

中美貿易戰

對台灣、美國市場

不同類型ETF的影響

LOGO

顏維君、郭庭瑜、李孟哲

/專案目標

中美貿易戰下
各類產業與投資人的
投資決策判定



2018/ 7 ~ 2019/ 5

美國公布2000億美元關稅
清單

美國開始針對2000億
美元清單課稅

中國公布600億美元關稅
清單

美國另外公布3000億
美元關稅清單

標普五百
美元
黃金
民生用品
能源
金融



台灣五十
台商五十
科技
金融
台積電

/ 專案概述

經濟日報



REUTERS

Reuters



Bloomberg



The New York Times

/ 研究步驟概述

- 一、決定研究標的及範圍 → 篩選ETF
- 二、新聞分類 → 找出正負面新聞並量化
- 三、關聯性顯著程度 → 統計迴歸分析

/ 研究步驟概述

- 一、用python、bs4爬蟲並整理資料
- 二、用nltk情緒分析、kmeans將新聞屬性分類
- 三、用r語言迴歸分析、python計算相關係數

研究步驟



一、決定研究標的及範圍： → 篩選ETF

- 不同地區
- 不同資產類別 ex. Bond, Equity...
- 指數型
- 不同產業 ex. FB科技、金融、摩台

資產類別					
股票型	台灣元大50 (0050)	元大台商50 (0054)			
產業類別	元大電子 (0053)	富邦科技 (0052)	富邦金融 (0059)		
非ETF	台積電 2330	鴻海 2317	南亞 1303	台塑 1301	國泰金 2882

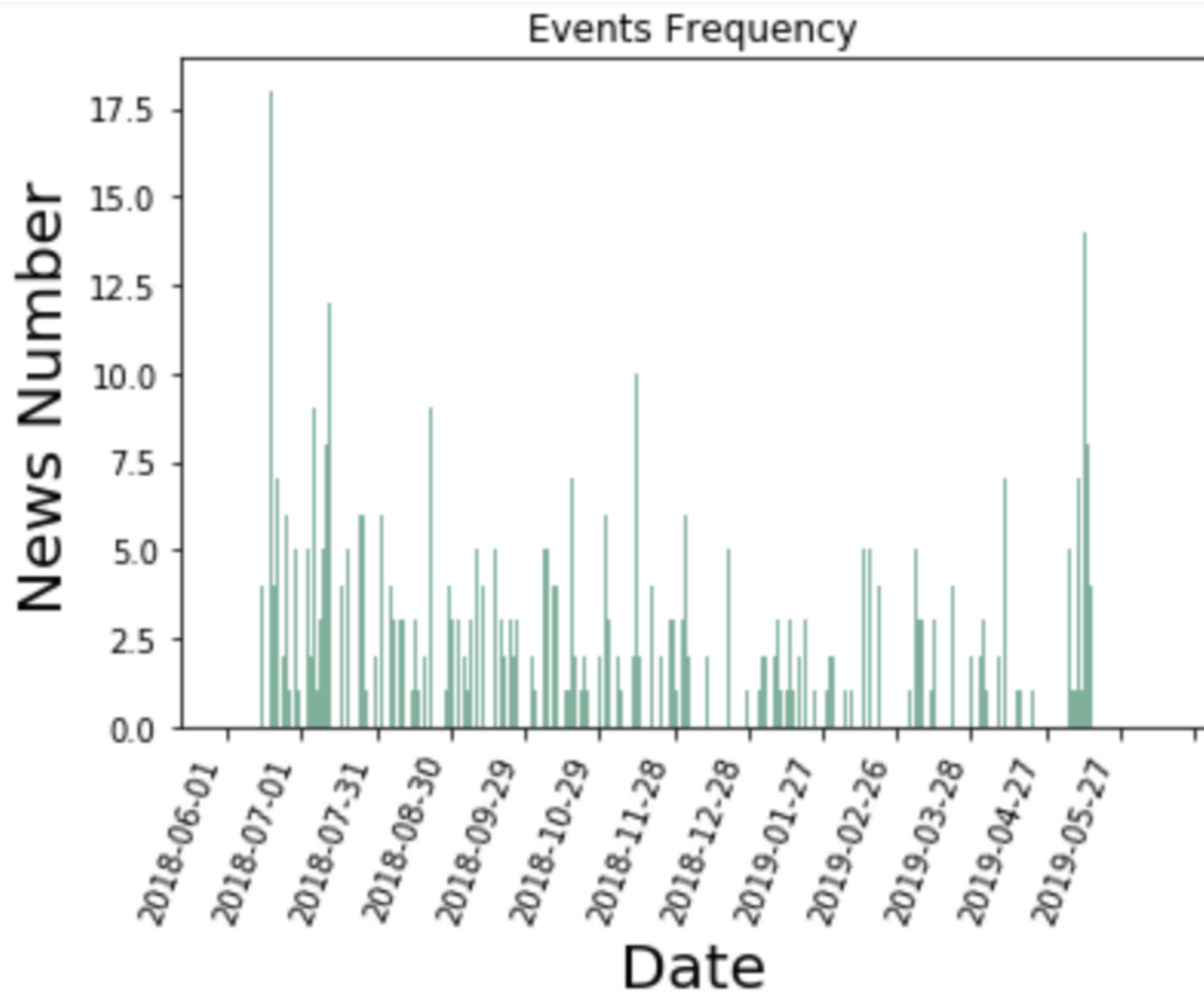
資產類別	Ticker	追蹤標的物
Equity	SPY	S&P 500
Bond	AGG	All-Term bonds
Currency	UUP	USD
Real Estate	VNQ	US real estate
Commodity	GLD	Gold Bullion
產業類別 (Sector)		
Consumer Staples	VDC	民生用品：Coca-Cola, Walmart, Costco...
Energy	VDE	Oil, Gas Corp.
Financials	VFH	JPM, Bank of America, Goldman Sachs...
Industrials	VIS	Boeing, 3M...
Materials	VAW	
Technology	VGT	Microsoft, Apple, Visa, Mastercard, Intel...
Telecom	VOX	Facebook, Alphabet, AT&T, Netflix...

