1. **Project Title**

Final Score

2. **Project Summary: It should be a 1-2 paragraph description of what your project is.**

Our project entails using the Last Two Minute Report (L2M) for every game played in the NBA since 2015 (which is when the L2M report was first created) to create a variety of visualizations for users to evaluate various metrics such as percentage of correct/incorrect calls for a certain team, evaluating the "home court advantage," seeing which players/teams get the most calls, and many more. We aim to allow users to break down NBA officiating to its finest parts and create even more transparency than the NBA currently provides. With difficult to read basketball jargon, large sets of data, and multiple barriers to entry, most avid NBA watchers don't even know how to access this data. And to be blunt, the current L2M data store is quite difficult to interpret. With the help of tabulated datasets, we aim to provide our visualizations real-time after a game has been processed by the NBA for their own L2M reports. Our intended user is average basketball fans who: a) want to learn more about how the NBA officiates, b) fantasy basketball players who hope to use a more holistic understanding of these hard-to-interpret basketball statistics to their advantage, c) players and coaches who hope to plan for certain strategic advantages when playing highly contested games.

While there are many more uses and potential users of the L2M data, we want to keep our project somewhat lighthearted and fun while still implementing the database design and implementation skills we learn in class.

3. **Description of an application of your choice. State as clearly as possible what you want to do. What problem do you want to solve, etc.?**

Kids from all over the world grow up watching and hoping to become an NBA player one day. With an astounding $10.58 billion in revenue, the NBA stands as one of the biggest and most entertaining sports leagues in the world. As the NBA has grown in popularity, the average player skill and competitiveness of the league have also increased. In these games of fine margins, external factors have impacted the outcomes of games, especially the officiating. With an increased amount of bad officiating in recent years, fans have begun to resent referees.

The Last Two Minute Report (L2M) is a crucial part of the NBA's efforts to increase transparency and understanding of the rules and processes that govern the game, providing fans, players and coaches with a more thorough picture of officiating,

acknowledging both missed and correct calls. By providing a play-by-play report regarding all calls and material non-calls in the last two minutes of the fourth quarter or any overtime period, L2M helps build a greater awareness and understanding of the rules and processes that govern the game, alongside adding accountability on the NBA to ensure fair regulation methods.

Using the data provided in official L2M reports, our goal is to use this data to identify biases in referees, stadiums, matchups, and more. We hope to use the datasets we find to help identify ways to improve the NBA's officiating and keep the games in the hands of the players instead of the referees. Additionally, we think creating some creative visualizations of different scenarios such as the importance of "home court advantage" would add another layer to this project that would prove extremely useful.

4. **What would be a good creative component (technically challenging function) that can improve the functionality of your application? (What is something cool that you want to include? How are you planning to achieve it?)**

One of the cool features we are adding to the app is a data visualization component that will allow users to see the data in various charts and graphs, and also see the data updating live as the games are happening. This would make the app more practical for daily use, as users can easily access and explore the data they are interested in, and also have a better experience understanding the data and discovering insights from it. We can achieve this by using a data visualization library when building the web application, which will provide us with the tools and functions to create and manipulate the visual elements.

One data visualization library we can use is D3, which is a popular and powerful library for creating interactive and dynamic data visualizations. D3 will allow us to quickly visualize our data in an effective manner, using various techniques such as scales, axes, transitions, animations, and more. With D3, we can create stunning and engaging data visualizations that will enhance our app and attract more users.

Another cool feature we hope to implement is predicting how contested a game will be between any given two teams based on given referees and the court they are playing on. This would add a fun, niche aspect to our project and allow us to use machine learning, an area we are all fascinated by. Based on our rudimentary research, XGBoost would be our preferred framework for using a NN to predict the refereeing results of a given game.

5. **Usefulness. Explain as clearly as possible why your chosen application is useful. What are the basic functions of your web application? (What can users of this website do? Which simple and complex features are there?). Make sure**

to answer the following questions: Are there any similar websites/applications out there? If so, what are they, and how is yours different?

