Executive Summary

Overview:

When it comes to playing sports, athletic performance is considered to be the most important quality athletes and coaches can concentrate on. Even small differences in athletic performance could mean the difference between a team victory or defeat. In order to address concerns with athletic performance, coaches and athletes should try and understand key underlying factors which determine performance. For the purpose of this analysis, specific focus will be placed on athlete fatigue.

The dataset from Rugby Canada Women's 7S consists of four different datasets collecting information with the goal of increasing performance levels. These datasets look at measuring fatigue externally(gps, rpe, games) and using subjective measures(wellness).

Analysis:

The main component of the analysis focused on creating a variable which measured fatigue on a score out of 100. We based the score on 2 main categories: lifestyle variables and training variables. Using mathematical transformations, we fit the values from the variables Focus Rating, Sleep Quality, and Irritability to be half the fatigue score. The variables Daily Load and Acute-Chronic Ratio were fitted to be the second half of the fatigue score. Once we formed the score, we used other variables from the data (some self-reported and some measured) to form a predictive random forest model, which listed variables of importance, including Duration, Objective Rating, Pain, and more. The out-of-bag error estimate was 4.932, and our percentage of variables explained was 39.51, neither of which are awful for a model on human characteristics. We then looked at the relationship between some of the most important variables according to the random forest model, all of which were only slight relationships, nothing too extraordinary.

We then pulled the data from the gps.csv file to look at the speeds of players and how they change over the course of a game. We found the average speed of players in the first and second halves of each game, and found whether they increased or decreased and by how much. We felt this was a good indicator of endurance, or during-activity fatigue. We then broke both our Fatigue Score and the Difference in Speed between Halves into categories: High and Low Fatigue Scores, and Big, Medium, and Small Average Speed Differences between Halves. The proportion of big speed differences is larger for the athletes with low fatigue score than those with high fatigue score.

We also used the gps data to create an animation that we compared side-by-side with a video of the game being played, as an illustration of the change in speeds and amount of energy used by the athletes.

Conclusions:

Our fatigue variable seemed effective at estimating the fatigue level for the athletes. We wanted the factors making up the fatigue score to be controllable aspects of the athletes life, so that if we could show that improving this fatigue score will increase athletic performance, there was a way to do so. Athletic performance is the culmination of many factors, but we've shown how improving the fatigue score will help improve endurance by some amount. We also believe the fatigue score provides a tool for coaches to monitor how their athlete's daily and training habits may be impacting their performance.