

Python - Analiza danych z modulem PANDAS

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

LAB - S05-L005 - stack unstack

1. Zaimportuj moduł pandas i numpy, nadaj im standardowe aliasy. Do zmiennej **fuel** wczytaj zawartość pliku **fuel.csv**. Pobierz następujące kolumny: **'Year','Make','Model','Range (FT1)','City Range (FT1)'**. Zdefiniuj multiindex w oparciu o kolumny **'Year','Make','Model'**, posortuj dane wg indeksu i wyświetl nagłówki obiektu data frame.
2. Polecenia, które za chwilę będziesz uruchamiać wymagają, aby w analizowanych danych nie było duplikatów, dlatego wykonaj następujące kroki (stosuj **inplace=True**):
 - usuń duplikaty ze względu na kolumny **'Year','Make','Model'**
 - usuń wiersze zawierające wartości **NaN**
 - utwórz multiindex w oparciu o **'Year','Make','Model'**
 - wyświetl obiekt **fuel**
3. Kolejne zadania wykonuj bez przypisywania wyników do zmiennej i bez **inplace=True**. W każdym z poniższych punktów staraj się za pomocą pojedynczego polecenia odpowiednio przebudować data frame **fuel**
 - skonwertuj nagłówki kolumn do kolejnego poziomu multiindeksu. Wynik powinien być serią danych
 - skonwertuj nagłówki kolumn do kolejnego poziomu multiindeksu. Wynik powinien być data frame
 - przesun poziom indeksu o nazwie **Model** do kolumn
 - przesun poziom indeksu o nazwie **Year** do kolumn
 - skonwertuj nagłówki kolumn do kolejnego poziomu multiindeksu, a poziom o nazwie **Year** przesun do kolumn (tabela w nagłówku powinna mieć tylko **Year**)
 - skonwertuj nagłówki kolumn do kolejnego poziomu multiindeksu, a poziom o nazwie **Make** przesun do kolumn (tabela w nagłówku powinna mieć tylko **Make**)
 - skonwertuj nagłówki kolumn do kolejnego poziomu multiindeksu, a poziomy o nazwie **'Year','Make'** przesun do kolumn (tabela w nagłówku powinna mieć tylko **Year i Make**)
 - skonwertuj nagłówki kolumn do kolejnego poziomu multiindeksu, a poziomy o nazwie **'Make','Year'** przesun do kolumn (tabela w nagłówku powinna mieć tylko **Make i Year**). Zadbaj o sortowanie kolumn w kolejności najpierw po marce samochodu (**Make**) a następnie po roku (**Year**)

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
fuel = pd.read_csv("./fuel.csv",
                  usecols=['Year', 'Make', 'Model', 'Range (FT1)', 'City Range (FT1)'])
fuel.head(10)
```

```
Out[1]:
```

	Year	Make	Model	Range (FT1)	City Range (FT1)
0	1984	Alfa Romeo	GT V6 2.5	0.0	0.0
1	1984	Alfa Romeo	GT V6 2.5	0.0	0.0
2	1984	Alfa Romeo	Spider Veloce 2000	0.0	0.0
3	1984	Alfa Romeo	Spider Veloce 2000	0.0	0.0
4	1984	AM General	DJ Po Vehicle 2WD	0.0	0.0
5	1984	AM General	DJ Po Vehicle 2WD	0.0	0.0
6	1984	AM General	FJ8c Post Office	0.0	0.0
7	1984	AM General	FJ8c Post Office	0.0	0.0
8	1984	American Motors Corporation	Eagle 4WD	0.0	0.0
9	1984	American Motors Corporation	Eagle 4WD	0.0	0.0

```
In [2]: fuel.drop_duplicates(subset=['Year', 'Make', 'Model'], inplace=True)
fuel.dropna(how='any', inplace=True)
fuel.set_index(['Year', 'Make', 'Model'], inplace=True)
```

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

```
Out[3]:
```

	Year	Make	Model	Range (FT1)	City Range (FT1)
1984	Alfa Romeo	GT V6 2.5	0.0	0.0	
		Spider Veloce 2000	0.0	0.0	
	AM General	DJ Po Vehicle 2WD	0.0	0.0	
		FJ8c Post Office	0.0	0.0	
	American Motors Corporation	Eagle 4WD	0.0	0.0	

```
In [4]: fuel.stack().head()
```

```
Out[4]: Year  Make      Model
1984  Alfa Romeo  GT V6 2.5      Range (FT1)      0.0
                                           City Range (FT1)  0.0
                                           Spider Veloce 2000  Range (FT1)      0.0
                                           City Range (FT1)  0.0
           AM General  DJ Po Vehicle 2WD  Range (FT1)      0.0
dtype: float64
```

```
In [5]: fuel.stack().to_frame().head()
```

```
Out[5]:
```

0				
Year	Make	Model		
1984	Alfa Romeo	GT V6 2.5	Range (FT1)	0.0
			City Range (FT1)	0.0
		Spider Veloce 2000	Range (FT1)	0.0
			City Range (FT1)	0.0
	AM General	DJ Po Vehicle 2WD	Range (FT1)	0.0

```
In [6]: fuel.unstack().head(5)
```

```
Out[6]:
```

Range (FT1)												...	City Range (FT1)		
Model		09-Mar	09-May	1-Ton Truck 2WD	100	100 Wagon	100 quattro	100 quattro Wagon	1000	128ci Convertible	128i	...	i-350 Crew Cab 4WD	i-370 Crew Cab 2WD	i-400 Crew Cab 4WD
Year	Make														
1984	AM General	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
	Alfa Romeo	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
	American Motors Corporation	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
	Aston Martin	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
	Audi	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN

5 rows × 5914 columns

```
In [7]: fuel.unstack(level=0).head(5)
```

```
Out[7]:
```

		Year	Range (FT1)										...	City Range (FT1)			
			1984	1985	1986	1987	1988	1989	1990	1991	1992	1993		2003	2004	2005	2006
Make	Model																
AM General	DJ Po Vehicle 2WD		0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN
	FJ8c Post Office		0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN
	Post Office DJ5 2WD		NaN	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN
	Post Office DJ8 2WD		NaN	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN
ASC Incorporated	GNX		NaN	NaN	NaN	0.0	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN

5 rows × 58 columns

```
In [8]: fuel.stack().unstack(level=0).head()
```

```
Out[8]:
```

		Year	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	...	2003	2004	2005	20
Make	Model																
AM General	DJ Po Vehicle 2WD	Range (FT1)	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN
		City Range (FT1)	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN
	FJ8c Post Office	Range (FT1)	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN
		City Range (FT1)	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN
	Post Office DJ5 2WD	Range (FT1)	NaN	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN

5 rows × 29 columns

```
In [9]: fuel.stack().unstack(level='Make').head()
```

```
Out[9]:
```

		Make	AM General	ASC Incorporated	Acura	Alfa Romeo	American Motors Corporation	Aston Martin	Audi	Aurora Cars Ltd	Autokraft Limited	Avant Motor Corpc
Year	Model											
1984	1000	Range (FT1)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
		City Range (FT1)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	18i	Range (FT1)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
		City Range (FT1)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	190 D 2.2/190 E 2.3	Range (FT1)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

5 rows × 127 columns

```
In [10]: fuel.stack().unstack(level=['Year', 'Make']).head()
```

```
Out[10]:
```

	Year	1984										...	20
	Make	Alfa Romeo	AM General	American Motors Corporation	Aston Martin	Audi	Avanti Motor Corporation	Bertone	Bill Dovell Motor Car Company	Bitter GmbH and Co. Kg	BMW	...	L
Model													
09- Mar	Range (FT1)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	
	City Range (FT1)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	
09- May	Range (FT1)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	
	City Range (FT1)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	
1-Ton Truck 2WD	Range (FT1)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	

5 rows × 1386 columns

```
In [11]: fuel.stack().unstack(level=['Year', 'Make']).sort_index(axis=1).head()
```

```
Out[11]:
```

	Year	1984											...	20
	Make	AM General	Alfa Romeo	American Motors Corporation	Aston Martin	Audi	Avanti Motor Corporation	BMW	Bertone	Bill Dovell Motor Car Company	Bitter Gmbh and Co. Kg	...	L	
Model														
09-Mar	Range (FT1)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	
	City Range (FT1)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	
09-May	Range (FT1)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	
	City Range (FT1)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	
1-Ton Truck 2WD	Range (FT1)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	

5 rows × 1386 columns

```
In [ ]:
```