

## Assumptions

- The majority of assumptions are already commented on in the main.py script, so, in this document I will briefly explain some common assumptions that are “global”.
- The whole project is made in python (with pandas and geopandas libraries, among others). We can export some csv to mongo or postgre for the queries but I decided to solve this task in the same script.
- All the tasks are made in the same script main.py.
- The code has a boolean Test, that lets me execute some pieces of code fastly (to execute the whole script Test should be set to False).
- I work with pandas in data cleaning and data normalization because I consider that it is easier to work with the latitude and longitude coordinates.
- I have decided to reduce the sampling frequency by “truncating” the datetime (minutes and seconds are the decimal values in this analogy). This is because reducing the sampling from minutes to hours reduces the quantity of points but without losing too much information.
- The normalization algorithm is well explained in the script, so I do not see the need to comment further.
- I normalize all the coordinates of the 4 datasets in order to always use the same reference system.