

# Southern Water Corporation

Solving an Organization's problem with Data Analytics

*“This project analyzed time series data from production performance over a period of time to predict pump failure in order to aid preventive maintenance schedule and optimize plant availability”*



# Problem Statement Worksheet (Hypothesis Formation)

**How will Southern Water Corporation optimize long term profitability and avoid asset failure whilst reducing maintenance cost by 20% year on year.**

H

## 1 Context

Southern Water Corporation is one of the largest desalination plants in Israel. It retails fresh desalination water for residential, public and private consumption.

Having witnessed a surge in demand from customers, she is taking a strategic overview of the business financial health to determine the impact that scaling up to meet this demand is having on maintenance and life of assets.

## 2 Criteria for success

- Identify a list of variables that may provide an indication of when the pumps may be failing
- Come up with a prototype linear equation that can be used to 'describe' what variables are closely related to pump failure

## 3 Scope of solution space

Focus on controlling operational costs and preventing untimely asset failure which will lead to eventual loss of revenues

## 4 Constraints within solution space

- Production Manager's goal of meeting production target may be at variance with Reliability Engineer's goal of shutting down the plant for maintenance.
- Data on past machine failures may not necessarily and accurately predict future failure dynamics.

## 5 Stakeholders to provide key insight

- Head of Finance –Joanne O'Neil
- Production Manager– Melanie Griggs
- Head of Analytics – Andrew Xu
- Chief scientist - Joanna Luez
- Reliability/Maintenance Engineer – John Lukes

## 6 Key data sources

- Income Statement – year to Date
- Production Data – year to Date
- Maintenance Data- year to Date
- Cash Flow Statement – year to Date

H

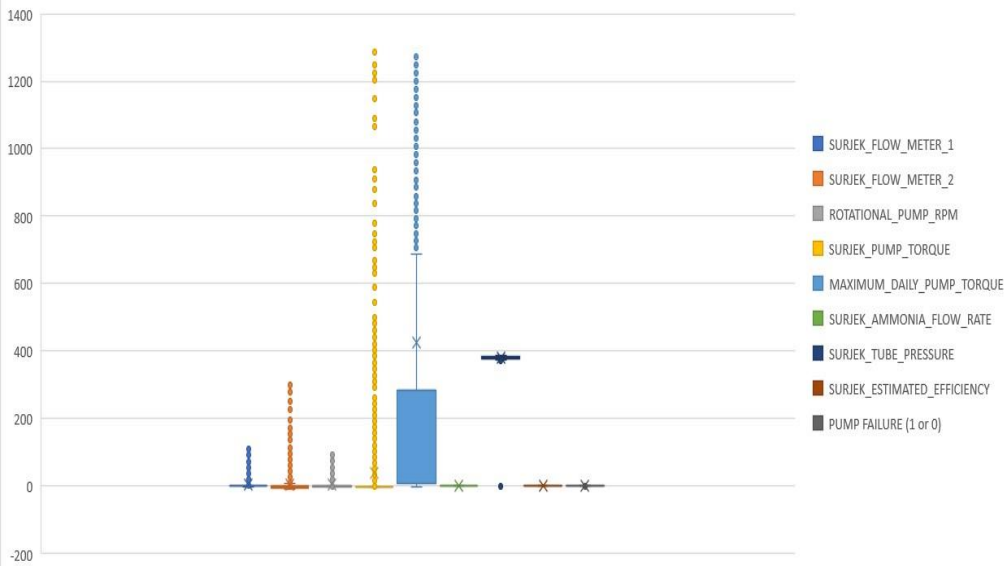
D

E

I

P

All Variables

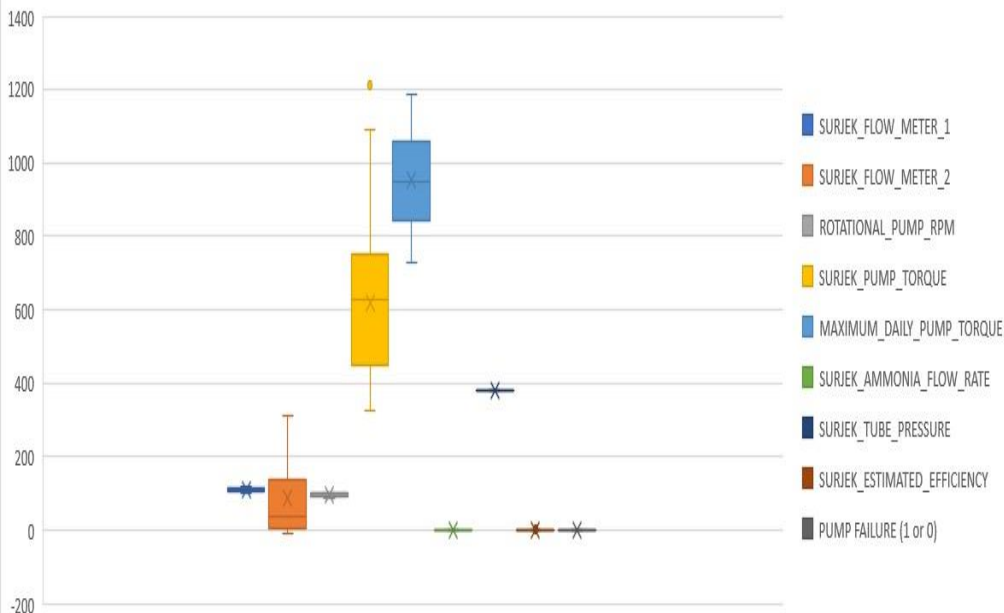


## Comments:

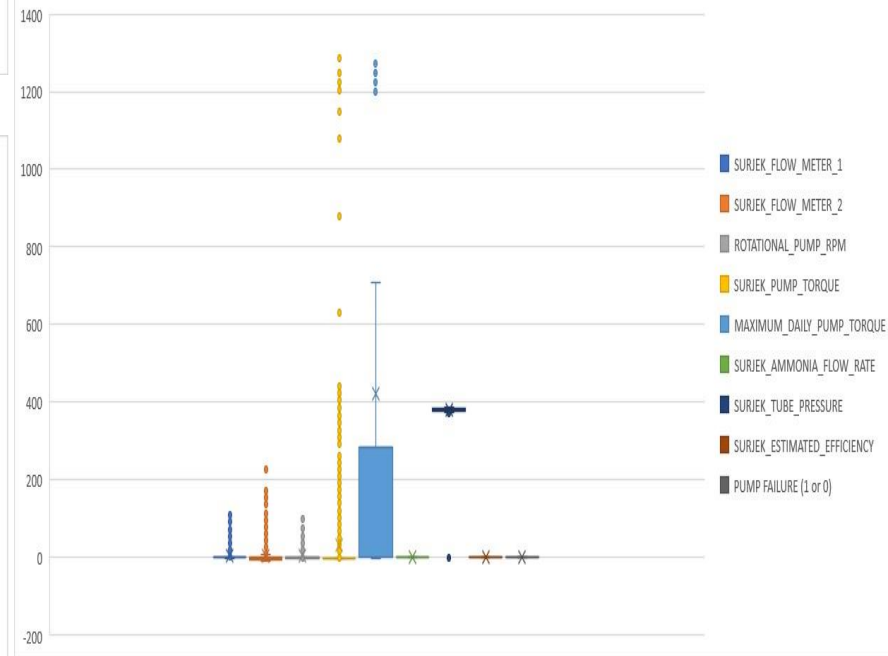
**All Variables** – **Max daily pump torque** and Surjek pump torque seems to stand out with highest data spreads

**Pump Failure mode =Yes/No** - **Max daily pump torque** and Surjek pump torque are still the ones moving significantly with the pump failure mode.

