# Exploring The Failures of The New York Knicks (2013-2022)

Daniel Liu

27/06/2023

#### Introduction

As an avid New York Knicks fan, this data analysis report attempts to gain further insight into why the New York Knicks have disappointed over the past decade in the NBA. The insights I uncover will help uncover key areas of improvement which will hopefully guide the Knicks to winning more games. I will approach this task through the Six Steps of Data Analysis Process.

### Questions of Interest

- 1. How has the rise in popularity of the 3 point shot impacted the Knicks?
- 2. What other key statistical areas of improvement are important?
- 3. Does the team salary play a role in the success of a team?

## Preparing data

I pulled data from a multitude of sources in conducting my analysis.

- 1. <u>NBA Database</u> is a dataset containing statistics on all 30 NBA teams, 4800+ NBA players and 60,000+ NBA games throughout history.
- 2. <u>NBA API</u> is an API I used to obtain data on NBA team statistics as opposed to having to wrangle the above datasets.
- 3. <u>NBA Salaries</u> is a dataset containing NBA player salaries between 2010-2020. As this analysis is conducted up until 2022, I scraped data from spotrac to obtain NBA team salaries between 2021-2022.

#### Extracting data using NBA API in Python

Obtaining advanced metrics for all teams from 2013-2022.

```
seasons = ['2013-14', '2014-15', '2015-16', '2016-17', '2017-18', '2018-19', '2019-20',
'2020-21', '2021-22', '2022-23']
team_metrics = []
for season in seasons:
    tmp = teamestimatedmetrics.TeamEstimatedMetrics(season = season).get_data_frames()[0]
    tmp["YEAR"] = season[:4]
    team_metrics.append(tmp)
team_metrics = pd.concat(team_metrics, ignore_index = True)
team_metrics.to_csv("team_metrics.csv", index = False)
```

Obtaining basic team statistics from 2013-2022.

```
def get_team_stats(individual_team_id: int):
    all_df = teamyearbyyearstats.TeamYearByYearStats(team_id = individual_team_id)
    all_df = all_df.get_data_frames()[0].tail(10)
    all_df["YEAR"] = all_df["YEAR"].apply(lambda year: year[:4])
    return all_df
```

```
team_data = []
for team_id in team_ids:
    team_info = get_team_stats(str(team_id))
    team_data.append(team_info)
all_team_df = pd.concat(team_data, ignore_index = True)
all_team_df.to_csv("all_team_stats.csv", index = True)
```

Obtaining New York Knick team statistics from 2013-2022.

```
NYK_stats_df = teamyearbyyearstats.TeamYearByYearStats(team_id = 1610612752)
NYK_stats_df = NYK_stats_df.get_data_frames()[0].tail(10)
NYK_stats_df["YEAR"] = NYK_stats_df["YEAR"].apply(lambda year: year[:4])
NYK_stats_df.to_csv("nyk_team_stats.csv", index = True)
```

#### Importing required files

## Processing data

In this section, I process and add to existing data frames to make it suitable for this analysis by:

- Making the extracted NBA API dataframes a consistent format with all other data frames.
- Manipulating the formatting of the team salaries dataframe and performing web scraping to retrieve up to date data.
- Filtering data to be within the analysis period (2013-2022).
- Initializing useful data frames for the future.

Cleaning column names from NBA API dataframes.

```
all_team_stats <- clean_names(all_team_stats)
nyk_team_stats <- clean_names(nyk_team_stats)
team_metrics <- clean_names(team_metrics)</pre>
```

I noticed there were 31 NBA teams as opposed to 30 in the all\_teams\_stats data frame.

```
length(unique(all_team_stats$team_name))
```

```
## [1] 31
```

This likely means a team changed their name throughout this period. To find this team, I must identify teams that didn't played the whole ten seasons under one singular team name.

```
all_team_stats %>%
  group_by(team_name) %>%
  count() %>%
  filter(n != 10)
```

As the Charlotte Bobcats became the Hornets in 2013 (i.e. team name change), I need to change the Bobcats name to the Hornets.

```
all_team_stats <- all_team_stats %>%
  mutate(team_name = recode(team_name, "Bobcats" = "Hornets"))

team_metrics <- team_metrics %>%
  mutate(team_name = recode(team_name, "Charlotte Bobcats" = "Charlotte Hornets"))
```

I combine individual player salaries into 30 NBA team salaries to keep the analysis on teams, not individual players.