/ 研究步驟



二、新聞分類： →找出正負面新聞並量化

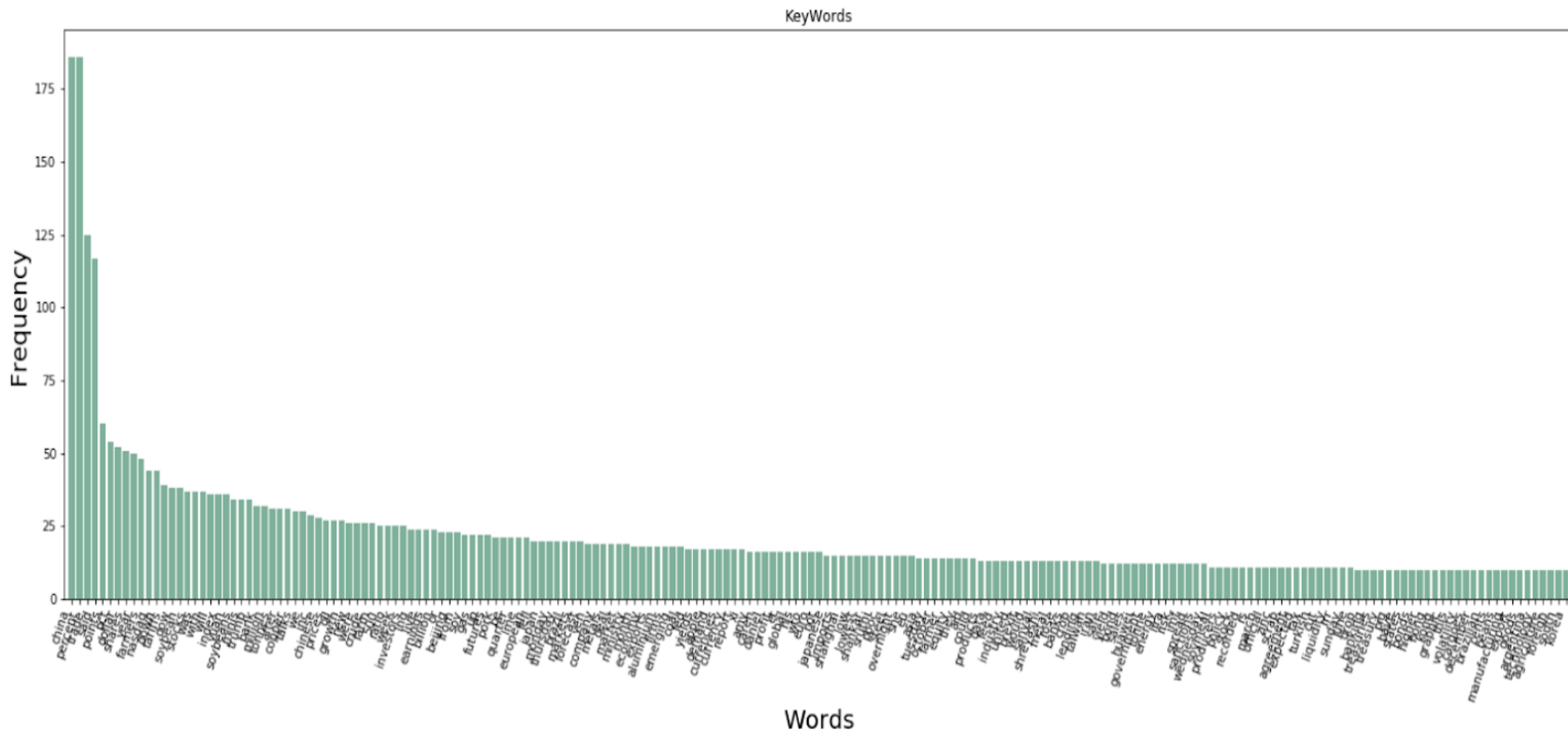
- 文字預處理：TF-IDF
- 關聯性分析：共現圖
- 情感分析 (nltk)、kmeans (關鍵詞分析)、手動分析



2018年6~7月以及2019年5月
為事件發生最頻繁的兩段時間

2018年6~7月：中美雙方公布
對彼此的關稅清單

2019年5月：美國先行對中國加
徵關稅，中國亦採取反制關稅，
議題並延燒至華為事件



詞頻最高的前幾名：

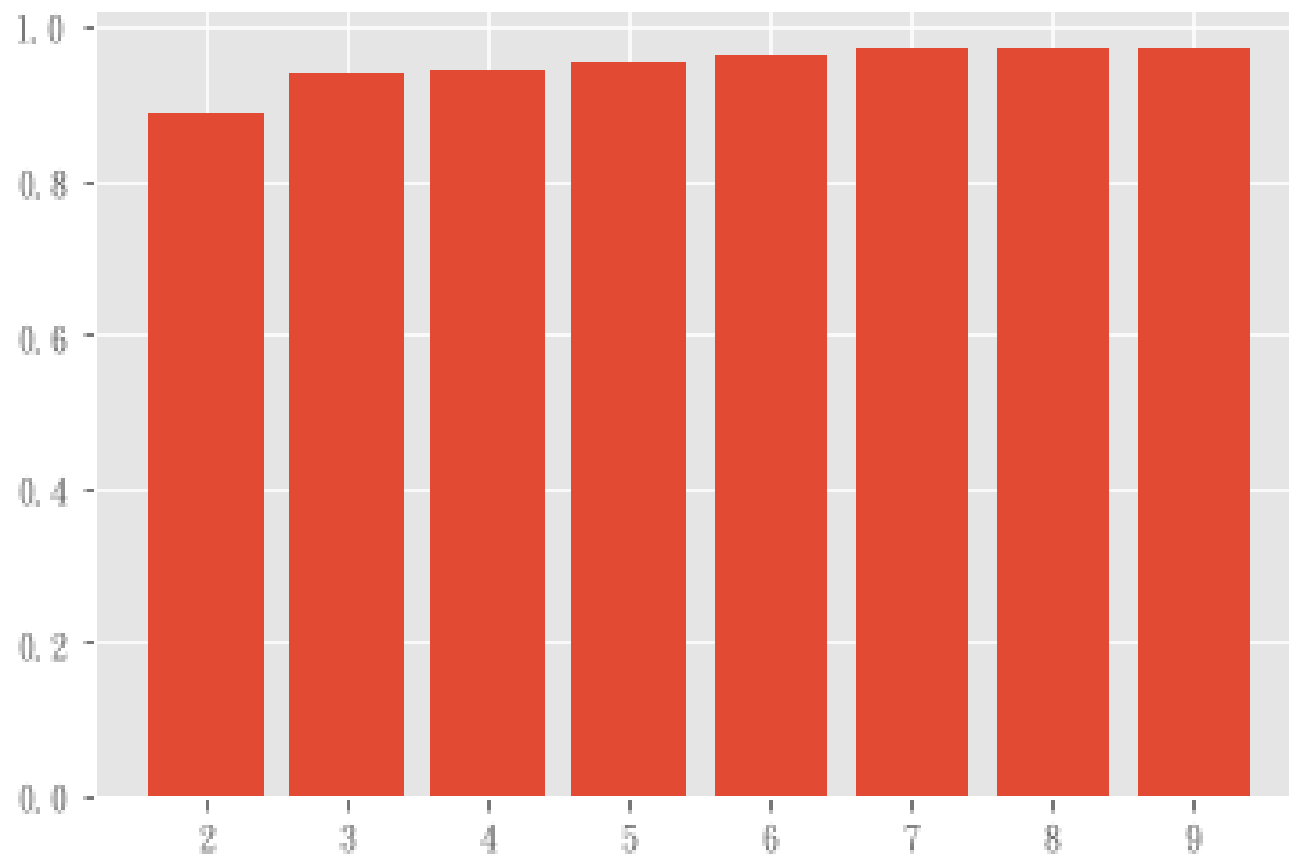
china, trade, dollar, yuan, shares, nasdaq, dow...

➤ 經過篩選得出的關鍵詞，可以初步看出中美貿易的影響：

1. 'shares' , 'nasdaq' , 'down' , 'dow' , 'stocks' ,
'investors'...：股市通常是最迅速反應消息的地方，可想而知，中美貿易對投資人必是操作的重要考量。
2. 'soybean' , 'copper' , 'gas' , 'oil' , 'coal' , 'aluminium' ,
'agriculture' , 'technology' , 'huawei'：根據中美雙方貿易的往來，可以發現農業、能源類產業以及高科技產業等常常是新聞的討論焦點。

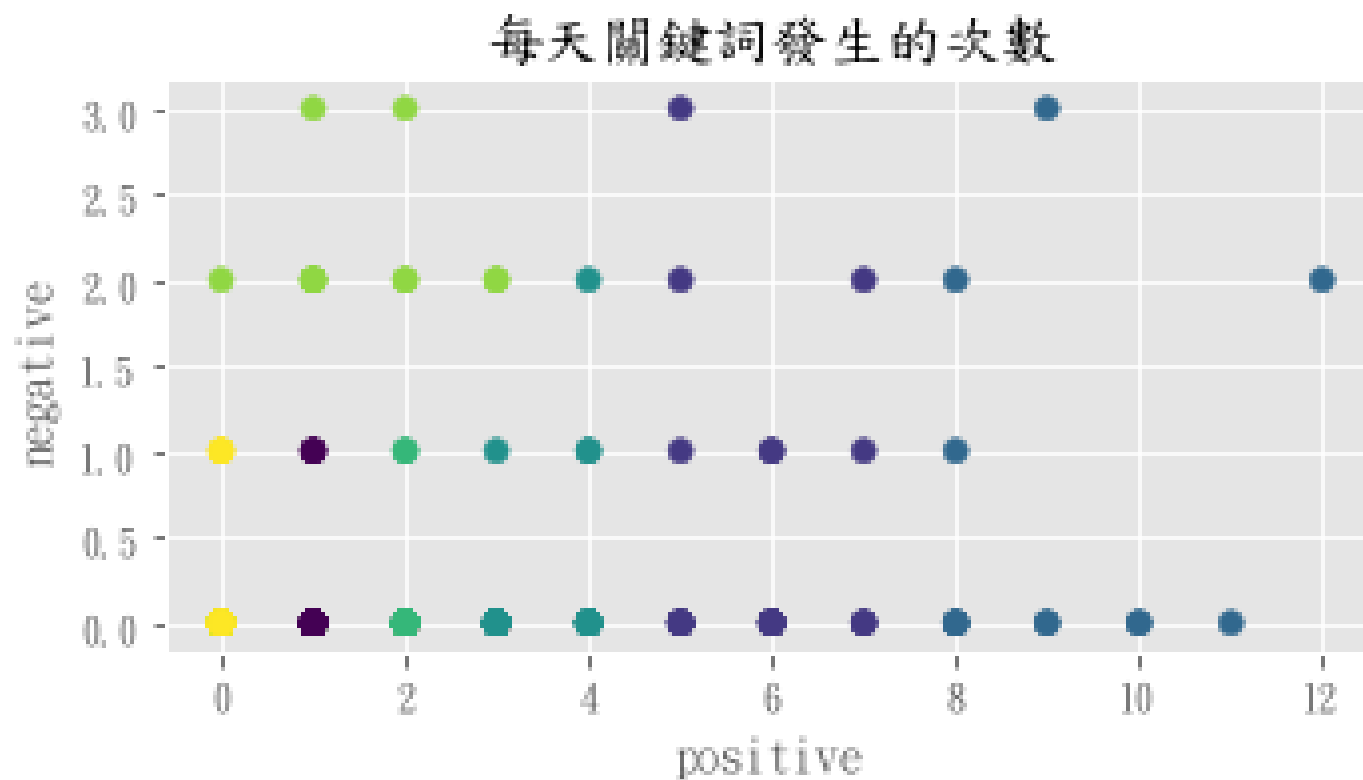
➤ 共現圖呈現關鍵詞群聚的現象，能看出某些重點議題：

1. {'aluminium', 'LME', 'copper', 'metals'}：中美貿易戰使得倫敦金屬交易所(LME)基本金屬期貨下跌
2. {'LNG', 'gas'}：中國提高液化天然氣(LNG)關稅，並減緩對其採購
3. {'soy', 'soybeans'}：中國是美國大豆最大進口國，因此貿易戰對美國大豆得影響甚鉅
4. {'nasdaq', 'down'}, {'dow', '500'}：可見貿易戰對美國股市造成劇烈震盪
5. {'dollar', 'yuan', 'bank'}：貿易戰和美元、人民幣匯率息息相關



K means分群分析

最佳分群數



K means分群分析

分七群

正負面關鍵詞清單內容過少

1

經濟日報用詞相當中性

2

nlTK嚴苛，正負面情緒詞少

3

資料來源過於單調

加資料、判斷句子情緒

/ 研究步驟



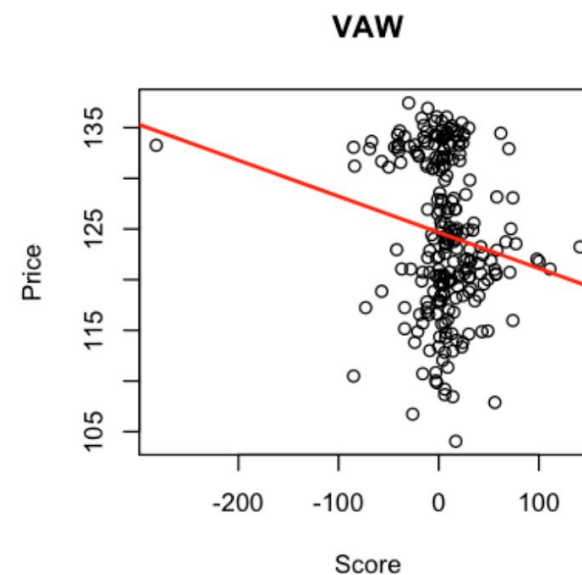
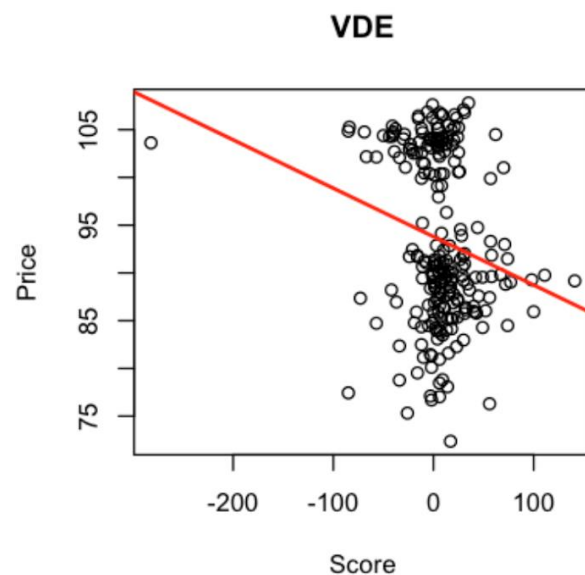
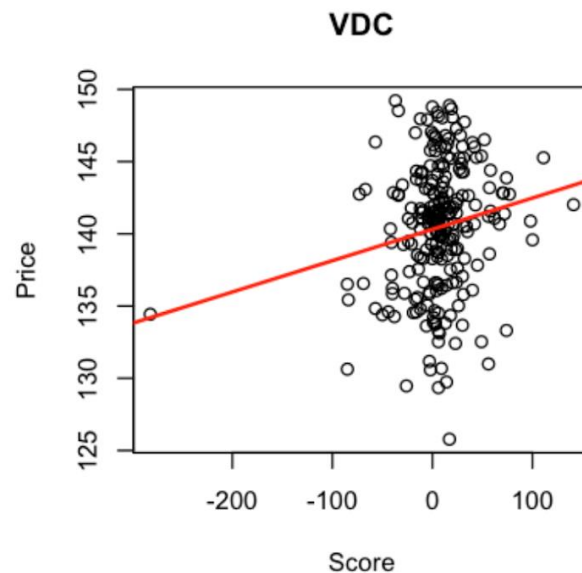
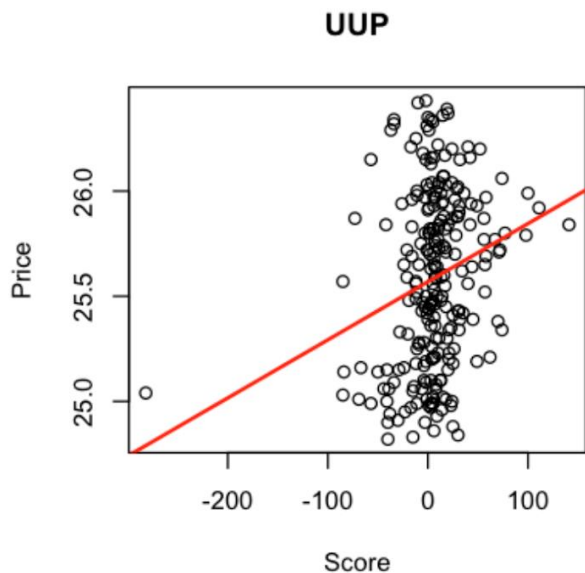
三、關聯性顯著程度： → 統計迴歸分析

- 市場漲跌幅和正負面新聞發生次數的關係

	A	B	C	D	E	F	
1		Date	Price	Change	Score	Square	
2	0	2018/6/15	24.9699993	0.01000023	-17	289	
3	1	2018/6/18	24.9699993	-0.0300007	3	9	
4	2	2018/6/19	25.0400009	-0.0299988	-282	79524	
5	3	2018/6/20	25.0599995	0.01999855	-44	1936	
6	4	2018/6/21	24.9899998	-0.0500011	-57	3249	
7	5	2018/6/22	24.8999996	-0.0200005	-40	1600	
8	6	2018/6/25	24.8600006	-0.0199986	6	36	
9	7	2018/6/26	24.9899998	0.06999969	4	16	
10	8	2018/6/27	25.1399994	0.09999847	-84	7056	
11	9	2018/6/28	25.1399994	-0.0100002	-50	2500	
12	10	2018/6/29	24.9400005	-0.0799999	-38	1444	
13	11	2018/7/2	25.0599995	-0.0500011	-6	36	
14	12	2018/7/3	24.9799995	-0.0200005	1	1	
15	13	2018/7/5	24.8999996	-0.0200005	-3	9	
16	14	2018/7/6	24.8199997	0.01000023	-40	1600	
17	15	2018/7/9	24.8299999	0.05999947	-15	225	
18	16	2018/7/10	24.8400002	-0.0599995	30	900	
19	17	2018/7/11	25.0100002	0.17000008	-69	4761	
20	18	2018/7/12	25.0300007	0	-85	7225	
21	19	2018/7/13	25	-0.0799999	-41	1681	
22	20	2018/7/16	24.9599991	0.00999832	14	196	
23	21	2018/7/17	25.0699997	0.06999969	-14	196	
24	22	2018/7/18	25.1000004	-0.0499997	3	9	

比對新聞與ETF開盤之共同日期，擷取日期、收盤價、股價波動幅度、情緒分數及其平方，進行分析

資產類別	Ticker	結果
Equity	SPY	統計上不顯著
Bond	AGG	統計上不顯著
Currency	UUP	統計上顯著，但影響非常小
Real Estate	VNQ	統計上不顯著
Commodity	GLD	統計上不顯著
產業類別 (Sector)		
Consumer Staples	VDC	統計上顯著，但影響非常小
Energy	VDE	統計上顯著，但影響非常小
Financials	VFH	統計上不顯著
Industrials	VIS	統計上不顯著
Materials	VAW	統計上顯著，但影響非常小
Technology	VGT	統計上不顯著
Telecom	VOX	統計上不顯著

