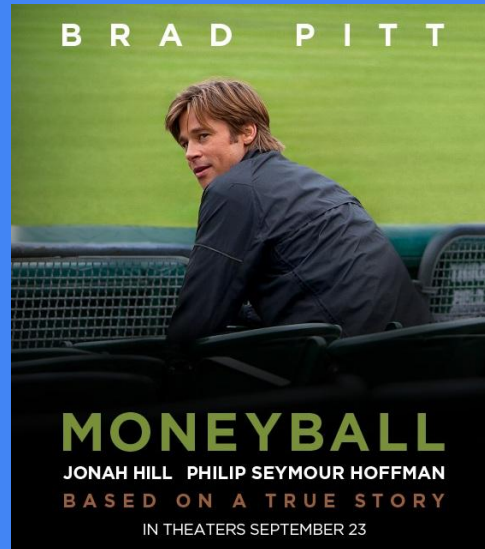# Data Analytical Application for cricket using Moneyball Techniques

Progress Report

Objective

Evolve a mathematical model to build the perfect team possible (statistically) to win a certain tournament(IPL).

❏ To optimize the auction strategies of teams, given the price and player constraints.
❏ To evolve our metrics to rank players against each other.
❏ Analyse previous auction strategies and comment on the effectiveness of the buy.

Data Scraping

A lot of unorganized data on the web(except Cricinfo).
(Python's BeautifulSoup used)

1. The database used is MongoDB
2. Initially, all Indian players (batsmen) database(all batting related stats) has been scraped.
3. All players(batsmen) bought in 2015 Auctions data (performance in 2015) scraped.

Literature Survey

After substantial research in this area, two kinds of approaches were seen:

1. Player pricing on various attributes(hedonic pricing)

2. Evaluating useful metrics to compare players, so that appropriate price analysis could be carried out.

Hedonic Pricing

Some research papers

1. "Player Pricing and Valuation of Cricketing Attributes: Exploring the IPL Twenty-Twenty Vision " by Siddhartha K. Rastogiand Satish Y. Deodhar analyses the hedonic pricing of players using linear regression on the varied dataset mined from Cricinfo and Wikipedia.
2. "A Hedonic Model of Player Wage Determination from the Indian Premier League Auction: Further Evidence" by Liam Lenten, Wayne Geerling, and Laszlo Konya focussed on analysing the inaugural 2008 IPL auctions, and commented on the efficiency of investments in a player.

Some research papers

3. "Player valuations in the Indian Premier League" by David Parker, Phil Burns and Harish Natarajan is a very well- written paper in this area. It analyses all aspects discussed so far, in a well- defined manner. It also discusses the aspect of marketability of a player being a factor in determining team decisions in auctions. It goes on to explain certain events like English players getting a different paycheque from the others, and the"icon" players having a different set of variables.

Metrics for comparison

Some research papers

1. "Analysis of Performance of Bowlers using Combined Bowling Rate" by Dibyojyoti Bhattacharjee and Darshan G Pahinkar explain how bowlers can be compared using metrics like combined bowling rate(harmonic mean of economy, strike rate, and average of bowler).

2. "A MCDM Approach for Evaluating Bowlers Performance in IPL" by Pabitra Kumar Dey, Dipendra Nath Ghosh and Abhoy Chand Mondal , a highly mathematical paper uses MCDA(multiple criteria decision analysis) to generate a ranking of players, based on the multiple criteria available.

Return of Investment (2015 Auctions)
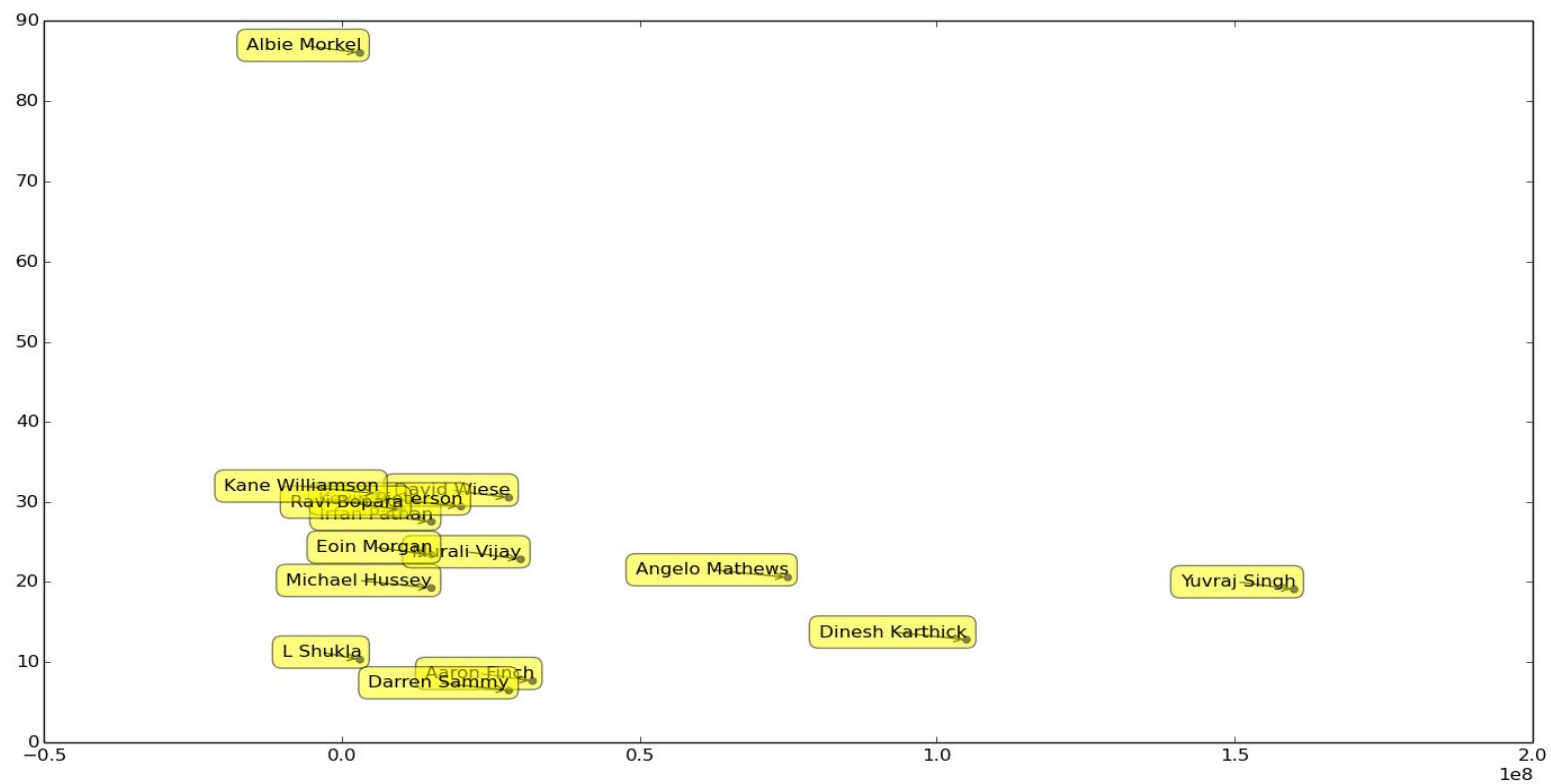
Average of player Vs