

PCTO in Coding & Data Science

CD: 50/50 - Coding Diversity

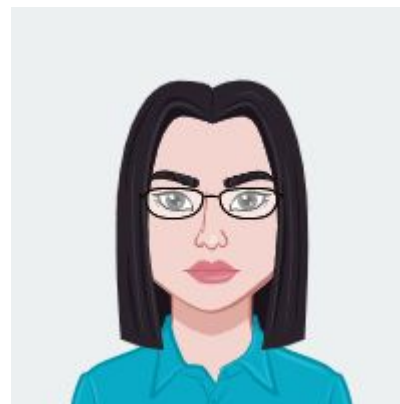
Liceo Scientifico S. Cannizzaro

14/03/2022

CD: 50/50
Coding Diversity



Liceo Cannizzaro



Elena & Federica

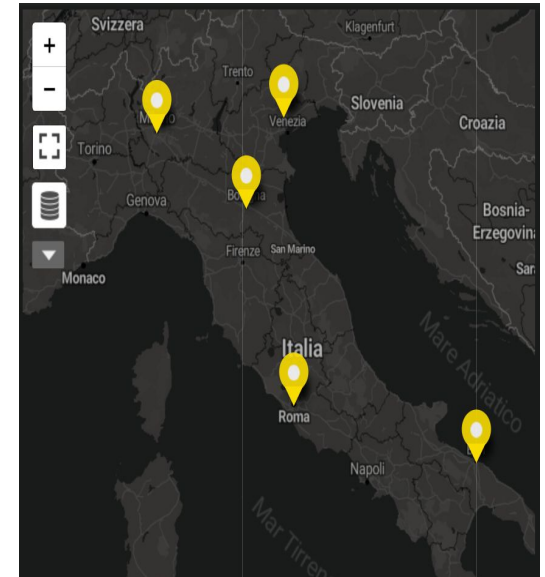
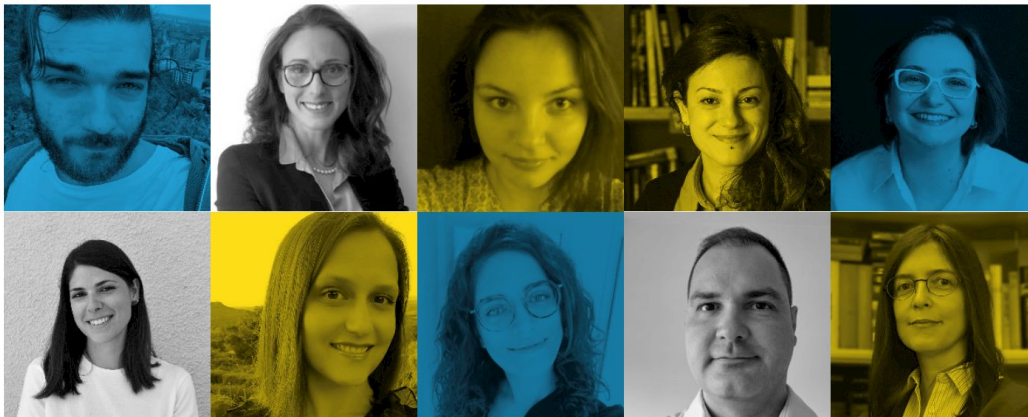
**Il data scientist è il terzo
profilo professionale più
richiesto a livello globale**

Si iscrivono a corsi di laurea STEM:

6 ragazzi su 10

2 ragazze su 10

**(STEM = Science, Technology, Engineering
and Mathematics)**



CD:50/50
CODING DIVERSITY

<https://cd5050.org/>

Alcune info pratiche

- Il **materiale** lo trovate qui:
<https://github.com/federicadaniele/PCTOcodingdataascience>
- Ogni giorno vi daremo **alcuni semplici esercizi** da provare a fare a casa **nel pomeriggio** e ne discuteremo la mattina dopo in classe
- Per **comunicare** possiamo utilizzare Telegram:
<https://t.me/+YrAVEnMYC6I2NmI0>

Modulo 1: Introduzione al coding

Modulo 2: Saper leggere e rappresentare i dati

Modulo 3: Basi di inferenza e analisi predittiva

Modulo 4: Basi di machine learning

Modulo 1: Introduzione al coding

1. L'unità base: il dato
2. Un insieme di dati: il database

Cos'è secondo voi un dato?

Alcuni esempi di dati...

Età = 25

Colore capelli = marroni

I dati possono avere diversi **formati**, *numerico* o *carattere*.

Questi sono esempi di **dati “strutturati”**. Poi esistono i **dati “non strutturati”**.

Una canzone, una foto.

Come rappresentaresti una canzone di Beyonce?





Wordcloud basata sui testi delle canzoni di Beyonce.



Fonte: <https://www.databasic.io/en/wordcounter/>

Definizioni di dato

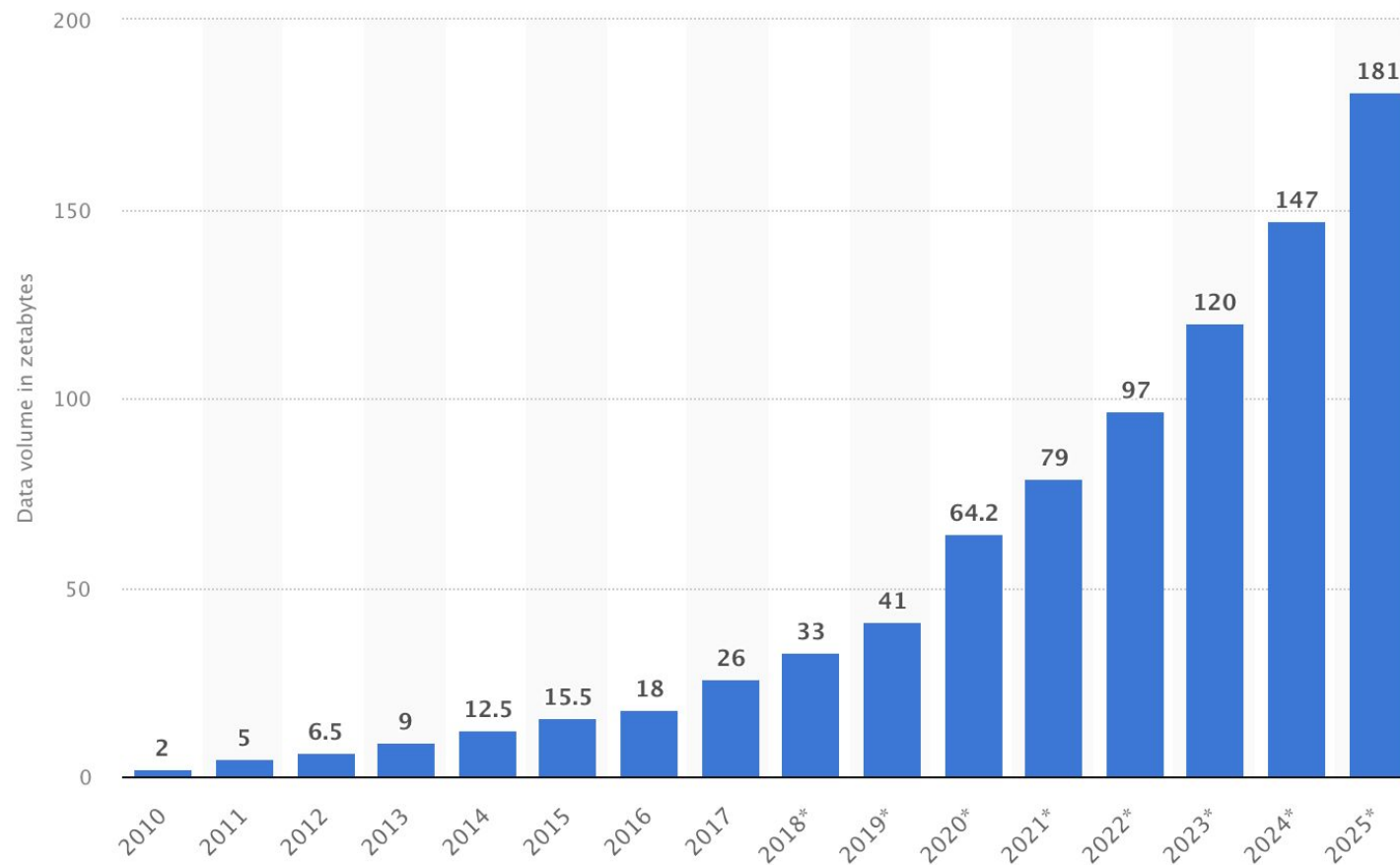
- Il termine **dato** deriva dal latino “*datum*”, che significa letteralmente “**fatto**” ed è un valore grezzo inserito in contesto e dotato di significato sulla base del quale è possibile ricavare un’**informazione**
- Nel corso degli anni è aumentata la capacità di “immagazzinare” dati **STORAGE**


- **Definizione “moderna” di dato:** qualsiasi informazione che può essere letta da un computer

Quanti dati produciamo in un dato istante?

Produciamo dati ogni volta che...

- Accendiamo la luce
- Timbriamo il biglietto dell'autobus
- Siamo presenti all'appello in classe
- Facciamo un acquisto con la carta di credito
- Pubblichiamo una storia su Instagram
- Facciamo una ricerca su Google
- ...

Nel 2020, ciascun essere umano ha creato in media 1.7 MB di dati AL SECONDO.



© Statista 2021

Fonte: <https://www.statista.com/statistics/871513/worldwide-data-created/>.

Uno zettabyte: 1 seguito da 21 zeri.

Nel 2018 1% dell'elettricità prodotta a livello globale era utilizzata per immagazzinare dati



La **data science** ci risolve il problema di cosa farci con questa montagna di dati!



“... quella disciplina orientata all'estrazione e analisi di grandi volumi di dati (“big data”) per mezzo di moderne tecniche/strumenti al fine di estrarne informazioni funzionali al miglioramento delle operazioni di una determinata organizzazione.”

Cosa ci si fa con la data science?

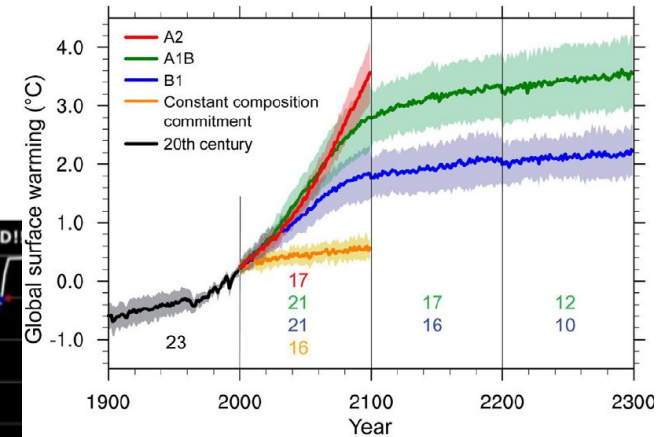
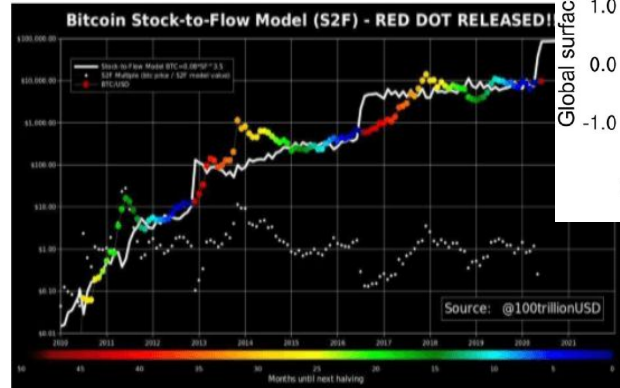
Raccomandazione di pagine/siti

Pubblicità targettizzata



Cosa si fa con la data science?

Previsioni (e non quelle del meteo!)



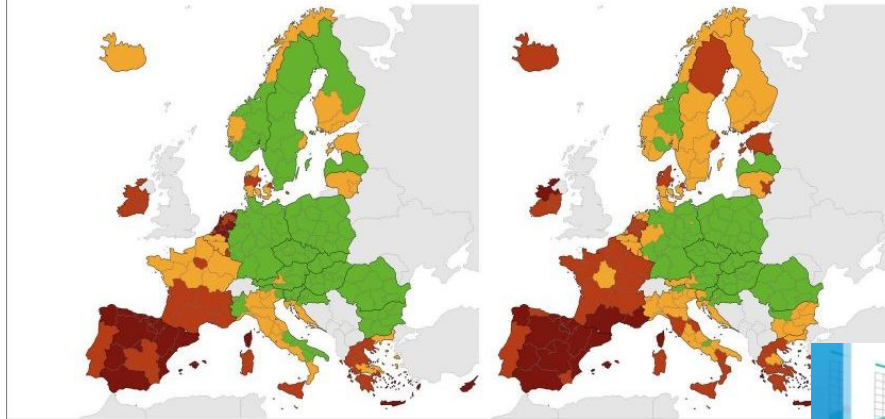
Intelligenza artificiale



Cosa si fa con la data science?

Statistica e analisi empirica

COME È CAMBIATA LA SITUAZIONE IN EUROPA IN DUE SETTIMANE: DAL 29 LUGLIO AL 12 AGOSTO



Cos'è la programmazione?

*“... il linguaggio che utilizziamo per comunicare/impartire comandi al **computer**.”*



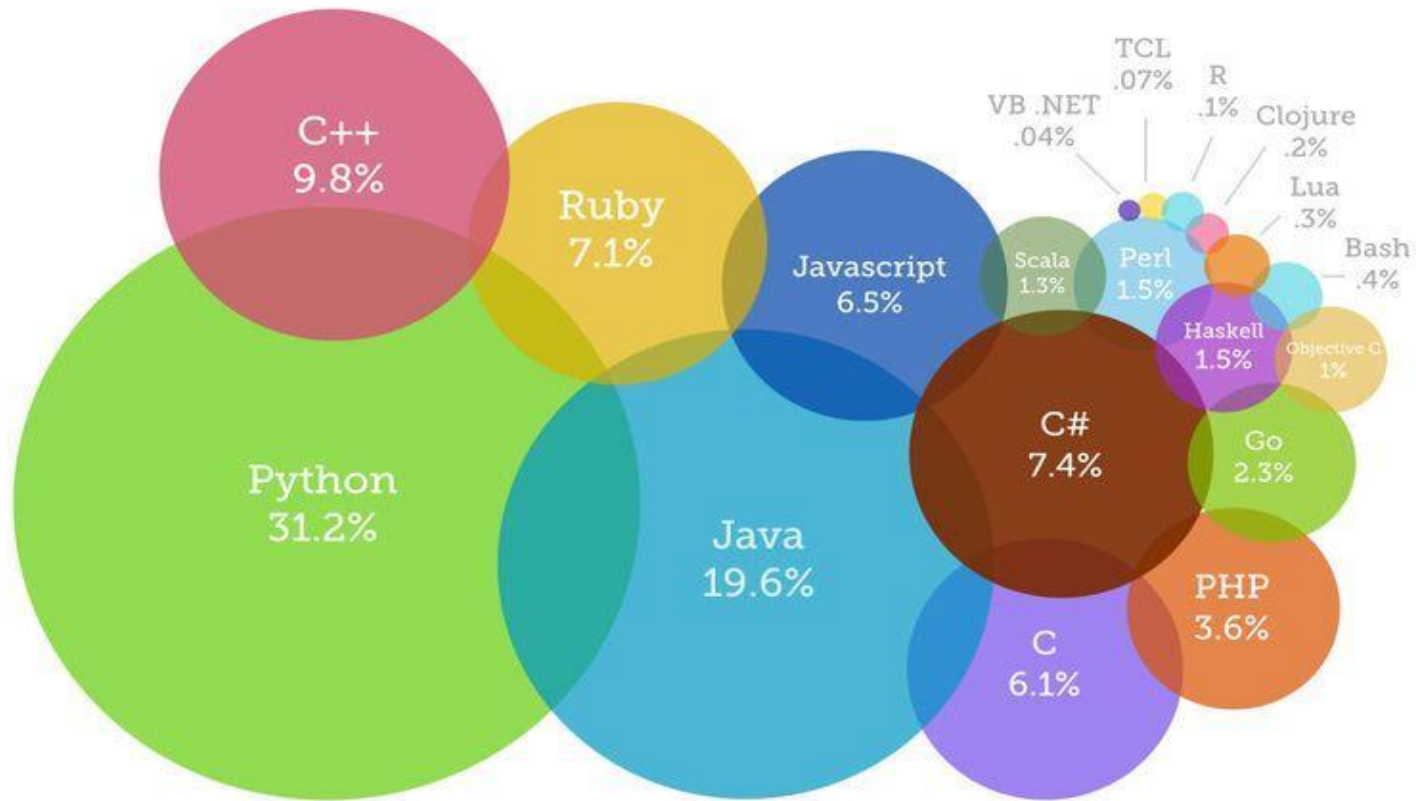
Her (2013)

La carta e la penna dei linguaggi di programmazione

1. **“script”**: foglio su cui “scriviamo” quando programmiamo
2. lo script è organizzato in righe, sulle righe si scrivono i **comandi**
3. l'insieme dei comandi si chiama **codice** (come un testo)

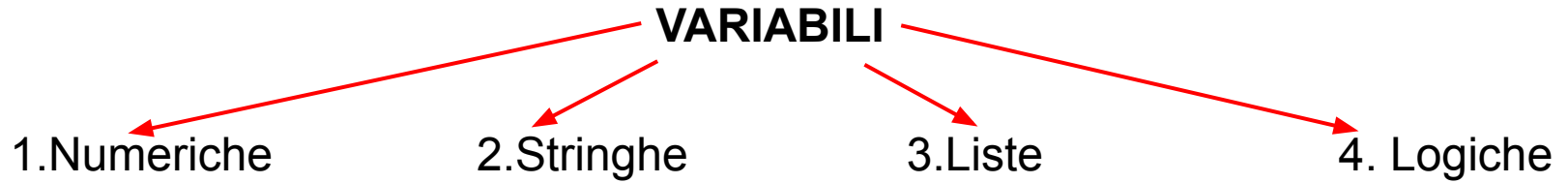
Spesso il codice è un insieme ordinato di comandi, come una storia che segue un suo filo temporale: si chiede al computer di eseguire prima i comandi all'inizio, scorrendo via via di riga in riga...

Esistono molti linguaggi di programmazione!



noi utilizziamo **Python!**

Gli elementi di un codice



I COMANDI CONDIZIONALI

Salvataggio dei dati in memoria: le variabili

Creare una variabile in un linguaggio di programmazione significa associare un nome a un contenuto (per es. un dato), e salvarlo nella memoria del programma.

Il contenuto lo possiamo creare sul momento o importarlo da una fonte esterna (per esempio un file Excel).

Esempi di variabili cui sono associati diversi tipi di dati:

```
nome = "John"
```

```
# Una stringa
```

```
eta = 31
```

```
# Una variabile numerica
```

```
maggiorenne = eta > 18
```

```
# Una variabile logica
```

```
mylist = [ 'Annie', 160, 'John', 170 ]
```

```
# Una lista
```

Cosa ci facciamo con i dati: le funzioni

→ dipendono da dei parametri

1. funzioni pre-definite: print(), dir(), type()

```
print("Hello world")
```

```
>> Hello world
```

2. funzioni scritte da noi

Dapprima la **definiamo**: *definire serve per riutilizzare*

```
def myage(x):
```

```
    print("La mia età è", x)
```

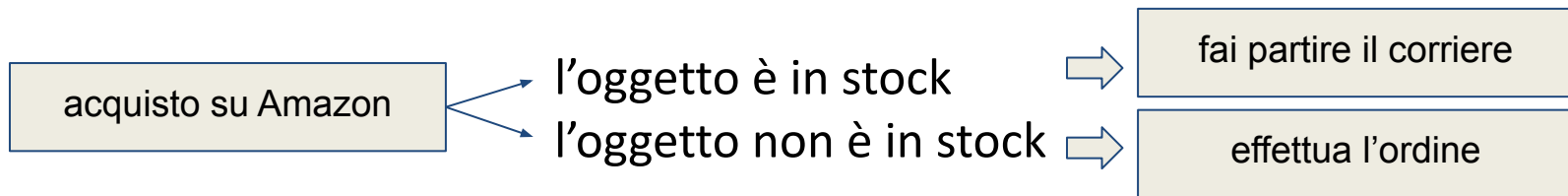
Dopo averla definita, la **appliciamo**:

```
myage(31)
```

```
>> La mia età è 31
```

I comandi condizionali

In molte situazioni può essere utile applicare dei comandi soltanto se si verificano determinate condizioni:



In Python, il tutto viene reso possibile dall'uso combinato di `if/else:`

```
if eta>18:
    maggiorenne = 1
else:
    maggiorenne = 0
print(maggiorenne)
>> 1
```

Modulo 1: Introduzione al coding

1. L'unità base: il dato
- 2. Un insieme di dati: il database**

Cos'è un database?

“...un insieme organizzato di dati.”

