Wireline Log Analysis

Introduction

The economic value of the Petroleum Industry is unparalleled, making it one of the most significant industries in the world. However, the operator companies responsible for making key decisions often face high costs in the form of millions of dollars paid to service companies. These costs can be attributed to the exorbitant rental fees charged for tools, which can run into thousands of dollars per hour, ultimately amounting to millions of dollars per day. Given the high stakes involved, it is essential for operator companies to make the most accurate decisions possible, as service providers are happy to receive payment for any decision made. Achieving this level of accuracy requires a deep understanding of the problem at hand, which primarily centers around the identification of oil reserves and their location through the analysis of logs.

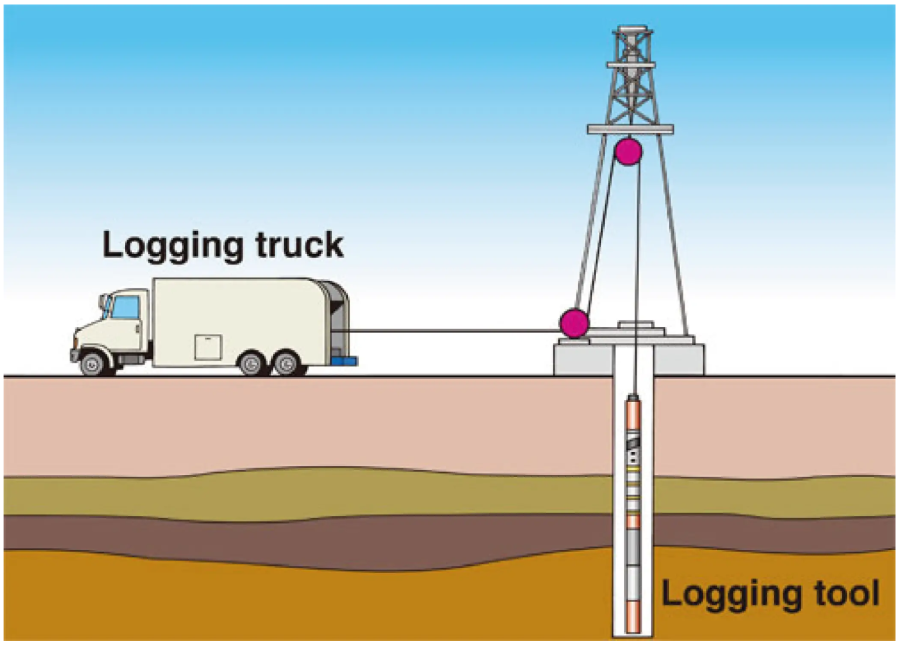
Scientific Background

Once an oil well has been drilled in a known oil reservoir, the next step is to use a wireline logging tool to gather comprehensive data about the well. This data is crucial in determining whether oil exists in the well and, if so, its precise location within the well. The wireline logging tool is lowered into the well and used to collect all available data, providing critical information for assessing the viability of the well. The typical logging tool contains the following sensors:

- Near, medium, and deep resistivity
- Self-potential
- Gamma-ray
- Neutron porosity
- Shear, and compressional sonic wave
- Photoelectric factor
- Bulk density

- Caliper
- Depth

However, the cost of using each sensor is calculated separately and several variations of the tools exist.



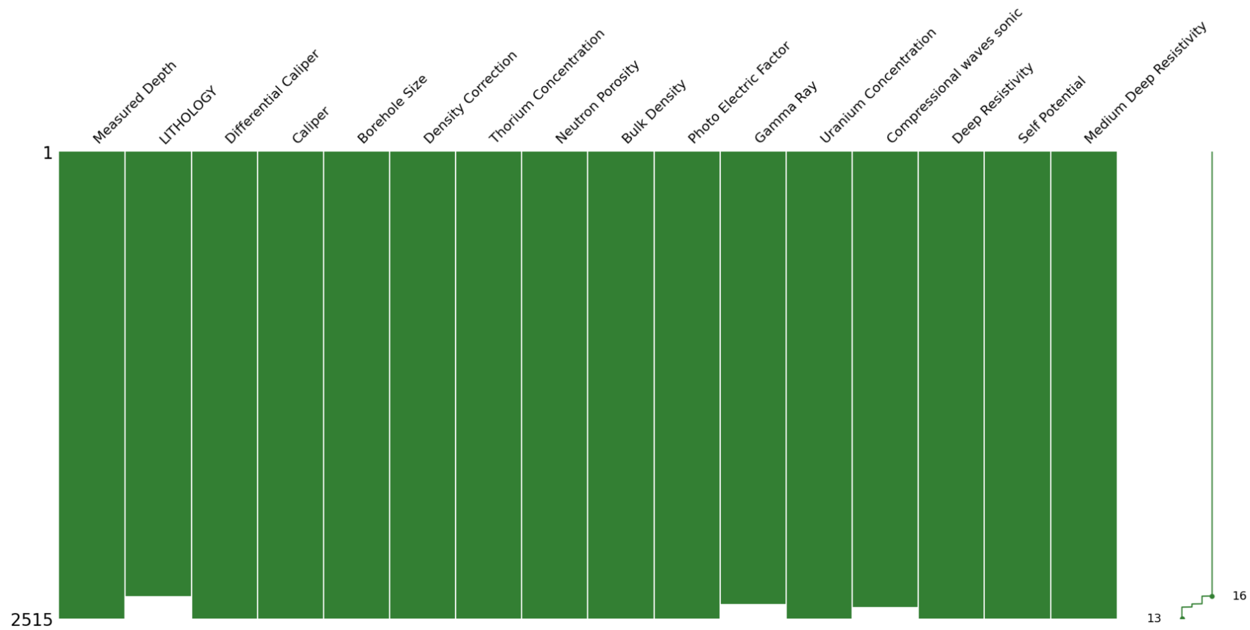
Dataset

- Norwegian government well logs dataset
- Force 2020 competition dataset

Objective

- Study the good nature
- Predict the lithology of the well
- Predict the possible oil or gas existence locations

Data Missingness



There are a few missing data that we can drop easily without affecting our analysis. Data missingness in oil logs is due to the lack of importance to these points in the well.

