

Data Analyst Challenge

You are provided with a historical dataset about Covid 19 cases worldwide for analysis in whatever format you choose. The observations provided span from 22/01/2020 to 29/08/2020. You are free to illuminate whichever insights interest you and apply your mathematical, programming, and/or data science knowledge.

The only specific challenge is to prepare and/or use the data to forecast historical results after 29/08/2020, for instance for a particular country, and to do your own additional research to determine if and when these forecasts are accurate. This can be traditional or modern methods.

Data

The data is available in CSV format from the link below.

<https://drive.google.com/file/d/1gWzXnXiByv4b88C9MnuwqspSD3tRtfxO/view?usp=sharing>

Format of Solution

Once you've downloaded the data, you can analyse and present it in whichever way you are most comfortable with. Examples may include a Python or R script, a JuPyter or other Python Notebook, an Excel spreadsheet, etc. Any dependencies should be included or installed in the Notebook, a requirements file, or an accompanying README.

What We're Looking For

We're principally interested in how you approach the data and present your analyses. The data will present various challenges, including organising and reformatting the data into a usable state, data enrichment, visualisation, and forecasting. Explaining your working with thorough and explanatory comments is a must, and if an additional README would help, then one should be provided. You do not need to send back the data, only any additional files you have created, but it should be clear where the data should be located such that the solution can load it.

We will be looking for analyses beyond the specified task, which could consist of descriptive analytics and visualisations, hypotheses of your own being tested, or additional modelling for e.g. prediction other than the forecasting task described.

Effort

You are being asked to tackle this challenge in your own time, so be mindful of how long you spend on it. The task overall should be scheduled across one week, but how much of that week you put into it is up to you. However this isn't a pass/fail test. There's no right answer! We're interested in technique and approach: how you prepare the data, how you apply basic data analytic techniques; how you generate and test hypotheses, hence the free reign. Eight hours is a reasonable upper limit to try and stay within; a few hours a challenging lower limit.

Your Feedback is Valued

This is a first draft of this particular technical challenge: it may evolve a little over time. Is it too easy? Too difficult? Could you do with a few more pointers? Were there any unnecessary time-wasting obstacles? Was the data too easy or too complex to work rapidly with? Whatever happens, and whether you finish it or not, we'd appreciate your feedback on it so that we can improve how we evaluate candidates in future. Many thanks.