



IPL DATA ANALYSIS

AGENDA

- ✓ Top Performing Teams
- ✓ Insights about Team's previous performances
- ✓ All-Star Team
- ✓ Recommendations
- ✓ Prediction Model for Team Win
- ✓ Appendix

Dataset Description

MATCHES.CSV

RangeIndex: 756 entries, 0 to 755

Data columns (total 15 columns):

#	Column	Non-Null Count	Dtype
---	-----	-----	-----
0	id	756 non-null	int64
1	season	756 non-null	int64
2	city	749 non-null	object
3	date	756 non-null	object
4	team1	756 non-null	object
5	team2	756 non-null	object
6	toss_winner	756 non-null	object
7	toss_decision	756 non-null	object
8	result	756 non-null	object
9	dl_applied	756 non-null	int64
10	winner	752 non-null	object
11	win_by_runs	756 non-null	int64
12	win_by_wickets	756 non-null	int64
13	player_of_match	752 non-null	object
14	venue	756 non-null	object

DELIVERIES.CSV

RangeIndex: 179078 entries, 0 to 179077

Data columns (total 21 columns):

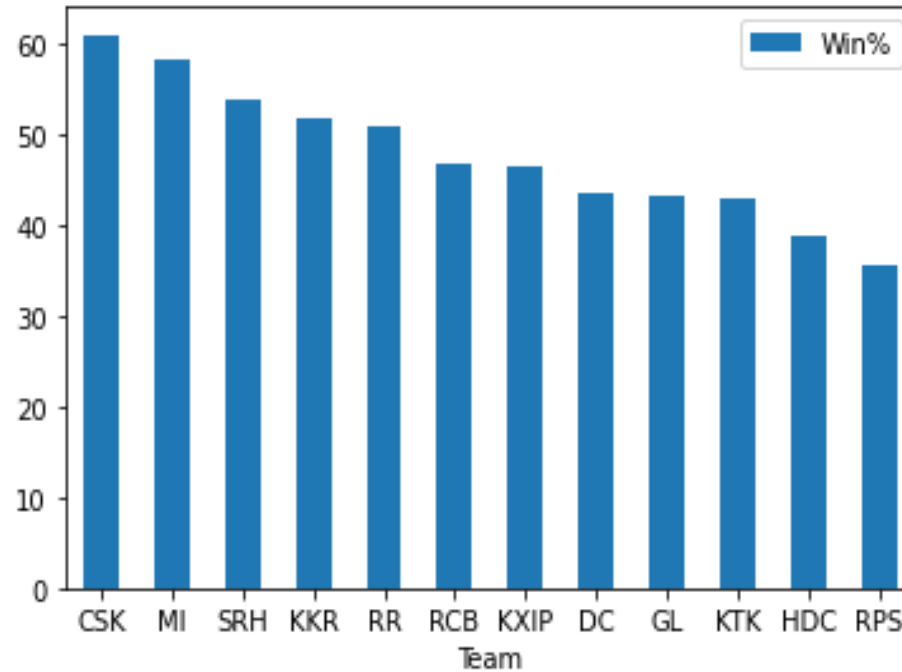
#	Column	Non-Null Count	Dtype
---	-----	-----	-----
0	match_id	179078 non-null	int64
1	inning	179078 non-null	int64
2	batting_team	179078 non-null	object
3	bowling_team	179078 non-null	object
4	over	179078 non-null	int64
5	ball	179078 non-null	int64
6	batsman	179078 non-null	object
7	non_striker	179078 non-null	object
8	bowler	179078 non-null	object
9	is_super_over	179078 non-null	int64
10	wide_runs	179078 non-null	int64
11	bye_runs	179078 non-null	int64
12	legbye_runs	179078 non-null	int64
13	noball_runs	179078 non-null	int64
14	penalty_runs	179078 non-null	int64
15	batsman_runs	179078 non-null	int64
16	extra_runs	179078 non-null	int64
17	total_runs	179078 non-null	int64
18	player_dismissed	8834 non-null	object
19	dismissal_kind	8834 non-null	object
20	fielder	6448 non-null	object

Key to be Used

Team Name	Abbreviation Used
Chennai Super Kings	CSK
Mumbai Indians	MI
Sunrisers Hyderabad	SRH
Kolkata Knight Riders	KKR
Rajasthan Royals	RR
Royal Challengers Bangalore	RCB
Kings XI Punjab	KXIP
Delhi Capitals	DC
Gujarat Lions	GL
Kochi Tuskers Kerala	KTK
Hyderabad Deccan Chargers	HDC
Rising Pune Supergiants	RPS

Top Performing Teams

Team	Total Matches	Total Wins	Win%
CSK	164	100	61.0
MI	187	109	58.3
SRH	108	58	53.7
KKR	178	92	51.7
RR	147	75	51.0
RCB	180	84	46.7
KXIP	176	82	46.6
DC	177	77	43.5
GL	30	13	43.3
KTK	14	6	42.9
HDC	75	29	38.7
RPS	76	27	35.5



