

TEXT ANALYSIS WITH PYTHON

Lesson 1 – 6/8/2021

a.y. 2020-2021



Lesson content

BUILDING A COMMON GROUND

- Why NLP in today's world: its applications
- Preliminaries
- Introduction to Jupyter Notebook
- Brief recap of Python basics
- Pandas: the essentials

Attendance registration

To track your presence in class (wherever you are), please:

- either use the app on your smartphone or tablet
- or go to this web page → www.unibocconi.it/attendance

using

- your own yoU@B credentials
- today's six-digit code*

If you have problems with the app:

- try to log out and then log in again
- if the problem persists, notify our tutor via chat who will manually register your presence



abcdef

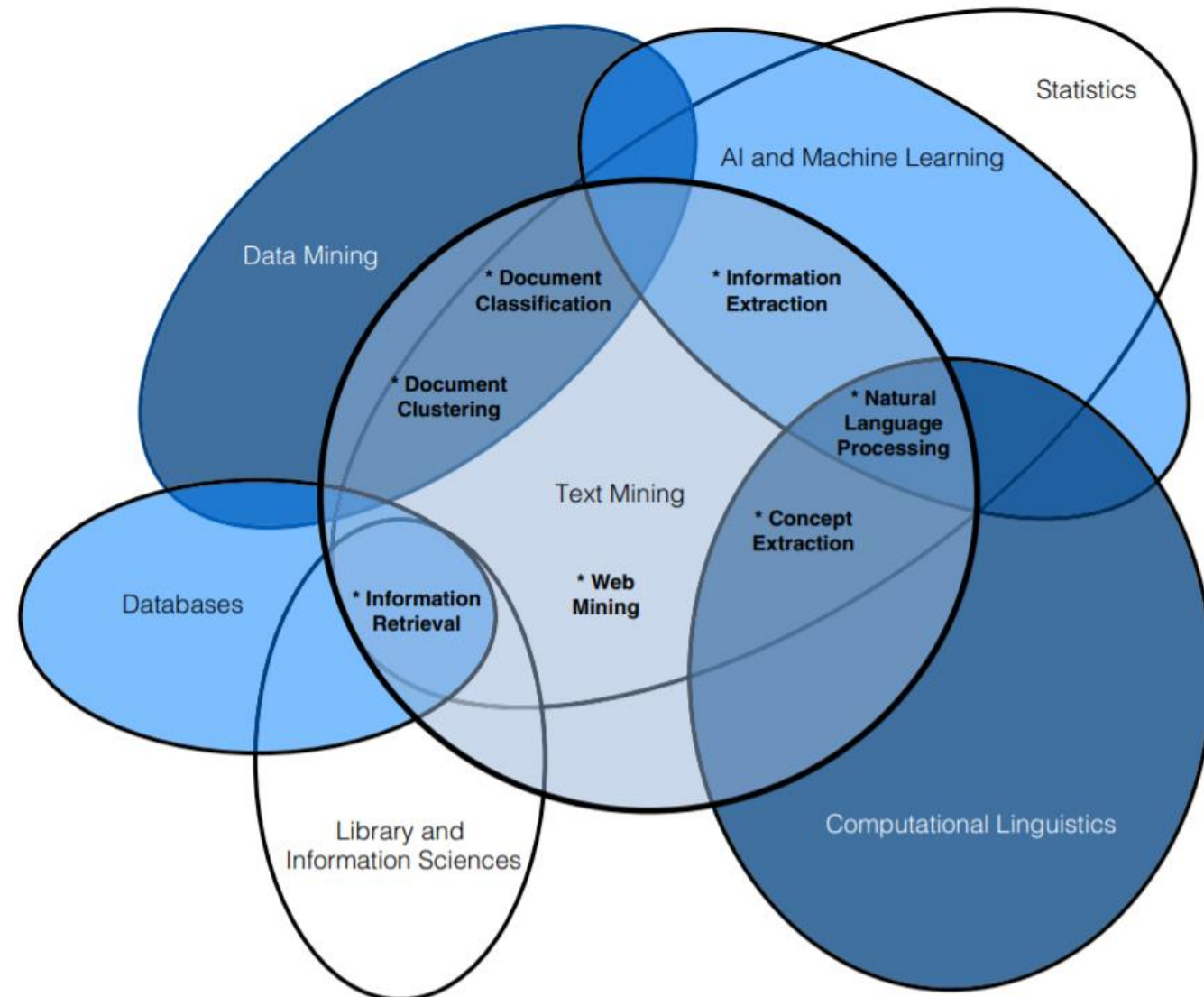
(() It will be shown in the classroom and will remain active only for 10 minutes starting from the first of you who will register*

WHY NLP TODAY?

The context

- More than 80% of the knowledge is made by not structured data = text
- Having such a wealth of material without knowing/studying it is equivalent to not having information available
- Information is of no value if it is not managed: big amount needs an automated approach
- Today 4 out of 5 companies still prefer to analyze (only) structured data

Text mining, its practice areas and the overlapping fields



Practical Text Mining and Statistical
Analysis for Non-Structured Text Data
Applications

G. Miner, D. Delen, J. Elder, A. Fast, T. Hill,
and R. Nisbet, Elsevier, January 2012

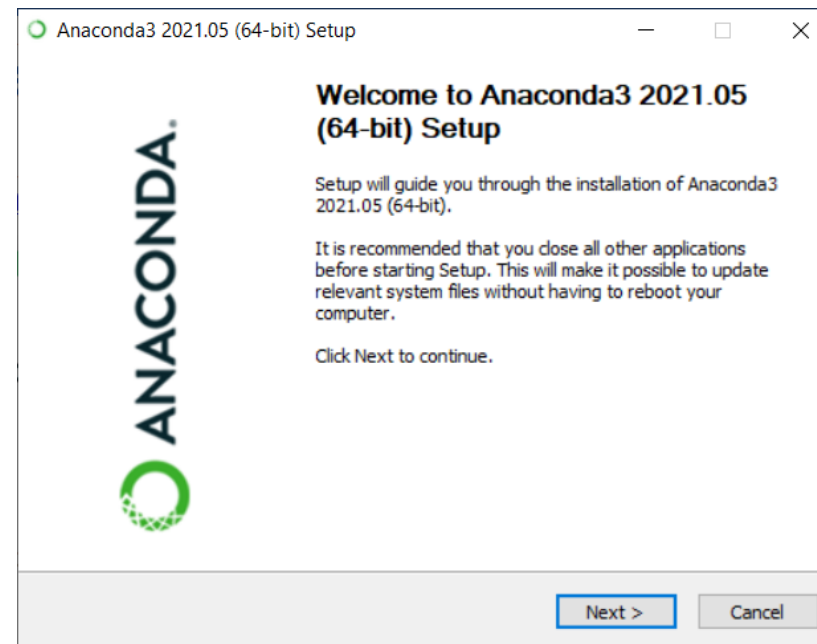
http://www.textanalyticsworld.com/wp-content/uploads/2012/03/PracticalTextMining_Excerpt.pdf



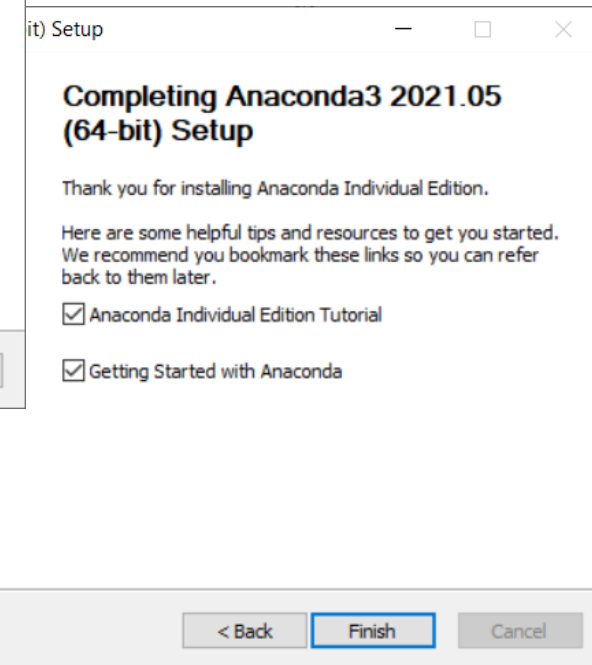
ANACONDA AND JUPYTER NOTEBOOKS

Install Anaconda

<https://www.anaconda.com/products/individual-b>



Accept all
default settings



Anaconda Navigator

Anaconda Navigator

File Help

ANACONDA.NAVIGATOR

Sign in

Refresh

Applications on base (root) Channels

Home

Environments

Learning

Community

Join Now

Discover premium data science content

Documentation

Anaconda Blog

Twitter YouTube LinkedIn

Jupyter Notebook
6.3.0
Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.
Launch

PyCharm Professional
A Full-Rledged IDE by JetBrains for both Scientific and Web Python development. Supports HTML, JS, and SQL.
Install

RStudio
1.1.456
A set of integrated tools designed to help you be more productive with R. Includes R essentials and notebooks.
Install

Orange 3
3.26.0
Component based data mining framework. Data visualization and data analysis for novice and expert. Interactive workflows with a large toolbox.
Install

Glueviz
1.0.0
Multidimensional data visualization across files. Explore relationships within and among related datasets.
Install

Spyder
4.2.5
Scientific PYTHON Development Environment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection Features
Launch

Qt Console
5.0.3
PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.
Launch

Powershell Prompt
0.0.1
Run a Powershell terminal with your current environment from Navigator activated
Launch

JupyterLab
3.0.14
An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.
Launch

IBM Watson Studio Cloud
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

Datalore
Online Data Analysis Tool with smart coding assistance by JetBrains. Edit and run your Python notebooks in the cloud and share them with your team.
Launch

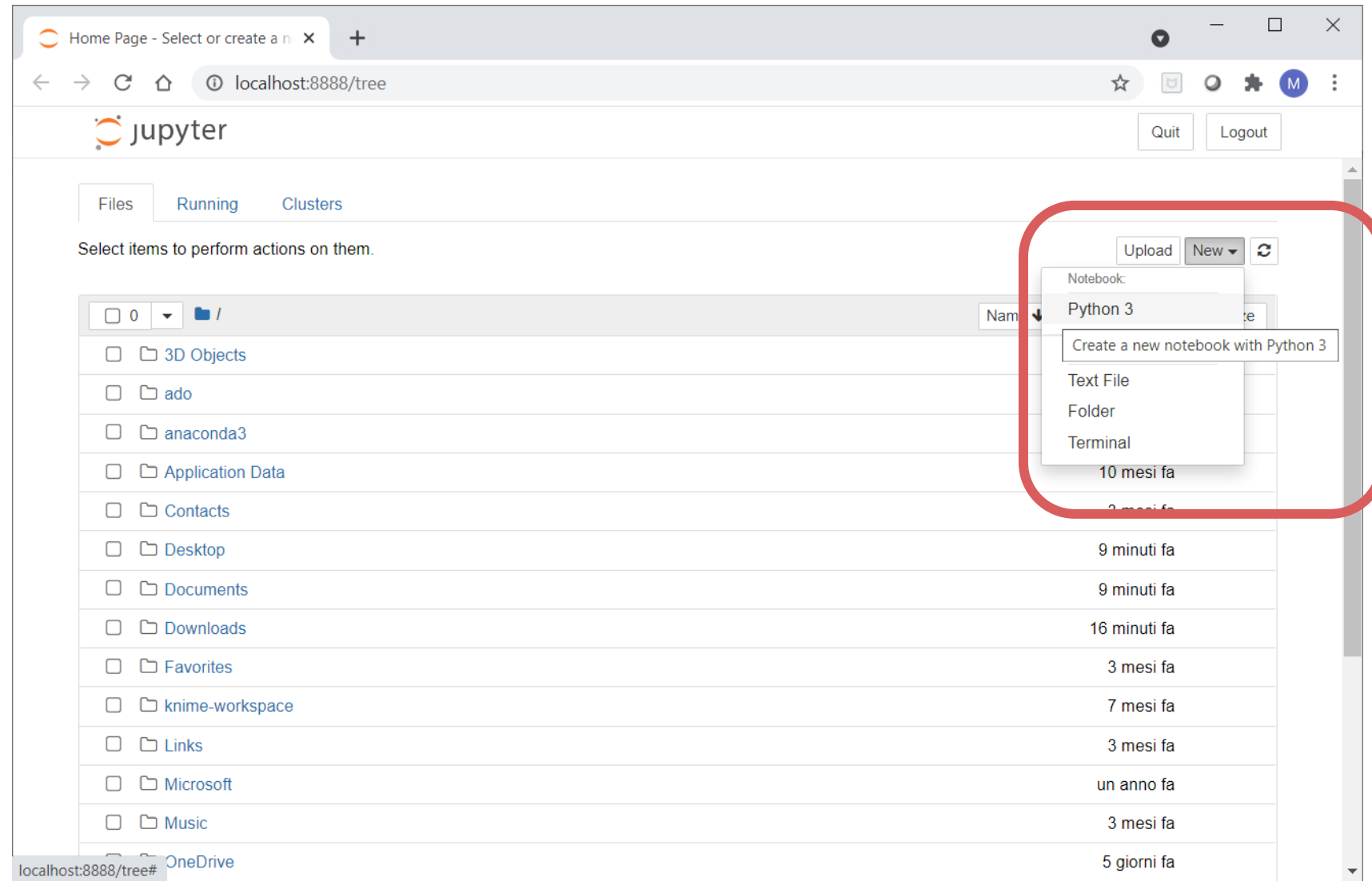
CMD.exe Prompt
0.1.1
Run a cmd.exe terminal with your current environment from Navigator activated
Launch

https://docs.anaconda.com/anaconda/user-guide/?utm_source=anaconda.com&utm_medium=individual-get-started

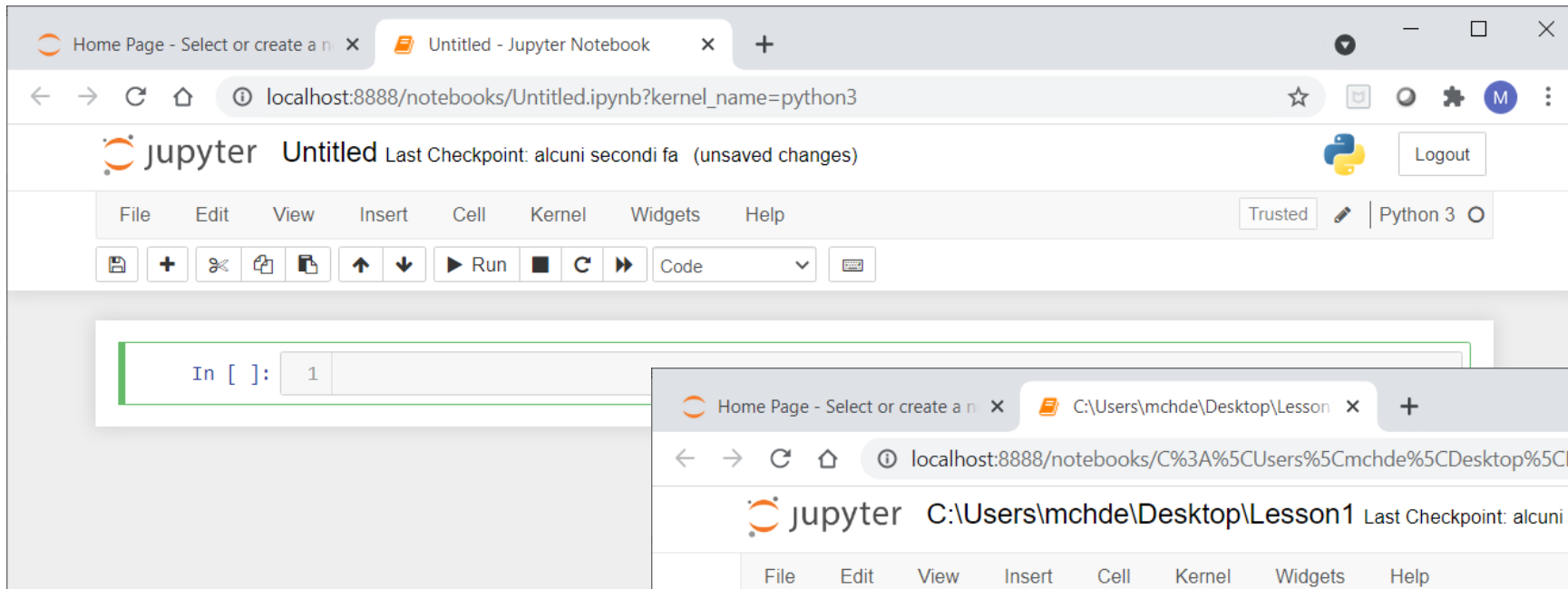


Notebook Dashboard

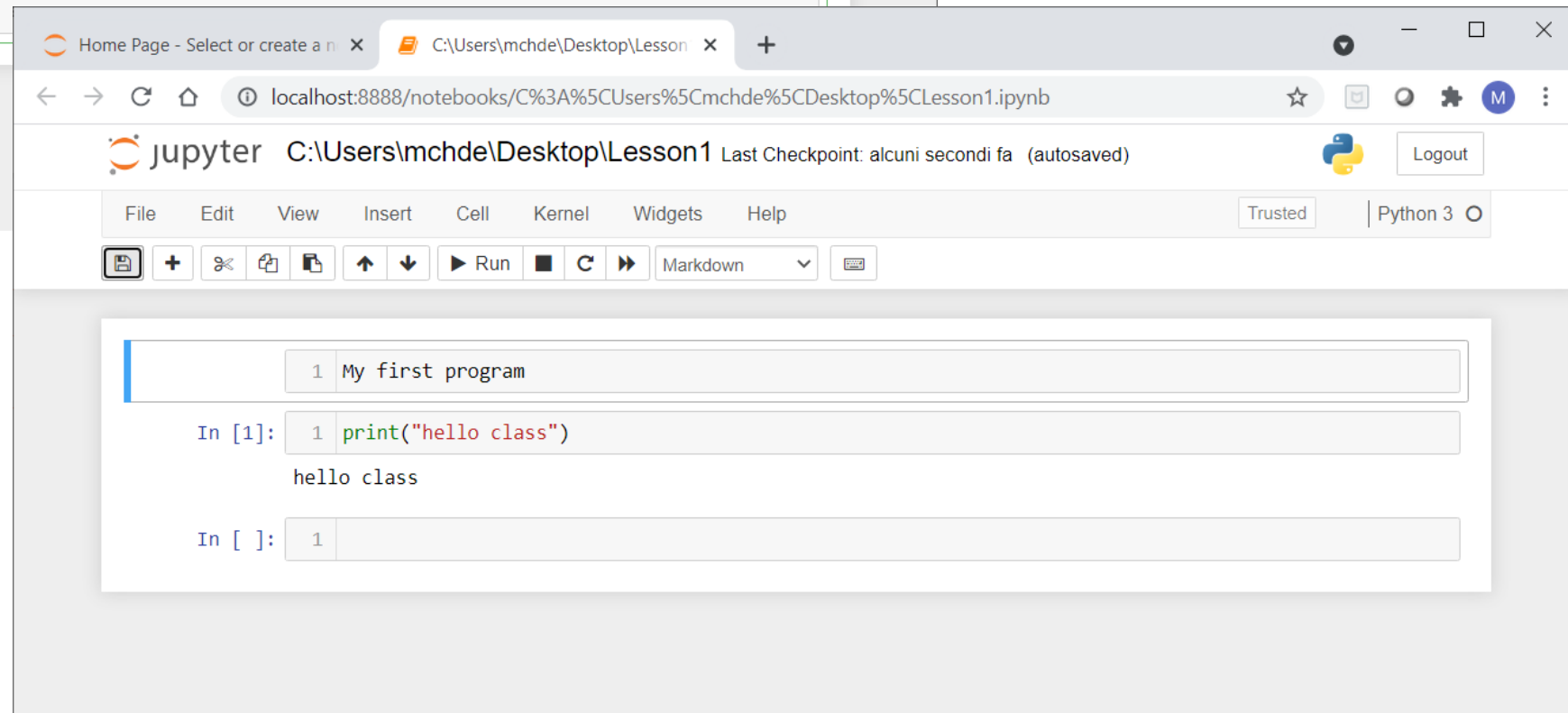
- It is mainly used to open notebook documents, and to manage the running kernels (visualize and shutdown).
- It has other features similar to a file manager, namely navigating folders and renaming/deleting files.



