

# MACHINE LEARNING WITH PYTHON

## WORKSHOP

### LAB REFERENCE

\*\*\*\*\* Jupyter / Pandas / Matplotlib / Seaborn \*\*\*\*\*

## 1 Jupyter Notebook / JupyterLab

Home page

<https://jupyter.org/>

Official documentation

<https://docs.jupyter.org/en/latest/>

### 1.1 Installation and execution

The most direct way to install is to use Pip

<https://jupyter.org/install>

With Anaconda distribution, Jupyter should already be preinstalled (check with `conda list` from an Anaconda prompt), so you can run it immediately

<https://problemsolvingwithpython.com/02-Jupyter-Notebooks/02.04-Opening-a-Jupyter-Notebook/>

Opening a Jupyter Notebook:

Note: When Jupyter is started (regardless of how this is achieved), an Anaconda / DOS command prompt is opened to show the status of the Jupyter server, including the URLs and files to access the main dashboard page at the server.

Note that you must use those URLs directly to access the main dashboard tree in a new tab, if you just type: <https://localhost:8888/tree>, you will still be prompted for the token show above.

When you start Jupyter, you will see the Notebook dashboard which shows you the files and sub-folders contained within Jupyter's start-up directory. If you are starting Jupyter from Anaconda

Navigator or directly from the Windows start menu, then the startup directory will typically be `C:\Users\Username`, and can be seen from the output of the prompt:

```
[I 17:20:37.110 NotebookApp] Serving notebooks from local directory: C:\Users\User
```

If you wish the start up directory to be a different folder (which will usually be the case most of the time), start Jupyter via the command `jupyter notebook` in a DOS / Anaconda prompt. In this case, the current working directory of the prompt becomes the startup directory. You can change this startup directory (or notebook working directory)

When a Jupyter notebook is saved as a `*.ipynb` file, it will be relative to this working directory.

Occasionally, when interacting with the Jupyter at the main dashboard or notebook, commands may not respond or behave as expected. This may be due to the Jupyter server not responding, so switching to the Anaconda / DOS command prompt showing the status of the server and pressing Enter a few times should resolve the matter.

## 1.2 General overview

<https://www.dataquest.io/blog/jupyter-notebook-tutorial/>

<https://problemsolvingwithpython.com/02-Jupyter-Notebooks/02.05-The-Jupyter-Notebook-Interface/>

<https://problemsolvingwithpython.com/02-Jupyter-Notebooks/02.07-Getting-Help-in-a-Jupyter-Notebook/>

## 1.3 Markdown guide for markdown cells

<https://www.datacamp.com/tutorial/markdown-in-jupyter-notebook>

<https://medium.com/analytics-vidhya/the-ultimate-markdown-guide-for-jupyter-notebook-d5e5abf728fd>

## 1.4 Magic commands

<https://towardsdatascience.com/top-8-magic-commands-in-jupyter-notebook-c1582e813560>

<https://www.aboutdatalog.com/post/top-8-magic-commands-in-jupyter-notebook>

# 2 Pandas

General reference

<https://python-course.eu/numerical-programming/introduction-to-pandas.php>

<https://sparkbyexamples.com/python-pandas-tutorial-for-beginners/>

<https://www.w3resource.com/pandas/index.php>

Official documentation

[https://pandas.pydata.org/docs/user\\_guide/index.html](https://pandas.pydata.org/docs/user_guide/index.html)

<https://pandas.pydata.org/docs/reference/index.html>

## 2.1 Basics

<https://www.learndatasci.com/tutorials/python-pandas-tutorial-complete-introduction-for-beginners/>

Topic: What's Pandas for, How does pandas fit into the data science toolkit?

