MUNER
MOTORVEHICLE
UNIVERSITY OF
EMILIA-ROMAGNA

Introduction to Python for data management and ML in automotive

September 5th, 2025 - Prof. Marchioni Alex
alex.marchioni@unibo.it

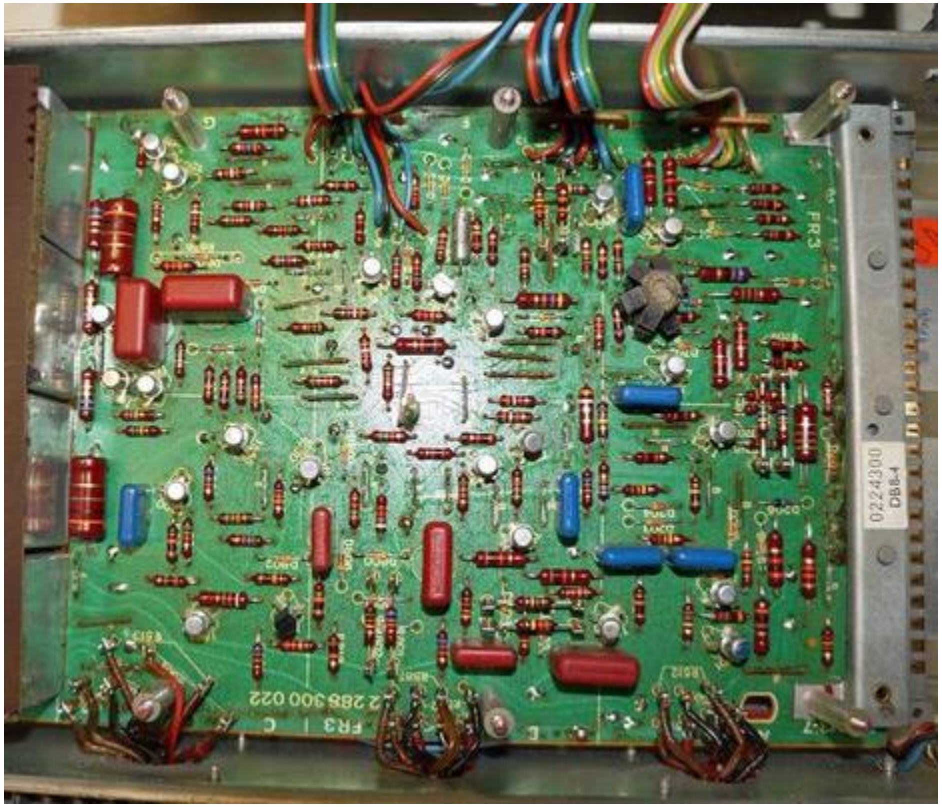
Why data processing?

The idea to include data processing manipulation in a car is not new

- ✓ A **carburetor** is an **analog computer** that adjust the fuel flow based on the intake air flow which is controlled by the throttle position, i.e., the wanted power.
- ✓ While the carburetor shown is fairly modern, similar principles date back to the origin of the internal combustion engine in the **1870s**.
- ✓ Unbelievably complex, a carburetor can be calibrated by means some **screws** or changing **interior parts** that regulate the flow of fuel or air.

Why data processing?

- ✓ This is a Bosch D-Jetronic (1969) used through the **70s**.
- ✓ This electronic circuit is a purely **analog computer** that takes a bunch of analog inputs such as heat, speed, throttle position, fuel pressure, to compute when and how long to open fuel injectors.
- ✓ It can be adjusted by varying the **variable resistors** on the underside of the board.



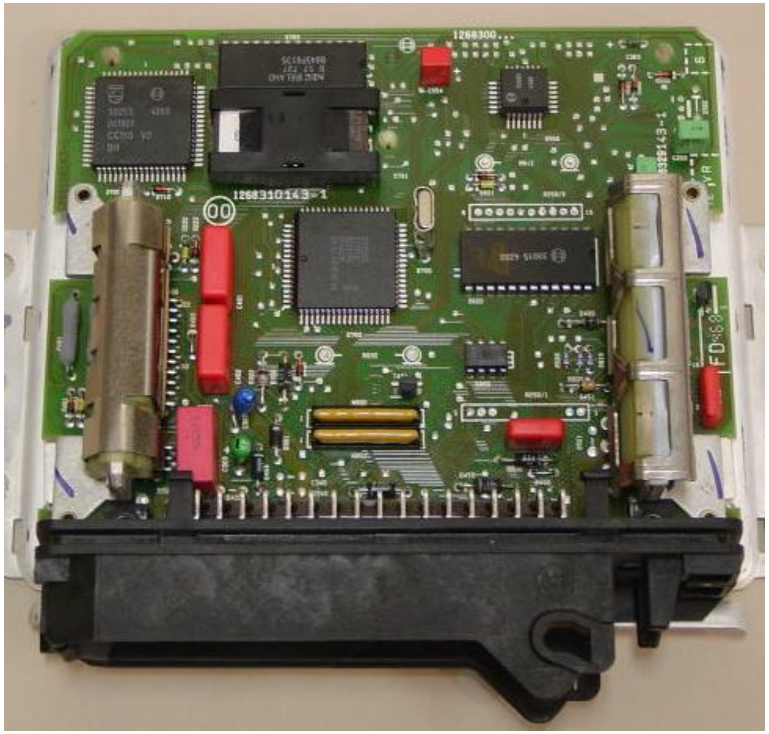
MOTORVEHICLE
UNIVERSITY OF
EMILIA-ROMAGNA

Why data processing?