Our project is designed to visualize and identify biases in the NBA's Last Two Minute (L2M) data. This is a unique approach that we believe is incredibly useful. The L2M data from NBA games is a well of endless raw information, especially about the crucial final moments of games. However, biases in this data can lead to skewed interpretations and potentially misinformed decisions. Our project addresses this by helping users gain a more accurate understanding of the game's dynamics, which is invaluable for analysts, commentators, coaches, players, and fans who are interested in the strategic aspects of the game.

In terms of functionality, users can upload or select specific L2M data sets for analysis on our platform. The application then processes this data, applying statistical methods to identify potential biases. These could be related to specific teams, players, referees, or types of calls. The results are visualized in an intuitive and interactive format, allowing users to explore the data and draw their own conclusions.

Another primary user who can create, update and delete from the platform is the NBA itself. Ideally, whenever a L2M report has been created, the NBA will update it to our platform in CSV format for us to use in our visualizations.

While there are other sports analytics platforms available, our project will stand due to readability and the focus on the L2M. Most other analytical platforms focus on player stats and efficiency metrics instead of officiating. To our knowledge, this is one of the first attempts to break down referee data and add elements of visualization and readability to it. This specific focus allows us to provide a more in-depth analysis than more general sports analytics tools. Our use of interactive visualizations makes the data more accessible, even to those without a background in data analysis or statistics.

In conclusion, we believe our application fills a unique niche in the sports analytics field. By providing tools to visualize and identify biases in NBA's L2M data, we can help users gain a more nuanced understanding of the game's final moments. This could lead to more informed discussions, strategies, and decisions, ultimately enhancing the game for everyone involved.

6. **Realness. We want you to build a real application. So, make sure to locate real datasets. Describe your data sources (Where is the data from? In what format [csv, xls, txt,...], data size [cardinality and degree], what information does the data source capture?). It would be hard to satisfy stage 2 requirements with one dataset. Thus, we strongly recommend identifying at least two different data sources for your project.**

We are very fortunate to have access to a number of large datasets that cover NBA stats, as they can provide us with valuable information and insights for our app. However, some of these datasets stand out more than others, because they cover a variety of markers that are not commonly found in other sources.

One of these datasets is a GitHub repository that contains numerous CSV files of data scraped from the NBA's Last Two Minute Reports (L2M). This dataset has both relevant and abundant data that can prove to be useful in building the backbone of our app, as it covers various aspects of the game, such as fouls, calls, reviews, and corrections. Some of the data in this dataset includes all the fouls in the game and the percentage of good calls by referees, which can help us analyze the performance and accuracy of the officiating. Another dataset that we found is a large CSV file from Kaggle2 that has all the calls, the referee that made them, and the exact time that the call was made for every game in the 2017-18 season. This dataset can help us set up the live data tracking that we could use to visualize the data live, as it allows us to see the frequency and distribution of calls throughout the game, and also the impact of each call on the outcome. By using these datasets, we can create a unique and engaging app that shows NBA game data in a new and exciting way.

Dataset 1:
https://github.com/atlhawksfanatic/L2M/tree/master

Dataset 2:
https://www.kaggle.com/datasets/thedevastator/nba-last-two-minute-game-reports?resource=download

7. **A detailed description of the functionality that your website offers. This is where you talk about what the website delivers. Talk about how a user would interact with the application (i.e., things that one could create, delete, update, or search for). Read the requirements for stage 4 to see what other functionalities you want to provide to the users. You should include:**

There are a few key components that make up our website:
   a. General landing page visualizations of trends in calls since 2015, general statistics on the amount of calls, and team splits
   b. On court visualizations of where correct calls (CC), incorrect calls (IC), correct no-calls (CNC) and incorrect no-calls (INC) occur on the court. People can filter by game, by team, and by player.

c. Graph visualizations of the amount of the trend of calls a team receives every season.
d. A "custom" section that allows for users to create their own custom matchups where we predict who will get what calls in a given game.

Although we still aim to expand/edit these functionalities, the team/player splits on the amount/type of calls they receive is critical to our website. Alongside that, we hope to create a new, more improved and customizable view of the current F2M report shown in the mockup below. We don't know exactly how we will format this updated view, but that is also another component.

