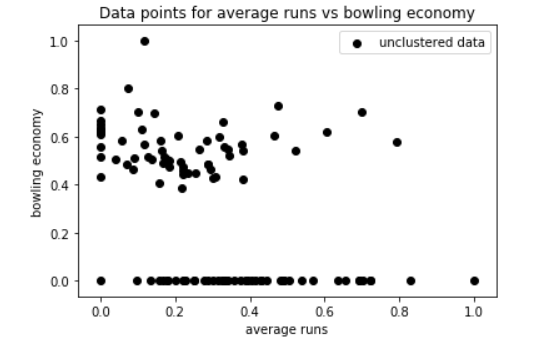
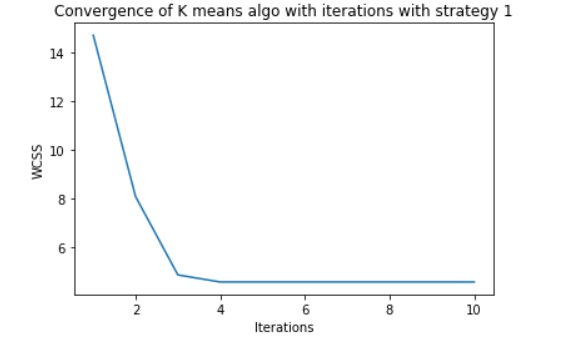
## Ans1 (a)

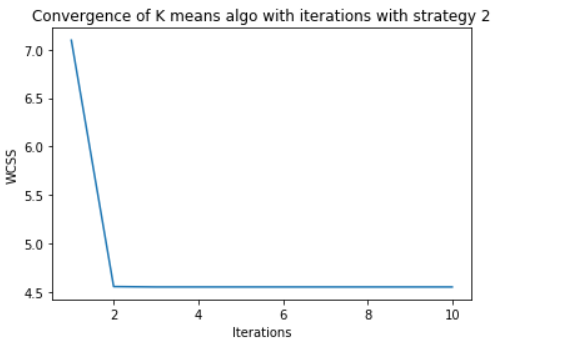
The Data for cricketers’ career history is plotted for average runs vs bowling economy.



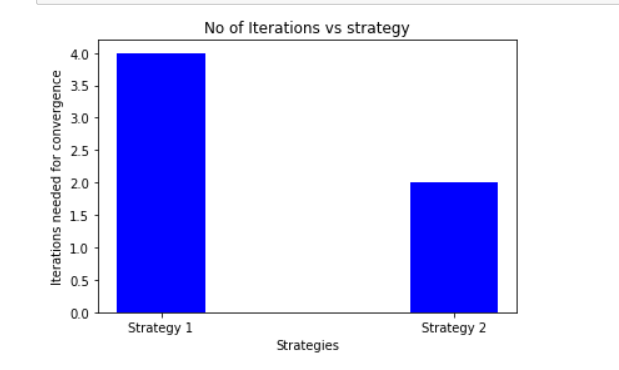
When the above data is clustered in 2 clusters using K-means algorithm with strategy 1, which is the centroids are initialized as mean of the data points plus a random vector, then the convergence of distortion functions(WCSS) with number of iterations is obtained as below:



When the above data is clustered in 2 clusters using K-means algorithm with strategy 2, which is the centroids are initialized as k-means++ method, then the convergence of distortion function (WCSS) with number of iterations is obtained as below:



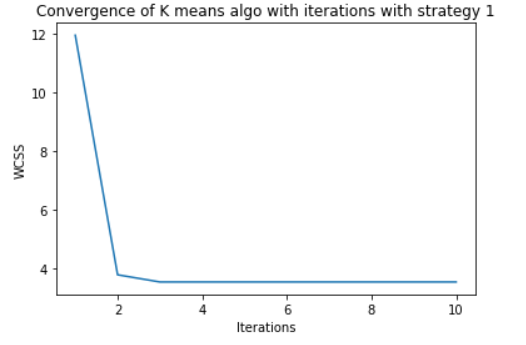
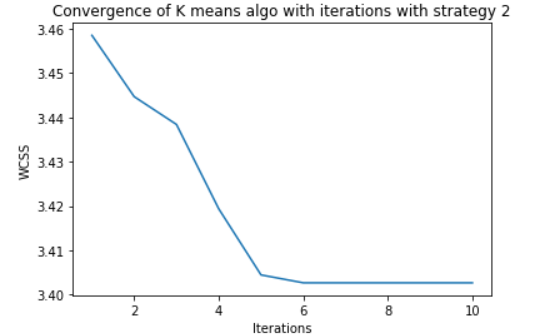
Number of iterations taken to converge for different strategies is plotted in bar plot as below:



