

Python - Analiza danych z modulem PANDAS

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

LAB - S10-L003 - excel

1. Zaimportuj moduł pandas i numpy nadaj im standardowe aliasy
2. Do zmiennej **mcdonalds** zaimportuj plik **mcdonalds.csv**, pobierz tylko kolumny **Category, Item, Calories i TotalFat**
3. Korzystając z context managera otwórz do zapisu plik **mcdonalds-excel.xlsx**. Plik jest otwierany na zapis mode = "w". Zmienna excelWriter będzie widoczna tylko w tej komórce w kodzie znajdującym się poniżej bloku with, dlatego wstrzymaj się z uruchomieniem tej linijki - najpierw, w tej komórce poniżej wpisz komendę opisaną w następnym kroku

with pd.ExcelWriter("mcdonalds-excel.xlsx", mode="w") as excelWriter:

4. Zapisz zawartość zmiennej mcdonalds do **excelWriter**. Nowy arkusz ma się nazywać **Menu**. Indeks nie powinien być zapisywany. Uruchom kod tej komórki.
5. Uruchomienie komórki otworzyło, zapisało dane i zamknęło plik. Możesz sprawdzić, jak wygląda teraz plik, ale po takim sprawdzeniu pamiętaj aby zamknąć Excela, bo inaczej dostęp do pliku może być zablokowany i kolejne komendy skończą się błędem.
6. Do zmiennej **imported1** zaimportuj ze skoroszytu **mcdonalds-excel.xlsx** arkusz **Menu**.
7. Zaktualizuj w **imported** zawartość kolumny **Calories** dzieląc ją przez **2**. W ten sposób dostaniesz nieco lżejsze menu :)
8. Korzystając z podobnej instrukcji, jak w kroku 3, zapisz zawartość **imported1** w **excelWriter**, dodając arkusz **Menu Light**. Ponieważ teraz dane są dopisywane do istniejącego pliku, to tryb "w" (write) powinien zostać zmieniony na "a" (append).
9. Zaktualizuj w **imported1** zawartość kolumny **Calories** dzieląc ją ponownie przez **2**. W ten sposób dostajemy znowu mniej kaloryczne dania :)
10. Znowu korzystając z instrukcji, jak w kroku 8, zapisz zawartość **imported1** w **excelWriter**, dodając arkusz **Menu Extra Light**
11. Do zmiennej **final** wczytaj zawartość wszystkich arkuszy ze skoroszytu **mcdonalds-excel.xls**
12. Sprawdź ile masz obiektów data frame w zmiennej **final**, wyświetlając jej długość
13. Wyświetl nagłówki zerowego, pierwszego i drugiego arkusza. Czy widzisz zmniejszającą się ilość kalorii w kolejnych arkuszach?

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]: mcdonalds = pd.read_csv("Mcdonalds.csv",
                                usecols=['Category', 'Item', 'Calories', 'TotalFat'])
mcdonalds.head()
```

Out[2]:

	Category	Item	Calories	TotalFat
0	Breakfast	Egg McMuffin	300	13.0
1	Breakfast	Egg White Delight	250	8.0
2	Breakfast	Sausage McMuffin	370	23.0
3	Breakfast	Sausage McMuffin with Egg	450	28.0
4	Breakfast	Sausage McMuffin with Egg Whites	400	23.0

```
In [3]: # użyj rozszerzenia xlsx, bo xls jest już wycofane (office 2003)
# zamiast osobnych operacji (1) otwórz plik (2) zapisz (3) zapisz
# używamy tzw. context manager (to ta składnia z with excelWriter...)
# context manager automatycznie otworzy plik,
#                                     a po wykonaniu zapisu go zamknie

with pd.ExcelWriter("mcdonalds-excel.xlsx", mode="w") as excelWriter:
    mcdonalds.to_excel(excelWriter, sheet_name="Menu", index=False)
```

```
In [4]: imported1 = pd.read_excel("mcdonalds-excel.xlsx", sheet_name="Menu")
```

```
In [5]: imported1["Calories"] = imported1["Calories"] / 2
```

```
In [6]: imported1.head()
```

Out[6]:

	Category	Item	Calories	TotalFat
0	Breakfast	Egg McMuffin	150.0	13.0
1	Breakfast	Egg White Delight	125.0	8.0
2	Breakfast	Sausage McMuffin	185.0	23.0
3	Breakfast	Sausage McMuffin with Egg	225.0	28.0
4	Breakfast	Sausage McMuffin with Egg Whites	200.0	23.0

```
In [7]: with pd.ExcelWriter("mcdonalds-excel.xlsx", mode="a") as excelWriter:
imported1.to_excel(excelWriter, sheet_name="Menu Light")
```

```
In [8]: imported1["Calories"] = imported1["Calories"]/2
```

```
In [9]: imported1.head()
```

Out[9]:

	Category	Item	Calories	TotalFat
0	Breakfast	Egg McMuffin	75.0	13.0
1	Breakfast	Egg White Delight	62.5	8.0
2	Breakfast	Sausage McMuffin	92.5	23.0
3	Breakfast	Sausage McMuffin with Egg	112.5	28.0
4	Breakfast	Sausage McMuffin with Egg Whites	100.0	23.0

```
In [10]: with pd.ExcelWriter("mcdonalds-excel.xlsx", mode="a") as excelWriter:
imported1.to_excel(excelWriter, sheet_name="Menu Extra Light")
```

```
In [11]: final = pd.read_excel("mcdonalds-excel.xlsx", sheet_name=[0,1,2])
```

```
In [12]: len(final)
```

Out[12]: 3

```
In [13]: final[0].head()
```

Out[13]:

	Category	Item	Calories	TotalFat
0	Breakfast	Egg McMuffin	300	13.0
1	Breakfast	Egg White Delight	250	8.0
2	Breakfast	Sausage McMuffin	370	23.0
3	Breakfast	Sausage McMuffin with Egg	450	28.0
4	Breakfast	Sausage McMuffin with Egg Whites	400	23.0

```
In [14]: final[1].head()
```

Out[14]:

	Unnamed: 0	Category	Item	Calories	TotalFat
0	0	Breakfast	Egg McMuffin	150.0	13.0
1	1	Breakfast	Egg White Delight	125.0	8.0
2	2	Breakfast	Sausage McMuffin	185.0	23.0
3	3	Breakfast	Sausage McMuffin with Egg	225.0	28.0
4	4	Breakfast	Sausage McMuffin with Egg Whites	200.0	23.0

```
In [15]: final[2].head()
```

Out[15]:

	Unnamed: 0	Category	Item	Calories	TotalFat
0	0	Breakfast	Egg McMuffin	75.0	13.0
1	1	Breakfast	Egg White Delight	62.5	8.0
2	2	Breakfast	Sausage McMuffin	92.5	23.0
3	3	Breakfast	Sausage McMuffin with Egg	112.5	28.0
4	4	Breakfast	Sausage McMuffin with Egg Whites	100.0	23.0

```
In [ ]:
```

```
In [ ]:
```