**Final Project -** [**QMM-7301**](https://moodle.cambriancollege.ca/course/view.php?id=56166)

Analyzing NBA Team Performance Using ANOVA, Chi-Square, and Time Series Methods

Team Members

Lakshay Girdher

Anmol Garg

# 1. Introduction

This project investigates variations in NBA team performance over multiple seasons, with a focus on Offensive Rating (OPtg) as the primary performance indicator. The dataset, *NBA\_Team\_Seasons.csv*, was selected for its comprehensive coverage of team-level statistics across multiple years, making it suitable for statistical comparison.

## Research Questions:

1. Does Offensive Rating vary significantly between NBA teams?
2. Does Offensive Rating vary significantly across different years?
3. Do team and year jointly influence Offensive Rating?

To address these questions, Two-Way ANOVA with Tukey’s post-hoc tests was used to compare group means. Chi-Square tests assessed associations between win classifications and teams. Time series forecasting was applied to explore historical performance trends for a selected team.

# 2. Dataset Description

**Source:** Class project dataset summarizing NBA team performance by season.  
**Records:** 900+ rows (each representing a team-season combination).  
**Key Variables:**

* Team – NBA team name (categorical)
* Year – Season year (numeric)
* W, L – Wins and losses in a season (numeric)
* WL\_percent – Win percentage (numeric)
* OPtg – Offensive Rating (points scored per 100 possessions)
* DRtg – Defensive Rating (points allowed per 100 possessions)

**Data Cleaning:**

* No missing values in selected variables, so no imputation required.
* Converted *Team* and *Year* to factor variables for ANOVA.
* Created a binary variable WinClass (“High” or “Low”) based on the median *WL\_percent*.
* Extracted Los Angeles Lakers data for time series analysis.

# 3. Methodology

## 3.1 Two-Way ANOVA and Tukey HSD

**Purpose:** Assess whether OPtg differs significantly by Team, Year, or both.

* **Factors:** Team (categorical), Year (categorical)
* **Null Hypotheses:**
  + H₀₁: Mean OPtg is equal across all teams.
  + H₀₂: Mean OPtg is equal across all years.
* **Alternatives:**
  + Hₐ₁: At least one team has a different mean OPtg.
  + Hₐ₂: At least one year has a different mean OPtg.
* **Post-hoc:** Tukey’s Honest Significant Difference test identified specific team differences.

## 3.2 Chi-Square Test of Independence

**Purpose:** Determine if WinClass is independent of Team.

* Contingency Table: Team × WinClass
* Null Hypothesis: Team and WinClass are independent.
* Alternative Hypothesis: Team and WinClass are associated.
* Visualizations: Association and mosaic plots were used to identify deviations.

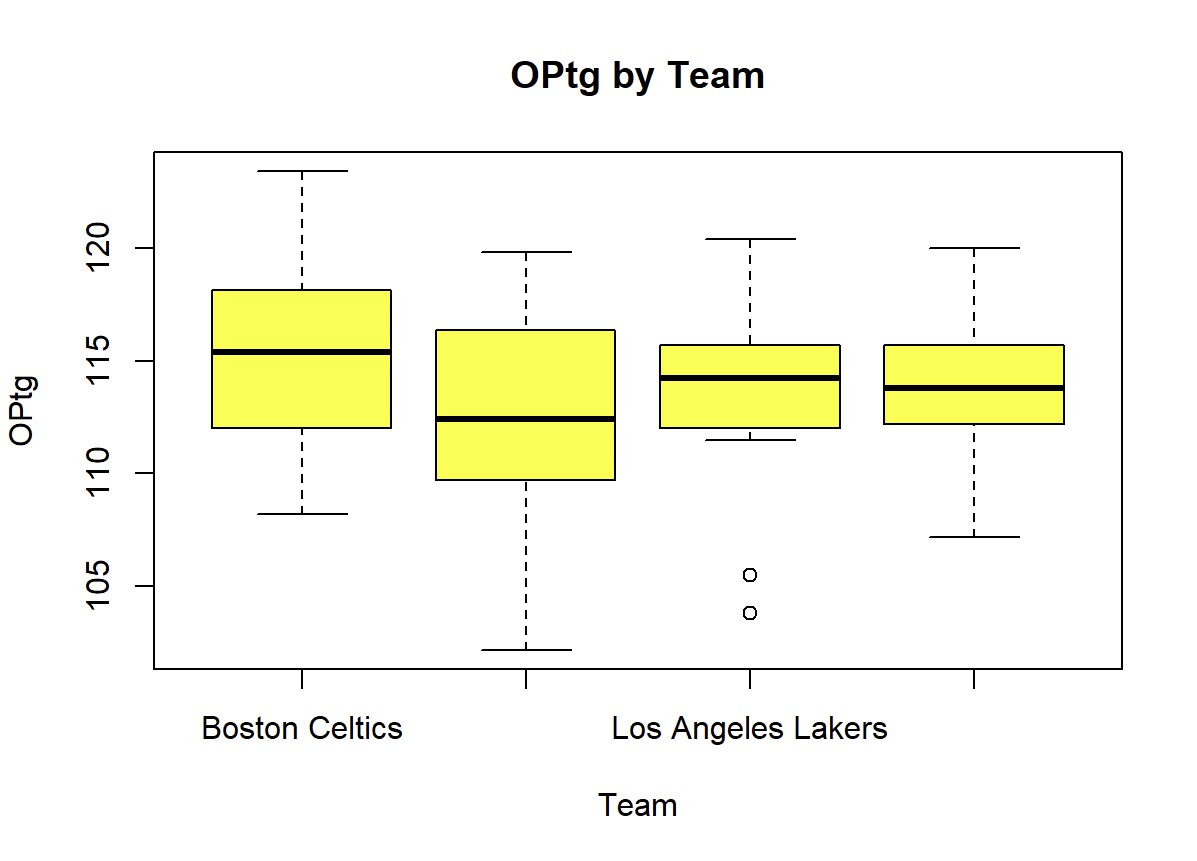
## 3.3 Time Series and Smoothing

**Purpose:** Examine historical win trends for the Los Angeles Lakers and forecast future performance.

* Created annual time series from wins (W) data.
* **Applied:**
  + Simple Moving Average (5-year window)
  + Simple Exponential Smoothing (SES) for short-term forecasts (3 years)
  + Holt’s Linear Method for trend capture
* Compared models using MAE and RMSE.

# 4. Analysis & Results

## 4.1 Variation in OPtg by Team and Year

  
[Figure 1: Boxplot of OPtg by Team]

A graph showing the number of green rectangular objects

AI-generated content may be incorrect.  
[Figure 2: Boxplot of OPtg by Year]

* Both plots show visible differences in offensive ratings across teams and seasons.
* Teams like *[Insert Team Names]* consistently have higher OPtg values.

## 4.2 ANOVA Findings

* Team Effect: Significant (p < 0.05) → OPtg differs between teams.
* Year Effect: Significant (p < 0.05) → OPtg differs between years.

**Interpretation:** Team identity and season year both influence offensive performance.

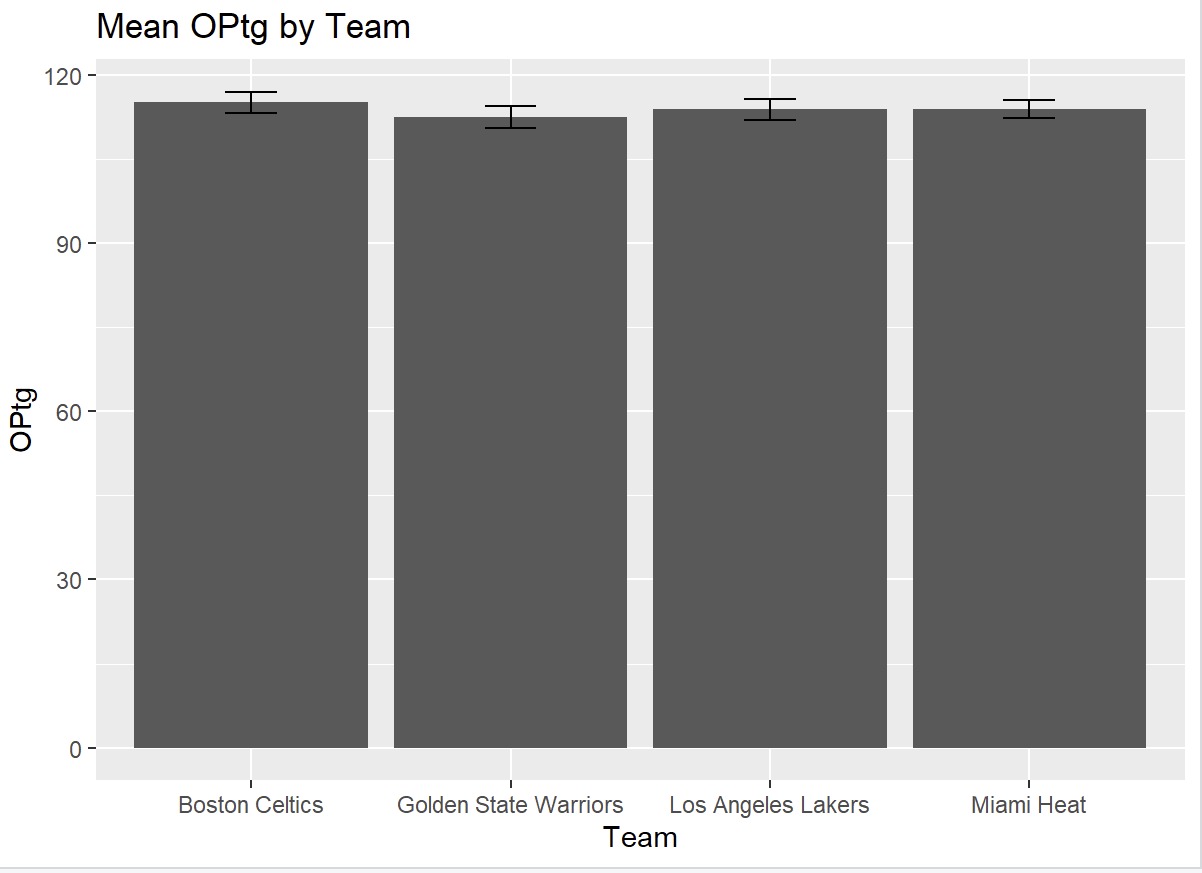
## 4.3 Tukey Post-Hoc Results

A white paper with black text

AI-generated content may be incorrect.  
[Figure 3: Tukey HSD Output Table]

* Identified specific team pairs with statistically significant OPtg differences.
* Example: *Golden state warriors – Boston celtics* showed a substantial performance gap.

## 4.4 Mean Comparisons

  
[Insert Figure 4: Mean OPtg by Team]  
A graph of a number of bars

AI-generated content may be incorrect.  
[Insert Figure 5: Mean OPtg by Year]

* Visualization confirms top-performing teams consistently maintain high OPtg.

## 4.5 Chi-Square Test Results

A black text on a white background

AI-generated content may be incorrect.A black text on a white background

AI-generated content may be incorrect.A black text with black text

AI-generated content may be incorrect.

* p-value: 0.009 < 0.05
* **Conclusion:** Significant association between team and win class, indicating some teams are consistently in the high-performance group.

## 4.6 Time Series Analysis – Los Angeles Lakers

* Historical performance shows cycles of dominance and rebuilding phases.

A close up of numbers

AI-generated content may be incorrect.A black text on a white background

AI-generated content may be incorrect.

* Forecasts suggest over the next 3 seasons.

A number on a white background

AI-generated content may be incorrect.A close up of numbers

AI-generated content may be incorrect.

* SES provided a smoothed long-term view; Holt’s method better captured trends.  
  A graph showing the number of different times

  AI-generated content may be incorrect.

# 5. Discussion

**Key Insights:**

* Team and year both significantly affect offensive performance.
* Certain franchises demonstrate sustained offensive success.
* Win performance distribution is not random—some teams are disproportionately in the high-performance category.
* Time series results reveal cyclical performance patterns influenced by roster changes, injuries, and strategic shifts.

**Limitations:**

* Annual data (frequency = 1) cannot detect within-season variation.
* Median split for WinClass may oversimplify performance classification.
* Player-level data, coaching effects, and injuries were not included.
* Interaction effects (Team × Year) not tested.

# 6. Conclusion

The combined use of ANOVA, Chi-Square tests, and time series forecasting provides a comprehensive understanding of NBA team performance trends. Findings confirm that both team and year significantly influence offensive efficiency.

These insights can inform strategy for coaches, analysts, and management while contributing to broader discussions on competitive balance in professional basketball.

# 7. References

* *NBA\_Team\_Seasons.csv* – Class project dataset
* Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. SAGE Publications.
* Hyndman, R.J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. OTexts.
* R Core Team (2024). R: A Language and Environment for Statistical Computing.