Case: Karrieremulighet Inmeta CoE Nordics: Data Scientist - Mustafa Jarallah
Tidspunkt: **Tirsdag, 15. august kl. 08:00 – 09:30**

Task 1:
You are given a dataset. (See other attachment). Your task is to develop a model that performs as best as you can.

We want you to go through the following steps:
>   1. EDA – Exploratory Data Analysis.
>   2. Develop a regression model. Use findings from the EDA to guide your choice of model, pre-processing etc.
>   3. Evaluate training metrics. You are free to choose metrics, but the final score should be given as MAE (Mean Absolute Error)

Iterate over this cycle a couple of times.

You must use Jupyter Notebooks to present the results from task 1.

Task 2:
Your task is to model you developed in Task 1 and operationalize it in Azure ML.
Requirements:
1. The model should be deployed with an HTTP REST endpoint. You should be able to demonstrate an interaction with the endpoint using Python / Jupyter. (Demonstration can be done from within the workspace environment. No public IP is required)
2. You should create a training script following best practices wrt. instrumentation / logging using MLFlow.
3. You should create a test script (scoring script) to test the trained model. (Due to the size of the dataset, you are allowed to use the same data as both test- and validation. You should be able to explain why this is a bad idea)
4. You should create AML pipeline (scripts or components) to perform model training.
5. You should create AML pipeline (scripts or components) to test the model.
6. Demonstrate or Discuss how you would automate (schedule) model training, testing and deployment of a newly trained model, including when to push the model to the Azure Model Registry.
7. Discuss other considerations you find important if any.

The work should be carried out in Azure / Azure ML.