

BRIEF INTRODUCTION TO PYTHON

DANIEL MALDONADO

FEBRUARY 2022

Contents

1	Introduction	2
1.1	Why MedPCPy	2
1.2	About this package	2
2	Brief introduction to Python	3
2.1	Installing Python	3
2.2	Integrated Development Environments	3
2.3	Virtual Environments	3
2.4	Python basics	5
2.4.1	Data types	5
2.4.2	Variable assignment	6
2.4.3	Nesting	6
2.4.4	Functions	7
2.4.5	Classes and objects	9
2.4.6	Installing libraries	10
3	This library	12
3.1	First steps	12
3.2	Importing the library	13
3.3	Declaring variables	13
3.4	Final use	14
4	Acknowledgments	17
5	Other resources	17

1 Introduction

1.1 Why MedPCPy

Within behavior analysis a common limiting factor is the inability to process data efficiently. Although given enough time most researchers would probably find ways to manage their data with varying degrees of success, many still struggle with manually converting their files to a more manageable format and making manual counts. Excel macros are a great tool for this, but they still require a great deal of involvement from the researcher, and given that they require human interaction they are prone to errors.

The motivation to make this library was to provide a free and accessible tool that allowed both new and seasoned researchers to quickly organize the data they need without spending time learning to code. This tool aims to be easy to learn and use, less error prone than regular scripting, and above all, free (as in *gratis* and as in *libre*) for everyone. Med Associates Inc.[®] provides their own software for a similar purpose, but it is prohibitively costly. Those who develop open tools share the belief that one cannot put a price neither on knowledge nor on the tools to produce it.

1.2 About this package

The team behind MedPCPy has extensively worked to make sure that the library works in all conditions. Bugs and special cases have been fixed to the best of our abilities. However—as is the case with anything programming-related—something could go wrong at some point. We encourage you to try and test the library in different cases, and to let us know when/if things go wrong so that we can improve the library and make it more useful for everyone.

If you find that something does not work as expected, or if you have any ideas on how to improve the library, please feel free to either open an issue on the project's [GitHub](#) page, or [email us](#).

2 Brief introduction to Python

Python is a general purpose, interpreted programming language. Although it is not as fast as other languages, such as *C*, its simple syntax makes it ideal for purposes as machine learning and artificial intelligence, among many others. It is easy to learn and very powerful. For those reasons it was chosen to develop this library.

2.1 Installing Python

Python can be installed on any machine by downloading it from its [official page](#), from the Microsoft Store, or from any Linux repository. Several tutorials on YouTube will be of help if any problems arise during the installation.

