

[Open in app](#)[Follow](#)

603K Followers



You have **2** free member-only stories left this month. [Upgrade for unlimited access.](#)

PYTHON

Mito: One of the Coolest Python Libraries You Have Ever Seen

Here is my take on this cool Python library and why you should give it a try



Ismael Araujo 1 day ago · 9 min read ★



[Open in app](#)

A few weeks ago, I posted a blog about DataRobot that became quite popular. The blog was very well received, achieving tens of thousands of views in the first week. And after that, I had planned to write about other Data Science-related subjects and was going to avoid writing about Python libraries for some time unless I had found something awe-inspiring. Well, I found it, and it's called Mito.

I heard about Mito a while ago but never had the chance to test it out, but recently I decided to give it a try, and it's impressive! But before anything, a little note: this is **not** sponsored content. I'm saying this because some developers have reached out to me in the past few weeks asking if I would do a paid partnership, so if you are one of those folks, I don't do any kind of paid partnerships. However, I do accept suggestions of subjects to write about. Now, let's go back to Mito.

Mito — save hours of work with a few clicks

Mito is a Python library that helps you perform data preparation, data cleaning, transformation, exploratory data analysis, creating graphs, etc. Through a GUI, you can get a lot done with just one or two lines of codes and a few clicks. They tried to create a similar experience as if you were using Microsoft Excel. Mito and Bamboolib (another Python library) have many similarities, such as having a GUI to make it easier for people to perform tasks, and they both create Python codes that you can copy and use anywhere, even if you don't have Mito installed.

If you haven't read my blog on Bamboolib, you can find it here:

Bamboolib: One of the Most Useful Python Libraries You Have Ever Seen

Here is my take on this cool Python library and why you should give it a try

towardsdatascience.com

That being said, let's put our hands on Mito!

Installation

The installation process is easy. For security reasons, I recommend you create an environment for Mito. Their website shows how you can create a Python environment. To do so, you can copy and paste one of the two codes into your terminal:

[Open in app](#)

```
python3 -m venv mitoenv
source mitoenv/bin/activate
python -m pip install mitoinstaller
python -m mitoinstaller install
```

Windows:

```
python3 -m venv mitoenv
mitoenv\Scripts\activate.bat
python -m pip install mitoinstaller
python -m mitoinstaller install
```

If you decide to go with a Conda Virtual Environment, you can copy and paste the following code:

```
conda create -n mitoenv python=3.8
conda activate mitoenv
python -m pip install mitoinstaller
python -m mitoinstaller install
```

After these steps, you should be good to go. Note that Mito will create a JupyterLab notebook file for you with the starting code. You can also initiate a JupyterLab notebook by typing `jupyter lab` in your Terminal. Now let's get our hands dirty.

First Steps

Initiating Mito couldn't be any easier. You just need to import it by typing `import mitosheet` and initiate it by typing `mitosheet.sheet()`. That's all the coding we will be using today!

[Open in app](#)

GIF by Author

Now let's import the dataset that we will be using. Click on `import`, find the file, and we are good to go. For this demonstration, I will be using the [Top Video Games 1995–2021 Metacritic](#) dataset from Kaggle. After importing the dataset, Mito shows us a few carts with what you can do with it.

The screenshot shows the Mito application interface. At the top, there is a code editor window containing the following Python code:

```
import mitosheet
mitosheet.sheet()
```

Below the code editor is a toolbar with various icons for data manipulation: UNDO, REDO, CLEAR, IMPORT, EXPORT, ADD COL, DEL COL, PIVOT, MERGE, GRAPH, SAVE, REPLAY, STEPS, DOCS, and FULLSCREEN. A vertical scroll bar is visible on the left side of the main workspace.

In the center of the workspace, a message reads:

You have not imported any data into Mito yet.
Pass a dataframe to the `mitosheet.sheet` call, or use the Import button in the toolbar above.

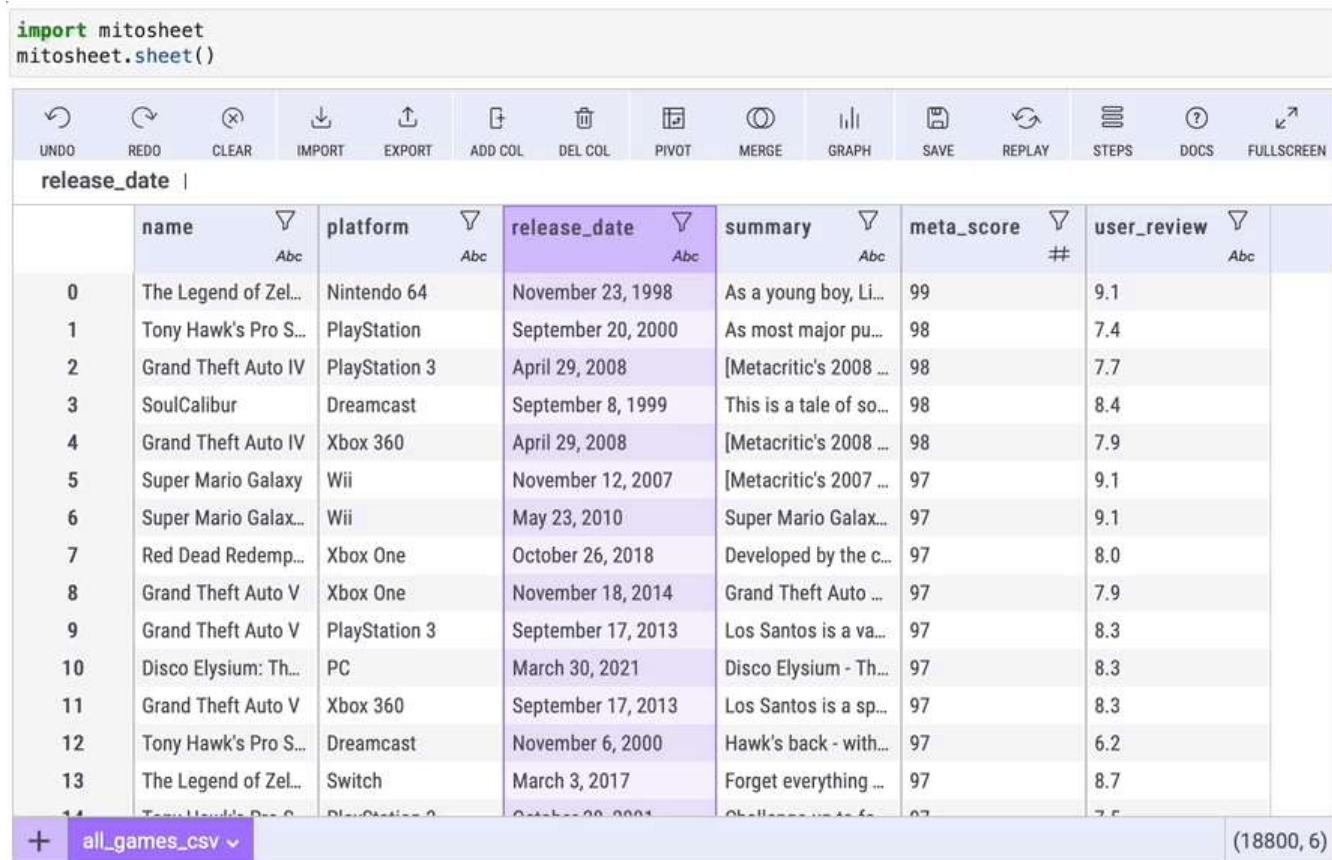
At the bottom left of the workspace is a small purple square with a white plus sign (+).

GIF by the Author

Data Preparation


[Open in app](#)

datatype, right below the column's name, and choose the new datatype. In the example below, I will change a string to a `datetime` datatype. You will see that the icon will change from *Abc* (which means that the datatype is a string) to a calendar icon. DAT EASY!



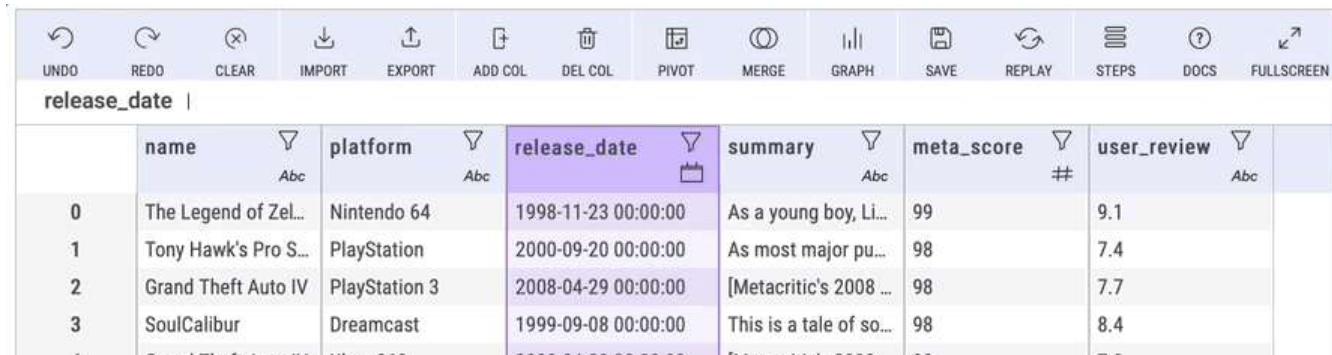
```
import mitosheet
mitosheet.sheet()
```

	name	platform	release_date	summary	meta_score	user_review
0	The Legend of Zel...	Nintendo 64	November 23, 1998	As a young boy, Li...	99	9.1
1	Tony Hawk's Pro S...	PlayStation	September 20, 2000	As most major pu...	98	7.4
2	Grand Theft Auto IV	PlayStation 3	April 29, 2008	[Metacritic's 2008 ...	98	7.7
3	SoulCalibur	Dreamcast	September 8, 1999	This is a tale of so...	98	8.4
4	Grand Theft Auto IV	Xbox 360	April 29, 2008	[Metacritic's 2008 ...	98	7.9
5	Super Mario Galaxy	Wii	November 12, 2007	[Metacritic's 2007 ...	97	9.1
6	Super Mario Galax...	Wii	May 23, 2010	Super Mario Galax...	97	9.1
7	Red Dead Redemp...	Xbox One	October 26, 2018	Developed by the c...	97	8.0
8	Grand Theft Auto V	Xbox One	November 18, 2014	Grand Theft Auto ...	97	7.9
9	Grand Theft Auto V	PlayStation 3	September 17, 2013	Los Santos is a va...	97	8.3
10	Disco Elysium: Th...	PC	March 30, 2021	Disco Elysium - Th...	97	8.3
11	Grand Theft Auto V	Xbox 360	September 17, 2013	Los Santos is a sp...	97	8.3
12	Tony Hawk's Pro S...	Dreamcast	November 6, 2000	Hawk's back - with...	97	6.2
13	The Legend of Zel...	Switch	March 3, 2017	Forget everything ...	97	8.7
14	Tony Hawk's Pro S...	PlayStation 2	October 29, 2001	Overall rating: 97	97	7.5

+ all_games_csv ~ (18800, 6)

They don't have an option where you can choose the format of the `datetime` datatype, so if any of the Mito developers ready this blog, I think it would be a nice extra feature if we can choose the format.

Also, it seems like the `user_review` column is a string. Let's fix this by changing it to a float.



	name	platform	release_date	summary	meta_score	user_review
0	The Legend of Zel...	Nintendo 64	1998-11-23 00:00:00	As a young boy, Li...	99	9.1
1	Tony Hawk's Pro S...	PlayStation	2000-09-20 00:00:00	As most major pu...	98	7.4
2	Grand Theft Auto IV	PlayStation 3	2008-04-29 00:00:00	[Metacritic's 2008 ...	98	7.7
3	SoulCalibur	Dreamcast	1999-09-08 00:00:00	This is a tale of so...	98	8.4


[Open in app](#)

