Predicting NBA Salaries During Free Agency

Vihaan Hari

Agenda

Topic Overview
01

Dataset
02

Exploration
03

Model
04

Limitations
05

Insights
06

Topic Overview

Free agency refers to the offseason period where players **switch** teams after being offered different **contracts** and options

Trades

Players are **exchanged** for one another along with cash or pick compensation

Signings

Players **leave** their current teams and **accept** a contract offered by another team

Cuts

Players are **taken off** the roster for their current team and are able to be signed

Significance

For 2023 alone, there is a **15% increase** in Unrestricted Free agents and **10% increase** in Restricted Free Agents.

→

- Roster Changes
- Fan Attendance
- Salary Cap Management

Intro to Dataset

2016 Free Agency Class (2015-16 Stats)

Basic Stats

Individual player performance metrics

Points per Game
Field Goal %
Assists per Game

Advanced Stats

Team-based player performance metrics

VORP
Win Shares
Player Efficiency Rating

Categorical Data

Data not recorded on Stats Sheet per game

Position
Home State
Type of Free Agent

Dataset Quality

Advantages

Mix of **quantitative** and **qualitative** data

38 unique features to describe metrics

Individual and **team** stats are accounted

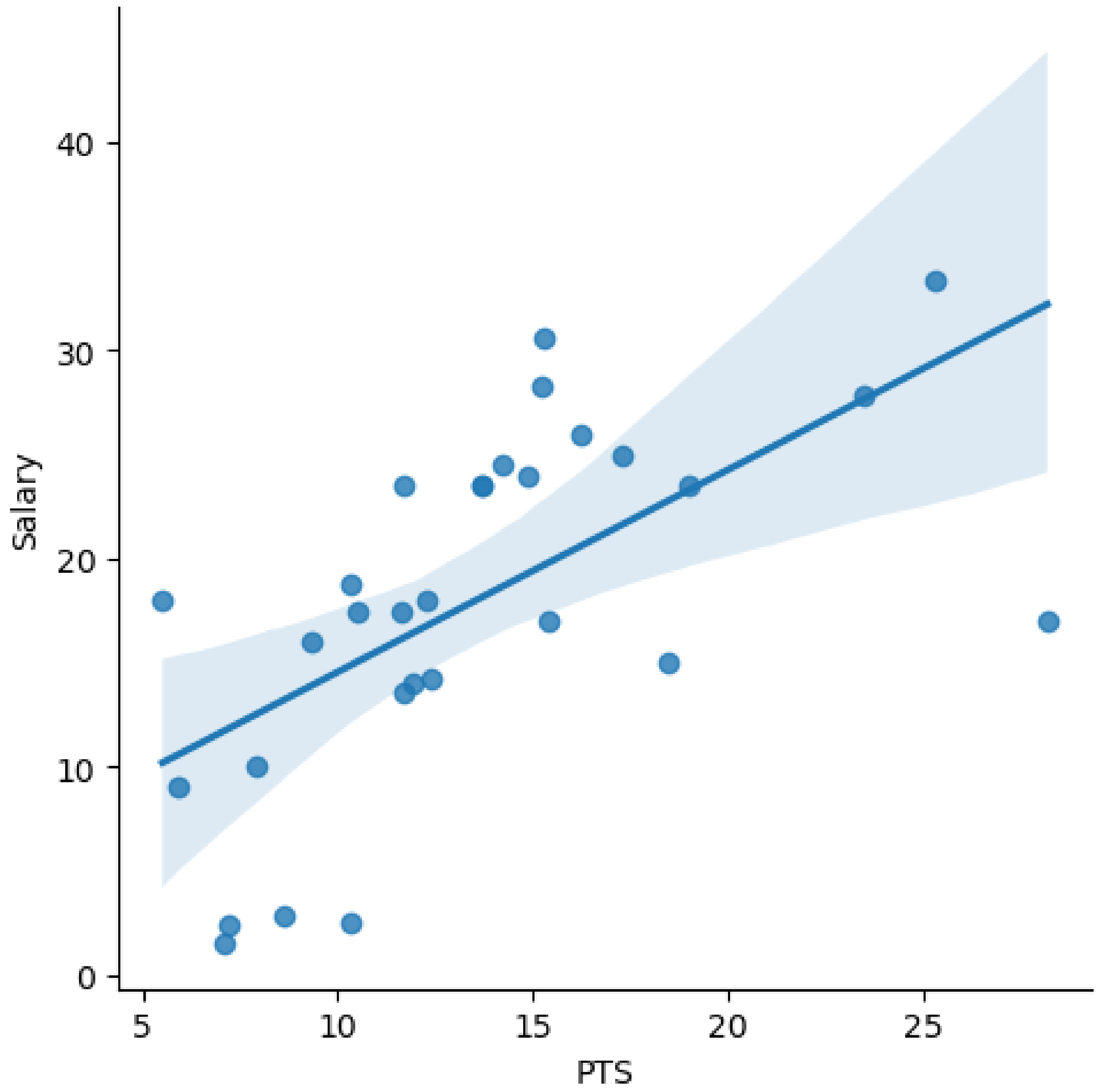
Disadvantages

Only **60** UFA in 2016 free agency cycle

