



# Unsupervised Learning: Clustering methods

Parcours Progis

Etudes, Medias, communication, Marketing

Bahareh Afshinpour

23.02.2026

# Suggested Reading

- Introduction to Machine Learning with Python, A GUIDE FOR DATA SCIENTISTS, Andreas C. Müller & Sarah Guido ,oreilly
- <https://medium.com/@esha.jagatia/k-means-clustering-in-social-media-management-a-statistical-approach-8284427b5ed1>
- <https://www.youtube.com/watch?v=8QCBI-xdeZI>

# Supervised vs unsupervised

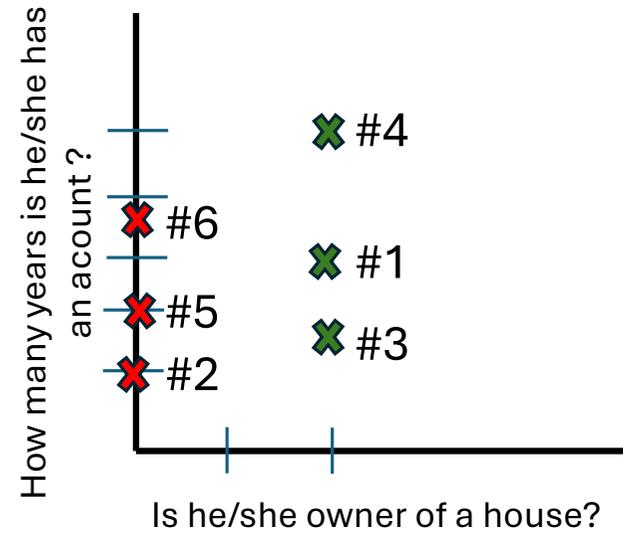
- In **supervised learning**, the aim is to learn a mapping from the input to an output whose correct values are provided by a supervisor.
- In **unsupervised learning**, there is no such supervisor and we **only** have **input data**. The aim is to find the regularities in the input.

Unsupervised Learning involves analysing **unlabeled data** to uncover hidden patterns or structures within data.

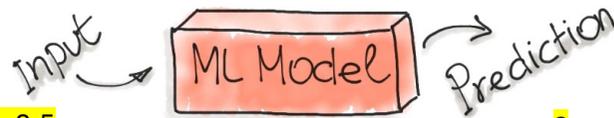
- Unsupervised learning subsumes all kinds of machine learning where there is no known output, no teacher to instruct the learning algorithm.
- In unsupervised learning, the learning algorithm is just shown the input data and asked to extract knowledge from this data.

# Example of a labeled dataset

	Is-owner	How many years	Back-loan
#1	1	3	Yes
#2	0	1	No
#3	1	1.5	Yes
#4	1	5	Yes
#5	0	2	No
#6	0	3.5	No



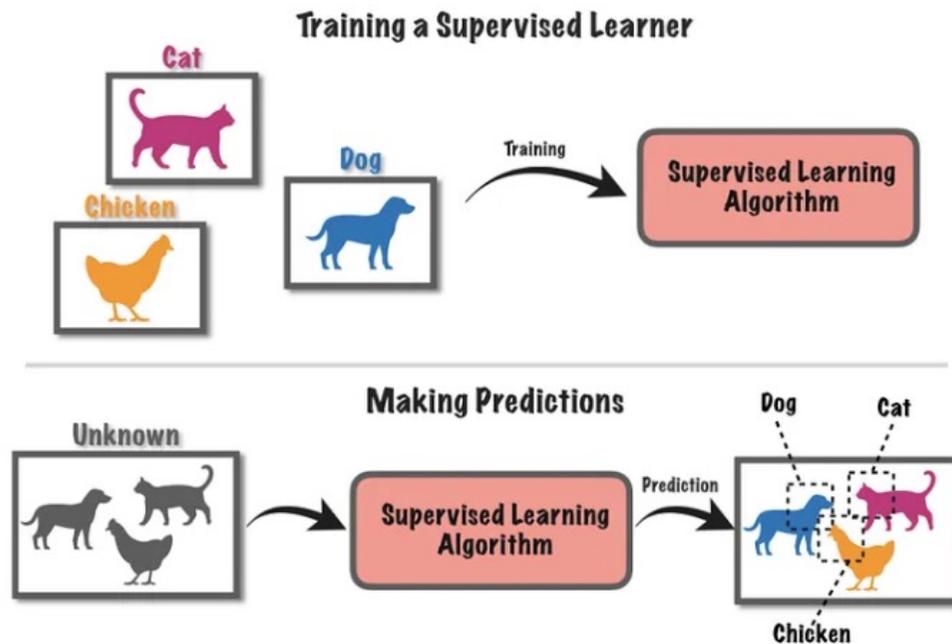
New client Is-owner=1 how many years=2.5 ?



New client Is-owner=1 how many years=2.5

?

# Supervised learning



<https://kasunprageethdissanayake.medium.com/artificial-intelligence-2-supervised-learning-unsupervised-learning-and-reinforcement-learning-7bf00c732e99>

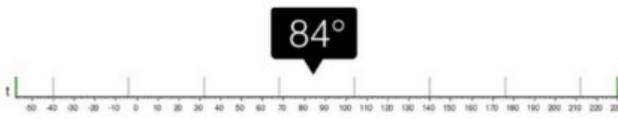
# Supervised Learning



## Regression



What will be the temperature tomorrow?



Fahrenheit

## Classification



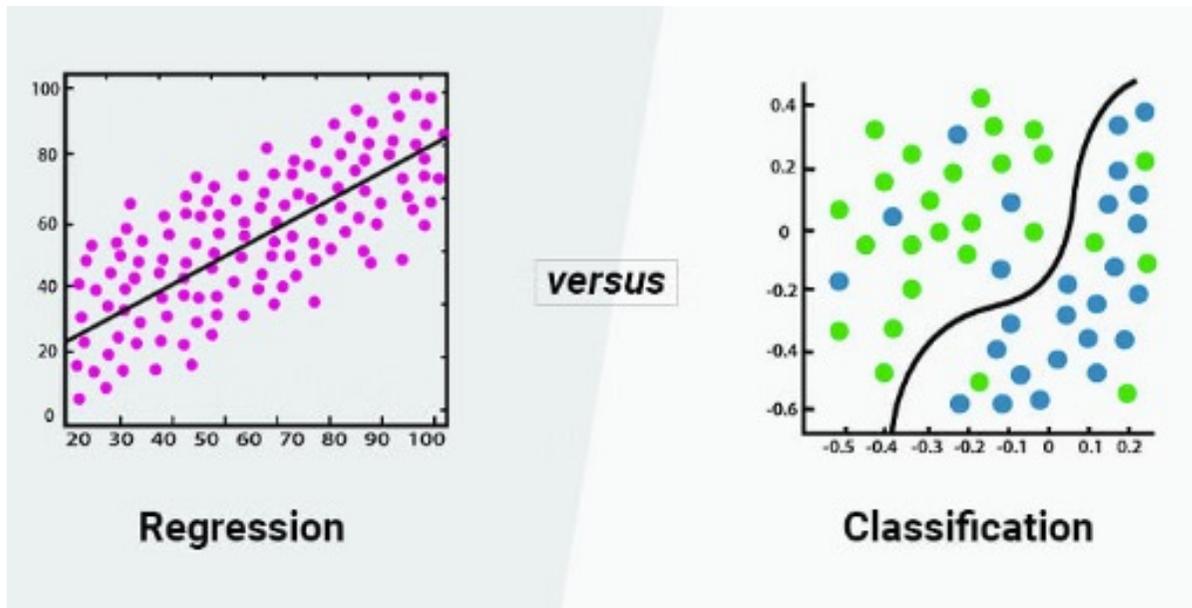
Will it be hot or cold tomorrow?



Fahrenheit

<https://www.enjoyalgorithms.com/blogs/classification-and-regression-in-machine-learning>

# Supervised learning



<https://www.simplilearn.com/regression-vs-classification-in-machine-learning-article>

# Regression vs Classification

Square Meters	Number of Bedrooms	Age of House (years)	Price (in \$1000s)
70	2	10	150
120	3	5	250
90	2	20	130
200	5	2	500
60	1	30	90

Goal: Predict the "Price"  
 (a **continuous** value)  
 Task: Regression

Email Length	Number of Links	Contains "Offer"	Is Spam (Label)
120	3	Yes	Yes
45	0	No	No
200	8	Yes	Yes
80	1	No	No

Goal: Predict "Is Spam"  
 (a **categorical** value: Yes/No).  
 Task: Classification

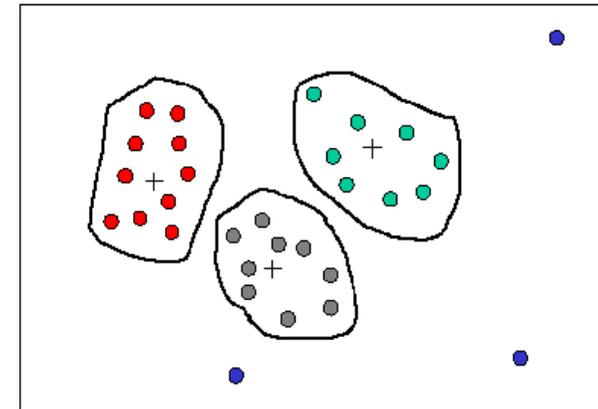
# Unsupervised learning

- Unsupervised learning is a type of machine learning in which models identify patterns or structures in data without using labeled outputs.

