

Project Title: Predicting NBA Team Wins Using Past 5 Seasons and Current Season Statistics

Team Members: Wyatt Wetzler, Leonel Marines, Rhea , Darien Hernandez

Topic & Motivation

Our project explores how NBA team performance metrics from the past five seasons—combined with early data from the current season—can be used to predict overall team success. This topic highlights how data analytics translates into meaningful insights within sports, revealing which performance factors most strongly influence win totals. By connecting statistical trends to real-world outcomes, we aim to identify patterns that can help analysts, coaches, and fans evaluate teams more objectively. We chose this topic because of our shared passion for sports analytics and the abundance of high-quality NBA data available online.

DATASET DESCRIPTION

Source & Format:

The dataset will be merged from two datasets. One from Kaggle in CSV format, containing team-level statistics for the last five NBA seasons (2019–2024) and another with partial data for the 2025–26 season.

Key Variables / Dimensions:

Team Name, Season, Wins (target), Points Per Game (PPG), Opponent PPG, Field Goal %, 3-Point %, Free Throw %, Offensive and Defensive Ratings (ORtg, DRtg), Rebounds (REB), Turnovers (TOV), and Point Differential (Pts For – Pts Against).

Fit for the Project:

These quantitative variables provide strong predictors for team success, enabling statistical modeling and performance comparison. If multiple tables are used, we will merge team statistics with win records using Team and Season as common keys to build a unified dataset.

PLAN FOR ANALYSIS & VISUALIZATION

The analysis will focus on predicting NBA team wins using a combination of supervised and unsupervised learning methods. Data will be cleaned, merged, and explored using Pandas, NumPy, and Matplotlib/Seaborn.

1. Exploratory Data Analysis (EDA):

- Summarize key variables and visualize relationships using heatmaps, scatterplots, and correlation matrices.

- Identify which metrics (e.g., ORtg, DRtg, PPG, TOV) have the strongest correlation with win totals.

2. Modeling Approach:

- Multiple Linear Regression → baseline prediction of total wins.
- K-Means Clustering → group teams by similar performance patterns (offense-heavy, defense-heavy, balanced).
- Decision Tree Regression → capture nonlinear relationships and rank the most influential statistics.

3. Model Evaluation:

- Assess model accuracy using R^2 , Mean Absolute Error (MAE), and feature importance plots.
- Compare results across methods to determine which provides the best predictive accuracy for the 2025–26 season.