Imagine that you have already started your job as a Data Scientist at Semrush.
The situation in the example below is exaggerated in some places and simplified in others, but generally, it looks like a real problem that may come to a Data Scientist.
We would like to see your thinking process, how deep you dig into data, and how you present the outcomes. Good luck!

Data Science task

Intro:

You have been provided with the data: `synthetical_payments.csv`
It's an extract with payment data over the last few years.

userId	billingCountry	transactionTime	product	price	amount	period
41851	United States	1420334701	PRO	99.99	1007.9	12
13575	India	1420383152	PRO	99.99	99.99	1
73971	India	1420461487	PRO	99.99	99.99	1
80119	United States	1421010897	GURU	199.99	2015.9	12
90456	United States	1421020614	GURU	199.99	2015.9	12

The context - users are paying for a subscription to the service at different levels (`product`: PRO→GURU→BUSINESS). The company is interested in users to pay as long as possible. Users can change the subscription level over time. Users can pay every month or can pay for the year at once and get a small discount.
Other fields in the data:

- `price` is the price of the product per month
- `amount` is the actual amount paid during the subscription period
- `period` is the duration of the subscription that a user paid for (# of months)
- `transactionTime` is the transaction timestamp in the Unix time format (seconds)

The task:

1. Check the data and perform EDA.
2. Forecast MRR (Monthly Recurring Revenue) for the next 3 months (starting from the last month in the data) in total by PRO/GURU/BUSINESS (3 time series), preferably with confidence intervals.

The result of completing a task is:

1. A completed repository with the project code meeting the following criteria:
    a. Reproducible.
    b. Containing comments on the steps taken.
    c. Designed according to the best practices you consider the most important.
2. A CSV file with a forecast.
3. A short report for the stakeholders from the business in any format on the quality/usability of the forecast. Additional comments are also welcome: a) any problems with the available data, b) what you would do if you had more resources/time/data (which ones?).

Deadlines and format for submitting the test task

The format is free - you can send as a link to the GitHub, shared drive, zip archive or as links to separate artifacts.

The deadline - usually we give 7 days (168 hours) for candidates to submit the results, but you can discuss any changes to this with us if necessary.

Don’t try to make it ideal, we all know it may take forever to dig the data. Try to do the best you can in one evening, following the "value for money" approach.