Overview of K-Means Example

Dataset Link:

<https://www.kaggle.com/sns224/geyser>

Dataset Content:

This dataset contains two columns. It gives information about the duration of eruptions (in minutes) and the interval between two eruptions (in minutes) from the famous Old Faithful hot water geyser in Yellowstone National park.

Problem:

We would like to make clusters of time series of eruptions in Old Faithful geyser in order to better understand the varying behavior upon eruption in different situations.

Solution:

We want to develop a clustering model using K-Means algorithm which allows us to cluster eruptions into different groups based on the duration of eruptions and the interval between eruptions.

Variables:

Independent variable🡪 Duration of eruptions, Interval between eruptions.

Dependent variable🡪 Unsupervised Learning technique (No dependent variable)