Out[2]:

Click here to toggle on/off the raw code.

# Well Exploration

# DEMO BEGIN

# Understanding the training process for Well Top Picking

Well top picking is an essential process in the O&G exploration process.

# 0. Preliminaries

I am a Geoscientist who knows the basics of ML. I have a dataset of 500 well logs (https://www.youtube.com/watch?v=1RJY4hCiWC4 (https://www.youtube.com/watch?v=1RJY4hCiWC4)) and picking tops in each one of them is very time consuming. I know that in the O&G AI Workbench there are ML models for well top picking. I will try to apply them to one of my well logs to see if I can use them to pick the well tops for my dataset.
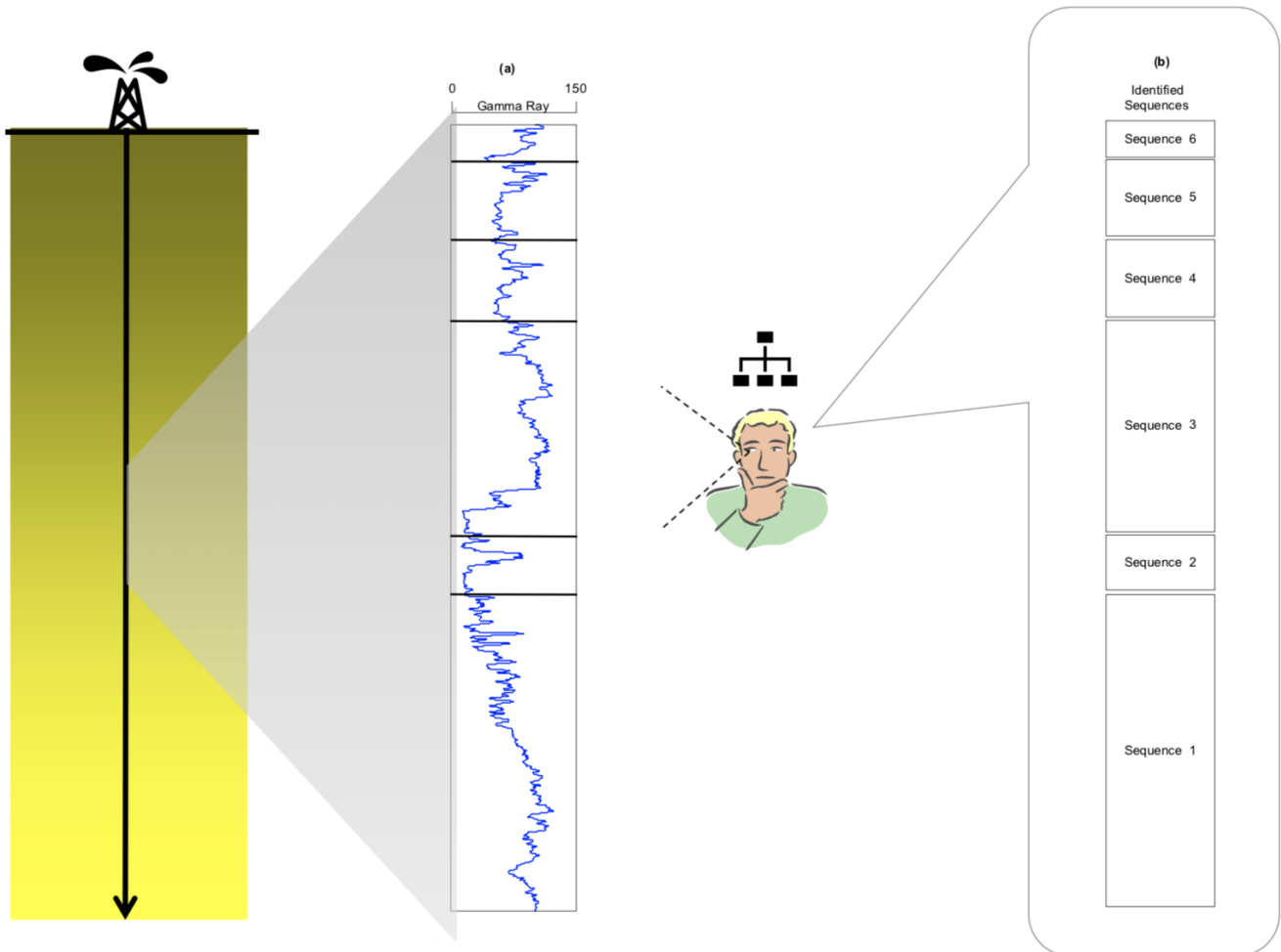
# 1. Analyzing my well log file

