



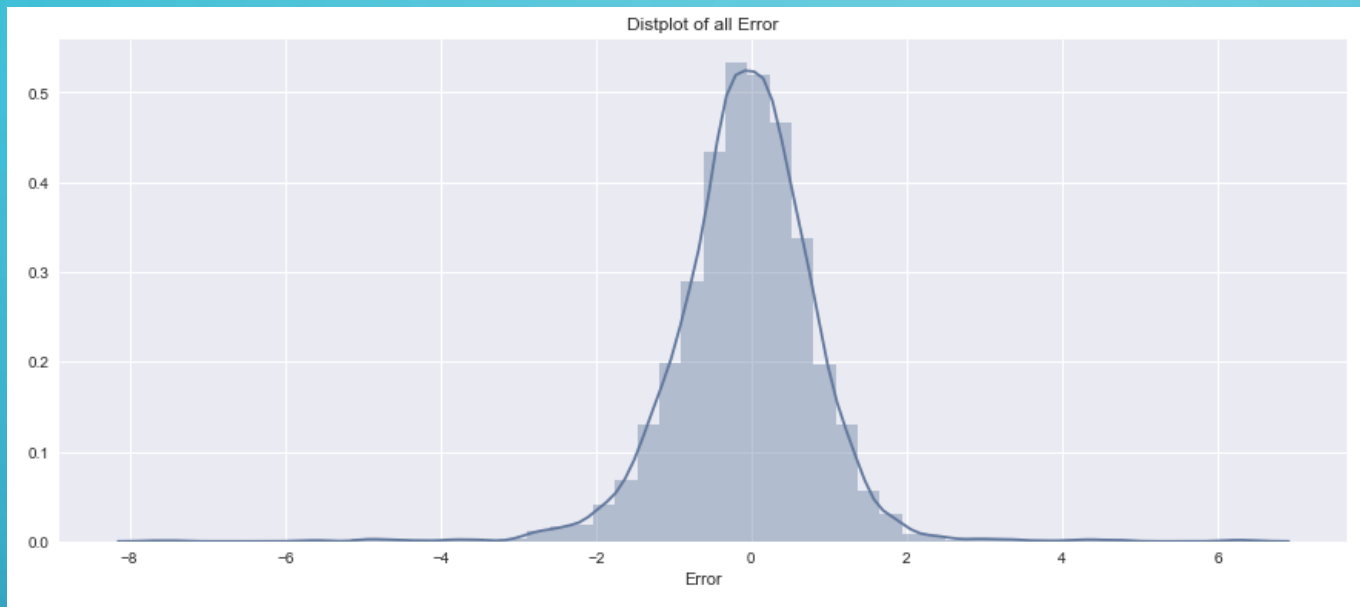
# PRODUCTION TIME SERIES – CONTROL AND PREDICTION OF PROCESS ERROR

AUSTIN BOCK

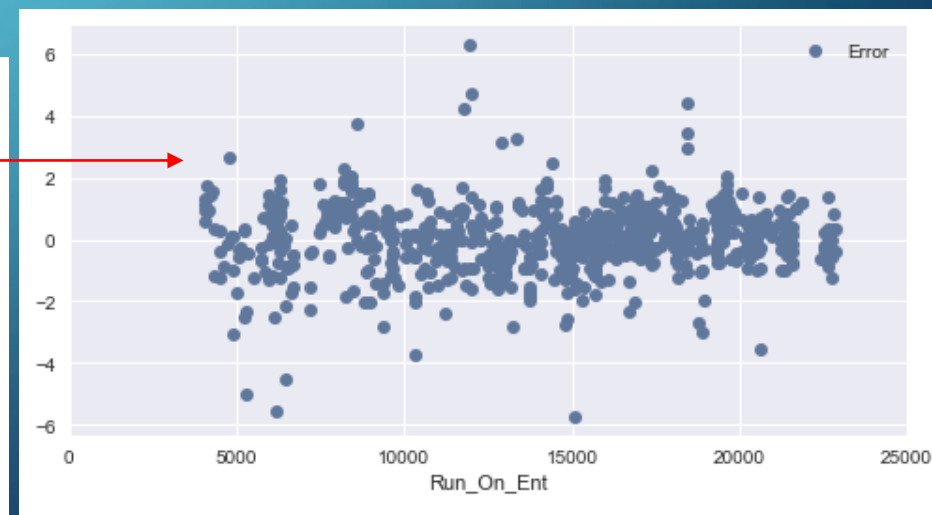
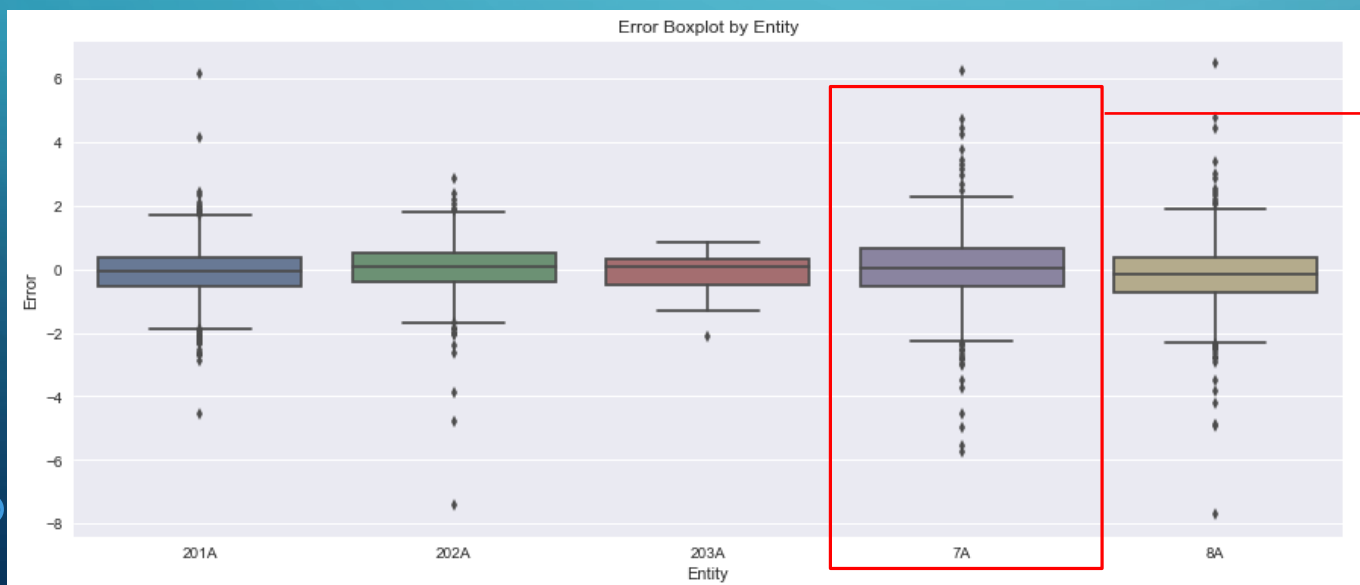
SPRINGBOARD CAPSTONE 1

AUGUST, 2017

# DATA

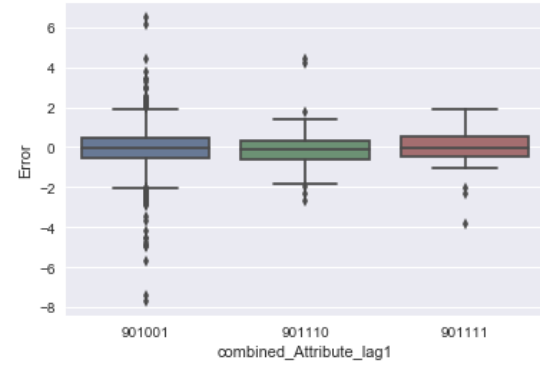


Data_Column	Count_Nulls
Error	67583
DIC_Design	65961
Fam_Type	65961
Attribute5	0
Attribute4	0
Attribute3	0
Attribute2	0
Attribute1	0
Feature4	0
Feature3	0
Feature2	0
Feature1	0
Op	0
Lot	0
Run_On_Ent	0
Entity	0