Next, I perform web-scraping to get 2021 and 2022 NBA salary data, binding it onto the current salaries dataframe.

```
scrape_salaries <- function(html_link, year){</pre>
  page <- read_html(html_link)</pre>
  table <- page %>%
    html_table() %>%
    data.frame() %>%
    as_tibble() %>%
    select(Team, Total.Cap) %>%
    rename("salary" = "Total.Cap",
         "team" = "Team") %>%
    mutate(across("salary", ~gsub("\\$", "", .))) %>%
    mutate(across("salary", ~gsub("\\,", "", .))) %>%
    mutate(season = year)
  table$salary <- as.numeric(table$salary)</pre>
  returned_df <- salaries %>%
    bind_rows(table)
  return(returned_df)
}
salaries <- scrape_salaries("https://www.spotrac.com/nba/cap/2022/", 2022)
salaries <- scrape_salaries("https://www.spotrac.com/nba/cap/2021/", 2021)</pre>
```

As the analysis is conducted between 2013-2022, I filter for data from 29/10/2013 (start of 2013 NBA season) till present

```
game_df$game_date <- as.Date(game_df$game_date)
game_df$season_id <- as.character(game_df$season_id)
game_df <- game_df %>%
  filter(game_date > "2013-10-29") %>%
  mutate(year = substring(season_id, 2))

play_by_play_df <- game_df %>%
  select(game_id, game_date) %>%
  right_join(play_by_play_df, on = "game_id") %>%
  filter(game_date > "2013-10-29")
```

Next, I prepare a data frame on each NBA team with their wins, losses and win percentage between 2013-2022. This will be used frequently.

Table 1: NBA teams with their wins and lossess (2013-2022)

team_id	full_name	team_name	year	gp	wins	losses	win_pct
1610612737	Atlanta Hawks	Hawks	2013	82	38	44	46.3
1610612738	Boston Celtics	Celtics	2013	82	25	57	30.5
1610612739	Cleveland Cavaliers	Cavaliers	2013	82	33	49	40.2
1610612740	New Orleans Pelicans	Pelicans	2013	82	34	48	41.5
1610612741	Chicago Bulls	Bulls	2013	82	48	34	58.5
1610612742	Dallas Mavericks	Mavericks	2013	82	49	33	59.8

## Analyzing data

In this section, I will demonstrate the shortcomings of the Knicks over the past decade and identify key areas of improvement moving forth.

### A disappointing decade

When you ask any NBA fan to name the worst teams from the past decade, chances are the Knicks will be named. It's not hard to see why considering the Knicks rank third last (28th) in win percentage over the past decade, winning less than 40% of their games.

Table 2: Knicks win rank

team_name	avg_win_pct	win_rank
Magic	36.10	30
Pistons	37.76	29
Knicks	38.98	28
Kings	41.05	27
Timberwolves	41.16	26

Not only are the Knicks failing to win, but they're playing less than inspiring basketball. 80% of words containing sentiment used by commentators when reacting to Knicks games were negative words such as "miss", "bad" and "terrible".

