



Data Science Tools Installation

BRI Audit Analytics

Pre-Workshop Guide
December 2023



Contents

Introduction	2
Installation	3
• Installing Python using <i>Anaconda</i>	
System Verification	6
• Verify Python Installation	
• Verify Conda Installation	
• Verify Jupyter Installation	
Library Installation	10
• Installing Packages and Running Jupyter Notebook	

■ ■ ■



Introduction

This guide is a resource for students at Algoritma to use in setting up their laptop or environment prior to the scheduled workshops. In this guide, students can find a list of prerequisites that will be consistently used throughout the entire course. These prerequisites are required to be **completed before** the start of the workshop.

For new students, we will run through the installation process to ensure that the necessary programming languages and tools - such as Python - are installed. The next section will then talk about methods on how to verify whether the installs were completed successfully.

For recurring students, we recommend repeating the System Verification section once more to confirm past completed installations.



Installation

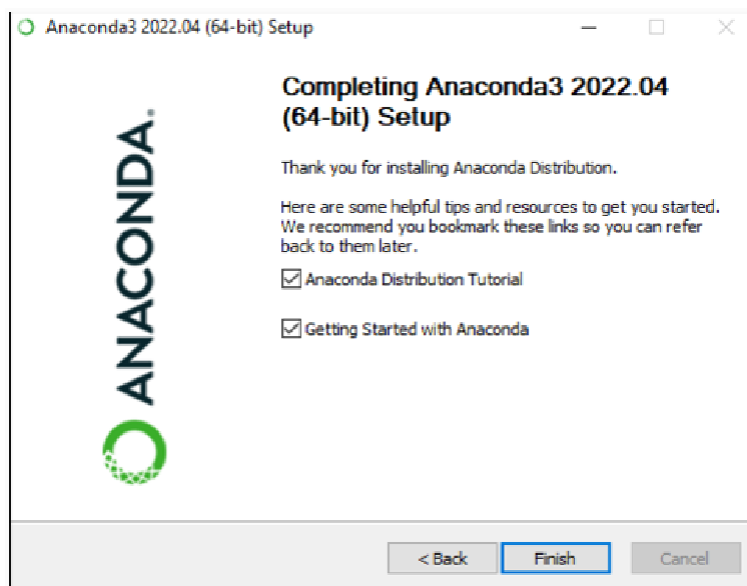
Installing Python using *Anaconda*

For our Python installation, we will be using and installing a package manager named *Anaconda*. With *Anaconda*, users will not only have Python installed but also will have the necessary packages (i.e. numpy, pandas) utilized in our workshops. Also, *Anaconda* would have included the installation of *Jupyter* - an open-source web application that allows you to create and share Python code. For alternative, you can also install *Miniconda*, the minimal Anaconda installer version that includes only conda, Python, the packages they depend on. Once opening the link below, **please choose Python version 3 for installation.**

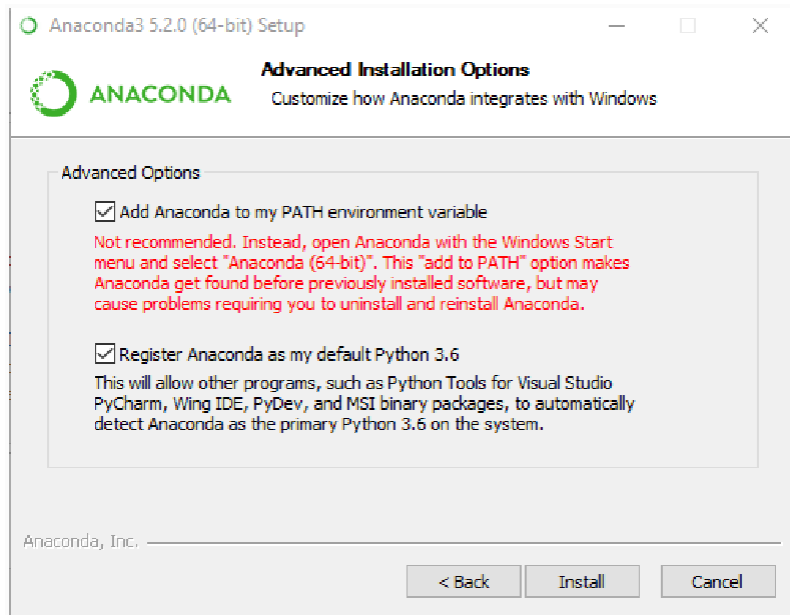
Use this link:

https://repo.anaconda.com/archive/Anaconda3-2022.10-Windows-x86_64.exe, or Choose your appropriate operating system. Make sure you choose **Python 3.10 version**

- Install Anaconda



- For Windows users make sure to check Add Anaconda to my PATH to environment and then wait until the installation finished



More info on *Anaconda*:

<https://docs.continuum.io/anaconda/#anaconda-navigator-or-conda>

Warning: For Windows operating systems, if you can't find the conda command from your Command Prompt please add the C:\User\Anaconda3\ and the C:\User\Anaconda3\Scripts\ to the environment variable as shown here:

<https://superuser.com/questions/949560/how-do-i-set-system-environment-variables-in-windows-10>

■ ■ ■

System Verification

- For Mac OS X and Linux-based OS: Open “*Terminal*”
- For Windows: Open “*Anaconda Prompt*”

Verify Python Installation:

1. Type the command `python`
2. If the installation was completed successfully, there should be a response which includes information on which Python version was installed as shown below. In this case, it appears the user installed Python version 2.7.13. **But make sure yours is 3.10 or above.**
3. To exit, enter the command `quit()` or use Ctrl-D

```
[Matthews-MacBook-Pro:~ matthewhamdani$ python
Python 2.7.13 |Anaconda 4.4.0 (x86_64)| (default, Dec 20 2016, 23:05:08)
[GCC 4.2.1 Compatible Apple LLVM 6.0 (clang-600.0.57)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
Anaconda is brought to you by Continuum Analytics.
Please check out: http://continuum.io/thanks and https://anaconda.org
>>>
```

Figure 1: python Response on Mac OS X Terminal

```
(base) C:\Users\dyahn>python
Python 3.8.3 (default, Jul 2 2020, 17:30:36) [MSC v.1916 64 bit (AMD64)] :: Anaconda, Inc. on win32
Type "help", "copyright", "credits" or "license" for more information.
>>>
```

Figure 2: python Response on Windows Command Prompt

Verify Anaconda Installation

1. Type the command `conda list` in your “Terminal” or “Anaconda Prompt”.
2. If the installation was completed successfully, your terminal will give a response of list of packages like the example below. In this case, the environment has Python version 3.8 pre-installed.
3. If your terminal do not give any response, please check the **Warning** in the installation section, if the problem still persist please contact mentor@algorit.ma via email for further help.

```
[Matthews-MacBook-Pro:~ matthewhamdani$ conda list
# packages in environment at /Users/matthewhamdani/anaconda2:
#
_license                  1.1                      py27_1
alabaster                  0.7.10                   py27_0
anaconda                   4.4.0                    np112py27_0
anaconda-client            1.6.3                    py27_0
anaconda-navigator         1.6.2                    py27_0
anaconda-project           0.6.0                    py27_0
appnope                    0.1.0                    py27_0
appscript                  1.0.1                    py27_0
asn1crypto                 0.22.0                   py27_0
astroid                    1.4.9                    py27_0
astropy                    1.3.2                    np112py27_0
babel                      2.4.0                    py27_0
backports                  1.0                      py27_0
backports_abc              0.5                      py27_0
beautifulsoup4             4.6.0                    py27_0
bitarray                   0.8.1                    py27_0
blaze                      0.10.1                   py27_0
bleach                     1.5.0                    py27_0
bokeh                      0.12.5                   py27_1
boto                       2.46.1                   py27_0
bottleneck                 1.2.1                    np112py27_0
```

Figure 3: conda list Response on Mac OS X Terminal

```
(base) C:\Users\dyahn>conda list
# packages in environment at C:\Users\dyahn\anaconda3:
#
# Name                                Version                                Build      Channel
_ipyw_jlab_nb_ext_conf               0.1.0                                py38_0
alabaster                             0.7.12                               py_0
anaconda                              2020.07                              py38_0
anaconda-client                       1.7.2                                py38_0
anaconda-navigator                    1.9.12                               py38_0
anaconda-project                      0.8.4                                py_0
appnope                               0.1.0                                pypi_0     pypi
argh                                   0.26.2                               py38_0
asn1crypto                            1.3.0                                py38_0
astroid                                2.4.2                                py38_0
astropy                               4.0.1.post1                          py38he774522_1
atomicwrites                           1.4.0                                py_0
attrs                                  19.3.0                               py_0
autopep8                              1.5.3                                py_0
babel                                  2.8.0                                py_0
backcall                               0.2.0                                py_0
backports                              1.0                                   py_2
backports.functools_lru_cache          1.6.1                                py_0
backports.shutil_get_terminal_size     1.0.0                                py38_2
backports.tempfile                     1.0                                   py_1
backports.weakref                       1.0.post1                            py_1
bar-chart-race                         0.1.0                                pypi_0     pypi
bcrypt                                  3.1.7                                py38he774522_1
beautifulsoup4                         4.8.0                                pypi_0     pypi
```

