Master Degree Program in

Data Science and Advanced Analytics

Business Intelligence

Final Project Report: NBA CASE STUDY

Ana Carolina Ottavi, 20220541
Carolina Bezerra, 20220392
Luca Loureiro, 20221750
Maria Arbelaez, 20221381

Group 14

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa
June, 2023

INDEX

1. Integration of the contents of the previous delivery

In the first delivery, we designed a Dimensional Model to support the different levels of business decision-makers for the NBA case study. This model served as the basis for our Power BI implementation, where we developed reports that answered the business questions and requirements of the NBA case study challenge.

1.1. Introduction and Presentation of Business

The case study for this project presents a sports company that operates a website dedicated to providing in-depth NBA data and statistics to its audience. The company's goal is to offer valuable insights into team and player performances, which can help website visitors and league bettors make well-founded decisions. The company aspires to boost user engagement and draw a wider audience to its platform.

Apart from providing data to website users, the company's employees also need efficient access to this information to perform their daily tasks. To achieve these objectives, the company has partnered with our team to develop an end-to-end self-service BI solution that will provide easy and efficient access to the NBA data of interest to both website users and company employees, supporting their respective decisions.

This solution will consist of an engaging dashboard that focuses on various aspects of NBA games, such as team and player stats, match official records, and historical performance trends, and that will generate accurate reports that answer the questions of each target user. The dashboard will cater to the requirements of both gambler users and company employees by presenting the data in an easy-to-understand and visually attractive way. Additionally, the implementation of row-level security in Power BI will ensure the confidentiality of the data. Filters restrict data access at the row level, and you can define filters within roles.

Ultimately, the Power BI solution will improve the decision-making process and overall experience of both gambler users and company employees on the website.

1.2. Identification of the Business Problem

The main problem faced by the sports company is that their website lacks a user-friendly tool for visitors and gamblers to analyze NBA statistics. Due to the absence of an efficient platform for data presentation and analysis, users find it difficult to obtain meaningful insights from the vast amount of available information.

To solve this problem, the company aims to create a Power BI solution that will provide valuable insights into team and player performance in an easy-to-understand and interactive format. With this solution, users can explore different aspects of the game, compare teams and players, and identify trends or patterns that can help them make better decisions. By enhancing the user experience, the company hopes to attract more website users, increase user engagement and build a loyal customer base.

Another challenge is the need for employees to conduct thorough and efficient analysis, mainly regarding metrics related to performance and evolution by team, but with more detail than what is offered to website users.

1.3. Business Questions

To ensure that the Power BI solution meets the needs of the website users, we have identified the following key business questions:

1. What are the main performance metrics by team per game? The main performance metrics include the number of home wins and losses, the number of away wins and losses, the points scored and the difference of points between visiting and home team.

2. What are the main statistics of the performance metrics by team per each season and in history?

3. How has each team evolved over time regarding the main performance metrics?

4. What is the average of the main performance metrics by player per game? The main performance metrics include points, assists, rebounds, steals and PIE.

5. What are the main attributes of each player? The main attributes include weight, height, position and current team.

- Note: in the requests, the client asks to see the evolution of each player throughout the season, however, in the datasets provided there is no information of the metrics per player per game. Without this data it is not possible to fulfill this request.

To ensure that the Power BI solution meets the needs of the company employees, we have identified the following key business questions:

1. How do the main performance metrics compare between teams, from a national and regional perspective? For example, how do average points from all the games in each season and in history compare between teams? The main performance metrics include the number of home wins and losses, the number of away wins and losses and the points scored.
2. What is the ratio of home wins versus away match wins of every team against each other team per season and in history?

3. How have the top 5 teams evolved by season in terms of average points?

4. What are the main attributes of each match official and how many games did each of them oversee per season and in history?

1.4. Description of the Source Data and Discovery Process

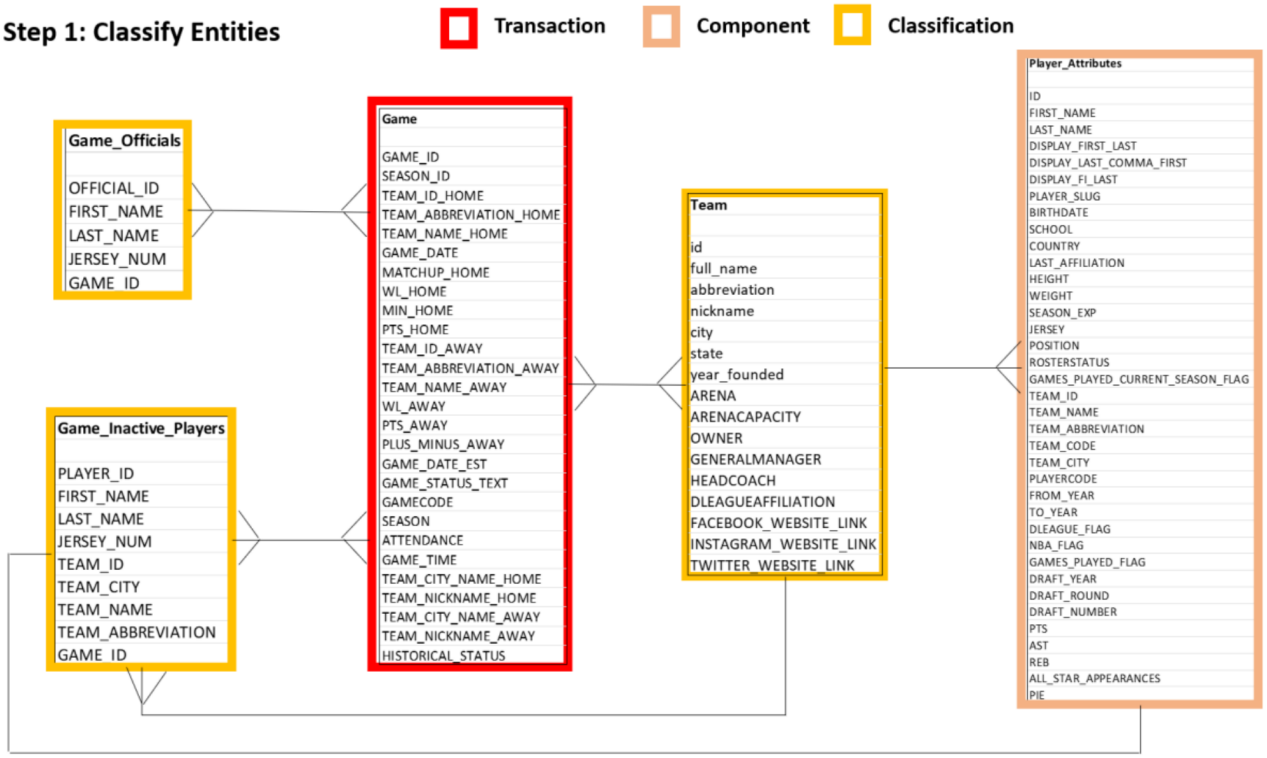
The data provided for the Power BI solution comprises five csv files that contain valuable information about different aspects of NBA games, players, officials, and teams. To ensure accurate insights in the Power BI solution, we will thoroughly explore these datasets during the data discovery process, understanding their structure, attributes, and relationships. Additionally, through the data cleaning process, we will ensure data quality and consistency. The following files are included:

File	Dimension	Contents / Description
Game	columns: 11 rows: 62.451	The last update was 2021/12/10. It contains details for each game in all available seasons, such as: date, location, teams involved, game time, viewer attendance and match results for home and away teams.
Player_Attributes	columns: 37 rows: 4.501	This file offers essential information about NBA players, including their names, birthday, nationality, school, height, weight, draft year, round and number, current team, jersey number, position, roster status, when they started playing for that team and when they left it, if they have played in the D league and/or the NBA, and if they have played in any game historically and in the current season. It also contains performance statistics such as average number of points scored per game, average assists per game, average rebounds per game, average PIE (Player Impact Estimate) per game and count of appearances in All-Star games.
Game_Officials	columns: 5 rows: 65.159	This file provides information about match officials associated with each game, such as their names and jersey numbers
Game_Inactive_Players	columns: 9 rows: 98.680	This file contains information about inactive players for each game, such as their names, jersey numbers and team. Those players were not eligible to play in those particular matches.
Team	columns: 21 rows: 31	This file provides information of each team, such as name, abbreviation, nickname, location, year of establishment, associated arena and its capacity, owner, general manager, head coach, D-league affiliation and social media html links.

By combining these data sources, the Power BI solution will answer the identified business questions, offering a comprehensive platform for analyzing NBA team and player performance.