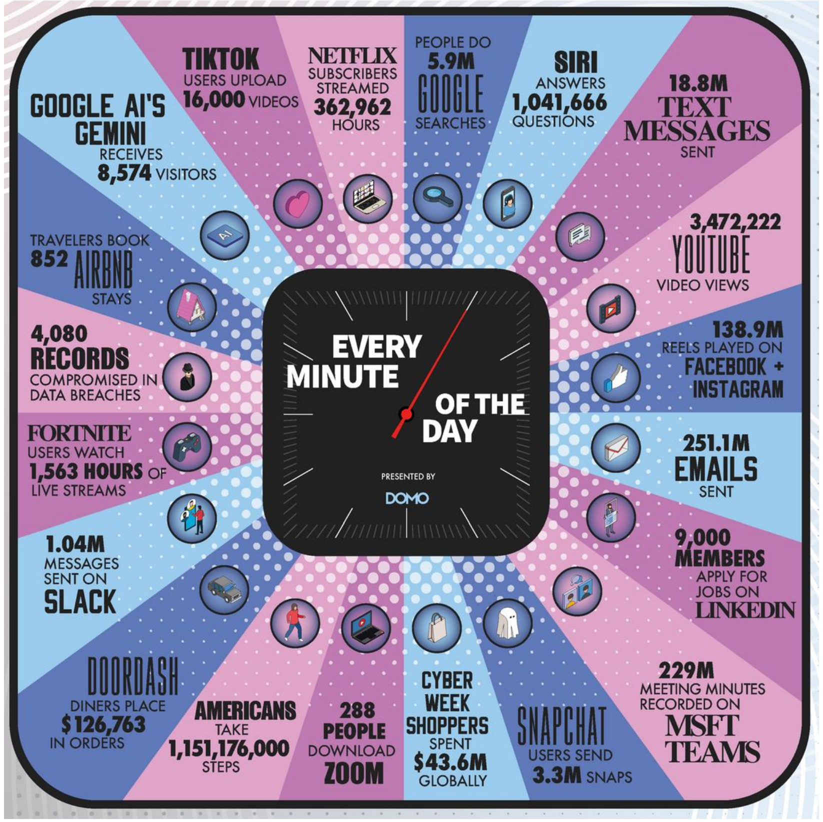
✓ This is a Bosch DME ECU (1984).

✓ This electronic circuit is a full **digital computer** that can execute programs stored in memory.
Not only it controls injection based on measurements of engine related quantities, but it also control ignition thus managing the entire combustion process.

✓ How the combustion process is controlled is determined by the program and its **parameters stored in memory**.



How much data?


TIKTOK USERS UPLOAD 16,000 VIDEOS
NETFLIX SUBSCRIBERS STREAMED 362,962 HOURS
PEOPLE DO 5.9M GOOGLE SEARCHES
SIRI ANSWERS 1,041,666 QUESTIONS
18.8M TEXT MESSAGES SENT
GOOGLE AI'S GEMINI RECEIVES 8,574 VISITORS
3,472,222 YOUTUBE VIDEO VIEWS
TRAVELERS BOOK 852 AIRBNB STAYS
138.9M REELS PLAYED ON FACEBOOK + INSTAGRAM
4,080 RECORDS COMPROMISED IN DATA BREACHES
EVERY MINUTE OF THE DAY
PRESENTED BY DOMO
FORTNITE USERS WATCH 1,563 HOURS OF LIVE STREAMS
251.1M EMAILS SENT
1.04M MESSAGES SENT ON SLACK
9,000 MEMBERS APPLY FOR JOBS ON LINKEDIN
DOORDASH DINERS PLACE $126,763 IN ORDERS
AMERICANS TAKE 1,151,176,000 STEPS
288 PEOPLE DOWNLOAD ZOOM
CYBER WEEK SHOPPERS SPENT $43.6M GLOBALLY
SNAPCHAT USERS SEND 3.3M SNAPS
229M MEETING MINUTES RECORDED ON MSFT TEAMS

domo.com/learn/infographic/data-never-sleeps-12

F1 car telemetry:

- ✓ 500/1000 sensors acquired with a rate of 100-10k sample per second.
- ✓ ~100 Mbps
- ✓ ~1 TB for each race

Outline

- ✓ Machine Learning
- ✓ Python
- ✓ Hands on
  - ✓ Intro to Python
  - ✓ Data management
  - ✓ Data visualization
  - ✓ Intro to ML

