Indian Premier League (IPL)

-Mithilesh Vaidya

Abstract

- The dataset, named ‘IPL’, has been downloaded from Kaggle. It contains ball-by-ball details of all 10 seasons of the popular cricket league, ‘IPL’.
- After extracting relevant information, I have put together a number of statistics and findings. To name a few:

1. Most runs/wickets
2. Should you bat first on winning the toss? In case of rain?
3. Most effective batsmen/bowlers in the death overs (Overs 16-20)

- Firstly, I framed questions which I felt would provide some interesting insight. Then, using all the libraries taught in this course, I imported the data and wrote the required code to discover some interesting answers to the framed questions!

Motivation

Being an avid cricket fan, right from my childhood, I have always been fascinated by IPL. The answers I have come up with are purely statistical results which may help the coach/captain make better, informed decisions throughout the course of the season.

eg. individual stats like ‘Economy Rate’ in death overs can guide the captain in handing over the ball to the team’s best death overs bowler in the last 5 overs.

For all the cricket enthusiasts, the revelations are intriguing and answer some common questions which we all ponder upon!

e.g. Does DL method favour the bowling team?

Dataset

- I downloaded the ‘IPL’ dataset from Kaggle.
- It contained two files: ‘matches.csv’ contained data which helped me answer team statistics while ‘deliveries.csv’ provided insight into individual stats and records.
- It has ball-by-ball data of 150,460 deliveries and the match statistics of a total of 636 matches spread across 10 seasons.
- Ball-by-ball data contained information about the batsman, bowler, runs scored, dismissal type, etc.
- Match data contained data about the venue, umpire, man of the match, etc.

Data Preparation and Cleaning

- After importing the data, I dropped a few columns which I felt would not provide any interesting insight. I also dropped those rows corresponding to matches which were washed out due to rain.
- Here are some problems I faced:

1. Since a few teams had multiple home grounds across various seasons, I had to aggregate them together for answering questions related to home advantage. Also, since the 2009 edition was held in South Africa, there was no question of home crowd. So, I dropped the season for calculating the home wins vs away wins statistics.
2. A few teams have been dissolved while new ones have come up throughout the 10-year history of IPL but the home stadium did not change. As a result, I had to split the dataset and do the calculations individually.

Research Questions

I have come up with answers to multiple questions:

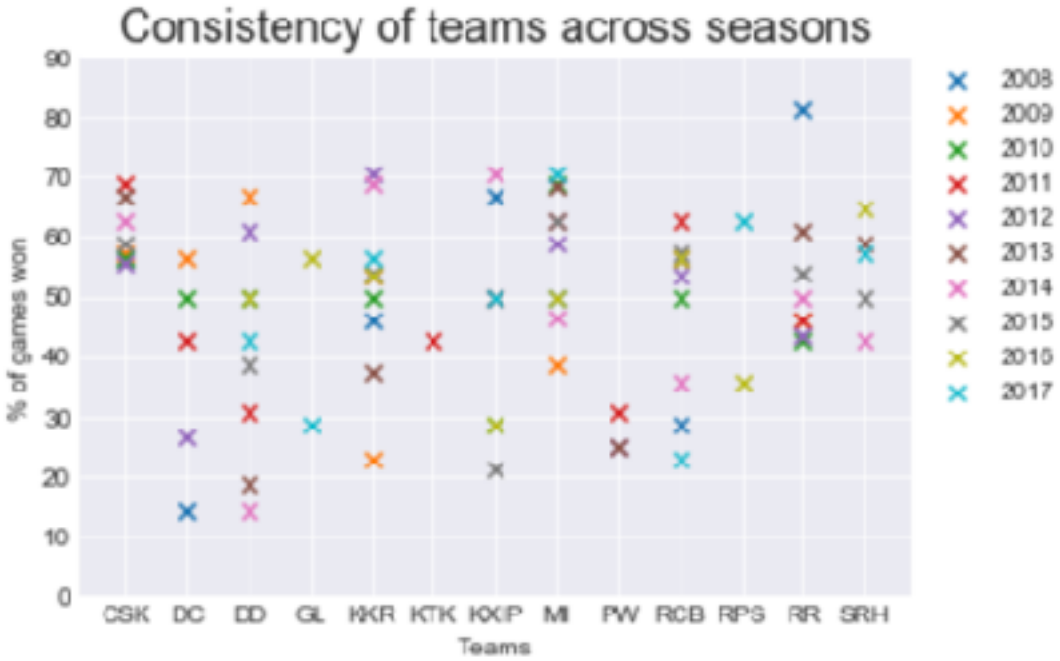
- Is DL method biased towards the field-first team?
- Does winning the toss influence the outcome of the game?
- Being a game which has favoured batsman over the last few years, is batting first a better option than bowling first?
- Who has the most runs/wickets/man of the match awards, aggregated over all the 10 seasons?
- Which batsman/bowler has performed well in the death overs?

Methods

- After importing the data, I took the required slices of the main dataset, required for answering a particular question.
- Then, I cleaned it up and arranged it in ascending/descending order based on the values in a particular column of interest.
- After obtaining the required data, I plotted pie/bar/line charts, depending on the type of information, using matplotlib.
- The graphs have been scaled and the y-axis limits are chosen appropriately, so as to gain more insight.

