# Analysis Report on FIFA Dataset

## INTRODUCTION

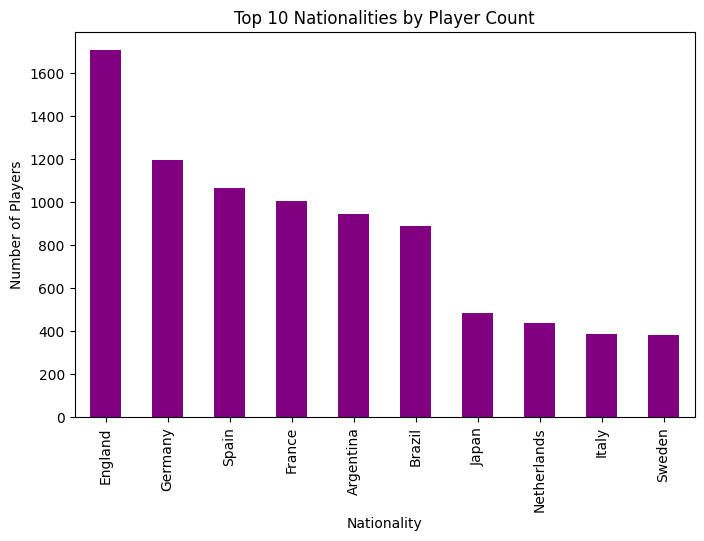
### In this report, I have analyzed the 2021 FIFA dataset to generate some key insights into the players' characteristics and how they influence their performance and wages. Digging into the dataset, we can see numerous details centered around players such as their positions, nationalities, wages, heights, weights, and attributes like agility, speed, and tackling.

### Objectives of this analysis:

1. **I want t**o investigate which factors impacting player wages, including position and nationality.
2. I will also explore the distribution and trends of players’ Nationalities, Weights, and wage
3. Understanding the relationship that exists between key attributes such as height, weight, acceleration, and agility and their interrelations and possible effects on performance and wages.
4. To understand how players' physical characteristics might influence their on-field performance and predict certain performance metrics.

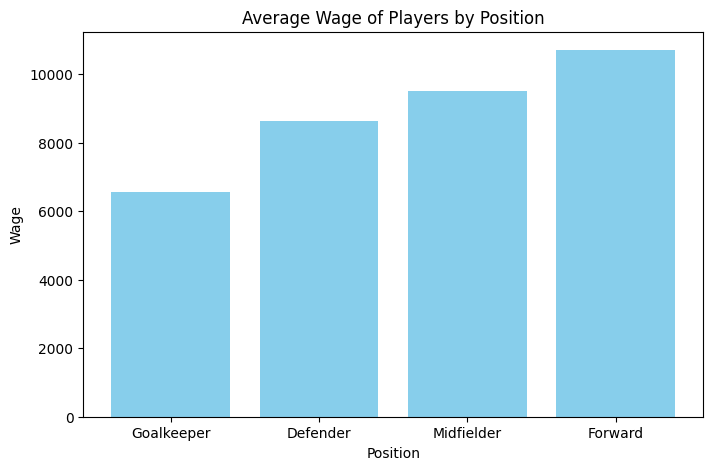
## Exploratory Data Analysis (EDA)

**Top 10 Nationalities by Player Count**: After cleaning the data I explored the distribution of players by nationality. I was able to identify the top 10 nationalities. Countries like England, Germany, Spain, France, and Argentina with total players of 1705, 1195, 1065, 1003, and 943 respectively, make up the first 5 of the top 10 countries. As can be seen in the chart below

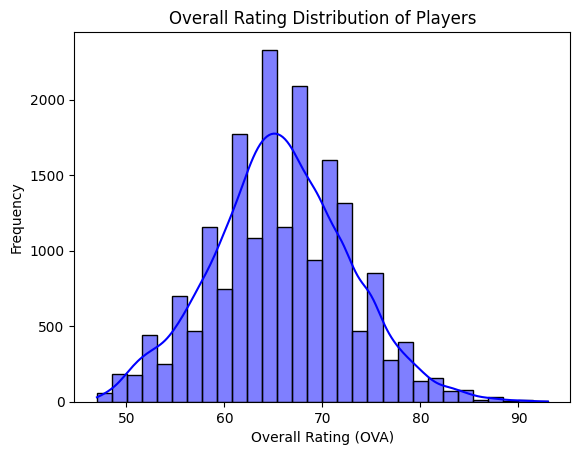


### Player Positions by Wage Analysis

The wage distribution among player positions was analyzed to understand which positions receive higher wages on average. A bar chart illustrates wage differences across various positions, providing insights into the most and least lucrative positions.



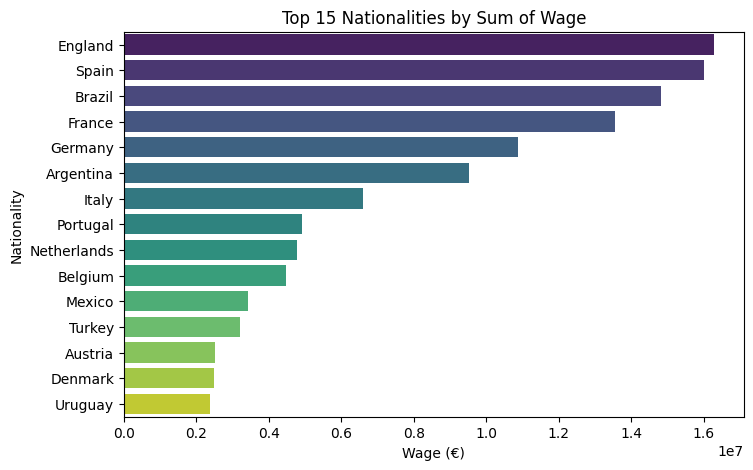
**Overall Rating Distribution**: There is a normal distribution of the players as most players fall between the rating of 60-80. The peak of the distribution suggests that the majority of players in the dataset have overall ratings of around 70, while fewer players have higher or lower ratings



### Descriptive Analysis of Player Nationality, Weight, and Wage

This section explores the players' nationalities relating to their wages and examines the distribution of player weights and wages.

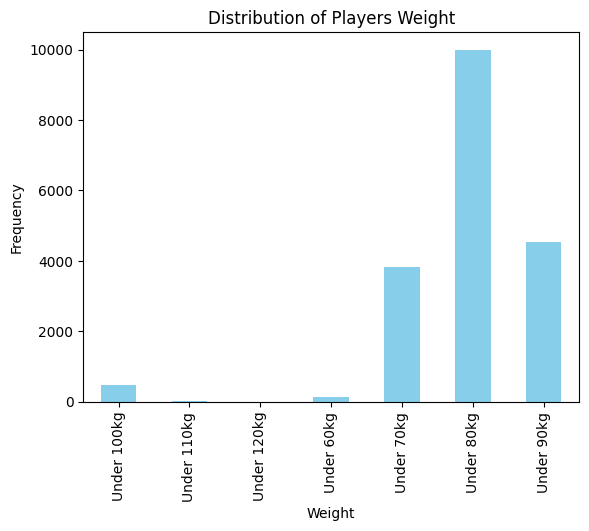
**Nationality:** European nationalities dominated the chart of the total earnings with a total of 10 teams in the top 15 while South America has 3 representatives in the top 15. We have countries like England, Spain, Brazil, France and many more.

 **Wage**:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Mean | STD | Min | Max | 25% | 50% | 75% |
| Wage | 9092 | 19707 | 0 | 560000 | 1000 | 3000 | 8000 |

The average wage of the players stands at €9092 with the minimum wage being €0.0 because some players have no club presently so they earn nothing. The maximum wage is €560,000. This huge gap between the minimum and the maximum is a result of many outliers that is many top players who command very high salaries compared to the average footballers. These players are from top nationalities and top football clubs.

**Weight:** The players' weight is grouped into 7 categories from under 60kg to under 120kg. Most of the players fall under the 80kg category. Under 70kg and under 90kg have a decent number too but there is a sharp decline as the weight increases with only a few of the players weighing above 100kg.



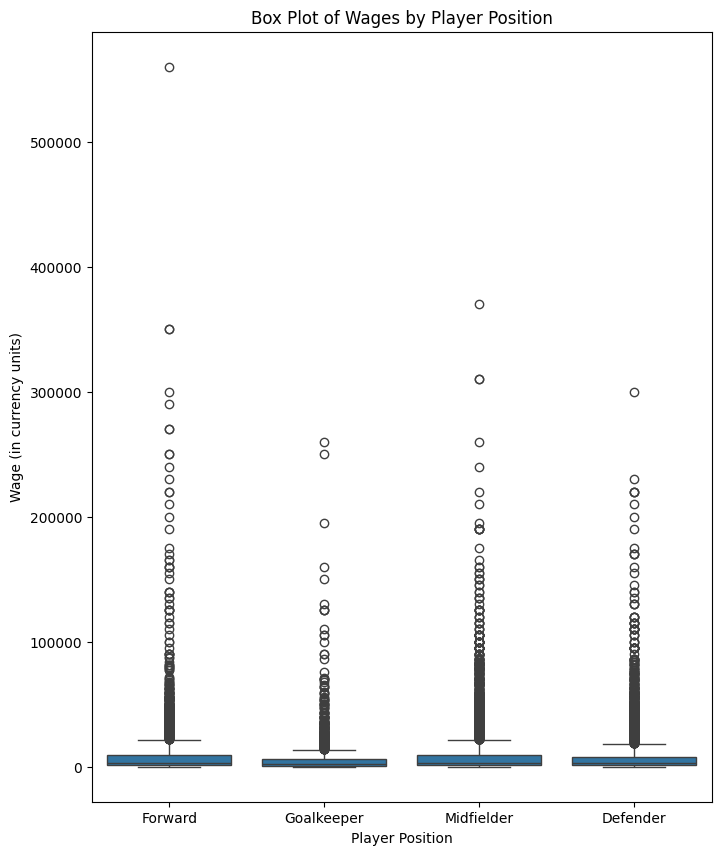
## Correlation Analysis

### Position and Wage Correlation

Analyzing the correlation between player positions and wages, the findings reveal which positions are likely to command higher wages. The table below and visualization shed light on the strength and nature of this relationship.

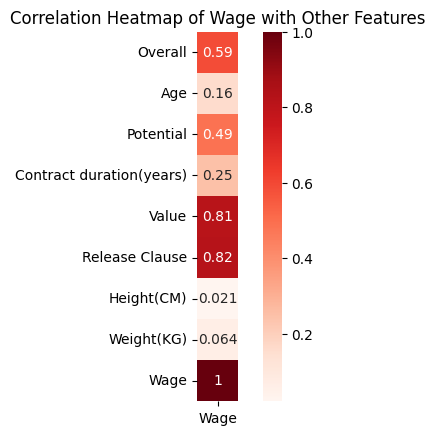
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Position | Count | Mean | STD | Min | Max | 25% | 50% | 75% |
| Forward | 3242 | €10,697 | €26,092 | €0.0 | €560,000 | 1000 | 3000 | 9000 |
| Midfielder | 7273 | €9,504 | €19,432 | €0.0 | 370000 | 1000 | 3000 | 9000 |
| Defender | 6389 | €8,632 | €17,127 | €0.0 | 300000 | 1000 | 3000 | 8000 |
| Goal-keeper | 2075 | €6,553 | €15,827 | €0.0 | 260000 | 750 | 2000 | 6000 |

The table above shows the difference in the average wage of the players across different positions. The forward players rank highest on the average earnings with a sum of €10,697 followed by the midfield players earning €9,504 on the average. The defenders' average wages stand at €8,632 with the goal-keepers coming last with an average sum of €6,553. This shows that the players who play further forward on the pitch like the forwards and midfielders tend to earn more than their defensive counterparts like the defenders and goalkeepers. The image below throws more light into this. We can see that there are many outliers but the forwards and midfielders have more than the rest.



### Predictive Factors for Player Wage

This analysis identifies factors that correlate with player wages. Factors with strong correlations are discussed in detail, supported by visualization.



## The heatmap above shows how player wages relate to different attributes in the dataset. Correlation values can go from -1 to 1, where numbers close to 1 mean a strong positive relationship, numbers close to -1 show a strong negative relationship, and values around 0 mean there’s little to no connection

* **Release Clause**: With **0.82** release clause has the highest positive correlation this suggests that the more the release clause of a player the more wage the player command
* **Value**: This stands close to the release clause with a positive correlation of **0.81**. this translates that player with higher values in the transfer market generally earns higher wages
* **Overall Rating**: Although low to the other two, a positive correlation of **0.59** indicates that players with higher overall ratings tend to earn more
* **Potential**: **The 0.49** correlation with potential indicates that players with higher potential are also likely to earn higher wages, although the relationship is not as strong as with market value or release clause
* **Contract Duration**: This has a weak **0.25** positive correlation with wage this may indicate that players on long contracts earn more but this is not very strong enough to draw a reasonable conclusion on that.
* **Age**: Age also has a very small positive relationship of **0.16** with the wage which suggests that there is a slight tendency for older players to earn more
* **Height**: At **-0.021** there is almost no existing relationship between how much a player earns and the height of that player.
* **Weight**: -**0.064**. Wage also has a very weak relationship with wage that it is not a significant factor in determining a player's wage

Nationality, club, and position were also evaluated as a factor to predict how much players earn below are the results of the test:

* **Wage vs Nationality: F-Statistic: 6.25, P-Value: 1.42e-119**. The F-statistics points to a weaker but significant relationship between nationality and wage. The p-value suggests that nationality has some effect on the players' wages although it is not strong compared to club and position
* **Wage vs Club: F-Statistic: 42.26, P-Value: 0.0**. With a high F-statistic and a p-value essentially at zero, the club has a strong and statistically significant relationship with wage. This result suggests that the club a player is associated with plays a major role in determining their wage, which makes sense as different clubs have varying budgets and wage structures
* **Wage vs Position: F-Statistic: 20.93, P-Value: 1.58e-13.** The F-statistic of 20.93 and the very low p-value (1.58 × 10⁻¹³) indicate a moderately strong and statistically significant relationship between wage and position. This suggests that a player’s position influences their wage, though not as strongly as the club they play for.

To get more insight into how we can predict the players' wages below is the result of a **regression analysis:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Coefficient | Std. Error | t-Statistic | P-value | 95% Confidence Interval |
| Age | 372.7 | 33.89 | 10.997 | 0.0 | (306.3, 439.1) |
| Overall | 312.9 | 29.26 | 10.693 | 0.0 | (255.5, 370.2) |
| Potential | 175.9 | 29.50 | 5.963 | 0.0 | (118.1, 233.8) |
| Contract duration | 349.3 | 36.11 | 9.673 | 0.0 | (278.5, 420.1) |
| value | 0.0018 | 0.000013 | 140.559 | 0.0 | (0.0018, 0.0020) |
| Height | 37.5 | 18.3 | 2.049 | 0.04 | (1.6, 73.4) |
| Weight | -17.8 | -17.8 | -0.977 | 0.329 | (-53.6, 17.9) |

**Age**: A positive coefficient of 372.7 indicates that for each additional year of age, a player’s wage increases by approximately €372.7 on average, suggesting a slight wage increase with experience or maturity.

**Overall Rating**: A positive coefficient of 312.9 means that for every one-point increase in overall rating, a player’s wage increases by €312.9 on average. This reflects the financial value placed on higher skill levels.

**Potential**: With a coefficient of 175.9, this suggests that each additional point in a player's potential rating is associated with an average wage increase of €175.9, highlighting how future value influences wages.

**Contract Duration**: The coefficient of 349.3 indicates that for each additional year in contract duration, a player’s wage increases by about €349.3, likely due to the security and commitment of longer-term contracts.

**Market Value**: The coefficient for market value is 0.0018, implying that each additional €1 in market value corresponds to a €0.0018 increase in wage. Although small, this indicates a proportional relationship between value and wage.

**Height (CM)**: The coefficient of 37.5 suggests that for each additional centimeter in height, a player’s wage increases by €37.5 on average, potentially due to physical advantages in the sport.

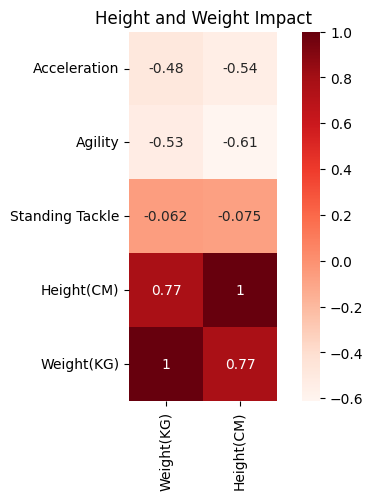
**Weight (KG)**: A coefficient of -17.8 indicates a slight wage decrease of €17.8 for each additional kilogram in weight, though this effect is not statistically significant (p-value > 0.05).

This regression analysis identifies key factors that influence player wages. Age, skill ratings, contract duration, and market value significantly contribute to wages. However, physical characteristics like height and weight have limited influence, indicating that skill and reputation are more critical in determining wages.

### Impact Analysis

### Height and Weight on Acceleration, Agility, and Standing Tackle

This section examines how a player’s height and weight impact their acceleration, agility, and standing tackle abilities. Using regression and correlation analyses, the results indicate the extent to which these physical factors affect specific skills, with relevant charts and tables.



## The heatmap above shows the following:

 **Height and Weight**:

* **Positive Correlation (0.77)**: There’s a strong positive correlation of 0.77 between height and weight, indicating that taller players tend to weigh more. This relationship is expected due to the natural link between body dimensions.

 **Acceleration and Physical Attributes**:

* **Negative Correlation with Height (-0.54)**: Taller players tend to have slightly lower acceleration scores, suggesting that increased height might hinder rapid movement.
* **Negative Correlation with Weight (-0.48)**: Heavier players also show a moderate negative correlation with acceleration, indicating that greater body mass can reduce speed over short distances.

 **Agility and Physical Attributes**:

* **Negative Correlation with Height (-0.61)**: There’s a moderate negative correlation between height and agility, meaning that taller players may be less agile. This could imply that shorter players can maneuver more quickly on the field.
* **Negative Correlation with Weight (-0.53)**: Similarly, heavier players are generally less agile, reinforcing the impact of lighter builds on agility.

 **Standing Tackle and Physical Attributes**:

* **Slight Negative Correlation**: Standing tackle ability shows minimal correlation with both height (-0.075) and weight (-0.062), suggesting that standing tackle proficiency is not significantly influenced by a player's physical attributes.

Here is a regression analysis to generate more insights: Weight:

|  |  |  |
| --- | --- | --- |
| Variable | Coefficient | Interpretation |
| Acceleration | -0.0685 | Each unit increase in acceleration is associated with a 0.0685 kg decrease in weight, on average. This suggests that players with higher acceleration tend to be lighter. |
| Agility | -0.2024 | A one-unit increase in agility is associated with a 0.2024 kg decrease in weight, indicating that more agile players tend to weigh less. |
| Standing Tackle | 0.0116 | A slight positive association where each unit increase in standing tackle ability corresponds to a 0.0116 kg increase in weight. This may imply that players who excel at tackling might have a marginally higher weight. |

Reveals that **acceleration** and **agility** have significant negative relationships with weight, suggesting that lighter players tend to be faster and more agile. Conversely, **standing tackle** shows a small positive relationship with weight, possibly indicating that a slightly higher weight might be beneficial for defensive roles.

Height:

|  |  |  |
| --- | --- | --- |
| Variable | Coefficient | Interpretation |
| Acceleration | -0.0547 | A one-unit increase in acceleration is associated with a 0.0547 cm decrease in height, on average, indicating that shorter players tend to have higher acceleration. |
| Agility | -0.2449 | Each unit increase in agility corresponds to a 0.2449 cm decrease in height, suggesting that more agile players tend to be shorter. |
| Standing Tackle | 0.0110 | |  | | --- | |  |  |  | | --- | | A positive coefficient, showing that each unit increase in standing tackle ability is associated with a 0.011 cm increase in height. This might indicate that taller players tend to excel slightly more in standing tackle. | |

This regression analysis shows that **acceleration** and **agility** are negatively associated with height, meaning shorter players generally exhibit higher agility and acceleration. On the other hand, **standing tackle** has a small positive relationship with height, suggesting that taller players might have a slight advantage in this skill.

## Conclusion

The analysis offers insights into the factors influencing player wages and skills in the FIFA dataset. Key findings include wage distribution by position, correlations with wage predictors, and the impact of physical attributes on certain skills, providing a comprehensive view of the dataset.