

# 金融資料爬蟲與資料庫 視覺化數據分析

洪祥發

# 目錄

1

專案介紹

2

數據與核心技術

3

專案數據流

4

專案展示

5

結論與未來發展

# 專案介紹

## 背景

### 1.科技驅動數據決策：

隨著大數據、AI、機器學習等技術進步，金融機構和投資者越來越重視數據分析來提高投資效益與風險管理。

### 2.資訊抓取與視覺化的重要性：

快速提取關鍵數據並以視覺化方式呈現，成為各行業實現精準決策的基礎。

## 目標：

1.利用爬蟲技術獲取股票相關資料，建立結構化的資料庫

2.進行視覺化與數據分析，根據結果撰寫洞察報告。

## 動機：

### 1.提高資料獲取效率：

利用爬蟲自動抓取金融數據，降低人力與時間成本，確保數據時效與完整性。

### 2.深入挖掘數據價值：

結構化整理龐大數據，發掘市場規律，支援投資策略與風險評估。

### 3.促進數據驅動決策：

透過資料庫與視覺化工具展示數據，幫助決策者直觀掌握市場動態。

# 數據與核心技術

數據來源:

goodinfo台灣股市資訊網、yahoofinance

數據期間:

2014~2024，共十年

數據頻率:

日資料

input資料:

- 股價（開盤價、最高價、最低價、收盤價）
- 基本面（EPS、ROE）
- 技術面（KD、RSI）
- 籌碼面（三大法人買賣超）

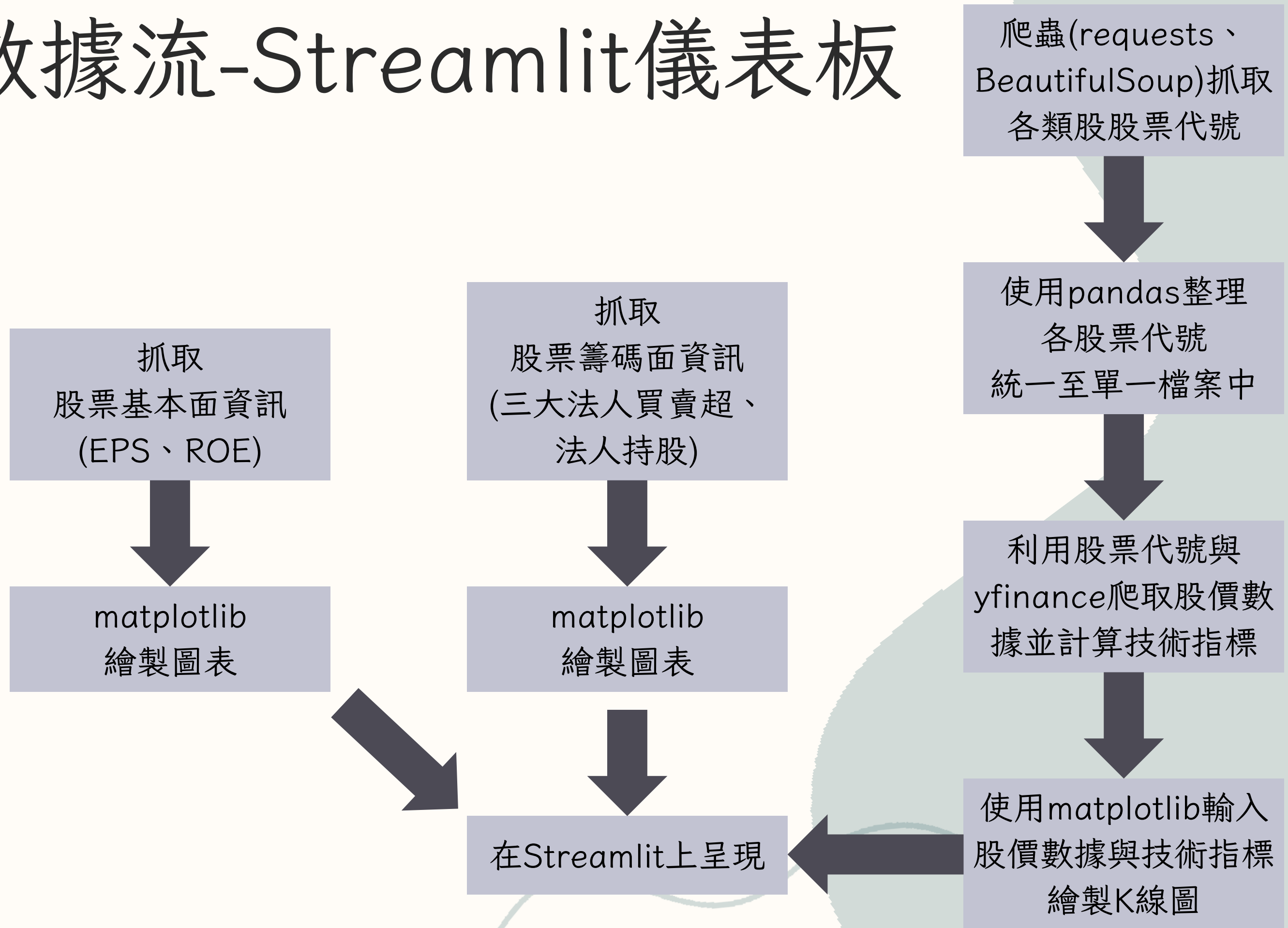
核心技術:

