

Sources de données

December 2022



Numpy datasets

- NumPy est une bibliothèque permettant de travailler avec de grands tableaux multidimensionnels et des matrices de données numériques.
- Il fournit plusieurs jeux de données que vous pouvez utiliser pour pratiquer des techniques de manipulation et de visualisation de données.
- Pour télécharger un jeu de données à partir de NumPy, vous pouvez utiliser la fonction de chargement pour charger un jeu de données à partir d'un fichier ou d'une URL.

```
import numpy as np

# Load the iris dataset from a URL
iris = np.load('https://archive.ics.uci.edu/ml/machine-learning-
databases/iris/iris.data')

# Load the iris dataset from a local file
iris = np.load('iris.data')
```

Pandas datasets

- Pandas est une bibliothèque pour travailler avec des données structurées.
- Il fournit plusieurs fonctions pour lire des données à partir de différents formats, y compris les bases de données CSV, Excel et SQL.
- Pour télécharger un jeu de données à l'aide de Pandas, vous pouvez utiliser la fonction `read_csv` pour lire un fichier CSV à partir d'une URL ou d'un fichier local

```
import pandas as pd

# Load the iris dataset from a URL
iris = pd.read_csv('https://archive.ics.uci.edu/ml/machine-learning-
databases/iris/iris.data')

# Load the iris dataset from a local file
iris = pd.read_csv('iris.data')
```

Scikit-learn datasets



- Scikit-learn est une bibliothèque pour l'apprentissage automatique en Python.
- Il fournit une variété de jeux de données que vous pouvez utiliser pour pratiquer les techniques d'apprentissage automatique. Pour télécharger un jeu de données à partir de Scikit-learn, vous pouvez utiliser la fonction `load_iris` pour charger le jeu de données d'iris.
- Pour obtenir la liste des jeux de données disponibles dans la bibliothèque scikit-learn, vous pouvez utiliser le module `datasets`, qui fournit des fonctions pour charger et récupérer des jeux de données populaires. Le module `datasets` dispose d'un certain nombre de fonctions pour charger des jeux de données, telles que `load_iris`, `load_boston` et `load_diabetes`.
- Vous pouvez utiliser la fonction `dir` pour obtenir une liste de toutes les fonctions du module `datasets`.

```
from sklearn.datasets import load_iris

# Load the iris dataset
iris = load_iris()
```

```
import sklearn.datasets

# Get a list of all the functions in the datasets module
dataset_functions = dir(sklearn.datasets)

# Print the list of functions
print(dataset_functions)
```

```
# Get help for the load_iris function
help(sklearn.datasets.load_iris)
```



Seaborn datasets

- Seaborn est une bibliothèque pour la visualisation de données en Python. Il fournit plusieurs jeux de données que vous pouvez utiliser pour pratiquer des techniques de visualisation de données.
- Pour télécharger un jeu de données à partir de Seaborn, vous pouvez utiliser la fonction `load_dataset` pour charger un jeu de données par nom.
- Pour obtenir une liste des jeux de données disponibles dans Seaborn, vous pouvez utiliser la fonction `get_dataset_names`, qui retourne une liste de tous les jeux de données disponibles dans Seaborn.
- Voici un exemple de la façon dont vous pouvez utiliser la fonction `get_dataset_names` pour obtenir une liste de jeux de données dans Seaborn

```
import seaborn as sns

# Load the iris dataset
iris = sns.load_dataset('iris')
```

```
# Load the iris dataset
iris = sns.load_dataset('iris')
```

```
import seaborn as sns

# Get a list of all the datasets available in Seaborn
dataset_names = sns.get_dataset_names()

# Print the list of datasets
print(dataset_names)
```

```
# Get help for the iris dataset
help(sns.load_dataset('iris'))
```



