**Goals:**

* Determine the most accurate method for forecasting PDP in Raisa
* Group wells for Securitization based on a 95% confidence requirement for the forecast when compared to the actual production

**Project Deadline:**

* **Friday December 14th, 2018**

**PDP Forecasting Methodology:**

* Arps Model (Hadeer) and Machine Learning Model (Salma)
* Both models will run in parallel on the same set of wells and have the same tasks. We will compare the accuracy from the two models and determine the best path forward when forecasting PDP for Raisa.

**Dataset to be used:**

Purpose of 3 Datasets:

* Help improve deterministic forecasting by grouping certain wells based on features
* Identifying wells that do not have commonalities, but grouping them improves precision of the forecast
* Improving the Machine Learning model

**Datasets:**

* Horizontal wells drilled before 2014 (including wells with and without Raisa interest across all 4 basins)
  + DS Team to query .IHS to obtain this data
  + DS Team should provide an API list for this dataset to the Engineers
  + 4 basins: DJ, Williston, Anadarko, Permian
* Horizontal wells drilled after 2014 (including wells with and without Raisa interest across all 4 basins)
  + DS Team to query .IHS to obtain this data
  + DS Team should provide an API list for this dataset to the Engineers
  + 4 basins: DJ, Williston, Anadarko, Permian
* Raisa Securitization Package (all wells in Raisa Reserves PDP database that will be used for 12/31 reserves)
  + **A duplication of the Raisa Reserves Aries database should be created. Auto-forecasts will be imported into this database for comparison and QC.**
  + **All wells within this duplicated database should be forecasted by the Engineers.**

**Requirements/Tasks:**

* **Q’C accuracy of PDP forecasts**
  + **The following tasks should be done on each of the 3 datasets described above.**
  + 3, 6, 9, 12 month data should be used in forecasting
    - 4 deterministic (P50) forecasts will be made for each well
      * One using only the first 3 months, one using the first 6 months, one using the first 9 months, and one using the first 12 months
  + Calculate monthly variance/error between the forecast and actual production data for each forecast type (3M, 6M, 9M, 12M) for the life of the well
    - The definition of the variance/error calculation should be defined and presented to the Engineering team
    - Summarize variance/error in histograms and tables
      * This tabular data should be calculated for each of the 4 forecasts in all datasets
    - Error/variance should be discounted over the life the of the well
      * Error/variance in the early time prediction of a well impacts cashflow much more than error/variance in the later time production.
      * Guidance from the Engineers can be provided on how to approach this if needed.
  + Only forecasts for the wells in the Raisa Securitization Package dataset (all wells to be included in the 12/31 reserves) should be exported
    - These forecasts will be imported into the Aries database (duplicate of Raisa reserves) for comparison against historical forecasts
    - This is the only well set that requires forecasts to be exported
    - Forecasts should be in the format that was provided to Sandy for the Jarvis report
  + Based on variance/error data, the Engineering team should determine the validity of forecasts from both the Arps model and the ML model
    - Desired value for acceptance: Average monthly error/variance distribution should be centered on 0 (both Non-absolute and Absolute)
    - A decision will be made to for which model to use going forward
  + Methodologies for improving the deterministic forecasting in both the Arps and ML models should be investigated
* **PDP Database**
  + After Engineering has signed-off on a forecasting method for deterministic (P50) forecasts, forecasts should be imported into the PDP Database
  + Spec for the PDP Database is found here:
    - [https://raisaoilandgas.sharepoint.com/:w:/g/EU6w7TWVgvVGuHZ3TCIlq1wBYzyC5FmlulvKRK-XnYSp5Q?e=VoNH4Y](https://raisaoilandgas.sharepoint.comw:/g/EU6w7TWVgvVGuHZ3TCIlq1wBYzyC5FmlulvKRK-XnYSp5Q?e=VoNH4Y)
  + Only P50 forecasts should be imported into the database for the first phase
  + The wells in the Raisa Securitization Package (wells in the Reserves database) should be flagged in the database with the Raisa Well ID
* **Probabilistic Forecasting**
  + Arps Model (Hadeer)
    - The Di\_SECANT of the closest 10 neighboring wells should be selected for a given target PDP
      * Selecting neighboring wells should follow the same criteria as the ML model (refer to rules in ML model)
    - MonteCarlo simulation should be run on the Di\_SECANT values of the 10 wells to determine the P10 and P90 Di\_SECANT values
    - The P10/P90 values from the simulation should be used to construct P10 and P90 forecasts for the target PDP
  + ML Model (Salma)
    - The ML model should be used to calculate the P10/P90 values for a given PDP
    - There will be little guidance from domain expertise in this task as to not bias or guide the model.
    - If the model produces inconclusive or irrelevant results, then logic similar to the MonteCarlo simulation can be followed.
  + P10/P90 forecasts from each model (Arps and ML) should be exported for the Raisa Securitization Package dataset only
    - The export should be in the same format provided by Sandy for Aries import
* **95% Confidence Determination**
  + Each well should be forecasted with a curve that has 95% confidence when compared to the actual production
  + If wells cannot be forecasted with a 95% confidence, then they should be flagged.
  + Wells that can be forecasted with 95% confidence should be flagged
    - Wells that meet the 95% confidence should be analyzed a grouped using cluster analysis or any other grouping methodology
    - API lists of the grouped wells should be exported
      * Location for exported lists to be determined by the Engineers