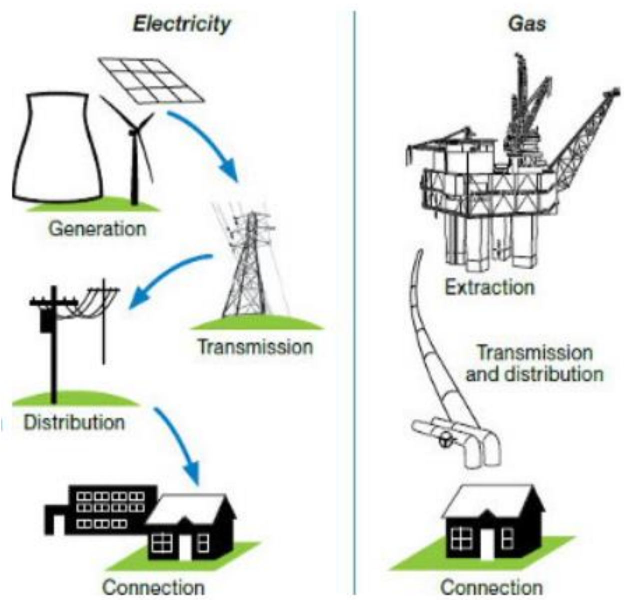
Capstone Project 1 – Boiler Feed Pump Monitoring Model

Christopher Hui

<u>Issue: Avoidance of generating assets being forced offline through the application of optimisation curves in the form of an envelope function</u>

"Electricity is the set of physical phenomena associated with the presence and motion of electric charge." It is the one of the foundational pillars underpinning society as we know it; from the electrons charging your mobile phone at night to the humble buzz of the industrial sector creating goods for the modern man to consume; energy underpins all of this. Synergy, our client, is Western Australia's largest energy supplier to the South West Interconnected System (SWIS). With a customer base spanning 255,064KM, Synergy is a major player in the Western Australian Energy Market with ~ 60% market share.



Being such a major player and in line with the Western Australia's government's policy of "Keeping the lights on", availability to supply energy, commonly known as the Availability Capacity Factor (ACF), is a key metric all energy suppliers pay attention to. The energy market operates in the below manner:

1) **Availability Payments -** Individual facilities are provided with 'availability payments'. These are credits that are provided to the company under the guarantee that it will provide energy when demanded and be available to dispatch. Think of it as an incentive for generators to stay available. The only circumstance in which availability payments are with-held or taken away is in the case where the generator experiences an unplanned outage which is commonly known as a **forced** outage.

2) **Forced Outages** - In the case a forced outage occurs, Reserve Capacity Payments are issued. These are financial penalties generators must pay for not being available. Depending on the time of day and the unit capacity, penalties can range from $0 AUD up to $1,000,000 + depending on the severity of the outage. Take for example, an outage in Summer where temperatures routinely soar to 40+ degrees, if a major generator goes out, then that means extra suppliers will have to supply energy into the market immediately which they may or be unable to do. As such,

Over the past 3 – 4 years, the upstream business has experienced several forced outages directly tied towards Boiler Feed Pump (BFP) failure with RCR costs averaging ~ $<Retracted due to confidentiality> AUD. To resolve this and reduce the number of forced outages tied to Boiler Feed Pumps (BFPs), I propose development of a BFP maintenance curve model (1). This model will project the Boiler Feed

Pump efficiency curve using actual plant data against the ‘expected’ efficiency of the BFP which is from its initial specifications.

The logic is as follows:

