**A Statistical Analysis of One Day International Cricket Matches**

Team Members:

Rahul Aakunuru (1209394573)

Aishwarya Pratap Singh (1211162781)   
Ayushi Jain (1204841348)

Nikhil Lohia (1211168085)

1. **Project Idea**  
   The purpose of our project is to analyze and visualize the statistical data for one day international (ODI) cricket matches and to derive specific patterns that affect a game’s outcome and a team’s performance under different conditions.  
     
   *Who would be interested in understanding this data?*  
   This data would be useful for match predictions, story-telling for a match commentary, betting, team management and team biddings.  
     
   *What would our target audience want to know about the data?*  
   We plan to address the following:

* Which country has the maximum number of average runs?
* How many times did the team winning toss win the match?
* How accurate is the phrase “catches win matches?”
* Analyze the performance of a team in different over intervals.
* Which countries play better in which over intervals?
* Which countries perform well in the last overs?
* What is the percentage of runs scored in boundaries?
* Which overs have more boundaries?
* Do boundaries determine whether a match is won?
* How do giving extra runs to a team determine whether it wins or loses?
* Percentage of matches won by teams
* Is there a home advantage for a country/team?
* How many runs must a team score to win 90% of the matches?

1. **Data**  
   We will collect data from around 1300 one day international (ODI) cricket matches played between 2007 and 2016. The data (represented in YAML format) can be downloaded from <http://www.cricsheet.org>.   
     
   Sample data formatinnings:

- 1st innings:

team: England

      deliveries:

        - 0.1:

           batsman: ME Trescothick

           bowler: DT Johnston

           non\_striker: EC Joyce

           runs:

             batsman: 0

             extras: 0

             total: 0

We have one data file per match (A total of 1306 files). The data in file is divided into two sections. First section contains match statistics like Venue, Date, Winner, Team played, toss winner etc. In the second section; for each ball, we have a data node. Each data node has information like, name of the bowler, name of the batsman, number of runs scored and extra runs conceded on that delivery, in the above format. We would parse these files for each match and extract some relevant features like total runs scored by the team, total number of fallen wickets, total of extra runs conceded by the bowling team, total number of boundaries, total of runs scored in boundaries, number of runs scored in each over and so on.

1. **Project Team**The workload will be divided as follows:Rahul Aakunuru/Aishwarya Pratap Singh**:**Find correlations and patterns in data. Analyze what factors affect a game’s outcome and determine the best possible way to visualize them.  
   Nikhil Lohia**:**Use D3JS/Tableau and related tools for data visualization.  
   Ayushi Jain**:**Perform data aggregation and cleaning. We have data (~ 160MB) in YAML format which needs to be parsed and converted to desirable formats (e.g. CSV, XML, JSON). Help Nikhil with D3JS and Tableau. Help Rahul and Aishwarya to visualize the data in best possible way.
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