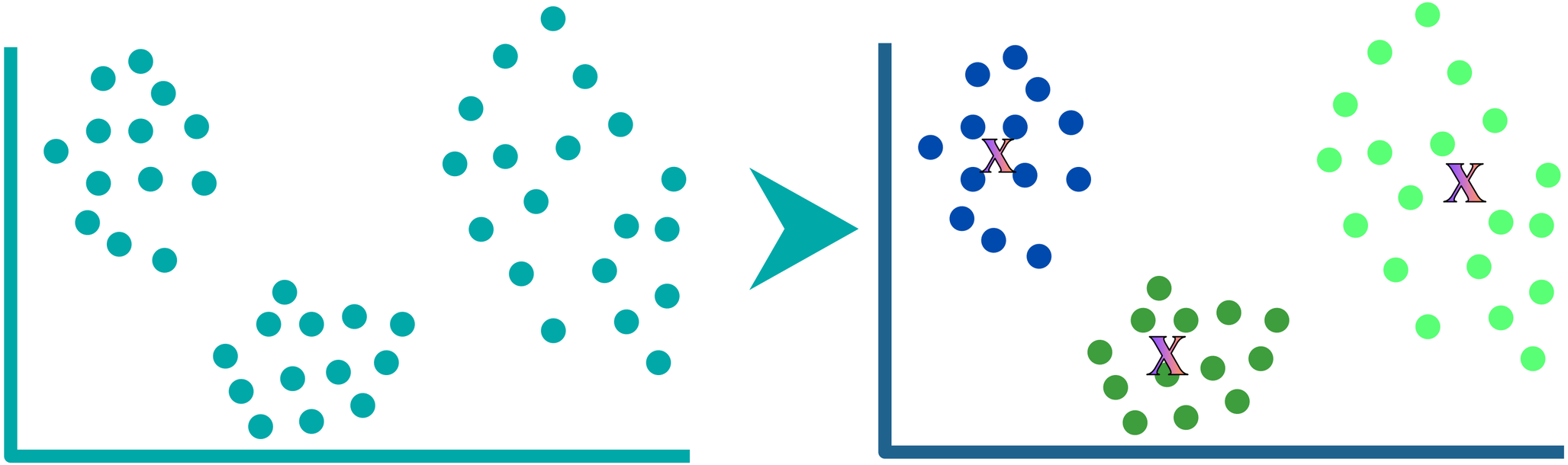
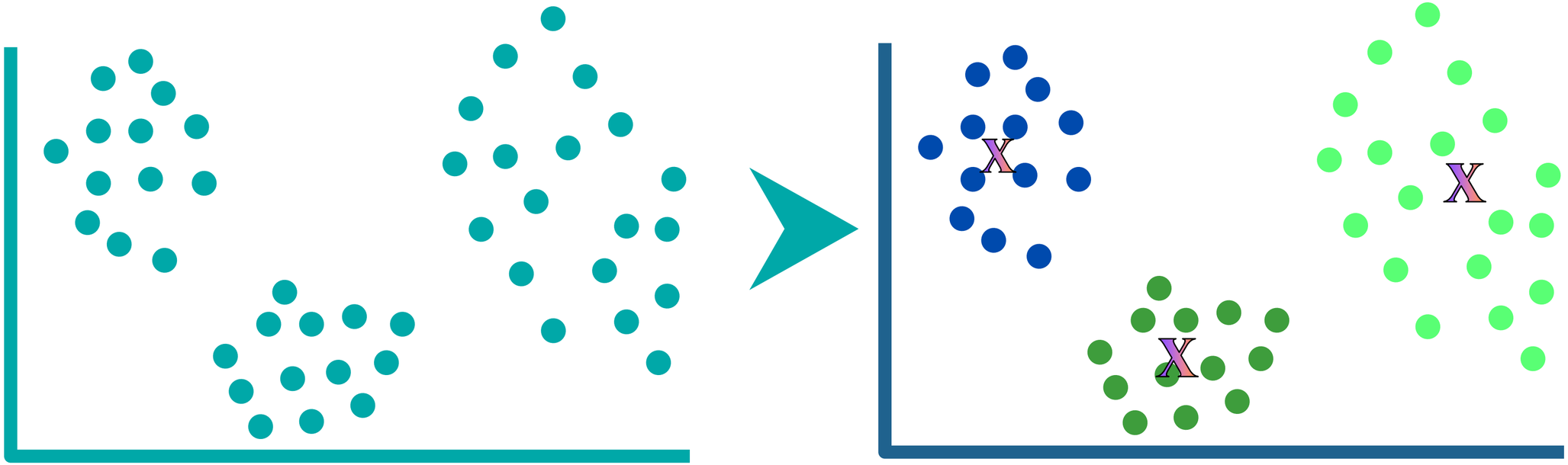
What is K-Means Clustering