1.5. Draft Design and Description of the Dimensional Model

To build the dimensional model for our BI solution we followed the Kimball methodology learned in class. First, we classified the entities present in the operational model:



The Team, Game_Officials and Game_Inactive_Players entities were classified as Classification entities because they all represent business processes that categorize data. The Game entity was classified as a Transaction entity as it represents a business process that captures one single event; in this case, it captures data for each single game. Finally, the Player_Attributes entity was classified as a Component entity because besides categorizing the players, it represents a business process made up of smaller components, as the metrics it contains for each player correspond to the average results of several games. Then we identified the hierarchies present in the current operational model that were relevant to the business questions we need to sort out with our BI solution:

- **Dim Date**

  For this table, the dates are hierarchized according to Year, Trimester, Month and Day.

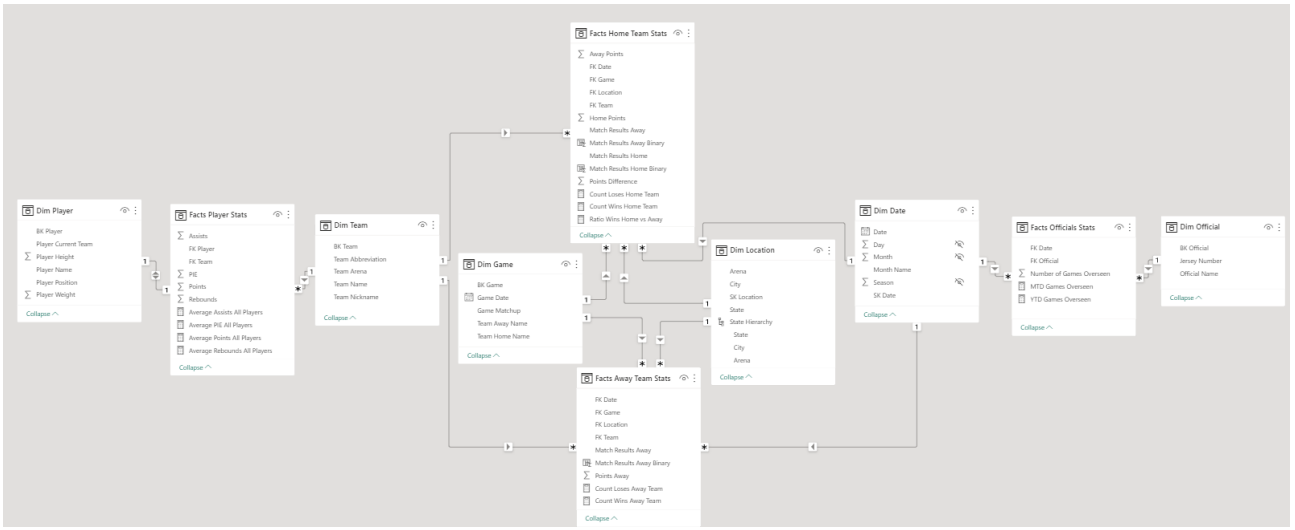
- **Dim Game**

  The dates for the NBA Games are classified by Year, Trimester, Month and Day.

- **Dim Location**

  The locations of the games are represented in Dim Location and are hierarchized by State, City and Arena.

And finally, we collapsed the hierarchies to build a dimensional model based on a constellation schema composed of facts and dimensions:



1.6. Data Modelling Methodology

The Kimball methodology was employed to create the Dimensional Model for our Power BI project. This methodology is a recognized and reliable approach that focuses on building data warehouses and business intelligence systems by identifying significant facts and dimensions and creating a star or constellation schema. The steps taken to develop the Dimensional Model were:

1. Identify the business processes: We thoroughly revised the scope file provided to comprehend the business objectives, requirements, and the specific questions the company wanted to address through the Power BI solution.
2. Declare the grain: By analyzing the business requirements, we determined the level of detail that the BI solution would provide to each target user. This was detailed in chapter 1.3. “Business Questions”.
3. Identify the dimensions: We examined the provided data sources and identified the dimensions that provide context to the facts. These dimensions will enable users to analyze and filter the data based on different attributes.
4. Identify the facts: We examined the provided data sources and identified the quantitative measures of interest.
5. Build the dimensional model: We developed a constellation schema with four fact tables, one regarding player performance metrics, two others regarding team performance metrics in home

and away games, and the third one regarding match official performance metrics; and six dimension tables (Player, Team, Location, Game, Date and Official). This schema enables efficient querying and reporting on the data and addresses the identified business questions.

By utilizing the Kimball methodology, we ensured that the Dimensional Model was optimized for both performance and ease of use, providing a strong foundation for the Power BI solution to offer valuable insights to the business stakeholders.

1.7. Description of the Main Components of the Dimensional Model

The main components of the developed Dimensional Model are the four Fact tables and the associated Dimension tables. Each table serves a specific purpose in providing context and information to answer the identified business questions.

1. **Fact table – Player Stats Facts:** This table contains the quantitative metrics of interest for each player. It includes foreign key columns to link it to the player and team dimension tables because they contain the information regarding the attributes of interest expressed in the business questions: weight, height, current team and position of each player.
2. **Fact table – Home Team Stats Facts:** This table contains the quantitative metrics of interest for the home team. It includes foreign key columns to link it to the team, game, date and location dimension tables because this will allow comparison between teams, games, locations and time.
3. **Fact table – Away Team Stats Facts:** This table contains the quantitative metrics of interest for the away team. It includes foreign key columns to link it to the team, game, date and location dimension tables because this will allow comparison between teams, games, locations and time.
4. **Fact table – Official Stats Facts:** This table contains the only quantitative metric of interest for each match official which is the number of games overseen. It includes foreign key columns to link it to the official, game and date dimension tables as this will allow querying the desired metric by season or history.
5. **Dimension table - Player:** This table provides context surrounding the events of the player performance fact table. It gives us the ‘who’ of those events, that is, the players. The PlayerID currently contained in the database is the primary key, therefore it is called a business key.
6. **Dimension table - Team:** This table provides context surrounding the events of the teams performance fact table. It gives us the ‘who’ of those events, that is, the teams. The TeamID currently contained in the database is the primary key, therefore it is called a business key.
7. **Dimension table - Location:** This table provides context surrounding the events of the team performance fact table. It gives us the ‘where’ of those events, that is, the state, city and arena. LocationID is the primary key, and since it was created by us, it is a surrogate key.
8. **Dimension table - Game:** This table provides context surrounding the events of the team performance and official performance fact tables. It gives us the ‘what’ of those events. The GameID currently contained in the database is the primary key, therefore it is called a business key.
9. **Dimension table - Date:** This table provides context surrounding the events of the team performance and official performance fact tables. It gives us the when of those events, including the season and date. DateID is the primary key, and since it was created by us, it is a surrogate key.
10. **Dimension table - Official:** This table provides context surrounding the events of the official performance fact table. It gives us the whole of those events, that is, the officials. The OfficialID currently contained in the database is the primary key, therefore it is called a business key.

Together, these components form a robust Dimensional Model that allows for efficient querying and reporting on the data, providing valuable insights to the business stakeholders and addressing their key business questions. This well-designed model serves as the foundation for the Power BI solution to deliver a user-friendly, interactive dashboard for the sports company's website users and league gamblers.

2. Description of the most important data engineering applied steps in Power BI

Now that the business requirements have been defined and the data source has been provided, we must proceed to the data engineering phase. In our project, we performed several data engineering steps in Power BI to ensure the quality and usability of our data. ETL operations account for a significant portion of data engineering activity. The ETL phase consists of three steps, which are Data Extraction, Data Transformation and Data Loading.

The Extraction stage entails obtaining information from one or more sources, such as databases, files, or external systems. Following that, the selected data is read from these sources.

After the data has been extracted, it must be converted into a format acceptable for the data warehouse. This transformation process comprises procedures such as data cleansing, integration, and aggregation. The goal is to verify that the data is consistent, correct, and formatted in accordance with the Data Warehouse schema.

The modified data is then placed into the data warehouse. The loading step entails entering the modified data into the DW's relevant tables or data structures.

2.1. Data Extraction

Data integration involves combining data from different sources and providing users with a unified view of these data. In our project, we integrated data from the 'Game', 'Game_Inactive_Players', 'Game_Officials', 'Player_Attributes', and 'Team' tables. This allowed us to create a comprehensive dataset for analysis, providing a more holistic view of NBA team and player performances.

2.2. Data Transformation

Data cleaning is the process of detecting and correcting (or removing) corrupt or inaccurate records from a dataset. In our project, we performed several data cleaning steps:

Removing Unnecessary Rows and Columns: In the 'Game' table, we removed the first two rows, which were not necessary for our analysis. We also made the first row the header for better data understanding and manipulation. We also removed columns that were replicates. For example this was the case in the Game table with the HOME_TEAM_ID and VISITOR_TEAM_ID columns that were replicates of TEAM_ID_HOME and TEAM_NAME_HOME; as well as in the Team table where ID, ABBREVIATION, NICKNAME, YEARFOUNDED and CITY, were replicates of id, full_name, abbreviation, nickname and city. We also deleted redundant columns, for example in the Player_Atribbutes table, were DISPLAY_LAST_COMMA_FIRST, DISPLAY_FI_LAST, PLAYER_SLUG and LAST_AFFILIATION, were just different formats to express the players names, country and school, which was data already contained in other columns. Finally, in the Game table, we also deleted the column GAME_TIME because it was 100% empty.

Fixing Data Types: Data types are crucial for accurate data analysis. We ensured that all columns in our tables had the correct data types. For example, numerical columns were set as integers or decimals, and categorical columns were set as text.

Dim player	BK Player	Whole Number
	Player Name	Text
	Player Height	Decimal Number
	Player Weight	Decimal Number
	Player Position	Text
	Player current team	Text
Dim Team	BK Team	Whole Number
	Team Name	Text
	Team Abbreviation	Text
	Team Nickname	Text
	Team Arena	Text
Dim Location	SK Location	Whole Number
	State	Text
	City	Text
	Arena	Text
Dim Date	SK Date	Whole Number
	Date	Date
	Day	Whole Number
	Month	Whole Number
	Month Name	Text
	Season	Whole Number
Dim Official	BK Official	Whole Number
	Official Name	Text
	Jersey Whole Number	Whole Number
Dim Game	BK Game	Whole Number
	Game Date	Date

	Game Matchup	Text
	Team Home Name	Text
	Team Away Name	Text
Facts Players Stats	FK Player	Whole Number
	FK Team	Whole Number
	Points	Decimal Number
	Assists	Decimal Number
	Rebounds	Decimal Number
	PIE	Decimal Number
Facts Home Team Stats	FK Team	Whole Number
	FK Game	Whole Number
	FK Location	Whole Number
	FK Date	Whole Number
	Match Results Home	Text
	Match Results Away	Text
	Home Points	Whole Number
	Away Points	Whole Number
	Points Difference	Whole Number
Facts Away Team Stats	FK Team	Whole Number
	FK Game	Whole Number
	FK Location	Whole Number
	FK Date	Whole Number
	Match Results Home	Text
	Match Results Away	Text
	Points Away	Whole Number
Facts Official Stats	FK Official	Whole Number
	FK Date	Whole Number
	Whole Number of Games Overseen	Whole Number

Removing Duplicate Rows: We removed duplicate rows in every table under the logic that no record can be exactly the same. For example, one player can't have exactly the same name, attributes and metrics. Or the same match up can't be played on exactly the same date, and so on.

Filtering Rows: We also did some row filtering taking into account the business requirements. In the game table, we filter out all rows with a value of 1 in the HISTORICAL_STATUS column, as the business specifically stated that they did not want to see information regarding historical seasons. In the Player_Attributes table we filtered out all players with inactive status in the ROASTERSTATUS column, as the interest of the business is in the active players. After doing this filtering, the ALL_STAR_APPEARANCES column was 100% empty, therefore we deleted the column.

Data transformation is the process of converting data from one format or structure into another format or structure. In our project, we created calculated columns using DAX (Data Analysis Expressions) to transform our data into a more useful format for analysis.

We created binary columns match results for home, 'Match Results Home Binary', and for away, 'Match Results Away Binary'. These columns indicate whether a team won (1) or lost (-1) a match. This transformation simplified our data and made it easier to visualize team performance.

In the Dim Official table we created a new column with the full name of the official, concatenating the columns FIRST_NAME and LAST_NAME of the original Official table. In the Facts Official Stats table we created a new column called Number of Games Overseen, using the Group By function which allowed us to count the games overseen per official per date.

2.3. Data Loading

3. Explanation of the model optimization steps applied

Model optimization is a crucial aspect of any data analysis project. It involves improving the performance and usability of the data model to ensure efficient and effective data analysis. In our project, we applied several model optimization steps in Power BI, which are detailed below:

3.1. Hierarchies

Hierarchies in Power BI allow users to drill down into more detailed data. We configured hierarchies correctly in our model, enabling users to navigate from high-level to more granular data. For example, users can drill down from season-level data to individual game data, providing a more detailed view of team and player performances.

3.2. Hidden Columns

To simplify the view and improve performance, we hid unnecessary columns in our tables. These are columns that are not needed for our analysis or visualization but are useful for data processing or calculations. In our model, we hid all the columns that were already included into the hierarchies. By hiding these columns, we made our model more user-friendly and efficient.

3.3. Formatting

We formatted our data to make it more readable and understandable. This included applying appropriate number formats, setting date formats, and using descriptive column names. Proper formatting not only makes the data easier to understand but also enhances the visual appeal of our reports and dashboards.

