untitled3

June 23, 2020

```python
In [1]: import numpy as np
        import pandas as pd
```

```python
In [2]: url = "nhanes_2015_2016.csv"
        da = pd.read_csv(url)
```

```python
In [3]: da.shape()
```

```
        ---------------------------------------------------------------------------

        TypeError                                 Traceback (most recent call last)

        <ipython-input-3-60bc473d046c> in <module>()
    ----> 1 da.shape()


        TypeError: 'tuple' object is not callable
```

```python
In [4]: da.shape
```

```
Out[4]: (5735, 28)
```

```python
In [5]: da.columns
```

```
Out[5]: Index(['SEQN', 'ALQ101', 'ALQ110', 'ALQ130', 'SMQ020', 'RIAGENDR', 'RIDAGEYR',
               'RIDRETH1', 'DMDCITZN', 'DMDEDUC2', 'DMDMARTL', 'DMDHHSIZ', 'WTINT2YR',
               'SDMVPSU', 'SDMVSTRA', 'INDFMPIR', 'BPXSY1', 'BPXDI1', 'BPXSY2',
               'BPXDI2', 'BMXWT', 'BMXHT', 'BMXBMI', 'BMXLEG', 'BMXARML', 'BMXARMC',
               'BMXWAIST', 'HIQ210'],
              dtype='object')
```

```python
In [6]: da.dtypes
```

```
Out[6]: SEQN            int64
        ALQ101        float64
        ALQ110        float64
        ALQ130        float64
```

```
SMQ020            int64
RIAGENDR          int64
RIDAGEYR          int64
RIDRETH1          int64
DMDCITZN        float64
DMDEDUC2        float64
DMDMARTL        float64
DMDHHSIZ          int64
WTINT2YR        float64
SDMVPSU           int64
SDMVSTRA          int64
INDFMPIR        float64
BPXSY1          float64
BPXDI1          float64
BPXSY2          float64
BPXDI2          float64
BMXWT           float64
BMXHT           float64
BMXBMI          float64
BMXLEG          float64
BMXARML         float64
BMXARMC         float64
BMXWAIST        float64
HIQ210          float64
dtype: object
```

In [8]: w = da["DMDEDUC2"]
        x = da.loc[:, "DMDEDUC2"]
        y = da.DMDEDUC2
        z = da.iloc[:, 9]

In [9]: w = da["DMDEDUC2"]

In [10]: x = da.loc[:, "DMDEDUC2"]

In [11]: print(x)

```
0        5.0
1        3.0
2        3.0
3        5.0
4        4.0
5        2.0
6        4.0
7        4.0
8        NaN
9        3.0
10       5.0
11       2.0
```

```
12      4.0
13      5.0
14      5.0
15      1.0
16      NaN
17      5.0
18      4.0
19      4.0
20      2.0
21      3.0
22      1.0
23      5.0
24      2.0
25      5.0
26      4.0
27      5.0
28      4.0
29      1.0
        ...
5705    5.0
5706    2.0
5707    5.0
5708    5.0
5709    2.0
5710    3.0
5711    1.0
5712    4.0
5713    1.0
5714    3.0
5715    5.0
5716    1.0
5717    5.0
5718    5.0
5719    1.0
5720    5.0
5721    5.0
5722    5.0
5723    2.0
5724    5.0
5725    5.0
5726    2.0
5727    1.0
5728    2.0
5729    5.0
5730    3.0
5731    5.0
5732    4.0
5733    1.0
```

```
5734      5.0
Name: DMDEDUC2, Length: 5735, dtype: float64


In [12]: print(w)

0         5.0
1         3.0
2         3.0
3         5.0
4         4.0
5         2.0
6         4.0
7         4.0
8         NaN
9         3.0
10        5.0
11        2.0
12        4.0
13        5.0
14        5.0
15        1.0
16        NaN
17        5.0
18        4.0
19        4.0
20        2.0
21        3.0
22        1.0
23        5.0
24        2.0
25        5.0
26        4.0
27        5.0
28        4.0
29        1.0
         ...
5705      5.0
5706      2.0
5707      5.0
5708      5.0
5709      2.0
5710      3.0
5711      1.0
5712      4.0
5713      1.0
5714      3.0
5715      5.0
```

```
5716      1.0
5717      5.0
5718      5.0
5719      1.0
5720      5.0
5721      5.0
5722      5.0
5723      2.0
5724      5.0
5725      5.0
5726      2.0
5727      1.0
5728      2.0
5729      5.0
5730      3.0
5731      5.0
5732      4.0
5733      1.0
5734      5.0
Name: DMDEDUC2, Length: 5735, dtype: float64
```

In [13]: print(y)

```
0         5.0
1         3.0
2         3.0
3         5.0
4         4.0
5         2.0
6         4.0
7         4.0
8         NaN
9         3.0
10        5.0
11        2.0
12        4.0
13        5.0
14        5.0
15        1.0
16        NaN
17        5.0
18        4.0
19        4.0
20        2.0
21        3.0
22        1.0
23        5.0
```

```
24      2.0
25      5.0
26      4.0
27      5.0
28      4.0
29      1.0
        ...
5705    5.0
5706    2.0
5707    5.0
5708    5.0
5709    2.0
5710    3.0
5711    1.0
5712    4.0
5713    1.0
5714    3.0
5715    5.0
5716    1.0
5717    5.0
5718    5.0
5719    1.0
5720    5.0
5721    5.0
5722    5.0
5723    2.0
5724    5.0
5725    5.0
5726    2.0
5727    1.0
5728    2.0
5729    5.0
5730    3.0
5731    5.0
5732    4.0
5733    1.0
5734    5.0
Name: DMDEDUC2, Length: 5735, dtype: float64
```