By plotting some properties of my well log file. My well log file is:

```
my_well_log_path = "data/W4_065_02_29_12(1AA00).las"
```
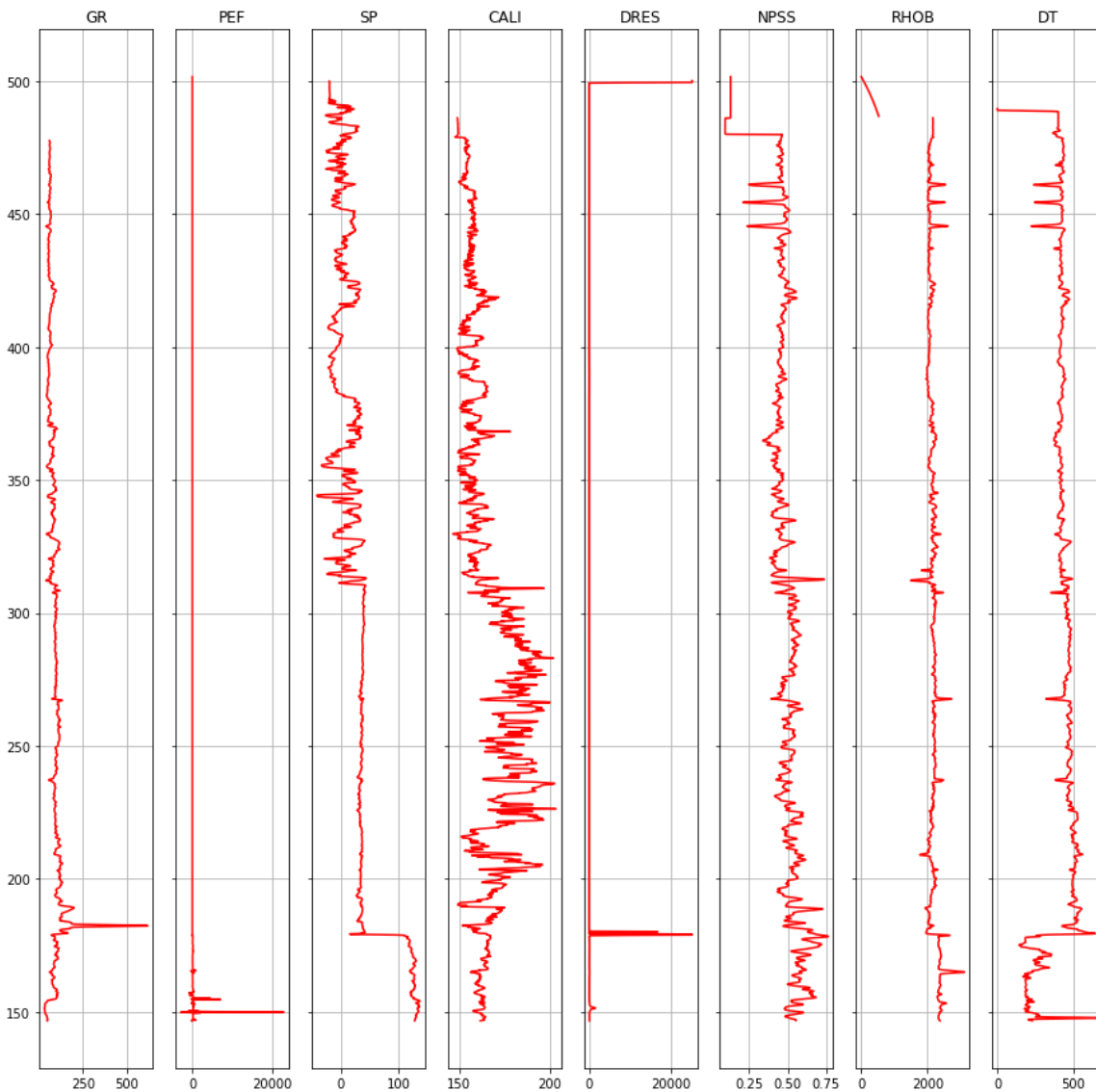


# 2. Retrieving the best ML model for Well Top picking in the IBM O&G AI Workbench platform

Get the model with highest `F1` measure

```
Best model path models/highest_all_f1.h5
```

# 3. Applying the best model to pick well tops in my well log

The apply model function returns the same well log with an additional property with the tops marked.

```
My well logs with tops picked: data/W4_065_02_29_12(1AA00).las.tops.
shf.las
```

# 4. Plotting my well log with the tops picked by the best ML model

----------

----------

----------

----------

----------

----------

----------

----------

----------
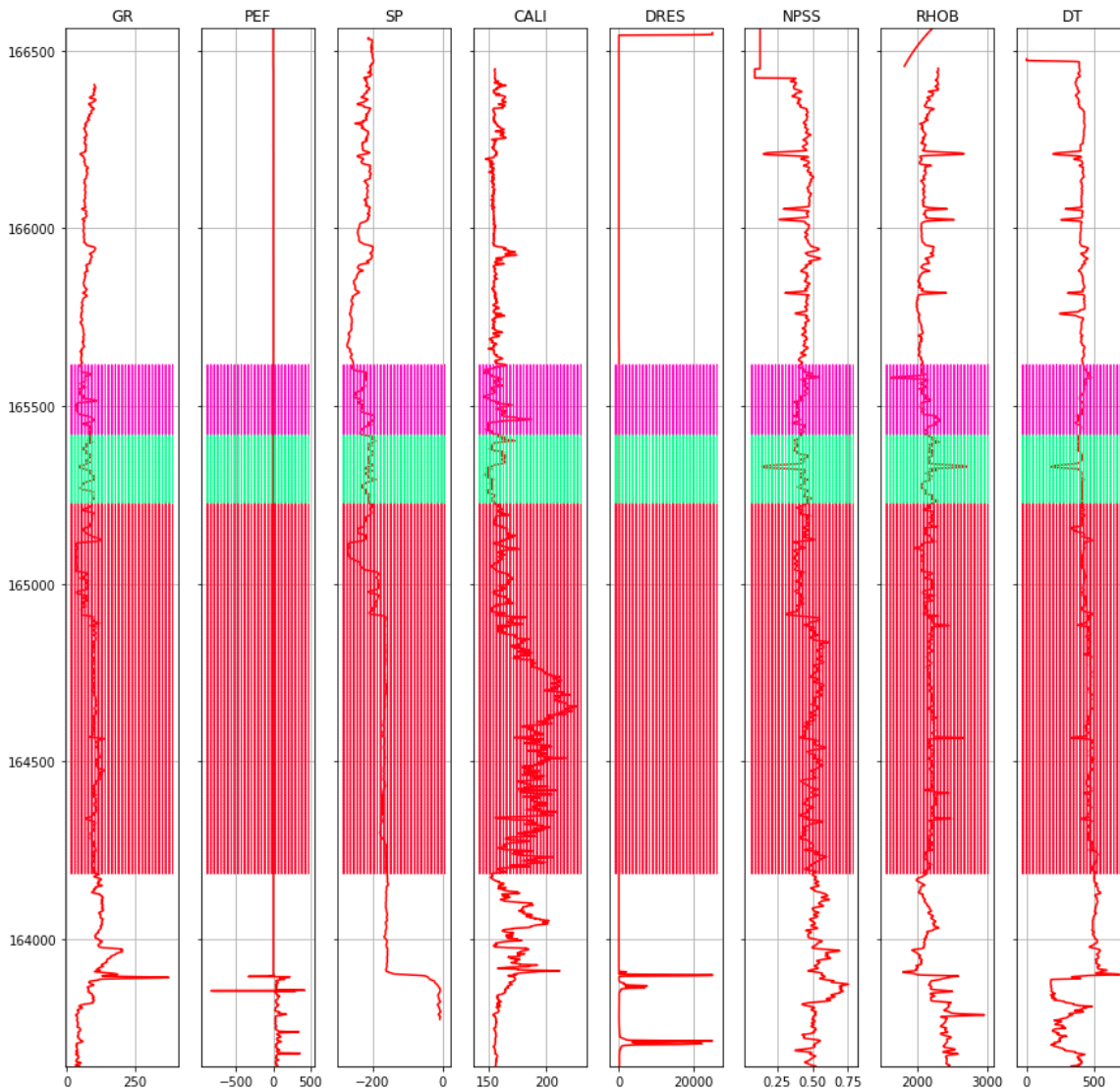
# 5. Trying to understand why the model did not work for my well log

