INFO 330

Tison, Kevin, Collin, Anshul

Our Dataset and Examples

- As the NBA finals are just getting started we thought it would be interesting to find and create a database around NBA playoff statistics
- This dataset displays information about player stats during the playoffs from 1950-2022.
- One of the key components of this dataset is player performance. It includes detailed information about each player, such as team names, position and the team they played for. It captures individual player performance in the playoffs, featuring essential metrics like points scored, rebounds, assists, steals, blocks, and shooting percentages
- An example is a column that shows the average minutes played as well as another column that shows the player age. We can use those two columns to find if there's a correlation between a player's age and time played in the playoffs.
- We can also do comparisons like the average player time in the playoffs in 1950 compared to 2022.

Datamodel: Entities and Relationships

- For our database we decided to make 5 tables to encompass our data.
- These 5 tables are offensive, defensive, season, games, and player.

The Offensive Table categorizes points, field goals, and points based off position.

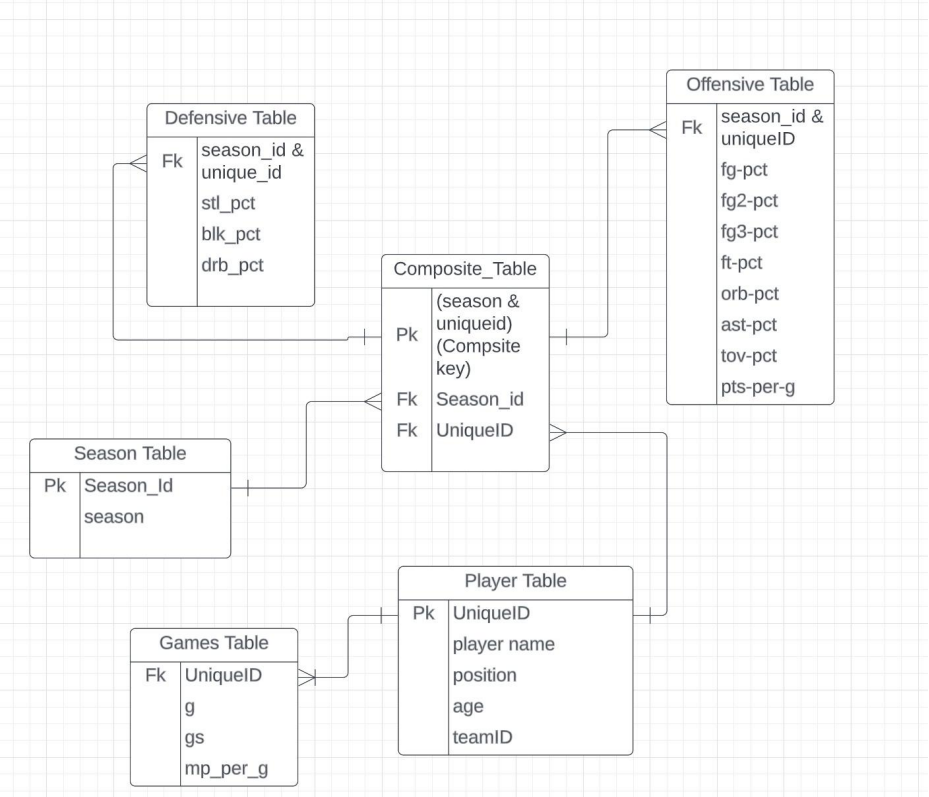
The Defensive Table identifies defensive statistics categorizing steals, blocks, and defensive rebounds

The Player Table helps to identify a unique ID for an individual's name, position, age, and team

The Games Table attributes all the games in the playoff series from 1950-2022, overall game stats, and minutes per each game (accounting for overtime)

The Season Table identifies a unique id for the year or season varying between 1950-2022

The Composite Table contains a composite key in order to map a season and name of player. It follows a one to many relationship because it allows access for offensive and defensive stats of an individual



Schema Designs, Patterns and Noticeable Characteristics

- Composite Key
  - Needed a composite key in order to uniquely map a specific players offensive and defensive stats. Each player has stats for each season and we need these two primary keys in order to find that information
- We started with 51 columns and dropped those that we didn't use in our queries because there was excessive data that would have made creating tables harder. Some examples of tables we dropped was information such as win share, which means one win is equivalent to three win shares. Another example column that we dropped was box plus minus, which means an "estimate of the points per 100 possessions that a player contributed above a league-average player, translated to an average team". We dropped these examples of tables because it didn't have value to us and our queries.
- This dataset had columns with pre-computed values like percentages. These were computed from other columns like fg_attempted and fg_made. We decided to only keep columns with percentages to do fewer calculations.

