

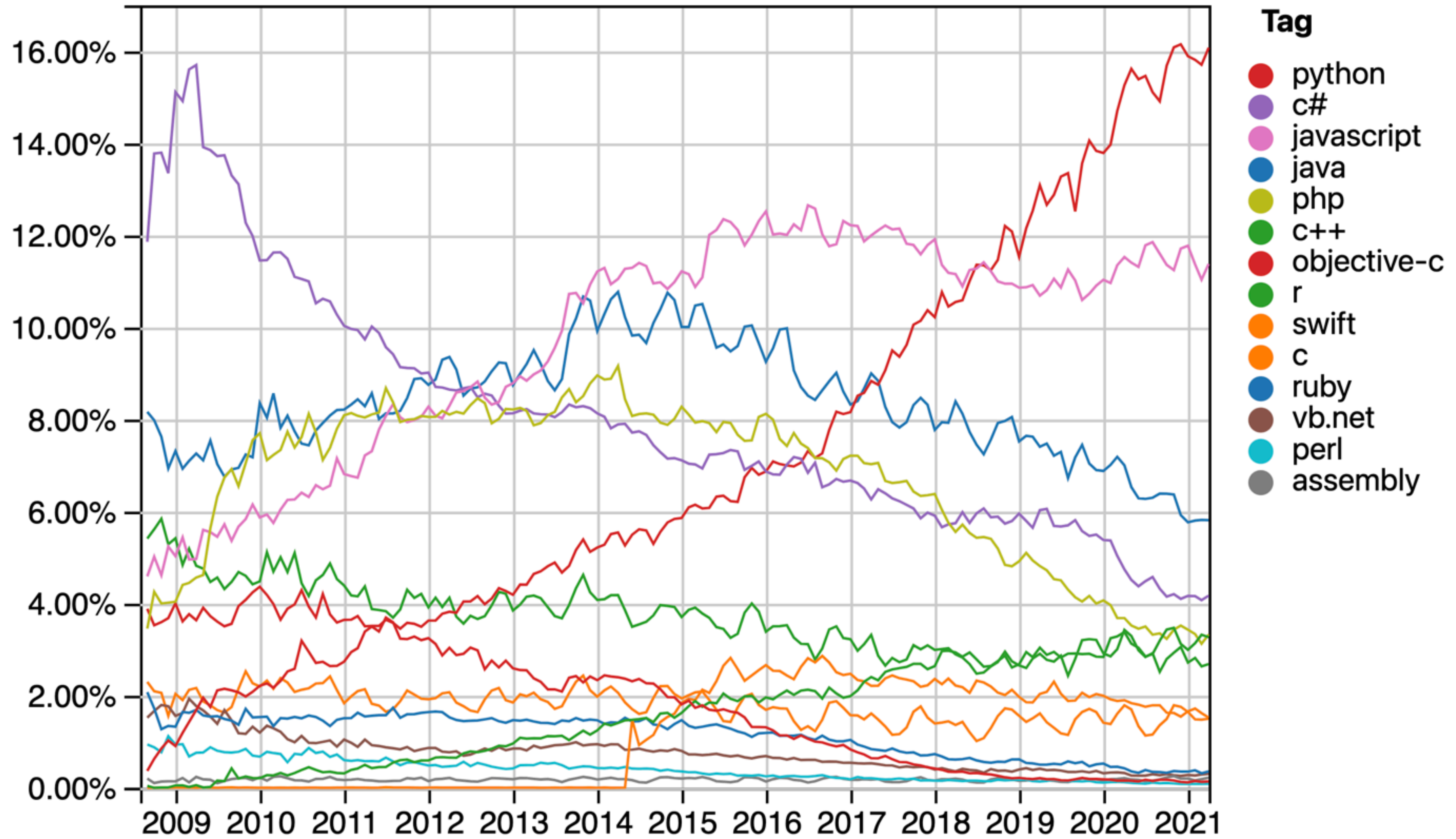
Hands-on Machine Learning

Faire de la Data Science en 2021

Martin Vielvoye 2021

Python et Jupyter Notebook

% of Stack Overflow questions that month



Pourquoi Python?

- Langage de programmation léger. Installation et maintenance facile.
- Syntax permissive et accessible. Rapide à apprendre. Permet de développer rapidement des applications.
- Langage Open-Source et une communauté très forte et présente.
- Énormément de bibliothèques et d'outils orientés data, data science, mathématiques et visualisation.
- Une communauté Scientifique très forte.
- Permet de manipuler les fichiers CSV et JSON facilement.
- Jupyter Lab et Notebook

Librairies Python

2021

- Pandas : Manipulation de tableaux et de données.
- NumPy : Manipulation 'scientifique' d'objets multidimensionnelles.
- Matplotlib : Outil de visualisations, de graphs et autres représentation graphique.
- SciKit-Learn : Possède beaucoup de modèles d'IA "prêt-à-utiliser" et de fonctions mathématiques, statistiques, complexes, etc.
- TensorFlow : Librairies de Machine Learning open-source.
Keras
PyTorch



**Jupyter
Notebook**

**Jupyter
Lab**

jupyter notebook

En local :

- Jupyter Notebook et Jupyter Lab.
- Possibilité de convertir les fichiers *.py* en *.ipynb*.
- Extension *notebook* dans VSCode et autres IDE.

En ligne :

- Jupyter Notebook et Google Colab :
 - Allocation de ressources et matériels accélérés : CPU, GPU et TPU.
 - Compatibilité avec google drive.
- Possibilité d'entraîner sur AWS, Azure et autres cloud-provider.

Démonstration



<https://colab.research.google.com/notebooks/welcome.ipynb?hl=fr>

Exemple de transformation de la donnée

Feature Scaling et Mean normalisation

Feature Scaling :

On divise chaque valeur par son écart (*valeur maximal - valeur minimal*).

-> On force toutes les valeurs a avoir un écart de 1.

Mean Normalisation :

On soustrait a chaque valeurs, la moyenne de toutes les valeurs.

-> La nouvelle moyenne est de 0.

$$x_i := \frac{x_i - \mu_i}{s_i}$$

μ_i : moyenne des valeurs x s_i : l'écart des valeurs ($\max(x) - \min(x)$)
--