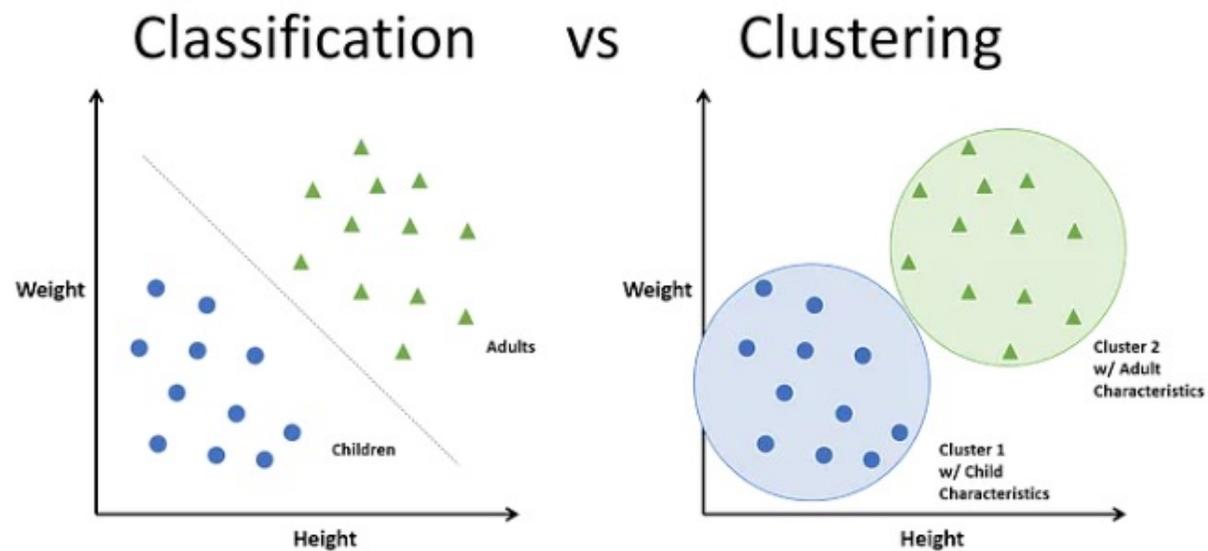
# Clustering

Clustering is a process of partitioning a set of data (or objects) in a set of meaningful sub-classes, called **clusters**

- Cluster:
  - ▶ a collection of data objects that are “similar” to one another and thus can be treated collectively as one group
  - ▶ but as a collection, they are sufficiently different from other groups

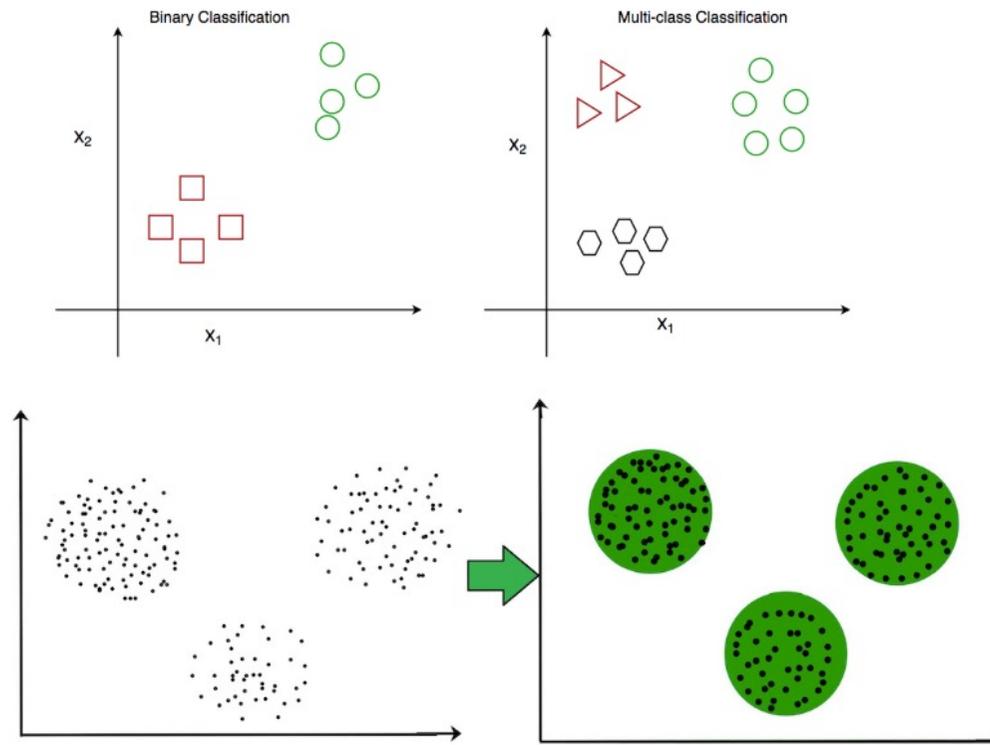


# Classification vs Clustering



<https://kevin-c-lee26.medium.com/machine-learning-101-classification-vs-clustering-e11b12c71243>

# Classification vs Clustering



<https://www.geeksforgeeks.org/ml-classification-vs-clustering/>

# Classification vs Clustering

Parameter	Classification	Clustering
Type	used for supervised learning	used for unsupervised learning
Basic	process of classifying the input instances based on their corresponding <b>class labels</b>	grouping the instances based on their <b>similarity</b> without the help of class labels
Need	it has labels so there is need of <b>training and testing</b> dataset for verifying the model created	there is no need of training and testing dataset
Example Algorithms	Logistic regression, K-NN, MLP, Decision Tree, Support vector machines, etc.	k-means clustering, agglomerative clustering, etc.

# Application of clustering

Where to use clustering?

- ✓ Document clustering
- ✓ Emotion detection
- ✓ Customer segmentation (Marketing)
- ✓ Image Processing and object detection
- ✓ Anomaly detection
- ✓ Genomics and bioinformatics
- ✓ Social network analysis and community detection

## Some common tasks

- We will look into two kinds of unsupervised learning in this course:
  - transformations of the dataset and clustering.

1. **Clustering**: grouping datapoints into clusters based on similarity

1. **Unsupervised transformations** of a dataset are algorithms that create a new representation of the data which might be easier for humans or other machine learning algorithms to understand.

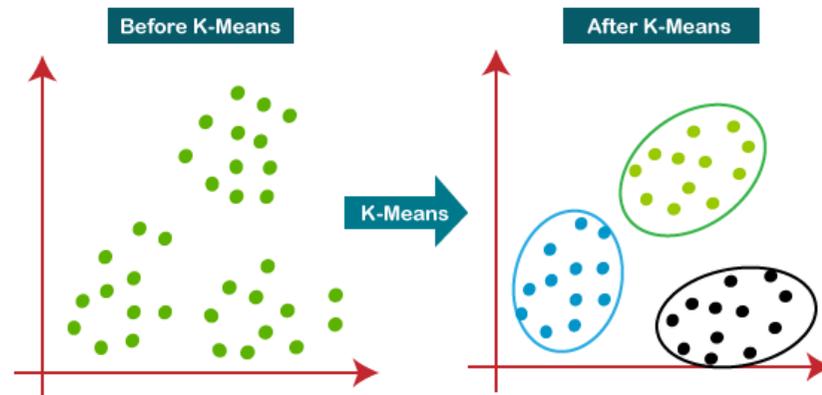
- dimensionality reduction

# Clustering Methodologies

- Two general methodologies
  - ▶ Partitioning Based Algorithms
  - ▶ Hierarchical Algorithms
- Partitioning Based
  - ▶ divide a set of  $N$  items into  $K$  clusters (top-down)
- Hierarchical
  - ▶ **agglomerative**: pairs of items or clusters are successively linked to produce larger clusters

# K-means clustering

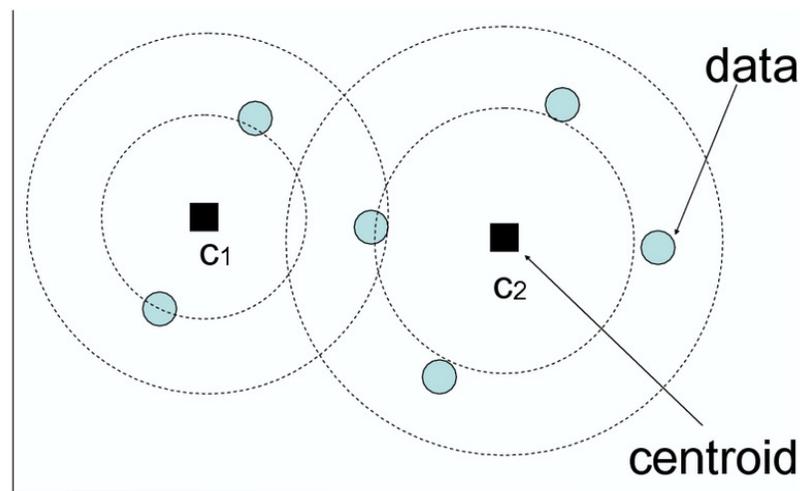
- The most widely used clustering algorithm.
- It partitions data into K distinct group based on feature similarity.
- Applications: Market segmentation, image compression, document clustering, and more.



