# AIML UNSUPERVISED LEARNING ASSIGNMENT 1

2. Write your own code for K-means algorithm using two attributes namely average\_runs and bowling\_economy. Take K=2. Plot clusters on a scatter plot with X and Y being the two attributes namely average\_runs and bowling\_economy, respectively. Color data points belonging to the first cluster with red and the second cluster with blue. Copy the plot diagram in the word document and interpret the output. [3 points]

a.

Chart, scatter chart

Description automatically generated

Once the K Means started running , the centroids were adjusted in every iteration to arrive with the final clusters shown above . We can clearly see two distinct clusters where the bowling economy is less vs where it high . There is one data point which looks out of place who has although visually seems to be part of cluster 0 but has been marked in cluster 1 . highlighted in yellow above.

3. Redo question-2 on different values of K = 2,3,4,5. For each case, draw the plot of clusters as stated above. Visualize these plots, copy the plot diagrams in the word document,  and comment on which is better clustering (and reasons) based on visualization only. [3 points]

a.

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

K=2 K=3

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

K=4 K=5

Based on the above visualizations , K=4 gives a better cluster for the reasons below

a) Cluster 0 might represent average players who might not be good batsmen or bowlers

b) Cluster 1 might represent pure batsmen as the bowling economy is higher and average runs is also higher

c) Cluster 2 might most probably represents pure bowlers as the bowling economy is less and average runs is also less

d) Cluster 3 might represent all rounders who are good batsmen and bowlers

4. Write a few lines in a word document about the interpretation of the best clusters obtained. Also write a few statements about how these clusters can be useful. [2 points]

a.K=5 and K=4 as per my analysis might denote the best clusters . K=4 is the best cluster for reasons stated above and K=5 might also be helpful to find out players who are not really good performers or are average performers denoted by Cluster 0 and 4 . Here both the clusters consists of players whose bowling economy is higher and batting average is lesser , hence these people might not be very good players in reality subject to more data scrutiny