

DatasetsLoader

The core of our SDK is the integration of Python's Polars library, chosen for its efficiency in handling large datasets. Polars enables quick data processing and manipulation, which is vital for data analysis and machine learning. Our [DatasetsLoader](#), built on Polars, offers an easy-to-use solution for loading various datasets, making the process smoother and more efficient for data-driven projects.

DatasetsLoader

Locating reliable, easily reproducible datasets can often be a challenge. A key aim of the Giza Datasets SDK is to simplify the process of accessing datasets of various formats and types. The most straightforward way to start is to explore the [Dataset Library](#) or use the [DatasetsHub](#).

Assuming that we have already know the name of the dataset we want to load, we can now use the `DatasetLoader` to load it.

...

```
Copy from giza_datasets import DatasetsLoader
```

Instantiate the DatasetsLoader object

```
loader=DatasetsLoader()
```

...

Depending on your device's configuration, it may be necessary to provide SSL certificates to verify the authenticity of HTTPS connections. You can ensure that all these certifications are correct by executing the following line of code:

...

```
Copy import certifi import os
```

```
os.environ['SSL_CERT_FILE']=certifi.where()
```

...

Lets load the dataset

...

```
Copy df=loader.load('yearn-individual-deposits')
```

```
df.head()
```

...

```
shape: (5, 7)
```

```
evt_block_time evt_block_number vaults token_contract_address token_symbol token_decimals value datetime[ns] i64 str
str i64 f64 2023-06-07 09:50:35 17427717 "0x3b27f92c0e21... "0xdac17f958d2e... "USDT" 6 14174.301085 2022-08-25
13:53:28 15409462 "0x3b27f92c0e21... "0xdac17f958d2e... "USDT" 6 38.046614 2022-08-25 07:13:02 15407745
"0x3b27f92c0e21... "0xdac17f958d2e... "USDT" 6 4620.369198 2022-11-19 03:41:35 16001443 "0x3b27f92c0e21...
"0xdac17f958d2e... "USDT" 6 969.687071 2022-12-30 18:34:11 16299403 "0x3b27f92c0e21... "0xdac17f958d2e... "USDT"
6 56.270566
```

Keep in mind that giza-datasets uses Polars (and not Pandas) as the underlying DataFrame library.

Success! We can now use the loaded dataset for ML development.

[Previous DatasetsHub](#) [Next Aggregated datasets](#)

Last updated 1 month ago