

ChangeDataFileTypes

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	\
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 [4]: drinks.dtypes # objects -> strings
```

```
Out[4]:
```

country	object
beer_servings	int64
spirit_servings	int64
wine_servings	int64
total_litres_of_pure_alcohol	float64
continent	object
dtype:	object

```
In [5]: drinks['beer_servings']=drinks.beer_servings.astype(float) # over-write drinks
```

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

```
Out[6]:
```

country	object
beer_servings	float64
spirit_servings	int64
wine_servings	int64
total_litres_of_pure_alcohol	float64
continent	object
dtype:	object

```

In [8]: drinks = pd.read_csv('http://bit.ly/drinksbycountry', dtype={'beer_servings':
In [10]: drinks.dtypes
Out[10]: country                object
        beer_servings          float64
        spirit_servings         int64
        wine_servings          int64
        total_litres_of_pure_alcohol float64
        continent              object
        dtype: object

In [11]: orders=pd.read_table('http://bit.ly/chiporders')
In [12]: orders.head()
Out[12]:   order_id  quantity                item_name \
0         1         1      Chips and Fresh Tomato Salsa
1         1         1                        Izze
2         1         1      Nantucket Nectar
3         1         1  Chips and Tomatillo-Green Chili Salsa
4         2         2      Chicken Bowl

        choice_description  item_price
0                        NaN      $2.39
1          [Clementine]      $3.39
2            [Apple]      $3.39
3                        NaN      $2.39
4  [Tomatillo-Red Chili Salsa (Hot), [Black Beans...  $16.98

In [13]: orders.dtypes # item_price is stored as string
Out[13]: order_id          int64
        quantity          int64
        item_name         object
        choice_description object
        item_price         object
        dtype: object

In [14]: orders.item_price.str.replace('$', '').astype(float).mean() # you still ha
        # even if you remove $ -> still an object until you cast it explicitly as

Out[14]: 7.4643357853742973

In [15]: orders.item_name.str.contains('Chicken').head()
Out[15]: 0    False
        1    False
        2    False
        3    False
        4     True
        Name: item_name, dtype: bool

```

```
In [16]: orders.item_name.str.contains('Chicken').astype(int).head() # converts false to 0
```

```
Out[16]: 0    0  
         1    0  
         2    0  
         3    0  
         4    1  
         Name: item_name, dtype: int64
```

```
In [ ]:
```