FIFA 19 Player Analysis

Mauro Mujica-Parodi III, Oluwafemi Shobowale, Eric Vistnes, Bae Min Jung, Tsu-Erh Lin

I. Abstract

This project aims to analyze players' attributes and position for every player registered in FIFA 2019 using multivariate analysis. Soccer players' performance and ability depend on multilateral elements such as technicality, tactics,physiological and psychological components. All these elements primarily depend on each other either. Principal component analysis (PCA) was used for dimensionality reduction and interpretation. Exploratory factor analysis (EFA) was used to understand the relationship between players' reputations and compensation. Canonical correspondence analysis (CCA) was used to determine the association between a player's skills attribute and the position ratings. Linear regression was used to determine what variables affect players' overall rating in FIFA and how vital the variables are. The PCA categorized the variables into five components (Attacking, Defending, Hype&Value, Players Physic, and Others). EFA proposed that reputation and compensation are somewhat related but not linearly. CCA proposed that a player's skills are indeed associated with their position rating and vice-versa. Linear regression tells us that age plays an essential factor to soccer players, and skills/attributes correlate to players' position. The analysis provides essential information about dimensionality reduction, more data insight, understanding the relationship between reputation and compensation, and what skill/ability affects players' overall score/ranking in FIFA 19.

Keywords:principal component analysis,exploratory factor analysis, canonical correspondence analysis, linear regression,multivariate analysis

II. Introduction

Soccer is the most popular sport globally, with an estimated fan base of 3.5 billion fans worldwide.Soccer is is played on a large field, a team is made up of 11 players per side. Goal post are located on the opposite side of the field each team tries to score the other by putting the ball on the post, while the other defends the goal post. The team with the most score wins the game after 90 minutes. Players who score the goals are called strikers while those who stop the attacker from getting into the goal post areas are called defenders, a player who stays in the middle of the field to assist the attackers and defenders are called midfielders and the player who protects the goal is called the goal-keeper. Being able to predict a soccer match is very important to

team managers, coaches, fans, scouts, betting companies, sponsors, and club owners.

All fans have different reasons why they love the sport. As a competitive game, soccer requires tactical, technical, and physical performance from players. Sports scientists have conducted studies to build models to help improve players' training or help distinguish between players of different abilities.

The worldwide love for soccer also carries over to the love of FIFA 19, a soccer video game. There is a new edition of FIFA that comes out every year, and fans pore over advanced statistics for each player in order to build their ideal teams. Some casual fans may simply add their favorite players to their ideal team, but more dedicated fans may use more advanced analysis to choose which players fit best on their team. FIFA 19 provides all the information needed to complete this task, overwhelming fans with numbers about the player's skills, positions, and more.

The game of soccer has many positions and corresponding skills. Each position requires different skills and some positions, such as goaltender, have almost no overlapping skills with positions such as attacker. This dichotomy makes it important to be able to understand how each individual score in FIFA 19 contributes to the overall player score. Without this understanding, it would be easy to dismiss skilled but specialized players who have a lower overall player ranking. Knowing the difference between bad players and specialized players can make the difference in a match, whether in real life or video games.

Because players are so specialized and can be skilled while still ranking poorly in FIFA 19, it is important to know what is similar between all good players. FIFA tracks an enormous amount of data on players and games, so some skilled players are bound to show commonalities. Though wage and reputation of a player are a good indicator, there are other signs, such as physical traits, that allow us to better generalize players.

Literature Review (Background on the application)

Many factors are relevant in determining the success of a soccer player. For instance, psychological skills such as interventions can influence the efficacy of sports (Thelwell, 2006[1]). The physical composition of the body is also relevant in the preparation of players for competitive performance (Carling, 2010[2]), it is an important component of fitness like running and jumping. Hence, in this research, we also focused on variables such as weight and height.

[1] Thelwell, R. C., Greenlees, I. A., & Weston, N. J. (2006). Using psychological skills training to develop soccer performance. *Journal of applied sport psychology*, *18*(3), 254-270.

[2] Carling, C., & Orhant, E. (2010). Variation in body composition in professional soccer players: interseasonal and intraseasonal changes and the effects of exposure time and player position. *The Journal of Strength & Conditioning Research*, *24*(5), 1332-1339.

In the game of soccer there are various skills, strategies, and physical elements that are required to perform well (Joo, C.H., & Seo, D. I, 2016[3]). For instance, a midfielder need agility and must be adept at long and short passing, and a defender should be able to jump high (Kim, 2000)[4]. We therefore must understand that skill levels among different positions can be dependent variables in and of themselves.

Additional studies (Abdullah, 2016[5]) incorporated the presence of competition, a teams' history and a players' behavior in previous matches as an input to forecast a soccer players' performance. The authors' perspective was that the nature of the problem limited the applicability of standard techniques and as such, they developed a soccer team and soccer player performance-forecasting model based on Artificial Neural networks. Simulation results indicated that the proposed model could be classified as a stable prediction model especially for soccer team's status and performance, achieving a high accuracy rate up to 95%.

Research Questions

We analyzed the FIFA 19 complete data set[6] from Kaggle. The dataset includes the detailed attributes, wages, value, weight, height, age, and position score for every player registered in the latest edition of FIFA 2019. The variables include seven categorical variables and 81 numerical variables that include over 650 teams and 18,200 players. In essence, the dataset contains every player from every team that plays under FIFA, and is constantly updated. Though this data is from a previous year, this allows future analysis using our work to apply this data to real world information such as wins and game statistics.

One of the primary goals of this research paper is to analyze how various features such as skills, positions, and stats, affect each players' performance. Some skills or positions may contribute to higher overall scores while other, just as important skills, may not contribute as heavily. In order to understand how to best rank players and build a team in FIFA, we must know how the scores are calculated.
Secondly, we focus on finding commonalities in the features of our data in order to reduce dimensionality. The data set is large and would benefit from combining features as possible while maintaining a large proportion of variance explained.

The third objective of this paper is to examine how typically non-related features share explained variance. In particular, are reputational features related to skills, positions, or even demographic information? This could lead to interesting insights in

[3] Joo, C. H., & Seo, D. I. (2016). Analysis of physical fitness and technical skills of youth soccer players according to playing position. Journal of exercise rehabilitation, 12(6), 548.
[4] Kim YK. A fitness profiles of the professional soccer players by each position. Korean J Sports Med. 2000;18:217–226.
[5] Abdullah, M. R., Maliki, A. B., Musa, R. M., Kosni, N. A., & Juahir, H. (2016). Intelligent prediction of soccer technical skill on youth soccer player's relative performance using multivariate analysis and artificial neural network techniques. *International Journal on Advanced Science, Engineering and Information Technology*, 6(5), 668.
[6] https://www.kaggle.com/karangadiya/fifa19

what leads to better reputation or compensation, or lead to rejecting that only certain positions in soccer lead to a better reputation.
Finally, we examine whether there are ways to combine the features of the data into factors or components that are not currently used by FIFA. Given that there may be correlation between features that we previously thought were unrelated, it could change the understanding of FIFA scores if there are new ways to characterize players.