In [14]: print(z)

```
0       5.0
1       3.0
2       3.0
3       5.0
4       4.0
5       2.0
```

6	4.0
7	4.0
8	NaN
9	3.0
10	5.0
11	2.0
12	4.0
13	5.0
14	5.0
15	1.0
16	NaN
17	5.0
18	4.0
19	4.0
20	2.0
21	3.0
22	1.0
23	5.0
24	2.0
25	5.0
26	4.0
27	5.0
28	4.0
29	1.0
	...
5705	5.0
5706	2.0
5707	5.0
5708	5.0
5709	2.0
5710	3.0
5711	1.0
5712	4.0
5713	1.0
5714	3.0
5715	5.0
5716	1.0
5717	5.0
5718	5.0
5719	1.0
5720	5.0
5721	5.0
5722	5.0
5723	2.0
5724	5.0
5725	5.0
5726	2.0
5727	1.0

```
5728      2.0
5729      5.0
5730      3.0
5731      5.0
5732      4.0
5733      1.0
5734      5.0
Name: DMDEDUC2, Length: 5735, dtype: float64
```

In [15]: `print(w.max())`
         `print(x.max())`
         `print(y.max())`
         `print(z.max())`

```
9.0
9.0
9.0
9.0
```

In [16]: `print(type(da))`
         `print(type(da.DMDEDUC2))`
         `print(type(da.iloc[2,:]))`

```
<class 'pandas.core.frame.DataFrame'>
<class 'pandas.core.series.Series'>
<class 'pandas.core.series.Series'>
```

In [17]: `x = da.iloc[3, :]`

In [18]: `print(x)`

```
SEQN          83735.0
ALQ101            2.0
ALQ110            1.0
ALQ130            1.0
SMQ020            2.0
RIAGENDR          2.0
RIDAGEYR         56.0
RIDRETH1          3.0
DMDCITZN          1.0
DMDEDUC2          5.0
DMDMARTL          6.0
DMDHHSIZ          1.0
WTINT2YR     102718.0
SDMVPSU           1.0
SDMVSTRA        131.0
```

```
INDFMPIR          5.0
BPXSY1          132.0
BPXDI1           72.0
BPXSY2          134.0
BPXDI2           68.0
BMXWT           109.8
BMXHT           160.9
BMXBMI           42.4
BMXLEG           38.5
BMXARML          37.7
BMXARMC          38.3
BMXWAIST        110.1
HIQ210            2.0
Name: 3, dtype: float64


In [19]: x = da.iloc[3:5, :]
         y = da.iloc[:, 2:5]

In [20]: print(x)
         print(y)

    SEQN  ALQ101  ALQ110  ALQ130  SMQ020  RIAGENDR  RIDAGEYR  RIDRETH1  \
3  83735     2.0     1.0     1.0       2         2        56         3
4  83736     2.0     1.0     1.0       2         2        42         4

   DMDCITZN  DMDEDUC2  ...  BPXSY2  BPXDI2   BMXWT   BMXHT  BMXBMI  BMXLEG  \
3       1.0       5.0  ...   134.0    68.0   109.8   160.9    42.4    38.5
4       1.0       4.0  ...   114.0    54.0    55.2   164.9    20.3    37.4

   BMXARML  BMXARMC  BMXWAIST  HIQ210
3     37.7     38.3     110.1     2.0
4     36.0     27.2      80.4     2.0

[2 rows x 28 columns]
    ALQ110  ALQ130  SMQ020
0      NaN     1.0       1
1      NaN     6.0       1
2      NaN     NaN       1
3      1.0     1.0       2
4      1.0     1.0       2
5      2.0     NaN       2
6      NaN     8.0       1
7      NaN     1.0       2
8      NaN     NaN       2
9      NaN     NaN       2
10     NaN     1.0       1
11     NaN     3.0       1
```

12	NaN	2.0	1
13	1.0	1.0	2
14	NaN	3.0	2
15	NaN	1.0	2
16	2.0	NaN	2
17	NaN	1.0	2
18	NaN	NaN	1
19	2.0	NaN	2
20	NaN	2.0	2
21	2.0	NaN	2
22	1.0	NaN	1
23	NaN	NaN	2
24	2.0	NaN	2
25	NaN	3.0	2
26	NaN	4.0	1
27	1.0	1.0	1
28	NaN	NaN	2
29	2.0	NaN	2
...	...	...	...
5705	1.0	NaN	1
5706	NaN	NaN	1
5707	NaN	3.0	2
5708	2.0	NaN	2
5709	NaN	NaN	1
5710	2.0	NaN	2
5711	NaN	1.0	1
5712	NaN	2.0	2
5713	NaN	4.0	2
5714	1.0	2.0	2
5715	NaN	NaN	2
5716	NaN	NaN	1
5717	NaN	3.0	1
5718	NaN	3.0	2
5719	2.0	NaN	1
5720	NaN	1.0	2
5721	NaN	2.0	2
5722	NaN	1.0	2
5723	1.0	NaN	1
5724	NaN	NaN	2
5725	2.0	NaN	2
5726	NaN	2.0	1
5727	1.0	NaN	2
5728	NaN	3.0	2
5729	2.0	NaN	2
5730	2.0	NaN	1
5731	2.0	NaN	2
5732	NaN	1.0	1
5733	NaN	NaN	1

```
5734     NaN     2.0        2

[5735 rows x 3 columns]


In [21]: print(pd.isnull(da.DMDEDUC2).sum())
         print(pd.notnull(da.DMDEDUC2).sum())

261
5474
```