Machine Learning

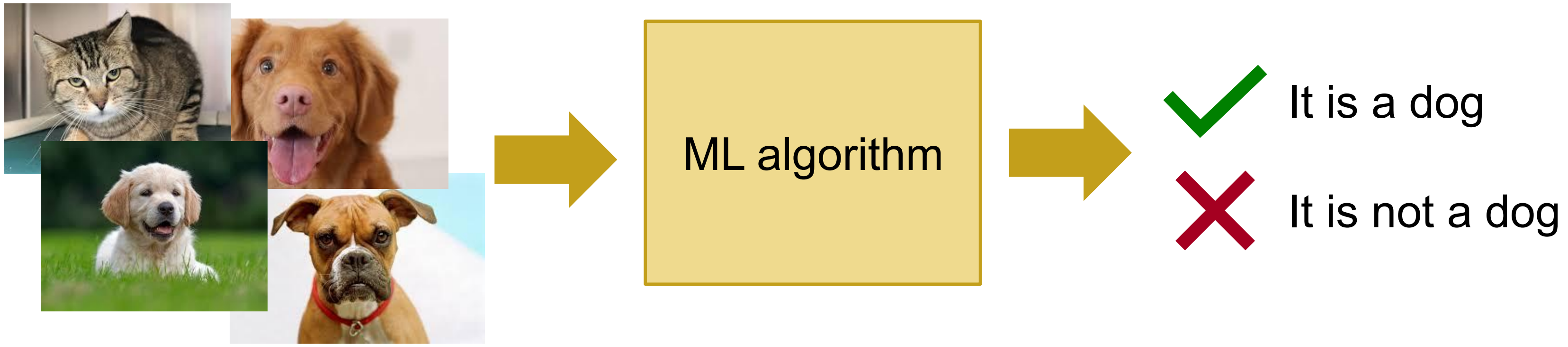
Machine Learning

A machine learning algorithm is an algorithm that can learn from data.

What do we mean by learning?

*“A computer program is said to learn from **experience** E with respect to some class of **tasks** T and **performance measure** P, if its performance at tasks in T , as measured by P , improves with experience E.”*

T. Mitchell, “Machine Learning”, McGraw-Hill, 1997



ML algorithm

✓ It is a dog

✗ It is not a dog

The task

Nearly all task that a machine learning algorithm can perform fall into one of the two category:

- ✓ **Classification**: specifying which of k categories some input belongs to.
  - ✓ Object Recognition in images. (application in medicine, industry, etc.)
  - ✓ Anomaly Detection is about classifying an input normal or abnormal.
- ✓ **Regression**: infer the relationship between input and output to predict a numerical value given some input.
  - ✓ applications in Finance forecasting, Trend Analysis, Marketing, Drug Response modeling, etc.
- ✓ **Synthesis and Sampling**: generating new examples that are similar to the ones that compose the machine's experience:
  - ✓ Generating text or images starting from a prompt (e.g., ChatGPT, DALL-E), replicating the style of a painter, etc.

The Experience

For a Machine, the Experience is the data set.

- ✓ In **Unsupervised Learning** the data set is a collection of examples with no additional information.
  - ✓ Finding structures and patterns in the data for tasks as density estimation, denoising, clustering or synthesis.
- ✓ In **Supervised Learning**, each example is paired with a label that is the correct answer the machine must learn.
  - ✓ Classification and Regression are supervised learning tasks.

This distinction that seems quite straightforward is often very blurred as many methods can fall in both categories or none of them. For example, **Reinforcement Learning** refers to algorithms that continuously learn by interacting with the environment.

The Performance Measure

A machine needs a **quantitative measure** to assess how well it is performing the task.

- ✓ It is often task specific.
  - ✓ For instance, Accuracy is ok for classification, but it does not fit regression.

In general, the dataset is split in two part:

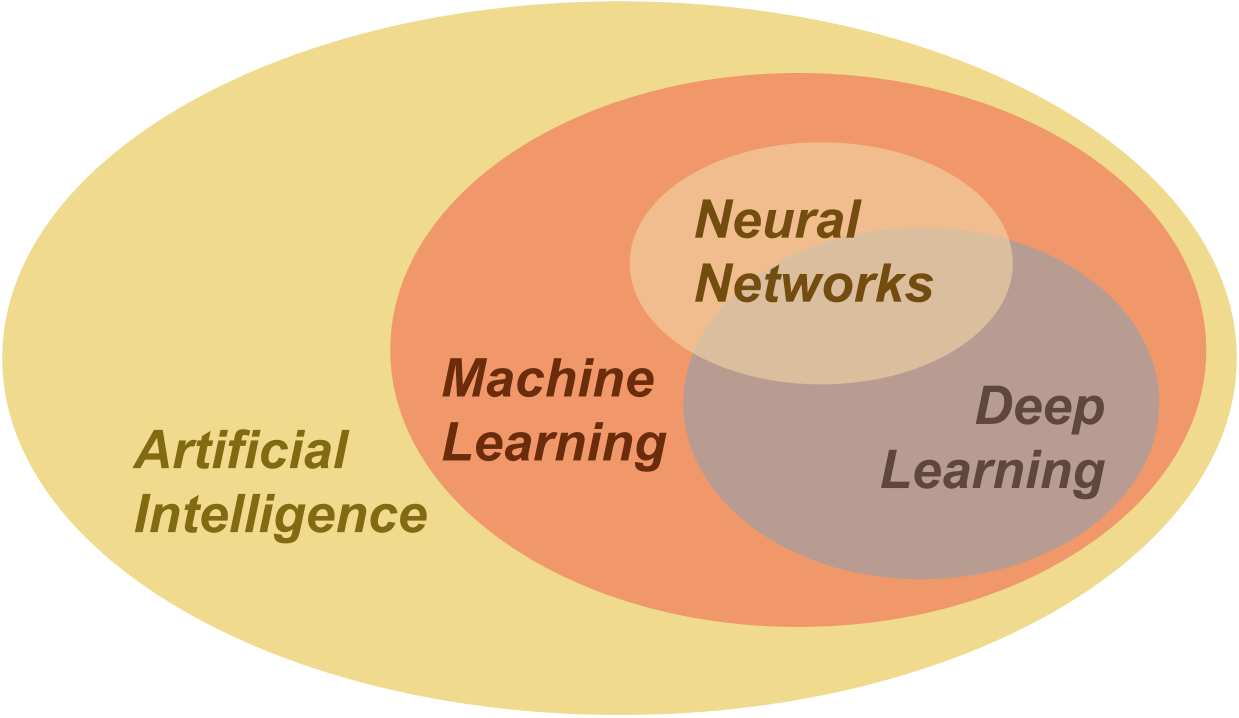
- ✓ **Training set**: used for learning (i.e., the experience)
- ✓ **Test set**: used for model evaluation. For a fair assessment, the performance must be measured on examples that are not part of the experience (a.k.a. training set).

Sometimes, there is the necessity for a **validation set** used to tune some hyperparameters – parameters that cannot be trained with the model.

Terminology

- ✓ **Artificial Intelligence (AI)**: is a broad concept referring to systems that can perform tasks that typically require human intelligence. Often, AI requires to reverse-engineer human abilities to emulate on machines.
- ✓ **Machine Learning (ML)**: main subset of AI regarding algorithms that can learn by themselves by observing examples (data).
- ✓ **Neural Networks (NN)**: Specific class of ML algorithm in which the idea consists in miming the functioning of biological neural networks of animals' brain.
- ✓ **Deep Learning (DL)**: Traditionally, data needed to be pre-processed before feeding a ML algorithm. Deep Learning refer to extremely complex ML methods that can directly process data as it is generated.

