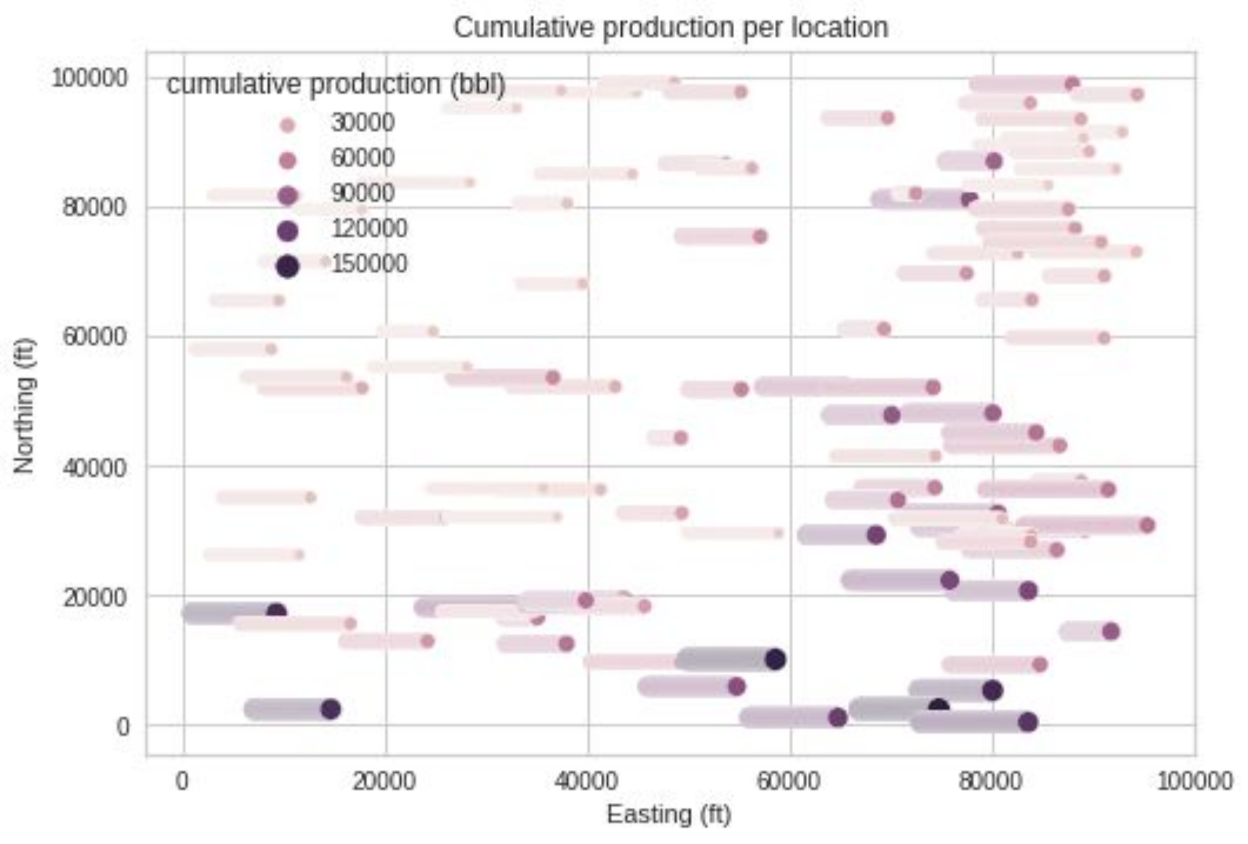
Hingle Basin



Why does our problem matter?

- Optimizing new drilling locations allows for more efficient investment.
- Optimal drill locations also reduce cost and improve access for customers.

Our First Attempt (computation)

1. Preprocessing our Data

The biggest step in our path was processing the data into a form that could be used to train a machine learning model.