8	Grand Theft Auto V	Xbox One	2014-11-18 00:00:00	Grand Theft Auto ...	97	7.9		
9	Grand Theft Auto V	PlayStation 3	2013-09-17 00:00:00	Los Santos is a va...	97	8.3		
10	Disco Elysium: Th...	PC	2021-03-30 00:00:00	Disco Elysium - Th...	97	8.3		
11	Grand Theft Auto V	Xbox 360	2013-09-17 00:00:00	Los Santos is a sp...	97	8.3		
12	Tony Hawk's Pro S...	Dreamcast	2000-11-06 00:00:00	Hawk's back - with...	97	6.2		
13	The Legend of Zel...	Switch	2017-03-03 00:00:00	Forget everything ...	97	8.7		
14	Tomb Raider	PlayStation 4	2018-03-06 00:00:00	It's time to go back...	97	7.7		

[+ all_games_csv ▾](#)

(18800, 6)

```
# MITO CODE START (DO NOT EDIT)

from mitosheet import * # Import necessary functions from Mito
register_analysis('UUID-d2da963e-00f8-451c-9f62-2de33e040cef') # Let Mito know which analysis is being run

# Imported /Users/ismaeltrevi/Desktop/Mito/all_games.csv
import pandas as pd
all_games_csv = pd.read_csv(r'/Users/ismaeltrevi/Desktop/Mito/all_games.csv')

# Changed release_date from object to datetime
all_games_csv['release_date'] = pd.to_datetime(all_games_csv['release_date'], infer_datetime_format=True, errors='raise')

# MITO CODE END (DO NOT EDIT)
```

GIF by the author

Did you see that a line of code was added to the cell below Mito's GUI? That's because Mito also gives you the code to use it in another notebook, even if you don't have Mito installed. If you are learning to code, but don't know how to do something, this is a fantastic tool. Another cool thing about it is that it automatically adds comments in the cell so that anyone can read it.

Just to make a quick comparison, Bamboolib can automatically recognize that the `user_review` column should be a float and not an integer, but Mito cannot, so if you try to change a column with digits to an integer, it will throw you an error. Not a significant problem, though.

[Open in app](#)

GIF by the Author

Renaming columns

Changing the column's name couldn't be any simpler. Just like you do in Excel, you will just click on the column's name and edit. Awesome, right? I worked on a project a while ago, and I had to change the name of 200 columns, and it took me quite some time. I can't think how much time I would have saved if I knew Mito.



GIF by the author

Dropping columns

Dropping columns is another thing that Mito makes ridiculously simple. Just select the column, and click on `DEL COL`. That's it!

[Open in app](#)

GIF by the author

Undoing actions

Mito makes deleting columns so easy that someone could delete more columns than they would like. If you are using Jupyter notebook without Mito, what would you do? Run all the code again? Who has time for that? Just click on `UNDO`, and you can undo any actions. Did you use `UNDO` a little too much? No problem! Just click on `REDO`, and you will undo the undo. Sorry, I had to say it.



[Open in app](#)

GIF by the author

Using multiple DataFrames

Another very cool feature is that you can work with multiple DataFrames at the same time. There are two ways to do it. You can click on `IMPORT` and upload files from your computer, or, in case you have the DataFrames in your notebook, you can add their names inside the parenthesis, just like the following code.

```
mitosheet.sheet(df1, df2)
```



GIF by the author

For the example above, I created two DataFrames. One where `user_review > 9` and one where `user_review < 9`.

Data Transformation

[Open in app](#)

filter. In the example below, I'm filtering to see only games where `user_review` is larger than 9.5. You can also create filter groups where you can select multiple conditionals.



GIF by the author

Pivot Table

`.groupby()` is very useful for data analysis and, even if you are a pro user, you need to confess that grouping data in Python can be a little time-consuming. If you create one or two `.groupby()` you are probably ok, but doing multiple times every day can become annoying. Now, Mito doesn't have a `.groupby()` option. It had `.pivot_table()` instead, which works similarly. But what's the difference between them?

In short, `.pivot_table()` allows you to aggregate data into more shapes. For example, you can choose the index, column, and rows values. `.groupby()` will create tables where the given dimensions are placed into columns, and the rows will be created for each combination of those dimensions.

Now that we understand the difference between them, you can create a pivot table by clicking on PIVOT, selecting the columns you want to create a group by, the stats you want to see, and voilá!

[Open in app](#)

GIF by the author

Here's the code created by Mito for the `.pivot_table()` that we just did. That's quite some code and we did all that with a few clicks.

```
# Pivoted all_games_csv into df2
unused_columns =
all_games_csv.columns.difference(set(['platform']).union(set([])).union(set({'name', 'user_review', 'meta_score'})))
tmp_df = all_games_csv.drop(unused_columns, axis=1)
pivot_table = tmp_df.pivot_table(
    index=['platform'],
    values=['meta_score', 'name', 'user_review'],
    aggfunc={'meta_score': ['mean'], 'name': ['count'],
    'user_review': ['mean']}
)
```

That's a nice feature, but what it's really cool to me is that you get the group by in a new tab. You can navigate back in forth through tabs and easily compare results.

[Open in app](#)

GIF by the author

Creating conditional columns

You can create columns based on conditionals with very few steps using IF. In the example below, I will create a new column saying if the user_review column is above nine or not. To do so, I clicked on any value in the user_review column, clicked on ADD COL, and typed `=IF(user_review,'Yes', 'No')`. The way it works is `=IF(column, value_IF_True, value_IF_False)`, which is pretty similar to what you would do in Excel.



[Open in app](#)

Now, let's explore the code that Mito generated:

```
all_games_csv.rename(columns={"new-column-1inv": "above_9"},  
inplace=True)  
  
all_games_csv['above_9'] = IF(all_games_csv['user_review'] > 9,  
'Yes', 'No')
```

I must confess that I didn't know that we could use `IF` as is in the code snippet. I would normally use a `.lambda()` function. Good to know!

Data Exploration

Mito makes data exploration intuitive. Just click on the column, click on the filter icon, then click on `Summary`. You will see information such as a graph of the data distribution, and descriptive statistics such as mean, standard deviation, number of null values, etc.



GIF by the author

It works for numerical and categorical, in case you are wondering.

[Open in app](#)

Data Visualization

If you need to create some basic graphs, you can do that as well. For example, you can create a bar plot with a few clicks, and you can generate the code used to generate the graph if you want to edit it. Just click on `GRAPH`, then select the chart type and the x-axis and y-axis.



GIF by the author

Final Thoughts

[Open in app](#)

people to start using Python. Python can be a little tricky and frustrating for those who are beginning, and I believe that Mito can help people learn and develop their skills.

Even if you have been using Python for years, you can learn a thing or two by observing the code generated by Mito. Now, is it for everyone? That's a tricky question, and it will depend on what you need, but I think even people who have been coding for years can take advantage of some of the features and, instead of typing 15 lines of code, maybe you can achieve the same in a few clicks. Nevertheless, try it out and let me know what you think about it. Happy Coding!

You might also like:

Bamboolib: One of the Most Useful Python Libraries You Have Ever Seen

Here is my take on this cool Python library and why you should give it a try

[towardsdatascience.com](https://towardsdatascience.com/bamboolib-one-of-the-most-useful-python-libraries-you-have-ever-seen-5261dbfedceb)

5 Python Libraries That You Don't Know, But Should

[towardsdatascience.com](https://towardsdatascience.com/5-python-libraries-that-you-don-t-know-but-should-5a2a2f3a2a2f)

3 Awesome Python Libraries That You Should Know About

Discover three cool libraries and why you should try them right now

[towardsdatascience.com](https://towardsdatascience.com/3-awesome-python-libraries-that-you-should-know-about-5a2a2f3a2a2f)

[Open in app](#)

Sign up for The Variable

By Towards Data Science

Every Thursday, the Variable delivers the very best of Towards Data Science: from hands-on tutorials and cutting-edge research to original features you don't want to miss. [Take a look.](#)

[Get this newsletter](#)

Emails will be sent to pedrotechel@gmail.com.

[Not you?](#)

[Data Science](#) [Python](#) [Programming](#) [Coding](#) [Artificial Intelligence](#)

[About](#) [Write](#) [Help](#) [Legal](#)

Get the Medium app

