The problem:

Mechanical induced disruptions (e.g., delays and cancellations) have a big impact for commercial airline operators as they affect their revenue streams and erode customer trust and retention. In an average year, at least 30% of all disruptions are caused by mechanical reasons, with environmental/inclement weather accounting for the rest.

Preventive maintenance of complex assets like aircrafts and their major assemblies (e.g., engines, APUs, landing gear, etc.) is generally based on fixed intervals. However, this is not an optimal approach as a high number of part removals end up being unscheduled and unscheduled maintenance means less flight time for the aircraft.

This emphasizes the need for better predictions on:

- Component failure/remaining operating life

Your assignment (Part 1):

- Given the attached data set, your goal is to build a ML model to predict unscheduled removals (i.e., when a part will fail) at the component level (i.e., serial number).
- Hint: You could apply survival analysis techniques to predict probability of survival (or failure) for each part at any given cycle count.

Your solution should be able to demonstrate:

- Knowledge of the different stages of a data science project and implementing a ML pipeline.
- Familiarity with data exploration and manipulation techniques/libraries in Python

Notes:

- There are not expected or "correct" answers for this part and predictive performance is NOT the main objective of this part. The purpose of this section is to present/discuss your e2e ML pipeline based on the chosen technique.
- This is a coding assignment; your solution should be presented in a Jupyter notebook for demonstration and discussion.

Your assignment (Part 2):

- Based on the solution provided in **Part 1,** your goal is to implement the necessary steps to bring the ML pipeline into production.
- Hint: provide API endpoints to perform inference (prediction) calls. Optionally, you may include a training pipeline.

Your solution should be able to demonstrate:

- Ability to transform a notebook-based/experimental solution into a repetable/reproducible and deployable ML pipeline.
- Knowledge on applying operation-focused MLOps solutions, model deployment and serving, testing, logging, APIs, etc.

Notes:

- This is a coding assignment to simulate the implementation of an inference pipeline in a production setting. You may use any tools/techniques of your choice (flask, mlflow, kedro, Argo, docker, etc.).