Snapshot of the algorithm and data collected for the RNN-LSTM with attention

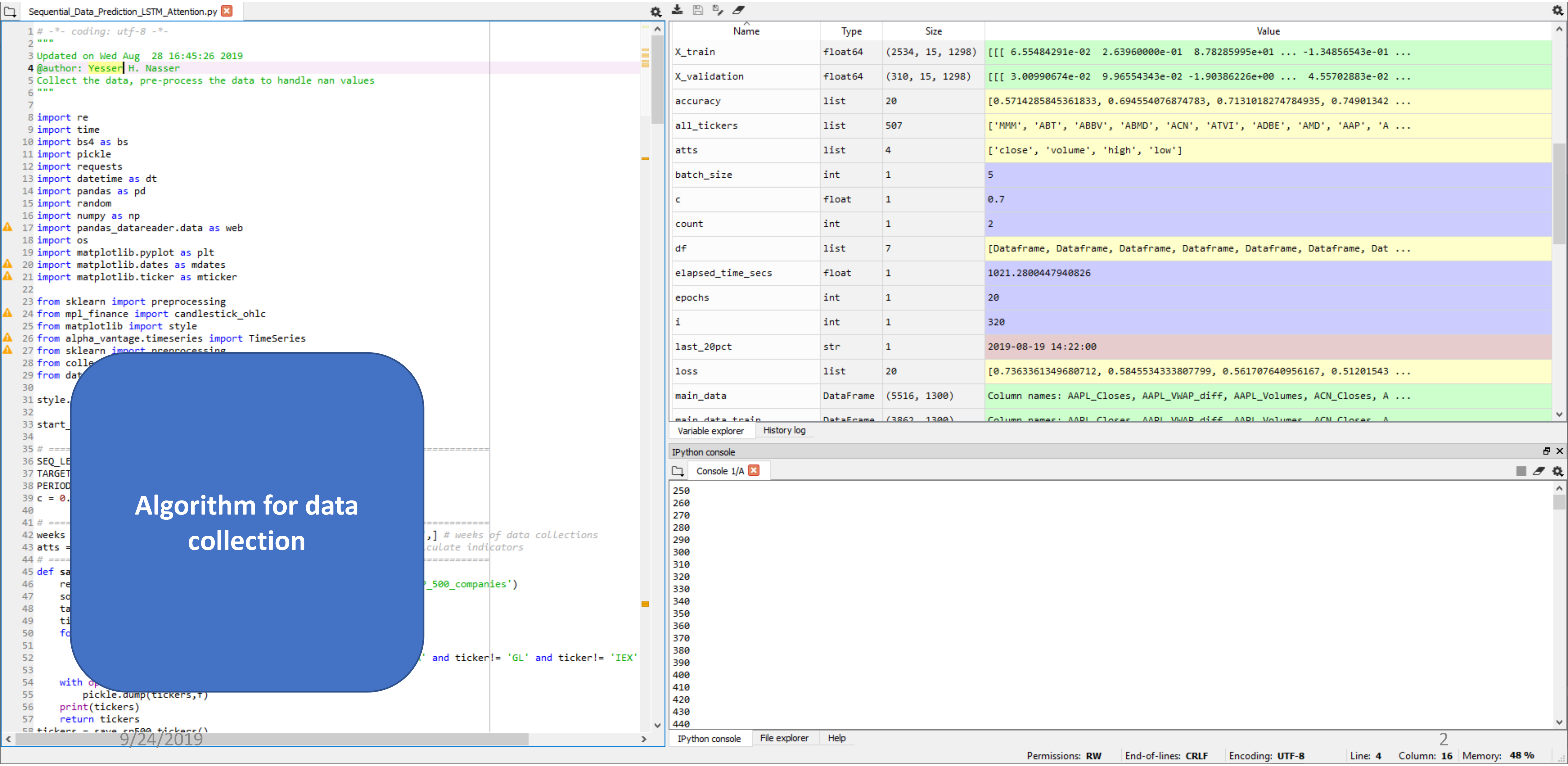
Ongoing Application: High Frequency Quant Trading strategies using machine learning.

Author: Yesser Nasser, PhD

Quantitative Data Scientist – Machine Learning Engineer

Yesser.Nasser@icloud.com

Algorithm Part 1: data collection



Sequential_Data_Prediction_LSTM_Attention.py

```python
    return tickers
tickers = save_sp500_tickers()

# append additonal tickers
symbols_1 = ['SPY','IWM','DIA','IEF','TLT','GLD','SLV','USD',]
for symbol in symbols_1:
    if symbol not in tickers:
        tickers.append(symbol)

def preprocess_price_volume (df,Keep,week):
    df = df.interpolate()
    if Keep == 'volume':
        df = df.round(0)
        df = df.fill
    else:
        for co
            d
            m
            (                                    ), data_inter[~mask])

    df.colum
    print(df
    df.to_cs

# =========
#========= o
def compile_
    with ope

        main_
        #sto
        for
                                                 all the no-alphabet i


                                                 ker name





            i

        preproce

# ====================== execute the function ==============================
for week in weeks:
    To_Keep = ['close','volume', 'high', 'low', 'open']
    for tokeep in To_Keep:
        compile_data(tokeep,week)
## ========================================================================
##=========================== unit test ====================================
```

- Compiling data
- Building Dataframes
- Sequences

Name	Type	Size	Value
X_train	float64	(2534, 15, 1298)	[[[6.55484291e-02 2.63960000e-01 8.78285995e+01 ... -1.34856543e-01 ...
X_validation	float64	(310, 15, 1298)	[[[3.00990674e-02 9.96554343e-02 -1.90386226e+00 ... 4.55702883e-02 ...
accuracy	list	20	[0.5714285845361833, 0.694554076874783, 0.7131018274784935, 0.74901342 ...
all_tickers	list	507	['MMM', 'ABT', 'ABBV', 'ABMD', 'ACN', 'ATVI', 'ADBE', 'AMD', 'AAP', 'A ...
atts	list	4	['close', 'volume', 'high', 'low']
batch_size	int	1	5
c	float	1	0.7
count	int	1	2
df	list	7	[Dataframe, Dataframe, Dataframe, Dataframe, Dataframe, Dataframe, Dat ...
elapsed_time_secs	float	1	1021.2800447940826
epochs	int	1	20
i	int	1	320
last_20pct	str	1	2019-08-19 14:22:00
loss	list	20	[0.7363361349680712, 0.5845534333807799, 0.561707640956167, 0.51201543 ...
main_data	DataFrame	(5516, 1300)	Column names: AAPL_Closes, AAPL_VWAP_diff, AAPL_Volumes, ACN_Closes, A ...
main_data_train	DataFrame	(3862, 1300)	Column names: AAPL_Closes, AAPL_VWAP_diff, AAPL_Volumes, ACN_Closes, A

Variable explorer | History log

IPython console

Console 1/A

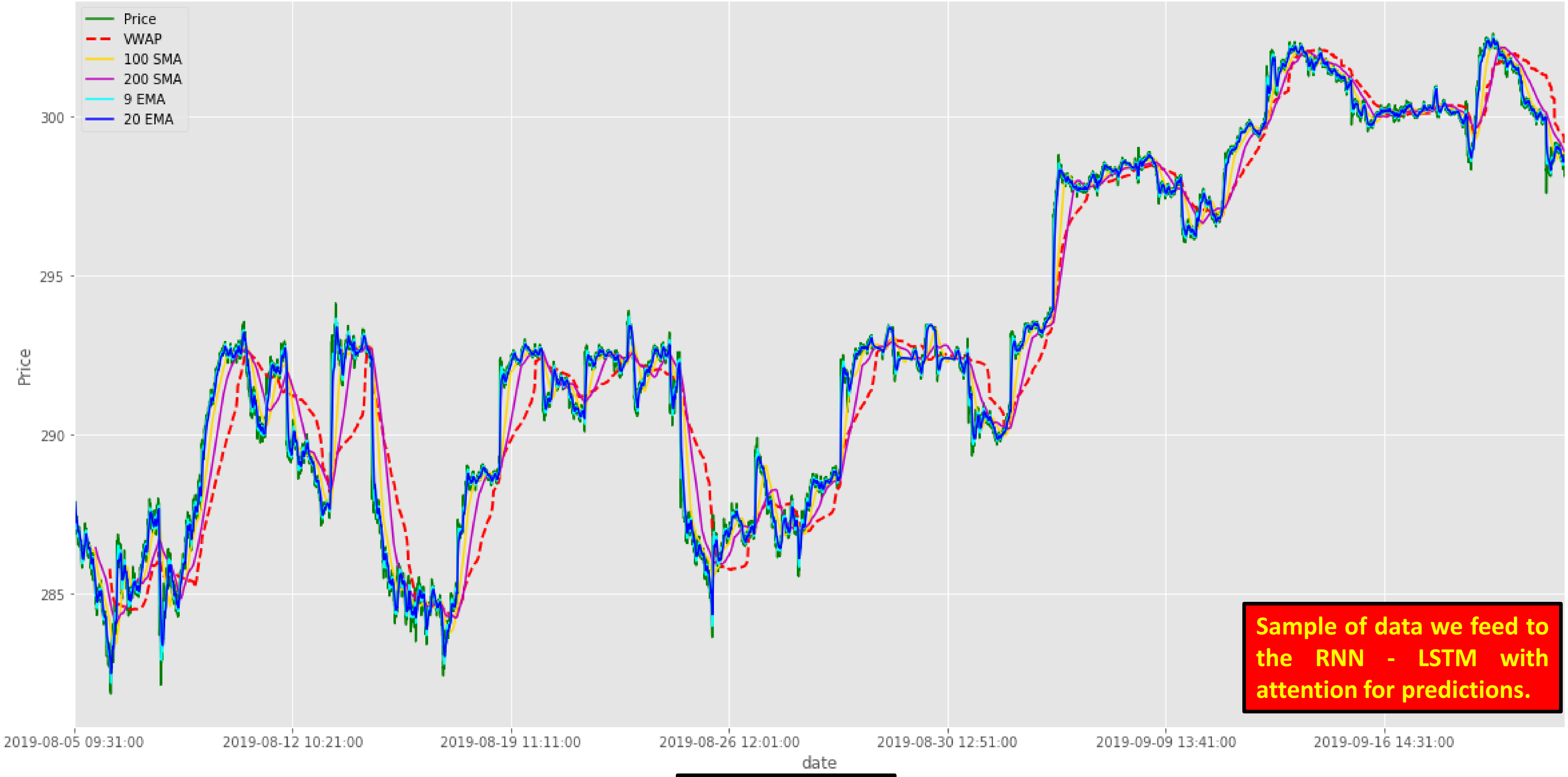
```
460
470
480
490
500
                     MMM_low  ABT_low  ABBV_low  ...   GLD_low  SLV_low    USD_low
date                                             ...
2019-08-12 09:31:00   163.02    86.11     65.20  ...  141.7165  15.8500  39.850000
2019-08-12 09:32:00   163.00    86.11     65.26  ...  141.6550  15.8450  40.037300
2019-08-12 09:33:00   162.68    86.21     65.24  ...  141.6600  15.8499  39.965475
2019-08-12 09:34:00   162.25    86.24     65.23  ...  141.7203  15.8501  39.893650
2019-08-12 09:35:00   162.60    86.24     65.08  ...  141.7495  15.8500  39.821825

[5 rows x 507 columns]
0
10
20
30
40
50
```

IPython console | File explorer | Help

Permissions: RW End-of-lines: CRLF Encoding: UTF-8 Line: 4 Column: 16 Memory: 49 %

SPY

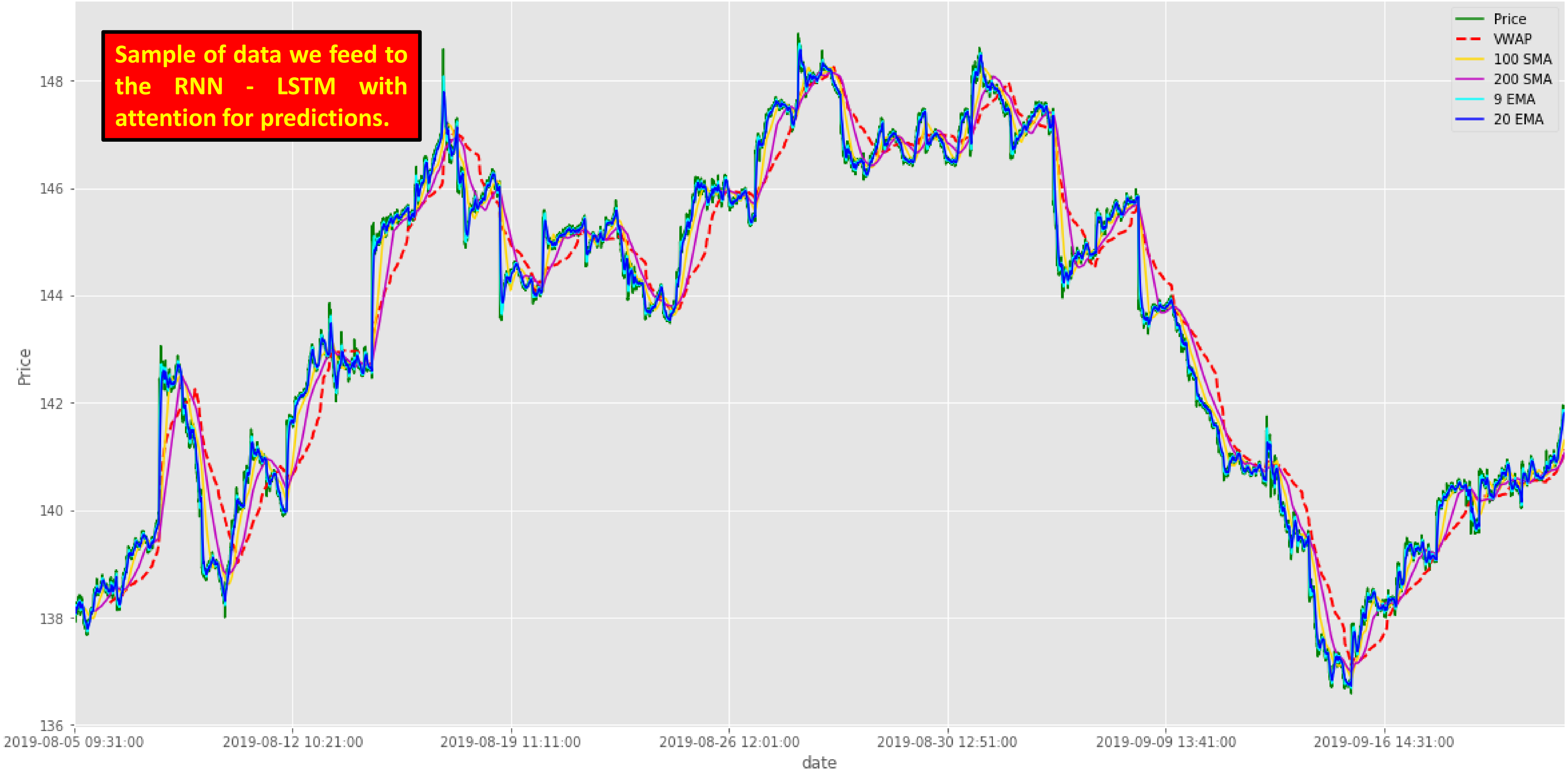


Sample of data we feed to the RNN - LSTM with attention for predictions.

Very resent data!

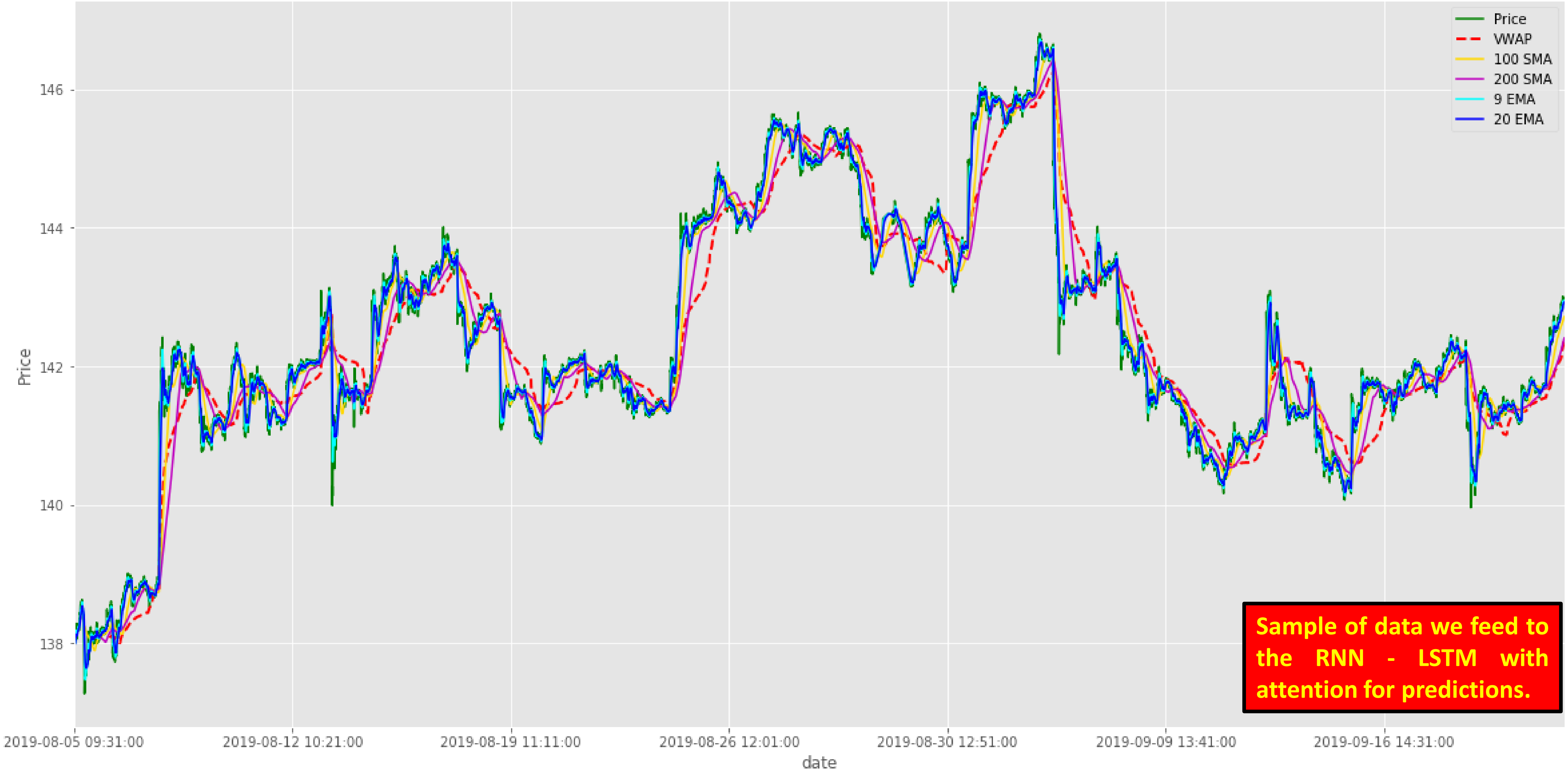
N.B: Time frame : 1 min

TLT



N.B: Time frame : 1 min

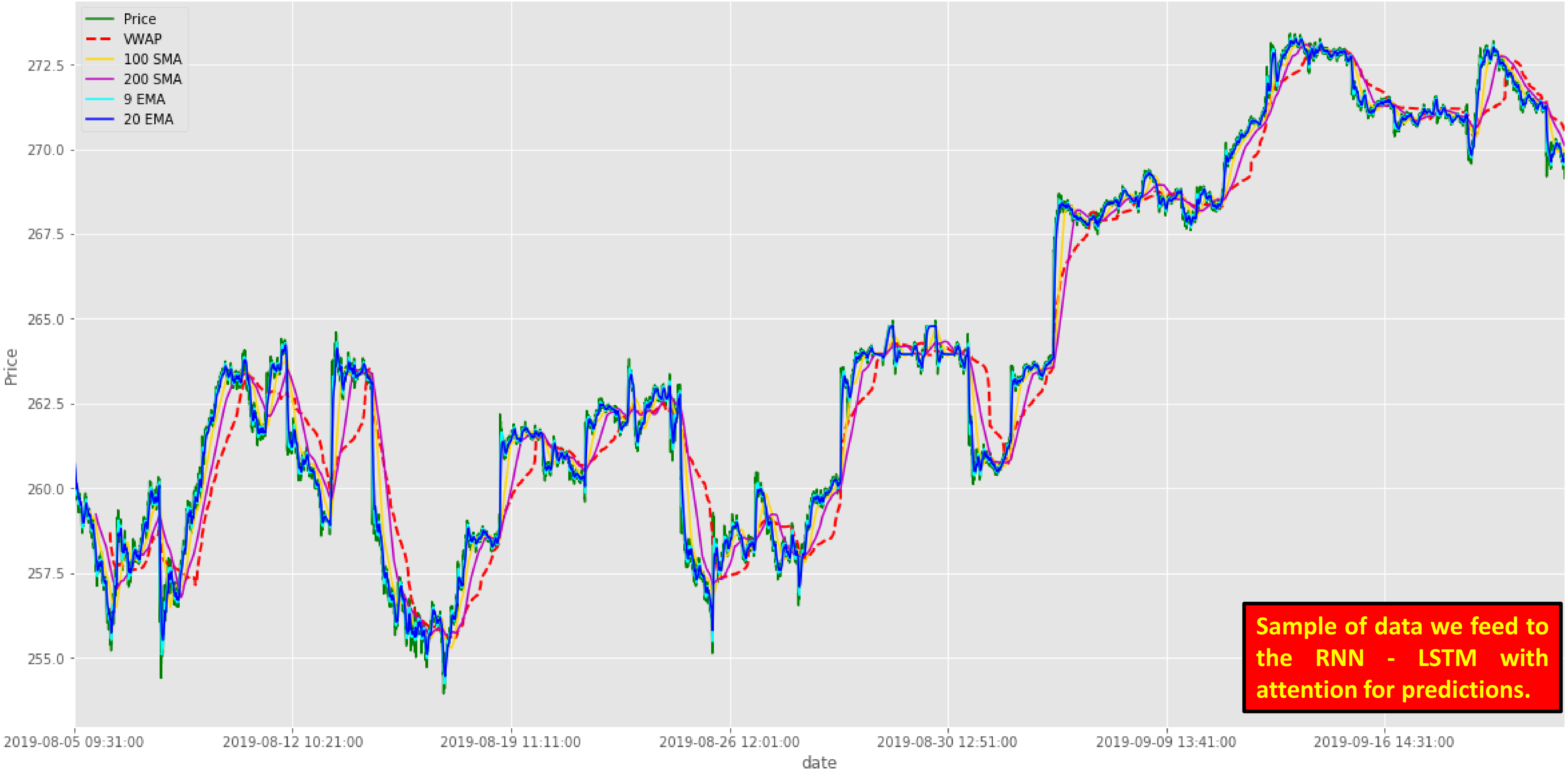
GLD



Sample of data we feed to the RNN - LSTM with attention for predictions.

N.B: Time frame : 1 min

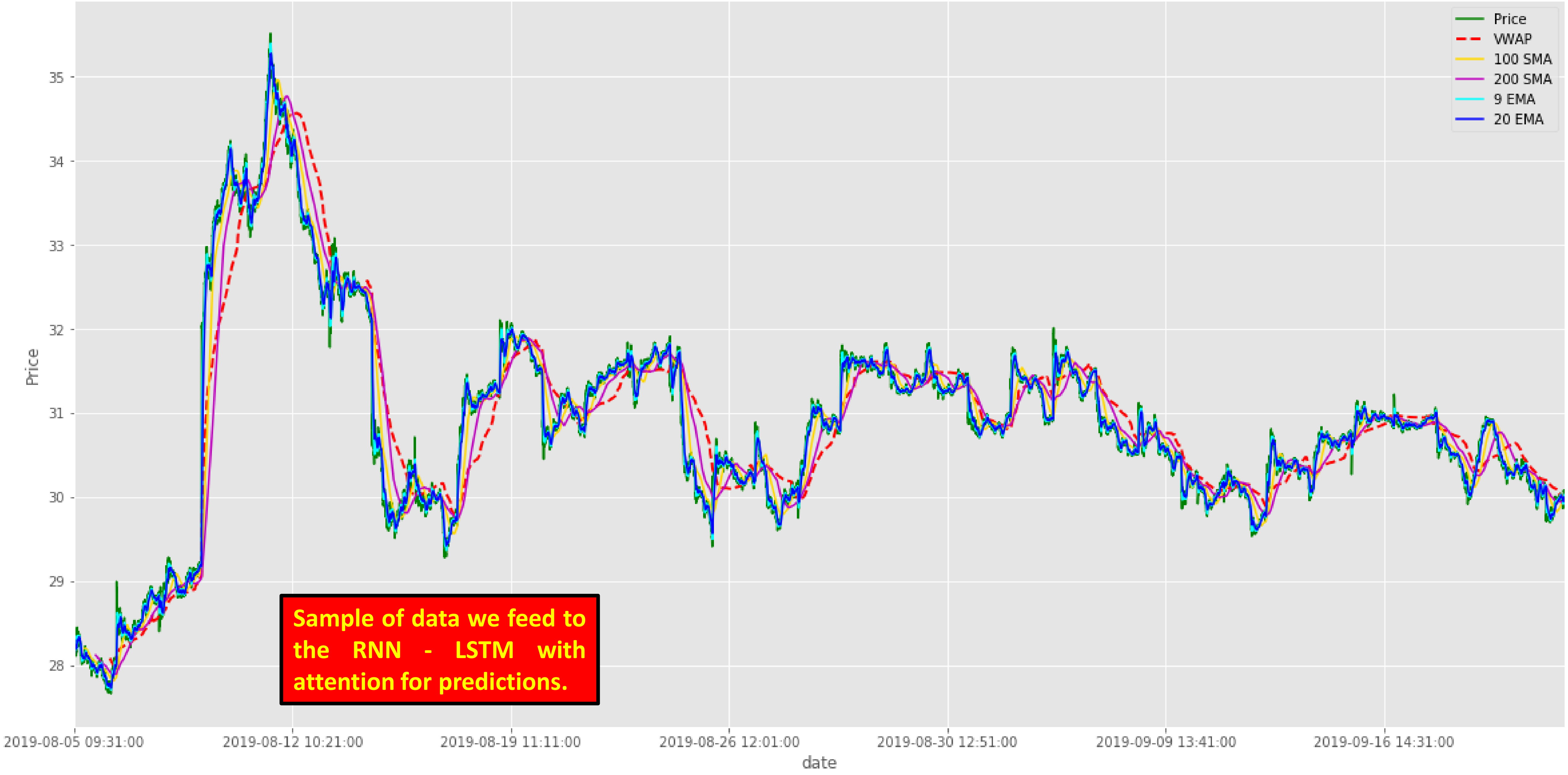
DIA



Sample of data we feed to the RNN - LSTM with attention for predictions.

N.B: Time frame : 1 min

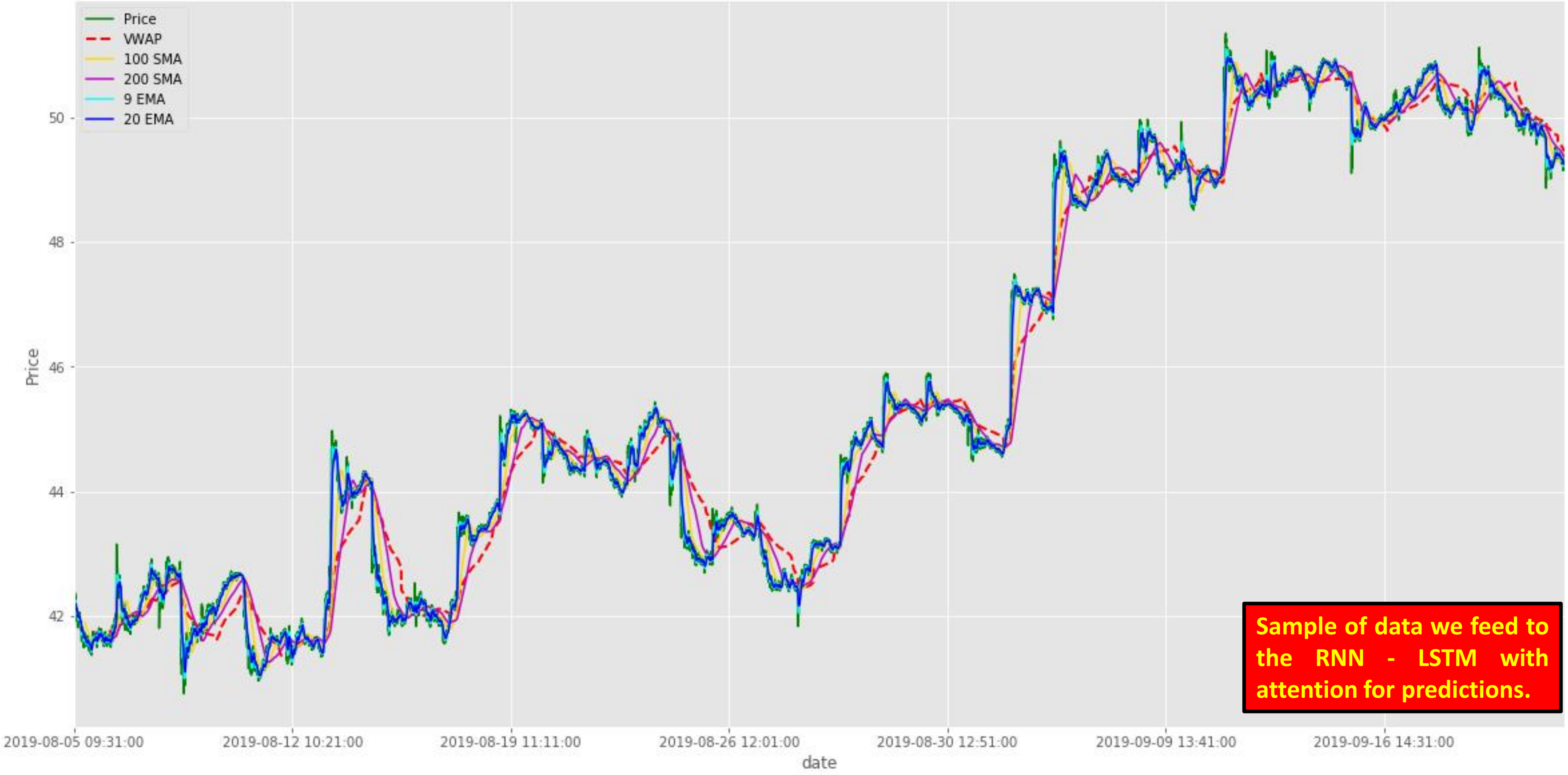
Stock Price - AMD



Sample of data we feed to the RNN - LSTM with attention for predictions.

N.B: Time frame : 1 min

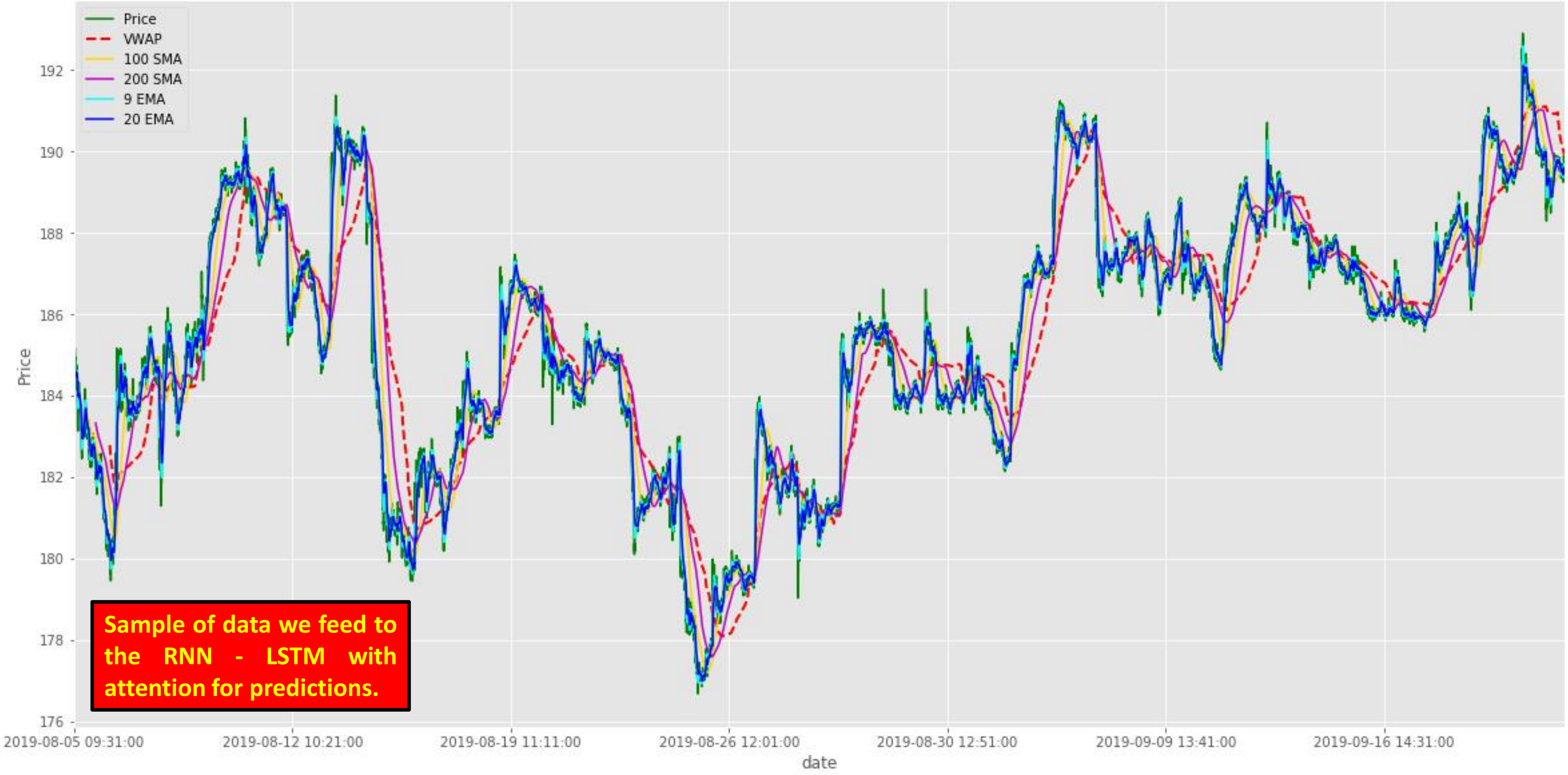
Stock Price - MU



Sample of data we feed to the RNN - LSTM with attention for predictions.

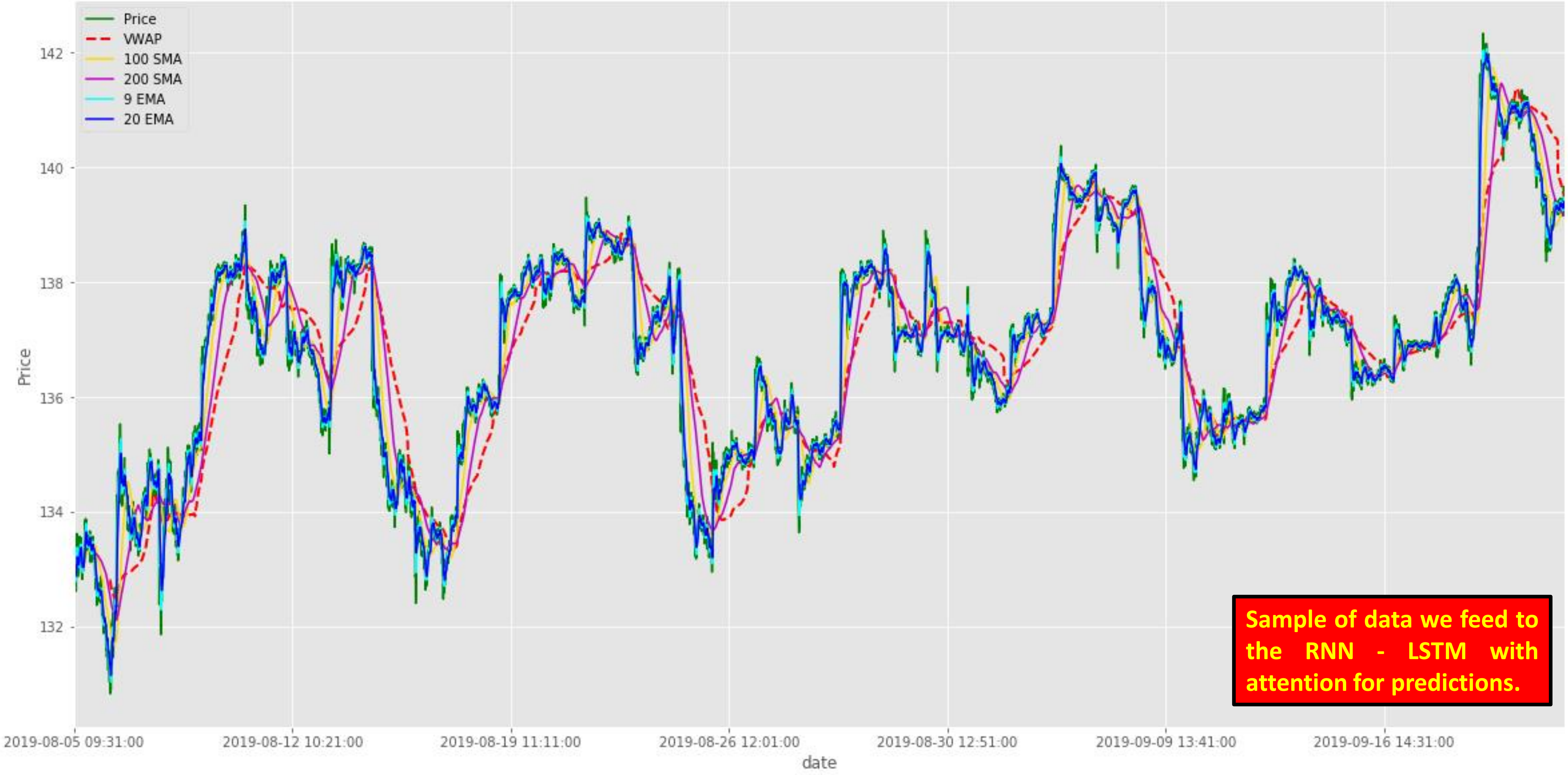
N.B: Time frame : 1 min

Stock Price - FB



N.B: Time frame : 1 min

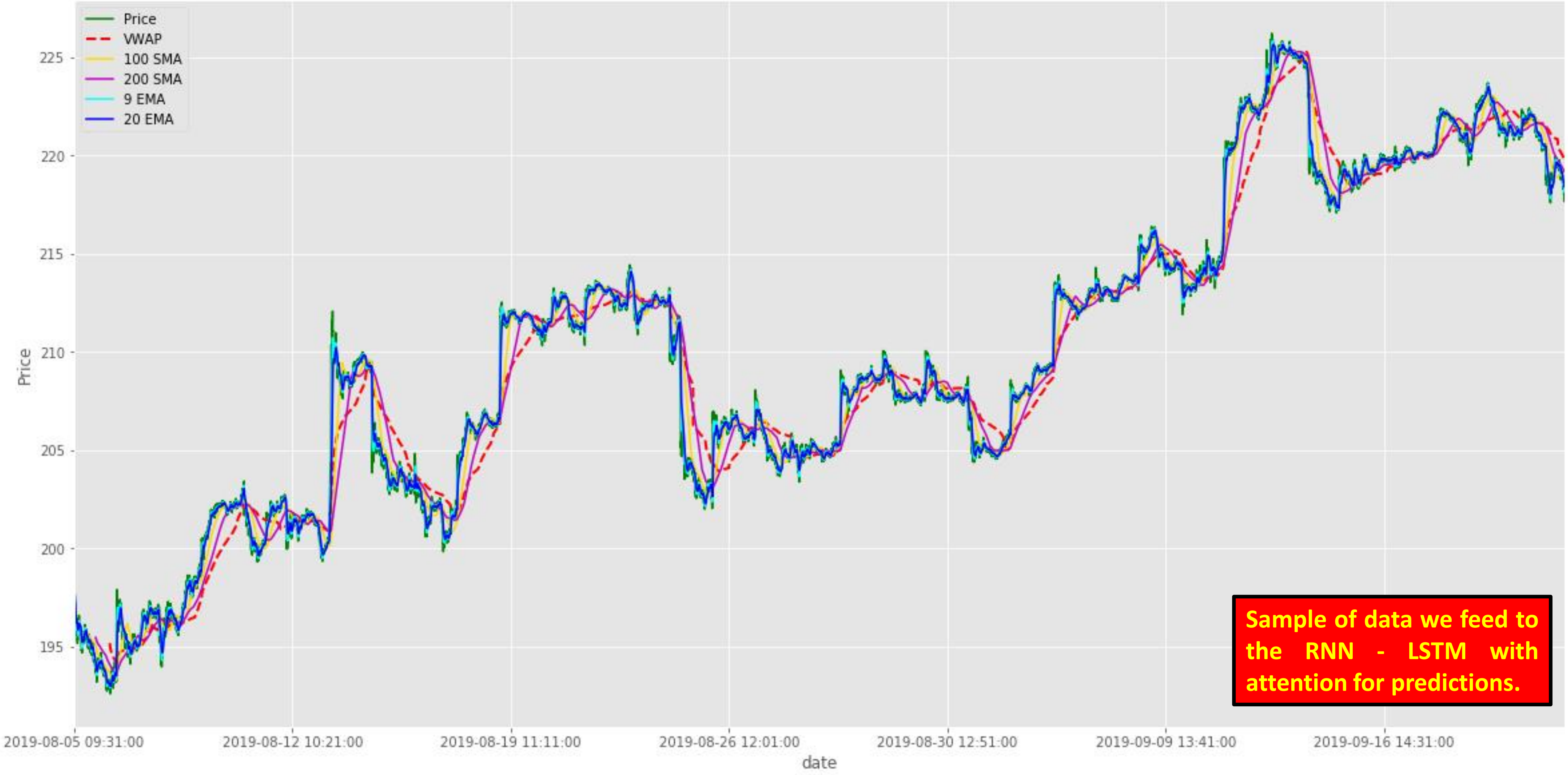
Stock Price - MSFT



Sample of data we feed to the RNN - LSTM with attention for predictions.

N.B: Time frame : 1 min

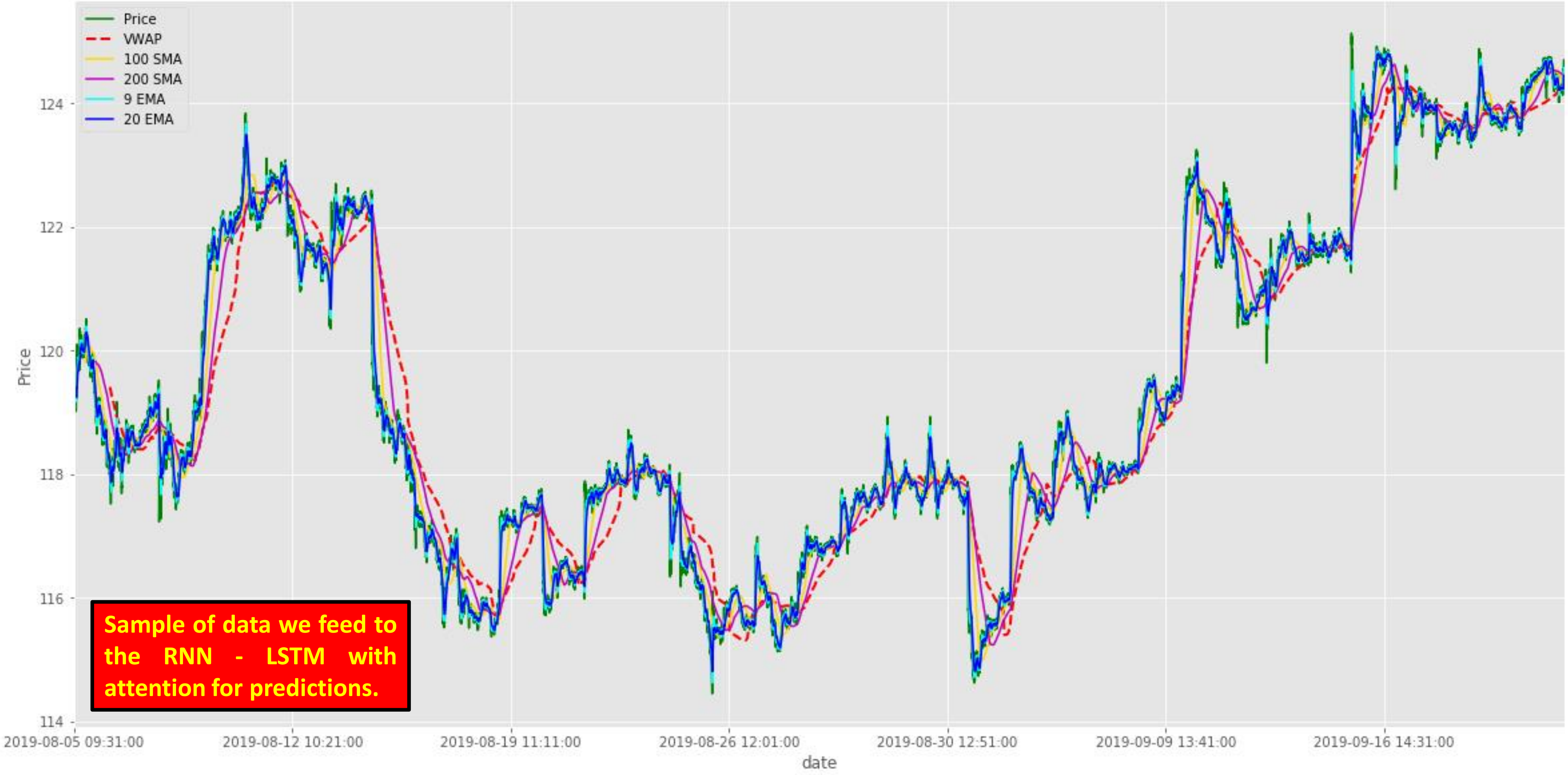
Stock Price - AAPL



Sample of data we feed to the RNN - LSTM with attention for predictions.