Table 3: Knicks sentiment amongst commentators (2013-2022)

sentiment	n	proportion
negative positive	$31310 \\ 7726$	$\begin{array}{c} 0.8020801 \\ 0.1979199 \end{array}$

Compare this to a league average of 75% between 2013-2022 and you find that the Knicks are certainly not exciting with their style of play

```
knitr::kable(play_by_play_df %>%
    select(game_id, homedescription) %>%
    na.omit() %>%
    unnest_tokens(
        output = word, input = homedescription, token = "words") %>%
    anti_join(get_stopwords()) %>%
    inner_join(get_sentiments("bing")) %>%
    count(sentiment) %>%
    mutate(proportion = n / sum(n)), booktabs = T,
    caption = "League wide sentiment amongst commentators (2013-2022)",
    format = "latex", position = "h!")
```

Table 4: League wide sentiment amongst commentators (2013-2022)

sentiment	n	proportion
negative positive	774947 257918	$0.7502888 \\ 0.2497112$

The average NBA fan could have already predicted these results and so the question that should be asked is what can be done now and in the future to bring the Knicks glory over the next decade?

#### The key is the three

The tiles highlighted by the red box in fig 1 below show the correlation between common NBA statistics and win percentage. On the surface, one might believe that points would yield the highest correlation with winning given the goal of basketball is to score more points than your opponents.

Interestingly, the 3 point shot has the greatest correlation with winning at 0.58 among traditional NBA statistics

```
all_team_stats %%
  select(win_pct, fg3_pct, reb, ast, stl, tov, blk, pts) %>%
  cor_mat() %>%
  gather(-rowname, key = cor_var, value = r) %>%
  ggplot(aes(x = rowname, y = cor_var, fill = r)) +
  geom_tile() +
  labs(x = "variables", y = "variables") +
  scale_fill_gradient(low = "light yellow", high = "dark green") +
  geom_text(aes(label = r)) +
  geom_segment(aes(x = 0.5,xend = 8.5,y = 7.5,yend = 7.5),colour="#FF4500", size = 1.5) +
  geom_segment(aes(x = 0.5,xend = 8.5,y = 8.5,yend = 8.5),colour="#FF4500", size = 1.5) +
  geom_segment(aes(x = 0.5,xend = 0.5,y = 7.5,yend = 8.5),colour="#FF4500", size = 1.5) +
  geom_segment(aes(x = 8.5,xend = 8.5,y = 7.5,yend = 8.5),colour="#FF4500", size = 1.5)
```

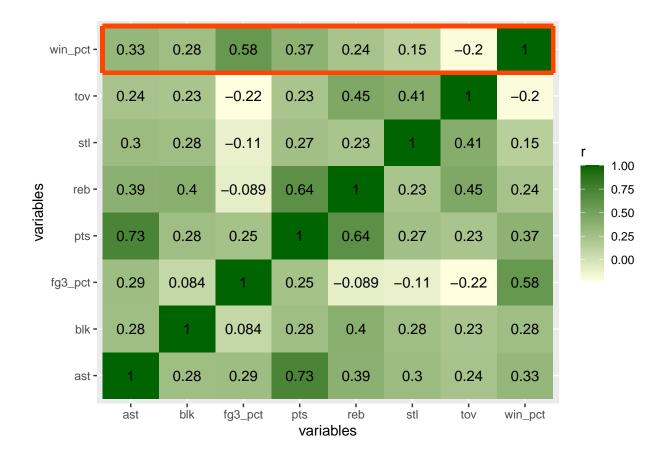


Figure 1: Correlation matrix of common NBA statistics

It's clear that a moderate positive correlation exists between win percentage and 3 point percentage, suggesting the better a teams 3 point percentage, the more likely that team will achieve a greater win percentage.

