

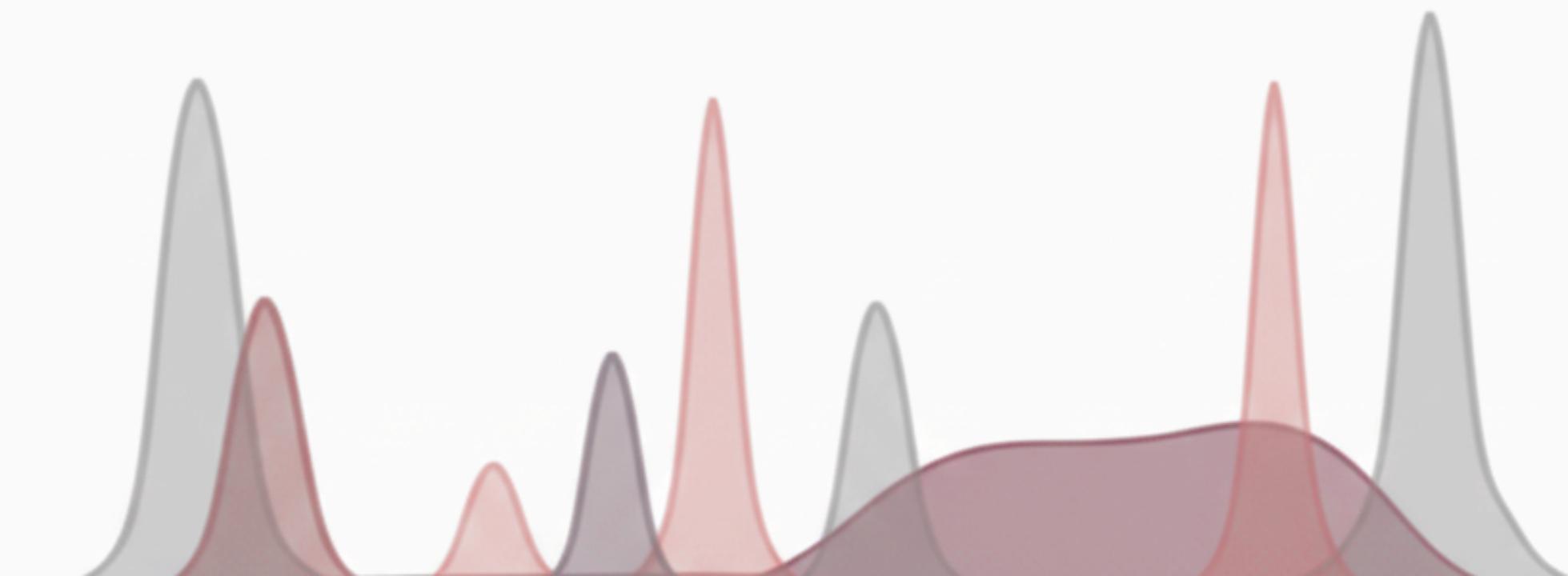


DBSCAN ALGORITHM

Principi e modelli della percezione,

A.A. 2025/2026

Sofia Arbani, Dora Magyari, Miriana Stella Malnati



Cos'è DBSCAN

Density-Based Spatial Clustering of Applications with Noise

- Cluster: regioni ad alta densità
- Rumore: regioni più vuote

Due parametri fondamentali:

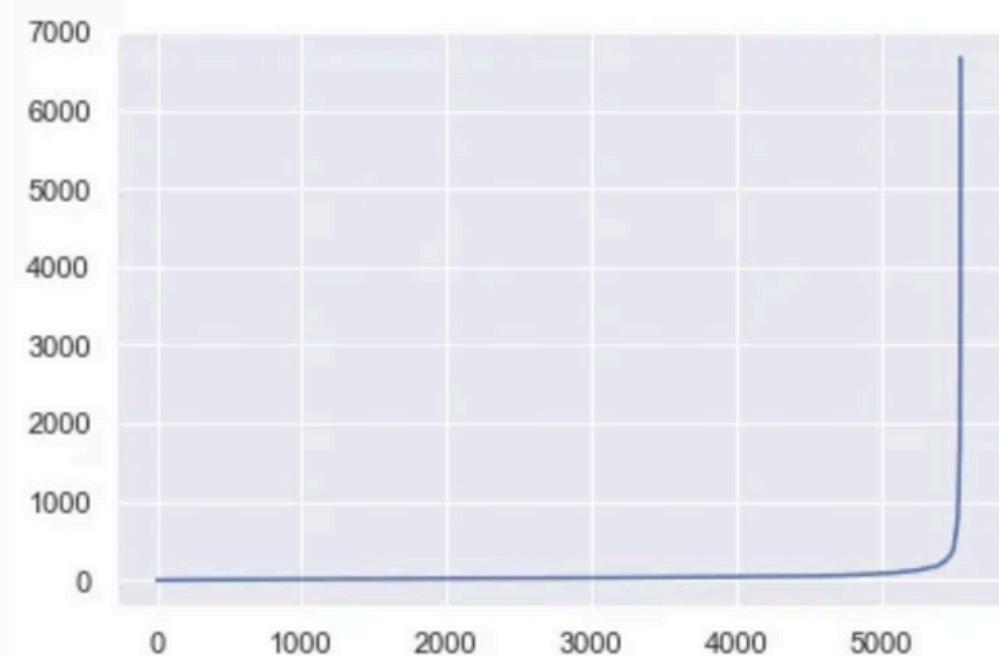
- **minPts (o Min Samples)**

Numero minimo di punti che devono stare dentro il raggio per dire che quella zona è “densa”.

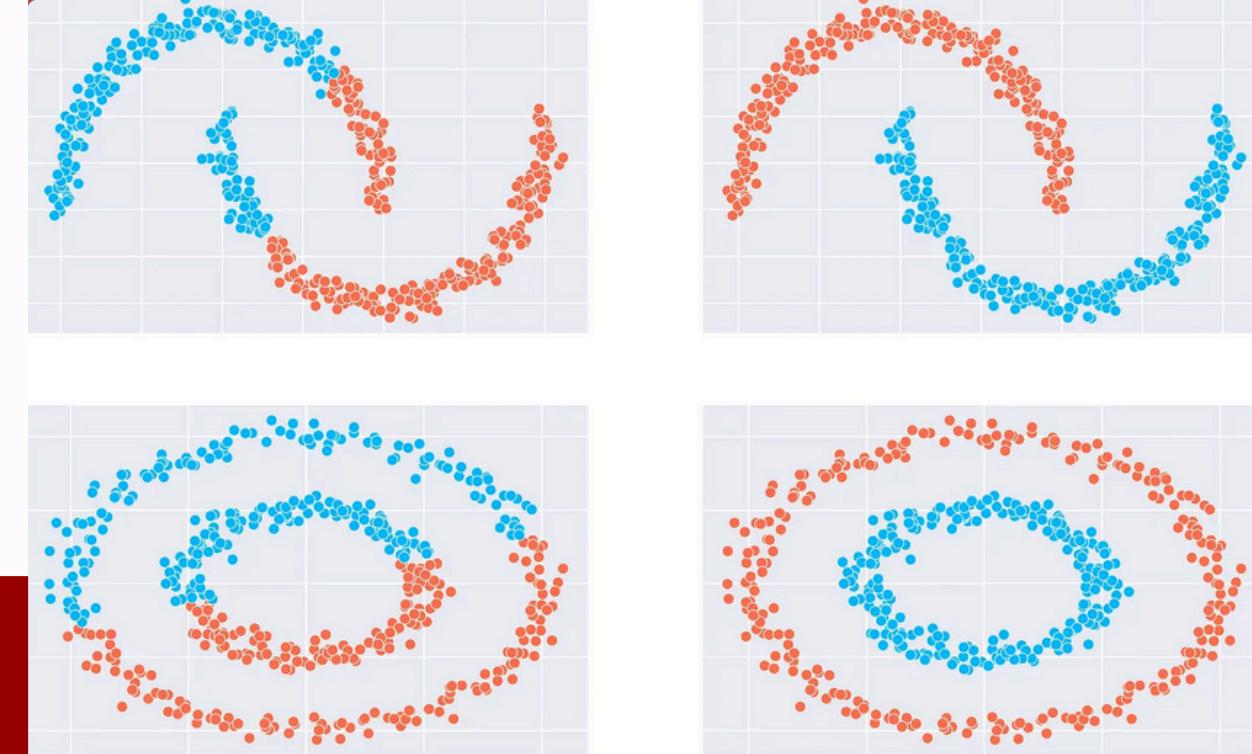
- $\text{minPts} \geq \text{dimensione dello spazio} + 1$

- **ϵ — epsilon**

raggio di ricerca intorno a un punto → “knee finding” o “elbow finding”



KMEANS	DBSCAN
Cluster rotondi / sferici basati sulla distanza dal centroide	Cluster di forme qualsiasi anche irregolari o “a macchia”
Bisogna scegliere K prima	Non richiede K, lo trova da solo
Soffre molto gli outlier	Identifica rumore (punti -1) in modo naturale

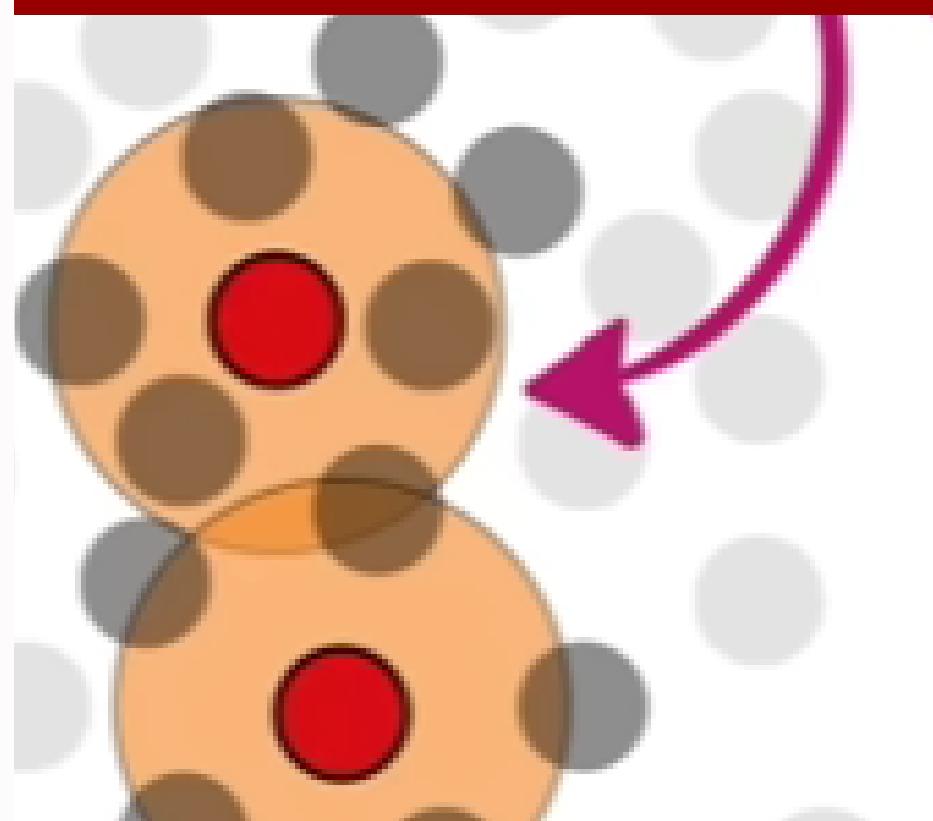


Classificazione dei Punti

Espansione del Cluster

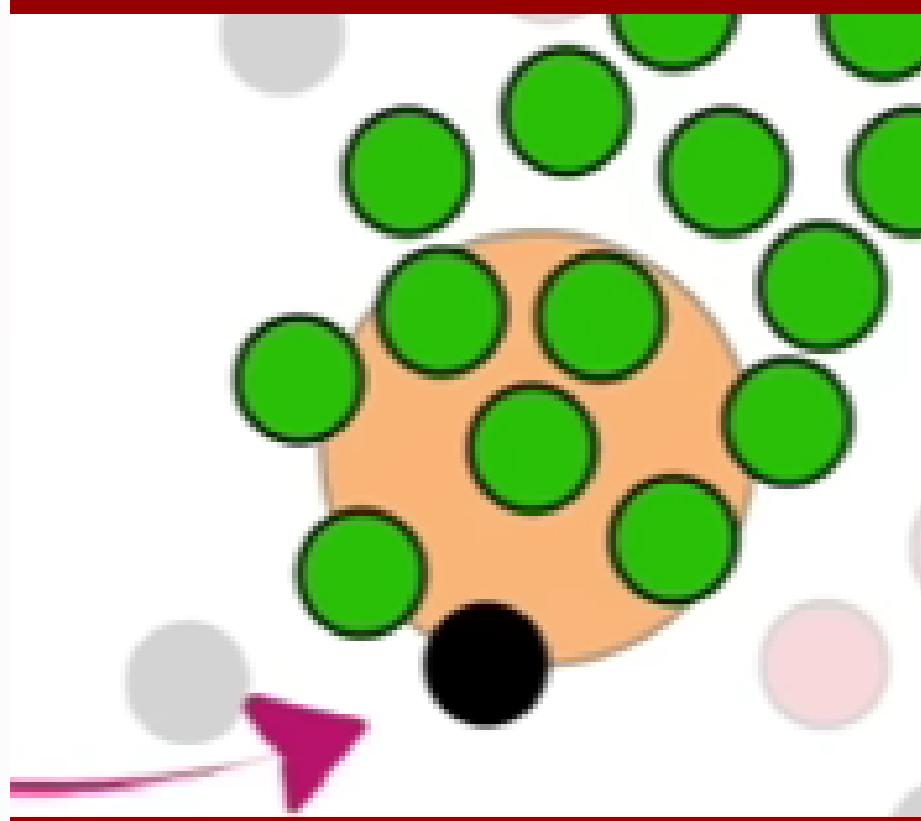
Iterazione

Core Point



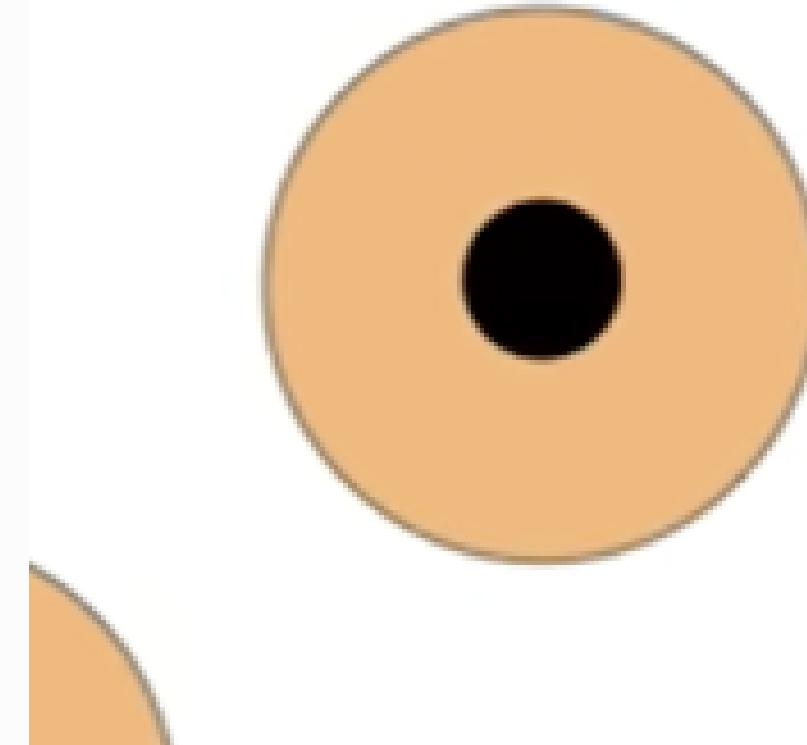
Punti che hanno un numero di vicini entro il raggio epsilon uguale o superiore a minPts - generano e fanno crescere il cluster

Border Point



Punti che si trovano entro il raggio epsilon di un core point, ma che da soli hanno meno di minPts vicini

Noise Point



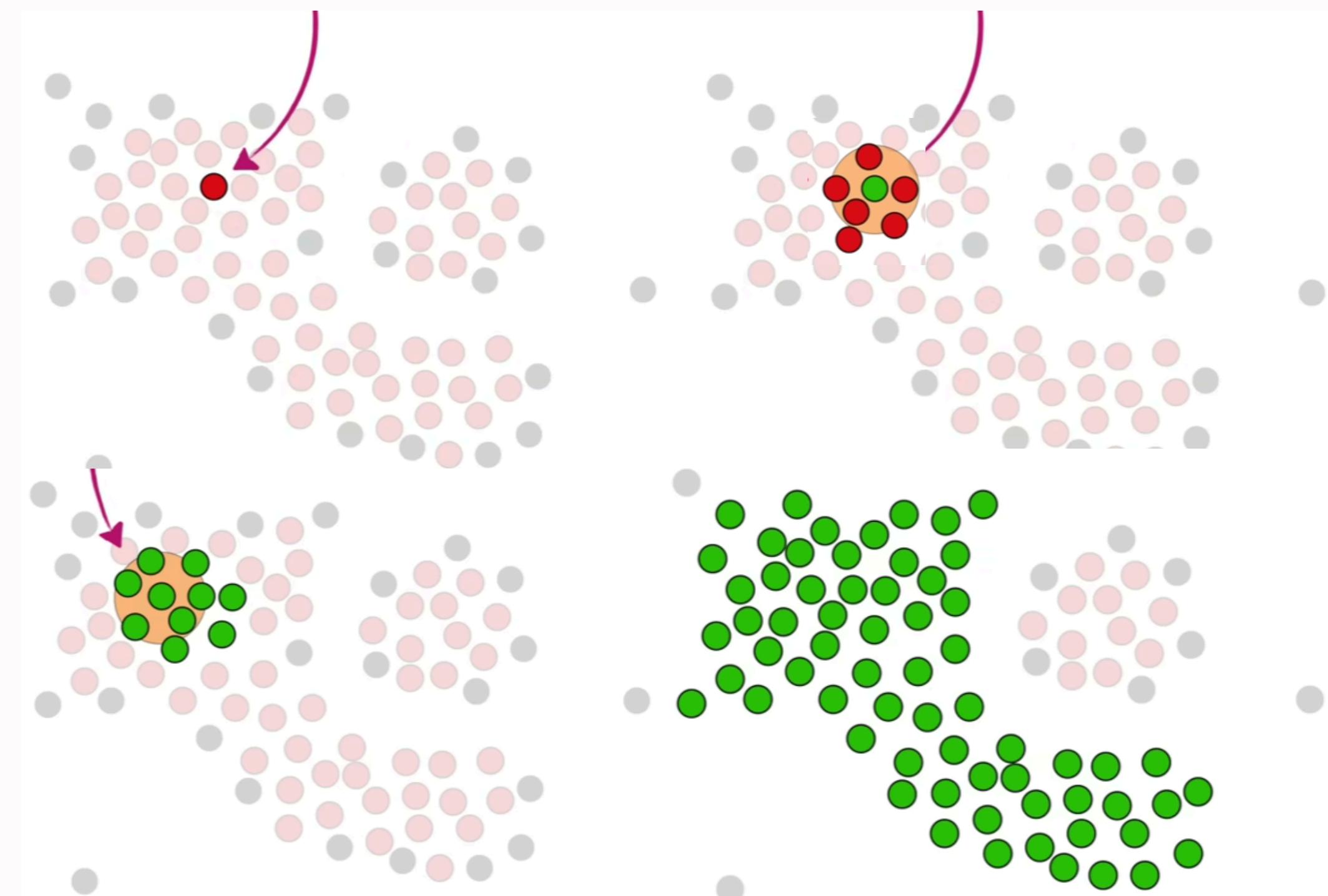
Punti che non rientrano nell'intorno di nessun core point

Classificazione dei Punti

Espansione del Cluster

Iterazione

1. Sceglie a caso uno tra i core points.
2. Si aggiungono al cluster tutti i punti entro eps .
3. Per ogni punto aggiunto:
 - se è un core point \rightarrow il cluster si espande tramite i suoi vicini
 - se non è core \rightarrow entra nel cluster, ma non lo amplia
4. L'espansione del cluster continua finché non ci sono più core point collegati



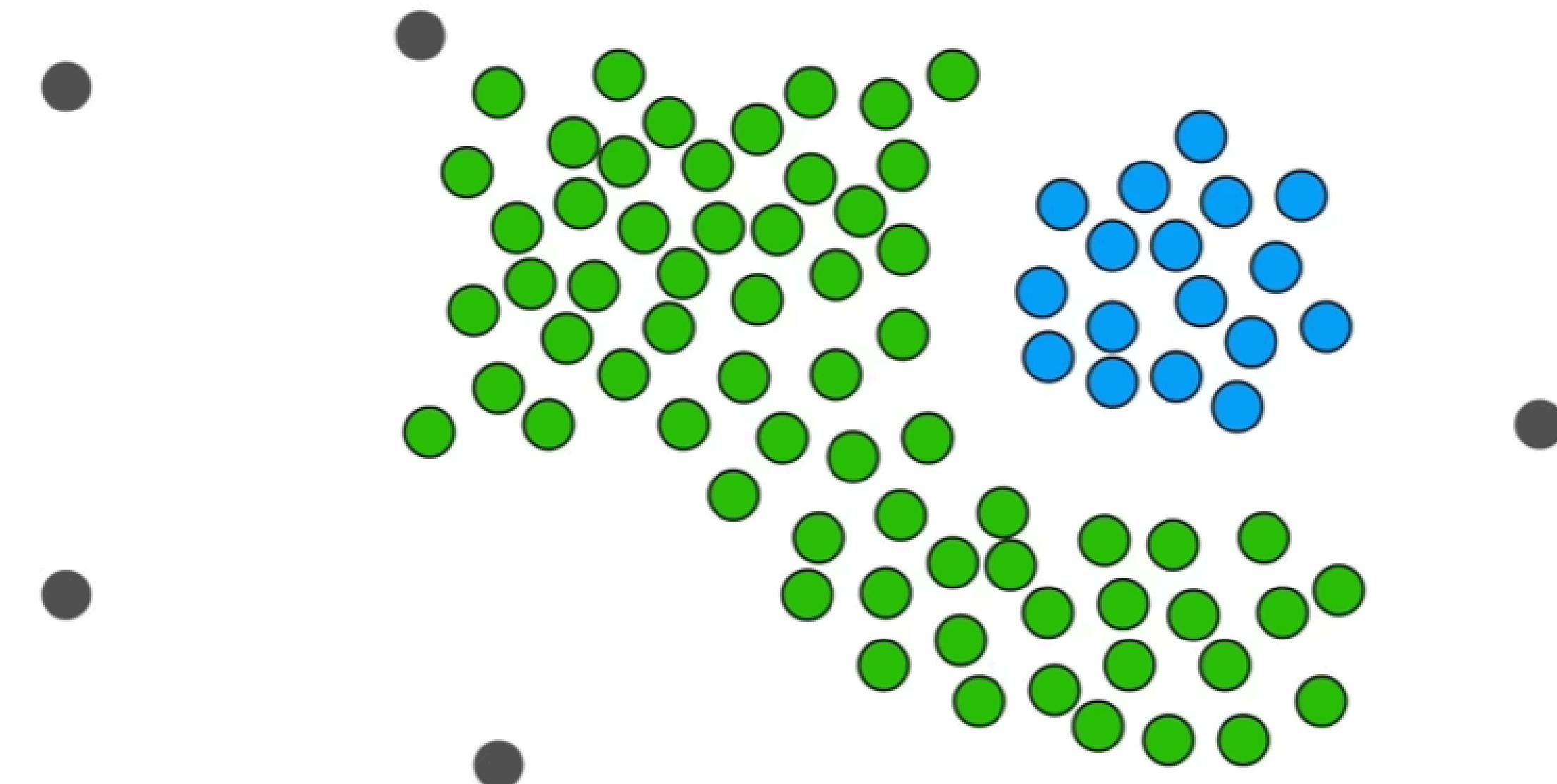


Classificazione dei Punti

Espansione del Cluster

Iterazione

Il procedimento viene ripetuto finché non ci sono più core point da espandere, individuando così tutti i cluster.
I punti che non vengono assegnati a nessun cluster vengono automaticamente classificati come outlier



IL PROGETTO



Dataset scelto

Quantità settimanali acquistate di 800 prodotti nell'arco di 52 settimane.

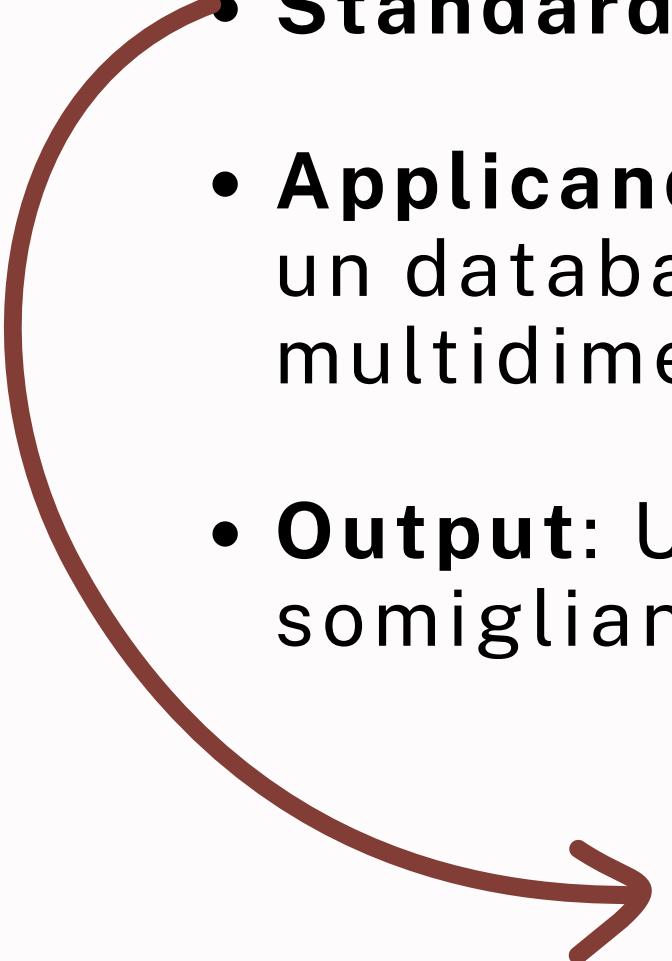


L'obiettivo

Identificare gruppi di prodotti che condividono lo stesso pattern di vendita settimanale per ottimizzare le strategie di riassortimento e identificare gli outlier.

Comprendere il dataset e Standardizzazione

- **Input:** 800 prodotti (righe) monitorati per 52 settimane (colonne).
- **Standardizzazione**
- **Applicando l'algoritmo:** Ogni prodotto cessa di essere una riga in un database e diventa un punto preciso in uno spazio multidimensionale a 52 dimensioni.
- **Output:** Una mappa dove la distanza tra due punti riflette la somiglianza nell'andamento temporale delle vendite.



