# SQL Test for Analytics Engineer

### **Getting the Data**

NBA Team Data is available from several freely accessible sources. We recommend that you use this following guide, and pull the data from this API:<https://api-sports.io/documentation/nba/v2#section/Introduction>

Please include the scripts you used to gather the data and to clean it into the following formats.

When you load the data note that we only care about the teams that are in the standard league and are NBA teams.

### **Data Assumptions**

# Assume you have two tables:

# games:

# game\_id (INT)

# season (YEAR)

# date (DATE)

# home\_team (VARCHAR)

# away\_team (VARCHAR)

# home\_score (INT)

# away\_score (INT)

# teams:

# team\_id (INT)

# team\_name (VARCHAR)

# conference (VARCHAR)

# division (VARCHAR)

### **Test Tasks**

#### **Task 1: Basic Data Extraction**

# Write a query to retrieve the top 10 highest-scoring games in the last decade.

1. You have a games table with columns including date, home\_team, away\_team, home\_score, and away\_score.
2. The total score for a game is the sum of home\_score and away\_score.
3. Focus on games played in the last decade.
4. Use the season column to filter games that occurred within the last 10 years.

#### **Task 2: Win-Loss Records**

# Write a query to calculate the win-loss record for each team over the last decade.

#### **Task 3: Team Performance by Season**

# Write a query to calculate the average points scored by each team per season over the last decade.

#### **Task 4: Conference Analysis**

# Write a query to determine which conference (East or West) has had the most wins in the last decade.

#### **Task 5: Detailed Game Analysis**

# Write a query to find the team with the highest average margin of victory in the last decade.

Understanding the Margin of Victory:

1. The margin of victory is calculated as the difference between the points scored by the winning team and the points scored by the losing team in each game.
2. For each game, you will need to calculate this margin for both home and away games.

#### **Task 6: Analyzing Team Performance Over Multiple Seasons**

# In NBA basketball, "average points allowed per game" is a important statistical concept. It refers to the average number of points a team allows its opponents to score in each game over a specified period. This metric is important for analyzing a team's defensive performance. It helps to understand how well a team is preventing its opponents from scoring.

# For example, if a team plays 10 games and allows a total of 950 points across those games, their average points allowed per game would be 95 (950 total points / 10 games).

# Write a query to analyze the performance trends of NBA teams over the last decade. For each team, calculate the average points scored per game and the average points allowed per game for each season. Use CTEs to structure your query.

* Use CTEs to separate the data preparation and calculation steps.
* Display the results showing the team name, season, average points scored per game, and average points allowed per game, ordered by team name and season.

### **Instructions**

# You have 2 hours to complete all tasks.

# Write your queries in a SQL script file.

# Ensure your queries are well-formatted and optimized.

# You may add comments to explain your thought process if necessary.

# If you encounter any issues or have assumptions, document them clearly.

### **Bonus**

Write a python program that dynamically loads this data from one of the available sources like API-Sports.io and loads this data into duckdb and/or parquet.