**List of opensource tools to boost data analysis with Python Pandas library**



Pandas is a Python library for data analysis. [Wes McKinney](http://wesmckinney.com/pages/about.html) started it in 2008 out of a need for a effective and flexible quantitative analysis tool, pandas has grown into one of the most popular Python libraries.

**[pandas documentation - pandas 1.5.3 documentation](https://pandas.pydata.org/docs/" \t "_blank)**

[The reference guide contains a detailed description of the pandas API. The reference describes how the methods work and…](https://pandas.pydata.org/docs/" \t "_blank)

[pandas.pydata.org](https://pandas.pydata.org/docs/" \t "_blank)

In my search, I found the following open-source python libraries that will immensely boost your Pandas workflow the moment you start using them.

pandas-profiling: Generate EDA report of data in one-line of code.

* Link: [https://bit.ly/pd-profiling](https://substack.com/redirect/ab131132-bcfa-45bd-bd46-d93346c6f6a3?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

D-Tale: Visualizer for pandas dataframe.

* Link: [https://bit.ly/py-dtale](https://substack.com/redirect/73ec7c75-1288-4ca0-bf81-cb7181ea003c?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

PandasML: Pandas, sklearn and matplotlib integrated.

* Link: [https://bit.ly/pandasml](https://substack.com/redirect/7de6bb6a-f192-4513-887b-f07972f96625?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

Jupyter-Datatables: Enrich the default preview of a DataFrame in jupyter notebook.

* Link: [https://bit.ly/jupy-dtables](https://substack.com/redirect/deafd974-64ba-4854-ac54-ba3fc5c605d6?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

Pandera: Provides a flexible and expressive API for performing data validation on dataframe-like objects to make data processing pipelines more readable and robust.

* Link: [https://pandera.readthedocs.io/en/stable/#](https://pandera.readthedocs.io/en/stable/)

SummaryTools: Boost the describe() method in Pandas.

* Link: [https://bit.ly/summarytools](https://substack.com/redirect/1bda78ce-58e3-4c9e-95a8-e95f29416af6?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

Sidetable: Charge the value\_counts() method in Pandas.

* Link: [https://bit.ly/py-sidetable](https://substack.com/redirect/7bee70a0-85f5-4b23-b226-f46c9e7095d1?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

Sketch: Generate code/insights about data by asking questions in natural language.

* Link: [https://bit.ly/py-sketch](https://substack.com/redirect/e8604512-7a90-4bba-b54d-5c84e50a535f?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

Deepchecks: Generate a comprehensive validation report of your data.

* Link: [https://bit.ly/deepchks](https://substack.com/redirect/25a5d38b-0536-4068-b91f-77cf7192b3e3?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

Pandas Flavor: Extend Pandas to attach methods to the dataframe object.

* Link: [https://bit.ly/pd-flavor](https://substack.com/redirect/a11d0688-aca4-4708-a432-228d35413804?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

Pandarallel: Parallelize Pandas across multiple CPU cores.

* Link: [https://bit.ly/pandarallel](https://substack.com/redirect/f3c667a6-594d-48f6-b39d-c3af005525df?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

Geopandas: Work with Geospatial data in Pandas.

* Link: [https://bit.ly/geo-pd](https://substack.com/redirect/fc472610-94f7-4cdd-8b4e-dafb438716b4?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

DuckDB: Run SQL queries on dataframes.

* Link: [https://bit.ly/duckdb](https://substack.com/redirect/6fa7bf8e-66dc-4cef-a92e-0fed8204af4f?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

Modin: Boost Pandas’ performance up to 70x by modifying the import.

* Link: [https://bit.ly/modin-guide](https://substack.com/redirect/2c366973-530c-4705-b349-4ed95a15410a?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

PivotTableJS: Create pivot tables by using drag and drop functionality.

* Link: [https://bit.ly/pivottablejs](https://substack.com/redirect/3325e9b2-1b46-440b-a7c2-7d28f6629570?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

Missingno: Visualize missing values in your dataset.

* Link: [https://bit.ly/py-missingno](https://substack.com/redirect/b5b75a35-0216-4e09-bc89-096d4e2680cc?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

Pandas Alive: Create animated charts for pandas dataframes.

* Link: [https://bit.ly/pd-alive](https://substack.com/redirect/1dee9e7b-1ef9-456b-a9c0-6d34d1d6db6d?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

Skimpy: Supercharge the describe() method in Pandas.

* Link: [https://bit.ly/py-skimpy](https://substack.com/redirect/85a9a2b1-946e-4104-ab65-1ab8a045e6b3?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

Pandas-log: Debug pandas pipeline using step-by-step logging.

* Link: [https://bit.ly/py-log](https://substack.com/redirect/85517045-6eed-4cb6-9f6b-613d269c56d9?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

tsflex: Process time series and perform feature extraction.

