



## Hackathon Task Document

### About the Dataset

This dataset contains features of around 18,000 football players. You are expected to complete the following tasks. Use only numerical attributes to cluster the data.

[Click here to download the dataset](#)

### Task 1 - Data Visualization and Data Preparation

You can use matplotlib and seaborn for visualization. This is an open ended task. For your comfort, below mentioned are the basic visualization you can start with:

- Plot histogram of count of players based on some attributes like height.
- Plot histogram and kde plots on attributes like 'Value' and 'Wage'
- Use visualization techniques to find outliers like Ronaldo and Messi
- Distribution of players in different clubs/countries on the basis of some attributes
- And many more....

Try to perform below mention data transformations:

- Remove inconsistencies from 'Value' and 'Wage' columns
- Missing Values and Outliers treatment

### Task 2 - K Means

1. Implement k-means clustering. Choose  $k = 3, 5, 7$
2. Use only numeric attributes to cluster
3. Use elbow method and Silhouette Score to get optimal number of clusters
4. Analyse the results/clusters formed(in every case above) based on the following parameters:
  - a. How good are the clusters? Use inter and intra class similarities to measure the goodness of clusters
  - b. Find any hidden patterns if any

### Task 3 - Hierarchical Clustering

1. Cluster the data using Agglomerative method of your choice
2. Plot dendrogram
3. Analyse the cluster formed based on the following parameters:
  - a. How good are the clusters? Use inter and intra class similarities to measure the goodness of clusters
  - b. Find any hidden patterns if any

### Task 4 - DBScan



1. Use DBScan to cluster the data
2. DBScan requires 2 parameters - epsilon and minPts. Show all the experiments performed to arrive at the final epsilons and minPts
3. Analyse the cluster formed
  - a. How good are the clusters? Use inter and intra class similarities to measure the goodness of clusters
  - b. Find any hidden patterns if any

**NOTE:** Find the Instruction Document for the test [here](#). It will help you understand how to participate and submit this test.