<https://www.kaggle.com/code/mohammadbolandraftar/cluster-analysis-using-python-k-means>

# K-Means Clustering

- Separate the objects (data points) into K clusters.
- Cluster center (centroid) = the average of all the data points in the cluster.
- Assigns each data point to the cluster whose centroid is nearest (using distance function.)



# K-Means Algorithm

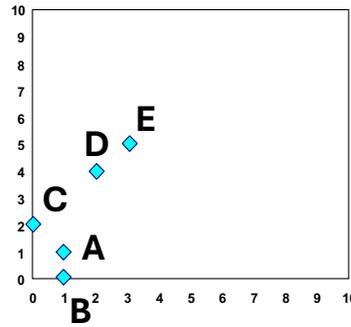
1. Assign K random points in the feature space as initial cluster centroids.
2. Based on distance from K centroids, assign nearest points to clusters.

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

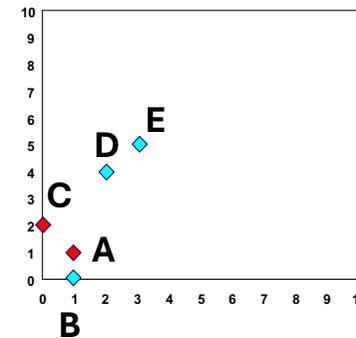
1. Recalculate the positions of the K centroids.
2. Repeat Steps 2 & 3 until the group centroids no longer move.

# Example

i	x1	x2
A	1	1
B	1	0
C	0	2
D	2	4
E	3	5

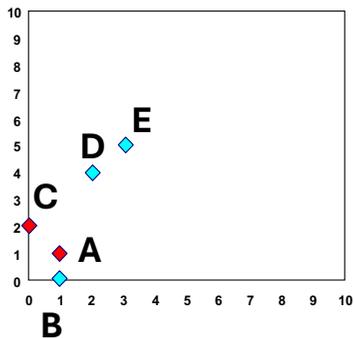


We can pick A and C as the two center to start.



# Example

i	x1	x2
A	1	1
B	1	0
C	0	2
D	2	4
E	3	5



Let's figure out how far away (distance) all of these points are from A and C:

Center1=A=(1,1)

Center2=C=(0,2)

A- Center1= $\sqrt{(1-1)^2+(1-1)^2}=0$

A- Center2= $\sqrt{(1-0)^2+(1-2)^2}=1.4$

B- Center1= $\sqrt{(1-1)^2+(1-0)^2}=0$

B is near Ceter1(A)

B- Center2= $\sqrt{(1-0)^2+(0-2)^2}=2.2$

C- Center1= $\sqrt{(0-1)^2+(2-0)^2}=2.2$

C- Center2= $\sqrt{(0-0)^2+(2-2)^2}=0$

D- Center1= $\sqrt{(2-1)^2+(4-0)^2}=\sqrt{17}=4.1$

D is near Ceter2(C)

D- Center2= $\sqrt{(2-0)^2+(4-2)^2}=\sqrt{8}=2.8$

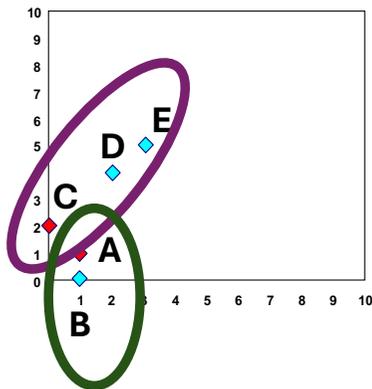
E is near Ceter2(C)

E- Center1= $\sqrt{(3-1)^2+(5-0)^2}=\sqrt{29}=5.3$

E- Center2= $\sqrt{(3-0)^2+(5-2)^2}=\sqrt{18}=4.2$

# Example

i	x1	X2	cluster
A	1	1	1
B	1	0	1
C	0	2	2
D	2	4	2
E	3	5	2



We need to find the average of each group, which is the average of all those members in that group.

Center1 = avg A,B =  $((1,1)+(1,0))/2 = (1,0.5)$

Center2 = AVG C,D,E =  $((0,2),(2,4),(3,5))/3 = (1.7, 3.7)$

**A**- Center1 =  $\text{sqrt}((1-1)^2+(1-0.5)^2) = 0.5$

A is near Ceter1

**A**- Center2 =  $\text{sqrt}((1-1.7)^2+(1-3.7)^2) = 2.8$

**B**- Center1 =  $\text{sqrt}((1-1)^2+(1-0.5)^2) = 0.5$

B is near Ceter1

**B**- Center2 =  $\text{sqrt}((1-1.7)^2+(0-3.7)^2) = 3.7$

**C**- Center1 =  $\text{sqrt}((0-1)^2+(2-0.5)^2) = 1.8$

C is near Ceter1

**C**- Center2 =  $\text{sqrt}((0-1.7)^2+(2-3.7)^2) = 2.4$

**D**- Center1 =  $\text{sqrt}((2-1)^2+(4-0.5)^2) = 3.6$

**D**- Center2 =  $\text{sqrt}((2-1.7)^2+(4-3.7)^2) = 0.42$

D is near Ceter2

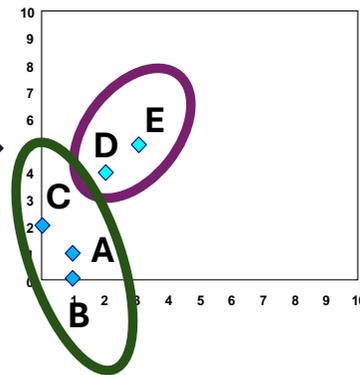
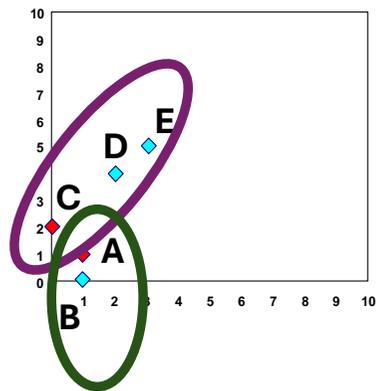
**E**- Center1 =  $\text{sqrt}((3-1)^2+(5-0.5)^2) = 4.9$

E is near Ceter2

**E**- Center2 =  $\text{sqrt}((3-1.7)^2+(5-3.7)^2) = 1.8$

# Example

i	x1	X2	cluster
A	1	1	1
B	1	0	1
C	0	2	1
D	2	4	2
E	3	5	2

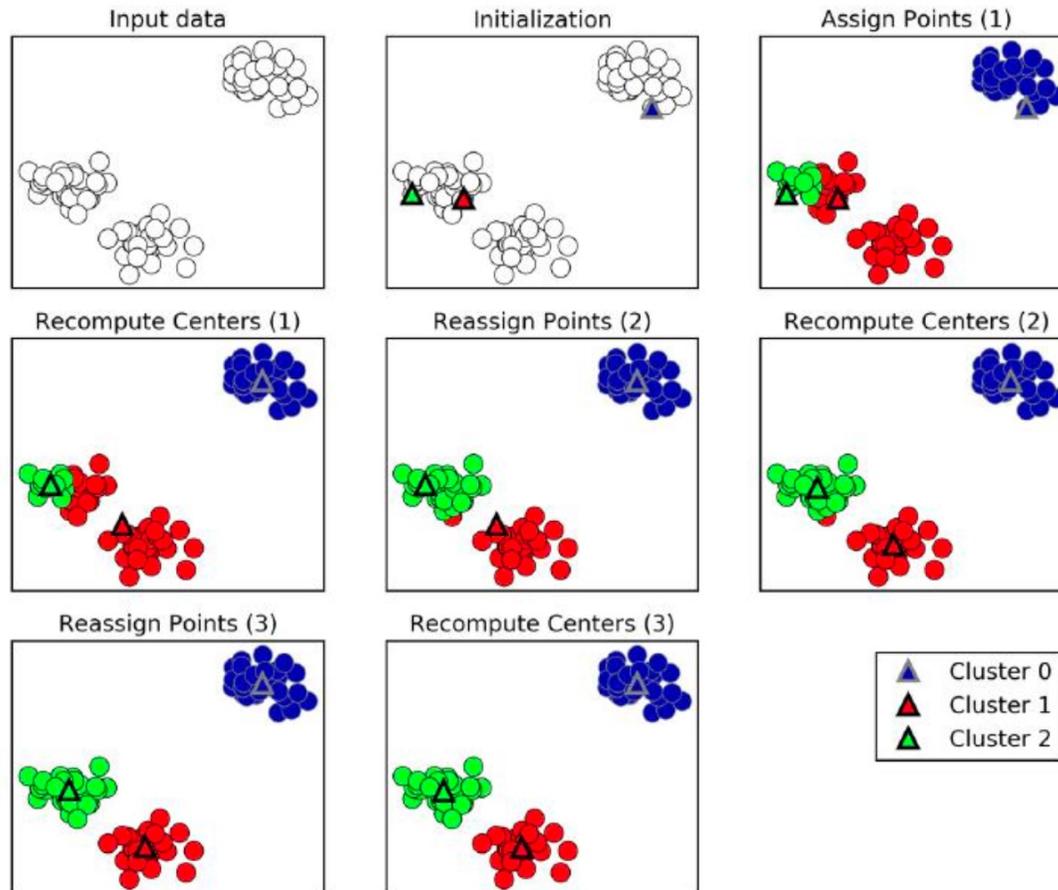


# Example

- We keep going until we get the same result as our previous step. Since it doesn't seem to have changed, we can stop and report the clusters. (centers no change)

Each cluster of data reveals a unique narrative. By identifying these natural groupings, we gain valuable insights into the underlying structure of our data.

