**Problem statement-** Through our analyses, we aim to achieve the following objectives:

1. Predict if a footballer has high or low potential (Classification problem)
2. Predict the overall rating of a footballer (Regression problem)

**Dataset-** The dataset to be considered contains various attributes of European soccer players which is a blend of both categorical and numeric variables. Adhering to the SEMMA approach, we will conduct data processing and modelling techniques on these attributes to accurately predict the two target variables as mentioned below. The attribute ‘player\_api\_id’ acts as a primary key and can be used to merge the dataset with other tables. We will also be using player demographics datasets (country, team, league) for various exploratory data analyses.

**Observations-** 183,978 rows, 38 columns

**Predictor Attributes-** id, player\_fifa\_api\_id, player\_api\_id, date, overall\_rating, potential, preferred\_foot, attacking\_work\_rate, defensive\_work\_rate, crossing, finishing, heading\_accuracy, short\_passing, volleys, dribbling, curve, free\_kick\_accuracy, long\_passing, ball\_control, acceleration, sprint\_speed, agility, reactions, balance, shot\_power, jumping, stamina, strength, long\_shots, aggression, interceptions, positioning, vision, penalties, marking, standing\_tackle, sliding\_tackle, gk\_diving, gk\_handling, gk\_kicking, gk\_positioning, gk\_reflexes

**Target Variables-**

1. For the classification problem: Player\_potential
2. For the regression problem: Overall\_rating

**Dataset source-**

<http://sofifa.com/> : Players and teams attributes from EA Sports FIFA games.

<https://www.kaggle.com/hugomathien/soccer> : European soccer database.