Indian Premier League - EDA

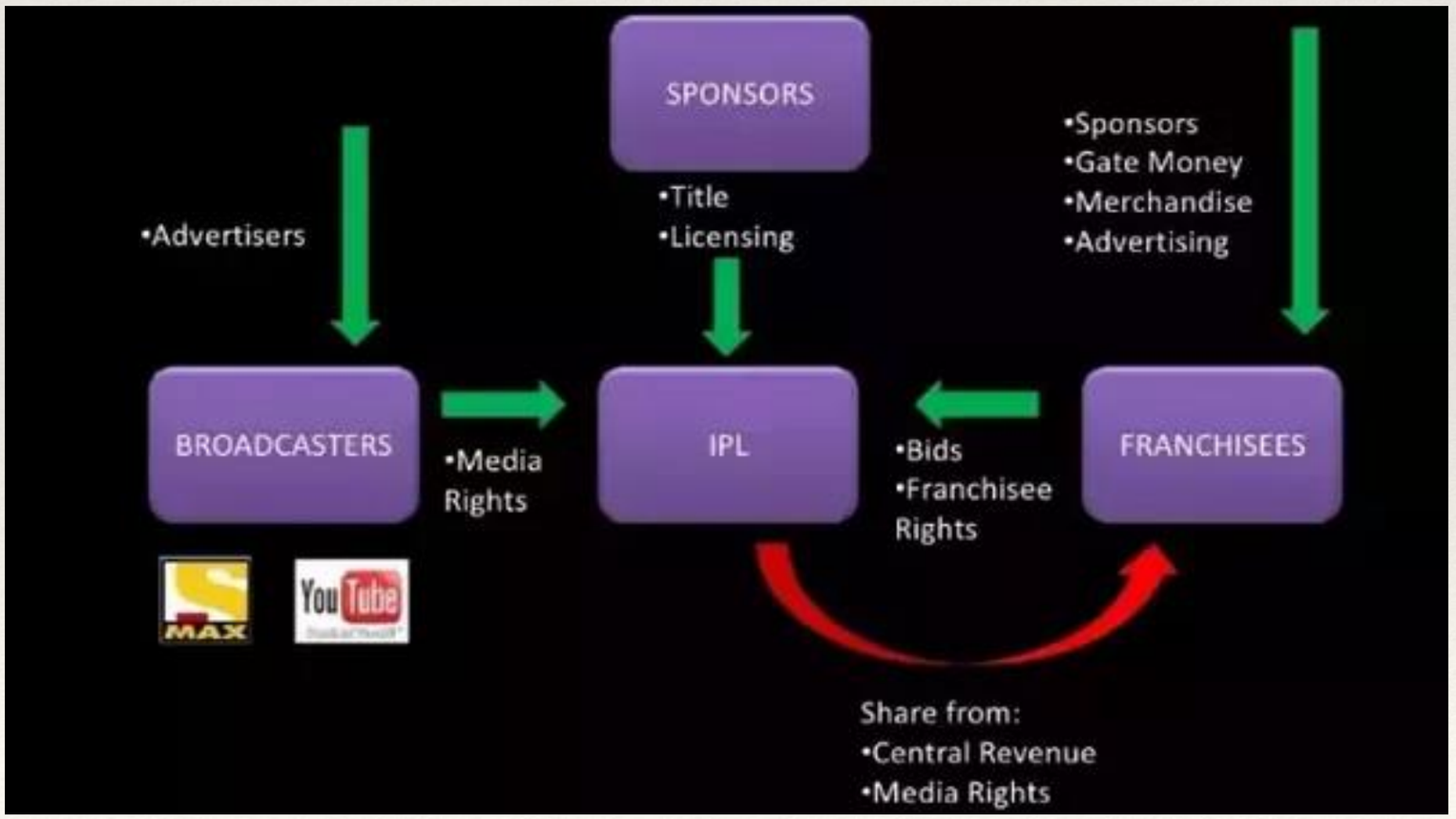
by Kamal Upadhyaya

Table of Content

IPL

The Indian Premier League (IPL), officially Vivo Indian Premier League for sponsorship reasons, is a professional Twenty20 cricket league in India contested during April and May of every year by teams representing Indian cities and some states. The league was founded by the Board of Control for Cricket in India (BCCI) in 2008, and is regarded as the brainchild of Lalit Modi, the founder and former commissioner of the league. IPL has an exclusive window in ICC Future Tours Programmes.

IPL is ***allowing corporate India into the dressing room of Indian cricket.*** IPL's business master stroke is an amazing synergy of entertainment, cricket and industry. When the largest democracy in the world is mad about cricket, IPL is a paradise for investors.



IPL - EDA

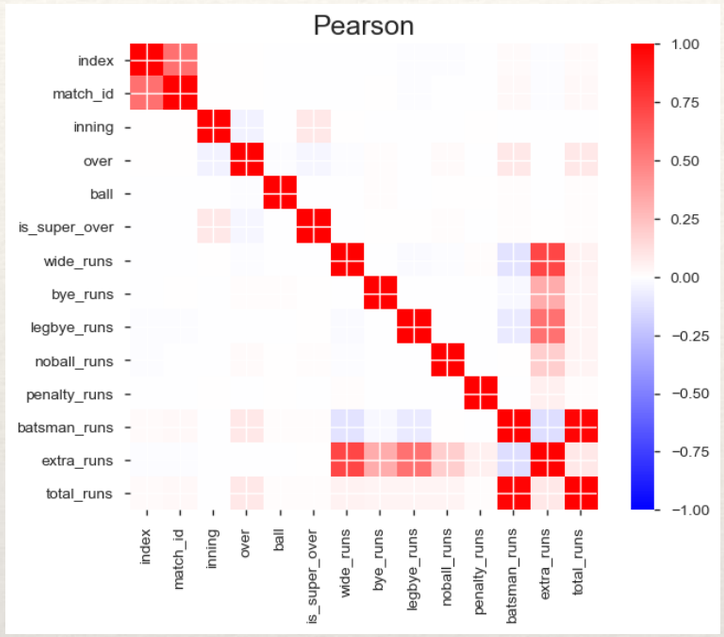
- Exploratory data analysis (EDA) is an approach to analyze data sets to summarize their main characteristics, often with visual methods. A statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modeling or hypothesis testing task

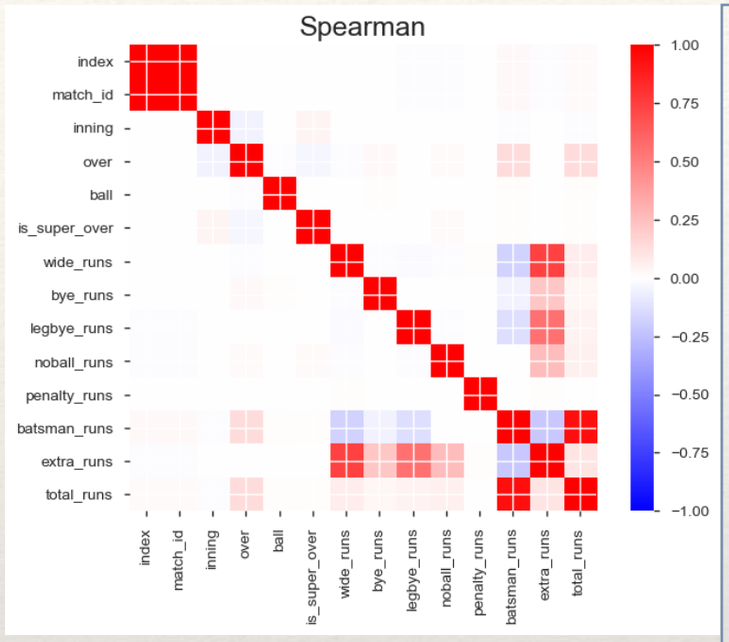
- With this Exploratory Data Analytics on IPL, we have tried to explore IPL data on three dimension bowling, batting and team performance. provide observations with each step in order to explain thoroughly on the approach. Based on the observation, found answers to some of the questions and couple of hypothesis.