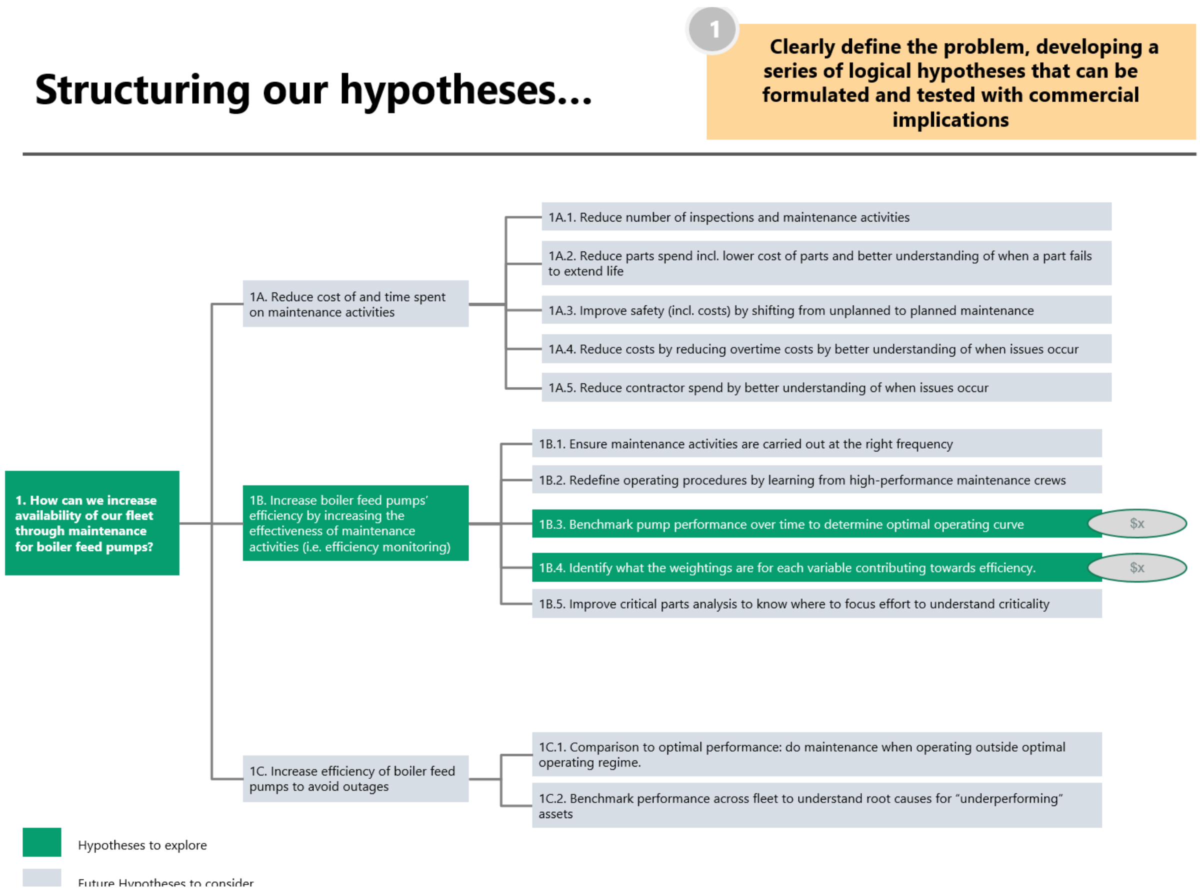
When an dip in the efficiency curve is noted, this should be observed and reviewed by the asset engineers to identify sub-optimal operation parameters which are linked to pump failure. Through continuous monitoring of the ‘optimisation’ curve, defective behaviour will be noted before an asset may fail, allowing proactive maintenance to be carried out.

Currently, we do not have active Boiler Feed Pump Curve models for monitoring our thermal fleet. Once the model is developed, this can be applied both in a forward propagation manner as well as a backward propagated manner to identify whether efficiency % deviation is a strong indicator of BFP failure and if so, adapt our maintenance strategies accordingly.

With this context in mind, we can propose the following hypotheses to explore.

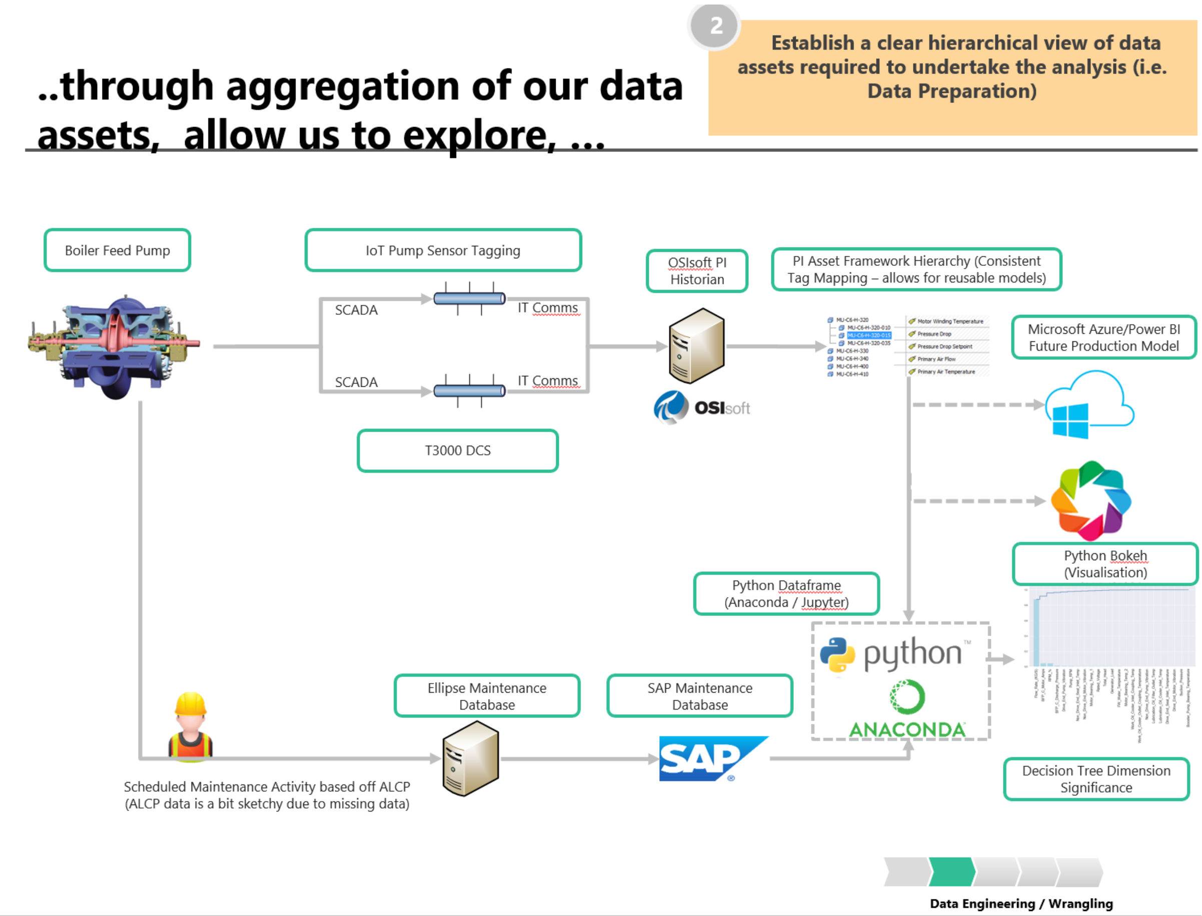
A) Can we increase BFP efficiency by increasing the effectiveness of maintenance activities through the benchmarking of pump performance over time to determine an optimal operating curve?

