



MODULE CODE: DBA3702

Descriptive Analytics with R
Project Submission



Submitted By: SA2 Team 7
Power NBA Analytics

GROUP MEMBERS	MATRIC NUMBER
Corneles Jaemes D'silva	A0201737L
Janine Heng Sze Teng	A0217704J
John Alec Mendoza Branzuela	A0201504B
Kaung Htet Zaw	A0218258Y
Yong Wan Xin	A0222435N

1. Executive Summary	2
1.1 Background and Objectives	2
1.2 Team Win Rates	2
1.3 Team vs NBA Stats	2
1.4 Weighted ranks within Team	2
1.5 Player Statistics within Team	2
1.6 Player Valuations	2
2. Company Description	2
3. Market Research	3
3.1 Market Growth	3
3.2 Current Market Practices	3
3.3 Pain Points of Market Practices	4
4. Data Processing	5
4.1 Datasets	5
4.2 Data Cleaning	5
5. Our App	6
5.1 Overview of App	6
5.2 Team Win Rate across Seasons	7
5.3 Team vs NBA Statistics	8
5.4 Teams' Statistics	9
5.4.1 Weighted Rank Within Team	9
5.4.2 Player Statistics Within Team	10
5.5 Salary vs Performance	10
6. Marketing	11
7. Financials	12
7.1 Revenue assumptions	12
7.2 Cost assumptions	12
7.2 Overall projections	12
8. Conclusion	12
9. References	13

1. Executive Summary

1.1 Background and Objectives

The National Basketball Association (NBA) is a professional basketball league in North America, consisting of 30 teams from across 21 states of the USA as well as Canada. Power NBA Analytics is a startup focused on helping aspiring commentators improve their content on NBA through the data visualisation our dashboard offers. This market opportunity comes from the increasing popularity of alternative commentaries during live streamed sports events. Compared to existing solutions, our solution focuses on mass customization at very low costs.

1.2 Team Win Rates

Our first line graph compares two selected team's win rates per season ID against the league's average. This gives the commentator an opportunity to talk about the trend over time to critique certain decisions made by managers/coaches which happened when win rates changed.

1.3 Team vs NBA Stats

This spider plot compares the team's performance on a multitude of performance metrics against the league average. The commentator can choose to toggle the opposing team as well to see how each matches up.

1.4 Weighted ranks within Team

This customizable bar chart allows the commentator to input his ideas of what an overall good player looks like as defined by NBA statistics. Then, he can understand how the players in a team compare against one another and comment on how their strengths or weaknesses may come to complement each other. He could also talk about how players with a lower score could be trained to improve and comment on their progress over time.

1.5 Player Statistics within Team

This bar chart shows the relative performance of players in a team given a single metric. It could be used by commentators to further elaborate on what skills a player might be lacking in and how they might improve on that or how the other teammates could cover that weakness.

1.6 Player Valuations

During the break times of NBA games, the commentator could discuss possible replacements for certain players given their salaries and weighted skill set. They could also mention a fair salary for the players and comment on their justification for players' salaries.

2. Company Description

Our company, Power NBA Analytics is a new startup based in Singapore, where all our employees have a strong and keen interest in the NBA. Our mission is to help visualise the NBA in a way that helps aspiring commentators easily interpret past data and make discerning views through visualisations. Our vision as a company is to be the resource that anyone can turn to when needing important data analytics about the NBA.

3. Market Research

3.1 Market Growth

As the third most popular sport in the world, NBA has 2.2 billion fans globally (Das, 2022). Anticipating the popularity of player statistics among its fans, the NBA launched partnerships with Microsoft, SAP, and Turner sports to provide minute-level statistics for each game (Owusu, 2022). Currently, game data is being collected by many sources, including courtside cameras and even from players' wearable fitness trackers (Owusu, 2022). These devices provide information for machine learning, enhancing coaches' abilities to improve strategies and players' performance (Petra, 2020). Analytics can also be utilised for contract negotiation or refining in-game decisions and performance among others (NBASTuffer, 2022). Live commentators can also bring up these statistics and comment on these decisions and their expected outcomes to make it more interesting for their listeners.

As the NBA grew in popularity, the market for live sports broadcast changed. In a consumer report for live broadcast, it was revealed that sports fans' satisfaction for live broadcasts was only 39% for current solutions (Deloitte, 2019). This opened up opportunities for companies like Kiswe, which offer alternative audio commentaries. In 2021, Kiswe powered half a billion fans with multiple concurrent sport commentary streams in different languages (Nordeen, 2021), exemplifying the market's potential. Power NBA Analytics believes that this would open up a new market for aspiring commentators to demand low-cost, mass-customised data visualisations to provide better content for their listeners.

