



Aziz



Emil



Jorge



Pedro



Walid

# Fifa Metrics FC

Sponsored by:



# Mission Statement:

- › Analyze how player attributes can shape their overall value in the market.
- › Analyze how FIFA rates their players, clubs and countries.

# Questions:

1. What are the top 10 players with highest growth potential?
2. What are the top 10 Nations & top 10 Clubs based on Player Value?
3. How do positions/wages/overall ratings vary in each league?
4. What is the relationship between market value, wage and overall performance?
5. What is the relation between player height and their overall ratings?

# Dataset:

- › FIFA 19 Dataset from Kaggle
- › 89 columns of attributes for each player,
- › 30,000+ Players/rows
- › Shooting, Defending, Passing, Wages etc
- › Data generated from <https://sofifa.com>

# Example of Data:



**K. Mbappé (ID: 231747)** FIFA 19 AUG 29, 2019  
Kylian Mbappé  **ST RW** Age 19 (Dec 20, 1998) 5'10" 161lbs

<b>88</b> Overall Rating	<b>95</b> Potential	Value €82M	Wage €100K
--------------------------	---------------------	------------	------------

Preferred Foot **Right**

International Reputation 3 ★

Weak Foot 4 ★

Skill Moves 5 ★

Work Rate **High/ Medium**

Body Type **Normal**

Real Face **Yes**

Release Clause **€168.1M**

#Speedster #Dribbler #Acrobat

 **Paris Saint-Germain**  
**84** ★★★★★  
Position **LS**  
Jersey Number **7**  
Joined **Jul 1, 2018**  
Contract Valid Until **2022**

 **France**  
**85** ★★★★★  
Position **RM**  
Jersey Number **10**

# Example of Data:

## Attacking

- 77 Crossing
- 88 Finishing
- 77 Heading Accuracy
- 82 Short Passing
- 78 Volleys

## Mentality

- 62 Aggression
- 38 Interceptions
- 88 Positioning
- 82 Vision
- 70 Penalties
- 86 Composure

## Skill

- 90 Dribbling
- 77 Curve
- 63 FK Accuracy
- 73 Long Passing
- 91 Ball Control

## Defending

- 34 Marking
- 34 Standing Tackle
- 32 Sliding Tackle

## Movement

- 96 Acceleration
- 96 Sprint Speed
- 92 Agility
- 87 Reactions
- 83 Balance

## Goalkeeping

- 13 GK Diving
- 5 GK Handling
- 7 GK Kicking
- 11 GK Positioning
- 6 GK Reflexes

## Power

- 80 Shot Power
- 75 Jumping
- 83 Stamina
- 74 Strength
- 79 Long Shots

## Traits

- Finesse Shot
- Flair
- Speed Dribbler (CPU AI Only)
- Technical Dribbler (CPU AI Only)

# Data Cleaning: Numeric Values

- › Adjusting Wage, Value and Release Clause columns to numeric fields
  - › €226.5 M to 226500000.00
  - › €170.5 K to 170500.00

# Data Cleaning: Simplified Position

- › Combining all of the positions into 4 simplified positions

**Goalkeeper:** GK

**Defender:** LWB, RWB, LB, LCB, CB, RCB, RB

**Midfielder:** LAM, CAM, RAM, LM, LCM, CM, RCM, RM, LDM, CDM, RDM

**Attacker:** LS, ST, RS, LW, LF, CF, RF, RW

- › Added new column to DataFrame called "Simplified Position"



# Data Cleaning: Creating Leagues

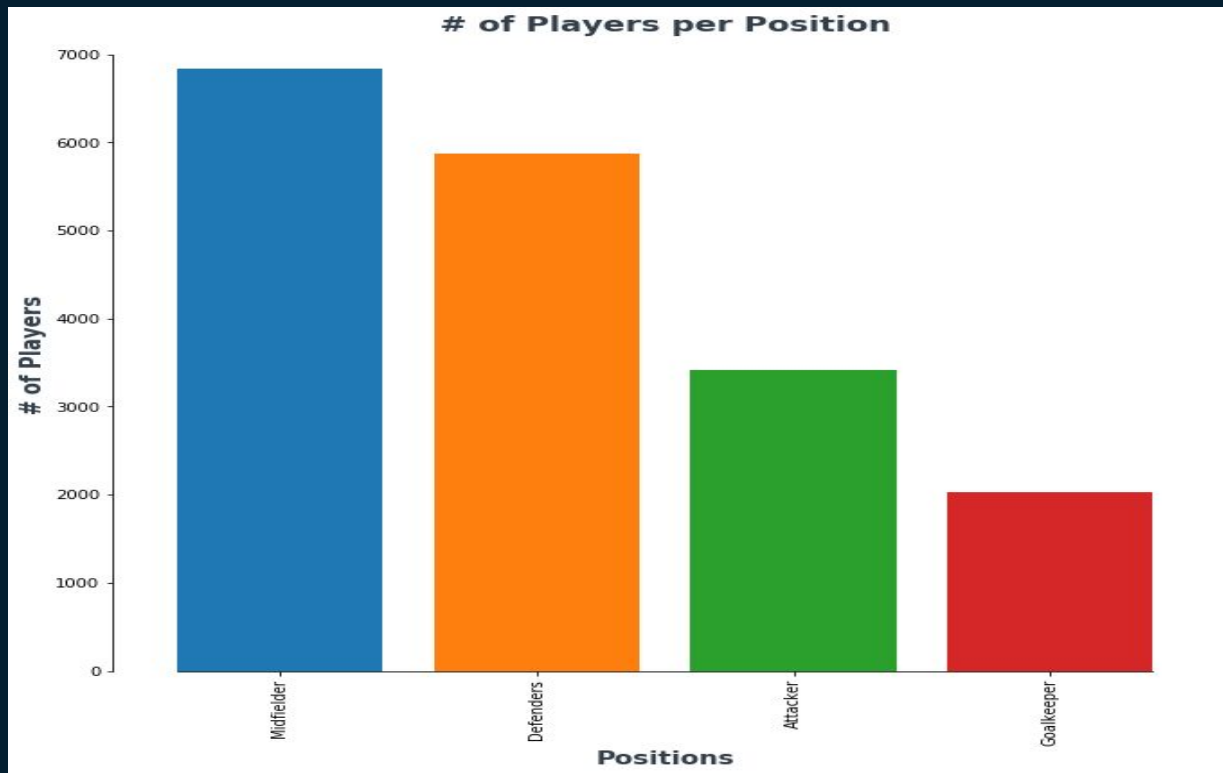
- › Placed all club teams in their respective leagues

```
la_liga = ['Deportivo Alavés', 'FC Barcelona', 'Atlético Madrid', 'Real Madrid',  
premier_league = ['Manchester City', 'Liverpool', 'Chelsea', 'Tottenham Hotspur', 'Ar  
seria_a = ['Juventus', 'Napoli', 'Atalanta', 'Internacional', 'Milan', 'Roma', 'T  
bundes_liga = ['FC Bayern München', 'Borussia Dortmund', 'RB Leipzig', 'Bayer 04  
league_1 = ['Paris Saint-Germain', 'LOSC Lille', 'Olympique Lyonnais', 'AS Saint-
```

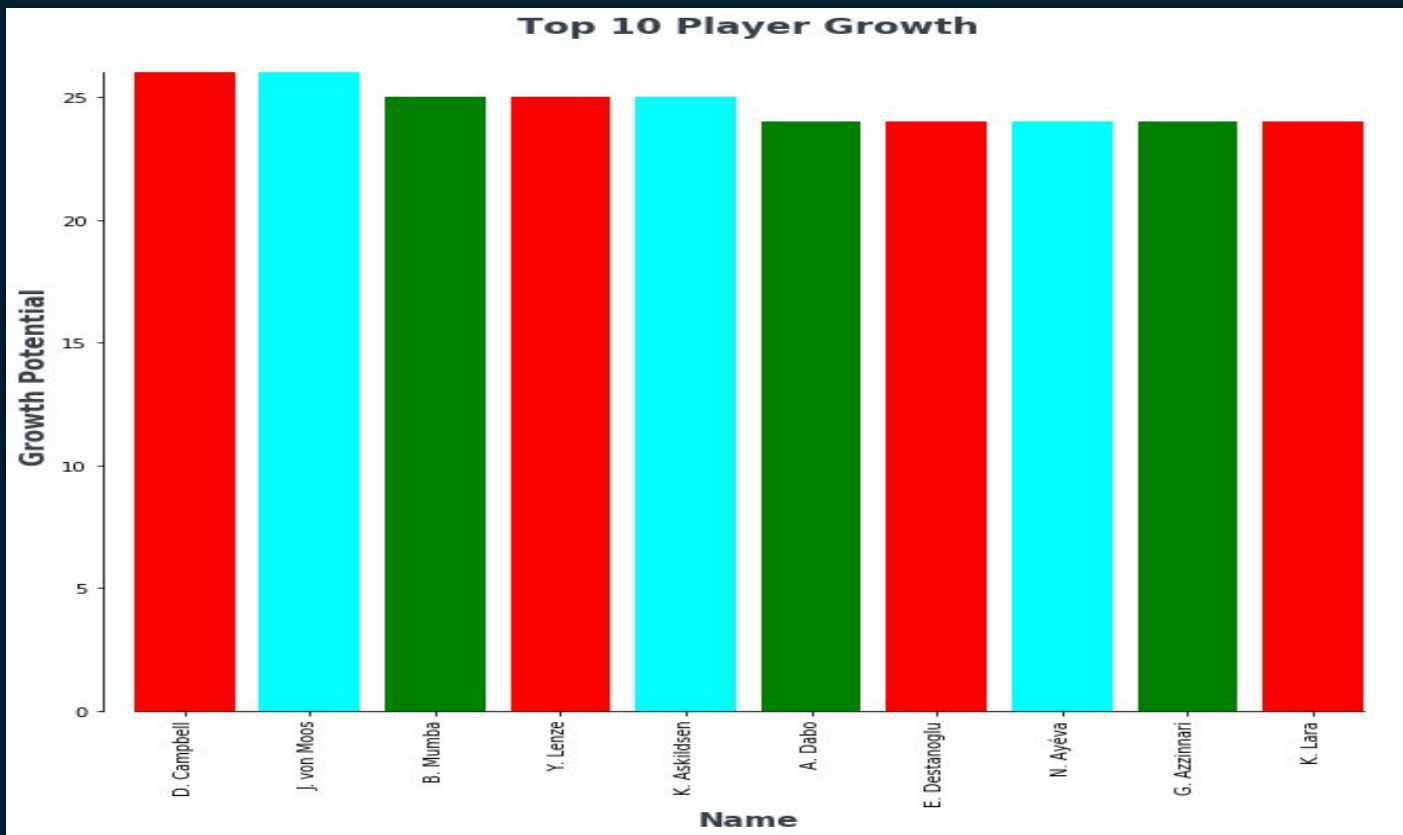
›

- › Added a new column to the DataFrame called "League"

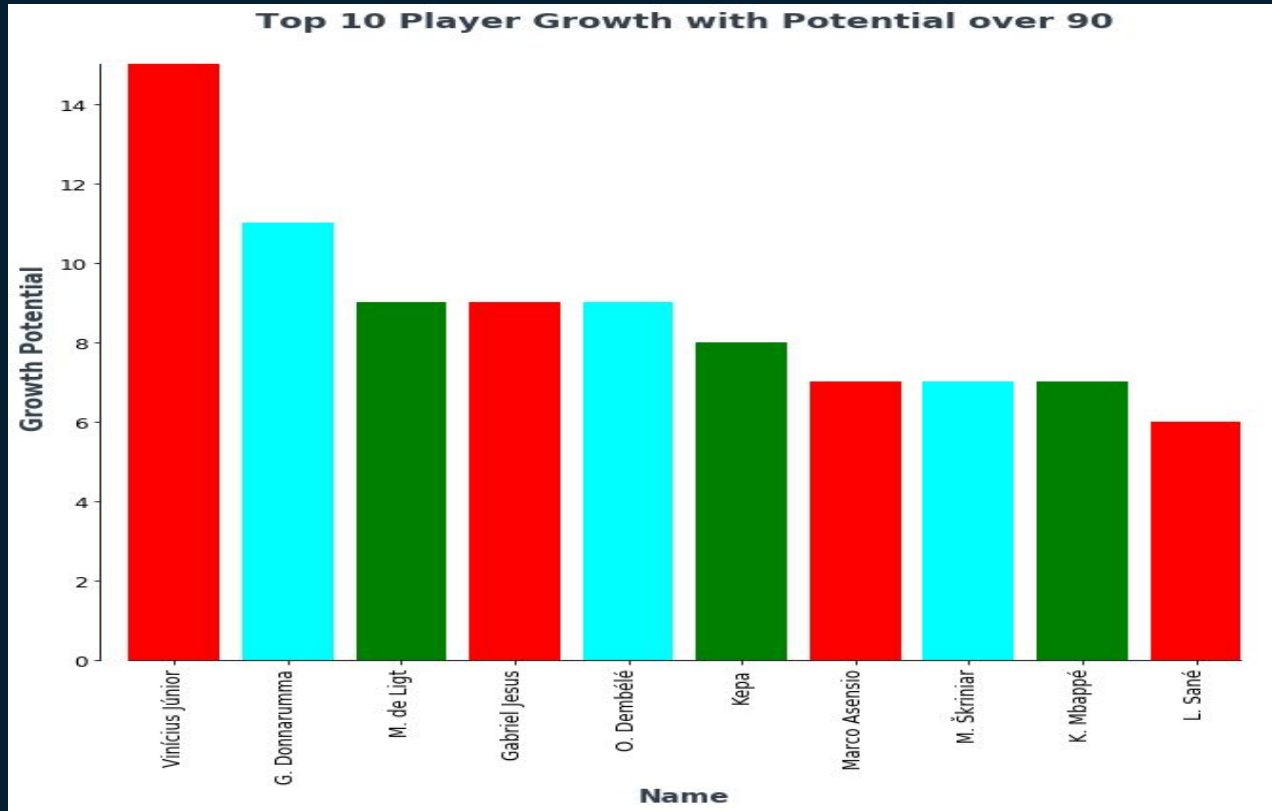
# Players per Position



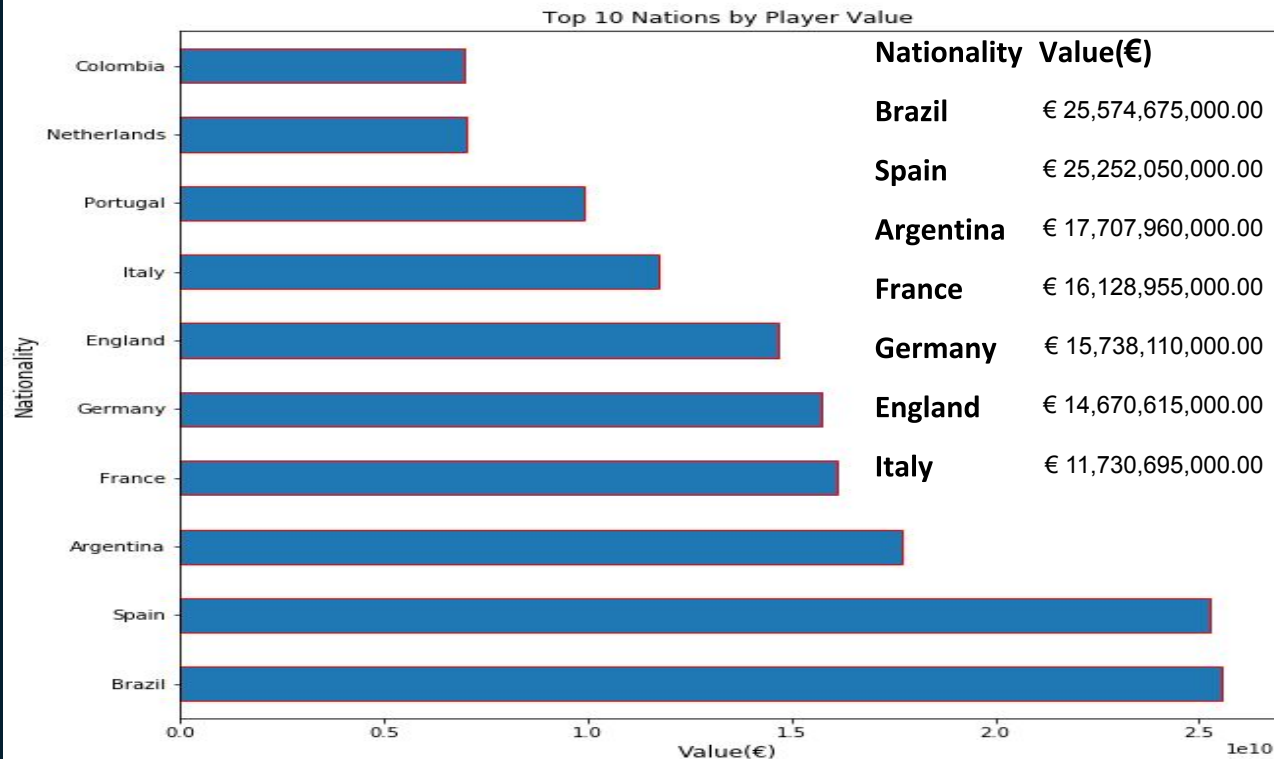
# Top 10 Player Growth



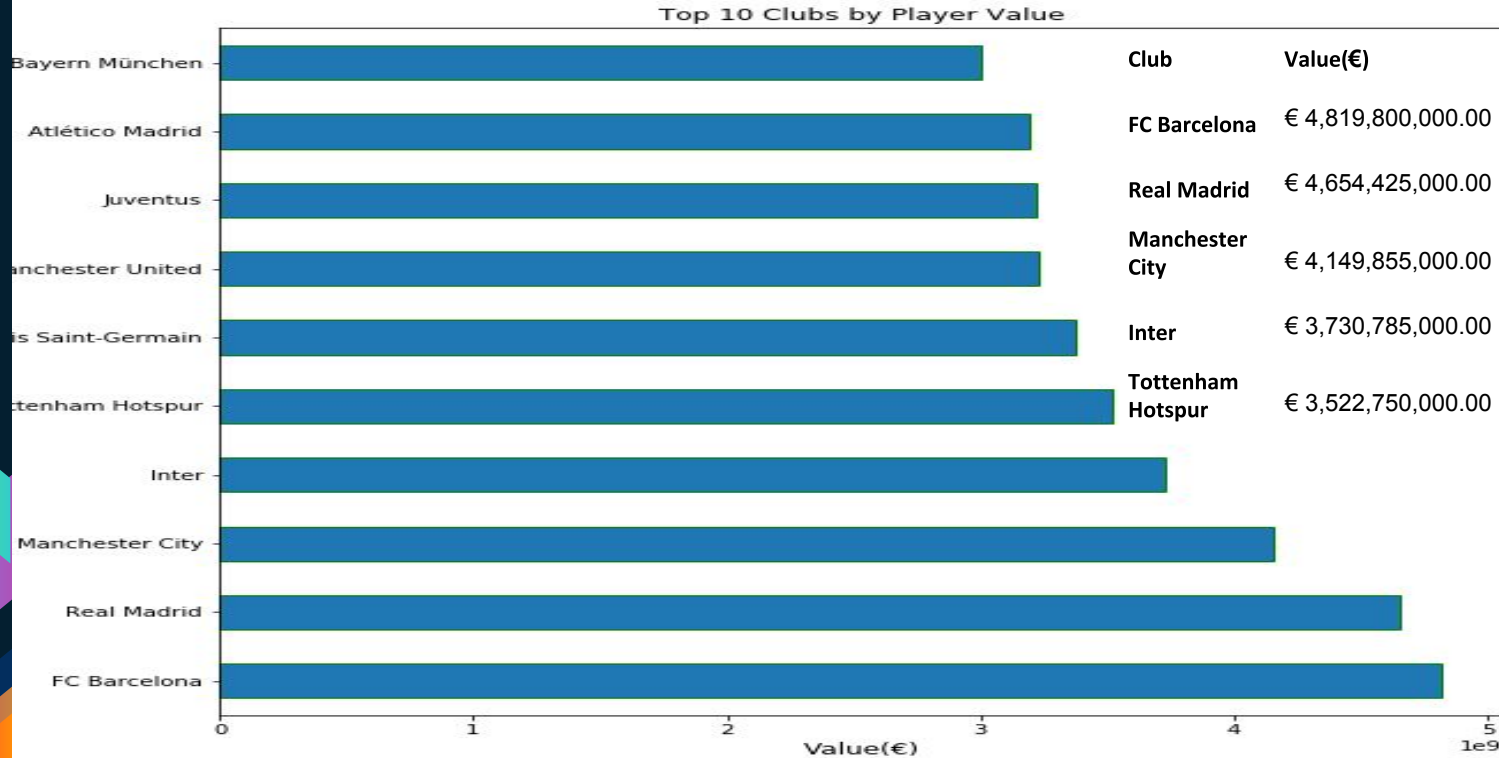
# Top 10 Player Growth with Potential over 90



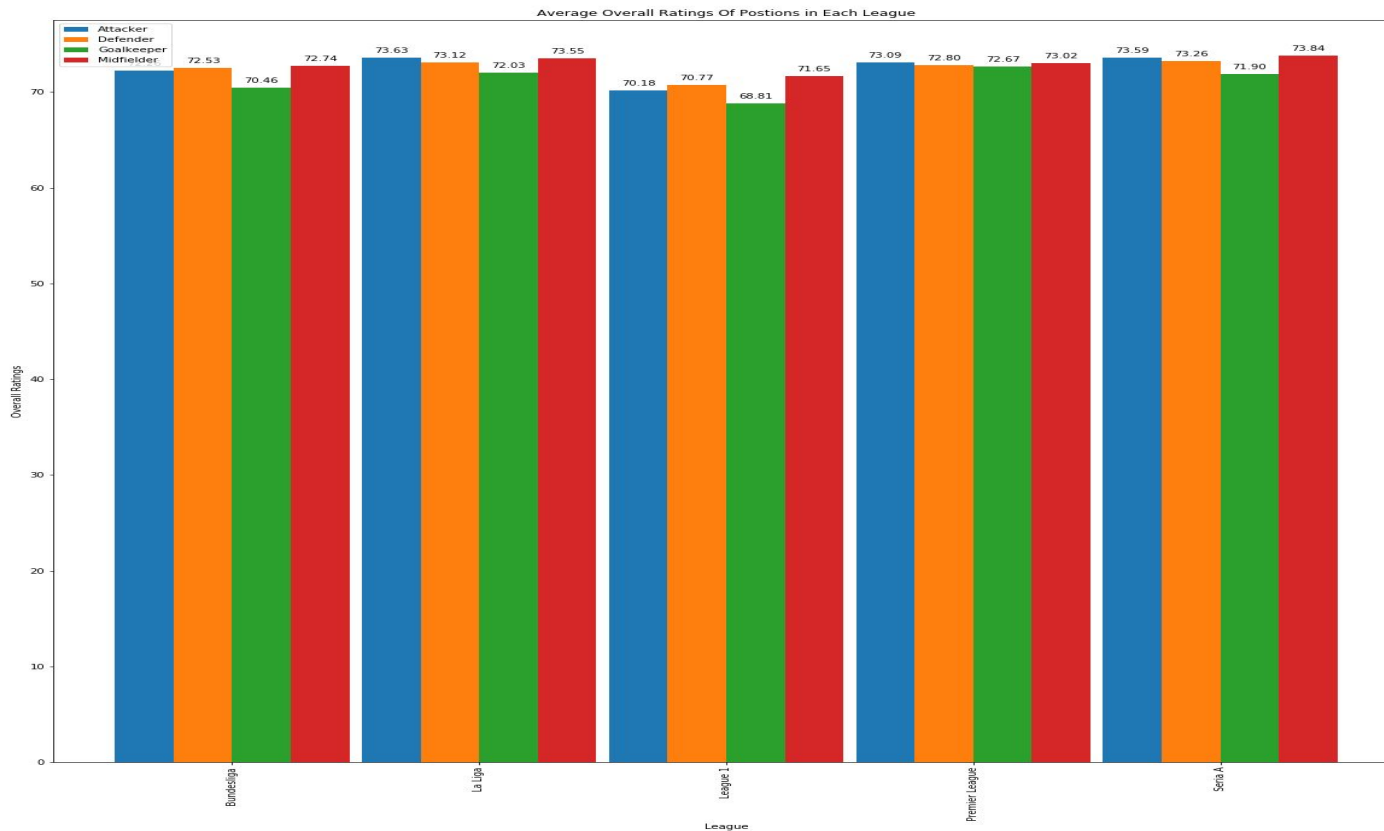
# Top 10 Nations by Player Value(€)



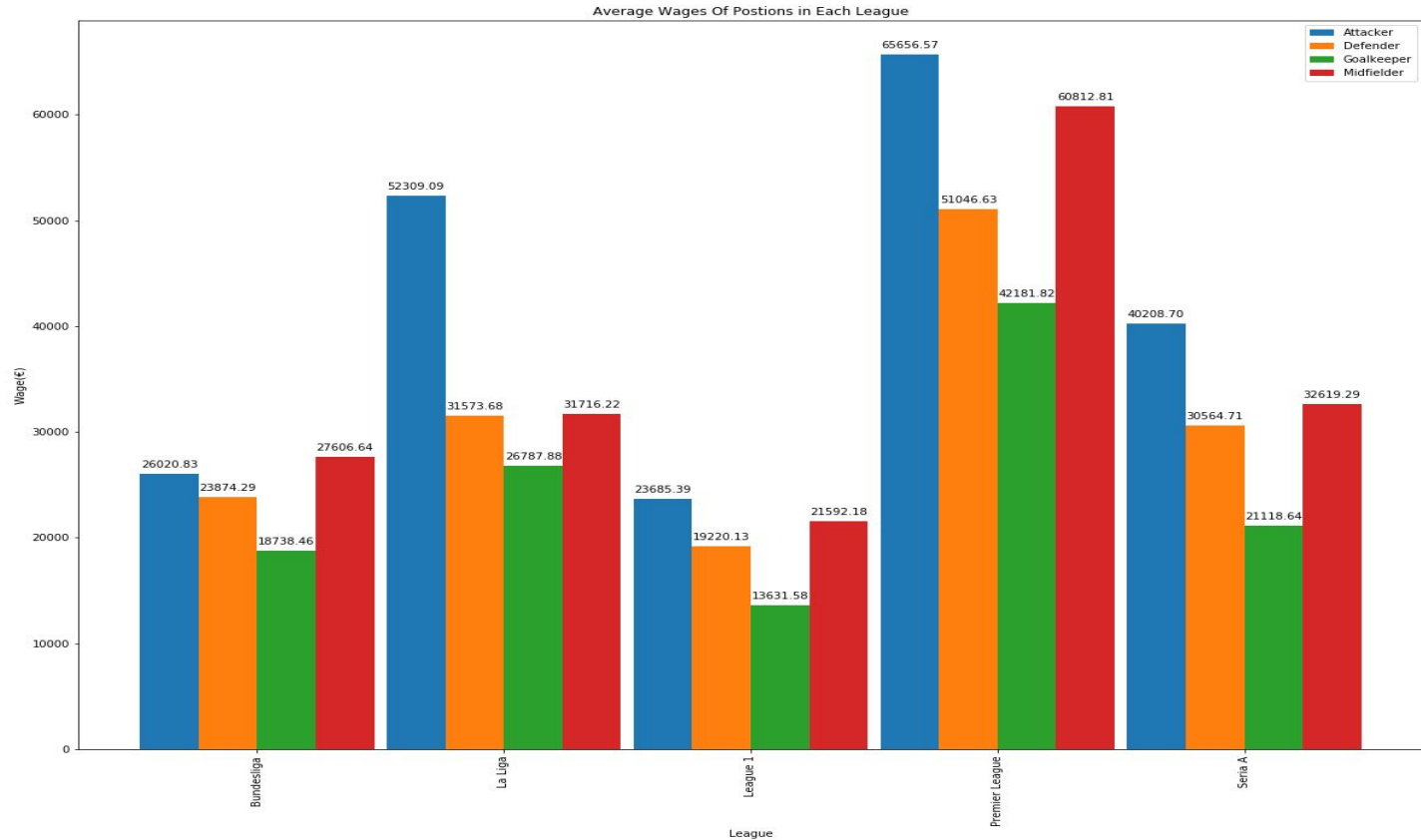
# Top 10 Clubs by Player Value(€)



# Average Overall Rating of Positions in Each League



# Average Wage of Positions in Each League

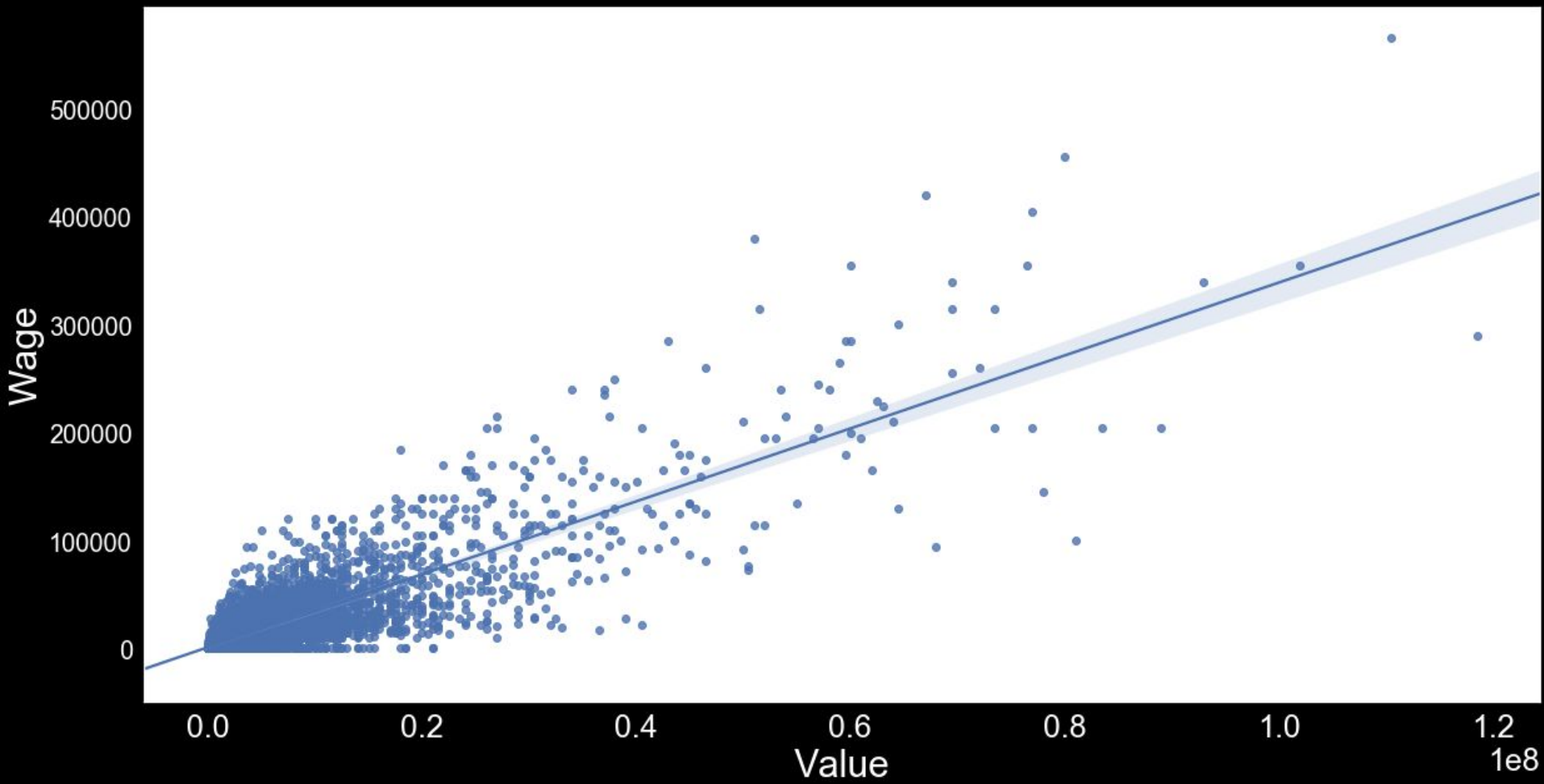




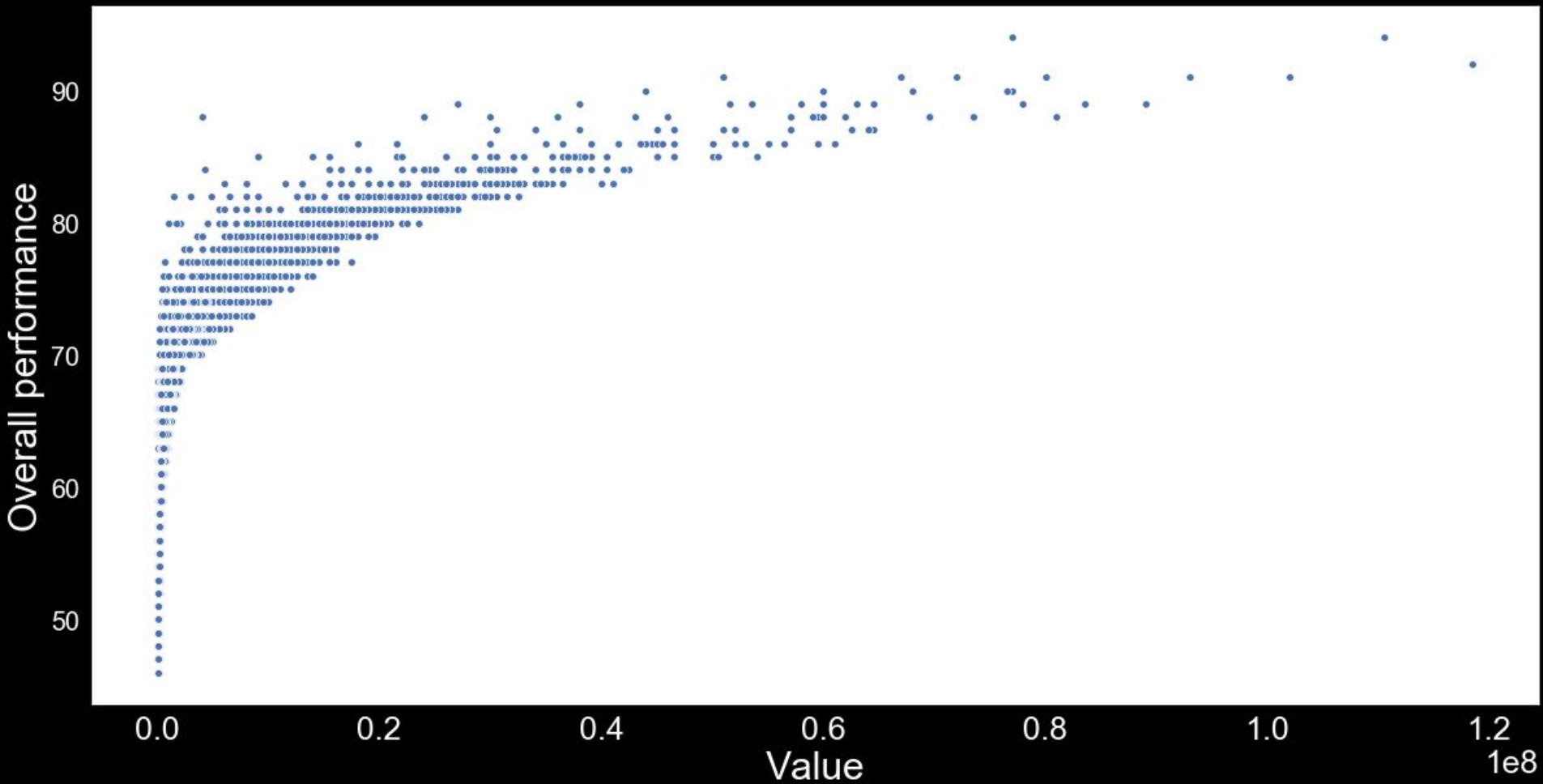
# Player's market Value, Wage and Overall Performance

What is the relationship between market value, wage and overall performance?

# Wage VS Value distribution



# Overall performance VS Value



# Overall performance VS Wage

Overall performance

90

80

70

60

50

0

100000

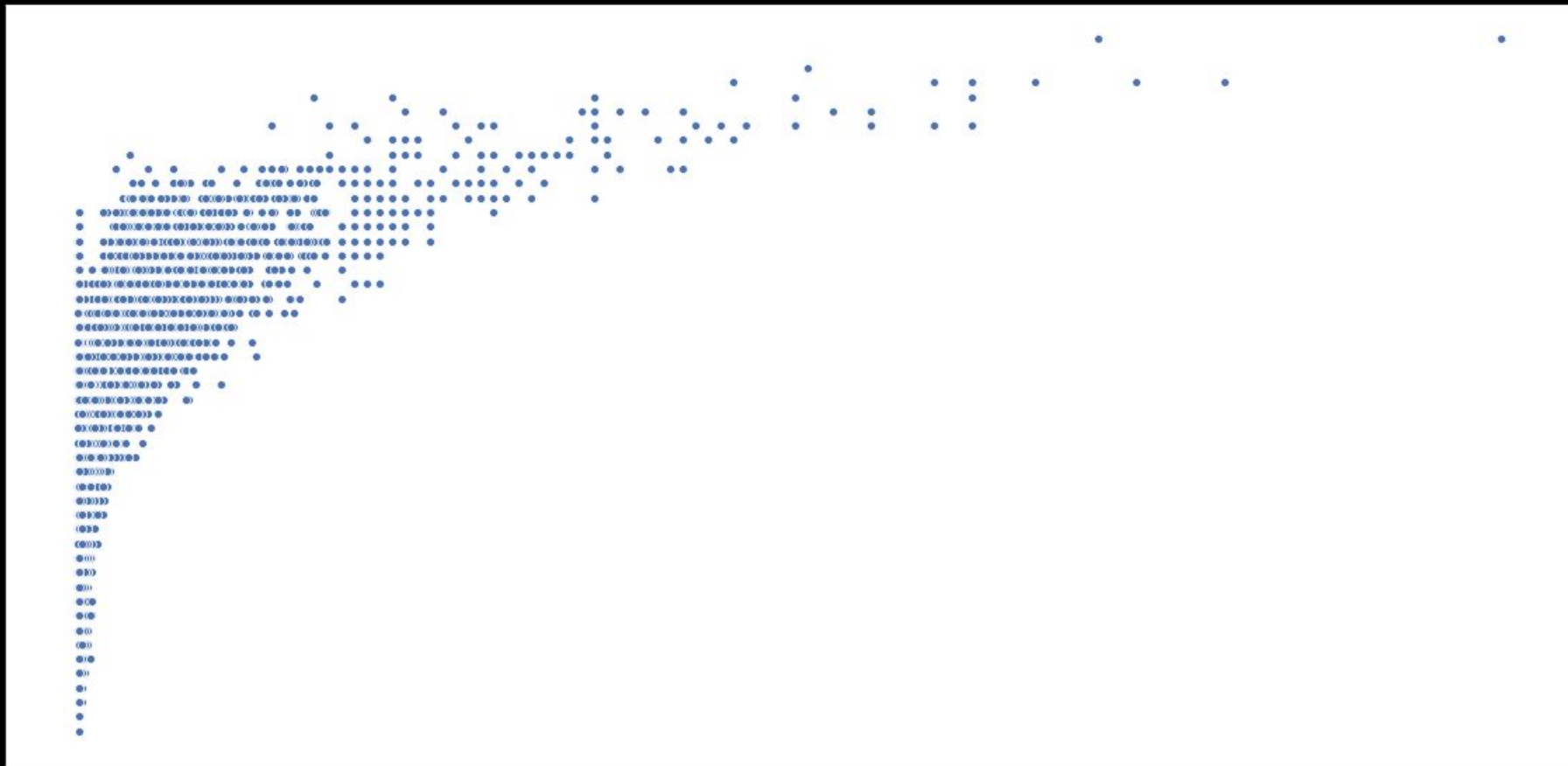
200000

300000

400000

500000

Wage



# Player's Value VS Skill Moves

Skill Moves

5  
4  
3  
2  
1  
0

0.0

0.2

0.4

0.6

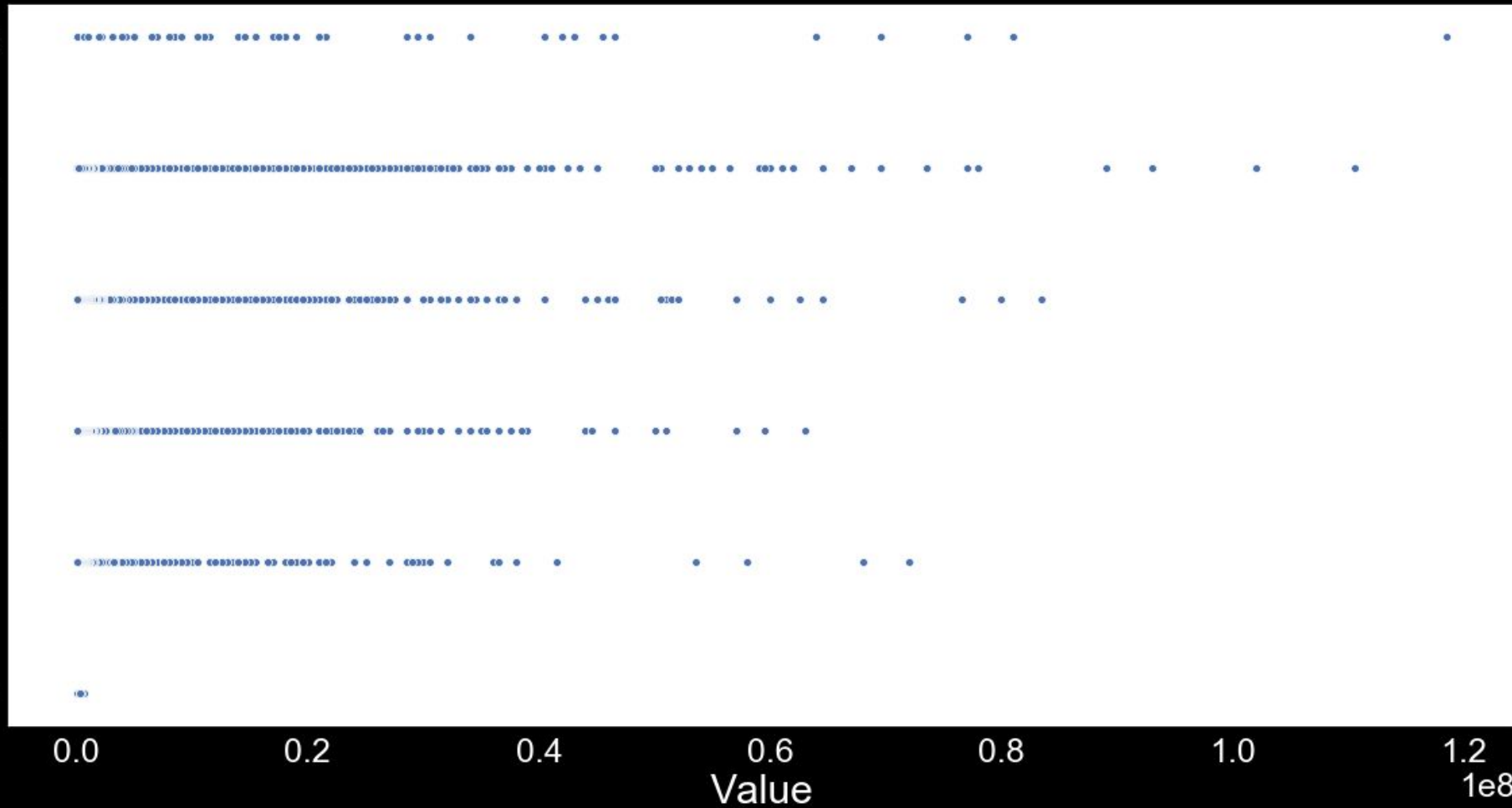
0.8

1.0

1.2

1e8

Value



## Regression Analysis - Value vs. Overall performance, Wage, Potential, Skill moves, and Age

```
# Regression value(dependent var.), overall performance, wage, potential, skill moves and age as independent variables
df = df.fillna(0)
x = np.array(df[['Overall','Wage','Potential','Skill Moves', 'Age']])
y = np.array(df['Value'])
model = LinearRegression().fit(x, y)
r_sq = model.score(x, y)
print('coefficient of determination:', r_sq)
print('intercept:', model.intercept_)
print('slope:', model.coef_)
y_pred = model.predict(x)
print('predicted response:', y_pred, sep='\n')
y_pred = model.intercept_ + np.sum(model.coef_ * x, axis=1)
print('predicted response:', y_pred, sep='\n')
print('median diff Value:', (y_pred-df['Value']).median())
print('median diff Overall:', (y_pred-df['Overall']).median())
```

coefficient of determination: 0.7836334584782609

intercept: -9723019.622792415

slope: [ 2.36218320e+05 1.82036365e+02 -9.50909932e+03 1.01320772e+05  
-1.92860539e+05]

predicted response:

[ 1.08864800e+08 7.94545811e+07 5.94074954e+07 ... -1.95895894e+06  
-2.14231038e+06 -2.18566816e+06]

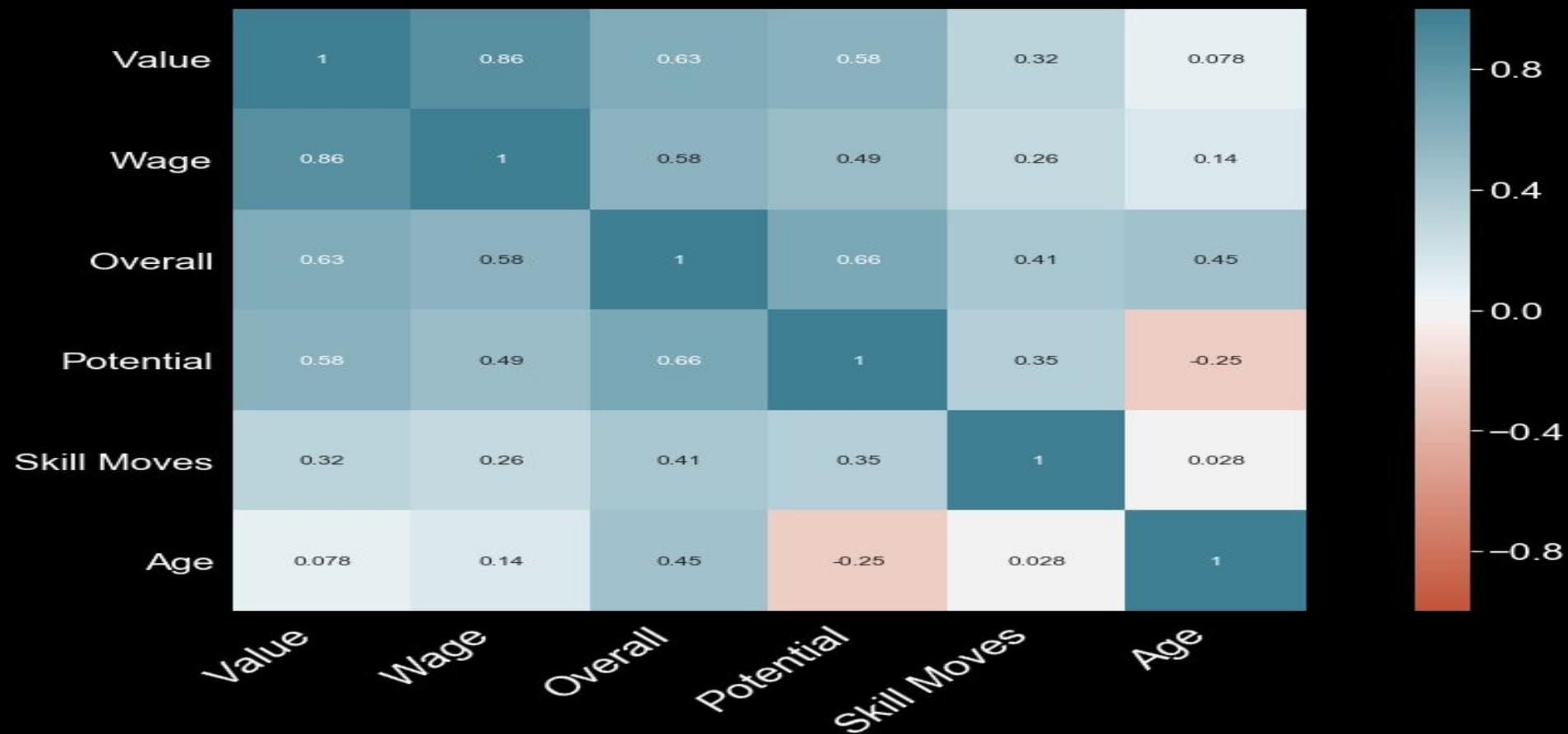
predicted response:

[ 1.08864800e+08 7.94545811e+07 5.94074954e+07 ... -1.95895894e+06  
-2.14231038e+06 -2.18566816e+06]

median diff Value: 96146.89623352326

median diff Overall: 1179331.1768911183

# Matrix correlation



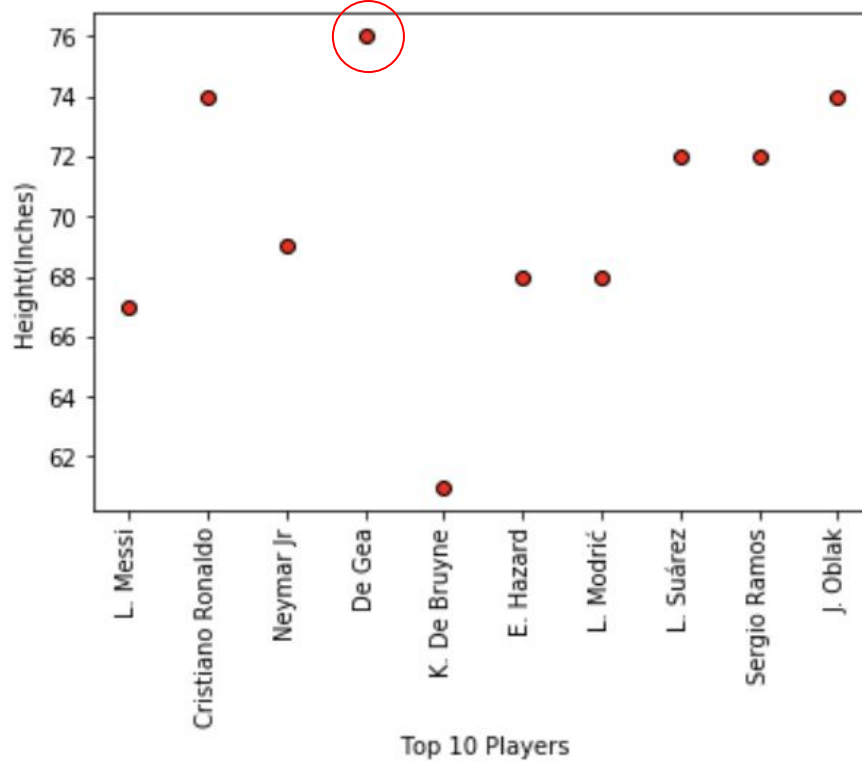
# Does FIFA strongly use height to a player's overall rating?



The player on the right is taller, is he valued more in the world of FIFA?



## The best 10 players in the world vs height

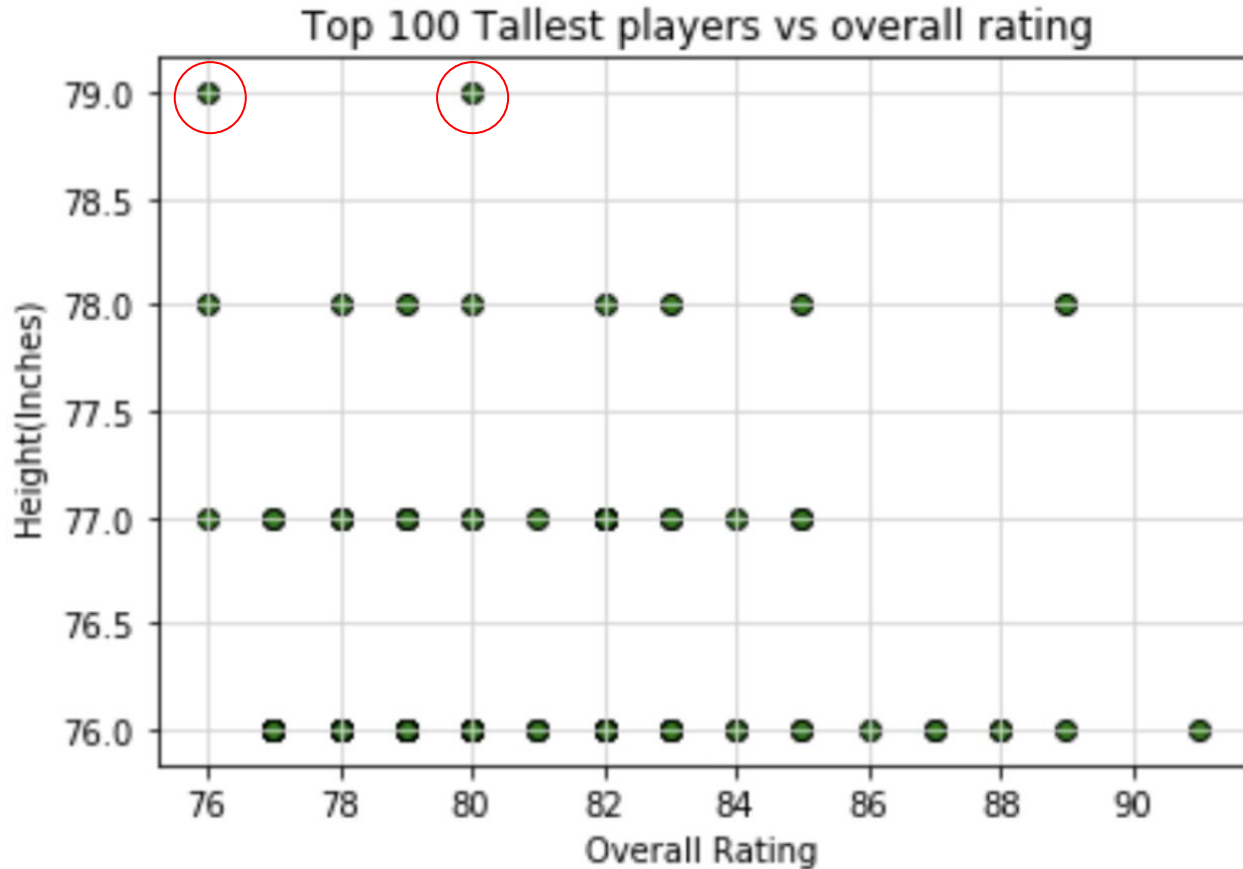


Rating increases

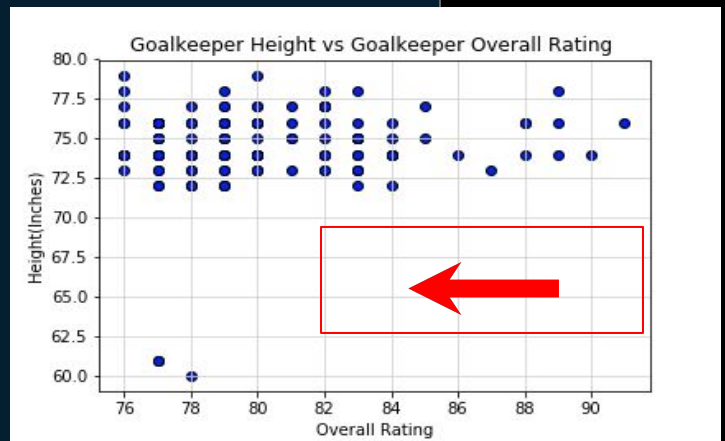
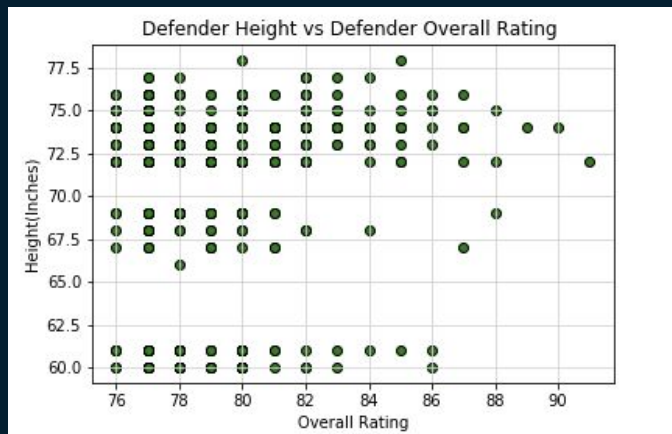
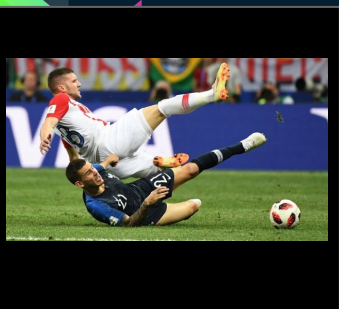
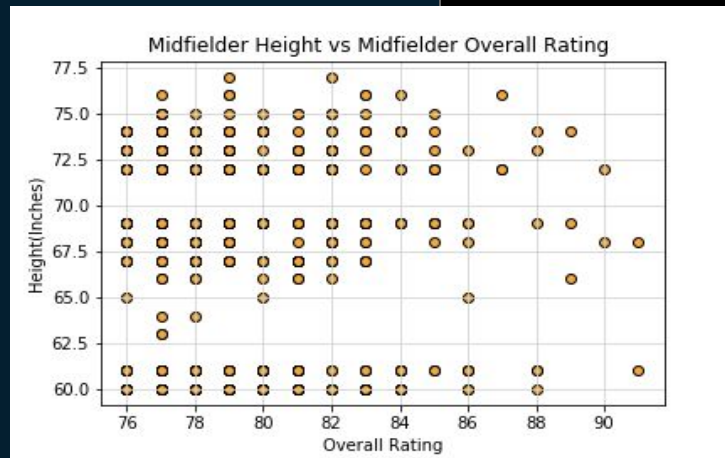
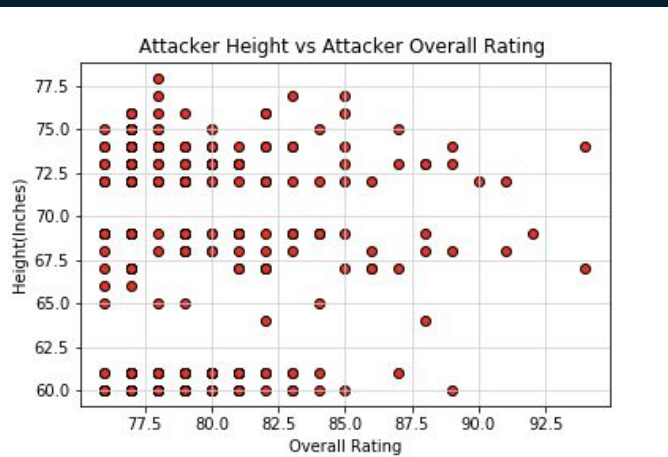


**David DeGea, best rated tallest player is a goalkeeper.**  
**Only players in soccer allowed to use their hands**

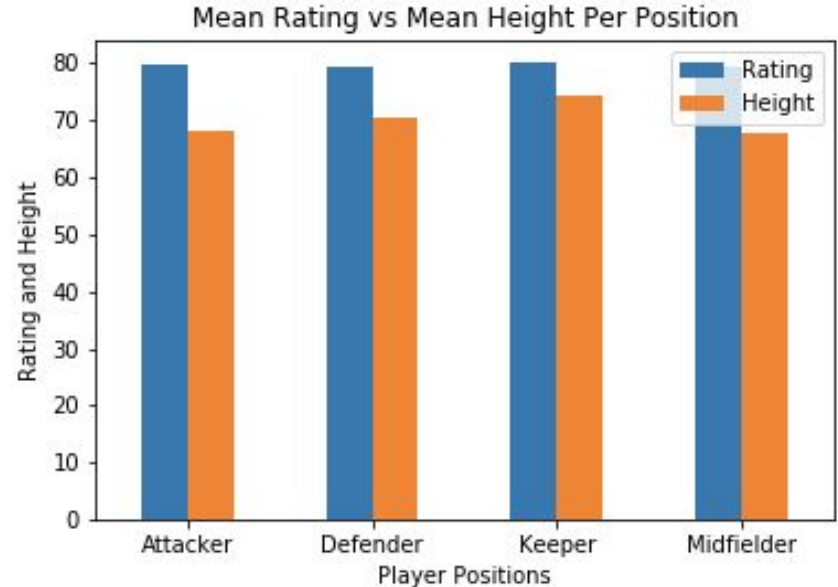
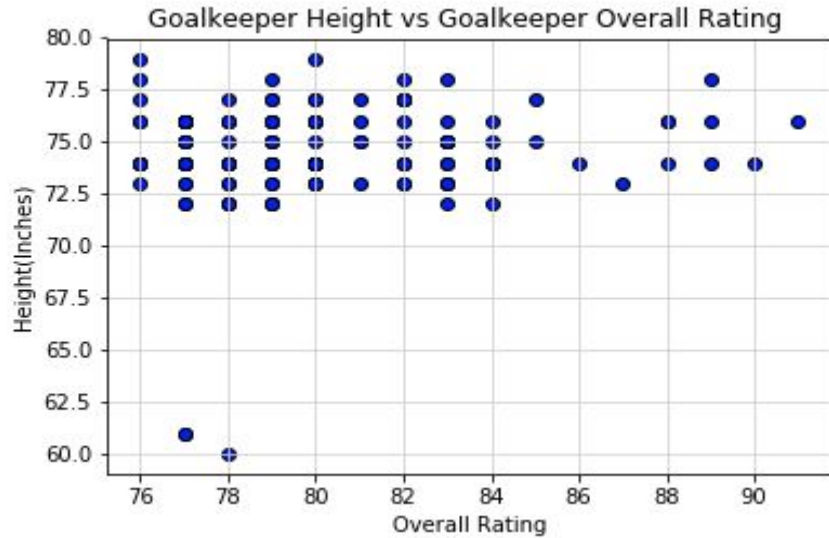
# The Tallest 100 players and their rating



# Player Positions Height vs Rating



# Looking at goalkeepers closely



```
In [52]: gb_keeper_mean = gb_keeper.mean()  
gb_keeper_mean
```

```
Out[52]: Overall      80.246269  
Height(Inches)  74.492537  
dtype: float64
```

The height mean is  
taller by 4 inches  
than other positions



**Height is important for a goalkeeper in football,  
the best rated are taller .**



## Limitations:

- › Data only focused on 2018-19 player data
- › Data did not contain columns for leagues so we only focused on top 5 leagues in top 5 leagues
- › Can't quantify some attributes players possess
- › Can't quantify the impact coaches have on players

# Summary:

- › D. Campbell and J. Von Moos will see the most overall improvement.
- › Vinicius Junior has the most potential to be the next superstar.
- › There is high correlation between value and wages. Medium correlation between overall performances, potential and financial returns. Low to negative correlation among skill moves, age and financial metrics.
- › Players positions & wages in each league does not affect their overall ratings.
- › A clubs value does not correlate to a players rating or wages.
- › South America outperformed in talent by nation based on player value however the highest player value at the club level came from Europe.
- › A player's overall rating is not evidently affected by their own height, unless you are a goalkeeper playing on the world stage!

Thank you!