<https://www.w3resource.com/pandas/index.php>

What is Pandas

Pandas is built on top of NumPy, which means the NumPy package is a dependency of it (which is imported internally). For e.g. the values of a column in a DataFrame is of type [ndarray](#)

<https://www.tutorialspoint.com/does-pandas-depend-on-numpy>

Differences between pandas and numpy

<https://www.geeksforgeeks.org/difference-between-pandas-vs-numpy/>

<https://www.learndatasci.com/tutorials/python-pandas-tutorial-complete-introduction-for-beginners/>

Core components of pandas: Series and DataFrames

## 2.2 Series

### 2.2.1 Creating a series

<https://sparkbyexamples.com/pandas/create-pandas-series-in-python/>

<https://www.w3resource.com/pandas/series/series.php>

### 2.2.2 Accessing values from a series

<https://sparkbyexamples.com/pandas/get-value-from-pandas-series/>

## 2.3 DataFrames

General list of all functions / properties

[https://www.w3schools.com/python/pandas/pandas\\_ref\\_dataframe.asp](https://www.w3schools.com/python/pandas/pandas_ref_dataframe.asp)

<https://www.w3resource.com/pandas/dataframe/dataframe.php>

### 2.3.1 Data types for DataFrames

To read and understand more

[https://pbpython.com/pandas\\_dtypes.html](https://pbpython.com/pandas_dtypes.html)

<https://medium.com/dunder-data/data-types-and-missing-values-aa0003e51b27>

### 2.3.2 Properties / attributes

List of main properties and summary of their use

<https://www.plus2net.com/python/pandas-dataframe-attributes.php>

<https://www.geeksforgeeks.org/dataframe-attributes-in-python-pandas/>

<https://pythontic.com/pandas/dataframe-attributes/introduction>

#### 2.3.2.1 *columns*

[https://www.w3schools.com/python/pandas/ref\\_df\\_columns.asp](https://www.w3schools.com/python/pandas/ref_df_columns.asp)

Other approaches to obtaining column names / labels of a dataframe

<https://www.geeksforgeeks.org/how-to-get-column-names-in-pandas-dataframe/>

#### 2.3.2.2 *index*

[https://www.skytowner.com/explore/pandas\\_dataframe\\_index\\_property](https://www.skytowner.com/explore/pandas_dataframe_index_property)

<https://pythonexamples.org/get-index-of-pandas-dataframe/>

<https://sparkbyexamples.com/pandas/pandas-get-index-from-dataframe/>

This also contains other ways to obtain the index

#### 2.3.2.3 *values*

<https://www.w3resource.com/pandas/dataframe/dataframe-values.php>

[https://www.skytowner.com/explore/pandas\\_dataframe\\_values\\_property](https://www.skytowner.com/explore/pandas_dataframe_values_property)

Note:

<https://towardsdatascience.com/pandas-df-to-numpy-array-c1a9e7d8585f>

Topic: Avoid using df.values

#### 2.3.2.4 *ndim*

<https://www.w3resource.com/pandas/dataframe/dataframe-ndim.php>

[https://www.skytowner.com/explore/pandas\\_dataframe\\_ndim\\_property](https://www.skytowner.com/explore/pandas_dataframe_ndim_property)

Note: This is only always either 1 (for Series) or 2 (for DataFrame). Multi-dimensional structures (more than 2) are only available in NumPy

#### 2.3.2.5 *shape*

<https://www.w3resource.com/pandas/dataframe/dataframe-shape.php>

<https://pythonexamples.org/pandas-dataframe-shape/>

#### 2.3.2.6 *size*

<https://www.w3resource.com/pandas/dataframe/dataframe-size.php>

[https://www.skytowner.com/explore/pandas\\_dataframe\\_size\\_property](https://www.skytowner.com/explore/pandas_dataframe_size_property)

<https://blog.hubspot.com/website/pandas-dataframe-size>

This demonstrates accessing size, shape and ndim

#### 2.3.2.7 *dtypes*

<https://www.geeksforgeeks.org/python-pandas-dataframe-dtypes/>

<https://datatofish.com/data-type-pandas-dataframe/>

#### 2.3.2.8 *at, iat*

<https://www.w3resource.com/pandas/dataframe/dataframe-at.php>

<https://www.geeksforgeeks.org/python-pandas-dataframe-at/>

[https://www.skytowner.com/explore/pandas\\_dataframe\\_at\\_property](https://www.skytowner.com/explore/pandas_dataframe_at_property)

<https://www.w3resource.com/pandas/dataframe/dataframe-iat.php>

<https://www.geeksforgeeks.org/python-pandas-dataframe-iat/>

<https://dataindependent.com/pandas/pandas-set-dataframe-values-at-iat/>

`at` accesses a single value based on row / column labels lookup (similar to `loc`)

`iat` accesses a single value based on row / column integer positions (similar to `iloc`)

You can use these two properties to both access and set a single value in the DataFrame.

The main point here is that this approach only references a single value, and if you wish to reference multiple rows / columns, then use either [loc / iloc](#)

#### 2.3.2.9 *loc, iloc*

To select rows / columns based on integer indexing, use the `.iloc` function.

To select rows / columns based on label names, can use the `.loc` function.

[https://www.w3schools.com/python/pandas/ref\\_df\\_iloc.asp](https://www.w3schools.com/python/pandas/ref_df_iloc.asp)

<https://www.shanelynn.ie/pandas-iloc-loc-select-rows-and-columns-dataframe/>

<https://sparkbyexamples.com/pandas/pandas-iloc-usage-with-examples/>

<https://www.w3resource.com/pandas/dataframe/pandas-dataframe-iloc.html>

Note:

Arguments to `iloc` must be a numeric index value for either row or column portion

[https://www.w3schools.com/python/pandas/ref\\_df\\_loc.asp](https://www.w3schools.com/python/pandas/ref_df_loc.asp)

<https://www.w3resource.com/pandas/dataframe/dataframe-loc.php>

<https://towardsdatascience.com/a-python-beginners-look-at-loc-part-1-cb1e1e565ec2>

Note:

`.loc` will always interpret integers as labels, not as integer positions of the index (you should use `.iloc` for that purposes).

There is another option called `ix` which is now deprecated

<https://deeplearning.lipingyang.org/pandas-iloc-ix-loc-explained/>

## 2.4 Creating a dataframe

The most common ways

<https://towardsdatascience.com/4-different-ways-to-create-a-pandas-dataframe-91bca1ff31bb>

<https://medium.com/analytics-vidhya/5-different-ways-to-create-pandas-dataframe-in-python-aa2a037d9c03>

More comprehensive

<https://www.geeksforgeeks.org/different-ways-to-create-pandas-dataframe/>

## 2.5 Accessing rows / columns

### 2.5.1 head, tail

<https://www.w3resource.com/pandas/dataframe/dataframe-head.php>

[https://www.w3schools.com/python/pandas/ref\\_df\\_tail.asp](https://www.w3schools.com/python/pandas/ref_df_tail.asp)

[https://www.skytownner.com/explore/pandas\\_dataframe\\_tail\\_method](https://www.skytownner.com/explore/pandas_dataframe_tail_method)

<https://www.geeksforgeeks.org/select-first-or-last-n-rows-in-a-dataframe-using-head-and-tail-method-in-python-pandas/>

### 2.5.2 Selecting columns with column labels

<https://pythonhow.com/how/access-column-pandas-dataframe/>

[https://www.skytownner.com/explore/accessing\\_columns\\_of\\_a\\_dataframe\\_using\\_column\\_labels\\_in\\_pandas](https://www.skytownner.com/explore/accessing_columns_of_a_dataframe_using_column_labels_in_pandas)

This demonstrates a variety of approaches in greater details

<https://www.geeksforgeeks.org/how-to-select-multiple-columns-in-a-pandas-dataframe/>

### 2.5.3 Selecting a specific cell

<https://sparkbyexamples.com/pandas/pandas-get-cell-value-from-dataframe/>

### 2.5.4 Boolean indexing

<https://www.geeksforgeeks.org/boolean-indexing-in-pandas/>

<https://lifewithdata.com/2022/02/11/pandas-how-boolean-indexing-works-in-pandas/>

