playerdata

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

Data Scientist Intern

Technical Task

Aim:

Convert raw GPS data into tactical metrics that a coach would be able to understand and utilise.

Instructions and details:

Generate 2 metrics using the raw GPS data.

1. Initially:
    - Remove noise from the data.
    - Remove data outside the pitch boundaries.
    - Split the pitch into 3 areas or more.
2. First Metric:
    - Find the number of High Intensity events within each of the areas you have created per participant.
    - A High Intensity event is defined as a participant holding a speed between 19.8km/h to 25.1km/h for more than 0.5s.
3. Second Metric:
    - Propose a metric derived from the available data that can offer tactical insights for coaches following a match session.
    - This metric should provide valuable information and does not need to use the previously defined areas.

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 an explanation of the tactical metrics.

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