

Python Pandas

- `pip3 install pandas`

- soporte para excel:

```
pip3.9 install openpyxl # needed to load Excel .xlsx files
pip3.9 install xlrd # needed to load Excel old .xls files)
```

- Estilos: https://docs.bokeh.org/en/latest/docs/user_guide/styling/plots.html#ug-styling-plots

varios

importar en dataframe

- `pandas.read_csv('data.txt', header=None)`
- `pandas.read_json('data.json')`

columnas

- asignar nombres de columna

```
df = pandas.read_csv('data.txt', header=None)
df.columns = ["Col1", "Col2", "Col3"]
```

- asignar indice:

```
df.set_index("nom_col", inplace=True, drop=False)
```

- `inplace`: los cambios se aplican en el mismo objeto (variable), Por defecto, los cambios se deberán guardar en otra variable.
- `drop`: Con `False`, no saca la columna de los datos para convertirla en índice

manipulación datos

Para acceder a datos en el dataframe es como un sistema de coordenadas

label based indexing

- `loc[start_value_row;final_value_row,start_column_header:final_column_header]`

```
df.loc["valor": "valor", "columna": "columna"]
list(df.loc[:, "Columna"]) # devuelve como lista los valores de la columna, para todas las filas
```

- se puede indicar solo un valor en los rangos

position based indexing

`iloc[start_value_row;final_value_row,start_column_header:final_column_header]`

```
df.iloc[1:4,2:5]
```

- `iloc` se comporta como cualquier **slice** en python, el último valor no está incluido. Así, en el ejemplo cogería las filas 1,2,3 y las columnas 2,3,4

borrar columnas

- `drop(«fila»|«columna»)`
- los cambios no son `inplace`

```
df.drop("filas",0)
df.drop(df.index[0:3],0) # elimina las filas 0,1,2
```

pandas + bokeh ejemplos

```
# segundo bokeh plot con pandas
from bokeh.plotting import figure
from bokeh.io import output_file, show
import pandas

df=pandas.read_csv("https://pythonizing.github.io/data/bachelors.csv")

x=df["Year"]
y=df["Engineering"]

output_file("Line.html")

f=figure()
f.line(x,y)

show(f)
```

```
import pandas

from bokeh.plotting import figure, output_file, show

df=pandas.read_excel("https://github.com/pythonizing/data/blob/master/verlegenuken.xlsx", sheet_name=0)
df["Temperature"]=df["Temperature"]/10
df["Pressure"]=df["Pressure"]/10

p=figure(width=500,height=400,tools='pan')

p.title.text="Temperature and Air Pressure"
p.title.text_color="Gray"
p.title.text_font="arial"
p.title.text_font_style="bold"
p.xaxis.minor_tick_line_color=None
```

```
p.yaxis.minor_tick_line_color=None
p.xaxis.axis_label="Temperature (°C)"
p.yaxis.axis_label="Pressure (hPa)"

# old versions:
# p.circle(df["Temperature"],df["Pressure"],size=0.5)

p.scatter(x=df["Temperature"],y=df["Pressure"],size=0.5)
output_file("Weather.html")
show(p)
```

- desde Jupyter Notebook problema leyendo fichero desde URL

From:
<https://miguelangel.torresegea.es/wiki/> - **miguel angel torres egea**

Permanent link:
<https://miguelangel.torresegea.es/wiki/development:python:pandas?rev=1728566732>

Last update: **10/10/2024 06:25**

