Price definitions for additional/richer analysis

1 - Intraday range (normalized): This uses the High and Low daily price in OHLC data and normalizes using the Close price. You can use Adjusted versions of the prices here.

$$range\_pct_t = \frac{High_t - Low_t}{Close_t}$$

You can also add the 30-day mean and standard deviation to use with the above range value:

Mean intraday range:

$$\mu_{range} = \text{mean}(range\_pct_1..range\_pct_{30})$$

Std of intraday range:

$$\sigma_{range} = \text{std}(range\_pct_1..range\_pct_{30})$$

Advantage of using the above is the following two classes can be separated more effectively:

- **STATIONARY** (look for low mean and low std)
- **OSCILLATING** (look for higher range metrics)

Also noisy periods in the data will manifest themselves with high std values

2 - Overnight gap (normalized): This compares the Opening price on a day with the Closing price of the previous day as below. Log gap is also consistent with log returns and hence the ln() below.

$$gap\_pct_t = \ln\left(\frac{Open_t}{Close_{t-1}}\right)$$

The above metric can also be aggregated across a 30-day window as follows:

Mean absolute gap:

$$\mu_{|gap|} = \text{mean}(|gap\_pct_2..gap\_pct_{30}|)$$

Std gap:

$$\sigma_{gap} = \text{std}(gap\_pct_2..gap\_pct_{30})$$

Gap frequency (tail event rate): Given that we have defined a threshold ‘g’, which is the 90th percentile of |gap_pct| across all windows/days, then:

$$f_{gap} = \frac{\#\{t : |gap\_pct_t| \geq g^{\backslash *}\}}{29}$$

The above calculations help distinguish “oscillation” caused by frequent overnight jumps vs intraday reversals, and can push ambiguous windows into **OTHER** class.

3 - Volume-derived context metrics (optional): Because volume is often heavy-tailed, log() volume is used as follows. The following metrics give additional context about regime changes and abnormal trading activity without requiring finance assumptions.

$$v_t = \ln(Volume_t + 1)$$

Mean log volume:

$$\mu_v = \text{mean}(v_1..v_{30})$$

Volume volatility:

$$\sigma_v = \text{std}(v_1..v_{30})$$

Volume spike rate:

$$f_{vspike} = \frac{\#\{t : v_t \geq v^{*}\}}{30}$$