USOPC Performance Analyst Data Project

I started by loading the data into a python notebook and doing some exploratory data analysis. This helped me understand what values were present for each athlete, as well as see missing values and start to explore the relationships between variables.

It was at this point in the process that I noted that travel stress and sport specific training volume had very few observations (50% or less than total observations), which eventually led me to exclude these variables from later analysis.

I also took the time to isolate a single athlete, ‘Athlete 2’, to see the observations of an individual athlete. At this point, I plotted the various wellness metrics to see if there were any evident trends over time, while also plotting a correlation matrix to note the interrelatedness of various metrics. Notably, there is a gap in wellness metrics between July 1 and August 1, which happens to also be the time during which several competitions took place. Several pairs of variables were highly correlated: Fatigue and Travel Hours, Sleep Hours and Sleep Quality, Motivation and Travel Hours, and Fatigue and Motivation.

After exploring the data, I then joined the competition data with the wellness metrics and plotted the combined data for ‘Athlete 2’ to once again check for any obvious correlations or trends. Notably, the times for Heat 1 and Heat 2 for that athlete in any given event were nearly identical, and as expected, times for Heat 1 and Heat 2 were highly correlated with the total event time.

As an elite mountain biker myself, it was easy for me to think about this performance data in the context of mountain biking. With varying terrain/length/course conditions, as well as the potential variance of other variables outside of the athlete's individual wellness and performance (what competitors are there, their performance, etc.), the confidence of any conclusions drawn from the result data is limited.

For example, any correlation between wellness factors and improved (reduced) times may be incorrect, as the duration/distance of the course may be more a factor in athlete times across events than are the wellness metrics. If the target variable is athlete ranking, improvements in ranking may also be more a factor of what other athletes competed and their resulting performances than a representation of that individual athlete’s wellness metrics.

However, with these limitations in mind, I proceeded to ask the question: “What wellness metrics could potentially be used to predict an athlete's performance (by rank or time) in an event?” Specifically, I wanted to give coaches and athletes some actionable insights for guiding their training in the lead-up to competitions.

As a cyclist, I also know that some training platforms (Training Peaks) have a metric (TSS) based on power (typically measured by a power meter), that, when exponentially calculated over time, can be used to measure chronic training load (CTL), acute training load (ATL), and training stress balance (TSB), also sometimes referred to as form. While this dataset does not include TSS (or the power metrics necessary to calculate TSS), I thought perhaps a variation of these metrics could be calculated to help measure the cumulative training load of athletes leading up to competition.

In this data, fatigue is self-reported here the athlete, meaning it is a more subjective variable than power would be, but my thought was to use the fatigue metric, exponentially weighted with the starting values used by Training Peaks in calculating their CTL from TSS. After calculating CTL, ATL, and TSB for each athlete based on their reported fatigue, I attempted to predict the athlete’s competition time and ranking based on their wellness metrics.

In order to do this, I aggregated the mean of the previous 7 day's fitness metrics (Stress, Motivation, Sleep Quality, Resting HR, Fatigue, Sleep Hours, Soreness, ATL, and CTL), as well as the prior day's TSB for each athlete's event date. I then fit the data to both a multiple regression model and a random forest model for each target variable. I did this for the 7-day aggregated data, as well as 30-day aggregated data, and non-aggregated data.

Data **Target Variable**	**Model**	**R^2**	**MSE**	**OOB**	**F-Statistic** **P-Value**	**Statistically Significant Coefficients** (Multiple Regression Model) *or* **Most Important Features** (Random Forest model)
7-Day Aggregated Athlete Rank	Multiple Regression	0.26			1.833 0.0870	None
7-Day Aggregated Athlete Rank	Random Forest	0.758	10.004	0.048		Fatigue Motivation ATL
30-Day Aggregated Athlete Rank	Multiple Regression	0.298			2.451 0.0207	Resting HR Soreness
30-Day Aggregated Athlete Rank	Random Forest	0.741	11.332	-0.164		TSB ATL Sleep Quality
Non-Aggregated Athlete Rank	Multiple Regression	0.205			2.023 0.068	TSB CTL Fatigue ATL Sleep Hours Soreness
Non-Aggregated Athlete Rank	Random Forest	0.192	34.808	-0.153		Resting HR Motivation Sleep Hours
7-Day Aggregated Athlete Time	Multiple Regression	0.204			1.336 0.245	Sleep Quality CTL
7-Day Aggregated Athlete Time	Random Forest	0.808	295.593	-0.415		CTL ATL Stress
30-Day Aggregated Athlete Time	Multiple Regression	0.097			0.619 0.775	None
30-Day Aggregated Athlete Time	Random Forest	0.736	402.80	-0.478		Stress Resting HR TSB
Non-Aggregated Athlete Time	Multiple Regression	0.085			0.731 0.646	Stress Resting HR
Non-Aggregated Athlete Time	Random Forest	0.070	1405.543	-0.39		Soreness Sleep Hours Stress

Both multiple regression models would indicate that the 7-day aggregation of wellness data is not statistically significant when trying to predict event performance (either by rank or time). The 30-day aggregation of wellness data is only statistically significant when predicting athlete time in an event (which, due to the

potential variance in event distance may be in itself an unreliable outcome). Without aggregating the data, the model is statistically significant when predicting rank, but still only explains a small amount of the variance in rank.

The random forest models seem to perform significantly better when using 7- or 30- day aggregated data in comparison to the non-date-aggregated data. The best performing model (as measured by MSE) was the random forest model using 7-day aggregated data to predict athlete rank. This model indicated that fatigue, motivation, and acute training load (ATL) are the three most important features when predicting athlete rank.

This can lead to several conclusions:
1. For the athlete (and coach), the numbers are not the final say. If wellness numbers seem to be poor prior to an event, the athlete can still enter the event with confidence--knowing that wellness metrics do not consistently correlate with performance outcomes.
2. For researchers (and athletes and coaches), more data should be gathered, especially as regards the mental state of athletes going into events. There may be a more definitive correlation between mental state and performance than between wellness metrics and performance. This is where the use of an athlete management system to record mental wellness metrics prior to an event (perhaps a brief questionnaire regarding feelings of preparedness, confidence, etc.) could be useful in enhancing the collection and analysis of project data. It would also be helpful to have day-of-event wellness data (the same data that is collected daily, but during the season and on event day), as gaps in the data collected certainly cast doubt on any conclusions that might be drawn from existing data.

With these conclusions in mind, I created a dashboard for the athlete and coaches to help them visualize this data (with the added CTL, ATL, and TSB metrics) in the lead-up and preparation for competitions. I created one dashboard for individual athletes to see their wellness metrics, and then a second dashboard where a coach could view all athletes simultaneously. The third dashboard gives the 7-day aggregated data alongside historical athlete event rank, with reference lines based on the results of the best-performing random forest model so coaches and athletes can see at a glance where their wellness metrics fall in comparison to what the model’s threshold values were for predicting improvement in rank at upcoming competitions.