**Group 5 - Project Proposal**



Group members:

## Khuseyma, Mohamed, Abdullahi, Motasim

**Project outline**

The concept we will be using is a custom related topic. We will be looking into the athletes within the NBA organisation, specifically their stats throughout the seasons and the results of utilizing predictive ML algorithms with reduced dimensionality NBA player data. Advanced statistical analysis has become a dominant trend in sports.

K-means clustering are to be applied to a dataset drawn from the aggregate season statistics for a single NBA season preferably (2020-2021) or (2021-2022) as those athletes are still active.

**Structure of the project**

1. The first section will provide background on common NBA statistical analysis. What data is valid?
2. The following section will describe the manner in which the data was organized, and the algorithms were applied
3. the third section will present the results.
4. The final section then provides conclusions to the work and possible future approaches.

**Dataset that will be used for the project:**

(The one we were initially meant to use) Kaggle link: [Nba 2020-2021 Season Player Stats | Kaggle](https://www.kaggle.com/datasets/umutalpaydn/nba-20202021-season-player-stats)

(The one we ended up using) https://www.kaggle.com/datasets/drgilermo/nba-players-stats-20142015/code?resource=download

**Question that we are researching**

1. Can machine learning algorithms be used to predict an NBA players statistics?
2. Do a player’s statistics directly effect a team’s winning %?
3. Do the minutes played by a player have any effect on their statistics?
4. Is there any relation between a player’s points per game and their height?

**Project task delegation:**

* Data cleansing (Motasim)
* Data model implementation/R-squared value (Abdullahi)
* K means and PCA/Proposal (Mohamed)
* Data model optimisation/PowerPoint presentation (Khuseyma)