

# Data\_Mining\_Instruction

*Xuan*

*November 28, 2018*

## Read Me

This is an appending document in my personal Overwatch Project to help you understand how I built my dataset.

Please note that the ranking system is a live update system. This means that everytime you run my code, it is highly possible that you acquire a different dataset from mine. The datasets I applied are stored in the repo, for your reference.

The recording time of the ranking system in my dataset is 11/30/2018.

## Code Sample

```
library(rvest)
library(tidyverse)

# web scrapping
htmlana <- "https://www.overbuff.com/heroes/ana/rankings"
anaplayerid <- htmlana %>% read_html() %>% html_nodes(".sortable a") %>%
  html_attr("href") %>% as.data.frame()
anaplayerid <- unique(anaplayerid)
rownames(anaplayerid) <- c(1:100)

# check the dataset, and we detect a problem that special ids
# do not show correctly. For simplicity, we remove these
# rows.
anaplayerid <- data.frame(anaplayerid[-c(18, 48, 75, 99), ])
colnames(anaplayerid) <- "ID"
rownames(anaplayerid) <- c(1:96)
anaplayerid[, 1] <- as.character(anaplayerid[, 1])

# create links to their main pages
linkhead <- "https://www.overbuff.com"
linktailana <- "/heroes/ana?mode=competitive"
anaplayerid$PLAYER <- NA
for (i in 1:96) {
  anaplayerid$PLAYER[i] <- unlist(strsplit(anaplayerid[i, 1],
    split = "/", fixed = TRUE))[4]
  # list player-id if you are interested in some specific
  # players
  anaplayerid[i, 1] <- sprintf("%s%s", linkhead, anaplayerid[i,
    1], linktailana)
}
colnames(anaplayerid) <- c("URL", "ID")

# Game Record create empty data frame
```

```

GameRecordS13Ana <- as.data.frame(matrix(ncol = 22, nrow = 1,
NA))
colnames(GameRecordS13Ana) <- c("GAME_PLAYED", "WIN_RATE", "ON_FIRE",
"ELIS", "OBJ_KILLS", "OBJ_TIME", "DAMAGE", "HEALING", "DEATH",
"WEAPON_ACC", "SCOPED_ACC", "ENEMY_SLEPT", "OFF_ASSISTS",
"DEF_ASSISTS", "BOOST_ASSISTS", "ED", "VOTING_CARDS", "MEDALS",
"GOLD", "SILVER", "BRONZE", "PLAYER_ID")

for (i in 1:96) {
  tryCatch({
    url <- anaplayerid[i, 1]
    templink <- read_html(url)
    a <- templink %>% html_nodes(".player-heroes .value") %>%
      html_text()
    a <- as.data.frame(matrix(ncol = 21, a, byrow = TRUE))
    colnames(a) <- c("GAME_PLAYED", "WIN_RATE", "ON_FIRE",
      "ELIS", "OBJ_KILLS", "OBJ_TIME", "DAMAGE", "HEALING",
      "DEATH", "WEAPON_ACC", "SCOPED_ACC", "ENEMY_SLEPT",
      "OFF_ASSISTS", "DEF_ASSISTS", "BOOST_ASSISTS", "ED",
      "VOTING_CARDS", "MEDALS", "GOLD", "SILVER", "BRONZE")
    a$PLAYER_ID <- anaplayerid[i, 2]
    GameRecordS13Ana <- rbind.data.frame(a, GameRecordS13Ana) #add game records each by each
  }, error = function(e) {
  }) #avoid error situations
}

# convert factors to numeric values
GameRecordS13Ana$WIN_RATE <- lapply(GameRecordS13Ana$WIN_RATE,
  function(x) as.numeric(sub("%", "", x)))
GameRecordS13Ana$WIN_RATE <- as.numeric(GameRecordS13Ana$WIN_RATE)/100
GameRecordS13Ana <- GameRecordS13Ana %>% filter(WIN_RATE < 1.1) #remove outliers
GameRecordS13Ana$GAME_PLAYED <- as.numeric(GameRecordS13Ana$GAME_PLAYED)
GameRecordS13Ana$WIN <- round(GameRecordS13Ana$GAME_PLAYED *
  GameRecordS13Ana$WIN_RATE, 0)
GameRecordS13Ana$LOSS <- GameRecordS13Ana$GAME_PLAYED - GameRecordS13Ana$WIN
GameRecordS13Ana$ED <- as.character(GameRecordS13Ana$ED)
GameRecordS13Ana$ED <- as.numeric(GameRecordS13Ana$ED)
GameRecordS13Ana$MEDALS <- as.character(GameRecordS13Ana$MEDALS)
GameRecordS13Ana$MEDALS <- as.numeric(GameRecordS13Ana$MEDALS)

# write down the file
write.csv(GameRecordS13Ana, file = "GameRecordS13Ana.csv")

## Now we do this again to get the dataset for Reinhardt mains
## web scrapping
htmlrein <- "https://www.overbuff.com/heroes/reinhardt/rankings"
reinplayerid <- htmlrein %>% read_html() %>% html_nodes(".sortable a") %>%
  html_attr("href") %>% as.data.frame()
reinplayerid <- unique(reinplayerid)
rownames(reinplayerid) <- c(1:100)
reinplayerid <- as.data.frame(reinplayerid[-51, ])
colnames(reinplayerid) <- "ID"

