

The background is a solid dark grey color. Scattered around the central text are white line-art illustrations of various items: a pen and a highlighter in the top left; a wallet in the top center; a tablet or laptop in the top right; a pair of headphones on the right side; a cup of coffee on the left side; a calculator in the bottom right; a pair of earbuds with a cord in the bottom left; and a laptop in the bottom center. The central text is in a large, white, sans-serif font.

# 這些上市電子股 你買對了嗎？

第1組 許皓維 孫佑鈞 林孟萱 洪季芃

# 問題定義與說明

- 以巴菲特聞名的「價值投資」為概念發想，提出以公司**基本面(財報)**預測公司**中長期股價走勢**。
- 目前普遍投資人觀察一間公司營運狀況時，多數**僅參考EPS、殖利率、營收月增率與營收年增率**等零星變數，但財報變數項目眾多，如何**提取更多能反應股價的變數**為本研究目的之一。

# |資料來源與前處理步驟

- 研究對象：上市電子股族群。
- 資料來源：
  - TEJ財報資料。
  - FinMind 股價資料<sup>[1]</sup>。
- 資料區間：2010年1月至2020年12月，以**季**為單位。

[1] <https://finmindtrade.com/>

# 資料來源與前處理步驟

## • 資料前處理

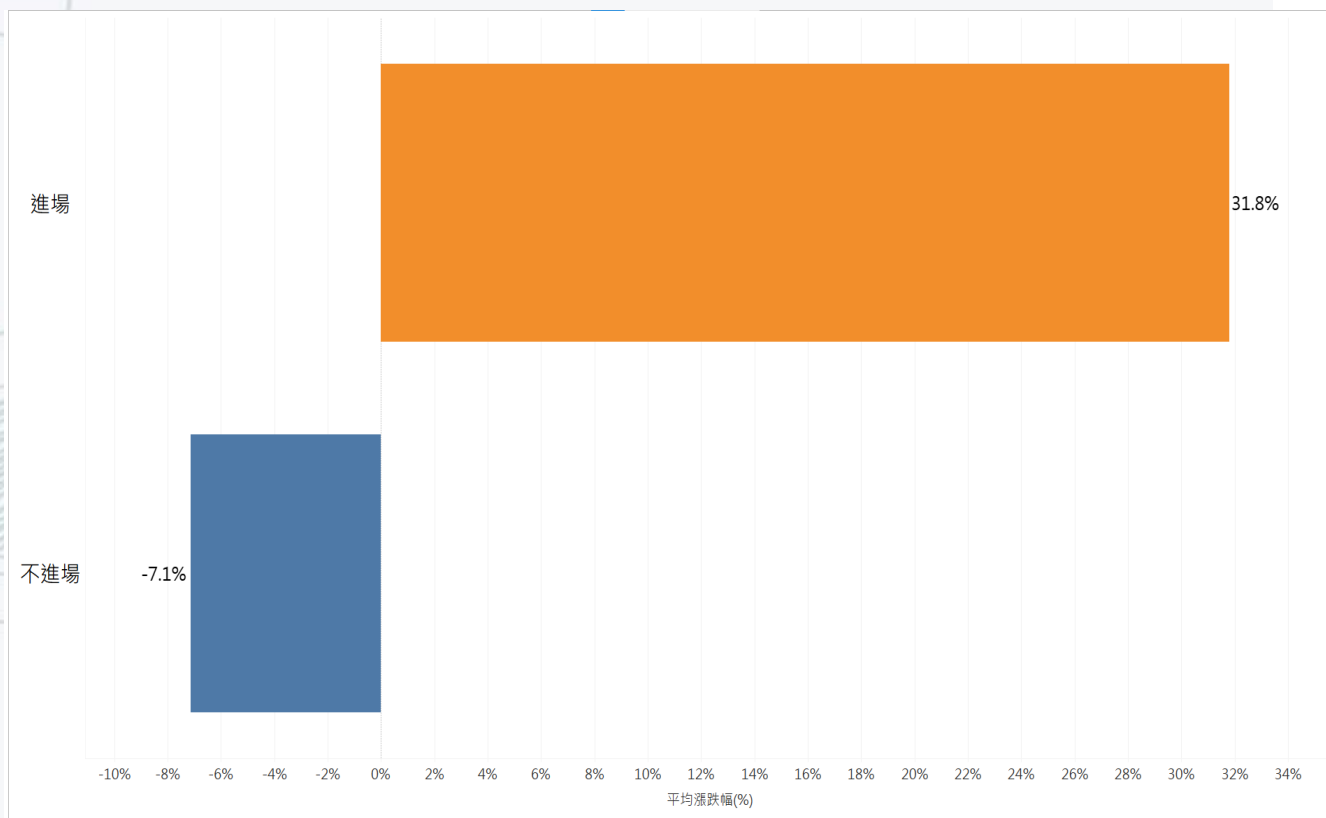
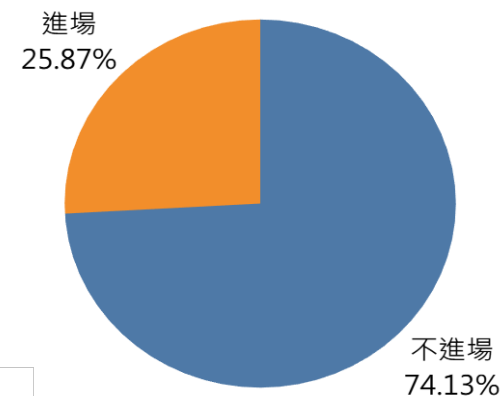
- 合併各公司三大報表(資產負債表、損益表、現金流量表)、獲利指標及成長率指標。
- 計算各公司單季營收與季營收成長率。
- 股價：日K轉換為季K，計算該季與前季的股價漲跌幅。
- 進場訊號：根據股價季成長幅度是否達標10%。
  - 1→股價漲跌幅 $\geq 10\%$
  - 0→股價漲跌幅 $< 10\%$

