
DataBrewer Documentation

Release 0.1.1

Rolando (Max) Espinoza

May 05, 2017

Contents

1	DataBrewer	3
1.1	Quickstart	3
1.2	Datasets	4
1.3	Roadmap	4
1.4	Contributing	5
2	Installation	7
2.1	Stable release	7
2.2	From sources	7
3	History	9
3.1	0.1.1 (2017-05-05)	9
3.2	0.1.0 (2017-05-05)	9
4	Indices and tables	11

Contents:

CHAPTER 1

DataBrewer

The missing datasets manager.

- Free software: MIT license
- Documentation: <https://databrewer.readthedocs.org>.

Databrewer let you search and discover datasets. Inspired by Homebrew, it creates and index of known datasets that you can download with a single command. It will provide an API to allow to do the same in, for example, a IPython notebook so you no longer have to manually download datasets.

Quickstart

Install databrewer:

```
pip install databrewer
```

Update the recipes index:

```
databrewer update
```

Search for some keywords:

```
databrewer search nyc taxi
```

Example output:

```
andresmh-nyc-taxi-trips - NYC Taxi Trips. Data obtained through a FOIA request
nyc-tlc-taxi           - This dataset includes trip records from all trips
                        completed in yellow and green taxis in NYC in 2014 and
                        select months of 2015.
```

Let's check the nyc-tlc-taxi dataset:

```
databrewer info nyc-tlc-taxi
```

We can either download the entire dataset (which is huge!):

```
databrewer download nyc-tlc-taxi
```

Or just a few files in the dataset, or select a subset:

```
databrewer download "nyc-tlc-taxi[green][2014-*]"
```

Note: Note that `*` is the standard glob operator and `[green]` acts as selector. The selectors depends on how the recipe is defined. When using selectors you must enclose the name in quotes in most shells.

Finally you need to know where the files are located for further processing:

```
databrewer download "nyc-tlc-taxi[green][2014-*]"
```

Example output:

```
/Users/rolando/.databrewer/datasets/nyc-tlc-taxi/green_tripdata_2014-01.csv
/Users/rolando/.databrewer/datasets/nyc-tlc-taxi/green_tripdata_2014-02.csv
/Users/rolando/.databrewer/datasets/nyc-tlc-taxi/green_tripdata_2014-03.csv
/Users/rolando/.databrewer/datasets/nyc-tlc-taxi/green_tripdata_2014-04.csv
/Users/rolando/.databrewer/datasets/nyc-tlc-taxi/green_tripdata_2014-05.csv
/Users/rolando/.databrewer/datasets/nyc-tlc-taxi/green_tripdata_2014-06.csv
/Users/rolando/.databrewer/datasets/nyc-tlc-taxi/green_tripdata_2014-07.csv
/Users/rolando/.databrewer/datasets/nyc-tlc-taxi/green_tripdata_2014-08.csv
/Users/rolando/.databrewer/datasets/nyc-tlc-taxi/green_tripdata_2014-09.csv
/Users/rolando/.databrewer/datasets/nyc-tlc-taxi/green_tripdata_2014-10.csv
/Users/rolando/.databrewer/datasets/nyc-tlc-taxi/green_tripdata_2014-11.csv
/Users/rolando/.databrewer/datasets/nyc-tlc-taxi/green_tripdata_2014-12.csv
```

Datasets

The aim is to index known and not-so-known datasets. There is no plans to standardize the dataset format as we want to keep it as published by the authors.

Recipes

Datasets are defined in recipes which contains information about the dataset and where to find it.

These recipes are community maintained and hosted in the [databrewer-recipes](#) repository.

Roadmap

- Include an API. For now it only provides a CLI-interface but in the near future it will include an API so you can search, download and load datasets directly in your Python code.

Contributing

You can help by the following means:

- Spread the word!
- Request missing features
- Report usability issues or bugs
- Request missing datasets

See `CONTRIBUTING.rst` for more information.

Stable release

To install DataBrewer, run this command in your terminal:

```
$ pip install databrewer
```

If you don't have [pip](#) installed, this [Python installation guide](#) can guide you through the process.

From sources

The sources for DataBrewer can be downloaded from the [Github repo](#).

You can either clone the public repository:

```
$ git clone git://github.com/rolando/databrewer
```

Or download the [tarball](#):

```
$ curl -OL https://github.com/rolando/databrewer/tarball/master
```

Once you have a copy of the source, you can install it with:

```
$ pip install -e .
```


0.1.1 (2017-05-05)

Fix packaging issues.

0.1.0 (2017-05-05)

- First release on PyPI.

CHAPTER 4

Indices and tables

- `genindex`
- `modindex`
- `search`