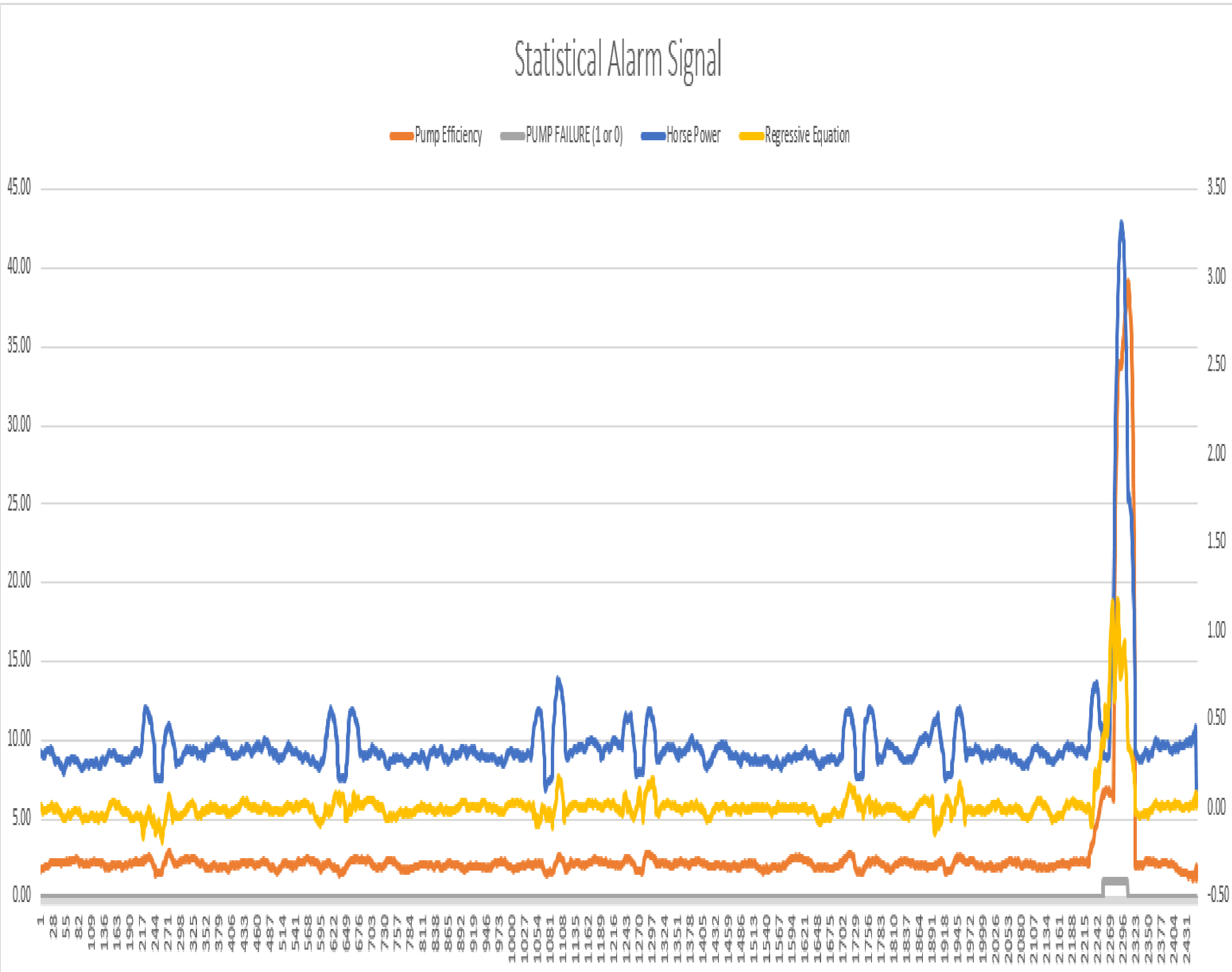
Statistics Analysis Report

Objectives:

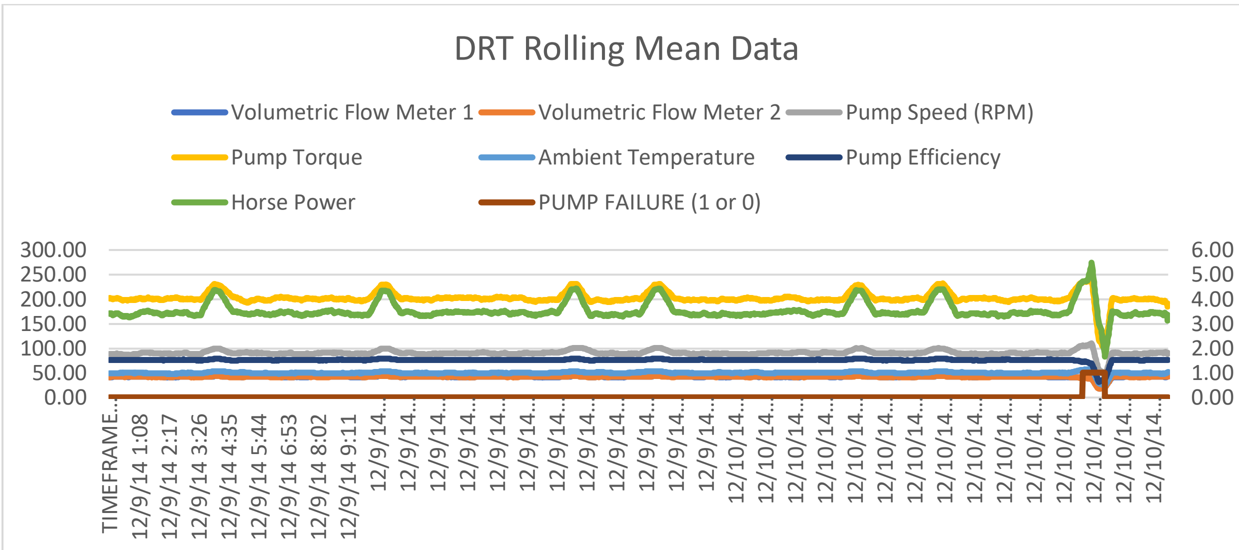
1. Using a line-series chart, show the Multivariate Regression Equation Values you've developed versus the Pump Failure (1 or 0) and Using a Column Chart, show the correlations ranked from highest-to-lowest for the Rolling Standard Deviation Dataset
2. Create a Rolling Standard Deviation Plot and Rolling Mean Plot
3. Create a Rolling Standard Deviation Plot (Pump Failure = 1) , Rolling Standard Deviation Plot (Pump Failure = 0), Rolling Mean Plot (Pump Failure = 1), and Rolling Mean Plot (Pump Failure = 0)
4. Raw Dataset Correlation against Pump Failure, Rolling Mean Dataset Correlation against Pump Failure, and Rolling Standard Deviation Dataset Correlation against Pump Failure
5. Create a column-chart from the Multivariate Regression which shows the Regression Coefficients ordered from highest to lowest (i.e. The variable with the largest co-efficient has the highest rate-of-change with respect to Pump Failure)

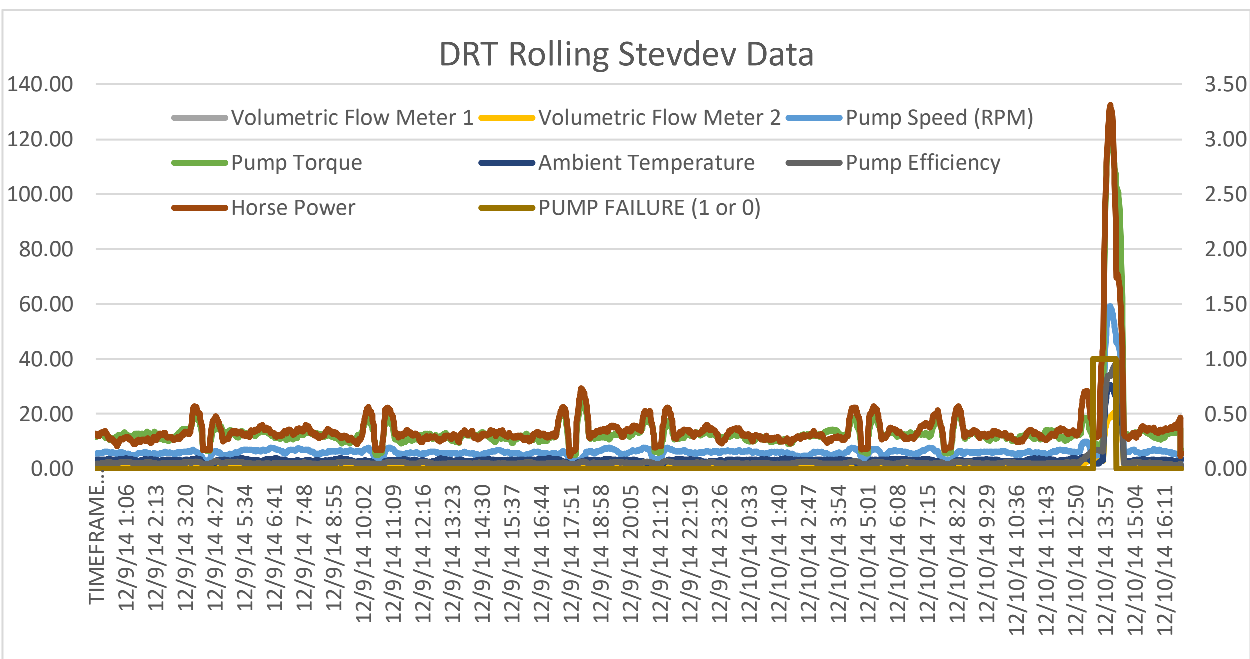
Analysis:

1. 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.

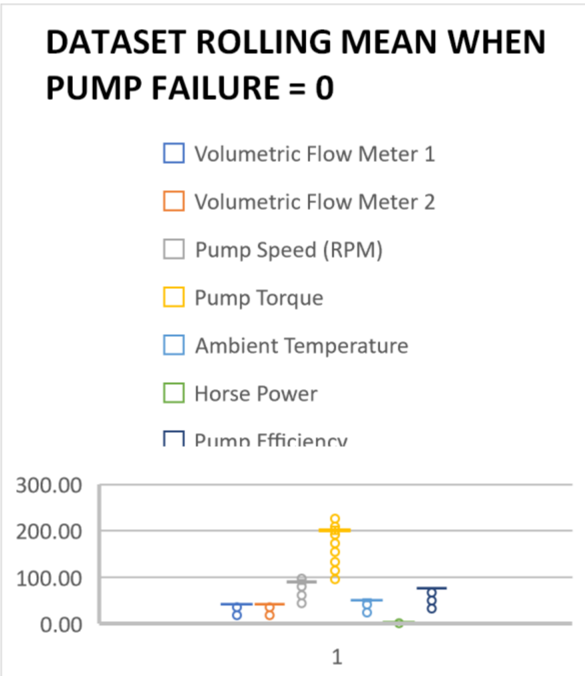


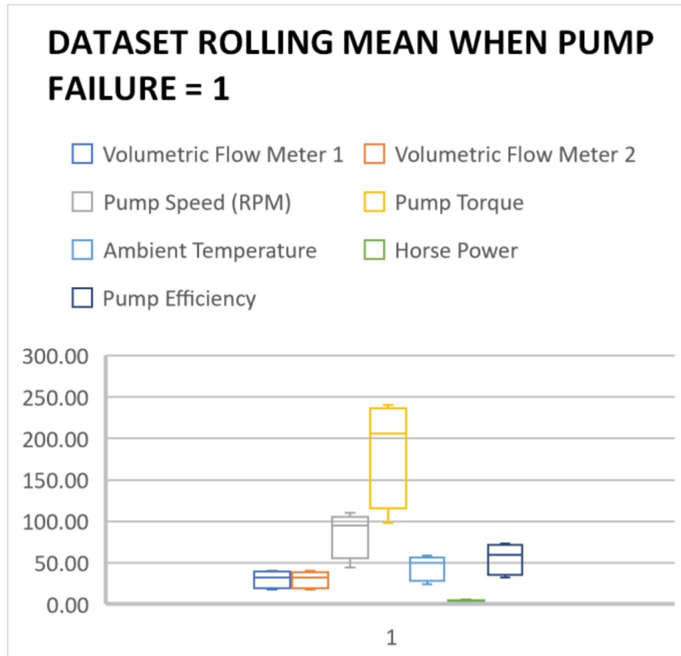
2. Descriptive Analysis has enabled us to clearly identify particular signature abnormalities showing clear signature changes in both Rolling Standard Deviation and Rolling Mean Datasets when observed over the respective failure period of interest.

