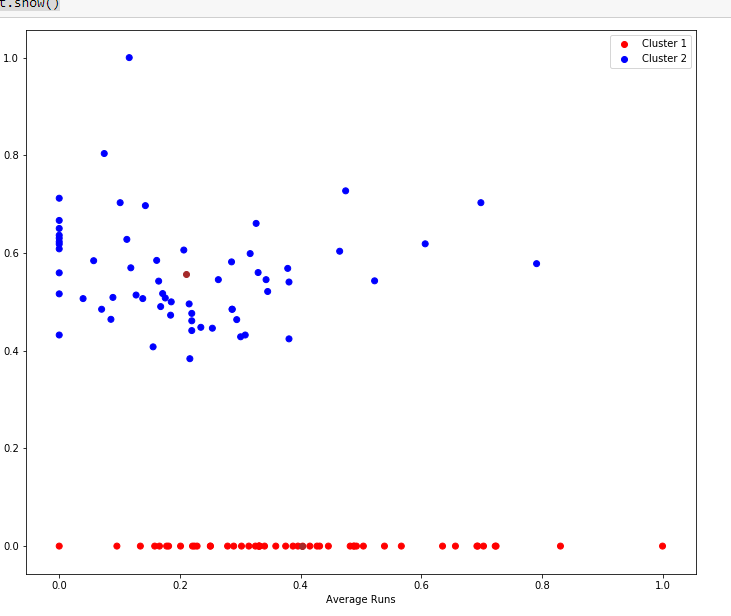
**K Means Assignment 1**

**Question 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]**

K=2

Bo

The above scatter plot is between Average runs(x-axis) and Bowling Economy(y-axis) of a cricketers dataset.

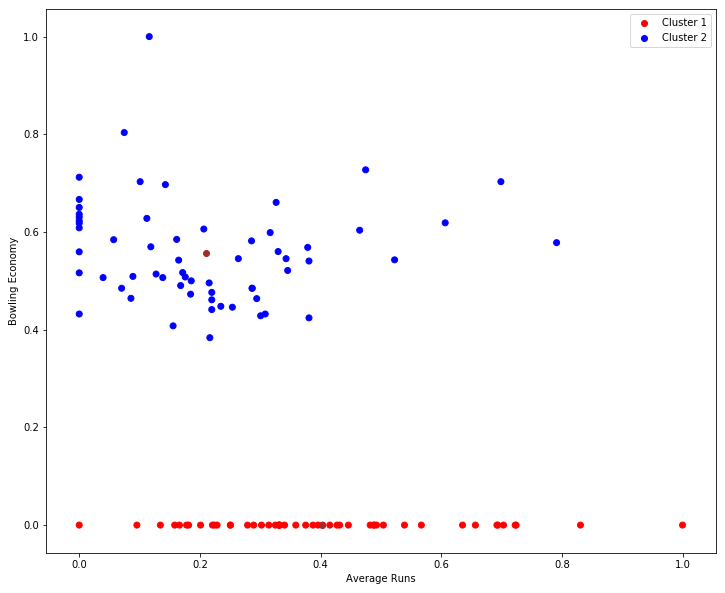
Cluster 1 (Red) and Cluster 2 (Blue) is the result of K means clustering with K=2.

The above visualization can be used in below scenarios.

* If team needs batsmen then we can choose from Cluster 2 with the average runs greater than 0.5
* If the team needs All Rounder then they can choose from Cluster 1 where average runs > 0.5 also if team needs bowler then they can choose from cluster 1

**Question 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]**

i. K=2



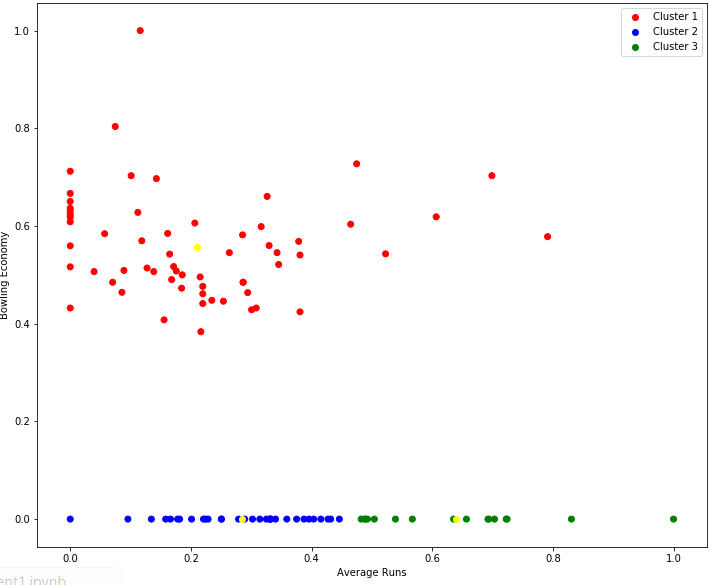
The above scatter plot is between Average runs(x-axis) and Bowling Economy(y-axis) of a cricketers dataset.

Cluster 1 (Red) and Cluster 2 (Blue) is the result of K means clustering with K=2.

The above visualization can be used in below scenarios.

