**-Predicting Earthquakes with data science**

**Exploratory data analysis**

The dataset is rather large, with approximately 10 GB of training data, so the dataset was quite large (Dask is very useful in such situations where your computer's RAM is not enough to look at the dataset). Training data consists of 630 million notes with only two features: audio data, which is seismic signal, and failure time, which is what we are trying to predict.

The total acoustic data in the training set. Each of the large spikes represents an earthquake event, and there are about 16 earthquake events in the training set.

**Feature Extraction**

After preprocessing the data into segments, how do we actually solve this regression problem with a single acoustic feature? Problems like this are common in the real world, where data scientists try to make predictions from a given time series or detect signals in the time series. In either case, we have to find features from the time series to perform any kind of regression problem. Some of the features are simple aggregate features like the mean of a segment, IQR, standard deviation, and etc. However, to get better results, we also have to use more complicated features that are prevalent in time series analysis.