⊘ **Milestone 4** | NBA Performance Analysis

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.

In this Milestone, you'll take a shot at being a sports analyst! You'll explore how NBA team performance and playing styles have changed over the past seventeen seasons. Using real data, you'll uncover trends in scoring, shot selection, and strategy. As a data analyst supporting a coaching staff, your goal is to identify patterns tied to team success and make evidence-based recommendations. Your insights could help inform decisions around training, resource allocation, and game strategy—just like in a real-world analytics role.

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 <mark>yellow boxes</mark> for the queries you write and <mark>blue boxes</mark> 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. Please <u>don't *ever*</u> remove (paste your query below 👇) or (write your **answer** below 👇). These help your Evaluator!

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.

HINT: 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 total_games,
    min(season) as first_season,
    max(season) as last_season
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.

(paste your query below 👇)

```
select
    round(avg(pts_home), 2) as avg_home_score,
    round(avg(pts_away), 2) as avg_away_score,
    round(avg(home_team_win), 2) as avg_home_wins
from
    nba.games
```

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?

(write your **answer** below 👇)

Home teams are more likely to win as they win 59 percent of games and on average score 3 more points.

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.

(paste your query below 👇)

```
select
   season,
   round(avg(pts_home), 2) as avg_home_score,
   round(avg(pts_away), 2) as avg_away_score,
   round(avg(home_team_win), 2) as avg_home_wins
from
   nba.games
group by
   season
order by
   season asc
```

What can you say about the trend in these values over the years?

(write your **answer** below 👇)

From 2004 to 2020 the average scoring increased from approximately 98 points per game to 112 points per game and also the average home wins dropped from 60 to 55 percent showing home court advantage is not as strong as it once was.

D. As a data analyst, your job isn't just to observe patterns—it's to help others understand the story behind the data. Now that you've identified a trend in home team win rates over time, it's time to dig deeper into why this might be happening.

Based on ChatGPT's response, what's one possible explanation for the decline in home-court advantage?

(write your **answer** below 👇)

One possible explanation ChatGPT said was that the game has transitioned to more 3 point focused scoring so a familiar home environment is not as important and also since then there has been improvements in travel, training, and also consistency with the refs for each game.

E. Add two more summaries to your query from part C, to get the average 3-point shot rate for both away and home teams.

(paste your query below 👇)

```
select
    season,
    round(avg(pts_home), 2) as avg_home_score,
    round(avg(pts_away), 2) as avg_away_score,
    round(avg(home_team_win), 2) as avg_home_wins,
    round(avg(pct_3p_home), 2) as avg_home_3pt,
    round(avg(pct_3p_away), 2) as avg_away_3pt
from
    nba.games
group by
    season
order by
    season asc
```

Do these values change over time?

No, the average 3pt for both home and away have been consistent the past 15 years with only really a 0.01 fluctuation throughout the seasons.

— 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. If done correctly, your query should result in a table with 510 rows.

(paste your query below 👇)

```
select
    team_home,
    season,
    round(avg(home_team_win), 2) as avg_home_wins,
    round(avg(pct_3p_home), 2) as avg_home_3pt
from
    nba.games
group by
    team_home,
    season
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.)

(paste your query below 👇)

```
select
    team_home,
    season,
    round(avg(home_team_win), 2) as avg_home_wins,
    round(avg(pct_3p_home), 2) as avg_home_3pt
from
    nba.games
where
    season >= '2018'
group by
    team_home,
    season
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)? **HINT:** You'll get this from the output of the SQL app interface, rather than directly from the query.

Try this prompt: I need to modify this SQL query to count how many teams had an average three-point shooting percentage of at least 37% (0.37) from 2018 onward: <INSERT QUERY FROM 2B> How should I adjust my query to return only teams that meet this threshold, and should I use

(paste your query below 👇)

```
select
   team_home,
   season,
   round(avg(home_team_win), 2) as avg_home_wins,
   round(avg(pct_3p_home), 2) as avg_home_3pt
from
   nba.games
where
   season >= '2018'
group by
   team_home,
   season
having
   avg(pct_3p_home) >= 0.37
order by
   season asc
```

(write your **answer** below 👇)

25

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.

(paste your query below 👇)

```
select
   team_home,
   season,
   round(avg(home_team_win), 2) as avg_home_wins,
```

```
    round(avg(pct_3p_home), 2) as avg_home_3pt
  from
    nba.games
  where
    season >= '2018'
  group by
    team_home,
    season
  having
    avg(pct_3p_home) >= 0.37
    and avg(home_team_win) < 0.5
  order by
    season asc
```

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.)

(write your **answer** below 👇)

> 2

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
    team_home,
    season,
```

```
    round(avg(home_team_win), 2) as avg_home_wins,
    round(avg(pct_3p_home), 2) as avg_home_3pt
from
    nba.games
where
    season >= '2018'
group by
    team_home,
    season
having
    avg(pct_3p_home) <= 0.34
    and avg(home_team_win) < 0.5
order by
    season asc
```

(write your **answer** below 👇)

> 7

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 👇)

> The lower the 3pt percentage the more games a team loses and the higher the 3pt percentage the more games a team wins as we saw only 2 had a losing record with a high 3pt percentage while 7 had a losing record with a low 3pt percentage. The 3pt shot is crucial for winning games.

G. You've started to explore how 3-point shooting relates to team success—but basketball strategy is complex, and winning teams often rely on more than one approach.

Based on ChatGPT's response, what other metrics should we investigate to get a more complete picture of team success?

(write your **answer** below 👇)

> The other metrics ChatGPT said to investigate are other ways of scoring, as some teams can dominate with paint scoring and also scoring from the free throw line showing a more aggressive approach of getting to the paint to score or draw fouls. Also, there are 2 sides of the ball and another metric is how good a team defense is, showing what is their opponent's average field goal percentage or points allowed.

— 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 LevelUp 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.

   HINT: You should get a table with 510 rows.

(paste your query below 👇)

```
select
    team_home,
    season,
    round(avg(pts_home), 2) as avg_home_pts,
    round(avg(pct_3p_home), 2) as avg_3pt_home_prcnt,
    sum(home_team_win) as total_home_wins
from
    nba.games
group by
    team_home,
    season
order by
    season asc,
    team_home
```

B. Repeat Part A for the **away** team. You will have to be creative to calculate the number of away wins since the table only tells you whether or not the home team won the game.

If done correctly, your query should result in a table with 510 rows.

(paste your query below 👇)

```
select
    team_away,
    season,
    round(avg(pts_away), 2) as avg_away_pts,
    round(avg(pct_3p_away), 2) as avg_3pt_away_prcnt,
    sum(1 - home_team_win) as total_away_wins
from
    nba.games
group by
    team_away,
    season
```

```
order by
    season asc,
    team_away
```

— 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.