

IPL 2008-2017



KAUSTUBH SINGH

Statistical Questions

1. Most teams prefer batting second when they win the toss in IPL.
2. Winning toss increases chance of winning match.
3. Batting first if win the match, they usually win by 10-20 runs of margin.
4. Batting second if win the match, they usually win by 3-4 wickets.
5. Number of 3 runs ran by any batsman in IPL is not more than 10 times.
6. Find top 3 batsman who hitted most sixes to certain bowlers.

Variables used for analysis

Match.csv

- Toss_winner - String
- Toss_Name - String
- Match_winner - String
- Team1 - String
- Team2 - String
- Win_Type - String
- Win_Margin - Float

Deliveries.csv

- batsman - String
- Batsman_runs - integer
- Bowler - String
- total_runs - integer
- noball_runs - integer
- extra_runs - integer

Variables mean

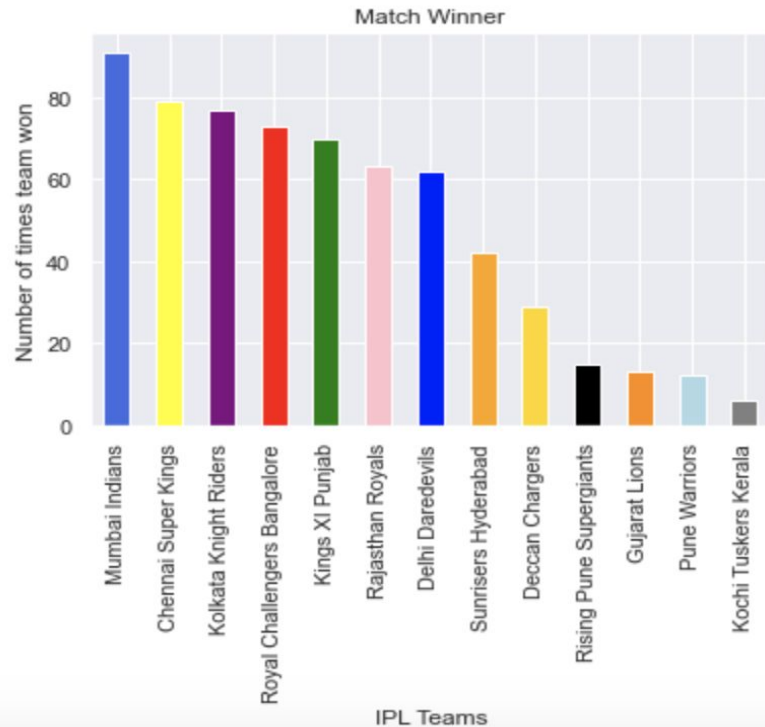
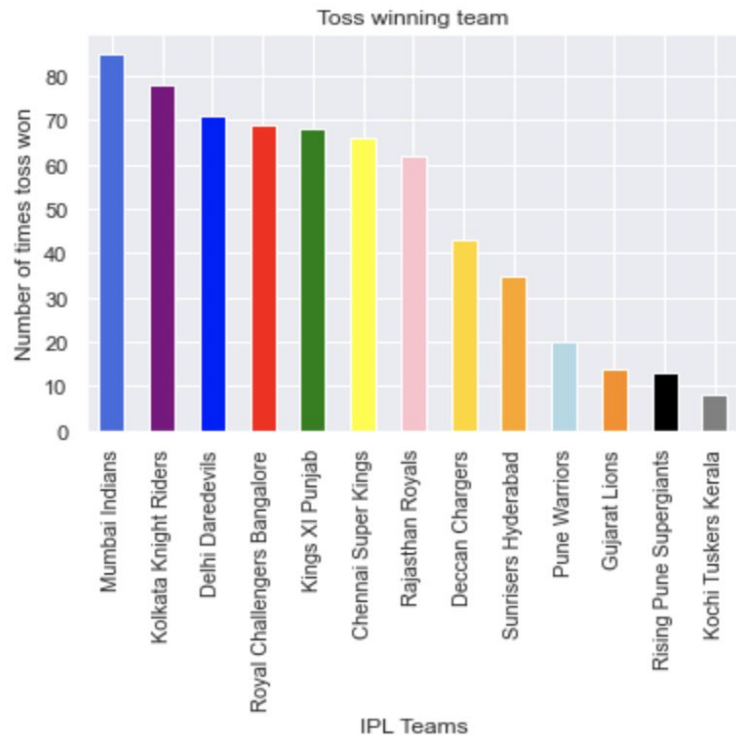
Match.csv

- Toss_winner - It will give team name who won the toss.
- Toss_Name - It will give information about whether team elected to bat or field first.
- Match_winner - It will give team name who won the match.
- Team1 - Team names
- Team2 - Team names
- Win_Type - runs/wickets gives us information about whether batting first team won or batting second.
- Win_Margin - runs/wickets

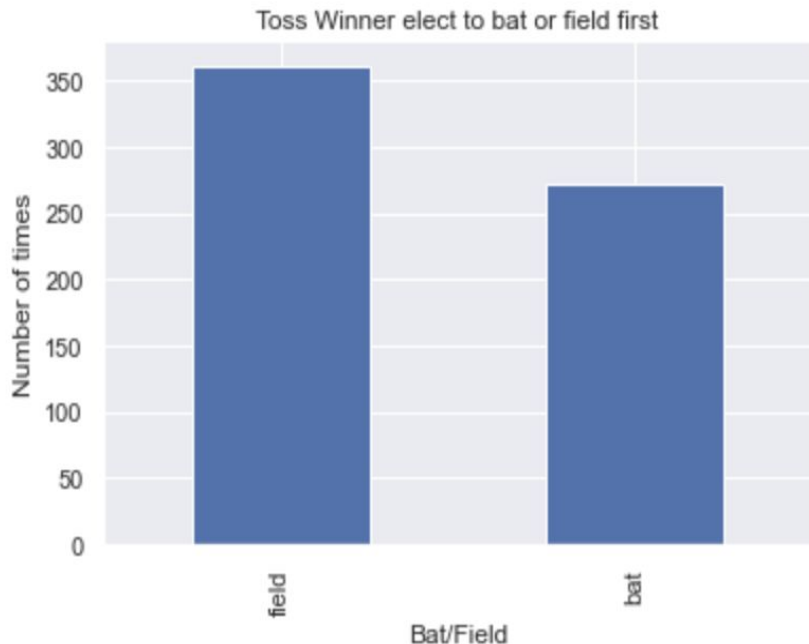
Deliveries.csv

- batsman - Batsman name
- Batsman_runs - Run scored by batsman
- Bowler - Bowler name
- total_runs - Total score by team
- noball_runs - no ball delivery
- extra_runs - extra runs

Histogram/Bar plot

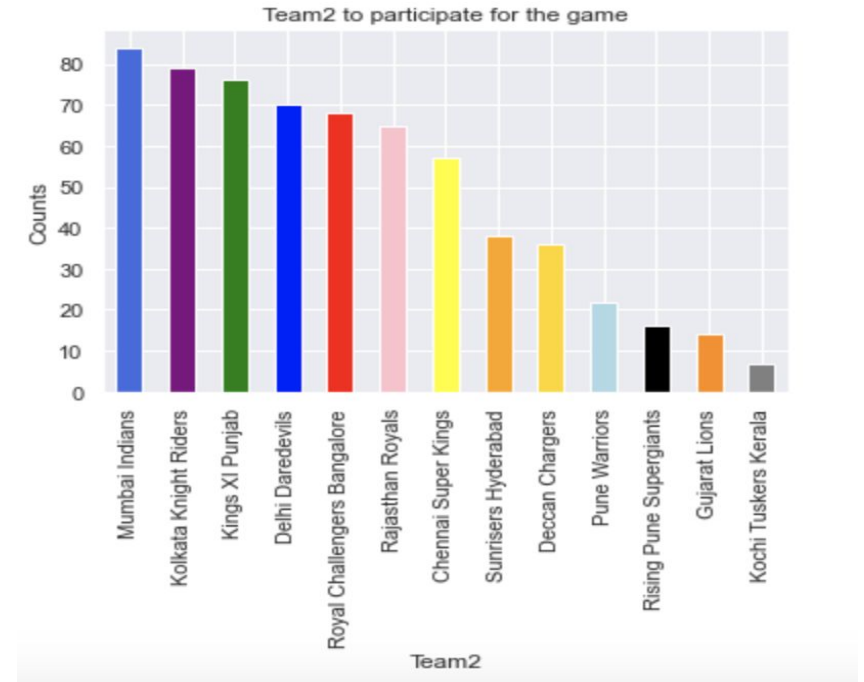
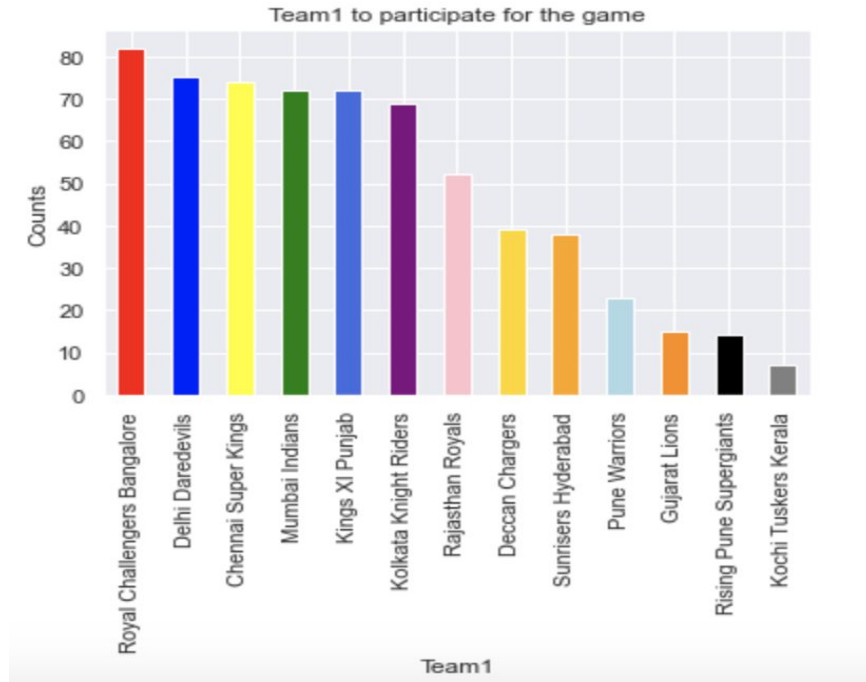


Toss Winner elect to bat or field first

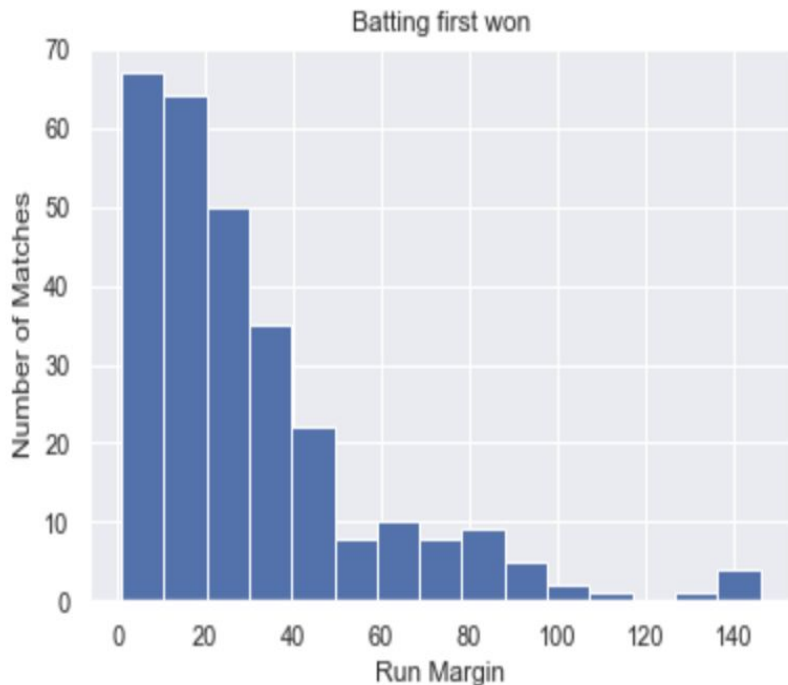


- This bar plot shows that toss winner highly elect to field first in IPL

Team1 and Team2 participate for the game

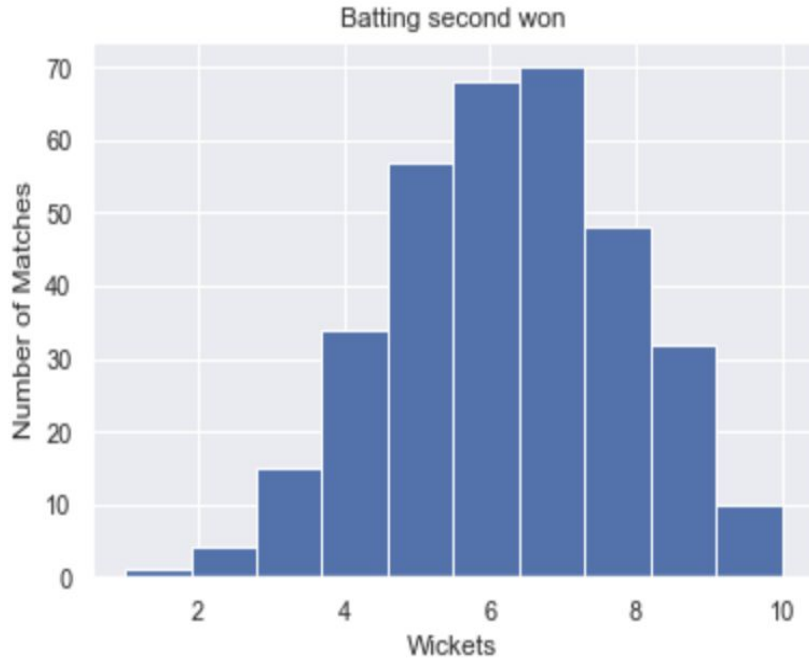


Team batted first won by margin



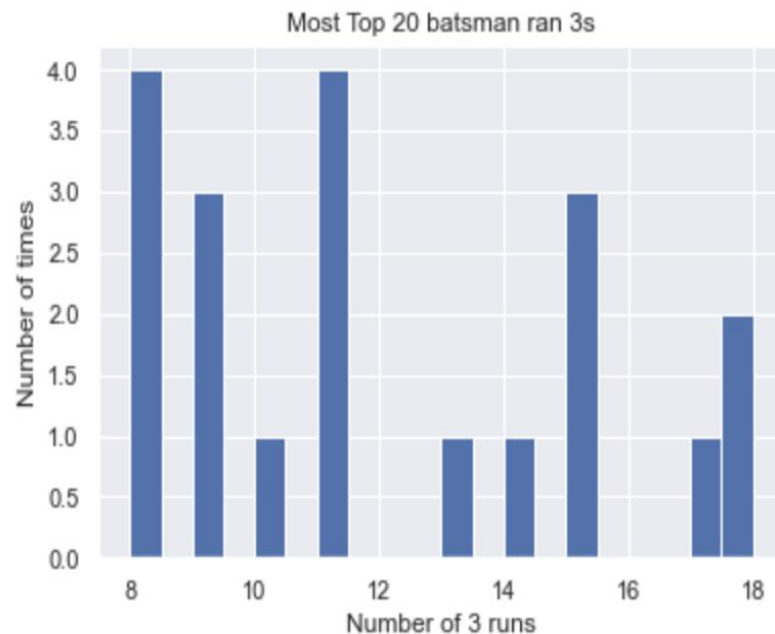
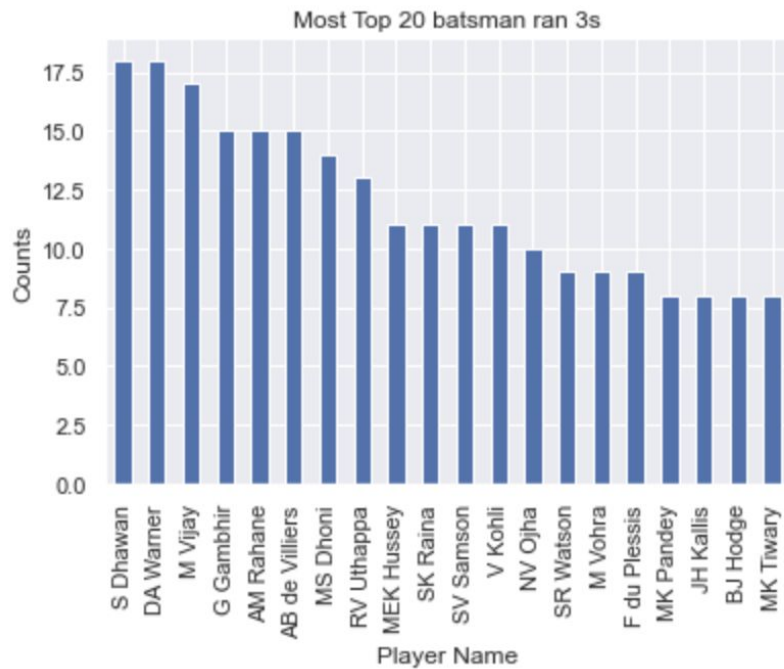
- Team batting first, if win, usually win by 1-40 runs of margin. Most like between 1 to 20.

Team batted second won by margin

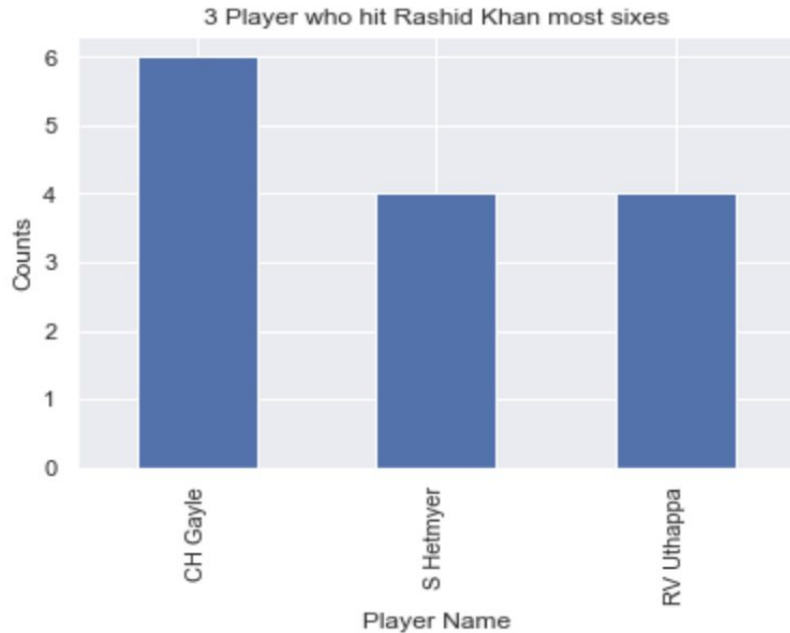


- Team batting second, if win, usually win by 6-7 wickets in hand.

Most Top 20 batsman ran 3 runs



3 Player who hit Rashid Khan most sixes



- Three batsman who hit Rashid Khan most sixes are CH Gayle, S Hetmyer and RV Uthappa.

Mean, Mode, Spread, and Tails for Variables

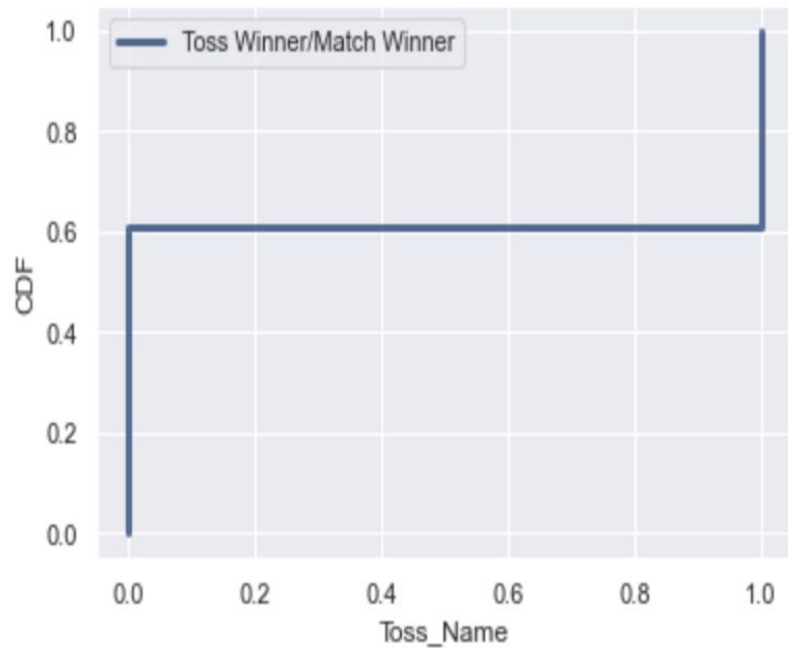
- Some of teams played very few match due to ban or not qualifying for later years. It made bar plot look going uniform and dipping down for those few teams.
- There were more than 17000 dot balls and when we look at it in terms of number of 3's or 6's it look like outlier but it is not since we are looking for data from 2008-2017 IPL and usually dot balls are higher than any other runs.
- Batting first won by more than 100 runs of margin which is possible but it is rare so we cannot consider it as outliers.

PMF



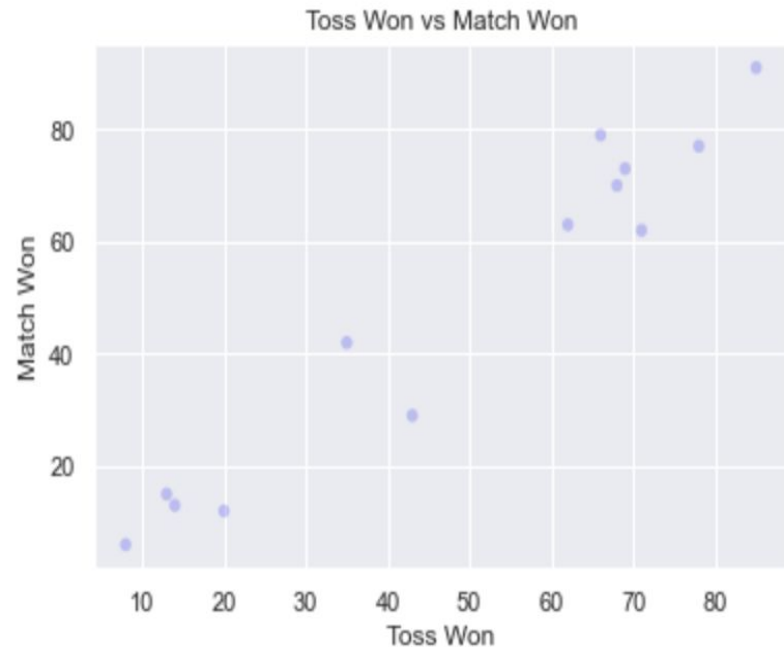
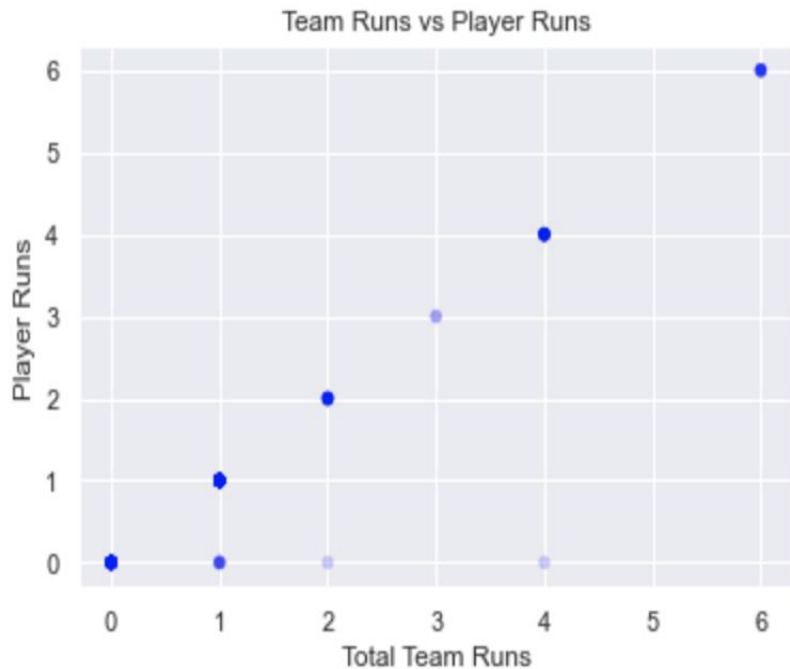
- Probability of winning match by fielding first is 61 percent and approximately 39 percent for batting first. But pitch and weather condition also plays vital role. Since this dataset does not have those we will consider above probability.

CDF

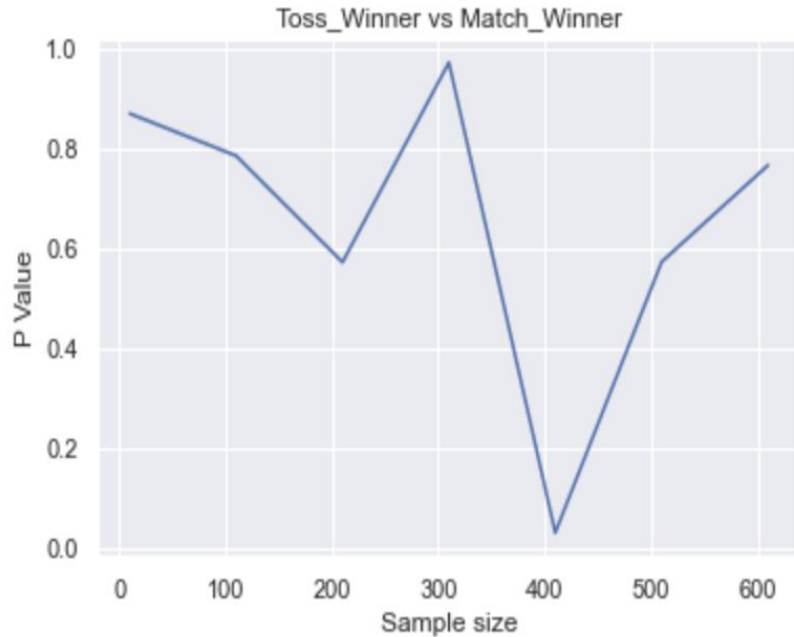


- CDF for winning toss, field or bat and winning match.

Scatter plot

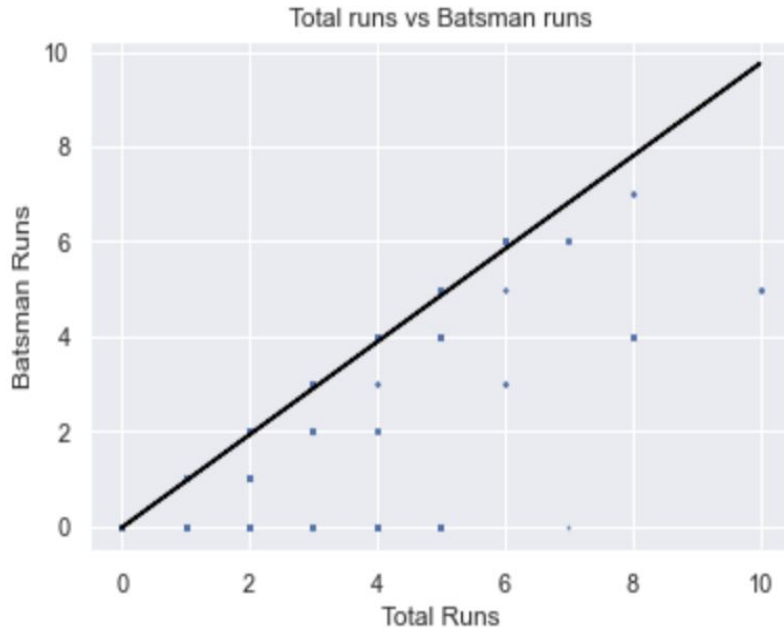


Test on hypothesis



- Test on Toss Winner and Match Winner by using P Value.

Regression Analysis



Linear Regression analysis for Total runs vs Batsman runs.

```
LinregressResult(slope=0.9790120256030593, intercept=-0.03945621184145365, rvalue=0.9772779660702506, pvalue=0.0, std  
err=0.0005017746216760767)
```

Conclusion

From 2008-2017 IPL, teams who won the toss elected to field first 361 times whereas 271 times elected to bat first. Out of 361 times they field first won 196 times and teams elected to bat first out of 271 they won 125 times.

Probability of winning match by fielding first is 61 percent and batting first is approximately 39 percent excluding all the ties.

Team batting first, if win, usually win by 1-40 runs of margin and it is very rare that they win by more than 100 runs.

Batting second, if win, usually win by 6-7 wickets in hand and rarely that they win by only one wicket left or without losing any wickets.

There were 12 batsman who ran 3 runs more than 10 times and 18 to be the most by two players.

Cont.

I created function which will take bowler name as input and shows top three batsman who hit that bowler most sixes.

I was struggling with dataset since the variables that I was working on contained string values. I wanted to do more in this project by exploring and joining other dataset together but due to time restriction was only able to do this much.

There was no assumption that made me felt incorrect.

Main challenge that I faced was with string values inside the columns which made comparison hard for me. I will have to research more about this and continue working on this project.