III. Methods

PCA

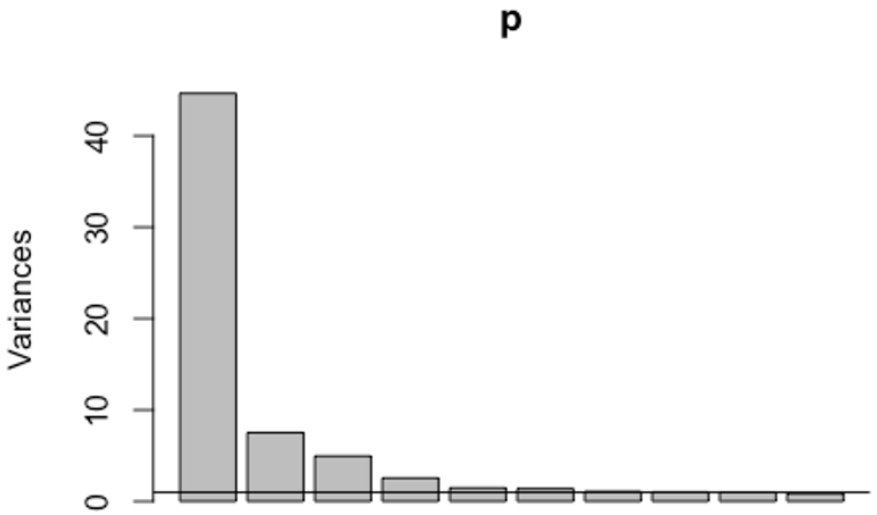
Three multivariate analyses were performed on the dataset, which is principal component analysis, exploratory factor analysis and canonical correlation analysis. Principal component analysis was applied to reduce the dimensionality in the data,interpret the data and identify variability among the skill and position. All the non-numeric variables were truncated from the original dataset to perform the analysis. Since all variables had different weights we had to normalize the data. By running the PCA we group positions and skills that have the same information, this also allows us to reduce the dataset with minimal loss.

Before performing the PCA, the dataset must be factorable. To confirm the factorability of the data, we used the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity. Barlett's test of sphericity determines if the correlation matrix is an identity matrix. The KMO indicates the proportion of variance in the variables caused by the underlying factor. A KMO close to 1.0 means factor analysis may be valid with data, and KMO less than 0.5, the results might not be accurate.

EFA

Before we began to conduct Exploratory Factor Analysis, we first conducted Principal Component Analysis to determine an optimal number of factors. As there is some multicollinearity in the data, we chose to use oblique rotation. Our tests of stability include Bartlett's test of Sphericity and Kaiser-Meyer-Olkin (KMO)

From our PCA, we found that we should use 5 factors in our Factor Analysis to properly explain the variance of the data. From the Scree plot below, the number is harder to gauge. We could choose between 3 and 6 factors easily, leaving us to test multiple configurations of factors and examine the variance explained. 9 factors explain 80% of the variance in the data, while 4 factors explain 79%. However, the gains of those last 5 factors start to become marginal, leaving us with the best fit of 4 factors.



< Figure 1: PCA Scree plot >

Though this diverges from the PCA that gives us 5 components, we can interpret the 4 factors easily and see that they explain different overarching factors. We used a factor loading cutoff of 0.4 and saw that most features were loaded into a factor- and features not loaded could be explained by features that were.

Canonical Correlation Analysis(CCA)

In the previous analysis, we were able to identify correlation between two variables. However, that does not tell us the correlation between groups of several variables. Therefore, we used CCA twice to see the correlation between multiple sets of variable groups using correlation coefficients.

CCA: Physical Traits and Player Skills

To run CCA, we needed to consider how to derive one variable from each variable group. We categorized variables into two groups: physical traits and player skills. Physical traits include six variables, which are weight, height_cm, strength, agility, stamina, and vision. Player skills include four variables, which are short_passing, long_passing, shot_power, and long_shots. Then we generated canonical correlation coefficients and scores to see how they are correlated.

CCA: Position Ratings and Player Skills

This piece of analysis was focused on the associations between two categories of variables: position ratings and player skills. Position ratings include 26 variables, which are *ls, st, rs, lw, lf, cf, rf, rw, lam, cam, ram, lm, lcm, cm, rcm, rm, lwb, ldm, cdm, rdm, rwb, lb, lcb, cb, rcb, and rb*. Player skills include 34 variables,

which are *crossing, finishing, heading_accuracy, short_passing, volleys, dribbling, curve, fk_accuracy, long_passing, ball_control, acceleration, sprint_speed, agility, reactions, balance, shot_power, jumping, stamina, strength, long_shots, aggression, interceptions, positioning, vision, penalties, composure, marking, standing_tackle, sliding_tackle, gk_diving, gk_handling, gk_kicking, gk_positioning, and gk_reflexes*. After creating the new dataframes from the resulting PCA scores, CCA was conducted using the *YACCA* library.

Regression

For linear regression, we want to answer the question that for all players what factors will be significant if we try to build a regression model to predict the player's potential score. On the other hand, in soccer, specialized training for players in different positions is required to improve their performance (Joo C.H, 2016[7]). Therefore, the other question is, will the result be different if we split players into three different positions (defender, midfield, offensive, and defensive positions).

To build the linear regression model, we used both manual and auto selection methods to find the best fit model and used cross-validation to train the model. To focus on players in different positions, we split players into three positions. Defender, we included center back (CB), fullback (LB, RB), and Wingback (LWB, RWB). Midfield positions included defensive midfielder (DM), central midfielder (CM), attacking midfielder (AM), and left/right midfielder (LM/RM). In the offensive position, we included striker (ST), center forward (CF), left winger (LW), right-winger (RW), right forward (RF), left forward (LF), right striker (RS), and left striker (LS). Finally, the last position is defender - Goalkeeper.

IV. Discussion and Results

Principal Component Analysis (PCA)

Before primary analysis, we did the factorability test on the dataset. Bartlett's test of sphericity (p=0.0001) proves that the data are not correlated. The kaiser-Mayer-Olkin (KMO=0.5) measure of sampling adequacy implies that other variables can explain the correlation between pairs of variables. After passing the factorability test, the dataset can undergo further analysis. After running the PCA, we used the scree plot to select five components. We use the "PROMAX" rotation to interpret the components.PC1(Attacking) this component has all attacking skills and positions, PC2(Defending) this component has all defending position and skill, PC3(Hype & Value) this component deals with players value, wage, international reputation, release clause, and potential.PC4(Player's physic) this component has all physical traits like; weight, height, strength, and balance. PC5(Others) this component has

[7] Joo, C. H., & Seo, D. I. (2016). Analysis of physical fitness and technical skills of youth soccer players according to playing position. *Journal of exercise rehabilitation*, *12*(6), 548.

other stand-alone features like age, reactions, jersey number, and composure. The 5 PCA tells a sum of 82% variance. The attacking component, the variable with the highest loading, is finishing(1.241); this implies that as an attacker scoring matters. In the attacking component, interception(1.159) has the highest score; this means that defenders must get the ball from strikers. In the value & hype component, the players' value(1.055) matters more. And when it comes to physics, weight (0.823) matters most, and for others, the player's reaction(0.956) is essential.

```
Loadings:
                       RC1    RC2    RC3    RC4    RC5
special                 0.612
skill_moves             0.815
ls                      0.990
st                      0.990
rs                      0.990
lw                      0.905
lf                      0.945
cf                      0.945
rf                      0.945
rw                      0.905
lam                     0.892
cam                     0.892
ram                     0.892
lm                      0.845
lcm                     0.717
cm                      0.717
rcm                     0.717
rm                      0.845
crossing                0.580
finishing               1.236 -0.516
heading_accuracy        0.604  0.417          0.486
short_passing           0.586
volleys                 1.125
dribbling               0.892
curve                   0.818
fk_accuracy             0.750
ball_control            0.827
acceleration            0.570                -0.401
sprint_speed            0.585
agility                 0.565                -0.495
shot_power              1.010
long_shots              1.048
positioning             1.062
vision                  0.695
penalties               1.099
gk_diving              -0.705
gk_handling            -0.701
gk_kicking             -0.695
gk_positioning         -0.701
gk_reflexes            -0.702
lwb                            0.706
ldm                            0.769
cdm                            0.769
rdm                            0.769
rwb                            0.706
lb                             0.783
lcb                            0.899
cb                             0.899
rcb                            0.899
rb                             0.783
aggression                     0.770
interceptions          -0.431  1.150
marking                        1.107
standing_tackle        -0.434  1.180
sliding_tackle         -0.490  1.199
potential                              0.846
value                                  1.053
wage                                   0.970
international_reputation               0.740
```

