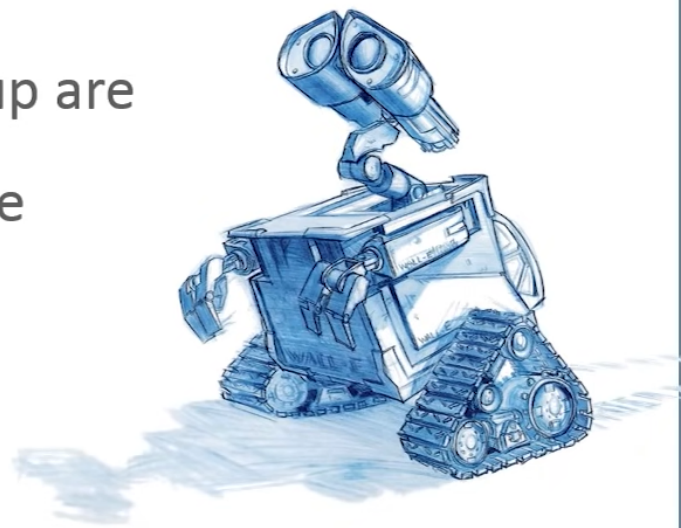
edureka!

What is Clustering?

"**Clustering** is the process of dividing the datasets into groups, consisting of similar data-points"

- Points in the same group are as similar as possible
- Points in different group are as dissimilar as possible



edureka!

What is Clustering?

Group of diners in a restaurant

Items arranged in a mall

edureka!

Where is it Used?

Recommendation System

Recommended Movies

Fickr's Photos

edureka!

How business use Clustering?

Retail Store

Banking

Insurance Companies

edureka!

Types of Clustering

- Exclusive Clustering
- Overlapping Clustering
- Hierarchical Clustering

Exclusive Clustering

- Hard Clustering
- Data Point / Item belongs exclusively to one cluster
- For Example: K-Means Clustering

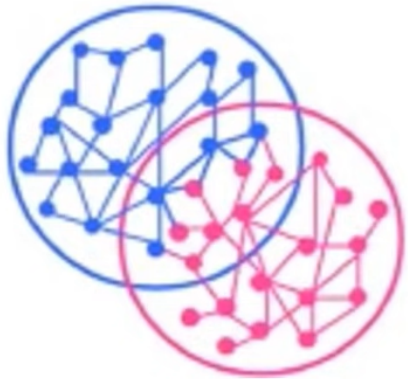
edureka!

Types of Clustering

- Exclusive Clustering
- Overlapping Clustering
- Hierarchical Clustering

Overlapping Clustering

- Soft Cluster
- Data Point/ Item belongs to multiple cluster
- For Example: Fuzzy/ C-Means Clustering