How does the K-Means Clustering Work



codefinity

Initial Data set

Step 1

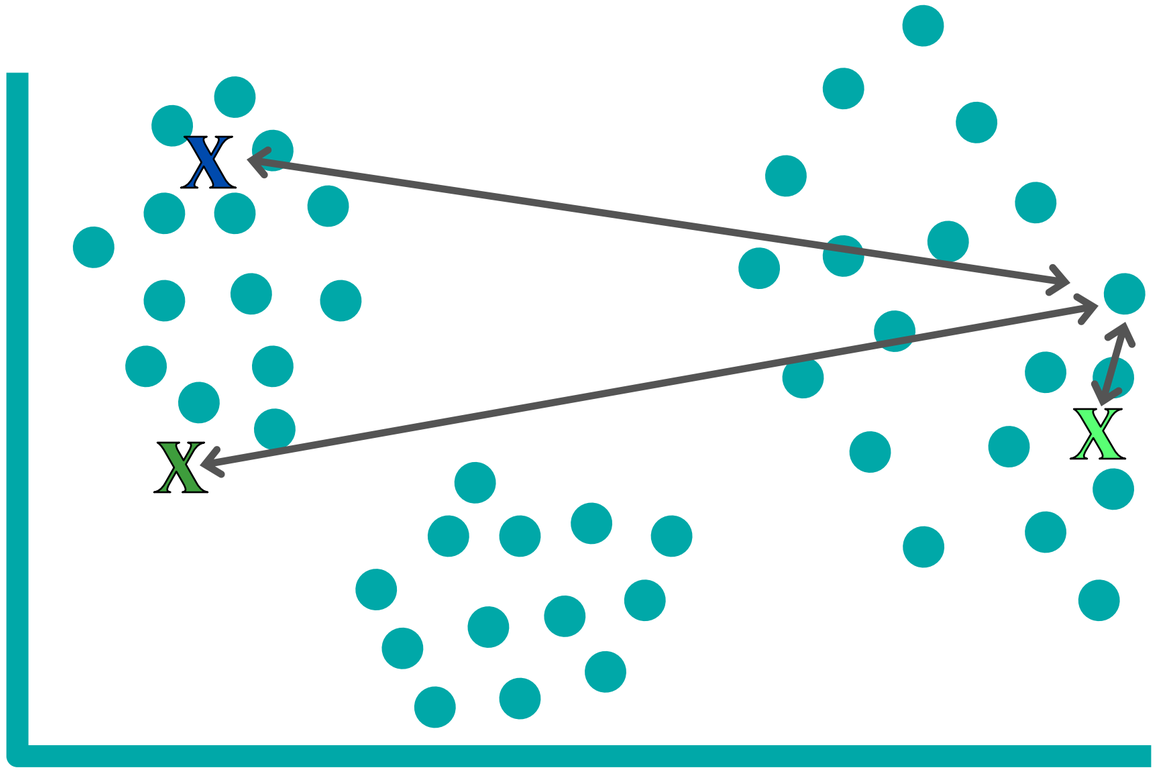
codefinity

Set Cluster Center Randomly

Step 2

codefinity

Calculate the Distance of the Points to the Clusters



Step 3

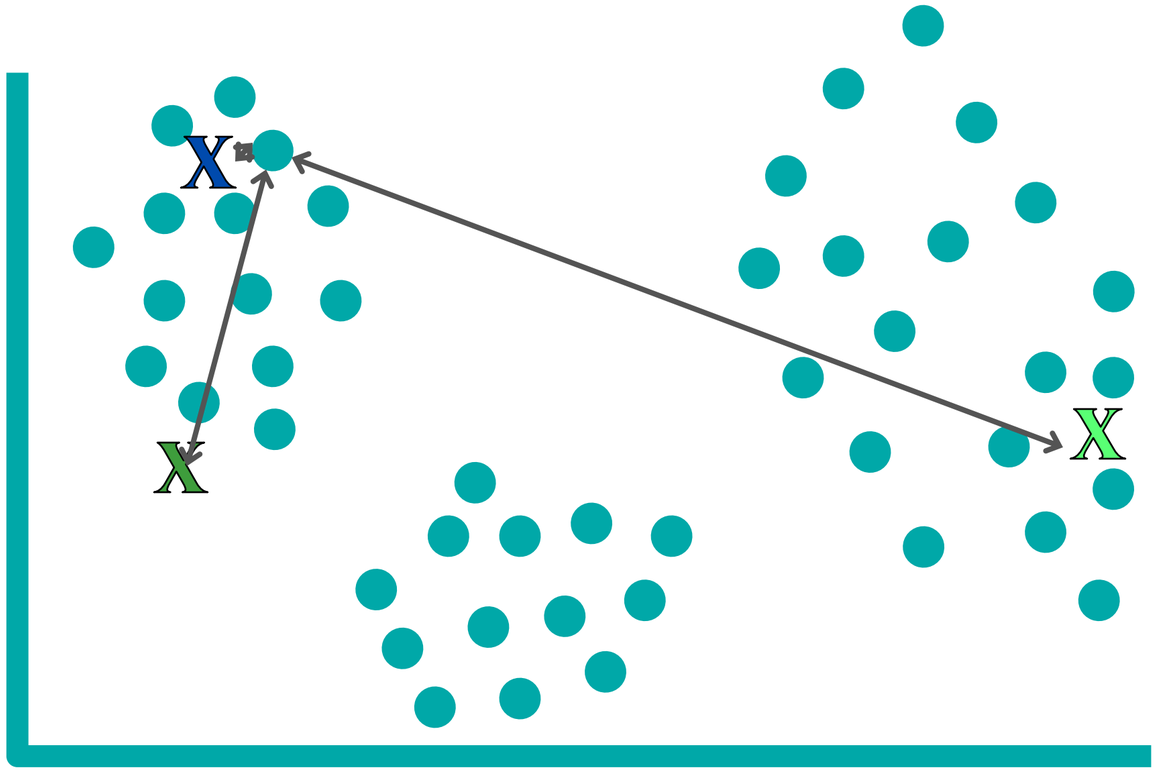
codefinity

Assign Point to the Closest Clusters

Step 3

codefinity

Assign Point to the Closest Clusters



Step 3

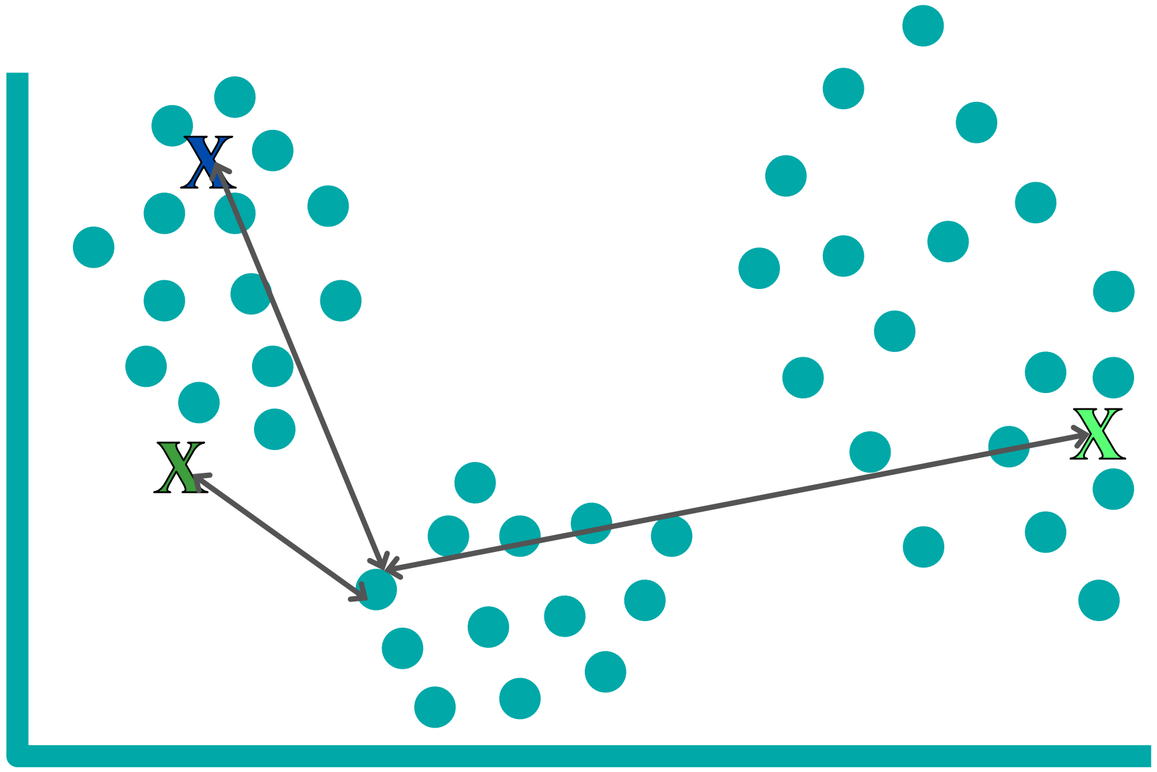
codefinity

Assign Points to Clusters

Step 3

codefinity

Assign Points to Clusters



Step 3

codefinity

Assign Points to Clusters

Step 3

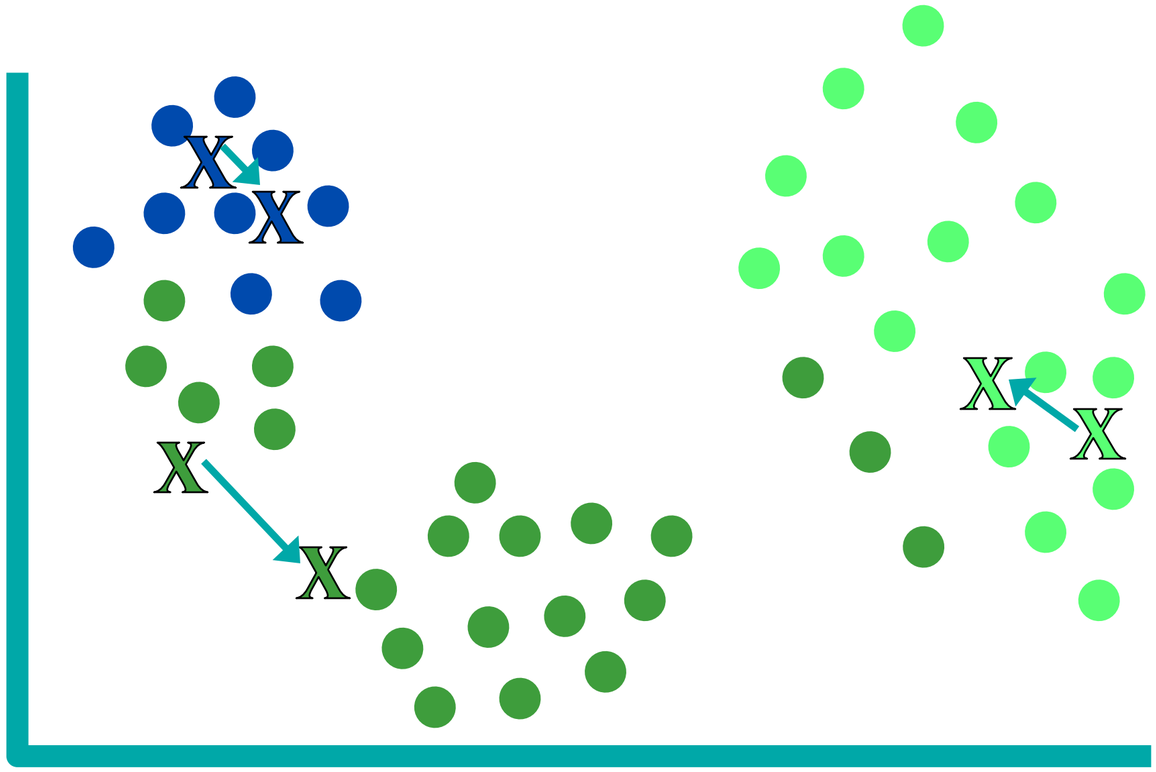
codefinity

Assign Points to Clusters

Step 4

codefinity

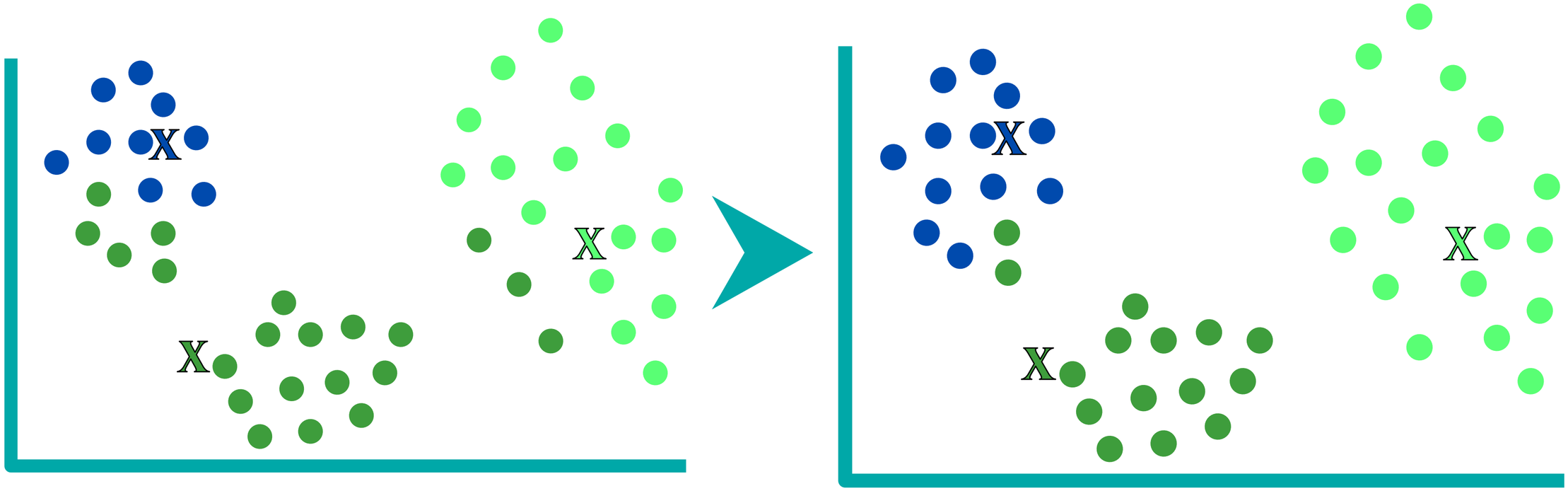
Calculate the Center of each Cluster



Step 5

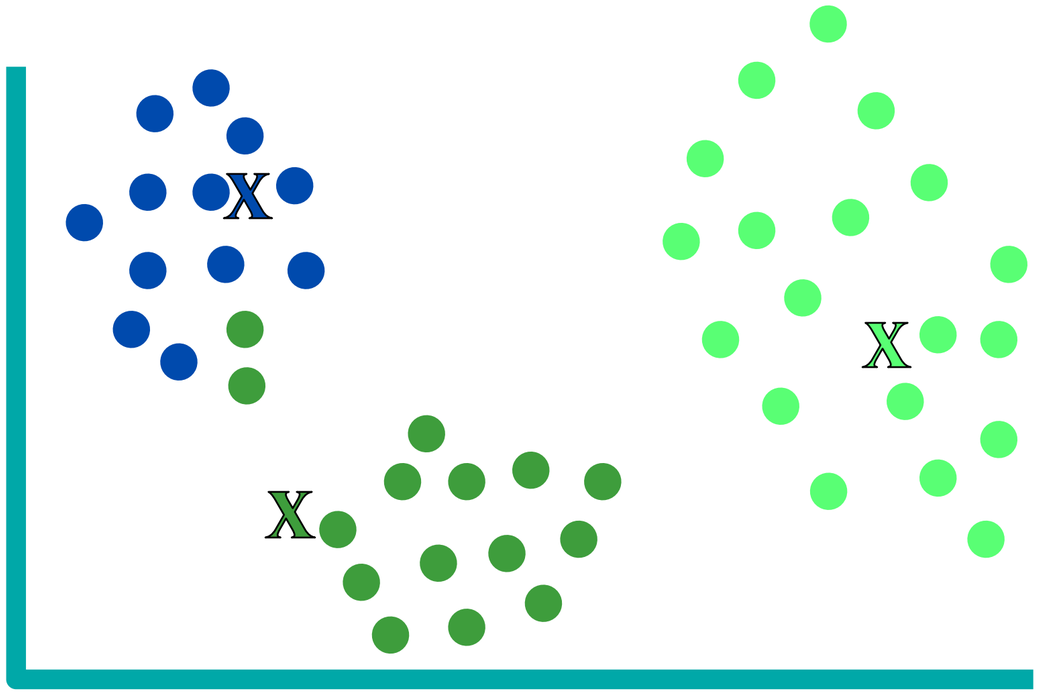
codefinity

Assign Points to the new Clusters

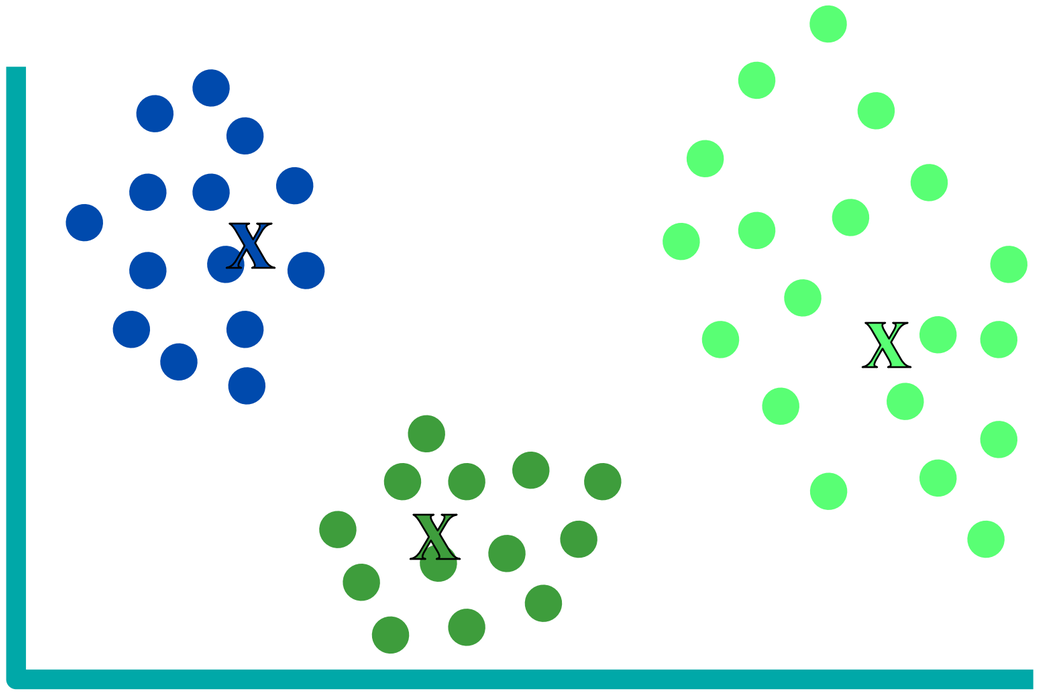


codefinity

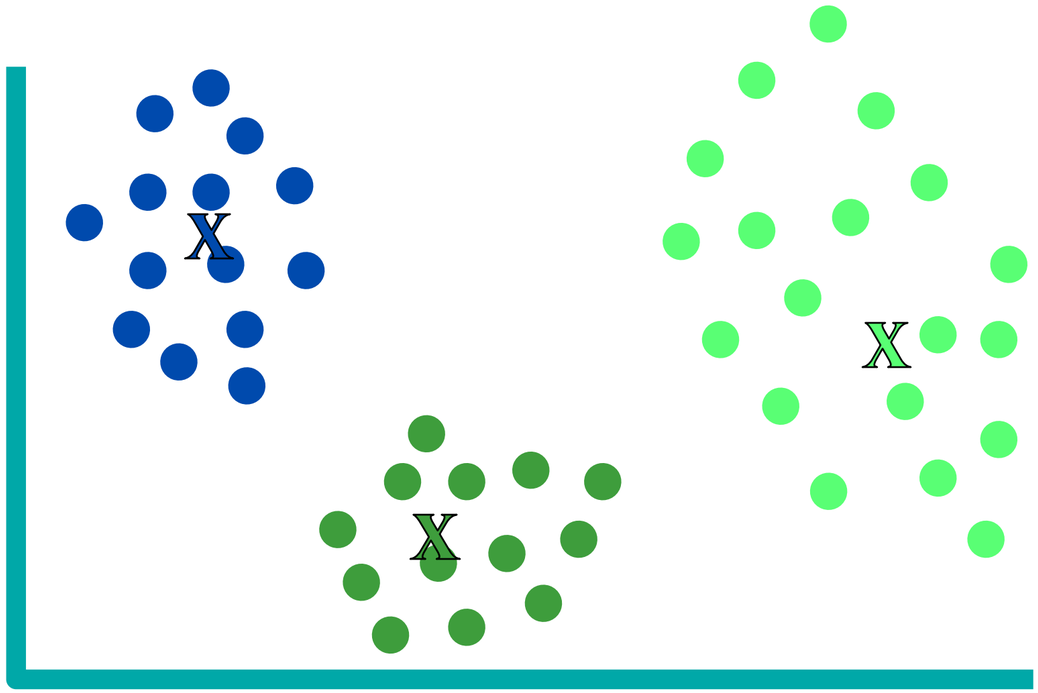
Repeat Step 4 and 5



Final Clusters



When to Stop?

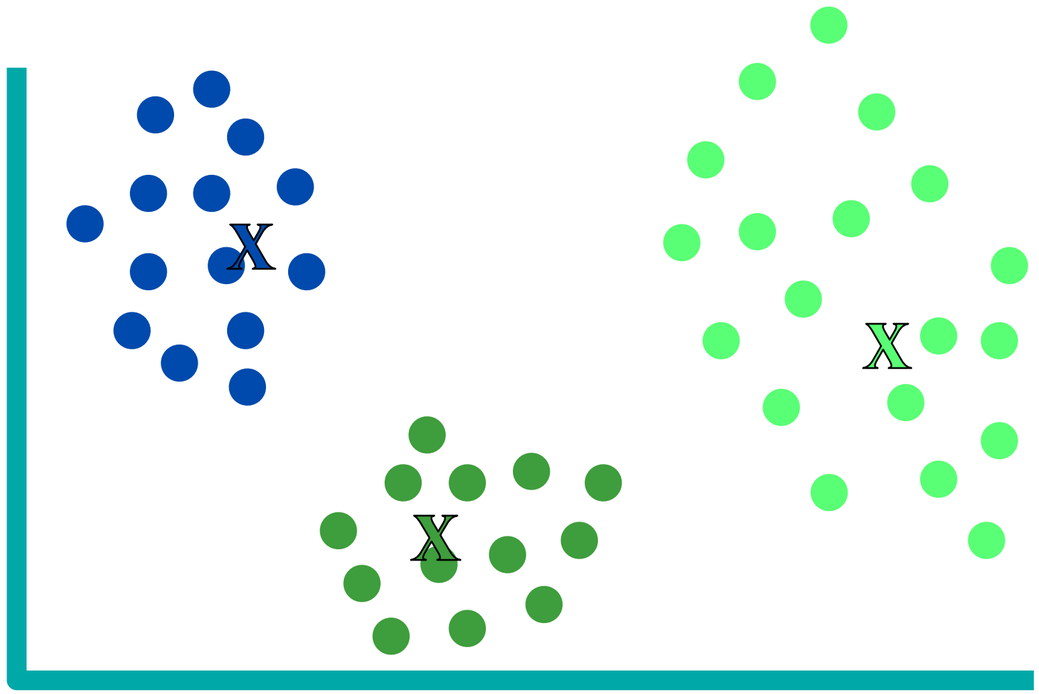


- When the change in the cluster centroids in close to zero
- After a limited number of iterations

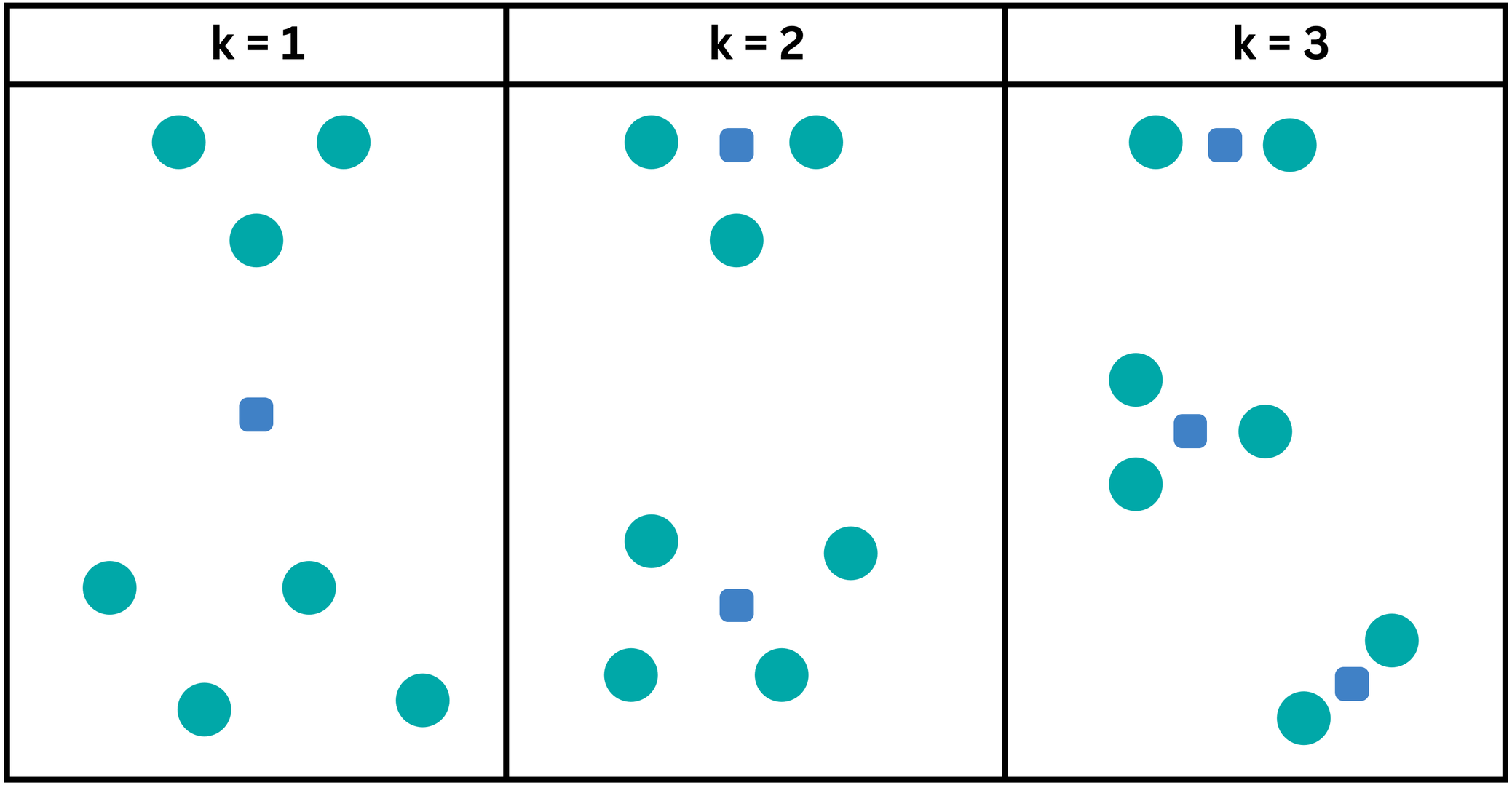
Optimal Cluster Number

2 Cluster

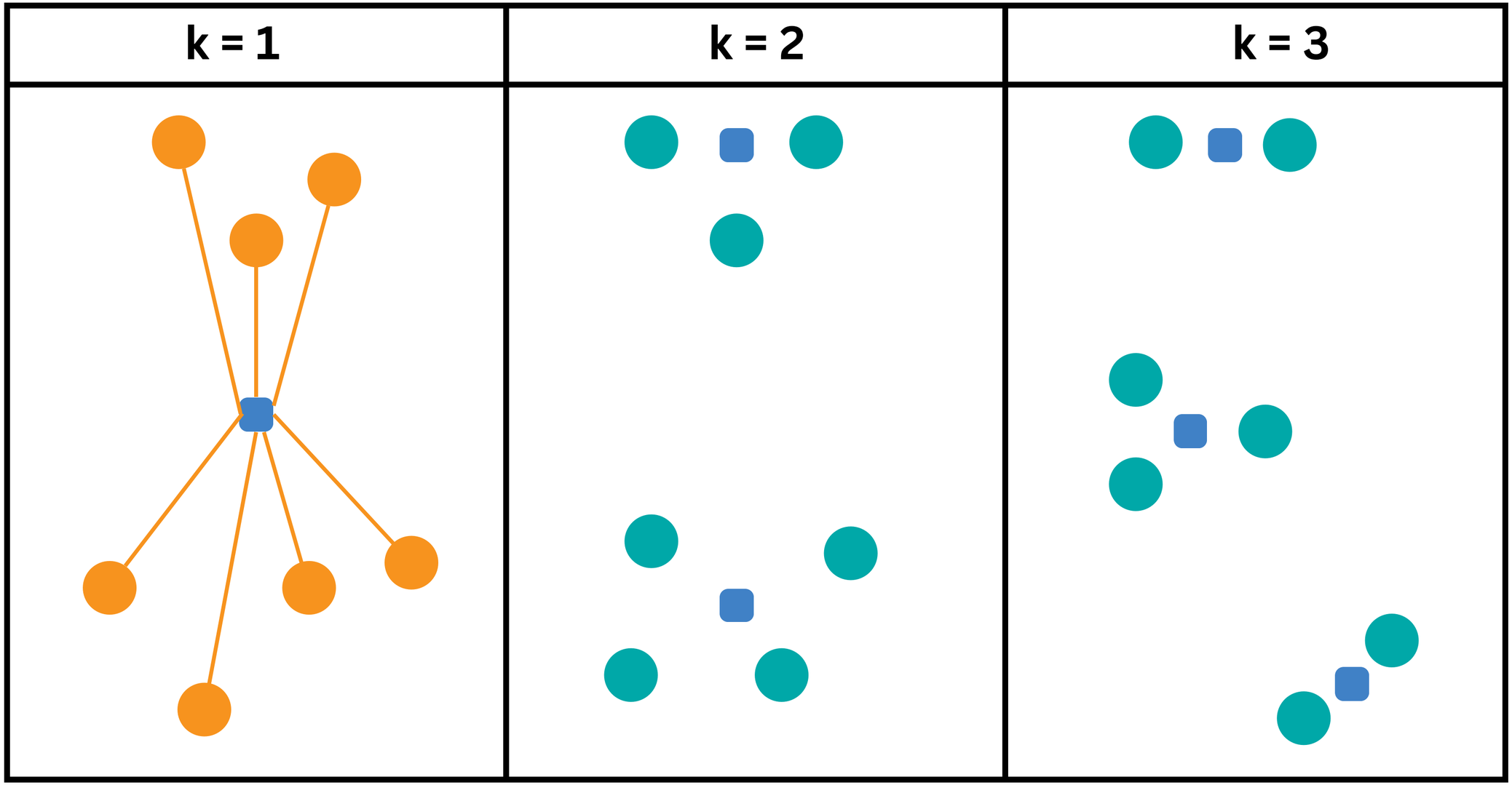
3 Cluster



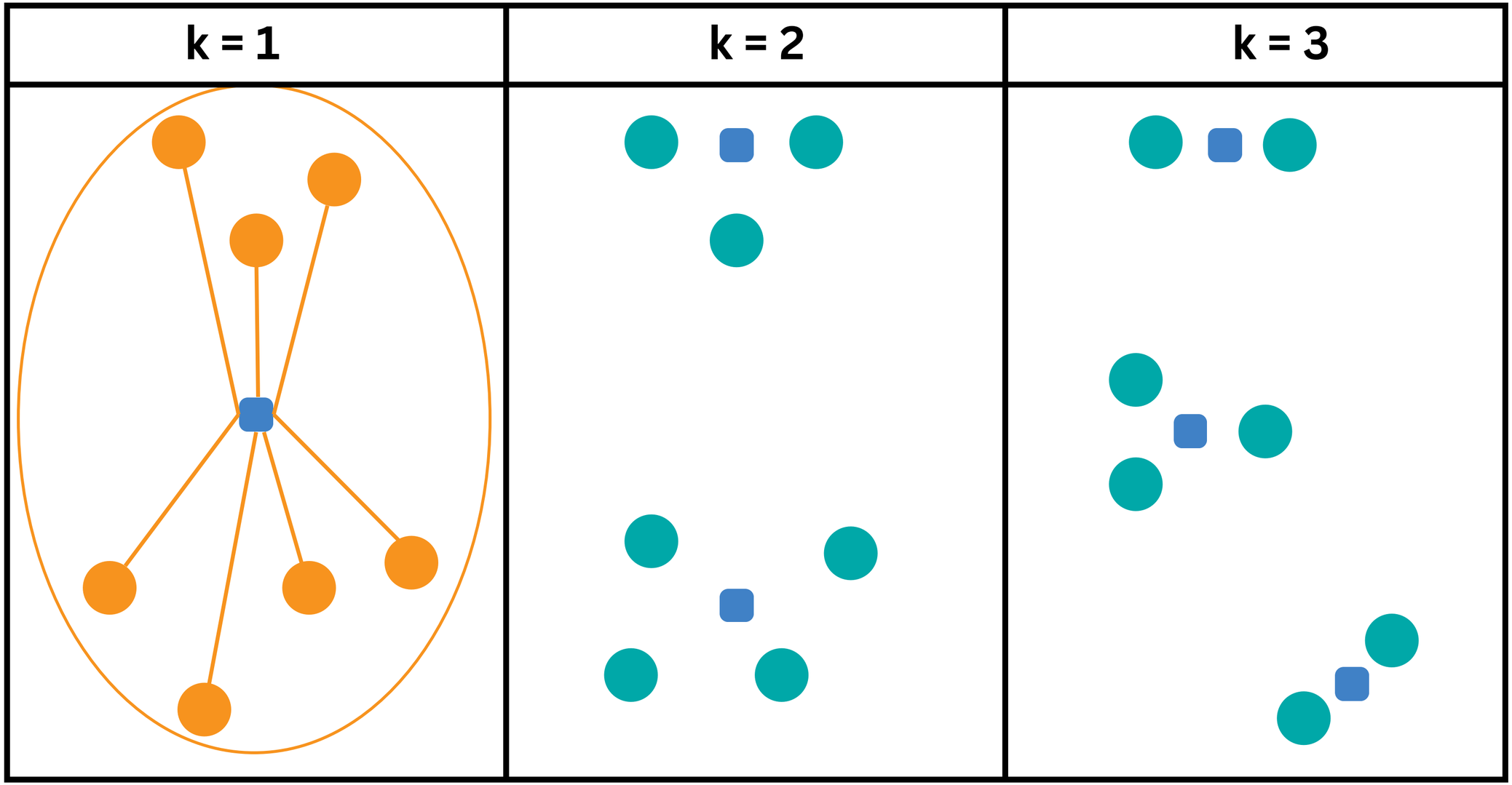
Within Sum of Squares (WSS)



