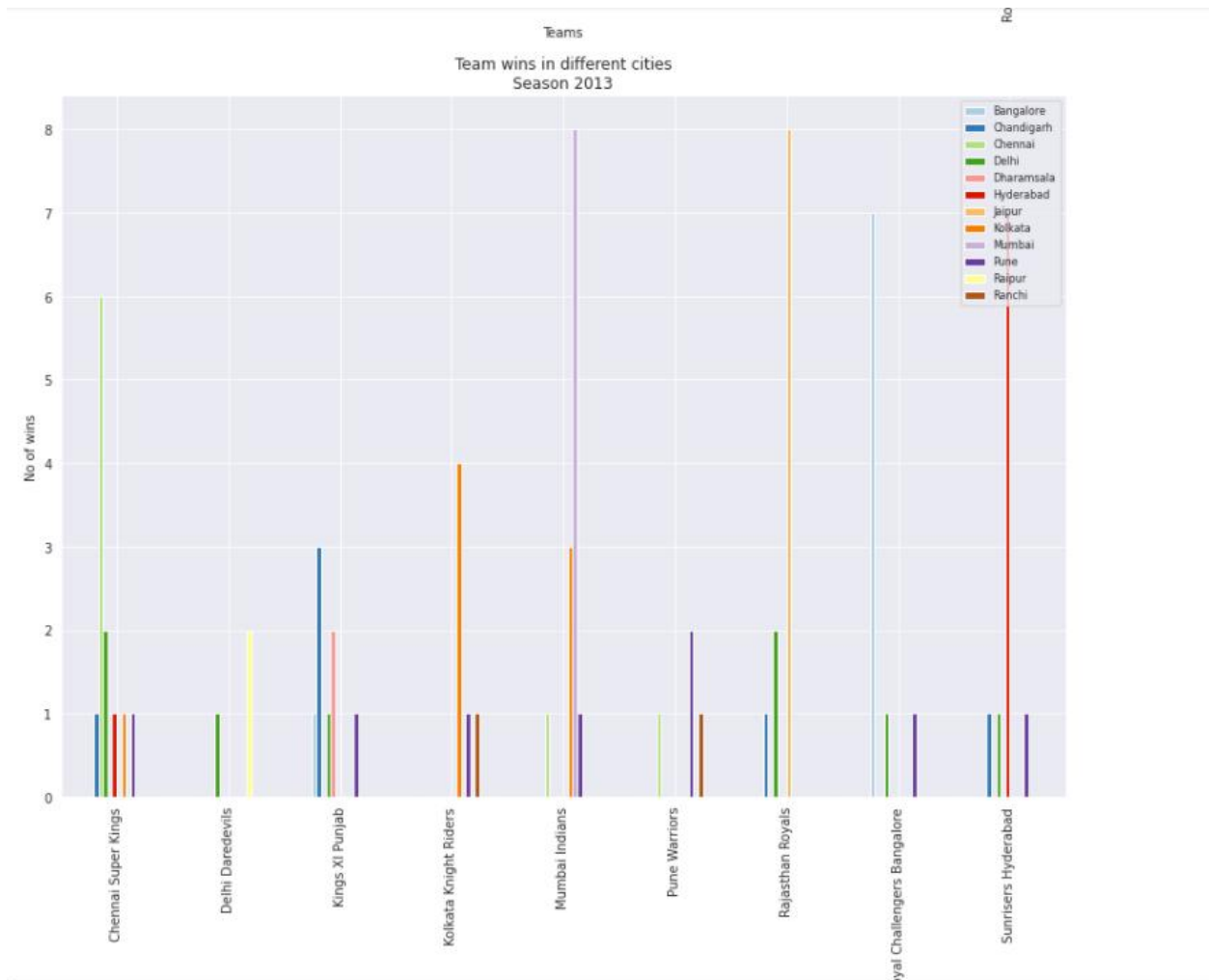
IPL Data Analysis



This plot indicates the number of wins for different IPL teams in different venues in IPL season 2013.

Here x-axis indicates the IPL teams and Y axis indicates the Number of matches won by IPL team in season 2013. Add different city is represented by different colors and its plotted.