# Effect of 3 Point % on Winning

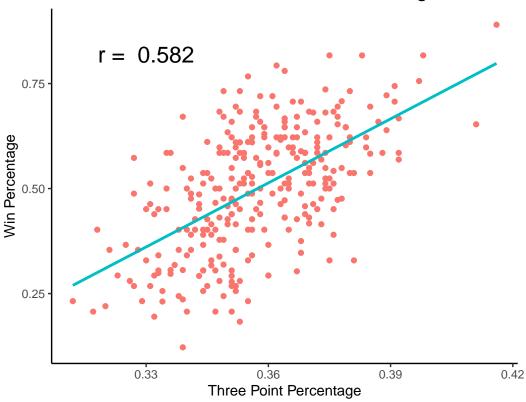


Figure 2: Three point percent vs win percent for NBA teams (2013-2022)

With the above analysis demonstrating how crucial 3 point percentage is in modern basketball, how have the Knicks performed in this key statistic?

Aside from two significant anomalies in 2013 and 2020, the Knicks have been a below average 3 point shooting team throughout the period 2013-2022.

## Change in three point percentage

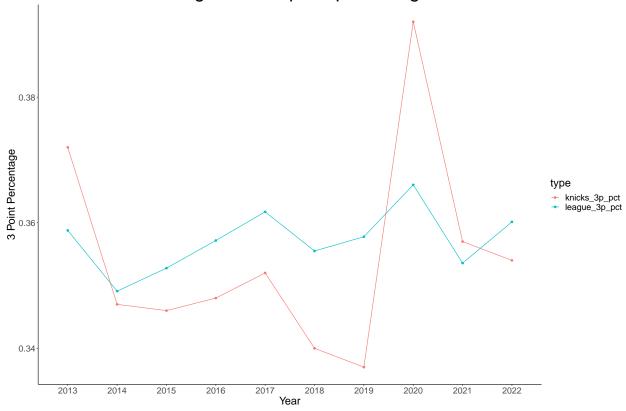


Figure 3: Knicks vs league avg 3 point percentage

Only 16 of the 30 NBA teams will make the "playoffs" to compete for the championship. When considering the importance of being at-least league average in terms of accumulating wins to make the playoffs and the significance of 3 point % on winning, the Knicks clearly have not placed enough of an emphasis on efficient 3 point shooting.

Not only have the Knicks shot poorly from three over the past decade, but they also attempt less threes compared to the league average.

The largest difference between the Knicks attempted threes in a season and the league average in that season was in 2019.

In 2019, the Knicks shot 600 less threes than league average and only won 31.8% of their games that season, indicating the importance of shooting more threes and with greater efficiency.

```
all team stats %>%
  group_by(year) %>%
  summarise(league_3s_attempted = mean(fg3a)) %>%
  left_join(nyk_team_stats, by = "year") %>%
  select(year, league_3s_attempted, fg3a) %>%
  rename("knicks_threes_attempted" = "fg3a") %>%
  pivot_longer(cols = c("league_3s_attempted", "knicks_threes_attempted"), names_to = "type", values_to
  ggplot(aes(x = as.factor(year), y = threes_attempted, col = type, group = type)) +
  geom_line() +
  geom_point() +
  theme_classic() +
  labs(title = "Change in three point attempts",
      x = "Year",
      y = "3 Pointers attempted") +
  theme(text = element_text(size = 20),
        plot.title = element_text(hjust = 0.5, size = 35))
```



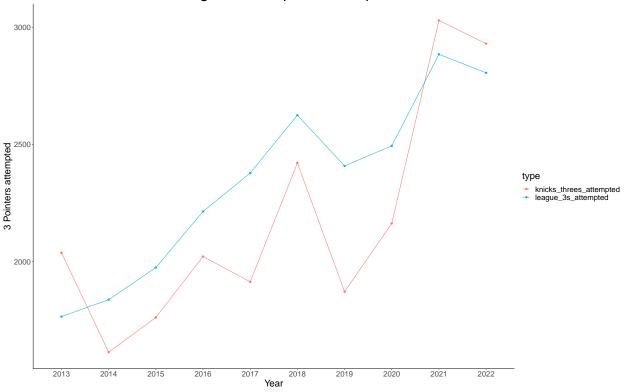


Figure 4: Knicks vs league avg 3 point attempts

One encouraging sign I observed from fig 3 is that over the past few seasons, the Knicks have been trending at or above league average in terms of three point efficiency, perhaps explaining an above 50 win percentage from 2020 onwards.

```
tail(nyk_team_stats, 3) %>% summarise(win_pct = sum(wins) / sum(gp) * 100)

## # A tibble: 1 x 1

## win_pct

## <dbl>
## 1 53.0
```

I recognize that 3 pointers are only one part of the game, but with how integral it is to the modern NBA, the Knicks ought to place greater emphasis on shooting the three ball more efficiency. This might occur through free agency signings, drafting players or training our current players to improve their three ball efficiency.

#### Prioritize defence or offense?

Net rating/efficiency is a formula that measures how efficient a team is on both offense and defense. Offensive net rating measures how many points a team allows per 100 possessions. Given the opponent the ball 100 times, how many points are they likely to score? The same principle applies with defensive net rating.

When exploring advanced NBA statistics, a 0.96 correlation between winning percentage and net efficiency stands out in fig 5 below. This should come as no surprise because the more you score and less your opponents score per 100 possessions, the better a chance you have at winning more games over the course of an 82 game NBA season. What is more interesting is whether defense or offensive is more important as we might like to know which is more of a priority when considering signing or drafting different players.

There is a moderate correlation of 0.62 between winning percentage and offensive efficiency rating and a moderate correlation of -0.52 between winning percentage and defensive efficiency rating. It's interesting to note that a teams offensive capabilities are slightly more linked with success compared to defense.

```
adv_cor_df %>%
    ggplot(aes(x = rowname, y = cor_var, fill = r)) +
    geom_tile() +
    labs(x = "variables", y = "variables") +
    scale_fill_gradient2(low = "#fb6767", high = "#3CB043") +
    geom_text(aes(label = r)) +
    geom_segment(aes(x = 0.5, xend = 9.5, y = 9.5, yend = 9.5), colour="#FF4500", size=1.5)+
    geom_segment(aes(x = 0.5, xend = 9.5, y = 8.5, yend = 8.5), colour="#FF4500", size=1.5)+
    geom_segment(aes(x = 0.5, xend = 0.5, y = 9.5, yend = 8.5), colour="#FF4500", size=1.5)+
    geom_segment(aes(x = 9.5, xend = 9.5, y = 9.5, yend = 8.5), colour="#FF4500", size=1.5)+
    theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

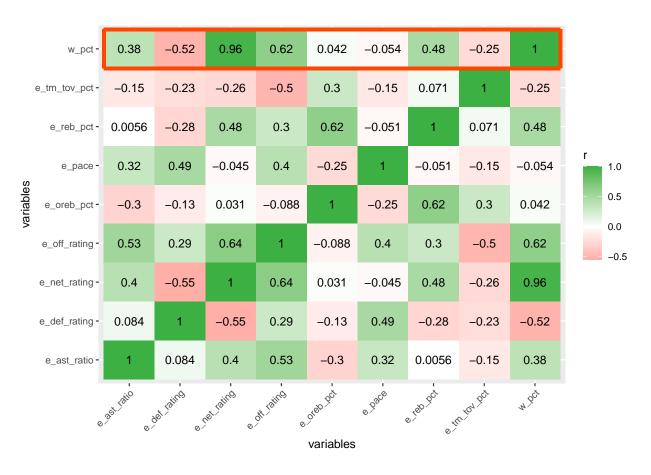
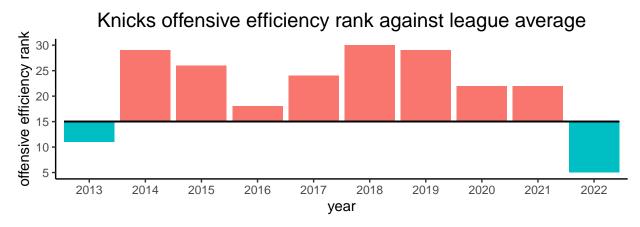


Figure 5: Correlation matrix of advanced NBA statistics

Now lets consider whether this difference in correlation between winning percentage and offensive efficiency rating compared to winning percentage and defensive efficiency has any impact on the Knicks winning.