3.2 Current Market Practices

Current sources of statistical analysis are NBA stats resources as well as blogs like NBASTuffer that focuses on plus-minus matrices, player and team performances as spatial analysis. As seen from Fig 1, fans or critics are able to obtain statistical data from NBA stats, where they are able to filter by the season, team, positions, as well as the basis of comparison (per game, per minute, per number of possessions etc.). Column headers can be clicked on to sort players according to the individual statistic. Specific players and teams can be typed into a search function to single out their stats.

PLAYER	TEAM	AGE	GP	W	L	MIN	USG%	%FGM	%FGA	%3PM	%3PA	%FTM	%FTA	%OREB	%DREB	%REB
1 Duane Washington Jr.	PHX	22	4	4	0	9.0	40.0	40.0	56.3	0.0	100	0.0	0.0	0.0	0.0	0.0
1 Trevor Keels	NYK	19	1	1	0	1.0	33.3	0.0	100	0.0	100	0.0	0.0	0.0	50.0	50.0
3 Cole Swider	LAL	23	1	0	1	2.0	25.0	0.0	50.0	0.0	50.0	0.0	0.0	0.0	50.0	50.0
3 Corey Kispert	WAS	23	1	0	1	18.0	20.9	0.0	20.0	0.0	50.0	22.2	20.0	0.0	7.1	6.3
3 Marko Simonovic	CHI	23	1	0	1	2.0	25.0	0.0	33.3	0.0	50.0	0.0	0.0	0.0	0.0	0.0

Fig 1: NBA Stats

Next, in NBA Stuffer as shown in Fig 2, the user is able to pick multiple players from the drop-down menu respectively, to compare.

FULL NAME ▾

TEAM ▾

POS ▾

Donte DiVincenzo
Andrew Wiggins

RANK	FULL NAME	TEAM	POS	AGE	GP	MPG	MIN%	USG%	TO%	FTA	FT%	2PA	2P%	3PA	3P%
1	Donte DiVincenzo	Gol	G	25.76	3	14	29.2	18.6	32	4	0.75	5	0.6	6	0.333
2	Andrew Wiggins	Gol	F	27.7	9	33.3	69.4	20.4	5.5	22	0.727	75	0.52	54	0.389

Fig 2: NBA Stuffer Individual Players

As shown in Fig 3 for team stats, users are able to see a brief summary of their key stats compared to their opponents' along with a simple visualisation. They are able to single out a specific team using a search function as well.

RANK	TEAM	CONF	DIVISION	GP	PTS/GM	aPTS/GM	PTS DIFF	PACE	OEFF	DEFF	EDIFF	SOS	rSOS
1	Atlanta	East	Southeast	9	116.4	115.3	1.1	101.1	113.9	112.8	1.1	3.94	0.32
2	Boston						3.5	98.1	118.3	114.9	3.4	0.32	-0.18
3	Brooklyn						-0.8	97.8	113.5	114.3	-0.8	0.71	0.18
4	Charlotte						-4.4	100.7	107	111.3	-4.3	4.44	0.53
5	Chicago						2	100.8	111.4	109.4	2	-0.89	0.29
6	Cleveland						12.1	96.8	115.9	103.8	12.1	-3.54	-0.03
7	Dallas						6	95.2	117.8	111.7	6.1	-3.90	-0.13
8	Denver						1.9	100.2	116.1	114.3	1.8	-0.45	-0.22
9	Detroit						-11.6	100.3	107	118.6	-11.6	4.87	0.97
10	Golden Sta						-4	103	112.8	116.7	-3.9	1.17	0.63
11	Houston						-9	101	106.9	115.8	-8.9	2.20	0.06
12	Indiana						-1.8	101.9	113.7	115.4	-1.7	-2.51	1.04
13	LA Clippers						-3.9	98.8	103.6	107.6	-4	-0.70	0.80
14	LA Lakers						-5.3	102.6	104.4	109.4	-5	1.03	0.28

Fig 3: NBA Stuffer Team

3.3 Pain Points of Market Practices

These sources are useful in providing data that is easily filtered and sorted. However, they do not come with much visualisations as most of the data is presented in a table form consisting of a large amount of data. This can hamper analysis and usage of the data to come up with insights. Moreover, the weightage of each statistic cannot be customised to provide an aggregated standing.

Other sources of analytics platforms are Opta Analytics, Digital Scout, Mocap Analytics, Elias Sports Bureau and more. These platforms are often used by team managers to come up with custom reports of their own team players and are charged at a premium. Common features offered by these platforms include team and individual player reports, shooting charts and percentages, division rankings etc.

Our team would like to ride on the rise in data analytics in the NBA and live sports broadcast, and come up with an app that allows for mass customisation of data visualisations. We hope that commentators are able to arrive at insightful interpretations to deliver better content.

