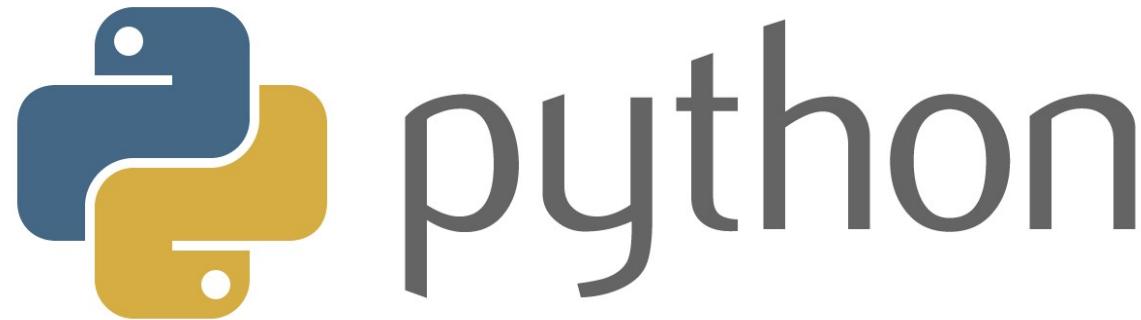




HERRAMIENTAS EN PYTHON PARA DATA SCIENCE

DSV @ mfrasquet



Python 2 (2.7 última versión 2010)

Python 3 (En desarrollo desde 2014 – 3.5.2)

Problemas de compatibilidad -> Virtual environment



El viaje.....

**OBTENER
DATOS**



**PREPARAR
DATOS**



**PROCESAR
DATOS**



OBTENER DATOS



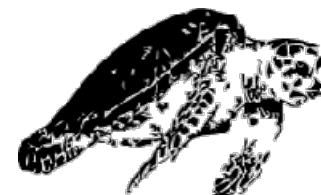
Bases de datos
“publicas”



Scraping & web crawling



(Compatible con Py3, pero
mejor Py2)



Requests
(urllib3)

LXML
HTML5lib

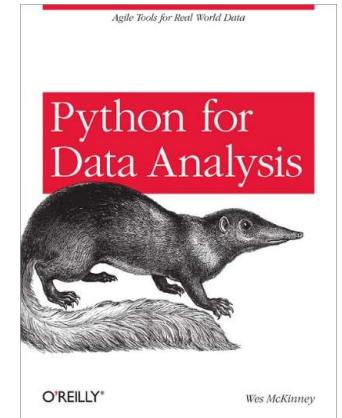
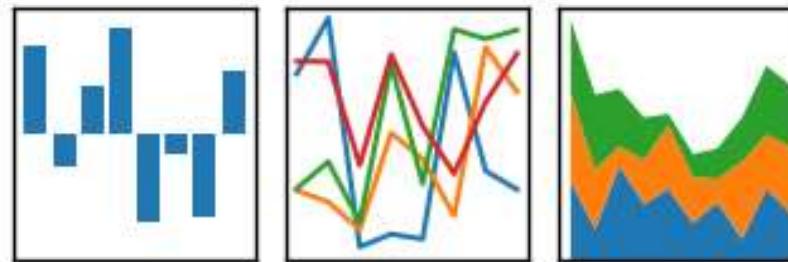
Beautiful Soup

PREPARAR DATOS

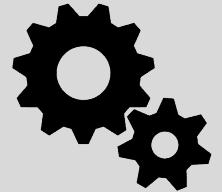


pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



PROCESAR DATOS



Basic Machine learning



CLASIFICACIÓN

K nearest neighbors

Support vector machine

Arboles de decisión

REGRESIÓN

Lasso

Least squares

Ridge

CLUSTERIZACIÓN

K-means

Deeplearning

TORCH

*FaceBook
Twitter*

THEANO

University of Montreal



Google



EXTRA 1- CONFERENCIAS



Pycones 2016

- <https://github.com/CAChemE/stochastic-optimization> "Optimización estocástica"-
- <https://github.com/intiveda/pyreglasasociacion> - "Reglas de Asociación"- <https://github.com/DiegoHueltes/bcolz-bquery-pycones2016> - "Bcolz"-
- https://github.com/manugarri/pycones_2016_talk_recommenders - A primer on recommenders"-
- <https://github.com/AeroPython/Taller-Algoritmos-Geneticos-PyConEs16> - "Algoritmos genéticos"-
- <https://github.com/Peque/PyConES-Spain-2016-osbrain> "OSBRain"- <https://github.com/Peque/PyConES-Spain-2016-trading> "Plataforma trading"-
- <https://github.com/javiromero/pycones2016> "Depurando como un pro"- <https://github.com/CAChemE/Pyomo-Diet> "Optimización dieta"



VIII Jornadas de Usuarios de R

Muchas ponencias interesantes

Taller paralelo - SVM Concepto y aplicaciones

Taller paralelo – Procesamiento de datos masivos con SParkR

