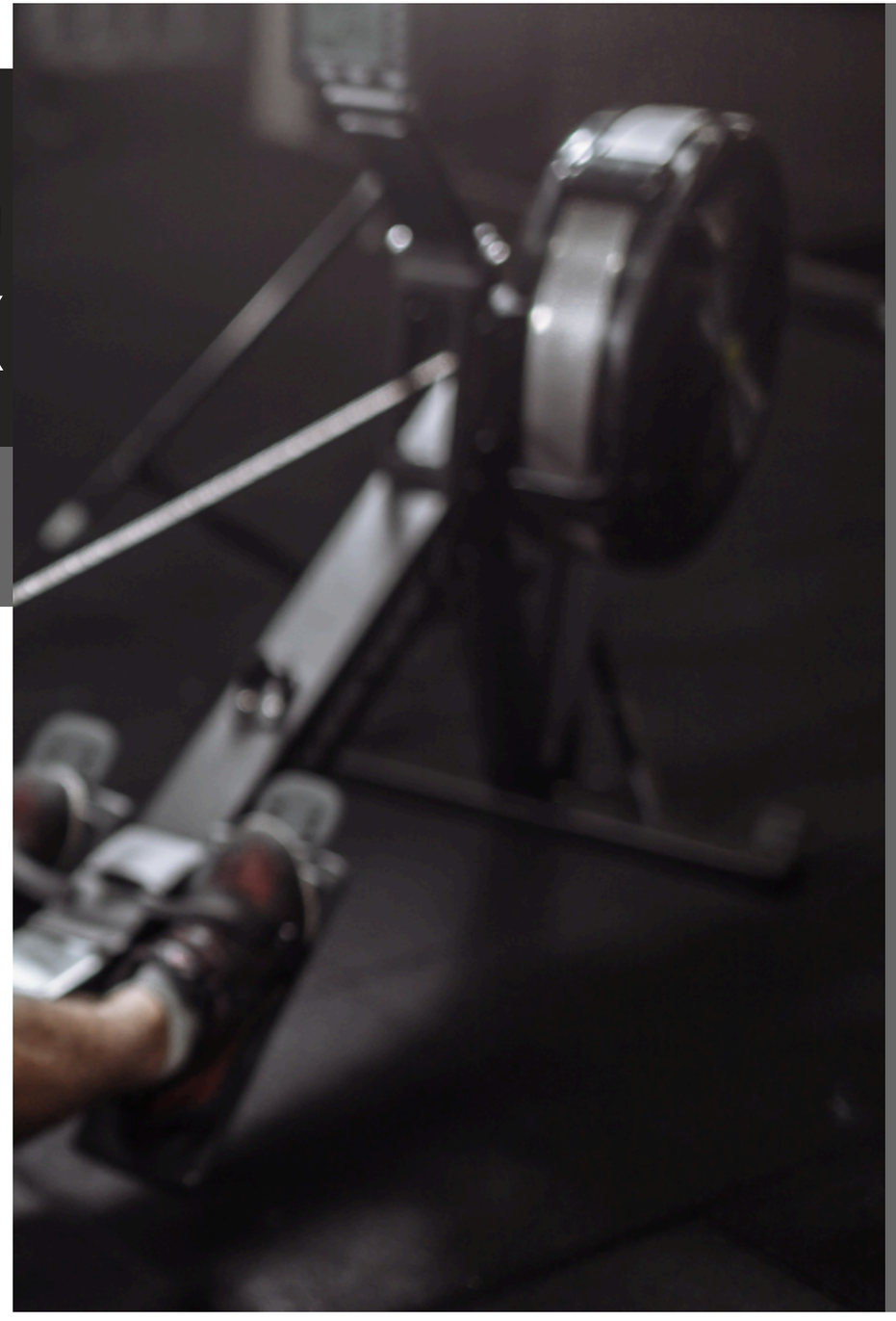
Data Science competes with HYROX

Student number: BP0289090



As a data science student with a passion for fitness and a personal connection to Hyrox - thanks to my husband who enjoys competing - I decided to dive deep into the data of this exciting, relatively new competition.

Hyrox is growing fast, but research around what contributes to success in these races is still limited. My curiosity led me to analyse the performance data of these incredible athletes to uncover what key factors contribute to achieving those impressive sub-one-hour finishes.

Additionally, I aimed to identify which variables can be used to predict a participant's finish time more accurately. Click the arrow on the right to begin exploring the analysis and discover the secrets behind Hyrox's top competitors.

This analysis is completed on male finishers, competing in the Pro category.

Summary

1. Predicting Finish Times:
   - Our analysis shows that later race segments, particularly **Run 4** and **Run 5**, are stronger predictors of overall finish times, as indicated by higher R-squared values (0.67 and 0.66, respectively). This suggests that performance in these segments is more closely linked to overall success.
   - Focusing on these key segments can help athletes improve their overall race time by targeting specific areas for performance enhancement.

Click here to check this out

2. Finish Time Predictor Tool:
   - The Finish Time Predictor tool provides a practical feature for athletes and coaches to estimate total race times based on accumulated times up to Run 4. This helps set realistic mid-race targets and adjust pacing strategies accordingly.
   - This tool is also beneficial for spectators, allowing them to better anticipate when an athlete will finish, enhancing the overall viewing experience.

Click here to check this out

3. Work and Run Times Analysis:
   - What makes these sub 1 hour finishers perform so well??
   - For competitors finishing in under one hour, certain workout stations—such as **Sled Push**, **Burpee Broad Jumps**, and **Sandbag Lunges**—show a significant opportunity for time savings. Improving performance in these stations could lead to faster overall race times.
   - In terms of running, **Runs 4 and 5** have shown the highest correlation with total race times, indicating these are critical segments to focus on for optimal performance.

Check work times

Check run times

Predicting Finish Times

Race time predictors often estimate your finish time based on the time you have spent competing so far. However, this approach is more challenging with Hyrox due to the wide variety of stations.

To gain insight into this, we can examine each run or workout station and calculate the R-squared value against the total finish time. R-squared is a statistical measure that shows how much of the variance in the dependent variable (finish time) is explained by the independent variable (each run or workout time). In other words, **it helps us determine how accurately we can predict a finish time based on performance in a single event.**

RSqd	Name
0.67	Run 4
0.66	Run 5
0.63	Run 7
0.62	Run 6
0.61	Run 8
0.60	Burpee Broad Jumps
0.59	Sandbag Lunges
0.58	Run 3
0.55	Run 2
0.55	Rowing
0.54	Wall Balls
0.52	Sled Pull
0.50	Farmers Carry
0.42	SkiErg
0.38	Sled Push
0.26	Run 1

Looking at the R-squared values for each run and workout station, we can see which activities are the best predictors of total finish time in Hyrox. Higher R-squared values indicate a stronger correlation with the overall race time. For example, **Run 4** and **Run 5** have the highest R-squared values at 0.67 and 0.66, respectively, suggesting that performance in these segments is more closely linked to the overall finish time. On the other hand, **Run 1** has the lowest R-squared value at 0.26, indicating it has less impact on predicting total race time. This could be due tot he fact that many runners run this pumped with race line adrenaline, and as such, it isn't an accurate reflection of their capability.

From this, we can infer that focusing on improving performance in the later runs and workouts could potentially lead to better overall race times, as these segments seem to have a more significant impact. Athletes and coaches can use this information to target specific areas for training, aiming to optimise performance where it counts the most for reducing total finish time.

That said, **it doesn't necessarily mean that you have to aim to run these faster**. In the next section we will learn that sub 1 hour finishers are actually spending a bigger proportion of their time on running that the finishers that are taking over an hour. So while we need to be aware that these later runs are a good indicator of your finish time, it may be wise to use this as a reminder that Hyrox is a great test on ones endurance. **Performance in the final run stages is therefore a great indicator of your overall ability.**

By the time an athlete reaches **Run 4**, we can make a fairly accurate prediction of their final finish time. While an R-squared value of 67% isn't overwhelmingly strong, it still provides a useful indicator. However, we wouldn't rely solely on Run 4's time to make predictions. At this point, the athlete has also completed several other stations, providing us with even more data points to improve accuracy. Proceed to the next page to use this predictor to estimate the final finish time based on the total time accumulated up to the end of Run 4.

