Kobe Bryant Career Stats Analysis

Mark Chen

2020/3/6

Kober Bryant NBA record (R)

To explore the NBA career of Kobe Bryant, the best player in our generation !

Import packages

```
library(tidyverse) # Data wrangle
library(lubridate) # Date
library(reticulate)
```

Import data

```
working_dir <- "/cloud/project/"
file_name <- "kobe.csv"
file_dir <- paste0(working_dir, file_name)
data <- read.csv(file_dir, stringsAsFactors = F)
```

Data Preprocessing

```
data <- mutate(data, shot_point = as.numeric(substring(shot_type, 1, 1)),
               shot_gain = shot_point * shot_made_flag,
               total_seconds_remaining = minutes_remaining * 60 + seconds_remaining)
data[, "game_date"] <- as.Date(data[, "game_date"] ,"%Y/%m/%d")
```

(1) Kobe's performance against any team.

Calculate the average 2 points, 3 points field goal rate, and gained scores against any team.

```
data_1 <-
  data %>%
  select(opponent, shot_type, shot_made_flag) %>%
  filter(opponent == "HOU") %>%
  group_by(shot_type) %>%
  summarise(FGp = sum(shot_made_flag)/length(shot_made_flag),
```

```
            shooting_number = n())
data_1
```

```
## # A tibble: 2 x 3
##   shot_type        FGp shooting_number
##   <chr>          <dbl>           <int>
## 1 2PT Field Goal 0.388            1350
## 2 3PT Field Goal 0.266             316
```

(2) Toughest Teams Kobe Has Competed against In His Career

Find the 5 teams making Kobe get the lowest average scores in whole career.

```
data_2 <-
  data %>%
  select(game_id, opponent, shot_gain) %>%
  group_by(game_id, opponent) %>%
  summarise(game_score = sum(shot_gain)) %>%
  group_by(opponent) %>%
  summarise(mean_score = mean(game_score)) %>%
  arrange(mean_score)
head(data_2,5)
```

```
## # A tibble: 5 x 2
##   opponent mean_score
##   <chr>         <dbl>
## 1 VAN            11.2
## 2 BKN            12.7
## 3 IND            14.0
## 4 UTA            14.1
## 5 NJN            14.2
```

(3) Kobe's clutch moments in NBA playoff.

Find the games in which Kobe got the highest scores during the last 3 minutes in the playoff.

```
data_3 <-
  data %>%
  select(game_id, opponent, playoffs, period, total_seconds_remaining, shot_gain) %>%
  filter(playoffs == 1, period == 4, total_seconds_remaining <= 3*60) %>%
  group_by(game_id, opponent) %>%
  summarise(total_scores = sum(shot_gain)) %>%
  arrange(desc(total_scores))
head(data_3,5)
```

```
## # A tibble: 5 x 3
## # Groups:   game_id [5]
##    game_id opponent total_scores
##      <int> <chr>           <dbl>
## 1 49900074 POR                 9
## 2 40100234 SAS                 8
## 3 40700402 BOS                 7
## 4 49900024 SAC                 7
## 5 40100402 NJN                 6
```

(4) Is Kobe the best closer?

List the Jump-shot Field Goal Percentage of Kobe in the last one minute in every season playoff.

```
data_4 <-
  data %>%
  select(season, opponent, playoffs, period, total_seconds_remaining,
         action_type, shot_made_flag) %>%
  filter(playoffs == 1, period == 4, total_seconds_remaining <= 1*60,
         action_type == "Jump Shot") %>%
  group_by(season) %>%
  summarise(FGp = sum(shot_made_flag)/ length(shot_made_flag))
data_4
```

```
## # A tibble: 15 x 2
##    season     FGp
##    <chr>    <dbl>
##  1 1996-97 0.4
##  2 1997-98 0
##  3 1998-99 0.143
##  4 1999-00 0.1
##  5 Apr-03  0.167
##  6 Aug-07  0
##  7 Dec-11  0.0909
##  8 Feb-01  0.5
##  9 Jan-00  0
## 10 Jul-06  0.5
## 11 Jun-05  0
## 12 Mar-02  0.333
## 13 Nov-10  0
## 14 Oct-09  0.333
## 15 Sep-08  0.143
```