1.網路爬蟲:requests、BeautifulSoup、yfinance

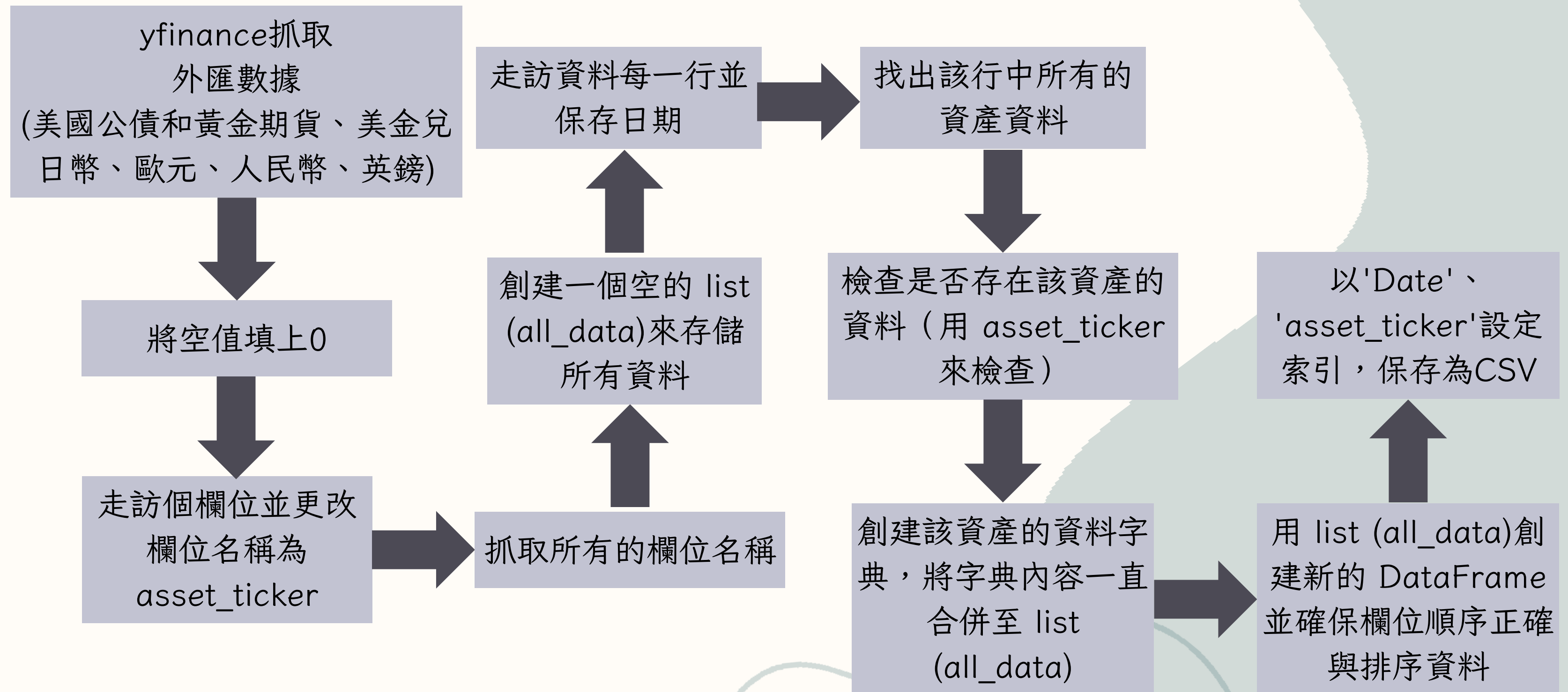
2.資料整理:pandas

3.視覺化呈現:matplotlib、Streamlit

# 專案數據流-Streamlit儀表板



# 專案數據流-外匯資料庫整理



# 專案展示-外匯資料庫整理結果

▲	A	B	C	D	E	F	G	H	I	J
1	Date	asset_ticker	Open	High	Low	Close	Adj Close	Volume	Dividends	Stock Splits
2	2014/12/22	Gold	1198	1198	1172.7	1179.7	1179.7	223	0	0
3	2014/12/22	TNX	2.185	2.188	2.155	2.162	2.162	0	0	0
4	2014/12/22	TYX	2.782	2.787	2.745	2.75	2.75	0	0	0
5	2014/12/22	USDCNY	6.2099	6.213	6.2062	6.2099	6.2099	0	0	0
6	2014/12/22	USDEUR	0.818	0.8181	0.8143	0.8179	0.8179	0	0	0
7	2014/12/22	USDGBP	0.6398	0.6409	0.6383	0.6398	0.6398	0	0	0
8	2014/12/22	USDJPY	119.495	119.992	119.318	119.497	119.497	0	0	0
9	2014/12/23	Gold	1179.3	1180.3	1175	1177.9	1177.9	232	0	0
10	2014/12/23	TNX	2.172	2.261	2.172	2.257	2.257	0	0	0
11	2014/12/23	TYX	2.755	2.854	2.755	2.852	2.852	0	0	0
12	2014/12/23	USDCNY	6.2116	6.217	6.2116	6.2116	6.2116	0	0	0
13	2014/12/23	USDEUR	0.818	0.822	0.8163	0.8181	0.8181	0	0	0
14	2014/12/23	USDGBP	0.6418	0.6456	0.6409	0.6419	0.6419	0	0	0
15	2014/12/23	USDJPY	120.142	120.788	120.044	120.15	120.15	0	0	0
