Codebook v 1.2.0

The Fjelstul World Cup Database

Joshua C. Fjelstul, Ph.D.

Datasets

tournaments

Tournaments

Description

This dataset records all World Cup tournaments. There is one observation per tournament. It includes the host of the tournament, the winner of the tournament, the start and end dates of the tournament, and information about the format of the tournament. There are 18 variables and 30 observations.

Variables

`key_id`	`integer` The unique ID number for the observation.
`tournament_id`	`text` The unique ID number for the tournament. Has the format `WC-####`, where the number is the year of the tournament.
`tournament_name`	`text` The name of the tournament.
`year`	`integer` The year of the tournament.
`start_date`	`date` The start date of the tournament in the format `YYYY-MM-DD`.
`end_date`	`date` The end date of the tournament in the format `YYYY-MM-DD`.
`host_country`	`text` The unique ID number for the country that hosted the tournament. References `team_id` in the `teams` dataset.

`winner`	`text` The name of the team that won the tournament.
`host_won`	`boolean` Whether one of the host countries won the tournament. Coded `1` if one of the host countries won and `0` otherwise.
`count_teams`	`integer` The number of teams that participated in the tournament.
`group_stage`	`boolean` Whether the match is a group stage match. Coded `1` if the match is a group stage match and `0` otherwise.
`second_group_stage`	`boolean` Whether there was a second group stage. Coded `1` if there was a second group stage and `0` otherwise.
`final_round`	`boolean` Whether there was a final round. Coded `1` if there was a final round and `0` otherwise.
`round_of_16`	`boolean` Whether there was a round of 16 stage. Coded `1` if there was a round of 16 stage and `0` otherwise.
`quarter_finals`	`boolean` Whether there was a quarter-finals stage. Coded `1` if there was a quarter-finals stage and `0` otherwise.
`semi_finals`	`boolean` Whether there was a semi-finals stage. Coded `1` if there was a semi-finals stage and `0` otherwise.
`third_place_match`	`boolean` Whether there was a third-place match. Coded `1` if there was a third-place match and `0` otherwise.
`final`	`boolean` Whether there was a final match. Coded `1` if there was a final match and `0` otherwise.

confederations

Confederations

Description

This dataset records all FIFA confederations. There is one observation per confederation. There are 5 variables and 6 observations.

Variables

`key_id`	`integer`	The unique ID number for the observation.
`confederation_id`	`text`	The unique ID number for the confederation. Has the format `CF-#`, where the number is a counter that is assigned with the confederations sorted in alphabetical order.
`confederation_name`	`text`	The name of the confederation.
`confederation_code`	`text`	The abbreviation for the confederation.
`confederation_wikipedia_link`	`text`	The Wikipedia link for the confederation.

teams

Teams

Description

This dataset records all teams who have participated in a World Cup match. There is one observation per team. It includes the 3-letter ISO code for each country, whether the country's men's and women's team have qualified for a tournament, the name of the country's national federation, the country's FIFA confederation, and links to the Wikipedia pages for their men's and women's teams, if the team has qualified for a tournament. There are 14 variables and 88 observations.

Variables

`key_id`	`integer` The unique ID number for the observation.
`team_id`	`text` The unique ID number for the team. Has the format `T-##`, where the number is a counter that is assigned with the data sorted by the year of the team's first tournament and then by the team's name.
`team_name`	`text` The name of the team.
`team_code`	`text` The 3-letter code for the team.
`mens_team`	`boolean` Whether the country's men's team has qualified for a tournament.
`womens_team`	`boolean` Whether the country's women's team has qualified for a tournament.

`federation_name`	`text`	The name of the team's federation.
`region_name`	`text`	The name of the region that the country is located in.
`confederation_id`	`text`	The unique ID number for the confederation. References `confederation_id` in the `confederations` dataset.
`confederation_name`	`text`	The name of the confederation.
`confederation_code`	`text`	The abbreviation for the confederation.
`mens_team_wikipedia_link`	`text`	The Wikipedia link for country's men's team. Coded `not applicable` if the country's men's team has not qualified for a tournament.
`womens_team_wikipedia_link`	`text`	The Wikipedia link for country's women's team. Coded `not applicable` if the country's women's team has not qualified for a tournament.
`federation_wikipedia_link`	`text`	The Wikipedia link of the team's federation.

players

Players

Description

This dataset records all players who have participated in a World Cup match, including players on the bench. There is one observation per player. It includes their name, their birth date, their sex, and a link to their Wikipedia page, if they have one. Note that it does not include their team, as a small number of players represent different countries in different tournaments. There are 13 variables and 10401 observations.

Variables

`key_id`	`integer` The unique ID number for the observation.
`player_id`	`text` The unique ID number for the player. Has the format `P-#####`, where the number is a randomly drawn, uniquely identifying number.
`family_name`	`text` The family name of the player.
`given_name`	`text` The given name of the player.
`birth_date`	`date` The birth date of the player in the format `YYYY-MM-DD`.
`female`	`boolean` Whether the player is female. Coded `1` if the player is female and `0` if the player is male.

`goal_keeper`	`boolean` Whether the player was a goal keeper. Coded `1` if the player was a goal keeper and `0` otherwise.
`defender`	`boolean` Whether the player was a defender. Coded `1` if the player was a defender and `0` otherwise.
`midfielder`	`boolean` Whether the player was a midfielder. Coded `1` if the player was a midfielder and `0` otherwise.
`forward`	`boolean` Whether the player was a forward. Coded `1` if the player was a forward and `0` otherwise.
`count_tournaments`	`integer` The number of tournaments that the player participated in.
`list_tournaments`	`text` A list of tournaments that the player participated in, separated by a comma.
`player_wikipedia_link`	`text` The name of the team of the player.

managers

Managers

Description

This dataset records all managers who have participated in a World Cup match. There is one observation per manager. It includes their name, their sex, their home country, and a link to their Wikipedia page, if they have one. There are 7 variables and 475 observations.

Variables

`key_id`	`integer` The unique ID number for the observation.
`manager_id`	`text` The unique ID number for the manager. Has the format `M-####`, where the number is a counter that is assigned with the data sorted by the year of the manager's first appearance, then by the manager's family name, and then by the manager's given name.
`family_name`	`text` The family name of the manager.
`given_name`	`text` The given name of the manager.
`female`	`boolean` Whether the manager is female. Coded `1` if the manager is female and `0` if the manager is male.
`country_name`	`text` The name of the manager's home country.
`manager_wikipedia_link`	`text` The Wikipedia link for the manager.

referees

Referees

Description

This dataset records all referees who have participated in a World Cup match. There is one observation per referee. It includes their name, their sex, their home country, their confederation, and a link to their Wikipedia page, if they have one. There are 10 variables and 493 observations.

Variables

`key_id`	`integer` The unique ID number for the observation.
`referee_id`	`text` The unique ID number for the referee. Has the format `R-####`, where the number is a counter that is assigned with the data sorted by the year of the referee's first appearance, then by the referee's family name, and then by the referee's given name.
`family_name`	`text` The family name of the referee.
`given_name`	`text` The given name of the referee.
`female`	`boolean` Whether the referee is female. Coded `1` if the referee is female and `0` if the referee is male.
`country_name`	`text` The name of the referee's home country.
`confederation_id`	`text` The unique ID number for the confederation. References `confederation_id` in the `confederations` dataset.

`confederation_name`	`text`	The name of the confederation.
`confederation_code`	`text`	The abbreviation for the confederation.
`referee_wikipedia_link`	`text`	The Wikipedia link for the referee.

stadiums

Stadiums

Description

This dataset records all stadiums that have hosted a World Cup match. There is one observation per stadium. It includes the country and city of the stadium, the approximate capacity of the stadium, and the link to the Wikipedia pages for the city and the stadium. There are 8 variables and 240 observations.

Variables

`key_id`	`integer` The unique ID number for the observation.
`stadium_id`	`text` The unique ID number for the stadium. Has the format `S-###`, where the number is a count that is assigned with the data sorted by country, then by city, then by the name of the stadium.
`stadium_name`	`text` The name of the stadium.
`city_name`	`text` The city in which the match was played.
`country_name`	`text` The name of the country in which the stadium is located.
`stadium_capacity`	`integer` The approximate capacity of the stadium.
`stadium_wikipedia_link`	`text` The Wikipedia link for the stadium.

`city_wikipedia_link`	`text` The Wikipedia link for the city in which the match was played.

matches

Matches

Description

This dataset records all World Cup matches. There is one observation per match per tournament. It includes the home team, the away team, the date of the match, the country, city, and stadium that the match was played in, the final score, the score margin for each team, whether the match went to extra time, whether there was a penalty shootout, the number of penalties scored in the shootout (if applicable), the result of the match (home team win, away team win, draw, replayed), and the winner (if applicable). There are 38 variables and 1248 observations.

Variables

`key_id`	`integer` The unique ID number for the observation.
`tournament_id`	`text` The unique ID number for the tournament. References `tournament_id` in the `tournaments` dataset.
`tournament_name`	`text` The name of the tournament.
`match_id`	`text` The unique ID number for the match. Has the format `M-####-##`, where the first number is the year of the tournament and the second number is a within-tournament counter that is assigned with the data sorted by the date of the match, then by the time of the match, then by the name of the group, and then by name of the home team.

`match_name`	`text` The name of the match.
`stage_name`	`enum` The stage of the tournament in which the match occurred. The possible values are: `first round`, `second round`, `group stage`, `round of sixteen`, `quarter-finals`, `semi-finals`, `third place match`, `final`. Note that not all values are applicable to all tournaments.
`group_name`	`text` The name of the group.
`group_stage`	`boolean` Whether the match is a group stage match. Coded `1` if the match is a group stage match and `0` otherwise.
`knockout_stage`	`boolean` Whether the match is a knockout stage match. Coded `1` if the match is a knockout stage match and `0` otherwise.
`replayed`	`boolean` Whether the match was replayed. Coded `1` if the match was replayed and `0` otherwise.
`replay`	`boolean` Whether the match was a replay. Coded `1` if the match was a replay and `0` otherwise.
`match_date`	`date` The date of the match in the format `YYYY-MM-DD`.
`match_time`	`integer` The time of the match in the format `HH:MM`.
`stadium_id`	`text` The unique ID number for the stadium. References `stadium_id` in the `stadiums` dataset.
`stadium_name`	`text` The name of the stadium.
`city_name`	`text` The city in which the match was played.
`country_name`	`text` The name of the country in which the match was played.
`home_team_id`	`text` The unique ID number for the home team. References `team_id` in the `teams` dataset.
`home_team_name`	`text` The name of the home team. See the `teams` dataset.
`home_team_code`	`text` The 3-letter code for the home team.

`away_team_id`	`text` The unique ID number for the away team. References `team_id` in the `teams` dataset.
`away_team_name`	`text` The name of the away team. See the `teams` dataset.
`away_team_code`	`text` The 3-letter code for the away team.
`score`	`text` The score of the match in the format `#-#`, where the first number is the score of the home team and the second number is the score of the away team.
`home_team_score`	`integer` The score of the home team.
`away_team_score`	`integer` The score of the away team.
`home_team_score_margin`	`integer` The score margin for the home team.
`away_team_score_margin`	`integer` The score margin for the away team.
`extra_time`	`boolean` Whether the match went to extra time. Coded `1` if the match went to extra time and `0` otherwise.
`penalty_shootout`	`boolean` Whether the match ended in a penalty shootout. Coded `1` if the match ended in a penalty shootout and `0` otherwise.
`score_penalties`	`text` The score of the penalty shootout in the format `#-#`. Coded `0-0` if there was not a penalty shootout.
`home_team_score_penalties`	`integer` The score of the home team in the penalty shootout. Coded `NA` if there was not a penalty shootout.
`away_team_score_penalties`	`integer` The score of the away team in the penalty shootout. Coded `NA` if there was not a penalty shootout.
`result`	`enum` The result of the match. The possible values are: `home team win`, `away team win`, `draw`, `replayed`.
`home_team_win`	`boolean` Whether the home team won the match. Coded `1` if the home team won the match and `0` otherwise.
`away_team_win`	`boolean` Whether the home team won the match. Coded `1` if the home team won the match and `0` otherwise.

`draw`	`boolean` Whether the match ended in a draw. Coded `1` of the match ended in a draw and `0` otherwise.

awards

Awards

Description

This dataset records all individual awards that are handed out to players. There is one observation per award. It includes the name of the award, the year the award was first introduced, and a description of the award. There are 5 variables and 8 observations.

Variables

`key_id`	`integer` The unique ID number for the observation.
`award_id`	`text` The unique ID number for the award. Has the format `A-#`, where the number is a counter.
`award_name`	`enum` The name of the award. The possible values are: `Golden Ball`, `Silver Ball`, `Bronze Ball`, `Golden Boot`, `Silver Boot`, `Bronze Boot`, `Golden Glove`, `Best Young Player`.
`award_description`	`text` A description of the award.
`year_introduced`	`integer` The year the award was first introduced.

qualified_teams

Qualified teams

Description

This dataset records all qualified teams. There is one observation per team per tournament. It includes the tournament, the team, and the performance of each team (the furthest stage reached). There are 8 variables and 625 observations.

Variables

`key_id`	`integer` The unique ID number for the observation.
`tournament_id`	`text` The unique ID number for the tournament. References `tournament_id` in the `tournaments` dataset.
`tournament_name`	`text` The name of the tournament.
`team_id`	`text` The unique ID number for the team. References `team_id` in the `teams` dataset.
`team_name`	`text` The name of the team.
`team_code`	`text` The 3-letter code for the team.
`count_matches`	`integer` The number of matches that the team played in the tournament.
`performance`	`text` The furthest stage of the tournament reached by the team.

squads

Squads

Description

This dataset records the composition of each squad. There is one observation per player per team per tournament. It includes the position of each player, the shirt number of each player (from 1954), the current club of each player, and a link to the Wikipedia page for the club, if it has one. There are 14 variables and 13843 observations.

Variables

`key_id`	`integer` The unique ID number for the observation.
`tournament_id`	`text` The unique ID number for the tournament. References `tournament_id` in the `tournaments` dataset.
`tournament_name`	`text` The name of the tournament.
`team_id`	`text` The unique ID number for the team. References `team_id` in the `teams` dataset.
`team_name`	`text` The name of the team of the player.
`team_code`	`text` The 3-letter code for the team.
`player_id`	`text` The unique ID number for the player. References `player_id` in the `players` dataset.

`family_name`	`text` The family name of the player.
`given_name`	`text` The given name of the player.
`shirt_number`	`integer` The shirt number of the player.
`position_name`	`enum` The position of the player. The possible values are: `goal keeper`, `defender`, `midfielder`, `forward`.
`position_code`	`enum` The code for the position of the player. The possible values are: `GK`, `DF`, `MF`, `FW`.

manager_appointments

Manager appointments

Description

This dataset records all manager appointments. There is one observation per manager per team per tournament. There are some teams that have co-managers. There are 10 variables and 637 observations.

Variables

`key_id`	`integer` The unique ID number for the observation.
`tournament_id`	`text` The unique ID number for the tournament. References `tournament_id` in the `tournaments` dataset.
`tournament_name`	`text` The name of the tournament.
`team_id`	`text` The unique ID number for the team. References `team_id` in the `teams` dataset.
`team_name`	`text` The name of the team of the manager.
`team_code`	`text` The 3-letter code for the team.
`manager_id`	`text` The unique ID number for the manager. References `manager_id` in the `managers` dataset.
`family_name`	`text` The family name of the manager.

`given_name`	`text`	The given name of the manager.
`country_name`	`text`	The name of the manager's home country.

referee_appointments

Referee appointments

Description

This dataset records all referee appointments. There is one observation per referee per tournament. This dataset only includes the main referee, not assistant referees, fourth officials, or video assistant referees.There are 10 variables and 668 observations.

Variables

`key_id`	`integer` The unique ID number for the observation.
`tournament_id`	`text` The unique ID number for the tournament. References `tournament_id` in the `tournaments` dataset.
`tournament_name`	`text` The name of the tournament.
`referee_id`	`text` The unique ID number for the referee. References `referee_id` in the `referees` dataset.
`family_name`	`text` The family name of the referee.
`given_name`	`text` The given name fo the referee.
`country_name`	`text` The name of the referee's home country.
`confederation_id`	`text` The unique ID number for the confederation. References `confederation_id` in the `confederations` dataset.

`confederation_name`	`text`	The name of the confederation.
`confederation_code`	`text`	The abbreviation for the confederation.

team_appearances

Team appearances

Description

This dataset records all team appearances. There is one observation per team per match per tournament. It includes whether the team is the home team or the away team, the number of goals for and against, the goal difference, whether there was a penalty shootout, penalties for and against (if applicable), and whether the team wins, loses, or draws. There are 37 variables and 2496 observations.

Variables

`key_id`	`integer` The unique ID number for the observation.
`tournament_id`	`text` The unique ID number for the tournament. References `tournament_id` in the `tournaments` dataset.
`tournament_name`	`text` The name of the tournament.
`match_id`	`text` The unique ID number for the match. References `match_id` in the `matches` dataset.
`match_name`	`text` The name of the match.
`stage_name`	`enum` The stage of the tournament in which the match occurred. The possible values are: `first round`, `second round`, `group stage`, `round of sixteen`, `quarter-finals`,

	`semi-finals`, `third place match`, `final`. Note that not all values are applicable to all tournaments.
`group_name`	`text` The name of the group.
`group_stage`	`boolean` Whether the match is a group stage match. Coded `1` if the match is a group stage match and `0` otherwise.
`knockout_stage`	`boolean` Whether the match is a knockout stage match. Coded `1` if the match is a knockout stage match and `0` otherwise.
`replayed`	`boolean` Whether the match was replayed. Coded `1` if the match was replayed and `0` otherwise.
`replay`	`boolean` Whether the match was a replay. Coded `1` if the match was a replay and `0` otherwise.
`match_date`	`date` The date of the match in the format `YYYY-MM-DD`.
`match_time`	`integer` The time of the match in the format `HH:MM`.
`stadium_id`	`text` The unique ID number for the stadium. References `stadium_id` in the `stadiums` dataset.
`stadium_name`	`text` The name of the stadium.
`city_name`	`text` The city in which the match was played.
`country_name`	`text` The name of the country in which the match was played.
`team_id`	`text` The unique ID number for the team. References `team_id` in the `teams` dataset.
`team_name`	`text` The name of the team.
`team_code`	`text` The 3-letter code for the team.
`opponent_id`	`text` The unique ID number for the team's opponent. References `team_id` in the `teams` dataset.
`opponent_name`	`text` The name of the team's opponent.

`opponent_code`	`text` The 3-letter code for the team's opponent.
`home_team`	`boolean` Whether the team was the home team. Coded `1` if the team was the home team and `0` otherwise.
`away_team`	`boolean` Whether the team was the away team. Coded `1` if the team was the away team and `0` otherwise.
`goals_for`	`integer` The number of goals scored by the team.
`goals_against`	`integer` The number of goals scored against the team.
`goal_differential`	`integer` The team's goal differential.
`extra_time`	`boolean` Whether the match went to extra time. Coded `1` if the match went to extra time and `0` otherwise.
`penalty_shootout`	`boolean` Whether the match ended in a penalty shootout. Coded `1` if the match ended in a penalty shootout and `0` otherwise.
`penalties_for`	`integer` The number of penalties scored by the opponent, if the match ended in a penalty shootout. Coded `0` if there was not a shootout.
`penalties_against`	`integer` The number of penalties scored by the team, if the match ended in a penalty shootout. Coded `0` if there was not a shootout.
`result`	`enum` The result of the match. The possible values are: `home team win`, `away team win`, `draw`, `replayed`.
`win`	`boolean` Whether the team won the match. Coded `1` if the team won the match and `0` otherwise.
`lose`	`boolean` Whether the team lost the match. Coded `1` if the team lost the match and `0` otherwise.
`draw`	`boolean` Whether the match ended in a draw. Coded `1` of the match ended in a draw and `0` otherwise.

player_appearances

Player appearances

Description

This dataset records all player appearances since 1970. There is one observation per player per team per match per tournament. It includes players who play in the match, including players who are in the starting eleven and players who come in as substitutes. FIFA match reports do not include information about substitutions before 1970. There are 21 variables and 27432 observations.

Variables

`key_id`	`integer` The unique ID number for the observation.
`tournament_id`	`text` The unique ID number for the tournament. References `tournament_id` in the `tournaments` dataset.
`tournament_name`	`text` The name of the tournament.
`match_id`	`text` The unique ID number for the match. References `match_id` in the `matches` dataset.
`match_name`	`text` The name of the match.
`match_date`	`date` The date of the match in the format `YYYY-MM-DD`.
`stage_name`	`enum` The stage of the tournament in which the match occurred. The possible values are: `first round`, `second round`, `group stage`, `round of sixteen`, `quarter-finals`,

	`semi-finals`, `third place match`, `final`. Note that not all values are applicable to all tournaments.
`group_name`	`text` The name of the group.
`team_id`	`text` The unique ID number for the team of the player. References `team_id` in the `teams` dataset.
`team_name`	`text` The name of the team of the player.
`team_code`	`text` The 3-letter code for the team of the player.
`home_team`	`boolean` Whether the team was the home team. Coded `1` if the team was the home team and `0` otherwise.
`away_team`	`boolean` Whether the team was the away team. Coded `1` if the team was the away team and `0` otherwise.
`player_id`	`text` The unique ID number for the player. References `player_id` in the `players` dataset.
`family_name`	`text` The family name of the player.
`given_name`	`text` The given name of the player.
`shirt_number`	`integer` The shirt number of the player.
`position_name`	`text` The name of the position of the player.
`position_code`	`text` A 2-letter or 3-letter code that indicates the position of the player.
`starter`	`boolean` Whether the player started the match. Coded `1` if the player started the match and `0` otherwise.
`substitute`	`boolean` Whether the player was a substitute. Coded `1` if the player was a substitute and `0` otherwise.

manager_appearances

Manager appearances

Description

This dataset records all manager appearances. There is one observation per manager per team per match per tournament. There are some teams that have co-managers. There are 17 variables and 2538 observations.

