1. Use S&P 500 futures tick data taken from

https://www.kaggle.com/datasets/finnhub/sp-500-futures-tick-data-sp. Save it as SP.csv. Remove the rows with 0 volume and then proceed.

A screenshot of a black screen

AI-generated content may be incorrect.

# Convert time column to string format first, then combine

sp['date\_time'] = pd.to\_datetime(sp['date'].astype(str) + ' ' + sp['time'].astype(str))

sp\_processed = sp[['date\_time', 'price', 'volume']].copy()

sp\_processed.columns = ['date\_time', 'price', 'volume']

from mlfinlab.data\_structures import time\_data\_structures

time\_bars = time\_data\_structures.get\_time\_bars(

    sp\_processed, resolution="D", verbose=False

)

A graph showing the growth of the stock market

AI-generated content may be incorrect.2. Form dollar bars for the data from Exercise 1 above.

from mlfinlab.data\_structures import standard\_data\_structures

dollar\_bars = standard\_data\_structures.get\_dollar\_bars(

    sp\_processed, threshold=1000000, batch\_size=100000, verbose=False

)

A graph showing a line graph

AI-generated content may be incorrect.

1. Apply a symmetric CUSUM filter (Chapter 2, Section 2.5.2.1) where the threshold is the standard deviation of daily returns (Snippet 3.1).

A screen shot of a computer code

AI-generated content may be incorrect.

Snippet 2.4 modified to take the series of emwa vols

A screen shot of a computer program

AI-generated content may be incorrect.

dv = getdailyVol(time\_bars['close'], span0=100)

tevents = getTEvents(time\_bars['close'], dv)

I don’t have a good idea how to align the index of daily emwa vols with Dollar Bars.

I did t events for time bars instead.

Charting time bars vs the T-Events series for the last 100 Time Bars

A graph with red and blue lines

AI-generated content may be incorrect.

I have snippets 3.2 Triple Barrier, 3.5 labeling and Size and 3.8 Dropping Under-populated Labels in GitHub

========LOOKS LIKE I’LL NEED A BIT LONGER TO UNWRAP THIS===========

(b) Use Snippet 3.4 on a pandas series t1, where numDays=1.

(c) On those sampled features, apply the triple-barrier method, where ptSl=[1,1]

and t1 is the series you created in part b.

(d) Apply getBins to generate the labels.

3. On data from Exercise 1 above, use Snippet 3.8 to drop rare labels