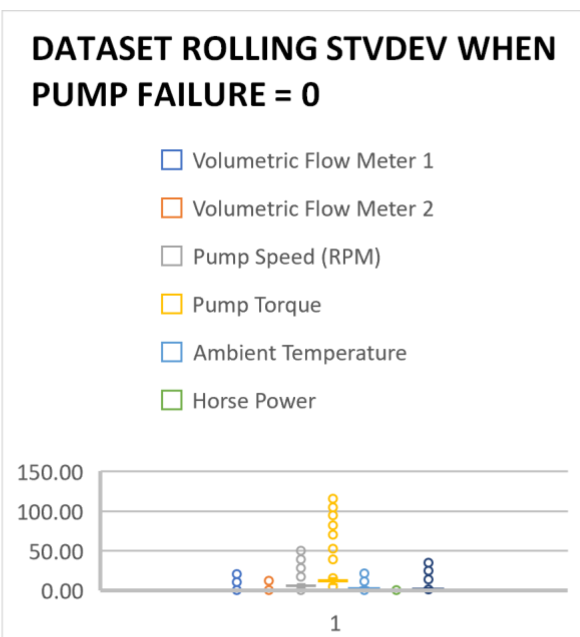


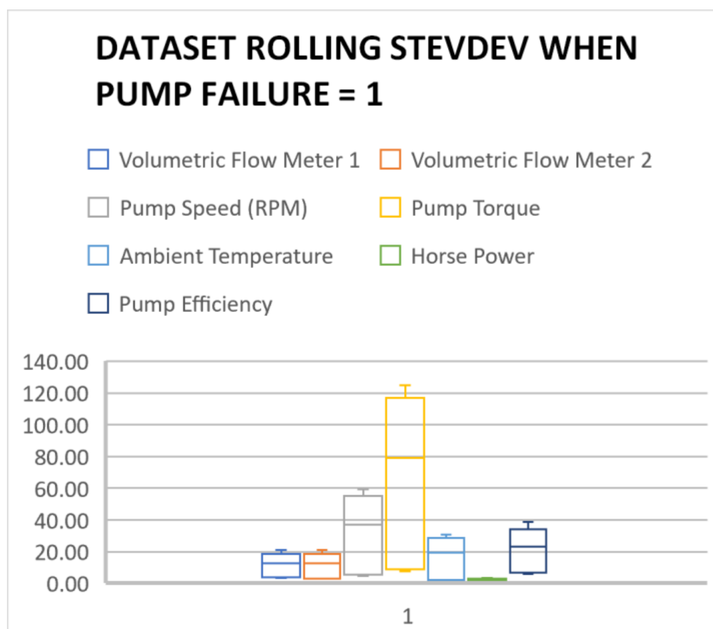


3. Compare to above all box plots the Rolling Mean when pump Failure = 0 there are many bottom outliers which means it has lower median values. On the other hand, Rolling Stdev when pump failure is = o the top outliers' values are higher than median. Compare to Rolling Mean when pump failure = 1 the top wiskers has low values than median and for Rolling Stdev when pump failure = 1 the top wisker has more than median.

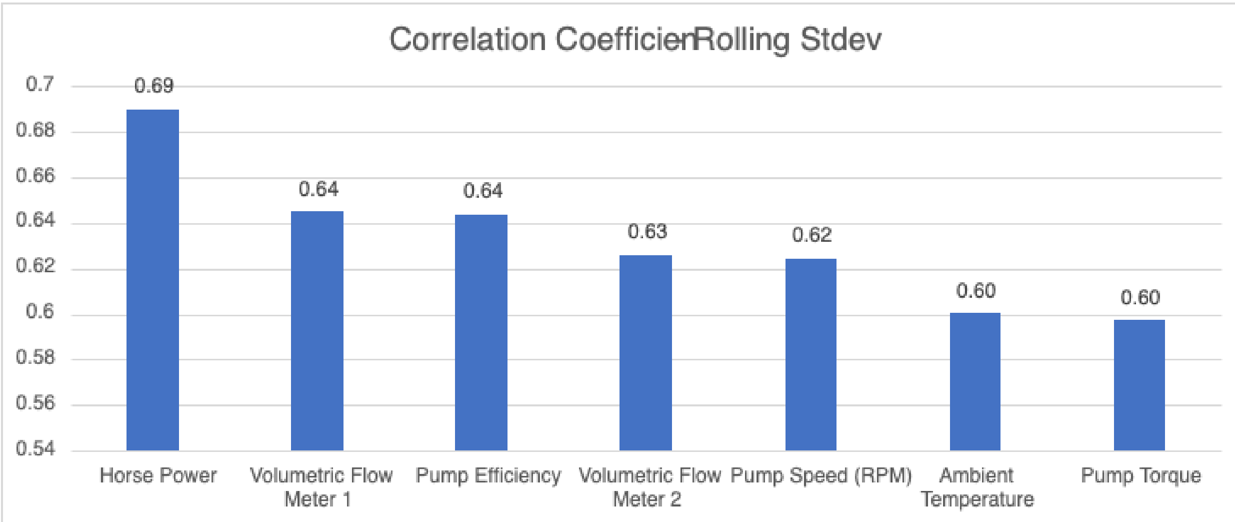


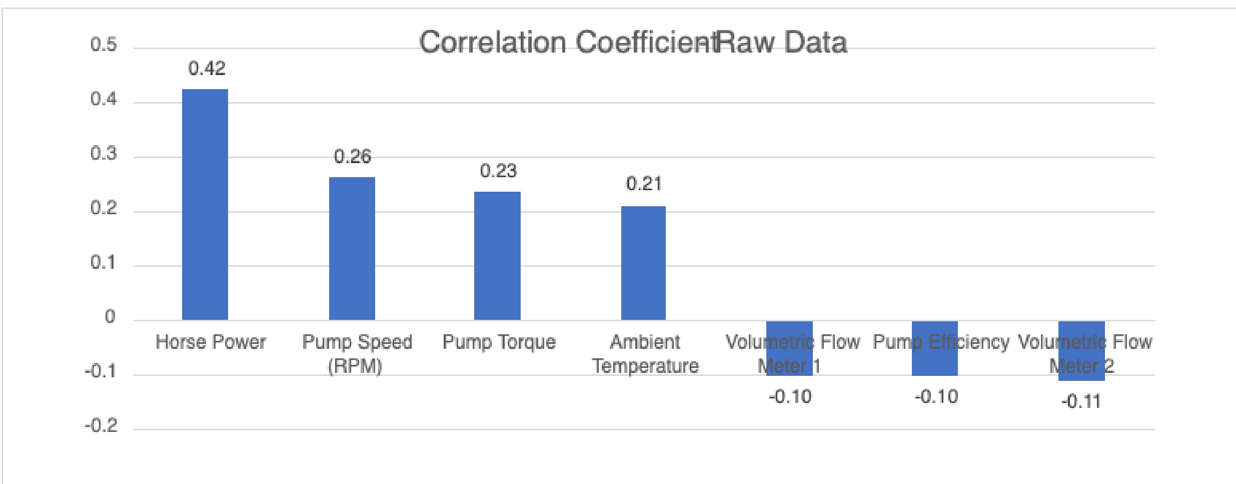


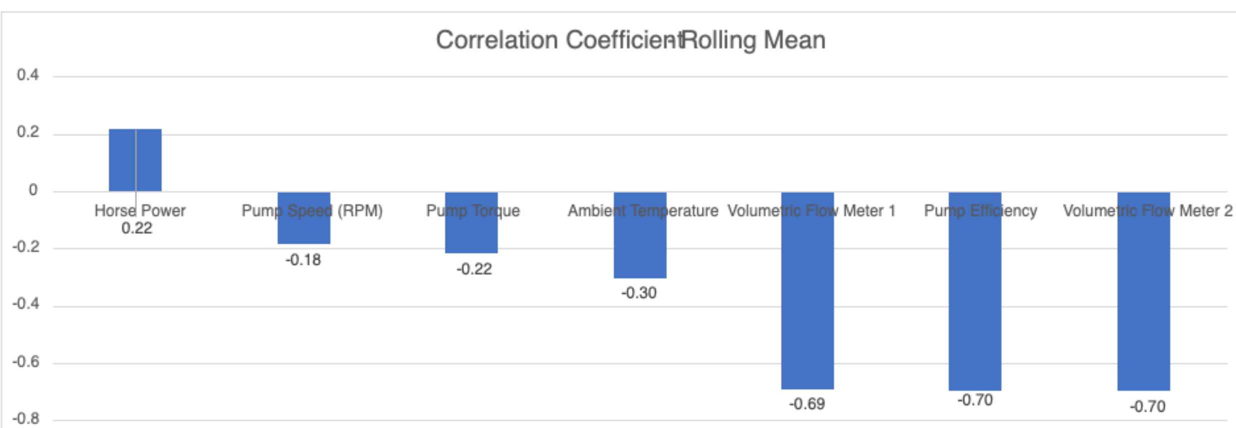




4. Correlation analyses across datasets yield interesting insights with pump efficiency and volumetric flow meter 2 negatively correlated with Pump Failure in the Rolling Mean Data, whilst Pump Efficiency and Volumetric flow 1 show a subsequently strong positive correlation in the Rolling Stdev Dataset.







