



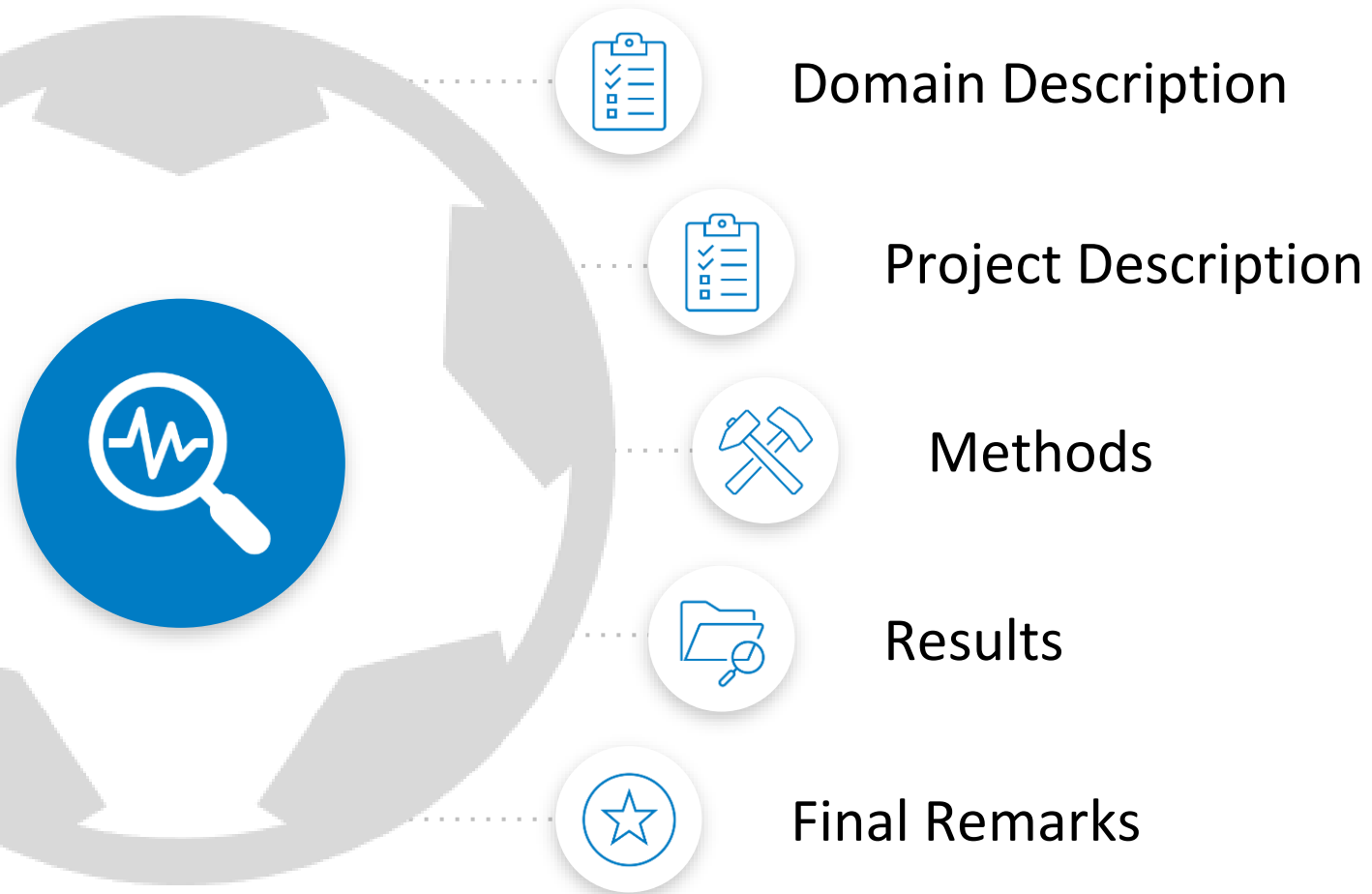
UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH



Analysis of football data

Process-oriented Data Science 23/24

Agenda



Domain Description

Key requirements for our project work



Data includes Games, Players & Team-specific data



Getting insights for coaches by Analyzing play patterns



Focus on Process Discovery rather than performance



Project Description

Our approach to analyzing the World Cup data taking Process Mining into account



Use general **statistics** from **Data Analytics** as a reference for **Process Mining**



Comparison of play for the 2 finalists **France** and **Croatia**



Analysis of the MVP Luka Modric compared to **2nd & 3rd best player**



Comparison of 2 completely different games involving the winner **France**

Methods

Overview of data and pre-processing including our selection of tools

Raw Data



Actions

Action
Team

Type

Possession Team

Result

Coordinates

Player

...

Pre-Processing



Adjust data format

Add data (e.g. zones)

Merge files

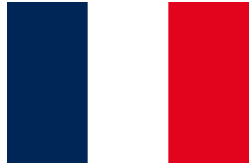
Remove unnecessary data

Analysis



Results

Comparison of two finalists



→ Actions:

- ◆ Attacks based more on **individual dribblings**
- ◆ Intensive **pressing** after unsuccessful passes or dribbles, often leading to **recovery**
- ◆ Shots after **deeper build-up** (low number of shots after first initial passes)

→ Players:

- ◆ **Not clear playmaking leader** - most active players Kante, Pogba, Pavard, Varane have similar amount of passes. No player has as thick edges as Modric for Croatia.

→ Pitch Zones:

- ◆ **Short passes** from defensive zones
- ◆ Most of the **possession losses** happen in **opponent's half**



→ Actions:

- ◆ Attacks based on **passes**, less dribblings
- ◆ More **fouls after miscontrol**
- ◆ More shots from **set pieces** and accidental (**w/o build up**)

→ Players:

- ◆ Team strongly dependent on **Luka Modric**, which received and performed most passes, is among players that start and end the ball possession sequences most often

→ Pitch zones:

- ◆ **Long passes** to transport the ball directly from defence to attack
- ◆ **Dispossessions in more central zone** of the field than France

Results

Analysis of MVP Luka Modric

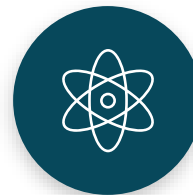
	Modric	Hazard	Griezmann
Pass Acc.	84%	83%	75%
Assists	1	2	2
Shots	10	18	21
On Target	5	13	15
Goals	2	3	3
Fouls	13	7	6



Modric is far **more important** for **opportunity creations** of his team



Outstanding success and **failure stats** when under **pressure**



Modric is more **important** for his **team** as the **core** of their **play**

Results

Comparison of 2 different games with the France team involved: FRA-ARG vs. FRA-DEN



→ Actions:

- ◆ more **Pressure**
- ◆ ending with **Fouls** - more intense game
- ◆ 20s less case duration - higher pace in game

→ Players:

- ◆ **Left-side** players more involved
- ◆ striker **Olivier Giroud** had more interactions with others
- ◆ **goalkeeper** highly involved in opportunity creation

→ Pitch Zones:

- ◆ France was playing more through **left-side** of the pitch
- ◆ 2nd half required more **defense** work

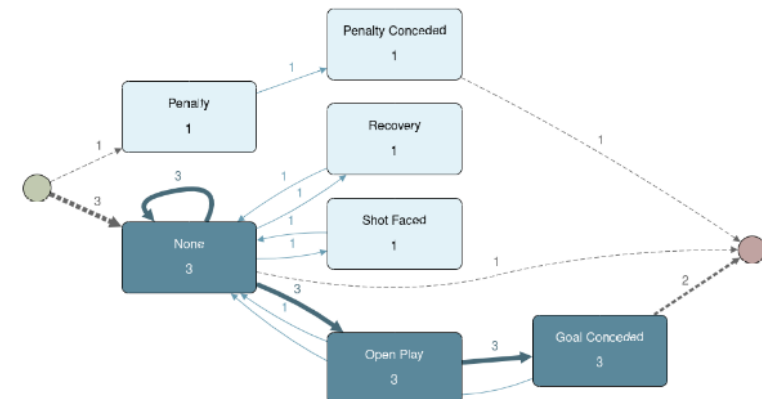


→ Actions:

- ◆ longer pass network
- ◆ **Clearance** is more dominant
- ◆ higher ball possession of France

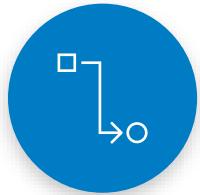
→ Players:

- ◆ **Kylian Mbappé** only substitute until 78' minute
- ◆ more player interaction in the defense



Closing Remarks

Reflections on the projects and tools used



In case **data** is **connected** to some **process**, data should also be **viewed** in a **Process Mining context** and **not only statistically** with Data Analytics.



Playing with **different perspectives** is interesting to obtain **new results** and get to **new findings**.



With **today's technology standards**, extended **functionality** to perform **Data Analytics** within **Process Mining tools** is **desired** or expected.

Thank you!

Any Questions?



Image Sources

- FIFA World Cup 2018: https://th.bing.com/th/id/OIP.KUpyt_FLyUX8pNUUoHgQkQHaHa?rs=1&pid=ImgDetMain
- Python: https://1.bp.blogspot.com/-ugMNNEeWN0g/X9WezhHRipI/AAAAAAAAJA4/LBtozO1knloe-kg8jBR_s4_dYaa2xncrQCLcBGAsYHQ/s982/pngfind.com-python-logo-png-626208.png
- CSV: <https://openscience.vn/images/icons/csv-13-512.png>
- Apromore: <https://www.leonardo.com.au/hs-fs/hubfs/ApromoreLogo.png?width=1626&height=328&name=ApromoreLogo.png>
- Disco: https://do3z7e6uuakno.cloudfront.net/uploads/event/logo/1004960/Disco_400.png
- Luka Modric: <https://th.bing.com/th?id=OSB.AAyHQxB.png&w=138&h=104&o=6&dpr=1.2&pid=SANGAM>
- Eden Hazard: <https://www.futcards.com.br/wp-content/uploads/2018/10/eden-hazard-chelsea.jpg>
- Antoine Griezmann: <https://th.bing.com/th?id=OSB.AAyB4Wc.png&w=138&h=104&o=6&dpr=1.2&pid=SANGAM>