2.2 Integrated Development Environments

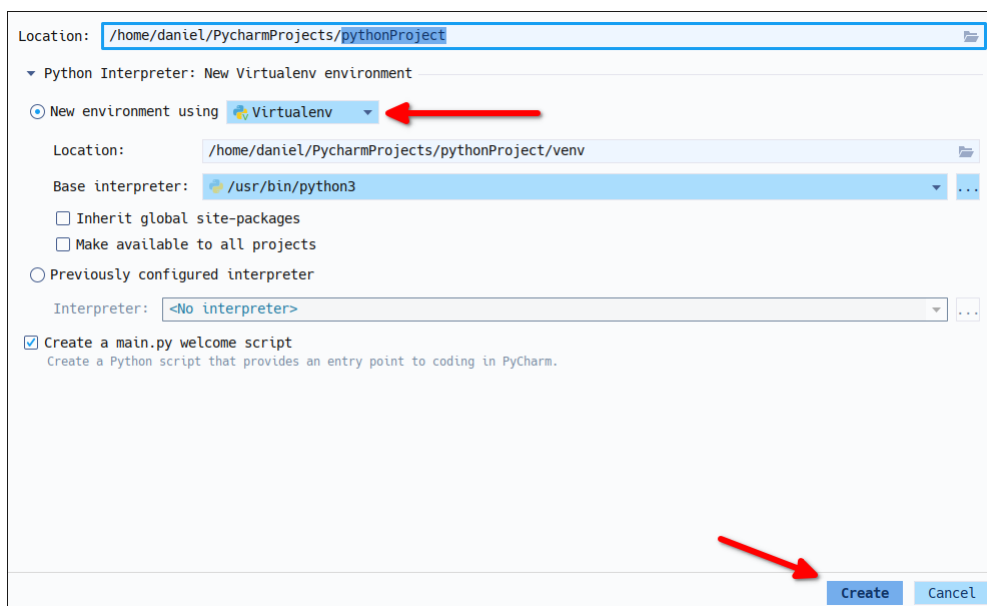
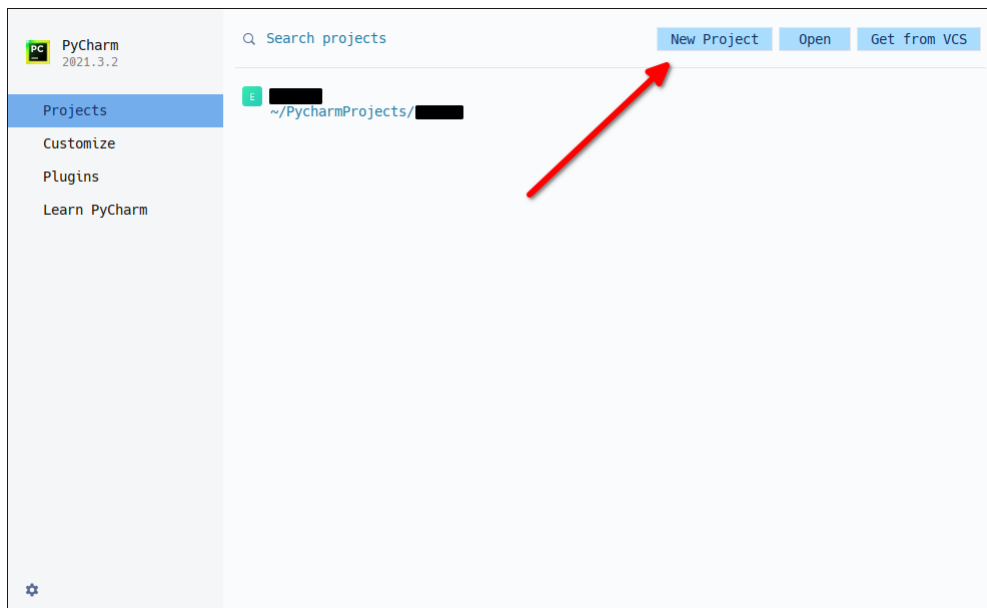
For beginner users and those who wish a working ready-to-use experience, several Integrated Development Environments (IDEs) are available. An IDE is a program which provides services to facilitate computer programming, such as syntax highlighting, code autocompletion, easy documentation access, and more. We recommend the community edition of [PyCharm](#) because it is free and can automatically manage virtual environments. Other common IDEs are [Spyder](#), and [Visual Studio Code](#). However, any IDE or text editor will be more than enough.

2.3 Virtual Environments

Python is a modular language, which means any number of third-party libraries can be installed on a system. This allows the use of an unlimited amount of tools created by the community for all sorts of tasks. That, however, can give rise to conflicts if, for example, a project in which someone is working requires version 2 of a library, but a different project requires version 3. In order to solve this issue, as of Python 3.3 the *virtualenv* module is included with every standard Python installation.

The *virtualenv* module allows the creation of *virtual environments*: containers isolated from the system-wide install which have their own Python executable and can have their independent set of installed Python packages. These packages are accessible only from within the virtual environment, and thus avoid version conflicts as long as projects with different version requirements are hosted in their own separate environments.

The creation of virtual environments is simple. And in fact the PyCharm IDE makes it almost automatic. To create a virtual environment in PyCharm one needs only to select the “New Project” button, choose a directory in which to create it, select “New environment using: Virtualenv”, and then click on “Create”:



To create a virtual environment without an IDE it is necessary to use the terminal or command line. On windows it can be accessed by pressing WindowsKey + R, and writing "cmd" on the prompt. Clicking "Ok" will then open the terminal.

On Mac, the terminal can be opened by clicking on the Launchpad icon, typing "terminal" and clicking on the Terminal program.

On Linux, I trust you.

Once the terminal is open, a virtual environment can be created by running the command

```
python -m venv [directory]
```

where [directory] is the path to the directory in which you will create the environment. After the virtual environment is created, it can be activated by running a script installed by venv. On windows, this is done with

```
myenv\Scripts\activate.bat
```

on either Mac or Linux this is done with

```
source myenv/bin/activate
```

where “myenv” is the name of the virtual environment directory just created.

