**Team Members:**

1. Logan Collins
2. Eli Burt
3. Austin Middlebrooks
4. Noelle Rocheteau
5. Rett Sams

**Problem Description:**

The objective of this project is to design, model, and implement a comprehensive relational database for the 2024 NFL season. The core focus of the system is the NFL Team entity, which is connected to related entities such as Players, Coaches, Referees, Broadcasters, and other operational components of the league.

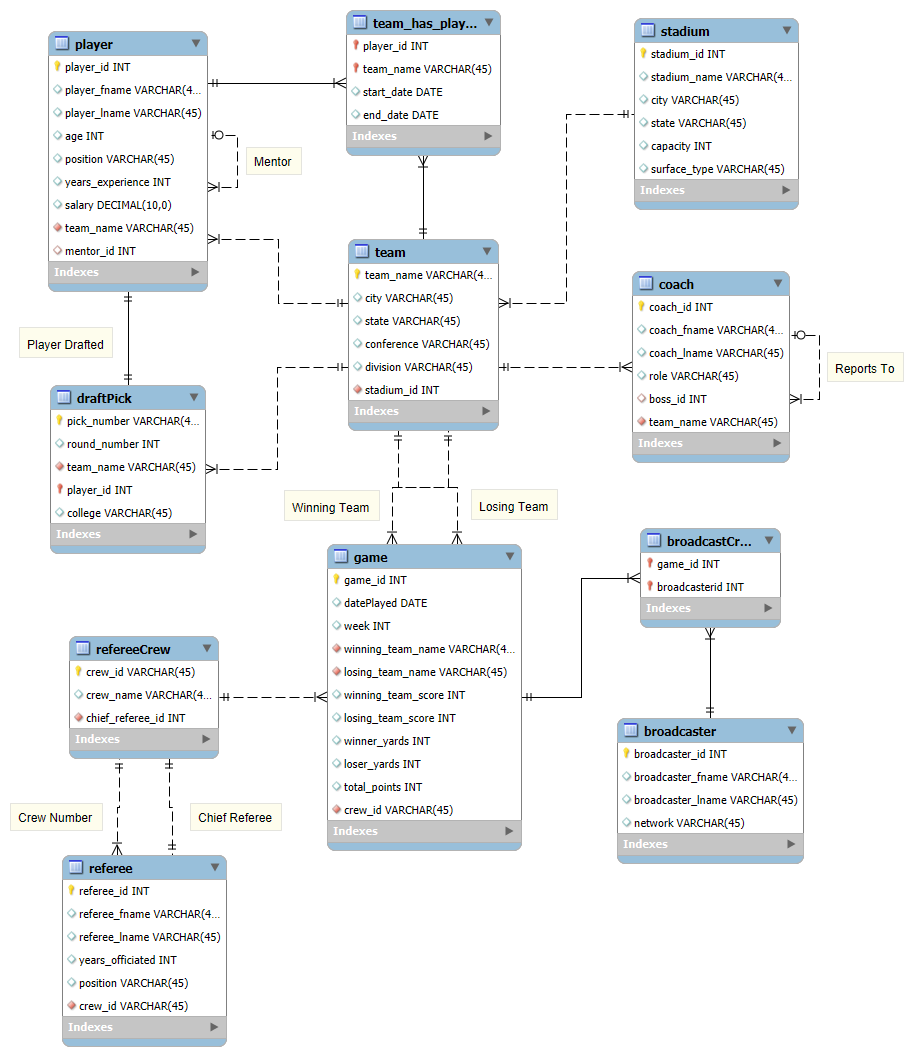
The project focuses on:

* Developing a complete Entity-Relationship model that accurately represents the structure and interactions of NFL organizations during the 2024 season.
* Populating the database with detailed and realistic season data, including team rosters, game schedules, officiating crews, broadcast assignments, and statistical outcomes.
* Writing and executing SQL queries to efficiently retrieve meaningful insights from the database, such as team performance metrics, player statistics, and game reporting details.

By the end of the project, we will have a fully functioning database system that provides valuable visibility into the dynamics and analytics of the 2024 NFL season.

Note that we could not find complete accurate data for Broadcasting Crews and Player Start + End Dates.

**Data Model:**



**Model Explanation:**

Our model is based on the structure of the NFL, particularly for the 2024 season. The team entity represents one of the thirty-two NFL teams. Each team can have many players, and a player can also have many teams (if a player gets cut or traded), which is shown by a many-to-many relationship. We used the associative entity team\_has\_player to track the start and end dates of when the player joined each team.

A team also has a stadium, but in some instances, teams share a stadium (NYJ and NYG or LAC and LAR). This is represented by a one-to-many relationship from stadium to team. Further, a team also has multiple coaches, and coaches can have only one team. Coaches do not get traded during a season, so they cannot be part of multiple teams. This is shown by a one-to-many relationship from team to coaches. Coaches also have a hierarchy—a coach can be the boss of many other coaches, but a coach can only have one boss. This is identified through a one-to-many recursive relationship.

A team plays 18 regular-season games each year. Therefore, in our model, a team has many games. A game can only have one winner and one loser. This is modeled by two one-to-many relationships. A game must also have referees. A game can have only one referee crew and a referee crew can officiate many games over the season. This is shown by a one to many relationship from refereeCrew to game. A referee Crew can have many referees but a referee can only have one referee crew that they work with. A referee crew also has one chief referee.

Similarly, a game can have more than one broadcaster, and a broadcaster can call more than one game each season. This is yet another many-to-many relationship, which creates the associative entity broadCastCrew, tracking which broadcasters call each game.

Players in the NFL can have mentors throughout the league who help them develop their game. A player can only have one mentor but can have many other players whom they mentor. This is shown by a one-to-many recursive relationship between players. We also modeled the 2024 NFL draft. A player can be only one draft pick, and a draft pick can be only one player, so we showed this relationship using a one-to-one relationship labeled draft pick. Teams, however, can have more than one draft pick, so to show this relationship, there is a one-to-many from team to draft.

**Data Dictionary:**

### **Table: player**

| **Column Name** | **Description** | **Data type** | **Size** | **Key** |
| --- | --- | --- | --- | --- |
| player\_id | Unique identifier for each player | INT | — | PK |
| player\_fname | Player first name | VARCHAR | 45 |  |
| player\_lname | Player last name | VARCHAR | 45 |  |
| age | Player age | INT | — |  |
| position | Player position | VARCHAR | 45 |  |
| years\_experience | Years of experience | INT | — |  |
| salary | Player salary | DECIMAL | 10,0 |  |
| team\_name | Current team name (if any) | VARCHAR | 45 | FK |
| mentor\_id | Player ID of mentor (self-reference) | INT | — | FK |

### **Table: team\_has\_player**

| **Column Name** | **Description** | **Data type** | **Size** | **Key** |
| --- | --- | --- | --- | --- |
| player\_id | Player in the stint | INT | — | PK, FK |
| team\_name | Team for the stint | VARCHAR | 45 | PK, FK |
| start\_date | Stint start date | DATE | — | PK |
| end\_date | Stint end date | DATE | — |  |

### **Table: stadium**

| **Column Name** | **Description** | **Data type** | **Size** | **Key** |
| --- | --- | --- | --- | --- |
| stadium\_id | Unique identifier for stadium | INT | — | PK |
| stadium\_name | Stadium name | VARCHAR | 45 |  |
| city | City | VARCHAR | 45 |  |
| state | State | VARCHAR | 45 |  |
| capacity | Seating capacity | INT | — |  |
| surface\_type | Playing surface type | VARCHAR | 45 |  |

### **Table: coach**

| **Column Name** | **Description** | **Data type** | **Size** | **Key** |
| --- | --- | --- | --- | --- |
| coach\_id | Unique identifier for coach | INT | — | PK |
| coach\_fname | Coach first name | VARCHAR | 45 |  |
| coach\_lname | Coach last name | VARCHAR | 45 |  |
| role | Coaching role/title | VARCHAR | 45 |  |
| boss\_id | Coach ID of manager they report to | INT | — | FK |
| team\_name | Team coached | VARCHAR | 45 | FK |

### **Table: team**

| **Column Name** | **Description** | **Data type** | **Size** | **Key** |
| --- | --- | --- | --- | --- |
| team\_name | Unique team name | VARCHAR | 45 | PK |
| city | Team city | VARCHAR | 45 |  |
| state | Team state | VARCHAR | 45 |  |
| conference | Conference | VARCHAR | 45 |  |
| division | Division | VARCHAR | 45 |  |
| stadium\_id | Home stadium | INT | — | FK |

### **Table: draftPick**

| **Column Name** | **Description** | **Data type** | **Size** | **Key** |
| --- | --- | --- | --- | --- |
| pick\_number | Overall pick number | VARCHAR | 45 | PK |
| round\_number | Draft round | INT | — |  |
| team\_name | Team making the pick | VARCHAR | 45 | FK |
| player\_id | Player selected | INT | — | FK |
| college | Player’s college | VARCHAR | 45 |  |

### **Table: game**

| **Column Name** | **Description** | **Data type** | **Size** | **Key** |
| --- | --- | --- | --- | --- |
| game\_id | Unique identifier for game | INT | — | PK |
| datePlayed | Date played | DATE | — |  |
| week | Week number | INT | — |  |
| winning\_team\_name | Winning team | VARCHAR | 45 | FK |
| losing\_team\_name | Losing team | VARCHAR | 45 | FK |
| winning\_team\_score | Winning team score | INT | — |  |
| losing\_team\_score | Losing team score | INT | — |  |
| winner\_yards | Total yards by winner | INT | — |  |
| loser\_yards | Total yards by loser | INT | — |  |
| total\_points | Total points scored | INT | — |  |
| crew\_id | Crew identifier |  |  |  |

### **Table: refereeCrew**

| **Column Name** | **Description** | **Data type** | **Size** | **Key** |
| --- | --- | --- | --- | --- |
| game\_id | Game being officiated | INT | — | PK, FK |
| crew\_id | Crew identifier | VARCHAR | 45 | PK |
| crew\_name | Crew name | VARCHAR | 45 |  |
| chief\_referee\_id | Chief referee for the crew | INT | — | FK |

### **Table: referee**

| **Column Name** | **Description** | **Data type** | **Size** | **Key** |
| --- | --- | --- | --- | --- |
| referee\_id | Unique identifier for referee | INT | — | PK |
| referee\_fname | Referee first name | VARCHAR | 45 |  |
| referee\_lname | Referee last name | VARCHAR | 45 |  |
| years\_officiated | Years officiated | INT | — |  |
| position | Field position (e.g., Umpire) | VARCHAR | 45 |  |
| crew\_id | Assigned crew | INT | — | FK |

### **Table: broadcastCrew**

| **Column Name** | **Description** | **Data type** | **Size** | **Key** |
| --- | --- | --- | --- | --- |
| game\_id | Broadcast game | INT | — | PK, FK |
| broadcaster\_id | Member of broadcast crew | INT | — | PK, FK |

