Basketball Operations Analyst

Prepared for: Redacted - Analyst, Basketball Operations

Prepared by: Jason Lee, Data Analyst

March 22, 2017

For my analysis I have used the R programming environment. Code for each question is included at the end.

1. Identify the top 5 teams & players in offensive rebounding. Describe the metric(s) used and explain why you used it.

To determine which teams are the best at offensive rebounding, I used the percentage of offensive rebounds. I took each teams' total number of offensive rebounds and divided it by the number of missed shots the team had total, including free throws that resulted in some type of rebound, to determine their percentage.

For the players I decided to use the rate at which they get offensive rebounds. I used this statistic to show which players were the best offensive rebounders based on their rate per 48 minutes of playing time instead of the percentage of rebounds they grabbed. I applied a filter to the players to take into account only players who averaged a minimum of 12 min of playing time per game to eliminate players who do not play much.

Team	% Offensive Rebounds
Unicaja Malaga	35.85%
Anadolu Efes Istanbul	32.55%
Maccabi FOX Tel Aviv	32.45%
Real Madrid	31.20%
Khimiki Moscow Region	31.01%

Player	Off Rebounds Per 48min
Trevor Mbakwe	7.35
Ognjen Kuzmic	6.96
Othello Hunter	6.52
Vladimir Stimac	6.31
Felipe Reyes	6.24

2. Identify the top 5 teams & players whose offensive rebounds most effectively yield 2nd chance points.

The team data was straight forward and did not need any filters. The player data I applied a filter to only include players who averaged more than 12 min per game and also attempted an average of 0.7 second chance field goal attempts per game. I felt the 0.7 mark eliminated the 100% FG shooting of certain players who only took a handful of shots.

Team	2nd Chance FG%
Pinar Karsiyaka Izmir	56.25%
Limoges CSP	55.00%
Anadolu Efes Istanbul	54.91%
Fenerbahce Istanbul	54.50%
Cedevita Zagreb	53.99%

Player	2nd Chance FG%
Christian Eyenga	87.50%
Colton Iverson	85.71%
Ali Traore	80.00%
Jan Vesely	76.32%
Kyle Hines	74.47%

3. What is the 2P%, 3P%, eFG% of shots after offensive rebounds?

2nd Chance 2P%	2nd Chance 3P%	2nd Chance eFG%
55.99%	30.15%	53.32%

4. What additional data would you acquire/use to predict likelihood of scoring 2nd chance points?

To help predict the likelihood of second chance points there would be several statistics that I would be interested in investing to discover if there were any correlation. Using the opponent's defensive rating and our team's offensive rating would help. What zone the second chance field goal attempt took place in. Opponents Defensive FG% for each zone and our team's FG% at each zone. If we really wanted to go deeper, we could look at the specific lineup that our team and the opponent has on the floor and use each of the before mentioned statistics for them to see if there is any noticeable interaction between lineups.

Also, the time between the offensive rebound and the field goal attempt could add some insight. Whether the game is played at home or on the road may also help because several teams play better at home. How many days rest before the game may factor in as well. If we knew how distance away from the defender when the field goal was attempted could help too.

5. (Bonus) What is the FG% of shots after offensive rebounds in each court zone (restricted area, (non-RA) in-the-paint, midrange, corner 3, above the break 3)?

To preface my results, I was not 100% positive about the coordinate locations of the different zones. The way I broke down the zones are as follows:

Restricted Area: Y coordinate less than 100 and the X coordinate is between 100 and -100.

Non-RA: Y coordinate less than 400 and the X coordinate is between 250 and -250.

Midrange: All other 2FGA.

Corner 3: Y coordinate less than 300.

Above the break 3: Y coordinate greater than or equal to 300.

Zone	2nd Chance FG%
Restricted Area	83.28%
In-The-Paint (non-RA)	50.74%
Midrange	32.68%
Corner 3	28.36%
Above the break 3	30.48%

6. (Bonus) Build a visualization to show FG% of shots after offensive rebounds.

My normal NBA shot charts I create pull from the nba.com website where the location data is on a different scale than the data I used for this project, so I want to apologize for the unpolished looking plot.
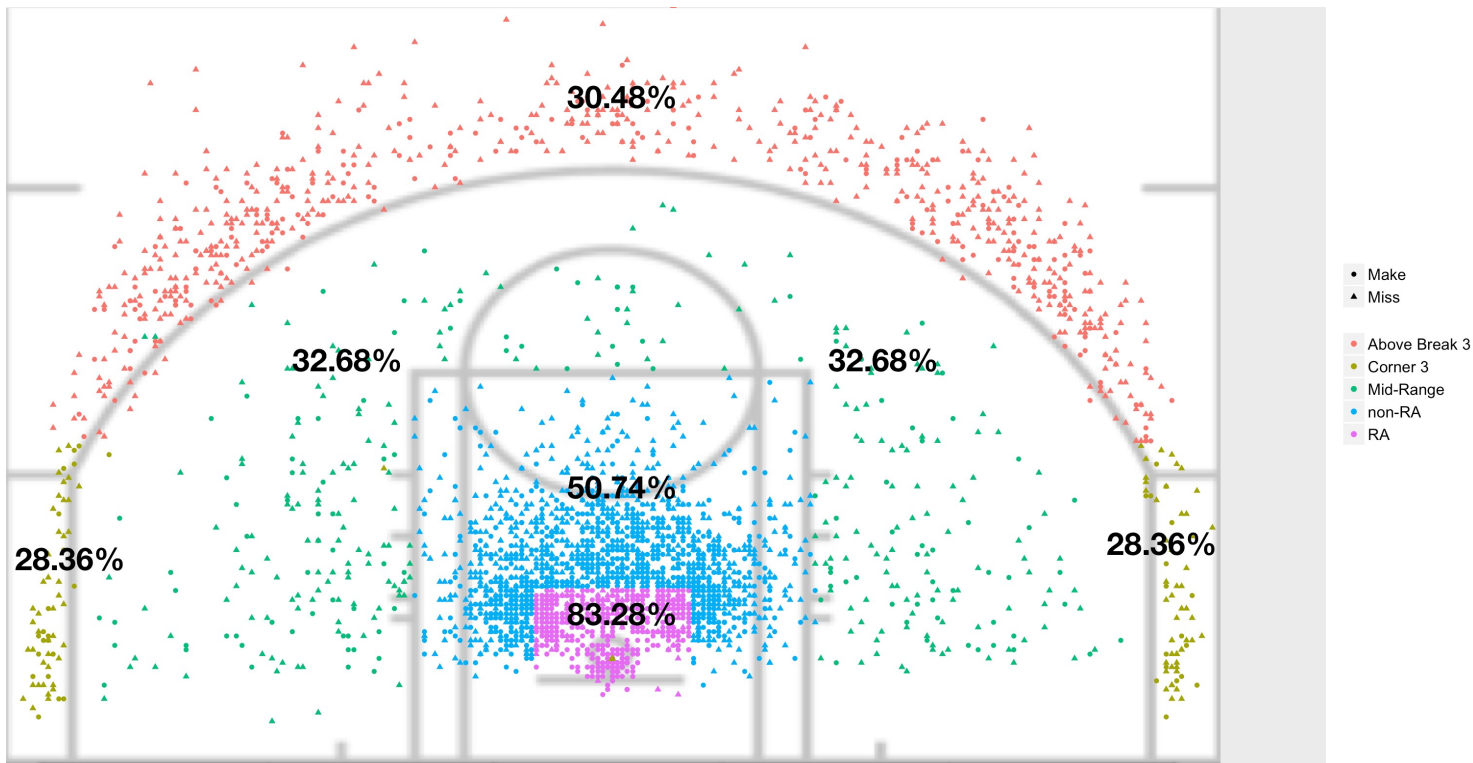
R CODE

```
####################################

#####  Phoenix Suns Analyst Project  #####

####################################


# load all packages to be used
library(rjson)
library(ggplot2)
library(grid)
library(gridExtra)
library(png)
library(jpeg)
library(RCurl)
library(hexbin)
library(plyr)
library(Hmisc)
library(tidyverse)

# Load Data
DataChallengeEuroGameRoster <- read.csv("~/Desktop/R Script/Basketball/NBA/DataChallenge2017/
DataChallengeEuroGameRoster.csv", stringsAsFactors=FALSE)
DataChallengeEuroGamePBP <- read.csv("~/Desktop/R Script/Basketball/NBA/DataChallenge2017/
DataChallengeEuroGamePBP.csv", stringsAsFactors=FALSE)

roster <- DataChallengeEuroGameRoster
plays <- DataChallengeEuroGamePBP

#'  1. Identify the top 5 teams & players in offensive rebounding.
#'      Describe the metrics used and explain why you used it.
```

```r
# adding variables to plays
plays$lastplay <- Lag(plays$event_desc_id, 1)
plays$putback <- ifelse(plays$lastplay == "O" & plays$event_desc_id == "2FGM" | plays$lastplay == "O" &
plays$event_desc_id == "3FGM", 1, 0) # 2nd chance points
plays$FTmiss <- ifelse(plays$lastplay == "FTA" & plays$event_desc_id == "O" | plays$lastplay == "FTA" &
plays$event_desc_id == "D", 1, 0) # free throws that resulted in a rebound
plays$secondChance <- ifelse(plays$lastplay == "O", 1, 0)

# gets only the offensive rebounds
oreb <- filter(plays, event_desc_id == "O")
oreb$oreb <- 1

# get how many misses
misses <- filter(plays, event_desc_id %in% c("2FGA", "2FGAB", "3FGA", "3FGAB")) # gets all misses
misses1 <- filter(plays, FTmiss == 1) # gets all the free throws missed that resulted in some type of rebound

# combine the misses dataframes
miss <- rbind(misses, misses1)
miss$miss <- 1

# gets the total rebounds for each team
oreb$team_id <- as.factor(oreb$team_id)
teamsOreb <- aggregate(oreb ~ team_id, oreb, sum)
teamsMiss <- aggregate(miss ~ team_id, miss, sum)

# gets total offensive rebounds per game for each team
teamOrebPerGame <- aggregate(oreb ~ team_id + gamecode, oreb, sum)

# gets total games played for each team
teamGames <- data.frame(table(teamOrebPerGame$team_id))
colnames(teamGames) <- c("team_id", "games") # rename columns

# combine team dataframes
teamOrebPercent <- merge(teamsOreb, teamsMiss, by = "team_id")
teamOrebPercent <- merge(teamOrebPercent, teamGames, by = "team_id")
```

```
# add percentage of offensive rebounds and also just for kicks add how many offensive rebounds each team avg
per game
