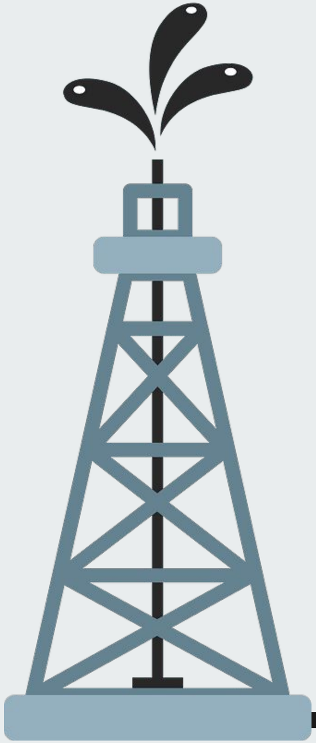


Petro Phreaks

Prediction of Peak Oil Production Rate

Challenge:

Chevron decision-makers are often required to optimally plan the development of wells, based on their expected performance in oil production. These decisions are expensive in nature and can influence Chevron’s standing amongst its competitors.

Our Goal:

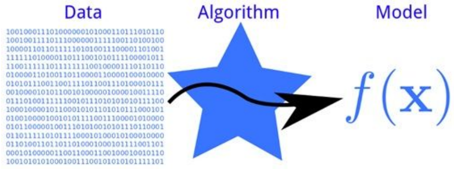
We aim to provide an accurate method of predicting a well’s peak oil production rate in order to help Chevron make data-informed , optimal decisions.

Process:

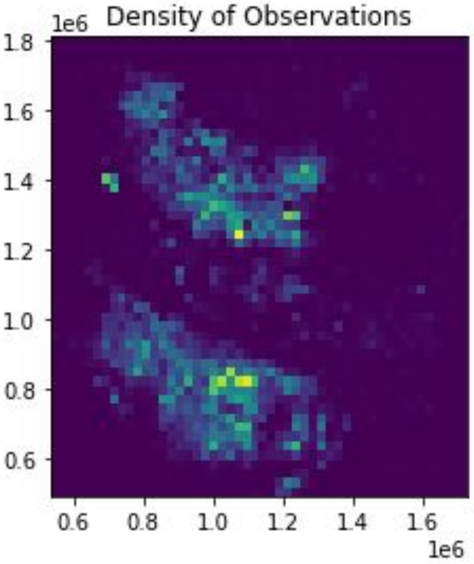
1. Data Wrangling
   a. How did we decide which columns and rows to keep?

2. Data Exploration
   a. What behaviors does the data exhibit?

3. Model Building
   a. What model will perform most accurately, defined by the lowest RMSE?



Density of Coordinate Data

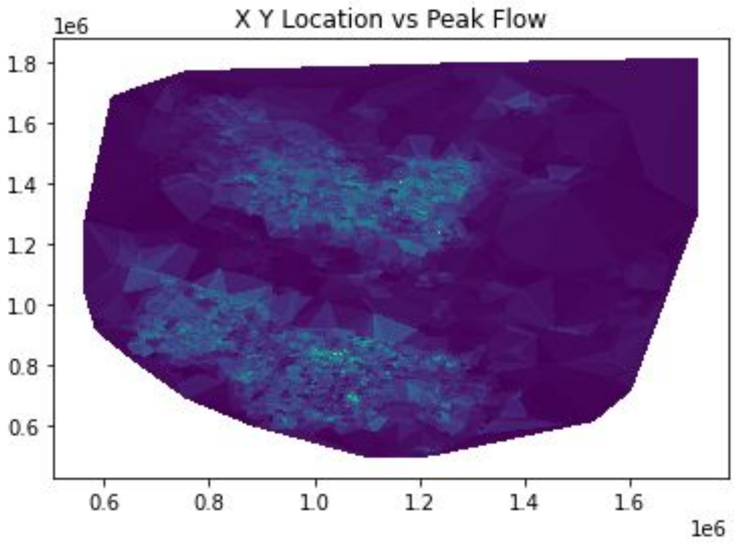


Histogram showing the X Y coordinates in the data set.

