

# **Introduction to Programming and Data Processing Module 1**

## **Lecture 1: Course Introduction**

# Outline

- ① Course introduction
- ② Sneak preview of Module 2
- ③ Overview to programming

## Note on the 2-modules structure

**2-modules structure:** [http://bit.ly/IPDPSSSA20\\_21](http://bit.ly/IPDPSSSA20_21)

As you know, this course is the first module of a teaching unit of two modules. Intuitively

- M1: Module 1 focuses on programming
- M2: Module 2 focuses on data analysis and machine learning

Students can attend single modules.

M1 gives the necessary background for M2

**These slides focus on M1**

## Note on the 2-modules structure

**2-modules structure:** [http://bit.ly/IPDPSSSA20\\_21](http://bit.ly/IPDPSSSA20_21)

As you know, this course is the first module of a teaching unit of two modules. Intuitively

- M1: Module 1 focuses on programming
- M2: Module 2 focuses on data analysis and machine learning

Students can attend single modules.

M1 gives the necessary background for M2

**These slides focus on M1**

**Previous edition, A.Y. 2019/2020**

M1 is a re-edition of course **Introduction to Programming in Python** held by us last year for the Allievi Ordinari of SSSA

- If you have attended it, and plan to attend M2, you can skip to it

# Course Responsible

- Course responsible: Andrea Vandin
  - ★ [andrea.vandin@santannapisa.it](mailto:andrea.vandin@santannapisa.it)
  - ★ Tenure-track Assistant Professor in Computer Science at Institute of Economics & EMbeDS @ SSSA
  - ★ Formerly:
    - ▶ Associate Professor in Computer Science at DTU Technical University of Denmark
      - Most related teaching activity: responsible for 3 years of course 'Programming in C++ for non-computer scientists', 250 students
- Co-lecturer: Daniele Licari
  - ★ [daniele.licari@santannapisa.it](mailto:daniele.licari@santannapisa.it)
  - ★ EMbeDS Data Engineer
  - ★ Great expert of Python

# Course References & Material

- Webpages of the course:
  - ★ [http://bit.ly/IPDPSSSA20\\_21](http://bit.ly/IPDPSSSA20_21)
    - ▶ Slides and examples from the lectures, further materials and links
  - ★ [http://bit.ly/IPDPSSSA20\\_21Repl](http://bit.ly/IPDPSSSA20_21Repl)
    - ▶ Weekly coding assignments
- Suggested books:
  - ★ M. Lutz, Learning Python;
  - ★ W. McKinney, Python for Data Analysis.
- Well-done tutorial: <https://docs.python.org/3/tutorial/>
- Software
  - ★ Python: <https://www.python.org/>
  - ★ Python editor: JupyterLab <https://jupyter.org/>
  - ★ Setup your machine: [http://bit.ly/IPDPSSSA20\\_21](http://bit.ly/IPDPSSSA20_21)

# Tentative Course Description

## This course will

introduce the students to the fundamental principles of (object-oriented) structured programming with basic applications to data processing. Using Python as reference language, the course starts from basic notions of programming (variables, data types, collections, control & repetition structures, functions & modules), up to basic data processing functionalities (loading, manipulation, and visualization of CSV data).

# Tentative Course Description

## **This course will**

introduce the students to the fundamental principles of (object-oriented) structured programming with basic applications to data processing. Using Python as reference language, the course starts from basic notions of programming (variables, data types, collections, control & repetition structures, functions & modules), up to basic data processing functionalities (loading, manipulation, and visualization of CSV data).

## **A student who has met the objectives of the course will**

acquire an understanding of the issues involved in computer programming, to be able to make informed decisions. The student will be able to write simple to medium python programs of various nature, including those for reading, manipulating and visualizing data.



# Tentative Learning Objectives

A student who has met the objectives of the course will be able to:

- select and use the correct data types and collections for the problem at hand
- use variables and operations
- use and describe control and repetition structures (if, loops)
- create and use functions and classes
- use libraries for File I/O, data manipulation, and data visualization
- use principles of structured program design and methods
- explain and apply the principles of abstract data types
- discuss Python-related issues in a clear and concise way, possibly using on-line platforms

# Evaluation

- Weekly coding assignments
  - ★ To be handed in via *Repl.it* at [http://bit.ly/IPDPSSSA20\\_21Repl](http://bit.ly/IPDPSSSA20_21Repl)
    - ▶ Automatic tests for your code and hints to fix bugs
    - ▶ (Soft) deadlines: before the following class
    - ▶ Feel free to contact us for support
  - ★ A fundamental learning tool of this course
- Oral Exam
  - ★ We will do an oral examination
    - ▶ starting from your solutions to the assignments
    - ▶ Another reason for doing your assignments!
  - ★ Date: TBD

# Tentative Lecture Plan

#	Date	Time	Topic
1	22/02	17:00-19:00	Course introduction
2	24/02	15:00-18:00	Data types & operations
3	26/02	15:00-18:00	Collections
4	01/03	15:00-18:00	Control and Repetition statements
5	05/03	15:00-18:00	Functions
6	08/03	15:00-18:00	Modules & Exceptions & Object Oriented Programming
7	12/03	15:00-18:00	Advanced libraries for data manipulation/visualization
-	TBD	TBD	Exam

## Further info

- No previous experience on computer programming required
- Previous experience in writing small programs is advantageous
- We might adjust the course level according to your expertise and feedback

## Further info

- No previous experience on computer programming required
- Previous experience in writing small programs is advantageous
- We might adjust the course level according to your expertise and feedback
- You will never learn programming if you don't practice it!
  - ★ Therefore you have to regularly do all the assignments

# Ideas for an Effective Course

## Live Programming & Assignments

We have blocks of 3 hours.

- First part:

Intro to week's topics & Live programming

- ★ Not many slides

- ★ Instead: we develop a few example programs

- ▶ Please have your laptop ready! [http://bit.ly/IPDPSSSA20\\_21](http://bit.ly/IPDPSSSA20_21)

- ▶ You find code in advance here

- Second part:

You consolidate your understanding working on the assignments

- ★ Begin working on the assignments with our live support if needed

- ★ Complete them offline before next class. Contact us if needed

# Ideas for an Effective Course

## Live Programming & Assignments

We have blocks of 3 hours.

- First part:

Intro to week's topics & Live programming

- ★ Not many slides

- ★ Instead: we develop a few example programs

- ▶ Please have your laptop ready! [http://bit.ly/IPDPSSSA20\\_21](http://bit.ly/IPDPSSSA20_21)

- ▶ You find code in advance here

- Second part:

You consolidate your understanding working on the assignments

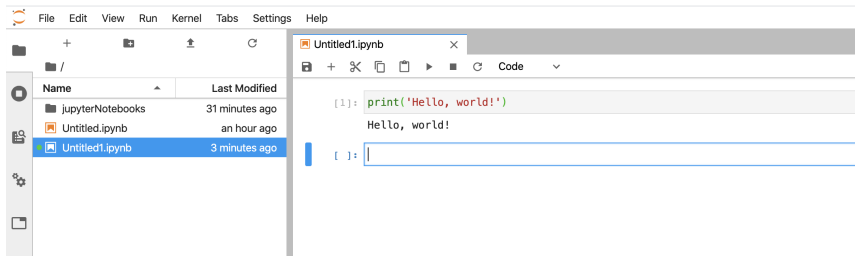
- ★ Begin working on the assignments with our live support if needed

- ★ Complete them offline before next class. Contact us if needed

However, we have the ambitious goal of covering many topics necessary to introduce you to programming and data processing in just 20 hours. Hence we might skip some second parts.

# Live Programming

Find the JupyterLab notebooks at [http://bit.ly/IPDPSSSA20\\_21](http://bit.ly/IPDPSSSA20_21)





# Repl.it

The screenshot shows the Repl.it web interface. At the top, the Repl.it logo is on the left, and the title 'Module01.01 - Hello, world!' is in the center. To the right of the title are links for 'my repls', 'student', 'notifications', 'gg talk', and 'arwandu (0)'. Below the title bar, there is a 'back to classroom' link, a 'run' button with a play icon, and a dropdown menu. The main area is a code editor with the following Python code:

```
1- # This program prints 'Hello, world!':  
2- print('Hello, world!')  
3-  
4- # Can you change it so that it prints the same, but in Italian?
```

At the bottom of the code editor, it says 'Python 3.8.1 (default, Feb 2 2020, 08:37:37)'. To the right of the code editor, there is a sidebar with the following content:

- Due:** Apr 20, 2020 02:59 pm
- submit** button
- Instructions from your teacher:**
- Statement**  
Write a program that prints the Italian translation of 'Hello, world!'. That is: 'Ciao, mondo!'
- Example input**
- Example output**
- Theory**  
Well, there is not much theory to use here.  
In the example, you can see how to print 'Hello, world!'.

- First time visit: [http://bit.ly/IPDPSSSA20\\_21ReplFirst](http://bit.ly/IPDPSSSA20_21ReplFirst)
- After that: [http://bit.ly/IPDPSSSA20\\_21Repl](http://bit.ly/IPDPSSSA20_21Repl)

# Repl.it

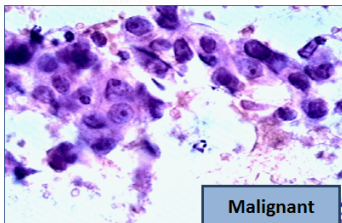
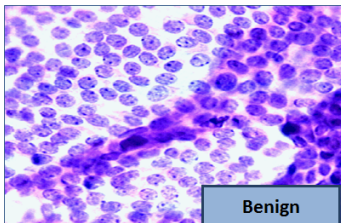
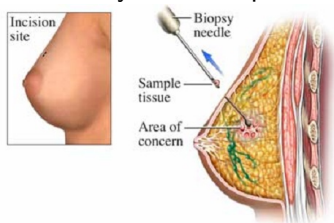
- A Repl.it team is a collection of assignments with autograding functionalities
- Your dashboard will be an ordered list of assignments.  
You can see the status of your assignments
- Our dashboard tells us the status of all your assignments

# Outline

- 1 Course introduction
- 2 Sneak preview of Module 2**
- 3 Overview to programming

## Sneak preview of Module 2

Starting from the competences developed in the first module, we will study how to apply data analysis techniques from Machine learning



Can we classify them automatically?

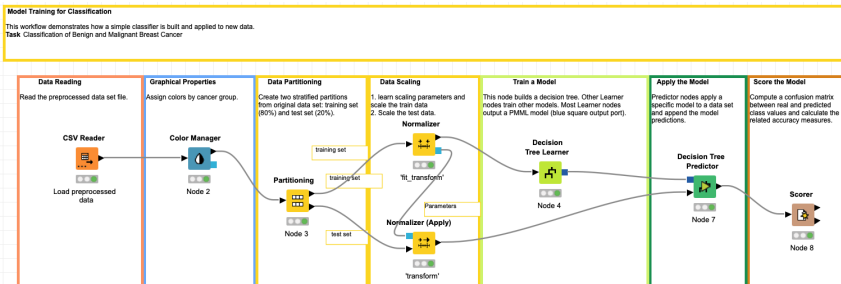
## Sneak preview of Module 2

We will go through a classic pipeline for these data analysis tasks

- with emphasis on data pre-processing.

We will use two alternative approaches

- Python: **main focus**
- Knime: a graphical workflow language



## Sneak preview of Module 2

### Evaluation

You will do the same on data of interest or on data on titanic sinking

- Would you have survived the sinking of the titanic?

# Outline

- ① Course introduction
- ② Sneak preview of Module 2
- ③ Overview to programming**

# What is a program?

- A sequence of code instructions to control a machine
  - ★ Input/output
  - ★ Mathematical operations
  - ★ Conditional and repetitive executions
- A recipe to instruct a machine to execute instructions.
  - ★ We can't use a *natural language*.
  - ★ We need a **programming language**



# Programming languages

## The 7 Most In-Demand Programming Languages of 2019

March 15, 2019

Aspiring developers need to know what languages to learn; they need to select the right education and work on a skill set that will impress future employers and land their dream job. So what are the top programming languages? And what is the best one to learn? We've compiled a list for you that highlights the most in-demand programming languages based off current job postings on the market.

