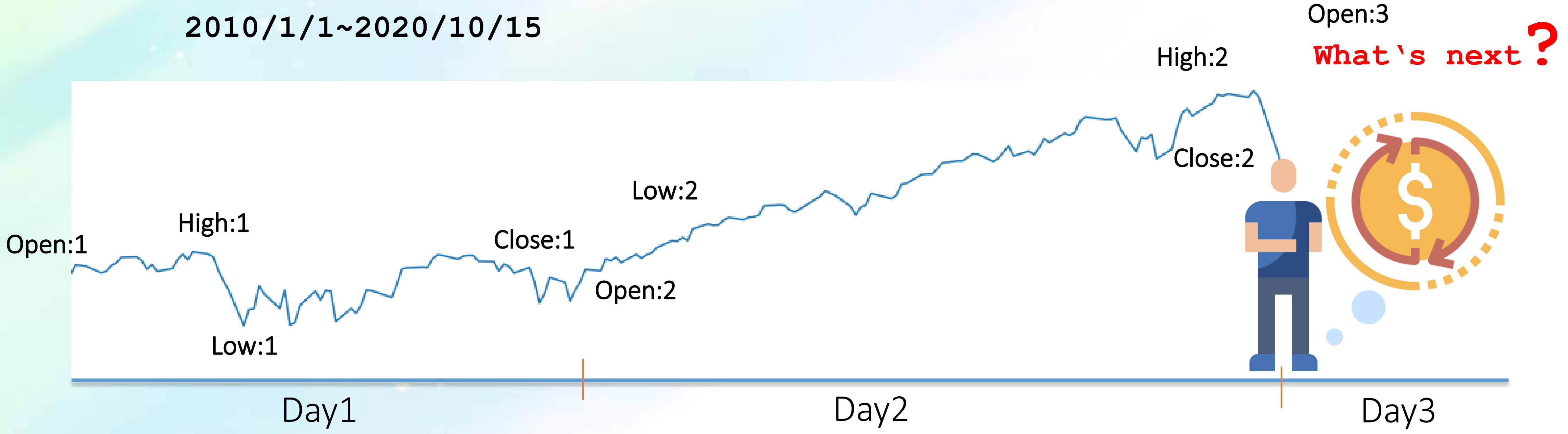
Transform Time-Series data to cross-section data

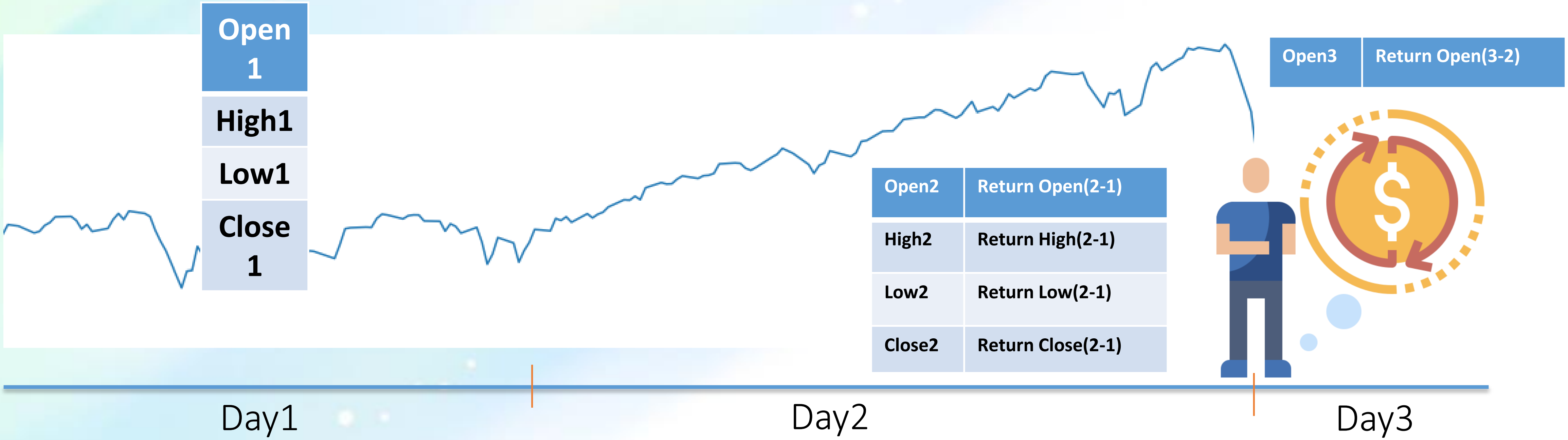
Duration:

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

What Information can I get while Day3 Open Price be known?