After the environment is activated, any packages that are installed will only be so in the local environment. Python scripts that are run while inside the environment will be able to use any packages installed in it.

To deactivate or exit a virtual environment one only needs to run “deactivate” in the terminal. PyCharm automatically opens and closes virtual environments when opening and closing projects.

2.4 Python basics

Providing an introductory course to Python is not the purpose of this guide, but aiming to be as accessible as possible, the very basics needed to understand and use the MedPCPy library will be explained.

2.4.1 Data types

Several data types are available in Python. The most common ones are **strings**, **integers**, **floats**, **lists**, **tuples**, **dictionaries**, and **booleans**.

Name	Type	Description	Examples
Strings	str	Ordered sequence of characters	"Hello, world" "123"
Integers	int	Whole numbers	0 399 1996
Floating points	float	Numbers with a decimal point	0.0 3.1415
Lists	list	Ordered sequence of objects	[1, 3.6, "one"]
Tuples	tup	Non-modifiable list	(10, "hello", 1.4)
Dictionaries	dict	Ordered Key:Value pairs	{"name": "Giorno", "age": 15}
Boolean	bool	Logical True/False values	True False

2.4.2 Variable assignment

All data types can be assigned to variables to make them more manageable.

Python

```
1 my_string = "Hello, world!"
2 my_int = 745
3 my_float = 100.0
4 my_list = [my_string, my_int, my_float]
5 my_tuple = (my_string, my_int, my_float, my_list)
6 my_dict = {
7     "name": "Joseph",
8     "nationality": "english",
9     "age": 18
10 }
11 my_bool = True
```

As you can see, having named variables instead of crude data is more readable. It also allows the use of a previously declared variable on the declaration of another one.

2.4.3 Nesting

Lists, tuples, and dictionaries allow nesting other lists, tuples or dictionaries within them with no limit to how deep the nesting can go.

Python

```
1 list_1 = [1, 2, 3, 4]
2 list_2 = [5, 6, 7, 8]
3 super_list = [list_1, list_2]
```

In this example the variable `super_list` would contain two items, both of which would be lists. I.e., the code above is equivalent to:

Python

```
1 super_list = [[1, 2, 3, 4], [5, 6, 7, 8]]
```

And the same is true for either tuples or dictionaries:

Python

```
1 super_tuple = (("a", "b", "c"), "d")
2 dictionary = {
3     "names": ["Jonathan", "Joseph", "Giorno"],
4     "nationalities": ["english", "american", "italian"]
5 }
```

This is also not limited to the same type of variable. That is, lists can contain tuples and dictionaries, tuples can contain lists and dictionaries, and so on.

Dictionaries can even be nested inside other dictionaries:

Python

```
1 planets = {
2     "earth": {"color": "blue", "population": "lots"},
3     "mars": {"color": "red", "population": "none"}
4 }
```

The analysis list used in the MEDPCPY library is a list comprised of dictionaries with other, nested dictionaries:

Python

```
1 analysis_list = [
2     # Measure 1
3     {"fetch": {"cell_row": 1,
4               "cell_column": 1,
5               "sheet": "Measure1",
6               "summary_column_dict": measure1_cols,
7               "offset": 0,
8               }},
9     ...
10    ]
```

Be mindful that the use of space in the declaration of lists, tuples or dictionaries has no syntactic purpose. The distribution is only to increase readability.

2.4.4 Functions

Functions are blocks of reusable code which perform actions that are needed frequently. They are declared only once but can be executed any time they are called. They help increase the readability and modularity of code.

Functions are declared with the **def** keyword, followed by the name which the user desires to assign to it, optionally the arguments in parentheses, and a semicolon. They can either only perform an action, or return a value which can then be assigned to a variable.

Python

```
1 def hello_function():
2     print("Pizza Mozzarella")
3
4 def my_sum_function(a, b):
5     return a + b
```

Then they are called with their name and the necessary arguments:

Python

```
1 hello_function()
2
3 x = 20
4 y = 30
5 result = my_sum_function(x, y)
6 print(result)
```

This would result in printing the string "Pizza Mozzarella", assigning the value 50 to the `result` variable, and then printing that variable as well:

Output

```
Pizza Mozzarella
50
```

Functions can have positional arguments (identified by their position) and keyword arguments (identified by a keyword). Additionally, if a default value for an argument is provided when the function is defined, then that argument becomes optional and, if it is not declared when the function is called, it will take the default value.

