Calculate Gas Consumption

Context:

In a steel manufacturing plant, steel coils are produced through a multi-stage process starting from molten metal generation, chemical composition adjustment, casting into ~12 m slabs, reheating, and sequential processing through roughing and finishing mills. These slabs are eventually rolled into coils ~1200 m in length and further processed via APL lines (e.g., CRM, APL, coil prep) to achieve final coil lengths of ~1500–3000 m.

The objective of this assignment is to **analyze the gas consumption per coil** using data given. Each production line contains a set of **gas flow sensors**, which vary in **number and type** depending on the production line (domain).

NOTE: In steel making, each step (for example melting, casting or rolling) is called a production line—in our data we call that step the 'Domain.

User had provided specific **formulas for calculating gas usage** from these sensors. The task is to:

- Apply these formulas to the sensor data.
- Aggregate the **total gas consumed per ProductId (coil)** per production line.

Input Table:

1. **Lengthbased_table- Timeseries Sensor Data**
    a. per-second readings for every sensor by coil and production line (domain)
    b. columns include:
    TagId- sensor name,
    Timestamp,
    Value- instant reading value,
    Date,
    unix_timestamp- Process start timestamp of the product,
    Domain- Production Line,
    ProductId- coil ID,
    ProductPosition- Coil's instantaneous start position wrt coil,
    Length- Total coil length

2. **Master UDM table- Sensor Metadata**
    a. details for each sensor in the given production line
    b. columns include:
    Identifier- Sensor Name (Tagid in Timeseries data),
    Description_EN,
    Description_Local ,
    Domain- Production Line,
    Type,
    Type_Parquet,
    Unit- Scientific unit in which the sensor value is being measured,
    Range_Lower_Limit,
    Range_Upper_Limit,
    Sampling_Frequency,
    Application,
    Technology,
    Sensor_Position_Relative- Position of sensor wrt a given reference point,
    Sensor_Position_Absolute- Position of sensor wrt the beginning of the coil

These tables let you filter out the gas-measurement sensors and aggregate their readings by coil and domain.

Output Table:

The output table is “metal_carbon_footprint” table.

productid	total_gas (Nm3)	gas_per_ton
Productid Number (coil)	Total gas consumption for the product (during all its process time in the line) with unit as Nm3	Total_gas divided by the weight of the product (to have the gas_per_ton)

Calculation Logic:

total_gas= Gas consumed by each coil= sum of all gas flow sensor readings during its processing.

= Value(GasFlowRate1_Z01) + Value(GasFlowRate1_Z02)+ .... + Value(GasFlowRate1_Z12)+ Value(GasFlowRate1_Z13)

For GN-CR-APL2 domain- There are 13 gas flow sensor which fetches gas flow value, and their names have some pattern as given below:
for eg.
GasFlowRate1_Z01 is called "GN_CR_APL4_Oven_Debiet_gas_Z01_ist" (as Identifier or Tagid) and similar for the other zones just change Z01 as Z02 and so on.

Steps to Complete the assignment:

1. Fetch the list of gas flow sensors(which are relevant for gas consumption) from Metadata.
2. Using the list of gas-flow sensors, pull their values from the per-coil time series. Calculate each coil's gas consumption by summing those readings during its processing, and write the results to the output table.

3. Create an IPython notebook that clearly outlines each step of the process. The code should be self-explanatory and well-structured. Ensure that the final output table is displayed at the end of the notebook

4. Bonus step: There can be missing timestamps in between, use the standard best practices to impute those missing gaps and then calculate the gas consumption per coil.

   NOTE: The Unit of "value" column in the lengthbased table can be seen using the metadata. For eg the sensor **"GN_CR_APL3_Oven_Debiet_gas_Z01_ist**" has unit of Nm3/h and the total_gas is in Nm3. Please take care of the unit while calculating the gas consumption.

PFA the Data Set: First try to use the complete Dataset: Complete_dataset
If unable to download then try with: sample_dataset