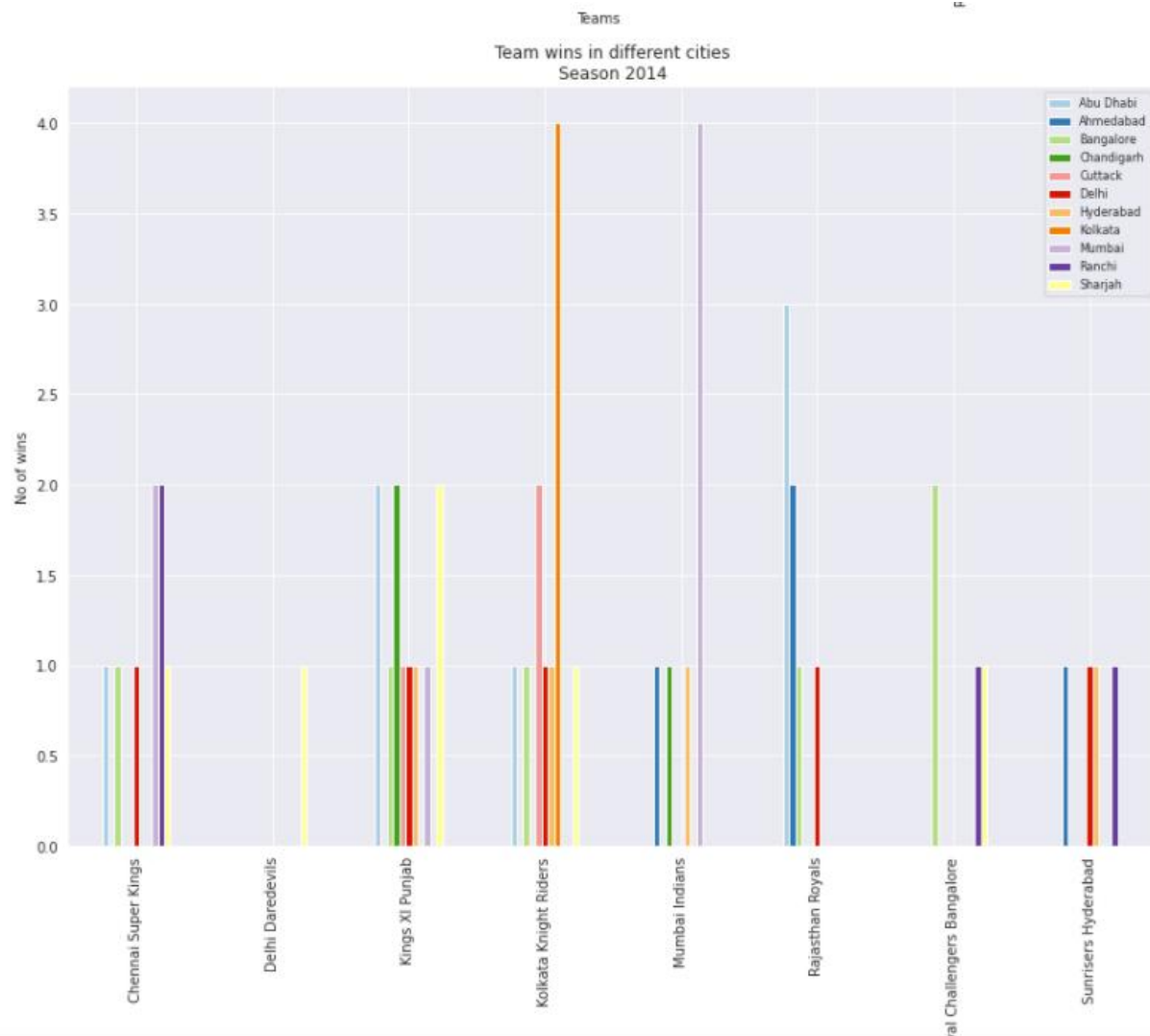
IPL Data Analysis



This plot indicates the number of wins for different IPL teams in different venues in IPL season 2014.

Here x-axis indicates the IPL teams and Y axis indicates the Number of matches won by IPL team in season 2014. Add different city is represented by different colors and its plotted.

