SOCY2169 Online Research Methods: Assessment 3 – Research Project

Introduction

The NBA (National Basketball Association, played in the USA) is the most important basketball league in the world, having more than 10 million viewers at its peak, not only from the United States, but also from different countries around the world. However, this figure is only reached in June, during the finals, as it averages 1.6 million viewers throughout the season with its worst moment in November when it is just beginning.

So, in an effort to provide more entertainment for its current fans and to attract new ones (and no secret, as the business that it is) especially during its lowest viewing period, the league has decided to create a new tournament within the regular season itself during November and December in an attempt to break the monotony of recent years in these months.

However, as with any new decision, there are those who support it and those who do not. Despite the publicity and the effort made to make people understand the format, it was not very intuitive, and many people complained at the beginning about the idea of more games being played (as it implies a higher risk of injury to the players, considering that there are already 82 regular season games plus playoffs). The reality is that the games in this tournament count for the regular season as well, except for the final, so only 2 teams will play a single extra game.

Taking into account that apparently the rejection of the initiative was due to misinformation at the time of the announcement, the objective of this project is to identify how the general perspective of NBA fans has changed 4 months before the start of the tournament and the same day of its start, as well as to identify the main reasons of the public that was (or is, if those reasons are still valid) against it; as these new proposals are important within the NBA community and can be fundamental for new ideas to be implemented in the future in search of the continuous improvement and expansion of the league.

Literature review and research questions

Clearly, new decisions in the NBA and the impact they have on the league and its environment are not in the interest of everyone, but all of this greatly affects the large number of fans, workers and players that make up the league. Within it, the importance of this new tournament is evident in the constant search that the NBA has been carrying out in recent years as an initiative to increase its reach, especially after the fall presented during the pandemic. Articles such as "The NBA is considering its own FA Cup. the results may not be pretty 2019" and "Overton, How the NBA can regain viewer engagement - the daily Utah Chronicle 2021" confirm the interest shown in an innovation around 2020.

Also, as a tournament that started only a few days ago and is still in play, no articles or research related to it can be found, but this represents a clear opportunity to draw one of the first conclusions about it.

Because all of this, it is obviously important to analyse the general opinion about it, in order to get the first answers to: Has the general opinion of some Reddit users and NBA fans about the new in-season tournament changed in a period of 4 months before its start?, if so, how has this changed?; what were the main reasons of the opposition for not supporting the initiative at the beginning?, have these reasons been maintained over time?

Data & methods

The data will be collected from two Reddit threads. The first one was posted on July 9, 2023, right after the official announcement of the tournament in the United States, with the literal title "How to you feel about the nba's new in season tournament?", this thread facilitates the information as it is directly asking for people's opinion about the tournament and has a considerable amount of answers that will allow us to carry out the analysis (this thread will be identified as "the first thread" and all the information with a "1" will be related to it.).

The second thread is called "In-season Tournament tips off in Indy" and was published on 4 November 2023, the same day as the tournament starts (as the name indicates) and although it does not ask any questions, we can see how the users who participate in it leave comments about what they think about the tournament. With a similar amount of activity, it

is the right candidate to use for the project (this thread will be identified as "the second thread" and all the information with a "2" will be related to it.).

Both threads were found by searching on Reddit the words "In-season tournament" and filtering by date to get the desired results for comparison.

Initially, the information was downloaded from the two Reddit threads, obtaining a total of 287 comments in the first thread and 424 in the second. With this, two activity networks were created, which, although it was not essential for the subsequent sentiment analysis, they are a good technique to graphically approach the downloaded information. Activity networks were chosen because although the actor networks also have the text of the comments as an edges' attribute, the activity network allows us to observe the distribution of the comments within each thread, as well as the responses to each reply in a better way. The networks can be seen in *Figure 1* and *Figure 2* and their node and edge information can be found below:

First thread:

- **Nodes:** 288
- **Edges:** 287

Second thread:

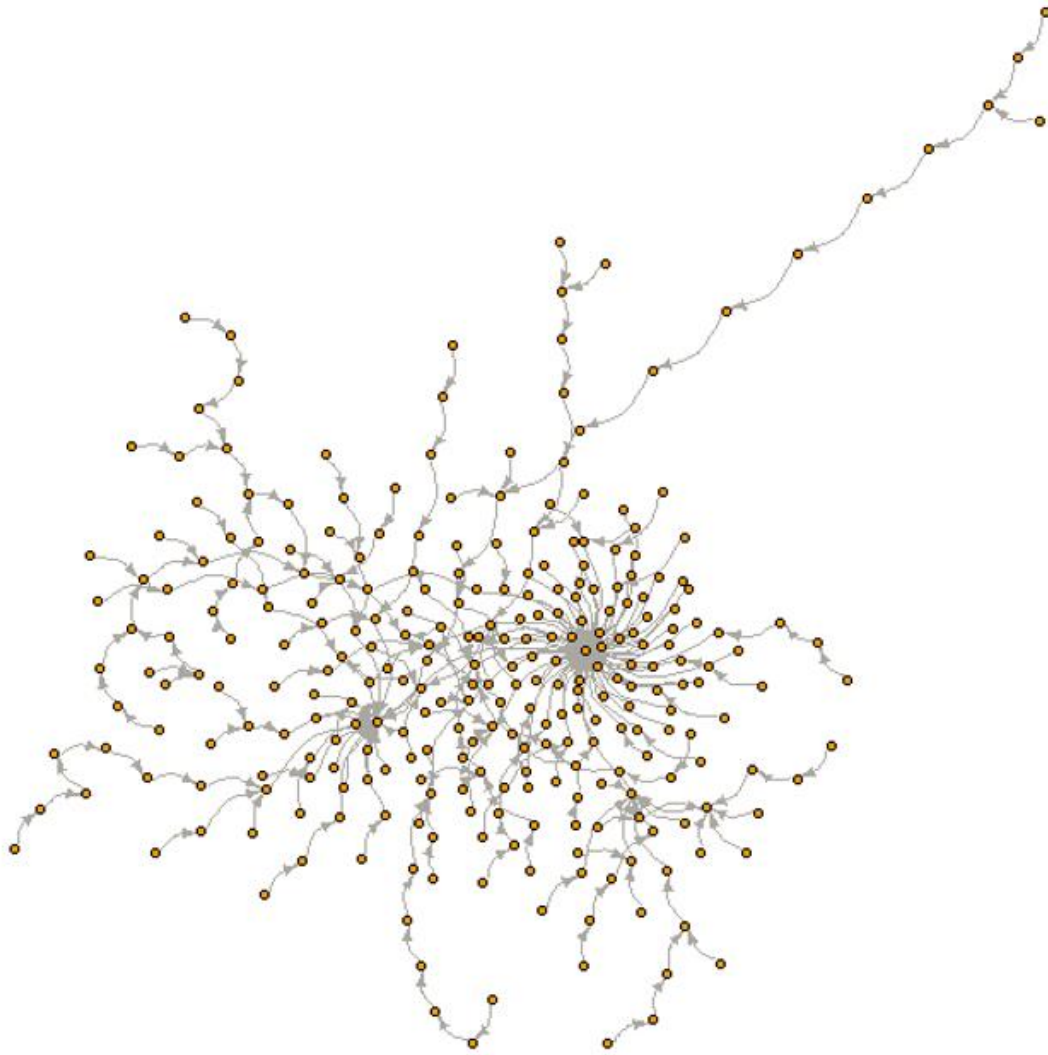- **Nodes:** 425
- **Edges:** 424

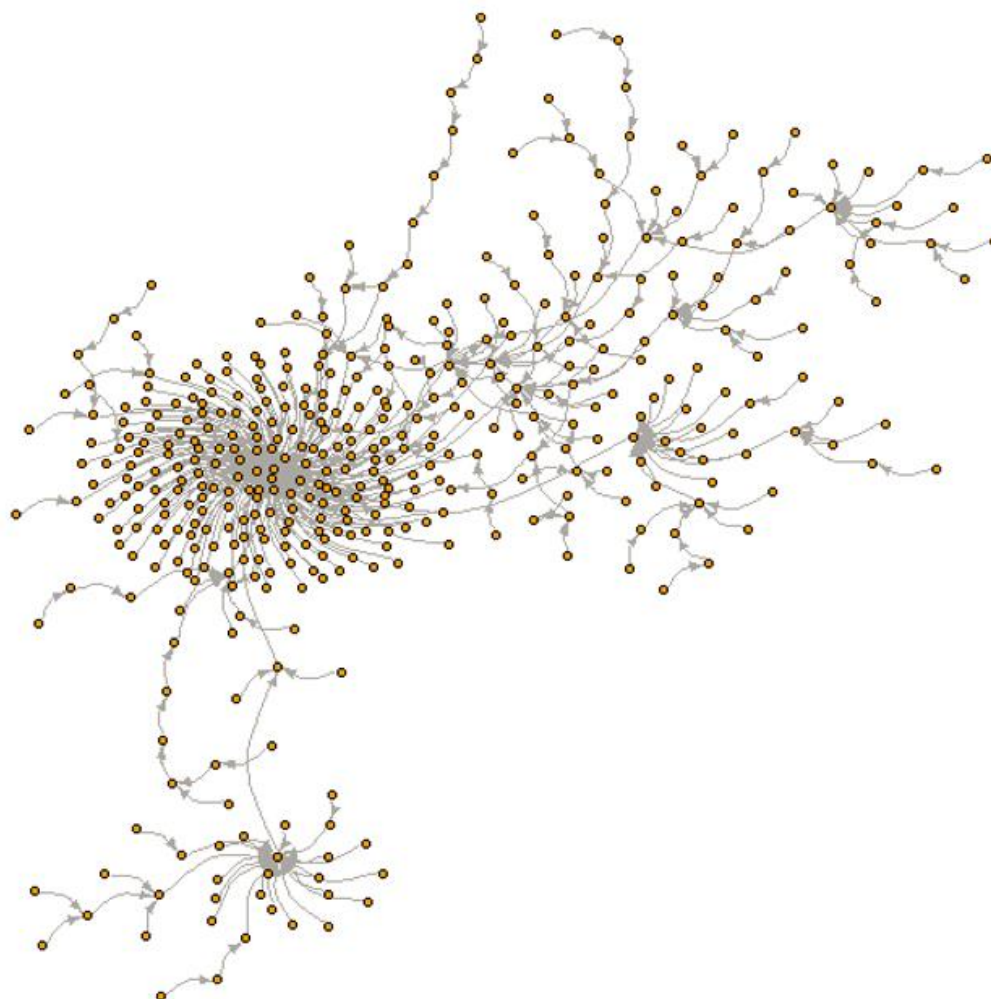*Figure 1. Thread 1 Activity network*

Figure 2. Thread 2 Activity network

Once this was done, a preliminary analysis was made to identify the main topics in each of the two threads, in order to make a first approach to the information they presented. In this technique, a corpus was created for each thread with a document for each comment collected, in addition to a basic cleaning by eliminating punctuation, separators, numbers, etc. Tokens were extracted by separating them by bigrams and trigrams, to finally make a word cloud with the 50 most frequent tokens.
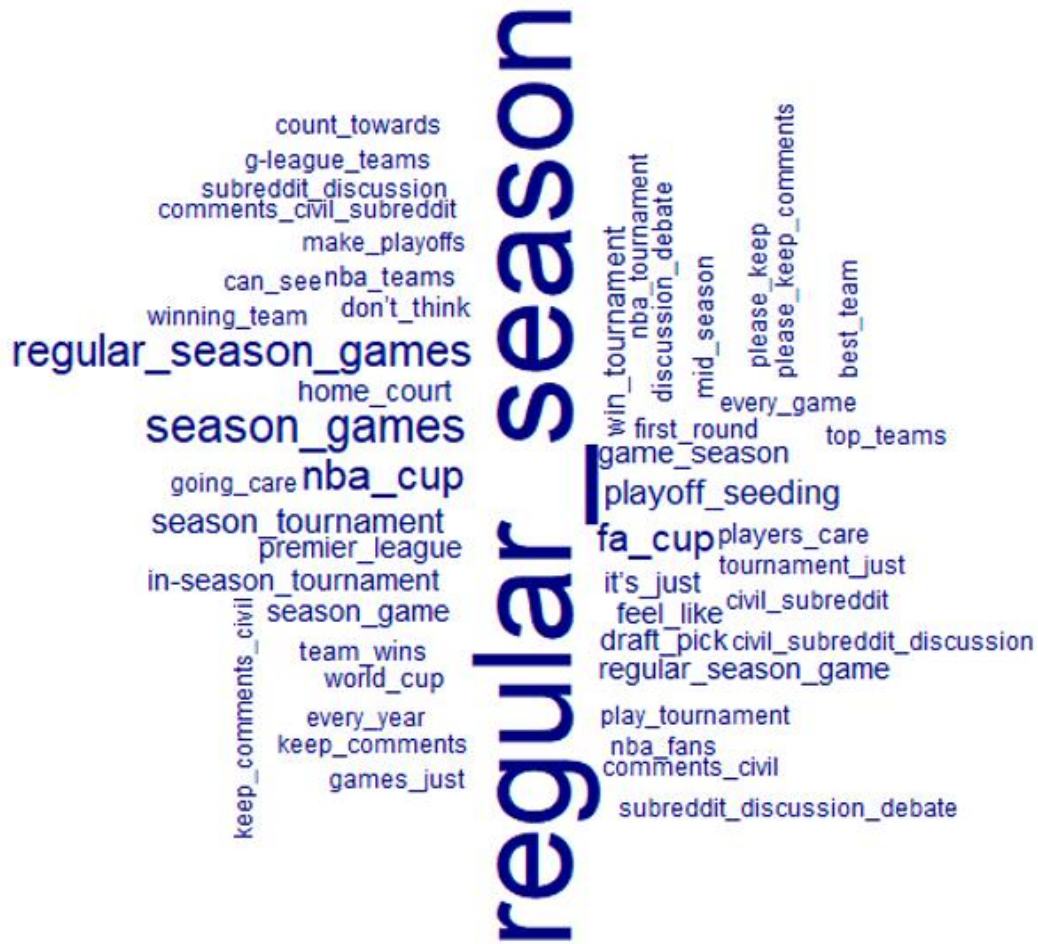
Figure 3. Thread 1 Main themes Word cloud

Figure 4. Thread 2 Main themes word cloud

In *Figure 3*, we see how in the first thread the main topics are about "regular season games", probably because at the time there were still evidently doubts about the tournament format, leading to many clarifications that in-season tournament games would also count as regular season games. In the second thread (*Figure 4*) some negative words are noted, apparently indicating a bad appearance of the courts, with phrases such as "court looks like" and "looks like shit". Clearly, it would be too hasty to draw any conclusions, but this process gave us an idea of what to expect going forward.

We want to know the main themes to identify the main reasons for supporting or not supporting the tournament, but this is why we want to relate these topics to a positive or negative sentiment towards the initiative, and this is why the best way is with using sentiment analysis. It is important to clarify that the whole process was done equally with

each of the two threads, so despite speaking in singular, each procedure was done with the first thread and then with the second.

To begin with, we started again from the corpus that we already had and tokenised it again, in this case it is not necessary to perform a thorough cleaning of the data because, when categorised by sentiment, in case there is for example a number or a URL, they will simply be assigned as "neutral" and then they will be removed. Each token was evaluated with the "data_dictionary_LSD2015", and with this, the "logit scale sentiment score" seen in class was calculated to give a score to the sentiment indicated by each comment. This was used to average the sentiment score for each comment and give us an approximate value of the overall sentiment within each thread.

As we are interested in people's reasons to justify their opinion on the tournament, and whether it is related to a feeling of support or rejection (positive or negative), it is a good time to conduct a sentiment analysis on the keywords. For this, we obtained the value of the range, which calculates the difference between negative and positive tokens in a comment and assigns it a value. In order to simplify the information, this value is qualified with the following criteria:

- **"Positive"** if range is greater than 0.
- **"Neutral"** if range is 0.
- **"Negative"** if range is less than 0.

Finally, a comparison word cloud is made with the most frequent words related to "positive" and "negative" (words categorised as "neutral" are not taken into account, as they do not express any feeling and therefore do not contribute to the research). Bigrams and trigrams were again used for the comparison word clouds. Although the use of unigrams gave us more tokens with a higher frequency, the information it gives is only single words, and although it is true that it gives us a hint of what it might mean, by using two- or three-word phrases we can get more information about people's reasons for their opinion, which is what we want, which is why the process was done in this way.

Results & discussion

As previously mentioned, the averages of the values obtained for the sentiment score in each thread were calculated, with the following results:

- 0.53 for the first thread.
- -0.108 for the second thread.

With these results we could say that popular opinion, at least as far as Reddit users who follow the NBA are concerned, has diminished from the announcement of the in-season tournament to the date of its start. However, it is possible that the reasons may provide us with a little more information, so let's look at the comparison word clouds for both threads before concluding completely.

negative

lottery_teams mid_season_tournament
try_win load_management
every_team make_games
every_game game_season
going_care playoff_spot playoff_seeding
cash_grab get_stale don't_think
feel_like top_teams
it's_just games_played just_don't
home_court

first round regular_season

team_like nba_cup tournament_just
give_team season_nba
cup_winner win_tournament team_wins_nba
rather_team rest_season
winning_team best_team want_win
g-league_teams nba_fans wins_nba_cup
top_talent seems_like number_seed
#1_pick playoffs_make wins_nba something_like
think_it's court_advantage normal_regular care_nba team_win
tournament_champ winning_tournament think_good
nba_tournament_champ home_court_advantage g_league
think_tournament normal_regular_season fa_cup

players care team wins tournament_stakes

positive

Figure 5. Thread 1 Comparison word cloud

negative

looks_like

positive

Figure 6. Thread 2 Comparison word cloud

Figure 5 shows us the most used keywords in thread 1 related to each sentiment, positive or negative, and seems to give us useful information that we might expect; nothing is striking. However, in *Figure 6*, related to thread 2, we can see a lot of talk about the visual appearance of the court, with phrases classified with negative sentiment such as "eyes hurt" or "bit distracting". While we might have expected something like this from the preliminary analysis of the prevailing themes, looking again at the "In-season Tournament tips off in

Indy" thread, we see that the original comment is accompanied by a video of the court's appearance, which is certainly eye-catching. While general opinion may indeed have declined overall, the video in the thread has undoubtedly diverted the public's attention and attracted the most related comments. This brings up a new question that leaves no room for calm conclusion: apart from the obvious dissatisfaction with how the tournament's special courts look, what do people now think about the rest of the in-season tournament aspects?

