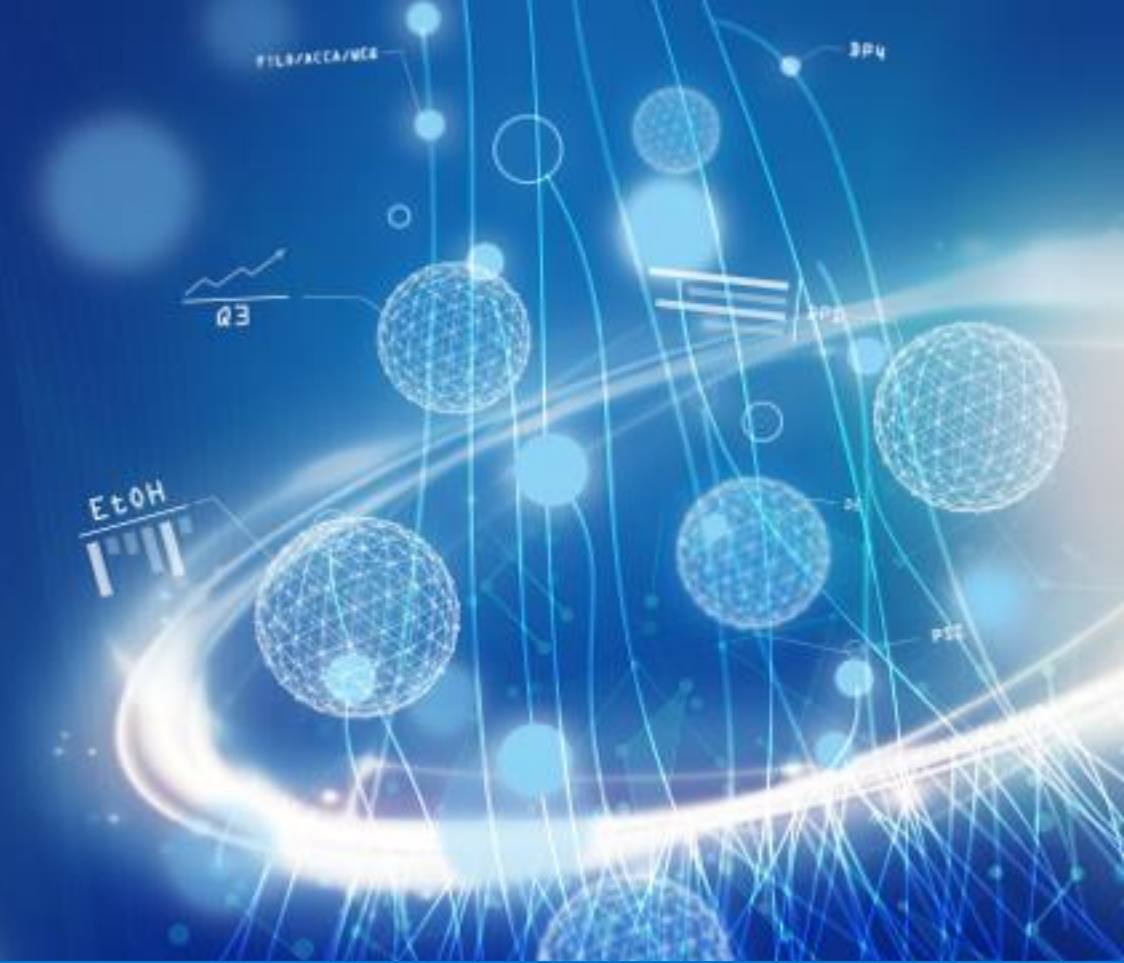


XCELIS® AI

FEATURE ENGINEERING

And Other Methods to Prep Lab Data for Integration with Machine Learning Models

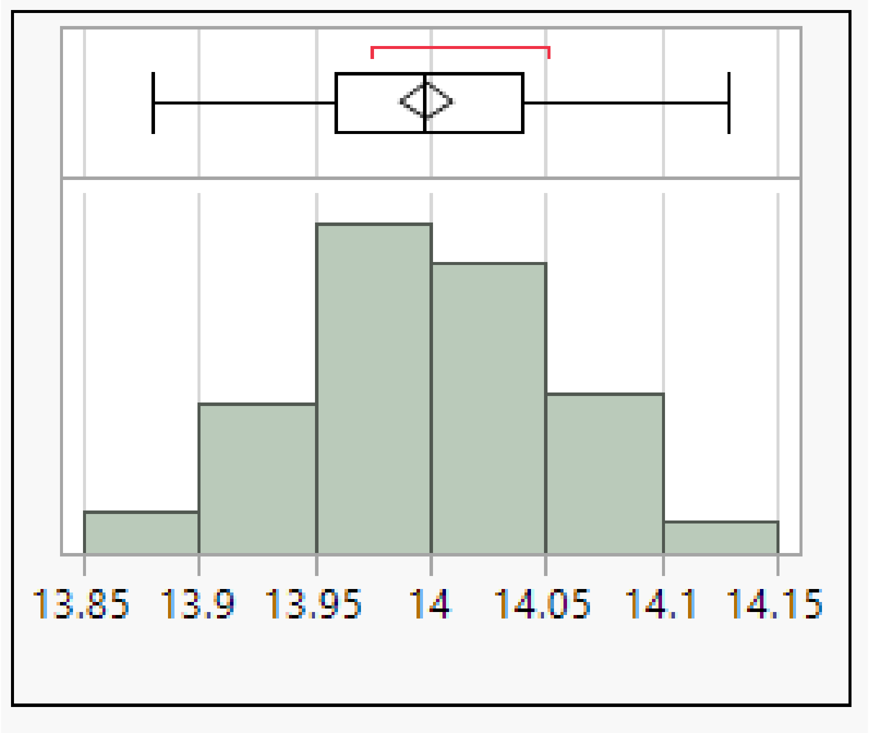
XCELIS® AI

Use **digital tools** to change how the ethanol industry collects, maintains, and analyzes data to gain more **accurate insights** in **less time** while utilizing **fewer resources**.

Advanced Data Solutions	Virtual Plant Technology
Maximize your now	*Predict your future*
Flexible data processing, customizable reporting, and advanced analytics	Predictive models driven by science and engineering fundamentals
Aimed to help you make the most of the data generated across your plant and reduce the time to actionable insight	Aimed to help you look to the future and make informed decisions about products and process changes
Data Processing and Advanced Analytics	***Dynamic Fermentation Modeling*** ***Holistic Process Modeling***

WHY DO WE NEED STATISTICS

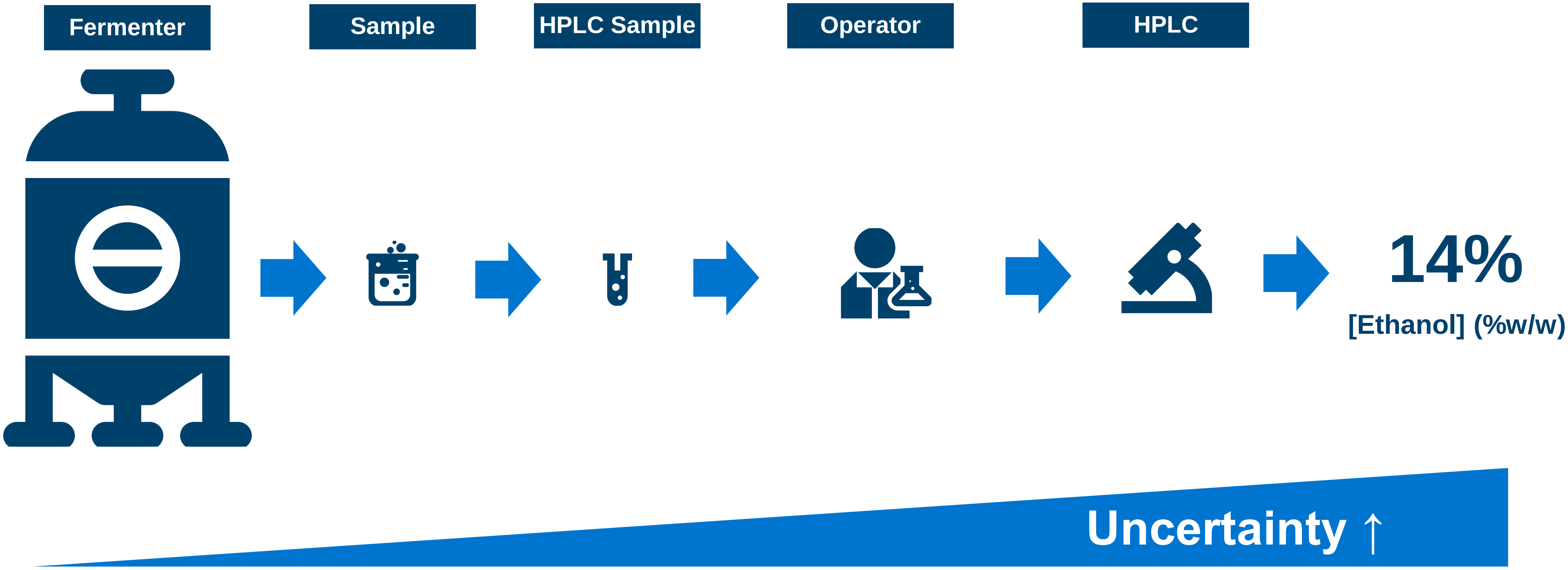
Dealing with Uncertainty



Public

WHY DO WE NEED STATISTICS

Dealing with Uncertainty from Samples



HPLC Data, Lab Data, Sieve Analysis, Yeast Health, Solids, etc. (anything where you are analyzing samples)

OBJECTIVES

What is the goal of the talk today?

What can *you* do for you?

What can *I* do for you?

What can *AI* do for you?

Images generated by Copilot

Public

OBJECTIVES

Data Upload	Performance Metric Evaluation	Trained/Validated ML Models
Data Cleaning and Outlier Analysis	Batch Linking	Model Evaluation
Process Screening	Fit Curve Analysis	Probabilistic Action Items
Feature Engineering	Direct/Indirect Feature Analysis	Local Data Analysis
Least Square Linear Models (Multiple Features)	Variability Analysis	Dynamic Time Warping
	Predictor Screening, Partition Modeling	

Images generated by Copilot

Public

DATA UPLOAD

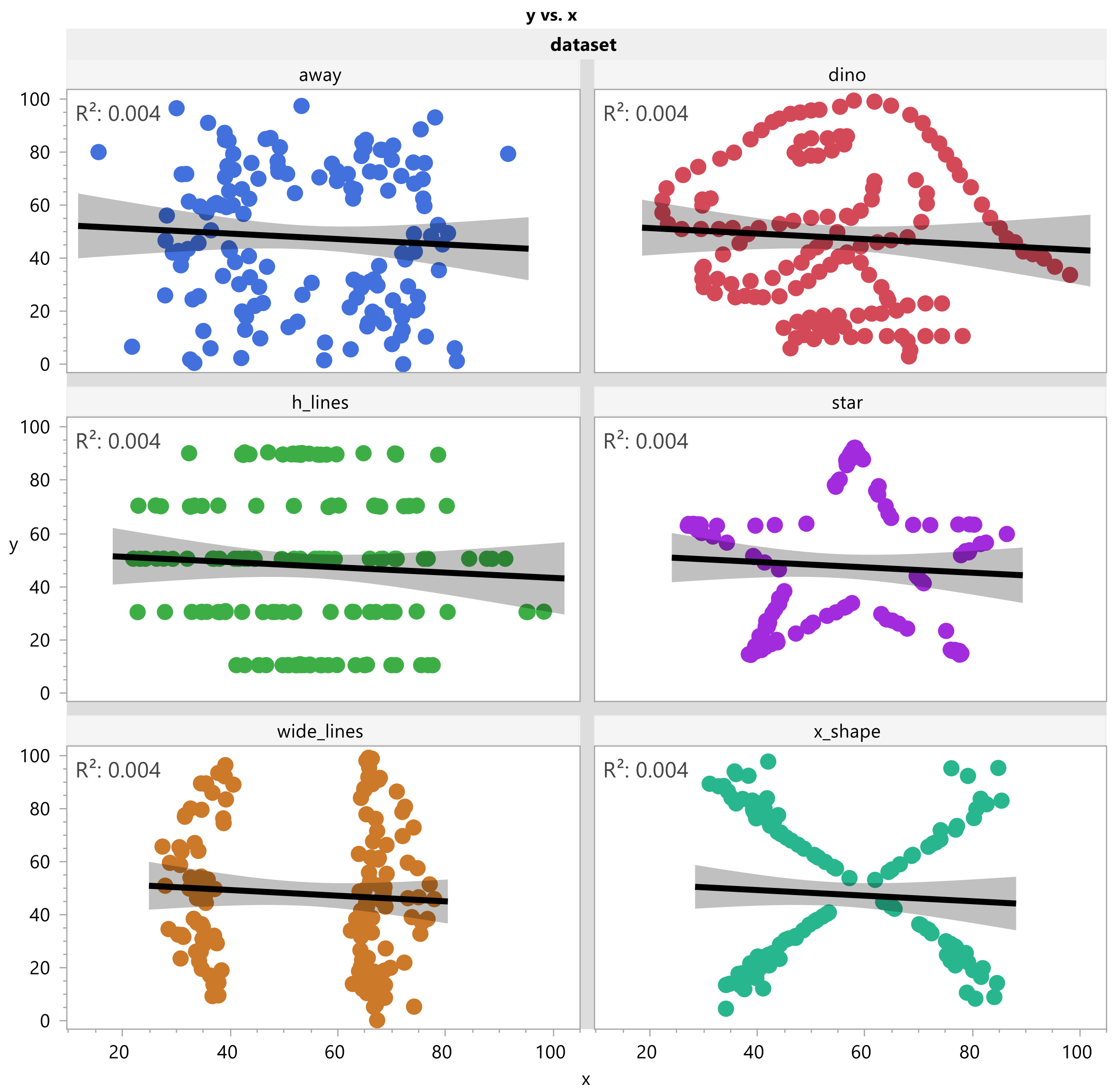
DATA UPLOAD

- **Utilizing coding to improve the process – Routine reports can be uploaded faster by coding (scripting, macros) to transform raw data into a workable data table.**
- **Consistency is key – Automated data is guaranteed to be consistent and not contain different features, columns, formats, etc.**
- **Faster Turnaround Time - Minimizes time spent uploading and cleaning, meaning more time spent analyzing data and providing insights.**

