

影響空氣品質的因素之分析

Group 10

B10601104 許喬雅

B10707032 楊政憲

B10715029 陳彥瑋

B10815034 吳承翰

前情提要

前情提要

自變量

天氣（氣溫、降水量、相對溼度、氣壓、風速）、
火力發電發電量、交通量、工業區數目與面積（新增）

因變量

AQI（空氣品質指數）

分析資料

全台22縣市2010~2018共9年每月的資料

分析方法

數據分析：找出相關性
機器學習：建立預測模型

可能的結果

1. | 火力發電與AQI關聯不大：需考慮**污染物排放量**&**擴散**因素
2. | 汙染源大多來自**境內**，而**電力**帶來的汙染只佔其中的一**小部分**
3. | 空氣汙染有**季節性**：多集中在10~3月

分析過程

資料採集、整理

資料來源

天氣 → 交通部中央氣象局 觀測資訊查詢系統

火力發電發電量 → 臺灣電力公司 過去電力供需資訊

交通量 → 交通部公路總局 公路交通量調查統計表

工業區面積 → 台灣工業用地供給與服務資訊網 工業區統計

採集方法

網路查詢

Python爬蟲，使用函式庫：selenium(爬取資料)、os(檔案分類)...

後續處理

挑選可用資料，整理為各縣市每月資料
透過Python處理

資料分析 — AQI & 天氣

總體相關係數

- 1. AQI與天氣關聯性不大
- 2. 降水量→低度負相關
- 3. 氣壓與相對溼度→高度正相關

	AQI大於100 之比率(%)	氣溫	降水量	相對 溼度	氣壓	風速
AQI大於100 之比率(%)	1					
氣溫	-0.0814	1				
降水量	-0.2065	0.2814	1			
相對溼度	-0.0208	0.0454	0.2103	1		
氣壓	0.0473	0.0951	0.0127	0.7901	1	
風速	0.0292	-0.0946	-0.1514	-0.0263	0.1212	1

氣壓→中度負相關

台中市	AQI大於100之比率(%)	氣溫	降水量	相對溼度	氣壓	風速
AQI大於100之比率(%)	1					
氣溫	-0.1295	1				
降水量	-0.1609	0.4495	1			
相對溼度	-0.1785	0.0853	0.5175	1		
氣壓	-0.5838	-0.1201	-0.1167	-0.2114	1	
風速	-0.0014	-0.5293	-0.3298	-0.4165	0.3359	1

風速→中度正相關

金門縣	AQI大於100之比率(%)	氣溫	降水量	相對溼度	氣壓	風速
AQI大於100之比率(%)	1					
氣溫	-0.2277	1				
降水量	-0.2763	0.2615	1			
相對溼度	-0.2534	0.4246	0.5116	1		
氣壓	0.1898	-0.9210	-0.3801	-0.6486	1	
風速	0.6108	-0.2609	-0.3352	-0.3581	0.3114	1

相對溼度→中度負相關

嘉義市	AQI大於100之比率(%)	氣溫	降水量	相對溼度	氣壓	風速
AQI大於100之比率(%)	1					
氣溫	-0.2152	1				
降水量	-0.2463	0.5255	1			
相對溼度	-0.4166	-0.0276	0.3222	1		
氣壓	0.2202	-0.9334	-0.6661	-0.0836	1	
風速	-0.3753	-0.2032	0.1118	0.0819	0.0803	1

風速→中度正相關

連江縣	AQI大於100之比率(%)	氣溫	降水量	相對溼度	氣壓	風速
AQI大於100之比率(%)	1					
氣溫	-0.0972	1				
降水量	-0.1910	0.0905	1			
相對溼度	-0.2434	0.2486	0.5075	1		
氣壓	0.3518	-0.8322	-0.3249	-0.4229	1	
風速	0.4025	-0.0133	-0.3047	-0.2141	0.3406	1