B) Can we increase BFP efficiency by increasing the effectiveness of maintenance activities through the identification of the key weightings of variables contributing towards efficiency?



Data Required:

With our hypotheses now defined, we can move onto the next step which is identification of what data sets we now need to extract. To create the BFP monitoring model, I will need access to the following data repositories that have been summarised in the below slide.



Data Required	***Source System***	***Hypotheses this data supports***	***Ease of Access (Low, Med, High)***	***Granularity of data required***
Boiler Feed Pump Target Specifications	Technical Documentation (Asset Management System)	Hypotheses A – To create an envelope function, we will first need to understand the vendor (OEM) specifications. From this, we can create a baseline curve model based off manufacturer specifications so deviations can be plotted.	**Med.** The documentation is located in archive files in the 1970's. This will require some exploration and further analysis to pull out the exact pump brand and pump specifications due to the pump manufacturer having been bought out by Alstrom which was then bought out by GE.	**N/A**

Boiler Feed Pump User Manual	Document Management – (Asset Management System)	Hypotheses A / B - This is essential if I am to understand which variables I need to extract data from.	**Low.** I liaise very frequently with the Asset Engineers so this information should be quick to come by and digest.	**N/A**
Boiler Feed Pump IoT Sensor Data / PI Tags	OsiSoft PI Data Historian (BFP Temperature, Discharge Pressure, Vibration, Pipe Flow, Motor Ampage, Shaft Speed, Generator Load etc.)	Hypotheses A/B - PI is our plant historian system which holds information on all of our assets provided the relevant PI tags have been made. Without this PI data, we cannot prove / disprove any of our hypotheses.	**Med.** I will need to liaise with the Pi Specialist to ensure I extract the correct data over the relevant time periods.	1 Hour grain across a 1 year period.
Boiler Feed Pump Maintenance Data	Ellipse / SAP	Hypotheses A/B - Ellipse / SAP is our Centralised Maintenance Management System (CMMS). This records when we have performed work overs on the pumps and the reason for this. We can join this information together with our BFP data to identify perhaps if there is any significant variance in efficiencies before a work over.	**Medium – High.** We have to create a SQL table to map the unique ID's in Ellipse and map this to the unique ID's in SAP. The equipment ID in Ellipse does not correspond to the Functional Location (FLOC) in SAP.	1 Hour grain across a 1 year period.

Cleansing and Wrangling of Data

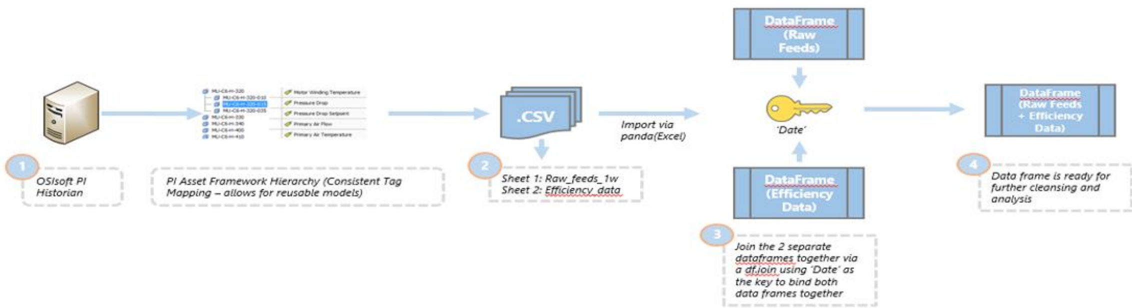


Figure 1 – Process that was undertaken to extract data required for BFP Analysis

The cleansing and wrangling of our data followed the process we have outlined above.

