



# Candidate Data Exercise & Presentation

---

**Objective:**

To identify signals that are important in determining pass/fail yield in a manufacturing process.

**Description:**

In this exercise we would like you to perform an analysis combining data from multiple sources and then create a presentation (keynote, powerpoint, etc) to communicate your approach and results. This presentation should include:

- The tool(s) you used to perform the analysis (scripting language, statistical packages, etc.)
- How you ingested data from different sources
- How you preprocessed and combined data
- Description of your analysis including any assumptions you made
- Visualizations and conclusions
- Brief description on how you would automate this solution in a production environment

**Data Sources**

The majority of the data can be found at <http://archive.ics.uci.edu/ml/datasets/SECOM>. However we'd also like you to supplement this data with the attached JSON file. You can assume that the index in the JSON file matches the ordered of the data found in the SECOM dataset.

A detailed description of the dataset and the problem can be found at the above website but in short, we are looking to see how you identify important features for predicting failure. There are 591 signals found online and an additional 5 signals found in the JSON file.

**Questions:**

For questions, please contact your recruiter and they will forward it to the appropriate party.