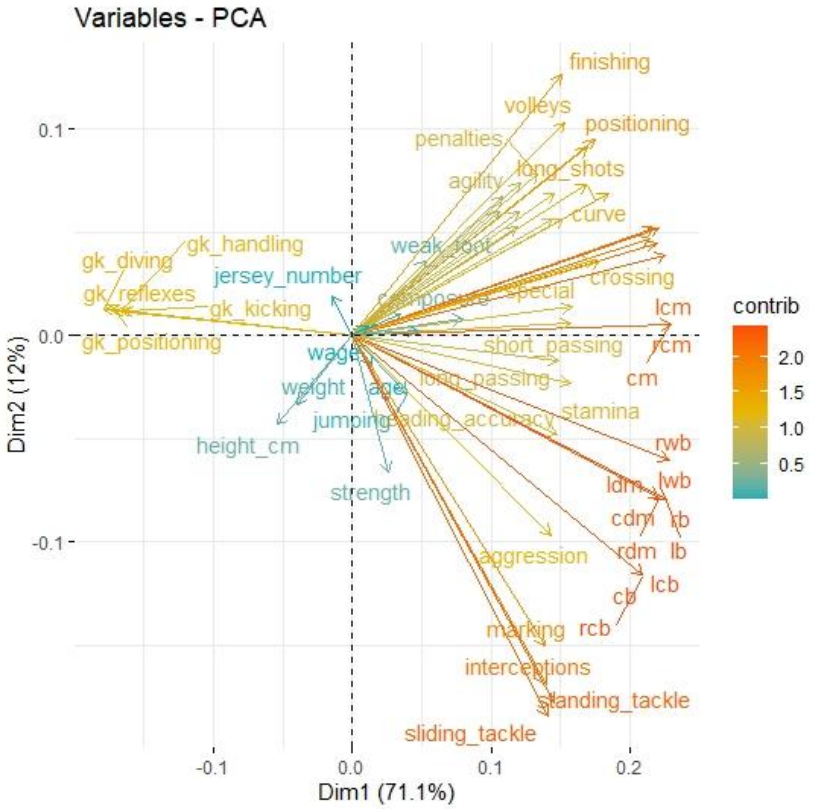
< Table 1: Variables >

release_clause			1.059		
weight				0.818	
balance				-0.685	
strength				0.812	
height_cm				0.818	
age			-0.403		0.961
overall			0.479		0.514
reactions					0.556
weak_foot	0.424				
jersey_number					-0.422
long_passing		0.460			
jumping					
stamina	0.409	0.490			
composure	0.466				0.462

< Table 2: PCA Loadings >

	Attacking	Defending	Hype&Value	Player's Pysic	Others
SS Loadings	30.964	15.811	4.979	3.758	3.260
Pro. Var	0.430	0.220	0.069	0.052	0.045
Cum. Var	0.430	0.650	0.719	0.771	0.816

< Table 3: PCA Loadings Summary >



< Figure 2: PCA result using eigenvalue >

Exploratory Factor Analysis (EFA)

Our factors can be interpreted as Position/Skill, Past Statistics, Compensation/Reputation, and Demographics. Position and Skill explain features of what position a player plays, as well as certain skills on the field such as passing or finishing. Past Statistics explains factors such as interceptions, tackling, and so on. Compensation and Recognition explain features such as wage, international reputation, and reactions. Finally, Demographics explain features such as height, weight, and physical features.

Below is the factor loading table with a cutoff of 0.4 that shows that even with generally strong loading, there was cross-loading among features. We choose the strongest loading, but some of the cross-loadings show interesting patterns.

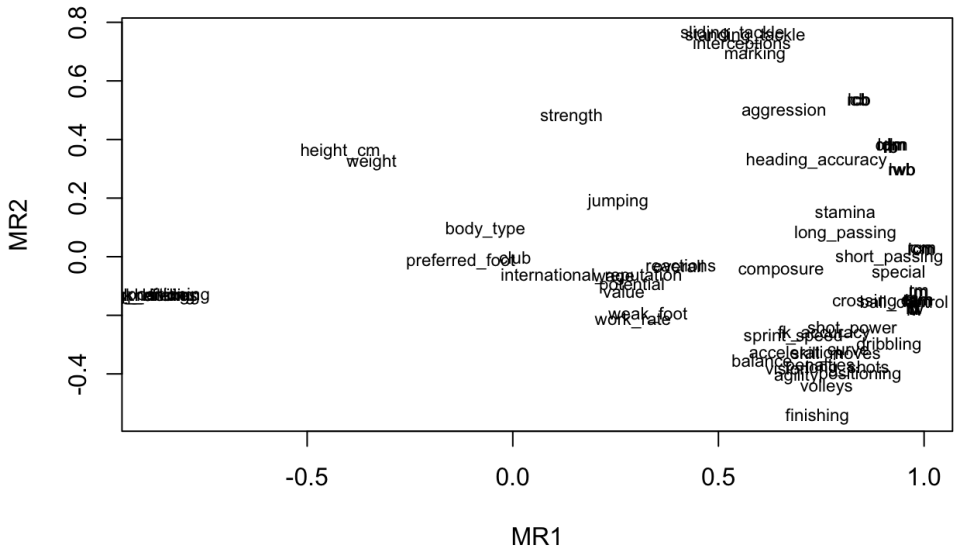
item	MR1	MR2	MR3	MR4
lcm	0.994			

cm	0.994			
rcm	0.994			
lm	0.987			
rm	0.987			
ldm	0.921	0.383		
cdm	0.921	0.383		
rdm	0.921	0.383		
lb	0.919	0.375		
rb	0.919	0.375		
short_passing	0.916			
dribbling	0.915			
gk_diving	-0.875		0.359	
gk_reflexes	-0.875		0.361	
gk_handling	-0.874		0.359	
gk_positioning	-0.871		0.364	
gk_kicking	-0.871		0.355	
crossing	0.855			
positioning	0.845	-0.401		
lcb	0.842	0.539		
cb	0.842	0.539		
volleys	0.763	-0.444		
fk_accuracy	0.757			
penalties	0.748	-0.371		
finishing	0.741	-0.543		
heading_accuracy	0.738	0.327		-0.381
agility	0.689	-0.407		
acceleration	0.688	-0.323		
sprint_speed	0.682			

vision	0.668	-0.381		
aggression	0.659	0.499		
composure	0.652		0.51	
balance	0.606	-0.354		0.448
weak_foot	0.329			
sliding_tackle	0.534	0.762		
standing_tackle	0.565	0.755		
interceptions	0.557	0.727		
marking	0.589	0.692		
overall	0.404		0.834	
reactions	0.409		0.716	
value			0.66	
wage			0.628	
international_reputation			0.572	
potential			0.523	
weight	-0.343	0.322	0.424	-0.527
strength		0.48	0.394	-0.517
height_cm	-0.419	0.361	0.336	-0.517

< Table 4: EFA Loadings >

These factors provide interesting insight into the FIFA 19 scores, as it shows that there is little to no differentiation between skills and positions in terms of how FIFA treats them. We could not group features in separate factors that better define the player, but rather, show that all the skills and positions define the player. Looking at a plot that shows explained variance in Factor 1 and Factor 2 shows that most features are explained by Factor 1.



