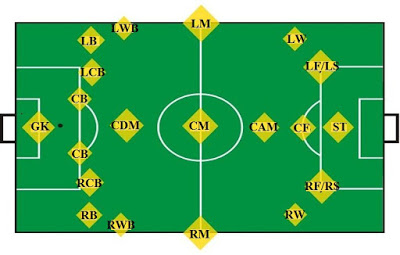
Soccer Stats Analysis:

Data Description:

The data set is comprised of 15000 soccer players with 108 columns from the Fifa website.

Data Preparation

In order to clean the data and reduce the amount of null values, I divided the data into 2 big groups: the field players(field\_clean.csv) and the goalkeepers(gk\_clean.csv). Also, the data set had a lot of columns with variables values; In other words the baseline for a player speed could be 80 + 3 for the adjusted monthly performance. For this data set I kept it to the baseline performance and removed the adjusted performance.



Data Analysis

For the data analysis, I narrowed down the 108 different columns to 16 essential ones and I attempted to determine how much each of the columns where correlated to the overall player rating.

Columns:

|  |  |
| --- | --- |
| age | The player’s age |
| Height\_cm | The player’s height in cm |
| Overall | The player’s overall rating |
| Weight\_kg | The player’s weight in kg |
| Potential | The player Potential |
| Value\_eur | The player value in euro |
| Wage\_eur | The player’s salary |
| International\_reputation | The player’s international performance |
| Weak\_foot | The player’s weak foot ability |
| Skill\_moves | The player’s skill move ability |
| Pace | The player’s pave |
| Shooting | The player’s shooting |
| Passing | The player’s passing |
| Dribbling | The player’s dribbling |
| Defending | The player’s defending ability |
| Physic | The player’s physical rating |

Results:

I grouped the results into 4 files: field\_corr.csv, attack\_corr.csv, defense\_corr.csv, midfield\_corr.csv. They each have 2 columns with the same headers: player\_attr(the player attribute) and corr\_to\_over(correlation to overall)

Observation and insights:

What we observe after running the correlation on every field player is that there is not a strong correlation between most of the attributes except for potential, value and salary. In other words, there isn’t a standard definition of an excellent player.

After we split up the correlation by positions (offense, midfield, defense) the picture becomes clearer.

We notice that passing is very important in the midfield with a correlation of over 88%, dribbling and shooting are the most important skills for attacker, and physicality is vital for every defender.