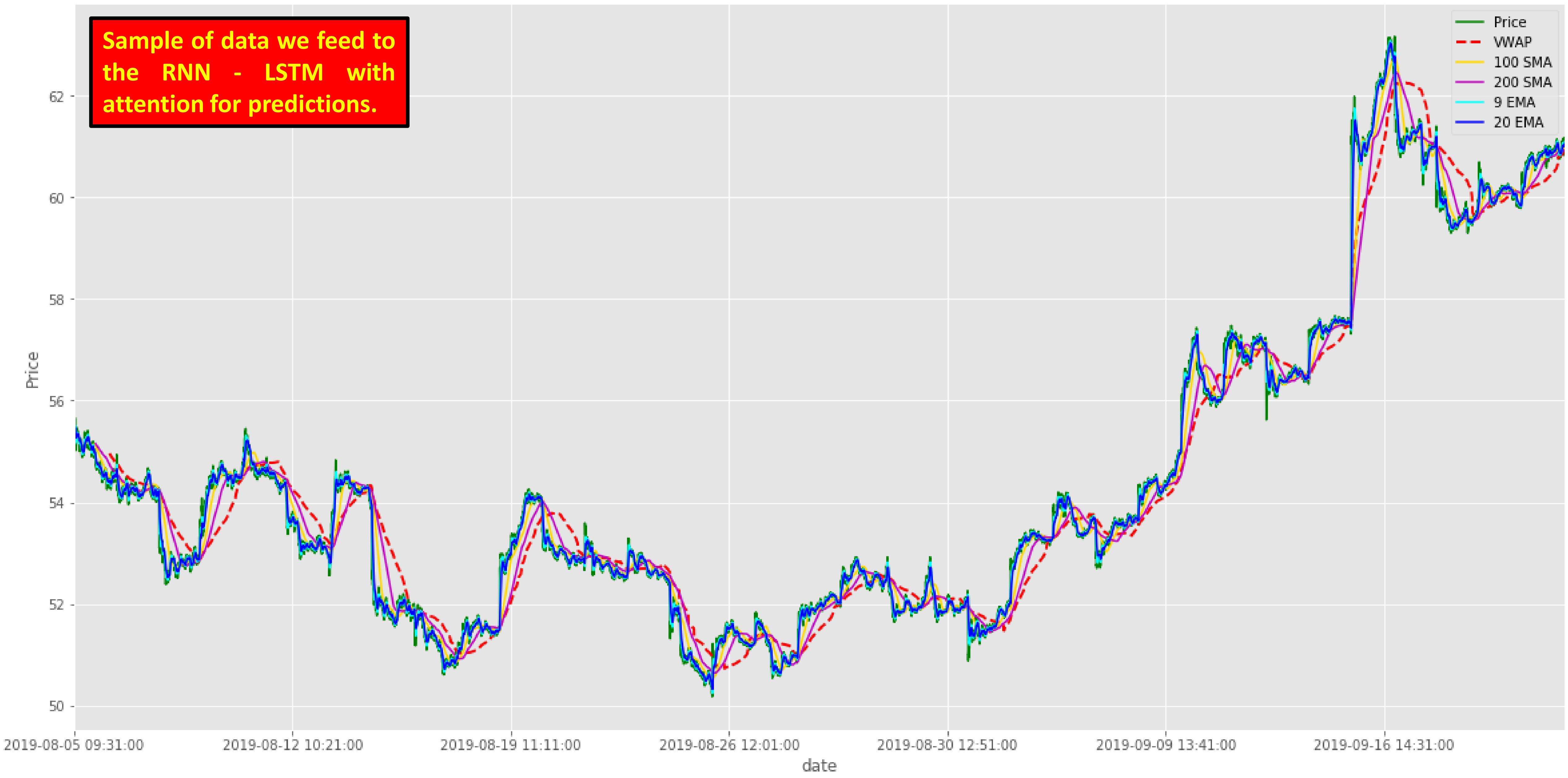
N.B: Time frame : 1 min

Stock Price - CVX



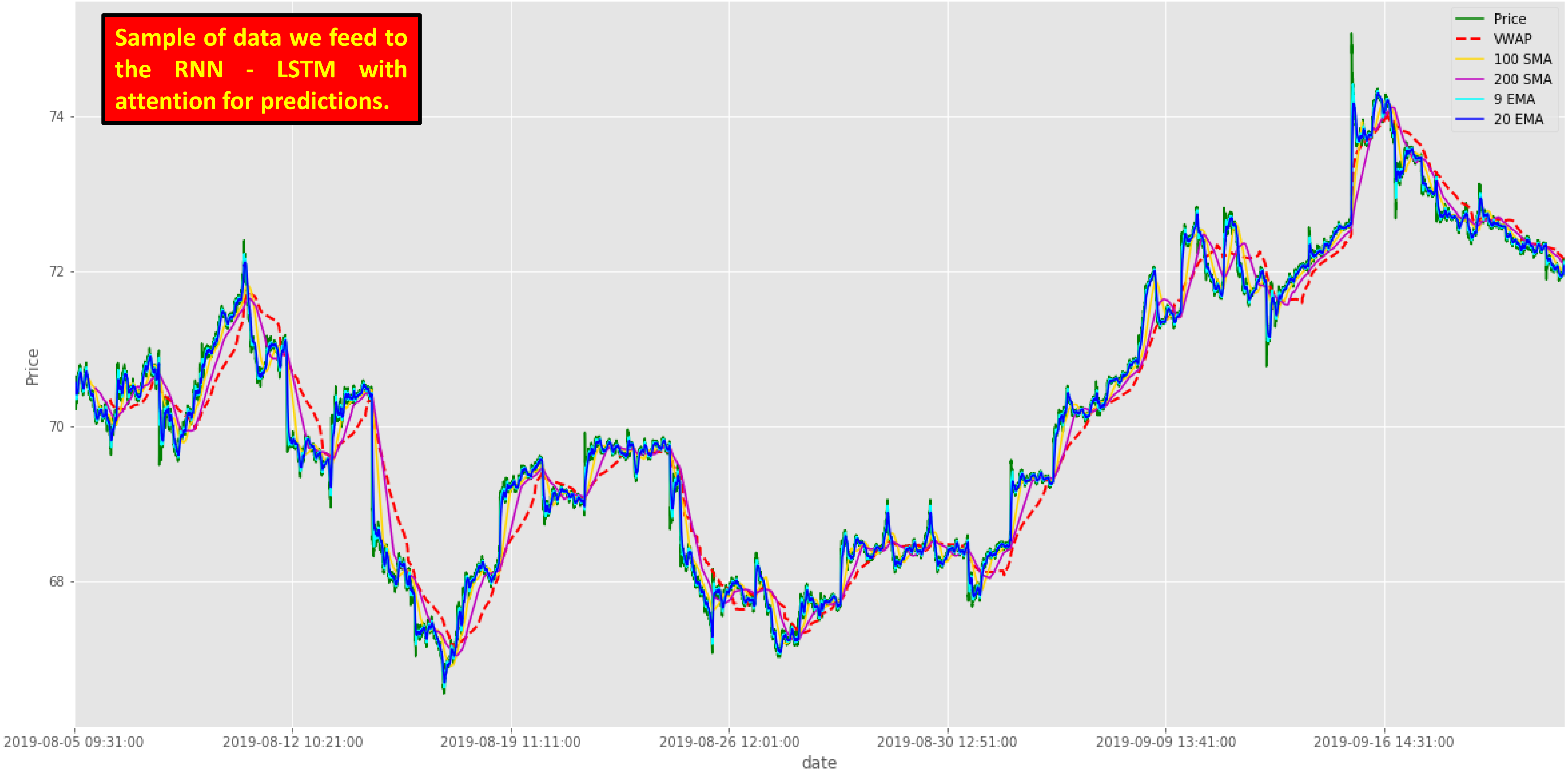
N.B: Time frame : 1 min

Stock Price - COP



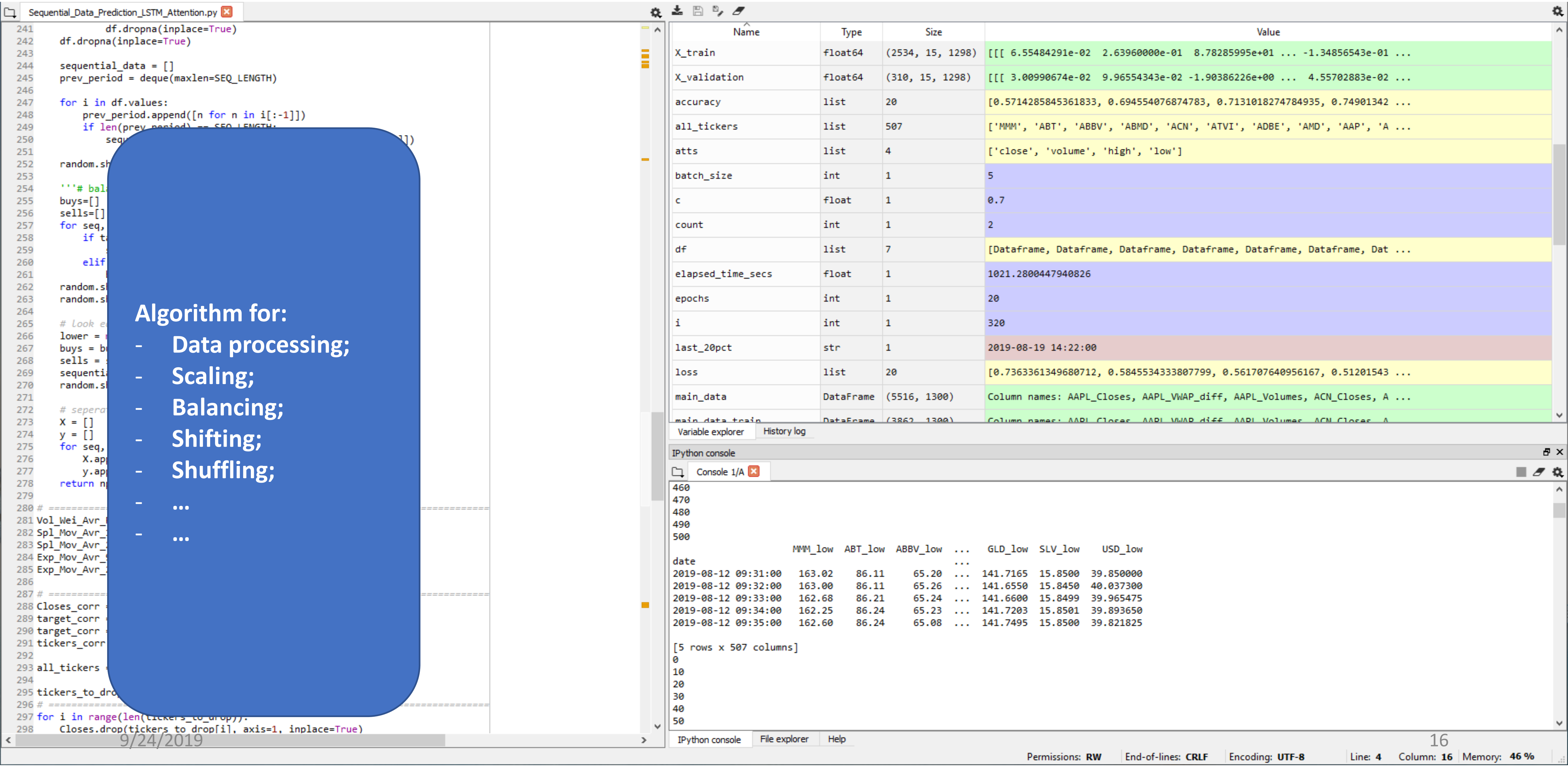
N.B: Time frame : 1 min

Stock Price - XOM

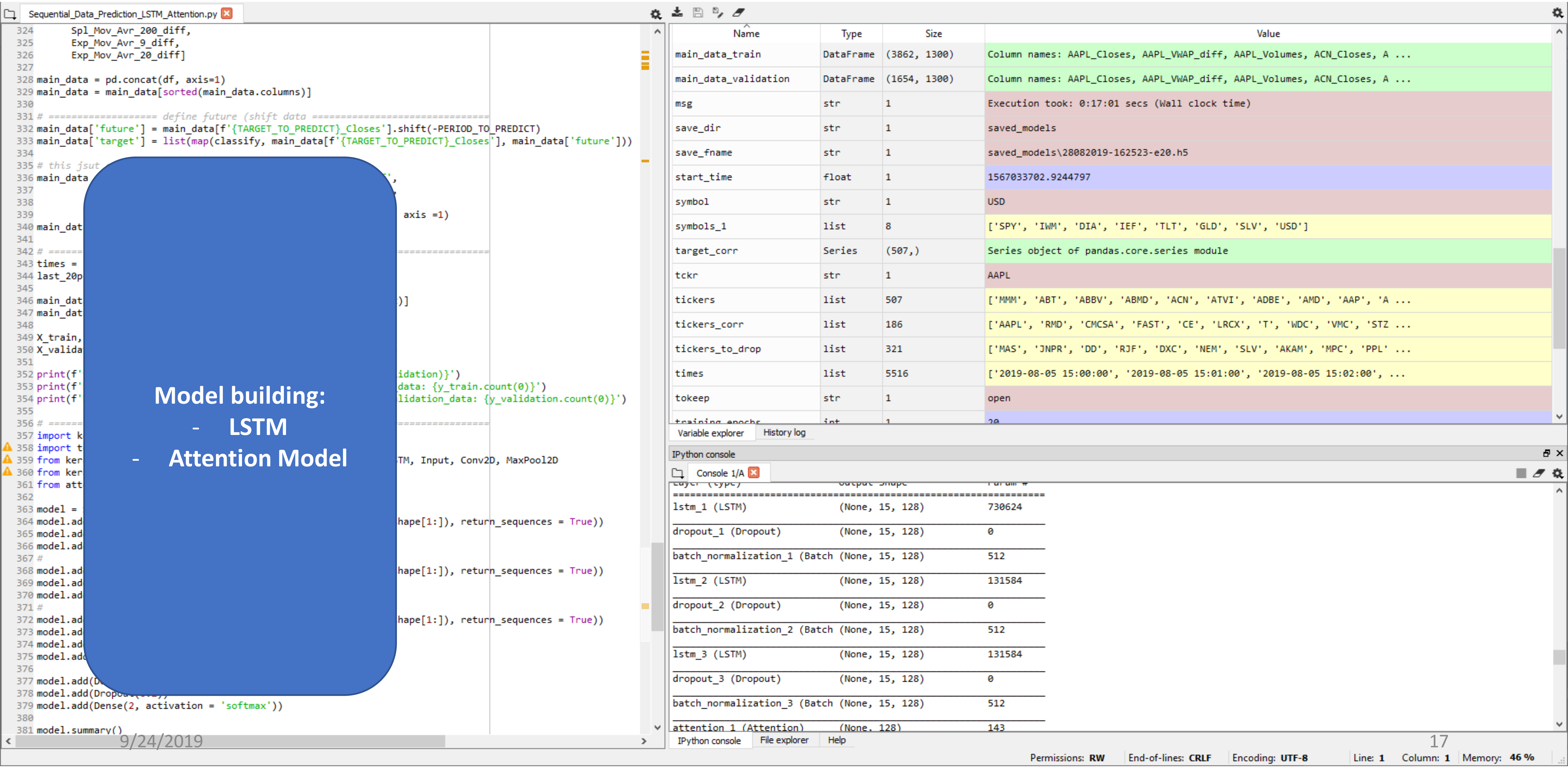


N.B: Time frame : 1 min

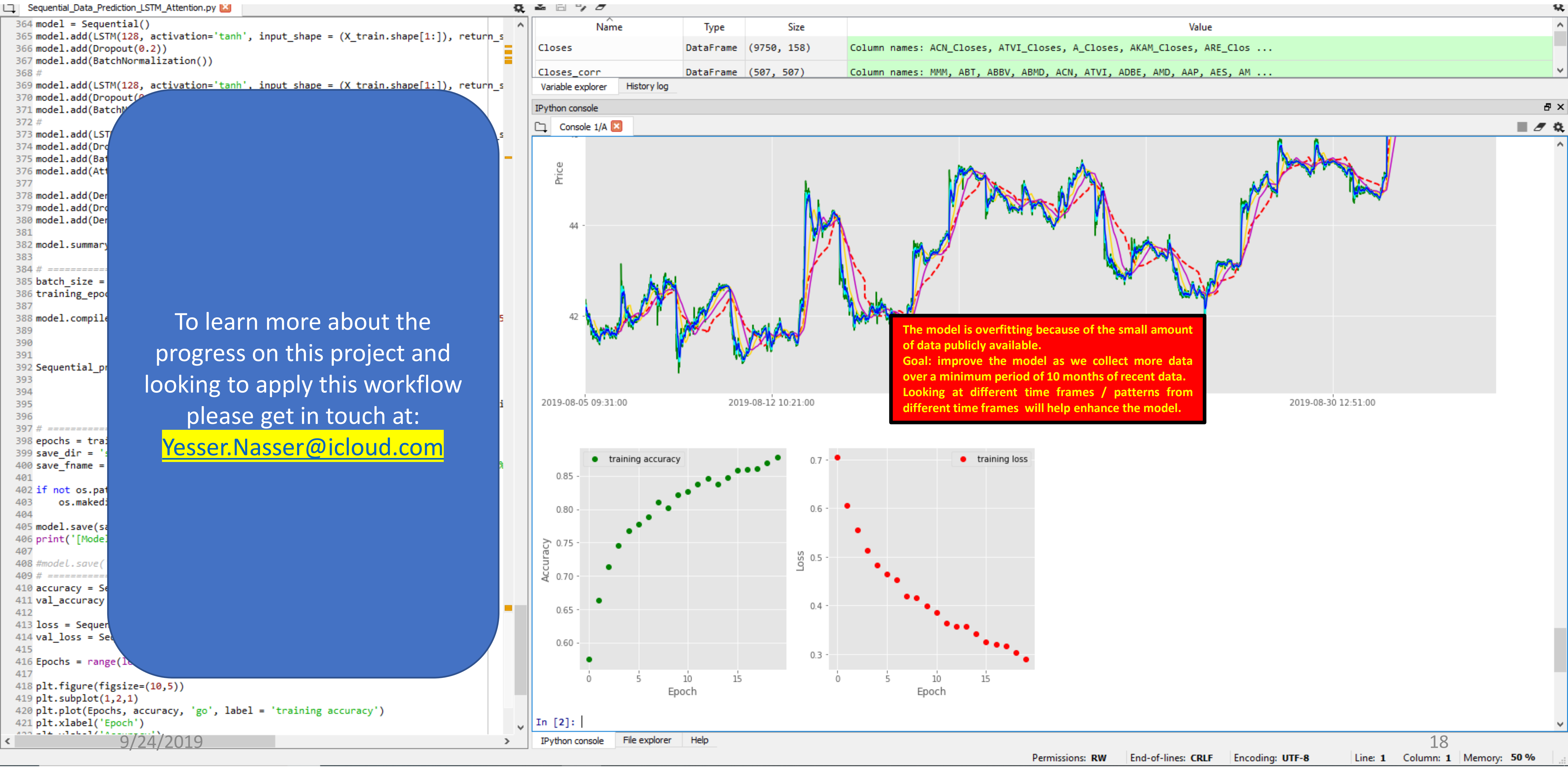
Algorithm Part 2: Data preprocessing



Algorithm Part 2: Model building



Algorithm Part 3: Running Model – preliminary results



To learn more about the progress on this project and looking to apply this workflow please get in touch at: Yesser.Nasser@icloud.com

The model is overfitting because of the small amount of data publicly available.
Goal: improve the model as we collect more data over a minimum period of 10 months of recent data.
Looking at different time frames / patterns from different time frames will help enhance the model.