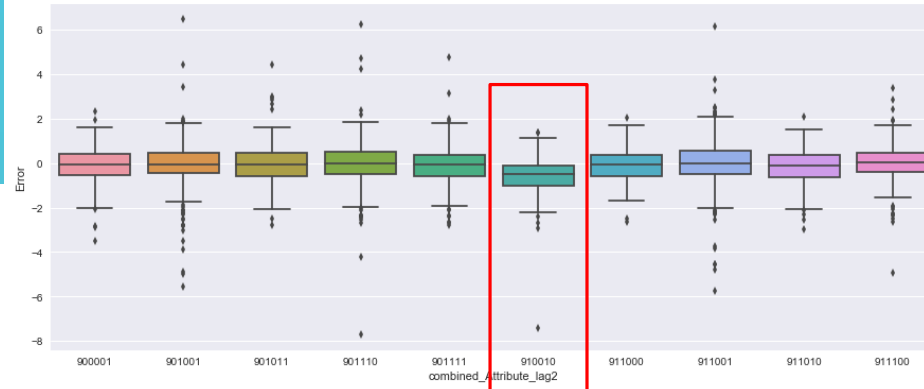


# DATA LAG SUMMARY

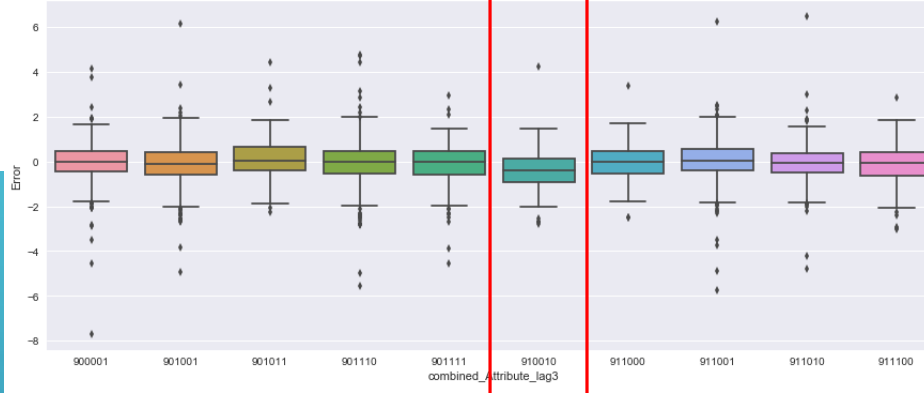
Error Boxplot by combined\_Attribute\_lag1 (n>50)



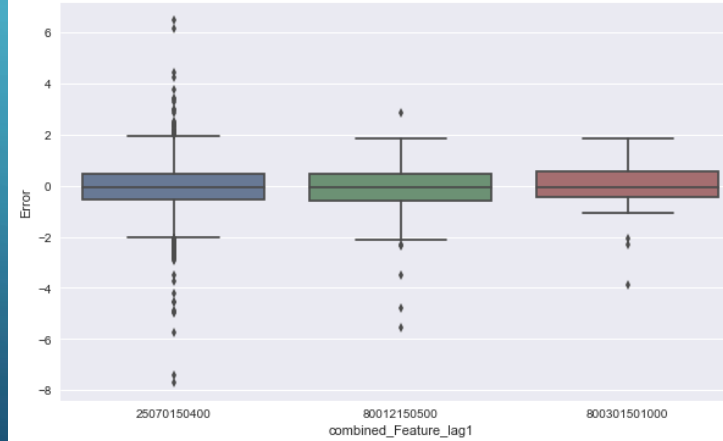
Error Boxplot by combined\_Attribute\_lag2 (n>100)



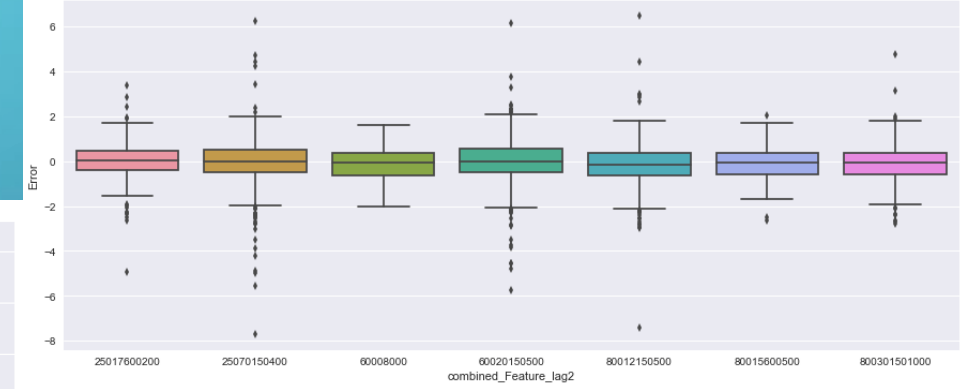
Error Boxplot by combined\_Attribute\_lag3 (n>100)



