# Volcanic Eruption Prediction

## Introduction

**Anticipating volcanic eruptions equivalently to the weather predictions** - Tens of thousands of lives saved, more timely evacuations and the damage mitigated. **Patterns of seismicity are difficult to interpret** - in very active volcanoes, current approaches predict eruptions only some minutes in advance.
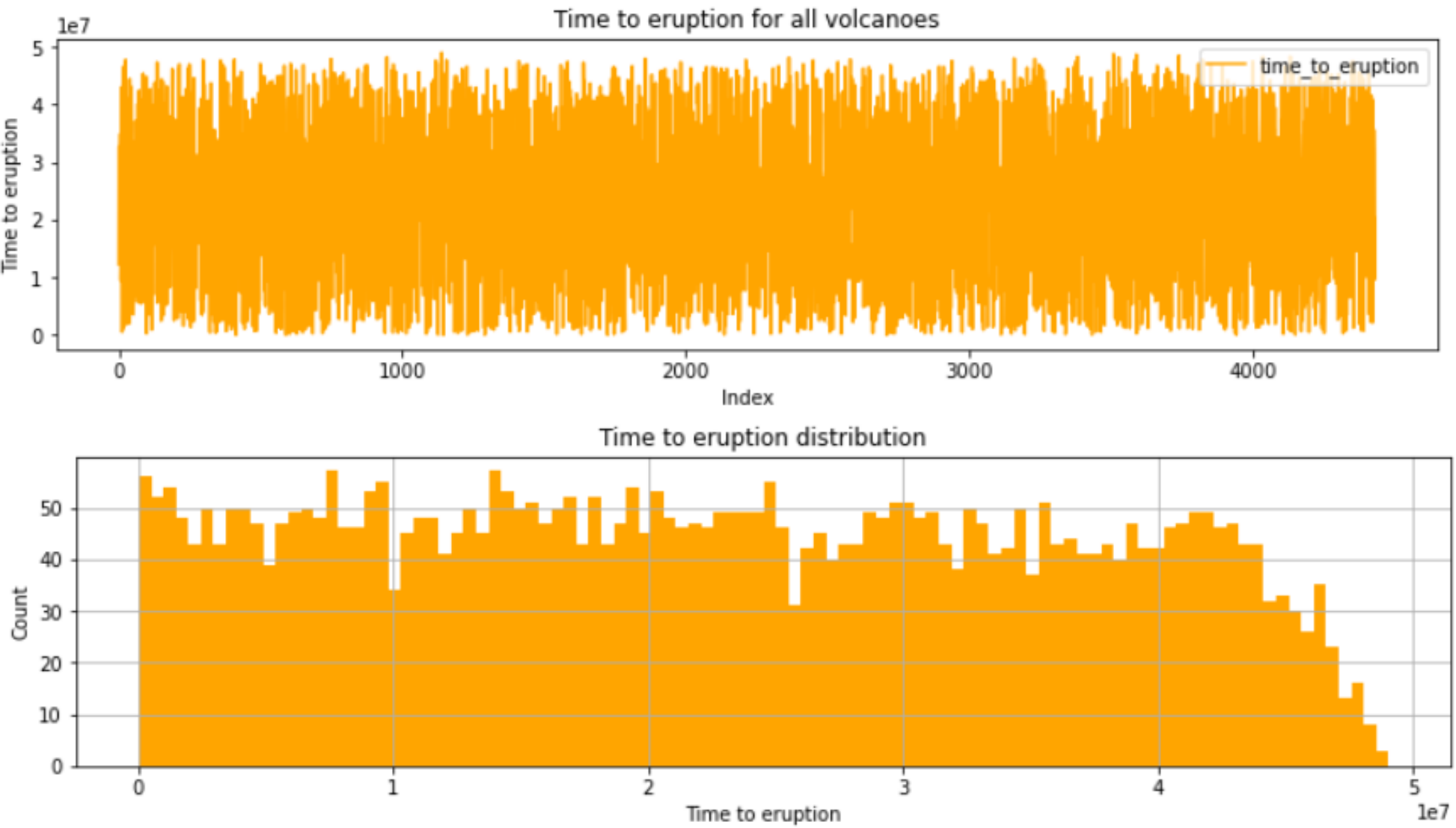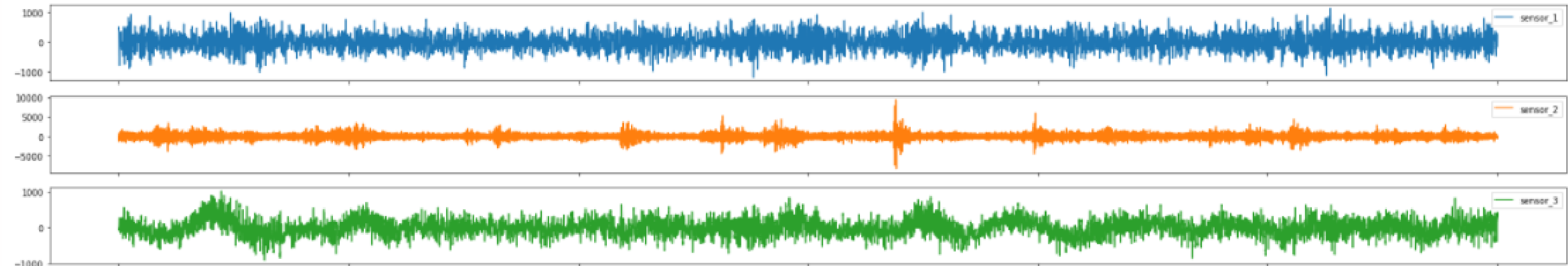
Therefore we are competing in Kaggle's **competition INGV - Volcanic Eruption Prediction** to put our data science skills to the test and try to predict the next eruptions of 8951 volcanoes.

## Geophysical Data

The given data in the competition was already split into training and testing data

**Train.csv** file - ID for the data segment that matches the name of the associated data file and the target value that describes the time until the next eruption

**Test and train datasets** - ten minutes of logs from ten sensors placed around a volcano. There are data files of **8951 volcanoes** and **every file has 60000 rows of seismic data**.

## Approach & Methods

**1.** Understanding the data, plotting, describing etc;

**2.** Systematizing the data - creating dictionaries with sensors' mean, max, min and std, creating the dataframes that we need;

**3.** Predicting - splitting actual training data to train and test (80/20), finding best parameters by randomized search, using the best model for predicting (scoring 'neg_mean_squared_error') **RandomForestRegressor(min_samples_split=8, n_estimators=400)**

## Goals & Results

**Our main goal was to predict when the next eruption of volcanoes would occur based on given geophysical data.** Our secondary goal was to be in the top 50% of the competitors. We acknowledged that the final percent would change, due to our courses project deadline being before the competitions ending date. Therefore we also do not know the correct answers to the predictions. On the 14th of December, **we reached 292th place** (434 teams) **with a score of 7304793, which means that our prediction is about 2h off. Hence we fulfilled the main goal, but did not succeed in getting to the top 50%.**



| | sensor_1 | sensor_2 | sensor_3 | sensor_4 | sensor_5 | sensor_6 | sensor_7 | sensor_8 | sensor_9 | sensor_10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 449 | 199 | -56 | -214 | 156 | 102 | -104 | -277 | 451 | -1116 |
| 1 | 465 | -123 | -47 | -202 | 24 | 104 | -119 | -179 | 464 | -1107 |
| 2 | 485 | -334 | -36 | -224 | 39 | 58 | -132 | -373 | 464 | -1070 |
| 3 | 520 | 124 | -20 | -199 | 10 | 0 | -147 | -360 | 459 | -1004 |



| 292 | IDS-group4-A&M |
|---|---|
| Your Best Entry ⬆ | |
| Your submission scored 7304793, | |

**Agnes Peetrik & Maiken Veedla**