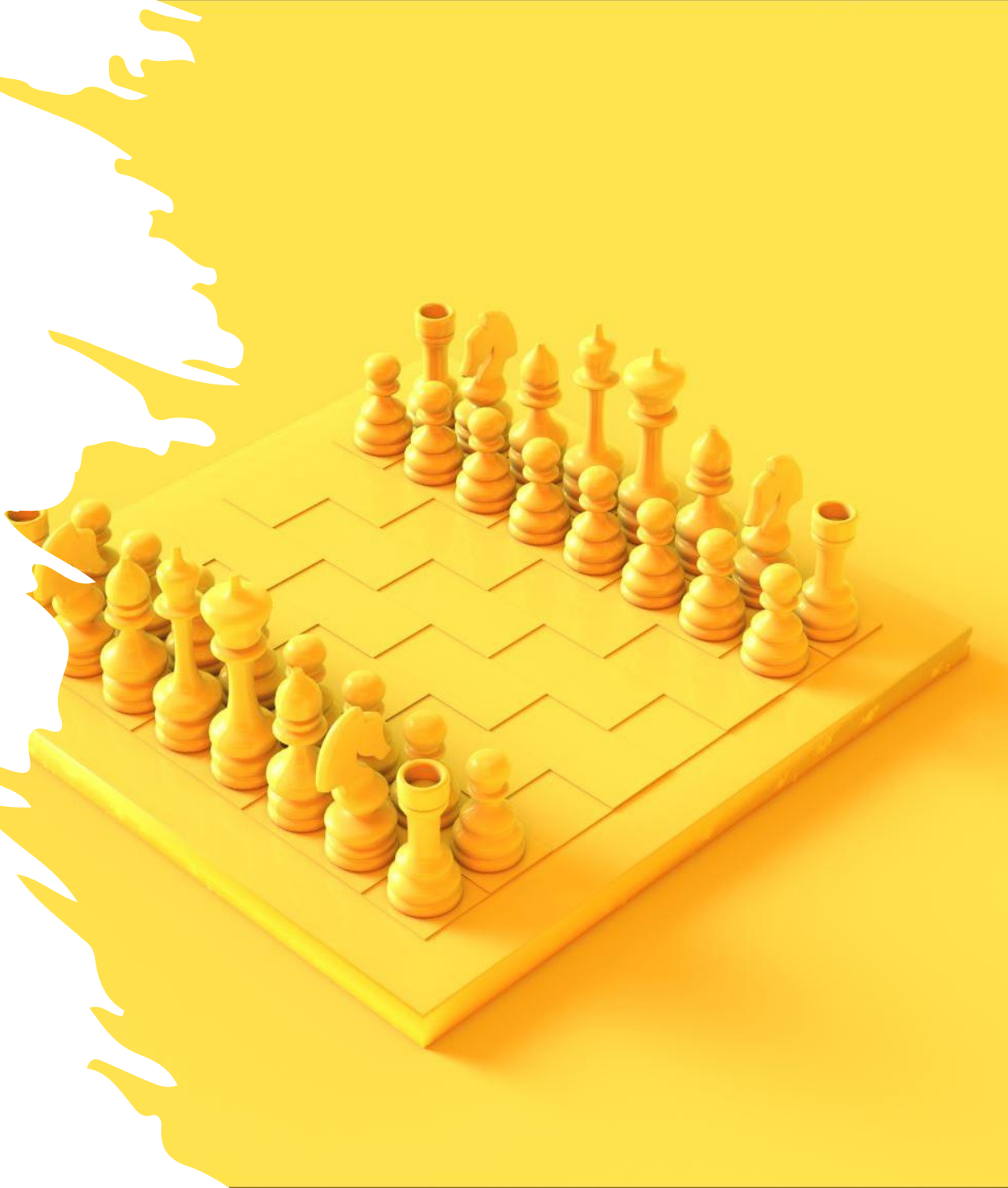
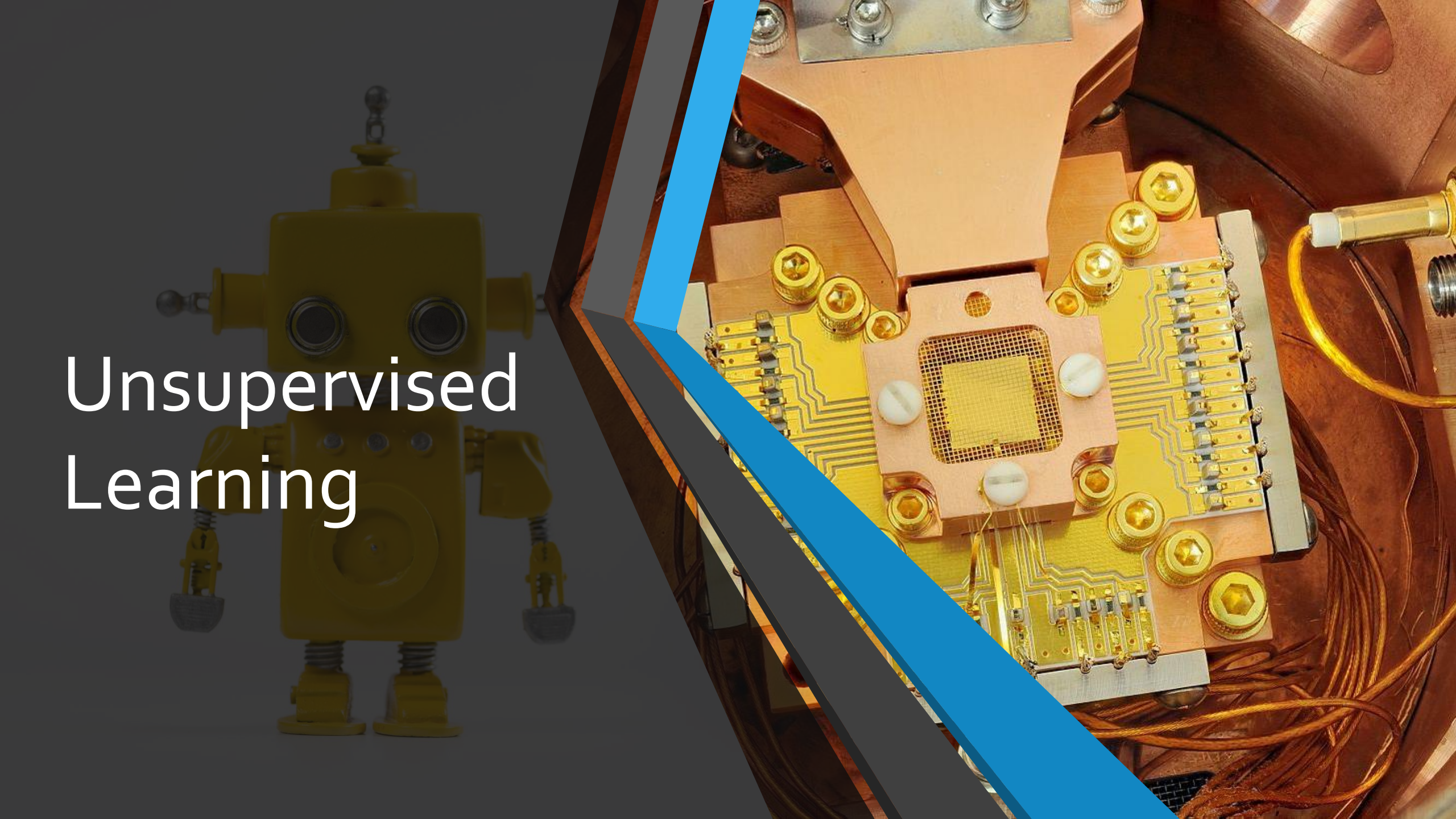
“LIFE IS LIKE A CHESS GAME YOU DON’T WANT TO WASTE A MOVE”



