

DS Final Project Proposal

依據天氣影響之下周菜價預測平台

Sorry my bad

- 曾柏誠 E24126717
- 何寬羿 C34104032
- 林業誠 E24105038

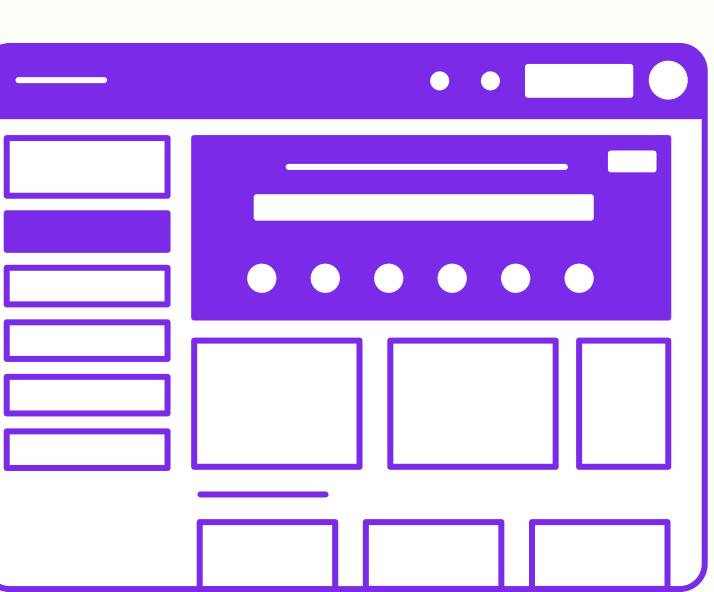
1. Introduction

Problems

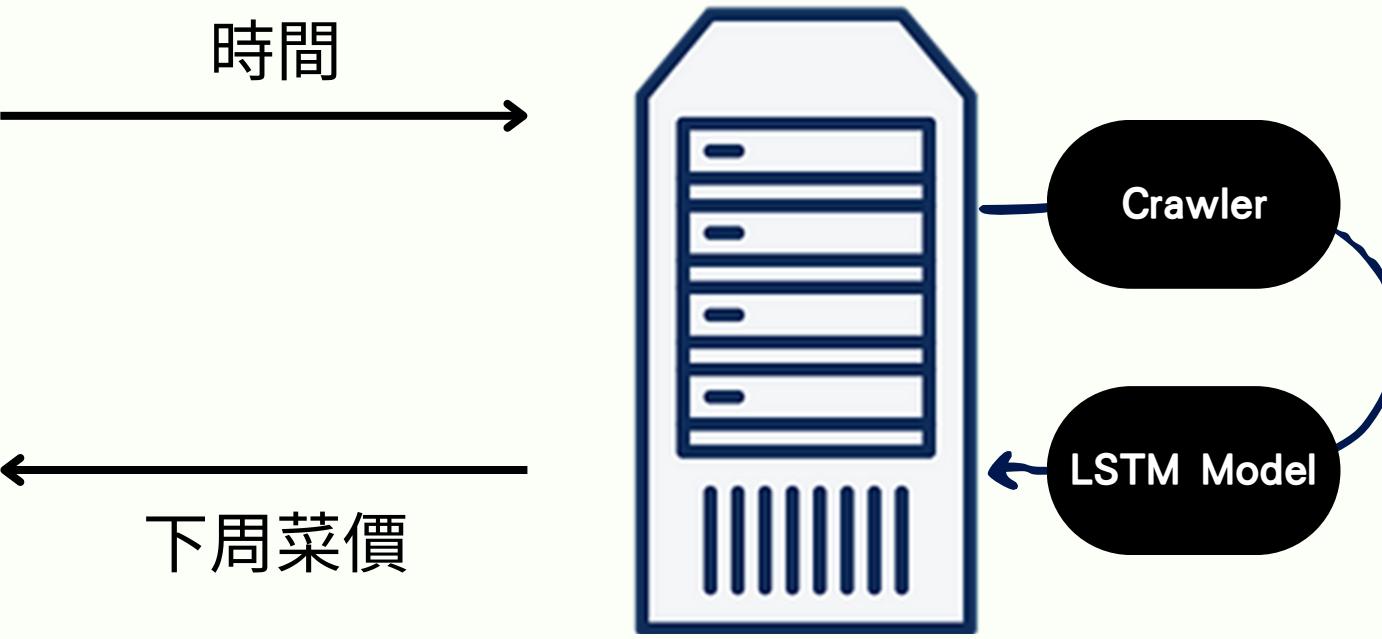
- 菜價會受天氣影響
- 缺乏預測菜價的平台

Goals - 菜價預測平台

- 易使用
- 誤差20%以內



GUI
(前端)



Methods
(後端)

Training

- Model : LSTM
- Input : 前一周蘿蔔菜價、全台平均
降水
- Output : 下一周蘿蔔菜價
- Dataset : Training/Validation/Test
--7:2:1 (2016~2024)
- Evaluation Metrics : RMSE

2. Data Description and Preprocessing

--資料來源網站

1. 天氣資料 (交通部中央氣象署-每日雨量)

- 特點：氣象資料豐富，更新頻率高、即時。
- 應用：捕捉天氣對農作物生長和供應的影響。

2. 菜價資料 (農業部-田邊好幫手)

- 特點：每日更新不同地區的蔬菜價格
- 應用：結合歷史菜價與天氣特徵進行價格預測

3. 颱風資料 (颱風資料庫)

