

# Python - Analiza danych z modulem PANDAS

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

## S02-L007 - LAS02-L007B - Bardziej zaawansowane filtrowanie

1. Zaimportuj moduły: pandas, numpy, matplotlib (tylko pyplot), math i nadaj im standardowe aliasy
2. Na stronie [https://en.wikipedia.org/wiki/European\\_countries\\_by\\_electricity\\_consumption\\_per\\_person](https://en.wikipedia.org/wiki/European_countries_by_electricity_consumption_per_person) ([https://en.wikipedia.org/wiki/European\\_countries\\_by\\_electricity\\_consumption\\_per\\_person](https://en.wikipedia.org/wiki/European_countries_by_electricity_consumption_per_person)) znajduje się informacja o zużyciu energii elektrycznej na osobę w Europie. Utwórz następujące zmienne:
  - **namesList** (nazwy krajów) o wartościach: ['Albania','Austria','Belarus',  
'Belgium','Bulgaria','Croatia','Cyprus','Czech Republic','Denmark','Estonia',  
'Finland','France','Germany','Greece','Hungary','Iceland','Ireland','Italy',  
'Latvia','Lithuania','Luxembourg','Macedonia','Malta','Montenegro','Netherlands',  
'Norway','Poland','Portugal','Romania','Russia','Serbia','Slovenia','Spain', 'Sweden','Switzerland','United Kingdom','Turkey','Ukraine']
  - **energy2010List** (zużycie energii na osobę w 2010 roku) o wartościach:  
[1947,8347,3564,8369,4560,3814,4623,6348,6328,6506,16483,7736,7264,5318,3876,  
51440,5911,5494,3230,3471,16830,3521,4171,5420,7010,24891,3797,4959,2551,  
6410,4359,6521,5707,14934,8175,2498,3550,5701]
  - **energy2012List** (zużycie energii na osobę w 2012 roku) o wartościach:  
[2118,8507,3698,7987,4762,3819,4057,6305,6039,6689,15687,7344,7270,5511,3919,  
53203,5665,5398,3588,3608,14696,3626,4761,5416,6871,23658,3899,4736,2604,  
6617,4387,6778,5573,14290,7886,2794,3641,5452]
3. Dla każdej z tych zmiennych utwórz obiekt serii o nazwach jak niżej, a potem wyświetl je:
  - namesSeries
  - energy2010Series
  - energy2012Series
4. W zmiennej mean2010 zapisz średnią wartość zużycia prądu w całej Europie w 2010 roku. Wyświetl tą wartość
5. W zmiennej mean2012 zapisz średnią wartość zużycia prądu w całej Europie w 2012 roku. Wyświetl tą wartość
6. W zmiennej filterAboveMean2010 zapisz wynik porównania wartości serii energy2010Series z wartością średnią mean2010. W porównaniu badaj, czy wartość zużycia prądu w danym kraju jest większa od wartości średniej za rok 2010
7. W zmiennej filterAboveMean2012 zapisz wynik porównania wartości serii energy2012Series z wartością średnią mean2012. W wyniku porównania badaj, czy wartość zużycia prądu w danym kraju jest większa od wartości średniej za rok 2012
8. Wyświetl kraje z serii namesSeries, które w roku 2010 i w roku 2012 zużywały prąd w ilości powyżej średniej za dany rok
9. W zmiennej filterBelowMean2010 zapisz wynik porównania wartości serii energy2010Series z wartością średnią mean2010. W porównaniu badaj, czy wartość zużycia prądu w danym kraju jest mniejsza od wartości średniej za rok 2010
10. Sprawdź czy są kraje, które w 2010 roku zużywały mniej prądu niż średnia za 2010, ale w 2012 znalazły się powyżej tej średniej
11. Czy z tej analizy można wyciągnąć jakieś wnioski (nawet dość oczywiste)?

## Rozwiązania:

Poniżej znajdują się propozycje rozwiązań zadań. Prawdopodobnie istnieje wiele dobrych rozwiązań, dlatego jeżeli rozwiążesz 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
import matplotlib.pyplot as plt
import math as math
```

