

In [1]:

```
1 import pandas as pd
2 import wget
3 import tabula
4 import re
5 import numpy as np
6 pd.set_option('display.max_colwidth', -1)
7 pd.set_option('display.max_rows', 600)
```

C:\Users\Laptop\Anaconda3\lib\site-packages\ipykernel_launcher.py:6: FutureWarning: Passing a negative integer is deprecated in version 1.0 and will not be supported in future version. Instead, use None to not limit the column width.

In [19]:

```
1 #lista nazw tygodni
2 tygodnie = ['A', 'B', 'C', 'D']
3 miesiace = ['01', '02', '03', '04', '05', '06', '07', '08', '09', '10', '11', '12']
4 lata = ['10', '11', '12', '13', '14', '15', '16', '17', '18', '19', '20']
5 mies_tyg = [['{}{}'.format(row, col) for col in tygodnie] for row in miesiace]
6 mies_tyg2 = sum(mies_tyg, [])
7
8 #print(mies_tyg2)
```

In [14]:

```
1 dodatek = []
2 for x in lata:
3     dana = x + "/G_" + x + "_"
4     dodatek.append(dana)
5 #print(dodatek)
```

In [13]:

```
1 srod_url = [['{}{}'.format(row,col) for col in mies_tyg2] for row in dodatek]
2 srod_url2 = sum(srod_url, [])
3 #print(srod_url2)
```

In [6]:

```
1 dane = pd.DataFrame()
```

In [42]:

```
1 dane['part_URL'] = srod_url2
2 #dane
```

In [11]:

```
1 dane['URL'] = 'http://wwold.pzh.gov.pl/oldpage/epimeld/grypa/20' + dane['part_URL'] +
```

In [47]:

```
1 del dane['part_URL']
```

In [27]:

```

1 dane['Plik'] = ""
2 dane['Rok'] = ""
3 dane['Miesiac'] = ""
4 dane['Tydzien'] = ""

```

In [50]:

```

1 dane['Rok'] = dane['URL'].str.slice(47,51,1)
2 dane['Miesiac'] = dane['URL'].str.slice(57,59,1)
3 dane['Tydzien'] = dane['URL'].str.slice(57,60,1)
4 dane['Plik'] = 'C:/Users/Laptop/Desktop/Grypa/Dane' + '/' + dane['Rok'] + '/' + dane['T
5 #dane

```

In [49]:

```

1 dane

```

Out[49]:

	URL	P
0	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01A.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/01A.p
1	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01B.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/01B.p
2	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01C.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/01C.p
3	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01D.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/01D.p
4	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_02A.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/02A.p
5	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_02B.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/02B.p
6	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_02C.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/02C.p
7	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_02D.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/02D.p
8	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_03A.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/03A.p
9	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_03B.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/03B.p

In [79]:

```
1 dane = dane.drop(dane[(dane.Rok == '2020')].index)
2 dane
```

Out[79]:

	URL	P
0	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01A.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/01A.p
1	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01B.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/01B.p
2	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01C.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/01C.p
3	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01D.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/01D.p
4	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_02A.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/02A.p
5	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_02B.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/02B.p
6	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_02C.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/02C.p
7	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_02D.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/02D.p
8	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_03A.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/03A.p
9	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_03B.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/03B.p

In [77]:

```
1 dane
```

Out[77]:

	URL	P
0	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01A.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/01A.p
1	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01B.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/01B.p
2	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01C.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/01C.p
3	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01D.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/01D.p
4	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_02A.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/02A.p
5	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_02B.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/02B.p
6	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_02C.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/02C.p
7	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_02D.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/02D.p
8	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_03A.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/03A.p
9	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_03B.pdf	C:/Users/Laptop/Desktop/Grypa/Dane/2010/03B.p

In [49]:

```
1 #podstawowy zbiór: Tydzien, URL, Adres
2 dane = pd.DataFrame()
3 dane['Tydzien'] = suma
4 dane['URL_2010'] = 'http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_' + dane[
5 dane['Adres'] = 'C:/Users/Laptop/Desktop/Grypa/Dane/2010/' + dane['Tydzien'] + '.pdf'
6 dane['Total'] = ""
7 #dane
```

In [52]:

```
1 #tworzenie list URL i Adres
2 lista_URL = dane['URL'].to_numpy()
3 lista_adres = dane['Plik'].to_numpy()
4 dic = dict(zip(lista_URL, lista_adres))
5 dic
```

Out[52]:

```
{'http://wwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01A.pdf': 'C:/U
sers/Laptop/Desktop/Grypa/Dane/2010/01A.pdf',
 'http://wwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01B.pdf': 'C:/U
sers/Laptop/Desktop/Grypa/Dane/2010/01B.pdf',
 'http://wwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01C.pdf': 'C:/U
sers/Laptop/Desktop/Grypa/Dane/2010/01C.pdf',
 'http://wwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01D.pdf': 'C:/U
sers/Laptop/Desktop/Grypa/Dane/2010/01D.pdf',
 'http://wwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_02A.pdf': 'C:/U
sers/Laptop/Desktop/Grypa/Dane/2010/02A.pdf',
 'http://wwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_02B.pdf': 'C:/U
sers/Laptop/Desktop/Grypa/Dane/2010/02B.pdf',
 'http://wwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_02C.pdf': 'C:/U
sers/Laptop/Desktop/Grypa/Dane/2010/02C.pdf',
 'http://wwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_02D.pdf': 'C:/U
sers/Laptop/Desktop/Grypa/Dane/2010/02D.pdf',
 'http://wwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_03A.pdf': 'C:/U
sers/Laptop/Desktop/Grypa/Dane/2010/03A.pdf'.
```

In [50]:

```
1 for x in range(48):
2     ad = dane.loc[x, 'Adres']
3     liczba = tabula.read_pdf(ad, area = (230, 84, 240, 180), pages=1)
4     liczba_str = str(liczba)
5     total = re.sub('[^0-9]', '', liczba_str)
6     #TEST['Total'] = total
7     dane.loc[x, "Total"] = total
8     print(total)
9
```

16614
15927
16607
15298
15820
16015
13345
12341
12749
16989
12656
15389
8318
8667
7106
7539
6133
6013
6870
7417
6640
6899
5534
5600
5408
3166
2570
2432
2351
2382
1806
2929
3604
6892
10398
12631
15750
19872
18634
18753
16372
18387
14448
20848
19792
23899
21455
23177

In [60]:

1 dane

Out[60]:

	Tydzien	URL_2010	
0	01A	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01A.pdf	C:/Users/Laptop/Desl
1	01B	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01B.pdf	C:/Users/Laptop/Desl
2	01C	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01C.pdf	C:/Users/Laptop/Desl
3	01D	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01D.pdf	C:/Users/Laptop/Desl
4	02A	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_02A.pdf	C:/Users/Laptop/Desl
5	02B	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_02B.pdf	C:/Users/Laptop/Desl
6	02C	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_02C.pdf	C:/Users/Laptop/Desl
7	02D	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_02D.pdf	C:/Users/Laptop/Desl
8	03A	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_03A.pdf	C:/Users/Laptop/Desl
9	03B	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_03B.pdf	C:/Users/Laptop/Desl
10	03C	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_03C.pdf	C:/Users/Laptop/Desl
11	03D	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_03D.pdf	C:/Users/Laptop/Desl
12	04A	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_04A.pdf	C:/Users/Laptop/Desl
13	04B	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_04B.pdf	C:/Users/Laptop/Desl
14	04C	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_04C.pdf	C:/Users/Laptop/Desl
15	04D	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_04D.pdf	C:/Users/Laptop/Desl
16	05A	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_05A.pdf	C:/Users/Laptop/Desl
17	05B	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_05B.pdf	C:/Users/Laptop/Desl
18	05C	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_05C.pdf	C:/Users/Laptop/Desl
19	05D	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_05D.pdf	C:/Users/Laptop/Desl
20	06A	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_06A.pdf	C:/Users/Laptop/Desl
21	06B	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_06B.pdf	C:/Users/Laptop/Desl
22	06C	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_06C.pdf	C:/Users/Laptop/Desl
23	06D	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_06D.pdf	C:/Users/Laptop/Desl
24	07A	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_07A.pdf	C:/Users/Laptop/Desl
25	07B	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_07B.pdf	C:/Users/Laptop/Desl
26	07C	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_07C.pdf	C:/Users/Laptop/Desl
27	07D	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_07D.pdf	C:/Users/Laptop/Desl
28	08A	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_08A.pdf	C:/Users/Laptop/Desl
29	08B	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_08B.pdf	C:/Users/Laptop/Desl
30	08C	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_08C.pdf	C:/Users/Laptop/Desl
31	08D	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_08D.pdf	C:/Users/Laptop/Desl
32	09A	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_09A.pdf	C:/Users/Laptop/Desl
33	09B	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_09B.pdf	C:/Users/Laptop/Desl

