playerdata

101 George Street,
EH2 3ES, Edinburgh,
United Kingdom

Data Scientist

Technical Task

Aim:

Convert raw GPS data into meaningful metrics that a coach would be able to understand (Total Distance, Distance at Speed Zone 5, and Top Speed).

Instructions and details:

Generate a leaderboard of athletes, for the stated 3 metrics.

- Speed Zone 5 is seen as 19.8km/h to 25.1km/h).
- Top speed is calculated as the highest speed achieved.
- Distance at Speed Zone 5 is cumulative distance achieved whilst running in said speed zone.
- Remove noise from the speed data.
- Filter the data using the pitch boundaries.
- Look into additional insights and metrics that could be found for the coach beyond what we have asked for, e.g what could be done using the ball data?

Dataset:

participation_id	*Time (s)*	*Pitch_x*	*Pitch_y*	*Speed(m/s)*

- 'participation_id' is a unique identifier for each athlete.
- 'time' is in seconds
- 'pitch_x' and 'pitch_y' are the positional data on the pitch for the athlete
- 'speed' is the speed of an athlete in metres/second (m/s)

The pitch ranges are as follows:

- 'pitch_x' from -52.5 to +52.5 (105m)
- 'pitch_y' from -34 to +34 (68m)

Deliverables:

- A GITHUB repo with all scripts used to analyse the data, and a README file with instructions on how to run them and any associated documentation.
- Short (<10 minute) presentation on how you approached the task, along with the leaderboard and findings of the task.

*Please include how long this task took you to complete, and how long you spent on each section (discovery/execution/presentation).