

Chapter 3:

Machine Learning Environment Setup

To be able to learn and practice machine learning, you will first need to get your Python/Anaconda, Scikit-Learn, and TensorFlow environment set up. The following chapter will take you through how to set up your environment. With your environment set up correctly, you will be able to follow the examples in this book.

Setting Up Python and Anaconda

Scikit-learn and TensorFlow can be used with Python 2.7 and above. They will also work with Anaconda 2.7 and higher. There are a few libraries that need to be installed after you have installed Python and Anaconda which are a requirement for installing Scikit-Learn or TensorFlow. These will be covered in the corresponding sections of this chapter.

Note:

- This book assumes that you are using the **Windows** 7 or higher operating system to download and install Python, Anaconda, Scikit-Learn, and TensorFlow. While MAC and Linux download and installation procedures differ, the exercises can still be used on these operating systems.
- If you already have Python on your system, you can skip the “**Installing Python**” section and go to the “**Installing Anaconda**” section if you wish to use Anaconda.
- If you have Python already installed on your machine, you may want to check if you have the following libraries installed:
 - NumPy v 1.18.4
 - SciPy v 1.4.1
 - Pip
- If you have the above libraries installed, or the version pertaining to the release of Python installed on your machine, skip to the “**Installing Anaconda**” section.
- If you are not installing Anaconda or already have it installed on your system, you can skip to “**Installing Scikit-Learn**” or “**Installing TensorFlow**” section should you already have one or the other installed.
- If you already have your Python, Scikit-Learn, and TensorFlow environment set, you can skip this chapter altogether and proceed to the next chapter “**Using Scikit-Learn.**” However, it will do you no harm checking or even upgrading your current system environment for Scikit-Learn and TensorFlow.

Installing Python

The latest version of Python is 3.8.3 and it can be retrieved from the Python.org website. For the sake of this book, version 3.8.3 will be used to keep the book up to date as at time of print. As it is open source software, it is free to download and use.

If you already have Python installed on your machine you can use it to follow along as long as it is 2.7 or higher. It is good to note that there may be differences between the earlier versions and the current ones. Please check the Python.org website for more details on the differences between the versions.

Download Python

On the Python.Org website go to the “**Downloads**” screen.



There is the option to download earlier releases under the “**Active Python Releases**” section of the page.

Active Python Releases				
For more information, visit the Python Developer's Guide.				
Python version	Release name	First released	End of support	Release schedule
3.8	bugfix	2019-10-14	2024-01	PSF-2019
3.7	bugfix	2019-09-17	2022-06-27	PSF-2019
3.6	security	2019-10-14	2021-11-22	PSF-2019
3.5	security	2018-09-18	2020-06-27	PSF-2018
3.4	end of life	2016-07-26	2019-03-24	PSF-2016

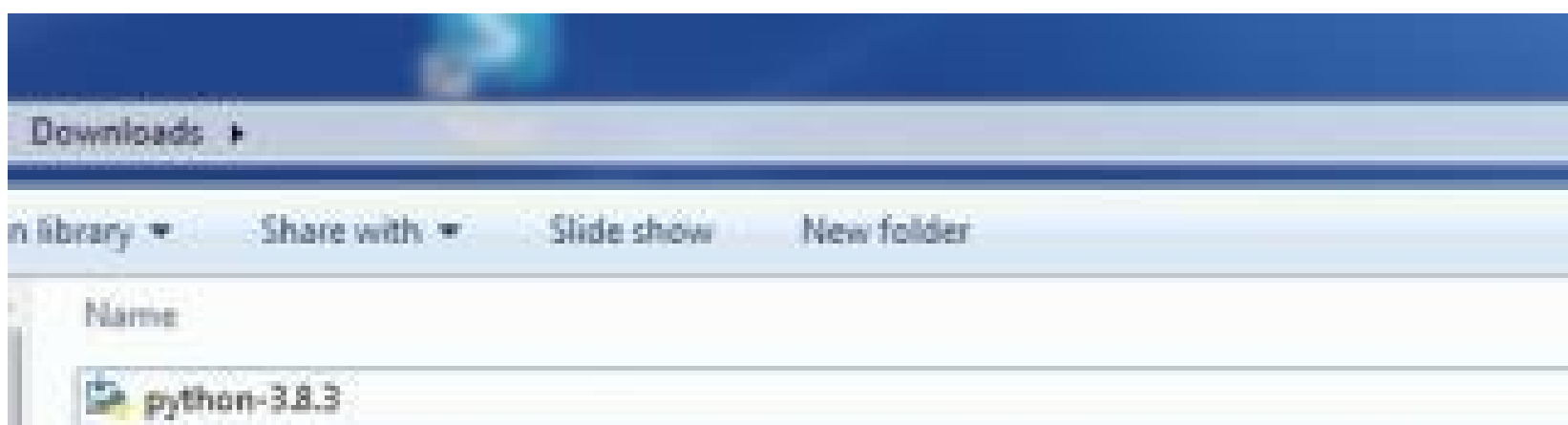
Looking for a specific release?		
Python releases by version number		
Release version	Release date	Click for more
Python 3.8.3	May 14, 2020	Download Release notes
Python 3.8.2rc1	April 29, 2020	Download Release notes

For this exercise you are going to use **Python 3.8.3**.

Choose the version from the “**Active Python Releases**” or click on the “**Download Python 3.8.3**” button beneath the “**Download the latest version for Windows**” section. Make sure it is the 64-bit version, as the current version of TensorFlow works best with the 64-bit.



The file will download into the “**Downloads**” folder on your computer.



Check that the file is there once the download has been completed.

Installing Python on Windows

From the “**Downloads**” folder, double click on the “**python-3.8.3**” executable file.

Click on the “**Run**” button from the “**Open File - Security Warning**” dialogue box that will appear.

Make sure to tick the “**Add Python 3.9 to Path**” at the bottom of the “**Install Python 3.8.3 (xx-bit)**” screen. Keep the default installation directory, unless you are a more advanced user and would like to install the software in another location. For the sake of this training exercise, the default directory has not been changed to simplify the process.

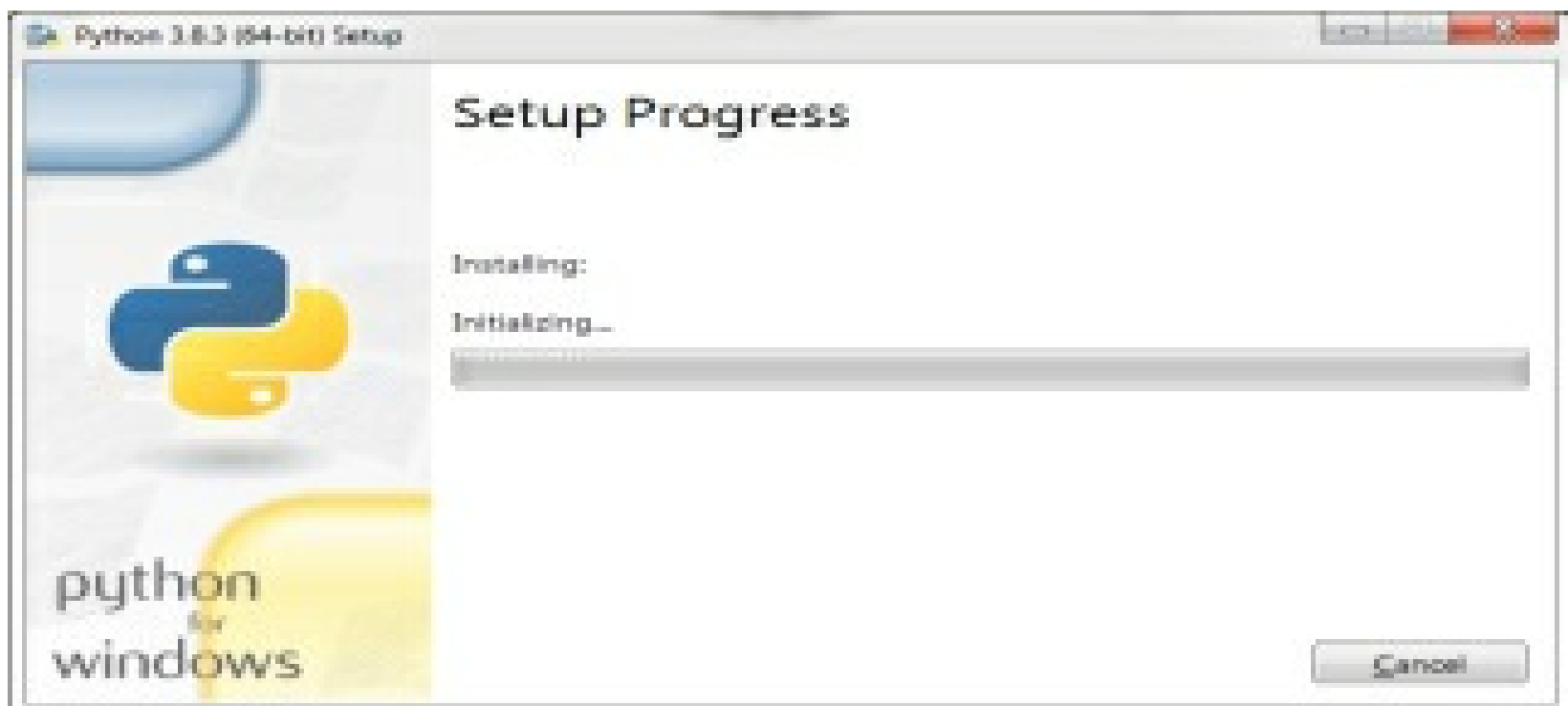
Click the “**Install Now**” option on the “**Install Python 3.8.3 (xx-bit)**” screen.



If the “**Access Control**” warning screen comes up, click on “**Yes**” to allow the application permission to install on

the system.

The “**Setup Progress**” screen will appear with the installation progress for installing Python.

