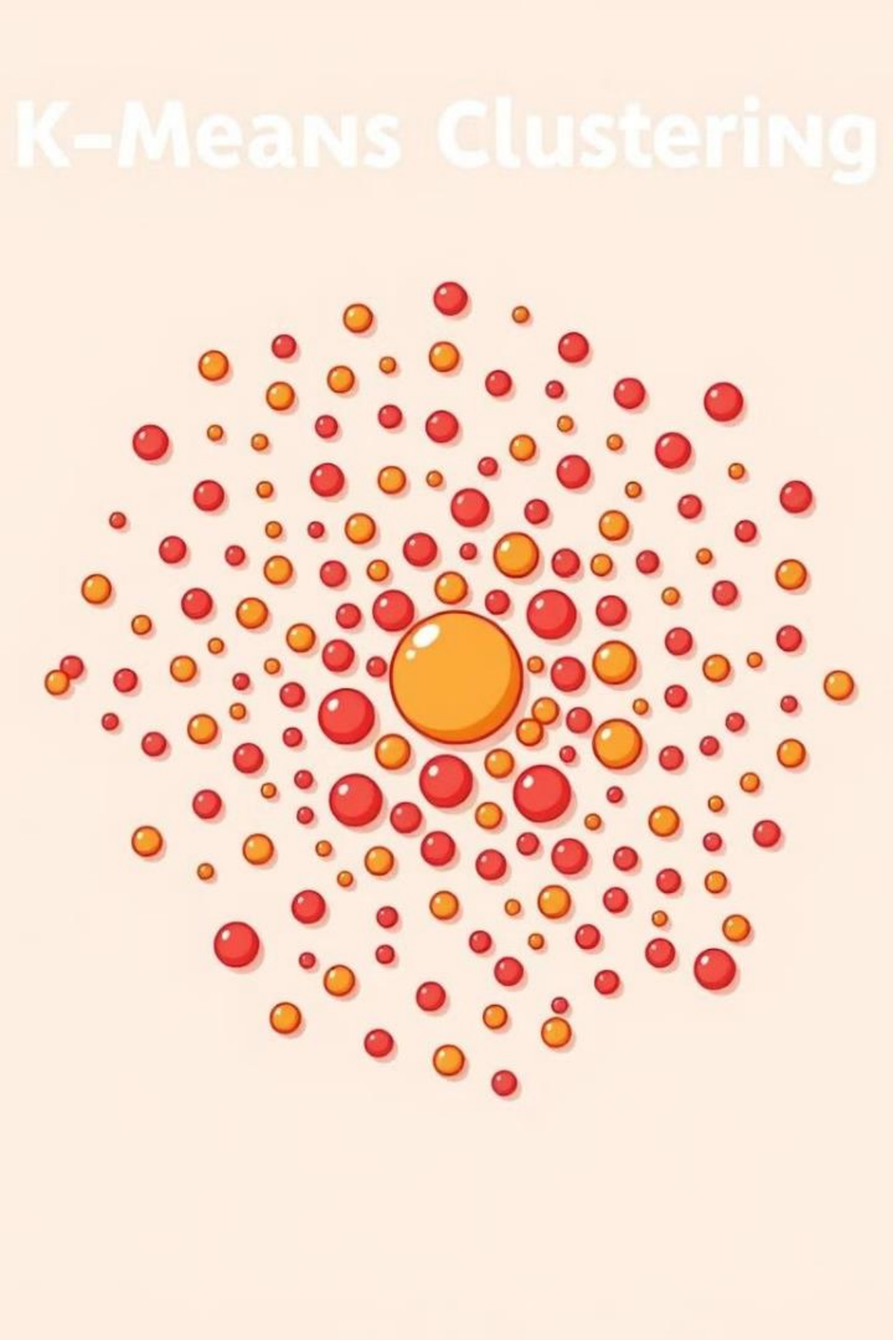



# K-means Clustering for Unsupervised Learning

Welcome to this presentation on K-means Clustering, a powerful technique for unsupervised learning.

