**Project Title:** **CRICKET - The Past, the Present and the Future...!**

**Introduction:**

Cricket, considered to be England’s national summer sport, is now a global sport being played throughout the world, particularly in Australia, India, Pakistan, the West Indies, and the British Isles. The origin of Cricket is believed to have begun possibly as early as the 13th century as a game in which country boys bowled at a tree stump or at the hurdle gate into a sheep pen. Although the game rules are very different, the basic concept of cricket is similar to that of baseball (a premier sport in the United States). Yet, not many Americans know much about cricket. People in the US often misapprehend cricket to be a complicated sport and ask me to explain it to them. Nothing could be further from the truth. Cricket is played with a bat and ball and involves two competing sides (teams) of 11 players. Teams bat in successive innings and attempt to score runs, while the opposing team fields and attempts to bring an end to the batting team's innings. After each team has batted an equal number of innings (either one or two, depending on conditions chosen before the game), the team with the most runs wins. Cricket currently is played under 3 formats - Test (5-day game), One Day International (ODI, 1-day 50-overs game) and more recently T20 (1-day 20-overs game).

**Data source:**

The dataset I’ll be using is publicly available at <https://cricsheet.org/downloads/>. I’ll be analyzing the datasets corresponding to Test Matches, One-day Internationals and T20 Internationals.

**Motivation & Data Analytics:**

As an Indian, I have lived the game of cricket all of my life. Cricket is my passion and it is considered to be a religion back in India. Cricket is most definitely a game of numbers and statistics. Hence, for my project, I have decided to dig deeper into the past, the present and the future prospects of the game. Specifically, I wish to accomplish the following:

1. Educate the people in the US about the game and the "numbers" that define the game using descriptive statistics and regressions.

2. Explore the evolution of cricket across different formats (3 formats - Test, ODIs and T20s) and its impact on the outcome considering many factors: Rule changes, mindset changes, equipment changes, game conditions, etc. (analytics grouped by specific factors)

3. Predict future game outcomes given current game conditions and statistics from the past e.g. Use ‘decision trees’ to predict the outcome given a team wins the toss and scores ‘X’ amount of runs or find how crucial is winning the toss in terms of winning the game across the 3 formats.

4. Track the impact of game popularity and means to promote the game in the US, e.g. track the progress of US cricket in terms of participation and outcomes.

5. The future of “Test” cricket. Is the game dying? – Is “Test” cricket now more result oriented compared to the past? Track viewership (on TV and at stadiums) to assess popularity of the game.