

Pandas est une bibliothèque open-source permettant la manipulation et l'analyse de données de manière simple et intuitive en Python. Elle a été développée par *Wes McKinney* en 2008 alors qu'il travaillait chez AQR Capital Management. À la fin de l'année 2009, elle a été mise en open source et est aujourd'hui activement utilisée dans le domaine de la [Big data](#) et de la data science car celle-ci **offre des performances et une productivité** élevée à ces utilisateurs.

L'une des forces de Panda est qu'il **se base sur** la très populaire bibliothèque [NumPy](#). Elle fournit **diverses structures de données** et **opérations pour le traitement de données numériques** et de **séries chronologiques**. En plus de cela, les données produites par Panda sont souvent utilisées comme données en entrée pour les fonctions de plotting de Matplotlib, l'analyse statistique en SciPy, les algorithmes de machine learning en Scikit-learn. Les data scientists l'utilisent pour le chargement, le traitement et l'analyse des données tabulaires (données stockées sous format .csv, .tsv ou .xlsx) à l'aide de requêtes de type SQL.

(ressources - peut être utile pour faire notre clustering)