Dataset: quake can be imported via library bds\_courseware:

pip install bds\_courseware

from bds\_courseware import read\_drive\_dataset

from bds\_courseware import print\_dataset\_description, print\_module\_datasets

from bds\_courseware import HOMEWORK\_DATASETS

print("Dataset names: ", HOMEWORK\_DATASETS.keys())

name = "quake"

df = read\_drive\_dataset(\*HOMEWORK\_DATASETS[name])

print(df.head(2))

print(df.shape)

The dataset is also available via link:

<https://drive.google.com/file/d/1S5stL1xz51y5QkGRHHfWqteacHDYOYXM/view?usp=sharing>

It contains information about earthquakes: coordinates, depth and strength. Your task is to cluster the data:

**Evaluation criteria (100% in total):**

* Exploratory analysis (there are some specific characteristics of features, which may change the result of the analysis)  **(10%)**
* Build K-Means with 15 clusters **(10%)**
* Determine optimal number of clusters for K-Means **(10%)**
* Create visualization for the obtained clusters **(10%)**
* Try other clustering algorithms; tune their hyperparameters to achieve better results. You should use at least two different algorithms **(20%)**
* Use labelled and unlabeled metrics to estimate quality of clusters you built. As ground truth use cluster labels from K-Means, choose most similar algorithm **(20%)**
* Explain the final choice of best clusterizations: give interpretation of clusters **(10%)**
* Visualize the best clusterizations in your opinion on the world map. **(10%)**