Here are the Top 7 programming languages with most job posting on [Indeed](https://www.indeed.com) as of January 2019:

- Java – 65,986 jobs
- Python – 61,818 jobs
- Javascript – 38,018 jobs
- C++ – 36,798 jobs
- C# – 27,521 jobs
- PHP – 16,890 jobs
- PERL – 13, 727 jobs

1	Java		11	MATLAB	
2	C		12	R	
3	Python		13	Perl	
4	C++		14	Assembly Language	
5	Visual Basic .NET		15	Swift	
6	Javascript		16	Go	
7	C#		17	Delphi/Object Pascal	
8	PHP		18	Ruby	
9	SQL		19	PL/SQL	
10	Objective-C		20	Visual Basic	

<http://www.codingdojo.com/blog/the-7-most-in-demand-programming-languages-of-2019>

# Programming languages

The index can be used to check whether your programming skills are still up to date or to make a strategic decision about what programming language should be adopted when starting to build a new software system. The definition of the TIOBE index can be found [here](https://www.tiobe.com/tiobe-index/).

Feb 2021	Feb 2020	Change	Programming Language	Ratings	Change
1	2	▲	C	16.34%	-0.43%
2	1	▼	Java	11.29%	-6.07%
3	3		Python	10.86%	+1.52%
4	4		C++	6.88%	+0.71%
5	5		C#	4.44%	-1.48%

<https://www.tiobe.com/tiobe-index/>

# The Python Programming language



- High-level: almost human readable. Abstracts from hardware
- Beginner-friendly: streamlined syntax:
  - ★ it is easy to write your *first programs*
- Free, open-source and multi-platform
- Developed since the 90s, therefore it has
  - ★ A wide community, and its popularity keeps increasing
  - ★ Many predefined software modules

# Python programs

- A sequence of python instructions to control a machine
- Python supports the most common programming styles
  - ★ Imperative: Statements are executed in sequence changing the state of the program (the variables)
  - ★ Procedural: The program is structured in reusable units named functions
  - ★ Object-oriented: The program is structured as a collection of interacting objects that send messages to each other.
  - ★ Functional: Statements are not written/executed as an ordered sequence of instructions. A computation is treated as the evaluation of a mathematical function.

# Variables

## Basic abstraction to represent units of data

A variable has a name and a value

- Names can contain any letter, number, or the underscore \_

False	await	else	import	pass
None	break	except	in	raise
True	class	finally	is	return
and	continue	for	lambda	try
as	def	from	nonlocal	while
assert	del	global	not	with
async	elif	if	or	yield

Note:

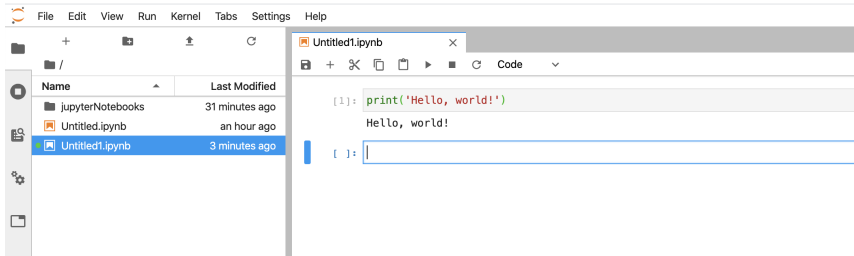
- ★ Cannot start with numbers
- ★ Cannot be a keyword
- ★ Names are case-sensitive

- We assign/update values to variables using assignment statements

```
month_number=3
month_name="April"
print("The number of",month_name,"is",month_number)
month_number=4
print("The number of",month_name,"is",month_number)
```

# Live Programming

Find the JupyterLab notebooks at [http://bit.ly/IPDPSSSA20\\_21](http://bit.ly/IPDPSSSA20_21)



# Configure your machine

If you have not done it yet

Follow the instructions in  
[http://bit.ly/IPDPSSSA20\\_21](http://bit.ly/IPDPSSSA20_21)

## “But it works . . .”





# “Can You Learn To Ski Without Lessons?”



<https://www.skibro.com/blog/en/can-you-learn-to-ski-without-lessons/>

Most of the times you get to the valley.  
**The problem is how you get there . . .**