總體相關係數

- 1. 只取設有火力發電廠的10個縣市之資料
- 2. AQI與火力發電量只有低度正相關性

	AQI大於100 之比率(%)	發電量
AQI大於100 之比率(%)	1	
發電量	0.1152	1

資料分析 — AQI & 火力發電量

縣市相關係數

- 1. 苗栗縣→中度正相關
- 2. 台中市→低度正相關
- 3. 其他縣市接近無相關

苗栗縣	AQI大於100 之比率(%)	發電量
AQI大於100 之比率(%)	1	
發電量	0.4829	1

台中市	AQI大於100 之比率(%)	發電量
AQI大於100 之比率(%)	1	
發電量	0.2801	1

資料分析—AQI & 交通量

總體相關係數

- 1. 以各縣市9年之**每日平均車流量**評估(非每月資料)
- 2. 各年度之AQI與每日平均車流量均為**低度正相關**
- 3. 臺北市、金門縣、連江縣資料缺失

備註：
A → 2010-2012每月平均AQI大於100比例(%)
B → 2013-2015每月平均AQI大於100比例(%)
C → 2016-2018每月平均AQI大於100比例(%)

	A	B	C	每日平均車流量
A	1			
B	0.9727	1		
C	0.8300	0.8306	1	
每日平均車流量	0.2761	0.2222	0.1567	1

資料分析 — AQI & 工業區數目與面積

總體相關係數

- 1. 以各縣市當前工業區數目、面積及佔各縣市面積比例評估
- 2. 工業區面積與2016-2018之每月平均AQI為低度正相關

備註：
A → 2010-2012每月平均AQI大於100比例(%)
B → 2013-2015每月平均AQI大於100比例(%)
C → 2016-2018每月平均AQI大於100比例(%)

	A	B	C	工業區數目	面積(平米)	工業區佔各縣市面積比例(%)
A	1					
B	0.8362	1				
C	0.7216	0.8259	1			
工業區數目	0.1749	0.0088	-0.0324	1		
面積(平米)	0.1583	0.1238	0.3374	0.2334	1	
工業區佔各縣市面積比例(%)	-0.0014	-0.0251	0.1990	0.1045	0.9131	1

資料分析 — AQI & 地理區域

相關矩陣

1. 北部地區→與鄰近縣市有高度正相關

備註：
>=0.7：高度相關
>=0.85：特別相關

北部地區	基隆市	臺北市	新北市	宜蘭縣	桃園市	新竹縣	新竹市
基隆市	1						
臺北市	0.8063	1					
新北市	0.9009	0.9264	1				
宜蘭縣	0.8061	0.7900	0.7993	1			
桃園市	0.8139	0.8113	0.8843	0.7615	1		
新竹縣	0.7878	0.7642	0.8309	0.7587	0.9634	1	
新竹市	0.6189	0.6110	0.6983	0.5813	0.8956	0.8986	1

資料分析 — A Q I & 地理區域

相關矩陣

- 1. 中部地區→與鄰近縣市有高度正相關
- 2. 東部地區→與鄰近縣市低度相關

備註：
>=0.7：高度相關
>=0.85：特別相關

中部地區	苗栗縣	臺中市	彰化縣	南投縣	雲林縣
苗栗縣	1				
臺中市	0.9262	1			
彰化縣	0.8934	0.9287	1		
南投縣	0.7772	0.8752	0.8988	1	
雲林縣	0.8415	0.9250	0.9444	0.9685	1

