

# Digital Apprenticeship Program



## Data Science/ Data Analytics challenge

Consider a dataset providing information on the functionality of infrastructure resources, for each water point it includes the name of the village it is in and its functional state. Implement a data processing module in **python** which takes a dataset URL as input and returns:

- The number of water points that are functional,
- The number of water points per community,
- The rank for each community by the percentage of broken water points.

There should be a top level object or function calculate(“http://...”), which returns a data structure with the above information, something like:

```
{
  number_functional: ...,
  number_water_points: {
    communityA: ...,
  },
  community_ranking: ...
}
```

But that’s just a suggestion and we can think of alternative, maybe better, data structures to use.

We will use a water point dataset that contains many columns, the relevant ones are:

```
communities_villages, water_functioning
```

The data can be found on this [URL](#)

When your solution is complete, upload it to github (if you do not have a github account you may create a free one). Submit your solution by sending us a link to it on github.