Kevin Benson – Director, Asset Risk (Kevin.Benson@pacificorp.com)
Jordan Pino – Data Science Manager (Jordan.Pino@pacificorp.com)

PACIFIC POWER

Data Science Assessment

Summary: This assessment is aimed at allowing us to understand your thought process when it comes to a broad Data Science question. There is no right or wrong answer or way to build a model, analyze the data, or interpret findings. Just be sure to think creatively about the tasks and use your experience in Data Science to guide you.

Datasets: *train.csv* and *test.csv* –> please use train.csv to train the model and test.csv to test the model.

Preferred Tools: Python (*i.e.* any IDE, Jupyter Notebook, or similar platform) and PowerPoint.

Tasks: Please see below for a list of tasks we'd like you to complete **independently** (please avoid discussing with any of your peers; you **MAY** use any online sources).

1. Please perform some **exploratory** analysis on the data (*i.e.* create some descriptive visualizations of the data).

2. Next, please note any oddities in the training dataset that might impact a predictive model and then attempt to clean some of those issues.

3. Once the data cleaning is completed, please create a **predictive model** that outputs energy consumption using the test dataset. Please select any algorithm you feel would work best or you can try a handful and determine which is the best given the data. Also, be sure you can explain why you selected a certain model and what led you to it.

4. Lastly, compile a short PowerPoint presentation with your results/figures (~5-6 slides, but feel free to add more if needed).

Delivery Method: Please email your PowerPoint presentation and either a .py of your code or a Jupyter Notebook to Jordan Pino (Jordan.Pino@pacificorp.com) and copy Kevin Benson (Kevin.Benson@pacificorp.com) when complete.