Foundations of Data Science using Python

Python and Libraries for Data Science Installation

Why Python?

Python offers all the skillsets that are required for Machine Learning.

- Simple and Consistent
- Libraries and Frameworks
- Platform Independence
- Great Community Base

Simple and Consistent

Python's simple syntax allows developers to write code that is reliable, concise and readable. The productivity is increased during the overall development of the process. Prototype can be built faster as the complex machine learning tasks and testing process can be done quickly.

Libraries and Frameworks

Another reason to learn python for machine learning is the availability of amazing libraries and frameworks of python. It is nothing but the pre-written code that can be used to solve common programming problem. Python has rich bank of libraries for machine learning, some of them are Tensor-Flow, Keras, Scikit-learn, Numpy, Scipy, Pandas, Seaborn etc. Numpy and Scipy are specifically for scientific and advance computing respectively. Pandas is generally used for Data analysis and Seaborn is specifically for Data visualization.

Platform Independence

It basically means one can freely shift from one machine to another without making changes to the actual code (on with minimal changes). It supports many platforms such as Windows, macOS, and Linux.

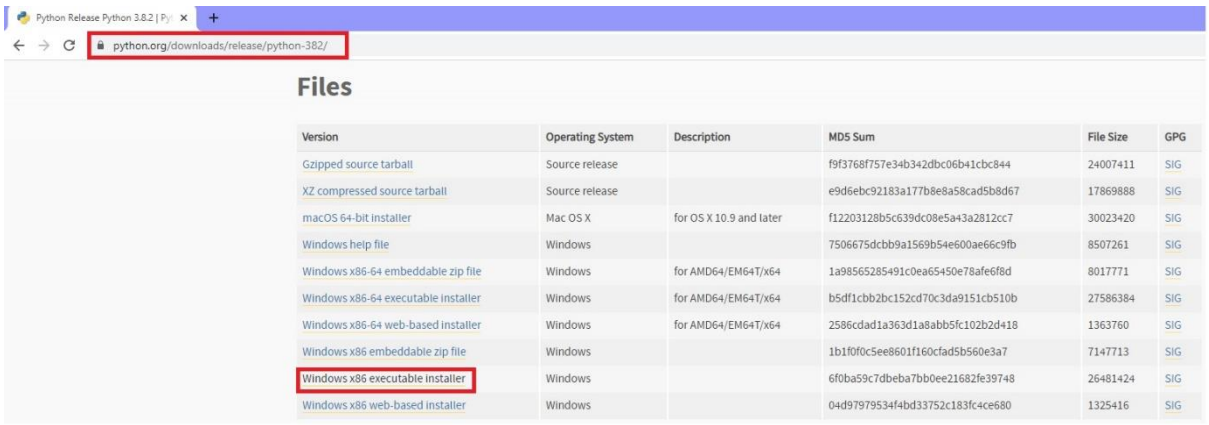
Great Community Base

The python's community has grown across the globe, and especially in the world of machine learning and data science. There are active communities that contribute to the large exchange of information which involves solutions to the problems. For any problem you come across, chances are very high that someone out there has already gone through that same problem and solved it successfully. Hence you can find guidance and bits of advice at any level of doubt. You won't be the only one who went through it. Also, you may come to know some of the best results according to your specifications, all you need to do it to turn to the huge python community.

Python Version 3.8.2 Installation Steps

1. Go to www.python.org/downloads for downloading the latest version of python

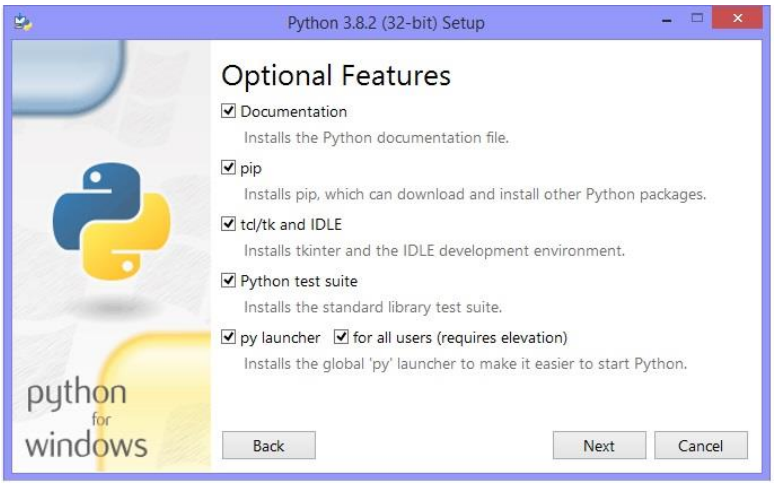
2. For python documentation, visit https://www.python.org/doc/
3. Select the file as per your machine.