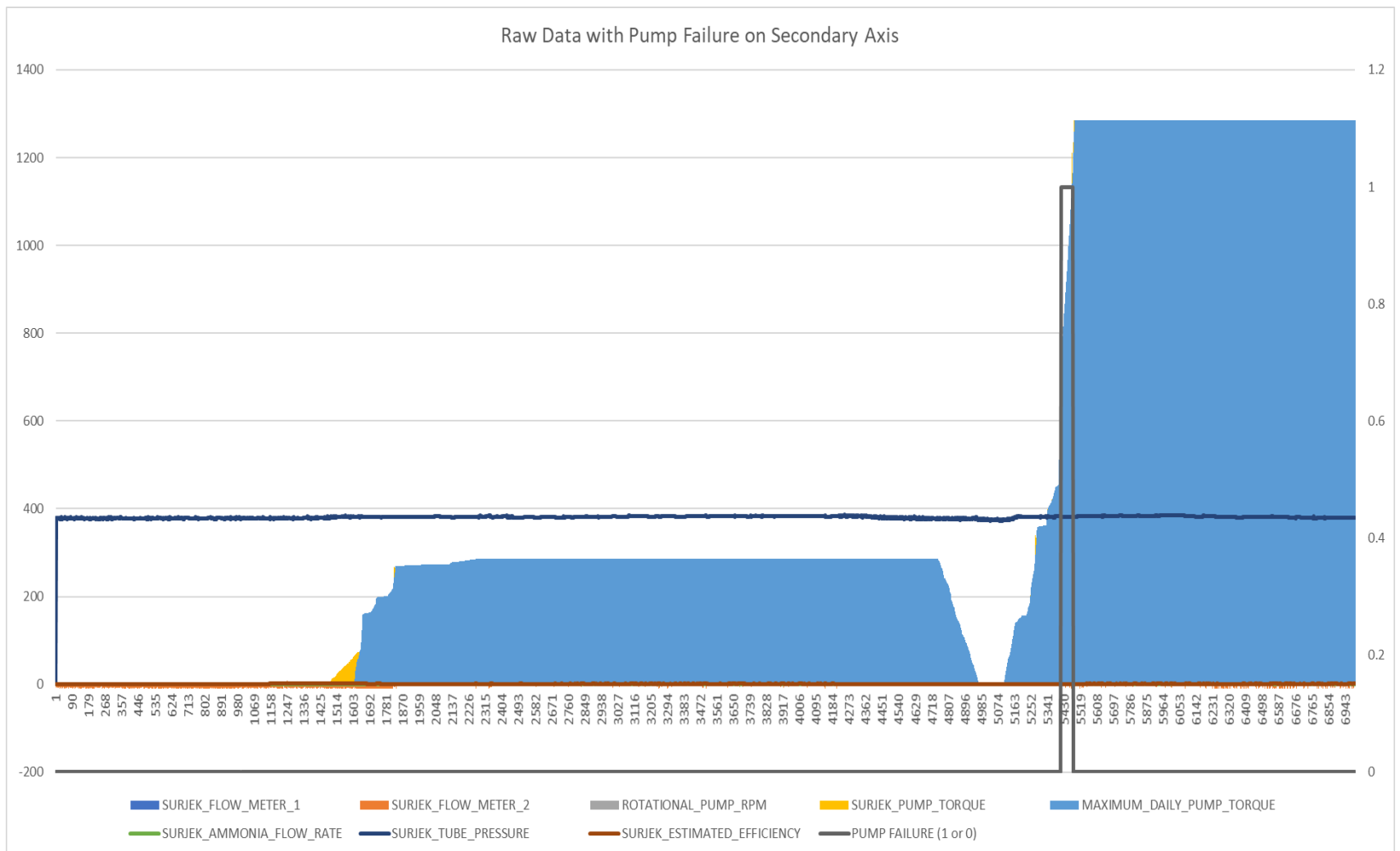
Pump Failure = Yes



Pump Failure = No



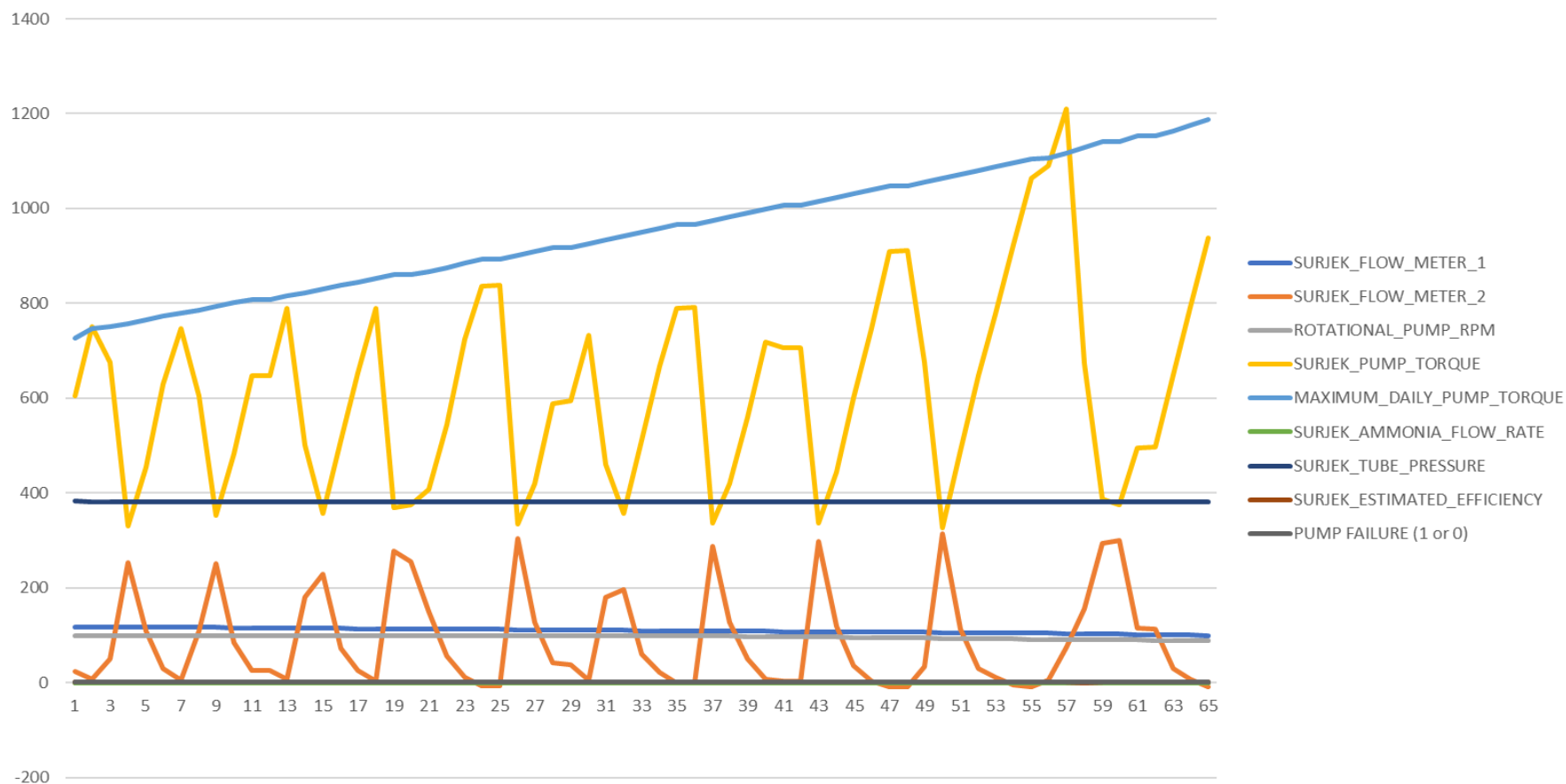
**Correlation/Causation**- It's safe to conclude that there's a correlation between **max\_daily\_pump\_torque** and Surjek\_Pump\_Torque with respect to pump failures but this does not necessarily depict causation.



### Comments:

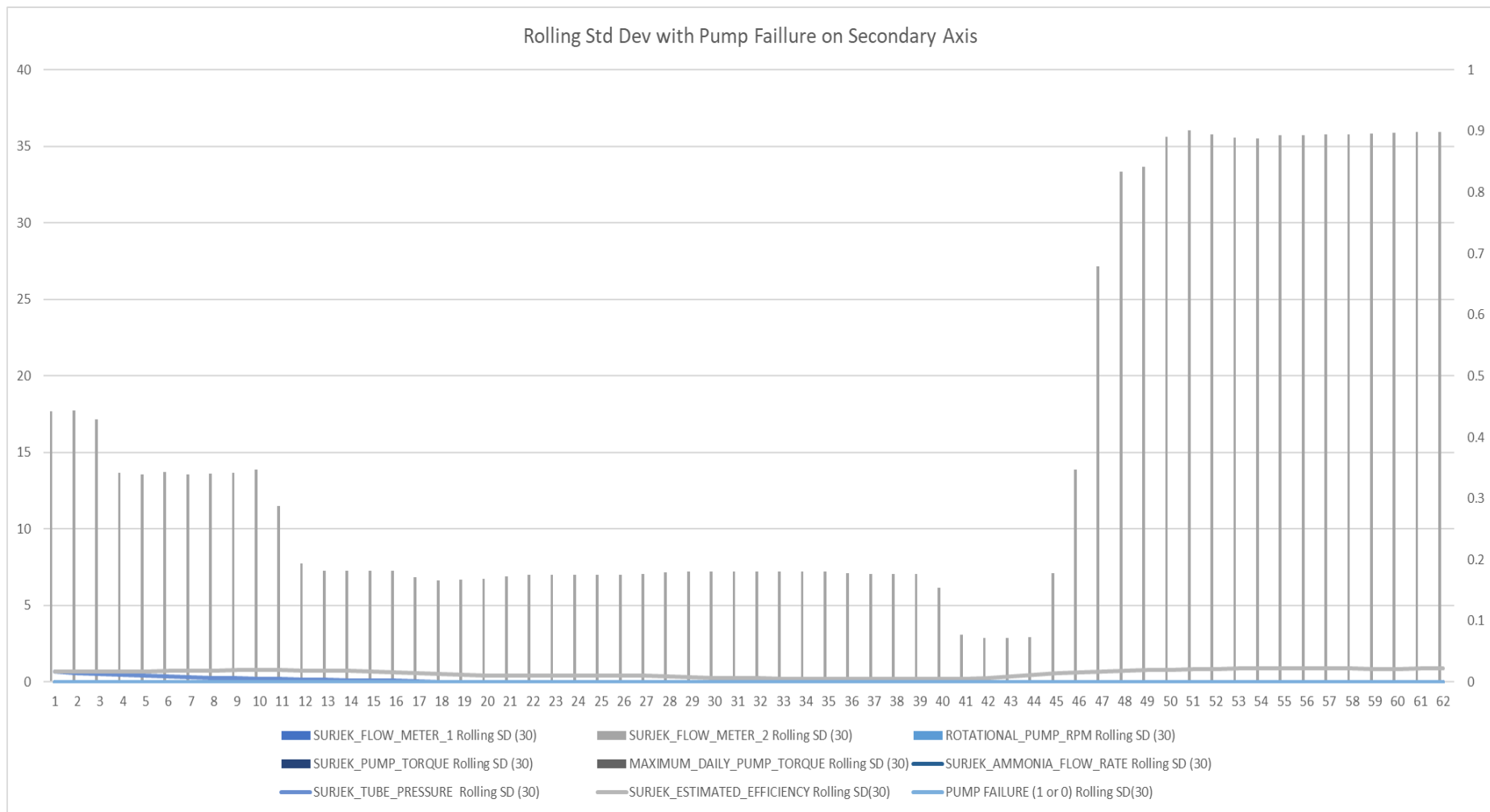
'Max daily pump torque' stands out as a major determinant of pump failure (blue region) and a little bit of surjek pump torque (yellow region)

**Plot re-failure** (all variables for the time period 10/12/2014 14:40 to 10/12/2014 14:45)



### Comments:

Only 'Max daily pump torque' seem to move slightly with the pump failure when we zoom in on pump failure mode while Surjek Pump torque has the maximum variation.

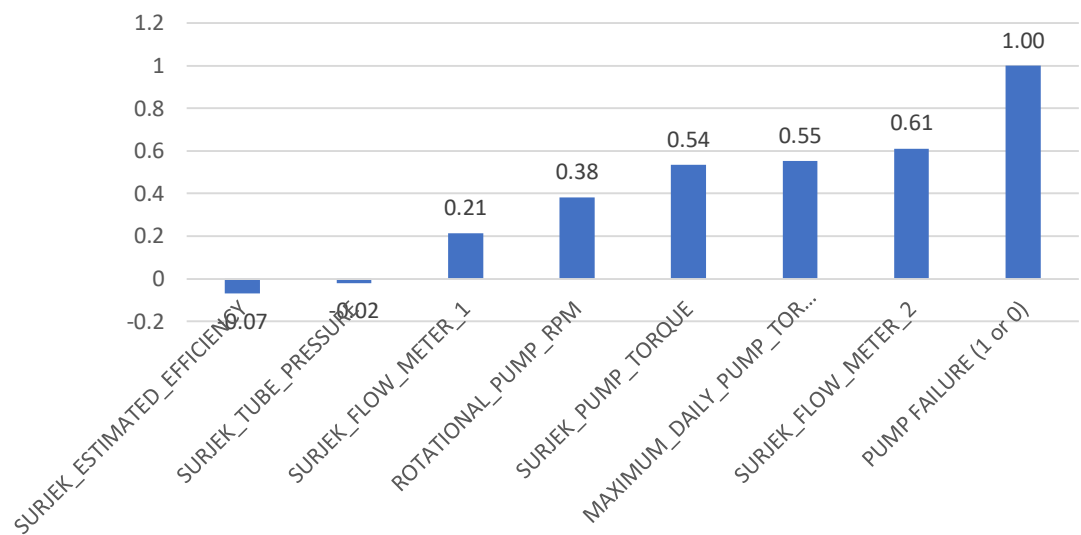


### Comments:

Only 'Surjek Flow Meter 2' stands out this time when zoomed in for time period 12/10/2014 14:40 to 12/10/2014 14:45. Bu there is no pump failure here.

