

# Python - Analiza danych z modułem PANDAS

www.udemy.com (R)

## LAB - S04-L002-Modyfikacja danych

1. Zaimportuj moduł pandas i nadaj mu standardowy alias. Do zmiennej **professions** wczytaj zawartość pliku **Prestige.csv**. Podczas wczytywania zdefiniuj indeks oparty o kolumnę name. Wyświetl nagłówek obiektu data frame.
2. W wierszu identyfikowanym indeksem **chemists** zmień wpis w kolumnie **type** na **scientist**. Wyświetl nagłówek.
3. Zbuduj serię o nazwie **isScientist**, zawierającą wartości True/False. True, jeśli odpowiadający wiersz w **professions** ma w kolumnie **type** wartość **scientist**, a False w przeciwnym wypadku. Wyświetl nagłówek
4. Nauka zyskuje na prestiżu! Zmień dla wszystkich zawodów posiadających w kolumnie **type** zapis **scientist** wartość 90 do kolumny **prestige**. Wyświetl nagłówek
5. Za prestiżem idą pieniądze. Zapisz dla wszystkich zawodów, które w kolumnie **type** mają wartość **scientists** do kolumny **income** wartość o 50% większą. Wyświetl nagłówek
6. Za prestiżem idzie więcej pieniędzy. Wykonaj jeszcze raz czynność z poprzedniego punktu ale podczas przypisywania wartości skorzystaj z notacji **\*=**

## 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
professions = pd.read_csv("Prestige.csv", index_col='name')
professions.head(5)
```

```
Out[1]:
```

	education	income	women	prestige	census	type
name						
gov.administrators	13.11	12351	11.16	68.8	1113	prof
general.managers	12.26	25879	4.02	69.1	1130	prof
accountants	12.77	9271	15.70	63.4	1171	prof
purchasing.officers	11.42	8865	9.11	56.8	1175	prof
chemists	14.62	8403	11.68	73.5	2111	prof

```
In [2]: professions.loc["chemists", "type"] = 'scientist'
professions.head()
```

Out[2]:

	education	income	women	prestige	census	type
name						
gov.administrators	13.11	12351	11.16	68.8	1113	prof
general.managers	12.26	25879	4.02	69.1	1130	prof
accountants	12.77	9271	15.70	63.4	1171	prof
purchasing.officers	11.42	8865	9.11	56.8	1175	prof
chemists	14.62	8403	11.68	73.5	2111	scientist

In [3]:

```
isScientist = (professions["type"] == 'scientist')
isScientist.head()
```

Out[3]:

name	
gov.administrators	False
general.managers	False
accountants	False
purchasing.officers	False
chemists	True
Name: type, dtype: bool	

In [4]:

```
professions.loc[isScientist,"prestige"] = 90
professions.head()
```

Out[4]:

	education	income	women	prestige	census	type
name						
gov.administrators	13.11	12351	11.16	68.8	1113	prof
general.managers	12.26	25879	4.02	69.1	1130	prof
accountants	12.77	9271	15.70	63.4	1171	prof
purchasing.officers	11.42	8865	9.11	56.8	1175	prof
chemists	14.62	8403	11.68	90.0	2111	scientist

In [5]:

```
professions.loc[isScientist,"income"] = professions["income"] *1.5
professions.head()
```

Out[5]:

	education	income	women	prestige	census	type
name						
gov.administrators	13.11	12351.0	11.16	68.8	1113	prof
general.managers	12.26	25879.0	4.02	69.1	1130	prof
accountants	12.77	9271.0	15.70	63.4	1171	prof
purchasing.officers	11.42	8865.0	9.11	56.8	1175	prof
chemists	14.62	12604.5	11.68	90.0	2111	scientist

In [6]:

```
professions.loc[isScientist,"income"] *= 1.5
professions.head()
```

Out[6]:

	education	income	women	prestige	census	type
name						
gov.administrators	13.11	12351.00	11.16	68.8	1113	prof
general.managers	12.26	25879.00	4.02	69.1	1130	prof
accountants	12.77	9271.00	15.70	63.4	1171	prof
purchasing.officers	11.42	8865.00	9.11	56.8	1175	prof
chemists	14.62	18906.75	11.68	90.0	2111	scientist