Python

```
1 def another_function(name = "Ford Prefect"):
2     print(f"Hello, my name is {name}")
3
4 another_function("Arthur Dent")
5 another_function()
```

This would print the following:

Output

```
Hello, my name is Arthur Dent
Hello, my name is Ford Prefect
```

Given that the optional argument `name` took the default value `Ford Prefect` on the second calling of the function.

2.4.5 Classes and objects

Python is an object-oriented programming language, which means that a good deal of its functionality is built around “objects” which contain data and code. In object-oriented programming the code is organized into reusable “blueprints” called **classes**, and those classes can be used to create individual instances of **objects**, which share the general structure of the class, but have their own particular content. Objects can have properties (data) and methods (functions) associated to them.

A class is declared with the **class** keyword, the name of the class, and a semicolon. Class attributes are declared in a special function called `__init__`, and methods associated with the class are declared as functions inside of it.

Python

```
1 class Dog:
2     def __init__(name, age):
3         self.dog_name = name
4         self.dog_age = age
5
6     def print_name():
7         print(f"""
8         Hello, my name is {self.dog_name},
9         and I am {self.dog_age} years old. Woof!
10        """)
```

Afterwards, the class can be used as a blueprint to create individual objects with the general properties of the class.

Python

```
1 my_dog = Dog("John", 5)
```

The creation of the object is similar to the calling of a function. In this example the end result would be an object of class `Dog` being assigned to the variable `my_dog`. This object would have two attributes and a method associated. The attributes can be changed with the notation `object.attribute = [new value]`.

Python

```
1 my_dog.dog_name = "Cat"
```

And the method can be used with the notation `object.method()`.

Python

```
1 my_dog.print_name()
```

This would result in printing the phrase

Output

```
Hello, my name is Cat,  
and I am 5 years old. Woof!
```

The core of the MedPCPy library is an object of a particular class called **Analyzer**. It has methods to get the sessions associated with a subject, convert files to .xlsx format, create documents, and compute the appropriate analyses.