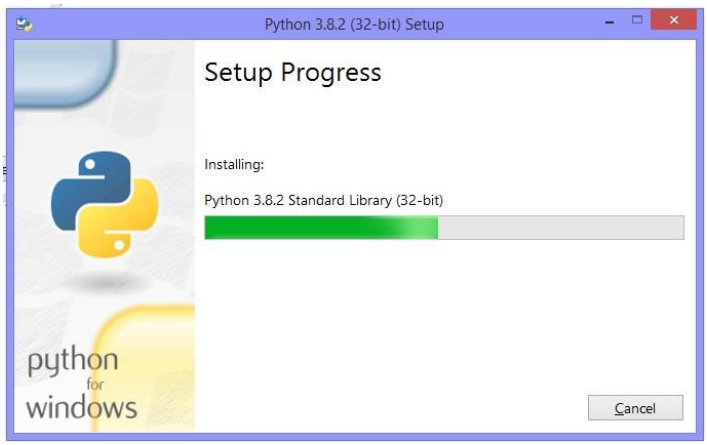
The name of the executable file is Python-3.8.2.exe

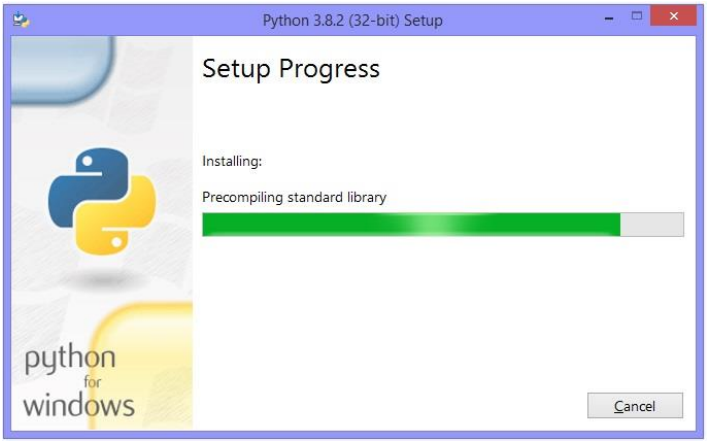
4. Launch the python installer and select the options of Custom installation for choosing the location and other features. Select the options as highlighted below:

5. Select all the optional features and click 'Next'

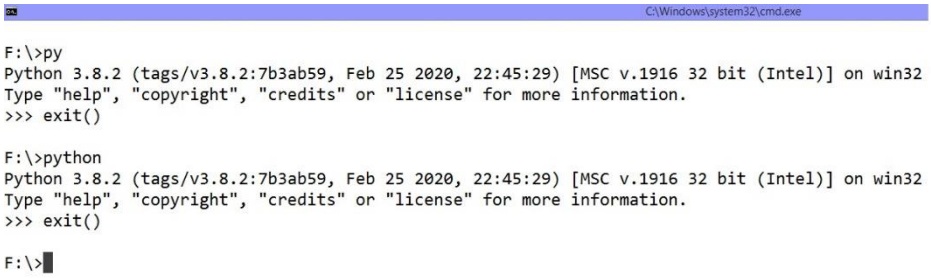


6. Select the advanced options, for example, add python to environment variables, custom location etc., as highlighted below and click 'Install', the Setup successful window appears once the installation is completed.





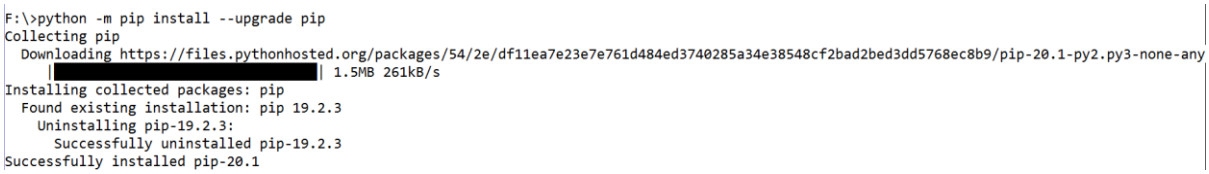
7. From the command prompt, launch python with the commands ‘py’ or ‘python’. If python is successfully installed, the python prompt ‘>>>’ will appear on the command prompt.



Python Libraries Installation Steps

8. For installing any python package, from the python prompt '>>>', execute the command 'pip install <package name>. For example, install Numpy by executing the command 'pip install numpy'

