Package 'NBADP'

May 4, 2023

Type Package

Title NBA Data Processing

Version 0.1.0

Author Yi LU <121090386@link.cuhk.edu.cn>, Wei SHI <120040085@link.cuhk.edu.cn>, Yutong LIU <120040056@link.cuhk.edu.cn>

Maintainer Yi LU <121090386@link.cuhk.edu.cn>

Description A package to predict and visualize corresponding full reflection of the players' performance in games by analyzing NBA data. The data is provided, tell 'NBADP' to reflect the players' performance. Then we can give some prediction or guidance for the team. Potential Function: 1 predict the probability of championship of each team this year by regression model, according to the data of previous years(regression); 2 predict the members of All-NBA first/ second/third team this year according to the personal data (PPG, RPG, APG, SPG, BPG. . .) (classification); 3 predict the DPOY of this year(classification); 4 analyze and visualize the personal state's change of Kawhi Leonard, from 2022 October to now, according to the personal data and his team's achievements(regression); 5 calculate player efficiency rating (PER) to collect or boil down all of a player's contributions into a number to rate every player's statistical performance.; 6 predict player's one year approximate salary that would earn in the NBA at his current level of ability according to the past players' data and the current system. (regression); 7 analyze and visualize the field goal percentage of the player(visualization); 8 visualize the 2 PT percentage and 3 PT percentage of all the players.(visualization); 9 analyze the best performance player in international countries in terms of the personal data (PPG, RPG, APG, SPG, BPG. . .) (classification); 10 analyze the top-3 players of the best cost–performance ratio (regression); 11 help give the starter's plan of Losangles Clippers for the coach according to the personal data of each player(linear programming)

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

Imports ggplot2, readr, dplyr

Depends R (>= 2.10)

Suggests roxygen2, devtools, knitr, rmarkdown, testthat (>= 3.0.0)

VignetteBuilder knitr

Config/testthat/edition 3

R topics documented:

---

best_nba_player   *Find the best NBA players based on their salaries and score stats*

Description

Find the best NBA players based on their salaries and score stats

Usage

```
best_nba_player(top, year)
```

Arguments

top	Integer, input the number of top NBA players
year	Integer, input the year to consider for score stats and salaries

Value

A dataframe of the top NBA players based on their salaries and score stats

Examples

```
best_nba_player(top = 4, year = 2021)
```

DataSalariesRaw *NBA_Salaries_1990_2023_*

Description

A dataset containing NBA players' salaries from 1990 to 2023. The variables are as follows:

Usage

DataSalariesRaw

Format

A data frame with 15857 rows and 5 variables:

...1 variable ranking

playerName player name

seasonStartYear the starting year of an NBA season

salary the actual salary of an NBA player

inflationAdjSalary the salary adjusted for inflation

DataScoreRaw *NBA_Player_Box_Score_Stats_1950_2022_*

Description

A dataset collected from box scores, containing detailed statistics on individual player performances in NBA games from 1950 to 2022. The variables are as follows:

Usage

DataScoreRaw

Format

A data frame with 25080 rows and 29 variables:

...1 variable ranking

Season the season in which the game was played

Game_ID a unique identifier for each game

PLAYER_NAME the name of the player

Team the team the player played for

GAME_DATE the date the game was played

MATCHUP the matchup of the game

WL the outcome of the game for the player's team

MIN the number of minutes the player played in the game

FGM the number of field goals made by the player

FGA the number of field goals attempted by the player

FG_PCT the player's field goal percentage (FGM/FGA)

FG3M the number of three-point field goals made by the player

FG3A the number of three-point field goals attempted by the player

FG3_PCT the player's three-point field goal percentage (FG3M/FG3A)

FTM the number of free throws made by the player

FTA the number of free throws attempted by the player

FT_PCT the player's free throws attempted by the player

OREB the number of offensive rebounds the player had in the game

DREB the numbver of defensive rebounds the player had in the game

REB the total number of rebounds the player had in the game

AST the number of assists the player had in the game

STL the number of steals the player had in the game

BLK the number of blocks the player had in the game

TOV the number of turnovers the player had in the game

PF the number of personal fouls the player scored in the game

PTS the total number of points the player scored in the game

PLUS_MINUS the point differential when the player was on the court

VIDEO_AVAILABLE a flag indicating whether or not video of the game is available for analysis

---

datasets	*NBA contract predict model datasets*

Description

A dataset containing the players' 2022-2023 salary and 2021-2022 NBA performance data. Also, those players are not have accurate contract for next season yet. The variables are as follows:

Usage

```
datasets
```

Format

A data frame with 128 rows and 51 variables:

Player player name

salarya contract's salary in 2022-2023

Guaranteed the amount of a player's remaining salary that is guaranteed

Pos position

Age player's age

G games

MP minutes played

PER player efficiency rating: a measure of per-minute production standardized such that the league average is 15 (-45.2–76.2)

TSp true shooting percentage: a measure of shooting efficiency that takes into account 2-point field goals, 3-point field goals, and free throw

X3PAr 3-point attempt rate: percentage of FG attempts from 3-point range

FTr Free Throw attempt rate: number of FT attempts per FG attempt

ORBp Offensive Rebound Percentage: an estimate of the percentage of available offensive rebounds a player grabbed while they were on the floor

DRBp Defensive Rebound Percentage: an estimate of the percentage of available defensive rebounds a player grabbed while they were on the floor

TRBp Total Rebound Percentage: an estimate of the percentage of available rebounds a player grabbed while they were on the floor

ASTp Assist Percentage: an estimate of the percentage of teammate field goals a player assisted while they were on the floor

STLp Steal Percentage: an estimate of the percentage of opponent possessions that end with a steal by the player while they were on the floor

BLKp Block Percentage: an estimate of the percentage of opponent two-point field goal attempts blocked by the player while they were on the floor

TOVp Turnover Percentage: an estimate of turnovers committed per 100 plays

USGp Usage Percentage: an estimate of the percentage of team plays used by a player while they were on the floor

OWS Offensive Win Shares: an estimate of the number of wins contributed by a player due to offense

DWS Defensive Win Shares: an estimate of the number of wins contributed by a player due to defense

WS Win Shares: an estimate of the number of wins contributed by a player

WSd48 Win Shares Per 48 Minutesan estimate of the number of wins contributed by a player per 48 minutes (league average is approximately .100)

OBPM Offensive Box Plus/Minus: a box score estimate of the offensive points per 100 possessions a player contributed above a league-average player, translated to an average team

DBPM Defensive Box Plus/Minus: a box score estimate of the defensive points per 100 possessions a player contributed above a league-average player, translated to an average team

BPM Box Plus/Minus: a box score estimate of the points per 100 possessions a player contributed above a league-average player, translated to an average team

VORP Value over Replacement Player: a box score estimate of the points per 100 Team possessions that a player contributed above a replacement-level (-2.0) player, translated to an average team and prorated to an 82-game season

GS game started

FG field goals per game

FGA field goal attempts per game

FGp field goal percentage

X3P 3-point field goals per game

X3PA 3-point field goal attempts per game

X3Pp 3-point field goal percentage

X2P 2-point field goals per game

X2PA 2-point field goal attempts per game

X2Pp 2-point field goal percentage

eFGp effective field goal percentage: This statistic adjusts for the fact that a 3-point field goal is worth one more point that a 2-point field goal

FT free throws per game

FTA free throw attempts per game

FTp free throw percentage

ORB offensive rebounds per game

DRB defensive rebounds per game

TRB total rebounds per game

AST assists per game

STL steals per game

BLK blocks per game

TOV turnovers per game

PF personal fouls per game

PTS points per game

---

field_nba_player	*Plot the field goal percentage (FGP) of NBA players for a given year and number of top players to display.*

Description

Plot the field goal percentage (FGP) of NBA players for a given year and number of top players to display.

Usage

```
field_nba_player(index = 30, year = number, desc = T)
```

Arguments

index	Integer, indicating the number of top players to display
year	Integer, indicating the year of the NBA season to plot FGP for
desc	Logical value, indicating whether to sort the players in descending order by FGP and games played (G)

Value

A plot of the top players with their FGP and G displayed

Examples

```
field_nba_player(index = 30, year = 2021, desc = TRUE)
```

Description

A dataset containing GSW competition performance on 2022-2023 Season. The variables are as follows:

Usage

GSW

Format

A data frame with 40112 rows and 19 variables:

game_id Game ID

period Period of Play

clock Game-clock at time of play

home home team

scoreHome Home team points at time of play

away Away team

scoreAway Away team points at time of play

playerNameI Player making the play

teamTricode Team of the player making the play

description The description of the play

actionType What kind of play

subType A more detailed description of an actionType

xLegacy The horizontal position of a player's shot

yLegacy The vertical position of a player's shot

shotDistance The distance from the basket of a field goal

isFieldGoal Whether the play is or isn't a field goal

shotVal The number of points that could be scored from a play

scoreVal The number of points that were scored on a play

location Which side of the court the play occurred

Kawhi.img   *An image of Kawhi Leonard*

Description

An image of Kawhi Leonard

Usage

```
Kawhi.img
```

Format

a png file

Kawhi.Leonard.state.analysis

analysis of state fluctuation of a superstar player: Kawhi Leonard

---

Description

analysis of state fluctuation of a superstar player: Kawhi Leonard

Usage

```
Kawhi.Leonard.state.analysis(
  data.category = c("MIN", "PTS", "REB", "FGM"),
 month.range = c("October", "November", "December", "January", "February", "March",
    "April")
)
```

Arguments

data.category	the data category that we want to their changes, they are set as "MIN","PTS","REB","FGM", "REB" as the default However, you can choose any data category in the data set. They include "GP W L MIN PTS FGM FGA FG% 3PM 3PA 3P% FTM FTA FT% OREB DREB REB AST TOV STL BLK PF FP DD2 TD3 BPM"
month.range	the month range of Kawhi Leonard. They are required to be a subset of ("October","November","December","January","February","March","April") with the same order.

Value

the state change of the player, as a form of data frame, column names are the month names, from October to April row names are the average points, rebounds, assists, steals, blocks, turnovers of that player

Examples

```
library(NBADP)
Kawhi.Leonard.state.analysis(data.category=c("GP","STL","BLK","BPM"),
month.range=c("December","January","February","March"))
```

---

lg | *NBA League Average data in each year*

Description

A dataset containing the league average level in many different variables. The variables are as follows:

Usage

    lg

Format

A data frame with 77 rows and 32 variables:

Rk time ranking (1–77)

Season the year of NBA season (1946-47–2022-23)

Lg name of league (NBA)

Age the average age of the league in each year (26.1–27.9)

Ht average height of the league in each year (6-4–6-7)

Wt average weight of the league in each year (195–223)

G average number of games(192–1230)

MP minutes played per game (240.5–242.2)

FG field goals per game (25.9–45.9)

FGA field goal attempts per game (75.4–109.4)

3P 3-point field goals per game (0.5–12.7)

3PA 3-point field goal attempts per game (2.0–35.2)

FT free throws per game (15.9–28.6)

FTA free throw attempts per game (21.7–38.3)

ORB offensive rebounds per game (9.7–15.1)

DRB defensive rebounds per game (28.4–34.8)

TRB total rebounds per game (41.0–73.5)

AST assists per game (6.9–26.3)

STL steals per game (7.2–9.6)

BLK blocks per game (4.3–5.6)

TOV turnovers per game (13.8–20.8)

PF personal fouls per game (19.3–28.8)

PTS points per game (67.8–118.8)

FG% field goals percentage (0.279–0.492)

3P% 3-point field goals percentage (0.238–0.367)

FT% free throw percentage (0.641–0.782)

Pace pace factor: an estimate of possessions per 48 minutes (88.9–107.8)

eFG% effective field goal percentage: This statistic adjusts for the fact that a 3-point field goal is worth one more point that a 2-point field goal (0.279–0.545)

TOV% turnover percentage: an estimate of turnovers committed per 100 plays (12.3–16.5)

ORB% offensive rebound percentage: an estimate of the percentage of available offensive rebounds a player grabbed while they were on the floor (22.2–33.5)

FT/FGA free throws per field goal attempt (0.171–0.333)

ORtg offensive rating: an estimate of points produced (players) or scored (teams) per 100 possessions (97.7–114.8)

---

nba_win_rate   *Predict NBA Team Win Rate*

Description

Predict NBA Team Win Rate

Usage

```
nba_win_rate(year = 2023, desc = T)
```

Arguments

year	Integer, the year to predict the win rate for, defaults to 2023
desc	Logical value, Whether to sort the output by descending predicted win rate

Value

A bar chart that shows the predicted win rate and the historic win rate for each team

Examples

```
nba_win_rate(year = 2023)
```

PER   *Player Efficiency Rate*

Description

We calculate player efficiency rating (PER) to boil down a player's performance.

Usage

```
PER(lgdata = lg, regulardata = X2021to2022data, yrs = "2021-22", player1, team)
```

Arguments

lgdata	dataset, a dataset that contain the variables that have league's basic information
regulardata	dataset, a dataset that contain the variables that have basic information of player's performance
yrs	string, the seasons that we want to calculate
player1	string, input the name of the player that we want to test
team	string, input the team that the player on

Value

a number

Examples

```
library(NBADP)
PER(lgdata = lg, regulardata = X2021to2022data, yrs = "2021-22", "LeBron James","LAL")
```

---

player_visual	*Visualize a player's shot percentage and distribution on a season*

Description

Visualize a player's shot percentage and distribution on a season

Usage

```
player_visual(teamdata = GSW, playerName, teamName)
```

Arguments

teamdata	dataset, a dataset that contain the team's all regular match data in 2022-2023 season
playerName	string, input the player's name
teamName	string, input the team that the player on

Value

The shot plot of the player in the whole regular season

Examples

```
library(NBADP)
player_visual(teamdata = GSW,"S. Curry","GSW")
```

---

pred_salary   *Make a salary forecast for 2023-2024*

Description

Make a salary forecast for 2023-2024

Usage

```
pred_salary(
  turn_over,
  points_per_game,
  two_point_field_goal_attempts_per_game,
  free_throws_per_game,
  assists_per_game,
  games_started,
  value_over_replacement_player,
  defensive_rebounds_per_game,
  minutes_played_per_game,
  three_point_field_goal_attempts_per_game,
  usage_percentage,
  offensive_box_plusorminus,
  defensive_win_shares
)
```

Arguments

turn_over   numeric, turnovers per game (TOV).

points_per_game

                  numeric, points per game (PTS).

two_point_field_goal_attempts_per_game

                  numeric, (2PA).

free_throws_per_game

                  numeric, free throws per game (FT).

assists_per_game

                  numeric, assists per game (AST).

games_started   numeric, games started (GS).

value_over_replacement_player

                  numeric, value over replacement player (VORP).

defensive_rebounds_per_game

                  numeric, defensive rebounds per game (DRB).

minutes_played_per_game

                  numeric, minutes played per game (MP).

three_point_field_goal_attempts_per_game

                  numeric, three point field goal attempts per game (3PA).

usage_percentage

                  numeric, usage percentage (USG%).

offensive_box_plusorminus

                  numeric, offensive box plus or minus (OBPM).

defensive_win_shares

                  numeric, defensive win shares (DWS).

Value

according to the player's performance in 2023-2024 to predict 2023-2024 salary

Examples

```
library(NBADP)
pred_salary(0.6, 3.9, 2.1, 0.4, 1.4, 6, 0, 0.8, 13.4, 1.4, 13.4, -2.9, 0.6)
```

---

shot   *NBA Shots Dataset(20211205)*

Description

A dataset containing the competition performance on date 2021/12/05. The variables are as follows:

Usage

shot

Format

A data frame with 690 rows and 16 variables:

...1 variable ranking

Unnamed: 0 variable ranking

Unnamed: 0.1 variable ranking

match_id game identifier. Unique for each matchup

shotX x coordinate of the shot position

shotY y coordinate of the shot position

quarter quarter of the game

time_remaining time remaining in the quarter

player name of the player who took the shot

team 3 letter abbreviation of the attacking team

made whether the shot resulted in a made basket or not

shot_type 2-pointer or 3-pointer

distance Distance (in ft) of shot from the basket. Dunks count as shots made from 0 distance

score game score after the shot

opp 3 letter abbreviation of the defending team

status situation of score (tied,trials,leads)

---

shot_visual   *Visualize the 2 PT percentage and 3 PT percentage of all the players*

Description

The plot that show the corresponding match shot result

Usage

```
shot_visual(data1 = shot, team1, opponent1)
```

Arguments

data1	dataset, dataset that contains all match data of the day
team1	string, input the team that we want to visualize
opponent1	string, input the non-home team of this match

Value

plot, containing the situation of whether the basketball shot is in or out

Examples

```
library(NBADP)
shot_visual(data1 = shot, team1 = "NOP",opponent1 = "NOP")
```

---

state.data.raw   *Analyze the state change of Kawhi Leonard the state fluctuation of Kawhi Leonard each month, from 2022 October to 2023 April*

Description

A data set containing different attributes of Kawhi Leonard in those months. The variables are as follows.

Usage

```
state.data.raw
```

Format

a data frame with 7 rows and 30 variables:

Player the name of the player(Kawhi Leonard)

Month the month that he plays(October–April)

TEAM the team that Kawhi Leonard plays for(LAC)

AGE the age of Kawhi Leonard(31)

GP the number of games that Kawhi Leonard plays

W the number of games that Kawhi Leonard wins

L the number of games that Kawhi Leonard loses

MIN the average time that Kawhi Leonard plays per game

PTS the average points that Kawhi Leonard gets per game

FGM the average field goals that Kawhi Leonard made per game

FGA the average field goals that Kawhi Leonard attempted per game

FG% the average field goal percentage of Kawhi Leonard per game

3pM the average three point field goals that Kawhi Leonard made per game

3pA the average three point field goals that Kawhi Leonard attempted per game

3p% the average three point field percentage percentage of Kawhi Leonard per game

FTM the average free throws that Kawhi Leonard made per game

FTA the average free throws that Kawhi Leonard attempted per game

FT% the average free throw percentage of Kawhi Leonard per game

OREB the average offensive rebounds that Kawhi Leonard gets per game

DREB the average defensive rebounds that Kawhi Leonard gets per game

REB the average total rebounds that Kawhi Leonard gets per game

AST the average assists that Kawhi Leonard gets per game

TOV the average turnovers that Kawhi Leonard makes per game

STL the average steals that Kawhi Leonard gets per game

BLK the average blocks that Kawhi Leonard gives per game

PF the average personal fouls that Kawhi Leonard has per game

FP the average fantasy points that Kawhi Leonard has per game

DD2 the number of double doubles that Kawhi Leonard has in the month

TD3 the number of triple doubles that Kawhi Leonard has in the month

BPM the value of Plus-Minus of Kawhi Leonard per game

X2021to2022data   *2021-2022 NBA Player Stats: Per Game*

Description

A dataset containing each player's details performance data in the 2021-2022 NBA regular season. The variables are as follows:

Usage

    X2021to2022data

Format

An object of class `spec_tbl_df` (inherits from `tbl_df`, `tbl`, `data.frame`) with 812 rows and 31 columns.

Details

@format A data frame with 812 rows and 31 variables:

Rk player ranking (1–605)

Player player name

Pos position

Age player's age (19–41)

Tm 3 letter abbreviation of the team

G number of games that the player join in (1-82)

GS game started (0–82)

MP minutes played per game(1–2854)

FG field goals per game (0–774)

FGA field goal attempts per game (0–1564)

FG% field goal percentage (0–1)

3P 3-point field goals per game (0–285)

3PA 3-point field goal attempts per game (0–750)

3P% 3-point field goal percentage (0–1)

2P 2-point field goals per game (0–724)

2PA 2-point field goal attempts per game (0–1393)

2P% 2-point field goal percentage (0–1)

eFG% effective field goal percentage: This statistic adjusts for the fact that a 3-point field goal is worth one more point that a 2-point field goal (0–1)

FT free throws per game (0–654)

FTA free throw attempts per game (0–803)

FT% free throw percentage (0–1)

ORB offensive rebounds per game (0–349)

DRB defensive rebounds per game (0–813)

TRB total rebounds per game (0–1019)

AST assists per game (0–737)

STL steals per game (0–138)

BLK blocks per game (0–177)

TOV turnovers per game (0–303)

PF personal fouls per game (0–286)

PTS points per game (0–2155)

Player-additional additional information

Index