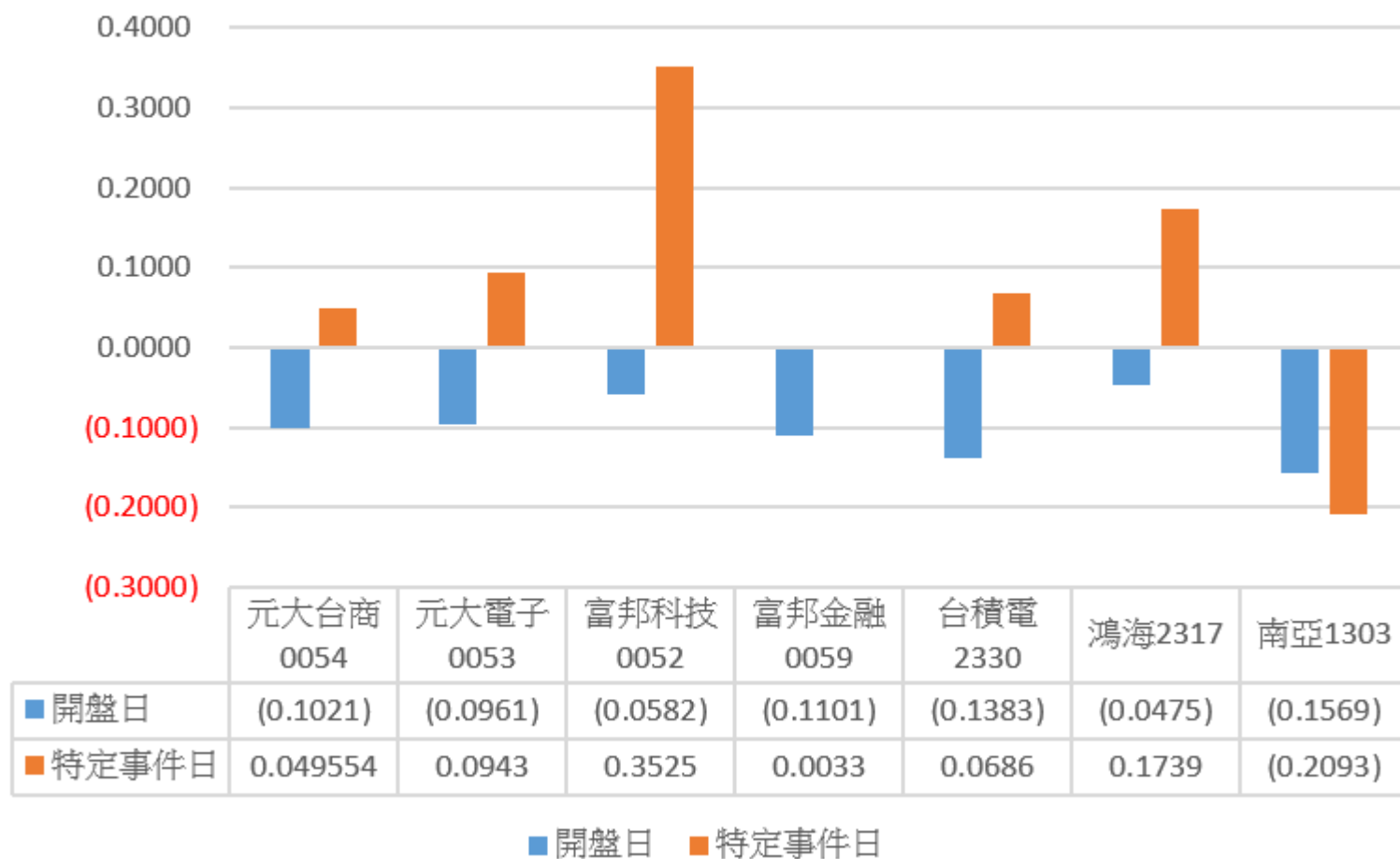


➤ 匯率、民生必需品、
能源、原物料類型的ETF
在統計上有顯著性，
但從圖可以發現，實際上
情緒分數的變動不太能解
釋價格的波動

元大台商0054	元大電子0053	富邦科技0052
(0.1021)	(0.0961)	(0.0582)
0.049554	0.0943	0.3525

富邦金融 0059	台積電 2330	鴻海 2317	南亞 1303
(0.1101)	(0.1383)	(0.0475)	(0.1569)
0.0033	0.0686	0.1739	(0.2093)

相關係數差異



/結論

迴歸分析出來的結果絕大多數在統計上都不顯著，推測以下原因：

1. 光靠新聞發生的情緒分數無法看出對價格的影響
2. 情緒分數計算問題
 - a. 斷詞：使用單字無法準確表達出正負面
 - b. 斷句：新聞的句子多半較長，偏中性
 - c. 模型配分標準不夠透明



Thank you for listening~

<https://github.com/mengjeelee/datasc/tree/master/%E5%B0%88%E6%A1%88>