

```
In [1]: import csv
        from string import ascii_uppercase
        import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
```

```
In [2]: data_report = pd.read_excel('總統-A05-4-候選人得票數一覽表-各投開票所(新北市).xls', skiprows=[0, 1, 3, 4], thousands=',')
        data_report
```

Out[2]:

	Unnamed: 0	Unnamed: 1	Unnamed: 2	(1)\n宋楚瑜\n余湘	(2)\n韓國瑜\n張善政	(3)\n蔡英文\n賴清德	Unnamed: 6	Unnamed: 7	Unnamed: 8	Unnamed: 9	Unnamed: 10	Unnamed: 11	Unnamed: 12	Unnamed: 13	
	0	總 計	NaN	NaN	112620	959631	1393936	2466187	28041	2494228	71	2494299	827160	3321459	75.0943
	1	板橋區	NaN	NaN	15719	123922	204134	343775	4051	347826	10	347836	107164	455000	76.4453
	2	NaN	留侯里	1124.0	58	430	557	1045	24	1069	0	1069	304	1373	77.8587
	3	NaN	流芳里	1125.0	50	345	534	929	6	935	0	935	311	1246	75.0401
	4	NaN	赤松里	1126.0	24	218	336	578	15	593	0	593	153	746	79.4906

	2584	NaN	烏來里	2300.0	38	363	438	839	5	844	0	844	324	1168	72.2603
	2585	NaN	烏來里	2301.0	47	454	304	805	9	814	0	814	330	1144	71.1538
	2586	NaN	孝義里	2302.0	3	63	112	178	4	182	0	182	51	233	78.1116
	2587	NaN	信賢里	2303.0	11	158	172	341	6	347	0	347	136	483	71.8427
	2588	NaN	福山里	2304.0	12	252	110	374	8	382	0	382	221	603	63.3499

2589 rows × 14 columns

```
In [3]: n_cols = data_report.columns.size
        n_candidates = n_cols - 11
        id_vars = ['town', 'village', 'office']
        candidates = list(data_report.columns[3:(3+n_candidates)])
        office_cols = list(ascii_uppercase[:8])
        col_names = id_vars + candidates + office_cols
        data_report.columns = col_names
```

```
In [4]: data_report
```

Out[4]:

	town	village	office	(1)\n宋楚瑜\n余湘	(2)\n韓國瑜\n張善政	(3)\n蔡英文\n賴清德	A	B	C	D	E	F	G	H	
	0	總 計	NaN	NaN	112620	959631	1393936	2466187	28041	2494228	71	2494299	827160	3321459	75.0943
	1	板橋區	NaN	NaN	15719	123922	204134	343775	4051	347826	10	347836	107164	455000	76.4453
	2	NaN	留侯里	1124.0	58	430	557	1045	24	1069	0	1069	304	1373	77.8587
	3	NaN	流芳里	1125.0	50	345	534	929	6	935	0	935	311	1246	75.0401
	4	NaN	赤松里	1126.0	24	218	336	578	15	593	0	593	153	746	79.4906

	2584	NaN	烏來里	2300.0	38	363	438	839	5	844	0	844	324	1168	72.2603
	2585	NaN	烏來里	2301.0	47	454	304	805	9	814	0	814	330	1144	71.1538
	2586	NaN	孝義里	2302.0	3	63	112	178	4	182	0	182	51	233	78.1116
	2587	NaN	信賢里	2303.0	11	158	172	341	6	347	0	347	136	483	71.8427
	2588	NaN	福山里	2304.0	12	252	110	374	8	382	0	382	221	603	63.3499

2589 rows × 14 columns

```
In [5]: filled_na = data_report.fillna(method='ffill')
        df = data_report.assign(town=filled_na)
        df
```

Out[5]:

	town	village	office	(1)\n宋楚瑜\n余湘	(2)\n韓國瑜\n張善政	(3)\n蔡英文\n賴清德	A	B	C	D	E	F	G	H	
	0	總 計	NaN	NaN	112620	959631	1393936	2466187	28041	2494228	71	2494299	827160	3321459	75.0943
	1	板橋區	NaN	NaN	15719	123922	204134	343775	4051	347826	10	347836	107164	455000	76.4453
	2	板橋區	留侯里	1124.0	58	430	557	1045	24	1069	0	1069	304	1373	77.8587
	3	板橋區	流芳里	1125.0	50	345	534	929	6	935	0	935	311	1246	75.0401
	4	板橋區	赤松里	1126.0	24	218	336	578	15	593	0	593	153	746	79.4906

	2584	烏來區	烏來里	2300.0	38	363	438	839	5	844	0	844	324	1168	72.2603
	2585	烏來區	烏來里	2301.0	47	454	304	805	9	814	0	814	330	1144	71.1538
	2586	烏來區	孝義里	2302.0	3	63	112	178	4	182	0	182	51	233	78.1116
	2587	烏來區	信賢里	2303.0	11	158	172	341	6	347	0	347	136	483	71.8427
	2588	烏來區	福山里	2304.0	12	252	110	374	8	382	0	382	221	603	63.3499

2589 rows × 14 columns

```
In [6]: df = df.dropna()
        df
```

Out[6]:

	town	village	office	(1)\n宋楚瑜\n余湘	(2)\n韓國瑜\n張善政	(3)\n蔡英文\n賴清德	A	B	C	D	E	F	G	H
2	板橋區	留侯里	1124.0	58	430	557	1045	24	1069	0	1069	304	1373	77.8587
3	板橋區	流芳里	1125.0	50	345	534	929	6	935	0	935	311	1246	75.0401
4	板橋區	赤松里	1126.0	24	218	336	578	15	593	0	593	153	746	79.4906
5	板橋區	黃石里	1127.0	35	248	423	706	13	719	0	719	230	949	75.7640
6	板橋區	挹秀里	1128.0	37	289	351	677	13	690	0	690	188	878	78.5877
...
2584	烏來區	烏來里	2300.0	38	363	438	839	5	844	0	844	324	1168	72.2603
2585	烏來區	烏來里	2301.0	47	454	304	805	9	814	0	814	330	1144	71.1538
2586	烏來區	孝義里	2302.0	3	63	112	178	4	182	0	182	51	233	78.1116
2587	烏來區	信賢里	2303.0	11	158	172	341	6	347	0	347	136	483	71.8427
2588	烏來區	福山里	2304.0	12	252	110	374	8	382	0	382	221	603	63.3499

2559 rows × 14 columns

In [7]:

```
df['town'].unique()  
strip_str = df['town'].str.replace('\u3000', '')  
df = df.assign(town=strip_str)  
df['town'].unique()
```

Out[7]: array(['板橋區', '三重區', '中和區', '永和區', '新莊區', '新店區', '樹林區', '鶯歌區', '三峽區',
 '淡水區', '汐止區', '瑞芳區', '土城區', '蘆洲區', '五股區', '泰山區', '林口區', '深坑區',
 '石碇區', '坪林區', '三芝區', '石門區', '八里區', '平溪區', '雙溪區', '貢寮區', '金山區',
 '萬里區', '烏來區'], dtype=object)

In [8]: df

Out[8]:

	town	village	office	(1)\n宋楚瑜\n余湘	(2)\n韓國瑜\n張善政	(3)\n蔡英文\n賴清德	A	B	C	D	E	F	G	H
2	板橋區	留侯里	1124.0	58	430	557	1045	24	1069	0	1069	304	1373	77.8587
3	板橋區	流芳里	1125.0	50	345	534	929	6	935	0	935	311	1246	75.0401
4	板橋區	赤松里	1126.0	24	218	336	578	15	593	0	593	153	746	79.4906
5	板橋區	黃石里	1127.0	35	248	423	706	13	719	0	719	230	949	75.7640
6	板橋區	挹秀里	1128.0	37	289	351	677	13	690	0	690	188	878	78.5877
...
2584	烏來區	烏來里	2300.0	38	363	438	839	5	844	0	844	324	1168	72.2603
2585	烏來區	烏來里	2301.0	47	454	304	805	9	814	0	814	330	1144	71.1538
2586	烏來區	孝義里	2302.0	3	63	112	178	4	182	0	182	51	233	78.1116
2587	烏來區	信賢里	2303.0	11	158	172	341	6	347	0	347	136	483	71.8427
2588	烏來區	福山里	2304.0	12	252	110	374	8	382	0	382	221	603	63.3499

2559 rows × 14 columns

In [9]:

```
df = df.drop(labels=office_cols, axis=1)  
df
```

Out[9]:

	town	village	office	(1)\n宋楚瑜\n余湘	(2)\n韓國瑜\n張善政	(3)\n蔡英文\n賴清德
2	板橋區	留侯里	1124.0	58	430	557
3	板橋區	流芳里	1125.0	50	345	534
4	板橋區	赤松里	1126.0	24	218	336
5	板橋區	黃石里	1127.0	35	248	423
6	板橋區	挹秀里	1128.0	37	289	351
...
2584	烏來區	烏來里	2300.0	38	363	438
2585	烏來區	烏來里	2301.0	47	454	304
2586	烏來區	孝義里	2302.0	3	63	112
2587	烏來區	信賢里	2303.0	11	158	172
2588	烏來區	福山里	2304.0	12	252	110

2559 rows × 6 columns

In [10]:

```
df = pd.melt(df, id_vars=id_vars, var_name='candidate_info', value_name='votes')  
df
```

Out[10]:

	town	village	office	candidate_info	votes
0	板橋區	留侯里	1124.0	(1)\n宋楚瑜\n余湘	58
1	板橋區	流芳里	1125.0	(1)\n宋楚瑜\n余湘	50
2	板橋區	赤松里	1126.0	(1)\n宋楚瑜\n余湘	24
3	板橋區	黃石里	1127.0	(1)\n宋楚瑜\n余湘	35
4	板橋區	挹秀里	1128.0	(1)\n宋楚瑜\n余湘	37
...
7672	烏來區	烏來里	2300.0	(3)\n蔡英文\n賴清德	438
7673	烏來區	烏來里	2301.0	(3)\n蔡英文\n賴清德	304
7674	烏來區	孝義里	2302.0	(3)\n蔡英文\n賴清德	112
7675	烏來區	信賢里	2303.0	(3)\n蔡英文\n賴清德	172
7676	烏來區	福山里	2304.0	(3)\n蔡英文\n賴清德	110

7677 rows × 5 columns