16	2014/12/24	Gold	1177	1179.3	1173.5	1173.5	1173.5	23	0	0
17	2014/12/24	TNX	2.284	2.301	2.261	2.264	2.264	0	0	0
18	2014/12/24	TYX	2.873	2.882	2.831	2.834	2.834	0	0	0
19	2014/12/24	USDCNY	6.2159	6.2211	6.2053	6.2159	6.2159	0	0	0
20	2014/12/24	USDEUR	0.8213	0.8216	0.8183	0.8214	0.8214	0	0	0
21	2014/12/24	USDGBP	0.6443	0.6449	0.6428	0.6442	0.6442	0	0	0
22	2014/12/24	USDJPY	120.676	120.699	120.277	120.69	120.69	0	0	0
23	2014/12/25	Gold	0	0	0	0	0	0	0	0
24	2014/12/25	TNX	0	0	0	0	0	0	0	0
25	2014/12/25	TYX	0	0	0	0	0	0	0	0
26	2014/12/25	USDCNY	6.2053	6.2053	6.2053	6.2053	6.2053	0	0	0
27	2014/12/25	USDEUR	0.8198	0.8202	0.8159	0.8198	0.8198	0	0	0
28	2014/12/25	USDGBP	0.6426	0.6433	0.6417	0.6426	0.6426	0	0	0
29	2014/12/25	USDJPY	120.399	120.459	119.999	120.459	120.459	0	0	0



# 專案展示-資料蒐集與處理

goodinfo台灣股市資訊網爬蟲結果:

代號	名稱	成交	財報 季度	市值 (億)	現金 (億)	收 帳款 (億)	存貨 (億)
1101	台泥	33.3	24Q3	2515	744	329	211
1102	亞泥	42.9	24Q3	1521	152	139	71.1
1103	嘉泥	17.9	24Q3	141	19.7	3.42	1.33
1104	環泥	29.2	24Q3	201	13.2	17.8	4.04
1108	幸福	14.5	24Q3	58.7	3.85	10.2	40.7
1109	信大	18.4	24Q3	62.8	30.3	7.45	10.1
1110	東泥	21.15	24Q3	121	4.51	7.4	12.5

