



Data Science Challenge

In this challenge, you will face a real problem we have at Trialing. You will have to get in the shoes of a data scientist at our startup to get the data we need.

You will have to make a Python script to **scrape** data from this website: <http://trialing-ds.s3-website-eu-west-1.amazonaws.com/>

Information you need to get

From this website, we need a CSV file with the following columns:

- name
- address
- latitude
- longitude
- country
- region
- city
- contact data

You can use your preferred Web Scaping library, but it should be done in Python.

Data Analysis

After getting this information, we need an analysis of clinical trials. You have a file **hospital_trials.csv**. The first column has a code for a clinical trial, the second column has the ID of the hospital. Please bear in mind each line is an association, so there can be multiple lines for the each clinical trial ID.

We want to answer this question: **What's the number of clinical trials per region?**
Please, use the plot/graph that you consider more appropriate. You can use your preferred system to show this: Python, Excel, PowerBI...

Hint: You don't know how to associate Hospitals to their ID? Look better in the website! 🤖

Deliverable

Once you've completed the challenge, please upload it all to a git repository (it can be GitHub or any other). Please, create a ReadMe explaining how to run your code.

Make sure to add your code as well as the results!



Advices

This should not be a difficult challenge. Clean the data, keep calm and don't get overwhelmed! It shouldn't take you more than a couple of hours to complete it. It's worth trying it!

Also, don't worry if hospital names, addresses and coordinates are don't match. Data has been shuffled and it's not part of the challenge to complete that. Assume it's correct.

Good luck!