



Synthesis of DJ Basin Horizontal Play Performance - A Data Science Approach

by Laura Elliott, Crossroads Geoscience Consulting

And Ajith Patnaik, QuadPay

Springboard Data Science Career Track
Capstone I Project / 2021



Outline

- Problem Definition
- Data Acquisition
- Data Wrangling
- Exploratory Data Analysis
- Data Preparation and Evaluation
- Feature Engineering
- Feature Selection
- Modelling - Machine Learning
- Conclusions
- Next Steps



Problem Definition

- Data Acquisition and Software Costs continue to increase
- Can we build a workflow using only publicly available data and open source software only to perform a typical Exploration & Production problem?
- What is the average well lifecycle span of a horizontal well in the DJ Basin?
- Can we accurately predict likelihood of future abandonment?

Data Acquisition

- Primary Data Used From [Colorado Oil & Gas Conservation Commission \(COGCC\)](#)
 - 21 Years of Yearly Production Reports (.csv files) 1999-2020
 - Shapefiles to Provide Well Attributes, Geometry, and Location
 - Well Surface Locations
 - Directional Bottomhole Locations
 - Directional Lines
- State and County Boundaries – [GADM.org](#)
- PLSS Township and Section Boundaries – [Bureau of Land Management \(BLM\)](#)
- Open-Source Software
 - [QGIS](#)
 - Python libraries - GeoPandas



Raw Data Attributes

- Update this



Data Wrangling

- Work performed in the following Jupyter notebooks – add links to notebooks
- Load, quality check, correct, and re-format / merge
- Iterative Process
- 90% of effort
- Data Issues Identified and Corrected:
 - Typographical Errors - Outliers
 - Missing Values – Ground Elevations
 - Duplicated Production Records and Records Assigned to N-COM
 - Missing Months of Known Production
 - Inconsistencies Between Data Sources – API Formatting, Status,
 - Associating the Correct Boreholes Between Datasets



Outliers

- Typographical Errors



Check for Missing Values

- Missing Values – Ground Elevations
- Missing Months of Known Production
- Production Records Assigned to N-COM



Removing Duplicate Rows

- Duplicated Production Records



Exploratory Data Analysis

- Used to quality check the data and id outliers
- Statistical Insights
 - Identify correlations and trends
 - Production Characteristics
- Build basis for prediction of future well behavior
- Plots created in Jupyter Notebooks using Python with the following libraries:
 - Matplotlib
 - Plotnine ggplot



Data Preparation / Evaluation

- Target Analysis – Sampling Strategy
- Scoring metric



Feature Engineering

- # of features
- List of features



Feature Selection

- Yada yada yada



Modelling – Machine Learning

- Baseline Modelling Binary Classification Type
- Model types employed
 - Random Forest
 - Ensemble decision tree / bagging technique



Random Forest Results

- Work performed in



Conclusions

- The success of this projects proves that publicly available data /open source software is a viable approach to E&P projects
- Expect a considerable effort on quality checking and correction as new data comes in



Next Steps

- Best practice if this were an on-going project would be to save corrected historical data in a SQL database
- Update periodically with only new data
- Incorporate more geologic and engineering data to improve predictions