# Introduction:

The primary objective of this data science project is to develop a comprehensive predictive model aimed at forecasting the development and production of oil, gas and water of a given basin.

# Approach:

The proposed approach involves developing a rule-based model that integrates rig data, historical information, and type curve analysis to generate diverse cycle times and production dates for inventories, new wells and wip wells within the target basin.

Drillout Model contains 4 sub – models:

# Cycle Time Model:

# Purpose:

The Cycle Time Model aims to analyze historical data and extract distinct cycle times for each well, both at the operator level and the type curve level. It helps in understanding the time taken for various stages of drilling and completion operations.

the Cycle Time Model determines the time duration for the following stages for each operator within each Type Curve area by analyzing historical data:

1. **Time taken from permit submission to permit approval:**
   * This stage involves the duration between the submission of permits for drilling operations and their approval by regulatory authorities. It encompasses the regulatory review process and any associated administrative tasks.
2. **Time taken from permit approval to spud:**
   * After obtaining permit approval, this stage represents the duration between the approval date and the commencement of drilling operations (spud). It includes logistical preparations, mobilization of equipment, and site-specific activities.
3. **Time taken from spud to rig release:**
   * Once drilling operations commence (spud), this stage measures the duration until the rig is released from the wellsite upon completion of drilling activities. It accounts for the drilling process, including drilling depth, formation characteristics, and operational challenges.
4. **Time taken from rig release to well completion:**
   * Following rig release, this stage represents the duration between the release of the drilling rig and the completion of well construction activities. It encompasses activities such as casing installation, cementing, perforation, and hydraulic fracturing (if applicable).
5. **Time taken from well completion to first production:**
   * After the completion of well construction, this stage measures the duration until the commencement of hydrocarbon production from the well. It includes post-completion operations.

# Assumptions:

1. **API14 Instances Linked to API10:**
   * The model assumes that multiple API14 instances (representing individual wells) are linked to a single API10 (representing a well pad or location).
   * When selecting the API10 for analysis, the model chooses the API10 associated with the API14 instance that has the maximum lateral length.
   * This assumption ensures that the selected API10 represents the most significant well within a given well pad or location, which may have the most substantial impact on cycle times.
2. **Handling Outliers:**
   * The model acknowledges the presence of outliers associated with different cycle times, which may distort the overall analysis if not appropriately addressed.
   * To mitigate the influence of outliers, the model opts to use the median value for each completion stage rather than relying on mean or other summary statistics.
3. **Handling Negative Completion to First Production Time:**
   * In general, we are getting completion to first prod time as negative, so we are assuming first prod date as completion date.

# Algorithm:

* + The model focuses on analyzing historical data spanning the past 5 years from the well completions table.
  + For most cycle times (permit submission to approval, permit approval to spud, rig release to well completion, and well completion to first production), information is extracted from the well completions table.
  + For the spud to rig release cycle time, the model retrieves data from the rigs history table instead of the well completions table.
  + The model computes the median value of all cycle times from the data.

# Rig Model

# `Purpose:

The Rig Model serves two main purposes: forecasting rig availability for drilling operations under each operator and facilitating the allocation and deallocation of rigs among operators within the basin. It provides insights into when rigs will be available to commence drilling activities and allows users to adjust the number of rigs allocated to each operator based on operational needs. By forecasting rig availability and enabling dynamic rig allocation.

# Assumptions:

In general, the number of active rigs varies over time, reflecting fluctuations in drilling activity and operational demand within the basin. However, for the current analysis, it is assumed that the number of active rigs remains constant over time based on user input.

# Algorithm:

The rig algorithm can works as follows:

1. **Historical Rig Data Analysis:**
   * The algorithm analyzes historical rig data from the past 2 years to understand the distribution of rigs among different operators at various points in time.
   * It determines the percentage of rigs allocated to each operator over time, providing insights into rig utilization patterns and operator preferences.
2. **Combination of Small Operators:**
   * To simplify the analysis and allocation process, the algorithm may combine smaller operators into a single category labeled as "Miscellaneous."
   * This consolidation helps streamline rig allocation decisions by reducing the complexity associated with managing numerous small operators individually.
3. **Allocation and Deallocation of Rigs:**
   * Based on user input and operational requirements, the algorithm facilitates the allocation and deallocation of rigs among operators.
   * Users can specify the desired number of rigs allocated to each operator, and the algorithm adjusts rig allocations accordingly.
   * This functionality allows operators to adapt rig deployment strategies to meet drilling targets, project timelines, and resource availability.
4. **Calculation of Rig Release Time:**
   * The algorithm leverages the Cycle Time Model to calculate rig release times for all rigs within the basin.
   * Using the cycle time data, it estimates the duration between rig deployment (spud) and rig release for each drilling operation.
   * This information helps operators plan drilling schedules, assess project timelines, and optimize resource utilization.

# Drilling & Scheduling Model

# Purpose

The purpose of the Drilling & Scheduling Model is to generate production dates for inventories (wells) within the basin. This is accomplished by leveraging data and insights from both the Cycle Time Model and the Rigs Model.

# Assumption

During drilling operations, rigs can be moved to any location within the type curve area without considerations of distance.

# Algorithm:

1. **Current Operator Completion:**
   * Initially, the rig operates under a specific operator within the typecurve area and completes all the inventories (wells) assigned to that operator based on their respective IRR rankings, starting with the highest-IRR inventory.
2. **Evaluation of Other Type Curves for Same Operator:**
   * After completing the inventories for the current operator, the model evaluates all other type curves for the same operator.
   * It calculates the ratio of undrilled inventories to active rigs on each type curve for that operator at the current moment in time.
3. **Selection of Best Type Curve:**
   * The model selects the best type curve for the operator based on the highest ratio of undrilled inventories to active rigs.
   * This selection aims to prioritize moving the rig to the type curve area with the greatest drilling demand relative to available drilling capacity.
4. **Completion of All Type Curves for Operator:**
   * The rig continues to move across type curve areas for the same operator, completing inventories until all type curves for that operator are exhausted.
5. **Move to Another Operator at Same Type Curve Area:**
   * Once all type curves for the current operator are completed, the rig is moved to another operator within the same type curve area.
   * The selection of the next operator is based on the maximum ratio of undrilled inventories to rigs among operators within the same type curve area.
   * This strategy aims to optimize rig utilization by prioritizing drilling activities in areas with the highest drilling demand relative to available rigs.

# Production Estimation Model

# Purpose:

The purpose of the Production Estimation Model is to provide comprehensive estimates of combined production for oil, gas, and water from the entire basin. This model is designed to offer insights into the production potential of various categories of wells within the basin, including new wells, producing wells, inventory wells, and work-in-progress (WIP) wells.

# Algorithm:

# WIP Wells:

The Production Estimation Model addresses work-in-progress (WIP) wells by first identifying the API number of the rig's current location. From there, it retrieves the first production dates for associated wells and determines their corresponding type curves. By merging this data with type curve production information, the model estimates the future production potential of WIP wells.

# Inventory Wells:

Similar to the approach for WIP wells, the Production Estimation Model addresses inventories by first identifying their current locations within the basin. By retrieving completion or drilling dates for each inventory. By integrating the production information of inventories with their respective type curve profiles, the model estimates their future production potential.

# Producing Wells:

In producing wells, there are two types of production data. For wells with existing forecasted values, these are directly incorporated into the model without alteration. However, for wells with data available only up to a certain date, a different approach is taken. The model assesses the latest production records for these wells and compares them with type curve production values, calculating an error factor. Using this factor, the model adjusts the forecasted values from the type curve production table for periods beyond the last recorded data, ensuring continuity and accuracy in the production estimates.

# New Wells:

For new wells with production data available up to a certain date, a similar methodology is applied. The model forecasts production for these wells using the same approach as described previously, which involves calculating an error factor by comparing the latest production records with type curve production values. This error factor is then used to adjust the forecasted values from the type curve production table for periods beyond the last recorded data, ensuring accurate and reliable production estimates for the new wells.

# Denormalization:

After obtaining total production values, all data is denormalized using lateral length as a reference. Previously, the values were normalized assuming a lateral length of 10,000 feet. To denormalize the data, the production values are adjusted based on the actual lateral length of each well. This ensures that the production estimates accurately reflect the characteristics of each individual well, accounting for variations in lateral length across the basin.