# 資料來源與前處理步驟

- 遺失值處理

- 考量到部分公司未使用IFRS前，其營收資訊不完整且不具可回溯性，因此排除無法取得公司完整財報之季度資料。
- 僅保留公司上市後的數據。若公司於2010年1月後上市，上市前的財報資料則不列入本次資料集範圍中。
- 其餘欄位遺失值以回溯歷史財報資料的方式進行填補，由組員分工合作完成。

# 資料EDA與視覺化解讀



# |如何分工

- 許皓維：  
資料蒐集、資料探勘、研究規劃、模型建置
- 孫佑鈞：  
資料清理、研究規劃、模型建置
- 林孟萱：  
資料蒐集、資料清理、特徵欄位提取、提供會計專業知識
- 洪季芃：  
資料蒐集、資料清理、特徵欄位提取、提供會計專業知識



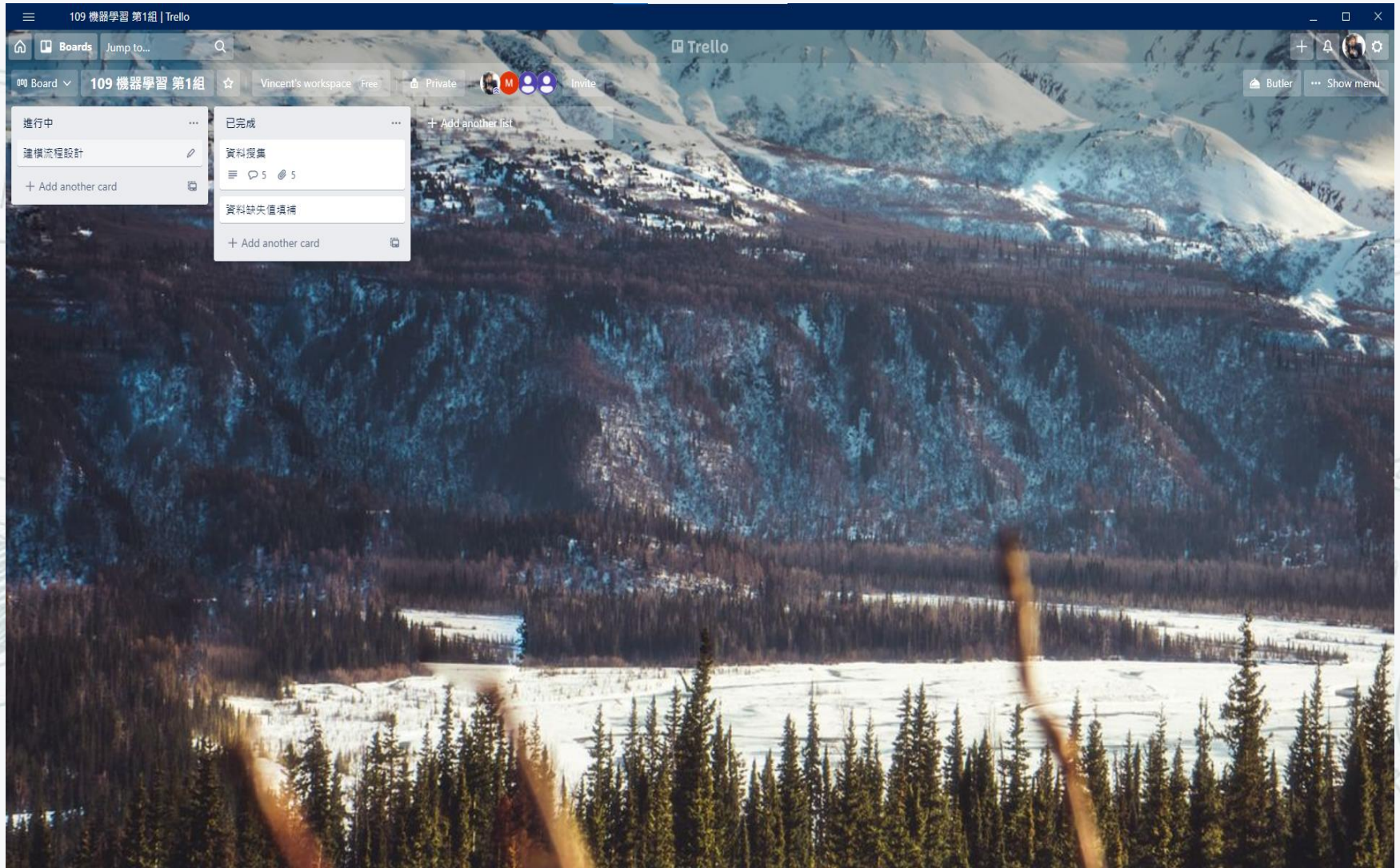


# |如何解決問題

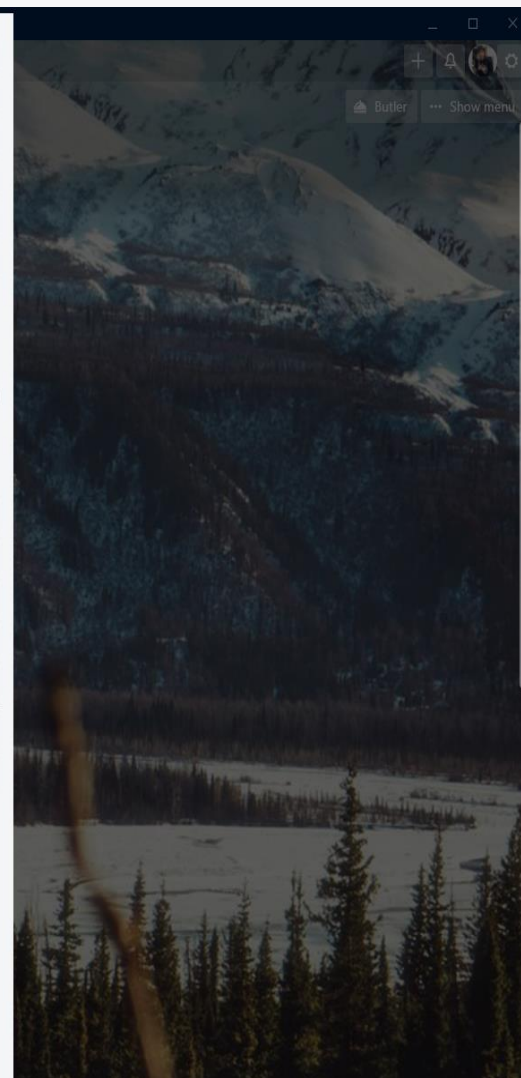
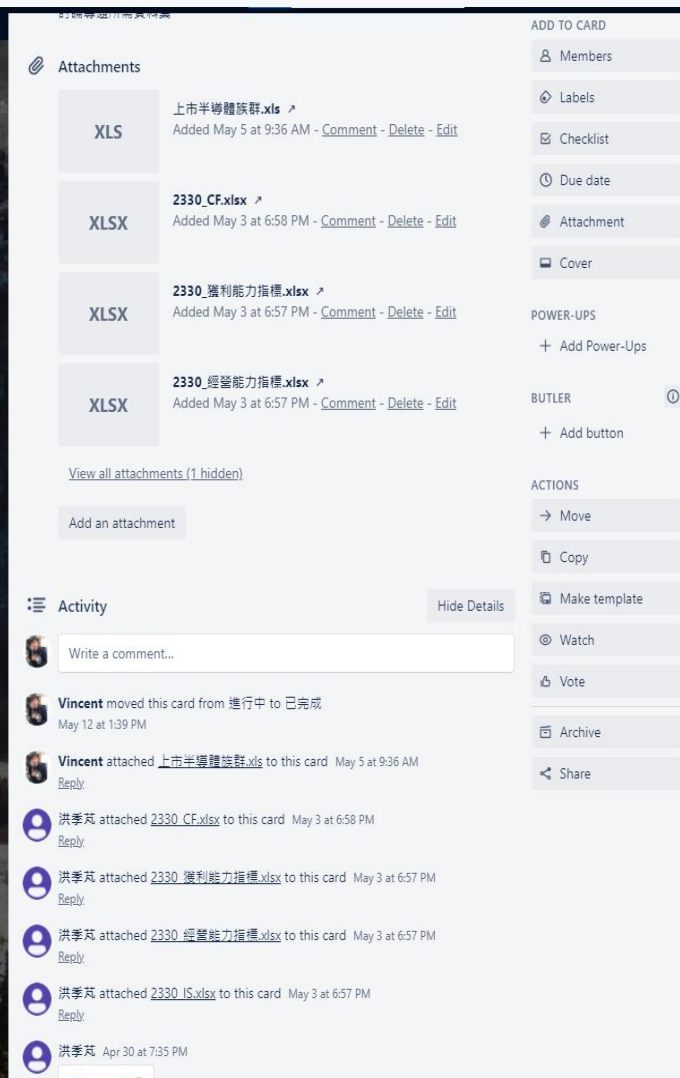
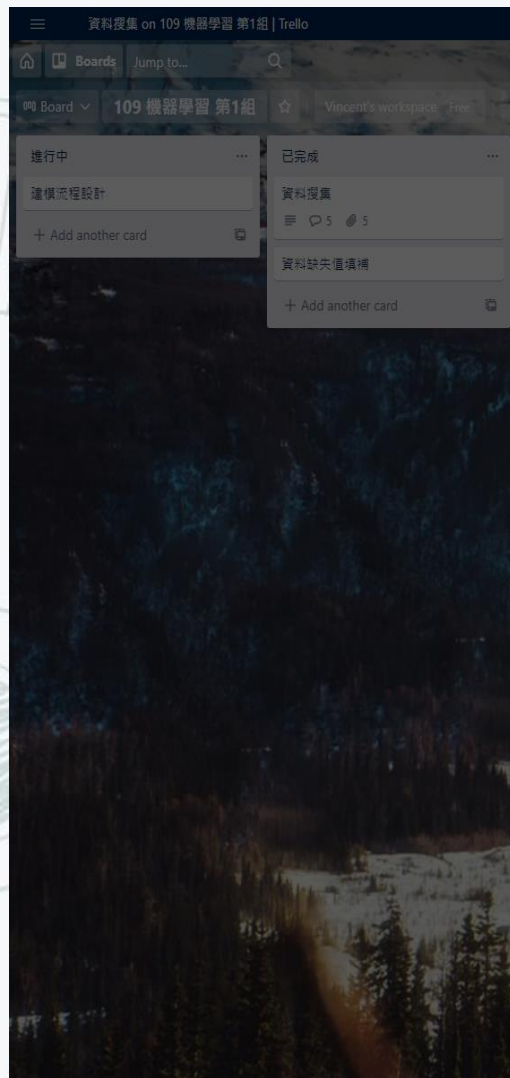
- 本組採用Trello與Line群組雙軌的討論路徑，從資料蒐集到模型所需的資料，由**四位組員齊心協力討論出最佳答案並分工進行**，力求最完整且具有參考性的預測股價進場方案。