4. Data Processing

4.1 Datasets

We obtained the *game_data*, *game_details_data* and *rankings* through an API client for www.nba.com called *nba_api* which can be found here: (https://github.com/swar/nba_api). *Game_details_data* consist of players' raw statistics for each game, whilst *game_data* includes match details like the season, date and time.

Our team was unable to find salary data in this api and hence we did some web scraping to obtain the *players_salary* dataset from source: http://www.espn.com/nba/salaries/_/year/2021/page/

We used a python library named *Beautiful Soup* to parse the HTML data from the website above. From there, we were able to obtain the salary data we required for the year 2020-2021. The reason for choosing this library and Python as our language to scrape the data was because we had experience using this library before and wanted to make sure that retrieving the data we need for the project was not a bottleneck in completing our project.

4.2 Data Cleaning

We first merged *game_data* and *game_details_data* to obtain the season ID for each row in the players' statistics. Next, we filtered out for games in the 2020 season, as well as for players who had court time (played for more than 0 minutes). The main dataset we will be working with for the App is *game_details_2020*.

```
# Get data
game_details_data <- read.csv("data/games_details.csv")
game_data <- read.csv("data/games.csv")
ranking <- read.csv("data/ranking.csv")
players_salary <- read_excel("data/player_salary.xlsx")

##1. Replace '' in MIN in game_details with '0' and Convert MIN column to format HH:MM instead of HH:MM and convert to numeric
game_details_data <- mutate(game_details_data, MINS = gsub(':', '.', MINS))
game_details_data$MINS[game_details_data$MINS==""] <- '0'
game_details_data$MINS <- as.numeric(game_details_data$MINS)
##2. Add new column to show the season of each game played by matching season to game_ID in game_details_data dataframe
game_season <- game_data[,c('GAME_ID', 'SEASON')]
game_details_season <- merge(game_details_data, game_season, by = "GAME_ID", all=T)
# get subset of data for 2020 season
game_details_2020 <- game_details_season[game_details_season$SEASON == 2020, ]
# get subset of data for players who played >0 MINS
game_details_2020 <- game_details_2020[game_details_2020$MINS > 0, ]
game_details_data_with_season <- merge(game_details_data, game_season, by = "GAME_ID", all=T)
nba_2020_avg <- game_details_data_with_season[game_details_data_with_season$MINS > 0 & game_details_data_with_season$SEASON == 2020,]
nba_2020_avg <- nba_2020_avg %>% select('FG_PCT', 'FG3_PCT', 'FT_PCT', 'PTS', 'REB', 'AST', 'STL', 'BLK', 'TO', 'SEASON') %>% group_by(SEASON)
nba_2020_avg <- sapply(nba_2020_avg, mean)
```

Fig 4: Snippet of data cleaning

The raw game statistics we looked at was field goal percentage (FG_PCT), 3-point field goal percentage (FG3_PCT), free-throw percentage (FT_PCT), points (PTS), rebounds (REB), steals (STL), blocks (BLK), and assists (AST). For the latter five statistics, we analysed them per minute as that is a more accurate way of assessing a player rather than using the absolute number.

To analyse the teams' performance, we filtered out the data by the team ID and aggregated the data.

5. Our App

5.1 Overview of App

Our App consists of 5 components. The first component of the App features a summary of a selected team's win rates across seasons. The second component presents users with a consolidated view of a team's performance compared against the NBA average. The third and fourth component allows users to analyse performance within a team, with respect to overall performance and individual skills. The last component features a valuation of players' performance against their respective salaries. Aside from the five features listed above, our application also includes a screenshot feature and an 'About Us' tab to enhance the user experience.