	team	winner
0	MI	4
1	CSK	3
2	KKR	2
3	RR	1
4	HDC	1
5	SRH	1

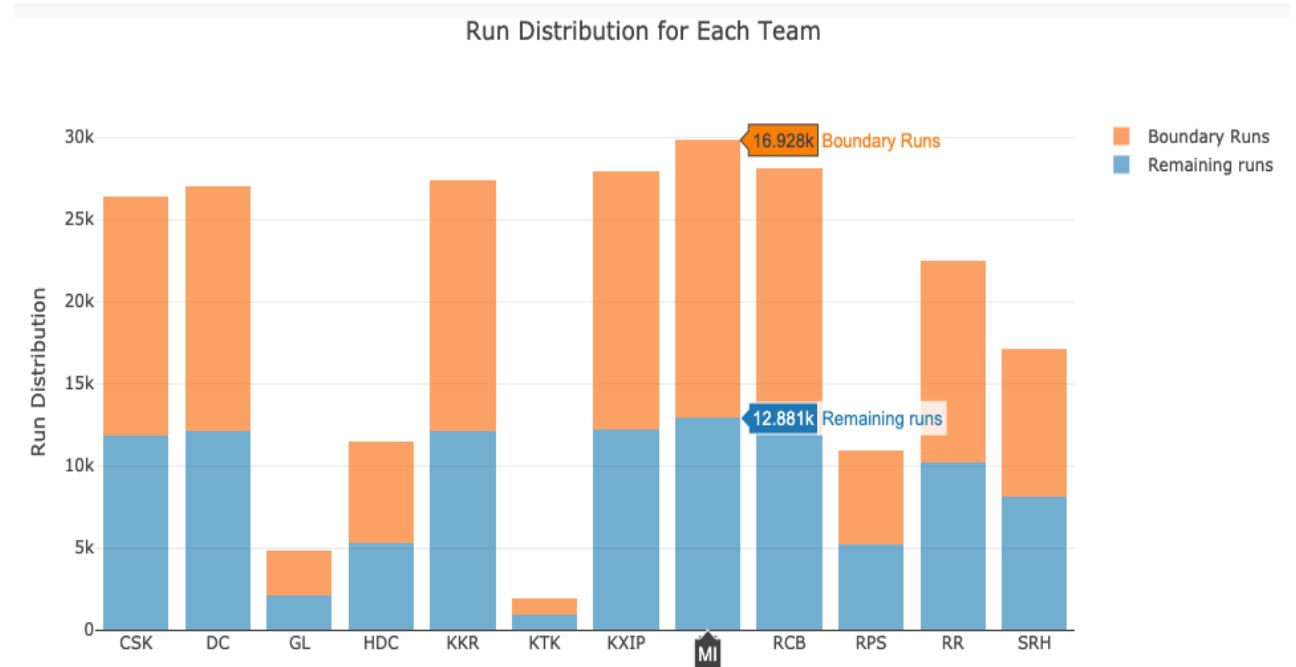
- By far CSK, MI and KKR are the best teams having a win percentage of more than 50% and they have successfully lifted the IPL trophy two or more times.

Team's Previous Performance

- ✓ Team-wise runs Distribution
- ✓ Powerplay middle over and death overs batting performance
- ✓ Over-Wise Runs leaked analysis
- ✓ Powerplay middle over and death overs bowling performance
- ✓ High Score frequency and win% analysis
- ✓ Top Performers