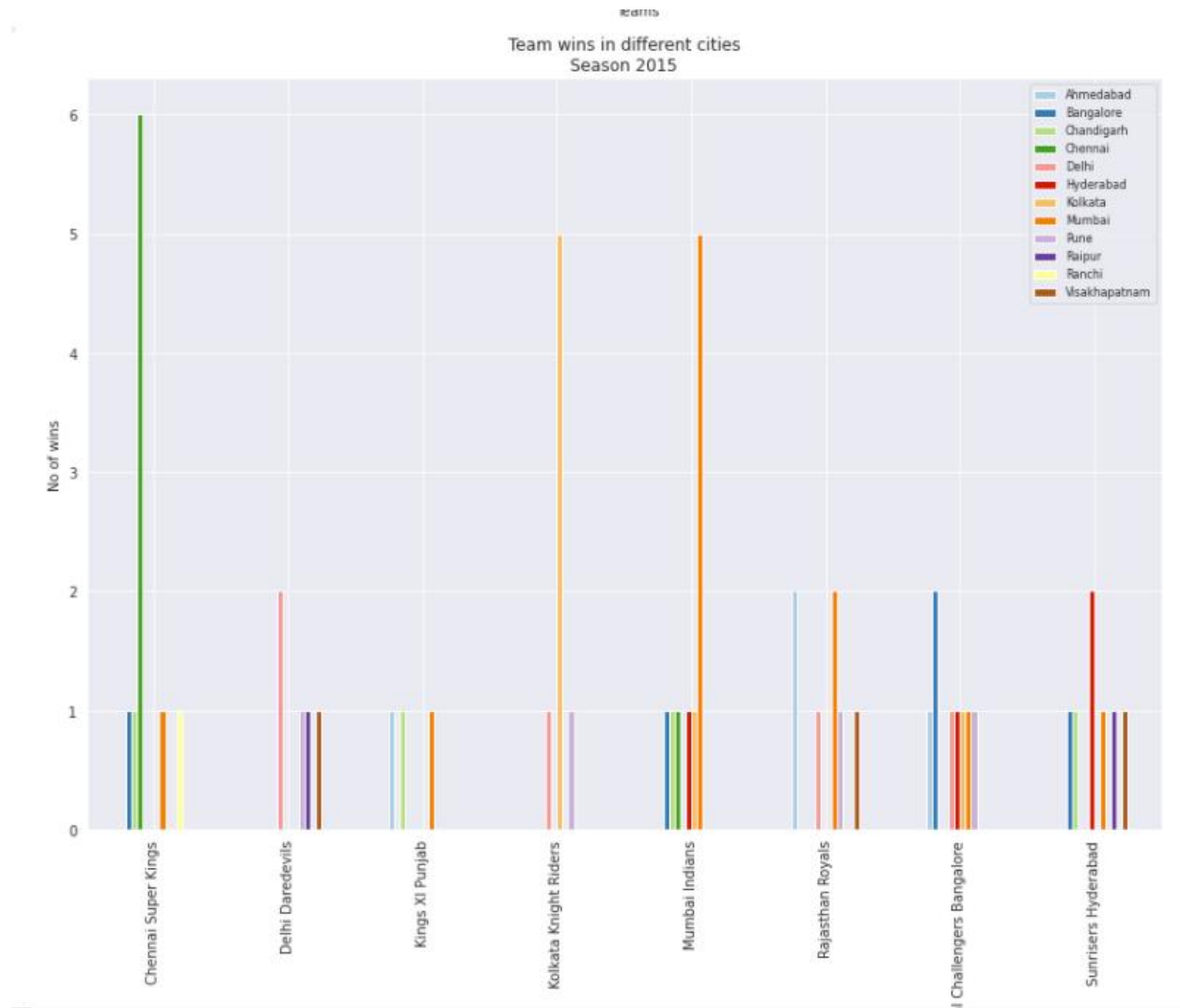
IPL Data Analysis



This plot indicates the number of wins for different IPL teams in different venues in IPL season 2015.

Here x-axis indicates the IPL teams and Y axis indicates the Number of matches won by IPL team in season 2015. Add different city is represented by different colors and its plotted.