* Link: [https://bit.ly/tsflex](https://substack.com/redirect/febd5ff1-e122-4797-bdcb-976cd67d51fb?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

Mars: A tensor-based framework for scaling numpy, pandas, scikit-learn, and Python functions.

* Link: [https://bit.ly/py-mars](https://substack.com/redirect/a59d8052-7905-43e4-84c8-79ee0e47b341?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

nptyping: Apply type hints for Pandas data frames.

* Link: [https://bit.ly/nptyping](https://substack.com/redirect/4b2dd702-df46-492f-bd9f-a922b6bef661?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

popmon: Profile your data to determine its stability.

* Link: [https://bit.ly/py-popmon](https://substack.com/redirect/dfc38211-c321-4400-b9db-eb5f0b13b71f?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

Gspread-pandas: Interact with Google sheets through pandas dataframes.

* Link: [https://bit.ly/pd-gsheets](https://substack.com/redirect/ee937350-b95c-433b-ba0c-f14c19553e98?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

pdpipe: Create pandas pipeline easily and intuitively.

* Link: [https://bit.ly/py-pdpipe](https://substack.com/redirect/e596eb25-063b-434a-ada6-a03b475ee01b?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

PrettyPandas: Prettify the dataframe when printed.

* Link: [https://bit.ly/PrettyPandas](https://substack.com/redirect/dfc1e2c4-d2e6-4f73-bfe2-c7b860cec0fb?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

Dora: An intuitive API for data cleaning, processing, feature selection, visualization, etc.

* Link: [https://bit.ly/py-dora](https://substack.com/redirect/4a742112-4805-493e-8415-e477399693a9?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

Pandapy: The speed of NumPy combined with Pandas’ elegance.

* Link: [https://bit.ly/pandapy](https://substack.com/redirect/566652ab-becf-4aca-82db-30ad282e4b7c?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

PyJanitor: A clean API for cleaning data.

* Link: [https://bit.ly/pyjanitor](https://substack.com/redirect/4d4c72ac-57cb-43ea-a4ce-ff3577c5a821?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

swifter: Speed-up the apply() method in Pandas.

* Link: [https://bit.ly/py-swifter](https://substack.com/redirect/b0d23fb1-2732-4fd9-9564-730f08b64ca4?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

Mito: Analyze data in Jupyter by editing a spreadsheet.

* Link:[https://bit.ly/mito-ds](https://substack.com/redirect/faa3342b-5dbd-42ee-a8b4-289ff69d6e9c?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

Visual Python: GUI-based Python code generator for data science

* Link: [https://bit.ly/visual-py](https://substack.com/redirect/f9c13f18-110f-4198-b543-6dc3bd1212fd?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

tqdm: Add progress bars to Pandas methods.

* Link: [https://bit.ly/tqdm-pd](https://substack.com/redirect/dcc6bf67-8646-4dd9-8162-0b184ace4dc5?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

Lux: Automatic data visualization.

* Link: [https://bit.ly/pd-lux](https://substack.com/redirect/0d1f71b3-3e08-4130-a998-5d41ced93841?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

AutoClean: Automated data preprocessing & cleaning.

* Link: [https://bit.ly/py-autoclean](https://substack.com/redirect/8c9bc102-c0f9-406e-8a0d-f231c74f7072?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

pytablewriter: Write a dataframe in various formats: AsciiDoc / CSV / HTML / JSON / LaTeX / Markdown / Excel / TOML / TSV / YAML, etc.

* Link: [https://bit.ly/pytablewriter](https://substack.com/redirect/559d6a66-49a4-4b91-8e2d-d988e702e5a3?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

itables: Pandas dataframes as interactive datatables.

* Link: [https://bit.ly/itables](https://substack.com/redirect/47aef467-132d-49f0-b4c7-7f60137c060e?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

PandasGUI: A GUI for Pandas dataframes.

* Link: [https://bit.ly/PandasGUI](https://substack.com/redirect/5efcdfa4-ecbb-4176-8910-530f05f6535f?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

tabula-py: Extract table from PDF into Pandas dataframe.

* Link: [https://bit.ly/tabulapy](https://substack.com/redirect/78a79be4-6e2a-4004-b02c-0c5637106bbb?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

Pingouin: Perform statistical testing on Pandas dataframe.

* Link: [https://bit.ly/pypingouin](https://substack.com/redirect/1f279f78-7fc4-4f26-9561-24f1483b1af3?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

Dexplot: Create many types of beautiful data visualizations with a simple, consistent, and intuitive API.

* Link: [https://bit.ly/dexplot](https://substack.com/redirect/7746fb67-3aca-4222-a515-4000c59d5cfb?j=eyJ1IjoiMjE5anF5In0.hV4AbmX97aK5ekEE8UX6EvvtfOlqa6rD3uEs97Ea90g)

A huge shout out to the creators of these libraries. These libraries really increases pandas usability. There might be many more libraries to make the tabular datasets easy to analyze in Python.

Happy data exploring everyone!