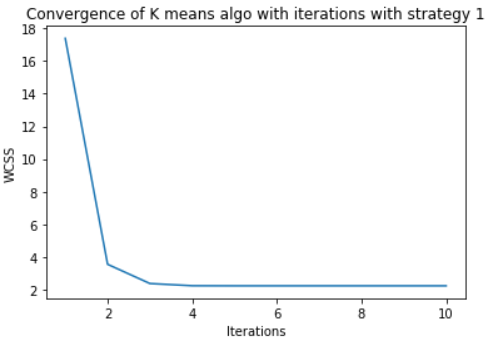
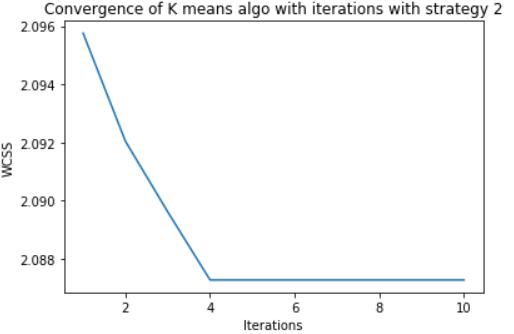
## Ans1 (b)

When the data is clustered in different number of clusters using K-means algorithm with strategy 1 and strategy 2, then the convergence of distortion function (WCSS) with number of iterations is obtained as:

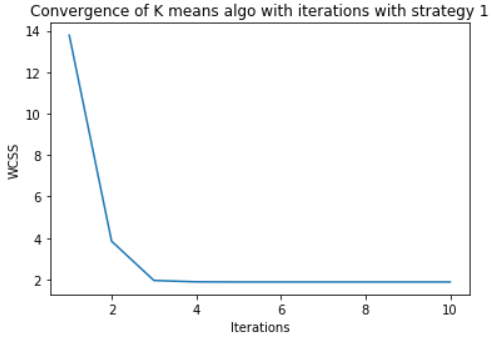
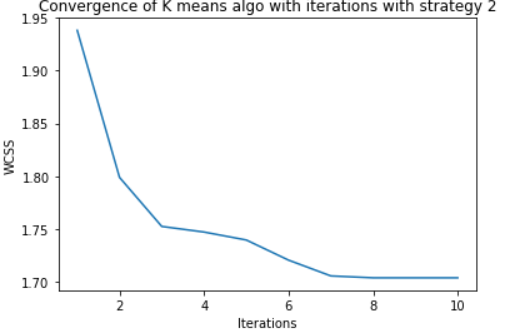
**For K =3:**

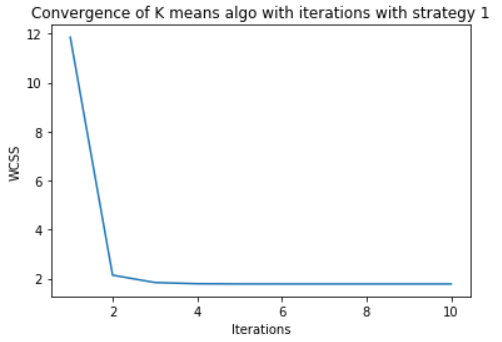
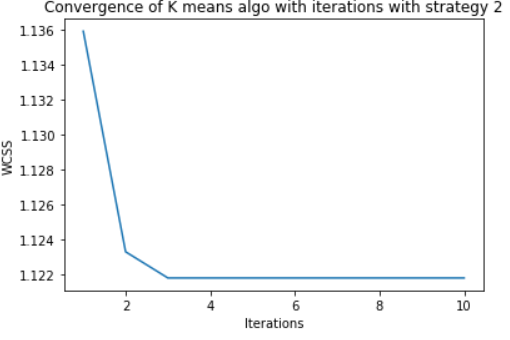
**For K =4:**

