Joey Goodman

Capstone Project – Proposal

August 7, 2016

In sports competition, emotions tend to run high. Sometimes, a player or coach may become so incensed that they are removed from the game, ejected by officiating personnel. It is quite possible that such a change in the competitive dynamics would be likely to reduce a team’s performance in the short term: if a key player or manager is ejected from the match, a lesser athlete or coach may be required to fill this role temporarily. However, it is also theoretically possible for ejections to serve strategically beneficial purposes as well, especially ejections of managers or coaches: a manager might initiate an argument with an official with the objective of getting ejected in order to “fire up” his players and improve their performance for the remainder of the game. So, do ejections improve or worse a team’s performance in the short run?

My Springboard capstone project seeks to investigate this puzzle within the context of baseball. This research topic falls within the field of baseball analytics, and answering my research question would be of strategic interest to a team seeking to optimize its in-game performance. Depending on whether an ejection causes a team to perform better or worse, it may make sense for a manager to bite his tongue, rather than engaging an umpire in an argument. However, it may also be the case that the effect of an ejection on team performance is dependent on what role the individual plays on the team or how much he contributes to the team’s success.

I plan to approach this question by analyzing 16 MLB seasons (2000 through 2015) of play-by-play data from the baseball data repository site Retrosheet. The data is downloaded as text files for each home team in each season. Though “play” records in these files – rows which capture changes in the game dynamics such as a hit or an out – can be processed fairly easily using code provided by retrosheet, data on ejections appears in these flat files as “comment” records. Though comment records include miscellaneous commentary about each game, I have written my own Python script to pull out comments that specifically refer to ejections and processed all ejections occurring between 2000 and 2015 seasons into a single CSV file. These data have been loaded into a PostgreSQL database in a table called *Ejections,* which has about 3,400 rows. Play-by-play data from these 16 seasons have been loaded into a table called *Event*, which has almost 3.1 million rows.

The main way I seek to approach this research question is by comparing “windows” within a game that occur just before and immediately after an ejection. Given that an ejection occurs, how is a team’s final score (or runs allowed) impacted, compared to the average for non-ejection games? In other words, how discriminative is an ejection as a feature toward a team scoring more runs, allowing fewer runs, and/or winning games?

After completing my analysis, potential deliverables might include code, a presentation deck, as well as a short article-length write-up, perhaps in the style of what you might find in the sports section of FiveThirtyEight.