AI Academy Capstone Project: Predicting NBA Salaries

Mason Wagner

Agenda

1 Overview & Project Goals

2 Data & Preparation

3 Regression Models

4 Iteration and Results

5 Conclusion

Business Overview

Industry
Sports/
Entertainment

Client
National
Basketball
Association

Location
Brooklyn,
New York

Project Goals

Problem	Understanding how much NBA players should be paid a year.
Reasoning	Top basketball talent can come at a steep price.
Other Considerations	Make sure players are properly valued per production.

Datasets

NBA Player Statistics from 1984-2018

NBA Player Salaries from 1984-2018

NBA League Salary Cap from 1984-2018

Data Preparation

Combining and Cleaning Databases

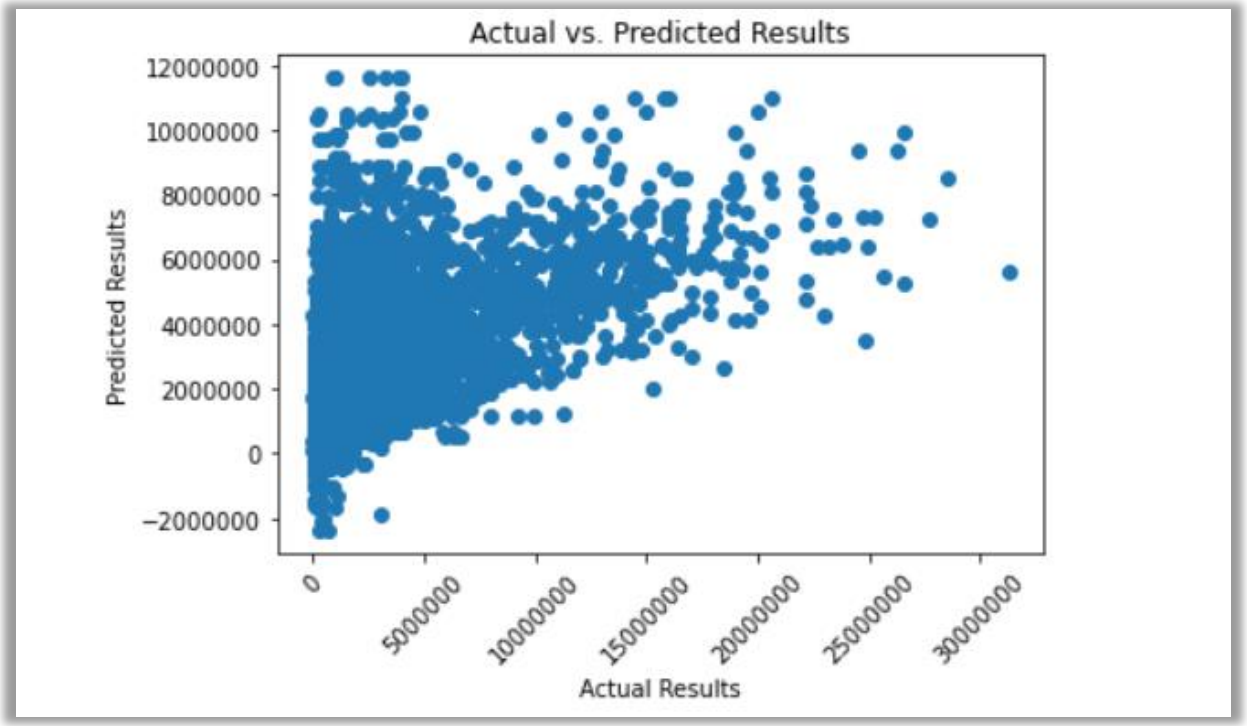
Results

- Points Per Game
- Career Assists
- Career FG Percentage
- Career FG3 Percentage
- Career FT Percentage
- Career Games
- Career Points
- Salary
- Salary Cap

Using Python and Pandas to combine datasets.

One complete Database containing all stats and calculated columns.

