

新北市不動產實價登錄資料分析

N1066744 夜二技二甲 張孝如

研究目的：

欲以迴歸模型評估影響房價的各種特徵，並計算出各種顯著程度

Step1: 匯入新北市 107 年度實價登錄檔案資料集

```
In [96]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

%matplotlib inline
plt.style.use('ggplot')
plt.rcParams['font.family']='Microsoft JhengHei' #顯示中文
plt.rcParams['axes.unicode_minus']=False #正常顯示負號

pd.set_option("display.max_columns",30) #設定pandas最多顯示出30個欄位資訊

df = pd.read_csv('f_lvr_land_a.csv', encoding='utf8') #新北實價登錄，下載的檔案用筆記本另存 csv，編碼選
df.sample(3) #隨機前3筆
```

Out[96]:

鄉鎮市區	交易標的	土地區段位置	土地移轉面積	都市土地使用分區	非都市土地使用分區	非都市土地使用編定	交易年月日	交易筆數	移轉層數	總樓層數	建物型態	主要用途	主要建材	建築完成年月	建物移轉總面積	建物現況格局-房	建物現況格局-廳	建物現況格局-衛	建物現況格局-隔間
237	新莊區	房地(土地+建物)+車位	13.47	住	NaN	NaN	1071020	土地1建物1單位1	九層	十二層	住宅大樓(11層含以上有電梯)	住家用	鋼筋混凝土造	0940906	136.19	3	2	2	有

先將所有變數引入，看前幾筆資料是否引入無誤

```
In [59]: df2 = df[['鄉鎮市區', '交易標的', '土地區段位置建物區段門牌', '土地移轉總面積平方公尺', '都市土地使用分區',
               '非都市土地使用分區', '非都市土地使用編定', '交易年月日', '交易筆棟數', '移轉層次', '總樓層數',
               '建物型態', '主要用途', '主要建材', '建築完成年月', '建物移轉總面積平方公尺', '建物現況格局-房',
               '建物現況格局-廳', '建物現況格局-衛', '建物現況格局-隔間', '有無管理組織', '總價元', '單價元平方公尺',
               '車位類別', '車位移轉總面積平方公尺', '車位總價元', '備註', '編號'
               ]]
df2.head(3) # 看前3行資料
```

Out[59]:

	鄉鎮市區	交易標的	土地區段位置 建物區段門牌	土地移轉總面積 平方公尺	都市土地 使用分區	非都市土地 使用分區	非都市土地 使用編定	交易年月 日	交易筆棟 數	移轉層 次	總樓層 數	建物型 態	主要 用途	主
0	The villages and towns urban district	transaction sign	land sector position building sector house num...	land shifting total area square meter	the use zoning or compiles and checks	the non-metropolis land use district	non-metropolis land use	transaction year month and day	transaction pen number	shifting level	total floor number	building state	main use	t m.
1	土城區	房地(土地+建物)	新北市土城區裕生路83巷1~30號	21.28	住	NaN	NaN	1071107	土地2建物1車位0	四層，陽台	五層	公寓(5樓含以下無電梯)	住家用	網
2	板橋區	房地(土地+建物)	新北市板橋區忠孝路48巷4弄61~90號	26.2	住	NaN	NaN	1071015	土地1建物1車位0	五層	五層	公寓(5樓含以下無電梯)	住家用	網

```
In [69]: df.head(3)
```

Out[69]:

	鄉鎮市區	交易標的	土地區段位置 建物區段門牌	土地移轉總面積 平方公尺	都市土地 使用分區	非都市土地 使用分區	非都市土地 使用編定	交易年月 日	交易筆棟 數	移轉層 次	總樓層 數	建物型 態	主要 用途	主
0	The villages and towns urban district	transaction sign	land sector position building sector house num...	land shifting total area square meter	the use zoning or compiles and checks	the non-metropolis land use district	non-metropolis land use	transaction year month and day	transaction pen number	shifting level	total floor number	building state	main use	t m.
1	土城區	房地(土地+建物)	新北市土城區裕生路83巷1~30號	21.28	住	NaN	NaN	1071107	土地2建物1車位0	四層，陽台	五層	公寓(5樓含以下無電梯)	住家用	網
2	板橋區	房地(土地+建物)	新北市板橋區忠孝路48巷4弄61~90號	26.2	住	NaN	NaN	1071015	土地1建物1車位0	五層	五層	公寓(5樓含以下無電梯)	住家用	網

計算筆數

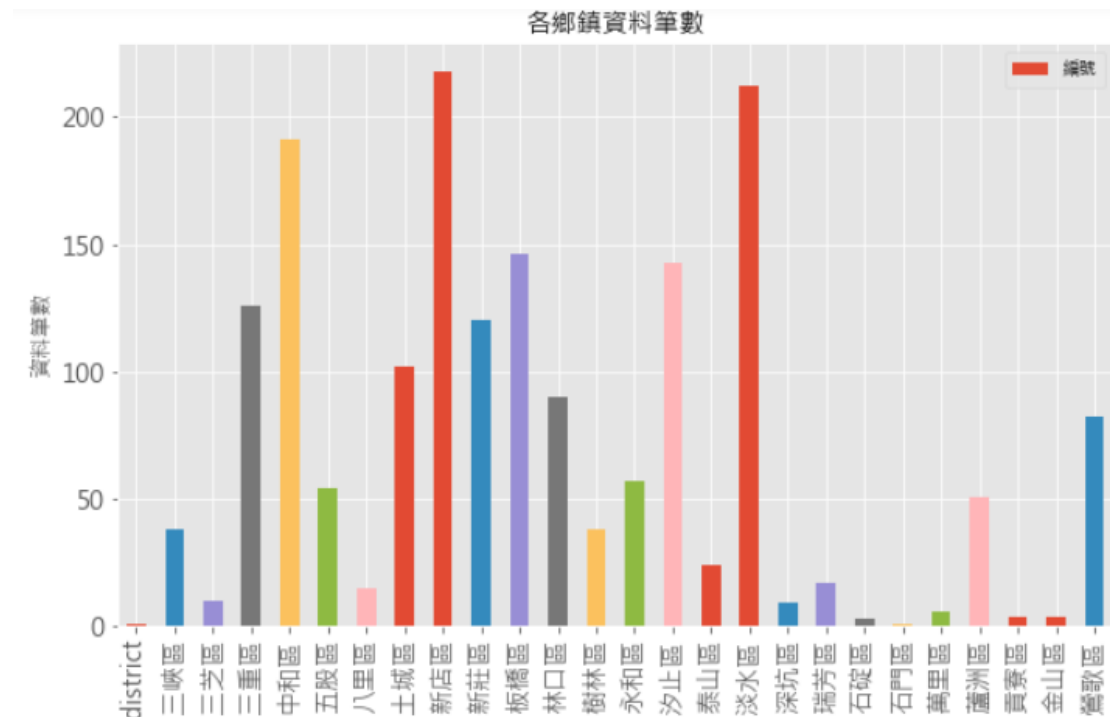
```
In [74]: df['鄉鎮市區'].count() # 計算鄉鎮市區筆數
```

Out[74]: 1762

圖示各鄉鎮資料筆數

```
In [75]: # 圖示各鄉鎮資料筆數
ax = df.groupby('鄉鎮市區').count().plot(kind='bar', y='編號', figsize=(10,6), fontsize=14, title='各鄉鎮資料筆數')
ax.set_ylabel('資料筆數')
```

Out[75]: Text(0,0.5,'資料筆數')



2. 原始資料集預處理

In [6]: `df.groupby('鄉鎮市區').mean()` # 以鄉鎮市區進行聚合，可查看所有鄉鎮市區資料

```

-----
DataError                                Traceback (most recent call last)
<ipython-input-6-162ddc862985> in <module>()
----> 1 df.groupby('鄉鎮市區').mean() # 以鄉鎮市區進行聚合，可查看所有鄉鎮市區資料

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\groupby\groupby.py in mean(self, *args, **kwargs)
   1304     nv.validate_groupby_func('mean', args, kwargs, ['numeric_only'])
   1305     try:
-> 1306         return self._cython_agg_general('mean', **kwargs)
   1307     except GroupByError:
   1308         raise

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\groupby\groupby.py in _cython_agg_general(self, how, alt, numeric_only, min_count)
   3972         min_count=-1):
   3973     new_items, new_blocks = self._cython_agg_blocks(
-> 3974         how, alt=alt, numeric_only=numeric_only, min_count=min_count)
   3975     return self._wrap_agged_blocks(new_items, new_blocks)
   3976

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\groupby\groupby.py in _cython_agg_blocks(self, how, alt, numeric_only, min_count)
   4044         if len(new_blocks) == 0:
-> 4045             raise DataError('No numeric types to aggregate')
   4047         # reset the locs in the blocks to correspond to our
   4048

DataError: No numeric types to aggregate

```

預處理失敗，且之後的步驟亦無法成功進行