teamOrebPercent$perc <- with(teamOrebPercent, round(100*(oreb/miss), 2))
teamOrebPercent$perGame <- with(teamOrebPercent, oreb/games)
teamOrebPercent <- teamOrebPercent[order(teamOrebPercent$perc, decreasing = TRUE),] # rearrange from
greatest to least


############### gets the top 5 teams for percentage of rebounds  #######
team.top5 <- teamOrebPercent[1:5,]



# gets the totals for offensive rebounds per player
playerOreb <- aggregate(oreb ~ player_id, oreb, sum)
playerTime <- aggregate(mp_as_int ~ player_id, roster, sum)

## there are 844 offensive rebounds that don't have a player listed with them

# combine dataframes
playerTeam <- roster[,4:5]
playerTeam <- merge(playerTeam, teamGames, by = "team_id", all.x = TRUE)
final.player <- merge(playerOreb, playerTime, by = "player_id", all.x = TRUE)
final.player <- merge(final.player, playerTeam, by = "player_id", all.x = TRUE)
final.player <- final.player[!duplicated(final.player$player_id),] # gets rid of duplicated rows from the merge

# add variables - seconds per game they played and offensive rebounds per 48 minutes (game)
final.player$secPerGame <- with(final.player, mp_as_int/games)
final.player$per48min <- with(final.player, (oreb/mp_as_int)*60*48)

# filter out players who play less than a quarter a game on average (12 minutes)
final.player <- filter(final.player, secPerGame >= 720)

## need to set a limit for rebounds... maybe like more than 48 min playing time
final.player <- final.player[order(final.player$per48min, decreasing = TRUE),]

# TOP 5 PLAYERS
player.top5 <- final.player[1:5,]
```

```r
#' 2. Identify the top 5 teams & players whose offensive rebounds
#'    most effectively yield 2nd chance points


# this gets us the second chance buckets
SecondChance <- filter(plays, secondChance == 1)

# splits up misses and makes from the play types
SecondChanceMiss <- filter(SecondChance, event_desc_id %in% c("2FGA", "2FGAB", "3FGA", "3FGAB")) # gets
all misses
SecondChanceMake <- filter(SecondChance, event_desc_id %in% c("2FGM", "3FGM")) # gets all misses

# sum up the misses and makes
teamSecondMake <- aggregate(secondChance ~ team_id, SecondChanceMake, sum)
teamSecondMiss <- aggregate(secondChance ~ team_id, SecondChanceMiss, sum)
colnames(teamSecondMake)[2] <- "make"
colnames(teamSecondMiss)[2] <- "miss"

# combine makes and misses dataframe
final.team2ndChance <- merge(teamSecondMake, teamSecondMiss, by = "team_id")

# find the FG% of second chance attempts
final.team2ndChance$perc <- with(final.team2ndChance, round(100*(make/(miss + make)), 2))

# reorders data
final.team2ndChance <- final.team2ndChance[order(final.team2ndChance$perc, decreasing = TRUE),]