Price(in tens of crores) paid in 2015 Auctions
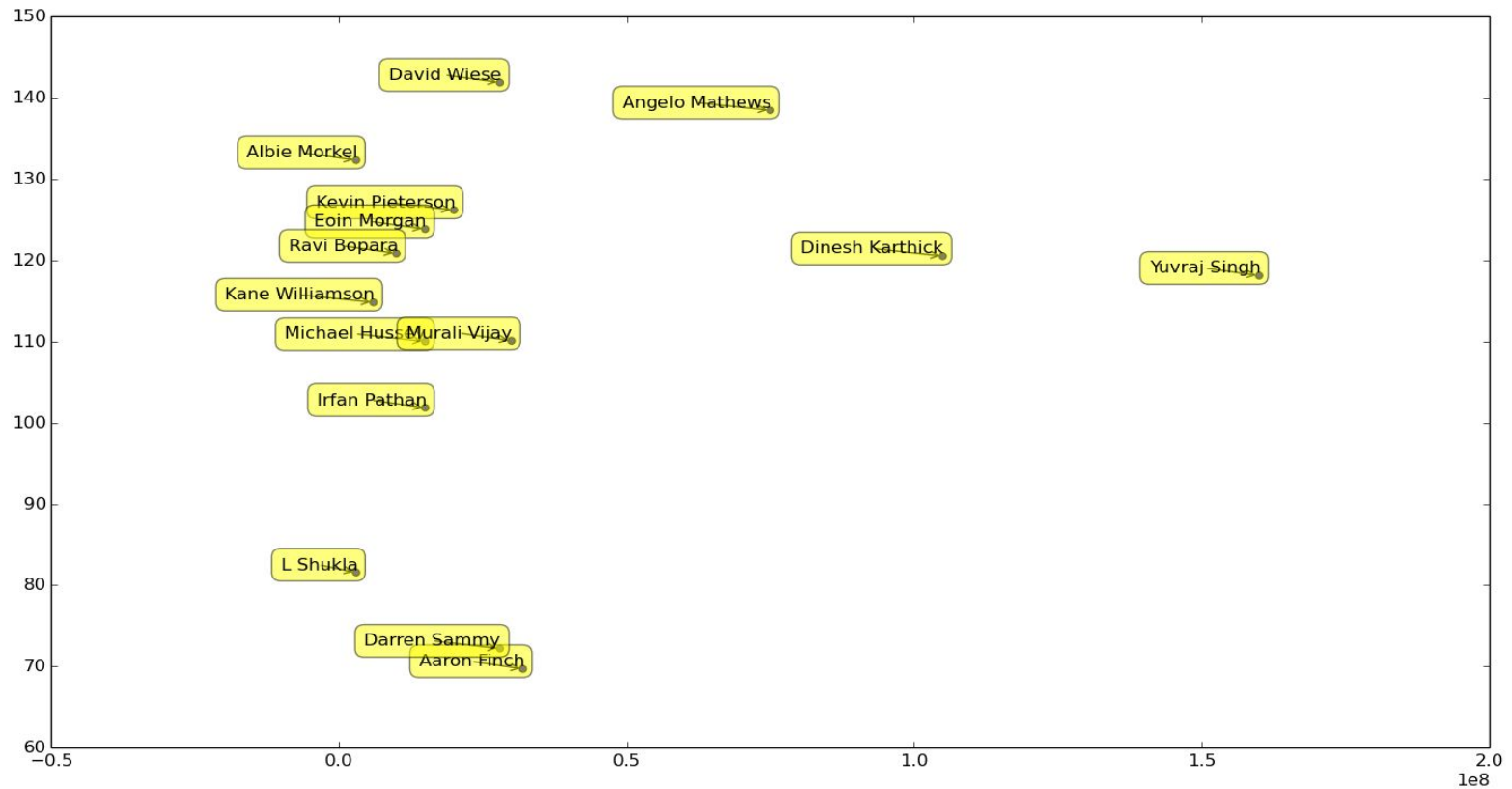
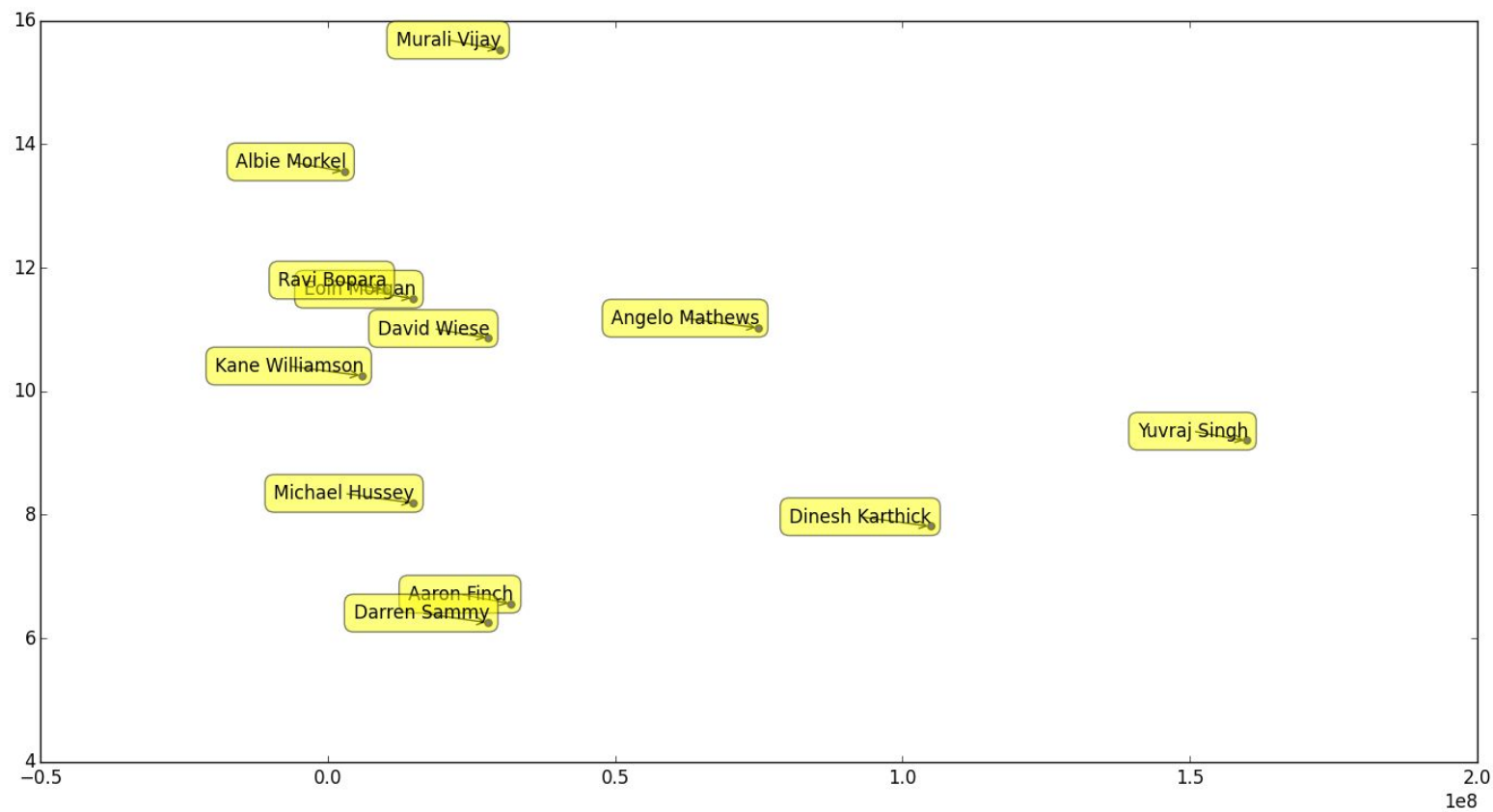Strike Rate of player Vs Price(in tens of crores) paid in 2015 Auctions

