Data Challenges

Successfully completing a data challenge is like the MVP of getting a job offer! *And the best way to get ready is to do as many practice data challenges as possible. There's no substitute for it: once you sit down with the prompt and the empty IPython notebook, you force yourself to think about all of the things that could go wrong with your analysis. You internalize the steps of the data challenge (1: State the problem. 2: Download the data. 3: Explore the data. 4: Clean the data. 5: Choose features. 6: Scale features, etc.). You also force yourself to actually look up some of the modeling packages in scikit-learn and use them for real -- you don't want to take up valuable time learning/debugging during an actual, timed data challenge. Finally, you will start recognizing common data challenge problems and will quickly get to the strategies that could be most helpful (because you will have solved similar problems before!).*

Preparing for the challenge

- Gain as many clues as you can from your interviewer
    - What do they want (you) to learn from the data?
    - What are some things that the company really cares about?
    - Clarify the audience for your report -- is this for c-level or technical teams?

Working with the data

Concentrate on 1) ingesting data, 2) cleaning it up, 3) implementing a model, and 4) interpreting your findings. Hitting all of these marks should be your goal, even if it means you can only work with a subset of the data.

- Use your time wisely -- Have a clear goal in mind and ensure you are working toward that goal
- Use the language you are most comfortable with
- Prepare simple visualizations of the data to better explain your results
- Work smart -- it may be more efficient to do simple calculations in Excel, create visualizations in R, etc.
- Your code should be reproducible

Getting ready to present

Before you present, practice being in the moment. You don't have to fake it til you make it -- you're already a data scientist, you've made it -- but a bit of role playing could help.

- Pretend that you are three months into your job at the company, and you've been asked to solve this data problem for your team.
  - How would you go about tackling the problem?
  - How would you interact with the 'audience' evaluating your findings?
    - Treat them as your colleagues, rather than as people testing you

Presenting your findings

- Briefly mention what you did to clean up and/or explore the data (summary statistics, correlations, etc.)
- Emphasize the bottom line, the value these insight could provide in terms of money or time saved
- Provide justification/demonstrate sound judgment for the choices you made
- Deliver your conclusions confidently
  - Do not focus on uncertainty or caveats inherent in the test/method (but discuss these if asked)
- Be prepared to discuss alternative approaches you might have taken
  - For example: ML vs. statistical/probabilistic vs. descriptive approaches

If you're asked to summarize your findings in writing:
- Use an outline similar to a technical paper -- introduction, methods, findings -- to help you get started
- Include simple visualizations, if you had time to make them
- Write in complete sentences and make sure to run a spell-check!