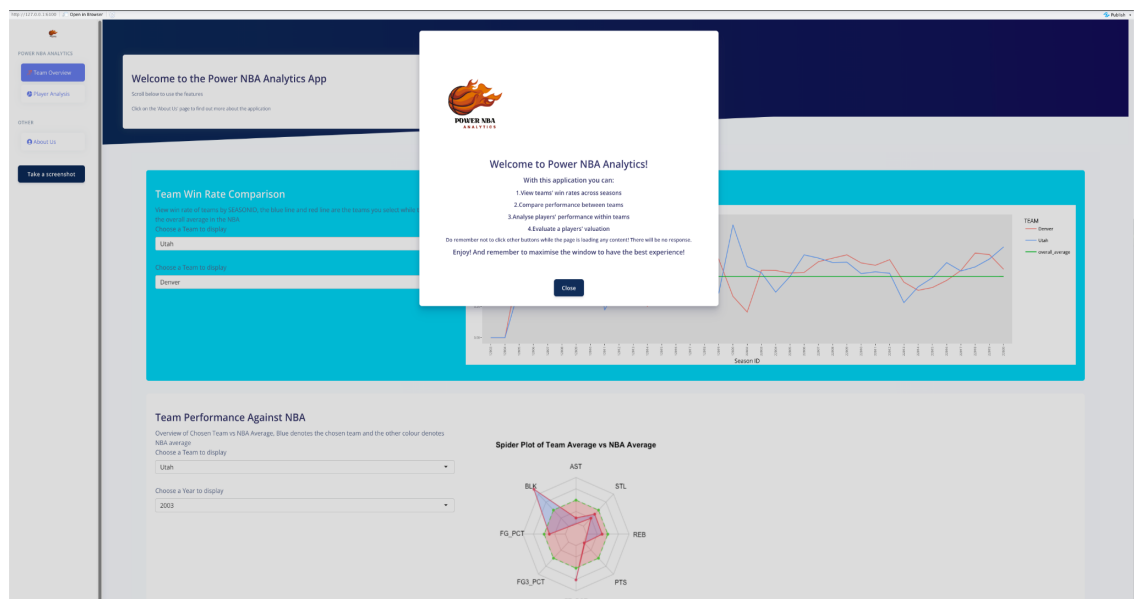


Fig 5: Application State when first started

5.2 Team Win Rate across Seasons

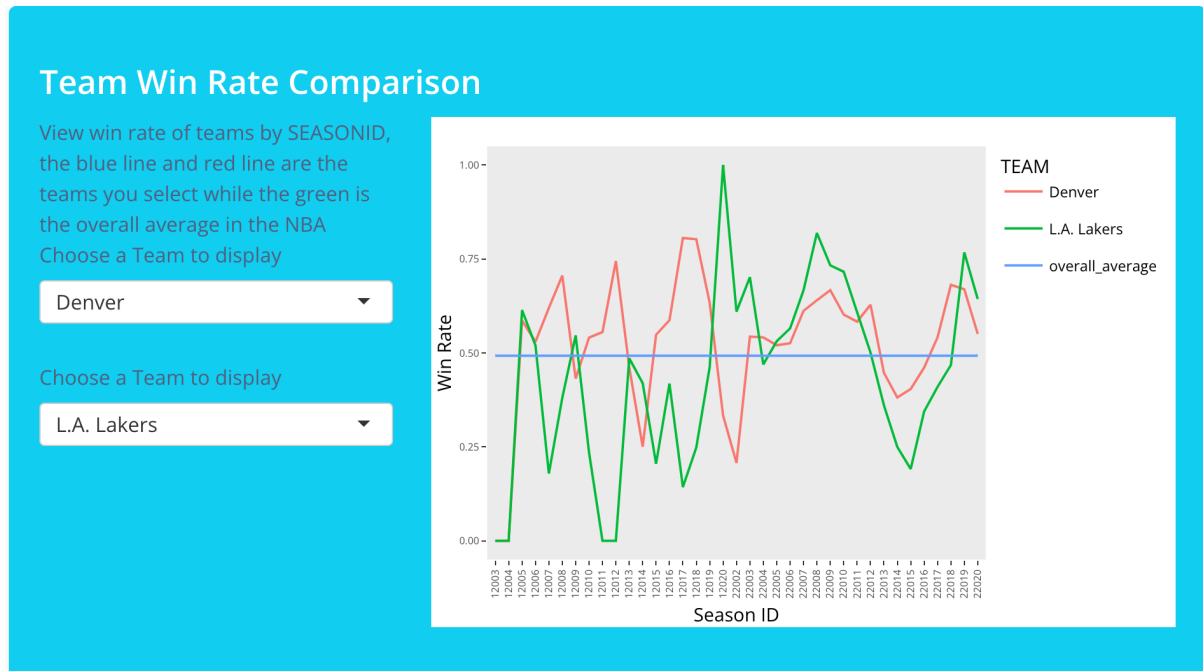


Fig 6: Win Rates of two team Across Seasons

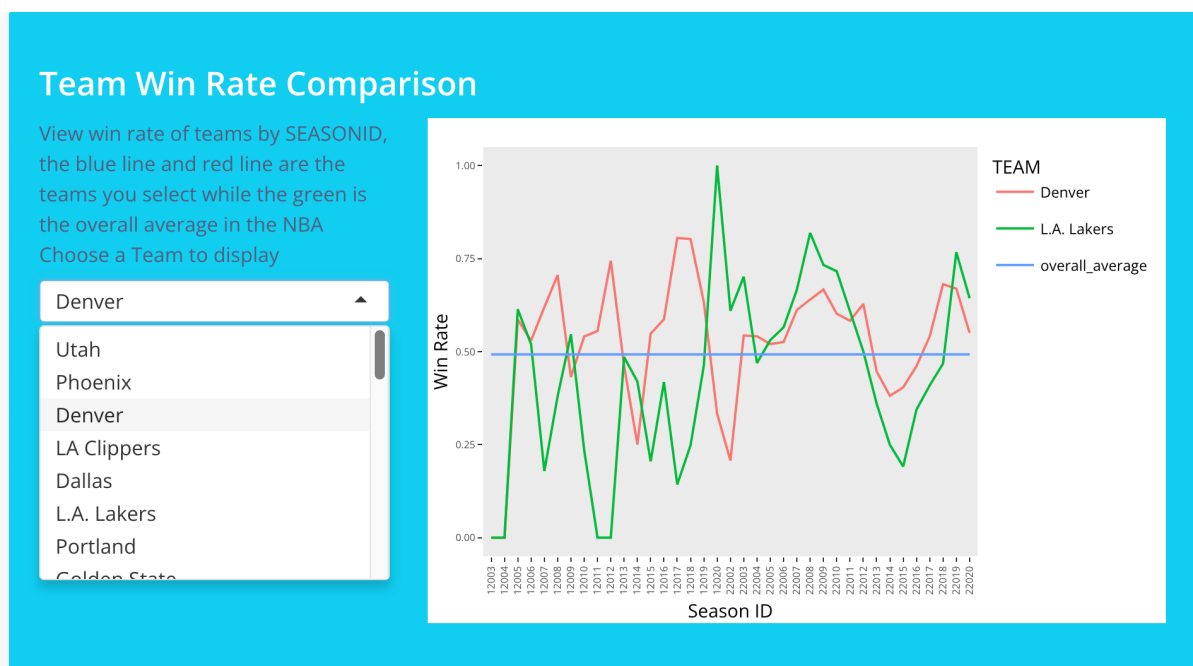


Fig 7: Drop down lists

Question(s) to answer: How did two teams perform in previous seasons?

In the first feature of the app as seen in Fig 6, users will be able to select two teams and visualise the team's average win rates across different seasons of the NBA.

The teams' average win rate is compared against the league's average win rate across seasons for a better gauge of the teams' performance within the league. Amateur commentators may use this feature of the app to gain a broad understanding of two team's performance across seasons to evaluate or comment on their current performance. They could also look for periods where the win rate changed drastically to comment on certain decisions or actions which the managers or coaches took during that period.

The graph shown depends on the user's input from the drop down list of team names as seen in Fig 7. The app then renders a line graph based on the teams selected and shows the trend of both teams average win rate through the season that they played in. For example, the