Taille previligier pour les DataSets :

Régression : 1,000 - 10,000 lignes

Classification : 1,000 - 50,000 lignes

Clustering : 1,000 - 20,000 lignes

Deep Learning (DNN, CNN, RNN, etc.)

DataSet Choisie:

* <https://archive.ics.uci.edu/dataset/360/air+quality> (Regression)
  + - 10k lignes
    - 15 features
    - Missing Values : Yes
    - Choisir la Target entre plusieurs
  + <https://archive.ics.uci.edu/dataset/849/power+consumption+of+tetouan+city> (Regression)
    - 50k Lignes
    - 6 Feature
    - Trois Target
  + <https://archive.ics.uci.edu/dataset/374/appliances+energy+prediction> (Regression)
    - 30k lignes
    - 28 Features
    - Missing Values : No
* <https://archive.ics.uci.edu/dataset/357/occupancy+detection(Classification)>
  + - 20k lignes
    - 6 features
    - Missing Values : No
    - Target Occupation 0 ou 1
  + <https://www.kaggle.com/datasets/priyamchoksi/100000-diabetes-clinical-dataset> (Classification)
    - 100k lignes
    - 16 Features
    - Missing Values : No
    - Target : Diabete 0 ou 1
* <https://archive.ics.uci.edu/dataset/352/online+retail> (Clustering)
  + - 500k lignes
    - 8 features
    - Missing Values : No
  + <https://archive.ics.uci.edu/dataset/963/ur3+cobotops>
    - 7k lignes
    - 20 features
  + Intervention Vaines DataSet (Classification, Clustering) DataSet de l’entreprise
    - 1M lignes
    - 30 Features
    - Missing Values : Yes

DataSert non choisie:

* <https://www.kaggle.com/datasets/mlg-ulb/creditcardfraudClustering> (Classification, Clustering)
  + - 300k lignes
    - 10 features
    - Missing Values : No
    - Pas assez d’uniformité
  + <https://archive.ics.uci.edu/dataset/235/individual+household+electric+power+consumption> (Regression, Clustering)
    - 2M de lignes
    - 9 Features
    - Missing Values : Yes
    - Trop de lignes