

```
In [67]: import numpy as np
import pandas as pd
pd.set_option('display.max_columns', 10) # panda set_option helps to set how many columns to show in display.
```

```
In [68]: df= pd.read_csv('C:\\Users\\A S U S\\Desktop\\aaa.csv')
```

```
In [69]: df
```

Out[69]:

	SEQN	PEASCST1	PEASCTM1	PEASCCT1	BPXCHR	...	CSXSLTRT	CSXSLTRG	CSXNAF
0	73557	1	620.0	NaN	NaN	...	62.0	1.0	NaN
1	73558	1	766.0	NaN	NaN	...	28.0	1.0	NaN
2	73559	1	665.0	NaN	NaN	...	49.0	1.0	NaN
3	73560	1	803.0	NaN	NaN	...	NaN	NaN	NaN
4	73561	1	949.0	NaN	NaN	...	NaN	NaN	NaN
...
9808	83727	1	611.0	NaN	NaN	...	NaN	NaN	NaN
9809	83728	1	124.0	NaN	110.0	...	NaN	NaN	NaN
9810	83729	1	679.0	NaN	NaN	...	55.0	1.0	NaN
9811	83730	1	381.0	NaN	72.0	...	NaN	NaN	NaN
9812	83731	1	498.0	NaN	NaN	...	NaN	NaN	NaN

9813 rows × 224 columns



```
In [11]: col_names= df.columns
col_names
```

```
Out[11]: Index(['SEQN', 'PEASCST1', 'PEASCTM1', 'PEASCCT1', 'BPXCHR', 'BPAARM',
               'BPACSZ', 'BPXPLS', 'BPXPULS', 'BPXPTY',
               ...,
               'CSXLEA0D', 'CSXS0A0D', 'CSXGRA0D', 'CSX0N0D', 'CSXNGS0D', 'CSXS
               LTRT',
               'CSXSLTRG', 'CSXNART', 'CSXNARG', 'CSAEFFRT'],
              dtype='object', length=224)
```

```
In [12]: keep = [column for column in col_names if 'BMX' in column]
```

```
In [13]: keep
```

```
Out[13]: ['BMXWT',
          'BMXRECUM',
          'BMXHEAD',
          'BMXHT',
          'BMXBMI',
          'BMXLEG',
          'BMXARML',
          'BMXARMC',
          'BMXWAIST',
          'BMXSAD1',
          'BMXSAD2',
          'BMXSAD3',
          'BMXSAD4']
```

```
In [15]: df_BMX= df[keep]
          df_BMX.head()
```

```
Out[15]:
```

	BMXWT	BMXRECUM	BMXHEAD	BMXHT	BMXBMI	BMXLEG	BMXARML	BMXARMC	BMXWA
0	78.3	NaN	NaN	171.3	26.7	39.2	40.2	35.3	1
1	89.5	NaN	NaN	176.8	28.6	40.0	41.5	34.7	1
2	88.9	NaN	NaN	175.3	28.9	40.0	41.0	33.5	1
3	32.2	NaN	NaN	137.3	17.1	33.5	29.5	21.0	
4	52.0	NaN	NaN	162.4	19.7	36.3	37.5	25.2	

```
df.loc[:, keep]
```

Out[16]:

	BMXWT	BMXRECUM	BMXHEAD	BMXHT	BMXBMI	BMXLEG	BMXARML	BMXARMC	BMXWRT
0	78.3	NaN	NaN	171.3	26.7	39.2	40.2	35.3	171.3
1	89.5	NaN	NaN	176.8	28.6	40.0	41.5	34.7	176.8
2	88.9	NaN	NaN	175.3	28.9	40.0	41.0	33.5	175.3
3	32.2	NaN	NaN	137.3	17.1	33.5	29.5	21.0	137.3
4	52.0	NaN	NaN	162.4	19.7	36.3	37.5	25.2	162.4
...
9808	71.8	NaN	NaN	171.3	24.5	41.4	35.2	29.9	171.3
9809	11.3	85.0	NaN	84.4	15.9	NaN	16.5	14.7	84.4
9810	89.6	NaN	NaN	162.3	34.0	40.0	37.6	37.0	162.3
9811	22.8	NaN	NaN	119.1	16.1	NaN	26.0	19.0	119.1
9812	42.3	NaN	NaN	148.1	19.3	35.2	31.7	25.0	148.1

