

# Cluster Analysis User Guide

Cluster Analysis is a QGIS 3 plugin to perform clustering based on numerical values on vector layers with any geometry type. The algorithms currently available are Agglomerative Hierarchical and K-Means from the library scikit-learn. Besides, the plugin provides functionalities for dimensionality reduction (highly correlated features, constant and quasi constant features), feature selection, data scaling, and clustering evaluation. For additional information, refer to the brief guide in each section of the UI, and to the file "Plugin implementation" (in particular Chapter 5) in the plugin repository (<https://github.com/folini96/Cluster-Analysis-plugin/tree/main>) for a detailed description of each functionality.

## Installation

The plugin is not officially published in the QGIS folder yet, for this reason the installation must be performed manually.

### Install dependencies

Install the libraries not available in QGIS or OSGeo4w Python (scipy, pandas, scikit-learn). To install the libraries on Windows follow these steps in the same order:

Open OSGeo4w Shell installed with QGIS as Administrator and type:

```
$ py3_env
$ python -m pip install scipy
$ python -m pip install pandas
$ python -m pip install scikit-learn
```

In Linux or Mac OS X, QGIS uses the standard Python installation (independent of QGIS), so you can install the missing libraries using the Terminal (<https://packaging.python.org/tutorials/installing-packages/>).

### Add the plugin to QGIS

To manually install the plugin in QGIS you can perform the following steps:

- Download the entire plugin repository from:  
<https://github.com/folini96/Cluster-Analysis-plugin/tree/main>
- Open QGIS 3 and go to Settings → User Profiles → Open Active User Folder
- In the User Folder go to python → plugins
- Add the plugin to this folder
- In QGIS 3, go to Plugins → Manage and Install Plugins, and make sure that Cluster Analysis is checked in the Installed tab

A new icon for Cluster Analysis will be added to the Plugins menu and on the QGIS main panel.

### Save and load experiments

In the Results section of the plugin, it is possible to save the experiments of the current session in a text file, and to load older ones. The saved files are stored in the Experiments folder inside the plugin folder. To access the plugin folder from QGIS go to Settings → User Profiles → Open Active User Folder → python → plugins. The load function opens by default the Experiments folder, but it is possible to open files stored in other locations. It is not recommended to modify a file and load it in the plugin.

## Configuration file

In the plugin folder you can find a json file called Configuration. From this file it is possible to modify some settings of the algorithms in the plugin. To effectively change the settings, after updating the file you have to reload the plugin either with Plugin Reloader or by restarting QGIS. The parameters are:

- *frequency\_cut*: the threshold for the ratio of the most common value to the second most common value, used in the quasi-constant feature elimination. Requires a numerical value greater than 0;
- *unique\_cut*: the threshold for the ratio of distinct values to the number of total samples, used in the quasi-constant feature elimination. Requires a percentage value expressed as a number between 0 and 1;
- *entropy\_iterations*: the number of random samples used for the sampling entropy algorithm. Requires a positive integer, it is suggested to use a number greater than 35. The execution of sampling entropy takes longer as this number grows;
- *sample\_size*: the number of points in every random sample for the sampling entropy algorithm; Requires a positive integer, the quality of the selection gets lower the smaller the sample size is, while the execution takes longer as this number grows;
- *graph\_max\_cluster*: the max number of clusters used when plotting WSS and BSS trends. Requires a positive integer;
- *distance*: distance measure used in hierarchical clustering, the accepted values are ‘euclidean’ or ‘manhattan’.