Try out the Finish Time Predictor

Predicting Finish Times

Choosing the accumulated time taken to complete up to **Run 4** provides the most accurate estimate of the total finish time. This midpoint can serve as an effective target to guide you towards your final goal.

Additionally, it is a great tool for spectators keeping track of friends and family.

Select the time that you finish Run 4: 00:27:55

Estimated Finish Time

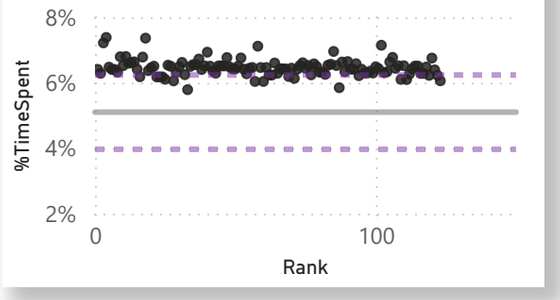
01:06:05

Work Times: <1 hour finishers

Sub 1 hour
Yes

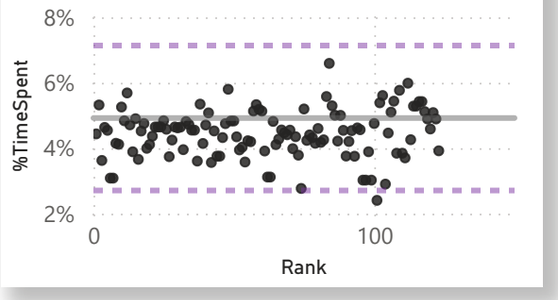
This page shows the proportion of time athletes spent on each work station, focusing on those who finished the race in under one hour. The purple dotted lines indicate the range where 90% of all finishers' times are expected to fall, providing a benchmark for identifying what would be an unexpectedly low proportion of time.
You can filter the data in the top right corner to view sub-1-hour finishers, non-sub-1-hour finishers, or all finishers. By selecting sub-1-hour finishers, the analysis reveals which stations these faster athletes spent a lesser proportion of their time on. For example, sub-1-hour finishers tend to spend significantly less time on the Sled Push/Pull, Burpee Broad Jump, and Sandbag Lunges stations. This suggests that these are "time saving stations" and that reducing your time on these stations (more training) could be key to achieving faster overall race times.