Chaque cluster de données révèle une histoire unique. En identifiant ces groupements naturels, nous obtenons des informations précieuses sur la structure sous-jacente de nos données.

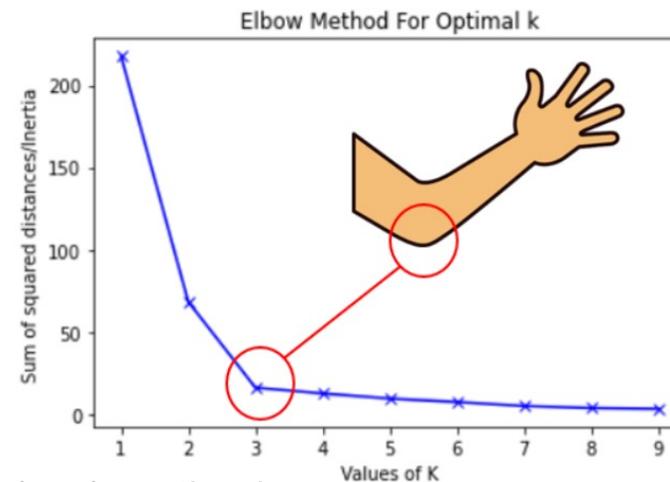


# Questions

- How to create 'good' clusters?
- How many clusters do we need?

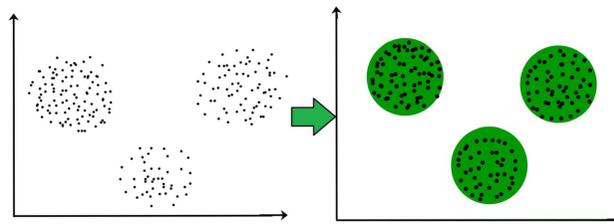
# How many clusters – Elbow method

- Find the optimal number of clusters  $K$  by minimizing the within-cluster sum of squares
- Elbow method:
  - The point where the rate of decrease sharply slows down (resembles an « elbow») is considered the optimal  $K$ .
  - The elbow method calculates the sum of the square of the points and calculates the average distance.



# What Is Good Clustering?

- A **good clustering** method will produce high quality clusters with
  - Low **intra-class similarity**
  - High **inter-class similarity**
- The quality of a clustering result depends on both the similarity measure used by the method and its implementation.
- The quality of a clustering method is also measured by its ability to discover some or all of the hidden patterns.



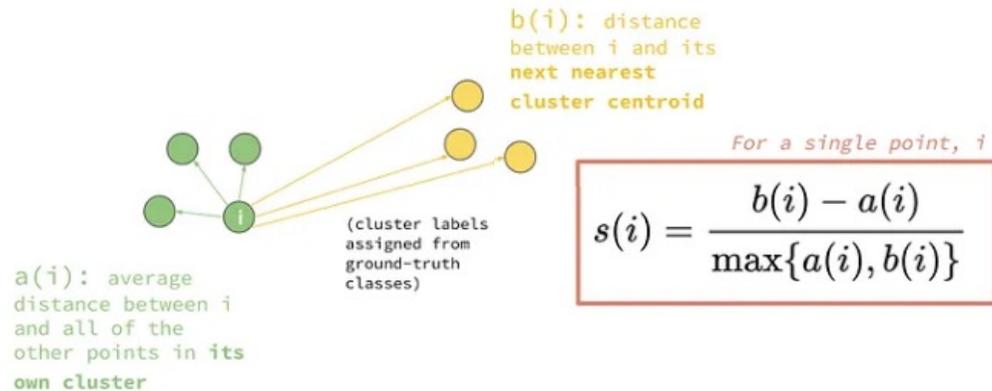
# Objective of k means Clustering

- The main objective of k-means clustering is to partition your data into a specific number (k) of groups
- **Grouping similar data points**
- **Minimizing within-cluster distance**
- **Maximizing between-cluster distance**

# Evaluation - Silhouette Score

- It provides a quantitative way to measure how well data points are grouped and separated into clusters.
- **+1**: Perfect clustering — the point is far from other clusters and well-matched to its own cluster.
- **0**: Borderline clustering — the point is equidistant between two clusters.
- **-1**: Poor clustering — the point is closer to another cluster than its own.

# Silhouette Score

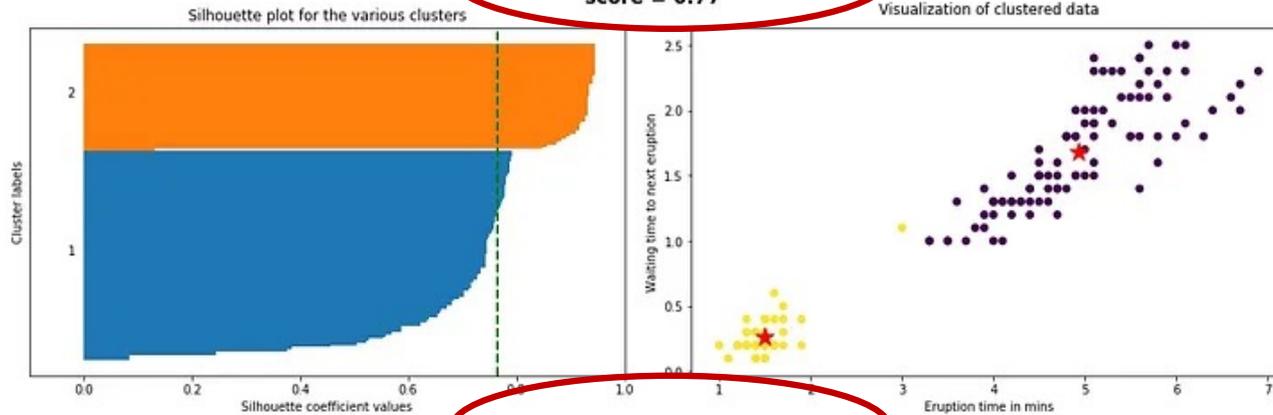


It considers two aspects:

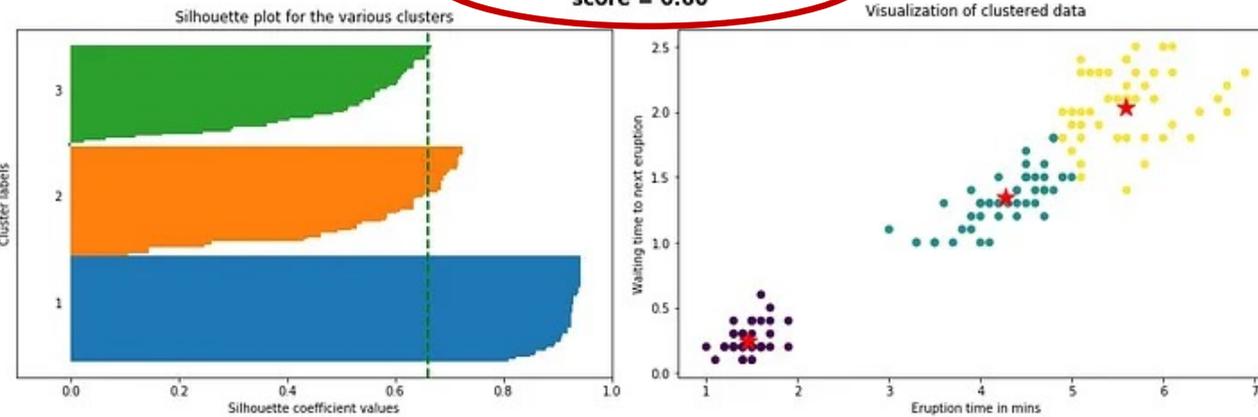
- 1. Cohesion ( $a(i)$ ):** How close a data point is to other points in its own cluster.
- 2. Separation ( $b(i)$ ):** How far a data point is from the points in the nearest neighboring cluster.

# Silhouette Score

Silhouette analysis using k = 2  
score = 0.77



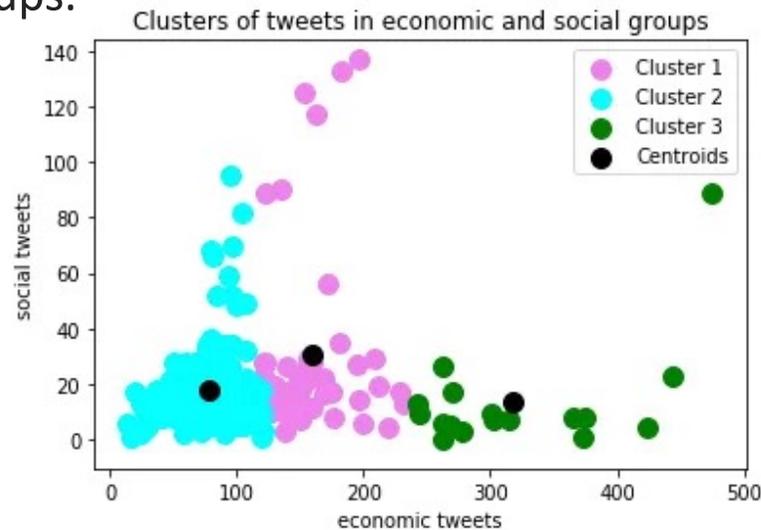
Silhouette analysis using k = 3  
score = 0.66



we can clearly see that how plot and score are different according to  $n\_cluster(k)$ . So, we can easily choose **high score** and **number of k** via silhouette analysis technique

## Example of k-Means-Tweets Clustering

- 4 major categories: *Economic, Social, Cultural and Health* then performing KMeans cluster analysis on the groups.