Queries Explained

- The queries we developed for this project encompassed a wide variety of aspects of players and team performances. We developed 8 unique queries that would help gain a better understanding of the NBA and how the game has changed through the years.

Anshul’s Queries

- **What was the average points scored per position?**

This inquiry examines player performance in the NBA playoffs from 1950 to 2022, focusing on scoring contributions from the five traditional basketball positions: Point Guard, Shooting Guard, Small Forward, Power Forward, and Center. The goal is to identify potential scoring trends across these positions and gain insights into strategic tendencies. This analysis could reveal whether specific positions, with their unique roles and responsibilities, tend to accumulate more points for their teams.

```sql
-- 1st and 2nd queries
SELECT player, MAX(pts_per_g) AS max_points
-- Retrieves players with the highest average points per game (PTS) in playoffs
FROM playoffStats
GROUP BY player
ORDER BY max_points DESC;

SELECT pos, AVG(fg_pct) AS avg_fg_percentage
-- Calculates average field goal percentage (FG%) for each position (pos) in playoffs
FROM playoffStats
GROUP BY pos;
```

- **Which player has had the best field goal percentage per minute played all time?**

This query aims to find the NBA playoff player with the highest field goal percentage per minute played. Field goal percentage is a key basketball statistic that measures a player's efficiency in converting shots attempted into successful shots made. By considering minutes played, this query evaluates a player's scoring effectiveness during their time on the court. It seeks to identify players who maximize their scoring opportunities within their limited playing time, going beyond total points scored.

Kevin's Query:

- ***What are the average minutes played per team in the playoffs per game***
  - •By executing this query, we can assess the average playing time of each NBA team during the playoffs. Analyzing this data provides valuable insights into a team's playoff experience and overall organizational knowledge. This information holds significance as it sheds light on how newer teams handle the demanding pressure situations and workload management during the playoffs. Recognizing these factors enables teams to strategize effectively, make informed decisions, and pursue success in their quest for championships.
- ***What players have played on the greatest number of teams in the playoffs?***
  - From this query we are able to analyze the players that have made it to the NBA playoffs that have been on the greatest number of teams, we are able to analyze player movement and team interactions during the postseason. By looking at this we can understand how players value trades and strategic decisions impact teams' success in the playoffs. This query provides valuable insight into player mobility, team dynamics and the influence of individual players in playoff games and the building of NBA superteams. For the future if we were to continue working with this dataset it may be interesting to include the amount of championships or conference title wins that this player has won as well in order to get more direct casual relationship from this.

```sql
--Finds the top 10 players that have been on the most amount of playoff teams
SELECT pt.player, COUNT(DISTINCT pt.teamID) AS team_count
FROM player_table AS pt
GROUP BY pt.player
HAVING team_count > 1
ORDER BY team_count DESC
LIMIT 10;

--Finds the average minutes per game in the playoff for each team
Select pt.team_id, AVG(gt.mp_per_g)
From game_table as gt
Join player_table as pt on pt.unique_id = gt.unique_id
Group by pt.team_id;
```

Collin's Queries:

Who are the top 10 players with the best defense based on their steal percentage?
The SQL query retrieves the top 10 players with the best defense based on their steal percentage from the player table. It selects the player name and steal percentage columns, sorts the results in descending order by steal percentage, and limits the output to only the top 15 players. By executing this query, we obtain a list of the top performers in terms of defense, specifically their names and steal percentages. This information is valuable for evaluating players' defensive capabilities and identifying those who excel at creating turnovers and disrupting opponents' offensive plays.

```
SELECT pt.player, dt.stl_pct
FROM defensive_table AS dt
JOIN composite_table AS ct ON ct.Comp_ID = dt.Comp_ID
JOIN player_table AS pt ON pt.UniqueID = ct.UniqueID
JOIN season_table AS st ON st.season_id = ct.season_id
WHERE dt.stl_pct IS NOT NULL AND dt.stl_pct <> '' AND st.season = 2022
ORDER BY dt.stl_pct DESC
LIMIT 15;
```

Calculate the average field goal 3 percentage (FG%) for each position (pos) in the playoffs:
This SQL query calculates the average field goal 3 percentage for each position in the NBA playoffs. It selects the position column and uses the Avg() function to compute the average FG%. The results are then grouped by position, offering valuable insights into the three-point shooting efficiency of each position during playoff matchups. This query provides beneficial information as it aims to validate whether the perception that a certain position in basketball is better at shooting 3s holds true. By examining the average FG% for each position, we can determine if this perception is based on actual statistical evidence or if it is merely influenced by star players who excel at shooting 3s.

```
SELECT pt.position, AVG(ot.fg3_pct) AS avg_fg3_pct
FROM offensive_table AS ot
JOIN composite_table AS ct ON ct.Comp_ID = ot.Comp_ID
JOIN player_table AS pt ON pt.UniqueID = ct.UniqueID
GROUP BY pt.position;
```