1 - Ski Erg



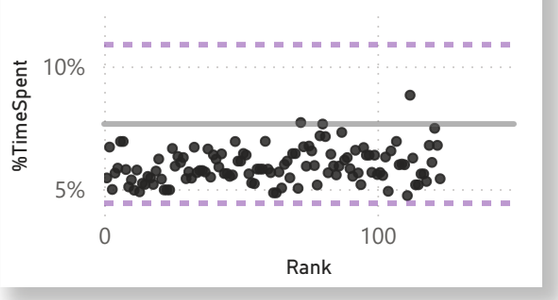
0
Exception Count

2 - Sled Push



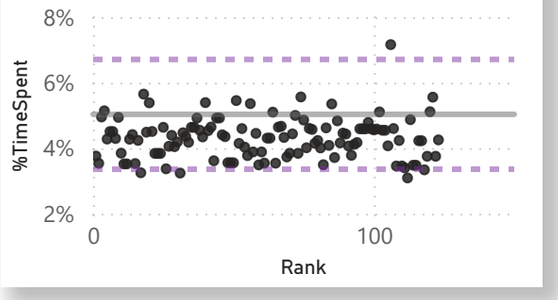
1
Exception Count

3 - Sled Pull



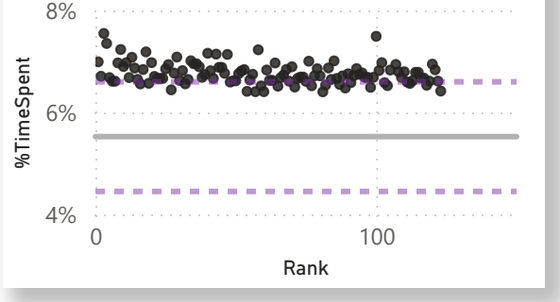
0
Exception Count

Burpee Broad Jump



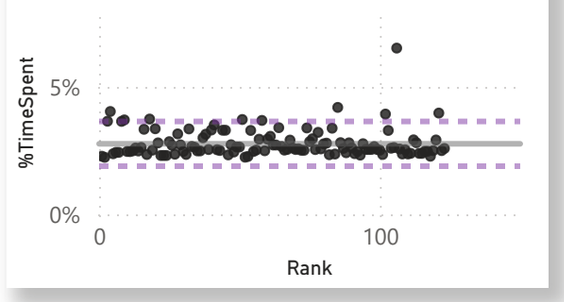
4
Exception Count

5 - Row



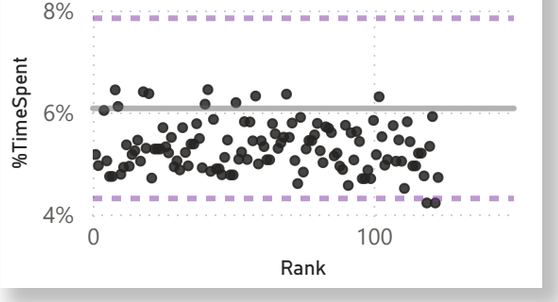
1
Exception Count

6 - Farmers Carry



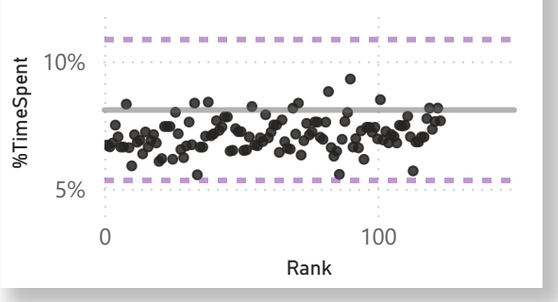
0
Exception Count

7 - Sandbag Lunges



4
Exception Count

8 - Wall Balls



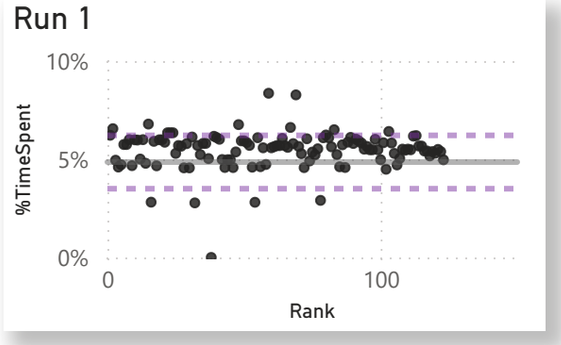
0
Exception Count

Run Times: <1 hour finishers

Sub 1 hour
Yes

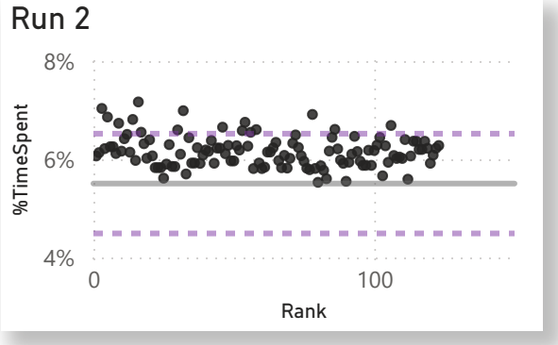
This page illustrates the run times for athletes who completed the race in under one hour. The graphs display the percentage of total time spent on each run segment by these sub-1-hour finishers. The grey lines represent the average percentage time spent across all runners in the dataset, while the purple dashed lines indicate the threshold range where 90% of all finishers' times are expected to fall.
It's important to note that the averages and thresholds are calculated based on the entire dataset, but only sub-1-hour finishers are shown here.
Observing these graphs, we see that sub-1-hour finishers tend to cluster above the average line on most run segments. This suggests they spend more proportionate time on these runs compared to their peers. For athletes looking to improve their performance and aim for sub-1-hour finishes, targeting the work stations rather than the run stations could be the way to go.