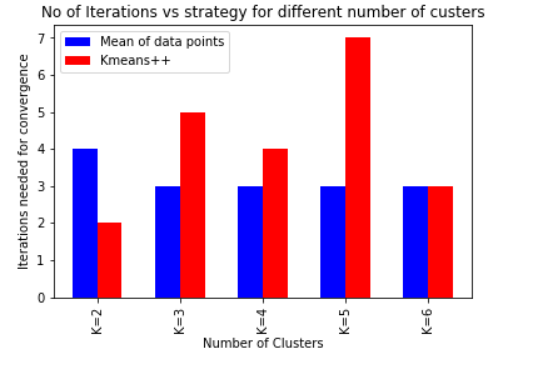
**For K =5:**

**For K =6:**

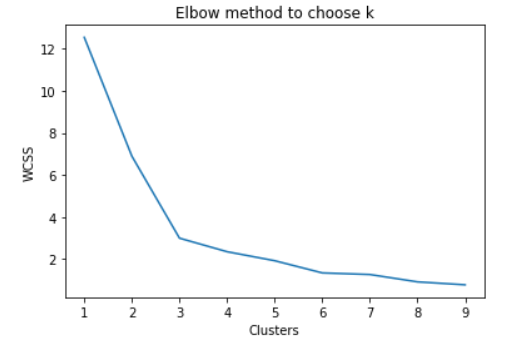
 

The bar plot for the number of iterations vs the number of clusters for both strategies is:



## Ans2.

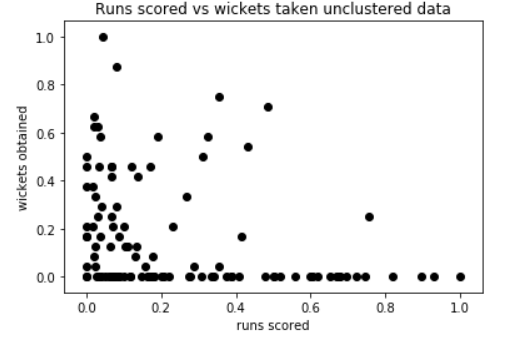
Using the elbow method to find the best value of K and considering four attributes as runs\_scored, average\_runs, wickets\_obtained, bowling\_economy :



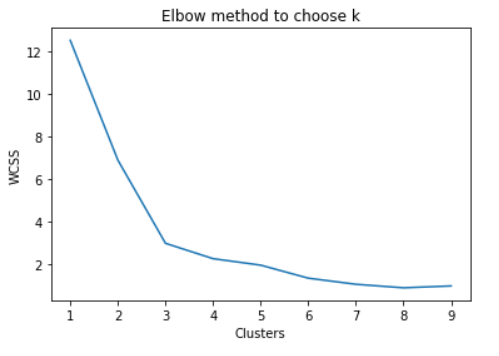
We can see from the elbow curve above that the best value for K is 3.

## Ans3.

We now plot the scatter plot for cricketers using features runs\_scored and wickets\_obtained.

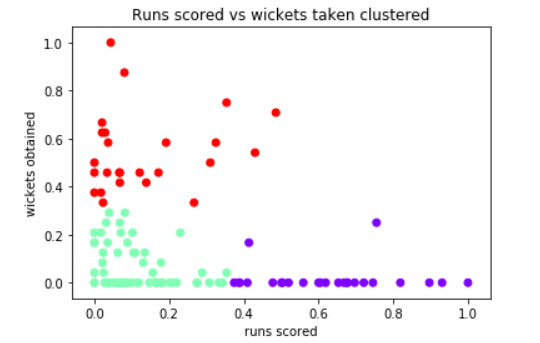


Using the elbow method to find the best number of clusters:



Based on the above diagram, we take the value of K = 3 to be best way to categorize players.

Players are categorized as:



The players highlighted in blue are the players which have high runs scored, so they are batsmen.

The players highlighted in cyan appear to be the players who are all-rounders.

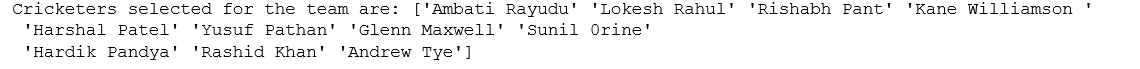
The players highlighted in red have high number of wickets obtained, so they are bowlers.

Now we create our team by selecting 4 batsmen, 4 bowlers and 3 all-rounders. They are selected from their respective clusters as:









The selected cricketers can be plotted in scatter plot as below based on the features taken:

