

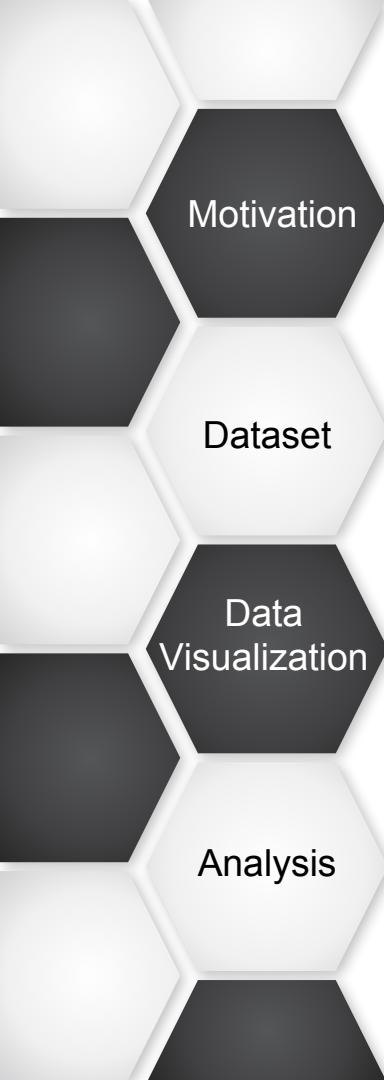
# **Visualisation & Analysis Of FIFA Player Attribute Dataset**

**TEAM 3**

**Sanjay  
John**

**Debosmit  
Majumder**

**Unnikrishnan  
Sivaprasad**



Motivation

Dataset

Data  
Visualization

Analysis

- **Sports Analytics is huge, will worth \$4.7 Billion by 2021.**
- **Our analysis techniques serve soccer; applications elsewhere**
- **Love for the game**
- **Using the open source SOFIFA dataset, collected from 25000+ matches**
- **Visual representation of the data provides useful insights**
- **Useful for both players and managers**
- **Insights obtained from visuals used for data analytics**
- **Fitting a machine learning model**

# Objective

- ❖ Scrape, organize and visualize the SOFIFA dataset
- ❖ Analyse useful trends, player correlation, recommending player positions to scouts

# Application

- ❑ Our analysis and visualizations will allow soccer clubs to make better decisions on player strategies
- ❑ Our machine learning model shows the importance of each attribute for each position
- ❑ These techniques can apply to any team sport using a similar dataset

# Dataset: Attributes

- ❑ Calculated by EA through weighted formula determined by experts
- ❑ Subjective data scored from 9000 coaches and scouts arranged by 300 editors
- ❑ Empirical data from sports analytics data scientists and outside agencies

# Dataset

20000  
Players

48  
Attributes

NAME	AG	OV.	PO	TEAM & CONTRACT	VALUE	WAGE
L. Messi CF ST RW	30	94	94	FC Barcelona 2004 ~ 2021	€118.5M	€565K
Cristiano Ronaldo LW ST	32	94	94	Real Madrid CF 2009 ~ 2021	€95.5M	€565K
Neymar LW	25	92	93	Paris Saint-Germain 2017 ~ 2022	€119.5M	€280K

Open Source dataset  
from SOFIFA

Scraped, cleaned  
and organized  
pandas dataframe

Attacking

77 Crossing  
95 Finishing  
71 Heading Accuracy  
88 Short Passing  
86 Volleys

Skill

97 Dribbling  
90 Curve  
92 FK Accuracy  
87 Long Passing  
96 Ball Control

Movement

92 Acceleration  
87 Sprint Speed  
90 Agility  
95 Reactions  
95 Balance

Power

85 Shot Power  
67 Jumping  
73 Stamina  
59 Strength  
88 Long Shots

ShortPassing	ShotPower	SlidingTackle	SprintSpeed	Stamina	StandingTackle	Strength	Vision	Volleys	age	club	club_pos	country	dob	full_name
88	85	26	87	73	28	59	92	86	30	RS	[{"Spain", "Argentina"}]	Jun 24, 1987	L. Messi	
83	94	23	91	92	31	80	85	88	32	LW	[{"Spain", "Portugal"}]	Feb 5, 1985	Cristiano Ronaldo	
82	80	33	90	78	24	53	83	83	25	LW	[{"France", "Brazil"}]	Feb 5, 1992	Neymar	

# Crawling and Scraping

## BeautifulSoup

L. Messi (ID: 158023)  
Lionel Messi Age 30 (Jun 24, 1987) 5'7" 159lbs  
Overall Rating 94 Potential 94 Value €118.5M Wage €565K

