As part of your evaluation as a candidate, we’d like you to work through a data problem (questions below).

Your overall goal for this exercise should be to produce results that are defensible and

appropriate to the problems. However, we do not expect you to deliver the perfect solutions to these problems. **Instead, you should aim to deliver work that you’d be comfortable implementing as a proof of concept, not a final product**. In general, approaches should balance quality and thoroughness, with expediency.

The dataset you will use is a public-use datafile from the Centers for Medicare & Medicaid Services (CMS) on Nursing Home Quality Measures. The file contains quarterly metrics about how healthcare providers performed on various industry-standard quality measures. You should use the csv contained in the zip file sent with this project. But, if you’d like further information on the dataset, you can find it in its public location:

https://www.healthdata.gov/dataset/mds-quality-measures

Note that we chose this dataset because it contains real and contemporary data; it has not been pre-processed at all. You should treat it skeptically before performing any analysis.

**Below are the tasks we’d like you to complete for this exercise:**

**Question 1.** Create a function that calculates the proportional change in a given Measure from one quarter to any future quarter. By default, this function should return results at the provider level. It should also be able to calculate the proportional change in aggregate Measures at the national, state, and ZIP code levels. When aggregating, the function should calculate the proportional change in the average of each group.

**Question 2.** Please code a model that solves one of the two following problems[[1]](#footnote-1). Your solution should only include data from Massachusetts providers; you can ignore data for other states. For the problem you choose *not* to solve, do not prepare any material, but please be ready to discuss potential approaches.

1. **User Story A:** As the head administrator of a hospital in Massachusetts, I want to use this data to forecast each of my providers’ expected Measure Scores for Q4[[2]](#footnote-2). I’ll then use this forecast to assess my providers’ Q4 performance relative to expectations.
   * As the athenahealth Data Science team, we would like to create a model that meets this client’s immediate needs. The model should output a forecasted Measure Score for each provider in the dataset.
2. **User Story B:** As the athenahealth User Experience (UX) team, we want to use this data to group our clients’ providers into similar “Provider Groups.” We’ll use this information to tailor our EMR interface for each Provider Group, in order to create a more personalized experience.
   * As the athenahealth Data Science team, we would like to create a model that supports the UX team’s needs. The model output should assign each provider in the dataset to a group.

**Question 3.** Please be prepared to *explain* how your code could be adapted to:

* use a future version of the dataset, covering a different time period (perhaps future quarters of data); and,
* handle the addition of a new provider-level column, such as “Time since medical license granted.”

**Logistics:**

You will be presenting your completed exercise to members of the Data Science team at athenahealth as a code walkthrough and discussion. You will not be giving a formal presentation, so there’s no need to prepare a separate PowerPoint or any other special content. Be prepared to review your code, its output, and the answers to the questions below in any format that you find easy to reference during the discussion. It’s perfectly fine to show a Jupyter Notebook, a session of R Studio, or a script. Note that while you’ll be going through your code live, you should not plan on *running* it live. Given that, if you’re going to present your code in a script, be sure to save the output beforehand, so you have it readily available to show during your presentation. You can expect the audience to be technically sophisticated.

1. Athenahealth HR will send you a copy of these prompts/directions and a zipped .csv data file 8 days before your in-person interview (one full week before it is due). The project is due via email to your HR contact by ***9am EST the day before*** your interview. The reason for this is so we have enough time to address any potential technical hurdles prior to your arrival.
   1. Note: even if you plan on presenting with your own computer, all your solutions are due at 9am EST on the day before the interview. All your content should be with your HR contact at this time.
   2. In cases where there are unavoidably less than 8 days between your phone screen and your in-person interview, we commit to getting you the materials as soon as we can, and then adjusting our deliverable expectations per that truncated window.
2. As soon as you receive the materials, make sure you can open all the files immediately, to ensure nothing was corrupted in transmission. If you have any trouble, alert your HR contact immediately.
3. The day of your in-person interview, there will be one 60 minute session with 3-6 members of the Data Science team where you’ll walk through your solutions - again, no formal presentation is necessary. As you’re describing your work, the audience may ask some specific questions, and a period will be reserved at the end of the session for further questions about the project as a whole.
4. It is your choice how you’d like to display your work during the interview. There are two possible ways:
   1. We want you to be comfortable; so you should f**eel free to bring your own computer to the interview.** Our projectors are compatible with VGA and HDMI outputs (other types can be arranged with sufficient notice).
   2. If you’re either unable to bring your computer, or you’d simply prefer not to, you’ll be able to use one of the DS team members’ computers. You can expect all your delivered files to be downloaded onto a Mac, which has Sublime Text, TextEdit, or PyCharm (Community Edition) installed.
5. To submit your finished product (again, due at ***9am EST the day before*** your interview), send an email with all the materials you produced back to your HR contact at athenahealth. In that message, please also state how you’d like to present your work.
   1. For example, if you completed the project using a python script, we’d recommend sending:
      1. the .py file,
      2. the saved output from that script as a text file, pdf, or other,
      3. and which computer you’d like to use during the interview.
         1. If you’re planning to bring your own computer in, please specify your output requirements (HDMI, VGA, other).
         2. If you’re planning to use one of our computers, then additionally specify your preferred IDE for any code you’d like to show. We’ll do our best to accommodate your needs.
   2. If you completed the project in a Jupyter Notebook, or a similar browser-formatted editor, then we recommend saving your session to a .html or .pdf file, and sending that in as well.

If you have questions, please use your judgement and document your assumptions. Good luck!

1. These problems are written as “user stories”, which is a concept from agile software development. Look it up! [↑](#footnote-ref-1)
2. Use Q1-Q4 as defined in the dataset; note that each provider’s Q4 may be a different period in time. [↑](#footnote-ref-2)