



Otrium

Analytics Engineer
Case Study

Hi there! 🖐️



Congratulations on moving forward to this stage of the interview process for our **Analytics Engineer (Recommendations)** role!



Here are your instructions;

- You will have **one week** from receipt of this exercise to complete the assignment.
- You should take approx. 2 - 3 hours to complete this exercise.
- The following tasks and set of questions are designed to help us understand how you might tackle an Analytics Engineering challenge not too dissimilar to the ones we face at Otrium.
- To help with the test we have provided a series of CSV files with some example data (next slide), this is not of the same scale as the “real world” but is enough that a reasonable example can be created. Where specific detail hasn't been provided please ask or feel free to use your initiative/make something up to get to a result.
- There is no right final answer, we are more interested in how you approach the problem, the techniques you use, and your ability to have a conversation with us about how you have completed the task.

What to do once it's complete?

- Please send your completed Case Study to Libby and Nikki (nikki.bueno@otrium.com) via email
- We will then assess your Case Study and reach out to you to organise a Feedback Session

We're very excited to see what you can do and thank you in advance for your efforts!

The Data:



- [Export_customers.csv](#) - a list of customer data
- [Export_products.csv](#) - a list of product data
- [Export_orders.csv](#) - a list of sales data
- [Export_views.csv](#) - a subset of product views from the website

The Task:



Otrium is looking to create a new recommender system. We have identified AWS Personalize as an off-the-shelf product that can allow us to implement something quickly using our existing data. The public documentation can be found [here](#) specifically the information on datasets and fields required for an ecommerce recommender are [here](#).

We would like you to create an example of how data could be prepared to ingesting into the Personalize system (note: you don't actually need to do the Personalize step, this is just about preparing the data)

Using the data provided and the information above, please:

1. Create a small DBT project locally to process and transform the data provided into final tables ready to be loaded into AWS Personalize
2. Consider how you might validate the quality of the data and any cleaning that might be required
3. Please make sure all dbt models have at least one test
4. Provide a copy of final dbt repo when completed either as a zip file or via a repo link

We are also interested in your thoughts on how we would take this example to production, for example:

- What additional data might be useful to help improve the recommender?
- What might the impact of bad data be on the final outcome?
- What considerations might there need to be for the larger volumes of data in a production environment?
- How often would this data need to be processed?
- How would you work with analysts + ML engineers to improve the process / outcomes?



Otrium

Best of luck!