

Python programming for data science

marco milanesio
MScDSAI

Who am I?

- 2010 - PhD in Turin (Italy)
 - Layering Multi-Purpose Applications over Structured and Dependable P2P Systems
- 2011 - 2017: Research Engineer @Eurecom
 - Distributed Storage
 - Network performances
- 2017 - now: Research Engineer @UniCA MSI
 - Distributed learning (@Inria)
 - MDLAB (@CHU Nice)
 - Corals (@IRCAN)
 - Shells (@CEPAM)
 - Knights of the Malta's Order (@CMMC)
 - Teaching (@Datahyking, @MscDSAI)
 - DATAZUR (@UniCA)
 - EEG (@LAPCOS)



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
-
- 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
 - mid term
 - software project
 - written test
 - (maybe) oral discussion

Contacts

`marcomilanesio.github.io`

`marco.milanesio@univ-cotedazur.fr`

Evaluation

Mid-term test

2 hours

a set of functions

(almost) Fully Automated

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
- D. Two files only

Written test

on the last day

1 or 2 hours

no computer

train your muscle memory

Oral discussion

subject to results in mid-term / written test

after the course is finished

single interview

discuss on the project / test / course topics

Questions

`marcomilanesio.github.io`

`marco.milanesio@univ-cotedazur.fr`