```
plot1 <- eff_rank_metrics %>%
  ggplot(aes(x = year, y = e_off_rating_rank)) +
  geom_col(aes(fill = e_off_rating_rank<0), position = "dodge", col = "transparent") +</pre>
  scale y continuous(labels = function (x) x + offset) + theme classic() +
  theme(legend.position = "none", plot.title = element_text(hjust = 0.5, size = 15)) +
  geom_segment(x = 0.55, xend = 10.45, y = 0, yend = 0) +
  labs(title = "Knicks offensive efficiency rank against league average",
       y = "offensive efficiency rank")
plot2 <- eff_rank_metrics %>%
  ggplot(aes(x = year, y = e_def_rating_rank)) +
  geom_col(aes(fill = e_def_rating_rank<0), position = "dodge", col = "transparent") +</pre>
  scale_y_continuous(labels = function (x) x + offset) + theme_classic() +
  theme(legend.position = "none", plot.title = element_text(hjust = 0.5, size = 15)) +
  geom_segment(x = 0.55, xend = 10.45, y = 0, yend = 0) +
  labs(title = "Knicks defensive efficiency rank against league average",
       y = "defensive efficiency rank")
plot_grid(plot1, plot2, nrow = 2)
```



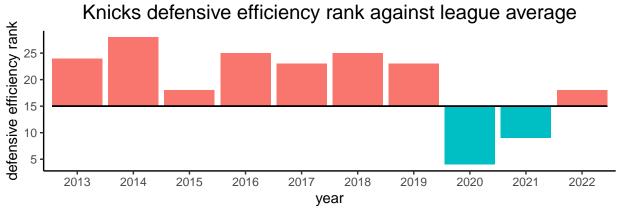


Figure 6: Knicks offensive and defensive efficiency ranking compared to league average of 15

From the two plots above in fig 6, one key observation is made:

• In no singular season have the Knicks been an above average team in both offensive and defensive efficiency. This implies might be a trade off between offense and defense, perhaps explained by the difficulty in finding elite players who can excel in both facets of basketball.

On the point of trade-offs, I would advise the Knicks to not give up one for the other but try to maximize their efficiency in total. As difficult as that sounds, both play a significant role in improving win percentage. This point is proven through table 5 below where:

- The knicks saw a reduction in defensive ranking between 2020-2021, holding offensive ranking constant leading to a reduction in win percentage
- The knicks saw a significant improvement in offensive ranking between 2021-2022, with a smaller drop in defensive ranking, leading to an improvement in win percentage.

The above two points demonstrate just how vital having a good ranking in both categories is to win percentage. This means the correlation difference between offense and defense as noted earlier should not be large enough to convince a team to entirely focus just offense, or just defense.

Table 5: Effects of changing efficiency ranking on win percentage

year	win_pct	e_off_rating_rank	e_def_rating_rank
2020	0.569	22	4
2021	0.451	22	9
2022	0.573	5	18

### Has the Knicks team salary hindered winning?

A common consensus in sports is that the more you spend, the better your team performs.

Despite its validity in practically every sport, there is no pattern to validate this claim for the 30 NBA teams over the 10 seasons played between 2013-2022.