<https://blog.finxter.com/pandas-boolean-indexing/>

## 2.6 Manipulation

### 2.6.1 drop

<https://www.freecodecamp.org/news/dataframe-drop-column-in-pandas-how-to-remove-columns-from-dataframes/>

<https://sparkbyexamples.com/pandas/pandas-drop-columns-examples/>

<https://sparkbyexamples.com/pandas/pandas-drop-first-column-from-dataframe/>

### 2.6.2 get\_dummies

[https://www.geeksforgeeks.org/python-pandas-get\\_dummies-method/](https://www.geeksforgeeks.org/python-pandas-get_dummies-method/)

<https://www.sharpsightlabs.com/blog/pandas-get-dummies/>

[https://dataindependent.com/pandas/pandas-get-dummies-pd-get\\_dummies/](https://dataindependent.com/pandas/pandas-get-dummies-pd-get_dummies/)

### 2.6.3 fillna

<https://www.geeksforgeeks.org/python-pandas-dataframe-fillna-to-replace-null-values-in-dataframe/>

<https://sparkbyexamples.com/pandas/pandas-dataframe-fillna-fill-nan-column-values-2/>

<https://www.kdnuggets.com/2023/02/optimal-way-input-missing-data-pandas-fillna.html>

<https://datagy.io/pandas-fillna/>

### 2.6.4 dropna

<https://www.w3resource.com/pandas/dataframe/dataframe-dropna.php>

<https://sparkbyexamples.com/python/pandas-dropna-usage-examples/>

<https://www.datasciencelearner.com/pandas-dropna-remove-nan-rows-python/>

<https://datagy.io/pandas-dropna/>

### 2.6.5 apply

<https://www.dataquest.io/blog/tutorial-how-to-use-the-apply-method-in-pandas/>

<https://www.digitalocean.com/community/tutorials/pandas-dataframe-apply-examples>

<https://towardsdatascience.com/introduction-to-pandas-apply-applymap-and-map-5d3e044e93ff>



### 2.6.6 reset\_index

[https://thatascience.com/learn-pandas/reset\\_index/](https://thatascience.com/learn-pandas/reset_index/)

<https://pynative.com/pandas-reset-index/>

<https://www.machinelearningplus.com/pandas/pandas-reset-index/>

<https://www.geeksforgeeks.org/pandas-how-to-reset-index-in-a-given-dataframe/>

## 2.7 Conversion

### 2.7.1 to\_numpy

<https://pythonexamples.org/convert-pandas-dataframe-to-numpy-array/>

[https://www.geeksforgeeks.org/pandas-dataframe-to\\_numpy-convert-dataframe-to-numpy-array/](https://www.geeksforgeeks.org/pandas-dataframe-to_numpy-convert-dataframe-to-numpy-array/)

[https://dataindependent.com/pandas/pandas-dataframe-to-numpy-array-df-to\\_numpy/](https://dataindependent.com/pandas/pandas-dataframe-to-numpy-array-df-to_numpy/)

### 2.7.2 squeeze

[https://www.w3schools.com/python/pandas/ref\\_df\\_squeeze.asp](https://www.w3schools.com/python/pandas/ref_df_squeeze.asp)

<https://www.geeksforgeeks.org/python-pandas-series-squeeze/>

### Converting

Note that the squeeze method call is optional when accessing the column directly

### 2.7.3 Converting a DataFrame / column to a NumPy array

<https://towardsdatascience.com/pandas-df-to-numpy-array-c1a9e7d8585f>

Key point: Avoid using the values property to perform the conversion

<https://sparkbyexamples.com/pandas/pandas-convert-column-to-numpy-array/>

<https://sparkbyexamples.com/pandas/convert-pandas-dataframe-to-numpy-array/>

## 2.8 Aggregation / grouping functionality

### 2.8.1 count

<https://www.geeksforgeeks.org/python-pandas-dataframe-count/>

<https://sparkbyexamples.com/pandas/pandas-dataframe-count-function/>

<https://www.w3resource.com/pandas/dataframe/dataframe-count.php>

#### 2.8.2 value\_counts

<https://towardsdatascience.com/9-pandas-value-counts-tricks-to-improve-your-data-analysis-7980a2b46536>