3.4. Summarization

We created summary measures to provide high-level insights into our data. These measures, created using DAX, include counts, sums, averages, and other aggregations. They allow users to quickly grasp key trends and patterns in the data without having to delve into the details.

3.5. Date Table

In Power BI, marking a table as a 'Date' table enables time intelligence calculations, which are crucial for analyzing trends over time. We marked the 'Date' column as a date column, enabling us to perform calculations such as year-to-date and month to date comparisons.

In conclusion, these model optimization steps were crucial in enhancing the performance and usability of our Power BI model. They ensured that our model was efficient, user-friendly, and capable of delivering valuable insights to answer the business questions of the NBA case study challenge.

4. Description and explanation of the measures and calculated columns created

We created several calculated columns and measures to answer the business needs:

Calculated Columns: We created binary columns for home and away match results. These columns indicate whether a team won (1) or lost (-1) a match. The purpose of creating this was to be able to design a visualization that showed with bars the wins or loses over time:

```
Match Results Home Binary = IF(TRIM('Facts Home Team Stats'[Match Results Home]) = "Win", 1, -1)
```

```
Match Results Away Binary = IF(TRIM('Facts Away Team Stats'[Match Results Away]) = "Win", 1, -1)
```

DAX Measures: We created measures to count the number of wins and losses for home and away teams, as well as a measure to calculate the ratio of home vs away wins for each team:

```
Count Wins Home Team = CALCULATE(COUNT('Facts Home Team Stats'[Match Results Home Binary]),
'Facts Home Team Stats'[Match Results Home Binary] = 1)
```

```
Count Loses Home Team = CALCULATE(COUNT('Facts Home Team Stats'[Match Results Home Binary]),
'Facts Home Team Stats'[Match Results Home Binary] = -1)

Count Wins Away Team = CALCULATE(COUNT('Facts Away Team Stats'[Match Results Away Binary]),
'Facts Away Team Stats'[Match Results Away Binary] = 1)

Count Loses Away Team = CALCULATE(COUNT('Facts Away Team Stats'[Match Results Away Binary]),
'Facts Away Team Stats'[Match Results Away Binary] = -1)
```

```
Ratio Wins Home vs Away = 'Facts Home Team Stats'[Count Wins Home Team] / 'Facts Away Team
Stats'[Count Wins Away Team]
```

We also created measures to get the average PIE, Points, Rebounds and Assists of all players so we could set this as a KPI goal for each specific player:

```
Average Assists All Players =
CALCULATE(
    AVERAGE('Facts Player Stats'[Assists]),
    ALLEXCEPT('Facts Player Stats', 'Facts Player Stats'[FK Player])
)
Average PIE All Players =
CALCULATE(
    AVERAGE('Facts Player Stats'[PIE]),
    ALLEXCEPT('Facts Player Stats', 'Facts Player Stats'[FK Player])
)

Average Points All Players =
CALCULATE(
    AVERAGE('Facts Player Stats'[Points]),
    ALLEXCEPT('Facts Player Stats', 'Facts Player Stats'[FK Player])
)

Average Rebounds All Players =
CALCULATE(
    AVERAGE('Facts Player Stats'[Rebounds]),
    ALLEXCEPT('Facts Player Stats', 'Facts Player Stats'[FK Player])
)
```

Additionally, to comply with the requirements for this project, two date intelligence measures were created, which count the number of games overseen by a selected official within a rolling 30-day window and 1-year window:

```
MTD Games Overseen = TOTALMTD(SUM('Facts Officials Stats'[Number of Games Overseen]),('Dim Date'
[Date]))

YTD Games Overseen = TOTALYTD(SUM('Facts Officials Stats'[Number of Games Overseen]),'Dim Date'
[Date])
```

These calculated columns and measures are crucial for answering the business questions, as they provide insights into team performance.

5. Description of the main technical aspects

Visualizations are critical in Power BI for displaying data in a clear and comprehensible manner. These visualizations enable users to effectively examine and analyze data by using a variety of technical elements. We employed a project on Power BI, which helps users to obtain important insights and make data-driven decisions by including features such as tooltips, multiple types of graphs, slicers, drillthrough capabilities, and predictive insights.

We created a drillthrough menu to make navigating seamless and dynamic. This menu lets users easily navigate our report pages. For instance, clicking on the match results menu item would instantly take users to a report page with detailed match data by region, team, date, results and points. There are also back and forward arrows that let the user navigate the pages of the report in order, as well as a link in the NBA logo present in all pages that takes the user back to the homepage.

