Python programming for data science

marco milanesio
MScDSAI 2023-2024

Who am I?

- PhD [2010]
  - Distributed systems
  - Network measurements and performances
  - Distributed storage
- @Inria
  - Image analysis
  - Distributed Learning
- @MSI - MDLab
  - Dev-ops
  - Databases
  - DNA sequencing
- @MSI - IRCAN
  - Computer vision
  - Deep Learning
  - Marine biology

Inria
inventeurs du monde numérique

Epione
e-patient / e-medicine

UNIVERSITÉ CÔTE D'AZUR
CENTER OF MODELING SIMULATION AND INTERACTIONS

UNIVERSITÉ CÔTE D'AZUR
MEDICAL DATA LABORATORY

Who are you?

- Curious, open minded
- You know how to use a PC
- Wanting to learn some cool stuff
- Not feared of tackling problems
- Not feared by errors
- (optional) Some coding experience
- (bonus) “LMGTFY” skills (and not eager to…)

- Ideal profile:
  - 30% data scientist: exploit the data
  - 70% software developer: you need to code

Course Overview

- Deep dive on Python3
  - 3.8 onwards, maybe up to 3.11
  - builtins
  - error handling
  - OO vs FP
  - concurrency
  - testing
- Data Manipulation
  - cleaning
  - preprocessing
- Basic data analysis
- Advanced data analysis
  - numpy, pandas
- Database
  - sqlalchemy

Course Overview

- Coding / Coding / Coding
- Brace yourself
- REPL + scripts + notebooks
- 10 lessons
  - ~ 20% lectures
  - ~ 80% live sessions
- Evaluation
  - Software project (3 months)
  - Written test (1-2 hour(s))
  - Oral discussion (30 min)

Contacts

https://github.com/marcomilanesio/python-course

marco.milanesio@univ-cotedazur.fr

Evaluation

Software project

Build a utility library
for your data science journey

Notes

A. Growing thing
B. There is no "correct" answer: it has to serve you
C. Testing

Written test

on the last day
1 or 2 hours
no computer
train your muscle memory

Oral discussion

after the course is finished
single interview
discuss on the project / test / course topics