

Introducción a Python.

Visualización y Preprocesamiento de datos.

Práctica 1

Requisitos: La distribución de Python Anaconda, con una versión de Python 3.6 o superior. Podés encontrar un tutorial sobre como instalar Anaconda y ejecutar python o jupyter en:

<http://facundoq.github.io/courses/aa2018/jupyter.html>

- 1- **Tutorial de Python.** Desde una terminal de Anaconda ubicarse en el directorio donde se encuentra los archivos con extensión “.ipynb” (notebooks) y escribir: “Jupyter notebook”. Desde el entorno web, acceder a la dirección del servidor (generalmente <http://localhost:8888/>), y abrir el tutorial “01_Tutorial de Python con Jupyter Notebook.ipynb” y realizarlo.
- 2- **Tutorial de NumPy.** Desde el entorno web, abrir el tutorial “02_Tutorial de Numpy.ipynb” y realizarlo.
- 3- El archivo “Sopas.xls” contiene información de diferentes sopas incluyendo el costo, las calorías, grasas, etc. Complete el código del archivo `sopas.py` para:
 - a. Calcular el valor promedio del atributo “grasa”.
 - b. Contar la cantidad de sopas del tipo “CC”.
 - c. Encontrar e imprimir la sopa con más sodio (y su valor).
Investigue las funciones que posee NumPy para calcular de forma eficiente el máximo, mínimo, promedio, etc.
- 4- El archivo “iris.csv” contiene información de 3 especies de flores.
 - a. Utilice la librería Pandas para abrir el archivo.
 - b. Haga un gráfico tipo “scatter_matrix” para visualizar todos los pares de variables.
 - c. Realice un histograma de todas las variables (`dataFrame.hist()`).
- 5- Realice el mismo procedimiento que para el ejercicio 4 pero con el *dataset* “Sopas.xls”.
- 6- Para el dataset “iris.csv” normalice cada variable (*feature*) con “min-max” y con “z-score”. Grafique con histogramas para verificar los resultados.