

CPS

DataScience con Python

Salve!

- Massimo Regoli
- regoli@uniroma2.it
 - [CPS2023] nel subject
- Studio:
 - Facoltà di Ingegneria,
 - Edificio Informazione,
 - Piano terra,
 - Stanza AT05
- Ricevimento:
 - Martedì su appuntamento
 - SOLO ONLINE

Il corso

- Lezioni
 - Mercoledì 11:30-13:15, aula
- Più che altro esercitazioni su alcuni interessanti dataset
- Il corso si baserà sull'ambiente *Jupyter Lab* nella configurazione presente nel pacchetto *Anaconda*
 - Ma nulla vieta di usare altre piattaforme
- Non verranno registrate le esercitazioni né fornito lo *streaming online*

L'esame

- Come avrete sentito dal prof. Monte, l'esame sarà diviso in scritto ed orale
 - Nella prova orale sarà richiesta la presentazione di un progetto di corso
 - Progetto che deve essere richiesto da un gruppo (*) attraverso Teams
 - Il progetto verrà consegnato il giorno dell'esame
 - È cosa buona e giusta che tutti i membri del gruppo facciano l'esame durante lo stesso appello
 - La discussione del progetto potrà avvalersi di un file di presentazione
 - Il contenuto della consegna dovrà comprendere:
 - Un manuale dei requirements e di uso dell'applicazione
 - Un file in python (notebook o altro) contenente tutto il codice
 - Tutti i file necessari per il funzionamento del progetto
 - Le richieste di progetto inizieranno verso dicembre

() dicesi gruppo un insieme di almeno 2 studenti*

Anaconda

Data science technology for a better world.

Anaconda offers the easiest way to perform Python/R data science and machine learning on a single machine. Start working with thousands of open-source packages and libraries today.

Download 

For Windows

Python 3.9 • 64-Bit Graphical Installer • 594 MB

Get Additional Installers

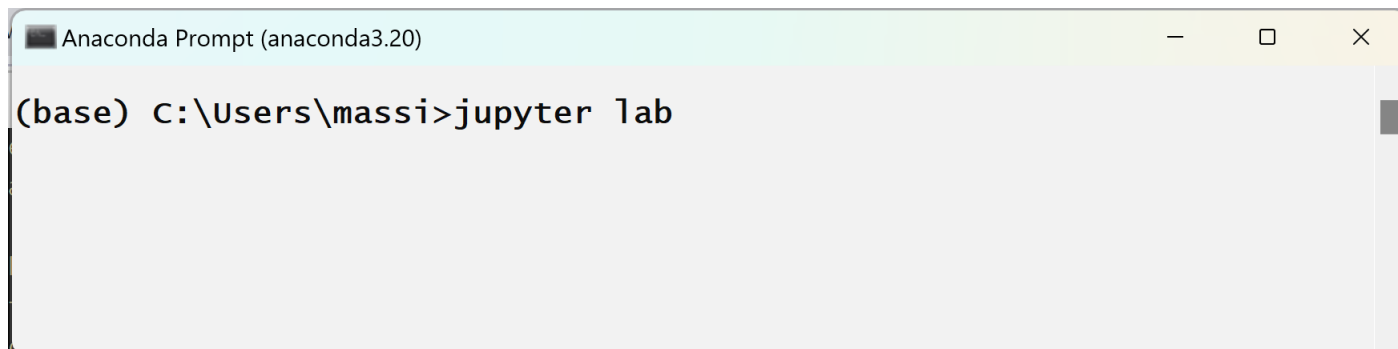
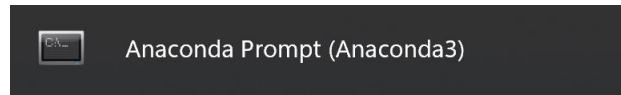


Jupyter Notebook

- *Jupyter Notebook* è un'applicazione Web open source che permette di creare e condividere documenti testuali interattivi, contenenti oggetti quali equazioni, grafici e codice sorgente eseguibile.
- *Jupyter* è diventato uno standard *de-facto* per data scientist perché offre la possibilità di realizzare, documentare e condividere analisi di dati all'interno di un framework [cit.]

