Pandas實作練習\_20201021

10 minutes to pandas

https://pandas.pydata.org/pandas-docs/stable/user\_guide/10min.html

產生兩個Pandas物件

[1]Pandas物件1

dates = pd.date\_range('20130101', periods=6)

df = pd.DataFrame(np.random.randn(6, 4), index=dates, columns=list('ABCD'))

根據以上的方式所產生的Pandas物件

A B C D

2013-01-01 0.469112 -0.282863 -1.509059 -1.135632

2013-01-02 1.212112 -0.173215 0.119209 -1.044236

2013-01-03 -0.861849 -2.104569 -0.494929 1.071804

2013-01-04 0.721555 -0.706771 -1.039575 0.271860

2013-01-05 -0.424972 0.567020 0.276232 -1.087401

2013-01-06 -0.673690 0.113648 -1.478427 0.524988

[2]Pandas物件2

df2 = pd.DataFrame({'A': 1.,

'B': pd.Timestamp('20121021'),

'C': pd.Series(1, index=list(range(4)), dtype='float32'),

'D': np.array([3] \* 4, dtype='int32'),

'E': pd.Categorical(["test", "train", "test", "train"]),

'F': 'foo'})

df2

題目1:顯示df2每一個欄位的資料型態 \_\_\_\_\_\_\_\_\_\_\_\_\_ df2.dtypes

題目2: 顯示df1的前五筆資料 df.head(5)

題目3: 顯示df2的後五筆資料 df.tail(5)

題目4: Display the index df.index

題目5:Display thecolumns: df.columns

題目6: 顯示df1的描述統計學結果 df.describe()

題目:資料轉換 df.T

原本資料

A B C D

2013-01-01 0.469112 -0.282863 -1.509059 -1.135632

2013-01-02 1.212112 -0.173215 0.119209 -1.044236

2013-01-03 -0.861849 -2.104569 -0.494929 1.071804

2013-01-04 0.721555 -0.706771 -1.039575 0.271860

2013-01-05 -0.424972 0.567020 0.276232 -1.087401

2013-01-06 -0.673690 0.113648 -1.478427 0.524988

轉換後的資料

2013-01-01 2013-01-02 2013-01-03 2013-01-04 2013-01-05 2013-01-06

A 0.469112 1.212112 -0.861849 0.721555 -0.424972 -0.673690

B -0.282863 -0.173215 -2.104569 -0.706771 0.567020 0.113648

C -1.509059 0.119209 -0.494929 -1.039575 0.276232 -1.478427

D -1.135632 -1.044236 1.071804 0.271860 -1.087401 0.524988

=======各種排序技術======================

題目:逆排序1 根據column df.sort\_index(axis=1, ascending=False)

排序後的資料如下:

D C B A

2013-01-01 -1.135632 -1.509059 -0.282863 0.469112

2013-01-02 -1.044236 0.119209 -0.173215 1.212112

2013-01-03 1.071804 -0.494929 -2.104569 -0.861849

2013-01-04 0.271860 -1.039575 -0.706771 0.721555

2013-01-05 -1.087401 0.276232 0.567020 -0.424972

2013-01-06 0.524988 -1.478427 0.113648 -0.673690

題目: 逆排序2 根據索引 df.sort\_index(axis=0, ascending=False)

題目: 根據column B 由小排到大 df.sort\_values(by='B')

A B C D

2013-01-03 -0.861849 -2.104569 -0.494929 1.071804

2013-01-04 0.721555 -0.706771 -1.039575 0.271860

2013-01-01 0.469112 -0.282863 -1.509059 -1.135632

2013-01-02 1.212112 -0.173215 0.119209 -1.044236

2013-01-06 -0.673690 0.113648 -1.478427 0.524988

2013-01-05 -0.424972 0.567020 0.276232 -1.087401

===== Selection各種資料選擇技術===========

題目: 挑出column A的所有欄位資料 df['A']

題目: 挑出前三筆的所有欄位資料 df[0:3]

題目: 挑出從20130102到20130104的所有欄位資料 df['20130102':'20130104']

題目: 挑出第一筆的所有欄位資料但不含索引

df.loc[dates[0]]

題目: 挑出所有筆數的column A與column B資料

df.loc[:, ['A', 'B']]

題目: slicing運算

挑出從20130102到20130104的column A與column B資料

df.loc['20130102':'20130104', ['A', 'B']]

題目:挑出單一點的兩種技術

挑出第一筆的column A資料

df.loc[dates[0], 'A']

df.at[dates[0], 'A']

======= 根據不同位置挑選出資料Selection by position ======

題目: 挑出第三筆的所有欄位資料但不含索引

df.iloc[3]

題目:挑選技術:完成底下的Selection

結果如下

A B

2013-01-04 0.721555 -0.706771

2013-01-05 -0.424972 0.567020

df.iloc[3:5, 0:2]

題目:

題目: