Airflow Summary

1. Airflow setup:
    a. Linux / Ubuntu -> No Linux, so Windows -> VirtualBox (with Ubuntu 20.04) o Vagrant.
    b. Follow installation steps: https://airflow.apache.org/docs/apache-airflow/stable/installation.html
    c. Specific steps:
        i. https://airflow.apache.org/docs/apache-airflow/stable/installation.html#installation-script
        ii.
2. Airflow:
    a. Two components:
        i. Scheduler:
            1. Run airflow scheduler
        ii. WebServer:
            1. Run: airflow webserver
                a. Go to your browser: http://localhost:8080
    b. It creates a DAG folder /home/<user>/airflow/dags (or dag)
    c. Create a file within /home/<user>/airflow/dags (or dag) and copy the following code: https://airflow.apache.org/docs/apache-airflow/stable/tutorial.html
    Note that t1, t2, and t3 are tasks as Bash Operator (it can be a shell command, script or whatever)

3. Practice:
    a. Each task functionality:
        i. (T0) Download D1: https://github.com/manuparra/MaterialCC2020/blob/master/humidity.csv.zip?raw=true
            1. bash_command='wget -O /tmp/datos/humidity.csv.zip https://github.com/manuparra/MaterialCC2020/blob/master/humidity.csv.zip?raw=true',
            2. (T2) bash_command='unzip /tmp/datos/humidity.csv.zip /tmp/datos/',
            3.
        ii. (T1) Download D2:
        iii. https://github.com/manuparra/MaterialCC2020/blob/master/temperature.csv.zip?raw=true
            1. bash_command='wget https://github.com/manuparra/MaterialCC2020/blob/master/temperature.csv.zip?raw=true',
            2. (T3) bash_command='unzip temperature.csv.zip',
        iv. Execution and test:

1. T0 >> T2 >> T1 >> T3
2. [T0,T1]>>[T2,T3]

v. T6:
    1. Reducing sample. Take just 40 rows from both files.

vi. Add one task more:
    a. T4: Extract from temperature, datetime and “San Francisco”
    b. T5: Extract from humidity, datetime and “San Francisco”
    2. [T0,T1]>>[T2,T3]>>T6>>[T4,T5]