Tison's Queries

- ***Does age affect average minutes played?***

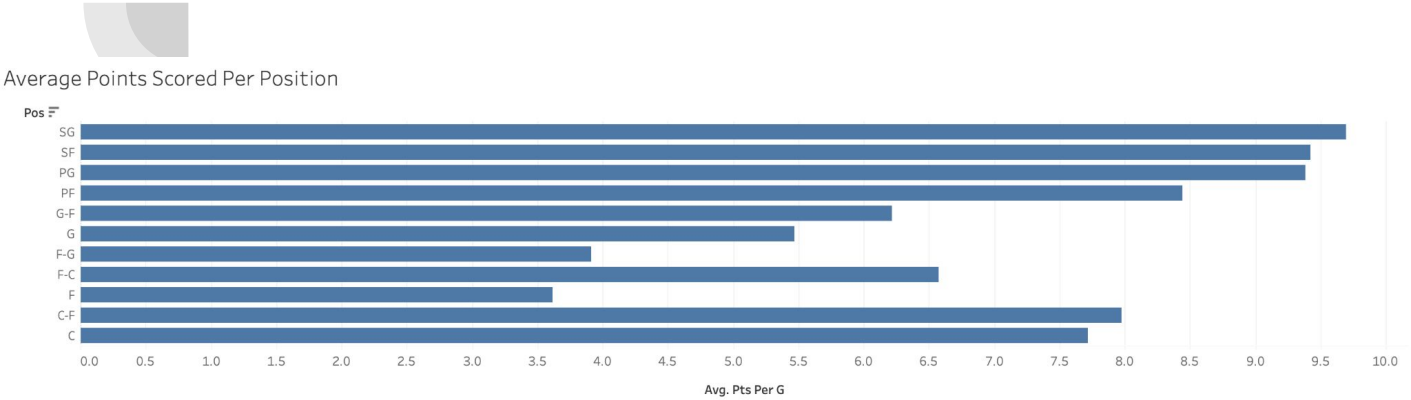
We wanted to see whether or not there was a correlation between a player's age and time off the bench. Age can also mean experience, which is a factor in if the coach wants to play them. For our query we selected the age and average minutes played from our playoff stat table. Then, we used order by which sorts in ascending order to see the average minutes played as the age gets older

- ***Retrieve the players with the highest average points per game (PTS) in the playoffs.***

For this query, we want to see players that scored the highest points on average per game in the playoffs. The database contains stats on players from 1950-2022. From there, we can take a look at who were the top 10 best scorers during the playoffs. The query selects the player and average points column, and then groups by player. Then, we ordered those average points in descending order to determine the 10 highest scorers in playoff history from 1950-2022.

