

SCRIPT CLUSTER

Dataset su cui si lavora: **Top 20 products**

Algoritmo che si usa: **k-means, con $k = 2$**

Leader / Follower (usiamo $k = 2$ perché vogliamo identificare questo comportamento)

Dati aggregati (somma delle vendite in volume e in valore) **rispetto al tempo anno per anno**

Ci serve per vedere come si comportano i leader anno per anno, chi rimane sempre e chi invece entra/esce dai leader. Rimangono sempre Moretti 66 Cl, Ichnusa non filtrata 50 Cl e Heineken 66 Cl (tranne l'ultimo anno, che però si vede che è lì al limite). Il primo anno viene inclusa anche la Peroni 66 Cl, rimane lì vicino ai leader durante il secondo anno (ma è già clusterizzata come follower) e cala drasticamente dal terzo anno, in particolare nell'ultimo anno ha vendite bassissime (outliers)

Dati rispetto alla serie temporale

1. Si usa la serie temporale così com'è, niente aggregazione
2. Cluster con funzione tscust, specifica delle time series, applicando k-means con distanze euclidee
3. Plot in scala semilogaritmica della time series, con i prodotti colorati in base al cluster: la Peroni 66 Cl viene classificata come leader, ma vediamo che le sue vendite diminuiscono vertiginosamente nell'ultimo periodo (outliers)

Dati aggregati (somma delle vendite in volume e in valore) **rispetto al prodotto su tutti gli anni**

Leader: Moretti 66 Cl, Heineken 66 Cl, Ichnusa non filtrata 50 Cl, coerentemente al cluster sulla time series e quello fatto anno per anno. In questo caso la Peroni 66 Cl viene esclusa dai leader, decidiamo di usare questa classificazione di Leader/Follower nella regressione perché ci sembra quella più 'robusta'

Risposta ChatGPT alla domanda “Perché usare la scala logaritmica nelle vendite di una time series?”

È spesso preferibile utilizzare la scala logaritmica quando si lavora con serie temporali di vendite o prezzi per diversi motivi principali:

1. Gestione delle Variazioni Percentuali

Quando si lavora con serie temporali che mostrano variazioni significative nei valori (come vendite o prezzi), le variazioni percentuali sono più rilevanti rispetto alle variazioni assolute. La scala logaritmica tende a comprimere le variazioni assolute più grandi, rendendo più evidenti le variazioni percentuali minori. Questo aiuta a visualizzare meglio le dinamiche relative delle serie temporali.

2. Stabilizzazione della Varianza

Le serie temporali che mostrano varianza crescente nel tempo (eteroschedasticità) possono beneficiare dall'essere trasformate su scala logaritmica. Questa trasformazione può aiutare a stabilizzare la varianza nel tempo, rendendo più uniforme la distribuzione dei dati rispetto alla media. Ciò è particolarmente utile quando si desidera applicare tecniche statistiche che presuppongono omoschedasticità (varianza costante).

3. Interpretazione delle Tendenze

Su una scala lineare, piccoli cambiamenti nei valori possono sembrare insignificanti, mentre su una scala logaritmica, gli stessi cambiamenti possono essere più visibili. Questo può facilitare l'identificazione di tendenze e pattern nascosti nelle serie temporali, consentendo una migliore interpretazione dei dati.

4. Comparazione tra Serie Temporali

La trasformazione logaritmica può rendere più semplice la comparazione tra diverse serie temporali che potrebbero variare notevolmente in termini di ampiezza. Portare i dati su scala logaritmica può aiutare a standardizzare le serie temporali e a rendere più facile rilevare le differenze relative tra di esse.

Considerazioni Pratiche

- **Logaritmo Naturale o Decimale:** La scelta tra logaritmo naturale (\ln) e logaritmo decimale dipende dal contesto e dalle preferenze dell'analista. Entrambi sono comuni, con il logaritmo naturale che è più frequentemente usato in analisi finanziarie e economiche.

- **Interpretazione dei Risultati:** È importante ricordare che la scala logaritmica modifica la percezione dei dati, quindi l'interpretazione dei risultati dovrebbe riflettere questa trasformazione. Ad esempio, un aumento costante su una scala logaritmica corrisponde a un tasso di crescita percentuale costante su una scala lineare.

In conclusione, l'uso della scala logaritmica è consigliato quando si lavora con serie temporali di vendite o prezzi per migliorare la visualizzazione, stabilizzare la varianza, facilitare l'interpretazione delle tendenze e consentire una comparazione più efficace tra le serie temporali.