1 Players

This dataset describes all players of the teams playing in seven national and international soccer competitions (Italian, Spanish, French, German, English first divisions, World Cup 2018, European Cup 2016). Each competition consists of the following fields:

- **birthArea**: geographic information about the player's birth area;
- **birthDate**: the birth date of the player, in the format "YYYY-MM-DD";
- **currentNationalTeamId**: the identifier of the national team where the players currently plays;
- **currentTeamId**: the identifier of the team where the player plays for. The identifier refers to the field "wyId" in a team document;
- **firstName**: the first name of the player;
- **lastName**: the last name of the player;
- **foot**: the preferred foot of the player;
- **height**: the height of the player (in centimeters);
- **middleName**: the middle name (if any) of the player;
- **passportArea**: the geographic area associated with the player's current passport;
- **role**: the main role of the player. It is a subdocument containing the role's name and two abbreviations of it;
- **shortName2**: the short name of the player;
- **weight**: the weight of the player (in kilograms);
- **wyId**: the identifier of the player, assigned by Wyscout.

2 Teams

This dataset describes all the soccer teams in seven prominent soccer competitions (Italian, Spanish, German, French and English first divisions, World Cup 2018, European Cup 2016). It consists of the following fields:

- **city**: the city where the team is located. For national teams it is the capital of the country;
- **name**: the common name of the team;
- **area**: information about the geographic area associated with the team;
- **wyId**: the identifier of the team, assigned by Wyscout;
- **officialName**: the official name of the team (e.g., Juventus FC);
- **type**: the type of the team. It is "club" for teams in the competitions for clubs and "national" for the teams in international competitions;

3 Competitions

This dataset describes seven major soccer competitions (Italian, Spanish, German, French, English first divisions, World cup 2018, European cup 2016). Each competition is a document consisting of the following fields:

- **area**: it denotes the geographic area associated with the league as a sub-document, using the ISO 3166-1 specification (https://www.iso.org/iso-3166-country-codes.html);
- **format**: the format of the competition. All competitions for clubs have value "Domestic league". The competitions for national teams have value "International cup";
- **name**: the official name of the competition (e.g., Italian first division, Spanish first division, World Cup, etc.);
- **type**: the typology of the competition. It is "club" for the competitions for clubs and "international" for the competitions for national teams (World Cup 2018, European Cup 2016);
- **wyId**: the unique identifier of the competition, assigned by Wyscout.

4 Coaches

The coaches data set describes all coaches of the clubs and the national teams of the seven competitions we make available. It consists of the following fields:

- **wyId**: the identifier of the coach, assigned by Wyscout.
- **shortName**: the short name of the coach;
- **firstName**: the first name of the coach;
- **middleName**: the middle name (if any) of the coach;
- **lastName**: the last name of the coach;
- **birthDate**: the birth date of the coach, in the format "YYYY-MM-DD";
- **birthArea**: geographic information about the coach's birth area;
- **passportArea**: the geographic area associated with the referee's current passport;
- **currentTeamId**: the identifier of the coach's team. The identifier refers to the field "wyId" in a team document.

5 Events

This dataset describes all the events that occur during each match. Each event refers to a ball touch and contains the following information:

- **eventId**: the identifier of the event's type. Each eventId is associated with an event name (see next point);
- **eventName**: tteamIdhe name of the event's type. There are seven types of events: pass, foul, shot, duel, free kick, offside and touch;
- **subEventId**: the identifier of the subevent's type. Each subEventId is associated with a subevent name (see next point);
- **subEventName**: the name of the subevent's type. Each event type is associated with a different set of subevent types;
- **tags**: a list of event tags, each one describes additional information about the event (e.g., accurate). Each event type is associated with a different set of tags;
- **eventSec**: the time when the event occurs (in seconds since the beginning of the current half of the match);
- **id**: a unique identifier of the event;
- **matchId**: the identifier of the match the event refers to. The identifier refers to the field "wyId" in the match dataset;
- **matchPeriod**: the period of the match. It can be "1H" (first half of the match), "2H" (second half of the match), "E1" (first extra time), "E2" (second extra time) or "P" (penalties time);
- **playerId**: the identifier of the player who generated the event. The identifier refers to the field "wyId" in a player dataset;
- **positions**: the origin and destination positions associated with the event. Each position is a pair of coordinates (x, y). The x and y coordinates are always in the range [0, 100] and indicate the percentage of the field from the perspective of the attacking team. In particular, the value of the x coordinate indicates the event's nearness (in percentage) to the opponent's goal, while the value of the y coordinates indicates the event's nearness (in percentage) to the right side of the field;
- **teamId**: the identifier of the player's team. The identifier refers to the field "wyId" in the team dataset.

6 Matches

This dataset describes all the matches made available. Each match is a document consisting of the following fields:

- **competitionId**: the identifier of the competition to which the match belongs to. It is a integer and refers to the field "wyId" of the competition document;

- **date** and **dateutc**: the former specifies date and time when the match starts in explicit format (e.g., May 20, 2018 at 8:45:00 PM GMT+2), the latter contains the same information but in the compact format YYYY-MM-DD hh:mm:ss;

- **duration**: the duration of the match. It can be "Regular" (matches of regular duration of 90 minutes + stoppage time), "ExtraTime" (matches with supplementary times, as it may happen for matches in continental or international competitions), or "Penalities" (matches which end at penalty kicks, as it may happen for continental or international competitions);

- **gameweek**: the week of the league, starting from the beginning of the league;

- label: contains the name of the two clubs and the result of the match (e.g., "Lazio - Internazionale, 2 - 3");

- **roundID**: indicates the match-day of the competition to which the match belongs to. During a competition for soccer clubs, each of the participating clubs plays against each of the other clubs twice, once at home and once away. The matches are organized in match-days: all the matches in match-day i are played before the matches in match-day i + 1, even tough some matches can be anticipated or postponed to facilitate players and clubs participating in Continental or Intercontinental competitions. During a competition for national teams, the "roundID" indicates the stage of the competition (eliminatory round, round of 16, quarter finals, semifinals, final);

- **seasonId**: indicates the season of the match;

- **status**: it can be "Played" (the match has officially finished), "Cancelled" (the match has been canceled for some reason), "Postponed" (the match has been postponed and no new date and time is available yet) or "Suspended" (the match has been suspended and no new date and time is available yet);

- **venue**: the stadium where the match was held (e.g., "Stadio Olimpico");

- **winner**: the identifier of the team which won the game, or 0 if the match ended with a draw;

- **wyId**: the identifier of the match, assigned by Wyscout;

- **teamsData**: it contains several subfields describing information about each team that is playing that match: such as lineup, bench composition, list of substitutions, coach and scores:

    - **hasFormation**: it has value 0 if no formation (lineups and benches) is present, and 1 otherwise;

    - score: the number of goals scored by the team during the match (not counting penalties);

    - **scoreET**: the number of goals scored by the team during the match, including the extra time (not counting penalties);

    - **scoreHT**: the number of goals scored by the team during the first half of the match;

    - **scoreP**: the total number of goals scored by the team after the penalties;

    - **side**: the team side in the match (it can be "home" or "away");

    - **teamId**: the identifier of the team;

    - **coachId**: the identifier of the team's coach;

    - **formation**:

        - **bench**: the list of the team's players that started the match in the bench and some basic statistics about their performance during the match (goals, own goals, cards);

        - **lineup**: the list of the team's players in the starting lineup and some basic statistics about their performance during the match (goals, own goals, cards);

        - **substitutions**: the list of team's substitutions during the match, describing the players involved and the minute of the substitution.

7 Events and subevents

Pass: Launch, High pass, Simple pass, Cross, Head pass, Hand pass, Smart pass
Duel: Air duel, Ground attacking duel, Ground loose ball duel, Ground defending duel
Free Kick: Free kick cross, Free Kick, Penalty, Corner, Goal kick, Free kick shot, Throw in
Foul: Protest, Time lost foul, Out of game foul, Hand foul, Violent Foul, Simulation, Foul, Late card foul
Others on the ball: Clearance, Acceleration, Touch
Shot: Shot, Save attempt, Reflexes, Save attempt
Goalkeeper leaving line: Goalkeeper leaving line
Offside
Interruption

8 Tags Ids

Tag	Label	Description
101	Goal	Goal
102	own_goal	Own goal
301	assist	Assist
302	keyPass	Key pass
1901	counter_attack	Counter attack
401	Left	Left foot
402	Right	Right foot
403	head/body	Head/body
1101	direct	Direct
1102	indirect	Indirect
2001	dangerous_ball_lost	Dangerous ball lost
2101	blocked	Blocked
801	high	High
802	low	Low
1401	interception	Interception
1501	clearance	Clearance
201	opportunity	Opportunity
1301	Feint	Feint
1302	missed ball	Missed ball
501	free_space_r	Free space right
502	free_space_l	Free space left
503	take_on_l	Take on left
504	take_on_r	Take on right
1601	sliding_tackle	Sliding tackle
601	anticipated	Anticipated
602	anticipation	Anticipation
1701	red_card	Red card
1702	yellow_card	Yellow card
1703	second_yellow_card	Second yellow card
1201	gb	Position: Goal low center
1202	gbr	Position: Goal low right
1203	gc	Position: Goal center
1204	gl	Position: Goal center left
1205	glb	Position: Goal low left
1206	gr	Position: Goal center right
1207	gt	Position: Goal high center
1208	gtl	Position: Goal high left
1209	gtr	Position: Goal high right
1210	obr	Position: Out low right
1211	ol	Position: Out center left
1212	olb	Position: Out low left
1213	or	Position: Out center right
1214	ot	Position: Out high center
1215	otl	Position: Out high left
1216	otr	Position: Out high right
1217	pbr	Position: Post low right
1218	pl	Position: Post center left
1219	plb	Position: Post low left
1220	pr	Position: Post center right
1221	pt	Position: Post high center
1222	ptl	Position: Post high left
1223	ptr	Position: Post high right
901	through	Through
1001	fairplay	Fairplay
701	lost	Lost
702	neutral	Neutral
703	won	Won
1801	accurate	Accurate
1802	not accurate	Not accurate