

Senior Data Engineer Candidate Take Home Test

Overview

The focus of this take home is around ingesting a mock Trustpilot Review dataset into a database and building a reporting API.

Problem Scenario

Trustpilot's legal team is one of the Data Governance & Compliance teams stakeholders. Sometimes they make ad-hoc data requests. While these requests can vary in exactly what information they need, we've spotted some trends in what they are asking for.

- Provide reviews for business X
- Provide reviews by user Y
- Provide user account information for user Z

To make dealing with these data requests easier, we want to build a simple PoC of an API that when called will return the data as a CSV file.

Requirements

The assignment is designed to be open to your interpretation and is language and technology agnostic.

We ask that you:

1. Ingest the Reviews CSV file provided into a database of your choice
2. If needed, model and programmatically clean up the data
3. Create an API that can be used to answer ad-hoc data requests
4. Submit your solution via **Github**

What Are We Looking For?

We want to give you the opportunity to show off your software engineering skills and coding best practices.

Examples of areas we care about:

- Thought Process
- Design Choices
- Performance Considerations
- Unit testing
- Readability (Code comments, etc)
- Productionisation
- Documentation/knowledge sharing
- Cost

Please pick the language that you feel most comfortable with. At Trustpilot we use a selection of languages (Python, Typescript, Java) and we are open to you choosing other languages as well.

We welcome the use of AI in helping create your solution to this challenge, please be ready to explain how you used it and which areas were most impacted by its use.

Be prepared to discuss your solution, how you could extend it and how you might run the application in a production environment.

How Long Should It Take?

This should take somewhere between 3 and 6 hours.