

---

# **openclean-notebook**

**New York University**

**Aug 15, 2021**



**CONTENTS:**

<b>1</b>	<b>Installation</b>	<b>3</b>
<b>2</b>	<b>Usage</b>	<b>5</b>
<b>3</b>	<b>openclean_notebook</b>	<b>7</b>
3.1	openclean_notebook package . . . . .	7



# openclean *NOTEBOOK*

This package provides a graphical user interface for **openclean** that can be used to visualize and manipulate datasets in notebook environments like Jupyter Notebooks.



## INSTALLATION

The package can be installed using `pip`.

```
pip install openclean-notebook
```

You can use the additional *[jupyter]* option to install the Python Jupyter package if you want to use the UI within a Jupyter Notebook.

```
pip install openclean-notebook[jupyter]
```

The notebook UI is a JavaScript bundle that is included in the installed package.





## USAGE

To use the notebook UI, an instance of the `openclean_notebook.engine.OpencleanAPI` is required. The API engine provides a namespace that manages a set of datasets that are identified by unique names. The engine is associated with an object repository that provides additional functionality to register objects like functions, lookup tables, etc.. The engine is also responsible for coordinating the communication with the JavaScript UI.

A helper function to create an instance of the openclean API is included in the `openclean_notebook` package. For example:

```
from openclean_notebook import DB
db = DB(basedir='.openclean', create=True)
```

In this example a new instance of the API engine is created that stores all dataset files in a local folder `.openclean`. The `create=True` flag ensures that a fresh instance is created every time the code (cell) is run.

The next step is to create a new dataset in the API, e.g., from a given data frame or data file. Each dataset has to have a unique name.

```
db.load_dataset(source=source='./data/bre9-aqqr.tsv.gz', name='covid-cases')
```

You can then either view and edit the full dataset using the notebook UI or (e.g., for performance reasons) a sample of the dataset. The recipe that is created from the interactions in the notebook UI can later be applied on the full dataset. In the example below we use a sample of 100 rows for display in the notebook UI.

```
db.edit('covid-cases', n=100)
```

For a full example please have a look at the [example notebook](#) that also shows how to register and run commands on the dataset.



## OPENCLEAN\_NOTEBOOK

### 3.1 openclean\_notebook package

#### 3.1.1 Subpackages

openclean\_notebook.controller package

Subpackages

openclean\_notebook.controller.spreadsheet package

Submodules

openclean\_notebook.controller.spreadsheet.base module

openclean\_notebook.controller.spreadsheet.data module

Submodules

openclean\_notebook.controller.comm module

openclean\_notebook.controller.html module

#### 3.1.2 Submodules

openclean\_notebook.engine module

openclean\_notebook.version module