

Python - Analiza danych z modulem PANDAS

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

LAB - S02-L009 - Standardowe metody Pythona a obiekty PANDAS

1. Zaimportuj moduły: pandas, numpy, matplotlib (tylko pyplot), math i nadaj im standardowe aliasy
2. Z pliku StackOverflowDeveloperSurvey.csv wczytaj kolumnę Salary jako Data Series i zapisz ją w zmiennej salary. Wczytując elementy pomiń te ankiety, w których respondenci nie podali kwoty wynagrodzenia (użyj dropna()). Wyświetl 5 pierwszych elementów serii
3. Ile osób odpowiedziało w ankiecie na temat otrzymywanego wynagrodzenia?
4. Jakie jest najmniejsze wynagrodzenie? A jakie jest największe wynagrodzenie?
5. Pięć pierwszych elementów serii zamień na listę (nie zapisuj jej w zmiennej, po prostu wyświetl)
6. Pięć pierwszych elementów serii zamień na słownik (nie zapisuj go, po prostu wyświetl)
7. W zmiennej listSalarySorted zapisz listę powstałą z posortowania serii salary malejąco
8. Wyświetl pięć pierwszych elementów listy listSalarySorted. (Pamiętaj że to nie seria i nie zadziała metoda head())
9. Zmień nazwę atrybutu name dla serii salary na "Salary of a person"
10. Wyświetl 5 pierwszych elementów serii salary

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]: salary = pd.read_csv("StackOverflowDeveloperSurvey.csv", usecols=["Salary"], squeeze=True)
salary.head()
```

```
Out[2]: 2      113750.0
14      100000.0
17      130000.0
18       82500.0
22      100764.0
Name: Salary, dtype: float64
```

```
In [3]: len(salary)
```

```
Out[3]: 12891
```

```
In [4]: min(salary)
```

```
Out[4]: 0.0
```

```
In [5]: max(salary)
```

```
Out[5]: 197000.0
```

```
In [6]: list(salary.head())
```

```
Out[6]: [113750.0, 100000.0, 130000.0, 82500.0, 100764.0]
```

```
In [7]: dict(salary.head())
```

```
Out[7]: {2: 113750.0, 14: 100000.0, 17: 130000.0, 18: 82500.0, 22: 100764.0}
```

```
In [8]: listSalarySorted = sorted(salary,reverse=True)
```

```
In [9]: listSalarySorted[:5]
```

```
Out[9]: [197000.0, 195000.0, 195000.0, 195000.0, 195000.0]
```

```
In [10]: salary.name = 'Salary of a person'
```

```
In [11]: salary.head()
```

```
Out[11]: 2      113750.0
14      100000.0
17      130000.0
18       82500.0
22      100764.0
Name: Salary of a person, dtype: float64
```