< Figure 3: EFA factor 1&2 Loadings >

However, the value of this insight is that the overall scores then do not properly take into account specialized players. As mentioned previously, goalkeepers are players with extremely specialized stats, and scores of 0 in many of the features in the data set. This does not make the player worthless, but rather, shows that the FIFA 19 score does not properly encapsulate the player's value. Luckily, other factors, such as Past Statistics or Compensation are means outside of FIFA's ranking system that can capture a player's worth.

Finally, an interesting insight into the data comes from looking further at the cross-loading. Weight and height both cross-load into the Reputation and Compensation Factor. In many ways, this is obvious as reputation is no doubt reliant on conventional attractiveness as well as skill, but skills do not weigh into the Reputation and Compensation Factor. Further, FIFA is filled with extremely athletic players, so height and weight are not a determining factor in conventional attractiveness vs other players. Overall, this leaves an interesting correlation to be looked into at a later date.

CCA

CCA: Physical Traits and Player Skills

```
> # get the corr matrices
> cor(X)
              weight    agility    stamina    strength      vision  height_cm
weight     1.0000000 -0.5342643 -0.2233167  0.61579779 -0.28411350  0.7546782
agility   -0.5342643  1.0000000  0.5687060 -0.23419873  0.59732736 -0.6215713
stamina   -0.2233167  0.5687060  1.0000000  0.26269371  0.47233542 -0.2883512
strength   0.6157978 -0.2341987  0.2626937  1.00000000 -0.04692916  0.5363115
vision    -0.2841135  0.5973274  0.4723354 -0.04692916  1.00000000 -0.3696723
height_cm  0.7546782 -0.6215713 -0.2883512  0.53631152 -0.36967235  1.0000000
> cor(Y)
              short_passing long_passing shot_power long_shots
short_passing     1.0000000    0.8957218  0.7718453  0.7617495
long_passing      0.8957218    1.0000000  0.6714256  0.6678472
shot_power        0.7718453    0.6714256  1.0000000  0.8892543
long_shots        0.7617495    0.6678472  0.8892543  1.0000000
```

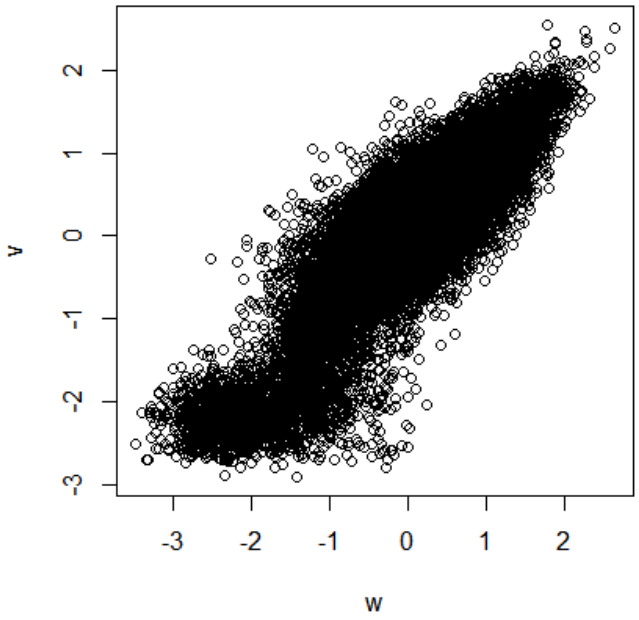
< Table 5-1: Correlation between X and Y >

```
> # cannonical correlates
> c = cancor(X, Y)
> c
$cor
[1] 0.8830137 0.3252802 0.2055534 0.1894048

$xcoef
                  [,1]          [,2]          [,3]          [,4]          [,5]          [,6]
weight    -2.510866e-05  1.249429e-04 -0.0001176027 -2.769320e-04  6.722576e-04 -3.268025e-04
agility    7.642236e-05  1.074836e-04 -0.0007526596 -2.472936e-04 -8.224752e-05  1.642179e-04
stamina    1.623814e-04 -3.436301e-04  0.0001292602  3.327755e-04  4.371401e-04  1.017048e-04
strength   9.698623e-05 -3.301640e-04 -0.0000137725 -3.610397e-04 -7.171104e-04 -2.935949e-04
vision     3.151845e-04  3.181561e-04  0.0004436498 -1.983421e-04 -3.673888e-05  9.641524e-05
height_cm -8.811992e-05 -9.597012e-05 -0.0001556629 -2.608353e-05 -1.557906e-04  1.864899e-03

$ycoef
                     [,1]          [,2]          [,3]          [,4]
short_passing 2.041809e-04 -0.0006889272 -7.854765e-04  0.0008194780
long_passing  9.696702e-05  0.0002956479  1.025468e-03 -0.0002080984
shot_power    5.887769e-06 -0.0005699909  5.128989e-05 -0.0008129335
long_shots    1.792396e-04  0.0008018741 -2.083242e-04  0.0002052363

$xcenter
   weight   agility   stamina  strength    vision height_cm
165.97913  63.50361  63.21995  65.31197  53.40090 181.25758

$ycenter
short_passing  long_passing    shot_power    long_shots
     58.68671      52.71193      55.46005      47.10997
```

< Table 5-2: Canonical correlation coefficients of X and Y >

First, we assigned variables to groups of X and Y. Then, we used the cor() function to generate the correlation coefficient matrix that does not classify groups. And we ran Canonical Correlation Analysis using the cancor() function. Based on the results, the first canonical correlation coefficient was 0.8830137, which means there is a high positive correlation between the two groups. Scatter Plot also shows us there is a positive correlation.

```
> cc_mm$cor
[1] 0.8830137 0.3252802 0.2055534 0.1894048
```