1) **System Data Extraction -** The first step required an extract from our OsiSoft PI System for the relevant engineering tags related to Boiler Feed Pump Activity. This included things such as Vibration, Motor Temperature, Lubrication, Linear Displacement etc. These values were all taken on a specific 1 hour grain which was a linear interpolation of the sampled data over a 60 minute period.

2) **CSVs** – Once the data had been extracted, we placed the information in two 'tables'. This is because certain information belongs in certain CSVs. While we did not stick to a predefined data model, we believe it is important to maintain separate tables for different types of data.

3) **Joining the data** – Using pd.read_csv and storing the csv into a dataframe is the first step. The second step was to join these two different data frames using the 'join' command, which

linked the two data frames together by a common index – in our case this is the date.

4) **Review the data** – Inspecting the data revealed that we had almost everything we needed for our Boiler Feed Pump Calculation with the exception of Total Head (Refer to the next slide)

Even with the data joined up, certain variables were still lacking. This requires us to create a new feature. In our case, one of the features required to measure pump degradation is called 'Total Hydraulic Head'. Total Hydraulic Head (THH) is calculated via the below formula:

$$\mathbf{THH} = \frac{\boldsymbol{P}}{\boldsymbol{\rho g}}$$

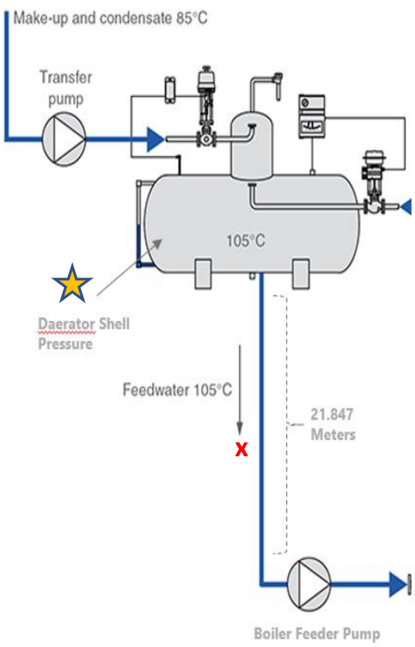
P is the differential pressure between the pumps discharge pressure versus the suction pressure;
ρ represents the density of the fluid (water);
g represents the constant, gravity (9.8 m/s^2).

Provided we have ***P*** this isn't a problem – however, apart from our traditional data cleansing (NA's, values > 0 etc.), we *do not have the **suction** pressure*. If we miss the suction pressure, we cannot calculate the rate of change in ***P*;** Hence calculation of the Total Hydraulic Head is not possible.

$$\mathbf{THH} = \frac{\boldsymbol{Dp} - \boldsymbol{Sp}}{\boldsymbol{\rho g}}$$

Dp is the discharge pressure of the fluid being discharged from the pump;
Sp is the suction pressure, or secondary pressure, from the fluid entering another section of the pump.

While we lack the ***Sp***, there is an alternative way this can be inferred. Reviewing the system specifications reveals another sub-system before the Boiler Feed Pump; the deaerator. For every Boiler Feed Pump it is connected to the deaerator system. The deaerator is a sub-system which strips oxygen from fluids (it de-aerates them). Looking into the Plant Information (PI) system, it is noted that there exists a sensor here with recordable deaerator pressure in kPa. With this in mind, we have the deaerator pressure + X which, in theory, should give us the ***Sp.*** Reviewing the system specifications via isometric drawings yields the distance from the deaerator through to the Boiler Feed Pump being **21.847 meters**.



Using the Standard Instrumentation metrics (1 kPa = 0.1019977m of head) , we can convert this height to an appropriate kPa.

Therefore, we have the following:

$$\boldsymbol{x\ kPa} = \frac{21.847m}{0.1019977}$$

$$\therefore x = 214.19\ kPa$$

Combining this with the deaerator pressure sensor enables us to recreate the $\boldsymbol{Sp}$ value which allows us to solve for THH. Our data set is now almost ready for exploratory data analysis with the exception of the below filters that need to be applied.

