

# Unit 7

## Data Science

Data science is the process of learning about the world using data and computation.

It is basically the study of data to extract meaningful insight about the population, resources or business. For these studies you can use mathematics, statistics, IA and computer engineering to analyze large amounts of data.

It includes:

- Exploring data sets (save, clean, explore)
- Using statistics
- Creating data visualizations
- Identifying patterns in data

What structure can we use to save data in Python?

Let's start by exploring data sets. We're going to use pandas which is a Python library for data manipulation and analysis

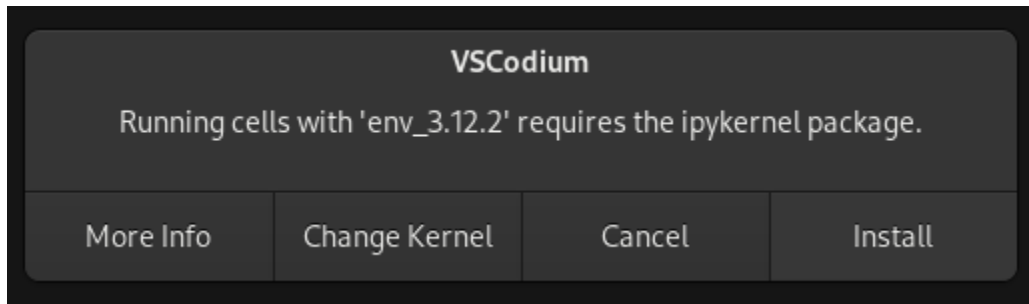
### Set up tools

- Install a new python venv.
- Activate the environment.
- Install the package pandas (pip install pandas).
- If you are using vsodium, create a file with the extension .ipynb. This is the extension for a jupyter notebook.

Open the notebook, maybe vsodium is going to ask you to install an extension (jupyter), install it.

I will advise you to create another .py (example: test.py) file, just to set up the env for this file, so you will be able to select this env as a kernel for the notebook (otherwise the env does not show up as a selection for the kernel).

When you execute a cell in the notebook, vscodeium will ask you to install the ipykernel in the env, click on install:



- If you are **NOT** using vscodeium, you will need to use a jupyter lab server. So go ahead, and install the jupyterlab (pip install jupyterlab), and then launch a jupyter server with the command jupyter lab, and create a notebook.