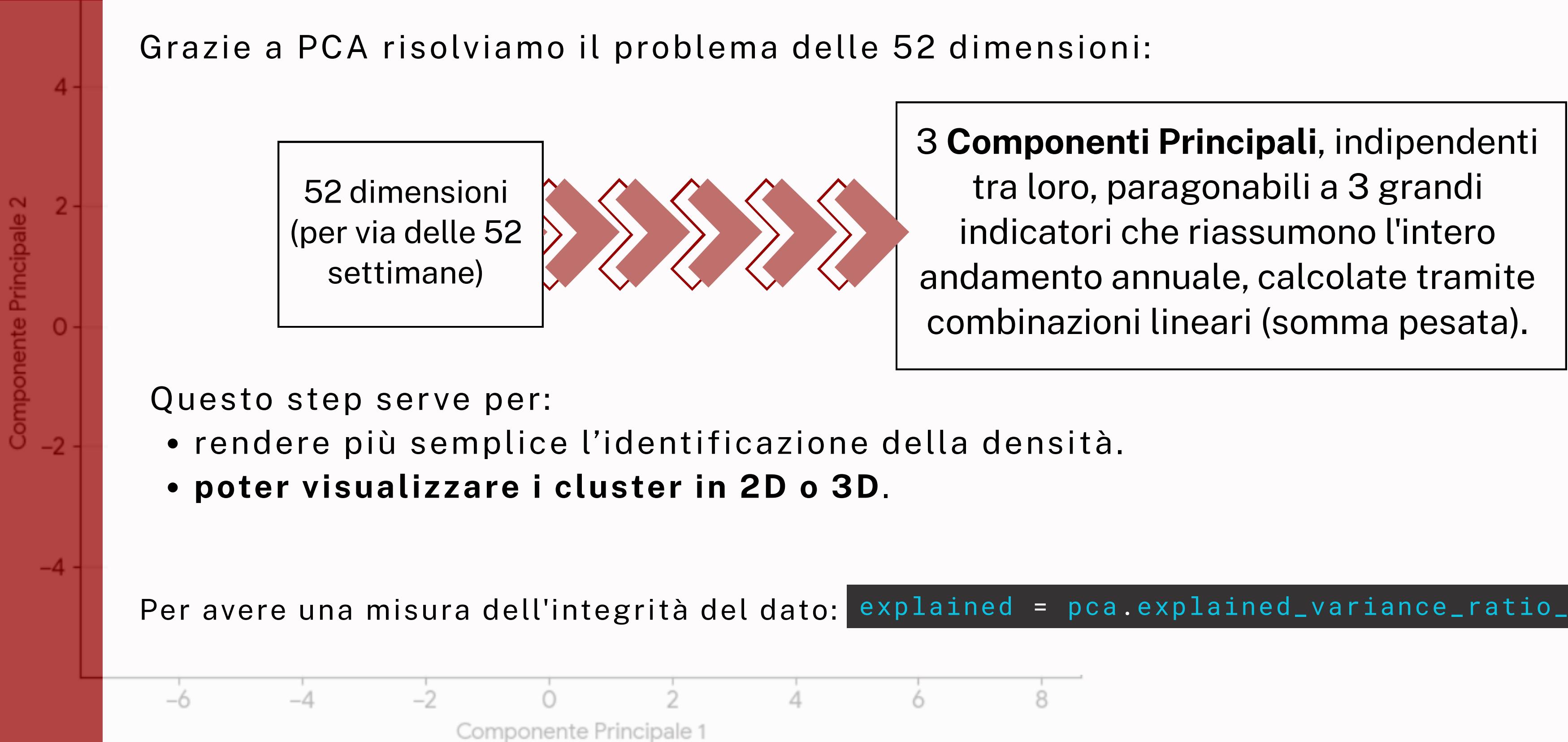
```
scaler = StandardScaler()
```

porta tutte le settimane sulla **stessa scala metrica**, permettendo a ogni settimana di avere lo stesso peso matematico.

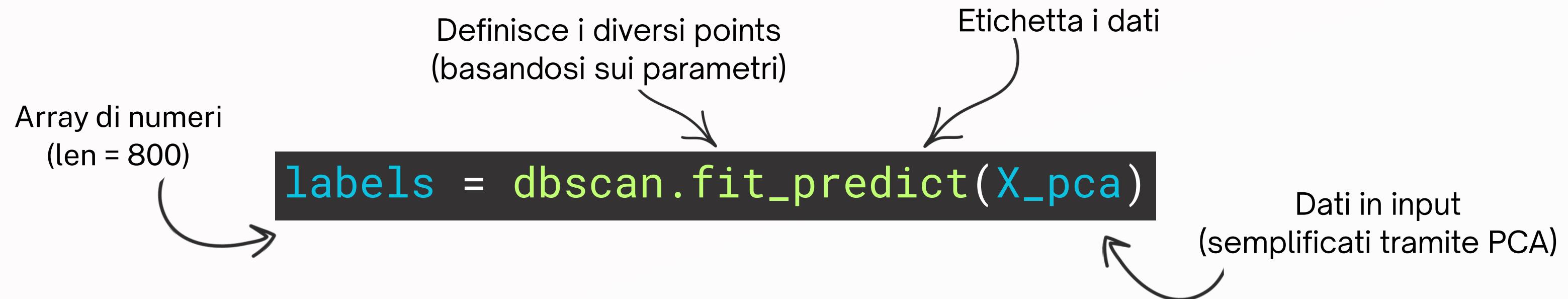
PCA - Principal Component Analysis

```
pca = PCA(n_components=3)
```

Grazie a PCA risolviamo il problema delle 52 dimensioni:



L'applicazione del DBSCAN



Da questo momento, ogni prodotto porta con sé una etichetta che lo identifica. Così facendo abbiamo ottenuto una mappatura completa che distingue automaticamente i cluster dal rumore.

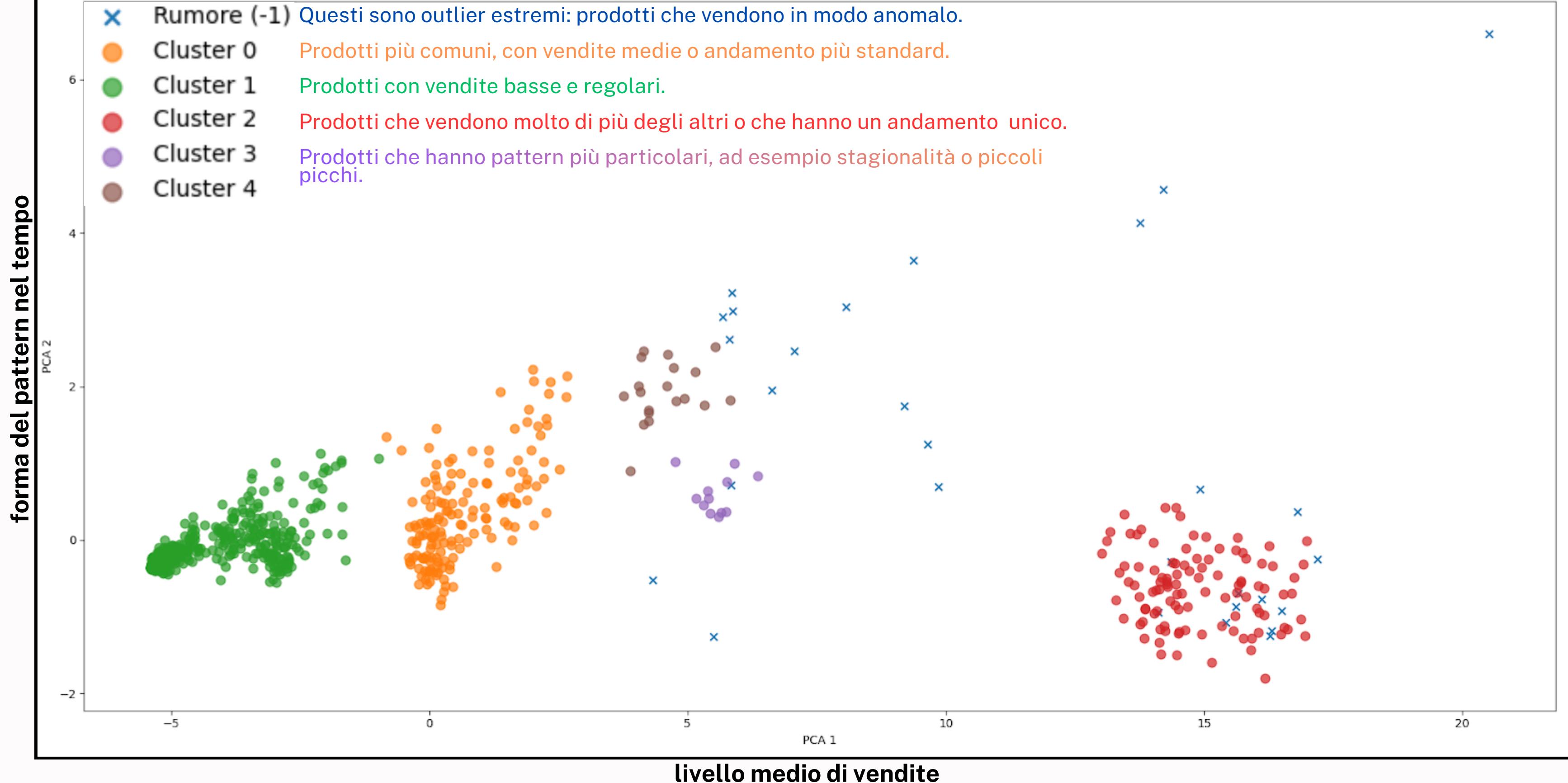
Esempio: [0, 0, 1, -1, 0, 2, -1, ...]

- Il primo prodotto appartiene al **Cluster 0**.
- Il terzo prodotto appartiene al **Cluster 1**.
- Il quarto prodotto (**-1**) è **Rumore/Outlier**.

SCATTER PLOT

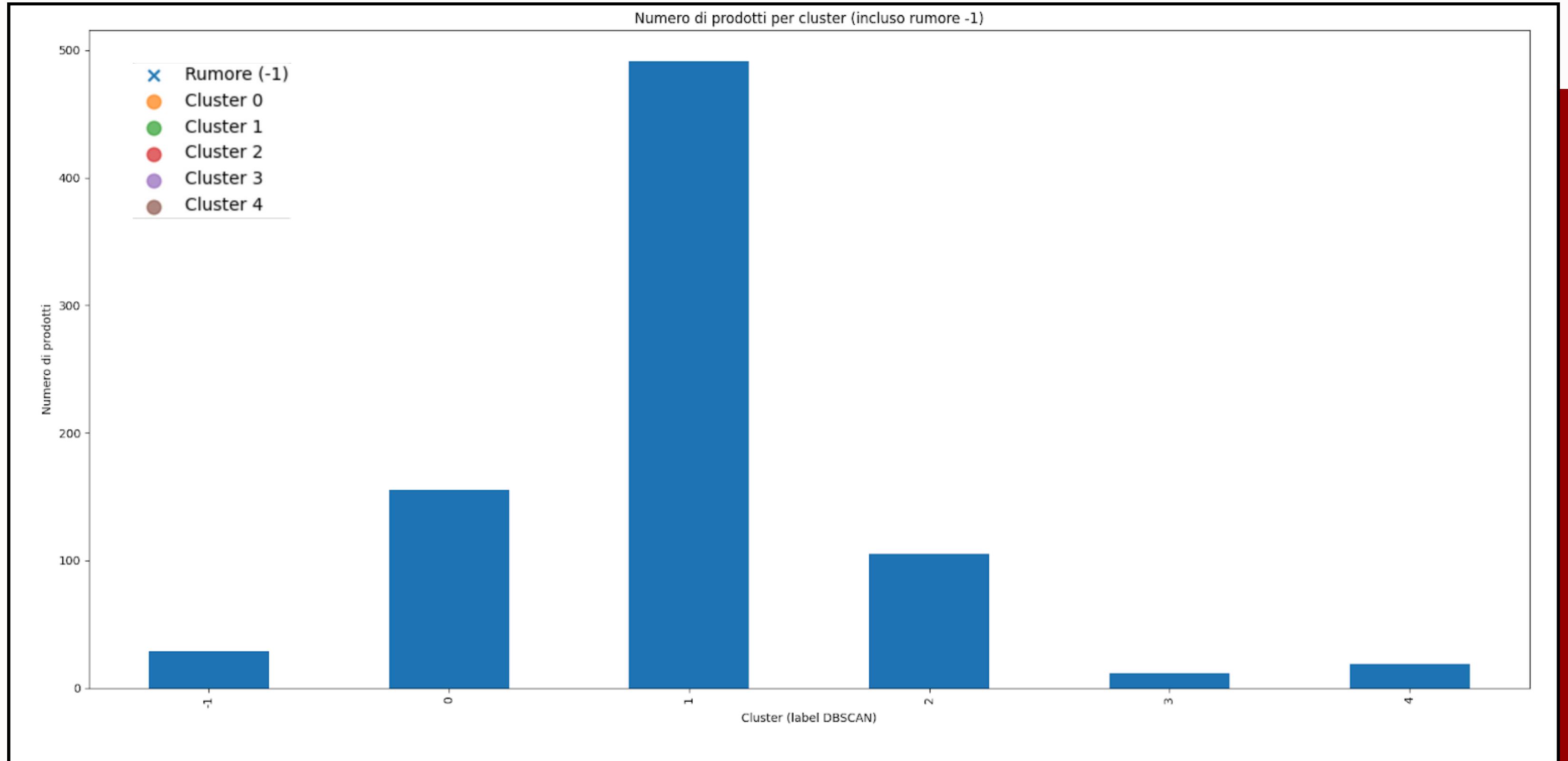
Comportamento di vendita simile durante l'anno

DBSCAN sui prodotti (prime 2 componenti PCA)



