Pythons Pandas Library:

Leveraging Open-Source Tools for Data Analysis

Samuel Gelman,

Weizmann Institute: LSCF, Bioinformatics unit



Programming

Laboratory Information Management System (LIMS)

Programming Pipelines for Analysis (UTAP and more)





Deep Sequencing Data Analysis (NGS)

Weizmann LSCF

Bioinformatics unit

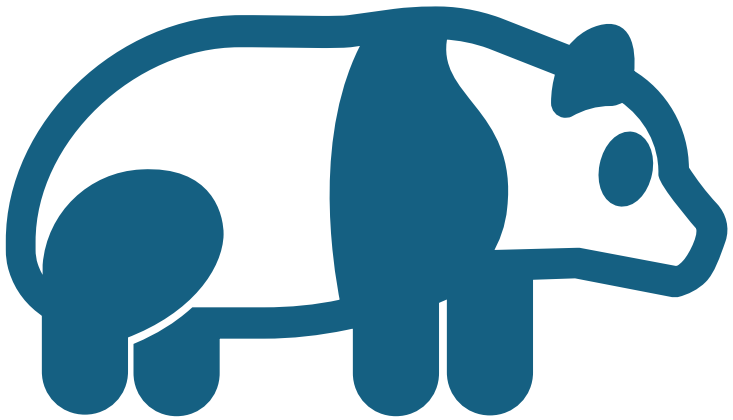

Sequence Analysis and CRISPR Design

Statistics, Machine Learning & Artificial Intelligence



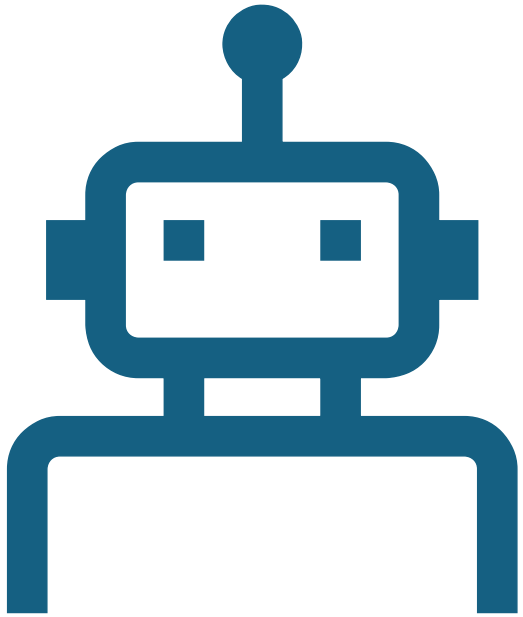
What will I learn?

- The what?
- The how?



What

- What is Pandas and where does it fall in relation to Python as a language.
- Useful tools which help make the most out of Pandas.
- Capabilities and functionalities.



How

- How to work with Pandas through a hands-on approach.
- Familiarizing and experimenting with its capabilities.
- How to dream

Outline

- Pandas Ecosystem.
- Birds-eye-view of Pandas and the workbook
- Getting our hands dirty and writing some code.

Other tools

- A few other tools to help us along the way

Anaconda

- Anaconda streamlines the setup of Python environments.
- Anaconda is the backbone for setting up the Python environment needed to run our Pandas workshop.
- Supports many languages.

Jupyter

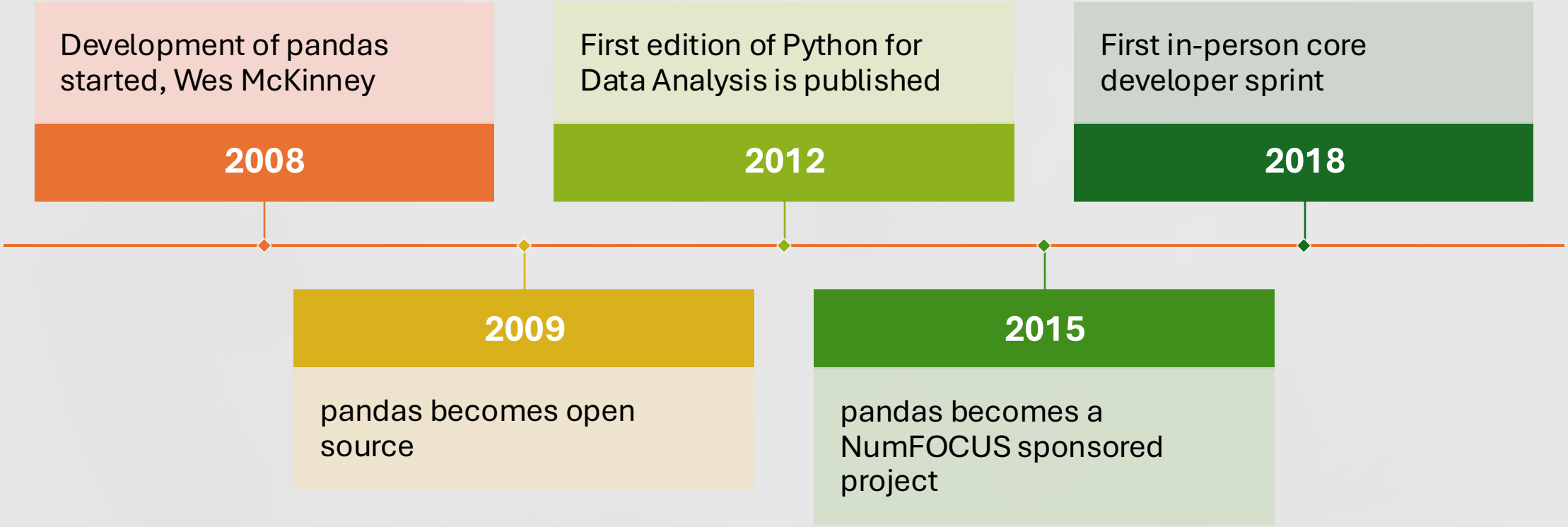
- Jupyter provides an interactive computing environment where you can write, run, and visualize code.
- Widely adopted in academia and industry.
- This is where we will be doing all our hands-on work.

Pandas namesake?

Panel Data + Python Data Analysis = Pandas

What is Pandas?

Short History



Series

One-dimensional labeled array capable of holding any data type.

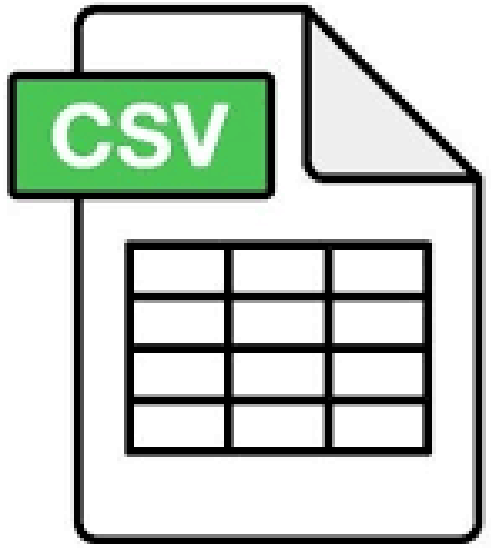
DataFrames:

A DataFrame in Pandas is a two-dimensional labeled data structure capable of holding data of various types in a tabular format

DataFrames can be thought of as a collection of Pandas Series where each Series represents a column.

With many built-in methods!!

Loading in data.



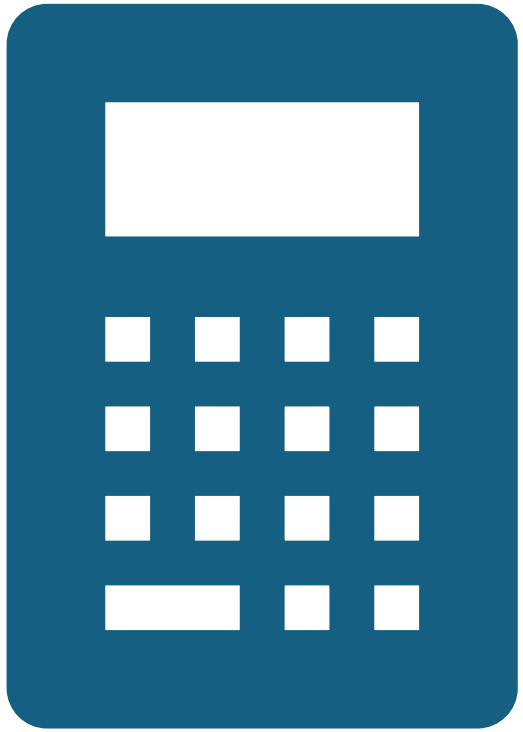
Subsets of the data

COLUMN

INDEX

LOC & ILOC

SLICING

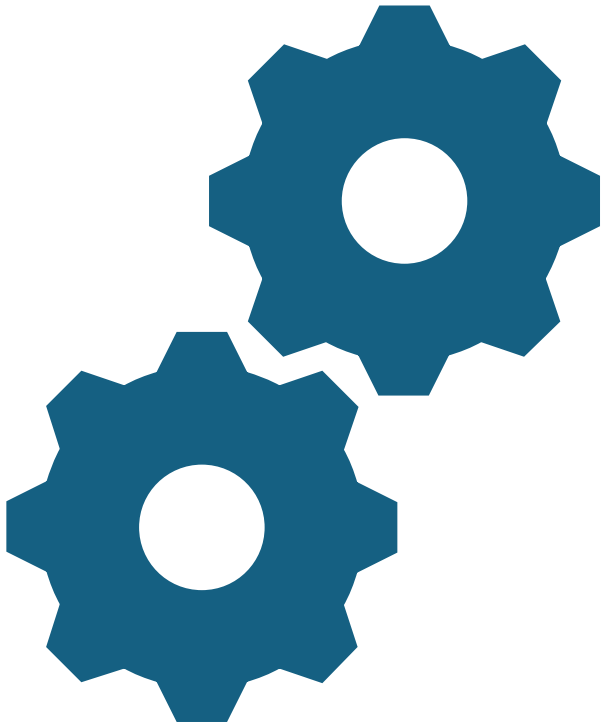


Mathematical Operations

Masks

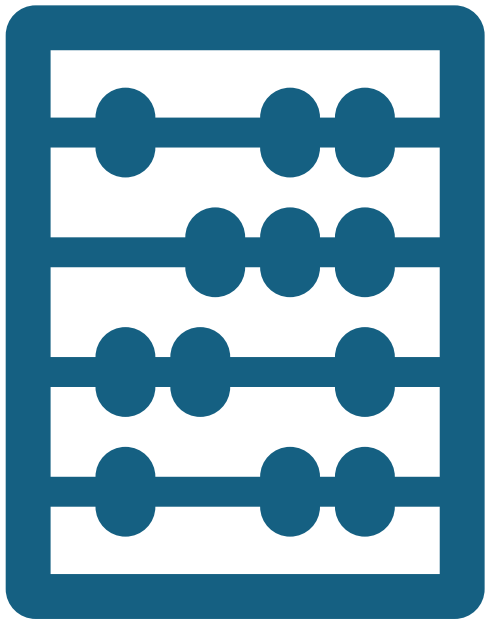
Time

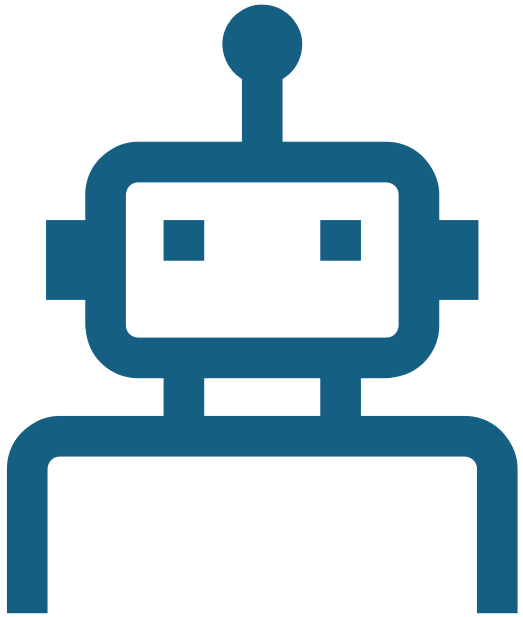
Groupby



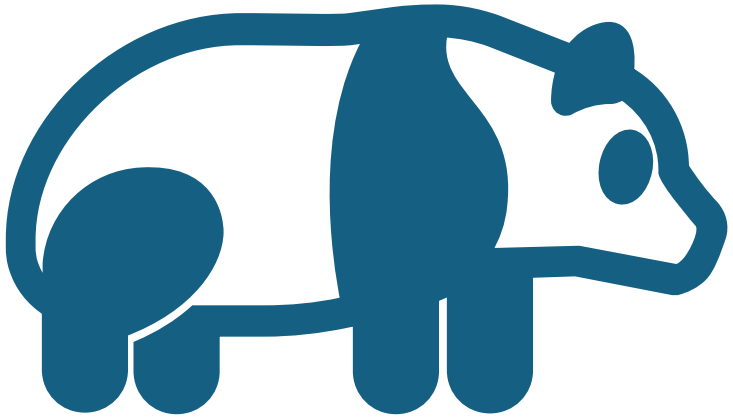
Feature Engineering

One-Hot Encoding





Machine Learning Algorithms



Let's Go!