BrainStation

Prepared by Anastassia Troitskaia

September 16 2024

NBA Salary Prediction & Analysis

Deep diving into our data with clear direction

Content

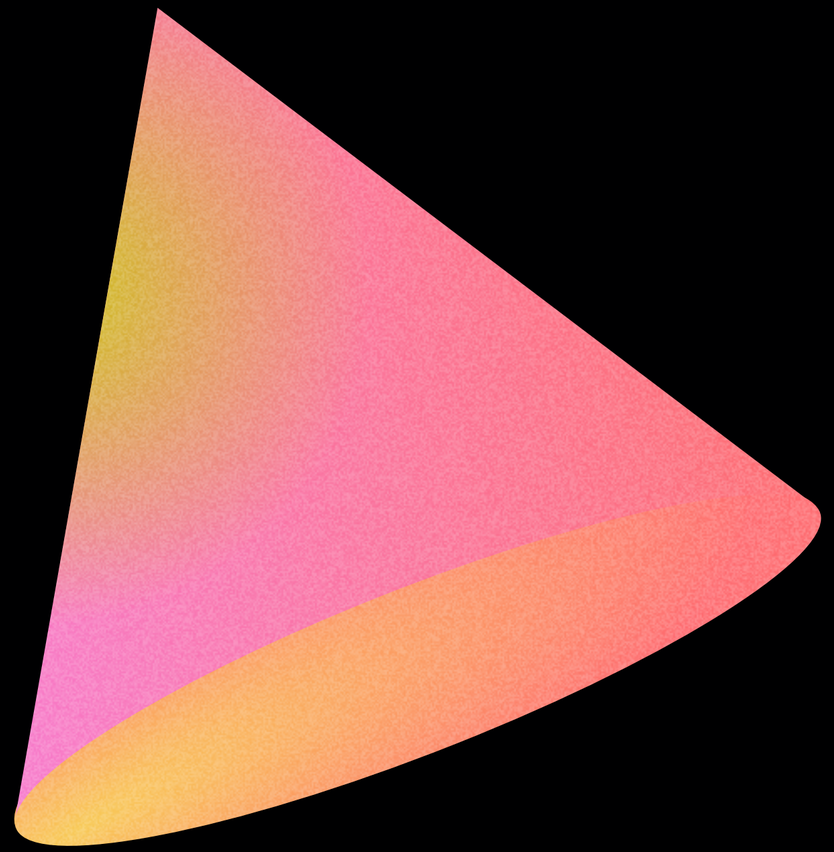
Introduction

Purpose:

To predict NBA salaries for college basketball players based on their performance statistics and historical data, aiding NBA teams in draft and salary decisions.

Description:

This project tackles the challenge of translating college performance into professional success, using machine learning and comprehensive data from current and former NBA players to improve player evaluation and financial planning.



Back to Content Page

Recap

Previous Issues

- Missing Values in many different categories
- Data has a lot of issues with data types
- Many values were entered as string object and etc.

How was it handled?

- Lot’s of data cleaning had to be performed
- Missing values imputation
  - KNN imputation
- Convert all the data types from string object
- Data such as Draft Team, College Name had to be encoded

Back to Content Page

Game Statistics

- PTS (Points): Key offensive statistic that often correlates with a player's value and potential salary.
- G (Games Played): This indicates durability and consistency, which are valuable traits for NBA teams.
- MP (Minutes Played): Shows how much a player is trusted on the court, which can be indicative of their skill level and potential value.
- FG% (Field Goal Percentage), 3P% (Three-Point Percentage), FT% (Free Throw Percentage): These shooting percentages are crucial indicators of a player's offensive efficiency.
- TRB (Total Rebounds): Rebounding is a valuable skill that contributes to both offense and defense.
- AST (Assists): Important for guards and indicates playmaking ability.
- STL (Steals) and BLK (Blocks): These defensive stats can significantly impact a player's value.
- Draft Year: This could potentially capture trends in salary over time or how recent draft picks are valued.

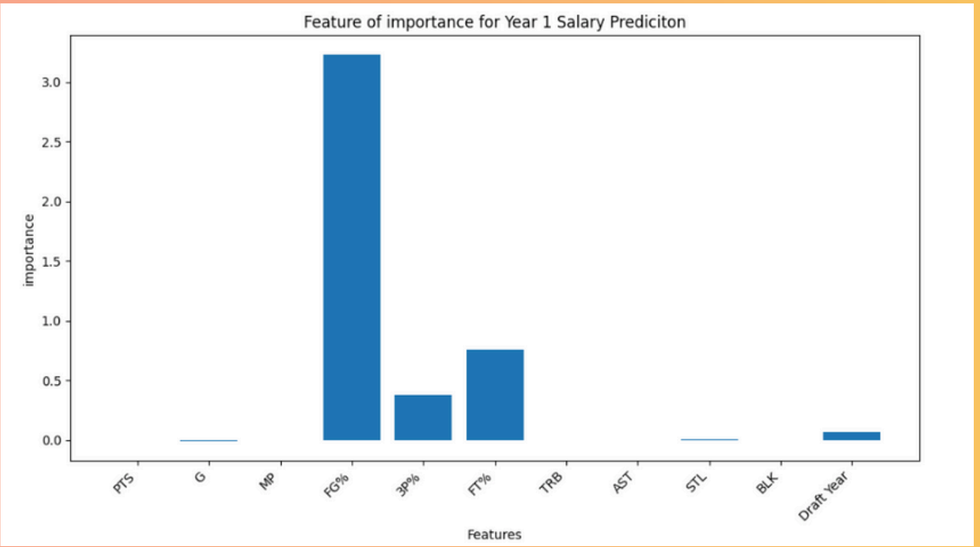
Player Physical Statistics

- Height: Important for positioning, especially for centers and forwards.
- Wingspan: Indicates defensive reach and ability to block shots or intercept passes.
- Weight: Shows a player's bulk, which can affect physicality in the game.
- Vertical Leap: Measures jumping ability, critical for rebounds and scoring near the basket.
- Agility and Speed: These metrics are tested in combine drills like the lane agility drill and three-quarter sprint.
- Standing Reach: Measures how high a player can reach without jumping, important for defense and rebounding.

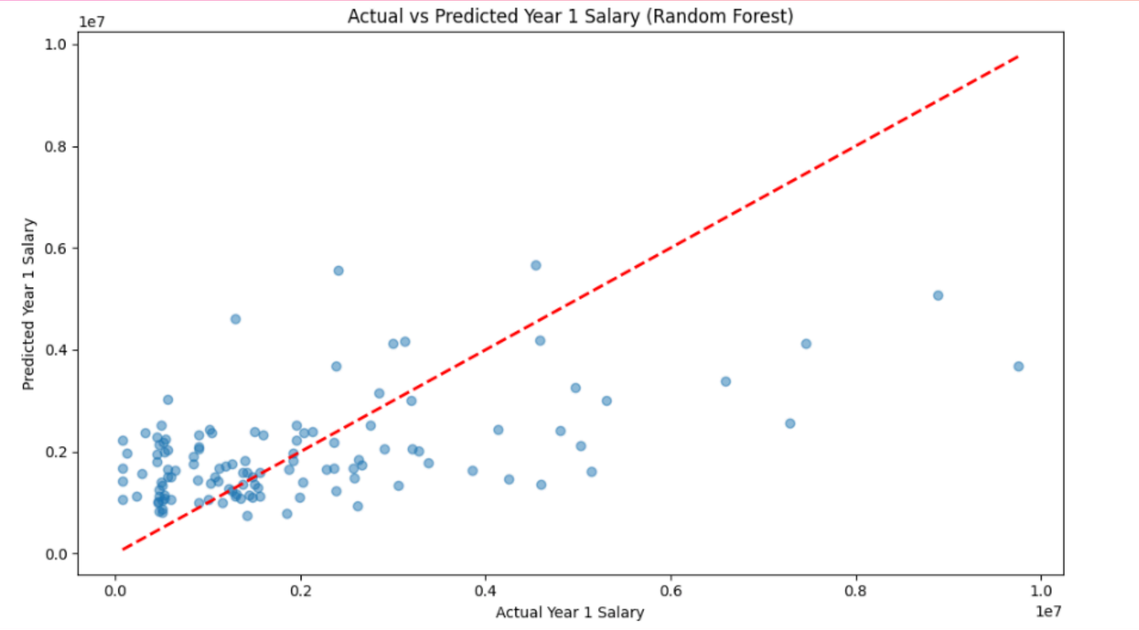
Linear Regression Model

- **RMSE ~ *$1,577,872* -> average prediction error**
- **R-squared ~ 0.2417 or 24.17%**
- **Very poor predictive power**

Features Of Importance



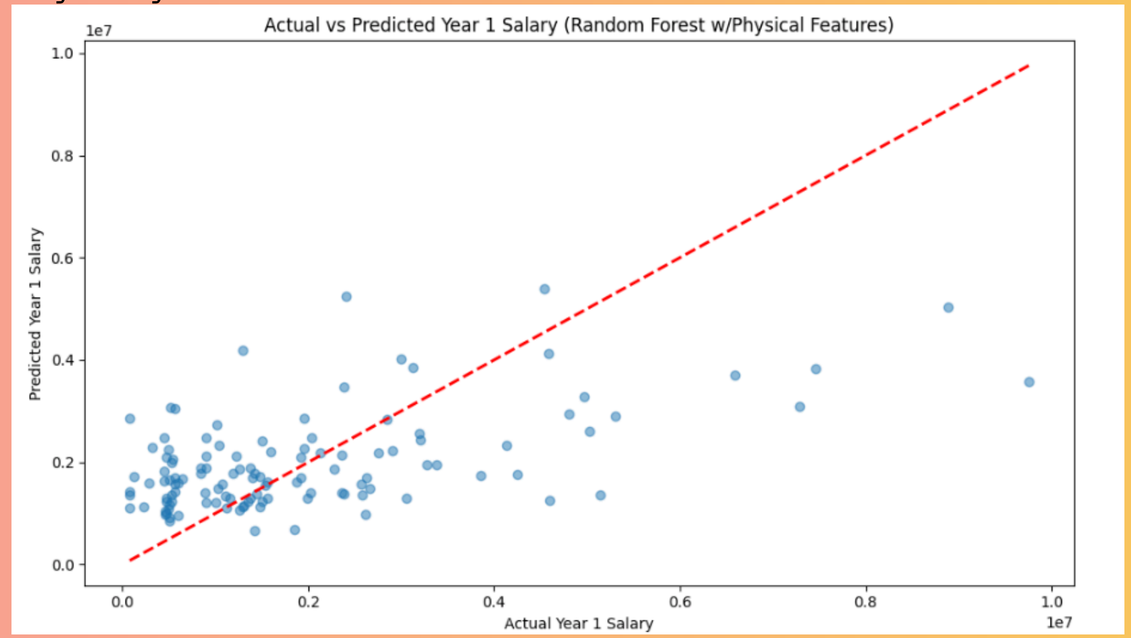
Random Forest Model



- **The RMSE has slightly improved from *$1,577,871* to *$1,526,530*, a reduction of about 3.25%.**
- **The R-squared has increased from 0.2417 (~24%) to 0.2902 (~29%) , an improvement of about 20%**

Random Forest Model with additional features

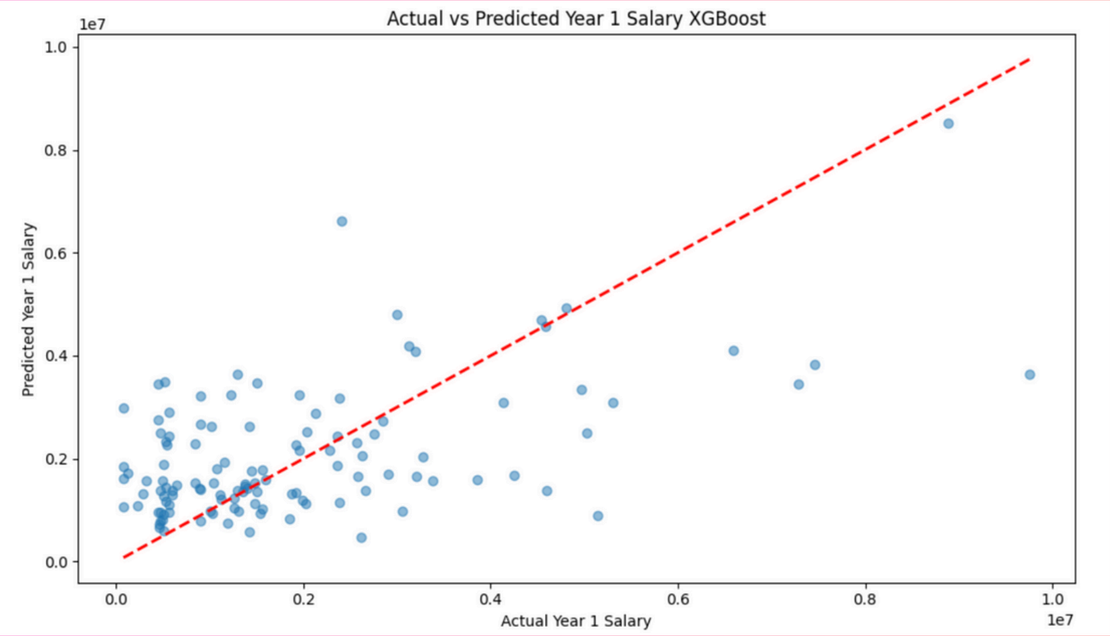
Player Physical Statistics



- **RMSE: *$1,507,924***
- **R-squared: 0.3074, which is ~ 30.74%**

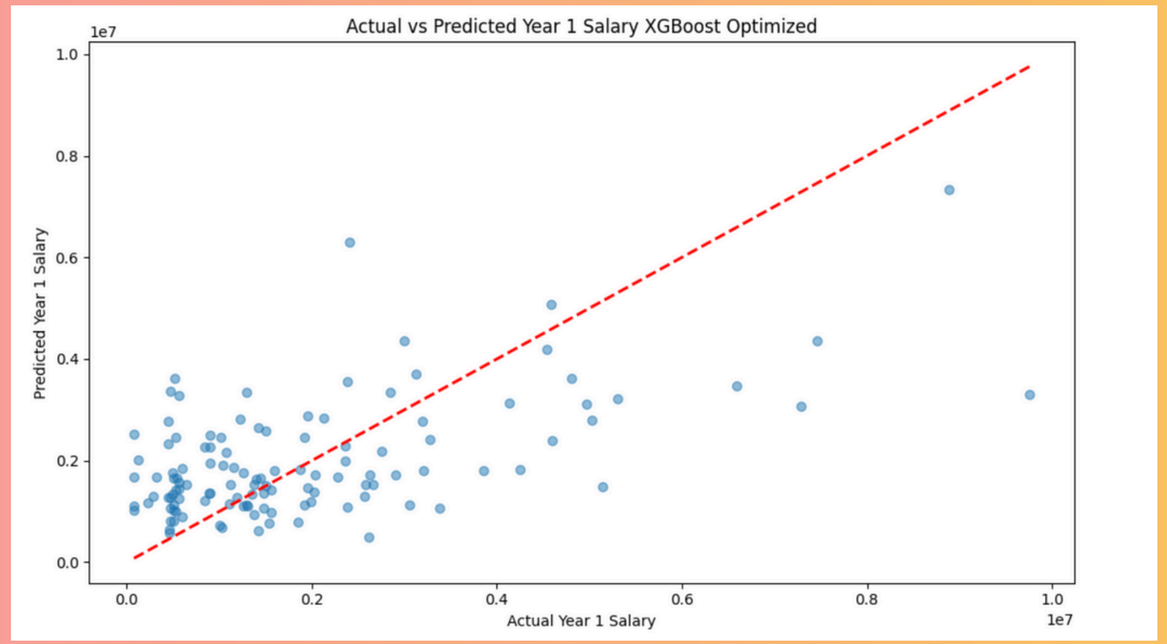
XGBoost Model

- **Root Mean Squared Error: *$1,543,885***
- **R-squared: 0.27**

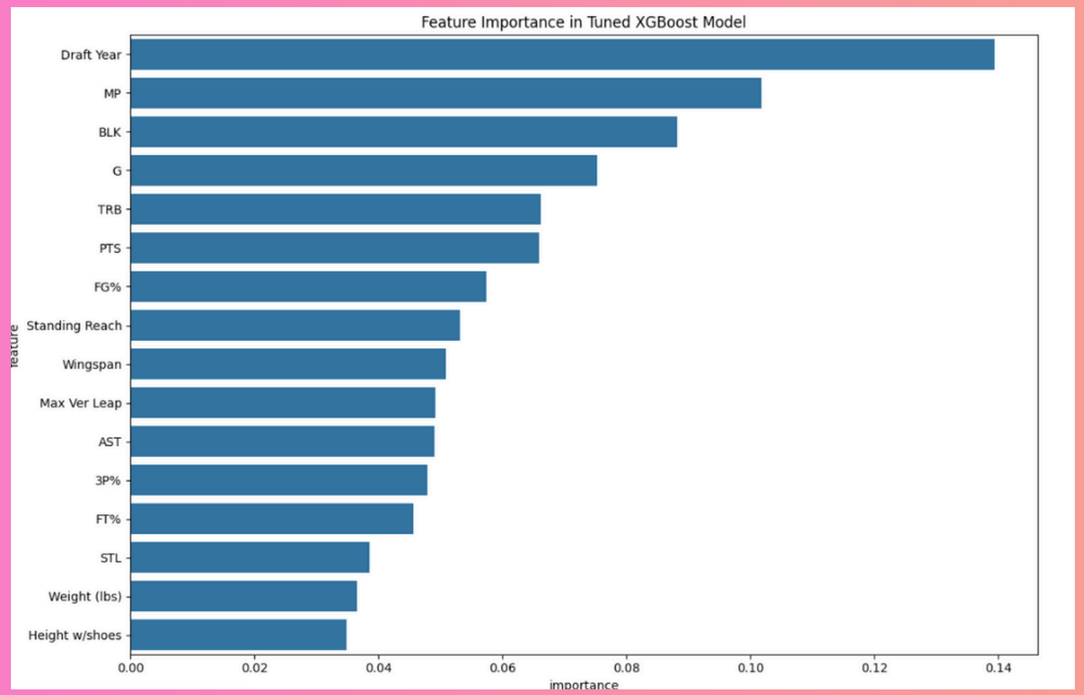


XGBoost Optimized

- **Tuned XGBoost - RMSE: *$1,501,051 imporved from random forest model : $1,507,924***
- **Tuned XGBoost - R-squared: 0.31, about 31% from R-squared of Random forest ~ 30.74%**



