Project Proposal: Predicting MLB Players Salaries

Earlier last year, the Los Angeles Angels signed Mike Trout to a 12 year, $430 million contract. This is the highest amount ever paid to a player in the history of the MLB. The Angels made an investment with the hopes that he will bring even greater value over the next 12 years in the form of ticket sales and merchandise sales. But, if it doesn't turn out as planned and he gets injured or begins to perform poorly, they would be at a loss. So, how can they be sure they're getting their money's worth?

A rookie, who is called up to play at the big league level is automatically given a minimum starting salary of $500,000. After a couple of years, they are able to begin partially negotiating their own contracts in the form of salary arbitration. If they are unable to come to an agreement with the ownership, independent third parties are paid large sums to arbitrate. How can a player know their value to a potential team to maximize their negotiation position for a contract?

In this project, I plan to build a model to predict player salaries based on their previous year stats which can be used both by sports franchises, to minimize their risk, and by players, to help them determine their value. The data will be scraped from 'http://www.thebaseballcube.com/', a baseball data warehouse. The final deliverables for the project will be the code in a github repository alongside the final report and slide deck.