

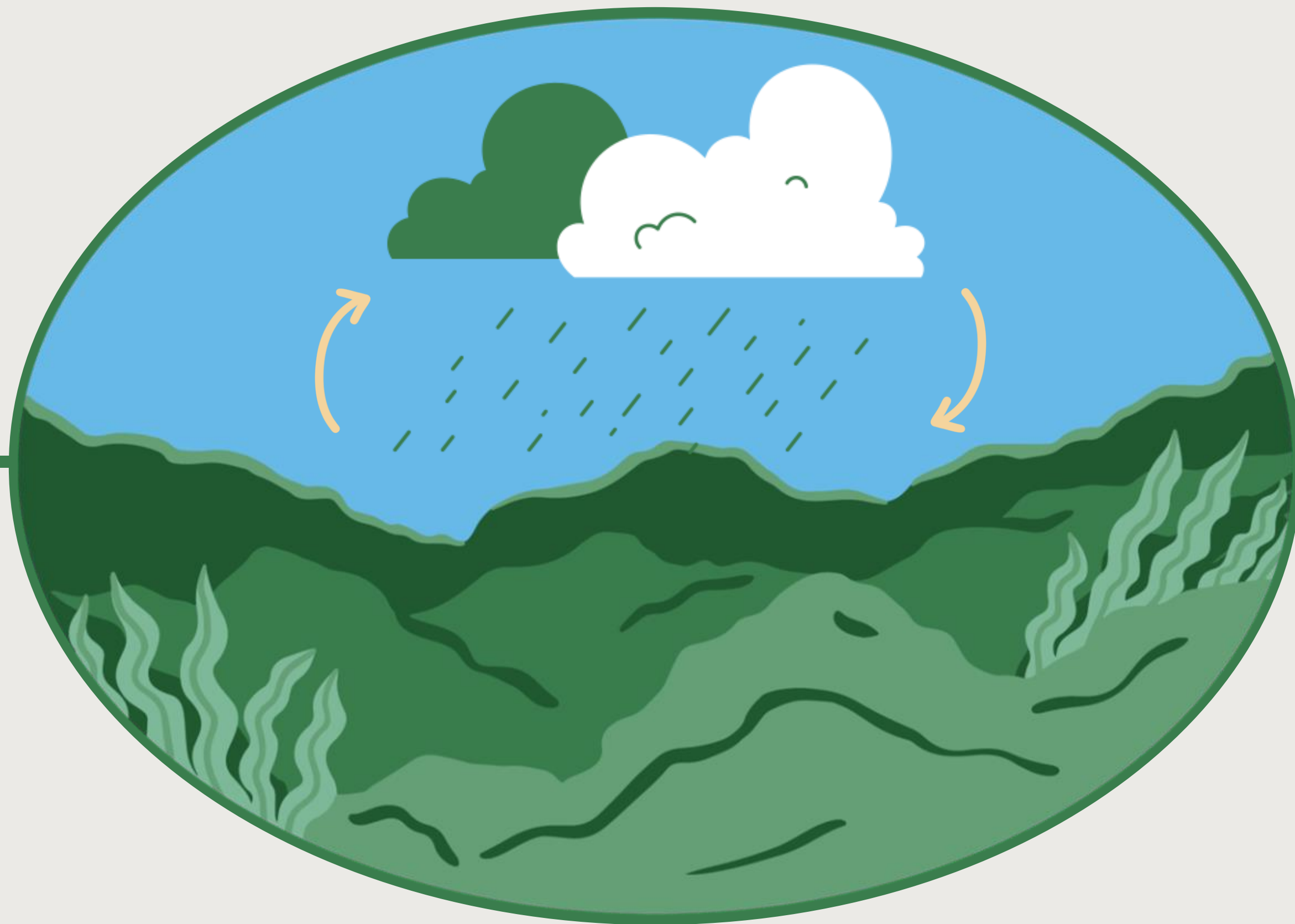
資料科學概論 — 期末專題

● 利用雨量預測空氣品質指標

● 組別：熬夜寫Code來杯Java

● 組員：江曉明、陳柏淮、楊承翰



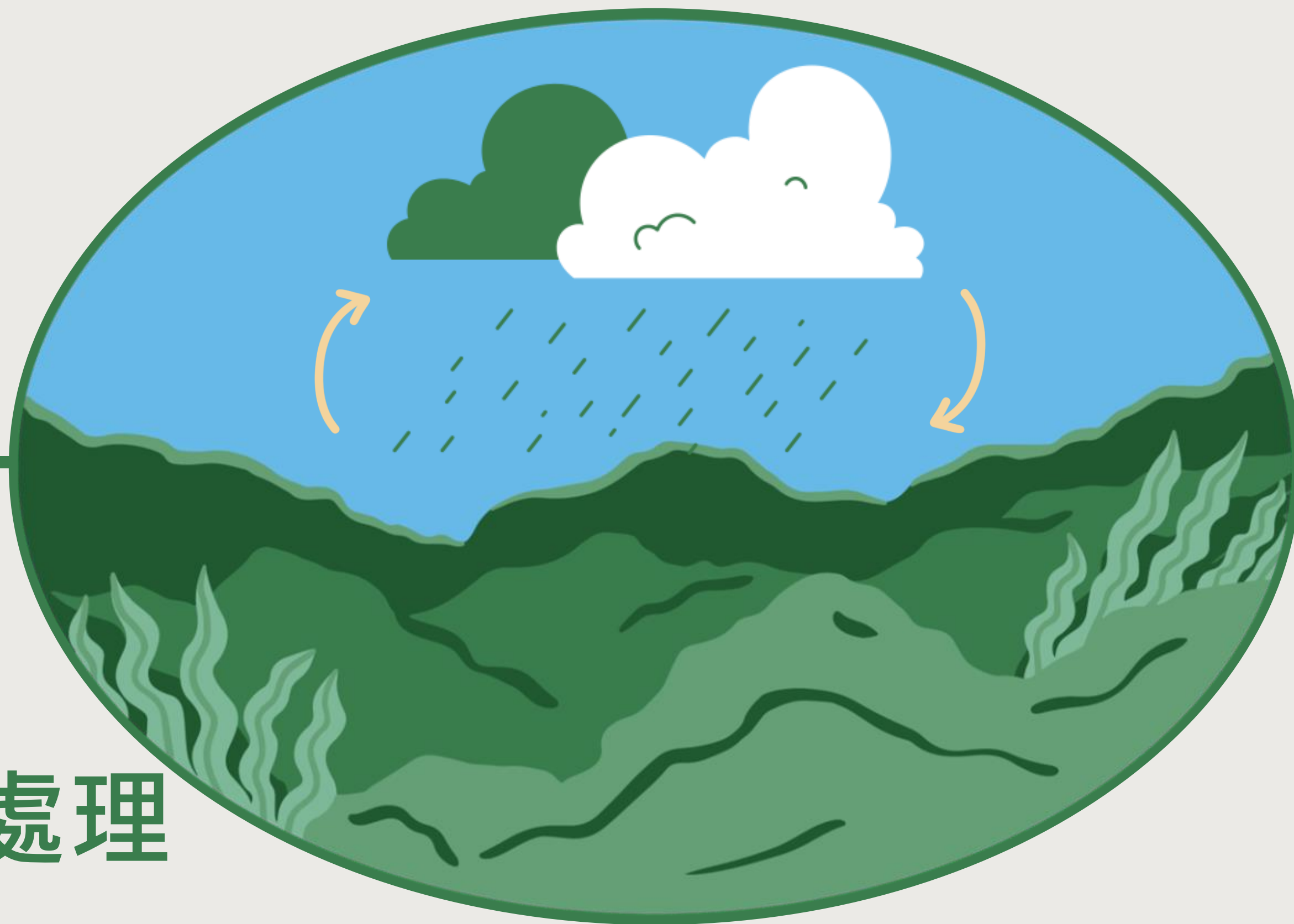


主題介紹

主題介紹

探討降雨量對空氣品質的影響

希望做出一個可以預測AQI的模型



資料蒐集與處理

資料蒐集與處理—資料蒐集

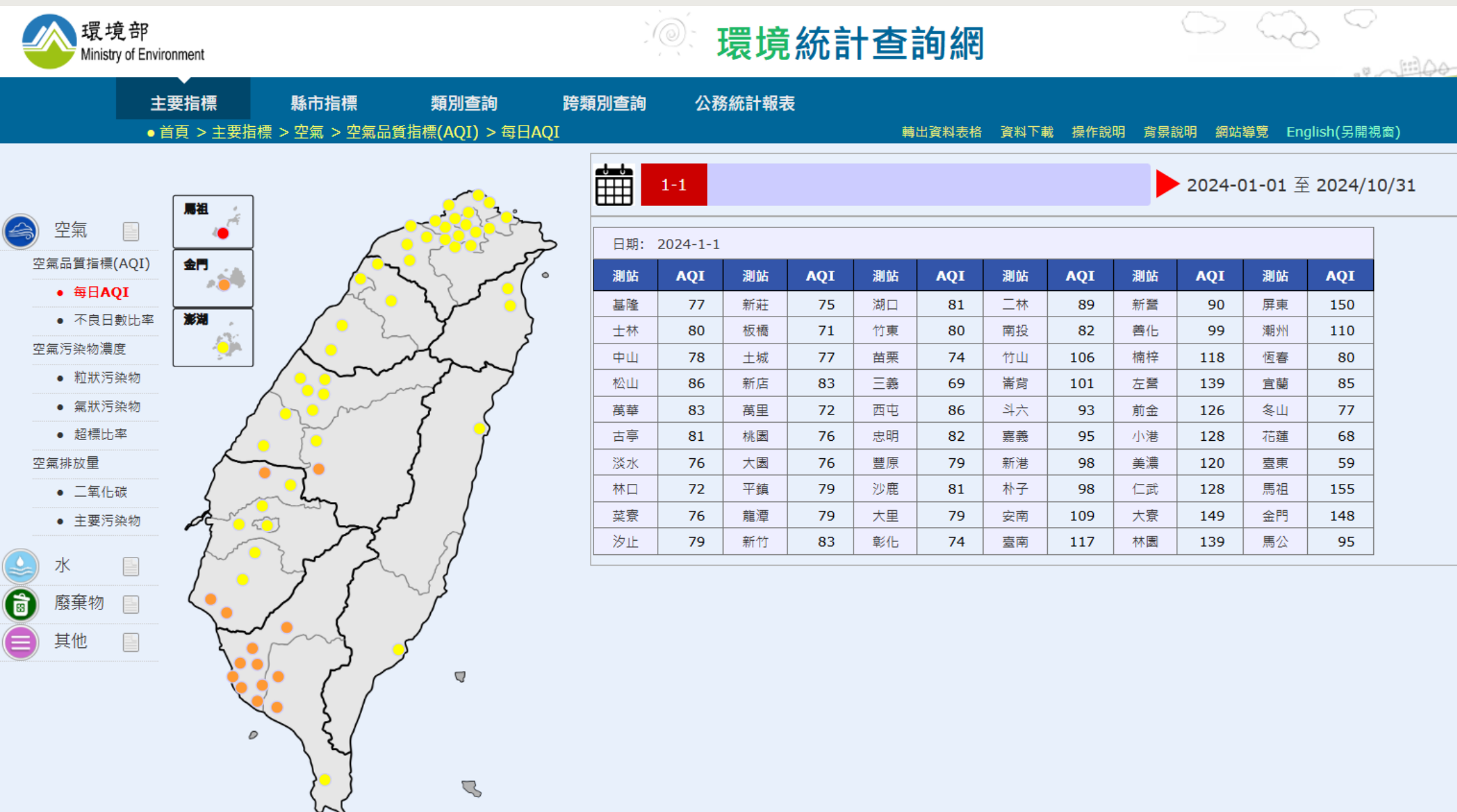
資料來源

環境部-統計查詢網

環境部-空氣品質指標(AQI)(歷史資料)

中央氣象署-每日降雨量

資料蒐集與處理—資料蒐集(1)



資料日期

2017年1月至
2024年9月底

測站數量

60個

Date, Location, AQI

2017-01-01, 基隆, 37.0

2017-01-01, 新莊, 37.0

2017-01-01, 湖口, 45.0

2017-01-01, 二林, 71.0

2017-01-01, 新營, 100.0

2017-01-01, 屏東, 153.0

2017-01-01, 士林, 39.0

2017-01-01, 板橋, 31.0

2017-01-01, 竹東, 63.0

2017-01-01, 南投, 105.0

2017-01-01, 善化, 111.0

2017-01-01, 潮州, 134.0

2017-01-01, 中山, 36.0

2017-01-01, 土城, 36.0

2017-01-01, 苗栗, 58.0

2017-01-01, 竹山, 119.0

2017-01-01, 楠梓, 138.0

2017-01-01, 恆春, 44.0

2017-01-01, 松山, 37.0

2017-01-01, 新店, 41.0

2017-01-01, 三義, 67.0

2017-01-01, 崙背, 95.0

2017-01-01, 左營, 111.0

2017-01-01, 宜蘭, 35.0

2017-01-01, 萬華, 40.0

資料蒐集與處理—資料蒐集(2)

AQI 指標

良好

0~50

普通

51~100

對敏感族群不健康

101~150

對所有族群不健康

151~200

非常不健康

201~300

危害

301~400

危害

401~500

O₃

(ppm)

8小時平均值

0.000 - 0.054

0.055 - 0.070

0.071 - 0.085

0.086 - 0.105

0.106 - 0.200

(2)

(2)

O₃

(ppm)

小時平均值(1)

-

-

0.125 - 0.164

0.165 - 0.204

0.205 - 0.404

0.405 - 0.504

0.505 - 0.604

PM_{2.5}

(µg/m³)

24小時平均值

0.0 - 15.4

15.5 - 35.4

35.5 - 54.4

54.5 - 150.4

150.5 - 250.4

250.5 - 350.4

350.5 - 500.4

PM₁₀

(µg/m³)

24小時平均值

0 - 50

51 - 100

101 - 254

255 - 354

355 - 424

425 - 504

505 - 604

CO

(ppm)

8小時平均值

0 - 4.4

4.5 - 9.4

9.5 - 12.4

12.5 - 15.4

15.5 - 30.4

30.5 - 40.4

40.5 - 50.4

SO₂

(ppb)

小時平均值

0 - 20

21 - 75

76 - 185

186 - 304⁽³⁾

305 - 604⁽³⁾

605 - 804⁽³⁾

805 - 1004⁽³⁾

NO₂

(ppb)

小時平均值

0 - 30

31 - 100

101 - 360

361 - 649

650 - 1249

1250 - 1649

1650 - 2049

即時空氣品質指標 (即時 AQI)

為提供預警參考，本部發布即時空氣品質指標 (AQI) 值，並提供每小時監測數據，而即時空氣品質指標 (AQI) 計算方式如下：

各測項即時濃度依下列公式計算後，再對應下表得出 O₃、PM_{2.5}、PM₁₀、CO、SO₂、NO₂ 等 6 個測項之即時副指標值，再取出其中最大值為即時空氣品質指標，該最大值測項即為指標污染物：

資料蒐集與處理—資料蒐集(2)

空氣品質指標(AQI)(歷史資料)

每小時提供各測站之空氣品質指標 (AQI)

資料集詮釋資料		<input type="checkbox"/>	資料集名稱	下載格式
資料集代碼	AQX_P_488	<input type="checkbox"/>	空氣品質指標(AQI)(歷史資料)	CSV JSON XML
資料集所屬單位	環境部監測資訊司	<input type="checkbox"/>	空氣品質指標(AQI)(歷史資料) (2024/11)	CSV JSON XML
資料集類別	大氣	<input type="checkbox"/>	空氣品質指標(AQI)(歷史資料) (2024/10)	CSV JSON XML
環境保護計畫類別	大氣環境	<input type="checkbox"/>	空氣品質指標(AQI)(歷史資料) (2024/09)	CSV JSON XML
聯合國永續發展目標	良好健康和福祉	<input type="checkbox"/>	空氣品質指標(AQI)(歷史資料) (2024/08)	CSV JSON XML
		<input type="checkbox"/>	空氣品質指標(AQI)(歷史資料) (2024/07)	CSV JSON XML
		<input type="checkbox"/>	空氣品質指標(AQI)(歷史資料) (2024/06)	CSV JSON XML
		<input type="checkbox"/>	空氣品質指標(AQI)(歷史資料) (2024/05)	CSV JSON XML
		<input type="checkbox"/>	空氣品質指標(AQI)(歷史資料) (2024/04)	CSV JSON XML
		<input type="checkbox"/>	空氣品質指標(AQI)(歷史資料) (2024/03)	CSV JSON XML
		<input type="checkbox"/>	空氣品質指標(AQI)(歷史資料) (2024/02)	CSV JSON XML

sitename, county, aqi, pollutant, status, so2, co, o3, o3_8hr, pm10, pm2.5, no2, nox, n...

嘉義, 嘉義市, 75.0, 細懸浮微粒, 普通, 2.9, 0.41, 16, 24.0, 45.0, 24, 15.0, 18.0, 3.2, 2.2, 1...

頭份, 苗栗縣, 35.0, nan, 良好, 2.4, 0.26, 25, 31.0, 19.0, 12, 15.0, 17.0, 1.4, 1.5, 59.0, "2...

線西, 彰化縣, 44.0, nan, 良好, 4.2, 0.36, 28, 34.0, 38.0, 14, 12.0, 13.0, 1.3, 5.1, 41.0, "2...

新竹, 新竹市, 35.0, nan, 良好, 1.0, 0.32, 27, 30.0, 18.0, 13, 6.0, 5.5, -0.5, 2.1, 54.0, "20...

竹東, 新竹縣, 30.0, nan, 良好, 1.5, 0.28, 19, 27.0, 18.0, 11, 9.2, 10.0, 1.3, 0.5, 106.0, "2...

新營, 臺南市, 62.0, 細懸浮微粒, 普通, 2.9, 0.29, 20, 26.0, 62.0, 22, 13.0, 14.0, 1.3, 2.5, 1...

湖口, 新竹縣, 34.0, nan, 良好, 2.4, 0.2, 33, 37.0, 15.0, 10, 7.9, 9.0, 1.1, 3.5, 42.0, "2017...

龍潭, 桃園市, 30.0, nan, 良好, 1.1, 0.23, 31, 32.0, 26.0, 4, 6.7, 7.6, 0.9, 3.5, 52.0, "2017...

彰化, 彰化縣, 33.0, nan, 良好, 2.5, 0.26, 25, 25.0, 25.0, 16, 10.0, 11.0, 0.4, 3.3, 330.0, "2...

平鎮, 桃園市, 33.0, nan, 良好, 0.4, 0.28, 33, 36.0, 27.0, 5, 8.7, 10.0, 1.3, 2.8, 59.0, "20...

善化, 臺南市, 58.0, 懸浮微粒, 普通, 3.5, nan, 16, 25.0, 59.0, 18, 16.0, 17.0, 1.5, 3.1, 325...

觀音, 桃園市, 36.0, nan, 良好, 4.8, 0.28, 35, 35.0, 22.0, 6, 8.0, 8.3, 0.4, 6.8, 47.0, "2017...

大園, 桃園市, 31.0, nan, 良好, 4.6, 0.2, 31, 33.0, 29.0, 6, 13.0, 15.0, 2.3, 4.7, 49.0, "20...

西屯, 臺中市, 44.0, nan, 良好, 1.4, 0.19, 24, 29.0, 29.0, 17, 15.0, 15.0, 0.0, 2.9, 48.0, "2...

桃園, 桃園市, 32.0, nan, 良好, 2.5, 0.23, 36, 35.0, 16.0, 6, 6.2, 7.5, 1.3, 3.3, 95.0, "2017...

大同, 臺北市, 36.0, nan, 良好, 2.2, 0.41, -, nan, 44.0, 14, 18.0, 28.0, 10.0, nan, nan, "20...

松山, 臺北市, 33.0, nan, 良好, 1.1, 0.28, 30, 28.0, 19.0, 10, 13.0, 14.0, 1.7, 2.0, 81.0, "2...

安南, 臺南市, 64.0, 懸浮微粒, 普通, 2.7, 0.27, 21, 30.0, 66.0, 17, 12.0, 12.0, 0.1, 2.8, 30...

古亭, 臺北市, 31.0, nan, 良好, 1.1, 0.28, 30, 34.0, 9.0, 8, 12.0, 14.0, 1.7, 3.5, 84.0, "20...

朴子, 嘉義縣, 66.0, 懸浮微粒, 普通, 1.6, 0.17, 31, 32.0, 51.0, 12, 6.6, 8.2, 1.6, 2.6, 38.0...

萬華, 臺北市, 25.0, nan, 良好, 1.8, 0.26, 31, 27.0, 22.0, 9, 13.0, 15.0, 2.2, 3.5, 107.0, "2...

志明, 臺中市, 35.0, nan, 良好, 1.5, 0.28, 21, 23.0, 6.0, 8, 15.0, 17.0, 2.3, 2.6, 16.0, "20...

中山, 臺北市, 22.0, nan, 良好, 2.7, 0.33, 29, 24.0, 17.0, 10, 14.0, 16.0, 2.2, 2.8, 120.0, "2...

崙背, 雲林縣, 59.0, 細懸浮微粒, 普通, 3.9, 0.2, 24, 30.0, 32.0, 16, 12.0, 13.0, 1.3, 2.6, 61...

大里, 臺中市, 31.0, nan, 良好, 2.6, 0.4, 14, 19.0, 20.0, 3, 22.0, 24.0, 2.1, 2.1, 31.0, "20...

沙鹿, 臺中市, 33.0, nan, 良好, 2.3, 0.22, 28, 31.0, 28.0, 13, 9.7, 10.0, 0.6, 4.3, 25.0, "20...

士林, 臺北市, 33.0, nan, 良好, 8.0, 0.17, 34, 36.0, 21.0, 6, 9.8, 10.0, 0.7, 3.8, 108.0, "20...

斗六, 雲林縣, 65.0, 細懸浮微粒, 普通, 2.2, 0.29, 18, 27.0, 40.0, 12, 13.0, 14.0, 1.2, 1.5, 1...

南投, 南投縣, 53.0, 細懸浮微粒, 普通, 2.7, 0.42, 7.7, 17.0, 41.0, 16, 25.0, 28.0, 3.2, 1.2...

豐原, 臺中市, 38.0, nan, 良好, 2.4, 0.31, 16, 27.0, 15.0, 12, 16.0, 18.0, 1.5, 1.8, 313.0, "2...

新港, 嘉義縣, 56.0, 懸浮微粒, 普通, 4.1, 0.24, 23, 28.0, 59.0, 12, 9.5, 10.0, 0.9, 2.6, 314...

二崙, 苗栗縣, 38.0, 良好, 1.4, 0.26, 24, 28.0, 18.0, 13, 6.0, 10.0, 0.5, 3.0, 30.0, "20...

資料日期

2017年1月至
2024年11月底

資料蒐集與處理—資料蒐集(3)

2024年臺北氣象站逐日雨量資料

單位：毫米

	一月	二月	三月	四月	五月	六月	七月	八月	九月	十月	十一月	十二月
1	-	-	11.5	1.5	7.5	3.5	56.0	-	-	-	6.0	-
2	-	-	7.5	0.5	9.5	43.5	16.5	-	-	14.5	T	-
3	5.0	T	T	-	-	0.5	T	-	-	22.0	-	T
4	-	T	-	4.5	T	T	-	-	-	6.0	T	-
5	-	1.0	-	T	T	25.5	-	T	-	1.0	T	-
6	-	T	3.0	1.0	-	6.0	9.0	-	-	-	T	-
7	-	16.0	11.5	4.0	-	T	2.0	T	24.5	14.5	T	T
8	T	13.0	-	19.5	-	-	51.0	T	-	23.0	T	T
9	2.5	12.5	T	1.0	-	16.0	23.5	17.5	-	0.5	0.5	T
10	T	-	T	5.5	-	-	34.0	8.5	1.5	T	-	T
11	-	-	23.0	-	-	T	T	3.0	-	-	1.5	-
12	-	-	2.0	-	16.0	-	-	13.5	-	-	6.5	T
13	-	-	-	-	14.0	-	T	0.5	5.5	-	6.5	T
14	-	-	T	T	0.5	1.5	-	16.5	33.0	-	T	T
15	4.5	-	0.5	-	-	T	-	31.0	-	T	T	-

- rainfall
- Changhua_Precipitation.csv
 - Chiayi_Precipitation.csv
 - Hsinchu_Precipitation.csv
 - Hualian_Precipitation.csv
 - Kaohsiung_Precipitation.csv
 - Keelung_Precipitation.csv
 - Kinmen_Precipitation.csv
 - Matsu_Precipitation.csv
 - Miaoli_Precipitation.csv
 - NewTaipei_Precipitation.csv
 - Penghu_Precipitation.csv
 - Pingtung_Precipitation.csv
 - SunMoonLake_Precipitation.csv
 - Taichung_Precipitation.csv
 - Tainan_Precipitation.csv
 - Taipei_Precipitation.csv
 - Taitung_Precipitation.csv
 - Taoyuan_Precipitation.csv
 - Yilan_Precipitation.csv
 - Yunlin_Precipitation.csv

```
Serial,Precipitation(mm)
20170101,0
20170102,0
20170103,0
20170104,0
20170105,0
20170106,0
20170107,0
20170108,0
20170109,0
20170110,0
20170111,0
20170112,0
20170113,0
20170114,1.3
20170115,0
20170116,0
20170117,0
20170118,0
20170119,0
20170120,0
20170121,0
20170122,0
20170123,0
20170124,0
20170125,0
```

