

# KickOff

A visual analytic system  
for analyzing football  
leagues in a given season

Visual Analytics

Master Of Science in Engineering in CS



# The team

Let's introduce ourselves...



*@sergio  
\_picca*

Worked on the player view and on the part of the data dealing the players (merging phase).



*@andrea\_  
napoletani*

Worked on the team view, collecting relevant information from the data about matches.

# Outline of the talk



## 1. Motivation

Explaining the reasons behind the chosen topic.



## 2. Related Work

Show some other application of football analytics.



## 3. Methodology

Road to the final system: the dataset and the views.



## 4. Result, conclusions

Pros and cons at the end of the journey.

# Motivation

## Soccer Analytics

Nowadays soccer analysis is acquiring interest and it became a **real business**.

Teams **pay money** to get data **to know their opponents** and Pay tv tell you everything about a single player.



# Related works

## Other football analytic systems

Even if we did not take inspiration from a paper or some experiment, we refer here to systems that provide similar analysis about football statistics.

### Pappalardo et al.



Described the “largest open collection of soccer-logs”[1], containing the spatio-temporal events.

### Wyscout



“Wyscout provides a toolset specifically designed for football professionals.”[2]  
(From the website)

# Methodology

## *Dataset and ...*

Our dataset is made up by the statistics of the major European leagues in the season 2018/19.

## *... its main structure*

Our dataset was obtained by merging the Pappalardo's dataset with the FIFA19's one. We have info about:



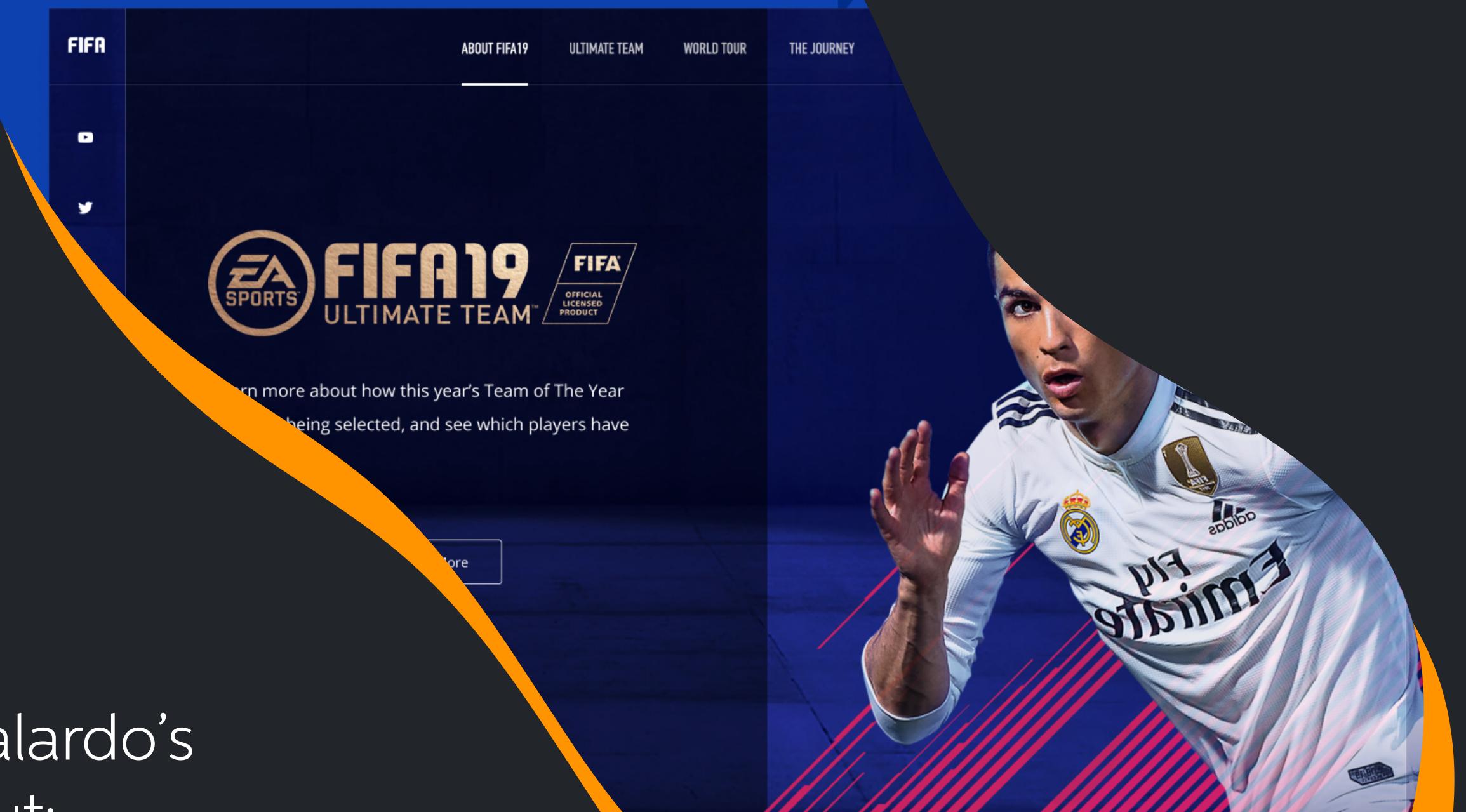
INFO ABOUT  
MATCHES



INFO ABOUT  
TEAMS



INFO ABOUT  
PLAYERS



# Tools & pipeline



## Python

A (**quite long**) preprocessing step was necessary.



## Nodejs

For the **backend**, use of ejs for better readability of the code.



## D3js

All the **frontend** part.

We came up with **3440 players**, with **55 attributes** and ...

... a .csv file for each major league, containing more or less **20 teams** each with **81 attributes**.

AS index = 197 300 (we have 5 major leagues)

# The system

Two main views



## *Players view*

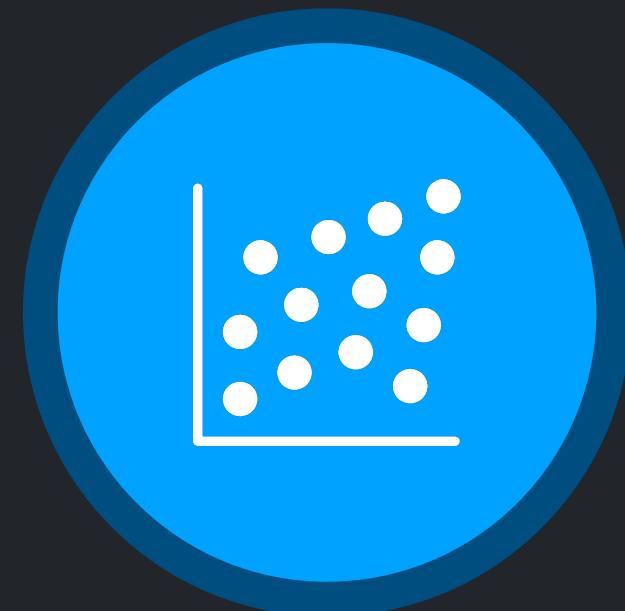
Focusing on the skills of all players in an European team. We go from a **general overview to particular comparisons** among players.



## *Teams view*

We investigate the performance of teams in the season 2018/19. In particular we look for **similarity among teams**.

# Players view



## Scatter plot

Display players according to **minutes played** and received scores.



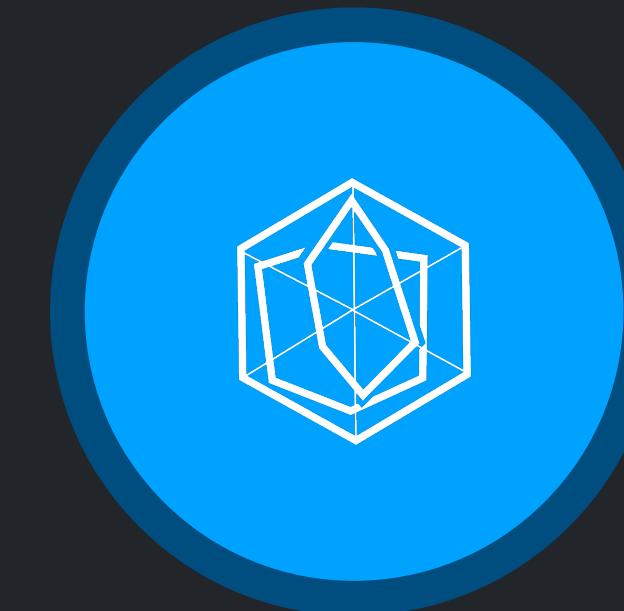
## Parallel coord.

Display players' **skills**, with the possibility of choosing the set of skills.



## Barchart

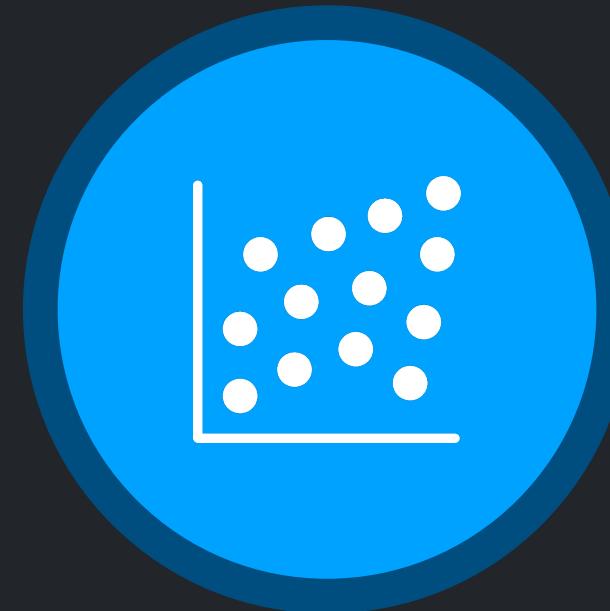
Display the **overall value** of selected players, from the above views.



## Radar chart

Detailed analysis of **select players'** skills in order to make comparisons.

# Teams view



## Scatter plot

Display of teams cluster (application of **K-means**).



## Matrix

Display the teams in the league, looking for **similarities**.



## Bar chart

Display the **team trend** during the season (both rounds).



## Reduction

We applied **reduction algorithms** (PCA,MDS, t-SNE).

# Lesson learned



## *What we learned?*

- ✓ The importance of **colors**.
- ✓ Usage of distortion.
- ✓ Dealing with **large dataset**.
- ✓ Importance of **dimensionality reduction**.



## *Issues along the way!*

- ✗ Finding the **right dataset**.
- ✗ Dataset merge phase.
- ✗ Defining the **right interactions**.
- ✗ D3 versions compatibility

# Conclusion & future work

What next?

## A toy example

We may do a real system  
acquiring more data, **but is not  
easy.**

## More than just one season

Extend our system to **more than  
one season.**

## Do interviews

Listen to domain expert in order  
to know what kind of interaction  
they really need.



That's it.

# Thank you.



**GitRepository**

 [Repository](#)



**Documentation**

 [Web-doc](#)



**Working system**

 [Azure-web-site](#)

# References

1. Luca Pappalardo, Paolo Cintia, Alessio Rossi, Emanuele Massucco, Paolo Ferragina, Dino Pedreschi, Fosca Giannotti: **A public data set of spatio-temporal match events in soccer competitions**, Scientific data, (2019) 6:236 | <https://doi.org/10.1038/s41597-019-0247-7>.
2. **Soccer match event dataset**, <https://wyscout.com> .
3. **FIFA19 dataset**, <https://www.kaggle.com/karangadiya/fifa19/data#> .
4. **Wyscout**, <https://wyscout.com> .
5. **Nodejs**, <https://nodejs.org/it/> .
6. **D3js**, <https://d3js.org> .