東部地區	花蓮縣	臺東縣
花蓮縣	1	
臺東縣	0.2161	1

資料分析 — A Q I & 地理區域

相關矩陣

- 1. 南部地區→除澎湖縣外，其他地區與鄰近縣市特別正相關
- 2. 離島地區→與鄰近縣市特別正相關

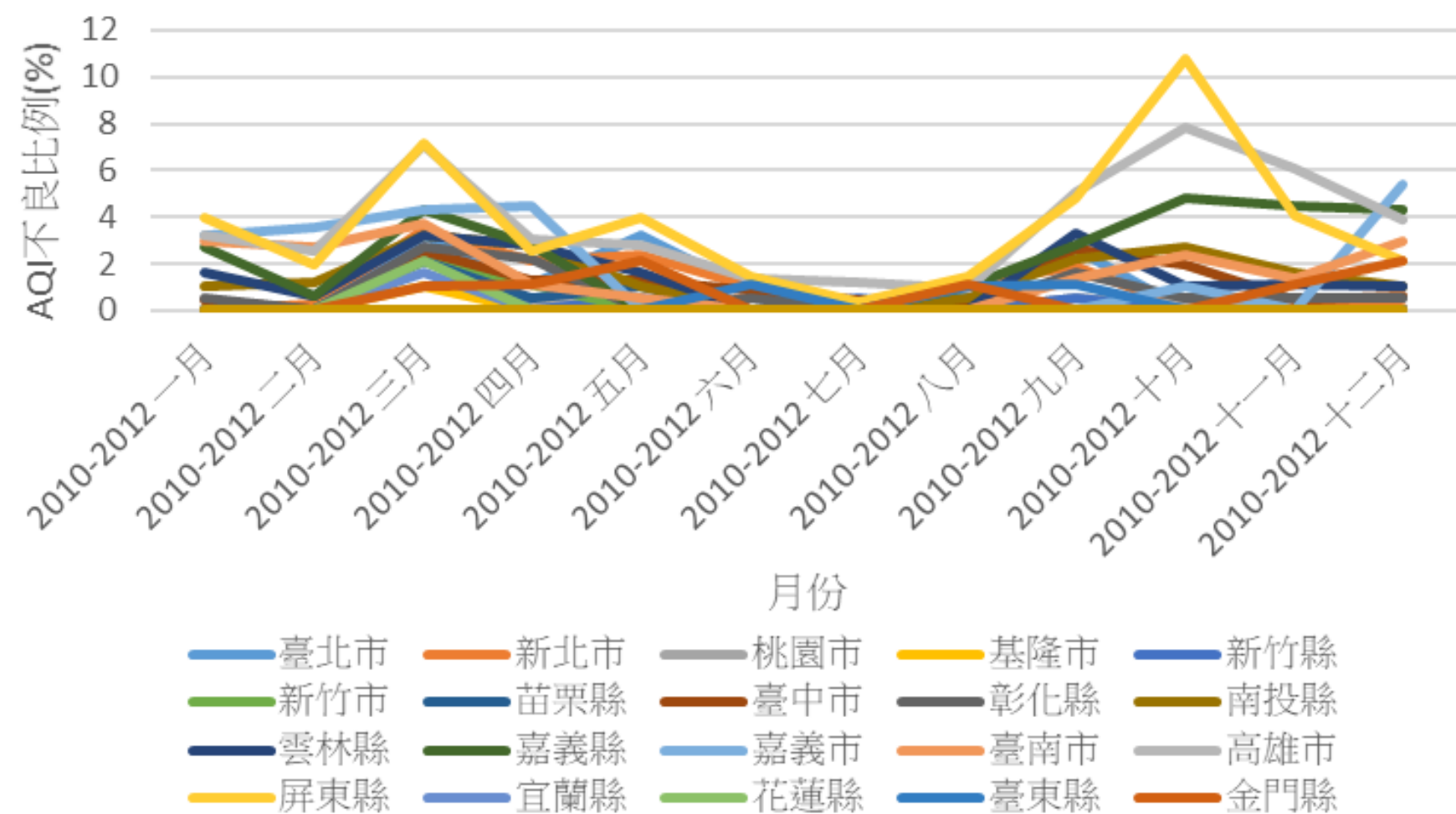
備註：
>=0.7：高度相關
>=0.85：特別相關

南部地區	嘉義縣	嘉義市	臺南市	高雄市	屏東縣	澎湖縣
嘉義縣	1					
嘉義市	0.9400	1				
臺南市	0.9418	0.9836	1			
高雄市	0.8925	0.9412	0.9616	1		
屏東縣	0.8546	0.9090	0.9328	0.9754	1	
澎湖縣	0.6417	0.6044	0.6318	0.5712	0.5526	1