2. Create a Prediction Model for Production

After preprocessing our data, we trained a ExtraTrees Regression Model which gave predictions with upwards of 65% accuracy.

3. Decide on a strategy for computing new well permutations

After our model was created, the next step was to find which parameters would be best predicted with our location and rock property data.

4. Choosing the 10 Wells

After choosing which parameters to predict and which to iterate through, we were able to use the model we created in step 2 to predict the 10 best locations for our wells.

First Method

- After preprocessing, we had our model predict the natural log of our cumulative production.
- Using an ExtraTreesRegressor, we achieved a Score of 0.83 in our testing data set.
- From these data points, we decided to create a database of roughly 2 million permutations of our regression's hyperparameters. We created multiple regressors to predict the rock properties at any northing and easting value.
- Each permutation was stored as a row within an SQLite database.

Budgeting Computer Resources

We made a tier-list showing which features were most/least important:

Tier 1

- pr - pump rate **Iterated** 32 iteations
- ppf - proppant per stage **Iterated** 32 iterations

Tier 2

- easting **Iterated** 40 iterations
- perm **Predicted**

Tier 3

- northing **Iterated** 25 iterations

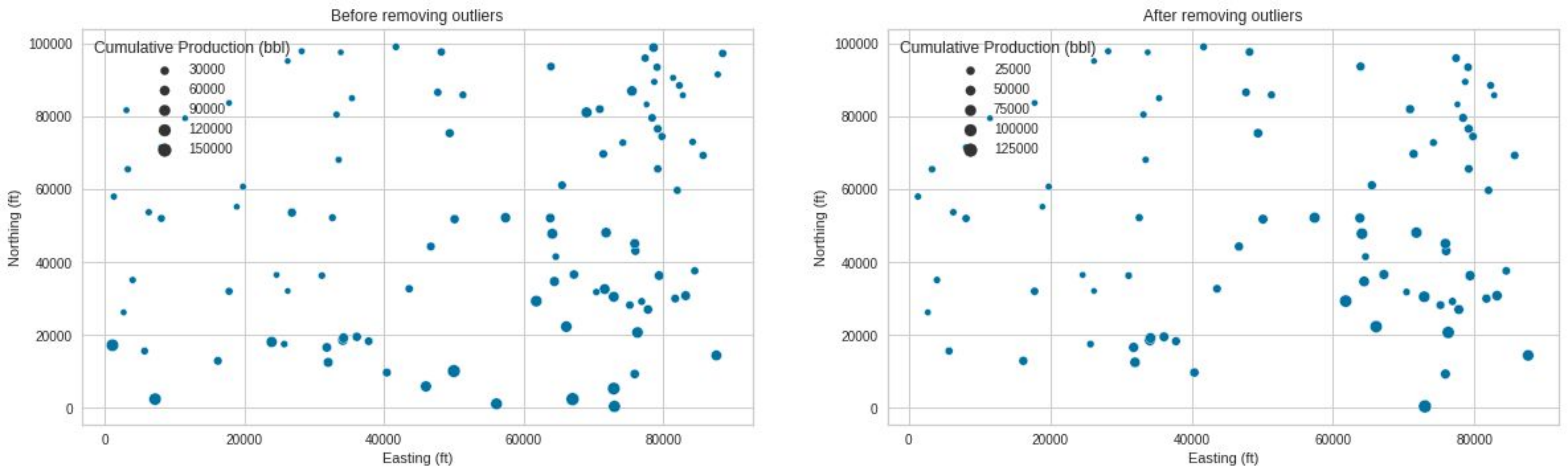
We based our iterations on the relative importance of each hyper-parameter.

First Attempt Obstacles

- We decided to implement automated decline curve analysis and create an exponential and hyperbolic equation for production. This meant that we had to integrate over our chosen equation rather than use the provided formula.

- In an attempt to improve our regressors' accuracy and computate better synthetic data we found that we needed to remove outliers.

