**Problem Statement**

We are looking to predict spot prices over time using defined weather variables. Since we have well defined independent variables and a dependent variable, this would be a supervised machine learning exercise. The dependent variable is natural gas spot prices which is recorded at the day level. The objective of this problem is to predict these prices over a year, so this would qualify as a classic linear regression problem.

**Features**

The independent variables are daily average temperature (in Celsius), daily average pressure (in inches of mercury) and daily average relative humidity (percentage). This data was obtained from Weather Underground using their APIs. The raw data had readings for every 20 minutes, it was averaged at the day level prior to modelling to match the dependent variable, spot price which is set at the day level.

**Technique**

This looks to be a classic regression problem, so our plan of action will be to build different linear regression models and evaluate them. If the results are not in line with expectations, we will also look at regression trees

**Evaluation**

As with any linear regression exercise, we will use R-squared, adjusted R-squared, RMSE, SSE, F-statistic and p-value to measure the strength of the model.