If the pump is to be defined as 'on' then:

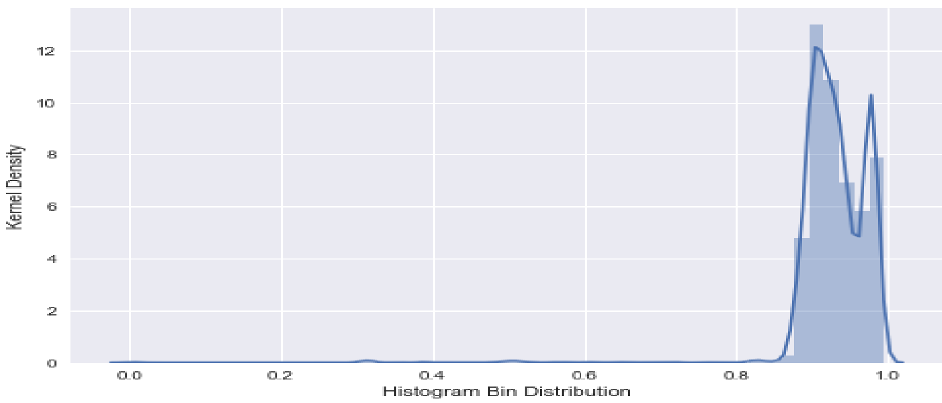
- Flow rate must be > 0
- Pump RPM must be > 0
- Total Head must be > 0

After applying these filters and returning the dataframe object, we are now ready for exploratory data analysis.

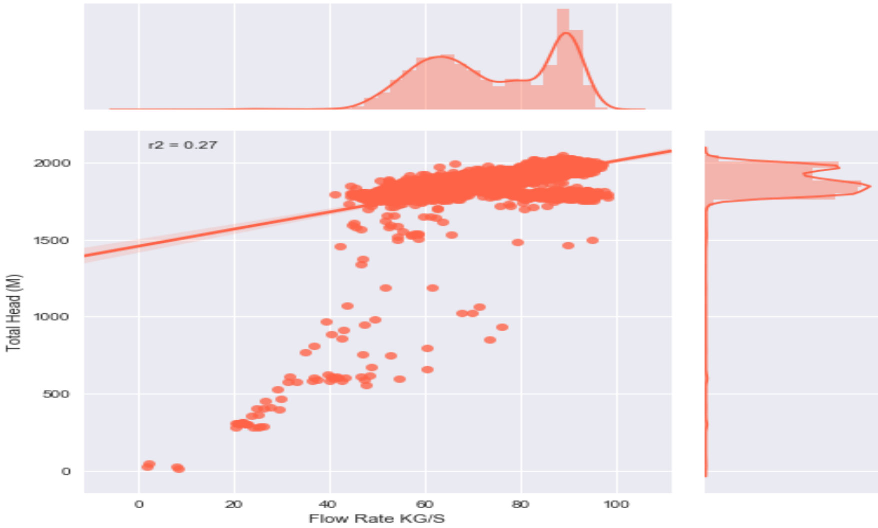
Exploratory Data Analysis

A) Can we increase BFP efficiency by increasing the effectiveness of maintenance activities through the benchmarking of pump performance to determine an optimal operating curve?

1. Based off Original Equipment Manufacturer (OEM) standards, it is recommended a pump curve be developed for each of the operating conditions of the pump at 25%, 50%, 75%, 90% and 100% as a function of speed - RPM (Revolutions Per Minute). **Before we take this as the gospel truth, it's worthwhile for us to test whether this is valid or not; do RPM populations exist at 25%,50%,75%,90% and 100%?** Plotting RPM via a Kernel Density Estimate reveals the majority of observations are actually between 80% and 100%, with the majority centred around 90%. Taking this into account, it is worthwhile to focus our analysis at the 90% RPM speed.

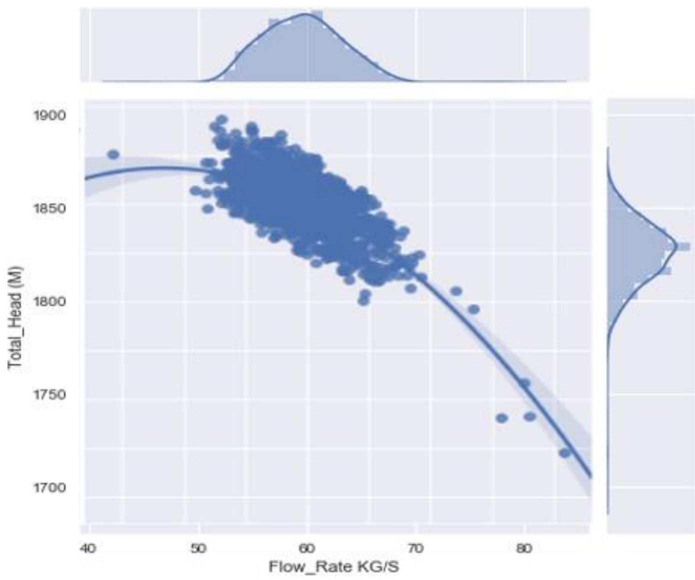


2. With the knowledge that 90% RPM speed is where our analysis should be focused on; we will investigate this further. Two variables which should have a strong correlation with each other is the THH versus the flow rate. For every type of pump there is a specific THH and flow rate which is optimal. In other words, we should see a relatively strong correlation. ***Does there exist a strong hypotheses between THH and flow rate?***



Looking at the above plot, it is evident a weak correlation is present between THH and flow rate. Additionally a bimodal distribution is present. Knowledge of the asset would dictate that we should *expect to see a strong correlation between both variables alongside a normal distribution.* ***Would this become apparent if we filter our data at a 90% RPM?***

3. Filtering our dataset to a 90% RPM reveals a completely different view. Rather than seeing a scattered concentration of points, we find a concentrated plot with which a curve could be fit through. Through the application of a 'curve' fitting algorithm, we identify that a 2nd order polynomial equation yields a line of best fit which cuts through our data smoothly, also yielding a normal (gaussian) distribution. This represents the THH vs. Flow Rate Curve which identifies the 'expected' THH based off any flow rate.



