

PDAI

Programming & Data Analytics & AI

Module 1

Lecture 1: Course Introduction

Outline

- 1 Course introduction**
- 2 Sneak preview of Module 2**
- 3 Let's Kahoot!**
- 4 Overview to programming**

Note on the 3-modules structure

3-modules structure:

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

- M1: Module 1 focuses on programming 'well'
- M2-ML: Module 2-ML focuses on data analysis and ML
- M2-PM: Module 2-PM focuses on process-oriented data science

Students can attend single modules. M1 is a requirement for the M2s

These slides focus on M1

Note on the 3-modules structure

3-modules structure:

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

- M1: Module 1 focuses on programming 'well'
- M2-ML: Module 2-ML focuses on data analysis and ML
- M2-PM: Module 2-PM focuses on process-oriented data science

Students can attend single modules. M1 is a requirement for the M2s

These slides focus on M1

Previous editions: A.Y. 2019/2020

Introduction to Programming in Python held by us for the Allievi Ordinari of SSSA

- Was a preliminary version of current M1
- Planned and run online due to COVID :(

Note on the 3-modules structure

3-modules structure:

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

- M1: Module 1 focuses on programming 'well'
- M2-ML: Module 2-ML focuses on data analysis and ML
- M2-PM: Module 2-PM focuses on process-oriented data science

Students can attend single modules. M1 is a requirement for the M2s

These slides focus on M1

Previous editions: A.Y. 2020/2021

Intro to Programming & Data Processing 1

Two modules

- M1 was a preliminary version of the current M1
- M2 was a 10-hours version of the current M2-ML

Note on the 3-modules structure

3-modules structure:

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

- M1: Module 1 focuses on programming 'well'
- M2-ML: Module 2-ML focuses on data analysis and ML
- M2-PM: Module 2-PM focuses on process-oriented data science

Students can attend single modules. M1 is a requirement for the M2s

These slides focus on M1

Previous editions: A.Y. 2021/2022

Programming & Data Analytics

Two modules

- M1 was a preliminary version of the current M1
- M2 was a preliminary version of the current M2-ML

Note on the 3-modules structure

3-modules structure:

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

- M1: Module 1 focuses on programming 'well'
- M2-ML: Module 2-ML focuses on data analysis and ML
- M2-PM: Module 2-PM focuses on process-oriented data science

Students can attend single modules. M1 is a requirement for the M2s

These slides focus on M1

Previous editions: A.Y. 2022/2023

Programming & Data Analytics & AI

Two modules

- The course reached maturity
- However, we keep on refining and updating it yearly

Note on the 3-modules structure

3-modules structure:

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

- M1: Module 1 focuses on programming 'well'
- M2-ML: Module 2-ML focuses on data analysis and ML
- M2-PM: Module 2-PM focuses on process-oriented data science

Students can attend single modules. M1 is a requirement for the M2s

These slides focus on M1

Previous editions: A.Y. 2023/2024

Programming & Data Analytics & AI

Two modules

- The course reached maturity
- However, we keep on refining and updating it yearly

Note on the 3-modules structure

3-modules structure:

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

- M1: Module 1 focuses on programming 'well'
- M2-ML: Module 2-ML focuses on data analysis and ML
- M2-PM: Module 2-PM focuses on process-oriented data science

Students can attend single modules. M1 is a requirement for the M2s

These slides focus on M1

This edition: A.Y. 2024/2025

Programming & Data Analytics & AI

Three modules

- The course reached maturity
- We added a new M2 (M2-PM). It alternates yearly with the 'classic' M2 (M2-ML).

