

隨堂練習5

題目：

台灣的觀光遊憩據點遍佈各地，假設臺東縣政府想與旅遊業者透過遊客數據來分析各景點的遊客趨勢，進而制定更有效的觀光政策與行銷策略。現在，如果你是一名數據分析師，負責處理台灣某旅遊報告提供的觀光遊憩據點訪客數據，你被指派任務是：

目標：

1. 讀取一份包含各縣市旅遊景點遊客統計的 **CSV** 文件，這份文件的欄位名稱部分缺失，因此需要進行適當的欄位修正。
2. 將景點名稱(遊憩據點)作為索引，方便查詢與分析。
3. 修正欄位名稱，使數據結構更清晰易讀。
4. 輸出整理後的前幾筆資料，以確認處理是否正確。

輸入格式

本題無輸入格式

輸出格式

縣市		英文名稱 \	
遊憩據點			
NaN	NaN	Location	
東部海岸富岡地質公園及加路蘭休憩區		臺東縣	Taitung County
三仙台遊憩區及周邊地區		臺東縣	Taitung County
八仙洞遊憩區		臺東縣	Taitung County
遊憩地點英文名稱		108年7月	\
遊憩據點			
NaN			
Scenic Spots		2019/7	
東部海岸富岡地質公園及加路蘭休憩區		East Coast Fugang Geopark	
And Jialulan Recreat...		65771	
三仙台遊憩區及周邊地區		Sanxiantai recreation area	
and surrounding areas		64872	
八仙洞遊憩區		Baxian	
Caves Recreation Area		11551	
	108年8月	108年9月	108年10月
108年11月	108年12月	109年1月	...
\			
遊憩據點			
...			
NaN	2019/8	2019/9	2019/10
2019/11	2019/12	2020/1	...
東部海岸富岡地質公園及加路蘭休憩區		51654	4770976751
40976	40877	91235	...
三仙台遊憩區及周邊地區		52134	40093 56200
40454	57901	...	
八仙洞遊憩區		9482	11625 11476
6124	14833	8815	
...			
	112年11月	112年12月	113年1月
113年2月	113年3月	113年4月	113年5月
\			
遊憩據點			
NaN	2023/11	2023/12	2024/1
2024/2	2024/3	2024/4	2024/5
東部海岸富岡地質公園及加路蘭休憩區		41079	46334
325808	295242	227590	288382
250299			

三仙台遊憩區及周邊地區	32781	27492	125682
217132	155958	107730	116275
八仙洞遊憩區	5730	4576	24792
27668	16360	22334	36927
	113年6月	113年7月	小計
遊憩據點			
NaN	2024/6	2024/7	Total
東部海岸富岡地質公園及加路蘭休憩區	279951	313015	4891011
三仙台遊憩區及周邊地區	125704	146359	2994488
八仙洞遊憩區	22540	25166	531848

Hint:

```

1 df.rename(columns={'Unnamed: 1': '英文名稱', 'Unnamed:
2 3': '遊憩地點英文名稱'}, inplace=True)
df.set_index('遊憩據點', inplace=True)

```

步驟分解:

(1) 讀取 CSV 文件, 使用 `pd.read_csv()`, 在 **Pandas** 中, `pd.read_csv()` 是最常用來讀取 **CSV**(逗號分隔值) 文件, 而 **csv** 則是 **excel** 當中的一種下載儲存的方法。它能夠將 **CSV** 檔案轉換為 **DataFrame**(表格形式的數據結構), 方便進行分析和處理。

即

```

1 import pandas as pd
2 df = pd.read_csv('觀光遊憩據點按縣市及遊憩據點交叉分析.
csv')

```

(2) 使用 `rename()` 方法將 `Unnamed: 1` 改名為 "英文名稱", 將 `Unnamed: 3` 改名為 "遊憩地點英文名稱"。在 **Pandas** 中, `rename()` 方法可以用來重新命名欄位名稱(**columns**) 或 索引(**index**) 等等。

如下:

```
1 df.rename(columns={'Unnamed: 1': '英文名稱', 'Unnamed: 3': '遊憩地點英文名稱'}, inplace=True)
```

(3) 使用 `set_index()` 方法將 "遊憩據點" 欄位設為索引。

`set_index()` 可以用來 將某一欄(或多欄)設為 **DataFrame** 的索引，這在進行資料分析和查詢時非常有用。

```
1 ldf.set_index('遊憩據點', inplace=True)
```

(4) 使用 `print(df.head(3))` 來顯示結果的前三行。