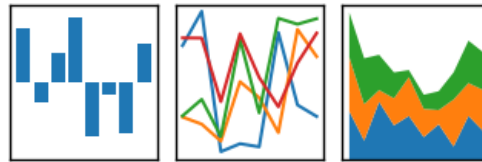


PandasBasic

2018 年 6 月 22 日

pandas
 $y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$



1 Pandas 基本使用

1.1 介紹

專門拿來處理表格的函式庫，一生懸命只為了處理表格而生

1.2 用途

1. 處理各種表格 (csv, excel... 等等)
2. 拿來表示機器學習的資料集 (每個列就是一個物品，每個行代表物品的每個特徵)
3. 可以快速結合常用的繪圖函式庫，直接畫出漂亮的圖形

1.3 安裝方法

1. 使用 PyCharm: PyCharm -> Settings -> Project -> Project Interpreter -> + -> (搜索)pandas -> Install Packages
2. 使用命令列: cd 到你安裝 Python 的資料夾 -> 輸入 `python -m pip install pandas`

1.4 官方文件

<http://pandas.pydata.org/pandas-docs/stable/>

1.5 可以處理的表格形式

<http://pandas.pydata.org/pandas-docs/stable/io.html>

1.6 目標

我們使用 Kaggle 的 TED 資料集來教你 Pandas 的基本操作

1.7 資料集位置

<https://www.kaggle.com/rounakbanik/ted-talks>

1. 需要登入才能下載
2. 只取裡面的 `ted_main.csv` 來做分析

1.8 Pandas 基本資料

1. 多個行 * 多個列 -> DataFrame
2. 一個行 * 多個列或者一個列 * 多個行 -> Series

```
In [1]: import pandas as pd
        # 為了顯示的漂亮，我刻意的把印出來的 row 和 column 只顯示六個
        # 大家練習的時候可以去掉下面兩行
        pd.set_option('display.max_rows', 6)
        pd.set_option('display.max_columns', 6)
```

1.9 讀取表格操作

使用 `read_` 表格形式來讀取，記得最好明確表示用 `utf-8` 來讀取網站檔案 (網路上的檔案通常使用 `utf-8` 來儲存)

注意: 如果是 windows 的一些檔案，內建的儲存編碼是 ANSI，用 `utf-8` 會失效，我們留待編碼篇好好說

1.9.1 read_csv 重要參數:

1. 必要參數: 檔案位置
2. 選用參數 (有預設值): 讀取使用編碼

1.9.2 read_csv 回傳值:

DataFrame

```
In [2]: df = pd.read_csv("ted_main.csv", encoding = "utf-8")
        df
```

```

Out[2]:
      comments      description  duration \
0      4553  Sir Ken Robinson makes an entertaining and pro...      1164
1       265  With the same humor and humanity he exuded in ...      977
2       124  New York Times columnist David Pogue takes aim...     1286
...      ...      ...      ...
2547      10  Science fiction visions of the future show us ...      651
2548      32  In an unmissable talk about race and politics ...     1100
2549       8  With more than half of the world population li...      519

      ...      title \
0      ...      Do schools kill creativity?
1      ...      Averting the climate crisis
2      ...      Simplicity sells
...      ...      ...
2547      ...      What intelligent machines can learn from a sch...
2548      ...      A black man goes undercover in the alt-right
2549      ...      How a video game might help us build better ci...

      url      views
0  https://www.ted.com/talks/ken_robinson_says_sc...  47227110
1  https://www.ted.com/talks/al_gore_on_averting_...  3200520
2  https://www.ted.com/talks/david_pogue_says_sim...  1636292
...      ...      ...
2547  https://www.ted.com/talks/radhika_nagpal_what_...  375647
2548  https://www.ted.com/talks/theo_e_j_wilson_a_bl...  419309
2549  https://www.ted.com/talks/karoliina_korppoo_ho...  391721

[2550 rows x 17 columns]

```

1.10 DataFrame 大小

1. 由於我們有兩個維度，所以以前習慣的 `len` 不能使用了，我們需要使用 `.shape` 來取得兩個維度
2. `.shape` 是一個 `tuple`，所以第一個元素 `[0]` 就是你的列數，第二個元素 `[1]` 就是你的行數

```
In [3]: df.shape
```

```
Out[3]: (2550, 17)
```

1.11 表格行篩選

篩選行的時候，我們就像在操作字典一樣，對你的 DataFrame 加上 []

