OGE Energy Corp

Technical Presentation

Date: 12/29/20

Presenter: John Dawson

In recent years, OGE Energy Corp customer retention has dropped 5% dues to energy stability complaints. As solar energy demand increases, asset stability and profitability are the areas of greatest concern.

Problem 1 Low AC conversion conversion at plants	There have been inconsistencies in the Solar Energy System and with an increase in Solar Energy demand, these inconsistencies could impact the reliability of inverters at our two operational plants.
Problem 2 With overwhelming support of Solar Power recently, demand is outpacing supply.	The two new solar farms are sold out. With high solar energy demand and recent suboptimal performance in inverters, the reliability of power generation could negatively impact solar client growth in the coming years.
Problem 3 Solar Cell In-Efficiency	The stability of current Solar Power Plants is of great concern. We are looking to predict power generation to help with better grid management as well as identify equipment with suboptimal performance. This would keep plants at peak performance and power generation.

Financial pressures can be limited through the identification of suboptimal inverters, which will isolate maintenance needs and costs, reducing unnecessary operational expenditures. Our target is a 15% year-over-year increase in revenue with a 10% reduction in maintenance costs over the 2022-2023 calendar year.

Issues to Explore	Why do we do this?
Which inverters experience zero DC Power generation during daylight hours?	Solar energy outages could be caused by weather issues, OGE forced outage reason, or failure of inverters. We are looking at historical data to see if company assets (inverters) are the problem.
Can we identify features that correlate to inverter failures?	We analyze the data with an assumption that inverter failure can be prevented. If we cannot prove this, we conclude that failure cannot be understood with the available data.
If correlations do exist with respect to inverter failure, can we create an equation to represent this for a single Inverter?	If we are able to model failure for a single inverter, we could scale modeling for other plants and inverters.
If an equation for a single Inverter can be created, can we scale this out for multiple inverters?	If we are able to scale modeling across inverters and plants, we may be able to mitigate failure.

Strategy: Use of descriptive and inferential statistical analysis to identify key variables for Solar Plant DC Power Generation and Inverter failure along with predictive signals for Inverter failure.

Descriptive Statistical Analysis

Box Plot

Outliers and Variance of Inverter readings from their median. Among all features, both plants had 3 with higher outliers and skewed distribution.

Key Variables:

1. DC Power (Both Plants)
2. AC Power (Both Plants)
3. Ambient Temperature (Plant 2)
4. Module Temperature (Plant 1)

Inferential Statistical Analysis

Correlation Analysis

Key Variables:

1. Irradiation
2. Ambient Temperature
3. Module Temperature
4. Daytime Hours

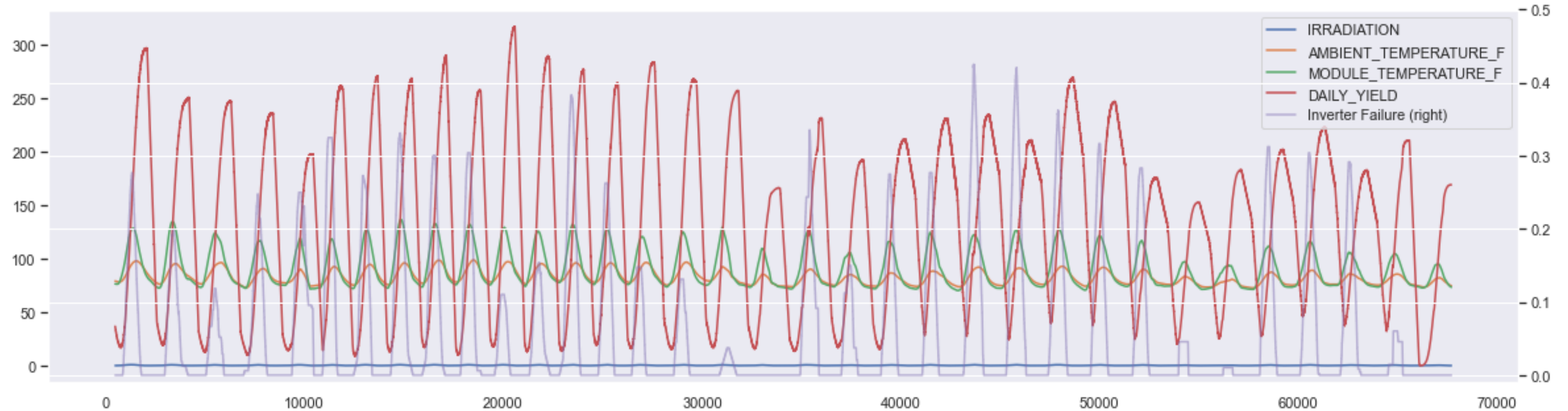
Regression Analysis

Identified 4 major power generation signals.

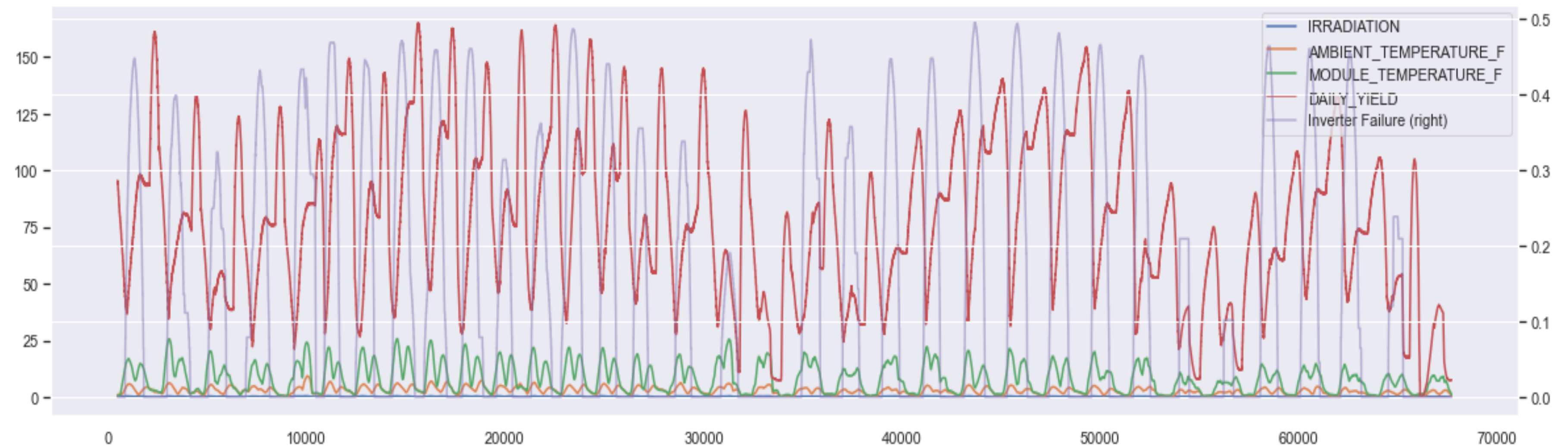
Proactive identification of abnormal Inverter behavior was not obtained due to limited timeframe of data as well as lack of reliable weather and climate data.

Descriptive Analysis has enabled us to identify abnormalities, showing clear changes in both Rolling Standard Deviation and Rolling Mean Datasets when observed over the inverter failure periods. Module and Ambient temperatures peak while Standard Deviation decreases during failure periods.

