Tkinter and Pandas

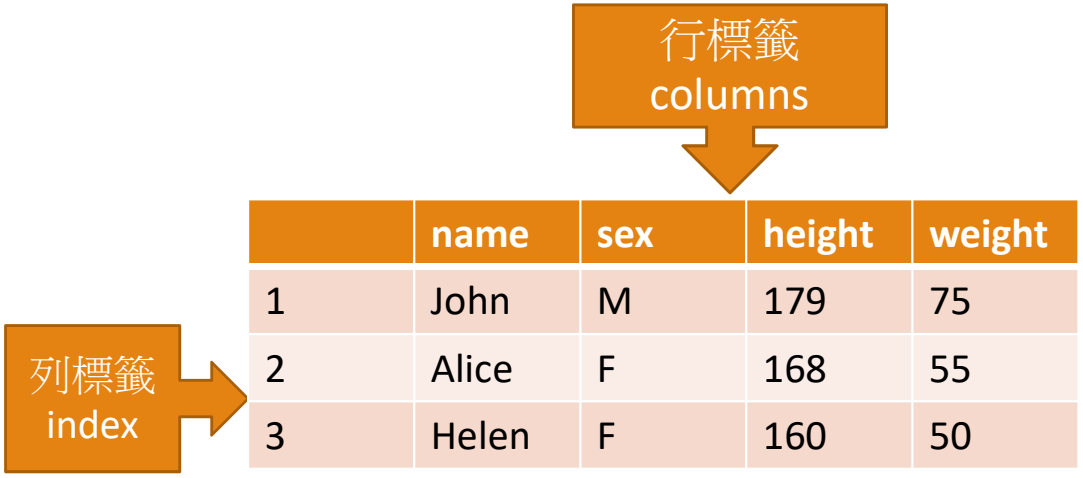
授課老師：邱淑怡

Outline

- Pandas module
  - Introduce dataframe
  - Dataframe operation
  - 10 minutes to pandas (https://pandas.pydata.org/pandas-docs/stable/user_guide/10min.html)
- Combine tkinter(user interface) with pandas(select data) for recommendation system/query

DataFrame

- Stored integer, float, string .....
- 2-dimensional type
- Dataframe contains 3 parts:
  1. index(列標籤)
  2. columns(行標籤)
  3. data



	name	sex	height	weight
1	John	M	179	75
2	Alice	F	168	55
3	Helen	F	160	50

```
import pandas as pd
```

df1=pd.DataFrame(data, index=None, columns=None,其他選擇性參數), data用來指定DataFrame的資料, index用來指定資料的列標籤, column用來指定資料的行標籤(變數名稱)

Basic information about the dataframe

```
display(Df2)

Df2.info()

Df2.describe()

Df2.columns

Df2.sort_values(by='a').head()
```

Read txt, csv, excel url files

1. Excel file: read_df = pd.read_excel("data2.xlsx",sheet_name='工作表2',header=0, index_col=0)
2. csv file: US2020_df = pd.read_csv("D:\\temp\\governors_county.csv",header=0)
3. File from website: df1=pd.read_csv(url_link)
4. Txt file:
   - txt_url = 'http://people.apache.org/~edwardyoon/kmeans.txt'
   - iris_df = pd.read_table(txt_url, sep = "\t")

Selection

Use indexing or label to select a single/multiple data

Methods	Introduction
dataframe.at	Use row/column labels to select a single data (selection by label)
dataframe.iat	Use 0-based indexing to select a single data (selection by position)
dataframe.loc	Use row/column labels to select multiple data (selection by label)
dataframe.iloc	Use 0-based indexing to select multiple data (selection by position)

Selection one data

df1.iat[0,0] #get one data at first row and first column

df1.at[1,’column_1’]

Selection multiple data

1. dataframe_name.iloc[0:3,2:5]:
   - 先篩選第幾筆到第幾筆資料，再選欄位
2. per_df.loc[‘2':'4’,column_1’:’column_3’])
   - loc: location
   - first index range, then column_name range

Selection data by columns

- A single column

  df_1[‘column_name’] or df_1.column_name

  ex: df_1.type.tolist() # 將type欄位的資料抓出來轉成list

- multiple columns

  df_1[[‘column_1’,’column_2’]]

Selection by conditions

- a single condition

```
print(df[df['salary']>50000])
```

- multiple conditions

```
c1 = df['class']=='class0'
c2 = df['height'] > 170
temp_df=df[(c1 & c2)]
(or) temp_df1=df[ (df[‘class’]==‘class0’) & (df[‘height’] > 170) ]
```

Delete some data by column/row labels

```
new_df1 = df.drop([“class"], axis=1)
display(new_df1)
new_df2= df.drop([‘3’,‘5’]) # axis預設值為0
new_df2=df2.drop(new_df2.index[0])  #remove first row
new_df2=df2.drop(new_df2.index[-1])  #remove last row
display(new_df2)
# 刪除空值(nan)
new_df3 = df.dropna()
```

messagebox (訊息視窗)

```
import tkinter as tk
from tkinter import messagebox
win=tk.Tk()
win.geometry("600x500")
def message():
    messagebox.showinfo("result","Hello! everyone!!")
button_1=tk.Button(win,text="show message",command=message)
button_1.pack()
```

Exercise

Question

- Read csv file
- 建立netflix平台的查詢介面
- Show result on messagebox