**DATA SCIENCE ASSIGNMENT**

Implementation of an unsupervised learning algorithm

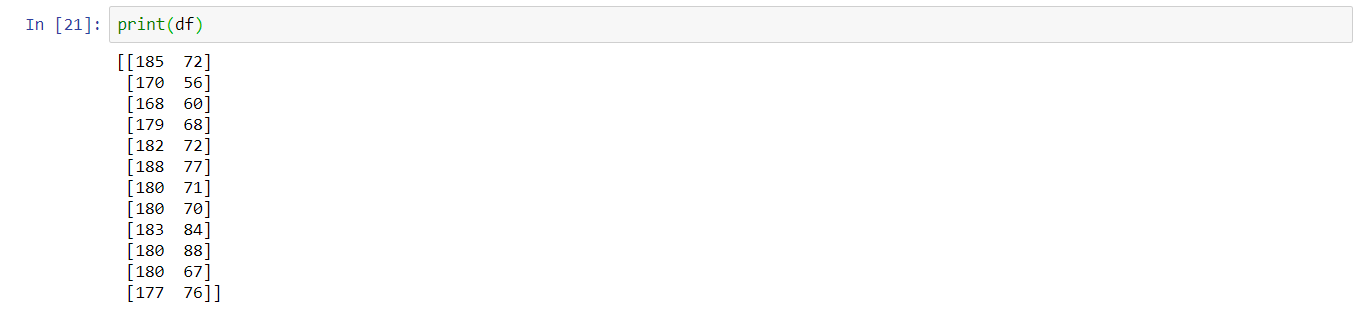
K-Means

By,

Yash Khakhar – 1MS16IS120

Vishal Chawan – 1MS16IS116

* ABSTRACT – The target of this project is to use an unlabelled data set and cluster the data using K-Means algorithm. The algorithm is expected to give us clusters in the data.
* INTRODUCTION - This algorithm will regroup **n**data points into **K**number of clusters. So given a large amount of data, we need to cluster this data into K clusters. A snapshot of the dataset is shown below.



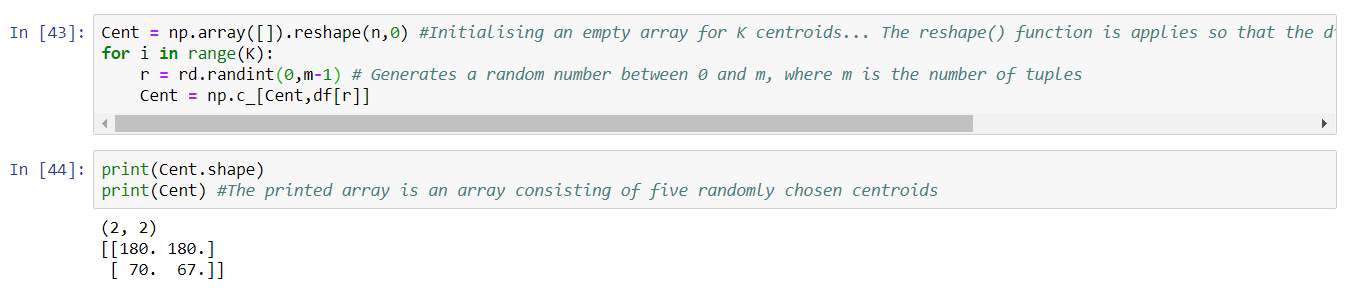
Given the dataset the algorithm will cluster customers on the basis of their heights and weights.

* **ALGORITHM** –

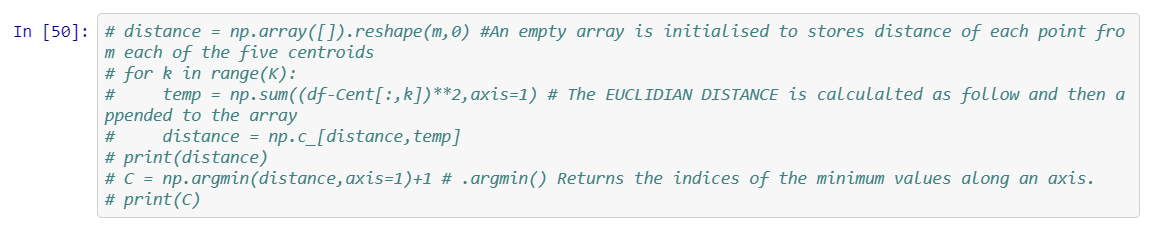
1. Randomly initialize the cluster centers of each cluster from the data points.
2. For each data point, compute the euclidian distance from all the centroids and assign the cluster based on the minimal distance to all the centroids.
3. Adjust the centroid of each cluster by taking the average of all the data points which belong to that cluster on the basis of the computations performed in the previous step.
4. Repeat till the specified number of iterations are done.

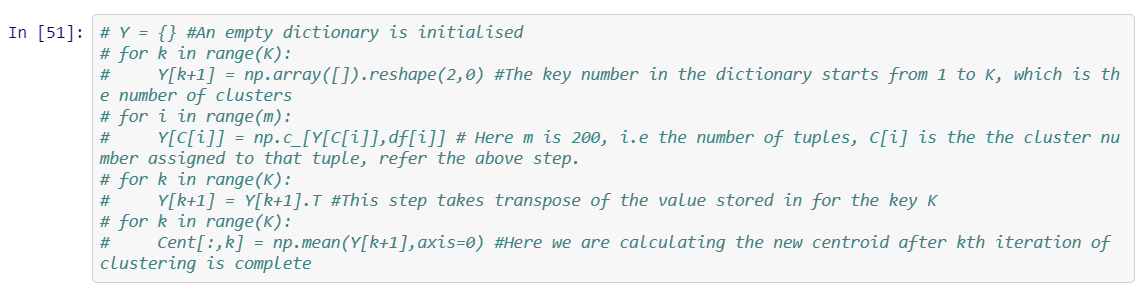
* **IMPLEMENTATION** –

1. Generating random centroids –



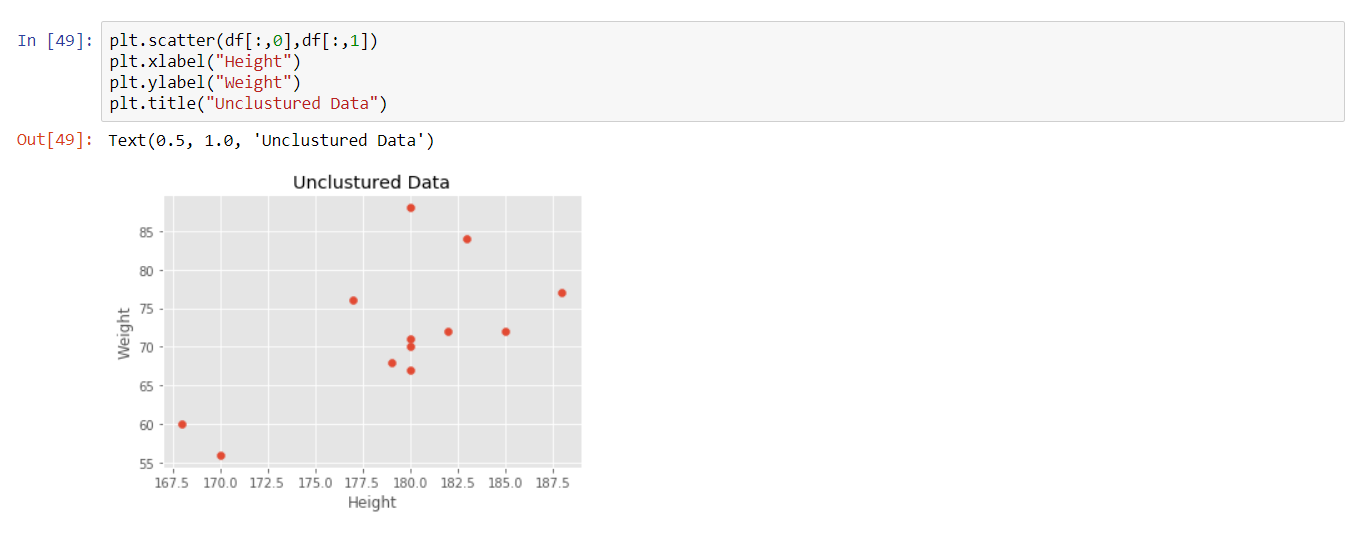
1. Calculating distance of each point from the generated centroids -



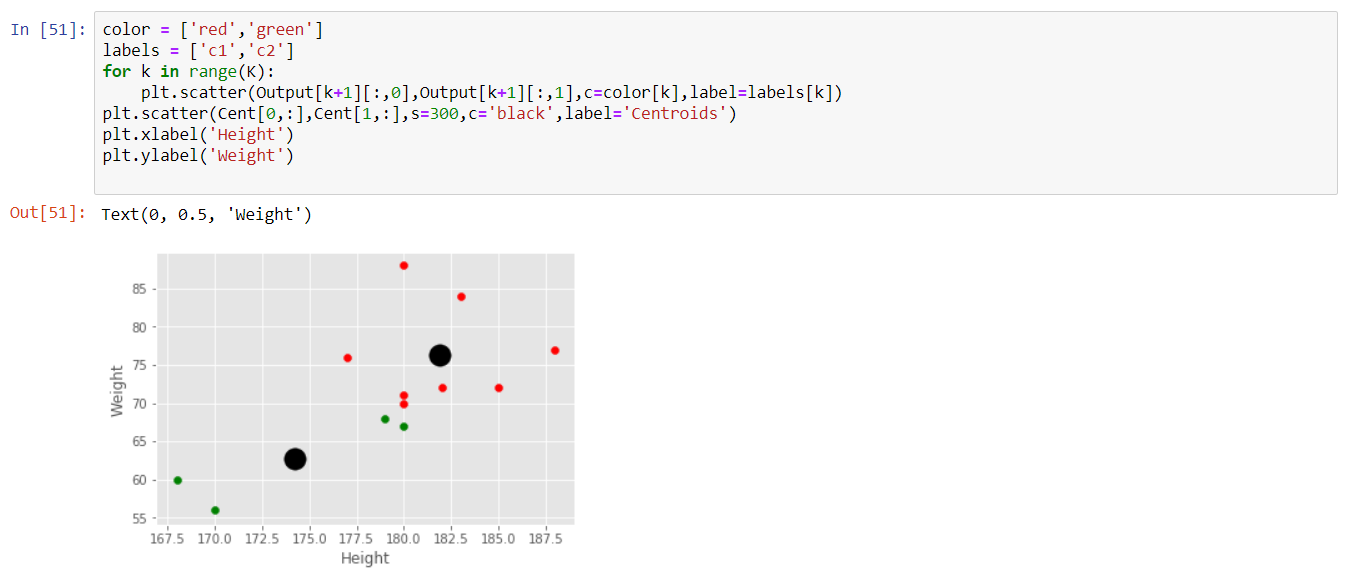
1. Classifies the points by making a dictionary, wherein the key is the cluster number and value for that key is the array of elements which belong to that cluster –

* **VISUALISATION** –

1. Unclustered Data -



1. Clustered Data –



* **CONCLUSION** – As above we can see that the given dataset has been dividing into two clusters one of green colour and another red, using the K-Means clustering algorithm. The thick black circles represent the cluster centroids.
* REFERENCES –

1. <https://medium.com/machine-learning-algorithms-from-scratch/k-means-clustering-from-scratch-in-python-1675d38eee42>
2. <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>
3. <https://pythonprogramming.net/k-means-from-scratch-machine-learning-tutorial/>