Conclusions

Attendance vs Record:

We found a .69 correlation between attendance and the team’s win/loss ratio in a season. We expected a stronger correlation between these values. We attributed this to ‘any weather’ fans who essentially show up and buy tickets regardless of the team’s performance. An example of this phenomena is the Chicago Cubs who show high attendance despite having a less than 50% win rate.

Attendance vs Temperature

We found essentially no correlation between temperature and attendance despite our expectation that there would be a significant impact. Our assumption is that this is due to tickets being purchased well in advance of games. Additionally, we used MLB reporting of attendance, which is truly ticket sales, and could not find publicly available data of true attendance at games. Season ticket holders are also counted as attending every game regardless.

Attendance vs Payroll

This relationship is high, as we expected it to be. This is probably caused by teams with higher ticket sales having greater income to spend on payroll. There are some significant outliers in this data, for instance, the New York Yankees spent an average of 2.14x of the average payroll and do have the highest average attendance per game but the St Louis Cardinals only spent 1.06x of the average payroll and still had 92% of the ticket sales of the Yankees.

Attendance vs Population

We assumed that this relationship was weaker than it turned out to be. Certain cities have a reputation as being “baseball cities” and while this did have a significant effect in some cases, there is a moderate correlation between population and attendance. We chose to use Combined Statistical Areas in order to pull larger geographic areas and collect population data for cities such as St Louis where a portion of the metropolitan area is across state lines. Because of this we did not include data for the Toronto Blue Jays and 4 US teams (Arizona, Tampa, Miami, and San Diego) which are not grouped by the Census into part of a Combined Statistical Area. We assume that this would slightly lower the correlation as Tampa and Miami have substantial populations but some of the lowest attendance in major league baseball.