1.11.1 單行操作

直接在 [] 加上你想要的標籤名字，由於一個維度變成 1，所以你會發現從 DataFrame 變成 Series 了 (印出來是不一樣的)

```
In [4]: df["comments"]
```

```
Out[4]: 0      4553
        1       265
        2      124
        ...
        2547     10
        2548     32
        2549      8
        Name: comments, dtype: int64
```

1.11.2 多行操作

你必須把想要的標籤集成一個 list 傳入行操作的 []，所以這裡兩個 [] 代表截然不同的意思
 1. 外面的 []: DataFrame 的行操作 2. 裡面的 []: 把標籤集合起來的 list

```
In [5]: df[ ["comments", "description", "url"] ]
```

```
Out[5]:
```

	comments	description \	url
0	4553	Sir Ken Robinson makes an entertaining and pro...	https://www.ted.com/talks/ken_robinson_says_sc...
1	265	With the same humor and humanity he exuded in ...	https://www.ted.com/talks/al_gore_on_averting_...
2	124	New York Times columnist David Pogue takes aim...	https://www.ted.com/talks/david_pogue_says_sim...
...	
2547	10	Science fiction visions of the future show us ...	
2548	32	In an unmissable talk about race and politics ...	
2549	8	With more than half of the world population li...	

```

...
2547 https://www.ted.com/talks/radhika_nagpal_what_...
2548 https://www.ted.com/talks/theo_e_j_wilson_a_bl...
2549 https://www.ted.com/talks/karoliina_korppoo_ho...

[2550 rows x 3 columns]

```

1.12 表格列篩選

1. 篩選列的時候，我們使用的是`.loc`(少用, 如果有自己創造列標籤才用得上), `iloc`(常用, 使用 `pandas` 幫你創的 0 開始的列標籤)
2. 使用 `iloc` 的時候會得到一個像是 `list` 的資料，接著就可以使用類似 `list` 的操作來操作
3. `.iloc -> ["第一筆資料", "第二筆資料", "第三筆資料", ..., "最後一筆資料"]`

1.12.1 單列篩選

在`.iloc` 這個列表加上 [座號]

```
In [6]: df.iloc[0]
```

```

Out [6]: comments                                4553
description   Sir Ken Robinson makes an entertaining and pro...
duration                                             1164
...
title                                                Do schools kill creativity?
url          https://www.ted.com/talks/ken_robinson_says_sc...
views                                              47227110
Name: 0, dtype: object

```

1.12.2 多列篩選

`.iloc` 後使用 [頭部座號 (包括): 尾部座號 (不包括)]

```
In [7]: # 取 10, 11, 12, 13, 14 共五筆資料
df.iloc[10:15]
```

```

Out [7]:      comments                                description  duration \
10      79  Accepting his 2006 TED Prize, Cameron Sinclair...      1414
11      55  Jehane Noujaim unveils her 2006 TED Prize wish...      1538
12      71  Accepting the 2006 TED Prize, Dr. Larry Brill...      1550

```

```

13      242  Jeff Han shows off a cheap, scalable multi-tou...      527
14      99  Nicholas Negroponte, founder of the MIT Media ...    1057

```

```

...                                     title \
10  ...      My wish: A call for open-source architecture
11  ...                                     My wish: A global day of film
12  ...                                     My wish: Help me stop pandemics
13  ...      The radical promise of the multi-touch interface
14  ...                                     One Laptop per Child

```

```

...                                     url      views
10  https://www.ted.com/talks/cameron_sinclair_on_...  1211416
11  https://www.ted.com/talks/jehane_noujaim_inspi...  387877
12  https://www.ted.com/talks/larry_brilliant_want...  693341
13  https://www.ted.com/talks/jeff_han_demos_his_b...  4531020
14  https://www.ted.com/talks/nicholas_negroponte_...  358304

```

```
[5 rows x 17 columns]
```

1.12.3 頭幾列篩選

Pandas 也提供一些讓你偷懶的函式，如果是要篩選頭幾列的話用 `head` 函式來篩選

In [8]: # 頭五列

```
df.head(5)
```

```

Out[8]:      comments                                     description  duration \
0      4553  Sir Ken Robinson makes an entertaining and pro...      1164
1       265  With the same humor and humanity he exuded in ...       977
2       124  New York Times columnist David Pogue takes aim...     1286
3       200  In an emotionally charged talk, MacArthur-winn...     1116
4       593  You've never seen data presented like this. Wi...     1190

...                                     title \
0  ...      Do schools kill creativity?
1  ...      Averting the climate crisis
2  ...      Simplicity sells
3  ...      Greening the ghetto

```

```

4      ...      The best stats you've ever seen

                                url      views
0  https://www.ted.com/talks/ken_robinson_says_sc...  47227110
1  https://www.ted.com/talks/al_gore_on_averting_...  3200520
2  https://www.ted.com/talks/david_pogue_says_sim...  1636292
3  https://www.ted.com/talks/majora_carter_s_tale...  1697550
4  https://www.ted.com/talks/hans_rosling_shows_t...  12005869

[5 rows x 17 columns]

```

1.12.4 尾幾列篩選

使用 `tail` 函式來做尾部的篩選

In [9]: # 尾五列

```
df.tail(5)
```

```

Out[9]:      comments      description  duration \
2545      17  Between 2008 and 2016, the United States depor...      476
2546      6  How can you study Mars without a spaceship? He...      290
2547     10  Science fiction visions of the future show us ...      651
2548     32  In an unmissable talk about race and politics ...     1100
2549      8  With more than half of the world population li...      519

      ...      title \
2545  ...  What we're missing in the debate about immigra...
2546  ...      The most Martian place on Earth
2547  ...  What intelligent machines can learn from a sch...
2548  ...      A black man goes undercover in the alt-right
2549  ...  How a video game might help us build better ci...

                                url      views
2545  https://www.ted.com/talks/duarte_geraldino_wha...  450430
2546  https://www.ted.com/talks/armando_azua_bustos_...  417470
2547  https://www.ted.com/talks/radhika_nagpal_what_...  375647
2548  https://www.ted.com/talks/theo_e_j_wilson_a_bl...  419309
2549  https://www.ted.com/talks/karoliina_korppoo_ho...  391721

```

```
[5 rows x 17 columns]
```

1.13 表格行+列篩選

1. 只要是 DataFrame 就可以使用上面的行或者列篩選
2. 所以你可以任意組合行列篩選，先 [] 再 .iloc[]，或者先 .iloc[] 再 []

In [10]: # 如果：後面不寫就是到底，：前面不寫就是從最頭開始

```
df[["comments", "description", "duration"]].iloc[:5]
```

```
Out[10]:
```

	comments	description	duration
0	4553	Sir Ken Robinson makes an entertaining and pro...	1164
1	265	With the same humor and humanity he exuded in ...	977
2	124	New York Times columnist David Pogue takes aim...	1286
3	200	In an emotionally charged talk, MacArthur-winn...	1116
4	593	You've never seen data presented like this. Wi...	1190

1.14 列過濾

1. 過濾操作是把符合我們期待的列留下來，不符合期待的列丟掉的一個操作
2. 核心概念是做一個跟我們的資料筆數一樣大的布林 list，對到 True 的資料留下，對到 False 的資料丟掉
3. (特別) 這時候一樣是對你的 DataFrame 加上 [], 把布林 list 丟進你的 [] 裡

In [11]: # 先做個實驗給你看，只取三列資料

```
test = df.iloc[:3]
test
```

```
Out[11]:
```

	comments	description	duration \
0	4553	Sir Ken Robinson makes an entertaining and pro...	1164
1	265	With the same humor and humanity he exuded in ...	977
2	124	New York Times columnist David Pogue takes aim...	1286

	...	title \
0	...	Do schools kill creativity?
1	...	Averting the climate crisis
2	...	Simplicity sells

	url	views
--	-----	-------


```

0 https://www.ted.com/talks/ken_robinson_says_sc... 47227110
1 https://www.ted.com/talks/al_gore_on_averting... 3200520
2 https://www.ted.com/talks/david_pogue_says_sim... 1636292

```

[3 rows x 17 columns]

```

In [12]: # 過濾，創一個三個大小的 True, False list
# 對到 True(第一三筆) 留下，對到 False(第二筆) 丟掉
test[ [True, False, True] ]

```

```

Out[12]:      comments                                description  duration \
0      4553  Sir Ken Robinson makes an entertaining and pro...      1164
2       124  New York Times columnist David Pogue takes aim...      1286

      ...                                title \
0      ...  Do schools kill creativity?
2      ...  Simplicity sells

                                url      views
0 https://www.ted.com/talks/ken_robinson_says_sc... 47227110
2 https://www.ted.com/talks/david_pogue_says_sim... 1636292

```

[2 rows x 17 columns]

```

In [13]: # 但我們不可能自己用手創一個 2000 多個元素的布林 list
# 所以我們藉由 pandas 的函式幫我們
# 先取出一個 Series 取 str 屬性得到字串 list
# 藉由 pandas 定義的 contains 對裡面每個元素做出布林判斷
bool_filter = df["description"].str.contains("Sir")
bool_filter

```

```

Out[13]: 0      True
1     False
2     False
...
2547  False
2548  False
2549  False
Name: description, dtype: bool

```

```
In [14]: # 帶入 DataFrame
```

```
df[bool_filter]
```

```
Out[14]:
```

	comments	description	duration \
0	4553	Sir Ken Robinson makes an entertaining and pro...	1164
15	325	Violinist Sirena Huang gives a technically bri...	1481
54	203	Speaking as both an astronomer and "a concerne...	1046
...
1978	64	The founder of Sirius XM satellite radio, Mart...	1264
2192	61	Trust: How do you earn it? Banks use credit sc...	491
2503	20	How smart can our machines make us? Tom Gruber...	586

	title \
0	Do schools kill creativity?
15	An 11-year-old's magical violin
54	Is this our final century?
...	...
1978	My daughter, my wife, our robot, and the quest...
2192	A smart loan for people with no credit history...
2503	How AI can enhance our memory, work and social...

	url	views
0	https://www.ted.com/talks/ken_robinson_says_sc...	47227110
15	https://www.ted.com/talks/sirena_huang_dazzles...	2702470
54	https://www.ted.com/talks/martin_rees_asks_is_...	2121177
...
1978	https://www.ted.com/talks/martine_rothblatt_my...	1304737
2192	https://www.ted.com/talks/shivani_siroya_a_sma...	1437353
2503	https://www.ted.com/talks/tom_gruber_how_ai_ca...	1139827

[11 rows x 17 columns]

```
In [15]: # 你仔細對照，你會發現 contains 是只要有 contains 那個字串就可以
# 並不一定是完整的一個字 (Sirena 也算有 contains Sir)
# 但我們可以使用格式 (正規表示式) 來結合 contains
# 記得在你格式字串前加上 r(不轉換任何東西，原始字串)
# 不然\b 會被當成 backspace，而不是兩個字
df[ df["description"].str.contains(r"\bSir\b") ]
```

```

Out[15]:
   comments      description  duration \
0      4553  Sir Ken Robinson makes an entertaining and pro...    1164
54      203  Speaking as both an astronomer and "a concerne...    1046
692     1234  In this poignant, funny follow-up to his fable...    1008
833      473  In this talk from RSA Animate, Sir Ken Robinso...     700
1502     634  Sir Ken Robinson outlines 3 principles crucial...    1151
1802      59  Sir Tim Berners-Lee invented the World Wide We...     403

   ...      title \
0      ...      Do schools kill creativity?
54      ...      Is this our final century?
692      ...      Bring on the learning revolution!
833      ...      Changing education paradigms
1502      ...      How to escape education's death valley
1802      ...      A Magna Carta for the web

   url      views
0  https://www.ted.com/talks/ken_robinson_says_sc...  47227110
54  https://www.ted.com/talks/martin_rees_asks_is_...  2121177
692  https://www.ted.com/talks/sir_ken_robinson_bri...  7266316
833  https://www.ted.com/talks/ken_robinson_changin...  1854997
1502  https://www.ted.com/talks/ken_robinson_how_to_...  6657858
1802  https://www.ted.com/talks/tim_berners_lee_a_ma...  1054600

[6 rows x 17 columns]

```

1.15 儲存表格

1. 非常簡單!!! 就跟 `read` 一樣，你想儲存什麼就讓你的 `DataFrame` 使用 `to_` 儲存格式
2. 一樣建議指定使用 `utf-8` 做儲存

1.15.1 `to_csv` 重要參數

1. 必要參數: 檔案位置
2. 選用參數 `encoding`(有預設值): 讀取使用編碼
3. 選用參數 `index`(有預設值 `True`): 要不要把 `pandas` 幫你產生的列編號寫進檔案, `True`: 寫, `False`: 不寫, 通常我會選 `False`

In [16]: # 用剛剛的 `filter` 過後的東西做個例子

```
filter_df = df[ df["description"].str.contains(r"\bSir\b") ]
# 儲存成 csv
filter_df.to_csv("filter.csv", encoding = "utf-8", index = False)
```

In [17]: # 把剛剛儲存的東西讀出來給你看看

```
pd.read_csv("filter.csv", encoding = "utf-8")
```

```
Out[17]:
```

	comments	description	duration	\
0	4553	Sir Ken Robinson makes an entertaining and pro...	1164	
1	203	Speaking as both an astronomer and "a concerne...	1046	
2	1234	In this poignant, funny follow-up to his fable...	1008	
3	473	In this talk from RSA Animate, Sir Ken Robinso...	700	
4	634	Sir Ken Robinson outlines 3 principles crucial...	1151	
5	59	Sir Tim Berners-Lee invented the World Wide We...	403	

	...	title	\
0	...	Do schools kill creativity?	
1	...	Is this our final century?	
2	...	Bring on the learning revolution!	
3	...	Changing education paradigms	
4	...	How to escape education's death valley	
5	...	A Magna Carta for the web	

	url	views
0	https://www.ted.com/talks/ken_robinson_says_sc...	47227110
1	https://www.ted.com/talks/martin_rees_asks_is_...	2121177
2	https://www.ted.com/talks/sir_ken_robinson_bri...	7266316
3	https://www.ted.com/talks/ken_robinson_changin...	1854997
4	https://www.ted.com/talks/ken_robinson_how_to_...	6657858
5	https://www.ted.com/talks/tim_berners_lee_a_ma...	1054600

[6 rows x 17 columns]

1.16 刪除行

1. 你可以使用類似字典的刪除方式，用 `del` 來直接刪除一行
2. 你可以使用 `drop` 來刪除多行，不過記得如果你想要讓 `df` 變成刪除過後的樣子要記得設定回去

In [18]: # 删除单行, 直接就修改了 *df*

```
del df['title']
df
```

```
Out[18]:
```

	comments		description	duration \
0	4553	Sir Ken Robinson makes an entertaining and pro...		1164
1	265	With the same humor and humanity he exuded in ...		977
2	124	New York Times columnist David Pogue takes aim...		1286
...
2547	10	Science fiction visions of the future show us ...		651
2548	32	In an unmissable talk about race and politics ...		1100
2549	8	With more than half of the world population li...		519

	tags \
0	['children', 'creativity', 'culture', 'dance', ...
1	['alternative energy', 'cars', 'climate change...
2	['computers', 'entertainment', 'interface desi...
...	...
2547	['AI', 'ants', 'fish', 'future', 'innovation', ...
2548	['Internet', 'TEDx', 'United States', 'communi...
2549	['cities', 'design', 'future', 'infrastructure...

	url	views
0	https://www.ted.com/talks/ken_robinson_says_sc...	47227110
1	https://www.ted.com/talks/al_gore_on_averting_...	3200520
2	https://www.ted.com/talks/david_pogue_says_sim...	1636292
...
2547	https://www.ted.com/talks/radhika_nagpal_what_...	375647
2548	https://www.ted.com/talks/theo_e_j_wilson_a_bl...	419309
2549	https://www.ted.com/talks/karoliina_korppoo_ho...	391721

[2550 rows x 16 columns]

In [23]: # 删除多行, *axis = 1* 指的是删除行的意思, *axis = 0* 是删除列的意思

```
df.drop(["url", "views"], axis = 1)
```

```
Out[23]:
```

	comments		description	duration \
0	4553	Sir Ken Robinson makes an entertaining and pro...		1164

1	265	With the same humor and humanity he exuded in ...	977
2	124	New York Times columnist David Pogue takes aim...	1286
...
2547	10	Science fiction visions of the future show us ...	651
2548	32	In an unmissable talk about race and politics ...	1100
2549	8	With more than half of the world population li...	519

	...	\
0	...	
1	...	
2	...	
...	...	
2547	...	
2548	...	
2549	...	

