**📌 Why is clustering useful?**

✅ **Clustering is most commonly used for:**

* **Exploratory analysis** — getting to know the data, spotting patterns, without testing a specific hypothesis.

✅ Less often, it can also help:

* Confirm prior beliefs (but classification is usually better for this).
* Show that known clusters have changed over time — signaling market or behavior shifts.

**📌 Types of analysis:**

1️⃣ **Exploratory analysis**

* First step: familiarize with data & patterns.
* Tools: visualization, descriptive stats, clustering.
* No hypothesis tested yet.

2️⃣ **Confirmatory / Explanatory analysis**

* Later stage: test hypotheses or explain phenomena.
* Tools: hypothesis tests, regression.

**📌 Clustering vs Classification:**

* **Classification** → you already know the groups, you just want to assign observations to them (supervised).
* **Clustering** → you discover the groups from scratch (unsupervised).

**📌 Real-life application: 4 clusters example**

We have 4 customer segments:

* 🟣 **Fans** → satisfied & loyal → keep them happy!
* 🟡 **Supporters** → loyal but not fully satisfied → improve satisfaction (e.g., shorter queues).
* 🔴 **Roamers** → satisfied but not loyal → improve loyalty (e.g., loyalty programs, discounts).
* 🔷 **Alienated** → neither satisfied nor loyal → least priority, harder to recover.

**📌 Strategies based on clusters:**

✅ Supporters: improve satisfaction so they become fans.  
✅ Roamers: improve loyalty through loyalty programs.  
✅ Alienated: low priority, as they’re disengaged and hard to win back.  
✅ Fans: maintain their high satisfaction and loyalty.

**📌 Advanced strategy:**

* Use fans’ demographic & behavioral data to find **similar people** and target them with ads → common in digital marketing by companies like Google, Amazon, Facebook.

**🔷 Summary:**

✨ Clustering helps:

* Understand customer segments.
* Prioritize actions based on group needs.
* Detect market changes.
* Target ads effectively by finding lookalikes.

**📚 Lesson: Clustering of Clustering**

Now that you already understand the basics of **clustering** — grouping similar data points — let’s talk about a more advanced topic:

**🔷 What is “Clustering of Clustering”?**

Clustering as a concept was originally invented by **anthropologists** to study the origins of humans.  
Later, it was applied in psychology, intelligence research, biology, and many other fields.

Nowadays, we distinguish between two broad approaches:

* **Flat Clustering**
* **Hierarchical Clustering**

**🟥 1️⃣ Flat Clustering**

Flat clustering methods (like **K-Means**) simply create a flat set of clusters:  
✅ You pick the number of clusters kkk,  
✅ Run the algorithm,  
✅ You get kkk groups.

No *hierarchy* or levels — just kkk clusters.

**🟥 2️⃣ Hierarchical Clustering**

Hierarchical clustering produces a **tree-like structure** of clusters:  
🔷 Big clusters are broken down into smaller sub-clusters (or vice versa),  
🔷 All clusters are *nested* within each other.

A classic example is **biological taxonomy:**

* Animal
  + Mammal
    - Dog
      * Labrador
    - Cat
  + Bird
    - Flying birds
    - Flightless birds

This tree of relationships is called a **Hierarchy of Clusters**.

**🟥 Two Types of Hierarchical Clustering:**

**1️⃣ Divisive (Top-Down)**

* Start with **all data points in one big cluster** (like all dinosaurs).
* Then split it into two clusters.
* Then split those into smaller ones.
* Continue splitting until each point is its own cluster.

⚠️ *Drawback*: At each step you must check many possible splits — computationally expensive!

**2️⃣ Agglomerative (Bottom-Up)**

* Start with **each point as its own cluster**.
* Merge the two closest clusters at each step.
* Continue merging until you end up with one big cluster.

✅ Easier and more efficient than divisive.

**🟥 How does this relate to K-Means?**

K-Means can be thought of as a *flat version of a divisive idea*:

* You start with k=1k=1k=1 and increase kkk, splitting into more clusters,
* Use the **Elbow Method** to decide where to stop.

So it mimics *divisive clustering* but faster and more practical.