代號	名稱	成交	財報 季度	市值 (億)	現金 (億)	收 帳款 (億)	存貨 (億)	動 資產
1201	味全	18.25	24Q3	92.4	13.1	30.4	17.8	45.2
1203	味王	42.25	24Q3	101	17.9	5.31	16.9	26
1210	大成	53.7	24Q3	480	77.6	77.1	88.1	161
1213	大飲	12.1	24Q3	6.85	0.18	0.61	0.32	0.79
1215	卜蜂	99.6	24Q3	294	4.74	29.2	22.3	39.9
1216	統一	85.6	24Q3	4864	1022	332	546	1739
1217	愛之味	11.85	24Q3	58.6	5.52	14.8	7.53	23.4
1218	泰山	21	24Q3	105	23.6	16.6	11.3	44.1
1219	福壽	16.75	24Q3	57.2	8.87	15.7	25.8	25.2
1220	台榮	17.2	24Q3	30.5	5.05	6.88	6.08	12.6
1225	福懋油	79.4	24Q3	182	14.1	17	18	31.1
1227	佳格	37.5	24Q3	343	33.1	46.8	54	120
1229	聯華	58	24Q3	992	50.5	22.8	15.5	146
1231	聯華食	119.5	24Q3	324	4	25.1	18.7	34.2



# 專案展示-資料蒐集與處理

goodinfo台灣股市資訊網個股

基本面抓取結果:

年度	股本 (億)	財報 評分	年度股價(元)				獲利金額(億)					ROE (%)	ROA (%)	EPS(元)		BPS (元)
			收盤	平均	漲跌	漲跌 (%)	營業 收入	營業 毛利	營業 利益	業外 損益	稅後 淨利			稅後 EPS	年增 (元)	
24Q3	2,593	89	1090	888	497	83.8	20,258	11,120	8,963	607	7,986	28.4(年 估)	18.2(年 估)	30.8	7.67	153.8
2023	2,593	89	593	543	144.5	32.2	21,617	11,751	9,215	577	8,385	26	16	32.34	-6.86	133.3
2022	2,593	89	448.5	516	-166.5	-27.1	22,639	13,484	11,213	229	10,165	39.6	23.4	39.2	16.19	113.0
2021	2,593	87	615	598	85	16	15,874	8,195	6,500	131	5,965	29.7	18.4	23.01	3.04	83.6
2020	2,593	89	530	379	199	60.1	13,393	7,111	5,668	180	5,179	29.8	20.6	19.97	6.65	71.3
2019	2,593	89	331	262	105.5	46.8	10,700	4,927	3,727	171	3,453	20.9	15.9	13.32	-0.22	62.5
2018	2,593	94	225.5	237	-4	-1.7	10,315	4,979	3,836	139	3,511	21.9	17.2	13.54	0.31	64.6
2017	2,593	91	229.5	210	48	26.4	9,774	4,948	3,856	106	3,431	23.6	17.7	13.23	0.34	58.7

# 專案展示-資料蒐集與處理

goodinfo台灣股市資訊網個股

籌碼面抓取結果:

期別	成交	漲跌	漲跌 (%)	成交量 (張)	外資					投信			自營商		
					買進 (張)	賣出 (張)	買賣超 (張)	持有 (張)	持股 比率	買進 (張)	賣出 (張)	買賣超 (張)	買進 (張)	賣出 (張)	買賣超 (張)
24/12/30	1090	0	0	27,437	18,143	15,534	2,610	#####	73.6	112	143	-30.8	283	516	-233
24/12/27	1090	5	0.46	26,152	13,304	6,529	6,775	#####	73.6	61	146	-84.7	346	251	94.6
24/12/26	1085	0	0	15,018	7,618	3,425	4,192	#####	73.6	187	140	47.5	156	72.9	83.1
24/12/25	1085	5	0.46	16,186	6,527	2,729	3,798	#####	73.6	127	76.5	50.5	1,510	209	1,301
24/12/24	1080	0	0	30,070	20,289	14,176	6,113	#####	73.6	286	298	-12.4	793	261	532
24/12/23	1080	45	4.35	32,800	24,870	7,169	17,701	#####	73.6	619	66.2	553	1,295	433	862
24/12/20	1035	-35	-3.27	56,392	23,385	45,590	-22,205	#####	73.5	780	777	3.41	823	323	499
24/12/19	1070	-15	-1.38	46,127	22,712	28,908	-6,196	#####	73.6	520	95.3	425	647	331	315
24/12/18	1085	10	0.93	47,331	35,131	22,568	12,564	#####	73.6	369	559	-190	1,026	1,934	-908
24/12/17	1075	-10	-0.92	40,034	25,568	22,818	2,750	#####	73.6	227	400	-173	611	659	-48
24/12/16	1085	20	1.88	43,680	33,095	14,328	18,767	#####	73.6	9	118	-109	278	1,197	-919
24/12/13	1065	5	0.47	24,330	13,432	14,327	-896	#####	73.5	4	183	-179	134	997	-863
24/12/12	1060	20	1.92	26,574	16,780	17,415	-635	#####	73.5	22	180	-158	171	264	-93.3
24/12/11	1045	-20	-1.88	26,949	11,077	18,354	-7,277	#####	73.5	528	102	427	562	970	-408