```
In [2]: namesList = ['Albania', 'Austria', 'Belarus', 'Belgium', 'Bulgaria', 'Croatia', 'Cyprus',
                    'Czech Republic', 'Denmark', 'Estonia', 'Finland', 'France', 'Germany', 'Greece',
                    'Hungary', 'Iceland', 'Ireland', 'Italy', 'Latvia', 'Lithuania', 'Luxembourg',
                    'Macedonia', 'Malta', 'Montenegro', 'Netherlands', 'Norway', 'Poland', 'Portugal',
                    'Romania', 'Russia', 'Serbia', 'Slovenia', 'Spain', 'Sweden', 'Switzerland',
                    'United Kingdom', 'Turkey', 'Ukraine']

namesSeries = pd.Series(namesList)
namesSeries
```

```
Out[2]: 0      Albania
1      Austria
2      Belarus
3      Belgium
4      Bulgaria
5      Croatia
6      Cyprus
7      Czech Republic
8      Denmark
9      Estonia
10     Finland
11     France
12     Germany
13     Greece
14     Hungary
15     Iceland
16     Ireland
17     Italy
18     Latvia
19     Lithuania
20     Luxembourg
21     Macedonia
22     Malta
23     Montenegro
24     Netherlands
25     Norway
26     Poland
27     Portugal
28     Romania
29     Russia
30     Serbia
31     Slovenia
32     Spain
33     Sweden
34     Switzerland
35     United Kingdom
36     Turkey
37     Ukraine
dtype: object
```

```
In [3]: energy2010List = [1947, 8347, 3564, 8369, 4560, 3814, 4623, 6348, 6328, 6506, 16483, 7736,  
                          7264, 5318, 3876, 51440, 5911, 5494, 3230, 3471, 16830, 3521, 4171, 5420,  
                          7010, 24891, 3797, 4959, 2551, 6410, 4359, 6521, 5707, 14934, 8175, 2498,  
                          3550, 5701]  
energy2010Series = pd.Series(energy2010List)  
energy2010Series
```

```
Out[3]: 0      1947  
        1      8347  
        2      3564  
        3      8369  
        4      4560  
        5      3814  
        6      4623  
        7      6348  
        8      6328  
        9      6506  
       10     16483  
       11      7736  
       12      7264  
       13      5318  
       14      3876  
       15     51440  
       16      5911  
       17      5494  
       18      3230  
       19      3471  
       20     16830  
       21      3521  
       22      4171  
       23      5420  
       24      7010  
       25     24891  
       26      3797  
       27      4959  
       28      2551  
       29      6410  
       30      4359  
       31      6521  
       32      5707  
       33     14934  
       34      8175  
       35      2498  
       36      3550  
       37      5701  
dtype: int64
```

```
In [4]: energy2012List = [2118,8507,3698,7987,4762,3819,4057,6305,6039,6689,15687,7344,
                        7270,5511,3919,53203,5665,5398,3588,3608,14696,3626,4761,5416,
                        6871,23658,3899,4736,2604,6617,4387,6778,5573,14290,7886,2794,
                        3641,5452]
energy2012Series = pd.Series(energy2012List)
energy2012Series
```

```
Out[4]: 0      2118
        1      8507
        2      3698
        3      7987
        4      4762
        5      3819
        6      4057
        7      6305
        8      6039
        9      6689
       10     15687
       11      7344
       12      7270
       13      5511
       14      3919
       15     53203
       16      5665
       17      5398
       18      3588
       19      3608
       20     14696
       21      3626
       22      4761
       23      5416
       24      6871
       25     23658
       26      3899
       27      4736
       28      2604
       29      6617
       30      4387
       31      6778
       32      5573
       33     14290
       34      7886
       35      2794
       36      3641
       37      5452
dtype: int64
```

```
In [5]: mean2010 = energy2010Series.mean()
mean2010
```

```
Out[5]: 7779.8421052631575
```

```
In [6]: mean2012 = energy2012Series.mean()
mean2012
```

```
Out[6]: 7706.815789473684
```

```
In [7]: filterAboveMean2010 = energy2010Series > mean2010
```

```
In [8]: filterAboveMean2012 = energy2012Series > mean2012
```

```
In [9]: namesSeries.where(filterAboveMean2010 & filterAboveMean2012).dropna()
```

```
Out[9]: 1      Austria
        3      Belgium
        10     Finland
        15     Iceland
        20    Luxembourg
        25     Norway
        33     Sweden
        34    Switzerland
        dtype: object
```

```
In [10]: filterBelowMean2010 = energy2010Series < mean2010
        namesSeries.where(filterBelowMean2010 & filterAboveMean2012).dropna()
```

```
Out[10]: Series([], dtype: object)
```

Wnioski: kraje w których konsumpcja prądu była wysoka, nadal pozostała wysoka. Kraje rozwinięte (zużywające dużo prądu) nie cofają się.

```
In [ ]:
```