Lec0606 – Project example

新竹市不動產實價登錄資訊-買賣案件| 政府資料開放平臺

https://data.gov.tw/dataset/67502

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]、使用分區或編定、非都市土地使用分區、非都市土地使用地、交易年月、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]、現況格局-房、現況格局-廳、現況格局-衛、現況格局-隔間、有無管理組織、總價[元]、單價[元/平方公尺]、車位...¶

Project 其他資料集

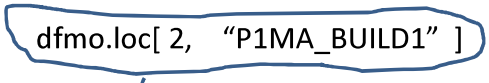
- 美國開放資料平台（https://www.data.gov/）
- 加州大學爾灣分校機器學習資料庫（http://archive.ics.uci.edu/ml/）
- Stanford Large Network Dataset Collection（https://snap.stanford.edu/data/）
- Kaggle（https://www.kaggle.com/）

被 google 買走

你聽過 Kaggle 嗎？Google 買下知名機器學習社群，加速推廣雲端AI

如何取 dataframe row/column 位置

- df.loc




```
dfmo[:5]
```

	P1LA_CF46	(沒有資料行名稱)2	P1LA_FArea	P1MA_BUILD1	P1MA_BUILD2	P1MA_BUILD3	P1MA_TOTPRICE
0	香山區	7209.0	114.860	3房	1廳	1衛	8500000
1	東區	NaN	75.900	3房	2廳	1衛	5408000
2	東區	NaN	0.000	NaN	NaN	NaN	13770000
3	東區	10102.0	103.851	1房	1廳	1衛	6900000
4	香山區	8705.0	28.392	1房	NaN	1衛	520000

df.loc : explicit index
df.iloc : implicit index, as if it is a simple Numpy array

如何修改欄位名稱

```
dfmo[:5]
```

	P1LA_CF46	(沒有資料行名稱)2	P1LA_FArea	P1MA_BUILD1	P1MA_BUILD2	P1MA_BUILD3	P1MA_TOTPRICE
0	香山區	7209.0	114.860	3房	1廳	1衛	8500000
1	東區	NaN	75.900	3房	2廳	1衛	5408000
2	東區	NaN	0.000	NaN	NaN	NaN	13770000
3	東區	10102.0	103.851	1房	1廳	1衛	6900000
4	香山區	8705.0	28.392	1房	NaN	1衛	520000

```
dfmo=dfmo.rename(columns={'P1LA_CF46':'鄉鎮市區',
                          '(沒有資料行名稱)2':'建築完成年月',
                          'P1LA_FArea':'建物移轉總面積[平方公尺]' })
dfmo[:6]
```

	鄉鎮市區	建築完成年月	建物移轉總面積[平方公尺]	P1MA_BUILD1	P1MA_BUILD2	P1MA_BUILD3	P1MA_TOTPRICE
0	香山區	7209.0	114.860	3房	1廳	1衛	8500000
1	東區	NaN	75.900	3房	2廳	1衛	5408000
2	東區	NaN	0.000	NaN	NaN	NaN	13770000
3	東區	10102.0	103.851	1房	1廳	1衛	6900000
4	香山區	8705.0	28.392	1房	NaN	1衛	520000
5	東區	7106.0	110.400	4房	2廳	1衛	5428000

一些常用dataframe 指令

- 刪除欄位
  - df_new=df.drop(“欲刪除的欄位 “ , axis=1)

std (the standard deviation)

- std: The standard deviation is the square root of the average of the squared deviations from the mean, i.e.,
  - VAR(X)=E[$(X\text{-}E[X])^2$]
  - STD(X)=VAR$(X)^{0.5}$

```
import math
series_X=dfmm['new']
std = math.sqrt(np.mean(abs(series_X -series_X.mean())**2))
print(std)
```