

Python - Analiza danych z modułem PANDAS

www.udemy.com (<http://www.udemy.com>) (R)

LAB - S05-L007 - melt

1. Zaimportuj moduł pandas i numpy nadaj im standardowe aliasy. Zaimportuj też datetime, timedelta i time, możesz skorzystać z poniższych poleceń:

```
from datetime import datetime
from datetime import timedelta
import time
```

2. Do wykonania zadań z wykorzystaniem polecenia melt będziemy korzystać z danych w postaci tabeli przestawenej. Uruchom poniższy kod, który przygotuje zmienną df o odpowiedniej strukturze:

```
df = pd.read_csv('./marathon_results_2016.csv', index_col='Bib',
                 usecols=['Bib', '40K', 'Half', 'Pace', 'Age', 'M/F', 'Country', 'State', 'City'])

df['40K'] = df['40K'].apply(pd.to_timedelta)
df['Half'] = df['Half'].apply(pd.to_timedelta)

df['TotalSeconds'] = df['40K'].apply(lambda x: timedelta.total_seconds(x))
df['HalfSeconds'] = df['Half'].apply(lambda x: timedelta.total_seconds(x))

df = df.pivot_table(index="Age", columns="M/F", values="TotalSeconds").head()
df.head()
```

3. Usuń indeks z obiektu df
4. Zamień dane do postaci tabeli korzystając z polecenia melt definiując kolumnę **Age** jako kolumnę indeksu
5. Dodaj do poprzedniego polecenia parametr, który spowoduje, że kolumna z wartościami będzie nazwana **TotalSeconds**
6. Dodaj do poprzedniego polecenia parametr, który spowoduje, że nowo utworzona kolumna będzie miała nagłówek **Sex**
7. Podobnie jak w punkcie drugim wykonaj następujący kod, który spowoduje utworzenie nieco innego obiektu w postaci pivot table:

```
df = pd.read_csv('./marathon_results_2016.csv', index_col='Bib',
                 usecols=['Bib', '40K', 'Half', 'Pace', 'Age', 'M/F', 'Country', 'State', 'City'])

df['40K'] = df['40K'].apply(pd.to_timedelta)
df['Half'] = df['Half'].apply(pd.to_timedelta)

df['TotalSeconds'] = df['40K'].apply(lambda x: timedelta.total_seconds(x))
df['HalfSeconds'] = df['Half'].apply(lambda x: timedelta.total_seconds(x))

df = df.pivot_table(index="Age", columns="M/F", values=["HalfSeconds", "TotalSeconds"]).head()
df.head()
```

8. Usuń indeks z obiektu df
9. Zamień dane do postaci tabeli korzystając z polecenia melt definiując kolumnę **Age** jako kolumnę indeksu
10. Zmień poprzednie polecenie tak, aby kolumna z wartościami (aktualnie nazwana value) zmieniła nazwę na **Time**

Dane pochodzą z <https://github.com/llimllib/bostonmarathon> (<https://github.com/llimllib/bostonmarathon>)
<https://www.kaggle.com/rojour/boston-marathon-2016-finishers-analysis/data>
(<https://www.kaggle.com/rojour/boston-marathon-2016-finishers-analysis/data>)

Rozwiązania:

Poniżej znajdują się propozycje rozwiązań zadań. Prawdopodobnie istnieje wiele dobrych rozwiązań, dlatego jeżeli rozwiązujesz zadania samodzielnie, to najprawdopodobniej zrobisz to inaczej, może nawet lepiej :)
Możesz pochwalić się swoimi rozwiązaniami w sekcji Q&A

In [1]:

```
import pandas as pd
import numpy as np
from datetime import datetime
from datetime import timedelta
import time
```

In [2]:

```
df = pd.read_csv('./marathon_results_2016.csv', index_col='Bib',
                 usecols=['Bib', '40K', 'Half', 'Pace', 'Age', 'M/F', 'Country', 'State', 'City'])

df['40K'] = df['40K'].apply(pd.to_timedelta)
df['Half'] = df['Half'].apply(pd.to_timedelta)

df['TotalSeconds'] = df['40K'].apply(lambda x: timedelta.total_seconds(x))
df['HalfSeconds'] = df['Half'].apply(lambda x: timedelta.total_seconds(x))

df = df.pivot_table(index="Age", columns="M/F", values="TotalSeconds").head()
df.head()
```

