**Project Overview**

This project applies petrophysical analysis techniques to predict **formation tops** using data from **offset wells** within the same hydrocarbon field. Formation tops are depth markers that indicate the boundaries between key geological formations. Accurately predicting these tops improves well placement, reduces drilling uncertainty, and enhances subsurface understanding.

By using **well log responses** (e.g., gamma ray, resistivity, density, neutron porosity), the project identifies patterns near known formation tops and builds **regression models** to estimate the top depths in wells where full interpretation may not yet be available.

**Objectives**

* Integrate well log data from multiple offset wells in a consistent format.
* Perform exploratory data analysis (EDA) to understand relationships between logs and formation tops.
* Train regression models to predict the depth of formation tops in new wells.
* Visualize predictions and assess model performance.
* Provide an interpretable, reproducible workflow for subsurface teams.

**Dataset Description**

* Source: Training data used for a petrophysical analysis software called Aspen Geolog
* Type of data: \*.las files and \*.txt files
* Data used:
  + Total vertical depth
  + Cartesian grid coordinates
  + Gamma Ray
  + Resistivity
  + Acoustic logs
  + Neutron Porosity
  + Formation top names and corresponding vertical depth

**🛠️ Technologies & Libraries Used**

* pandas, numpy – data handling
* matplotlib, seaborn – visualization
* scikit-learn – regression modeling
* lasio – reading LAS files

**Exploratory Data Analysis (EDA)**

* Data Cleaning (handling missing values, data types)
* Plotting of well curves
* Examination of mnemonics and the location in the .las files

**Regression vs Classification**

**Why Regression Was Used:**

This project uses **regression tools** because the target variable is **continuous** rather than **categorical**. Predicting the formation top depths based on offset well formation top depths and the geographic locations requires regression tools.

**Justification:**

* The objective is to predict or model a **numerical value which is the depth at which a particular formation top can be predicted at a particular location in the field**.
* Evaluation metrics used include **RMSE**, **MAE**, **R² score**, which are specific to regression tasks.
* Data relationships explored include **linear and non-linear correlations**, best understood through regression analysis.

**Well Location and Distance from Each Well Calculated:**

A screenshot of a computer

AI-generated content may be incorrect.

**Formation Tops DataFrame:**

A screenshot of a computer

AI-generated content may be incorrect.

**Well Log Information:**

A graph of a graph showing a variety of data

AI-generated content may be incorrect.

**Project Structure**

petrophysics- project/

data/ # Raw and cleaned datasets

notebooks/ # Jupyter notebooks for EDA and modeling

images/ # Saved plots and visualizations

README.md # Project documentation

**Conclusion & Next Steps**

* The key findings from the EDA are:
  + The signature of the well logs are similar to each other which helps to apply ML
  + The log data is limited and upon examining, it is not clear that the formation tops can be derived directly from the log signatures. Hence the offset data and the formation top depths were used directly instead without interpreting the log signature. More data from sampling, cores are necessary to develop a formation top prediction from the log signatures and other data.
  + Some wells have the location coordinates embedded in the ‘params’ section rather than the ‘wells’ section of the .las file
  + Some formation top names are in CAPS and need to be converted for consistency
  + The key features identified are Easting, Northing, Depth and the Formation Top.
  + Baseline regression models have been built and tested with Linear Regression, KNeighbors Regressor, Decision Tree Regressor, SVR, Ridge and Random Forest Regressor.
  + Next steps include model refinement, expansion of features to inter-distance between each well, and potentially applying more advanced regression techniques.