

# Exploration et Manipulation des données Tabulaires avec Python (Numpy et Pandas)

Python pour data scientists



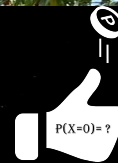


# Table des Matières



## Introduction

- ❑ Introduction aux Données Tabulaire
- ❑ Numpy et Pandas
- ❑ Exploration des Données avec Pandas
- ❑ Manipulation des Données avec Pandas
- ❑ Cas Pratique
- ❑ Visualisation des Données



# Introduction



## Pourquoi ce cours ?

- ✓ **Les données tabulaires sont omniprésentes** : feuilles Excel, bases SQL, fichiers CSV/JSON.
- ✓ **Elles sont essentielles pour** :
  - Comprendre les **tendances de marché**
  - Évaluer la **performance d'une organisation**
  - Analyser des **résultats scientifiques** ou économiques

## Ce que vous allez apprendre

- ✓ Charger des données avec pandas en python (CSV, Excel, JSON...)
- ✓ Nettoyer, transformer et rendre les données exploitables
- ✓ Effectuer des calculs
- ✓ Exporter et partager vos résultats de manière professionnelle

## À qui s'adresse ce cours ? ?

- Étudiants, professionnels ou curieux de la data
- Aucune expertise préalable nécessaire
- Idéal pour **poser les bases solides** en Data Science



# Table des Matières



- Introduction

- **Introduction aux Données Tabulaire**

- Numpy et Pandas

- Exploration des Données avec Pandas

- Manipulation des Données avec Pandas

- Cas Pratique

- Visualisation des Données



# Introduction aux Données Tabulaire



Qu'est-ce qu'une donnée tabulaire ?

Organisation en **lignes (observations)** et **colonnes (variables)**,  
comme dans :

- Fichiers **CSV**, **Excel**
- Bases de données **SQL**, **NoSQL**
- Index **Elasticsearch**

Nom	Ville	Entrées
Musée des Beaux-Arts	Lyon	334459
Musée d'Art Roger-Quillot	Clermont	78386
Musée Dauphinois	Grenoble	76413
Musée Alpin	Chamonix	35747
Musée d'art contemporain	Lyon	135000

# Introduction aux Données Tabulaire



## Pourquoi Python ?

### Flexibilité

- Traite des fichiers **locaux ou massifs**, formats variés (JSON, SQL, API...)
- S'intègre dans des **pipelines automatisés** et des **apps data**

### Connectivité étendue

- Bases SQL (PostgreSQL, MySQL, SQLite...)
- APIs REST, fichiers web, systèmes NoSQL (MongoDB...)
- Big Data (Spark, Hadoop)

### Bibliothèques puissantes

- **NumPy** : calcul rapide sur tableaux numériques (`ndarray`)
- **Pandas** : manipulation intuitive de données tabulaires (`DataFrame`, `Series`)  
Chargement, nettoyage, filtrage, jointures, séries temporelles





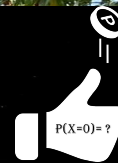
# Table des Matières



- ❑ Introduction
- ❑ Introduction aux Données Tabulaire

## ❑ Numpy et Pandas

- ❑ Exploration des Données avec Pandas
- ❑ Manipulation des Données avec Pandas
- ❑ Cas Pratique
- ❑ Visualisation des Données



# Numpy et Pandas



## Numpy : Calcul numérique

- **Performance** : Code optimisé en C sous le capot
- **Structure de base** : les tableaux `ndarray`
- **Calcul numérique** haute performance
- **Performances** : 100x plus rapide que les listes Python pour les opérations vectorielles.

## Pandas : Manipulation de données tabulaires

- **Structures de données principales** : `DataFrame` (tableau 2D)
- **Lecture/écriture** : CSV (`read_csv`), Excel (`read_excel`), SQL (`read_sql`), JSON (`read_json`), etc.
- **Sélection et filtrage** : **Colonnes, Lignes, Filtres booléens**
- **Productivité** : Moins de code pour des opérations complexes.
- **Interopérabilité** : Compatible avec NumPy, Matplotlib, Scikit-learn.
- **Optimisé pour les données tabulaires** (jusqu'à quelques Go).



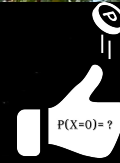
# Table des Matières



- ❑ Introduction
- ❑ Introduction aux Données Tabulaire
- ❑ Numpy et Pandas

## ❑ Exploration des Données avec Pandas

- ❑ Manipulation des Données avec Pandas
- ❑ Cas Pratique
- ❑ Visualisation des Données



# Exploration des Données avec Pandas



**Voir le tutoriel**

<https://github.com/openDataSenegal/Open-Learning-Data-Science/blob/main/Python/ExplorationManipulationDonnesTabulairesNumpyPandas.md>

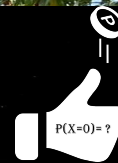
# Table des Matières



- ❑ Introduction
- ❑ Introduction aux Données Tabulaire
- ❑ Numpy et Pandas
- ❑ Exploration des Données avec Pandas

## ❑ Manipulation des Données avec Pandas

- ❑ Cas Pratique
- ❑ Visualisation des Données







## Voir le tutoriel

<https://github.com/openDataSenegal/Open-Learning-Data-Science/blob/main/Python/ExplorationManipulationDonnesTabulairesNumpyPandas.md>



# Table des Matières



- ❑ Introduction
- ❑ Introduction aux Données Tabulaire
- ❑ Numpy et Pandas
- ❑ Exploration des Données avec Pandas
- ❑ Manipulation des Données avec Pandas

## ❑ Cas Pratique

- ❑ Visualisation des Données





## Projet

<https://github.com/openDataSenegal/Open-Learning-Data-Science/blob/main/Python/ExplorationManipulationDonnesTabulairesNumpyPandas.ipynb>



# Table des Matières



- ❑ Introduction
- ❑ Introduction aux Données Tabulaire
- ❑ Numpy et Pandas
- ❑ Exploration des Données avec Pandas
- ❑ Manipulation des Données avec Pandas
- ❑ Cas Pratique



## ❑ Visualisation des Données





Pour visualiser les données, nous pouvons utiliser **Matplotlib** ou **Seaborn** en complément de Pandas



Conclusion

?????