5. Lastly, analysis of the model fit reveals that with a R Squared of 77% or .77, a linear model is a good fit for the data with variables X, Y and Z having the largest coefficients, indicative that these variables have the most immediate relationship with respect to Pump Failure behaviour.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.882543972
R Square	0.778883863
Adjusted R Square	0.778250552
Standard Error	0.067858879
Observations	2452

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	7	39.6430286	5.6632898	1229.8593	0
Residual	2444	11.2541982	0.00460483		
Total	2451	50.8972268			

	Coefficients	*Standard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-0.035779667	0.00351939	-10.1664396	8.2726E-24	-0.04268096	-0.02887837	-0.04268096	-0.02887837
Volumetric Flow Meter	0.050933944	0.01002225	5.08208765	4.0155E-07	0.03128097	0.07058692	0.03128097	0.07058692
Volumetric Flow Meter	-0.61136433	0.01997618	-30.6046732	2.033E-174	-0.65053631	-0.57219235	-0.65053631	-0.57219235
Pump Speed (RPM)	-0.017867336	0.0019615	-9.10901584	1.683E-19	-0.02171371	-0.01402096	-0.02171371	-0.01402096
Pump Torque	-0.018901995	0.00083707	-22.581109	1.055E-102	-0.02054344	-0.01726055	-0.02054344	-0.01726055
Ambient Temperature	0.018206887	0.00308731	5.89733444	4.2052E-09	0.01215288	0.0242609	0.01215288	0.0242609
Horse Power	0.764845388	0.02177434	35.1259961	3.552E-219	0.72214732	0.80754346	0.72214732	0.80754346
Pump Efficiency	0.34236589	0.01211294	28.2644687	2.592E-152	0.3186132	0.36611858	0.3186132	0.36611858

Conclusion:

The descriptive analysis conducted has allowed us to identify notable patterns and abnormalities in the datasets related to pump failure. Specifically, the Rolling Standard Deviation and Rolling Mean datasets exhibit distinct changes during the failure period of interest. Correlation analysis revealed interesting insights, such as negative correlations between pump efficiency and volumetric flow meter 2 with pump failure in the Rolling Mean Data, and strong positive correlation between pump efficiency and volumetric flow 1 in the Rolling Standard Deviation dataset.

Further segmentation of the data based on pump failure (0 or 1) highlighted significant differences in behavior. Pump torque, pump efficiency, and pump speed showed the largest variances between normal behavior and failure instances. Additionally, model fitting analysis indicated a good fit with a linear model, with an R-squared value of 0.77. Horsepower, pump speed, and pump torque were identified as the key variables with the largest coefficients, suggesting a significant relationship with pump failure behavior.

Overall, the descriptive and inferential statistical methodologies employed have proven effective in creating a proactive alarm system to accurately identify pump failures. Horsepower and pump efficiency have emerged as crucial variables of interest, with deviations of 15 HP and more than 3% PE serving as core thresholds for signaling potential issues.

Recommendations:

Based on the analysis, the following recommendations can be made to improve the proactive identification of pump failures:

- Implement a real-time monitoring system for the rolling standard deviation and rolling mean datasets. This will enable the detection of signature abnormalities that may indicate potential pump failures, allowing for early intervention to prevent further damage.
- Focus on improving pump efficiency by implementing regular maintenance programs and monitoring key performance indicators such as Horse Power and Pump Efficiency closely. Deviations of 15 HP and > 3% PE should be used as core signal thresholds to trigger alarms and initiate corrective actions.
- Segment the data using binary means to distinguish between normal behavior and pump failure. This will help identify key variables such as Pump Torque, Pump Efficiency, and Pump Speed, which have the largest variances, indicating that they are important variables to monitor closely.
- Use statistical models to predict and prevent pump failures. The analysis of the model fit revealed that a linear model is a good fit for the data, with Horse Power,

Pump Speed, and Pump Torque having the largest coefficients. Therefore, these variables should be used as key indicators for proactive pump failure detection.

- In summary, implementing a real-time monitoring system, focusing on improving pump efficiency, segmenting the data, and using statistical models can help improve the proactive identification of pump failures. These recommendations, when implemented effectively, can help prevent pump failures, increase equipment reliability, and improve operational efficiency.