- When computing new data we ran into memory limitations that caused us to use an SQLite database rather than a pandas dataframe to store our two million predicted permutations.



First Method Results

- We naively believed that the errors in estimating rock properties (around 0.98 accuracy) would be negligible when predicting cumulative production using our synthetic data.

- Depending on which types of regressors we chose, our model either overpredicted or unpredicted cumulative production by orders of magnitude.

- No matter how we tried to tune our regressors, we could not eliminate our "compounded-error" problem. After further research, we think this is an intrinsic problem of a one-step model.

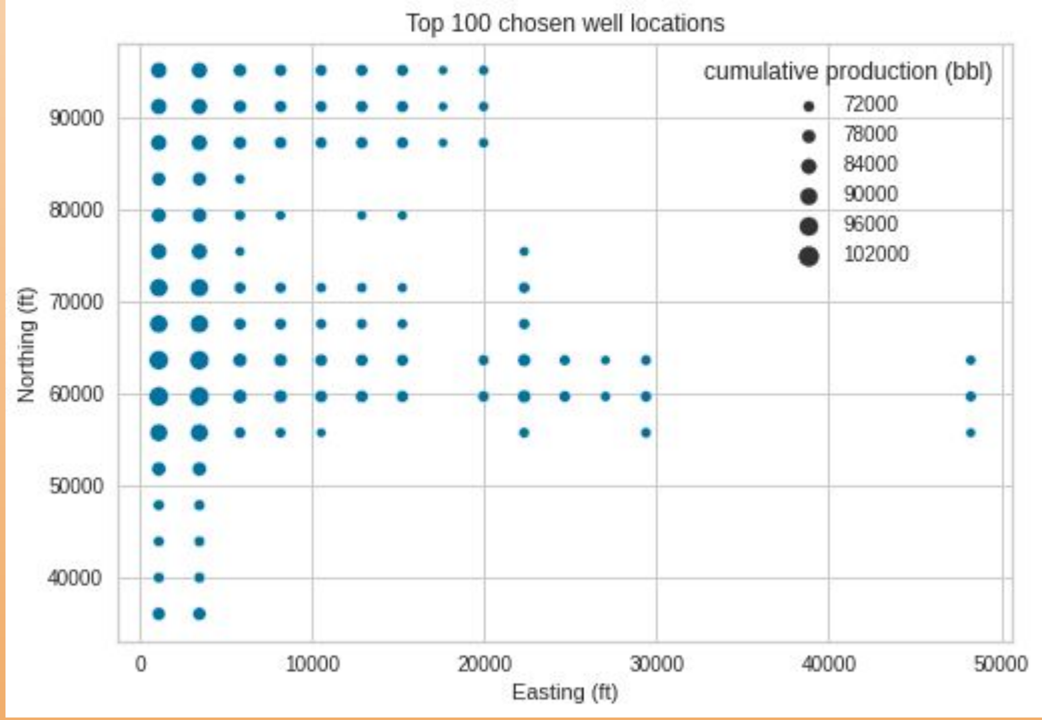
Second Method

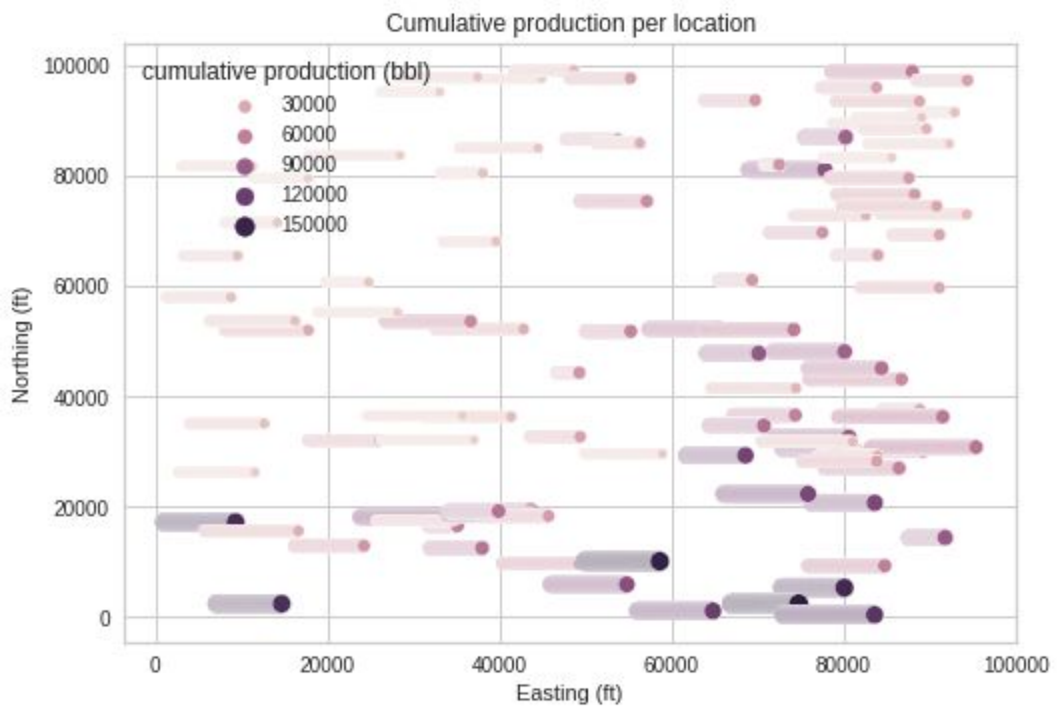
- To fix our compounded error problems, we trained our regressors to predict cumulative output using only the easting and northing features. With this method, we were able to get an accuracy of ~0.60 using a tuned RandomForestRegressor.

- Then, we utilized our regression to predict the cumulative output for all of our candidate wells and chose the top 10 that were at least two thousand feet away from one another.

Top 100 New Well Locations

These clearly did not match our results from our given data.

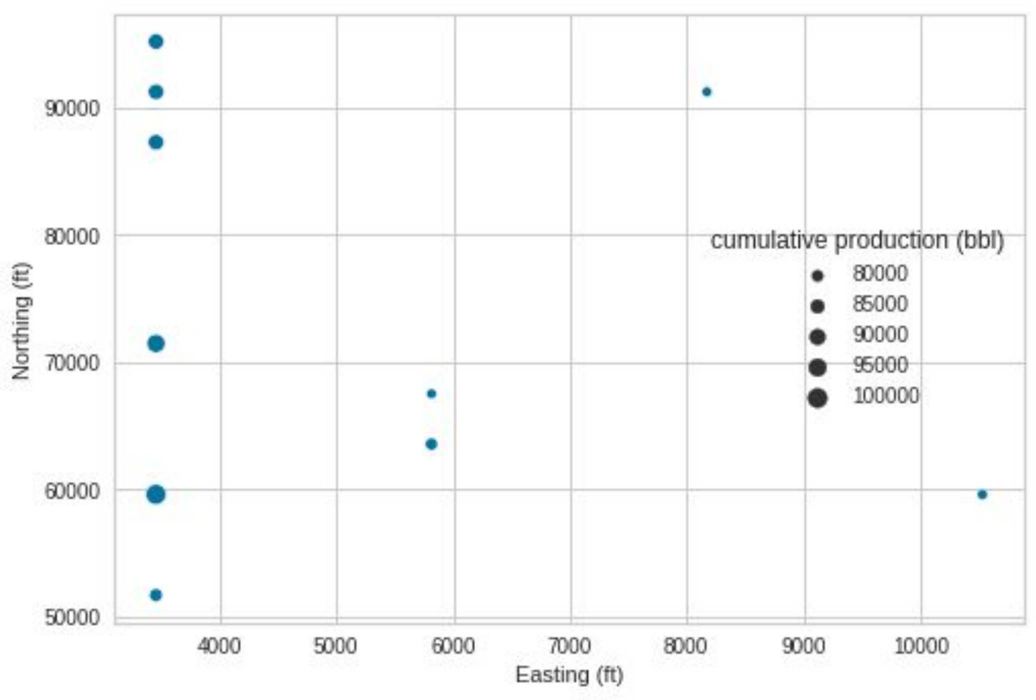




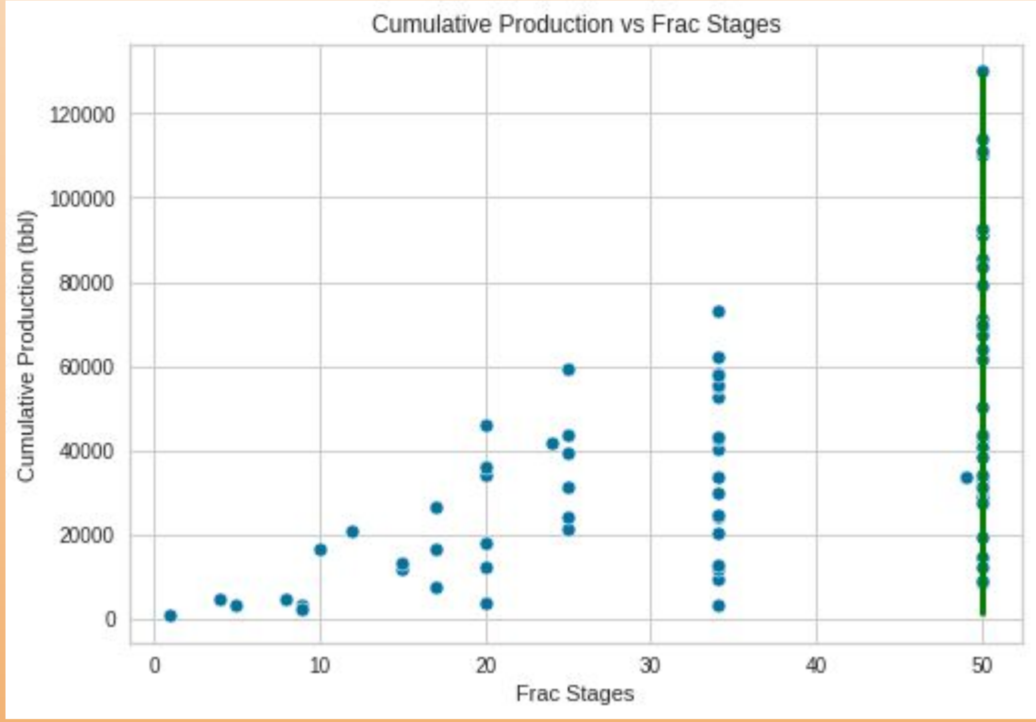
10 New Well Locations

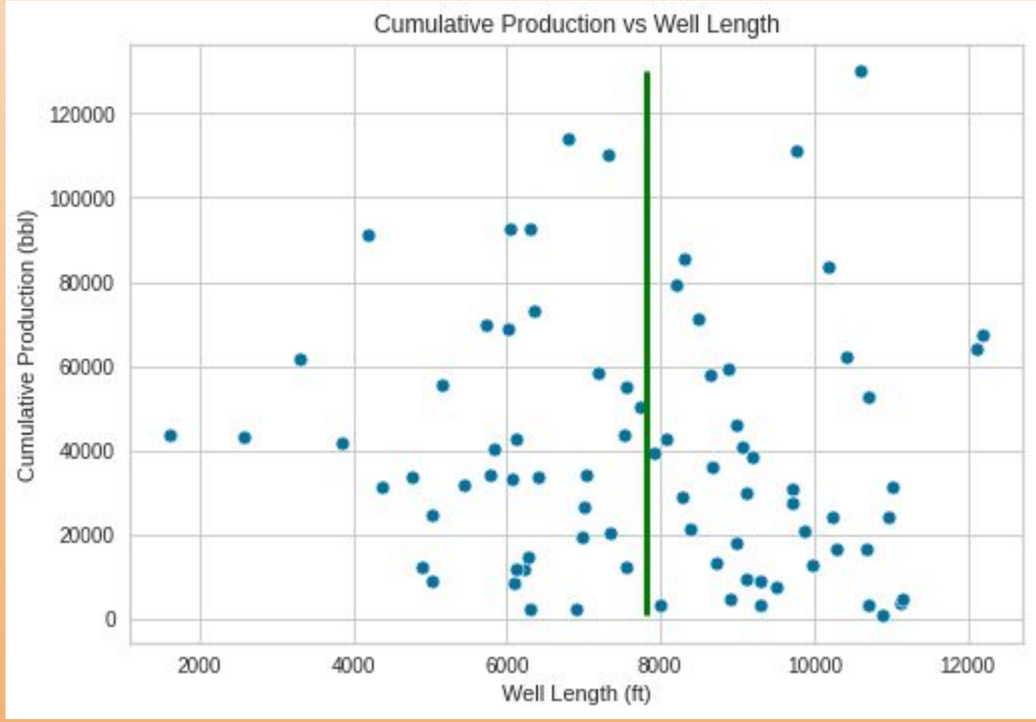
To choose our 10 locations we picked values within our top 40 predicted cumulative productions that were well-spaced from one another.

Easting (ft)	Northing (ft)	Pump Rate (cft/min)	Amt. Of Proppant (lbs)	# Frac Stages	Well Length (ft)	OOIP (bbl)	RR/EUR (bbl)
3457	59620	399	1.38E+06	50	7800	1148827.64	102505.08
3457	71464	399	1.38E+06	50	7800	1024674.50	96963.46
3457	87256	399	1.38E+06	50	7800	768876.52	89427.80
3457	91204	399	1.38E+06	50	7800	796789.30	89427.80
3457	95152	399	1.38E+06	50	7800	774819.86	89427.80
3457	51724	399	1.38E+06	50	7800	1736293.53	83504.37
5814	63568	399	1.38E+06	50	7800	978310.16	82293.42
10528	59620	399	1.38E+06	50	7800	1044814.20	78666.86
5814	67516	399	1.38E+06	50	7800	1026266.09	78002.65
8171	87256	399	1.38E+06	50	7800	701426.91	77946.59

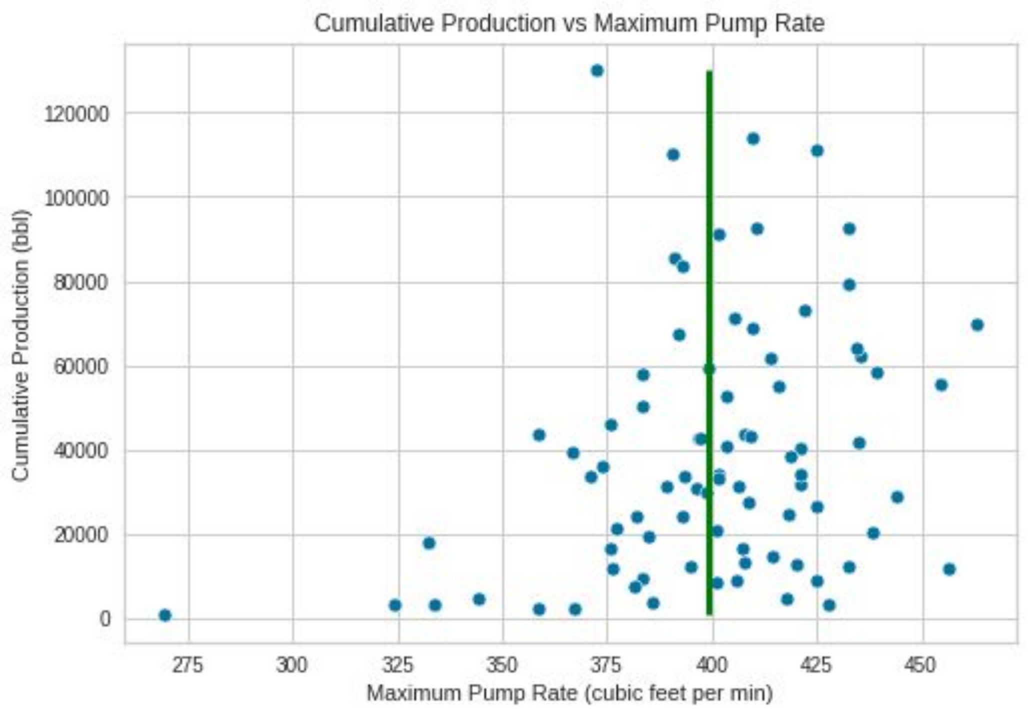


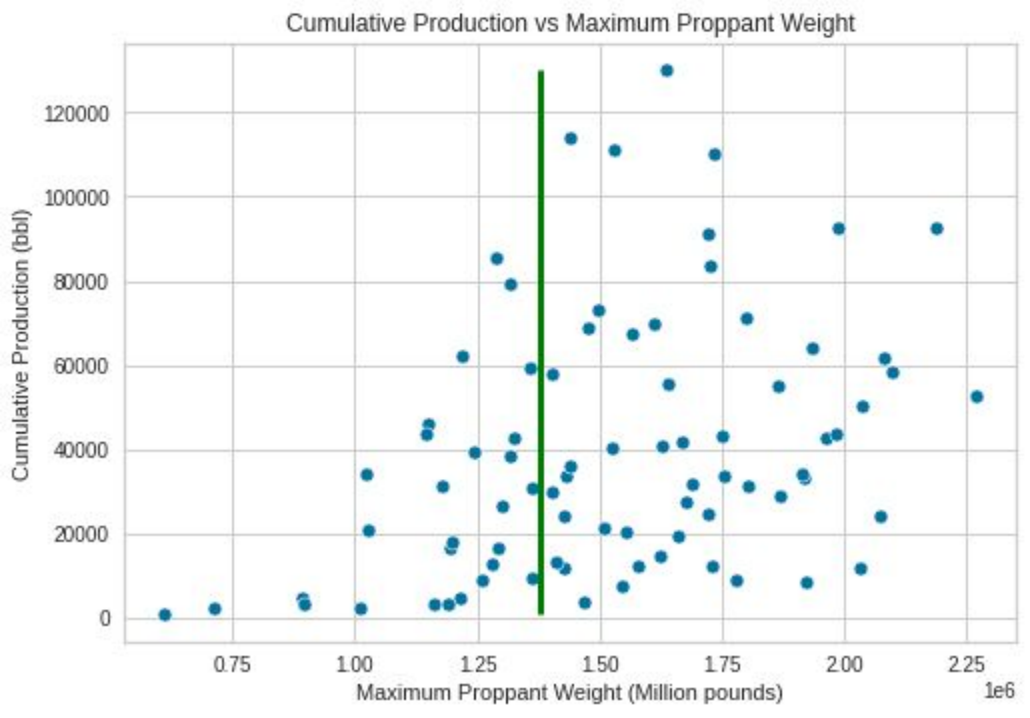
Feature Selection





Feature Selection





Experience Gained

- Managing memory and processing speed hardware limitations through integration with SQLite3.
- Implementing automated decline curve analysis.
- New knowledge of the limitations that come from interpolating synthetic data with a one-step model.

Our Improvement Ideas

We think that implementing a multi-step model instead of our single-step model would fix our compounding error problem and allow us to look at more features when predicting cumulative output.