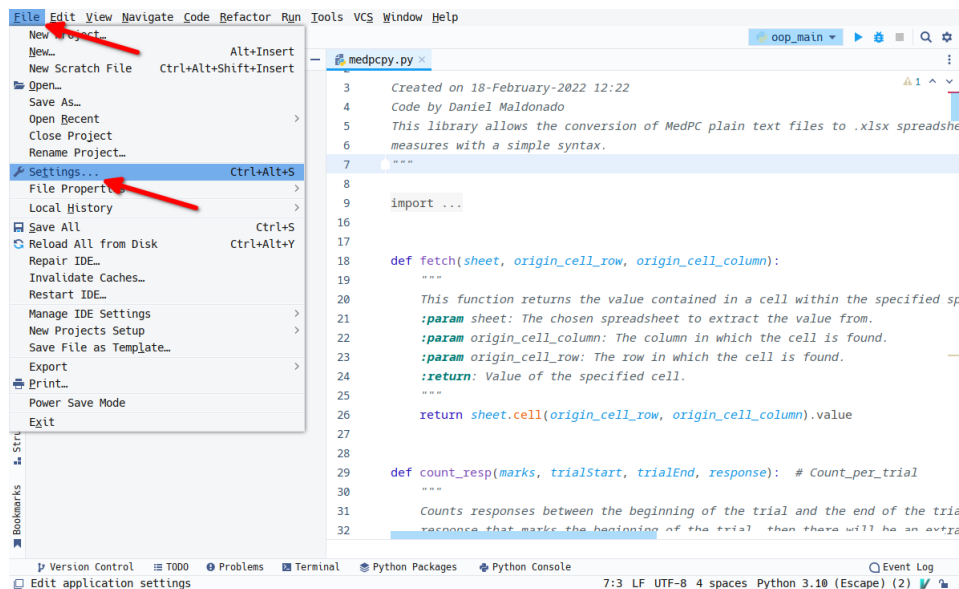
2.4.6 Installing libraries

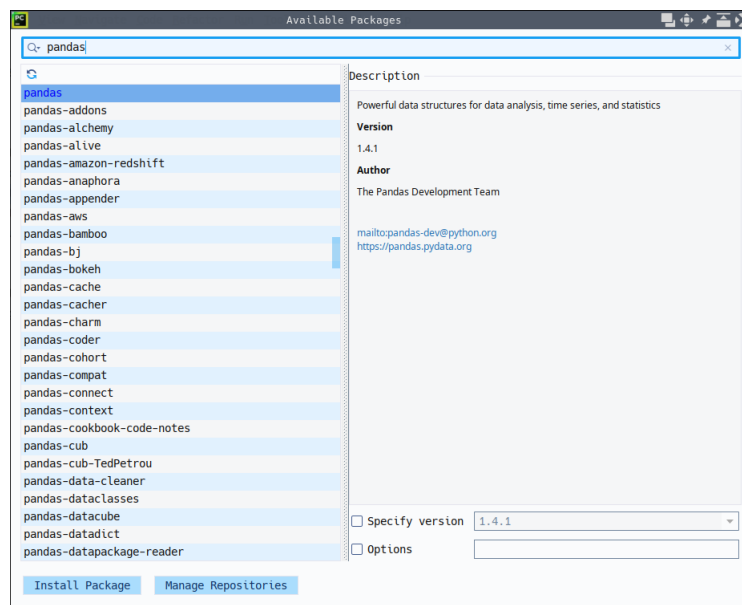
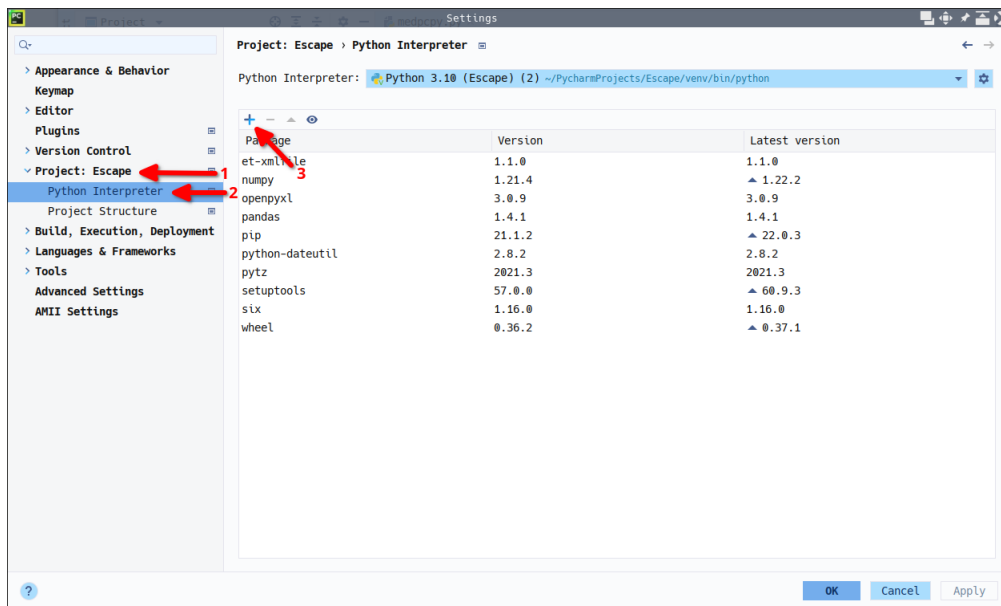
Python is a modular language, which means that its functionality can be extended by including code created by the community. This code is shared through libraries which users can easily install into their system and import into their projects.

Libraries can be installed from within the virtual environment by running the command

```
pip install [library-name]
```

on the terminal. PyCharm can handle the installation of libraries in a longer but more intuitive way if one is not accustomed to working in the terminal. To install a new library navigate to **File → Settings → Project → Python interpreter**. There, click on the + symbol, type the name of the library you wish to install, and click “Install”.





3 This library




To install this library run the command

```
pip install medpcpy
```

on the terminal, or install it from the interface of your IDE of choice. It will install all its required dependencies by default.

3.1 First steps

The first thing to do is to check that your system is ready to begin the process of file conversion and data extraction. To do this, you must create three separate folders in any part of your system that you desire. These folders will contain the raw files before the analysis (TemporaryDir), the raw files after the analysis (PermanentDir), and the already converted and processed .xlsx files (ConvertedDir). The reason why two separate folders for raw files are needed is that the library scans the temporary folder and uses its contents to determine the names of the subjects with at least one associated file and the numbers of the sessions to analyze. If all files are contained in the temporary folder, then all files will be processed every time the script is run. That is the reason why only new, unprocessed files must be placed in the temporary folder. It is worth noting that the directory names are arbitrary.