# TOP 5 Teams
top5.2ndChance <- final.team2ndChance[1:5,]



# gets the totals for offensive rebounds per player
player2ndMake <- aggregate(secondChance ~ player_id, SecondChanceMake, sum)
```

```r
player2ndMiss <- aggregate(secondChance ~ player_id, SecondChanceMiss, sum)
colnames(player2ndMake)[2] <- "make"
colnames(player2ndMiss)[2] <- "miss"

# merge dataframes
final.player2ndChance <- merge(player2ndMake, player2ndMiss, by = "player_id", all = TRUE)

# change NA to 0
final.player2ndChance$make <- ifelse(is.na(final.player2ndChance$make) == TRUE, 0,
final.player2ndChance$make)
final.player2ndChance$miss <- ifelse(is.na(final.player2ndChance$miss) == TRUE, 0, final.player2ndChance$miss)

# gets percentage
final.player2ndChance$attempts <- with(final.player2ndChance, miss + make)
final.player2ndChance$perc <- with(final.player2ndChance, make/attempts)

# combine dataframes
final.player2ndChance <- merge(final.player2ndChance, playerTime, by = "player_id", all.x = TRUE)
final.player2ndChance <- merge(final.player2ndChance, playerTeam, by = "player_id", all.x = TRUE)
final.player2ndChance <- final.player2ndChance[!duplicated(final.player2ndChance$player_id),] # gets rid of
duplicated rows from the merge

# add variables - seconds per game they played and offensive rebounds per 48 minutes (game)
final.player2ndChance$secPerGame <- with(final.player2ndChance, mp_as_int/games)
final.player2ndChance$attPerGame <- with(final.player2ndChance, attempts/games)

# filter out players who play less than a quarter a game (12 minutes)
final.player2ndChance <- filter(final.player2ndChance, secPerGame >= 720)
final.player2ndChance <- filter(final.player2ndChance, attPerGame >= .7)

# reorder data
final.player2ndChance <- final.player2ndChance[order(final.player2ndChance$perc, decreasing = TRUE),]

# TOP 5 PLAYERS 2nd chance points
player2nd.top5 <- final.player2ndChance[1:5,]
```

```r
# 3. What is the 2P%, 3P%, eFG% of shots after offensive rebounds?

# quick fix to combine blocked shots with regular misses
SecondChanceMiss[,13] <- sapply(SecondChanceMiss$event_desc_id, function(x) {gsub("B", "", x)})

# sum up the misses and makes
SecondMake <- aggregate(secondChance ~ event_desc_id, SecondChanceMake, sum)
SecondMiss <- aggregate(secondChance ~ event_desc_id, SecondChanceMiss, sum)

colnames(SecondMake)[2] <- "make"
colnames(SecondMiss)[2] <- "miss"

# add id for whether it is a 2 pointer or a 3 pointer
SecondMake$id <- c(2,3)
SecondMiss$id <- c(2,3)

# combine and add variables
final <- merge(SecondMake, SecondMiss, by = "id")
final$att <- with(final, make + miss)
final$perc <- with(final, round(100*(make/att),2))
final$perc
#[1] 0.5599229 0.3015134