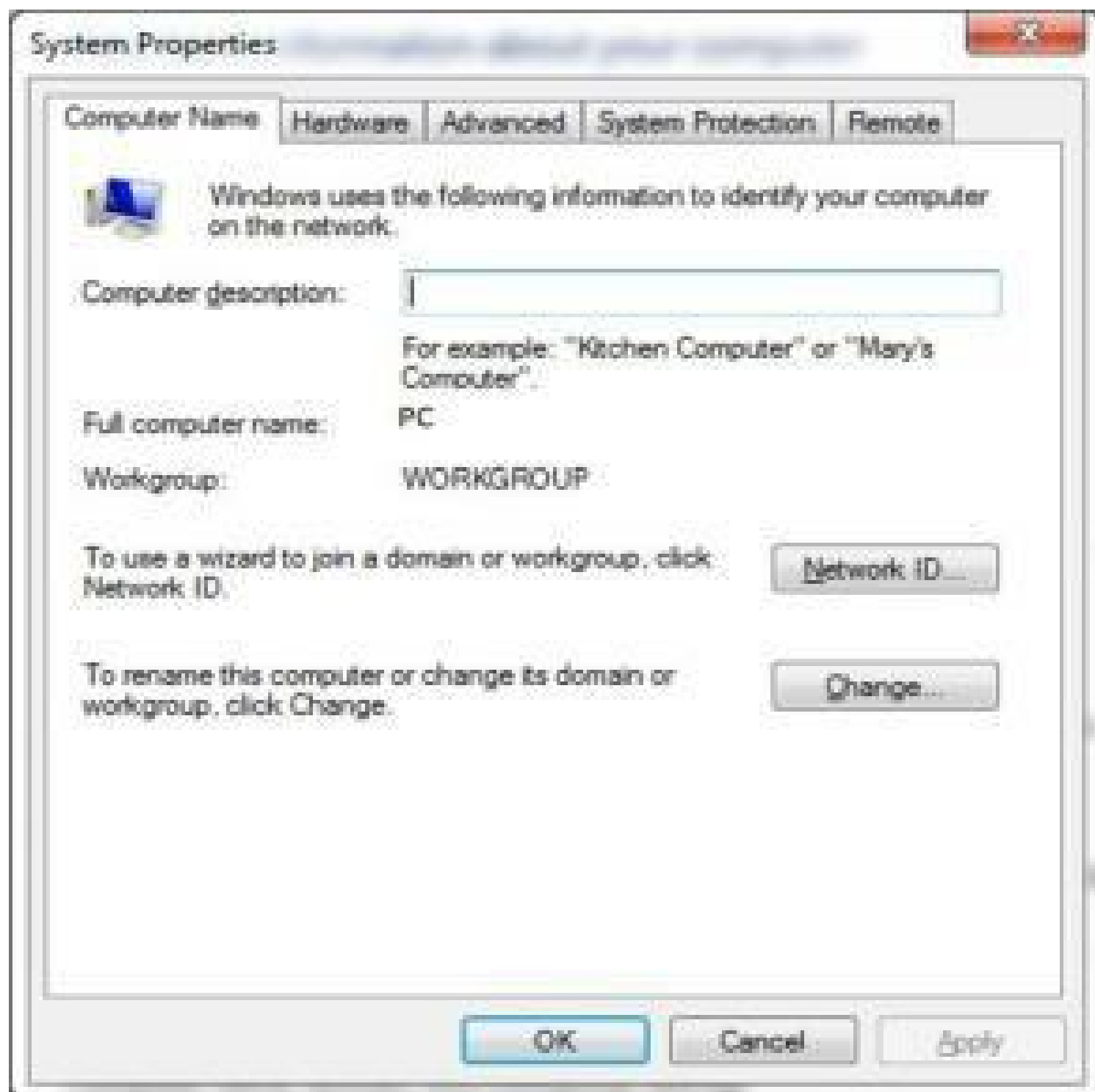


When Python is finished installing the “**Setup was successful**” screen will appear. Python will now be installed on your system.

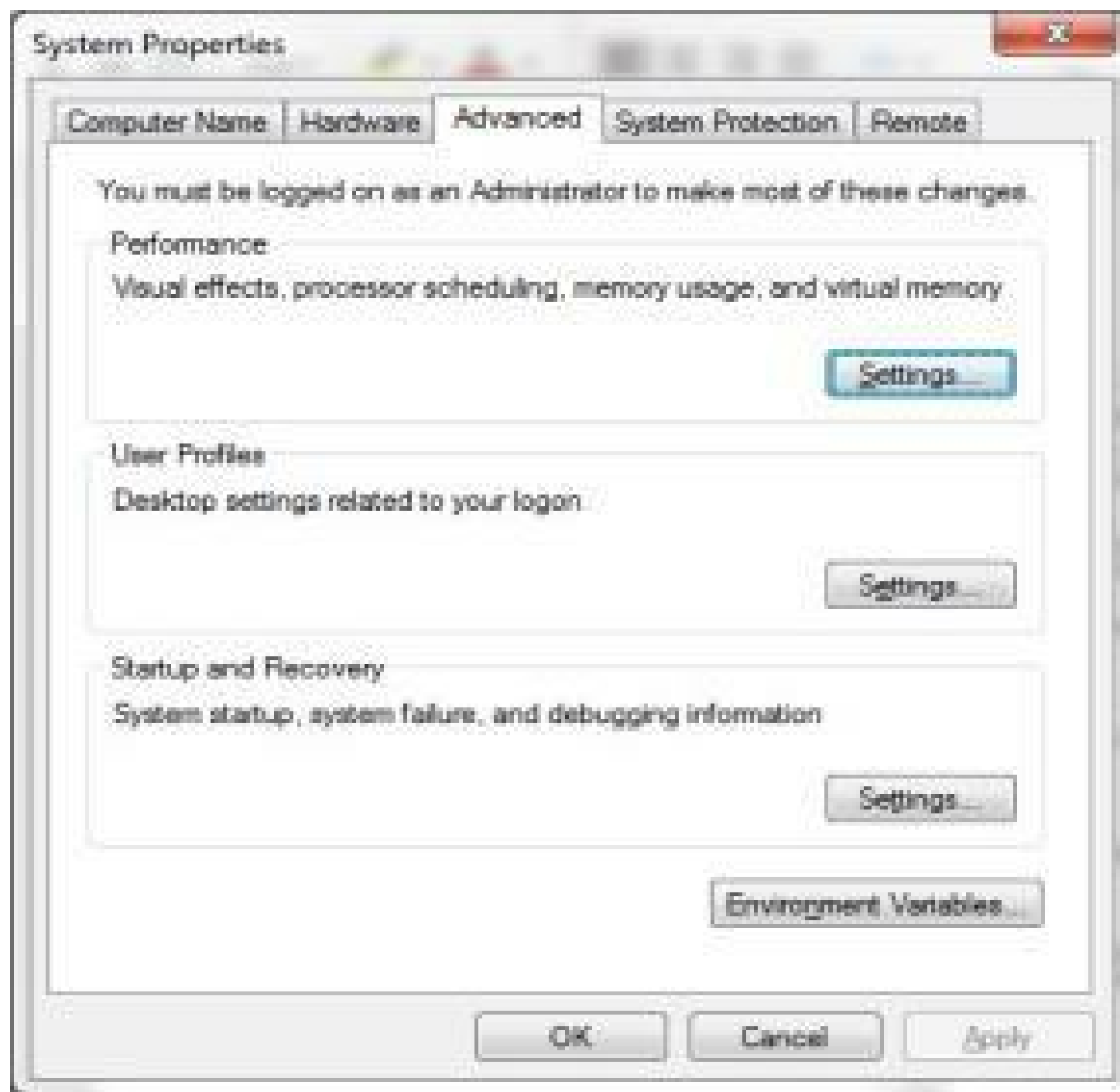


Setting Up the Python Environment

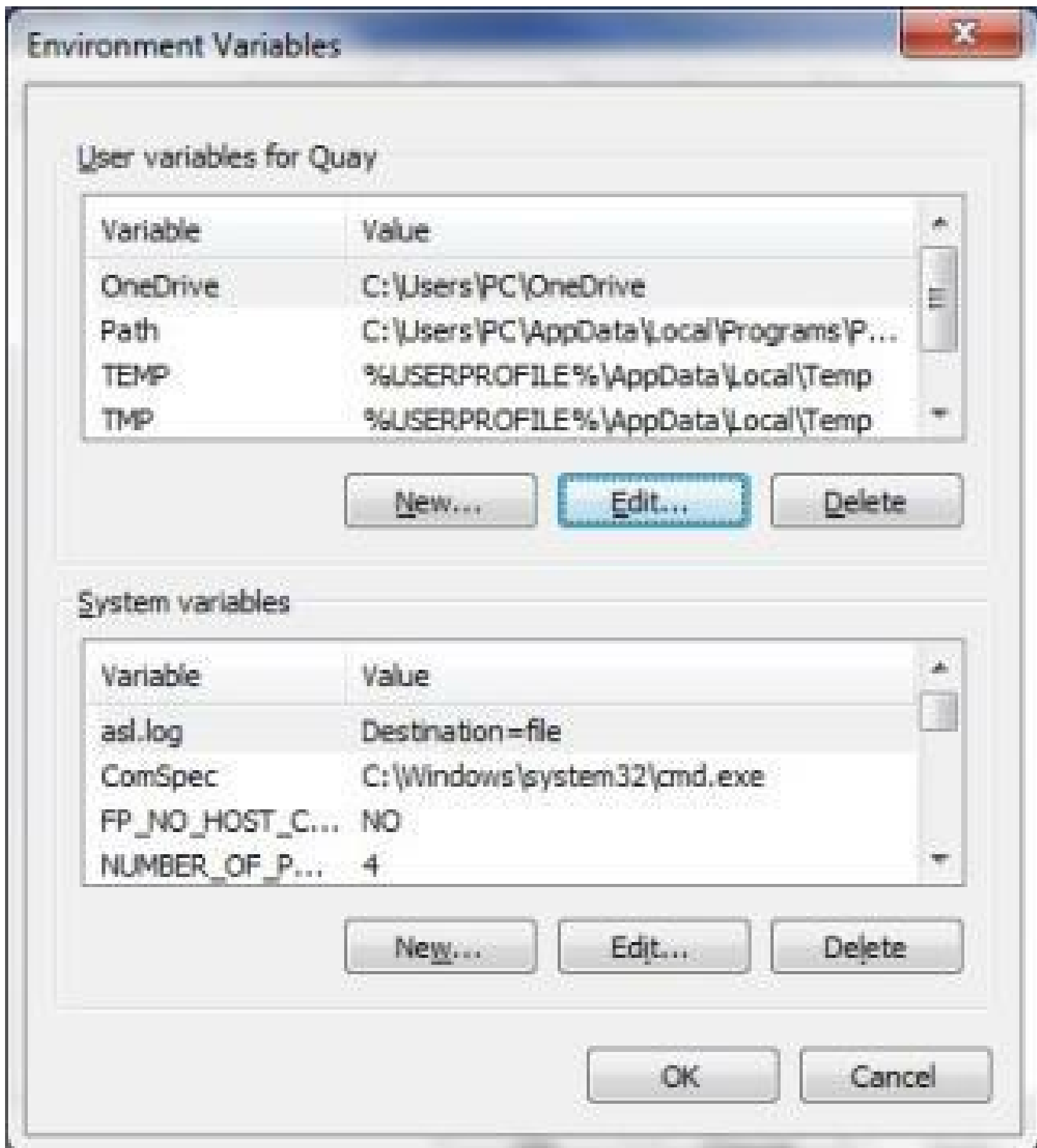
You will need to check to see if Python is running on the system. If you did not tick the “Add Python to Path” box, you will need to add it to the **System Environment Variable** path. To check if it was put into the **System Environment Variable** path, go to the “**Control Panel - System and Security - System Properties.**”



Click on the “**Advanced**” tab.



Click on the “**Environment Variables**” button.



Click on “**Path**” under “System variables” and then the “**Edit**” button.

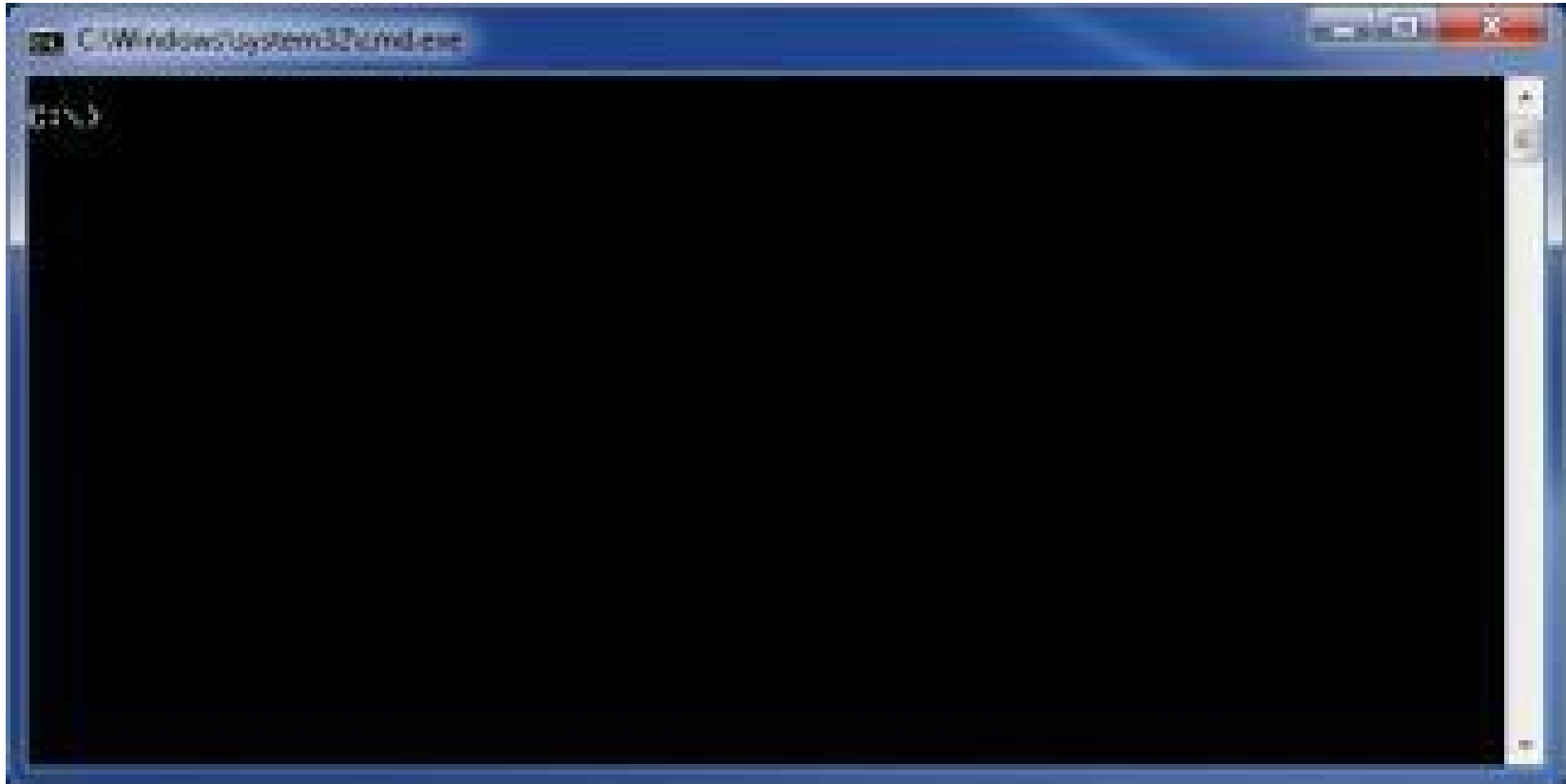
For systems earlier than Windows 10, you will need to place your cursor at the end of the “**Variable value**” path to check if the following line (or similar, depending on where you installed Python) appears. For Windows 10 you will add a new path to the list of variables.

The following variable path should appear in the “**Variable value**” box:

;C:\Python\Python 38-32\

If there is no reference to the Python.exe file, find the path to the Python executable file and type it in here as per the example above. **NOTE:** Do not forget to separate each system path variable entry with a “;” for Windows OS earlier than Windows 10, or it will create problems and not pick up the application. When you are done checking the System path variable click “**OK**” until you have exited the **System Variable** screen. You can also close down the “**Control Panel**” before moving on to test the Python installation.

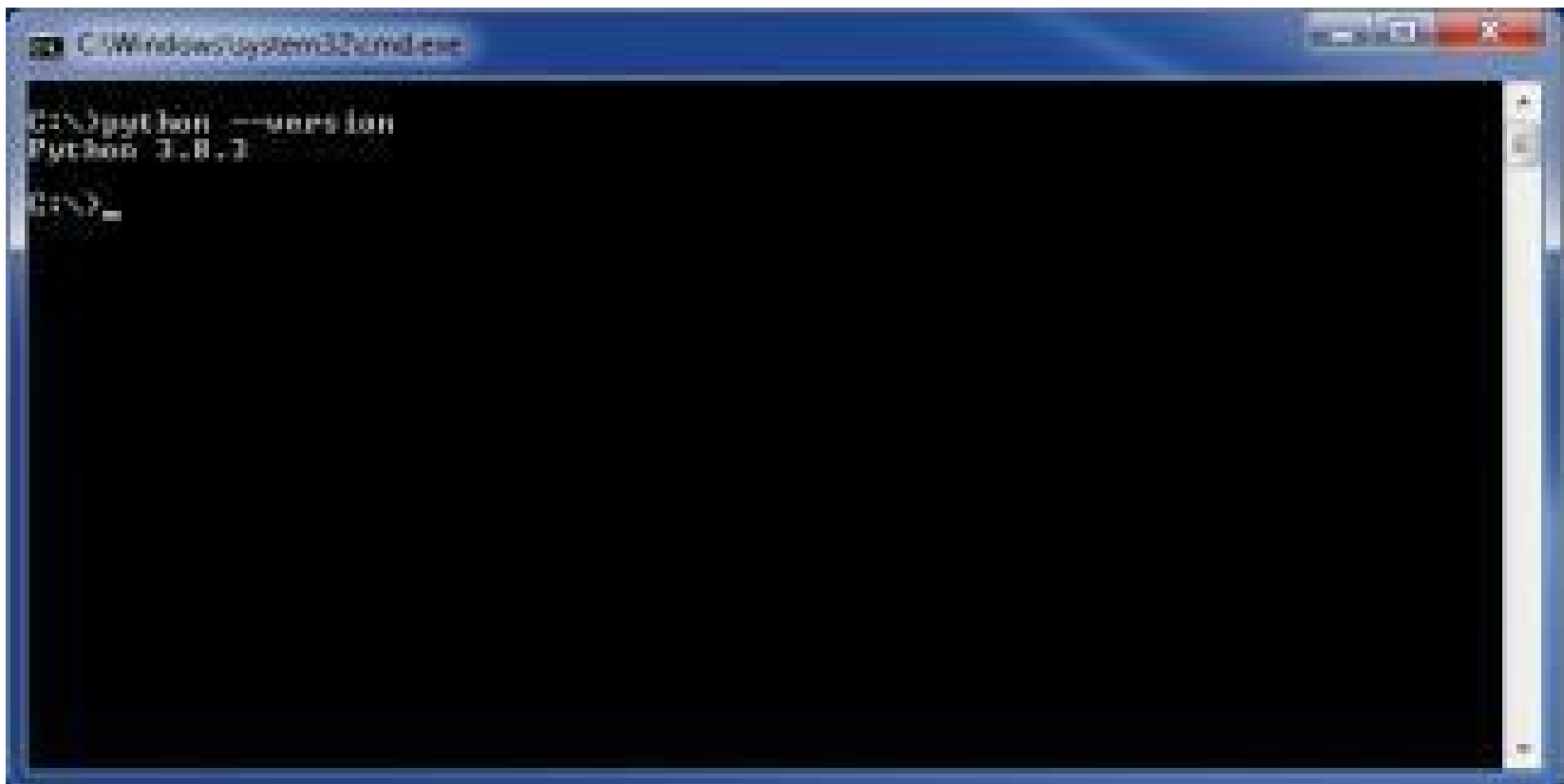
From the Start menu, type **cmd** in the search box and hit enter. This will load the command line screen to test that the Python installation installed correctly and can be accessed. The following screen (or similar) will appear:



To test the Python installation type the following into the command line:

```
python --version
```

The following or similar should appear on the screen:



The version of Python that was recently installed on the machine should appear beneath the command you typed in. If it does not appear there, you will need to check the **System Variable Path**.

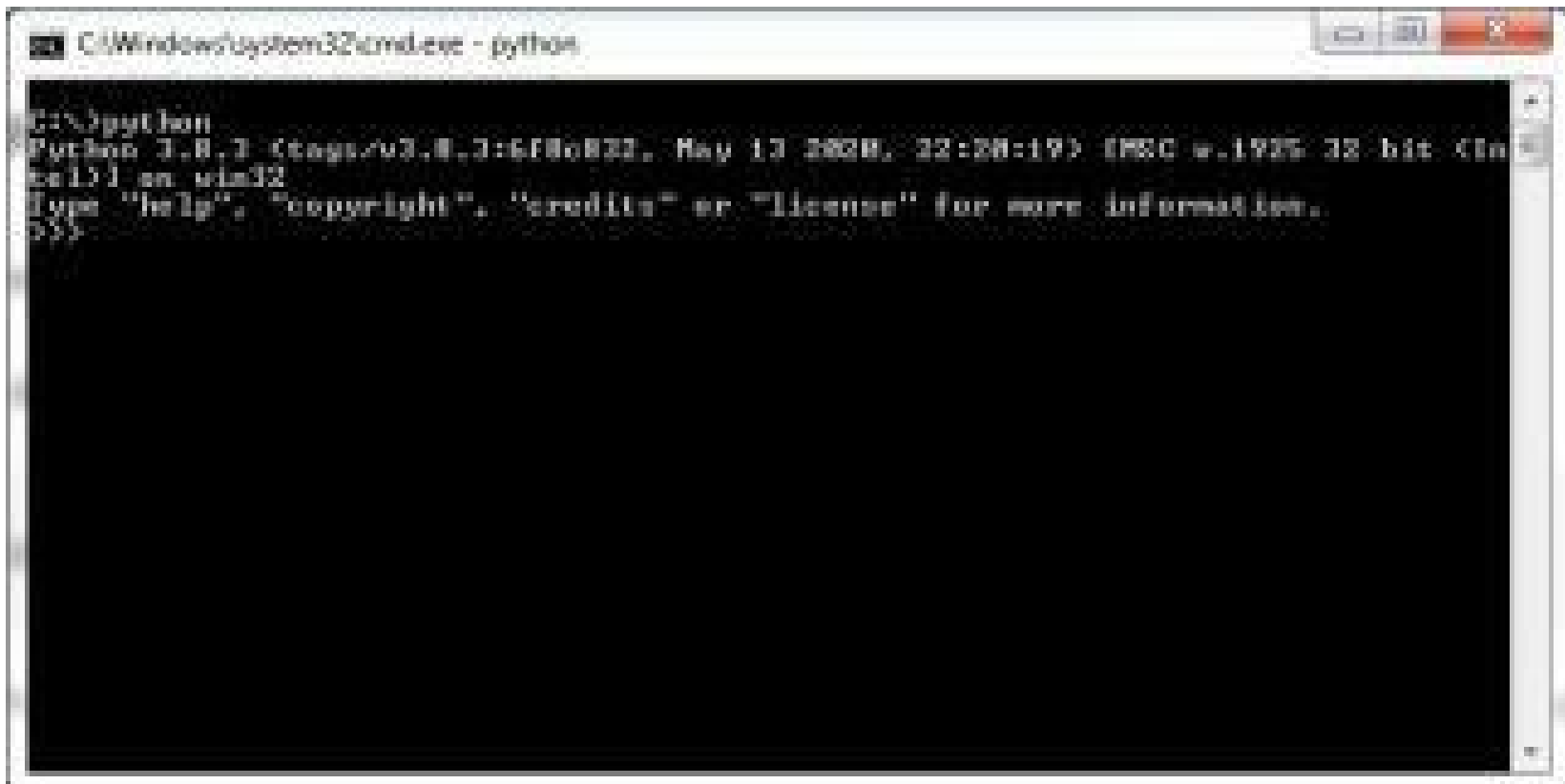
There are a few ways to load the Python command line:

- Create a desktop shortcut
- Access Idle or the Python command line from the “**Start**” menu “**Programs**” option
- Access Python through the **DOS** command line using **cmd** from the “**Start**” menu **run/search** facility.

For the sake of this exercise the **cmd** option is going to be used.

From the cmd command line prompt type:

```
python
```

If you are going to be using Python you will need to use **pip** (Preferred Installer Program). **Pip** is the package installer for Python and if you have Python version 2.7 and higher, it should already be installed with Python. To check that **pip** is installed, type the following into the Python command line:

```
pip --version
```

It should return the version of pip. If it is older than 20.1.1 (the new version at the time of writing this book is version 20.1.1) then **pip** will need to be upgraded. In order to be able to run **pip** from the command line, you can add the **pip** program path to the System Environment Variable Path as you did for the Python executable. The **pip** executable file is usually found in the following Python directory:

```
<Python installation path>\Python\Python38-32\Scripts
```

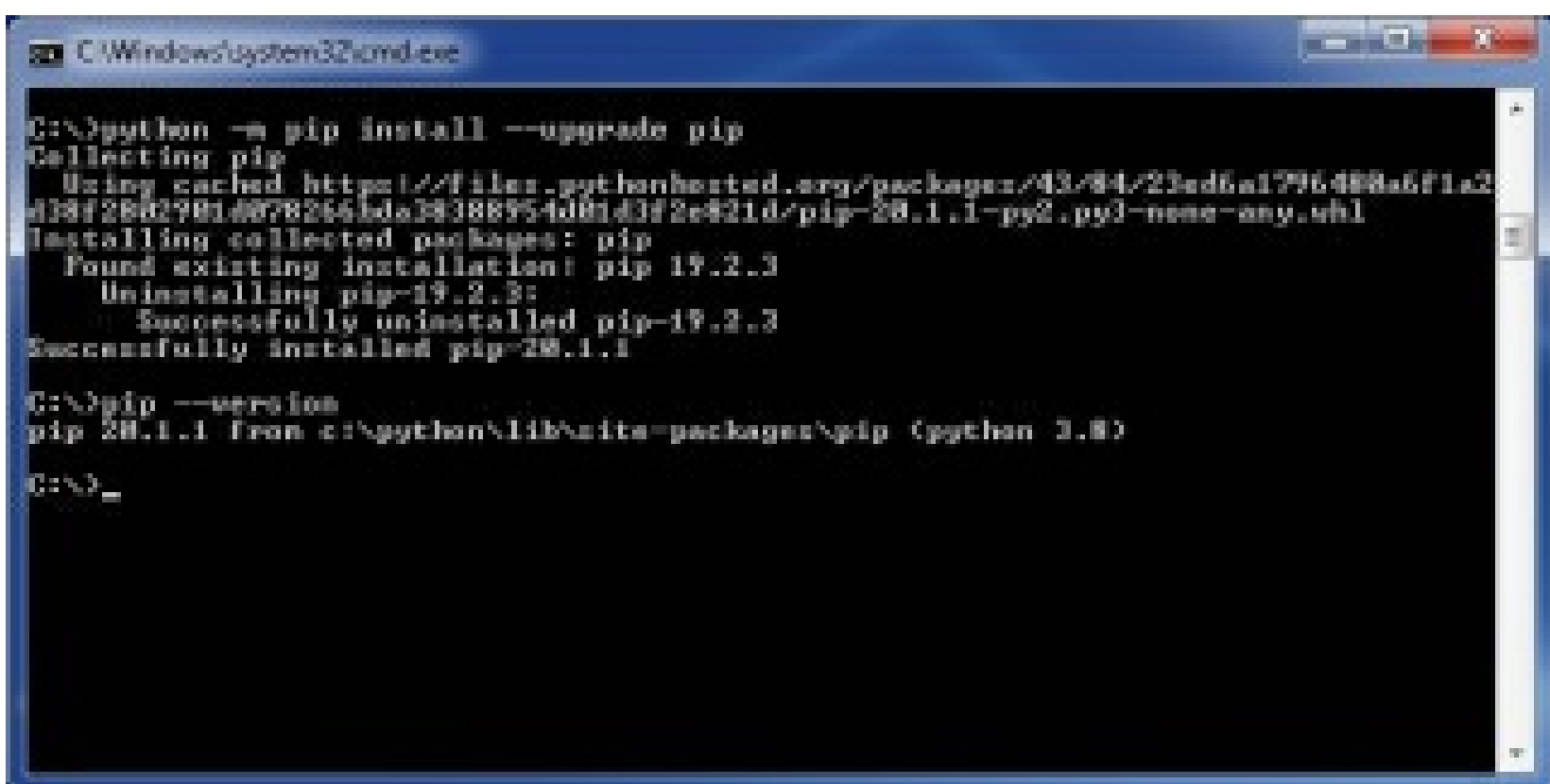
To upgrade **pip** type the following into the command line prompt:

```
python -m pip install --upgrade pip
```

When the **pip** upgrade is finished type:

```
pip --version
```

The version should be 20.1.1



The Python **pip** utility should now be the latest version.

Install NumPy With Pip

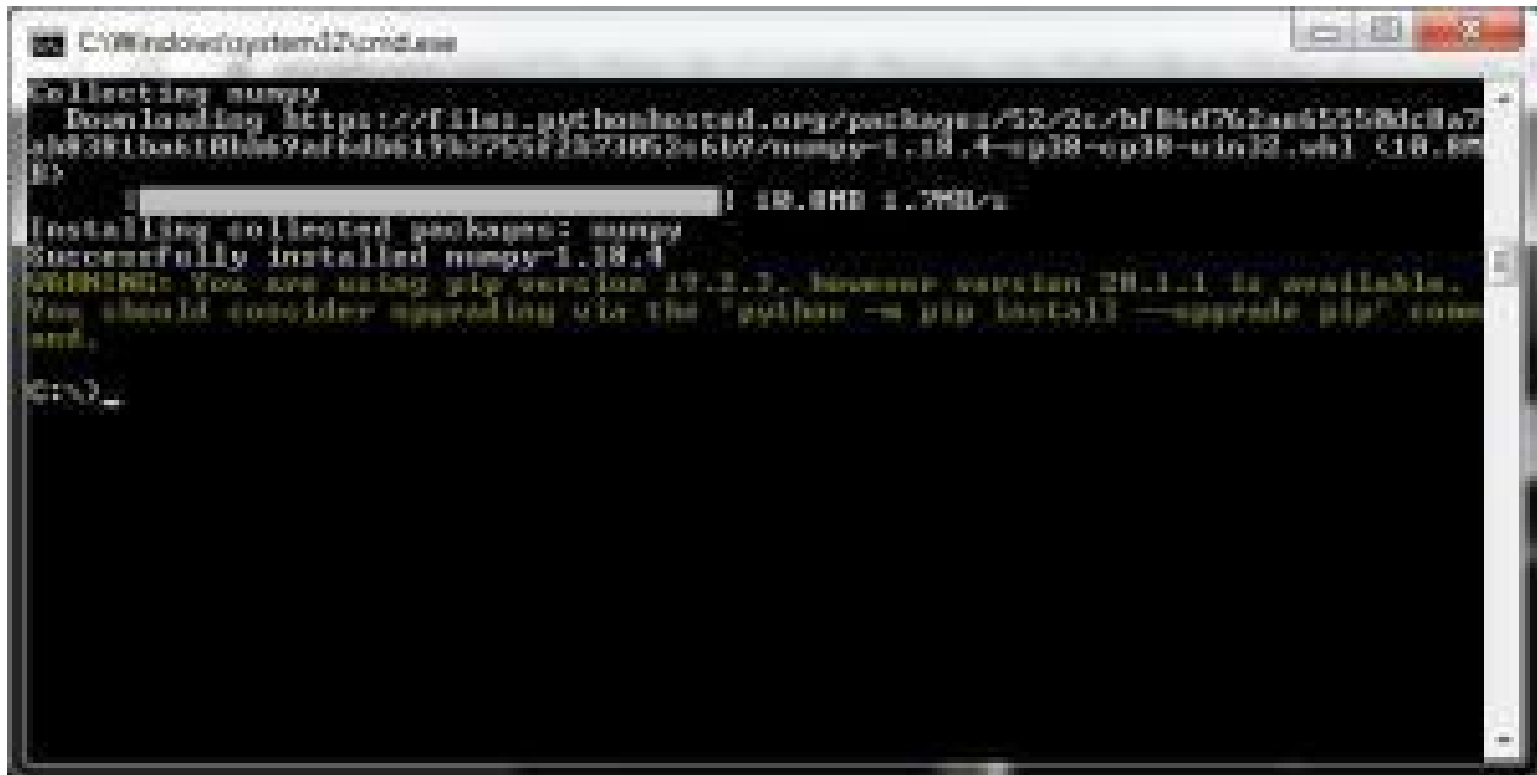
Now you need to check if the NumPy library is installed. NumPy is needed for the support of matrices, high-level math functions, and multi-dimensional arrays that are large. You can check if NumPy was pre-installed with Python by using the following test on the Python command line:

```
import numpy
```

If the command line returns an error “**No module named ‘numpy’**” you will need to get the library. At the command line prompt type the following:

```
pip install numpy
```

If you have pip loaded and referenced correctly in the System Variable Path, pip will download and install the NumPy library as per the image below.



```
C:\Windows\system32\cmd.exe
Collecting numpy
  Downloading https://files.pythonhosted.org/packages/52/2c/bf84d762a655586c8a7
4b8381ba618b69a6f4b617b2755f2b73852a6b9/numpy-1.18.4-cp38-cp38-win32.whl (18.8M
B)
  |#####| 18.8MB 1.7MB/s
Installing collected packages: numpy
Successfully installed numpy-1.18.4
WARNING: You are using pip version 19.2.3, however version 20.1.1 is available.
You should consider upgrading via the 'python -m pip install --upgrade pip' com
mand.
C:\>
```

Install SciPy With Pip

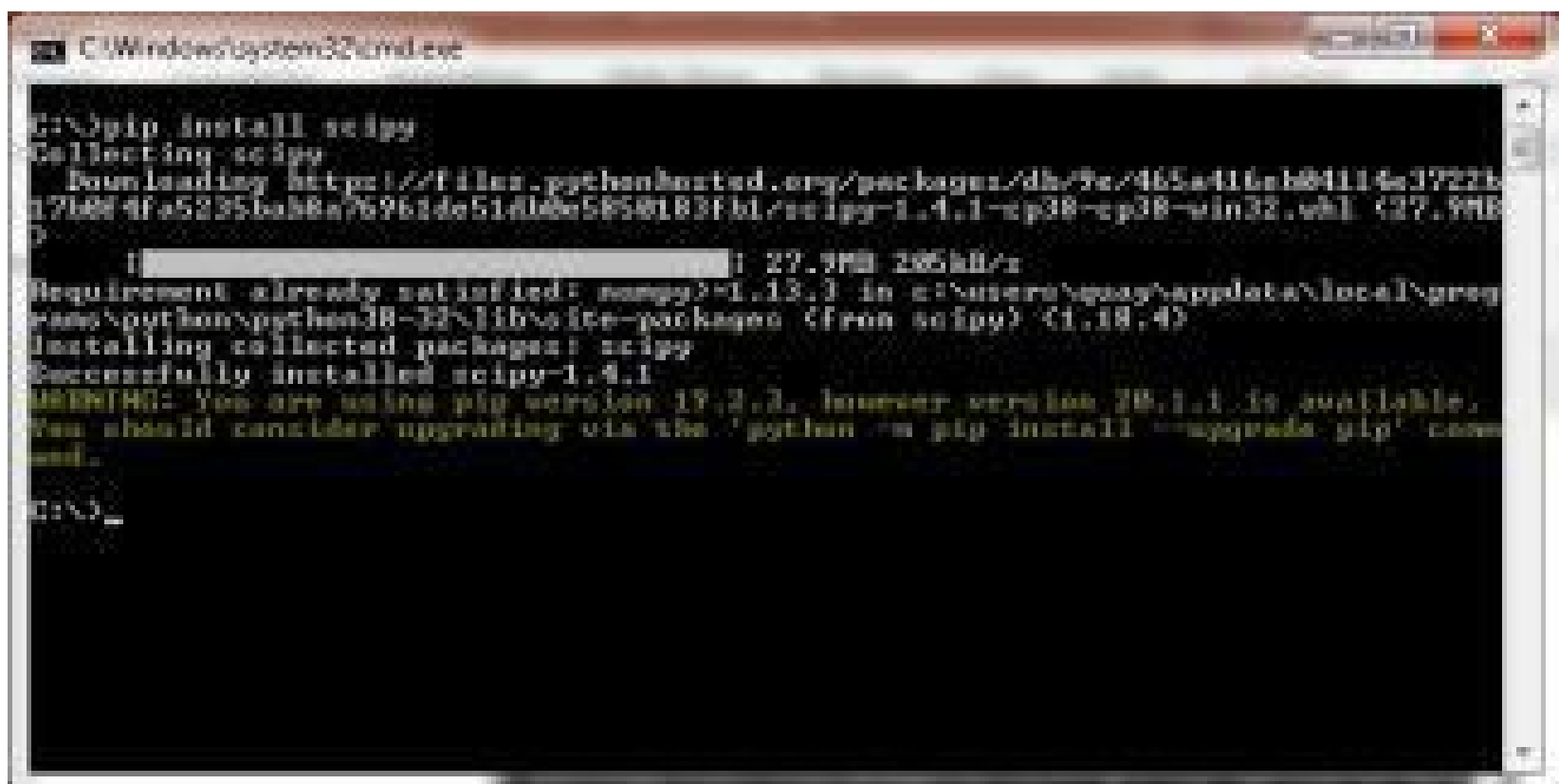
Now you need to check if the SciPy library is installed. You can check if SciPy was pre-installed with Python by using the following test on the Python command line:

```
import scipy
```

If the command line returns an error “**No module named ‘scipy’**” you will need to get the library. At the command line prompt type the following:

```
pip install scipy
```

If you have **pip** loaded and referenced correctly in the System Variable Path, **pip** will download and install the SciPy library as per the image below.



```
C:\Windows\system32\cmd.exe
C:\>pip install scipy
Collecting scipy
  Downloading https://files.pythonhosted.org/packages/d8/9e/465a416ab04114e3722b
17b884fa5235ba38a76961de51d0ba5858183f61/scipy-1.4.1-cp38-cp38-win32.whl (27.9M
B)
  |#####| 27.9MB 285kB/s
Requirement already satisfied: numpy>=1.13.3 in c:\users\quay\appdata\local\prog
rams\python\python38-32\lib\site-packages (from scipy) (1.18.4)
Installing collected packages: scipy
Successfully installed scipy-1.4.1
WARNING: You are using pip version 19.2.3, however version 20.1.1 is available.
You should consider upgrading via the 'python -m pip install --upgrade pip' com
mand.
C:\>
```

Installing Anaconda

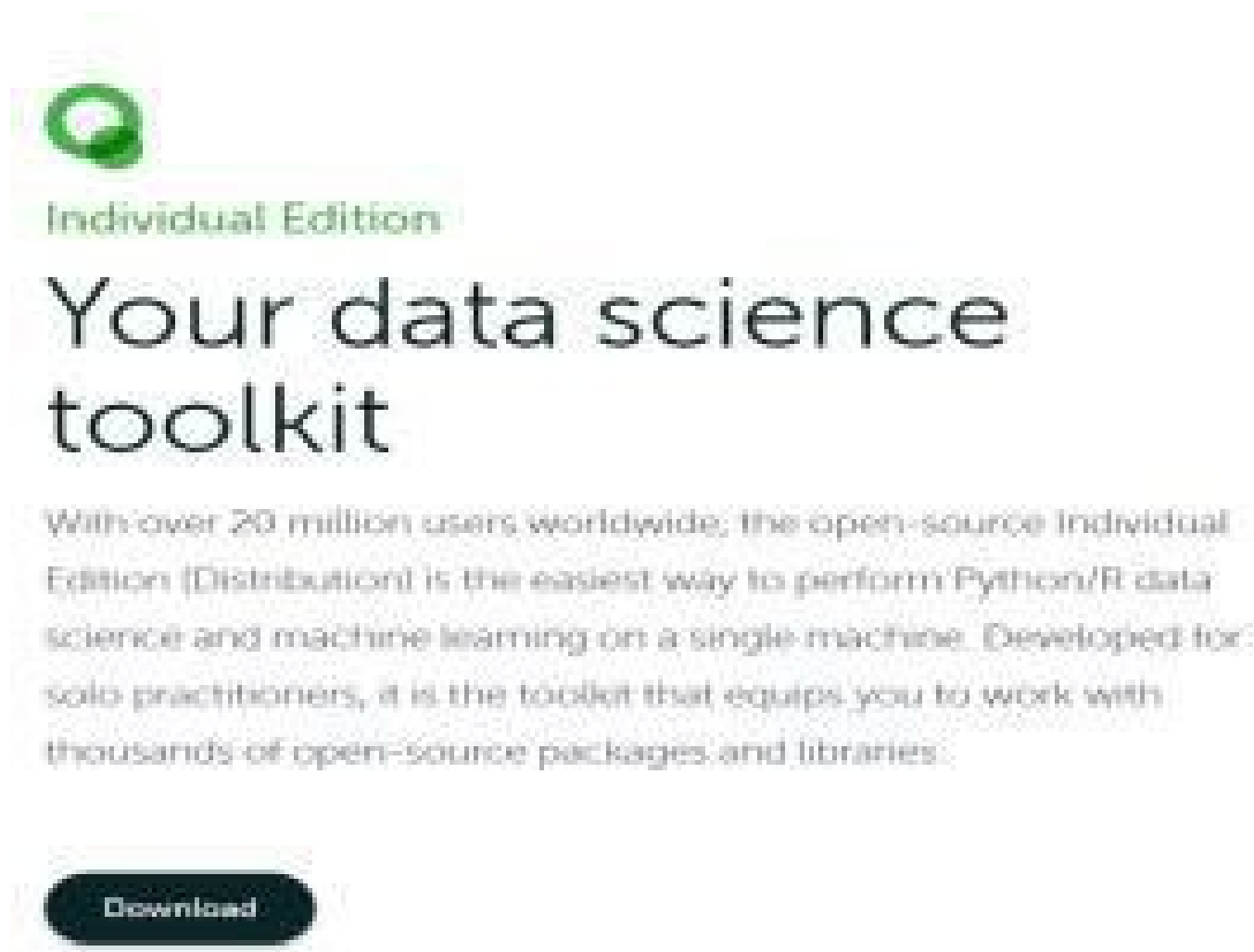
The latest version of Anaconda is 3.7 and it can be retrieved from the Anaconda.com website. For the sake of this book, version 3.7 64-bit will be used to keep the book up to date at time of print. The 64-bit version is advisable as TensorFlow works best with 64-bit. As TensorFlow is open source software, it is free to download and use.

If you already have Anaconda installed on your machine, you can use it to follow along as long as it is 2.7 or

higher. It is good to note that there may be differences between the earlier versions and the current ones. Please check the Anaconda.com website for more details on the differences between the versions.

Download Anaconda

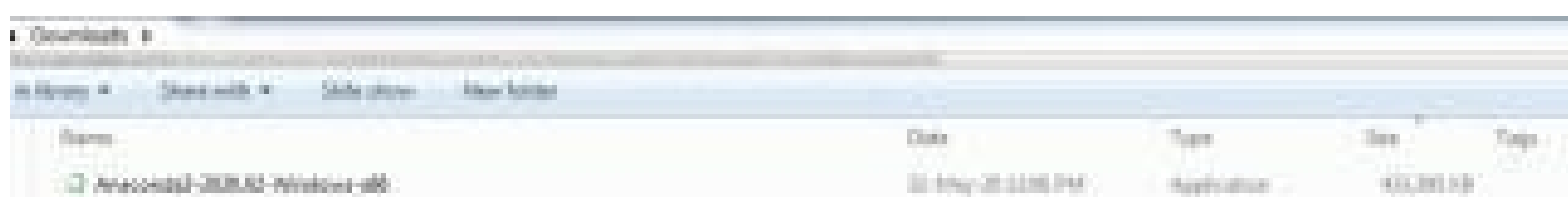
On the Anaconda.com website, scroll down the page until you find the “**Downloads**” button.



There is an option to download earlier releases under the “**Anaconda Installers**” section of the page.



For this exercise you are going to use the **Python 3.7** version which relates to your machine (i.e. 32-bit or 64-bit). Choose the version from the “**Anaconda Installers**” screen and it will start to download the installer file into the “**Downloads**” folder on the machine.



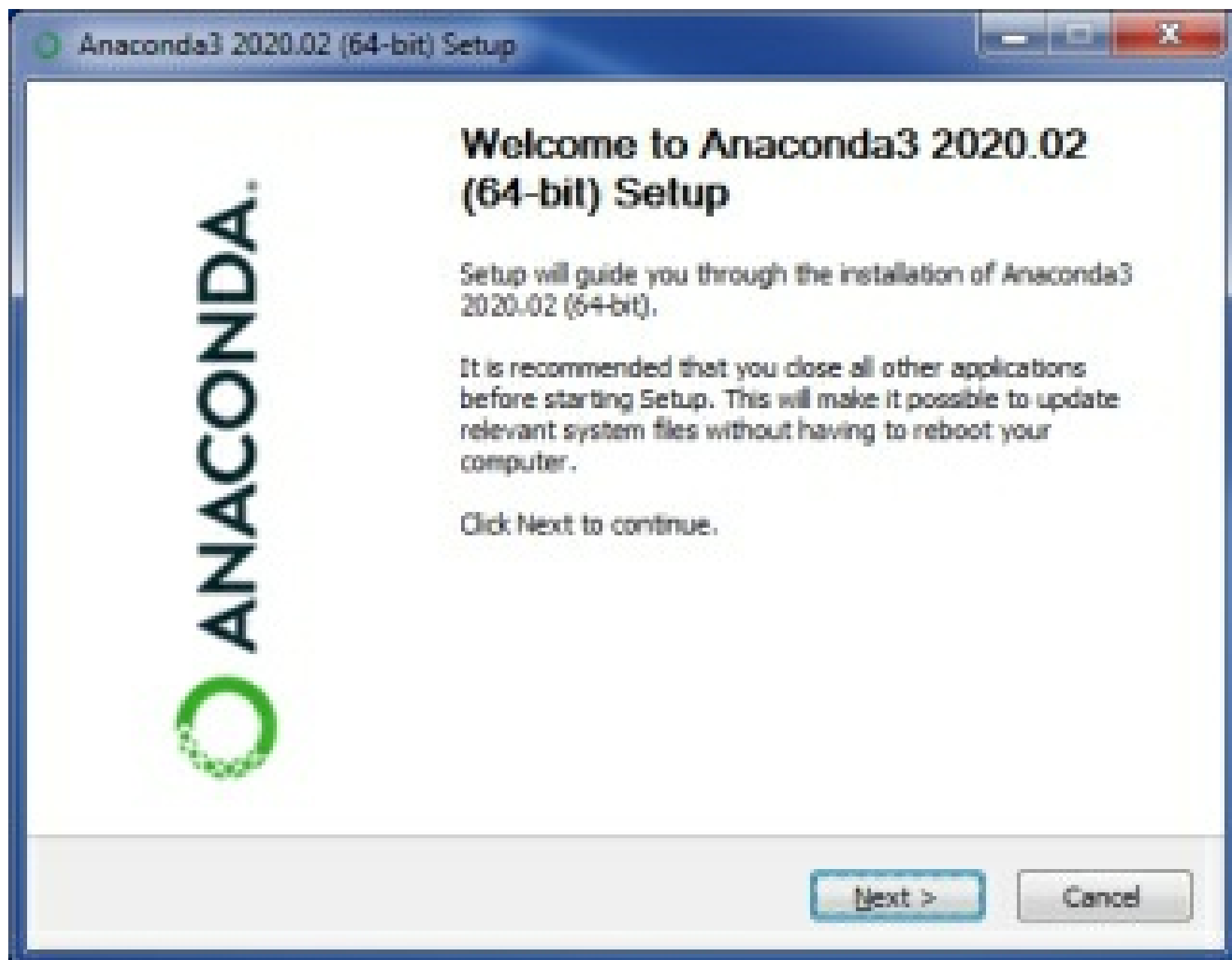
Check that the file is there once the download has been completed.

Installing Anaconda on Windows

From the “**Downloads**” folder, double click on the “**Anaconda3.2020.02-Windows-xx**” executable file.

Click on the “**Run**” button from the “**Open File - Security Warning**” dialogue box that will appear.

Click “**Next**” at the bottom of the “**Welcome to Anaconda3 2020.02 (xx-bit) Setup**” screen.



Click the “**I Agree**” button on the “**ANACONDA License Agreement**” screen.

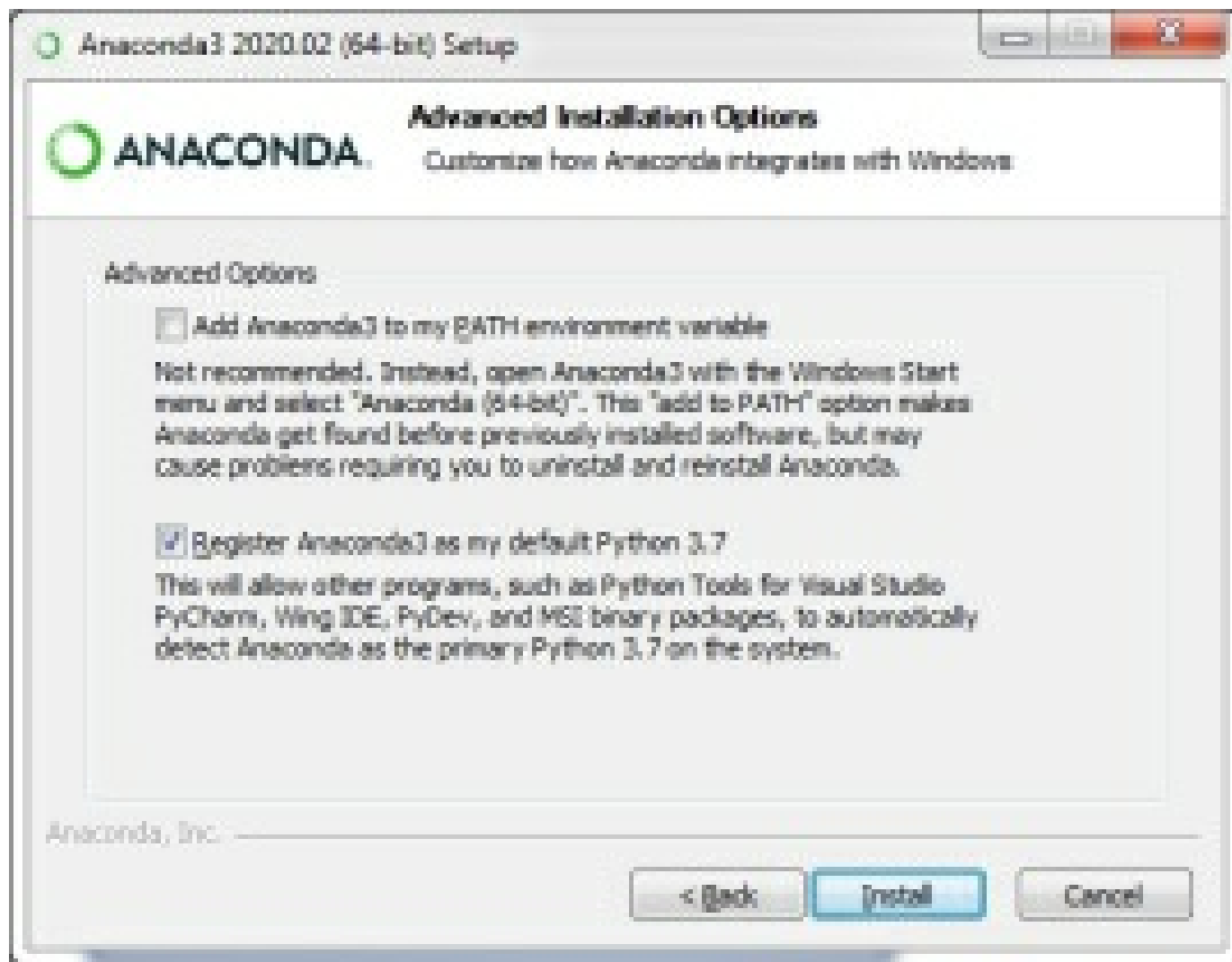
Choose either “**Just me (recommended)**” or “**All Users (requires admin privileges)**” depending on your preferences at the “**Select Installation Type**” screen.

At the “Choose Install Location” screen, it is recommended to keep the default installation directory or set it to the directory of your choice. The directory for the sake of this book will be:

C:\anaconda3

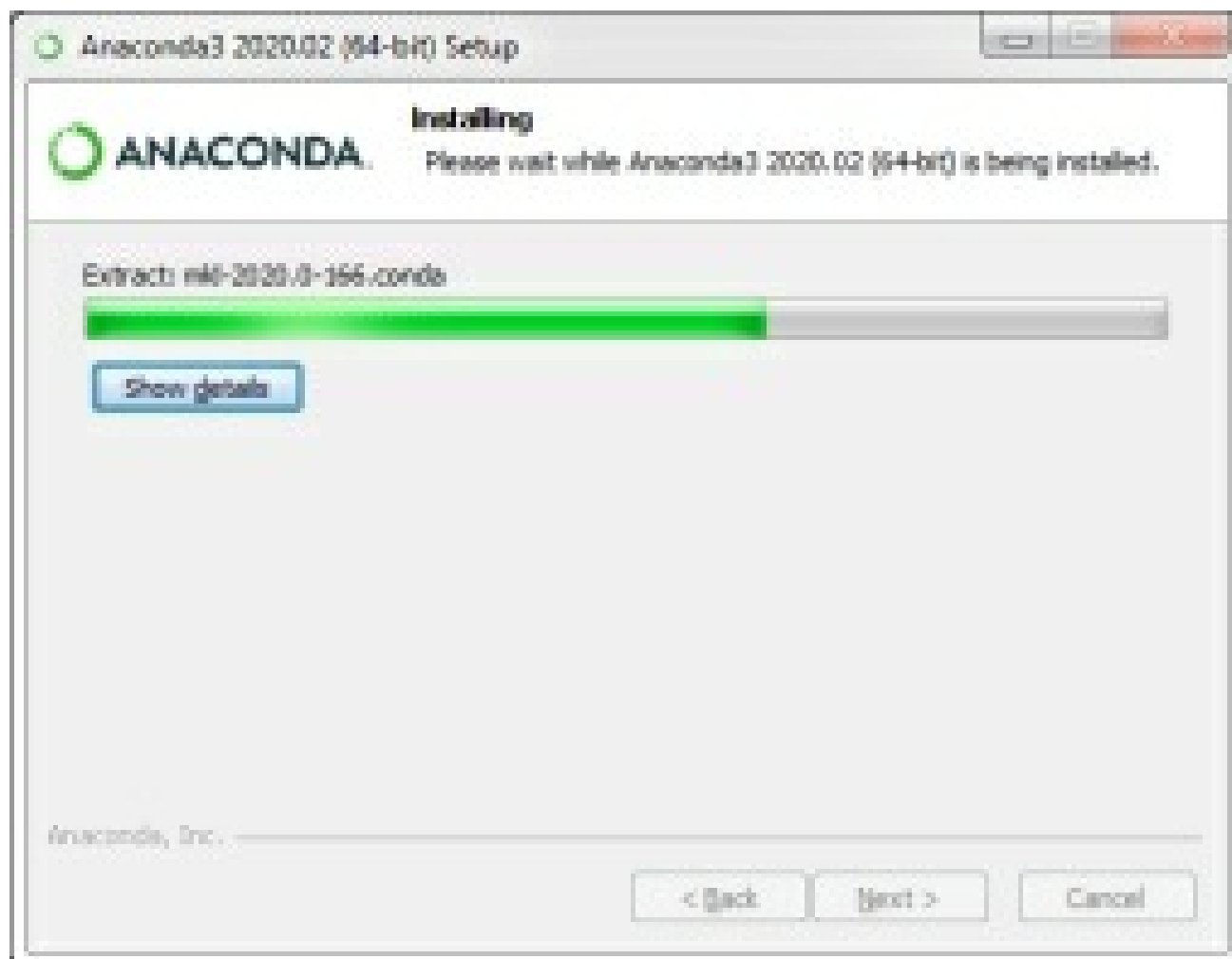
Click the “**Next**” button when you are ready to proceed with the installation.

At the “**Advanced Installation Options**” screen, leave the “Add Anaconda3 to my PATH environment variable” **UNCHECKED**. This will be added later after the installation is complete. You can leave the “**Register Anaconda3 as my default Python 3.7**” **CHECKED**.



Click the “**Install**” button at the bottom of the screen to continue the installation process. The installation will continue with a progress bar indicating the installation progress.

When Anaconda is finished installing the “**Installation Complete**” screen will appear. Anaconda will now be installed on your system.



Click the “**Next**” button which will take you to an informational screen.

Press the “**Next**” button to be taken to the “**Completing Anaconda3 2020.02 (xx-bit) Setup**” screen. Uncheck the “**Anaconda Individual Tutorial**” and “**Learn More About Anaconda**” then click the “**Finish**” button.

Setting Up the Anaconda Environment

There are a few ways to load the Anaconda command line:

- Create a desktop short-cut
- Access Anaconda command line from the “**Start**” menu “**Programs**” option

- Access Anaconda through the **DOS** command line using **cmd** from the “**Start**” menu **run/search** facility.

For the sake of this exercise the “**Start**” menu “**Programs**” option is going to be used.

Go to the “**Start**” menu “**Programs**” and find “**Anaconda3 (xx-bit)**” then open up the folder.

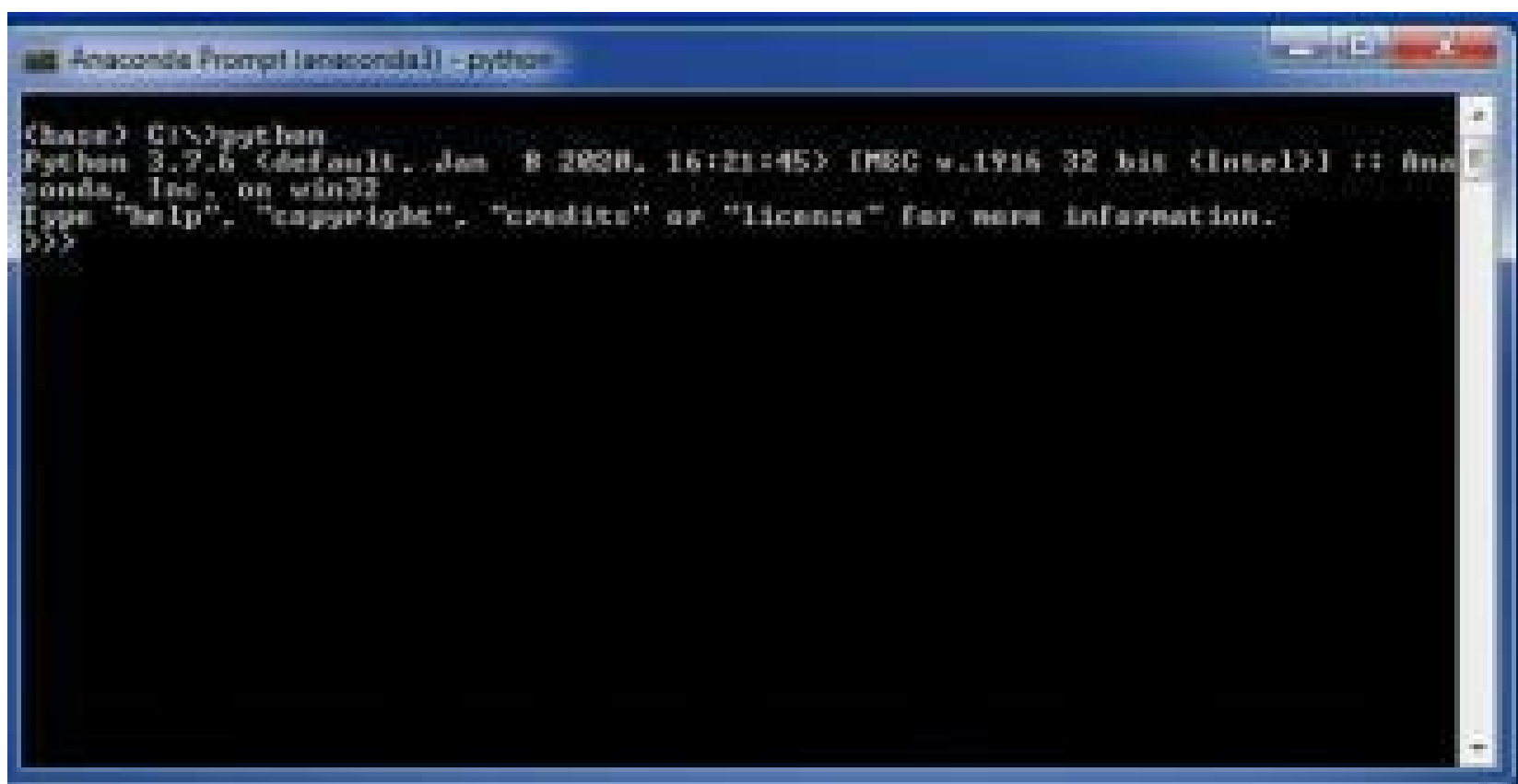
Choose “**Anaconda Prompt (anaconda3).**”



At the command line prompt type:

`python`

The Anaconda Python command line shell will load and look similar to the screen below.



If you want to test the Anaconda installation, import a file called “The Zen of Python” by Tim Peter by typing the following into the command line:

`import this`

Check that “**conda**” is installed in Anaconda, as it is the package installer for Anaconda.

If you have the **Anaconda Prompt** open and **Python** loaded, first exit the Python command line by typing the following:

`exit()`

This will take you back to the **Anaconda Prompt** and you can type in the following:

`conda --version`

This should show the version of conda. At the time of printing this book the latest version was 4.8.3. **Conda.exe** should come preloaded with the newer versions of Anaconda. However, they are not always the latest version of **conda** and may need to be upgraded. If the currently installed version of conda is older than version 4.8.3 then you should upgrade to the latest version.

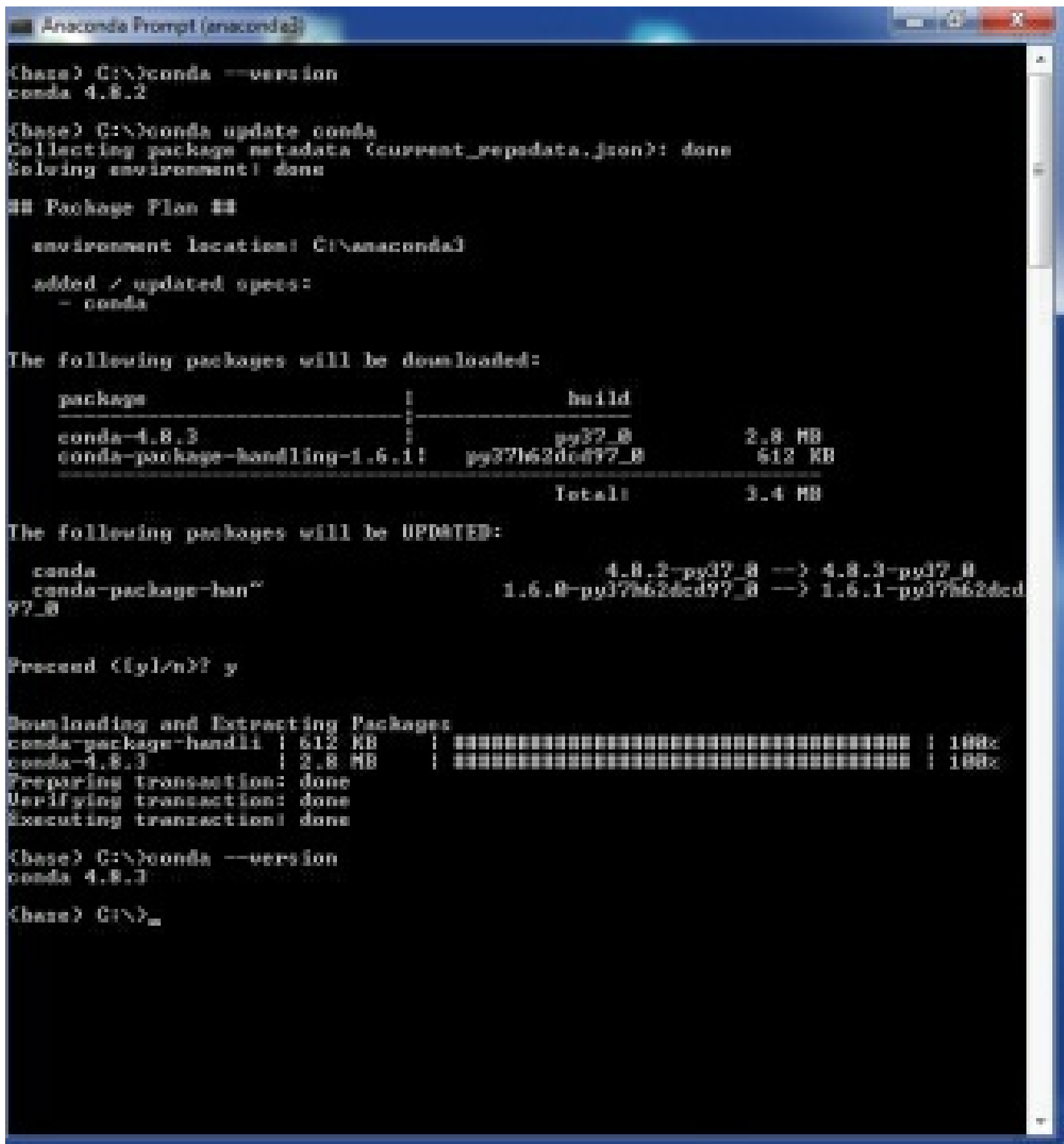
To upgrade **conda**, type the following into the **Anaconda Prompt** command line:

```
conda update conda
```

When the **conda** upgrade is finished type:

```
conda --version
```

The **conda** version should be 4.8.3

A screenshot of the Anaconda Prompt window. The terminal shows the command 'conda --version' returning 'conda 4.8.2'. Then 'conda update conda' is entered, followed by 'Collecting package metadata (current_repodata.json): done' and 'Solving environment: done'. A '## Package Plan ##' section shows the environment location as 'C:\anaconda3' and lists 'conda' as the package to be updated. A table lists the packages to be downloaded: 'conda-4.8.3' (2.8 MB) and 'conda-package-handling-1.6.1' (612 KB), with a total of 3.4 MB. Below this, it shows the packages to be updated: 'conda' from 4.8.2 to 4.8.3 and 'conda-package-handling' from 1.6.0 to 1.6.1. The user is prompted to proceed and types 'y'. The terminal then shows progress bars for downloading and extracting the packages, both at 100%. Finally, it shows 'Preparing transaction: done', 'Verifying transaction: done', and 'Executing transaction: done'. The command 'conda --version' is run again, returning 'conda 4.8.3'.

```
(base) C:\>conda --version
conda 4.8.2

(base) C:\>conda update conda
Collecting package metadata (current_repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: C:\anaconda3

added / updated specs:
- conda

The following packages will be downloaded:

package | build | size
-----|-----|-----
conda-4.8.3 | py37_0 | 2.8 MB
conda-package-handling-1.6.1 | py37h62ded97_0 | 612 KB
-----|-----|-----
Total: | 3.4 MB

The following packages will be UPDATED:

conda | 4.8.2-py37_0 --> 4.8.3-py37_0
conda-package-handling | 1.6.0-py37h62ded97_0 --> 1.6.1-py37h62ded97_0

Proceed [y/n]? y

Downloading and Extracting Packages:
conda-package-handling | 612 KB | 100%
conda-4.8.3 | 2.8 MB | 100%
Preparing transaction: done
Verifying transaction: done
Executing transaction: done

(base) C:\>conda --version
conda 4.8.3

(base) C:\>
```

The Anaconda **conda** utility should now be the latest version.

Install NumPy and SciPy With Conda

To find out if NumPy and SciPy are loaded, type in the following for a full list of loaded modules:

```
conda list
```

Scroll the list to ensure that Numpy and SciPy are loaded. If they are not loaded, you will have to install them by typing in the following at the **Anaconda Prompt**:

- **To download NumPy**
conda install numpy
- **To download SciPy**
conda install scipy

Installing Scikit-Learn

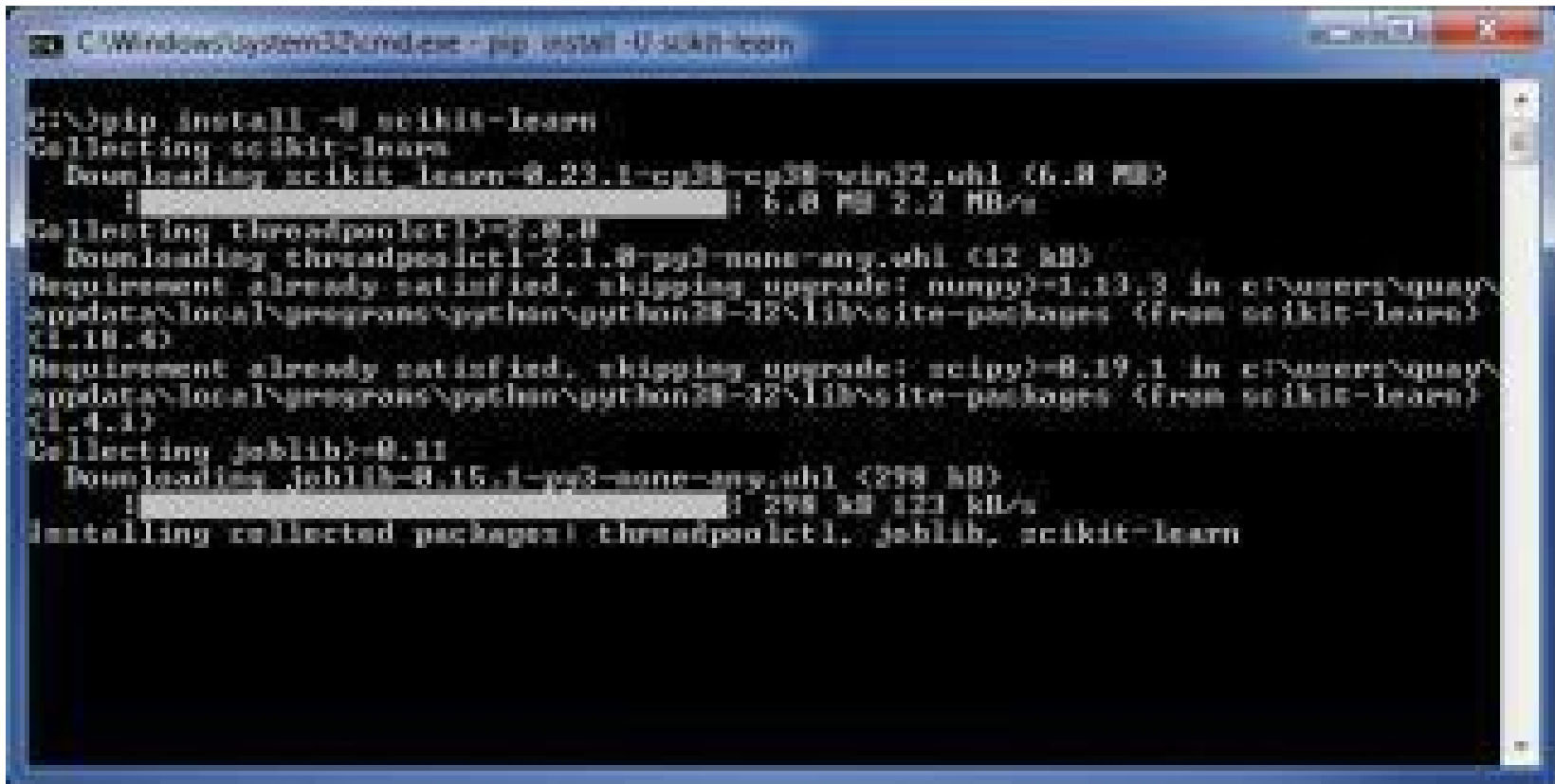
Either Python **pip** or Anaconda **conda** can be used to install Scikit-learn.

