



# Data 301 Project

By Brendan Callender



# Project Goal

- Profile European football players based on in-game statistics
- Determine role from actions rather than position
- Scouting Purposes
  - Can be used to find replacement players who can play similar role



# The Data

Data Acquired from



- Observational Units
  - Players from top 5 leagues for 2021-2022 Season
  - Non Goalkeepers
  - Minimum 1000 minutes Played
- Data From FbRef.com
  - Data includes shooting, passing, defense, possession, and shot creation variables



# Machine Learning Plan

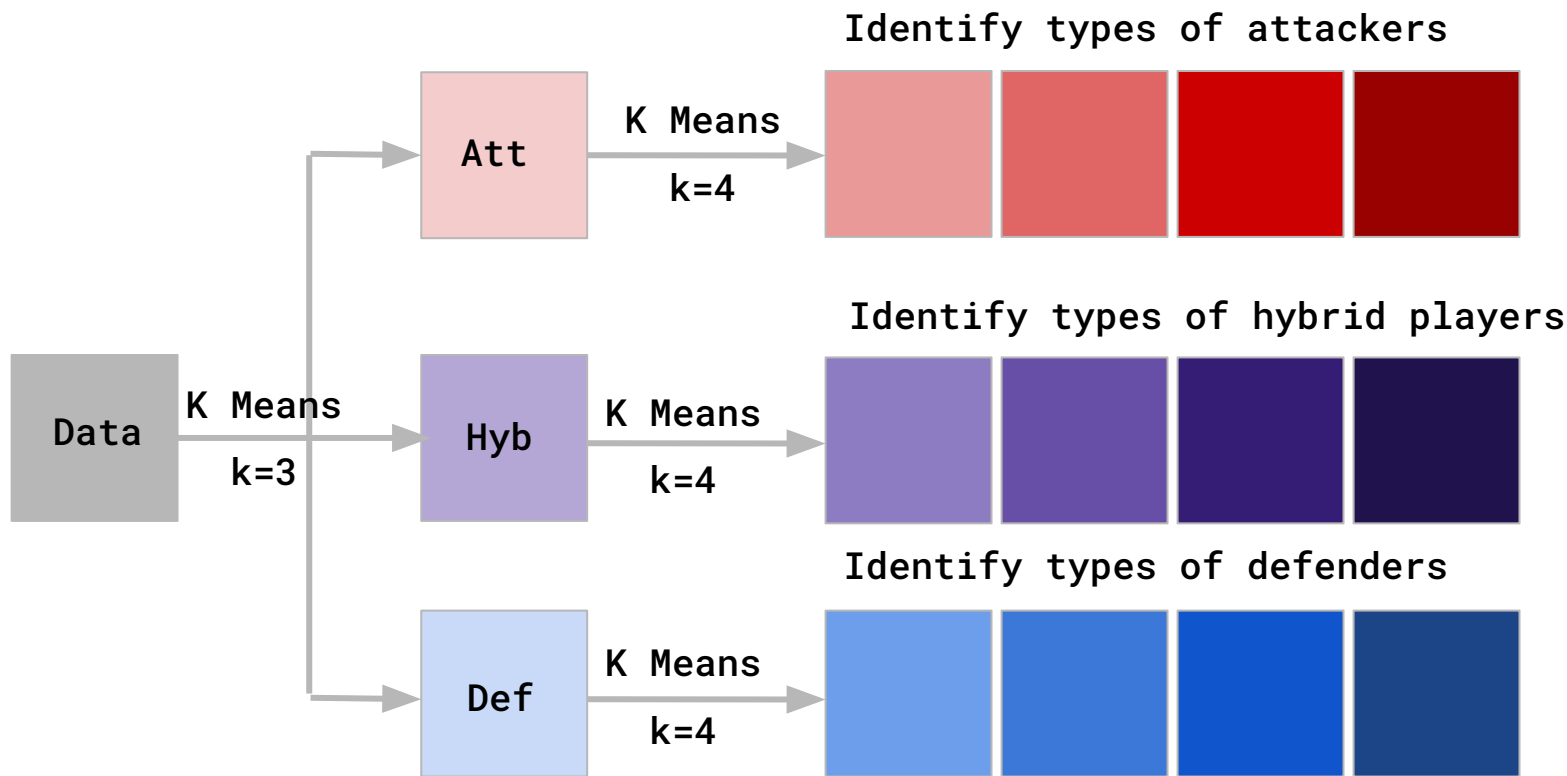
## **Hierarchical K Means Clustering Method:**