# Correlation and Regression Analysis

Correlation with Pump Failure



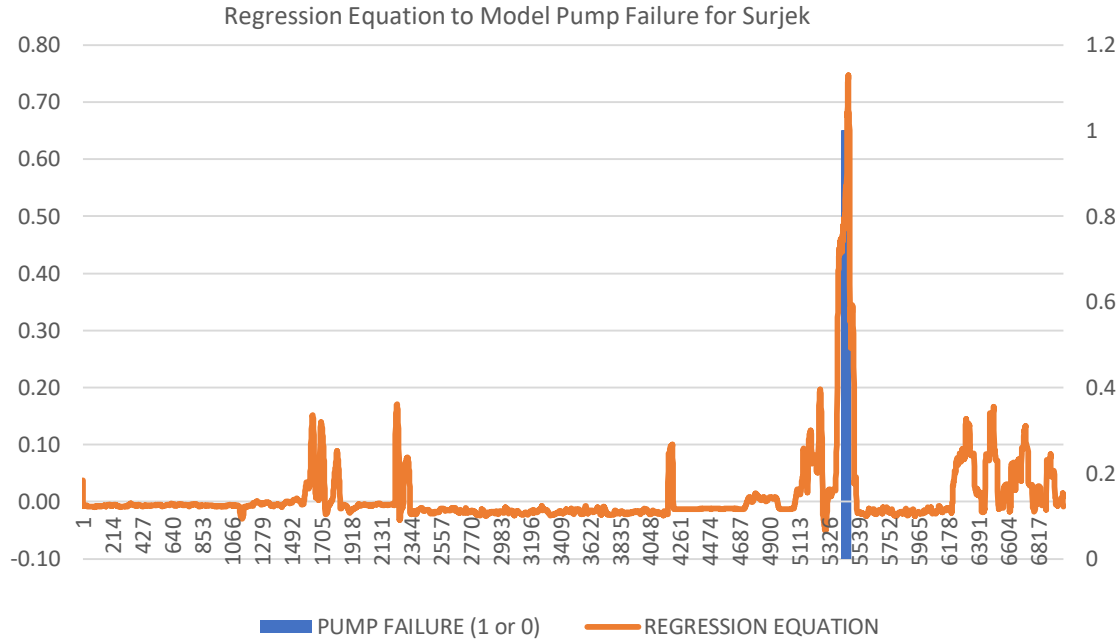
Correlation > 0.5 is very significant

SURJEK_ESTIMATED_EFFICIENCY	-0.06984408
SURJEK_TUBE_PRESSURE	-0.02052327
SURJEK_FLOW_METER_1	0.21342086
ROTATIONAL_PUMP_RPM	0.382259028
SURJEK_PUMP_TORQUE	0.53529471
MAXIMUM_DAILY_PUMP_TORQUE	0.553287393
SURJEK_FLOW_METER_2	0.611275356
PUMP FAILURE (1 or 0)	1
SURJEK_AMMONIA_FLOW_RATE	#DIV/0!

Correlation Matrix

	SURJEK_FLOW_METER_1	SURJEK_FLOW_METER_2	ROTATIONAL_PUMP_RPM	SURJEK_PUMP_TORQUE	MAXIMUM_DAILY_PUMP_TORQUE	SURJEK_AMMONIA_FLOW_RATE	SURJEK_TUBE_PRESSURE	SURJEK_ESTIMATED_EFFICIENCY	PUMP_FAILURE (1 or 0)
SURJEK_FLOW_METER_1	1								
SURJEK_FLOW_METER_2	0.092376911	1							
ROTATIONAL_PUMP_RPM	0.247884943	0.15905218	1						
SURJEK_PUMP_TORQUE	0.586672049	0.405336665	0.411534798	1					
MAXIMUM_DAILY_PUMP_TORQUE	0.154925109	0.516477227	0.295968285	0.458389364	1				
SURJEK_AMMONIA_FLOW_RATE	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	1			
SURJEK_TUBE_PRESSURE	-0.008312236	-0.069016452	-0.014058326	-0.02662968	-0.006651472	#DIV/0!	1		
SURJEK_ESTIMATED_EFFICIENCY	-0.101375313	0.101283675	-0.01110262	-0.061754828	-0.185131261	#DIV/0!	-0.15220936	1	
PUMP_FAILURE (1 or 0)	0.21342086	0.611275356	0.382259028	0.53529471	0.553287393	#DIV/0!	-0.020523274	-0.069844078	1

# Regression Equation to Model Pump Failure



## Summary Statistics of Correlation and Regression Analysis

	Coefficients
Intercept	-0.013022
SURJEK_FLOW_METER_1	-0.001472
SURJEK_FLOW_METER_2	0.0028563
ROTATIONAL_PUMP_RPM	0.0046253
SURJEK_PUMP_TORQUE	0.0005505
MAXIMUM_DAILY_PUMP_TORQUE	0.001635
SURJEK_AMMONIA_FLOW_RATE	0
SURJEK_TUBE_PRESSURE	0.0004774
SURJEK_ESTIMATED EFFICIENCY	-0.019583

## Multivariate Regression Equation to Model Pump Failure

$$Y (\text{Pump Failure}) = -0.002(\text{SURJEK\_FLOW\_METER\_1}) + 0.003(\text{SURJEK\_FLOW\_METER\_2}) + 0.005(\text{ROTATIONAL\_PUMP\_RPM}) + 0.001(\text{SURJEK\_PUMP\_TORQUE}) + 0.002(\text{MAXIMUM\_DAILY\_PUMP\_TORQUE}) + 0.001(\text{SURJEK\_TUBE\_PRESSURE}) - 0.020(\text{SURJEK\_ESTIMATED\_EFFICIENCY}) - 0.013$$

In summary our overall analysis reveals that SW Corp need to pay particular attention to the following variables which are strongly correlated with pump failure:

# SURJEK\_PUMP\_TORQUE

# MAXIMUM\_DAILY\_PUMP\_TORQUE and

# SURJEK\_FLOW\_METER\_2