Executive Summary

Many baseball enthusiasts have the desire to visit all thirty Major League Baseball (MLB) stadiums in their lifetime. An even more thrilling and challenging accomplishment though would be to watch a game in all of those thirty stadiums in one season. To quench this thirst of accomplishment, this study attempts to minimize both time and distance traveled to watch a game in every stadium during the 2017 regular season.

To accurately portray both the time and distance aspect of the problem, the full MLB schedule was pulled from the official MLB website and driving distances were downloaded from http://dailybaseballdata.com/base/sched4.html. This led to two data sets used in the model. The schedule data was a 30x168 data set and the distance data was a 30x30 data set, leading to 151,200 variables. Although the goal of the study was to provide as realistic of an answer as possible, several assumptions were made to simplify this initial study. The first of these was that all travel would be completed by car. The second, and possibly most unrealistic, was that all stadiums could be driven to in less than one day from any other stadium so that a game would not be missed due to travel time. This can be thought of as instantaneous travel times. The third assumption used in this study was the starting point being at a given stadium and not at a neutral site.

Using this basic logic, the problem was setup as a binary math programming problem. The minimization problem described above was constrained by logically forcing the path to start and leave node one. Formulation also forced travelers to continue on a connected path and not transport to other cities without actually leaving previous cities. It also ensured that a game was being played on the days spent at each city.

Due to the large amount of binary variables in the model, there was an anticipation of large run times for math programming problem. Our analysis sought to account for this limitation with an experiment of optimality and feasibility condition. Three levels of optimality conditions were tested however none of these attempts met project deadlines when invoked in a LINGO model. To improve solver time, only a portion of the schedule (60 and 80 days) was considered in the trip optimization, however this instance did not meet project deadlines either.

We mitigated this issue by creating a heuristic to find a near optimal route of travel. The heuristic arbitrarily chooses a starting location, determines the closest stadium to travel to that has a game the following day and has not been visited yet, and iterates for thirty games. Using this approach our suboptimal solution to the problem indicated a 12594 mile trip in 42 days. This heuristic should be advanced by comparing solutions from instances where the starting point is at each different stadium. If given additional time to solve this problem, an optimal solution may be more tractable if solved in GAMS or another more advanced solver. Future additions to our methodology should include breaking down our third assumption that driving time is instantaneous and our first assumption that travel can only be completed by driving.