

Python - Analiza danych z modulem PANDAS

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

LAB - S02-L010 - Sortowanie danych data series

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 pominięte te ankiety, w których respondenci nie podali kwoty wynagrodzenia (użyj dropna()). Wyświetl 5 pierwszych elementów serii
3. Posortuj dane wg wynagrodzenia malejąco (wyświetl tylko 5 pierwszych pozycji).
4. Posortuj dane wg wynagrodzenia rosnąco (wyświetl tylko 5 pierwszych pozycji).
5. Zmień parametry sortowania tak, aby lista została posortowana wewnętrznie na stałe wg wynagrodzenia malejąco. Wyświetl kilka początkowych elementów, aby przekonać się że obiekt jest teraz posortowany jak należy.
6. Posortuj dane wg indeksu malejąco
7. Zapisz w zmiennej maxSalaries 100 największych wynagrodzeń z serii Salary
8. Zapisz w zmiennej minSalaries 100 najniższych wynagrodzeń z serii Salary
9. Wyznacz średnią wartość z wynagrodzeń znajdujących się w serii maxSalaries
10. Wyznacz średnią wartość z wynagrodzeń znajdujących się w serii minSalaries

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

```
Out[3]: 31600      197000.0
        27943      195000.0
        6743       195000.0
        41398      195000.0
        1696       195000.0
        Name: Salary, dtype: float64
```

```
In [4]: salary.sort_values().head()
```

```
Out[4]: 51144      0.0
        23953      0.0
        23907      0.0
        44841      0.0
        2951       0.0
        Name: Salary, dtype: float64
```

```
In [5]: salary.sort_values(inplace=True,ascending=False)
salary.head()
```

```
Out[5]: 31600      197000.0
        27943      195000.0
        6743       195000.0
        41398      195000.0
        1696       195000.0
        Name: Salary, dtype: float64
```

```
In [6]: salary.sort_index(ascending=False).head()
```

```
Out[6]: 51390      40000.000000
        51387      58000.000000
        51382      32258.064516
        51378     107526.881720
        51371      74193.548387
        Name: Salary, dtype: float64
```

```
In [7]: maxSalaries = salary.sort_values(ascending=False).head(100)
```

```
In [8]: minSalaries = salary.sort_values().head(100)
```

```
In [9]: maxSalaries.mean()
```

```
Out[9]: 180198.1045586777
```

```
In [10]: minSalaries.mean()
```

```
Out[10]: 67.08388907133795
```