

Summary of Dataset

For our project, we have chosen to use data from the 2024-2025 NBA season. The dataset includes game-by-game stats on every player, team data, player data, draft data, combine data, schedule data, and coach data. The complete dataset is drawn from eight different files with a total of 36092 rows and 151 columns. The size of the dataset is a great fit for this project as it is large and diverse enough to be broken down into 10 or more tables, while still being manageable and complex enough to make some interesting queries. In the later stages we will trim attributes that aren't relevant to our needs.

As mentioned, there are 8 different files that this database will be taking information from. The `common_player_info.csv` file contains 4,171 rows with 33 attributes describing player details, including `person_id`, `height`, `weight`, `birthdate`, `school`, and `position`. The `draft_combine_stats.csv` file has 1,202 rows and 47 attributes of player measurements and combine stats from the NBA Draft Combine, such as `player_id`, `season`, `standing_vertical_leap`, `max_vertical_leap`, and `bench_press`. The `draft_history.csv` file has draft outcomes for players from 1947–2023, with 7,990 rows and 14 attributes, including `person_id`, `season`, `round_number`, `round_pick`, and `overall_pick`. The `player.csv` file holds a simpler set of player records from the 1947–2023, that contains 4,831 rows with 5 attributes, including `id`, `first_name`, `last_name`, `full_name`, and `is_active`. The `team.csv` file contains 30 rows representing all NBA teams, with 7 attributes such as `id`, `full_name`, `abbreviation`, `city`, `state`, and `year_founded`. These csv files were taken from <https://www.kaggle.com/datasets/wyattowalsh/basketball>.

The `database_24_25.csv` contains detailed game-by-game stats for every player in the 2024-2025 season, with 16512 rows, and 25 attributes. Some attributes include, `Player`, `Tm` (team), `Opp` (opponent), `Res` (result), `MP` (minutes played), `FG` field goals made, `FGA` field goal attempts, `ORB` offensive rebounds, and many other performance metrics. This csv file was taken from <https://www.kaggle.com/datasets/eduardopalmieri/nba-player-stats-season-2425>.

The `coaches_24_25.csv` contains 35 rows and 22 attributes that summarize the 2024-2025 coaches. Attributes include `Coach` (name), `Tm` (team), `Seasons_with_franchise`, `Seasons_overall`, and attributes about coaching performance in the regular season and playoffs, including games coached, games won, and games lost. This csv file was taken from https://www.basketball-reference.com/leagues/NBA_2025_coaches.html.

Lastly, the `NBA_24_25_schedule.csv` file records the season schedule with 1321 rows and 12 attributes, such as `date`, `Start` (time start ET), `Vistor`, `Home`, `Away_PTS`, `Home_PTS`, and other attributes such as attendance and arena name. This csv file was taken from https://www.basketball-reference.com/leagues/NBA_2025_games.html. The pandas library was used to read each month's schedule into one csv file.