

Python - Analiza danych z modulem PANDAS

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

LAB - S07-L006 - left & right 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_del_2.csv**. Pobierz tylko kolumny **ProductID**, **ProductName**, **SupplierID****
4. Należy odnaleźć tych dostawców z **suppliers**, którzy nie dostarczają żadnych produktów. W tym celu wykonuj kolejne kroki
5. W zmiennej **sup_prod** zapisz wynik zwykłego złączenia **suppliers** i **products** w oparciu o kolumnę **SupplierID**. Skorzystaj z parametru **indicator**.
6. Wyświetl ilość elementów w **suppliers**, **products** i **sup_prod**. Powinieneś uzyskać wyniki: 29/68/68
7. W zmiennej **sup_prod_left** zapisz wynik złączenia typu **left** z **suppliers** do **products** w oparciu o kolumnę **SupplierID**. Skorzystaj z parametru **indicator**
8. Wyświetl ilość elementów w **suppliers**, **products** i **sup_prod**. Powinieneś uzyskać wyniki: 29/68/70. Oznacza to, że "idąc od suppliers do products", dla 2 dostawców nie udało się znaleźć żadnych produktów, które byłyby przez nich dostarczane.
9. Wyświetl te wiersze z **sup_prod_left**, które w kolumnie **"merged"** posiadają wartość **left_only**
10. Jeżeli chcesz tak przekształć operacje 7-9, aby to samo zadanie zrealizować korzystając z **right merge**

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 [15]: import pandas as pd
import numpy as np
```

```
In [16]: suppliers = pd.read_csv("../northwind-mongo-master/suppliers.csv", usecols=["SupplierID", "CompanyName", "City"])
```

```
In [17]: suppliers.head()
```

Out[17]:

	SupplierID	CompanyName	City
0	1	Exotic Liquids	London
1	2	New Orleans Cajun Delights	New Orleans
2	3	Grandma Kelly's Homestead	Ann Arbor
3	4	Tokyo Traders	Tokyo
4	5	Cooperativa de Quesos 'Las Cabras'	Oviedo

```
In [18]: products = pd.read_csv("../northwind-mongo-master/products_del_2.csv", usecols=["ProductID", "ProductName", "SupplierID"])
```

```
In [19]: products.head()
```

```
Out[19]:
```

	ProductID	ProductName	SupplierID
0	1	Chai	1
1	2	Chang	1
2	6	Grandma's Boysenberry Spread	3
3	7	Uncle Bob's Organic Dried Pears	3
4	8	Northwoods Cranberry Sauce	3

```
In [20]: sup_prod = suppliers.merge(products, how='inner', on='SupplierID', indicator=True)
```

```
In [21]: print(len(suppliers))
print(len(products))
print(len(sup_prod))
```

```
29
68
68
```

```
In [ ]:
```

```
In [8]: sup_prod_left = suppliers.merge(products, how='left', on='SupplierID', indicator=True)
```

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

```
Out[9]:
```

	SupplierID	CompanyName	City	ProductID	ProductName	_merge
0	1	Exotic Liquids	London	1.0	Chai	both
1	1	Exotic Liquids	London	2.0	Chang	both
2	2	New Orleans Cajun Delights	New Orleans	NaN	NaN	left_only
3	3	Grandma Kelly's Homestead	Ann Arbor	6.0	Grandma's Boysenberry Spread	both
4	3	Grandma Kelly's Homestead	Ann Arbor	7.0	Uncle Bob's Organic Dried Pears	both

```
In [10]: print(len(suppliers))
print(len(products))
print(len(sup_prod_left))
```

```
29
68
70
```

```
In [11]: filter = sup_prod_left["_merge"] == "left_only"
sup_prod_left[filter]
```

```
Out[11]:
```

	SupplierID	CompanyName	City	ProductID	ProductName	_merge
2	2	New Orleans Cajun Delights	New Orleans	NaN	NaN	left_only
67	27	Escargots Nouveaux	Montceau	NaN	NaN	left_only

```
In [12]: prod_sup_right = products.merge(suppliers, how='right', on='SupplierID', indicator=True)
```

```
In [13]: print(len(suppliers))
         print(len(products))
         print(len(prod_sup_right))
```

```
29
68
70
```

```
In [14]: filter = prod_sup_right["_merge"] == "right_only"
         prod_sup_right[filter]
```

Out[14]:

	ProductID	ProductName	SupplierID	CompanyName	City	_merge
68	NaN	NaN	2	New Orleans Cajun Delights	New Orleans	right_only
69	NaN	NaN	27	Escargots Nouveaux	Montceau	right_only

```
In [ ]:
```