**PANDAS**

**Pandas** is used for data manipulation, analysis and cleaning. Python pandas are well suited for different kinds of data, such as: Tabular data with heterogeneously-typed columns. Ordered and unordered time series data.



Pandas is the most popular python library that is used for data analysis. It provides highly optimized performance with back-end source code is purely written in C or Python. **It** is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool  
built on top of the python programming language.

We can analyze data in pandas with:

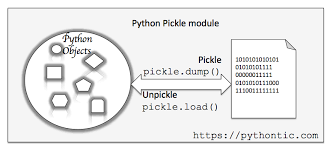
1. Series

2. Data Frames

**PYTHON PICKLE**

Python **PICKLE module** is used for serializing and de-serializing a Python object structure. ... Pickling is a way to convert a python object (list, dictionary, etc.) into a character stream. The idea is that this character stream contains all the information necessary to reconstruct the object in another python script.



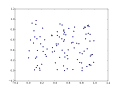
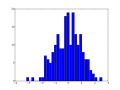


Python pickle module is used for serializing and de-serializing a Python object structure. Any object in Python can be pickled so that it can be saved on disk. What pickle does is that it “serializes” the object first before writing it to file. Pickling is a way to convert a python object (list, dictionary, etc.) into a character stream. The idea is that this character stream contains all the inform

**MATPLOTLIB**

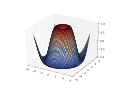
Matplotlib is a [plotting](https://en.wikipedia.org/wiki/Plotter) [library](https://en.wikipedia.org/wiki/Library_(computer_science)) for the [Python](https://en.wikipedia.org/wiki/Python_(programming_language)) programming language and its numerical mathematics extension [NumPy](https://en.wikipedia.org/wiki/NumPy). It provides an [object-oriented](https://en.wikipedia.org/wiki/Object-oriented_programming) [API](https://en.wikipedia.org/wiki/API) for embedding plots into applications using general-purpose [GUI toolkits](https://en.wikipedia.org/wiki/GUI_toolkit) like [Tkinter](https://en.wikipedia.org/wiki/Tkinter), [wx Python](https://en.wikipedia.org/wiki/WxPython), [Qt](https://en.wikipedia.org/wiki/Qt_(software)), or [GTK+](https://en.wikipedia.org/wiki/GTK%2B). There is also a [procedural](https://en.wikipedia.org/wiki/Procedural_programming) "pylab" interface based on a [state machine](https://en.wikipedia.org/wiki/State_machine) (like [OpenGL](https://en.wikipedia.org/wiki/OpenGL)), designed to closely resemble that of [MATLAB](https://en.wikipedia.org/wiki/MATLAB), though its use is discouraged, [Scipy](https://en.wikipedia.org/wiki/SciPy) makes use of Matplotlib.

Pyplot is a Matplotlib module which provides a MATLAB-like interface.Matplotlib is designed to be as usable as MATLAB, with the ability to use Python, and the advantage of being free and open-source.

**SCATTER PLOT**

HISTOGRAM

**K-MEANS CLUSTERING AND ALGORITHM**

**3-D PLOT**

We are given a data set of items, with certain features, and values for these features (like a vector). The task is to categorize those items into groups. To achieve this, we will use the k-Means algorithm; an unsupervised learning algorithm.

**Overview**

(It will help if you think of items as points in an n-dimensional space).  The algorithm will categorize the items into k groups of similarity. To calculate that similarity, we will use the Euclidean distance as measurement.

The algorithm works as follows:

1. First we initialize k points, called means, randomly.
2. We categorize each item to its closest mean and we update the mean’s coordinates, which are the averages of the items categorized in that mean so far.
3. We repeat the process for a given number of iterations and at the end, we have our clusters.

The points mentioned above are called means, because they hold the mean values of the items categorized in it. To initialize these means, we have a lot of options. An intuitive method is to initialize the means at random items in the data set. Another method is to initialize the means at random values between the boundaries of the data set.

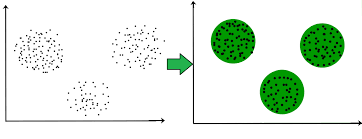


FIG . CLUSTERING IN MACHINE LAERNING