1. Fit initial K means cluster to separate players who primarily attack, defend, or do both
2. Fit separate, more specialized K means clusters on each group
  - a. Emphasize attack variables for attack group, defense for defense...
3. Examine distinct groups by comparing cluster means

## **Cosine Similarity Method: (Secondary Project)**

1. I also created a function to find most similar players to a given player by looking at Cosine Similarity

# Machine Learning Overview



All relevant variables were converted to per 90 minute values

# Model Variables

## Possession:

Def pen touches  
Def 3rd touches  
Mid 3rd touches  
Att 3rd touches  
Att pen touches  
Dribbles attempted  
Passes received  
Prog passes received

## Goal Creation:

Assists  
Expected Assists  
Passes leading to shots  
Dribbles leading to shots  
Shots from def action  
Shot creating actions  
Goal creating actions

## Defense:

Tackles  
Defensive 3rd Tackles  
Middle 3rd Tackles  
Attacking 3rd Tackles  
Interceptions  
Blocks  
Clearances  
Times Dribbled Past

## Passing:

Passes Attempted  
Short Passes  
Med Passes  
Long Passes  
Progressive Passes  
Passes into final 3rd  
Passes into att pen  
Crosses  
Completion %

## Shooting:

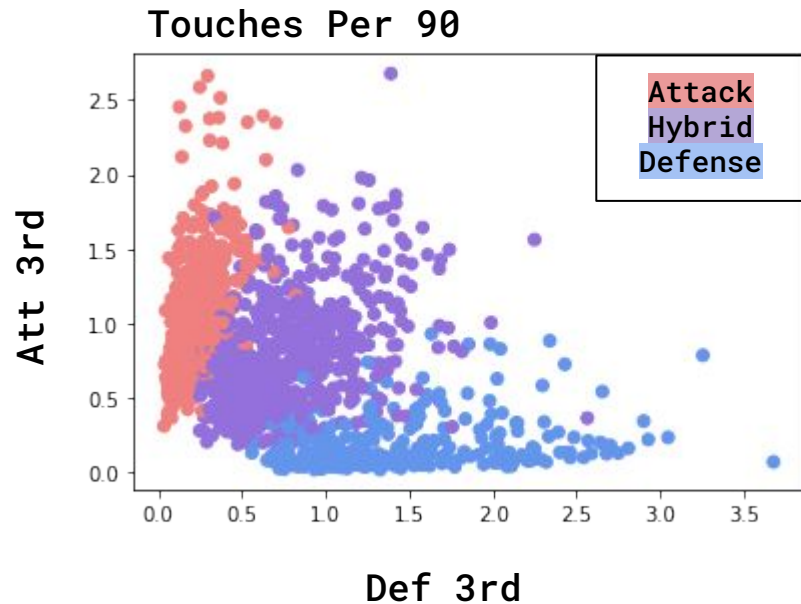
Shots  
Shots on Target  
Goals  
Expected Goals

## Key:

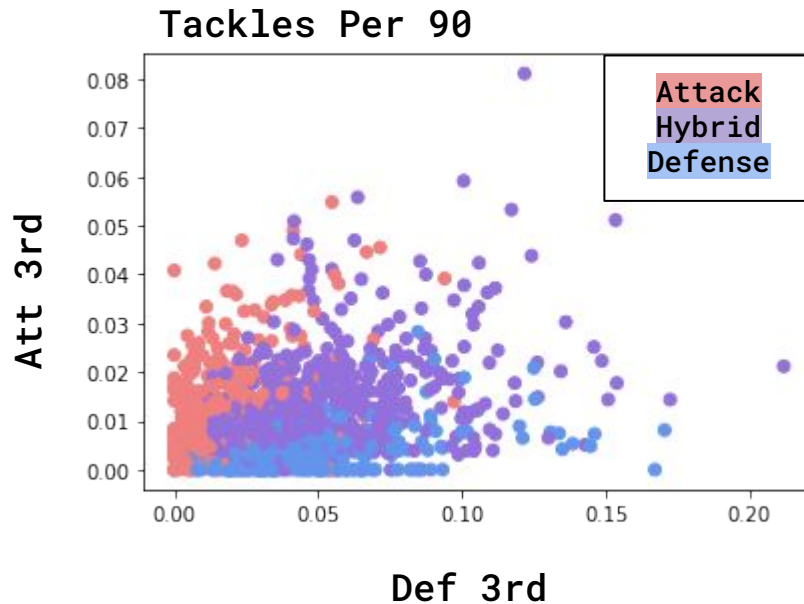
- Expect to be high for attacking players
- Expect to be high for defensive players
- Expect to be high for hybrid players

# First Cluster Results

Possession:



Defense:



## Second Cluster Results

# Attack

## Group 1: Pure Goalscorers

- Low pass volume
- High Goals/90, Low Assists/90
- Low dribbles attempted
- Low goal creation numbers

## Group 2: Forward Creators

- High pass volume
- Moderate Goals/90, High Assists/90
- High creation from dribbles and passing
- High Key Pass, Progressive Pass, Final 3rd Pass

## Group 3: Complete Forwards

- High Goals/90, High Assists/90
- Moderate pass volume
- High shots and High Penalty Box Touches
- High creation from dribbles
- Low defensive numbers

## Group 4: Counter Attackers

- Low Goals/90, Low Assists/90
- Moderate pass volume
- High defensive numbers
- High creation from dribbles
- Low shot volume

# Notable Attacking Players

## Forward Creators:

Kevin De Bruyne  
Raphinha  
**Lionel Messi**  
Neymar  
Ousmane Dembele  
Jamal Musiala



## Pure Goalscorers:

**Christian Pulisic**  
Pierre-Emerick Aubameyang  
Marcus Rashford  
Olivier Giroud  
Dusan Vlahovic



## Complete Forwards:

Gabriel Jesus  
Son Heung-min  
Cristiano Ronaldo  
Vinicius Junior  
Karim Benzema  
**Kylian Mbappe**



## Counter Attackers:

Leandro Trossard  
Emile Smith Rowe  
Antoine Griezmann  
Gabriel Martinelli  
Jesper Lindstrom  
**Timothy Weah**

# Defense

## Group 1: Front Foot Defenders

- Low pass volume
- High cross volume
- High tackle volume
- High defensive numbers in Att 3rd
- High dribble volume

## Group 2: Defensive Walls

- Moderate pass volume
- Low dribbled past
- High defensive penalty touches
- Moderate clearances and blocks
- High long pass volume

## Group 3: No-nonsense Defenders

- Low pass volume
- High blocks, clears, and interceptions
- Low dribble volume
- Low passes received

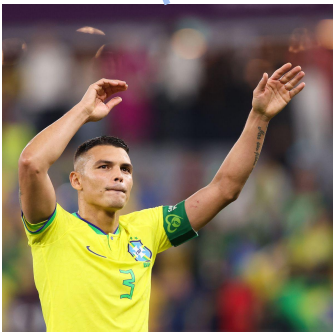
## Group 4: Ball Playing Defenders

- High pass volume (short, med, and long)
- High passes volume into final 3rd
- Moderate tackle volume
- High volume of touches
- High dribbles

# Notable Defensive Players

## Ball Playing Defenders:

Aymeric Laporte  
Ruben Dias  
John Stones  
Josko Gvardiol  
Nico Schlotterbeck  
**Thiago Silva**



## Front Foot Defenders:

Christian Romero  
Lukas Klostermann  
**Ricardo Rodriguez**  
Anel Ahmedhodzic

## Defensive Walls:

**Virgil van Dijk**  
Raphael Varane  
Eder Militao  
Matthijs de Ligt  
Kalidou Koulibaly



## No-Nonsense Defenders:

Fabian Schar  
**Maya Yoshida**  
Kurt Zouma  
Armel Bella Kotchap  
Victor Lindelof



# Hybrid

## Group 1: Defensive Progressors

- Low Goals/90 and low Assists/90
- High pass volume
- High Tackle volume
- High Middle 3rd touches
- Low creativity numbers

## Group 2: Box to Box Creators

- High Goals/90 and High Assists/90
- High pass volume
- High key pass and pass into penalty area
- Low defensive numbers
- High goal creation from passes

## Group 3: Defensive Containers

- Low Goals/90 Assists/90
- Low pass volume
- High tackle volume
- High clearances and interceptions
- Low creativity numbers

## Group 4: Box to Box Goal Threats

- High Goals/90, moderate Assists/90
- Low pass volume
- Moderate defensive numbers
- High att 3rd and att penalty touches
- High dribble volume

# Notable Hybrid Players

## Box to Box Goal Threats:

Gavi

**Weston McKennie**

Sergino Dest

Jude Bellingham

Sandro Tonali



## Defensive Progressors:

Casemiro

**Tyler Adams**

Sergio Busquets

Benjamin Pavard

Declan Rice

Rodri



## Defensive Containers:

**Takehiro Tomiyasu**

Scott McTominay

Kieran Tierney

Wilfred Ndidi

## Box to Box Creators:

Trent Alexander-Arnold

Luka Modric

Joshua Kimmich

Bruno Fernandes

Marco Verratti

**Achraf Hakimi**

Alphonso Davies



Thank You  
Any Questions?

