* **Project Title:**

Machine Learning of the Equinor – Volve Production Data set

* **Source of the Dataset** (Equinor Volve Dataset)**:**

The Volve oil field is located in Block 15/9 in the southern part of the Norwegian North Sea, at a water depth of around 80m. It is situated approximately 200km west of Stavanger and 8km from the Sleipner Ost Field. The dataset contains information from 7 production wells in the Volve Field for a period from 2008 to 2014. The dataset could be found in the following [Link](https://www.equinor.com/en/how-and-why/digitalisation-in-our-dna/volve-field-data-village-download.html)

* **Scope of the Project:**

1. To build a prediction model to determine downhole pressure of a test well.
2. To predict the choke size for a given gas and oil rate.

* **Assumptions for Data Analysis:**

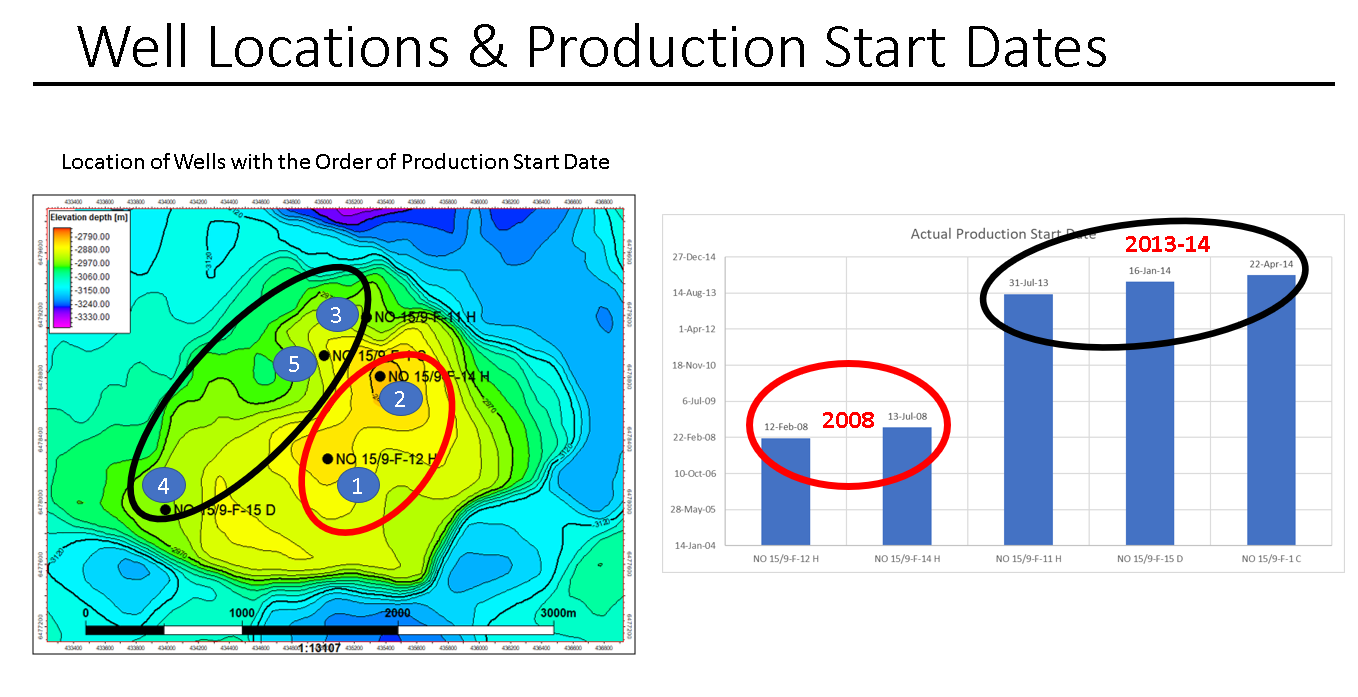
1. We have done analysis of different variables for the 7 wells in the dataset. Two of those wells did not have any production at all, so we dropped them. Among the five candidate wells, we found two clusters of wells, considering production conditions such as locations of the wells in the field and production period (Figure 1 ). 2 wells are situated at the structural crest of the field and have been producing from 2008. The other 3 wells are situated on the flanks of the structure and has production from 2013-14 period. All these well could be separated by faults that may compartmentalize the reservoir (hydrocarbon tank).We have formed two clusters in our dataset based on this clustering. Also, during production period there are days of production shut-in probably for mechanical or pressure maintenances. Those days have been eliminated from this analysis.

Figure 1. Locations of the 5 wells and their time periods of production, showing the two clusters of wells. Wells 1 and 2 are on crest and 3-5 are on the flank reservoir.

1. Each row is an instant of time, during the production period, in chronological order, that gives us the daily production rate as a function of the other variables on that day. Our models will be able to predict the choke size and the downhole pressure of the well, based on other variables. On any given instant, these outputs are independent of the previous outputs, they are mutually independent and all samples belong to the same dataset and identical to each other.

* **Task Breakdown:**

1. Extraction and Understanding of the production and subsurface dataset. Analyze the various production parameters from each well and establish correlations and variances. (Pairwise plots and correlation plot- Heat map)
2. Transforming the dataset – Performing some data wrangling and clean up.
3. Split data into testing and training datasets
4. Performing Linear and Non-Linear Regression models with Regularization (Lasso shrinkage)
5. Residual Plotting.
6. - Grid SearchCV.
7. Creating a visualization dashboard to predict the downhole pressure and choke size for certain given input parameters, with an accuracy of our models.