**Cohort A Team 12**

**PLAY-BY-PLAY, PLAYER-BY-PLAYER – COLLEGE BASKETBALL ANALYTICS**

**Problem Statement and Motivation:**

The primary objective to do a sports-based project, is to explore and implement our findings and machine learning techniques within the sports industry. Our team is passionate and has experience in various sports which led us to choose this sports dataset. The entrenched culture of college sports in the United States is a key marker of a student-athlete's potential prior to their progression into professional sports. This culture serves as a source of local pride and personal accomplishment. The goal in the project is to assess the performance metrics of players from 2009-2021, extracting insights that can enhance our understanding.

**Dataset and Data Source:**

Kaggle - College Basketball: Dataset Shape - (61062 rows × 65 columns) [Click here](https://www.kaggle.com/datasets/adityak2003/college-basketball-players-20092021?select=DraftedPlayers2009-2021.xlsx)

**Proposed Methodologies:**

* **K-Means Clustering**: We'll use k-means clustering with an elbow plot to determine the optimal number of clusters. Players in the NCAA Basketball dataset will be grouped based on their proximity to centroids, which are recalculated iteratively until stable or until a set number of iterations is reached. This approach will help categorize players with similar attributes.
* **t-SNE Dimensionality Reduction**: To handle the diverse and potentially non-linear relationships in our dataset, we'll apply t-SNE for dimensionality reduction. This technique will help us identify significant variables by projecting data points to a lower-dimensional space, maintaining the proximity of similar points and aiding in the effective grouping of players.
* **Principal Component Analysis (PCA)**: We will use PCA for further dimensionality reduction, focusing on attributes that define player types, like shooting efficiency. This method will identify the most significant attributes in explaining variance, allowing us to create a model that clusters players based on their key abilities.

**Business Relevance:**

* **Drafting for professional teams**: College basketball players from differing conferences go to the NBA or different leagues overseas to play for professional teams.
* **Transfer window**: Allows coaches to decide which key players to acquire and offer scholarships during the transfer period.
* **Strategy**: Coaches can form different strategies for different games based on the opponents’ style of playing.
* **Sponsorships**: Companies can create Name, Image, & Likeness deals with top athletes to promote their products to fan bases.