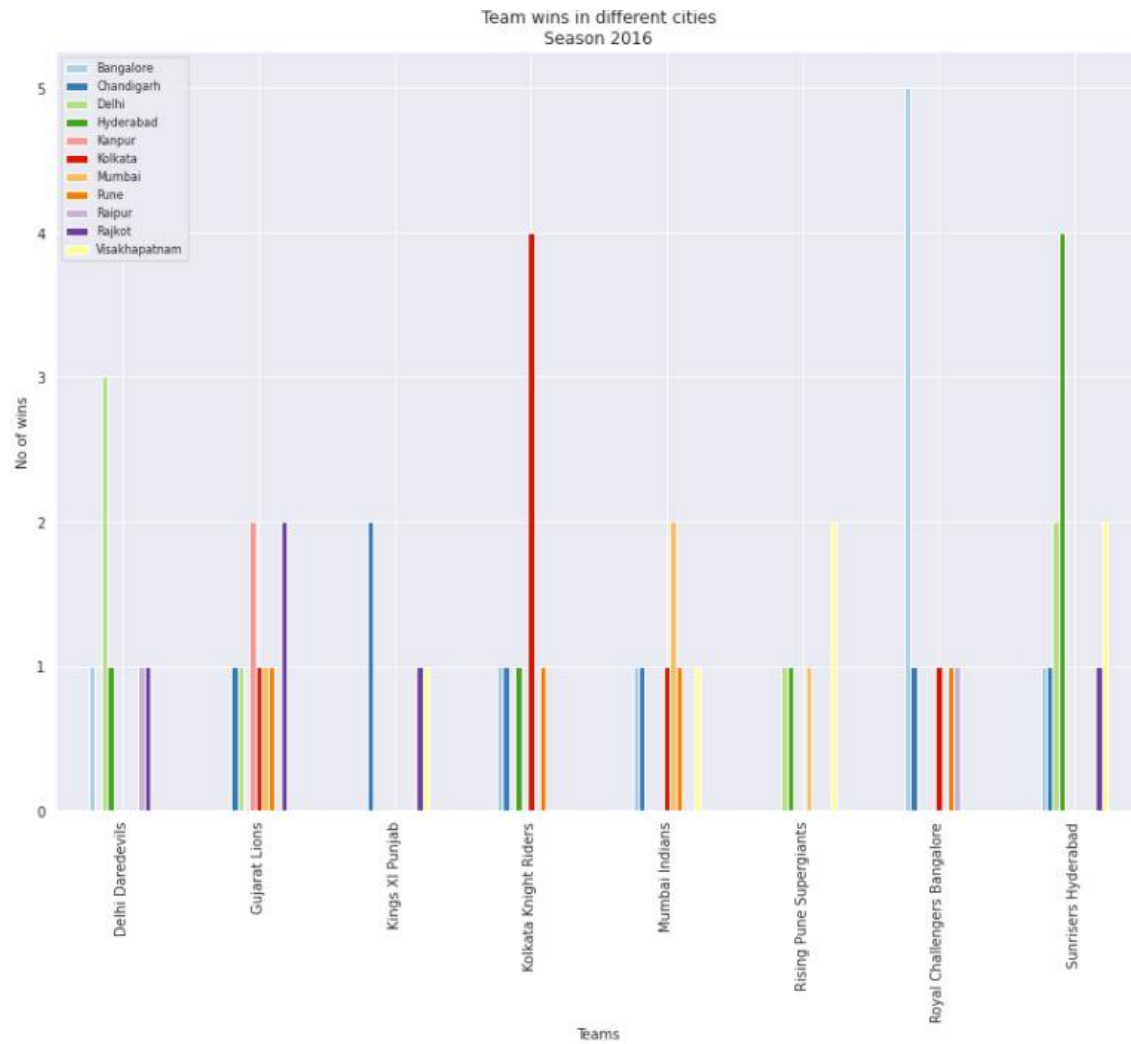
IPL Data Analysis



This plot indicates the number of wins for different IPL teams in different venues in IPL season 2016.

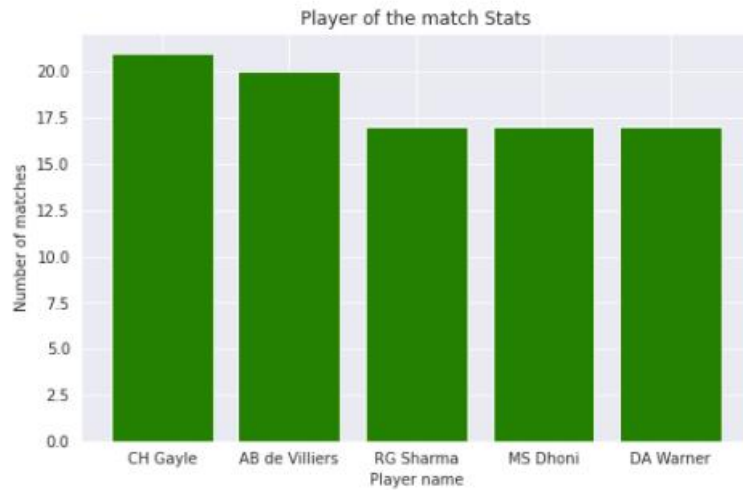
Here x-axis indicates the IPL teams and Y axis indicates the Number of matches won by IPL team in season 2016. Add different city is represented by different colors and its plotted.

IPL Data Analysis

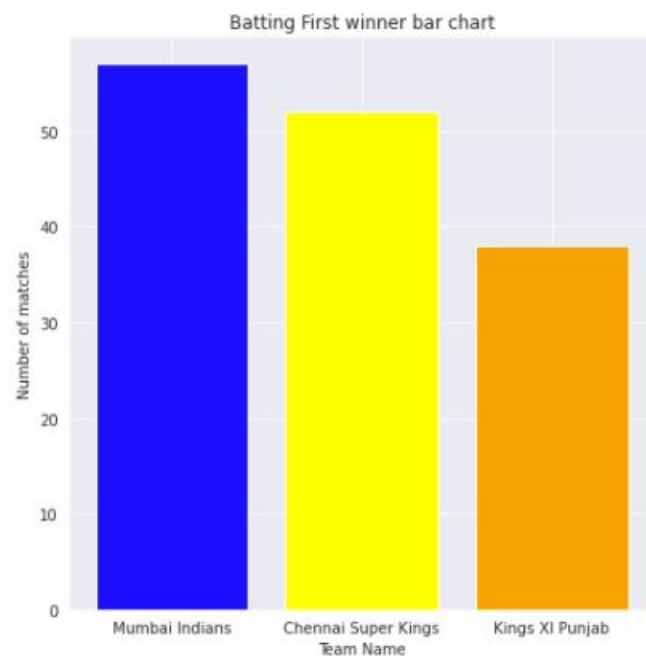


This Plot indicates the Top5 batsmen in IPL, number of the matches is represented in Y-axis and Name of Player of the match in x axis.

