



KAGGLE

# JPX Tokyo Stock Exchange Prediction

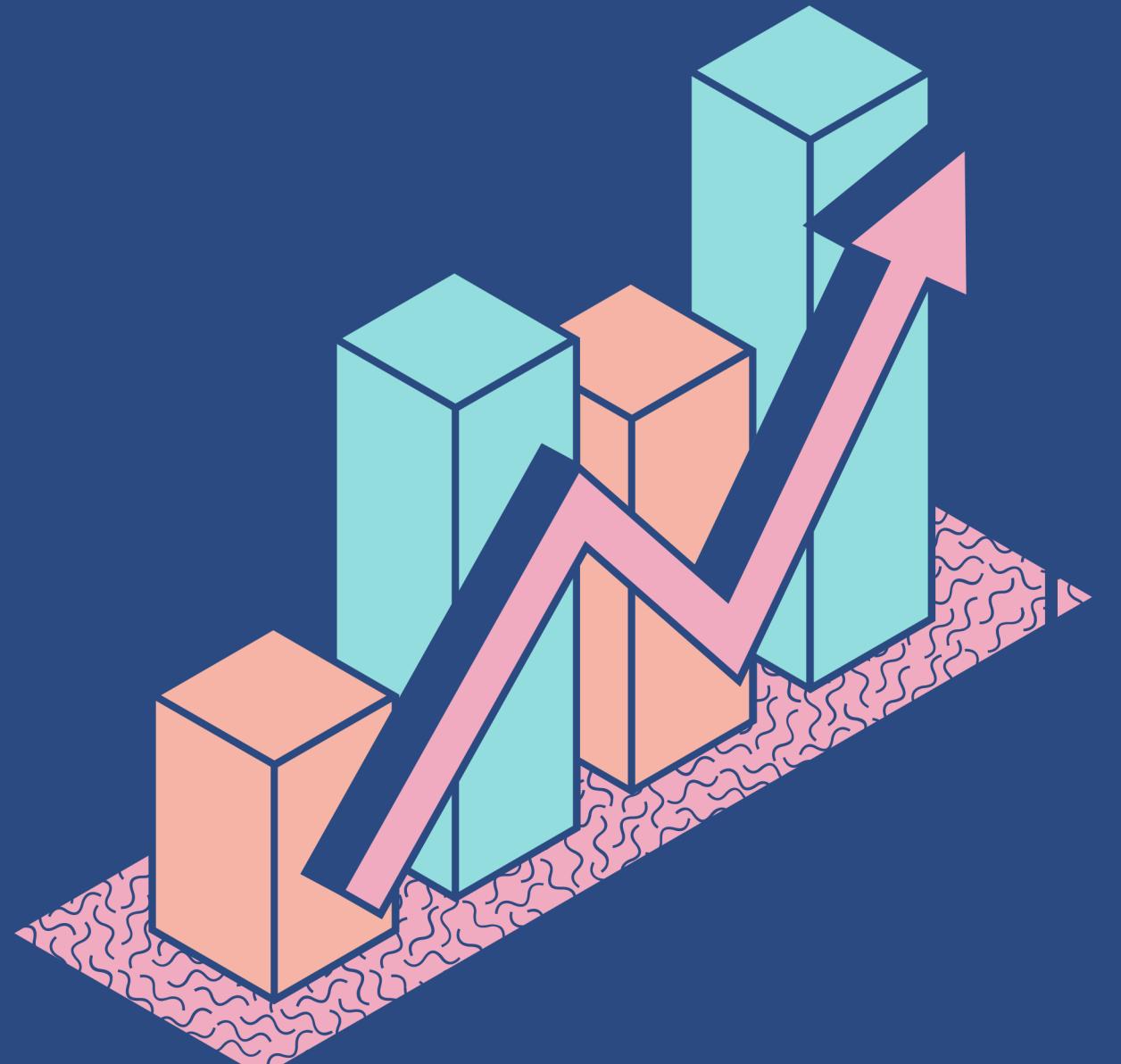
Explore the Tokyo market with your data science skills

# 比賽簡介



比賽將使用日本市場提供的金融數據，  
對大約2000支股票建構投資組合，  
每個參與者必須對股票預期回報進行排名，  
並根據前兩百名和後兩百名股票的回報差異進行評估。

# 評估



1. 對某一天活躍的股票進行排名
2. 單日收益將排名最高（例如 0 至 199）的 200 隻股票視為已購買
3. 排名最低（例如 1999 至 1800）的 200 隻股票視為做空
4. 根據股票的排名對股票進行加權
5. 並假設股票在第二天購買並在次日出售
6. 計算投資組合的總回報

# 數據

## **stock\_prices.csv**

2,000 種最常交易的股票包括每隻股票的每日收盤價和目標欄

## **trades.csv**

前一個工作週的交易量匯總

## **options.csv**

基於大盤的各種期權狀態的數據

## **financials.csv**

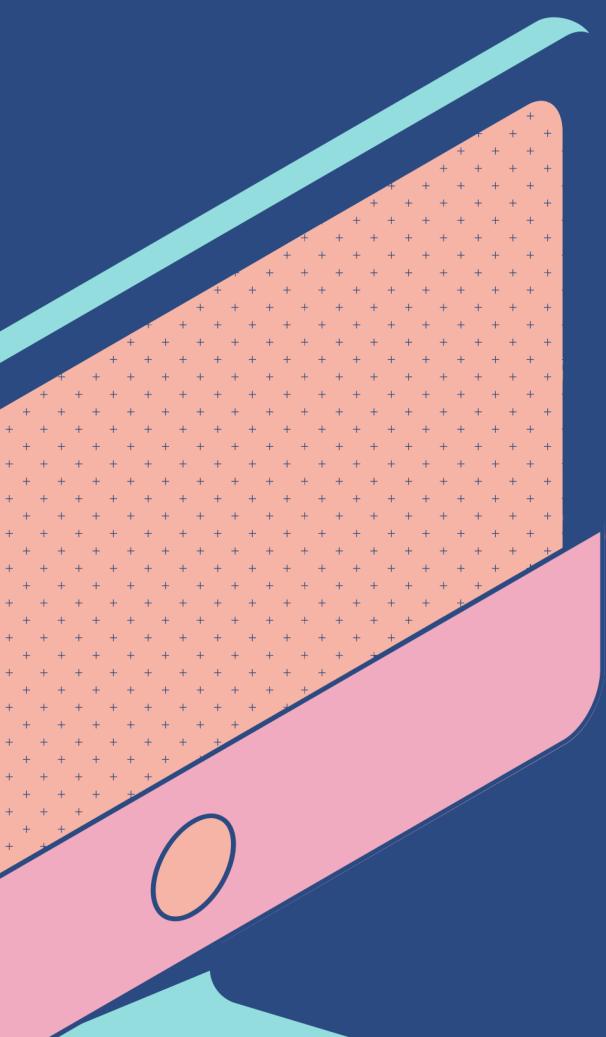
季度收益報告的結果

## **secondary\_stock\_prices.csv**

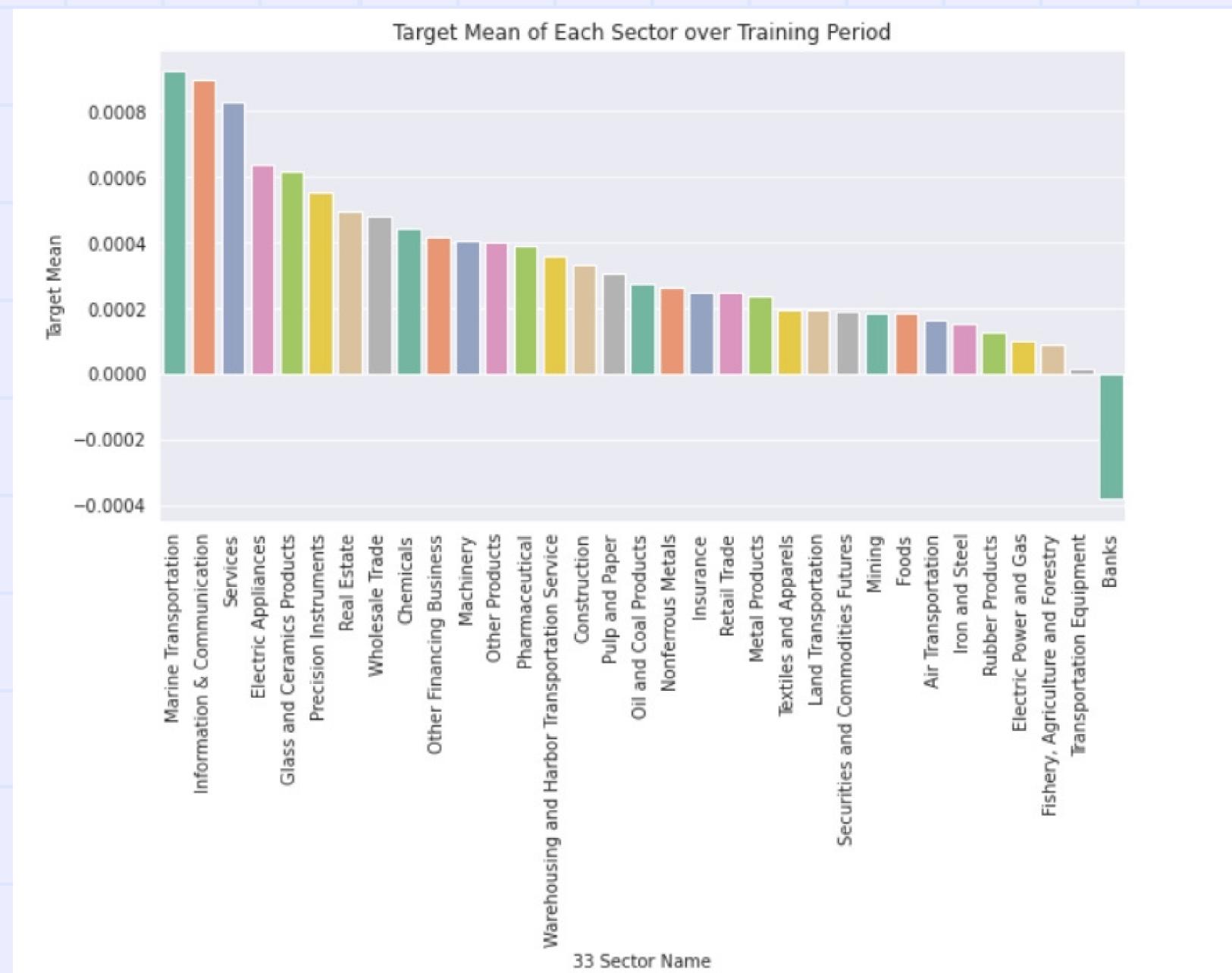
流動性較差的證券數據

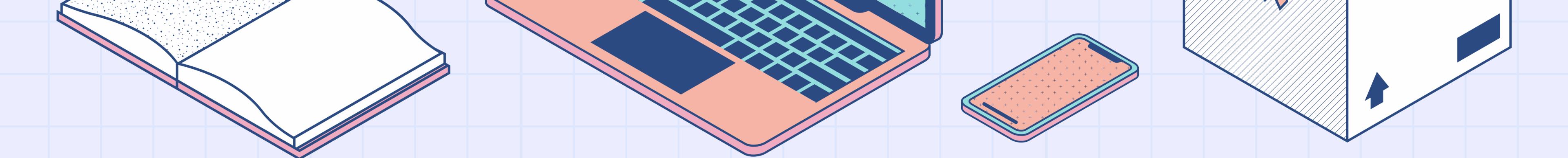
## **stock\_list.csv**

SecuritiesCode對應公司名稱，及公司一般訊息



# 股票清单stock\_list.csv





# 查看訓練數據stock\_prices.csv

	ExpectedDividend	Open	High	Low	Close	Target	RowId	Date	SecuritiesCode	Volume	AdjustmentFactor	SupervisionFlag
	NaN Ratio	99.191222	0.326169	0.326169	0.326169	0.326169	0.010204	0.0	0.0	0.0	0.0	0.0
	RowId	Date	SecuritiesCode	Open	High	Low	Close	Volume	AdjustmentFactor	ExpectedDividend	SupervisionFlag	Target
