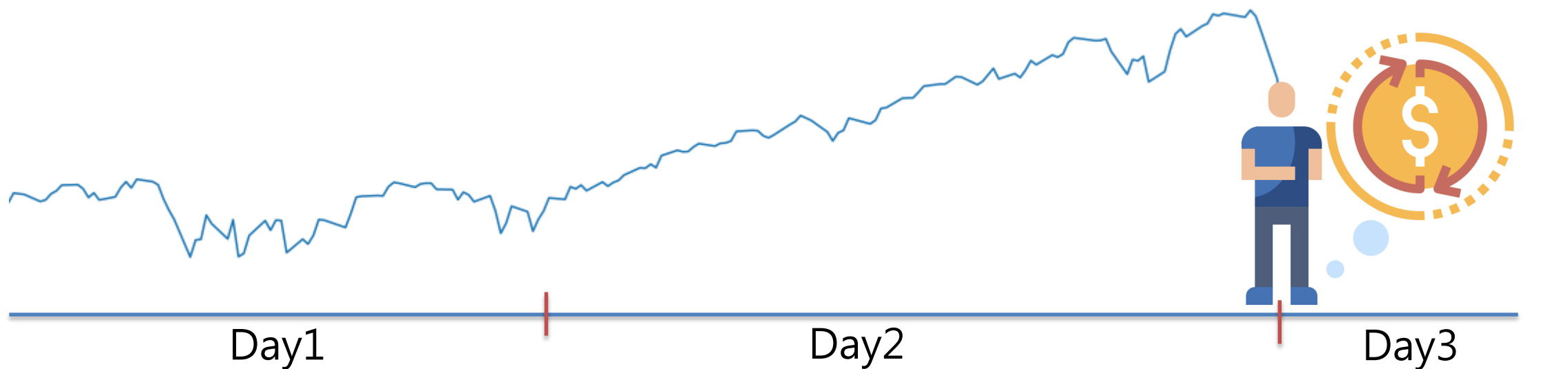


# 行為財務人工智能神經網路庶民理財模型

Ethan Wu

Duration:

2010/1/1~2020/10/15



	TxDate	StockID	Open	High	Low	Close	Adj Close	Volume	type	有價證券名稱	市場別	產業別	公開發行/上市(櫃)/發行日
0	2010-01-04	1101	27.1635	27.483101	27.083599	27.4032	13.858723	10321349.0	stock	台泥	上市	水泥工業	1962-02-09
1	2010-01-05	1101	27.6429	28.601601	27.563000	28.3619	14.343570	60016780.0	stock	台泥	上市	水泥工業	1962-02-09
2	2010-01-06	1101	28.3619	29.080900	28.282000	28.9611	14.646606	44831404.0	stock	台泥	上市	水泥工業	1962-02-09
3	2010-01-07	1101	29.0410	29.080900	28.361900	28.4018	14.363750	18095530.0	stock	台泥	上市	水泥工業	1962-02-09
4	2010-01-08	1101	28.4018	28.601601	28.082300	28.2820	14.303162	13307856.0	stock	台泥	上市	水泥工業	1962-02-09