Course Responsible

- Course responsible: Andrea Vandin
 - ★ andrea.vandin@santannapisa.it
 - ★ Associate Professor in Computer Science at Institute of Economics & L'EMbeDS @ SSSA,
Adj. Associate Professor at DTU Technical University of Denmark
 - ★ Former Associate Professor in Computer Science at DTU
 - ▶ *Programming in C++ for non-computer scientists*, 250 students
 - ★ Prof. Daniele Giachini will give a class on the use of object-oriented programming in economic ABMs
- Teaching Assistant: Sima Sarv Ahrabi
 - ★ Sima.SarvAhrabi@santannapisa.it
 - ★ L'EMbeDS Data Scientist (Python, ML, process mining ...)
- Co-designer of previous editions: Daniele Licari
 - ★ Now at Banca d'Italia. Previously: EMbeDS Chief Data Engineer

Related courses

We also teach related courses

- PhD Students in Computer Science of Gran Sasso Science Institute, L'Aquila [2020-NOW]
- Master Students in Economics, University of Pisa & SSSA, Pisa [2023-NOW]
- Online certification program Python4Finance, Politecnico Di Milano Graduate School of Management, Milan [2024-NOW]
- Master Students in Finance Politecnico Di Milano Graduate School of Business, Milan [2023]
- PhD students of Scuola Superiore Normale, Pisa [2020-2022]

Course References & Material

- Webpage of the course:
 - ★ https://bit.ly/PDAI_24_25
 - ▶ Slides and examples from the lectures, further materials and links
 - ▶ 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/>
 - ★ Setup your machine: https://bit.ly/PDAI_24_25
 - ▶ Python editors: JupyterLab, Colab, Visual Studio Code

Course Description - M1

This module will

introduce students to the fundamental principles of structured programming, with applications to data processing and analysis.

- It starts from basic notions of programming (data types, collections, control structures, functions & modules, OOP),
- Progresses to data processing functionalities (loading, manipulation, and visualization of CSV data).

It will teach you how to **program well**

Course Description - M1

This module will

introduce students to the fundamental principles of structured programming, with applications to data processing and analysis.

- It starts from basic notions of programming (data types, collections, control structures, functions & modules, OOP),
- Progresses to data processing functionalities (loading, manipulation, and visualization of CSV data).

It will teach you how to **program well**

A student who has met the objectives of the course will get

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 good quality and of various nature, including those for reading, manipulating and visualizing data.

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 and describe variables, operations, and control structures (if, loops)
- create and use functions and classes
- use libraries for I/O, data manipulation, and data visualization
- use principles of structured program design and methods
- discuss Python-related issues in a clear and concise way, possibly using on-line platforms

Evaluation

- Regular coding assignments
 - ★ Available at https://bit.ly/PDAI_24_25
 - ▶ Every class comes with a set of related assignments
 - ▶ We have built an online framework for automatically testing your code and getting hints
 - ▶ (Soft) deadlines: before the following class
 - ▶ We will try to allocate time at the end of classes to work on them
 - ▶ You will have to send to us your solutions before the exam
 - ★ 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

You find it on the website
https://bit.ly/PDAI_24_25

Computing, Data Analysis & Modeling @ L'EMbeDS

This website collects materials from courses on the **(compressed and indexed) storage, search, analysis, and modeling of Big Data** offered at the Sant'Anna School of Advanced Studies by [L'EMbeDS](#) members for the academic year 2024/2025

These courses provide skills and notions critical for the research promoted by [L'EMbeDS](#).

Target audience:

- The principal target audience for these courses are the honor students (*allievi ordinari*) of the Sant'Anna School. However, the courses are open to all interested students – including BS and MSc level students from other academic institutions in the Pisa area.
- A parallel and crucial target is students from doctoral programs (internal or external to the School) who may benefit from any or all of the content being offered – depending on their background and research interest. Examples of internal doctoral programs are the [PhD in Economics](#) and the [PhD in Health Science, Technology and Management](#). An example of the external doctoral program is the [Italian National PhD in AI for Society](#)

For further information on single courses, please follow the links in the right-sidebar

CSIR

- [Slides, code and other material](#)

DMPD

- [Slides, code and other material](#)

ISE

- [Slides, code and other material](#)

PDAI

- [Slides, code and other material](#)
- [Setup your machine](#)
- [Extra coding exercises](#)

ASM

- [Slides, code and other material](#)

SLLD

- [Slides, code and other material](#)

Further info

- No previous experience on computer programming required
- Previous experience in writing small programs is advantageous

Further info

- No previous experience on computer programming required
- Previous experience in writing small programs is advantageous
- 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 often have blocks of 3 hours.