資料日期

2017年1月至
2024年11月底

縣市數量

20個

資料蒐集與處理—資料處理

爬取的AQI 並沒有2021年5月的資料

→ 使用下載的資料集進行補充。

下載的資料集是以小時為單位計算AQI

→ 對該天的空氣汙染物取平均，作為該日空氣汙染物的數值

下載的資料集超過60個測站

→ 刪掉其餘的測站，只著重在能夠爬取AQI的測站

訓練集與測試集

→ 選擇 2017年~2024年9月底為訓練集，選擇 2024年10月、11月作為測試集

資料蒐集與處理—資料處理

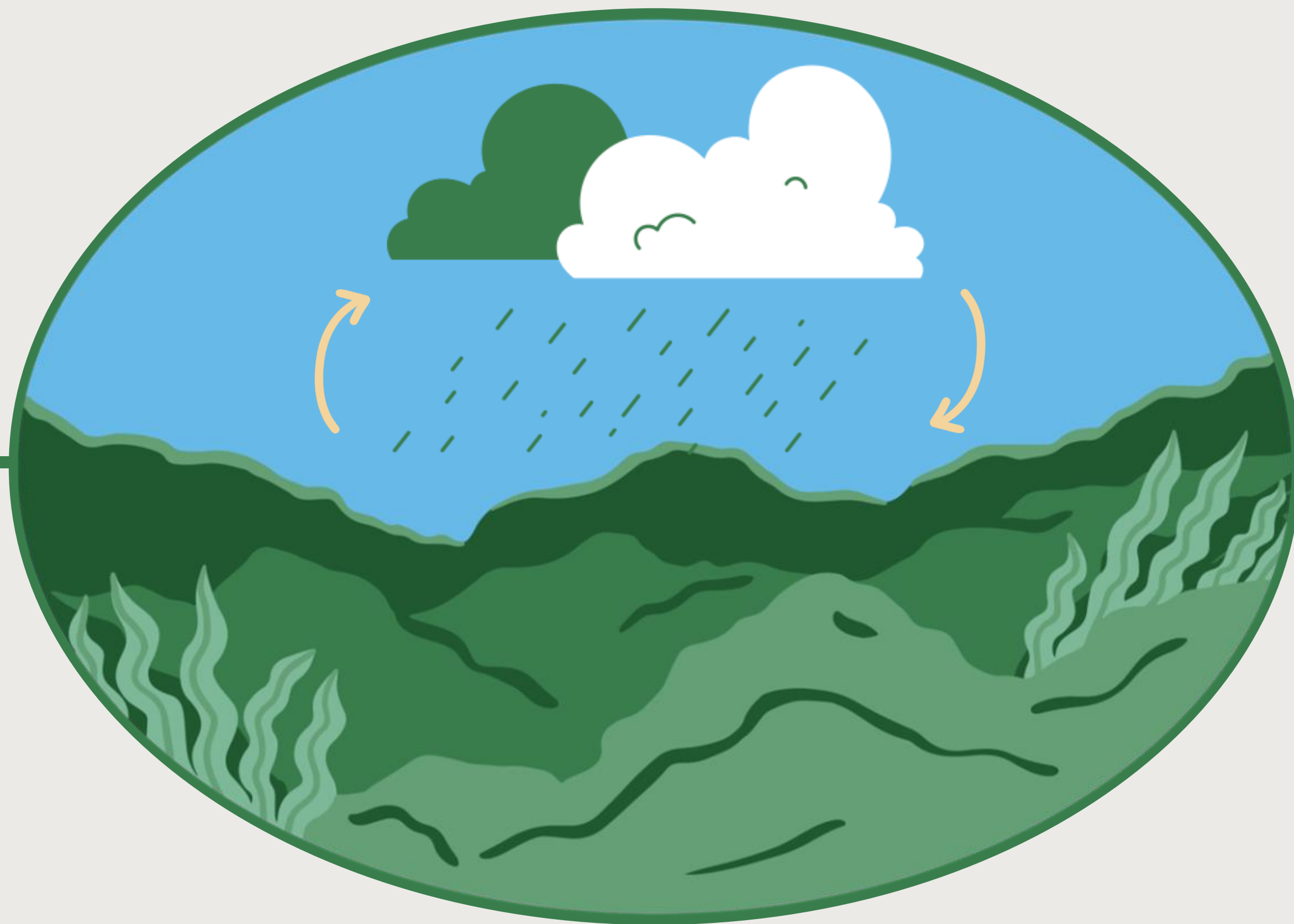
測站名稱 (sitename)	縣市 (county)	空氣品質指標 (AQI)	主要污染物 (pollutant)	空氣品質狀態 (status)
二氧化硫 (SO ₂)	一氧化碳 (CO)	臭氧 (O ₃)	臭氧8小時平均 (O ₃ _8hr)	懸浮微粒 (pm10)
細懸浮微粒 (pm2.5)	二氧化氮 (NO ₂)	氮氧化物 (NO _x)	一氧化氮 (NO)	風速 (windspeed)
風向 (winddirec)	資料建立日期 (datacreationdate)	單位 (unit)	一氧化碳 8小時平均 (CO_8hr)	細懸浮微粒平均 (pm2.5_avg)
懸浮微粒平均 (pm10_avg)	二氧化硫平均 (SO ₂ _avg)	經度 (longitude)	緯度 (latitude)	測站編號 (siteid)

資料蒐集與處理—資料處理

測站名稱 (sitename)	縣市 (county)	空氣品質指標 (AQI_2)	主要污染物 (pollutant)	空氣品質狀態 (status)
二氧化硫 (SO ₂)	一氧化碳 (CO)	臭氧 (O ₃)	臭氧8小時平均 (O ₃ _8hr)	懸浮微粒 (pm10)
細懸浮微粒 (pm2.5)	二氧化氮 (NO ₂)	氮氧化物 (NO _x)	一氧化氮 (NO)	風速 (windspeed)
風向 (winddirec)	資料建立日期 (datacreationdate)	單位 (unit)	一氧化碳 8小時平均 (CO_8hr)	細懸浮微粒平均 (pm2.5_avg)
懸浮微粒平均 (pm10_avg)	二氧化硫平均 (SO ₂ _avg)	經度 (longitude)	緯度 (latitude)	測站編號 (siteid)
降雨量 (Precipitation)	隔日的空氣品質指標 (next_AQI)			

資料蒐集與處理—資料處理

```
datacreationdate,sitename,county,aqi,aqi_2,so2,co,o3,o3_8hr,pm10,pm2.5,no2,nox,no,windspeed,winddirec,co_8hr,pm2.5_avg,pm10_avg,Precipitation(mm),next_aqi
2017-01-01,三義,苗栗縣,64.0,67.0,1.579,0.328,33.917,32.125,46.25,21.542,7.258,8.196,0.954,2.129,94.042,0.358,20.833,46.875,0.0,81.0
2017-01-02,三義,苗栗縣,75.083,81.0,1.495,0.314,29.708,28.208,60.87,27.348,9.583,10.883,1.208,2.508,79.875,0.325,25.333,58.25,0.0,52.0
2017-01-03,三義,苗栗縣,60.792,52.0,1.921,0.255,26.917,27.917,38.042,16.625,8.779,10.292,1.5,2.979,39.292,0.263,19.792,42.333,0.0,67.0
2017-01-04,三義,苗栗縣,60.125,67.0,1.771,0.288,29.725,31.083,40.333,21.667,8.5,9.708,1.217,1.696,84.25,0.258,19.5,37.042,0.0,120.0
2017-01-05,三義,苗栗縣,111.792,120.0,2.392,0.351,31.667,30.458,85.25,42.875,11.567,13.183,1.654,1.771,186.25,0.354,39.958,77.25,0.0,79.0
2017-01-06,三義,苗栗縣,90.25,79.0,1.967,0.342,26.4,26.792,66.083,27.417,12.262,14.183,1.929,2.504,60.167,0.346,31.292,73.875,0.0,79.0
2017-01-07,三義,苗栗縣,75.958,79.0,2.121,0.396,19.817,21.042,60.0,26.792,13.988,17.779,3.762,1.588,145.112,0.383,25.542,59.542,0.0,64.0
2017-01-08,三義,苗栗縣,67.625,64.0,1.804,0.36,24.0,23.875,39.75,20.917,11.183,13.208,1.975,2.938,39.567,0.346,22.542,42.833,0.0,61.0
2017-01-09,三義,苗栗縣,59.565,61.0,1.63,0.478,33.391,31.087,35.957,19.957,7.961,9.517,1.696,4.896,39.348,0.47,19.13,35.478,0.0,80.0
2017-01-10,三義,苗栗縣,73.833,80.0,1.854,0.408,36.792,37.292,45.542,26.458,7.904,9.658,1.783,3.733,38.625,0.421,24.917,42.708,0.0,88.0
2017-01-11,三義,苗栗縣,86.913,88.0,1.618,0.392,29.409,30.348,54.909,30.727,12.105,13.957,1.776,2.957,39.304,0.409,30.087,53.478,0.0,71.0
2017-01-12,三義,苗栗縣,84.167,71.0,1.592,0.34,23.083,22.667,48.167,24.792,12.338,13.6,1.179,3.4,43.583,0.367,28.958,54.042,0.0,29.0
2017-01-13,三義,苗栗縣,30.417,29.0,1.338,0.417,21.208,21.417,20.917,9.042,14.321,15.454,1.154,4.487,40.417,0.383,9.375,22.042,0.0,43.0
2017-01-14,三義,苗栗縣,37.75,43.0,1.262,0.39,37.333,34.792,20.75,12.708,6.975,8.392,1.383,5.454,40.25,0.421,11.25,18.583,8.0,32.0
2017-01-15,三義,苗栗縣,36.083,32.0,0.721,0.21,32.417,34.0,15.583,18.0,4.317,5.792,1.483,4.379,40.083,0.229,9.75,18.375,1.0,30.0
2017-01-16,三義,苗栗縣,27.435,30.0,1.257,0.182,27.652,28.783,13.739,7.087,5.196,6.596,1.383,3.887,40.348,0.2,6.783,12.348,0.0,50.0
2017-01-17,三義,苗栗縣,47.417,50.0,1.217,0.256,22.904,23.042,31.083,15.583,10.912,14.979,4.008,2.583,66.125,0.229,14.625,28.667,0.0,37.0
2017-01-18,三義,苗栗縣,38.458,37.0,1.881,0.318,36.833,36.583,87.381,31.18,33.14,11.188,1.813,2.838,41.0,0.381,18.485,38.167,0.0,54.0
```



