Research Scientist – Pandemics

Applicant assessment

*This assessment is designed to provide you an opportunity to demonstrate analytical thinking and technical skills relevant to the Research Scientist role on the Pandemics team at IHME. There is not a single “correct” answer; rather, use the challenge below to show how you would approach a problem, and develop an initial attempt at implementing your proposed methodology.*

# Instructions

You have 24 hours to complete this assessment; it is likely to take 3 – 4 hours. Use the final question below to indicate areas for future development or improvements to your approach.

Use the data attached to answer the questions below. For each question please provide the following:

1. A short written answer to the question; this can be typed in a separate document.
2. Where applicable, visuals to show your results, as described in each question. This can be a separate file for each question, but please be conscious of file size and navigability (i.e. do not create a separate file for each location).
3. The code used to create the results and figures –either delivered via a GitHub repo, or attach the files to the submission.

# Prompt

You have been provided actual data from the covid-19 pandemic, as reported by state and national governments. These data have already had spikes or errors in reporting corrected and redistributed. Attached are data from the US (excluding Washington state):

* Cumulative confirmed (reported) cases of Covid 19, by location by date
* Cumulative reported deaths of covid-19, by location by date
* Cumulative hospitalizations for Covid-19, by location by date (NOTE: hospitalization data is not available for all 50 states, only some)
* Population of each location

These are the only 4 inputs that IHME uses for our first-stage deaths model, which produces 14-day forecasts that are then used in an SEIR model.

# Questions

* 1. What is the relationship between cases, hospitalizations, and deaths? Describe how these indicators relate to each other, and visualize the relationships in at least 2 different ways.
  2. Fit a curve of daily deaths, utilizing these inputs. Describe the approach you used and visualize the results.
  3. Create projections for 14-days after the last observed data point. Visualize the result. Describe the benefits and limitations of your approach. Where do you think this approach has performed particularly well? What types of situations cause your model to struggle?
  4. Lastly, describe future areas of exploration or improvement for your approach. If you had more time, what would you do next?