Data Profiling

- The initial data profiling provides following insights:
- There are 164750 observations of 22 variables. 13 numeric and 8 object types.
- Below variables have high level of Null values obviously due to low number of dismissals compared to number of deliveries.
  - player_dismissed
  - dismissal_kind
  - fielder
- Removed duplicate records
- Replaced Rising Pune Supergiant with Rising Pune Supergiants
- Below teams have played in IPL. Few teams have played less seasons, thus having less data.
  - Sunrisers Hyderabad
  - Mumbai Indians
  - Gujarat Lions
  - Kings XI Punjab
  - Chennai Super Kings
  - Deccan Chargers
  - Pune Warriors
  - Royal Challengers Bangalore
  - Rising Pune Supergiants
  - Kolkata Knight Riders
  - Delhi Daredevils
  - Rajasthan Royals
  - Kochi Tuskers Kerala

Data Profiling



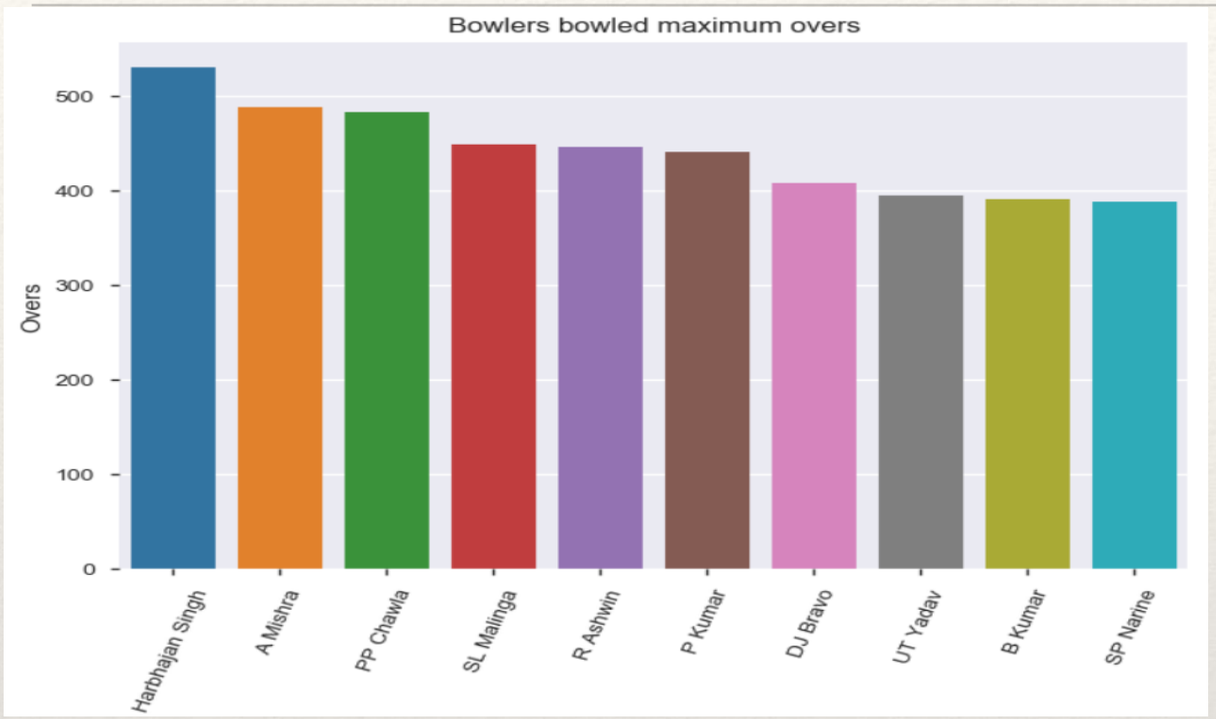


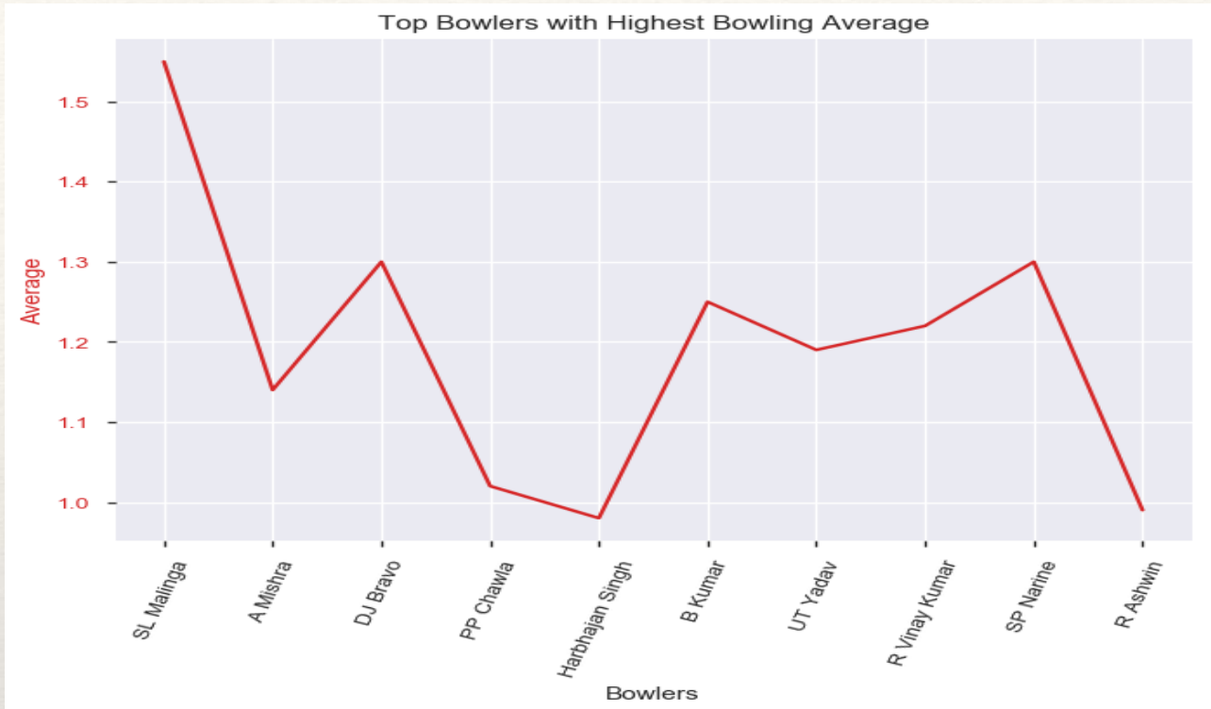
Sample

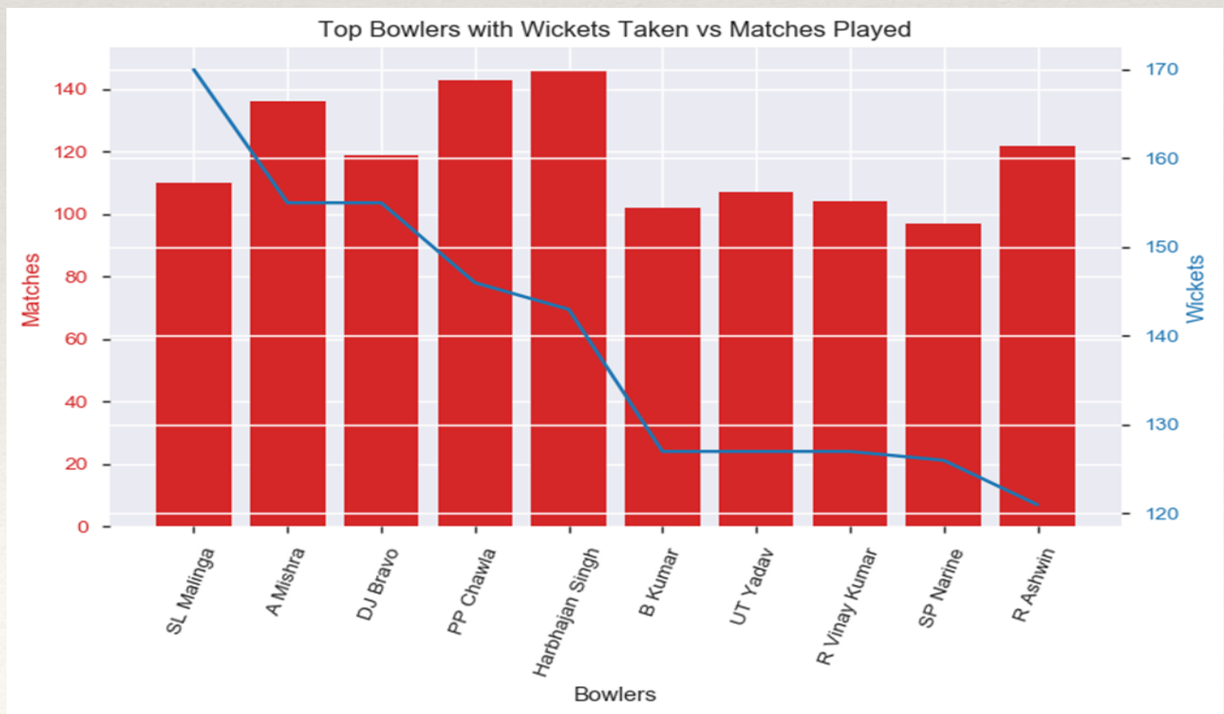
	match_id	inning	batting_team	bowling_team	over	ball	batsman	non_striker	bowler	is_super_ov
0	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	1	DA Warner	S Dhawan	TS Mills	0
1	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	2	DA Warner	S Dhawan	TS Mills	0
2	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	3	DA Warner	S Dhawan	TS Mills	0
3	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	4	DA Warner	S Dhawan	TS Mills	0
4	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	5	DA Warner	S Dhawan	TS Mills	0

- Dataset Info:
  - Variables: 22
  - Observations:164745
  - Total Missing (%):13.0%
  - Size in memory : 27.7 MiB
  - Average record size:176.0 B
- Variables types:
  - Numeric:11
  - Categorical:8
  - Boolean:2
  - Date:0
  - Text (Unique):0
  - Rejected:1
  - Unsupported:0
- Total runs are highly correlated with batsman_runs.
- Wide, Bye, Legbye, noball have high correlation with extra_runs.
- Penalty _runs have only two observations, so can be discarded.

Bowling Analysis

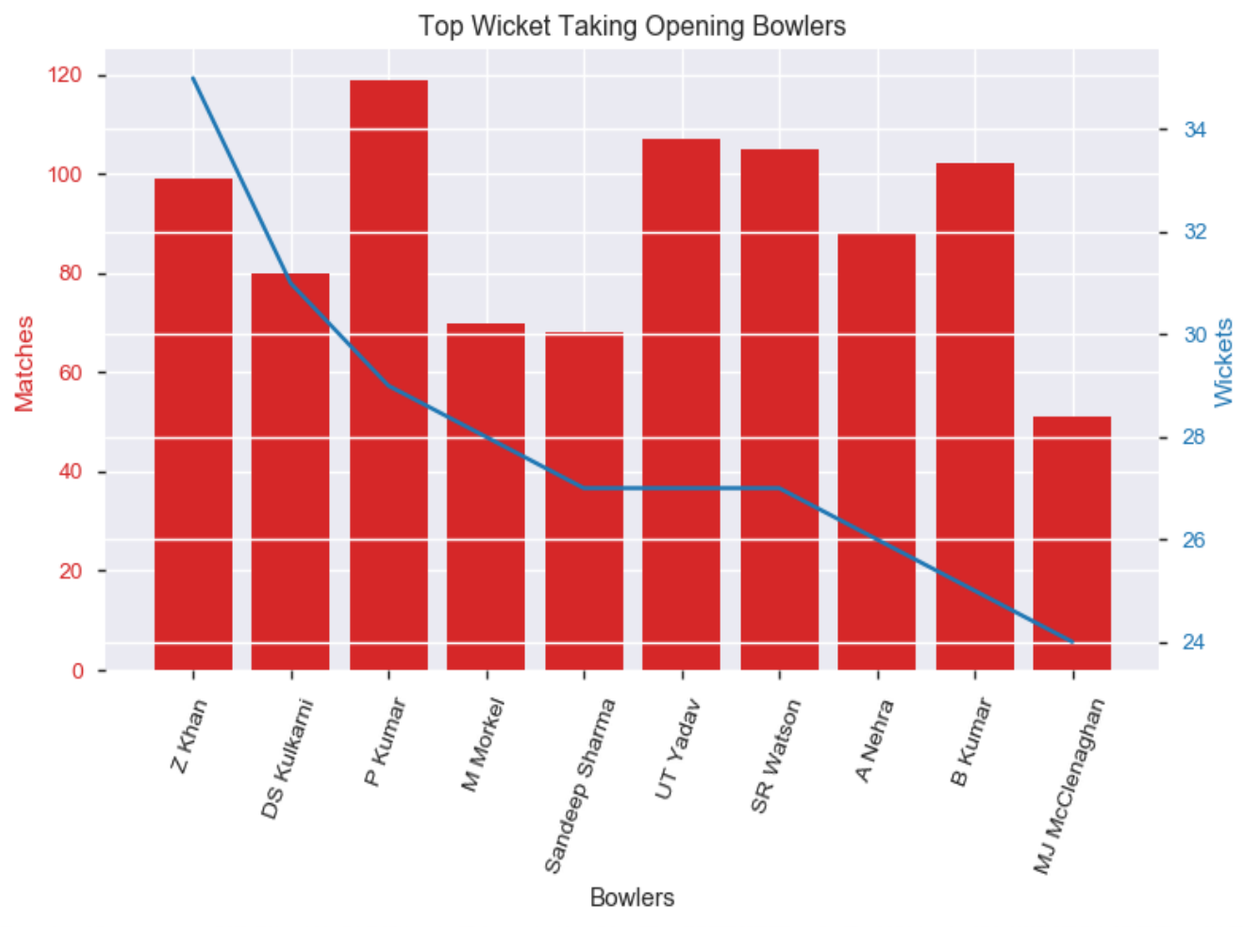


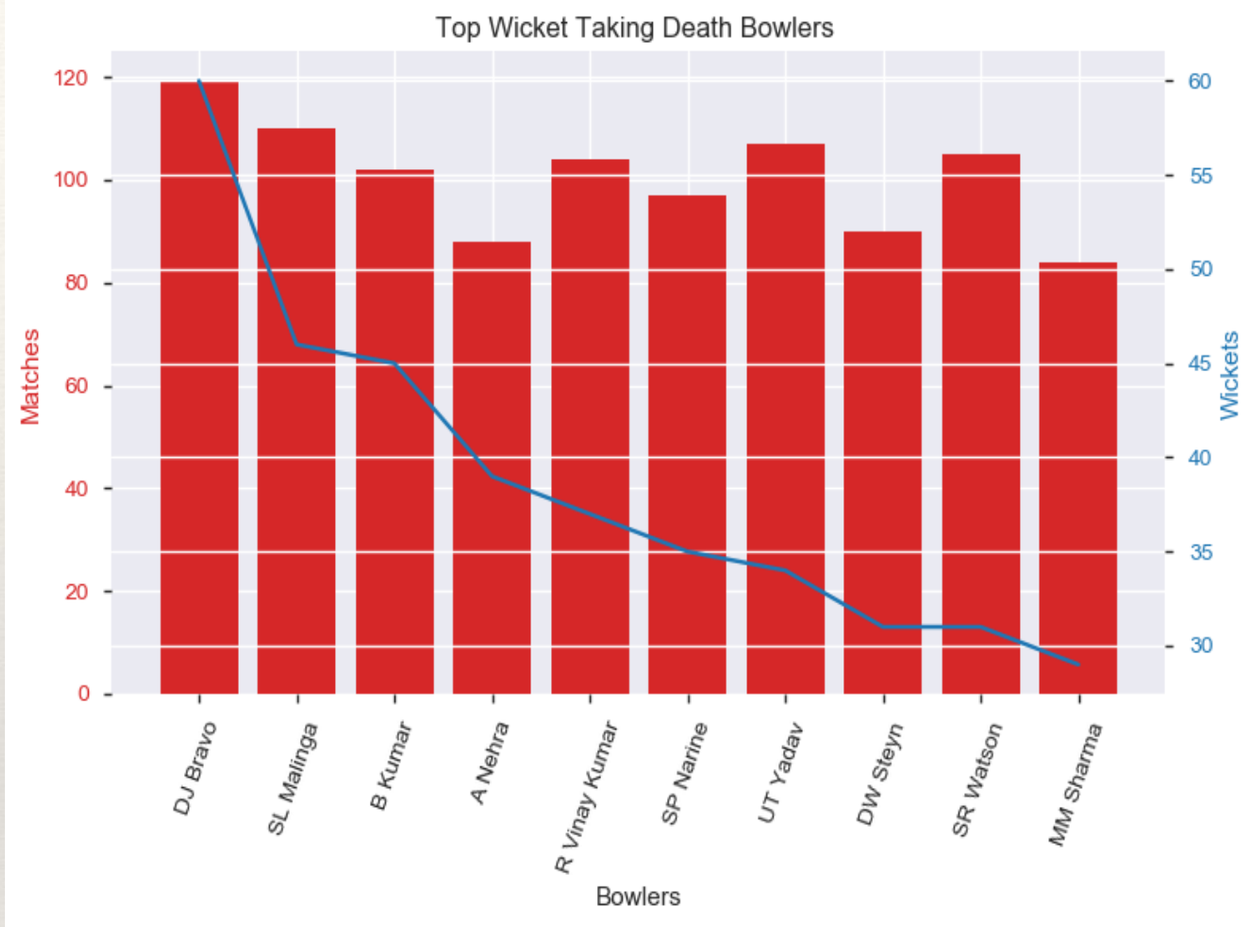




- Harbhajan Singh has bowled maximum number of overs at 530 which is way ahead of Amit at 488 and Piyush at 482.
- Malinga has the best average of 1.55 followed by Bravo and Narine at 1.3 and B Kumar at 1.25.
- Harbhajan, Piyush and Amit have played 146,143 and 136 matches respectively. Malinga has taken 170 wickets followed by Amit and Bravo at 155 each.

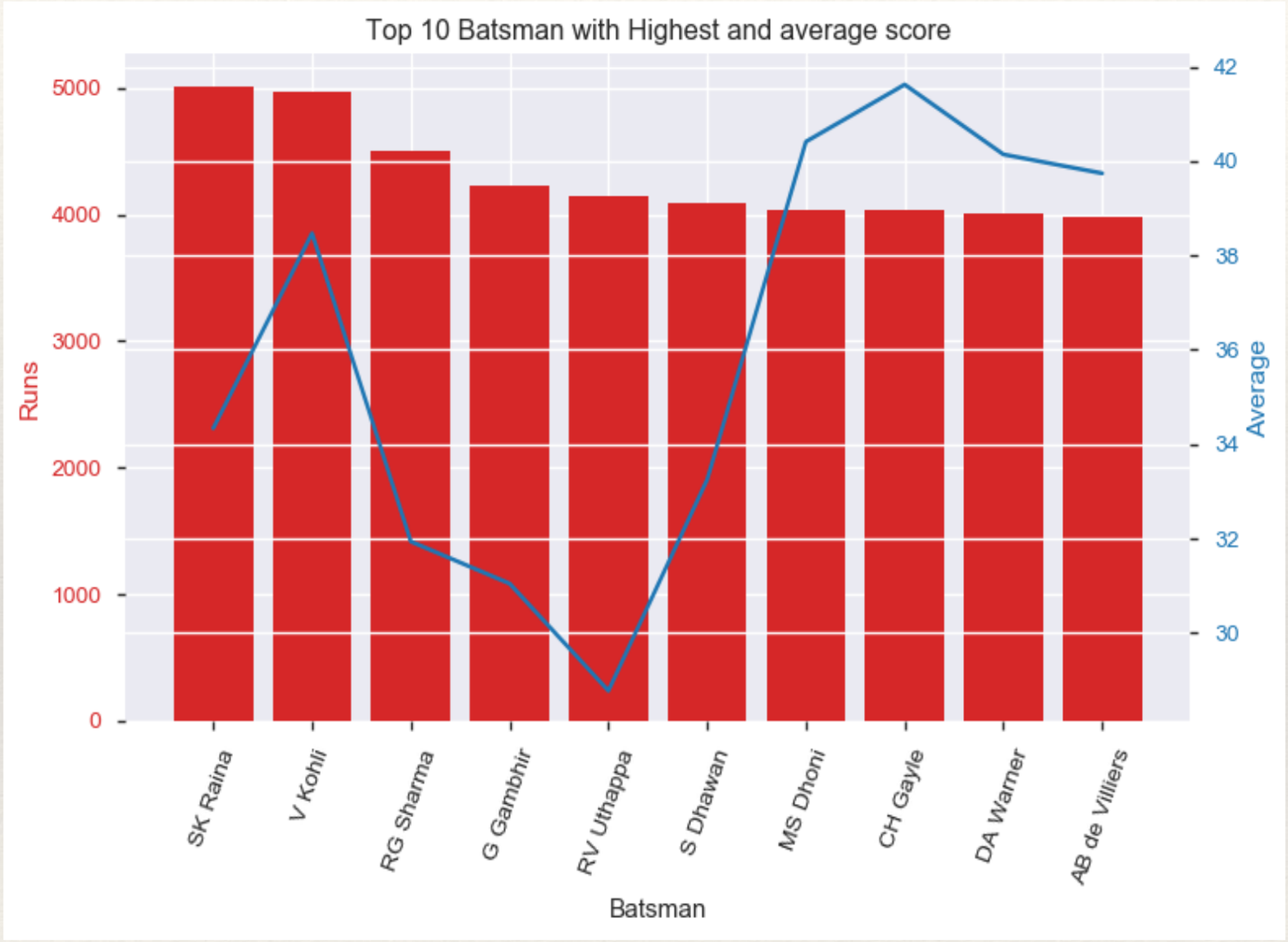
Bowling Analysis





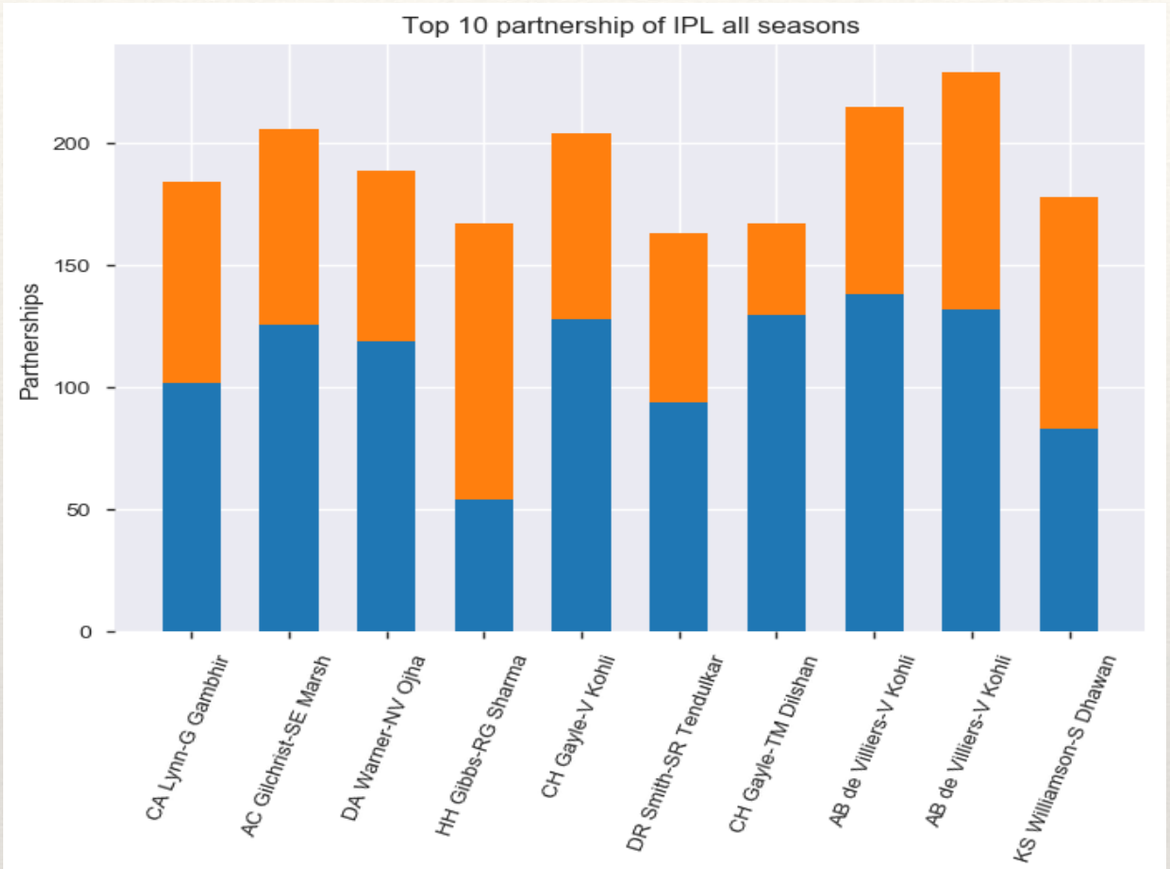
- Bowlers above in the graph are the 10 best opening bowlers. Zahir Khan has been the leading wicket taker in first 6 overs with 35 wickets in 99 matches. He is followed by Kulkarni with 31 wickets in 80 matches.
- Bowlers in right graph are best death overs bowlers. Bravo leading the chart at 60 wickets in 119 matches followed by Malinga and B Kumar.
- Umesh Yadav, Watson, Nehra, B Kumar have been leading wickets takers in both opening and death overs