Public

DATA CLEANING
OUTLIER ANALYSIS

ALWAYS GRAPH YOUR DATA

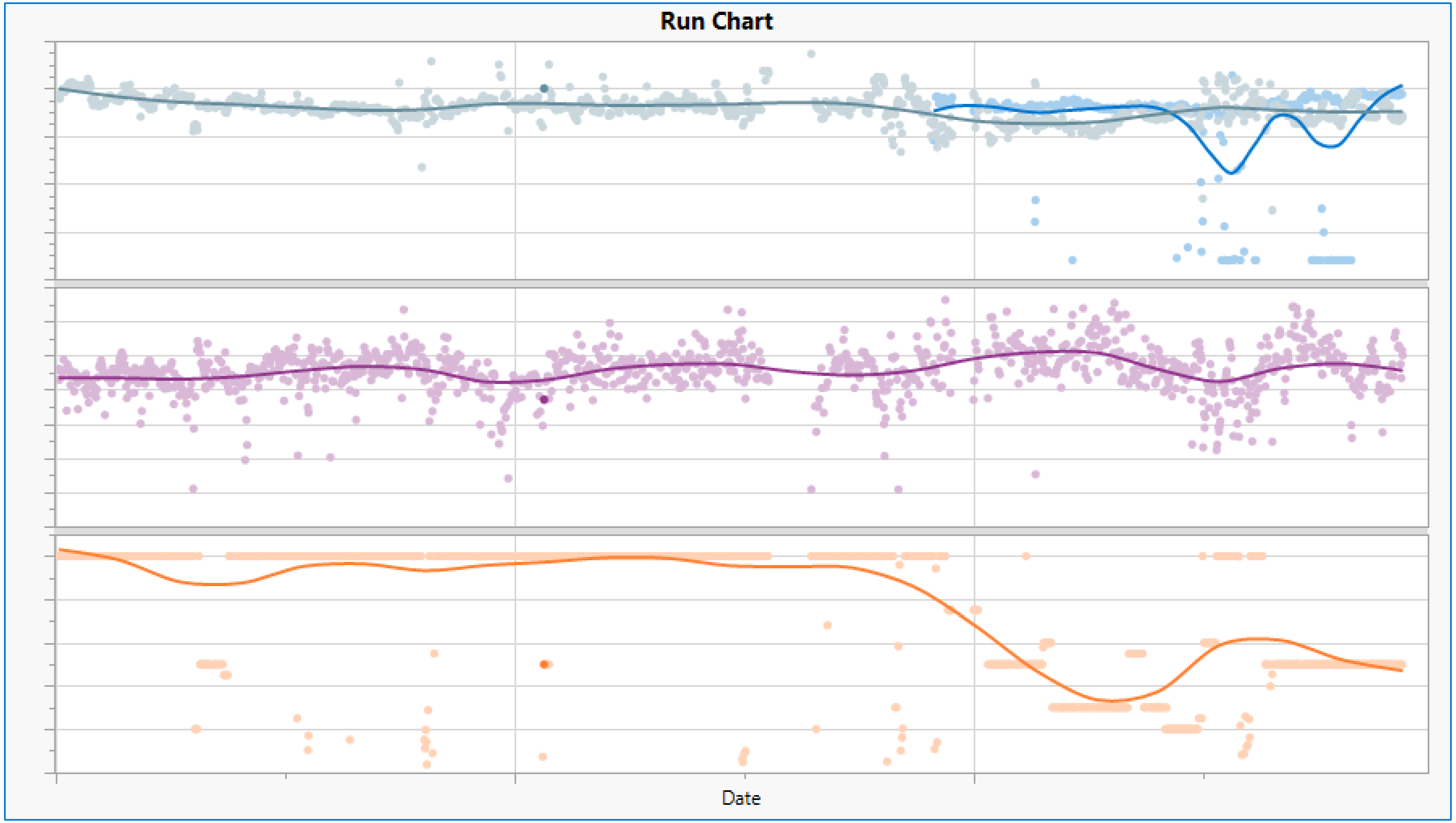


These graphs all show the exact same output in terms of R-squared. Are they the same?

Always graph your data to validate assumptions and make sure you are not being misled

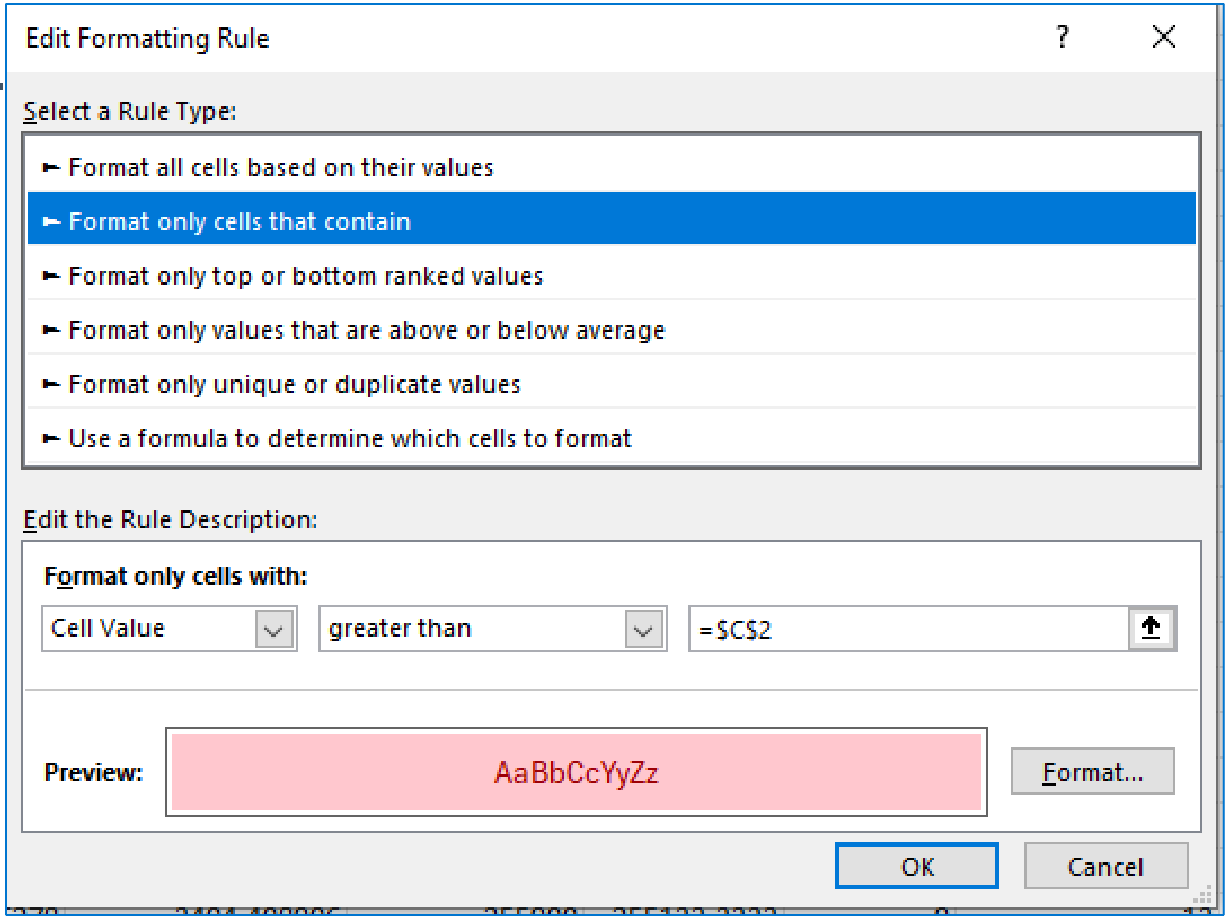
ALWAYS GRAPH YOUR DATA

Run charts are important to see if changes occur in a shift, or randomly.



DATA CLEANING

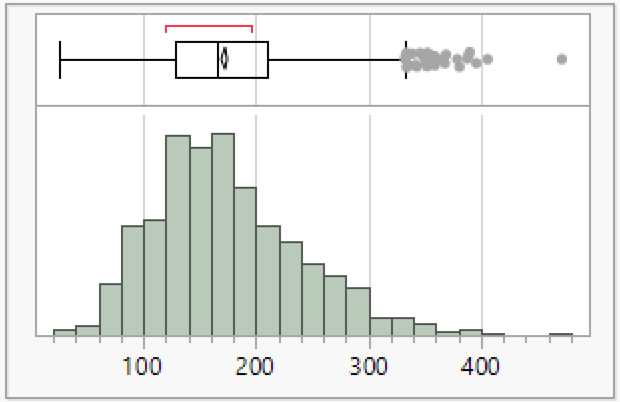
- **Model Outputs are only as good as model inputs.**
- **Outliers exist. Understanding what to do with them is where you can transform raw data into insights.**
- **Don’t blindly discard the bad data. Ensure that data is reviewed and corrected.**
- **Site-Specific Range Checks**
  - Ensure that data entered will be within a valid range
  - Works best for universal features (i.e. 1.0<pH<14.0)
  - Some data historians have the ability to set conditions when extracting data. [Example Later]



DATA CLEANING

Example – Outlier Calculation

Cell Count 12Hr



Quantiles

100.0%	maximum	472
99.5%		360.02
97.5%		313.275
90.0%		263
75.0%	quartile	209.75
50.0%	median	165
25.0%	quartile	128
10.0%		94.9
2.5%		72
0.5%		53
0.0%	minimum	25

Summary Statistics

Mean	172.28279
Std Dev	63.17966
Std Err Mean	1.3038518
Upper 95% Mean	174.83962
Lower 95% Mean	169.72597
N	2,348
N Missing	656

Q = 3 (This can be any number, it just multiplies the outlier distance)

$\text{Quantile}_{\text{Low}}$ = 0.25 (25% Quantile) = 128

$\text{Quantile}_{\text{High}}$ = 0.75 (75% Quantile) = 210

Outlier Distance = Q*(Value(Quantile$_{(1-x)}$) – Value(Quantile$_{(x)}$))

Outlier Distance = 3*(210-128) = 246

Low Outlier ≤ Value(Quantile$_{(x)}$) - Outlier Distance
Low Outlier ≤ 128 – 246
Low Outlier ≤ -118 (Here, the obvious case would be to set a minimum outlier value of **≤ 0**)

High Outlier ≥ Value(Quantile$_{(1-x)}$) + Outlier Distance
High Outlier ≥ 210 + 246
High Outlier ≥ **456** (This is a reasonable value for a high-end outlier)

DATA CLEANING

Setting up Filters in Pi Datalink

	A	B
1	EPN	401-T1ENZ-VL
2	Average	0.037422151
3	Std Dev	0.018518085
4	High Limit	1.00
5	Average + 3 Std Dev	0.09
6	Average - 3 Std Dev	-0.02
7	Low Limit	0.00
8		
9	High Filter	0.09
10	Low Filter	0.00
11	Filter Expression	'401-T1ENZ-VL'

```
fx  ="'"&B1&"' <"&B9&" AND '"&B1&"' >"&B10
```

Show steps of how to get to the Filter Expression

1. **For each tag (EPN), set up the average, Std Dev. Pi Calc, cell B2, B3**
2. **Manually set high and low limits in cells B4 and B7.**
3. **Calculate average + 3 Std Dev. Excel formula in Cell B5, B6.**
4. **Set high and low filter as excel equations.**

1. HIGH = Min(B4,B5)
2. LOW = Max(B6:B7)

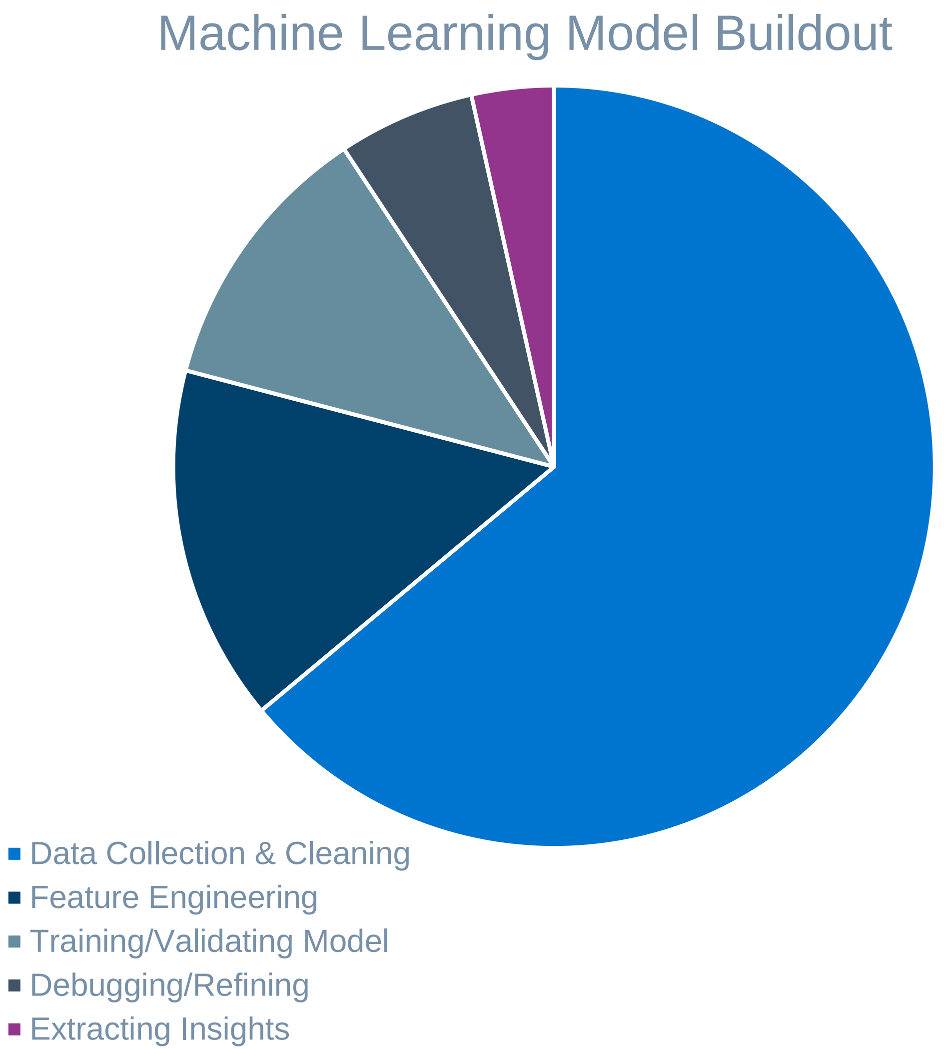
Check with your data historian to see if the filter expression feature is available.

FEATURE ENGINEERING

FEATURE ENGINEERING

Machine learning technique used to create new variables. Used to enhance model accuracy.

- **Summary Statistics**
  - Mean, Standard Deviation, Min/Max, Median, N
  - Graphing these can show clearer trends than raw data.
- **Feature Creation**
  - Total Sugars, Ethanol/Solids, Ethanol/Glycerol
  - Supplements raw data going into a model.
  - Improves model accuracy and improves insights for data analytics.