(5) Kobe's "hot hand" performance

List the longest consecutive games which Kobe got at least 33% FGP.

```
continuous_detect <- function (x) {
  #check
  if (class(x)[1] == "tbl_df"){
    x <- pull(x)
  }
  ini <- 1; con_df <- c(); end <- length(x)
  while (ini != end) {
    if ( x[ini] == 0) {
      ini <- ini + 1
    }else{
      start <- ini
      while (x[ini] == 1 ) {
        ini <- ini + 1
      }
      df <- data.frame(start_index = start, end_index = ini - 1, interval = ini - start )
      con_df <- rbind(con_df, df)
    }
  }
  con_df
}
data_5_intme <-
  data %>%
  select(game_id, game_date, shot_made_flag) %>%
  group_by(game_id, game_date) %>%
  summarise(FGp = sum(shot_made_flag)/ n()) %>%
  arrange(game_date) %>%
  mutate(meet_goal = if_else(FGp < 0.33, 0, 1) )
dt <- continuous_detect(data_5_intme[,4])
date <- dt %>% arrange(desc(interval)) %>% head(3)
calender <- lapply(1:nrow(date), function(i)
  as.character(c(data_5_intme$game_date[date[i,1]], data_5_intme$game_date[date[i,2]])))
data_5 <- as.data.frame(cbind(do.call(rbind,calender), date$interval))
colnames(data_5) <- c("starting_date", "ending_date", "interval")
data_5
```

```
##   starting_date ending_date interval
## 1    2000-02-01  2000-03-12       19
## 2    2001-06-15  2001-12-07       19
## 3    2002-01-14  2002-02-17       16
```

(6) Kobe's worst status in the games.

List the games which Kobe got the lowest FGP and gained more scores in the first half than the second one.

```
FsurpassS_id <-
  data %>%
  select(game_date, period, shot_gain) %>%
  mutate(half = ifelse(period <= 2, "first", "second")) %>%
  group_by(game_date, half) %>%
  summarise(half_scores = sum(shot_gain)) %>%
  group_by(game_date) %>%
  summarise(gains_diff =  half_scores[1] - half_scores[2]) %>%
  filter(gains_diff > 0)
data_6 <-
  data %>%
  select(game_date, period, opponent, shot_gain, shot_made_flag) %>%
  filter(game_date %in% FsurpassS_id$game_date, period %in% 1:4) %>%
  group_by(game_date, opponent) %>%
  summarise(FGp = sum(shot_made_flag)/n(),
            points = sum(shot_gain)) %>%
  inner_join(FsurpassS_id) %>%
  arrange(FGp)
head(data_6,3)
```

```
## # A tibble: 3 x 5
## # Groups:   game_date [3]
##   game_date  opponent    FGp points gains_diff
##   <date>     <chr>     <dbl>  <dbl>      <dbl>
## 1 2015-11-24 GSW      0.0714      3          3
## 2 2003-12-21 PHX      0.0833      2          2
## 3 2016-04-05 LAC      0.0833      2          2
```

(7) Is Kobe a persistent bricklayer?

List the maximum of Kobe's continuous missing shot in one game.

```
continuous_time <- 3
continuous_max_intvl <- function (x) {
  ini <- 1; end <- length(x); interval <- 0;interval_list <- c()
  if ( end < 2) {
    if ( x[1] == 0){
      interval <- interval + 1
    }
    interval
  }else{
    interval_list <-  c(interval_list, interval)
    while (ini != end + 1 ) {
      if ( x[ini] == 1) {
        ini <- ini + 1
      }else{
        interval <- 0
        while ((x[ini] == 0) & (ini != end + 1)){
          interval <- interval + 1
          ini <- ini + 1
        }
        interval_list <- c(interval_list, interval)
        }
    }
    max(interval_list)
  }
}
data_7_ini <-
  data %>%
  select(game_date, period, minutes_remaining, shot_made_flag) %>%
  arrange(game_date, period) %>%
  group_split(game_date)
total_intvl <-unlist(lapply(1:length(data_7_ini), function(i)
  continuous_max_intvl(data_7_ini[[i]]$shot_made_flag) ) )
total_date <- unlist(lapply(1:length(data_7_ini), function(i)
  as.character(data_7_ini[[i]]$game_date[1])))
total_date <- as.Date(total_date ,"%Y-%m-%d")
date_intvl <- data.frame(game_date = total_date, interval = total_intvl)
index <- date_intvl %>% arrange(desc(interval)) %>% head(continuous_time)
data_7 <-
  data %>%
  select(game_date, opponent, shot_gain) %>%
  filter(game_date %in% index$game_date) %>%
  group_by(game_date, opponent) %>%
  summarise(points = sum(shot_gain)) %>%
  inner_join(index) %>%
  arrange(desc(interval))
data_7
```

```
## # A tibble: 3 x 4
## # Groups:   game_date [3]
##   game_date  opponent points interval
```

```
##   <date>     <chr>     <dbl>    <dbl>
## 1 2010-04-02 UTA           8       16
## 2 2015-12-01 PHI          10       16
## 3 2003-03-12 DET          12       15
```