Terminology



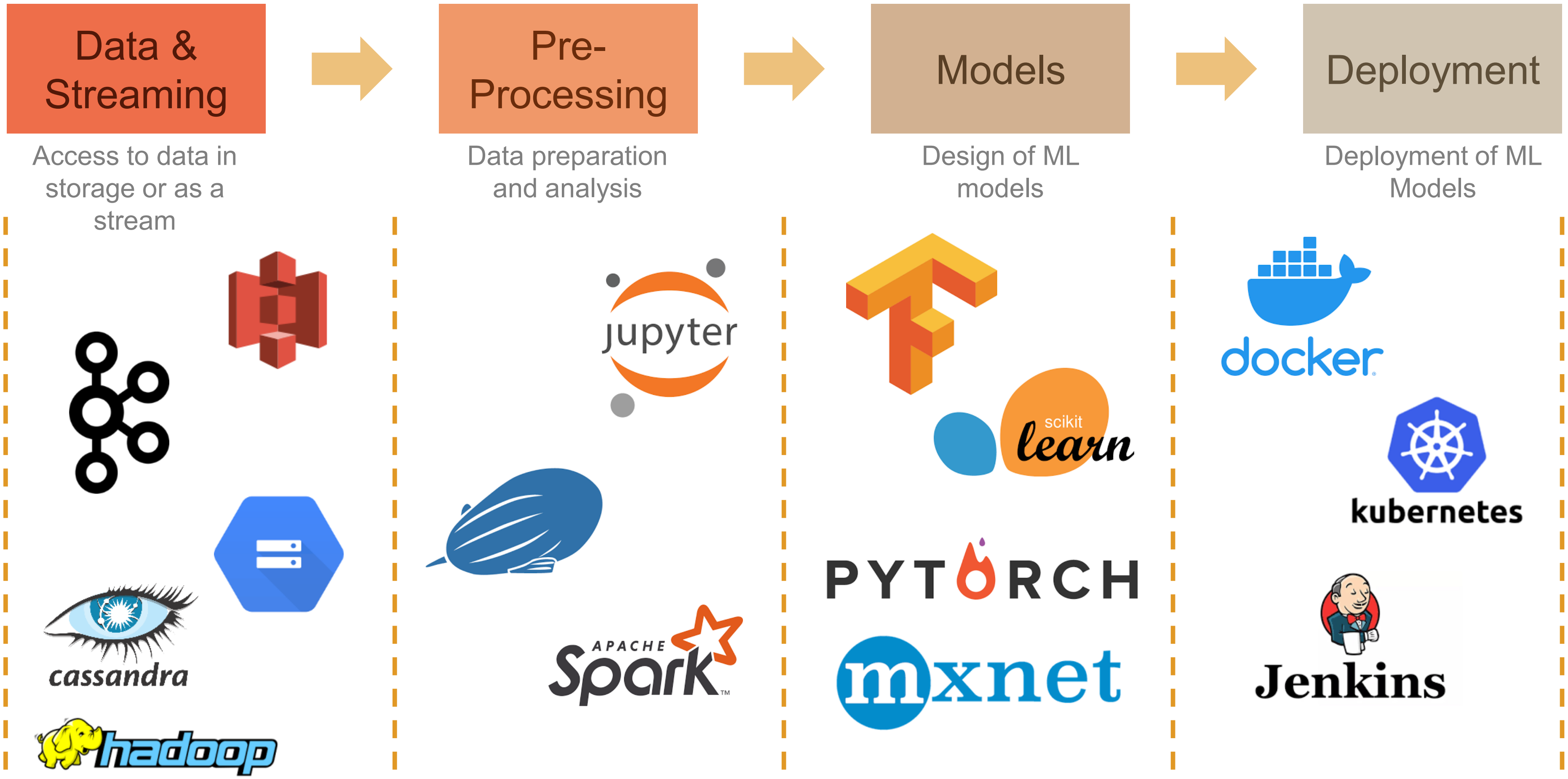
ML in Automotive

- ✓ **Design**
  - ✓ Acceleration of Rendering
  - ✓ Acceleration of FEM simulations
  - ✓ Generation of alternative solutions
- ✓ **Manufacturing**
  - ✓ Schedule/process optimization
  - ✓ Identification of defects in components
  - ✓ Plant Predictive maintenance

ML in Automotive

- ✓ **Supply Chain**
  - ✓ forecasting and replenishment optimization
  - ✓ Supplier ranking
- ✓ **Quality Control**
  - ✓ Automatic assessment of the quality of produced parts and ensembles
  - ✓ Automatic defect detection, classification and prediction
- ✓ **Driver Behavior**
  - ✓ Advanced Driver Assistance Systems (ADAS)
  - ✓ Driver Attention Alert

Machine Learning Pipeline



Data & Streaming → Pre-Processing → Models → Deployment

Data & Streaming: Access to data in storage or as a stream

Pre-Processing: Data preparation and analysis

Models: Design of ML models

Deployment: Deployment of ML Models

Python

Why Python

Python is an Interpreted and Object-Oriented Programming Language.

WHY Python?

✓Simple syntax

✓Very flexible

✓Highly extensible

✓Cross-platform

✓Open-source with a huge community

Google says: *Python where we can, C++ where we must*

Applications

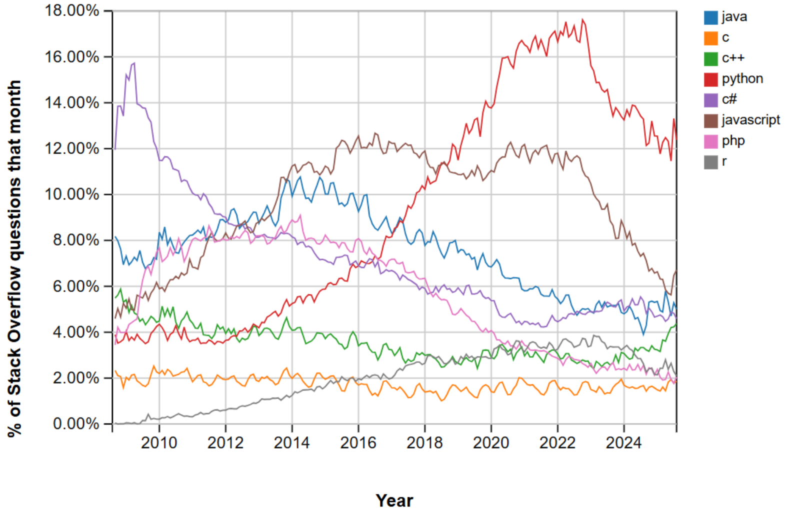
- ✓ Web and Internet Development
- ✓ Scientific and Numeric
- ✓ Education
- ✓ Desktop GUIs
- ✓ Software Development
- ✓ Business Applications

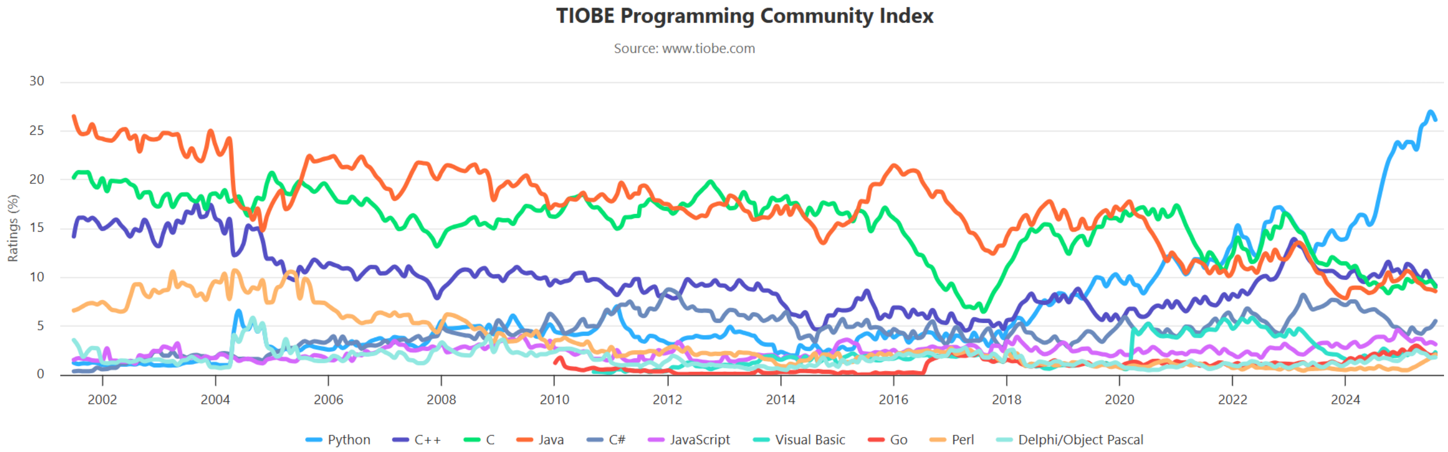
Basically anything, **like English for spoken languages**

Popularity

Python is the **most popular** (or second most) programming language over several popularity indexes (Stack Overflow, GitHub, RedMonk, TIOBE)

It is having a further boost nowadays as it is the **most used for AI** development and the **most generated by AI** tools.

