

# Rice Datathon 2024

## Chevron Track

TEAM PEAK

Fernando Recinos, Clark Horak, Igor Lucic, Konstantin Iginiamre



# Data Cleaning

- Null values for “OilPeakRate” field
- Reduced columns based upon importance
- Removed outlier data

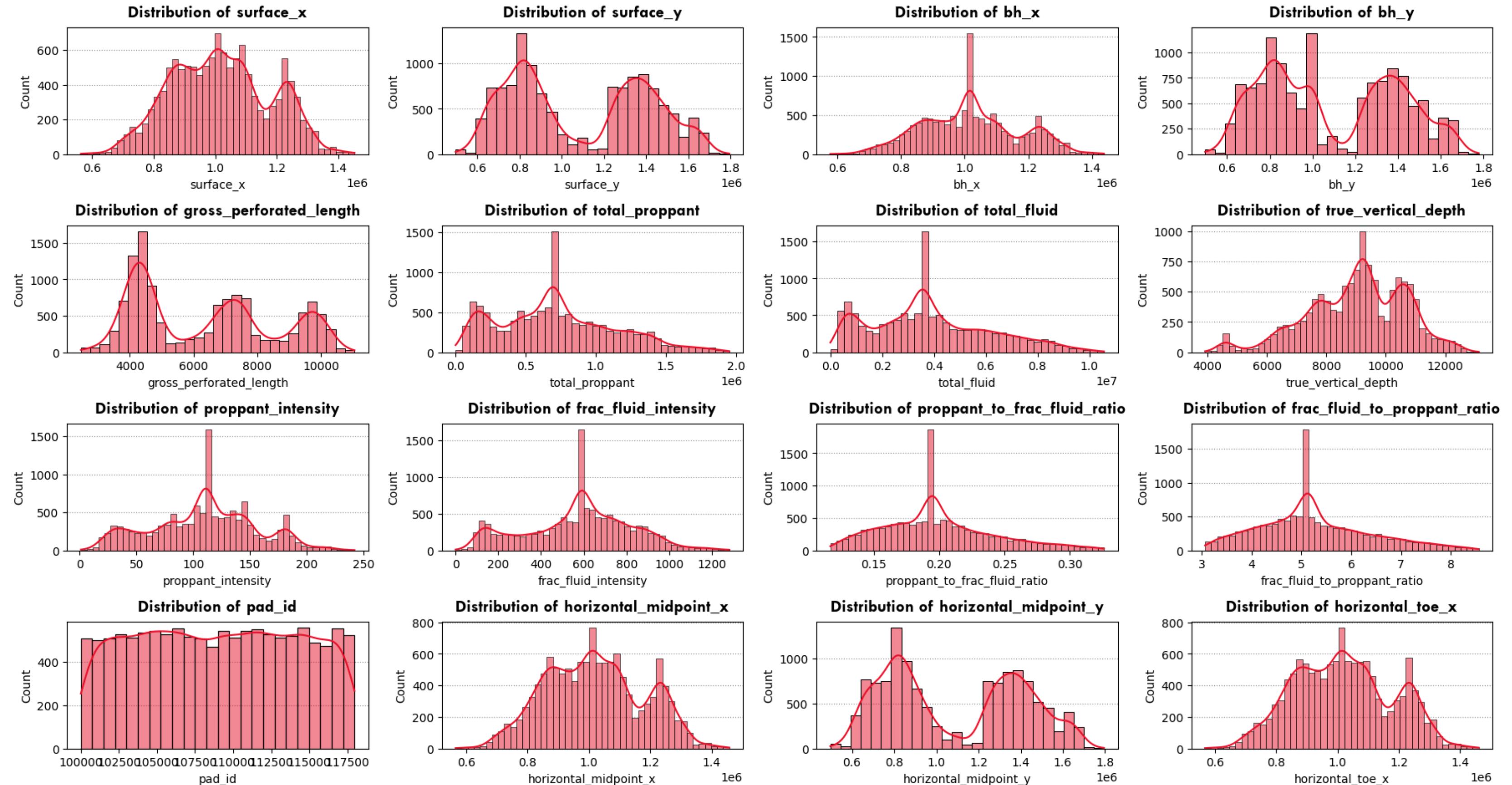
Before Cleanup

	surface_x	surface_y	OilPeakRate
0	1383493.751	717329.237	46.623
1		711916.273	59.750
2	1379915.895	717845.006	10.786
3	1383567.443	706640.201	
4	1378962.585	702623.979	102.310

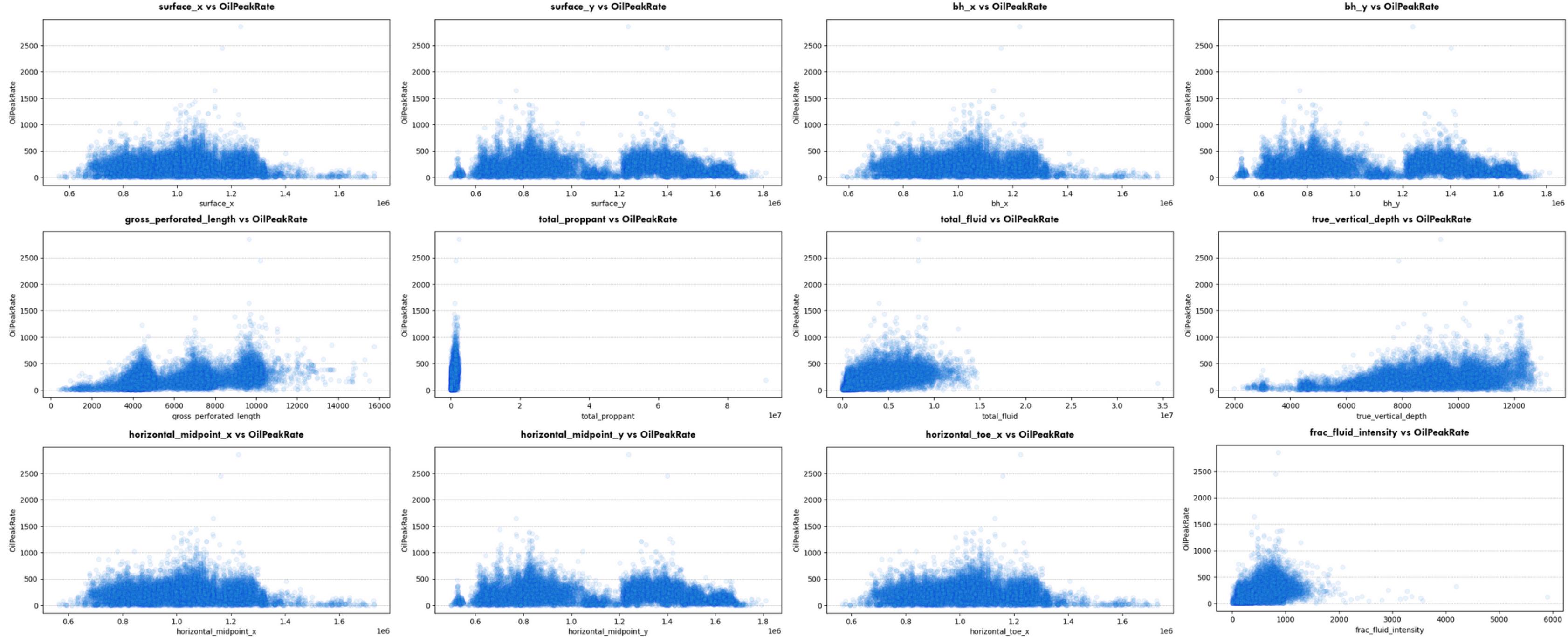
After Cleanup

	surface_x	surface_y	OilPeakRate
	1383493.751	717329.237	46.623
		711916.273	59.750
	1379915.895	717845.006	10.786
	1378962.585	702623.979	102.310

# Exploratory Data Analysis

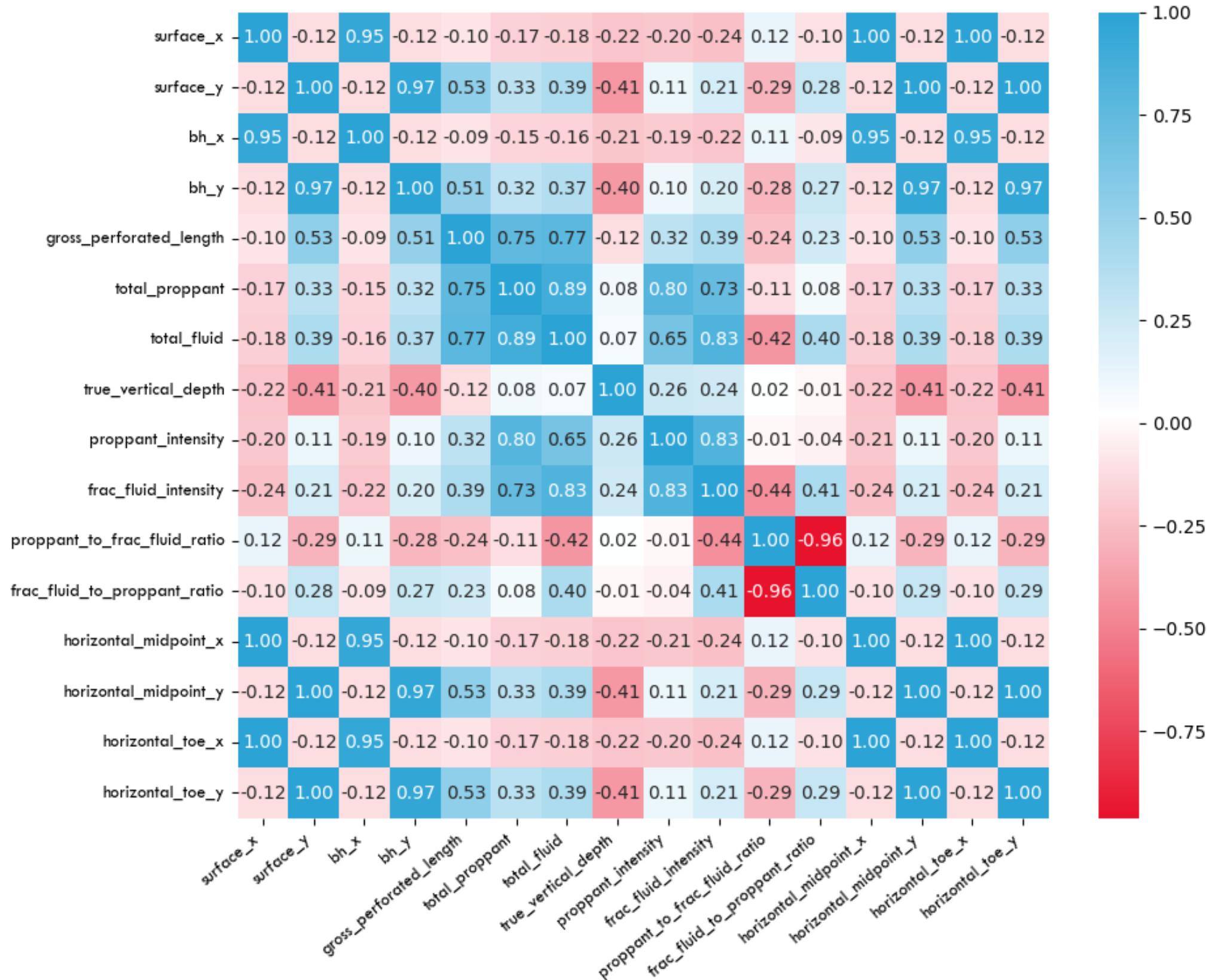


# Exploratory Data Analysis



# Exploratory Data Analysis

## Correlation Matrix Heatmap



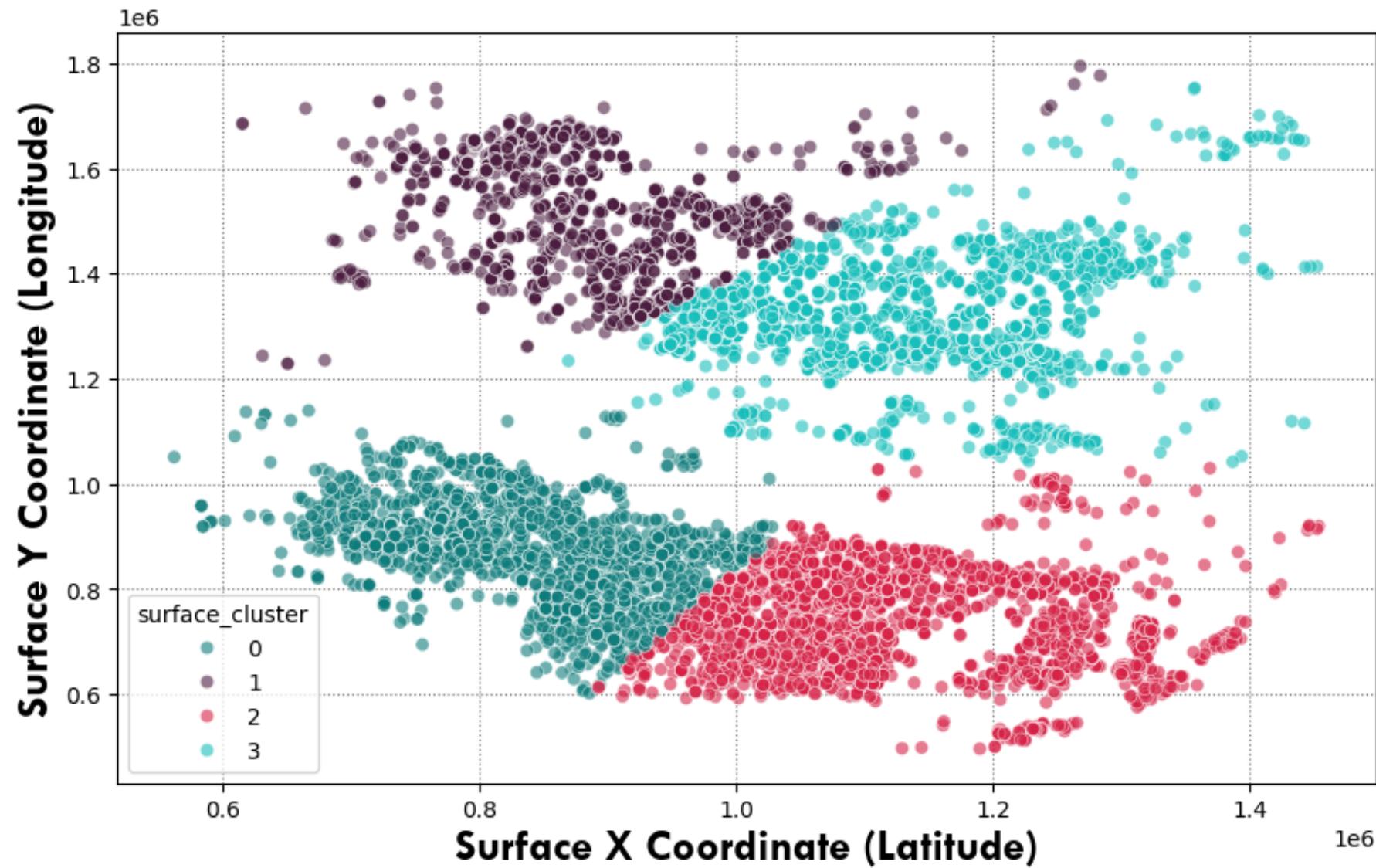
# Methods for Analysis

- Statistical learning for predictor significance on target value
- Hyper-parameter tuning for random forest model
- Clustering based on various locations
- Variable forward selection based upon p-values

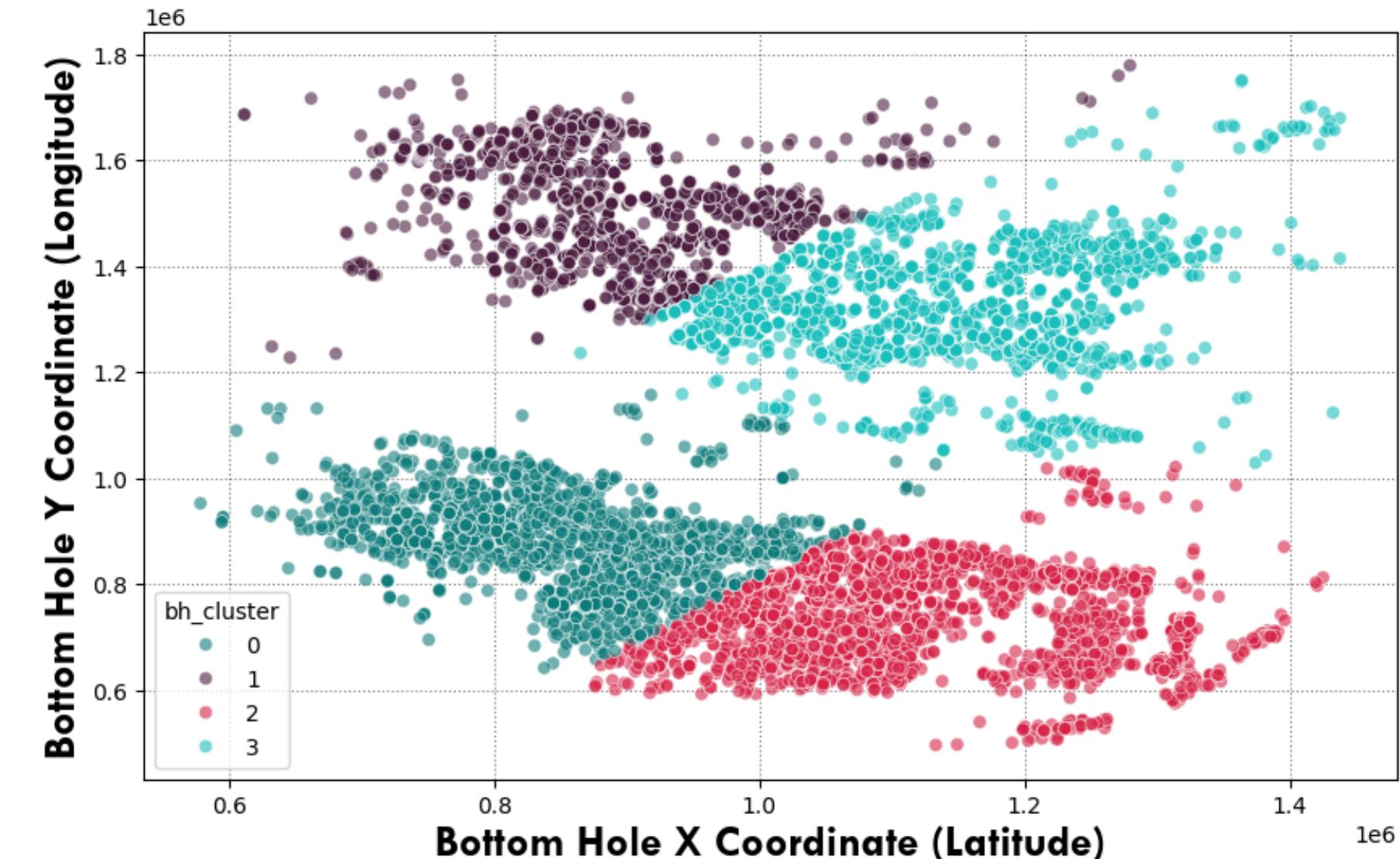
# Findings and Analysis

Redundancy within coordinate-based variables

**Surface Clusters**



**Bottom Hole Cluster Locations**



# Findings and Analysis

Reducing complexities based on various statistical values (coef, t-value, etc.)

OLS Regression Results						
Dep. Variable:	OilPeakRate	R-squared:	0.370			
Model:	OLS	Adj. R-squared:	0.369			
Method:	Least Squares	F-statistic:	430.6			
Date:	Sun, 21 Jan 2024	Prob (F-statistic):	0.00			
Time:	14:30:34	Log-Likelihood:	-75187.			
No. Observations:	12490	AIC:	1.504e+05			
Df Residuals:	12472	BIC:	1.505e+05			
Df Model:	17					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	-220.8576	35.561	-6.211	0.000	-290.562	-151.153
surface_x	0.0001	6.1e-06	17.807	0.000	9.67e-05	0.000
surface_y	-6.832e-05	3.91e-06	-17.475	0.000	-7.6e-05	-6.07e-05
standardized_operator_name	0.0339	0.008	4.013	0.000	0.017	0.050
gross_perforated_length	0.0279	0.002	15.257	0.000	0.024	0.031
total_proppant	-5.812e-05	1.35e-05	-4.290	0.000	-8.47e-05	-3.16e-05
total_fluid	6.782e-06	2.51e-06	2.698	0.007	1.85e-06	1.17e-05
true_vertical_depth	0.0190	0.001	28.067	0.000	0.018	0.020
ffs_frac_type	0.8893	0.873	1.019	0.308	-0.821	2.600
proppant_intensity	0.6191	0.104	5.962	0.000	0.416	0.823
frac_fluid_intensity	0.0454	0.020	2.230	0.026	0.005	0.085

# Predictive Accuracy

High RSME with Low Overfitting

Testing RMSE: 79.66243489776772

Actual	Predicted
202.557968	258.906587
479.709258	386.224862
233.678608	279.396442
76.353328	57.582255
409.607206	373.773184

TEAM PEAK

Rulz!

Thank u 4 watching :)