Figure 4: conda list Response on Windows Command Prompt

■ ■ ■

Verify Jupyter Installation:

1. For Mac OS X or Linux-based OS: Type the command `jupyter notebook` in your “Terminal”. For windows OS find your jupyter notebook in the windows search or use `Anaconda Prompt` and type in `jupyter notebook`.
2. If the installation was completed successfully, Jupyter would have started a server connection and automatically opened a new window in your browser.
3. If it does not open automatically, the *Terminal* or *Command Prompt* would have provided a URL link for you to open in your browser manually.
4. If neither of these options occurred, repeat the *Anaconda* installation process.
5. To shutdown the server and exit, use Ctrl-C then type `y` to confirm or `n` to cancel

```
[Matthews-MacBook-Pro:~ matthewhamdani$ jupyter notebook
[I 11:06:53.460 NotebookApp] Serving notebooks from local directory: /Users/matthewhamdani
[I 11:06:53.460 NotebookApp] 0 active kernels
[I 11:06:53.460 NotebookApp] The Jupyter Notebook is running at: http://localhost:8888/?token=61604128eaa7ddfff4048ce8a1e271959149f48584c04363
[I 11:06:53.460 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 11:06:53.461 NotebookApp]

Copy/paste this URL into your browser when you connect for the first time,
to login with a token:
http://localhost:8888/?token=61604128eaa7ddfff4048ce8a1e271959149f48584c04363
0:97: execution error: "http://localhost:8888/tree?token=50a7a657dceba621e9932c7b1ab4873bdf5073a646505819" doesn't
understand the "open location" message. (-1708)
```

Figure 5: jupyter notebook Response on Mac OS X Terminal

```
(base) C:\Users\dyahn>jupyter notebook
[W 13:12:20.273 NotebookApp] Error loading server extension jupyter_http_over_ws
Traceback (most recent call last):
  File "C:\Users\dyahn\anaconda3\lib\site-packages\notebook\notebookapp.py", line 1670, in init_server_extensions
    mod = importlib.import_module(modulename)
  File "C:\Users\dyahn\anaconda3\lib\importlib\_init_.py", line 127, in import_module
    return _bootstrap._gcd_import(name[level:], package, level)
  File "<frozen importlib._bootstrap>", line 1014, in _gcd_import
  File "<frozen importlib._bootstrap>", line 991, in _find_and_load
  File "<frozen importlib._bootstrap>", line 973, in _find_and_load_unlocked
ModuleNotFoundError: No module named 'jupyter_http_over_ws'
[I 13:12:20.312 NotebookApp] [jupyter_nbextensions_configurator] enabled 0.4.1
[I 13:12:20.424 NotebookApp] JupyterLab extension loaded from C:\Users\dyahn\anaconda3\lib\site-packages\jupyterlab
[I 13:12:20.424 NotebookApp] JupyterLab application directory is C:\Users\dyahn\anaconda3\share\jupyter\lab
[I 13:12:20.427 NotebookApp] Serving notebooks from local directory: C:\Users\dyahn
[I 13:12:20.427 NotebookApp] The Jupyter Notebook is running at:
[I 13:12:20.427 NotebookApp] http://localhost:8888/?token=0648ecf70743a074c963e220d7027792e0c4f2e4b55b90f6
[I 13:12:20.427 NotebookApp] or http://127.0.0.1:8888/?token=0648ecf70743a074c963e220d7027792e0c4f2e4b55b90f6
[I 13:12:20.428 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 13:12:20.461 NotebookApp]
```

Figure 6: jupyter notebook Response on Windows Anaconda Prompt

Library Installation

Installing Packages and Running Jupyter Notebook:

1. After the anaconda installation has been verified, please re-open anaconda prompt and install all the required packages:

Data Analytics			
Material	List Library	Version	Installation
<ul style="list-style-type: none">• P4DA• DWV• PS	pandas	2.0.0/Latest	pip install pandas
	numpy	1.26.1/Latest	pip install numpy
	ipykernel	6.24.0/Latest	pip install ipykernel
	matplotlib	3.8.1/Latest	pip install matplotlib
	yfinance	0.2.31/Latest	pip install yfinance
Data Science			
Material	List Library	Version	Installation
<ul style="list-style-type: none">• Regression Model• Unsupervised Learning• Classification in Machine Learning	openpyxl	3.1.2/Latest	pip install openpyxl
	scikit-learn	1.3.2/Latest	pip install scikit-learn
	scipy	1.11.3/Latest	pip install scipy
	seaborn	0.13.0/Latest	pip install seaborn
	statsmodels	0.14.0/Latest	pip install statsmodels
	gower	0.1.2/Latest	pip install gower
	plotly	5.18.0/Latest	pip install plotly
	imbalanced-learn	0.11.0/Latest	pip install imbalanced-learn
	nbformat	5.9.2/Latest	pip install nbformat

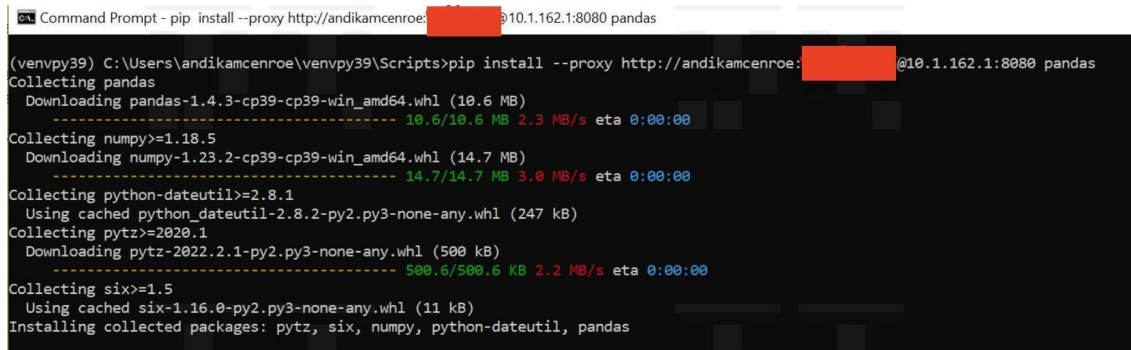
Command Installation:

```
pip install pandas numpy matplotlib ipykernel seaborn openpyxl  
scikit-learn plotly imbalanced-learn gower statsmodels nbformat
```

Optional Command Installation :

If an error occurred during installation using the previous command installation, we can install one by one package that we would like to install.

```
pip install pandas -> install
pip install numpy -> install
etc
```



```
Command Prompt - pip install --proxy http://andikamcenroe:10.1.162.1:8080 pandas

(venvpy39) C:\Users\andikamcenroe\venvpy39\Scripts>pip install --proxy http://andikamcenroe:10.1.162.1:8080 pandas
Collecting pandas
  Downloading pandas-1.4.3-cp39-cp39-win_amd64.whl (10.6 MB)
    ----- 10.6/10.6 MB 2.3 MB/s eta 0:00:00
Collecting numpy>=1.18.5
  Downloading numpy-1.23.2-cp39-cp39-win_amd64.whl (14.7 MB)
    ----- 14.7/14.7 MB 3.0 MB/s eta 0:00:00
Collecting python-dateutil>=2.8.1
  Using cached python_dateutil-2.8.2-py2.py3-none-any.whl (247 kB)
Collecting pytz>=2020.1
  Downloading pytz-2022.2.1-py2.py3-none-any.whl (500 kB)
    ----- 500.6/500.6 KB 2.2 MB/s eta 0:00:00
Collecting six>=1.5
  Using cached six-1.16.0-py2.py3-none-any.whl (11 kB)
Installing collected packages: pytz, six, numpy, python-dateutil, pandas
```

Figure 7: Pip Installation Response on Anaconda Prompt

2. Start your jupyter by run this command: `jupyter notebook`



Figure 8: Jupyter Page on Web Browser