DATTABOT

**DATA ENGINEER – TECHNICAL TEST**

1. Imagine you are given a Python script from a project that you must handle. You notice that the scripts are not optimized yet, and it takes time to run in the production environment. So, you initiate the process of optimizing them. Please refer to the scripts.py
   a. Since you do not have access to the data, please create dummy data that would be needed to optimize this script.
   b. Since you do not have a documentation for the script, please create proper documentation (input, output, objective, limitation, parameter of script, etc.)
   c. Optimize these scripts to improve running time.
   d. Record the number of rows in your dummy data and the running time before and after the improvements.
   e. If you have any additional optimization options beyond point C, what else can you do to optimize these scripts?

2. Imagine that in your upcoming project, you are asked to preprocess employee data from a certain company. Please refer to the dataset.xlsx.
   a. From the given dataset, clean the table according to these criteria:
      - Employee ID should be unique. Please make sure this criterion is fulfilled.
      - Create a standard format for the following columns: Name, Gender, Education Level, Job Category, Age, Score, Phone Number, Domicile, and First Day of Employment.
      - Convert the employee's age into their birth date.
      - Feel free to use your creativity to improve the table and make it cleaner.
   b. Create a data flow or process flow diagram of your preprocessing script as a documentation for internal team (you can use draw io or another tools)
   c. [Optional] the head of HR wants to know their employee situation. Create a visualization (using BI tools is better) that can provide insight for the head of HR using the cleaned dataset.

Please use Python for this task. Send us the script, and feel free to use Jupyter Notebook or any IDE you are comfortable with.