	Tydzien	URL_2010	
34	09C	http://www.wold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_09C.pdf	C:/Users/Laptop/Desl
35	09D	http://www.wold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_09D.pdf	C:/Users/Laptop/Desl
36	10A	http://www.wold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_10A.pdf	C:/Users/Laptop/Desl
37	10B	http://www.wold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_10B.pdf	C:/Users/Laptop/Desl
38	10C	http://www.wold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_10C.pdf	C:/Users/Laptop/Desl
39	10D	http://www.wold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_10D.pdf	C:/Users/Laptop/Desl
40	11A	http://www.wold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_11A.pdf	C:/Users/Laptop/Desl
41	11B	http://www.wold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_11B.pdf	C:/Users/Laptop/Desl
42	11C	http://www.wold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_11C.pdf	C:/Users/Laptop/Desl
43	11D	http://www.wold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_11D.pdf	C:/Users/Laptop/Desl
44	12A	http://www.wold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_12A.pdf	C:/Users/Laptop/Desl
45	12B	http://www.wold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_12B.pdf	C:/Users/Laptop/Desl
46	12C	http://www.wold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_12C.pdf	C:/Users/Laptop/Desl
47	12D	http://www.wold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_12D.pdf	C:/Users/Laptop/Desl

In [59]:

```
1 dane['Total2'] = pd.to_numeric(dane['Total'], errors='ignore')
```

In [61]:

```
1 suma = dane['Total2'].sum()
2 print(suma)
```

550442

In [43]:

```
1 #####
2 #####
3 #####
4 #####
```


In [39]:

```

1 #zbiór testowy
2 TEST = dane.iloc[0:4, 0:4]
3 TEST.reset_index(drop=False, inplace=True)
4 TEST['Total']=""
5 TEST

```

Out[39]:

	index	Tydzien	URL_2010
0	0	01A	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01A.pdf C:/Users/Laptop/I
1	1	01B	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01B.pdf C:/Users/Laptop/I
2	2	01C	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01C.pdf C:/Users/Laptop/I
3	3	01D	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01D.pdf C:/Users/Laptop/I

In [6]:

```

1 #pojedynczy odczyt
2 liczba = tabula.read_pdf('C:/Users/Laptop/Desktop/Grypa/Dane/2010/01B.pdf', area = (230, 84, 240, 180), pages=1)
3 liczba_str = str(liczba)
4 total = re.sub('[^0-9]', '', liczba_str)
5 print(total)

```

15927

In [40]:

```

1 for x in range(4):
2     ad = TEST.loc[x, 'Adres']
3     liczba = tabula.read_pdf(ad, area = (230, 84, 240, 180), pages=1)
4     liczba_str = str(liczba)
5     total = re.sub('[^0-9]', '', liczba_str)
6     #TEST['Total'] = total
7     TEST.loc[x:x+1, 4:5] = total
8     print(total)
9

```

16614

15927

16607

15298

In [41]:

```
1 #reset indeksu
2 #poczytać o dodawaniu wartości do konkretnej komórki
3 TEST
```

Out[41]:

	index	Tydzien	URL_2010
0	0	01A	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01A.pdf C:/Users/Laptop/I
1	1	01B	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01B.pdf C:/Users/Laptop/I
2	2	01C	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01C.pdf C:/Users/Laptop/I
3	3	01D	http://wwwold.pzh.gov.pl/oldpage/epimeld/grypa/2010/G_10_01D.pdf C:/Users/Laptop/I

