In [1]:

```
import numpy as np
```

In [4]:

```
import os
```

In [6]:

```
def compute_euclidean_distance(point, centroid):
    return np.sqrt(np.sum((point - centroid)**2))
```

In [9]:

```
def assign_label_cluster(distance, data_point, centroids):
    index_of_minimum = min(distance, key=distance.get)
    return [index_of_minimum, data_point, centroids[index_of_minimum]]
```

In [11]:

```
def compute_new_centroids(cluster_label, centroids):
    return np.array(cluster_label + centroids)/2
```

In [13]:

```
def iterate_k_means(data_points, centroids, total_iteration):
    label = []
    cluster_label = []
    total_points = len(data_points)
    k = len(centroids)

    for iteration in range(0, total_iteration):
        for index_point in range(0, total_points):
            distance = {}
            for index_centroid in range(0, k):
                distance[index_centroid] = compute_euclidean_distance(data_points[in
            label = assign_label_cluster(distance, data_points[index_point], centroi
            centroids[label[0]] = compute_new_centroids(label[1], centroids[label[0]

            if iteration == (total_iteration - 1):
                cluster_label.append(label)

    return [cluster_label, centroids]
```

In [14]:

```
def print_label_data(result):
    print("Result of k-Means Clustering: \n")
    for data in result[0]:
        print("data point: {}".format(data[1]))
        print("cluster number: {} \n".format(data[0]))
    print("Last centroids position: \n {}".format(result[1]))
```

In [16]:

```
def create_centroids():
    centroids = []
    centroids.append([5.0, 0.0])
    centroids.append([45.0, 70.0])
    centroids.append([50.0, 90.0])
    return np.array(centroids)
```

In [19]:

```
    filename = "income.csv"
    data_points = np.genfromtxt(filename, delimiter=",")
    centroids = create_centroids()
    total_iteration = 100
    [cluster_label, new_centroids] = iterate_k_means(data_points, centroids, total_i
    print_label_data([cluster_label, new_centroids])
    print()
```

```
Result of k-Means Clustering:

data point: [nan nan]
cluster number: 0

data point: [5.0e+01 1.5e+05]
cluster number: 0

data point: [2.e+01 3.e+04]
cluster number: 0

data point: [1.8e+01 2.0e+04]
cluster number: 0

data point: [   17. 10000.]
cluster number: 0

data point: [4.3e+01 1.2e+05]
cluster number: 0

data point: [2.7e+01 3.5e+04]
cluster number: 0

data point: [3.e+01 5.e+04]
cluster number: 0

data point: [3.5e+01 6.0e+04]
cluster number: 0

data point: [6.0e+01 1.8e+05]
cluster number: 0

data point: [3.9e+01 1.0e+05]
cluster number: 0

data point: [2.9e+01 4.0e+04]
cluster number: 0

data point: [2.7e+01 3.7e+04]
cluster number: 0

data point: [4.8e+01 1.1e+05]
cluster number: 0

data point: [2.5e+01 3.0e+04]
cluster number: 0

Last centroids position:
 [[nan nan]
 [45. 70.]
 [50. 90.]]
```

```
In [ ]:
```