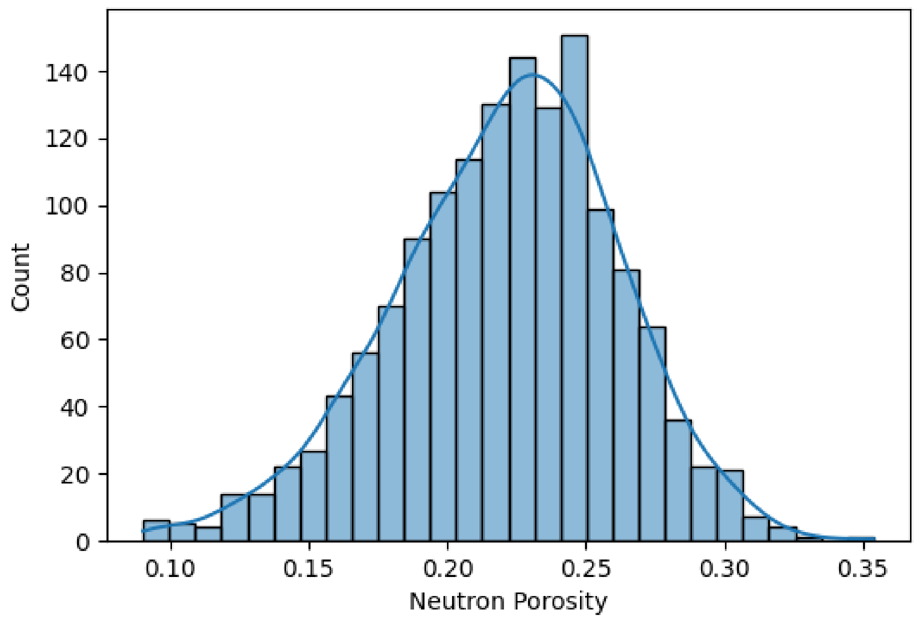
Q1: What are the sensor reading distributions?

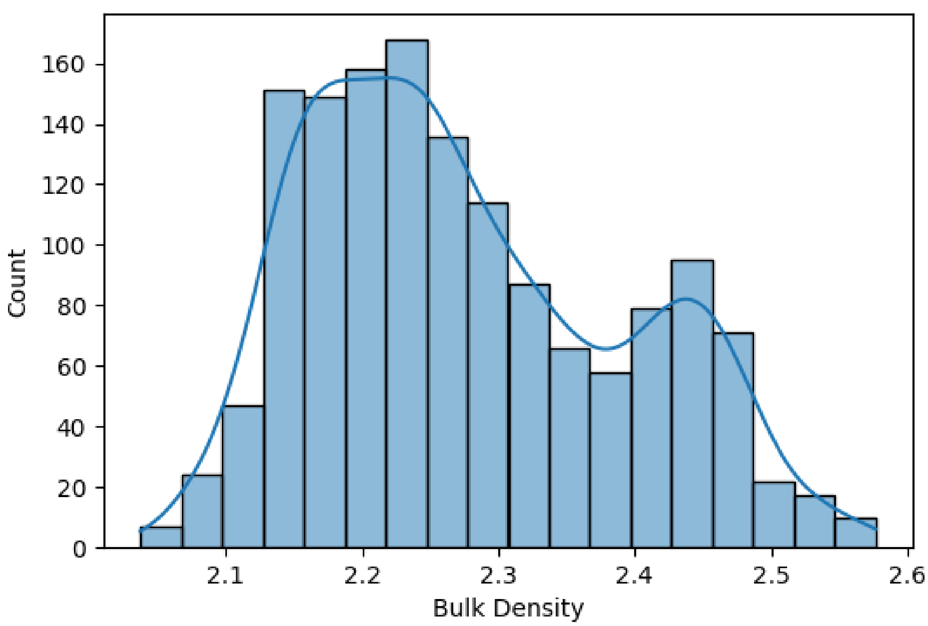
Purpose of the question: Understand the nature of the data

Expectation: we expect that the distribution of each sensor reading is dependent on the well's structure and geological formation however the sensor readings are expected to be within an expected range:

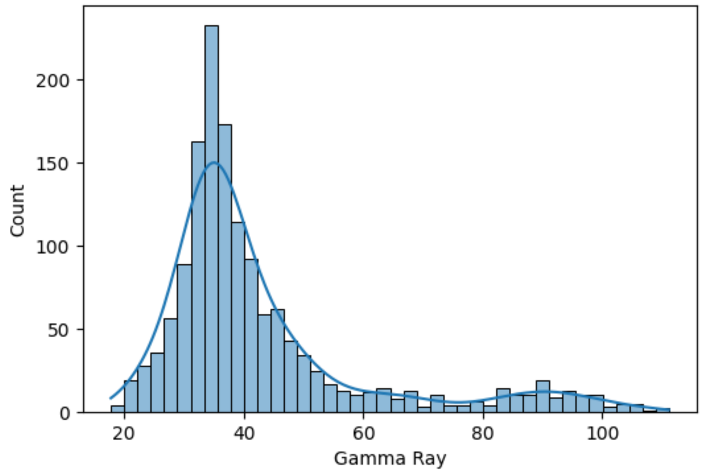
- Resistivity ranges from fractions to thousands of ohms (depending on the rock formation and presence of water)
- Different resistivity types distributions are close to each other
- In the case of using thorium or uranium sensors, they take a distribution close to gamma rays' due to the common nature of radioactive materials

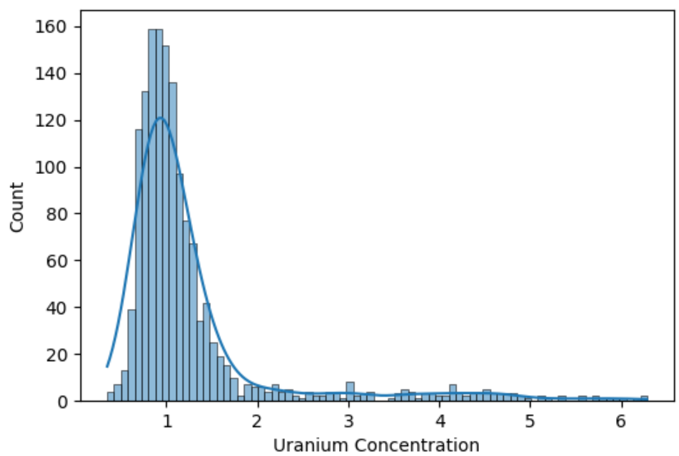
Answer: we plot the different readings as histograms

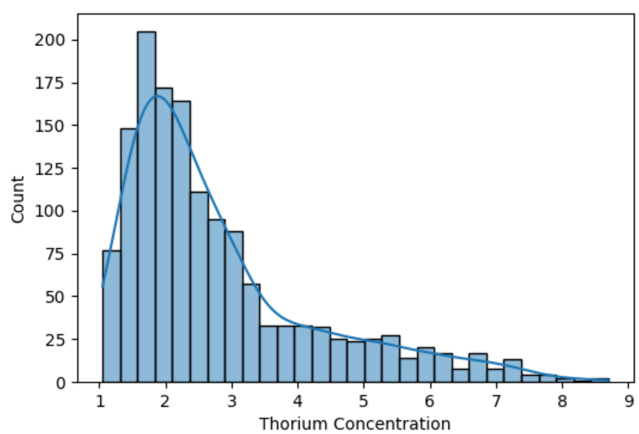




Observation: Neutron porosity ranges from **0%** to **35%** while bulk density ranges from 2 g/cm^3 to 2.6 g/cm^3 which are very important ranges used in detecting crossovers that express the presence of hydrocarbons





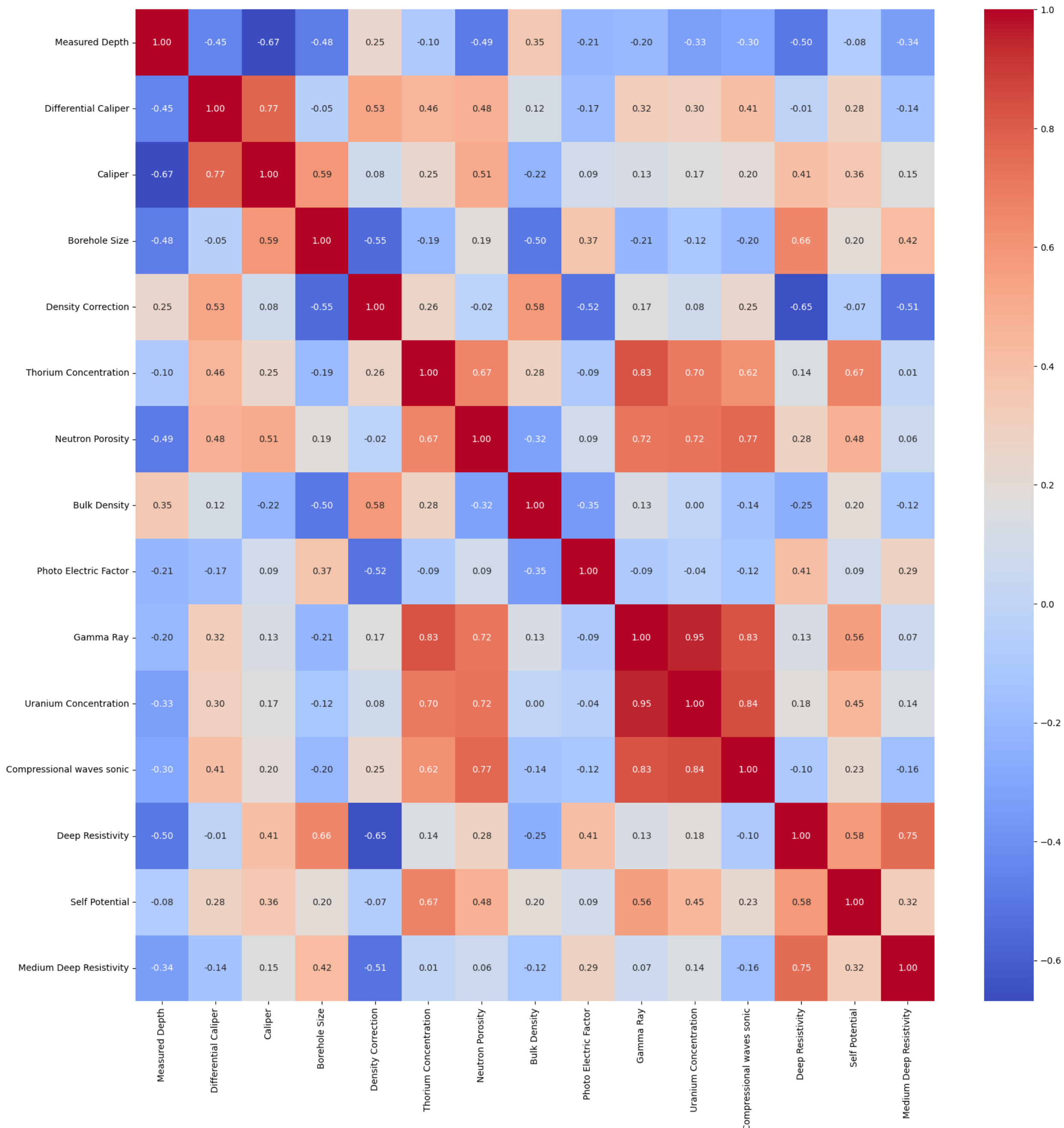


Observation: all radioactive materials have the same distribution

Q2: How correlated are the sensor readings?

Purpose of the question: Understand how redundant is the data

Expectation: We expect to find all types of resistivity readings to be highly correlated. Also, the radioactive materials should be somehow correlated with each other.



Observation:

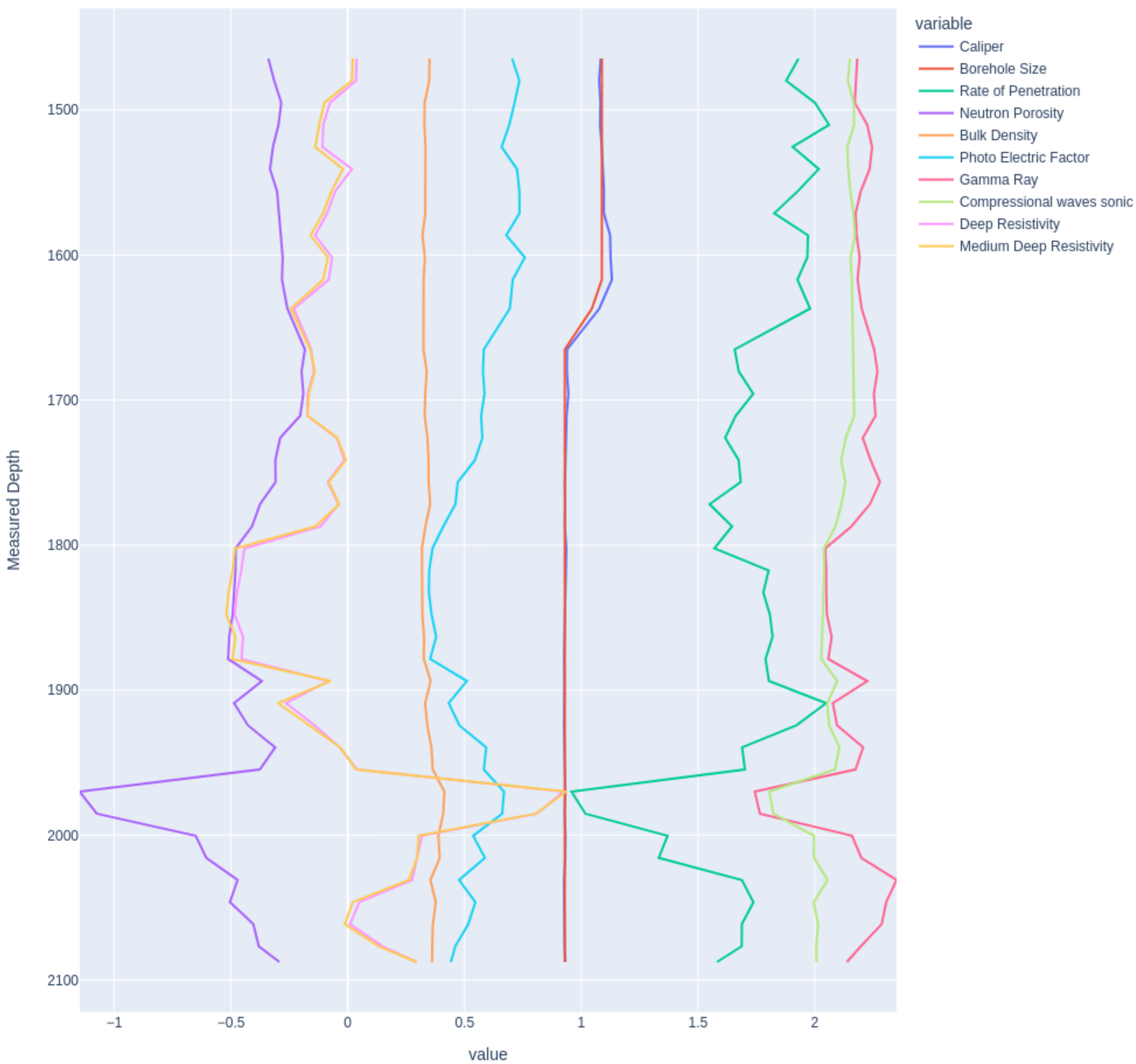
- Shear wave sonic and Compressional wave sonic are highly correlated
- Uranium Concentration or Thorium Concentration are highly correlated with Gamma Rays as all of them emit Gamma Rays
- Caliper and Borehole size are highly correlated which is expected
- Neutron porosity and compressional wave sonic (or Shear wave sonic) are highly correlated which gives an indication that we can give up on the 'Neutron porosity' sensor as it is highly expensive and dangerous (nuclear hazard)
- Gamma Rays are not always highly correlated with Neutron porosity (and in turn with compressional wave sonic ..) but sometimes they are depending on the formation
- Deep Resistivity, Medium Deep Resistivity, Shallow Resistivity, and Micro Resistivity are relatively highly correlated because generally, they measure the same value

Q3: How does sensor reading change with depth?

Purpose of the question: Find out if the depth is affecting the other columns (is depth a feature)

Expectation: Sensor readings are affected by lithology regardless of the depth

Answer: To find how readings are affected by the depth we divided the oil height into equal partitions and took the mean of all the readings in this partition and plotted the results



Observation:

- Neutron Porosity tends to decrease with depth
- Gamma tends to decrease with depth
- Resistivity rends to increase with depth but generally it fluctuates
- Other readings keep their mean value which means they are important only when they have anomalies

Q4: The means of gamma rays and neutron porosity are significantly different in shale and sandstone in the same well. Test the hypothesis.

Purpose of the question: Find out if neutron porosity and gamma rays are the main features of classifying the lithology as believed

Expectations: We expect neutron porosity to have different means for shale and sandstone but not gamma rays

Answer: we have:

- Null hypothesis: both means of gamma rays and neutron porosity aren't significantly different in shale and sandstone
- Alternative hypothesis: both the means of gamma rays and neutron are significantly different in shale and sandstone

We did the hypothesis test as following:

1. For each well in dataset
    a. Extract the sandstone and shale
    b. Do a 2-sample t-test using the gamma rays in shale and sandstone
    c. Do a 2-sample t-test using the neutron porosity in shale and sandstone

Results: all the p-values in all tests are less than 0.05 thus we can reject the null hypothesis and conclude that the means are significantly different

Q5: Can we predict lithology accurately based on wireline well logs data?

Purpose of the question: Classify the layers of lithology in the well to find the sandstone

Expectations: we can classify the lithology accurately using supervised machine learning

Answer: We transformed the data using a standard scaler then we trained several models:

Model	Accuracy
Logistic Regression	0.86
SVC	0.95
Random Forest	0.96

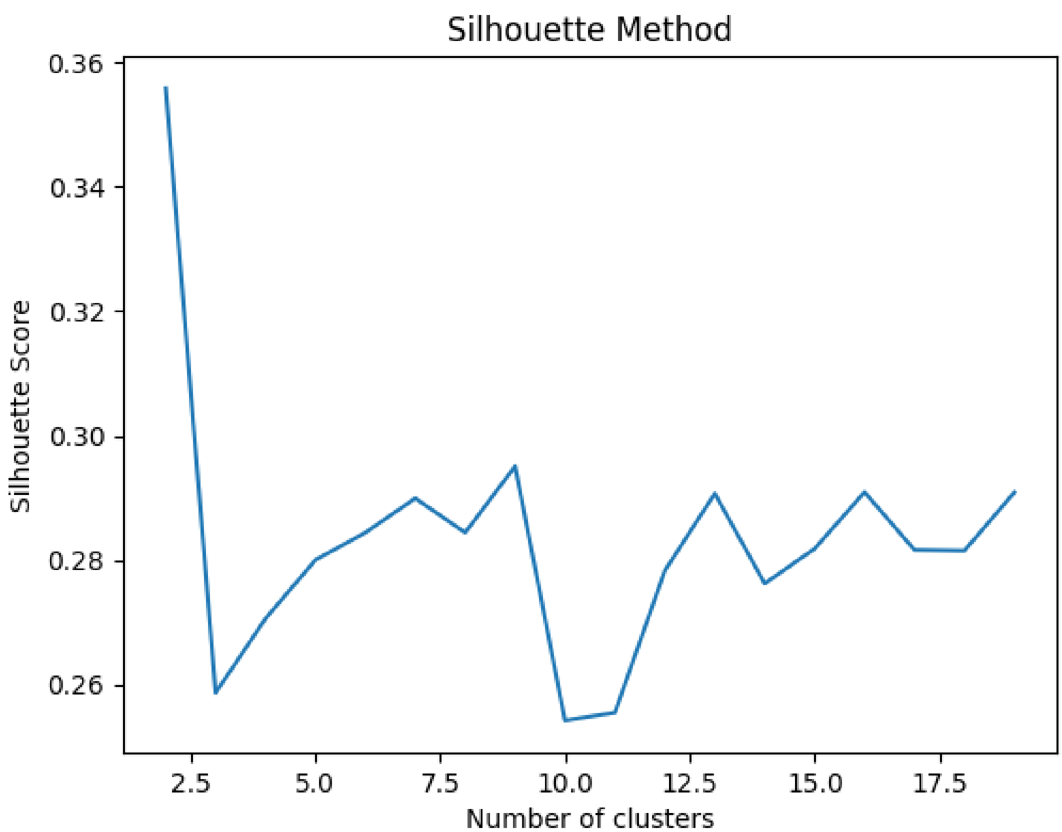
Observation: Random forest is the most accurate classifier for this type of data