<https://medium.com/swlh/tweets-classification-and-clustering-in-python-b107be1ba7c7>

The plot above indicates most of the users share more economy-centred tweets compared to social tweets. There's a few who try to maintain a balance between the categories.

## Applications of K-Means in Social Media Management

- **Audience Segmentation**

Social media platforms cater to diverse audiences with different interests and behaviors. K-means clustering helps group users based on: Age, gender, location, Likes, shares, comments interests and Common topics they interact with.

- **Hashtag and Content Strategy Optimization**

Marketers use K-means to analyze hashtags and content performance by clustering posts based on: Engagement metrics (likes, shares, comments), Content themes (tech, fitness, fashion, etc.) and Posting time.

## Applications of K-Means in Social Media Management

- **Influencer Identification**

K-means helps businesses identify and categorize influencers based on: Engagement rates , Follower demographics. This ensures brands collaborate with the right influencers for maximum impact.

- **Sentiment Analysis and Customer Insights**

- K-means clustering helps analyze user comments, reviews, and discussions to track sentiment trends. By grouping text data into positive, neutral, and negative clusters, brands can Improve customer service.

# Challenges - Data distribution

- When using **K-means clustering** on complex distributions, there are a key challenge:
- **Assumption of spherical clusters:** K-means assumes that clusters are spherical (or roughly circular). This means it struggles when the data has clusters with different shapes, densities, or sizes (e.g., elongated, concentric, or overlapping clusters).
- If your data consists of non-convex clusters (like crescent shapes or clusters with holes), K-means will struggle to identify them correctly.

# Example of k-means clustering

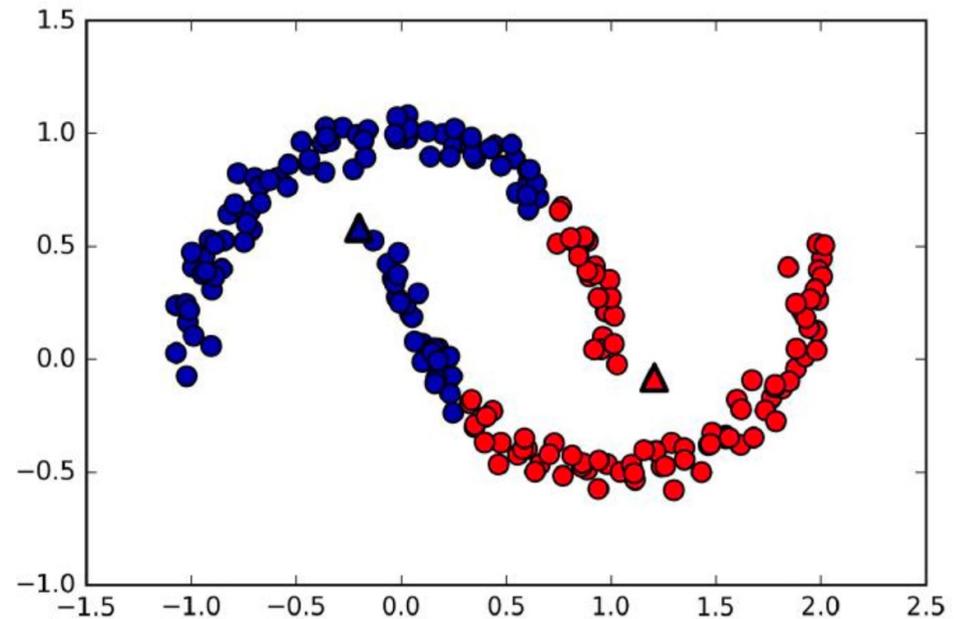
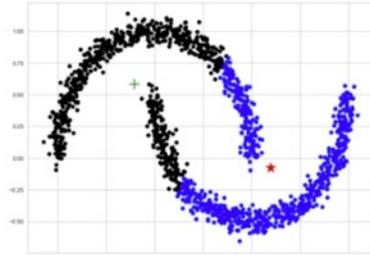
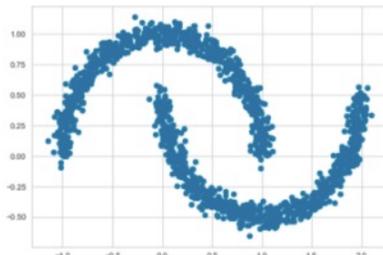
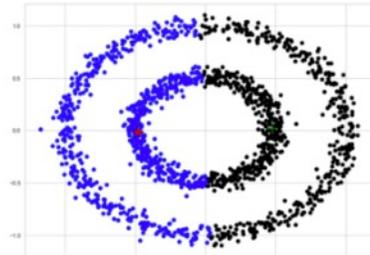
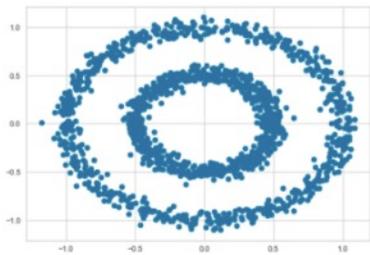
```
df = pd.DataFrame(data)

# 2. Scaling (Important for K-Means!)
scaler = StandardScaler()
df_scaled = scaler.fit_transform(df)

# 3. Running K-Means
# Let's try to find 3 groups (e.g., Low, Medium, High Risk)
kmeans = KMeans(n_clusters=3, random_state=42)
df['Cluster'] = kmeans.fit_predict(df_scaled)
```

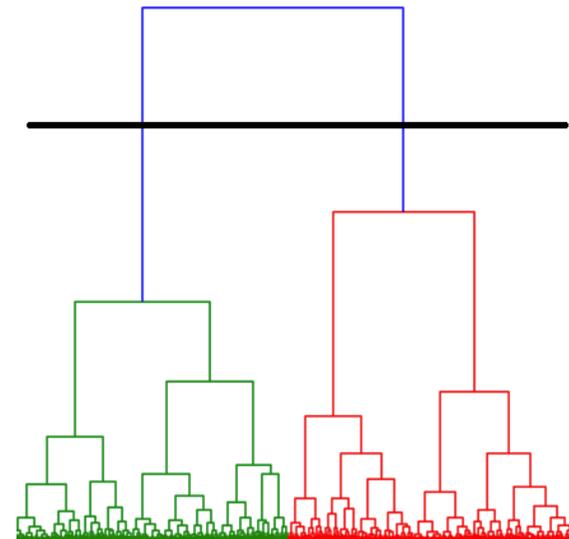
# Challenges - Data distribution

- the two\_moons data



# Challenges - Data distribution

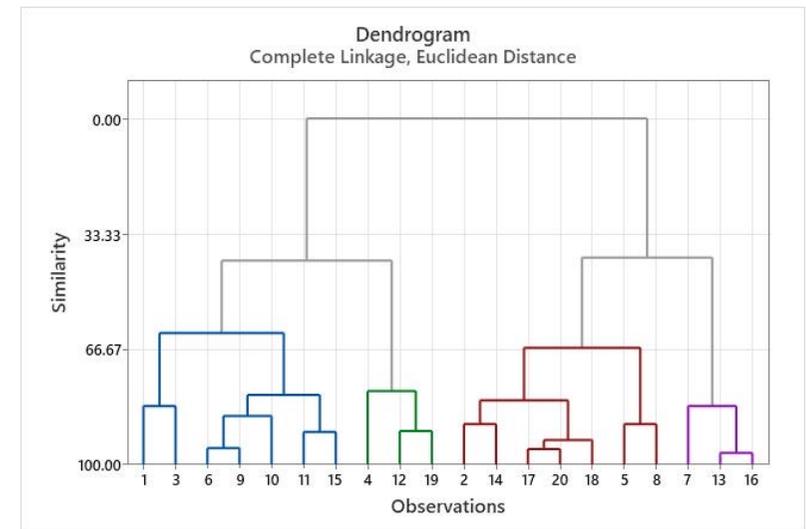
**Agglomerative Clustering:** Can be more flexible than K-means and handle complex shapes, though still limited with respect to density differences.



