

Python - Analiza danych z modułem PANDAS

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

LAB - S02-L015 - Określanie indeksu podczas wczytywania pliku CSV

1. Zaimportuj moduł pandas i nadaj mu standardowy alias.
2. Do zmiennej fortune500 wczytaj dane z pliku Fortune_500_2017.csv. Wczytaj kolumny Rank i Title. Kolumna Rank powinna stać się indeksem. Wyświetl nagłówek otrzymanej serii danych.
3. Wyświetl 10 pierwszych firm znajdujących się najwyżej w klasyfikacji Fortune 500
4. Wyświetl 20 ostatnich firm znajdujących się w klasyfikacji Fortune 500
5. Wczytaj jeszcze raz plik, ale tym razem indeksem ma być kolumna Title, a wartościami kolumna Employees opisująca ilość pracowników w każdej firmie.
6. Sprawdź ilu pracowników pracuje dla IBM, Alphabet (Google), Facebook i Apple
7. Wyświetl informacje o ilości pracowników dla firm od IBM do Intel

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
```

```
In [2]: #fortune500 = pd.read_csv("Fortune_500_2017.csv", usecols=["Rank", "Title"],
#                               index_col="Rank", squeeze=True)
fortune500 = pd.read_csv("Fortune_500_2017.csv", usecols=["Rank", "Title"],
                        index_col="Rank").squeeze()

fortune500.head()
```

```
Out[2]: Rank
1          Walmart
2  Berkshire Hathaway
3           Apple
4       Exxon Mobil
5       McKesson
Name: Title, dtype: object
```

```
In [3]: fortune500[:10]
```

```
Out[3]: Rank
1          Walmart
2    Berkshire Hathaway
3          Apple
4        Exxon Mobil
5        McKesson
6    UnitedHealth Group
7        CVS Health
8    General Motors
9          AT&T
10         Ford Motor
Name: Title, dtype: object
```

```
In [4]: fortune500[-20:]
```

```
Out[4]: Rank
481        Booz Allen Hamilton Holding
482                        Chemours
483    Western & Southern Financial Group
484                        Celanese
485        Windstream Holdings
486                        Seaboard
487                        Essendant
488                        Apache
489                        Airgas
490        Kelly Services
491        Liberty Media
492        Rockwell Collins
493    Robert Half International
494                        CH2M Hill
495                        Big Lots
496        Michaels Cos.
497        Toll Brothers
498                        Yahoo
499        Vistra Energy
500        ABM Industries
Name: Title, dtype: object
```

```
In [5]: #fortune500 = pd.read_csv("Fortune_500_2017.csv", usecols=["Title", "Employees"],
#                                     index_col="Title", squeeze=True)
fortune500 = pd.read_csv("Fortune_500_2017.csv", usecols=["Title", "Employees"],
                           index_col="Title").squeeze()

fortune500.head()
```

```
Out[5]: Title
Walmart          2300000
Berkshire Hathaway 367700
Apple             116000
Exxon Mobil       72700
McKesson          68000
Name: Employees, dtype: int64
```

```
In [6]: fortune500['IBM']
```

```
Out[6]: 414400
```

```
In [7]: fortune500['Apple']
```

```
Out[7]: 116000
```

```
In [8]: fortune500['Alphabet']
```

```
Out[8]: 72053
```

```
In [9]: fortune500['Facebook']
```

```
Out[9]: 17048
```

```
In [10]: fortune500['IBM':'Intel']
```

```
Out[10]: Title
          IBM          414400
          State Farm Insurance Cos.    68234
          Phillips 66          14800
          Johnson & Johnson    126400
          Procter & Gamble    105000
          Valero Energy          9996
          Target          323000
          Freddie Mac          5982
          Lowe's          240000
          Dell Technologies    138000
          MetLife          58000
          Aetna          49500
          PepsiCo          264000
          Archer Daniels Midland    31800
          UPS          335520
          Intel          106000
          Name: Employees, dtype: int64
```