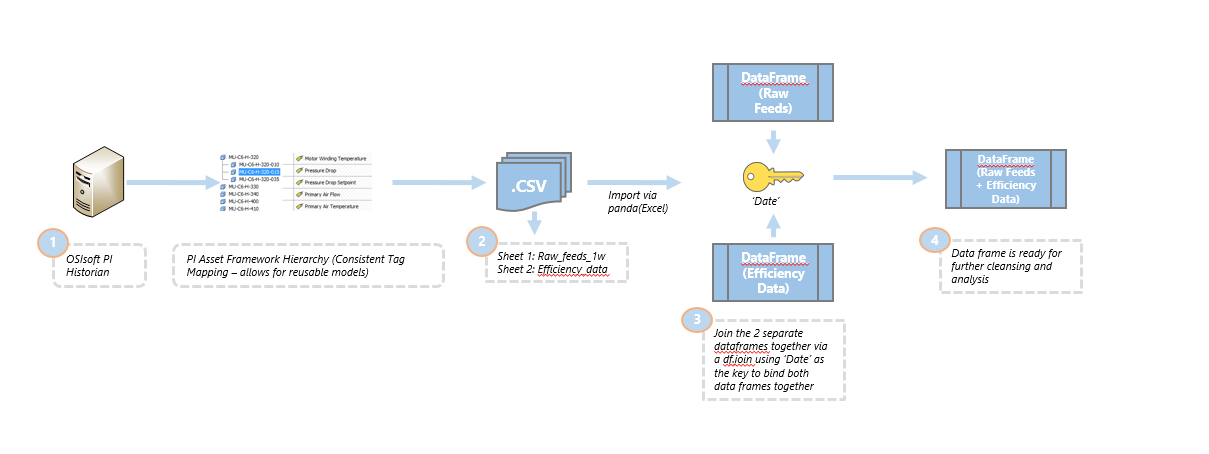
**Data Cleansing**

For the Boiler Feed Pump Project, the data is sourced in our OsiSoft PI System. The Plant Information (PI) System is a data historian which includes information across all our plant named in a specific manner (i.e. Unit-UnitNumber-System-Sub-System-TagIndex-Secondary-Index)

Having imported pandas as pd, we can use the following command to read in the data:  
*file = pd.excel(‘BFP\_data.csv’).* Once we have identified the ‘sheet\_names’ that include the data we wanted, we used the file.sheet\_names information and then passed this information as an argument to the .parse() method.

We have to specify ‘skiprows=1’ in the .parse() method due to the structure of the PI Data. Otherwise, the system-tag names will be included in our data frame which isn’t useful.



Upon further inspection of the dataframe, we had to provide new names to the df.columns field because of 2 reasons:  
1) We needed to skip the first row of data as this was in a non-user-friendly format (i.e. MPS-5-FW-FT-1045-XQ01).

2) We needed to rename the existing column headers to a standardised and understandable format for the stakeholders (engineers) reading this



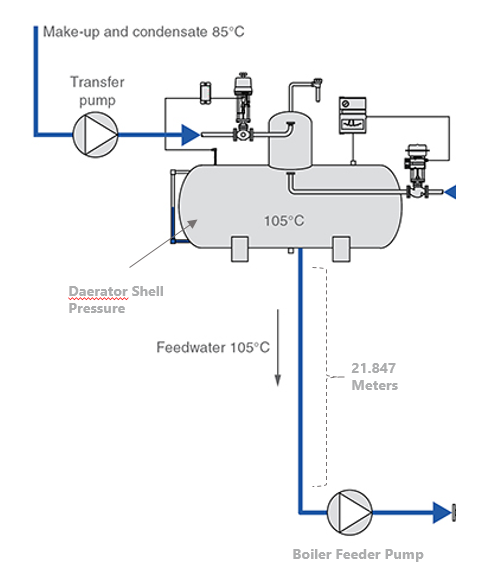
Before renaming of the df.columns can occur, we need to remove all the additional columns that were included in the data frame with null values. These were removed using the .dropNA() command. The key words how=’All’ ensures that all data frame columns (axis=1) will be dropped from the data set.

Once this had been completed, we needed to create new intermediate variables from the data that we now have.

We had to create the following 6 variables:

* Suction Pressure - Pressure variable required for the calculation of the Total Head.
* Total Head – This represents the Total Pressure in the pump system.
* RPM % - This is the ‘Actual Revolutions Per Minute’ divided by 6,412 which is the target RPM for the pump design.
* Net Input – This is the amount of energy that is being fed into the motor.
* Net Output – This is the amount of energy that is output by the motor.
* Efficiency % - A ratio of how efficient our motor is at converting energy.

Suction Pressure is needed if we are to calculate Total Head (Total Pressure in the pipe system). However, we did not have the ‘suction pressure’ tag in PI. We only had the discharge pressure. Without the suction pressure, the differential pressure cannot be calculated. Reviewing the engineering drawings, we were able to identify another signal that we could use as a substitute for suction pressure – this is the Daerator Pressure. This is represented in the below drawing.



Every Boiler Feed Pump is connected to the Daerator. The Daerator has a signal called Daerator Shell Pressure recorded in kPa (kiloPascals). Measuring the height from the Daerator to the Boiler Feed Pump gives us a height of 21.847 Meters. Using the standard international table of measurements, we can conver this height (Meters) into an appropriate kPa. 1 kPa = 0.10199773339984 Meters. Therefore, using this logic we have:

Using the Daerator Shell Pressure and combining this with our newly discovered height-to-kPa measure, we have a new variable now called Suction Pressure. This is done in the below line of code:



This approach has been repeated for the remaining 5 variables. Once this process was complete, the last check was to apply a filter on the data frame to remove periods where the pump is not on. We have applied a simple filter to do so.

This can be done with the below line which checks to see that:

**Flow Rate is greater than 0** – If there is no flow rate, there is no water supply, hence the pump cannot be on.

**Pump RPM > 0** – If there are no revolutions, then the pump is not in operation.

**Total Head > 0** – If the Total Pressure in the system is not greater than 0, then the pump cannot be on as pressure is required to transfer fluid from one location to another. With no ‘head’, this cannot occur.