Best practice

- Io preferisco *jupyter lab*
 - Interfaccia migliore e moderna
 - Pluginable
- Lanciare *jupyter lab* tramite *Anaconda Navigator* può risultare un processo lungo e farraginoso, meglio usare il prompt dei comandi (sia Windows, Linux, MacOS)



Jupyter lab

The screenshot displays the JupyterLab web interface in a browser window. The address bar shows the URL `http://localhost:8888/lab/tree/jupyter/cps23`. The interface is divided into several panels:

- File Browser (Left Panel):** Contains a search bar labeled "Filter files by name" and a file list for the directory `/ jupyter / cps23 /`. The list includes:

| Name | Last Modified |
|------------------------|---------------|
| heart_2020_cleaned.csv | 18 days ago |
| heart.ipynb | 18 days ago |
- Launcher (Right Panel):** Displays a grid of icons for launching different environments and file types under the heading `jupyter/cps23`.
 - Notebook:** Includes icons for Python 3 (ipykernel), Julia 1.4.2, Kotlin, and R.
 - Console:** Includes icons for Python 3 (ipykernel), Julia 1.4.2, Kotlin, and R.
 - Other:** Includes icons for Terminal, Text File, Markdown File, Julia File, Python File, and R File.

The bottom status bar shows "Simple" mode, a file count of "0", and a kernel count of "7". The "Launcher" label is visible in the bottom right corner.

Notebook - 1

The screenshot displays the JupyterLab web interface. On the left, a file browser shows the directory structure: `/jupyter / cps23 /`. It lists three files: `CPS01.ipynb` (modified seconds ago), `heart_2020_cleaned.csv` (modified 18 days ago), and `heart.ipynb` (modified 18 days ago). The main area on the right is the notebook editor, titled `CPS01.ipynb`. It contains the following content:

Esercitazione 1

Notebook di Jupyter Lab

Un notebook di jupyter lab permette di alternare codice python a notazioni in formato markdown.

Il formato markdown permette di inserire delle annotazione formattate testualmente co effetti come:

1. *corsivo*
2. **grassetto**
3. ***corsivo grassetto***
4. `testo verbatim per codice`
5. Stili di
 - A. Titolo (un #)
 - B. Sottotitolo (due ##)
 - C. ecc.
6. Elenchi numerati
 - o elenchi
 - puntati
 - tutti e due anche in
 - forma nidificata
 - come qui

The bottom status bar indicates the interface is in 'Simple' mode, using 'Python 3 (ipykernel)', and the current cell is at 'Ln 1, Col 1'.

Notebook - 2

The screenshot displays the JupyterLab web interface in a browser. The top bar shows the URL `http://localhost:8888/lab/tree/jupyter/cps23/CPS01.ipynb`. The left sidebar contains a file browser for the `/jupyter/cps23/` directory, listing `CPS01.ipynb` (modified seconds ago), `heart_2020...` (18 days ago), and `heart.ipynb` (18 days ago). The main area is split into two panes. The left pane shows the notebook code with the following Python code:

```
[2]: import random  
x = [1, 2, 3, 4, 5]  
random.shuffle(x)  
  
[ ]:
```

The right pane is the console, titled "Console 5", showing the output of the code:

```
Python 3.7.11 (default, Jul 27 2021, 09:42:29) [MSC v.1916 64 bit (AMD64)]  
Type 'copyright', 'credits' or 'license' for more information  
IPython 7.31.1 -- An enhanced Interactive Python. Type '?' for help.  
  
[3]: x  
[3]: [5, 3, 4, 2, 1]
```

The bottom status bar indicates the interface is in "Simple" mode, using "Python 3 (ipykernel)", and shows the current position as "Ln 1, Col 1" in the `CPS01.ipynb` file.

Alzate la mano

1. Chi conosce Python?
2. Chi conosce Jupyter Lab
3. Sapete cos'è un codice in formato Markdown?
4. Chi conosce Numpy e/o Pandas?

Per *conoscenza* intendo che siate capaci di scrivere un qualcosa complesso in quel linguaggio o libreria

Librerie

- Librerie di base per Data Science
 - [math](#)
 - Raccolta di funzioni matematiche
 - [random](#)
 - Generatori di numeri casuali
 - [datetime](#)
 - Gestione delle date
 - [time](#)
 - Gestione del tempo
 - [json](#)
 - Raccolta di funzioni per la conversione dict \Leftrightarrow json

Librerie

- Librerie avanzate per Data Science
 - [numPy, sciPy](#)
 - Calcolo vettoriale (alla Matlab)
 - numpy.random (best random)
 - [pandas](#)
 - Analisi dati
 - [matplotlib](#)
 - Grafici
 - *seaborn*
 - Grafici
 - [sqlite3, dataset, mysql.connector](#)
 - database

Librerie

- Librerie specifiche per Data Science
 - [BeautifulSoup](#)
 - Scraping Web
 - [Selenium](#)
 - Web Browser automation

Librerie

- Altre utili librerie
 - [requests](#)
 - Accesso a internet