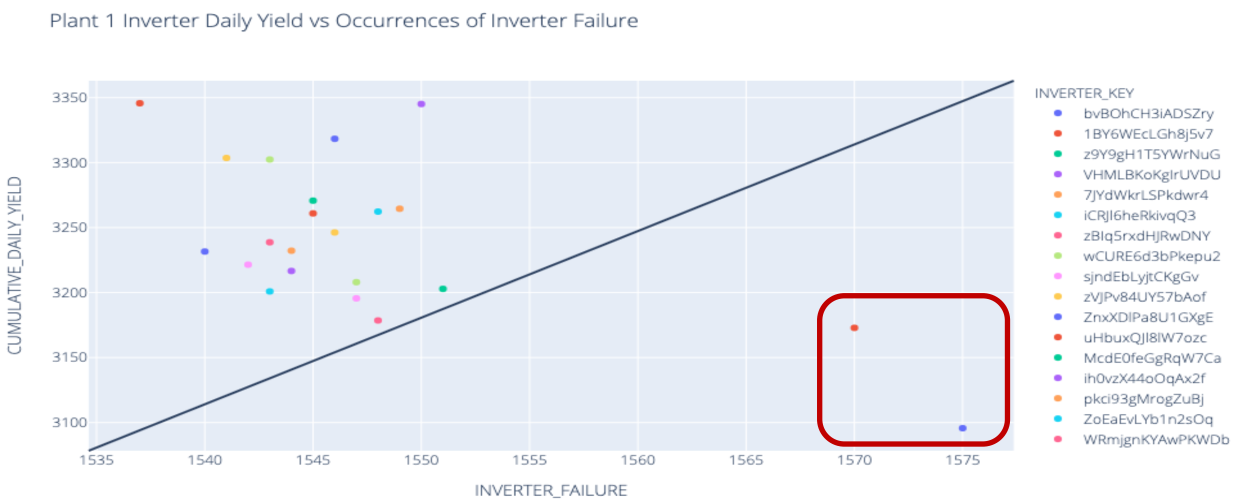
Rolling Mean Plot

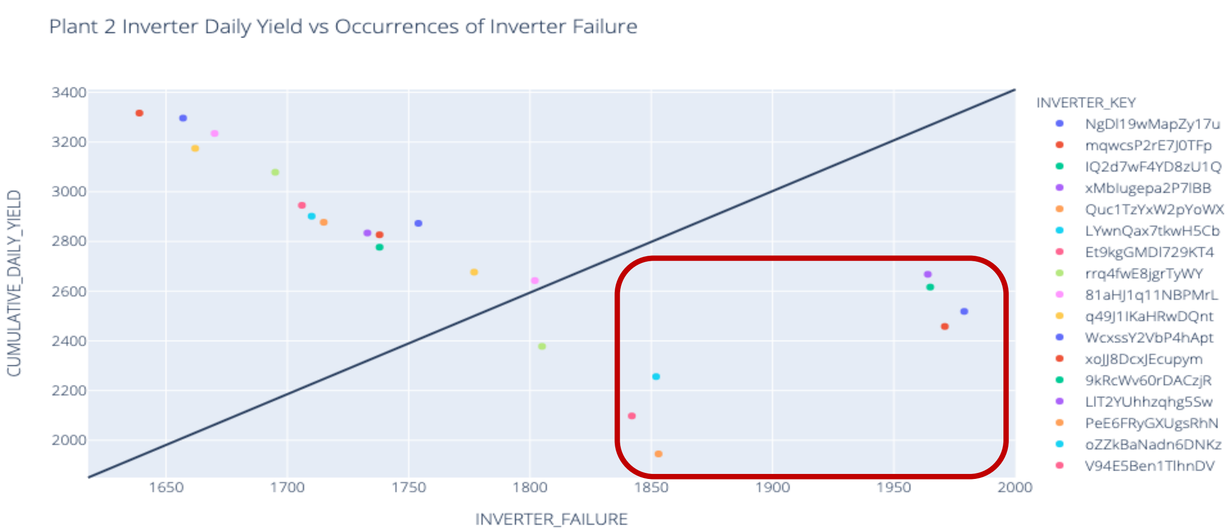


Rolling Standard Deviation Plot



Further analysis of the dataset identified a 54% and 52.2% decrease in outages from May to June for plants 1 and 2 respectively. Suboptimal inverters for both plants remain an issue, even with significant decreases month over month.





Key Insights

Identifying inverters with lower performance is of the greatest importance for OGE.

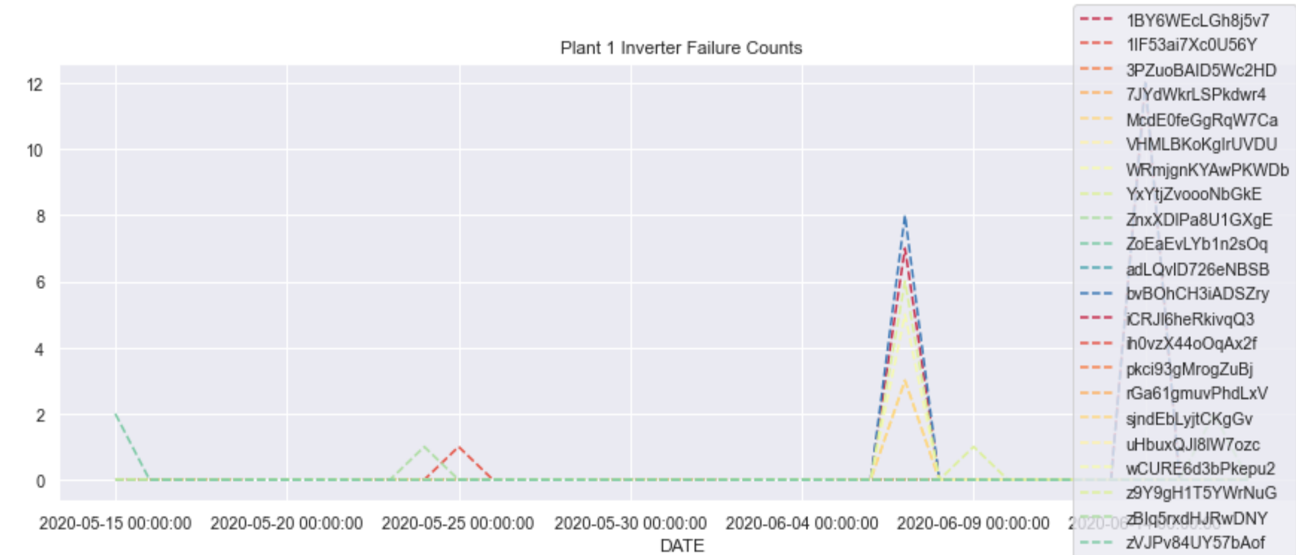
Inverters in the lower right portion of each chart have higher instances with zero power and lower cumulative daily yield. These inverters are the least efficient, reliable, and profitable for OGE.

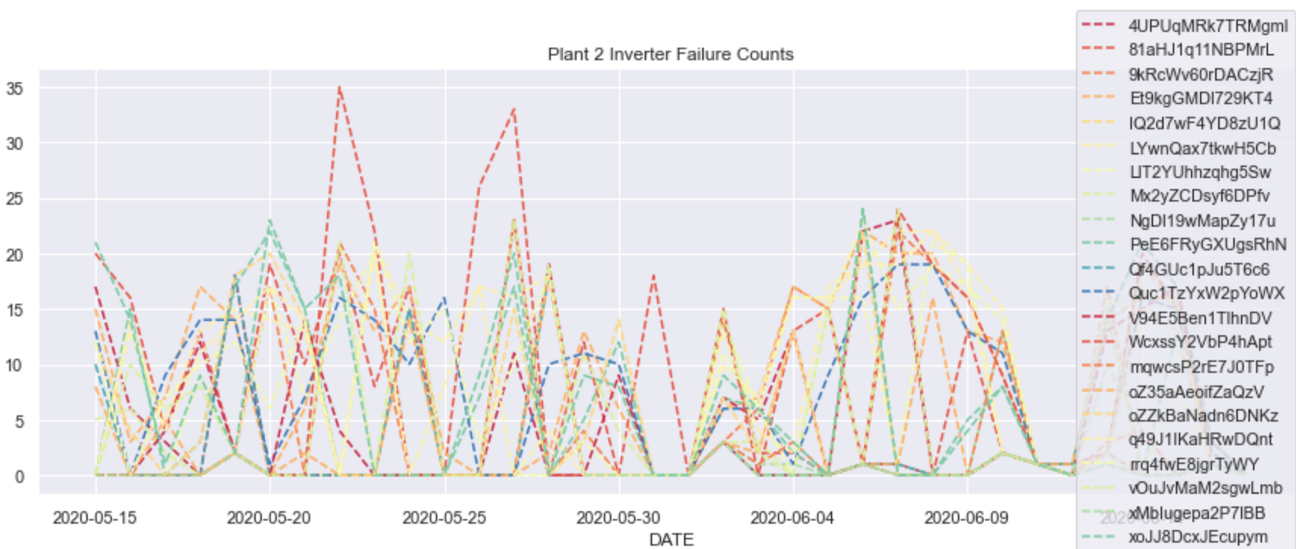
Suboptimal inverters decrease DC to AC conversion rates, negatively impacting plant efficiency and reliability.

Energy reliability has the highest influence on customer loyalty and long term profitability of the solar program.

Source: Solar Power Generation Data (www.kaggle.com)

From May to June, Plant 1 saw significant decreases in Failure counts across all inverters. However, plant 2 had increases in failure events for 3 inverters.





Key Insights

3 inverters with lower performance need to be looked at further to access the next steps.

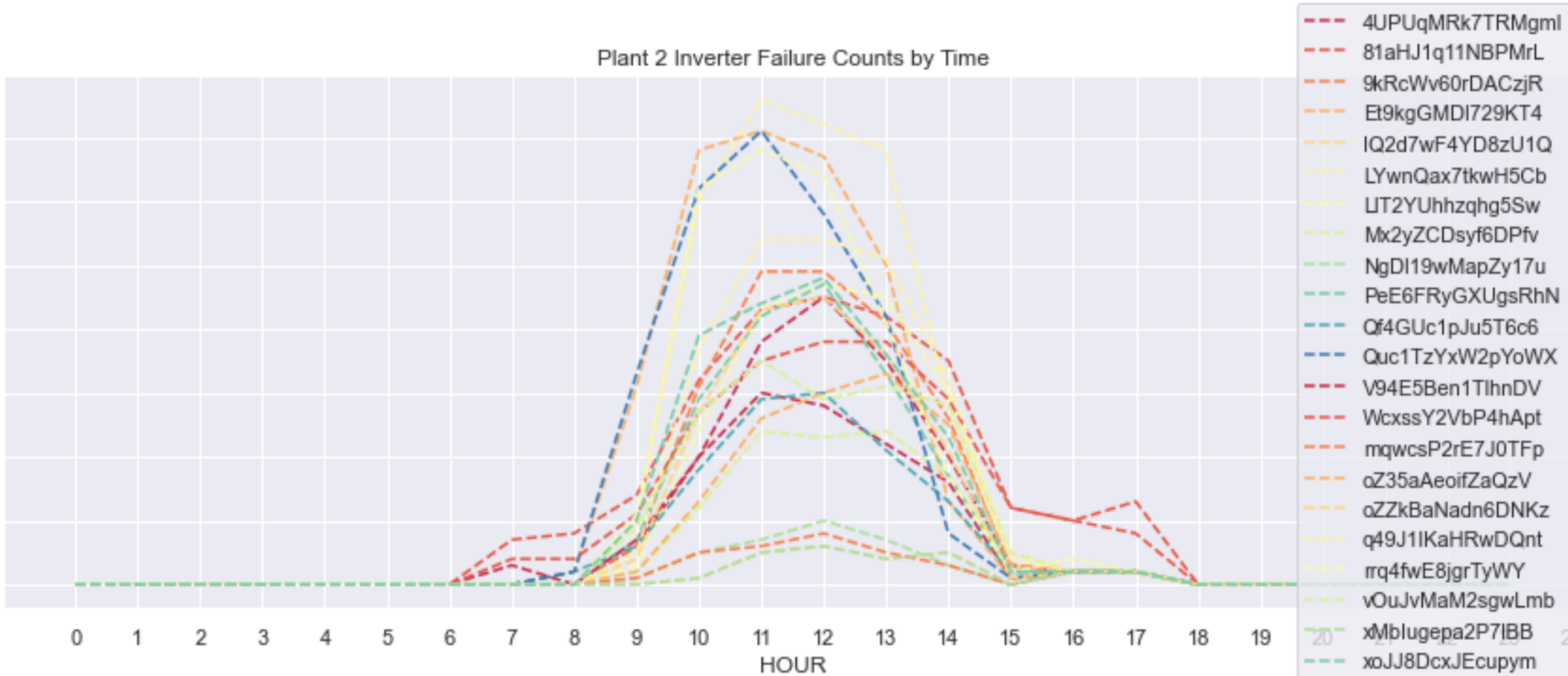
9kRcWv60rDACzjR had a 33% increase in failure events.

LIT2YUhhzqhg5Sw had a 51% increase in failure events.

V94E5Ben1TlhnDV
LIT2YUhhzqhg5Sw had a 67.5% increase in failure events.

Source: Solar Power Generation Data (www.kaggle.com)

Inverters with suboptimal performance have been identified at both plants. However, focus of our analysis is on Plant 2 which has more frequent inverter failures. With further analysis, proactive identification of failing inverters could lead to mitigation and growth plans moving forward.



