

General instructions:

This task requires geospatial reasoning and has a time dimension to it.

- Please only use Python
- Clarification questions are strongly encouraged (please email Miki Uryszek at muryszek@synmax.com).
- After code submission, we will have a technical round where you will walk us through your code and reasoning. We will ask follow-up questions.
- You are free to use Google, Stackoverflow, LLMs, and any other resource you can find online.
- For submission, please submit your code and output files to muryszek@synmax.com.

Task Background:

The fracking revolution in the US in the 2010s has resulted in a huge increase in gas and oil production, mainly because the technique allows for hydrocarbon flow to be stimulated directly from the source rock formation, rather than requiring direct connection to a conventional reservoir. Such 'unconventional' wells have a drastic production profile where the output is greatest in the first few months and then declines sharply. The production of a particular well can depend on many variables: source rock formation, technology used to drill/frack the well, length of the wellbore, previous fracking activity in the area, density of wells in the area, and many other factors.

Task 1:

You are given three datasets:

1. well-level production ("ds_interview_well_prod.parquet"): Reported gross production by well API. Each well reports its monthly total production for oil, and gas.
2. well properties ("ds_interview_well_props.parquet"): Properties of individual wells by well API, like horizontal length, surface lat/lon, bottom lat/lon, depth, etc.
3. well formations - ("ds_interview_formations.parquet"): Rock formations for selected wells.

Your task will be to forecast individual well declines for wells found in the state of Texas up to and including February 2025 (for all wells found in the production file). The data comes from the respective state regulator who receive it from the operators themselves. Famously, this data is of poor quality, e.g. some wells will have a last reported production in June of 2024 but they're still actively producing to this day, not all wells will be 'unconventional' wells and some of them might not be wells at all. It's advised to do your own research about typical declines curves and use that knowledge to check the quality of the data.

While this forecast is at the well level, often we care about the aggregate production. Please calculate the state totals (along with uncertainty estimates) for the following time period (September - November 2024), consider how you'd sanity check your production totals.

Task 2: Deployment

You do not need to provide code for this - we will ask you about how you might deploy this workflow in the cloud. If you have never done cloud deployment before, it would be great to read up on this.

Think about how you might handle new, messy data coming in every week/month, and how to handle performance issues.