# Split Train/Test data

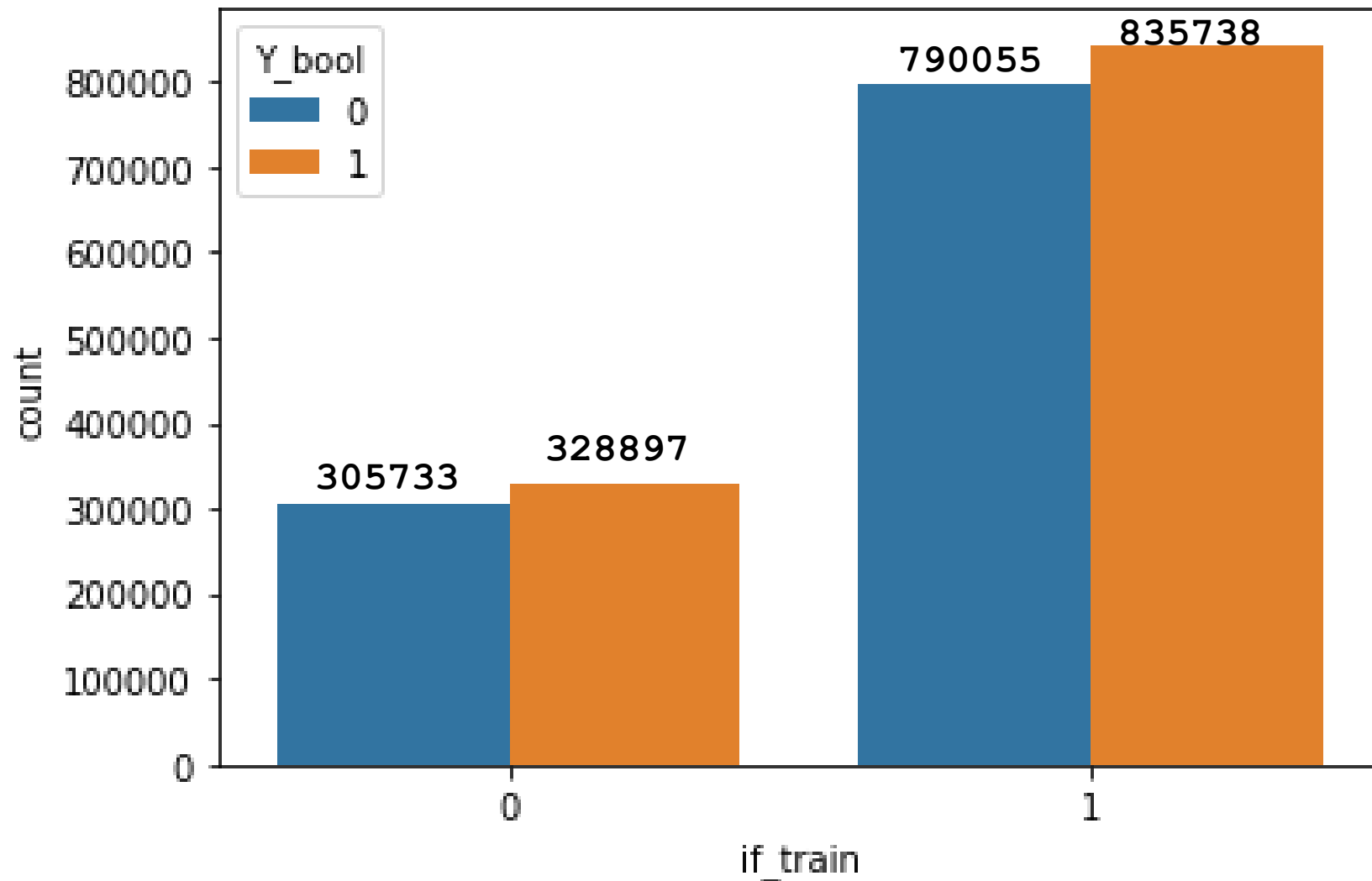
Test Duration:

2018/1/2~2020/10/15

Train Duration:

2010/1/1~2017/12/29

Trend跟我一樣



Dataset Shape: (2288783, 23)

	Name	dtypes	Missing	Uniques	First Value	Second Value	Third Value	Entropy
0	TxDate	datetime64[ns]	0	2650	2010-01-04 00:00:00	2010-01-05 00:00:00	2010-01-06 00:00:00	11.36644
1	StockID	int64	0	950	1101	1101	1101	9.8469
2	Open	float64	7368	254932	27.1635	27.6429	28.3619	14.27262
3	High	float64	7368	259958	27.4831	28.6016	29.0809	14.33147
4	Low	float64	7368	255687	27.0836	27.563	28.282	14.31288
5	Close	float64	7368	258628	27.4032	28.3619	28.9611	14.33336
6	Adj Close	float64	7368	868154	13.8587	14.3436	14.6466	18.82828
7	Volume	float64	7368	959554	1.03213e+07	6.00168e+07	4.48314e+07	17.26461
8	type	object	0	2	stock	stock	stock	0.03970
9	有價證券名稱	object	9757	943	台泥	台泥	台泥	9.83935
10	市場別	object	9757	1	上市	上市	上市	0.00000
11	產業別	object	9757	28	水泥工業	水泥工業	水泥工業	4.48537
12	公開發行/上市(櫃)/發行日	datetime64[ns]	9757	708	1962-02-09 00:00:00	1962-02-09 00:00:00	1962-02-09 00:00:00	8.77633

Step1: Stock information Clustering

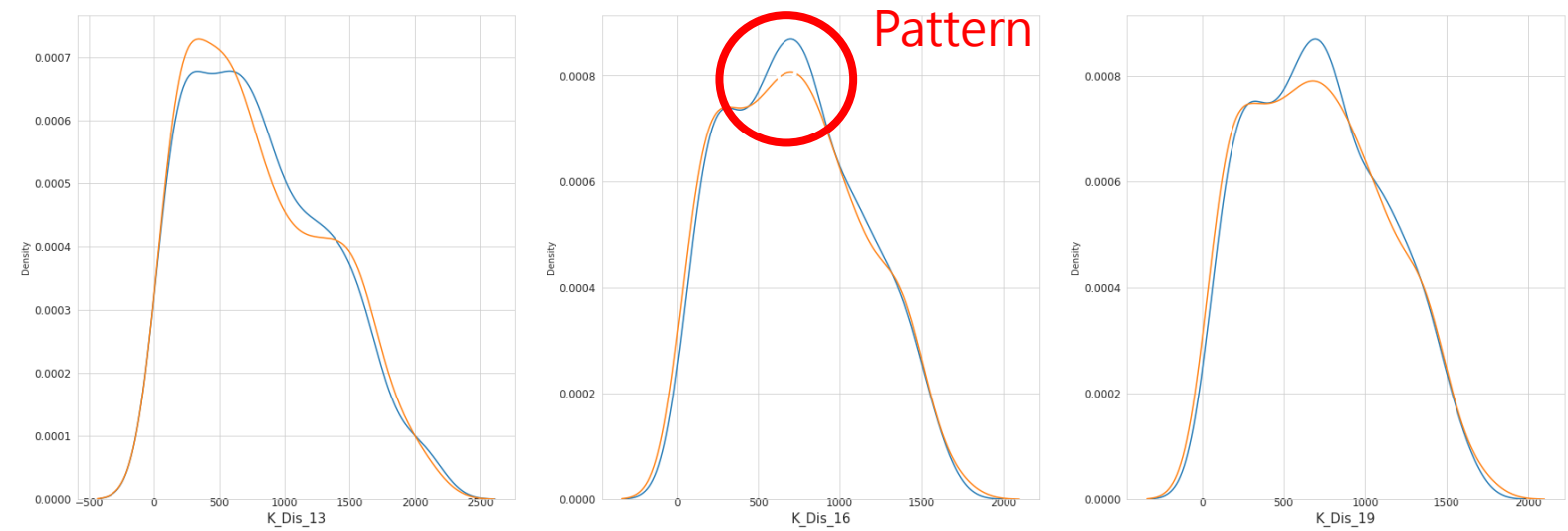
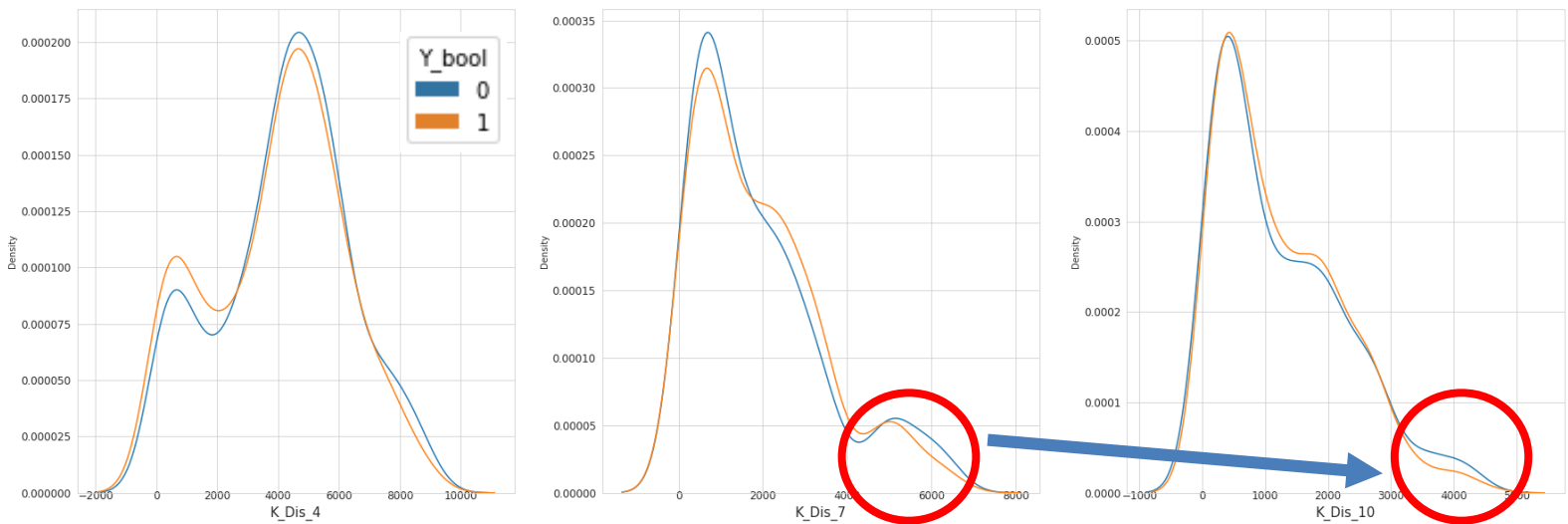
大市值 Group :



組數越多, Pattern越穩定

	StockID	Open	High	Low	Close	Adj Close
count	453.0	453.000000	453.000000	453.000000	453.000000	453.000000
mean	3008.0	3482.229581	3504.216336	3407.737307	3425.584989	3174.811207
std	0.0	1151.787326	1155.651750	1140.939156	1143.179969	1109.788865
min	3008.0	2080.000000	2080.000000	2005.000000	2010.000000	1815.583370
25%	3008.0	2525.000000	2525.000000	2455.000000	2455.000000	2246.780030
50%	3008.0	3075.000000	3090.000000	2980.000000	3000.000000	2748.621830
75%	3008.0	4445.000000	4495.000000	4385.000000	4430.000000	4163.034180
max	3008.0	5995.000000	6015.000000	5955.000000	5970.000000	5655.832520

f(不同Group|K Dis 4)



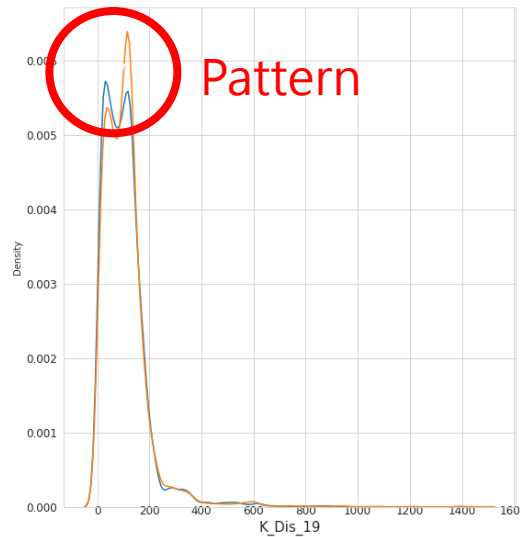
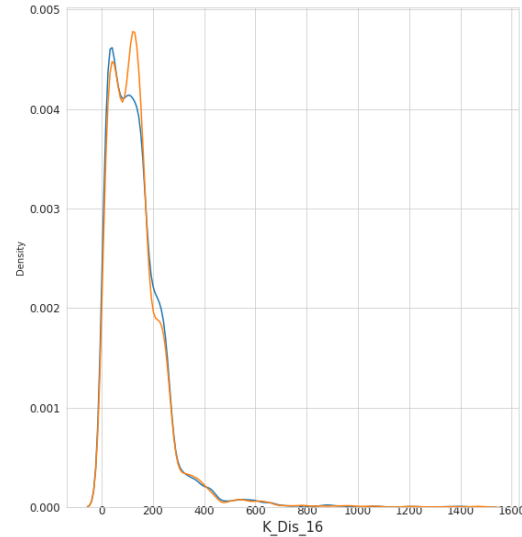
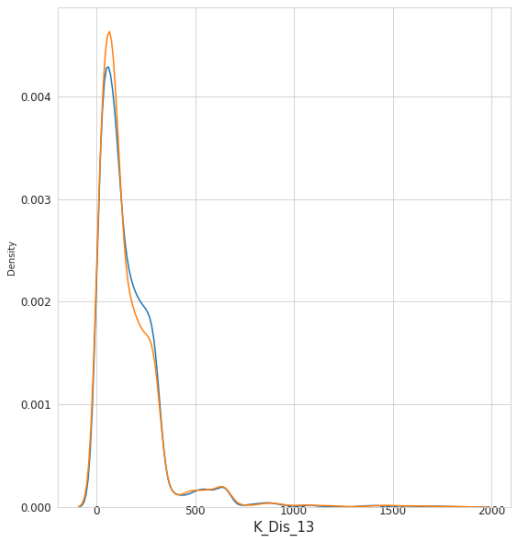
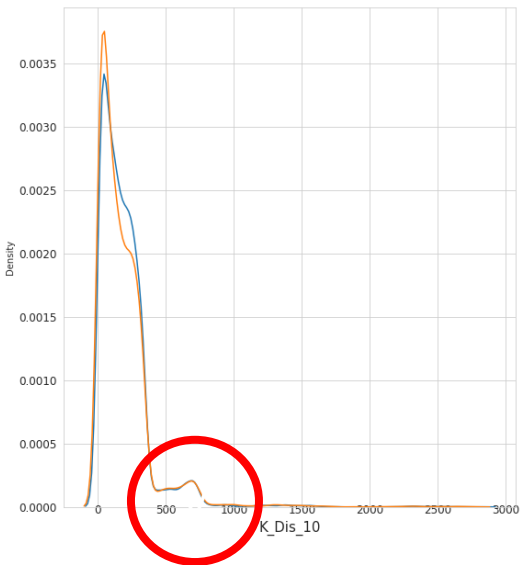
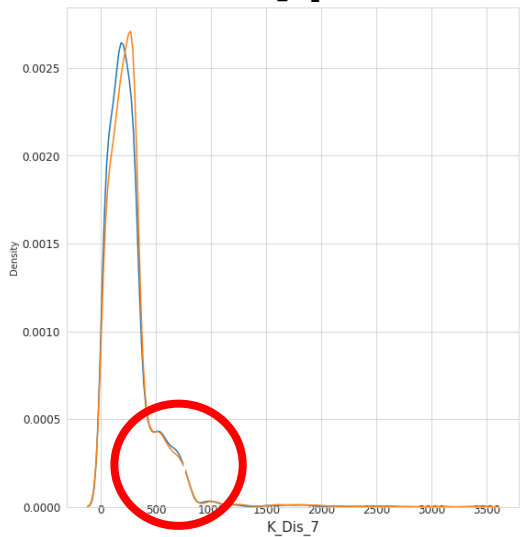
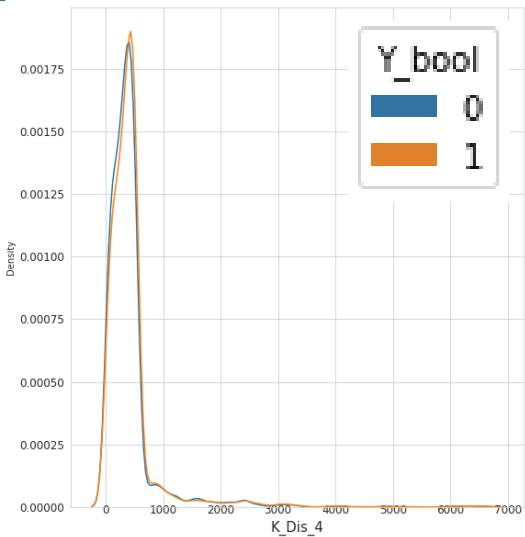
# 中型市值 Group :



