**A Statistical Analysis of Cricket ODI’s**

Team Members:

Rahul Aakunuru (1209394573)

Aishwarya Pratap Singh (1211162781)  
Ayushi Jain (1204841348)

Nikhil Lohia (1211168085)

1. **Project Idea**  
   To analyze and visualize the statistical data for one day international cricket matches and derive specific patterns that affect a game’s outcome as well as performance of a team under different conditions.  
     
   - Who would be interested in understanding this data?  
   This data would be useful in match predictions, story-telling for a match commentary, betting, team management and team biddings.  
     
   - What would these people want to know about the data?  
   The following set of questions can be answered:-

* Which country has the maximum average runs?
* How many times does the toss affect the outcome of a match?
* How true is the phrase “catches win matches”?
* Analyze the performance of a team in different over intervals.
* Which country plays better in which over intervals?
* Which country performs good in the last overs?
* What is the percentage of runs scored in boundaries?
* Which overs have more boundaries?
* Do boundaries determine if you win a match?
* Are there any outliers in the observation that more boundaries result in winning a game?
* Analysis of extra runs given by different countries.
* How do the extras given determine winning of a team?
* Number of matches played vs won.
* Is there a home advantage for a country/team?
* After scoring how many runs has a team won 90% of the matches

1. **Data**  
   We plan to collect data on around 1300 one day international cricket matches played between 2007 – 2016. The data consists of important match statistics like the teams playing, toss decision, venue, date, man of the match and ball by ball updates in a YAML format.  
   The data is extracted from the website http://www.cricsheet.org  
     
   Sample Data formatteam: England

deliveries:

- 0.1:

batsman: ME Trescothick

bowler: DT Johnston

non\_striker: EC Joyce

runs:

batsman: 0

extras: 0

total: 0

1. **Project Team  
     
   Rahul Aakunuru and Aishwarya Pratap Singh:** Finding co-relation and patterns between data. Analyzing what factors affect a game’s outcome and determining the best possible way to visualize them.  
   **Nikhil Lohia:** Using D3JS/Tableau and related tools for data visualization.  
   **Ayushi Jain:** Data aggregation and cleaning. We have the data (~ 160MB) in a YAML format and it needs to parsed to obtain it in a desirable format (xml/csv etc.).
2. **Related Work**

[1] Segel, E., Heer, J. 2010. Narrative visualization: telling stories with data. IEEE Transactions on Visualization and Computer Graphics

[2] G. Pingali, A. Opalach, Y. Jean, and I. Carlbom, “Visualization of sports

using motion trajectories: Providing insights into performance, style,

and strategy,” in Proc. IEEE Visualization, vol. 4, 2001

[3] A. Cox and J. Stasko, "SportVis: Discovering Meaning in Sports Statistics through Information Visualization," Proc. Conf. Compendium of the IEEE Symp.

[4] C.G. Healey, “Choosing Effective Colours for Data Visualization,” Proc. IEEE Visualization,

[5] C.G. Healey, “On the Use of Perceptual Cues and Data Mining for Effective Visualization of Scientific Datasets,” Proc. Graphics Interface

[6] Prasant Nair, “Editing the worm graph alias ball by ball data and predicting the winner for cricket,” International Journal of Physical Education, Sports and Health 2016

[7] J.M. Patten and K.-L. Ma, “A Graph Based Interface for Representing Volume Visualization Results,” Proc. Graphics Inter-face