# 55.99% for 2FG and 30.15% for 3FG

## eFG%
round(100*(final$make[1] + (final$make[2] * 1.5))/sum(final$att),2)
# 53.32% eFG%


#'  5. (bonus) What is the FG% of shots after offensive rebounds in each
#'      court zone (restricted area, (non-RA) in-the-paint, midrange, corner 3,
#'      above break 3)?


# add a flag
SecondChanceMake$result <- "Make"
SecondChanceMiss$result <- "Miss"
```

```
shots <- rbind(SecondChanceMake, SecondChanceMiss)

# separate 2s and 3s to make it easier to split up the zones
twos <- filter(shots, event_desc_id %in% c("2FGA", "2FGAB", "2FGM")) # gets all misses
threes <- filter(shots, event_desc_id %in% c("3FGA", "3FGAB","3FGM")) # gets all misses

# set zones for 3s
threes$zone <- ifelse(threes$coord_y < 300, "Corner 3", "Above Break 3")

# set zones for 2s
twos$zone <- "Mid-Range"
twos$zone <- ifelse(twos$coord_y < 100 & twos$coord_x > -100 & twos$coord_x < 100, "RA", twos$zone)
twos$zone <- ifelse(twos$coord_y < 400 & twos$coord_y >= 100 & twos$coord_x > -250 & twos$coord_x < 250,
"non-RA", twos$zone)
twos$zone <- ifelse(twos$coord_y <= 100 & twos$coord_x <= -100 & twos$coord_x > -250, "non-RA",
twos$zone)
twos$zone <- ifelse(twos$coord_y <= 100 & twos$coord_x >= 100 & twos$coord_x <= 250, "non-RA",
twos$zone)

# combine 2s and 3s
shotsZone <- rbind(twos, threes)

## get percentages for shot zones
zonePerc <- aggregate(secondChance ~ zone + result, shotsZone, sum)
# split up the dataset by make or miss
zonePerc1 <- zonePerc[1:5,]
zonePerc2 <- zonePerc[6:10,]
# combine them back together
zonePerc <- merge(zonePerc1, zonePerc2, by = "zone")
colnames(zonePerc)[3] <- "make"
colnames(zonePerc)[5] <- "miss"
zonePerc <- zonePerc[,-c(2,4)]
## add variables
zonePerc$att <- with(zonePerc, make + miss)
zonePerc$perc <- with(zonePerc, round(100*(make / att),2))

### gives us the percentages for all the zones
```

zonePerc[,c(1,5)]

#' 6. (bonus) Build a visualization to show FG% of shots after offensive rebounds.

```r
# half court image
court <- rasterGrob(readPNG("~/Desktop/R Script/Basketball/NBA/halfcourt.png"),
                    width=unit(1,"npc"), height=unit(1,"npc"))

# plot using NBA court background and colour by shot zone
ggplot(shotsZone, aes(x=coord_x, y=coord_y)) +
  annotation_custom(court, -750, 750, -100, 900) +
  geom_point(aes(colour = zone, shape = result)) +
  xlim(-800, 800) +
  ylim(-100, 1000) +
  theme(line = element_blank(),
        axis.title.x = element_blank(),
        axis.title.y = element_blank(),
        axis.text.x = element_blank(),
        axis.text.y = element_blank(),
        legend.title = element_blank(),
        legend.text=element_text(size = 12),
        plot.title = element_text(size = 17, lineheight = 1.2, face = "bold"))
```