[https://www.sharpsightlabs.com/blog/pandas-value\\_counts/](https://www.sharpsightlabs.com/blog/pandas-value_counts/)

[https://www.w3resource.com/pandas/series/series-value\\_counts.php](https://www.w3resource.com/pandas/series/series-value_counts.php)

#### 2.8.3 groupby

<https://www.geeksforgeeks.org/python-pandas-dataframe-groupby/>

<https://sparkbyexamples.com/pandas/pandas-groupby-explained-with-examples/>

<https://sparkbyexamples.com/pandas/pandas-groupby-count-examples/>

<https://towardsdatascience.com/all-pandas-groupby-you-should-know-for-grouping-data-and-performing-operations-2a8ec1327b5>

#### 2.8.4 crosstab

<https://www.geeksforgeeks.org/pandas-crosstab-function-in-python/>

<https://pbpython.com/pandas-crosstab.html>

<https://dataindependent.com/pandas/pandas-cross-tab-pd-crosstab/>

<https://towardsdatascience.com/3-easy-ways-to-crosstab-in-pandas-4123383bfbf2>

#### 2.8.5 idxmin

[https://www.w3schools.com/python/pandas/ref\\_df\\_idxmin.asp](https://www.w3schools.com/python/pandas/ref_df_idxmin.asp)

<https://www.geeksforgeeks.org/python-pandas-dataframe-idxmin/>

#### 2.8.6 Getting the row count

<https://mljar.com/blog/pandas-dataframe-row-count/>

## 2.9 Visualization functions

### 2.9.1 hist

<https://sparkbyexamples.com/pandas/pandas-plot-a-histogram>

<https://dataindependent.com/pandas/pandas-histogram-dataframe-hist/>

### 2.9.2 plot

<https://www.dataquest.io/blog/plot-dataframe-pandas/>

<https://realpython.com/pandas-plot-python/>

<https://neptune.ai/blog/pandas-plot-deep-dive-into-plotting-directly-with-pandas>

## 2.10 Mathematical functions

### 2.10.1 sum

<https://www.geeksforgeeks.org/python-pandas-dataframe-sum/>

<https://sparkbyexamples.com/pandas/pandas-sum-dataframe-columns/>

<https://dataindependent.com/pandas/pandas-sum-pd-dataframe-sum/>

## 2.11 Statistical functions

### 2.11.1 mean

<https://www.geeksforgeeks.org/python-pandas-dataframe-mean/>

<https://sparkbyexamples.com/pandas/pandas-series-mean-function/>

<https://dataindependent.com/pandas/pandas-mean-get-average-pd-dataframe-mean/>

Getting column average or mean

<https://sparkbyexamples.com/pandas/pandas-get-column-average-mean/>

### 2.11.2 sample

<https://dataindependent.com/pandas/pandas-sample-pd-dataframe-sample/>

<https://www.w3resource.com/pandas/dataframe/dataframe-sample.php>

<https://www.statology.org/pandas-sample-with-replacement/>

<https://datagy.io/pandas-sample/>

## 2.12 General functions

### 2.12.1 info

<https://www.w3resource.com/pandas/dataframe/dataframe-info.php>

<https://www.geeksforgeeks.org/python-pandas-dataframe-info/>

### 2.12.2 describe

<https://www.w3resource.com/pandas/dataframe/dataframe-describe.php>

<https://www.geeksforgeeks.org/python-pandas-dataframe-describe-method/>

## 2.13 Working with strings / text

<https://www.aboutdatablog.com/post/10-most-useful-string-functions-in-pandas>

<https://towardsdatascience.com/using-string-methods-in-pandas-5e4509ff1f5f>

<https://towardsdatascience.com/pandas-string-operations-explained-fdfab7602fb4>

## 2.14 Reading files

### Reading files in general (including CSV)

<https://www.geeksforgeeks.org/how-to-read-text-files-with-pandas/>

#### Detailed outline

<https://realpython.com/pandas-read-write-files>

### 2.14.1 read\_csv

[https://www.geeksforgeeks.org/python-read-csv-using-pandas-read\\_csv/](https://www.geeksforgeeks.org/python-read-csv-using-pandas-read_csv/)

<https://datatofish.com/import-csv-file-python-using-pandas/>

<https://towardsdatascience.com/how-to-read-csv-file-using-pandas-ab1f5e7e7b58>

## 2.15 Overall summary / tutorials

Further tutorials to explore ???