**🟥 Agglomerative Hierarchical Clustering: Step-by-Step**

✅ Start: Each observation is its own cluster — NNN clusters.  
✅ Measure the similarity (e.g., Euclidean distance) between all clusters.  
✅ Merge the two most similar clusters — now N−1N-1N−1 clusters.  
✅ Repeat: merge closest clusters each time until only 1 cluster remains.

**🟥 Dendrogram: The Hierarchical Tree**

The result of hierarchical clustering is visualized as a **dendrogram**:  
🌳 A tree diagram that shows:

* How clusters merge at each level.
* The nested structure of all possible clusterings.

You can “cut” the dendrogram at any level to choose how many clusters you want.

**🟥 Summary Table:**

| **Feature** | **Flat Clustering** | **Hierarchical Clustering** |
| --- | --- | --- |
| Method | K-Means | Agglomerative / Divisive |
| Output | Fixed kkk clusters | Tree of nested clusters |
| Hierarchy? | ❌ No | ✅ Yes |
| Visualization | Scatter plot | Dendrogram |

**📚 Lesson: Understanding a Dendrogram**

**🔷 What is a Dendrogram?**

A **dendrogram** is a tree-like diagram that shows how clusters are merged step by step in **hierarchical clustering**.  
It helps you:  
✅ Visualize how close or far clusters are from each other.  
✅ Decide how many clusters to choose.

**🟥 The Example: Countries Clustered by Latitude & Longitude**

We look at a dendrogram created for several countries based on their **longitude and latitude (standardized)**.  
⚠️ *Note:* Standardization in this example didn’t change much — sometimes it does, sometimes it doesn’t.

**🟥 Step by Step: How Clusters Merge**

Initially:

* Each country is its own cluster.

**🔷 First merge:**

🇩🇪 **Germany** & 🇫🇷 **France**

* They are the most similar based on the features.
* At this stage: 5 clusters remain.

**🔷 Next merge:**

**Germany-France** cluster joins 🇬🇧 **UK**

* Now: 4 clusters remain.

**🔷 Next merge:**

🇨🇦 **Canada** & 🇺🇸 **USA**

* North America cluster forms.

**🔷 Then:**

The **Europe cluster (Germany-France-UK)** and **North America cluster (Canada-USA)** merge.

* Australia 🇦🇺 is still alone.

**🔷 Finally:**

🇦🇺 Australia joins the rest → all become 1 cluster.

**🟥 What the distances mean:**

* The **height of the link** (vertical distance in the dendrogram) shows how different the clusters are.
* Example:
  + Germany & France merge quickly → they’re very similar.
  + USA & Canada also merge early → similar.
  + Europe & North America take longer to merge → less similar.
  + Australia joins last → very different from all others.

**🟥 How to decide the number of clusters:**

* You “cut” the dendrogram horizontally at some height.
* The number of vertical lines you cut through equals the number of clusters.

**Examples:**

✂️ Cut at a low level → 2 clusters: Australia & Rest.  
✂️ Cut higher → 3 clusters: Europe, North America, Australia.  
✂️ Cut even higher → 4 clusters… and so on.

✅ General advice:  
If the distance (height) between two merges is **big**, it’s often a good place to stop and cut.

**🟥 Why dendrograms are useful:**

✔️ Show all possible ways clusters can merge → better understanding of the data.  
✔️ You don’t need to set kkk (number of clusters) in advance.  
✔️ Different linkage methods are available (e.g., Ward’s method) — you can pick what works best for your data.  
✔️ Great for small datasets and exploratory analysis.

**🟥 But… why not always use them?**

⚠️ **Main drawback: Scalability.**

* With 1,000+ observations:
  + The dendrogram becomes too messy to interpret.
  + The computations become very slow.

Compared to **K-means**:

* K-means is much faster and scales well to large datasets.
* But K-means does not show a hierarchy or allow you to explore different levels of clustering.

**🟥 Pros & Cons:**

| **✅ Pros** | **❌ Cons** |
| --- | --- |
| Visual & intuitive | Doesn’t scale to big data |
| No need to pre-set kkk | Hard to read with many observations |
| Many linkage methods | Computationally expensive |