

Python - Analiza danych z modulem PANDAS

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

LAB - S07-L007 - opcje polecenia merge

1. Zaimportuj moduł pandas i numpy nadaj im standardowe aliasy
2. Do zmiennej **suppliers** zaimportuj zawartość pliku **suppliers.csv**. Pobierz tylko kolumny **SupplierID**, **CompanyName**, **City**
3. Do zmiennej **products** zaimportuj zawartość pliku **products.csv**. Pobierz tylko kolumny **ProductID**, **ProductName**, **SupplierID****
4. W kolejnych zadaniach będziesz łączyć ze sobą **suppliers** i **products**, a do połączenia wykorzystasz kolumnę **SupplierID**. Kolumna ta może delikatnie zmienić nazwę lub stać się indeksem. W zależności od tego musisz w odpowiedni sposób zmieniać parametry i wybierać odpowiednie polecenia:
5. Do data frame **suppliers** dołącz dane z **products** (tak jak to opisano powyżej)
6. W **suppliers** zdefiniuj **SupplierID** jako indeks (pamiętaj o parametrze inplace). Następnie do data frame **suppliers** dołącz dane z **products**
7. W **products** zdefiniuj **SupplierID** jako indeks (pamiętaj o parametrze inplace). Następnie do data frame **suppliers** dołącz dane z **products**
8. W **suppliers** zresetuj indeks (pamiętaj o parametrze inplace). Następnie do data frame **suppliers** dołącz dane z **products**
9. W **products** zresetuj indeks (pamiętaj o parametrze inplace), a potem zmień nazwę kolumny **SupplierID** na **supplier_id**. Następnie do data frame **suppliers** dołącz dane z **products**

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]: suppliers = pd.read_csv("../northwind-mongo-master/suppliers.csv", usecols=["SupplierID",
```

```
In [3]: products = pd.read_csv("../northwind-mongo-master/products.csv", usecols=['ProductID', '
```

```
In [4]: suppliers.merge(products, on='SupplierID').head(3)
```

Out[4]:

	SupplierID	CompanyName	City	ProductID	ProductName
0	1	Exotic Liquids	London	1	Chai
1	1	Exotic Liquids	London	2	Chang
2	1	Exotic Liquids	London	3	Aniseed Syrup

```
In [5]: suppliers.set_index("SupplierID", inplace=True)
```

```
In [6]: suppliers.merge(products, left_index=True, right_on="SupplierID").head(3)
```

```
Out[6]:
```

	CompanyName	City	ProductID	ProductName	SupplierID
0	Exotic Liquids	London	1	Chai	1
1	Exotic Liquids	London	2	Chang	1
2	Exotic Liquids	London	3	Aniseed Syrup	1

```
In [7]: products.set_index("SupplierID", inplace=True)
```

```
In [8]: suppliers.join(products).head(3)
```

```
Out[8]:
```

	CompanyName	City	ProductID	ProductName
SupplierID				
1	Exotic Liquids	London	1	Chai
1	Exotic Liquids	London	2	Chang
1	Exotic Liquids	London	3	Aniseed Syrup

```
In [9]: suppliers.reset_index(inplace=True)
```

```
In [10]: suppliers.merge(products, left_on="SupplierID", right_index=True).head(3)
```

```
Out[10]:
```

	SupplierID	CompanyName	City	ProductID	ProductName
0	1	Exotic Liquids	London	1	Chai
0	1	Exotic Liquids	London	2	Chang
0	1	Exotic Liquids	London	3	Aniseed Syrup

```
In [11]: products.reset_index(inplace=True)
```

```
In [12]: products.rename({'SupplierID':'supplier_id'}, axis=1, inplace=True)
```

```
In [13]: suppliers.merge(products, left_on="SupplierID", right_on="supplier_id").head(3)
```

```
Out[13]:
```

	SupplierID	CompanyName	City	supplier_id	ProductID	ProductName
0	1	Exotic Liquids	London	1	1	Chai
1	1	Exotic Liquids	London	1	2	Chang
2	1	Exotic Liquids	London	1	3	Aniseed Syrup

```
In [ ]:
```