* If team needs batsmen then we can choose from Cluster 2 with the average runs greater than 0.5
* If the team needs All Rounder then they can choose from Cluster 1 where average runs > 0.5 also if team needs bowler then they can choose from cluster 1

ii. K=3



The above scatter plot is between Average runs(x-axis) and Bowling Economy(y-axis) of a cricketers dataset.

Cluster 1 (Red) with Bowling economy>0 ,

Cluster 2 (Blue) with Bowling economy=0 and Average Runs<0.5,

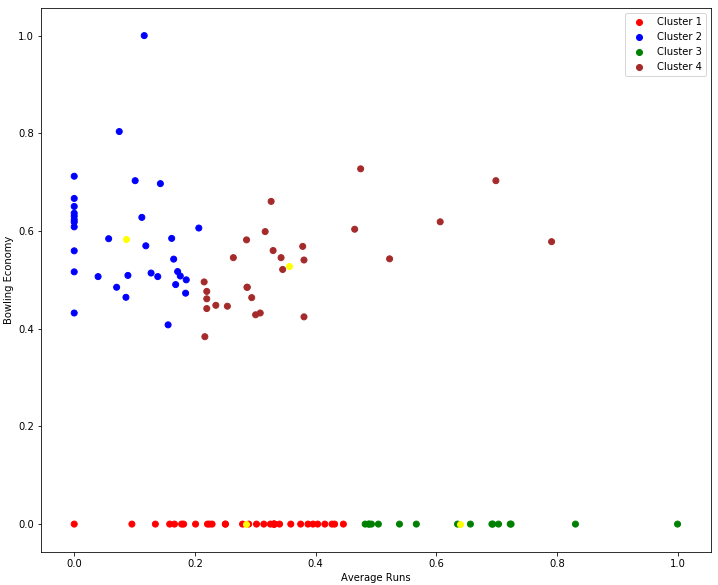
Cluster 3(Green) with Bowling economy=0 and Average Runs>0.5

is the result of K means clustering with K=3.

The above visualization can be used in below scenarios.

* If team needs batsmen then they can choose from Cluster 3
* If team needs All Rounder they can choose from Cluster 1 with average runs > 0.5

iii. K=4

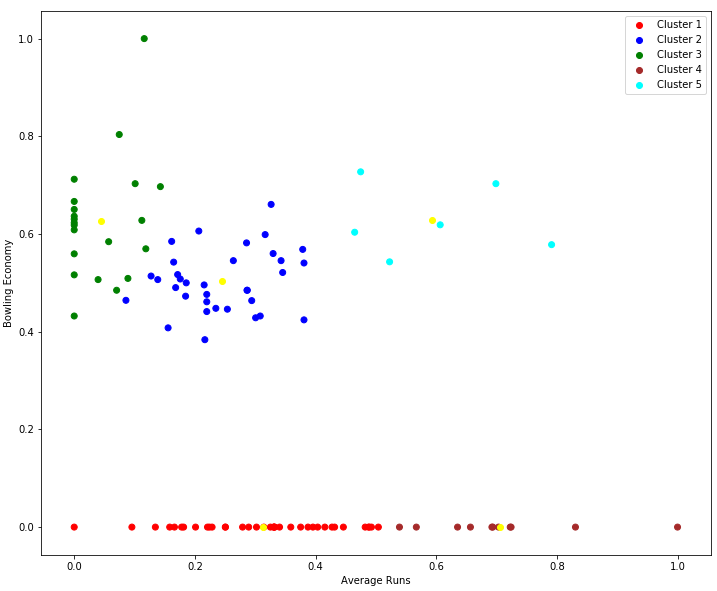


The above scatter plot is between Average runs(x-axis) and Bowling Economy(y-axis) of a cricketers dataset with K=4 clustering.

The above visualization can be used in below scenarios.

* If the team needs All rounder with high batting average and high bowling economy then they can choose from Cluster 4
* If team needs high bowling economy then they can choose from Cluster 2
* If team needs batsmen then they can choose from cluster 3

iv. K=5



The above scatter plot is between Average runs(x-axis) and Bowling Economy(y-axis) of a cricketers dataset with K=5 clustering.

Cluster 1 with Average Runs greater than 0.5 and Bowling Economy Zero

Cluster 4 with Average runs greater than 0.4 and Bowling Economy > 0.4

Cluster 5 with Bowling Economy greater than 0.4

The above visualization can be used in below scenarios.

* If team needs batsmen then they can choose from Cluster 1
* If team needs All Rounder they can choose from Cluster 4
* If team needs bowler, they can choose from Cluster 3

Among these K=2,3,4,5, K=5 clustering is the best of all , as the clusters formed when K=5 give more clear info on a cricketers performance which helps in choosing the most suitable player as per the requirement.

**Question 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]**

Among these K=2,3,4,5, K=5 clustering is the best of all , as the clusters formed when K=5 give more clear info on a cricketers performance which helps in choosing the most suitable player as per the requirement.

When K=5 the clusters formed are

* Cluster 5 → Standard batting average
* Cluster 4 → Good batting and bowling
* Cluster 2 → Low batting average
* Cluster 3 → High bowling average
* Cluster 1 → High batting Average

These clusters can be used in below scenarios:

* If team needs All Rounder then they can choose from Cluster 4
* If team needs good bowler then they can choose from Cluster 3
* If team needs batsmen then they can choose from Cluster 1
* Cluster 2 are low batting average players who might need more training and can be used as extra players.