- First part:

Intro to new topics & Live programming

- ★ No slides
- ★ Interactive *notebooks* mixing presentation material and code
 - ▶ Please have your laptop ready! https://bit.ly/PDAI_24_25
 - ▶ You find code in advance here

- Second part:

You consolidate your understanding working on the assignments

- ★ Begin working on the assignments with our support if needed
- ★ Complete them offline before next class. Contact us if needed

Ideas for an Effective Course

Live Programming & Assignments

We often have blocks of 3 hours.

- First part:

Intro to new topics & Live programming

- ★ No slides
- ★ Interactive *notebooks* mixing presentation material and code
 - ▶ Please have your laptop ready! https://bit.ly/PDAI_24_25
 - ▶ You find code in advance here

- Second part:

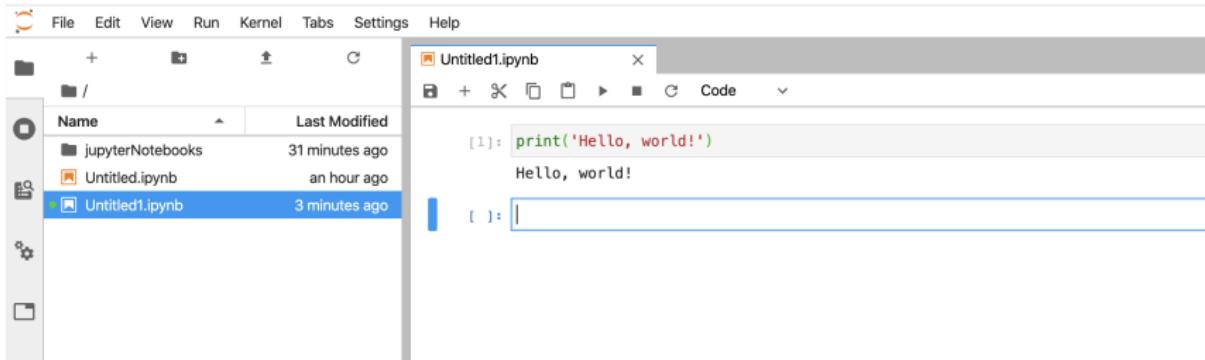
You consolidate your understanding working on the assignments

- ★ Begin working on the assignments with our 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 analytics in just 20 hours. Hence we might skip some second parts.

Live Programming

Find the JupyterLab notebooks at https://bit.ly/PDAI_24_25



Assignments on Colab (or locally in your machine)

01ConsoleIOandVariables_Assignments

File Edit View Insert Runtime Tools Help

Table of contents

RUN, BUT DO NOT MODIFY

Assignment 01.01: Sum of three numbers

Write your solution here

Run the following cells to perform the provided tests

TEST 1

Assignment 01.02: Hello, name1 name2

Write your solution here

Run the following cells to perform the provided tests

TEST Renzo Lucia

TEST Andrea Daniele

TEST Daniele Andrea

Assignment 01.03: Hello, name1 and name2

Write your solution here

Run the following cells to perform the provided tests

TEST Renzo Lucia

TEST Andrea Daniele

Assignment 01.01: Sum of three numbers

Statement

Write a program that prints the Italian translation of 'Hello, world!'.

That is: 'Ciao, mondo!'

Example input

Example output

Ciao, mondo!

Theory

Well, there is not much theory to use here.

In the example, you can see how to print 'Hello, world!'.

Write your solution here

- Each lecture comes with a set of simple coding assignments
 - ★ Links available in the wiki page for slides and further material

Colab

- Colab is a Google service similar to Google docs
 - ★ but for python notebooks.
 - ★ no installation required
- Each set of assignments is actually a python notebook
- We implemented in Colab autograding functionalities
 - ★ You can test your own solution
 - ★ Immediate feedback!
- If you prefer, you can also download them as jupyter notebooks,
 - ★ like the ones we use in class

Colab: auto-testing

▼ Write your solution here

- Do not change the first line (`def ...():`)
- Maintain the given indentation
- You can run some tests by yourself by decommenting the last line

```
[3] def asgn01_01Hello_world():
    # This program prints 'Hello, world!':
    #print('Hello, world!')

    # Can you change it so that it prints the same,
    #but in Italian?

    print('Ciao, mondo')

#You can test independently your solution by executing the following line
#asgn01_01Hello_world()
```

▼ Run the following cells to perform the provided tests



TEST 1

```
Test []
The program prints 1 lines as expected.

Line 0
Test FAILED
    Expected: Ciao, mondo!
    Actual   : Ciao, mondo
```

Colab: auto-testing

▼ Write your solution here

- Do not change the first line (`def ...():`)
- Maintain the given indentation
- You can run some tests by yourself by decommenting the last line

```
[ ] def asgn01_01Hello_world():
    # This program prints 'Hello, world!':
    #print('Hello, world!')

    # Can you change it so that it prints the same,
    #but in Italian?

    print('Ciao, mondo!')

    #You can test independently your solution by executing the following line
    #asgn01_01Hello_world()
```

▼ Run the following cells to perform the provided tests



TEST 1

```
Test []
The program prints 1 lines as expected.
Line 0
Expected and actual output match: Ciao, mondo!
Test PASSED!
```

With all of this . . .

. . . we hope to convince you on the importance of programming, and the importance of doing it well

“But it works . . .”



“Can You Learn To Ski Without Lessons?”



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

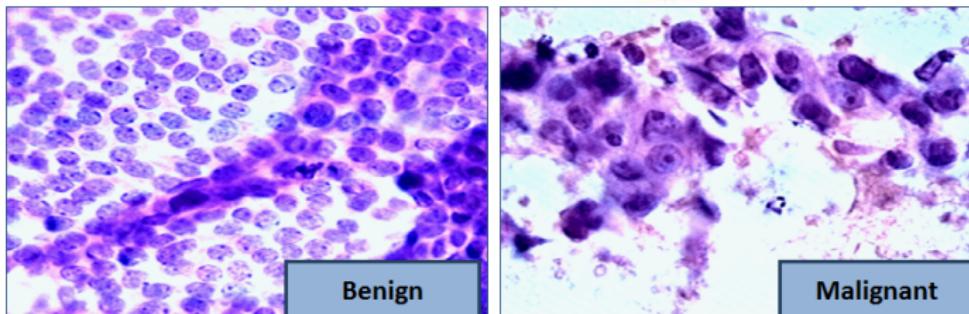
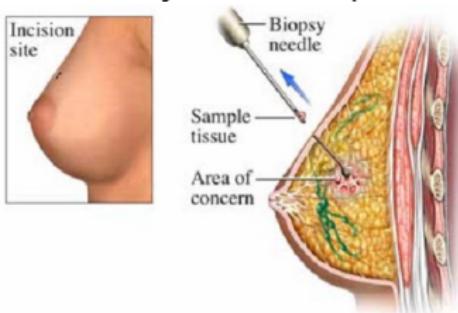
Most of the times you get to the valley.
The problem is how you get there . . .

Outline

- ① Course introduction
- ② **Sneak preview of Module 2**
- ③ Let's Kahoot!
- ④ Overview to programming

Sneak preview of Module 2 ML - next A.Y.

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 ML - next A.Y.

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

- With emphasis on how to do *good* data analysis projects
- We explain how to avoid common pitfalls of *bad* analysis projects

Sneak preview of Module 2 PM - THIS A.Y.

M2-PM (Process mining)

Advanced research-oriented data science topic
Use data to study the data generating process

- Inefficiencies or problems in the process?

1. *Lucy* takes your order

2. *Lucy* notes down your address

3. *Lucy* notes down your preferred payment method

4. *Luigi* prepares your burger

5. *Lucy* grabs your can of soda

6. *Luigi* puts your burger in a box

7. *Lucy* wraps your order

8. *Mike* delivers your order

1. *Randy* takes your order

2. *Randy* notes down your preferred payment method

3. *Randy* notes down your address

4. *Luigi* prepares your burger

5. *Luigi* puts your burger in a box

6. *Randy* wraps your order

7. *John* delivers your order

What is the process underlying my data?¹



¹Example from <https://pm4py.fit.fraunhofer.de/>

Sneak preview of Module 2 PM - THIS A.Y.

M2-PM (Process mining)

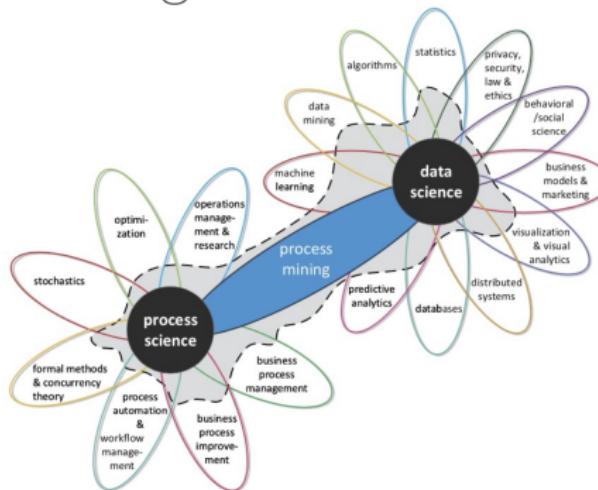
This is a hot topic. With interest in academia and in industry.
Interdisciplinary research.

What is Process Mining?

Process mining is a family of techniques linking data science and process management to support the analysis of operational processes based on event logs.

The goal of process mining is to turn event data into insights and actions.
Process mining is an integral part of data science, fueled by the availability of data and the desire to improve processes.

E.g., process mining uses event data to automatically discover a process model by observing events recorded by some enterprise system



Sneak preview of Module 2 PM - THIS A.Y.

M2-PM (Process mining)

This is a hot topic. With interest in academia and in industry.
Interdisciplinary research.

What can we do with Process Mining?



With PM we can
find the actual
process is
different from
the one we
designed!

Picture by Koen Olsthoorn

Our traces are the process logs...

Sneak preview of Module 2 PM - THIS A.Y.

M2-PM (Process mining)

This is a hot topic. With interest in academia and in industry.
Interdisciplinary research.

Challenges in Legal Process Discovery*

Hugo A. López¹

Department of Computer Science
University of Copenhagen**

Abstract. One of the main promises of process conformance is the opportunity to align normative processes (i.e. how the process *should* behave) and event logs (i.e. how does the process *actually* behaves). Results of conformance checking are valid as long as normative processes correspond to actual norms. Recent developments advocate the use of Natural Language Processing (NLP) to process model discovery from texts. We present a series of challenges in textual process discovery that limit its applicability to real norms. The challenges emerges from experiences with legal practitioners in the digitalization of administrative processes in Danish and Italian municipalities, and they need to be solved in order to provide accurate normative processes that reflect the intent of laws.

Key words: Natural Language Processing, Normative Processes, Process Discovery, Process Conformance

Business Process Compliance using Reference Models of Law

Hugo A. López^{1,3}, Søren Debois², Tjits Slaats¹, and Thomas T. Hildebrandt¹

¹ Software, Data, People & Society Section
Department of Computer Science
Copenhagen University, Denmark
[\(hala,slaats,hilde\)@di.ku.dk](mailto:(hala,slaats,hilde)@di.ku.dk)

² Computer Science Department, IT University of Copenhagen, Denmark
debois@itu.dk

³ DCR Solutions A/S, Denmark

Abstract. Legal compliance is an important part of certifying the correct behaviour of a business process. To be compliant, organizations might hard-wire regulations into processes, limiting the discretion that workers have when choosing what activities should be executed in a case. Worse, hard-wired compliant processes are difficult to change when laws change, and this occurs very often. This paper proposes a model-driven approach to process compliance and combines a) reference models from laws, and b) business process models. Both reference and process models

The process of a law?

Does my process comply with law?

Sneak preview of Module 2 PM - THIS A.Y.

M2-PM (Process mining)

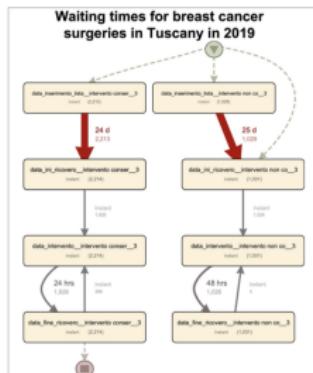
This is a hot topic. With interest in academia and in industry.
Interdisciplinary research.

Process Mining and Clinical Pathways: an application to Breast cancer data in Tuscany

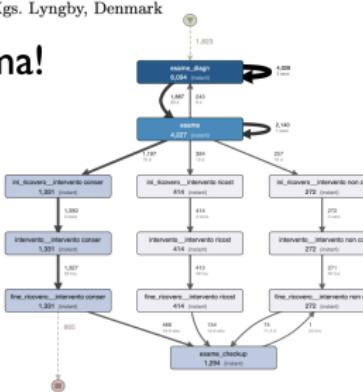
Francesca Ferrè¹, Chiara Seghieri¹, Andrea Burattin², and Andrea Vandin^{1,2}

¹ Sant'Anna School of Advanced Studies, Pisa, Italy

² DTU Technical University of Denmark, Kgs. Lyngby, Denmark



Now also Sima!



Sneak preview of Module 2 PM - THIS A.Y.

M2-PM (Process mining)

This is a hot topic. With interest in academia and in industry.
Interdisciplinary research.

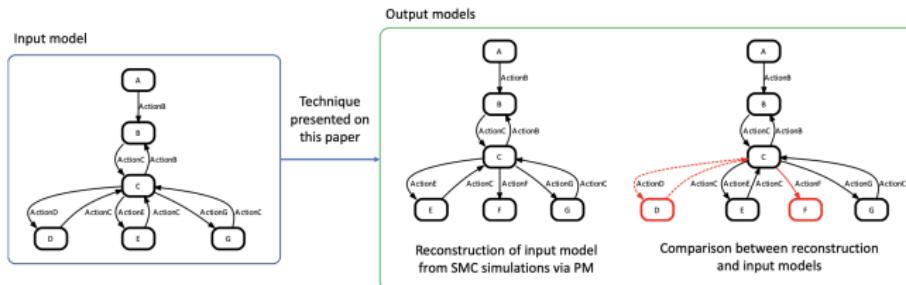
White-box validation of quantitative dynamic product lines by statistical model checking and process mining

Roberto Casaluce^a, Andrea Burattin^b, Francesca Chiaromonte^{a,c}, Alberto Lluch Lafuente^b, Andrea Vandin^{a,b,*}

^aInstitute of Economics and EMbeDS, Sant'Anna School of Advanced Studies, Pisa, Italy.

^bDTU Technical University of Denmark, Lyngby, Denmark.

^cDept. of Statistics and Huck Institutes of the Life Sciences, Penn State University, USA



Sneak preview of Module 2 PM - THIS A.Y.

M2-PM (Process mining)

This is a hot topic. With interest in academia and in industry.
Interdisciplinary research.



ICPM 2024 General Info Conference Code of Conduct Sponsors

Registration Program Awards Pictures

Welcome to Denmark!



Sneak preview of Module 2 PM - THIS A.Y.

M2-PM (Process mining)

This is a hot topic. With interest in academia and in industry.
Interdisciplinary research.

Diamond Sponsor



Gold Sponsors



Platinum Sponsors



MEHRWERK

Sponsors



Startup Sponsors



Sneak preview of Module 2

Evaluation

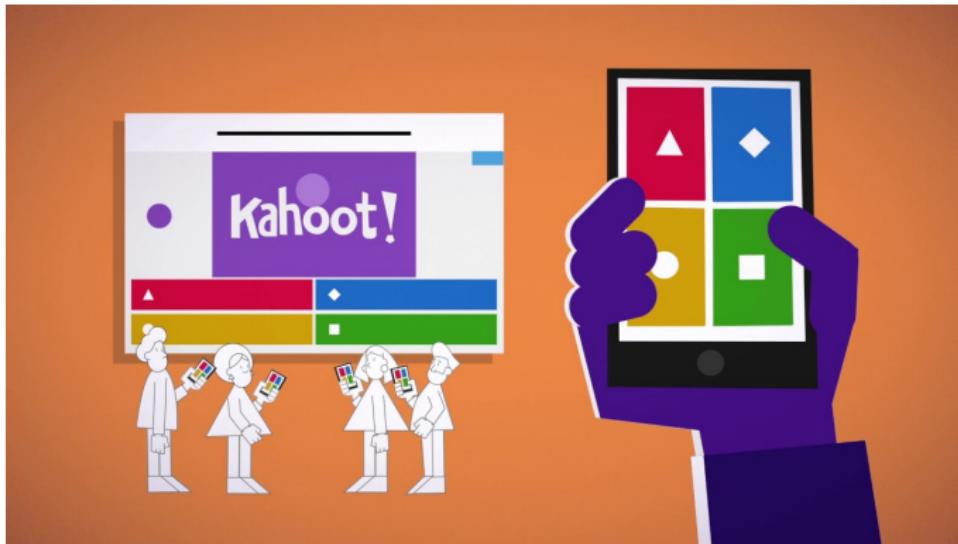
You will do a data-driven project on these topics.

- We will provide some datasets
- You can bring your own!

Outline

- ① Course introduction
- ② Sneak preview of Module 2
- ③ Let's Kahoot!
- ④ Overview to programming

Let's play a game on Kahoot!



- Using your smartphone or a second monitor
- Visit www.kahoot.it
- Type the code we will give you during the class

Outline

- ① Course introduction
- ② Sneak preview of Module 2
- ③ Let's Kahoot!
- ④ 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](#) 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

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

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](#).

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%

Programming languages

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

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](#).

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%

May 2021	May 2020	Change	Programming Language	Ratings	Change
1	1		C	13.38%	-3.68%
2	3	▲	Python	11.87%	+2.75%
3	2	▼	Java	11.74%	-4.54%
4	4		C++	7.81%	+1.69%
5	5		C#	4.41%	+0.12%

Programming languages

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

Jan 2023	Jan 2022	Change	Programming Language	Ratings	Change
1	1		 Python	16.36%	+2.78%
2	2		 C	16.26%	+3.82%
3	4		 C++	12.91%	+4.62%
4	3		 Java	12.21%	+1.55%
5	5		 C#	5.73%	+0.05%
6	6		 Visual Basic	4.64%	-0.10%
7	7		 JavaScript	2.87%	+0.78%
8	9		 SQL	2.50%	+0.70%
9	8		 Assembly language	1.60%	-0.25%
10	11		 PHP	1.39%	-0.00%
11	10		 Swift	1.20%	-0.21%

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:

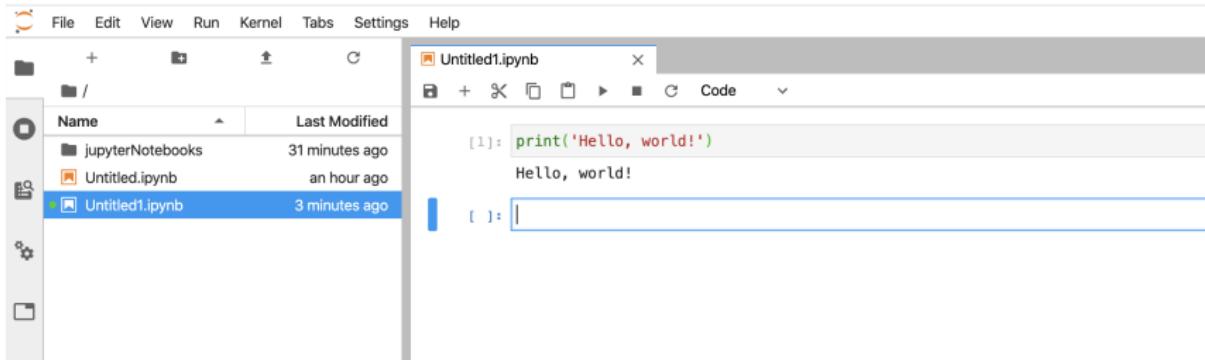
- ★ Cannot start with numbers
- ★ Cannot be a keyword
- ★ Names are case-sensitive

- 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 https://bit.ly/PDAI_24_25



Configure your machine

If you have not done it yet

Follow the instructions in
https://bit.ly/PDAI_24_25