Name	Size	Type	Date Modified
 ConvertedDir	0 items	Folder	Mon 28 Feb 2022 12:04:25 PM CST
 PermanentDir	0 items	Folder	Mon 28 Feb 2022 12:04:20 PM CST
 TemporaryDir	0 items	Folder	Mon 28 Feb 2022 12:04:13 PM CST

Then, you must check that your files' names follow the format

```
[subject name][separator][session number]
```

to guarantee that the library is able to read them. The **separator** is a character or group of characters that, as you can guess, separates the subject name and session number. It can be any character or group of characters, but there must be a separator. For example, if the files were named "Subject_1_session_1", "Subject_1_session_2", etc., the **separator** would be "_session_", since that is what is placed between the subject name and the session number. More specifically, "Subject_1" would be the subject name, "_session_" would be the separator, and "1" would be the session number.

3.2 Importing the library

The MedPCPy library can be imported to the current working script with the line

Python

```
1 from medpcpy import *
```

The star "*" makes all the functions and classes available to the current project without the need to write the name of the library every time they are used.

3.3 Declaring variables

Several variables have to be declared before the **Analyzer** object. These are the summary file name, directory paths, subject names, column dictionaries, sheet names, and the analysis list. A full explanation of their requirements (particularly for the column dictionaries and the analysis list, whose names are less intuitive) is provided in the [readme](#) file of the GitHub [repository](#). Here only an example declaration is provided:

Python

```
1 file = "your_summary_filename.xlsx"
2 temp_directory = "/path/to/your/temporary/directory/"
3 perm_directory = "/path/to/your/permanent/raw/directory/"
4 conv_directory = "/path/to/your/converted/directory/"
5 subjects = ["Subject1", "Subject2", "Subject3"]
6 measure1_cols = {"Subject1": 2, "Subject2": 4, "Subject3": 6}
7 sheets = ["Measure1"]
8
9 analysis_list = [
10     # Measure 1
11     {"fetch": {"cell_row": 1,
12                "cell_column": 1,
13                "sheet": "Measure1",
14                "summary_column_dict": measure1_cols,
15                "offset": 0,
16                }},
17 ]
```

After that, the **Analyzer** object can be declared, though it won't yet be ready to analyze data.

Python

```
1 my_analyzer = Analyzer(fileName=file,  
2                       temporaryDirectory=temp_directory,  
3                       permanentDirectory=perm_directory,  
4                       convertedDirectory=conv_directory  
5                       subjectList=subjects, suffix="_",  
6                       sheets=sheets,  
7                       analysisList=analysis_list,  
8                       relocate=False)
```

After its declaration, the `Analyzer` object must be used to convert at least one file to .xlsx format with the `.convert()` method.

Python

```
1 my_analyzer.convert()
```

Then, any one of the converted files must be manually inspected in order to determine the columns in which both the time and the marks (explained in the [readme](#) file) are located. Each mark is associated with a particular time of occurrence since the beginning of the session. They are assumed to be written on a Med array with the format “TIME.MARK”, and after conversion they are written on two adjacent columns of the same length.

After that, the `Analyzer` object can be fully declared with the mark and time columns, and is now ready to analyze data.

Python

```
1 my_analyzer = Analyzer(fileName=file,  
2                       temporaryDirectory=temp_directory,  
3                       permanentDirectory=perm_directory,  
4                       convertedDirectory=conv_directory  
5                       subjectList=subjects, suffix="_",  
6                       sheets=sheets,  
7                       analysisList=analysis_list,  
8                       timeColumn="0",  
9                       markColumn="P",  
10                      relocate=False)
```

3.4 Final use

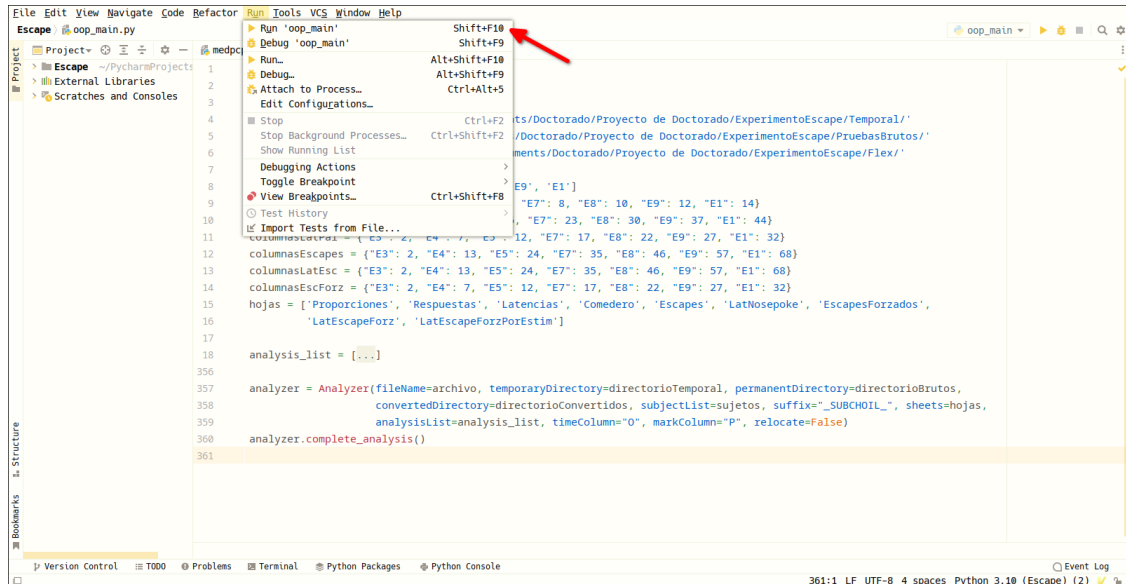
After the `Analyzer` object is declared with all its relevant variables, the analysis of an arbitrary number of sessions and subjects can be done with a single click. The analysis

is called with the `.complete_analysis()` method:

Python

```
my_analyzer.complete_analysis()
```

To run the analysis the user only needs to place their files on the temporary directory and run the script. This can be done by going to the “Run” menu and clicking on “Run [name of your script]” or by pressing **Shift + F10** (or **Ctrl + r** on Mac).



Individual `.xlsx` files with labeled lists of all declared measures will be created for each session of every subject.

	A	B	C	D	E	F	G	H	I	J	K	L
1	FrontLever1		FrontLever2		BackLever1		BackLever2		LatFrontLevers		LatBackLevers	
2	7		3		7		8		0.65		0.45	
3	1		4		7		1		0.6		0.45	
4	2		1		4		3		1.05		0.45	
5	1		1		3		9		0.6		0.45	
6	2		2		2		5		0.9		0.4	
7	3		2		2		5		0.9		0.45	
8			1				2		1.2		0.45	
9			3				6		0.9		0.45	
10			2				4		2.4		0.4	
11			1				2		1		0.4	
12			1				6		1.1		0.45	
13			1				2		0.85		0.4	
14			1				3		0.75		0.55	
15			4				3		1.5		0.4	
16			3				1		1.7		0.4	
17			1				2		1.3		0.35	
18			3				4		1.45		0.4	
19			1				4		0.95		0.45	
20			0				1		1.6		0.35	
21			1				5		1.8		0.45	
22			2				3		1.05		0.5	
23			0				6		1		0.5	
24			1				1		2.2		0.55	
25			2				2		3.3		0.45	
26									1		0.45	
27									1.55		0.45	
28									0.8		0.45	
29									1.8		0.45	
30									1.6		0.5	
31									0.95		0.45	
32												
33												

And the summary file will contain central tendency measures. Each session will occupy one row (or column) in the summary file, and each column (or row) will contain the data