9. In case of upgrading the pip, execute the command 'python –m pip install –upgrade'

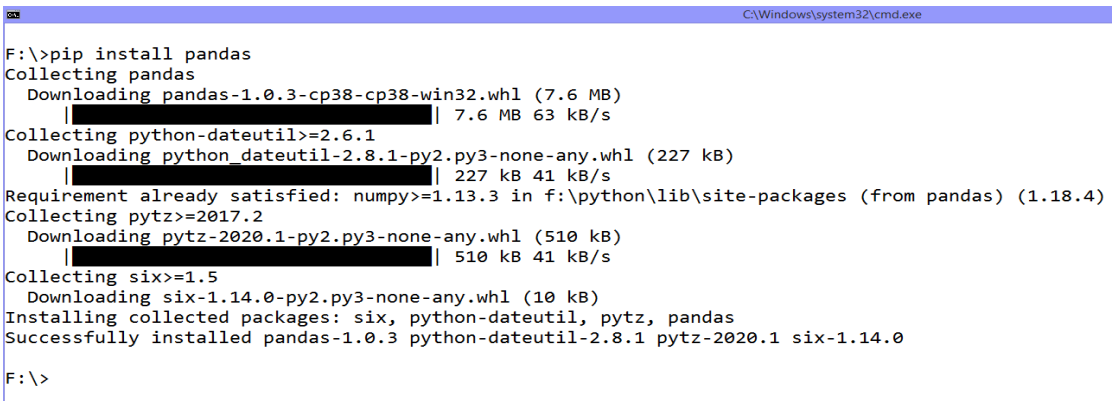


10. Install all the essential packages like Scipy, Pandas, Matplotlib, Scikit-learn, Seaborn using pip install command

```
F:\>pip install scipy
Collecting scipy
  Downloading scipy-1.4.1-cp38-cp38-win32.whl (27.9 MB)
     |████████████████████████████████| 27.9 MB 125 kB/s
Requirement already satisfied: numpy>=1.13.3 in f:\python\lib\site-packages (from scipy) (1.18.4)
Could not build wheels for numpy, since package 'wheel' is not installed.
Installing collected packages: scipy
Successfully installed scipy-1.4.1

F:\>pip install wheel
Collecting wheel
  Downloading wheel-0.34.2-py2.py3-none-any.whl (26 kB)
Installing collected packages: wheel
Successfully installed wheel-0.34.2

F:\>
```



```
F:\>pip install matplotlib
Collecting matplotlib
  Downloading matplotlib-3.2.1-cp38-cp38-win32.whl (9.0 MB)
     |████████████████████████████████| 9.0 MB 154 kB/s
Collecting kiwisolver>=1.0.1
  Downloading kiwisolver-1.2.0-cp38-none-win32.whl (43 kB)
     |████████████████████████████████| 43 kB 128 kB/s
Collecting cycler>=0.10
  Downloading cycler-0.10.0-py2.py3-none-any.whl (6.5 kB)
Requirement already satisfied: python-dateutil>=2.1 in f:\python\lib\site-packages (from matplotlib) (2.8.1)
Requirement already satisfied: numpy>=1.11 in f:\python\lib\site-packages (from matplotlib) (1.18.4)
Collecting pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1
  Downloading pyparsing-2.4.7-py2.py3-none-any.whl (67 kB)
     |████████████████████████████████| 67 kB 120 kB/s
Requirement already satisfied: six in f:\python\lib\site-packages (from cycler>=0.10->matplotlib) (1.14.0)
Installing collected packages: kiwisolver, cycler, pyparsing, matplotlib
Successfully installed cycler-0.10.0 kiwisolver-1.2.0 matplotlib-3.2.1 pyparsing-2.4.7

F:\>
```

```
F:\>pip install scikit-learn
Collecting scikit-learn
  Downloading scikit_learn-0.22.2.post1-cp38-cp38-win32.whl (5.7 MB)
     |████████████████████████████████| 5.7 MB 42 kB/s
Requirement already satisfied: scipy>=0.17.0 in f:\python\lib\site-packages (from scikit-learn) (1.4.1)
Requirement already satisfied: numpy>=1.11.0 in f:\python\lib\site-packages (from scikit-learn) (1.18.4)
Collecting joblib>=0.11
  Downloading joblib-0.14.1-py2.py3-none-any.whl (294 kB)
     |████████████████████████████████| 294 kB 94 kB/s
Installing collected packages: joblib, scikit-learn
Successfully installed joblib-0.14.1 scikit-learn-0.22.2.post1

F:\>
```

```
F:\>pip install seaborn
Collecting seaborn
  Downloading seaborn-0.10.1-py3-none-any.whl (215 kB)
     |████████████████████████████████| 215 kB 32 kB/s
Requirement already satisfied: numpy>=1.13.3 in f:\python\lib\site-packages (from seaborn) (1.18.4)
Requirement already satisfied: pandas>=0.22.0 in f:\python\lib\site-packages (from seaborn) (1.0.3)
Requirement already satisfied: scipy>=1.0.1 in f:\python\lib\site-packages (from seaborn) (1.4.1)
Requirement already satisfied: matplotlib>=2.1.2 in f:\python\lib\site-packages (from seaborn) (3.2.1)
Requirement already satisfied: python-dateutil>=2.6.1 in f:\python\lib\site-packages (from pandas>=0.22.0->seaborn) (2.8.1)
Requirement already satisfied: pytz>=2017.2 in f:\python\lib\site-packages (from pandas>=0.22.0->seaborn) (2020.1)
Requirement already satisfied: kiwisolver>=1.0.1 in f:\python\lib\site-packages (from matplotlib>=2.1.2->seaborn) (1.2.0)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in f:\python\lib\site-packages (from matplotlib>=2.1.2->seaborn) (2.4.7)
Requirement already satisfied: cycler>=0.10 in f:\python\lib\site-packages (from matplotlib>=2.1.2->seaborn) (0.10.0)
Requirement already satisfied: six>=1.5 in f:\python\lib\site-packages (from python-dateutil>=2.6.1->pandas>=0.22.0->seaborn) (1.14.0)
Installing collected packages: seaborn
Successfully installed seaborn-0.10.1

F:\>
```