IPL Data Analysis



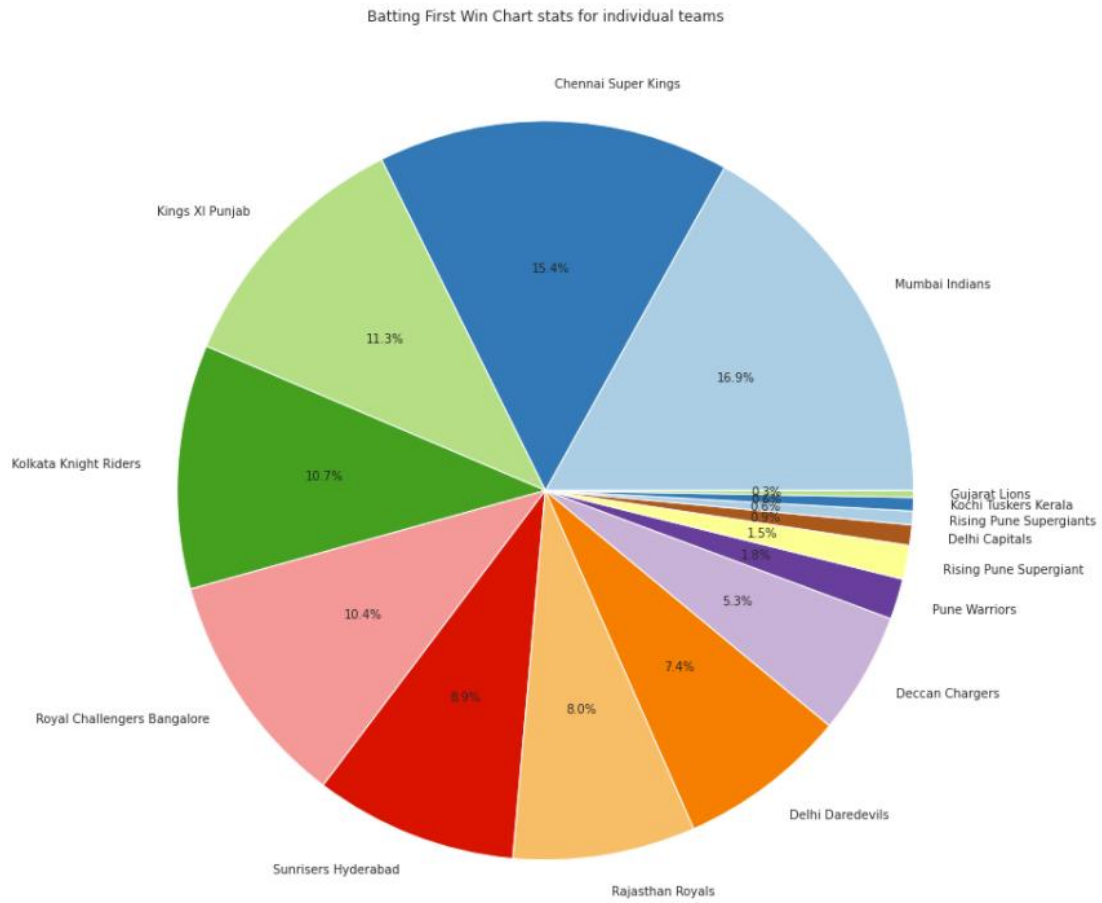
This plot indicates which top 3 teams have won matches maximum batting first. This can be considered as the data for future prediction of who will win the match.



Batting First Win Chart stats for individual teams

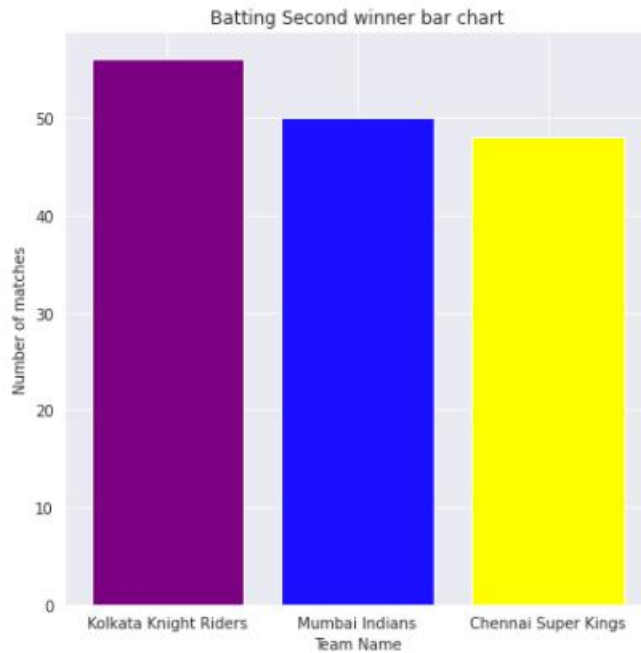
This Pie chart indicates all the IPL team wins based on Batting first.

IPL Data Analysis



This bar plot indicates which 3 teams have won maximum number of matches batting second.

IPL Data Analysis



Code:

Please run the code in Google colab. Have attached the code. Have attached the files which is used to read the data. Dataset files.

Code References:

<https://towardsdatascience.com/analysing-ipl-data-to-begin-data-analytics-with-python-5d2f610126a>

<https://medium.com/ai-in-plain-english/ipl-data-analysis-using-python-b6a0dac0a076>

<https://www.kaggle.com/nulldata/begin-your-data-analysis-in-python-with-ipl-data>

IPL Data Analysis

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.simplefilter(action = "ignore", category = FutureWarning)

%matplotlib inline

ipl_matches_df = pd.read_csv("matches.csv")
ipl_deliveries_df = pd.read_csv("deliveries.csv")
#drop umpire3 column, umpire 3 column has Nan values hence this column is
dropped.
ipl_matches_df.drop(['umpire3'], axis = 1, inplace = True)

# to avoid conflict amonf team names Team names and cricket stadium names
are replaced with a common name throughout the dataset.
ipl_matches_df.team1.replace({'Rising Pune Supergiants' : 'Rising Pune Sup
ergiant'}, regex=True, inplace=True)
ipl_matches_df.team2.replace({'Rising Pune Supergiants' : 'Rising Pune Sup
ergiant'}, regex=True, inplace=True)
ipl_matches_df.winner.replace({'Rising Pune Supergiants' : 'Rising Pune Su
pergiant'}, regex=True, inplace=True)
ipl_matches_df.venue.replace({'Feroz Shah Kotla Ground':'Feroz Shah Kotla'
,
                                'M Chinnaswamy Stadium':'M. Chinnaswamy Stadium',
                                'MA Chidambaram Stadium, Chepauk':'M.A. Chidambaram St
adium',
                                'M. A. Chidambaram Stadium':'M.A. Chidambaram Stadium'
,
                                'Punjab Cricket Association IS Bindra Stadium, Mohali
':'Punjab Cricket Association Stadium',
```

IPL Data Analysis

```
'Punjab Cricket Association Stadium, Mohali':'Punjab
Cricket Association Stadium',
      'IS Bindra Stadium':'Punjab Cricket Association Stadi
um',
      'Rajiv Gandhi International Stadium, Uppal':'Rajiv Gan
dhi International Stadium',
      'Rajiv Gandhi Intl. Cricket Stadium':'Rajiv Gandhi Int
ernational Stadium'}, regex=True, inplace=True)
ipl_deliveries_df.replace('Bangalore', 'Bengaluru', inplace = True)

#fill missing values of city details with venue details for city Column
ipl_matches_df['city'].fillna(ipl_matches_df['venue'], inplace = True)

#fill result details inplace for winnner column in ipl_matches_df datafram
e.
ipl_matches_df['winner'].fillna(ipl_matches_df['result'], inplace = True)

#fill result details inplace of True fir player of the match data
ipl_matches_df['player_of_match'].fillna(ipl_matches_df['result'], inplace
= True)

#fill unknown inplace of Nan for Umpire1 data
ipl_matches_df['umpire1'].fillna('unknown', inplace = True)

#fill unknown inplace of Nan for Umpire2 data
ipl_matches_df['umpire2'].fillna('unknown', inplace = True)

ipl_matches_df_new = ipl_matches_df
ipl_matches_df_new["match_id"] = ipl_matches_df["id"]

#Merging both matches and deliveries.csv data
ipl_match_deliveries_df = pd.merge(ipl_deliveries_df, ipl_matches_df, how=
'left', on=['match_id'])

# take the total list of seasons from matches and deliveries csv file.
season_list = list(ipl_match_deliveries_df["season"].unique())
#Sort the season data list, 2008 to 2019 are listed.
season_list.sort()

#iterate through each season
for season in season_list:
    # season = 2017
    #Bastmens are sorted in Alphabetical order and Runs scored by all batsme
n during each IPL season are listed.
```

IPL Data Analysis

```
batting_tot=ipl_match_deliveries_df.loc[ipl_match_deliveries_df['season']
]== season].groupby('batsman').apply(lambda x:np.sum(x['batsman_runs'])).
reset_index(name='Runs')

#batsmens are sorted as per their runs, batsmen with Highest score come
first.
batting_sorted=batting_tot.sort_values(by='Runs',ascending=False)

#top 20 batsmens with highest runs are listed.
top_batsmen=batting_sorted[:20]
print('The Top 20 Batsmen across different teams in '+str(season)+':\n'
,top_batsmen)

width1 = 30
height1 = 10
width_height_1 = (width1, height1)

plt.figure(figsize=width_height_1)

# plot top 20 batsmens
plt.bar(top_batsmen['batsman'],top_batsmen['Runs'])
#we can see e red dor on the bar graph after the plt.scatter() is execut
ed.
plt.scatter(top_batsmen['batsman'],top_batsmen['Runs'],color='r')
# xticks and yticks are for the text being printed on the plots where w
e specify the fontsize and the angle to which the text needs to be rotated
.
plt.xticks(rotation=60)
plt.xticks(fontsize=30)
plt.xlabel('Top 20 Batsmen across all teams in IPL-
Seasons '+str(season),size=20)
plt.ylabel('Runs Scored',size=20)
plt.title('Top 20 Batsmen across all teams in IPL-
Seasons '+str(season),size=20)
plt.show()