default team is 'Utah'. If the user selects the team L.A Clipper, the graph changes to match the team as seen above in Fig 7. The red line is the average win rate through the entire NBA for all the seasons.

5.3 Team vs NBA Statistics

Question(s) to answer: How does the selected team perform compared to the rest of the league based on the default statistics?

Under this feature, users can select a team and season to toggle a spider plot superimposing the chosen team's performance against the league's performance in a chosen season. The NBA average is taken as the baseline data (100%) to compare with the chosen team's performance. The value for each statistic is obtained by taking the mean across all games for fair comparison. Using the spider plot as seen in Fig 6 then allows the user to understand how the team is faring in each metric compared to the NBA average.

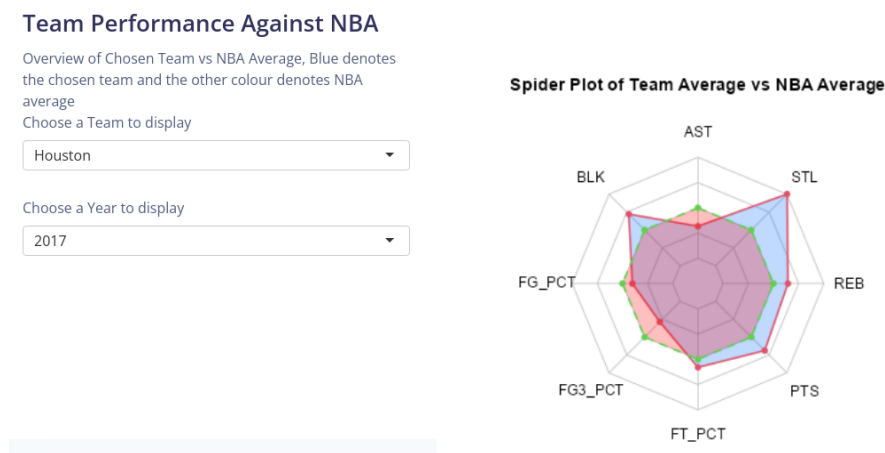


Fig 8: Comparison of Chosen Team vs NBA Average

Commentators can use this feature to understand and predict how a team will be performing for their upcoming games. They can easily see from the visualisation in Fig 8, which component the team is performing better or worse than the NBA average. From here, commentators might want to analyse performance within a team.

