

Analisi dati di CORAM

Prof. Francesco de Palma

Unisalento & INFN Lecce

Francesco.depalma@unisalento.it

Introduzione all'analisi

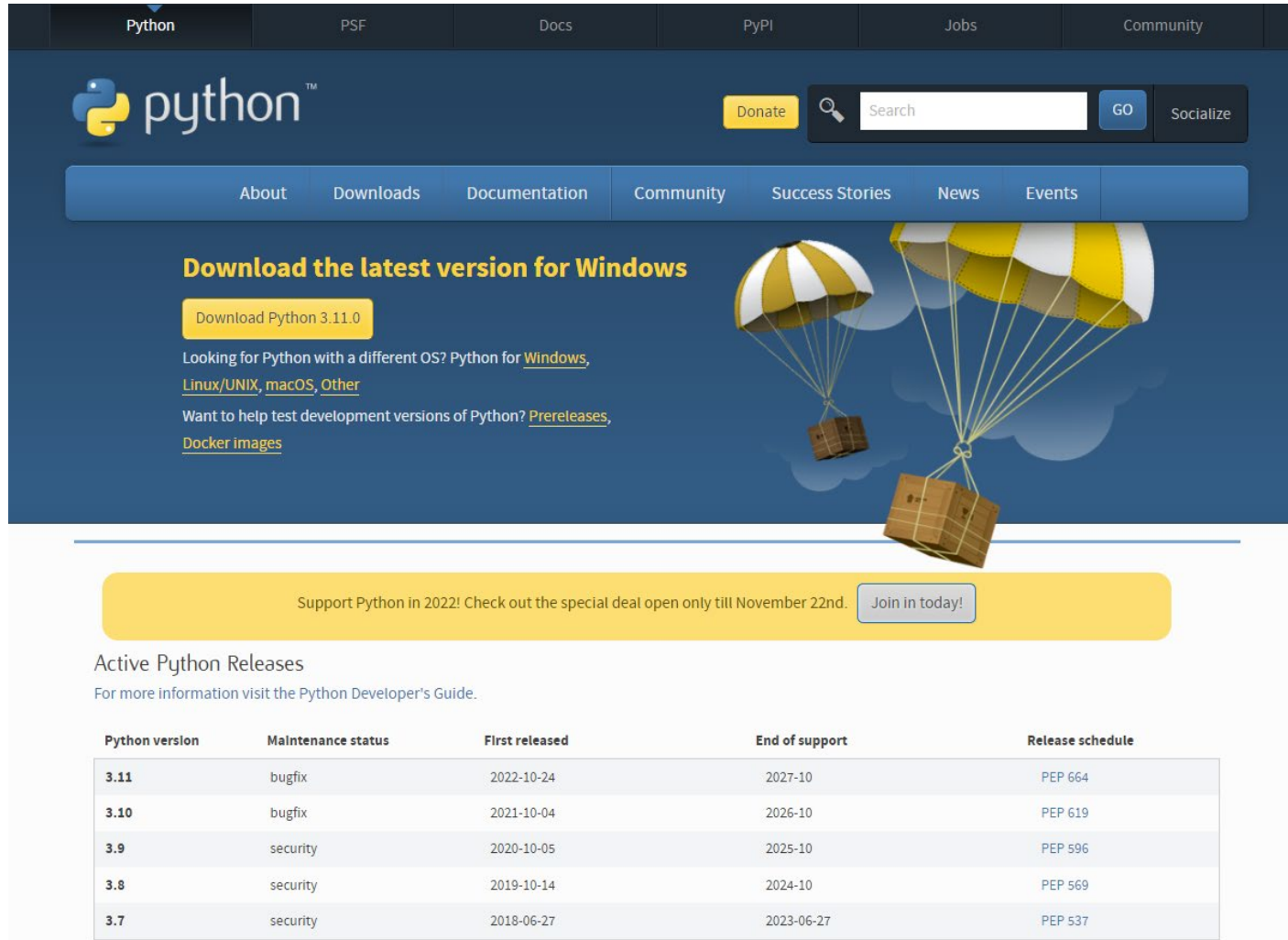
- Per fare l'analisi utilizzeremo un semplice linguaggio in *python* che verrà eseguito tramite il *browser* in una *macchina virtuale* di google.
- Un analisi simile potrebbe essere svolta tramite excel del pacchetto office (o altri fogli di calcolo). Questo tipo di analisi era meno facilmente distribuibile ed utilizzabile durante questo evento.