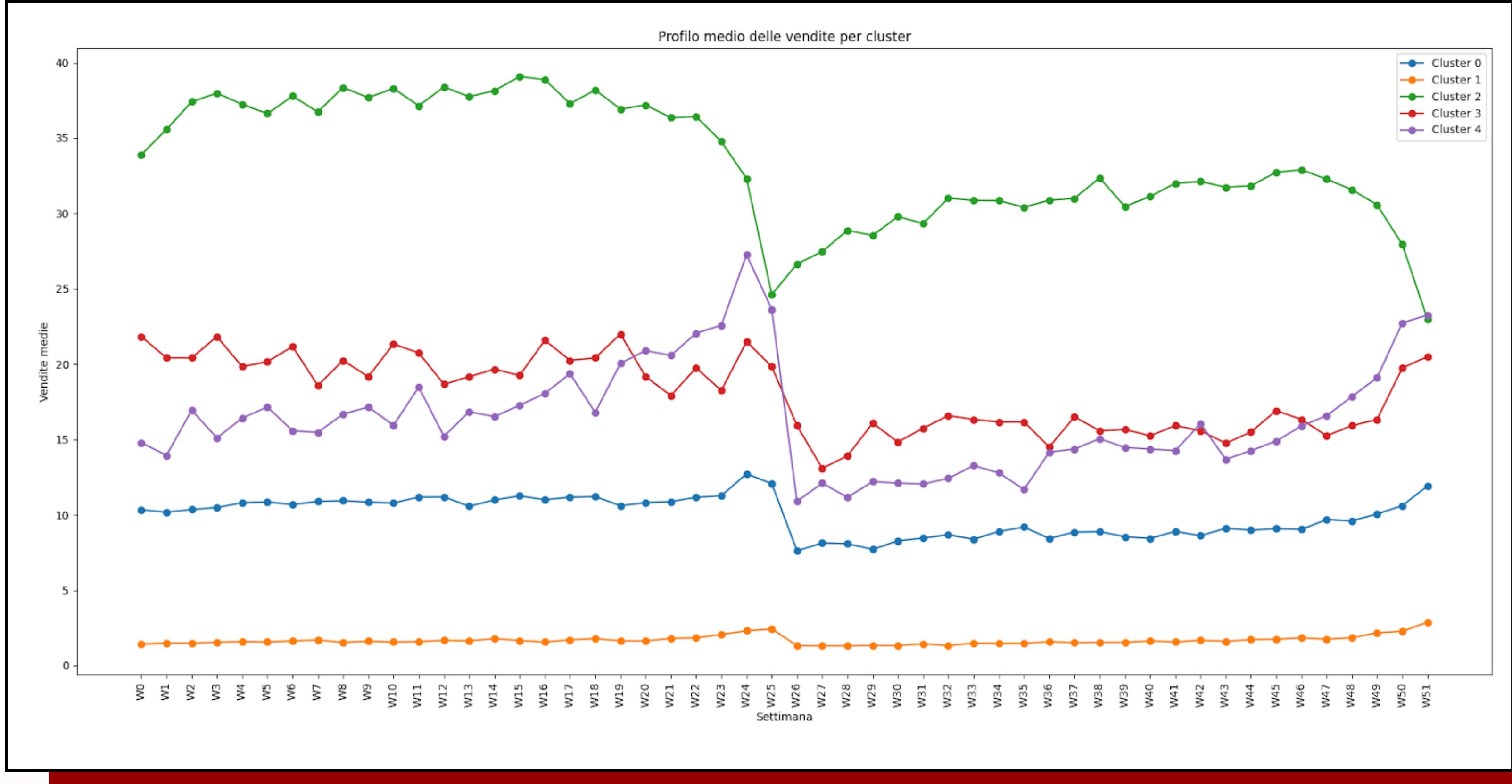
ISTOGRAMMA

Peso di ciascun cluster



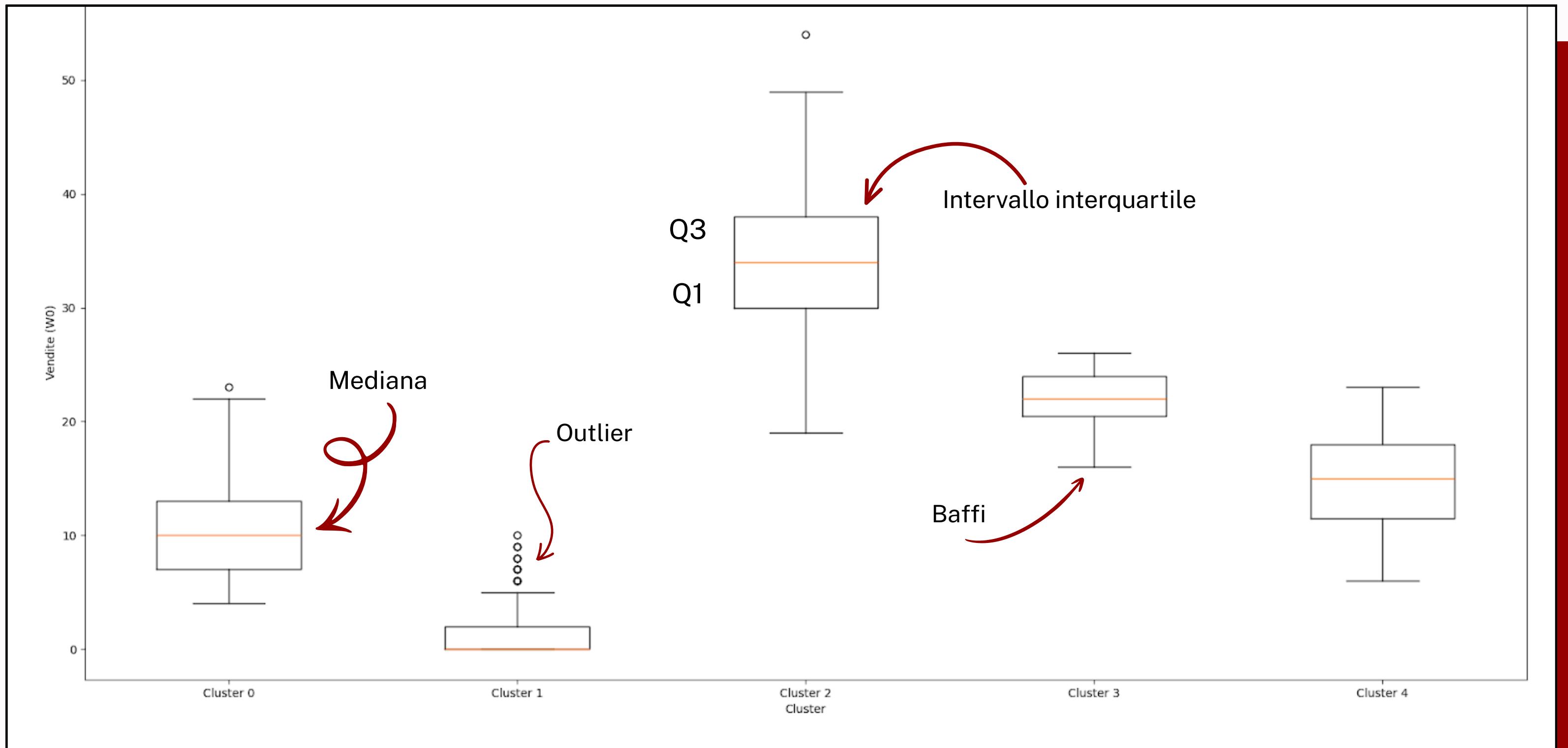
LINEPLOT

Come vendono i prodotti durante l'anno:



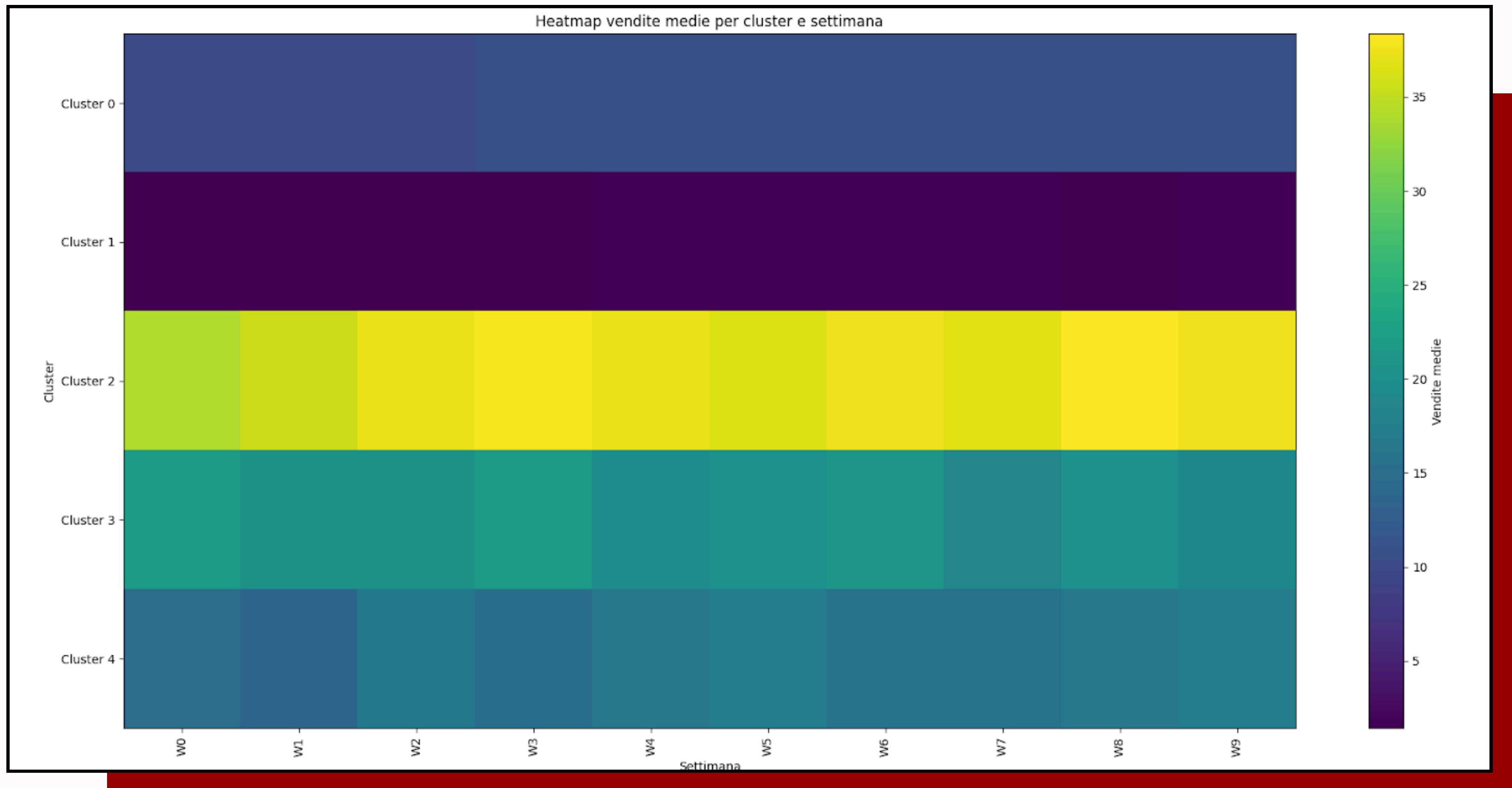
BOXPLOT

Distribuzione delle vendite per cluster nella settimana W0



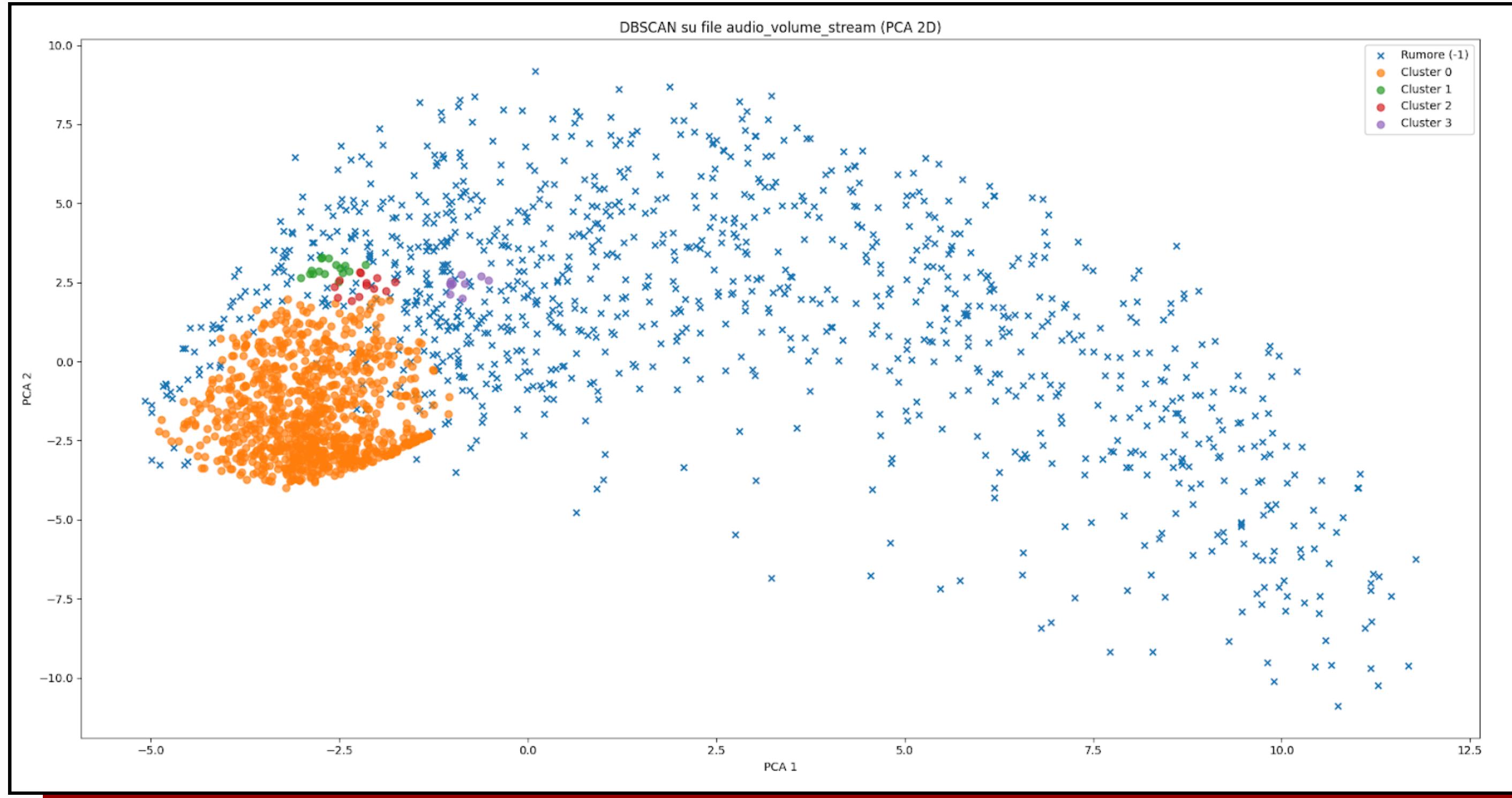
HEATMAP

Vendite medie per cluster e settimana



SVANTAGGI

- L'algoritmo non riesce più a distinguere zone dense da zone vuote.
- Non trova gruppi reali.
- Finisce per classificare praticamente tutto come rumore.





Grazie per
l'attenzione