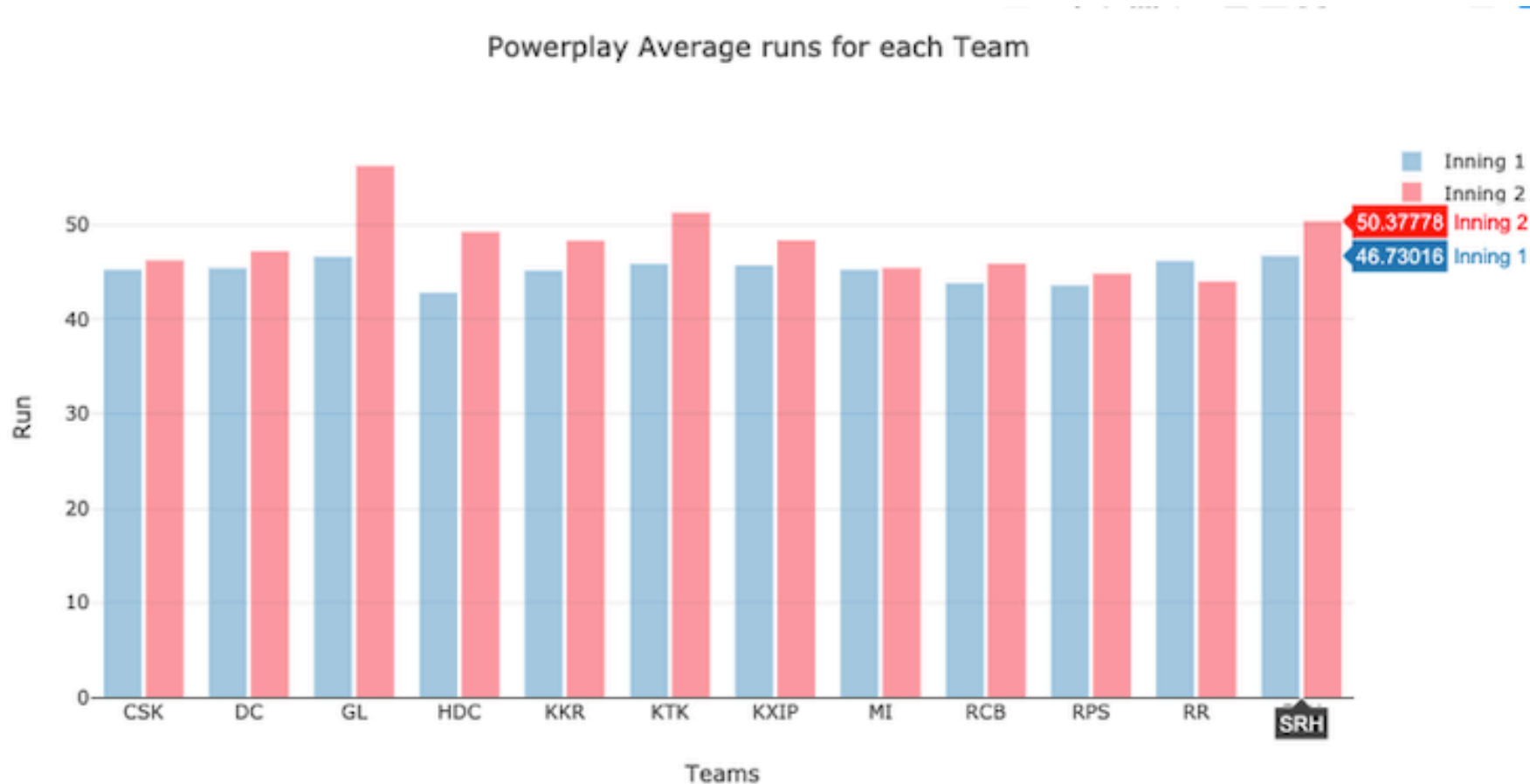
Run Distribution

batting_team	6"s	4"s	total_runs	total_boundaries	boundary%	Avg Score/innings
CSK	5838	8772	26418	14610	55.303202	161.085366
DC	5328	9600	27018	14928	55.252054	159.406417
GL	930	1840	4862	2770	56.972439	157.953704
HDC	2400	3828	11463	6228	54.331327	154.039326
KKR	5580	9736	27419	15316	55.859076	152.591837
KTK	318	680	1901	998	52.498685	156.255556
KXIP	5856	9832	27893	15688	56.243502	158.482955
MI	6576	10352	29809	16928	56.788218	152.644068
RCB	6792	9440	28126	16232	57.711726	162.066667
RPS	2118	3572	10891	5690	52.244973	135.785714
RR	4086	8140	22431	12226	54.504926	152.840000
SRH	3198	5776	17059	8974	52.605663	143.302632



- RCB and CSK have highest average score made in an innings , indicating consistency in their batting order.
- MI and RCB have highest boundary percentage amongst all the teams , indicating presence of power hitters in their batting line-up.

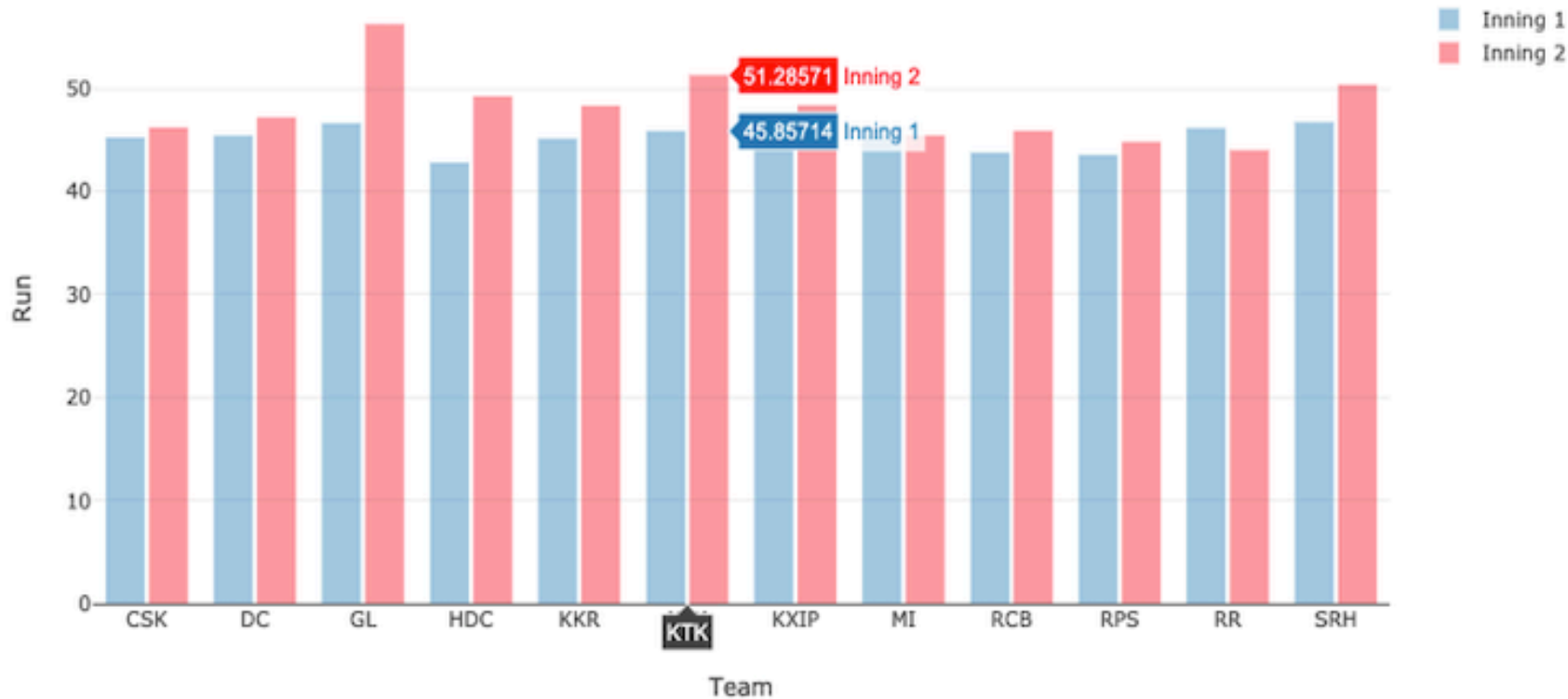
Powerplay(Overs 1-6) Batting Performance