# Hierarchical clustering

- It is a Clustering method that create a **hierarchical tree** of objects.
- The tree represents the relationships between the objects.
- It is Bottom-up approach:
  - Since each observation consider as a cluster
  - Then they combine together based on their similarities to form clusters
- The result of a hierarchical clustering is called a dendrogram

<https://support.minitab.com/en-us/minitab/help-and-how-to/statistical-modeling/multivariate/how-to/cluster-observations/interpret-the-results/all-statistics-and-graphs/dendrogram/>

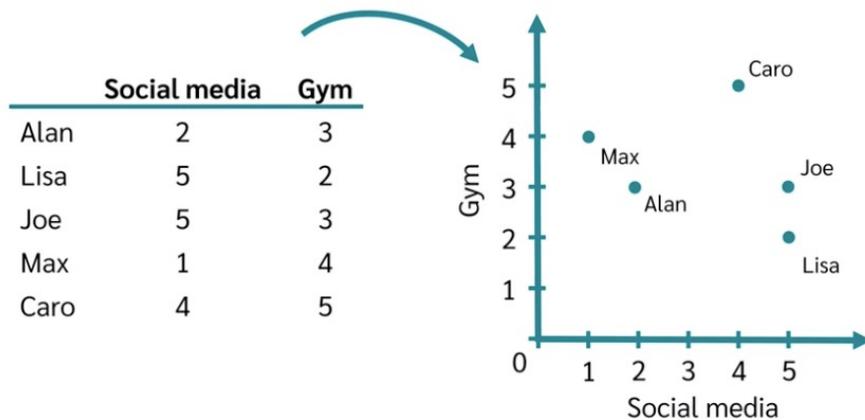


# Steps of hierarchical clustering

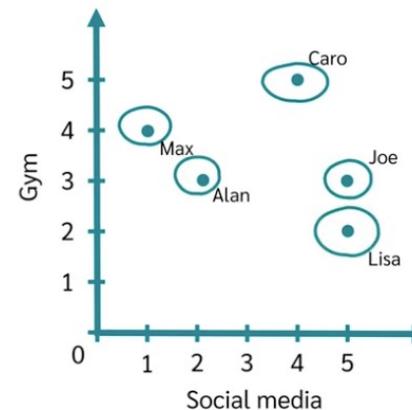
- Start by assigning each item to a cluster
- Let the distances (similarities) between the clusters the same as the distances (similarities) between the items they contain.
- Find the closest (most similar) pair of clusters and merge them into a single cluster, so that now you have one cluster less.
- Compute distances (similarities) between the new cluster and each of the old clusters.
- Repeat steps 2 and 3 until all items are clustered into K number of clusters

# Steps of hierarchical clustering

We asked people how many hours a week they spend on social media platforms and in the gym.



The **first step** is to assign a **cluster** to each **individual point**.

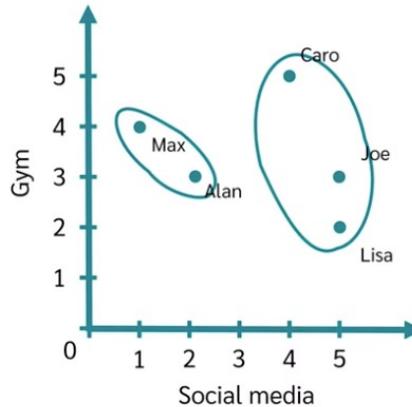
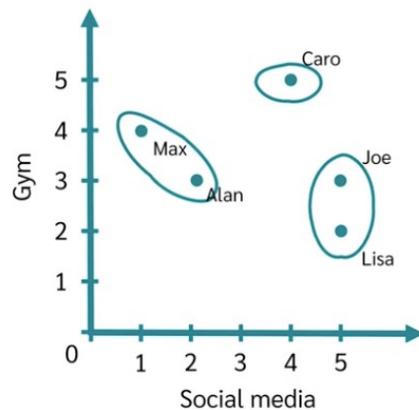


# Steps of hierarchical clustering

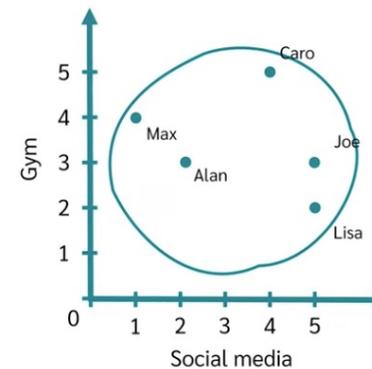
We asked people how many hours a week they spend on social media platforms and in the gym.

The **goal** now is:

Gradually merge more and more clusters,



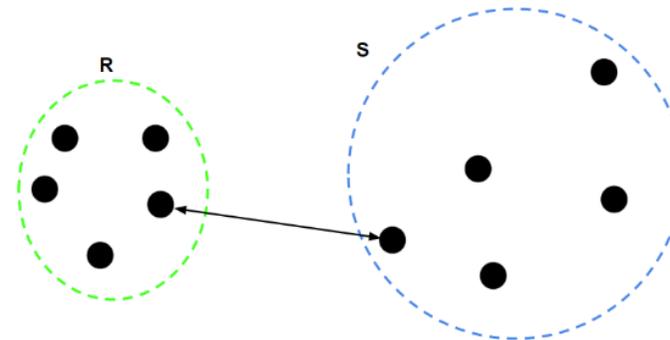
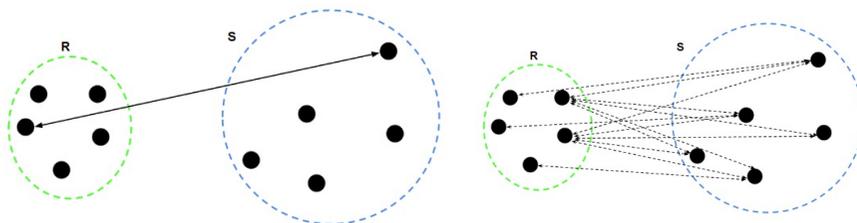
until finally **all points** are in **one cluster**.



<https://www.youtube.com/watch?v=8QCBI-xdeZI>

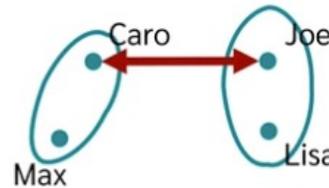
# Measuring similarity between clusters

- There are some method to calculate distances between clusters.
- The most commonly use method is **single linkage** method.
- If you have two clusters, the distance between these two, is the shortest possible distance
- There are other method like :
  - Average Linkage
  - And complete linkage



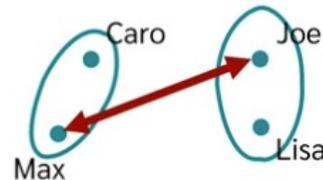
# Measuring similarity between clusters

**Single-linkage** uses the distance between the **closest elements** in the cluster.



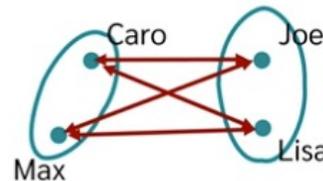
So, the distance between Caro and Joe.

**Complete Linkage** uses the distance between the **most distant elements** of the cluster.



So, between Max and Joe.

**Average linkage** uses the average of all pairwise distances.

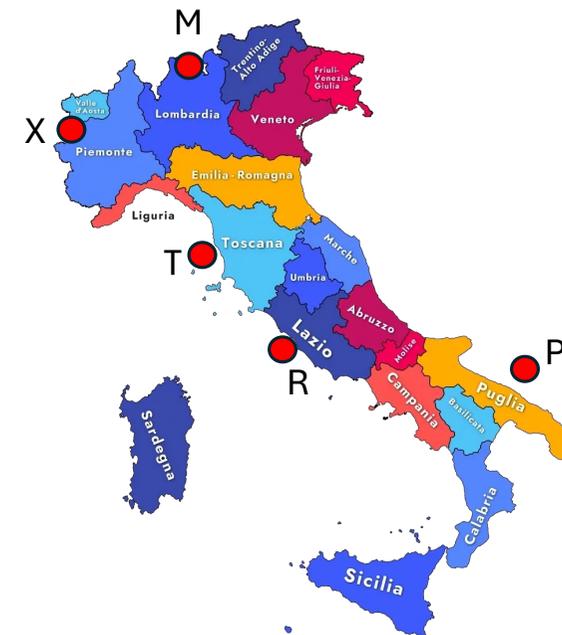


From each combination the distance is calculated and from it the average value.

# Example of Hierarchical clustering

Distances in kilometers between some italian cities.

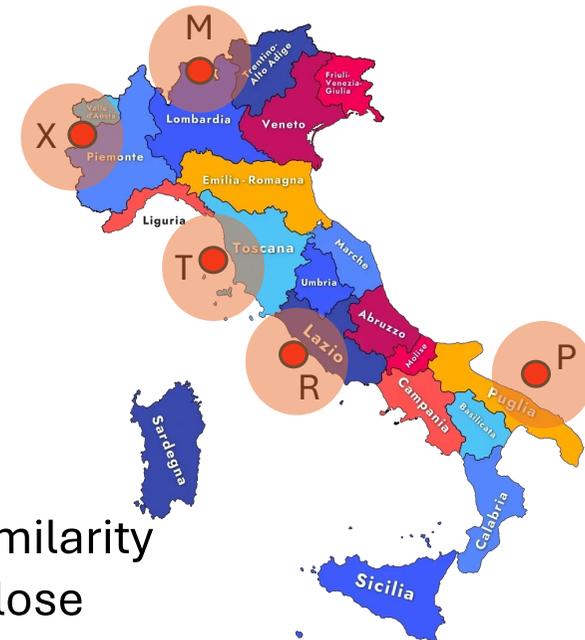
	P	T	M	R	X
P	0	662	877	412	996
T	662	0	295	268	400
M	877	295	0	564	138
R	412	268	564	0	669
X	996	400	138	669	0



The goal here is to find cities that are closest.

