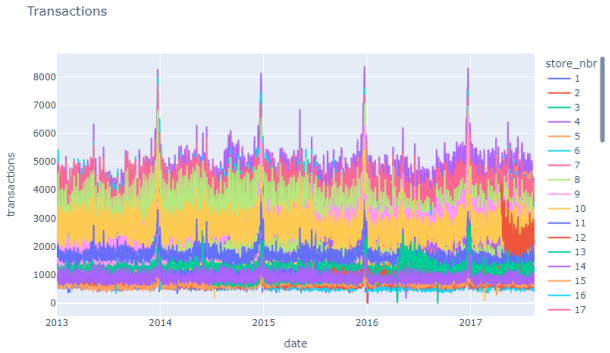
【Lab3】Store Sales - Time Series Forecasting

Deep 柴

1. 資料觀察
   A. Sales by store_nbr 畫出的圖，顯示store_nbr對於sales影響的影響很大



2. 資料前處裡
   A. 以(store_nbr,family) 作為Key，將sales跟onpromotion作為sequence
      - 會變成一個 1684 rows × 3564 columns的Matrix

	sales_1_BEAUTY	onpromotion_1_BEAUTY	sales_1_BABY CARE	onpromotion_1_BABY CARE	sales_1_LADIESWEAR	onpromotion_1_LADIESWEAR	sales_1_BEV
2013-01-01	0.0	0	0.0	0	0.0	0	
2013-01-02	2.0	0	0.0	0	0.0	0	
2013-01-03	0.0	0	0.0	0	0.0	0	

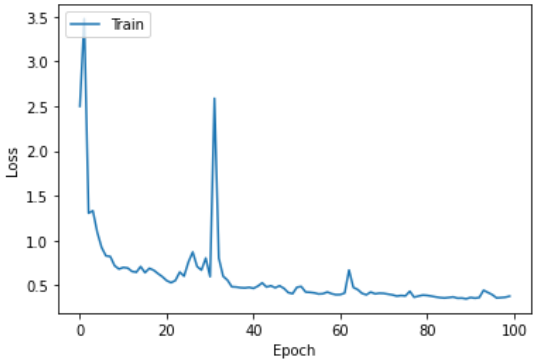
      - `generateInput(X_ori)` function處理這個部分
   B. 用60天去預測後16天，所以迴圈將Model Input處理成下圖的shape

```
trainX shape: (1609, 60, 3564)
trainY shape: (1609, 16, 1782)
```

      - `generater(df_X)`function處理這個部分
3. 模型訓練: 以下紀錄做過的嘗試
   A. 只取2015-01-01~2017-08-15的資料:
      - 原因
        - 看趨勢圖，2015-01-01~2017-08-15的資料較有規律
        - 第一次嘗試，以比較小的資料量減少訓練時間
          - 模型Summary如右圖
        - Loss: Train了100個epoch，約6s/epoch

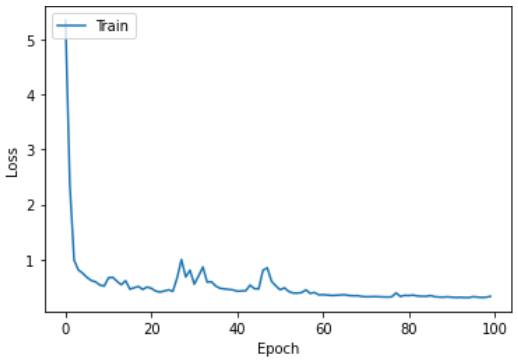
Layer (type)	Output Shape	Param #
conv1d_39 (Conv1D)	(None, 19, 100)	1782100
conv1d_40 (Conv1D)	(None, 17, 100)	30100
conv1d_41 (Conv1D)	(None, 16, 100)	20100
lstm_138 (LSTM)	(None, 16, 100)	80400
dropout_136 (Dropout)	(None, 16, 100)	0
lstm_139 (LSTM)	(None, 16, 100)	80400
dropout_137 (Dropout)	(None, 16, 100)	0
lstm_140 (LSTM)	(None, 16, 100)	80400
dropout_138 (Dropout)	(None, 16, 100)	0
lstm_141 (LSTM)	(None, 16, 100)	80400
dropout_139 (Dropout)	(None, 16, 100)	0
time_distributed_22 (TimeDistributed)	(None, 16, 1782)	179982

Total params: 2,333,882
Trainable params: 2,333,882
Non-trainable params: 0



B. 用全部的日期進行預測:

- 一樣使用A的模型結構
- Loss: Train了100個epoch，約12s/epoch



- 結果: 最好
  - 判斷是因為有學習到比較多的趨勢

4. 結果