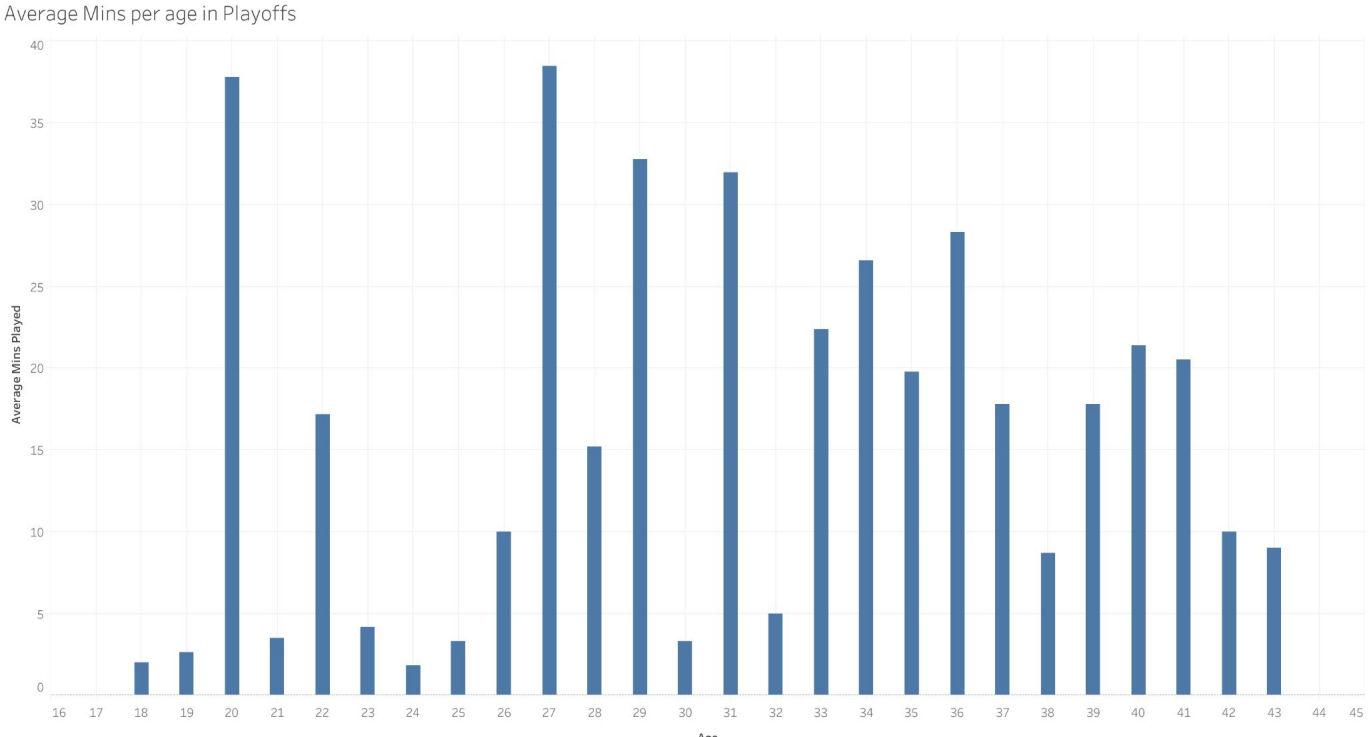
```sql
-- 3rd & 5th queries

SELECT age, mp_per_g AS average_mins_played
-- find the age and average points per game for players
FROM playoff_stat
GROUP BY age
ORDER BY age;
-- order by sorts the records in ascending order by default
-- so we can see the avg mins played as players get older

SELECT player, pts_per_g AS average_pts_per_game
FROM playoff_stat
GROUP BY player
ORDER BY average_pts_per_game DESC
-- limit 10 in descending order to restrict the output
-- to the top 10 players with the highest average points per game
LIMIT 10;
```

Anshul's Visualization



For this visualization we displayed the query that finds the average number of points scored per position in the playoffs. We can see through the descending order that Shooting Guards tend to score the most points in a playoff series game. This information can be applicable especially to recruiters for player evaluation, team building, and strategy development during the series. Understanding the averages of points per position can also enable those who are seeking talent to join teams as well. Overall this correlation visualized has numerous applications that are beneficial for recruiters and even sports fanatics.