- KXIP and DC put up a decent show (48 runs at 8 RPO) during their powerplay overs, displaying their prowess in top order.
- CSK and MI tend to take a slightly less risky approach during this period scoring around 44 runs at 7.3 RPO.
- Almost all teams lose around 2 wickets during the powerplay(refer to appendix)

Death overs(16-20) Batting Performance

Death Overs Average runs for each Team



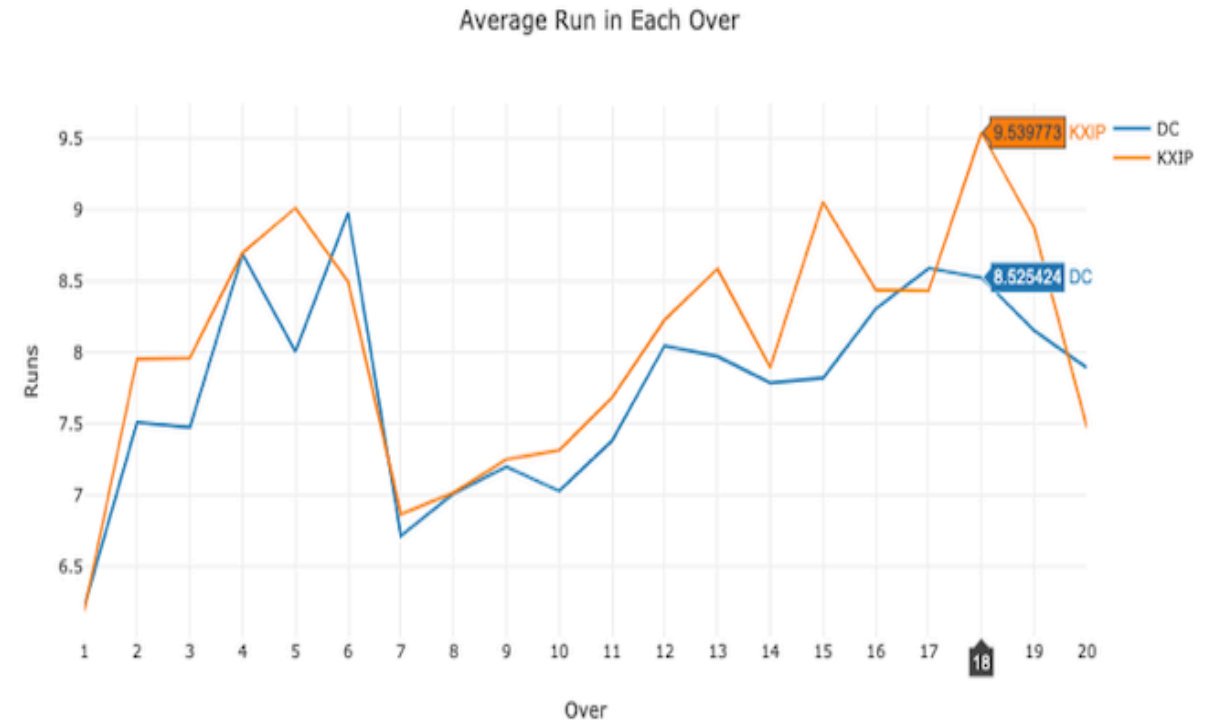
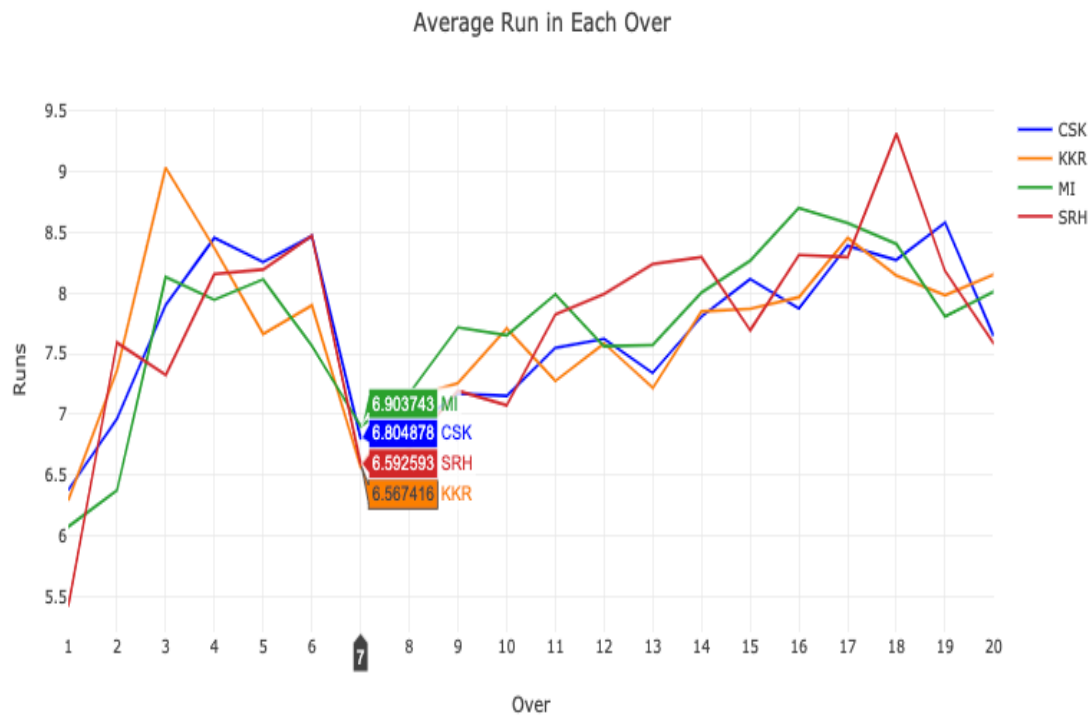
- KKR and SRH really push their accelerators during death overs(48 runs at 9.6 RPO) during their powerplay overs , displaying their prowess in their finishers.
- CSK and MI score at a decent rate of 9 RPO during the due end of an inning.