# 專案展示-資料蒐集與處理

goodinfo爬蟲結果處理展示:  
使用pandas將各類股中的股票與  
資訊統一至單一檔案中。

	代號	名稱	成交	財報 季度	市值 (億)	現金 (億)	收 帳款 (億)	存貨 (億)	動 資產 (億)	財報 評分	ck_category
0	2816	旺旺保	32.1	24Q3	71.8	42.5	9.52	0	0	67	保險業
1	2832	台產	28.95	24Q3	105	34.2	9.5	0	0	67	保險業
2	2850	新產	99.8	24Q3	315	125	25.4	0	0	67	保險業
3	2851	中再保	25.95	24Q3	208	93.3	12.3	0	0	61	保險業
4	2852	第一保	25.1	24Q3	75.6	18.4	4.87	0	0	58	保險業
5	2867	三商壽	6.55	24Q3	354	469	122	0	0	30	保險業
6	5878	台名	46.8	24Q3	11.7	0.39	0.77	0	3.65	45	保險業
7	2323	中環	11.4	24Q3	124	23.9	17.8	30.1	132	43	光電業
8	2349	鍊德	15	24Q3	104	31.7	12.2	15.3	47.2	47	光電業
9	2374	佳能	46.2	24Q3	134	23.1	10.4	9.09	34.5	55	光電業
10	2393	億光	79	24Q3	350	55.8	61.2	10.8	123	64	光電業
11	2406	國碩	16.7	24Q3	58.6	21.2	31.8	13.2	53.6	28	光電業
12	2409	友達	15.65	24Q3	1200	633	313	335	955	38	光電業
13	2426	鼎元	24.65	24Q3	74.1	10.7	8.85	6.61	21.7	51	光電業
14	2429	銘旺科	29.35	24Q3	25.1	7.2	0.21	0.45	7.45	43	光電業
15	2438	翔耀	39.2	24Q3	25.6	0.94	2.12	0.5	3.05	28	光電業
16	2466	冠西電	33.95	24Q3	58.2	1.46	4.11	5.35	5.57	23	光電業
17	2486	一詮	110.5	24Q3	259	13.8	23.6	13.1	39.6	43	光電業
18	2489	瑞軒	19.5	24Q3	133	45.1	58.6	31.9	143	43	光電業



# 專案展示-資料蒐集與處理

yfinance爬蟲結果處理展示:

爬取數據後將所有個股數據與  
技術指標統整至同一CSV檔中。

Stock_Code	date	open	high	low	close	volume	dividends	stock splits	RSI	ceil	floor	K	D
2349	2024/12/9	15.05	15.1	14.6	14.8	8917440	0	0	36.928979	70	40	13.72549	21.143791
2349	2024/12/10	14.75	15	14.7	14.75	3479043	0	0	36.014838	70	40	10.088552	16.434746
2349	2024/12/11	14.75	14.9	14.55	14.7	4321284	0	0	35.003935	70	40	10.272375	11.362139
2349	2024/12/12	14.75	14.95	14.5	14.5	3996765	0	0	30.949847	70	40	6.3508065	8.903911
2349	2024/12/13	14.45	14.5	14.25	14.25	3519907	0	0	26.477723	70	40	3.125	6.5827272
2374	2014/12/16	19.1	20.1	19.1	19.92	1169173	0	0		70	40		
2374	2014/12/17	19.92	20.16	19.51	19.51	682068	0	0		70	40		
2374	2014/12/18	19.75	20.04	19.75	19.86	701864	0	0		70	40		
2374	2014/12/19	20.27	20.27	19.75	19.92	643675	0	0		70	40		
2374	2014/12/22	19.92	20.1	19.75	19.8	352731	0	0		70	40		
2374	2014/12/23	19.8	19.92	19.69	19.75	227955	0	0		70	40		
2374	2014/12/24	19.75	19.92	19.69	19.75	410320	0	0		70	40		
2374	2014/12/25	19.75	19.98	19.75	19.98	354531	0	0	52.459016	70	40		
2374	2014/12/26	19.98	20.1	19.86	19.98	171566	0	0	52.459016	70	40		
2374	2014/12/29	20.27	20.33	19.92	19.92	345533	0	0	49.167734	70	40		
2374	2014/12/30	20.16	20.16	19.63	19.63	408521	0	0	36.318735	70	40		
2374	2014/12/31	19.63	19.69	19.51	19.51	281945	0	0	32.250087	70	40		
2374	2015/1/5	19.57	19.75	19.45	19.75	269947	0	0	46.289663	70	40	11.363636	23.256065
2374	2015/1/6	19.69	20.04	19.57	19.86	425917	0	0	51.647533	70	40	26.893939	18.308081