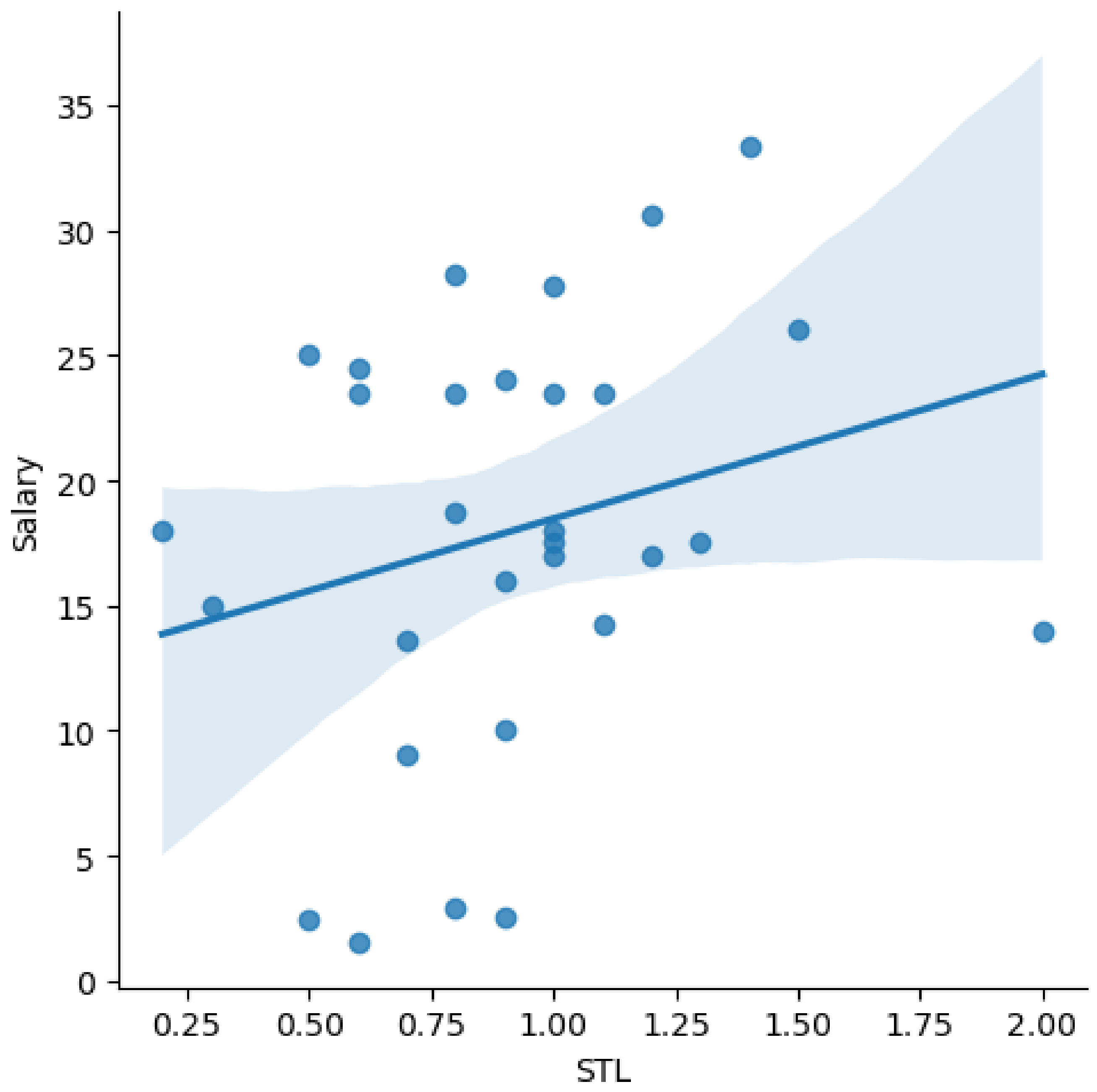
Limited number of **advanced** metrics

Lack of standardization from player **types**

Basic Stat Exploration

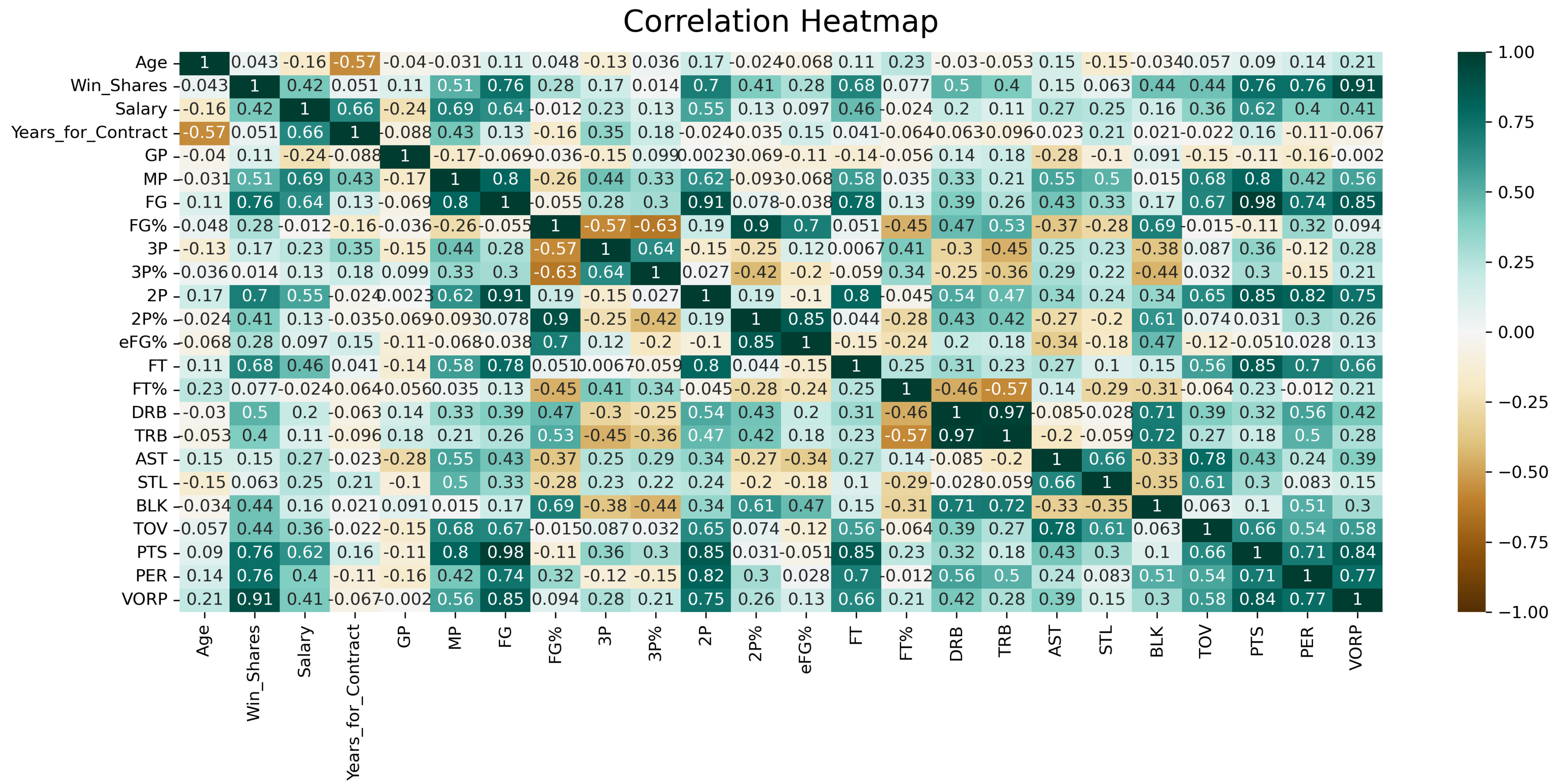


r=0.687 indicating a moderate positive relationship between PPG and Salary



r=0.247 indicating a weak positive relationship between Steals and Salary

Correlation Heat Map



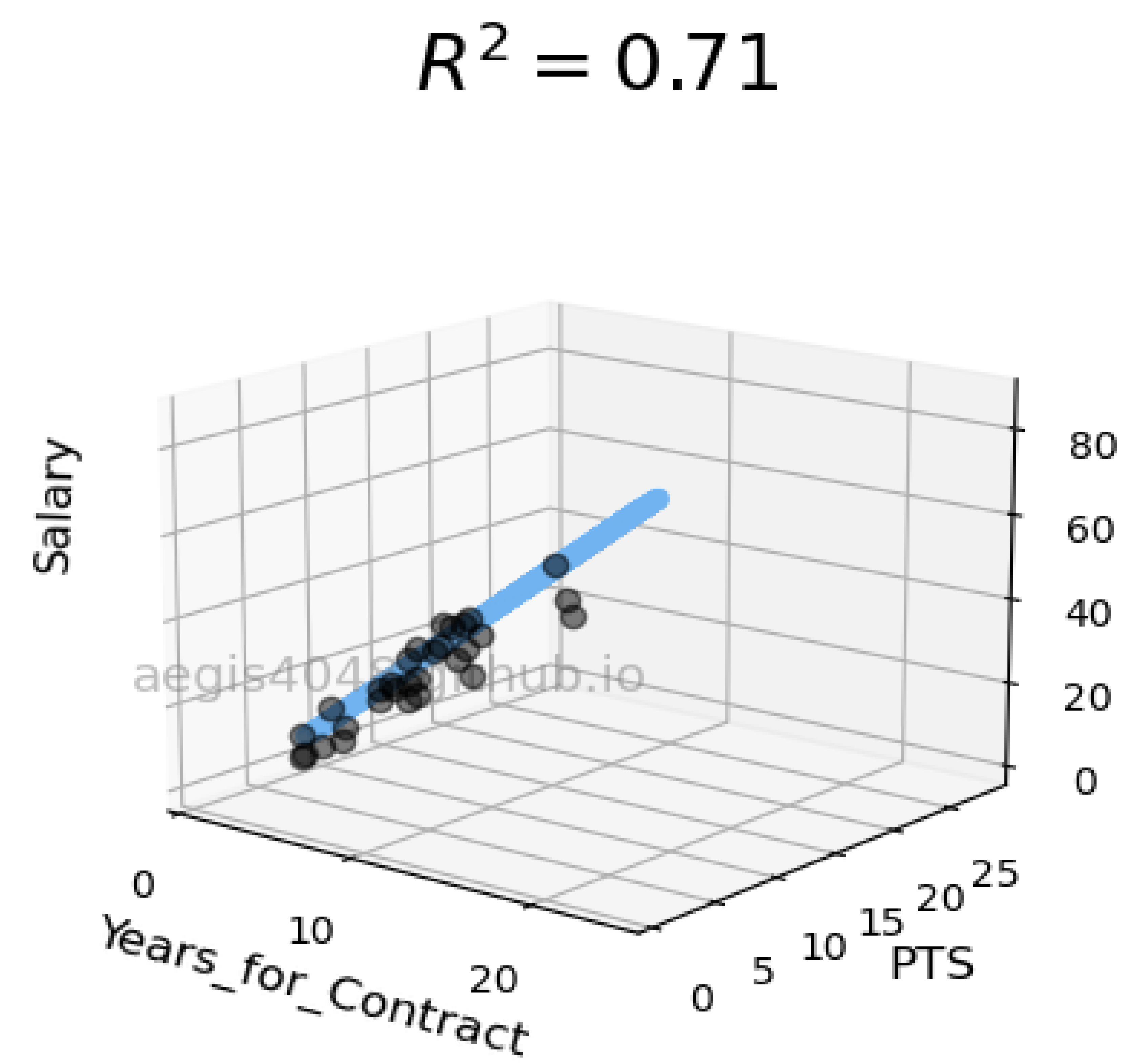
Includes basic, advanced, and categorical data