Public

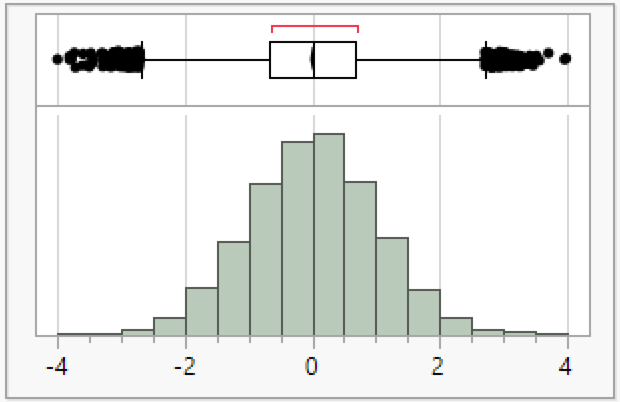
FEATURE ENGINEERING

Imputation	How to handle missing values. Numeric: Fill in with mean, median, mode. Fix with a calculated value (binned values). Categorical: Label text with “Missing” or “Blank”
Outliers	Remove the values. Quickest method, but the model will lose possibly valuable data. Smart Replace. Impute. Linear regression models susceptible to outliers. Cap value with an arbitrary max or min value based on distribution.
Transform	Log Scale – Convert skewed distributions into a more “normal” distribution. Helps with outliers. Binning – Create intervals that bin the float values into integers. Bins can have varying sizes. Scaling – This normalizes (0-1) or standardizes (0 mean, 1 variance) the range of all features.
One-Hot Encoding Response-Coding	Convert finite data (like categorical) into integers. Works well for Ferm #s, Mash Trains, Trial Conditions. Binary: Converts to 0 or 1. Works well for Pass/Fail, Above/Below average. Use for rare occurrences. Response Coding – creating conditional summary statistics.

RANDOMNESS

- Always add randomness to the models, even when evaluating linear regressions.
- Do not assign correlation to features with lower R^2 than random.

Random Normal



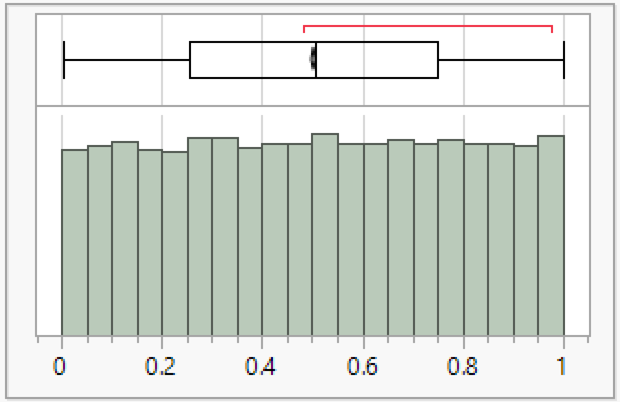
Quantiles

100.0%	maximum	3.9723024
99.5%		2.5697008
97.5%		1.9473761
90.0%		1.2726215
75.0%	quartile	0.6783057
50.0%	median	0.0124269
25.0%	quartile	-0.667821
10.0%		-1.275999
2.5%		-1.951007
0.5%		-2.52479
0.0%	minimum	-3.984277

Summary Statistics

Mean	0.0041718
Std Dev	0.9953178
Std Err Mean	0.0053641
Upper 95% Mean	0.0146857
Lower 95% Mean	-0.006342
N	34,429
N Missing	0

Random Uniform



Quantiles

100.0%	maximum	0.9999834
99.5%		0.9947546
97.5%		0.9749291
90.0%		0.9012684
75.0%	quartile	0.7518406
50.0%	median	0.5043985
25.0%	quartile	0.2558546
10.0%		0.1023137
2.5%		0.0260997
0.5%		0.0057587
0.0%	minimum	5.4082e-5

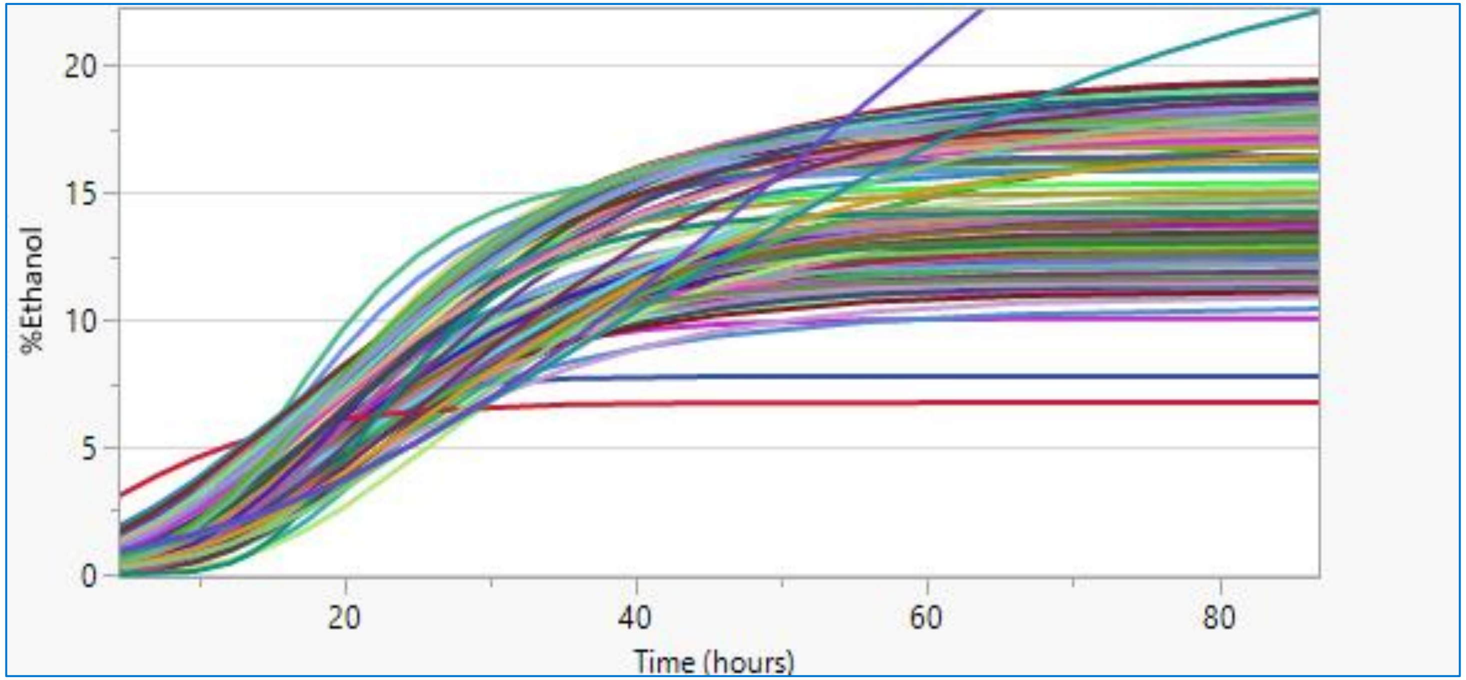
Summary Statistics

Mean	0.5026797
Std Dev	0.2879275
Std Err Mean	0.0015517
Upper 95% Mean	0.5057212
Lower 95% Mean	0.4996382
N	34,429
N Missing	0

FIT CURVE ANALYSIS

Trend any HPLC analyte across fermentation time.

Better understand fermentation kinetics.



Great way to incorporate feature engineering into models. For Ethanol, get information like:

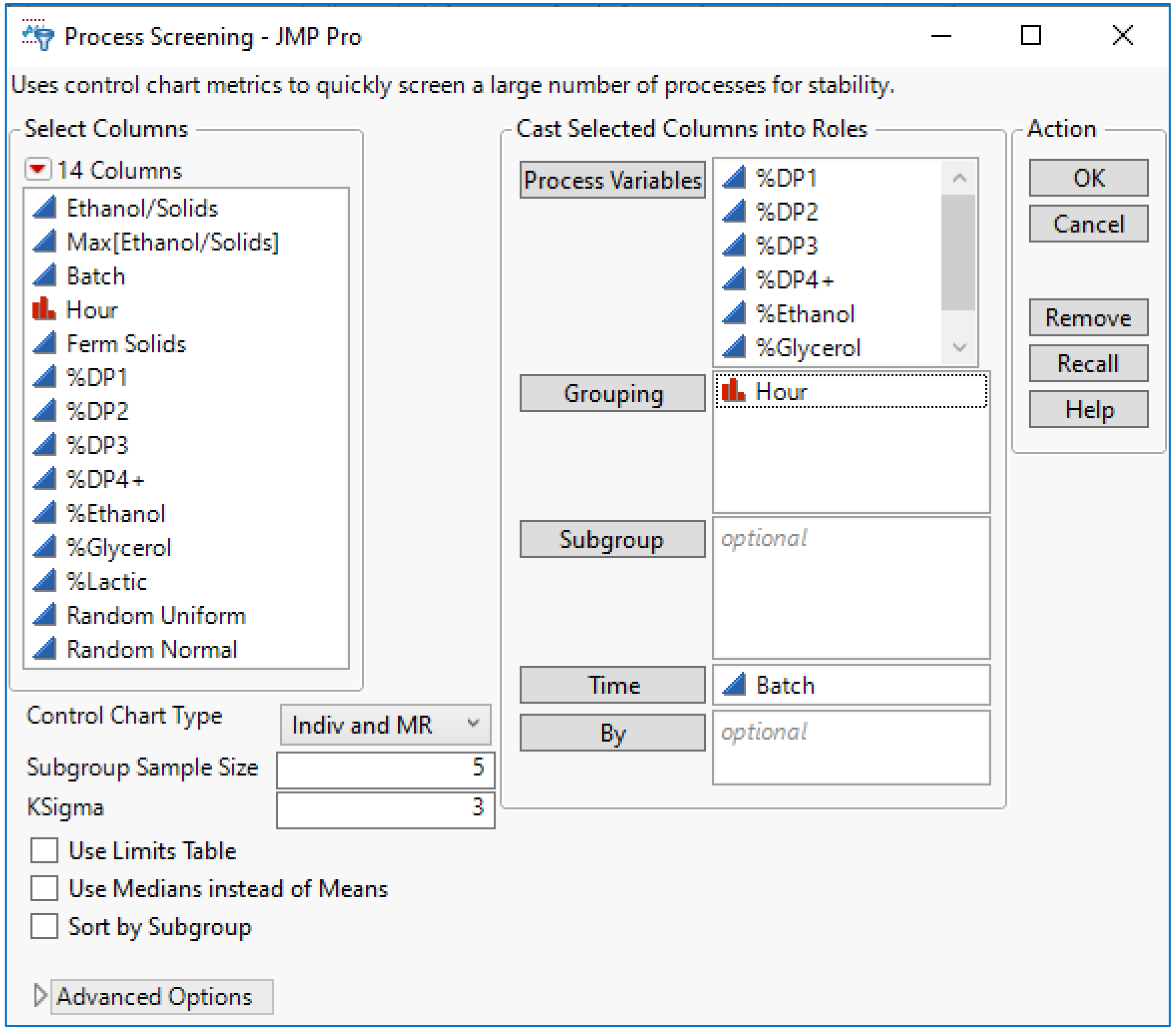
Inflection point – when ethanol production rate of growth is negative.

Asymptote – Peak ethanol values based on the curve.

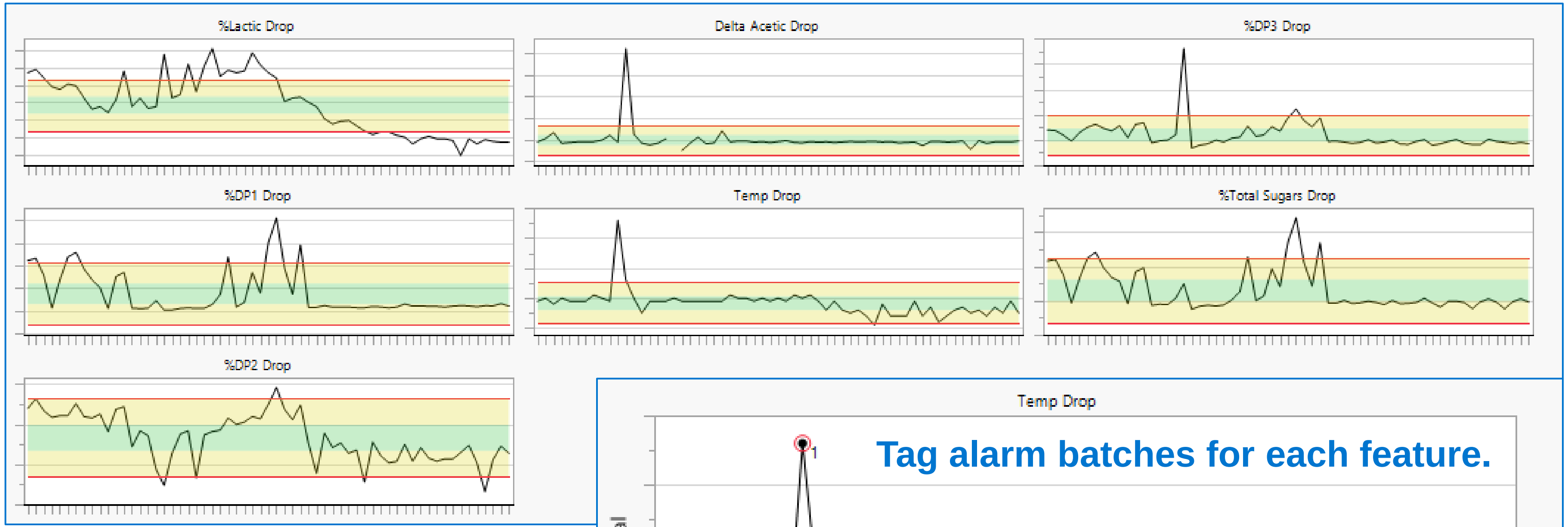
PROCESS SCREENING

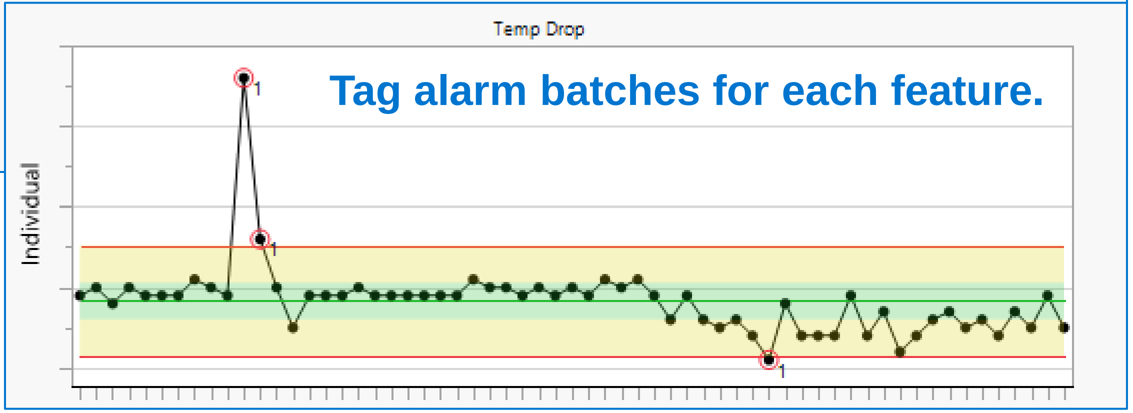
PROCESS SCREENING

- Process Screening Setup can be scripted for Fermentation.
- Good practice to have a data filter for Month/Year to see local changes.



PROCESS SCREENING





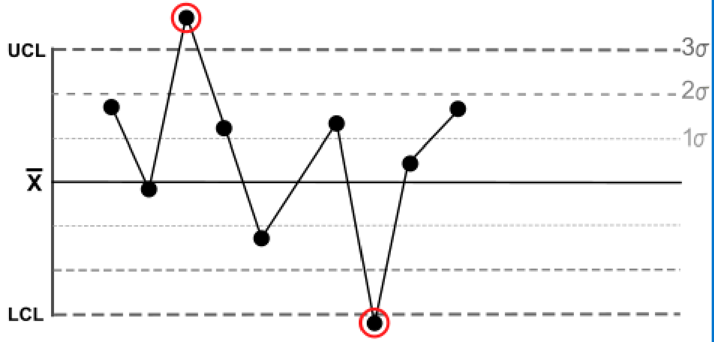
Sort Features by Stability

Column	Stability Index
%Lactic	3.33
Delta Acetic	2.45
%DP3	2.10
%DP1	2.08
Temp	1.94
%Total Sugars	1.88
%DP2	1.84
%Glycerol	1.65
pH	1.56
Brix	1.45
%DP4+	1.39
Delta Glycerol	1.35
%Acetic	1.32
Delta Lactic	1.27

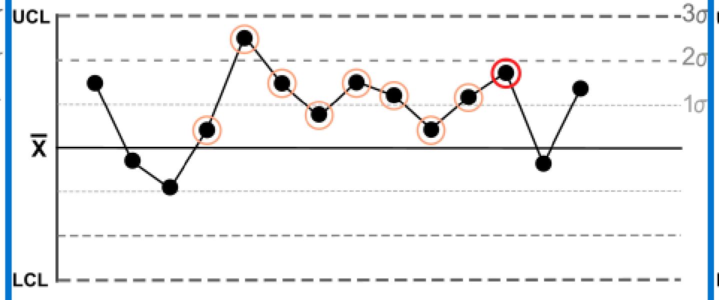
PROCESS SCREENING

Nelson Rules

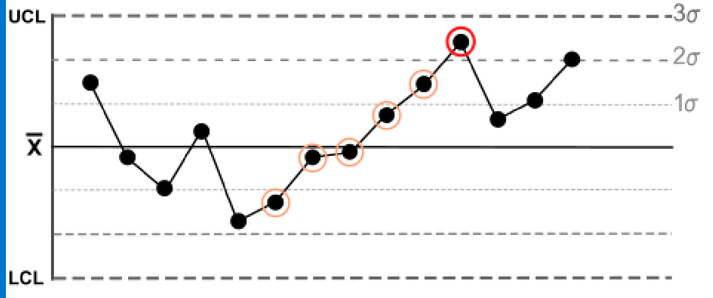
Rule 1: One point is more than 3 standard deviations from the mean



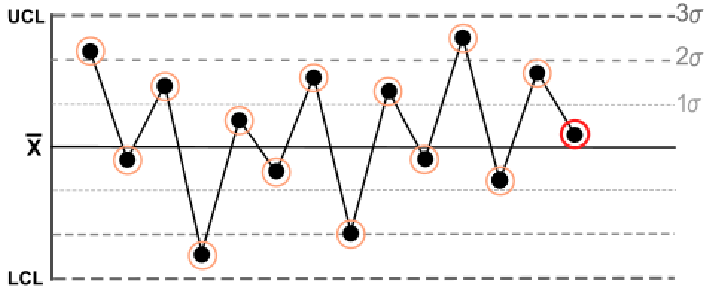
Rule 2: Nine (or more) points in a row are on the same side of the mean



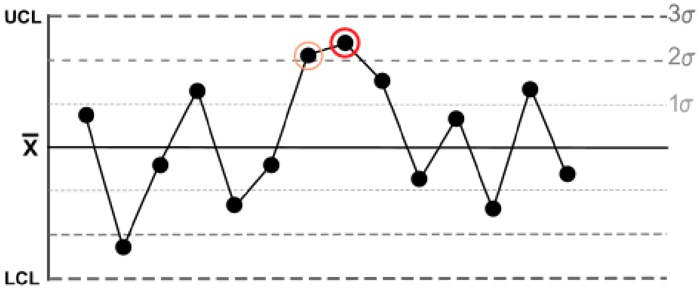
Rule 3: Six (or more) points in a row are continually increasing (or decreasing)



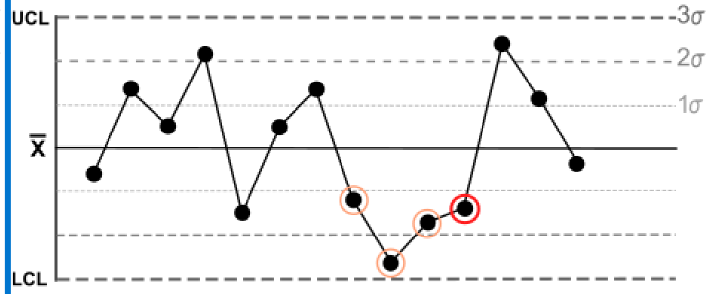
Rule 4: Fourteen (or more) points in a row alternate in direction, increasing then decreasing



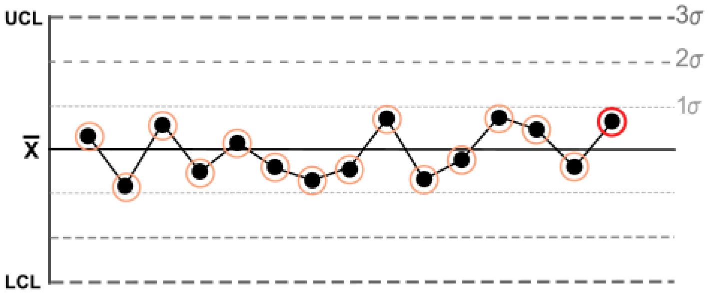
Rule 5: Two (or three) out of three points in a row are more than 2 standard deviations from the mean in the same direction



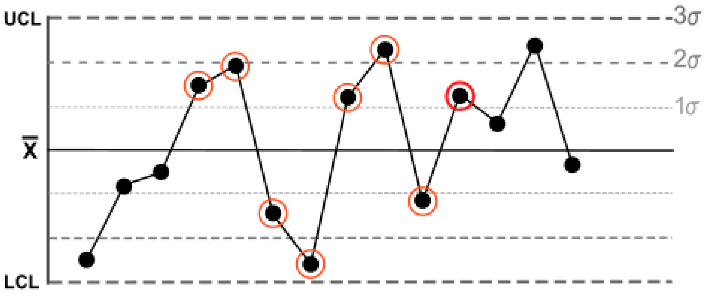
Rule 6: Four (or five) out of five points in a row are more than 1 standard deviation from the mean in the same direction



Rule 7: Fifteen points in a row are all within 1 standard deviation of the mean on either side of the mean

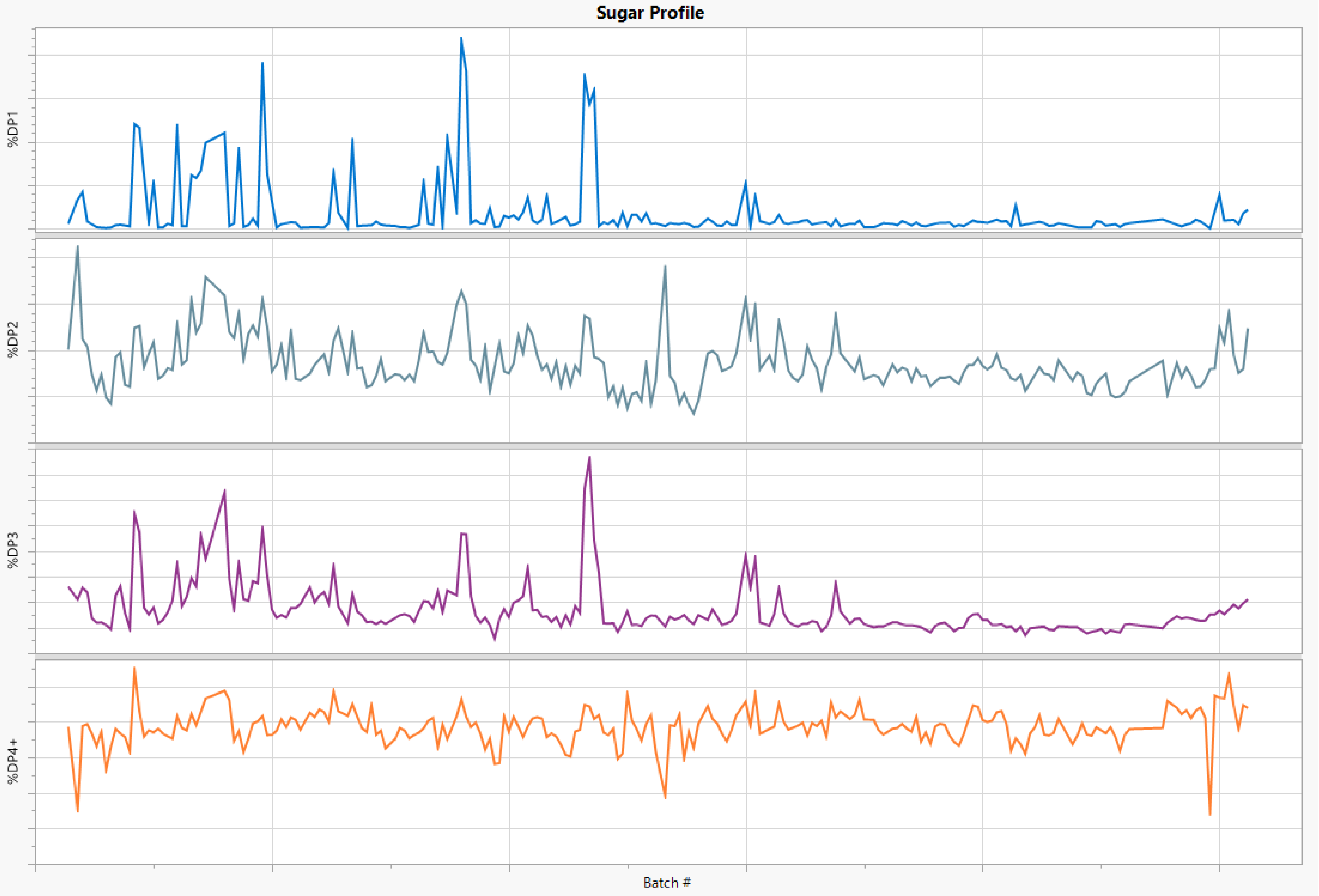


Rule 8: Eight points in a row exist with none within 1 standard deviation of the mean and the points are in both directions from the mean



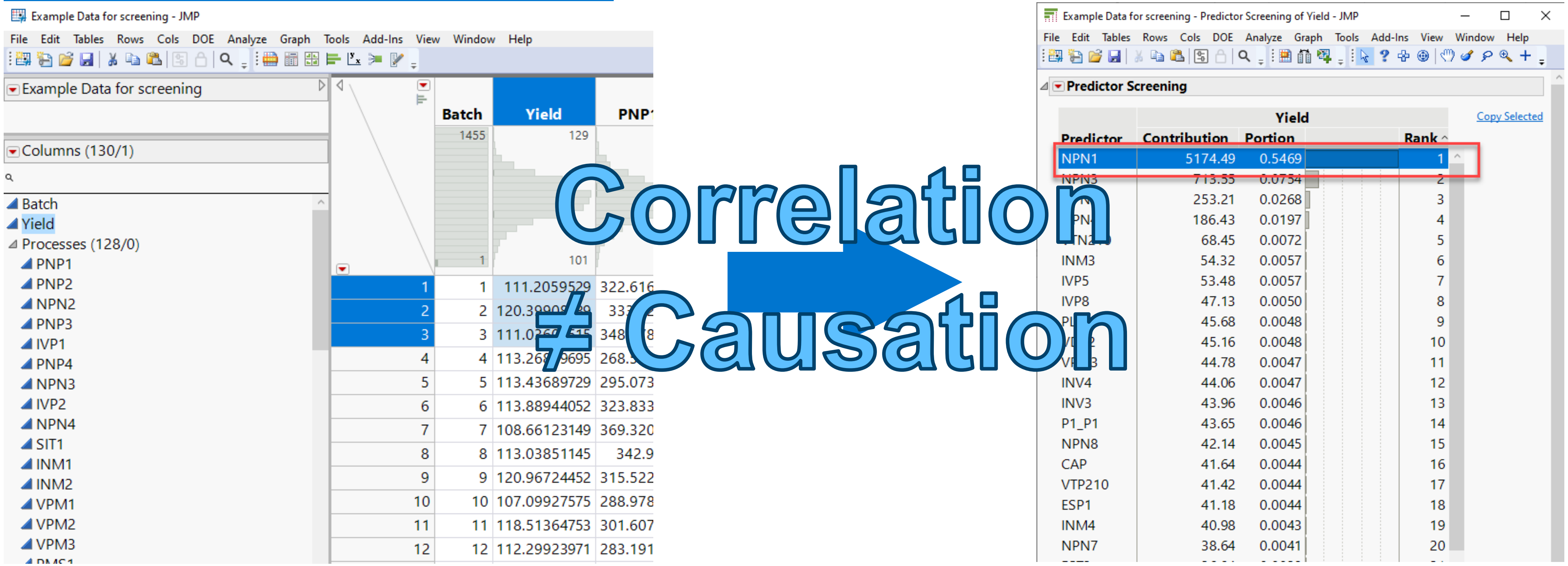
Highlight the rules that show trends and outliers.

PREDICTOR SCREENING


Sugar Profile
%DP1
%DP2
%DP3
%DP4+
Batch #

SCREENING FOR IMPORTANT VARIABLES

Predictor Screening



- Quickly go from many variables to find the most important variables, or the ones that correlate with the target.
- Include Random Uniform and Random Normal in the screening. Focus on variables above Random.

VARIABILITY ANALYSIS

Case Study

iff

VARIABILITY ANALYSIS

Overview

Goal

Estimate how much of the variability in the selected performance metric is stemming from each product type, and how big the residual is for a specific dataset.

Model

A **linear mixed model** is generalization of a linear regression model, where the mixed model can also take grouping in the data into account.

Assumption

Random effect coefficients and residuals are drawn from a normal distribution. The variance of this normal distributions tells us how much of the variability in data can be assigned to that grouping.

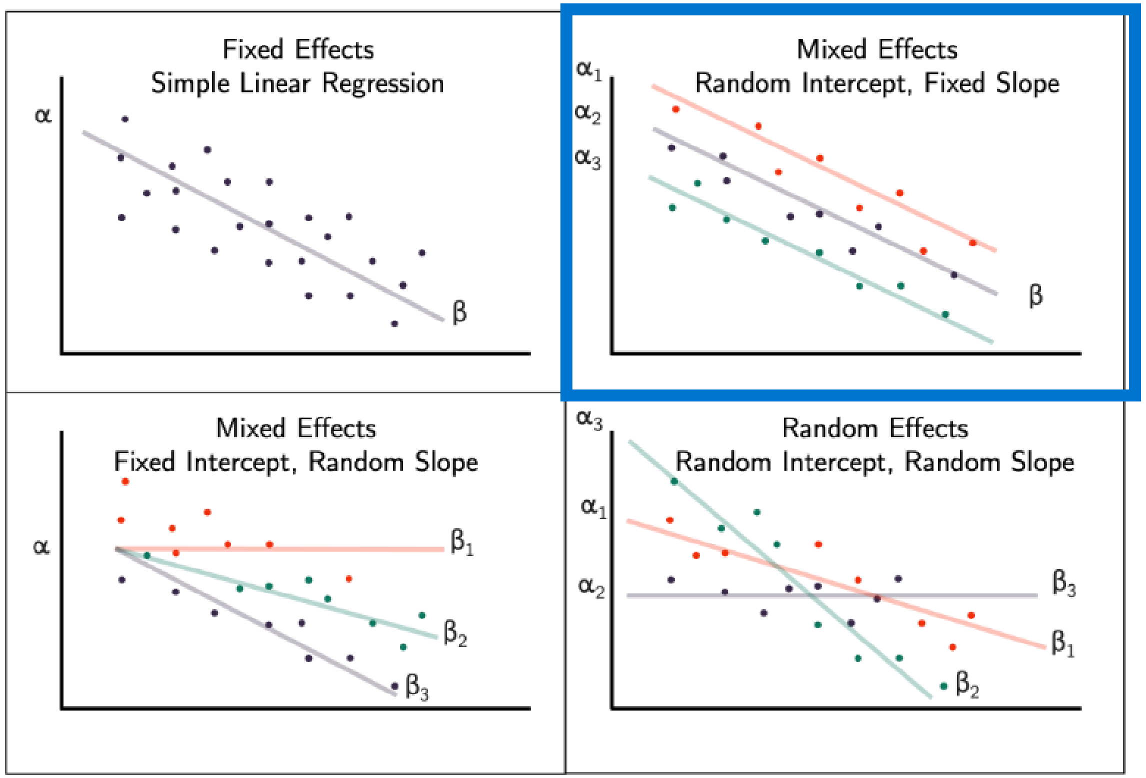


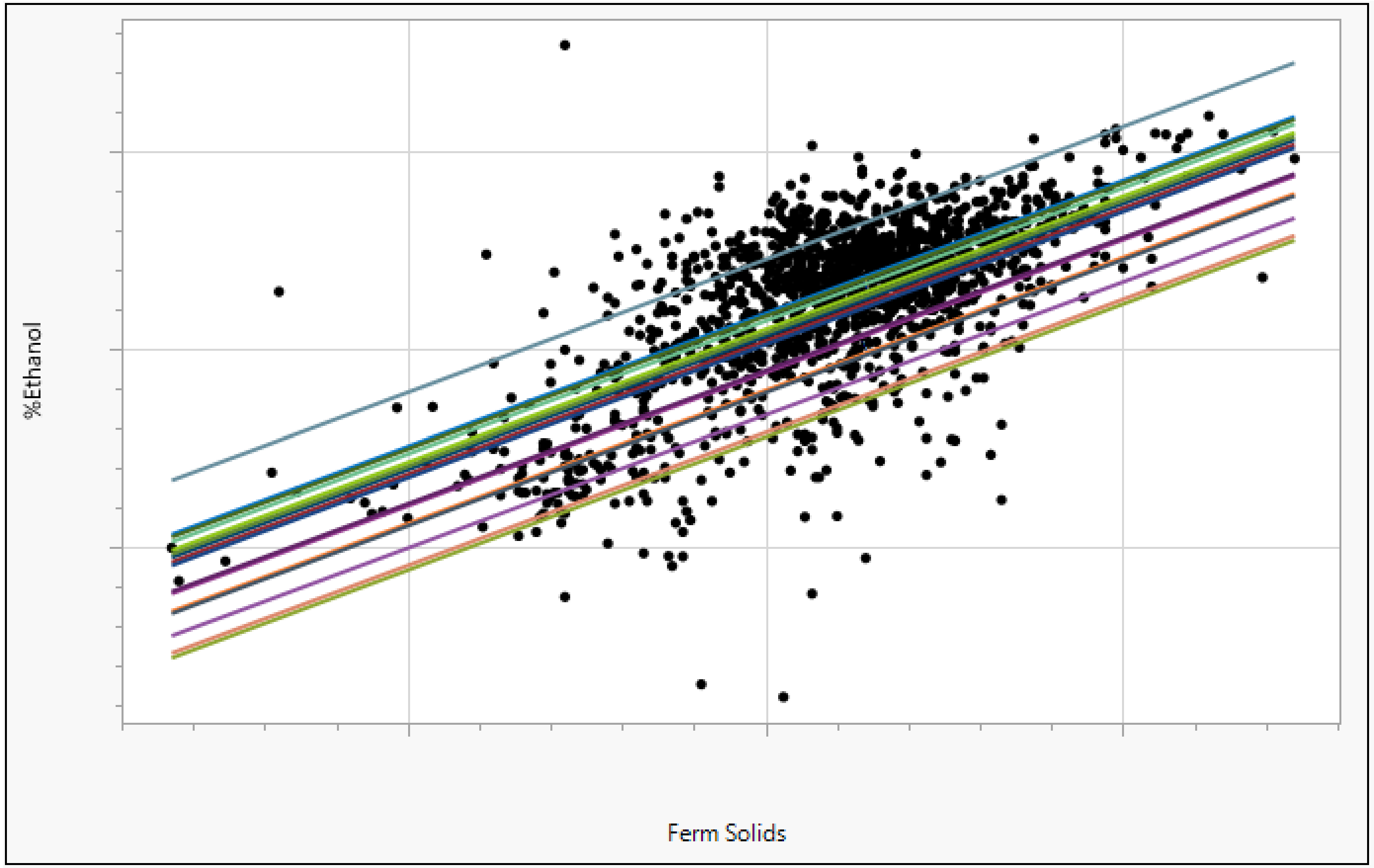
Image source: https://towardsdatascience.com/how-linear-mixed-model-works-350950a82911

EXAMPLE: LINEAR MIXED MODEL

Grouping: Yeast type

Here Ethanol Drop is the y-metric, Ferm solids is the fixed effect, and Yeast Type is the Random effect, adding a type dependent **random intercept.**

These random intercepts are drawn from a normal distribution. The variance of the normal distribution compared to the unexplained variability left in Ethanol Drop, tells us how much of the variability in Ethanol drop, not accounted for by Ferm Solids, can be accounted for by the Yeast type.



This chart shows which yeast types performed better. Let's discover what else are the core drivers of performance.

EXAMPLE: LINEAR MIXED MODEL

Grouping: Yeast

After a fit to Ferm Solids, variability left in Ethanol Drop is attributed to

- 47% Prop Add Yest Type
- 53% Residual

Residual will include other product changes, as well as process changes, corn quality changes etc.

It should be considered that process changes aligned in time with Yeast Type changes, will appear as an effect of the Yeast Type. This is why multiple changes during product trials make results tricky to interpret!

Variance Component Estimates are based on a model fit to historical data. Hence changes in data or in the model construction will change the outcome.

If we want to further understand how **other products** or interactions between products impact the performance, we need to include all products in the linear mixed model (see next slide).

Core driver of process variability	Description	Size of Variability
Yeast	Variability in Ethanol caused by differences between Yeasts	47%
Residual: Process parameters, Corn quality, etc.	Differences in Ethanol caused by all other features.	53%

Public

LINEAR MIXED MODEL: INCL. ALL PRODUCTS

REML Variance Component Estimates

Random Effect	Pct of Total
GA Type	3.058
Ferm Add Antibiotics Type	1.277
Prop Add Yeast Type	19.932
Prop Add Antibiotics Type	0.000
Prop Adds Prop GA Type	0.000
GA Type*Ferm Add Antibiotics Type*Prop Adds Prop GA Type*Prop Add Antibiotics Type*Prop Add Yeast Type	2.986
GA Type*Ferm Add Antibiotics Type	0.000
GA Type*Prop Adds Prop GA Type	0.220
Ferm Add Antibiotics Type*Prop Adds Prop GA Type	0.030
GA Type*Ferm Add Antibiotics Type*Prop Adds Prop GA Type	0.451
GA Type*Prop Add Antibiotics Type	0.000
Ferm Add Antibiotics Type*Prop Add Antibiotics Type	0.000
GA Type*Ferm Add Antibiotics Type*Prop Add Antibiotics Type	4.967
Prop Adds Prop GA Type*Prop Add Antibiotics Type	2.205
GA Type*Prop Adds Prop GA Type*Prop Add Antibiotics Type	1.793
Ferm Add Antibiotics Type*Prop Adds Prop GA Type*Prop Add Antibiotics Type	0.000
GA Type*Ferm Add Antibiotics Type*Prop Adds Prop GA Type*Prop Add Antibiotics Type	0.000
GA Type*Prop Add Yeast Type	0.000
Ferm Add Antibiotics Type*Prop Add Yeast Type	0.000
GA Type*Ferm Add Antibiotics Type*Prop Add Yeast Type	2.352
Prop Adds Prop GA Type*Prop Add Yeast Type	17.266
GA Type*Prop Adds Prop GA Type*Prop Add Yeast Type	0.000
Ferm Add Antibiotics Type*Prop Adds Prop GA Type*Prop Add Yeast Type	0.000
GA Type*Ferm Add Antibiotics Type*Prop Adds Prop GA Type*Prop Add Yeast Type	0.000
Prop Add Antibiotics Type*Prop Add Yeast Type	9.818
GA Type*Prop Add Antibiotics Type*Prop Add Yeast Type	0.000
Ferm Add Antibiotics Type*Prop Add Antibiotics Type*Prop Add Yeast Type	0.014
GA Type*Ferm Add Antibiotics Type*Prop Add Antibiotics Type*Prop Add Yeast Type	0.000
Prop Adds Prop GA Type*Prop Add Antibiotics Type*Prop Add Yeast Type	0.000
GA Type*Prop Adds Prop GA Type*Prop Add Antibiotics Type*Prop Add Yeast Type	2.113
Ferm Add Antibiotics Type*Prop Adds Prop GA Type*Prop Add Antibiotics Type*Prop Add Yeast Type	6.549
Residual	24.970
Total	100.000

LINEAR MIXED MODEL: ALL PRODUCTS

Core drivers of ETHANOL variability

Core driver of process variability	Description	Size of Variability	Model insights
Yeast	Variability in Ethanol caused by differences between Yeasts	47%	• **Yeast is a key driver of process variability.**
GA*Yeast	Variability in Ethanol caused by differences in the specific combination of GA and Yeast		• **The interaction between yeast and GA type is a key driver of process variability.**
Antibiotics*Yeast	Variability in Ethanol caused by differences in the specific combination of Antibiotics and Yeast		• **The interaction between yeast and Antibiotics also has some impact on process variability,** though significantly less than yeast and GA type combination.
Residual: Process parameters, Corn quality, etc.	Differences in Ethanol caused by the process conditions and corn quality differences.	25%	• **Process setpoints are not consistently achieved** in production and tend to have variation in fermentation time, percent backset and more. • **Corn quality** is not well accounted for. It could potentially be a big unknown driver of ethanol variability

CREATE YOUR OWN LEAST SQUARES LINEAR MODELS

LEAST SQUARES LINEAR MODEL

Find the most impactful features

One feature

$Y = mx+b$

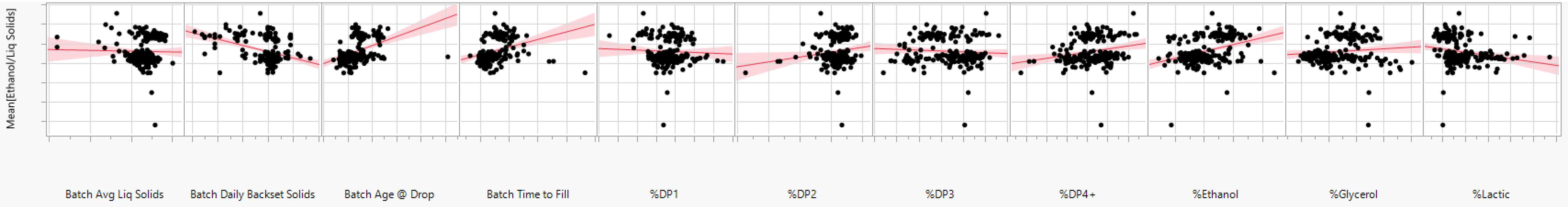
Two features

$Y = m_1x_1+m_2x_2+b$

Multiple features

$Y = b+m_1x_1+m_2x_2+\ldots$

Each feature individually has a poor linear fit with ethanol yield. How can you make a formula that fits all of these features into one equation?



LEAST SQUARE MODELS

Setup in Microsoft® Excel®

True = Calculated y-intercept
FALSE = y-intercept = 0

Each feature should be one column.

The Y-metric should also be a column.

Formula = LINEST([known_ys],[known_xs],TRUE,TRUE)

True = Full statistics table
FALSE = slope and y-intercept only

[Known_ys] will be the column of the Y metric

[known_xs] can be 1 or multiple columns of features

The output table will be 4 rows by and span all columns.

Let's do an example.
How well can we predict Ethanol/Solids based on 12-hr HPLC + Solids?

LEAST SQUARE MODELS

Early Ferm Example Setup

`=LINEST(C8:C215,D8:N215,TRUE,TRUE)` **Formula**

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
7	Batch	Hour	Yield	Liq Solids	Backset Sc	Age	Fill Time	DP1	DP2	DP3	DP4	Ethanol	Glycerol	Lactic
8	1257	12Hr	0.40	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
9	1258	12Hr	0.39	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
10	1259	12Hr	0.42	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
11	1263	12Hr	0.43	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
12	1264	12Hr	0.43	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Y-Metric

Features

Public

LEAST SQUARE MODELS

Early Ferm Example Setup

=LINEST(C8:C215,D8:N215,TRUE,TRUE)

Slopes are in reverse order of features!

Y-Intercept

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1			Slopes	-0.16295	-0.02186	0.051117	0.03916	-0.01938	0.023773	0.012179	0.009681609	2.20016E-06	0.001208	-0.01501	0.237098
2			Standard Error	0.030747	0.017902	0.005472	0.00239	0.005329	0.003866	0.002823	0.00146896	0.000161237	0.001519	0.001678	0.056984
3			R^2 Value	0.725107	0.033067	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
4			F Value	47.00031	196	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
5			RSS	0.565302	0.21431	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
6															
7	Batch	Hour	Yield	Liq Solids	Backset Sc	Age	Fill Time	DP1	DP2	DP3	DP4	Ethanol	Glycerol	Lactic	
8	1257	12Hr	0.40												
9	1258	12Hr	0.39												
10	1259	12Hr	0.42												
11	1263	12Hr	0.43												
12	1264	12Hr	0.43												

Y-Metric

Features

RSS = Regression of sum of squares

LEAST SQUARE MODELS

Early Ferm Example Setup

=LINEST(C8:C215,D8:N215,TRUE,TRUE)

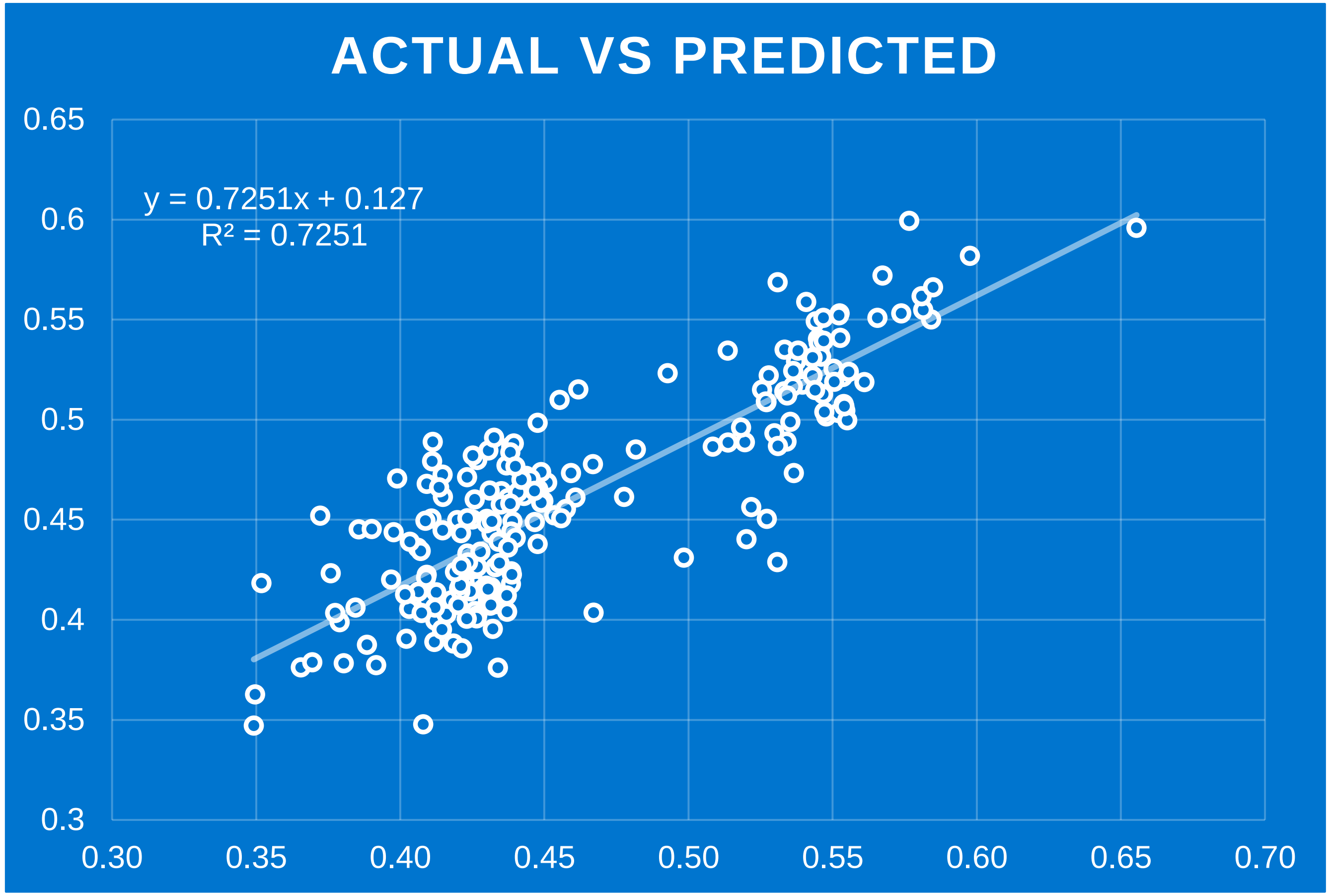
b (y-intercept)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1				-0.16295	-0.02186	0.051117	0.03916	-0.01938	0.023773	0.012179	0.009681609	2.20016E-06	0.001208	-0.01501	0.237098
2				0.030747	0.017902	0.005472	0.00239	0.005329	0.003866	0.002823	0.00146896	0.000161237	0.001519	0.001678	0.056984
3				0.725107	0.033067	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
4				47.00031	196	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
5				0.565302	0.21431	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
6															
7	Batch	Hour	Yield	Liq Solids	Backset Sc	Age	Fill Time	DP1	DP2	DP3	DP4	Ethanol	Glycerol	Lactic	Pred. Yield
8	1257	12Hr	0.40												
9	1258	12Hr	0.39												
10	1259	12Hr	0.42												
11	1263	12Hr	0.43												
12	1264	12Hr	0.43												

Predicted Yield = $m_1x_1 + m_2x_2 + m_3x_3 + \ldots + b$

LEAST SQUARE MODELS

Results

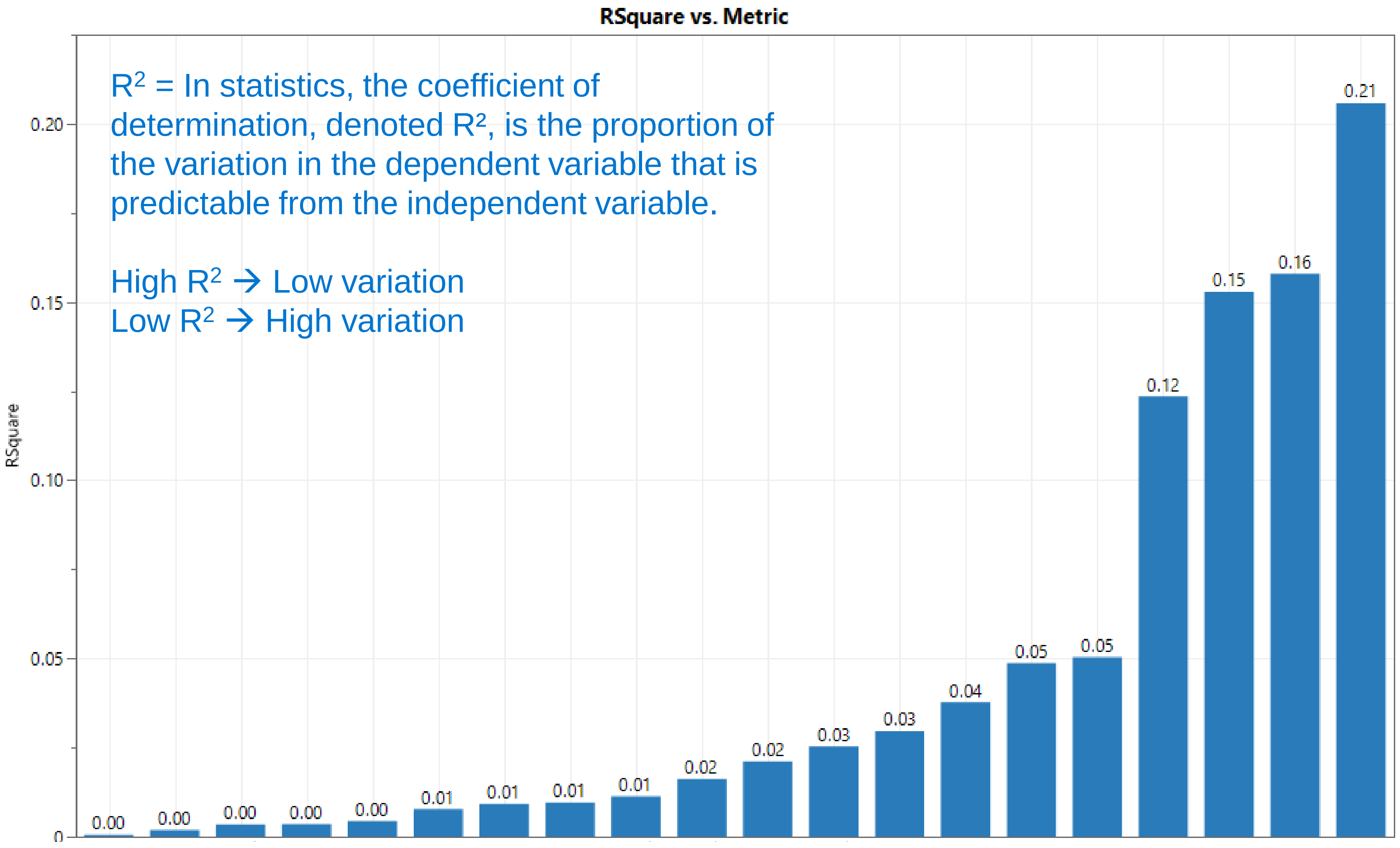


What can you do with results:

- Plot how changing certain features affect average performance.
- Interpret which features have the largest impact on performance (largest slope).
- Rerun model over time and see how the results shift.
- Create models specific to certain products to optimize various features.

PERFORMANCE METRIC EVALUATION

PERFORMANCE METRIC SELECTION



R^2 = In statistics, the coefficient of determination, denoted R^2, is the proportion of the variation in the dependent variable that is predictable from the independent variable.

High R^2 → Low variation
Low R^2 → High variation

Features

Minimize the deviation in your metric for better insights.

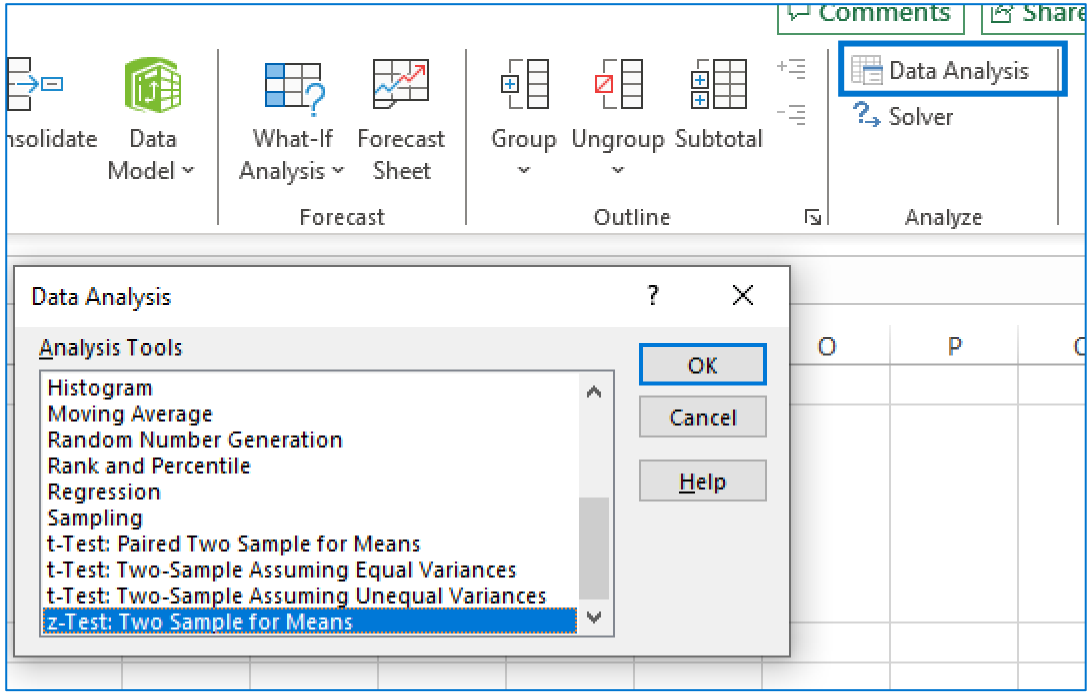
Evaluate multiple metrics; basic and advanced.

Evaluating the best metric allows for the most success with product trials and process changes.

TRIAL EVALUATION

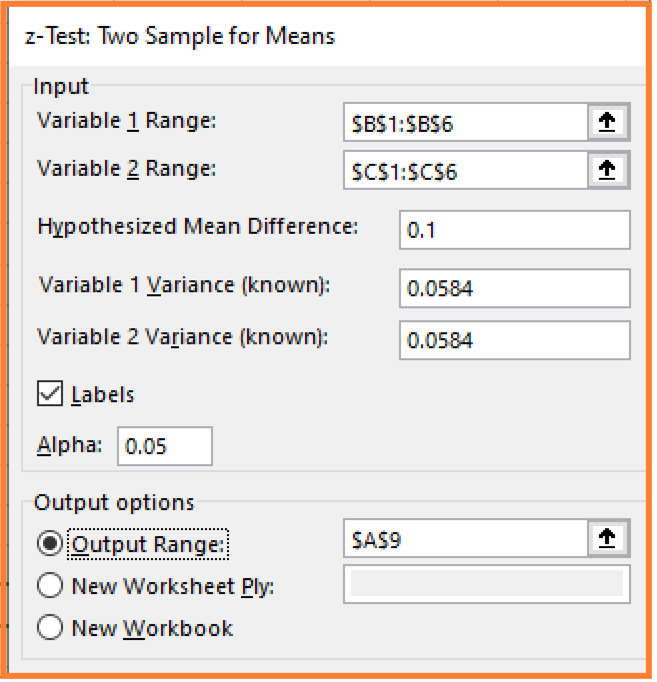
Z-test in Microsoft® Excel®

To access:



Inputs & Results:

	A	B	C
1		baseline	trial
2		13	13.1
3		13.1	13.2
4		13.5	13.6
5		12.8	12.9
6		12.9	13
7	Variance	0.0584	0.0584
8			
9	z-Test: Two Sample for Means		
10			
11		*baseline*	*trial*
12	Mean	13.06	13.16
13	Known Variance	0.0584	0.0584
14	Observations	5	5
15	Hypothesized Mean Difference	0.1	
16	z	-1.31	
17	P(Z<=z) one-tail	0.10	
18	z Critical one-tail	1.64	
19	P(Z<=z) two-tail	0.19	
20	z Critical two-tail	1.96	

