### lof

最も近いk個の隣接点を基準に各点の密度を計算し、隣接点との相対的な密度比率からLOF（Local Outlier Factor）指数を算出します。

#### 構文

lof [k=INT] FIELD, ... [by GRP\_FIELD, ...]

必須パラメータ

**FIELD, ...**

整数、実数、日付などの数値で構成される値を含むフィールドリスト。区切り文字はカンマ（,）です。

オプションパラメータ

**eps=DOUBLE**

データ間の最小距離調整係数（デフォルト: 0.00001）。データ間距離の合計が無限大に発散しないように調整するために使用します。

**k=INT**

計算に使用する隣接ノード数（デフォルト: 10）。

**[by GRP\_FIELD\_1, ...]**

LOF指数の計算に使用するグルーピングフィールドリスト。区切り文字はカンマ（,）です。このオプションはFIELD, ...の後に指定してください。

by句を利用してグループごとにスコアリングを行う場合、各グループのレコード数は隣接ノード数（k=INTで指定した値）より多くなければなりません。隣接ノード数よりグループのレコード数が少ない場合、すべての点が1つのクラスタとして扱われるため、LOF指数（**\_lof**フィールド）が意図通りに計算されません。

#### 説明

各レコードごとに\_lofフィールドにLOF指数を計算し、この値は以下のように分類できます：

* 値が1より大きい場合（LOF(k) > 1）：クラスタの外側に位置します。1より大きいほど異常値（anomaly）である可能性が高くなります。
* 値が1に近い場合（LOF(k) ≈ 1）：クラスタの境界に位置します。
* 値が1より小さい場合（LOF(k) < 1）：クラスタの内部に位置します。

#### 使用例

**sepal\_length**および**sepal\_width**フィールドの値を基準に異常値を算出します（ダウンロード: <https://raw.githubusercontent.com/illinois-cse/data-fa14/gh-pages/data/iris.csv>）。

wget url="https://raw.githubusercontent.com/illinois-cse/data-fa14/gh-pages/data/iris.csv" | eval line = split(line, "\n") | explode line | split sep="," sepal\_length,sepal\_width,petal\_length,petal\_width,species| eval sepal\_length = double(sepal\_length), sepal\_width = double(sepal\_width)| lof sepal\_length, sepal\_width| search \_lof > 2