Unsupervised Learning

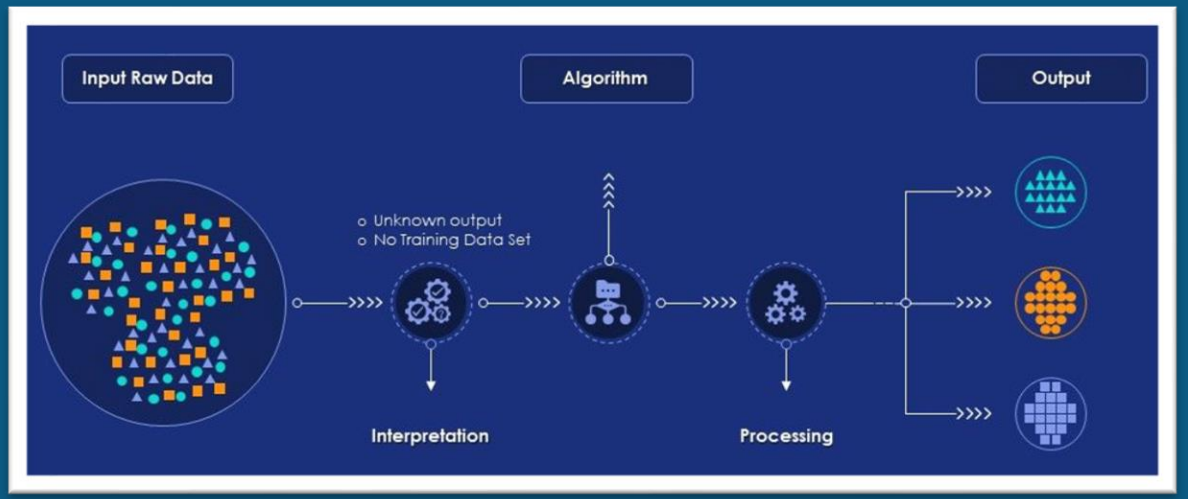


- What is Unsupervised ML?
- What is Clustering ?
- K-Means Clustering

What is Unsupervised Learning?

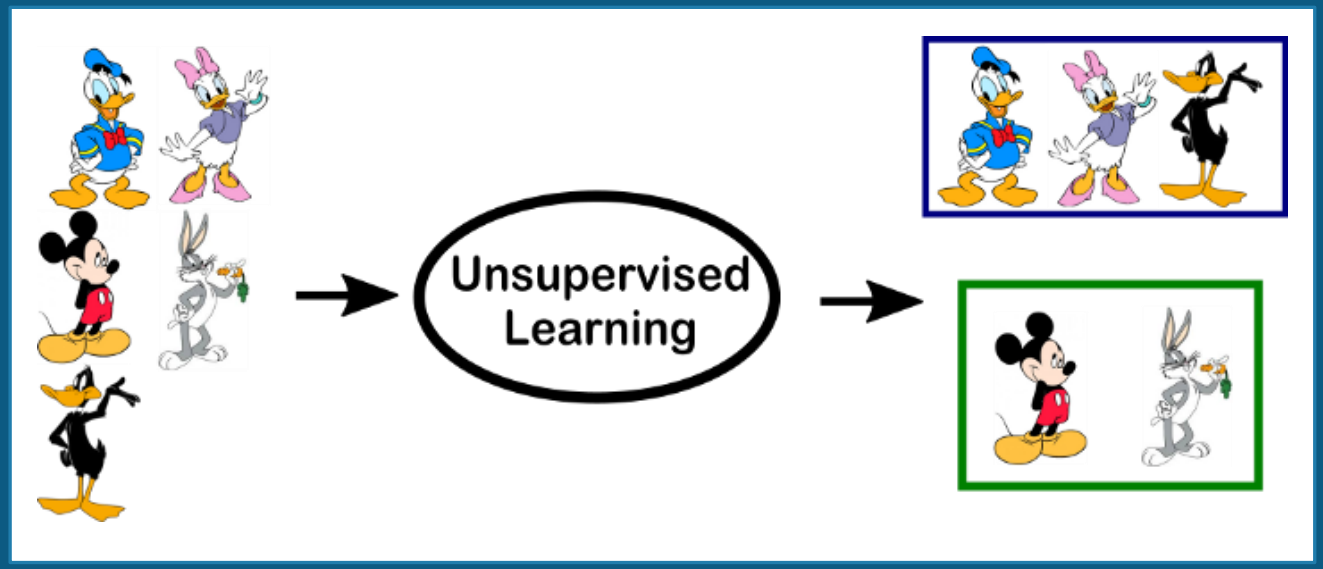
What is Unsupervised ML ?

- In Unsupervised Learning, the algorithm is provided with a finite set of data which does not contains the right answers for each of the input values.
- The goal in such unsupervised learning problems may be to discover groups of similar examples within the data, where it is called **clustering**, or to determine how the data is distributed in the space, known as **density estimation**.



Practical Example

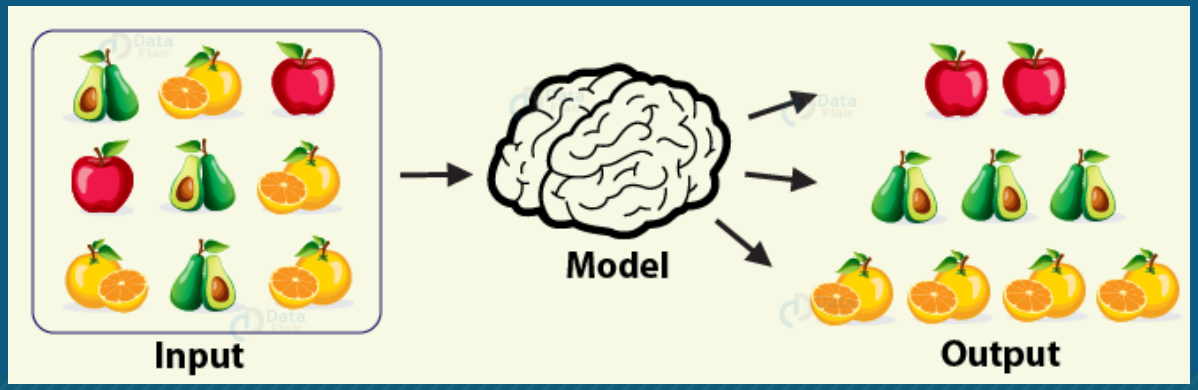
- In the example shown , we have given some characters to our model which are ‘Ducks’ and ‘Not Ducks’. In our training data, we don’t provide any label to the corresponding data. The unsupervised model can separate both the characters by looking at the type of data and models the underlying structure or distribution in the data to learn more about it.



What is Clustering?

What is Clustering ?

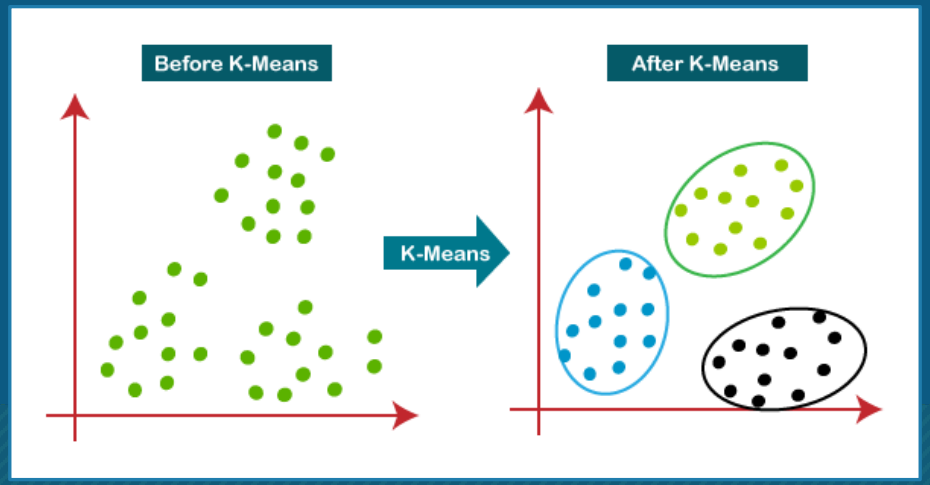
- A clustering problem is where we group similar data according to a pattern in data, such as grouping customers by purchasing behavior.
- In this type of unsupervised learning, the algorithm is given with unlabeled data and it works to find patterns in the data.
- The algorithm groups the data based on similarities into different clusters.



What is K-Means Clustering?

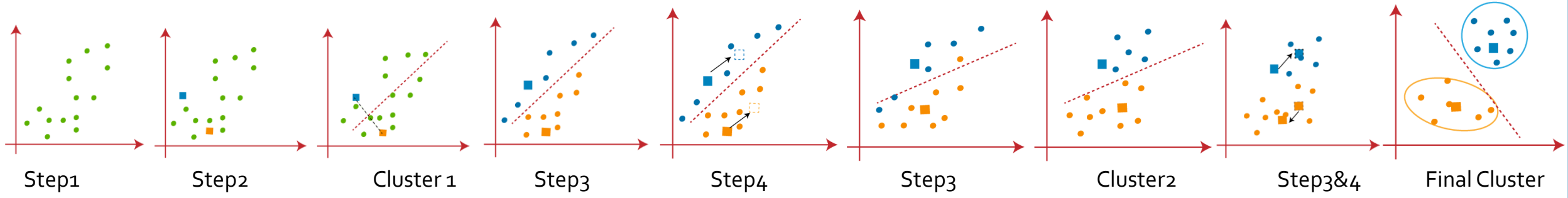
What is K-Means Clustering ?

- It is a type of clustering algorithm which divides the data points into different clusters based on randomly selected centroids. The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters.
- The algorithm takes the unlabeled dataset as input, divides the dataset into k-number of clusters, and repeats the process until it does not find the best clusters. The value of k should be predetermined in this algorithm.
- It halts creating and optimizing clusters when either :-
  - The centroids have stabilized — there is no change in their values because the clustering has been successful.
  - The defined number of iterations has been achieved.



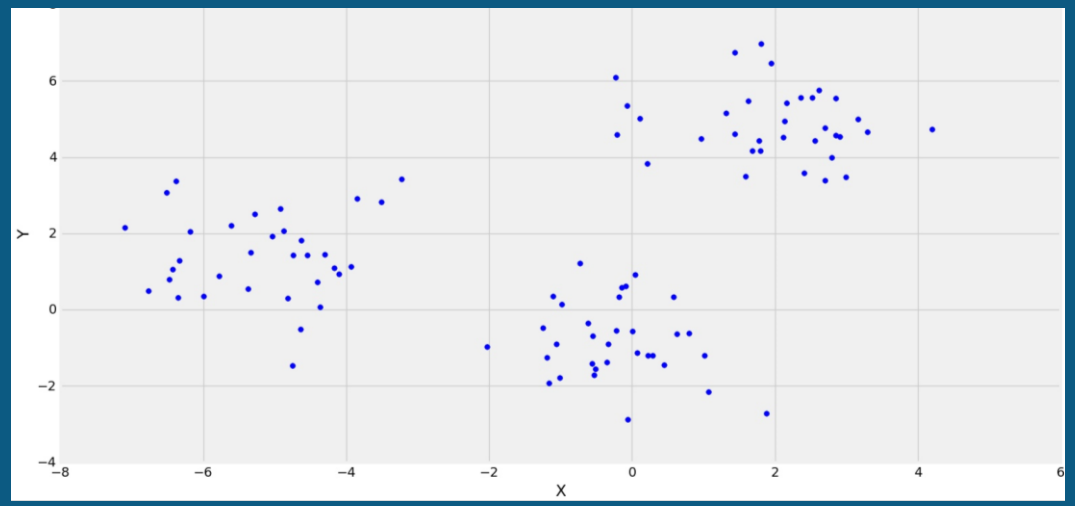
How does K-Means Clustering algorithm work ?

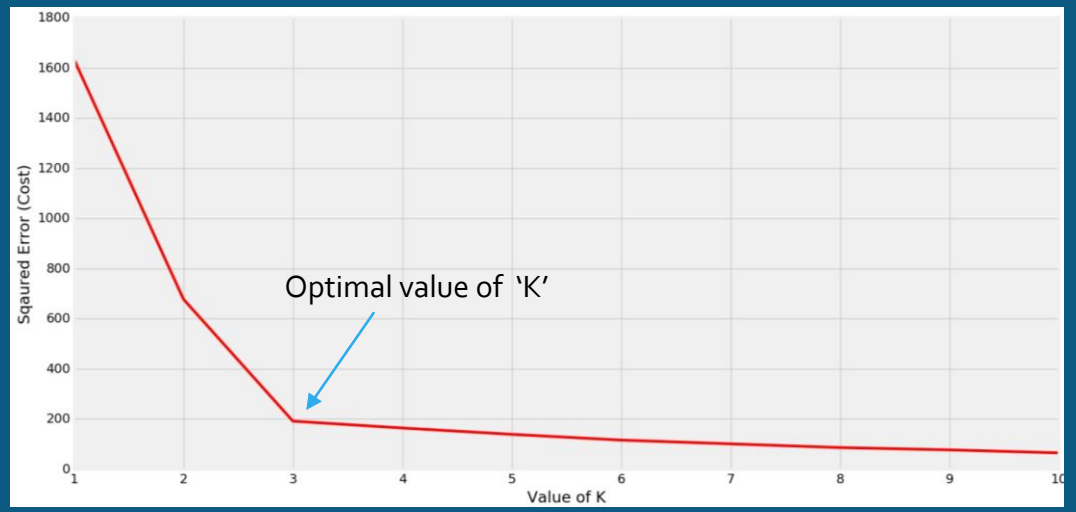
- Choose the number of clusters ‘K’.
- Select ‘K’ random points from the data as centroids.
- Assign all the points to the closest cluster centroid.
- Recompute the centroids of newly formed clusters.
- Repeat the above two steps until the centroid is stabilized or maximum number of iterations are reached.



How to choose the value of 'K' ?

- In order to choose the value of 'K' the most popular method used is **Elbow Method.**
- The basic idea behind this method is that it plots the various values of cost with changing *k*. As the value of *K* increases, there will be fewer elements in the cluster. So average distortion will decrease. The lesser number of elements means closer to the centroid. So, the point where this distortion declines the most (forming elbow) is the **elbow point**.





Thank You