Day	Y	Open	Open_pre	Open_return	Close	Close_pre	Close_return
1	X	10	X	X	10.5	X	X
2	11.5-11	11	10	(11/10)-1	11.5	10.5	(11.5/10.5)-1
3	12.5-12	12	11	(12/11)-1	12.5	11.5	(12.5/11.5)-1

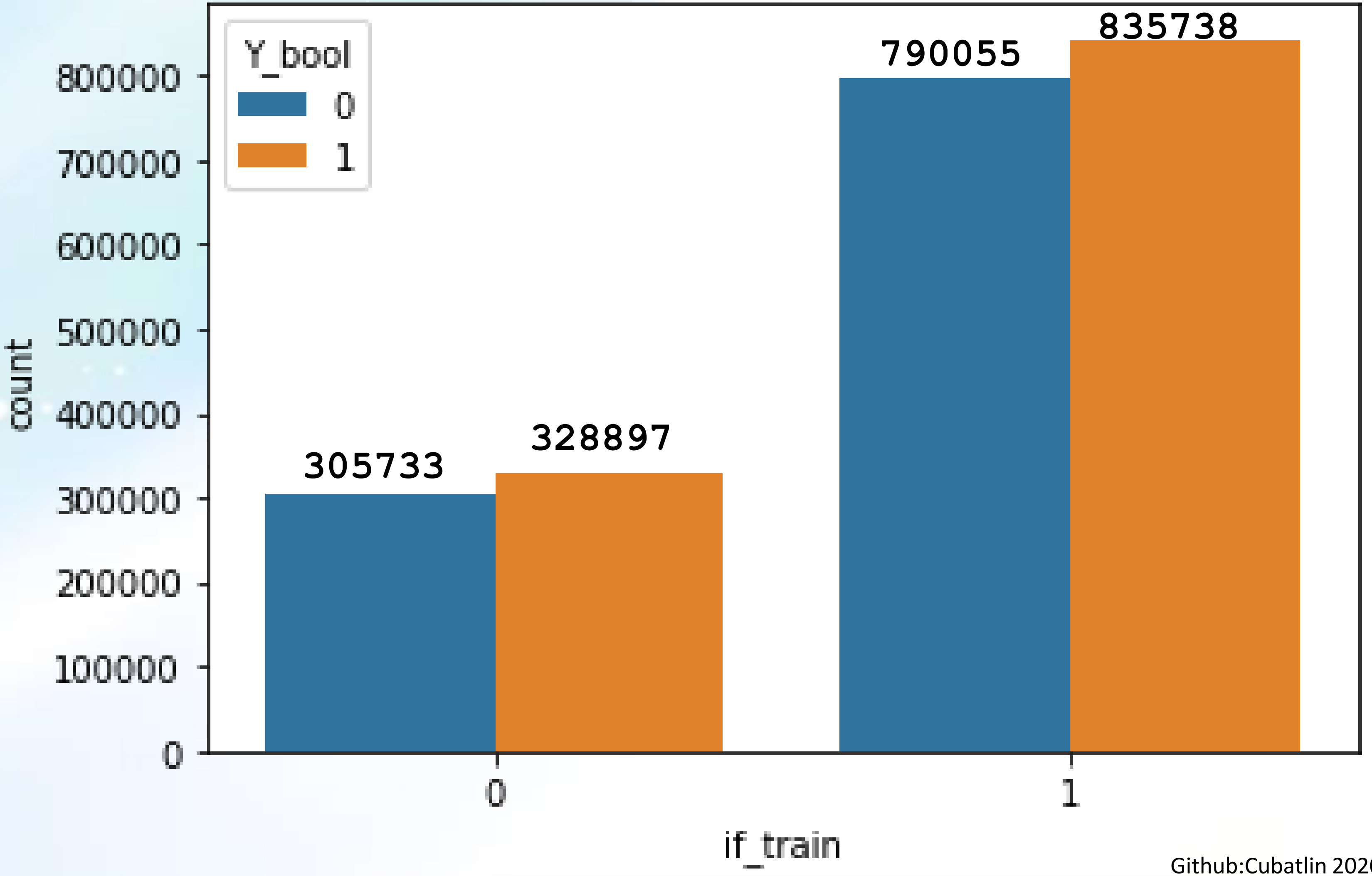
Split Train/Test data

Test Duration:

2018/1/2~2020/10/15

Train Duration:

2010/1/1~2017/12/29



Data Description

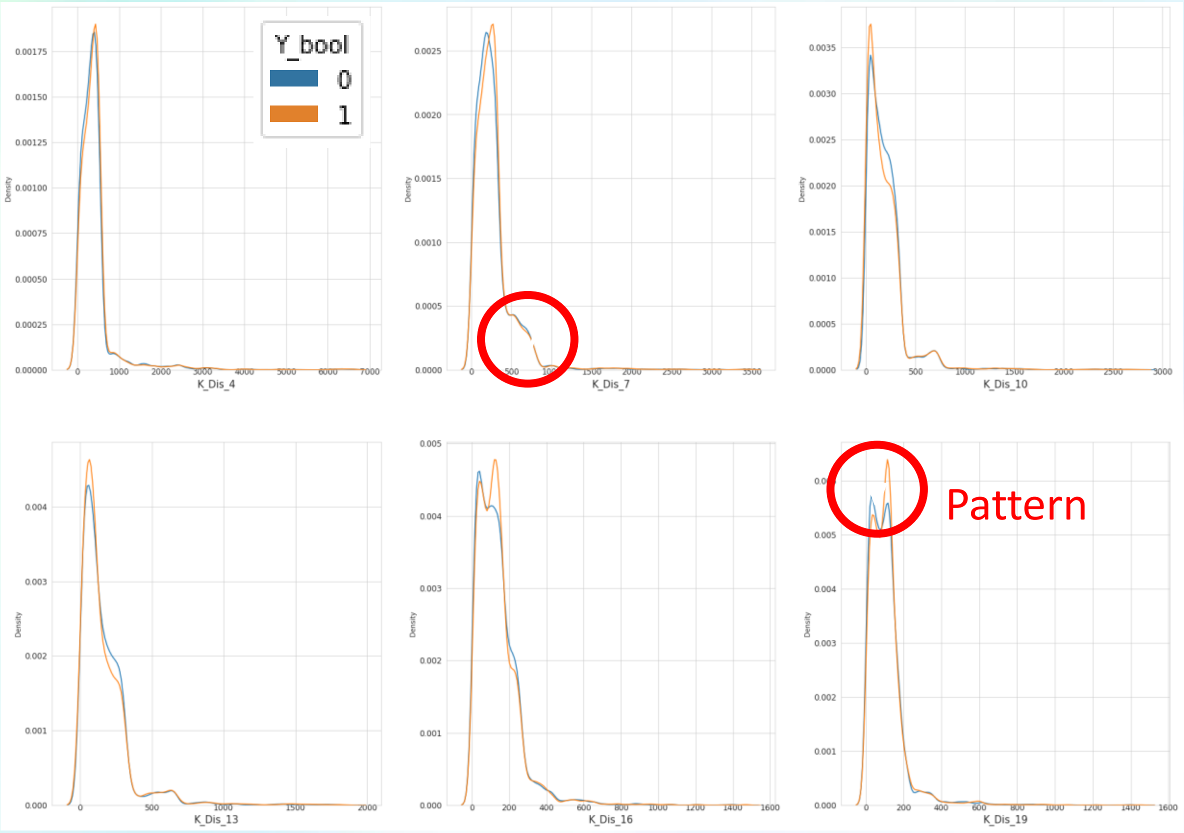
Dataset Shape: (2288783, 23)

	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

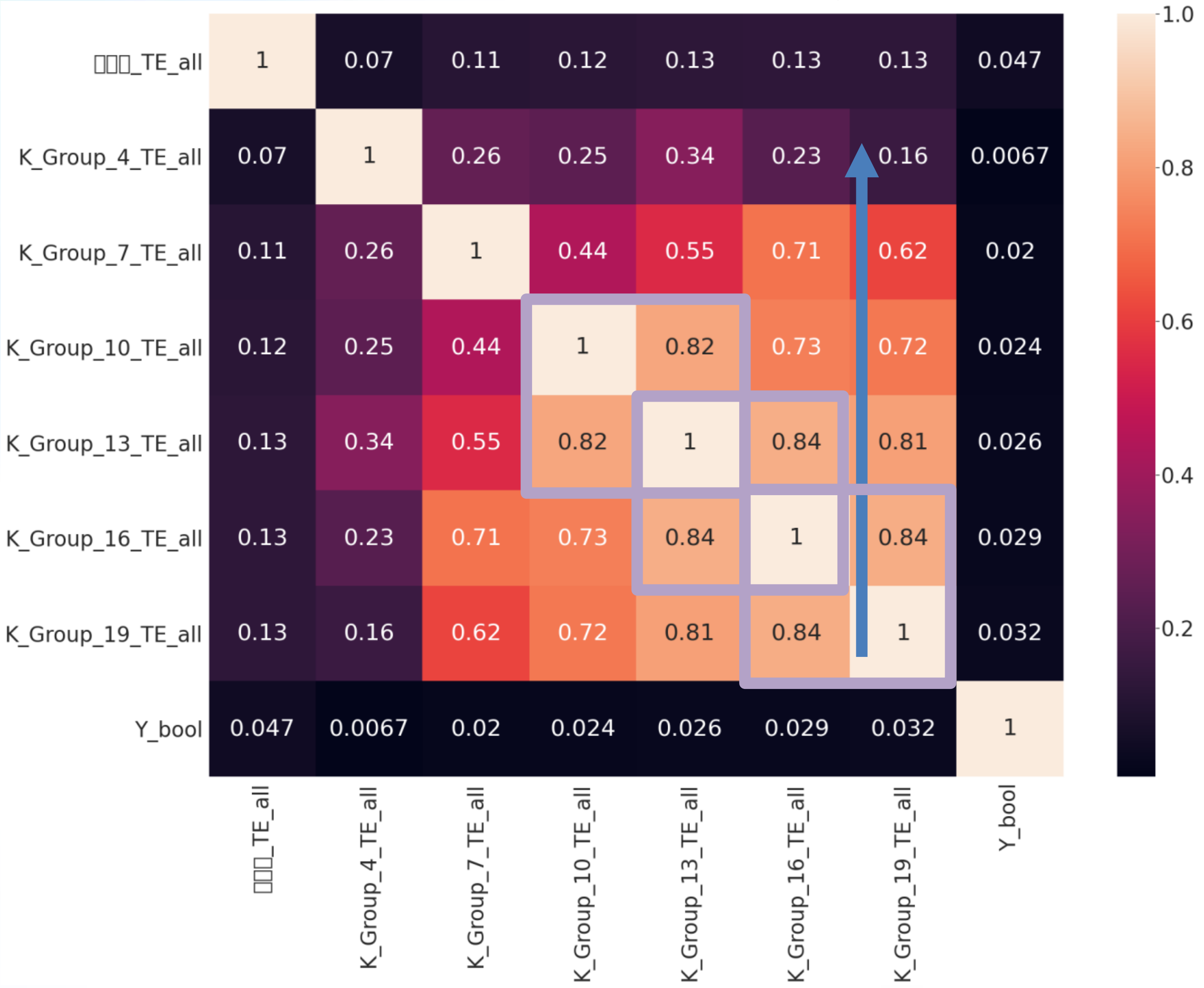
Stock Information Clustering

Grouping Correlation

f(Grouping|K_Dis_4)

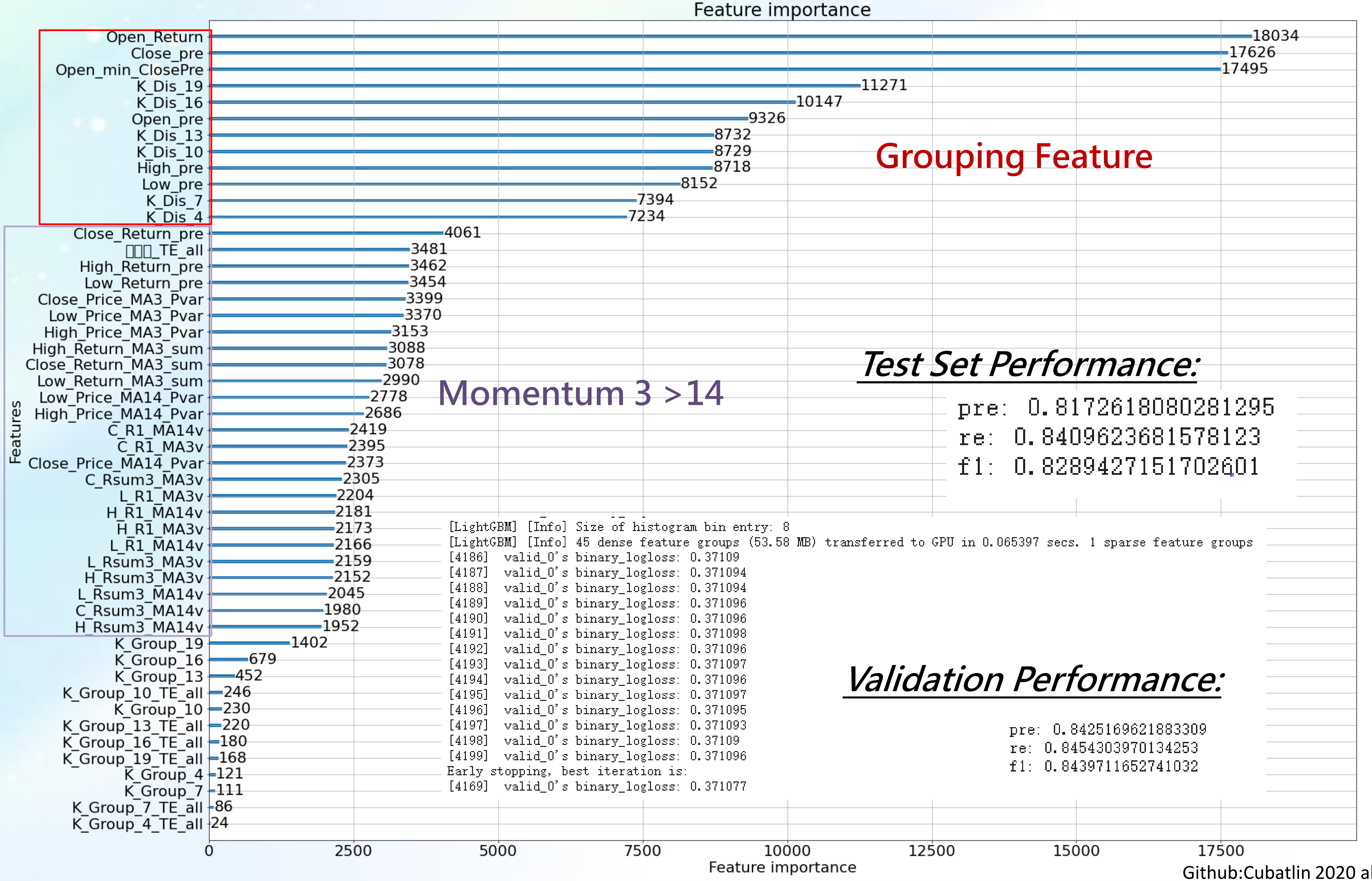


Diminishing from 19 to 4

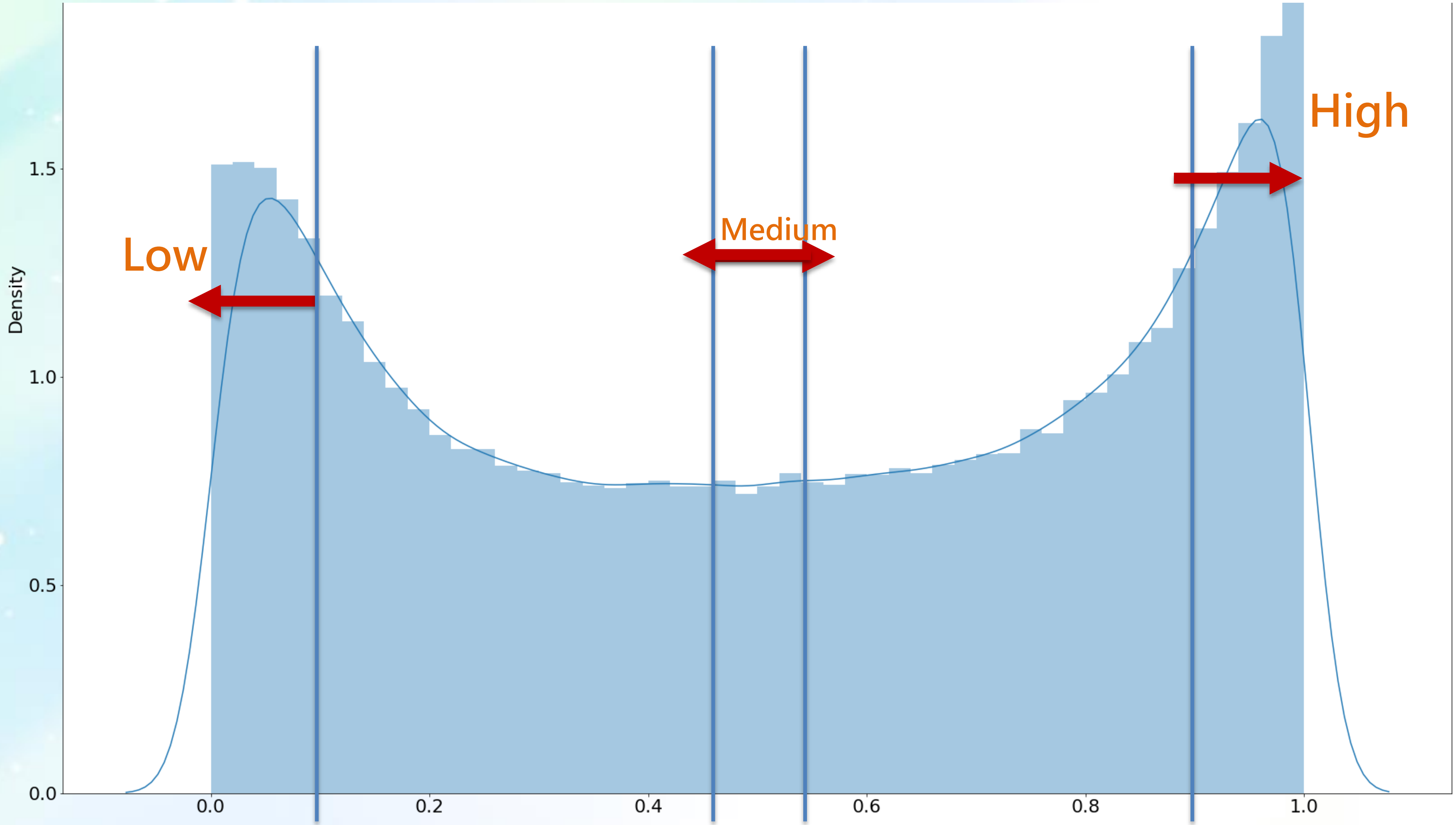


Turn K-bar information to Momentum features

Model Evaluation



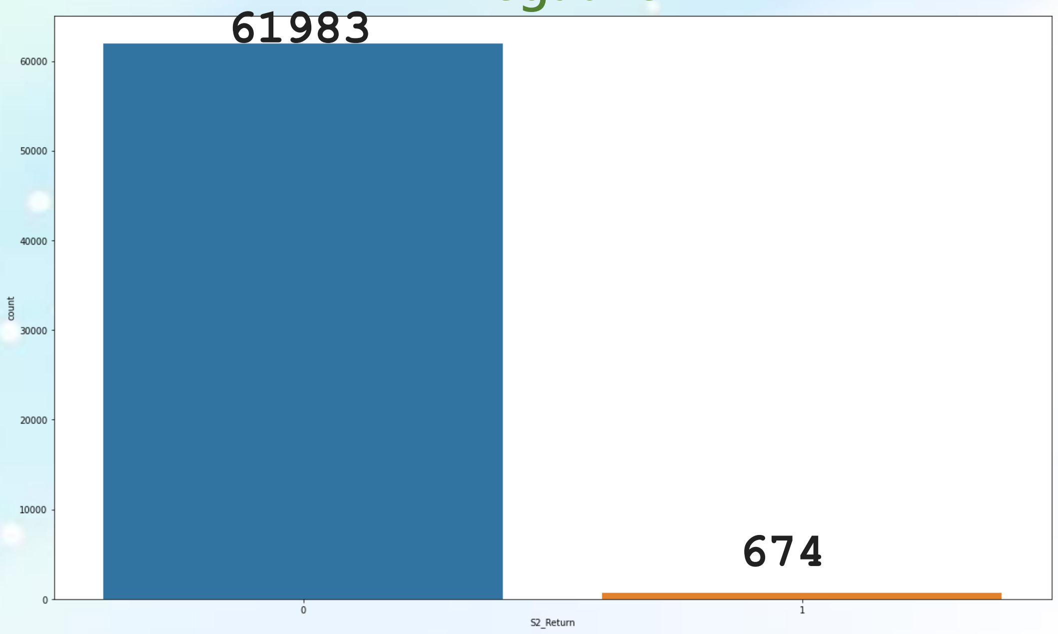
Test Data Predict Probability Distribution



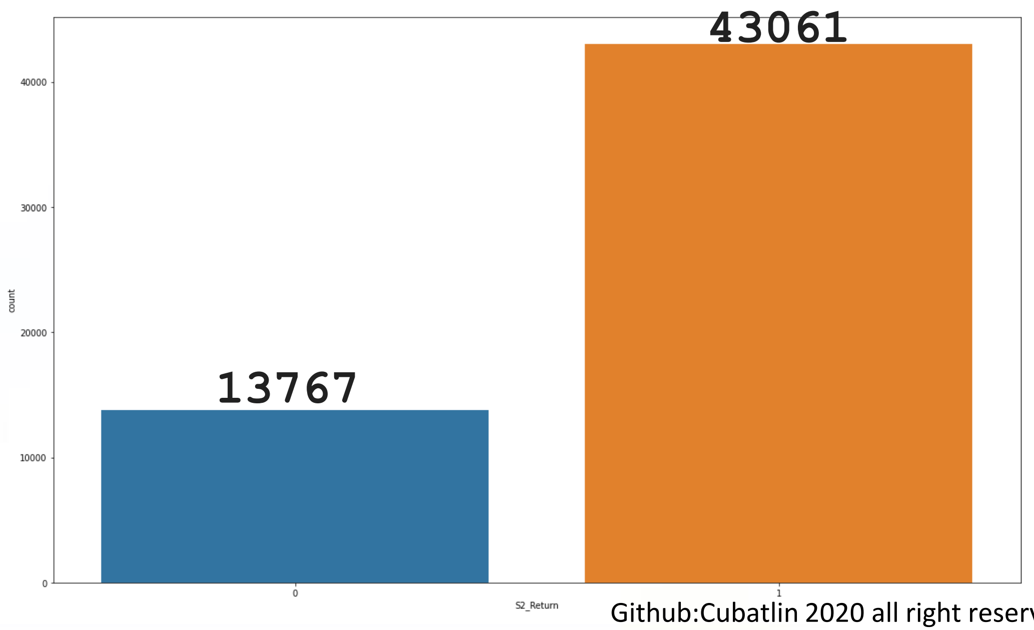
Day-Lag Trading – Monthly Cumulative Return

	TxDate	StockID	有價證券名稱	Open_pre	Open	High	Low	Close	S2_Return	if_neg	if_med	if_pos	pos_1	med_1	neg_1
	2020-09-28	1101	台泥	40.50	40.75	40.95	40.65	40.95	0.011111	0	1	0	0.0	0.0	0.0
	2020-09-29	1101	台泥	40.75	41.20	41.60	41.00	41.20	0.011043	0	0	0	0.0	1.0	0.0
	2020-09-30	1101	台泥	41.20	41.60	41.80	41.25	41.40	0.004854	0	0	0	0.0	0.0	0.0
	2018-01-05	1102	亞泥	28.50	28.75	29.35	28.75	29.35	0.029825	0	0	1	0.0	0.0	0.0
	2018-01-08	1102	亞泥	28.75	29.45	30.35	29.45	30.30	0.053913	0	0	0	1.0	0.0	0.0

Negative

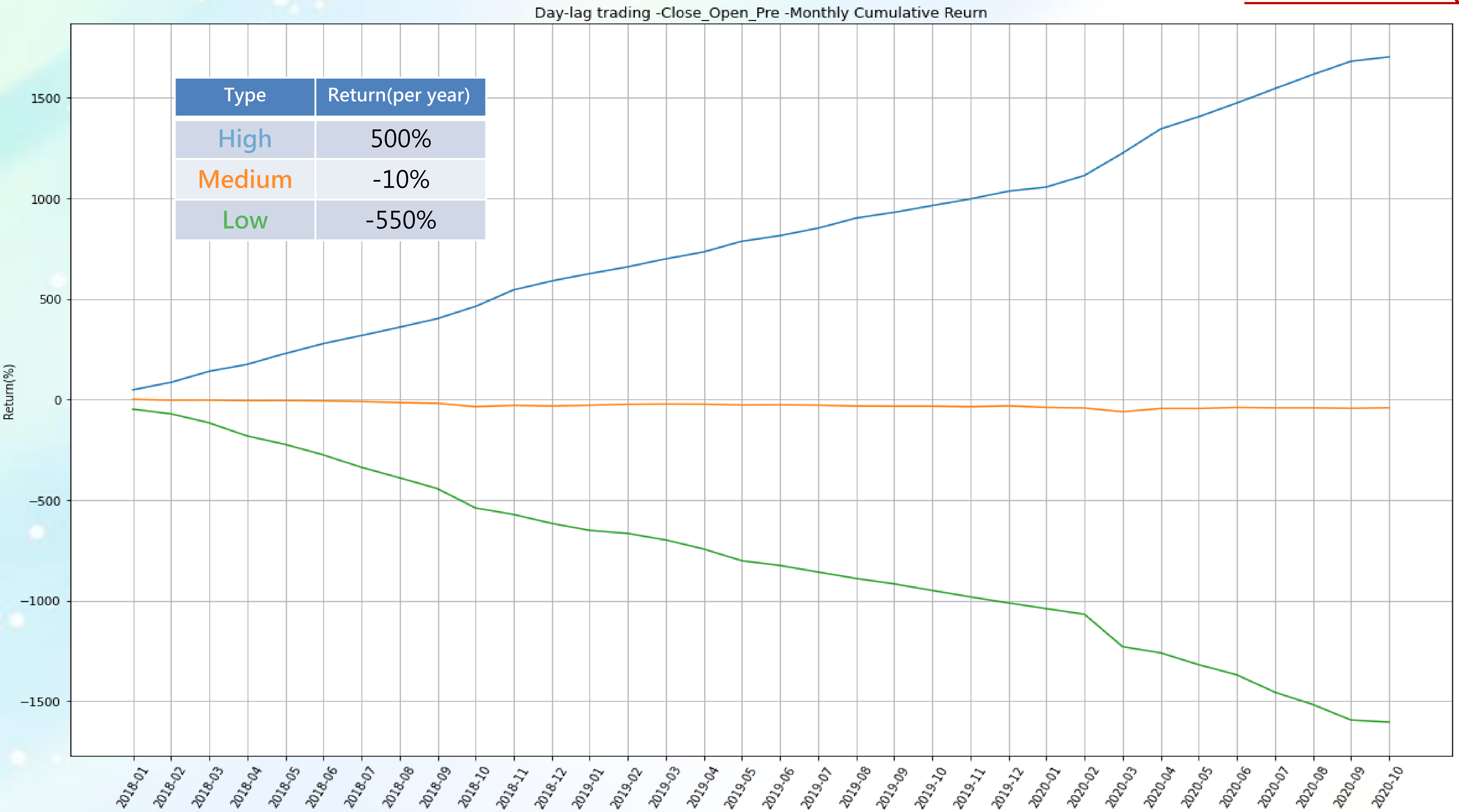


Positive



Day-Lag Trading –Monthly Cumulative Return

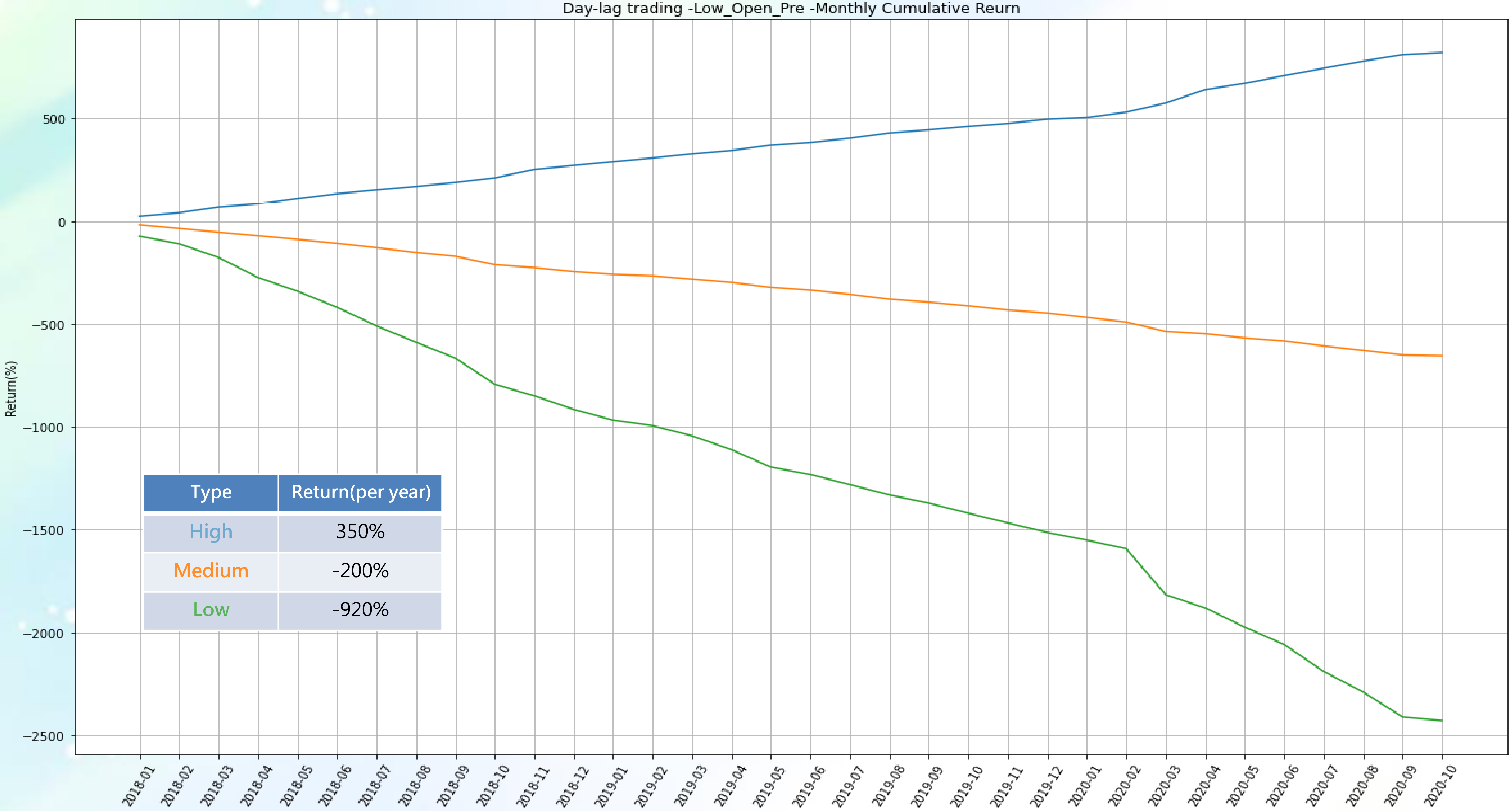
<u>Close - Cost(Open)</u>



Type	Return(per year)
High	500%
Medium	-10%
Low	-550%

Day-Lag Trading –Monthly Cumulative Return

Low-Cost(Open)



Type	Return(per year)
High	350%
Medium	-200%
Low	-920%