Introduzione a python

- **Python** è un linguaggio di programmazione di *alto livello ed interpretato*.
- Alto livello, ovvero, progettato per essere facilmente comprensibili dagli esseri umani, fino a includere alcuni elementi del linguaggio naturale. Per essere eseguiti da un calcolatore, questi programmi devono essere tradotti o interpretati da un altro programma.
- Interpretato, ovvero, che i programmi che andremo a scrivere non dovranno essere compilati in linguaggio macchina, ma l'interprete si occuperà di eseguire un'istruzione per volta.
- Inoltre python è un linguaggio orientato agli oggetti, ciò ci permette di interagire con diversi oggetti e porta dei benefici come una più facile gestione del codice e una migliore organizzazione.

Come ottenere python?

<https://www.python.org/downloads/>



The screenshot shows the Python.org website. At the top, there is a navigation bar with links for Python, PSF, Docs, PyPI, Jobs, and Community. Below this is the Python logo, a search bar, and a 'Socialize' button. A main navigation bar contains links for About, Downloads, Documentation, Community, Success Stories, News, and Events. The main content area features a large banner with the text 'Download the latest version for Windows' and a 'Download Python 3.11.0' button. Below this, there are links for other operating systems and pre-releases. A yellow banner at the bottom of the main content area promotes a special deal for 2022. Below the banner is a section titled 'Active Python Releases' with a link to the Python Developer's Guide. A table lists the active Python versions, their maintenance status, first release date, end of support date, and release schedule.

Python version	Maintenance status	First released	End of support	Release schedule
3.11	bugfix	2022-10-24	2027-10	PEP 664
3.10	bugfix	2021-10-04	2026-10	PEP 619
3.9	security	2020-10-05	2025-10	PEP 596
3.8	security	2019-10-14	2024-10	PEP 569
3.7	security	2018-06-27	2023-06-27	PEP 537

Perché python?

- E' un linguaggio libero e gratuito con una community molto attiva
- E' portabile, ovvero può essere usato su svariate piattaforme
- E' molto semplice da usare ed intuitivo
- E' ricco di librerie o moduli estremamente utili (noi useremo le più comuni: numpy, pandas, matplotlib) ma ne esistono svariate ed è possibile sempre scriverne di nuove
- E' veloce, nonostante sia interpretato e quindi inevitabilmente meno veloce di un linguaggio compilato (alcuni moduli sono implementati in C e compilati per maggiore efficienza)
- Interagisce facilmente con altri linguaggi
- E' ampiamente utilizzato in moltissime analisi di fisica (e non solo)

PYPL Popularity of Programming Language

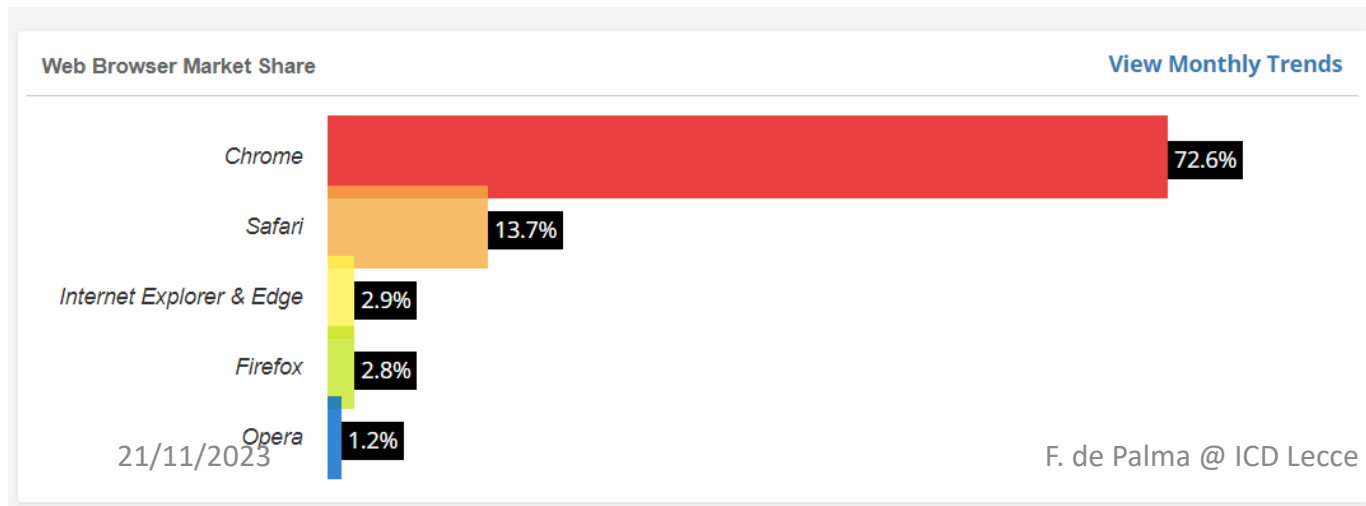
Worldwide, Nov 2023 :