Kaggle.com

- Pour charger un jeu de données Kaggle directement dans un bloc-notes Jupyter, vous devez disposer d'un compte Kaggle et installer l'API Kaggle. Voici les étapes à suivre pour charger un jeu de données Kaggle dans un bloc-notes Jupyter :
- Créez un compte Kaggle et connectez-vous.
- Accédez à la page du jeu de données sur Kaggle et cliquez sur le bouton « Copier la commande API » pour obtenir la commande API pour télécharger le jeu de données.
- Dans votre terminal ou invite de commandes, exécutez la commande API pour télécharger le jeu de données. Cela téléchargera le jeu de données dans votre répertoire de travail actuel.
- Installez l'API Kaggle en exécutant la commande suivante:



Kaggle.com

```
pip install kaggle
```

```
import kaggle

# Authenticate the API
api = kaggle.api.KaggleApi()
api.authenticate()
```

```
# Load the dataset into a Pandas DataFrame
df = kaggle.api.dataset.load_dataset('dataset-owner/dataset-name')

# Display the first few rows of the DataFrame
df.head()
```

- Dans votre bloc-notes Jupyter, importez les bibliothèques nécessaires et authentifiez l'API à l'aide de la classe KaggleApi.
- Utilisez la classe Dataset pour charger le jeu de données dans un DataFrame Pandas.
- Remplacez dataset-owner et dataset-name par le propriétaire réel et le nom du jeu de données que vous souhaitez charger.



Kaggle.com

```
# Download the dataset to your local machine
kaggle.api.dataset.download_dataset('dataset-owner/dataset-name', path='.')

# Read the dataset into a Pandas DataFrame
import pandas as pd
df = pd.read_csv('dataset-name.csv')

# Display the first few rows of the DataFrame
df.head()
```

- Vous pouvez également utiliser la fonction `download_dataset` pour télécharger le jeu de données sur votre ordinateur local, puis utiliser Pandas pour lire le jeu de données dans un DataFrame.
- Encore une fois, remplacez `dataset-owner` et `dataset-name` par le propriétaire réel et le nom du jeu de données que vous souhaitez charger.

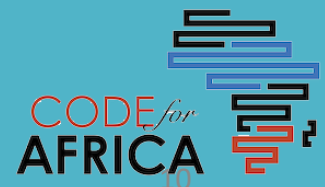
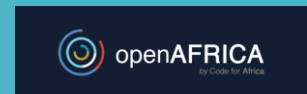
University of California at Irvine (UCI)



- Accès direct aux bases de données UCI:
 - <https://archive.ics.uci.edu/ml/machine-learning-databases/>

Sources for Africa datasets (1)

1. The African Development Bank: The African Development Bank has a data portal with a variety of datasets related to Africa, including economic, social, and development indicators. You can access these datasets at <https://dataportal.afdb.org/>.
2. The African Data Portal: The African Data Portal is a collaboration between the African Development Bank and the United Nations Economic Commission for Africa, and it provides access to a wide range of datasets related to Africa. You can access these datasets at <https://www.africadata.org/>.
3. The African Union: The African Union has a data portal with a variety of datasets related to Africa, including data on political, economic, and social indicators. You can access these datasets at <https://au.int/en/dataportal>.
4. openAFRICA aims to be the largest independent repository of open data on the African continent. openAFRICA is not a government portal. Instead, it's a grassroots initiative, maintained by Code for Africa, as a public service. The platform is available as a free resource for ordinary citizens, civil society organizations, civic activists, the media, and government agencies <https://africaopendata.org/>



Sources for Africa datasets (2)

5. The World Bank: The World Bank provides a wide range of datasets related to Africa, including economic, social, and development indicators. You can access these datasets through the World Bank's Open Data portal at <https://data.worldbank.org/region/africa>.
6. The United Nations: The United Nations has a number of databases and datasets related to Africa, including data on population, health, and development indicators. You can access these datasets at <https://unstats.un.org/unsd/databases>.
7. The United States Census Bureau: The United States Census Bureau has a number of datasets related to Africa, including data on population, demographics, and economic indicators. You can access these datasets at <https://www.census.gov/data/international-data/africa.html>.