- 特點：導致菜價急劇波動的因素之一
- 應用：增強模型預測的準確性

2. Data Description and Preprocessing

--爬取方法：Selenium爬蟲程式

```
from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.common.keys import Keys
from selenium.webdriver.support.ui import WebDriverWait,Select
from selenium.webdriver.support import expected_conditions as EC
from selenium.common.exceptions import TimeoutException
```

```
# 打開目標網站
url = "https://rdc28.cwa.gov.tw/TDB/public/typhoon_list/"
driver.get(url)

# 等待表格加載完成
time.sleep(5)

# 找到表格元素
try:
    table = driver.find_element(By.ID, 'typhoon_list_table')
    rows = table.find_elements(By.TAG_NAME, 'tr')

    # 提取表格內容
    data = []
    for i, row in enumerate(rows):
        cells = row.find_elements(By.TAG_NAME, 'th' if i == 0 else 'td')
        row_data = []
        for cell in cells:
            text = cell.text.strip()
            # 處理換行符號：颱風生命期間用 ~，颱風名稱移除換行
            if i > 0: # 跳過標題行
                if "颱風生命期間" in row_data: # 假設對應到特定欄位
                    text = text.replace("\n", "~")
                elif "颱風名稱" in row_data:
                    text = text.replace("\n", " ")
            row_data.append(text)
        if row_data:
            data.append(row_data)
except Exception as e:
    print(f"An error occurred: {e}")
```

我們在爬取菜價、天氣、颱風資料時，發現使用傳統的 BeautifulSoup 無法正確爬到我們需要的資料。因此，我們決定使用 Selenium，透過模擬人在操作網頁的方式來爬取訓練及預測時所需的資料。

2. Data Description and Preprocessing

以下為爬取到的歷年颱風資料(僅附上部分)：

	年份	颱風編號	颱風名稱	颱風生命期間	\
0	2024	202426	帕布(PABUK)	2024-12-23 06:00~2024-12-25 00:00	
1	2024	202425	天兔(USAGI)	2024-11-11 18:00~2024-11-16 03:00	
2	2024	202424	萬宜(MAN-YI)	2024-11-09 06:00~2024-11-20 00:00	
3	2024	202423	桔梗(TORAJI)	2024-11-09 06:00~2024-11-14 06:00	
4	2024	202422	銀杏(YINXING)	2024-11-03 18:00~2024-11-12 06:00	

	颱風生命期間中心最低氣壓(hPa)	颱風生命期間中心最大風速(m/s)	颱風生命期間最大7級風暴風半徑(km)	\
0	1000	18	80	
1	945	45	150	
2	900	60	180	
3	965	38	150	
4	930	51	180	

	颱風生命期間最大10級風暴風半徑(km)	警報發布報數
0	---	---
1	50	19
2	90	---
3	50	---
4	90	---

2. Data Description and Preprocessing

填補闕漏值：

The screenshot shows a search interface for agricultural product prices. The search bar contains 'LA1'. The date range is set from '113/12/26' to '113/12/26'. The search results table has columns: 交易日期 (Trading Date), 類別 (Category), 名稱 (Name), 市場 (Market), 上價 (High Price), 中價 (Middle Price), 下價 (Low Price), 平均價 (Average Price), and 交易量 (Trading Volume). A note at the bottom states: '請勿輸入特殊字元（標點符號及空格） 輸入名稱或代號，例如：鳳梨、B4'.

註

- 更新時間為每日14:00。
- “-”表示雨量為0。
- “T”表示雨跡，降水量小於0.5mm。
- “X”表無紀錄值或儀器故障。

• 在農業部田邊好幫手-農產品交易行情中，也能發現有部分天數是查詢不到菜價的，如上圖的2024年12月26日就查無菜價資料。

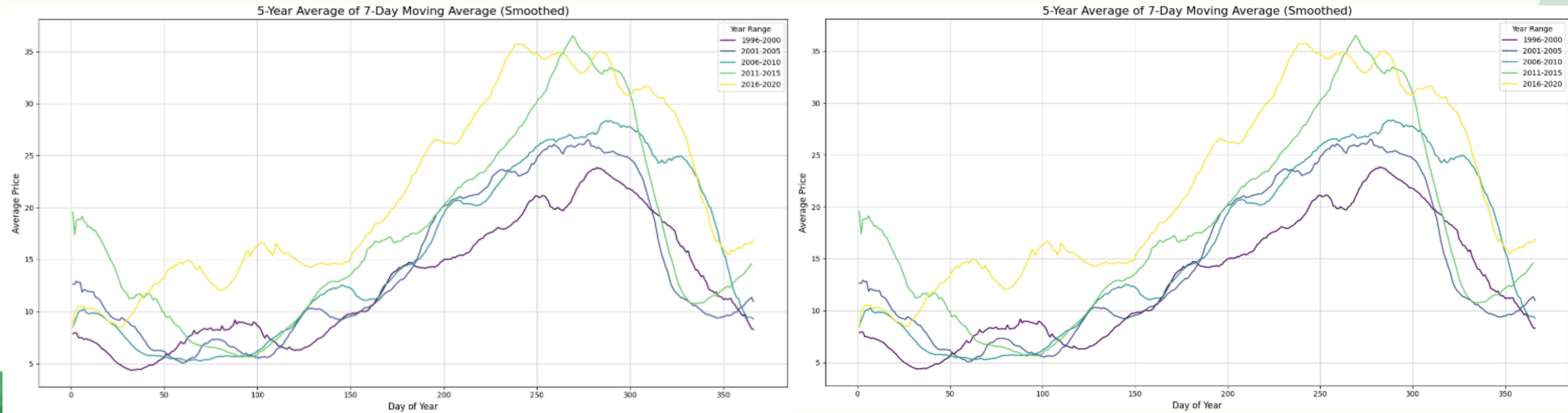
• 中央氣象署-每日雨量資料中有註明：若表格中雨量為“T”代表降水量小於0.5mm，若雨量為“X”代表無紀錄值或儀器故障。

由於以上因素，我們爬下來的資料存在許多闕漏值。我們採用以下方法填補資料缺漏：

- (1) 對於降水量為“T”(小於0.5mm)的數據，我們將其視為降水量0.0mm
- (2) 對於降水量為“X”(無紀錄值或儀器故障)的數據以及無菜價紀錄之日期，我們使用內插法填補。

2. Data Description and Preprocessing

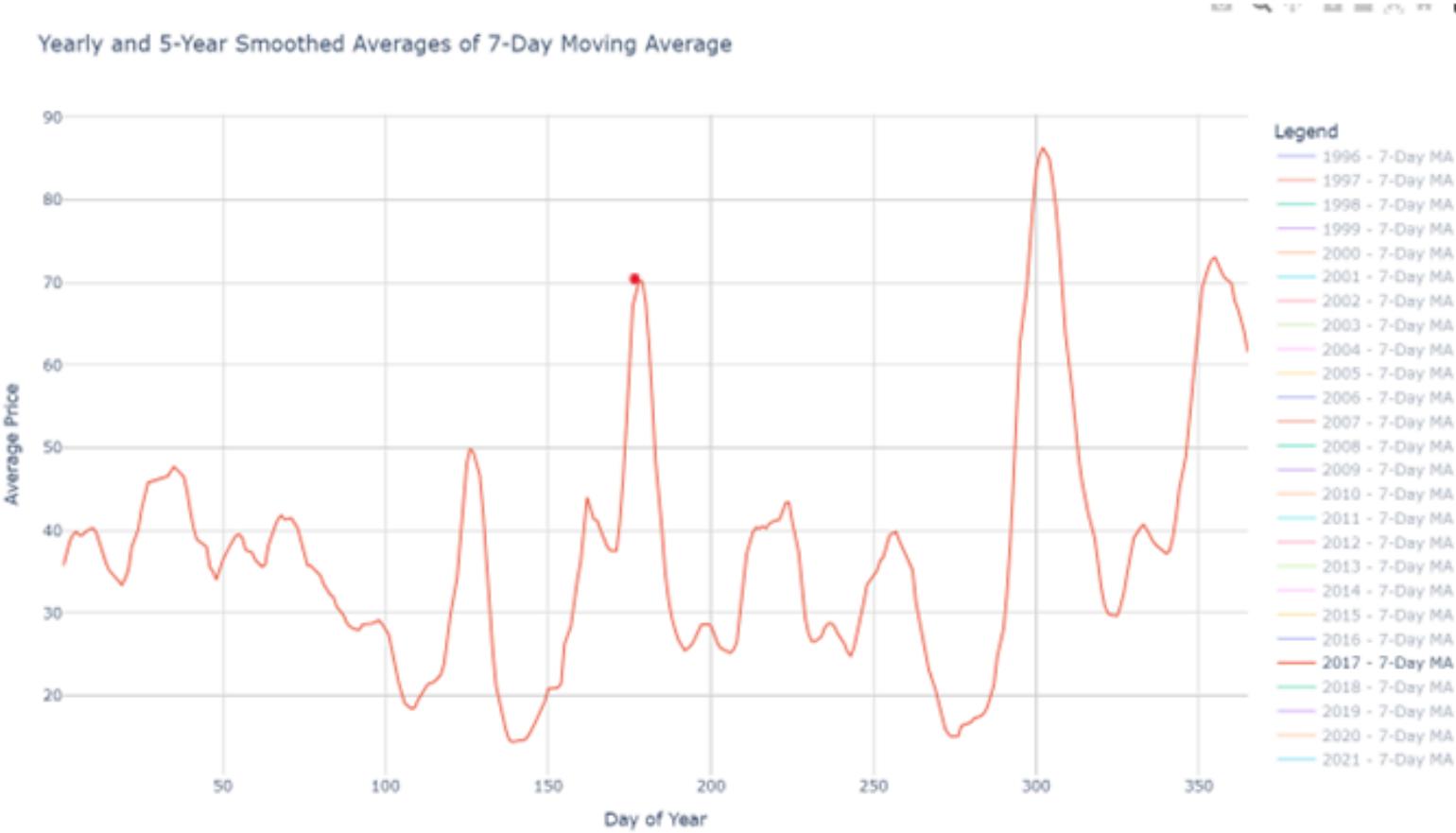
1) 果菜種類的市場特性



- 上圖為空心菜的歷年價格，可以看出部分除了春季些微較高外，無明顯趨勢

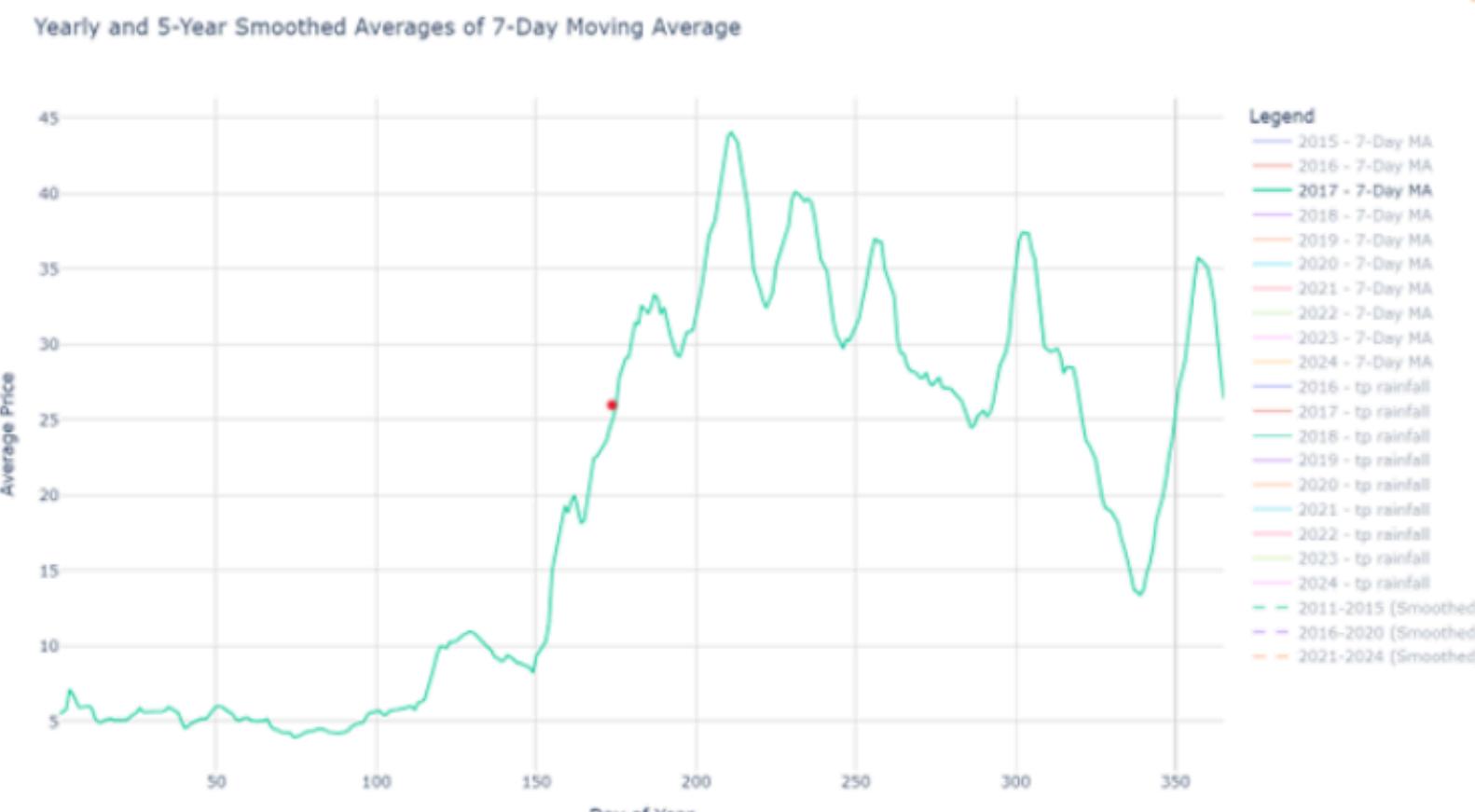
- 上圖為高麗菜(品種:初秋)歷年價格，可以明顯看出在秋冬之季價格最高，春夏之季最低

2. Data Description and Preprocessing



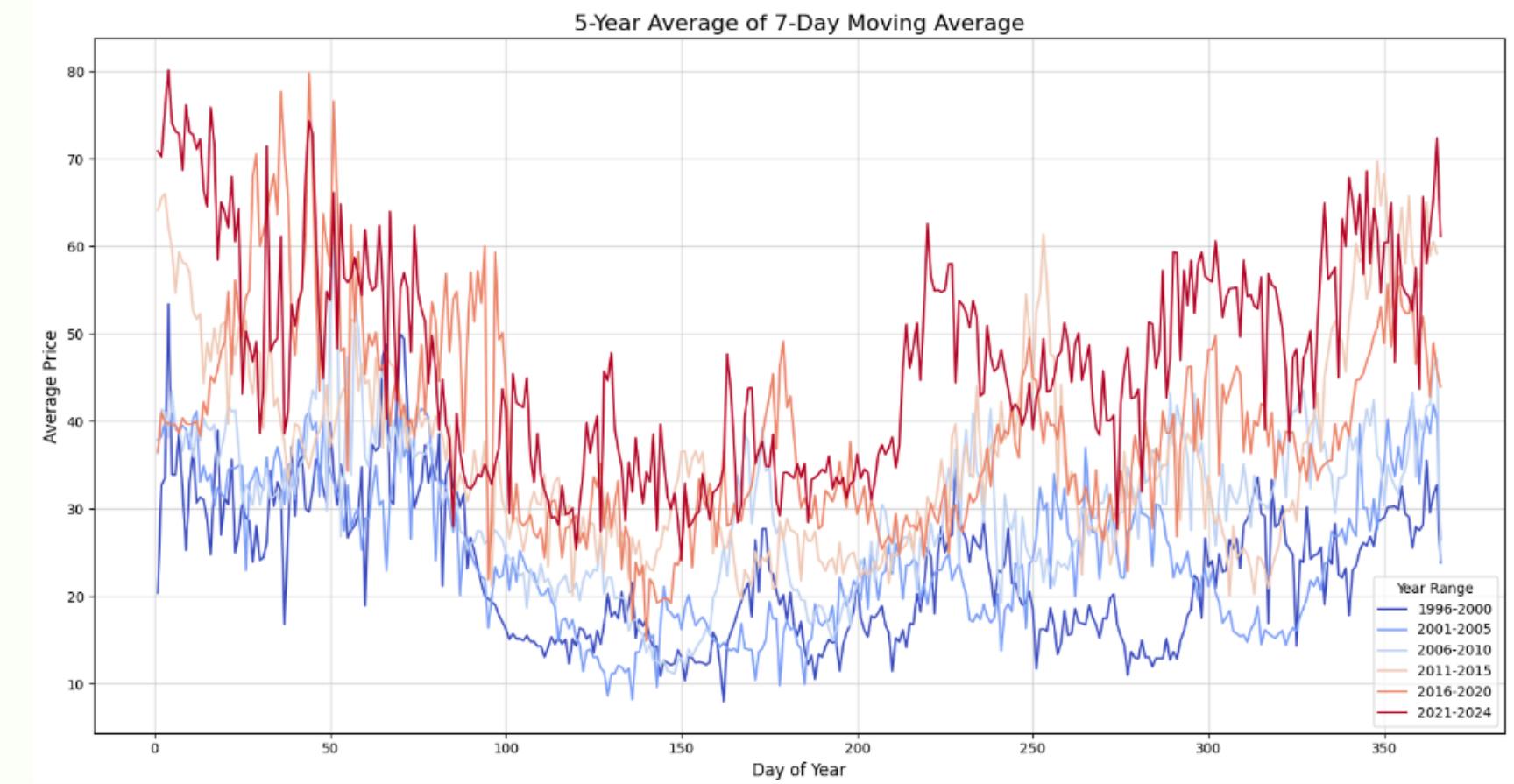
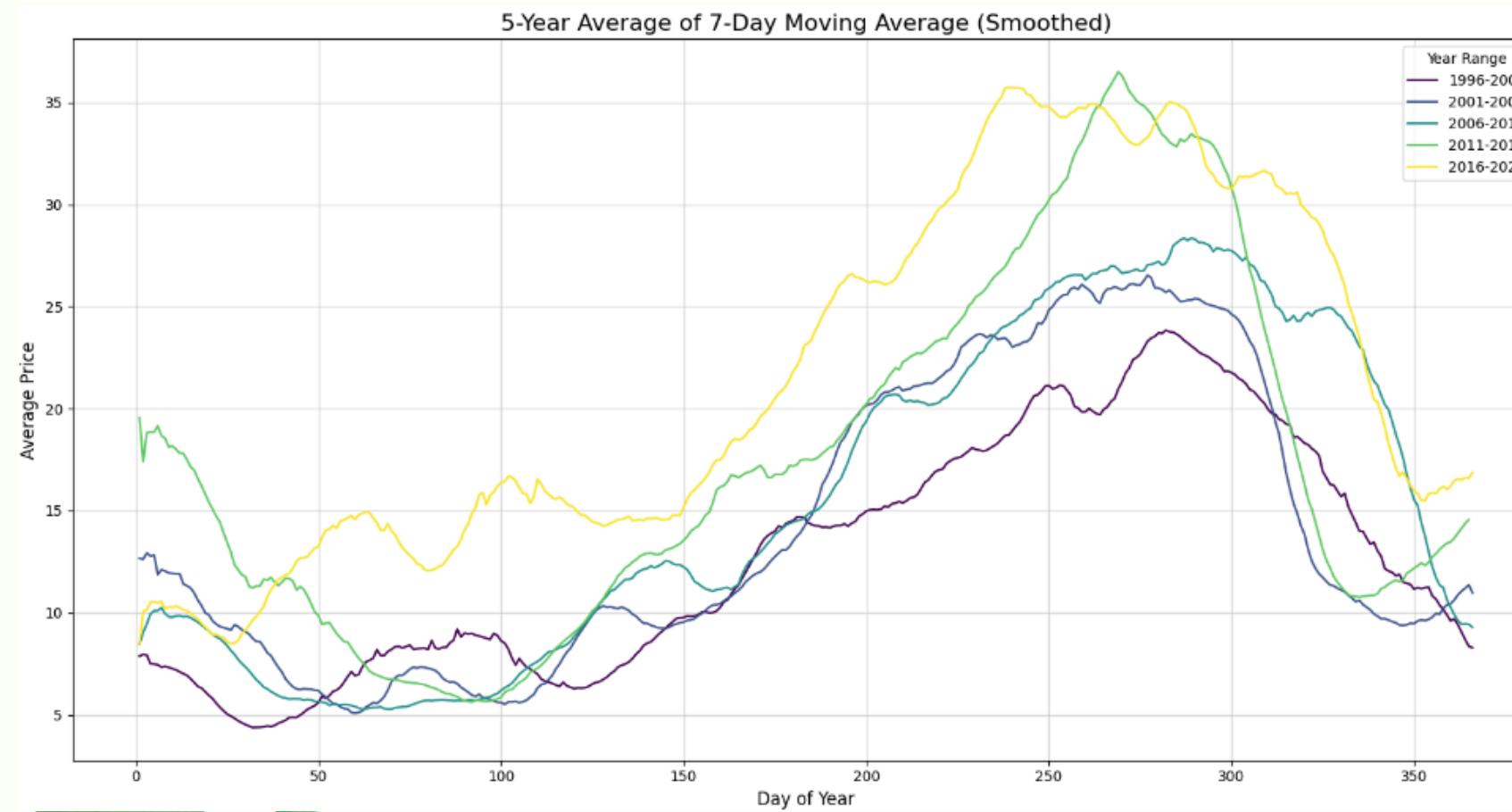
2) 植株受天災的影響程度

- 上圖（空心菜價格）可以看到在豪雨時價格上漲相當劇烈
- 下圖（高麗菜價格）可以看到在豪雨時雖然有價格上漲，但無明顯變化
- 綜合以上考量，我們最後選擇將焦點放在空心菜價格的預測上，因為他在遇到風雨時，價格有比較明顯的反應，而且他做為台灣的日常青菜，一年四季市場也較為穩定，也不像波菜、高麗菜有多個品種輪種的情形。



2. Data Description and Preprocessing

3) 通膨



通膨造成的影響也相當明顯，在20年來漲幅落在數成到一倍不等

2. Data Description and Preprocessing

期望價格 = $mean(adjust\ inflation(\text{歷史價格}))$

```
excepel_price = df_ma.copy()
excepel_price['weight'] = excepel_price['year'].map(every_year_average_price)

def calculate_custom_weighted_average(group):
    # 計算基準值：價格除以各自權重後取均值
    baseline_mean = (group['avg_price_ma7'] / group['weight']).mean()
    # 根據每年權重調整值
    group['special_avg'] = baseline_mean * group['weight']
    return group[['year', 'day_of_year', 'special_avg']]

# 按 day_of_year 分組並計算
special_average_df = (
    excepel_price.groupby('day_of_year')
    .apply(calculate_custom_weighted_average)
    .reset_index(drop=True)
)
```

資料前處理

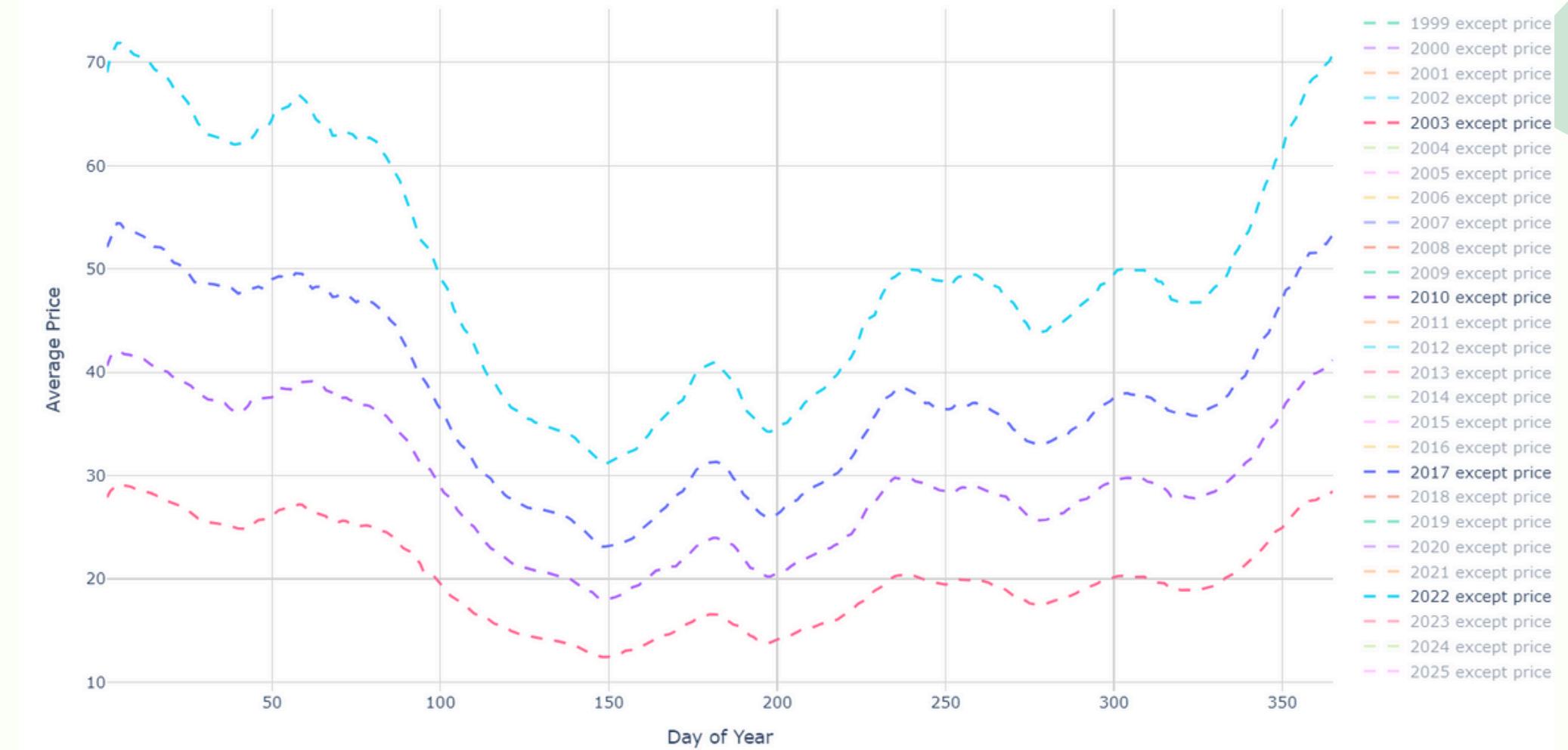
我們選擇將「實際價格」整理成「與期望價格的相差」，而期望價格的算法是先考量每年不同通膨的因素，將每年的價格用線性的方式調到相同的基準（讓他們每年的平均價格相同），之後取平均值。

但是因為歷史價格有部分的缺失且波動值極大，有先將其做一次7天的平滑，而平均後的價格在繪製成圖表時也因為波動依舊劇烈，所以在最後又疊加了一層14天的平滑，所以最後變成：

期望價格 = $14daysmooth(mean(7daysmooth(adjust\ inflation(\text{歷史價格}))))$

2. Data Description and Preprocessing

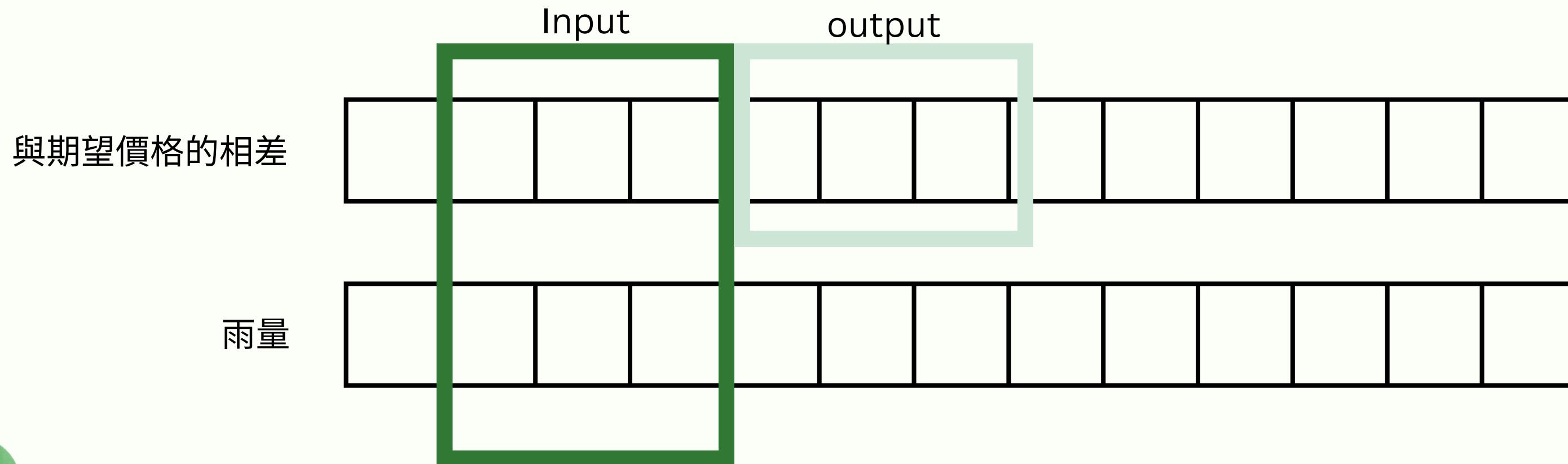
而之所以要大費周章地將實際價格換成「與期望價格的相差」是因為我們希望這樣可以減少輸入日期的干擾，讓模型只有「雨量」與「先前價格資訊」作為輸入，更可讓模型可以直接判斷現在為偏高或偏低的價格。



(上圖為推算出的各年預期價格的曲線，我們可以看到在考量通膨後，預期價格也有所不同)

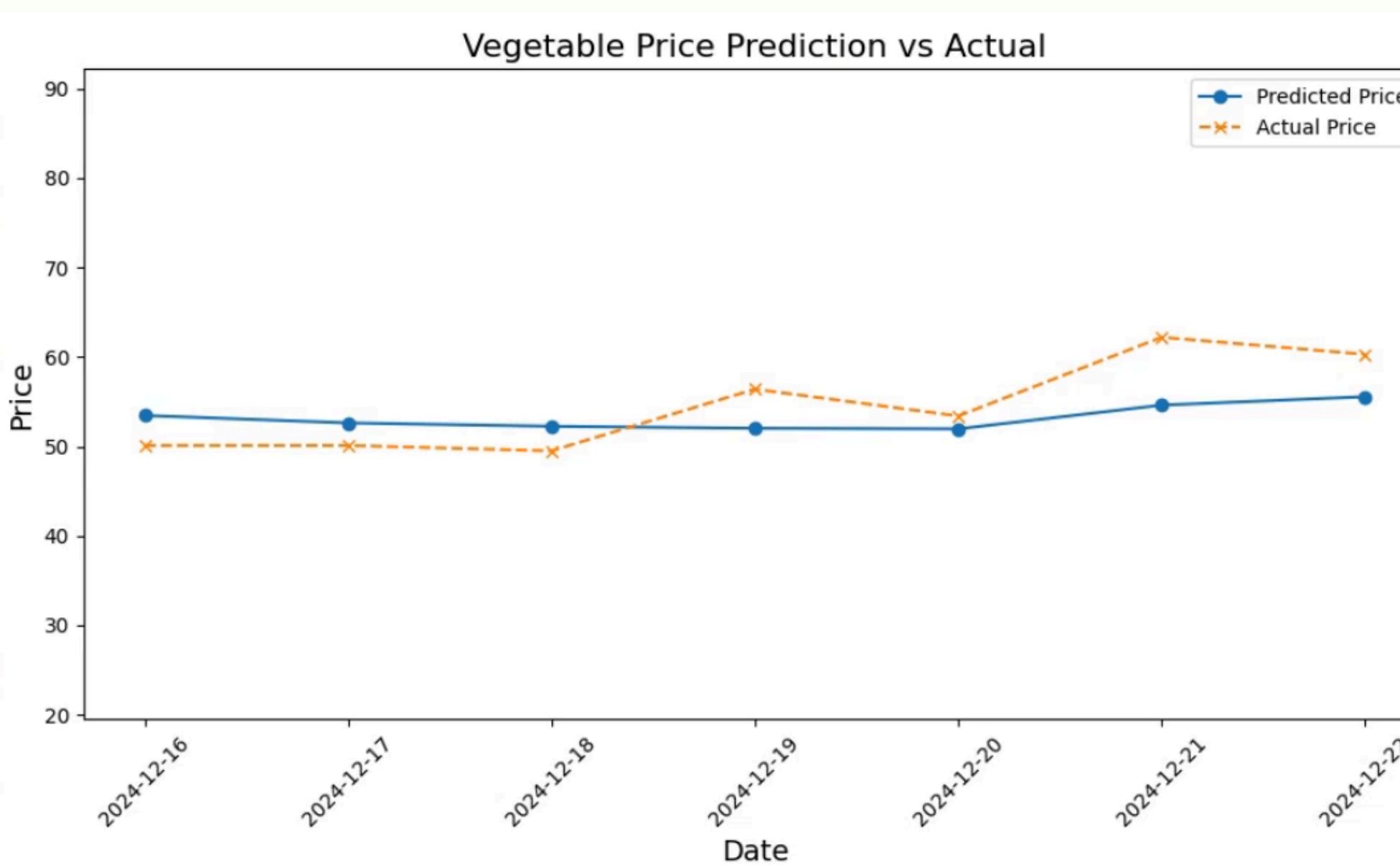
3. Methodology Details

選取訓練資料、使用LSTM



4. Evaluation and Results

在這樣的方法下，誤差約可以控制在2成以內(RMSE ≈ 12)



Epoch 37/50	
62/62	0s 4ms/step - loss: 68.0984 - val_loss: 134.3673
Epoch 38/50	
62/62	0s 3ms/step - loss: 66.4690 - val_loss: 135.6890
Epoch 39/50	
62/62	0s 3ms/step - loss: 62.8229 - val_loss: 142.1145
Epoch 40/50	
62/62	0s 4ms/step - loss: 62.0584 - val_loss: 135.9332
Epoch 41/50	
62/62	0s 3ms/step - loss: 59.9546 - val_loss: 137.8671
Epoch 42/50	
62/62	0s 3ms/step - loss: 62.9439 - val_loss: 142.6102
Epoch 43/50	
62/62	0s 3ms/step - loss: 58.4223 - val_loss: 136.5250
Epoch 44/50	
62/62	0s 3ms/step - loss: 55.6084 - val_loss: 146.7313
Epoch 45/50	
62/62	0s 4ms/step - loss: 55.0417 - val_loss: 142.5346
Epoch 46/50	
62/62	0s 3ms/step - loss: 52.9411 - val_loss: 141.4600
Epoch 47/50	
62/62	0s 3ms/step - loss: 54.2953 - val_loss: 138.1782
Epoch 48/50	
62/62	0s 3ms/step - loss: 53.8569 - val_loss: 146.4849
Epoch 49/50	
62/62	0s 4ms/step - loss: 53.7806 - val_loss: 145.6363
Epoch 50/50	
62/62	0s 4ms/step - loss: 53.8230 - val_loss: 145.4563
16/16	0s 2ms/step - loss: 152.1919
Test Loss (mse): 145.45628356933594	