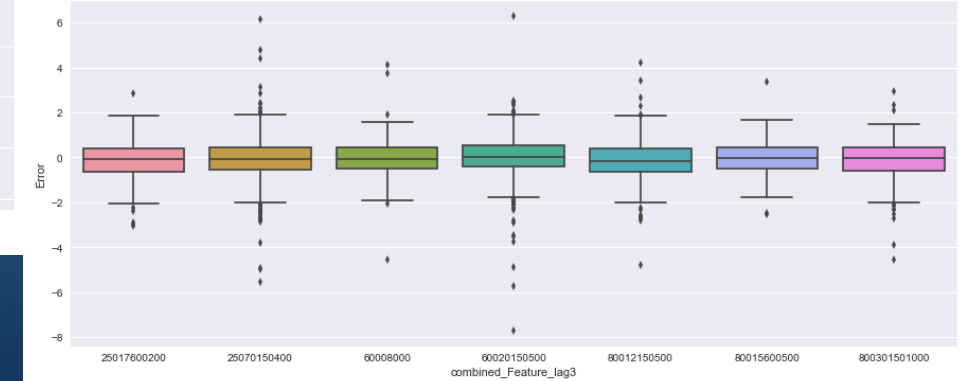
Error Boxplot by combined\_Feature\_lag1 (n>50)



Error Boxplot by combined\_Feature\_lag2 (n>100)



Error Boxplot by combined\_Feature\_lag3 (n>100)



# TUKEY HSD – 910010

```
mc = MultiComparison(CAL2['Error'], CAL2['combined_Attribute_lag2'])
|
Multiple Comparison of Means - Tukey HSD,FWER=0.05
=====
group1 group2 meandiff lower upper reject
-----
900001 910010 -0.518 -0.8024 -0.2335 True
901001 910010 -0.5299 -0.8072 -0.2526 True
901011 910010 -0.5594 -0.8673 -0.2515 True
901110 910010 -0.5653 -0.8429 -0.2878 True
901111 910010 -0.4675 -0.7425 -0.1926 True
910010 911000 0.462 0.1126 0.8113 True
910010 911001 0.5879 0.3159 0.86 True
910010 911010 0.4072 0.1115 0.7029 True
910010 911100 0.6203 0.3352 0.9054 True
-----
['900001' '901001' '901011' '901110' '901111' '910010' '911000' '911001'
 '911010' '911100']
```

```
mc = MultiComparison(CAL3['Error'], CAL3['combined_Attribute_lag3'])
|
Multiple Comparison of Means - Tukey HSD,FWER=0.05
=====
group1 group2 meandiff lower upper reject
-----
900001 910010 -0.3634 -0.6291 -0.0977 True
901001 910010 -0.325 -0.5723 -0.0777 True
901011 910010 -0.5151 -0.8247 -0.2055 True
901110 910010 -0.3872 -0.6374 -0.1369 True
901111 910010 -0.3309 -0.5951 -0.0666 True
910010 911001 0.4425 0.1761 0.7088 True
910010 911010 0.336 0.0526 0.6194 True
910010 911100 0.2823 0.0138 0.5508 True
-----
['900001' '901001' '901011' '901110' '901111' '910010' '911000' '911001'
 '911010' '911100']
```

Customer has clear outlier for one particular combined attribute (910010):

Recommendation to investigate run rules to prevent this sequence and improve overall control of process error.

# PREDICTION

Model:	Linear SGD	Linear Lasso	Linear Ridge
CV MSE score	0.9238	0.9341	0.9231
explained variance	0.0218	0	0.0234
R <sup>2</sup> coef of determination	0.0215	-0.0003	0.0232

Model:	Random Forest	SVM poly	kNN
CV MSE score	0.8714	0.8714	0.9152
explained variance	-0.1319	0.0028	0.0054
R <sup>2</sup> coef of determination	-0.1395	-0.0155	0.0053

All models indicate good fitting but very low explained variance.  
Likely cause is unaccounted for noise sources in the data.

# SUMMARY

- Goals and assumptions:

- Find any correlations between lagged data and larger |Error| which would be suitable for implementation of run rules in the production line.
- See if any models can predict Error based on previous runs through the tool.
- Key initial assumptions: other noise/error sources are not important (may/may not be true)

- Findings and recommendations:

- very strong outliers in the Error found that are suitable to perform DOE to prove run rules are indeed needed. \*Customer was able to confirm this with DOE and begin pilot to prevent occurrence
- Regressions yield models w  $\sim 0.9$  MSE and are scores are reproducible on test
  - none of the models are very good at explaining variance in the data. Conclusion is that the assumption of other noise/errors sources being unimportant is likely false. Recommendation to further improve the model is to retrieve other sources of data that can be incorporated into the model.
- Recommendation for next phase of modeling is to incorporate Heat maps of optical emission spectra for the fixed setup run (aka lag1) and each critical plate that measures Error.