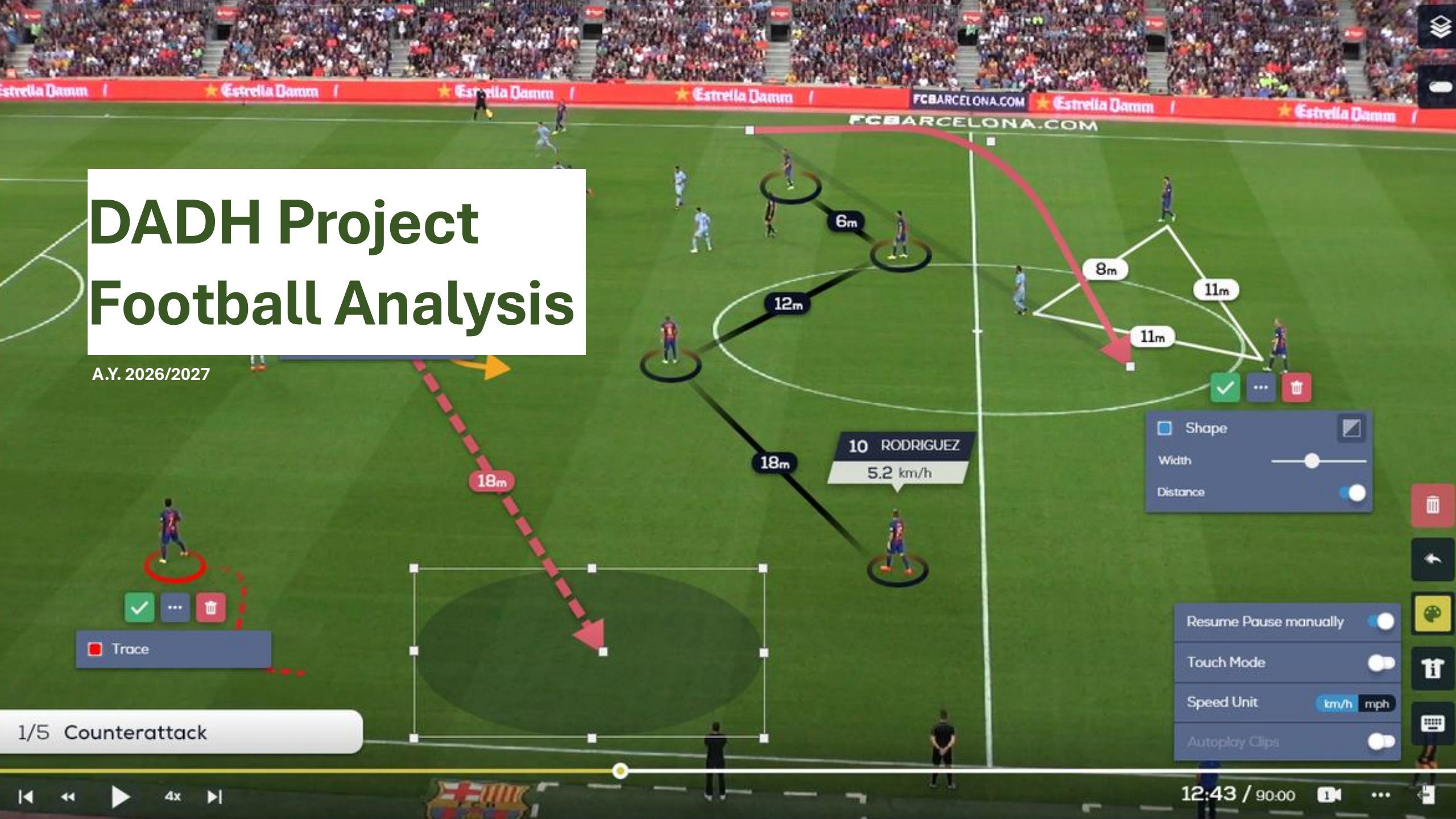


DADH Project Football Analysis

A.Y. 2026/2027



The data

The repository is composed of different kinds of data that describe the matches as well as the players.

Competitions

Available
competitions and
seasons

Lineups

Players and their
roles in each match

360

Spatial context for
selected events

Matches

Matches, teams,
scores, and match
information

Events

Actions performed
by players during
each match

Physical Info

About the players
About their runnings

Competitions.json

The first file to consider is ***competitions.json***.

It provides information about the competitions and seasons available in the dataset, together with their ***identifiers***. These identifiers are important for navigating the repository and linking the information available across the different data sources.

Competitions: Liga, Champions Leagues etc...

```
, {  
  "competition_id" : 16,  
  "season_id" : 2,  
  "country_name" : "Europe",  
  "competition_name" : "Champions League",  
  "competition_gender" : "male",  
  "competition_youth" : false,  
  "competition_international" : false,  
  "season_name" : "2016/2017",  
  "match_updated" : "2024-02-13T02:37:32.205154",  
  "match_updated_360" : "2021-06-13T16:17:31.694",  
  "match_available_360" : null,  
  "match_available" : "2024-02-13T02:37:32.205154"  
}, {
```

Matches.json

The competition and season identifiers allow you to select the corresponding JSON file in the matches folder.

- Each JSON file refers to one specific competition and season and may contain multiple matches.
- For each match, you can find information such as the **match identifier**, date, teams, score, competition stage, stadium, referee, managers, and other match-related information.

```
"away_team" : {  
  "managers" : [ {  
    "name" : "France"  
  }  
] },  
"home_score" : 0,  
"away_score" : 2,  
"match_status" : "available",  
"match_status_360" : "scheduled",  
"last_updated" : "2025-12-16T17:04:24.176407",  
"last_updated_360" : "2021-06-13T16:17:31.694",  
"metadata" : {  
  "data_version" : "1.1.0",  
  "shot_fidelity_version" : "2",  
  "xy_fidelity_version" : "2"
```

Lineups.json

Using the **match_id**, you can access the JSON file in the **lineups** folder. Each lineup file refers to a single match and contains information about the two teams and the players involved in the match.

- For each player, you can find information such as the player_id, name, jersey number, country, playing position(s), and cards received.
- The position information also indicates when a player occupied a specific role during the match, allowing you to identify position changes and substitutions.

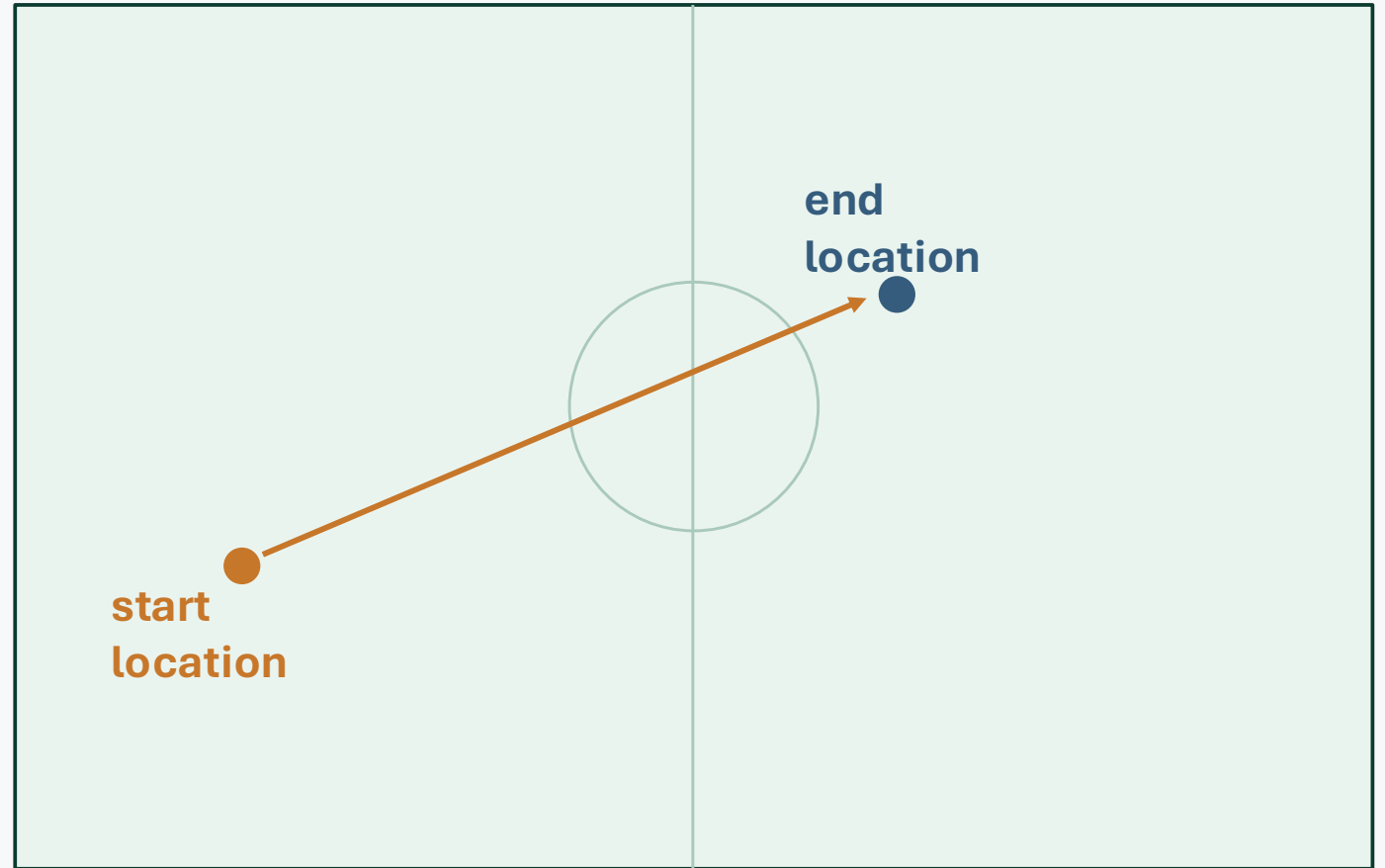
```
"lineup" : [ {  
  }, {  
    "name" : "Norway"  
  },  
  "cards" : [ ],  
  "positions" : [ {  
    "position_id" : 24,  
    "position" : "Left Center Forward",  
    "from" : "82:37",  
    "to" : "83:45",  
    "from_period" : 2,  
    "to_period" : 2,  
    "start_reason" : "Substitution - On (Tactical)",  
    "end_reason" : "Tactical Shift"  
  }, {  
    "position_id" : 13,  
    "position" : "Right Center Midfield",
```

Events.json

Using the **match_id**, you can access the JSON file in the events folder.

Each event file refers to a single match and contains the **sequence of actions** that occurred during the game.

- For each event, you can find information such as the event type, player, team, time, possession, position, and location on the pitch.
- Depending on the event type, additional information is available. For example, passes include information about length, direction, recipient, and outcome, while shots include information about location, expected goals (xG), technique, and outcome.
- Events may also contain contextual information, such as whether the player was under pressure when performing the action.



Events.json

There are many event types, depending on the action performed, such as passes, shots, dribbles, carries, pressures, and duels.

Each event type may contain different information describing the specific action, such as its location, outcome, duration, direction, or other event-specific characteristics.

Examples:

Pass: length, angle, recipient, end location, outcome, cross, assist

Shot: xG, technique, body part, end location, outcome

```
"lineup" : [ {  
  }, {  
    "name" : "Norway"  
  },  
  "cards" : [ ],  
  "positions" : [ {  
    "position_id" : 24,  
    "position" : "Left Center Forward",  
    "from" : "82:37",  
    "to" : "83:45",  
    "from_period" : 2,  
    "to_period" : 2,  
    "start_reason" : "Substitution - On (Tactical)",  
    "end_reason" : "Tactical Shift"  
  }, {  
    "position_id" : 13,  
    "position" : "Right Center Midfield",
```

360.json

Using the match_id, you can access the file in the 360 folder, if available.

- Each file contains freeze frames associated with selected events of the match.
- For each freeze frame, you can observe the locations of the visible players and the visible area of the pitch.
- The event_uuid links each freeze frame to the corresponding event in the events file.

We have info such as:

- **Location:** position of the player on the pitch [x, y]
- **Teammate:** indicates whether the player is a teammate or an opponent
- **Actor:** indicates whether the player is performing the action
- **Keeper:** indicates whether the player is a goalkeeper

```
}, {  
  "teammate" : true,  
  "actor" : false,  
  "keeper" : false,  
  "location" : [ 61.25664852014171, 56.417276328163084 ]  
}, {  
  "teammate" : true,  
  "actor" : false,  
  "keeper" : false,  
  "location" : [ 61.76994556042512, 30.461066072962247 ]  
} ]  
}, {  
  "event_uuid" : "7b823cb0-f2d0-49fb-b7ba-1c2941fd8752",  
  "visible_area" : [ 11.10157189302178, 80.0, 41.282959729,  
  "freeze_frame" : [ {  
    "teammate" : true,  
    "actor" : false,
```

Info_physical player_per_years.csv

This CSV provides additional player-level information collected over time. The same player may appear in multiple rows because observations can refer to different years or dates.

It contains info such as:

- **athlete_id**: player identifier used to link this file with the others
- **age_years**: player age at the time of the observation
- **stature_cm**: player height
- **body_mass_kg**: player body mass
- **body_mass_index**: BMI
- **playing_role**: player position or set of positions
- **dominant_foot**: preferred foot
- **endurance_rating**: ability to sustain physical effort
- **strength_rating**: physical strength
- **sprint_rating**: running speed
- **acceleration_rating**: ability to increase speed quickly
- **agility_rating**: ability to change direction quickly
- **jump_rating**: jumping ability
- **reaction_rating**: reaction ability
- **composure_rating**: ability to maintain performance in demanding situations

Players_running.csv

This CSV contains physiological measurements collected during incremental running tests for a subset of players.

- **athlete/player identifier:** links the test to the corresponding football player
- **test stage:** consecutive stage of the incremental running test
- **heart rate:** cardiovascular response during exercise
- **oxygen uptake (VO₂):** oxygen consumption during exercise
- **blood lactate:** lactate concentration measured during the test
- **other physiological measurements:** additional information collected during the test, when available

Be careful with the data!

The repository is large, heterogeneous, and not all information is available everywhere.

- Some values are missing
- Some records may be duplicated or repeated over time
- Some variables may be unclear or difficult to interpret
- Different players, seasons, and competitions may have **different coverage**
- Some competitions may contain **only a subset of the expected matches**
- The different data sources may **not overlap completely**
- Some values may be **unexpected, inconsistent, or outliers**
- Files have **different granularities**, so they cannot always be merged directly

Be careful with the data!

Work with the data step by step

Trying to load the entire repository at once may fail (too much memory).

- Load and inspect **one data source at a time**
- Process only the files needed for the current analysis
- Save **intermediate cleaned datasets** when useful
- Merge different sources only after understanding their structure and keys

Before starting the analysis, understand what the data actually contain.

Project Groups and Research Questions

- Form a group of **3 or 4 students**.
- Each group must choose **one project**, associated with a specific **research question**.
- Each research question is designed for a specific group size: **3 or 4 students**.
- A group of **3 students may choose a project designed for 4 students**.
- A group of **4 students cannot choose a project designed for 3 students**.
- Each research question (project) can be selected by **at most 2 groups**.
- If two groups select the same research question, they must work **independently** and develop their own analysis, methodological choices, results, and report.
- **Choose the project that best matches your group size and interests!**

Project Groups and Research Questions

- Each project is composed by 4 tasks:
 - Data Understanding & Pre-processing
 - Clustering
 - Classification
 - Time series analysis

The research question will guide the entire project: different research questions will require different choices, analyses, and solutions.

Project	Main research question	Students
Male and Female Player Behaviour in Football	Do male and female football players exhibit different behavioural, physical, and physiological profiles?	4
Player Career Evolution in Football	Do football players maintain the same playing profile throughout their career?	3
Behaviour Under Pressure in Football	How does pressure affect player actions and behaviour in football?	3
Player Behaviour Across Playing Time	How does player behaviour and performance change with increasing playing time during a football match?	3
Goalkeeper Profiles in Football	Do goalkeepers exhibit distinctive behavioural, physical, and physiological profiles compared with other football players?	3
Defender Profiles in Football	Do defenders exhibit distinctive behavioural, physical, and physiological profiles compared with other football players, and how do these profiles differ across defensive roles?	4
Midfielder Profiles in Football	Do midfielders exhibit distinctive behavioural, physical, and physiological profiles, and how do these profiles differ across midfield roles?	4
Forward Profiles in Football	Do forwards exhibit distinctive behavioural, physical, and physiological profiles, and how do these profiles differ across offensive roles?	3

Tentative Project Work & Check Schedule

- **05/10 (Monday)** Project Work
- **02/11**(Monday) Project Work
- **04/11** (Wednesday) **Mandatory Check**
- **09/12** (Wednesday) Project Work
- **14/12** (Monday) **Mandatory Check**

NB: These dates are tentative and may be subject to changes during the course.