# 專案展示-視覺化

Streamlit 視覺化儀表板：  
依照輸入的股票代號、圖表類型與  
日期區間顯示圖表。

## 專案實作

請輸入股票代號

選擇圖表類型

K線圖

K線圖

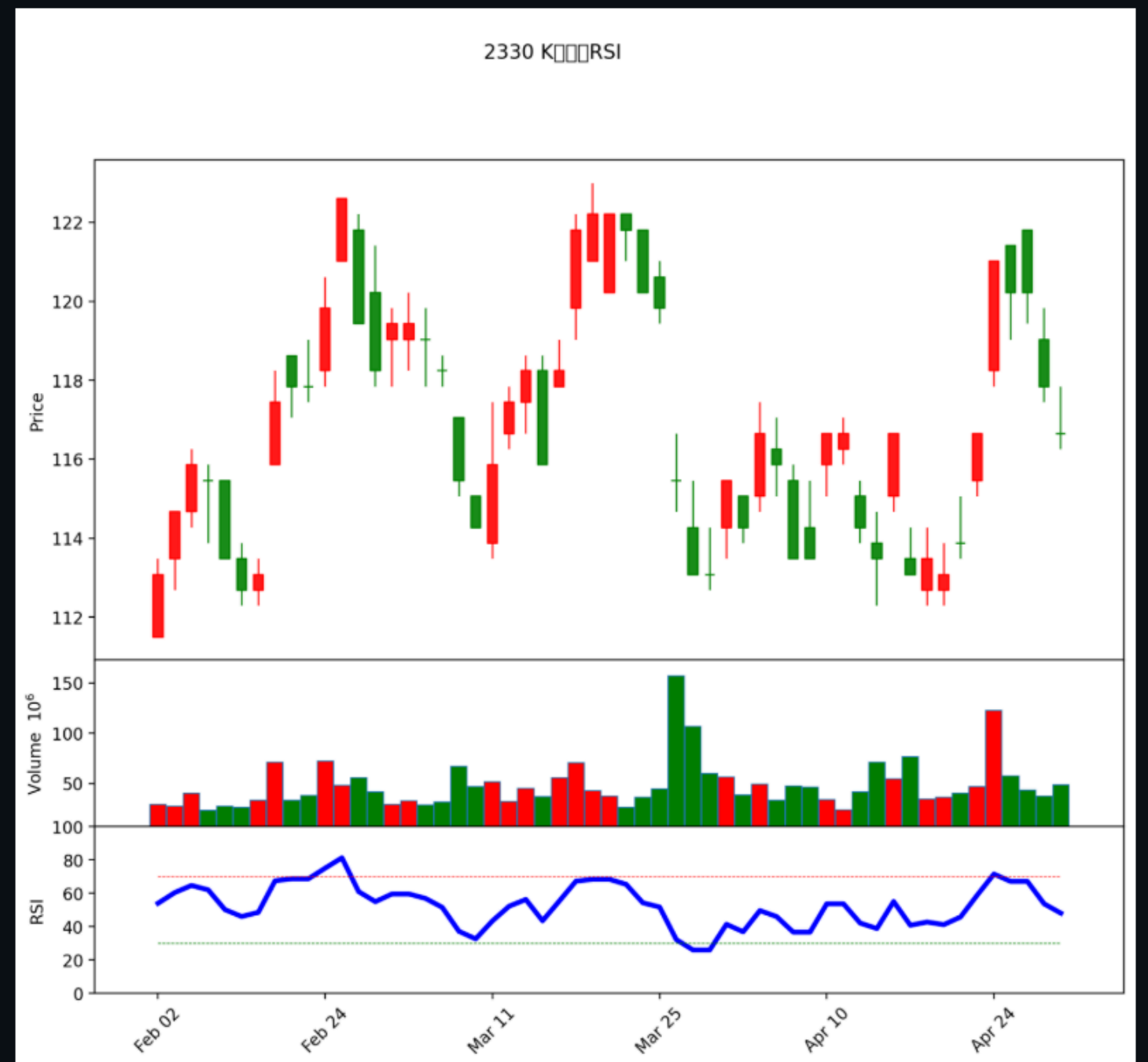
走勢圖

選擇顯示日期區間

2015-01-01 - 2015-01-31

顯示圖表

Deploy



# 專案展示-視覺化

Streamlit 視覺化儀表板：  
也可選擇基本面與籌碼面的圖表類型。

## 基本面分析

選擇基本面圖表類型

年度ROE條形圖

顯示基本面圖表

## 籌碼面分析

選擇籌碼面圖表類型

三大法人買賣超金額分布圖

顯示籌碼面圖表

# 專案展示-視覺化

Streamlit 視覺化儀表板：  
基本面與籌碼面的圖表展示。

## 基本面分析

選擇基本面圖表類型

EPS成長趨勢圖

顯示基本面圖表



## 籌碼面分析

選擇籌碼面圖表類型

法人持股變化

顯示籌碼面圖表





# 專案展示-數據分析

以台積電(2330)為例，近一個月來看，5MA都在20MA之上呈現多頭趨勢，即便股價跌破20MA，隔日也能站回5MA，具有上漲的支撐信號。

基本面方面，因資料來源只有年頻率的數據，並無2024年數據，所以只能看出台積電在2020後基本面比前幾年都好，無法確定近期基本面趨勢是否穩定，也無法從基本面看出是否能夠支持股價的成長。

而籌碼面，從法人籌碼流向來看外資近期都是買超的，法人持股也呈現上升趨勢，顯示資金正在進場。

最後個人認為短期內股價具備上漲信號，且法人籌碼流向顯示資金流入，對股價有正向支撐。然而，由於缺乏2024年基本面數據，近期基本面是否穩定還需要進一步觀察。



# 結論與未來發展

## 結論

在本專案中，成功構建了一個從數據獲取、資料庫整合到數據視覺化呈現的金融數據Web應用程式。透過網路爬蟲技術，自動抓取並清洗大量金融數據，設計出便於查詢的資料庫架構，再利用視覺化工具將複雜數據轉換成直觀圖表，有效輔助使用者快速分析數據並做出決策。

在實作過程中，我學習到了如何解決網頁表格原始碼難以定位的問題、如何優化資料庫以及如何從零開始開發Web應用程式，這不僅提升了個人的技術能力，也讓我深入理解數據從採集、處理到展示的全流程。即便在資料即時更新與數據質量控制等方面仍有改進空間，但這些挑戰促使我累積了寶貴的經驗，為未來奠定了堅實基礎。

## 未來發展

### 1. 程式碼優化與資料庫擴展：

- 程式碼：熟悉模組的運用方式，學習更多在數據分析時會使用到的工具。
- 資料抓取與處理：精進爬蟲技術及數據清洗的辦法，進一步提升數據抓取和預處理速度。
- 擴展資料來源：擴大數據來源範圍，包含更多金融平台、新聞網站，來豐富資料庫內容。

### 2. 使用者體驗與互動介面：

- 視覺化Web優化：持續改進數據視覺化介面，使數據展示更直觀、互動性更佳。
- 即時動態更新：加入即時資料更新機制，保證使用者能在第一時間獲得最新市場資訊。