### kmeans

Classifies the input record into k clusters based on Euclidean distance.

#### Syntax

kmeans [OPTIONS] FIELD, ...

Required Parameter

**FIELD, ...**

Name of the fields to be calculated, separated by a comma (,).The field value must be numeric, and any input record whose specified field value is not numeric is ignored. Up to 100,000 input records are allowed. The command classifies the records into the N number of clusters (N starting from 1) and assigns them to the **\_cluster** field. If there are more than 100,000 valid input records, it ignores records after 100,000.

Optional Parameter

**k=INT**

Number of clusters (default: 3)

**iter=INT**

Number of times to repeat kmeans (default: 100,000)

#### Usage

You can test the operation method of the kmeans command with iris data, which is often quoted in machine learning. Run the classification using length and width and compare it to the name of the actual species (download: <https://github.com/illinois-cse/data-fa14/blob/gh-pages/data/iris.csv>).

csvfile /opt/logpresso/iris.csv| eval sepal\_length = double(sepal\_length), sepal\_width = double(sepal\_width)| kmeans k=4 iter=100000 sepal\_length, sepal\_width