

Getting Started with Python and Anaconda - Beginner's Guide

Welcome to your data bootcamp! This guide will help you set up Python and Anaconda on your computer and verify everything is working correctly.

What is Anaconda?

Anaconda is a free, open-source distribution of Python that includes all the essential tools for data science. It comes with:

- Python programming language
- Jupyter Lab (where you'll write your code)
- Popular data science libraries like pandas and numpy
- A package manager to install additional tools

Step 1: Download Anaconda

1. Visit the official Anaconda website: <https://www.anaconda.com/download>
2. The website will automatically detect your operating system (Windows, macOS, or Linux)
3. Click the **Download** button for the latest version
4. The file is large (around 500 MB), so the download may take several minutes

Important: Make sure you download the Python 3.x version (not Python 2.x)

Step 2: Install Anaconda

For Windows Users:

1. Locate the downloaded file (usually in your Downloads folder) - it will be named something like `Anaconda3-2024.xx-Windows-x86_64.exe`
2. Double-click the installer to run it
3. Click **Next** when the setup wizard appears
4. Click **I Agree** to accept the license agreement
5. Select **Just Me** (recommended) and click **Next**
6. Choose the installation location (the default is fine) and click **Next**
7. **Important:** Check the box that says **Add Anaconda to my PATH environment variable** (even though it's not recommended, it makes things easier for beginners)
8. Click **Install** and wait for the installation to complete (this may take 10-15 minutes)

9. Click **Finish** when done

For macOS Users:

1. Locate the downloaded file (usually in your Downloads folder) - it will be named something like `Anaconda3-2024.xx-MacOSX-x86_64.pkg`
2. Double-click the `.pkg` file to start the installer
3. Click **Continue** through the introduction screens
4. Click **I Agree** to accept the license agreement
5. Select your hard drive and click **Continue**
6. Click **Install** (you may need to enter your password)
7. Wait for the installation to complete
8. Click **Close** when finished

For Linux Users:

1. Open your terminal
2. Navigate to the directory where you downloaded the file
3. Run the command: `bash Anaconda3-2024.xx-Linux-x86_64.sh` (replace with your actual filename)
4. Press **Enter** to continue
5. Read the license agreement and type **yes** to accept
6. Press **Enter** to confirm the installation location
7. Type **yes** when asked to initialize Anaconda
8. Close and reopen your terminal

Step 3: Verify the Installation

1. **Windows:** Open the **Anaconda Prompt** from your Start menu
macOS/Linux: Open your **Terminal** application
2. Type the following command and press Enter:

```
python --version
```

1. You should see something like `Python 3.11.5` (your exact version number may differ)
2. Type this command to check conda:

```
conda --version
```

1. You should see something like `conda 24.1.2`

If both commands work, congratulations! Anaconda is installed correctly.

Step 4: Launch Jupyter Lab

Jupyter Lab is where you'll write and run your Python code. Here's how to open it:

1. **Windows:** Open **Anaconda Navigator** from your Start menu, then click the **Launch** button under Jupyter Lab *OR* open **Anaconda Prompt** and type:
2. **macOS/Linux:** Open your **Terminal** and type:
3. Press **Enter** and wait a few seconds
4. Jupyter Lab will automatically open in your web browser
5. You should see a file browser on the left and a launcher tab in the main area

Note: Don't close the Anaconda Prompt or Terminal window while using Jupyter Lab - it needs to stay open.

Step 5: Create Your First Notebook

1. In the Jupyter Lab launcher, click on the **Python 3** button under "Notebook"
2. A new notebook file will open with an empty cell
3. The notebook will be named "Untitled.ipynb" - you can rename it by right-clicking on the name at the top


Step 6: Test Your Installation with Code

Now let's verify that pandas and numpy are working correctly. Follow these steps carefully:

Import pandas

1. In the first empty cell, type:

```
import pandas as pd
```

1. Press **Shift + Enter** to run the cell (or click the play button  in the toolbar)
2. If successful, a new empty cell will appear below and you won't see any error messages

Import numpy

1. In the new cell, type:

```
import numpy as np
```

1. Press **Shift + Enter** to run the cell
2. Again, if successful, you won't see any errors

Test pandas

Let's create a simple DataFrame to make sure pandas works:

1. In the next cell, type:

```
# Create a simple DataFrame
data = {'Name': ['Alice', 'Bob', 'Charlie'],
        'Age': [25, 30, 35],
        'City': ['New York', 'London', 'Paris']}

df = pd.DataFrame(data)
print(df)
```

1. Press **Shift + Enter**
2. You should see a nicely formatted table with names, ages, and cities

Test numpy

Let's perform a simple numpy operation:

1. In the next cell, type:

```
# Create an array and perform calculations
numbers = np.array([1, 2, 3, 4, 5])
print("Original array:", numbers)
print("Mean:", np.mean(numbers))
print("Sum:", np.sum(numbers))
print("Array squared:", numbers ** 2)
```

1. Press **Shift + Enter**
2. You should see the array and the results of various calculations

Understanding What You Just Did

-

import pandas as pd: This loads the pandas library and lets you refer to it as "pd" (a common shorthand)

- **import numpy as np:** This loads the numpy library and lets you refer to it as "np"
- **pd.DataFrame():** Creates a data table (like an Excel spreadsheet)
- **np.array():** Creates a numpy array (a list of numbers optimized for calculations)
- **np.mean(), np.sum():** Functions to calculate average and sum

Troubleshooting Common Issues

Issue: "jupyter: command not found" or "conda: command not found"

Solution: The installation path wasn't added correctly. Try:

- Restart your computer
- Reopen Anaconda Prompt (Windows) or Terminal (macOS/Linux)
- If still not working, you may need to reinstall Anaconda

Issue: "ModuleNotFoundError: No module named 'pandas'" or similar

Solution: The package might not be installed. Run this in a cell:

```
!conda install pandas numpy -y
```

Issue: Jupyter Lab won't open in browser

Solution:

- Check if any firewall is blocking it
- Try copying the URL from the terminal and pasting it manually in your browser
- Look for a line that says `http://localhost:8888/lab?token=...`

Next Steps

Congratulations! You've successfully:

- Installed Anaconda
- Launched Jupyter Lab
- Created your first notebook
- Verified pandas and numpy are working

You're now ready to start learning Python for data science!

Tips for Beginners

- **Save your work frequently:** Click the save icon or press **Ctrl+S** (Windows/Linux) or **Cmd+S** (macOS)
- **Run cells in order:** Code cells should generally be run from top to bottom
- **Don't worry about errors:** Everyone makes mistakes! Read the error message - it often tells you what went wrong
- **Restart if needed:** If things get messy, go to **Kernel 'Restart Kernel** in the menu
- **Keep the terminal open:** Don't close the Anaconda Prompt or Terminal window while using Jupyter Lab

Shutting Down Jupyter Lab

When you're done working:

1. Save your notebook
2. Go to **File 'Shut Down** in Jupyter Lab
3. Close the browser tab
4. In the Anaconda Prompt or Terminal window, press **Ctrl+C** twice to stop the server
5. Close the terminal window

Welcome to your data science journey!