第金 Golden

一股 Stock

屆獎 Awards

PTT股版績效回測及生態分析

好想吃午餐

經濟四 王詩蘋

國企四 林潔玟

財金四劉益嘉

經濟二 劉恩甜

研究動機:相信股版標的文會不會變韭菜?











實踐方法

1. 爬蟲:

- 股版[標的]文(2018.01.01-2021.06.01), 留言數、推&噓數、有無附圖等。
- 標的文數前 200 名個股近三年的每日收盤價

2. 資料清洗:

- 符合格式(標的、分類、分析/正文、進退場機制)、清除雜字、判別多空、長短線
- 依據分類判斷多空, 依據關鍵字判斷長短, 優先判斷長線, 剩餘則為短線
- 長:長線、長期投資、長期、不停利、不停損、長投、放(但不是"放空)
- 短:短線、(有"停利""停損"但沒有"不停利""不停損")、放空、退場、出場
- 設定投資期間並計算績效,長線時間250天,短線時間5天,截至6/10為止

3. 資料分析:

相關性分析、情緒分析、文本分析

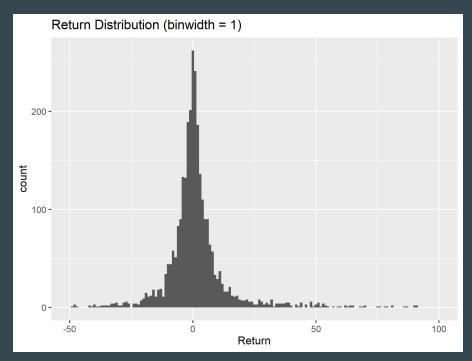
- 簡單敘述性統計
- 投資期間敏感度分析
- 股王排名
- 績效與留言數、推&嘘、有無附圖
- 績效與情緒分數
- 股價與情緒分數
- 詞頻表
- 文字雲
- Topic Modeling

簡單敘述性統計

- 總計 5012 篇標的文、多/空=4058/954、長線/短線=888/4124
- 平均每位鄉民發了 2.09 篇標的文
- 平均每篇(可計算績效的)標的文的年化報酬率為15.65%, n = 2907
- 發文數量前10名的作者, 其平均年化報酬率為26.18%
- 各年度發文篇數

2018:1342篇、2019:1045篇、2020:1503篇、2021(上半年):1122篇

簡單敘述性統計



*	累積報酬率趨近常態分配,	七尺名杨晔	店
•	希惧拟即平烟灯市恐刀趾,	但尼罗 奥州	且。

Target_name <chr></chr>	n <int></int>
台積電	174
大盤	154
長榮	76
鴻海	59
國巨	54
穩懋	49
聯電	44
聯發科	40
華新科	35
群創	35

投資期間敏感度分析:短期年化有顯著差異、長期則無

短線	3天	5天	7天
平均累積報酬率	0.26%	0.33%	0.21%
平均年化報酬率	21.89%	16.42%	7.43%

長線	250天	300天	350天
平均累積報酬率	11.52%	15.33%	18.08%
(完整天數)	10.86%	14.35%	13.32%
平均年化報酬率	11.52%	12.77%	12.92%
(完整天數)	10.86%	11.96%	9.52%

績效與大盤比較:股版平均報酬率未必能超越大盤

• 以年化報酬率進行比較

	2018	2019	2020	2021(至6/10)
大盤	-8.6%	23.3%	22.8%	16.57%
PTT股版	23.78%	8.5%	24.65%	1.7%
發文前50名 平均報酬	32.58%	7.5%	32.31%	17.8%
績效前50名 平均報酬	460.21%	304.41%	591.45%	689.71%

股王排名

● 最佳績效獎(累積報酬率)

時間	標的	作者	操作	績效
2020/11/23	長榮	cinghsang	長多	287.67%
2020/5/8	圓剛	ALAN	長多	255.31%
2020/4/28	聯電	dust	長多	247.62%
2020/12/31	高端疫苗	米食主義者	長多	226.15%
2020/3/17	群創	南茂哥	長多	188.65%

股王排名

• 最佳效率獎(年化報酬率)

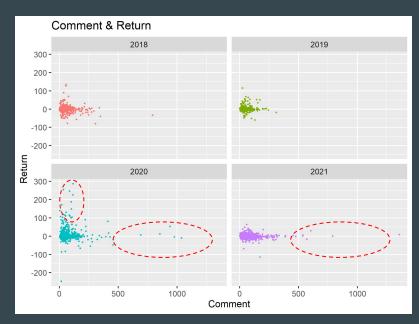
時間	標的	作者	操作	績效
2020/4/27	康那香	還敢皮啊	短多	2391.17%
2020/6/17	愛普	麻瓜	短多	2299.84%
2020/7/17	美德醫療-DR	高級本省人	短多	1859.67%
2021/5/30	高端疫苗	酵公菌	短空	1833.62%
2020/4/20	寶齡富錦	股票戰船	短多	1828.61%

股王排名

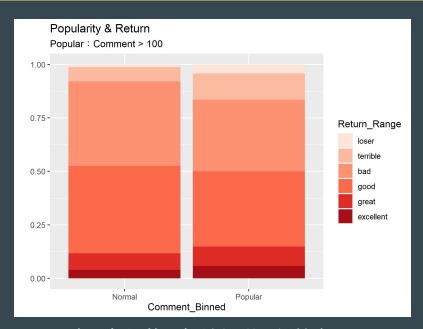
● 股海明燈獎(發文數超過5篇且累積報酬率最高)

作者	篇數	平均績效
南茂哥	9	42.87%
Mr.King	11	29.73%
好 欣	13	23.04%
Chupei	9	19.56%
馬芙丸5566	7	18.09%

績效與留言數

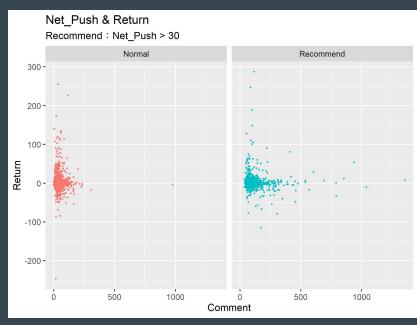


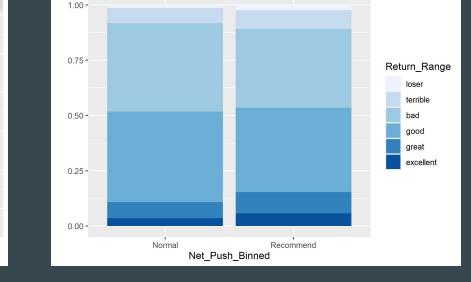
- ❖ 2020後標的文熱度高
- ❖ 2020高報酬標的文比其他年多,推測為 疫情關係



- ❖ 高人氣標的文極端報酬比例較高
- ◆ 報酬率以30%、10%、0%、-10%、-30%為 界

績效與推&嘘:無顯著差異





Net Push & Return

Recommend: Net Push > 30

❖ 高淨推數文資料點較為分散、報酬範圍 較大

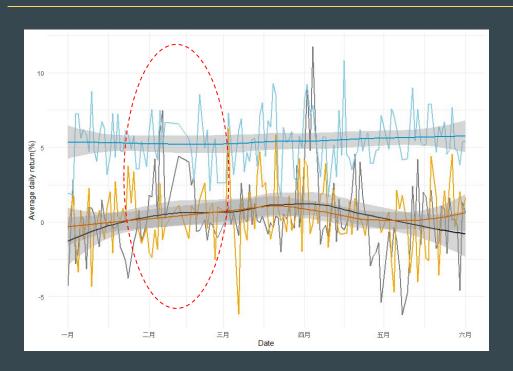
❖ 高淨推數文極端報酬比例稍高

績效與有無附圖:無顯著差異





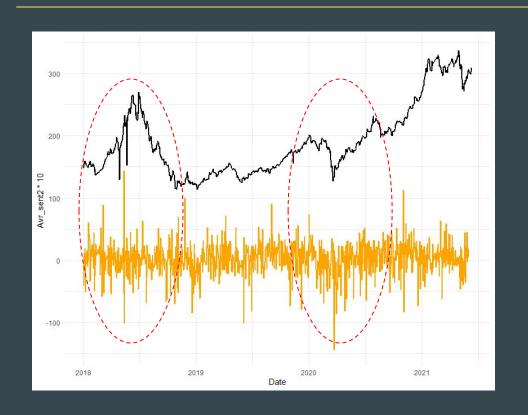
績效與情緒分數:變動趨勢相近



- ❖ 用詞強烈程度同時包含正向及負 向情緒,因此分數較高
- ◆ 情緒相關分數變動與績效變動趨 勢相近
- 使用中文維度型情感字典
- 用詞強烈程度=Arousal mean * abs (Valence mean 5)
- 情緒正向/負向程度=Arousal mean* (Valence mean - 5)
- 黑:每日平均績效
- 淺藍:用詞強烈程度
- 橘:情緒正向/負向程度

以2021年為例

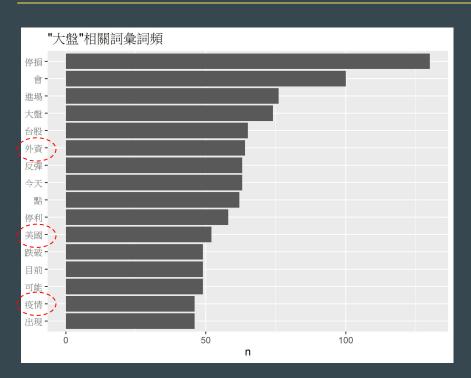
股價與情緒分數:放大情緒波動後可見與股價存在關係

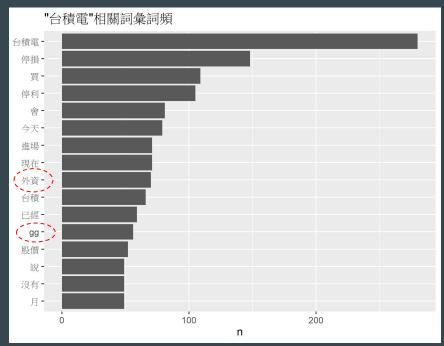


◆ 在2018年股價兩次急跌以及2020 年疫情爆發時最為明顯

- 黑:股價
- 橘:情緒正向/負向程度
- 以最熱門10檔每日股價(佔所有資料1/3左右) /情緒正向/負向程度(*10倍以看出波動大小)

詞頻表





*

文字雲:歷年無特別明顯變動

```
市場
突破
消息面
```

2018:「外資」詞頻進入前五

2019:「美國」跌榜

消息面 「貿易戰」上榜

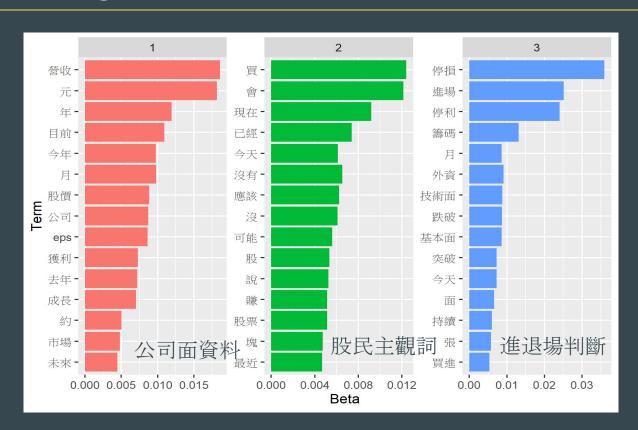
文字雲:歷年無特別明顯變動



自 很多 因此 買進

◆ 2021:「疫情」持續上榜

Topic Modeling



結論:To be 韭菜 or not to be, that is the question.

- 股版年化報酬率每年未必能超越大盤
- 累積報酬率趨近常態分配
- 股市變動與PTT內文情緒分數趨勢相近,但目前尚未觀察出因果關係
- 標的文最終的績效表現與文章熱門程度(推、 留言)無顯著關係
- 透過詞頻及文字雲觀察股市熱門字詞歷年無明顯變動



分工表

組員	分工內容	投入程度
詩蘋	Google爬蟲、PTT文章爬蟲 資料清洗、書面報告	6
益嘉	股價爬蟲、資料清洗、 績效計算分析、Topic Model 口頭報告	6
恩甜	詞頻表、文字雲、文本相似度分析	6
潔玟	情感分析、績效與發文數分析 口頭報告	6



感

謝

聆

Adapted from 第54屆金馬獎