5. Demo

菜價預測系統

請輸入預測起始日期 (格式: YYYYMMDD):
20241216

預測結束日期 (自動計算):
20241222

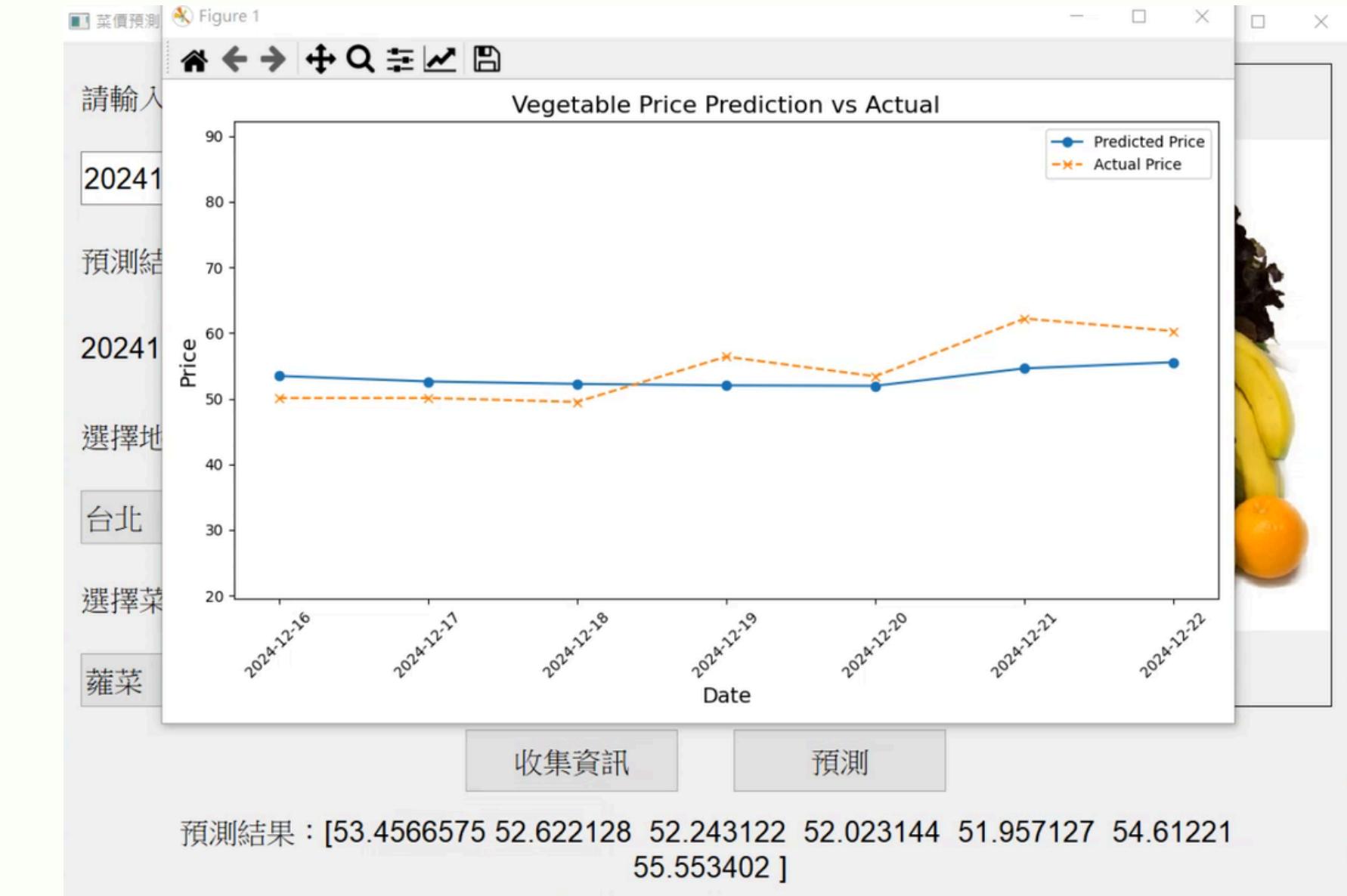
選擇地區:
台北

選擇菜類:
蘿蔔

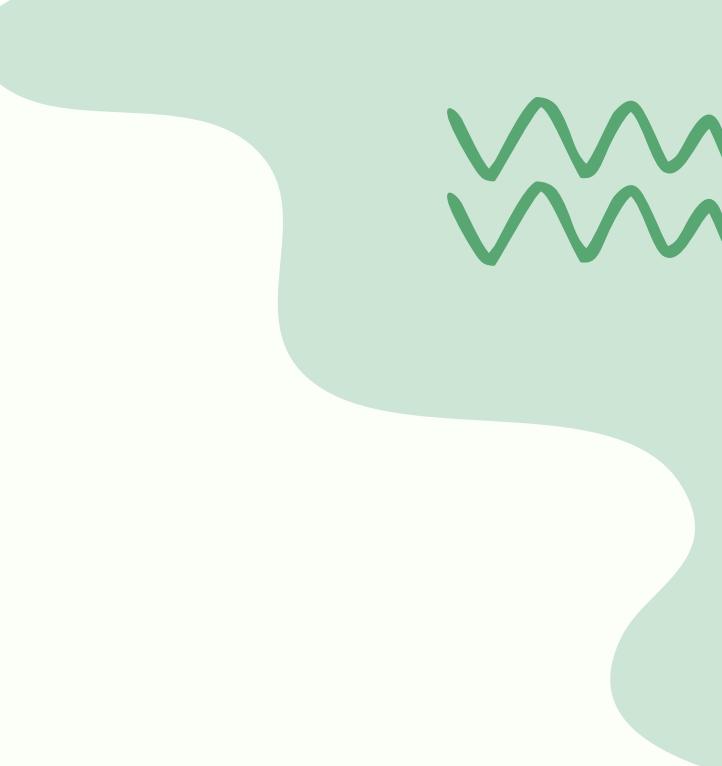
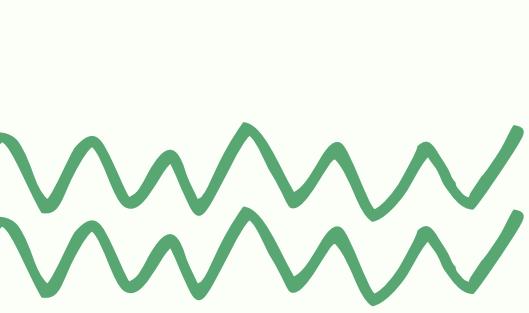


收集資訊 預測

預測前



預測後



6. Conclusions and Novelty

- 首次預測蕹菜
 - 特徵採用天氣+近期菜價
 - 預測時間長度達7天
 - 設有具拓展性之平台
- 

7. Contribution of each team member

組員名稱	貢獻
電機系116曾柏誠	<ol style="list-style-type: none">詳細分析資料，題目訓練&測試模型報告製作
資訊系114何寬羿	<ol style="list-style-type: none">使用網路爬蟲爬取資料訓練&測試模型報告製作
資訊系114林業誠	<ol style="list-style-type: none">建構UI圖像化使用者介面訓練&測試模型影片製作

Thank You