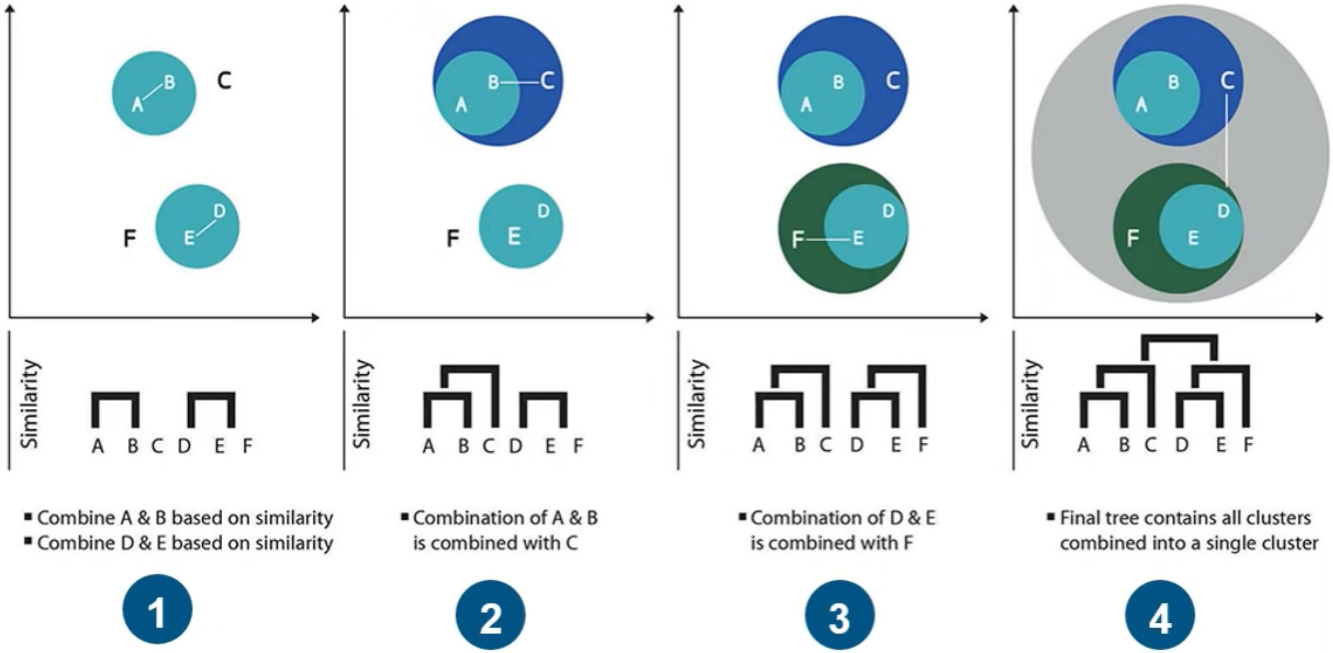


edureka!

Types of Clustering

- Exclusive Clustering
- Overlapping Clustering
- Hierarchical Clustering

Hierarchical Clustering

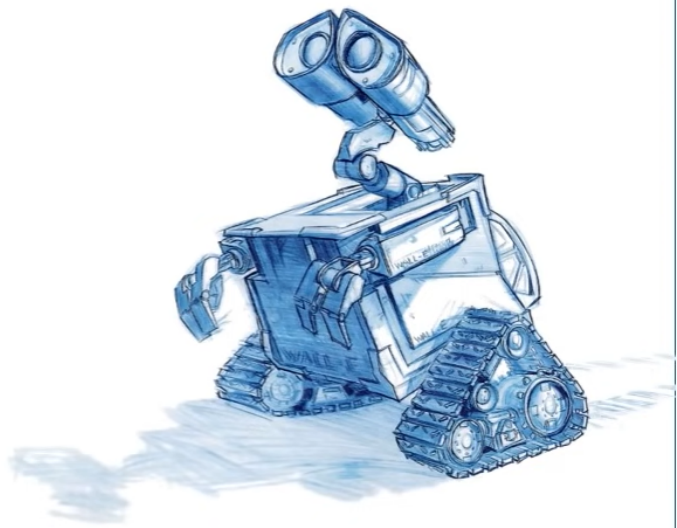


edureka!

What is K-Means Clustering?

"**K-Means** is a clustering algorithm whose mail goal is to group similar elements or data points into a cluster."

