

# Project: Modeling Production Capability

---

## Overview

This exercise involves analyzing production data from various factories across multiple cities. These factories produce a chemical compound measured in hectoliters, with all values rounded to the nearest integer.

The goal is to develop a model that can probabilistically predict future production amounts. Initially, we aim to model daily production. Subsequently, simulations will be employed to determine the distribution of production over multiple days.

For the current scope, it is assumed that production on specific days is independent and identically distributed (i.i.d.). Consequently, to simulate the total production over a 5-day period, one can simulate 5 values  $p_1, p_2, p_3, p_4, p_5$ , and the sum  $p = \sum_{i=1}^5 p_i$  represents a realization of simulating production over a 5-day period.

## Data

In the **data** folder, two types of files are present. The first, *master\_data.json*, contains information on the Maximal Sustainable Rate (MSR) at the production plant. The second type is located in *daily\_production/city/date.json*. These files provide details on production for each date, including the following values:

- *DoW*: Day of the Week.
- *hour*: Hour of registration.
- *minute*: Minute of registration.
- *date*: Date of registration.
- *maintenance*: Yes/No, indicating scheduled maintenance on this date.
- *prod\_loss*: Unexpected production loss on this date.
- *prod\_loss\_perc*: Percentage of MSR lost.
- *production*: Total production on this date.

## Detailed assignment

### Assignment: Setting Up Project and Developing a Production Simulation Program

In this assignment, you will follow a series of steps to set up a project structure, manage version control using Git, create a virtual environment with Anaconda, and develop a Python program for simulating a production process.

#### Step 1: Download and Organize Data

Begin by downloading all relevant data and organizing it into a designated folder within your project structure.

#### Step 2: Initialize Git Repository

Turn the project folder into a Git repository on your personal GitHub account to track changes and collaborate effectively. You can add the data into this git repo.

### Step 3: Create Virtual Environment with Anaconda

Utilize Anaconda to establish a virtual environment for your project. Ensure that you install only the necessary packages required for the project.

### Step 4: Implement Python Code for Data Reading

Write Python code to read and process all data acquired in **Step 1**, ensuring efficient data handling within your project.

### Step 5: Identify Suitable Distribution for Production Process on $n$ days

Explore and determine an appropriate probability distribution that best fits the characteristics of the production process on  $n$  days. You should make a distinction between the case of only a limited number of days and a lot of days, try to base this choice on some well chosen visualizations. You may assume that production on each is i.i.d.~with an unknown distribution.

### Step 6: Develop Simulation Program

Create a Python program capable of simulating the production process over a user-defined period,  $n$  days. The program should offer flexibility for users to choose the duration of the simulation and the number of times you want to simulate this number of days. That is, a function call `simulate(n_days = 7, num_runs = 10 ** 3)` should return  $10^{**3}$  values which all are distributed as the production on 7 days.

**Step 7: Use your simulation** Use your simulation program to simulate the production probabilities over a time period of 7 days. Also generate a plot where you show the shape of  $F(x) = P\{X \leq x\}$ , with  $X$  the random variable that equals the production over 7 (i.i.d.) days. Your code should look something like:

```
import matplotlib.pyplot as plt
import numpy as np

data = simulation(7, 10 ** 3)
F_empirical_data = cumulative_distribution_function(data)
plt.plot(data, Fdata)
```

### Step 8: Share your solution

Compose a *README.md* file to document and guide users through your project. Include clear instructions on setting up the project, and executing the simulation. Provide concise explanations of the project structure and include an environment *YAML* file that I can use to setup a virtual environment.

By completing these steps, you will establish a well-organized project structure, manage your code using version control, create a virtual environment for efficient development, and develop a simulation program to model the production process effectively.