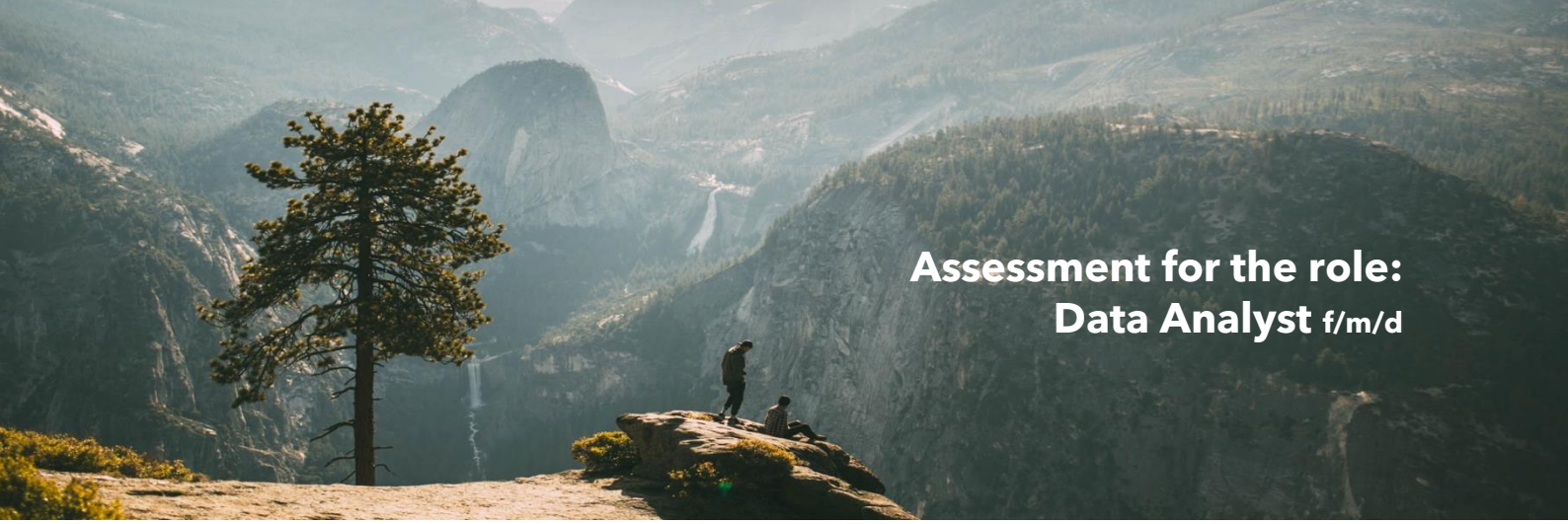

Assessment for the role: Data Analyst f/m/d

Briefing:

The ecolytiq sustainability platform uses financial transaction data and a multi-level enrichment procedure to provide CO2 impact data of bank users for data analysis.

Given is a huge set of transaction records from multiple bank clients from multiple banks for around one year. Instead of customer ids and personal data, we know only account-ids, the category for which the money was spent, the amount and the currency together with the calculated CO2 footprint for this transaction. No additional information about the person is available at this point in time.

In a separate data set we collect the answers from bank users to particular questions. Such questions tell us more about the habits of users. We are flexible in defining such questions and answer options.

A third data set will contain information about the usage of content items. We plan to persist when a given hint or insights item has been shown to a bank user, together with the rating or evaluation of the usefulness of this information.

Our goal is to extract valuable properties from the payment history. This starts with details of a particular account. The extracted results should tell us more about the behavior of the person which uses the particular banking card for digital payments and ideally, also more about the behavior changes we may be able to observe using the payment history.

Example 1:
Transactions in the category flights and trains are aggregated over 3 months. A comparison over all possible 3-month periods over all cards (accounts identified by account-id) can be used to find the distribution of the flight-train ratio values, eventually, there are clusters visible. Eventually, spending on flights and trains is correlated.

Example 2:
How does the average CO2 value per EUR spent evolve over time on a monthly or quarterly resolution? How stable are those values and can we find clusters?

Example 3:
The amount spent for gas or diesel drops significantly and in the same period, the amount or the nr of bought train/bus tickets increases. Does this change pattern indicate that a person switched from car to public transport? What should we ask the customer to clarify this situation? How can we correlate the deployed hints and insights with this potential behavior change?

Task 1
Please describe your approach towards clustering of bank users in Example 1.

Task 2
Please provide the analysis query (SQL/Python Pandas) for Example 2.

Task 3
Please provide the analysis query (SQL/Python Pandas) for Example 3.

Procedure
Please create a high-level project plan for your approach to solving the tasks.

The plan should show:
- the expected deliverables
- a timeline for tasks
- an estimation of the effort for work packages
- all materials you need for solving the task (we share all digital assets in a Github repository)

We will be happy to support you in answering questions to clarify the task and to give you early feedback on your thoughts, presented in the plan-review session. This plan review session is optional for you. You can request a review via email to: mirko.kaempf@ecolytiq.com or jobs@ecolytiq.com

(1) Please describe the analysis approach for Task 1 – it is important that you make clear your line of thoughts, but code / pseudo code is welcome.

(2) Please provide the analysis query for Task 2 and Task 3 either as SQL query or as Python scripts, using Pandas DataFrames.

(3) Please commit and push your results incl. your presentation slide-deck as PDF to your assigned Github-repository.

After an internal review of your results is done, you might meet some of our team members for a final presentation.

We will be happy to be inspired! ;-)

Please send the result to jobs@ecolytiq.com – you can directly send it as respond to our last email.

If you have any questions, please don't hesitate to reach out.

Background and Hints:

This task is for our candidates for the Data Analysis role in all levels.

We do not expect to get a production ready solution within just a view hours, but rather we intend to see how you handle such a complex not exactly specified task. Please do not invest more than one hour in the planning and preparation phase.

From a junior data analyst we would expect at least a project sketch and some pseudo code to illustrate how your approach looks like. All tasks should be addressed and only Task 2 must be solved with code.

An experiences data analyst should be able to solve all the tasks within a week and 30 to 45 min consultation with our team.

The presentation slides are not expected in a conference-ready quality, but rather on a day-2-day conversation level. Please, think about a presentation in a customer meeting, where you have to show your progress, not the final result of a research project.