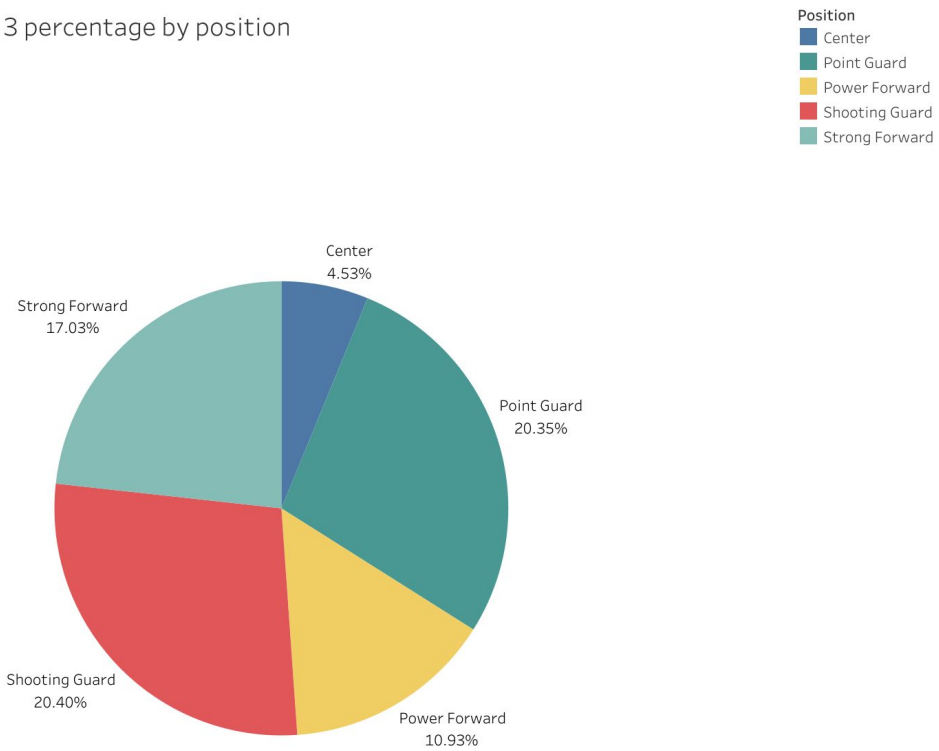
Tison's Visualization



For this visualization, we displayed the query that is trying to find the correlation between age and minutes played. We can see from the results that the average minutes were varied but had a higher average from age 27 to about 36. This can mean many things such as that is around the age of a players' prime, they have more experience which is why they have more play time than early 20s, as well the ages when a player is the healthiest. We can also pull interesting information from this bar graph like players in their 40s still getting playing.

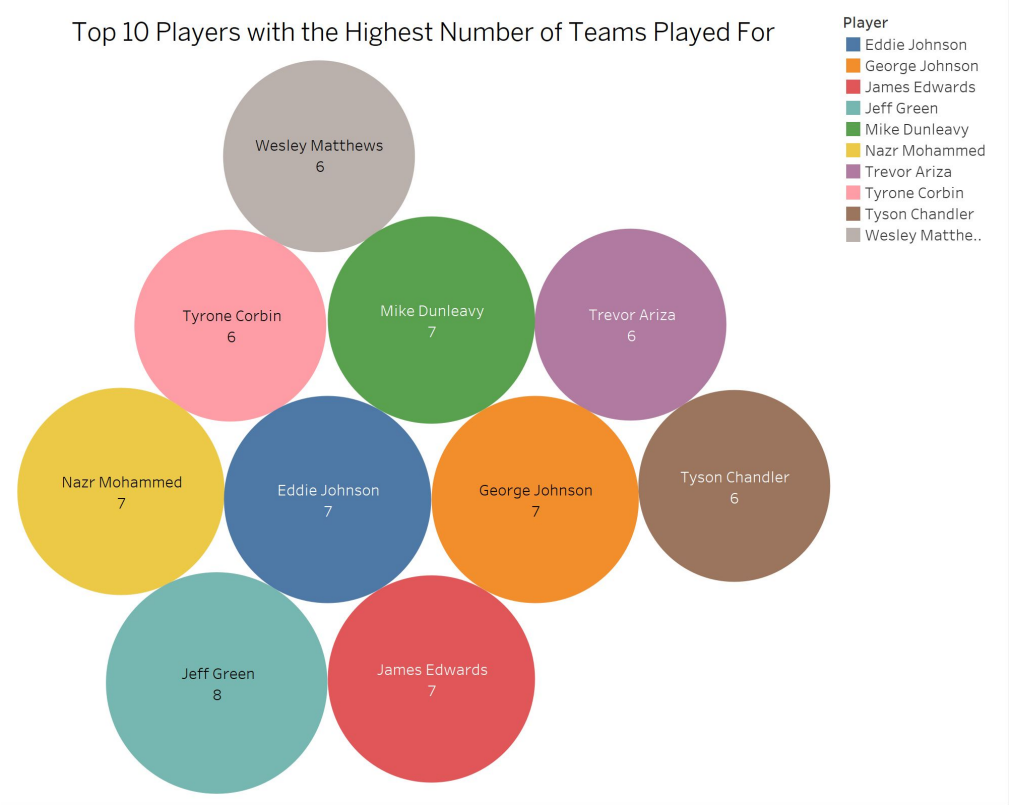
Collin’s Visualization

Average Field Goal 3 percentage by position



Our findings confirm that the two guard positions, widely considered the best shooting positions, indeed exhibit the highest field goal 3 percentage in the NBA playoffs. This visualization presents the average FG% for each position, shedding light on the three-point shooting efficiency during playoff matchups. By validating the perception regarding shooting abilities across positions, it provides valuable insights into whether statistical evidence supports the belief that certain positions excel at shooting three-pointers. This analysis also considers the influence of star players who may contribute to this perception.

Kevin’s Visualization



From this visualization we are able to see the top 10 NBA players that have played on the most amount of playoff qualifying teams. We can see that the player with the most amount of different teams played for in the playoffs is Jeff Green with a total of 8 teams. If we take a look into NBA analytics from the past years we can maybe gain a better insight on which of these players had a positive impact on their team, or maybe even looking into reasons why they did move to that many yrs,d. Some possible reasons could include “ring chasing” or potentially player chemistry with the team.