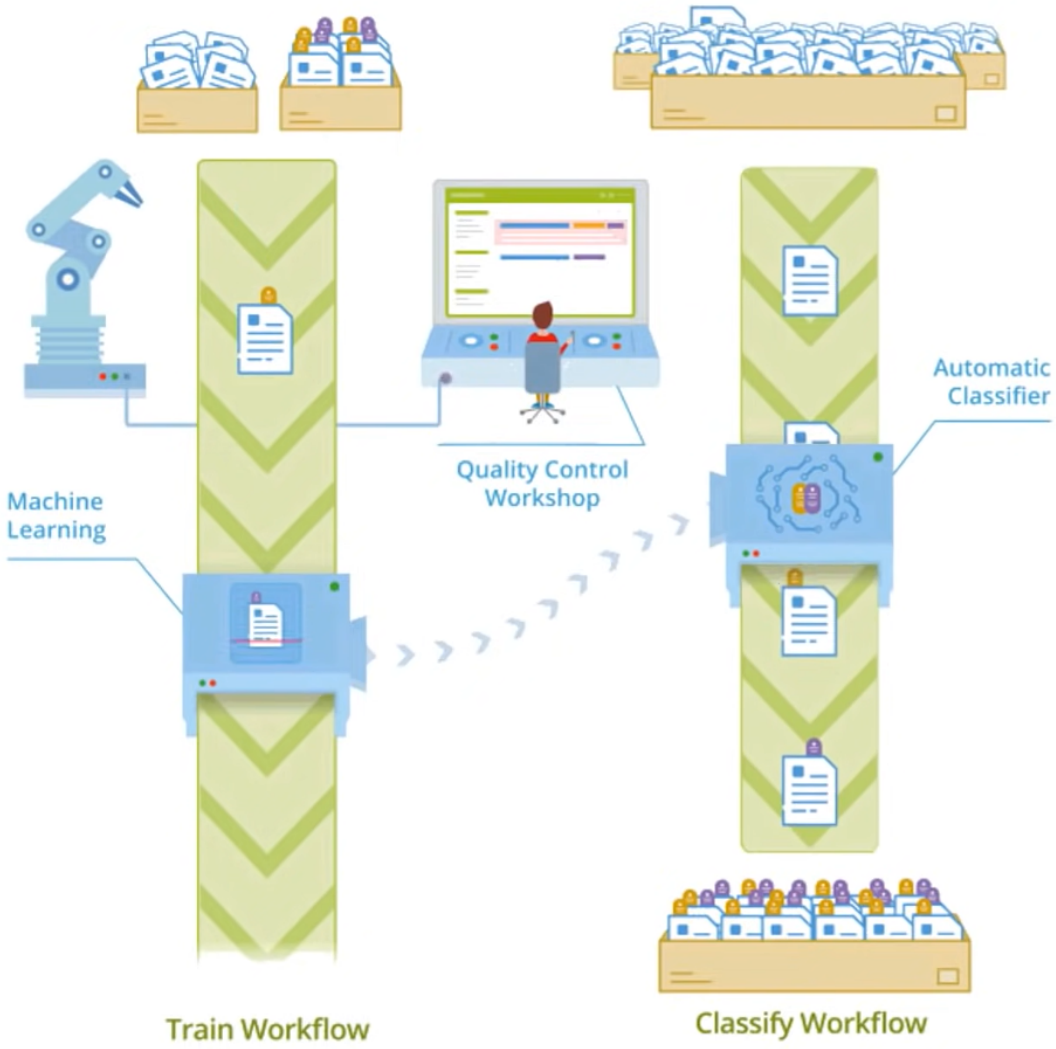
NOTE: 'K' in K-Means represent the number of clusters



edureka!

Where Can I apply K-Means?

https://gifer.com/en/Ckp3



Document Classifier

edureka!

K-Means Algorithm

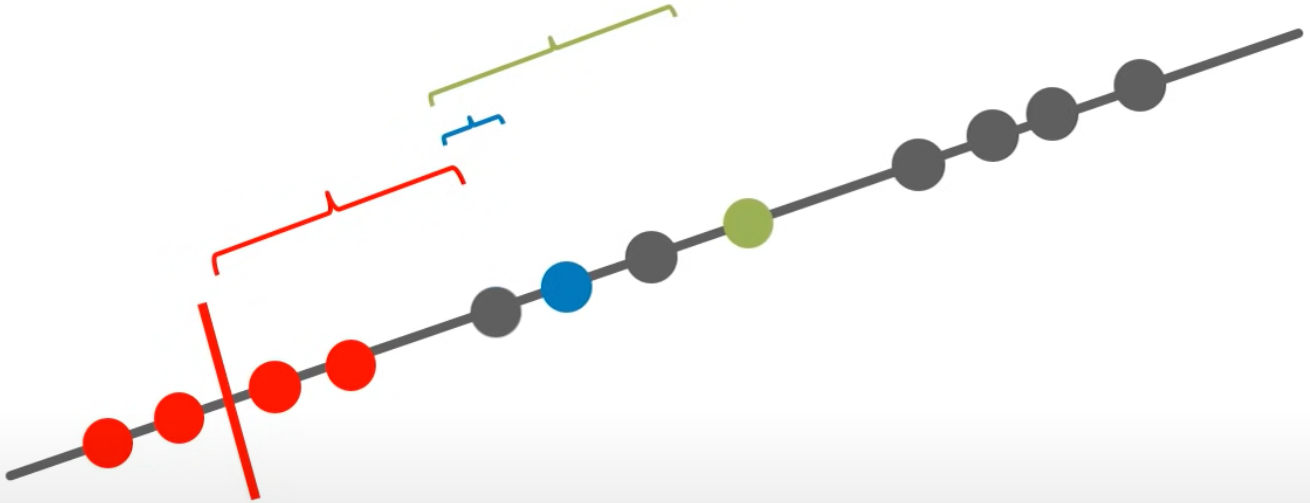
- **Step 1:** Select the number of clusters to be identified,

  i.e select a value for K =3 in this case

- **Step 2:** Randomly select 3 distinct data point

K-Means Algorithm

- Measure the distance
- Assign the point to the nearest cluster
- Calculate the cluster mean using the new point



edureka!

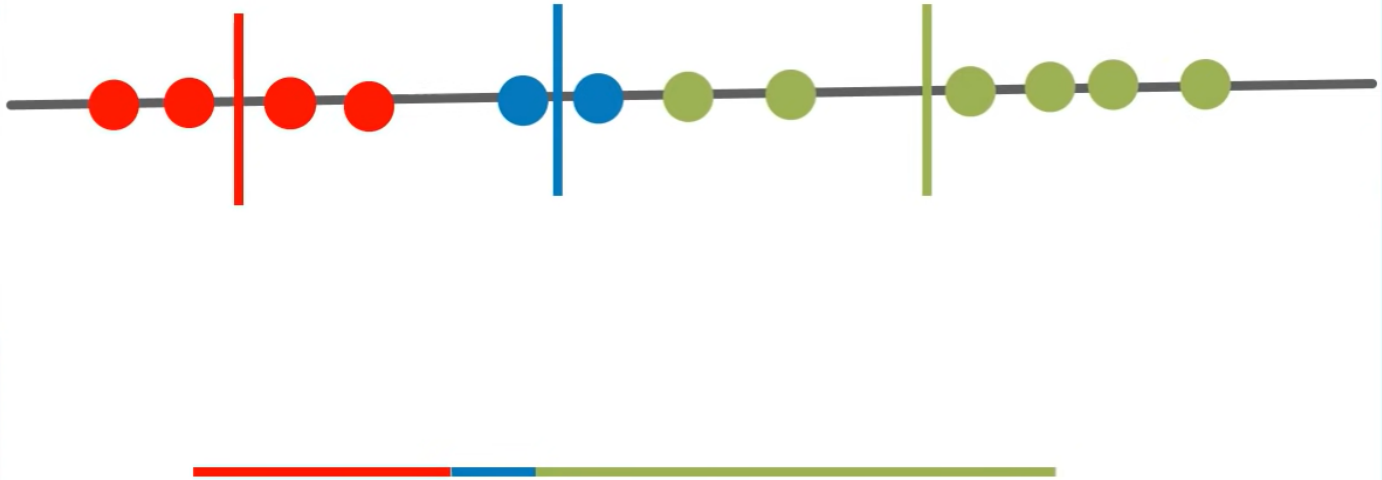
K-Means Algorithm

Result from 1st iteration

Original/Expected Result

edureka!

K-Means Algorithm



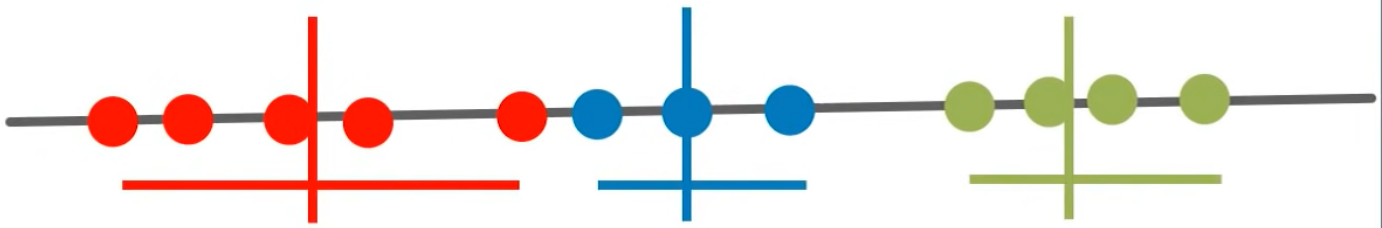
Total variation within the cluster

According to the K-Means Algorithm it iterates over again and again unless and until the data points within each cluster stops changing

K Means Clustering Algorithm | K Means Example in Python | Machine Learning Algorithms | Edureka

K-Means Algorithm

Algorithm picks 3 initial clusters and adds the remaining points to the cluster with the nearest mean, and again recalculating the mean each time a new point is added to the cluster

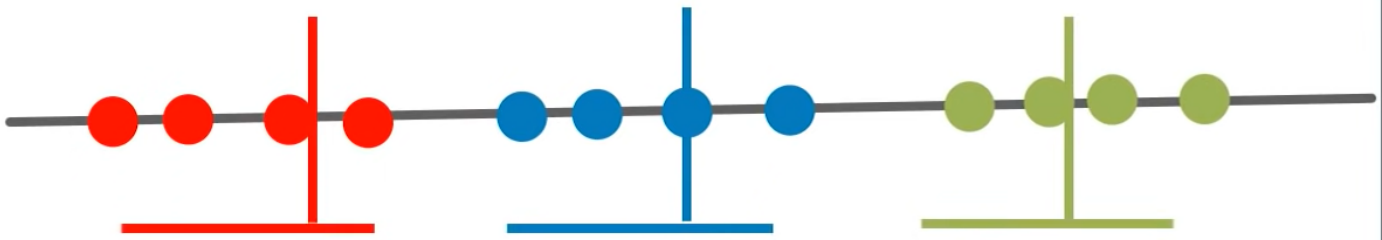


15:24 / 27:04

Scroll for details

K Means Clustering Algorithm | K Means Example in Python | Machine Learning Algorithms | Edureka

K-Means Algorithm

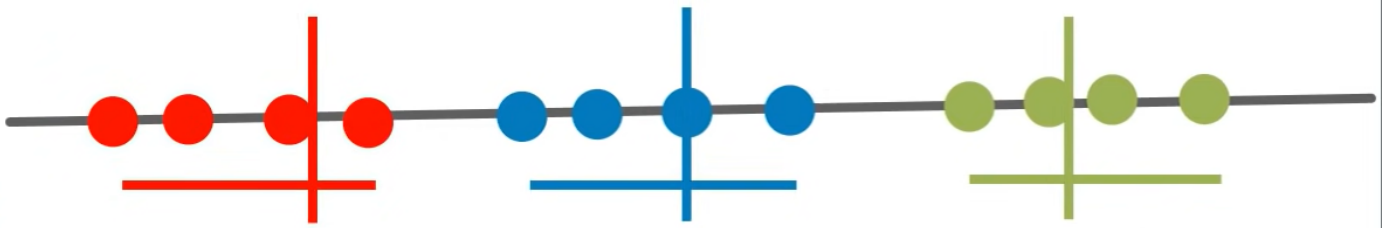


15:33 / 27:04

Scroll for details

K-Means Algorithm

The algorithm can now compare the result and select the best variance out of it



1st Iteration

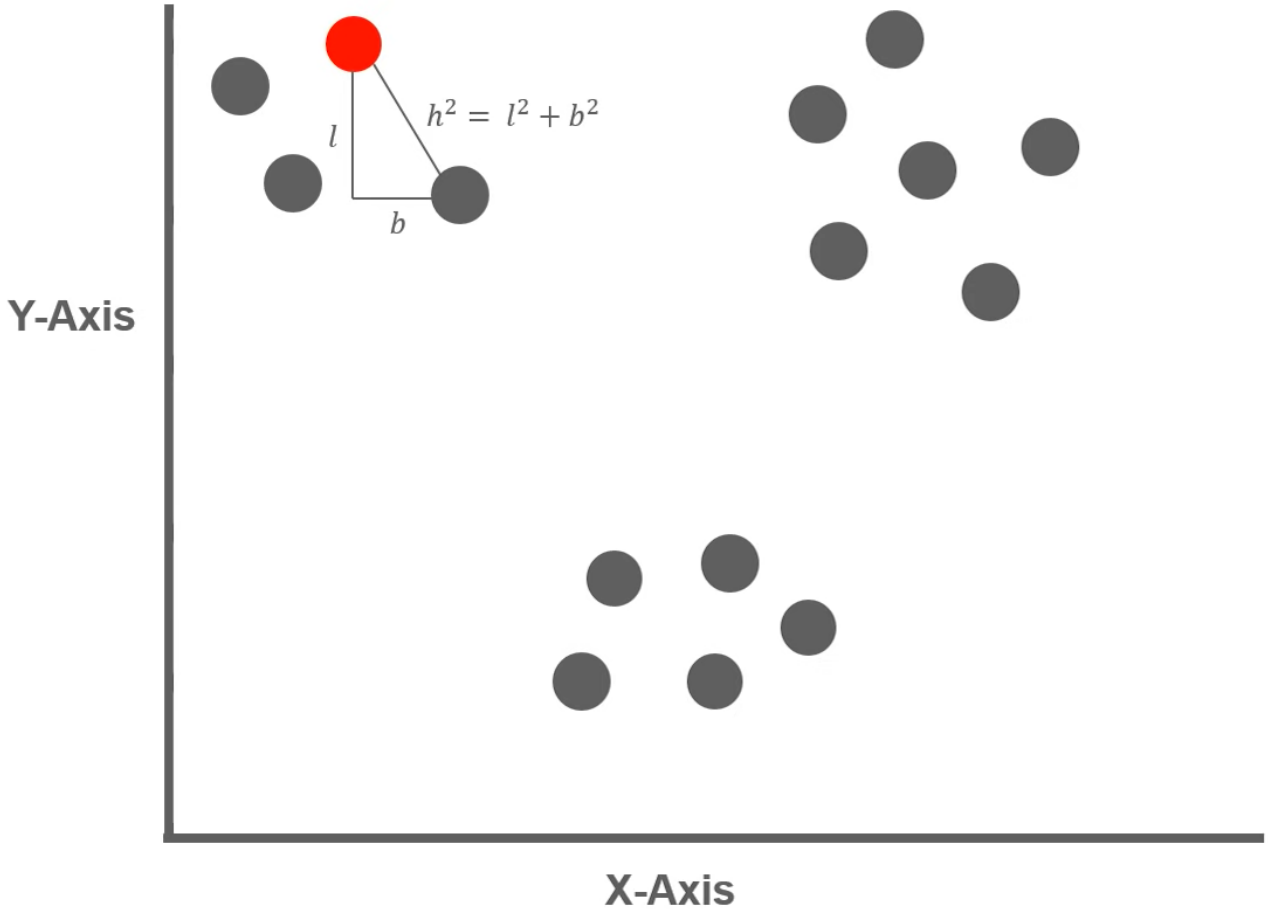
2nd Iteration

3rd Iteration

edureka!