We added a dynamic tooltip functionality to the map visualization of NBA teams as part of our Power BI project. We have included tooltips that show the average points scored by each side at home and away games. When you hover over a specific region, like Florida, the tooltip displays the average points scored by teams in that region, like the Miami Heat and Orlando Magic. The use of this functionality allows users to acquire a deeper understanding of NBA team performance across different geographical locations.

We have effectively created Key Performance Indicators for the performance of individual players across four different metrics as part of our report view. We created key performance indicators for average points per game, average rebounds per game, average assists per game, and average PIE (Player Impact Estimate) per game. This approach allows users to compare individual players' performance against the overall average.

More so, we created a line chart that illustrates any team's history and expected points per season. We projected the selected team's score for 10 years using Power BI's predictive capabilities. This forecast helps with long-term planning by predicting the team's scoring trends.

In designing the match results report page, our objective was to provide users with a highly interactive experience. To achieve this, we incorporated slicers that enable users to select specific teams and dates, catering to individuals who have particular games in mind. In addition to the slicers, we included cards to visually present the winners and losers of each game, showcasing the teams' respective points and the point differentials, which allows users to quickly assess the results and gain insights into the performance of individual teams.

In our Power BI report page dedicated to teams' evolution analysis, we have implemented slicers for selecting specific teams and seasons. The selected team's home and away games are visualized for comparison. Users can compare the team's performance in different situations using these visualizations. We can check their wins and losses through a certain season, the trend of points scored, the count of losses and wins and the average points overall for that team in a certain season.

Within our Power BI report, we have dedicated a page called "Teams Regional Analysis & Top 5" to provide a comprehensive overview for employees. This page focuses on the regional analysis of teams and team performance metrics, displaying average points scored, wins and losses, home vs. away performance ratio, and the top 5 teams with the highest points. This analysis aids in evaluating teams success, identifying trends, and making more informed decisions within the sports company.

We have dedicated a page specifically for player analysis. This page allows users to select a player's name using a slicer, providing a personalized view of their performance metrics and attributes. Users can slice a player's name to see their performance numbers and qualities on this page. This analysis helps users assess player performance and understand their contributions during games.

We have devoted an entire page to official analysis. This page enables users to select an official's name and season using slicers in order to view their performance metrics in detail. As for visualizations, we can see cards with the jersey number and number of games overseen per official according to the season. To provide a broader perspective, the page includes two bar charts: one depicting the rolling 30 days games overseen and the other showing the rolling one-year games overseen. These visualizations offer a trend analysis of the official's involvement over time, providing a comprehensive understanding of their short and medium-term performance.

Overall, these visualizations in Power BI provide a powerful way to organize and present data, enabling users to quickly access and understand relevant information and dive deeper into specific areas of interest.

6. Analysis and discussion of the outputs of the project

The following are the business questions and their respective answers based on the data:

Website Users Business Questions:

1. What are the main performance metrics by team per game? The main performance metrics include the number of home wins and losses, the number of away wins and losses, the points scored and the difference of points between visiting and home team.
2. What are the main statistics of the performance metrics by team per each season and in history?
3. How has each team evolved over time regarding the main performance metrics?

   Questions 1 to 3 are answered in the Match Results and Team Evolution Analysis pages of the report. Both pages show the main performance metrics per team and per game. The first one shows the metrics for a specific game, while the second one shows the history and evolution of metrics for several games (for all history or for a specific season, depending on the filter applied). For example, if a user wants to search the game results of the Atlanta Hawks vs. Boston Celtics on February 24, 2021, they will see that the Atlanta Hawks won the game by scoring 127 points vs. 112 points, which is a difference of 15 points. On the other hand, if a person wants to see, for example, the evolution of the main metrics for the Atlanta Hawks in season 2021, they can see that in home games they had 24 wins, 11 losses, and 115 average points, while in away games they had 15 wins, 20 losses, and 108 average points. They will also see that for home games, the team had more losses in the first half of the season and more wins in the second half, and accordingly, the number of points scored per game had a positive trend during the season. As for away games, the team also had more losses in the first half of the season and more wins in the second half, but the trend in the number of points scored per game was stable.

4. What is the average of the main performance metrics by player per game? The main performance metrics include points, assists, rebounds, and PIE.
5. What are the main attributes of each player? The main attributes include weight, height, position and current team.