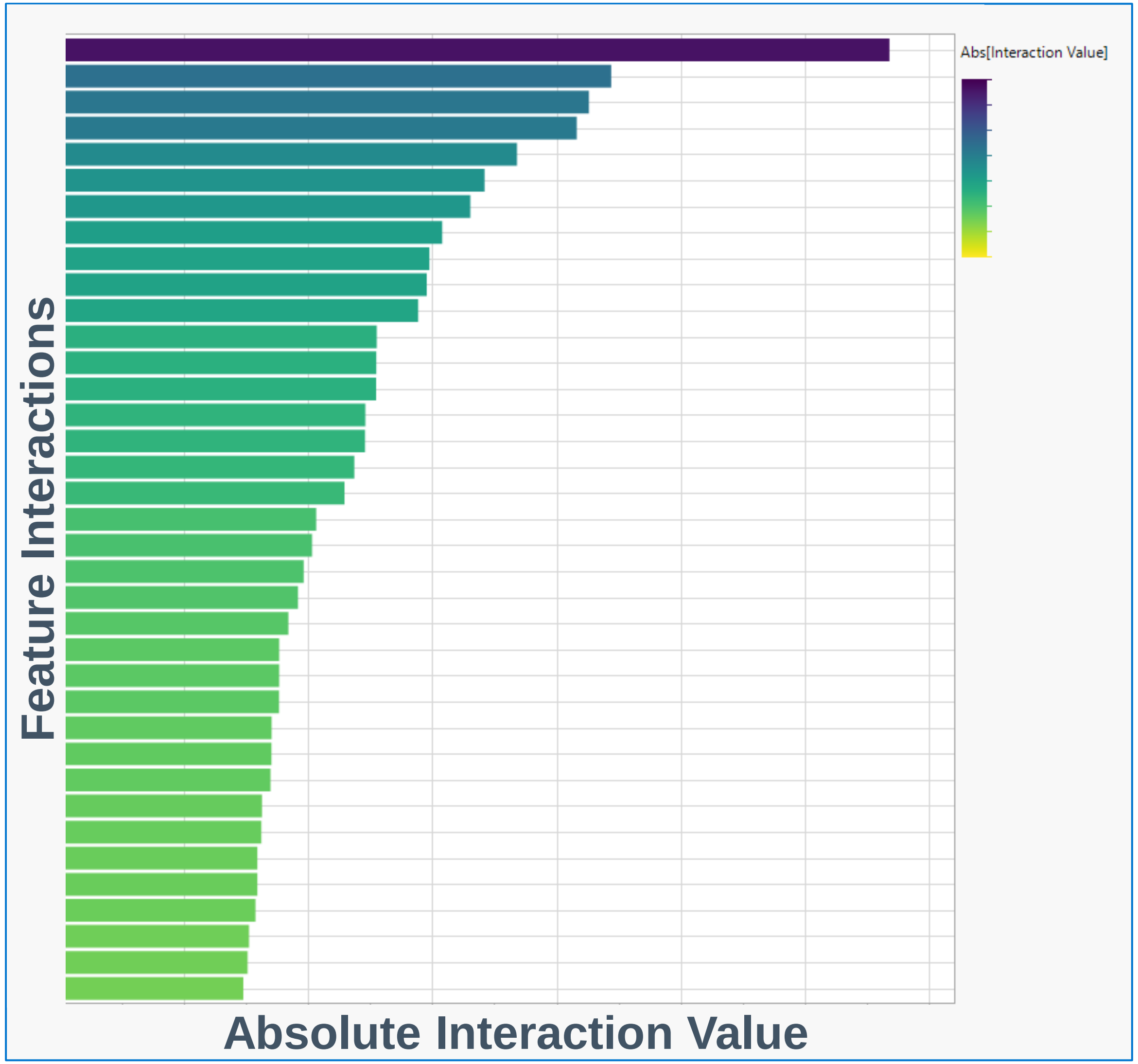


Based on the two-tail p-value, we can conclude with 81% certainty that the trial data exceeds the baseline data by 0.1.

PROCESS UNIT MACHINE LEARNING

PROCESS UNIT MACHINE LEARNING

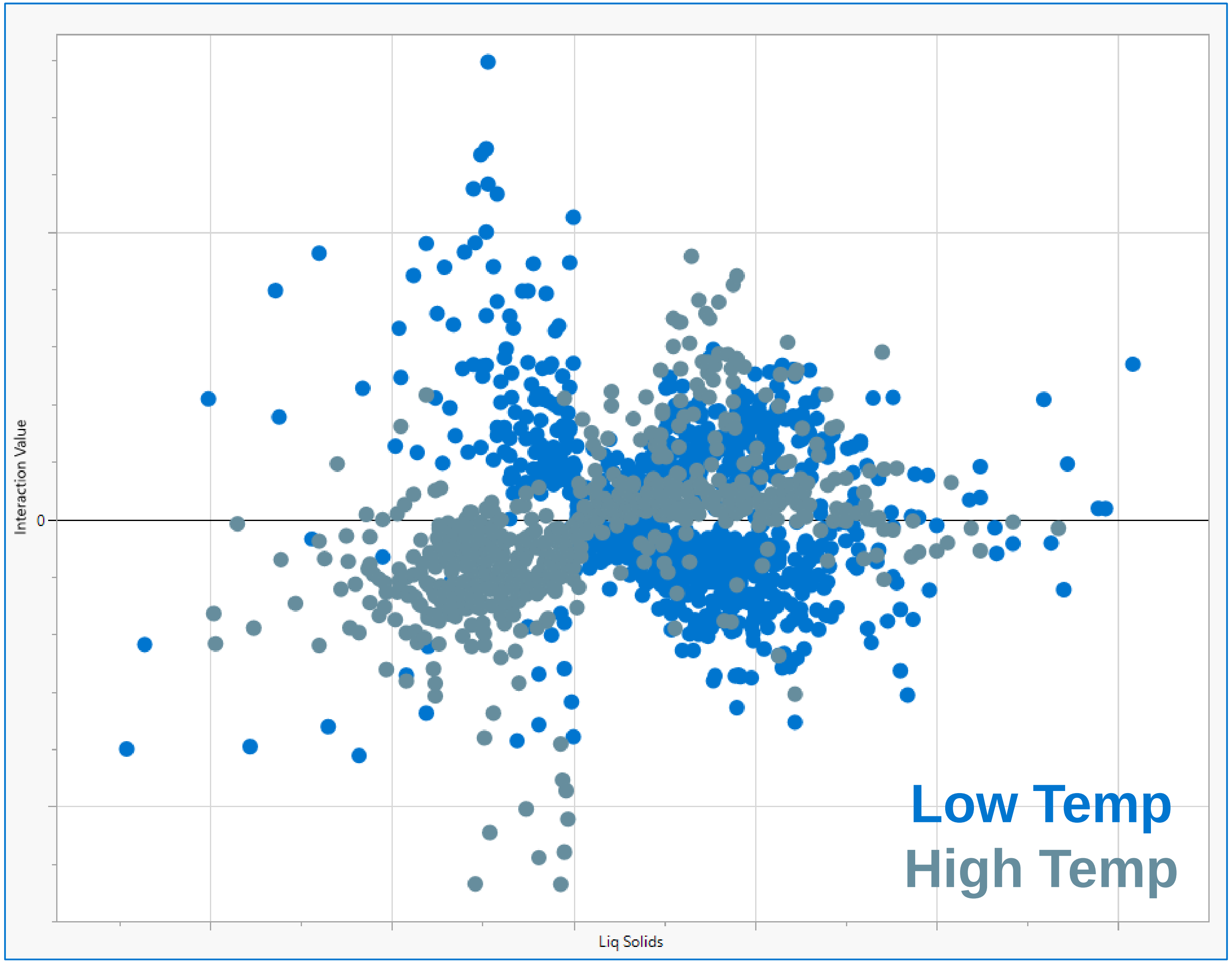
Interactions



In this model: Liq Solids and Slurry Temperature have the highest interaction value.

Interaction is 50% higher than the second highest interaction.

PROCESS UNIT MACHINE LEARNING



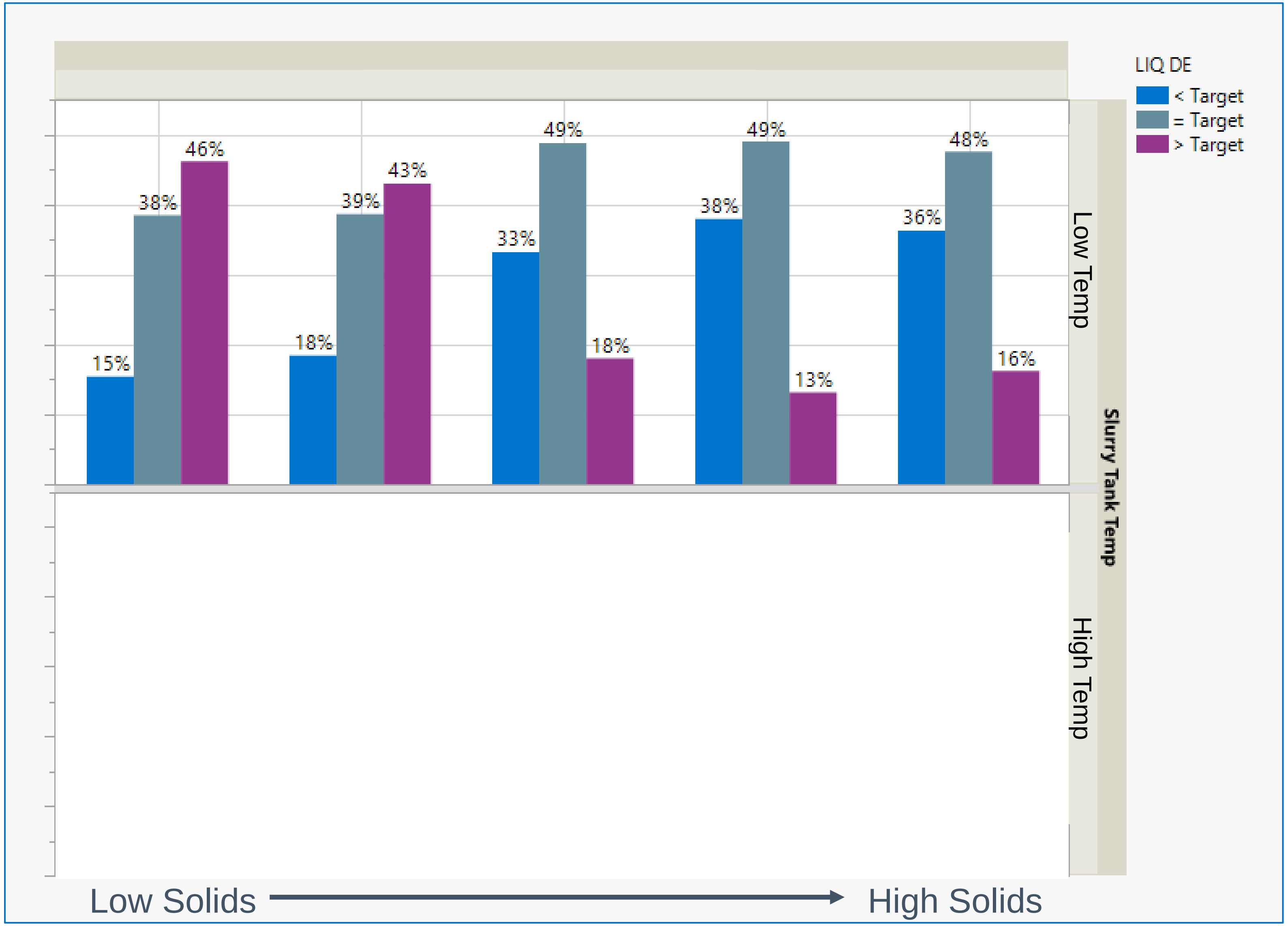
Feature Effect

Interaction Effect

Public

PROCESS UNIT MACHINE LEARNING

Probabilities



DYNAMIC TIME WARPING

DYNAMIC TIME WARPING

How it Works

Create a report of all available prop data, gathering data at 15-minute intervals.

Each prop is given a unique ID consisting of tub name, date, and hour of set time.

Create 5 distinct phases, based on the tub level, temperature, temperature controller output, and rate of change (direction and magnitude).

Phases:

1 = Wash

2 = Empty

3 = Fill

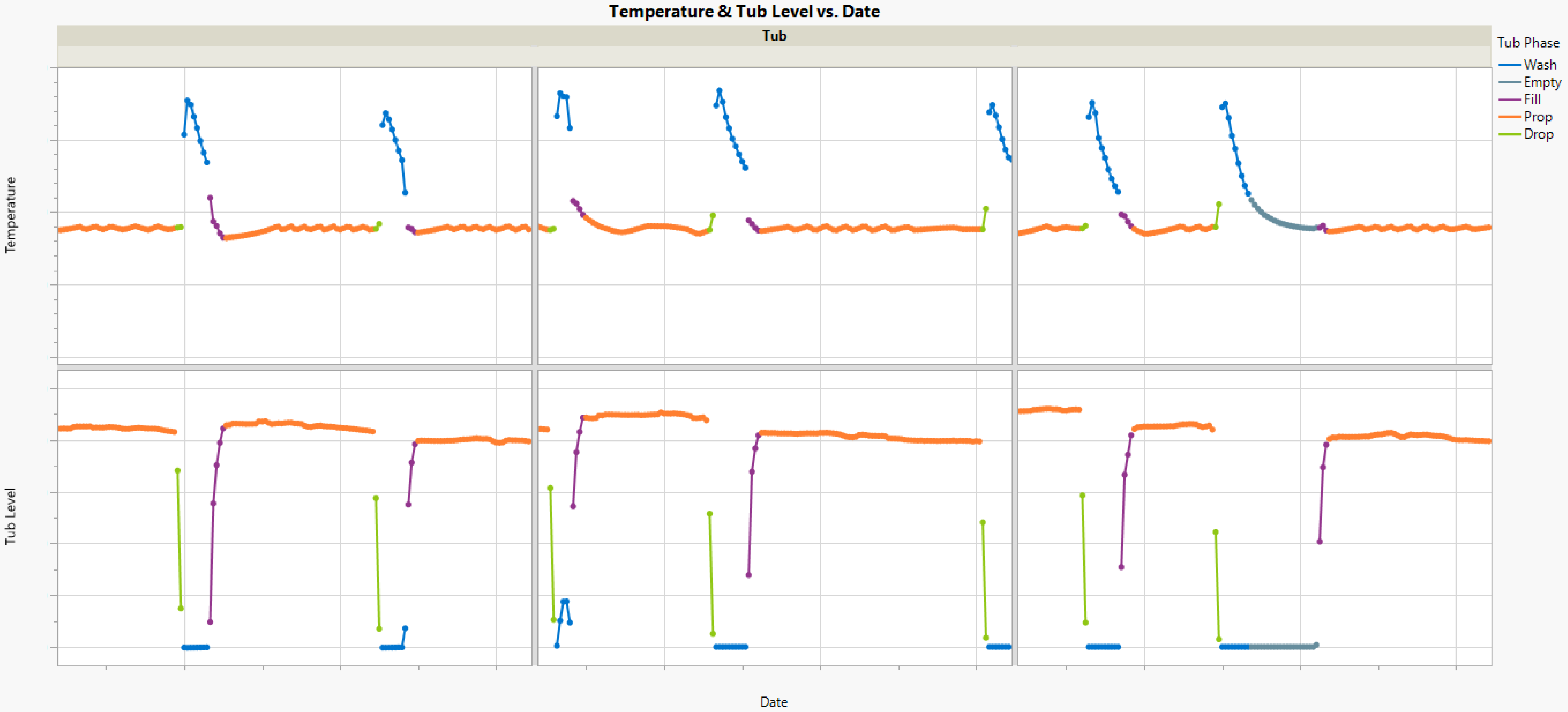
4 = Prop

5 = Drop

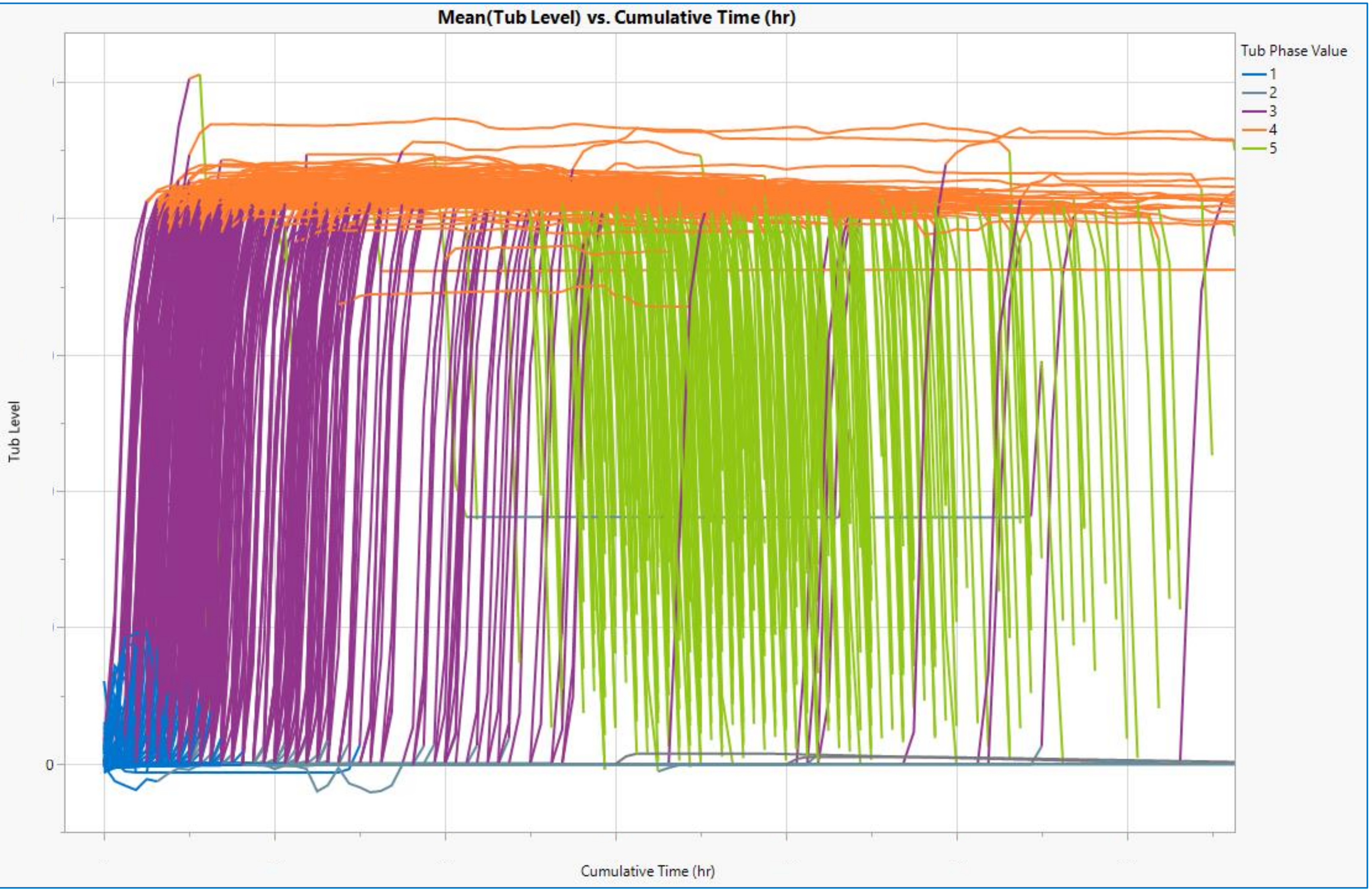
Find a better way to evaluate time-series data by overlaying all prop batches.

Then separate the prop duration into phases to evaluate separately.

HOW IT WORKS


Temperature & Tub Level vs. Date
Tub
Tub Phase
Wash
Empty
Fill
Prop
Drop
Temperature
Tub Level
Date

DYNAMIC TIME WARPING

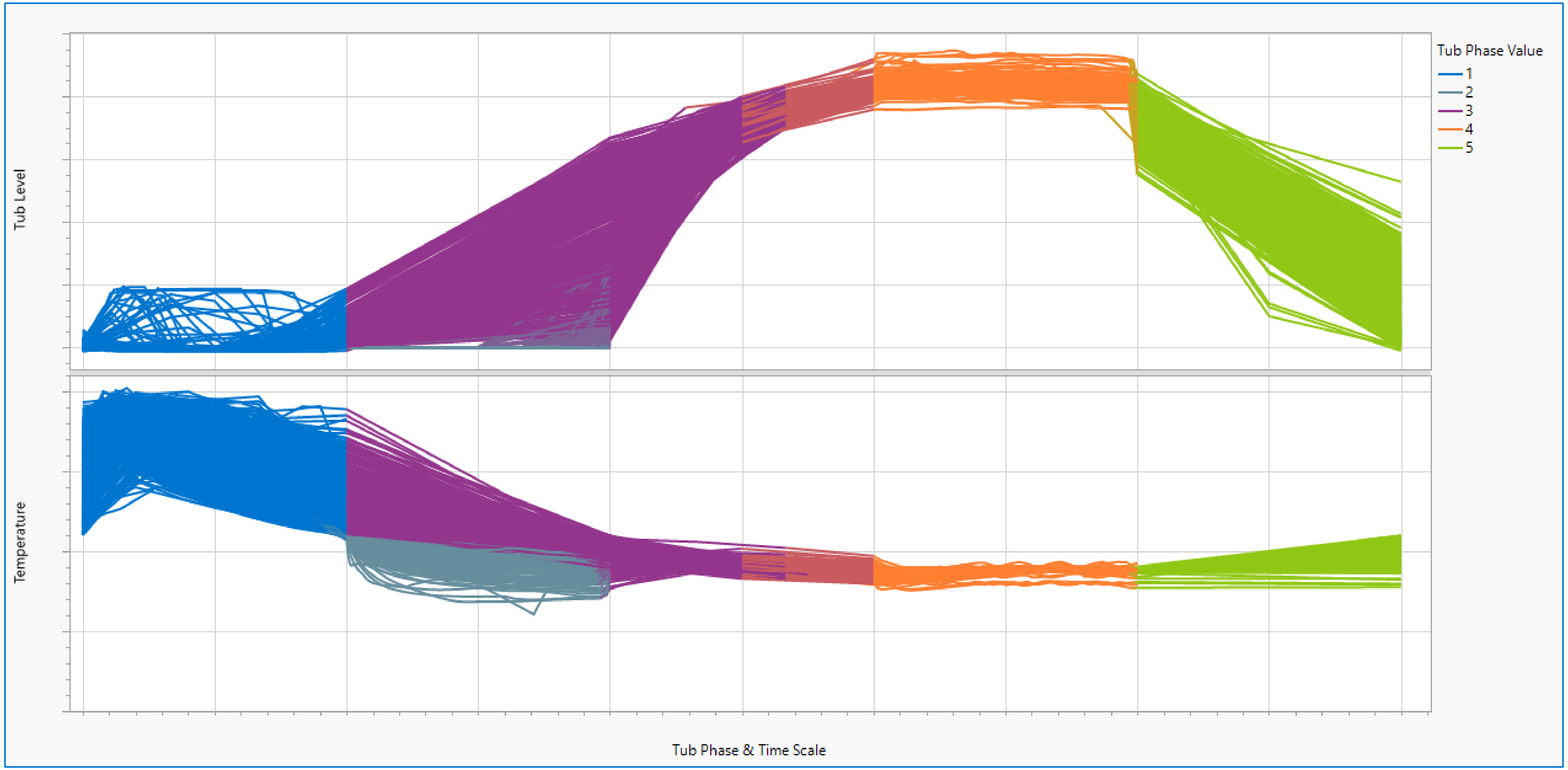


Evaluating props based on time can be difficult. The cumulative times do not always line up.

To best evaluate these phases, we need to group these phases based on relative time (phase time).

DYNAMIC TIME WARPING

By evaluating each phase of each prop by completeness, we have a better understanding of which features have the largest impact. This process of phasing the prop time is called “dynamic time warping”.



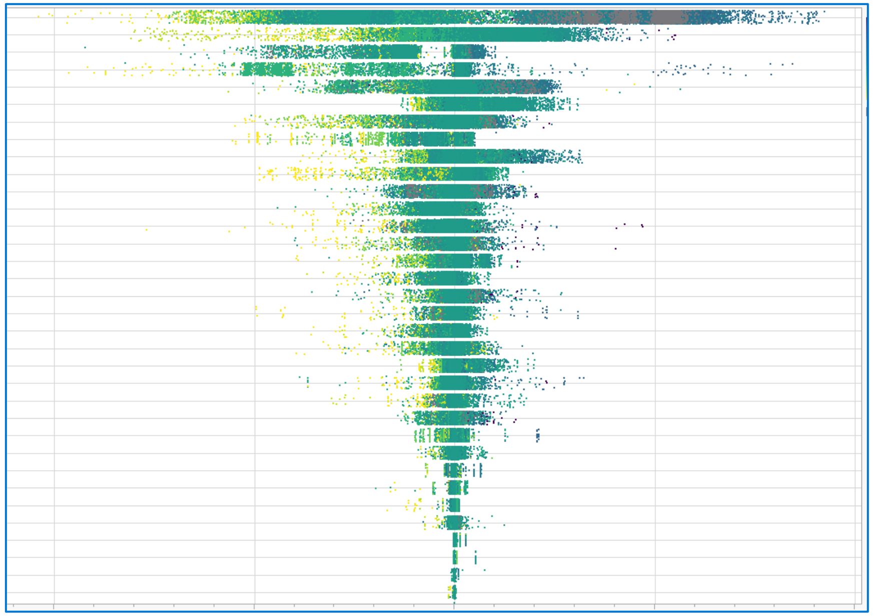
Now, instead of feeding *prop hours* into the model, we can use the better *phase hours*.

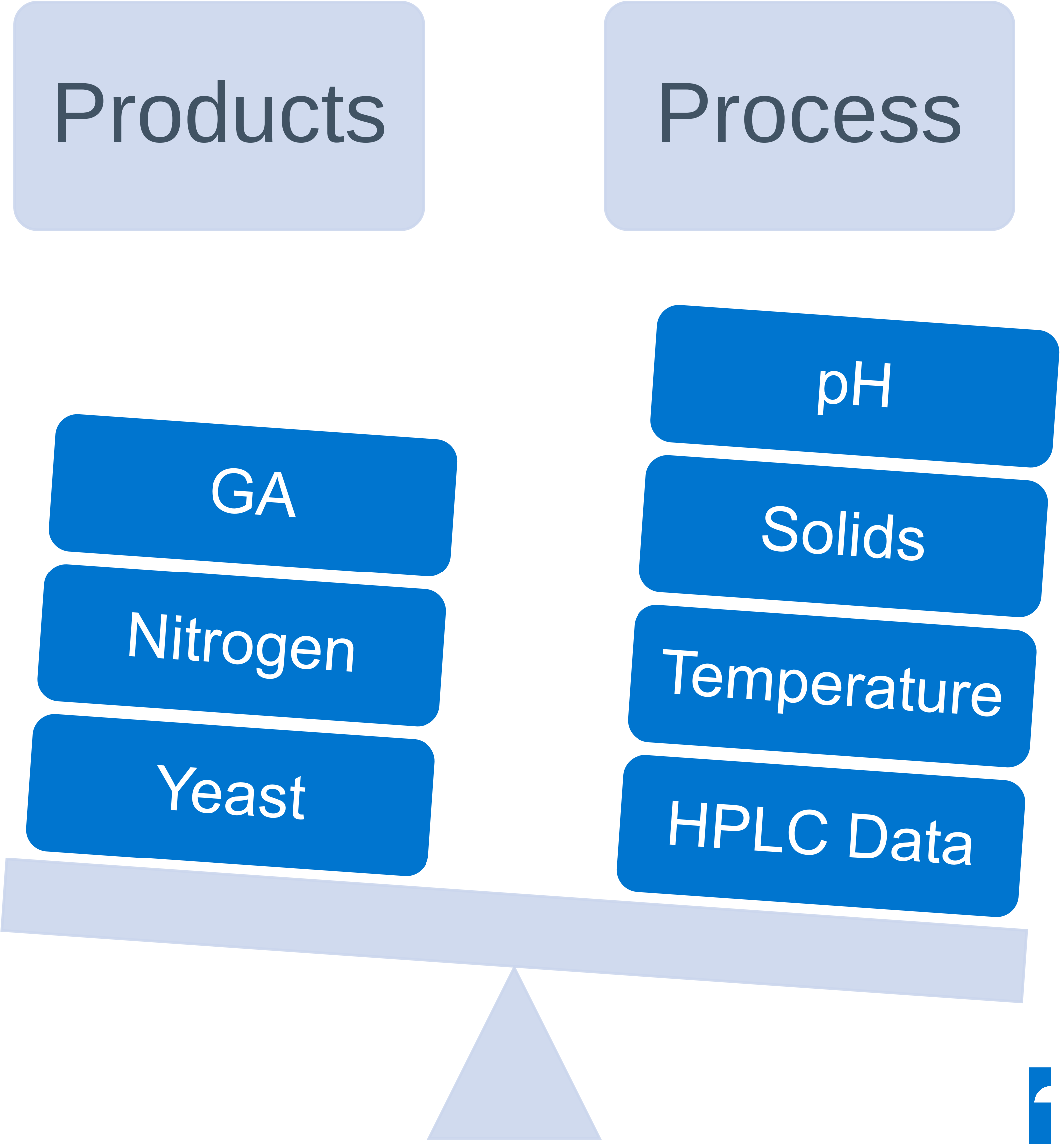
Phase hours also lets us evaluate each phase individually.

DYNAMIC TIME WARPING

Model Results

Gain insights on more than just HPLC from a couple of samples per prop.
Analyze adjustments to both products and process at the same time.
If the goal is viable cell count, what levers can we pull to collectively raise the cell count enough to decrease yeast volume pitched?





SUMMARY

What Can YOU Do?

Data Upload/Cleaning
Feature Engineering
Least Square Linear Models
Trial Evaluation

What Can I Do?

Automated Data Reporting
Process Screening
Fit Curve Analysis
Power Explorer/Trial Evaluation

What Can AI Do?

Predictor Screening
Variability Analysis
Partition Models
Trained/Validated Models
Dynamic Time Warping

Stay connected

www.xcelis.com

https://www.linkedin.com/showcase/xcelis-ethanol-solutions/

Grain Changers Community

Join Grain Changers for exclusive content for our valued partners.

Register for free at:

www.xcelis.com/grain-changers/