Run 1



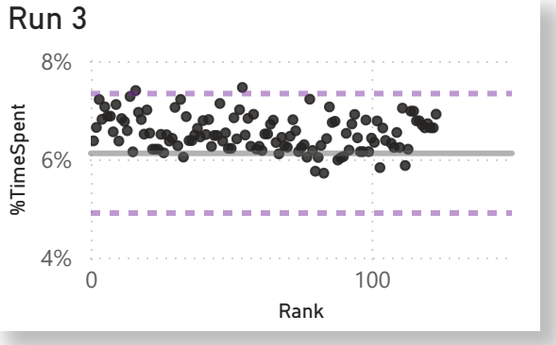
5
Exception Count

Run 2



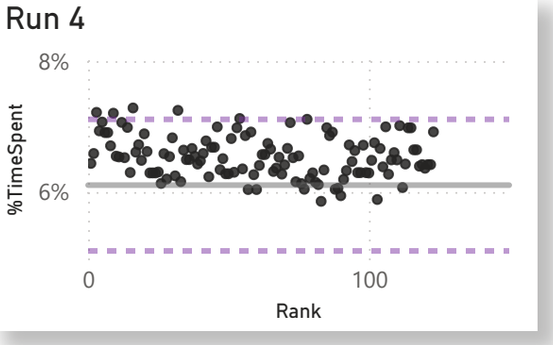
0
Exception Count

Run 3



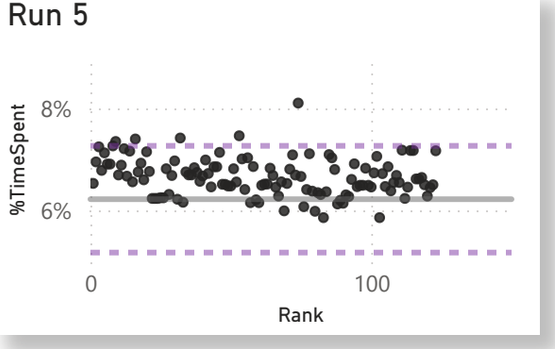
0
Exception Count

Run 4



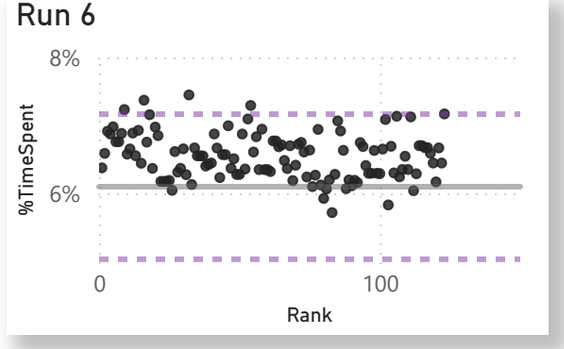
0
Exception Count

Run 5



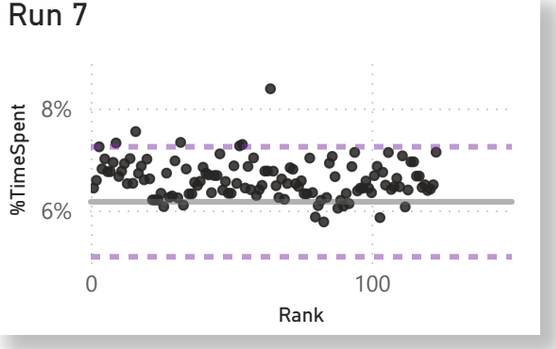
0
Exception Count

Run 6



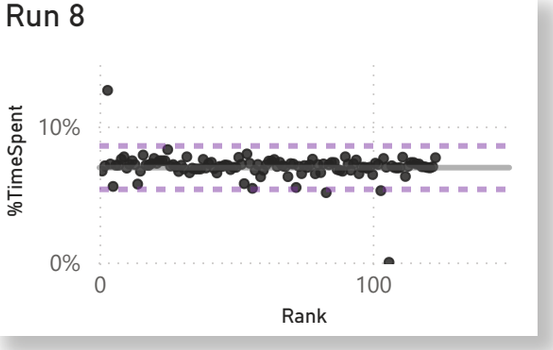
0
Exception Count

Run 7



0
Exception Count

Run 8



3
Exception Count