9813 rows × 13 columns

```
index_bool = np.isin(df.columns, keep)
```

```
index_bool
```

```
Out[18]: array([False, False, False, False, False, False, False, False, False,  
                False, False, False, False, False, False, False, False, False,  
                False, False, False, False, False, False, True, False, True,  
                False, True, False, True, False, True, False, True, False,  
                True, False, True, False, True, False, True, True, True,  
                True, False, False, False, False, False, False, False, False,  
                False, False, False, False, False, False, False, False, False,  
                False, False, False, False, False, False, False, False, False])
```

[illegible]

```
In [21]: waist_median = pd.Series.median(df_BMX['BMXWAIST'])
         waist_median
```

Out[21]: 87.8

```
In [24]: df_BMX[df_BMX['BMXWAIST'] > waist_median].head()
```

Out[24]:

	BMXWT	BMXRECUM	BMXHEAD	BMXHT	BMXBMI	BMXLEG	BMXARML	BMXARMC	BMXWA
0	78.3	NaN	NaN	171.3	26.7	39.2	40.2	35.3	1
1	89.5	NaN	NaN	176.8	28.6	40.0	41.5	34.7	1
2	88.9	NaN	NaN	175.3	28.9	40.0	41.0	33.5	1
5	105.0	NaN	NaN	158.7	41.7	34.2	36.2	41.8	1
7	93.4	NaN	NaN	161.8	35.7	37.1	39.3	38.0	1

```
In [25]: # Lets add another condition, that 'BMXLEG' must be less than 32
condition1 = df.BMX['BMXWAIST'] > waist_median
```

```
condition2 = df_BMX['BMXLEG'] < 32
df_BMX[condition1 & condition2].head() # Using [] method
# Note: can't use 'and' instead of '&'
```

Out[25]:

	BMXWT	BMXRECUM	BMXHEAD	BMXHT	BMXBMI	BMXLEG	BMXARML	BMXARMC	BMX
56	55.4	NaN	NaN	145.9	26.0	30.0	33.7	27.5	
77	60.9	NaN	NaN	156.6	24.8	30.8	32.0	27.7	
108	70.7	NaN	NaN	142.6	34.8	30.7	33.8	37.0	
138	123.3	NaN	NaN	163.3	46.2	31.4	37.7	44.1	
180	78.4	NaN	NaN	161.4	30.1	31.9	35.3	30.2	

In [26]:

```
df_BMX.loc[condition1 & condition2, :].head()

# Using df.loc[] method for loc, rows come first then columns
# note that the conditions are describing the rows to keep
```

Out[26]:

	BMXWT	BMXRECUM	BMXHEAD	BMXHT	BMXBMI	BMXLEG	BMXARML	BMXARMC	BMX
56	55.4	NaN	NaN	145.9	26.0	30.0	33.7	27.5	
77	60.9	NaN	NaN	156.6	24.8	30.8	32.0	27.7	
108	70.7	NaN	NaN	142.6	34.8	30.7	33.8	37.0	
138	123.3	NaN	NaN	163.3	46.2	31.4	37.7	44.1	
180	78.4	NaN	NaN	161.4	30.1	31.9	35.3	30.2	

In [27]:

```
df_BMX.head()
```

Out[27]:

	BMXWT	BMXRECUM	BMXHEAD	BMXHT	BMXBMI	BMXLEG	BMXARML	BMXARMC	BMXWA
0	78.3	NaN	NaN	171.3	26.7	39.2	40.2	35.3	1

	BMXWT	BMXRECUM	BMXHEAD	BMXHT	BMXBMI	BMXLEG	BMXARML	BMXARMC	BMXWA
1	89.5	NaN	NaN	176.8	28.6	40.0	41.5	34.7	1
2	88.9	NaN	NaN	175.3	28.9	40.0	41.0	33.5	1
3	32.2	NaN	NaN	137.3	17.1	33.5	29.5	21.0	
4	52.0	NaN	NaN	162.4	19.7	36.3	37.5	25.2	

In [28]: df_BMX

Out[28]:

	BMXWT	BMXRECUM	BMXHEAD	BMXHT	BMXBMI	BMXLEG	BMXARML	BMXARMC	BMXWA
0	78.3	NaN	NaN	171.3	26.7	39.2	40.2	35.3	
1	89.5	NaN	NaN	176.8	28.6	40.0	41.5	34.7	
2	88.9	NaN	NaN	175.3	28.9	40.0	41.0	33.5	
3	32.2	NaN	NaN	137.3	17.1	33.5	29.5	21.0	
4	52.0	NaN	NaN	162.4	19.7	36.3	37.5	25.2	
...	
9808	71.8	NaN	NaN	171.3	24.5	41.4	35.2	29.9	
9809	11.3	85.0	NaN	84.4	15.9	NaN	16.5	14.7	
9810	89.6	NaN	NaN	162.3	34.0	40.0	37.6	37.0	
9811	22.8	NaN	NaN	119.1	16.1	NaN	26.0	19.0	
9812	42.3	NaN	NaN	148.1	19.3	35.2	31.7	25.0	

9813 rows × 13 columns

In [29]: *# Lets make a small dataframe and give it a new index so can more clearly see the differences between .loc and .iloc*

```
tmp = df_BMX.loc[condition1 & condition2, :].head()
```

```
tmp.index = ['a', 'b', 'c', 'd', 'e'] # If you use different years than
2015-2016, this my give an error. Why?
```

```
tmp
```

Out[29]:

	BMXWT	BMXRECUM	BMXHEAD	BMXHT	BMXBMI	BMXLEG	BMXARML	BMXARMC	BMXW
a	55.4	NaN	NaN	145.9	26.0	30.0	33.7	27.5	1
b	60.9	NaN	NaN	156.6	24.8	30.8	32.0	27.7	
c	70.7	NaN	NaN	142.6	34.8	30.7	33.8	37.0	1
d	123.3	NaN	NaN	163.3	46.2	31.4	37.7	44.1	1
e	78.4	NaN	NaN	161.4	30.1	31.9	35.3	30.2	1

In [30]:

```
tmp
```

Out[30]:

	BMXWT	BMXRECUM	BMXHEAD	BMXHT	BMXBMI	BMXLEG	BMXARML	BMXARMC	BMXW
a	55.4	NaN	NaN	145.9	26.0	30.0	33.7	27.5	1
b	60.9	NaN	NaN	156.6	24.8	30.8	32.0	27.7	
c	70.7	NaN	NaN	142.6	34.8	30.7	33.8	37.0	1
d	123.3	NaN	NaN	163.3	46.2	31.4	37.7	44.1	1
e	78.4	NaN	NaN	161.4	30.1	31.9	35.3	30.2	1

In [36]:

```
tmp.loc['a':'c', 'BMXLEG']
```

Out[36]:

a	30.0
b	30.8
c	30.7

Name: BMXLEG, dtype: float64

```
In [37]: tmp.loc[['a','c'],'BMXLEG']
```

```
Out[37]: a    30.0  
c    30.7  
Name: BMXLEG, dtype: float64
```

```
In [39]: tmp.iloc[[0, 2],5]
```

```
Out[39]: a    30.0  
c    30.7  
Name: BMXLEG, dtype: float64
```

```
In [46]: tmp.iloc[0:3,2] = [0]*3  
tmp
```

```
Out[46]:
```

	BMXWT	BMXRECUM	BMXHEAD	BMXHT	BMXBMI	BMXLEG	BMXARML	BMXARMC	BMXW/
a	55.4	NaN	0.0	145.9	26.0	30.0	33.7	27.5	1
b	60.9	NaN	0.0	156.6	24.8	30.8	32.0	27.7	
c	70.7	NaN	0.0	142.6	34.8	30.7	33.8	37.0	1
d	123.3	NaN	NaN	163.3	46.2	31.4	37.7	44.1	1
e	78.4	NaN	NaN	161.4	30.1	31.9	35.3	30.2	1

```
In [47]: tmp.iloc[:,2]  
tmp
```

```
Out[47]:
```

	BMXWT	BMXRECUM	BMXHEAD	BMXHT	BMXBMI	BMXLEG	BMXARML	BMXARMC	BMXW/
a	55.4	NaN	0.0	145.9	26.0	30.0	33.7	27.5	1
b	60.9	NaN	0.0	156.6	24.8	30.8	32.0	27.7	
c	70.7	NaN	0.0	142.6	34.8	30.7	33.8	37.0	1
d	123.3	NaN	NaN	163.3	46.2	31.4	37.7	44.1	1

	BMXWT	BMXRECUM	BMXHEAD	BMXHT	BMXBMI	BMXLEG	BMXARML	BMXARMC	BMXW/
e	78.4	NaN	NaN	161.4	30.1	31.9	35.3	30.2	1

```
In [61]: tmp.loc['a':'c', 'BMXBMI'] = [2]
tmp.loc[:, 'BMXBMI']
```

```
Out[61]: a      2
b      2
c      2
d     20
e     20
Name: BMXBMI, dtype: int32
```

```
In [49]: tmp
```

```
Out[49]:
```