**by Nisha A K**

# Introduction to Unsupervised Learning

## Unlabeled Data

Unsupervised learning deals with data that lacks pre-defined labels or categories.

## Pattern Discovery

The goal is to uncover hidden patterns, structures, and relationships within the data.

# The Need for Clustering Algorithms

### Organize Chaos

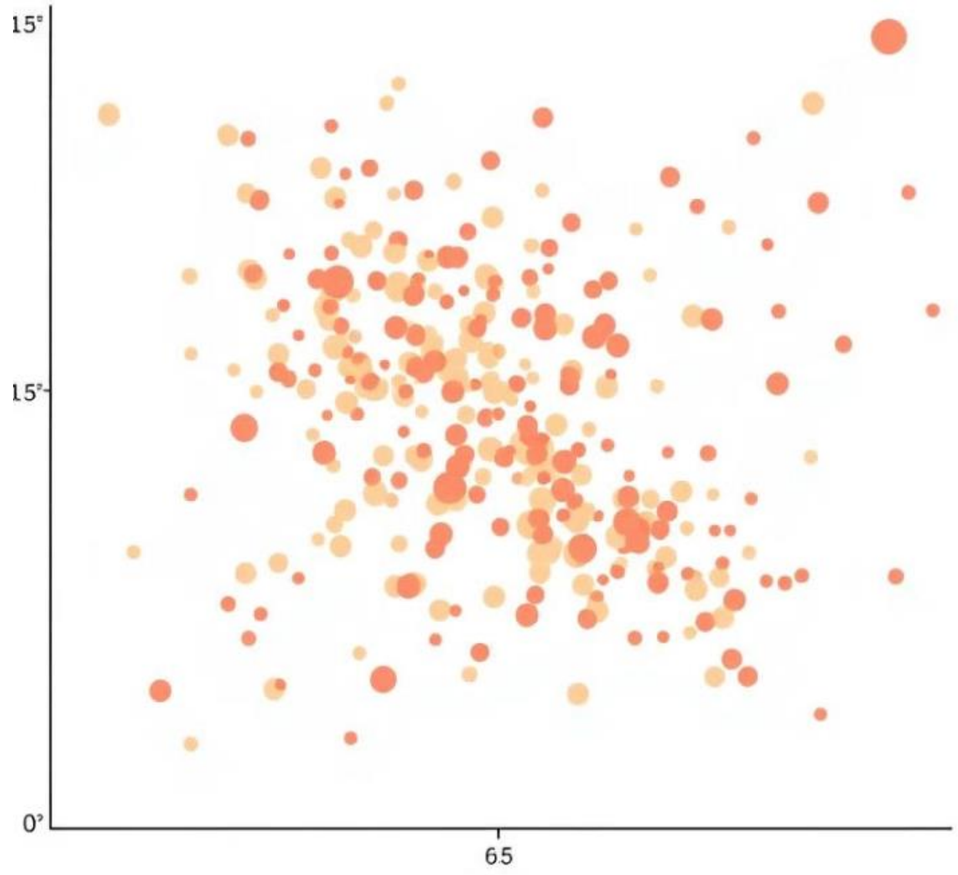
Clustering algorithms help to group similar data points together.

### Identify Groups

They provide insights into natural groupings or segments within data.

### Simplify Complexity

Clustering helps to make sense of complex datasets by identifying meaningful patterns.



# Understanding the K-means Algorithm

### Iterative Process

K-means involves an iterative process of assigning data points to clusters and updating cluster centroids.

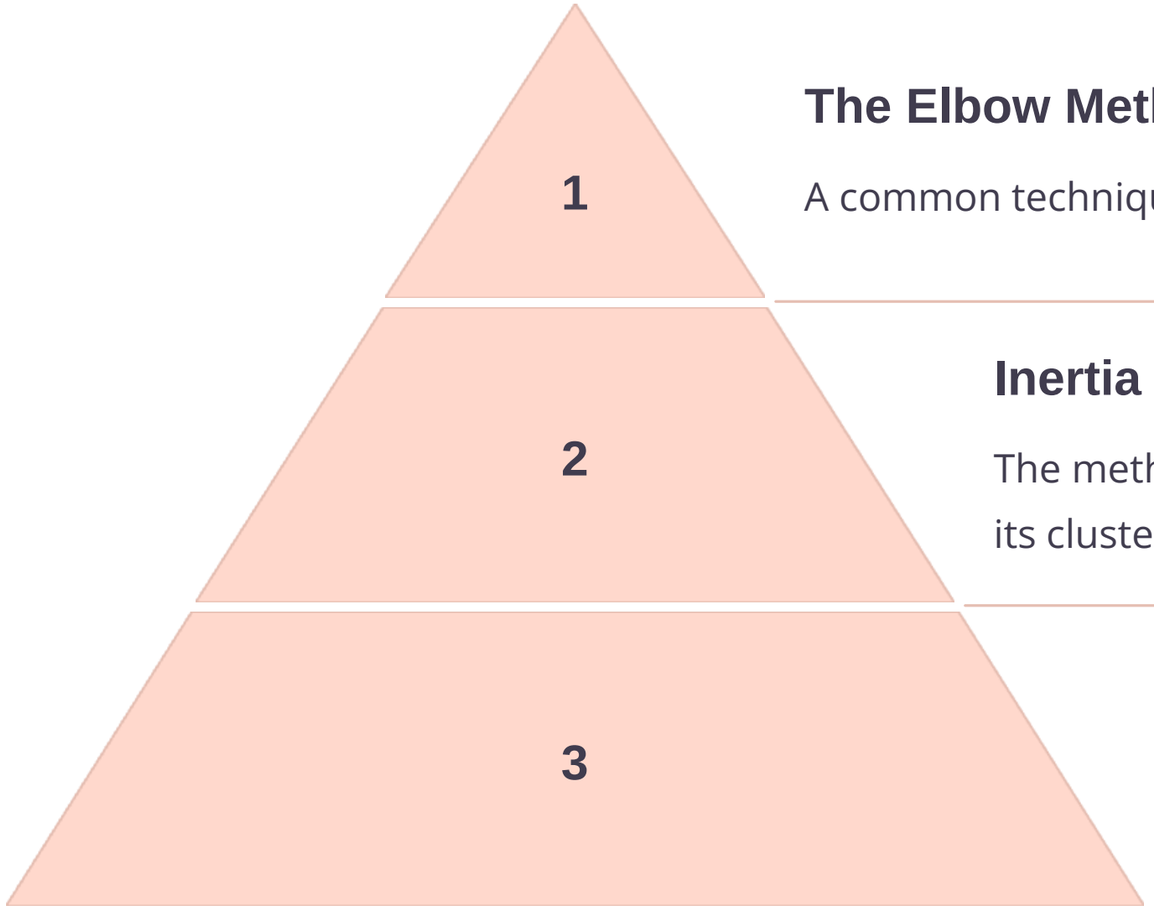
### Centroid-Based

The algorithm uses cluster centroids, representing the center of each cluster, to guide the assignment process.

### Distance Minimization

Data points are assigned to the cluster with the closest centroid, minimizing the overall distance between points and their cluster centers.

# Choosing the Optimal Number of Clusters (K)




### The Elbow Method

A common technique to determine the optimal number of clusters.

### Inertia

The method measures the sum of squared distances of each point to its cluster centroid.

### Optimal K

The "elbow" point in the graph represents the ideal number of clusters.

# Initializing Cluster Centroids

1

### Random Initialization

Centroids are randomly chosen from the dataset.

2

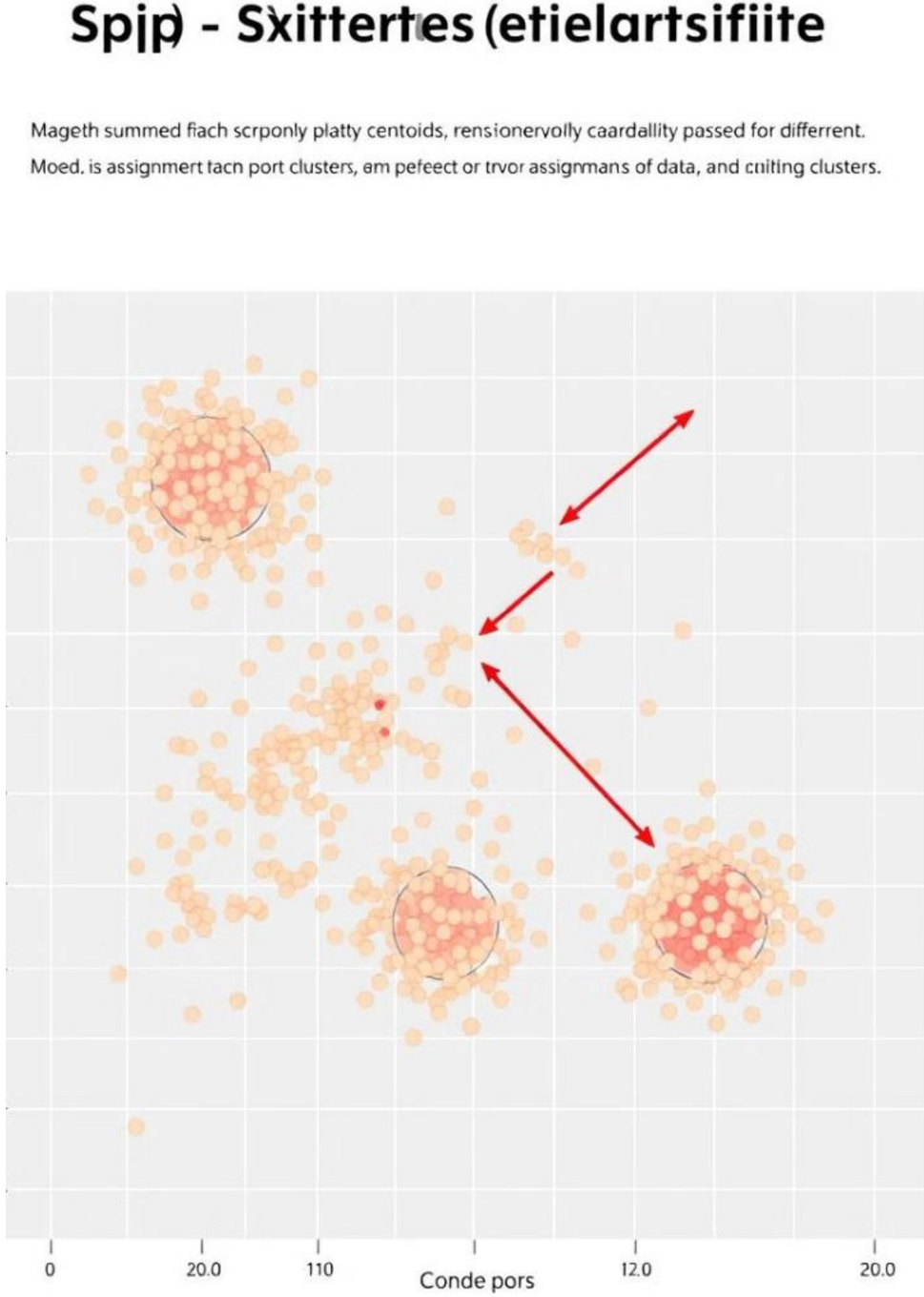
### K-means++

A more sophisticated initialization method, designed to improve convergence.

3

### Initial Assignment

Data points are initially assigned to the closest centroid.



# Iterative Optimization of Cluster Assignments

1

## Recalculating Centroids

Centroids are recalculated as the average of all data points assigned to each cluster.

2

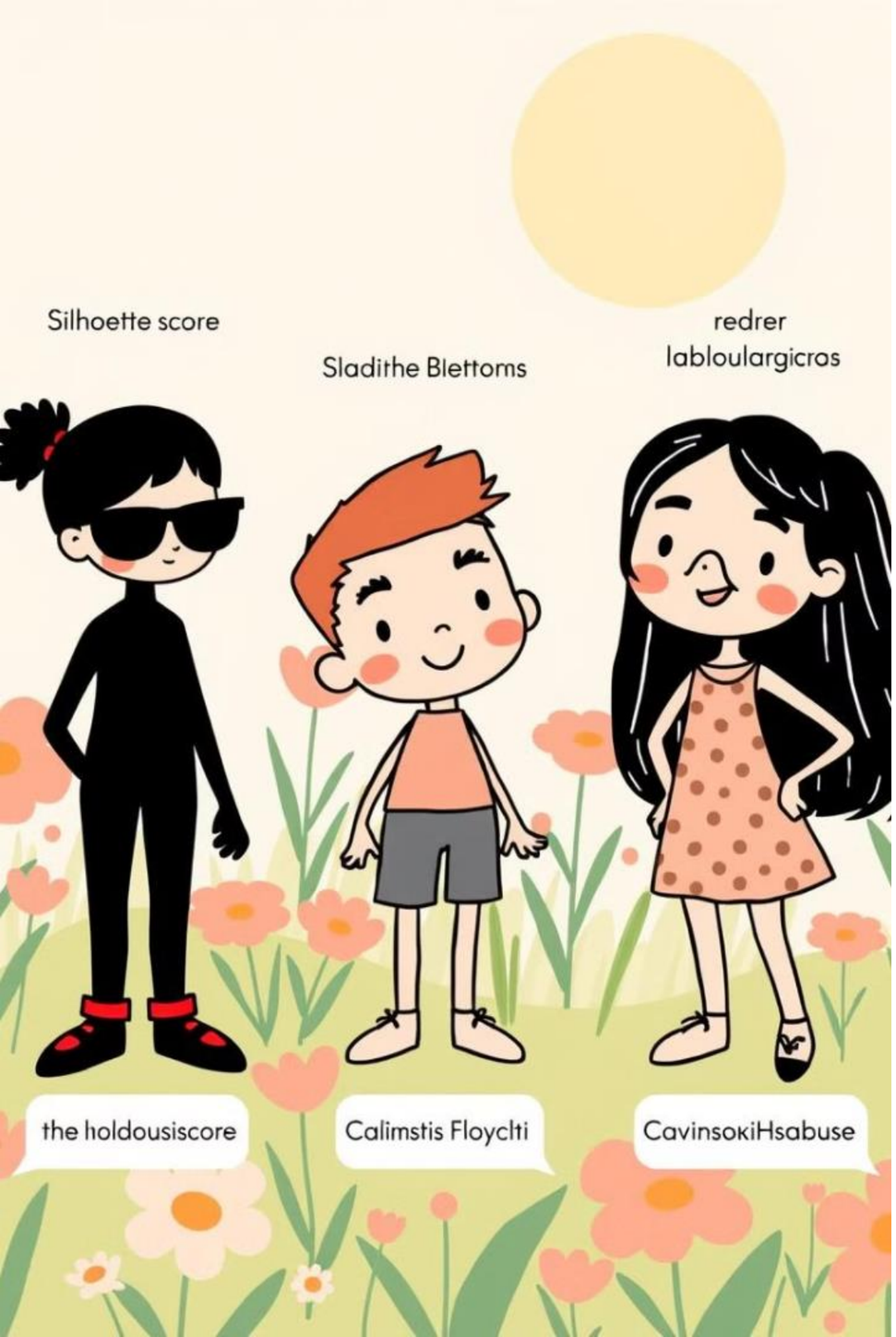
## Reassigning Points

Data points are reassigned to the closest centroid based on the updated centroids.

3

## Convergence

The algorithm continues iterating until cluster assignments stabilize, indicating convergence.




# Evaluating Cluster Quality

## Silhouette Score

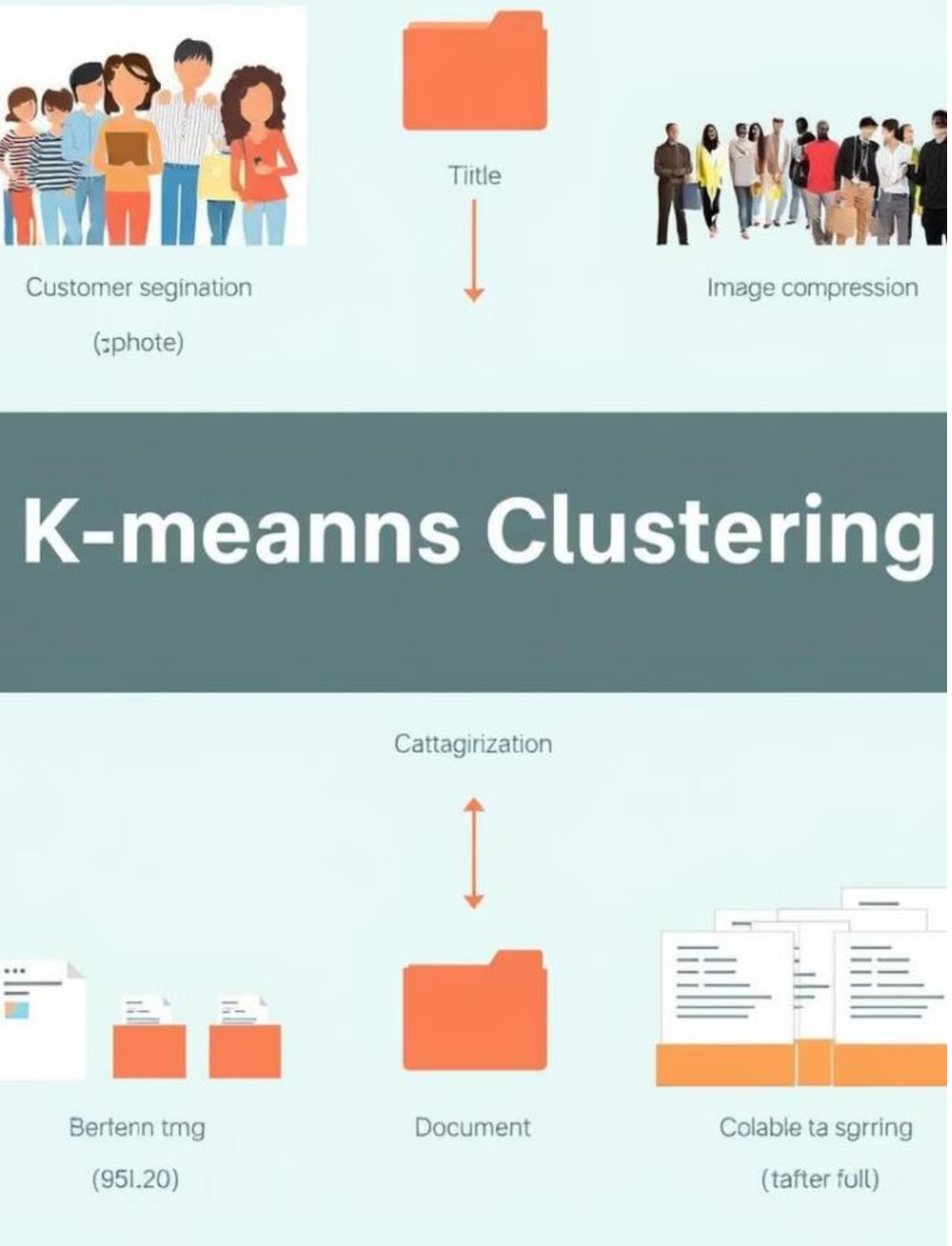
Measures how similar a data point is to its own cluster compared to other clusters.

## Davies-Bouldin Index

Evaluates the ratio of within-cluster distances to between-cluster distances.

## Calinski-Harabasz Index

Measures the ratio of between-cluster variance to within-cluster variance.




# Applications of K-means Clustering

## 1

### Customer Segmentation

Group customers into distinct segments based on their purchasing behavior.

## 2

### Image Compression

Reduce the number of colors in an image, resulting in a smaller file size.

## 3

### Document Categorization

Organize documents into different categories based on their content.

# Conclusion and Key Takeaways

K-means clustering is a powerful tool for unsupervised learning. It helps us uncover hidden patterns in unlabeled data, offering valuable insights for various applications.