4. Similar to above, the same principle is applied for our feature variable efficiency which is plotted on a secondary Y-axis against flow rate. Plotting this against our Total Head yields a benchmarked pump performance curve which corresponds to a set efficiency and flow rate. **This can then be used as a baseline performance curve to monitor pump performance.**

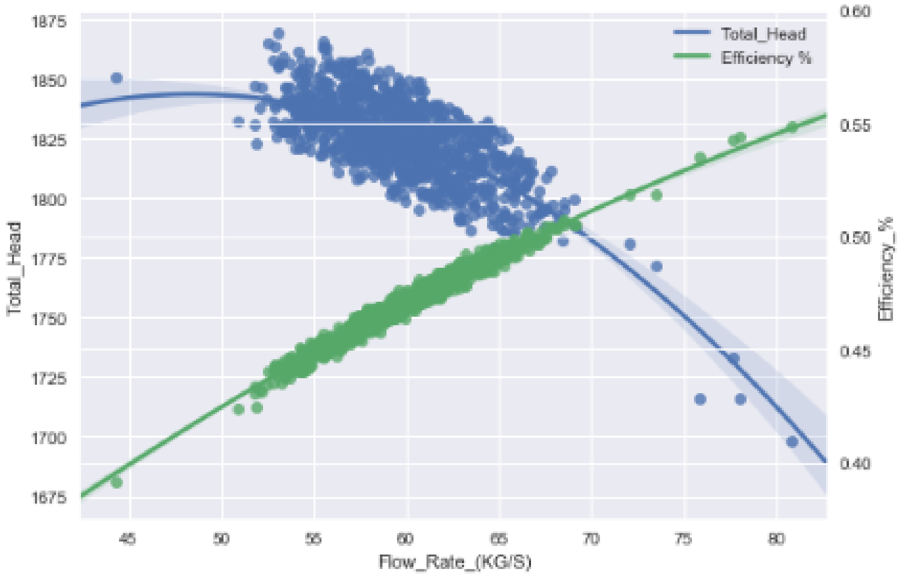


Figure 1 – Benchmark Pump Performance Curve (2017)

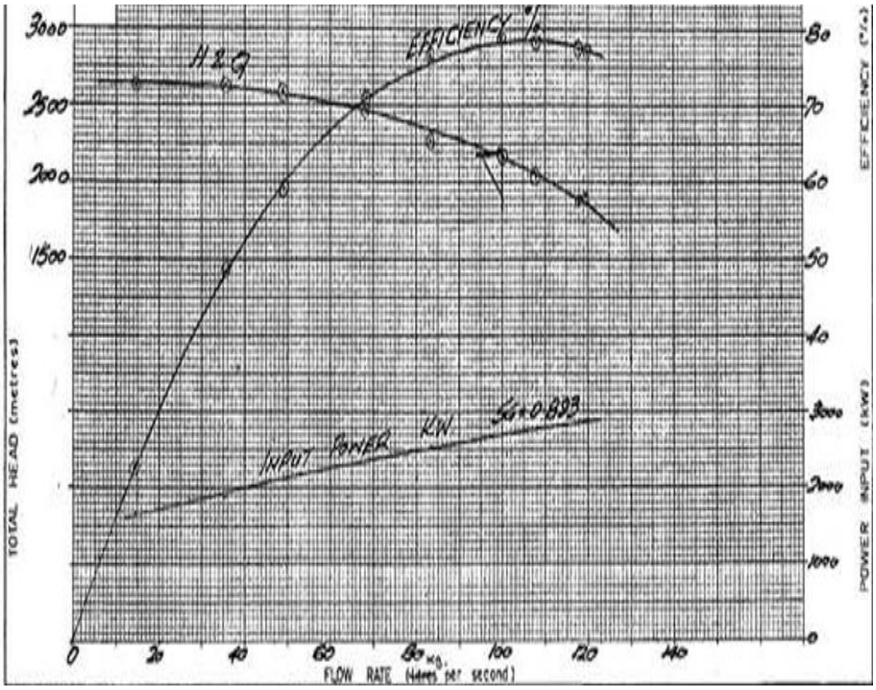


Figure 2 – Benchmark Pump Performance Curve (1960's)