0	20170104_1301	2017-01-04	1301	2734.0	2755.0	2730.0	2742.0	31400	1.0	NaN	False	0.000730
1	20170104_1332	2017-01-04	1332	568.0	576.0	563.0	571.0	2798500	1.0	NaN	False	0.012324
2	20170104_1333	2017-01-04	1333	3150.0	3210.0	3140.0	3210.0	270800	1.0	NaN	False	0.006154
3	20170104_1376	2017-01-04	1376	1510.0	1550.0	1510.0	1550.0	11300	1.0	NaN	False	0.011053
4	20170104_1377	2017-01-04	1377	3270.0	3350.0	3270.0	3330.0	150800	1.0	NaN	False	0.003026
...	...	...	...	...	...	...	...	...	...	...	...	...
2332526	20211203_9990	2021-12-03	9990	514.0	528.0	513.0	528.0	44200	1.0	NaN	False	0.034816
2332527	20211203_9991	2021-12-03	9991	782.0	794.0	782.0	794.0	35900	1.0	NaN	False	0.025478
2332528	20211203_9993	2021-12-03	9993	1690.0	1690.0	1645.0	1645.0	7200	1.0	NaN	False	-0.004302
2332529	20211203_9994	2021-12-03	9994	2388.0	2396.0	2380.0	2389.0	6500	1.0	NaN	False	0.009098
2332530	20211203_9997	2021-12-03	9997	690.0	711.0	686.0	696.0	381100	1.0	NaN	False	0.018414