Installing Scikit-Learn with Python Pip Without a Virtual Environment

To install Scikit-Learn with Python and ensure it is the latest version that is compatible with your Python installation, open a **cmd** screen and type the following:

```
pip install -U scikit-learn
```

The following screen will appear when installing Scikit-Learn.



```
C:\Windows\system32\cmd.exe - pip install -U scikit-learn
C:\>pip install -U scikit-learn
Collecting scikit-learn
  Downloading scikit_learn-0.23.1-cp38-cp38-win32.whl (6.8 MB)
    |#####| 6.8 MB 2.2 MB/s
Collecting threadpoolctl>=2.0.0
  Downloading threadpoolctl-2.1.0-py3-none-any.whl (12 kB)
Requirement already satisfied, skipping upgrade: numpy>=1.13.3 in c:\users\quay\appdata\local\programs\python\python38-32\lib\site-packages (from scikit-learn) (1.18.4)
Requirement already satisfied, skipping upgrade: scipy>=0.19.1 in c:\users\quay\appdata\local\programs\python\python38-32\lib\site-packages (from scikit-learn) (1.4.1)
Collecting joblib>=0.11
  Downloading joblib-0.15.1-py3-none-any.whl (298 kB)
    |#####| 298 kB 123 kB/s
Installing collected packages: threadpoolctl, joblib, scikit-learn
```

Allow the installation to run to completion, which could take a few minutes.

Installing Scikit-Learn with Python Pip in a Virtual Environment

If you have an Anaconda installation, it is best to create a different virtual environment to run Scikit-Learn for Python in. To run this from the command line prompt, follow these instructions:

Install the virtual environment for Python:

```
pip install virtualenv
```

Create the virtual environment for Scikit-Learn:

```
python -m venv sklearn-venv
```

Once the environment has been set, type the following into the command line prompt to activate the sklearn-venv environment:

```
sklearn-venv\Scripts\activate
```

Once that script has run, you can then install Scikit-learn with the same commands as used above:

```
pip install -U scikit-learn
```

Allow the package installer to successfully install the library.

Checking the Python Pip Scikit-Learn Installation

To check the Scikit-Learn **pip** installation type the following into the command line once the installation has completed successfully:

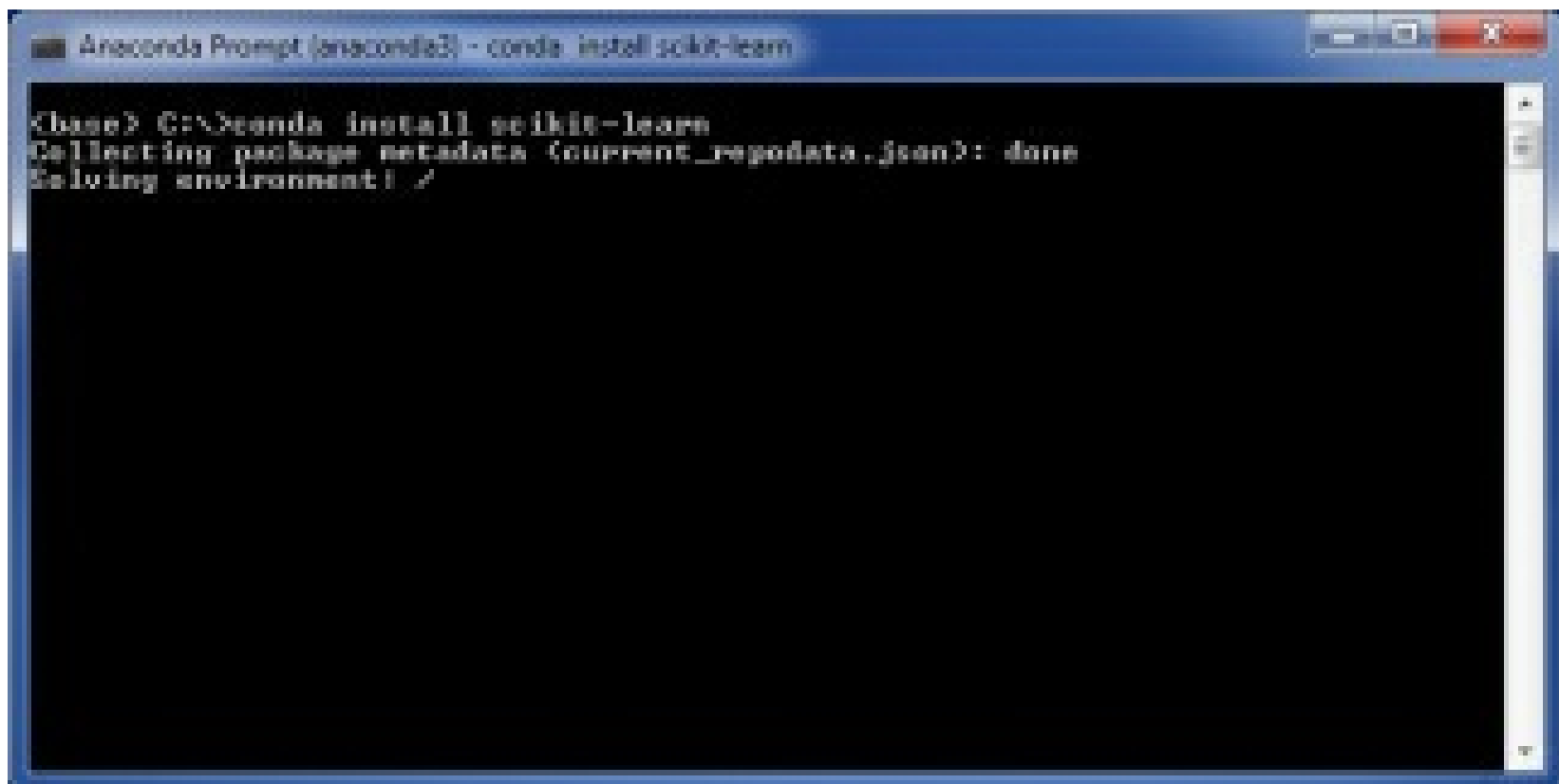
```
python -m pip show scikit-learn
```

This will bring up the version of Scikit-Learn (at the time of writing this book the version was 0.23.1) and where the library was installed.

Installing Scikit-Learn with Anaconda Conda Without a Virtual Environment

To install Scikit-Learn with Anaconda and ensure it is the latest version that is compatible with your Anaconda installation, run the **Anaconda Prompt** and type the following:

```
conda install scikit-learn
```

```

Anaconda Prompt (anaconda3) - conda: install scikit-learn

(base) C:\>conda install scikit-learn
Collecting package metadata (current_repodata.json): done
Solving environment: /

```

Allow the installation to complete. This can take several minutes.

Installing Scikit-Learn with Anaconda Conda in a Virtual Environment

If you have a Python environment, then it is best to create an environment to install and run Scikit-Learn for Anaconda that will not interfere with your Python installations.

To create a SciKit Learn environment from the **Anaconda Prompt**, do the following:

```
conda create -n sklearn-env
```

The next step is to activate the environment by typing the following into the **Anaconda Prompt**:

```
activate sklearn-env
```

Next, install Scikit-Learn into the Anaconda virtual environment. At the **Anaconda Prompt**, type the following:

```
conda install scikit-learn
```

Allow the installation to run through to completion.

Checking the Anaconda Conda Scikit-Learn Installation

To check the Scikit-Learn **conda** installation, type the following into the command line once the installation has completed successfully:

```
conda list scikit-learn
```

This will bring up the version of Scikit-Learn (at the time of writing this book the version was 0.21.1) and where the library was installed.

Installing TensorFlow

TensorFlow is a framework that is required for deep learning. It comes with a myriad of functionalities that allow a system to carry out deep learning functions. TensorFlow also has APIs that allow for the interaction between nearly all the most popular programming languages. The programming languages include Rust, Haskell, Java, C++, and Go.

As with most Python or Python compatible libraries, TensorFlow is free to download and use from the Internet as it is open source software. TensorFlow can be installed using either Python **pip** or Anaconda **conda** package installation programs. For the sake of this book the CPU-only version of TensorFlow is to be installed.

Installing TensorFlow With Python Pip Without a Virtual Environment

To install TensorFlow with Python and ensure it is the latest version that is compatible with your Python installation, open a **cmd** screen and type the following:

```
pip install -U tensorflow
```

The above command will help you install the **CPU** only version for TensorFlow. The **CPU** version of TensorFlow is faster and is a much easier system for beginners to work with. It is also recommended for the simpler machine learning models. When you are first starting out with machine learning model's, **CPU** is a lot easier to work with to design and train these types of models.

If you need to install a **GPU** version for TensorFlow, type the following command into the **cmd** command line:

```
pip install -U tensorflow-gpu
```

This will install the TensorFlow **GPU** version for your Windows system. **GPU** is a more advanced modelling system of TensorFlow and is used for huge amounts of data and graphics/images. It is also used to work with more complex tasks than **CPU** and can be quite a bit slower than **CPU**.

Installing TensorFlow With Python Pip with a Virtual Environment

If you created the sklearn-venv environment when you installed Scikit-Learn, then you should install TensorFlow into that environment. You can activate that environment by typing in the following at the **cmd** prompt:

```
sklearn-venv\Scripts\activate
```

Follow the same installation process for installing Scikit-Learn as you would when you are installing it without the virtual environment.

Installing TensorFlow With Anaconda Conda Without a Virtual Environment

To install TensorFlow with Anaconda and ensure it is the latest version that is compatible with your Anaconda installation, open the **Anaconda Prompt** screen and type the following:

```
conda install tensorflow
```

The above command will help you install the **CPU** only version for TensorFlow.

If you need to install a **GPU** version for TensorFlow, type the following command into the **Anaconda Prompt** command line:

```
conda install tensorflow-gpu
```

This will install the TensorFlow **GPU** version for your Windows system.

Installing TensorFlow With Anaconda Conda With a Virtual Environment

If you have already created a conda virtual environment for Scikit-Learn, then use the same environment. If you have not, follow the same procedure to create an Anaconda environment as shown in the “**Installing Scikit-Learn in Anaconda with a Virtual Environment**.” Once you have the Anaconda virtual environment, activate it using the following command:

```
activate sklearn-env
```

Install TensorFlow by typing in:

```
conda install tensorflow
```

This will install the TensorFlow CPU version for your Windows system.