Over Wise Runs given by Bowling Team

over	CSK	DC	KKR	KXIP	MI	RCB	RR	SRH
1	6.378049	6.214689	6.292135	6.187500	6.074866	6.333333	6.850340	5.416667
2	6.963415	7.508475	7.365169	7.954545	6.374332	7.561111	7.176871	7.592593
3	7.902439	7.474576	9.033708	7.960227	8.133690	8.188889	8.217687	7.324074
4	8.457317	8.689266	8.370787	8.698864	7.946524	7.961111	7.687075	8.157407
5	8.256098	8.005650	7.662921	9.011364	8.112299	8.666667	8.292517	8.194444
6	8.469512	8.977401	7.898876	8.494318	7.566845	8.033333	8.170068	8.472222
7	6.804878	6.711864	6.567416	6.863636	6.903743	6.894444	6.346939	6.592593
8	6.871951	7.011299	7.134831	7.017045	7.171123	7.577778	7.122449	6.712963
9	7.170732	7.197740	7.258427	7.250000	7.716578	7.822222	7.061224	7.194444
10	7.152439	7.028249	7.707865	7.312500	7.652406	7.427778	7.000000	7.074074
11	7.548780	7.378531	7.275281	7.681818	7.989305	7.461111	7.510204	7.824074
12	7.621951	8.045198	7.584270	8.227273	7.561497	7.227778	7.619048	7.990741
13	7.341463	7.971751	7.219101	8.585227	7.572193	7.772222	7.741497	8.240741
14	7.804878	7.785311	7.848315	7.892045	8.005348	7.961111	8.360544	8.296296
15	8.115854	7.819209	7.870787	9.051136	8.267380	8.450000	7.353741	7.694444
16	7.871951	8.305085	7.966292	8.437500	8.700535	8.461111	7.945578	8.314815
17	8.390244	8.587571	8.455056	8.431818	8.577540	9.377778	8.312925	8.296296
18	8.274390	8.525424	8.146067	9.539773	8.406417	8.855556	8.782313	9.314815
19	8.579268	8.152542	7.983146	8.875000	7.807487	8.866667	8.517007	8.185185
20	7.646341	7.892655	8.151685	7.471591	8.010695	7.816667	8.061224	7.583333

- MI and SRH are the most economical ones in powerplay (Overs 1-6).
- RCB, DC and KXIP leaks the most runs in the death overs (16-20).
- CSK is relatively economical(7.2 RPO) during their middle overs , but tend to get expensive during the death overs.
- MI is quite consistent in giving away runs(7.1 RPO) till the 15th over. After this , it's all the way up till almost 8.4 RPO till the 20th over.

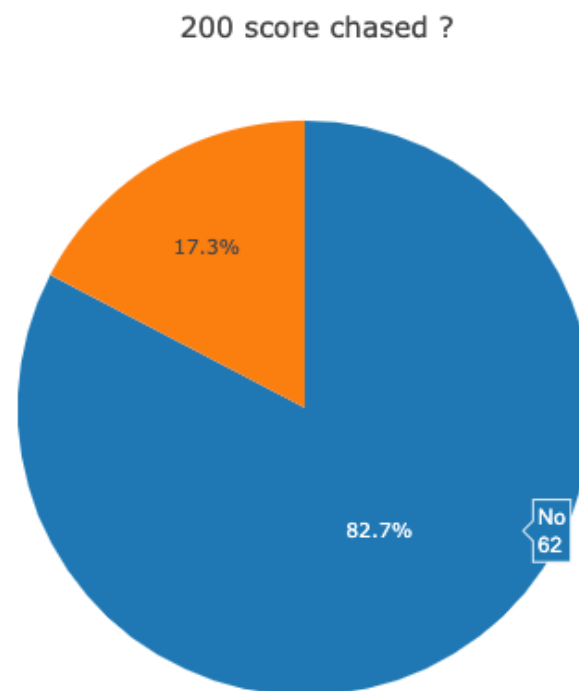
Powerplay, middle overs and death overs bowling



- Top Tier teams are relatively more economical(in the range 6.5-8 runs per over) during the middle overs(7-15).
- DC and KXIP tend to leak more runs(in the range 7-9 runs per over) during their middle overs.