< Figure 4: scatter plot >

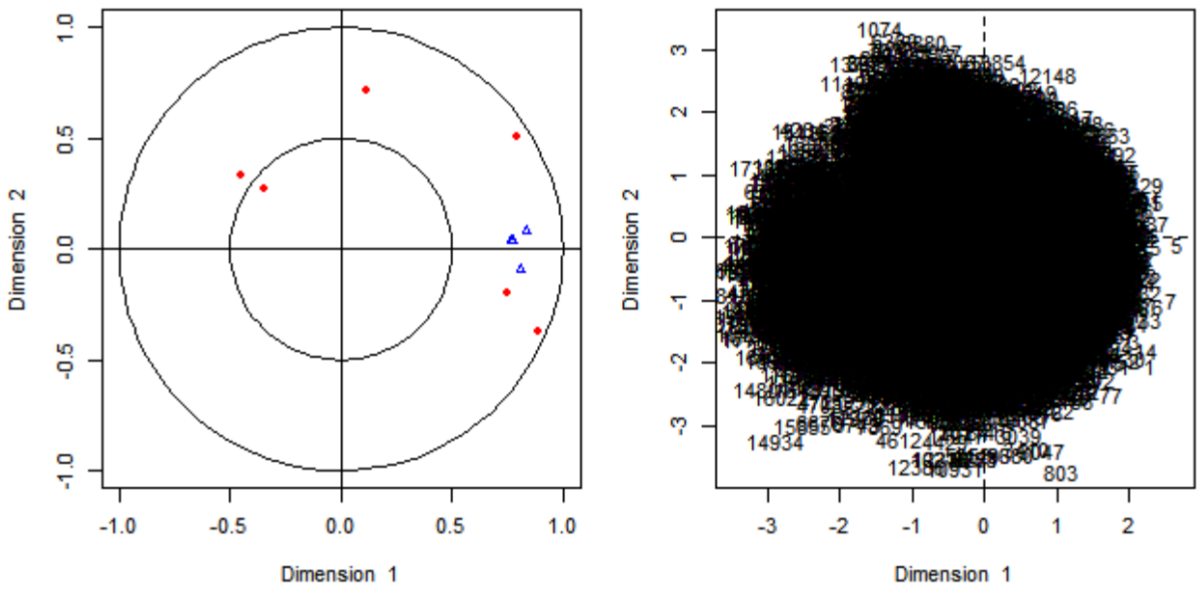
```
> #XCoef Correlations
> cc_mm$xcoef
                  [,1]        [,2]         [,3]         [,4]
weight    -0.003383432 -0.01683626  0.015847153 -0.037317032
agility    0.010298038 -0.01448359  0.101422098 -0.033323208
stamina    0.021881160  0.04630472 -0.017418024  0.044842034
strength   0.013069052  0.04449014  0.001855866 -0.048650690
vision     0.042471620 -0.04287205 -0.059782526 -0.026726926
height_cm -0.011874303  0.01293213  0.020975831 -0.003514798
>
> #YCoef Correlations
> cc_mm$ycoef
                      [,1]        [,2]         [,3]         [,4]
short_passing 0.0275137097  0.09283405  0.105844234  0.11042599
long_passing  0.0130664631 -0.03983903 -0.138183454 -0.02804160
shot_power    0.0007933864  0.07680720 -0.006911395 -0.10954412
long_shots    0.0241528317 -0.10805384  0.028072031  0.02765592
```

< Table 6: coefficient correlations coefficients >

Second, we calculated the $xscores and $yscores. This means the scores of canonical variables that each case value is plugged into canonical variables. The next result of corr.X.xscores and corr.Y.yscores shows the correlation between canonical variables and each variable. From the results, we can see that the correlation of vision is 0.8927782 in the case of the first canonical variable score of X, which shows a high positive correlation. For the first canonical variable score of Y, short_passing has a correlation coefficient value of 0.9487528, which is also a high positive correlation.

```
> loadings_mm$corr.X.xscores
                [,1]       [,2]        [,3]       [,4]
weight    -0.3416288  0.2693888 -0.13000587 -0.7648039
agility    0.7481801 -0.2022537  0.60974581  0.1561074
stamina    0.7960080  0.5081894  0.08560146  0.2306152
strength   0.1163136  0.7155657 -0.13258996 -0.6617067
vision     0.8927782 -0.3704165 -0.20543278 -0.1326217
height_cm -0.4486372  0.3334380 -0.19823336 -0.5503401
>
> #Correlation Y Scores
> loadings_mm$corr.Y.yscores
                   [,1]       [,2]        [,3]       [,4]
short_passing 0.9487528  0.2542330 -0.02142616  0.1864796
long_passing  0.8824073  0.1107277 -0.44334270  0.1119996
shot_power    0.8739912  0.1165730  0.14042794 -0.4503666
long_shots    0.9191953 -0.2721268  0.20537172 -0.1971026
```

< Table 7: scores of X and Y >



< Figure 5: A basic visualization of the canonical correlation >

Lastly, we also used the Yacca package for CCA. Based on the results, 33% of physical traits explain skills, and 65% of skills are already in physical traits.

CCA: Position Ratings and Player Skills

Before primary analysis, we did the factorability test on the datasets. Bartlett's test of sphericity (p=0.0001) proves that the data are not correlated. The kaiser-Mayer-Olkin measure of sampling adequacy resulted in a value of KMO = 0.97 indicating the sampling is adequate.

```
        Bartlett's Test of Sphericity

Call: bart_spher(x = skills_6)

     X2 = 1007678.394
     df = 528
p-value < 2.22e-16
```

```
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = skills_6)
Overall MSA =  0.97
MSA for each item =
```

< Table 8: Bartlett's Test of Sphericity results>

As there were too many variables within each group to conduct CCA, PCA was conducted as a pre-processing step to CCA to both groups to reduce the dimensionality in the data.

After running PCA on the full skills dataset, the result was four components with no cross-loading issues from the skills dataset with four variables dropped. The variables *Sprint_speed, jumping,* and *stamina* were removed due to increasing the threshold. R*eactions* was manually removed as five components resulted in a light component (i.e., less than three variables), and increasing the cutoff to remove the cross-loading would have resulted in removing too many additional variables.

```
> p3 = psych::principal(skills_5, rotate="promax", nfactors=4, scores=TRUE)
> print(p3$loadings, cutoff=.53, sort=T)

Loadings:
                  RC1    RC2    RC3    RC4
crossing           0.568
finishing          0.926
short_passing      0.629
volleys            0.943
dribbling          0.671
curve              0.819
fk_accuracy        0.809
long_passing       0.567
ball_control       0.682
reactions          0.895         0.653
shot_power         0.871
long_shots         0.938
positioning        0.812
vision             0.965
penalties          0.851
composure          0.903
aggression                0.656
interceptions             1.008
marking                   0.942
standing_tackle           1.008
sliding_tackle            1.024
heading_accuracy                -0.651
gk_diving                        0.845
gk_handling                      0.846
gk_kicking                       0.842
gk_positioning                   0.850
gk_reflexes                      0.846

acceleration                           0.578
agility                                0.602
balance                                0.730
strength                              -0.826
sprint_speed
jumping
stamina
```

< Table 9: PCA results of variables >

The resulting loading with the aforementioned variables removed captured 76.7% of the variability which was deemed acceptable. Evaluating the variables associated with the four components, RC1 looks to capture skill with the ball, RC3 looks to capture a player's abilities when playing in the net, RC2 looks to capture defensive ability, and RC4 looks to capture their movement abilities.

```
                       RC1   RC3   RC2   RC4
SS loadings         11.161 5.762 5.680 2.712
Proportion Var       0.338 0.175 0.172 0.082
Cumulative Var       0.338 0.513 0.685 0.767
```

< Table 10: four RCs >

PCA on the positions dataset was more straightforward and required less manipulation to meet our needs resulting in two components. The resulting loading captured 87.9% of the variability which was deemed acceptable. Evaluating the variables associated with the two components, RC1 looks to capture offensive positions while RC2 looks to capture defensive positions

```
> print(p30$loadings, cutoff=.4, sort=T)

Loadings:
    RC1    RC2
ls   0.975
st   0.975
rs   0.975
lw   0.989
lf   1.000
cf   1.000
rf   1.000
rw   0.989
lam  0.956
cam  0.956
ram  0.956
lm   0.911
lcm  0.711
cm   0.711
rcm  0.711
rm   0.911
lwb         0.765
ldm         0.875
cdm         0.875
rdm         0.875
rwb         0.765
lb          0.873
lcb         1.074
cb          1.074
rcb         1.074
rb          0.873

                 RC1   RC2
SS loadings    14.028 8.826
Proportion Var  0.540 0.339
Cumulative Var  0.540 0.879
```

< Table 11: components scores >

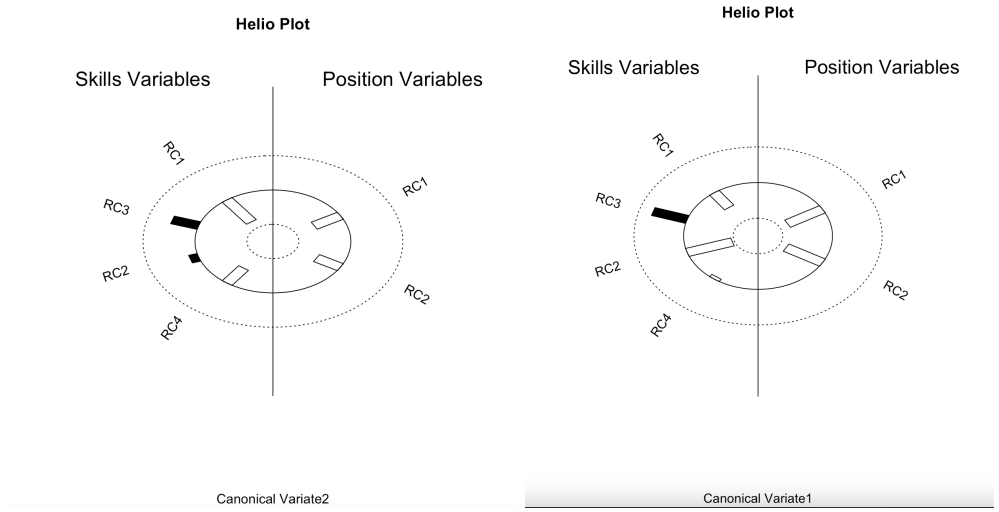
Datasets for the skills and position component scores were created; with the pre-processing complete, we were now able to conduct CCA using the *YACCA* library. Based on the results, the first canonical correlation coefficient was 0.9946760, which means there is a high positive correlation between the two groups. The results from Bartlett's Chi-Squared Test suggest that we can invalidate the null hypothesis that the canonical correlations are all equal to zero.

```
Bartlett's Chi-Squared Test:

          rho^2      Chisq df    Pr(>X)
CV 1 9.8938e-01 1.5252e+05  8 < 2.2e-16 ***
CV 2 9.7885e-01 7.0009e+04  3 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

< Table 12: Chi-Squared Test results >

In reviewing the correlations between the first canonical variate for skills and the position variables, we can conclude that two out of the four skills variables positively contribute to the first variate, with RC3 being the primary driver, while both position variables negatively contribute with near equal weighting to the first variate.

In reviewing the correlations between the second canonical variate for skills and the position variables, we can conclude that one out of the four skills variables positively contribute to the first variate, with RC3 again being the primary driver, while both position variables once again negatively contribute with near equal weighting to the second variate.



< Figure 6: visualization using yacca >

Based on the results, 78% of skills explain position, and ~99% of position are already in skills. From this, we learned that skills are greatly related to a player’s position rating, which supports our theoretical hypothesis that the two are associated. Given the high Redundancy Coefficients, one could come to a strong hypothesis that each set of variables could be used to predict the other variant.

```
> library(yacca)
> c3 = cca(skills_7, position_4)
> summary(c3)

Canonical Correlation Analysis - Summary


Canonical Correlations:

     CV 1      CV 2
0.9946760 0.9893702

Shared Variance on Each Canonical Variate:

     CV 1      CV 2
0.9893803 0.9788535

Bartlett's Chi-Squared Test:

          rho^2      Chisq df    Pr(>X)
CV 1 9.8938e-01 1.5252e+05  8 < 2.2e-16 ***
CV 2 9.7885e-01 7.0009e+04  3 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


      Canonical Variate Coefficients:

             X Vars:
                CV 1       CV 2
      RC1  0.06167320 -0.8404199
      RC3  0.37137132  0.2593235
      RC2 -0.79996458  0.6557564
      RC4  0.02600578 -0.1533795

             Y Vars:
               CV 1       CV 2
      RC1  0.3809103 -0.6065745
      RC2 -0.4400579  0.6144538


      Structural Correlations (Loadings):

             X Vars:
                CV 1       CV 2
      RC1 -0.51025080 -0.7953805
      RC3  0.72030949  0.5326735
      RC2 -0.95761703  0.1683741
      RC4 -0.08048939 -0.5411369

             Y Vars:
               CV 1       CV 2
      RC1 -0.8130055 -0.5822561
      RC2 -0.8468661 -0.5318062


      Fractional Variance Deposition on Canonical Variates:
```

```
Fractional Variance Deposition on Canonical Variates:

        X Vars:
           CV 1       CV 2
RC1 0.260355879 0.63263011
RC3 0.518845756 0.28374106
RC2 0.917030367 0.02834983
RC4 0.006478542 0.29282915

        Y Vars:
         CV 1      CV 2
RC1 0.6609779 0.3390221
RC2 0.7171822 0.2828178


Canonical Communalities (Fraction of Total Variance
Explained for Each Variable, Within Sets):

        X Vars:
      RC1       RC3       RC2       RC4
0.8929860 0.8025868 0.9453802 0.2993077

        Y Vars:
RC1 RC2
  1   1


Canonical Variate Adequacies (Fraction of Total Variance
Explained by Each CV, Within Sets):


         X Vars:
       CV 1      CV 2
  0.4256776 0.3093875

         Y Vars:
     CV 1    CV 2
  0.68908 0.31092


  Redundancy Coefficients (Fraction of Total Variance
  Explained by Each CV, Across Sets):


         X | Y:
       CV 1      CV 2
  0.4211571 0.3028451

         Y | X:
       CV 1      CV 2
  0.6817622 0.3043451


  Aggregate Redundancy Coefficients (Total Variance
  Explained by All CVs, Across Sets):

         X | Y: 0.7240021
         Y | X: 0.9861073
```

< Table 13: CCA process >

Regression

The linear regression model results are shown below (*Tabel 3, Fig 2-6*) as we compared all the models for different positions. First of all, age has a relatively high negative influence on the model. According to *Stephen P., 2008,* relative age effects (RAEs) have been known to impact a soccer player's status in Germany. Hence, we can assure that age's impact on players' performance is a significant indicator in real-world soccer data. Besides, for variables selected in the defender position model, we can tell that skills like short passing, ball control, interception, standing tackle, and sliding tackle are critical. However, long passing has a negative impact on the model. It is not surprising that players in defender positions usually won't make long passes except central defenders. Same as the crossing skills, usually left/right wingers need this skill during the game.

For midfield positions, we can see that the players' psychological components, such as reactions and composure, are relatively essential characters.Players in offensive positions have to be good at finishing skills since their main job is to finish a goal. On the contrary, the goalkeeper's model finishing shows a negative impact.

All Players	Defender Position	Midfield Position	Offensive Position	Defensive (GoalKeeper)
age	age	age	age	age
work_rate	crossing	crossing	finishing	finishing
heading_accuracy	heading_accuracy	heading_accuracy	heading_accuracy	sprint_speed
short_passing	short_passing	short_passing	short_passing	reactions
curve	curve	volleys	dribbling	jumping
long_passing	long_passing	dribbling	curve	aggression
acceleration	ball_control	fk_accuracy	ball_control	positioning
reactions	reactions	ball_control	acceleration	penalties
balance	shot_power	acceleration	reactions	gk_diving
jumping	jumping	reactions	stamina	gk_handling
stamina	stamina	shot_power	aggression	gk_kicking
positioning	strength	jumping	positioning	gk_positioning
penalties	aggression	aggression	penalties	gk_reflexes
composure	interceptions	vision	composure	
marking	penalties	penalties		
gk_positioning	composure	composure		
	marking	marking		
	standing_tackle	gk_diving		
	sliding_tackle	height_cm		
	gk_reflexes			

< Table 14: Linear Regression - Selected Variables >

