

DatasetsHub

A key aim of the Giza Datasets SDK is to simplify the process of searching the existing collection of datasets of various purposes, formats and sources. The most straightforward way to start is to use the [DatasetsHub](#), the search and query feature for the Giza Datasets library. Using the DatasetsHub, you can search through the datasets within your ML development environment.

Dataset Object

Before using the `DatasetsHub`, it's useful to first understand Datasets themselves. Datasets in `giza_datasets` are represented as [Dataset Class](#), which include details about a dataset such as the dataset's name, description, link to its documentation, tags, etc. You can query information about a given dataset with `DatasetsHub`

DatasetsHub

The `DatasetsHub` class provides methods to manage and access datasets within the Giza Datasets library. Before we delve deeper into various methods, let's import the `DatasetsHub` and instantiate a `DatasetsHub` object.

```
...
```

```
Copy from giza_datasets import DatasetsHub
hub = DatasetsHub()
```

```
...
```

Now we can call different `DatasetsHub` methods.

Use the `show()` method to print a table of all datasets in the hub:

```
...
```

```
Copy hub.show()
```

```
...
```

Use the `list()` method to get a list of all datasets in the hub:

```
...
```

```
Copy datasets = hub.list()
print(datasets)
```

```
...
```

Use the `get()` method to get a `Dataset` object with a given name:

```
...
```

```
Copy dataset = hub.get('tvl-fee-per-protocol')
```

```
...
```

Use the `describe()` method to print a table of details for a given dataset:

```
...
```

```
Copy hub.describe('tvl-fee-per-protocol')
```

```
...
```

Use the `list_tags()` method to print a list of all tags in the hub.

```
...
```

```
Copy hub.list_tags()
```

```
...
```

Use the `get_by_tag()` method to get a list of `Dataset` objects with the given tag.

```
...
```

```
Copy hub.get_by_tag('Liquidity')
```

...

Great! Now we can use [DatasetLoader](#) to load our selected datasets.

[Previous Overview](#) [Next DatasetsLoader](#)

Last updated 1 month ago