High Score(200+) Analysis

inning	batting_team	total_times
1	RCB	16
1	CSK	11
1	MI	10
1	KKR	9
1	KXIP	9
1	SRH	8
1	DC	6
1	RR	5
1	GL	1
1	HDC	1



- 76 out of 103 times , 200+ score has been achieved while batting first.
- 83% scores of 200+ have never been chased , making it one of the most important factor for a sure shot win.
- RCB ,CSK and MI leads the pack with maximum 200+ scores.
- DC and RR have breached the 200 score mark less than 6 times.

Top Performers of IPL

Top 10 Batsmen

	batting_team	Player	player_runscored
0	Royal Challengers Bangalore	V Kohli	5616
1	Chennai Super Kings	SK Raina	4779
2	Chennai Super Kings	MS Dhoni	4081
3	Royal Challengers Bangalore	AB de Villiers	3895
4	Mumbai Indians	RG Sharma	3837
5	Sunrisers Hyderabad	DA Warner	3461
6	Royal Challengers Bangalore	CH Gayle	3389
7	Kolkata Knight Riders	G Gambhir	3223
8	Rajasthan Royals	AM Rahane	2957
9	Mumbai Indians	KA Pollard	2940
10	Sunrisers Hyderabad	S Dhawan	2693

Top 10 Bowler

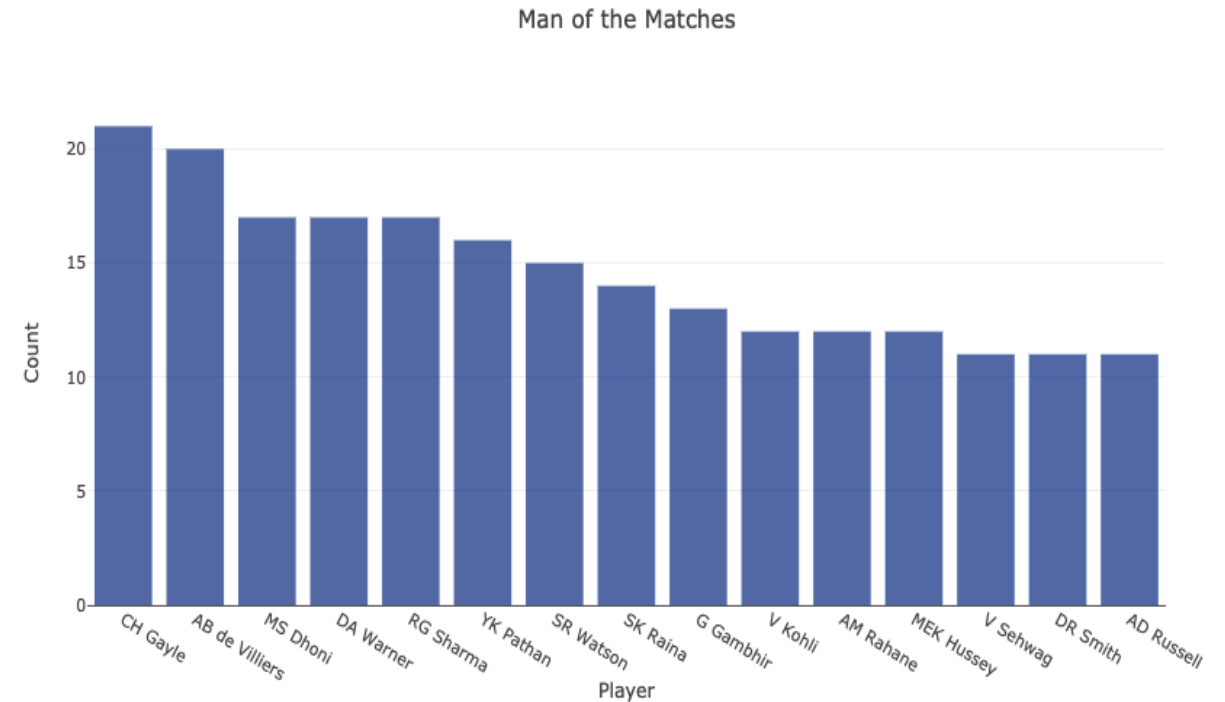
	bowling_team	Player	player_dismissed
0	Mumbai Indians	SL Malinga	188
1	Kolkata Knight Riders	SP Narine	137
2	Mumbai Indians	Harbhajan Singh	136
3	Chennai Super Kings	DJ Bravo	119
4	Sunrisers Hyderabad	B Kumar	116
5	Royal Challengers Bangalore	YS Chahal	102
6	Chennai Super Kings	R Ashwin	99
7	Mumbai Indians	JJ Bumrah	91
8	Delhi Daredevils	A Mishra	91
9	Kings XI Punjab	PP Chawla	89
10	Chennai Super Kings	RA Jadeja	86