```
Coefficients:
                   Estimate Std. Error  t value Pr(>|t|)
(Intercept)       5.098e+01  3.936e-01  129.530  < 2e-16 ***
age              -8.383e-01  7.312e-03 -114.653  < 2e-16 ***
value             2.068e-07  5.648e-09   36.609  < 2e-16 ***
work_rate2        8.759e-01  1.757e-01    4.986 6.24e-07 ***
work_rate3        3.671e-01  1.254e-01    2.928 0.003421 **
work_rate4        9.679e-01  2.039e-01    4.746 2.09e-06 ***
work_rate5       -1.195e+00  6.272e-01   -1.906 0.056712 .
work_rate6        7.920e-01  2.040e-01    3.883 0.000103 ***
work_rate7        3.735e-01  1.402e-01    2.665 0.007708 **
work_rate8        4.677e-01  1.672e-01    2.798 0.005147 **
work_rate9       -3.153e-02  1.198e-01   -0.263 0.792361
ldm               4.315e-02  3.718e-02    1.161 0.245834
lb                1.418e-01  2.982e-02    4.754 2.01e-06 ***
lcb               1.890e-02  2.803e-02    0.674 0.500082
crossing         -2.140e-02  4.528e-03   -4.727 2.30e-06 ***
heading_accuracy  6.627e-02  4.948e-03   13.392  < 2e-16 ***
short_passing     5.909e-02  6.675e-03    8.852  < 2e-16 ***
volleys          -1.341e-03  3.600e-03   -0.372 0.709599
dribbling         1.911e-02  5.316e-03    3.595 0.000325 ***
curve             9.371e-03  3.349e-03    2.798 0.005145 **
long_passing     -2.985e-02  5.832e-03   -5.118 3.12e-07 ***
ball_control      1.139e-01  6.932e-03   16.429  < 2e-16 ***
acceleration      1.274e-02  4.347e-03    2.930 0.003396 **
reactions         1.928e-01  5.352e-03   36.019  < 2e-16 ***
balance          -9.545e-03  3.126e-03   -3.053 0.002267 **
jumping           4.996e-03  2.818e-03    1.773 0.076302 .
stamina          -5.146e-02  3.628e-03  -14.183  < 2e-16 ***
long_shots       -8.109e-03  3.411e-03   -2.377 0.017442 *
interceptions    -3.380e-02  4.772e-03   -7.083 1.47e-12 ***
positioning      -4.362e-02  3.763e-03  -11.593  < 2e-16 ***
penalties         2.893e-02  3.369e-03    8.587  < 2e-16 ***
composure         1.140e-01  4.023e-03   28.341  < 2e-16 ***
marking           1.294e-02  4.008e-03    3.229 0.001244 **
sliding_tackle   -2.967e-02  6.078e-03   -4.882 1.06e-06 ***
gk_diving         6.289e-02  7.786e-03    8.077 7.07e-16 ***
gk_handling       8.282e-02  7.821e-03   10.589  < 2e-16 ***
gk_kicking        3.473e-02  7.282e-03    4.770 1.86e-06 ***
gk_positioning    1.003e-01  7.681e-03   13.057  < 2e-16 ***
gk_reflexes       5.946e-02  7.751e-03    7.671 1.80e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.258 on 16604 degrees of freedom
Multiple R-squared:  0.7202,    Adjusted R-squared:  0.7195
F-statistic:  1125 on 38 and 16604 DF,  p-value: < 2.2e-16
```

```
Linear Regression

16643 samples
   17 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 14978, 14978, 14980, 14980, 14978, 14979, ...
Resampling results:

  RMSE      Rsquared   MAE
  3.358911  0.7018505  2.631655
```

< Figure 7: Linear Regression - All Players Model Results >

```
Residuals:
    Min      1Q  Median      3Q     Max
-8.2513 -1.6858 -0.1944  1.5197 14.6525

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)      43.529613   0.467676  93.077  < 2e-16 ***
age              -0.917384   0.009592 -95.641  < 2e-16 ***
crossing         -0.013068   0.004067  -3.213 0.001320 **
heading_accuracy  0.054952   0.005958   9.223  < 2e-16 ***
short_passing     0.059219   0.007829   7.564 4.59e-14 ***
curve             0.012910   0.004030   3.203 0.001366 **
long_passing     -0.012227   0.005535  -2.209 0.027199 *
ball_control      0.043503   0.007052   6.169 7.39e-10 ***
reactions         0.087846   0.008294  10.592  < 2e-16 ***
shot_power        0.010604   0.003609   2.938 0.003313 **
jumping           0.014398   0.003477   4.141 3.51e-05 ***
stamina          -0.016965   0.004104  -4.134 3.62e-05 ***
strength          0.011758   0.004832   2.433 0.014993 *
aggression        0.029150   0.004760   6.124 9.79e-10 ***
interceptions     0.114214   0.009222  12.385  < 2e-16 ***
penalties         0.012111   0.004177   2.899 0.003755 **
composure         0.038169   0.006317   6.043 1.62e-09 ***
marking           0.142400   0.007767  18.335  < 2e-16 ***
standing_tackle   0.138345   0.013540  10.218  < 2e-16 ***
sliding_tackle    0.088490   0.011662   7.588 3.81e-14 ***
gk_reflexes      -0.043965   0.011367  -3.868 0.000111 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.528 on 5257 degrees of freedom
Multiple R-squared:  0.812,     Adjusted R-squared:  0.8113
F-statistic:  1135 on 20 and 5257 DF,  p-value: < 2.2e-16
```

```
Linear Regression

5278 samples
  21 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 4750, 4751, 4750, 4750, 4750, 4751, ...
Resampling results:

  RMSE      Rsquared   MAE
  2.524889  0.8119338  1.965278
```

< Figure 8: Linear Regression - Defender Position Model Results >

```
Residuals:
     Min       1Q   Median       3Q      Max
-10.7848  -2.0723  -0.2371   1.8487  17.3017

