



Lab | 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.

PROMPT: In this lab, 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?

– 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.

(paste your query below 📌)

```
SELECT
    COUNT(*) AS n_games,
    MIN(season) AS first_szn,
    MAX(season) AS last_szn
```

```
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?

(paste your query below 📌)

```
SELECT
    AVG(pts_home) AS avg_pts_home,
    AVG(pts_away) AS avg_pts_away,
    AVG(home_team_win) AS win_rate
FROM nba_games
```

(write your **answer** below 📌)

The average points scored per game is higher for the home team (103.2) than the visiting (100.4), which suggests that there is a home-team advantage: in other words, the home team wins more games than the away team. This is also shown in the home team win rate, which says that the home team wins almost 59% (0.5896) of their games.

- 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?

(paste your query below 📌)

```
SELECT
    season,
    AVG(pts_home) AS avg_pts_home,
    AVG(pts_away) AS avg_pts_away,
    AVG(home_team_win) AS home_win_rate
FROM nba_games
GROUP BY season
ORDER BY season ASC
```

(write your **answer** below 📌)

Game scores have been generally increasing over the period shown in the data, with an especially large jump in 2018 through 2020.

On the other hand, the home team advantage seems to have decreased over time, starting at around 60% from 2004 to 2012, and 58–59% from 2013 to 2018. 2019 and 2020 saw a much more reduced home team advantage, at 53.7% in 2019 and 54.5% in 2020. Generally, it seems like years with smaller differences in average score between the home and away teams have a smaller home team advantage, which makes sense.

As for why there may have been such a drop in home team advantage for 2019–20 and 2020–21, there were restrictions late in the 2019 season and throughout the 2020 season where games were played with limited capacity to no spectators at all. Perhaps a lack of a cheering crowd (or jeering crowd, for a visiting team) dampened their normal effects on the players!

- 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?

(paste your query below 📌)

```
SELECT
    season,
    AVG(pts_home) AS avg_pts_home,
    AVG(pts_away) AS avg_pts_away,
    AVG(home_team_win) AS home_win_rate,
    AVG(pct_3p_home) AS pct_3p_home,
    AVG(pct_3p_away) AS pct_3p_away
FROM nba_games
GROUP BY season
ORDER BY season ASC
```

(write your **answer** below 📌)

Despite the fact that the average points per game is increasing over time, there doesn't seem to be a strong trend in 3-point shooting percentage.

The increase in scoring can't really be accounted for in the data we have here. Perhaps teams are just making more 3-point shots, but maintaining the same accuracy. Or maybe the game is being played faster, and teams are making more shots in general.

– 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.

Note: you will not be looking at the away team in this analysis.

(You should get a table with 510 rows.)

(paste your query below 📌)

```
SELECT
    season,
    team_home,
    AVG(home_team_win) as home_win_rate,
    AVG(pct_3p_home) as avg_3p_rate
FROM nba_games
GROUP BY
    season,
    team_home
```

- 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.)

(paste your query below 📌)

```
SELECT
    season,
    team_home,
    AVG(home_team_win) as home_win_rate,
    AVG(pct_3p_home) as avg_3p_rate
FROM nba_games
WHERE season >= '2018'
GROUP BY
    season,
    team_home
```

- 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.)

(paste your query below 📌)

```
SELECT
    season,
    team_home,
    AVG(home_team_win) as home_win_rate,
    AVG(pct_3p_home) as avg_3p_rate
FROM nba_games
WHERE season >= '2018'
GROUP BY
    season,
    team_home
HAVING
    AVG(pct_3p_home) >= 0.370
```

(write your **answer** below 📌)

Twenty-five (non-unique) teams hit 3-pointers at a better than 37% rate from 2018–19 to 2020–21.

- 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. How many teams had a losing record while having a high 3-point shot percentage? (As with the previous part, you'll read this from the SQL app interface instead of directly from the query.)

(paste your query below 📌)

```
SELECT
    season,
    team_home,
```

```
    AVG(home_team_win) as home_win_rate,  
    AVG(pct_3p_home) as avg_3p_rate  
FROM nba_games  
WHERE season >= '2018'  
GROUP BY  
    season,  
    team_home  
HAVING  
    AVG(pct_3p_home) >= 0.370  
    AND AVG(home_team_win) < 0.5
```

(write your **answer** below 📌)

Only two teams had a losing record while having a high 3-point shot percentage. That's 8% of the teams we found in part C.

- 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.)

(paste your query below 📌)

```
SELECT  
    season,  
    team_home,  
    AVG(home_team_win) as home_win_rate,  
    AVG(pct_3p_home) as avg_3p_rate  
FROM nba_games  
WHERE season >= '2018'  
GROUP BY  
    season,  
    team_home  
HAVING  
    AVG(pct_3p_home) <= 0.340
```



```
AND AVG(home_team_win) < 0.5
```

- Remove the last line in the query to get just the teams with low 3-point accuracies regardless of win record.

(write your **answer** below 📌)

There are eleven (non-unique) teams with a 3-point shot accuracy of 34% or less, and seven of them have losing records.

- F. What conclusions can you draw from your analysis in the previous Parts regarding the relationship between 3-point shot rates and team success?

(write your **answer** below 📌)

Only 64% of the teams with a low 3-point shooting rate have a losing record, which suggests that a low 3-point rate is not as harmful as a high 3-point rate is beneficial, since only two teams had a losing record while having a high 3-point shot percentage. Another possible interpretation is that there are successful strategies or game play styles that aren't focused around the 3-point shot.

– LevelUp: 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.)

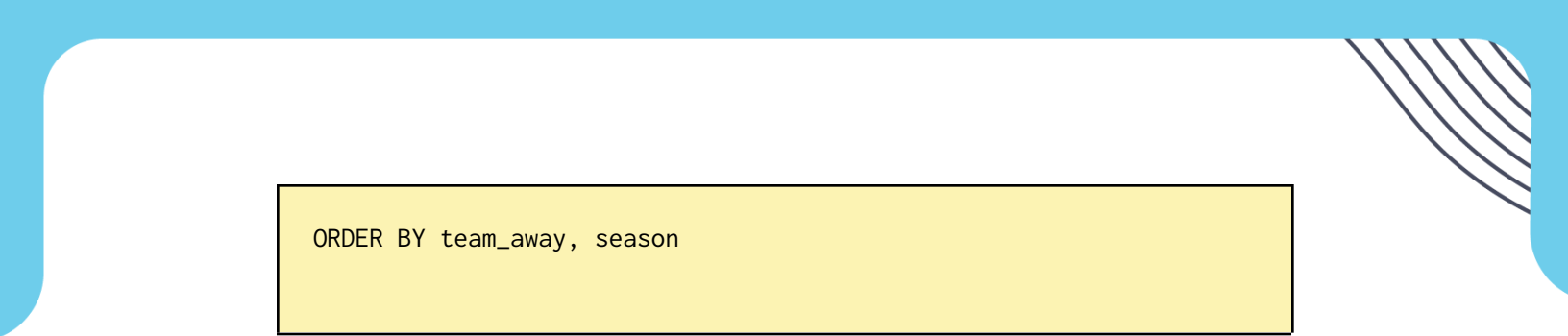
(paste your query below 📌)

```
SELECT
    team_home,
    season,
    AVG(pts_home) as avg_pts_home,  AVG(pct_3p_home) as
avg_3p_home,
    SUM(home_team_win) as home_team_wins
FROM nba_games
GROUP BY team_home, season
ORDER BY team_home, 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.)

(paste your query below 📌)

```
SELECT
    team_away,
    season,
    AVG(pts_away) as avg_pts_away,
    AVG(pct_3p_away) as avg_3p_away,
    -- only one method below is needed, but both are valid to -
    calculate away wins
    COUNT(home_team_win) - SUM(home_team_win) as num_away_wins,
    SUM(1 - home_team_win) as num_away_wins_2
FROM nba_games
GROUP BY team_away, season
```



```
ORDER BY team_away, season
```