#pick the bowlers with highest number of wickets, sort the list of bowlers
as per totsals number of wickets they have taken and plot it.
for season in season_list:
    bowling_wickets=ipl_match_deliveries_df[ipl_match_deliveries_df['dismiss
al_kind']!='run out']
    bowling_tot=bowling_wickets.loc[bowling_wickets['season']== season].gro
upby('bowler').apply(lambda x:x['dismissal_kind'].dropna()).reset_index(na
me='Wickets')
    #lists the wickets taken by each bowler.
```

IPL Data Analysis

```
bowling_wick_count=bowling_tot.groupby('bowler').count().reset_index()
# bowlers with highest wickets are placed first and the list proceeds in
descending order.
bowling_top=bowling_wick_count.sort_values(by='Wickets',ascending=False)
#lists top20 highest wicket takers.
top_bowlers=bowling_top.loc[:,['bowler','Wickets']][0:20]
width1 = 30
height1 = 10
width_height_1 = (width1, height1)
plt.figure(figsize=width_height_1)
print('The Top Wicket Takers across different teams in the Seasons '+str
(season)+' are:\n',top_bowlers)
plt.bar(top_bowlers['bowler'],top_bowlers['Wickets'],color='r')
plt.plot(top_bowlers['bowler'],top_bowlers['Wickets'],color='g')
plt.xticks(rotation=60)
plt.xticks(fontsize=30)
plt.xlabel('Top 20 Bowlers across different teams in IPL-
Seasons '+str(season),size=20)
plt.ylabel('Wickets Taken',size=20)
plt.title('Top 10 Bowlers across different teams in IPL-
Seasons '+str(season),size=20)
plt.show()

#plots Player Name with respect how many Player fo the match award he has
won in that particular IPL season.
for season in season_list:
    width1 = 30
    height1 = 10
    width_height_1 = (width1, height1)
    plt.figure(figsize=width_height_1)
    ax = sns.countplot("player_of_match", data = ipl_matches_df.loc[ipl_matc
hes_df['season'] == season],order = ipl_matches_df.loc[ipl_matches_df['sea
son'] == season]['player_of_match'].value_counts()[0:20].index,palette='vir
idis')
    plt.title("Total number of Player of the match in season "+str(season))
    # plt.xticks(rotation=90, ha = 'right')
    plt.xticks(rotation=60)
    plt.xticks(fontsize=20)
    plt.ylabel('Number of Player of the match across different teams in sea
son '+str(season),size=20)
    plt.xlabel('Name of the top 20 Player of the match across different tea
ms in season '+str(season),size=20)
    #annotation_plot(ax,0.08,1)
    plt.show()
```


IPL Data Analysis

```
# Plots how many number of matches a team has won on electing to field on
winning a toss and how many number of matches a team has won on electing to
bat on winning a toss.
#for each IPL season the we have the above plot.
for season in season_list:
    width1 = 30
    height1 = 10
    width_height_1 = (width1, height1)
    plt.figure(figsize=width_height_1)
    ax = sns.countplot("winner", data = ipl_matches_df.loc[ipl_matches_df['s
eason'] == season], hue = 'toss_decision',order = ipl_matches_df.loc[ipl_m
atches_df['season'] == season]['winner'].value_counts().index,palette='vir
idis')
    plt.title("Total number of wins in season "+str(season))
    plt.xticks(rotation=45, ha = 'right')
    plt.xticks(fontsize=20)
    plt.ylabel("Total Number of Matches in season "+str(season),size=20)
    plt.title('Team winning Toss Decision VS Match Winning the match in IPL-
Seasons '+str(season),size=20)
    #annotation_plot(ax,0.08,1)
    plt.show()

#plot Wins by teams in different cities here.
ipl_matches_df = pd.read_csv("matches.csv")

ipl_matches_df["type"] = "pre-qualifier"
for year in range(2008, 2017):
    final_match_index = ipl_matches_df[ipl_matches_df['season']==year][-
1:].index.values[0]

ipl_matches_df.groupby(["type"])["id"].count()
ipl_matches_df.head()

ipl_deliveries_df = pd.read_csv("deliveries.csv")
ipl_deliveries_df.head()

# Innings match and corresponding runs are being grouped.
team_score = ipl_deliveries_df.groupby(['match_id', 'inning'])['total_runs
'].sum().unstack().reset_index()

# id ,season, city,date,team1, team2, toss_winner,toss_decision,result,dl_
applied,umpire1,umpire2,umpire3,type,match_id,1,2 are grouped.
matches_data_merged = pd.merge(ipl_matches_df, team_score, left_on = 'id',
    right_on = 'match_id', how = 'outer')
```

IPL Data Analysis

```
#innings, match id and extra runs are being grouped.
team_extras = ipl_deliveries_df.groupby(['match_id', 'inning'])['extra_runs'].sum().unstack().reset_index()

#Appending team_extras to matches_data_merged obtained from ipl_matches_df
matches_data_merged = pd.merge(matches_data_merged, team_extras, on = 'match_id', how = 'outer')

