

Newsela Senior Analytics Engineering Take-Home Challenge

Hi there! Welcome to the take-home portion of our interview process.

This challenge is designed to simulate a typical task you might encounter on our team. We estimate it should take no more than couple hours to complete. We'll be evaluating your submission for the following:

- the technical approach of the solutions
- the structure, readability, extendability your SQL
- query performance and cost considerations
- data quality of your results

During the follow up Technical Interview please come prepared to discuss your solution and answer follow up questions about your submission.

Good luck! We're excited to see your work.

What will you need?

- A Github account
 - We ask that you return your submission in a hosted Github public repo, which should be accessible by users in the Newsela Github Organization
- A Google account
 - For this challenge you'll need to be able to access a public dataset hosted by Google. The dataset is free to query, but you'll need to be authenticated with Google to access it.

The Task - Stack Overflow Post Analysis

Take a look at this [public BigQuery Dataset of stackoverflow posts](#).

Please provide queries which answer the following prompts

1. What tags on a Stack Overflow question lead to the most answers and the highest rate of approved answers for the current year? What tags lead to the least? How about combinations of tags?
2. For posts which are tagged with only 'python' or 'dbt', what is the year over year change of question-to-answer ratio for the last 10 years? How about the rate of approved answers on questions for the same time period? How do posts tagged with only 'python' compare to posts only tagged with 'dbt'?

3. Other than tags, what qualities on a post correlate with the highest rate of answer and approved answer? Feel free to get creative

How to return your submission

1. Please organize your solutions in a Github repo which is accessible by members of the Newsela Github Organization.
2. Please include a README file which explains the results of each prompt as well as an explanation of your technical approach to each prompt.
3. The repo should also contain the executable queries which were used to obtain your results. Please include code comments in your queries as needed.
4. Please email a link to your repository to carlie.copeland@newsela.com
5. We ask that you please complete and return this exercise within approximately 1 week from the day you receive it.