模型建構



基於不同模型

使用 Regression NN

使用 LSTM

使用其它模型

基於不同預測輸出

預測隔日的AQI

先預測隔日的空氣污染
指標，進而推算AQI

基於區域劃分

以測站為單位建構模型

以縣市為單位建構模型

模型建構方法

江曉明

使用 Regression NN

預測隔日的AQI

以測站為單位建構模型

陳柏淮

使用其它模型

先預測隔日的空氣污染
指標，進而推算AQI

以縣市為單位建構模型

楊承翰

使用 LSTM

先預測隔日的空氣污染
指標，進而推算AQI

以測站為單位建構模型

模型建構—江曉明

輸入特徵

月份 (Month)	前日降雨量 (Precipitation)	空氣品質指標 (AQI)	一氧化碳 (CO)	一氧化碳8小時平均 (CO_8hr)
一氧化氮 (NO)	二氧化氮 (NO ₂)	氮氧化物 (NO _x)	臭氧 (O ₃)	臭氧8小時平均 (O ₃ _8hr)
懸浮微粒 (pm10)	細懸浮微粒 (pm2.5)	懸浮微粒平均 (pm10_avg)	細懸浮微粒平均 (pm2.5_avg)	二氧化硫 (SO ₂)

輸出特徵

空氣品質指標 (next_AQI)

模型建構—江曉明

使用的模型：Regression NN + XGBoost

Regression NN模型結構

```
class RegressionNN(nn.Module):

    def __init__(self, input_size, output_size, hidden_size = [64, 128, 256, 1024, 1024
, 512, 128, 128, 64]):
        super(RegressionNN, self).__init__()

        self.hidden_layer_cnt = len(hidden_size)
        self.fc = nn.ModuleList()
        self.bn = nn.ModuleList()

        self.fc.append(nn.Linear(input_size, hidden_size[0]))
        self.bn.append(nn.BatchNorm1d(hidden_size[0]))

        for i in range(1, len(hidden_size)):
            self.fc.append(nn.Linear(hidden_size[i - 1], hidden_size[i]))
            self.bn.append(nn.BatchNorm1d(hidden_size[i]))

        self.fc.append(nn.Linear(hidden_size[-1], output_size))

        self.relu = nn.ReLU()                # ReLU function
        self.dropout = nn.Dropout(0.3)       # Dropout to avoid overfitting

    def forward(self, x):

        for i in range(self.hidden_layer_cnt):
            x = self.fc[i](x)
            x = self.bn[i](x)
            x = self.relu(x)
            x = self.dropout(x)
        x = self.fc[-1](x)

        return x
```

XGBoost 模型結構

```
param_dist = {
    'learning_rate': uniform(0.01, 0.3),
    'max_depth': randint(3, 10),
    'n_estimators': randint(100, 501),
    'subsample': uniform(0.6, 0.4),
    'colsample_bytree': uniform(0.6, 0.4),
    'reg_alpha': uniform(0, 1),
    'reg_lambda': uniform(1, 2),
}

xgb_model = xgb.XGBRegressor(
    objective='reg:absoluteerror',
    random_state=random_seed
)

random_search = RandomizedSearchCV(
    estimator = xgb_model,
    param_distributions = param_dist,
    n_iter = 50,
    scoring = 'neg_mean_absolute_error',
    cv = 3,
    verbose = 1,
    random_state = random_seed,
    n_jobs = -1
)
```

模型建構—江曉明

```
regression_y = regression_nn.predict(regression_model, X)
xgboost_y = xgboost.predict(xgboost_model, X)
X['regression'] = regression_y.cpu()
X['xgboost'] = xgboost_y
```

輸入特徵

月份 (Month)	前日降雨量 (Precipitation)	空氣品質指標 (AQI)	一氧化碳 (CO)	一氧化碳8小時 平均 (CO_8hr)
一氧化氮 (NO)	二氧化氮 (NO ₂)	氮氧化物 (NO _x)	臭氧 (O ₃)	臭氧8小時平均 (O ₃ _8hr)
懸浮微粒 (pm10)	細懸浮微粒 (pm2.5)	懸浮微粒平均 (pm10_avg)	細懸浮微粒平 均 (pm2.5_avg)	二氧化硫 (SO ₂)
Regression NN 預測結果	XGBoost 預測結果			

輸出特徵

空氣品質指標 (next_AQI)

模型建構—江曉明

```
ensemble_regression, train_losses, val_losses = regression_nn.build_model(X, y  
, ensemble_path, epochs = 150)
```


模型建構—陳柏淮

輸入特徵

當日降雨量 (next_Precipitation)	前日降雨量 (Precipitation)	空氣品質指標 (AQI)	一氧化碳 (CO)	一氧化碳8小時平均 (CO_8hr)
一氧化氮 (NO)	二氧化氮 (NO ₂)	氮氧化物 (NO _x)	臭氧 (O ₃)	臭氧8小時平均 (O ₃ _8hr)
懸浮微粒 (pm10)	細懸浮微粒 (pm2.5)	風速 (Windspeed)	風向 (Winddirec)	二氧化硫 (SO ₂)
當前縣市的 各個測站	年份 (Year)	月份 (Month)	日 (Day)	

輸出特徵

二氧化硫 (next_SO ₂)	一氧化碳 (next_CO)	臭氧8小時平均 (next_O ₃ _8hr)	臭氧 (next_O ₃)	細懸浮微粒 (next_pm2.5)	懸浮微粒 (next_pm10)	二氧化氮 (next_NO ₂)
---------------------------------	-------------------	---------------------------------------	------------------------------	-----------------------	---------------------	---------------------------------

模型建構—陳柏淮

使用的模型：XGBoost

模型結構

```
xgb = XGBRegressor(  
    verbosity=1,  
    objective='reg:absoluteerror',  
    random_state=42  
)  
# Define parameter distributions for random search  
param_distributions = {  
    "n_estimators": [100, 200, 300, 400, 500],  
    "max_depth": [3, 5, 7, 9, 11],  
    "learning_rate": [0.01, 0.05, 0.1, 0.2],  
    "subsample": [0.6, 0.8, 1.0],  
    "colsample_bytree": [0.6, 0.8, 1.0],  
    "gamma": [0, 0.1, 0.2, 0.3],  
}
```

RandomizedSearchCV 參數

```
# Perform random search  
rand_search = RandomizedSearchCV(  
    xgb,  
    param_distributions,  
    n_iter=50,  
    scoring='neg_mean_absolute_error',  
    cv=5,  
    verbose=0,  
    n_jobs=-1,  
    random_state=42,  
)
```

使用 RandomizedSearchCV 以及交叉驗證找出最佳超參數，並設定目標函數為平均絕對誤差（MAE）。

模型建構—陳柏淮

預測隔日的空氣汙染物指標後，
就利用AQI公式計算隔日的AQI。

```
def compute_predicted_aqi(row):
    so2 = row["predicted_next_so2"]
    co = row["predicted_next_co"]
    o3_8hr = row["predicted_next_o3_8hr"]
    o3_1hr = row["predicted_next_o3"]
    pm10 = row["predicted_next_pm10"]
    pm25 = row["predicted_next_pm2.5"]
    no2 = row["predicted_next_no2"]

    o3_8hr_i = calc_sub_index("o3_8hr", o3_8hr)
    o3_1hr_i = calc_sub_index("o3_1hr", o3_1hr)
    o3_i = max(o3_8hr_i, o3_1hr_i)

    pm25_i = calc_sub_index("pm2.5", pm25)
    pm10_i = calc_sub_index("pm10", pm10)
    co_i = calc_sub_index("co", co)
    so2_i = calc_sub_index("so2", so2)
    no2_i = calc_sub_index("no2", no2)

    aqi = max(o3_i, pm25_i, pm10_i, co_i, so2_i, no2_i)
    return aqi
```

```
breakpoints = {
    "o3_8hr": [
        (0.000, 0.054, 0, 50),
        (0.055, 0.070, 51, 100),
        (0.071, 0.085, 101, 150),
        (0.086, 0.105, 151, 200),
        (0.106, 0.200, 201, 300),
    ],
    "o3_1hr": [
        (0.125, 0.164, 101, 150),
        (0.165, 0.204, 151, 200),
        (0.205, 0.404, 201, 300),
        (0.405, 0.504, 301, 400),
        (0.505, 0.604, 401, 500),
    ],
    "pm2.5": [
        (0.0, 15.4, 0, 50),
        (15.5, 35.4, 51, 100),
        (35.5, 54.4, 101, 150),
        (54.5, 150.4, 151, 200),
        (150.5, 250.4, 201, 300),
        (250.5, 350.4, 301, 400),
        (350.5, 500.4, 401, 500),
    ],
    "pm10": [
        (0, 50, 0, 50),
        (50, 100, 51, 100),
        (101, 254, 101, 150),
        (255, 354, 151, 200),
        (355, 424, 201, 300),
        (425, 504, 301, 400),
        (505, 604, 401, 500),
    ],
}
```

```
"co": [
    (0.0, 4.4, 0, 50),
    (4.5, 9.4, 51, 100),
    (9.5, 12.4, 101, 150),
    (12.5, 15.4, 151, 200),
    (15.5, 30.4, 201, 300),
    (30.5, 40.4, 301, 400),
    (40.5, 50.4, 401, 500),
],
"so2": [
    (0, 20, 0, 50),
    (21, 75, 51, 100),
    (76, 185, 101, 150),
    (186, 304, 151, 200),
    (305, 604, 201, 300),
    (605, 804, 301, 400),
    (805, 1004, 401, 500),
],
"no2": [
    (0, 30, 0, 50),
    (31, 100, 51, 100),
    (101, 360, 101, 150),
    (361, 649, 151, 200),
    (650, 1249, 201, 300),
    (1250, 1649, 301, 400),
    (1650, 2049, 401, 500),
],
}
```