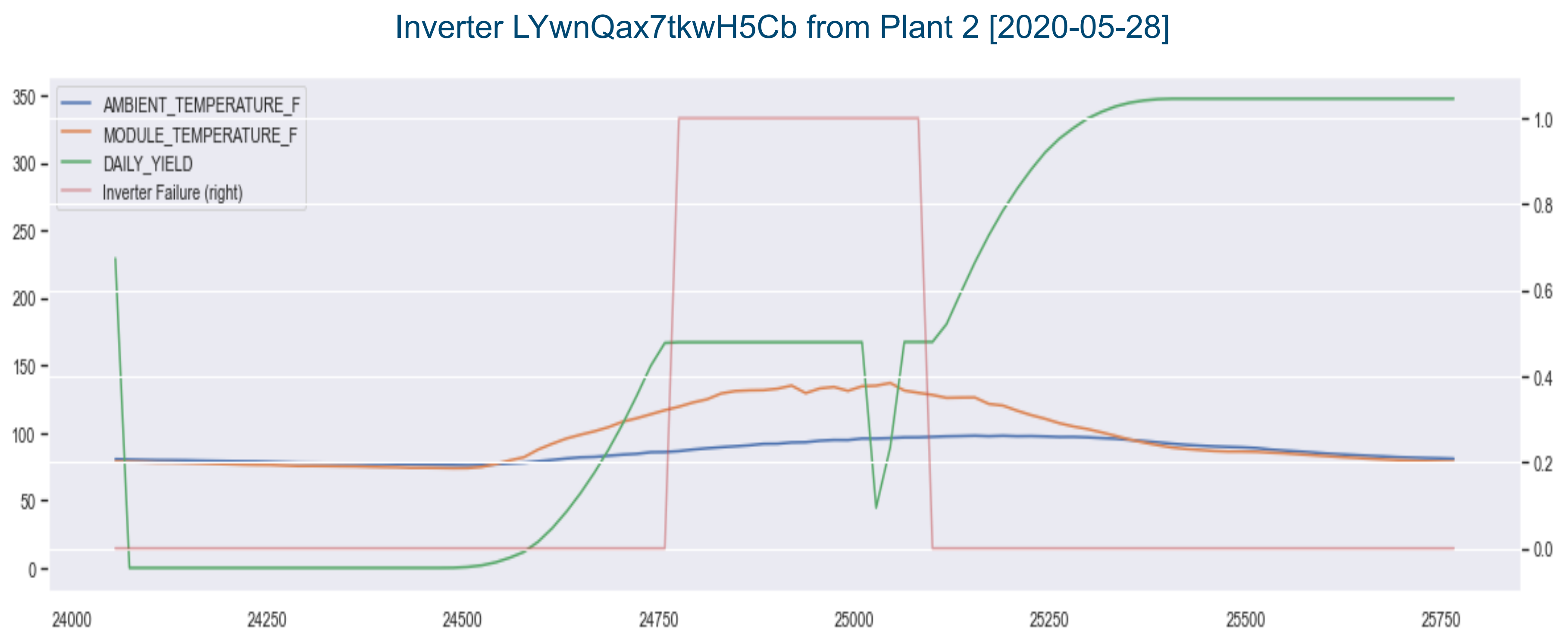
Key Insights

Identification of inverters with low performance is of the greatest importance for OGE.

Both plants have noted Inverter failure, which leads to unreliable and inefficient power generation, diminishing profitability for OGE. Midday has higher instances of failure across multiple inverters, which could be an anomaly. Identifying these errors in sensors is necessary to identify true failure events and mitigate energy losses.

Source: Solar Power Generation Data (www.kaggle.com)

Inverter LYwnQax7tkwH5Cb had 320 failure instances, the highest of all inverters from both plants. We see failure takes place when Daily Yield, Module Temperature, and Ambient Temperature are increasing, which makes sense due to higher failures during the middle of the day.



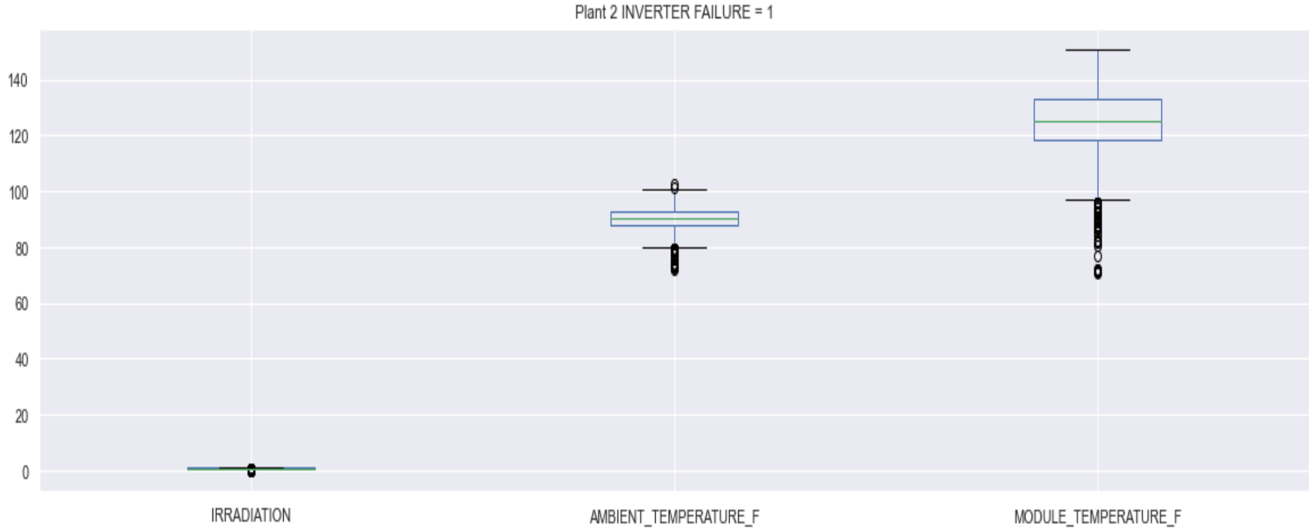
Key Insights

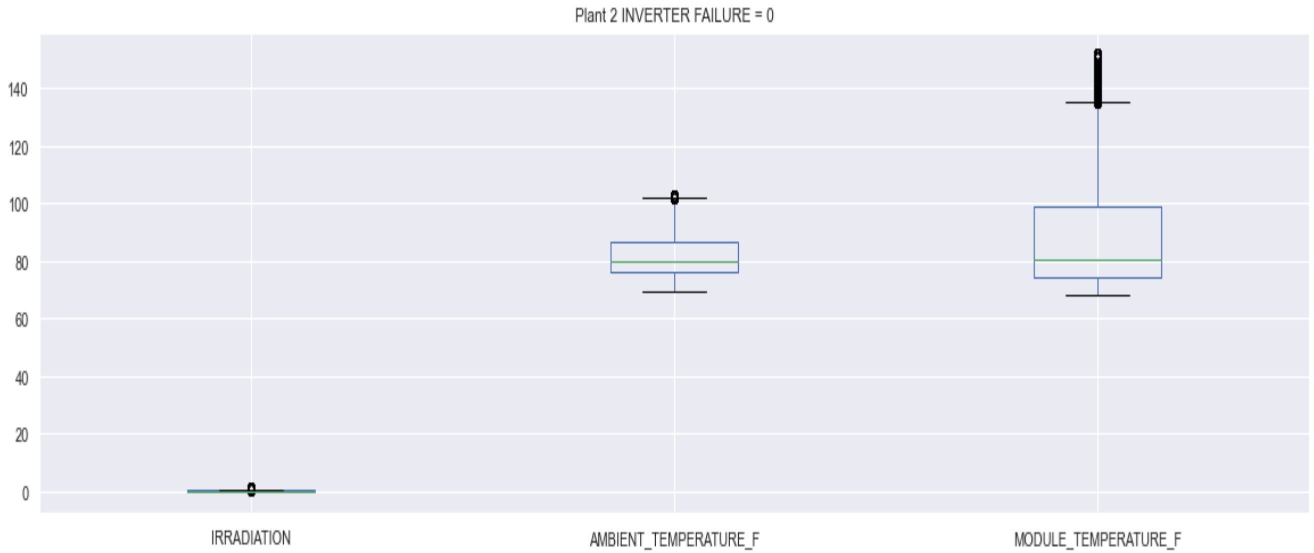
Module Temperature and Ambient Temperature appear to increase relative to the inverter failure event period. These variables could be key indicators for inverter failure.

There is a flattening of Daily Yield just prior inverter failure and a decrease before quickly increasing toward the end of the inverter failure event period, and then increases further after the failure.

Source: Solar Power Generation Data (www.kaggle.com)

Further segmentation of the data via binary means (Inverter Failure = 0 or 1) illustrated through box plots, show a clear signature difference between that of normal behaviour and that of Failure with Module Temperature showing the largest variance.

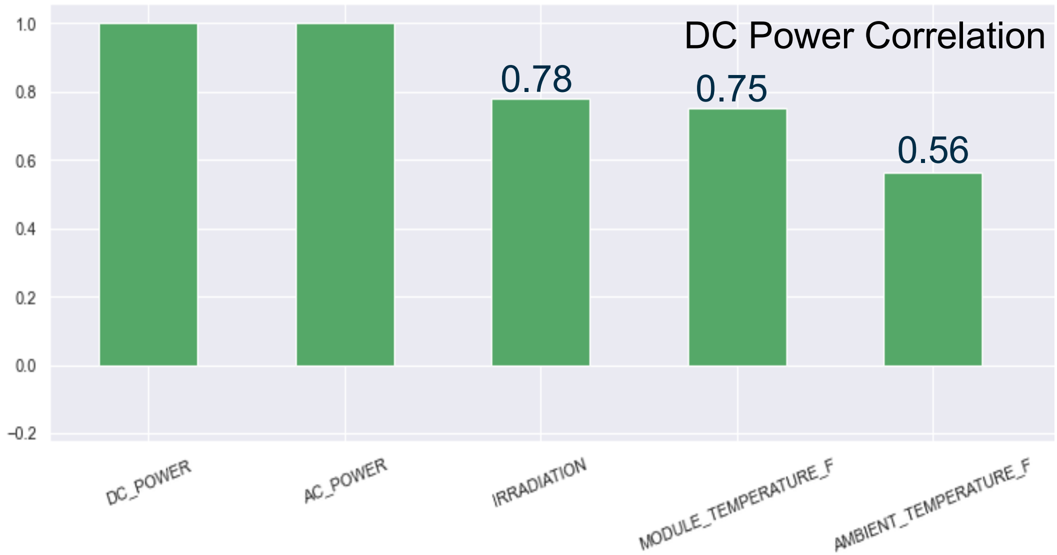


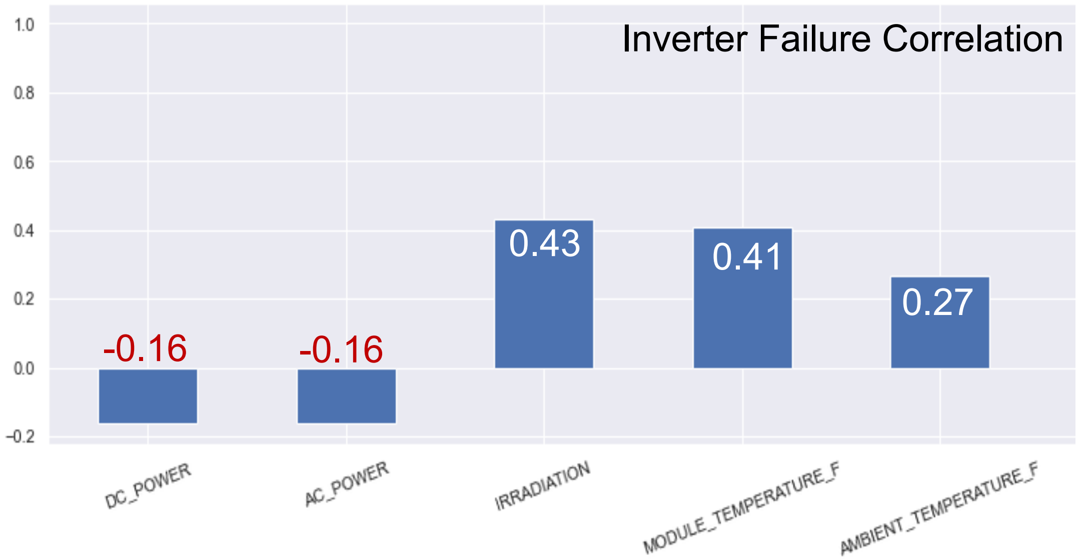


Key Insights

- IQR for Module Temperature during failure spans from 120 to 130 while it is much lower during non-failure periods (75-100)
- Module Temperature Mean > 40 degrees above that of non inverter failure periods.
- Ambient Temperature has remained mostly consistent with overall and non-failure data, though a diminishing of IQR and higher mean have been noted.

Correlation analyses indicate a positive correlation for DC Power Generation and Inverter Failure with Irradiation, Module Temperature, Ambient Temperature.