Top Performers of IPL

Top 10 All-Rounder

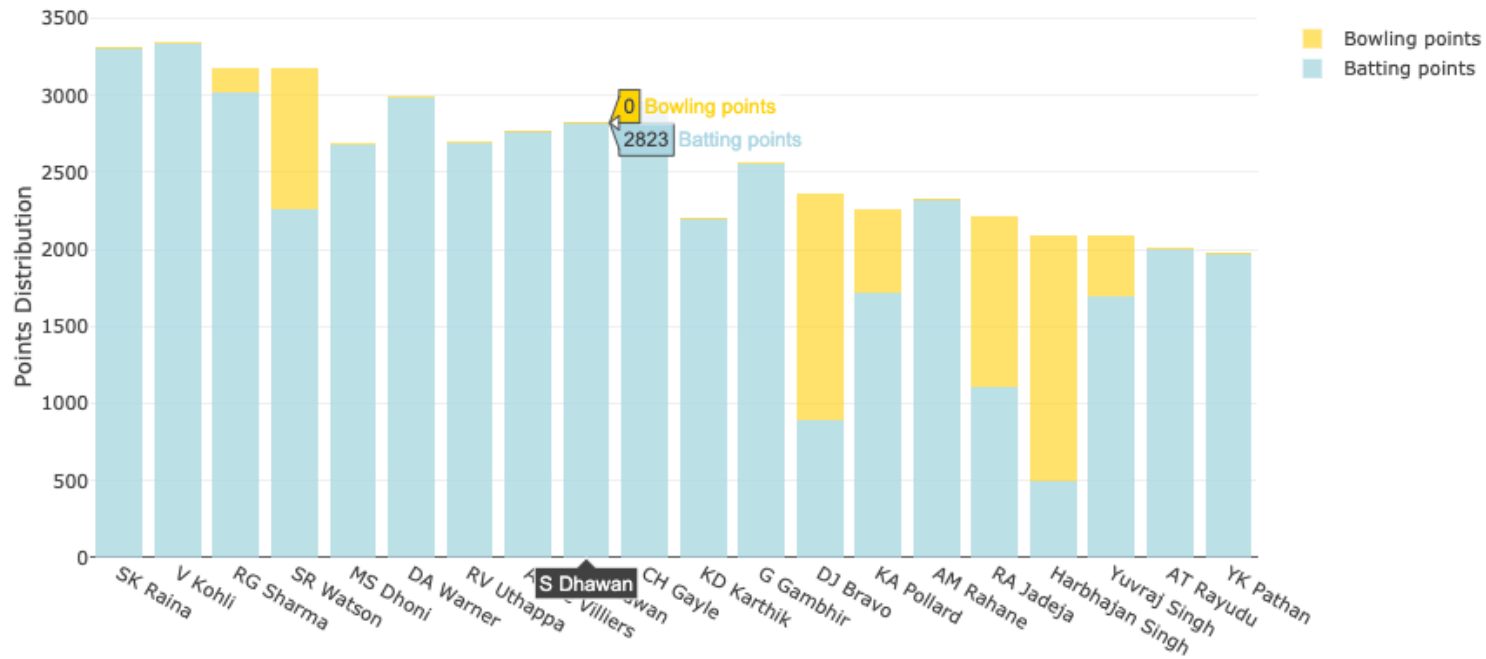
	batting_team	Player	player_runscored	player_dismissed
0	Mumbai Indians	KA Pollard	2940	67
1	Rajasthan Royals	SR Watson	2489	72
2	Chennai Super Kings	SR Watson	1018	72
3	Kolkata Knight Riders	AD Russell	1466	59
4	Kolkata Knight Riders	JH Kallis	1353	48
5	Royal Challengers Bangalore	JH Kallis	1214	48
6	Mumbai Indians	HH Pandya	1200	45
7	Chennai Super Kings	DJ Bravo	995	119
8	Mumbai Indians	KH Pandya	960	42
9	Chennai Super Kings	RA Jadeja	942	86
10	Chennai Super Kings	JA Morkel	871	86

Maximum Player of the match award




All-Star Team

Points Distribution of Top Players



- Point System (refer :Dream11)
 - Batsmen : 8 points : each 100 ; 4points : each 50 ; 1 point for each six ; 0.5 points for each four ; *0.5 for remaining runs
 - Bowler : 10 points for bowled , 14 points for caught and bowled. 4 points 4-w haul and 8 points for 5 wicket haul. 4 points for maiden ,3 points for Economy <4RPO , 2 points for 5<Eco>4 , 1 point for 6<Eco<5 , -1 for 10<Eco<9 , -2 for 11<E<10 , -2 for Eco>11
- Based on venue, fitness and playing conditions , best 11 can be selected from these 20 performers.

Recommendations for losing teams

- 
- ✓ Scoring 200+ while batting first ensures 80% chance of a victory. Team like DC and KXIP ,at the bottom of the table, should aim for a 200+ score while batting first.
 - ✓ Top Teams like MI and CSK concede less runs during middle overs, whereas DC and KXIP go for more. They should capitalize these middle overs so that there is less burden on death over bowlers.
 - ✓ Top Teams like SRH and CSK, are quite consistent with their Run rate throughout the innings, whereas teams like KXIP and DC burst through the first six overs and during the death overs , remaining silent during the middle ones.
 - ✓ Very few players from losing team in top performers. They can plan for the next auction using player data and aim for players with decent strike rate to accelerate during middle overs , and economical bowlers for decent middle overs spells.

