









Manipulation de données

Introduction

Xavier Gendre 

Plan de la formation

- Mise en place des outils de travail 
- Réseau et gestion de versions avec Git 
- Manipulation de données élémentaires 
- Bases de données relationnelles (SQL) 
- Format JSON et requêtes sur API 
- Web scraping 
- MongoDB, une base de données NoSQL 
- Atelier 

Onyxia

Onyxia est le datalab du Genes (Groupe des Écoles Nationales d'Économie et Statistique)

Tous les services utiles y sont disponibles :

- Jupyter
- VSCode
- MongoDB
- ...



Mise en place de l'environnement de travail :

1. Se connecter sur [Onyxia](#).
2. Visiter le [Catalogue de services](#).
3. Lancer le service **Vscode-python** (ou un autre *Service interactif*) avec les options par défaut.
4. Le navigateur redirige vers [Mes services](#).
5. Attendre la fin du processus de création du service.
6. Copier le mot de passe et ouvrir le service.

Le datalab offre également d'autres fonctionnalités :

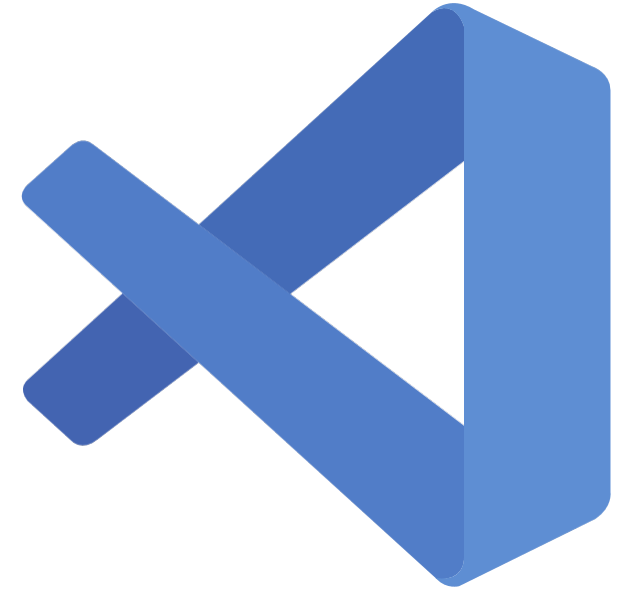
- lien avec des comptes Gitlab, Github et Kaggle,
- stockage de secrets,
- stockage de fichiers,
- exploration de fichiers Parquet et CSV,
- ...

Suppression des services

Les *services interactifs* (VSCode, Rstudio et Jupyter) de plus de 15 jours sont supprimés automatiquement le week-end. Les codes et les données doivent être sauvegardés pour être conservés (dans l'espace de stockage de fichiers, par exemple).

VSCode

VSCode (*Visual Studio Code*) est un environnement de développement développé par Microsoft et publié sous une double licence MIT et propriétaire.



Il propose de nombreux outils utiles pour développer du code dans plusieurs langages (coloration syntaxique, débogage, intégration de Git, ...).

Les fonctionnalités de VSCode peuvent être enrichies par des *extensions*. Plusieurs extensions sont déjà installées dans le service **Vscode-python** de Onyxia et l'environnement Python dispose des principaux modules utiles en Data Science.

Pour installer une extension, un [Marketplace](#) est disponible et accessible depuis la barre latérale dans VSCode.



Exemple : **Zip Tools**

Utile pour décompresser les supports de la formation et compresser les fichiers à sauvegarder 😊

Un **notebook** est une interface de programmation interactive qui contient des sections textuelles et des cellules de code informatique. Ce format n'est pas idéal en production mais il s'avère particulièrement bien adapté pour la phase de recherche et pour l'enseignement.

Les notebooks Python ont une extension **ipynb** et peuvent être manipulés depuis différents environnements dont VSCode.

Au démarrage d'un notebook, un noyau (*kernel*) doit être sélectionné. Dans le service **Vscode-python** de Onyxia, celui de l'environnement **base** convient parfaitement.

Quelques raccourcis utiles pour travailler avec des notebooks :

- **Ctrl+Enter** : exécuter la cellule courante,
- **Shift+Enter** : exécuter la cellule courante et passer à la suivante,
- **A / B** : créer une cellule avant / après la cellule courante,
- **X / C** : couper / copier la cellule courante,
- **V** : coller une cellule après la cellule courante,
- **DD** : supprimer la cellule courante,
- ...

C'est parti !