**A STATISTICAL ANALYSIS OF CRICKET ONE DAY INTERNATIONALS**

Outline: Cricket is an extremely popular game played by over a hundred countries and one day international (ODI) is the most popular format. A single game of cricket has the possibility to generate a lot of data. The performance of the players, the fielding patterns, the bowling variations, the dynamics of a venue, the performance of the team throughout the games, all these factors determine the outcome of a match. All this data can be overwhelming to deduce some insightful results unless we know a co-relation between the factors. The project aims to find some metrics which can investigate the performance of a team as the game and years progress.

The project starts by asking some basic questions like the popularity of the game through the time, the number runs scored at a venue, which country has the maximum runs and so on. It then goes on to ask some insightful questions and debunk myths like how is the toss related to an outcome? Is there a thing as home advantage for a team? How many runs does a team need to increase its chances of winning? Do extra runs really matter? These questions are answered through some interesting visualizations like stacked bar graphs, heat maps, sunburst sequences, choropleth maps and scattered plots.

Data: The project is based upon data from 1300 ODI’s played during the years 2005-2016. Ball by ball updates are available for each game along with venue, date, winner, teams played, toss winner etc.

Tasks Planned to do:

* As a part of the project, I am involved in the task of data cleaning and aggregation so that it can be brought in the desired format. The data is present in YAML format and it is to be converted to csv format with the following fields

Date, Team1, Team2, Venue, Toss, Winner, Over (ex. 3.2), Runs, Extras

* Once the correlation between the variables is determined, I will be implementing a web template how we can present the data along with a layout of all the visualizations such that they relate to each other.
* Using D3 and related tools I plan to

- create a sunburst visualization to determine which overs see the most number of boundaries.

- create a stacked bar graph to determine if winning a toss affects winning the match.

This is not a definite list of tasks to be performed. As the project progresses more co-relations could be determined and visualized.