1.

Window Size	Step	MSE
5	5	14.860
5	10	150.154
10	5	23.617

We can see that the when step size is smaller the window size, that is to said there are certain overlapped between sequent train data, model performs usually well. If the step size is over the window size meaning that some data are ignored, the model perform terrible due to the inconsecutive data. Also, with the window size = 5, step = 5 which is relatively small value performs the best in this task, addressing that the small range of fluctuation is more important than large trend of the value in this data.

2.

(1)

Window Size	Step	MSE
5	5	1094.636
5	10	1357.874
10	5	1095.619

By directly adding feature “Volumes” the MSE increasing, which might address that this information provide large noise to model for learning the trend.

(2)

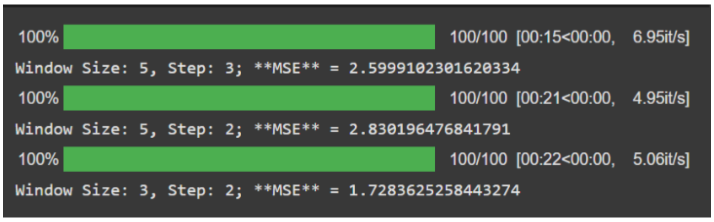
Window Size	Step	MSE
5	5	12.346
5	3	6.612
5	2	2.504
3	5	22.259
3	3	8.550
3	2	1.893

According to the results from question 1, smaller window size with step size being smaller than the window size help as to get better performance with all features excluding “Volume”. Thus, we can get when the window size is 3 and the step size is 2, its MSE is the smallest. The reason why this combination does well is because the way we create our data. Since there is only on rapidly increasing trend in last of this sequence data, smaller window size and step size provide more data which after we product the mean operation, the big error would be diluted by the large number of test data.

3.

Window Size	Step Size	MSE (before)	MSE (after)
5	3	6.612	2.537
5	2	2.504	2.723
3	2	1.893	1.338

According to the table, normalized the value by columns help decreasing the MSE to make model perform even better. In turn, we can conclude with the normalization, model is more stable. Noted that with the normalization, we can consider the feature “Volume” during training with less harmfulness of the model performance. However, the result still state that the consideration of “Volume” would disrupt the model to capture features.



4.

I don’t think the statement is correct. According to the experimental results above, we can clearly discern that when step size is less than window size, which there is overlap between windows, the model performs better. To the best of my knowledge, there are few articles discussing about the size of step. Since in real time series forecasting problem, more data provide usually more information. They usually set step size as 1, even in review article they do so Lim, Bryan, and Stefan Zohren. "Time-series forecasting with deep learning: a survey." *Philosophical Transactions of the Royal Society A* 379.2194 (2021): 20200209.

5.

Teng, Xiao, et al. "Enhancing Stock Price Trend Prediction via a Time-Sensitive Data Augmentation Method." *Complexity* 2020.1 (2020): 6737951.

In this paper, they decomposed time series data into components at different frequence level, so called discrete wavelet transform (DWT). They remained the low frequency part and adding noise with interpolation in high frequency one, consequently composed them together and gained the new augmented data. In summary, this method enhance the ability of LSTM on stock price prediction task.

6.

(i) Convolution-based models use fixed window size. Due to the sliding window mechanism which rely on the overlap segments for capturing features, the size of window has to be carefully chosen.

(ii) Fixed window size is often used to limit computational complexity and mitigate issues like vanishing gradients, but it could set longer window size compare to convolution-based model.

(iii) The optimal window size may vary depending on the specific characteristics of the time series data and the forecasting horizon. But the window size is fixed as well for considering how much historical data have to be taken into consideration. Due to the self-attention mechanism, we could select even larger window size and only have to worry about the computational limitation.