In order to address this, I found a new thread on Reddit called "The NBA In-Season Tournament will turn into a joke?" posted on November 3, 2023, one day before the thread used earlier. This thread was not used from the beginning because while it is true that it may give more room for a broader discussion on different opinions, it only has 197 comments. However, in this situation, it is useful to complement the information already collected before. After performing exactly the same procedure as with the first two threads and downloading the information from the comments, we arrive once again at the following comparison word cloud shown in *Figure 7*.
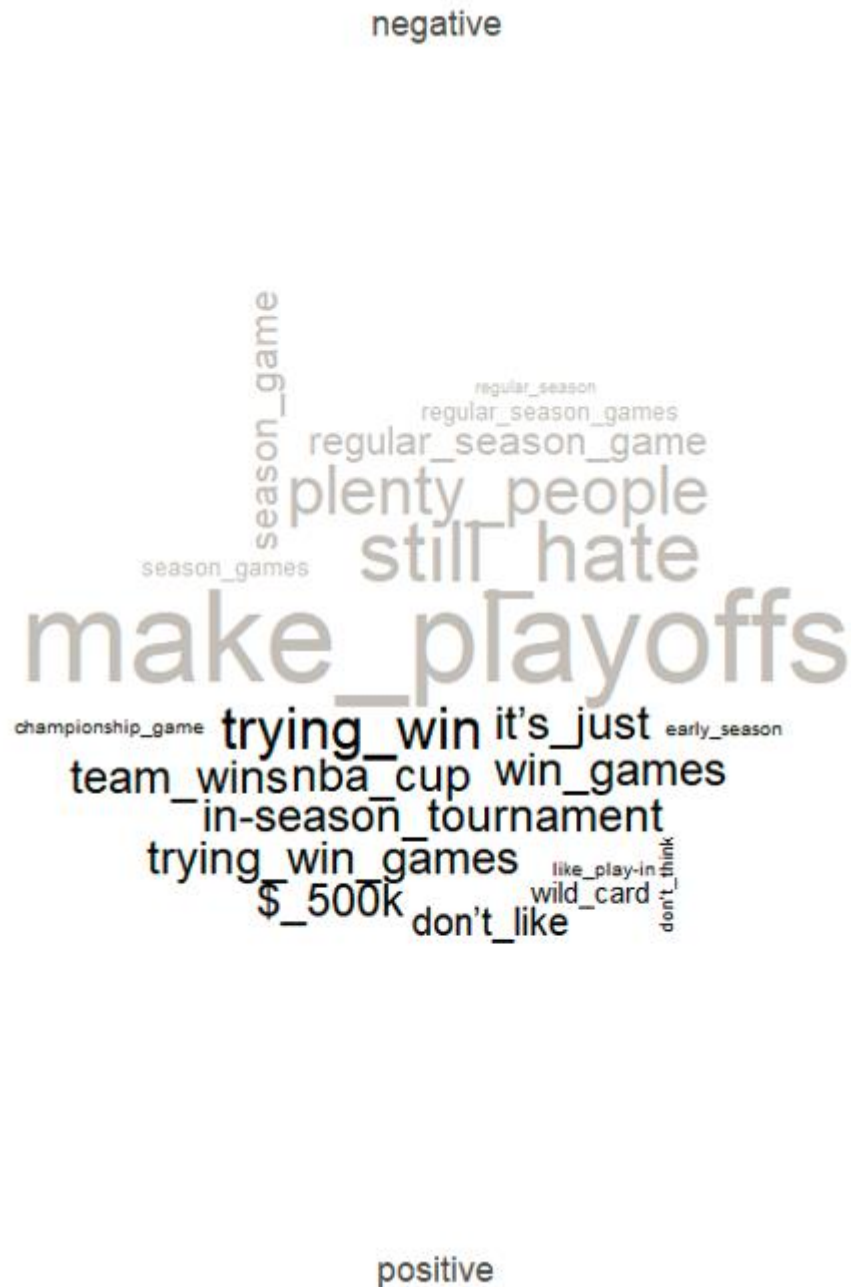
Figure 7. Thread 3 Comparison word cloud

The average logit sentiment score for this new thread was 0.2, a slight improvement from the previous one. Also, in *Figure 7* we do not find anything related to the courts, but rather, we see a high frequency of bigrams and trigrams such as "regular season game", "season games" or "regular season" indicating perhaps a lack of information as to how the tournament format worked, as previously speculated and mentioned.