Prediction Model for Team Win

```
In [130]: model = RandomForestClassifier(n_estimators=100)
outcome_var = ['winner']
predictor_var = ['team1', 'team2', 'venue', 'toss_winner', 'city', 'toss_decision']
classification_model(model, df, predictor_var, outcome_var)
```

```
<ipython-input-128-97cc6167b4c0>:2: DataConversionWarning: A column-vector y was passed when a 1d array was expected.
Please change the shape of y to (n_samples,), for example using ravel().
    model.fit(data[predictors], data[outcome])
```

Accuracy : 85.581%

```
#'team1', 'team2', 'venue', 'toss_winner', 'city', 'toss_decision'
team1='RCB'
team2='KKR'
toss_winner='RCB'
input=[dicVal[team1], dicVal[team2], '14', dicVal[toss_winner], '2', '1']
input = np.array(input).reshape((1, -1))
output=model.predict(input)
print(list(dicVal.keys())[list(dicVal.values()).index(output)]) #find key by value search output
```

KKR

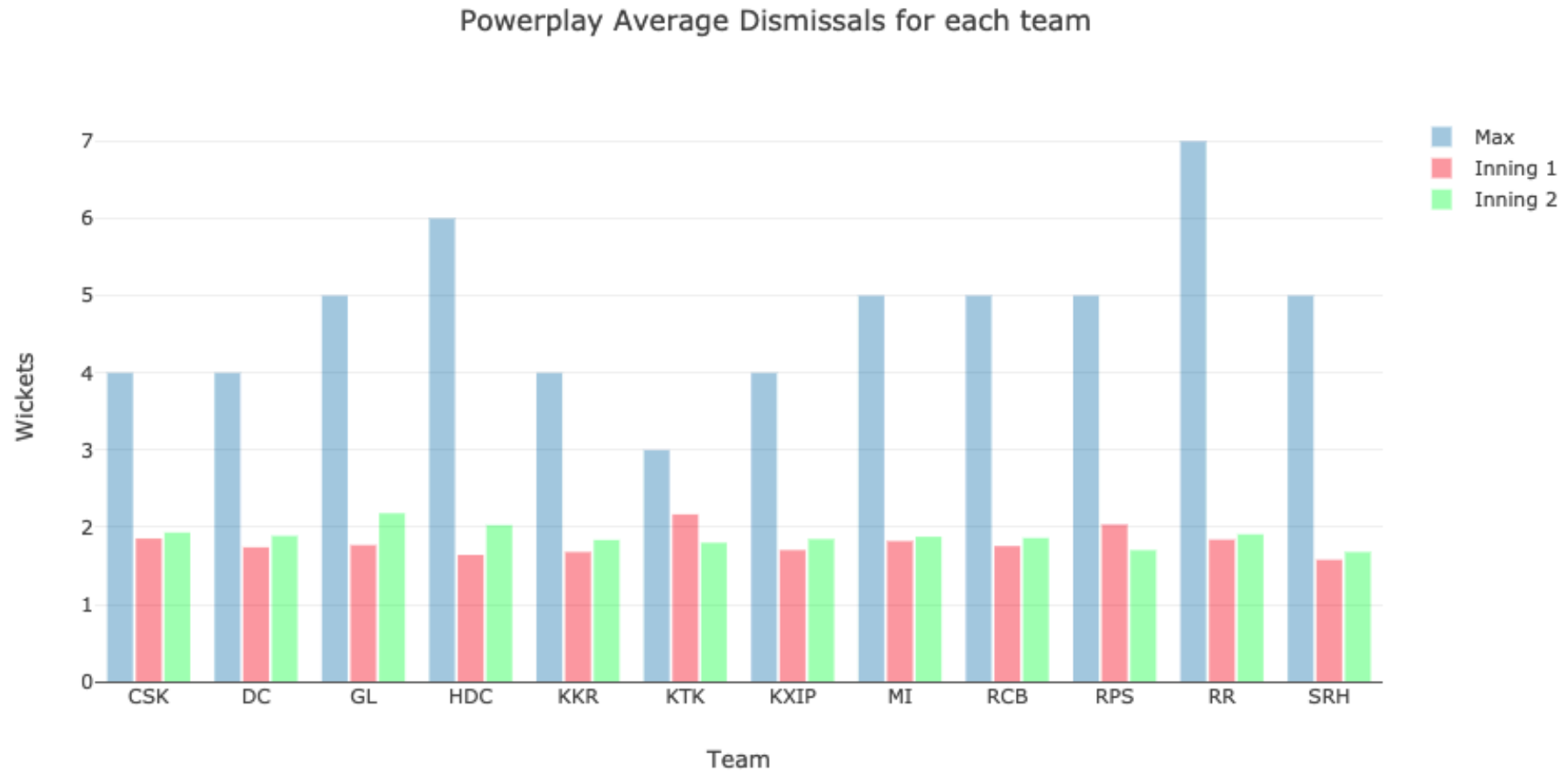
- Features Used : team1, team2, venue(pitch condition), city(weather condition), toss_winner(luck factor), toss_decision.
- Random forest classifier emerged as the best model with an accuracy of 85%.

APPENDIX

- ✓ Powerplay Dismissals Team-Wise
- ✓ Point calculation for All-Star Team



Powerplay Average Dismissals for each team



Point Calculation for All-Star

Batsman	Points	Sixes	Fours
SK Raina	3150	195	495
V Kohli	3149	191	482
RG Sharma	2866.5	194	431
CH Gayle	2795	327	376
DA Warner	2781	181	459
S Dhawan	2675	96	526
AB de Villiers	2606.5	214	357
RV Uthappa	2597	156	436
MS Dhoni	2594	207	297

Bowler	Dismissals	Points
SL Malinga	170	1720
A Mishra	156	1572
Harbhajan Singh	149	1530
PP Chawla	149	1506
DJ Bravo	147	1502
B Kumar	133	1350
R Ashwin	125	1262
SP Narine	122	1248
UT Yadav	110	1100



Email : rakesh.soni1998@gmail.com

Date : 14, December 2020