

# Moneyball: A Cricket Analysis

Team Aces (Kunal Jain, Arjun Benoy, Joyson Chacko George, Nawang Bhutia)

## Motivation and About the Project

Cricket is treated almost as a religion in India, which gathers millions around this sport. In this project, we tried to get insights into this field from the Data Science perspective.

The project has been divided into two parts:

### Part A: Analyse the importance of Toss in winning a Game of cricket.

Data parsed from ESPN Cricinfo website was used in this part to answer the question- To what extent the Toss is going to impact the result of the game?

### Part B: Selecting the World XI Team for the ICC T20I World Cup 2022.

The IPL ball to ball data of the past 5 years, available on Kaggle, was used to select the best 11 players for the World XI team based on the data and statistics.

## Data and Labels

1. For part A, data was parsed from the Cricinfo website to create a dataframe with the match results, toss results and teams playing.
2. For part B, we used last 5 years' IPL ball-by-ball data from Kaggle. It contained ball by ball data of all matches: Batsman on strike, Bowler, Extras, Runs, Wicket, BallNumber, etc.

## Methodology

### Part A

Web scraping

Data cleaning and preprocessing

Data Visualization

### Part B

Top order Batsmen selection

Middle order batsmen selection

Bowlers and Tail Enders Selection

## Results

### Part A

- The impact of Toss on winning the game is different for different Teams. There are teams which utilizes the toss to its advantage, and there are teams who cannot utilize the Toss efficiently.
- The Teams like India, England, South Africa and Bangladesh are found to be utilizing the toss to turn the match in their favour
- Australia, West Indies, Sri Lanka and Zimbabwe are showing a negative trend with respect to toss results

### Part B

The selections were as follows:

- Top Order Batsmen: JC Butler, DA Warner, Virat Kohli
- Middle Order Batsmen: KL Rahul, DA Miller, AD Russell
- Bowlers: Jasprit Bumrah, L Ngidi, DL Chahar, A Nortje

## Conclusion and Future Work

- Utilizing the power of data exploration and analysis we were able to find out various **interesting insights** about cricket, we were able to understand the importance of Toss and how it provides an edge to most of the international teams
- Utilizing the IPL dataset we were able to **identify the best players** of the game
- We can improve upon our player selection by providing **greater importance to recent players** who have performed well.
- Most bowlers were fast pacers and not spinners, this was due to the **lack of role column** in the dataset, we can correct this by finding out **more details about the players by scraping** various cricket websites to gain the required information.

## References

1. <https://www.espncricinfo.com/>
2. <https://www.kaggle.com/dataset/s/vora1011/ipl-2008-to-2021-all-match-dataset>
3. [BeautifulSoup documentation](#)