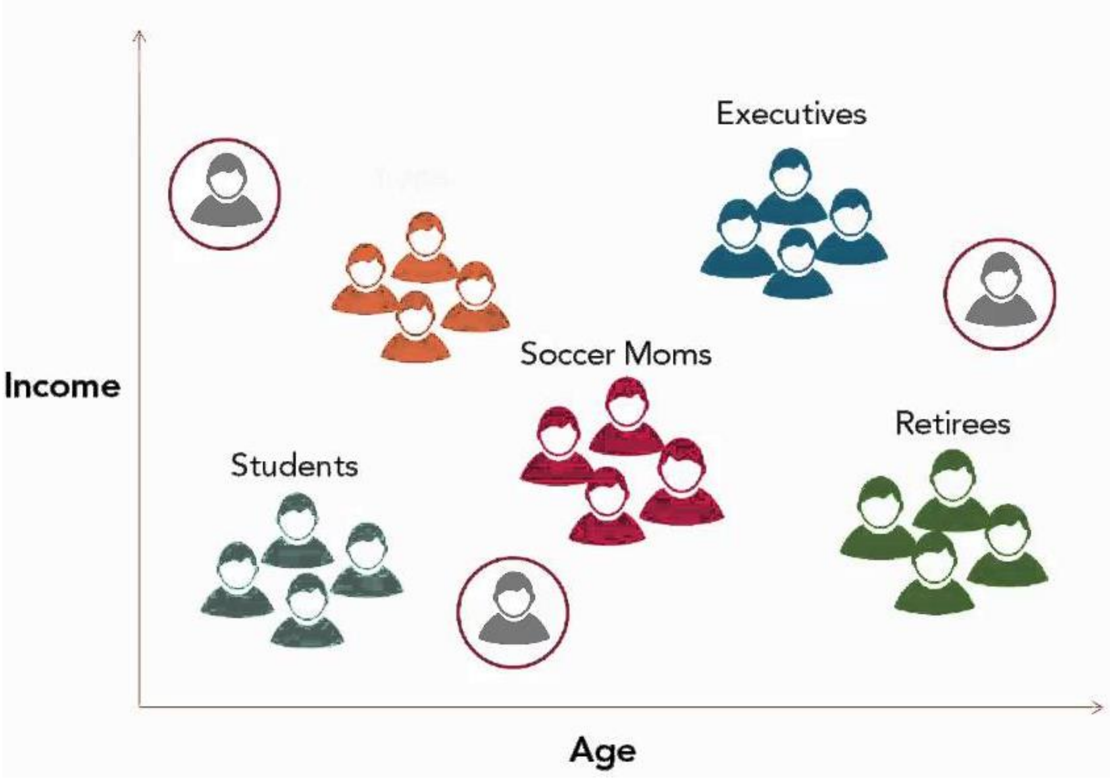
Customer Segmentation

- Customer segmentation is the process of grouping customers based on similar characteristics for the purpose of targeted marketing and resource allocation. Clustering is a technique used for customer segmentation that groups similar data points together and is classified under unsupervised machine learning. The K-means algorithm is one example of a clustering algorithm commonly used in customer segmentation. It groups data points into k clusters based on their proximity to a centroid.



Personalization has significant positive effects

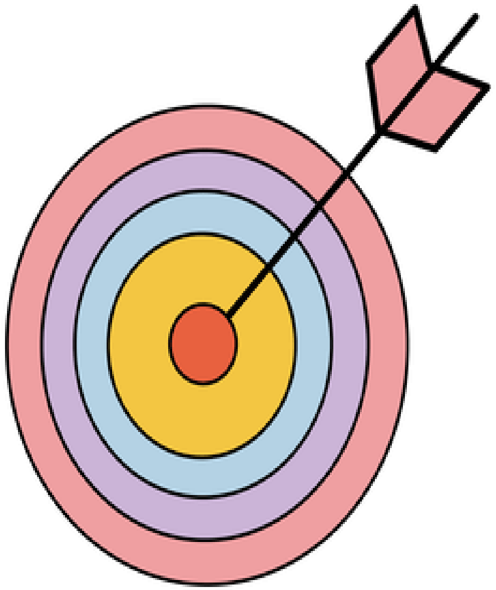
After a consumer has a personalized shopping experience, however:

44% Will be likely to become a repeat buyer

32% Will be likely to leave a positive review

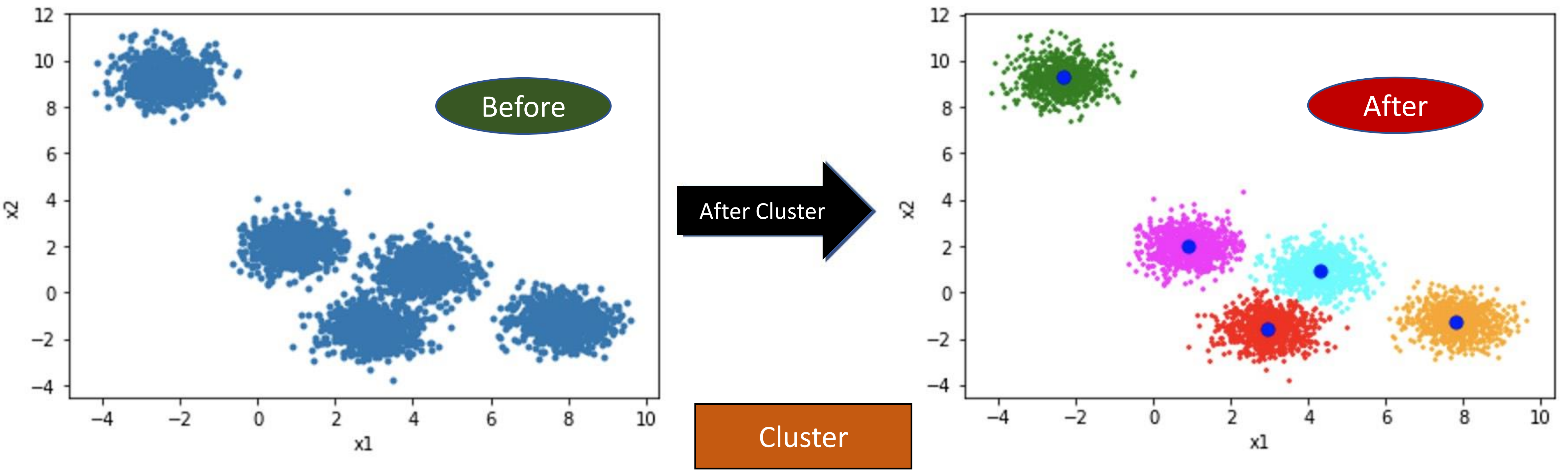
39% Will be likely to tell friends and family

22% Will be likely to post a positive comment on social media



Unsupervised Learning: *Clustering*

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups.



Applications of Clustering: Real-World Scenarios

Clustering is a widely used technique in the industry. It is actually being used in almost every domain, ranging from banking to recommendation engines, document clustering to image segmentation.

- Customer Segmentation
- Document Clustering
- Image Segmentation
- Recommendation Engines

Perform K-Means Cluster

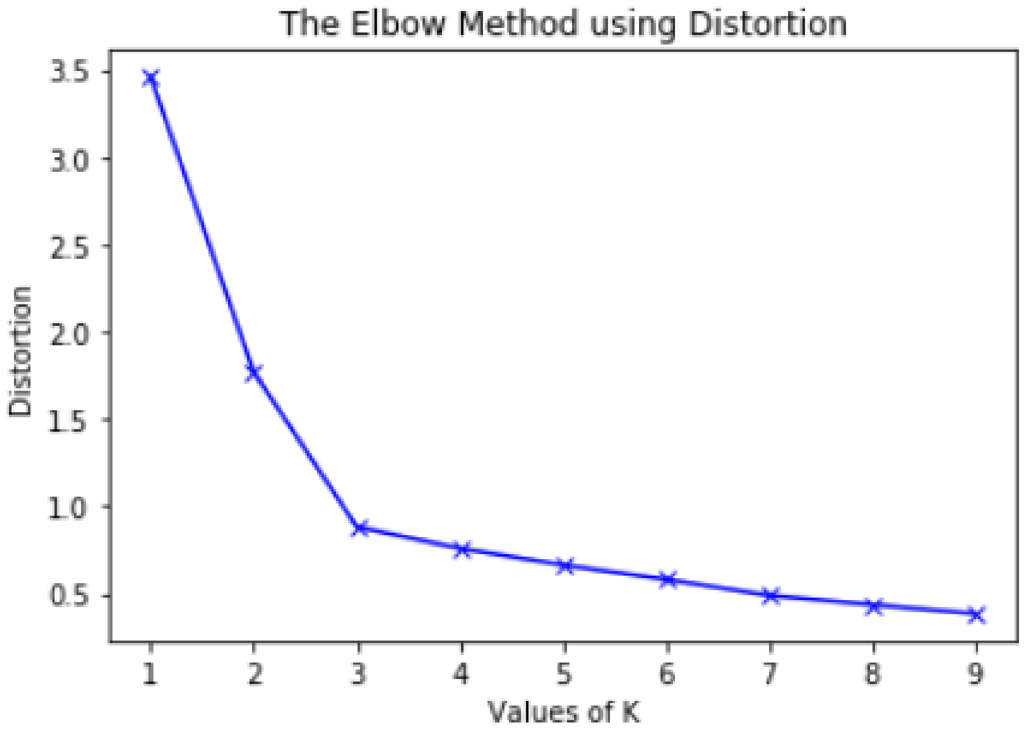
AI Quest

Tasks:

The k-means cluster algorithm mainly performs two important tasks:

- Determines the best value for K center points or centroids by an iterative process.
- Assigns each data point to its closest k-center (also called centroid). Those data points which are near to the particular k-center, create a cluster.

Elbow Method for optimal value of k in KMeans



Perform K-Means Cluster

AI Quest

Steps:

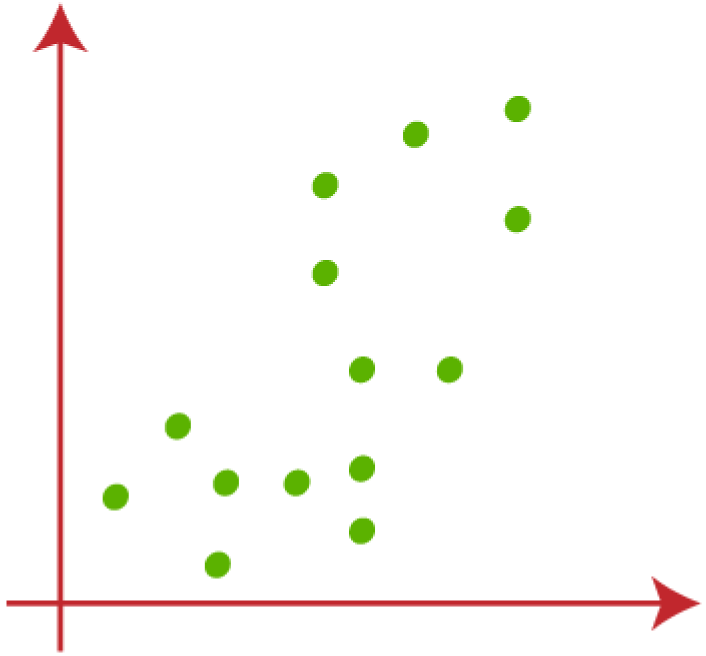
How does the K-Means Algorithm Work?

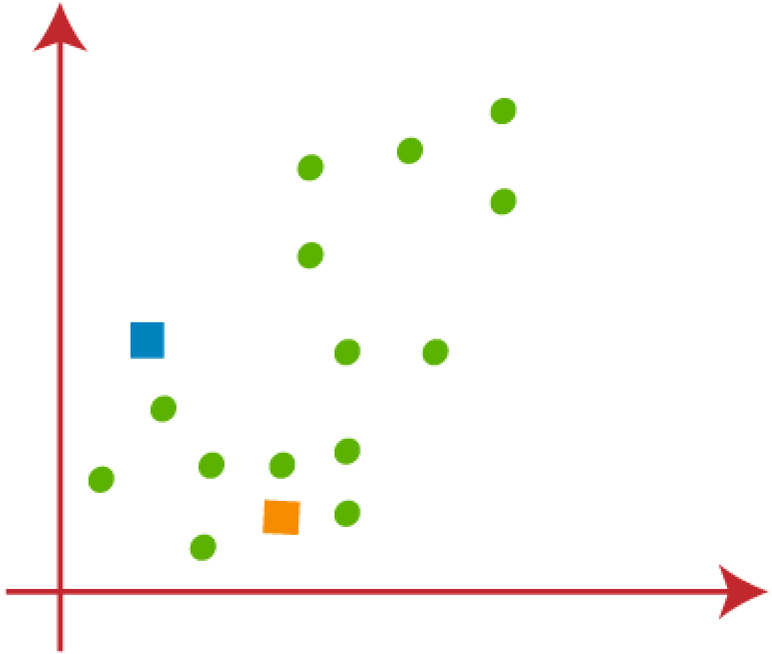
The working of the K-Means algorithm is explained in the below steps:

- Step-1: Select the number K to decide the number of clusters.
- Step-2: Select random K points or centroids.
- Step-3: Assign each data point to their closest centroid, which will form the predefined K clusters.
- Step-4: Calculate the variance and place a new centroid of each cluster.
- Step-5: Repeat the third steps, which means reassign each data point to the new closest centroid of each cluster.
- Step-6: If any reassignment occurs, then go to step-4 else go to FINISH.
- Step-7: The model is ready.

Perform K-Means Cluster

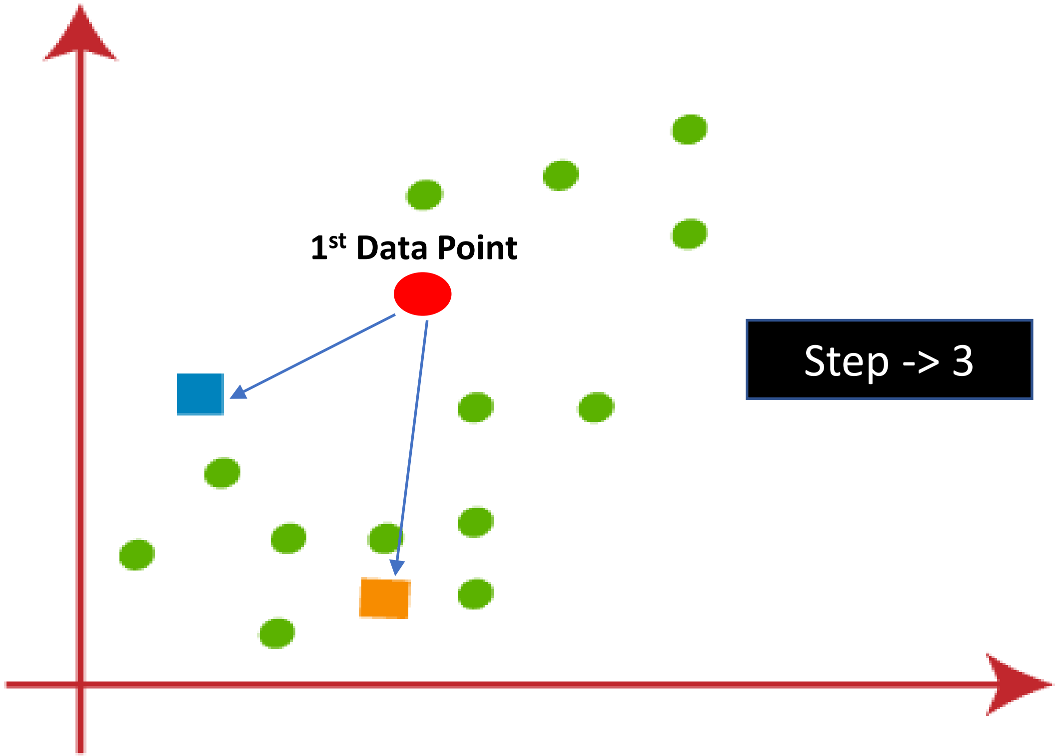
Assume, we have two variables M1 and M2. The X & Y axis scatter plot of these 2 variables is given below:



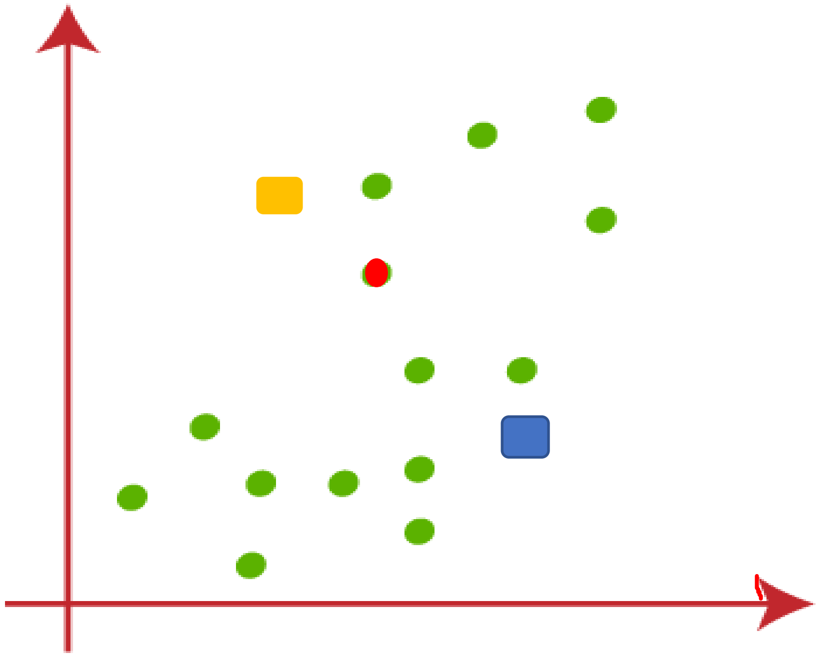


Note: We need to choose some random k points or centroid to form the cluster. These points can be either the points from the dataset or any other point.

Perform K-Means Cluster



Perform K-Means Cluster



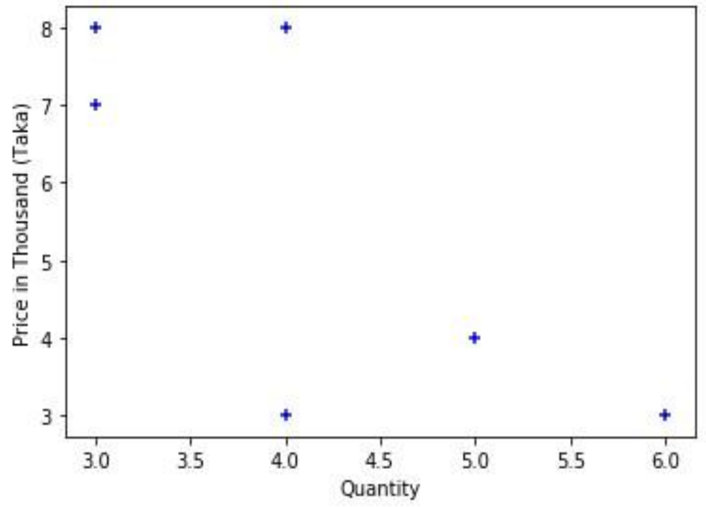
Javapoint.com

Let's see an EXAMPLE

Products	Quantity	Price(K)
FaceWash	3	7
Cream	5	4
Shoes	4	3
Bags	4	8
Jacket	6	3
Shirt	3	8

In [6]:
```
plt.xlabel('Quantity')
plt.ylabel('Price in Thousand (Taka)')
plt.scatter(dataframe['Quantity'], dataframe['Price(K)'],marker='+',color='blue')
```

Out[6]:
```
<matplotlib.collections.PathCollection at 0x2a9e524a048>
```



Products	Quantity	Price(K)
FaceWash	3	7
Cream	5	4
Shoes	4	3
Bags	4	8
Jacket	6	3
Shirt	3	8

$C1 = (3,7)$ and $C_2 = (5,4)$

* For First data point $(3,7)$ Facewash :

Distance from $C_1 = 0$ * C_1

Distance from $C_2 = \sqrt{(5-3)^2 + (4-7)^2}$

$= \sqrt{9+9}$

$= 4.24$

Products	Quantity	Price(K)
FaceWash	3	7
Cream	5	4
Shoes	4	3
Bags	4	8
Jacket	6	3
Shirt	3	8

$c1 = (3, 7)$ and $c_2 = (5, 4)$

* For First data point (3, 7) . Facewash :

Distance from $c_1 = 0$ * c_1

Distance from $c_2 = \sqrt{(5-3)^2 + (4-7)^2}$

$= \sqrt{9+9}$

$= 4.24$

* For second data point (5, 4) . Cream :

Distance from $c_1 = \sqrt{(5-3)^2 + (4-7)^2}$

$= \sqrt{4+9}$

$= \sqrt{13} = 3.60$

Distance from $c_2 = \sqrt{0}$ * c_2

Products	Quantity	Price(K)
FaceWash	3	7
Cream	5	4
Shoes	4	3
Bags	4	8
Jacket	6	3
Shirt	3	8

$c1 = (3,7)$ and $c_2 = (5,4)$

* For First data point (3,7), Facewash :

Distance from $c_1 = 0$ * c_1

Distance from $c_2 = \sqrt{(5-3)^2 + (4-7)^2}$

$= \sqrt{9+9}$

$= 4.24$

* For second data point (5,4), Cream :

Distance from $c_1 = \sqrt{(5-3)^2 + (4-7)^2}$

$= \sqrt{4+9}$

$= \sqrt{13} = 3.60$

Distance from $c_2 = \sqrt{0}$ * c_2

* For third data point (4,3) shoes :

Distance from $c_1 = \sqrt{(4-3)^2 + (3-7)^2}$

$= \sqrt{1+16}$

$= 4.123$

Distance from $c_2 = \sqrt{(4-5)^2 + (3-4)^2}$

$= \sqrt{1+1}$

$= 1.41$ * c_2

So new centroid $= \left(\frac{5+4}{2}, \frac{4+3}{2}\right)$

$c_2 = (4.5, 3.5)$

Products	Quantity	Price(K)
FaceWash	3	7
Cream	5	4
Shoes	4	3
Bags	4	8
Jacket	6	3
Shirt	3	8

$c_1 = (3, 7)$ and $c_2\,(4.5, 3.5)$

For 4th data point (4, 8) bags :

Distance from $c_1 = \sqrt{(4-3)^2 + (8-7)^2}$

$= \sqrt{1+1}$

$= 1.41$ * c_1

Distance from $c_2 = \sqrt{(4-4.5)^2 + (8-3.5)^2}$

$= 0.25 + 20.25$

$= 20.50$

$\therefore$ New centroid $= \left(\frac{3+4}{2}, \frac{7+8}{2}\right)$

$c_1 = (3.5, 7.5)$

Products	Quantity	Price(K)
FaceWash	3	7
Cream	5	4
Shoes	4	3
Bags	4	8
Jacket	6	3
Shirt	3	8

For 5th data point (6,3) Jacket :

Distance from $c_1 = \sqrt{(6-3.5)^2 + (3-7.5)^2}$

$= 6.25 + 20.25$

$= 26.5$

Distance from $c_2 = \sqrt{(4.5-6)^2 + (3-3.5)^2}$

$= (2.25 + 0.25)$

$= 2.50$ $*$ c_2

New centroid $= \left(\frac{5+4+6}{3}, \frac{4+3+3}{3} \right)$

$c_2 = (5, 3.33)$

Products	Quantity	Price(K)
FaceWash	3	7
Cream	5	4
Shoes	4	3
Bags	4	8
Jacket	6	3
Shirt	3	8

$c_1 = (3, 7)$ and $c_2\,(4.5, 3.5)$

<u>For 4th data point (4,8) bags :</u>

Distance from $c_1 = \sqrt{(4-3)^2 + (8-7)^2}$

$= \sqrt{1+1}$

$= 1.41$ $\circledast$ c_1

Distance from $c_2 = \sqrt{(4-4.5)^2 + (8-3.5)^2}$

$= 0.25 + 20.25$

$= 20.50$

$\therefore$ New centroid $= \left(\frac{3+4}{2}, \frac{7+8}{2}\right)$

$c_1 = (3.5, 7.5)$

<u>For 5th data point (6,3) Jacket :</u>

Distance from $c_1 = \sqrt{(6-3.5)^2 + (3-7.5)^2}$

$= 6.25 + 20.25$

$= 26.5$

Distance from $c_2 = \sqrt{(4.5-6)^2 + (3-3.5)^2}$

$= (2.25 + 0.25)$

$= 2.50$ $\circledast$ c_2

New centroid $= \left(\frac{5+4+6}{3}, \frac{4+3+3}{3}\right)$

$c_2 = (5, 3.33)$

Products	Quantity	Price(K)
FaceWash	3	7
Cream	5	4
Shoes	4	3
Bags	4	8
Jacket	6	3
Shirt	3	8

$c_1 = (3.5, 7.5)$ and $c_2 = (5, 3.33)$

For 6th data point (3,8) shirt :

$$\text{Distance from } c_1 = \sqrt{(3-3.5)^2 + (8-7.5)^2}$$

$$= \sqrt{.25 + .25}$$

$$= 0.70 \quad * \quad c_1$$

$$\text{Distance from } c_2 = \sqrt{(3-5)^2 + (8-3.33)^2}$$

$$= \sqrt{4 + 2.16}$$

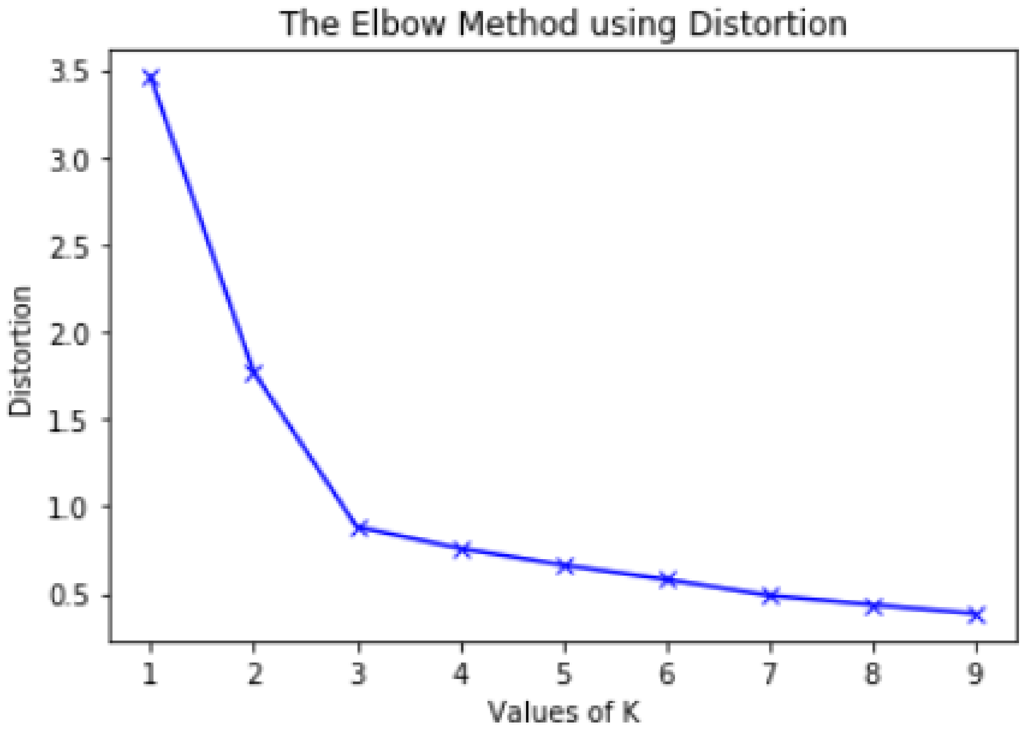
$$= 2.48$$

$$\text{New centroid} = \left(\frac{3+4+3}{3}, \frac{7+8+8}{3}\right)$$

$$c_1 = (3.33, 7.67)$$

$$c_2 = (5, 3.33)$$

Elbow Method for optimal value of k in KMeans



Applications of Clustering: Real-World Scenarios

- A retail store is looking to get insight about it's customers by identifying patterns and similarities in customer data, specifically age, income, score and gender, etc. To achieve this, the store plans to use clustering techniques to group customers into distinct segments, and then analyze each cluster to better understand the needs of each customer segment. The ultimate goal is to use this information to create targeted marketing strategies and improve overall customer satisfaction.