

Influence of physical characteristics on the potential, skill and skill, potential on the wages given to the players

Abstract:

FIFA is an organization that hosts the World Cup every four years, which is a hugely popular event with a global audience. This data set includes information about players from many nations, which allows me to examine patterns in relation to the most well-known game. Univariate analysis, Bivariate analysis is used to illustrate the distribution of target variables, relationship among different variables and obtain a sense of the relationship.

Introduction:

FIFA is one of the most well-known organizations with a large following. The proposed data set (Raghav, 2019) contains details of the players names as well as physical characteristics like height, which is recorded in cm, weight recorded in kgs, body type which is divided into 3 categories normal, lean and stocky body type, potential, skill of players, wages. The goal of this study is to look at the available patterns in players' physical characteristics in relation to their potential, talent, and salaries. This will also assist businesses in implementing methods that will help athletes reach their full potential.

Relevant Work:

(Li, 2017) investigates the difficulties of using the data mining algorithm to sports and presents a way for using the data mining algorithm to analyze the technical and tactical components of a sporting event. The experiment's outcomes show that the scheme may be determined from intuitive observation, but that simple measurements for the coach cannot be established. Because of the importance of the data and the capacity to access the key action conversion phase, the software for mining sports game important information plays a role. (Wang, 2018) In response to an actual demand for system analysis, this study creates and develops a data mining analysis system of college students' physical performance and physical ability test data based on data mining technology in physical performance management and physical ability analysis application. The method leverages test data from college players' physical ability as an analysis item. It uses data mining technology to optimize athletes' physical ability index. We present the concept and method of association rule mining, as well as algorithms suited for physical index operation analysis by college students. Then, using this method, a rule tree with specific cases will be created, and athletes' physical abilities will be assessed, resulting in a connected and usable physical teaching reference.

(Hannes Lepschy, 2021) Understanding the mechanisms that underpin football success is critical for coaches, players, managers, journalists, and other stakeholders. This understanding is as important in football as it is in other sports but determining what leads to success in football remains a challenge. Various attempts have been made to identify and quantify performance indicators, but the results vary and are partly inconsistent. The majority of studies focused on domestic leagues comprised of club teams, while studies on national team performance at tournaments are scarce. In recent years, only 11 studies involving data from a World Cup's success factors have been published. Using an elaborate statistical approach, the goal of this exploratory study is to identify the success factors for the FIFA

World Cup 2018 in Russia and the FIFA World Cup 2014 in Brazil. A results-based approach will be used to investigate 29 variables. This will be the first study to consider market value as a factor in the success of a FIFA World Cup

Strategy, Methods and Tools:

The objective is to use different technologies to get useful insights from the dataset.

R:

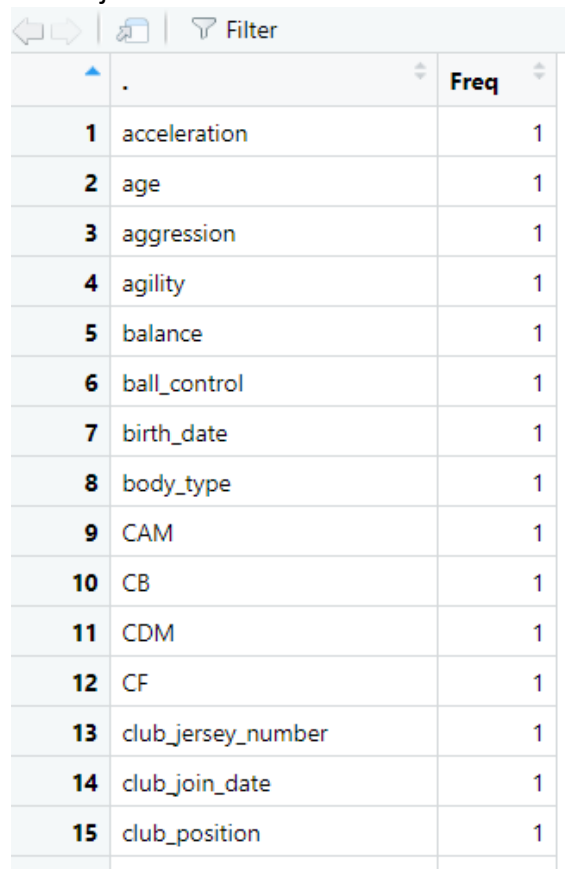
First, we clean the data set using R. Cleaning the data set involves cleaning of data by removing null values, unwanted data, from the dataset, removing duplicate values from the dataset. In this dataset, since there are many missing values replaced original value with unknown string. Data cleaning is a vital step to perform this process. After removing the unwanted data, the data is split into 3 data frames and written into new csv file to create new dataset that is relevant as per the research questions.

```

Rows: 17,954
Columns: 92
$ id          <dbl> 158023, 190460, 195864, 198219, 201024, 203376, 23174~
$ name        <chr> "L. Messi", "C. Eriksen", "P. Pogba", "L. Insigne", "~
$ full_name    <chr> "Lionel Andrés Messi Cuccittini", "Christian Dannema~
$ birth_date   <date> 1987-06-24, 1992-02-14, 1993-03-15, 1991-06-04, 1991~
$ age          <dbl> 31, 27, 25, 27, 27, 27, 20, 30, 32, 32, 30, 26, 26, 2~
$ height_cm    <dbl> 170.18, 154.94, 190.50, 162.56, 187.96, 193.04, 152.4~
$ weight_kgs   <dbl> 72.1, 76.2, 83.9, 59.0, 88.9, 92.1, 73.0, 69.9, 92.1,~
$ positions    <chr> "CF,RW,ST", "CAM,RM,CM", "CM,CAM", "LW,ST", "CB", "CB~
$ nationality   <chr> "Argentina", "Denmark", "France", "Italy", "Senegal",~
$ overall_rating <dbl> 94, 88, 88, 88, 88, 88, 88, 89, 89, 89, 89, 89, 89, 8~
$ potential    <dbl> 94, 89, 91, 88, 91, 90, 95, 89, 89, 89, 89, 90, 92, 9~
$ value_euro   <dbl> 110500000, 60500000, 73000000, 62000000, 60000000, 50

```

FIG 1 DATASET



		Freq
	.	
1	acceleration	1
2	age	1
3	aggression	1
4	agility	1
5	balance	1
6	ball_control	1
7	birth_date	1
8	body_type	1
9	CAM	1
10	CB	1
11	CDM	1
12	CF	1
13	club_jersey_number	1
14	club_join_date	1
15	club_position	1

FIG 2 SHOWING THE NULL VALUES

```

rq1_df<- data %>% select("full_name","height_cm","weight_kgs","body_type","potential")
head(rq1_df)
write.csv(rq1_df,"C:\\Users\\sahit\\OneDrive\\Documents\\Courses-1Sem\\AIT580\\Assignments\\Fi

rq2_df<- data %>% select("full_name","height_cm","body_type","weight_kgs","skill")
head(rq2_df)
write.csv(rq2_df,"C:\\Users\\sahit\\OneDrive\\Documents\\Courses-1Sem\\AIT580\\Assignments\\Fi

rq3_df<- data %>% select("full_name","height_cm","weight_kgs","body_type","skill","potential")
write.csv(rq3_df,"C:\\Users\\sahit\\OneDrive\\Documents\\Courses-1Sem\\AIT580\\Assignments\\Fi
head(rq3_df)

```

FIG 3 SHOWING THE DATASET SPLIT INTO 3 DATAFRAMES

Python:

Post data cleaning in R, I have used python to perform analysis on the cleaned dataset. After loading the datasets into python, I have tried to analyze the patterns of our target variable potential, skill and wages

