Take Home Test

Science roles

Thank you for your interest in **Natcap**. This task is designed to avoid whiteboard coding or other artificial technical interview processes, and allow you to work in your natural science habitat!

Please spend **no more than 4 hours** on this task. You may not get all of it done. That's OK. The task is ordered deliberately: please don't skip part 1 and dive straight to part 2!

We will test your solution, so please provide setup instructions - for example, let us know which version of Python you are using. We're working on Macs, and can flip to Linux or use Docker. If you're working on Windows, we'll do our best!

When you're done, send an email back to us with a zip or tarball containing:

- Your script, .py or notebook form is fine
- Any supporting documentation you think we need
- A slide deck of the outputs from Part 1, 2, and 3
- A recorded presentation of the slide deck produced in Part 1, 2, and 3; details outlined in Part 4

A few pointers:
- Please complete this task using Python. We use Python at Natcap and cannot provide support for production code written in other languages
- We know AI (ChatGPT in particular) is now a useful tool for coding and design documents, and it's absolutely fine to use AI to help you
- If you've never done this kind of task before, start by looking at the Python libraries [rasterio](), [shapely](), and [GeoPandas](); you mat also want to take a look at [GDAL]().

And with that, here are the tasks…

Land use change is a major driver of biodiversity loss. There are many data products which can help us measure and monitor land use change. Environmental Systems Research Institute (ESRI) produces an annually classified land cover product, which gives land cover classifications for 9 different land cover classes globally at 10m resolution. Their data can be viewed here.

We have taken some of these data and extracted an area which covers Spain. You can download the data you need for this task from this Google Driver folder (spain_2020.tif, spain_2021.tif, spain_2022.tif). The README.txt file contains a dictionary for the landcover classes.

This is an image dataset in tiff format, specifically GeoTIFF. We need to manipulate this data to extract particular regions or sites that we are interested in — "areas of interest" — defined by a vector polygon in a GeoJSON.

In the Google Drive folder, within folders "Part 1" and "Part 2" there are some example GeoJSONs of interest that we will use to test your script, and that you can use to try out your code.

Part 1

We are working with a client that operates multiple sites across Spain. They have identified one site (id_1), acquired in 2020, for a detailed analysis. The sustainability team has asked us to assess land cover changes since their acquisition, with the goal of presenting the findings to their leadership team.

Please create a deck containing 1-2 slides that gives the following details:

1. A summary of the land cover at id_1 in the year they acquired the site (2020), for example the total area of each land cover class, and the total area of all natural lands vs non-natural lands
2. Details of how the land cover has changed since they acquired the site, including differences in the area of different land cover classes between 2020 to 2021, and 2021 to 2022

3. At least one type of visualisation (please feel free to do more) which shows how the land cover has changed over time, what what might be driving that change

Part 2

The client would now like us to do the same analysis across all of their sites in Spain. Adapt the script you made for Part 1 to run across all the GeoJSONs in the folder called Part 2.

1. Save this as a separate script so we can see your changes from Part 1
2. Add 1-2 slides to the deck you made in Part 1. These new slides should highlight any sites (GeoJSONs) that may be of particular interest to the client

Part 3

If you were to apply your script to a global landcover dataset, made of hundreds of landcover tiles and thousands of sites (GeoJSONs), how would you change it?

1. Don't write any code for this part!
2. Add 1 slide to the previously made deck on what challenges you might face
3. Add another slide to the deck summarising the kinds of tools, libraries or approaches you would use to do this

Part 4

Record a presentation of the slide deck you created in Part 1, 2, and 3. This should be **no more than 5 minutes**, and is your opportunity to talk through the results, as well as expand and explain the statistics and visualisations chosen, and the challenges and changes you would make in Part 3.