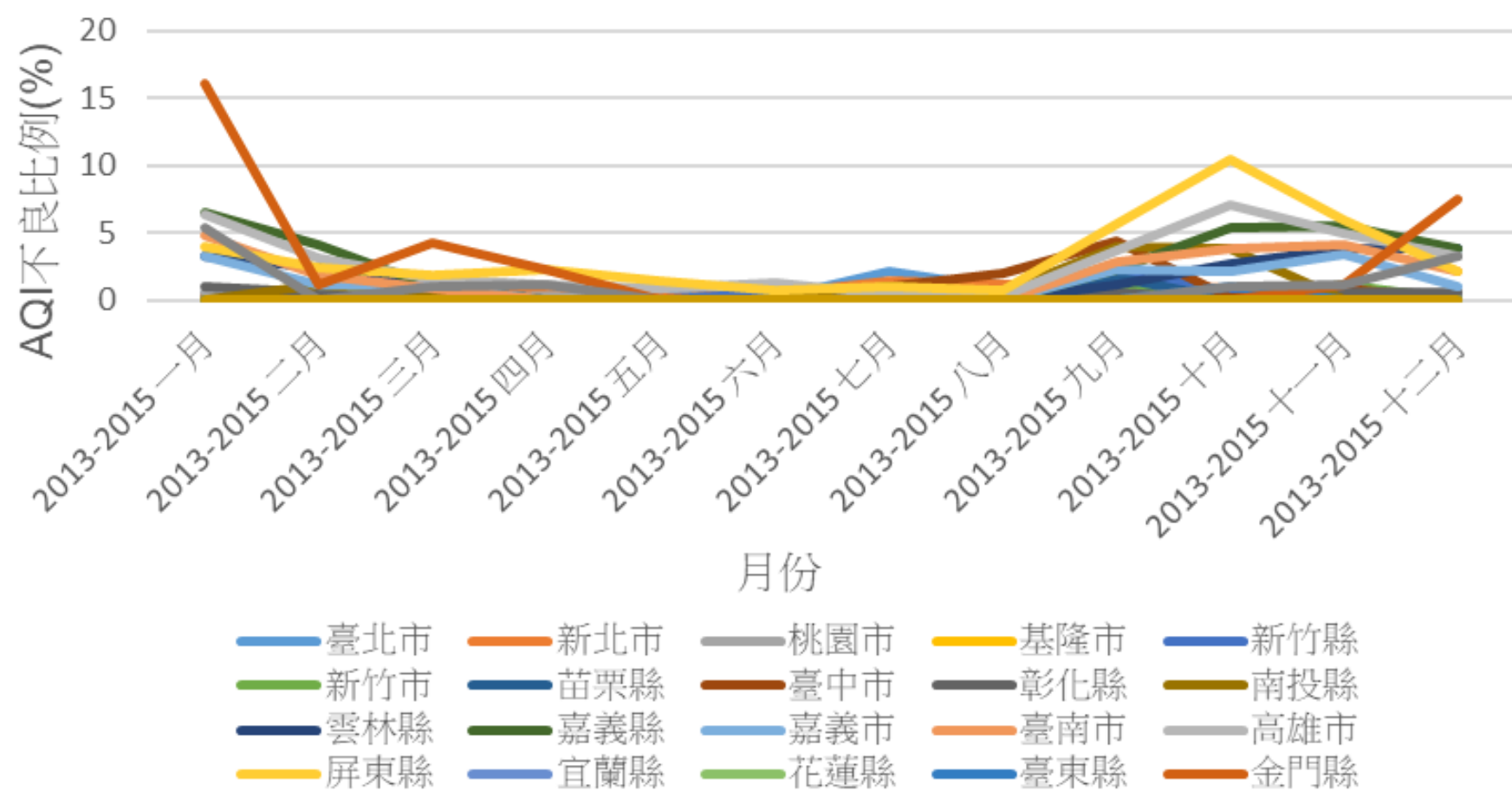
離島地區	金門縣	連江縣
金門縣	1	
連江縣	0.9167	1

資料分析—AQI & 月份

2010-2012各縣市每月平均AQI折線圖

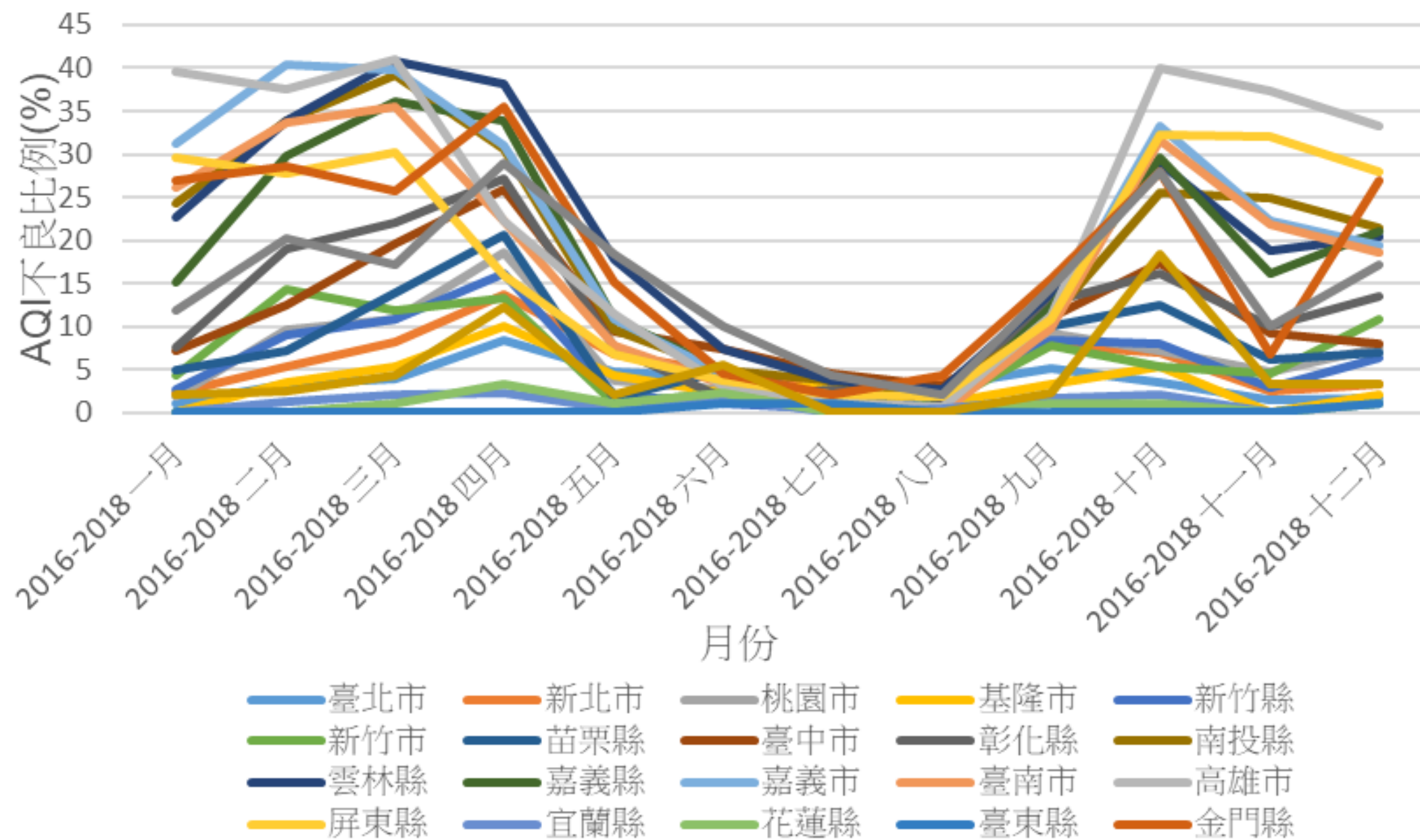


2013-2015各縣市每月平均AQI折線圖



資料分析—AQI & 月份

2016-2018各縣市每月平均AQI折線圖



兩者關係

1. 各年度之AQI不良比例→
10月~隔年4月較高，
6月~8月較低

機器學習

機器學習

學習內容

AQI大於100的天數比例，以是否大於0%來為分界點

使用方法

邏輯迴歸 (Logistic Regression)