Years for Contract, MP, FG, and PTS have the **highest** correlation

Efficiency has **low correlation** to salary size

Advanced Stats have **moderate** correlation

Regression Model



Used **Years** for Contract and **PPG** to perform a multiple variable regression to see the importance of these two features with respect to **Salary** size

71% of variance in determining salary size can be accounted by the two features and they have a r value of 0.843, indicating a **strong** relationship with salary

Other variables that had a significant impact were VORP, PER, and eFG% which contributed to **68%** of variance being accounted for with a r value of 0.826, indicating a **strong** relationship

Predictive Model

Feature Selection

PER

VORP

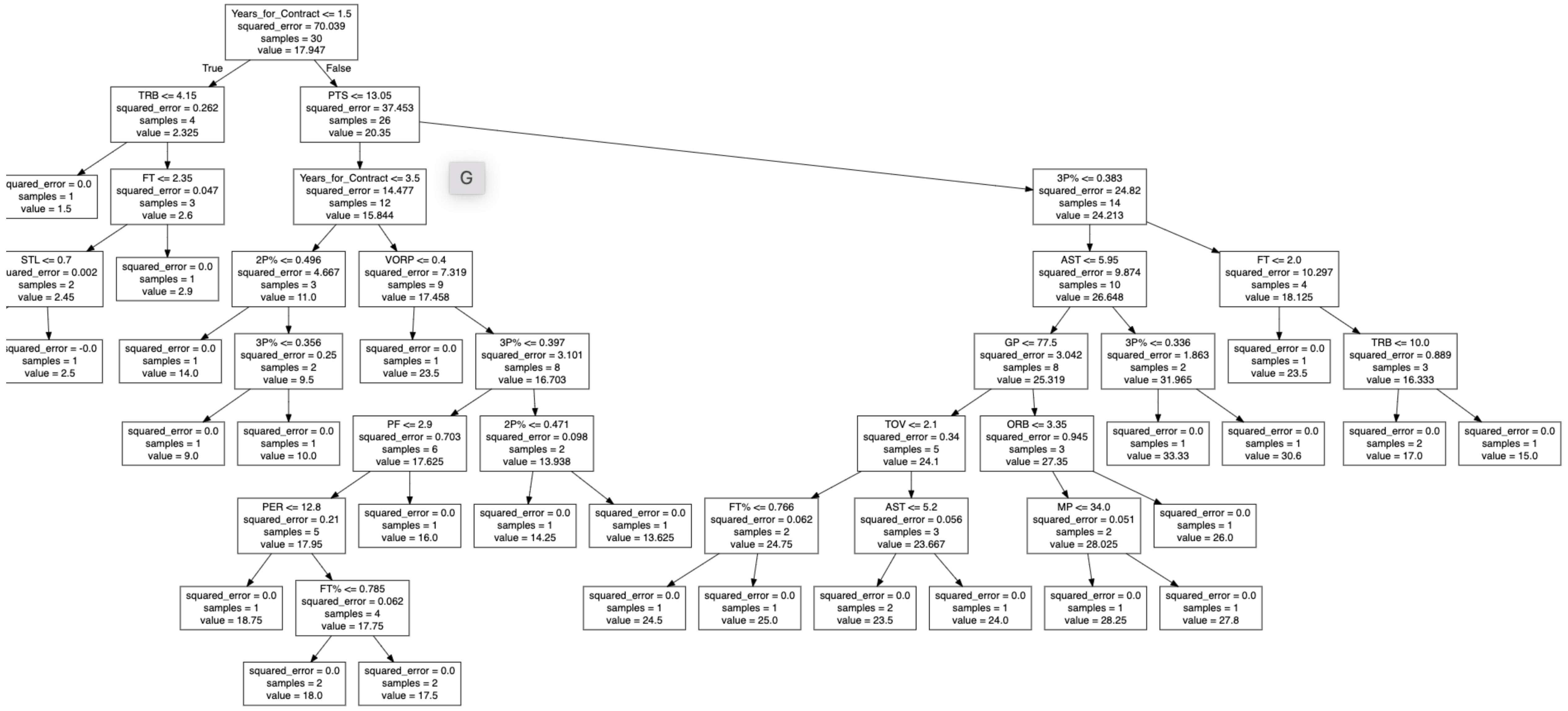
Field Goal %

Games Played

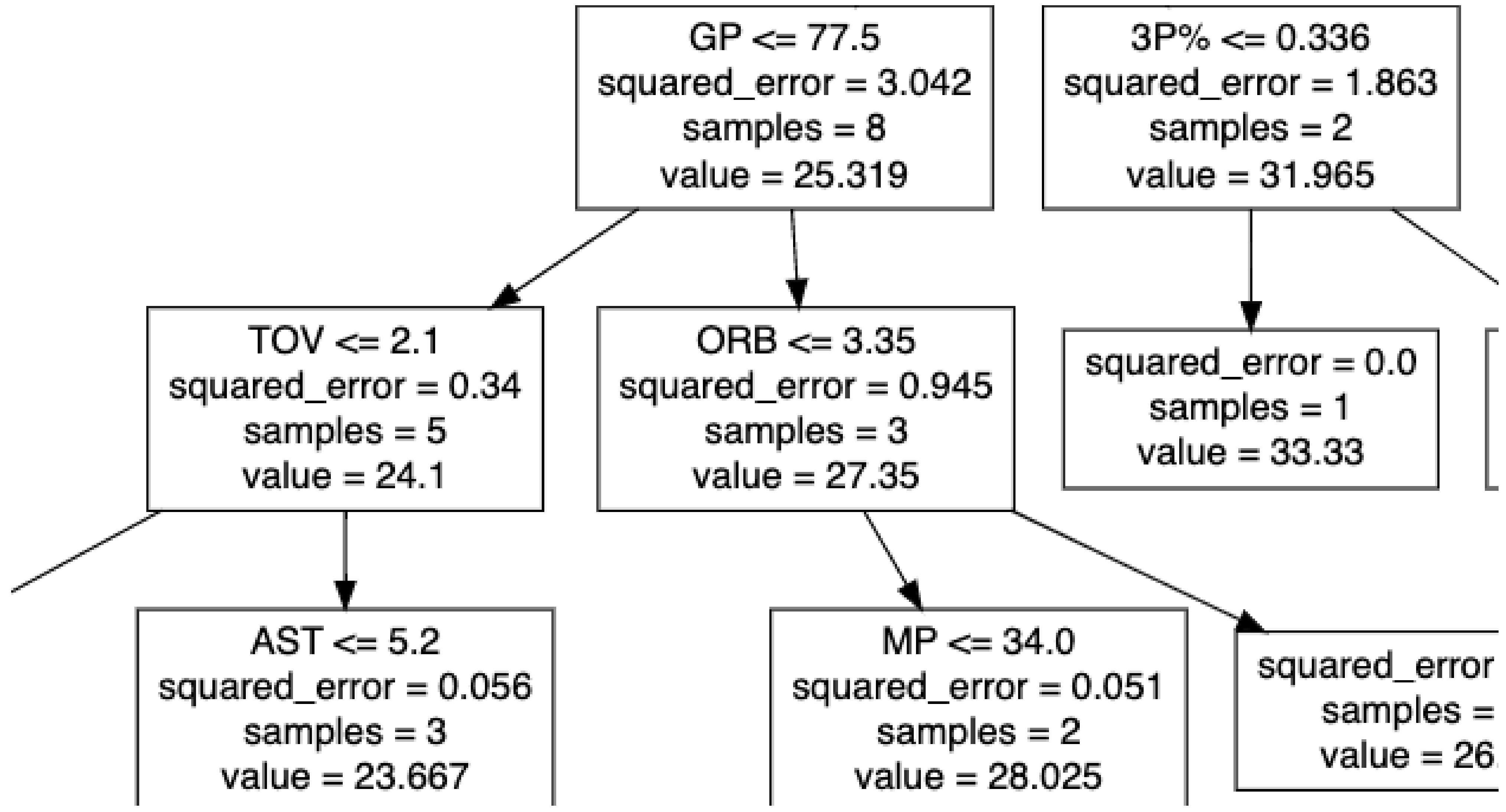
Minutes Played

Points per Game

Years on Contract

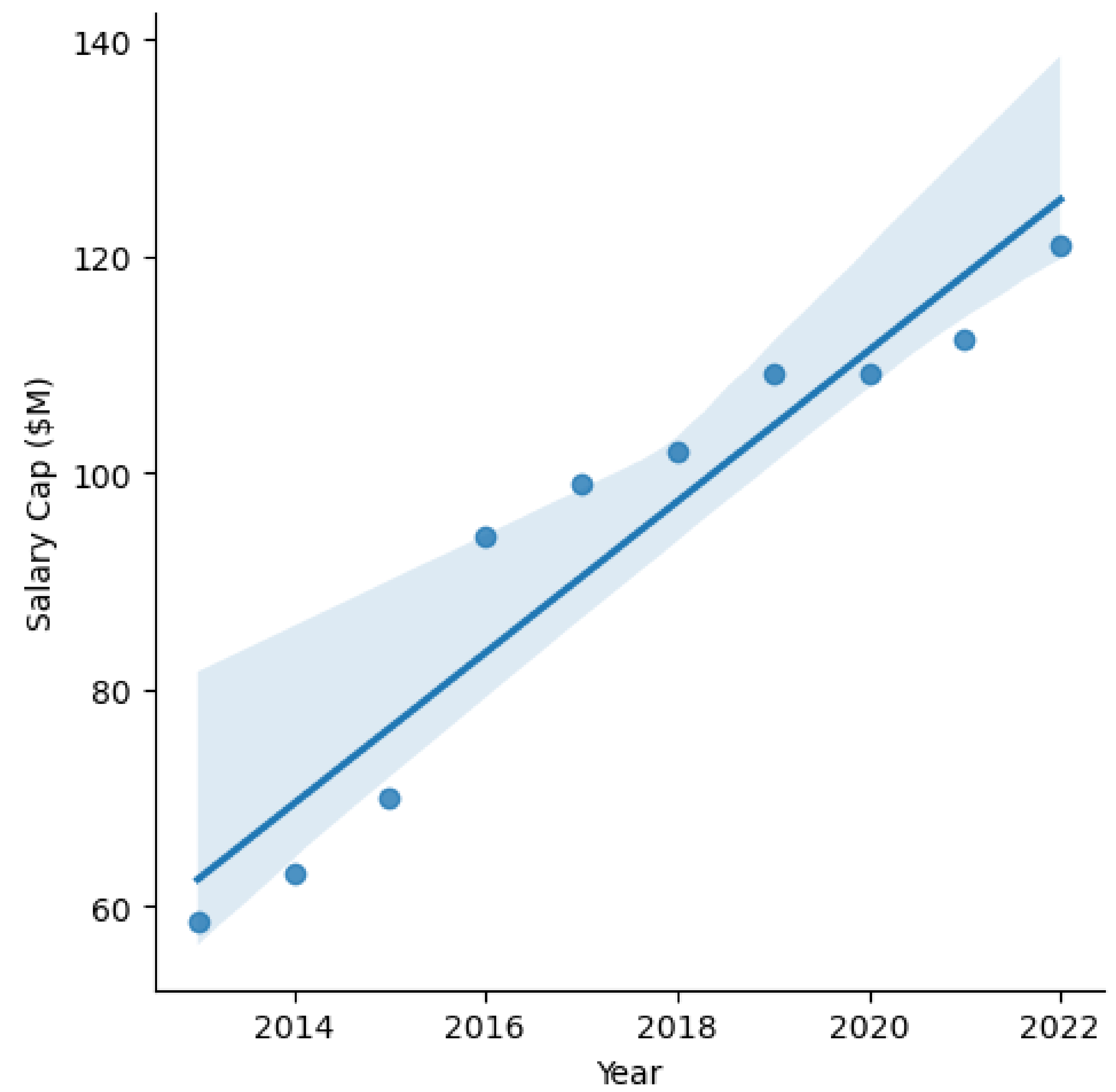


Decision Tree



The Regressor Decision Tree uses Squared Error and number of samples that fit the description of classes to come up with a predicted salary value

Limitations



Every year Salary Cap has **increased** meaning the results from the predictions will need to reflect growing cap space

Change in Nature of Game

More Three Point Shooting
Reliance on Star Power
Less Defensive Intensity
Faster Possessions

Dataset Limitations

Lack of Advanced Stats
Limited Player Sample Size
Untrackable Data not Realized
External Player Situations

Insights

01

Points per Game, Minutes Played, and Field Goals Attempted had the **highest importance** when determining salary size

02

Defensive Stats such as **Steals** and **Blocks** per Game do not play much importance when offering a contract, suggesting an offense heavy game-style

03

Front Offices recognize that **longer contract** terms go hand in hand with **salary size** and increase chances of player signings

04

Players do not need to excel at every skill; moreover, they must become a **specialist** at one trait. Being skilled at purely playmaking can earn a large contract

Business Application

Front Office

The front office, in charge of the team's structure, can **predict** which players they will need to **build** a championship roster

Fan Sales

Fans sales will depend on whether their favorite **star players** are still on the team or signed with **another**

Player Releases

If a star player signs with a new team, they often launch a **signature clothing brand** with a company like Nike or Adidas

Appendix

```python
from sklearn.tree import DecisionTreeRegressor
from sklearn.model_selection import train_test_split
from sklearn import metrics


X = player[['Years_for_Contract','GP','MP','FG%','3P%','2P%',
           'eFG%','FT','FTA','FT%','ORB','DRB','TRB','AST','STL','BLK','TOV','PF','PTS','PER','VORP']] # Features
y = player.Salary # Target variable

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=1)

regressor = DecisionTreeRegressor(random_state = 0)
regressor.fit(X, y)

y_pred = regressor.predict([[4,70,35,0.3,0.28,0.51,0.52,2.43,3.3,0.74,1.51,8,10,2.88,0.91,0.77,1.69,2.2,27,10,2.14]])
y_pred
```

```
/Users/vihaanhari/opt/anaconda3/lib/python3.9/site-packages/sklearn/base.py:450: UserWarning: X does not have valid feature name
s, but DecisionTreeRegressor was fitted with feature names
array([24.5])
```

Code Snippet for Decision Tree Regressor. Takes **all features** to make a tree that **predicts salary** for any combination of values of these features