



# NBA Player Salary Prediction Model

**Micah Matsuda**  
Springboard Data Science Bootcamp  
Capstone Project

# Contents

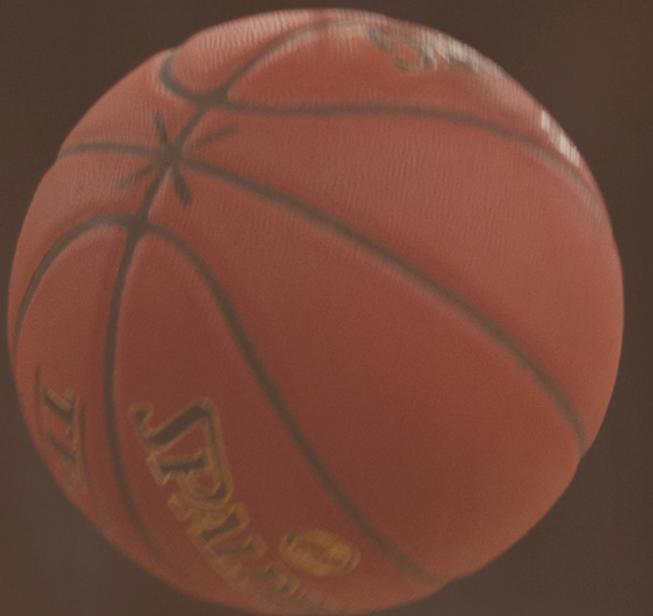
Goals

Exploratory Data Analysis

Key Findings

Modeling Results & Analysis

Conclusion



# What determines an NBA player's salary?

NBA's revenue topped \$10 Billion for the first time last year.

What does this mean for players?  
What determines a players salary?

\$40M+

is going to players like LeBron and Steph Curry.

Is this reflected fairly?



A basketball hoop and a basketball are positioned on the left side of the frame. The hoop is made of orange rim and white netting. The basketball is orange with black stripes and has the word "RAWLINGS" printed on it.

# Exploratory Data Analysis (EDA)



# EDA

## The goal is to answer:

What are some features that determine salary?

Which ones have more impact compared to the rest?

Does position affect a higher salary?

On average, what is the age group that has the highest salaries?

Which team has the highest total salary?

### △ Joint plots

EDA

Draws a plot of two variables with bivariate and univariate graphs (scatter plot and histograms)

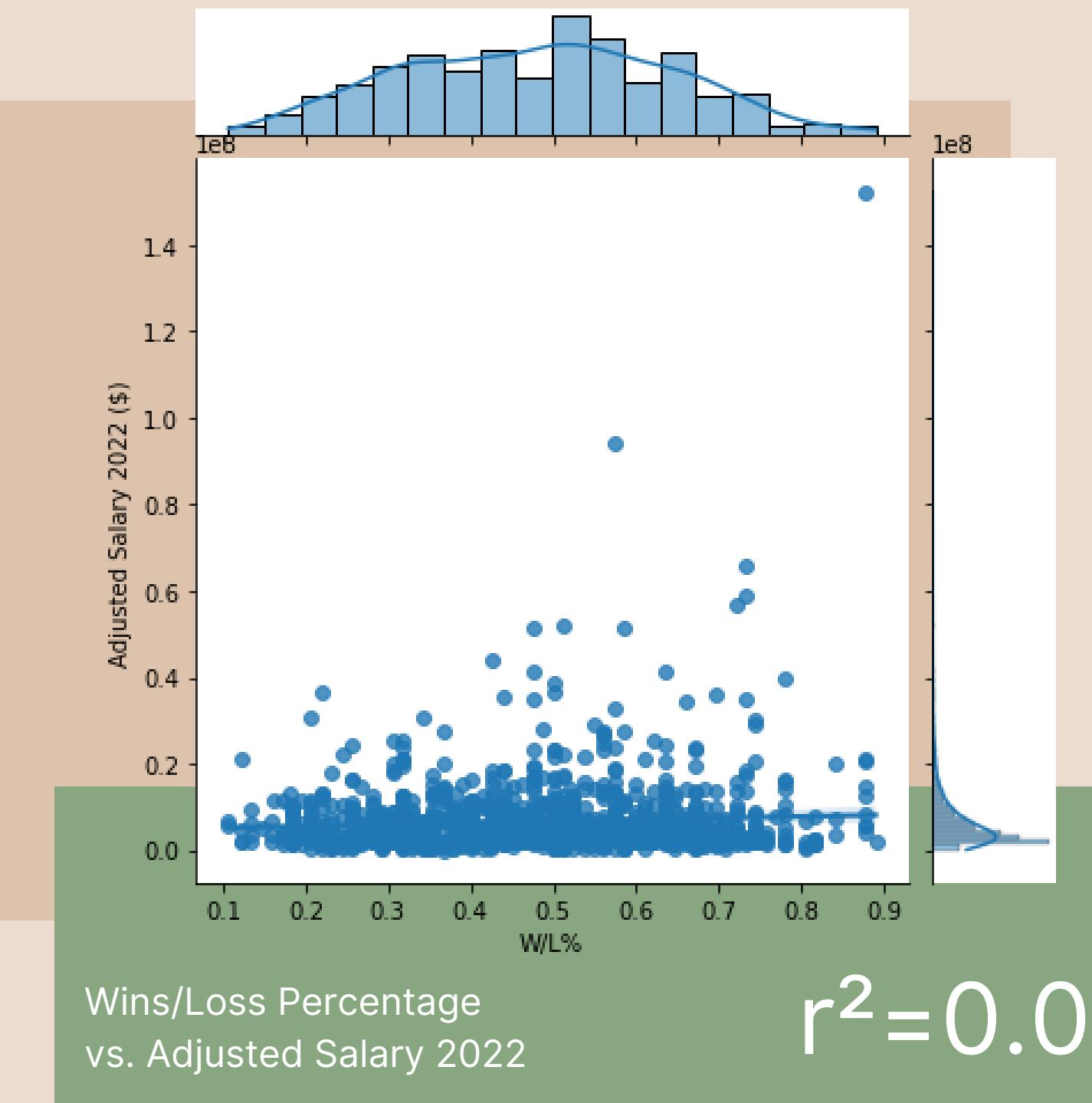
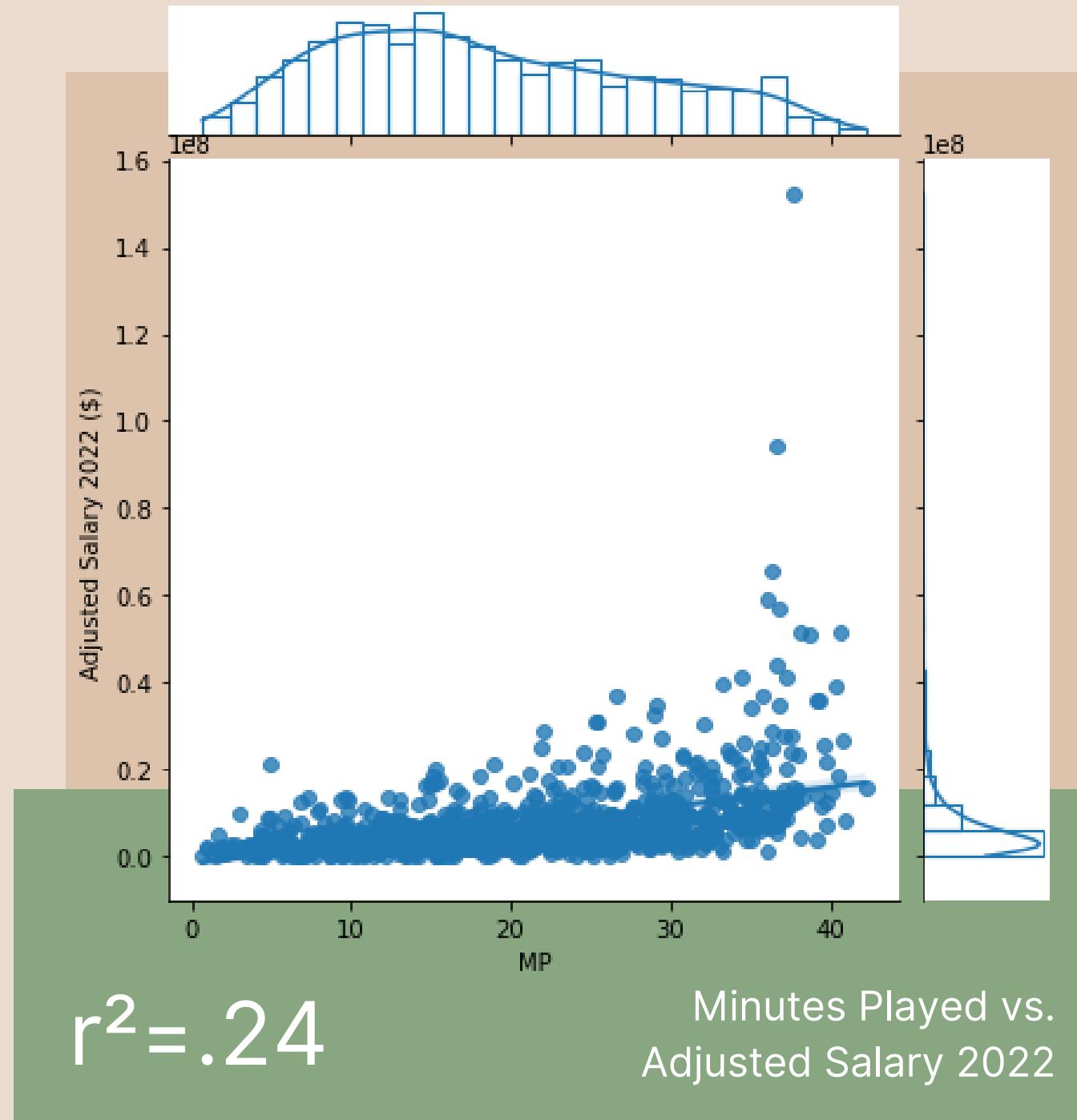
### △ Box plots

EDA

(or box-and-whisker plot) shows the distribution of data in a way that is comparing variables or across levels of a categorical variable

# Joint Plots

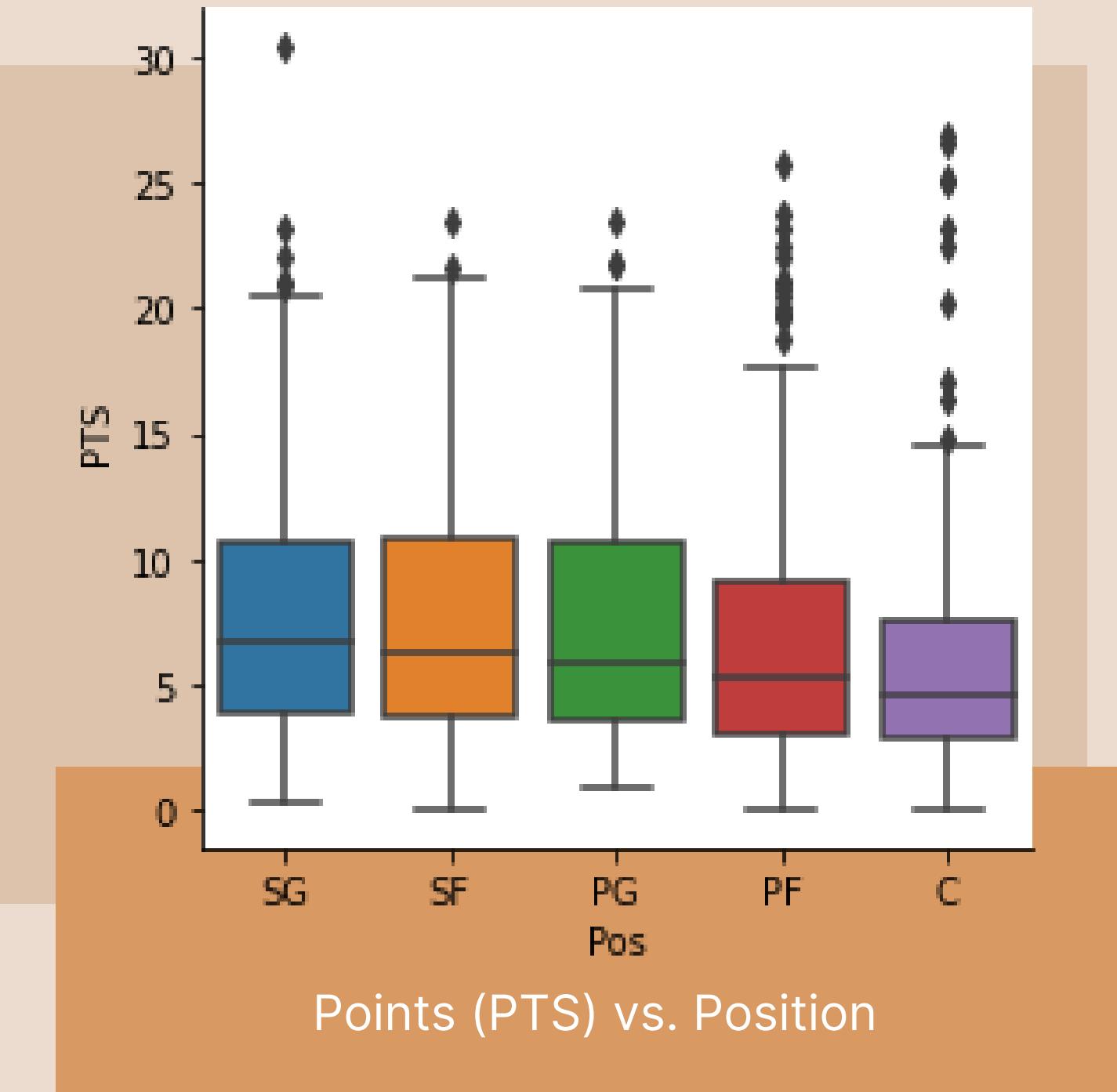
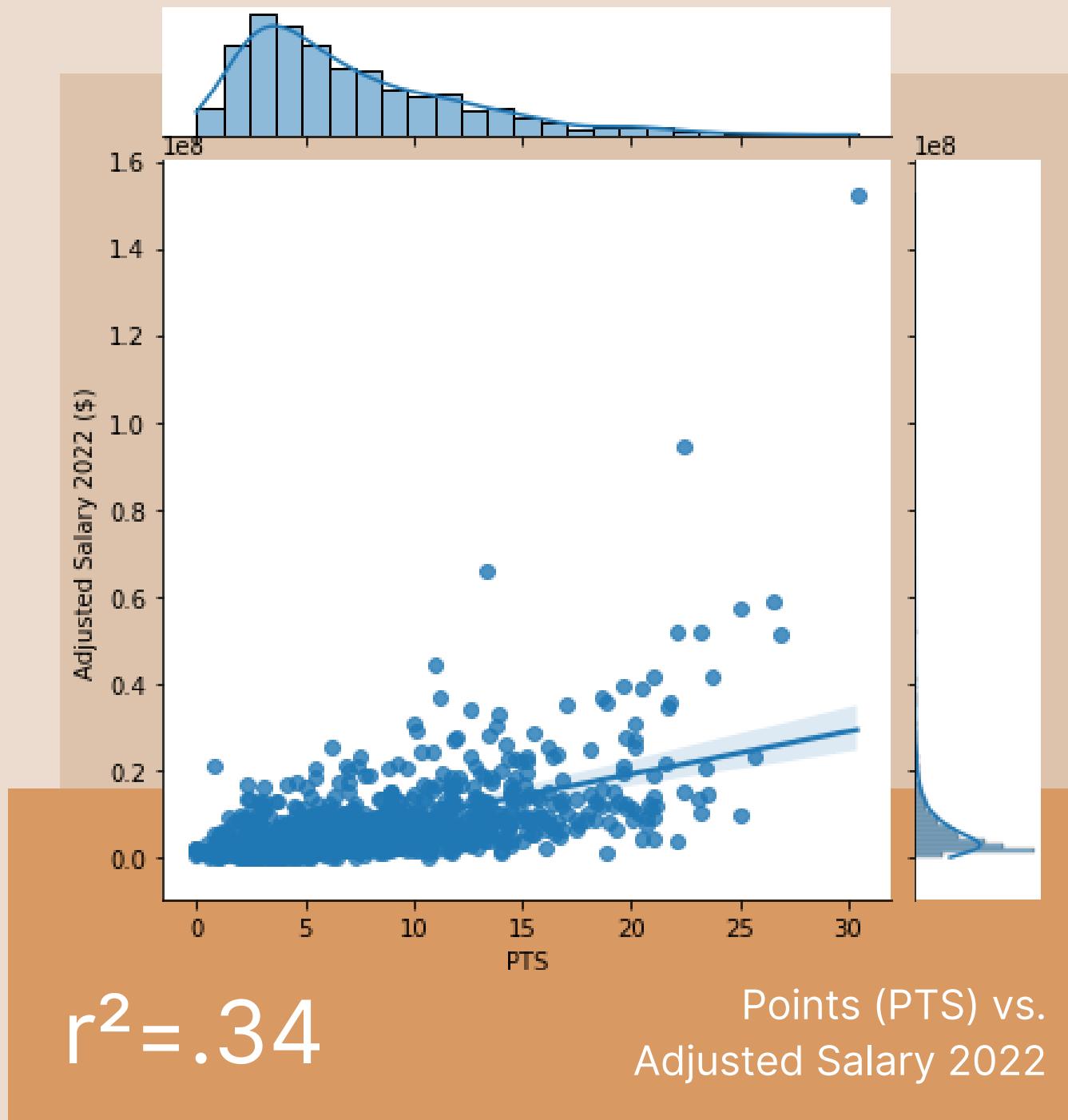
Selecting the features that are assumed to be most important in contribution to salary →



# Points (PTS)

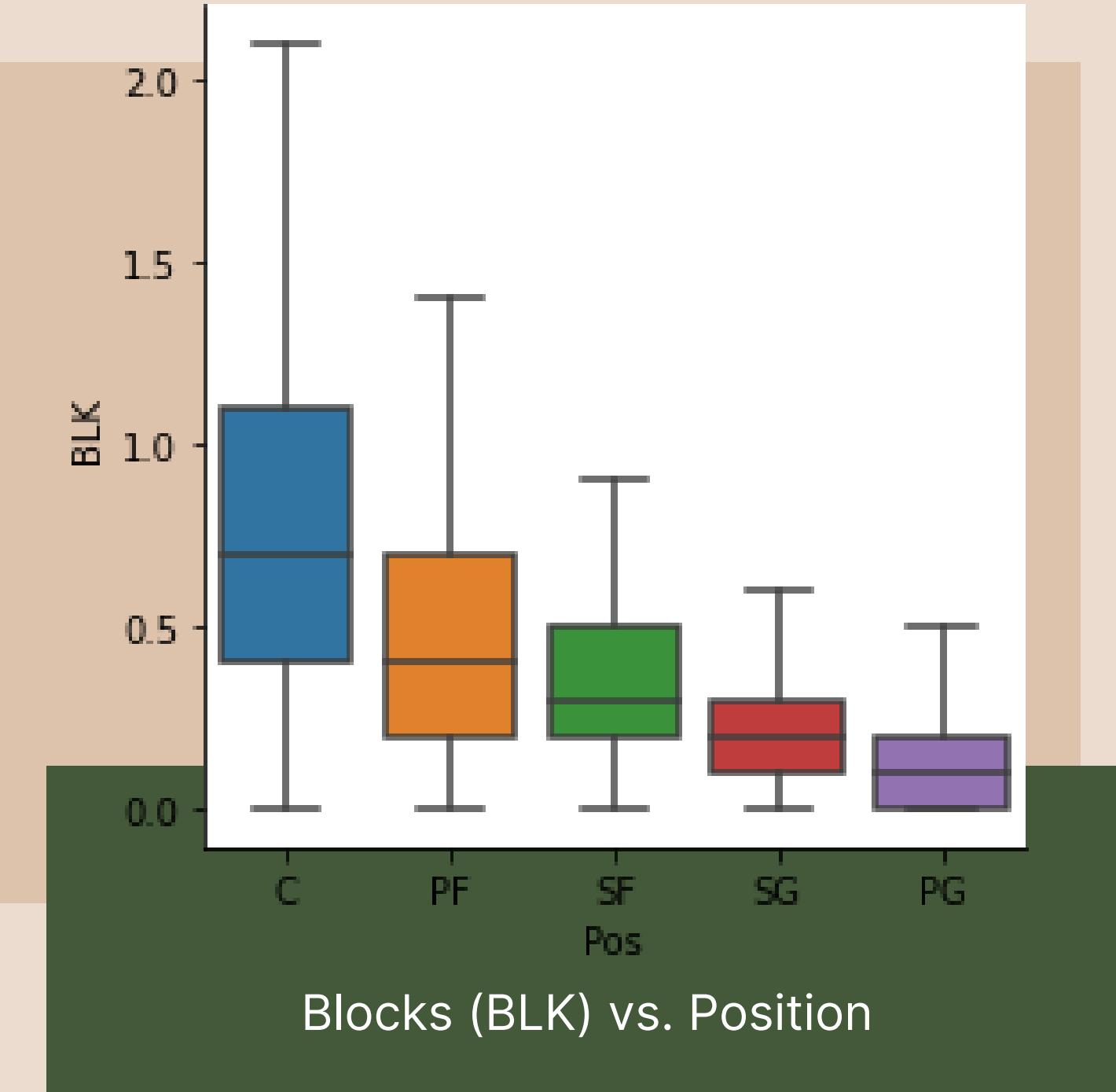
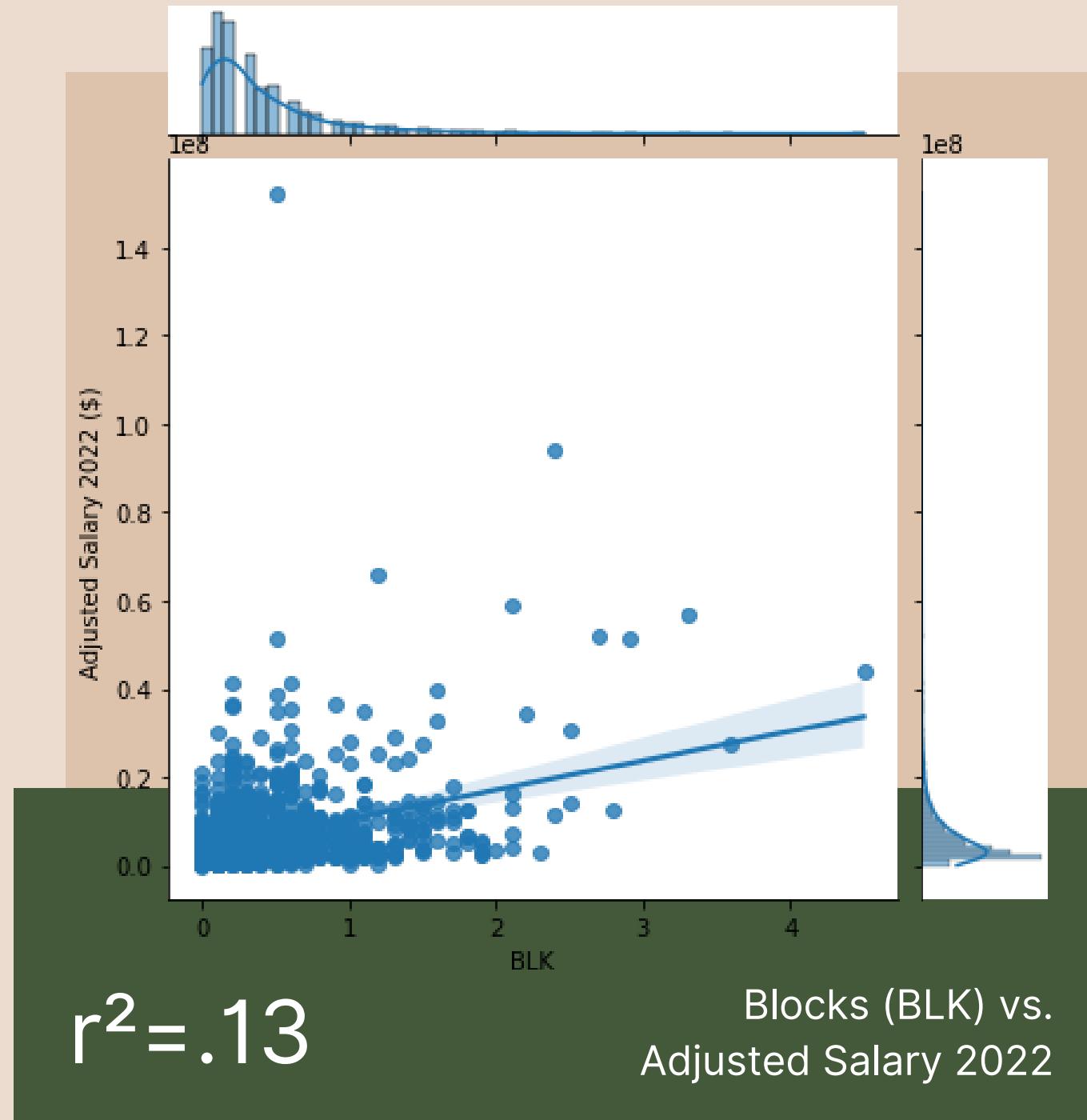
EDA

Points vs. Adjusted Salary 2022 compared by Position →



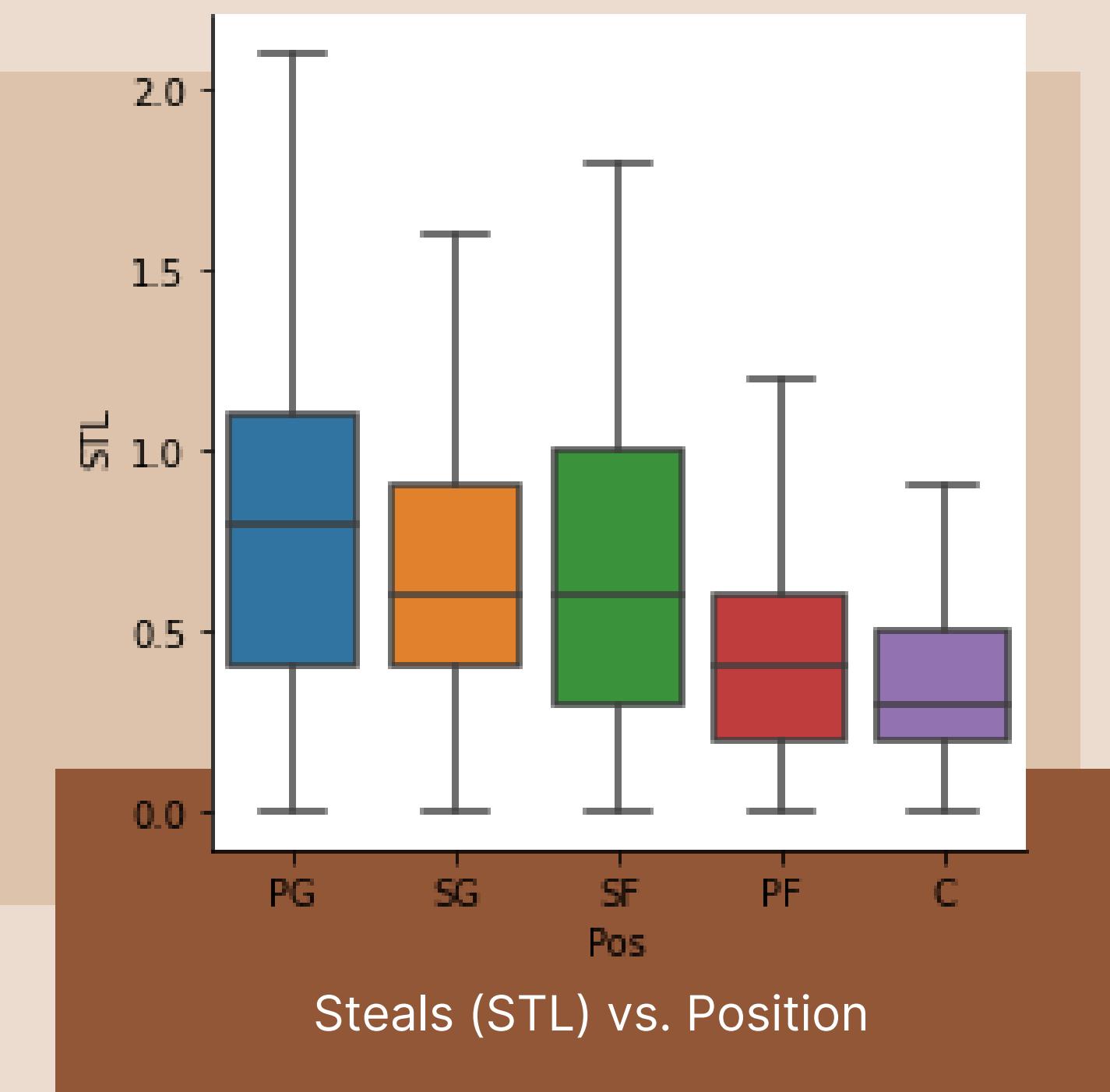
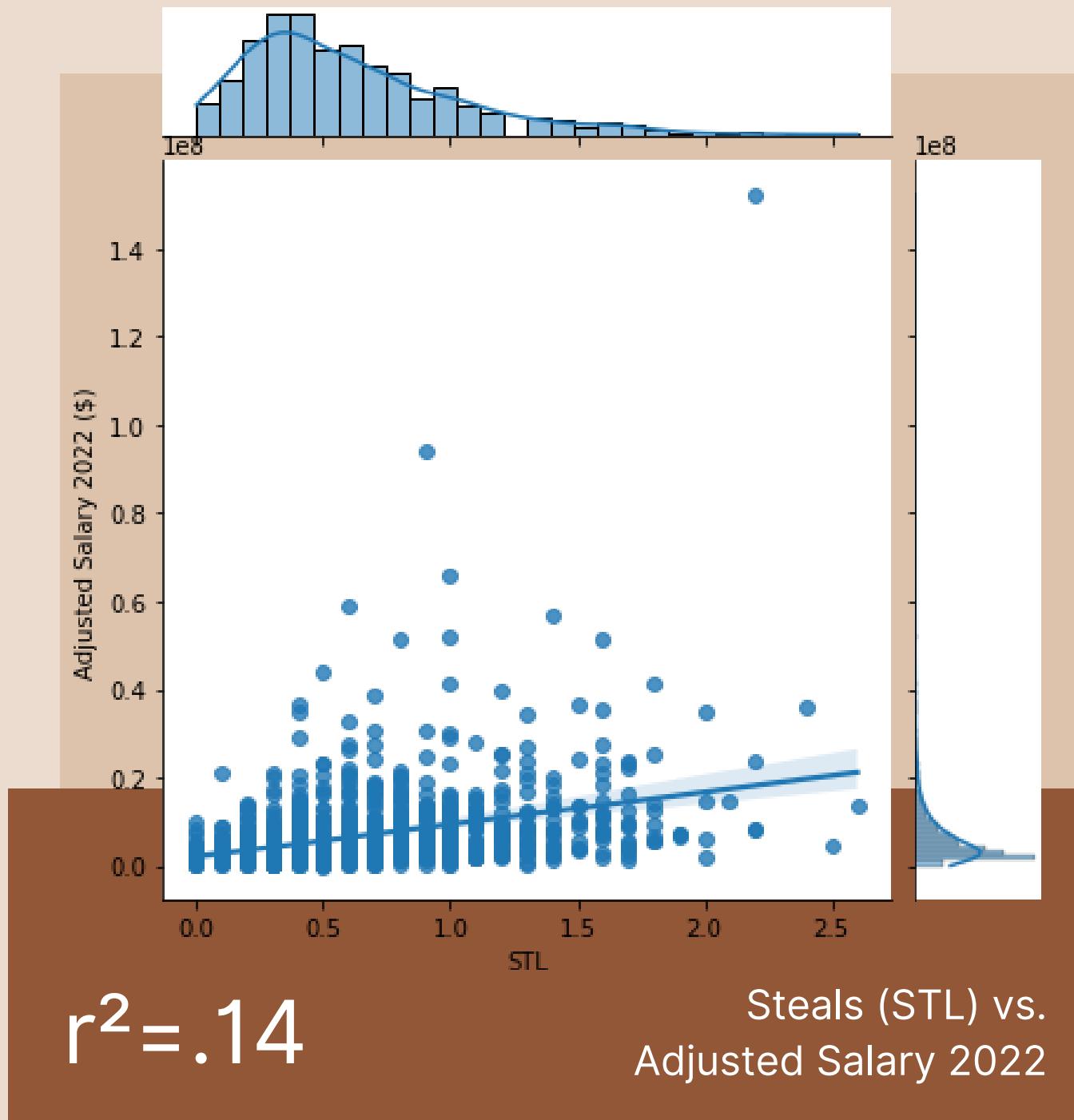
# Blocks (BLK)

Blocks vs. Adjusted Salary 2022 compared by Position →



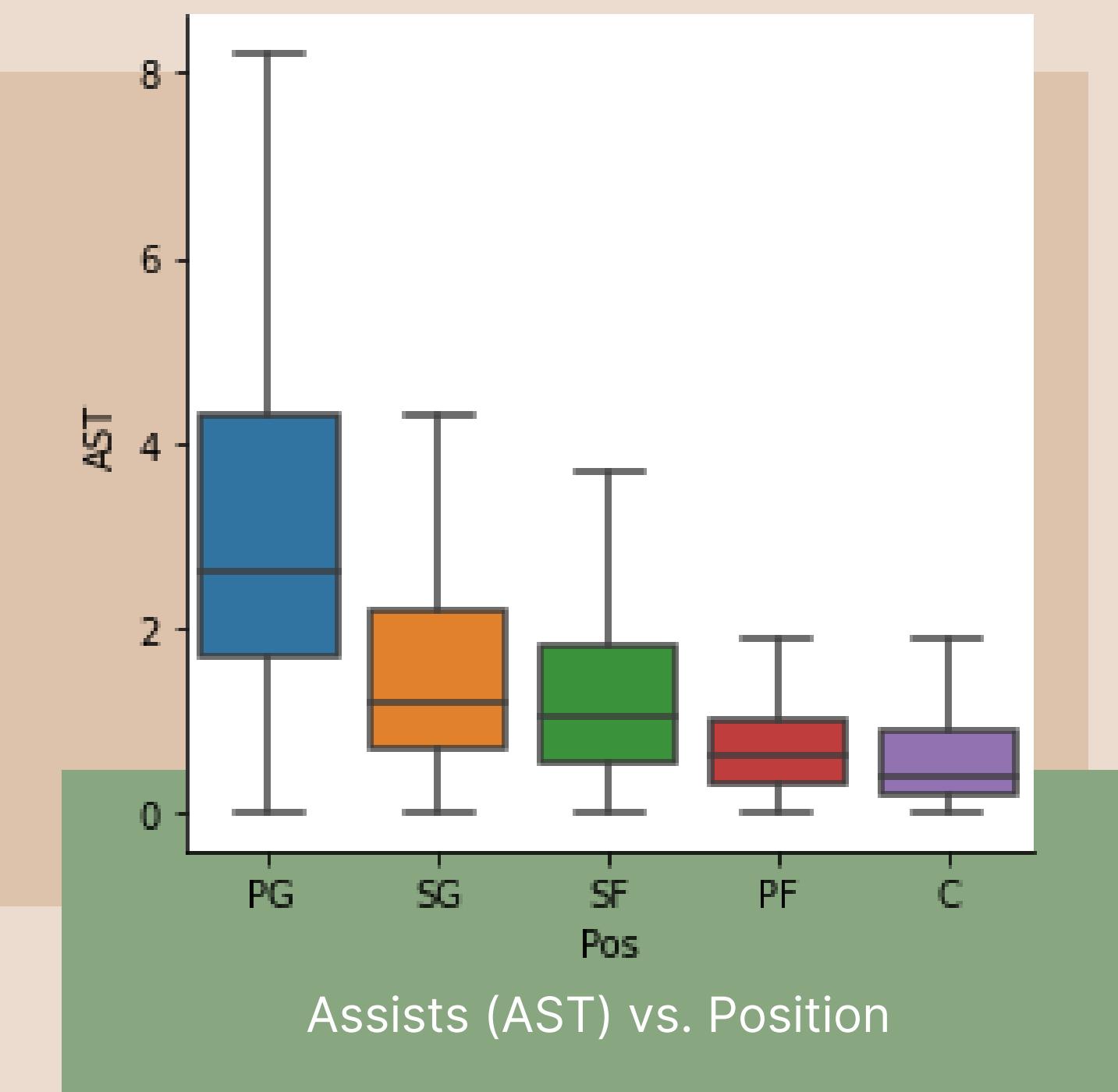
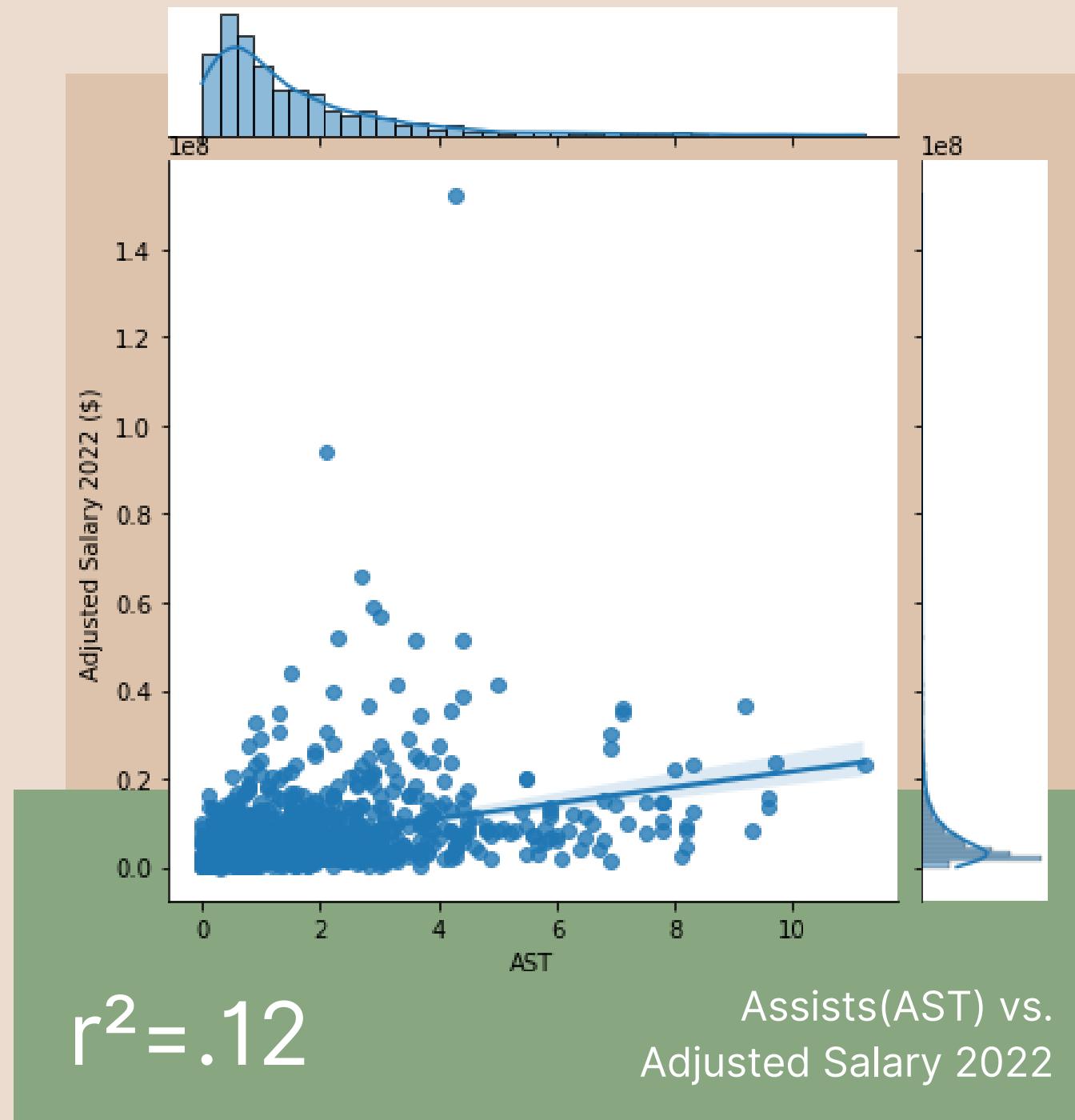
# Steals (STL)

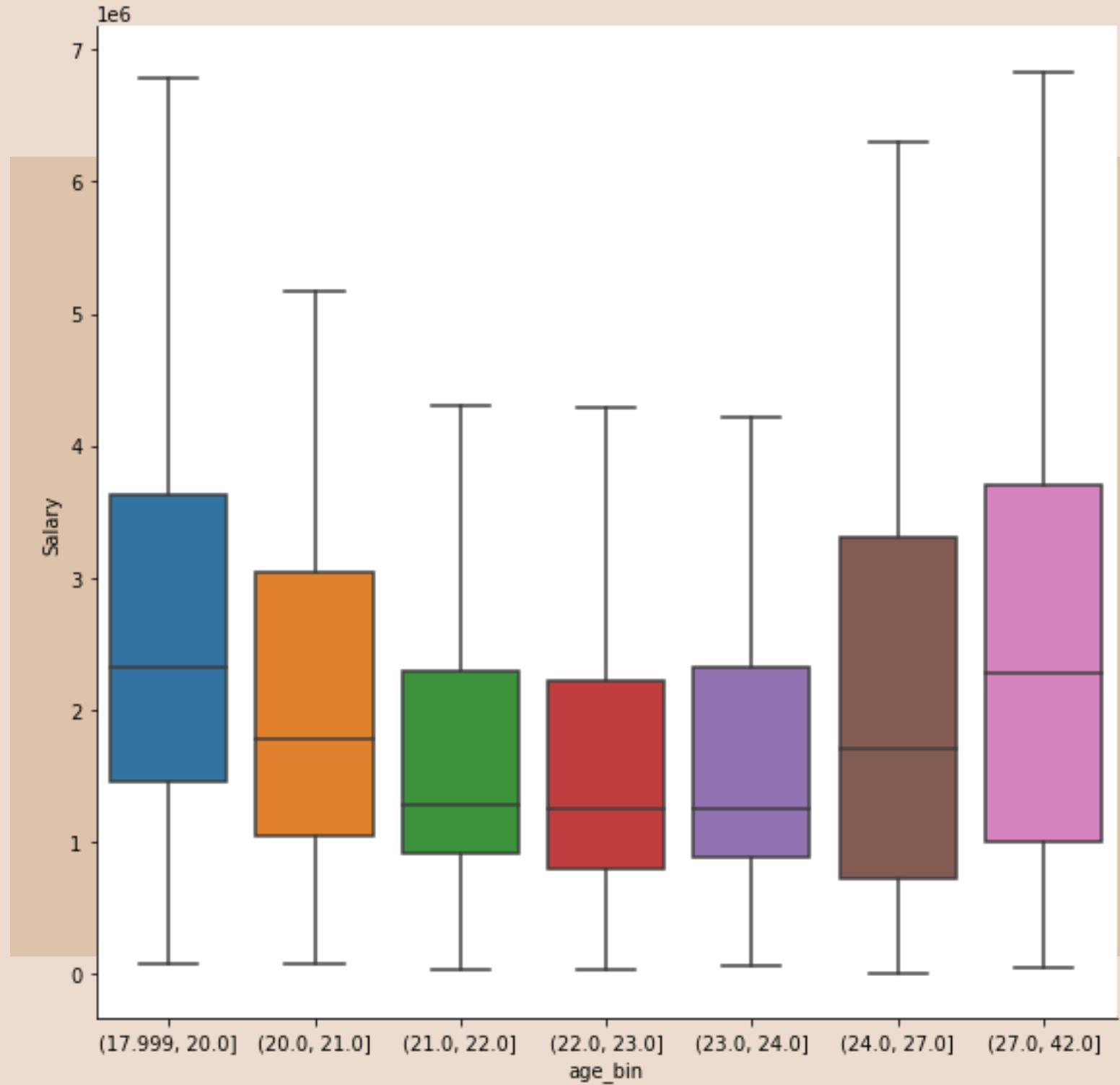
Steals vs. Adjusted Salary 2022 compared by Position →



# Assists (AST)

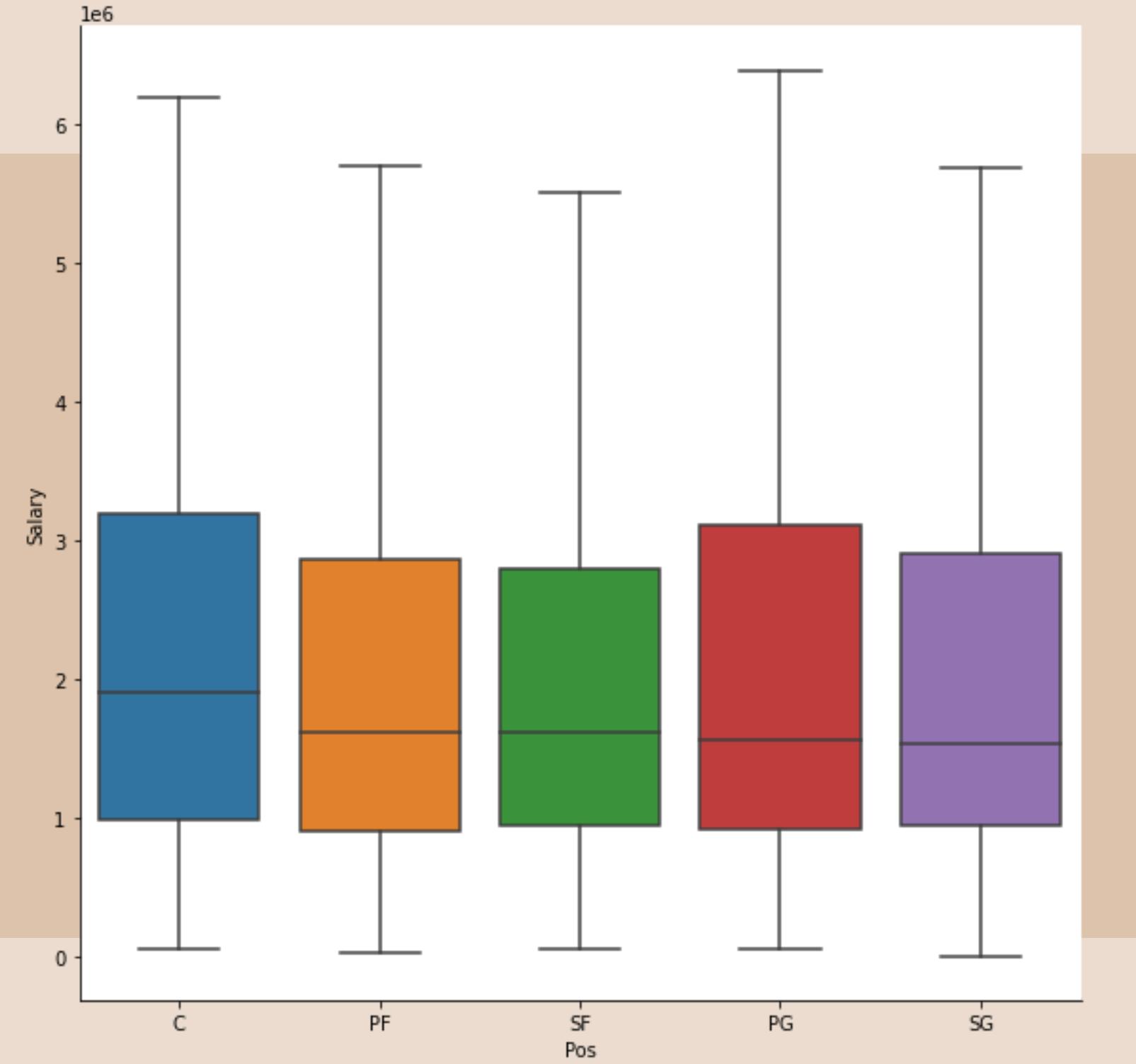
Assists vs. Adjusted Salary 2022 compared by Position →





## Age vs. Salary

- Aged bins ranges
- Age rule before 2006
- Longevity of players



## Position vs. Salary

- Center and Point Guard
- Median Values

# Correlation Heatmap



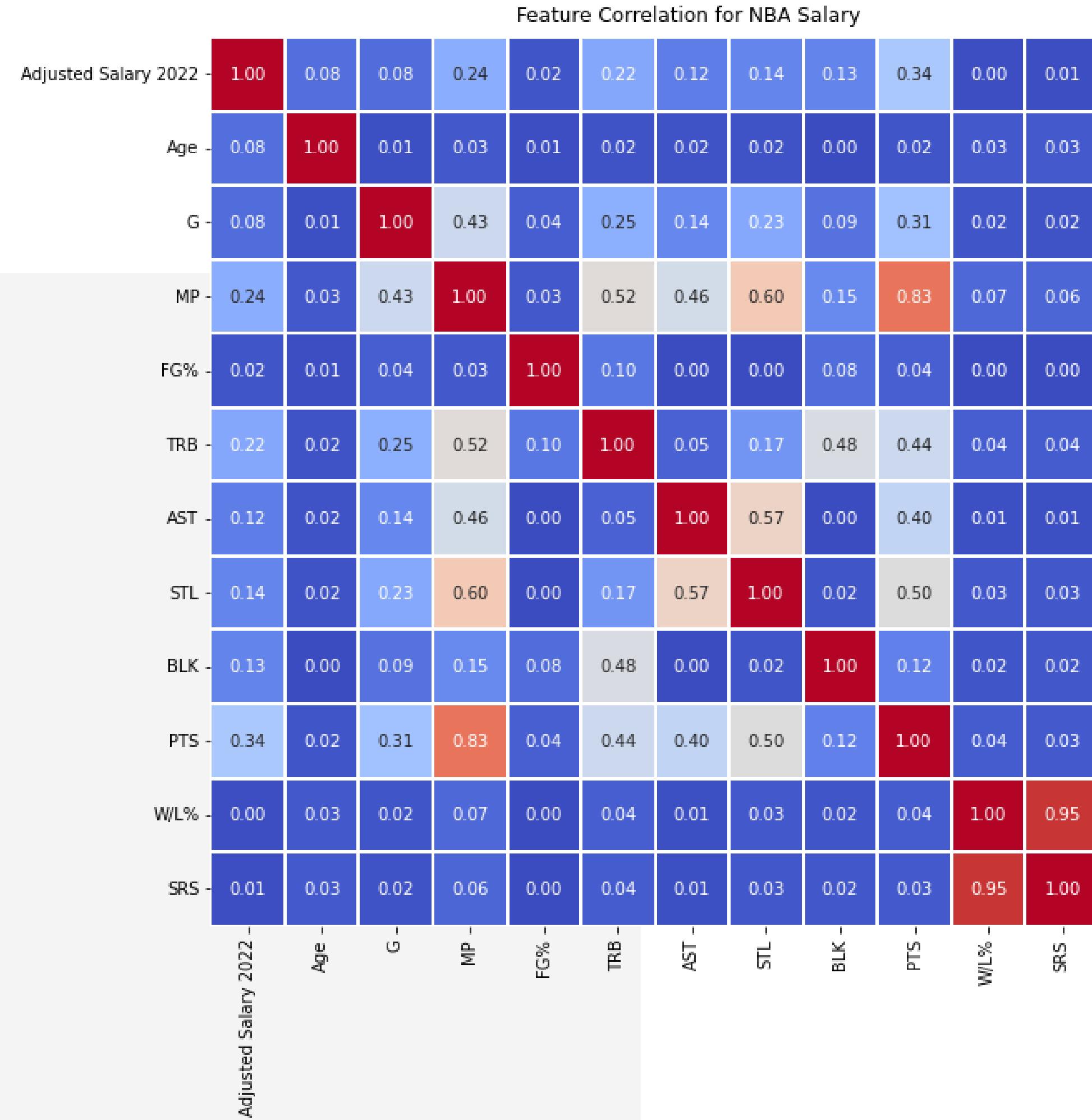
Low Correlation between features and Salary



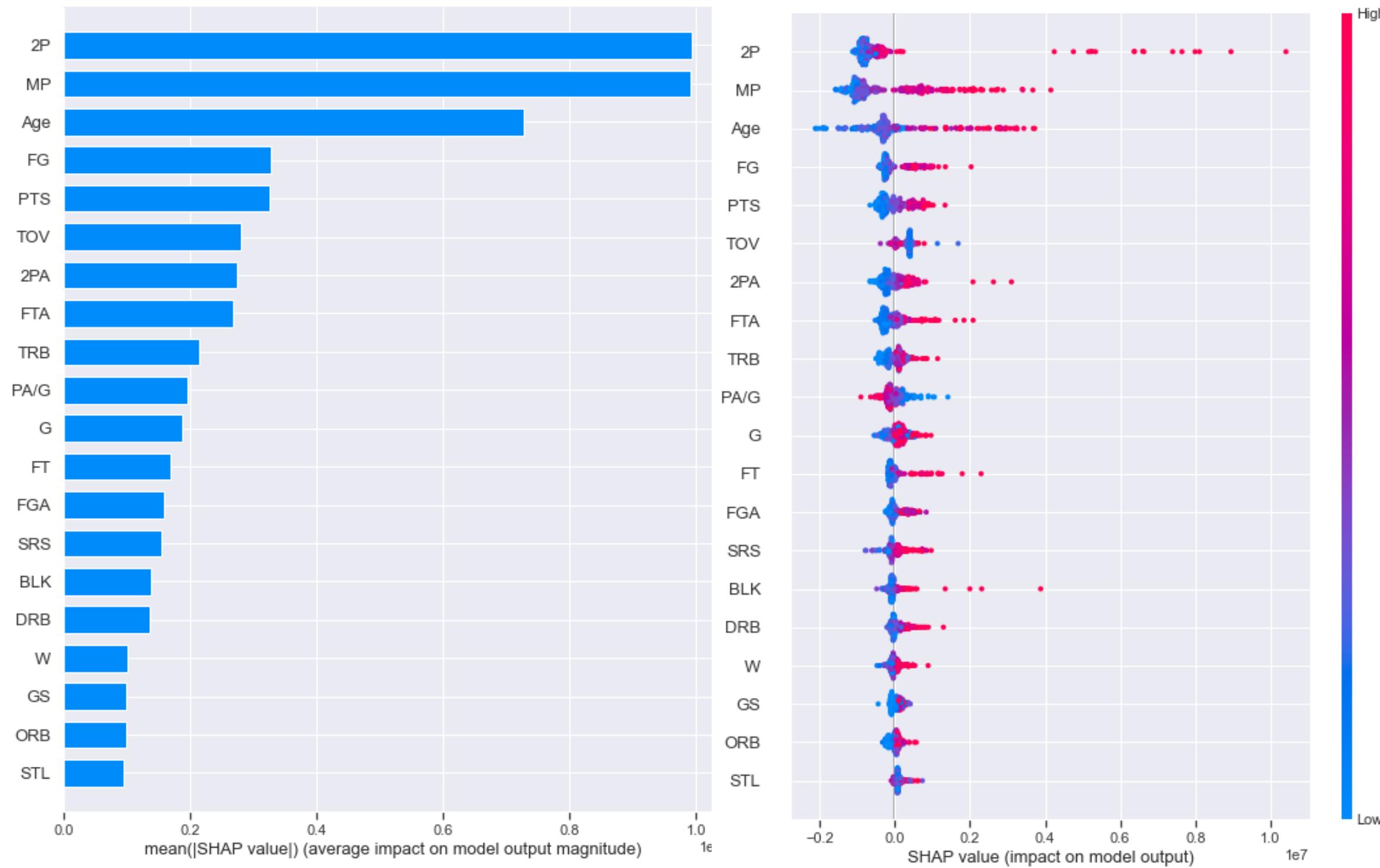
Able to use all features in the modeling process



Noteworthy high correlation features



# SHAP Feature Importance



2P  
High Importance

MP  
High importance, low correlation

Age  
is of importance too, longer career reflects higher numbers

# Key Findings

## The importance of Feature Importance

When looking on the surface, one would think points, wins/losses, or even the amount of games played during a season would contribute to a player's future salary. While there is some correlation featured on the heat map, the impact differed in the feature importance graphs.

Assists, blocks, and steals all matter in a game, however in the overall impact on their salary, there is little correlation to the impact it has on salary.

A photograph of a basketball hoop and a basketball. The hoop is made of white mesh and is attached to a silver metal rim. The basketball is orange with black stripes and has the word "PALLONE" printed on it.

# Modeling Results and Analysis

# Dummy Model

# Ridge Linear Model

# Random Forest Regressor



**Mean Absolute Percentage  
Error**  
**304.7%**

**Mean Squared Error**  
**11369831**

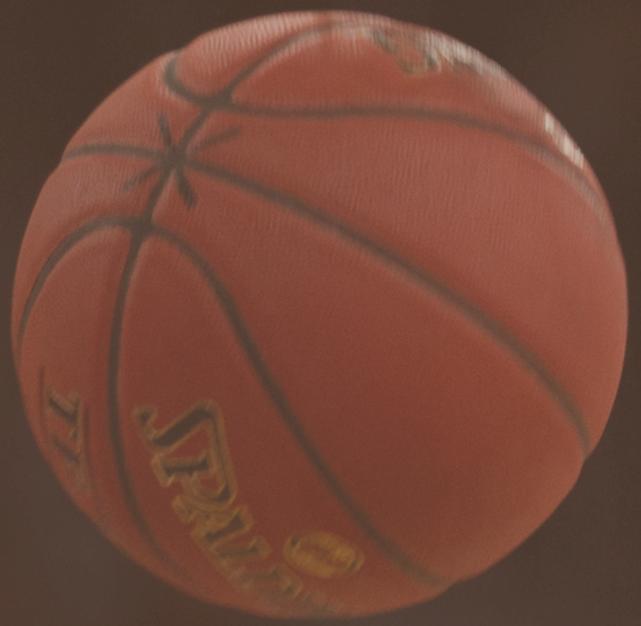
**Mean Absolute Percentage  
Error**  
**172.1%**

**Mean Squared Error**  
**7147890**

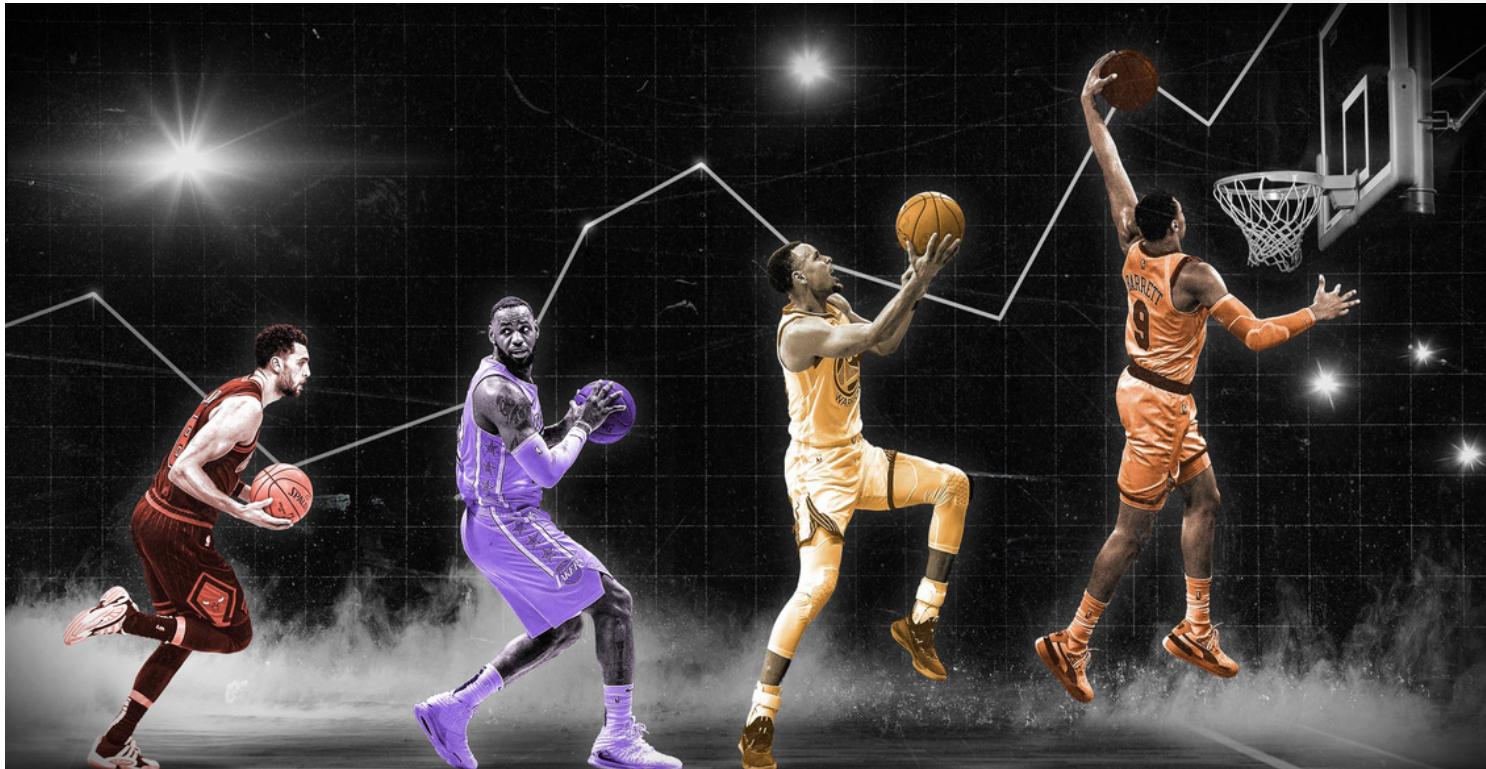
**Mean Absolute Percentage  
Error**  
**166.9%**

**Mean Squared Error**  
**8326467**

# Conclusion



# To look out for next time



## Time Value in Money

Inflation – values could be different in EDA



## Feature Selection

Could've added more, but also could have taken out a few that did not matter

## More machine learning models

More work done earlier for models to work better and develop more ML models

## Fine Tuning

Take more time to hypertune parameters, but still a lot more work would need to be done earlier to achieve that

# Thank you.

