Data Engineering Challenge

Challenge

Demonstrate your data engineering skills:
Harmonize as many soil organic carbon samples as possible from the RaCa dataset in a single dataset. The RaCa dataset can be found and downloaded through this link:

RaCa Data Tables

Requirements

1. The harmonized dataset must contain **georeferenced** samples, meaning that each sample should have at least one pair of coordinates (latitude and longitude) associated with it.
2. Furthermore, each sample should have a reference **top- and bottom-depth**.
3. Avoid duplicate data.
4. Focus on the following columns:

From the RaCa samples:
'Bulkdensity','SOC_pred1', 'BDmeasured', 'BDmethod', 'Texture', 'fragvolc', 'c_tot_ncs', 'n_tot_ncs', 's_tot_ncs', 'caco3', 'Measure_BD',

From the RaCa pedons:
'SOCstock5', 'SOCstock30', 'SOCstock100'

Deliverables

Present a summary of the compiled dataset in a jupyter notebook or python script.
Please, send us your compiled dataset and the jupyter notebook / pythons script for review.

Looking forward to your submission, we thank you and wish you the best of success.