	BMXWT	BMXRECUM	BMXHEAD	BMXHT	BMXBMI	BMXLEG	BMXARML	BMXARMC	BMXW/
a	55.4	NaN	0.0	145.9	1.0	30.0	33.7	27.5	1
b	60.9	NaN	0.0	156.6	1.0	30.8	32.0	27.7	
c	70.7	NaN	0.0	142.6	1.0	30.7	33.8	37.0	1
d	123.3	NaN	NaN	163.3	46.2	31.4	37.7	44.1	1
e	78.4	NaN	NaN	161.4	30.1	31.9	35.3	30.2	1

```
In [50]: tmp['BMXBMI'] = range(0, 5)
tmp
```

```
Out[50]:
```

	BMXWT	BMXRECUM	BMXHEAD	BMXHT	BMXBMI	BMXLEG	BMXARML	BMXARMC	BMXW/
a	55.4	NaN	0.0	145.9	0	30.0	33.7	27.5	1
b	60.9	NaN	0.0	156.6	1	30.8	32.0	27.7	
c	70.7	NaN	0.0	142.6	2	30.7	33.8	37.0	1

	BMXWT	BMXRECUM	BMXHEAD	BMXHT	BMXBMI	BMXLEG	BMXARML	BMXARMC	BMXW/
d	123.3	NaN	NaN	163.3	3	31.4	37.7	44.1	1
e	78.4	NaN	NaN	161.4	4	31.9	35.3	30.2	1

In [51]: tmp

Out[51]:

	BMXWT	BMXRECUM	BMXHEAD	BMXHT	BMXBMI	BMXLEG	BMXARML	BMXARMC	BMXW/
a	55.4	NaN	0.0	145.9	0	30.0	33.7	27.5	1
b	60.9	NaN	0.0	156.6	1	30.8	32.0	27.7	
c	70.7	NaN	0.0	142.6	2	30.7	33.8	37.0	1
d	123.3	NaN	NaN	163.3	3	31.4	37.7	44.1	1
e	78.4	NaN	NaN	161.4	4	31.9	35.3	30.2	1

In [56]: *# We will get a warning when using the [] method with conditions to set new values in our dataframe*

```
tmp[tmp.BMXBMI > 2]['BMXBMI'] = [10]*2 # Setting new values to a copy of tmp, but not tmp itself
tmp
```

You can see that the above code did not change our dataframe 'tmp'. This

D:\Anaconda\lib\site-packages\ipykernel_launcher.py:3: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

This is separate from the ipykernel package so we can avoid doing imports until

Out[56]:

	BMXWT	BMXRECUM	BMXHEAD	BMXHT	BMXBMI	BMXLEG	BMXARML	BMXARMC	BMXW
a	55.4	NaN	0.0	145.9	0	30.0	33.7	27.5	1
b	60.9	NaN	0.0	156.6	1	30.8	32.0	27.7	
c	70.7	NaN	0.0	142.6	2	30.7	33.8	37.0	1
d	123.3	NaN	NaN	163.3	3	31.4	37.7	44.1	1
e	78.4	NaN	NaN	161.4	4	31.9	35.3	30.2	1

In [55]: tmp.BMXWT

Out[55]: a 55.4
b 60.9
c 70.7
d 123.3
e 78.4
Name: BMXWT, dtype: float64

In [60]: *# The correct way to do the above is with .loc or .iloc*

```
tmp.loc[tmp.BMXBMI > 2, 'BMXBMI'] = [20]  
tmp
```

Now contains the chances

Out[60]:

	BMXWT	BMXRECUM	BMXHEAD	BMXHT	BMXBMI	BMXLEG	BMXARML	BMXARMC	BMXW
a	55.4	NaN	0.0	145.9	0	30.0	33.7	27.5	1
b	60.9	NaN	0.0	156.6	1	30.8	32.0	27.7	
c	70.7	NaN	0.0	142.6	2	30.7	33.8	37.0	1
d	123.3	NaN	NaN	163.3	20	31.4	37.7	44.1	1

	BMXWT	BMXRECUM	BMXHEAD	BMXHT	BMXBMI	BMXLEG	BMXARML	BMXARMC	BMXW/
e	78.4	NaN	NaN	161.4	20	31.9	35.3	30.2	1

In [62]:

tmp

Out[62]:

	BMXWT	BMXRECUM	BMXHEAD	BMXHT	BMXBMI	BMXLEG	BMXARML	BMXARMC	BMXW/
a	55.4	NaN	0.0	145.9	2	30.0	33.7	27.5	1
b	60.9	NaN	0.0	156.6	2	30.8	32.0	27.7	
c	70.7	NaN	0.0	142.6	2	30.7	33.8	37.0	1
d	123.3	NaN	NaN	163.3	20	31.4	37.7	44.1	1
e	78.4	NaN	NaN	161.4	20	31.9	35.3	30.2	1

In []: