



Materiale didattico per partecipante al corso **“TECNICO ESPERTO NELL’ANALISI E NELLA VISUALIZZAZIONE DEI DATI”** – Rif.P.A. 2021-15998/RER – approvata con DGR n. 1263 del 02/08/2021 di IFOA – Istituto Formazione Operatori Aziendali

# Definizione di dati

- I dati sono rappresentazioni originarie, cioè non interpretate, di un fenomeno, evento, o fatto, effettuate attraverso simboli o combinazioni di simboli, o di qualsiasi altra forma espressiva legate a un qualsiasi supporto
- Dati sono rappresentazioni di eventi o fatti
  - Non interpretate (originarie)
  - Attraverso simboli (o combinazioni di simboli)
  - Contenute su supporti (forma espressiva)

# Definizione di informazione

L'informazione deriva da un dato, o più verosimilmente da un insieme di dati, che sono stati sottoposti a un processo di interpretazione che li ha resi significativi per il destinatario

- L'informazione è:
  - insieme di dati
  - interpretati
  - comprensibili per il destinatario

Cosa e' il **MACHINE LEARNING** ?



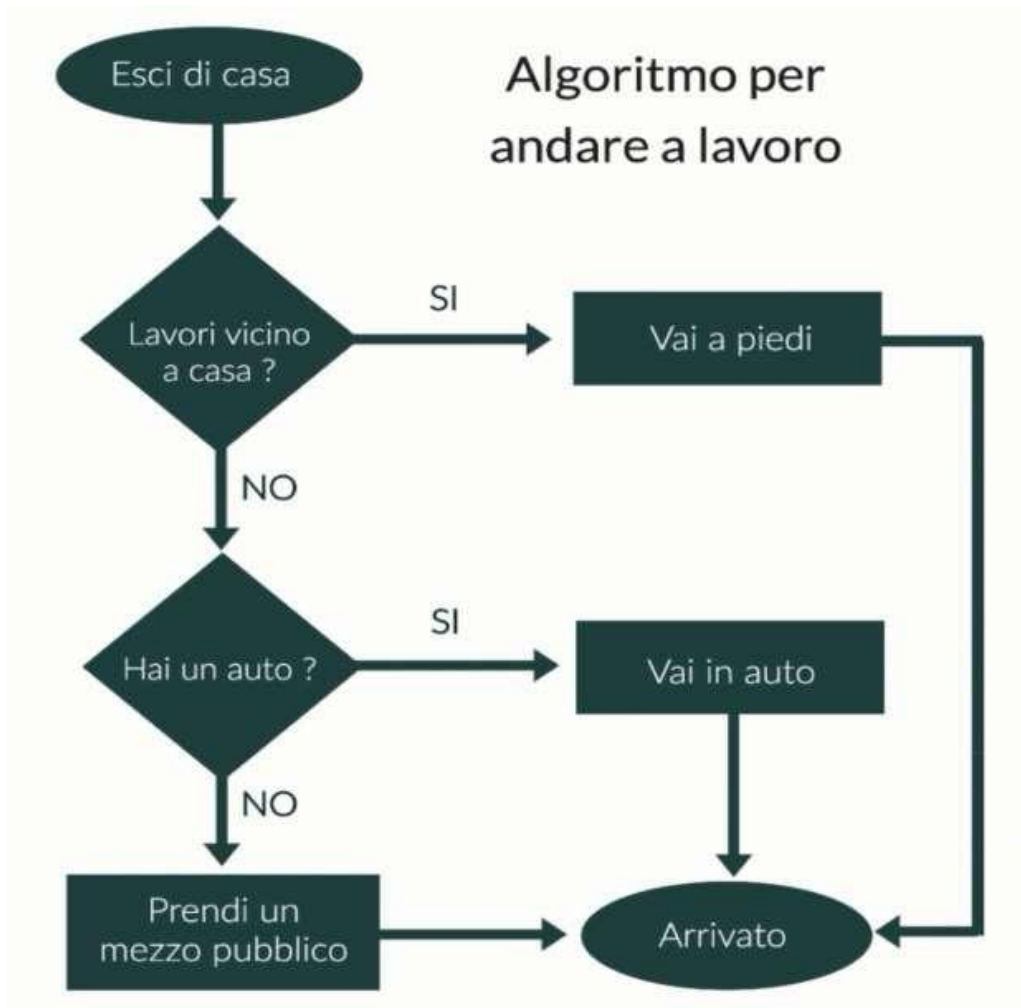
## Cosa e' il machine learning

"E' il settore dell'intelligenza artificiale che studia come dare ai computer l'abilità di imparare senza essere esplicitamente programmati" - 1959



**ARTHUR SAMUEL**

Pioniere dell'AI  
Inventore del termine "Machine learning"



E' un gatto 😊

## Programmazione classica

Hard coding delle regole



Dati di input



Risultato

## Machine learning

Input data



L'algoritmo impara le  
regole



Risultato

**Analisi di rischio**

**Recommender  
system**

**Elaborazione del  
linguaggio parlato**

**Riconoscimento di  
oggetti**

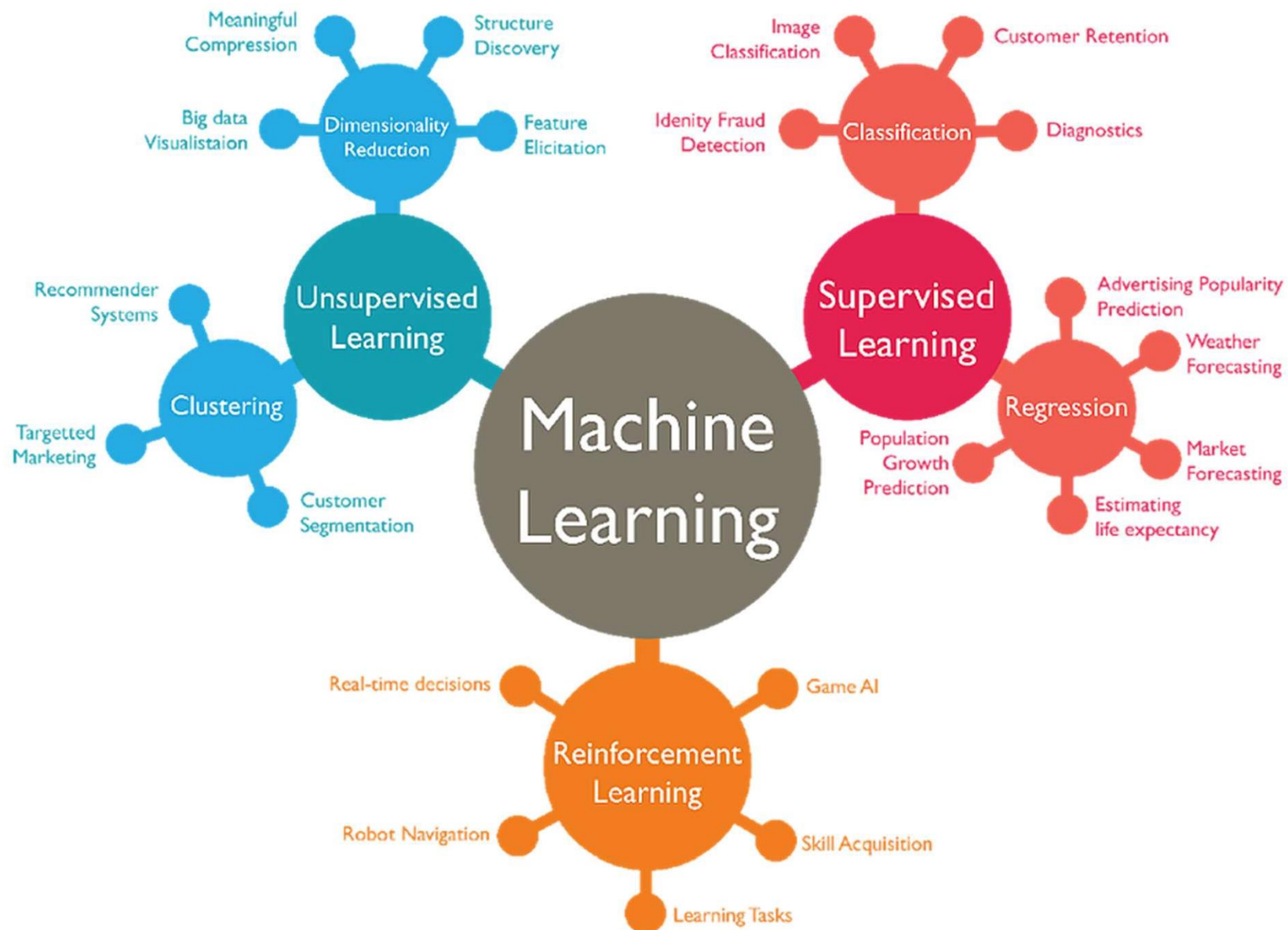
**Veicoli a guida  
autonoma**

**Fraud detection**

**Customer  
segmentation**

**Diagnosi mediche**





Il machine learning si basa su statistica e probabilità



92% gatto

8% tigre





**1997** - Deep Blue batte  
il campione di scacchi  
Garry Kasparov



**2008** - Watson vince a  
Jeopardy! contro i  
campioni Ken Jennings  
e Brad Rutter



**2016** - AlphaGo batte  
Go il 18 volte campio  
del mondo Lee Sedor

## PERCHE' OGGI ?



MAGGIORE POTENZA DI CALCOLO DISPONIBILE



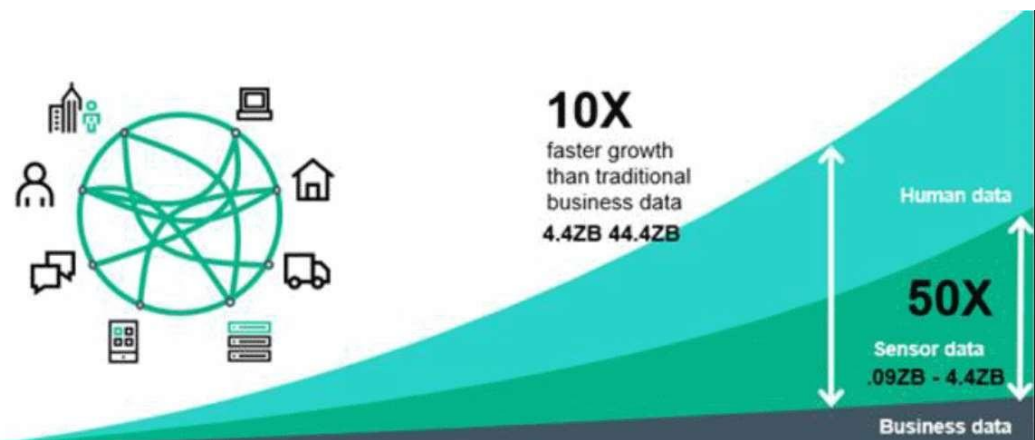
## DISPONIBILITA' DI ENORMI QUANTITA' DI DATI

# Perché lo si utilizza?

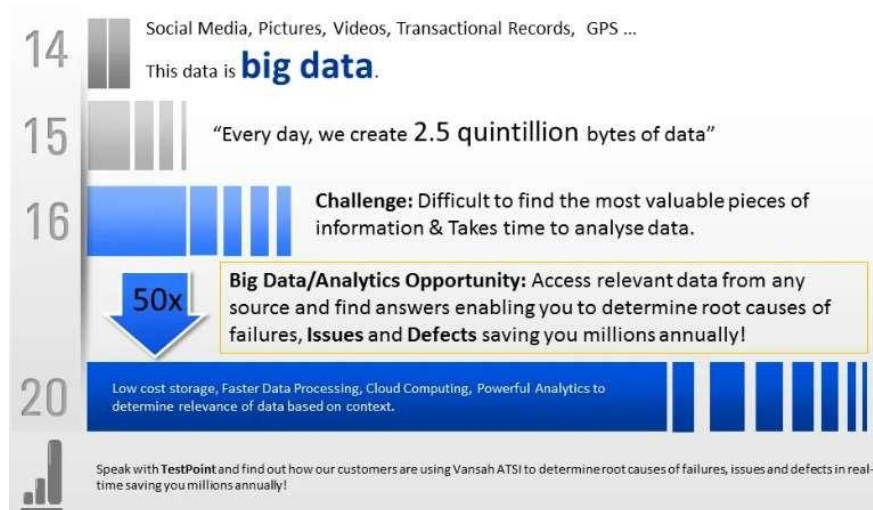
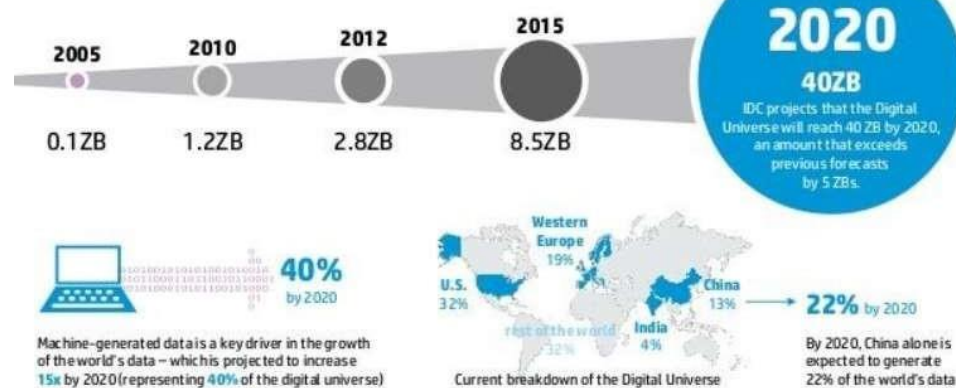
Il machine learning aiuta le aziende a dare un senso ai loro dati, indipendentemente dalla loro dimensione e dal settore di mercato in cui operano.

- Comprendere velocemente le informazioni
- Mostrare i fenomeni che determinano certi andamenti
- Evidenziare i trend emergenti
- Identificare relazioni e pattern nascosti
- Condividere le proprie scoperte con gli altri





## Machine Generated Data - 40% of the Digital Universe

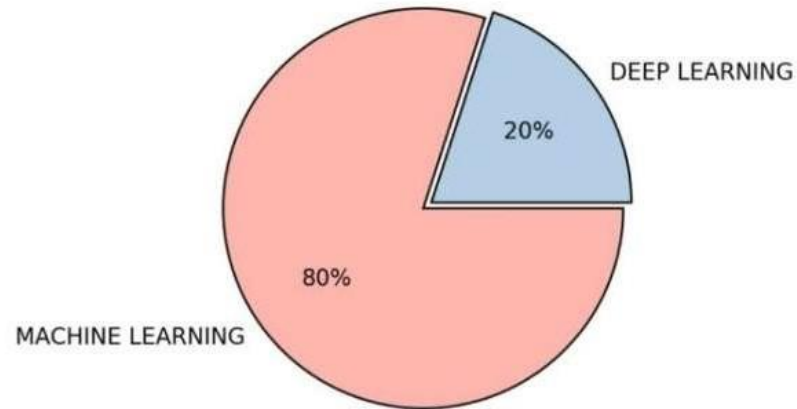




?

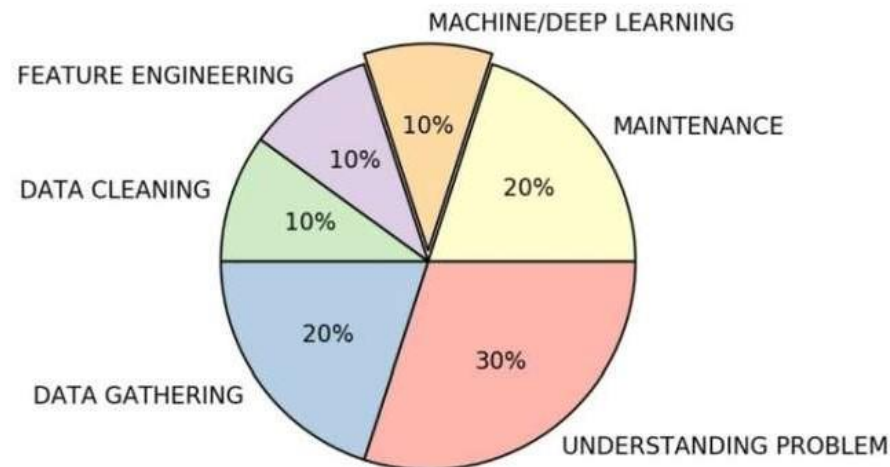
## DATA SCIENTIST JOB - EXPECTATION

@drangshu



Follow: Dr. Angshuman Ghosh

## DATA SCIENTIST JOB - REALITY







## Understanding the business problem

- You should ask relevant questions which makes you understand the problem which you are going to solve
- You should ask multiple WHY? questions and get answers from the client or the stakeholder or the person who told you to do the project.

# #2

## Data acquisition

- After deciding what features or metrics to use to solve the business problem.
- Next step is to gather the data.
- You may use sources like Databases, API's, Web scraper, online repositories etc...



## Data preparation

- This step involves 2 important things Data cleaning, Data transformation.
- Data cleaning is like check missing values, inconsistency datatypes, duplicate values etc..  
(Check our post on data pre-processing to see what are the most used techniques)
- Data transformation is a process of modifying the data based on predefined rules.

# #4

## Exploratory data analysis

- EDA helps you to understand what exactly you can do with the data.
- This is the most important step.
- Through EDA you can find what features are the most important in the model building.
- You can also find useful insights through EDA.



## **Data modeling**

- This is the most important part where you will be finding the model the best fits the business requirement.
- You will be doing multiple iterations on the test and train data to find the best performing model.

# #6

## Visualization and communication

- This is where you will show all the things which you did and found during the previous steps to your client, stakeholders or the person who gave you the project.
- You will be creating reports or dashboards to show your business finding in a powerful way (visualizations) to make them understand easily.



## Deploy & maintenance

- Test your best performing model multiple times before deploying it into production.
- You will be using reports and dashboards for real-time analytics.
- It is also important to monitor the model performance in the real world and retraining it if the performance degrades.

# Requirements:

- Comprendere i dati che desideri elaborare, incluse le loro dimensioni e cardinalità (l'unicità dei valori dei dati in una colonna).
- Determinare cosa stai processando e quale tipo di informazioni vuoi estrapolare.
- Conoscere la tua audience e comprendere come elabora l'informazione, o cosa si aspetta.
- Utilizzare una visualizzazione che trasmetta le informazioni nella maniera più appropriata e semplice per la tua audience.





**SHIT IN**

## **CAUTION: BAD DATA**



**BAD DATA QUALITY  
MAY RESULT IN  
FRUSTRATION AND  
LEAD TO DROP  
KICKING YOUR  
COMPUTER**

Evitare Errori Grossolani!





# WHAT IS PYTHON?



- ▶ A back end programming language
- ▶ High-level & approachable for beginners
- ▶ Has a welcoming & established community

## Used for tasks like:



## Used by companies like:

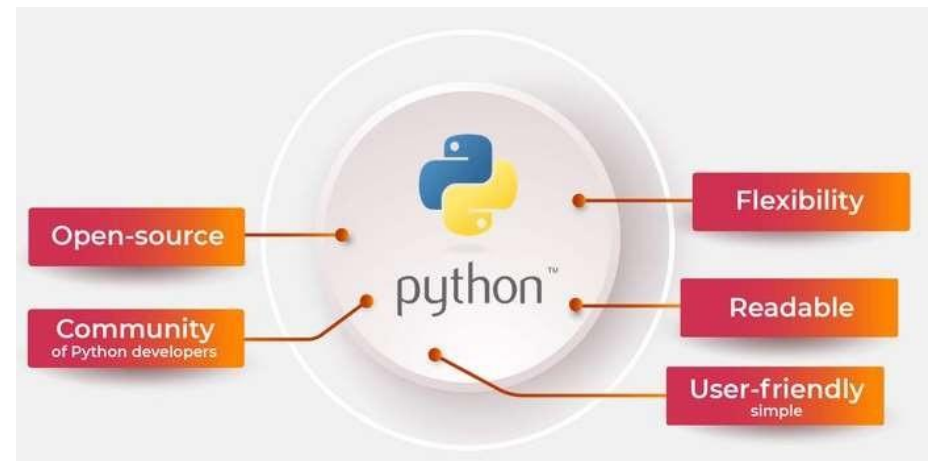


## Used with frameworks like:

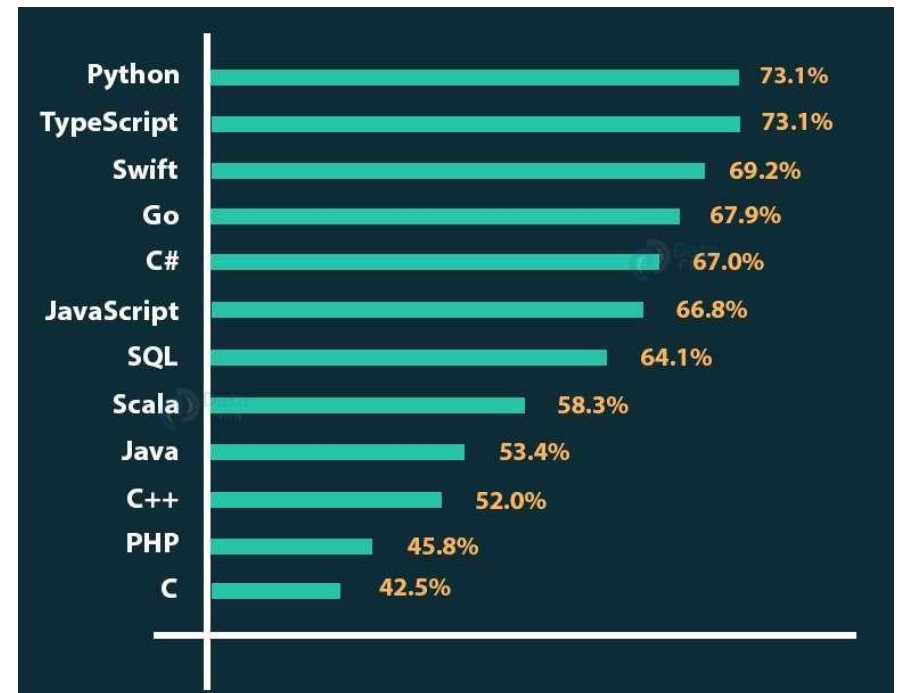
django



Flask



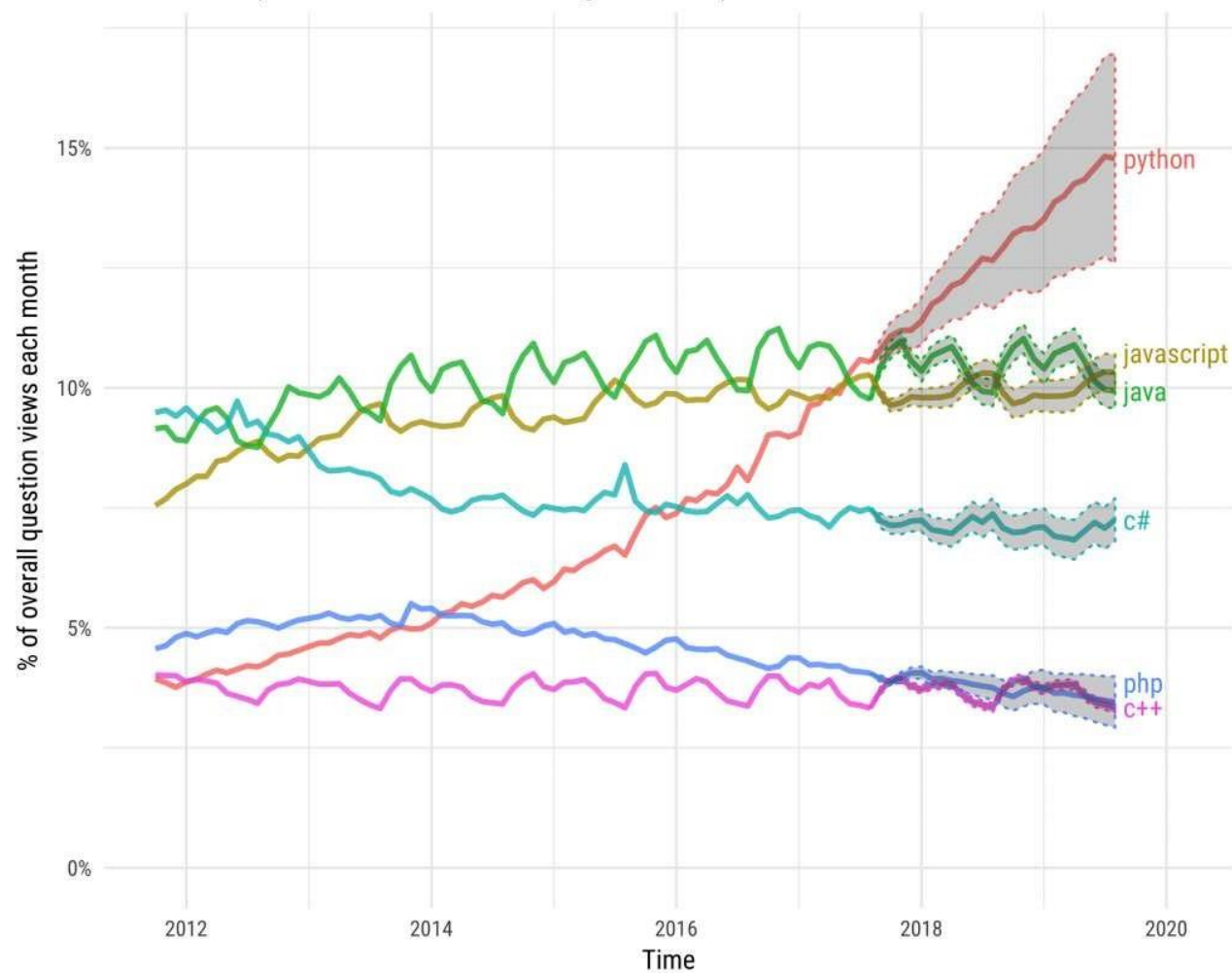
# PYTHON





## Projections of future traffic for major programming languages

Future traffic is predicted with an STL model, along with an 80% prediction interval.