觀察內容

準確度 (Accuracy)

召回率 (Recall)

機器學習 — 邏輯迴歸

1. 可將資料進行「二元分類」
利用邏輯迴歸，將「AQI大於100的天數」進行「二元分類」
2. 將比率大於0的標示為「1」，比率小於0的標示為「0」
1代表空氣比較不好，0代表空氣比較好

機器學習—邏輯迴歸

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m \text{Cost}(h_{\theta}(x^{(i)}), y^{(i)})$$

$$\text{Cost}(h_{\theta}(x), y) = \begin{cases} -\log(h_{\theta}(x)) & \text{if } y = 1 \\ -\log(1 - h_{\theta}(x)) & \text{if } y = 0 \end{cases}$$

$$h(\theta) = \frac{1}{1 + e^{-\theta^T x}}$$

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta) \quad \Rightarrow \quad \theta_j := \theta_j - \alpha \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

機器學習 — 混淆矩陣

使用混淆矩陣 (Confusion Matrix) 的召回率 (Recall) 進行分析

預測情況 \ 實際情況	True (1)	False (0)
True (1)	TP(True Positive)	FP(False Positive)
False (0)	FN(False Negative)	TN(True Negative)

機器學習 — 召回率 (Recall)

1.
$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$
$$= (\text{預測是真, 實際也是真}) / (\text{所有真實發生真的情況})$$

2. 因為台灣一整個月天氣好的時候約為62% (62%都是0)
所以我們想盡可能的預測天氣壞的情況

機器學習 — 結果

1.

準確度 (Accuracy) : 65.99%

召回率 (Recall) : 33.04%

2.

準確度以二元分類來講，不算高

召回率更是不太準

由此可知無法用前面選擇的資料製作精準的模型

分析結果

天氣

1. 總體而言，AQI與天氣幾乎沒有關聯，只有與降水量比較有相關性（相關係數： -0.2065 ）
2. 就區域來說，各縣市與個別量測值有相關性，如：
 - 台中市與氣壓（相關係數： -0.5838 ）
 - 嘉義市與相對溼度（相關係數： -0.4166 ）
 - 金門縣與風速（相關係數： 0.6108 ）
 - 連江縣與風速（相關係數： 0.4025 ）

火力發電量

1. 總體而言，AQI與火力發電量幾乎沒有關聯（相關係數：0.1152）
2. 就區域來說，有些縣市有較高的相關性，如：
苗栗縣→中度正相關（相關係數：**0.4829**）
台中市→低度正相關（相關係數：**0.2801**）

交通量

1. 總體來說，AQI與交通量關聯不大（相關係數：0.2761、0.2222、0.1567）

工業區數目與面積

1. 總體而言，AQI與工業區數目幾乎沒有關聯（相關係數： $-0.03 \sim 0.17$ ）
2. AQI與工業區面積關聯性不大（相關係數： $0.12 \sim 0.33$ ）

地理區域

1. 鄰近的區域在相同的時間點對於AQI的值會有非常高度的相關，但東部地區卻呈現低度相關（相關係數： 0.2161 ）

月份

1. AQI與月份有顯著的關係，在大約10月~隔年4月會有較高的AQI不良比例，而在6月~8月時比例較低

分析結論

各縣市之AQI受不同因素影響



與預期結果比較

1. 依各地區氣候地形條件的不同，對於AQI會有不同的影響。例如：金門&風速
2. AQI與火力發電量確實沒什麼關聯，與預期結果相符
3. AQI受季節因素的影響，與預期結果相符

預期結果

1. 火力發電與AQI關聯不大：需考慮污染物排放量&擴散因素
2. 汙染源大多來自境內，而電力帶來的汙染只佔其中的一小部分
3. 空氣汙染有季節性：多集中在10~3月

其他發現

1. AQI與交通量關聯性不大
2. AQI與工業區數量、面積看起來也沒有關聯性
3. 相近的區域可能因為有類似的地形氣候，因此會有較高的關聯性；
至於東部地區由於地形較為複雜，所以與附近縣市有不一樣的結果