組數越多, Pattern越穩定

	StockID	Open	High	Low	Close
count	10709.000000	10709.000000	10709.000000	10709.000000	10709.000000
mean	3371.935662	377.256631	379.568336	368.963918	371.172136
std	1836.934841	166.225351	167.443947	162.188861	163.052005
min	1476.000000	247.000000	248.455000	234.000000	235.000000
25%	2227.000000	284.914000	286.500000	278.682010	280.500000
50%	2498.000000	327.500000	329.976010	320.086000	322.000000
75%	3673.000000	406.884000	409.000000	398.500000	400.952000
max	9921.000000	2065.000000	2070.000000	2035.000000	2035.000000

**f**(不同Group|K Dis 4)

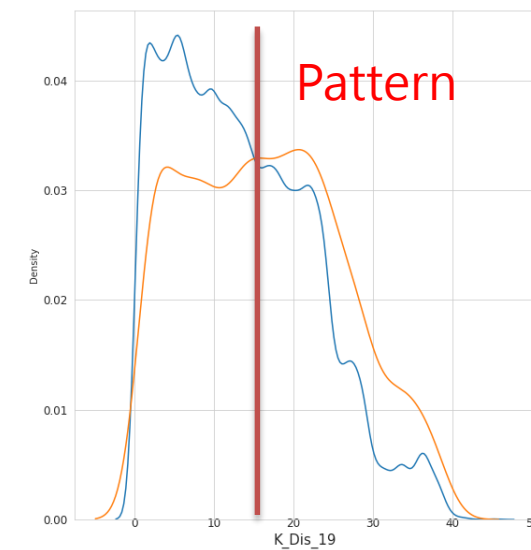
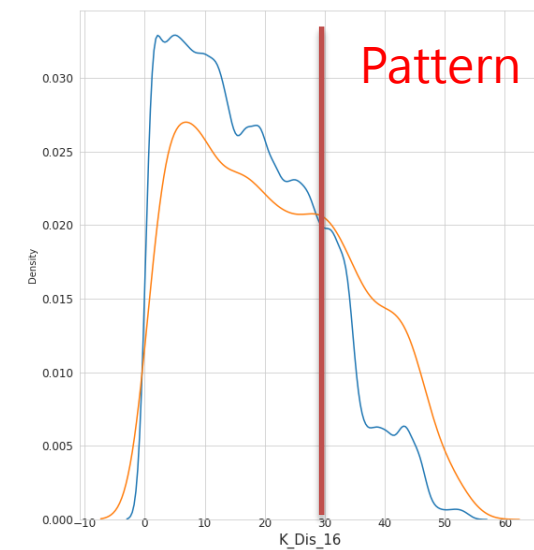
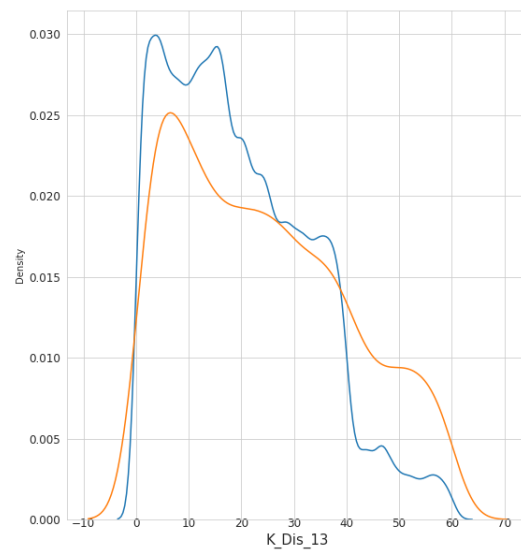
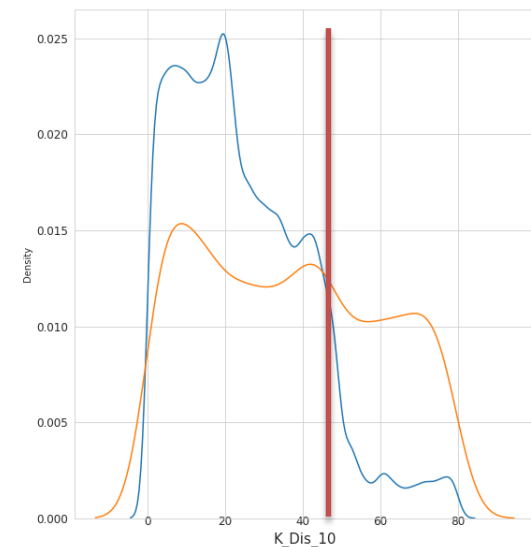
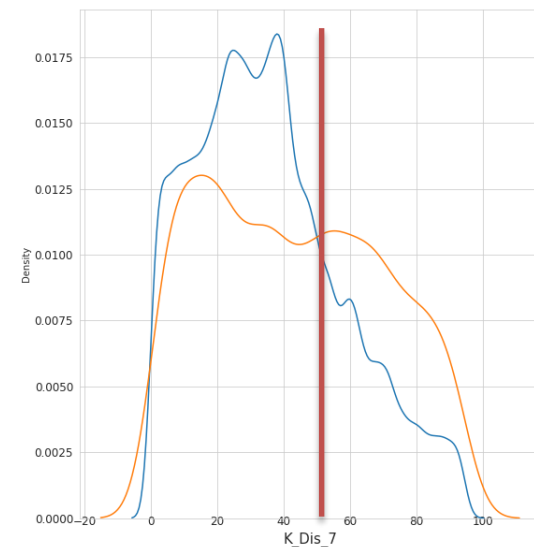
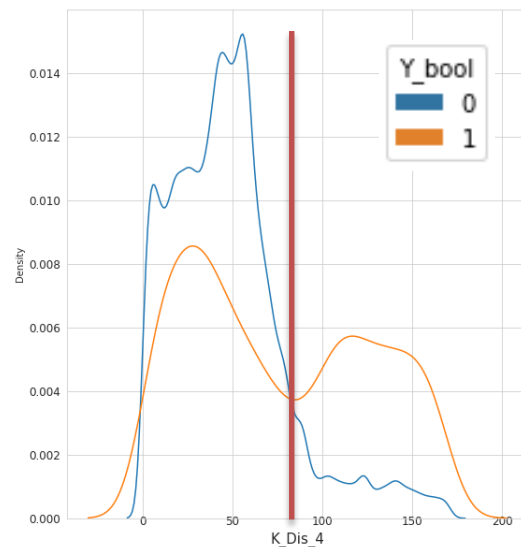


# 小型市值 Group :



Pattern一直都在,  
但應可Ensemble抓去更多  
細微特徵

**$f$ (不同Group|K Dis 4)**

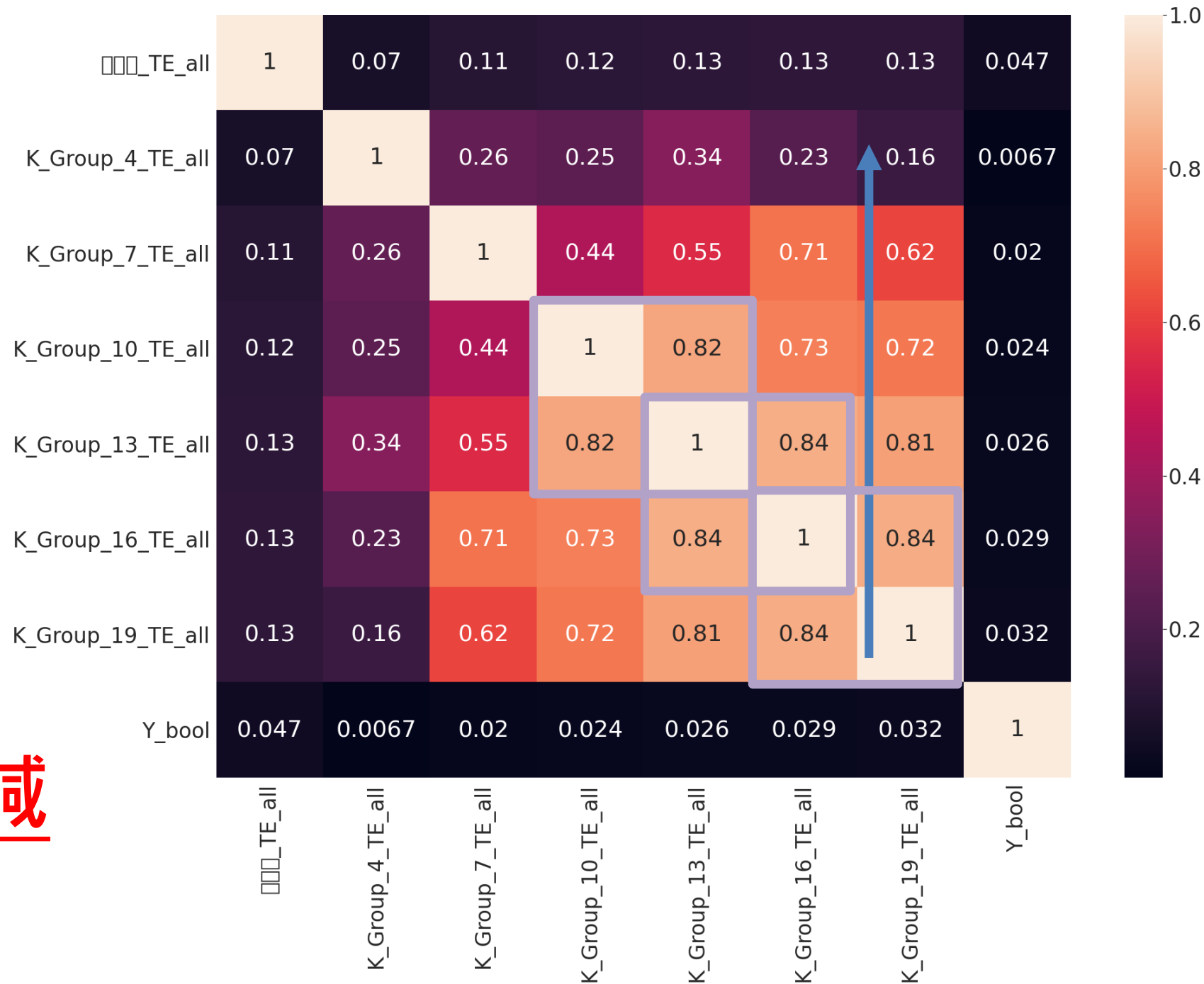


	StockID	Open	High	Low	Close
count	687240.000000	687240.000000	687240.000000	687240.000000	687240.000000
mean	3440.647349	25.151667	25.275494	24.682782	24.811403
std	2258.812614	15.277821	15.356455	14.994905	15.072366
min	50.000000	0.067640	0.067640	0.050730	0.050730
25%	2009.000000	13.068725	13.137800	12.820600	12.900000
50%	2537.000000	21.049800	21.156450	20.647100	20.750000
75%	3711.000000	34.300000	34.482800	33.669300	33.839100
max	9958.000000	71.000000	71.600000	67.115400	67.840300

# Correlation Grouping

10,13,16,19 Encode Value  
線性強度高(基本上是遞減)

做 **維度縮減**





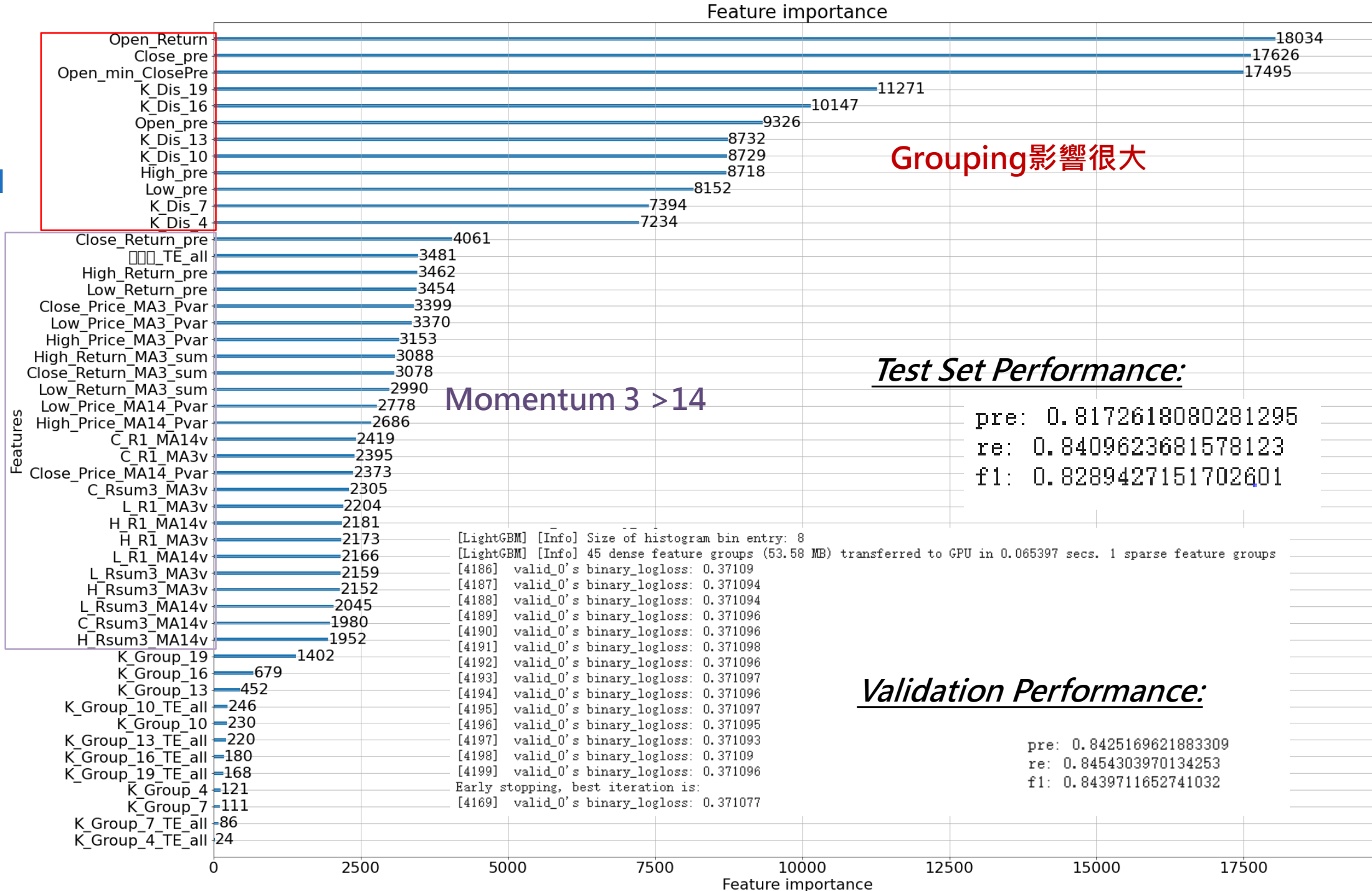
## Step2: Turn K-bar information into Momentum Feature.

依照Step1. Grouping製造依組別不同的Momentum強度


建立不同Scale的Feature Map來卷積重要的訊號圖示



# 庶民 Model



# 庶民model小節

- 透過庶民ML找Pattern可以讓  的隨機賭局勝率有效提升至80%
- Momentum指標:日線較週線有參考意義
- Cluster後距離越遠的可以有效分辨出各Group之標得的momentum訊號
- 小型股應用Ensemble Model,不宜放在一起train
- CNN Feature Map對多頭趨勢較有效

研究限制: 有Server應該可以做得不庶民一點^^

- 尚未考慮除權息等季節性因素
- Model要上線需考慮重量問題
- 財報中業績表現在多久時間會影響下一季價格
- 可以加入Volume等代表成交量的變數