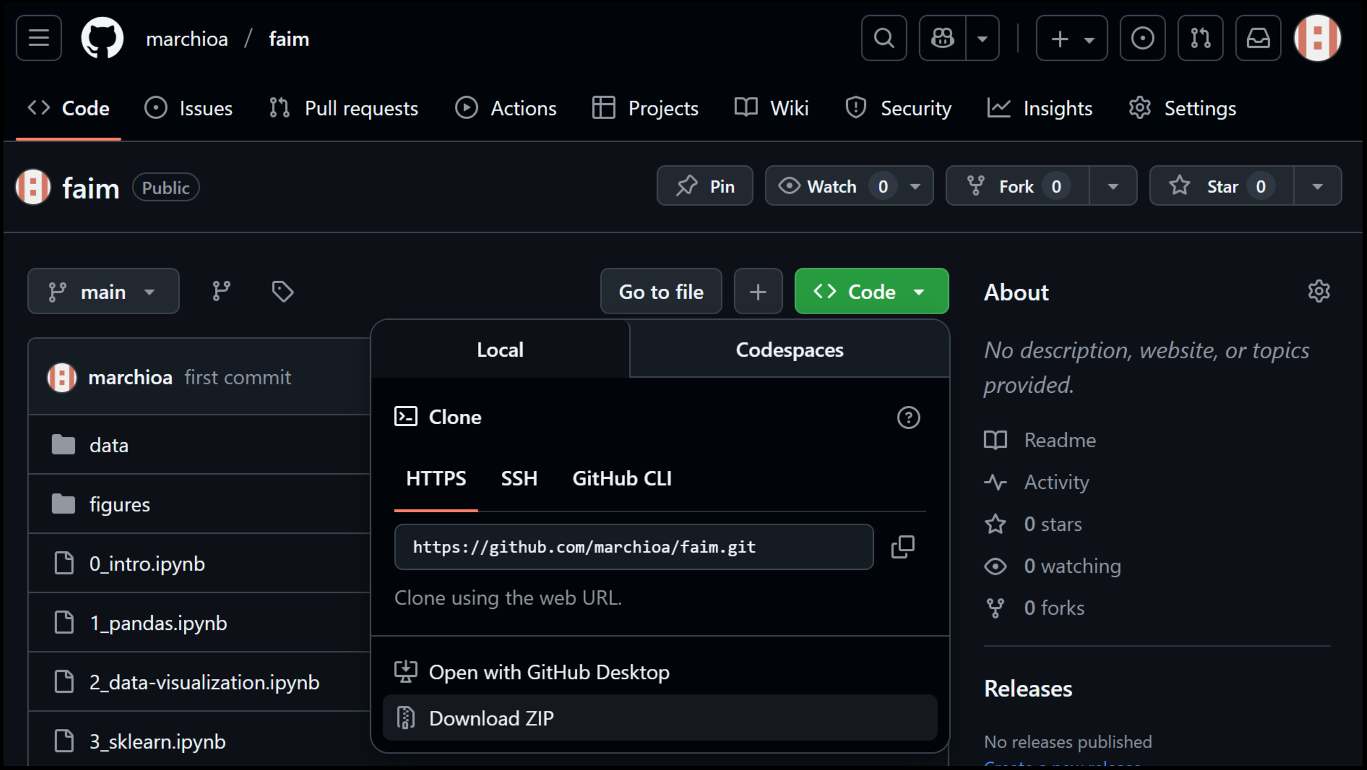




Framework set up

GitHub Repository

- ✓ Download (or clone) the repository in github.com/marchioa/faim



marchioa / faim

marchioa first commit

- data
- figures
- 0_intro.ipynb
- 1_pandas.ipynb
- 2_data-visualization.ipynb
- 3_sklearn.ipynb

```
https://github.com/marchioa/faim.git
```

Clone using the web URL.

No description, website, or topics provided.

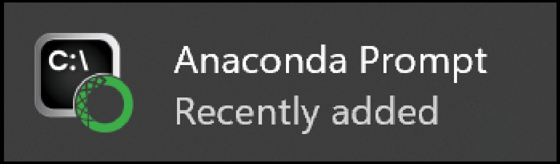
No releases published

Conda

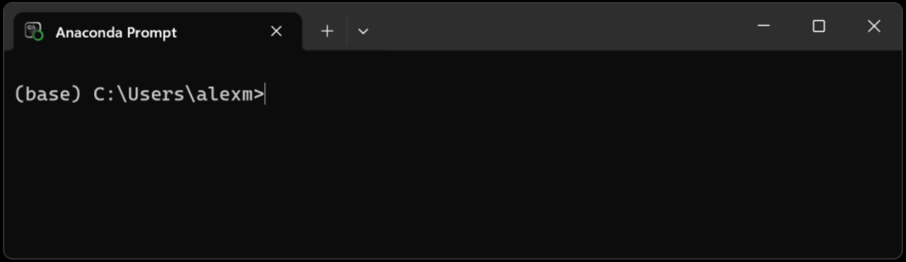
- ✓ Download the latest version of **miniconda** from here anaconda.com/download/success

- ✓ Open a conda prompt
  - ✓ On either Windows or macOS, find the **Anaconda Prompt** in the list of your programs and open it.
  - ✓ On Linux, open a **Terminal** and run the following command:

```
conda activate base
```



Anaconda Prompt
Recently added



```
(base) C:\Users\alexm>
```

Creating an environment

✓ Now we create an **environment** by running the command

```
conda create --name faim
```

✓ Once created, we need to **activate** the environment with the command

```
conda activate faim
```

✓ Before installing package let us add **conda-forge** as download.

```
conda config --add channels conda-forge
```

Installing Python libraries

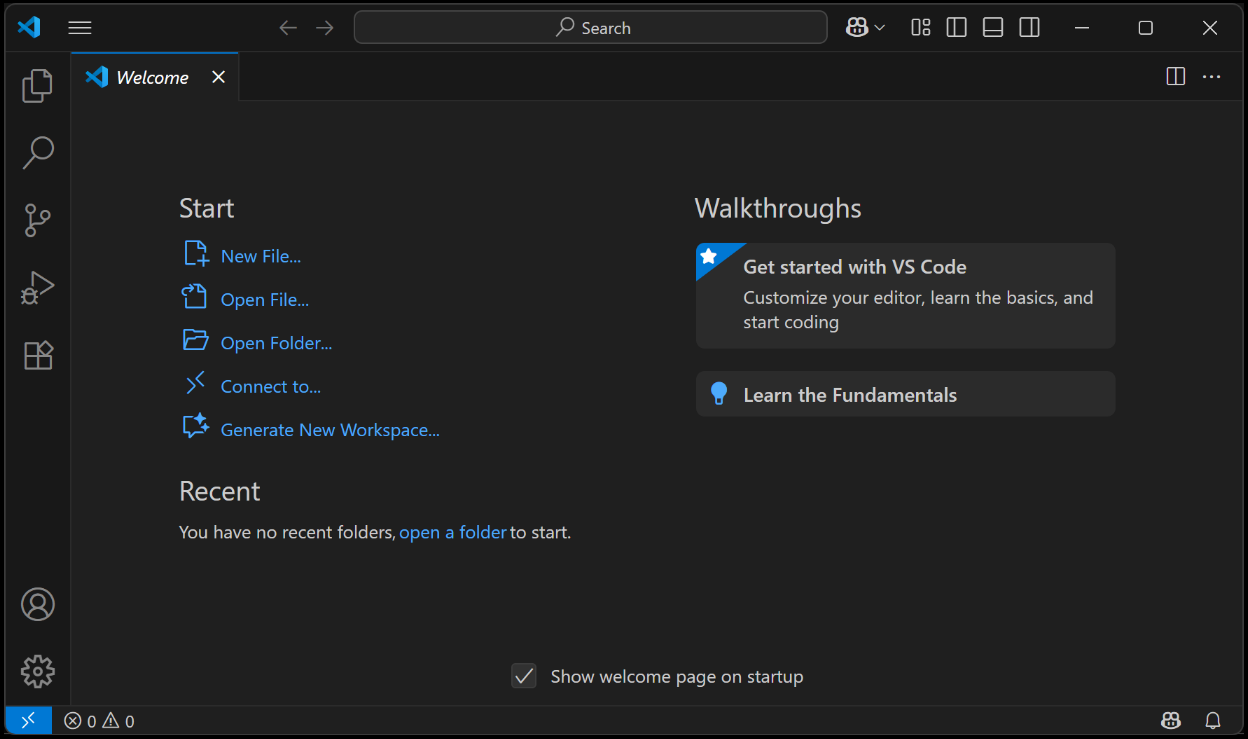
✓ In the `faim` environment, we can install some packages with the following commands:

```
conda install scipy
conda install matplotlib
conda install pandas
conda install seaborn
conda install folium
conda install scikit-learn
conda install ipykernel
conda install ipympl
```

It will take a while depending on your machine and internet connection.

Installing VSCode

- ✓ Download and install the latest version of **Visual Studio Code**
  code.visualstudio.com/download



Welcome

Start

New File...

Open File...

Open Folder...

Connect to...

Generate New Workspace...

Recent

You have no recent folders, open a folder to start.

Walkthroughs

Get started with VS Code

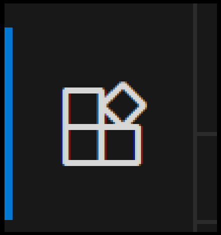
Customize your editor, learn the basics, and start coding

Learn the Fundamentals

☑ Show welcome page on startup

Install Extensions

✓ On the left toolbar, click the Extension Icon

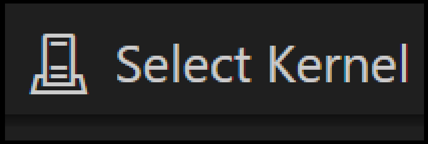


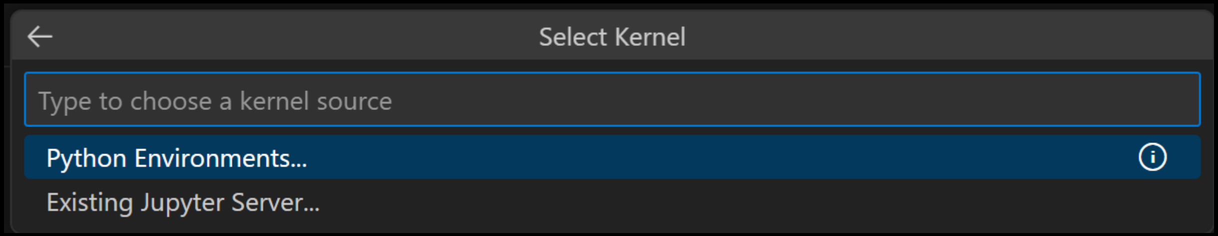
✓ Search and Install the following extensions:

- ✓ Python
- ✓ Jupyter

Open and Run a Notebook

- ✓ Open folder
- ✓ Select file `0_intro.ipynb`
- ✓ On the upper right corner click on “Select Kernel”
- ✓ Select the kernel in `faim` conda environment

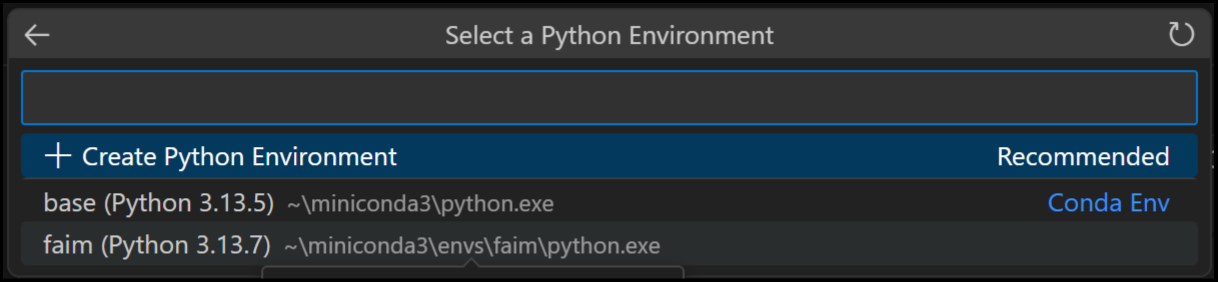



Select Kernel

Type to choose a kernel source

Python Environments...

Existing Jupyter Server...


Select a Python Environment

+ Create Python Environment Recommended

base (Python 3.13.5) ~\miniconda3\python.exe Conda Env

faim (Python 3.13.7) ~\miniconda3\envs\faim\python.exe

Let's start!