Create a new notebook



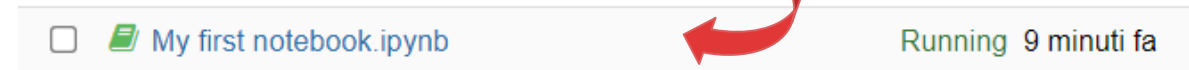
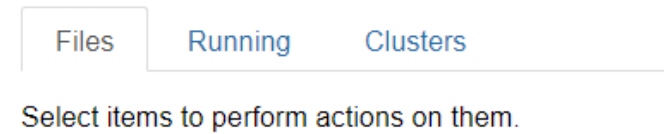
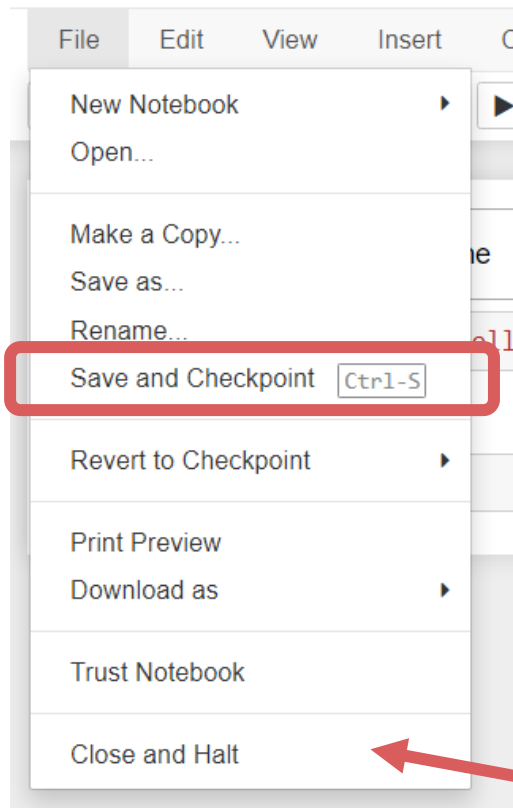
<https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/index.html>



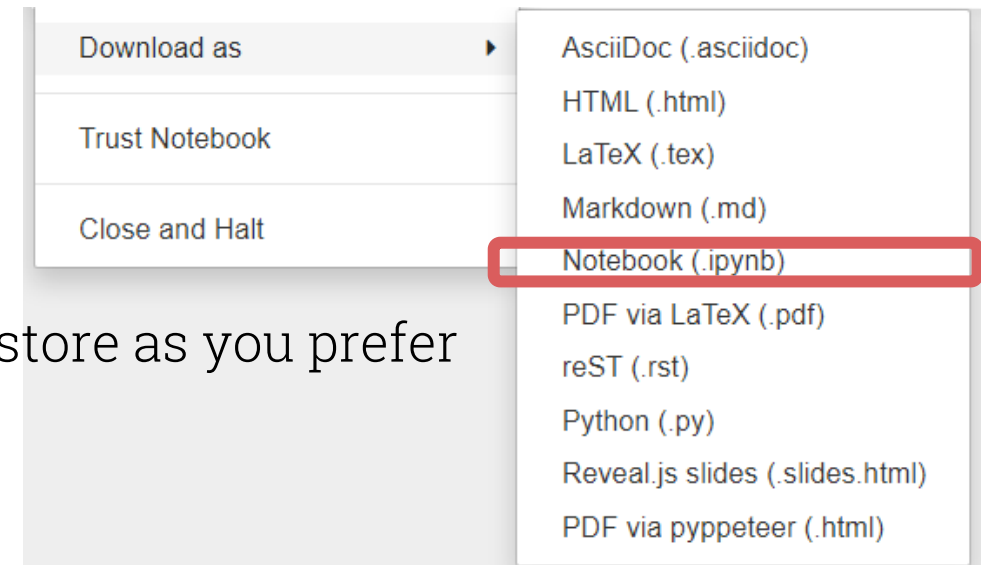
Open, save and download

- Open a notebook from the Dashboard

- Save it



- Download and store as you prefer
- After Close it



AND NOW LET'S DO SOME CODING...