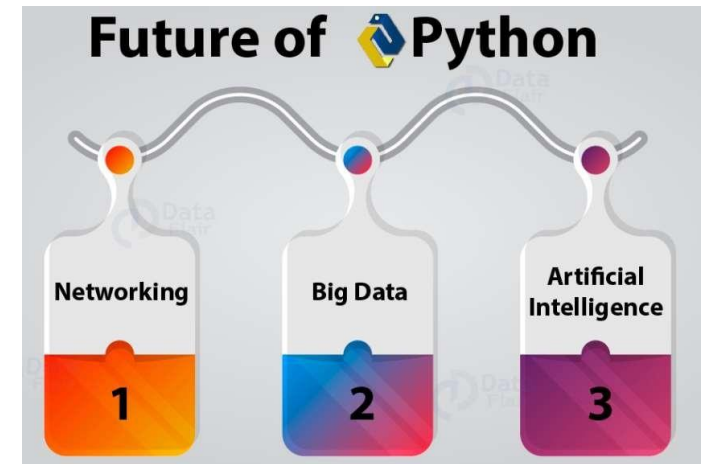
# Top Companies using Python





# WHY Python

01	Simplicity
02	Large Community
03	High Demand-Supply Ratio
04	Large Number of Frameworks
05	Chosen Language for AI and ML
06	Make your own DIYs



X	python
Lines of Code: 105	Lines of Code: 10
Wall Time: 7.89s	Wall Time: 0.83s
More complex	More simple



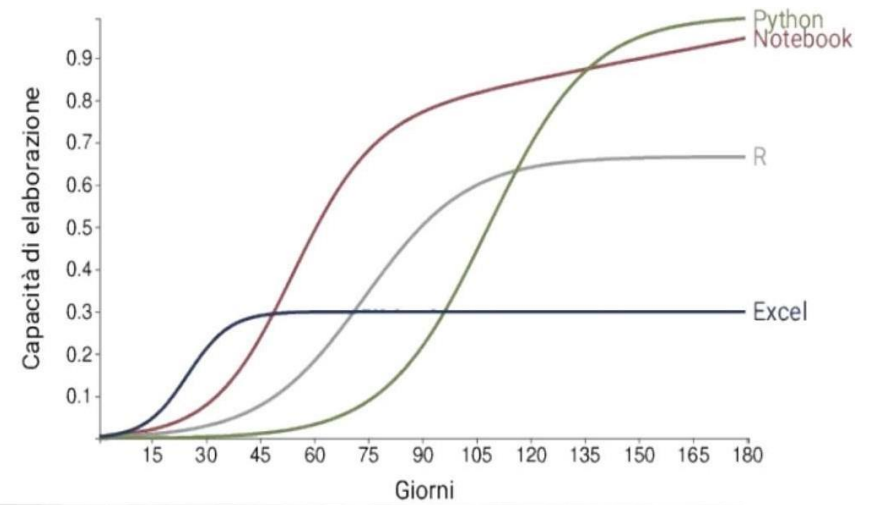


- Python is Easy
- More Functions – Less Code
- Perfect Language for Building Prototypes
- Great Flexibility
- Perfect Language in case you're on a Budget
- The Internet of Things (IoT) + Python Combination
- A Lot of useful Frameworks (Django, Flask)

"Python Vs Java!"	"Python Vs C++!"
<b>Python</b>	
Print ("Python Vs Java!")	Print ("Python Vs C++!")
<b>Java</b>	<b>C++</b>
<pre>public class PythonVsJava {     public static void main(String[] args)     {         System.out.println("Python Vs Java!");     } }</pre>	<pre>#include &lt;iostream&gt; void main() {     cout &lt;&lt; "Python Vs C++!"; }</pre>

**Gratificazione  
e  
soddisfazione**

**Noia e  
frustrazione**





Learn  
**programming**  
for **future work**



Learn  
**programming**  
to understand  
**programming jokes**