

Python - Analiza danych z modulem PANDAS

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

LAB - S02-L008 - Import danych do obiektu series

1. Zaimportuj moduły: pandas, numpy, matplotlib (tylko pyplot), math i nadaj im standardowe aliasy
2. Korzystając z pliku StackOverflowDeveloperSurvey.csv zaimportuj do zmiennej education kolumnę o nazwie "FormalEducation" jako jednokolumnowy obiekt Data Frame
3. Sprawdź typ otrzymanej zmiennej.
4. Wyświetl tylko 5 pierwszych wierszy z education.
5. Powtórz akcję tak jak w pkt. 2, ale tym razem zadбай o to , aby education było typu data series.
6. Sprawdź typ otrzymanej zmiennej.
7. Wyświetl 10 ostatnich wierszy z education
8. Do zmiennej allInfo zaimportuj cały plik danych (wszystkie kolumny), jako Data Frame
9. Zainicjuj obiekt country typu data series, pobierając dane z kolumny "Country" obiektu allInfo.
10. Wyświetl 5 pierwszych wierszy obiektu country
11. (zastosowanie i mała powtórka z poprzednich lekcji) Utwórz serię filterOnlyUSA, która będzie inicjowana w oparciu o serię countries. filterOnlyUSA ma zawierać w wierszu wartość True , jeżeli w serii countries wartość to "United States"
12. Wyświetl 5 pierwszych pozycji z serii filterOnlyUSA
13. Wyświetl dane z serii education w taki sposób, że:
 - pokazywane będą tylko informacje o ukończonej szkole tylko dla osób z country = "United States"
 - wartości NaN mają być usunięte
 - należy pokazać tylko 10 górnych wierszy
 - wszystko to powinno być zrobione w jednym poleceniu (method chaining)

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
import matplotlib.pyplot as plt
import math as math
```

```
In [2]: education = pd.read_csv("StackOverflowDeveloperSurvey.csv",
```

```
usecols=['FormalEducation'])
```

```
In [3]: type(education)
```

```
Out[3]: pandas.core.frame.DataFrame
```

```
In [4]: education.head()
```

```
Out[4]:
```

	FormalEducation
0	Secondary school
1	Some college/university study without earning ...
2	Bachelor's degree
3	Doctoral degree
4	Master's degree

```
In [5]: education = pd.read_csv("StackOverflowDeveloperSurvey.csv",
                                usecols=['FormalEducation'],
                                squeeze=True)
```

```
In [6]: type(education)
```

```
Out[6]: pandas.core.series.Series
```

```
In [7]: education.tail(10)
```

```
Out[7]:
```

51382	Master's degree
51383	I never completed any formal education
51384	Bachelor's degree
51385	Master's degree
51386	Some college/university study without earning ...
51387	Bachelor's degree
51388	Master's degree
51389	Some college/university study without earning ...
51390	Bachelor's degree
51391	Bachelor's degree

Name: FormalEducation, dtype: object

```
In [8]: allInfo = pd.read_csv("StackOverflowDeveloperSurvey.csv")
```

```
In [9]: country = allInfo["Country"]
```

```
In [10]: country.head()
```

```
Out[10]:
```

0	United States
1	United Kingdom
2	United Kingdom
3	United States
4	Switzerland

Name: Country, dtype: object

```
In [11]: filterOnlyUSA = country == "United States"
```

```
In [12]: filterOnlyUSA.head()
```

```
Out[12]:
```

```
0      True
1     False
2     False
3      True
4     False
```

```
In [13]: education.where(filterOnlyUSA).dropna().head(10)
```

```
Out[13]: 0      Secondary school
          3      Doctoral degree
          6      Master's degree
          10     Bachelor's degree
          15     Primary/elementary school
          17     Bachelor's degree
          18     Bachelor's degree
          25     Master's degree
          27     Bachelor's degree
          29     Bachelor's degree
          Name: FormalEducation, dtype: object
```