Findings

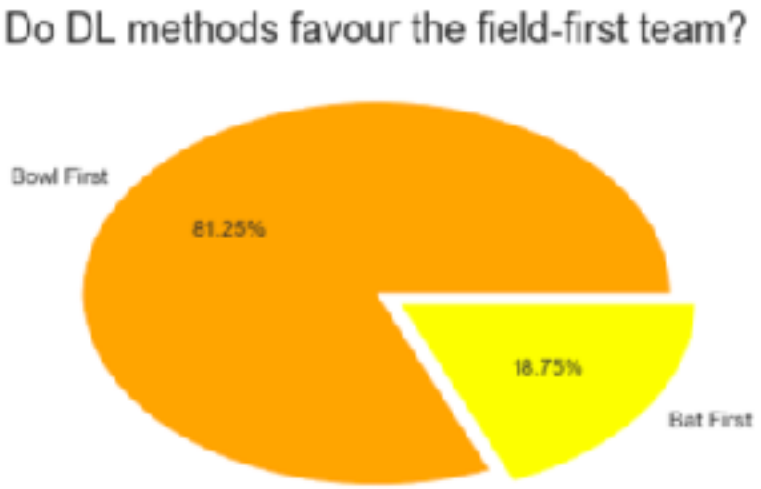
Which team has been consistent throughout the 10-year history of IPL?



As we can observe, CSK has been the most consistent while teams like DD and KXIP have had their share of ups and downs.

Findings

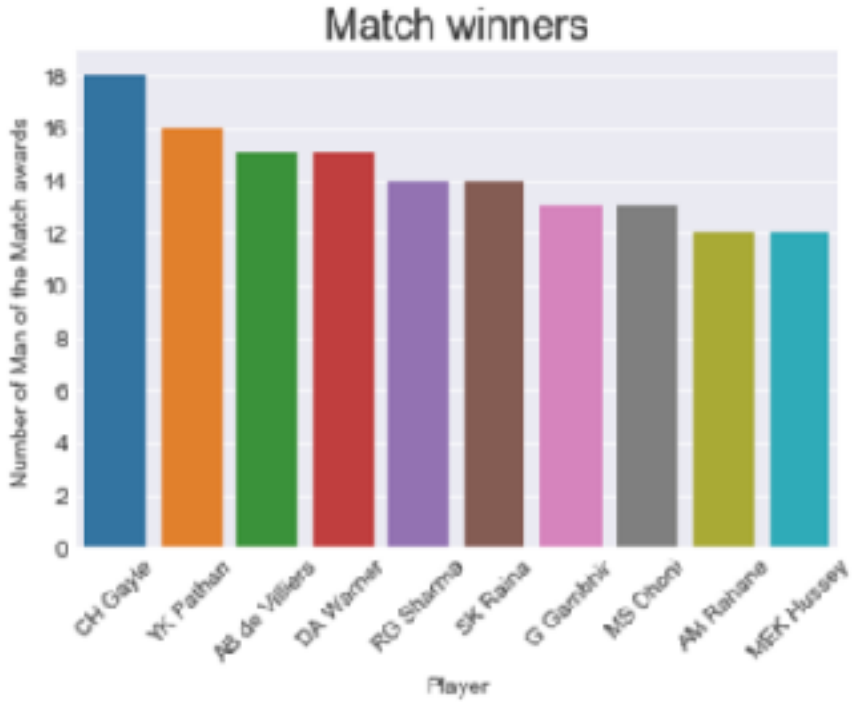
In rain-affected games, which team is more likely to win? The one which bats first or the one which bowls first?



The pie chart clearly indicates that the team which fields first wins majority of the games affected by rain. Maybe its time to revisit the Duckworth-Lewis rules and make appropriate changes.

Findings

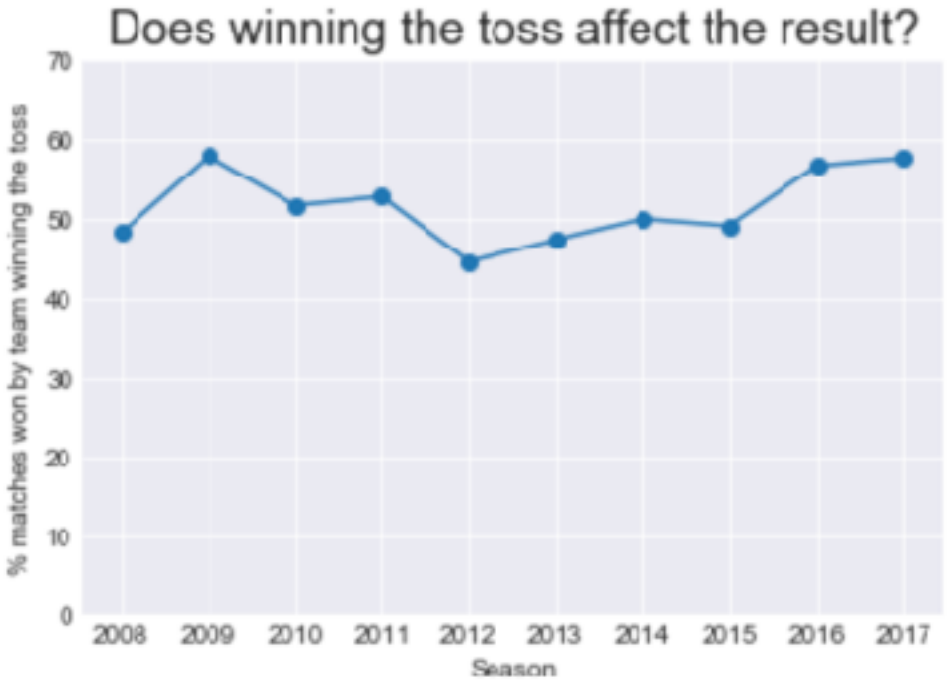
Who has won the maximum Man-of-the-Match awards?



The ‘Universal Boss’ Chris Gayle has been a match-winner in a record 18 games. Other big names like AB de Villiers and David Warner have also won many matches for their sides!

(Some more)Findings

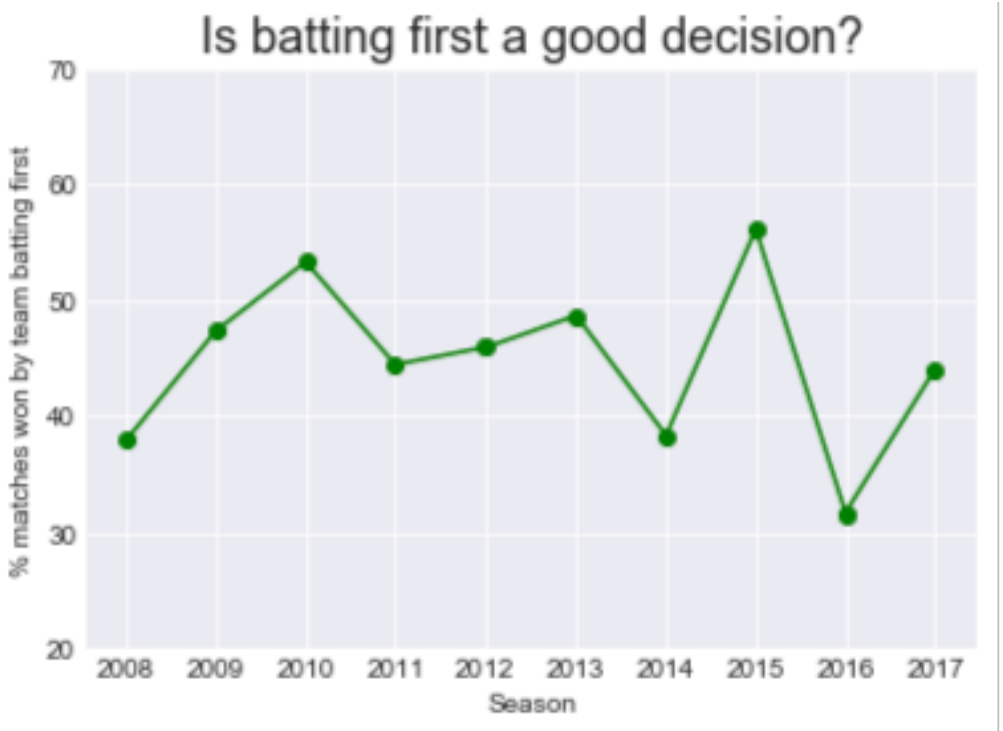
Should you be happy when your team wins the toss?



Not really. Numbers say that there is approximately a 50-50 chance of winning the game after winning the toss. Sounds counter-intuitive, doesn't it?

(Some more)Findings

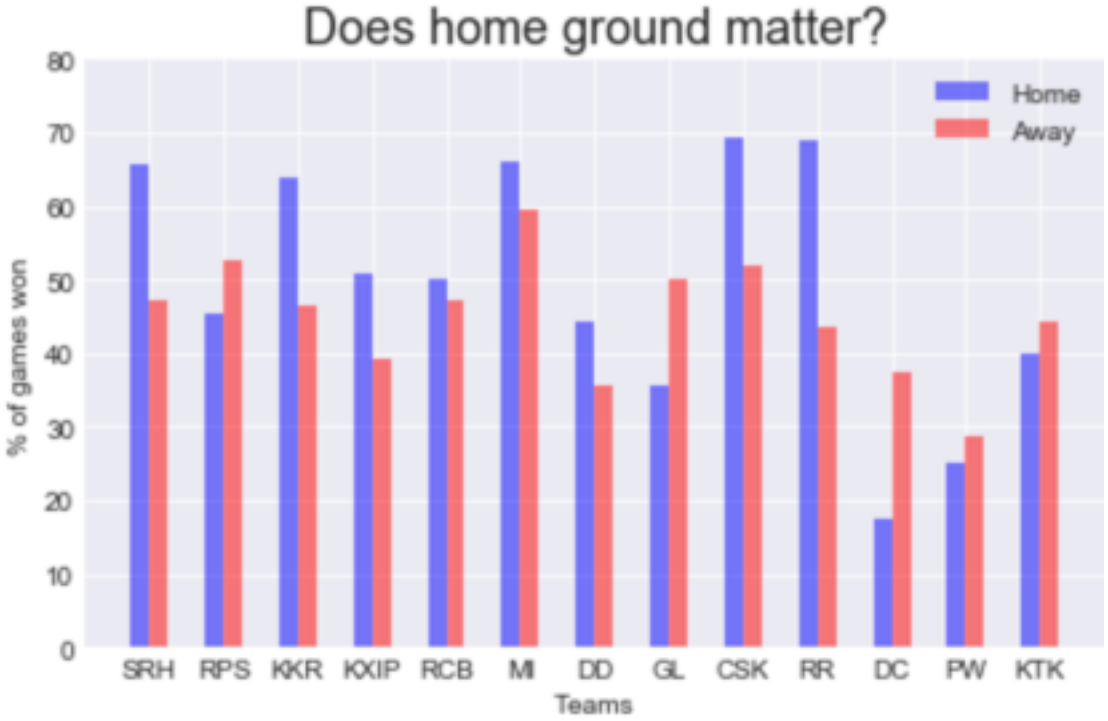
Say you win the toss. Should you bat?



As we average out the results over all the seasons, it does not make much of a difference. Probabilities of successfully defending and chasing the target are almost equal.

(Some more)Findings

How much difference does the home venue make?

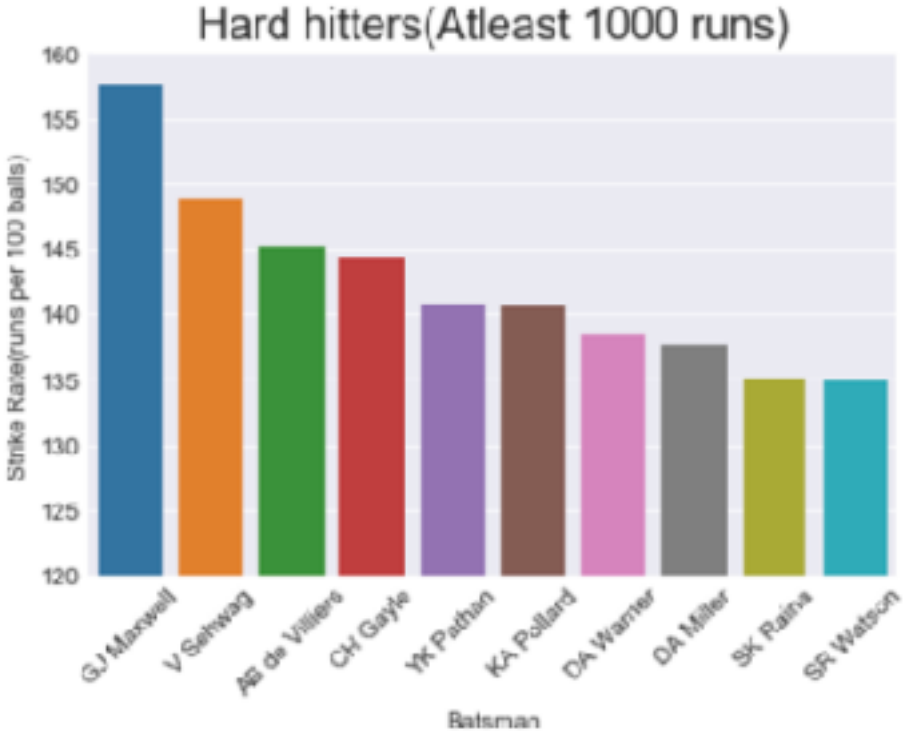


As we can see, the presence of home crowd is an important factor in determining the result.

Note that the 2009 edition has not been considered for this data because it was held in South Africa, where no team had a home advantage!

(Some more)Findings

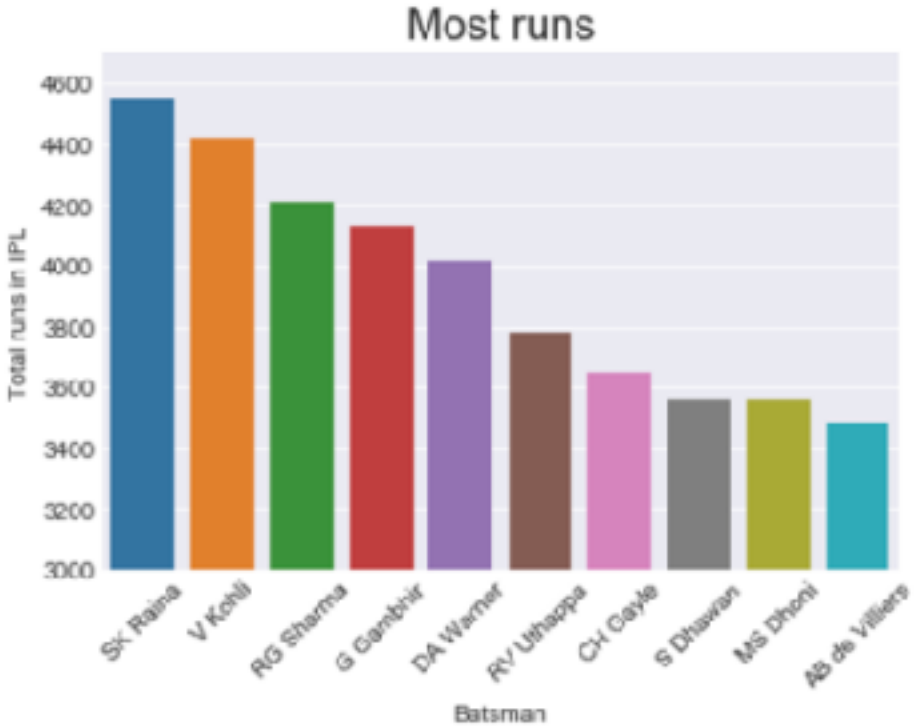
Who are some of the most ferocious batsmen?



All the familiar names like Maxwell, Sehwag and de Villiers have made it to the list of batsman with the highest strike rate (Runs per 100 balls faced)!

Findings(Please bear with me)

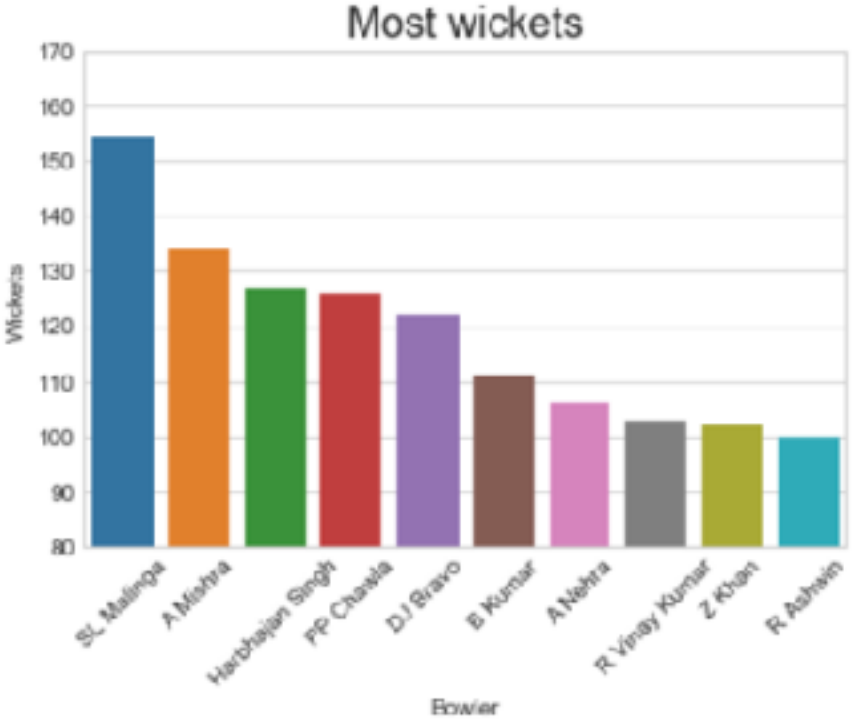
Here is a list of the batsmen with most runs (all seasons combined)



Raina and Kohli have been the backbone of their respective teams while the inclusion of MS Dhoni is astonishing as he comes out to bat down the order! (generally 6 down)

Findings(Please bear with me)

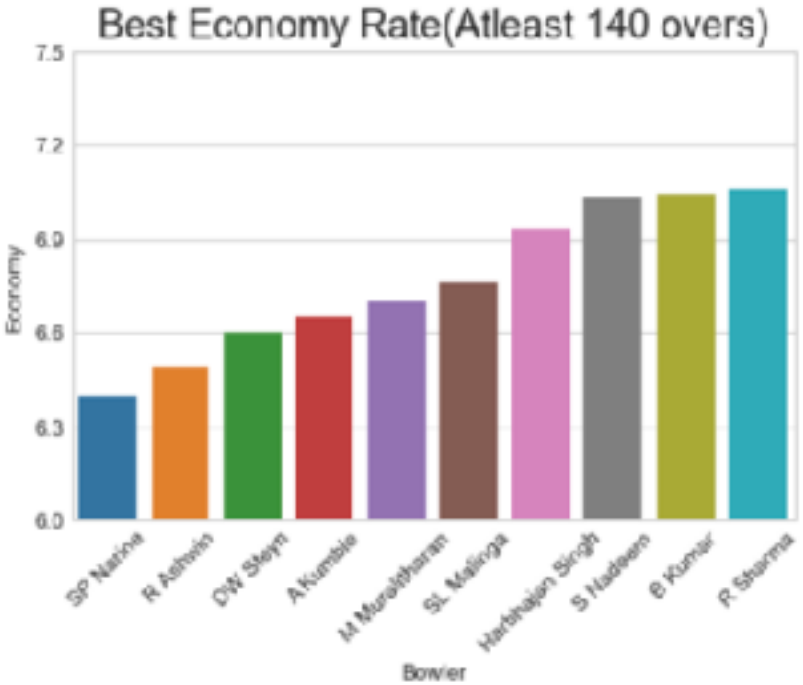
Enough of batsmen stats. Let's talk about bowlers now.



'Slinga Malinga' tops the list, with a lead of almost 20 wickets! An interesting observation is that both spinners and fast bowlers have made it to the list(I haven't explicitly denoted the type of bowler, but I've been watching them since I was a kid. So trust me!). It clearly suggests that a good balance of both ensures a healthy bowling attack!

Findings(Please bear with me)

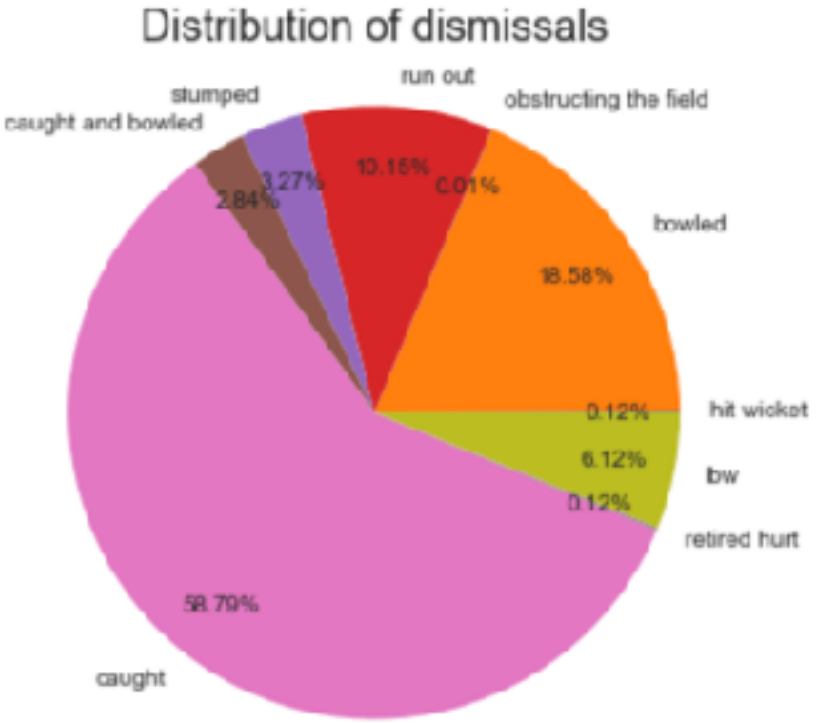
Since T20 is fast-paced, economy is as important as number of wickets



The chart is dominated by spinners and so when the captain needs to slow down the batting team and break their momentum, it seems that spinners are a better option than fast bowlers!

Findings(Please bear with me)

What are the chances of getting bowled vs getting caught? LBW? Have a look:



You probably saw that coming. In an era where blindly swinging the bat has become common, most batsman get caught. The tournament has also seen a handful of rare dismissals such as ‘obstructing the field’ and ‘hit wicket’.

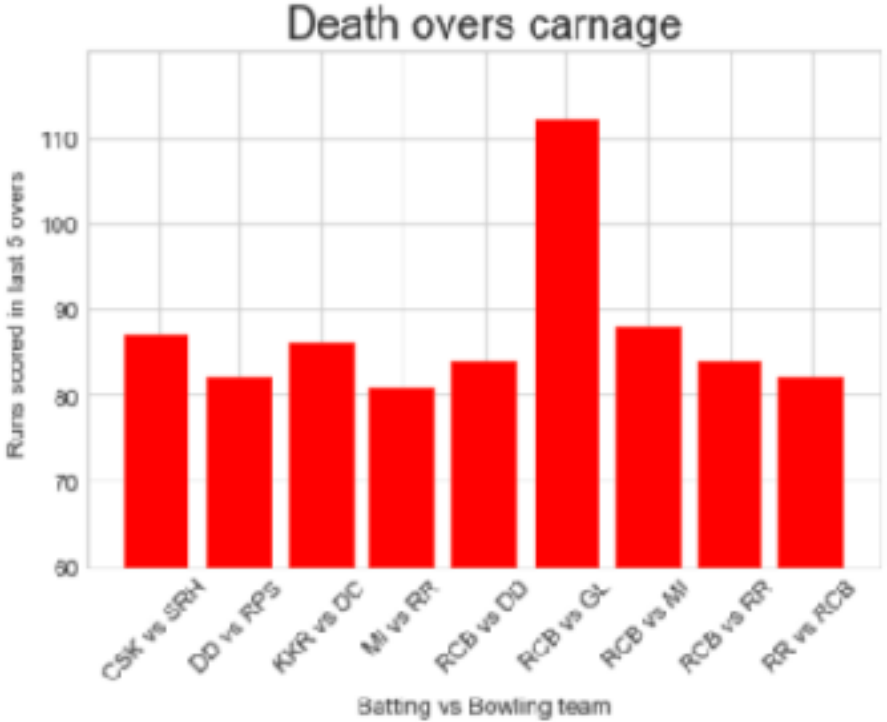
(Some more)Findings

Let’s talk about mammoth totals.

Looks like CSK, KXIP and RCB have hard-hitting batsmen in their team, which have helped them set up daunting totals for the opponent!

(Some more)Findings

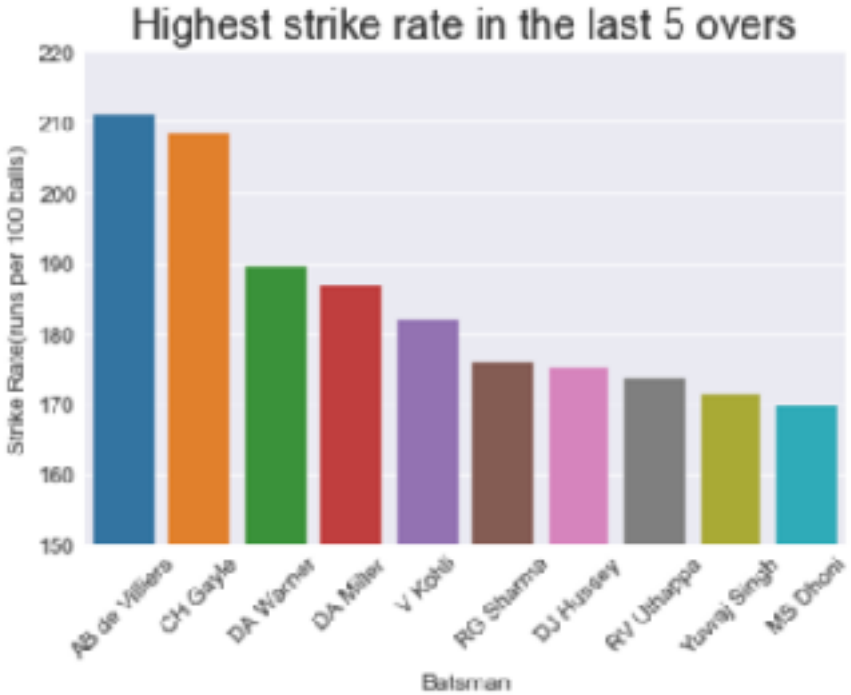
I personally enjoy watching the last 5 overs of an innings. Here's why:



As you might have guessed, the death overs are the most exciting as the batsman cut loose and swing their bat at almost every ball. I feel sorry for the bowlers who get smacked out of the park!

(Some more)Findings

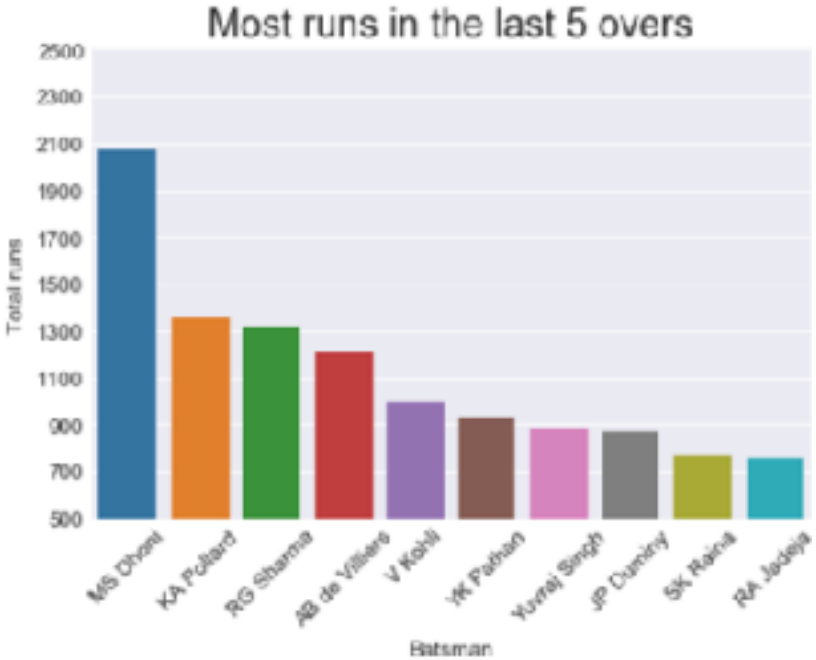
And which batsman are responsible for that previous graph?



As a captain, I would desperately try and get ABD, Gayle or Warner out in the first 15 overs. Otherwise, my bowlers are going to get smoked!

(Some more)Findings

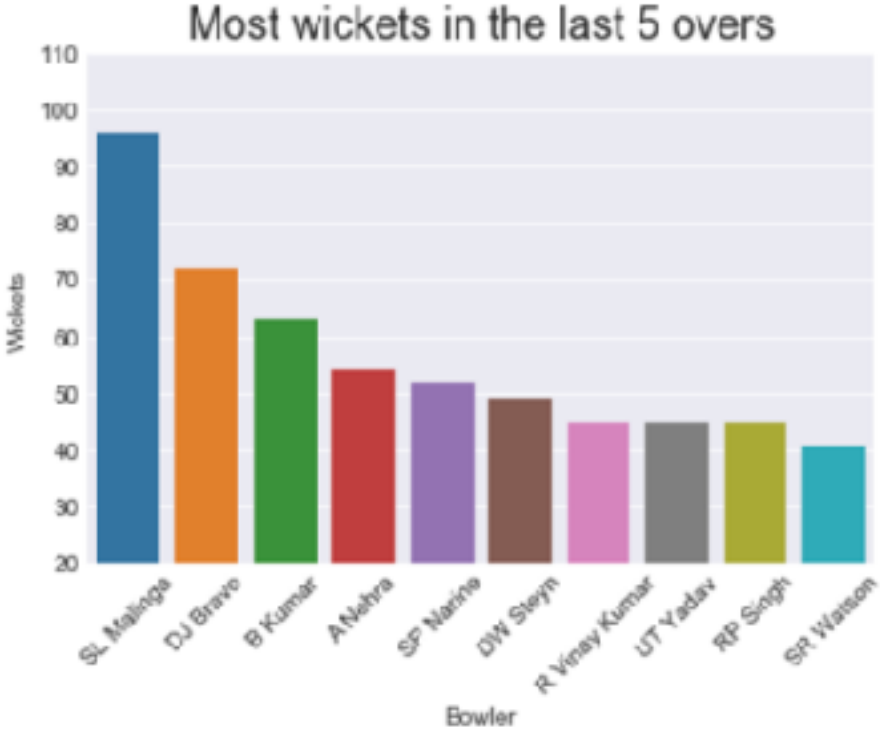
Which batsmen have contributed the most to the total in the death overs?



It had to be MS Dhoni! There is a reason why they call him the ‘Best Finisher’ in the world! The perfect graph to shut ‘Dhoni Haters’ up!

(Some more)Findings

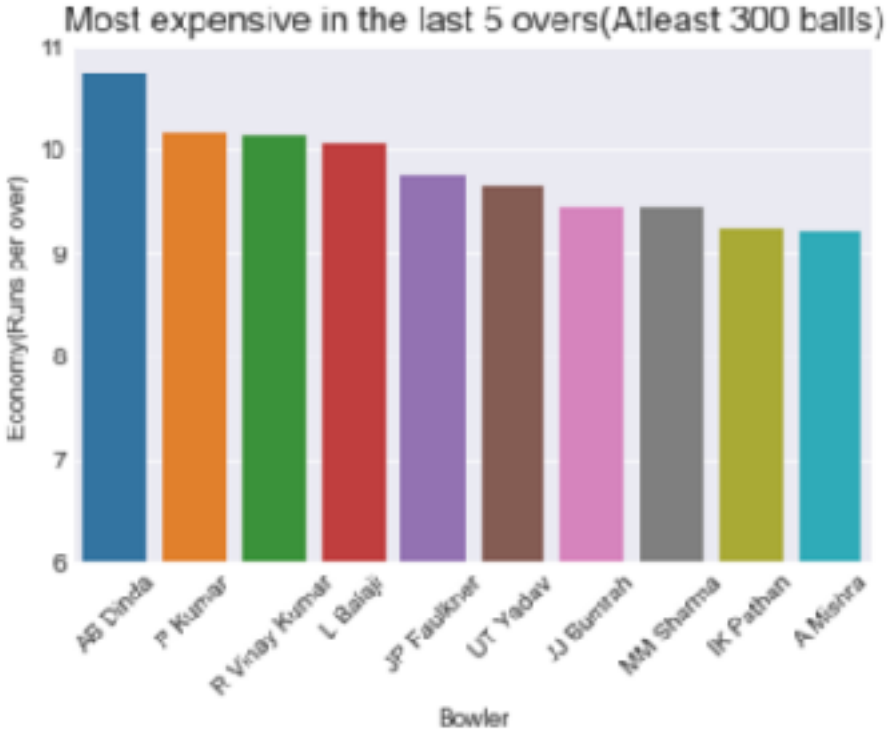
What if you want to get those hard-hitters OUT during death overs?



My study suggests the mentioned bowlers, also knows as ‘Death Over Specialist’, as they have a strong record of getting wickets in the last 5 overs. Interesting observation- most of them are fast bowlers!

(Some more)Findings

I mentioned I feel sorry for the bowlers who bowl in the last 5 overs. Who was I talking about?



Barring A Mishra, all the bowlers are fast bowlers! Looks like we just discovered an interesting stat! If you need wickets: go for fast bowlers! In desperate need of defending a total: go for spinners!

Limitations

My analysis has some serious limitations. To name a few:

- Runs scored during washed out matches haven't been considered since I directly dropped those rows from my dataset (It won’t make a noticeable difference since only a handful of matches were washed out)
- More number of wickets does not necessarily imply that a bowler is good. The quality of the batsman who was dismissed and the situation/pressure in which the bowler managed to get him out is an important factor which has not been captured by my analysis

Conclusions

Most of the answers to the questions are self-explanatory and are discussed along with the graphs.

But one important take-away is the fact that there are some interesting patterns which the coach should be aware of and implement changes to their strategy in order to improve their game. Since loads of data is available, it is now possible to approach every game, every opponent and every batsman/bowler with a different mindset. We can prepare a tailored approach for every situation based on previous history!

Acknowledgements

- I would like to thank Manas for compiling and sharing such valuable data, for free on kaggle.com
- I would also like to thank my brother for asking me some interesting questions whose answers have provided valuable insight.
- Last but not the least, a big thank you to the instructors of this course for all the lectures, quizzes and assignments, without which this project would not have been possible!

References

No research papers were referred to for completing this project.