Insights:

- Two main clusters visible on Y axis
- We cannot standardize coordinates using a mean

Peak Flow Against Location Data



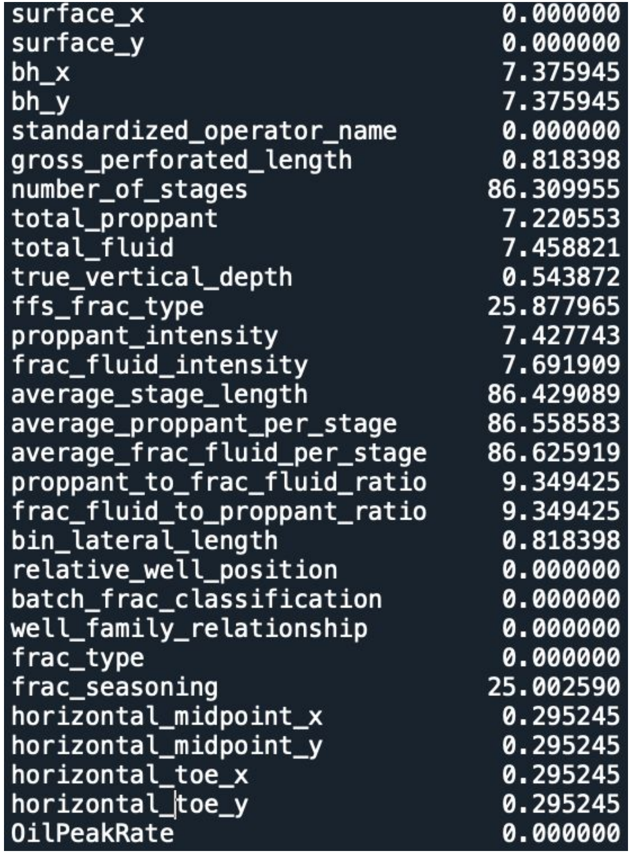
Graph showing Peak flow at each coordinate.

Insights:

- Two main regions with higher peak flow, with some noise
- Coordinates are important factor in peak flow

Data Wrangling Process

- Removed columns that had over 25% of observations with NANs
- Removed all observations with NAN for Oil Peak Rate
- Removed frac_fluid_to_proppant_ratio with value of infinity
- Scaled X and Y based normalizing across min and max
- Standardised remaining based on deviations from variable mean
- Used Dummy variables to represent categorical variables

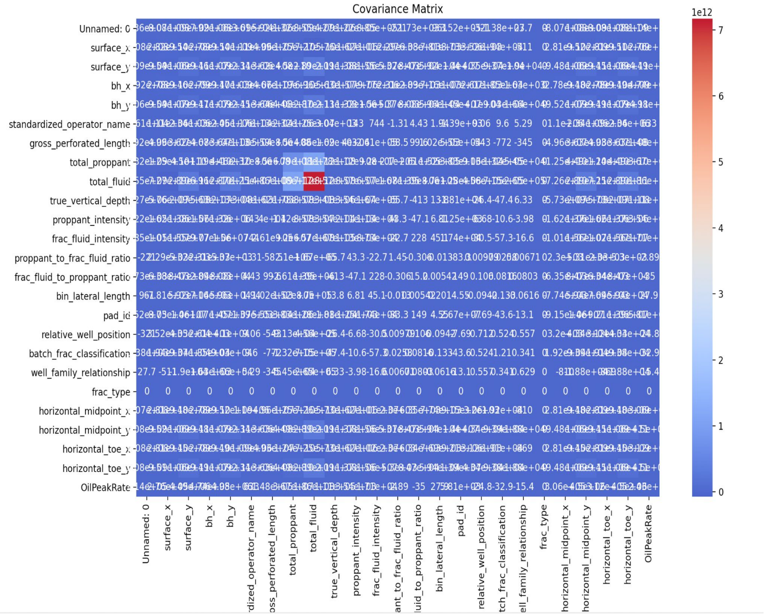


```
surface_x                         0.000000
surface_y                         0.000000
bh_x                              7.375945
bh_y                              7.375945
standardized_operator_name        0.000000
gross_perforated_length           0.818398
number_of_stages                 86.309955
total_proppant                    7.220553
total_fluid                       7.458821
true_vertical_depth               0.543872
ffs_frac_type                    25.877965
proppant_intensity                7.427743
frac_fluid_intensity              7.691909
average_stage_length             86.429089
average_proppant_per_stage       86.558583
average_frac_fluid_per_stage     86.625919
proppant_to_frac_fluid_ratio      9.349425
frac_fluid_to_proppant_ratio      9.349425
bin_lateral_length                0.818398
relative_well_position            0.000000
batch_frac_classification         0.000000
well_family_relationship          0.000000
frac_type                         0.000000
frac_seasoning                   25.002590
horizontal_midpoint_x             0.295245
horizontal_midpoint_y             0.295245
horizontal_toe_x                  0.295245
horizontal_toe_y                  0.295245
OilPeakRate                       0.000000
```

Variance Plots

- Top 5 Covariances Pairs
    - 'total_proppant', 'total_fluid'
    - 'total_fluid', 'horizontal_midpoint_y'
    - ‘total_fluid', 'horizontal_toe_y'
    - 'surface_y', 'total_fluid'
    - ‘bh_y', 'total_fluid'

- 5 Features with Highest Variance with OilPeakRate
    - total_fluid: 1.860112e+08
    - total_proppant: 3.667437e+07
    - bh_x: 4.735879e+05
    - horizontal_toe_x: 3.118518e+05
    - horizontal_midpoint_x: 3.056808e+05



Model Building:

In our model building process, we developed the four distinct models:

Gradient Boosted Regression:
Mean RMSE: 102.21

Neural Network:
Mean RMSE: 141.95

Linear Regression:
Mean RMSE: 119.49

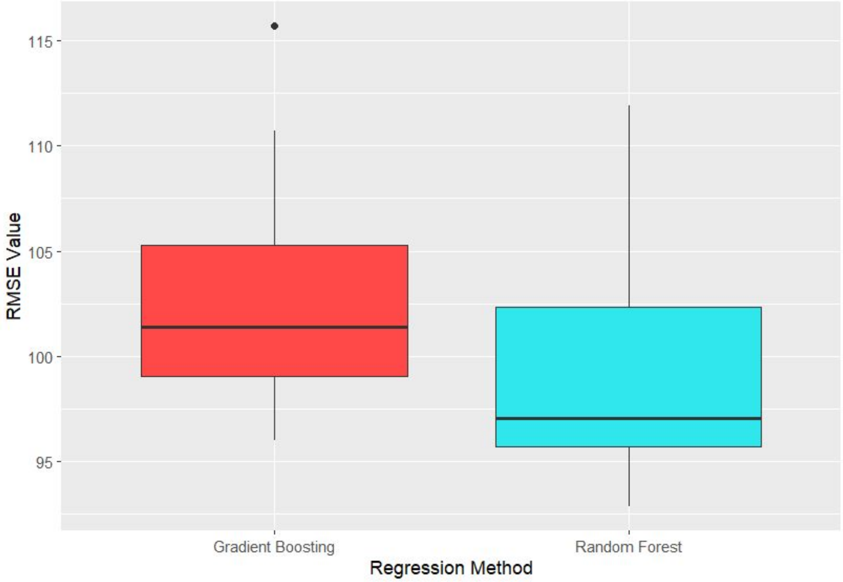
Random Forest Regression
Mean RMSE: 98.82

Hypothesis Testing

H_0: $RMSE_{rf} = RMSE_{gb}$ H_A: $RMSE_{rf} \neq RMSE_{gb}$

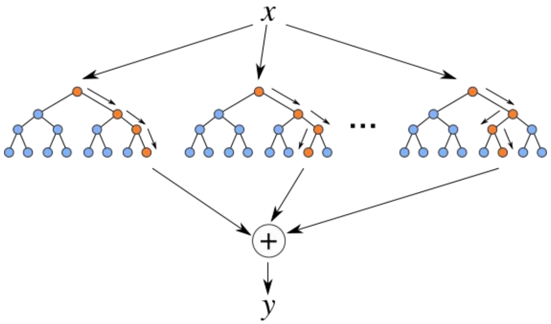
- Results
  - Welch Two Sample t-test on difference of means
  - p-value = 8.51×10^{-5}
  - 95% confidence interval: [1.76, 5.05]

RMSE Value Achieved vs. Regression Method



Conclusion

- Random Forest Regressor was optimal model
- Hypothesis test showed Random Forest consistently performed better than other models
- Potential Future Investments:
  - Feature engineering
  - More complete dataset
  - Different geographical factors (seismic reports, etc.)



Thank You!