Introducción a Pandas

¿Por qué usar Pandas?

Cuando se quiere trabajar con datos en tablas o datos estructurados (R dataframe, SQL, Excel, ...):

- Importar datos
- Limpieza de datos
- Exploración de datos
- Procesamiento para análisis
- Análisis de datos (acompañado de scikit-learn, statsmodels, ...)

Estructura de datos básicas

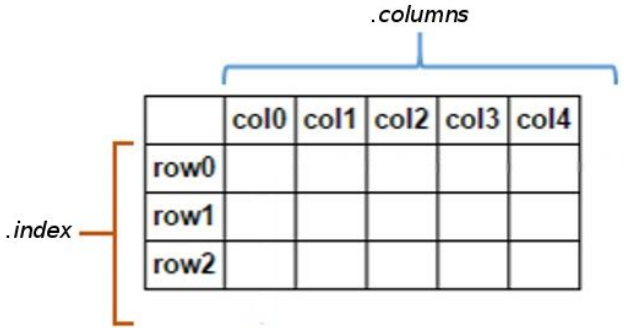
Pandas tiene 2 estructuras básicas, ambas contruidas a partir de NumPy arrays:

- Series object

```
s = pd.Series([0.1, 0.2, 0.3, 0.4])
```

- DataFrame object

```
data = {'country': ['Belgium', 'France', 'Germany', 'Netherlands', 'United Kingdom'],
        'population': [11.3, 64.3, 81.3, 16.9, 64.9],
        'area': [30510, 671308, 357050, 41526, 244820],
        'capital': ['Brussels', 'Paris', 'Berlin', 'Amsterdam', 'London']}
countries = pd.DataFrame(data)
```



Index y values

Las dos estructuras tienen índice y values

```
s = pd.Series([0.1, 0.2, 0.3, 0.4])
s.index
Out[8]:
Int64Index([0, 1, 2, 3], dtype='int64')
```

```
s = pd.Series([0.1, 0.2, 0.3, 0.4])
s.values
Out?
```

```
df = pd.DataFrame({'data':range(10)})
df.index
Out:
RangeIndex(start=0, stop=10, step=1)
```

```
df =
pd.DataFrame({'data':range(10),'data2':range(10,20)})
df.values
Out?
```

Operaciones básicas sobre DataFrame

```
df.columns
Out:
Index([u'data', u'data2'], dtype='object')
```

```
df.dtypes
Out[29]:
data             int64
data2             int64

df.info()

df.set_index('data')
df['data2']
```

Cambiando el índice:
`df.set_index('data')`

Eligiendo una única columna
`df['data2']`

Reordenando:

```
df.sort_index(by='data2', ascending=False)
```

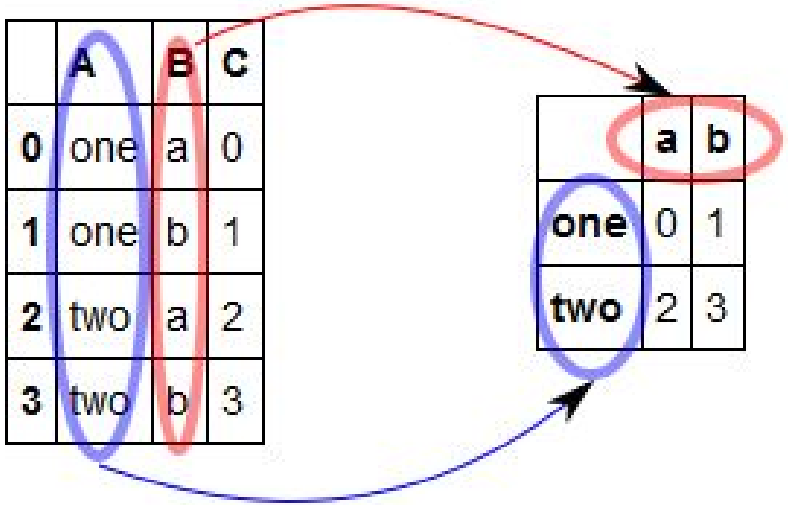
Reshape

df = df.set_index(['A','B'])

		C
A	**B**	
one	**a**	0
	b	1
two	**a**	2
	b	3

df.unstack()

B	**a**	**b**
A		
one	0	1
two	2	3



Leyendo y escribiendo archivos

http://pandas.pydata.org/pandas-docs/stable/api.html#input-output

```
df = pd.read_excel(open(filename))

df = pd.read_csv(open(filename))

df = pd.read_json(open(filename))
```

```
df.to_csv('archivo.csv',index=False,encoding='utf-8')

df.to_excel('archivo.xls')
```

¡A trabajar!