Python - Analiza danych z modułem PANDAS

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

LAB - S10-L001 - Import

1. Zaimportuj moduł pandas i numpy nadaj im standardowe aliasy
2. Strona https://www.europeandataportal.eu/data/en/dataset/http-ec-europa-eu-eurostat-web-products-datasets-isoc_cimobi_frq (https://www.europeandataportal.eu/data/en/dataset/http-ec-europa-eu-eurostat-web-products-datasets-isoc_cimobi_frq) zawiera informacje na temat wykorzystania mobilnego internetu w EU. Dane w postaci CSV znajdują się pod adresem http://ec.europa.eu/eurostat/estat-navtree-portlet-prod/BulkDownloadListing?file=data/isoc_cimobi_frq.tsv.gz&unzip=true (http://ec.europa.eu/eurostat/estat-navtree-portlet-prod/BulkDownloadListing?file=data/isoc_cimobi_frq.tsv.gz&unzip=true) W kolejnych punktach będziesz pobierać ten plik. Gdyby podany wyżej link nie był już aktualny, to skorzystaj z dołączonego pliku **mobile_internet_use_isoc_cimobi_frq.tsv**. Znaczenie symboli określających różne rodzaje połączeń znajdziesz pod adresem http://dd.eionet.europa.eu/vocabulary/eurostat/indic_is/view?E_SECINONE=&page=6 (http://dd.eionet.europa.eu/vocabulary/eurostat/indic_is/view?E_SECINONE=&page=6)
3. Zadeklaruj zmienną **url** wskazującą na adres z danymi
4. Zaimportuj dane do zmiennej **df**. Jako separator pól wykorzystaj tabulator
5. Do zmiennej **df_desc** zapisz data frame utworzony jako wynik rozbicia danych z kolumny **indic_is,ind_type,unit,time\geo** ze względu na przecinek. Skorzystaj z opcji konwertującej wynik do data frame (expand)
6. W **df_desc** zmień nazwy kolumn z 0,1,2,3 na **'indic_is','ind_type',unit','time'**
7. Korzystając z polecenia **join** dołacz do **df_desc** dane z **df**
8. Z **df_desc** usuń niepotrzebną już kolumnę **'indic_is,ind_type,unit,time\geo'**
9. Pod adresem https://examples.opendatasoft.com/explore/dataset/world-heritage-unesco-list/export/ (https://examples.opendatasoft.com/explore/dataset/world-heritage-unesco-list/export/) znajdują się dane dotyczące obiektów na liście UNESCO. Zaimportuj dane dostępne pod adresem https://examples.opendatasoft.com/explore/dataset/world-heritage-unesco-list/download/?format=json&timezone=Europe/Berlin (https://examples.opendatasoft.com/explore/dataset/world-heritage-unesco-list/download/?format=json&timezone=Europe/Berlin) Jeżeli plik byłby niedostępny, to skorzystaj z lokalnego pliku **world-heritage-unesco-list.json**

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]:
```
url = 'http://ec.europa.eu/eurostat/estat-navtree-portlet-prod/BulkDownloadListing?fil
```

In [3]:
```
df = pd.read_csv(url, delimiter = '\t').head()
df.head()
```

Out[3]:

	indic_is,ind_type,unit,time\geo	EU28	EU27	EU15	EA	BE	BG	CZ	DK	DE	...	SI	SK	FI	SE
0	I_IHDDAY,CB_EU_FOR,PC_IND,2012	15	15	16	12	15	: u	: u	22	7	...	: u	14 u	35 u	35 u
1	I_IHDDAY,CB_EU_FOR,PC_IND_IU3,2012	21	21	21	17	19	: u	: u	22	11	...	: u	17 u	36 u	40 u
2	I_IHDDAY,CB_EU_FOR,PC_IND_IUHD,2012	57	57	57	55	48	: u	: u	49 u	60	...	: u	: u	: u	: u
3	I_IHDDAY,CB_EXT_EU,PC_IND,2012	21	21	22	13	11	25 u	:	35	12	...	3	: u	22	39
4	I_IHDDAY,CB_EXT_EU,PC_IND_IU3,2012	28	28	29	19	15	28 u	:	39	14	...	5	: u	24 u	42

5 rows × 38 columns

In [16]:
```
pd.read_csv("mobile_internet_use_isoc_cimobi_frq.tsv",delimiter='\t').head()
```

Out[16]:

	indic_is,ind_type,unit,time\geo	EU28	EU27	EU15	EA	BE	BG	CZ	DK	DE	...	SI	SK	FI	SE
0	I_IHDDAY,CB_EU_FOR,PC_IND,2012	15	15	16	12	15	: u	: u	22	7	...	: u	14 u	35 u	35 u
1	I_IHDDAY,CB_EU_FOR,PC_IND_IU3,2012	21	21	21	17	19	: u	: u	22	11	...	: u	17 u	36 u	40 u
2	I_IHDDAY,CB_EU_FOR,PC_IND_IUHD,2012	57	57	57	55	48	: u	: u	49 u	60	...	: u	: u	: u	: u
3	I_IHDDAY,CB_EXT_EU,PC_IND,2012	21	21	22	13	11	25 u	:	35	12	...	3	: u	22	39
4	I_IHDDAY,CB_EXT_EU,PC_IND_IU3,2012	28	28	29	19	15	28 u	:	39	14	...	5	: u	24 u	42

5 rows × 38 columns

In [5]:
```
df_desc = df['indic_is,ind_type,unit,time\geo'].str.split(',',expand=True)
```

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

Out[6]:

	0	1	2	3
0	I_IHDDAY	CB_EU_FOR	PC_IND	2012
1	I_IHDDAY	CB_EU_FOR	PC_IND_IU3	2012
2	I_IHDDAY	CB_EU_FOR	PC_IND_IUHD	2012
3	I_IHDDAY	CB_EXT_EU	PC_IND	2012
4	I_IHDDAY	CB_EXT_EU	PC_IND_IU3	2012

In [7]:
```
df_desc.rename({0:'indic_is',1:',ind_type',2:',unit',3:'time'}, axis=1, inplace=True)
```

In [8]:
```
df_desc.head()
```

Out[8]:

	indic_is	,ind_type	,unit	time
0	I_IHDDAY	CB_EU_FOR	PC_IND	2012
1	I_IHDDAY	CB_EU_FOR	PC_IND_IU3	2012
2	I_IHDDAY	CB_EU_FOR	PC_IND_IUHD	2012
3	I_IHDDAY	CB_EXT_EU	PC_IND	2012
4	I_IHDDAY	CB_EXT_EU	PC_IND_IU3	2012

In [9]:
```
df_desc = df_desc.join(df)
```

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

Out[10]:

	indic_is	,ind_type	,unit	time	indic_is,ind_type,unit,time\geo	EU28	EU27	EU15	EA
0	I_IHDDAY	CB_EU_FOR	PC_IND	2012	I_IHDDAY,CB_EU_FOR,PC_IND,2012	15	15	16	12
1	I_IHDDAY	CB_EU_FOR	PC_IND_IU3	2012	I_IHDDAY,CB_EU_FOR,PC_IND_IU3,2012	21	21	21	17
2	I_IHDDAY	CB_EU_FOR	PC_IND_IUHD	2012	I_IHDDAY,CB_EU_FOR,PC_IND_IUHD,2012	57	57	57	55
3	I_IHDDAY	CB_EXT_EU	PC_IND	2012	I_IHDDAY,CB_EXT_EU,PC_IND,2012	21	21	22	13
4	I_IHDDAY	CB_EXT_EU	PC_IND_IU3	2012	I_IHDDAY,CB_EXT_EU,PC_IND_IU3,2012	28	28	29	19

5 rows × 42 columns

In [11]:
```
df.drop('indic_is,ind_type,unit,time\geo', axis=1, inplace=True)
```

In [12]:
```
df_desc.head()
```

Out[12]:

	indic_is	,ind_type	,unit	time	indic_is,ind_type,unit,time\geo	EU28	EU27	EU15	EA
0	I_IHDDAY	CB_EU_FOR	PC_IND	2012	I_IHDDAY,CB_EU_FOR,PC_IND,2012	15	15	16	12
1	I_IHDDAY	CB_EU_FOR	PC_IND_IU3	2012	I_IHDDAY,CB_EU_FOR,PC_IND_IU3,2012	21	21	21	17
2	I_IHDDAY	CB_EU_FOR	PC_IND_IUHD	2012	I_IHDDAY,CB_EU_FOR,PC_IND_IUHD,2012	57	57	57	55
3	I_IHDDAY	CB_EXT_EU	PC_IND	2012	I_IHDDAY,CB_EXT_EU,PC_IND,2012	21	21	22	13
4	I_IHDDAY	CB_EXT_EU	PC_IND_IU3	2012	I_IHDDAY,CB_EXT_EU,PC_IND_IU3,2012	28	28	29	19

5 rows × 42 columns

In [13]:
```
url = 'https://examples.opendatasoft.com/explore/dataset/world-heritage-unesco-list/dc
```

In [14]:
```
pd.read_json(url).head()
```

Out[14]:

	datasetid	fields	geometry	record_timestamp	
0	world-heritage-unesco-list	{'category': 'Cultural', 'short_description_en...	{'type': 'Point', 'coordinates': [26.69139, 53...	2017-10-17T17:34:52+02:00	ff1f5b718ce2ee87f18dfaf20610f257
1	world-heritage-unesco-list	{'category': 'Natural', 'short_description_en'...	{'type': 'Point', 'coordinates': [14.996666666...	2017-10-17T17:34:52+02:00	0fd1ad38878a73b01bf59ed430446e098
2	world-heritage-unesco-list	{'category': 'Cultural', 'short_description_en...	{'type': 'Point', 'coordinates': [-112.91611, ...	2017-10-17T17:34:52+02:00	fa66a8d51c28871277f1cdd9c0f2d09
3	world-heritage-unesco-list	{'category': 'Cultural', 'short_description_en...	{'type': 'Point', 'coordinates': [10.33333, 24...	2017-10-17T17:34:52+02:00	b16a410e99f603ef138909f6de916de
4	world-heritage-unesco-list	{'category': 'Natural', 'short_description_en'...	{'type': 'Point', 'coordinates': [114.91667, 4...	2017-10-17T17:34:52+02:00	bb09d483a4ca7411a334314fb13f1330

In [15]:
```
pd.read_json('world-heritage-unesco-list.json').head()
```

Out[15]:

	datasetid	fields	geometry	record_timestamp	
0	world-heritage-unesco-list	{'category': 'Cultural', 'short_description_en...	{'type': 'Point', 'coordinates': [26.69139, 53...	2017-10-17T17:34:52+02:00	ff1f5b718ce2ee87f18dfaf20610f257
1	world-heritage-unesco-list	{'category': 'Natural', 'short_description_en'...	{'type': 'Point', 'coordinates': [14.996666666...	2017-10-17T17:34:52+02:00	0fd1ad38878a73b01bf59ed430446e098
2	world-heritage-unesco-list	{'category': 'Cultural', 'short_description_en...	{'type': 'Point', 'coordinates': [-112.91611, ...	2017-10-17T17:34:52+02:00	fa66a8d51c28871277f1cdd9c0f2d09
3	world-heritage-unesco-list	{'category': 'Cultural', 'short_description_en...	{'type': 'Point', 'coordinates': [10.33333, 24...	2017-10-17T17:34:52+02:00	b16a410e99f603ef138909f6de916de
4	world-heritage-unesco-list	{'category': 'Natural', 'short_description_en'...	{'type': 'Point', 'coordinates': [114.91667, 4...	2017-10-17T17:34:52+02:00	bb09d483a4ca7411a334314fb13f1330

In [ ]: