**README: NBA Salary vs. Performance Modeling**

**Description**This project investigates how NBA player performance metrics — especially 3-point shooting — relate to player salaries. With the league's shift toward a pace-and-space offense and growing dependence on long-range shooting, the goal is to explore which shooting metrics influence salaries the most and whether player roles or positions affect this relationship.

Inspired by an earlier undergraduate research paper written by Maekala Turner for MATH 370 (Writing in Mathematics) at the University of Massachusetts Amherst, this work revisits the question using more advanced methods and recent data.

**Setup Instructions**

1. **Setup Environment:** Use RStudio or a Jupyter environment with the IRKernel.
2. **Install Required Libraries:**
   * tidyverse
   * readr, janitor, lubridate
   * broom, ggplot2, plotly
   * tidymodels, ranger, GGally, corrr
3. **Google Drive Setup:**
   * Mount your Google Drive if using Colab.
   * Set project directory: "/content/drive/My Drive/MTurner Portfolio/Salary vs. Performance Model"
   * Ensure all folders are existing: data files/, code & notebooks/, visuals & reports/

**Key Modules/Parts**

* **Data Import & Cleaning**: Load and clean salary and performance data, standardize names, and handle missing values.
* **EDA (Exploratory Data Analysis)**: Understand trends through scatter plots, boxplots, and correlation matrices.
* **Modeling**: Run linear and random forest regressions to estimate salary based on performance features.
* **Visualizations**: Output static or interactive plots to the visuals & reports/ folder.

**How to Use / Example Flow**

| # 1. Load raw salary + stats files salary\_df <- read\_csv("data files/nba\_salaries.csv") stats\_df <- read\_csv("data files/player\_stats\_2023.csv")  # 2. Clean + merge model\_df <- left\_join(clean\_names(salary\_df), clean\_names(stats\_df), by = c("player", "season"))  # 3. Run linear model lm\_model <- lm(salary ~ fg3\_pct + fg3a + efg\_pct + mp + ts\_pct, data = model\_df) summary(lm\_model)  # 4. Save visuals ggplot(model\_df, aes(x = fg3\_pct, y = salary)) +  geom\_point() + geom\_smooth(method = "lm") +  ggsave("visuals & reports/salary\_vs\_3pt.png") |
| --- |