# Example of Hierarchical clustering

	P	T	M	R	X
P	0	662	877	412	996
T	662	0	295	268	400
M	877	295	0	564	138
R	412	268	564	0	669
X	996	400	138	669	0



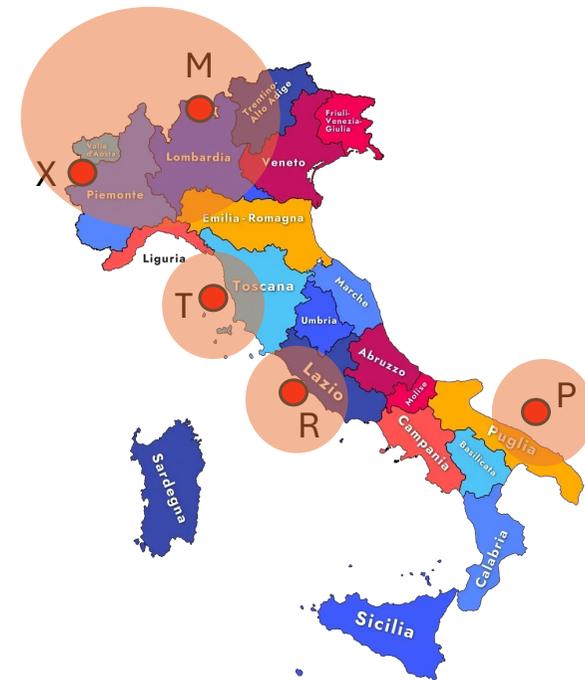
- Here, first we assume each city is the cluster
- Then we start to put cities together based on their similarity
- distance between x and M is 138 and they are very close
- So we have new cluster (M,X)
- We should update distance matrix

# Example of Hierarchical clustering

	P	T	M	R	X
P	0	662	877	412	996
T	662	0	295	268	400
M	877	295	0	564	138
R	412	268	564	0	669
X	996	400	138	669	0

I eliminated one row and one column

	P	T	M/X	R
P	0	662		412
T	662	0		268
M/X			0	
R	412	268		0

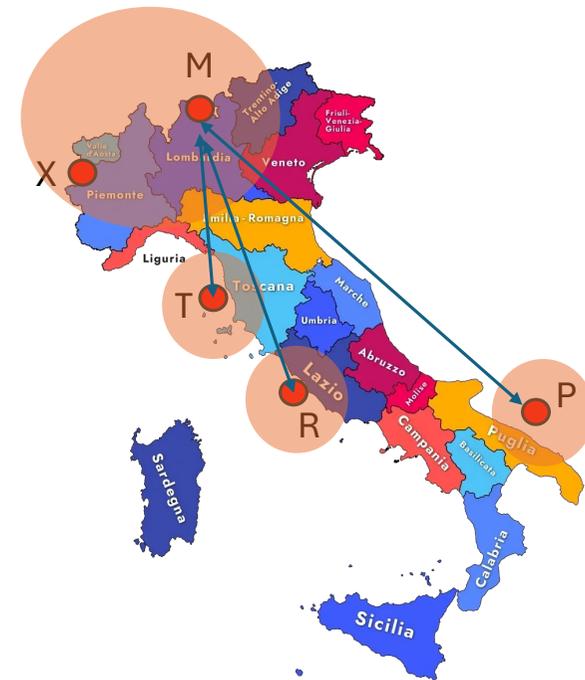


- We combine M and X and have a new cluster
- We need to calculate the distance from other cities to this new cluster.

# Example of Hierarchical clustering

	P	T	M	R	X
P	0	662	877	412	996
T	662	0	295	268	400
M	877	295	0	564	138
R	412	268	564	0	669
X	996	400	138	669	0

	P	T	M/X	R
P	0	662	877	412
T	662	0	295	268
M/X	877	295	0	564
R	412	268	564	0

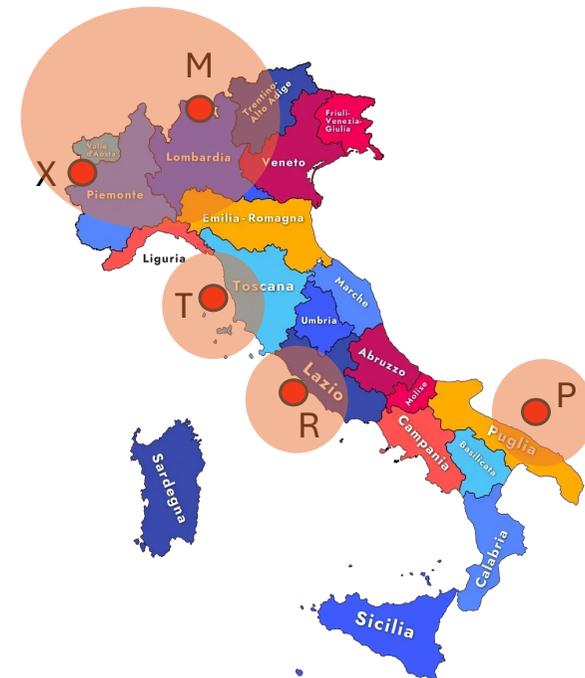
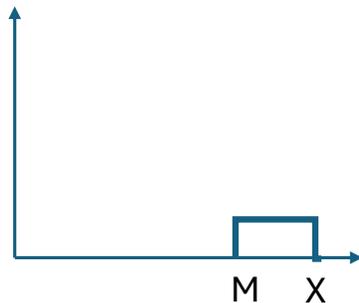


- Single linkage

# Example of Hierarchical clustering

	P	T	M/X	R
P	0	662	<b>877</b>	412
T	662	0	<b>295</b>	268
M/X	<b>877</b>	<b>295</b>	0	564
R	412	268	564	0

- We have to continue this process of combining
- And in each step we can create initial dendrogram



# Example of Hierarchical clustering

	P	T	M/X	R
P	0	662	<b>877</b>	412
T	662	0	<b>295</b>	268
M/X	<b>877</b>	<b>295</b>	0	564
R	412	268	564	0

- What are the next two cities that are the closest together?

	P	T	M/X	R
P	0	662	<b>877</b>	412
T	662	0	<b>295</b>	268
M/X	<b>877</b>	<b>295</b>	0	564
R	412	268	564	0

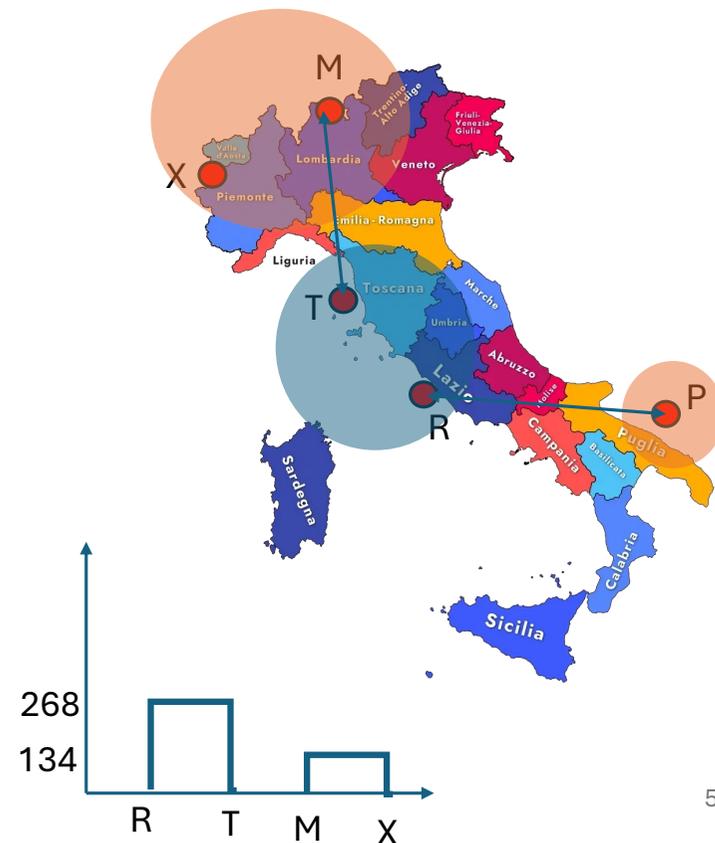


# Example of Hierarchical clustering

	P	T	M/X	R
P	0	662	877	412
T	662	0	295	268
M/X	877	295	0	564
R	412	268	564	0

	P	R/T	M/X
P	0	412	877
M/X	877	295	0
R/T	412	0	295

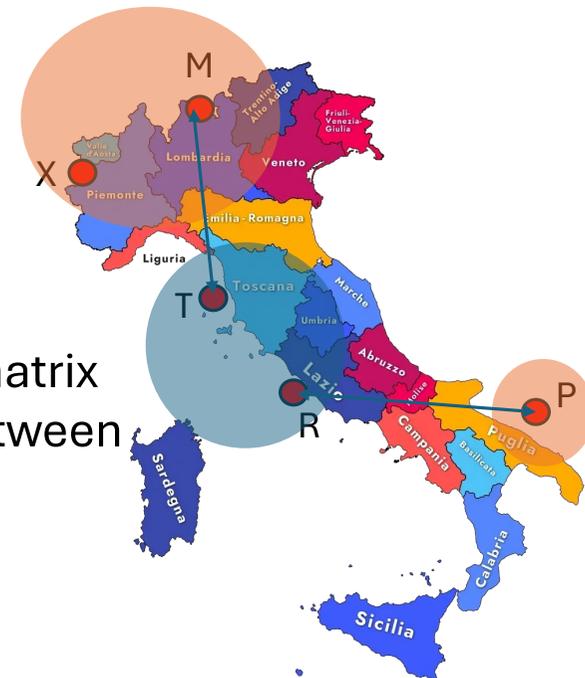
- Know we have a new cluster (R/T)
- In each step we refer to previous matrix



