

Python - Analiza danych z modulem PANDAS

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

LAB - S03-L005-LAB Oczyszczanie danych - dropna()

1. Zaimportuj moduł pandas oraz numpy i nadaj im standardowe aliasy. Do zmiennej **fuel** wczytaj zawartość pliku **fuel.csv**. Podczas wczytywania skorzystaj z dodatkowego argumentu **low_memory=False**, pobierz tylko następujące kolumny: 'Vehicle ID','Year','Make','Model','Class','Fuel Type','Combined MPG (FT1)'. Indekssem ma być kolumna **Vehicle ID** Wyświetl nagłówek tak utworzonego Data Frame
2. Wykonaj poniższe polecenia, żeby wprowadzić nowe wartości NaN do pierwszych wierszy zmiennej **fuel**

```
fuel.loc[27705, 'Class'] = np.NaN
```

```
fuel.loc[26561, 'Class'] = np.NaN
```

```
fuel.loc[27550, 'Fuel Type'] = np.NaN
```

3. Wyświetl nagłówek **fuel** żeby upewnić się, że wiersze rzeczywiście zawierają wartości **NaN**
4. Oczyszczyć dane stosując standardowe parametry polecenia **dropna()**
5. Oczyszczyć dane, usuwając kolumny zawierające chociaż jedną wartość **NaN**
6. Oczyszczyć dane, usuwając wiersze, które w kolumnie subset mają chociaż jedną wartość **NaN**
7. Wyświetl nagłówek oryginalnego obiektu **fuel**. Upewnij się, że wiersze i kolumny zawierające wartości **NaN** nadal są obecne w zmiennej **fuel**
8. Oczyszczyć dane tak jak w pkt. 6, ale tym razem zmiany mają być wykonane w oryginalnej zmiennej **fuel**
9. Wyświetl nagłówek oryginalnego obiektu **fuel**. Upewnij się, że wiersze zawierające w kolumnie **Class** wartość **NaN** rzeczywiście zostały usunięte

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=['Vehicle ID', 'Year', 'Make',
                                         'Model', 'Class', 'Fuel Type',
                                         'Combined MPG (FT1)'],
                  index_col = 'Vehicle ID')
fuel.head()
```

Out[1]:

	Year	Make	Model	Class	Fuel Type	Combined MPG (FT1)
Vehicle ID						
26587	1984	Alfa Romeo	GT V6 2.5	Minicompact Cars	Regular	20.0
27705	1984	Alfa Romeo	GT V6 2.5	Minicompact Cars	Regular	20.0
26561	1984	Alfa Romeo	Spider Veloce 2000	Two Seaters	Regular	21.0
27681	1984	Alfa Romeo	Spider Veloce 2000	Two Seaters	Regular	21.0
27550	1984	AM General	DJ Po Vehicle 2WD	Special Purpose Vehicle 2WD	Regular	17.0

```
In [2]: fuel.loc[27705, 'Class'] = np.NaN
fuel.loc[26561, 'Class'] = np.NaN
fuel.loc[27550, 'Fuel Type'] = np.NaN
```

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

```
Out[3]:
```

	Year	Make	Model	Class	Fuel Type	Combined MPG (FT1)
Vehicle ID						
26587	1984	Alfa Romeo	GT V6 2.5	Minicompact Cars	Regular	20.0
27705	1984	Alfa Romeo	GT V6 2.5	NaN	Regular	20.0
26561	1984	Alfa Romeo	Spider Veloce 2000	NaN	Regular	21.0
27681	1984	Alfa Romeo	Spider Veloce 2000	Two Seaters	Regular	21.0
27550	1984	AM General	DJ Po Vehicle 2WD	Special Purpose Vehicle 2WD	NaN	17.0

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

```
Out[4]:
```

	Year	Make	Model	Class	Fuel Type	Combined MPG (FT1)
Vehicle ID						
26587	1984	Alfa Romeo	GT V6 2.5	Minicompact Cars	Regular	20.0
27681	1984	Alfa Romeo	Spider Veloce 2000	Two Seaters	Regular	21.0
28426	1984	AM General	DJ Po Vehicle 2WD	Special Purpose Vehicle 2WD	Regular	17.0
27549	1984	AM General	FJ8c Post Office	Special Purpose Vehicle 2WD	Regular	13.0
28425	1984	AM General	FJ8c Post Office	Special Purpose Vehicle 2WD	Regular	13.0

```
In [5]: fuel.dropna(axis='columns').dropna().head()
```

```
Out[5]:
```

	Year	Make
Vehicle ID		
26587	1984	Alfa Romeo
27705	1984	Alfa Romeo
26561	1984	Alfa Romeo
27681	1984	Alfa Romeo
27550	1984	AM General

```
In [6]: fuel.dropna(subset=['Class']).head()
```

```
Out[6]:
```

	Year	Make	Model	Class	Fuel Type	Combined MPG (FT1)
Vehicle ID						
26587	1984	Alfa Romeo	GT V6 2.5	Minicompact Cars	Regular	20.0
27681	1984	Alfa Romeo	Spider Veloce 2000	Two Seaters	Regular	21.0
27550	1984	AM General	DJ Po Vehicle 2WD	Special Purpose Vehicle 2WD	NaN	17.0
28426	1984	AM General	DJ Po Vehicle 2WD	Special Purpose Vehicle 2WD	Regular	17.0
27549	1984	AM General	FJ8c Post Office	Special Purpose Vehicle 2WD	Regular	13.0

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

```
Out[7]:
```

	Year	Make	Model	Class	Fuel Type	Combined MPG (FT1)
Vehicle ID						
26587	1984	Alfa Romeo	GT V6 2.5	Minicompact Cars	Regular	20.0
27705	1984	Alfa Romeo	GT V6 2.5	NaN	Regular	20.0
26561	1984	Alfa Romeo	Spider Veloce 2000	NaN	Regular	21.0
27681	1984	Alfa Romeo	Spider Veloce 2000	Two Seaters	Regular	21.0
27550	1984	AM General	DJ Po Vehicle 2WD	Special Purpose Vehicle 2WD	NaN	17.0

```
In [8]: fuel.dropna(subset=['Class'], inplace=True)
```

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

```
Out[9]:
```

	Year	Make	Model	Class	Fuel Type	Combined MPG (FT1)
Vehicle ID						
26587	1984	Alfa Romeo	GT V6 2.5	Minicompact Cars	Regular	20.0
27681	1984	Alfa Romeo	Spider Veloce 2000	Two Seaters	Regular	21.0
27550	1984	AM General	DJ Po Vehicle 2WD	Special Purpose Vehicle 2WD	NaN	17.0
28426	1984	AM General	DJ Po Vehicle 2WD	Special Purpose Vehicle 2WD	Regular	17.0
27549	1984	AM General	FJ8c Post Office	Special Purpose Vehicle 2WD	Regular	13.0