模型建構—楊承翰

輸入特徵

前日降雨量 (Precipitation)	空氣品質指標 (AQI)	一氧化碳 (CO)	一氧化碳8小時平均 (CO_8hr)	一氧化氮 (NO)
二氧化氮 (NO ₂)	氮氧化物 (NO _x)	臭氧 (O ₃)	臭氧8小時平均 (O ₃ _8hr)	懸浮微粒 (pm10)
細懸浮微粒 (pm2.5)	懸浮微粒平均 (pm10_avg)	細懸浮微粒平均 (pm2.5_avg)	二氧化硫 (SO ₂)	

輸出特徵

輸出特徵與輸入特徵相同

模型建構—楊承翰

使用的模型：LSTM

LSTM 模型結構

```
model = Sequential([
    LSTM(128, input_shape=(seq_length, len(features)), return_sequences=True),
    LSTM(64, return_sequences=True),
    LSTM(32),
    Dense(len(features)),
])

model.compile(optimizer="adam", loss="mse", metrics=["mae"])

early_stopping = EarlyStopping(monitor="val_loss", patience=5, restore_best_weights=True)

model.fit(
    X_train_seq,
    y_train_seq,
    validation_data=(X_test_seq, y_test_seq),
    epochs=50,
    batch_size=32,
    callbacks=[early_stopping],
)
```

模型建構—楊承翰

利用前15日的資料，
預測出隔日的空氣汙染物指標後，
再利用AQI公式計算隔日的AQI。

```
def compute_predicted_aqi(row):
    so2 = row["predicted_next_so2"]
    co = row["predicted_next_co"]
    o3_8hr = row["predicted_next_o3_8hr"]
    o3_1hr = row["predicted_next_o3"]
    pm10 = row["predicted_next_pm10"]
    pm25 = row["predicted_next_pm2.5"]
    no2 = row["predicted_next_no2"]

    o3_8hr_i = calc_sub_index("o3_8hr", o3_8hr)
    o3_1hr_i = calc_sub_index("o3_1hr", o3_1hr)
    o3_i = max(o3_8hr_i, o3_1hr_i)

    pm25_i = calc_sub_index("pm2.5", pm25)
    pm10_i = calc_sub_index("pm10", pm10)
    co_i = calc_sub_index("co", co)
    so2_i = calc_sub_index("so2", so2)
    no2_i = calc_sub_index("no2", no2)

    aqi = max(o3_i, pm25_i, pm10_i, co_i, so2_i, no2_i)
    return aqi
```

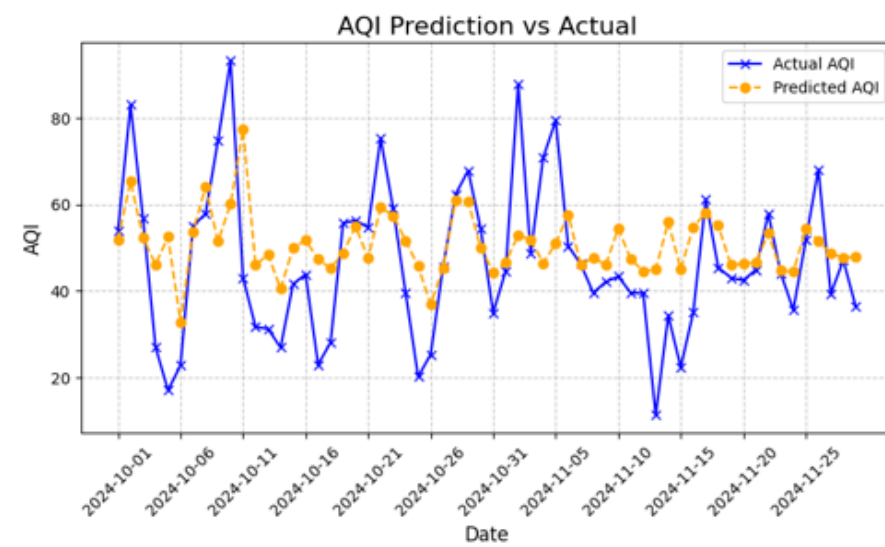
```
breakpoints = {
    "o3_8hr": [
        (0.000, 0.054, 0, 50),
        (0.055, 0.070, 51, 100),
        (0.071, 0.085, 101, 150),
        (0.086, 0.105, 151, 200),
        (0.106, 0.200, 201, 300),
    ],
    "o3_1hr": [
        (0.125, 0.164, 101, 150),
        (0.165, 0.204, 151, 200),
        (0.205, 0.404, 201, 300),
        (0.405, 0.504, 301, 400),
        (0.505, 0.604, 401, 500),
    ],
    "pm2.5": [
        (0.0, 15.4, 0, 50),
        (15.5, 35.4, 51, 100),
        (35.5, 54.4, 101, 150),
        (54.5, 150.4, 151, 200),
        (150.5, 250.4, 201, 300),
        (250.5, 350.4, 301, 400),
        (350.5, 500.4, 401, 500),
    ],
    "pm10": [
        (0, 50, 0, 50),
        (50, 100, 51, 100),
        (101, 254, 101, 150),
        (255, 354, 151, 200),
        (355, 424, 201, 300),
        (425, 504, 301, 400),
        (505, 604, 401, 500),
    ],
}
```

```
"co": [
    (0.0, 4.4, 0, 50),
    (4.5, 9.4, 51, 100),
    (9.5, 12.4, 101, 150),
    (12.5, 15.4, 151, 200),
    (15.5, 30.4, 201, 300),
    (30.5, 40.4, 301, 400),
    (40.5, 50.4, 401, 500),
],
"so2": [
    (0, 20, 0, 50),
    (21, 75, 51, 100),
    (76, 185, 101, 150),
    (186, 304, 151, 200),
    (305, 604, 201, 300),
    (605, 804, 301, 400),
    (805, 1004, 401, 500),
],
"no2": [
    (0, 30, 0, 50),
    (31, 100, 51, 100),
    (101, 360, 101, 150),
    (361, 649, 151, 200),
    (650, 1249, 201, 300),
    (1250, 1649, 301, 400),
    (1650, 2049, 401, 500),
],
}
```