Questions 4 and 5 are answered on the Player Analysis page of the report. This page lets the user select the name of the specific player they are interested in in the filter, and will consequently show them cards with all the metrics and attributes mentioned above. Moreover, for the metrics, it will let them compare how the player is performing compared to the average metrics of all players. For example, if a user is interested in Alec Burks, they will be able to see that the player weighs 214 lb, that his height is 78 ft and that he currently plays for the New York Nicks in the guard position. They will also be able to see that the player scored 12.6 points in average per game vs. an all players average of 10, that he made 4.3 rebounds in average per game vs. an all players average of 4, that he made 2.2 assists in average per game vs. an all players average of 2.2 and that his PIE score is 0.11 in average per game vs. an all players average of 0.09.

Employees Business Questions:

1. How do the main performance metrics compare between teams, from a national and regional perspective? For example, how do average points from all the games in each season and in history compare between teams? The main performance metrics include the number of home wins and losses, the number of away wins and losses and the points scored.
2. What is the ratio of home wins versus away match wins of every team against each other team per season and in history?

Questions 1 and 2 are answered in Team Regional Analysis and Points Map pages of the report. The first report page shows a table that has the team names in the rows and the metrics of interest in the columns. The filter allows the user to decide whether they want to see the metrics for all teams in the country or for the teams of a specific region/state and if they want to see the metrics regarding all history or for a specific season or seasons. The second page will show bubble maps where the bubbles correspond to the average points scored per game in each region/state and where the tooltips show the average points scored per game for the teams that belong to each region. For example, if an employee is interested in comparing the main performance metrics of the teams in California in season 2021, they will be able to see the following results:

Team	Avg Points Scored Home Games	Avg Points Scored Away Games	Wins Home	Wins Away	Losses Home	Losses Away	Ratio Wins Home vs Away
LA Clippers	111.12	103.06	42	16	24	50	2.63
Golden State Warriors	110.58	105.86	19	10	17	26	1.90
Sacramento Kings	108.18	111.88	10	17	23	16	0.59
Los Angeles Lakers	102.81	100.41	32	22	32	42	1.45

Furthermore, if an employee sees the Points Map page for season 2021, for example, they will see that in home games, Wisconsin, Maryland and Pennsylvania were the states where most points were scored per game on average, while in away games the best states were Minnesota, Oklahoma and Maryland. They will be able to see, for example, that in Wisconsin there is only one team called the Milwaukee Hawks and that on average they scored 120 points in home games and 117 in away games.

3. How have the top 5 teams evolved by season in terms of average points?

This question is answered on the Top 5 page of the report. The user will find a stacked bar chart here that displays the average points scored per game by the top five teams. In this case, the best five teams are: the Brooklyn Nets, the Golden State Warriors, the LA CLippers, the New Orleans

Pelicans and the Oklahoma City Thunder. In the graphic, the user will be able to see, for example, that the Golden State Warriors have been playing since 1996 and that the season where they scored more points on average was 2016 with 119 points, while the worst season was 1998 with 89 points on average.

4. What are the main attributes of each match official and how many games did each of them oversee per season and in history?

   This question is answered on the Official Analysis page of the report. Here, the user will be able to select one specific official and they will see three cards with their name, jersey number and number of games overseen during the season or seasons of interest. For example, if they pick Aaron Smith, they will see that his jersey number is 68 and that he has overseen 26 games in 2021 and 283 in all history. Additionally, the user will find two bar plots that display the number of games each official has overseen over the previous month and year. For example, they could see that there are six officials who have overseen the most games in the last 30 days (8 games); they are Aaron Smith, Brandon Adair, James Capers, Justin Van Duyne, Kane Fitzgerald and Zach Zarba. And the officials that have overseen the most games in the last year (52 games) are Aaron Smith and Dedrick Taylor.

7. Conclusion

Our project for the sports company has been a comprehensive journey, starting from understanding the business requirements to delivering a Power BI solution that caters to the needs of both the company's employees and its website users. The project's main aspects included identifying the business problem, formulating key business questions, understanding the source data, designing a dimensional model, applying data engineering steps in Power BI, and creating an interactive dashboard.

The project's key takeaways include the importance of a well-defined dimensional model, the value of meticulous data cleaning and transformation, and the power of a user-friendly dashboard in driving business decisions. The dimensional model, built using the Kimball methodology, served as the backbone of the Power BI solution, enabling efficient querying and reporting on the data. The data cleaning and transformation process ensured data quality and consistency, while the interactive dashboard provided valuable insights into team and player performances, enhancing the decision-making process for both the company's employees and website users.

In conclusion, the project underscores the importance of a well-designed BI solution in driving business decisions. It demonstrates how a thoughtful and meticulous approach to data cleaning, transformation, and modeling can result in a powerful tool that not only meets current business needs but also has the potential for future growth and sophistication as the company's data needs evolve.