Objective: Build Simple Linear regression model to predict the Net Hourly Electrical Energy Output of the plant.

Dataset Details

Combined Cycle Power Plant Data Set

The dataset given to you has been collected from a **Combined Cycle Power Plant** over 6 years (2006-2011).

A combined cycle power plant (CCPP) is composed of gas turbines (GT), steam turbines (ST) and heat recovery steam generators. In a CCPP, the electricity is generated by gas and steam turbines, which are combined in one cycle, and is transferred from one turbine to another. While the Vacuum is colected from and has effect on the Steam Turbine, the other three of the ambient variables effect the GT performance.

The variables in the dataset are as follows:

* __AmbTemp__ : Hourly average of ambient Temperature at the plant (Range 1.81°C and 37.11°C)

* __Exvac__ : Exhaust Vacuum is pumps air out of the steam turbine (Range 25.36-81.56 cm Hg)

* __AmbPress__ : Hourly average of ambient Pressure at the plant (Range 992.89-1033.30 milibar)

* __RelHum__ : Hourly average of ambient relative humidity at the plant (Range 25.56% to 100.16%)

* __ElOut__ : Net hourly Electrical Energy Output of the plant (Range 420.26-495.76 MW)

Perform following Steps on the data:

1. Read the Data into R**.**
2. How many Observations are there in the dataset? What are the data types of the variables?
3. Get some descriptive statistics of the variables in the dataset
4. How does the ambient temperature effect the electical energy output generation at the CCPP?
5. Plot the relationships between the explanatory and response variables
6. Plot a Correlation Plot
7. How many Missing Values does the dataset have?
8. How do you treat the Missing values?
9. Standardize (Scale and Center) the explanatory variables
10. Split the data into train and test data sets (70/30)

INS♦F
Inspire...Educate...Transform.

11. Build Models
- Build four simple linear regression models
- Note down the coefficients, intercepts, R^2 and p-values of the models you built.
- Plot the residual plots

12. Evaluate the Model on performance metrics