In [1]:
```
import pandas as pd
import numpy as np
import time
pd.options.display.max_rows = 100
```

Q1

In [2]:
```
filename = "iis_airbox_20200920.csv"
df = pd.read_csv(filename)
df.head()
```

Out[2]:

	device_id	SiteName	PM25	timestamp
0	08BEAC09FF5C	臺中市立長億高中(2019)	15	2020-09-20 00:00:00
1	74DA38C7D096	嘉義縣中埔國民小學(2017)	0	2020-09-20 00:00:00
2	08BEAC0A0842	桃園市市立羅浮國小(2019)	8	2020-09-20 00:00:00
3	08BEAC0A03C0	臺南市市立德高國小(2019)	16	2020-09-20 00:00:00
4	08BEAC0A01F4	臺南市市立重溪國小(2019)	36	2020-09-20 00:00:00

Q2取得臺中市測站資料

In [3]:
```
#1 利用pd.Series.str.contains()找出'臺中市' 是否包含在 df['SiteName'] 的每個元素裡面 df['SiteName'] 是 pd.Series 型態，所以在後面加上.str.co
mask1 = df['SiteName'].str.contains('臺中市')

#2 利用apply找出'臺中市' 是否包含在 df['SiteName'] 的每個元素裡面
mask2 = df['SiteName'].apply(lambda x: '臺中市' in x)


print("兩個判斷是否相等?：",(mask1==mask2).all())
```

```
兩個判斷是否相等?： True
```

In [4]:
```
mask1
```

Out[4]:
```
0          True
1         False
2         False
3         False
4         False
          ...
366259    False
366260    False
366261     True
366262     True
366263    False
Name: SiteName, Length: 366264, dtype: bool
```

In [5]:
```
df_tc = df[mask1]
print(df_tc)
```

```
           device_id          SiteName  PM25            timestamp
0       08BEAC09FF5C    臺中市立長億高中(2019)    15  2020-09-20 00:00:00
12      08BEAC0A04CC    臺中市立逢甲國小(2019)    14  2020-09-20 00:00:02
18      08BEAC0A0096    臺中市立成功國小(2019)    18  2020-09-20 00:00:03
32      08BEAC0A0162    臺中市立鹿峰國小(2019)    13  2020-09-20 00:00:05
38      08BEAC09FF88  臺中市立臺中啟聰學校(2019)    23  2020-09-20 00:00:06
...              ...                 ...   ...                  ...
366229  08BEAC0A005E    臺中市立梧南國小(2019)     9  2020-09-20 23:59:56
366244  08BEAC0A0058    臺中市立育英國中(2019)     8  2020-09-20 23:59:58
366251  08BEAC0A01D2    臺中市立大楊國小(2019)    14  2020-09-20 23:59:59
366261  08BEAC09FFDA    臺中市立崇德國中(2019)     5  2020-09-21 00:00:00
366262  08BEAC0A007A    臺中市立省三國小(2019)     7  2020-09-21 00:00:00

[61049 rows x 4 columns]
```

Q3.計算NaN數量

轉成wide table，並利用apply計算NaN數量

In [6]:
```
df_tc_wide = pd.pivot_table(df_tc, values = 'PM25', index = 'SiteName', columns = 'timestamp')

# 你也可以這樣寫
# df_tc_wide = df_tc.pivot_table(values = 'PM25', index = 'SiteName', columns = 'timestamp')
```

In [7]:
```
df_tc_wide
```

Out[7]:

timestamp	2020-09-20 00:00:00	2020-09-20 00:00:02	2020-09-20 00:00:03	2020-09-20 00:00:05	2020-09-20 00:00:06	2020-09-20 00:00:08	2020-09-20 00:00:09	2020-09-20 00:00:12	2020-09-20 00:00:15	2020-09-20 00:00:16	...	2020-09-20 23:59:47	2020-09-20 23:59:48	2020-09-20 23:59:49	2020-09-20 23:59:50	2020-09-20 23:59:51
SiteName																
臺中市立三光國中(2019)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN
臺中市立三光國小(2019)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN
臺中市立三和國小(2019)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN
臺中市立三田國小(2019)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN

In [8]:
```
# 使用 isnull()
df_tc_wide.isnull().apply(np.sum, axis = 1)
```

Out[8]:
```
SiteName
臺中市立三光國中(2019)    41497
臺中市立三光國小(2019)    41504
臺中市立三和國小(2019)    41510
臺中市立三田國小(2019)    41500
臺中市立上安國小(2019)    41498
                  ...
臺中市立龍泉國小(2019)    41609
臺中市立龍津國小(2019)    41504
臺中市立龍津高中(2019)    41499
臺中市立龍海國小(2019)    41503
臺中市立龍港國小(2019)    41501
Length: 272, dtype: int64
```

In [9]:
```
# 使用 dropna()
df_tc_wide.apply(lambda x:len(x)-len(x.dropna()), axis = 1)
```

Out[9]:
```
SiteName
臺中市立三光國中(2019)    41497
臺中市立三光國小(2019)    41504
臺中市立三和國小(2019)    41510
臺中市立三田國小(2019)    41500
臺中市立上安國小(2019)    41498
                  ...
臺中市立龍泉國小(2019)    41609
臺中市立龍津國小(2019)    41504
臺中市立龍津高中(2019)    41499
臺中市立龍海國小(2019)    41503
臺中市立龍港國小(2019)    41501
Length: 272, dtype: int64
```

Q4計算每個測站的日平均pm2.5

In [10]:
```
df_tc_wide.apply(np.nanmean, axis = 1)
```

Out[10]:
```
SiteName
臺中市立三光國中(2019)     9.536481
臺中市立三光國小(2019)    10.305310
臺中市立三和國小(2019)    16.159091
臺中市立三田國小(2019)    11.282609
臺中市立上安國小(2019)    13.767241
                  ...
臺中市立龍泉國小(2019)    11.074380
臺中市立龍津國小(2019)    11.393805
臺中市立龍津高中(2019)    12.948052
臺中市立龍海國小(2019)    10.933921
臺中市立龍港國小(2019)    11.257642
Length: 272, dtype: float64
```

In [ ]:
```

```