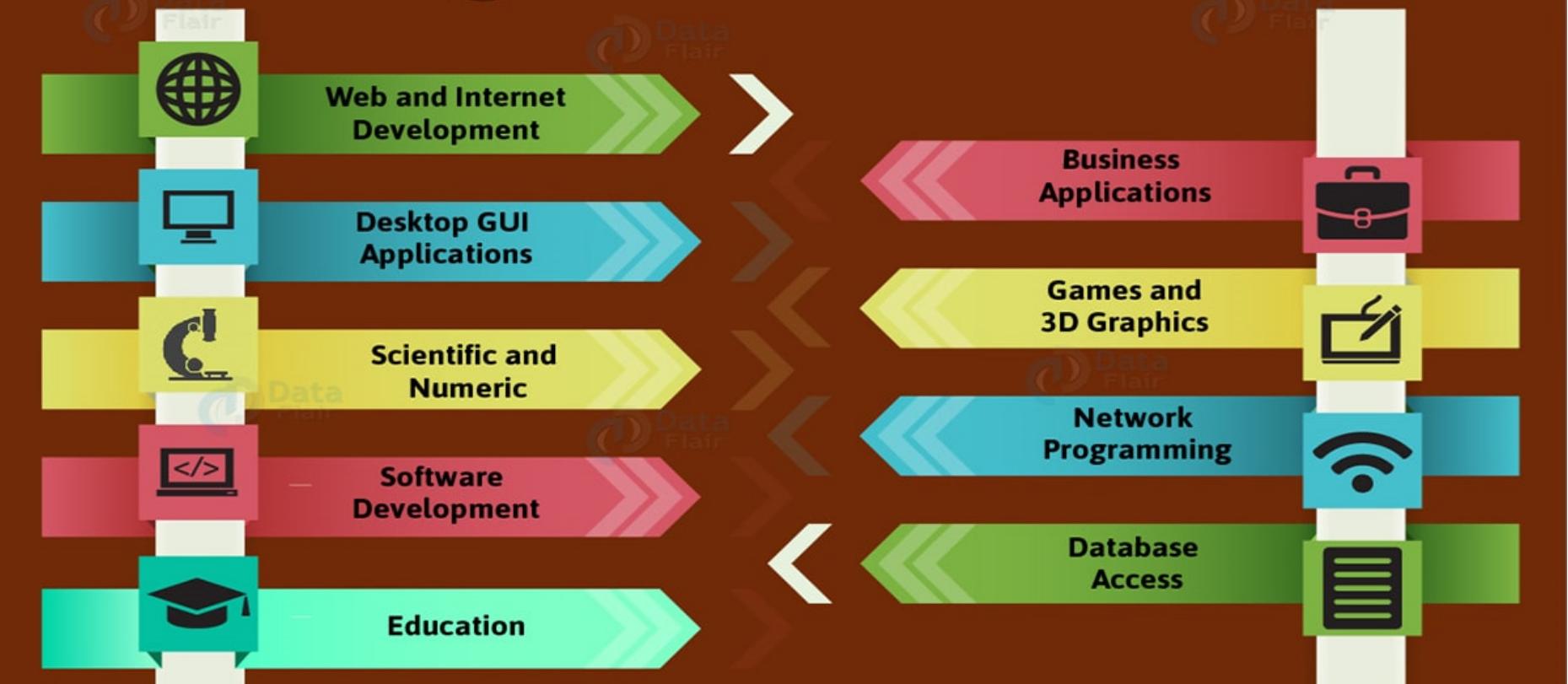


# Python + Machine Learning





# Python Applications



<https://data-flair.training/blogs/python-applications/>

# Top 10 Libraries in Python to Implement Machine Learning



Mantra Malhotra [Follow](#)

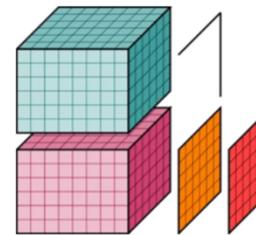
Sep 2, 2018 · 8 min read



<https://hackernoon.com/top-10-libraries-in-python-to-implement-machine-learning-12602cf5dc61>



StatsModels  
Statistics in Python



xarray



scikits-image  
image processing in python



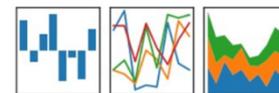
machine learning in Python



And many,  
many more...



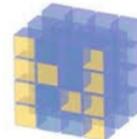
pandas  
 $y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$



matplotlib



bokeh



NumPy



IP[y]:  
IPython

python™



# Distribución Python



- Hay muchas distribuciones Python que incluyen todas las librerías.
- Se recomienda Anaconda
  - Principales características.
  - Gestión de paquetes via Conda o PIP
  - Varias GUI. Spyder, Jupiter, etc.
  - Versiones Linux, Mac y Windows. 32 y 64 bits.

Anaconda 2018.12 for Windows Installer

Python 3.7 version

[Download](#)

64-Bit Graphical Installer (614.3 MB) 

32-Bit Graphical Installer (509.7 MB)

Python 2.7 version

[Download](#)

64-Bit Graphical Installer (560.6 MB)

32-Bit Graphical Installer (458.6 MB)

<https://www.anaconda.com/distribution>

# Anaconda

The screenshot shows the Anaconda Distribution homepage. At the top, there's a navigation bar with links for Products, Why Anaconda?, Solutions, Resources, Company, Download, and a search bar. Below the navigation is a large green header section with the title "Anaconda Distribution" and the subtitle "The World's Most Popular Python/R Data Science Platform". A prominent "Download" button is located in the center of this green area. The main content area features a white background with a green decorative border on the right side. It contains a paragraph about the Anaconda Distribution, a bulleted list of features, and a grid of logos for various data science libraries. At the bottom, there are two sections for "Python 3.7 version" and "Python 2.7 version", each with download buttons and file sizes.

The open-source [Anaconda Distribution](#) is the easiest way to perform Python/R data science and machine learning on Linux, Windows, and Mac OS X. With over 11 million users worldwide, it is the industry standard for developing, testing, and training on a single machine, enabling *individual data scientists* to:

- Quickly download 1,500+ Python/R data science packages
- Manage libraries, dependencies, and environments with [Conda](#)
- Develop and train machine learning and deep learning models with [scikit-learn](#), [TensorFlow](#), and [Theano](#)
- Analyze data with scalability and performance with [Dask](#), [NumPy](#), [pandas](#), and [Numba](#)
- Visualize results with [Matplotlib](#), [Bokeh](#), [DataShader](#), and [HoloViews](#)

## Python 3.7 version

[Download](#)

64-Bit Graphical Installer (652.7 MB)  
64-Bit Command Line Installer (557 MB)

## Python 2.7 version

[Download](#)

64-Bit Graphical Installer (640.7 MB)  
64-Bit Command Line Installer (547 MB)

<https://www.anaconda.com/distribution/>

Anaconda Navigator

File Help

# ANACONDA NAVIGATOR

Sign in to Anaconda Cloud

Home Environments Learning Community Documentation Developer Blog Twitter YouTube GitHub

Applications on base (root) Channels Refresh

Icon	Name	Version	Description	Action
	JupyterLab	0.35.3	An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.	<a href="#">Launch</a>
	Notebook	5.7.4	Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.	<a href="#">Launch</a>
	Qt Console	4.4.3	PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.	<a href="#">Launch</a>
	Spyder	3.3.2	Scientific Python Development Environment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features	<a href="#">Launch</a>
	Glueviz	0.13.3	Multidimensional data visualization across files. Explore relationships within and among related datasets.	<a href="#">Install</a>
	Orange 3	3.17.0	Component based data mining framework. Data visualization and data analysis for novice and expert. Interactive workflows with a large toolbox.	<a href="#">Install</a>
	RStudio	1.1.456	A set of integrated tools designed to help you be more productive with R. Includes R essentials and notebooks.	<a href="#">Install</a>
	VS Code	1.31.1	Streamlined code editor with support for development operations like debugging, task running and version control.	<a href="#">Install</a>

# Introducción a Spyder

# Introducción

- Existen varios IDE que se pueden utilizar para desarrollar proyectos de ML, la gran mayoría son gratis y ya depende de cada desarrollador seleccionar el mas conveniente
  - Spyder
  - PyCharm
  - Thonny
  - Atom
  - Jupyter Notebook



[https://www.datacamp.com/community/tutorials/  
data-science-python-ide](https://www.datacamp.com/community/tutorials/data-science-python-ide)

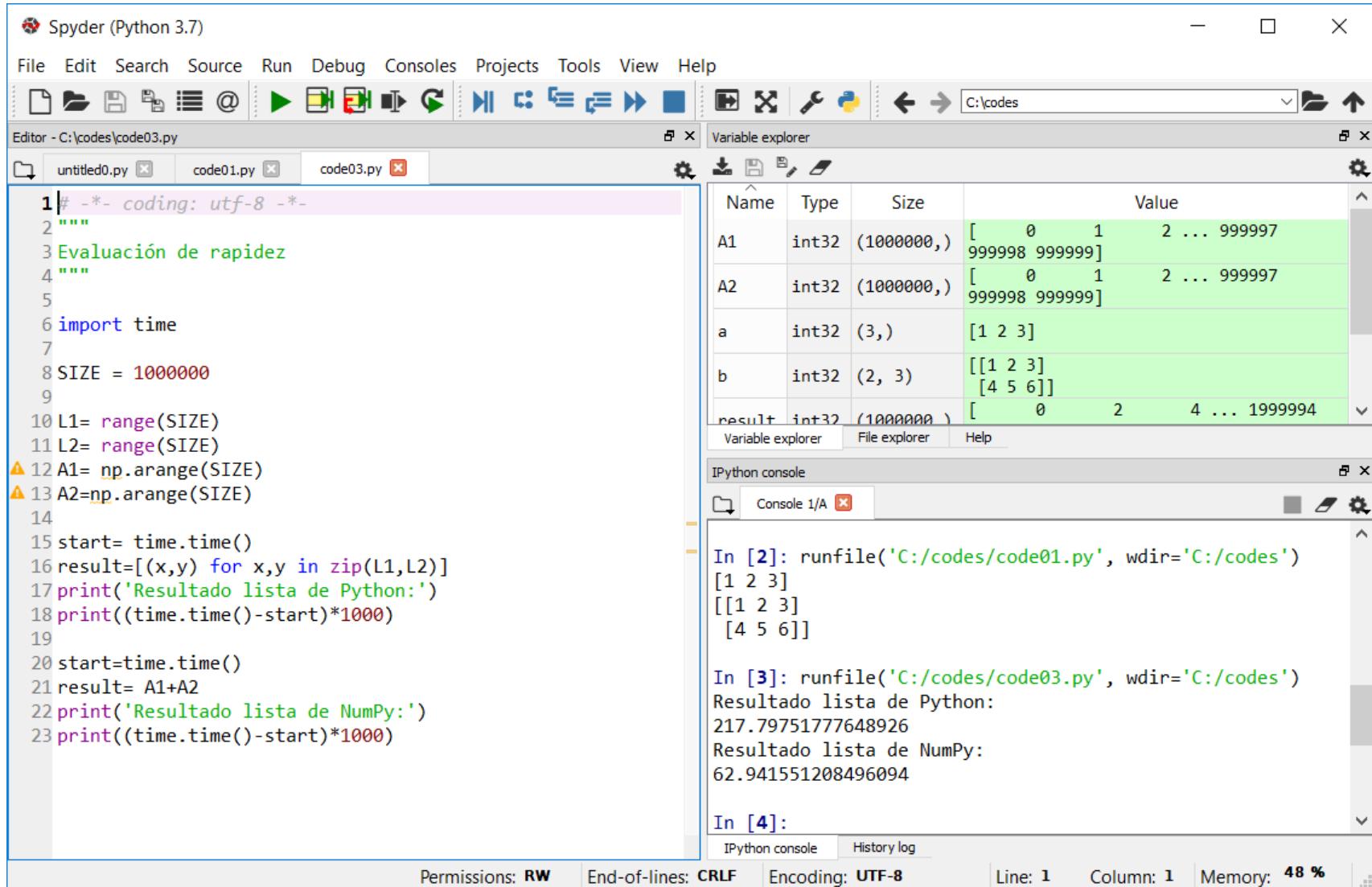
# Spyder

- Incluido en Anaconda
- Es de código abierto y esta escrito en Python para el desarrollo de Python con un enfoque en investigación, análisis de datos y creación de paquetes científicos
- Spyder (*Scientific PYthon Developoment EnviRonment*)
- Cuenta con una interfaz de usuario funcional con opciones interactivas, diseños personalizables y paneles intercambiables



<https://www.spyder-ide.org/>

# Apariencia de Spyder



# **Python - Librerías esenciales para el análisis de datos**

# Librerías esenciales para el análisis de datos

- Numpy
- Matplotlib
- Ipython
- Pandas
- SciPy
- Scikit-learn

- Python tiene listas, enteros, punto flotante, etc.
- Para cálculo numérico necesitamos más, allí aparece ...



[Métodos de numpy](#)



- Jupyter Notebook, es un entorno de trabajo interactivo que permite desarrollar código en Python de manera dinámica, a la vez que integrar en un mismo documento tanto bloques de código como texto, gráficas o imágenes
- Es un SaaS (*Software as a Service*) utilizado ampliamente en análisis numérico, estadística y machine learning, entre otros campos de la informática y las matemáticas

<https://jupyter.org/>

<https://azure.microsoft.com/es-mx/overview/what-is-saas/>



- Matplotlib es la librería más popular en Python para visualizaciones y gráficos. Puede producir gráficos de alta calidad dignos de cualquier publicación científica
- Algunas de las muchas ventajas que nos ofrece Matplotlib, incluyen:
  - Es fácil de aprender
  - Soporta texto, títulos y etiquetas en formato LATEX
  - Proporciona un gran control sobre cada uno de los elementos de las figuras, como ser su tamaño, el trazado de sus líneas, etc.
  - Permite crear gráficos y figuras de gran calidad en varios formatos: PNG, PDF, SVG, EPS, y PGF
- Matplotlib se integra con IPython, lo que proporciona un ambiente adecuado para las visualizaciones y la exploración de datos interactiva

# IP[y]:

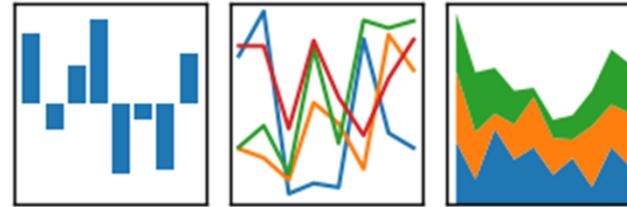
## IPython

- IPython es un shell interactivo que añade funcionalidades extra al modo interactivo incluido con Python, como resaltado de líneas y errores mediante colores, una sintaxis adicional para el shell, autocompletado mediante tabulador de variables, módulos y atributos
- Desde 2011, IPython ha sufrido una evolución muy importante con la aparición de IPython Notebook

The image displays two windows side-by-side. The left window is a standard terminal window titled 'C:\Anaconda3\python.exe'. It shows the Python 3.7.1 startup message, including the date (Dec 10 2018), time (22:54:23), and build information (MSC v.1915 64 bit (AMD64)). It also includes help instructions and a prompt starting with '>>>'. The right window is a Jupyter QtConsole window titled 'Jupyter QtConsole'. It has a menu bar with File, Edit, View, Kernel, Window, and Help. Below the menu is the version information 'Jupyter QtConsole 4.4.3' and the Python startup message. At the bottom, it shows the input prompt 'In [1]: |'.

# pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



- Pandas es una librería open source que aporta a Python unas estructuras de datos fáciles de usar y de alto desempeño, junto con un gran número de funciones esenciales para el análisis de datos
- Con Pandas se puede trabajar con datos estructurados de una forma más rápida y expresiva.



- SciPy es un conjunto de paquetes para distintos campos de la computación científica y el análisis numérico
- Algunos de los subpaquetes que incluye, son:
  - **scipy.integrate**: que proporciona diferentes funciones para resolver problemas de integración numérica
  - **scipy.signal**: para el análisis y procesamiento de señales
  - **scipy.optimize**: para los problemas de optimización y minimización.
  - **scipy.sparse**: para matrices dispersas y solucionar sistemas lineales dispersos
  - **scipy.linalg**: que proporciona funciones para resolver problemas de álgebra lineal.
  - **scipy.stats**: para el análisis de estadística y probabilidades.



- Scikit-learn es una librería especializada en algoritmos para data mining y machine learning
- Algunos problemas que pueden resolverse utilizando las herramientas de scikit-learn son:
  - **Clasificación:** Identificar las categorías a que cada observación del conjunto de datos pertenece.
  - **Regresión:** Predecire el valor continuo para cada nuevo ejemplo.
  - **Clustering:** Agrupación automática de objetos similares en un conjunto.
  - **Reducción de dimensionalidad:** Reducir el número de variables aleatorias a considerar.
  - **Selección de modelos:** Comparar, validar y elegir parámetros y modelos.
  - **Preprocesamiento:** Eliminación de outliers, normalización de datos.



- TensorFlow es una biblioteca más compleja para el cálculo numérico distribuido que utiliza grafos de flujo de datos.
- Permite entrenar y ejecutar ANN muy grandes de forma eficiente distribuyendo los cálculos a través de miles de servidores multi-GPU
- TensorFlow fue creado en Google y es compatible con muchas de sus aplicaciones de ML de gran escala
- Fue hecho código abierto en noviembre de 2015

<https://www.nextplatform.com/2015/12/11/inside-the-gpu-clusters-that-power-baidus-neural-networks/>

