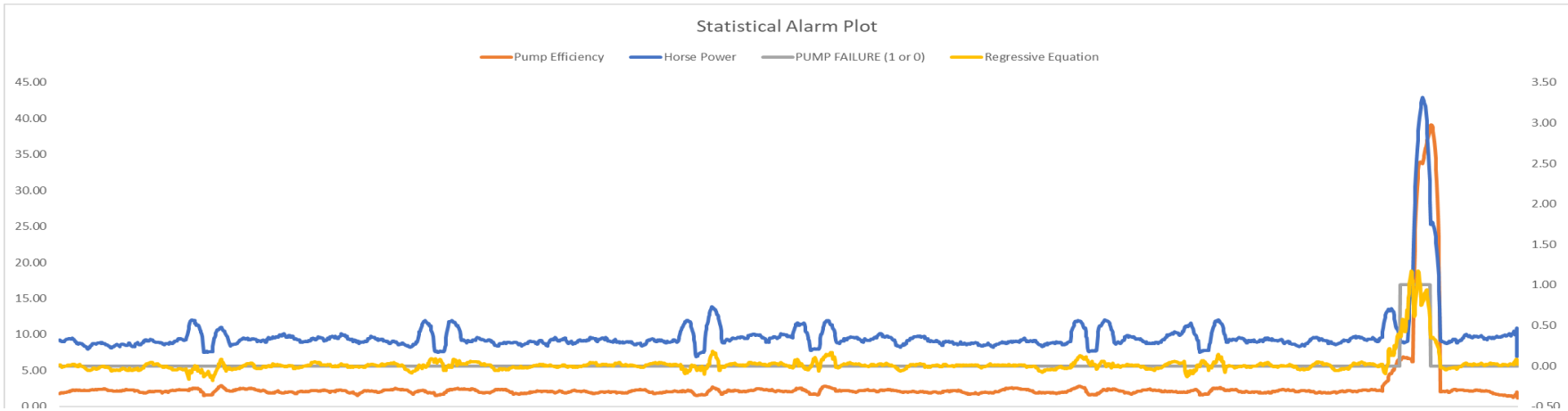
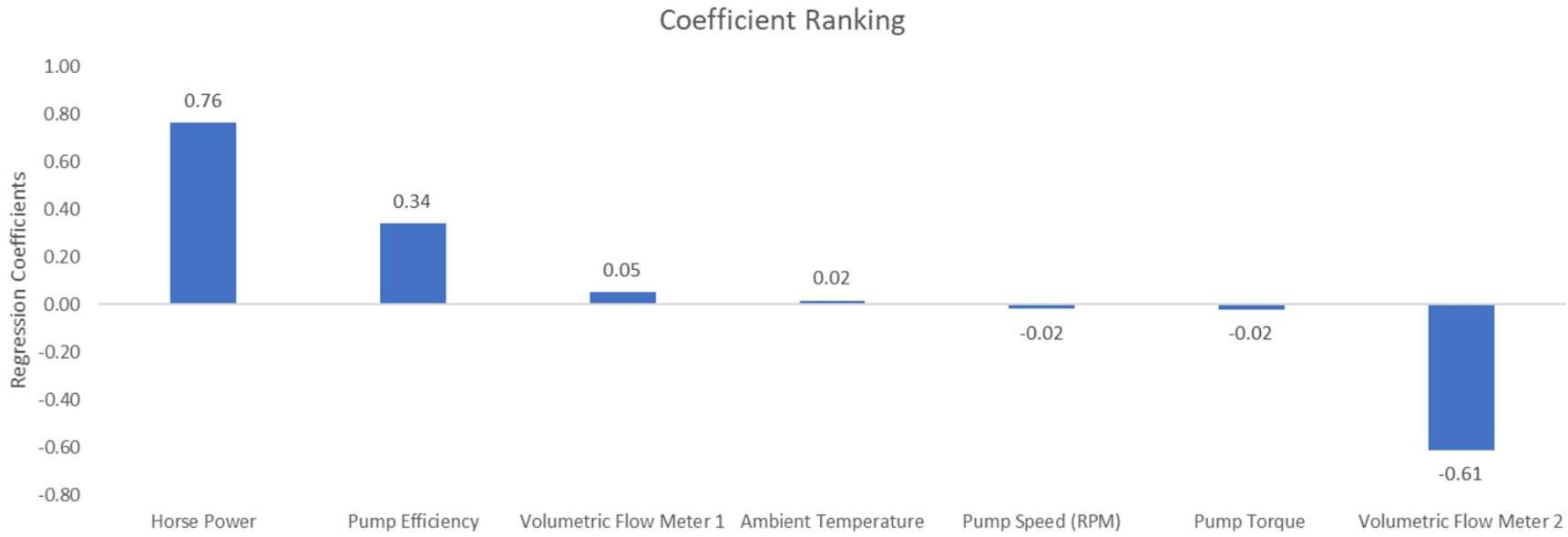
Multivariate Regression Analysis to Detect Pump Failure and Set Statistical Data Alarm for Southern Water Corp

By: Kidanemariam Hailu

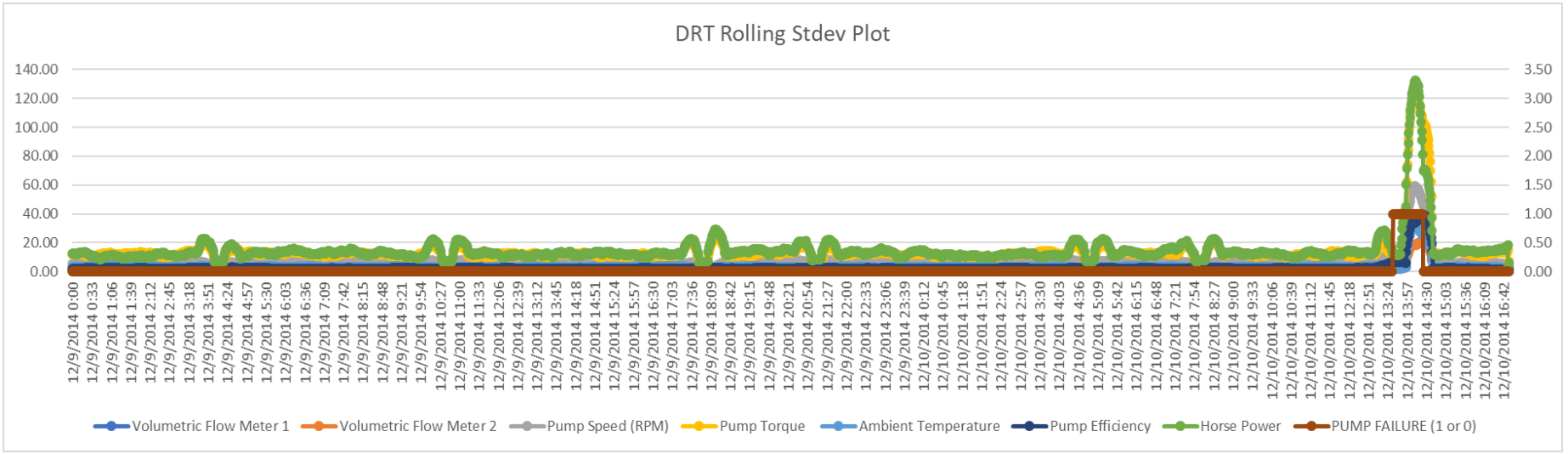
August 2022

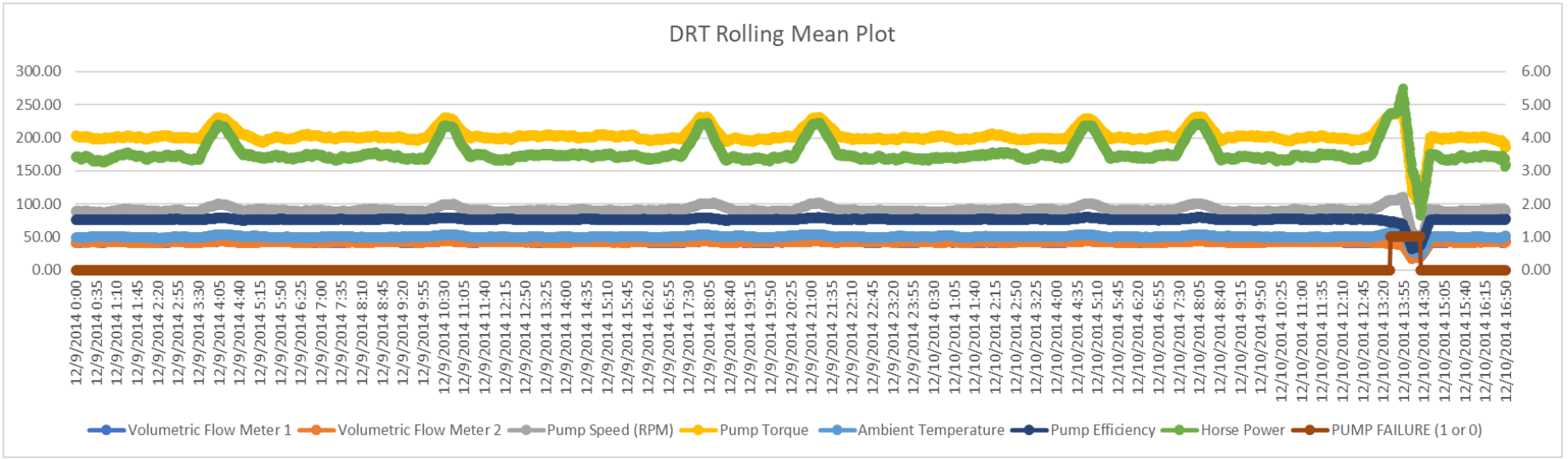
Descriptive and inferential statistical methodologies have proven effective in creating a proactive ‘alarm’, accurately identifying Pump Failures with Horse Power (HP) and Pump Efficiency (PE) emerging as key variables of interest with deviations of 15 HP and > 3 % PE being our core signal thresholds.



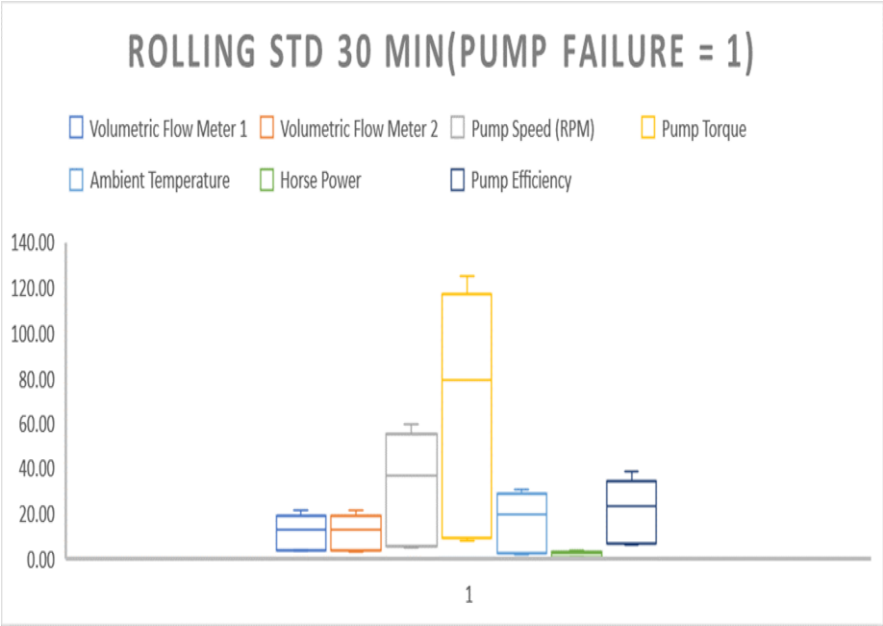


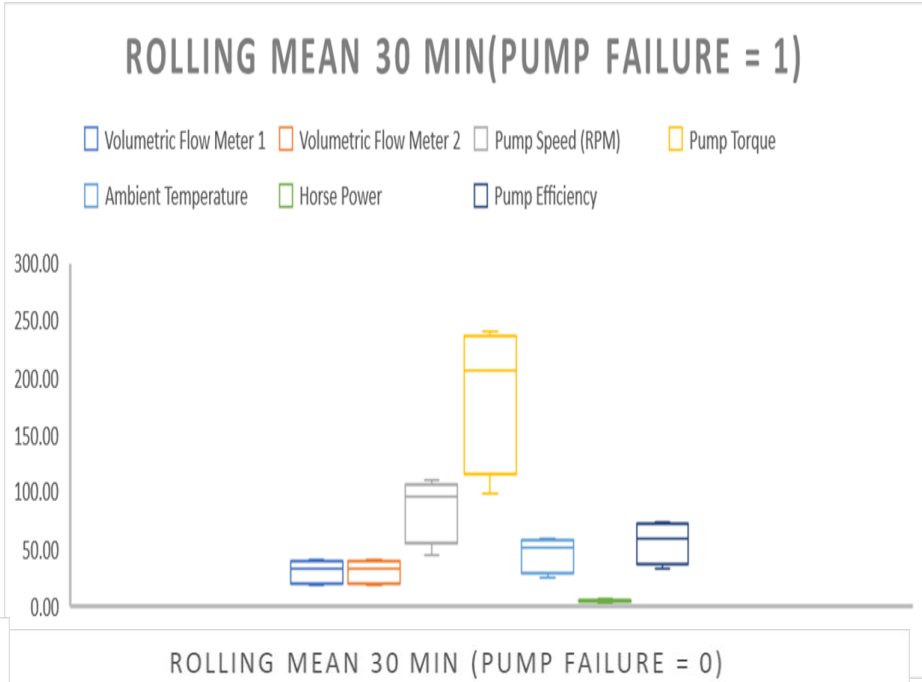
Descriptive Analysis has enabled us to clearly identify particular signature abnormalities showing clear signature changes in both Rolling Standard Deviation and Rolling Mean Datasets when observed over the respective failure period of interest.

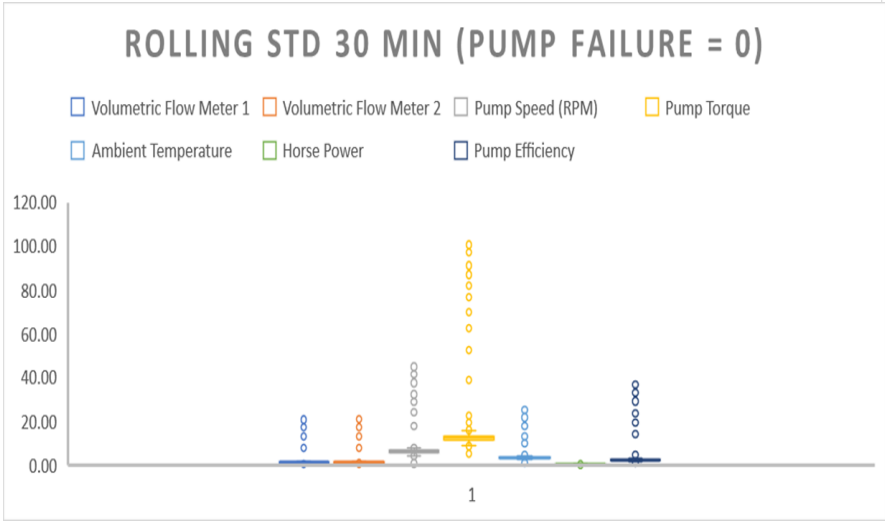


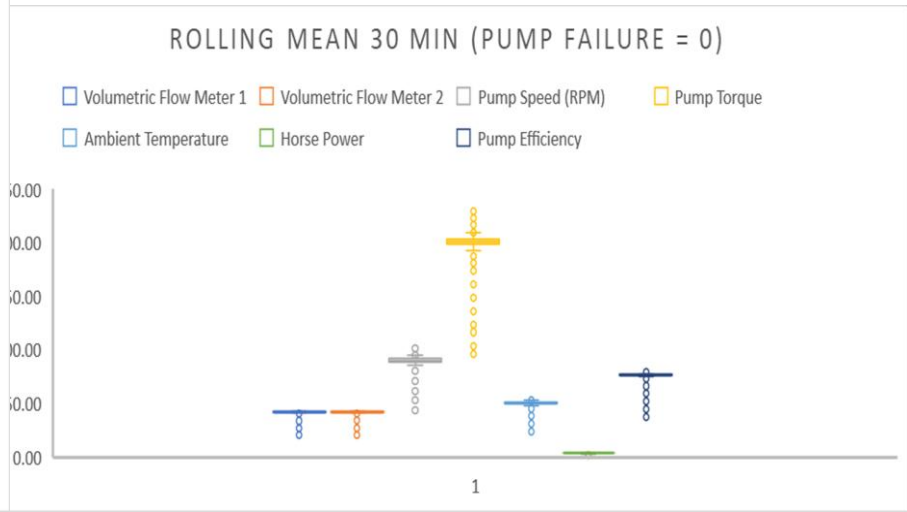


Further segmentation of the data via binary means (Pump Failure = 0 or 1) illustrated through BoxPlots, show a clear signature difference between that of normal behaviour and that of Failure with Pump Torque, Pump speed, Pump Efficiency showing the 3 largest variances.