```

```

reinplayerid$ID <- as.character(reinplayerid$ID)

# create links to their main pages
linkhead <- "https://www.overbuff.com"
linktailrein <- "/heroes/reinhardt?mode=competitive"
reinplayerid$PLAYER <- NA
for (i in 1:99) {
  reinplayerid$PLAYER[i] <- unlist(strsplit(reinplayerid[i,
    1], split = "/", fixed = TRUE))[4]
  # list player-id if you are interested in some specific
  # players
  reinplayerid[i, 1] <- sprintf("%s%s%s", linkhead, reinplayerid[i,
    1], linktailrein)
}
colnames(reinplayerid) <- c("URL", "ID")

GameRecordS13R <- as.data.frame(matrix(ncol = 21, nrow = 1, NA))
colnames(GameRecordS13R) <- c("GAME_PLAYED", "WIN_RATE", "ON_FIRE",
  "ELIS", "OBJ_KILLS", "OBJ_TIME", "DAMAGE", "DEATH", "SOLO_KILLS",
  "FINAL_BLOWS", "BLOCKED", "CHARGE", "ULT", "FIRE", "ED",
  "VOTING_CARDS", "MEDALS", "GOLD", "SILVER", "BRONZE", "PLAYER_ID")

for (i in 1:96) {
  tryCatch({
    url <- reinplayerid[i, 1]
    templink <- read_html(url)
    a <- templink %>% html_nodes(".player-heroes .value") %>%
      html_text()
    a <- as.data.frame(matrix(ncol = 20, a, byrow = TRUE))
    colnames(a) <- c("GAME_PLAYED", "WIN_RATE", "ON_FIRE",
      "ELIS", "OBJ_KILLS", "OBJ_TIME", "DAMAGE", "DEATH",
      "SOLO_KILLS", "FINAL_BLOWS", "BLOCKED", "CHARGE",
      "ULT", "FIRE", "ED", "VOTING_CARDS", "MEDALS", "GOLD",
      "SILVER", "BRONZE")
    a$PLAYER_ID <- reinplayerid[i, 2]
    GameRecordS13R <- rbind.data.frame(a, GameRecordS13R)
  }, error = function(e) {
  })
}

# convert factors to numeric values
GameRecordS13R$WIN_RATE <- lapply(GameRecordS13R$WIN_RATE, function(x) as.numeric(sub("%",
  "", x)))
GameRecordS13R$WIN_RATE <- as.numeric(GameRecordS13R$WIN_RATE)/100
GameRecordS13R <- GameRecordS13R %>% filter(WIN_RATE < 1.1)
GameRecordS13R$GAME_PLAYED <- as.character(GameRecordS13R$GAME_PLAYED)
GameRecordS13R$GAME_PLAYED <- as.numeric(GameRecordS13R$GAME_PLAYED)
GameRecordS13R$WIN <- round(GameRecordS13R$GAME_PLAYED * GameRecordS13R$WIN_RATE,
  0)
GameRecordS13R$LOSS <- GameRecordS13R$GAME_PLAYED - GameRecordS13R$WIN
GameRecordS13R$ED <- as.character(GameRecordS13R$ED)

```

```
GameRecordS13R$ED <- as.numeric(GameRecordS13R$ED)
GameRecordS13R$MEDALS <- as.character(GameRecordS13R$MEDALS)
GameRecordS13R$MEDALS <- as.numeric(GameRecordS13R$MEDALS)

# write down the file
write.csv(GameRecordS13R, file = "GameRecordS13R.csv")
```