

Python - Analiza danych z modulem PANDAS

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

LAB - S07-L003 - metoda join

1. Zaimportuj moduł pandas i numpy nadaj im standardowe aliasy
2. Do zmiennej **customer_names** zaimportuj z pliku **customers.csv** kolumny: **CustomerID** i **CompanyName**
3. Do zmiennej **customer_cities** zaimportuj z pliku **customers.csv** kolumny: **CustomerID**, **City** i **Country**
4. Wyświetl ilość elementów w obu tych zmiennych
5. Dodaj do obu zmiennych indeks na kolumnie **CustomerID**
6. Korzystając z polecenia **join** połącz dane z obu data frame. Dane mają być połączone "po indeksie"
7. Ile wierszy jest w wyniku złączenia?
8. Aby przećwiczyć inny rodzaj złączenia danych zresetuj indeks w **customer_names**
9. Korzystając z polecenia **join** połącz dane z obu data frame, ale tym razem wskaż jawnie, że do połączenia ma być wykorzystywana kolumna **CustomerID** z data frame **customer_names**
10. Ile jest wierszy w wyniku połączenia?

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 [13]: customer_names = pd.read_csv("northwind-mongo-master/customers.csv", usecols=['CustomerID', 'CompanyName'])
```

```
In [14]: customer_cities = pd.read_csv("northwind-mongo-master/customers.csv", usecols=['CustomerID', 'City', 'Country'])
```

```
In [21]: len(customer_names)
```

```
Out[21]: 91
```

```
In [22]: len(customer_cities)
```

```
Out[22]: 91
```

```
In [15]: customer_names.set_index('CustomerID', inplace=True)
```

```
In [16]: customer_cities.set_index('CustomerID', inplace=True)
```

```
In [18]: customer_names.join(customer_cities).head()
```

```
Out[18]:
```

	CompanyName	City	Country
CustomerID			
ALFKI	Alfreds Futterkiste	Berlin	Germany
ANATR	Ana Trujillo Emparedados y helados	México D.F.	Mexico
ANTON	Antonio Moreno Taquería	México D.F.	Mexico
AROUT	Around the Horn	London	UK
BERGS	Berglunds snabbköp	Luleå	Sweden

```
In [24]: len(customer_names.join(customer_cities))
```

```
Out[24]: 91
```

```
In [19]: customer_names.reset_index(inplace=True)
```

```
In [27]: customer_names.join(customer_cities, on='CustomerID').head()
```

```
Out[27]:
```

	CustomerID	CompanyName	City	Country
0	ALFKI	Alfreds Futterkiste	Berlin	Germany
1	ANATR	Ana Trujillo Emparedados y helados	México D.F.	Mexico
2	ANTON	Antonio Moreno Taquería	México D.F.	Mexico
3	AROUT	Around the Horn	London	UK
4	BERGS	Berglunds snabbköp	Luleå	Sweden

```
In [26]: len(customer_names.join(customer_cities, on='CustomerID'))
```

```
Out[26]: 91
```

```
In [ ]:
```