ECE 225
Probability and Statistics for Data Science
Final Project
Deriving Insights from Soccer Player Database

Apoorva Pushkar Gokhale - A53320452
Chaitanya Sharadchandra Patil - A53311586

Soccer is the most popular sport in the world. It consists of a seemingly simple set of rules, but has a lot of underlying intricacies.

In this project, we derive insights from a large dataset of soccer players. We make use of the comprehensive set of more than a 100 attributes made available by the FIFA player dataset to identify and visualize trends in the data, such as the correlation of certain attributes with others.

- The dataset boasts of certain key attributes that define the success of a soccer player such as the overall rating and the market value. Since we know that these depend on other attributes like age, speed, ball control, etc., we use the dataset to quantify the extent to which each of these attributes affect the overall rating and the market value of the player.
- We also make use of classification algorithms such as One-Versus-All Logistic Regression to identify the most important attributes that influence the best position that that player can play in.

Other insights obtained from the data include:

- By plotting the average market value per position we estimate the relative value and importance given to players' positions in the market.
- By plotting the average attributes of players belonging to a particular region we identify the talent pool the nation has to offer
- By calculating the variance in the rating of the player across various positions, we try to use it as an index to quantify the versatility of the player.
- We also find the top valued clubs based on the players that they have under contract.

Knowing these trends in data can greatly help team managers, administration, scouts as well as players, make important decisions for their club and the overall quality of the game worldwide.

Dataset Source : https://www.kaggle.com/stefanoleone992/fifa-20-complete-player-dataset