Data Engineering mini project for candidates

Working as a Data Engineer in the Pricing Data Science team in Maersk means a lot of exposure to data and lots of interaction, not only within the team itself, but also with other tech teams around Maersk and with business stakeholders. Given the wide variety of data sources and challenges, it is very important that you can go beyond the data engineering "mechanics". This includes being able to think creatively, anticipate challenges and problems, and clearly express the conclusions and insights from your work.

This activity sets out to evaluate you on three main aspects of Data Engineering
- How insightfully you handle a new dataset in a data domain that you may have not encountered before
- How well you structure and organize your work
- How well you communicate your work

This mini project is intended to simulate a data engineering scenario, where we as a team have a business challenge in scope, and you are given a relevant dataset to work on. We expect you to spend some hours (approximately 5 to 10) working on the task, and afterwards present your findings to a few people in the team. The presentation should be maximum 30 minutes, to leave time for about 20-30 minutes discussion afterwards.

Business Challenge

The fictional business problem is related to understanding different aspects of marine activity in Finland. Questions that may be relevant are for example
- Visualizing the number of Finnish registered ships per city, with a dot placed for each city's coordinate.
- Answer if the number of port calls made by Finnish ships correlate with population in the region?
- Answer if the number of ship registrations in Finnish cities correlate with the number of port calls?
- Understand the average dimensions of Finnish ships?
- Understand the average age of Finnish ships?
- Etc

Your role in the project is as data engineer, and you therefore need to provide the data foundation to be able to answer these questions – not necessarily to answer the questions yourself. The specific task is divided in two parts:

Understand the data

First of all, look at the datasets and make yourself familiar with them - just as you would do in a real-world situation. The datasets are not necessarily obvious and may have characteristics that generate many questions. In many cases, it's perfectly fine to make assumptions on anything – but be prepared to motivate or reason around them.

Prepare a presentation in any suitable format that clearly conveys your findings with the data and any open questions.

Outline a solution

Outline a solution for how to read, process and publish the data for usage, including how to handle historical data. Highlight preferred platforms, data processing principles and data modelling choices.

Include this topic in your presentation in a suitable format.

Finally, **be realistic in terms of effort**: the focus for this work does not necessarily need to be a fully developed solution, but to show your understanding and reasoning. You are free to make any simplifications and assumptions you want, but be transparent and be prepared to motivate these in a discussion.

Also be prepared to answer questions on your work during your presentation. This may be related to anything from specific methods used in code to reasoning about the data generating process. Again, we are looking for "insightful reasoning" rather than "correct" answers.

You are free to use any language or software you like, but we expect you to share your code with us prior to your presentation.

Datasets

Datasets are partly in Finnish – it's perfectly fine to use Google Translate to translate the contents… (See the Finnish language as an analogy of a real world setting with poor documentation and unclear abbreviations ☺)

Cities in Finland with population and coordinates
https://simplemaps.com/data/fi-cities

API that shows current port calls in Finland
https://meri.digitraffic.fi/api/v1/port-calls

All Finnish registered ships
https://eservices.traficom.fi/LicensesServices/Forms/Alusrekisteri.aspx?download=zip
(Metadata available here:
https://www.traficom.fi/sites/default/files/media/file/alusdata%20muuttujaluetto1.1.xlsx)

Mapping UN/LOCODE to Cities
https://service.unece.org/trade/locode/fi.htm