Codebook v 1.0

The Fjelstul English Football Database

Joshua C. Fjelstul, Ph.D.

Datasets

seasons

Description

This dataset records all seasons in the English Football League and the Premier League from inaugural season of the Football League (1888-89) through the most recent season (2021-22). There is one observation per season. It indicates the tier, division, subdivision, winner, and number of teams for each season.

Variables

`key_id`   `integer` The unique ID number for the observation.

`season_id`   `text` The unique ID number for the season. Has the format `S-####-#`, where the first number is the year in which the season started, the second number is the tier. In the third tier, from the 1921-22 season through the 1957-58 season, there were North and South subdivisions. These are indicated by a `-N` or `-S` suffix.

`season`   `integer` The year that the season started.

`tier`   `integer` The tier in English football. The possible values are `1` through `4`.

`division`   `text` The division in English football. For the current league structure, the possible values are `Premier League`, `Championship`, `League One`, and `League Two`. For previous

league structures, the possible values are `First Division`, `Second Division`, `Third Division`, and `Fourth Division`.

`subdivision` `text` The subdivision in English football. In the third tier, from the 1921-22 season through the 1957-58 season, there were North and South subdivisions. The possible values are `North`, `South`, and `None`.

`winner` `text` The name of the team that won the league.

`count_teams` `integer` The number of teams in the league (that actually played a match).

teams

Teams

Description

This dataset records all teams who have competed in the English Football League and the Premier League. There is one observation per team. It indicates the current name of the team, any former names of the team, whether the team is a former member of the Football League, whether the team is defunct, and the season that the team made their first appearance in the Football League.

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 appearance in the Football League and then by the team's name.

`team_name`   `text` The current name of the team.

`former_team_names`   `text` The former names of the team, separated by a comma. Coded `None` for teams that have not changed their name.

`current`   `boolean` Whether the team currently competes in the English Football League or the Premier League. Coded `1` if the team currently competes in these leagues and `0` otherwise.

`former`	`boolean`	Whether the team no longer competes in the English Football League or the Premier League (but still exists). Coded `1` if the team is no longer in these leagues and `0` otherwise.
`defunct`	`boolean`	Whether the team is defunct and no longer exists. Coded `1` if the team is defunct and `0` otherwise.
`first_appearance`	`integer`	The season that the team first competed in the Football League.

matches

Matches

Description

This dataset records all matches that have ever been played in the English Football League and the Premier League (1888-2022). There is one observation per match per season. It indicates the season, tier, division, and subdivision for the match, the score, the score margin for each team, and the result of the match (home team win, away team win, draw).

Variables

`key_id`   `integer` The unique ID number for the observation.

`season_id`   `text` The unique ID number for the season. References `season_id` in the `seasons` dataset.

`season`   `integer` The year that the season started.

`tier`   `integer` The tier in English football. The possible values are `1` through `4`.

`division`   `test` The division in English football. For the current league structure, the possible values are `Premier League`, `Championship`, `League One`, and `League Two`. For previous league structures, the possible values are `First Division`, `Second Division`, `Third Division`, and `Fourth Division`.

`subdivision`	`test`	The subdivision in English football. In the third tier, from the 1921-22 season through the 1957-58 season, there were North and South subdivisions. The possible values are `North`, `South`, and `None`.
`match_id`	`text`	The unique ID number for the match. Has the format `M-####-###`, where the first number is the season and second number is a within-season counter that is assigned with the data sorted by the name of the home team, then by the name of the away team.
`match_name`	`text`	The name of the match.
`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.
`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.
`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.
`result`	`enum`	The result of the match. The possible values are `home team win`, `away team win`, and `draw`.
`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.

appearances

Appearances

Description

This dataset records all appearances. There is one observation per team per match per season. It indicates whether the team is the home team or the away team, the number of goals for and against, the goal difference, whether the team wins, loses, or draws, and how many points the team earned from the match.

Variables

key_id   `integer` The unique ID number for the observation.

season_id   `text` The unique ID number for the season. References `season_id` in the `seasons` dataset.

season   `integer` The year that the season started.

tier   `integer` The tier in English football. The possible values are `1` through `4`.

division   `test` The division in English football. For the current league structure, the possible values are `Premier League`, `Championship`, `League One`, and `League Two`. For previous league structures, the possible values are `First Division`, `Second Division`, `Third Division`, and `Fourth Division`.

subdivision	`test`	The subdivision in English football. In the third tier, from the 1921-22 season through the 1957-58 season, there were North and South subdivisions. The possible values are `North`, `South`, and `None`.
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.
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.
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.
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_difference	`integer`	The team's goal difference.
result	`enum`	The result of the match. The possible values are `home team win`, `away team win`, and `draw`.
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.

`points` `integer` The number of points the team earned from the match. A team earns `0` points for a loss and `1` point for a draw. From the 1888-89 season through the 1980-81 season, teams earned `2` points for a win. Starting with the 1981-82 season, teams have earned `3` points for a win.

standings

Standings

Description

This dataset records all end-of-the-season standings. There is one observation per team per season. It indicates the final position of the team (accounting for 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, and the total number of points earned.

Variables

`key_id`　　　　　　　　　`integer`　The unique ID number for the observation.

`season_id`　　　　　　　`text`　The unique ID number for the season. References `season_id` in the `seasons` dataset.

`season`　　　　　　　　　`integer`　The year that the season started.

`tier`　　　　　　　　　　`integer`　The tier in English football. The possible values are `1` through `4`.

`division`　　　　　　　　`test`　The division in English football. For the current league structure, the possible values are `Premier League`, `Championship`, `League One`, and `League Two`. For previous league structures, the possible values are `First Division`, `Second Division`, `Third Division`, and `Fourth Division`.

`subdivision`	`test` The subdivision in English football. In the third tier, from the 1921-22 season through the 1957-58 season, there were North and South subdivisions. The possible values are `North`, `South`, and `None`.
`position`	`integer` The team's final position in the league.
`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.
`played`	`integer` The number of matches that the team played.
`wins`	`integer` The number of matches that the team won.
`draws`	`integer` The number of matches that the team drew.
`losses`	`integer` The number of matches that the team lost.
`goals_for`	`integer` The number of goals scored by the team.
`goals_against`	`integer` The number of goals scored against the team.
`goal_difference`	`integer` The team's goal difference.
`points`	`integer` The number of points that the team earned over the whole season (after any point adjustments).
`point_adjustment`	`integer` The number of points that were deducted by the league due to violations of league rules or added by the league due to forfeits.