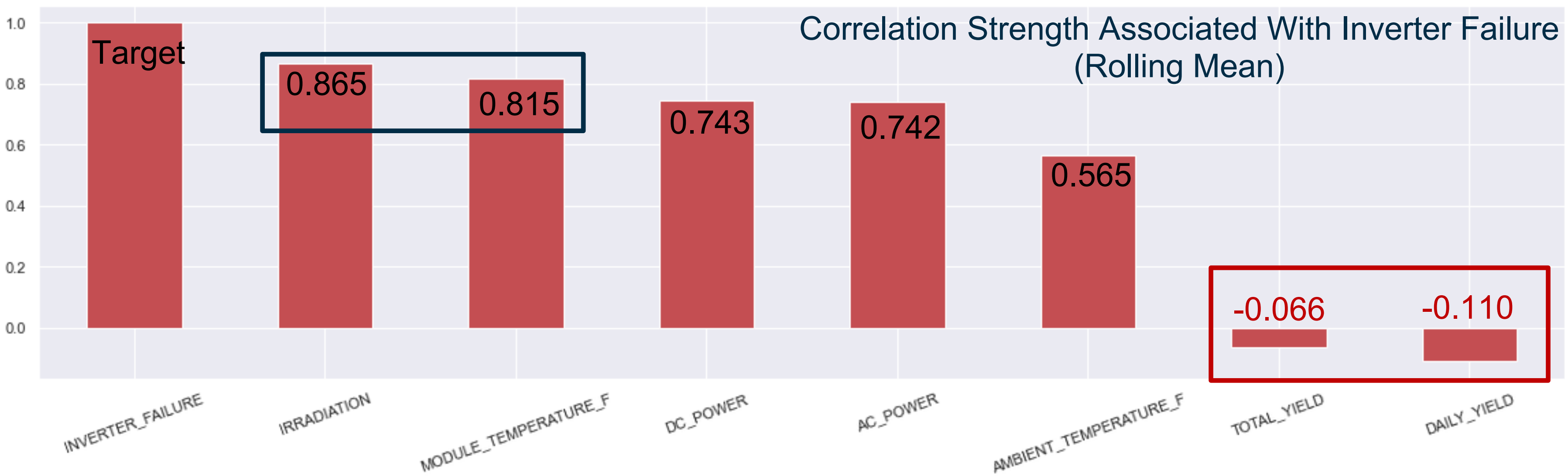


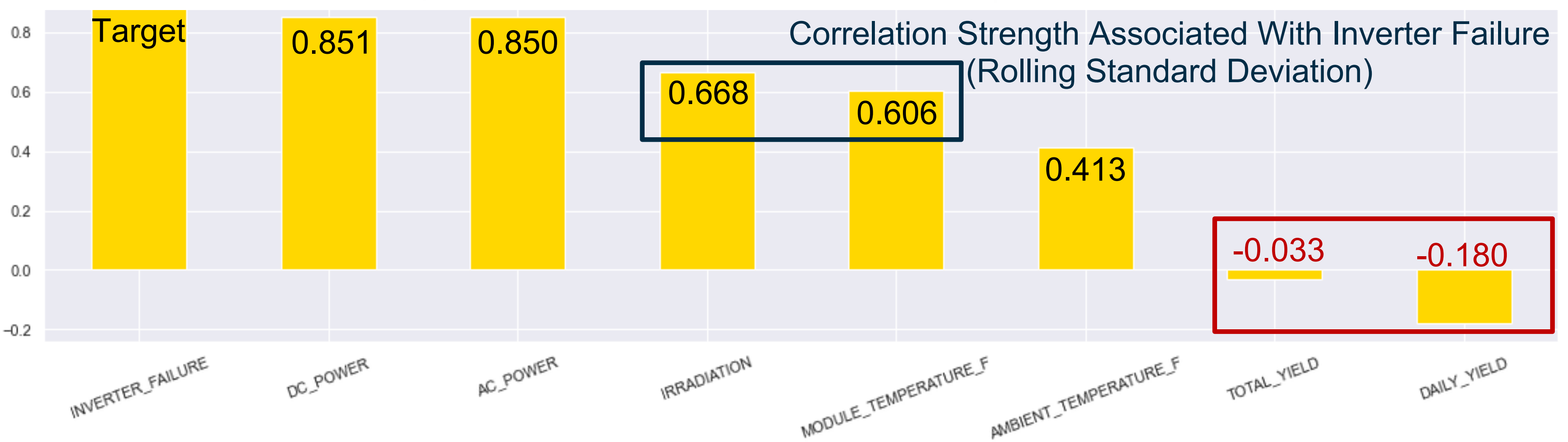


Key Insights

- Positive correlation between DC Power and both Irradiation and Module Temperature is strong.
- Positive correlation between DC Power and Ambient Temperature is moderate.
- Positive correlation between Inverter Failure and both Irradiation and Module Temperature is moderate.
- Positive correlation between DC Power and Ambient Temperature is weak.
- The dataset is limited to 1 month of data and does not include detailed weather data. Further analysis could proactive identification of failing inverters could lead to mitigation and growth plans moving forward.

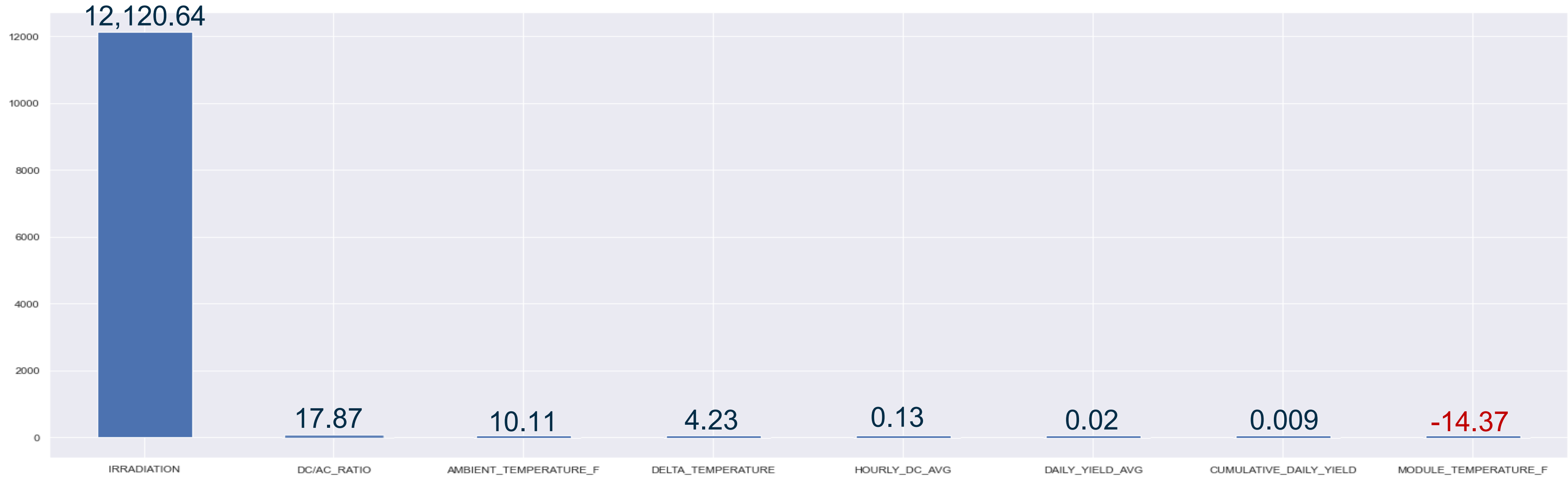
Correlation analyses across Mean and Standard Deviation datasets yield interesting insights with Daily and Total Yield negatively correlated with Inverter Failure, while Irradiation and Module Temperature shows a subsequently strong positive correlation. Irradiation and Module Temperature might be good predictors of the outcome variable.





Analysis of the statistical significance of variables contributing toward DC Power for Plant 1 reveals that with an R Squared of 0.982, a linear model is a good fit for Plant 1 and contributes key information for understanding DC Power.

Plant 1 Ranking of Variable Significance for DC Power (Linear Regression Coefficient)