for a single measure of a single subject. The labels of the summary file columns seen in the example are not automatically created. Rather, they must be manually written by the user.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	
1																					
2																					
3																					
4		10.83333	2.5	11.16667	9.333333				12.66667	0.625	17.83333	17.29167				10.66667	4.125	10.66667	11.91667		
5		13.16667	2.666667	10.5	9.875				11.66667	0.916667	15.66667	16.79167				10	5	12.66667	13.25		
6		14.25	5	9.5	11.16667				12.66667	1.125	17.16667	18.08333				7.333333	4.458333	20.33333	16.125		
7		13.16667	2.285714	11.75	8.727273				12.33333	1.333333	20	16.79167				12.33333	3.75	16.66667	15.79167		
8		12.5	1.541667	10.83333	9.333333				12.66667	1	18.66667	17.45833				12.66667	5.125	14.83333	13.95833		
9		12.5	1.789474	12.16667	10.52941				13.16667	0.875	17.5	14.875				10.66667	3.25	14.33333	15		
10		10.83333	2.75	12.16667	10.45833				13.83333	1.458333	16.5	16.33333				9.4	5.833333	16.33333	15.29167		
11		9.833333	1.958333	10.33333	9.666667				13.16667	1.25	18.16667	15.66667				12.83333	3.666667	17.66667	17.04167		
12		12.83333	2.291667	12.9	291667				14.5	1	18.33333	15				9	3.125	15.83333	14.33333		
13		13.5	2.458333	11.83333	9.041667				14.33333	1.583333	19.83333	18.95833				12.5	3.333333	15.83333	16		
14		12.83333	2	9.5	8.208333				16.5	1.208333	18.5	19.625				10.5	3.208333	15.5	15.45833		
15		14.66667	2.291667	10	10.25				14.33333	0.791667	17.66667	16.875				11.16667	5.083333	13.83333	11.66667		
16		12.83333	3.375	10.5	9.083333				12	1.416667	15.5	15.29167				10.83333	3.25	11.16667	10.625		
17		12.66667	4.791667	11.16667	9.791667				15.16667	1.083333	14.16667	15.79167				10.33333	3.583333	11.5	12.20833		
18		9.833333	3.208333	10.66667	10.125				16.16667	1.166667	17.33333	15.58333				10.66667	3.125	9.833333	10.91667		
19		13.66667	3.833333	10.5	9.541667				14.33333	1.041667	15	16				9.333333	6.916667	11.66667	14.45833		
20		14.16667	3.041667	11.9	083333				12.33333	0.958333	14.83333	14.08333				9.666667	4.625	8.333333	10.875		
21		2	0	9	12				14.83333	1.666667	16.5	15.375				10.16667	3.625	10.5	10.66667		
22		12.6	541667	11.66667	8.916667				0	0	0	19				9.5	4.875	13.16667	13.875		
23		13.5	3.083333	9.666667	8.875				15.5	1.083333	16	16.16667				8.666667	5.083333	11.66667	13.75		
24		11.33333	4.25	9.166667	8.708333				11.83333	0.75	16.83333	16.95833				6.833333	5.208333	15.5	15.79167		
25		11.83333	4.583333	11.16667	8.666667				12.16667	0.5	13.5	15.29167				8.166667	5.708333	15.5	15.875		
26		11.33333	3	9	9.125				11	1.083333	15.83333	15.91667				9.5	5.833333	18	15.58333		
27		11.25	2.5	8	8.8125				12	0.458333	13	15.33333				6.833333	4.833333	6.666667	7.708333		
28		9	1.611111	9.25	9.666667				12.16667	1.291667	15	15.08333				7.833333	5.583333	11.33333	10.08333		
29		12	1.434783	11.2	9				10.5	1.375	16.83333	13.79167				6.5	4.416667	7	8.791667		
30		12.66667	2.333333	9	9				9.5	1.791667	16	15				6.5	3.5	11.5	11.54167		
31		11.5	2.375	10.66667	8.5				11.5	0.75	17.83333	15.33333				5.333333	3.041667	7.5	7.625		
32		9.333333	3.458333	9	7.375				11.16667	2.041667	14.33333	15.16667				8	2.125	8	8.458333		
33		11.66667	3.5	7.666667	8.458333				9.666667	1.5	15.16667	14.75				5.666667	1.291667	10.33333	11.41667		
34																					
35																					
36																					
37																					
38																					
39																					
40																					
41																					
42																					
43																					

4 Acknowledgments

- Gabriela E. López-Tolsa, contributor, editor and tester. [Her Github](#)
- Rodrigo Uriel González-Torres, contributor, editor and tester. [His Github](#).

5 Other resources

- The official [Python documentation](#).
- Al Sweigart's [Automate the Boring Stuff with Python](#).
- Angela Yu's [100 Days of Code](#).

The team behind MedPCPy is not affiliated with any of these providers. However, their material proved to be useful to develop this library.