Batting Analysis

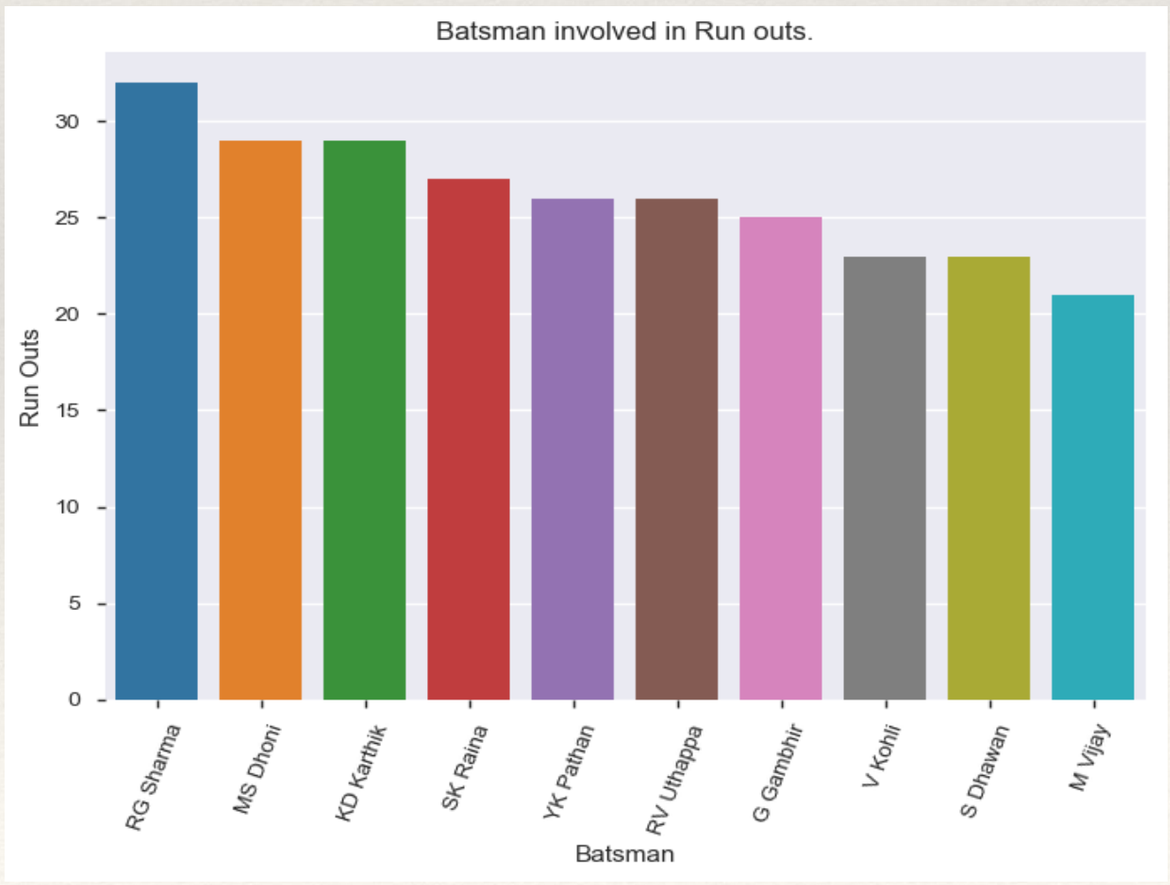


- Raina is leading run getter with over 5K runs in 172 matches followed by Kohli and Rohit.
- The players with highest batting average are Gayle, Dhoni & Warner with average above 40.
- Kohli is close to 5K runs with a strong batting average of over 38 which make him a consistent batsman.

Batting Analysis

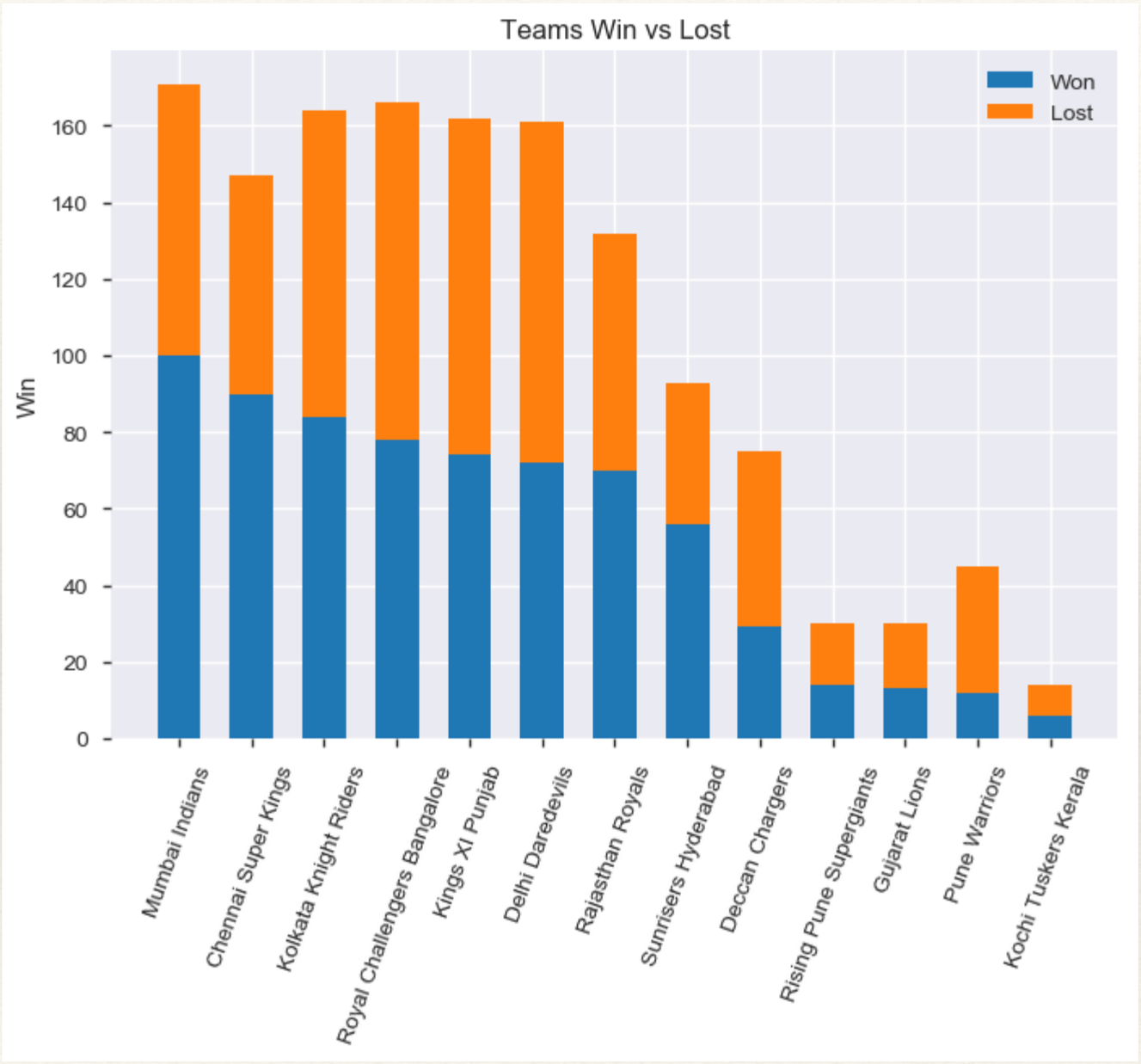


- This graph depicts the biggest partnership of all times in IPL.
- AB and Kohli had a top two biggest partnership of 229 and 215 runs followed by a partnership of 206 runs between Gilchrist and Marsh.
- Kohli featured in 3 of the top 4 partnerships. AB and Gayle were part of two partnerships each in top 10



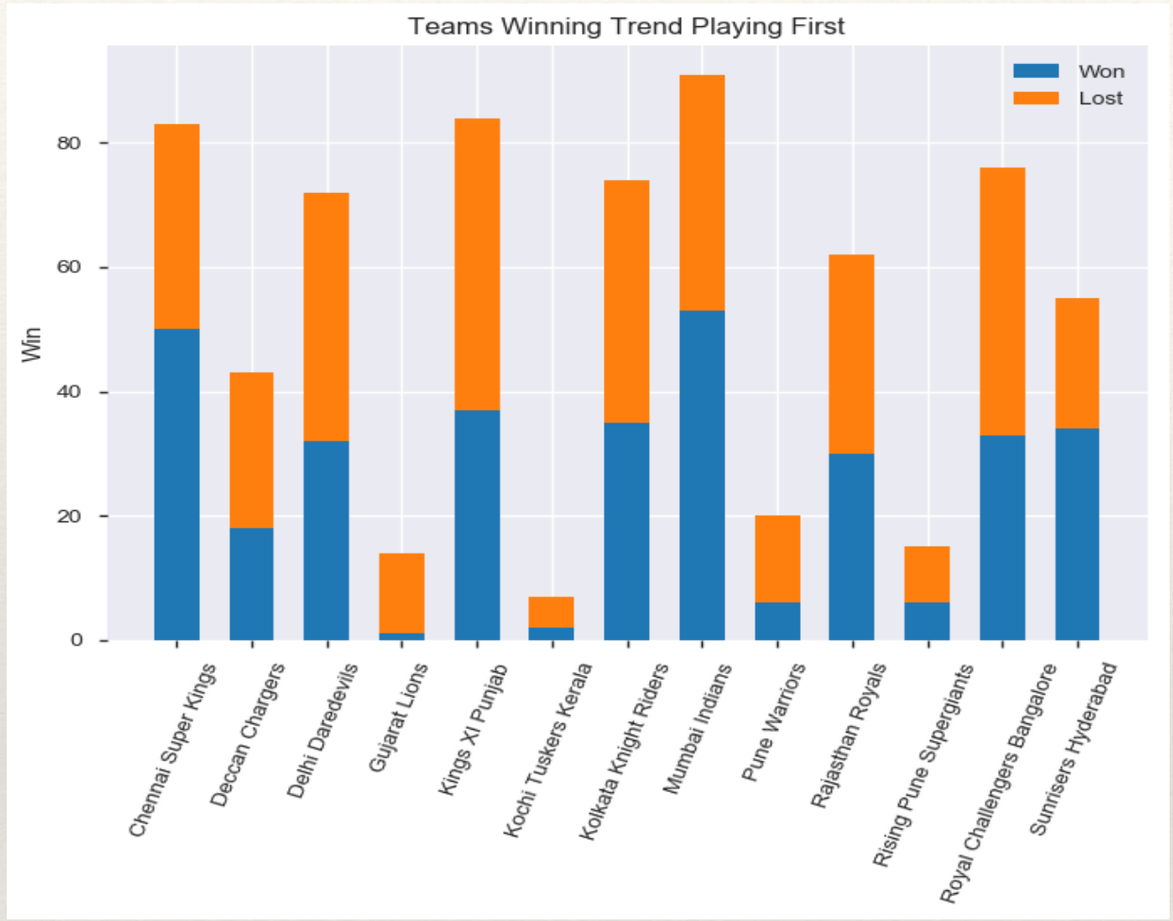
- Rohit, Dhoni, Karthik and Raina have been involved in runouts most number of times.
- Players featuring in the list need to work on their running between the wickets and other batsman have to be aware of the goof up these players have been part of, so as to be careful while sharing the pitch with them.

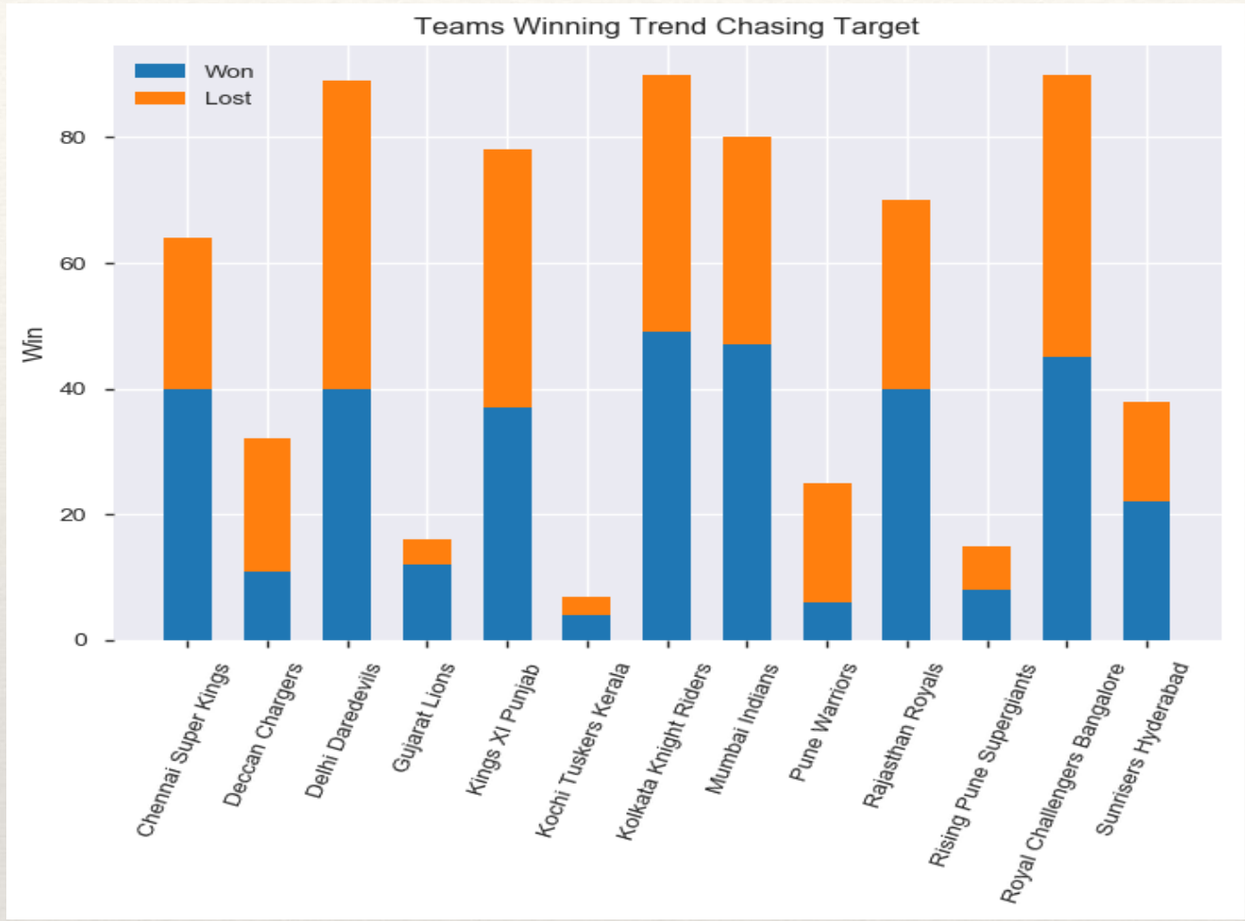
Team Performance



- Feature engineering was done to draw the result of each of the match.
- Mumbai Indian has won 100 matches out of 164 played followed by CSK and KKR.
- Winning percentage of CSK is 61% which is better than MI winning percent of 58%.
- SRH also has 60% winning percent but they have played less number of matches.
- Deccan chargers and Pune warriors have low winning percentage.

