

Python - Analiza danych z modulem PANDAS

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

LAB - S07-L005 - merge outer

1. Zaimportuj moduł pandas i numpy nadaj im standardowe aliasy
2. Do zmiennej **suppliers** zaimportuj zawartość pliku **suppliers_del_1.csv**. Pobierz tylko kolumny **SupplierID**, **CompanyName**, **City**
3. Do zmiennej **products** zaimportuj zawartość pliku **products_del_1.csv**. **Pobierz tylko kolumny** ProductID, ProductName, SupplierID**
4. W zmiennej **prod_and_sup** zapisz wynik złączenia (**inner**) z **products** do **suppliers**
5. Wyświetl informację o ilości wierszy w trzech wyżej wymienionych zmiennych: **suppliers**, **products** i **prod_and_sup**
6. Jeżeli wszystkie dane są spójne to liczba wierszy w products powinna być równa liczbie wierszy w prod_and_sup. Tutaj tak nie jest, więc podejrzewamy, że ktoś usunął jeden lub więcej wierszy w **suppliers**. Znajdź je!
 - w zmiennej **prod_and_sup_outer** zapisz wynik złączenia poleceniem **merge** dataframe **products** i **suppliers** (**outer**)
 - złączając dane dodaj parametr **indicator**, co pozwoli określić z której tabeli pochodzi dany wiersz
 - utwórz zmienną - serię wartości logicznych, która pozwoli zidentyfikować tylko wiersze, które znajdowały się po lewej stronie (**left_only**)
 - wyświetl informacje o tych produktach, dla których nie jest znany dostawca
7. Podejrzewamy, że w **suppliers** mogą być dostawcy, którzy nie dostarczają żadnych produktów (kandydaci do usunięcia z tego obiektu). Znajdź ich
 - korzystaj ze zmiennej **prod_and_sup_outer**
 - utwórz zmienną - serię wartości logicznych, która pozwoli zidentyfikować te wiersze, które znajdują się po prawej stronie (**right_only**)
 - wyświetl informacje o tych dostawcach, dla których nie udało się znaleźć żadnych produktów

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_del_1.csv", usecols=["Sup
```

```
In [3]: suppliers.head()
```

```
Out[3]:
```

	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 [4]: products = pd.read_csv("../northwind-mongo-master/products_del_1.csv", usecols=['ProductID', 'ProductName', 'SupplierID'])
```

```
In [5]: products.head()
```

```
Out[5]:
```

	ProductID	ProductName	SupplierID
0	1	Chai	1
1	2	Chang	1
2	3	Aniseed Syrup	1
3	4	Chef Anton's Cajun Seasoning	2
4	5	Chef Anton's Gumbo Mix	2

```
In [6]: prod_and_sup = products.merge(suppliers, on="SupplierID")
```

```
In [7]: prod_and_sup.head()
```

```
Out[7]:
```

	ProductID	ProductName	SupplierID	CompanyName	City
0	1	Chai	1	Exotic Liquids	London
1	2	Chang	1	Exotic Liquids	London
2	3	Aniseed Syrup	1	Exotic Liquids	London
3	4	Chef Anton's Cajun Seasoning	2	New Orleans Cajun Delights	New Orleans
4	5	Chef Anton's Gumbo Mix	2	New Orleans Cajun Delights	New Orleans

```
In [8]: print(len(suppliers))
print(len(products))
print(len(prod_and_sup))
```

```
27
75
70
```

```
In [9]: prod_and_sup_outer = products.merge(suppliers, on="SupplierID", how="outer", indicator=True)
```

```
In [10]: prod_and_sup_outer.head()
```

```
Out[10]:
```

	ProductID	ProductName	SupplierID	CompanyName	City	_merge
0	1.0	Chai	1	Exotic Liquids	London	both
1	2.0	Chang	1	Exotic Liquids	London	both
2	3.0	Aniseed Syrup	1	Exotic Liquids	London	both
3	4.0	Chef Anton's Cajun Seasoning	2	New Orleans Cajun Delights	New Orleans	both
4	5.0	Chef Anton's Gumbo Mix	2	New Orleans Cajun Delights	New Orleans	both

```
In [11]: filter_right_only = prod_and_sup_outer["_merge"] == "right_only"
prod_and_sup_outer[filter_right_only]
```

```
Out[11]:
```

	ProductID	ProductName	SupplierID	CompanyName	City	_merge
75	NaN	NaN	27	Escargots Nouveaux	Montceau	right_only

```
In [12]: filter_left_only = prod_and_sup_outer["_merge"] == "left_only"
prod_and_sup_outer[filter_left_only]
```

```
Out[12]:
```

	ProductID	ProductName	SupplierID	CompanyName	City	_merge
27	22.0	Gustaf's Knäckebröd	9	NaN	NaN	left_only
28	23.0	Tunnbröd	9	NaN	NaN	left_only
48	36.0	Inlagd Sill	17	NaN	NaN	left_only
49	37.0	Gravad lax	17	NaN	NaN	left_only
50	73.0	Röd Kaviar	17	NaN	NaN	left_only

```
In [ ]:
```