

**Task:**

To implement a simple algorithm to cluster people according to their weight and height. The data set includes a list of people with their weights and heights like so:

Person	Weight (in kg)	Height (in inches)
Person1	70	70
Person 2	75	80
Person 3	120	85

You can plot the data as a 2 dimensional data. Weight being one dimension and height being the other dimension. Weight can range from a minimum of 50kg to 150kg. Height can range from a minimum of 38inches to 90inches

**Algorithm:**

The algorithm (called K-means clustering) will cluster data into K groups goes as such:

1. Start with K clusters. Each cluster is defined by its center point which will start of as random weight and random height. Pick random numbers from within the corresponding ranges defined above.
2. For each person  
Calculate distance to center of each cluster using formula  

$$\text{distance} = \sqrt{(w_{\text{person}} - w_{\text{center}})^2 + (h_{\text{person}} - h_{\text{center}})^2}$$

where  $w_{\text{person}}$  = weight of person,  
 $h_{\text{person}}$  = height of person  
 $w_{\text{center}}$  = weight of cluster center point,  
 $h_{\text{center}}$  = height of cluster center point
3. Assign the person to the cluster with the shortest distance to center point of cluster
4. After end of step 2, you will end up with K clusters each assigned with a set of people
5. For each cluster, set the weight and height of the center point to the average of the people in the cluster  
 $w_{\text{center}} = (\text{sum of weight of each person in cluster}) / (\text{number of people in cluster})$   
 $h_{\text{center}} = (\text{sum of height of each person in cluster}) / (\text{number of people in cluster})$
6. Repeat steps 2 to 5 for 1000 iterations, then print out following information for each cluster.
  1. weight and height of center of cluster.
  2. list of people in cluster.

**Files:**

- person\_data.xls – dataset with list of person weights and heights, feel free to convert
- class\_diagram.pdf – suggested class diagram for implementing the task. Feel free to improvise on this design or use your own design.
  - If you do use the class diagram the sequence works as follows:
    - KmeansTestDriver class has the main method used to execute the test
    - In the main method, read the data file (you can convert file to csv to make it easier to read) and create List<Person> from the data file
    - in KmeansTestDriver create an instance of KmeansClusterCreator as such -
      - new KmeansClusterCreator(numberOfClusters, new EuclideanDistanceCalculator())
    - Call the createCluster() method on this instance with the list of persons read from data file
    - Print return value from List<Cluster> as returned by this method