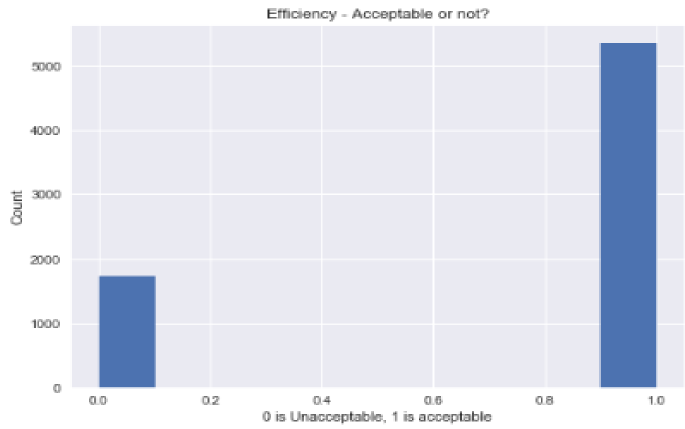
B) Can we increase BFP efficiency by increasing the effectiveness of maintenance activities through the identification of the key weightings of variables contributing towards efficiency?

The current issue is that operators monitor a variety of operational tags – the goal should be to focus on just a handful of key tags which impact efficiency and the associated weightings besides these tags. We will look at exploring this next.

1. With pump efficiency being one of the key measures of pump degradation – it is essential we understand the variables which impact pump efficiency. The equation which calculates pump efficiency is :

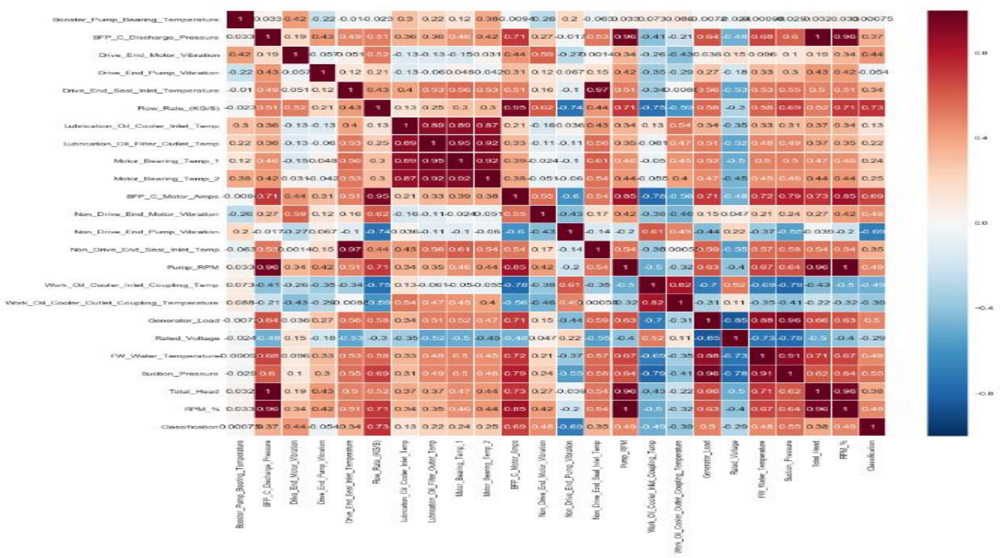
$$Net\ Efficiency = Net\ Output\ /\ Net\ Input$$

2. **The first question we need to ask is – of the baseline efficiency measure of 48%, how many times are we above this or below this?** Observing the below, we can see that we have unacceptable efficiency almost 25% of the time – with this in mind, can we understand what variables are contributing towards efficiency?

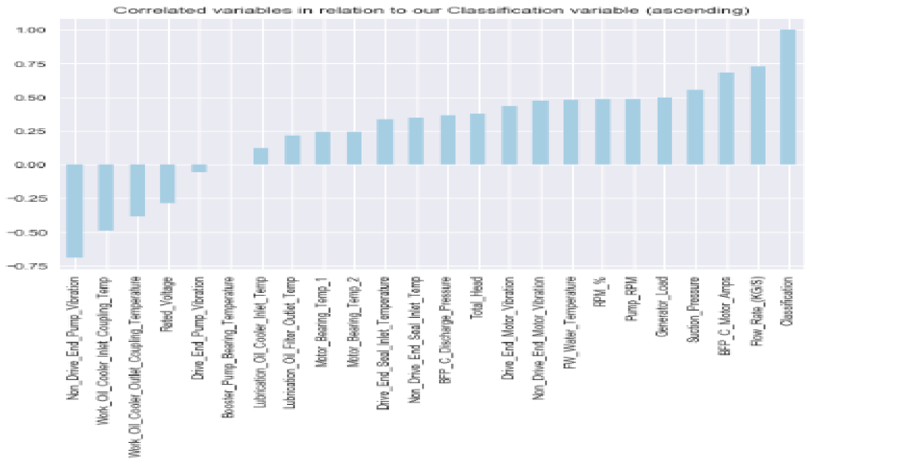



```
The ratio of acceptable efficiency to unacceptable is:
0.76 % acceptable efficiency
0.24 % unacceptable efficiency
```