5.4 Teams' Statistics

This feature allows users to analyse a team's play.

5.4.1 Weighted Rank Within Team

Question(s) to answer: Who are the better performing and underperforming players in a team?

In this component of the app, users select a team and are able to visualise each player's overall performance (weighted rank) within the team.

Weighted Ranks within Team

View weighted ranks within the team
chosen

Choose a Team to display

WAS

FG_PCT

0.2

FG3_PCT

0.3

FT_PCT

0.1

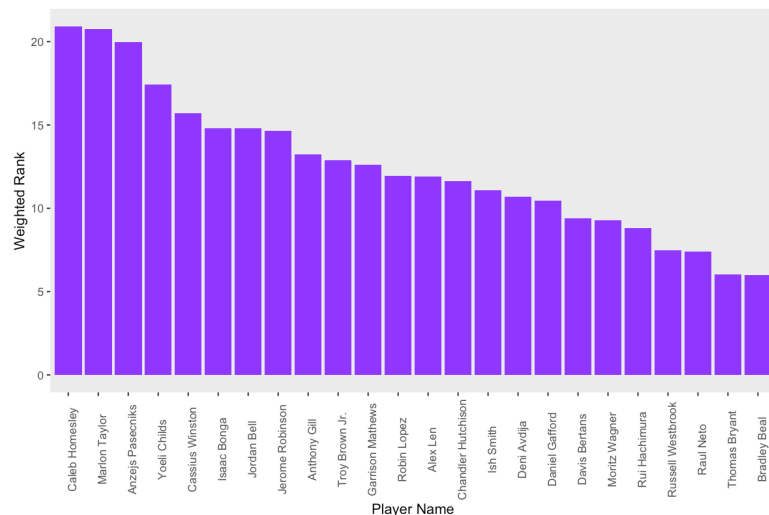


Fig 9: Weighted Rank of Players in a Selected Team

The players' performances are summarised by their weighted rank per skill (statistic). Specifically, players are ranked within the team for each skill, with a higher rank (which is denoted by smaller number) corresponding to a higher value for the statistic of the skill and better performance of the skill. An overall performance metric is then derived by taking the weighted average of the rank of each statistic for each player.

Users are able to decide and select the weights for the rank of each statistic (by keying in a value $0 < \text{weight} < 1$, summing up to 1) according to what they deem to be more important in a team's play or their opinion of the current basketball environment. For example, based on their opinion of the shift in NBA basketball trends, they might favour 3 pointers over mid-range shots. Therefore, users may choose to place a greater weight on the rank for FG3_PCT.

The app will then generate a bar plot of the overall performance of each player in the team as summarised by their weighted rank per statistic as seen in Fig 9. The graphs are ordered in descending order for ease of identifying the underperforming and better performing players.

Amateur commentators may use the information to make their commentary more insightful and interesting by adding comments on a player's performance within his team, or even commenting on a coach's decision to use a particular player. Alternatively, amateur commentators may select the better performing or underperforming players to conduct additional research on, finding out the story behind a player's over or underperformance.

Based on the overall performance of the players, the commentators may want to look into players' performance by individual indicators.

5.4.2 Player Statistics Within Team

Question(s) to answer: Which player(s) to target for each skill of concern?

Under this component of the app, users select a team and a statistic of concern and are able to visualise each player's performance in the selected aspect within the team as well as identify the underperforming few.