	related_talks	speaker_occupation	\
0	[{'id': 865, 'hero': 'https://pe.tedcdn.com/im...}	Author/educator	
1	[{'id': 243, 'hero': 'https://pe.tedcdn.com/im...}	Climate advocate	
2	[{'id': 1725, 'hero': 'https://pe.tedcdn.com/i...}	Technology columnist	
...	
2547	[{'id': 2346, 'hero': 'https://pe.tedcdn.com/i...}	Robotics engineer	
2548	[{'id': 2512, 'hero': 'https://pe.tedcdn.com/i...}	Public intellectual	
2549	[{'id': 2682, 'hero': 'https://pe.tedcdn.com/i...}	Game designer	

	tags
0	['children', 'creativity', 'culture', 'dance', ...]
1	['alternative energy', 'cars', 'climate change...]
2	['computers', 'entertainment', 'interface desi...]
...	...
2547	['AI', 'ants', 'fish', 'future', 'innovation', ...]
2548	['Internet', 'TEDx', 'United States', 'communi...]
2549	['cities', 'design', 'future', 'infrastructure...]

[2550 rows x 14 columns]

1.17 (高階技巧) 列過濾

1. 有時候利用 pandas 有的函式難以完成我想要的過濾, 這時候我可以自定義我的過濾流程
2. 對你的 Series 使用 `apply` 來過濾
3. (重要) 你的過濾流程最後一定要回傳 `True` or `False`

```
In [19]: # 我想把 tag 欄位裡的字串拿出來並且轉換成一個 list
# 再檢查某個字串也沒有在 list 裡
# 你的過濾流程的第一個參數, pandas 會幫你傳入
# 就是你每一格的資料, 也就是 element = 每一格的資料
def tag_filter(element):
    # 利用 eval 把字串當成 python 程式執行, 變成一個 list
    tag_list = eval(element)
    if 'children' in tag_list:
        return True
    else:
        return False
# 讓你看看做出來的 bool list
bool_filter = df["tags"].apply(tag_filter)
bool_filter
```

```
Out[19]: 0      True
1     False
2     False
...
2547  False
2548  False
2549  False
Name: tags, dtype: bool
```

```
In [20]: # 來過濾你的 DataFrame, 並且我們只看 tags, description 兩列
df[bool_filter][ ['description', 'tags'] ]
```

```
Out[20]:
```

	description \
0	Sir Ken Robinson makes an entertaining and pro...
14	Nicholas Negroponte, founder of the MIT Media ...
152	Author and illustrator Maira Kalman talks abou...
...	...
2479	Sixty-five million people were displaced from ...

```

2482 "We have seen advances in every aspect of our ...
2525 Could it be wrong to help children in need by ...

```

```

                                tags
0      ['children', 'creativity', 'culture', 'dance',...
14     ['children', 'design', 'education', 'entrepren...
152    ['art', 'children', 'culture', 'design', 'ente...
...
2479  ['TED Books', 'activism', 'big problems', 'chi...
2482  ['children', 'global issues', 'humanity', 'ide...
2525  ['TEDx', 'activism', 'children', 'family', 'po...

[143 rows x 2 columns]

```

In [21]: # 更進階定義，讓你在使用的時候可以再多帶入參數

```

def tag_filter(element, filter_tag):
    tag_list = eval(element)
    if filter_tag in tag_list:
        return True
    else:
        return False

# 你放在 apply 後面的參數，pandas 會幫你丟進你的過濾過程
# 但是參數名字就要對到過濾流程的參數
bool_filter = df["tags"].apply(tag_filter, filter_tag = 'Asia')
bool_filter

```

```

Out[21]: 0      False
         1      False
         2      False
         ...
        2547    False
        2548    False
        2549    False
         Name: tags, dtype: bool

```

In [22]: df[bool_filter][["description", "tags"]]

```

Out[22]:                                description \
4      You've never seen data presented like this. Wi...

```



```

117    Researcher Hans Rosling uses his cool data too...
359    Reporter Jennifer 8. Lee talks about her hunt ...
...
1621    The developed world holds up the ideals of cap...
1948    The former prime minister of Australia, Kevin ...
2310    Americanization and globalization have basical...

```

tags

```

4      ['Africa', 'Asia', 'Google', 'demo', 'economic...
117    ['Africa', 'Asia', 'Google', 'economics', 'glo...
359    ['Asia', 'business', 'culture', 'exploration',...
...
1621    ['Africa', 'Asia', 'china', 'democracy', 'econ...
1948    ['Asia', 'United States', 'china', 'politics']
2310    ['Africa', 'Asia', 'Europe', 'Foreign Policy',...

```

[26 rows x 2 columns]