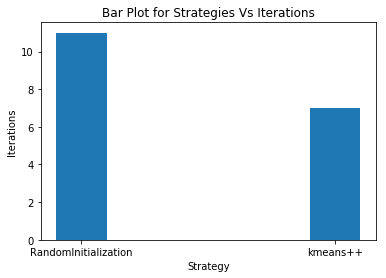
**#1**For the K-means algorithm using two attributes namely average\_runs and bowling\_economy, which you wrote in assignment-1, extend it with different strategies (take at least two strategies) for initial choice of cluster centroid. Take K=2.

(a) Which strategy achieves better convergence (iterations count) ? Draw bar plot (X-axis has different strategies and Y-axis has iterations count).  [2 points]

(b) Is it true for other values of K = 3,4,5,6 ? On the same bar plot, draw bars with different colors for these different values of K. [2 points]

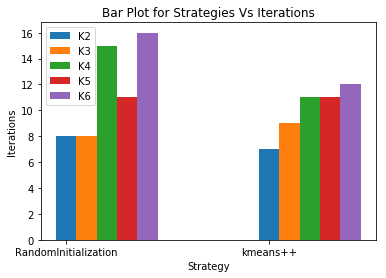
a) The following is the Bar Plot for Strategies Vs Iterations for Random Initialization as well as k means++

Even though the k means++ alogrithm is computationally intensive it achieves better convergence. So the trade off that happens for being computationally intensive is still fine as along as the convergence is better and faster.

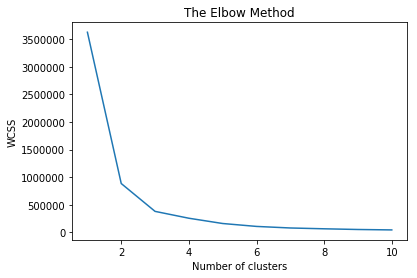


b) For different values of K=2, K=3,K=4,K=5, K=6 the following are the results.

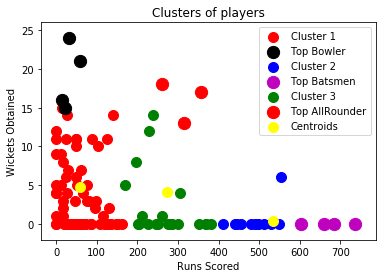
Kmeans++ certainly is a better algorithm compared to Random intialization though being computationally intensive it achieves better convergence.



**2** Use the elbow method to find the best value of K. Consider four attributes runs\_scored, average\_runs, wickets\_obtained, bowling\_economy.   Plot the related graph as discussed in the video lecture. [2 points]



**#3** Your task is to create a team comprising 4 batsmen, 4 bowlers, and 3 all-rounders.  Use appropriate K value. Consider two attributes runs\_scored and wickets\_obtained. Take the criteria for selecting top batsmen on the basis of runs\_scored, top bowlers on the basis of wickets\_obtained and top all-rounders on the basis of average of runs\_scored and wickets\_obtained.



**Top Batsmen**

1. Kane Williamson
2. Rishabh Pant
3. Lokesh Rahul
4. Ambati Rayudu

**Top Bowlers**

1. Andrew Tye
2. Rashid Khan
3. Hardik Pandya
4. Sunil 0rine

**Top All Rounders**

1. Andre Russell
2. Shane Watson
3. Sunil 0rine