I know that for training any supervised learning ML model, there exists a training dataset and hyperparameters. I'll analyze the dataset to see how the model was trained.
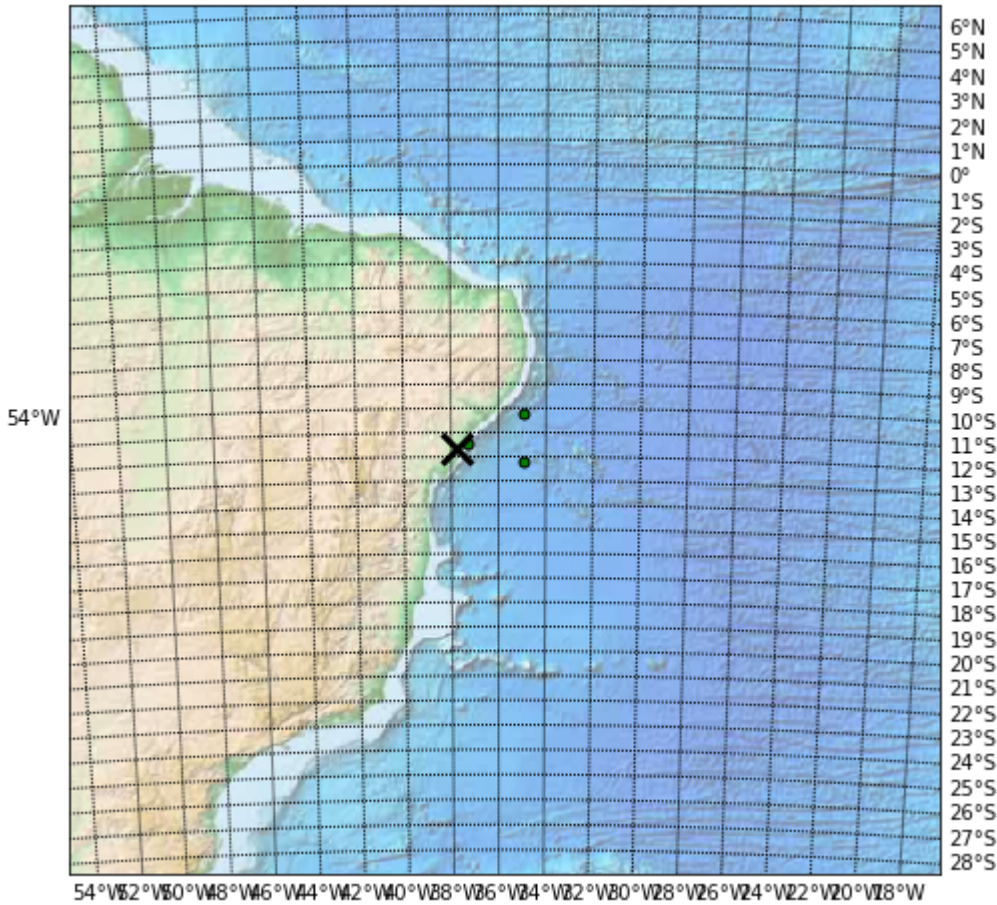
## 5.1 First hypothesis (H1):

**If my well log is far away (geographically) from well logs used in the training dataset, the ML model may not be generalized for my dataset.**

Then, I want to check the geographic position of the well logs used in the training dataset.