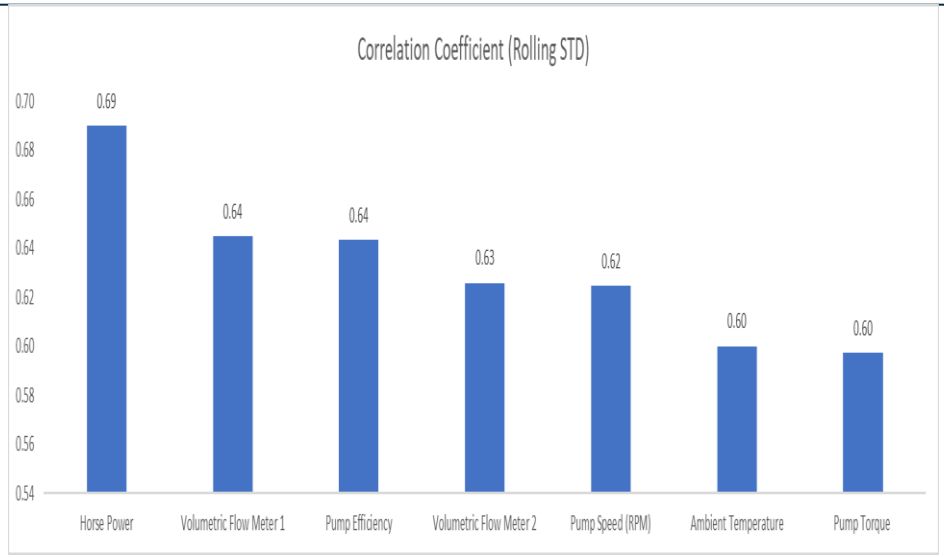


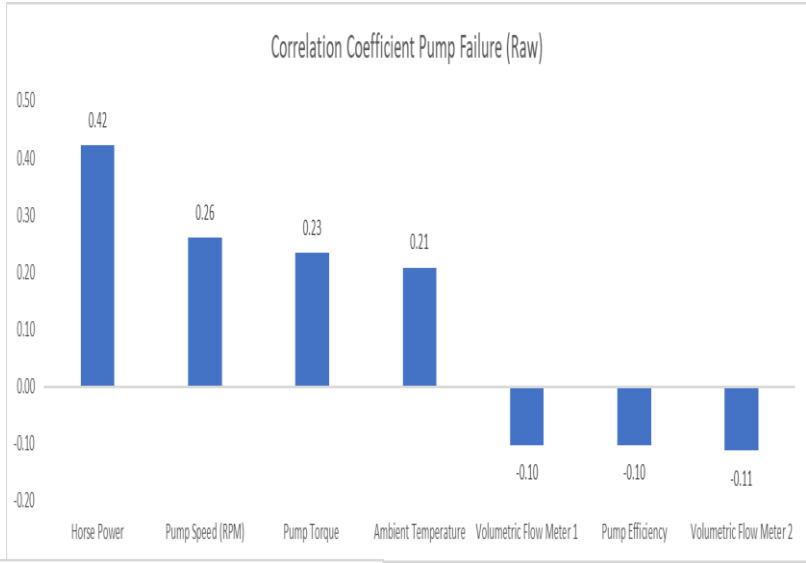


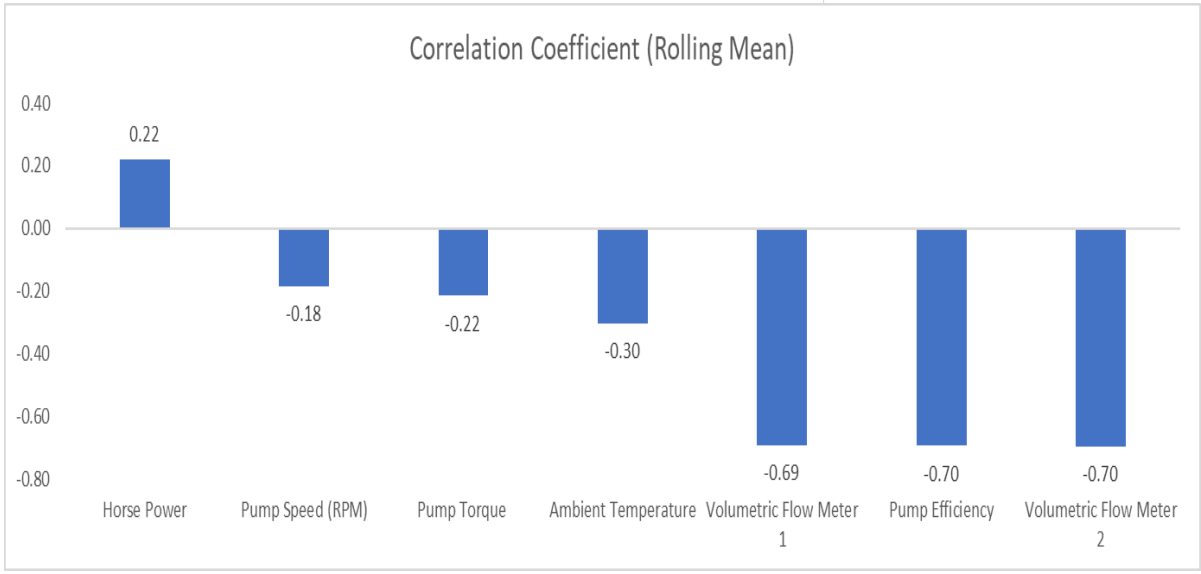




Correlation analyses across datasets yield interesting insights with Pump efficiency and Volumetric Flow Meter 2 negatively correlated with Pump Failure in the Rolling Mean Data, whilst Horse Power and Pump Efficiency show a subsequently strong positive correlation in the Rolling Stdev Dataset.







Lastly, analysis of the model fit reveals that with a R Squared of 77.9%, a linear model is a good fit for the data with variables Horse Power, Pump efficiency and Volumetric Flow Meter 2 having the largest coefficients, indicative that these variables have the most immediate relationship with respect to Pump Failure behaviour.