

TECHNICAL / PROGRAMMING TEST

FIELD EMPIRICS, RESEARCH VERTICAL

Assignment:

You have been provided with two abridged datasets from FiveThirtyEight's COVID-19 tracker. The datasets contain a list of opinion surveys fielded in the United States.

- [Dataset1](#) has opinion polls on the Trump administration's handling of the coronavirus pandemic.
- [Dataset2](#) has opinion polls measuring respondents' concerns, related either to the economy or infection from COVID-19.
- Kindly refer to the [Readme](#) for more details on the variables.

Task 1:

Your task is to familiarise yourself with the datasets, perform any necessary cleaning, and combine them into one dataset.

Before combining the datasets, here are a few things to keep in mind:

1. Since we are reporting aggregate statistics, we need to be able to weigh the responses from each poll, to generate a weighted average. This means that we should disregard polls that do not report sample sizes or report missing values in the response options. By response options we mean the variables that capture responses to the opinion polls.
2. For the purpose of this exercise, let us assume that our previous data indicates that a survey that has lower than 200 responses will be statistically underpowered, so please report results only for eligible studies.
3. Please ensure that the final dataset includes indicator variables that clearly mark which of the two original datasets the row originated from.

Task 2:

Once you have combined them, please generate the following outputs:

1. A summary table to display the frequency of approval polls and concern polls conducted by each pollster



2. Aggregated proportions of respondents approving of Donald Trump's performance, delineated by whether the poll was asked to Democrats, Republicans or Independents.
3. Aggregated proportions of respondents who are very worried or somewhat worried about the economy
4. Aggregated proportions of respondents who are very worried or somewhat worried about the infection from the coronavirus

Kindly share the following:

- 1. All your final results in a single PDF.**
- 2. Code for this exercise (starting from inputting the dataset to outputting the final results, along with all the mark-up/comments) - either as R file or a Do file**