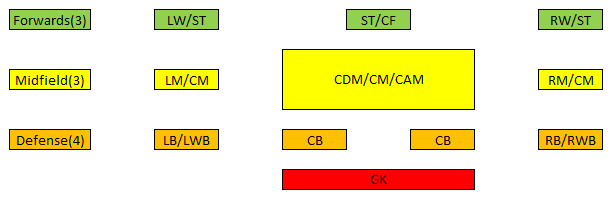
In this project, we will be looking at the player data provided by FIFA which contains information such as personal details, wages, physical attributes, technical skills, potential and their positional strengths. This is primarily data of FIFA 2018.Through this project, you will get a glimpse of insights behind the beautiful game and the kind of information and decisions a football manager goes through. Explore the data and attempt all the below asked questions in a step by step manner:

* Prepare a rank ordered list of top 10 countries with most players. Which countries are producing the most numbers of footballers that plays at this level?
* Plot the distribution of overall rating vs. age of players. Interpret what is the age after which a players stops improving?
* Which type of offensive players tends to get paid the most: the striker, the right-winger, or the left-winger? Visualize through a scatter plot
* Top 5 players for every preferred position in terms of overall as well as potential points. Who were the best in 2018? Who were destined to be the future superstars in that year?
* Which club(s) have the maximum share of players from England? Which club(s) have the maximum share of players from Spain? Which club(s) have the maximum share of players from Germany?
* Are the wages of a player influenced by the potential of a player? Check it out for players with age Between 16 to 28?
* Do Strikers score higher on "Aggression" than defenders do? Group both the set of players (from an overall score of 80 to 85) and compare their average aggression levels. Which particular position has the highest aggression as a given (players with an overall score of 80 to 90)?
* Which of the player characteristics (skills) are correlated on an aggregated level (check for players between overall score of 75 to 90). Comment on interesting and obvious insights.
* There is an additional worksheet provided called Football Leagues which have information about the clubs which are part of some of the famous football leagues such as EPL, Ligue1, Laliga, Bundesliga, Serie A, Eredivise,MLS, Premiera Liga. Use this information to create a new variable called League type. Map the clubs by leagues and put all the remaining clubs in a type called ‘Others’.
* Figure out the top 30 clubs by their average spend on wages. Create an additional variable called club category where category 1 represents the top 10 club based on the average wages, category 2 represents clubs that lie in the range of 11 – 20, category 3 represents the club that lies in the range of 21- 30 and category 4 represents all the left over clubs.
* Use the additional 2 variables that you have created to predict the wages of a player using linear regression. ( To use Categorical variables in regression try converting them to dummy variables, refer to the below link: <https://www.youtube.com/watch?v=r39JOUiFdIs>)
* As a National coach of France team you want to compare the national team of England, Spain, Italy and Germany to understand the competition. The formation of the teams is restricted to 4-3-3 (4 defenders, 3 midfielders, 3 forwards, 1 Goal Keeper) and players with overall value of more than 75 are preferred, now form the best team for each of the mentioned countries and compare them. Note down all the insights that you as a business analyst should share with the coach.

Note:-

* Pick out the best players based on the overall rating while selecting the players, for example if you need to pick 3 forwards then filter the data for all the forwards, sort them based on overall rating and pick the best 3, in case their overall rating ties, look at the potential rating, if that is also the same then choose on the basis of the value of the player
* When we talk about formation, the nomenclature is as follows (Number of defenders- Number of midfielders- number of forwards), goalkeeper will always be there in the team hence we don’t represent it in the formation.

4-3-3 formation:



Based on the positional requirement the above mentioned positions can be treated equally while shortlisting the squad. E.g. LW / ST – LW can be treated equally as striker