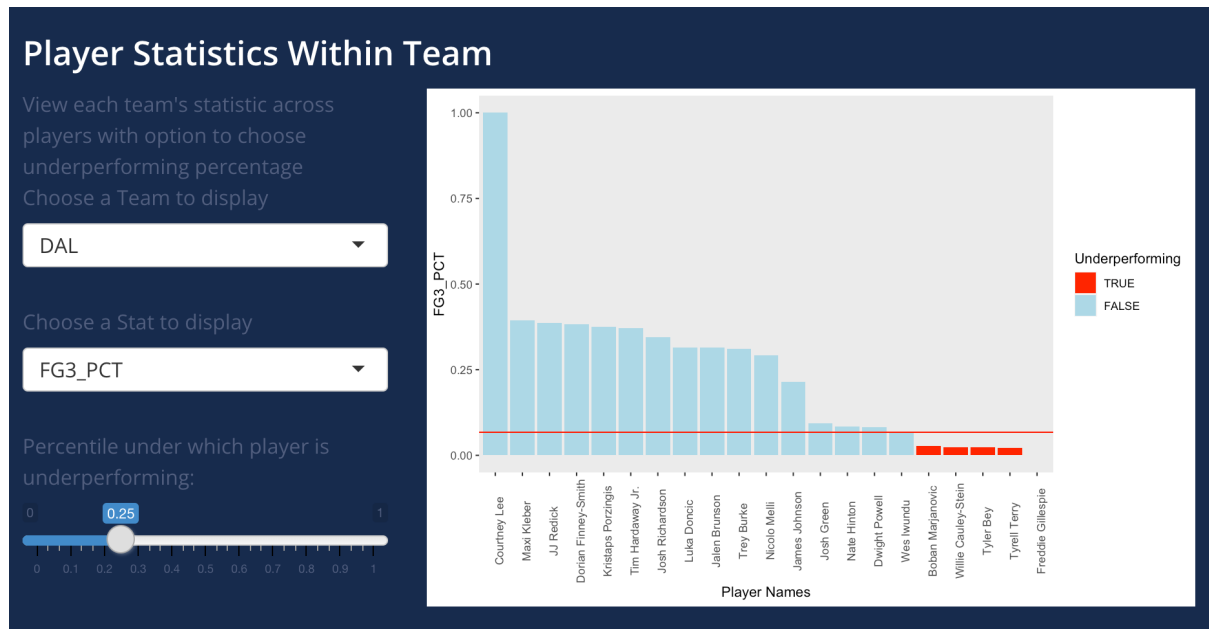


Fig 10: FG3_PCT of Players in a Selected Team

Users are able to select the statistic of concern from a drop down list. From the spider plot feature summarising a team's performance, users are able to identify areas in which the team is underperforming and may further narrow down their analysis in this component of the app. For example, the spider plots may identify that a team is underperforming in the aspect of 3 pointers (FG3_PCT). Users may then select FG3_PCT as the statistic of concern to analyse the team's players' performance in this aspect.

Users are also able to select the percentile under which a player is underperforming in the team. For example, a value of 0.25 selected on the slider would help to identify the bottom 25% in the team for the selected statistic as represented below the red line as seen above in Fig 10. Amateur commentators may use this information to recommend improvement strategies for the identified players accordingly. For example, commentators may recommend 3 point shooting drills for the underperforming players. Commentators are given the option to select this percentile as they may have their own definition of underperformance in the team. Alternatively, commentators may corroborate the overall underperforming players from the weighted rank component of the app with the underperforming players in the statistic of concern to make further recommendations such as to replace the player with a more cost effective player.

5.5 Salary vs Performance

Questions to answer:

- 1) How do we know if a player is properly compensated?
- 2) Which players are overpaid/underpaid considering their skillset?
- 3) Which players are good, cost-effective replacements for others?

In this feature of the app, users are able to visualise the salaries of each player in the league against their overall performance (as weighted by the commentators).

Player Valuation

View weighted ranks against mean salaries in 2020. The regression line could be an indication of fair value for a player of certain weighted rank. Any player paid more could be overvalued and switched for a player paid less. Use your mouse to highlight and Zoom into selected players, represented by the points.

FG_PCT

FG3_PCT

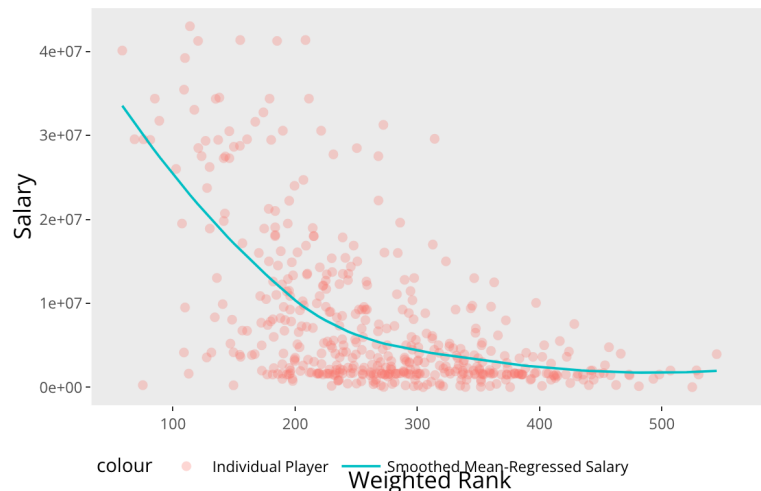


Fig 11: Scatter Plot of Player Salary Against Weighted Rank

In this Fig 11, users can quickly identify players through a dropdown list at the side. This will highlight the data point corresponding to the player's skill level and salary. A "weighted rank" of our 8 statistics can be decided based on user input. Users can then zoom into the player's skill level or salary level to identify potentially similarly skilled or paid players. This graph allows for customised zooming into regions. The blue line is a locally weighted smoothing regression line which shows an estimate of a "fair salary" for the given weighted rank which the user inputs.

Commentators are able to identify players in the league who have better performance and lower salaries, thus higher cost effectiveness, as compared to their player of concern. This may act as a baseline to help commentators make player replacement recommendations. The commentators are also able to identify the amount that players of a certain skill level should be fairly compensated, assuming that compensations are based on skills alone.

