

2024 Chevron Rice Datathon: Predicting Peak Oil Production Rate

Description

The production of oil, gas, and/or water, at either the individual well or across an entire field, varies over time. The graphical description of this production across time is known as a production profile. Having a robust qualitative and quantitative method for understanding and predicting production profiles for both our operated wells and our competitors' wells is essential for making informed decisions during the development of our assets.

In our unconventional assets, e.g. the Permian Basin or Haynesville Shale, wells tend to exhibit a particular type of production profile, and this production profile can generally be broken up into three parts:

1. The ramp up – the time from when the well is put on production (popped) until it reaches a peak production rate.
2. The peak – the point at which the well reaches its peak production rate.
3. The decline – the point after peak until the well is plugged and abandoned. Given enough time, all wells will exhibit declining production rates.

When planning the development of an asset, asset development teams are often faced with decisions that rely on making predictions of how wells will behave prior to them being popped. In challenge, you are tasked with developing methods to predict the oil production rate at peak.

Data

DISCLAIMER: The data provided for this challenge can be used only to develop a model for the challenge and should be discarded after the end of the challenge.

The challenge involves several files for your project. The files required to build and test your predictive model will be made available at the start of the Datathon. However, 30 minutes before the judging period starts, we will also release a scoring file that contains additional data for which you will run your final model and provide corresponding predictions. The following files will be available at the start of the Datathon:

- **training.csv:** contains raw data used to build, train and test your model.
- **data_dictionary.xlsx:** provides description of the variables in training.csv
- **submission_format.csv:** gives the required format of the model predictions file obtained by applying your model to the scoring.csv file.

The following file will be available 30 minutes before the Datathon judging begins:

- **scoring.csv:** used to generate model predictions required for us to score your model accuracy

- **submission_file.xlsx**: used to submit model predictions and calculate model accuracy. Input your model predictions into the “Predictions” sheet. After predictions are inserted, your model’s predictive accuracy will be calculated automatically so ***do not make any changes to the “RMSE” sheet.***

Project Submission

Please submit your Datathon project for judging. At minimum, the submission should include a link to a github repo that contains:

- your project code
- a summary of your methodology and findings
- submission_file.xlsx: the file in which you inserted the predictions outputted from running your model on the **scoring.xlsx** dataset

Evaluation Criteria

Projects will be evaluated holistically based on model accuracy the criteria outlined by Rice Datathon:

- **Data Wrangling**: How do you clean, manipulate, and process the data to make sure the final data used in the model is clean?
- **Visualization**: How do you use visualizations for exploratory data analysis and to communicate your findings?
- **Findings and Analysis**: How do you interpret the findings of your project? What analysis do you perform on your results?
- **Technical Skill**: Does your work use innovative and appropriate models, methodologies, or approaches?
- **Predictive Accuracy**: predictive accuracy will be measured by the Root Mean Squared Error (RMSE) in peak oil production rate predictions for the wells in the scoring.csv file:

$$RMSE = \sqrt{\frac{1}{n} * \sum_i (\hat{y}_i - y_i)^2}$$

Other areas that can be looked at include, but are not limited to feature generation, feature selection, and model selection/building.

Rules

1. Contestants must respect the privacy of the data and remove it from their computers upon completion of the competition.
2. Contestants' solution must be a model that is repeatable, adjusting model results manually (including “arbitrary” factors/constants) to tune model predictions is not permitted.
3. Contestants' prediction models can use only that data from the training.csv dataset provided to build the model.

Prizes

FIRST PLACE: Apple iPad Mini (Latest 6th Generation) 64GB, Wifi-6, All-Day Battery Life – Assorted colors for each winning team member (Space Gray, Pink, Purple, Starlight)

SECOND PLACE: Physical \$100 Amazon Gift Cards in Premium Gift Box for each winning team member