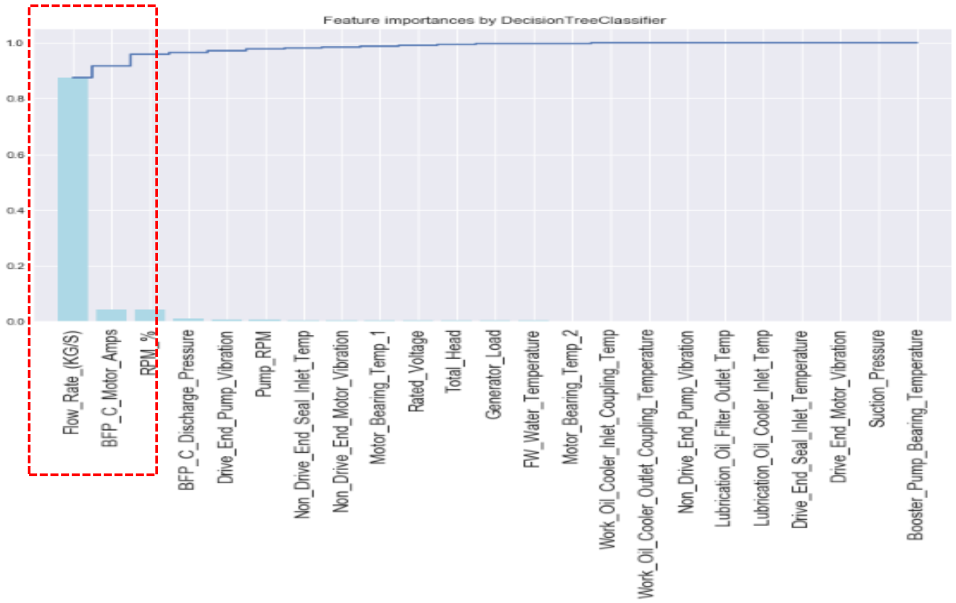
3. Now with our variable, Classification, being our labelled state where 1 is acceptable and 0 is unacceptable, we can ask our second question: **What are the variables which impact efficiency from a correlation score?**



4. The heatmap has highlighted a number of variables which have strong correlations with efficiency – however, the heatmap is a little difficult to read. Using the heatmap's dataframe, we chose the 'classification' criteria and printed this out into an ascending bar plot to make the identification of key variables contributing to efficiency, easier to interpret. We can see that flow rate, motor amps, suction pressure and generator load all seem to be contributing factors. However, this is purely a correlated relationship and does not necessarily dictate causation. Additionally, the correlation is too rough as a weighting. This brings us to our 3rd question: **is there any way to work out which variables contribute the most 'weight' to efficiency?**



5. Thankfully, most machine learning algorithms take dimensionality reduction into account. We are interested primarily in the key two to three variables which provide the greatest explanation of our variability. In this case, we chose to utilise a CART which represents a Classification And Regression Tree Model. Utilising this model reveals the below variables which explain the most variance related to explaining efficiency.



6. Reviewing the variables highlighted by the CART reveal **Flow Rate, Motor Ampage and RPM % being the 3 key variables contributing towards efficiency.** Consulting with the asset engineers and Subject Matter Experts – this makes perfect sense and instantly reduces the number of tags we need to monitor.

   Flow Rate – Every pump has a specific flow rate which needs to be adhered towards, else the pump will be worn away much quicker than expected. This is mentioned on the OEM standards. If we go above or below the safe thresh-holds, the efficiency will either be low, or less than ideal and start the process of degradation quicker than OEM plans.

   Motor Ampage – The amount of energy flowing into the motor is directly related to the amount of energy output from the motor. If the pump is consistently having a higher volume of energy funneled into it, then this means that the motor is working harder which is an indication of motor wear which would also be an indication of efficiency reductions which is a function of wear.

   RPM % - RPM is also an acceptbale indicator of efficiency degradation. Take for example, the more times the pump is rotating at a RPM below the acceptable thresh-hold, the more energy is required to maintain a particular flow rate. With more energy required to maintain a particular flow rate, impacts the amount of energy flowing into the motor. As such, all of these variables are intrinsically linked to one another and provide an accurate overview of the variables which contribute towards efficiency degradation.

<u>Conclusion:</u>

When we set out on this piece of work, we didn't have:

1. No optimisation curves to identify optimal curves to identify pump performance at less than ideal level
2. No weightings behind which variables contribute to pump efficiency amongst many other features which were trending

Now with the completion of this work, we have successfully developed and deployed trial optimisation curves for our boiler feed pumps as well as reducing our OsiSoft PI Monitoring of pump degradation to flow rate, motor ampage and RPM % from 15+ tags to just 3.