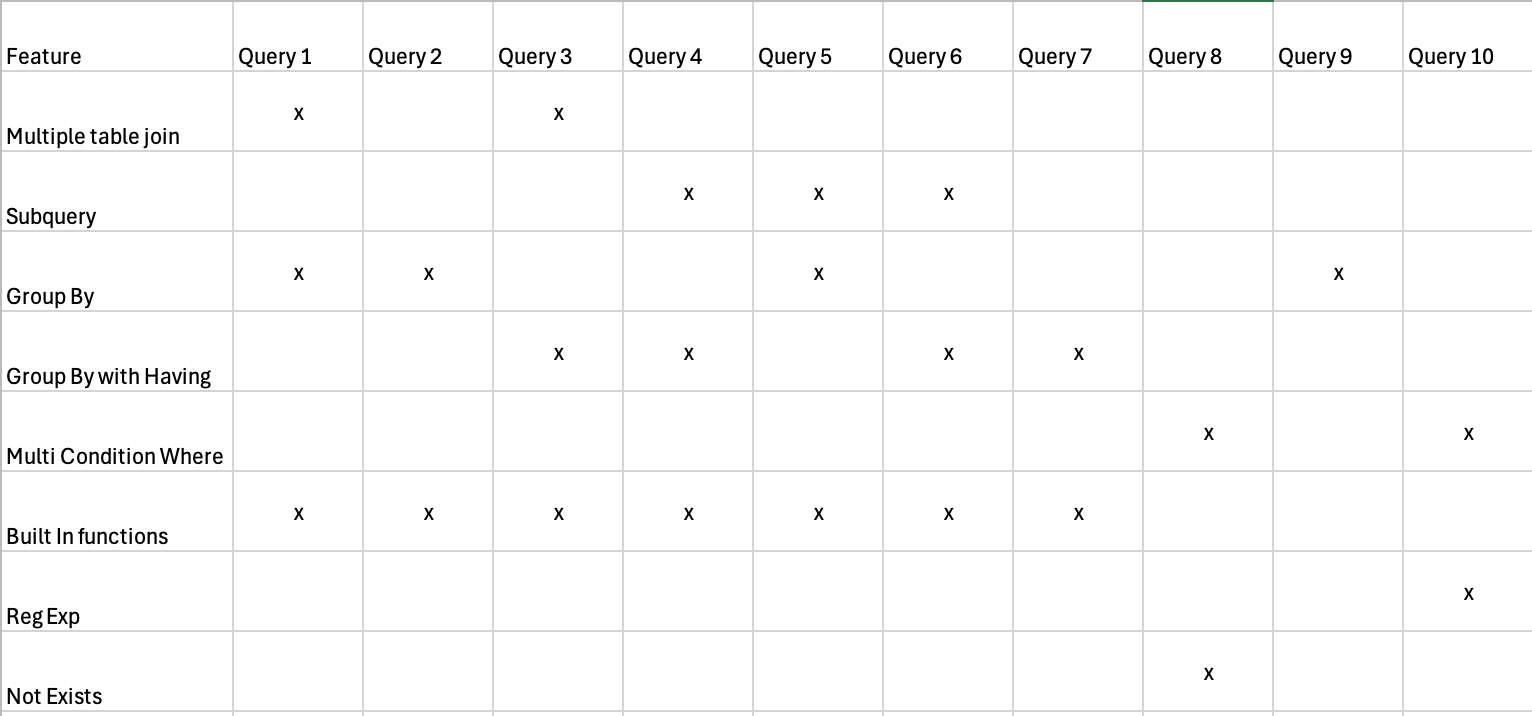
### 

### 

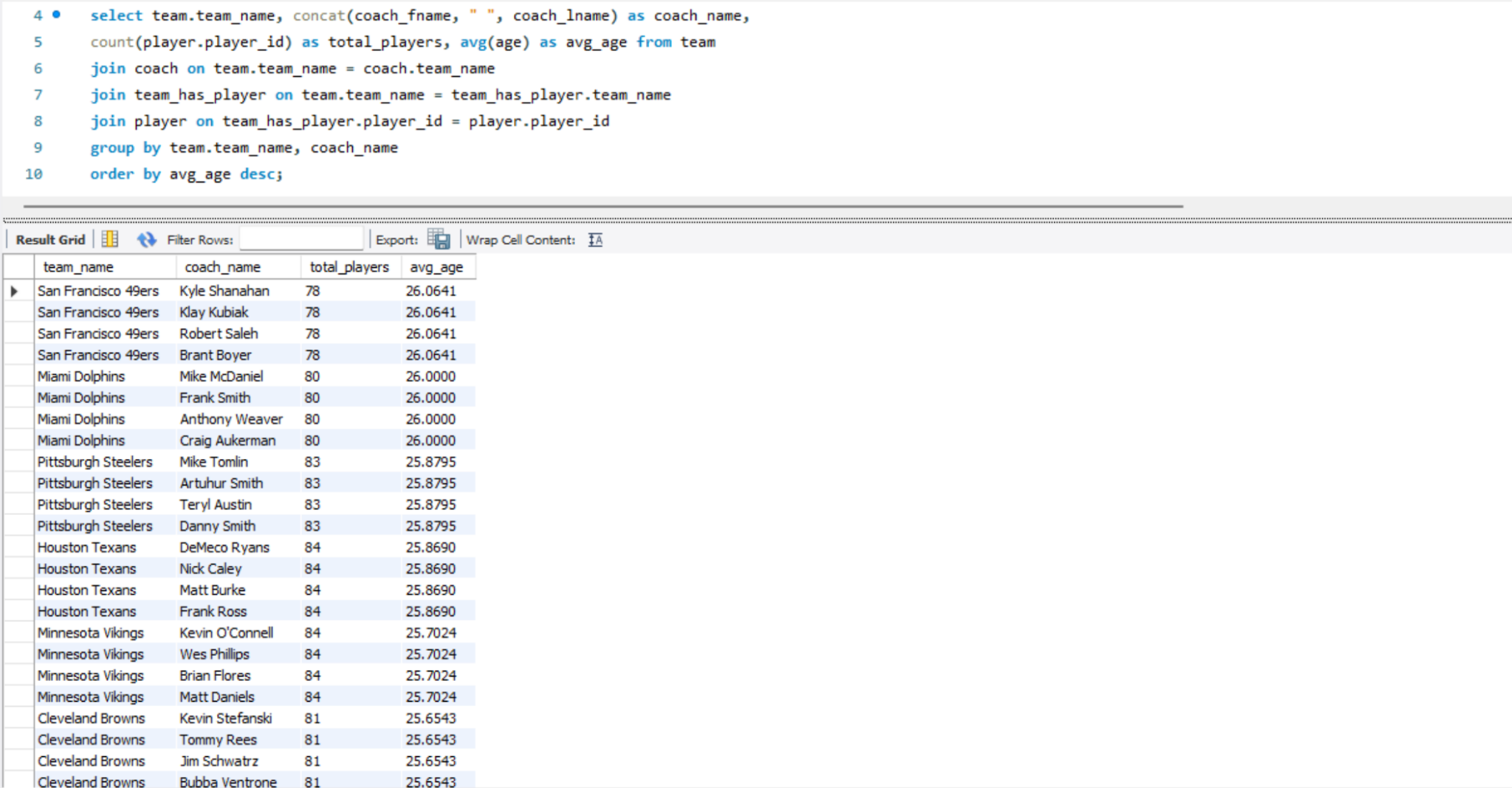
### **Table: broadcaster**

| **Column Name** | **Description** | **Data type** | **Size** | **Key** |
| --- | --- | --- | --- | --- |
| broadcaster\_id | Unique identifier for broadcaster | INT | — | PK |
| broadcaster\_fname | Broadcaster first name | VARCHAR | 45 |  |
| broadcaster\_lname | Broadcaster last name | VARCHAR | 45 |  |
| network | Network name | VARCHAR | 45 |  |