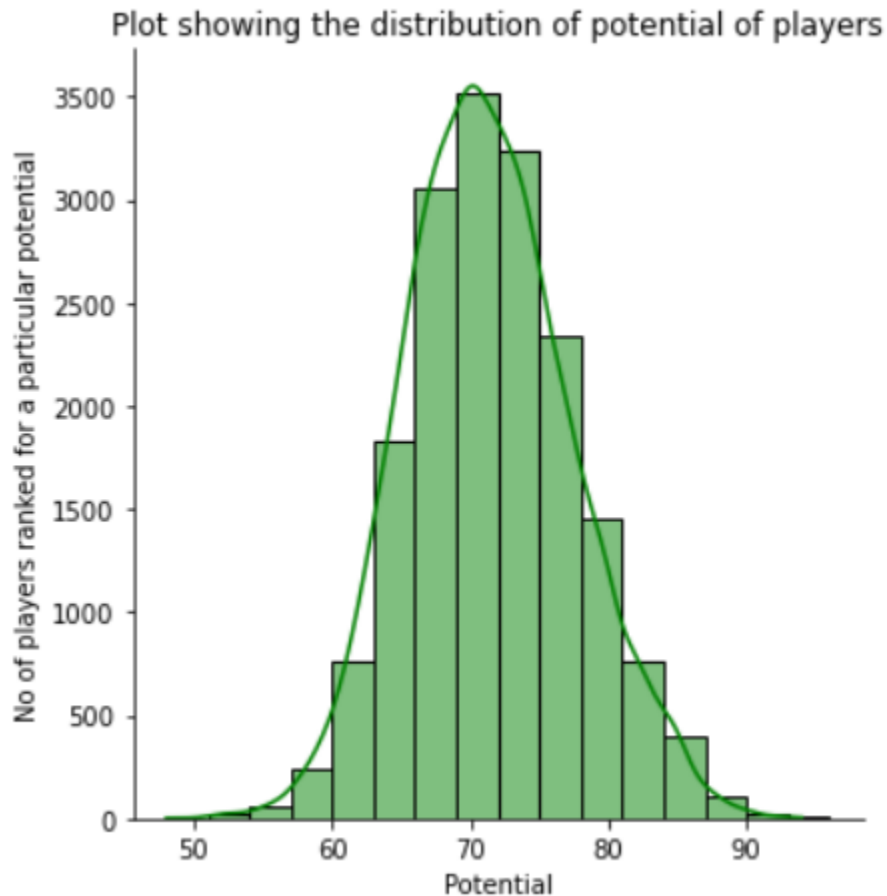


FIG 4 DISTRIBUTION OF POTENTIAL

From the above plot, the range of the target variable potential is fifty to ninety-four. We can deduce that the target variable potential is normally distributed in the dataset by looking at the plot

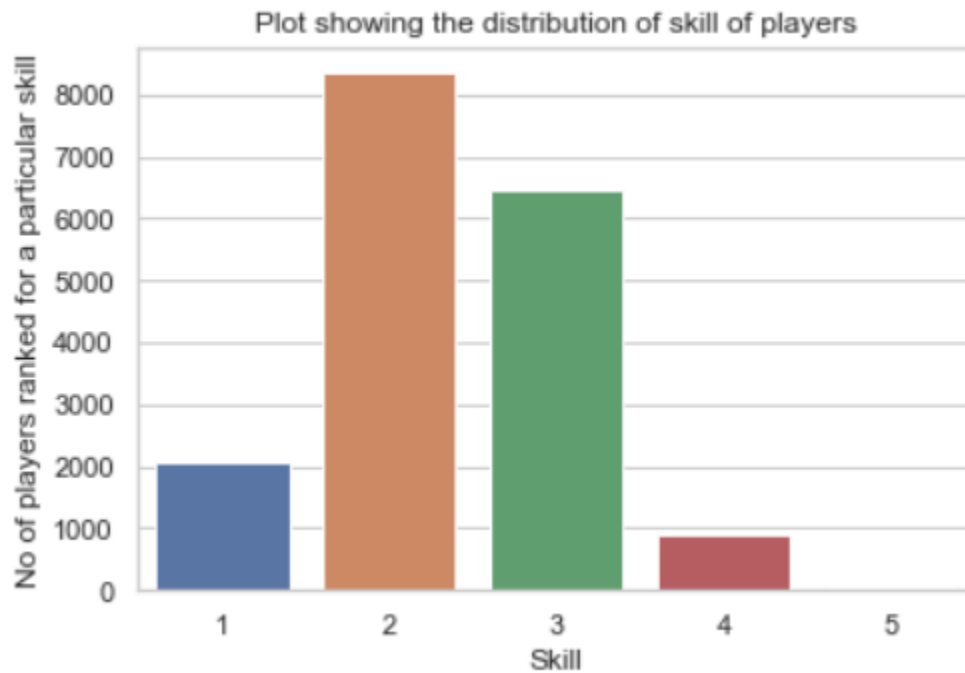


FIG 5 DISTRIBUTION OF SKILL

As the target variable skill is an ordinal variable, I have plotted bar plot to check the distribution of available values. From univariate analysis of the skill variable, it can be inferred that there are more players with ratings of 2 and 3, and less players with a skill level of 5. We can infer from the graphic that the data set's target variable skill is negatively skewed.

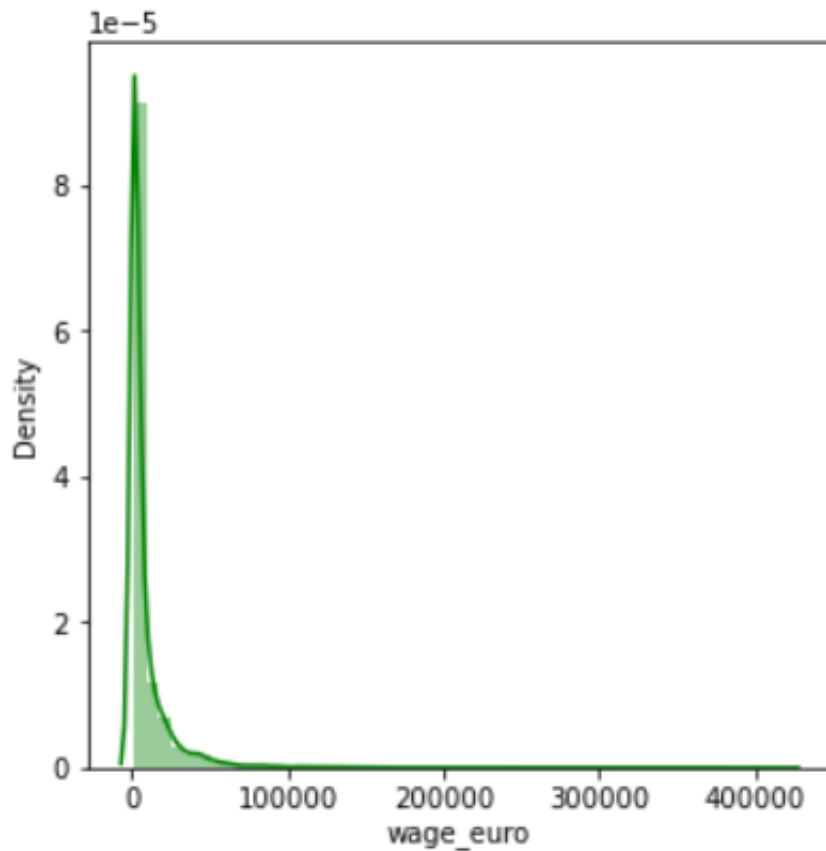


FIG 6 DISTRIBUTION OF WAGE

Since the wage_euro variable is a ratio type variable, I have used density plot to analyze the trend of the variable. From the above plot it is evident that the target variable wage is negatively skewed with more players having wages under the value 100000 euros.

SQL:

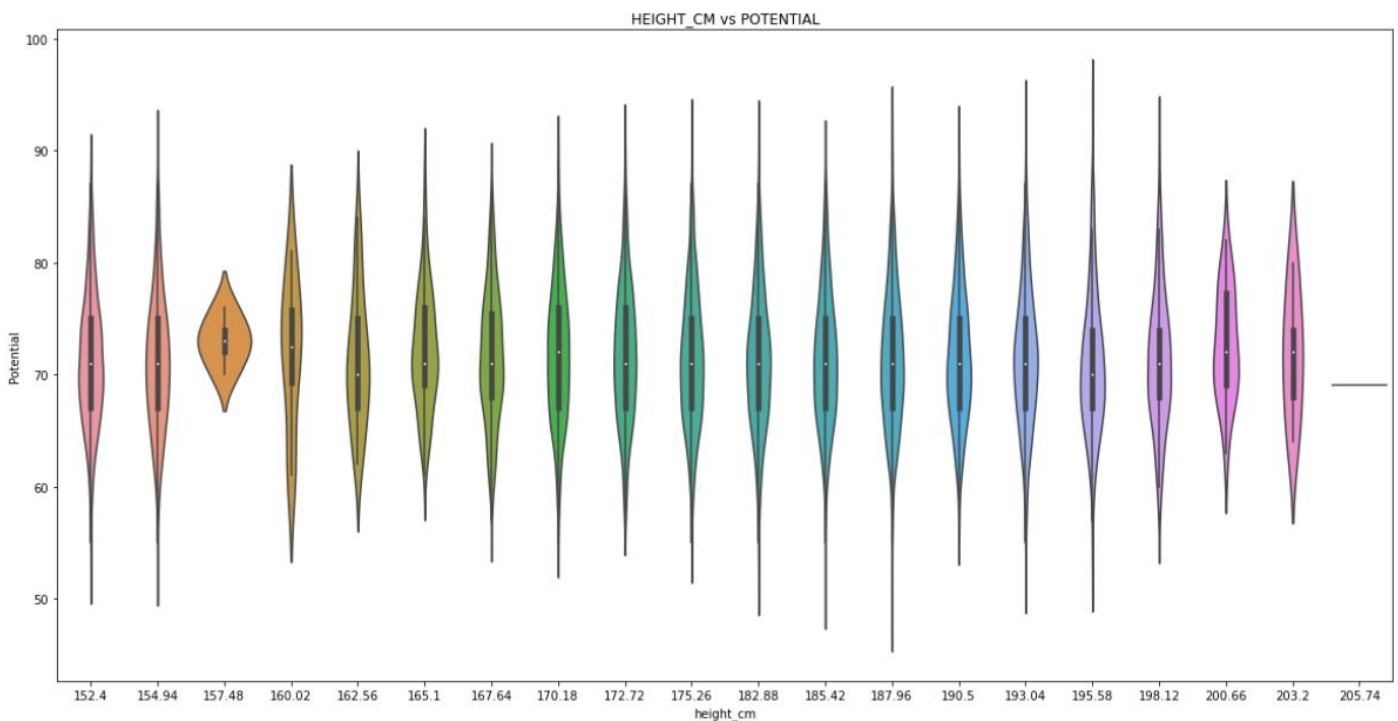
I have used sql to obtain more insights on the given data. Some of them are as below:

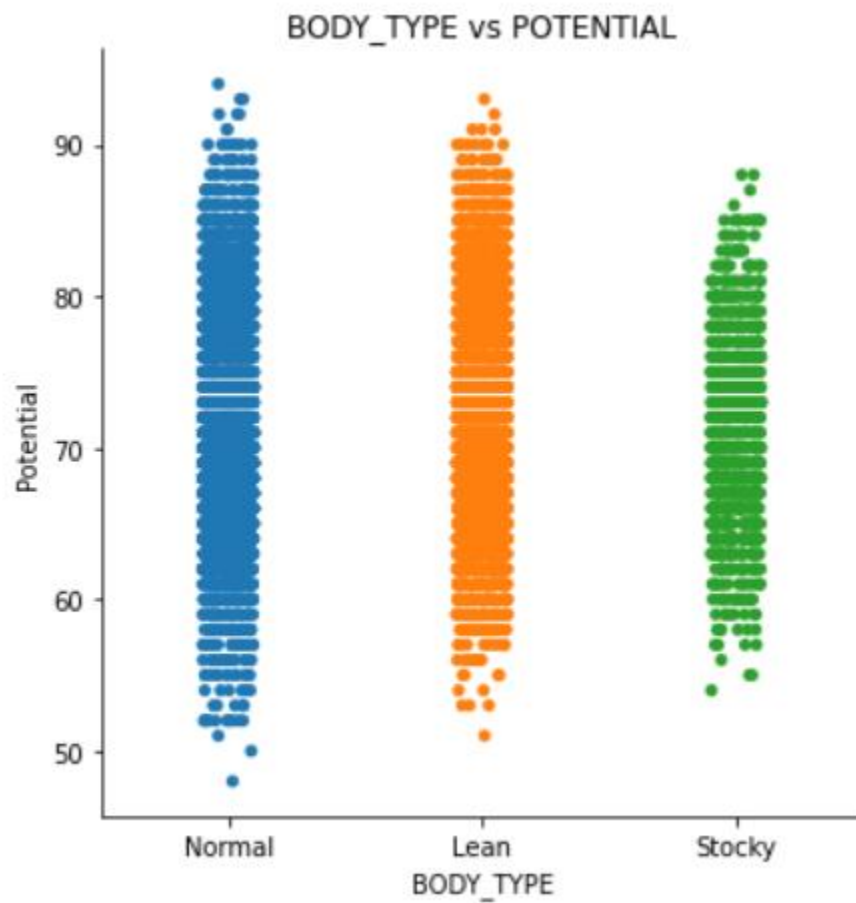
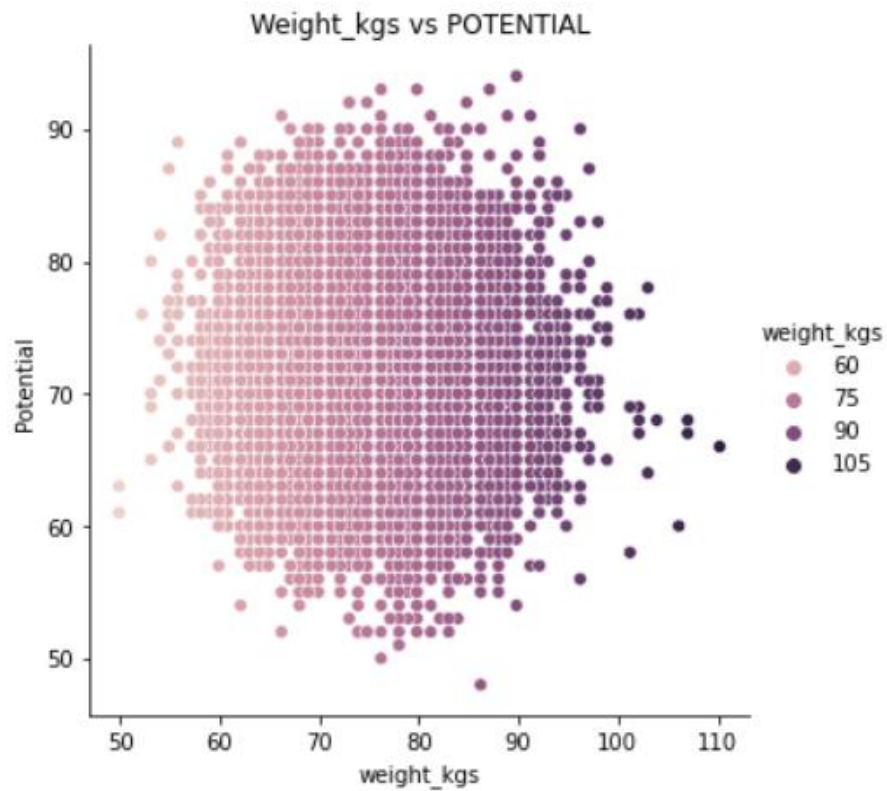
1. Selecting the players who have potential greater than 90.
2. Getting the average wage per country club.
3. Selecting the number of players per body type.
4. Selecting the distinct club teams available in the database
5. Selecting average height of players per team.

RESULTS:

In height cm vs Potential analysis, the potential is ranked from approximately 50-90 cm in the range of 150-160 cm. We can see a drop in potential values as we get higher in height, i.e., 200 and above. The weight kgs parameter vs target variable exhibits similar behavior. The range of potential is reduced as the weight grows by 90 kgs. The range of potential for normal and lean body types is about identical, however the range of potential for stocky body types is significantly less.

FIG 7 ANALYSIS OF POTENTIAL VS PHYSICAL FEATURES





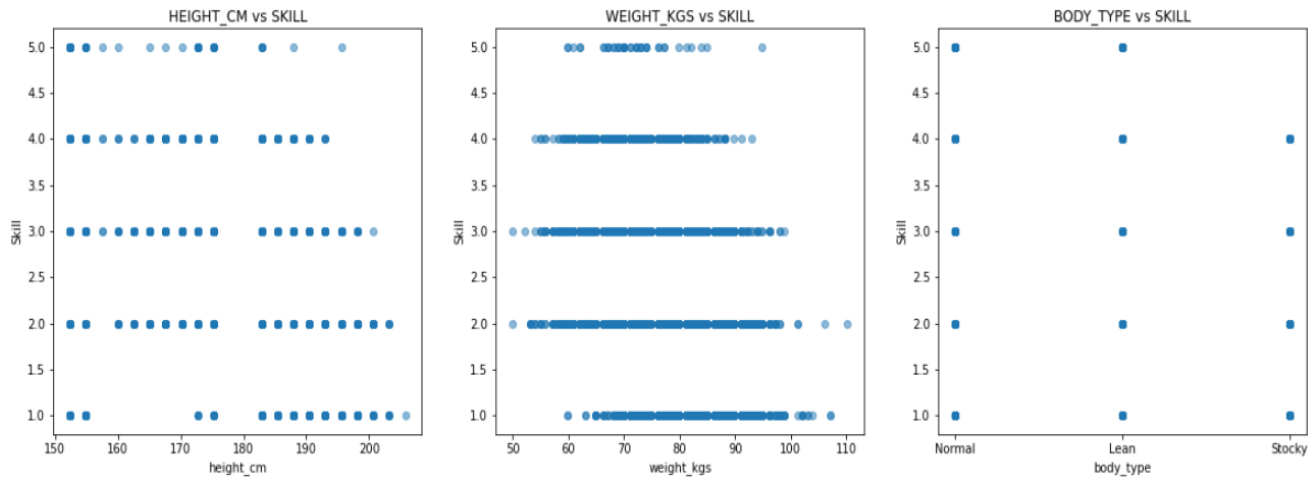


FIG 8 ANALYSIS OF SKILL VS PHYSICAL PARAMETERS

There is no discernible trend in the scatter plot of height cm vs Skill. When looking at the top players, the weight range is on the lower end, below 100. When comparing body type to skill, players with stocky bodies did not achieve higher skill ratings.

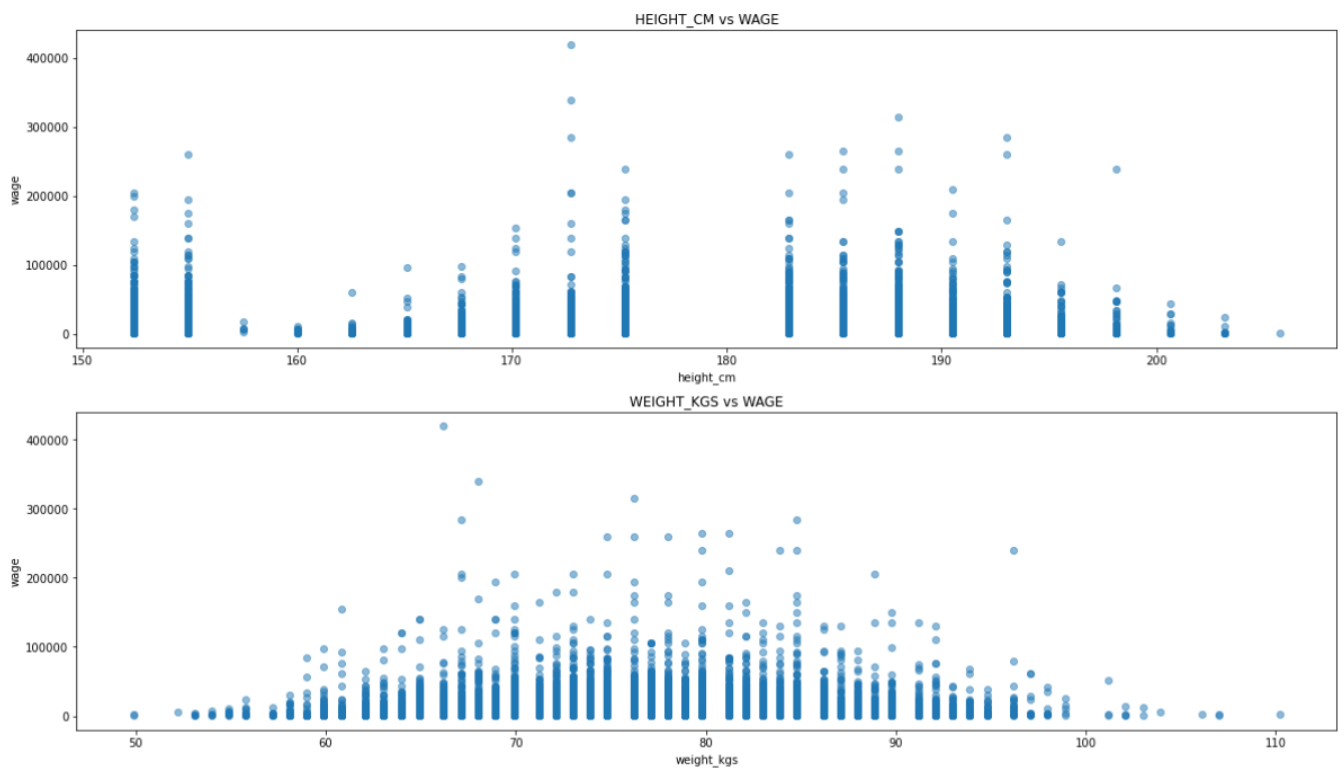


FIG 9 ANALYSIS OF WAGE VS HEIGHT_CM, WEIGHT_KGS

We have observed a trend where there is an increase in wages paid to the players as height increases to a certain point and as the height of the players increase above 190 cm, we can observe that the range of wages paid is less. In comparison of weight_kgs vs wage the range of wages paid to the players has a decreasing trend as the weight of players increases.

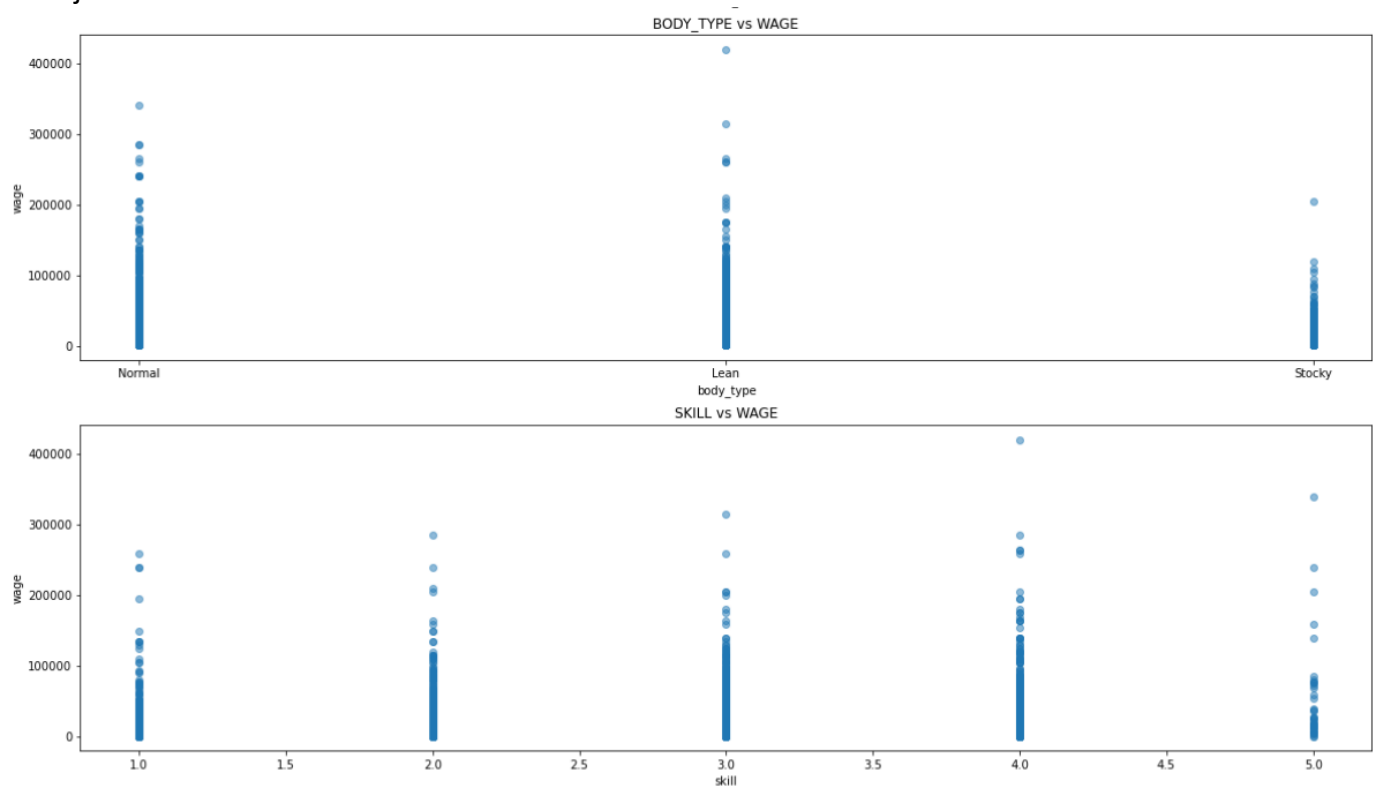


FIG 10 ANALYSIS OF WAGE VS BODY_TYPE, SKILL

In the analysis of body type vs wage, the range of players, the range of wages paid to the players with the normal and lean body type is higher in comparison to the wages paid to the players with stocky body type. In Skill vs wage analysis, the range of the wages paid to the players that is the difference of maximum wage vs minimum wage to the players is observed to be high for players with skill level 4.

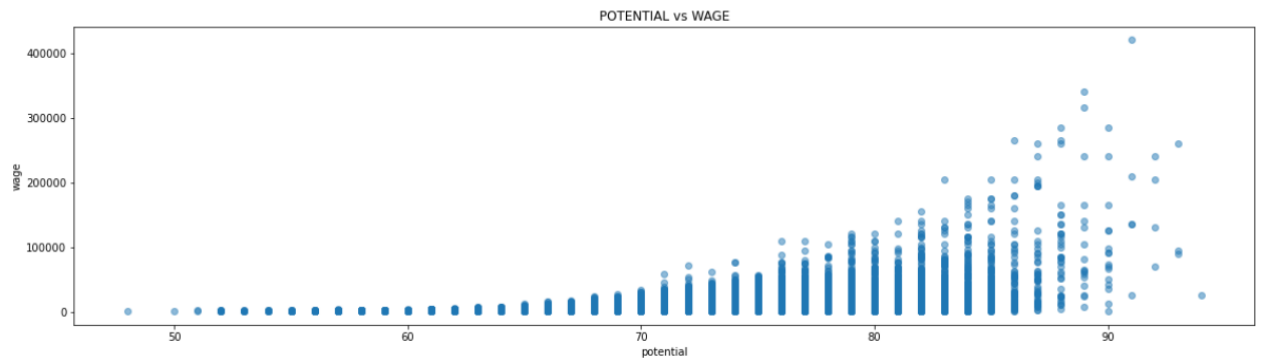


FIG 11 ANALYSIS OF WAGE VS POTENTIAL

In potential vs wage analysis, the players with the range of 80- 90 have more range that is the maximum vs minimum value of wages paid in comparison to the players who have potential greater than ninety.

References:

- 1) Hannes Lepschy, A. W. (2021). Success Factors in the FIFA 2018 World Cup in Russia and FIFA 2014 World Cup in Brazil.
Retrieved from https://www.researchgate.net/publication/349924065_Success_Factors_in_the_FIFA_2018_World_Cup_in_Russia_and_FIFA_2014_World_Cup_in_Brazil/link/605f5aaf458515e83476c8ee/download
- 2) Li, Z. (2017). Research and Analysis of Game Tactics Based on Data Mining Technology. Retrieved from <https://download.atlantis-pess.com/article/25882439.pdf>
- 3) Raghav. (2019). FIFA players dataset. Retrieved from <https://data.world/raghav333/fifa-players>
- 4) Wang, G.-X. (2018). Mining Analysis for College Students' Physical Performance and Fitness Tests. Retrieved from <https://ieeexplore.ieee.org/document/8531381>