To clarify this once and for all, I used the technique of keywords in context with the bigrams and trigrams just mentioned, to find the context in which they were found and if they were indeed related to lack of knowledge among fans. I found comments such as:

- "They count toward regular season record (except for the championship game), so they mean at minimum as much as any other game. And there are sizable monetary benefits, so they should mean even more to the players."
- "Do you still not understand that the games count towards the regular season standings?"

Thus, proving that there is still ignorance about the tournament, and this can influence the opinion about it.

When conducting research with information from Reddit threads, ethical concerns are indispensable. First, obtaining informed consent is challenging due to the public nature of these online discussions and the immense number of users that are involved, yet each person is also aware that the opinion they are giving is accessible to anyone on the web and we know that we are exposed to this when using any social network in a public way. Second, users frequently use pseudonyms, which makes it difficult to discern between people who are private and those who are public. Third, it is important to ensure the anonymity and data protection of those involved, however, in this document we did not have any information beyond the usernames of each user, which although they were downloaded with the comments, were not used at any time, as they were not important for our main purpose.

It is always important to fulfil ethical responsibilities with concerns in mind, and this was always the case in this project for research purposes only.

Conclusion

After all that has been analysed and discussed throughout this paper, we can finally conclude that the overall opinion on the in-season tournament has indeed decreased over time, as we observe how the average sentiment score of the current threads decreased compared to the first one, even when excluding the one that contained too much information about the courts. However, we can also say that the rejection of this thread is

partly due to the fact that there is still a certain level of misinformation about the tournament format, which can and hopefully will improve over time.

It is also important to make clear that there are several opportunities for improvement in the analysis conducted, as the threads have some difference in comments, which was further magnified with the addition of the third thread. It is also true that many methods were performed that greatly simplify the information, as assigning a sentiment to a comment only by the use or non-use of certain words can ignore the real intention and context of the reply, however, in terms of simplicity and efficiency it is a good approximation of the goal we wanted to achieve. Likewise, being a tournament that is still in play, it leaves open the possibility of complementing the research in the future when it is finished, perhaps decreasing the misinformation around it and giving an opinion that will be the most valid, as it will be the one obtained after having lived the in-season tournament completely.

References

1. *The NBA is considering its own FA Cup. the results may not be pretty* (2019) *The Guardian*. Available at: https://www.theguardian.com/sport/2019/dec/05/nba-midseason-tournament-proposal (Accessed: 15 November 2023).

2. No_Cardiologist_4705 (2023) *How to you feel about the NBA's new in season tournament?*, *Reddit*. Available at: https://www.reddit.com/r/nbadiscussion/comments/14ulg7z/how_to_you_feel_about_the_nbas_new_in_season/?sort=confidence (Accessed: 15 November 2023).

3. Overton, S. (2021) *How the NBA can regain viewer engagement - the daily Utah Chronicle*, *The Daily Utah Chronicle - The University of Utah's Independent Student Voice*. Available at: https://dailyutahchronicle.com/2021/04/12/nba-regain-viewer-engagement/ (Accessed: 15 November 2023).

4. RyanTannegod (2023) *The NBA in-season tournament will turn into a joke?*, *Reddit*. Available at: https://www.reddit.com/r/nba/comments/17m92df/the_nba_inseason_tournament_will_turn_into_a_joke/ (Accessed: 15 November 2023).

5. sewsgup (2023) *In-season tournament tips off in Indy*, *Reddit*. Available at: https://www.reddit.com/r/nba/comments/17n7wf2/inseason_tournament_tips_off_in_indy/ (Accessed: 15 November 2023).