K-Means Algorithm

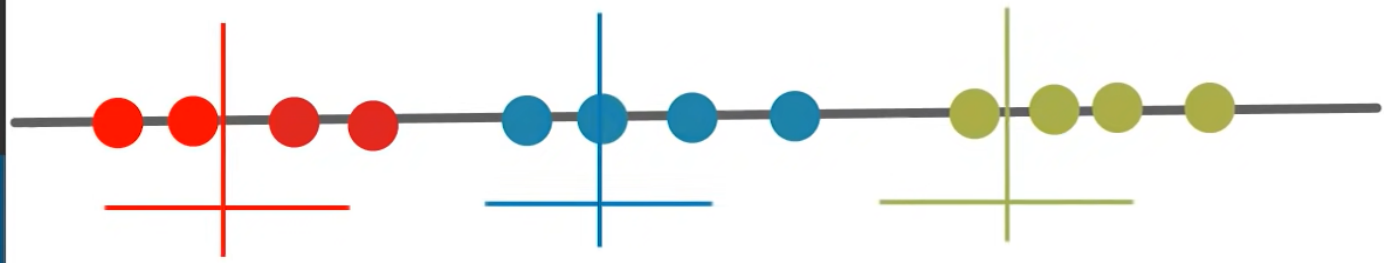
We will be using the Euclidean distance (in 2D its same as that of a Pythagorean Theorem)



edureka!

How will you find K value

Now try with K = 3



K = 1

K = 2

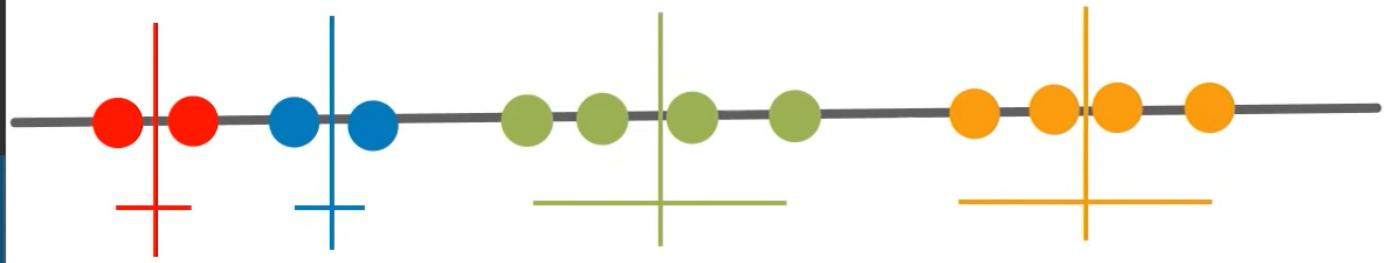
K = 3

edureka!

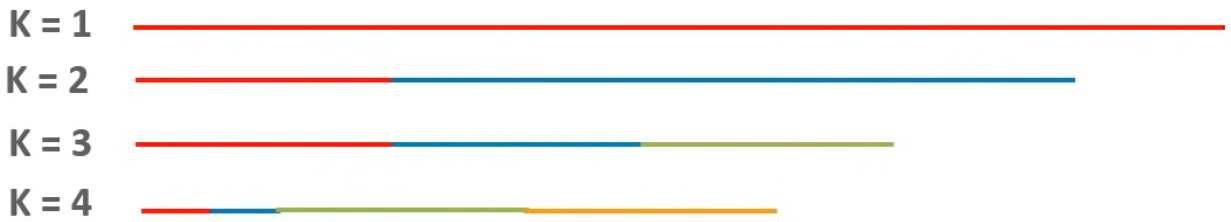
How will you find K value

Now try with K = 4

Each time you increase the cluster the variation decreases, no. of clusters = no. of data points then in that case the variation = 0



Total variation in K=4 is less than K =3



edureka!

How will you find K value