PandasIndex_01

November 17, 2016

In [1]:
```
import pandas as pd
```

In [2]:
```
drinks = pd.read_csv('http://bit.ly/drinksbycountry')
```

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

Out[3]:

	country	beer_servings	spirit_servings	wine_servings	total_litres_of_pure_alcohol	continent
0	Afghanistan	0	0	0	0.0	Asia
1	Albania	89	132	54	4.9	Europe
2	Algeria	25	0	14	0.7	Africa
3	Andorra	245	138	312	12.4	Europe
4	Angola	217	57	45	5.9	Africa

In [4]:
```
drinks.index
```

Out[4]:
```
RangeIndex(start=0, stop=193, step=1)
```

In [5]:
```
drinks.columns
```

Out[5]:
```
Index(['country', 'beer_servings', 'spirit_servings', 'wine_servings',
       'total_litres_of_pure_alcohol', 'continent'],
      dtype='object')
```

In [6]:
```
drinks.shape
```

Out[6]:
```
(193, 6)
```

In [7]:
```
# if header is not specified -> default header is index of ints
pd.read_table('http://bit.ly/movieusers', header=None, sep='|').head()
```

Out[7]:

	0	1	2	3	4
0	1	24	M	technician	85711
1	2	53	F	other	94043
2	3	23	M	writer	32067
3	4	24	M	technician	43537
4	5	33	F	other	15213

In [8]:
```
drinks[drinks.continent == 'South America']
```

Out[8]:

	country	beer_servings	spirit_servings	wine_servings	total_litres_of_pure_alcohol	continent
6	Argentina	193	25	221	8.3	South America
20	Bolivia	167	41	8	3.8	South America
23	Brazil	245	145	16	7.2	South America
35	Chile	130	124	172	7.6	South America
37	Colombia	159	76	3	4.2	South America
52	Ecuador	162	74	3	4.2	South America
72	Guyana	93	302	1	7.1	South America
132	Paraguay	213	117	74	7.3	South America
133	Peru	163	160	21	6.1	South America
163	Suriname	128	178	7	5.6	South America
185	Uruguay	115	35	220	6.6	South America
188	Venezuela	333	100	3	7.7	South America

In [9]:
```
# allows you to pull out value from a particular column when specifying row
drinks.loc[23, 'beer_servings']
```

Out[9]: 245

In [10]:
```
# set country as index => name of index is country
drinks.set_index('country', inplace=True)
drinks.head()
```

Out[10]:

country	beer_servings	spirit_servings	wine_servings
Afghanistan	0	0	0

```
Albania                    89              132               54
Algeria                    25                0               14
Andorra                   245              138              312
Angola                    217               57               45

                total_litres_of_pure_alcohol continent
country
Afghanistan                            0.0      Asia
Albania                                4.9    Europe
Algeria                                0.7    Africa
Andorra                               12.4    Europe
Angola                                 5.9    Africa
```

```
In [11]: drinks.index
```

```
Out[11]: Index(['Afghanistan', 'Albania', 'Algeria', 'Andorra', 'Angola',
                'Antigua & Barbuda', 'Argentina', 'Armenia', 'Australia', 'Austria'
                ...
                'Tanzania', 'USA', 'Uruguay', 'Uzbekistan', 'Vanuatu', 'Venezuela',
                'Vietnam', 'Yemen', 'Zambia', 'Zimbabwe'],
               dtype='object', name='country', length=193)
```

```
In [12]: drinks.columns
```

```
Out[12]: Index(['beer_servings', 'spirit_servings', 'wine_servings',
                'total_litres_of_pure_alcohol', 'continent'],
               dtype='object')
```

```
In [13]: drinks.shape
```

```
Out[13]: (193, 5)
```

```
In [14]: # by assigning countries as index, now just
         # sort data by Country -> more meaningful to us in sorting data
         drinks.loc['Brazil', 'beer_servings']
```

```
Out[14]: 245
```

```
In [16]: # clear out index name => 'country'
         drinks.index.name = None
         drinks.head()
```

```
Out[16]:              beer_servings  spirit_servings  wine_servings  \
Afghanistan               0                0               0
Albania                  89              132              54
Algeria                  25                0              14
Andorra                 245              138             312
Angola                  217               57              45
```

```
                total_litres_of_pure_alcohol continent
Afghanistan                            0.0      Asia
Albania                                4.9    Europe
Algeria                                0.7    Africa
Andorra                               12.4    Europe
Angola                                 5.9    Africa
```

In [17]:
```
# reset index to numbers and country back to one of columns
# always set name before resetting index
drinks.index.name = 'country'
drinks.reset_index(inplace=True)
drinks.head()
```

Out[17]:
```
       country  beer_servings  spirit_servings  wine_servings  \
0  Afghanistan              0                0              0
1      Albania             89              132             54
2      Algeria             25                0             14
3      Andorra            245              138            312
4       Angola            217               57             45

   total_litres_of_pure_alcohol continent
0                           0.0      Asia
1                           4.9    Europe
2                           0.7    Africa
3                          12.4    Europe
4                           5.9    Africa
```

In [19]:
```
drinks.describe()
```

Out[19]:
```
       beer_servings  spirit_servings  wine_servings  \
count     193.000000       193.000000     193.000000
mean      106.160622        80.994819      49.450777
std       101.143103        88.284312      79.697598
min         0.000000         0.000000       0.000000
25%        20.000000         4.000000       1.000000
50%        76.000000        56.000000       8.000000
75%       188.000000       128.000000      59.000000
max       376.000000       438.000000     370.000000

       total_litres_of_pure_alcohol
count                     193.000000
mean                        4.717098
std                         3.773298
min                         0.000000
25%                         1.300000
50%                         4.200000
75%                         7.200000
max                        14.400000
```

```
In [23]: drinks.describe().index

Out[23]: Index(['count', 'mean', 'std', 'min', '25%', '50%', '75%', 'max'], dtype='

In [24]: drinks.describe().columns

Out[24]: Index(['beer_servings', 'spirit_servings', 'wine_servings',
                'total_litres_of_pure_alcohol'],
               dtype='object')

In [26]: # pulled out 25% value in beer_servings of 20.0
         drinks.describe().loc['25%', 'beer_servings']

Out[26]: 20.0

In [ ]:
```