

Machine Learning and Programming in Python

Lecture for Master and PhD students

Chair of Data Science in Economics

Ruhr University Bochum

Summer semester 2024

Lecture 2

Outline

- How to download and install Python.
- What is Jupyter about?
- The main concepts of Python

- We use Python because it is easy to learn, friendly and we have thousand of online resources in case we need help.
- The syntax is simpler than in C++, Java, JavaScript...
- We are going to start from the beginner level, to progressively increase the complexity of the module.

- Code quality.
Highly readable code, making it reusable and maintainable.
- Developer Productivity.
Coding style is considered as easy and elegant, using a correct English syntax that makes everything simpler than in other software. Also, the amount of code is smaller.
- Code Portability.
The code can be imported by many used, with no need to change a line of code.
- Built-in and external libraries.
This will make your life easier.

Main Python fields

- Web application development

Python boom is mostly attributed to web development: Tasks' simplification, backend data sets, handling online information and protocols like http, ftp, etc.

- Data Science and Machine Learning

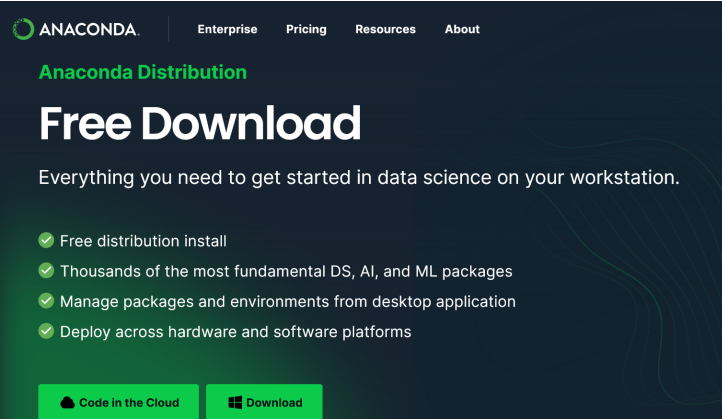
Machine learning and the automatisatisation of many economic processes (also business intelligence).

Installing Python either from python.org
(<https://www.python.org/downloads/>) or via the platform Anaconda
from anaconda.com

Recommendation to do it via Anaconda (see next slide)

The image displays two side-by-side screenshots of the Python.org website. The left screenshot shows the main 'python.org/downloads' page, featuring the Python logo, navigation links (About, Downloads, Documentation, Community, Success Stories, News, Events), and a prominent 'Download the latest version for Windows' button. Below this, it mentions 'Download Python 3.11.2' and provides links for different operating systems and pre-releases. The right screenshot shows the 'Python Releases for Windows' page, which lists 'Stable Releases' (Python 3.10.10 and 3.11.2) and 'Pre-releases'. It includes a note that Python 3.10.10 cannot be used on Windows 7 or earlier. The page lists download links for Windows embeddable packages (32-bit and 64-bit) and Windows installers (32-bit and 64-bit) for both stable and pre-release versions.

<https://www.anaconda.com/products/distribution>



The screenshot shows the Anaconda Distribution website. At the top, there is a navigation bar with the Anaconda logo and links for Enterprise, Pricing, Resources, and About. Below the navigation bar, the text "Anaconda Distribution" is displayed in green. The main heading "Free Download" is in large white font. Below this, a subtitle reads "Everything you need to get started in data science on your workstation." A list of four features is shown, each with a green checkmark icon: "Free distribution install", "Thousands of the most fundamental DS, AI, and ML packages", "Manage packages and environments from desktop application", and "Deploy across hardware and software platforms". At the bottom, there are two buttons: "Code in the Cloud" with a cloud icon and "Download" with a Windows logo icon.

ANACONDA

Enterprise Pricing Resources About

Anaconda Distribution

Free Download

Everything you need to get started in data science on your workstation.

- ✓ Free distribution install
- ✓ Thousands of the most fundamental DS, AI, and ML packages
- ✓ Manage packages and environments from desktop application
- ✓ Deploy across hardware and software platforms

Code in the Cloud Download

Anaconda Navigator
File Help













ANACONDA.NAVIGATOR

Home

Environments
Learning
Community

Applications on: base (root) Channels

Refresh

 CMD.exe Prompt 0.1.1 Run a cmd.exe terminal with your current environment from Navigator activated. Launch	 DataLore 1.1.1 Online Data Analysis Tool with smart coding assistance by JupyterLab, E66 and run your Python notebooks in the cloud and share them with your team. Launch	 IBM Watson Studio Cloud 1.0.0 IBM Watson Studio Cloud provides you the tools to analyze and visualize data, to cleanse and shape data, to create and train machine learning models, prepare data and build models, using open source data science tools or visual modeling. Launch	 JupyterLab 2.2.6 An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and architecture. Launch	 Jupyter Notebook 6.5.0 Web-based, interactive computing notebook environment. Edit and run human-readable documents describing the data analysis. Launch	 PowerShell Prompt 0.5.1 Run a PowerShell terminal with your current environment from Navigator activated. Launch
 Qt Console 4.1.7 PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical callouts, and more. Launch	 Spyder 4.1.1 Scientific Python Development Environment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features. Launch	 Glueviz 1.8.0 Multidimensional data visualization across files. Explore relationships within and among related datasets. Install	 Orange 3 3.16.0 Component-based data mining framework. Data visualization and data analysis for novice and expert. Interactive workflow with a large toolbox. Install	 PyCharm Professional 2021.1 A Full-Featured IDE by JetBrains for both Scientific and Web Python development. Supports HTML, JS, and SQL. Install	 RStudio 1.4.1108 A set of integrated tools designed to help you be more productive with R. Includes R essentials and notebooks. Install

ANACONDA NUCLEUS
[Join Now](#)

Discover premium data science content

Documentation

Anaconda Blog



Quit

Logout

Files

Running

Clusters

Select items to perform actions on them.

Upload

New

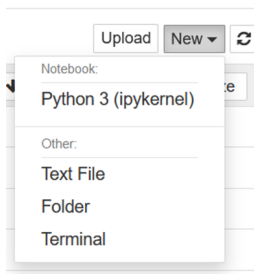


0 /		Name	Last Modified	File size
<input type="checkbox"/>	3D Objects		vor einem Jahr	
<input type="checkbox"/>	ado		vor 2 Jahren	
<input type="checkbox"/>	anaconda3		vor 10 Monaten	
<input type="checkbox"/>	Contacts		vor 2 Jahren	
<input type="checkbox"/>	Creative Cloud Files		vor 12 Stunden	
<input type="checkbox"/>	Desktop		vor 9 Minuten	
<input type="checkbox"/>	Documents		vor 2 Stunden	
<input type="checkbox"/>	Downloads		vor 8 Minuten	
<input type="checkbox"/>	Dropbox		vor 2 Monaten	
<input type="checkbox"/>	Favorites		vor 2 Jahren	
<input type="checkbox"/>	Jedi		vor 10 Monaten	
<input type="checkbox"/>	Links		vor 2 Jahren	
<input type="checkbox"/>	Music		vor 2 Jahren	
<input type="checkbox"/>	OneDrive		vor 2 Jahren	

Let's run it!

- Jupyter is both a notebook and a codebook; that is, it is used both to keep track of our script and to run estimates and algorithms in a friendly Python environment.
- The Dashboard gives you access to those files and folders created under the Jupyter environment, and you can change it if you prefer so.

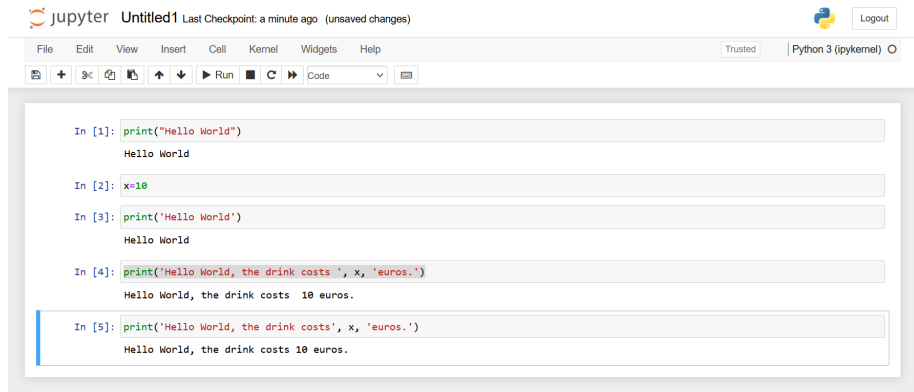
- “New”, gives you access to the different possibilities provided by Jupyter: Working in a Python environment, creating a new text file, or a folder.
- “Python 3” refers to your Jupyter Notebook.
- Your file will by default be saved as “.ipynb”. All the information generated, including strings, floats, lists, etc. will be saved in your notebook.



- Looks like the standard Microsoft Word environment (to make it more friendly).
- You have many options: Run your code, edit your text, view it, save it. . .
- Two concepts:
 - ▶ Kernel: The “engine” that is going to execute the code contained in our notebook.
 - ▶ Cell: What “contains” our text; to be fixed, modified, displayed, executed by Kernel...



1. Code Cell: Is the cell containing the code to be executed by our kernel. If everything is correct, the notebook is going to display the main results.



The screenshot displays the Jupyter Notebook interface. At the top, the title bar shows 'jupyter Untitled1' with a status message 'Last Checkpoint: a minute ago (unsaved changes)' and a 'Logout' button. Below the title bar is a menu bar with options: File, Edit, View, Insert, Cell, Kernel, Widgets, and Help. To the right of the menu bar are buttons for 'Trusted' and 'Python 3 (ipykernel)'. Below the menu bar is a toolbar with icons for file operations, navigation, and execution. The main area contains five code cells, each with a prompt 'In [n]:' followed by code and its output.

```
In [1]: print("Hello World")
Hello World

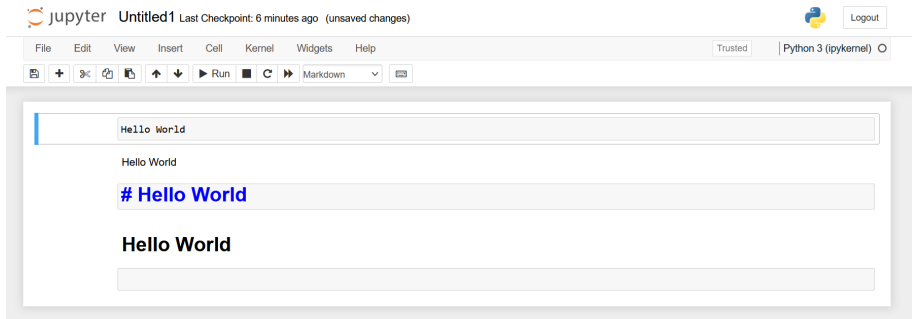
In [2]: x=10

In [3]: print('Hello World')
Hello World

In [4]: print('Hello World, the drink costs ', x, 'euros.')
Hello World, the drink costs  10 euros.

In [5]: print('Hello World, the drink costs', x, 'euros.')
Hello World, the drink costs 10 euros.
```

2. Markdown Cell: Contains text. Inserted just as it is, or with hashes. The output is displayed in a different way and not dependent on whether computations or logical operators are correct.

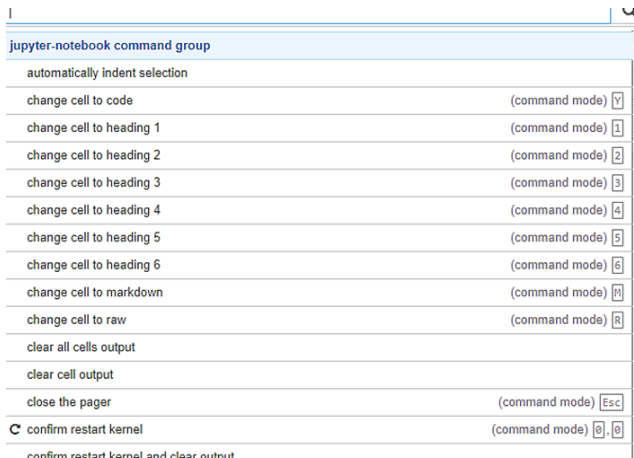


The screenshot displays the Jupyter Notebook web interface. At the top, the header shows the Jupyter logo, the file name 'Untitled1', and the status 'Last Checkpoint: 6 minutes ago (unsaved changes)'. On the right, there is a 'Logout' button and a Python logo. Below the header is a menu bar with options: File, Edit, View, Insert, Cell, Kernel, Widgets, and Help. To the right of the menu bar are buttons for 'Trusted' and 'Python 3 (ipykernel)'. Below the menu bar is a toolbar with icons for file operations (new, open, save, copy, paste), cell navigation (up, down), execution (run, stop, restart), and a dropdown menu currently set to 'Markdown'. The main content area contains a single cell with the text 'Hello World'. Below the input area, the rendered output is shown: 'Hello World' in plain text, followed by a blue header '# Hello World', and then a bold black header 'Hello World'. An empty text input field is visible at the bottom of the cell's output area.

Jupyter commands/ shortcuts: p

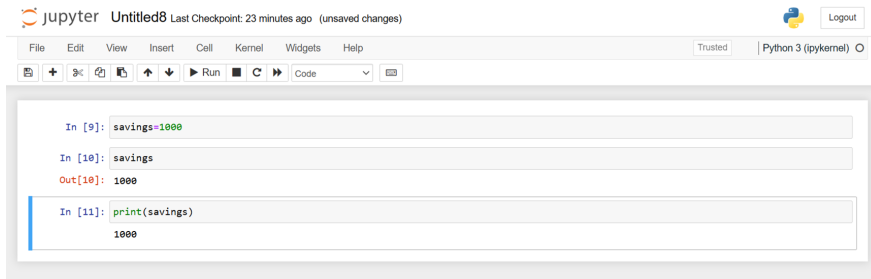
Run code: Ctrl and Enter

Print code/ Save as pdf: Ctrl and p



Variables:

Variables are what make our code and ideas feasible, in statistical and mathematical terms. We can set up the value of any variable just by “writing” it.



The screenshot shows a Jupyter Notebook window titled "Untitled8" with a last checkpoint of 23 minutes ago. The interface includes a top menu bar with options like File, Edit, View, Insert, Cell, Kernel, Widgets, and Help. Below the menu is a toolbar with icons for file operations, running, and code execution. The notebook content area displays three input cells:

```
In [9]: savings=1000
```

```
In [10]: savings
```

```
Out[10]: 1000
```

```
In [11]: print(savings)
```

```
1000
```


Data Types:

This refers to the format in which the variable has been created (string, float, etc.). In the previous example, savings was an integer.

There are many data types; floats, integers, strings, booleans, lists, tuples, sets, dicts, etc.

```
In [12]: type(savings)
```

```
Out[12]: int
```

```
In [13]: 2+3
```

```
Out[13]: 5
```

```
In [14]: "ab"+"cd"
```

```
Out[14]: 'abcd'
```