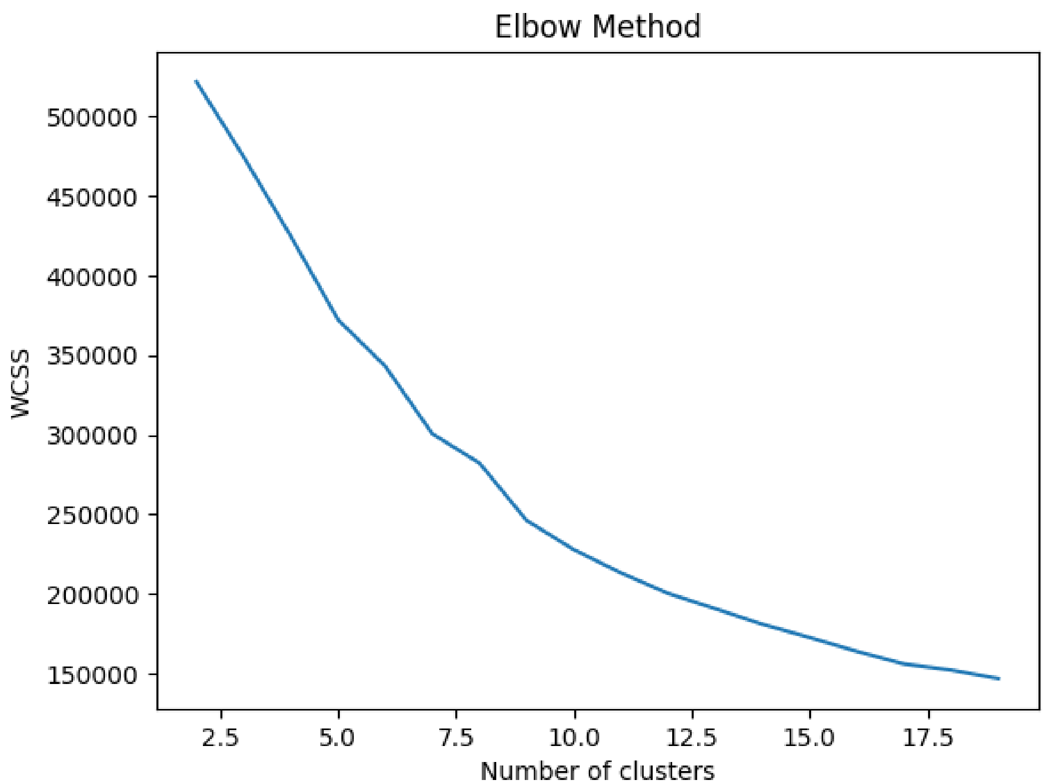
Q6: Can we predict lithology without labeled data?

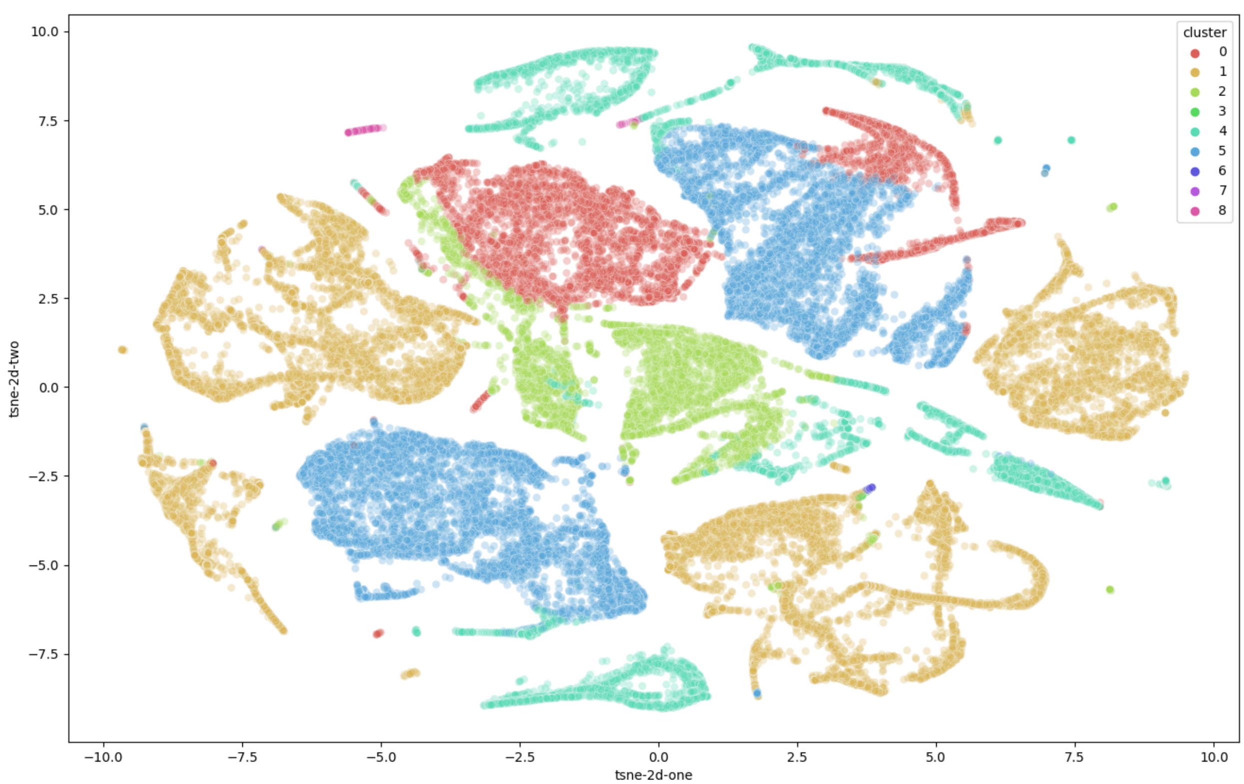
Purpose of the question: if the data labels are not accurate, can we have a better classification method

Expectation: we can cluster the data using unsupervised algorithms but clusters may not map to the classes in the supervised classification

Answer: We normalized the data using standard scaler then we tried clustering the data using K means clustering with different numbers of clusters and measured the silhouette score which was lowest at 10 an using the elbow method we found 9 to be a good number for clusters




Elbow Method
WCSS
500000
450000
400000
350000
300000
250000
200000
150000
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Number of clusters


cluster
0
1
2
3
4
5
6
7
8
tsne-2d-two
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0
tsne-2d-one

Observation: Clustering is very important in the case of lithologies however, it needs further analysis as we found out that sandstone count per cluster is high in more than one cluster as expected

```
# find the cluster with the most sandstone
df[df["LITHOLOGY"] == 0]["cluster"].value_counts()
```

```
[21] ✓ 0.0s
```

```
1    3911
2    1887
5     290
0     218
9     157
7     125
8      14
6       9
3       5
4       4
```

Q7: What are the most important features or well-log measurements contributing to lithology prediction?

Purpose of the question: can we reduce the readings acquired from the well to reduce the cost

Expectations: We can select a few features to reduce to predict the lithology with high accuracy to reduce the number of sensors used in logging and thus reduce the cost

Answer: We trained several random forests using iterative feature selection technique where we found the following accuracies

	Columns	Accuracy
0	['Shallow Resistivity']	0.540099
1	['Medium Deep Resistivity']	0.520050
2	['Deep Resistivity']	0.516804
3	['Bulk Density']	0.542104
4	['Gamma Ray']	0.654191
5	['Photo Electric Factor']	0.505919
6	['Compressional waves sonic']	0.588409
7	['Self Potential']	0.655528
8	['Density Correction']	0.534180
9	['Shallow Resistivity', 'Self Potential']	0.839603
10	['Medium Deep Resistivity', 'Self Potential']	0.852683
11	['Deep Resistivity', 'Self Potential']	0.862230
12	['Bulk Density', 'Self Potential']	0.831201
13	['Gamma Ray', 'Self Potential']	0.891827
14	['Self Potential', 'Photo Electric Factor']	0.805327
15	['Compressional waves sonic', 'Self Potential']	0.854210
16	['Density Correction', 'Self Potential']	0.775348
17	['Gamma Ray', 'Self Potential', 'Shallow Resistivity']	0.921902
18	['Medium Deep Resistivity', 'Gamma Ray', 'Self Potential']	0.925339
19	['Deep Resistivity', 'Gamma Ray', 'Self Potential']	0.925339
20	['Gamma Ray', 'Bulk Density', 'Self Potential']	0.925530
21	['Gamma Ray', 'Self Potential', 'Photo Electric Factor']	0.918083
22	['Gamma Ray', 'Compressional waves sonic', 'Self Potential']	0.920279
23	['Density Correction', 'Gamma Ray', 'Self Potential']	0.913787
24	['Gamma Ray', 'Bulk Density', 'Self Potential', 'Shallow Resistivity']	0.937273
25	['Medium Deep Resistivity', 'Gamma Ray', 'Bulk Density', 'Self Potential']	0.939087
26	['Deep Resistivity', 'Gamma Ray', 'Bulk Density', 'Self Potential']	0.941570
27	['Gamma Ray', 'Bulk Density', 'Self Potential', 'Photo Electric Factor']	0.938228
28	['Gamma Ray', 'Bulk Density', 'Compressional waves sonic', 'Self Potential']	0.935937
29	['Density Correction', 'Gamma Ray', 'Bulk Density', 'Self Potential']	0.935841
30	['Shallow Resistivity', 'Deep Resistivity', 'Self Potential', 'Gamma Ray', 'Bulk Density']	0.943479
31	['Deep Resistivity', 'Medium Deep Resistivity', 'Self Potential', 'Gamma Ray', 'Bulk Density']	0.943766
32	['Deep Resistivity', 'Self Potential', 'Gamma Ray', 'Bulk Density', 'Photo Electric Factor']	0.945770
33	['Compressional waves sonic', 'Deep Resistivity', 'Self Potential', 'Gamma Ray', 'Bulk Density']	0.943002
34	['Deep Resistivity', 'Self Potential', 'Gamma Ray', 'Density Correction', 'Bulk Density']	0.944052
35	['Gamma Ray', 'Shallow Resistivity', 'Deep Resistivity', 'Bulk Density', 'Self Potential', 'Photo Electric Factor']	0.946821
36	['Gamma Ray', 'Medium Deep Resistivity', 'Deep Resistivity', 'Bulk Density', 'Self Potential', 'Photo Electric Factor']	0.945770
37	['Gamma Ray', 'Compressional waves sonic', 'Deep Resistivity', 'Bulk Density', 'Self Potential', 'Photo Electric Factor']	0.948348
38	['Gamma Ray', 'Deep Resistivity', 'Density Correction', 'Bulk Density', 'Self Potential', 'Photo Electric Factor']	0.947966

Observation: Selecting Gamma Ray, Compressional waves sonic, Deep Resistivity, Bulk Density, Self Potential, Photo Electric as the only features gives us an accuracy of 0.9483482910063014 but reduces the cost of logging significantly from $75/m. To $30/m.. But this is regarding logging for lithology classification only.

Q8: "Geologists say that geology is not random, however, petroleum engineers say that it is random based on their experience" Given several well logs for the same reservoir (spatial domain), can we infer that there is no relation between the well location and its quality?

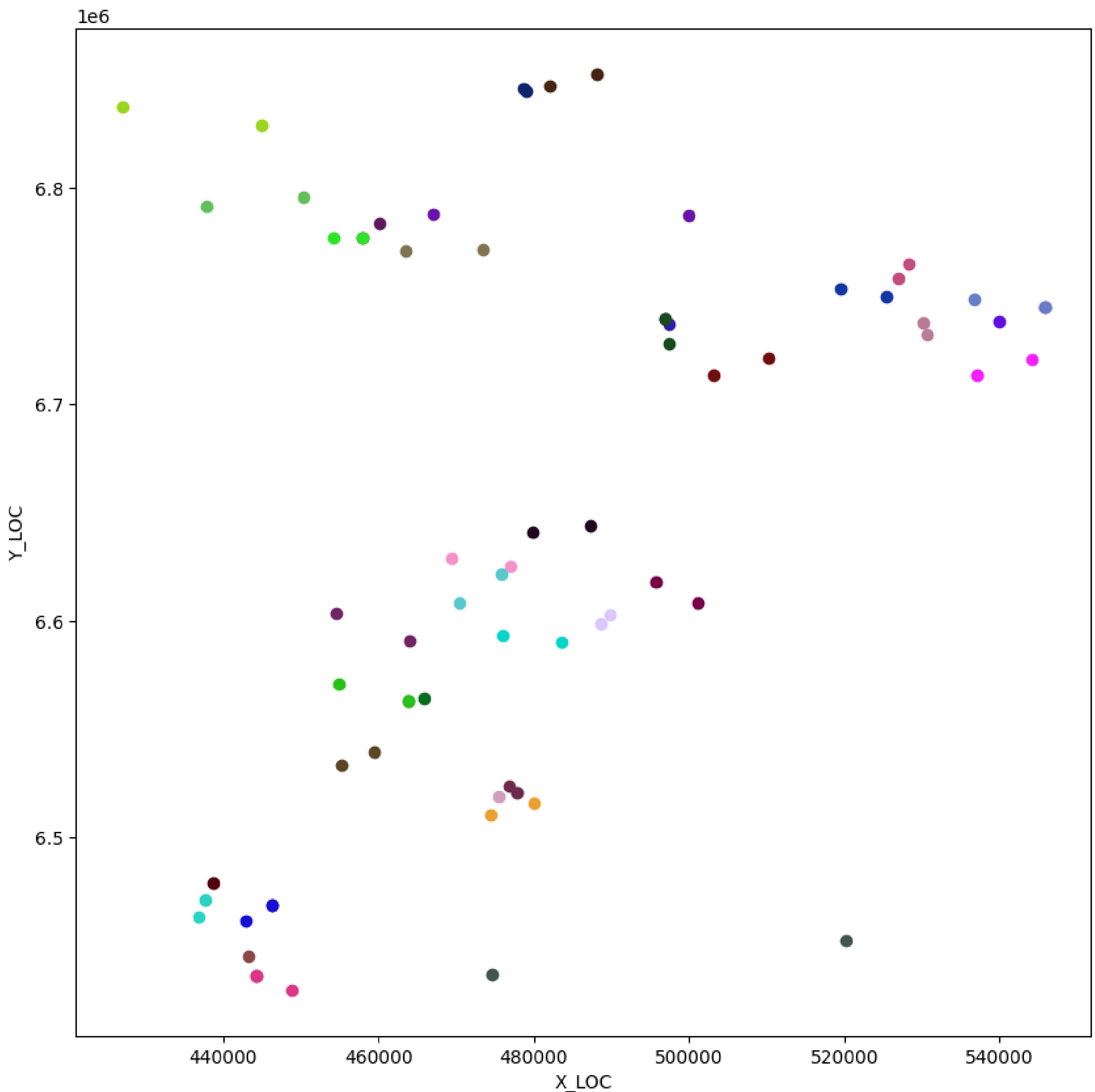
Purpose of the question: test the truth of the hypothesis

Expectation: We expected the null hypothesis to be true

Answer: We have:

- Null hypothesis: wells close to each other have the same distribution of lithology
- Alternative hypothesis: wells lithology is not the same regardless of how close they are to each other

We order the wells spatially by X, Y coordinates to find the closest wells to each other



Then we took each two close wells and tested the goodness of fit using the Kolmogorov Smirnov test. Kolmogorov Smirnov test is better than the chi-square here as it doesn't require the logs to have the same length in contrary to the chi-square.

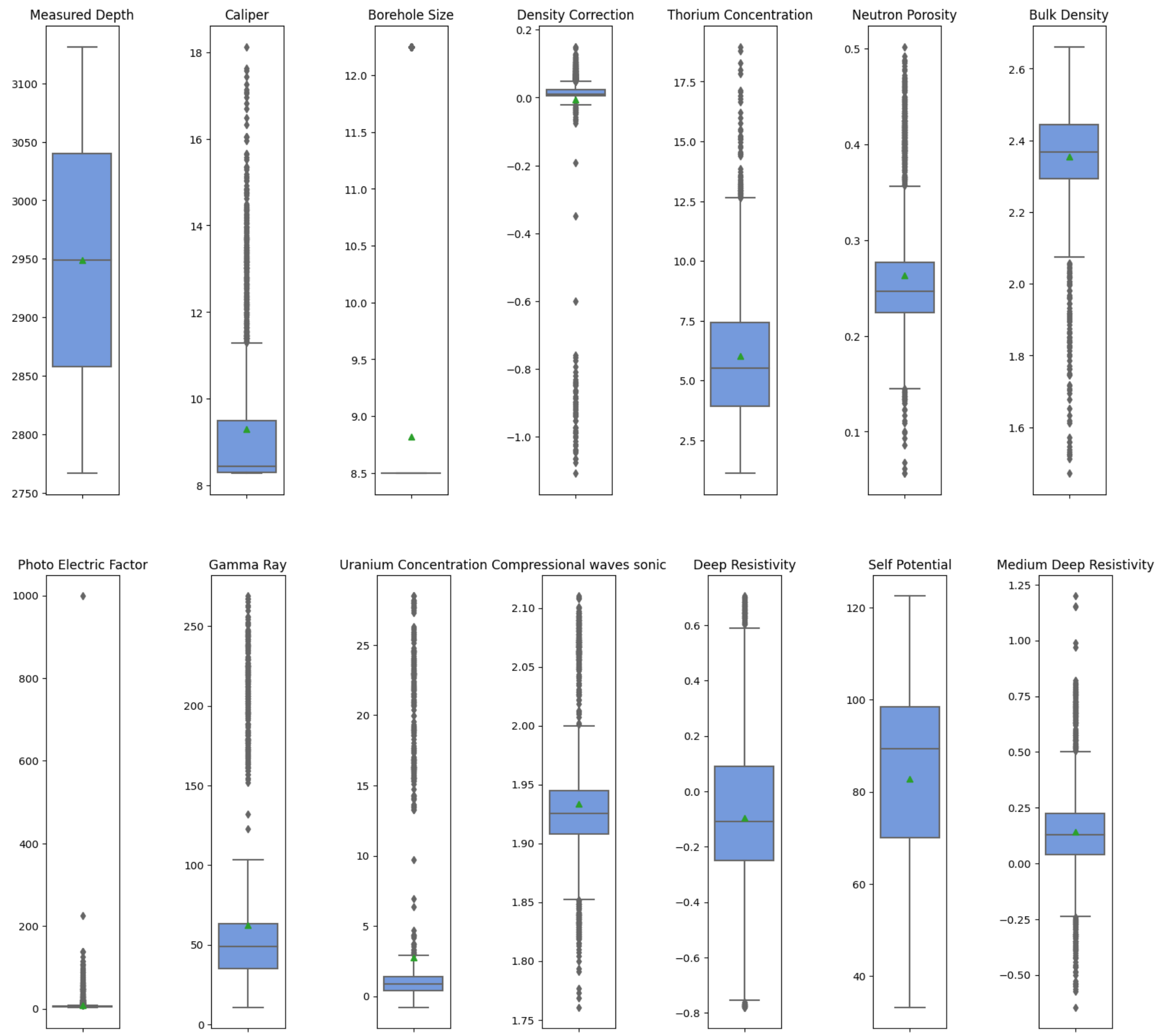
Observations: all the tests couldn't reject the null hypothesis thus the alternative is rejected

Q9: Do the noise and anomalies in sensor readings provide any information?

Purpose of the question: find the specific characteristics of the oil spots

Expectation: anomalies exist in nature but carry no useful meaning

Answer: Using IQD we plotted the anomalies as:



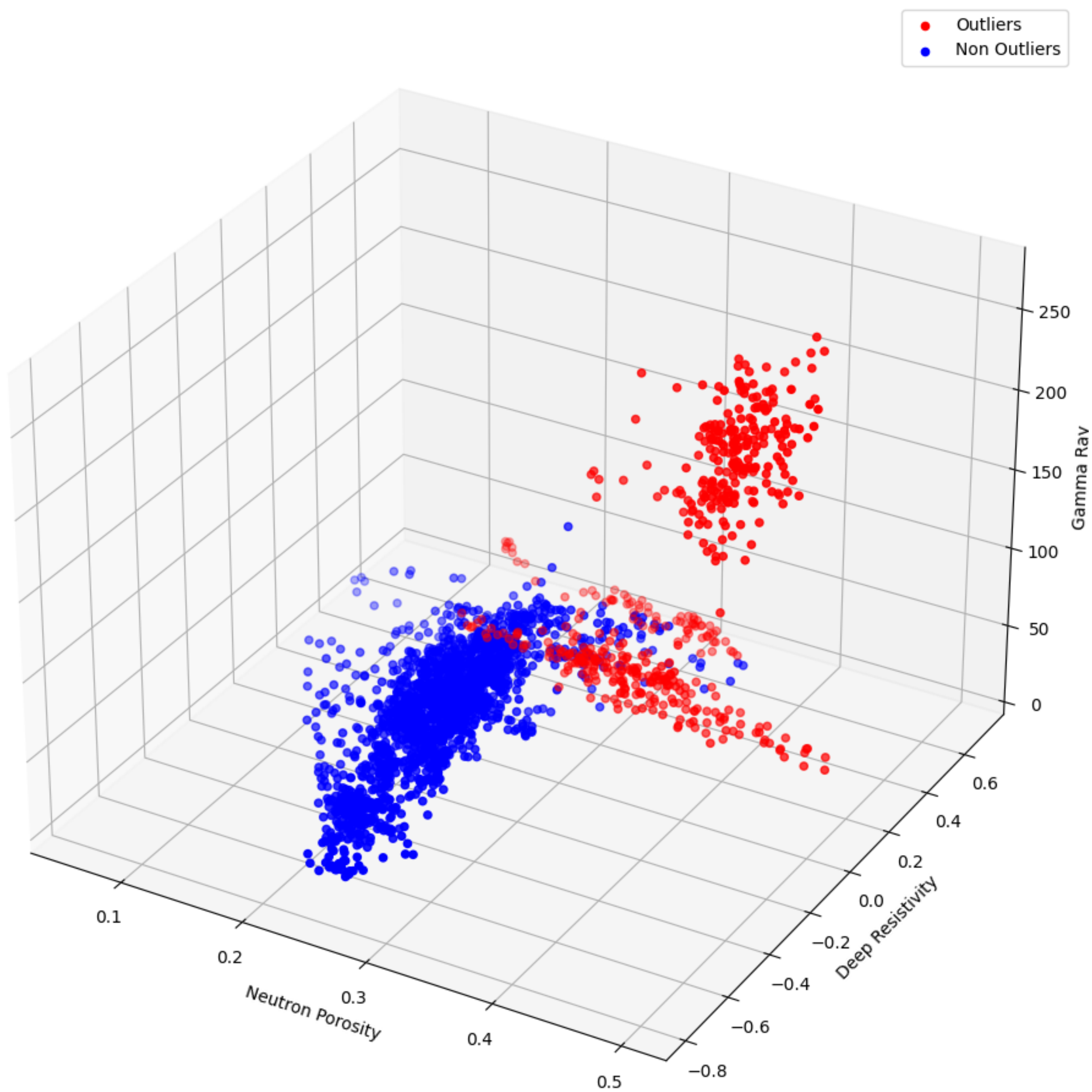
We can see there are so many anomalies but if we tried to extract data from the anomalies using IQR, it won't be possible as the number of anomalies is large. Another way to find anomalies is thresholding the points using the Euclidean distance or the Mahalanobis distance between them which is better for detecting anomalies in a multivariate dataset. As the Mahalanobis distance seeks to measure the correlation between variables and relaxes the assumption of the Euclidean distance, that all variables have the same weight in detecting outliers. Mahalanobis normalizes data before calculating distance.

MAHALANOBIS DISTANCE

same variable

$$M = \sqrt{\frac{(x_1 - \bar{x}_1)^2}{\sigma^2}}$$

We plotted the outliers against some sensor readings trying to understand what the anomalies are:



Observations: From the plot we can see that some outliers have:

- High resistivity → doesn't contain water
- Low gamma ray → may be sandstone
- High Neutron Porosity → not solid

Indicating that these outliers may be gas or oil

Q10: Based on a prediction, what are the sweet spots of the existence of crude oil?

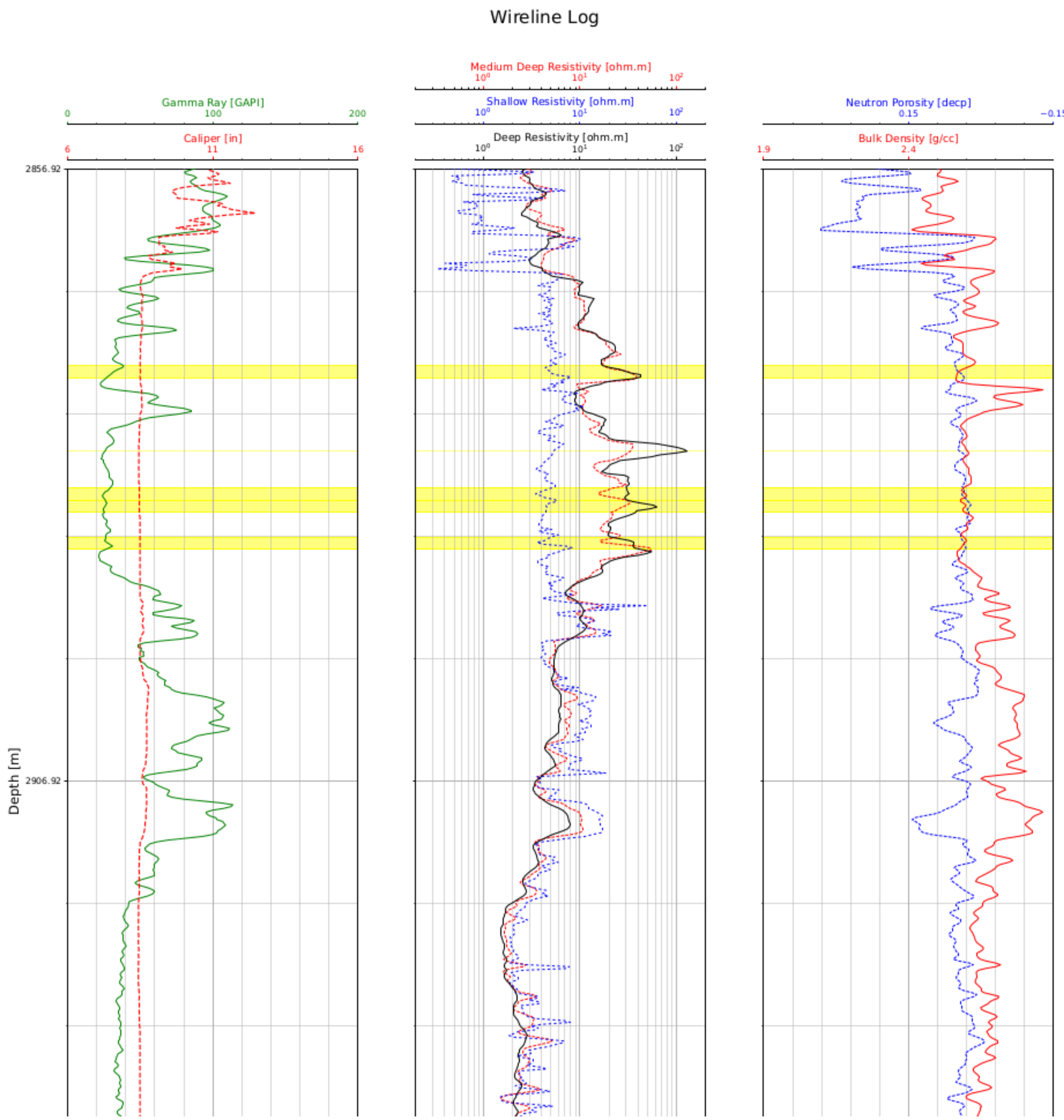
Purpose of the question: interpret the analysis results into a good log plot

Expectations: we can allocate the positions of possible oil or gas existences

Answer: using the facts we extract from the analysis we can say that oil has the following characteristics:

- Exists in sandstone (for it porosity)
- Has low gamma rays
- Has high resistivity
- Has a crossover between neutron porosity and bulk density: this fact is taking from petroleum references and it is based on the fact that a crossover indicates high density with high porosity which means the existence of hydrocarbon

We highlighted the positions with these characteristics in the wireline log and plotted the log in a standard 3 track log



Observation: The highlighted positions have the highest probability of finding oil/gas