We are absolutely thrilled you are interested to join us at Rohlik Group. By now we have spoken to you about your experience and personality in great detail. Next step is to find out how you tackle a task in reality.

**Data Engineer for Finance, People & Culture**

**Main Task**Visit <https://www.kaggle.com/olistbr/brazilian-ecommerce/home?select=olist_orders_dataset.csv> to review the schemas of the transaction data tables provided by Olist.

1. Cohorts summary
   1. Write pseudo-SQL code which you would use to calculate the number of customers per cohort.
   2. Adjust the code from the previous table to split the cohorts according to customers’ state.
   3. Adjust the code from step a) to contain the number of delivered orders per cohort per month.
2. Payment methods summary
   1. Write pseudo-SQL code which you would use to prepare a table which would contain the frequency of occurrence of a product category in all orders grouped by customers’ state.
   2. Adjust the code from the previous step by adding the payment type to it, however, only list the 3 most common payment types, and group the rest under other. Pivot the table to contain the payment type categories in columns.
3. Process structuring
   1. Imagine you have a table that contains transactional data in rows. The table has the following structure:
      1. Month
      2. Provider
      3. Code 1
      4. Code 2
      5. Numeric value
   2. The company would like to map the code values to specific aggregate measures (let’s assign them values A, B, C, D, E), however, this mapping can change over time. The past versions of the mapping, though, need to stay in place. How would you set up the mapping process in order to allow business users to make changes to the mapping if necessary in a way that would not require your help whenever the mapping changes?

**Bonus Task**

1. Pull the full list of Rick and Morty characters and episodes using a custom built extractor (built in Python or R) from the API endpoints available on <https://rickandmortyapi.com/>.
2. Transform the JSON responses into separate flat tables.
3. Prepare a table that will contain the following columns:
   1. episode
   2. episode\_name
   3. character\_name
4. Prepare a summary table in which you sort the trios that appear together most often based on their frequency in all episodes in a descending order, however, the trios must not contain Rick and Morty together. Specify the number of occurrences per trio per season in columns.

**Format**

We are not strictly set on the form of the presentation. Whatever works for you and you are comfortable with, works for us. So if you just want to have an open forum discussion without any materials, we are in. If you want to present any documents to us, we will provide the tools for it or bring your own. Just let us know.