python

April 3, 2021

[2]:
```
import pandas as pd
```

[3]:
```
df1 = pd.read_excel("/Users/zhengyuanrui/Desktop/  /  .xlsx")
```

[4]:
```
df1.head()
```

[4]:

					IP	1	\
0	1	2021/3/23 8:23:24	57	NaN	117.136.84.198(-)	1	
1	2	2021/3/23 8:37:29	134	NaN	221.192.178.123(-)	1	
2	3	2021/3/23 9:28:19	102	NaN	117.136.84.17(-)	2	
3	4	2021/3/23 9:49:03	169	NaN	223.104.240.157(-)	2	
4	5	2021/3/23 9:54:10	140	NaN	117.136.72.68(-)	2	

	2	3	4	…	24	\
0	2	1	2	…	6	
1	4	1	2	…	5	
2	4	2	1	…	4	
3	2	1	2	…	5	
4	2	1	2	…	6	

	25	26	27	\
0	5	5	4	
1	4	4	4	
2	4	4	4	
3	4	4	3	
4	4	4	4	

	28	29	\
0	5	5	
1	3	3	
2	4	2	
3	3	4	
4	4	3	

	30	31	32		
0	5	4	5	122	
1	4	3	4	102	

```
2                    2                    2                    5   97
3                    4                    4                    4  103
4                    3                    4                    4   84

[5 rows x 39 columns]
```

[5]:
```
df1.rename(columns={' ': 'ID', '    ':'time', '  ': 'use_time',' ':'source',
                    '  ':'detail_source',' IP':'IP','1   ':'gender', '2   ':
↪'grade',
                     '3  ':'residence', '4    ':'school'}, inplace=True)
```

[6]:
```
df1
```

[6]:
```
      ID                time use_time source detail_source  \
0      1   2021/3/23 8:23:24       57                  NaN
1      2   2021/3/23 8:37:29      134                  NaN
2      3   2021/3/23 9:28:19      102                  NaN
3      4   2021/3/23 9:49:03      169                  NaN
4      5   2021/3/23 9:54:10      140                  NaN
..   …                  …      …    …             …
522  523  2021/3/26 19:44:48      258                  NaN
523  524  2021/3/26 20:01:30       44                  NaN
524  525  2021/3/27 10:06:41      113                  NaN
525  526  2021/3/27 11:19:26       83                  NaN
526  527  2021/3/27 12:59:22      342                  NaN

                         IP  gender  grade  residence  school  …  \
0     117.136.84.198( - )       1      2          1       2  …
1    221.192.178.123( - )       1      4          1       2  …
2      117.136.84.17( - )       2      4          2       1  …
3    223.104.240.157( - )       2      2          1       2  …
4      117.136.72.68( - )       2      2          1       2  …
..                     …      …    …         …      …  …
522    117.136.72.12( - )       2      4          2       2  …
523   111.206.214.44( - )       2      4          1       1  …
524    117.136.0.144( - )       2      1          1       1  …
525   117.136.45.145( - )       2      4          2       1  …
526     183.8.144.39( - )       2      4          1       1  …

     24          25             \
0                    6                    5
1                    5                    4
2                    4                    4
3                    5                    4
4                    6                    4
..                 …                    …
522                  4                    4
```

523	7	5
524	4	4
525	6	5
526	7	5

	26	27	\
0		5	4
1		4	4
2		4	4
3		4	3
4		4	4
..		…	…
522		4	4
523		5	5
524		3	3
525		5	5
526		5	5

	28	29	\
0		5	5
1		3	3
2		4	2
3		3	4
4		4	3
..		…	…
522		4	4
523		5	5
524		3	3
525		4	3
526		5	3

	30	31	32	\
0		5	4	5
1		4	3	4
2		2	2	5
3		4	4	4
4		3	4	4
..		…	…	…
522		2	3	3
523		5	5	5
524		2	2	2
525		2	5	5
526		3	3	3

0	122
1	102

```
2     97
3    103
4     84
..   …
522   95
523  121
524   88
525  113
526  108

[527 rows x 39 columns]
```

[7]:
```
df1.columns = ['id', 'time', 'use_time', 'source','detail_source',
               'ip','gender','grade','residence','school','sns1','sns2',
               'sns3','sns4','sns5','sns6','ses1','ses2','ses3','ses4',
               'ses5','ses6','ses7','ses8','ses9','ses10','cses1','cses2',
               'cses3','cses4','eib1','eib2','eib3','eib4','eib5','eib6',
               'eib7','eib8','total']
```

[8]:
```
df1.head()
```

[8]:
```
   id              time use_time source detail_source  \
0   1  2021/3/23 8:23:24       57                NaN
1   2  2021/3/23 8:37:29      134                NaN
2   3  2021/3/23 9:28:19      102                NaN
3   4  2021/3/23 9:49:03      169                NaN
4   5  2021/3/23 9:54:10      140                NaN

                      ip  gender  grade  residence  school  …  cses4  eib1  \
0   117.136.84.198( - )       1      2          1       2  …      6     5
1  221.192.178.123( - )       1      4          1       2  …      5     4
2    117.136.84.17( - )       2      4          2       1  …      4     4
3  223.104.240.157( - )       2      2          1       2  …      5     4
4   117.136.72.68( - )        2      2          1       2  …      6     4

   eib2  eib3  eib4  eib5  eib6  eib7  eib8  total
0     5     4     5     5     5     4     5    122
1     4     4     3     3     4     3     4    102
2     4     4     4     2     2     2     5     97
3     4     3     3     4     4     4     4    103
4     4     4     4     3     3     4     4     84

[5 rows x 39 columns]
```

[9]:
```
df2 = df1[["id",'gender','grade','school','sns1','sns2',
               'sns3','sns4','sns5','sns6','ses1','ses2','ses3','ses4',
               'ses5','ses6','ses7','ses8','ses9','ses10','cses1','cses2',
```

```
                'cses3','cses4','eib1','eib2','eib3','eib4','eib5','eib6',
                'eib7','eib8']]
```

[10]:
```
df2.head()
```

[10]:

	id	gender	grade	school	sns1	sns2	sns3	sns4	sns5	sns6	…	cses3	\
0	1	1	2	2	5	4	5	4	5	5	…	7	
1	2	1	4	2	5	4	5	3	4	2	…	5	
2	3	2	4	1	5	4	5	4	2	5	…	5	
3	4	2	2	2	4	4	4	4	4	3	…	5	
4	5	2	2	2	2	2	1	1	2	1	…	5	

	cses4	eib1	eib2	eib3	eib4	eib5	eib6	eib7	eib8
0	6	5	5	4	5	5	5	4	5
1	5	4	4	4	3	3	4	3	4
2	4	4	4	4	4	2	2	2	5
3	5	4	4	3	3	4	4	4	4
4	6	4	4	4	4	3	3	4	4

[5 rows x 32 columns]

[20]:
```
df2['ses3'] = 5-df2['ses3']
df2['ses5'] = 5-df2['ses5']
df2['ses9'] = 5-df2['ses9']
df2['ses10'] = 5-df2['ses10']
```

```
<ipython-input-20-c216746a1a5d>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df2['ses3'] = 5-df2['ses3']
<ipython-input-20-c216746a1a5d>:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df2['ses5'] = 5-df2['ses5']
<ipython-input-20-c216746a1a5d>:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df2['ses9'] = 5-df2['ses9']
```

```
<ipython-input-20-c216746a1a5d>:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df2['ses10'] = 5-df2['ses10']
```

[26]:
```
df2['snst'] = df2['sns1'] + df2['sns2'] +␣
 ↪df2['sns3']+df2['sns4']+df2['sns5']+df2['sns6']
df2["sest"] = df2['ses1'] + df2['ses2'] + df2['ses3'] + df2['ses4'] +␣
 ↪df2['ses5'] + df2['ses6'] + df2['ses7'] + df2['ses8'] + df2['ses9'] +␣
 ↪df2['ses10']
df2['csest'] = df2['cses1'] + df2['cses2']+ df2['cses3']+ df2['cses4']
df2['eibt'] = df2['eib1'] + df2['eib2'] + df2['eib3'] + df2['eib4'] +␣
 ↪df2['eib5'] + df2['eib6'] + df2['eib7'] + df2['eib8']
```

```
<ipython-input-26-46c4d9ffd206>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df2['snst'] = df2['sns1'] + df2['sns2'] +
df2['sns3']+df2['sns4']+df2['sns5']+df2['sns6']
<ipython-input-26-46c4d9ffd206>:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df2["sest"] = df2['ses1'] + df2['ses2'] + df2['ses3'] + df2['ses4'] +
df2['ses5'] + df2['ses6'] + df2['ses7'] + df2['ses8'] + df2['ses9'] +
df2['ses10']
<ipython-input-26-46c4d9ffd206>:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df2['csest'] = df2['cses1'] + df2['cses2']+ df2['cses3']+ df2['cses4']
<ipython-input-26-46c4d9ffd206>:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df2['eibt'] = df2['eib1'] + df2['eib2'] + df2['eib3'] + df2['eib4'] +
```

```
df2['eib5'] + df2['eib6'] + df2['eib7'] + df2['eib8']
```

[27]:
```
df2.head()
```

[27]:

	id	gender	grade	school	sns1	sns2	sns3	sns4	sns5	sns6	…	eib4	\
0	1	1	2	2	5	4	5	4	5	5	…	5	
1	2	1	4	2	5	4	5	3	4	2	…	3	
2	3	2	4	1	5	4	5	4	2	5	…	4	
3	4	2	2	2	4	4	4	4	4	3	…	3	
4	5	2	2	2	2	2	1	1	2	1	…	4	

	eib5	eib6	eib7	eib8	seshh	snst	sest	csest	eibt
0	5	5	4	5	8	28	28	25	38
1	3	4	3	4	8	23	31	22	29
2	2	2	2	5	6	25	27	17	27
3	4	4	4	4	6	23	27	22	30
4	3	3	4	4	6	9	28	22	30

[5 rows x 37 columns]

[29]:
```
df3 = df2[["id",'gender','grade','school','snst','sest','csest','eibt']]
```

[32]:
```
df3
```

[32]:

	id	gender	grade	school	snst	sest	csest	eibt
0	1	1	2	2	28	28	25	38
1	2	1	4	2	23	31	22	29
2	3	2	4	1	25	27	17	27
3	4	2	2	2	23	27	22	30
4	5	2	2	2	9	28	22	30
..	…	…	…	…	…	…	…	…
522	523	2	4	2	22	26	18	28
523	524	2	4	1	28	34	28	40
524	525	2	1	1	25	23	17	22
525	526	2	4	1	26	30	24	34
526	527	2	4	1	24	31	28	32

[527 rows x 8 columns]

[ ]: