JUNIOR DATA ANALYST TECHNICAL TEST

Scenario: A reputed football club (FC Test) are looking to explore the relationship between GPS and Wellness data collected over five pre-seasons. Your task is divided into three sections:

# GENERATE DATA

Your first task is to generate dummy data for the two templates, GPS and Wellness. FC Test have a squad consisting of 20 players and their high-performance staff have managed to collect some pre-season data over 10 weeks starting on the first day of year 2014, 2015, 2016, 2017 and 2018.

The columns and restrictions on the templates are as follows:

## GPS data

GPS data is collected daily. The columns and restrictions are:

1. Date
2. Player Name
3. Session Duration – in minutes ranging between 45 and 90.
4. Player Load – arbitrary unit ranging between 300 and 600. FC Test is of the belief that Player load has a medium - high positive correlation (r > 0.4) with distance covered.
5. Number of Accelerations – ranging between 50 and 200.
6. Total Distance – in metres with a maximum value of 10,000.
7. Rate of Perceived Exertion (RPE) – a value between 1 and 10 submitted by the player.

## WELLNESS data

Wellness data is collected daily. The columns and restrictions are:

1. Date
2. Player Name
3. Sleep Hours – absolute value. Use realistic values.
4. Motivation – scale of 1 to 10
5. Soreness – scale of 1 to 10
6. Player position – one of 5 values; GK, CB, WM, CM, CF

Save the data as two separate csv files.

# SQL QUERYING

Assuming the 2 files from A above are written into a SQL reporting database, write a script to:

* Show the average player Load per position.
* Show the proportion of centre forwards that cover more than 5000m in training at least 75% of the time.
* Show what day of the week players are most sore

To investigate the effect of sleep on RPE, create another table (a subquery), that shows Date, Player Name, Player Position, Heart Rate and Sleep Hours from the day before.

# VISUALIZATION

Using your skills and experience, build a dashboard to summarize data and further explore the relationships between GPS and Wellness data. Where possible, you are advised to add any notes that might help explain your vision and applicability of your analysis.

Please submit your solution over email. Your solution must contain:

* A python or R script used to generate the dummy data.
* Two csv files containing generated data for GPS and Wellness
* A file containing your SQL script.
* A report built with a data visualization tool such as Power BI or Tableau.