## Plotting geographic coordinates of my well log and some of the well logs used to train this model

```
Clipping input data to the valid range for imshow with RGB data
([0..1] for floats or [0..255] for integers).
```



# Conclusion: They are near. Hypothesis H1 is rejected.

## 5.2 Second hypothesis (H2):

**Despite being near, if my well log is within a different Oil basin than the basin of the wells used in the training dataset, the ML model may not be generalized for my dataset.**

I know that **my well logs are located in the `Recôncavo` basin**. Now I want to check the basin of the well logs used in the training dataset.

```
The basin of the training datasets of this trained model is: Sergipe
e Alagoas
```
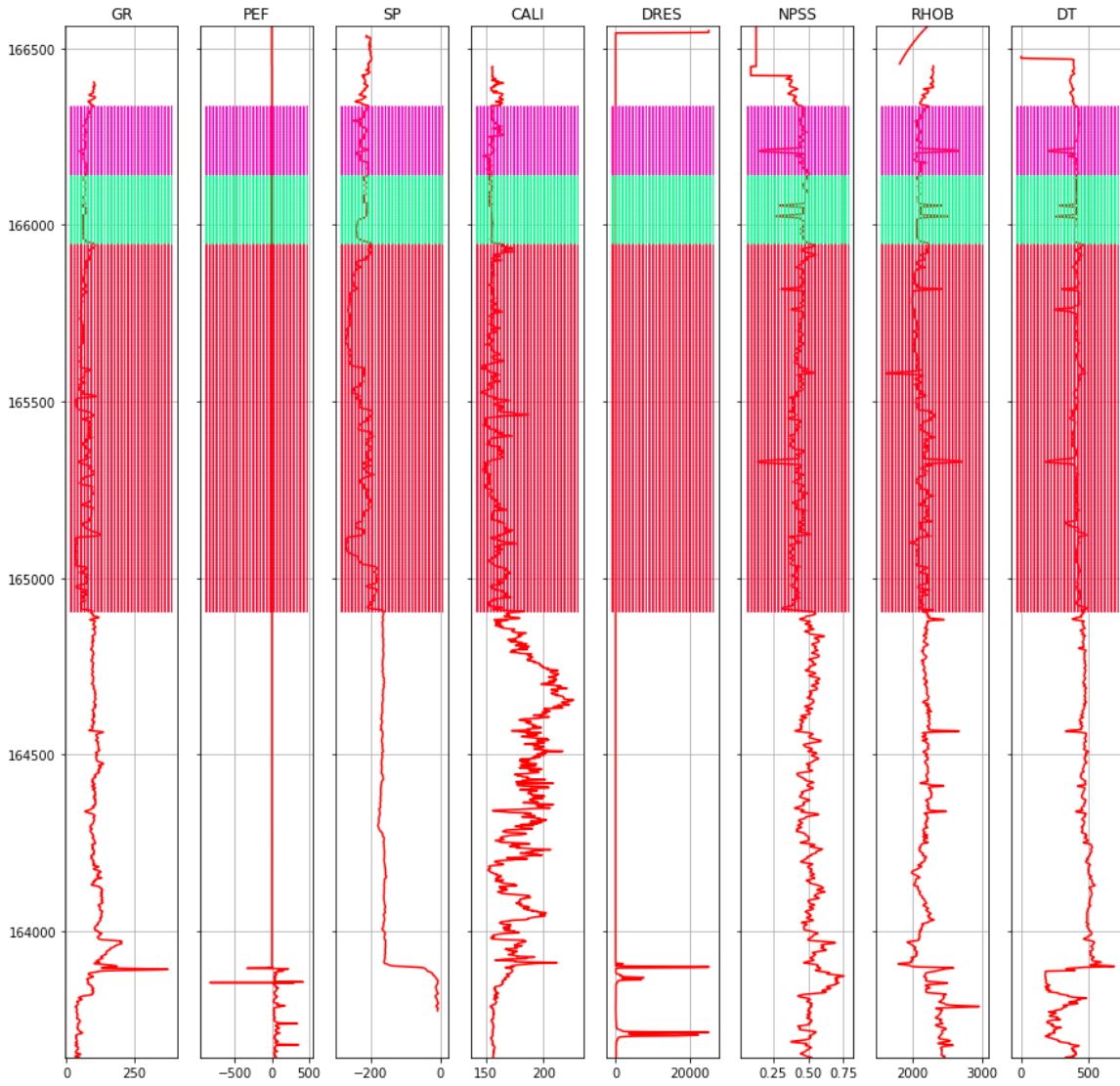
I want to plot the well logs used for the training datasets in the two basins: in the basin of my well log (Reconcavo Baiano, blue) and the basin for the well logs in the other basin (Sergipe-Alagoas, Red)
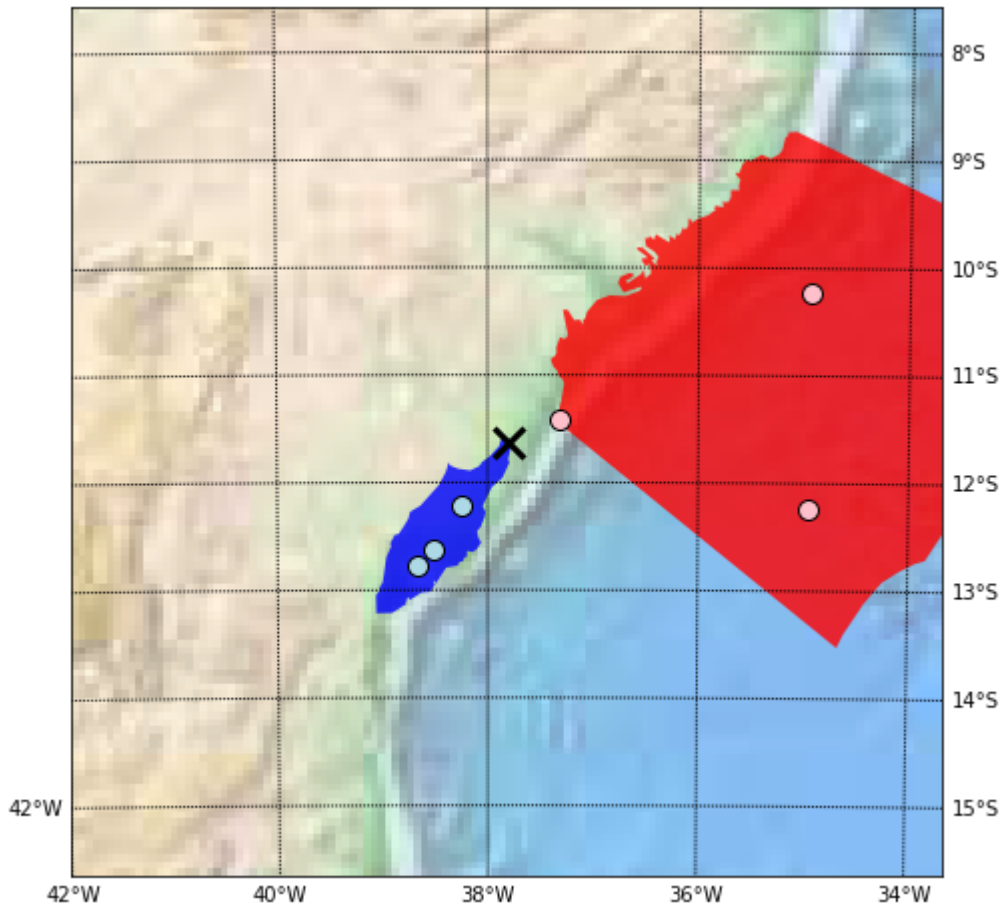


Conclusion: My well log is in a different basin than the well logs used for the best ML model, then the hypothesis is not rejected.

# 6. Retrieving the best ML model *trained with well logs in the same basin* of my well log

# 7. Plotting my well log with the tops picked by the ML model above

----------------------------------------------------------------

# Demo end