TOOL

Core Python Data Science Ecosystem Components

Python Standard Library	
Where:	https://docs.python.org/3/library/index.html
What:	A diverse set of modules included in all Python distributions providing a wide range of functions and the ability to interact with the operating system.
When:	When you need to manipulate files on your computer, perform some mathematical operations, read files in some standard formats, process text strings that you have read in from a file, figure out where your program is spending most of its time when it runs... the list goes on.

IPython	
Where:	ipython.org
What:	An extension of the default Python interpreter that provides powerful interactive capabilities to enhance your productivity.
When:	When you need to interactively explore the data with which you are working or develop new data science pipelines by figuring out the tools you need and how they can best work with each other.

Jupyter	
Where:	jupyter.org
What:	A system that enables you to create and share documents integrating code, documentation, commentary, results, and data visualizations.
When:	When you want to put together a coherent and reproducible analysis documenting a problem on which you are working, its solution, and the results of your analyses.

NumPy	
Where:	numpy.org
What:	The fundamental package for scientific computing in Python, providing multidimensional arrays and a variety of functions for acting on the data stored in arrays.
When:	When you want powerful number-crunching capabilities to work with array data, or as part of the larger Python data science ecosystem, which builds new operations on top of the core functionality provided by NumPy.

SciPy	
Where:	scipy.org
What:	An integrated set of packages for computational mathematics, science, and engineering, providing convenient Python interfaces to many well-established algorithms for scientific computing
When:	When you want to find the roots of an equation, identify the parameters at the minimum value of a function, integrate a differential equation, fit your data to a model, do some signal or image processing, interpolate between two data points... the list goes on.

Matplotlib	
Where:	matplotlib.org
What:	A 2D plotting package that produces publication-quality figures in a variety of formats, with a high degree of customization.
When:	When you need to make plots of your data, including line plots, scatter plots, bar charts, heatmaps, histograms, etc.

Seaborn	
Where:	seaborn.pydata.org
What:	A data visualization package based on Matplotlib that provides a high-level interface for characterizing statistical properties of data.
When:	When you need to visualize structure in your data set, overlaying different components of data into informative visual summaries, with the ability for customization through Matplotlib.

Bokeh	
Where:	bokeh.org
What:	A data visualization package that enables the development of interactive, web-based data visualizations.
When:	When you want to be able to display and interact visually with complex data sets or serve up such visualizations through the web.

Pandas	
Where:	pandas.pydata.org
What:	A fast, powerful, flexible, and easy-to-use data analysis and manipulation tool, especially suited to working with tabular data such as is contained in spreadsheets.
When:	When you need to read in data from an Excel spreadsheet or CSV file and then process it further by writing your own analysis code.

Scikit-learn	
Where:	scikit-learn.org
What:	A powerful and comprehensive package for doing machine learning in Python, with support for many algorithms as well as testing and cross-validation procedures.
When:	When you want to build predictive models from data, such as classification or regression pipelines (supervised learning) or clustering and dimensionality reduction analyses (unsupervised learning).

Tensorflow	
Where:	tensorflow.org
What:	A powerful framework for doing neural network-based deep learning.
When:	When you want to build machine learning pipelines, such as those that can recognize handwritten letters and digits or identify cats in images on the web.

Caffe	
Where:	caffe.berkeleyvision.org
What:	A powerful framework for doing neural network-based deep learning.
When:	When you want to build machine learning pipelines, such as those that can recognize handwritten letters and digits or identify cats in images on the web.

PyTorch	
Where:	pytorch.org
What:	A powerful framework for doing neural network-based deep learning.
When:	When you want to build machine learning pipelines, such as those that can recognize handwritten letters and digits or identify cats in images on the web.

NetworkX	
Where:	networkx.github.io
What:	A package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks.
When:	When you are analyzing data describing networks (such as the electrical grid, airline transporation networks, metabolic networks in cells, etc.) or when you want to characterize the web of relationships between different components in a system you are analyzing.

Statsmodels	
Where:	statsmodels.org
What:	A package for the statistical analysis of data.
When:	When you want to do statistical data exploration, develop statistical models of data and perform analyses with those models (as in a regression analysis), or conduct statistical tests to assess the significance of relationships you have uncovered.

Scikits	
Where:	www.scipy.org/scikits.html
What:	A broad collection of add-on packages for SciPy, hosted and developed separately but providing a wide variety of more application-specific functionality that can be of use. Scikit-learn is probably the most widely known, but there are many others.
When:	When you want to do image processing (scikit-image), create virtural environments (scikit-vi), geostatistical analysis (scikit-gstat)... the list goes on.

PyPI	
Where:	pypi.org
What:	The Python Package Index: a repository of software for the Python programming language.
When:	When you want to figure out whether a package might already be available to support your work so that you don't have to reinvent the wheel.