The First Linear Regression Model



Results

- **R^2: 0.2159289140732693**
- **MAE: 2,514,196.181567376**
- **RMSE: 3,636,710.895508616**

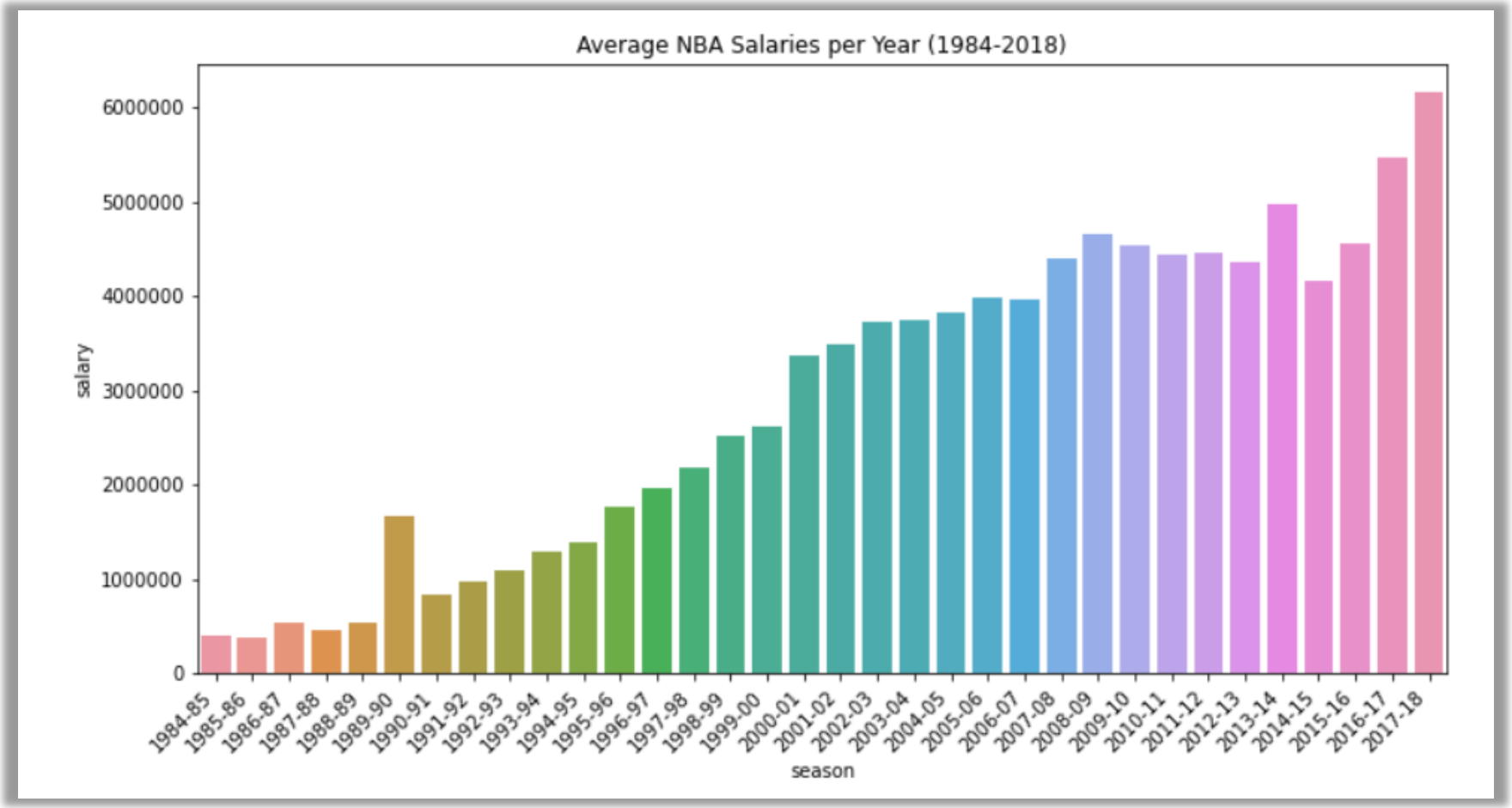
What does this mean?

The linear regression model was able to explain only 21.6% of the variance in Salaries. The MAE shows us that this model is off by about 2,514,196.18 in a given prediction, and the RMSE is off by 3,636,710.89 in a given prediction.

