# Football Player Segmentation

**Abstract:**

Clustering is one of the most common exploratory data analysis technique used to get an intuition about the structure of the data. It can be defined as the task of identifying subgroups in the data such that data points in the same subgroup (cluster) are very similar while data points in different clusters are very different. In other words, we try to find homogeneous subgroups within the data such that data points in each cluster are as similar as possible according to a similarity measure such as euclidean-based distance or correlation-based distance.

**Problem Statement:**

Use kmeans clustering algorithm to check how the football players groups based on common attributes or similar interests.

**Variable Description:**

Complete Dataset including :

- Player personal attributes (Nationality, Club, Photo, Age, Value, etc.)
- Player performance attributes (Overall, Potential, Aggression, Agility, etc)
- Player preferred position and ratings at all position

**Scope:**

- Dealing with missing values
- Visual data exploration
- Data Pre-processing using R
- Using kmeans algorithms to check how well the football players cluster

**Learning Outcome:**

The students will get a better understanding of how to find a pattern in the data using K-means clustering algorithm and evaluate the model using R.