Data Science Challenge

Challenge

Demonstrate your data science and engineering skills:
Combine and harmonize the following RaCA datasets: **RaCA samples**, **RaCA general locations** and **RaCA SOC pedons** from:

RaCa Data Tables

Conduct an exploratory data analysis on the dataset, discover and present correlations and patterns in the data.

Requirements

1. Identify common keys/indices to merge the data from the three datasets.
2. Pre-process/transform the SOC pedons dataset (wrt. the sampling depth) to merge it with the RaCA samples dataset.
3. Merge and harmonize the three datasets, so that each entry is **georeferenced** (has a **lat/lon** coordinate pair) and has a **top and bottom depth** in cm associated with it.
4. Avoid/remove duplicate data.
5. Highlight some correlations between the data and distinguish between numerical and categorical variables. For a bonus: check for spatial autocorrelation of variables.
6. Focus on the following columns:

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

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

Deliverables

Present a summary of the compiled dataset in a jupyter notebook.
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.