|  |  |
| --- | --- |
| Date | 12 September 2022 |
| Team ID | PNT2022TMID45267 |
| Project Name | Global Sales Data Analytics |

**Loading the Dataset**

**Your data can be stored in various places; they can be on your local machine’s disk, in a Github repository, and in in-memory data structures like Python dictionaries and Pandas DataFrames. Wherever a dataset is stored, Datasets can help you load it.**

**This guide will show you how to load a dataset from:**

**The Hub without a dataset loading script**

**Local loading script**

**Local files**

**In-memory data**

**Offline**

**A specific slice of a split**

**Hugging Face Hub**

**Datasets are loaded from a dataset loading script that downloads and generates the dataset. However, you can also load a dataset from any dataset repository on the Hub without a loading script! Begin by creating a dataset repository and upload your data files. Now you can use the load\_dataset() function to load the dataset**

**From a python dictionary**

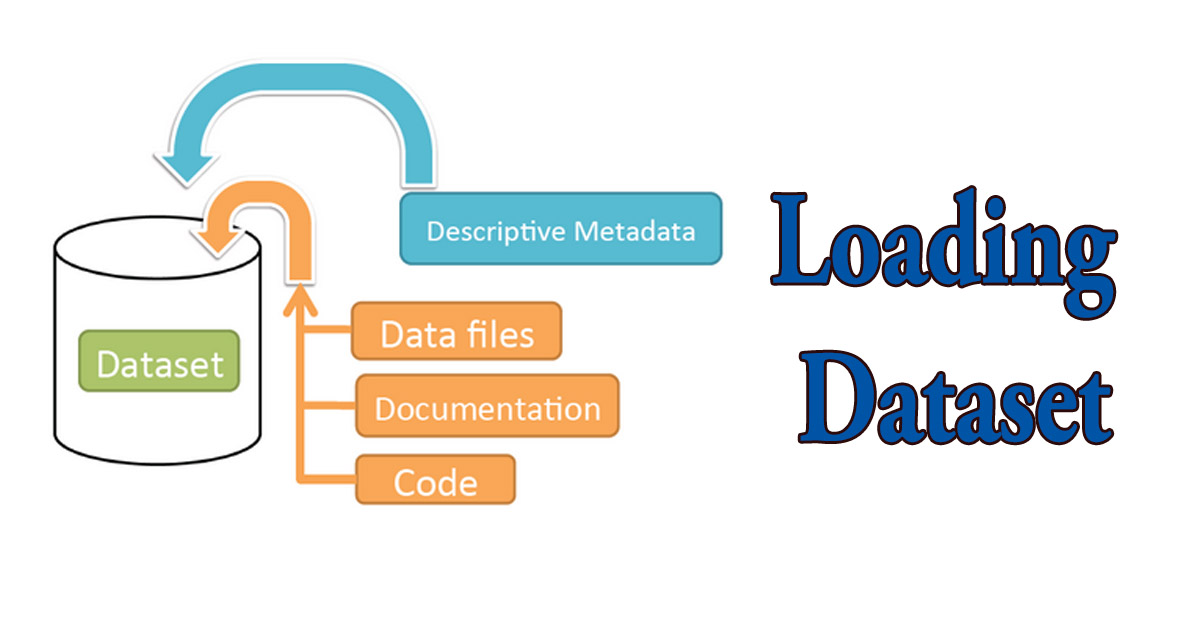
**Let’s say that you have already loaded some data in a in-memory object in your python session:**

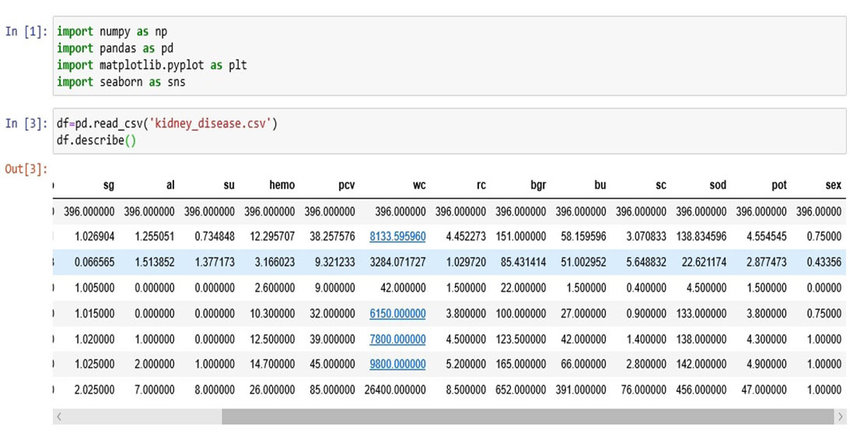
**>>> my\_dict = {‘id’: [0, 1, 2],**

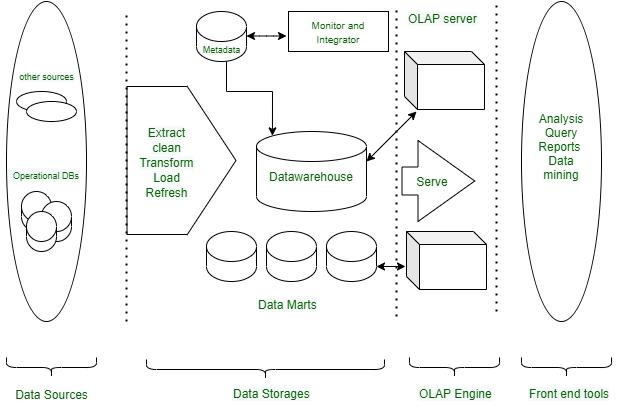
**>>> ‘name’: [‘mary’, ‘bob’, ‘eve’],**

**>>> ‘age’: [24, 53, 19]}**

**You can then directly create a datasets.Dataset object using the datasets.Dataset.from\_dict() or the datasets.Dataset.from\_pandas() class methods of the datasets.Dataset class:**

****

****

****

**If the DataSet already contains rows, the incoming data from the data source is merged with the existing rows. The Load method can be used in several common scenarios, all centered around getting data from a specified data source and adding it to the current data container (in this case, a DataSet ).**