2332531 rows × 12 columns

# 數據觀察

**stock\_prices.csv**

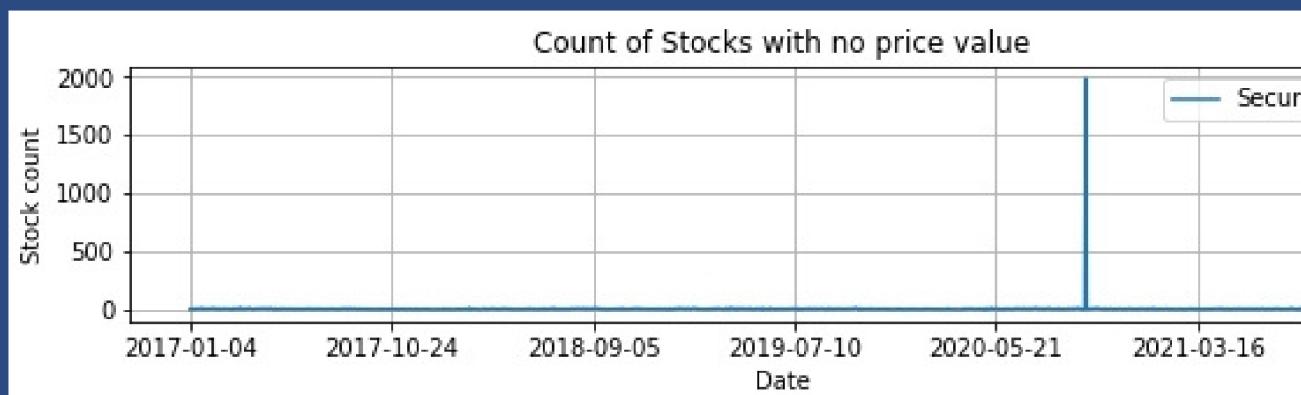
共有1202天資料  
從2017-01-04 到 2021-12-03

**supplemental\_stock\_prices.csv**

共有135 天的資料  
從2021-12-06 到 2022-06-24

93% 的股票中都有 1202 天的數據  
最少資料天數為232

99% 的股票中都有 135 天的數據  
最少的資料天數為94

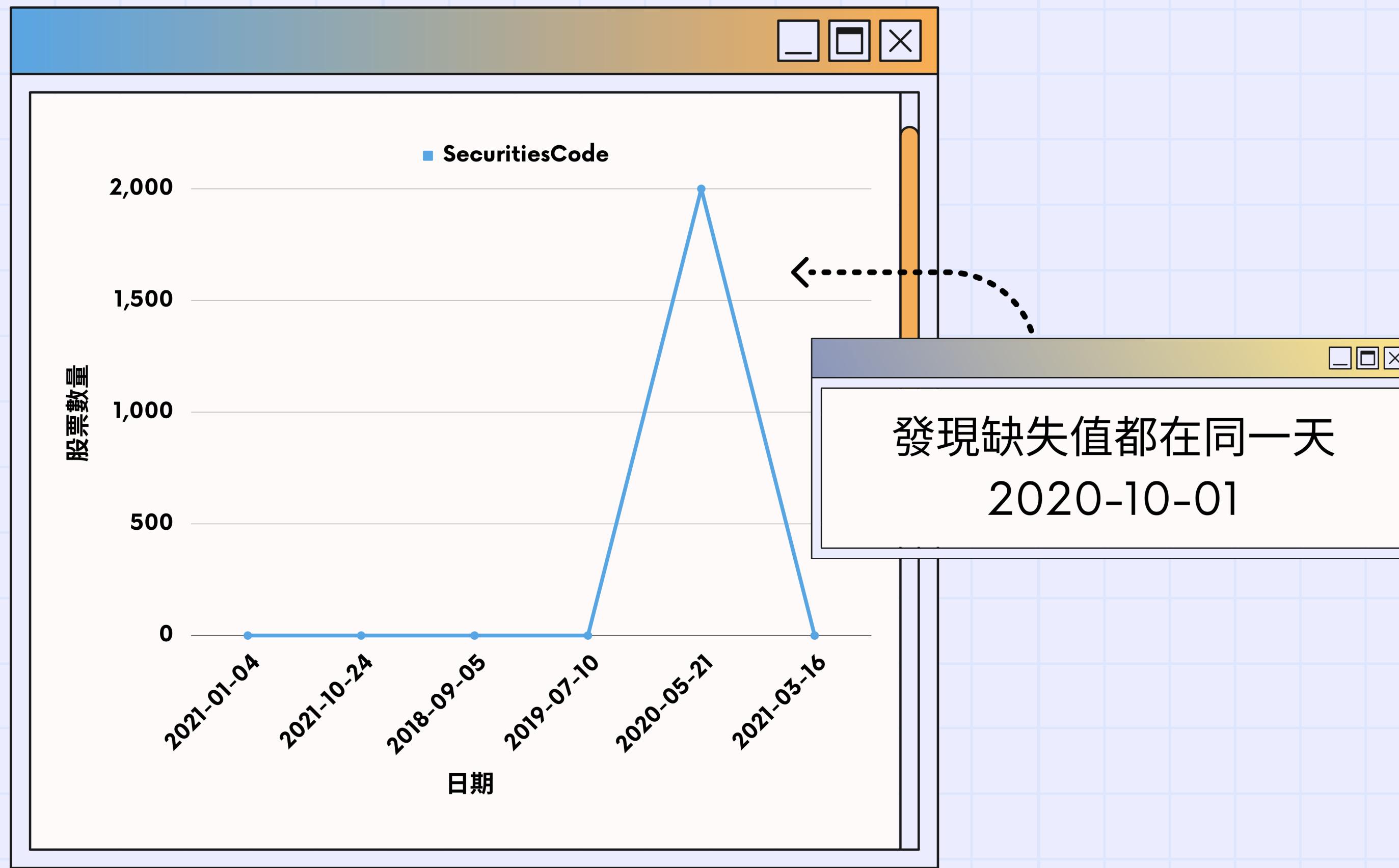


共有7608筆缺失價格資料



共有818筆缺失價格資料

# 有多少隻股票在哪一天有價格缺失值



**JPX October 1, 2020**

Oct. 19, 2020 **JPX** **TSE**

## The Failure of Equity Trading System on October 1, 2020

[Share](#) [Tweet](#)

TSE would like to express our sincerest apologies to investors and the people related to the market f...  
of the equity trading system, arrowhead on October 1, 2020.

TSE has published a report on the sequence of events surrounding the incident, the cause of the failu...  
prevent a recurrence.

In addition, with the aim of considering how it can develop its response to system failures and its rule...  
this incident, TSE will establish a "Council for Recurrence Prevention Measures" comprised of market...

For the details, please refer to the attached files.

# 計算Target

Target為明天收盤價和後天收盤價變化率

Target = (後天收盤價 - 明天收盤價) / 明天收盤

$$r_{(k,t)} = \frac{C_{(k,t+2)} - C_{(k,t+1)}}{C_{(k,t+1)}}$$



itiesCode	Open	High	Low	Close	Volume	AdjustmentFactor	ExpectedDividend	SupervisionFlag	Target	明天收盤	後天收盤	自行計算的Target
1301	2734.0	2755.0	2730.0	2742.0	31400	1.0	NaN	False	0.000730	2738.0	2740.0	0.000730
1301	2743.0	2747.0	2735.0	2738.0	17900	1.0	NaN	False	0.002920	2740.0	2748.0	0.002920
1301	2734.0	2744.0	2720.0	2740.0	19900	1.0	NaN	False	-0.001092	2748.0	2745.0	-0.001092
1301	2745.0	2754.0	2735.0	2748.0	24200	1.0	NaN	False	-0.005100	2745.0	2731.0	-0.005100
1301	2748.0	2752.0	2737.0	2745.0	9300	1.0	NaN	False	-0.003295	2731.0	2722.0	-0.003295
...	...	...	...	...	...	...	...	...	...	...	...	...
1301	2975.0	2984.0	2951.0	2951.0	14300	1.0	NaN	False	0.003793	2900.0	2911.0	0.003793
1301	2953.0	2997.0	2900.0	2900.0	20500	1.0	NaN	False	0.007558	2911.0	2933.0	0.007558
1301	2909.0	2936.0	2909.0	2911.0	10600	1.0	NaN	False	0.016706	2933.0	2982.0	0.016706
1301	2949.0	2973.0	2933.0	2933.0	15200	1.0	NaN	False	-0.003689	2982.0	NaN	NaN
1301	2983.0	2983.0	2965.0	2982.0	11400	1.0	NaN	False	0.031639	NaN	NaN	NaN

# 計算某日股票排名

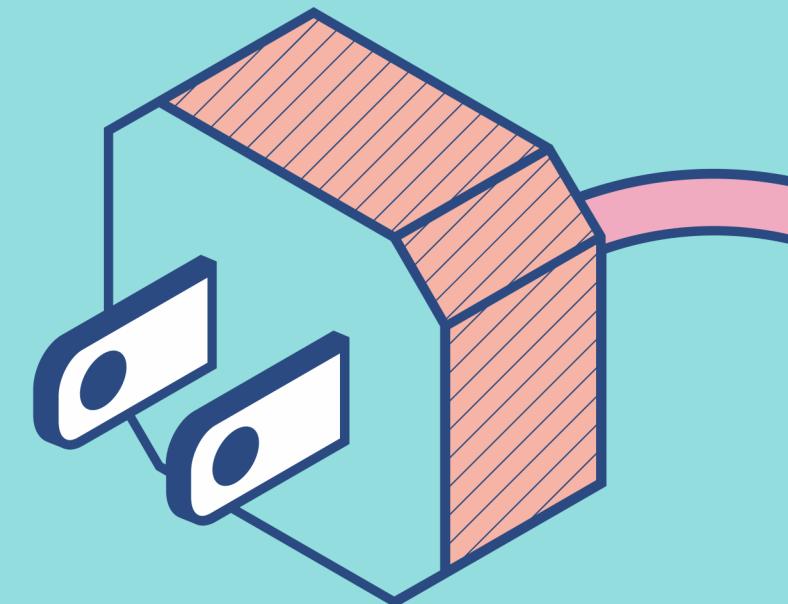
將Target由大到小排序推算出排名  
rank值越小，變化率越大，買入越有利可圖。 rank值越大，變化幅度越大，賣掉有利可圖

	RowId	Date	SecuritiesCode	Open	High	Low	Close	Volume	AdjustmentFactor	ExpectedDividend	SupervisionFlag	Target	Rank
0	20211202_4699	2021-12-02	4699	2037.0	2149.0	2037.0	2143.0	31000	1.0	NaN	False	0.231481	0.0
1	20211202_4488	2021-12-02	4488	5900.0	5990.0	5690.0	5700.0	57100	1.0	NaN	False	0.149254	1.0
2	20211202_8057	2021-12-02	8057	4710.0	4820.0	4615.0	4640.0	29800	1.0	NaN	False	0.123110	2.0
3	20211202_9632	2021-12-02	9632	8110.0	8150.0	8000.0	8030.0	3900	1.0	NaN	False	0.107011	3.0
4	20211202_6958	2021-12-02	6958	510.0	525.0	502.0	517.0	639200	1.0	NaN	False	0.103113	4.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
1995	20211202_4483	2021-12-02	4483	8200.0	8590.0	8180.0	8410.0	320300	1.0	NaN	False	-0.085480	1995.0
1996	20211202_6879	2021-12-02	6879	731.0	745.0	702.0	707.0	735500	1.0	NaN	False	-0.087248	1996.0
1997	20211202_4168	2021-12-02	4168	4280.0	4390.0	4200.0	4335.0	72600	1.0	NaN	False	-0.094595	1997.0
1998	20211202_3994	2021-12-02	3994	7490.0	7500.0	7060.0	7070.0	735000	1.0	NaN	False	-0.094980	1998.0
1999	20211202_4169	2021-12-02	4169	7160.0	7870.0	7110.0	7520.0	783000	1.0	NaN	False	-0.127424	1999.0

2000 rows × 13 columns

# API迭代資訊

prices	options
某一天2,000 種最常交易的股票資訊，不包括 Target	基於大盤的各種期權狀態的數據
financials	trades
季度收益報告的結果	前一個工作週的交易量匯總
secondary_prices	sample_prediction
流動性較差的證券數據	提交範例，要更改其Rank欄位資訊



# 提交內容

按日期和證券代碼（股票）預測排名

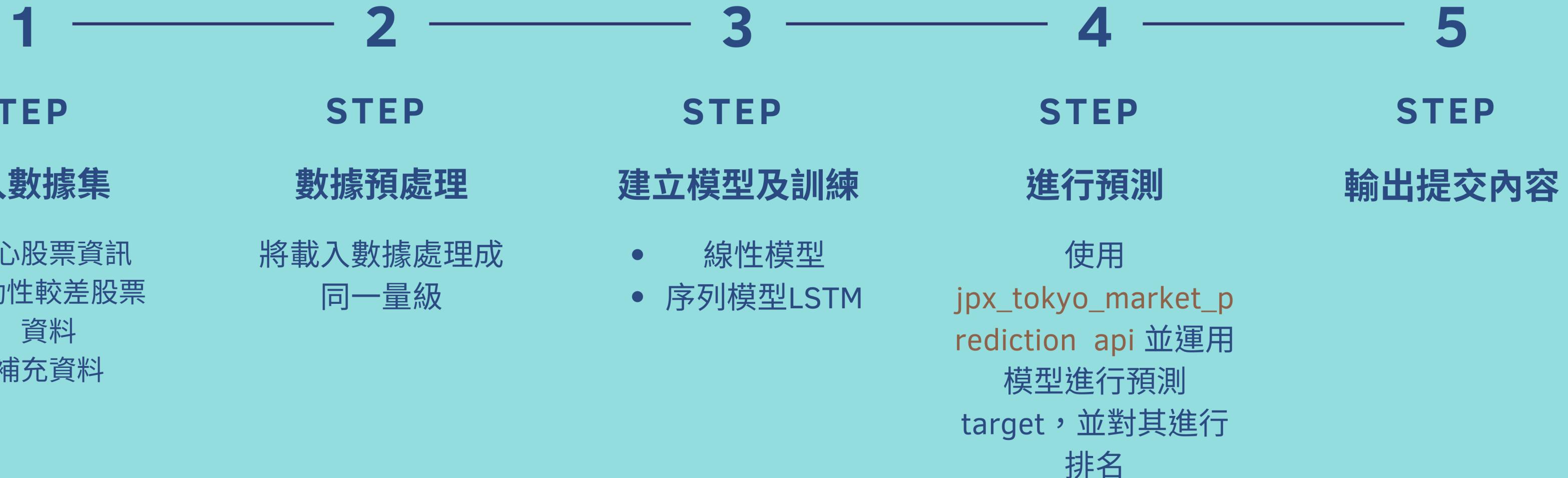
Rank表示2000隻股票預期收益的排名



Date	# Securities...	# Rank
2021-12-06	1301	0
2021-12-06	1332	1
2021-12-06	1333	2
2021-12-06	1375	3
2021-12-06	1376	4
2021-12-06	1377	5
2021-12-06	1379	6
2021-12-06	1381	7
2021-12-06	1407	8
2021-12-06	1413	9
2021-12-06	1414	10
2021-12-06	1417	11
2021-12-06	1419	12
2021-12-06	1429	13
2021-12-06	1431	14

sample\_submission.csv

# 我的想法



# 參考來源

<https://www.kaggle.com/competitions/jpx-tokyo-stock-exchange-prediction>

<https://www.kaggle.com/code/smeitoma/jpx-competition-metric-definition>

<https://www.kaggle.com/code/chumajin/easy-to-understand-the-competition>

<https://www.kaggle.com/code/kellibelcher/jpx-stock-market-analysis-prediction-with-lgbm>

<https://medium.com/@rehman.aimal/analysis-of-stock-pricing-data-of-jpx-tokyo-stock-exchange-ae7dfa35ed5d>

<https://www.kaggle.com/code/cv13j0/jpx-tokyo-stock-exchange-prediction-xgboost>

<https://www.kaggle.com/code/abaojiang/jpx-detailed-eda>

感謝聆聽