**Queries:**

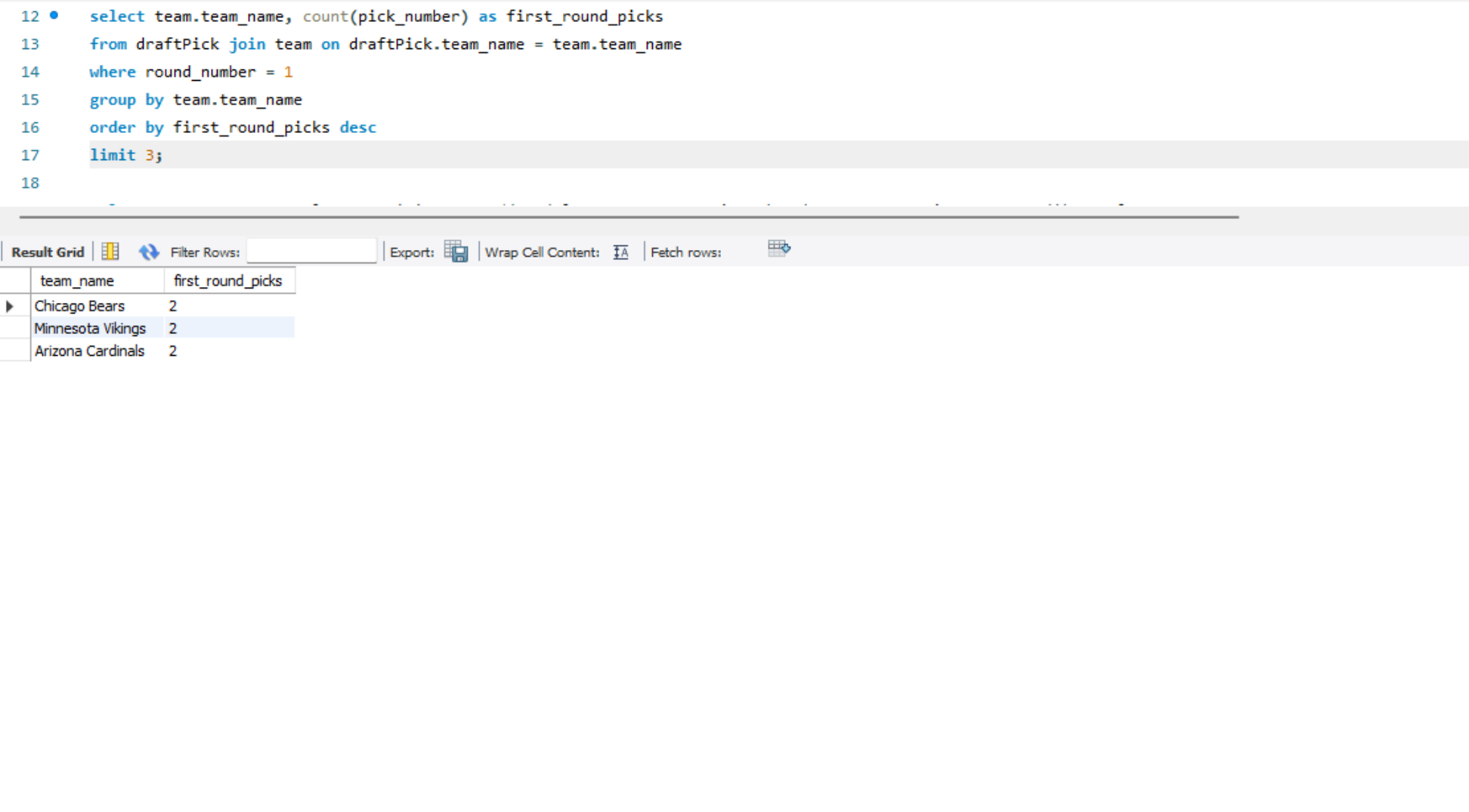


1. Write a query to list the coach, total number of players, and average player age by team.



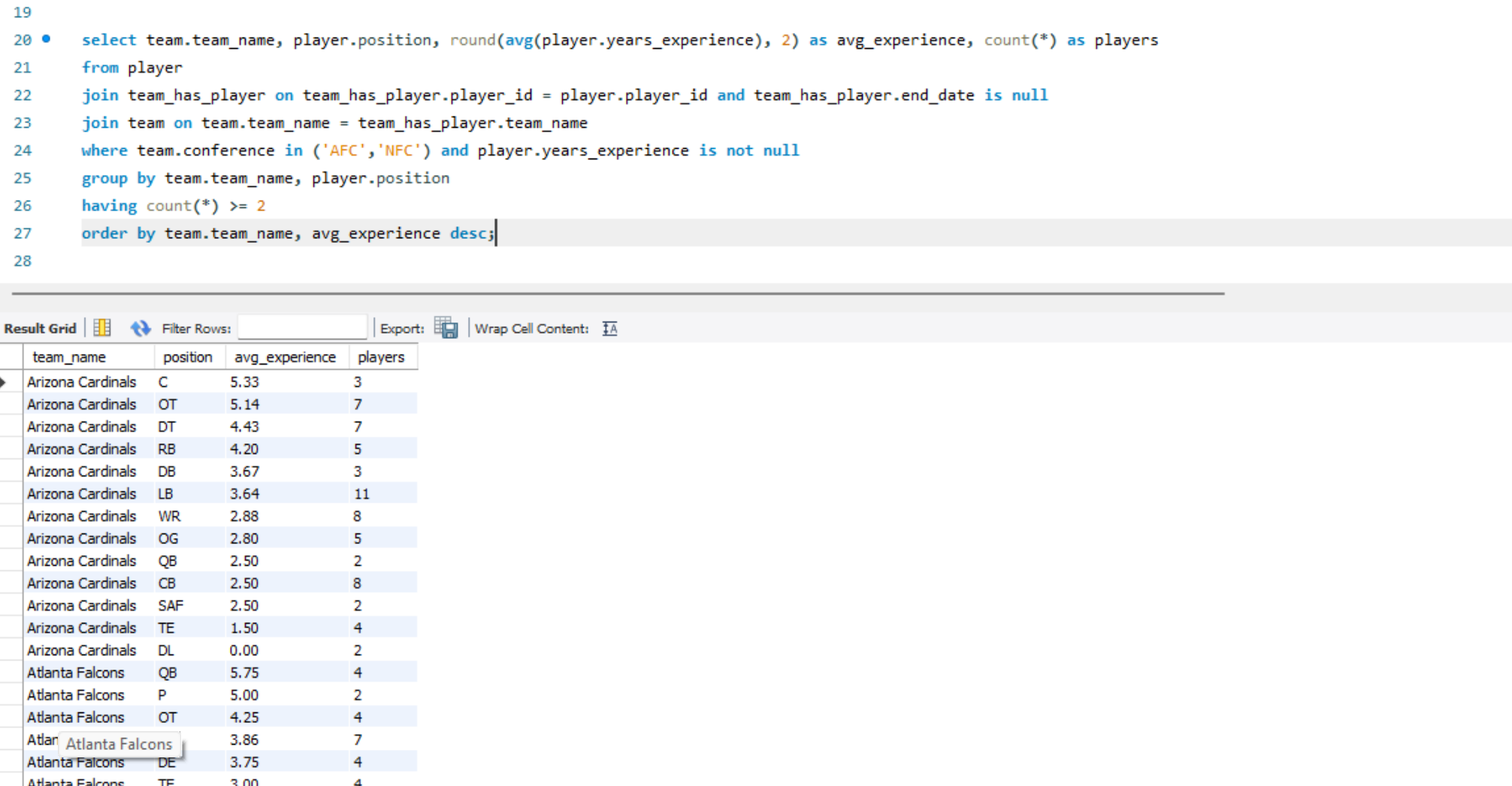
This query is meaningful because it shows roster and coaching information to provide information about the composition of each team. It highlights how team size and player age may differ under different coaches, which can be valuable for identifying roster balance, experience distribution, and potential leadership styles within each organization

1. Write a query that finds the top 3 teams with the most first round picks.



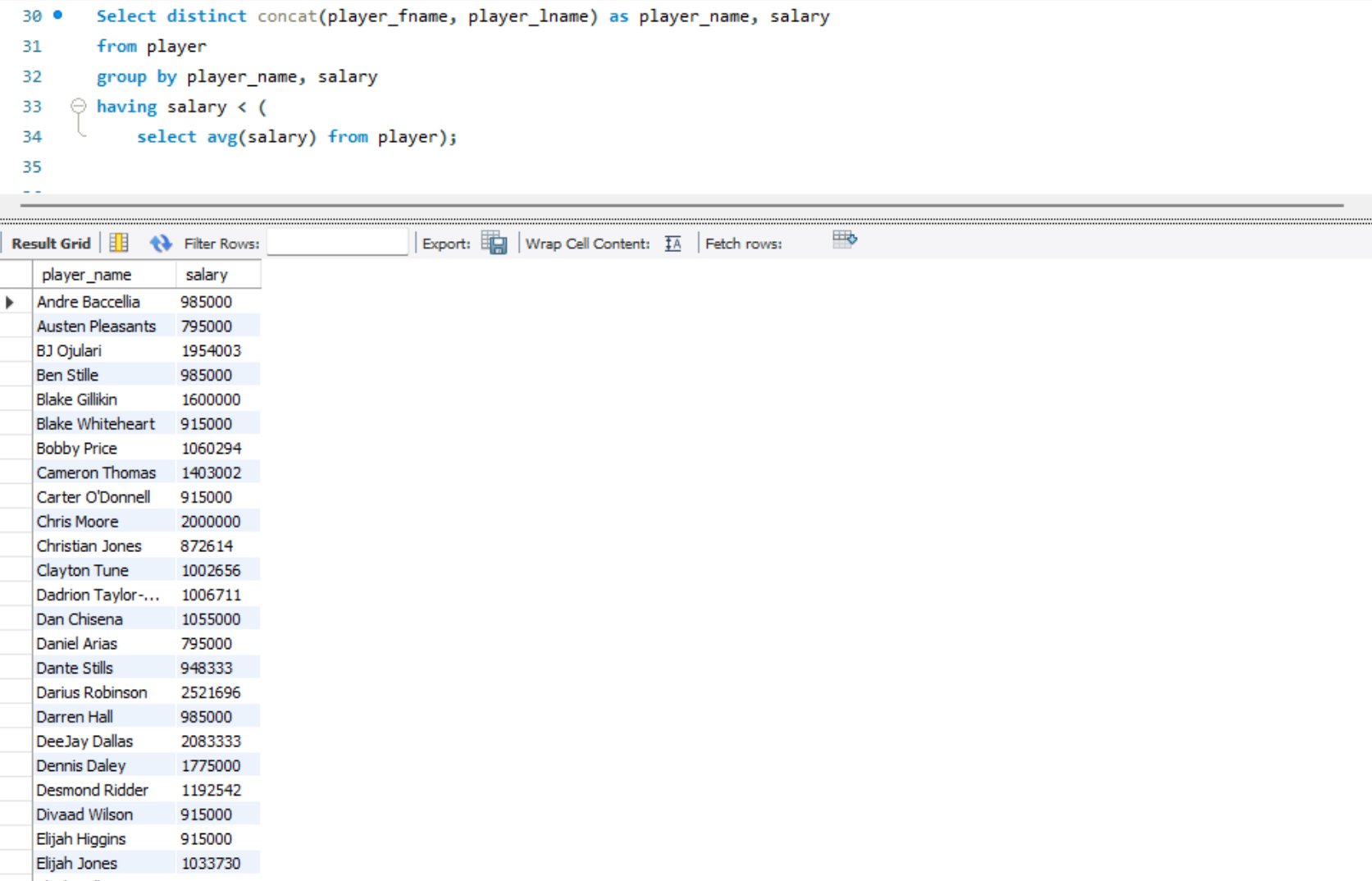
This query identifies which teams have had the most first round draft selections. By grouping draft data by team and filtering for only first round picks, it reveals which organizations have had the greatest access to top tier talent. This information is meaningful because it can help analyze how draft opportunities relate to team performance, roster strength, and overall franchise strategy.

1. Write a query that lists the avg years of experience by team & position(only show positions with ≥ 2 active players).



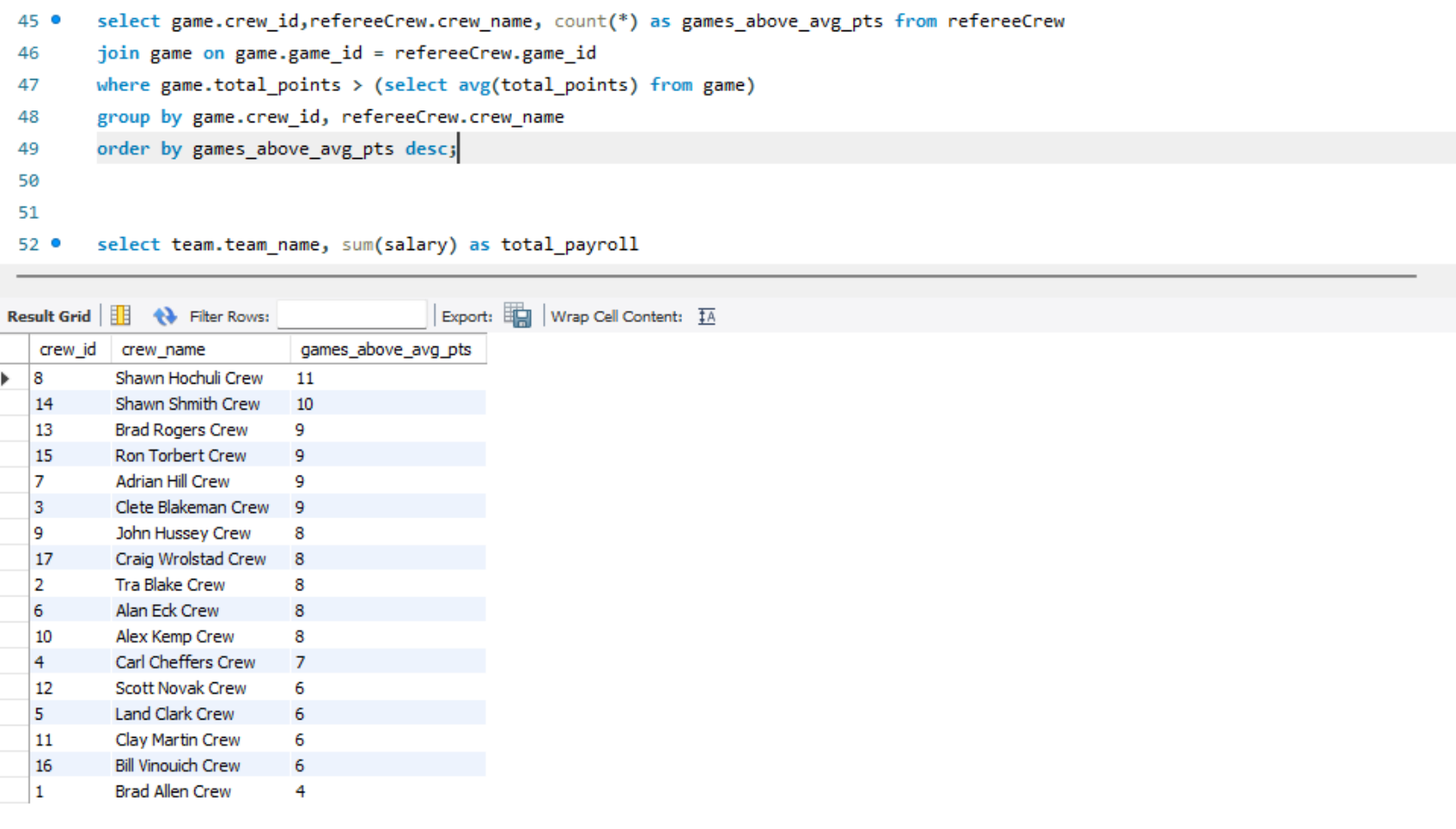
This query calculates the average years of experience for each position within every team, while only including positions that have at least two active players. It is meaningful because it provides insight into roster depth and experience balance across positions. Teams can use this data to evaluate whether certain positions rely more on seasoned veterans or less experienced players, which is valuable for assessing player development and recruitment strategies.

4) Write a query that retrieves players earning below average salary.



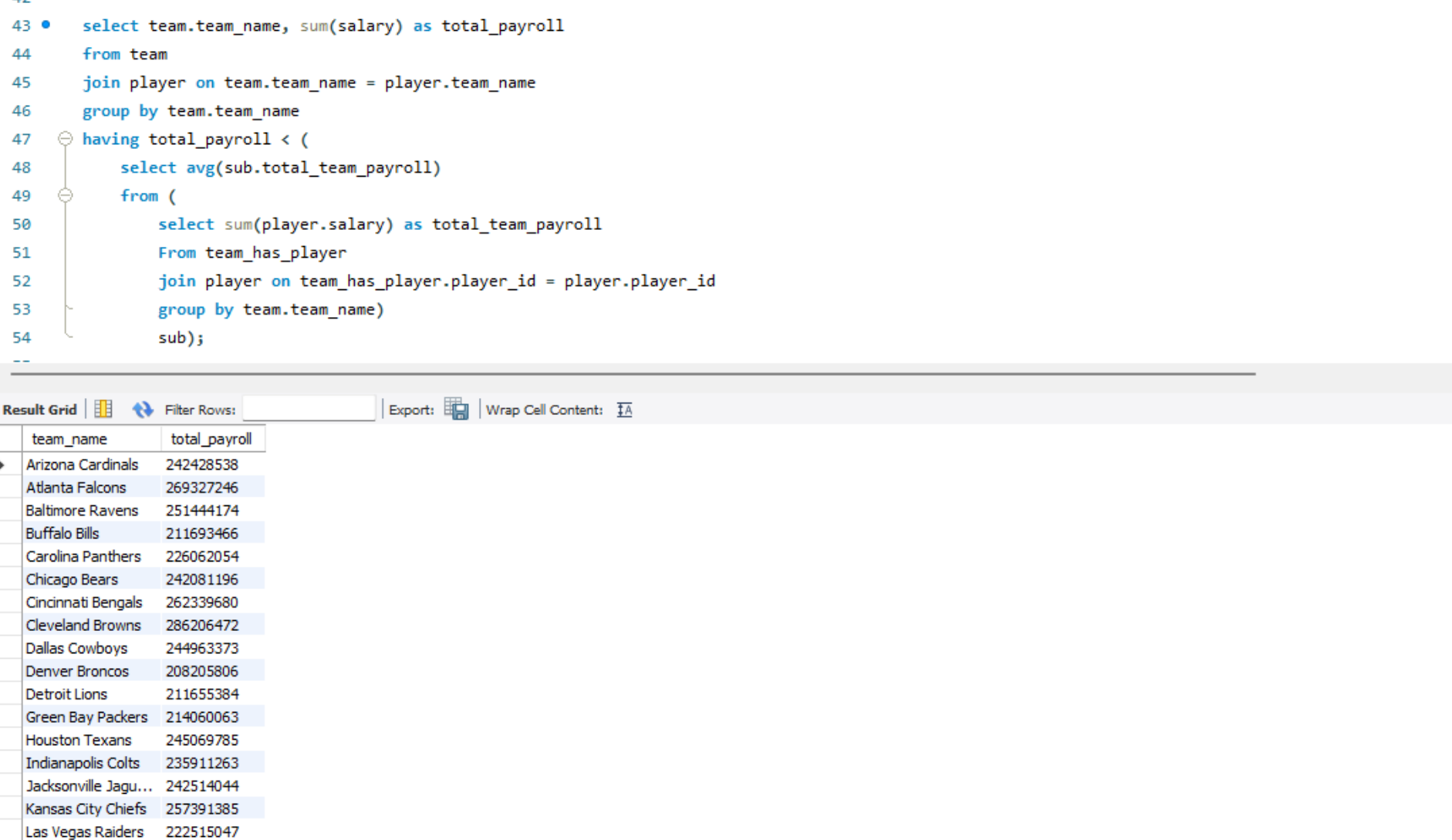
This query retrieves all players whose salaries are below the league wide average. It uses a subquery to calculate the average salary and compares each player’s pay against that benchmark. This information is meaningful because it helps identify undervalued or budget efficient players, which can assist teams and analysts in evaluating payroll distribution, player value, and potential opportunities for salary adjustments or contract negotiations.

5) Write a query that lists referee crews that officiated above-league-average scoring games.



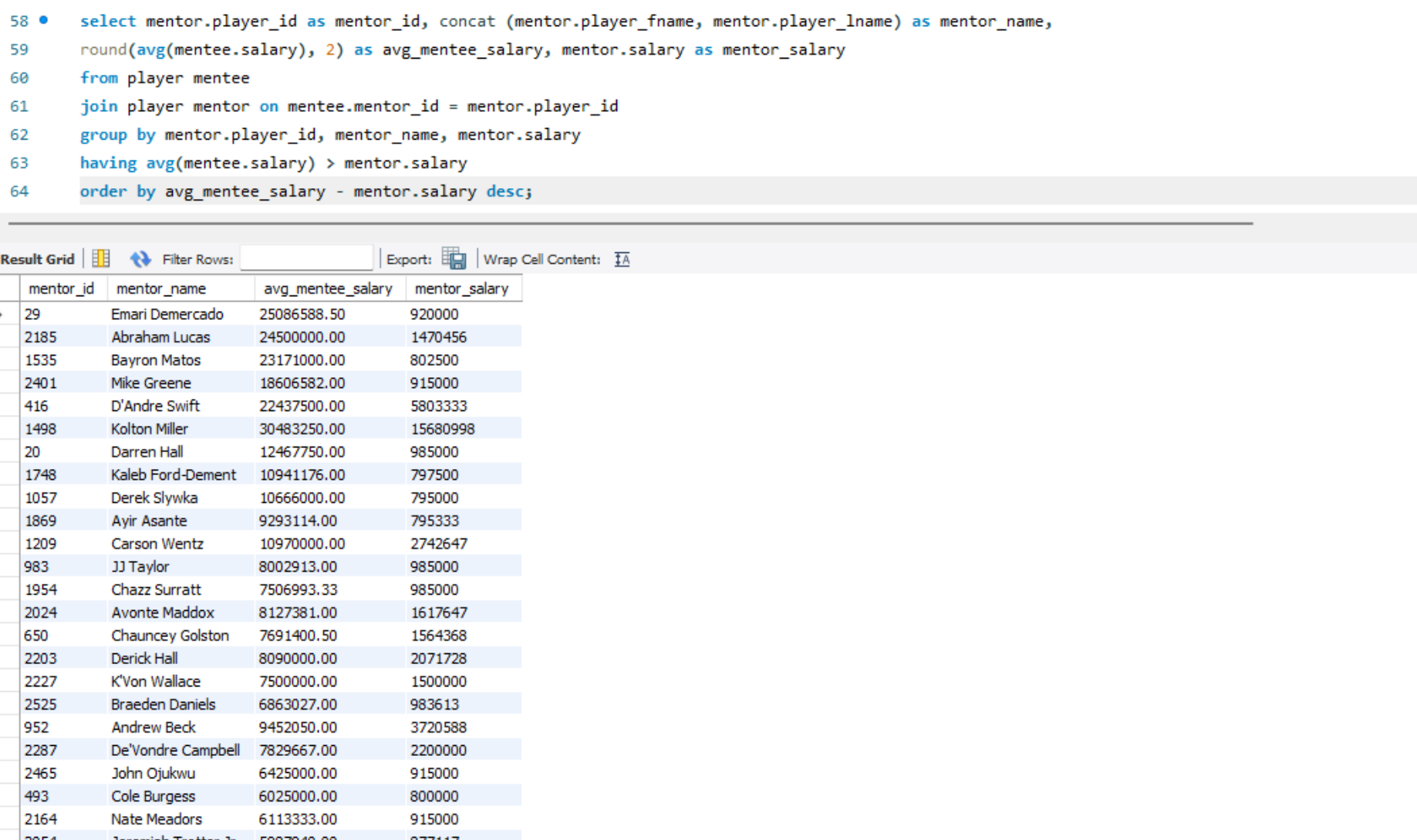
This query finds referee crews that have officiated games where the combined total points scored were higher than the league average. It joins the refereeCrew and game tables and compares each game’s total points to the league wide average calculated in a subquery. This is meaningful because it helps identify officiating crews associated with higher-scoring games, which can offer insights into game tempo, officiating tendencies, or styles that may influence scoring outcomes.

6) Write a query to find teams with below average payroll.



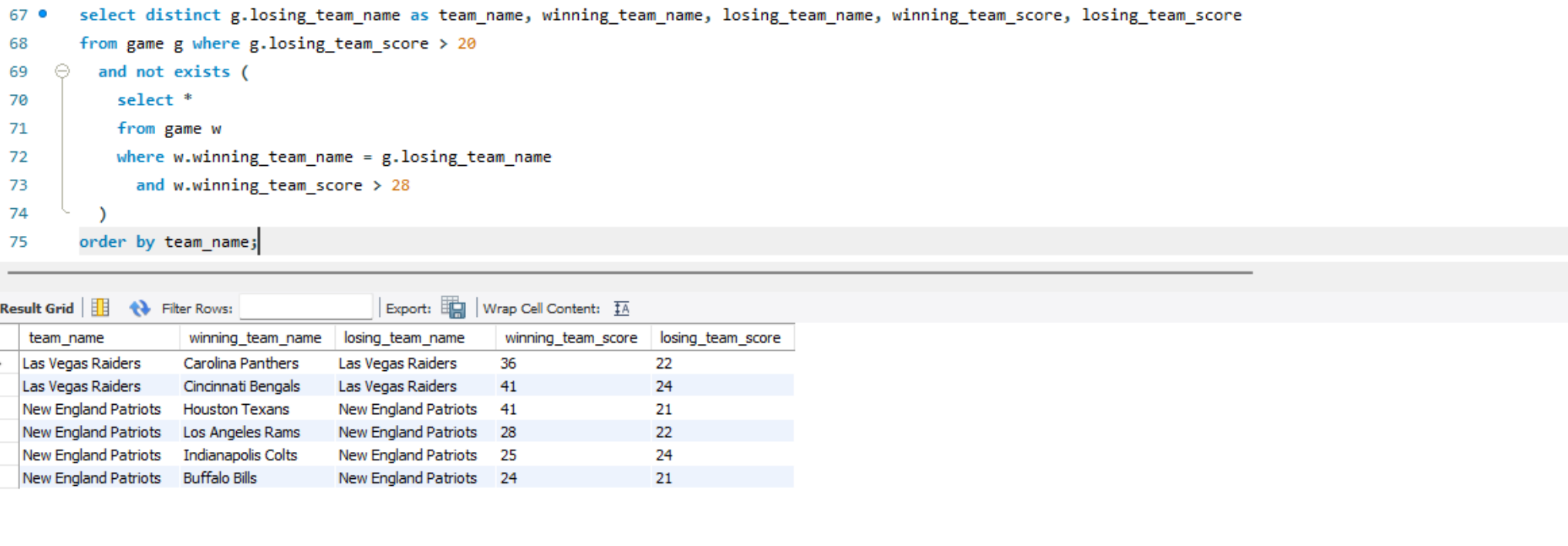
This query calculates each team’s total payroll by summing the salaries of active players, then compares those totals to the league-wide average payroll. It uses a subquery to determine the average across all teams. This is meaningful because it identifies organizations operating with lower payrolls, offering insight into financial efficiency, team spending strategies, and how resource allocation might correlate with overall team performance.

7) Write a query that lists mentors whose mentees (on average) out-earn the mentor.



This query identifies mentors whose mentees have a higher average salary than they do. It uses a self join on the player table to link each mentor with their mentees and compares the mentor’s salary to the average of their mentees’ salaries. This is meaningful because it highlights mentorship relationships where mentees have advanced beyond their mentors financially, offering insights into the effectiveness of player development, mentorship success, and potential leadership influence within teams.

8) Write a query to retrieve teams that lost even when scoring > 20, and where the winning team scored more than 28 points.



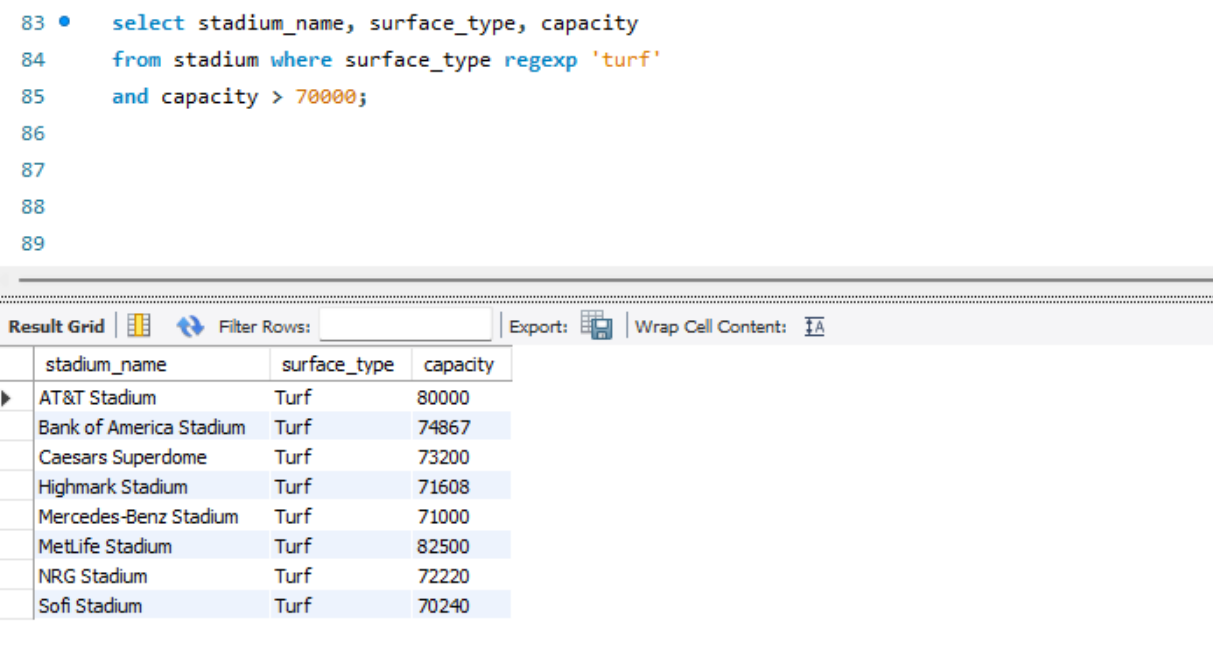
In games where the other team scored more than 28 points, this query returns teams that lost even though they scored more than 20 points. It focuses on high scoring losses, exposing teams with strong offenses but weak defenses. This is significant because it sheds light on game results and defensive flaws that may not be visible from win and loss records. This can assist analysts and coaches in identifying areas that require improvement.

9) Retrieve the names of parent categories that have subcategories.



Because it shows which coaches hold leadership positions within their organizations, this query is helpful. It gives information about the coaching hierarchy of the team and the division of duties by displaying which coaches have subordinates. Analysts and management can use this data to track mentorship relationships, comprehend organizational structure, and determine which coaches might have more management or influence over their teams.

10) Write a query to list the stadiums with a turf surface and capacity greater than 70,000.



This query retrieves all stadiums that have a turf playing surface and can hold more than 70,000 spectators. It filters data on surface type and seating capacity, then orders the results from largest to smallest. This is meaningful because it highlights the biggest turf based stadiums in the league, which can be useful for event planning, monitoring fan capacity, and identifying potential venues where the Superbowl and Pro Bowl may be played.

**Database Information**

Name of the database: cs\_jam37624