* **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.

* **Analysis of the Project:**

1. Downhole pressure is an important criterion of a well. Downhole gauge is an instrument that measures the downhole pressure of the well. However, the downhole gauges in each well constitute significant cost during full-field development or with passing time, sometimes could be worn-out / malfunction due to corrosion and gunk buildup or mechanical reasons. When the pressure gauge malfunctions, operators usually do not replace it due to the expense of the process and the lag in the hydrocarbon production that would mount significant costs. However, since it is not replaced, a well loses the record of the downhole pressure ever since it starts malfunctioning. To eliminate this, we would like to develop a model that would be able to predict the downhole pressure of the well. We would like to understand the dependencies of the different variables and what influences the downhole pressure most, so we can use those variables to predict the downhole pressure.
2. There is a physics that determines the dependencies of the different variables in the dataset with the output. Operators would always like to achieve minimum water to oil ratio. That means, for a given choke size, the output of water should be minimum, and output of oil would be the maximum. Through our project, we would like to build a second model that would be able to predict the choke size for a given oil and water rate. Operators, usually recalculate optimum choke sizes for every month of their production period to keep the water to oil ratio minimum and if we are able to predict that without requiring them to do these routine tests, this would save them the man hour and the associated expenses.

* **Assumptions for Data Analysis:**

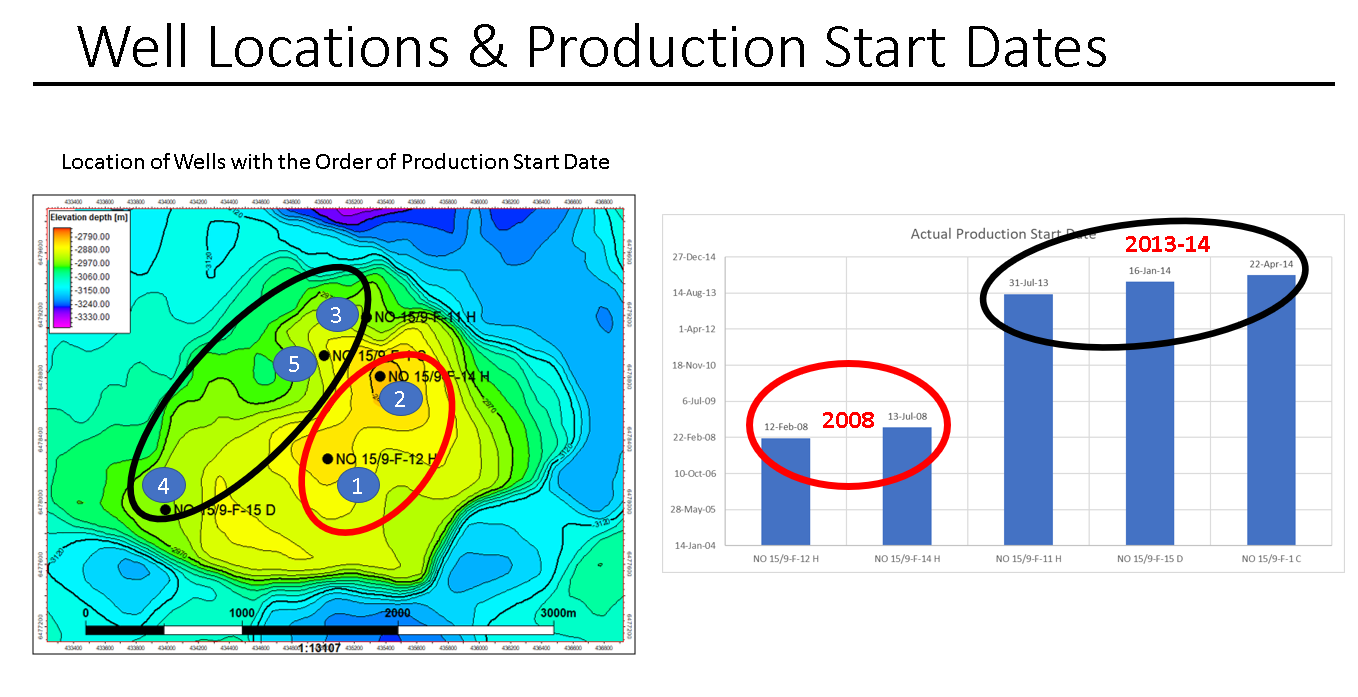
1. We have done analysis of different variables for the 7 wells in the dataset. Two of those wells did not have any production at all, so we dropped them. Among the five candidate wells, we found two clusters of wells, considering production conditions such as locations of the wells in the field and 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. However, in the proposed prediction models, we plan to compare production controls and volumes among the wells, rather than doing their individual time series analysis considering each of these events or rows are IID (Independent and identically distributed). 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.
8. Each of the wells belong to the same distribution- do hypothesis testing on the datasets, the mean is same for the output, they come from same statistical distbn. Frequency plot.