# Example of Hierarchical clustering

	P	R/T	M/X
P	0	412	877
M/X	877	295	0
R/T	412	0	295

- We continue this process until we get to a 2 by 2 matrix
- To update matrix the minimum distance is now between M/X and R/T – 295
- So we combine R/T and M/X cluster

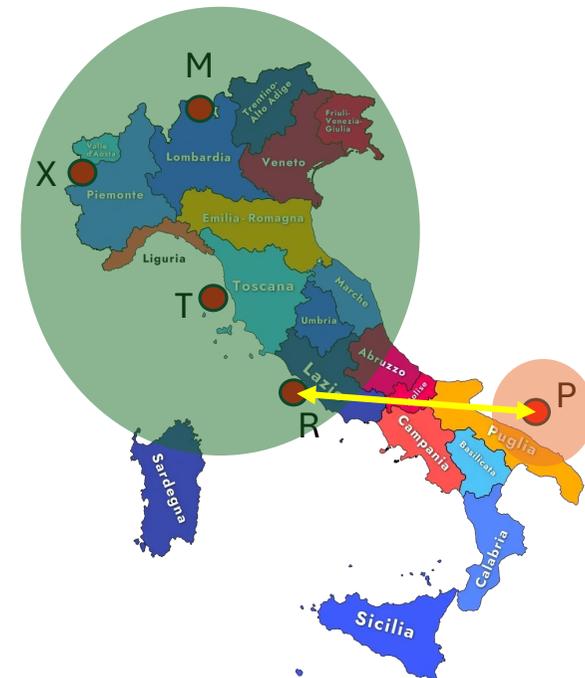


# Example of Hierarchical clustering

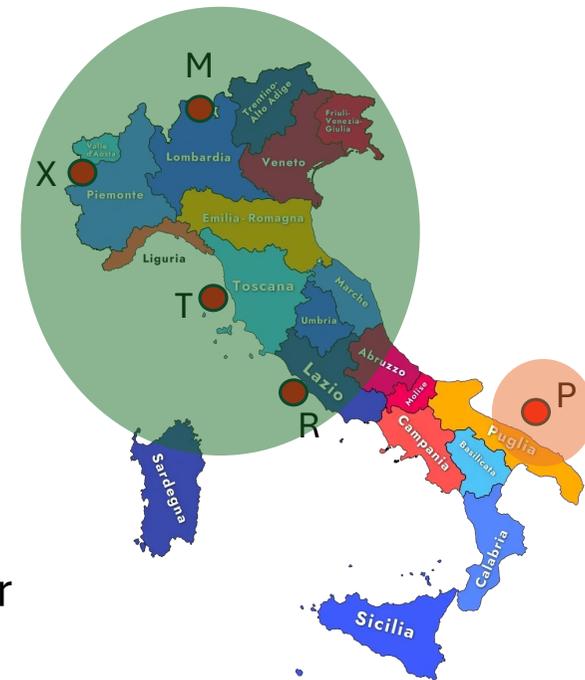
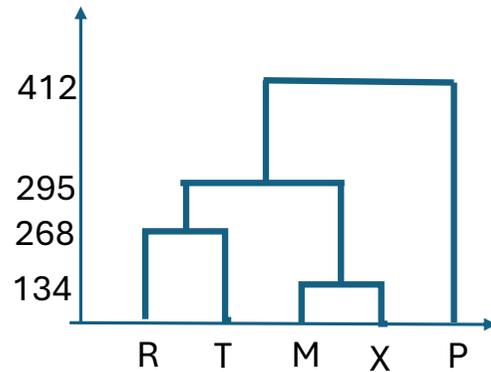
	<b>P</b>	<b>R/T</b>	<b>M/X</b>
<b>P</b>	0	412	877
<b>M/X</b>	877	295	0
<b>R/T</b>	412	0	295

	<b>P</b>	<b>R/T/M/X</b>
<b>P</b>	0	412
<b>R/T/M/X</b>	412	0

- So we combine R/T and M/X cluster
- We update the matrix
- P is near to R (single linkage)

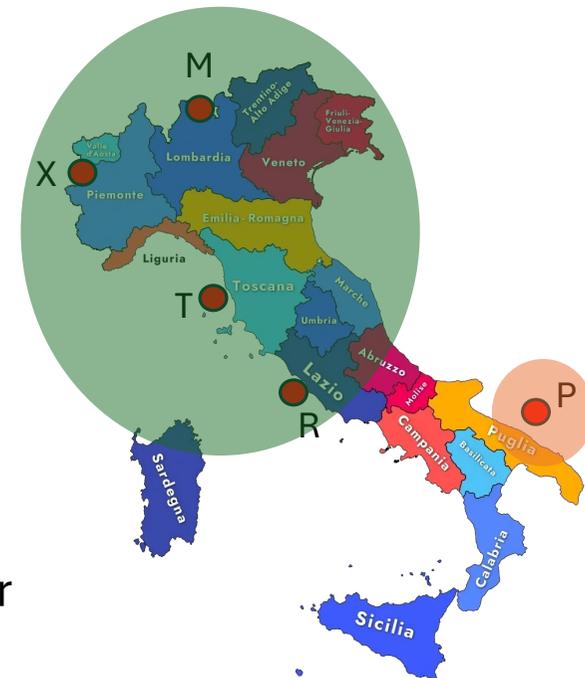
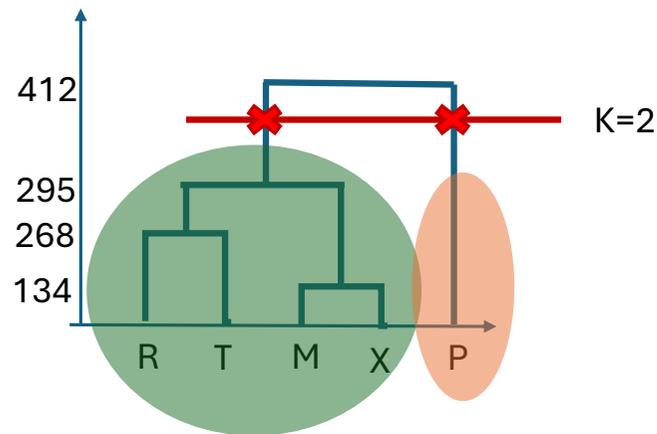


# Example of Hierarchical clustering



- We know that the length of this bar shows how far apart two clusters are from each other.
- To define the number of groups we have to create horizontal lines that hit the vertical lines.

# Example of Hierarchical clustering

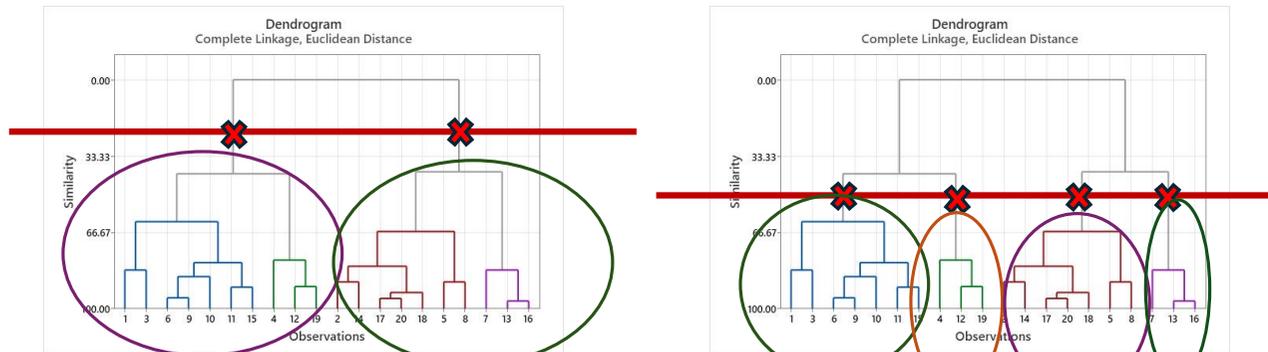


- We know that the length of this bar shows how far apart two clusters are from each other.
- To define the number of groups we have to create horizontal lines that hit the vertical lines.



# Hierarchical clustering

- We know that the length of the bar in the dendrogram shows how far apart two clusters are from each other.
  - The longer this length is, the more far apart two clusters are from each other.
- How can I use dendrogram to perform the grouping of observations?
  - Create horizontal lines that hit the vertical lines
  - Each of the vertical lines lead us to a new cluster



## Hierarchical clustering versus k-means clustering

- ✓ If there is a specific number of clusters in the dataset, but the group they belong to is unknown, choose K-means
- ✓ With a **large number** of variables, **K-means** compute faster
- ✓ It is easier to determine the number of clusters by hierarchical clustering's dendrogram

**END**