1. A low-fidelity UI mockup: What do you imagine your final application's interface might look like? A PowerPoint slide or a pencil sketch on a piece of paper works!
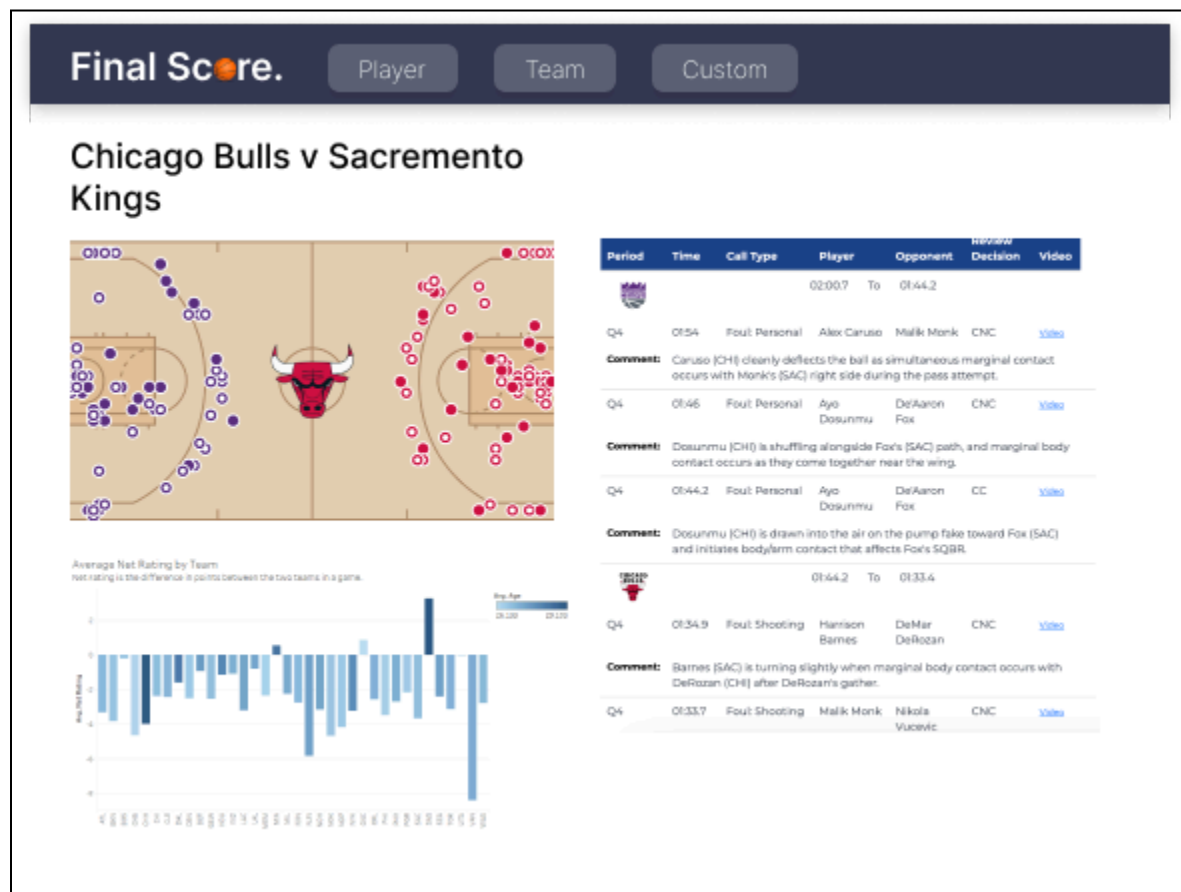
2. Project work distribution: Who will be responsible for each of the tasks or subtasks?

Explain how backend systems will be distributed across members. Be as specific as possible as this could be part of the final peer evaluation metrics.

    i.   Satyam Singh:
1. Full-stack focused on data visualization and properly implementing the D3 library to show live data.
2. Working on creating and deleting data from the database which will be done by the NBA.

    ii.   Shivam Syal:
1. Full-stack focused on data visualization and implementing the D3 library to show live data.
2. Working on creating and removing data from the database which will be done by the NBA.

    iii.   Yash Mandavia:
1. Full-stack focused on data prediction using old data and reflecting that back to the database.
2. Working on updating and reading data from the database which will be done by both the NBA and users.

    iv.   Sameer Komoravolu:
1. Back-end optimization and making sure the web app has the fundamentals to function properly.
2. Working on properly implementing the datasets into the web app.