

# Python - Analiza danych z modulem PANDAS

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

## LAB - S07-L004 - metoda merge

1. Zaimportuj moduł pandas i numpy nadaj im standardowe aliasy
2. Do zmiennej **f\_2015** zaimportuj z pliku **Fortune\_500\_2015.csv** informacje o firmach zakwalifikowanych na listę Fortune 500 w roku 2015. Wybierz tylko kolumny **rank** i **company**
3. Do zmiennej **f\_2016** zaimportuj z pliku **Fortune\_500\_2016.csv** informacje o firmach zakwalifikowanych na listę Fortune 500 w roku 2016. Wybierz tylko kolumny **rank** i **company**
4. W tym zadaniu będziemy się zastanawiać, czy są takie firmy, które:
  - znalazły się na liście w 2015 i w 2016
  - znalazły się na liście w 2015, ale nie w 2016
  - znalazły się na liście w 2016, ale nie w 2015
5. W zmiennej **companies\_2015\_2016** zapamiętaj wynik złączenia **f\_2015** i **f\_2016** łącząc ze sobą tylko te wiersze, które można odnaleźć jednocześnie w obu data frame. Złączenie ma się odbyć w oparciu o kolumnę **company**
6. W wyniku złączenia powinny się znajdować kolumny określające pozycję w rankingu w 2015 i 2016 roku. Korzystając z odpowiedniego parametru spraw, aby ranking za rok 2015 znajdował się w kolumnie o nazwie **rank\_2015**, a ranking za rok 2016 w kolumnie **rank\_2016**
7. Wyświetl z **companies\_2015\_2016** wiersz odpowiadający za pozycję firmy **Microsoft** w 2015 i 2016 roku
8. Teraz w zmiennej **companies\_2015\_2016** zapamiętaj wynik złączenia **f\_2015** i **f\_2016** łącząc ze sobą tylko te wiersze, które można odnaleźć jednocześnie w obu data frame lub tylko w 2015 lub tylko w 2016. Złączenie ma się odbyć w oparciu o kolumnę **company**. Polecenie powinno zwrócić również informacje o pochodzeniu wiersza (kolumna **\_merge**)
9. W zmiennej **only\_2015** zapamiętaj dane tylko tych firm, które znalazły się na liście Fortune 500 w roku 2015, ale nie znalazły się na niej w 2016
10. W zmiennej **only\_2016** zapamiętaj dane tylko tych firm, które znalazły się na liście Fortune 500 w roku 2016, ale nie znalazły się na niej w 2015

## 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
```

```
In [2]: f_2015 = pd.read_csv("Fortune_500_2015.csv", usecols=['rank', 'company'])
        f_2015.head()
```

Out[2]:

	rank	company
0	1	Walmart
1	2	Exxon Mobil
2	3	Chevron
3	4	Berkshire Hathaway
4	5	Apple

```
In [3]: f_2016 = pd.read_csv("Fortune_500_2016.csv", usecols=['rank', 'company'])
        f_2016.head()
```

Out[3]:

	rank	company
0	1	Walmart
1	2	Exxon Mobil
2	3	Apple
3	4	Berkshire Hathaway
4	5	McKesson

```
In [4]: companies_2015_2016 = f_2015.merge(f_2016, on="company", how="inner")
        companies_2015_2016.head()
```

Out[4]:

	rank_x	company	rank_y
0	1	Walmart	1
1	2	Exxon Mobil	2
2	3	Chevron	14
3	4	Berkshire Hathaway	4
4	5	Apple	3

```
In [5]: companies_2015_2016 = f_2015.merge(f_2016, on="company", how="inner", suffixes=['_2015',
        companies_2015_2016.head())
```

Out[5]:

	rank_2015	company	rank_2016
0	1	Walmart	1
1	2	Exxon Mobil	2
2	3	Chevron	14
3	4	Berkshire Hathaway	4
4	5	Apple	3

```
In [6]: companies_2015_2016[companies_2015_2016["company"] == "Microsoft"]
```

Out[6]:

	rank_2015	company	rank_2016
30	31	Microsoft	25

```
In [7]: companies_2015_2016 = f_2015.merge(f_2016, on="company", how="outer", suffixes=['_2015',
companies_2015_2016.head()
```

Out[7]:

	rank_2015	company	rank_2016	_merge
0	1.0	Walmart	1.0	both
1	2.0	Exxon Mobil	2.0	both
2	3.0	Chevron	14.0	both
3	4.0	Berkshire Hathaway	4.0	both
4	5.0	Apple	3.0	both

```
In [8]: only_2015 = companies_2015_2016[companies_2015_2016["_merge"] == "left_only"]
```

```
In [9]: only_2015.head()
```

Out[9]:

	rank_2015	company	rank_2016	_merge
83	84.0	Safeway	NaN	left_only
94	95.0	DirectTV	NaN	left_only
124	125.0	Alcoa	NaN	left_only
164	165.0	Kraft	NaN	left_only
174	175.0	TRW Automotive Holdings	NaN	left_only

```
In [10]: only_2016 = companies_2015_2016[companies_2015_2016["_merge"] == "right_only"]
```

```
In [11]: only_2016.head()
```

Out[11]:

	rank_2015	company	rank_2016	_merge
500	NaN	Arconic	126.0	right_only
501	NaN	Kraft Heinz	153.0	right_only
502	NaN	Performance Food Group	185.0	right_only
503	NaN	Hertz Global Holdings	269.0	right_only
504	NaN	PayPal Holdings	307.0	right_only

```
In [ ]:
```