Furthermore, the red dots in fig 7 below each of which indicate a salary and win percentage for one of the New York Knicks 10 seasons between 2013-2022 demonstrates no visible relationship either.

# Team salary against win percentage

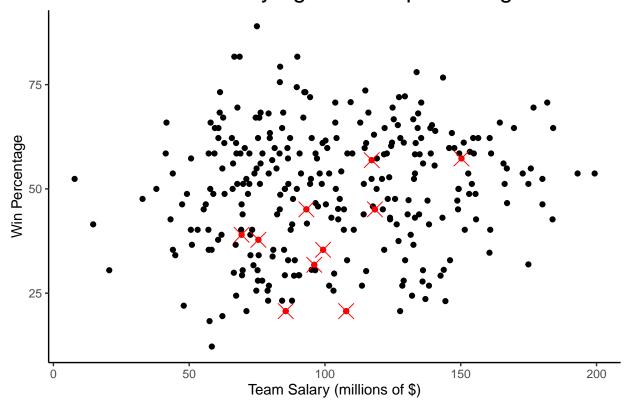


Figure 7: Team salary vs win percentage (2013-2022)

Rather than focusing on how total monetary spending directly relates to win percentage, we can look into how the Knicks team salary has compared to league average between 2013-2022 and its impact on winning.

The data shows that the Knicks have slowly reduced their team salary relative to league average between 2013-2022. Honing in on 2019-2022, the Knicks were well below league average in terms of spending and won 47.8% of games. Compare this to 2012-2018 when the Knicks were at or above league average salary but won just 33.1% of their games.

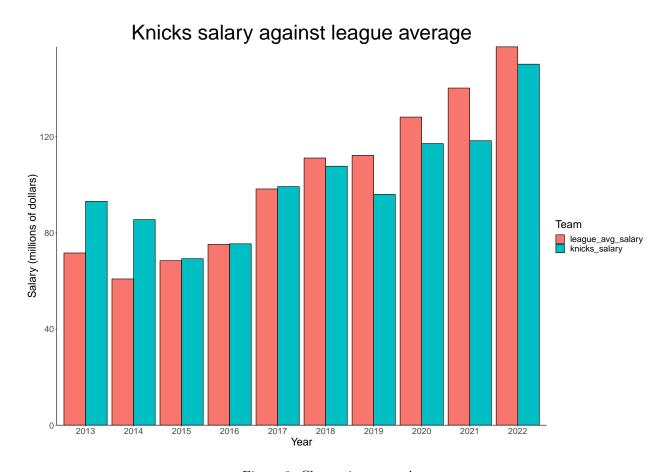


Figure 8: Change in team salary

Table 6: Knicks win percentage (2012-2018)

win\_pct 0.3311667

Table 7: Knicks win percentage (2019-2022)

win\_pct 0.47775

This salary analysis suggests that by keeping team salary below league average, the Knicks have tended to perform better. This could be for a multitude of reasons including:

- The development of young players drafted by the Knicks on low salaries. With the emergence of RJ Barrett, Immanuel Quickley and Mitchell Robinson as reliable players for the Knicks over the past few years, the Knicks have been able to keep players on rookie contracts. A lower salary, at least in the short term for the Knicks suggests good development of young talent who contribute to winning as there's less need to spend big on veteran players.
- Better management of player contracts. Since 2018, the Knicks have signed high calibre players like Julius Randle and Jalen Brunson on bargain contracts. The identification of under-valued talent has afforded the Knicks flexibility in trades and free agency, something the Knicks lacked between 2012-2017. A result of this includes the Josh Hart mid-season trade, propelling the Knicks to a 57.3 win percentage season in 2022.

## Acting on my findings

To conclude, I will come up with my top three recommendations that I believe will help the Knicks win in the future based on my analysis.

- 1. Invest available salary and draft picks in obtaining <u>"Three 'n D"</u> players who shoot the three pointer efficiently and play excellent defense. This is a sought-after mold that will add to three ball and defensive efficiency, two key metrics linked with success.
- 2. Focus equally on defense and offense, whether it be through signing particular players or in training drills. Being even just average in both facets guarantees a base level of success the Knicks have been consistently deprived of between 2013-2022.
- 3. Don't overspend on veterans in free agency but sign nice complementary pieces to our current star players Julius Randle and Jalen Brunson, allowing for flexibility when it comes to trading or signing players.