



# Data Scientist

## Challenge Project

Please use any open source software or tools available to complete this project.

We encourage you to work in a similar way as you would in a professional environment to produce what you consider to be a production-ready project.

Enclosed with this description are two scenes captured by ESA's Sentinel 2 (scene 1: S2A\_MSIL1C\_20230223T112111\_N0509\_R037\_T30UVF\_20230223T145910 and scene 2: S2B\_MSIL1C\_20230102T113359\_N0509\_R080\_T30UWG\_20230102T121121), processed by Open Cosmos. The images are in the GeoTIFF format and contain all the necessary metadata. The names of the files correspond to the [Sentinel 2 bands](#). The rights to the images remain with the original holder and you must delete them when the project is completed. In addition, a GeoJSON file is provided containing points that identify inland water bodies visible in the image with an ID.

We expect you to spend around 4 to 10 hours on this project. Don't worry if you don't finish it, just be prepared to explain what you would have done with more time.

### **Deliverables**

You must provide the source, configuration, and instructions required to run the project. Assume that the reader has fundamental knowledge of software development and data science but may not have prior experience with the tools you choose to use.

The raster and vector output files must be encoded with a format supported by GDAL "built-in by default" with "geo-referencing support". See the supported lists: [supported raster formats](#) and [supported vector formats](#).

Finally, you are asked to submit a brief document or a set of slides that explains your interpretation of the project, the decisions you made, and the reasoning behind them. You should also outline the next steps you would take if you had more time to work on the project.

## Project description

The image provided contains all the bands from the Sentinel 2 satellite on the level 1C. The area in the image contains land, sea, and inland water bodies. In addition, a GeoJSON file is provided containing points that identify inland water bodies visible in the image with an ID.

The provided solution should be able to execute from start to finish to generate the results with no manual steps or person input required. The solution should also be able to run on any provided Sentinel L1C image and corresponding GeoJSON and generate reasonable results.

The objectives of the project are to:

- Crop all bands from the first scene to the following bounding box: [-3.480290297664652,54.26510479276385,-2.9010711619639267,54.61995328561707] and the second scene to the following bounding box: [-2.815247,55.102730,-1.450195,55.553495] (expressed in lng/lat format in EPSG:4326 - WGS84, located in the UK).
- Calculate the NDVI and the NDWI on the cropped images. Deliver both indexes for each scene as a part of the project.
- With or without the use of those indexes, extract the inland water bodies shapes into a vector format.
- Using the GeoJSON file provided, assign the corresponding ID to the water body shapes identified and remove the ones that do not have an ID. Deliver the final result in a vector format.

## Evaluation

We will be evaluating the following:

- That each of the outcomes meet the objectives, and to what extent they do so.
- The code or configuration itself. How clear, readable, and maintainable it is.
- The reasoning provided for each decision taken, including tool selection.

Remember that this is an exercise for you to shine and show your best skills:

- Any additional contribution or idea that you wish to implement will be positively valued.
- You can use any tool(s) of choice for any of these additional contributions that you would like to make.

## GeoJSON

For S2A\_MSIL1C\_20230223T112111\_N0509\_R037\_T30UVF\_20230223T145910:

```
{
  "type": "FeatureCollection",
  "features": [
    {
      "type": "Feature",
      "properties": {
        "id": "Lake A"
      },
    },
  ],
}
```

```

    "geometry": {
      "coordinates": [
        -3.1490311933004023,
        54.57730915429053
      ],
      "type": "Point"
    }
  },
  {
    "type": "Feature",
    "properties": {
      "id": "Lake B"
    },
    "geometry": {
      "coordinates": [
        -2.934163125692578,
        54.379649943576
      ],
      "type": "Point"
    }
  },
  {
    "type": "Feature",
    "properties": {
      "id": "Lake C"
    },
    "geometry": {
      "coordinates": [
        -3.070901438647411,
        54.345261884865664
      ],
      "type": "Point"
    }
  }
]
}

```

For S2B\_MSIL1C\_20230102T113359\_N0509\_R080\_T30UWG\_20230102T121121:

```

{
  "type": "FeatureCollection",
  "features": [

```

```
{
  "type": "Feature",
  "properties": {
    "id": "Lake D"
  },
  "geometry": {
    "coordinates": [
      -2.4840234328177644,
      55.18324796070513
    ],
    "type": "Point"
  }
},
{
  "type": "Feature",
  "properties": {
    "id": "Lake E"
  },
  "geometry": {
    "coordinates": [
      -1.6251065778267844,
      55.282989205515236
    ],
    "type": "Point"
  }
}
]
```