#No of wins by team and season in each city
x, y = 2008, 2017
while x < y:
    #Season wise how many matches each team has won at a particular venue.
    wins_percity = matches_data_merged[matches_data_merged['season'] == x].groupby(['winner', 'city'])['match_id'].count().unstack()
    plot = wins_percity.plot(kind='bar', stacked=False, title="Team wins in different cities\nSeason "+str(x), figsize=(14, 10))
    sns.set_palette("Paired", len(matches_data_merged['city'].unique()))
    plot.set_xlabel("Teams")
    plot.set_ylabel("No of wins")
    plot.legend(loc='', prop={'size':8})
    x+=1

sns.set_style("darkgrid")
plt.rcParams['figure.figsize'] = (14, 8)

# specifies number of times a Player has won player_of_match award.
ipl_matches_df['player_of_match'].value_counts()

#Top 10 player_of_match are being considered
ipl_matches_df['player_of_match'].value_counts()[0:10]

ipl_matches_df['player_of_match'].value_counts()[0:10]

#Names of the top 10 Players are present in the list.
list(ipl_matches_df['player_of_match'].value_counts()[0:10].keys())

plt.figure(figsize=(8,5))

#While Plotting the Graph X-Axis is considered with Player Names and Y-axis is considered with Number of Matches count. Bar Graph indicates number of Player_of_match award won by each Player.
plt.bar(list(ipl_matches_df['player_of_match'].value_counts()[0:5].keys()), ipl_matches_df['player_of_match'].value_counts()[0:5], color="green")
```

IPL Data Analysis

```
plt.title("Player of the match Stats")
plt.xlabel("Player name")
plt.ylabel("Number of matches")
plt.show()
plt.figure(figsize=(8,5))

#indicates the number of matches which are tie, number of matches with no
result and number of matches with normal result win or loose.
ipl_matches_df["result"].value_counts()

#indicates the number of Toss win by each Team in IPL in all seasons.
ipl_matches_df["toss_winner"].value_counts()

#win_by_runs is considered for Batting first. Here whole Dataframe of ipl_
matches_df is considered where win_by_runs is not equal to 0.
batting_first=ipl_matches_df[ipl_matches_df['win_by_runs']!=0]

#Lists how many time sBatting first Each team has won.
batting_first["winner"].value_counts()

plt.figure(figsize=(7,7))
#color plot is done as Purple for Blue for Mumbai Indians, Yellow for Chen
nai Super Kings, Orange for Knigs11 Punjab.
# Here X Axis is denoted by the team names and Y axis represents number of
wins each team has got.
# batting_first["winner"].value_counts()[0:3].keys() provides the 3 Team n
ames and list(batting_first["winner"].value_counts()[0:3]) provides the nu
mber of matches won by each team batting first.
plt.bar(list(batting_first["winner"].value_counts()[0:3].keys()),list(batt
ing_first["winner"].value_counts()[0:3]),color=["blue","yellow","orange"])
plt.title("Batting First winner bar chart")
plt.xlabel("Team Name")
plt.ylabel("Number of matches")
plt.show()

plt.figure(figsize=(14,14))
#list(batting_first["winner"].value_counts() specifies the list of number
of matches won and list(batting_first["winner"].value_counts().keys() spec
ifies the Team names corresponding to number of matches won.
#pie chart is plotted for all the teams with the win percentage.
plt.pie(list(batting_first["winner"].value_counts()),labels=list(batting_f
irst["winner"].value_counts().keys()),autopct="%0.1f%%")
plt.title("Batting First Win Chart stats for individual teams")
plt.show()
```

IPL Data Analysis

```
#win_by_wickets specifies batting second team has won the match. Populates
all the data of ipl_matches_df Dataframe where win_by_wickets!=0.
batting_second=ipl_matches_df[ipl_matches_df["win_by_wickets"]!=0]

# Gives the first five rows of the dataframe.
batting_second.head()

plt.figure(figsize=(12,12))

#specifies the count of matches won by each team batting second in the entire IPL season.
batting_second["winner"].value_counts()

plt.figure(figsize=(7,7))
#color plot is done as Purple for Kolkatta Knight Riders, Blue for Mumbai Indians, Yellow for Chennai Super Kings.
# Here X Axis is denoted by the team names and Y axis represents number of wins each team has got.
# batting_second["winner"].value_counts()[0:3].keys() provides the 3 Team names and list(batting_second["winner"].value_counts()[0:3]) provides the number of matches won by each team batting second.
plt.bar(list(batting_second["winner"].value_counts()[0:3].keys()),list(batting_second["winner"].value_counts()[0:3]),color=["purple","blue","yellow"])
plt.title("Batting Second winner bar chart")
plt.xlabel("Team Name")
plt.ylabel("Number of matches")
plt.show()
```

Conclusion:

Data Visualization of IPL Dataset gives us a detailed analysis of all the Players and their performances. All the Plots give us a detailed analysis of the IPL dataset. Plots help in determining the Top 20 batsmen , Top 20 bowlers, Team winning details in different cities, Winning team based on Toss, Match winning prediction. These plots help in critical decision making for all the teams.