Variables

`key_id`	`integer` The unique ID number for the observation.
`tournament_id`	`text` The unique ID number for the tournament. References `tournament_id` in the `tournaments` dataset.
`tournament_name`	`text` The name of the tournament.
`match_id`	`text` The unique ID number for the match. References `match_id` in the `matches` dataset.
`match_name`	`text` The name of the match.
`match_date`	`date` The date of the match in the format `YYYY-MM-DD`.
`stage_name`	`enum` The stage of the tournament in which the match occurred. The possible values are: `first round`, `second round`, `group stage`, `round of sixteen`, `quarter-finals`,

	`semi-finals`, `third place match`, `final`. Note that not all values are applicable to all tournaments.
`group_name`	`text` The name of the group.
`team_id`	`text` The unique ID number for the team of the manager. References `team_id` in the `teams` dataset.
`team_name`	`text` The name of the team of the manager.
`team_code`	`text` The 3-letter code for the team of the manager.
`home_team`	`boolean` Whether the team was the home team. Coded `1` if the team was the home team and `0` otherwise.
`away_team`	`boolean` Whether the team was the away team. Coded `1` if the team was the away team and `0` otherwise.
`manager_id`	`text` The unique ID number for the manager. References `manager_id` in the `managers` dataset.
`family_name`	`text` The family name of the manager.
`given_name`	`text` The given name of the manager.
`country_name`	`text` The name of the manager's home country.

referee_appearances

Referee appearances

Description

This dataset records all referee appearances. There is one observation per referee per match per tournament. There are 15 variables and 1248 observations.

Variables

`key_id`	`integer` The unique ID number for the observation.
`tournament_id`	`text` The unique ID number for the tournament. References `tournament_id` in the `tournaments` dataset.
`tournament_name`	`text` The name of the tournament.
`match_id`	`text` The unique ID number for the match. References `match_id` in the `matches` dataset.
`match_name`	`text` The name of the match.
`match_date`	`date` The date of the match in the format `YYYY-MM-DD`.
`stage_name`	`enum` The stage of the tournament in which the match occurred. The possible values are: `first round`, `second round`, `group stage`, `round of sixteen`, `quarter-finals`, `semi-finals`, `third place match`, `final`. Note that not all values are applicable to all tournaments.

`group_name`	`text` The name of the group.
`referee_id`	`text` The unique ID number for the referee. References `referee_id` in the `referees` dataset.
`family_name`	`text` The family name of the referee.
`given_name`	`text` The given name of the referee.
`country_name`	`text` The name of the referee's home country.
`confederation_id`	`text` The unique ID number for the confederation. References `confederation_id` in the `confederations` dataset.
`confederation_name`	`text` The name of the confederation.
`confederation_code`	`text` The abbreviation for the confederation.

goals

Goals

Description

This dataset records all goals. There is one observation per goal. It indicates the team that scored the goal, player who scored the goal, the team of the player who scored the goal (to account for own goals), minute of the goal, and whether the goal was scored in the run of play by the opposition, was an own goal, or was a penalty. This dataset does not include converted penalties in a penalty shootout. There are 26 variables and 3637 observations.

Variables

`key_id`	`integer` The unique ID number for the observation.
`goal_id`	`text` The unique ID number for the goal. Has the format `G-####`, where the number is a counter that is assigned with the data sorted by the match ID, then the minute of the goal.
`tournament_id`	`text` The unique ID number for the tournament. References `tournament_id` in the `tournaments` dataset.
`tournament_name`	`text` The name of the tournament.
`match_id`	`text` The unique ID number for the match in which the goal occurred. References `match_id` in the `matches` dataset.
`match_name`	`text` The name of the match in which the goal occurred.

`match_date`	`date` The date of the match in the format `YYYY-MM-DD`.
`stage_name`	`enum` The stage of the tournament in which the match occurred. The possible values are: `first round`, `second round`, `group stage`, `round of sixteen`, `quarter-finals`, `semi-finals`, `third place match`, `final`. Note that not all values are applicable to all tournaments.
`group_name`	`text` The name of the group.
`team_id`	`text` The unique ID number for the team that scored the goal. References `team_id` in the `teams` dataset. For own goals, this is the team that is awarded the goal, not the team of the player who scored the own goal.
`team_name`	`text` The name of the team of the player who scored the goal.
`team_code`	`text` The 3-letter code for the team of the player who scored the goal.
`home_team`	`boolean` Whether the team was the home team. Coded `1` if the team was the home team and `0` otherwise.
`away_team`	`boolean` Whether the team was the away team. Coded `1` if the team was the away team and `0` otherwise.
`player_id`	`text` The unique ID number for the player who scored the goal. References `player_id` in the `players` dataset.
`family_name`	`text` The family name of the player who scored the goal.
`given_name`	`text` The given name of the player who scored the goal.
`shirt_number`	`integer` The shirt number of the player who scored the goal.
`player_team_id`	`text` The unique ID number for the team of the player who scored the goal. References `team_id` in the `teams` dataset. For own goals, this is the team of the player who scored the own goal, not the team that is awarded the goal.
`player_team_name`	`text` The name of the team of the player who scored the goal.

`player_team_code`	`text` The 3-letter code for the team of the player who scored the goal.
`minute_label`	`text` The minute of the match in which the goal occurred in the format `#'` or `#'+#'`.
`minute_regulation`	`integer` The minute of regulation time in which the substitution occurred.
`minute_stoppage`	`integer` The minute of stoppage time in which the goal occurred. Coded `0` if the substitution did not occur during stoppage time.
`match_period`	`enum` The period of the match in which the goal occurred. The possible values are: `first half`, `first half, stoppage time`, `second half`, `second half, stoppage time`, `extra time, first half`, `extra time, first half, stoppage time`, `extra time, second half`, `extra time, second half, stoppage time`, `after extra time`.
`own_goal`	`boolean` Whether the goal was an own goal. Coded `1` if the goal was an own goal and `0` otherwise.
`penalty`	`boolean` Whether the goal was a penalty that occurred during the game, as opposed to during a penalty shootout. Coded `1` if the goal was a penalty that occurred during the game and `0` otherwise.

penalty_kicks

Penalty kicks

Description

This dataset records all penalty kicks taken during penalty shootouts. There is one observation per penalty kick. This dataset does not include attempted penalty kicks during matches. It indicates minute of each kick, the player who took the kick, and whether the penalty was converted. There are 19 variables and 396 observations.

Variables

`key_id`	`integer` The unique ID number for the observation.
`penalty_kick_id`	`text` The unique ID number for the penalty kick. Has the format `PK-####`, where the number is a counter that is assigned with the data sorted by the match ID, then the minute of the penalty kick.
`tournament_id`	`text` The unique ID number for the tournament. References `tournament_id` in the `tournaments` dataset.
`tournament_name`	`text` The name of the tournament.
`match_id`	`text` The unique ID number for the match in which the penalty kick occurred. References `match_id` in the `matches` dataset.
`match_name`	`text` The name of match in which the penalty kick occurred.
`match_date`	`date` The date of the match in the format `YYYY-MM-DD`.

`stage_name`	`enum` The stage of the tournament in which the match occurred. The possible values are: `first round`, `second round`, `group stage`, `round of sixteen`, `quarter-finals`, `semi-finals`, `third place match`, `final`. Note that not all values are applicable to all tournaments.
`group_name`	`text` The name of the group.
`team_id`	`text` The unique ID number for the team of the player who took the penalty kick. References `team_id` in the `teams` dataset.
`team_name`	`text` The name of the team of the player who took the penalty kick.
`team_code`	`text` The 3-letter code for the team of the player who took the penalty kick.
`home_team`	`boolean` Whether the team was the home team. Coded `1` if the team was the home team and `0` otherwise.
`away_team`	`boolean` Whether the team was the away team. Coded `1` if the team was the away team and `0` otherwise.
`player_id`	`text` The unique ID number for the player who took the penalty kick. References `player_id` in the `players` dataset.
`family_name`	`text` The family name of the player who took the penalty kick.
`given_name`	`text` The given name of the player who took the penalty kick.
`shirt_number`	`integer` The shirt number of the player who took the penalty kick.
`converted`	`boolean` Whether the penalty kick was converted. Coded `1` if the penalty kick was converted and `0` otherwise.

bookings

Bookings

Description

This dataset records all bookings, including yellow cards and red cards, since 1970. The modern system of yellow and red cards was introduced in 1970. There is one observation per booking. It indicates the minute of each booking, the player who was booked, whether the booking was a yellow card or a red card, whether the card was a second yellow card, and whether the player was sent off as a result of the booking. There are 26 variables and 3178 observations.

Variables

`key_id`	`integer` The unique ID number for the observation.
`booking_id`	`text` The unique ID number for the booking. Has the format `B-####`, where the number is a counter that is assigned with the data sorted by the match ID, then the minute of the booking.
`tournament_id`	`text` The unique ID number for the tournament. References `tournament_id` in the `tournaments` dataset.
`tournament_name`	`text` The name of the tournament.
`match_id`	`text` The unique ID number for the match in which the booking occurred. References `match_id` in the `matches` dataset.
`match_name`	`text` The name of the match in which the booking occurred.

`match_date`	`date` The date of the match in the format `YYYY-MM-DD`.
`stage_name`	`enum` The stage of the tournament in which the match occurred. The possible values are: `first round`, `second round`, `group stage`, `round of sixteen`, `quarter-finals`, `semi-finals`, `third place match`, `final`. Note that not all values are applicable to all tournaments.
`group_name`	`text` The name of the group.
`team_id`	`text` The unique ID number for the team of the player who was booked. References `team_id` in the `teams` dataset.
`team_name`	`text` The name of the team of the player who was booked.
`team_code`	`text` The 3-letter code for the team of the player who was booked.
`home_team`	`boolean` Whether the team was the home team. Coded `1` if the team was the home team and `0` otherwise.
`away_team`	`boolean` Whether the team was the away team. Coded `1` if the team was the away team and `0` otherwise.
`player_id`	`text` The unique ID number for the player who was booked. References `player_id` in the `players` dataset.
`family_name`	`text` The family name of the player who was booked.
`given_name`	`text` The given name of the player who was booked.
`shirt_number`	`integer` The shirt number of the player who was booked.
`minute_label`	`text` The minute of the match in which the booking occurred in the format `#'` or `#'+#'`.
`minute_regulation`	`integer` The minute of regulation time in which the booking occurred.

`minute_stoppage`	`integer` The minute of stoppage time in which the booking occurred. Coded `0` if the substitution did not occur during stoppage time.
`match_period`	`enum` The period of the match in which the booking occurred. The possible values are: `first half`, `first half, stoppage time`, `second half`, `second half, stoppage time`, `extra time, first half`, `extra time, first half, stoppage time`, `extra time, second half`, `extra time, second half, stoppage time`, `after extra time`.
`yellow_card`	`boolean` Whether the booking was a yellow card. Coded `1` if the card is a yellow card and `0` otherwise.
`red_card`	`boolean` Whether the booking was a red card. Coded `1` if the card is a red card and `0` otherwise.
`second_yellow_card`	`boolean` Whether the booking was a second yellow card. Coded `1` if the booking is a second yellow and `0` otherwise.
`sending_off`	`boolean` Whether the booking resulted in the player being sent off. Coded `1` if the player was sent off and `0` otherwise.

substitutions

Substitutions

Description

This dataset records all substitutions since 1970. FIFA match reports do not include information about substitutions before 1970. There is one observation per player per substitution. It indicates the minute of the substitution, the player who went off, and the player who came on. There are 24 variables and 10222 observations.

Variables

`key_id`	`integer` The unique ID number for the observation.
`substitution_id`	`text` The unique ID number for the substitution. Has the format `S-####`, where the number is a counter that is assigned with the data sorted by the match ID, then the minute of the substitution, then whether the player is going off.
`tournament_id`	`text` The unique ID number for the tournament. References `tournament_id` in the `tournaments` dataset.
`tournament_name`	`text` The name of the tournament.
`match_id`	`text` The unique ID number for the match in which the substitution occurred. References `match_id` in the `matches` dataset.
`match_name`	`text` The name of the match in which the substitution occurred.

`match_date`	`date` The date of the match in the format `YYYY-MM-DD`.
`stage_name`	`enum` The stage of the tournament in which the match occurred. The possible values are: `first round`, `second round`, `group stage`, `round of sixteen`, `quarter-finals`, `semi-finals`, `third place match`, `final`. Note that not all values are applicable to all tournaments.
`group_name`	`text` The name of the group.
`team_id`	`text` The unique ID number for the team of the player who was substituted. References `team_id` in the `teams` dataset.
`team_name`	`text` The name of the team of the player who was substituted.
`team_code`	`text` The 3-letter code for the team of the player who was substituted.
`home_team`	`boolean` Whether the team was the home team. Coded `1` if the team was the home team and `0` otherwise.
`away_team`	`boolean` Whether the team was the away team. Coded `1` if the team was the away team and `0` otherwise.
`player_id`	`text` The unique ID number for the player who was substituted. References `player_id` in the `players` dataset.
`family_name`	`text` The family name of the player who was substituted.
`given_name`	`text` The given name of the player who was substituted.
`shirt_number`	`integer` The shirt number of the player who was substituted.
`minute_label`	`text` The minute of the match in which the substitution occurred in the format `#'` or `#'+#'`.
`minute_regulation`	`integer` The minute of regulation time in which the substitution occurred.

`minute_stoppage`	`integer` The minute of stoppage time in which the substitution occurred. Coded `0` if the substitution did not occur during stoppage time.
`match_period`	`enum` The period of the match in which the substitution occurred. The possible values are: `first half`, `first half, stoppage time`, `second half`, `second half, stoppage time`, `extra time, first half`, `extra time, first half, stoppage time`, `extra time, second half`, `extra time, second half, stoppage time`, `after extra time`.
`going_off`	`boolean` Whether the player was going off the field. Coded `1` if the player was going off and `0` otherwise.
`coming_on`	`boolean` Whether the player was coming on the field. Coded `1` if the player was coming on and `0` otherwise.

host_countries

Host countries

Description

This dataset records all host countries. There is one observation per host country per tournament. A tournament can have multiple host countries. It indicates the performance of each host country (the furthest stage reached). There are 7 variables and 31 observations.

Variables

`key_id`	`integer` The unique ID number for the observation.
`tournament_id`	`text` The unique ID number for the tournament. References `tournament_id` in the `tournaments` dataset.
`tournament_name`	`text` The name of the tournament.
`team_id`	`text` The unique ID number for the team. References `team_id` in the `teams` dataset.
`team_name`	`text` The name of the team.
`team_code`	`text` The 3-letter code for the team.
`performance`	`text` The furthest stage of the tournament reached by the host country's team.

tournament_stages

Tournament stages

Description

This dataset records the stages in each tournament. There is one observation per stage per tournament. It indicates the name of the stage, whether the stage was a group stage or a knockout stage, if the stage was a group stage, whether there were unbalanced groups, the start and end dates of the stage, and how many matches there were in the stage, how many teams participated in each stage, how many games were scheduled, how many replays there were, how many walkovers there were, and how many playoffs there were. There are 16 observations and 155 observations.

Variables

`key_id`	`integer` The unique ID number for the observation.
`tournament_id`	`text` The unique ID number for the tournament. References `tournament_id` in the `tournaments` dataset.
`tournament_name`	`text` The name of the tournament.
`stage_number`	`integer` The number of the stage.
`stage_name`	`enum` The stage of the tournament in which the match occurred. The possible values are: `first round`, `second round`, `group stage`, `round of sixteen`, `quarter-finals`, `semi-finals`, `third place match`, `final`. Note that not all values are applicable to all tournaments.

`group_stage`	`boolean` Whether the match is a group stage match. Coded `1` if the match is a group stage match and `0` otherwise.
`knockout_stage`	`boolean` Whether there was a knockout stage. Coded `1` if there was a knockout stage and `0` otherwise.
`unbalanced_groups`	`boolean` Whether there were unbalanced groups. Coded `1` if there were unbalanced groups and `0` otherwise.
`start_date`	`date` The start date of the stage in the format `YYYY-MM-DD`.
`end_date`	`date` The end date of the stage in the format `YYYY-MM-DD`.
`count_matches`	`integer` The number of matches in the stage.
`count_teams`	`integer` The number of teams that participated in the stage.
`count_scheduled`	`integer` The number of games that were scheduled in the stage.
`count_replays`	`integer` The number of replays in the stage.
`count_playoffs`	`integer` The number of playoff games in the stage.
`count_walkovers`	`integer` The number of walkovers in the stage.

groups

Groups

Description

This dataset records the names of the groups for each group stage. There is one observation per group per group stage per tournament. Some tournaments have multiple group stages. It indicates the stage, the name of the group, and how many teams were in the group. There are 7 variables and 117 observations.

Variables

`key_id`	`integer` The unique ID number for the observation.
`tournament_id`	`text` The unique ID number for the tournament. References `tournament_id` in the `tournaments` dataset.
`tournament_name`	`text` The name of the tournament.
`stage_number`	`integer` The number of the stage.
`stage_name`	`enum` The stage of the tournament in which the match occurred. The possible values are: `first round`, `second round`, `group stage`, `round of sixteen`, `quarter-finals`, `semi-finals`, `third place match`, `final`. Note that not all values are applicable to all tournaments.
`group_name`	`text` The name of the group.

`count_teams`	`integer`	The number of teams in the group.

group_standings

Group standings

Description

This dataset records group standings for each group stage. There is one observation per team per group per group stage per tournament. Some tournaments have multiple group stages. It includes the final position of the team (factoring in tie breakers), the name of the team, the number of matches played, the number of wins, the number of losses, the number of draws, the number of goals for, the number of goals against, the goal difference, the total number of points earned, and whether the team advanced out of the group. There are 19 variables and 626 observations.

Variables

`key_id`	`integer` The unique ID number for the observation.
`tournament_id`	`text` The unique ID number for the tournament. References `tournament_id` in the `tournaments` dataset.
`tournament_name`	`text` The name of the tournament.
`stage_number`	`integer` The number of the stage.
`stage_name`	`enum` The stage of the tournament in which the match occurred. The possible values are: `first round`, `second round`, `group stage`, `round of sixteen`, `quarter-finals`, `semi-finals`, `third place match`, `final`. Note that not all values are applicable to all tournaments.

`group_name`	`text` The name of the group.
`position`	`integer` The team's final position in the group.
`team_id`	`text` The unique ID number for the team. References `team_id` in the `teams` dataset.
`team_name`	`text` The name of the team.
`team_code`	`text` The 3-letter code for the team.
`played`	`integer` The number of matches that the team played in the group.
`wins`	`integer` The number of matches that the team won in the group stage.
`draws`	`integer` The number of matches that the team drew in the group stage.
`losses`	`integer` The number of matches that the team lost in the group stage.
`goals_for`	`integer` The number of goals scored by the team in the group stage.
`goals_against`	`integer` The number of goals scored against the team in the group stage.
`goal_difference`	`integer` The team's goal difference in the group stage.
`points`	`integer` The number of points that the team earned in the group.
`advanced`	`boolean` Whether the team advanced out of the group. Coded `1` if the team advanced and `0` otherwise.

tournament_standings

Tournament standings

Description

This dataset records the final standings for each tournament. There is one observation per position per tournament. The top four teams are ranked. In most tournaments, these are the winner of the final, the loser of the final, the winner of the third-place match, and the loser of the third-place match. There are 7 variables and 120 observations.

Variables

`key_id`	`integer` The unique ID number for the observation.
`tournament_id`	`text` The unique ID number for the tournament. References `tournament_id` in the `tournaments` dataset.
`tournament_name`	`text` The name of the tournament.
`position`	`integer` The place of the team in the final standings.
`team_id`	`text` The unique ID number for the team. References `team_id` in the `teams` dataset.
`team_name`	`text` The name of the team.
`team_code`	`text` The 3-letter code for the team.

award_winners

Award winners

Description

This dataset records all award winners. There is one observation per player per award per tournament. Some awards are shared by multiple players. It indicates the name of the award, the player(s) who won the award, the team of the player(s) who won the award, and whether the award was shared. There are 12 variables and 200 observations.

Variables

`key_id`	`integer` The unique ID number for the observation.
`tournament_id`	`text` The unique ID number for the tournament. References `tournament_id` in the `tournaments` dataset.
`tournament_name`	`text` The name of the tournament.
`award_id`	`text` The unique ID number for the award. References `award_id` in the `awards` dataset.
`award_name`	`enum` The name of the award. The possible values are: `Golden Ball`, `Silver Ball`, `Bronze Ball`, `Golden Boot`, `Silver Boot`, `Bronze Boot`, `Golden Glove`, `Best Young Player`.
`shared`	`boolean` Whether the award was shared between multiple players. Coded `1` if the award was shared and `0` otherwise.

`player_id`	`text` The unique ID number for the player who won the award. References `player_id` in the `players` dataset.
`family_name`	`text` The family name of the player who won the award.
`given_name`	`text` The given name of the player who won the award.
`team_id`	`text` The unique ID number for the team. References `team_id` in the `teams` dataset.
`team_name`	`text` The name of the team of the player who won the award.
`team_code`	`text` The 3-letter code for the team of the player who won the award.