# Trello



# 討論記錄





# Github

Search or jump to...

Pull requests Issues Marketplace Explore

🔔 + 🌐

haowei0215 / Machine-Learning-Course

👁 Unwatch 1 ⭐ Star 0 🍴 Fork 0

<> Code ⓘ Issues 🔄 Pull requests ⚙ Actions 📁 Projects 📖 Wiki 🔒 Security 📈 Insights ⚙ Settings

main Machine-Learning-Course / Data Cleansing.py / <> Jump to

Go to file

...

haowei0215 Update Code and Data

Latest commit 598dd7f 3 minutes ago 🕒 History

1 contributor

109 lines (92 sloc) 5.73 KB

Raw Blame 🗑 📄 📝

```
1 import pandas as pd
2 import numpy as np
3 import glob, math
4
5 # 合併月營收資料
6 val = pd.read_excel(r"./Data/營收/上市電子業_營收.xlsx")
7 val['年份'] = val['年月'].apply(lambda x: x[:4])
8 val['月份'] = val['年月'].apply(lambda x: int(x[5:]))
9 val['季'] = val['年月'].apply(lambda x: math.ceil(int(x[5:])/3))
10 val_season = val.groupby(['代號', '年份', '季'])['單月營收(千元)'].count().to_frame('紀錄').reset_index()
11 val_season = val_season[val_season['紀錄']==3]
12 save_list = list(zip(val_season['代號'], val_season['年份'], val_season['季']))
13
14 new_val = pd.DataFrame()
15 for x, y, z in save_list:
16     a = val['代號']==x
17     b = val['年份']==y
18     c = val['季']==z
19     tmp = val[(a)&(b)&(c)]
20     new_val = pd.concat([new_val, tmp], 0)
21
22 season = new_val.groupby(['代號', '年份', '季'])['單月營收(千元)'].sum().to_frame('單季營收(千元)').reset_index()
23 new_val = pd.merge(new_val, season, on=['代號', '年份', '季'], how='left')
24 new_val['單季營收成長'] = new_val['單月營收(千元)'].shift(-2)
25 new_val['單季營收成長'] = (new_val['單季營收成長'] - new_val['單月營收(千元)']) / new_val['單月營收(千元)']
26 new_val['單季營收成長'] = new_val.apply(lambda r: r['單季營收成長'] if r['月份'] in [1, 4, 7, 10] else None, 1)
```

**THE END**