6. Marketing

To boost the adoption of our application, we would consider partnerships with Kiswe, an interactive live broadcast video application, to provide these data analytics capabilities to all their NBA commentators. There are two options we can utilise to provide this service: Kiswe could either purchase our services at a bulk discounted price or offer our services as an add-on to commentators who can subscribe using a monthly premium.

We could also boost the market awareness for Power NBA Analytics through partnership with NBA influencers like ballislife or CliveNBAParody across Instagram and Youtube. These accounts have 3.6 and 1.95 million followers on Youtube respectively, which could widen the app reach if commentaries are made using our app's visuals.

Since this app is focussed on the mass-market, we could also open it for download on Google Play or Apple stores for any fans to download and do their own analyses.

7. Financials

7.1 Revenue assumptions

Given the app's B2C nature, we could offer the application at a price of \$1 per download as most apps on the Google app store, 62% of them, are priced below \$2 (Statista, 2022). We estimate that in our first year, the app downloads could reach up to 500/day (Bluecloud, 2015). Assuming we charge \$1 per download, that would be \$182,500 in revenue a year.

7.2 Cost assumptions

It costs about \$40,000 to develop an app with medium complexity in 4-6 months (Dogtiev, 2022). Since we are building apps for the Apple Store and Google Play, that would cost a total of \$80,000. To maintain the app, it would cost around \$250-\$500 per month (Georgiou, 2022), so we take the mid-point \$350 per month as a conservative estimate. In total, the first year costs are expected to hit \$40,000 for the first 5 months followed by \$350 for the next 7 months, totalling \$42,450. In the next 5 years, assuming just maintenance costs matter, each year's costs would be $\$350 \times 12 \text{ months} = \$4,200$.

7.2 Overall projections

Given the uncertainty in projecting beyond 3-5 years for a new company, we estimate that Power NBA's net profit would be: \$140,050 in the first year and \$178,300 in the next few years.

8. Conclusion

To conclude, our application is expected to be highly popular among the commentators at Kiswe. We provide affordable data analytics capabilities with mass-customization; perfect for any aspiring commentator. In the future, we hope to leverage on the client base to better understand what other data points would help enhance their commentaries. At Power NBA Analytics, we hope to grow our capabilities with our customer's needs.

9. References

- NBA Stuffer. (2022). *Analytics Movement*. Retrieved November 11, 2022, from <https://www.nbastuffer.com/analytics101/nba-analytics-movement/>
- Sponsor United. (2022). *NBA 2021-22 Marketing & Partnerships Annual Report*. Retrieved November 11, 2022, from <https://sponsorus.com/nba-2021-22-report/#:~:text=NBA%20teams%20surpassed%20%241.2B,a nd%20the%20emergence%20of%20Cryptocurrency>
- Owusu, T. (2022, September 28). *Microsoft, NBA to Deliver New App Basketball Fans Will Love*. TheStreet. Retrieved from <https://www.thestreet.com/investing/nba-unveils-next-generation-of-fan-engagement>
- Petra. (2020, April 21). *How data analytics is revolutionizing the NBA*. HBS Digital Initiative. Retrieved from <https://d3.harvard.edu/platform-digit/submission/how-data-analytics-is-revolutionizing-the-nba/>
- NBAstuffer. (2022). *Analytics Movement*. Retrieved from <https://www.nbastuffer.com/analytics101/nba-analytics-movement/>
- Deloitte. (2019). *Enhancing Digital Fan Engagement*. Retrieved from <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/technology-media-telecommunication/us-enhancing-digital-fan-engagement.pdf>
- Nordeen, C. (2021, March 17). *Kiswe Powers Over Half a Billion Fan Engagements Around the Globe in 2020*. Retrieved from https://www.prweb.com/releases/kiswe_powers_over_half_a_billion_fan_engagements_around_the_globe_in_2020/prweb17802046.htm
- Ceci, L. (2022, November 11). *Paid app price distribution in the Google Play store as of November 2022*. Statista. Retrieved from <https://www.statista.com/statistics/271109/average-price-android-apps/>
- How Many Downloads Should My App Get?* Bluecloud Solutions. (2015, February 6). Retrieved from <http://www.bluecloudsolutions.com/2015/02/06/how-many-downloads-should-my-app-get/>
- Dogtiev, A. (2022, October 31). *App Development Cost*. BusinessofApps. Retrieved from <https://www.businessofapps.com/app-developers/research/app-development-cost/>
- Georgiou, M. (2022, June 30). *Cost of Mobile App Maintenance in 2022 and Why It's Needed*. Imaginovation insider. Retrieved from <https://imaginovation.net/blog/importance-mobile-app-maintenance-cost/>