Rank	Change	Language	Share	1-year trend
1		Python	27.99 %	+0.0 %
2		Java	15.91 %	-0.8 %
3		JavaScript	9.18 %	-0.3 %
4	↑	C/C++	6.76 %	+0.2 %
5	↓	C#	6.67 %	-0.3 %
6		PHP	4.86 %	-0.3 %
7		R	4.45 %	+0.4 %
8		TypeScript	2.95 %	+0.1 %

Created by analyzing how often language tutorials are searched on Google.
<https://pypl.github.io/PYPL.html>

Cos'è un Browser?

- *Il browser Web, o navigatore Web, è un'applicazione per l'acquisizione, la presentazione e la navigazione di risorse sul Web.*
- I browser più comuni sono Google Chrome, Safari, Edge, Firefox e Opera
- Colab funziona con la maggior parte dei browser principali ed è testato in modo approfondito con le ultime versioni di Chrome, Firefox e Safari.



<https://www.w3counter.com/globalstats.php>

Cos'è una macchina virtuale?

- Una **macchina virtuale** (denominata *guest*) viene creata all'interno di un ambiente digitale (denominato *host*). In un *host* possono esistere contemporaneamente più macchine virtuali.
- Le macchine virtuali sono software che offrono le stesse funzionalità dei computer fisici. Come questi ultimi, eseguono delle applicazioni e un sistema operativo.
- Tuttavia, le macchine virtuali sono file digitali che vengono eseguiti su un computer fisico e si comportano come un computer fisico. In altre parole, le macchine virtuali si comportano come sistemi digitali separati.

Perché usiamo una macchina virtuale?

- Per rendere semplice la condivisione e non farmi installare software specifico sul vostro PC!
- Abbiamo bisogno unicamente di un browser per fare l'analisi!
- Utilizzeremo le risorse messe a disposizione da Google

Cos'è google colab?

- Colab, o "Colaboratory", ti permette di scrivere ed eseguire Python nel tuo browser con:
 - Nessuna configurazione necessaria
 - Accesso alle GPU senza costi
 - Condivisione semplificata
- E' un sistema pensato per il machine learning, data analysis e la divulgazione
- In maniera dettagliata, Colab è un servizio di «jupyter notebook» che non richiede nessuna installazione ed usa le GPU (e TPU) di google
- <https://research.google.com/colaboratory/faq.html> .

Che librerie utilizzeremo?

- Una libreria informatica, è un insieme di codice pre-scritto, funzioni, procedure e classi che possono essere utilizzate da un programma per svolgere determinate operazioni senza dover scrivere quel codice da zero.
- Noi utilizzeremo:
 - Numpy: la libreria fondamentale per l'analisi scientifica in python
 - Pandas: una libreria utile per la manipolazione dei dati
 - Matplotlib: una libreria utile per la visualizzazione dei dati

NumPy 

 pandas

matplotlib 

Cosa faremo con colab?

- Caricheremo le librerie necessarie
- Inseriamo i dati di **CORAM**
- Li analizziamo facendo le medie e le frequenze per doppie, triple e quadruple
- Li disegniamo in funzione di angolo, coseno e coseno al quadrato
- Valuteremo i parametri della retta (polinomio di grado 1: $y = p_1x + p_0$) che meglio descrive i punti e quanto i punti lungo la retta si discostano dal valore misurato, minimizzando:

$$E = \sum_{j=0}^k |p(x_j) - y_j|^2$$

- Salviamo i parametri ed il residuo finale di ciascun fit

Dove troviamo il colab di oggi?

- Qui:

<https://l.infn.it/icdlecce22>

- Dovrete aver fatto l'accesso con un vostro account google.
- Vi uscirà:
- **Avviso: l'autoring di questo blocco note non è stato eseguito da Google.**
- Rispondete: Esegui comunque

Modifica e visualizzazione

- Potete editare il codice come volete (ed ovviamente potrebbe non funzionare più dopo le modifiche)
- Se vorrete salvare il file modificato esso verrà salvato sul **vostro** google drive
- Per eseguire un blocco di codice basta cliccare sulla freccetta:



```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from matplotlib.pyplot import plot, figure, clf
from IPython.display import display, clear_output, Image
import math
```

