

✓ Milestone 5 | NBA Statistics

INTRODUCTION: Stats for players and teams have long been a part of professional sports, but since the 2000s, data analytics has become an increasingly important part of developing and running a successful sports team. This revolution in data has also resulted in new ways of measuring what it means for a player or team to be effective.

If you're feeling a little rusty on the details of professional basketball, here's how it's played:

A game of basketball is played between two teams, each with five players. The objective is to score more points than the opposing team by shooting a ball through a hoop/basket. Players can score for their team in a variety of ways – point values are assigned to the location of the shot.

A basket made from inside the "three-point line" is worth two points, while a shot made from beyond the line is worth three points. "Free throws" can also be awarded to a player or a team when the opposing team commits a foul or breaks a rule. These are worth one point each.

HOW IT WORKS: Follow the prompts in the questions below to investigate your data. Post your answers in the provided boxes: the **yellow boxes** for the queries you write and **blue boxes** for text-based answers. When you're done, export your document as a pdf file and submit it on the Milestone page – see instructions for creating a PDF at the end of the Milestone.

RESOURCES: If you need hints on the Milestone or are feeling stuck, there are multiple ways of getting help. Attend Drop-In Hours to work on these problems with your peers, or reach out to the HelpHub if you have questions. Good luck!

PROMPT: In this Milestone, you'll be looking at the way that professional basketball in the NBA has changed over seventeen recent seasons. If you were a coach in the league, what could you say about how the game is being played, and what are the most successful teams doing to be successful?

SQL App: [Here's that link](#) to our specialized SQL app, where you'll write your SQL queries and interact with the data.

– Data Set **Description**

The NBA games dataset (`nba.games`) contains information about 23 335 games played from the 2004 season through the 2020 season. There are eighteen columns in the dataset, of which the following will be used in the Milestone:

- **season** - Starting year for the season the game was played. For example, games that are part of the 2010–11 season will have a season value of 2010, even if they are played in 2021.
 - **team_home, team_away** - Full name of the home and visiting teams, respectively. Names will always reflect their current franchise names, even if they were known by a different name in prior years.
 - **pts_home, pts_away** - Number of points scored by the home and visiting teams, respectively, in each game.
 - **home_team_win** - Flag indicating whether the home team won (1) or the visiting team won (0).
 - **pct_3p_home, pct_3p_away** - Percentage of 3 point shots made by the home team and away team, respectively.
-

– Task 1: Game Statistics Trends Over Time

- A. Start by calculating the total number of rows and the first & last seasons in the dataset. This should be done in one query. If done correctly, the number of games is 23 335, the first season represented is 2004, and the last season represented is 2020.

```
SELECT
COUNT (*) AS total_amount_of_games_in_2004_and_2020,
MIN (season) AS _2004_,
MAX (season) AS _2020_
FROM nba.games
```

- B. Write a query that returns the average score from the home team, away team, and the average of the home_team_win column. The average of the home_team_win column can be interpreted as the win rate for the home team. What do these values tell you about what you can expect from the result of a random NBA game?

```
SELECT
AVG(pts_home) AS average_home_score,
AVG(pts_away) AS average_away_score,
AVG(home_team_win) AS home_team_win_rate
FROM nba.games
```

The average win rate for the home team is 58.96%. I would expect the results of a random NBA game to be an interesting game! Either the home or away team could win, however the home team would have a slight advantage. Also, the amount of points that

each team would score would be in the high nineties or the low hundreds.

- C. Modify your query from part B, so that the average scores from the home team, away team, and the home team win rate are grouped by each NBA season. Sort your output so that the seasons are ordered chronologically. What can you say about the trend in these values over the years?

```
SELECT season,  
AVG(pts_home) AS average_home_score,  
AVG(pts_away) AS average_away_score,  
AVG(home_team_win) AS home_team_win_rate  
FROM nba.games  
GROUP BY season  
ORDER BY season ASC
```

The amount of average points that were scored at home was consistently greater than the average amount of points scored away each season as time progressed.

Another trend over the years, was the amount of average points that were scored for both teams slowly increased as time progressed.

Lastly the average percentage of win rate for the home team stayed around 60% until 2012 when the average win rate then started to decrease as time progressed.

- D. Add two more summaries to your query from part C, to get the average 3-point shot rate for both away and home teams. Do these values change over time?

```
SELECT season,  
AVG(pts_home) AS average_home_score,  
AVG(pts_away) AS average_away_score,  
AVG(home_team_win) AS home_team_win_rate,  
AVG(pct_3p_home) AS percentage_of_3point_shots_made_at_home,  
AVG(pct_3p_away) AS percentage_of_3point_shots_made_away  
FROM nba.games  
GROUP BY season  
ORDER BY season ASC
```

The average percentage of 3 shots that were made by both the home and away team stayed very balanced as time progressed, both achieving almost the same percentage each season. Also the average percentage of 3 shots made by both teams stayed between the ranges of 34 and 36% over the years.