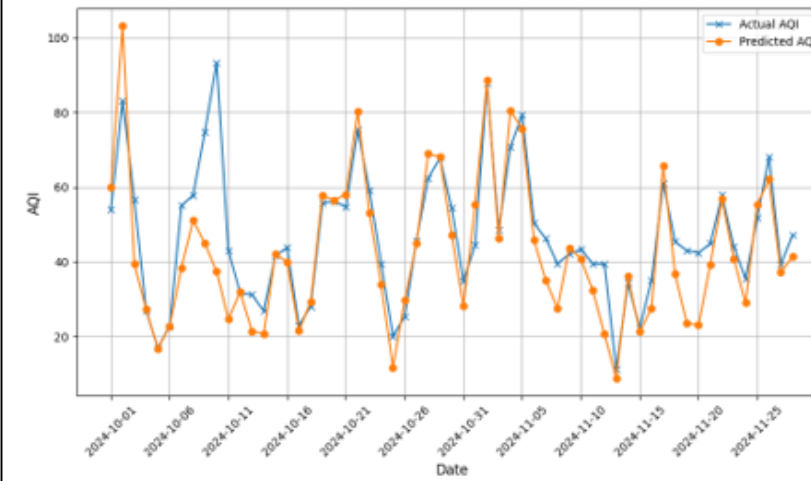
結果比較

萬里

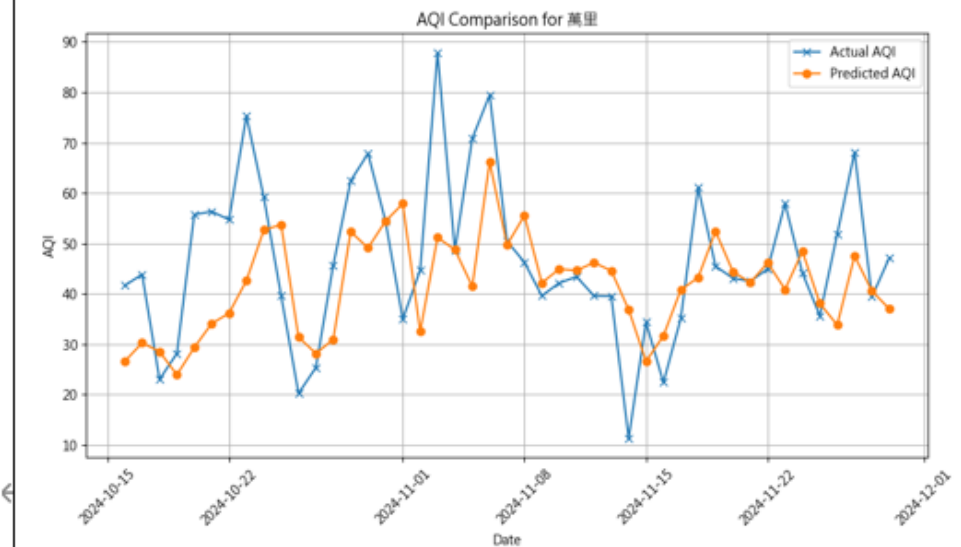
江曉明



陳柏淮

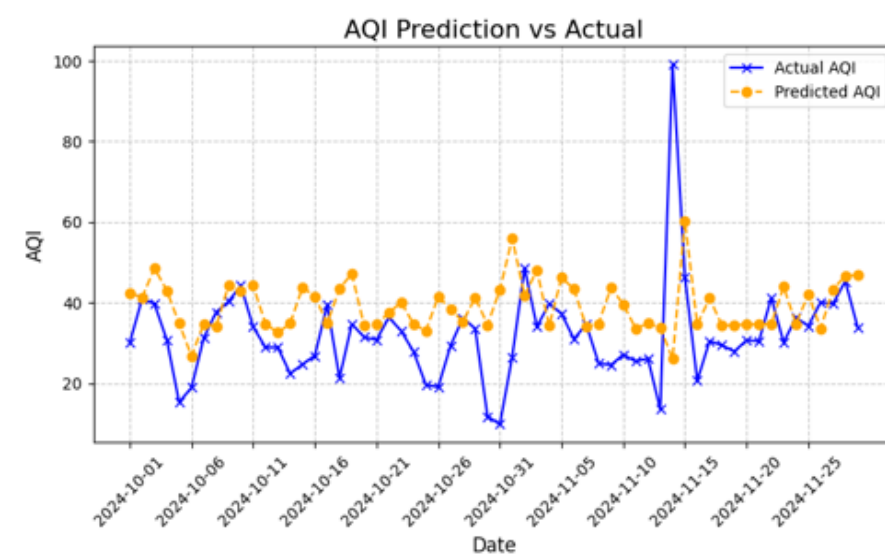


楊承翰

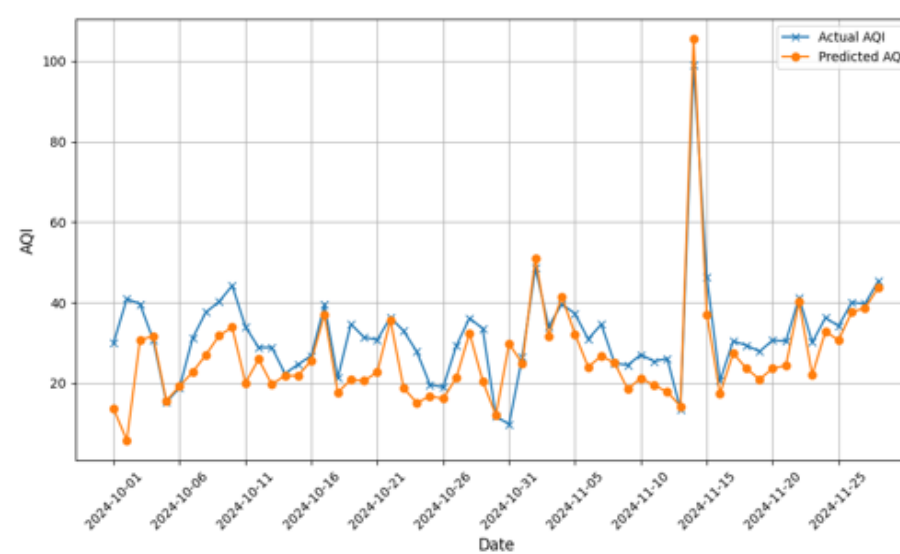


萬華

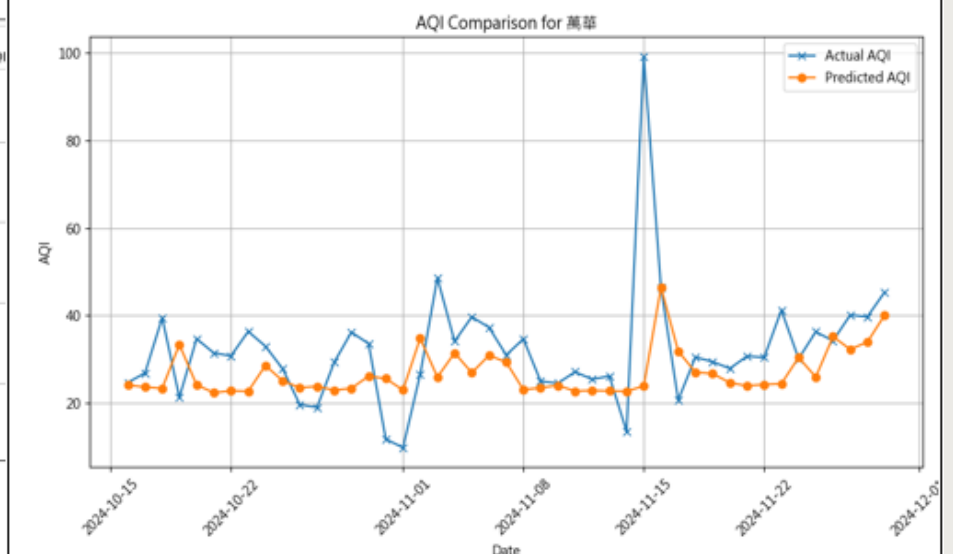
江曉明



陳柏淮



楊承翰



結果分析

為何陳柏淮的模型效果較好？

原因一
以縣市為單位訓練模型

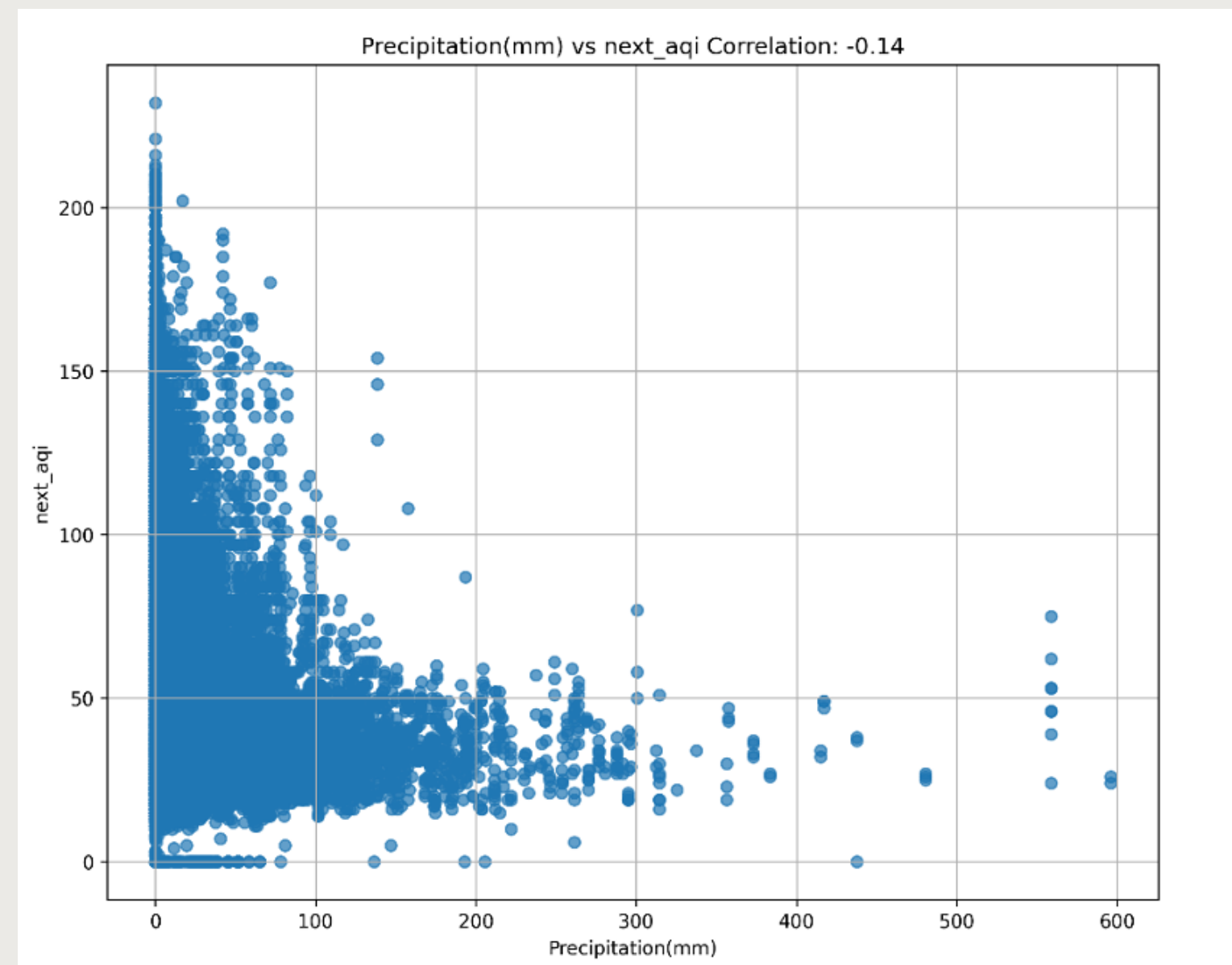
原因二
採用先預測空氣汙染物指標再計算
隔日AQI的方式

即時空氣品質指標 (即時 AQI)

