data dictionary

Smith College NFL Team

4/2/2020

full_data_with_thurs

Description of dataset:

This dataset combines `games` and `pbp` using a `left_join` of `games` with `pbp` by `game_key`. Both `games` and `pbp` have been cleaned using `janitor::clean_names()`. This dataset also contains a binary variable, `is_thurs` that indicates whether or not the game occured on a thursday. Note that for the years 2007 - 2010, many columns include only NA values. This is because the `pbp` data.frame did not include data for these years.

Variables

- `game_key`: (factor) *unique code corresponding to a game*
- `season`: (factor) *season during which the game took place. Denoted by year (2007, 2008 ... 2019).*
- `season_type` : (factor) *representing the type of season. In the case of this dataset, all games are "regular"*
- `game_number` : (integer) *number of game in the season. There are 256 total games in one season*
- `week` : (factor) *numerical value for week within the season. There are a max of 17 weeks in 1 season*
- `game_date` : (Date) *date of game*
- `game_day` : (factor) *day of the week during which the game took place*
- `game_site` : (factor) *site of game. May be city or name of home team*
- `start_time` : (factor) *time of the start of the game*
- `home_team` : (character) *name of home team*
- `visit_team`: (character) *name of visiting team*
- `stadium` : (character) *name of stadium at which game was played*
- `location` : (character) *location of the game in the form, "city, state"*
- `stadium_type` : (factor) *type of stadium at which game was played. Examples include "Dome, Retractable Roof.."*
- `turf` : (factor) *type of surface (i.e. turf) on which game was played. (ex. natural grass, field turf)*
- `time_zone` : (factor) *time zone of the location of the game*

- `game_length` : (factor) *length of game in hours and minutes*
- `game_weather`: (character) *description of weather on the day of the game*
- `temperature` : (integer) *temperature at game time*
- `humidity` : (integer) *humidity at game time (as a percent)*
- `wind_speed` : (character) *description of wind speed at game time*
- `wind_direction` : (character) *description of direction of wind*
- `wind_chill` : (factor) *description of windchill at start of game. Usually denoted by temperature*
- `play_id` : (integer) *identifies play within each game (unique for each play). It is roughly based on the time in the game, but this is not exact. Ordered from the beginning of the game until the end of the game*
- `home_club_code` : (factor) *code for the home team. Usually an abbreviation of the city or state of the team*
- `visitor_club_code` : (factor) *code for the visiting team. Usually an abbreviation of the city or state of the team*
- `quarter` : (factor) *period of the game. A typical game is 4 quarters, overtime consistutes a 5th*
- `possession_team` : (factor) *team with the ball on the play of interest*
- `play_description` : (character) *a description of the play, includes important players and play-specific details*
- `game_clock` : (integer) *time on clock in seconds*
- `down` : (factor) *down of the play. A down represents a team's attempt to make it a specified amount of yardage. Teams usually get 4 downs to make make it*
- `distance` : (integer) *in yards*
- `field_position` : (factor) *abbreviation for field position, usually associated with team code*
- `yard_line` : (integer) *marking on field where play begins*
- `pass_result` : (factor) *result of pass, C = complete, I = incomplete, IN = interception, 2 other unidentified levels*
- `play_result`: (integer) *yardage gained or lost on the play*
- `special_teams_play_type` : (factor) *if special teams play, the type of team on the field*
- `home_score_before_play` : (integer) *Score variable that changes within a game*
- `visitor_score_before_play` : (integer) *Score variable that changes within a game*
- `home_score_after_play` : (integer) *Score variable that changes within a game*
- `visitor_score_after_play` : (integer) *Score variable that changes within a game*
- `offense_play_result` : (integer) *yardage gained or lost after an offensive play*
- `play_nullified_by_penalty` : (factor) *if a play result is cancelled due to a penalty (Y = Yes or N = No)*
- `kick_return_yardage` : (integer) *yardage gained or lost on a kick return*

- `kick_return_yardage_no_penalties` : (integer) *yardage gained or lost on a kick return with no penalties called on the play*
- `pass_length` : (integer) *length of pass attempted in yards*
- `yards_after_catch` : (integer) *number of yards gained or lost after receiver makes the catch*
- `kickoff_length` : (integer) *number of yards of the kickoff*
- `kickoff_result` : (factor) *result of the play*
- `wp_home` : (integer) *win probability for the home team at the start of the play. This uses an in-house formula that accounts for several of the characteristics about the plays and games*
- `delta_wp_home` : (integer) *change in win probability for the home team on the play of interest*
- `ep` : (integer) *expected points for the possesion team on the play*
- `timo` : (factor) *timeouts reamining for the offense*
- `timd` : (factor) *timeouts remaining for the defense*
- `absolute_yard_line` : (integer) *represents how far a time is from its own end zone
- `is_goal_to_go` : (integer) *is it a "goal to go" situation
- `is_scoreing_play` : (factor) *binary variable representing whether or not the play is a scoring play*
- `is_thurs` : (factor) *binary variable representing whether or not the game occured on a thursday. If thursday, is_thurs = 1*