



Data Scientist exercise

Thanks for your interest in the role, and congratulations on passing our initial sift.

As part of the next stage of our hiring process we like to ask candidates to do an exercise to give us a sense of how you think and how you approach data. This task is fairly typical of the kind of work you'll be doing here at Stack.

We'll confirm the deadline for submission when we schedule the interview with you.

Task description

Description

We would like you to use a simple MRP (Multilevel Regression and Poststratification) model to estimate 2019 UK General Election results:

1. Download population counts for interlocked demographic groups resident in each UK Parliament constituency in England and Wales from the "Create a custom dataset" page of the ONS website. This will be your poststratification frame.
 - a. The demographic categories included should be age, highest level of qualification and sex
2. Use your choice of publicly available data from the [British Election Study](#) to estimate the percentage of people voting for the Conservatives, Labour, Liberal Democrats or Other parties in each cell of your poststratification frame.
 - a. **Hint:** To correctly estimate support across demographic groups, you will also need to estimate differential turnout rates. The BES Random Probability Survey might be a useful resource for [this](#).
 - b. You are welcome to use any publicly available data or resources to help complete this task.
3. Scale your estimates so that the estimated results in each constituency match the true results of the 2019 General Election in that constituency. The results are also available on the British Election Study website, [here](#).

Please submit the final version of your poststratification frame, complete with estimates of 2019 voting patterns, along with the R scripts or Python code you used to construct it. This can be in the form of a Github repository or Google drive/Dropbox folder.