為提供預警參考，本部發布即時空氣品質指標 (AQI) 值，並提供每小時監測數據，而即時空氣品質指標 (AQI) 計算方式如下：
各測項即時濃度依下列公式計算後，再對應下表得出 O_3 、 $PM_{2.5}$ 、 PM_{10} 、 CO 、 SO_2 、 NO_2 等 6 個測項之即時副指標值，再取出其中最大值為即時空氣品質指標，該最大值測項即為指標污染物：

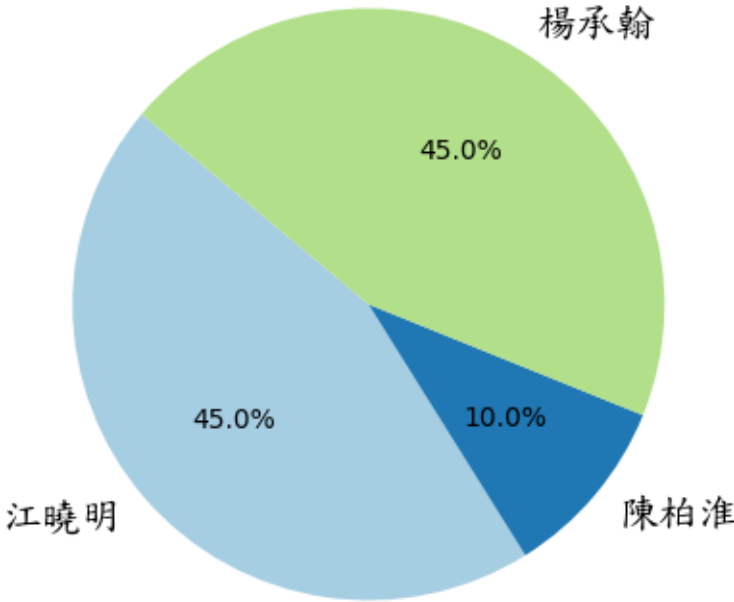
總結

- 雨量對空氣品質的影響：
我們發現雨量確實對空氣品質有正面影響，且降雨量越多，改善效果越明顯。但將雨量作為特徵預測空氣品質時，其效果並不突出。
- 預測空氣品質指標的最佳方法

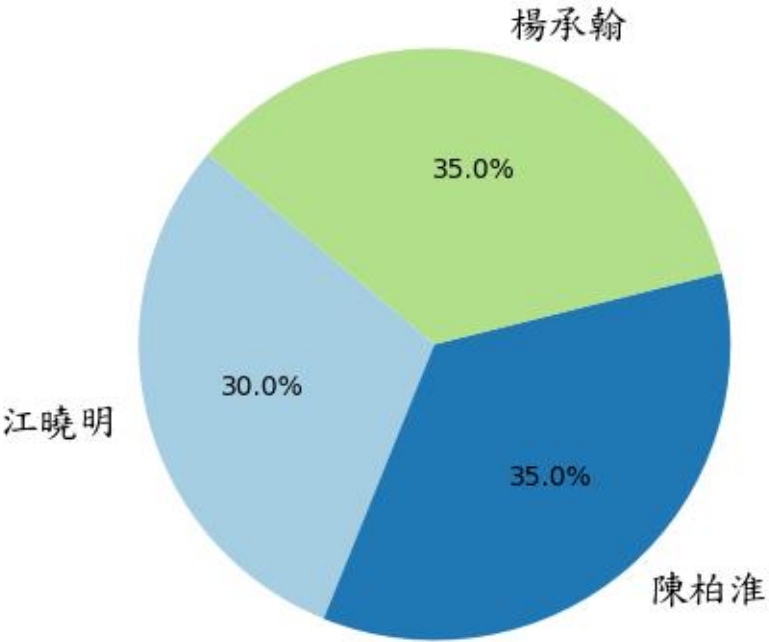


分工

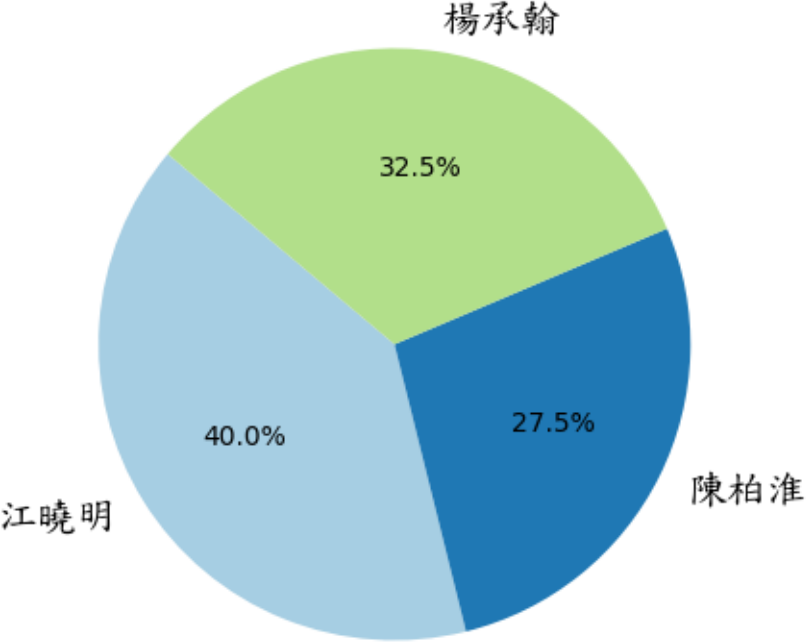
資料蒐集與處理

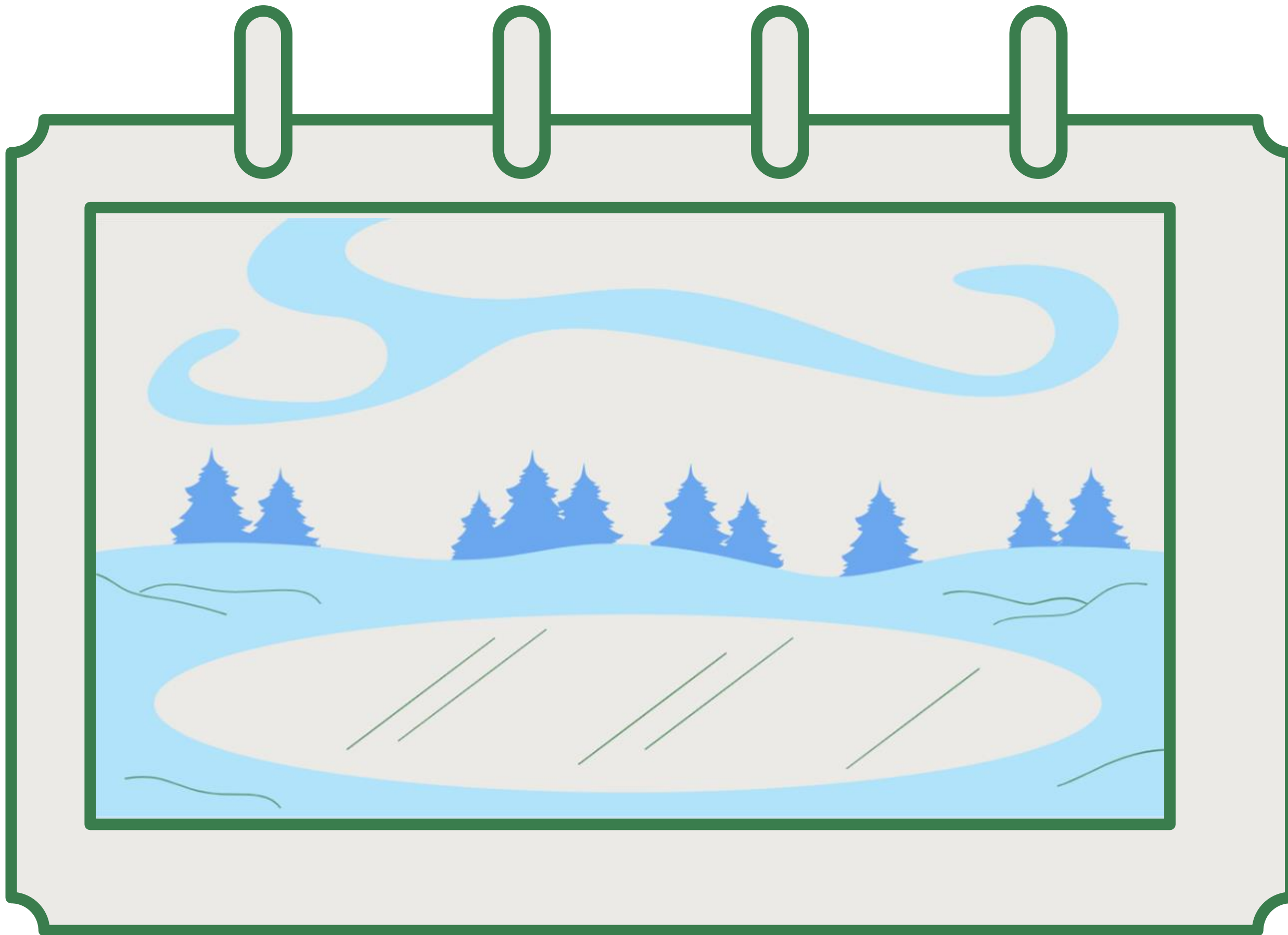


程式撰寫



報告撰寫





Thank you

Write a closing statement
or call-to-action here.