Realization about My Methodology

189.5%
Increase in the dollar's value since 1984

Inflation had to be considered to make this model more efficient.



Standardizing the Target Variables

Salary / Salary Cap

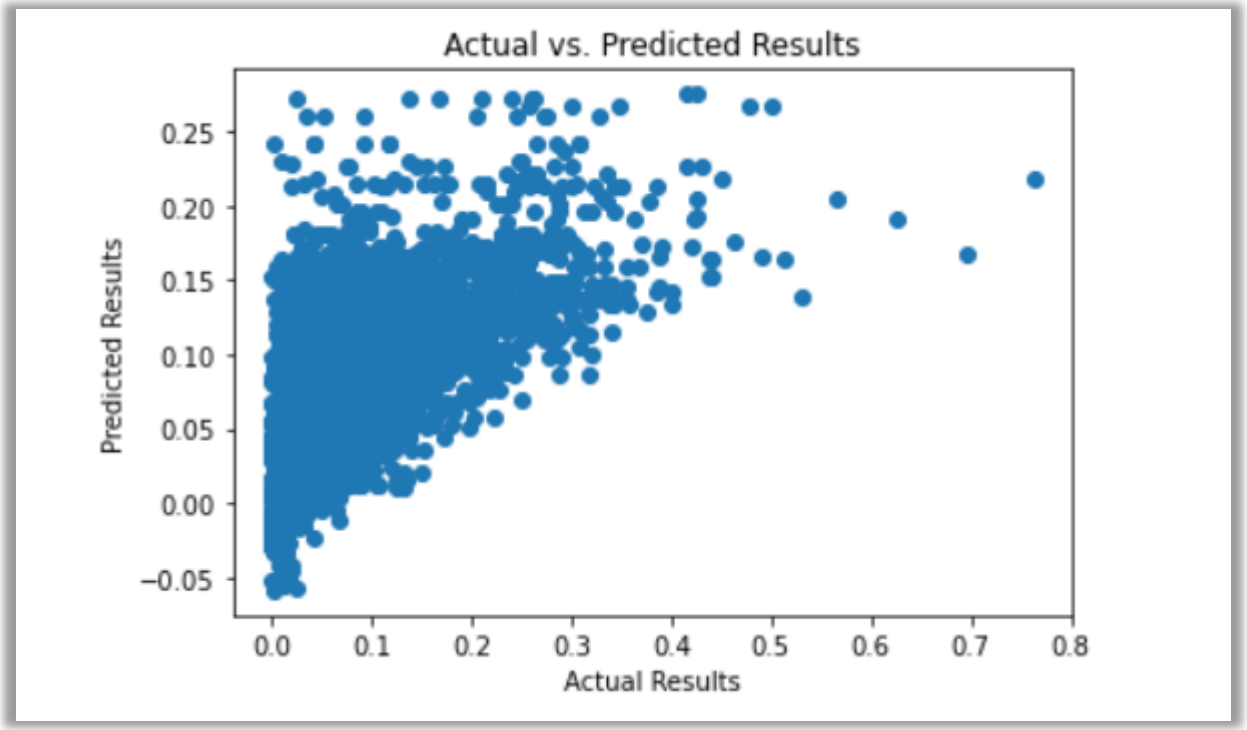
=

% of Salary Cap

Why is this important?

It standardized the dataset so that there is less variation in the target variable (Salary)

The Second Linear Regression Model



Results

- **R^2: 0.40514277465705 47**
- **MAE: 0.04422569210199696**
- **RMSE: 0.0627377252886066**

What does this mean?

The linear regression model was able to explain 40.5% of the variance in Salary cap percentages. The MAE shows us that this model is off by about 4.4% in a given prediction, and the RMSE is off by 6.3% in a given prediction.

Conclusion and Recommendations.

Simply considering inflation and standardizing the target variable of Salaries, improved my model by two-fold. Although it's still not perfect, potentially looking into datasets with more variables could be useful.

My recommendation would be to:

- Investigate finding a more robust dataset like sources to improve the model (Both extensive basketball databases) such as:
  - Kenpom
  - Basketballreference.com
- Explore another model type like XG boost or Random Forest.

Thank you!
Questions?