### Attacking

77 Crossing  
95 Finishing

### Skill

97 Dribbling  
90 Curve

### Movement

92 Acceleration  
87 Sprint Speed

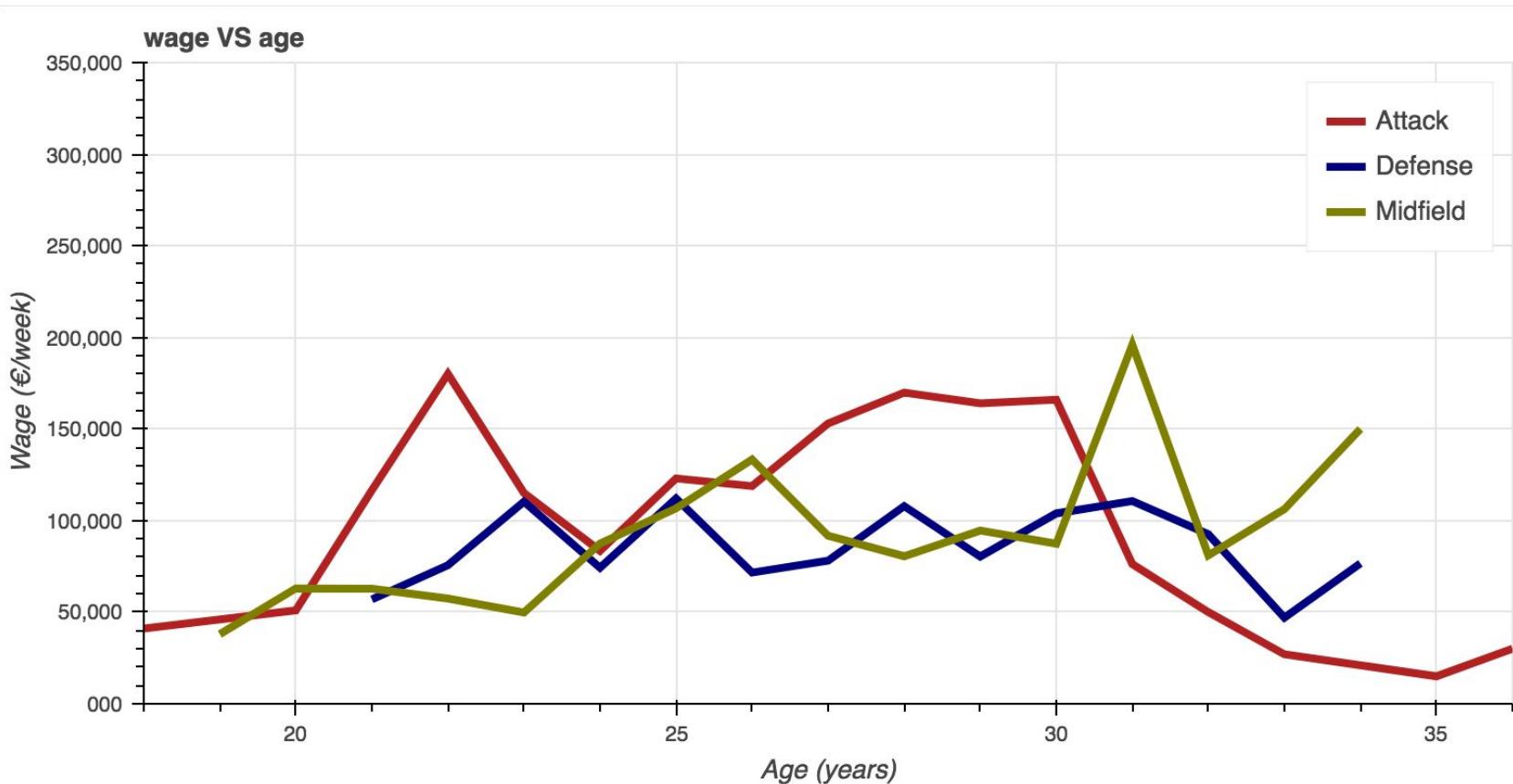
### Attacking

77 Crossing  
95 Finishing  
71 Heading Accuracy  
88 Short Passing  
86 Volleys

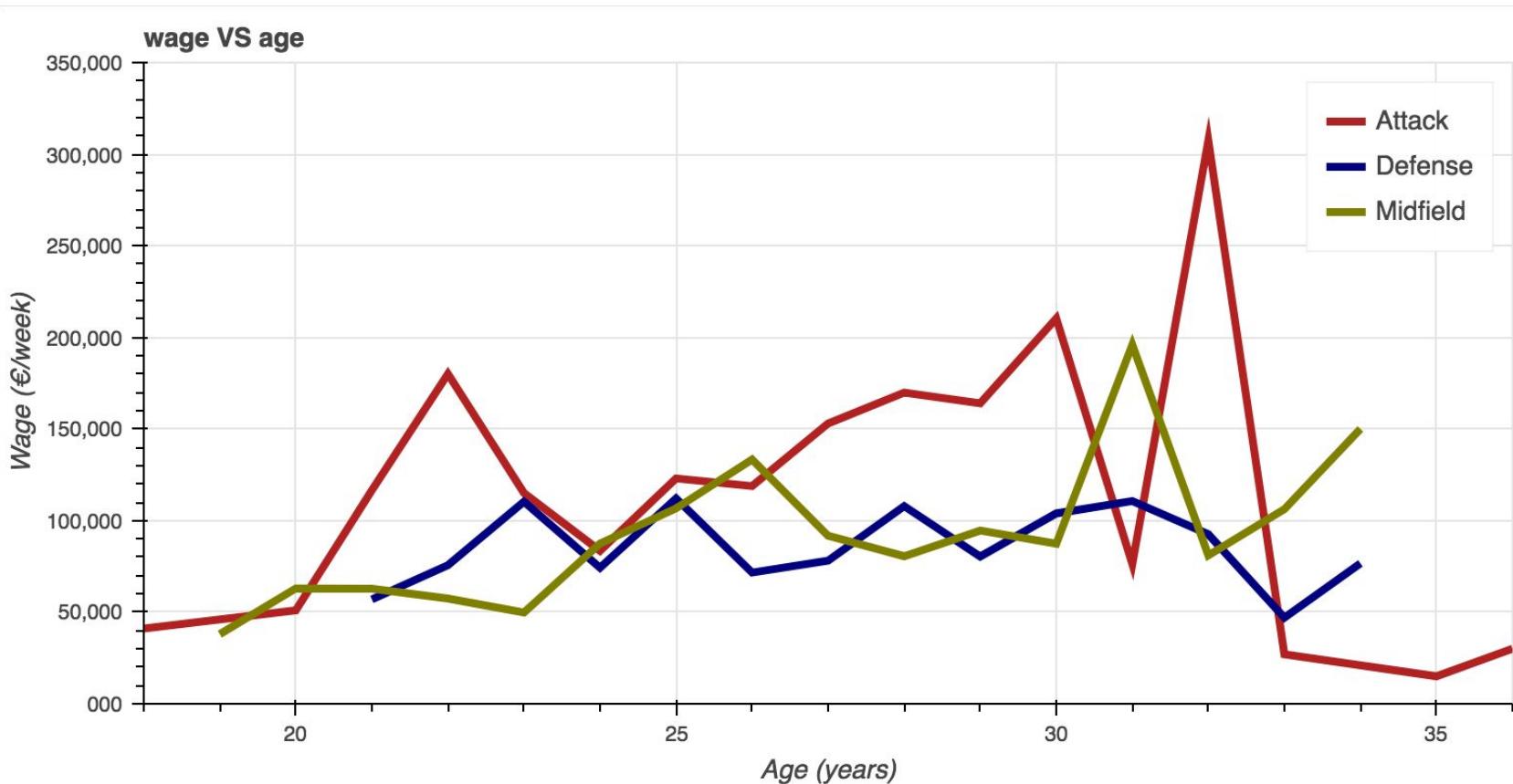
```
<!DOCTYPE html>
<html lang="en">
  >#shadow-root (open)
  ><head>...</head>
  ><body class="...>
    ><header>...</header>
    ><div class="sub-header">...</div>
    ><div class="hide" id="version-calendar">...</div>
    ><div class="container">...</div>
    ><script>...</script>
    ><section class="container">
      ><section class="columns">
        ><article class="column">
          ><div class="player">...</div>
        ><div class="columns">
          ... <div class="column col-3 mb-20">...</div> == $0
          ><div class="column col-3 mb-20">...</div>
          ><div class="column col-3 mb-20">
            ><div>
              ><h5>Movement</h5>
              ><ul class="pl1">...</ul>
            ></div>
          ></div>
          ><div class="column col-3 mb-20">...</div>
          ><div class="columns">...</div>
          ><div class="toast mb-20">...</div>
        ><div id="hit" class="hide">...</div>
```

# Visualization

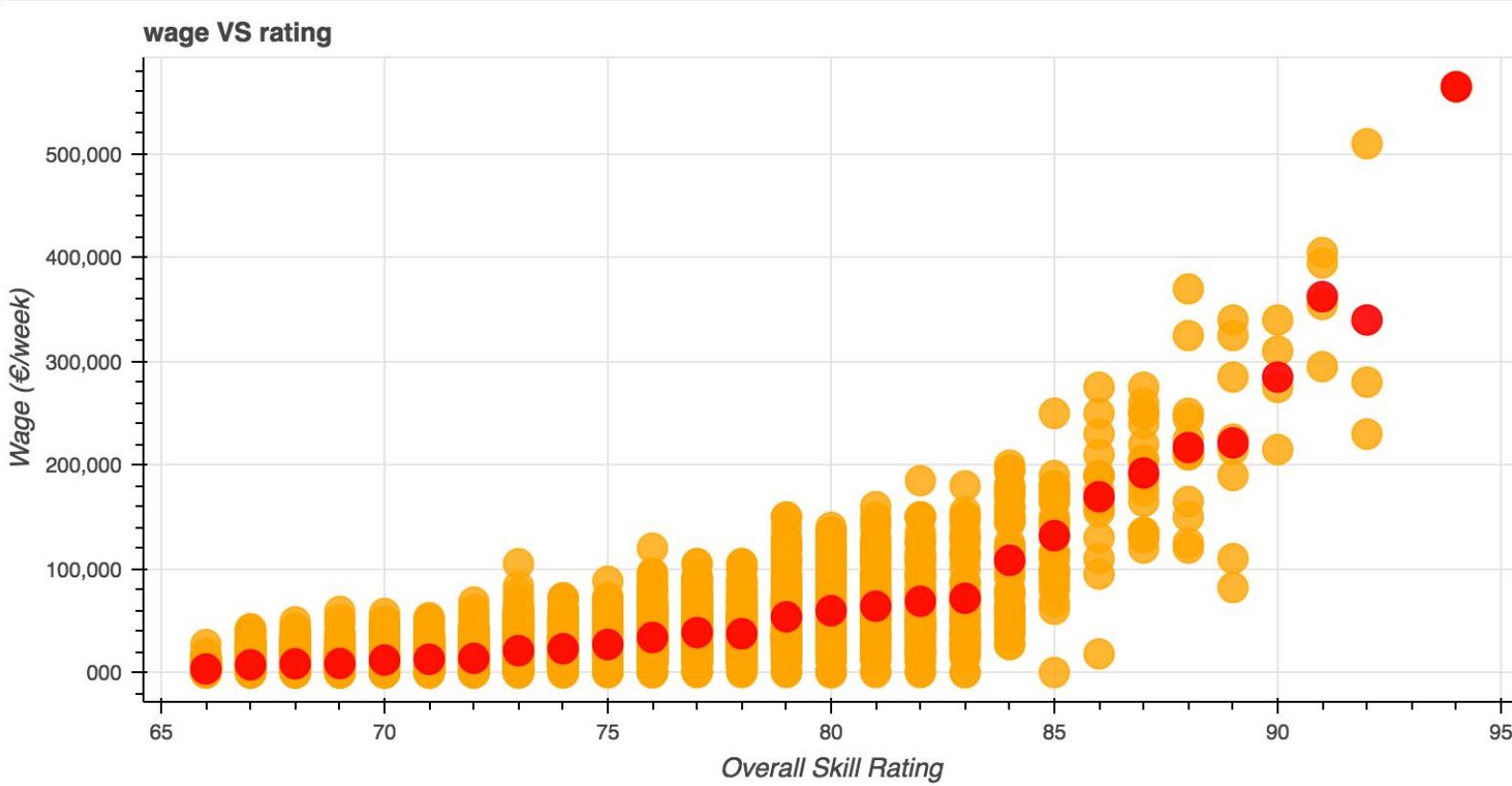
# Line Plot - Wage VS Age



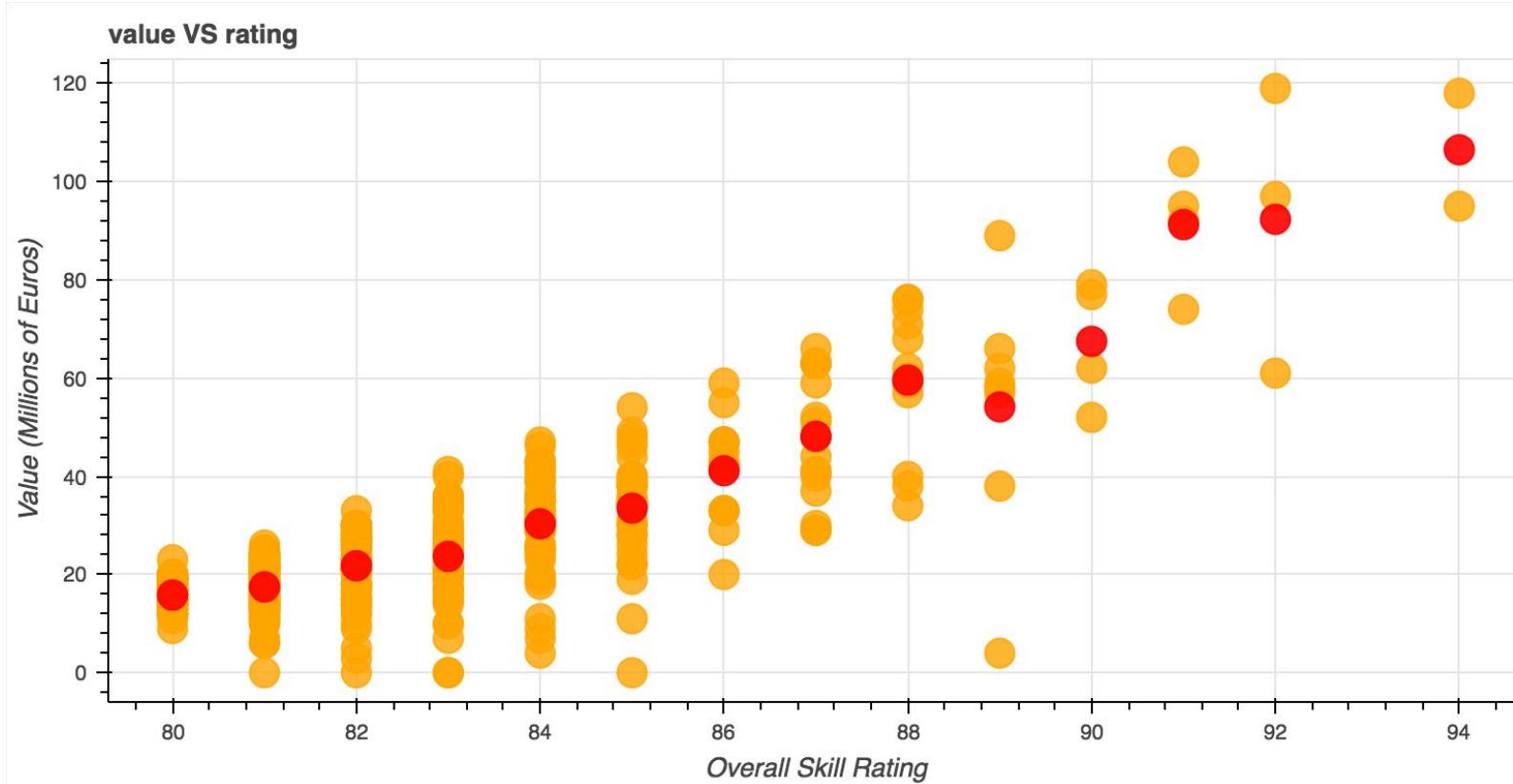
# Line Plot - Wage VS Age



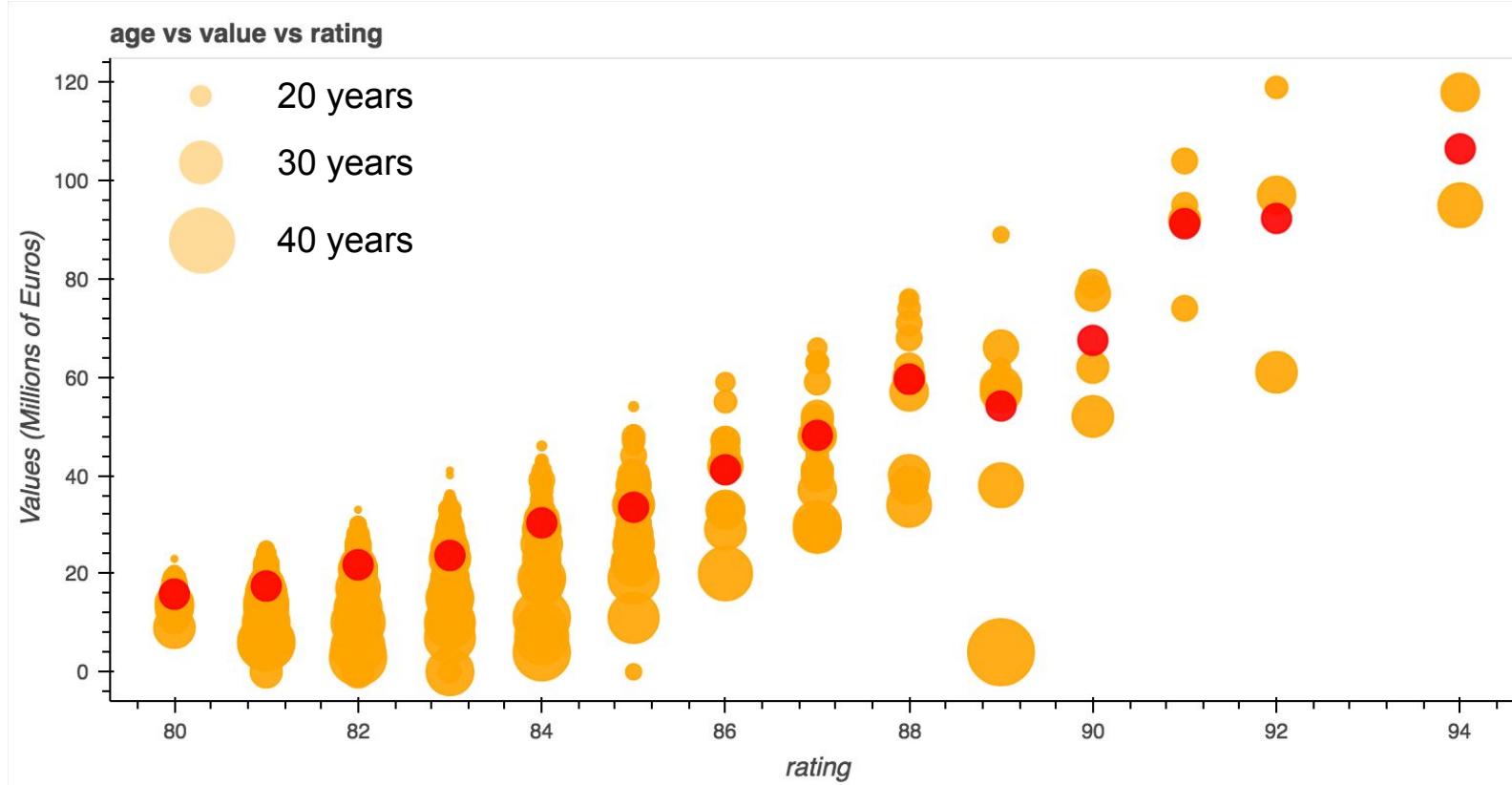
# Scatter Plot - Wage VS Rating



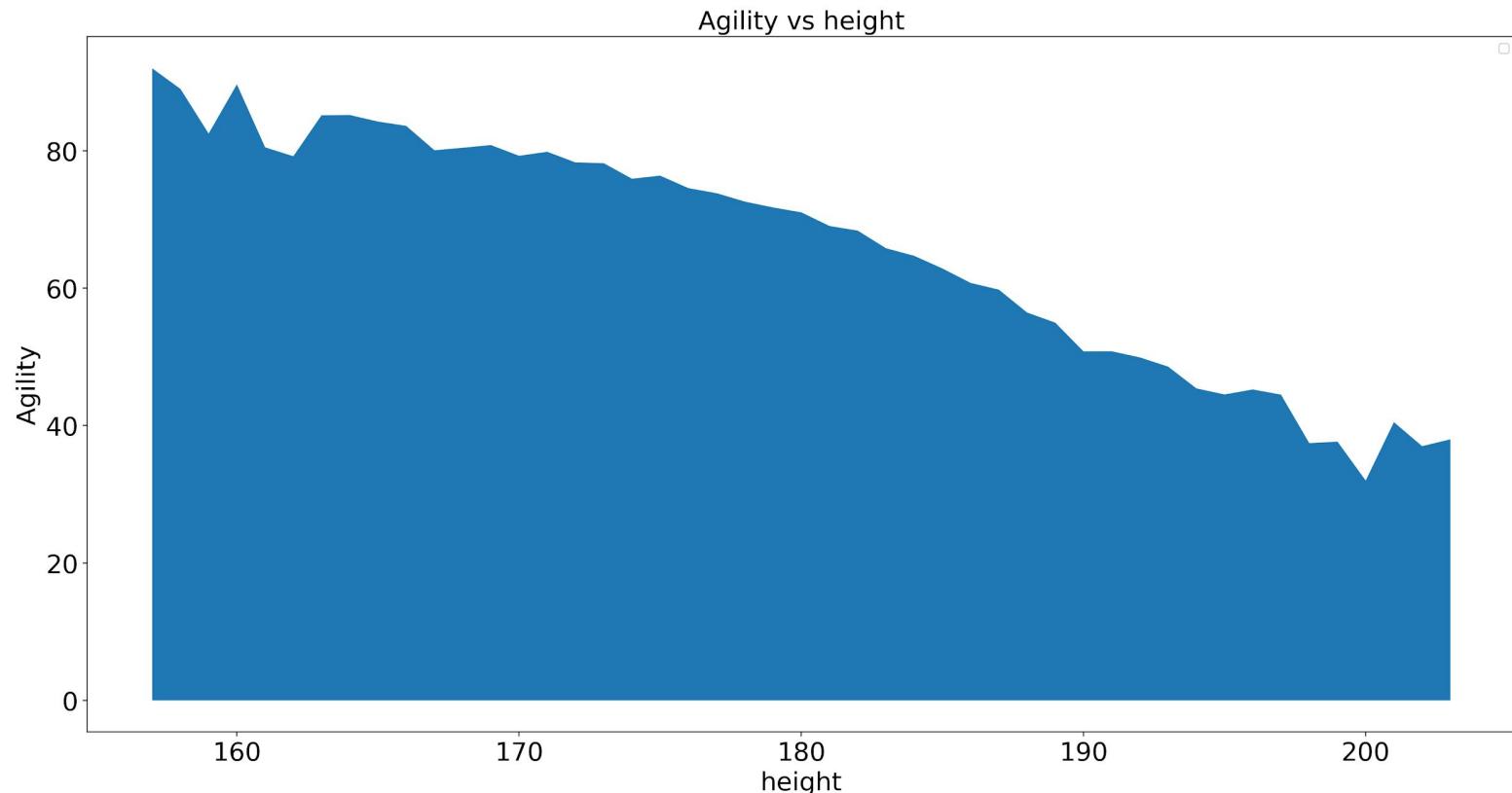
# Scatter Plot - Value VS Rating



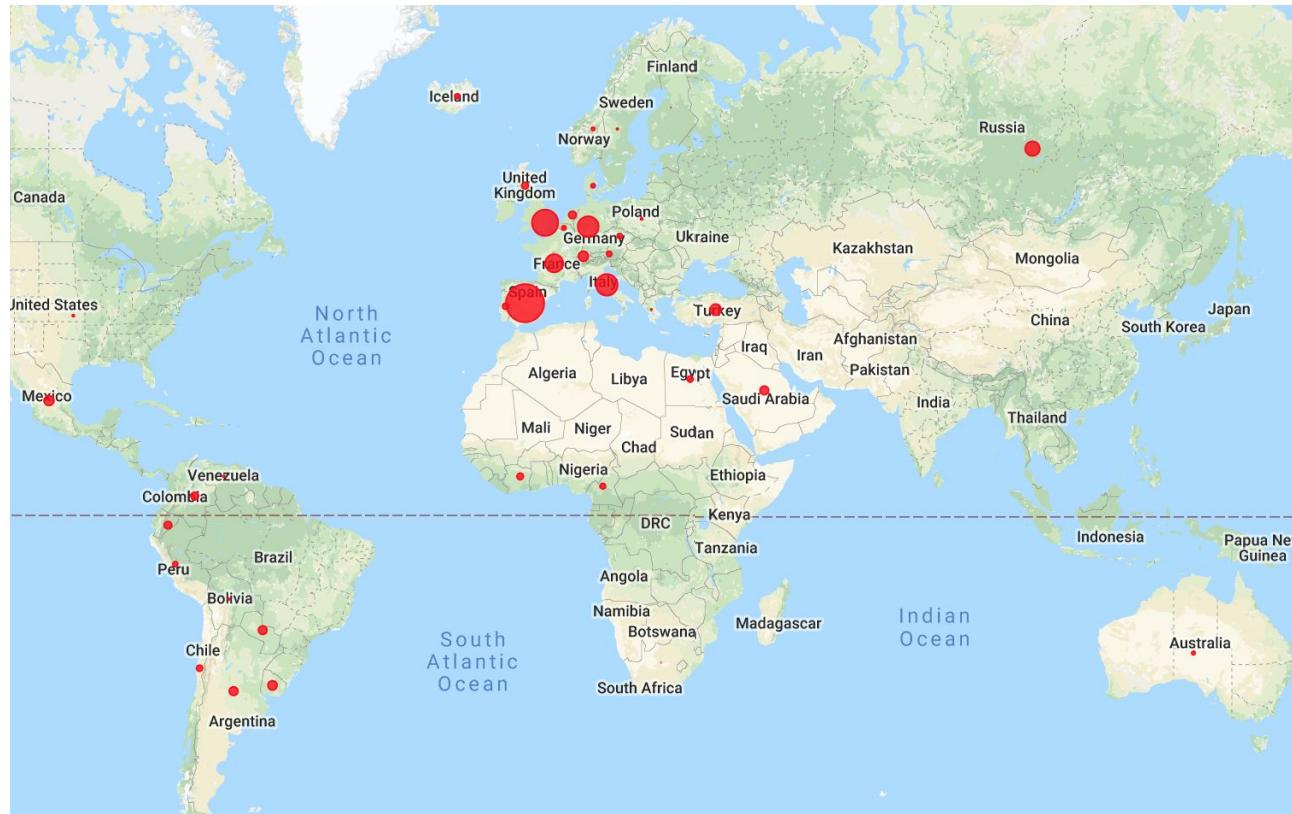
# Scatter Plot - Value VS Rating



# Bar Plot - Agility VS Height



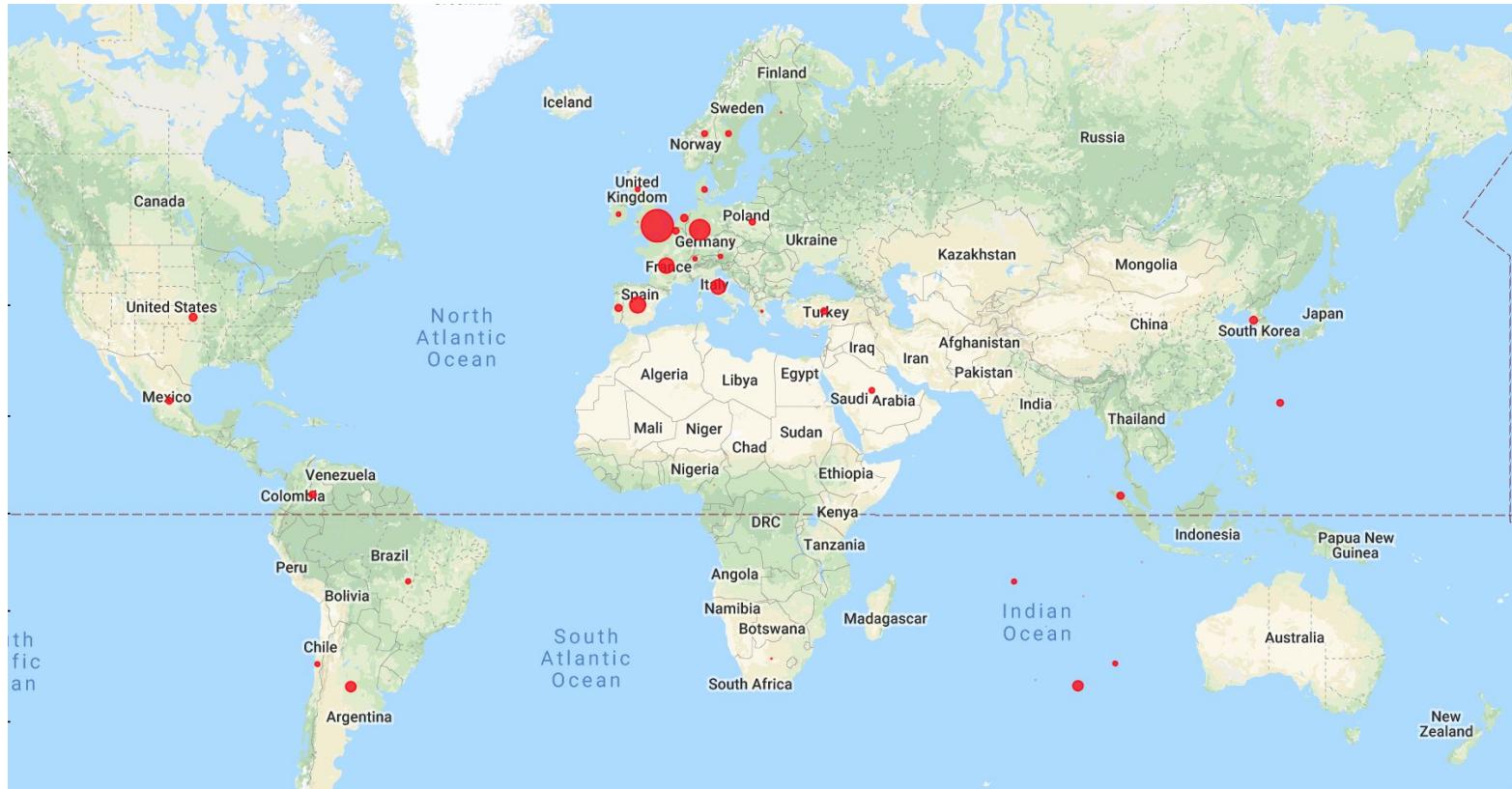
# World Map - Where Is The Money?



# Where Is The Money? - Europe



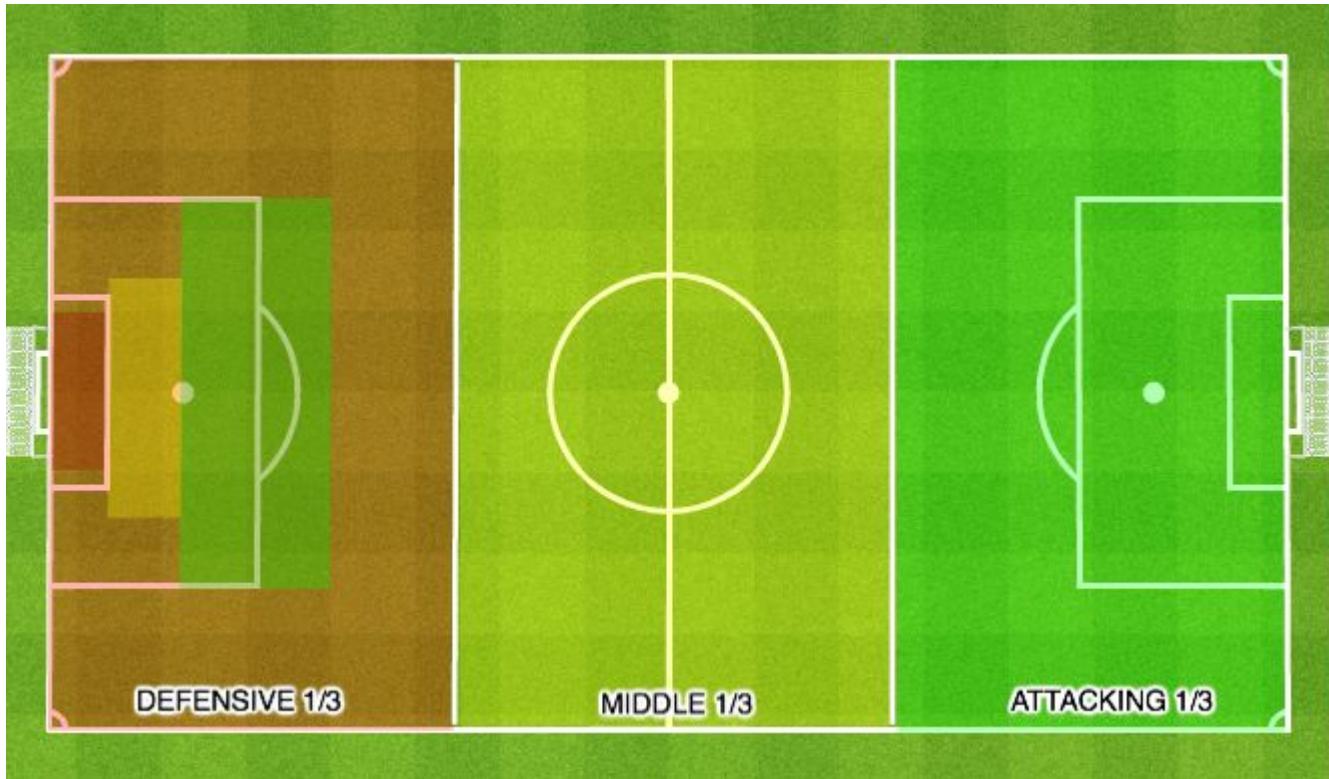
# World Map - Player Origin



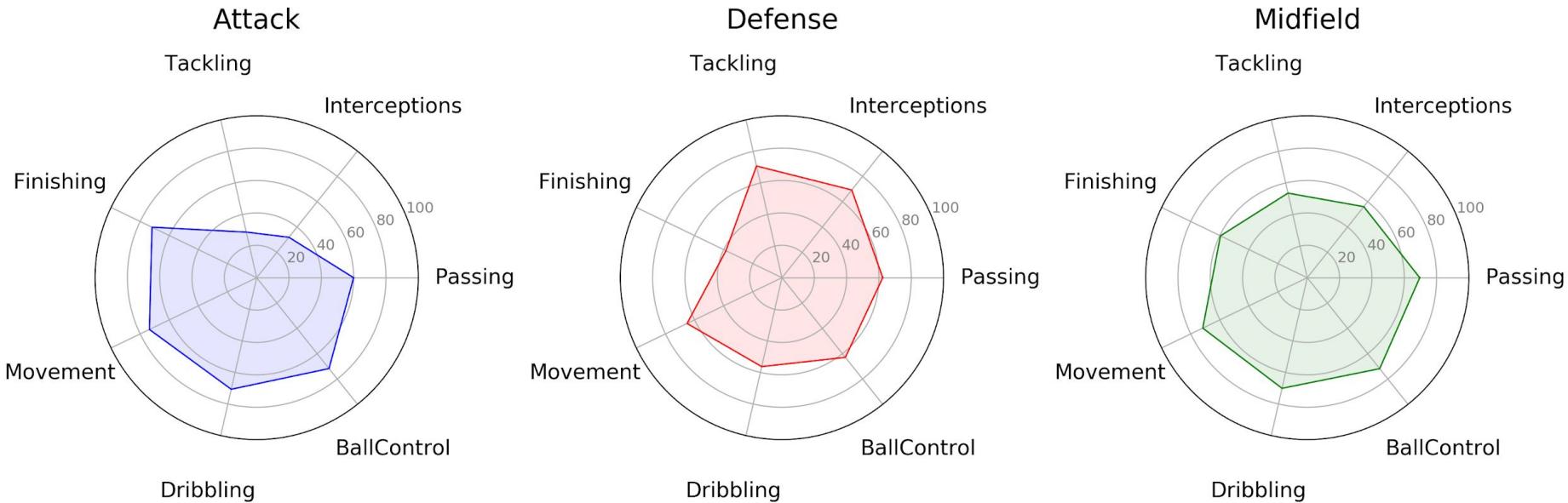
# Player Origin - Europe



# Positions



# Radar Plots



# Radar Plots

All Positions

Tackling

Interceptions

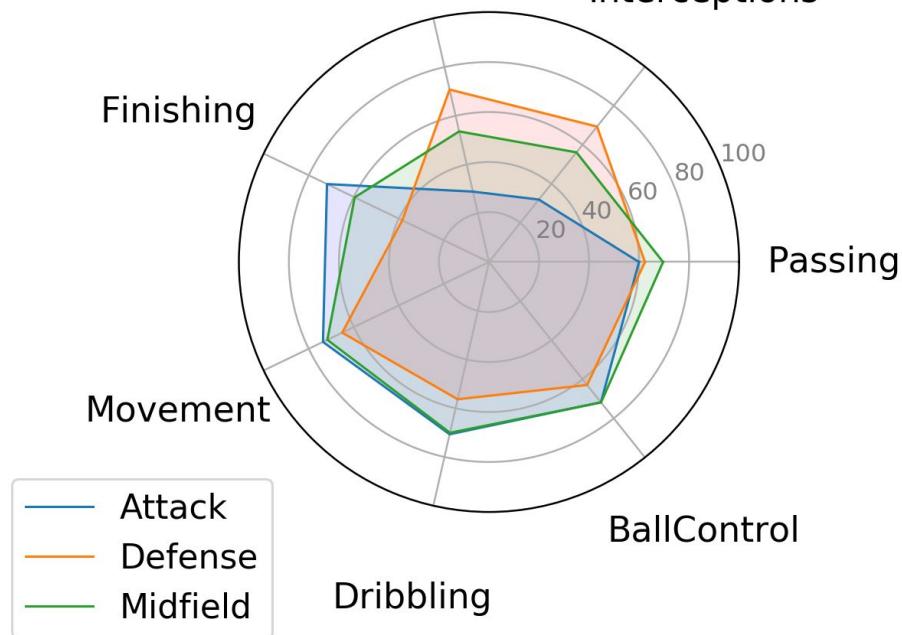
Finishing

Passing

Movement

Ball Control

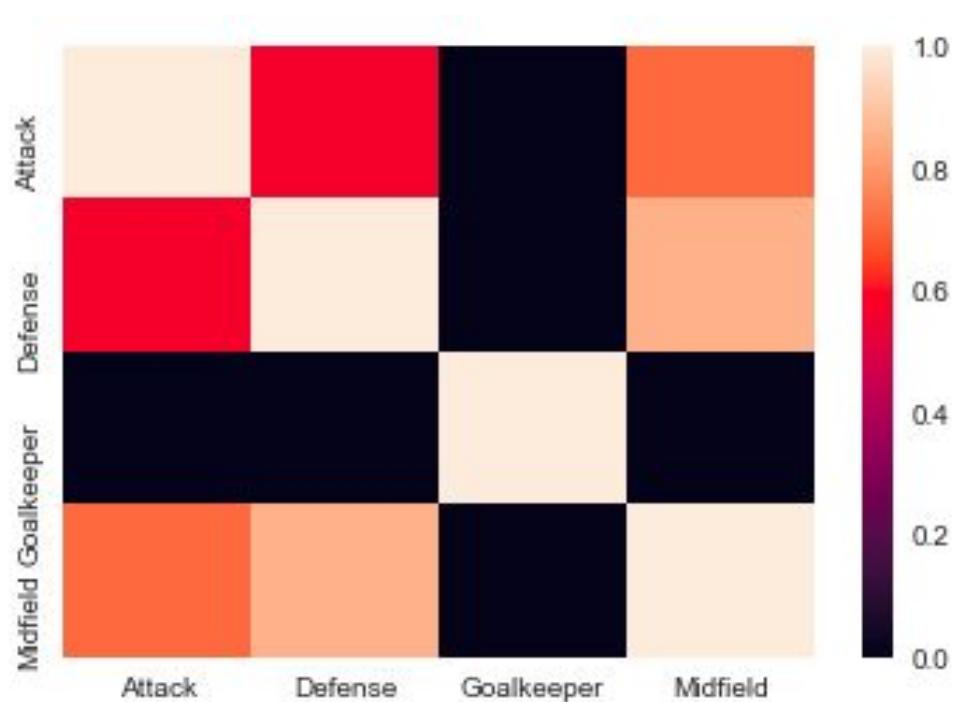
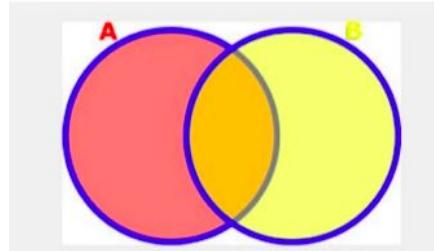
Dribbling



# Similarity Matrix & Heatmap

	Attack	Defense	Goalkeeper	Midfield
Attack	1.000000	0.571429	0.0	0.714286
Defense	0.571429	1.000000	0.0	0.857143
Goalkeeper	0.000000	0.000000	1.0	0.000000
Midfield	0.714286	0.857143	0.0	1.000000

Jaccard Similarity



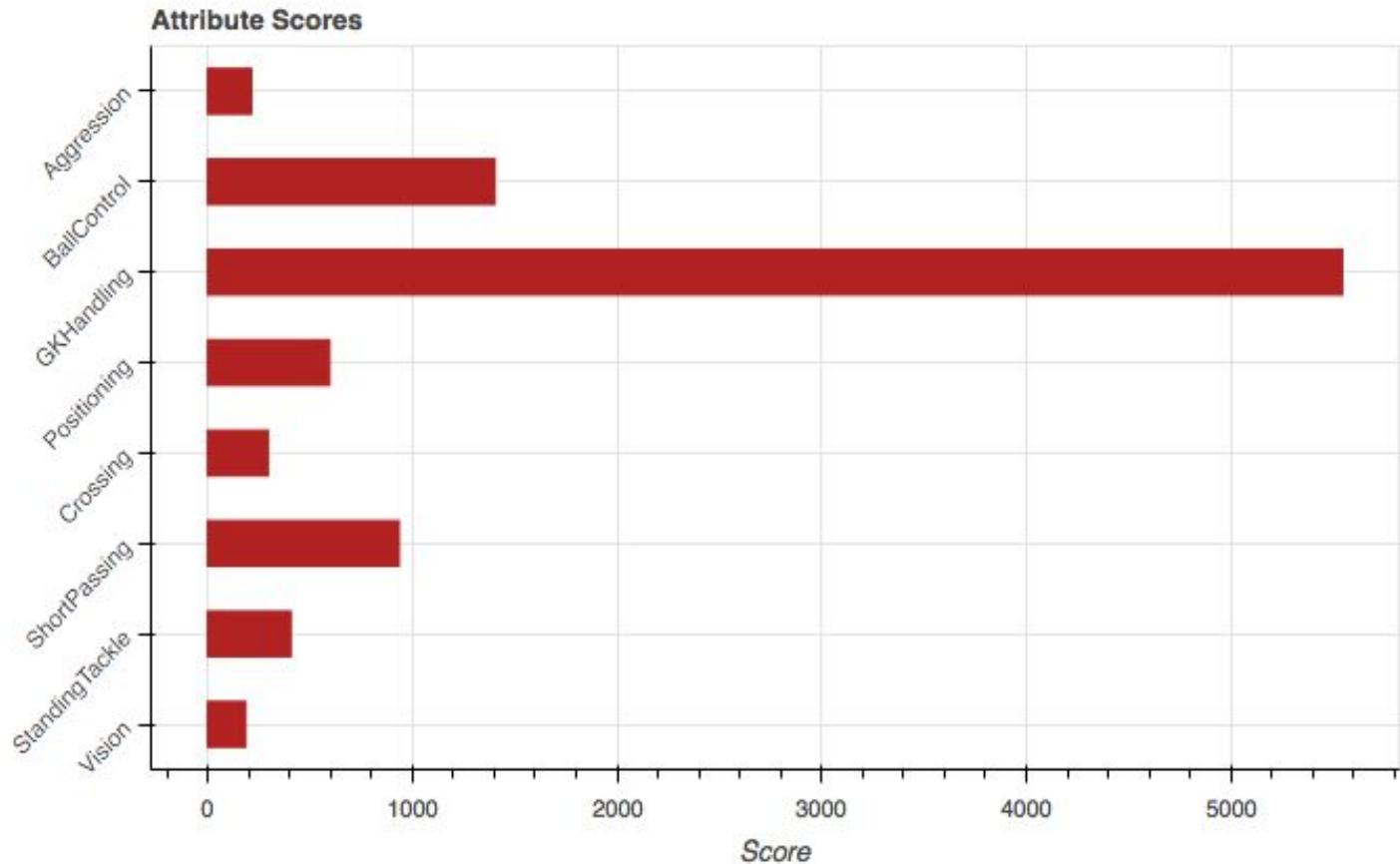
# Feature Selection

- ❑ Selected 8 features from 40 features to classify player into attacker, midfielder, defender or goalkeeper, using RFE or Recursive Feature Elimination

# Feature Selection

- ❑ RFE recursively removes features, builds a model using the remaining attributes and calculates the model accuracy. In this way, the function removes the most irrelevant feature in each step
- ❑ Also, selected 8 features based on our domain knowledge to analyse players in each position

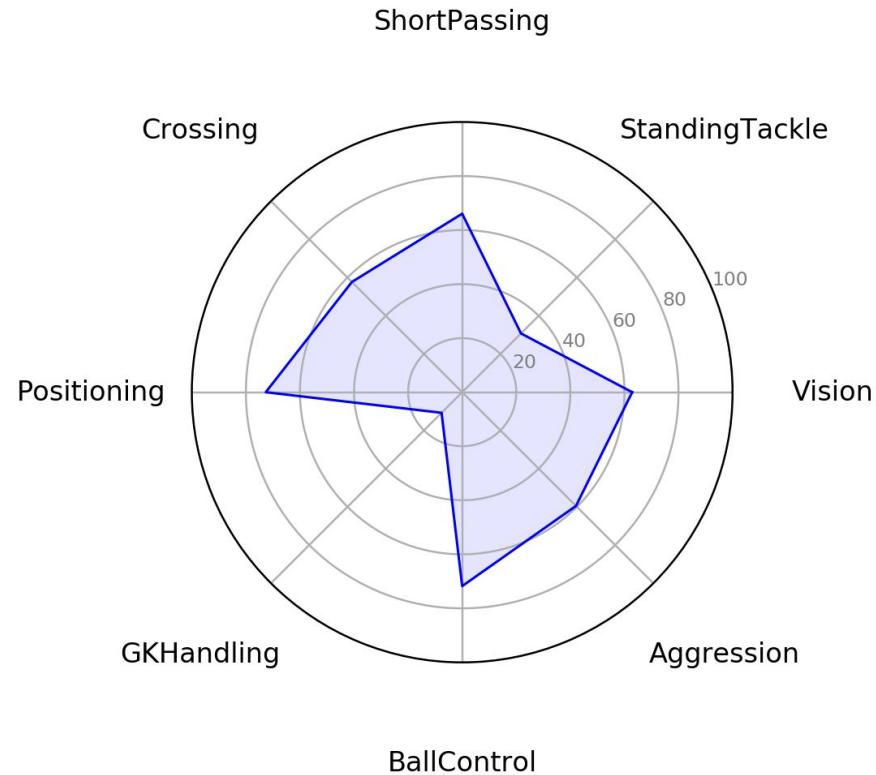
# Features vs Scores



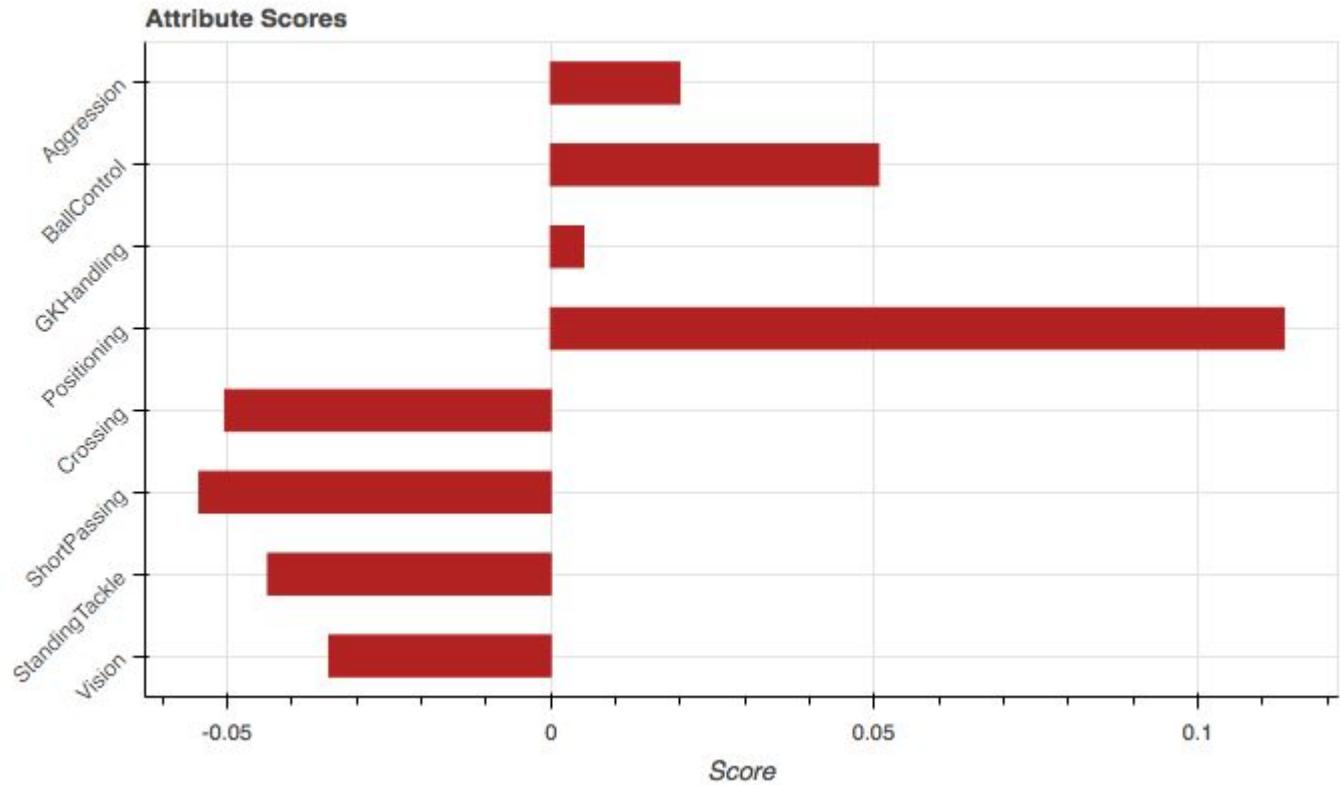
# SVM Model

- ❑ Fitted a supervised learning model, SVM to the SOFIFA dataset.
- ❑ Retrieved the coefficients of the model to find the feature importance for each position.

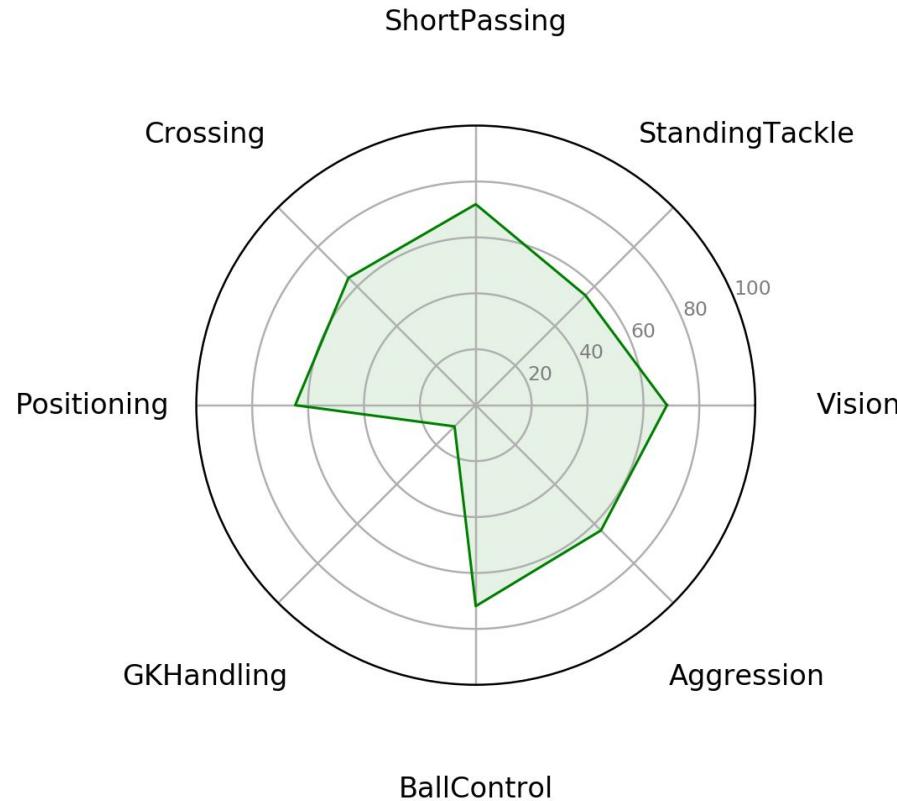
# Attacker



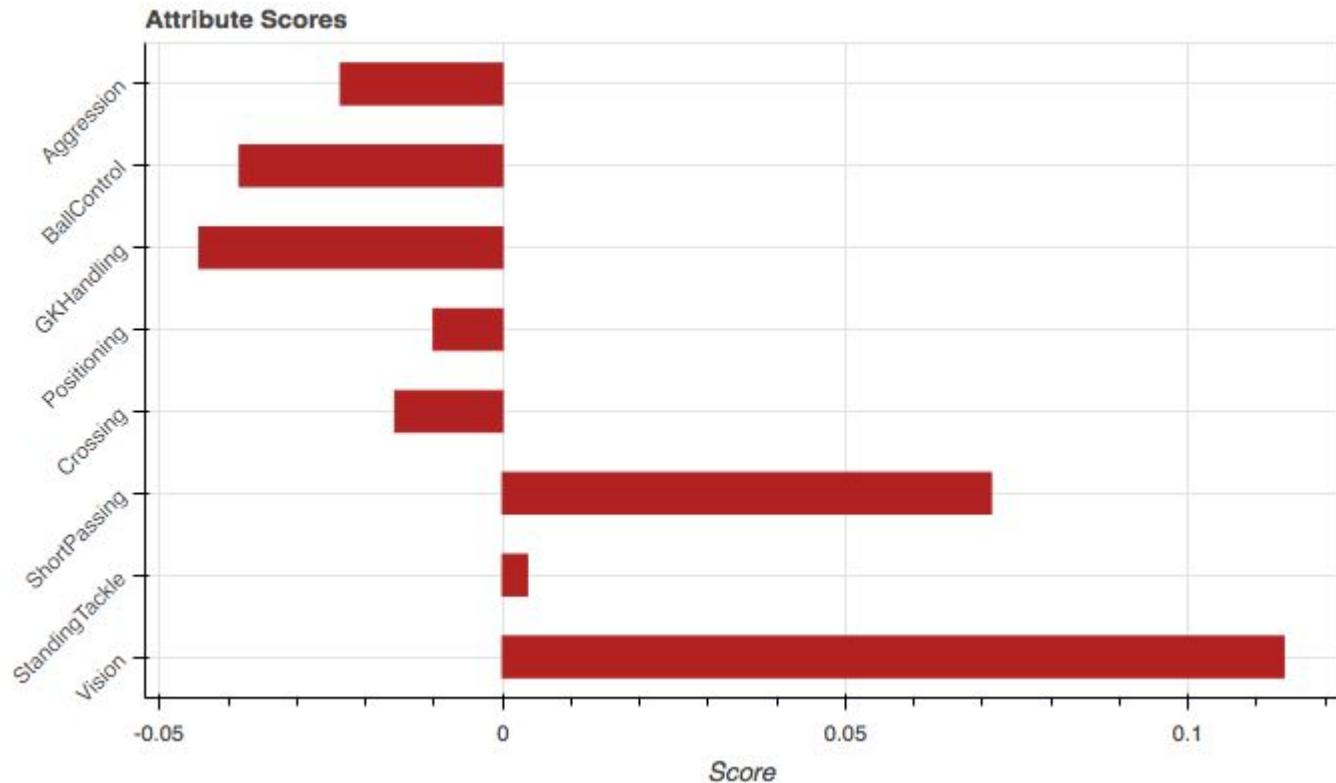
# Coefficients vs Feature



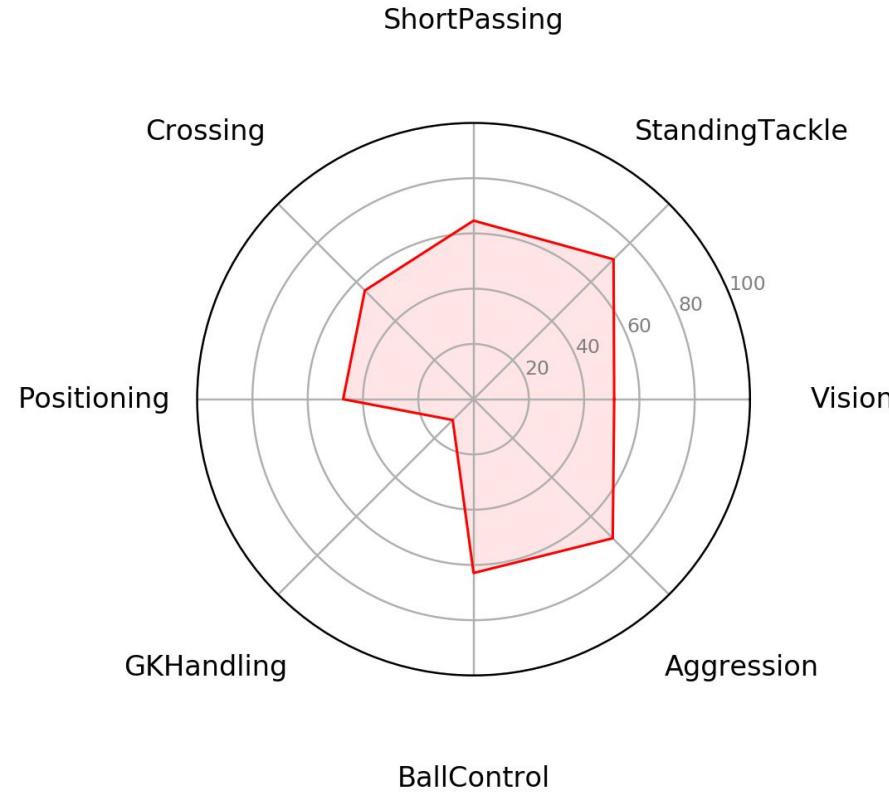
# Midfielder



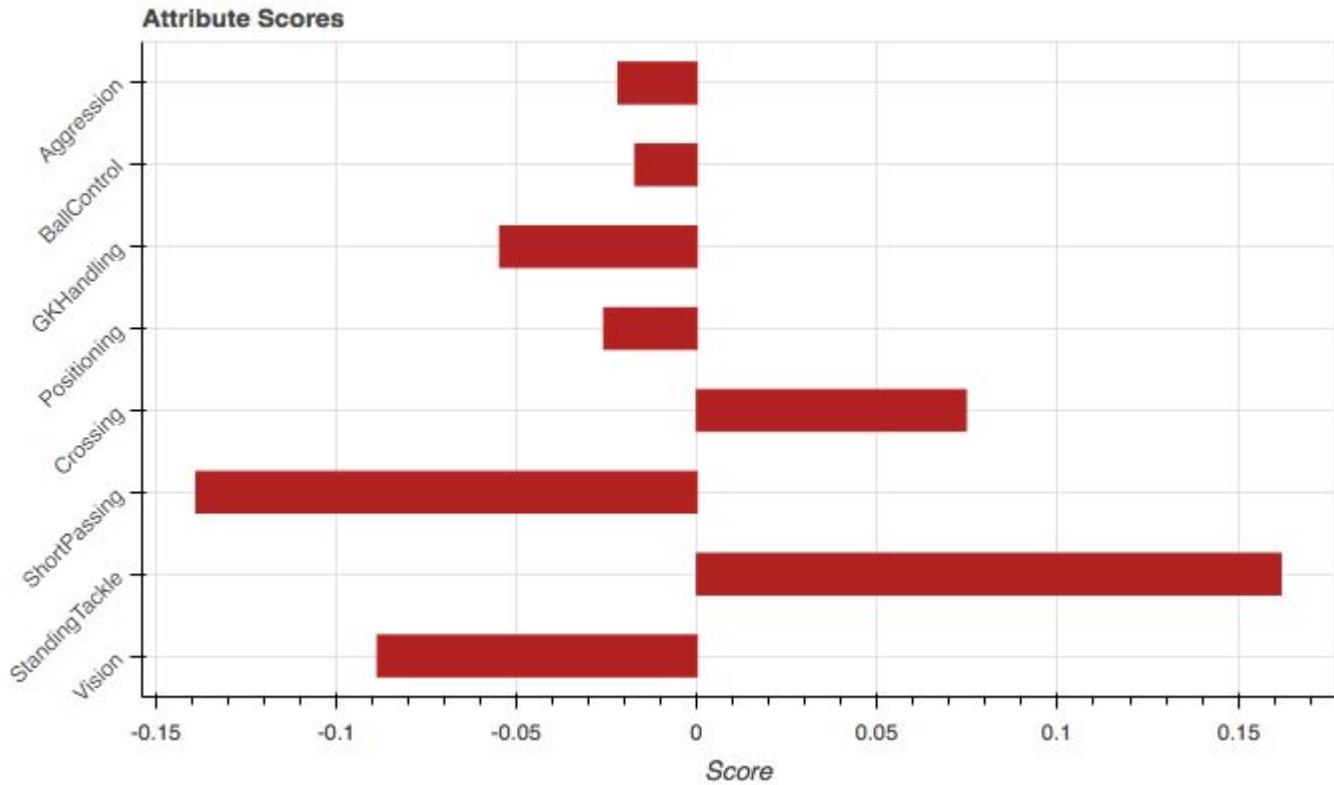
# Coefficients vs Feature



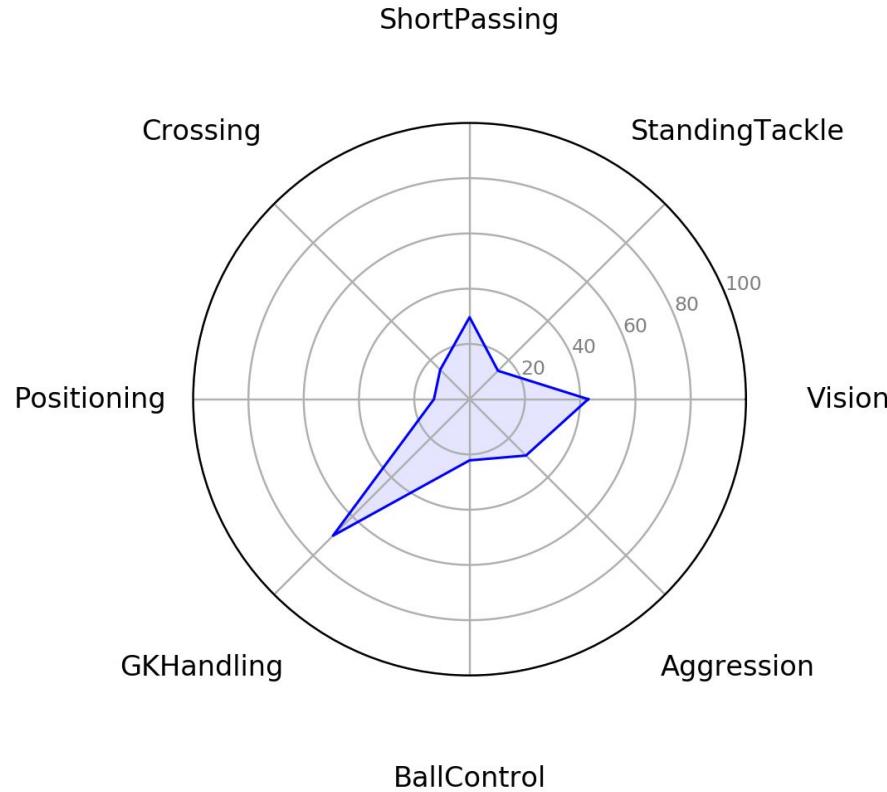
# Defender



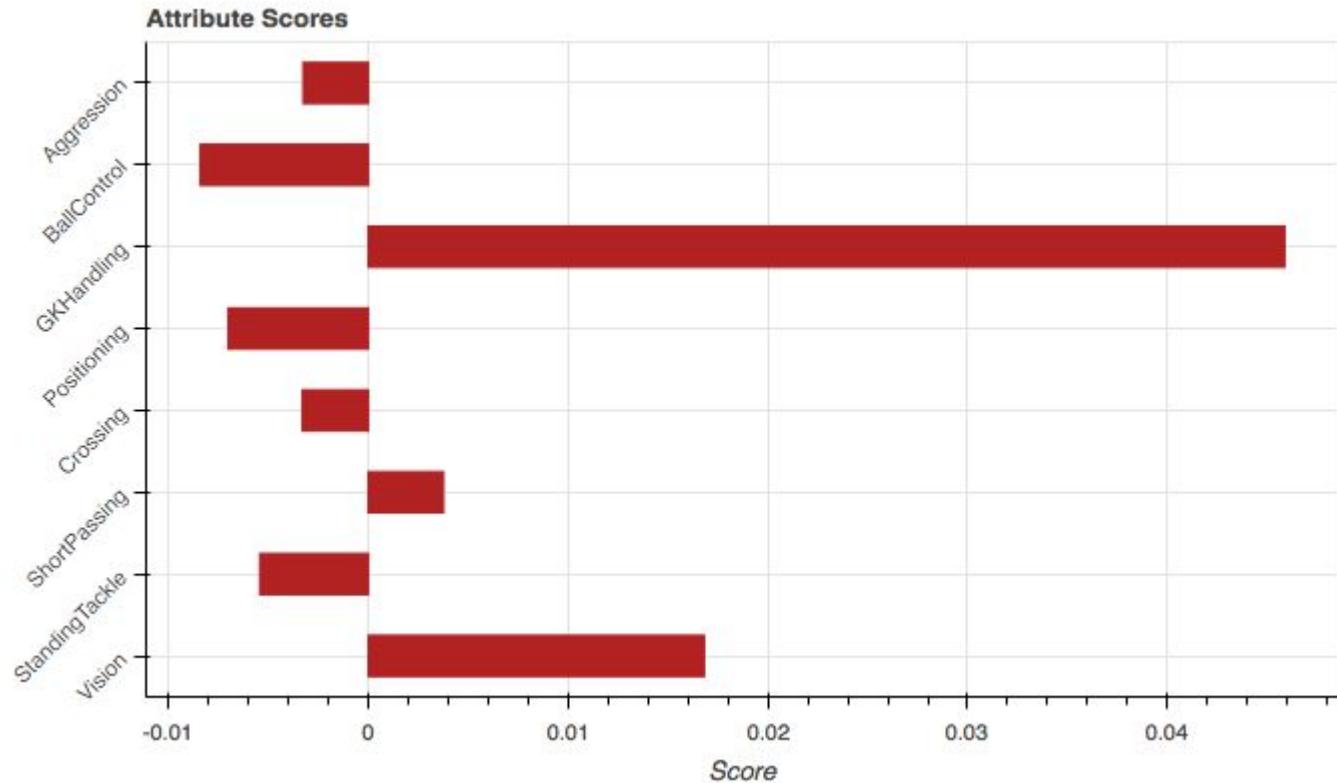
# Coefficients vs Feature



# Goalkeeper



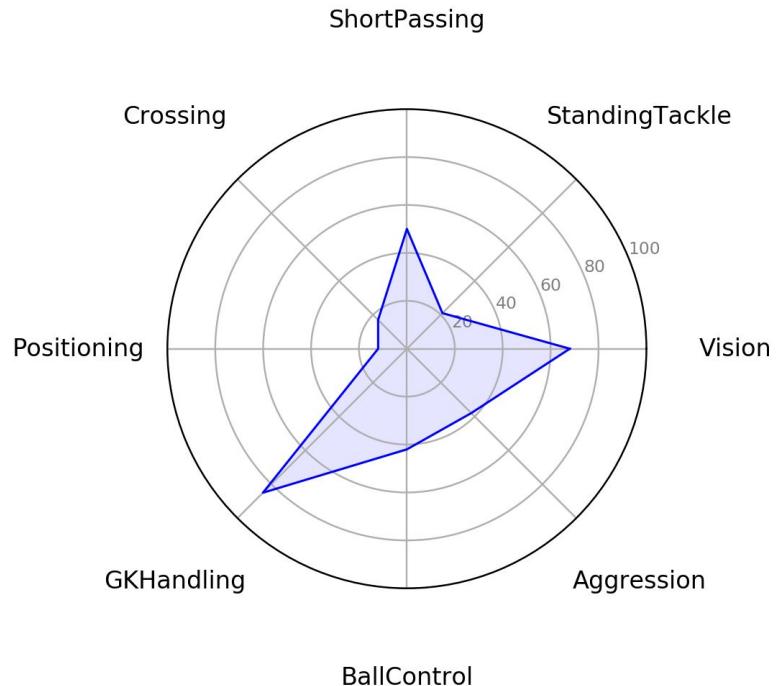
# Coefficients vs Feature



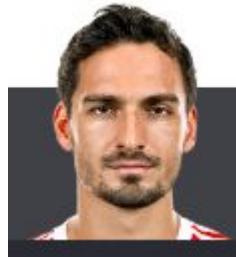
# Position recommendation



**De Gea**  
**Based on his skill,  
the model  
recommends that  
he should be a  
GOALKEEPER**

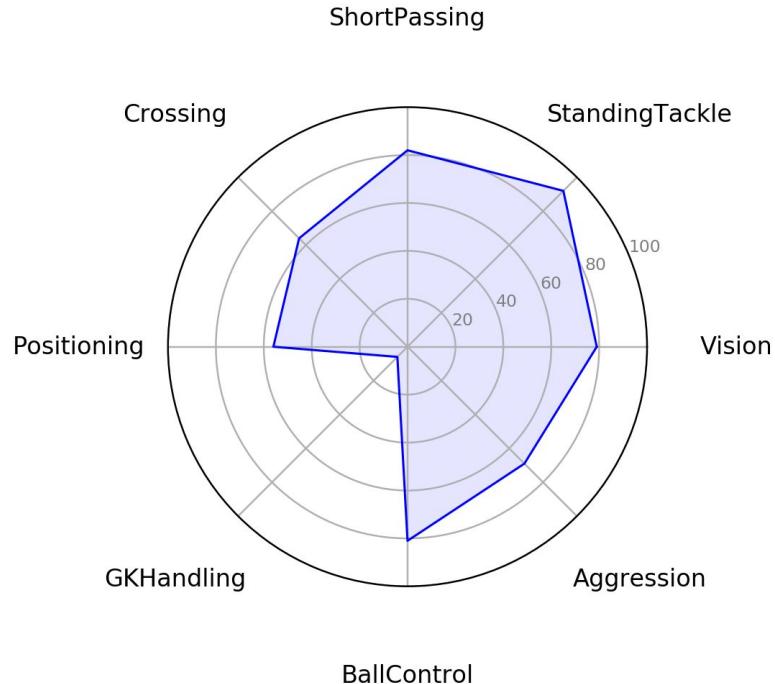


# Position recommendation



**M. Hummels**

**Based on his skill,  
the model  
recommends that  
he should be a  
DEFENDER**



# Conclusion

- ❑ We found out which attributes are important for each position
- ❑ Designed a model which recommends position of a player based on his skills
- ❑ Speed and Height are inversely correlated (unlike other sports)
- ❑ Younger players are valued more because of their longevity
- ❑ A geographical representation showed Europe to be the Mecca of soccer