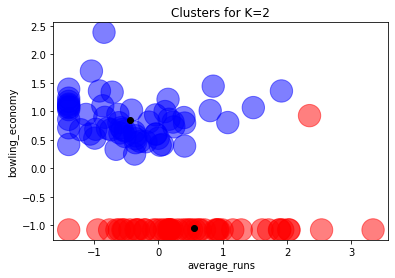
**2020AIML003 Unsupervised Learning Assignment1**

Visualisation of clusters when K=2:

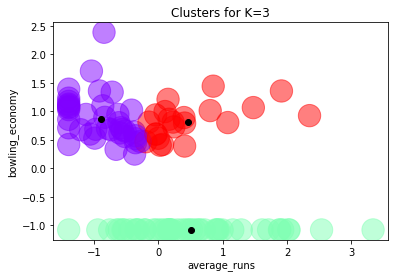
Note: Colours RED and BLUE chosen as per requirement!

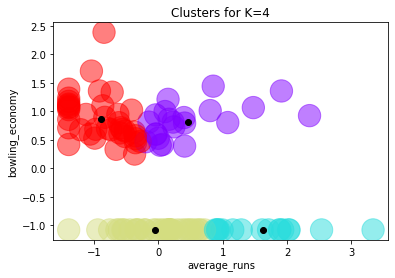


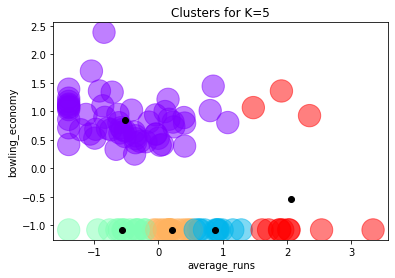
Interpretation of above output, when K=2:

* From the visualisation above, we can see clearly 2 clusters
* Also, when K=2, the K-Means algorithm has identified and separated the two clusters very well
* Barring one data point which should have been a RED (instead shown/marked as BLUE) all other data points seem to have been correctly segregated.

Visualisation of clusters when K=3, 4 & 5:

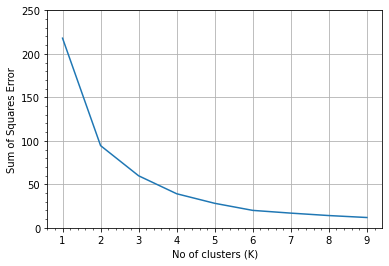






Comparing results when K=2, 3, 4 & 5:

* We see from the visualisation above mainly 2 clusters
* From the Squared of Sum Errors obtained for different values of K (from 1=10) we’ve plotted the graph below
* Clearly the Elbow of the graph indicates that when K=4/5 there the error has fallen down sharply and for K>5, there is no significant fall in the error
* So, K=4/5 seems to be most optimal value just going by the elbow method



Interpretation of the results obtained and how these clusters can be used

* We very well know that average runs and bowling economy figures in cricket are two completely independent things and hence no correlation is expected to be seen between them
* The visualization of the clusters also clearly indicates 2 clusters/groups that relate to the above fact.
* Hence, the clustering done when K=2, seems to be the best / most meaningful, though just based on error, K=4/5 seems to be better one.