

# 데이터 분석 패키지

---



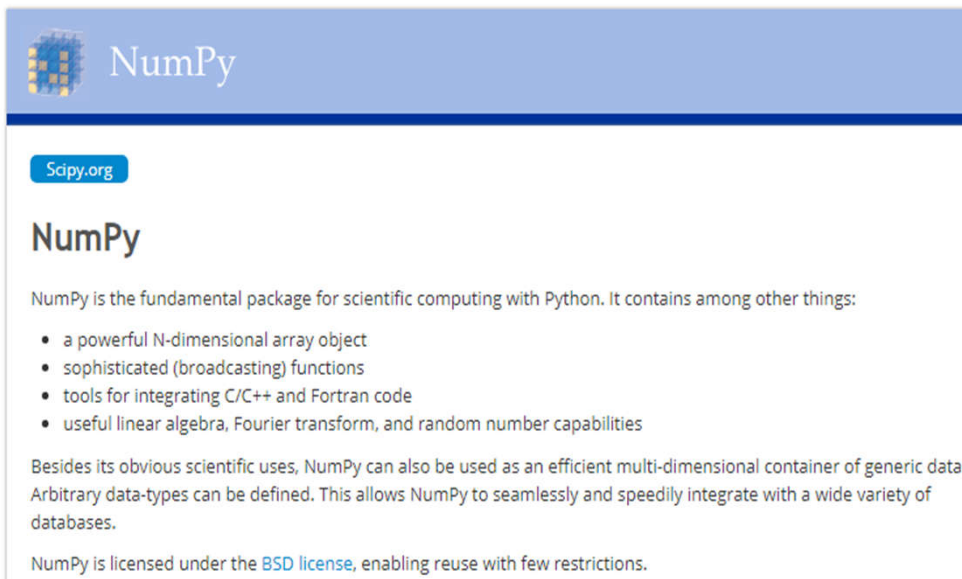
# 파이썬 주요 데이터분석 패키지들 소개



# 주요 패키지 소개

## ▪ Numpy

- <http://www.numpy.org/>
- C로 구현된 고성능 수치계산을 위한 라이브러리.
- 고성능 과학연산을 위한 패키지로 데이터분석, 머신러닝등에 필수로 사용된다
- 벡터, 행렬 연산관련 다양한 기능을 제공.
- Pandas, matplotlib, scikit-learn의 기반이 되는 패키지.

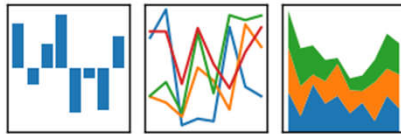


# 주요 패키지 소개

## ▪ Pandas

- 데이터 분석시 가장 많이 다루게 되는 표 형태의 데이터셋을 쉽게 다루는 기능을 제공하는 패키지.
- csv나 excel파일을 읽어 분석하는데 사용하는 다양한 기능을 제공.
- <https://pandas.pydata.org>

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



[home](#) // [about](#) // [get pandas](#) // [documentation](#) // [community](#) // [talks](#) // [donate](#)

## Python Data Analysis Library

*pandas* is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the [Python](#) programming language.

*pandas* is a [NumFOCUS](#) sponsored project. This will help ensure the success of development of *pandas* as a world-class open-source project, and makes it possible to [donate](#) to the project.

A Fiscally Sponsored Project of

**NUMFOCUS**  
OPEN CODE = BETTER SCIENCE

## VERSIONS

### Release

0.23.4 - August 2018

[download](#) // [docs](#) // [pdf](#)

### Development

0.24.0 - 2018

[github](#) // [docs](#)

### Previous Releases

0.23.3 - [download](#) // [docs](#) // [pdf](#)

0.23.2 - [download](#) // [docs](#) // [pdf](#)

0.23.1 - [download](#) // [docs](#) // [pdf](#)

0.23.0 - [download](#) // [docs](#) // [pdf](#)

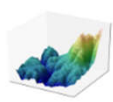
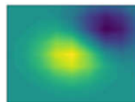
# 주요 패키지 소개 – 시각화 패키지

- matplotlib
  - 파이썬 시각화의 기본이 되는 패키지
  - <https://matplotlib.org>
- Seaborn
  - matplotlib를 기반으로 하는 시각화 패키지
  - matplotlib보다 좀더 쉽게 그래프를 그릴 수 있도록 한다.
  - <https://seaborn.pydata.org>



[home](#) | [examples](#) | [tutorials](#) | [API](#) | [docs](#) »

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter notebook, web application servers, and four graphical user interface toolkits.



seaborn 0.9.0 Gallery Tutorial API Site Page Search

## seaborn: statistical data visualization

Seaborn is a Python data visualization library based on [matplotlib](#). It provides a high-level interface for drawing attractive and informative statistical graphics.

For a brief introduction to the ideas behind the library, you can read the [introductory notes](#). Visit the [installation page](#) to see how you can download the package. You can browse the [example gallery](#) to see what you can do with seaborn, and then check out the [tutorial](#) and [API reference](#) to find out how.

### Contents

- [Introduction](#)
- [Release notes](#)
- [Installing](#)
- [Example gallery](#)

### Features

- [Relational: API | Tutorial](#)
- [Categorical: API | Tutorial](#)
- [Distributions: API | Tutorial](#)
- [Regressions: API | Tutorial](#)

# 주요 패키지 소개

- scikit-learn
  - 머신러닝(기계학습) 관련 기능을 제공하는 패키지.
  - <https://scikit-learn.org>



## Classification

Identifying to which category an object belongs to.

**Applications:** Spam detection, Image recognition.

**Algorithms:** SVM, nearest neighbors, random forest, ... — Examples

## Regression

Predicting a continuous-valued attribute associated with an object.

**Applications:** Drug response, Stock prices.

**Algorithms:** SVR, ridge regression, Lasso, ... — Examples

## Clustering

Automatic grouping of similar objects into sets.

**Applications:** Customer segmentation, Grouping experiment outcomes

**Algorithms:** k-Means, spectral clustering, mean-shift, ... — Examples