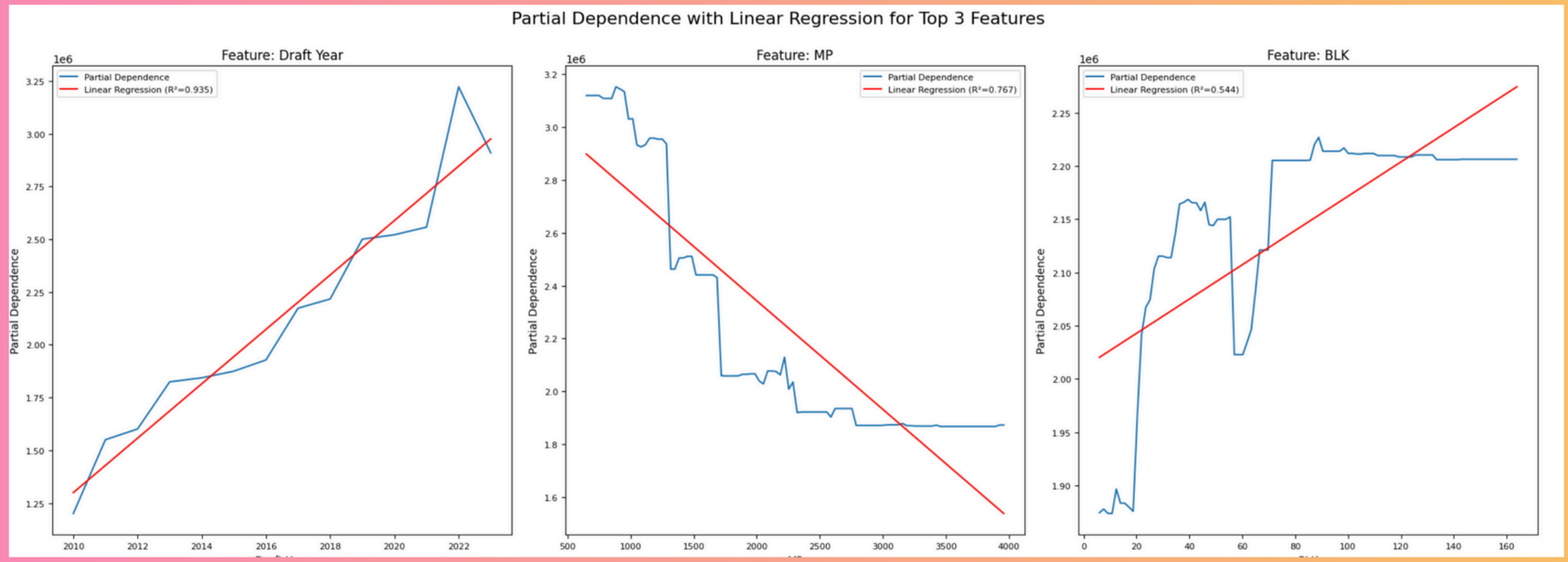
XGBoost Optimized



Feature Importance

- **Draft Year**
- **MP (Minutes played): Career college**
- **BLK (Blocks): These defensive stats can significantly impact a player's value.**
- **G Games**
- **TRB (Total Rebounds): Rebounding is a valuable skill that contributes to both offense and defense.**

Partial Dependences



- Draft Year:
  - Shows a clear upward trend
  - The relationship appears to be non-linear, with a steeper increase in recent years.
- MP (Minutes Played):
  - The relationship is complex.
  - There's a sharp drop in predicted salary for very low MP, then a rise until about 1000 minutes, followed by a gradual decline.
  - "Sweet spot" where additional minutes don't necessarily translate to higher salaries.
- BLK (Blocks):
  - The relationship appears to be mostly positive but non-linear.
  - There's a sharp increase in predicted salary as blocks increase from 0 to about 20, then a more gradual increase afterwards.

https://www.kaggle.com/code/dansbecker/partial-dependence-plots/

Next Steps:

- The dataset needs to be larger, so I need to get more data
- Feature Engineering:
  - Identify key college performance metrics that might influence NBA salary (e.g., points per game, rebounds, assists, shooting percentages).
  - Generate new features
- Ensemble Methods: Try to combine this tuned XGBoost model with the Random Forest model
- Try a Neural Network model again

Back to Content Page

Thank You!

Back to Content Page