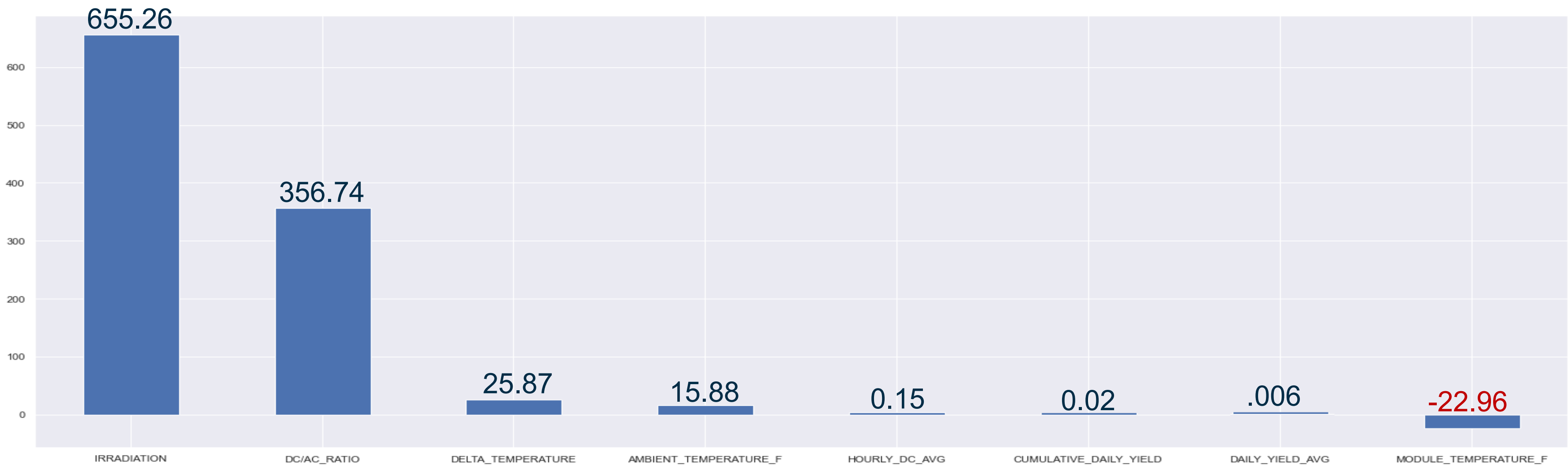


Key Insights

- All variables are statistically significant and have p values < .05
- For every unit increase in Irradiation (W/m²), DC Power is estimated to increases by 12,121 kW.
- For every unit increase in DC/AC Ratio (kW), DC Power is estimated to increases by 17.87 kW.
- For every unit increase in Ambient Temperature (°F), DC Power is estimated to increases by 10.11 kW.
- For every unit increase in Delta Temperature (°F), DC Power is estimated to increases by 4.23 kW.
- For every unit increase in Hourly DC Average (kW), DC Power is estimated to increases by 0.13 kW.
- For every unit increase in Daily Yield Average (kW), DC Power is estimated to increases by 0.02 kW.
- For every unit increase in Cumulative Daily Yield (kW), DC Power is estimated to increases by 0.009 kW.
- For every unit increase in Module Temperature (°F), DC Power is estimated to decrease by 14.37 kW.

Analysis of the statistical significance of variables contributing toward DC Power for Plant 2 reveals that with an R Squared of 0.771, a linear model is a good fit for Plant 2. Linear regression does contribute key information for understanding DC Power generation at Plant 2.

Plant 2 Ranking of Variable Significance for DC Power (Linear Regression Coefficient)

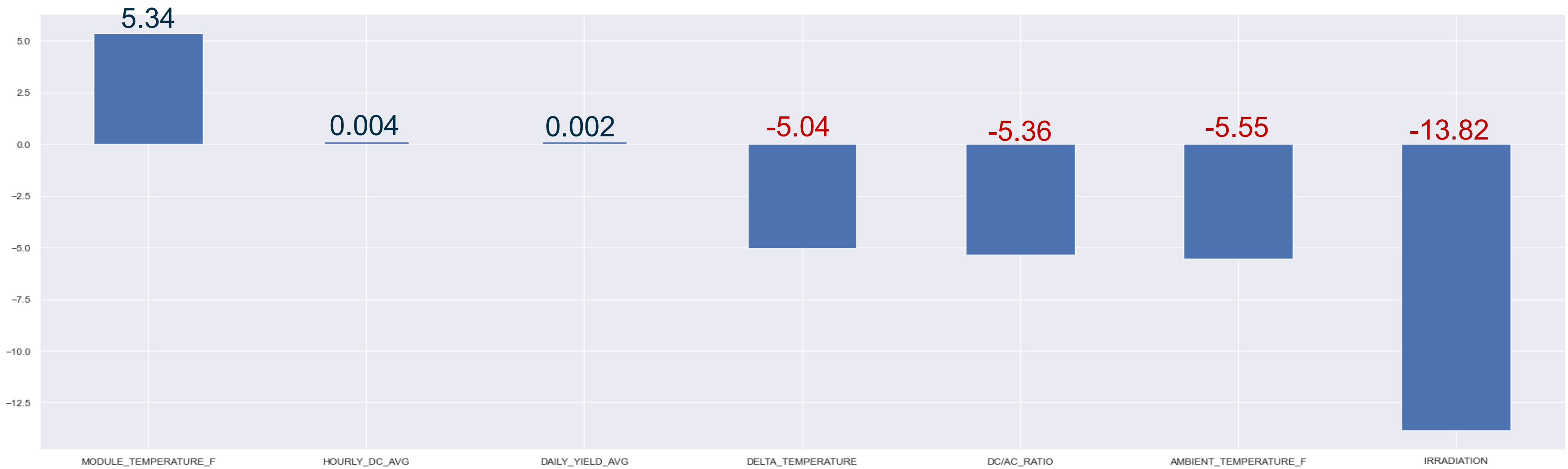


Key Insights

- All variables are statistically significant and have p values < .05
- For every unit increase in Irradiation (W/m²), DC Power is estimated to increases by 655.26 kW.
- For every unit increase in DC/AC Ratio (kW), DC Power is estimated to increases by 356.74 kW.
- For every unit increase in Delta Temperature (°F), DC Power is estimated to increases by 25.87 kW.
- For every unit increase in Ambient Temperature (°F), DC Power is estimated to increases by 15.88 kW.
- For every unit increase in Hourly DC Average (kW), DC Power is estimated to increases by 0.15 kW.
- For every unit increase in Cumulative Daily Yield (kW), DC Power is estimated to increases by 0.02 kW.
- For every unit increase in Daily Yield Average (kW), DC Power is estimated to increases by 0.006 kW.
- For every unit increase in Module Temperature (°F), DC Power is estimated to decrease by 22.96 kW.

Analysis of the statistical significance of variables contributing toward Inverter Failure for Plant 1 returned a Pseudo R Squared of 1.000, which would indicate a good fit for the dataset. However, with p-values greater than 0.05, none of our variables are statistically significant.

Plant 1 Ranking of Variable Significance for Inverter Failure (Logistic Regression Coefficient)

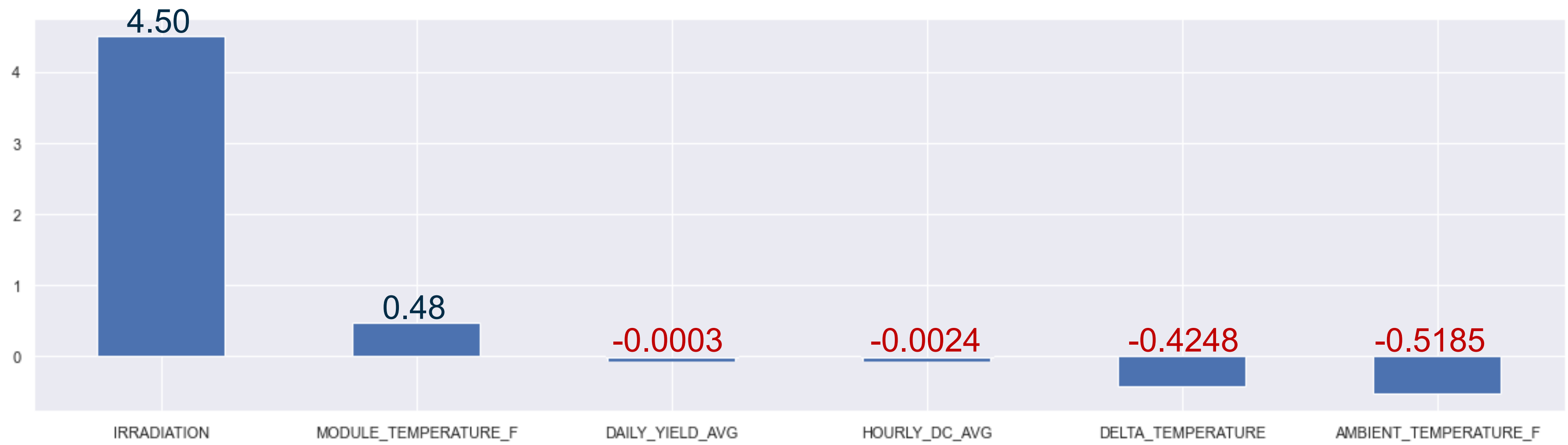


Key Insights

- Logistic Regression returned a Pseudo R Squared of 1.000. High explanatory power indicates a perfect fit or 100% of variation in Inverter Failure is explained by independent variables.
- Independent variable p-values > 0.05. A high conditional probability indicates there is no statistically significant dependence of Inverter Failure with respect to model variables.
- Logistic Regression indicates relationship exists between Inverter Failure and independent variables. Our model explains variation within the data, though it is not statistically significant.

