Ryan Thompson

Idea 2

As a second idea I would like to investigate something about professional baseball. I would love to build a model that could predict whether a team would make the playoffs or not. There are also other avenues I could take regarding professional baseball, such as building a model that could predict whether a player was an all-star or not based on certain statistics. For the purpose of this paper, I will focus on the idea of building a model that could predict a playoff clinching season.

There is so much available data on professional baseball. First, MLB’s official website has statistics going back years for every qualified player and team in baseball. You can also analyze both basic statistics and advanced statistics to evaluate a player or team. The data is not hard to transfer from their official website into an excel file which can than be put into jupyter for python.

With the current track I see myself taking, I believe I would build a model that could predict if a team makes the playoffs or not based on certain characteristics and statistics that team possesses. For example, I would take every team from the last 10-15 years, and assign them a binary variable, 1 if they made the playoffs, 0 if not. From there, I would run analysis and tests on each statistic to see which one lead to the team making the playoffs.

With this information, the stakeholders that would be most affected would first be front offices of baseball teams. These teams would be given this model and would build their clubs based around the results. Additionally, players would be greatly affected. Their salaries and contracts would fluctuate depending on what the model values a good team to need. If it is shown that on base percentage is the top predictor of a playoff team, then teams would prioritize players that get on base.