

## **Capstone 2 Data Project Proposal – Jonny Pearce – 11 January 2023**

### **International Men's Cricket – Predicting the Market Value of Individual Cricketers**

Cricket is a major international sporting business. Identifying players or teams with the highest chance of success is highly lucrative business, both for traditional cricketing formats, but particularly for national or international cricketing franchises. Although a global sport, the largest international cricket market is in India, based upon its own domestic cricket product, but also upon the India Premier League (IPL), an annual international cricket franchise in which players from all the over the world participate in an auction through which each franchise/team purchases its players for each season – see here for more information - <https://www.iplt20.com>

Choosing the right players for your team (whether for international competition or commercial competition) is big business. Being able to predict the right value for the best players and knowing which ones to avoid is crucial to both sporting and economic success. Using the data below, I will try to create a predictive model based on previous performance to indicate which cricketing attributes/performance influence pricing and use this to assess the correct value of players whether in the current system or not.

#### ***Monetary values***

The majority of international players are also centrally contracted to their own countries cricket boards, with varying rates of contractual pay – a summary of this data is provided here for the top tier cricket playing nations - <https://www.totalsportal.com/cricket/central-contracts-in-cricket/>

Auction values for the 2023 IPL season took place in December 2022 and price for each sold player is available, indicating name and player type (and are contained in separate spreadsheets).

#### ***Performance data***

Complete data on all players' performance stats and averages up to 2019 can be found here, in all different formats: <https://www.kaggle.com/datasets/mahendran1/icc-cricket>.

Batting and bowling data for all IPL seasons, highlighting attributes of (for batting) nos of matches/innings, highest scores, averages, strike rates, sixes, fours, and (for bowling) nos of matches, overs, runs conceded, wickets, averages, economy rate, strike rate are available. Also existing value rating for current season players available. All this data is in separate spreadsheets.

The outcomes of games/matches with relevant players can be found here: <https://cricsheet.org/downloads/>. This can provide another measure of value for players, in respect of their contribution to won, lost or drawn matches, as an additional measure to individual performance metrics.

#### **Initial steps**

It will be necessary to review and clean the data, combining from a number of different sources to ensure that it is internally consistent, complete, while also understanding and adjusting for some of the differences between different match performance stats (ODIs,

versus T20s, versus Tests, versus IPL), and understanding player inside and outside the IPL.