1. Dataset description.

These are three datasets that we will be using during the project (There will be more).

The structure of them is as follows (columns):

- ts (the time series corresponding to each sample)
- status (the status of the machine that determines its condition)
- faults (the list of faults in each machine)
- machine_name
- report_date
- measurement_date

The report date and measurement date were created during the preparation of the dataset, the two are equal and you can remove one of them and work with only one of them.

The datasets are in Pickle format, you will need the pandas library to read and load it.

The velocity datasets contain high range and low range readings. The difference is that high range readings cover a wider frequency band therefore some faults that occur at high frequencies appear better in the high range readings. Other faults occur at low frequencies and appear better in the low range readings.

Each time series in the ts column contains 12 channels, their structure is as follows:

- Each machine consists of a motor, and this motor is connected to either a pump or a fan.
- There is a total of 4 bearings in each machine that are measured.
- Each bearing generates signals in the three directions of space (horizontal, vertical, and axial).
- This means that the each '3' channels in the time series is a specific bearing.
- The first three are motor bearing non drive end. (The bearing furthest from the load)
- Second three are motor bearing drive end. (The bearing closest to the load and the coupling)
- Third three are pump bearing drive end.
- Fourth three are pump bearing non drive end.

2. Things you may want to do.

- Learn how to load the dataset into a notebook.
- Learn how to manipulate it and perform operations on it.
- Do exploratory analysis on the data (for example: How many times did fault 'X' occur?)
- For each machine, what was the most common fault in it?
  (Nearly all the machines contain several types of faults, which types of faults occurred together the most?)

- What is the count of the status?

- For each machine, you can try to do a bar chart to visualize the distribution of their status along with the time of the year.

- How many points are in each ts? Are all of them of the same length?
- Are there any NaN values in them?

- Make a function that calculates the FFT for each time series. What operations do you need to do on the time series before you apply FFT on it? (search for Hanning window)
- Try to make a function that gets as input 'the type of fault' you want to see, then the function accesses the dataset and extracts the time series corresponding to that fault, calculates the FFT and plots the time series along with the FFT.
- What meaningful features can you extract from the time series? for example: mean, standard deviation, kurtosis, skewness ...etc.
- What meaningful features can you extract from the FFT? for example: Dominant frequencies, rms ... etc.
- Can you make another data frame with new columns that contains those features you extracted?