Out[2]:

M/F	F	M
Age		
18	16050.666667	14554.615385
19	15351.958333	12156.529412
20	13835.931818	12111.100000
21	14681.027027	12408.360465
22	14366.421053	11872.666667

In [3]:

```
df.reset_index(inplace=True)
df.head()
```

Out[3]:

	M/F	Age	F	M
0	18	16050.666667	14554.615385	
1	19	15351.958333	12156.529412	
2	20	13835.931818	12111.100000	
3	21	14681.027027	12408.360465	
4	22	14366.421053	11872.666667	

In [4]:

```
df.melt(id_vars="Age").head()
```

Out[4]:

	Age	M/F	value
0	18	F	16050.666667
1	19	F	15351.958333
2	20	F	13835.931818
3	21	F	14681.027027
4	22	F	14366.421053

In [5]:

```
df.melt(id_vars="Age", value_name="TotalSeconds").head()
```

Out[5]:

	Age	M/F	TotalSeconds
0	18	F	16050.666667
1	19	F	15351.958333
2	20	F	13835.931818
3	21	F	14681.027027
4	22	F	14366.421053

In [6]:

```
df.melt(id_vars="Age", value_name="TotalSeconds", var_name="Sex").head()
```

Out[6]:

	Age	Sex	TotalSeconds
0	18	F	16050.666667
1	19	F	15351.958333
2	20	F	13835.931818
3	21	F	14681.027027
4	22	F	14366.421053

In [7]:

```
df = pd.read_csv('./marathon_results_2016.csv', index_col='Bib',
                 usecols=['Bib', '40K', 'Half', 'Pace', 'Age', 'M/F', 'Country', 'State', 'City'])

df['40K'] = df['40K'].apply(pd.to_timedelta)
df['Half'] = df['Half'].apply(pd.to_timedelta)

df['TotalSeconds'] = df['40K'].apply(lambda x: timedelta.total_seconds(x))
df['HalfSeconds'] = df['Half'].apply(lambda x: timedelta.total_seconds(x))

df = df.pivot_table(index="Age", columns="M/F", values=["HalfSeconds", "TotalSeconds"]).head()
df.head()
```

Out[7]:

	HalfSeconds		TotalSeconds	
	M/F	F	M	F
Age				
18	7999.888889	7078.538462	16050.666667	14554.615385
19	7473.666667	5817.764706	15351.958333	12156.529412
20	6871.045455	5716.800000	13835.931818	12111.100000
21	7148.554054	5929.627907	14681.027027	12408.360465
22	7091.097744	5751.000000	14366.421053	11872.666667

In [8]:

```
df.reset_index(inplace=True)
df.head()
```

Out[8]:

M/F	Age		HalfSeconds		TotalSeconds	
			F	M	F	M
0	18	7999.888889	7078.538462	16050.666667	14554.615385	
1	19	7473.666667	5817.764706	15351.958333	12156.529412	
2	20	6871.045455	5716.800000	13835.931818	12111.100000	
3	21	7148.554054	5929.627907	14681.027027	12408.360465	
4	22	7091.097744	5751.000000	14366.421053	11872.666667	

In [9]:

```
df.melt(id_vars="Age").head()
```

Out[9]:

Age		NaN	M/F	value
0	18	HalfSeconds	F	7999.888889
1	19	HalfSeconds	F	7473.666667
2	20	HalfSeconds	F	6871.045455
3	21	HalfSeconds	F	7148.554054
4	22	HalfSeconds	F	7091.097744

In [10]:

```
df.melt(id_vars="Age", value_name='Time').head()
```

Out[10]:

Age		NaN	M/F	Time
0	18	HalfSeconds	F	7999.888889
1	19	HalfSeconds	F	7473.666667
2	20	HalfSeconds	F	6871.045455
3	21	HalfSeconds	F	7148.554054
4	22	HalfSeconds	F	7091.097744

In []:

In []: