NBA Data Description

Ben Moolman, Craig Orman, Ethan Pross

```r
library(dplyr)
library(tidyverse)
library(knitr)

original_tbl <- read.csv("../NBA-BoxScores-2023-2024.csv")
# original_tbl$COMMENT[original_tbl$COMMENT == ""] <- NA
# original_tbl$COMMENT <- factor(original_tbl$COMMENT)

original_tbl <- mutate(original_tbl,
                       START_POSITION = na_if(START_POSITION, ""),
                       START_POSITION = factor(original_tbl$START_POSITION),
                       COMMENT = na_if(COMMENT, ""),
                       COMMENT = factor(original_tbl$COMMENT),
                       MIN = na_if(MIN, ""),
                       )

# Unknown variables requiring decoding
# -min
```

Data Set Description:

- There are 30 columns and 32,385 rows. 7 columns are identifiers, 1 column is a row index, 2 are characteristic and comment columns, and then we have 20 numerical statistic columns.

- Each row significes a unique tuple of (game, player), that is to say there is a row for stats of each player in every game.

- There are 1230 unique games, 30 unique teams, and 586 unique players. 7 of the

- There are 5984 rows where all the columns are NA due to the player being out for some reason listed in the comments. No missing values in identifiers.

- Presumably, missing values in the starting position column refer to a non-starter player.

Prior Analysis on this dataset

- There appears to be one analysis submitted for this Kaggle dataset, Done by Adam Briggs, they seem to have done mostly visuals and descriptive statistics, looking at best player on the team, and top performing player in various metrics.

Important notes

- Missing data sampling can be done on games where the player did not play, using priors to consider the effect of previous games and the team.

Variable	Description
GAME_ID	Unique Identifier of the Game.
TEAM_ID	Unique Identifier of the Team. For Each GAME_ID there are only two TEAM_ID.
TEAM_ABBREVIATION	Abbreviation of the Team (e.g. GWS - Golden State Warriors).
TEAM_CITY	NA
PLAYER_ID	Unique Identifier of the Player.
PLAYER_NAME	Complete Name (Name and Surname) of the Player who played the game.
NICKNAME	Nickname of the Player who played the game.
START_POSITION	Position in which the player started the game (If populated, the player started the game).
COMMENT	NA
MIN	Number of minutes in which the player played the game.
FGM	Number of Field Goals (both 2 and 3 Points) Made by the player in the game.
FGA	Number of Field Goals (both 2 and 3 Points) Attempted by the player in the game.
FG_PCT	Percentage of Field Goals (both 2 and 3 Points) by the player in the game.
FG3M	Number of 3 Points Made by the player in the game.
FG3A	Number of 3 Points Attempted by the player in the game.
FG3_PCT	Percentage of 3 Points by the player in the game.
FTM	Number of Free Throws Made by the player in the game.
FTA	Number of Free Throws Attempted Made by the player in the game.
FT_PCT	Percentage of Free Throws by the player in the game.
OREB	Number of Offensive Rebounds reached by the player in the single game.
DREB	Number of Defensive Rebounds reached by the player in the single game.
REB	Number of Total Rebounds (Defensive and Offensive Rebounds) reached by the player in the single game.
AST	Number of Assists done by the player in the single game.
STL	Number of Steals done by the player in the single game.
BLK	Number of Blocks done by the player in the single game.
TO	Number of Turnovers done by the player in the single game.
PF	Number of Personal Fouls done by the player in the single game.

Variable	Description
PTS	Number of Points done by the player in the single game.
PLUS_MINUS	Number of Plus Minus done by the player in the single game.

```r
summary(original_tbl)
```

```
##        X              GAME_ID            TEAM_ID          TEAM_ABBREVIATION
##  Min.   : 0.00   Min.   :22300001   Min.   :1.611e+09   Length:32385
##  1st Qu.: 6.00   1st Qu.:22300303   1st Qu.:1.611e+09   Class :character
##  Median :13.00   Median :22300611   Median :1.611e+09   Mode  :character
##  Mean   :12.71   Mean   :22300613   Mean   :1.611e+09
##  3rd Qu.:19.00   3rd Qu.:22300921   3rd Qu.:1.611e+09
##  Max.   :29.00   Max.   :22301230   Max.   :1.611e+09
##
##   TEAM_CITY          PLAYER_ID       PLAYER_NAME         NICKNAME
##  Length:32385      Min.   :   2544   Length:32385      Length:32385
##  Class :character   1st Qu.:1626246   Class :character   Class :character
##  Mode  :character   Median :1629637   Mode  :character   Mode  :character
##                     Mean   :1337627
##                     3rd Qu.:1630583
##                     Max.   :1642013
##
##  START_POSITION                     COMMENT           MIN
##   :20085                                :26401   Length:32385
##  C: 2460      DNP - Coach's Decision : 5159   Class :character
##  F: 4920      DND - Injury/Illness   :  560   Mode  :character
##  G: 4920      DNP - Injury/Illness   :  148
##               NWT - Injury/Illness   :   27
##               NWT - League Suspension:   23
##               (Other)                :   67
##       FGM              FGA             FG_PCT            FG3M
##  Min.   : 0.000   Min.   : 0.000   Min.   :0.000   Min.   : 0.000
##  1st Qu.: 1.000   1st Qu.: 3.000   1st Qu.:0.286   1st Qu.: 0.000
##  Median : 3.000   Median : 7.000   Median :0.455   Median : 1.000
##  Mean   : 3.929   Mean   : 8.284   Mean   :0.436   Mean   : 1.196
##  3rd Qu.: 6.000   3rd Qu.:12.000   3rd Qu.:0.588   3rd Qu.: 2.000
##  Max.   :25.000   Max.   :47.000   Max.   :1.000   Max.   :12.000
##  NA's   :5984     NA's   :5984     NA's   :5984    NA's   :5984
##      FG3A            FG3_PCT            FTM              FTA
##  Min.   : 0.000   Min.   :0.000   Min.   : 0.000   Min.   : 0.000
##  1st Qu.: 1.000   1st Qu.:0.000   1st Qu.: 0.000   1st Qu.: 0.000
##  Median : 3.000   Median :0.222   Median : 1.000   Median : 1.000
##  Mean   : 3.271   Mean   :0.265   Mean   : 1.587   Mean   : 2.024
##  3rd Qu.: 5.000   3rd Qu.:0.500   3rd Qu.: 2.000   3rd Qu.: 3.000
##  Max.   :23.000   Max.   :1.000   Max.   :24.000   Max.   :32.000
##  NA's   :5984     NA's   :5984    NA's   :5984     NA's   :5984
##     FT_PCT            OREB             DREB              REB
##  Min.   :0.000   Min.   : 0.000   Min.   : 0.000   Min.   : 0.000
##  1st Qu.:0.000   1st Qu.: 0.000   1st Qu.: 1.000   1st Qu.: 1.000
##  Median :0.333   Median : 1.000   Median : 2.000   Median : 3.000
```

```
## Mean   :0.415   Mean   : 0.983   Mean   : 3.074   Mean   : 4.057
## 3rd Qu.:1.000   3rd Qu.: 1.000   3rd Qu.: 4.000   3rd Qu.: 6.000
## Max.   :1.000   Max.   :15.000   Max.   :20.000   Max.   :31.000
## NA's   :5984    NA's   :5984     NA's   :5984     NA's   :5984
##     AST              STL              BLK              TO
## Min.   : 0.000   Min.   :0.000    Min.   : 0.000   Min.   : 0.000
## 1st Qu.: 0.000   1st Qu.:0.000    1st Qu.: 0.000   1st Qu.: 0.000
## Median : 2.000   Median :0.000    Median : 0.000   Median : 1.000
## Mean   : 2.485   Mean   :0.696    Mean   : 0.479   Mean   : 1.202
## 3rd Qu.: 4.000   3rd Qu.:1.000    3rd Qu.: 1.000   3rd Qu.: 2.000
## Max.   :23.000   Max.   :7.000    Max.   :10.000   Max.   :11.000
## NA's   :5984     NA's   :5984     NA's   :5984     NA's   :5984
##     PF               PTS           PLUS_MINUS
## Min.   :0.000    Min.   : 0.00    Min.   :-58
## 1st Qu.:1.000    1st Qu.: 3.00    1st Qu.: -7
## Median :2.000    Median : 9.00    Median :  0
## Mean   :1.745    Mean   :10.64    Mean   :  0
## 3rd Qu.:3.000    3rd Qu.:16.00    3rd Qu.:  7
## Max.   :6.000    Max.   :73.00    Max.   : 46
## NA's   :5984     NA's   :5984     NA's   :5984
```