→ il principale strumento del **data scientist**

Esempi:

La macchina di Laura è nera

→ **NON** è un database

La macchina di Laura è nera

Le giraffe sono dei mammiferi

→ **NON** è un database

La temperatura di lunedì è stata di 15 °C

nera, rossa, grigia

→ **NON** è un database

La macchina di Laura è nera

La macchina di Tobia è rossa

La macchina di Elena è grigia

→ **è un database**

Gli elementi costitutivi di un database

Quali sono gli **elementi costitutivi** di un database?

- Le **osservazioni**: Laura, Tobia, Elena, etc. → *le unità di cui registriamo una o più informazioni*
- Le **variabili**: colore della macchina, modello della macchina, a quanti anni ha acquistato la macchina → *identificano le diverse informazioni che registriamo per ciascuna unità*

Le variabili possono essere di vari **tipi**:

- **“stringa”**: colore/modello della macchina
- **numeriche**: età al momento dell’acquisto della prima macchina
- ...

Righe e colonne di un database

Diagram illustrating the structure of a database table, showing the relationship between observations (rows) and variables (columns).

Variabili (Variables) are represented by the columns.

Osservazioni (Observations) are represented by the rows.

Mammals								
Mammals (27 cases)								
in- dex	Mammal	Order	LifeSpan (years)	Height (meters)	Mass (kg)	Sleep (hours)	Speed (km/h)	Habitat
1	African ...	Probosc...	70	4	6400	3	40	land
2	Asian El...	Probosc...	70	3	5000	4	40	land
3	Big Bro...	Chiropt...	19	0.1	0.02	20	40	land
4	Bottlen...	Cetacea	25	3.5	635	5	37	water
5	Cheetah	Carnivo...	14	1.5	50	12	110	land
6	Chimpa...	Primate	40	1.5	68	10		land
7	Domest...	Carnivo...	16	0.8	4.5	12	50	land
8	Donkey	Perisso...	40	1.2	187	3	50	land

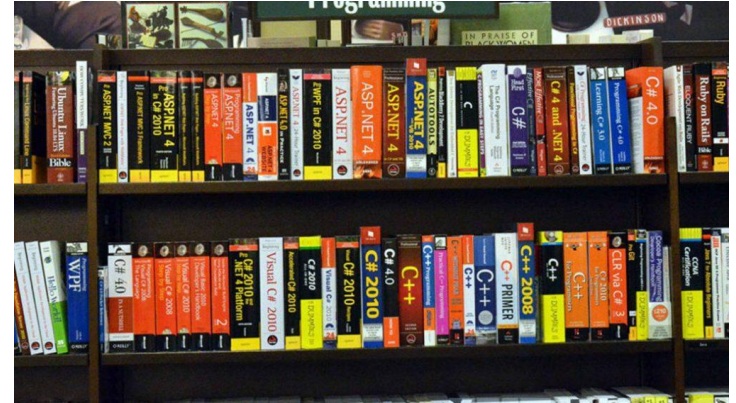
Tante osservazioni quante sono le **righe** di un database.

Tante variabili quante sono le **colonne**.

Database da esplorare: <https://codap.concord.org/app/static/dg/en/cert/index.html>

Le librerie

Nei linguaggi di programmazione cosiddetti ad “alto livello”, una **libreria** è un insieme di **funzioni** esterne finalizzate a un determinato tipo di analisi, che tu puoi prendere e utilizzare.



Quando scarichi il software che ti consente di iniziare usare un certo linguaggio di programmazione sul tuo computer, le librerie sono incluse.

L'importante è specificare all'inizio del foglio di comandi, il **codice**, quali librerie vuoi volta per volta analizzare!

Esempi:

- librerie che ti consentono di svolgere operazioni matematiche
- librerie che rendono più facile l'analisi di database

Un bel risparmio di tempo!
(anche se non sempre è possibile)



Al lavoro con il primo nostro database!



Checklist: cosa abbiamo imparato

1. Che cos'è un dato
2. Che cos'è la data science
3. Che cos'è un linguaggio di programmazione
4. Cosa sono le librerie
5. Cos'è un comando
6. Cos'è una variabile
7. Tipi di dati: stringhe, variabili numeriche, variabili logiche, liste, etc.
8. Cos'è una funzione
9. Comandi condizionali
10. Cos'è un database e i suoi elementi costitutivi: osservazioni (righe) e variabili (colonne) del database