

Course

- CS-E4002
- Course materials
- Your points
- Code Vault

This course has already ended.

« Sparse data with Scipy Course materials Exercises »

CS-E4002 / Functions and Sparse Data / Sparse data with Pandas

Sparse data with Pandas

The Pandas library also provides handling of [sparse data structures](#). It is mainly used to optimize the memory usage of a sparse dataframe.

```
>>> import pandas as pd
>>> df = pd.DataFrame(dense_matrix)
>>> sdf = df.astype(pd.SparseDtype("float", 0.))
```

The functionalities of such a dataframe are the same but some functions might take longer, especially the printing functions, since the table needs to be reconstructed from the sparse information to be able to print.

Note that to build a dataframe from a sparse matrix a specific function has to be used: [from_spmatrix\(\)](#). Notice that it is a function of Pandas' sparse module.

```
>>> import scipy.sparse
>>> mat = scipy.sparse.eye(3)
>>> pd.DataFrame.sparse.from_spmatrix(mat)
   0  1  2
0  1.0 0.0 0.0
1  0.0 1.0 0.0
2  0.0 0.0 1.0
```

« Sparse data with Scipy Course materials Exercises »