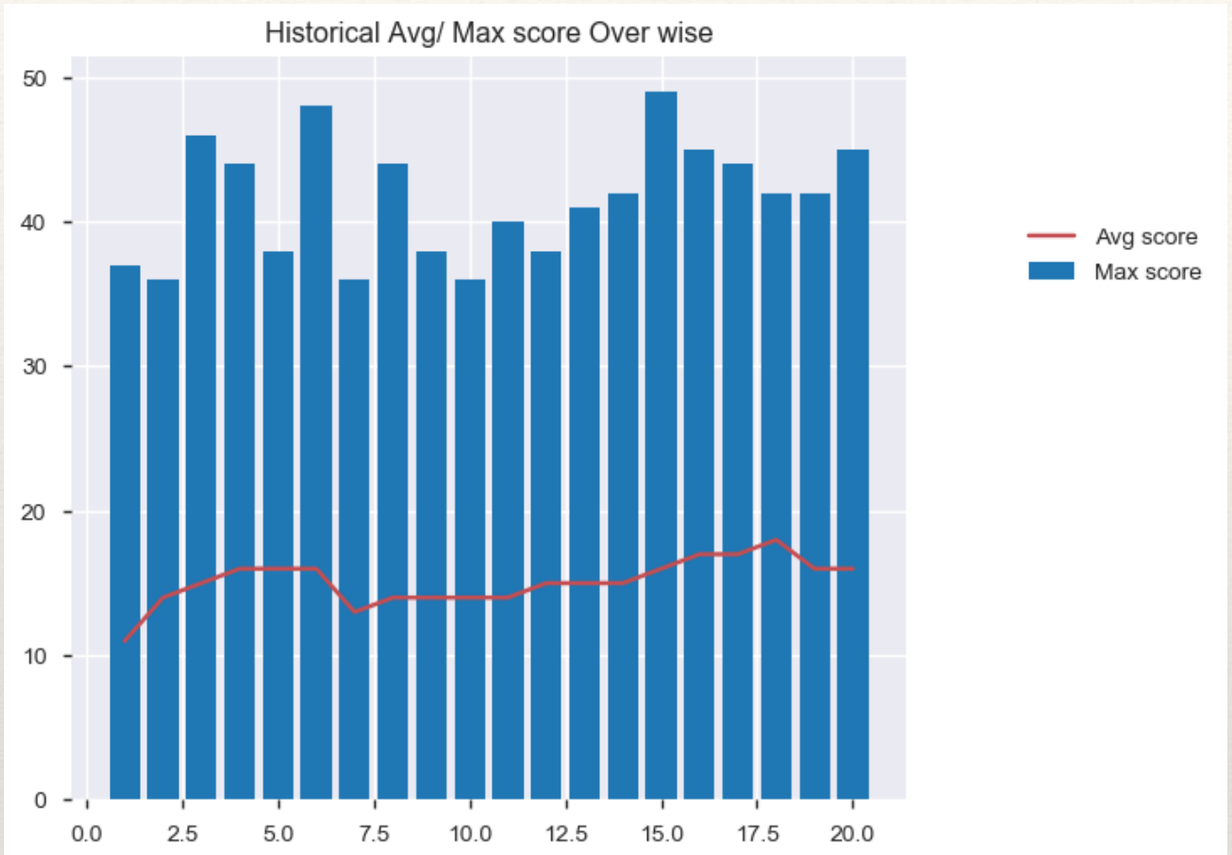
Team Performance





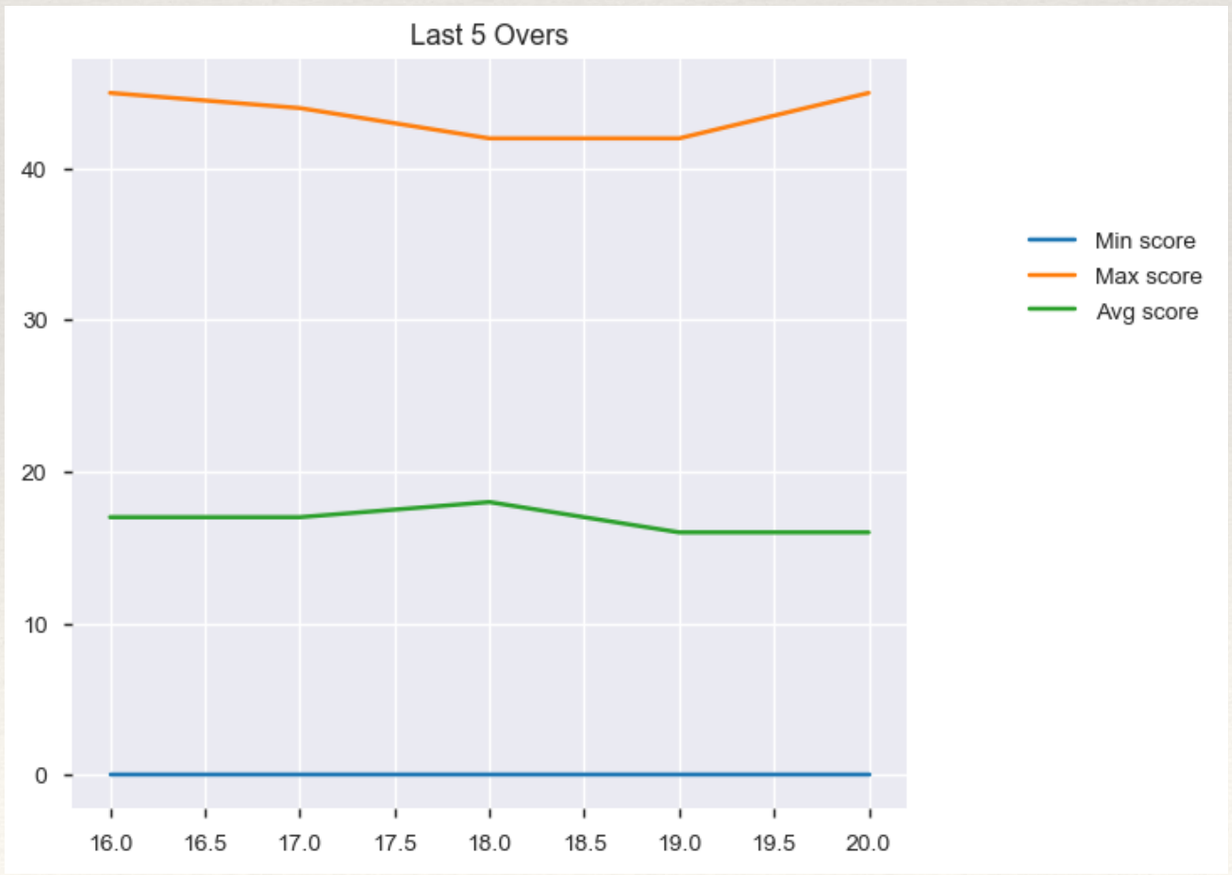
- Teams like CSK, SRH, DCH, PW have won more games setting target
- Teams like RR, KTK, KKR, RPS, RCB, KXIP have performed better chasing targets
- Teams like MI, GL & DD have performed equally good playing first or second.

Score Analysis

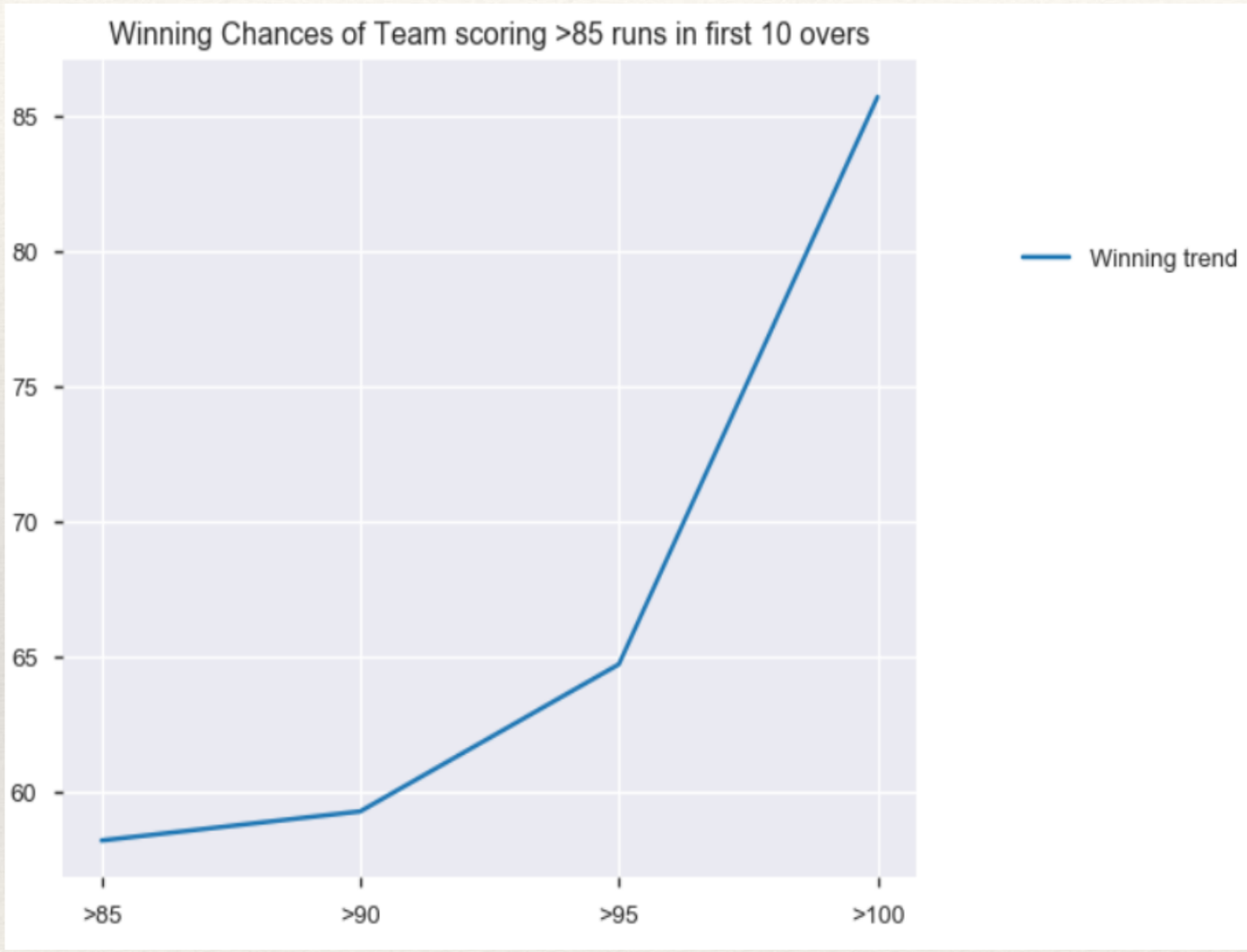


- The max runs scored in a over is 49 followed by 48 and 46 runs.
- Most of the runs are scored in last overs.
- The average runs scored in last 5 overs is relatively higher than other overs.
- It is also observed that more runs are scored in 4th, 5th and 6th over as the field restrictions are still on and batsman is set after playing first three overs.

- The graph shows that average score in last five overs average around 18 runs with a little plateau in 18th over.
- There have been times when 0 runs were scored in death overs and at times more than 40 runs were scored.



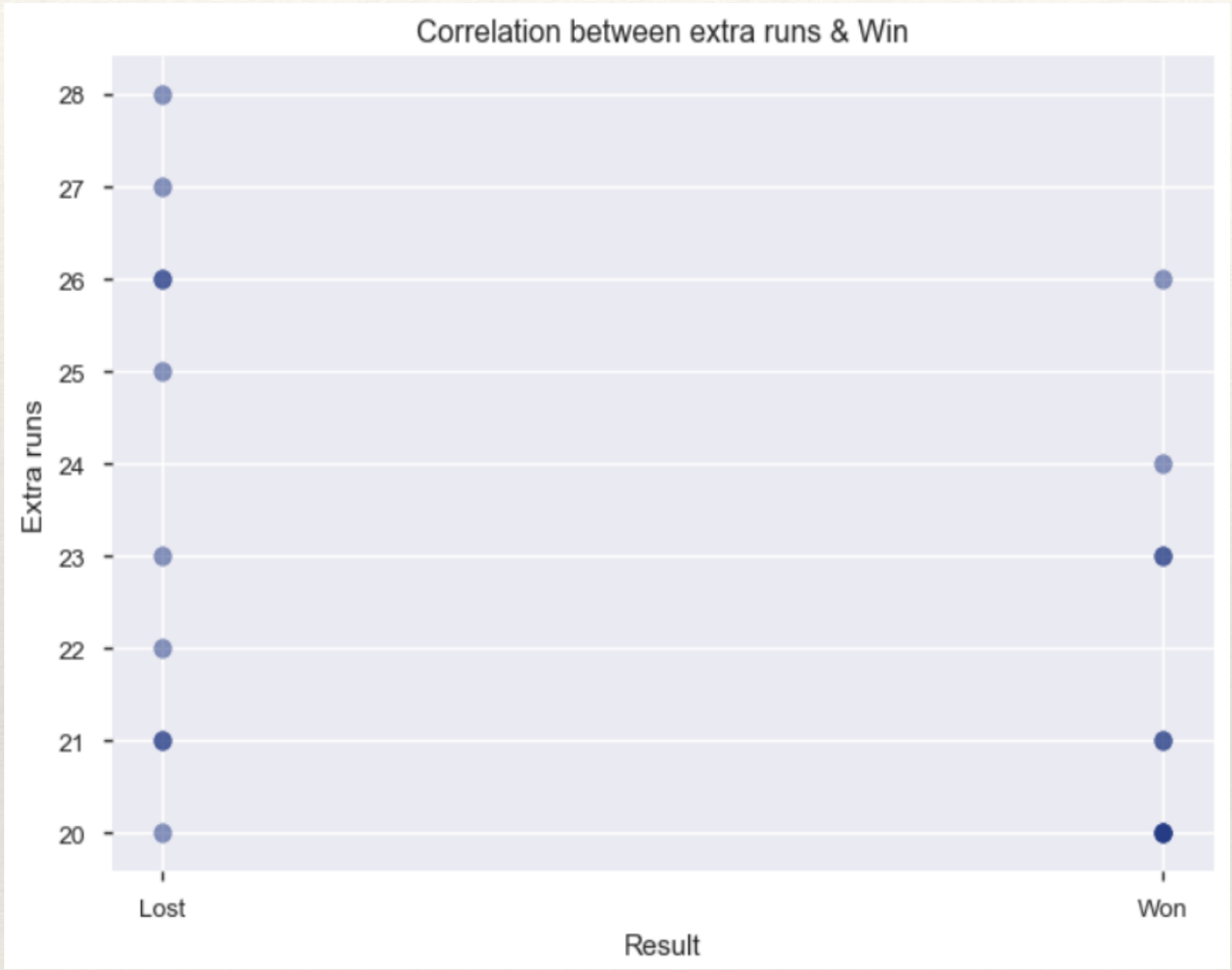
Hypothesis: Team batting first scoring >95runs in first 10 overs has high probability of winning matches.



The data from all the 600+ matches was analyzed which clearly shows the winning chances of team batting first and scoring more than 100 runs in first 10 overs have 86% chances of winning and the teams scoring between 95 and 100 have 65% chances of winning.

So higher you score in first 10 overs, more are the chances of setting up tough target for chasing team and improve the chances of winning match.

Hypothesis: Extra runs given by Fielding team is negatively impact the chances of Win.



It is evident from the above chart that higher the extra runs given, more are the chances of loosing match with very few exceptions. The extra runs gives the boost to the opponent team and they become mentally stronger to overcome the pressure.

*If both teams gave high extra runs, the team giving higher extras in a match is considered.

Conclusion

The sample used in this EDA exercise was of considerable size and was sufficient to clearly conclude on below aspects:

Bowling:

- Harbhajan and Piyush have bowled the max overs and Malinga has been the most successful bowler. Malinga has the best bowling average followed by Bravo and Narine.
- Zahir Khan and Kulkarni are the top successful opening bowlers whereas Bravo, Malinga and B Kumar are most successful death bowlers.

Batting:

- Raina is the highest runs scorer followed by Kohli and Rohit.
- Gayle, Dhoni & Warner are very consistent batsmen with batting average above 40 runs.
- AB and Kohli had a top two biggest partnership of 229 and 215 runs. Kohli featured in 3 of the top 4 partnerships. AB and Gayle were part of two partnerships each in top 10.
- Rohit, Dhoni, Karthik and Raina have been involved in highest number of runouts and need to improve their running between the wickets.

Team:

- Feature engineering was done to draw the result of each of the match.
- Mumbai Indian has won 100 matches out of 164 played followed by CSK and KKR. But winning percentage of CSK is 61% which is better than MI winning percent of 58%. SRH also has 60% winning percent but they have played less number of matches.
- CSK, SRH, DCH & PW are better setting targets than chasing
- RR, KTK, KKR, RPS, RCB & KXIP performs better chasing targets
- MI, GL & DD performs same batting first or chasing targets

Runs Scoring Pattern:

- The data shows most of the runs are scored in last overs. The average runs scored in last 5 overs is relatively higher than other overs. It is also observed that more runs are scored in 4th, 5th and 6th over as the field restrictions are still on and batsman is set after playing first three overs.

Hypothesis 1 : Team batting first scoring >95runs in first 10 overs has high probability of winning matches is true. The data exploration clearly shows the winning chances of Team batting first and scoring more than 100 runs in first 10 overs have 86% chances of winning and the teams scoring between 95 and 100 have 65% chances of winning. So higher you score in first 10 overs, more are the chances of winning.

Hypothesis 2 : Extra runs given by Fielding team negatively impacts the chances of Win stood true. It is evident from the above chart that higher the extra runs given, more are the chances of loosing match with very few exceptions. The extra runs gives the boost to the opponent team and they become mentally stronger to overcome the pressure.