Data Scientist (Applied R+D)
Take home exercise

Introduction
Given the dataset provided with this exercise, you should examine trends and offer insights based on analysis. You may choose the analytical methods that are appropriate to the task.

The exercise revolves around providing insights for the facilities management of an office building based on Internet of Things data that represents sensor measurements recorded from interior rooms of the building. The data is sampled from two floors, one on top of the other.

The features collected include indoor temperature, humidity and CO_2 levels for various rooms within each space (iaq.json). The data is collected by multiple sensors, each denoted by its own *sensor_id* (as specified in JSON files). A separate JSON file is provided indicating the position of each sensor within the floor plans (floors.json). Limited outdoor data is also recorded, including temperature, humidity, and windspeed (oaq.json).

Facilities managers are responsible for trying to maintain indoor temperature and CO_2 levels at specific values to meet the building operation goals. The required operational values are outlined in the *Resources* section of this document.

Expectations
- Display knowledge of basic data wrangling and data modelling tasks.
- Considering the time limits of the exercise, prioritise the analyses to make meaningful suggestions.
- Demonstrate appropriate analytical suggestions and vary presentation style depending on the audience.

Steps
You are asked to perform an exploratory data analysis, and then to make suggestions to the facilities managers based on trends and suggestions that they should take into consideration.
1. Document your thought process, assumptions, analysis, and results in your Jupyter notebook. Use Python or R.
2. Prepare and check the data to make sure it is ready for forecasting (forecasting is not part of the exercise).
3. Present a narrative describing your method and results to a non-technical audience that might include facility managers and other stakeholders. Export your presentation in PDF format; this can be based on PowerBI, Tableau dashboards or similar platforms. Please include narrative where appropriate.
4. Add all deliverables to a public or private Github repo. If using a private repo, invite ardgithub@fosterandpartners.com.
5. Send a link to the repo to ardgithub@fosterandpartners.com.

Deliverables
- Jupyter notebook with code and documentation.
- PDF file that includes presentation of findings to non-technical users.

Time estimate

5-6 hours depending on experience level.

Resources

1. **Indoor air quality data [iaq.json]**
   The data is provided on a 5-minute interval for the period 23 JAN 2023 to 19 FEB 2023.
   JSON dataset for indoor air quality:
   https://fpardrecruiting002st.z33.web.core.windows.net/CodeEvaluation/DataScientist/iaq.json

Building operation goals for indoor air quality - winter months		
Type	Operation goal during work hours	Comment
Temperature	At or above 21°C	Sensors installed at ceiling height. Temperature at desk height is estimated to be 1.5°C lower than the recorded data.
Humidity	At or above 20 %	
CO2	At or below 1000 ppm	

2. **Outdoor air quality [oaq.json]**
   The data is provided on an hourly interval for the period 23 JAN 2023 to 19 FEB 2023 for the location of the building.
   JSON dataset for outdoor air quality:
   https://fpardrecruiting002st.z33.web.core.windows.net/CodeEvaluation/DataScientist/oaq.json

3. **Sensor locations [floors.json]**
   JSON document with sensor location for the 3rd and 4th floor, relative to the provided floor-plans' images size.
   https://fpardrecruiting002st.z33.web.core.windows.net/CodeEvaluation/DataScientist/floors.json

4. **Floorplans in PNG format for H-3F and H-4F – please refer to "floors.json" for mapping sensor locations**
   1. H-3F.png
      https://fpardrecruiting002st.z33.web.core.windows.net/CodeEvaluation/DataScientist/H-3F.png
   2. H-4F.png
      https://fpardrecruiting002st.z33.web.core.windows.net/CodeEvaluation/DataScientist/H-4F.png