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      46.160869   1.598238  28.882  < 2e-16 ***
age              -0.946448   0.011609 -81.525  < 2e-16 ***
crossing          0.013707   0.006170   2.221 0.026370 *
heading_accuracy  0.029644   0.005227   5.671 1.49e-08 ***
short_passing     0.159115   0.010719  14.844  < 2e-16 ***
volleys          -0.015178   0.005624  -2.699 0.006978 **
dribbling         0.076505   0.011332   6.751 1.62e-11 ***
fk_accuracy      -0.014350   0.004667  -3.075 0.002117 **
ball_control      0.234090   0.014000  16.721  < 2e-16 ***
acceleration      0.019239   0.005438   3.538 0.000407 ***
reactions         0.114581   0.008237  13.911  < 2e-16 ***
shot_power        0.026780   0.006398   4.185 2.89e-05 ***
jumping          -0.010587   0.003977  -2.662 0.007786 **
aggression        0.012109   0.004081   2.967 0.003018 **
vision            0.040870   0.008303   4.922 8.80e-07 ***
penalties         0.035380   0.005642   6.271 3.86e-10 ***
composure         0.083655   0.007220  11.586  < 2e-16 ***
marking           0.018908   0.003740   5.055 4.44e-07 ***
gk_diving        -0.027864   0.012985  -2.146 0.031925 *
height_cm        -0.020679   0.008134  -2.542 0.011046 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.97 on 5436 degrees of freedom
Multiple R-squared:  0.7759,    Adjusted R-squared:  0.7751
F-statistic: 990.5 on 19 and 5436 DF,  p-value: < 2.2e-16
```

```
Linear Regression

5456 samples
  19 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 4910, 4911, 4911, 4911, 4908, 4911, ...
Resampling results:

  RMSE      Rsquared   MAE
  2.976387  0.7747034  2.356673
```

< Figure 9: Linear Regression - Midfield Position Model Results >

```
Residuals:
    Min      1Q  Median      3Q     Max
-9.5657 -1.9365 -0.3002  1.7038 12.7860

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      43.752035   0.602835  72.577  < 2e-16 ***
age              -1.012031   0.014217 -71.183  < 2e-16 ***
finishing         0.152349   0.013028  11.694  < 2e-16 ***
heading_accuracy  0.031998   0.007073   4.524 6.31e-06 ***
short_passing     0.087188   0.010416   8.370  < 2e-16 ***
dribbling         0.082899   0.013822   5.998 2.24e-09 ***
curve             0.017428   0.006016   2.897   0.0038 **
ball_control      0.153810   0.015989   9.620  < 2e-16 ***
acceleration      0.025625   0.006067   4.224 2.47e-05 ***
reactions         0.070409   0.010790   6.525 7.93e-11 ***
stamina          -0.047244   0.005824  -8.112 7.15e-16 ***
aggression        0.016602   0.004129   4.021 5.93e-05 ***
positioning       0.129736   0.012766  10.163  < 2e-16 ***
penalties         0.055880   0.007409   7.542 6.07e-14 ***
composure         0.058378   0.009694   6.022 1.93e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.757 on 3029 degrees of freedom
Multiple R-squared:  0.8048,    Adjusted R-squared:  0.8039
F-statistic: 892.3 on 14 and 3029 DF,  p-value: < 2.2e-16
```

```
Linear Regression

3044 samples
  14 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 2738, 2739, 2741, 2740, 2740, 2740, ...
Resampling results:

  RMSE      Rsquared  MAE
  2.764527  0.803161  2.20027
```

< Figure 10: Linear Regression - Offensive Position Model Results >

```
Residuals:
    Min      1Q  Median      3Q     Max
-8.1886 -1.7506 -0.3162  1.4127 12.2996

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)    32.520989   0.584023  55.684  < 2e-16 ***
age            -0.776111   0.014590 -53.195  < 2e-16 ***
finishing      -0.064139   0.019495  -3.290  0.00102 **
sprint_speed   -0.026557   0.007350  -3.613  0.00031 ***
reactions       0.070268   0.010512   6.684 3.05e-11 ***
jumping         0.024626   0.006122   4.022 5.99e-05 ***
aggression      0.023256   0.008327   2.793  0.00528 **
positioning    -0.044585   0.019028  -2.343  0.01922 *
penalties       0.022320   0.009828   2.271  0.02326 *
gk_diving       0.195953   0.019537  10.030  < 2e-16 ***
gk_handling     0.232099   0.017043  13.618  < 2e-16 ***
gk_kicking      0.038641   0.011870   3.255  0.00115 **
gk_positioning  0.213664   0.016526  12.929  < 2e-16 ***
gk_reflexes     0.150248   0.018839   7.975 2.61e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.581 on 1886 degrees of freedom
Multiple R-squared:  0.8357,    Adjusted R-squared:  0.8346
F-statistic: 737.9 on 13 and 1886 DF,  p-value: < 2.2e-16
```

```
Linear Regression

1900 samples
  13 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 1708, 1710, 1711, 1711, 1711, 1710, ...
Resampling results:

  RMSE      Rsquared   MAE
  2.588786  0.8331734  1.998319
```

< Figure 11: Linear Regression - Defensive Position Model Results >

Limitations

The limit of the study is that the data in the FIFA 19 players dataset is based on FIFA 19 game and thus is not entirely based on real life. Much of the data is based on the

players' skills, but it is subjective. An ideal way to apply this data would be to use our work on the skills and positions and apply it to real-world data.
The dataset we used has already created a scoring system such as overall score, potential, and international reputation in this research. In the future, our goal would be to predict and analyze different variables that are less understood in the dataset. Some examples would be wage, value, or even skills.
Additionally, for linear regression, we analyzed the regression. We found that there might be some related variables to use second-order or third-order terms to improve the model. Finally, we hope to use these FIFA statistics to analyze win rates in real-world games.

V. Conclusion

Overall, all analyses point out the importance of player performance in soccer. We reduced the dimensionality of the dataset and discovered five new components based on similar/shared variance. The analysis pointed out that finishing is an essential attribute for an attacker, while interception is most important for a defender. Weight, and height both cross-load into the Reputation and Compensation Factor. In many ways, this is obvious as reputation is no doubt reliant on conventional attractiveness and skill, but skills do not weigh into the Reputation and Compensation Factor.
We discovered what feature affects individual positions/performance and what each feature has in common with players' positions. One vital discovery is how age negatively affects players' performance; the older a player gets, performance reduces. Also, physical traits such as weight are not greatly related to soccer skills. Age, skills, and work rate are all critical features in determining a player's potential. Weight surprisingly explains similar variance as reputational features.
The various analyses performed on this dataset have a wide range of usage; club managers can improve their team or build a massive defense against opposition. Players can use this as a matrix to enhance a specific area or skill set they don't perform well. Countries can see what player can be called upon for the Olympics or any national tournament.

REFERENCE

1. Moura, F. A., Santana, J. E., Vieira, N. A., Santiago, P. R., & Cunha, S. A. (2015). Analysis of SOCCER PLAYERS' Positional variability during the 2012 UEFA EUROPEAN Championship: A case study. Journal of Human Kinetics, 47(1), 225–236. https://doi.org/10.1515/hukin-2015-0078
2. Abdullah, M. R., Maliki, A. B., Musa, R. M., Kosni, N. A., & Juahir, H. (2016). Intelligent prediction of soccer technical skill on youth soccer player's relative performance using multivariate analysis and artificial neural network techniques. International Journal on Advanced Science, Engineering and Information Technology, 6(5), 668. https://doi.org/10.18517/ijaseit.6.5.975
3. Al-Shebany, M. A., Amshaher, J. M., & Mohammed, H. (2021). Soccer team performance forecasting using artificial neural network. Journal of Pure & Applied Sciences, 20(1), 192–196. https://doi.org/10.51984/jopas.v20i1.1306
4. Joo, C. H., & Seo, D. I. (2016). Analysis of physical fitness and technical skills of youth soccer players according to playing position. Journal of exercise rehabilitation, 12(6), 548. https://doi.org/10.12965/jer.1632730.365
5. Kim YK. A fitness profiles of the professional soccer players by each position. *Korean J Sports Med.* 2000;18:217–226.