# Task : Clustering based identification

Data: Date and time, OCHL, volume , open interest

## Deciding Features (X):

- ☐ Normalization of position of the OHCL candle:
  - ☐ Find the moving mean of past n candles (close, open , 0.5(C+O))
  - ☐ (O, C, H, L ) - moving mean
  - ☐ Vedika's idea…….
- ☐ Moving mean
- ☐ Gradients 1st, 2nd …. Of moving mean
- ☐ Volume …. Their gradients
- ☐ Differences between OHLC <> OHCL
- ☐ Time of candle (hour + min)
- ☐ Week day
- ☐ Vedika's Idea…….

- ☐ Clustering part: k-mean :
- ☐ H-clustering
- ☐ Latent space VAEs

Let's say we get k no of clusters
KPI: identify the clusters with better Y var.

## Deciding the Y var. (close - open)

- ☐ Occurrence of Green/ Red in  n future candles (categorical type)
  - How many time it is green and red, neutral
  - Also decide the green/red based on threshold
- ☐ Amount of change in the open and close - for n future candles (numerical)

Required a table:

- Cluster number - Y_cat  - Y_num