**Introduction:**

We decided to do a project centered around the game of baseball. Our goal was to predict the rounds in which each MLB playoff team lost, from 1995-2016.

Here is a quick overview of the game of baseball:

Baseball is a **“bat-and-ball”** game played between two teams of nine payers each, who take turns **batting** and **fielding**. The batting team attempts to score runs by hitting a ball that is thrown by the **pitcher** with a bat swung by the **batter**, then running counter-clockwise around a series of 4 bases: first base, second base, third base, and home plate. A run is scored when a player advances around the bases and returns to home plate and touches the base. Players on the batting team take turns hitting against the pitcher of the fielding team, which tries to prevent runs by getting hitters “out” in any several ways. A player on the batting team who reaches a base safely can later attempt to advance to subsequent bases during teammates’ turns batting, such as a “hit” or by other means. The teams switch between batting and fielding whenever the fielding team records three outs. One turn batting for both teams, beginning with the vising team, constitutes one “inning”. A game is composed of nine innings, and the team with the greater number of runs at the end of the game wins. (In *Wikipedia*. Retrieved December 8th, 2016.)

In Major League Baseball, there are two leagues: 1) The National League and, 2) The American League. Each league has three divisions: 1) East, 2) Central, and 3) West. Every team plays a 162-game regular season. At the end of the regular season, the leaders of each division are guaranteed a playoff spot, totaling 6 playoff teams. The two teams with the next best records in each league play in a single-elimination “wild card” game, adding two more playoff teams. The respective playoff rounds that each league competes in individually are 1) the Division Series and, 2) the Championship Series. The National League Championship Series winner and the American League Championship Series winner will face off against each other in the **“World Series”.** The Division Series is out of 5 games, so the winner must win 3 games. The Championship Series and World Series are out of 7 games, so the winners must win 4 games in these rounds. Therefore, a team must win 11 playoff games to win the world series.

Baseball has been a pioneer in the use of analytics to predict performance in sports. Theo Epstein, the General Manager of the Chicago Cubs and former General Manager of the Boston Red Sox has been using analytics to operate his teams for the past few decades, and it has paid off in a huge way, winning the World Series with both teams. He is now one of the most legendary front-office managers in all of sports. We decided to follow his path and create a World Series Prediction Model.

We tested multiple predictive modeling techniques such as linear regression, logistic regression, and KNN – Nearest Neighbor to predict the number of playoff games won by each playoff team. The team that was closest to 11 playoff wins was deemed the “World Series Winner”. The team with the 2nd highest number of playoff wins was deemed the “Championship Series Winner”, and the team with the third highest number of playoff wins was deemed the “Division Series Winner.” We decided to use data from the past 20 years, including 1995, because the “Wild Card” playoff system began in 1995 which changed the number of playoff teams from six to eight. The modeling technique with the best accuracy score was KNN – Nearest Neighbor. Below, you can see a distribution of the playoff teams of the past 20 years, and the playoff rounds in which lost/won.