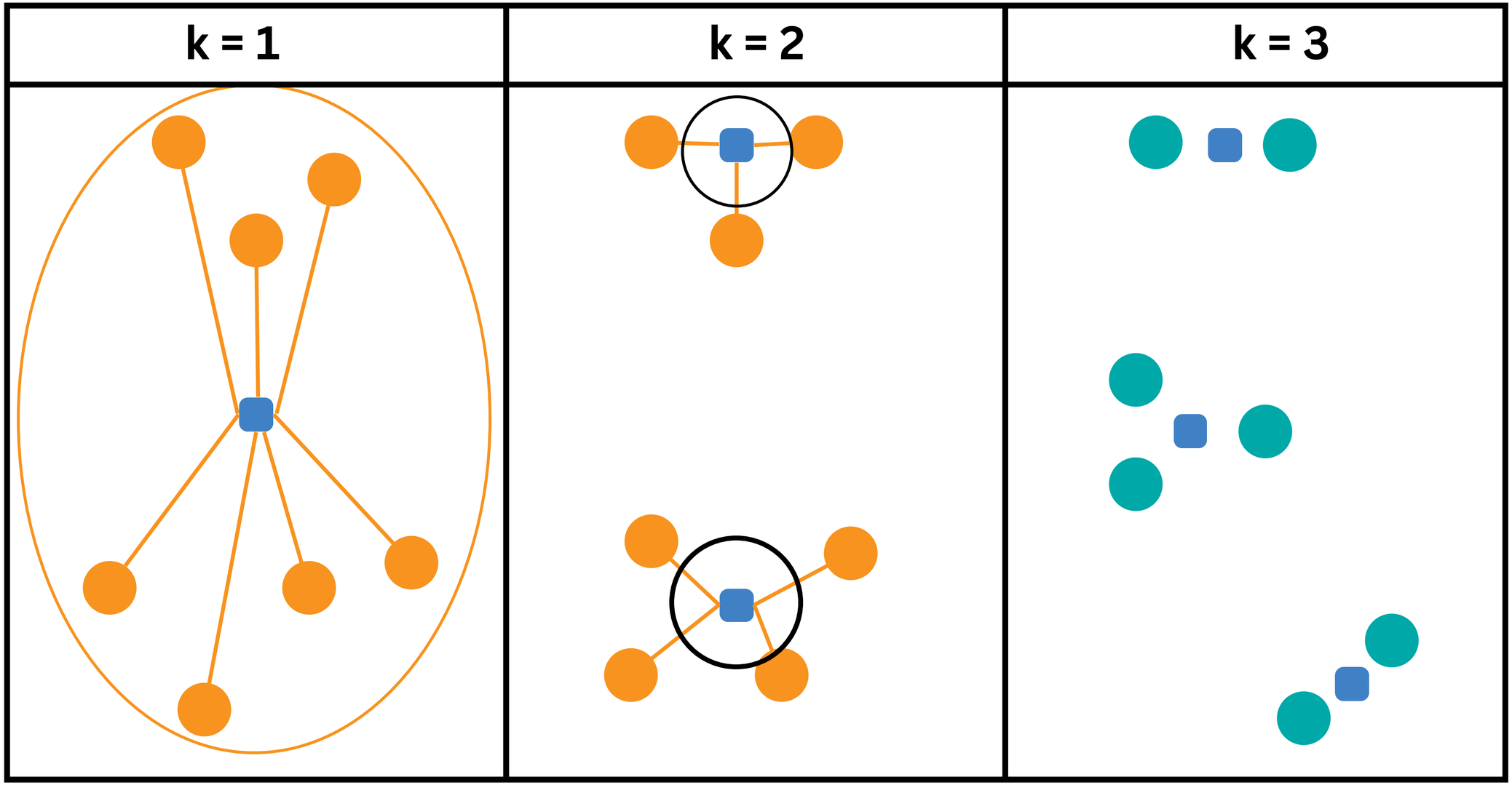
Within Sum of Squares



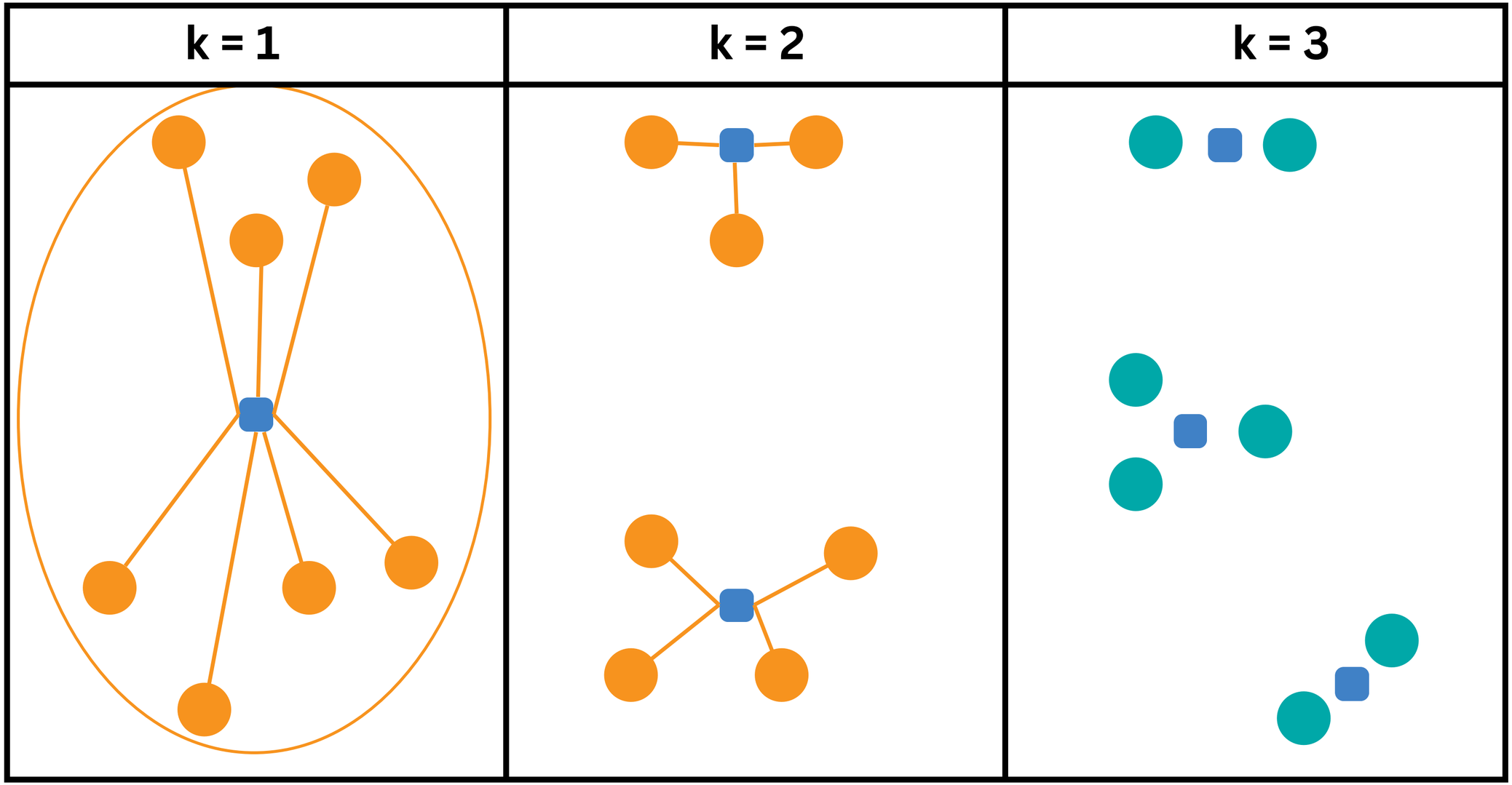
Within Sum of Squares



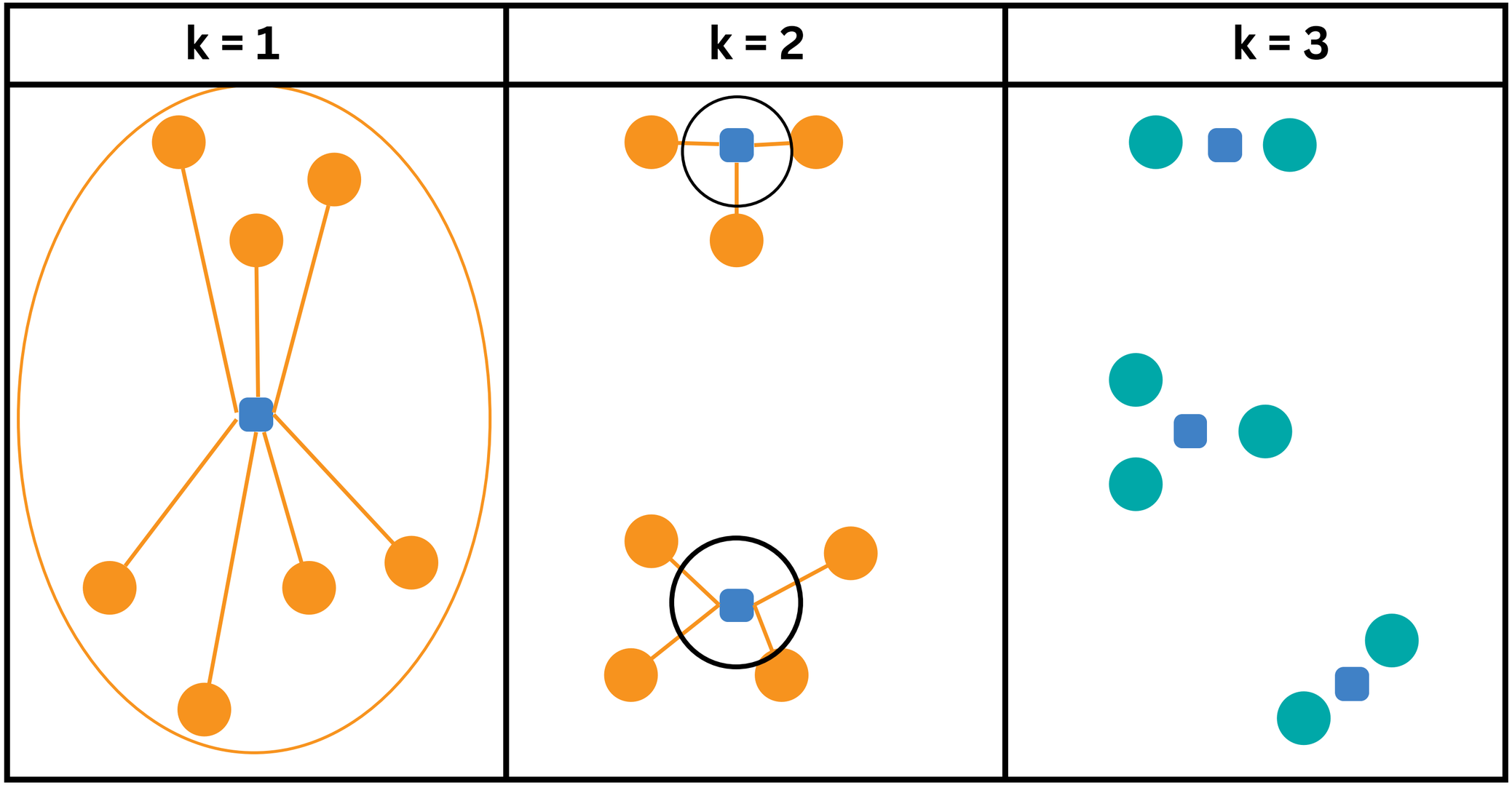
Within Sum of Squares (WSS)



Within Sum of Squares (WSS)

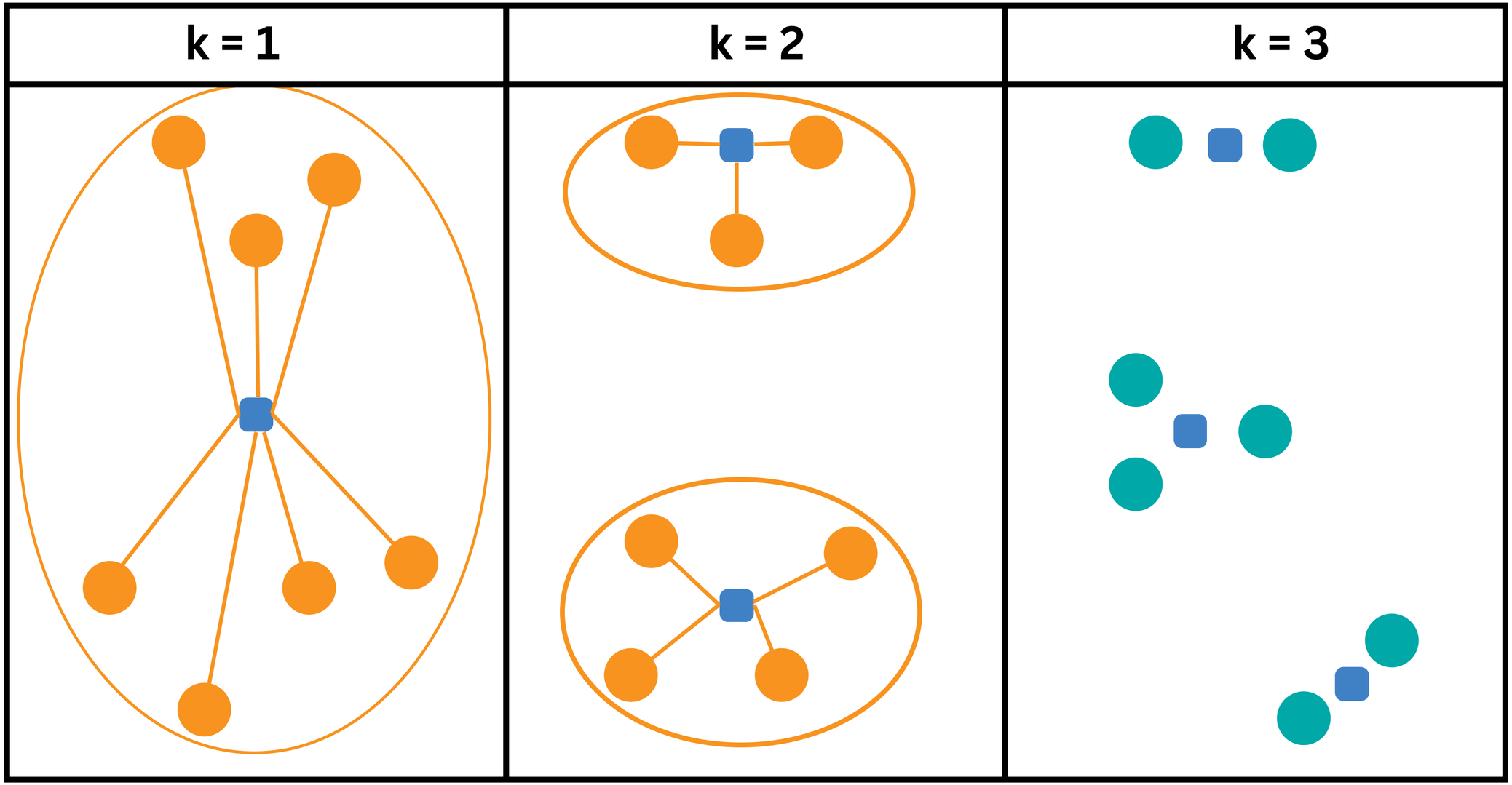


Within Sum of Squares (WSS)

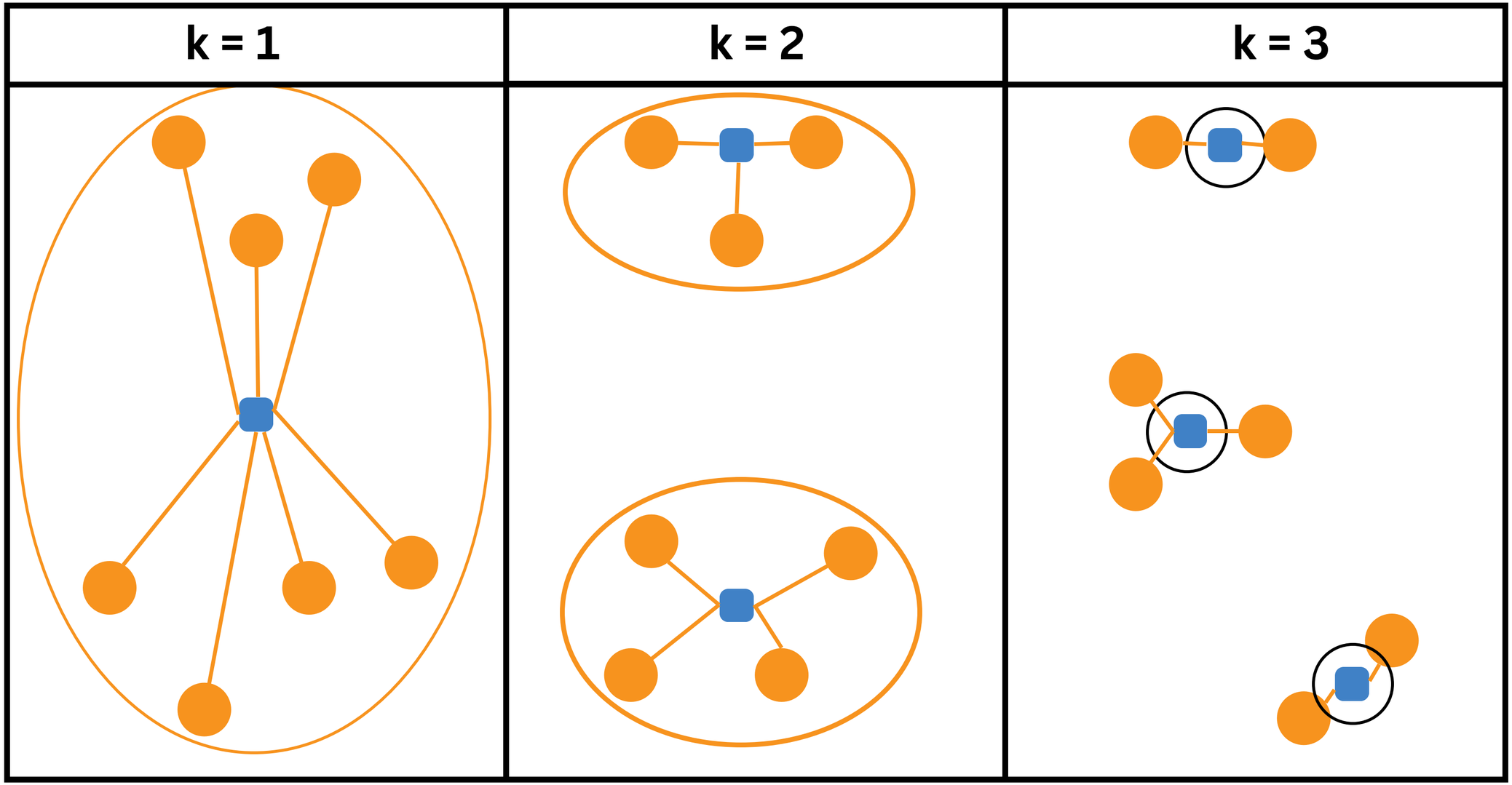


codefinity

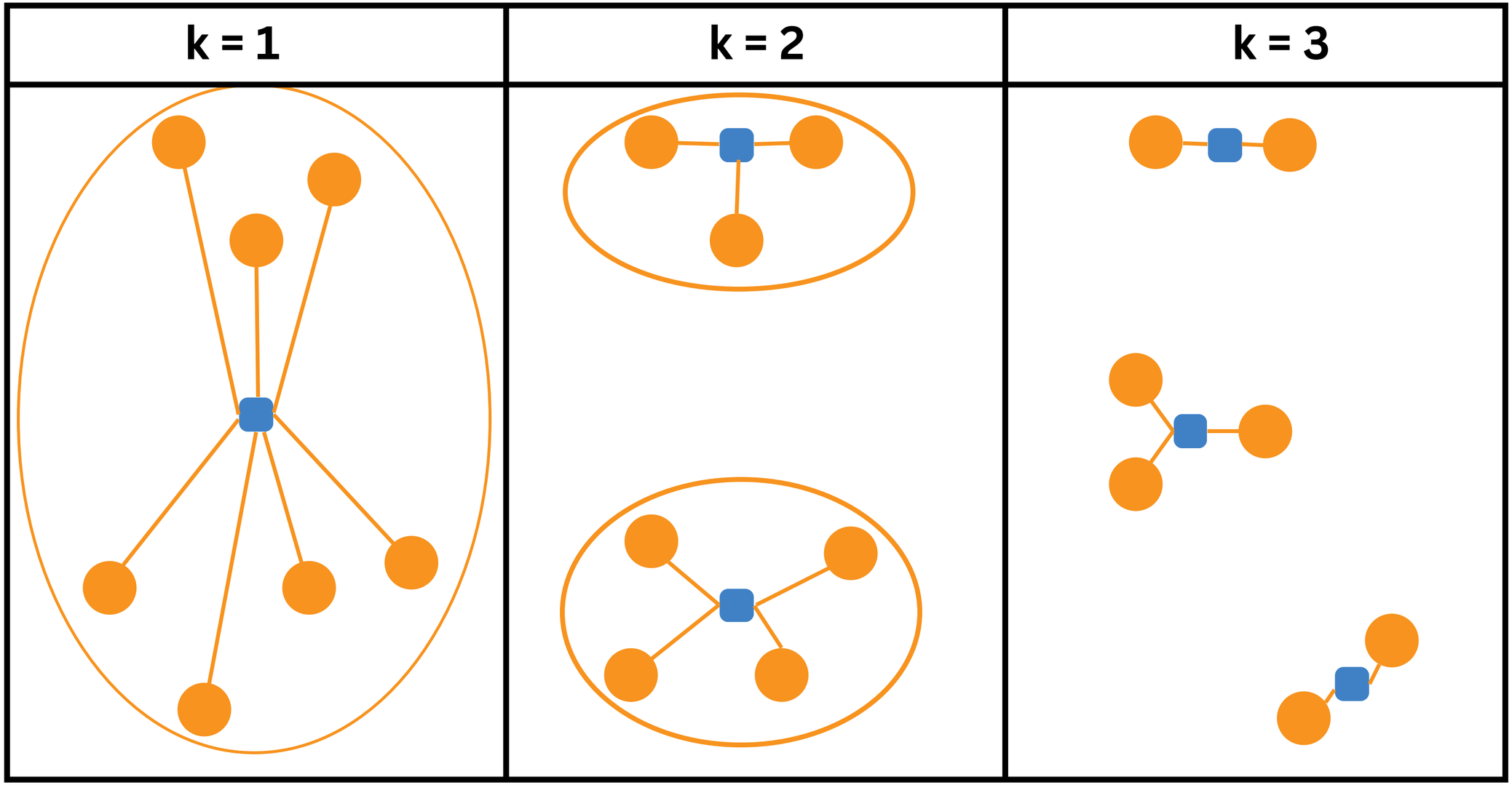
Within Sum of Squares (WSS)



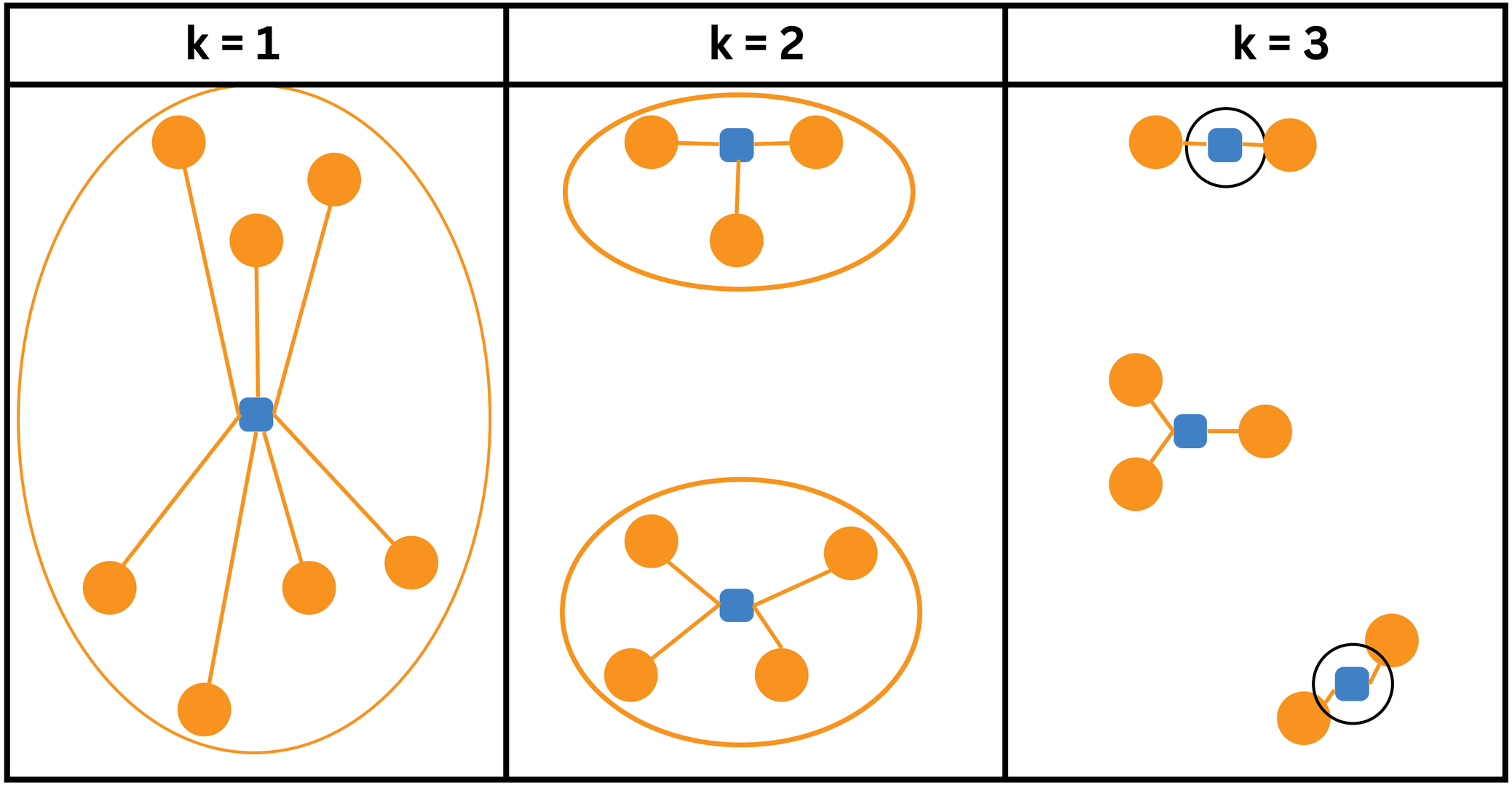
Within Sum of Squares (WSS)



Within Sum of Squares (WSS)

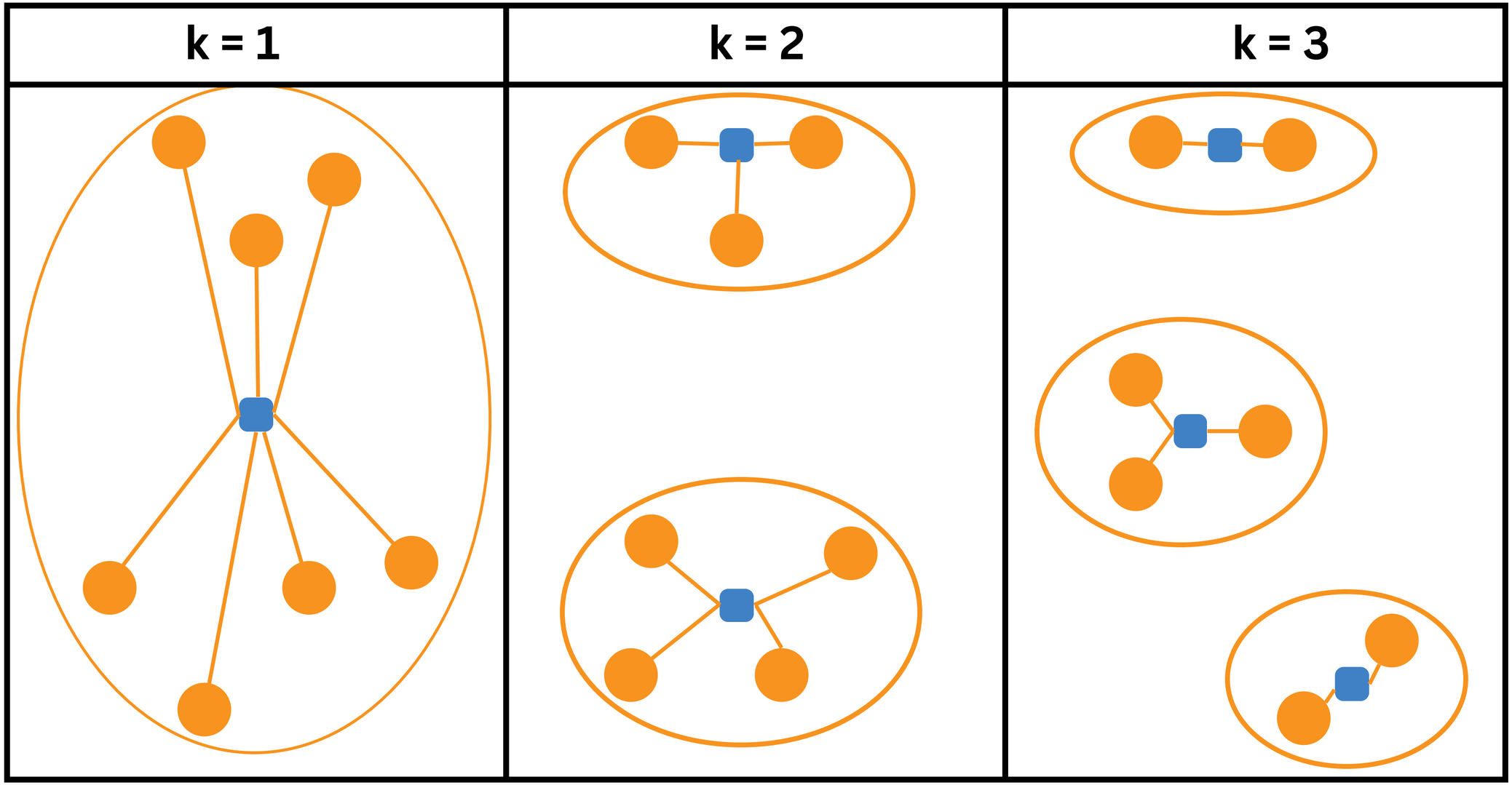


Within Sum of Squares (WSS)



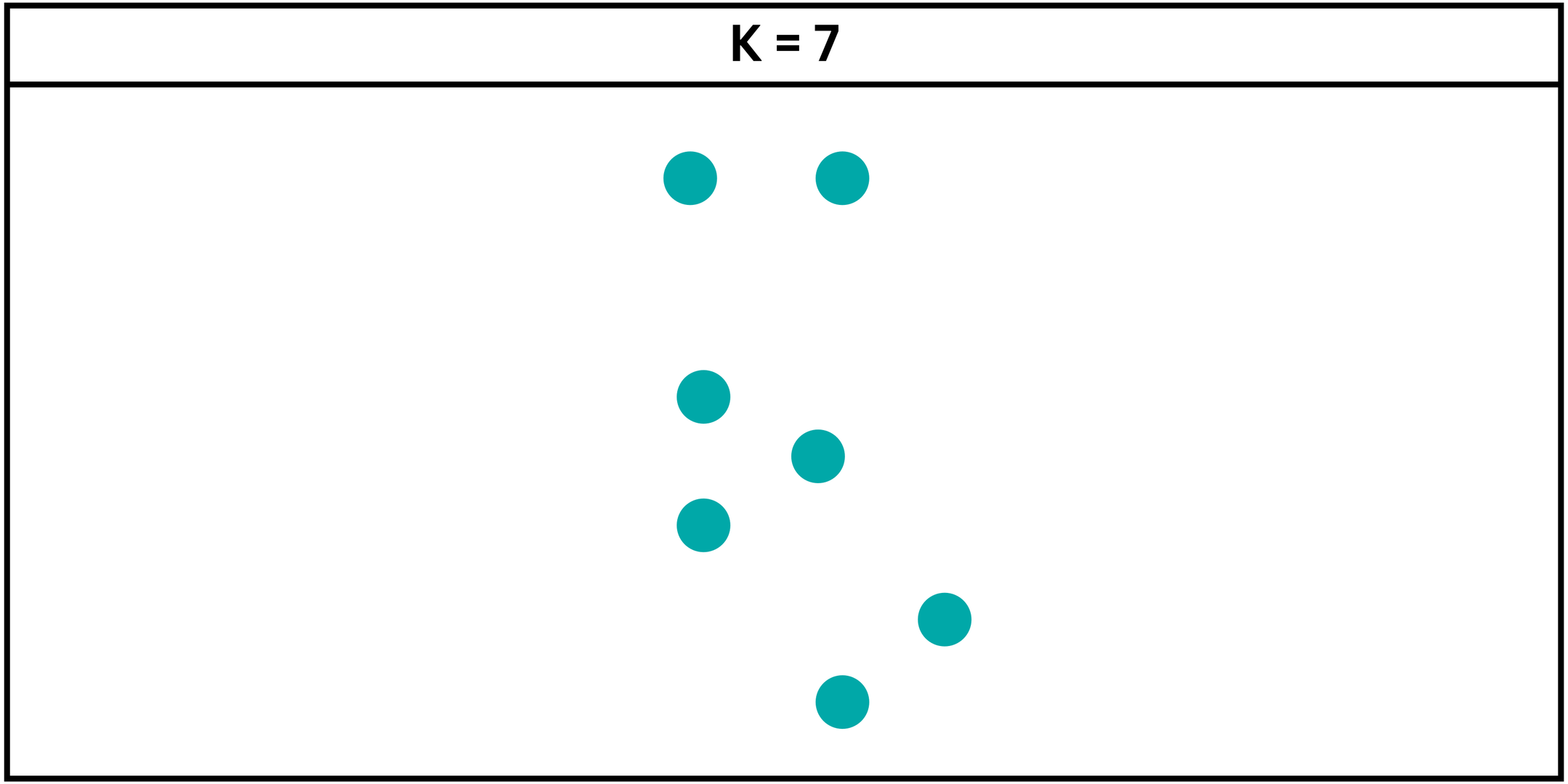
codefinity

Within Sum of Squares (WSS)



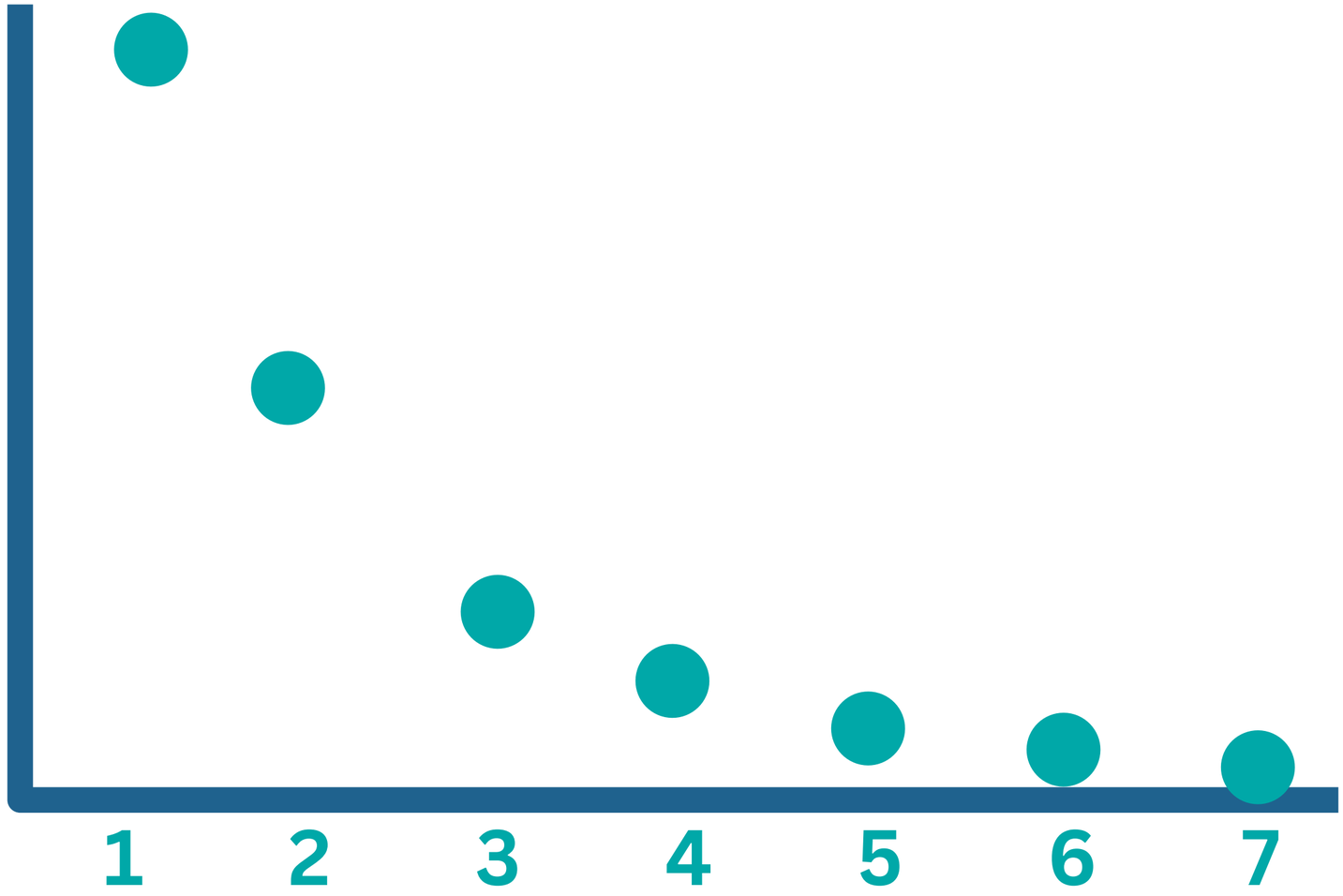
codefinity

Within Sum of Squares (WSS)



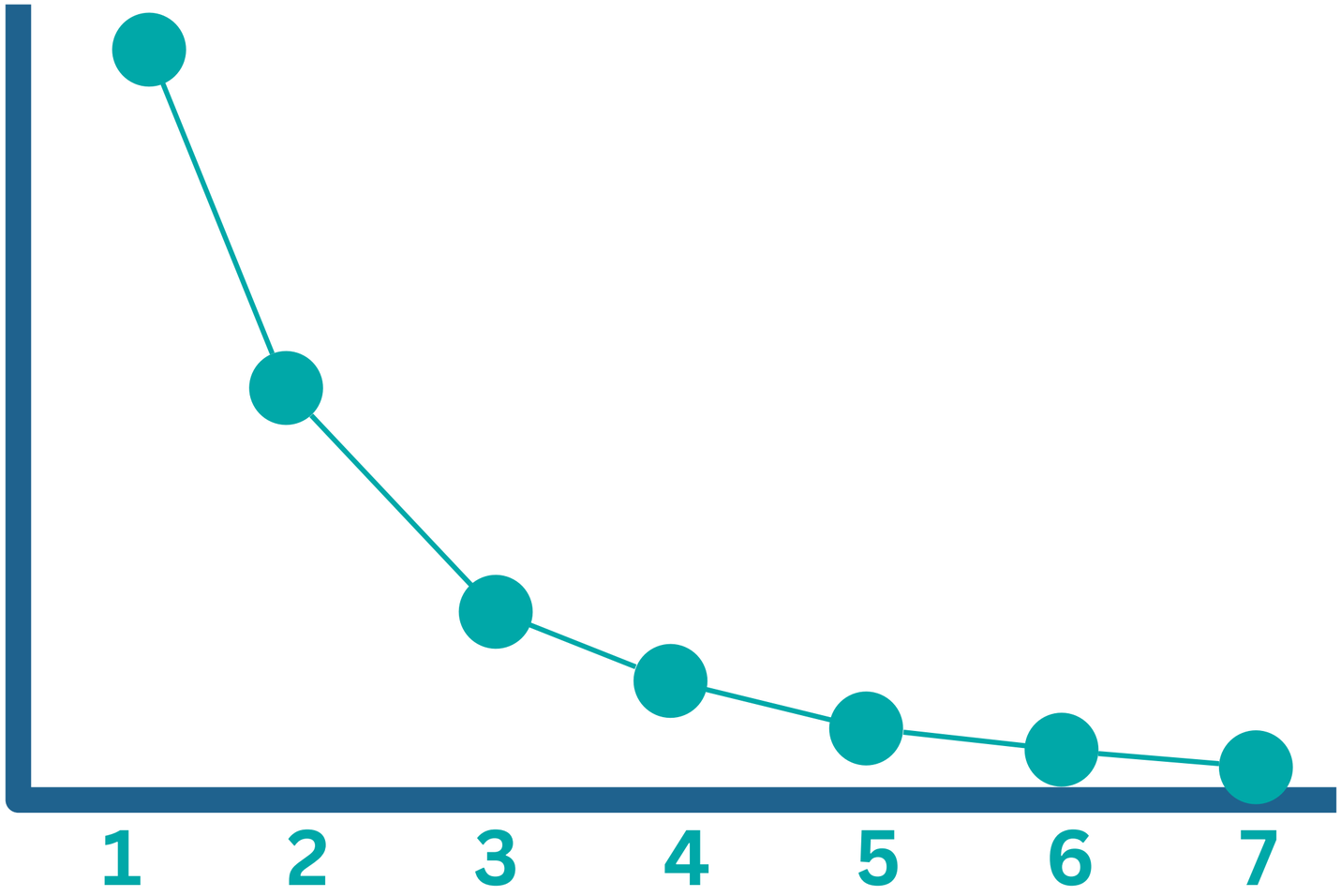
codefinity

WSS Plot



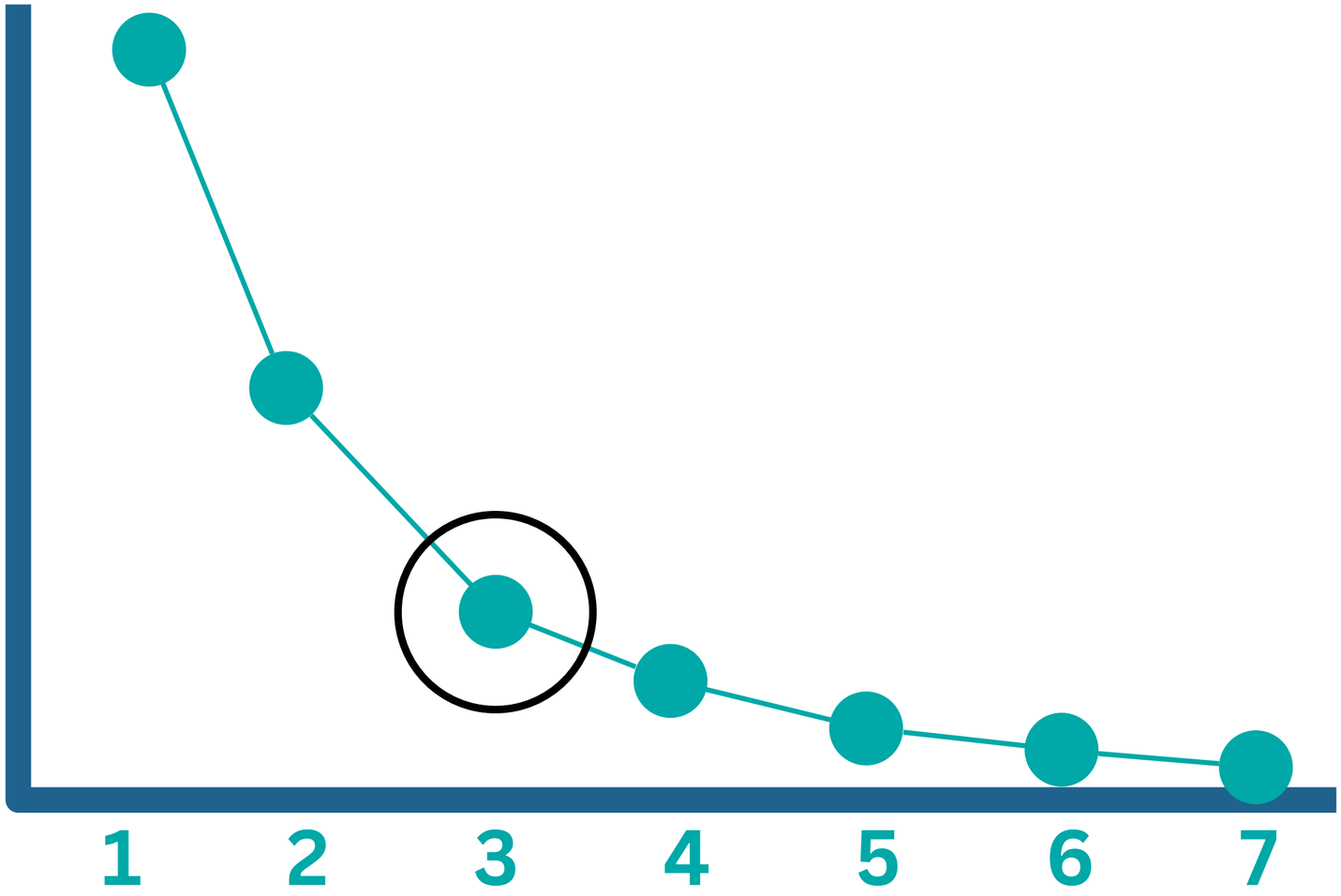
codefinity

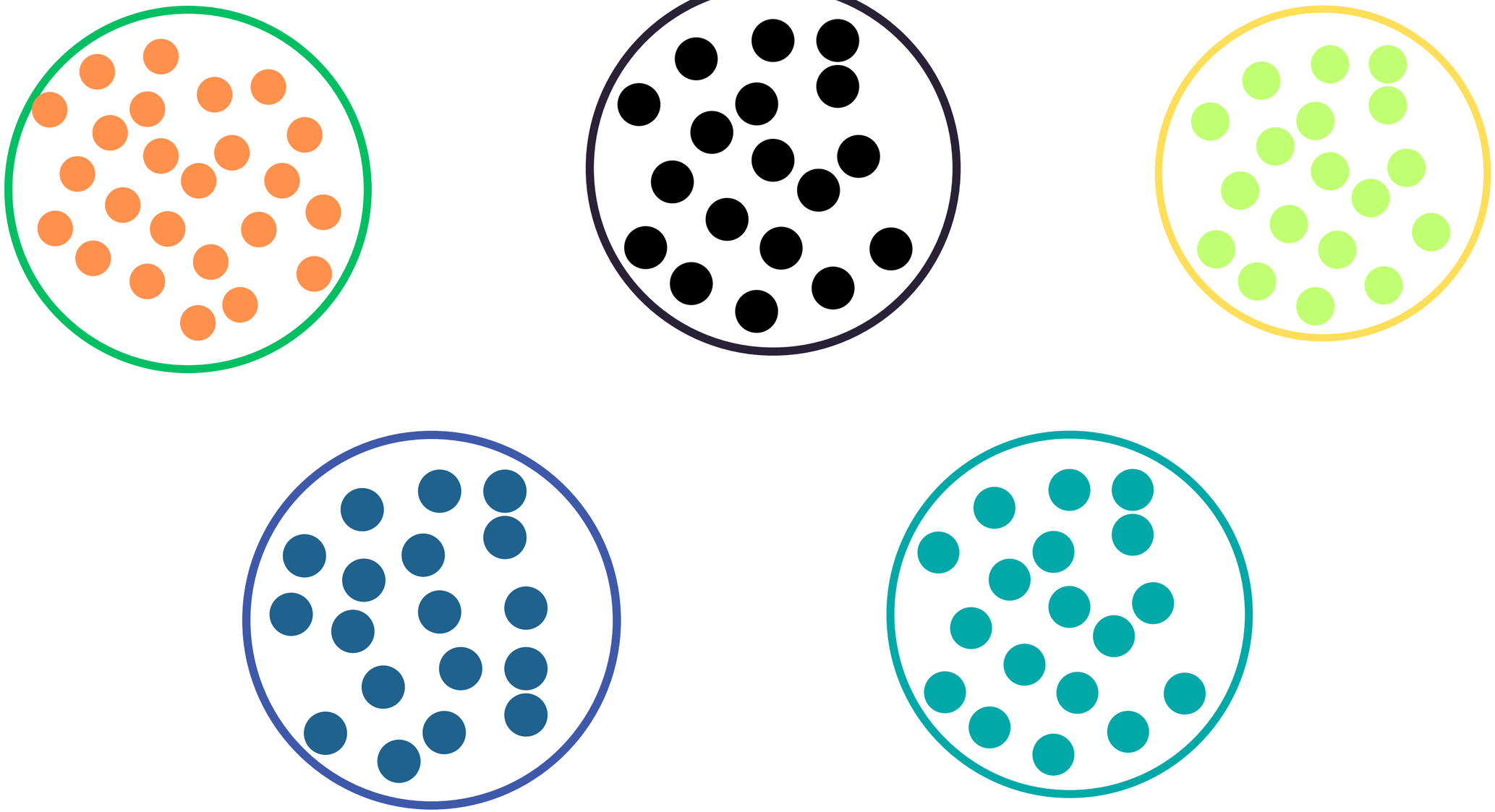
WSS Plot



codefinity

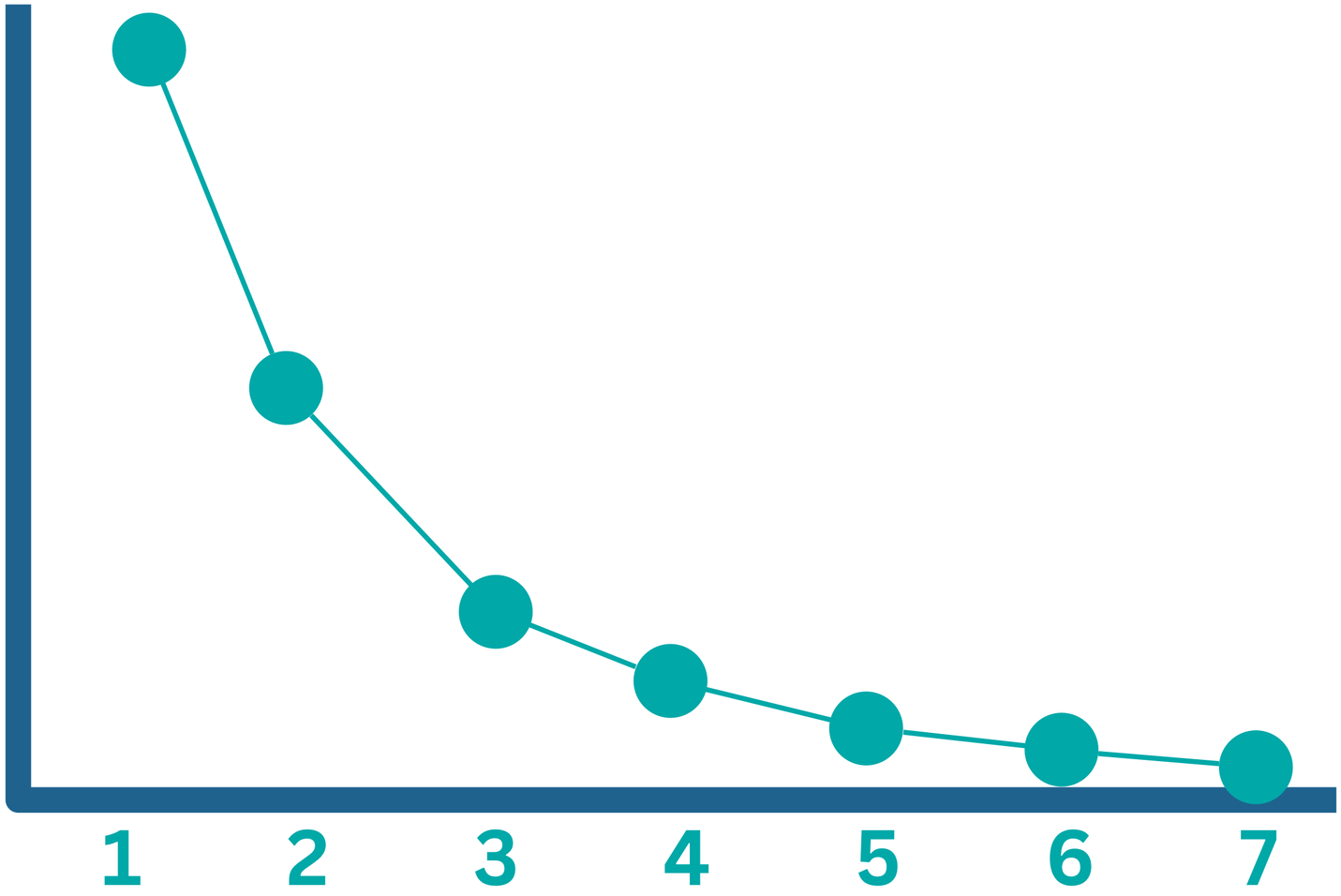
WSS Plot





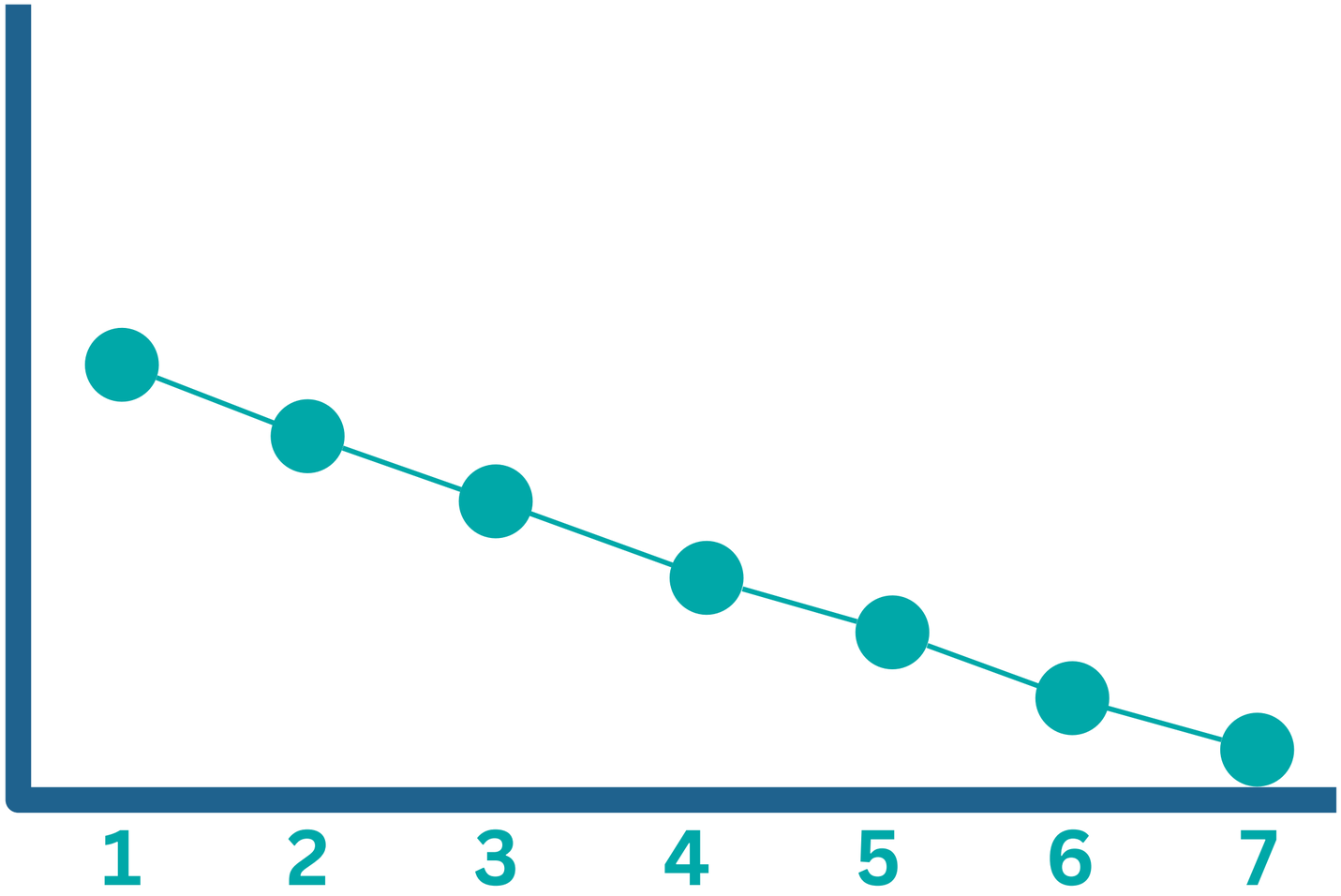
codefinity

WSS Plot

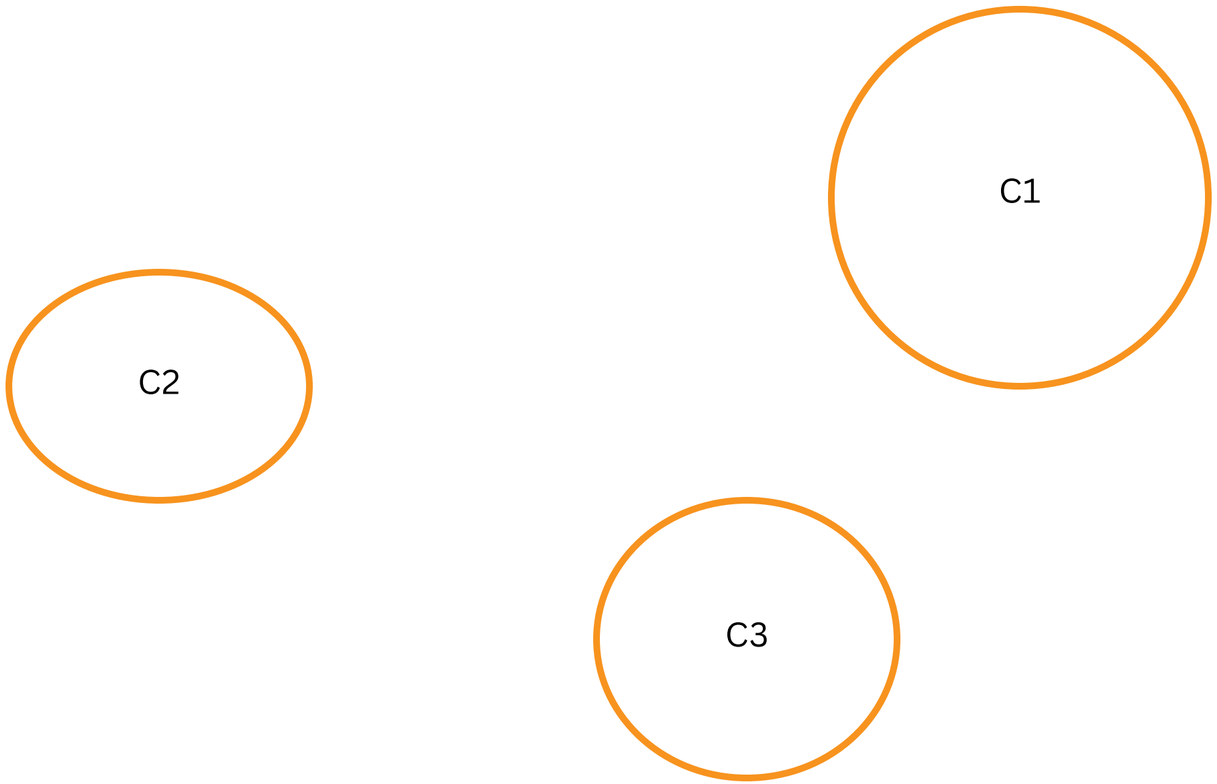


codefinity

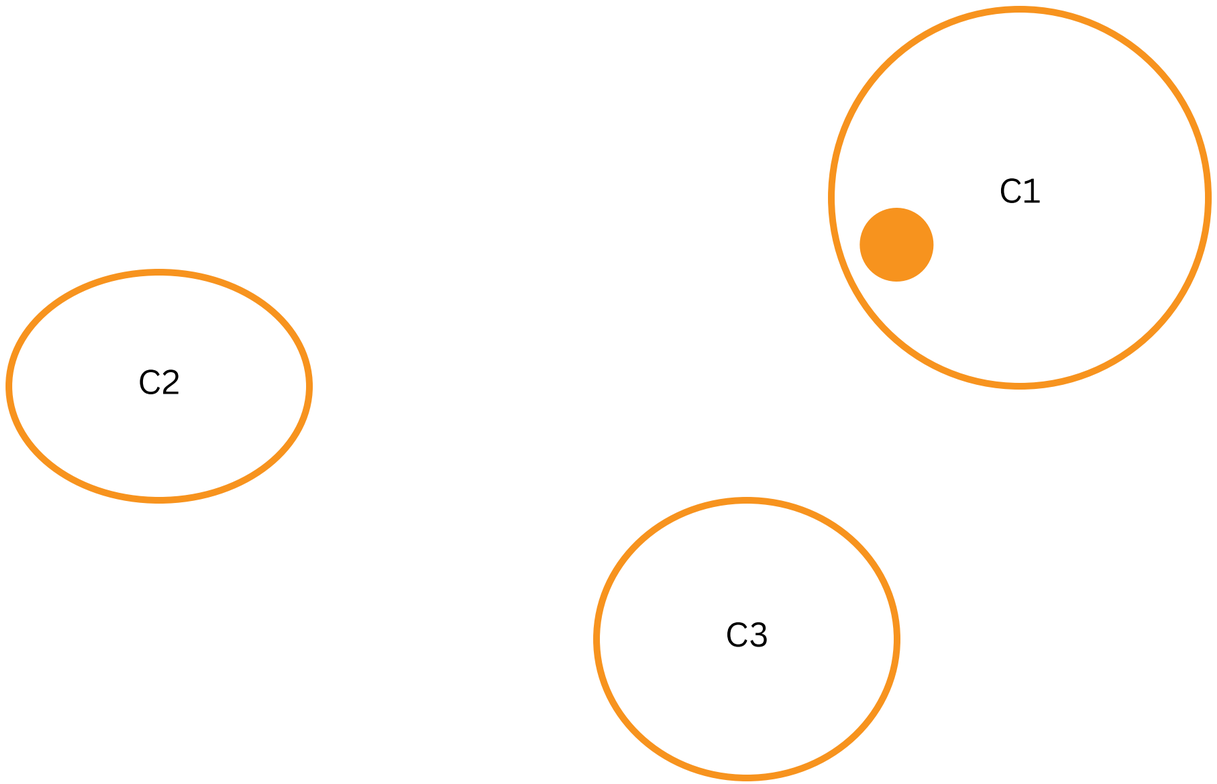
WSS Plot



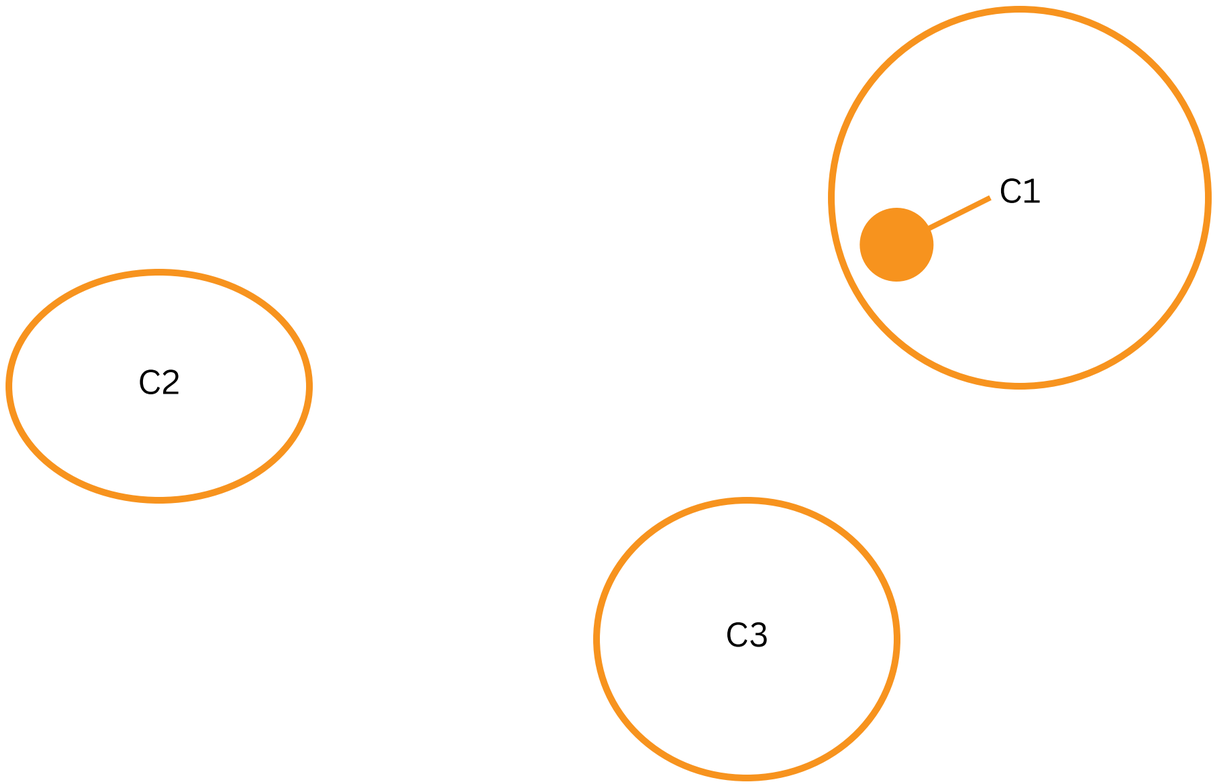
Silhouette Width



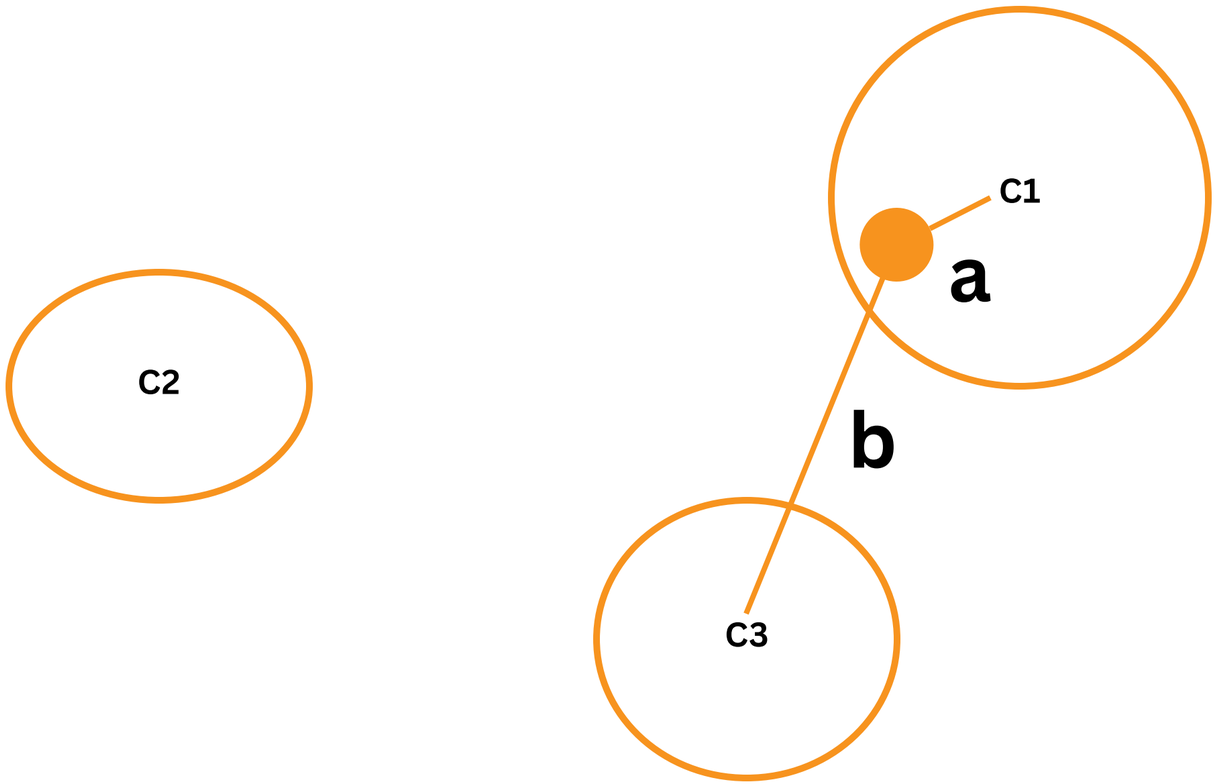
Silhouette Width



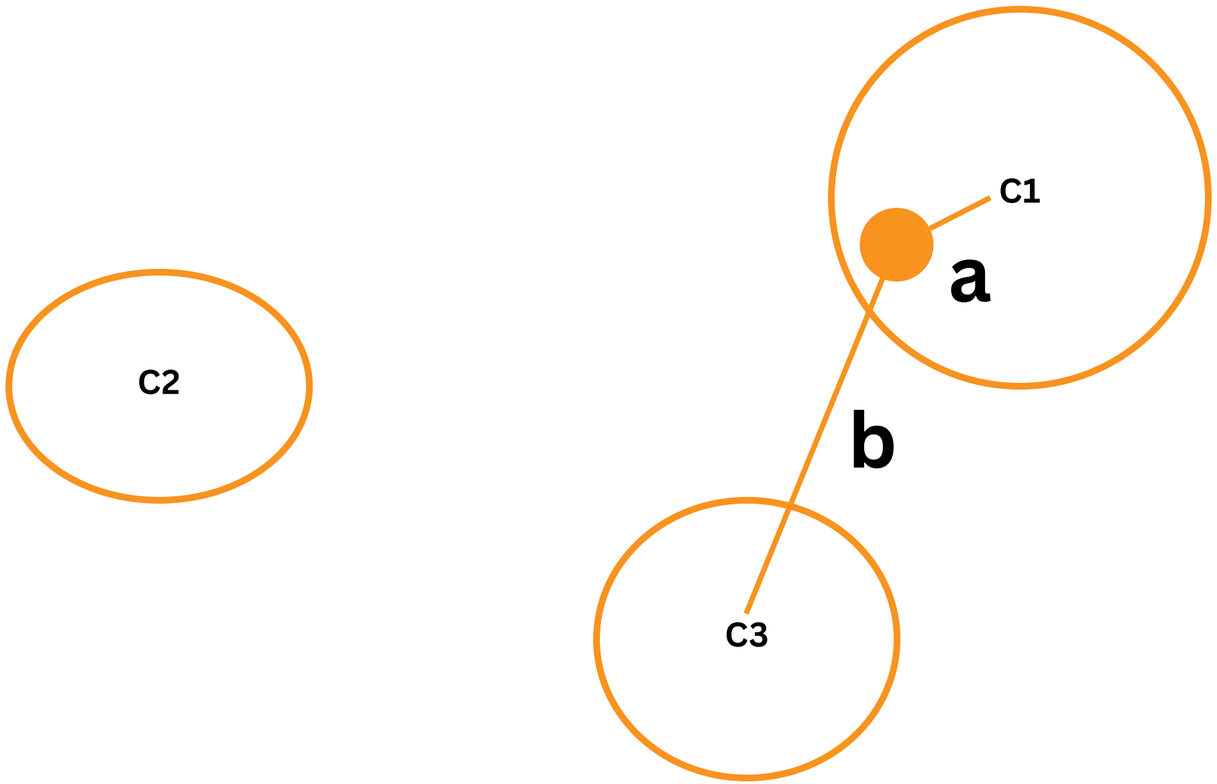
Silhouette Width



Silhouette Score



Calculating Silhouette Score



Score = b-a/max(a,b)

Range = [-1, 1]