

DATA SCIENCE

MÓDULO 1

Entornos (Docker
Kitematic y Jupiter
Notebooks).
Repaso de Python.

1 Repasar los conceptos aprendidos durante el pre-work

2 Incorporar nociones básicas de Docker y Kitematic

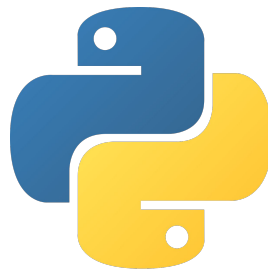
3 Familiarizarse con el entorno de Jupyter Notebook

4 Incorporar nociones básicas de POO

PYTHON



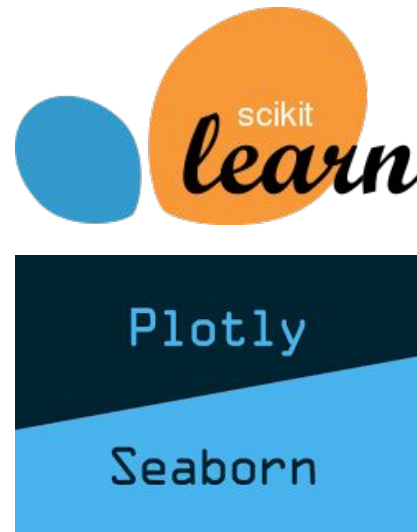
- Es un lenguaje de programación interpretado cuya filosofía hace hincapié en una sintaxis muy limpia y un código legible.
- Es un lenguaje muy sencillo y fácil de aprender.
- Su sintaxis es fácil de entender puesto que es cercana al lenguaje natural.
- Es de código abierto, nos permite crear nuestras aplicaciones e involucrarnos en su desarrollo.
- Tiene una gran comunidad dispuesta a ayudarnos con nuestro código.



Son implementaciones de recursos que se incorporan en un entorno para aprovechar su funcionalidad.

Por ahora, podemos dividir las en aquellas que nos sirven para:

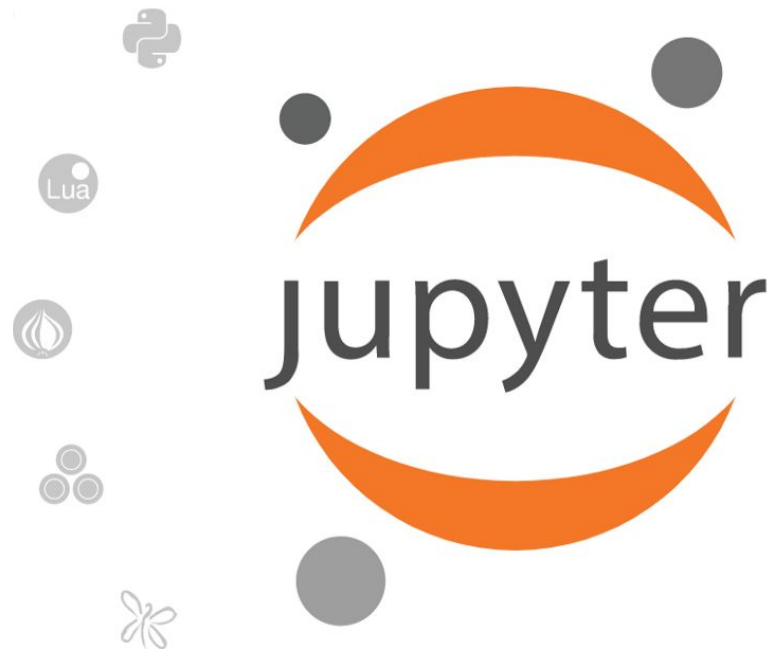
- Realizar operaciones matemáticas complejas.
- Realizar representaciones de resultados (gráficas).
- Manipular otros recursos del sistema.



Algunas de las bibliotecas que vamos a usar en este curso:

- **NumPy:** Permite trabajar de manera eficiente con operaciones matemáticas sobre arrays.
- **Pandas:** Se usa para análisis y manipulación de datos como tablas (dataframes).
- **Matplotlib:** Es una librería de visualización, para gráficos y tableros.
- **Seaborn:** Otra librería de visualización, basada en matplotlib.

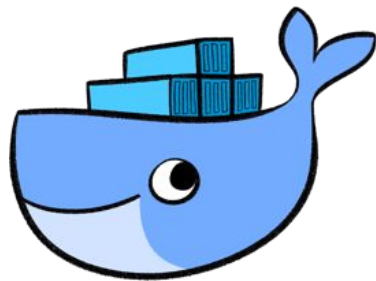
- Es un entorno de trabajo interactivo que permite:
 - desarrollar código (Python, R, Java, etc) de manera dinámica.
 - Integrar en un mismo documento tanto bloques de código como texto, gráficas o imágenes.
- Utilizado ampliamente en análisis numérico, estadística y machine learning, entre otros campos de la informática y las matemáticas.



DOCKER Y KITEMATIC



- Docker es una herramienta que se utiliza para aislar entornos a nivel de sistema operativo, en linux.
- Estos entornos se conocen como contenedores y se crean a partir de imágenes.
- Para poder hacerla funcionar en windows, debemos incluir software que nos permita ejecutar un kernel linux.
- Docker trae estas herramientas para la configuración de la virtualización de un kernel linux en windows se conoce como docker-machine.



- Es una aplicación gráfica para la gestión y ejecución de Docker.
- Permite buscar las imágenes (está conectado al repositorio docker-hub), levantar, inicializar, detener, y borrar los contenedores, configurar redes entre contenedores, etc.
- Aún se encuentra en fase de desarrollo (actualmente está en versión alpha) y presenta algunos problemas.





+NEW : Construir un nuevo contenedor a partir de una imagen

Stop/Start : Para detener o iniciar un contenedor

Restart : Reinicia un contenedor

Exec : Abre una terminal al interior del contenedor

Así se ve docker-hub en Kitematic

The screenshot shows the Kitematic Docker Hub interface. On the left, under 'Containers', there are two containers: 'examples_db_1' (mysql:5.7) and 'examples_wordpress_1' (wordpress:latest). The 'examples_wordpress_1' container is highlighted with a purple box, and a red arrow points from it to the text 'Contenedor' and a red arrow points from it to the text 'Nombre de la imagen'. The main area is titled 'Buscador de imágenes' (Image Search) and shows a grid of recommended Docker images. The 'official rethinkdb' image is highlighted with a purple box. A red arrow points from the search bar at the top to the 'Buscador de imágenes' title. A green arrow points from the 'CREATE' button of the 'official rethinkdb' image to the text 'Repositorio'. A red arrow points from the 'CREATE' button of the 'official rethinkdb' image to the text 'Bajar imagen'. The 'official rethinkdb' image details are shown on the right, with the text 'Detalles del repo' in green.

Containers

examples_db_1
mysql:5.7

examples_wordpress_1
wordpress:latest

Contenedor

Nombre de la imagen

Search for Docker Images from Docker Hub

FILTER BY All Recommended My Repos My Images

Recommended

kitematic
hello-world-nginx
A light-weight nginx container that demonstrates the features of Kitematic
♡ 81 ↓ 3M ... CREATE

ghost
official
ghost
Ghost is a free and open source blogging platform written in JavaScript
♡ 591 ↓ 6M ... CREATE

official
jenkins
Official Jenkins Docker image
♡ 2.9K ↓ 27M ... CREATE

official
redis
Redis is an open source key-value store that functions as a data structure server.
♡ 3.9K ↓ 288M ... CREATE

official
rethinkdb
RethinkDB is an open-source, document database that makes it easy to build and scale realtime...
♡ 411 ↓ 7M ... CREATE

official
solr
Solr is the popular, blazing-fast, open source enterprise search platform built on Apache...
♡ 400 ↓ 3M ... CREATE

official
elasticsearch
Elasticsearch is a powerful open source search and analytics engine that makes data easy to...
♡ 2.3K ↓ 77M ... CREATE

official
postgres
The PostgreSQL object-relational database system provides reliability and data integrity.
♡ 3.7K ↓ 65M ... CREATE

Repositorio

Detalles del repo

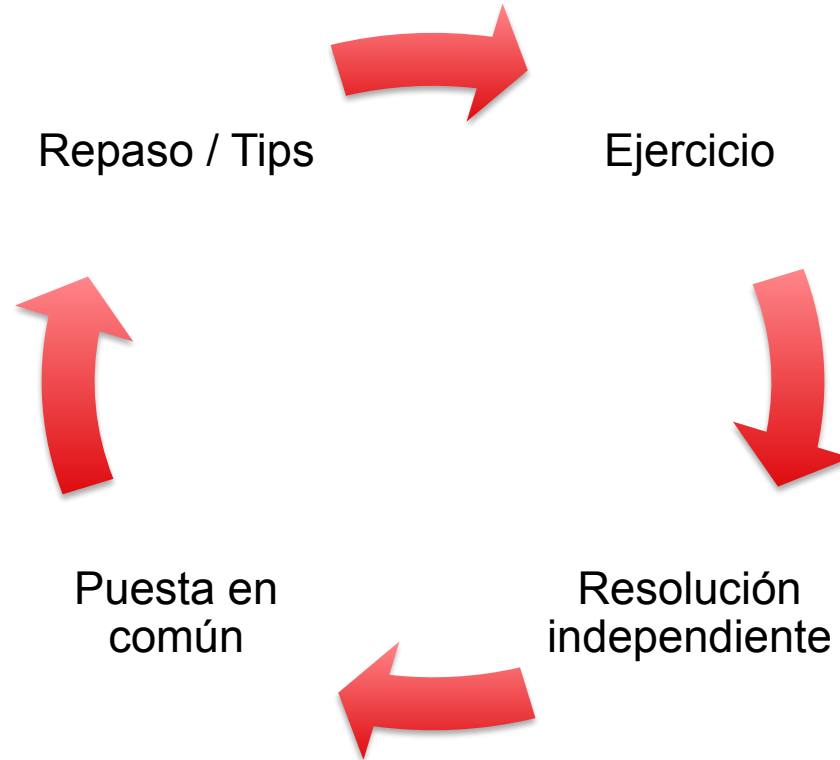
Bajar imagen

Recomendamos que

1. Instalar un sistema operativo (para arrancar, Ubuntu 18.04 LTS es optimo <https://www.ubuntu.com/download/desktop>)
2. Instalar <https://docs.docker.com/install/linux/docker-ce/ubuntu/> (No usar ninguna otra forma de instalacion. Ej: snap)
3. Instalar Jupiter Notebooks (para arrancar, la opción de pip3 es la mas recomienda <https://jupyter.org/install>)
4. Descargar el contenedor de Data Science de Digital House (con el comando `docker run -p 8888:8888 --name notebooks dsdh/data`)

REPASO DE PYTHON





- **Lección 1: Fundamentos Python** (1. Líneas y espacios 2. La consola 3. Comando print() 4. Strings 5. Concatenar strings 6. Manejo de errores 7. Variables 8. Números 9. Operaciones 10. Operaciones con variables 11. Tipos de división 12. Comentarios 13. Booleanos)
- **Lección 2: Listas, diccionarios, tuplas** (1. Listas, 2. Acceder a los elementos de una lista 3. Operaciones con listas 4. Método append() 6. Método remove() 7. Diccionarios 8. Accediendo a los valores de un diccionario 9. Actualizando diccionarios 10. Tuplas)
- **Lección 3: Bucle for** 5. Método range() 6. Contador 7. Acumulador 8. Integrando
- **Lección 4: Condicionales** 1. if 2. else 3. Encadenar condiciones 4. Expresión condicional 5. Operadores lógicos 6. Indentación y condicionales anidados
- **Lección 5: Funciones** 2. Definiendo funciones 4. Argumentos 6. Funciones y listas 7. Integrando estructuras 8. Combinando funciones
- **Lección 6: Listas por comprensión**
- **Lección 7: Diccionarios por comprensión**

Un conjunto descrito por **extensión** es aquel que enumera uno a uno a todos sus elementos.

Por ejemplo,

$$C = \{2, 4, 6, 8, 10\}$$

$$S = \{1, 4, 9, 16, 25, 36, 49, 64, 81, 100\}$$

Un conjunto descrito por **comprensión** es aquel que determina las propiedades que caracterizan a sus elementos

Por ejemplo,

$$C = \{n \mid n \text{ es un número par y } 1 \leq n \leq 10\}$$

$$S = \{n \mid n = k^2, k \text{ es natural y } n \leq 100\}$$

La sintaxis de las listas por comprensión es:

[*expresión* for *nombre* in *lista* [if *condicion*]]

[*n* for *n* in *lista* if *n%2==0*]

- **Facilidad y velocidad para el data scientist:** Las Listas por comprensión en Python son una sintaxis simple y poderosa que, una vez dominada, permite una manipulación rápida, eficiente e intuitiva de tipos de datos como arreglos o listas.
- **Código conciso y fácil de leer:** Puede ser que las listas por comprensión resulten confusas en un principio. Sin embargo, una vez adquirido el hábito y comprendidas tornan un código complejo en uno conciso y fácil de leer.
- **Reemplaza las estructuras iterativas (while / for):** Las listas por comprensión son esencialmente un reemplazo para sentencias de control iterativas. Compararemos las alternativas de utilizar y no utilizar comprensión para ilustrar su funcionamiento y sus ventajas.

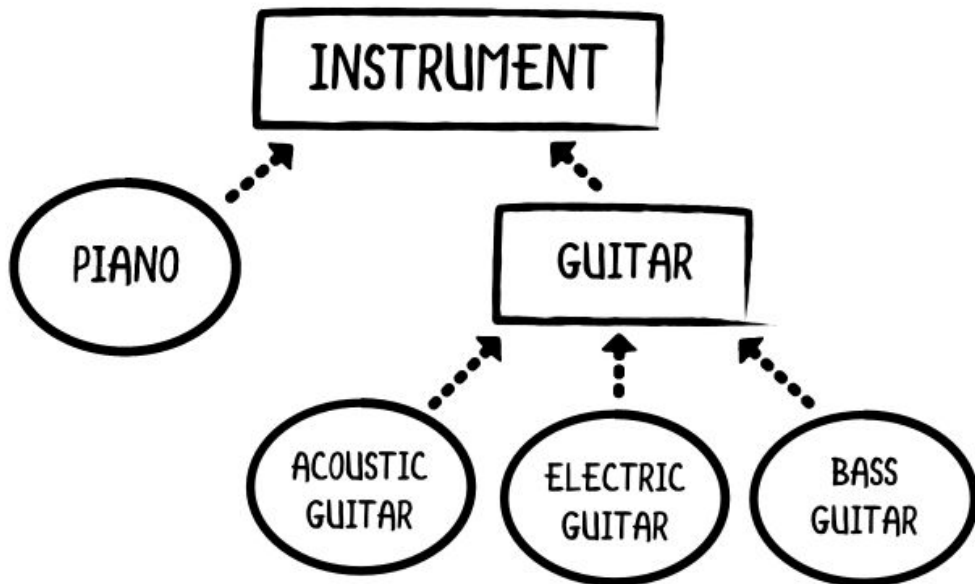
Programación Orientada a Objetos (POO)



Los objetos puede ser de distintos tipos:

- Listas []
- Tuplas ()
- Strings ""
- Clase (**Class**)
 - ◆ Funcion (**def**)
 - Variable (*cualquier*)

Para que un objeto exista, tengo que cargarlo en memoria en un proceso llamado instanciación.



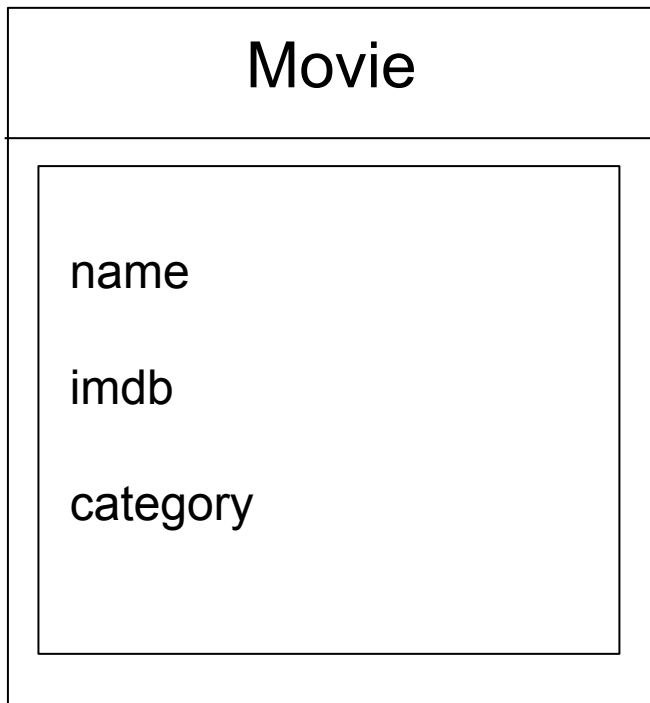
PRÁCTICA GUIADA



— Practiquemos estos conceptos con una práctica guiada

PRÁCTICA INDEPENDIENTE





Qué es una movie en nuestra práctica independiente?

Una diccionario que contiene características de una película.

Es una primer intuición de un 'objeto' en Python.

— Apliquemos lo aprendido en una práctica independiente