First and foremost, I feel the need to stress that our primary focus was utilizing all data provided, cleaning and joining appropriately, and using statistical methods to normalize our data.

Before trying to develop the hypothesis, we decided we needed some metric in which we could measure some sort of relative performance. We decided the numerous amount of GPS data would be the best source for this metric.

**We choose speed to be our overall performance metric for the following reasons:**

- If running a field goal, you will have high sustained speed over a long period of time
- The faster player has a huge advantage, and is harder to tackle
- If you get tackled, you will have a sudden decrease in speed
- Positive speed means your team is winning the scrum
- Average Speed is correlated with play-time, meaning coaches select for higher speed

**Cleaning Data**

We had multiple rows with the same date and player ID tracking multiple different exercises per row. We collapsed this into one row containing all the data. Among new columns we created total exercises for that day, total RPE, total duration. Any column that had more than 20% NA's were completely removed.

In order to ensure all wellness data rated on a 0-7 scale can be compared accurately, we normalized all individualised data using a z-score. After cleaning the data, we calculated the sum and mean for all our final variables and merged them.

An average rugby game is 12 minutes and our total time played for all players / 7 (the number of players on the field at any given moment) was ~13 - 16 with no outliers. Given rounding errors, potential overtime, and the data showing many games with game clocks going up to 16:00, we feel this is an extremely accurate metric for differentiating time played and time on the bench.

**Developing Hypothesis**

We were unable to predict speed using all of our variables and multivariate logistic regression.

Due to the level of detail on our data, we have developed a very large correlation matrix and statistical t tests. We have a numerous amount of variables that are correlated to other variables with 90% confidence. This data could be immensely useful for the coaches to run experiments.

We have shown that play time is largely correlated with average speed, which shows that coaches choose players with higher speed to play. Worth researching that average speed may not actually affect chances of winning in any significant manner. (We saw no correlation between average speed and winning within our data. However, a larger sample size would be necessary to draw firm conclusions on how average speed affects chances of winning).