Analysis of the statistical significance of variables contributing toward Inverter Failure for Plant 2 returned a Pseudo R Squared of 0.3941. For Plant 2, 40% of the observed variation can be explained by the logistic model inputs. Irradiation has an exponential increase of 4.5 on log likelihood of failure for every W/m² observed.

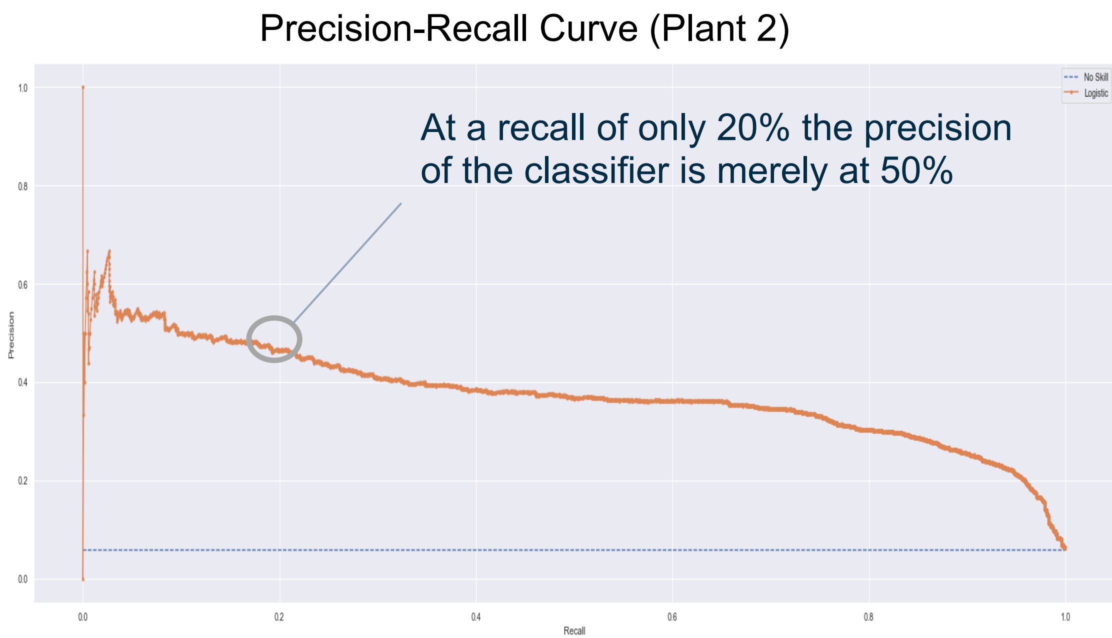
Plant 2 Ranking of Variable Significance for Inverter Failure (Logistic Regression Coefficient)

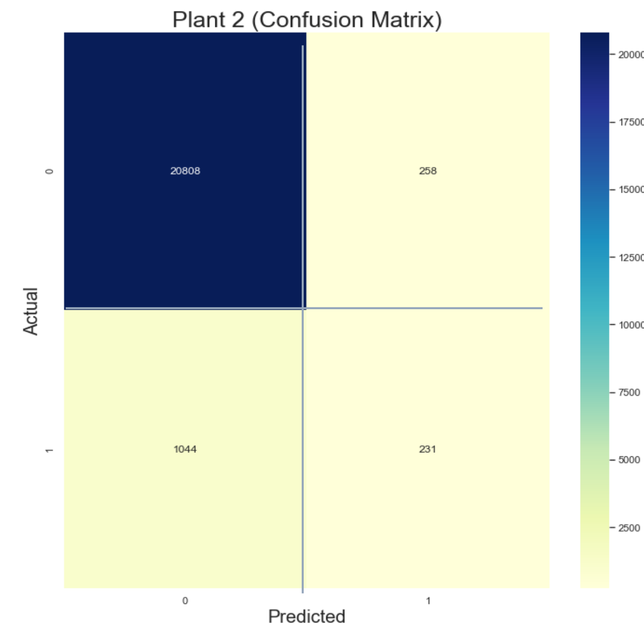


Key Insights

- All variables are statistically significant and have p values < .05
- For every unit increase in Irradiation (W/m²), the log odds of Inverter Failure increases by 8933.65%.
- For every unit increase in Module Temperature (°F), the log odds of Inverter Failure increases by 160.58%.
- For every unit increase in Daily Yield Average (kW), the log odds of Inverter Failure decrease by 0.03%.
- For every unit increase in Hourly DC Average (kW), the log odds of Inverter Failure decrease by 0.25%.
- For every unit increase in Delta Temperature (°F), the log odds of Inverter Failure decrease by 34.32%.
- For every unit increase in Ambient Temperature (°F), the log odds of Inverter Failure decrease by 40.17%.

Under Model Development, Prediction, and Cross-validation of Plant 2 data we find that our model accounts for 40% of the variance (Pseudo R-square of 0.3952 and Mean 5-Fold R Squared of 0.3964). We did not have a meaningful positive rate for Inverter Failure Prediction with our model.





Key Insights

- Precision for 0 (Non-Failure) is 0.95. Precision for 1 (Failure) is 0.47. Precision is the proportion of True Positives from all positives (True and False). We want precision greater than 0.47 for Inverter Failure.
- Recall is 0.99 for 0 (Non-Failure) and 0.18 for 1 (Failure). Recall is the proportion of True Positives that were identified correctly. At 0.18 our model is far from correctly identifying Inverter Failure.
- We used the Precision-Recall Curve which plots precision against recall. We can see that precision and recall are both lower than ideal. Our model returned a PR AUC of 0.376 and an F1-score of 0.262, which are both lower than we want.

Next Steps

Recommendations

Maintenance to 9kRcWv60rDACzjR, LIT2YUhhzqhg5Sw, and V94E5Ben1TlhnDV inverters at Plant 2 since they have seen increases in failure events from last year.

Inspection of the LYwnQax7tkwH5Cb inverter at Plant 2 since it has significantly more failure events than inverters across both operational plants.

Gather more data for further analysis. We specifically recommend gathering weather-specific datapoints for a more in-depth analysis of inverter failure trends. We also recommend gathering data beyond the one-month period we currently have since this does limit scope.

Increase in Irradiation and Module Temperature increase the log odds of Inverter Failure by 8933.65% and 160.58% respectively. *Analysis of the Temperature Coefficient of Pmax for OGE panels is recommended since increases in heat above an optimal Module Temperature can reduce output efficiency and lead to inverter failure. With such analysis, we should be able to predict energy loss for every °F above optimal Module Temperature for OGE Solar Panels.*

Future Scope of Work

We have been working on an ARIMA Model Time Series Forecasting on datasets for both plants. We have successfully forecasted multiple days of energy generation. This is helpful for power grid management. As more time-series and weather data is available, we will continue to iterate our modeling for the proactive identification of inverter failure and higher accuracy.