– Task 2: Investigating 3-point Shooting

The average three-point shot rate is about 35.4% over the entire dataset. Let's write some queries to investigate just how important a high three-point shot rate is in terms of winning games.

- A.** Write a query that returns the average home team win rate and average three-point percentage at home grouped by home team name and season. (You should get a table with 510 rows.)

```
SELECT season, team_home,  
AVG(home_team_win) AS home_team_win_rate,  
AVG(pct_3p_home) AS  
avg_percentage_of_3point_shots_made_at_home
```

```
FROM nba.games
GROUP BY season, team_home
ORDER BY season ASC
```

- B.** Modify your query so we are only looking at results from 2018 or later. Remember, the season column is a text field – don't forget your quotes! (This should reduce your results down to 90 rows.)

```
SELECT season, team_home,
AVG(home_team_win) AS home_team_win_rate,
AVG(pct_3p_home) AS
avg_percentage_of_3point_shots_made_at_home
FROM nba.games
WHERE season >= '2018'
GROUP BY season, team_home
ORDER BY season ASC
```

- C.** Add another expression to your query to answer the following question: How many teams had a three-point shot rate of at least 37% (i.e. 0.37)? (You'll get this from the output of the SQL app interface, rather than directly from the query.)

```
SELECT season, team_home,
AVG(home_team_win) AS home_team_win_rate,
AVG(pct_3p_home) AS percentage_of_3point_shots_made_at_home
FROM nba.games
WHERE season >= '2018'
GROUP BY season, team_home
HAVING AVG (pct_3p_home) >= '.37'
ORDER BY season, AVG(pct_3p_home) ASC
```

The amount of teams that had a 37% or more three point shot average that was during the season 2018 to 2020 are 25 teams.

- D. Add an additional condition to your query to filter to teams with a losing record (win rate < 0.5), in addition to having a high three-point shot rate. (As with the previous part, you'll read this from the SQL app interface instead of directly from the query.)

```
SELECT season, team_home,  
AVG(home_team_win) AS home_team_win_rate,  
AVG(pct_3p_home) AS percentage_of_3point_shots_made_at_home  
FROM nba.games  
WHERE season >= '2018'  
GROUP BY season, team_home  
HAVING AVG (pct_3p_home) >= '.37' AND AVG(home_team_win) <  
' .50'  
ORDER BY season, AVG(pct_3p_home) ASC
```

The amount of teams that had a 37% or more three point shot average that was during the season 2018 to 2020 and had a losing record is 2 teams.

- E. Repeat parts C and D, but this time filtering to teams that had a low 3-point shooting rate of 34% (0.34) or less. How many teams had this low of a 3-point accuracy, and how many of these teams had a losing record? (Paste only the query that answers the last question into the query box.)

```
SELECT season, team_home,  
AVG(home_team_win) AS home_team_win_rate,  
AVG(pct_3p_home) AS percentage_of_3point_shots_made_at_home  
FROM nba.games  
WHERE season >= '2018'  
GROUP BY season, team_home  
HAVING AVG (pct_3p_home) <= '.34' AND AVG(home_team_win) <  
' .50'  
ORDER BY season, AVG(pct_3p_home) ASC
```

The amount of teams that had a 34% or less three point shot average that was during the season 2018 to 2020 and had a losing record is 7 teams.

– Level Up: Building the Team Stats table

Most of the time when working with data, you will have to build summary tables yourself. This is done to alleviate storage costs, especially since these tables have to constantly be updated via a SQL query. In this Level Up you'll create two summary tables, one for the home team and one for the away team. Although you don't have the ability to join these tables (yet!), you can come back to this LevelUp and create the full table once you have learned how to join data.

- A. Write a query that returns the average number of home points scored, average 3 point percentage for the home team, and the number of wins for each team and season combination in the `nba.games` table. (You should get a table with 510 rows.)


```
SELECT team_home AS team_name,  
season,  
AVG(pts_home) AS avg_home_points,  
AVG(pct_3p_home) AS avg_3point_percentage_home,  
SUM(home_team_win) AS home_team_wins  
FROM nba.games  
GROUP BY team_home, season  
ORDER BY season
```

- B.** Repeat part A but now do the same thing for the **away** team. Note that you will have to get a little creative to calculate the number of away wins since the table only tells you whether or not the home team won the game. (You should get a table with 510 rows.)

```
SELECT team_away AS team_name,  
season,  
AVG(pts_away) AS avg_away_points,  
AVG(pct_3p_away) AS avg_3point_percentage_away,  
COUNT(*) - SUM(home_team_win) AS away_wins  
FROM nba.games  
WHERE home_team_win = 0  
GROUP BY team_away, season  
ORDER BY season
```

– Submission

Great work completing this Milestone! To submit your completed Milestone, you will need to download / export this document as a PDF and then upload it to the

Milestone submission page. You can find the option to download as a PDF from the File menu in the upper-left corner of the Google Doc interface.