<https://www.learndatasci.com/tutorials/python-pandas-tutorial-complete-introduction-for-beginners/>

<https://www.datacamp.com/tutorial/pandas-tutorial-dataframe-python>

<https://www.kaggle.com/learn/pandas>

[https://pandas.pydata.org/pandas-docs/stable/getting\\_started/tutorials.html](https://pandas.pydata.org/pandas-docs/stable/getting_started/tutorials.html)

[https://www.w3schools.com/python/pandas/pandas\\_csv.asp](https://www.w3schools.com/python/pandas/pandas_csv.asp)

## 3 Matplotlib

General reference

<https://python-course.eu/numerical-programming/overview-of-matplotlib.php>

[https://www.w3schools.com/python/matplotlib\\_intro.asp](https://www.w3schools.com/python/matplotlib_intro.asp)

<https://www.geeksforgeeks.org/matplotlib-tutorial/>

<https://www.datacamp.com/tutorial/matplotlib-tutorial-python>

### 3.1 plot

<https://www.activestate.com/resources/quick-reads/how-to-display-a-plot-in-python/>

### 3.2 subplots

[https://www.w3schools.com/python/matplotlib\\_subplot.asp](https://www.w3schools.com/python/matplotlib_subplot.asp)

<https://www.geeksforgeeks.org/matplotlib-pyplot-subplots-in-python/>

<https://jakevdp.github.io/PythonDataScienceHandbook/04.08-multiple-subplots.html>

### 3.3 scatter

<https://www.geeksforgeeks.org/matplotlib-pyplot-scatter-in-python/>

<https://jakevdp.github.io/PythonDataScienceHandbook/04.02-simple-scatter-plots.html>

<https://realpython.com/visualizing-python-plt-scatter/>

### 3.4 hist

[https://www.w3schools.com/python/matplotlib\\_histograms.asp](https://www.w3schools.com/python/matplotlib_histograms.asp)

<https://www.geeksforgeeks.org/matplotlib-pyplot-hist-in-python/>

<https://towardsdatascience.com/histograms-with-pythons-matplotlib-b8b768da9305>

## 4 Seaborn

### General reference

<https://seaborn.pydata.org/tutorial.html>

<https://www.geeksforgeeks.org/python-seaborn-tutorial/>

<https://elitedatascience.com/python-seaborn-tutorial>

### 4.1 pairplot

<https://www.geeksforgeeks.org/python-seaborn-pairplot-method/>

<https://towardsdatascience.com/seaborn-pairplot-enhance-your-data-understanding-with-a-single-plot-bf2f44524b22>

### 4.2 distplot

<https://www.digitalocean.com/community/tutorials/seaborn-distplot>

<https://www.digitalocean.com/community/tutorials/seaborn-distplot>

### 4.3 heatmap

<https://www.geeksforgeeks.org/seaborn-heatmap-a-comprehensive-guide/>

<https://pythonbasics.org/seaborn-heatmap/>

### 4.4 countplot

<https://www.geeksforgeeks.org/countplot-using-seaborn-in-python/>

<https://www.sharpsightlabs.com/blog/seaborn-countplot/>

<https://towardsdatascience.com/hands-on-python-data-visualization-seaborn-count-plot-90e823599012>

### 4.5 boxplot

<https://www.geeksforgeeks.org/box-plot-visualization-with-pandas-and-seaborn/>

<https://www.geeksforgeeks.org/boxplot-using-seaborn-in-python/>

<https://towardsdatascience.com/creating-boxplots-with-the-seaborn-python-library-f0c20f09bd57>