

# Computer Assisted Audit Techniques - Ejemplos de USO

---

## Demo: análisis de calidad de datos con Python

Para este demo se va a usar Docker. Entonces conviene tenerlo instalado antes de comenzar se comparte el siguiente enlace a su tutorial oficial de instalación.

- Instalar docker. [Tutorial](#)

El demo también podría ser ejecutado sin contar con Docker. En este caso sería necesario contar con una instalación de Python 3, desde ahí existen dos opciones:

- Generar un [entorno virtual](#) para instalar ahí los paquetes necesarios (recomendado).
- Instalar los paquetes necesarios vía pip en la instalación global de python.

## Generar la imagen a utilizar

Como se menciona en la presentación, hay ocasiones en las que conviene utilizar una imagen propia para que contenga todo lo necesario para un proyecto. O al menor establecer un único método para personalizarla.

Esto se puede hacer a través de un archivo **Dockerfile**. Vamos a crearlo con el siguiente contenido:

```
FROM python:3

WORKDIR /src

RUN pip3 install jupyterlab
RUN pip3 install pandas
```

## Qué es eso?

- `FROM python:3` sirve para indicar la imagen base sobre la que se va a trabajar.
- `WORKDIR /src` indica que esa ubicación será en la que se van a ejecutar los siguientes comandos del archivo.
- `RUN pip3 install [paquetes]` es la orden para ejecutar la instalación de los paquetes a utilizar en el proyecto en la imagen.

En este momento, después de generar el archivo, se podría generar una imagen con sus características, para eso se tiene que ejecutar el siguiente comando (**estando en la misma ubicación** del Dockerfile):

```
docker build -t jupyter .
```

### Descripción del comando:

- `build` es la orden para que docker genere una nueva imagen.
- `-t jupyter` es el parámetro a utilizar para definir el nombre de la imagen a crear.
- `.` es la indicación de la ubicación del Dockerfile que genera la imagen.

Con la imagen generada se puede pasar a ejecutarla, pero en este caso se deberá hacer con el siguiente comando:

```
docker run --rm -it -v $(pwd):/src -p 8080:8080 jupyter:latest jupyter
notebook --ip=0.0.0.0 --port=8080 --allow-root
```

**Recomendación:** es conveniente ejecutar esta instrucción estando posicionado sobre un directorio desde el que se pueda llegar a la carpeta donde se encuentran estos archivos.

### Consideraciones previas

Se va a analizar un dataset obtenido a partir de una base de datos a auditar utilizando para ello una serie de evaluaciones implementadas en lenguaje Python dispuestas en una libreta Jupyter. El dataset se encuentra en la carpeta **datasets** y las libretas en la carpeta **libretas**.

Por otra parte, en la carpeta **docs** se encuentra la documentación de referencia que representa el conocimiento del dominio sobre el cual está basado el dataset. Además, encontrarán un archivo guía para desarrollar el análisis.

## Pasos a seguir

1. Cuando inicia el contenedor con Jupyter al final muestra una URL a utilizar, acceder a ella desde el navegador web
2. Ingresar a la carpeta **work** en donde va a estar espejada la carpeta actual del demo
3. Abrir la libreta **Calidad de datos**
4. Ejecutar las diferentes celdas de la libreta e ir documentando sus resultados
5. Generar el reporte del uso de la herramienta para el registro de las evidencias que se encuentren

## Apagado del contenedor

```
Ctrl + C
```

### Referencias

[Documentación Jupyter Pandas](#)