Work to be done!

1. Devise accurate metrics to compare the player performances.
2. Generate Auction Strategies based on the given constraints. For Eg. In the case of Baseball, the statisticians found, if the team made 'x' number of runs over all matches in the league, they had a high probability of qualifying. So, they bought those players which gave them the 'x' runs in 'total'.
3. To experiment with data and find some such insight in the cricketing domain.

Metrics

Continuous Adjusted Average : Just looking at average isn't enough. Consistency is to be considered as well. Gini Coefficient measures the inequality among values of the 'run' distribution.

$$G = \frac{\sum_i \sum_j |x_i - x_j|}{2 \sum_i \sum_j x_i}$$
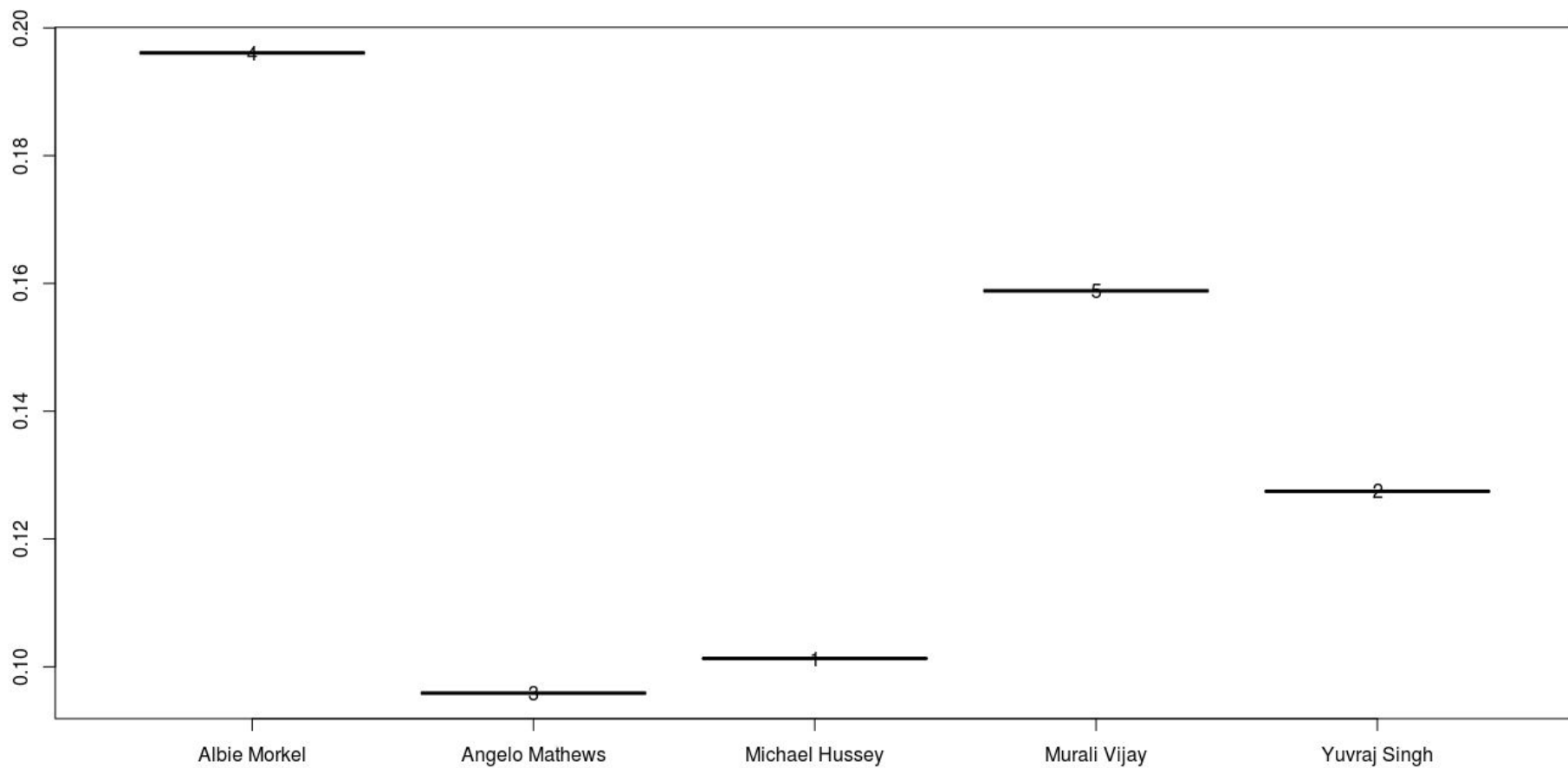
CAA = Average*(1-G)

Continuous Adjusted Average(CAA) of player Vs Price(in tens of crores) paid in 2015 Auctions

Metrics

The contribution of each player in the score of the team also says a lot about the batsman's impact on the match.

The ratio of batsman's score and team's score is averaged over all matches that the player has played.

Research on custom metrics(batsmen)

1. "Reflecting Against Perception: Data Analysis of IPL Batsman" by Amit Kumar , Dr. Ritu Sindhu
2. Some Issues in the Calculation of Batting Averages:Ranking (and Re-Ranking) the Top 50 Batsmen in Test Cricket, 1877-2006 by Vani K. Borooah & John Mangan

Ranking by metrics

Now, to test these metrics, further I'll implement techniques like PCA(Principal Component Analysis) to rank the players.

The high variance metrics will matter more, while very little varying metrics will be ignored(given low importance).

Source -> "An Introductory Application of Principal Components to Cricket Data" by Ananda B. W. Manage and Stephen M. Scariano

Entropy to measure competitiveness

For my proposed ultimate aim, to build teams based on data-driven decisions, an interesting measure is entropy.

Entropy is calculated based on the number of points won at the end of the tournament -> Indicative of competitiveness

Source -> "Measuring Competitive Balance in Sports using Generalised Entropy with an Application to English Premier League Football" by Vani Borooah, John Edward Mangan

Entropy to measure competitiveness

How we can use it?

To find out how unpredictable our current team is, and then building upon which players to hire to improve our predictability. Currently working upon getting team statistics.

Thank You!