Python 重組資料表的結構

資料表的結構重組

raw_data

NAME	PRODUCT	YEAR	PRICE	AMOUNT
BIG	A	2020	100	5
BIG	B	2020	200	6
BIG	C	2020	300	7
BIG	A	2021	400	8
BIG	B	2021	500	9
BIG	C	2021	600	10
MEDIEN	A	2020	700	11
MEDIEN	B	2020	800	12
		~		
SMALL	C	2020	1,500	
SMALL	B	2021	1,600	
SMALL	C	2021	1,700	
SMALL	A	2021	1,800	

reshaped_raw_data

	PRODUCT	A		B		C	
		PRICE	AMOUNT	PRICE	AMOUNT	PRICE	AMOUNT
NAME	YEAR						
BIG	2020	100	5	200	6	300	7
	2021	400	8	500	9	600	10
MEDIEN	2020	700	11	800	12	900	13
	2021	1000	14	1100	15	1200	16
SMALL	2020	1300	17	1400	18	1500	19
	2021	1800	22	1600	20	1700	21

Step 1 : 讀取 Excel 檔

pandas.read_excel( )

Step 2 : 設定 DataFrame 的索引(將欄位轉成索引)

pandas.DataFrame.set_index( )

Step 3 : 重新排列 DataFrame 資料表

pandas.DataFrame.stack( )

pandas.DataFrame.unstack( )

Step 4 : DataFrame 寫入 Excel 檔

pandas.DataFrame.to_excel( )

Step 1 : pandas.read_excel( )

In [2]:
```
xlsx_path = 'D:\RPA_UiPath\Python x RPA\Reshape the Data\Input\SampleData.xlsx'

raw_data = pd.read_excel( xlsx_path )
raw_data
```

Out[2]:

	NAME	PRODUCT	YEAR	PRICE	AMOUNT
0	BIG	A	2020	100	5
1	BIG	B	2020	200	6
2	BIG	C	2020	300	7
3	BIG	A	2021	400	8
4	BIG	B	2021	500	9
5	BIG	C	2021	600	10
6	MEDIEN	A	2020	700	11
7	MEDIEN	B	2020	800	12
8	MEDIEN	C	2020	900	13
9	MEDIEN	A	2021	1000	14

Step 2 : pandas.DataFrame.set_index( )

In [3]:
```
raw_data.set_index( ['NAME','YEAR','PRODUCT'], inplace=True )
raw_data
```

Out[3]:

			PRICE	AMOUNT
NAME	YEAR	PRODUCT		
BIG	2020	A	100	5
		B	200	6
		C	300	7
	2021	A	400	8
		B	500	9
		C	600	10
MEDIEN	2020	A	700	11
		B	800	12
		C	900	13
	2021	A	1000	14

Step 3 : pandas.DataFrame.stack( ) pandas.DataFrame.unstack( )

```
In [4]: reshaped_raw_data = raw_data.stack().unstack([2,3])
        reshaped_raw_data
```

Out[4]:

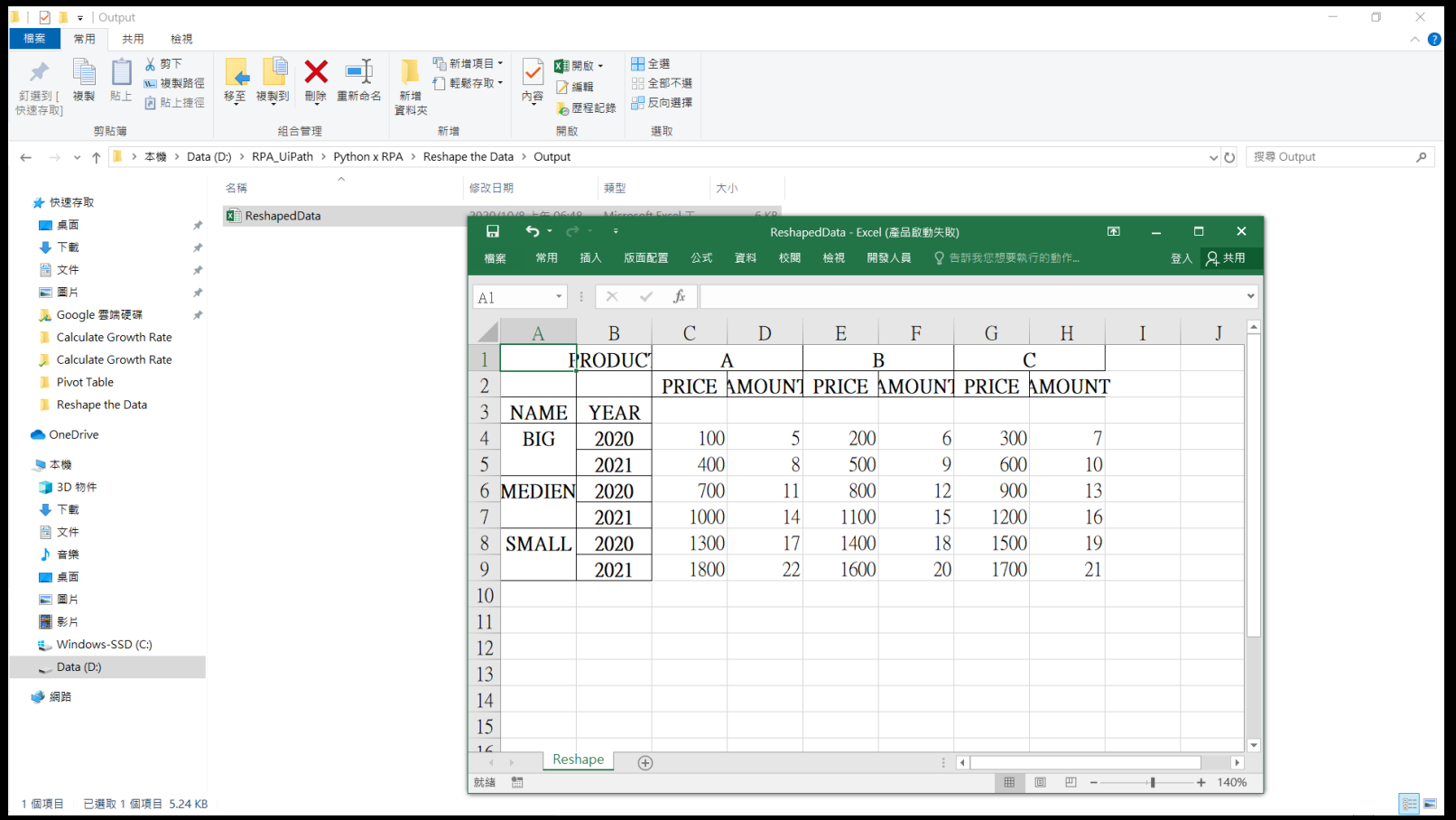
	PRODUCT	A		B		C	
		PRICE	AMOUNT	PRICE	AMOUNT	PRICE	AMOUNT
NAME	YEAR						
BIG	2020	100	5	200	6	300	7
	2021	400	8	500	9	600	10
MEDIEN	2020	700	11	800	12	900	13
	2021	1000	14	1100	15	1200	16
SMALL	2020	1300	17	1400	18	1500	19
	2021	1800	22	1600	20	1700	21

Step 4 : pandas.DataFrame.to_excel( )

```
In [5]: output_path = 'D:\RPA_UiPath\Python x RPA\Reshape the Data\Output\ReshapedData.xlsx'

        reshaped_raw_data.to_excel( output_path, sheet_name = 'Reshape' )
```

Output



	A	B	C	D	E	F	G	H
1		PRODUCT	A		B		C	
2			PRICE	AMOUNT	PRICE	AMOUNT	PRICE	AMOUNT
3	NAME	YEAR						
4	BIG	2020	100	5	200	6	300	7
5		2021	400	8	500	9	600	10
6	MEDIEN	2020	700	11	800	12	900	13
7		2021	1000	14	1100	15	1200	16
8	SMALL	2020	1300	17	1400	18	1500	19
9		2021	1800	22	1600	20	1700	21